Haeberli

Human Protein Data

WILEY-VCH

Human Protein Data

Edited by André Haeberli

WILEY-VCH

Weinheim · New York · Chichester · Brisbane · Singapore · Toronto

André Haeberli, Ph. D.
Professor of Biochemistry
Department of Medicine
University Hospital of Bern
Inselspital
CH-3010 Bern
Switzerland

Cover illustration: Molecular model of α_1-proteinase inhibitor (antitrypsin) based on that of intact ovalbumin. In the inhibitor the reactive centre loop (in red) will partially fold back into the A-sheet of the molecule (in blue) hinging on the glutamate at position 342 (first residue in red) that is mutated to a lysine in the common Z mutant. Figure prepared by Dr. C. J. Marshall. (See contribution *Alpha-1-Proteinase Inhibitor* by M. C. Owen and R. W. Carrell.)

Library of Congress Card No. applied for

A catalogue record for this book is available from the British library

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Human protein data / ed. by André Haeberli. - Weinheim ; New York ; Chichester ; Toronto ; Brisbane ; Singapore : WILEY-VCH
ISBN 3-527-30093-7

Pt. A - 1., bound ed. - 1998

Printing and Bookbinding: Strauss Offsetdruck GmbH, D-69509 Mörlenbach

Preface

The present bound edition of *Human Protein Data* includes the description of 314 human proteins. With this number of human proteins described the compendium is now becoming of broad interest and makes it a valuable handbook for a quick reference to the most pertinent data on one particular human protein.

Human Protein Data was started in a loose-leaf form with a 1st Installment in June 1992. The 1st Installment contained descriptions of 158 human proteins. It was followed by a 2nd Installment in November 1993 with 49 new proteins, and in June 1995 with the 3rd Installment containing descriptions of 56 new proteins.
With the 4th and 5th Installments revisions of 95 contributions from the 1st Installment and 41 from the 2nd Installment have been printed. In addition 26 contributions have been reviewed by the authors and been declared as still actual (indicated in the list of proteins by two asterisks). Thus the total number of proteins revised comes to 162 of the original 207 contributions collected in the first two Installments. The goal, already planned at the beginning, namely to revise the contributions every 4 to 5 years has therefore largely been realized. It is the continuous willingness of the authors to help realizing the compendium, that so many revisions of contributions could be collected in these two installments.
It is with the 6th Installment, published in March 1998, that the compendium has been expanded with the descriptions of 51 new human proteins.

It is part of the introduction to the 1st Installment (1992) of *Human Protein Data*, which gives the basic idea on how this project was started:

> Overwhelming progress in the last twenty years of molecular biology has made available the amino acid sequences of thousands of proteins of human, animal, bacterial and viral origin. The elucidation of the human genome will add an immense body of knowledge on DNA sequences and ultimately on amino acid sequences of proteins.
> High resolution two-dimensional gel electrophoresis has already localized over 3400 human proteins as products of the transcription of the genetic information. More than 1000 proteins have been identified and there is no doubt that in the next few years this number will increase. The three-dimensional structure of some 500 proteins is more or less well established. The information on protein structure and function is expanding in scope and depth. Superfamilies of proteins are being recognized with similar amino acid sequence motifs and structures as well as comparable function.
>
> This extraordinary accumulation of knowledge in molecular biology and possibly the shift of a vast number of scientists into this field has caused a change in the methods of biochemistry. A few years ago a newly discovered protein first had to be isolated. Its physico-chemical parameters and its amino acid sequence were then determined. Crystallization and the determination of the three-dimensional structure by X-ray crystallography was the ultimate goal in its characterization. Physiological and pathological characteristics were explored simultaneously. Today many proteins are only perceived by their function, and ultimately their DNA and amino acid sequence is determined by cDNA technology.
> It is the aim of the present compendium to bring together information from both fields and to present the data in a condensed and distinct way.
>
> The idea of creating *Human Protein Data* emerged during several years of my work as a biochemist in the Department of Internal Medicine at the Inselspital, Bern, Switzerland. I was frequently faced with questions on characteristics and biological functions of proteins. Sometimes the questions were easy to answer, but in other instances they involved hours of research in review and original articles or in databases.
> At this point I started collecting data on proteins relevant to my own field of research, blood coagulation. Since this particular field alone is growing fast, and the accumulation of data overwhelming, it soon became difficult to keep track of the data on coagulation proteins, and almost impossible on other human proteins.

Since my data on coagulation proteins was never complete and up-to-date, nor did I have the time to collect data on proteins of other fields, I started the search for an existing collection of protein data. Apart from the well-known and frequently used databases on amino acid and DNA sequences, which contain little information on physico-chemical parameters and even less on biological or pathological functions, there are only a few books containing all this information. One is the Handbook of *Biochemistry and Molecular Biology*, edited by G.D. Fasman, with much helpful information on physico-chemical data on proteins of human, animal and plant origin. Two other well-known books are *Molecular Biology of Human Proteins with special Reference to Plasma Proteins* (last edition 1966), by Schultze and Heremans, and a series of five volumes, *The Plasma Proteins*, edited by F.W. Putnam (last edition 1975 - 1987), both giving relatively short reviews on single proteins. Unfortunately, they all share the same disadvantage of not having been updated frequently enough.

At some point I realized that a concise collection of recent data on human proteins simply did not exist. I started working out the idea of what such a collection should look like. I asked one of my colleagues, Miha Furlan, to create with me a sample page on fibrinogen, a protein that we have both been interested in for many years.
With this sample page and a list of the names of the first one hundred proteins to be included in a starting volume I contacted Hans F. Ebel, Editor of chemistry and biochemistry books at VCH Publishers in Weinheim, Germany. After several lively discussions and evaluations of the project with VCH they agreed to publish *Human Protein Data* as a compendium.

I would like to express my thanks especially to my secretary Christine von Grünigen, who coordinates the secretarial and editorial work with more than 300 authors. I would also like to acknowledge Dr. Hans-Joachim Kraus, Karin Dembowsky and their staff at WILEY-VCH for their continuous support.

March 1998 André Haeberli

Introduction

The two-volume set includes the descriptions on 314 human proteins and thus all proteins described in the Installments 1-6 of the loose-leaf form of *Human Protein Data*.
Preceding the protein data sheets, the complete, most recent addresses of all authors including phone, FAX and E-mail address are given. Following the address list, there is a list of synonyms and abbreviations of all proteins covered in the six issues. This list will help the reader find his way to the protein he might be looking for.
A computer program at the end of the compendium contains part of the information and allows searches through the data, e.g., sorted lists of proteins with increasing molecular weight or increasing IEP.
The statement **„Unknown"** in a specific field means either that this field is **not applicable** or that the specific value has **not yet been determined**.

Remarks on the specific fields

Although the titles to the single fields are self-explanatory, a short summary as to the content of each field is provided here:

Synonyms	Many proteins have several synonyms and some of the proteins are still recognized by their older names. All common synonyms are mentioned and listed in front of the compendium.
Abbreviations	The most commonly used abbreviations are listed. All abbreviations are indexed in the synonym and abbreviation list.
Classifications	Classifications of the proteins such as electrophoretic mobility, protein classes and/or the EC No. for enzymes are given.
Description	A short and concise description of the protein is given. If the protein is synthesized in different organs, reference to identity or nonidentity is given.
Structure	The structures known from physico-chemical measurements, from electron microscopy, NMR or crystallography are described here. For many proteins comprehensible models are added at the end of the data sheets.
Molecular Weight	In general the molecular weight of the intact protein is given first. In addition the molecular weights of subunits, polypeptide chains or discrete fragments with biological signifcance are included. The methods used for the determination are given in parentheses.
Sedimentation Coeff.	The sedimentation coefficient is given in Svedberg (S) units.
Isoelectric Point	For a homogeneous protein the precise IEP is given; if the protein is heterogeneous (e.g. glycoproteins) the range or multiple IEP's are given. Special conditions for the determination of the IEP (e.g. urea) are given in parentheses.
Extinction Coeff.	The extinction coefficient is either given as E (280 nm, 1%, 1cm) or as molar extinction coefficient or both. The extinction coefficients for subunits or polypeptide chains are given as well.
Enzyme Activity	The mode of action of enzymes is described. The protein's enzyme activity is classified according to the rules of the IUB.
Coenzymes/Cofactors	Coenzymes and cofactors related to the protein are listed and their role is explained.

Substrates	The most important biological substrates for enzymes or inhibitors are listed. The most valuable chromogenic substrates used to determine the enzyme activity in vitro may be mentioned.
Inhibitors	The most important biological inhibitors and their mode of action are briefly described. The most relevant and useful synthetic inhibitors are listed.
Biological Functions	The concise biological function is described here.
Physiology/Pathology	The major physiological role of the protein is described. The consequences of reduced concentration, overproduction, or the absence of the protein are mentioned.
Degradation	The most important pathways and products of degradation and any biological activity associated with the degradation products are described.
Genetics/Abnormalities	The prominent genetic facts are noted and abnormal forms of the protein are described. The gene localization is indicated.
Half-Life	The biological half-life is given. This applies specially for circulating proteins.
Concentration	The concentration of the protein is given for different body fluids or even in cells. Ranges of normal concentrations are mentioned.
Isolation Method(s)	The most successful method(s) along with the best biological source for isolating the protein is briefly described.
Amino Acid Sequence	This paragraph in general does not include the whole amino acid sequence, since there are many databases containing this information. Instead, particularly prominent or physiologically important parts of the amino acid sequence and sequence homologies with other proteins or families of proteins are mentioned.
Disulfides/SH-Groups	The number of disulfide bonds and their location are indicated. The number of free SH-groups is mentioned.
General References	The most important references and if possible recent reviews on the protein are given.
Ref. for DNA/AA Sequences	References to DNA or amino acid sequences are given here. Citation of accession numbers and the respective databank are given where applicable.

List of Proteins, with Contributors
Installments 1 to 6

ß-N-acetyl-D-hexosaminidase	* Don J. Mahuran and Roderick B.C. Tse
	Urs Brodbeck
Acetylcholinesterase	* Kay Tanaka and Yasuyuki Ikeda
Acyl-CoA dehydrogenase	* Theodore Peters, Jr.
Albumin	* Bendicht Wermuth
Alcohol dehydrogenase	** Regina Pietruszko
Aldehyde dehydrogenase, E1, liver cytoplasm	** Regina Pietruszko
Aldehyde dehydrogenase, E2, liver mitochondria	* Regina Pietruszko and Gloria Kurys
Aldehyde dehydrogenase, E3, liver cytoplasm	** Regina Pietruszko
Aldehyde dehydrogenase, stomach cytoplasm	Russ Hille
Aldehyde oxidase	Oleg A. Barski and Kurt M. Bohren
Aldehyde reductase	Kurt M. Bohren and Oleg A. Barski
Aldose reductase	Harvey Rubin and Michael Plotnick
Alpha-1-antichymotrypsin	* Bo Åkerström
Alpha-1-microglobulin	Maurice C. Owen and Robin W. Carrell
Alpha-1-proteinase inhibitor	* Willi Jahnen-Dechent and Werner Müller-Esterl
Alpha2-HS-glycoprotein	* Lars Sottrup-Jensen
Alpha-2-macroglobulin	Dudley K. Strickland
Alpha-2-macroglobulin receptor	* Nobuo Aoki
Alpha-2-plasmin inhibitor	Michio Ogawa
Alpha-amylase	** Peter R. Clements and John J. Hopwood
Alpha-L-iduronidase	** Hans Sjöström and Ove Norén
Aminopetidase N	Greet C. Vanhoof and Filip Goossens
Aminopeptidase P	Wilma Wasco
Amyloid β-protein precursor	* Bert L. Vallee
Angiogenin	* James F. Riordan
Angiotensin-converting enzyme	William T. Tse and Samuel E. Lux
Ankyrin	** Jui-Yoa Chang
Antithrombin	* Gunther M. Fless and Angelo M. Scanu
Apolipoprotein(a)	** Lawrence Chan
Apolipoprotein A-I	** Lawrence Chan
Apolipoprotein A-II	* Richard B. Weinberg
Apolipoprotein A-IV	Lawrence Chan
Apolipoprotein B-48	** Lawrence Chan
Apolipoprotein B-100	Lawrence Chan
Apolipoprotein B messenger RNA editing protein	* Vassilis I. Zannis and Dimitris Kardassis
Apolipoprotein CI and CIV	* Eleni E. Zanni and Vassilis I. Zannis
Apolipoprotein CII	* Dimitris Kardassis and Eleni E. Zanni
Apolipoprotein CIII	Carlos López-Otín
Apolipoprotein D	* Stanley C. Rall, Jr.
Apolipoprotein E	Landon S. King and Peter Agre
Aquaporin-1	* Krzysztof Palczewski and Larry A. Donoso
Arrestin	Denis M. Grant and Urs A. Meyer
Arylamine N-acetyltransferase	* Erika Sandmeier and Philipp Christen
Aspartate aminotransferase	Hermann Passow
Band 3 protein	* Eric G. Berger and Martine Malissard
Beta 1,4-galactosyltransferase	Lennart E. Lögdberg
Beta-2-microglobulin	Mahin D. Maines
Biliverdin reductase	* Oksana Lockridge
Butyrylcholinesterase	* Peter J. Späth
C1 inhibitor	* Michael Loos and Franz Petry
C1q complement protein	* Gérard J. Arlaud
C1r complement protein	Reinhard Burger
C3 complement protein	

* revised 1997 ** revised 1997, but unchanged

* revised 1997 ** revised 1997, but unchanged

* revised 1997 ** revised 1997, but unchanged

* revised 1997 ** revised 1997, but unchanged

* revised 1997 ** revised 1997, but unchanged

* revised 1997 ** revised 1997, but unchanged

List of Contributors and Addresses
Installments 1 to 6

Komandoor E. Achyuthan
Noble Center for Biomedical Research
Oklahoma Medical Research Foundation
Oklahoma City, OK 73104
U.S.A.
Plasma factor XIII

Daniel Aeschlimann
Department of Medicine
University of Wisconsin
4459 Medical Science Center
1300 University Avenue
Madison, WI 53706
U.S.A.
Phone: +1 608 262 1375
FAX: +1 608 263 4969
Tissue transglutaminase

Peter Agre
Johns Hopkins University
Department of Biological Chemistry
725 N. Wolfe Street
Baltimore, MD 21205-2185
U.S.A.
Phone: +1 410 955 3154
FAX: +1 410 955 3149
E-mail: peteragre@mail.bs.jhu.edu
Aquaporin-1

Syed S. Ahmad
Thrombosis Research Center
Temple University School of Medicine
3400 North Broad Street
Philadelphia, PA 19140
U.S.A.
Phone: +1 215 221 4375
FAX: +1 215 221 2783
Factor IX

Bo Åkerström
Department of Medical
and Physiological Chemistry
Lund University
P.O. Box 94
S-22100 Lund
Sweden
Phone: +46 46 222 85 78
FAX: +46 46 222 40 22
E-mail: bo.akerstrom@medkem.lu.se
Alpha-1-microglobulin

Shizuo Akira
Department of Biochemistry
Hyogo College of Medicine
1-1 Mukogawa-cho
Nishinomiya, Hyogo 663
Japan
Phone: +81 0798 45 6355
FAX: +81 0798 46 3164
Interleukin 6

Sefik S. Alkan
Ciba Geigy Ltd.
Department of Allergy/Immunology
R-1056-407
CH-4002 Basel
Switzerland
Phone: +41 61 697 5384
FAX: +41 61 697 3968
Interferon gamma

Fabrice Allain
Laboratoire de Biochimie
Université des Sciences et Technologies de Lille
F-59655 Villeneuve d'Ascq Cédex
France
Phone: +33 20 434 155
FAX: +33 20 436 555
Cyclophilin-20

Lena Avedissian
Ross 1025
Johns Hopkins University School of Medicine
720 Rutland Avenue
Baltimore, MD 21205
U.S.A.
Phone: +1 410 955 5454
FAX: +1 410 955 0185
Erythropoietin receptor

José M. Andreu
Centro de Investigaciones Biologicas C.S.I.C
Velazquez 144
E-28006 Madrid
Spain
Phone: +34 1 585 4380
FAX: +34 1 562 7518
Tubulin

Nobuo Aoki
Ina Red Cross Blood Center
Ina-Machi, Kitaadachi-Gun
Saitama-Ken 362
Japan
Phone: +81 48 722 1211
FAX: +81 48 721 6404
E-mail: aoki.medl.1@med.tmd.ac.jp
Alpha-2-plasmin inhibitor

Gérard J. Arlaud
Laboratoire d'Enzymologie Moléculaire
Institut de Biologie Structurale
41, Avenue des Martyrs
F-38027 Grenoble Cédex
France
Phone: +33 76 884 981
FAX: +33 76 885 494
E-mail: arlaud@ibs.fr
C1r complement protein

John P. Atkinson
Washington University School of Medicine
Division of Rheumatology
Department of Internal Medicine
Campus Box 8045
660 South Euclid
St. Louis, MO 63110
U.S.A.
Phone: +1 314 362 8392
FAX: +1 314 362 1366
E-mail: jatkinso@imgate.wustl.edu
CD46

Angelo M. Azzi
Institut für Biochemie und Molekularbiologie
Universität Bern
Bühlstrasse 28
CH-3012 Bern
Switzerland
Phone: +41 31 631 41 31
FAX: +41 31 631 37 37
Cytochrome C oxidase

Marco Baggiolini
Theodor-Kocher-Institut
University of Bern
Freiestrasse 1
CH-3000 Bern 9
Switzerland
Phone: +41 31 631 41 43
FAX: +41 31 631 37 99
E-mail: baggiolini@tki.unibe.ch
Interleukin 8

Frank A. Baglia
Thrombosis Research Center
Temple University School of Medicine
3400 North Broad Street
Philadelphia, PA 19140
U.S.A.
Phone: +1 215 221 4375
FAX: +1 215 221 2783
Factor XI

Teaster T. Baird
Department of Biochemistry
Duke University Medical Center
Box 3711
Durham, NC 27710
U.S.A.
Phone: +1 919 684 8976
FAX: +1 919 684 4050
Carbonic anhydrase II

Milagros Balbín
Departemento de Biologia Funcional
Universidad de Oviedo
E-33006 Oviedo
Spain
Phone: +34 85 104 202
FAX: +34 85 103 534
Cystatin D

Oleg A. Barski
Baylor College of Medicine
The Diabetes Research Center
Texas Children's Hospital
1102 Bates, Suite 830, MC 3-2353
Houston, TX 77030-2399
U.S.A.
Phone: +1 713 770 3771
FAX: +1 713 770 3766
E-mail: obarski@bcm.tmc.edu
Aldose Reductase
Aldehyde Reductase

Gerhard Baumann
Center for Endocrinology and Molecular Medicine
Northwestern University Medical School
303 East Chicago Avenue
Chicago, IL 60611
U.S.A.
Phone: +1 312 503 4128
FAX: +1 312 908 9032
E-mail: gbaumann@nwu.edu
Somatotropin
Somatotropin-receptor & somatotropin binding protein

Robert C. Baxter
Kolling Institute of Medical Research
Royal North Shore Hospital
St Leonards, NSW 2065
Australia
Phone: +61 2 9926 8486
FAX: +61 2 9926 8484
Email: robaxter@med.su.oz.au
Insulin-like growth factor I
Insulin-like growth factor II
Insulin-like growth factor binding protein-1
Insulin-like growth factor binding protein-2
Insulin-like growth factor binding protein-3
Insulin-like growth factor binding protein-4
Insulin-like growth factor binding protein-5
Insulin-like growth factor binding protein-6

Earl P. Benditt
Department of Pathology SJ-60
University of Washington
Seattle, WA 98195
U.S.A.
Phone: +1 206 543 7472
FAX: +1 206 685 3662 or 543 3644
Serum amyloid A

Eric G. Berger
Physiologisches Institut
Universität Zürich
Winterthurerstr. 190
CH-8057 Zürich
Switzerland
Phone: +41 1 257 50 70
FAX: +41 1 364 05 64
E-mail: egberger@physiol.unizh.ch
Beta 1,4-galactosyltransferase

Barbara Bertocci
Faculté de Médicine Necker U.I. 373
156, Rue de Vaugirard
F-75730 Paris Cédex 15
France
Phone: +33 1 4061 5681
FAX: +33 1 4061 5590
Catechol-o-methyltransferase

Julie Bielicki
Lysosomal Diseases Research Unit
Department of Chemical Pathology
Women's and Children's Hospital
North Adelaide, SA 5006
Australia
Phone: +61 8 204 7293
FAX: +61 8 204 7100
Glucuronate-2-sulphatase
Iduronate-2-sulphatase
N-Acetylgalactosamine-4-sulphatase
N-Acetylgalactosamine-6-sulphatase

Francesco Blasi
Unita di Genetica Molecolare
HSR Ospedale S. Raffaele DIBIT
Via Olgettina 60
I-20132 Milano
Italy
Phone: +39 2 2643 4832
FAX: +39 2 2643 4844
E-mail: blasif@dibit.hsr.it
Urokinase plasminogen activator receptor

Kurt M. Bohren
Baylor College of Medicine
The Diabetes Research Center
Texas Children's Hospital
1102 Bates, Suite 830, MC 3-2353
Houston, TX 77030-2399
U.S.A.
Phone: +1 713 770 3759/70
FAX: +1 713 770 3766
E-mail: kbohren@bcm.tmc.edu
Aldose Reductase
Aldehyde Reductase

Raymond P. Boot-Handford
Dept. Biochemistry and Molecular Biology Sciences
University of Manchester
Stopford Building
Oxford Road
Manchester, M13 9PT
United Kingdom
Phone: +44 61 275 5117
FAX: +44 61 275 5082
Collagen type X

Walter Born
Research Laboratory for Calcium Metabolism
Depts. of Orthopaedic Surgery and Medicine
Forchstrasse 340
CH-8008 Zürich
Switzerland
Phone: +41 1 386 1650
FAX: +41 1 386 1652
Calcitonin and procalcitonin
Calcitonin gene-related peptide

Christoph Borner
Present address unknown
Protein kinase C

Sylviane Boularand
Laboratoire Neurobiologie Cellulaire et Moléculaire
C.N.R.S.
1, Avenue de la Terasse
F-91198 Gif sur Yvette
France
Phone: +33 1 6 982 3636
FAX: +33 1 6 982 3580
Tryptophan hydroxylase

Stephen O. Brennan
Molecular Pathology Laboratory
Christchurch School of Medicine
Christchurch Hospital
Christchurch
New Zealand
Phone: +64 3 364 0549
FAX: +64 3 364 0750
Furin

Ad Brinkman
Vlieland Hoever 63
NL-3137 GV Vlaardingen
The Netherlands
Phone: +31 10 475 1069
Insulin-like growth factor binding protein-1

Urs Brodbeck
Institut für Biochemie und Molekularbiologie
Bühlstrasse 28
P.O. Box 98
CH-3000 Bern 9
Switzerland
Phone: +41 31 631 41 11
FAX: +41 31 631 37 37
E-mail: brodbeck@mci.unibe.ch
Acetylcholinesterase

George J. Broze, Jr.
Division Hematology/Oncology
Medical Center
Jewish Hospital at Washington University
216 South Kingshighway Blvd.
St.Louis, MO 63110
U.S.A.
Phone: +1 314 362 8811
FAX: +1 314 362 8813
Tissue factor pathway inhibitor

Peter Bruckner
Institut für Physiologische Chemie
Universität Münster
Waldeyerstrasse 15
D-48149 Münster
Germany
Phone: +49 251 835 591
FAX: +49 251 835 596
E-mail: pibi@uni-muenster.dc
Collagen type IX

H. Franklin Bunn
Room 919 Thorn Bldg.
Brigham and Women's Hospital
Shattuck St.
Boston, MA 02115
U.S.A.
Phone: +1 617 732 5841
FAX: +1 617 732 6088
Hemoglobin

Ann Burchell
Ninewells Hospital and Medical School
Department of Obstetrics and Gynaecology
University of Dundee
Dundee DD1 9SY
United Kingdom
Phone: +44 1382 632 445
FAX: +44 1382 566 617
E-mail: a.burchell@dundee.ac.uk
Glucose-6-phosphatase
Glucose transport protein 1
Glucose transport protein 2
Glucose transport protein 3
Glucose transport protein 4
Glucose transport protein 5
Glucose transport protein 7

Brian Burchell
Ninewells Hospital and Medical School
Department of Obstetrics and Gynaecology
University of Dundee
Dundee DD1 9SY
United Kingdom
Phone: +44 1382 660 111
FAX: +44 1382 644 620
UDP-glucuronosyltransferases

Reinhard Burger
Robert-Koch-Institut
Bundesgesundheitsamt
Nordufer 20
D-13353 Berlin
Germany
Phone: +49 30 4547 2356
FAX: +49 30 4547 2602
C3 complement protein

R. Duncan Campbell
MRC Immunochemistry Unit
Department of Biochemistry
University of Oxford
South Parks Road
Oxford OX1 3QU
United Kingdom
Phone: +44 865 275 354
FAX: +44 865 275 729
C4 complement protein

Ernesto Carafoli
Laboratorium für Biochemie
ETH-Zentrum
Universitätsstrasse 16
CH-8092 Zürich
Switzerland
Phone: +41 1 256 3011
FAX: +41 1 252 6323
Calcium adenosinetriphosphatase (plasma membrane)

Robin W. Carrell
MRC Center
Department of Haematology
University of Cambridge
Hills Road
Cambridge CB2 2QH
United Kingdom
Phone: +44 223 336 788
FAX: +44 223 336 827
Alpha-1-proteinase inhibitor
Cortisol binding globulin

Giuseppe Castaldo
Dip. di Biochimica e Biotecnologie Medica
Il Facoltà di Medicina e Chirurgia
Università degli Studi di Napoli "Federico II"
Via Sergio Pansini 5
I-80131 Napoli
Italy
Phone: +39 81 746 3135
FAX: +39 81 746 3650
Gamma-glutamyltransferase

George H. Caughey
Cardiovascular Research Institute
University of California San Francisco
San Francisco, CA 94143-0911
U.S.A.
Phone: +1 415 476 9920
FAX: +1 415 476 9749
E-mail: ghc@itsa.ucsf.edu
Tryptase
Chymase

Joseph M. Chalovich
East Carolina University
School of Medicine
Dept. of Biochemistry
Greenville, NC 27858-4354
U.S.A.
Phone: +1 919 816 2675
FAX: +1 919 816 3383
Caldesmon

Lawrence Chan
Department of Cell Biology
Baylor College of Medicine
One Baylor Plaza
Houston, TX 77030
U.S.A.
Phone: +1 713 798 4478
FAX: +1 713 798 8764
E-mail: lchan@bcm.tmc.edu
Apolipoprotein A-I
Apolipoprotein A-II
Apolipoprotein B-100
Apolipoprotein B messenger RNA edit. prot.
Apolipoprotein B-48

Ronald E. Chance
Lilly Research Laboratories
The Lilly Corporate Center
Mail Drop 1543
Indianapolis, IN 46285
U.S.A.
Phone: +1 317 276 4233
FAX: +1 317 276 9159
Insulin and proinsulin

Jui-Yoa Chang
Ciba-Geigy Ltd.
K-121.1.04
CH-4002 Basel
Switzerland
Phone: +41 61 696 1111
FAX: +41 61 696 8313
Antithrombin

Julie Chao
Department of Biochemistry and
Molecular Biology
Medical University of South Carolina
171 Ashley Avenue
Charleston, SC 29425-2211
U.S.A.
Phone. +1 803 792 4321
FAX: +1 803 792 1627
E-mail: chaoj@musc.edu
Kallistatin
Prostasin
Tissue Kallikrein

Lee Chao
Department of Biochemistry and
Molecular Biology
Medical University of South Carolina
171 Ashley Avenue
Charleston, SC 29425-2211
U.S.A.
Phone. +1 803 792 4321
FAX: +1 803 792 1627
E-mail: chaol@musc.edu
Kallistatin
Prostasin
Tissue Kallikrein

Wai Yiu Cheung
St. Jude Children's Research Hospital
Department of Biochemistry
Memphis, TN 38101
U.S.A.
Phone: +1 901 522 0563
FAX: +1 901 525 8025
Calmodulin

Arcadio Chonn
Present address unknown
Clusterin

Philipp Christen
Biochemisches Institut der Universität Zürich
Winterthurerstr.190
CH-8057 Zürich
Switzerland
Phone: +41 1 257 5511(60)
FAX: +41 1 363 7947
E-mail: chris@biocfebs.unizh.ch
Aspartate aminotransferase

Thomas Ciardelli
Dept. of Pharmacology and Toxicology
Dartmouth Medical School
7650 Remsen Building
Hanover, NH 03756-3835
U.S.A.
Phone: +1 603 650 1826
FAX: +1 603 650 1129
Interleukin-2

David M. Clarke
Department of Medicine
University of Toronto
Medical Sciences Building, Room 7342
Toronto, Ontario M5S 1A8
Canada
Phone: +1 416 978 1105
FAX: +1 416 978 1105
P-glycoprotein

Peter R. Clements
Lysosomal Diseases Research Unit
Women's and Children's Hospital
North Adelaide, SA 5006
Australia
Phone: +61 8 204 7293
FAX: +61 8 204 7106
α-L-iduronidase

Kenneth J. Clemetson
Theodor-Kocher-Institut
Freiestr.1
CH-3012 Bern
Switzerland
Phone: +41 31 631 41 48
FAX: +41 31 631 37 99
Platelet basic protein
Platelet membrane glycoprotein Ibα (GPIbα)
Platelet membrane glycoprotein Ibβ (GPIbβ)
Platelet membrane glycoprotein V (GPV)
Platelet membrane glycoprotein IX (GPIX)

Inge Clemmensen
Department of Clinical Biochemistry
Statens Seruminstitut
Copenhagen
Denmark
Tetranectin

Désiré J. Collen
Center for Molecular and Vascular Biology
Bldg. Teaching and Research
Campus Gasthuisberg University of Leuven
Herestraat 49
B-3000 Leuven
Belgium
Phone: +32 16 215 775
FAX: +32 16 215 990
Tissue-type plasminogen activator

Nancy E. Cooke
Department of Medicine and Genetics
University of Pennsylvania
700 Clinical Research Building
422 Curie Boulevard
Philadelphia, PA 19104-6144
U.S.A.
Phone: +1 215 898 4425
FAX: +1 215 898 0189
E-mail. cooke@hmivax.humgen.upenn.edu
Vitamin D binding protein
Cysteine-rich protein

Paola Costanzo
Facoltà di Medicina e Chirurgia
Dipartimento di Biochimica e Biotecnologie Mediche
Università degli Studi di Napoli "Federico II"
Via Sergio Pansini 5
I-80131 Napoli
Italy
Phone: +39 81 746 3147
FAX: +39 81 746 3650
Fructose-1,6-bisphosphate aldolase

Robert R. Crichton
Department de Chimie
Biochimie (BIO)
Université Catholique de Louvain
B-1348 Louvain-la-Neuve
Belgium
Phone: +32 10472 794
FAX: +32 10472 836
Ferritin

Björn Dahlbäck
University Hospital
Department of Clinical Chemistry
S-20502 Malmö
Sweden
Phone: +46 40 33 1501
FAX: +46 40 33 7044
C4b-binding protein
Protein C
Protein S

Mose Da Prada
Hoffmann-La Roche PRPN
Grenzacherstrasse
CH-4002 Basel
Switzerland
Phone: +41 61 688 3071
FAX: +41 61 688 4484
Catechol-O-methyltransferase

Lisa de Conciliis
Facoltà di Medicina e Chirurgia
Dipartimento di Biochimica e Biotecnologie Mediche
Università degli Studi di Napoli "Federico II"
Via Sergio Pansini 5
I-80131 Napoli
Italy
Phone: +39 81 746 3135
FAX: +39 81 746 3650
Glyceraldehyde-3-phosphate dehydrogenase

Jean Delaunay
INSERUM U299
Centre Hospitalier de Bicêtre
78, rue du Général Leclerc
F-94275 Le Kremlin Bicêtre
France
Phone: +33 1 46 70 89 89
FAX: +33 1 46 70 64 46
Spectrin

Ingrid A. De Meester
University of Antwerp
Department of Pharmaceutical Sciences
Universiteitsplein 1
B-2610 Wilrijk
Belgium
Phone: +32 38 20 2727
FAX: +32 38 20 2745
E-mail: idmeest@uia.ua.ac.be
CD26

Mandeep Dhadly
Lipid Research Laboratory
VA Wadsworth Medical Center
Building 113, Room 312
Los Angeles, CA 90073
U.S.A.
Phone: +1 310 824 4395
FAX: +1 310 478 4538
Hormone-sensitive lipase

Richard G. DiScipio
La Jolla Institute for Exp. Medicine
11077 N.Torrey Pines Rd.
La Jolla, CA 92037
U.S.A.
Phone: +1 619 587 8788 ext 249 or 375
FAX: +1 619 458 9072
E-mail: richard@ljiem.org
C7 complement protein

William A. Dittman
Durham V.A. Medical Center
P.O. Box 111-G
508 Fulton Street
Durham, NC 27705
U.S.A.
Phone: +1 919 286 6944
FAX: +1 919 286 6823
Thrombomodulin

Larry A. Donoso
Wills Eye Hospital
900 Walnut Street
Philadelphia, PA 19107-5598
U.S.A.
Phone: +1 215 928 3268
FAX: +1 215 592 4628
Arrestin

Maria T. Dours-Zimmermann
Institut für Klinische Pathologie
Universitätsspital
Schmelzbergstrasse 12
CH-8091 Zürich
Switzerland
Phone: +41 1 255 39 45
FAX: +41 1 255 45 08
Versican

Ann D. Dunn
Health Sciences Center
University of Virginia
P.O. Box 511
Charlottesville, VA 22908
U.S.A.
Phone: +1 804 924 5929
FAX: +1 804 296 9275
Thyroglobulin

John T. Dunn
Division Endocrinology
University of Virginia
P.O. Box 511
Charlottesville, VA 22908
U.S.A.
Phone: +1 804 924 5929
FAX: +1 804 296 9275
Thyroglobulin

Paul F. Edgar
MRC Center
Department of Haematology
University of Cambridge
Hills Road
Cambridge CB2 2QH
United Kingdom
Phone: +44 223 336 788
FAX: +44 223 336 827
Cortisol binding globulin

Randy C. Eisensmith
Department of Cell Biology
Room T-705A
Baylor College of Medicine
One Baylor Plaza
Houston, TX 77030
U.S.A.
Phone: +1 713 798 4530
FAX: +1 713 798 8019
E-mail: randye@bcm.tmc.edu
Phenylalanine hydroxylase

Joyce A. Eldering
Medizinische Poliklinik
Freiburgstrasse 3
CH-3010 Bern
Switzerland
Phone: +41 31 632 88 90
FAX: +41 31 381 34 12
Lipocortin I

Caroline A. Enns
Department of Cell Biology L215
Oregon Health Sciences University
3181 SW Sam Jackson Park Rd
Portland, OR 97201
U.S.A.
Phone: +1 503 494-5845
FAX: +1 503 494 4253
E-mail: ennsca@ohsu.edu
Transferrin receptor

Nils Eriksen
Department of Pathology Box 357335
University of Washington
Seattle, WA 98195-7335
U.S.A.
Phone: +1 206 543 7499
FAX: +1 206 685 3662
Serum amyloid A

Doriano Fabbro
Abteilung Pharma Forschung
K-125.405
Oncology-Virology
Ciba-Geigy
CH-4002 Basel
Switzerland
Phone: +41 61 696 76 81
FAX: +41 61 696 38 35
E-mail: doriano.fabbro@chbs.mhs.ciba.com
Protein kinase C

David L. Farley
Ciba-Geigy Ltd.
K 125.902
CH-4002 Basel
Switzerland
Phone: +41 61 696 5864
FAX: +41 61 696 7228
e-mail: farley@fmi.ch
Leukocyte cathepsin G
Leukocyte elastase

Anna Maria Ferraris
IST - Istituto di Oncologia Clinice e Sperimentale
University of Genova
Viale Benedetto XV,10
I-16132 Genova
Italy
Phone: +39 10 353 4933
FAX: +39 10 509 052
Catalase
Glucose-6-phosphate dehydrogenase

Carol A. Fierke
Department of Biochemistry
Duke Uiversity Medical Center
Box 3711
Durham, NC 27710
U.S.A.
Phone: +1 919 684 8976
FAX: +1 919 684 4050
Carbonic anhydrase II

Catherine G. Figarella
Groupe de Recherche sur les Glandes Exocrines
Faculté de Médécine
27, Boulevard Jean Moulin
F-13385 Marseille Cedex 05
France
Phone: +33 91 78 82 60
FAX: +33 91 78 68 95
Trypsin(ogen) 1 and Trypsin(ogen) 2

Jan A. Fischer
Research Laboratory for Calcium Metabolism
Departments of Orthopaedic Surgery and Medicine
Forchstrasse 340
CH-8008 Zürich
Switzerland
Phone: +41 1 386 16 51
FAX: +41 1 386 16 52
Calcitonin and procalcitonin
Calcitonin gene-related peptide
Parathyroid hormone
Parathyroid hormone-related protein

Gunther M. Fless
University of Chicago
5841 S. Maryland
Mail Code MC 5041
Chicago, IL 60637
U.S.A.
Phone: +1 312 702 0369
FAX: +1 312 702 4534
Apolipoprotein(a)

Giuliana Fortunato
Facoltà di Medicina e Chirurgia
Dipartimento di Biochimica e Biotecnologie Mediche
Università degli Studi di Napoli "Federico II"
Via Sergio Pansini 5
I-80131 Napoli
Italy
Phone: +39 81 746 3135
FAX: +39 81 746 3650
Creatine kinase
Lactate dehydrogenase
Superoxide dismutase

Craig Freeman
Lysosomal Diseases Research Unit
Department of Chemical Pathology
Women's and Children's Hospital
North Adelaide, SA 5006
Australia
Phone: +61 8 204 7101/7293
FAX: +61 8 204 7100
Glucuronate-2-sulphatase

Brigitte M. Frey
Abteilung für Nephrologie
Inselspital
CH-3010 Bern
Switzerland
Phone: +41 31 632 94 76
FAX: +41 31 632 94 44
Lipocortin I

Miha Furlan
Inselspital Bern
Central Hematology Laboratory
University of Bern
CH-3010 Bern
Switzerland
Phone: +41 31 632 90 23
FAX: +41 31 632 93 66
Fibrinogen
von Willebrand factor

Gian Franco Gaetani
Ematologia Oncologica
Dipartimento di Oncologia
Clinica e Sperimentale
Università di Genova
Largo Rosanna Benzi, 10
I-16132 Genova
Italy
Phone: +39 10 522 0580
FAX. +39 10 509 052
E-mail: gaetani@mbox.ulisse.it
Catalase
Glucose-6-phosphate dehydrogenase

John A. Galloway
The Lilly Corporate Center
Lilly Research Laboratories
Mail Drop 2128
Indianapolis, IN 46285
U.S.A.
Phone: +1 317 276 4813
FAX: +1 317 276 6766
Insulin and proinsulin

Chandramowli Ganesch
Molecular Biophysics Unit
Indian Institute of Science
Bangalore - 560 012
India
Phone: +91 80 309 2612
FAX: +91 80 334 1683
E-mail: ganesh@mbu.iisc.ernet.in
Myoglobin

Gianni Garotta
Human Genome Sciences, Inc.
9410 Key West Avenue
Rockville, MD 20850
U.S.A.
Phone: +1 301 309 8504
FAX: +1 301 294 4843
Interferon gamma receptor

Wolfgang Gebhard
Klinikum Grosshadern
HNO-Forschungslabor
Ludwig-Maximilians-Universität München
Marchioninistrasse 15
D-81377 München
Germany
Phone: +49 8970 953 890
FAX: +49 8970 044 418
Inter-alpha-inhibitor

Hans-Hermann Gerdes
Institut für Neurobiologie
Universität Heidelberg
Im Neuenheimer Feld 364
D-69120 Heidelberg
Germany
Phone: +49 6221 56 38 17
FAX: +49 6221 56 37 00
Chromogranin A
Chromogranin B
Secretogranin II

Alison L. Gibb
University Department of Pharmacology
Mansfield Road
Oxford OX1 3QT
United Kingdom
Phone: +44 865 271597
FAX: +44 865 271853
Complement receptor type I

Vincent Goffin
INSERM Unité 344
Endocrinologie Moléculaire
Faculté de Médecine Necker
156, rue de Vaugirard
F-75730 Paris Cedex 15
France
Phone: +33 1 40 61 53 10
FAX: +33 1 43 06 04 43
E-mail: goffin@necker.fr
Prolactin
Prolactin receptor

Filip J. Goossens
University of Antwerp
Department of Pharmaceutical Sciences
Universiteitsplein 1
B-2610 Wilrijk
Belgium
Phone: +32 3 820 2727
FAX: +32 3 820 2745
E-mail: fgooss@uia.ua.ac.be
Prolyl oligopeptidase

Gregory A. Grabowski
Division of Human Genetics
Childrens' Hospital Medical Center
University of Cincinnati
3333 Burnet Avenue
Cincinnati, OH 45229-3039
U.S.A.
Phone: +1 513 559 7290
FAX: +1 513 559 7297
E-mail: grabgo@chmcc.org
Glucocerebrosidase

Denis M. Grant
Division of Clin. Pharmacol. & Toxicol.
Hospital for Sick Children
555 University Avenue
Toronto, Ontario M5G IX8
Canada
Phone: +1 416 813 5175
FAX: +1 416 813 7562
E-mail: grant@sickkids.on.ca
Arylamine N-acetyltransferase

Charles S. Greenberg
Duke University Medical Center
Box 2603
Durham, NC 27710
U.S.A.
Phone: +1 919 684 6703
FAX: +1 919 684 4670
E-mail: green032@mc.duke.edu
Plasma factor XIII

Anders O. Grubb
Department of Clinical Chemistry
Lund University Hospital
S-22185 Lund
Sweden
Phone: +46 46 173 964
FAX: +46 46 189 114
E-mail: anders.grubb@klinkem.lu.se
Cystatin C
Cystatin D

Markus G. Grütter
Novartis Pharma AG
Pharmaceutical Division
CH-4002 Basel
Switzerland
Transforming growth factor-beta 3 (1,2)

Danilo Guerini
Laboratorium für Biochemie
ETH-Zentrum
Universitätsstrasse 16
CH-8092 Zürich
Switzerland
Phone: +41 1 256 3011
FAX: +41 1 252 6323
Calcium adenosinetriphosphatase (plasma membrane)

Arabinda Guha
Mount Sinai School of Medicine
One Gustave Levy Place
P.O Box 1269
New York, NY 10029
U.S.A.
Phone: +1 212 241 6083
FAX: +1 212 860 7032
Tissue factor

Odette M. Guy-Crotte
Groupe de Recherche sur les Glandes Exocrines
Faculté de Médécine
27, Boulevard Jean Moulin
F-13385 Marseille Cedex 05
France
Phone: +33 91 78 82 60
FAX: +33 91 78 68 95
Trypsin(ogen)1 and Trypsin(ogen) 2

John G. Haddad
Department of Medicine
University of Pennsylvania
611 Clinical Research Building
422 Curie Boulevard
Philadelphia, PA 19104-6144
U.S.A.
Phone: +1 215 898 5028
Vitamin D binding protein

Robert E. Handschumacher
Prof. Emeritus & Sr. Scientist
Department of Pharmacology
School of Medicine
Sterling Hall of Medicine
333 Cedar Street
P.O. Box 208066
New Haven, CT 06520-8066
U.S.A.
Phone: +1 203 785 4385
FAX: +1 203 785 7670
Cyclophilin-18
Cyclophilin-40

Paul Hargrave
Department of Ophthalmology
University of Florida
Gainesville, FL 32610-0284
U.S.A.
Phone: +1 352 392 9098
FAX: +1 352 392 0478
E-mail: Hargrave@eyel.eye.ufl.edu
Opsin

Yoshitaka Hayashi
The University of Chicago
MC 3090
5841 South Maryland Avenue
Chicago, IL 60637
U.S.A.
Phone: +1 312 702 6939
FAX: +1 312 702 6940
Thyroxine-binding globulin

Catherine P.M. Hayward
McMaster University
Department of Pathology, HSC2N32
1200 Main Street West
Hamilton, Ontario L8N 3Z5
Canada
Phone: +1 905 521 2100
FAX: +1 905 521 2338
E-mail: haywrdc@fhs.mcmaster.ca
Multimerin

Gerhard Heinrich
Biomolecular Medicine
Evans 604
University Hospital
88 E. Newton Street
Boston, MA 02118
U.S.A.
Phone: +1 617 638 6010 / 6013
FAX: +1 617 638 6009
Nerve growth factor

Benjamin Hemmens
Karl-Franzens Universität Graz
Institut für Pharmakologie und Toxikologie
Universiätsplatz 2
A-8010 Graz
Austria
Phone: +43 316 380 8858
FAX: +43 316 380 9890
E-mail: hemmens@kfunigraz.ac.at
Nitric oxide synthase

Carl-Henrik Heldin
Ludwig Institute for Cancer Research
Biomedical Center
P.O. Box 595
S-75124 Uppsala
Sweden
Phone: +46 18 174 146
FAX: +46 18 506 867
E-mail: c-h.heldin@licr.uu.se
Platelet-derived growth factor

Dirk F. Hendriks
University of Antwerp
Department of Pharmaceutical Sciences
Universiteitsplein 1
B-2610 Antwerpen (Wilrijk)
Belgium
Phone: +32 3 820 27 27
FAX: +32 3 820 27 45
E-mail: hendriks@uia.ua.ac.be
Carboxypeptidase N
Procarboxypeptidase U

Christoph H. Heusser
Pharmaceuticals Division
Research Department
R 1056.4.09
CH-4002 Basel
Switzerland
Phone: +41 61 697 54 89
FAX: +41 61 697 88 56
CD 23
Interleukin 4

John S. Hill
Lipid Research
VA Wadsworth Medical Center
Bldg. 113, Room 312
Los Angeles, CA 90073
U.S.A.
Phone: +1 310 824 4395
FAX: +1 310 478 4538
Hepatic lipase

Andreas Hillarp
Department of Clinical Chemistry
University of Lund
University Hospital
S-20502 Malmö
Sweden
Phone: +46 40 333155
FAX: +46 40 337044
E-mail: andreas.hillarp@klkemi.mas.lu.se
C4b-binding protein
Protein S

Russ Hille
Department of Medical Biochemistry
333 Hamilton Hall
1645 Neil Avenue
Ohio State University
Columbus, OH 43210-1218
U.S.A.
Phone: +1 614 292 3545
FAX: +1 614 292 4118
E-mail: hille.1@osu.edu
Xanthine oxidoreductase
Aldehyde oxidase

Patricia M. Hinkle
Department of Pharmacology and Physiology
University of Rochester Medical Center
601 Elmwood Avenue, Box 711
Rochester, NY 14642
U.S.A.
Phone: +1 716 275 4933
FAX: +1 716 461 0397
E-mail: hinklep@pharmacol.rochester.edu
TRH Receptor

Erich Hochuli
Hoffmann-La Roche AG
PTFB, Bldg. 66/508
CH-4000 Basel
Switzerland
Phone: +41 688 34 00
FAX: +41 688 14 48
Interferon alpha

John J. Hopwood
Lysosomal Diseases Research Unit
Department of Chemical Pathology
Women's and Children's Hospital
North Adelaide, SA 5006
Australia
Phone: + 61 8 2047101 / 7293
FAX: + 61 8 2047100
α-L-iduronidase
Iduronate-2-sulphatase
Glucuronate-2-sulphatase
N-Acetylgalactosamine-4-sulphatase
N-Acetylgalactosamine-6-sulphatase

Leon W. Hoyer
American Red Cross Blood Services
Holland Laboratory
15601 Crabbs Branch Way
Rockville, MD 20855
U.S.A.
Phone: +1 301 738 0623
FAX: +1 301 738 0553
Factor VIII

Wieland B. Huttner
Institut für Neurobiologie
Universität Heidelberg
Im Neuenheimer Feld 364
D-69120 Heidelberg
Germany
Phone: +49 6221 56 82 18
FAX: +49 6221 56 37 00
Chromogranin A
Chromogranin B
Secretogranin II

René E. Humbel
Kirchrain 8
CH-8816 Hirzel
Switzerland
Phone: +41 1 729 95 41
Insulin-like growth factor I
Insulin-like growth factor II

James N. Ihle
St. Jude Children's Research Hospital
Department of Biochemistry
332 N Lauderdale
Memphis, TN 38105
U.S.A.
Phone: +1 901 522 0422
FAX: +1 901 525 8025
Interleukin 3

Yasuyuki Ikeda
National Cardiovascular Center Research Institute
5-125 Fujishiro-dai
Suita City, Osaka 565
Japan
Phone: +81 6 833 5012
Acyl-CoA dehydrogenase

Stephan Immenschuh
Zentrum für Innere Medizin
Abt. Gastroenterologie/Endokrinologie
Georg-August-Universität Göttingen
Robert-Koch-Str. 40
D-37075 Göttingen
Germany
Phone: +49 551 398 279
FAX: +49 551 398 279
Hemopexin

Satoko Isemura
Nippon Dental University
Junior College at Niigata
Hamaurachou 1-8
Niigata, Niigata
Japan
Cystatin S, SN, SA

Yoshifumi Itoh
Department of Cancer Cell Research
Institute of Medical Science
University of Tokyo
4-6-1 Shirokanedai, Minato-ku
Tokyo 108-0071
Japan
Phone: +81 3 5449 5259
FAX: +81 3 5449 5414
E-mail: yitoh@ims.u-tokyo.ac.jp
Membrane-type 1 matrix metalloproteinase

Craig M. Jackson
American Red Cross Southeastern
Michigan Region
P.O. Box 33351
Detroit, MI 48232-5351
U.S.A.
Phone: +1 313 494 2719
FAX: +1 313 833 4461
Factor X
Prothrombin

Mikko Järvinen
Department of Dermatology
University of Turku
SF-20520 Turku
Finland
Phone: +358-21-611611
FAX: +358-21-611610
Cystatin A
Cystatin B

Willi Jahnen-Dechent
Institute for Physiological Chemistry
University of Mainz
Duesbergweg 6
D-55128 Mainz 1
Germany
Phone: +49 6131 395 890
FAX: +49 6131 395 792
Alpha2-HS-glycoprotein

Bradford A. Jameson
Jefferson Cancer Institute
Department of Pharmacology
10 Walnut Street
Philadelphia, PA 19107
U.S.A.
Phone: +1 215 955 4564
FAX: +1 215 923 2117
Platelet factor 4

Dieter E. Jenne
Department of Neuroimmunology
Max-Planck Institute for Psychiatry
D-82152 Planegg-Martinsried
Germany
Phone: +49 89 8578 3588
FAX: +49 89 8578 3777
E-mail: Jenne@vms.biochem.mpg.de
C6 complement protein

Richard J. Jenny
Haematologic Technologies Inc.
Pinewood Plaza
Box 1021
57 River Road
Essex Jct., VT 05452
U.S.A.
Phone: +1 802 878 1777
FAX: +1 802 878 1776
Factor V

Ana Jonas
University of Illinois at Urbana-Champaign
Department of Biochemistry
College of Medicine
190 Medical Sciences Building
506 South Mathews Avenue
Urbana, IL 61801
U.S.A.
Phone: +1 217 333 0452
FAX: +1 217 333 8868
E-mail: a-jonas@uiuc.edu
Lecithin Cholesterol Acyltransferase

Kristopher Josephson
University of Alabama at Birmingham
Center for Macromolecular Crystallography
265 Basic Health Sciences
1918 University Boulevard
Birmingham, AL 35294-0005
U.S.A.
Phone: +1 205 934 8990
FAX: +1 205 934 0480
Interleukin-10

Dimitris Kardassis
Section of Molecular Genetics
Boston University Medical Center
80 East Concord Street, R420
Boston, MA 02118
U.S.A.
Phone: +1 617 638 5081
FAX: +1 617 638 5141
Apolipoprotein CI
Apolipoprotein CIII

Hiroshi Kawasaki
The Tokyo Metropolitan Institute of Medical Science
Department of Molecular Biology
3-18-22 Honkomagome
Bunkyo-Ku
Tokyo 113
Japan
Phone: +81 3 8232101
FAX: +81 5 6856609
Calpain

Paul A. Kelly
INSERM Unité 344
Endocrinologie Moléculaire
Faculté de Médecine Necker
156, rue de Vaugirard
F-75730 Paris Cedex 15
France
Phone: +33 1 40 61 53 10
FAX: +33 1 43 06 04 43
E-mail: kelly@necker.fr
Prolactin receptor

A. John Kenny
Le Presbytère
F-34390 St. Etienne d' Albagnan
France
Phone: +33 467 97 1537
FAX: +33 467 97 1537
Endopeptidase-24.11

Lynda J. Kieffer
Present address unknown
Cyclophilin-40

Masakazu Kikuchi
Faculty of Science and Engineering
Ritsumeikan University
1916 Noji-cho
Kusatsu, Shiga 525-77
Japan
Phone: +81 775 61 2843
FAX: +81 775 61 2659
E-mail: kikuchi@bkc.ritsumei.ac.88
Lysozyme

Erich Kilchherr
Pharmaceuticals Division
Research Department
R-1056.2.25
CH-4002 Basel
Switzerland
Phone: +41 61 697 69 64
FAX: +41 61 697 39 68
CD 23
Interleukin 4

Landon S. King
Johns Hopkins University
Division of Pulmonary and
Critical Care Medicine
720 Rutland Avenue / Ross 858
Baltimore, MD 21205-2196
U.S.A.
Phone: +1 410 955 3467
FAX: +1 410 955 0036
E-mail: isking@welchlink.welch.jhu.edu
Aquaporin-1

Tadamitsu Kishimoto
Department of Medicine III
Osaka University Medical School
Fukushima-Ku
1-1-50 Fukushima
Osaka City 553
Japan
Phone: +81 6 451 0489
FAX: +81 6 451 8870
Interleukin 6

Uday Kishore
MRC Immunochemistry Unit
Department of Biochemistry
University of Oxford
South Parks Road
Oxford OX1 3QU
United Kingdom
Phone: +44 01865 275 353
FAX: +44 01865 275 729
Mannose-binding lectin
Surfactant Protein A
Surfactant Protein D

Walter Kisiel
School of Medicine
Department of Pathology
University of New Mexico
Albuquerque, NM 87131
U.S.A.
Phone: +1 505 277 0094
FAX: +1 505 277 0593
Factor VII

Hynda K. Kleinman
NIH, NIDR 30/407
9000 Rockville PK
Bethesda, MD 20892
U.S.A.
Phone: +1 301 496 4069
FAX: +1 301 402 0897
E-mail: kleinman@yoda.nidr.nih.gov
Laminin

Teresa Krakauer
Dept. of Molec. Biology and Immunology
USAMRIID
Fort Detrick
Frederick, MD 21702-5011
U.S.A.
Phone: +1 301 619 4733
FAX: +1 301 619 2348
Interleukin 1

Egbert K.O. Kruithof
Hôpital Cantonal Universitaire de Genève
Département de Médecine Interne
Division d'Angiologie et d'Hémostase
CH-1211 Genève
Switzerland
Phone: +41 22 372 97 58
FAX: +41 22 372 92 99
E-mail: egkr@diogenes.hcuge.ch
Plasminogen activator inhibitor type 2

Satish Kumar
University of Oklahoma Health Sciences Center
941 Stanton L. Young Blvd., BSEB 302
Oklahoma City, OK 73104
U.S.A.
Phone: +1 405 271 3920
FAX: +1 405 271 3191
E-mail: satish-kumar@uokhsc.edu
Tamm-Horsfall protein

Gloria Kurys
Center of Alcohol Studies
Busch Campus
Rutgers University
Allison Rd.
Piscataway, NJ 08855-0969
U.S.A.
Phone: +1 908 932 3643
FAX: +1 908 932 5944
Aldehyde dehydrogenase, E3, liver cytoplasm

Bernhard Lämmle
Inselspital
Central Hematology Laboratory
University of Bern
CH-3010 Bern
Switzerland
Phone: +41 31 632 33 02
FAX: +41 31 632 93 66
Factor XII

Thung S. Lai
Duke University Medical Center
Box 2603
Durham, NC 27710
U.S.A.
Phone: +1 919 684 6703
FAX: +1 919 684 4670
E-mail: lai00002@mc.duke.edu
Plasma Faktor XIII

Anne-Marie Lambeir
University of Antwerp
Department of Pharmaceutical Sciences
Universiteitsplein 1
B-2610 Wilrijk
Belgium
Phone: +32 38 20 2727
FAX: +32 38 20 2745
E-mail: lambeir@uia.ua.ac.be
CD26

Annie Lamouroux
Laboratoire de Neurobiologie Cellulaire et Moléculaire
C.N.R.S.
1, Avenue de la Terrasse
F-91198 Gif sur Yvette
France
Phone: +33 1 698 23636
FAX: +33 1 698 23580
Dopamine beta-hydroxylase
Tyrosine hydroxylase

S.K. Alex Law
MCR Immunochemistry Unit
Department of Biochemistry
South Parks Road
Oxford OX1 3QU
United Kingdom
Phone: +44 865 275350
FAX: +44 865 275729
C4 complement protein

Jack Lawler
Vascular Research Division
Department of Pathology
Brigham and Women's Hospital
LMRC-Room 414
221 Longwood Avenue
Boston, MA 02115
U.S.A.
Phone: +1 617 732 6005
FAX: +1 617 732 5933
E-mail: lawler@mbcrr.harvard.Edu
Cartilage oligomeric matrix protein
Thrombospondin-1
Thrombospondin-2
Thrombospondin-3
Thrombospondin-4

Daniel A. Lawrence
Department of Biochemistry
Holland Laboratory
15601 Crabbs Branch Way
Rockville, MD 20855
U.S.A.
Phone: +1 301 517 0356
FAX: +1 301 738 0794
E-mail: LawrencD@U.S.A..redcross.org
Plasminogen activator inhibitor Type 1

Elizabeth C. Ledgerwood
Molecular Pathology Laboratory
Christchurch School of Medicine
Christchurch Hospital
Christchurch
New Zealand
Phone: +64 3 364 0549
FAX: +64 3 364 0750
Furin

Monica X. Li
Department of Biochemistry
University of Alberta
Edmonton, Alberta T6G 2H7
Canada
Phone: +1 403 492 3006
FAX: +1 403 492 0886
E-mail: monica.li@ualberta.ca
Troponin C

Wei Li
Yale University School of Medicine
Department of Pharmacology, SHM
333 Cedar Street
P.O. Box 3333
New Haven , CT 06510-8066
U.S.A.
Phone: +1 203 785 4383
FAX: +1 203 785 7670
Cyclophilin-18

Stephen A. Liebhaber
Howard Hughes Medical Institute
Departments of Genetics and Medicine
University of Pennsylvania
437A Clinical Research Building
422 Curie Boulevard
Philadelphia, PA 19104-6145
U.S.A.
Phone: +1 215 898 7834
FAX: +1 215 898 1257
Cysteine-rich protein

H. Roger Lijnen
Center for Molecular and Vascular Biology
Campus Gasthuisberg O & N
University of Leuven
Herestraat 49
B-3000 Leuven
Belgium
Phone: +32 16 215 775
FAX: +32 16 215 990
Histidine-rich glycoprotein

Hans Lilja
Malmö General Hospital
Department of Clinical Chemistry
S-21401 Malmö
Sweden
Phone: +46 40 331 424
FAX: +46 40 337 043
E-mail: hans.lilja@klkemi.mas.lu.se
Prostate specific antigen
Semenogelin I
Semenogelin II

Sven Lindskog
Department of Biochemistry
University of Umeå
S-90187 Umeå
Sweden
Phone: +46 90 165 387
FAX: +46 90 167 661
E-mail: slg@chem.umea.se
Carbonic anhydrase I

M. Kathryn Liszewski
Washington University School of Medicine
Division of Rheumatology
Department of Internal Medicine
Campus Box 8045
660 South Euclid
St. Louis, MO 63110
U.S.A.
Phone: +1 314 362 8392
FAX: +1 314 362 1366
E-mail: liszwski@im.wustl.edu
Membrane cofactor protein (CD46)

Oksana Lockridge
Eppley Institute
University of Nebraska Medical Center
600 S. 42nd St.
Omaha, NE 68198-6805
U.S.A.
Phone: +1 402 559 6032
FAX: +1 402 559 4651
E-mail: olockrid@unmc.edu
Butyrylcholinesterase

Lennart Lögdberg
Sandoz Research Institute
59 Route 10
East Hanover, NJ 07936
U.S.A.
Phone: +1 201 503 8638
FAX: +1 201 503 6870
Beta-2-microglobulin

George L. Long
Department of Biochemistry
B-418 Given Building
School of Medicine
University of Vermont
Burlington, VT 05405
U.S.A.
Phone: +1 802 656 0351
FAX: +1 802 862 8229
Osteonectin

Tip W. Loo
Department of Medicine
University of Toronto
Medical Sciences Building, Room 7342
Toronto, Ontario M5S 1A8
Canada
Phone: +1 416 978 1105
FAX: +1 416 978 1105
E-mail:
P-glycoprotein

Michael Loos
Institute of Medical Microbiology and Hygiene
Johannes Gutenberg-University Mainz
Augustusplatz / Hochhaus
D-55101 Mainz
Germany
Phone: +49 6131 173 139
FAX: +49 6131 173 439
E-mail: petrf000@goofy.zdv.uni-mainz.de
C1q complement protein

Carlos López-Otín
Departamento de Biologia Funcional
Facultad de Medicina
Universidad de Oviedo
E-33006 Oviedo
Spain
Phone: +34 8510 4201
FAX: +34 8510 3564
E-mail: clo@dwarf1.quimica.uniovi.es
Zn.alpha-2-glycoprotein
Apolipoprotein D

David J. Loskutoff
VB-3, Department of Vascular Biology
The Scripps Research Institute
10550 North Torrey Pines Rd.
La Jolla, CA 92037
U.S.A.
Phone: +1 619 784 7125
FAX: +1 619 784 7353
Plasminogen activator inhibitor type 1

Douglas M. Lublin
Associate Prof. of Pathology and Medicine
Washington University School of Medicine
660 South Euclid Avenue
Box 8118
St. Louis, MO 63110
U.S.A.
Phone: +1 314 362 8849
FAX: +1 314 362 3016
E-mail: lublin@labmed.wustl.edu
Decay accelerating factor

Gregory P. Lunstrum
Research Department
Shriners Hospital for Crippled Children
3101 SW Sam Jackson Park Road
Portland, OR 47201
U.S.A.
Phone: +1 503 221 1537
FAX: +1 503 221 3451
Collagen type XII
Collagen type XIV

Samuel E. Lux
Division of Hematology/Oncology
Children's Hospital, Enders 7
300 Longwood Avenue
Boston, MA 02115
U.S.A
Phone: +1 617 355 7904
FAX: +1 617 355 7262
E-mail: lux@a1.tch.harvard.edu.
Ankyrin

James A. Magner
Human Hospital-Michael Reese
Division of Endocrinology
Lake Shore Dr. at 31st Street
Chicago, IL 60616
U.S.A.
Phone: +1 312 791 3156
Thyrotropin

Don J. Mahuran
The Hospital for Sick Children
Research Institute
Division of Neurosciences
9th Floor, Elm Wing Room 9142
555 University Avenue
Toronto, Ontario M5G IX8
Canada
Phone: +1 416 813 6161
FAX: +1 416 813 5086
E-mail: hex@sickkits.on.ca
β-N-acetyl-D-hexosaminidase

Mahin D. Maines
Department of Biophysics
University of Rochester School of Medicine
575 Elmwood Avenue
Rochester, NY 14642
U.S.A.
Phone: +1 716 275 2441
FAX: +1 716 256 2591
Billiverdin reductase
Heme oxygenases

Rajneesh Malhotra
MRC Immunochemistry Unit
Department of Biochemistry
University of Oxford
South Parks Road
Oxford OX1 3QU
United Kingdom
Phone: +44 865 275 352
FAX: +44 865 275 729
Collectin receptor

Martine Malissard
Physiologisches Institut
Universität Zürich
Winterthurerstr. 190
CH-8058 Zürich
Switzerland
Phone: +41 257 50 11
FAX: +41 364 05 64
Beta 1,4-galactosyltransferase

Jacques Mallet
Laboratoire de Génétique Moléculaire de la Neurotransmission et des Processus Neuro-dégénératifs (L.G.N.)
UMR 9923 C.N.R.S.
Bâtiment CERVI
Hôpital de la Pitié Salpêtrière
83, Boulevard de l'Hôpital
F-75013 Paris
France
Phone: +33 1 42 17 75 32
FAX: +33 1 42 17 75 33
E-mail: mallet@infobiogen.fr
Dopamine beta-hydroxylase
Tryptophan hydroxylase
Tyrosine hydroxylase

Johan Malm
Malmö General Hospital
Department of Clinical Chemistry
S-21401 Malmö
Sweden
Phone: +46 40 333 536
FAX: +46 40 337 043
E-mail: johan.malm@klkemi.mas.lu.se
Prostate specific antigen
Semenogelin I
Semenogelin II

Kenneth G. Mann
Department of Biochemistry
B-418 Given Building
School of Medicine
University of Vermont
Burlington, VT 05405
U.S.A.
Phone: +1 802 656 0335
FAX: +1 802 862 8229
E-mail: glong@moose.uvm.edu
Osteonectin

Ned Mantei
Laboratorium für Biochemie II
ETH-Zentrum
Universitätsstrasse 16
CH-8092 Zürich
Switzerland
Phone: +41 1 256 30 50
FAX: +41 1 252 87 44
Lactase-phlorizin hydrolase

Fritz Märki
Present address unknown
Phospholipase A_2
Phospholipase C

Joseph A. Martial
Laboratory of Molecular Biology and Genetic Engineering
University of Liège
Institut de Chimie, Allée du 6 Août
B-4000 Sart-Tilman
Belgium
Phone: +32 4 366 33 11
FAX: +32 4 366 29 68
E-mail:jmartial@ulg.ac.be
Prolactin

Janet L. Martin
Kolling Institute of Medical Research
Royal North Shore Hospital
St Leonards, NSW 2065
Australia
Phone: +61 2 9926 8486
FAX: +61 2 9926 8484
Insulin-like growth factor I
Insulin-like growth factor II
Insulin-like growth factor binding protein-1
Insulin-like growth factor binding protein-2
Insulin-like growth factor binding protein-3
Insulin-like growth factor binding protein-4
Insulin-like growth factor binding protein-5
Insulin-like growth factor binding protein-6

Masaaki Matsushima
Protein Engineering Research Institute
6-2-3 Furuedai
Suita, Osaka 565
Japan
Phone: +81 6 872 8201
FAX: +81 6 872 8210
Lysozyme

Bernd Mayer
Karl-Franzens Universität Graz
Institut für Pharmakologie und Toxikologie
Universiätsplatz 2
A-8010 Graz
Austria
Phone: +43 316 380 5567
FAX: +43 316 380 9890
Nitric oxide synthase

J. Hugh McDowell
Department of Ophthalmology
University of Florida
Gainesville, FL 32610-0284
U.S.A.
Phone: +1 352 392 9097
FAX: +1 352 392 0478
Opsin

Dennis L. Merat
St. Jude Children's Research Hospital
Department of Biochemistry
Memphis, TN 38101
U.S.A.
Phone: +1 901 522 0436
FAX: +1 901 525 8025
Calmodulin

Jiri Mestecky
Medical Center
Departments of Microbiology and Medicine
University of Alabama at Birmingham
University Station
757 Bevill Biomedical Research Building
845 19th Street South
Birmingham, AL 35294-2170
U.S.A.
Phone: +1 205 934 2225
FAX: +1 205 934 3894
Immunoglobulin A
J chain
Secretory component

Urs A. Meyer
Department of Pharmacology
Biocenter
University of Basel
Klingelbergstrassse 70
CH-4056 Basel
Switzerland
Phone: +41 61 267 22 21
FAX: +41 61 267 22 08
Arylamine N-acetyltransferase

Robert Q. Miao
Department of Biochemistry and
Molecular Biology
Medical University of South Carolina
171 Ashley Avenue
Charleston, SC 29425-2211
U.S.A.
Phone: +1 803 792 4321
FAX: +1 803 792 1627
E-mail: miaoq@musc.edu
Kallistatin

John J. Mieyal
Department of Pharmacology
Case Western Reserve University
School of Medicine
2109 Adelbert Road
Cleveland, OH 44106-4965
U.S.A.
Phone: +1 216 368 3383
FAX: +1 216 368 3395
E-mail: jjm5@po.cwru.edu
Thioltransferase

Peer Mittl
Novartis Pharma AG
Pharmaceutical Division
CH-4002 Basel
Switzerland
Phone: +41 61 696 4432
FAX: +41 61 696 9301
E-mail: mittl@fmi.ch
Transforming growth factor-beta 3 (1,2)

Jean Montreuil
Laboratoire de Biochimie
Université de Sciences et Technologies de Lille
F-59655 Villeneuve d'Ascq Cédex
France
Phone: +33 20 434 884
FAX: +33 20 436 555
Lactoferrin

Mahnaz Moradi-Améli
IBCP - CNRS UPR 412
7 Passage du Vercors
F-69367 Lyon Cédex 07
France
Phone: +33 7 272 2666
FAX: +33 7 272 2602
E-mail: Jbdm@ibcp.fr
Collagen type V
Collagen type XI

Evan H. Morgan
Department of Physiology
University of Western Australia
Nedlands, WA 6009
Australia
Phone: +61 9 380 3399
FAX: +61 9 380 1025
Transferrin

Richard F. Mortensen
Department of Microbiology
Ohio State University
484 W. 12th Avenue
Columbus, OH 43210
U.S.A.
Phone: +1 614 292 3360
FAX: +1 614 292 8120
E-mail: mortensen.3@osu.edu
C-reactive protein

Dean F. Mosher
Department of Medicine (Hematology)
University of Wisconsin 4285 MSC
1300 University Avenue
Madison, WI 53706
U.S.A.
Phone: +1 608 262 1576
FAX: +1 608 262 2327
Fibronectin

Roman Muff
Research Laboratory for Calcium Metabolism
Departments of Orthopaedic Surgery and Medicine
Forchstrasse 340
CH-8008 Zürich
Switzerland
Phone: +41 1 386 16 51
FAX: +41 1 386 16 52
Parathyroid hormone
Parathyroid hormone-related protein

Ursula Muller-Eberhard
Medical College
Cornell University
1300 York Avenue
New York, NY 10021
U.S.A.
Phone: +1 212 746 3421
FAX: +1 212 746 8458
Hemopexin

Werner Müller-Esterl
Institute for Physiological Chemistry
University of Mainz
Duesbergweg 6
D-55128 Mainz 1
Germany
Phone: +49 6131 395 890
FAX: +49 6131 395 792
E-mail: muellere@mzdmza.zdv.uni-mainz.dc
Alpha2-HS glycoprotein
Kininogens

William M. Nauseef
Department of Medicine
University of Iowa
200 Hawkins Drive
Iowa City, IA 52242
U.S.A.
Phone: +1 319 356 1739
FAX: +1 319 356 4600
E-mail: nauseefw@intmed-po.int-med.uiowa.edu
Myeloperoxidase

Yale Nemerson
Mount Sinai School of Medicine
One Gustave Levy Place
P.O. Box 1269
New York, NY 10029
U.S.A.
Phone: +1 212 241 6083
FAX: +1 212 860 7032
Tissue factor

Christopher B. Newgard
Medical Center, Center for Diabetes Research
Department of Biochemistry
University of Texas Southwestern
5323 Harry Hines Blvd.
Dallas, TX 75235
U.S.A.
Phone: +1 214 688 2930
FAX: +1 214 688 8291
Glycogen phosphorylase

Nicos A. Nicola
P.O. Royal Melbourne Hospital
The Walter and Eliza Hall Institute for Medical Research
Centre for Cellular Growth Factors
Parkville, VA 3050
Australia
Phone: +61 3 345 2526
FAX: +61 3 347 0852
Granulocyte colony-stimulating factor (G-CSF)
Granulocyte-macrophage colony-stimulating factor (GM-CSF)
Leukemia inhibitory factor (LIF)
Macrophage colony-stimulating factor (M-CSF)

Stefan Niewiarowski
Department of Physiology
Temple University School of Medicine
3400 N Broad Street
Philadelphia, PA 19140
U.S.A.
Phone: +1 215 221 4408
FAX: +1 215 221 4003
E-mail: stni@astro.ocis.temple.edu
Platelet factor 4

Ove Norén
IMBG, Biochemistry Department C
Blegdamsvej 3
DK-2200 Copenhagen N
Denmark
Phone: +45 3532 7793
FAX: +45 3536 7980
E-mail: noren@biobase.dk
Aminopeptidase N

Michio Ogawa
Department of Surgery II
Kumamoto University Medical School
1-1-1 Honjo
Kumamoto 860
Japan
Phone: +81 96 373 5209
FAX: +81 96 371 4378
Alpha-amylase
Pancreatic secretory trypsin inhibitor

Joost J. Oppenheim
National Institute of Health
Laboratory of Molecular Immunoregulation
BRMP, FCRDC-NCI
Frederick, MD 21702-1201
U.S.A.
Phone: +1 301 846 1551
FAX: +1 301 846 1673
Interleukin 1

Denis O'Shaughnessy
I.C.R.C. Roche
4, Route de la Rivière
F-67382 Lingolsheim Strasbourg
France
Phone: +33 88 774 418
FAX: +33 88 774 400
Interferon alpha

Maurice C. Owen
Molecular Pathology Laboratory
Department of Clinical Biochemistry
Christchurch Hospital
Christchurch
New Zealand
Phone: +64 3 640 546
FAX: +64 3 640 750
Alpha-1-proteinase inhibitor

Krzysztof Palczewski
University of Washington
Ophthalmology, Box 356485
Seattle, WA 98195-6485
U.S.A.
Phone: +1 206 543 9074
FAX: +1 206 543 4414
E-mail: palczews@u.washington.edu
Arrestin
Rhodopsin kinase

Michael K. Pangburn
Department of Biochemistry
University of Texas Health Center
P.O. Box 2003
Tyler, TX 75710
U.S.A.
Phone: +1 903 877 7663
FAX: +1 903 877 7558
Factor B
Factor I

Manuel A. Pascual
Massachusettes General Hospital
Fruit Street
Boston, MA 02114
U.S.A.
Phone: +1 617 726 3772
FAX: +1 617 228 6306
Factor D

Hermann Passow
Max-Planck-Institut für Biophysik
Abteilung für Zellphysiologie
Heinrich Hoffmann-Strasse 7
D-60528 Frankfurt/Main 71
Germany
Phone: +49 6996 769 358 / 359
FAX: +49 6996 769 350
Band 3 protein (AE1)

Mark B. Pepys
Hammersmith Hospital
Department of Medicine
Royal Postgraduate Medical School
Du Cane Road
London W12 ONN
United Kingdom
Phone: +44 181 383 3202
FAX: +44 181 749 7478
E-mail: mpepys@rpms.ac.uk
Serum amyloid P component

Stephen. J. Perkins
Howard Hughes Medical Institute
Department of Medicine
Division of Rheumatology
Washington University
School of Medicine
St. Louis, MO 63110
U.S.A.
Immunoglobulin M

Anders Peter
Malmö General Hospital
Department of Clinical Chemistry
S-21401 Malmö
Sweden
Phone: +46 40 337 219
FAX: +46 40 337 043
E-mail: anders.peter@klkemi.mas.lu.se
Prostate specific antigen
Semenogelin I
Semenogelin II

Theodore Peters, Jr.
Research Institute
The Mary Imogene Bassett Hospital
Cooperstown, NY 13326
U.S.A.
Phone: +1 607 547 3673
FAX: +1 607 547 3061
E-mail: tedp@U.S.A..net
Albumin

Franz Petry
Institute of Medical Microbiology and Hygiene
Johannes Gutenberg University Mainz
Augustusplatz / Hochhaus
D-55101 Mainz
Germany
Phone: +49 6131 173 139
FAX: +49 6131 173 439
E-mail: petrf000.@goofy.zdv.uni-mainz.de
C1q complement protein

Michael Pieper
Universität Bielefeld
Universitätsstr. 25
D-33615 Bielefeld
Germany
Phone: +49 521 106 2081
FAX: +49 521 106 6014
Neutrophil collagenase

Regina Pietruszko
Center of Alcohol Studies
Busch Campus
Rutgers University
Allison Rd.
Piscataway, NJ 08855-0969
U.S.A.
Phone: +1 908 445 3643
FAX: +1 908 445 3500
Aldehyde dehydrogenase, E1, liver cytoplasm
Aldehyde dehydrogenase, E2, liver mitochondria
Aldehyde dehydrogenase, E3, liver cytoplasm
Aldehyde dehydrogenase, stomach cytoplasm
Glutamic γ-semialdehyde dehydrogenase, liver mitochondria

Taina Pihlajaniemi
Department of Medical Biochemistry
University of Oulu
Kajaanintie 52 A
SF-90220 Oulu
Finland
Phone: +358 81 537 5803
FAX: +358 81 537 5810
Collagen type XIII
Collagen type XV

Teresa J.T. Pinheiro
University of Oxford
Biochemistry Department
South Parks Road
Oxford OX1 3QU
United Kingdom
Phone: +44 1865 275 270
FAX: +44 1865 275 259
Cytochrome C

Michael Plotnick
Department of Medicine
School of Medicine
University of Pennsylvania
Philadelphia, PA 19104
U.S.A.
Alpha-1-antichymotrypsin

Krishna Prasadan
University of Oklahoma Health Sciences Center
941 Stanton L. Young Blvd., BSEB 302
Oklahoma City, OK 73104
U.S.A.
Phone: +1 405 271 3920
FAX: +1 405 271 3191
E-mail: krishna-prasadan@uokhsc.edu
Tamm-Horsfall protein

Klaus T. Preissner
Max-Planck-Gesellschaft
Kerckhoff-Klinik
Haemostasis Research Unit
Sprudelhof 11
D-61231 Bad Nauheim
Germany
Phone: +49 6032 345 719
FAX: +49 6032 345 707
Vitronectin

David H. Presky
Hoffmann-La Roche Inc.
340 Kingsland Street
Nutley, NJ 07110-1199
U.S.A.
Phone: +1 201 235 3544
FAX: +1 201 235 5046
Interleukin 12

David Puett
Department of Biochemistry
Life Sciences Building
University of Georgia
Athens, GA 30602
U.S.A.
Phone: +1 706 542 1676
FAX: +1 706 542 0182
E-mail: puett@bchiris.biochem.uga.edu
Choriogonadotropin

Stanley C. Rall, Jr.
Gladstone Institute of Cardiovascular Disease
P.O. Box 419100
San Francisco, CA 94141-9100
U.S.A.
Phone: +1 415 826 7500
FAX: +1 415 285 5631
E-mail: stanley-rall@quickmail.ucsf.edu
Apolipoprotein E

Basil Rapoport
Thyroid Molecular Biology Unit (111 T)
4150 Clement St.
San Francisco, CA 94121
U.S.A.
Phone: +1 415 476 5984
FAX: +1 415 752 6745
Thyrotropin receptor

Lars Rask
Department of Cell Research
Swedish University of Agricultural Sciences
P.O. Box 7055
S-750 07 Uppsala
Sweden
Phone: +46 18 673 286
FAX: +46 18 673 389
Retinol-binding protein

Samuel Refetoff
Department of Medicine
University of Chicago
MC 3090
5841 South Maryland Avenue
Chicago, IL 60637
U.S.A.
Phone: +1 312 702 6939
FAX: +1 312 702 6940
E-mail: refetoff@medicine.bsd.uchicago.edu
Thyroxine-binding globulin

Marko Rehn
Department of Medical Biochemistry
University of Oulu
Kajaanintie 52A
SF-90220 Oulu
Finland
Phone: +358 8573 5840
FAX: +358 8573 5811
E-mail: marko.rehn@oulu.fi
Collagen type XVIII

Kenneth B.M. Reid
MRC Immunochemistry Unit
Department of Biochemistry
University of Oxford
South Parks Road
Oxford OX1 3QU
United Kingdom
Phone: +44 01865 275 353
FAX: +44 01865 275 729
Properdin
Mannose-binding lectin
Surfactant Protein A
Surfactant Protein D

Karen Reue
Lipid Research Laboratory
VA Medical Center
Building 113, Room 312
Los Angeles, CA 90073
U.S.A.
Phone: +1 310 824 4395
FAX: +1 310 478 4538
Carboxyl ester lipase
Hormone-sensitive lipase

Egon E. Rickli
Institut für Biochemie
Universität Bern
Freiestrasse 3
CH-3012 Bern
Switzerland
Phone. +41 31 631 43 41
FAX: +41 31 631 44 99
Plasminogen

Ari Rinne
Department of Dermatology
University of Turku
Kiinamyllynkatu 4-8
SF-20520 Turku
Finland
Phone: +358 216 11611
FAX: +358 216 11610
Cystatin A
Cystatin B

Riitta Rinne
Department of Dermatology
University of Turku
Kiinamyllynkatu 4-8
SF-20520 Turku
Finland
Phone: +358 216 11611
FAX: +358 216 11610
Cystatin A
Cystatin B

James F. Riordan
Harvard Medical School
Seeley G. Mudd Building
250 Longwood Avenue
Boston, MA 02115
U.S.A..
Phone: +1 617 432 1367
FAX: +1 617 566 3137
E-mail: jriordan@warren.med.harvard.edu
Angiotensin converting enzyme

Louise M. Rosenbaum
IBCP-CNRS VPR 412
7, Passage du Verors
F-69367 Lyon Cédex 07
France
Phone: +33 7 272 2619
FAX: +33 7 272 2602
Collagen type VII

Eric D. Roush
Department of Biochemistry
Duke University Medical Center
Box 3711
Durham, NC 27710
U.S.A..
Phone: +1 919 684 8976
FAX: +1 919 684 4050
Carbonic anhydrase II

Harvey Rubin
Department of Medicine
School of Medicine
Universtiy of Pennsylvania
Philadelphia, PA 19104
U.S.A.
Phone: +1 215 662 6475
FAX: +1 215 573 4858
E-mail: rubinh@mail.med.upenn.edu
Alpha-1-antichymotrypsin

Michael W. Russell
Medical Center
Department of Microbiology, BBRB 734
University of Alabama at Birmingham
845 19th Street South
Birmingham, AL 35294-2170
U.S.A.
Phone: +1 205 934 4480
FAX: +1 205 934 3894
Immunoglobulin A
J chain
Secretory component

Lucia Sacchetti
Dip. di Biochimica e Biotecnologie Mediche
Facoltà di Medicina e Chirurgia
Università degli Studi di Napoli "Federico II"
Via Sergio Pansini 5
I-80131 Napoli
Italy
Phone: +39 81 746 3135
FAX: +39 81 746 3650
Creatine kinase
Gamma-glutamyltransferase
Lactate dehydrogenase
Superoxide dismutase

Ikuko Saito
Dept. Hygienic Chemistry
Fac. Pharmaceutical Sciences
Hokkaido University
Kita-ku, Sapporo 060
Japan
Phone: +81 11 706 3244
FAX: +81 11 706 4989
Glutathione peroxidase, cellular
Glutathione peroxidase, extracellular

Francesco Salvatore
Facoltà di Medicina e Chirurgia
Dipartimento di Biochimica e Biotecnologie Mediche
Università degli Studi di Napoli "Federico II"
Via Sergio Pansini 5
I-80131 Napoli
Italy
Phone: +39 81 746 3135
FAX: +39 81 746 3650
Fructose-1,6-bisphosphate aldolase
Glyceraldehyde-3-phosphate dehydrogenase

Colin J. Sanderson
Head, Molecular Immunology Division
TVWT Institute for Child Health Research
P.O. Box 855
Perth, WA 6872
Australia
Phone: +61 619 340 8504
FAX: +61 619 388 3414
E-mail: colin@ichr.uwa.edu.au
Interleukin 5

Erika Sandmeier
Biochemisches Institut der Universität Zürich
Winterthurerstrasse 190
CH-8057 Zürich
Switzerland
Phone: +41 1 257 55 11 (67)
FAX: +41 1 363 79 47
Aspartate aminotransferase

Maria J. Saraiva
Bioquimica Instituto de Ciências Biomedicas
Universidade do Porto
P-4000 Porto
Portugal
Phone: +351 2 606 6149
FAX: +351 2 606 6106
Transthyretin

Hiroyuki Satoh
Present address unknown
Hemopexin

Angelo M. Scanu
Department of Medicine
University of Chicago
5841 S. Maryland Mail Code MC 5041
Chicago, IL 60637
U.S.A.
Phone: +1 312 702 1775
FAX: +1 312 702 4534
Apolipoprotein(a)

Johann Schaller
Institut für Biochemie
Freiestrasse 3
CH-3012 Bern
Switzerland
Phone: +41 31 631 43 52 or 53
FAX: +41 31 631 48 87
E-mail: schaller@ibc.unibe.ch
Plasminogen

Simon Scharpé
University of Antwerpen
Laboratory of Medical Biochemistry
Universiteitsplein 1
B-2610 Wilrijk
Belgium
Phone: +32 3 820 27 18
FAX: +32 3 820 27 45
E-mail: scharpe@uia.ua.ac.be
Dipeptidyl peptidase IV (CD 26)
Aminopeptidase P
Prolyl oligopeptidase
Procarboxypeptidase U

Katinka A. Schatteman
University of Antwerp
Department of Pharmaceutical Sciences
Universiteitsplein 1
B-2610 Wilrijk
Belgium
Phone: +32 3 820 2727
FAX: +32 3 820 2745
E-mail: schattem@uia.ua.ac.be
Carboxypeptidase N
Procarboxypeptidase U

Jürg A. Schifferli
Laboratory of Immunonephrology
Medizinische Univeristätsklinik B
Department Innere Medizin
Kantonsspital Basel
CH-4000 Basel
Switzerland
Phone +41 61 265 42 92
FAX +41 61 265 53 90
Factor D

Michael C. Schotz
VA Wadsworth Medical Center
Bldg. 113, Room 312
Los Angeles, CA 90073
U.S.A.
Phone: +1 310 268 3927
FAX: +1 310 478 4538
E-mail: schotz1@ucla.edu
Lipoprotein lipase

Motoharu Seiki
Department of Cancer Cell Research
Institute of Medical Science
University of Tokyo
4-6-1 Shirokanedai, Minato-ku
Tokyo, 108-0071
Japan
Phone: +81 3 5449 5255
FAX: +81 3 5449 5414
E-mail: mseiki@ims.u-tokyo.ac.jp
Membrane-type 1 matrix metalloproteinase

Giorgio Semenza
Laboratorium für Biochemie
ETH-Zentrum
Universitätsstr. 16
CH-8092 Zürich
Switzerland
Phone: +41 1 632 3133
FAX: +41 1 632 1089
E-mail: semenza@bc.biol.ethz.ch
Lactase-phlorizin hydrolase

Anindita Sen
Duke University Medical Center
Box 3892
Durham, NC 27710
U.S.A.
Phone: +1 919 684 4343
FAX: +1 919 684 3590
E-mail: a.sen@cellbio.duke.edu
Caldesmon

Lei Shen
University Hospital
Department of Clinical Chemistry
S-20502 Malmö
Sweden
Phone: +46 40 33 1501
FAX: +46 40 33 7044
Protein C

Michael Silverberg
Division of Allergy, and Clinical Immunology
Department of Medicine
Health Sciences Center
SUNY Stony Brook, NY 11794-8161
U.S.A.
Phone: +1 516 444 2347
FAX: +1 516 444 3475
E-mail: michaels@epo.hsc.sunysb.edu
Prekallikrein

David N. Silverman
Department of Pharmacology & Therapeutics
University of Florida
Gainesville, FL 32610-0267
U.S.A.
Phone: +1 904 392 3556
FAX: +1 904 392 9696
Carbonic anhydrase III

Edith Sim
Department of Pharmacology
University of Oxford
Mansfield Road
Oxford OX1 3QU
United Kingdom
Phone: +44 865 271 596
FAX: +44 865 271 853
E-mail: esim@morbioc.ox.ac.uk.
Complement receptor type 1

Robert B. Sim
MRC, Immunochemistry Unit
Department of Biochemistry
University of Oxford
South Parks Road
Oxford OX1 3QU
United Kingdom
Phone: +44 865 275 351
FAX: +44 865 275 729
E-mail: rbsim@bioch.ox.ac.uk
Collectin receptor
Factor H

Peter J. Sims
The Blood Research Institute
The Blood Center of Southeastern Wisconsin
P.O. Box 2178
Milwaukee, WI 53201-2178
U.S.A.
Phone: +1 414 937 3862
FAX: +1 414 937 6284
CD 59

Annalisa Siri
Istituto Nazionale per la Ricerca sul Cancro
Laboratory of Cell Biology
Viale Benedetto XV, 10
I-16132 Genova
Italy
Phone: +39 10 353 4901
FAX: +39 10 352 855
E-mail: lzardi@cisi.unige.it
Tenascin

Hans Sjöström
Panum Institute
Department of Biochemistry C
University of Copenhagen
Blegdamsvej 3C
DK-2200 Copenhagen N
Denmark
Phone: +45 31 357 900
FAX: +45 35 367 980
Aminopeptidase N

Frantisek V. Skvaril
Pelikanweg 34
CH-3074 Muri b. Bern
Switzerland
Phone: +41 31 951 48 04
FAX: +41 31 371 40 44 c/o Labor Weissenbühl
Immunoglobulin G

William S. Sly
Department of Biochemistry and Molecular
Biology
St. Louis University
Health Siences Center
School of Medicine
1402 S. Grand Blvd.
St. Louis, MO 63104
U.S.A.
Phone: +1 314 577 8131
FAX: +1 314 776 1183
E-mail: slyws@wpogate.slu.edu
Carbonic anhydrase IV

Michael F. Smith
University of Virginia
Division of Gastroenterology
Building MR4, Box 1031
Charlottesville, VA 22908
U.S.A.
Phone: +1 804 924 2573
FAX: +1 804 243 6169
E-mail: mfs3k@virginia.edu
Interleukin-1 receptor antagonist

James M. Sodetz
Department of Chemistry and Biochemistry
University of South Carolina
Columbia, SC 29208
U.S.A.
Phone: +1 803 777 6625
FAX: +1 803 777 9521
E-mail: sodetz@psc.sc.edu
C8 complement protein

Lars Sottrup-Jensen
Department of Molecular Biology
Bldg. 130 University of Aarhus
DK-8000 Aarhus C
Denmark
Phone: +45 86 125 177
FAX: +45 86 196 226
E-mail: Lsj@mbio.aan.dk
Alpha-2-macroglobulin

Peter J. Späth
ZLB Central Laboratory
Blood Transfusion Service
Swiss Red Cross
CH-3000 Bern 22
Switzerland
Phone: +41 31 330 02 16
C1-inhibitor

Geneviève Spik
Laboratoire de Biochimie
Université des Sciences et Technologies de Lille
F-59655 Villeneuve d'Ascq Cédex
France
Phone: +33 20 434 155
FAX: +33 20 436 555
Lactoferrin
Cyclophilin-20

Jerry L. Spivak
Johns Hopkins University School of Medicine
720 Rutland Avenue
Traylor 924
Baltimore, MD 21205
U.S.A.
Phone: +1 410 955 5454
FAX: +1 410 955 0185
E-mail: jlspivak@welchlink.welch.jhu.edu
Erythropoietin
Erythropoietin receptor

Erwin E. Sterchi
Institute of Biochemistry and Molecular Biology
University of Bern
Bühlstrasse 28
CH-3000 Bern 9
Switzerland
Phone: +41 31 631 41 11
FAX: +41 31 631 37 37
E-mail: sterchi@mci.unibe.ch
Sucrase-isomaltase

Alvin S. Stern
Hoffmann-La Roche, Inc.
340 Kingsland Street
Nutley, NJ 07110-1199
U.S.A.
Phone: +1 201 235 5129
FAX: +1 201 235 3805
Interleukin 12

Dudley K. Strickland
The Jerome H. Holland Laboratories
Department of Biochemistry
American Red Cross
15601 Crabbs Branch Way
Rockville, MD 20855
U.S.A.
Phone: +1 301 738 0726
FAX: +1 301 738 0794
Alpha-2-macroglobulin receptor

Koichi Suzuki
Institute of Molecular and Cellular Biosciences
University of Tokyo
1-1-1Yayoi, Bunkyo-ku
Tokyo 113
Japan
Phone: +81 3 3812 2111
FAX: +81 3 3813 0654
Calpain

Koji Suzuki
Department of Molecular Pathobiology
Mie University School of Medicine
Tsu-city, Mie 514
Japan
Phone: +81 592 315 036
FAX: +81 592 315 209
E-mail: suzuki@doc.medic.mie-u.ac.jp
Protein C inhibitor

Brian D. Sykes
Department of Biochemistry
University of Alberta
Edmonton, Alberta T6G 2H7
Canada
Phone: +1 403 492 5460
FAX: +1 403 492 0886
E-mail: Brian.Sykes@ualberta.ca
Troponin C

Kazuhiko Takahashi
Dept. Hygienic Chemistry
Fac. Pharmaceutical Sciences
Hokkaido University
Kita-ku, Sapporo 060
Japan
Phone: +81 11 706 3244
FAX: +81 11 706 4989
Glutathione peroxidase, cellular
Glutathione peroxidase, extracellular

Kenji Takahashi
School of Life Science
Laboratory of Molecular Biochemistry
Tokyo University of Pharmacy and Life Science
Horinouchi, Hachioji
Tokyo 192-03
Japan
Phone: +81 426 76 7146
FAX: +81 426 76 7149
E-mail: kenjitak@ls.toyaku.ac.jp
Pepsinogen

Kay Tanaka
Department of Human Genetics
Yale University School of Medicine
333 Cedar Street
New Haven, CT 06510
U.S.A.
Phone: +1 203 785 2659
FAX: +1 203 785 3363
E-mail: tanakak@biomed.med.yale.edu
Acyl-CoA dehydrogenases

Gerald Thiel
Institute for Genetics
University of Köln
Zülpicherstrasse 47
D-50674 Köln
Germany
Phone: +49 221 470 4847
FAX: +49 221 470 5172
Ca^{2+}/calmodulin dependent protein kinase II
Synapsin I, Synapsin II

J. Terrig Thomas
Department of Biochemistry and Molecular Biology
University of Manchester
School of Biological Sciences
Stopford Building
Oxford Road
Manchester M13 9PT
United Kingdom
Phone: +44 61 275 5102
FAX: +44 61 275 5082
Collagen type X

Paul J. Thornalley
Dept. of Biological and Chemical Sciences
University of Essex
Central Campus Wivenhue Park
Colchester CO4 3SQ
United Kingdom
Phone: +44 206 873 010
FAX: +44 206 872 592
E-mail: thorp@essex.ac.uk
Glyoxalase I, Glyoxalase II

Douglas M. Tollefsen
Hematology Division
Washington University Medical School
660 South Euclid Avenue
P.O. Box 8125
St. Louis, MO 63110
U.S.A.
Phone: +1 314 362 8830
FAX: +1 314 362 8826
E-mail: tollefsen@visar.wustl.edu
Heparin cofactor II

Birgitta Tomkinson
Biomedical Center
Department of Veterinary Medical Chemistry
Swedish University of Agricultural Sciences
P.O. Box 575
S-75123 Uppsala
Sweden
Phone: +46 18 174 158
FAX: +46 18 550 762
E-mail: birgitta.tomkinson@vmk.slu.se
Tripeptidyl peptidase II

Beat Trueb
M.E. Müller Institut für Biomechanik
University of Bern
P.O. Box 30
CH-3010 Bern
Switzerland
Phone: +31 632 87 26
FAX: +31 632 49 99
E-mail: trueb@mem.unibe.ch
Collagen Type VI

Harald Tschesche
Universität Bielefeld
Fakultät für Chemie
Universitätsstr. 25
D-33615 Bielefeld
Germany
Phone: +49 521 106 2081
FAX: +49 521 106 6014
E-mail: Tschesche@chema.uni-bielefeld.de
Neutrophil collagenase
Neutrophil lipocalin

Jürg Tschopp
Institute of Biochemistry
University of LaU.S.A.nne
155, Chemin des Boveresses
CH-1066 Epalinges
Switzerland
Phone: +41 21 316 57 38
FAX: +41 21 316 57 05
C9 complement protein
Clusterin

Roderick B.C. Tse
The Hospital for Sick Children
Research Institute
Division of Neurosciences
555, University Avenue
9th Floor, Elm Wing, Room 9142
Toronto, Ontario M5G 1X8
Canada
Phone: +1 416 813 5766
FAX: +1 416 813 5086
β-N-acetyl-D-hexosaminidase

William T. Tse
Division of Hematology/Oncology
Children's Hospital, Enders 7
300 Longwood Avenue
Boston, MA 02115
U.S.A.
Phone: +1 617 355-7904
FAX: +1 617 355 7262
E-mail: tsew@a1.tch.harvard.edu
Ankyrin

Philip W. Tucker
Institute for Cell. and Mol. Biology
University of Texas
Austin, TX 78712
U.S.A.
Phone: +1 512 475 7705
FAX: +1 512 475 7707
E-mail: philtucker@mail.utexas.edu
Immunoglobulin D

Bert L. Vallee
CBBSM, Mudd 106
Harvard Medical School
250 Longwood Avenue
Boston, MA 02115
U.S.A.
Phone: +1 617 432 1367
FAX: +1 617 566 3137
E-mail: bvallee@warren.med.harvard.edu
Angiogenin

Michel van der Rest
Institut de Biologie Structurale
41, Avenue des Martyrs
F-38027 Grenoble Cedex 1
France
Phone: +33 7688 3482
FAX: +33 7688 9484
E-mail: mv@Ibs.fr
Collagen Introduction
Collagen type I
Collagen type II
Collagen type V
Collagen type XI

Greet C. Vanhoof
Janssen Pharmaceutica N.V.
Turnhoutseweg, 30
B-2340 Beerse
Belgium
Phone: +32 14 60 6183
FAX: +32 14 60 6111
E-mail: gvanhoof@janbe.jnj.com
Aminopeptidase P
Prolyl oligopeptidase

Robert L. Van Etten
Department of Chemistry
Purdue University
West Lafayette, IN 47907-1393
U.S.A.
Phone: +1 317 494 5274
FAX: +1 765 494 0239
E-mail: vanetten©chem.purdue.edu
Erythrocyte Acid Phosphatase

Raghavan Varadarajan
Molecular Biophysics Unit
Indian Institute of Science
Bangalore - 560012
India
Phone: +91 80 309 2612
FAX: +91 80 334 1683
E-mail: varadar@mbu.iisc.ernet.in
Myoglobin

Cees Vermeer
Department of Biochemistry
University of Limburg
P.O. Box 616
NL-6200 MD Maastricht
The Netherlands
Phone: +31 43 388 1682
FAX: +31 43 367 0992
Vitamin K-dependent carboxylase
Vitamin K epoxide reductase

Thomas Vorherr
Laboratorium für Biochemie
ETH-Zentrum
Universitätsstrasse 16
CH-8092 Zürich
Switzerland
Phone: +41 1 256 30 11
FAX: +41 1 252 63 23
Calcium adenosinetriphosphatase (plasma membrane)

Abdul Waheed
St. Louis University
Health Sciences Center
Department of Biochemistry
and Molecular Biology
School of Medicine
1402 S. Grand Blvd.
St. Louis, MO 63104
U.S.A.
Phone: +1 314 577 8131
FAX: +1 314 776 1183
Carbonic anhydrase IV

Nigel P.C. Walker
Hauptlaboratorium - A30
BASF AG
Carl -Bosch-Str. 38
D-67056 Ludwigshafen
Germany
Phone: +49 621 604 2638
FAX: +49 621 602 0440
Tumor necrosis factor

Peter N. Walsh
Thrombosis Research Center
Department of Medicine and Biochemistry
Temple University School of Medicine
3400 North Broad Street
Philadelphia, PA 19140
U.S.A.
Phone: +1 215 707 4375
FAX: +1 215 707 3005
E-mail: pnw@astro.ocis.temple.edu
Factor IX
Factor XI

Mark R. Walter
University of Alabama at Birmingham
Center for Macromolecular Crystallography
268 Basic Health Sciences
1918 University Boulevard
Birmingham, AL 35294-0005
U.S.A.
Phone: +1 205 934 9279
FAX: +1 205 934 0480
E-mail: walter@mimas.cmc.uab.edu
Interleukin-10

Alfred Walz
Theodor-Kocher-Institut
University of Bern
Freistrasse 1
CH-3000 Bern 9
Switzerland
Phone: +41 31 6314166
FAX: +41 31 6313799
Interleukin 8

Cindy Wang
Department of Biochemistry and
Molecular Biology
Medical University of South Carolina
171 Ashley Avenue
Charleston, SC 29425-2211
U.S.A.
Phone. +1 803 792 4321
FAX: +1 803 792 1627
E-mail: wangc@musc.edu
Tissue Kallikrein

Lee-Ho Wang
Division of Hematology
Department of Internal Medicine
University of Texas School of Medicine
Houston, TX 77030
U.S.A.
Thromboxane synthase

Roberta J. Ward
Department of Clinical Biochemistry
Kings College
London
United Kingdom
Ferritin
Haemosiderin

Robin A. Warren
Department of Cell Biology L215
Oregon Health Sciences University
3181 SW Sam Jackson Park Rd.
Portland, OR 97201
U.S.A.
Phone: +1 503 494-5845
FAX: +1 503 494 4253
E-mail: warrenr@ohsu.edu
Transferrin receptor

Wilma Wasco
Genetics and Aging Unit
Massachusetts General Hospital
Harvard Medical School
149 13th Street
Charlestown, MA 02129-9142
U.S.A.
Phone: +1 617 726 8307
FAX: +1 617 726 5677
E-mail: wasco@helix.mgh.harvard.edu
Amyloid β-protein precursor

Anthony Watts
University of Oxford
Biochemistry Department
South Parks Road
Oxford OX1 3QU
United Kingdom
Phone: +44 1865 275 268
FAX: +44 1865 275 234
Cytochrome C

Richard B. Weinberg
Medical Center Boulevard
The Bowman Gray School of Medicine
Winston-Salem, NC 27157
U.S.A.
Phone: +1 910 716 4612
FAX: +1 910 716 6376
E-mail: weinberg@bgsm.cdu
Apolipoprotein A-IV

Bendicht Wermuth
Department of Clinical Chemistry
University of Bern
Inselspital
CH-3010 Bern
Switzerland
Phone: +41 31 632 29 85
FAX: +41 31 632 48 62
E-mail: bwermuth@insel.ch
Alcohol dehydrogenase
Carbonyl reductase
Ornithine carbomoyltransferase

Rick A. Wetsel
Institute of Molecular Medicine
The University of Texas
Health Science Center-Houston
2121 W. Holcombe Blvd.
Houston, TX 77030
U.S.A.
Phone: +1 713 792 5495
FAX: +1 713 745 1929
C5 complement protein

David A. Williams
Herman B Wells Center for Pediatric Research
Howard Hughes Medical Institute
Research Laboratories
Indiana University School of Medicine
Cancer Research Building
1044 W. Walhut St., Room 402c
Indianapolis, IN 46202-5225
U.S.A.
Phone: +1 317 274 8960
FAX: +1 317 274 8679
E-mail: dwilliam@indyvax.iupui.edu
Interleukin-11

Howard Wong
Lipid Research
VA Wadsworth Medical Center
Bldg 113, Room 312
Los Angeles, CA 90073
U.S.A.
Phone: +1 310 268 3927
FAX: +1 310 478 4538
Hepatic lipase
Lipoprotein lipase

Savio L.C. Woo
Room T-721
Department of Cell Biology
Baylor College of Medicine
One Baylor Plaza
Houston, TX 77030
U.S.A..
Phone: +1 713 798 6080
FAX: +1 713 790 1211
Phenylalanine hydroxylase

Walter A. Wuillemin
Inselspital
Central Hematology Laboratory
University of Bern
CH-3010 Bern
Switzerland
Phone: +41 31 632 94 14
FAX: +41 31 632 93 66
Factor XII

Yoko Yamagata
Laboratory of Neurochemistry
National Institute for Physiological Sciences
Myodaiji, Okazaki 444
Japan
Phone: +81 564 557 822
FAX: +81 564 527 913
Ca^{2+}/calmodulin-dependent protein kinase II

Jack X. Yu
Department of Biochemistry and
Molecular Biology
Medical University of South Carolina
171 Ashley Avenue
Charleston, SC 29425-2211
U.S.A.
Phone. +1 803 792 4321
FAX: +1 803 792 1627
Prostasin

Elèni E. Zanni
Section of Molecular Genetics
Whitaker Cardiovascular Institute
Boston University Medical Center
Center for Advanced Biomedical Research
700 Albany Street, W-509
Boston, MA 02118-2394
U.S.A.
Phone: +1 617 638 5085
FAX: +1 617 638 5141
Apolipoprotein CII
Apolipoprotein CIII

Vassilis I. Zannis
Section of Molecular Genetics
Whitaker Cardiovascular Institute
Boston University Medical Center
Center for Advanced Biomedical Research
700 Albany Street, W-509
Boston, MA 02118-2394
U.S.A.
Phone: +1 617 638 5085
FAX: +1 617 638 5141
Apolipoprotein CI and CIV
Apolipoprotein CII
Apolipoprotein CIII

Luciano Zardi
Istituto Nazionale per la Ricerca sul Cancro
Laboratory of Cell Biology
Viale Benedetto XV, 10
I-16132 Genova
Italy
Phone: +39 10 352 855
FAX: +39 10 352 999
Tenascin

Dieter R. Zimmermann
Institut für Klinische Pathologie
Universitätsspital
Schmelzbergstrasse 12
CH-8091 Zürich
Switzerland
Phone: +41 1 255 39 45
FAX: +41 1 255 45 08
Versican

Volker Zölzer
Universität Bielefeld
Universitätsstr. 25
D-33615 Bielefeld
Germany
Neutrophil lipocalin

Synonyms and Abbreviations Installments 1 to 6

Synonym, Abbreviation	See under (if no listing here, then see left column)
1-pyrroline-5-carboxulate dehydrogenase	Glutamic-γ-semialdehyde dehydrogenase, liver mitochondria
1,4-D-glucan:orthophosphate D-glucosyltransferase	Glycogen phosphorylase
2-sulpho-L-iduronate-2-sulphatase	Iduronate-2-sulphatase
2.5S NGF	Nerve growth factor
4-sulphate sulphatase	N-acetylgalactosamine-4-sulphatase
4-sulpho-N-acetylgalactosamine sulphatase	N-acetylgalactosamine-4-sulphatase
4-sulphatase	N-acetylgalactosamine-4-sulphatase
4-S	N-acetylgalactosamine-4-sulphatase
9.5S α_1-glycoprotein	Serum amyloid P component
11S protein	C1 q complement protein
19S γ-globulin	Immunoglobulin M
20 kDa homologous restriction factor (HRF20)	CD 59
26 kDa protein	Interleukin 6
43K protein	Osteonectin
48 kDa protein	Arrestin
48 K	Arrestin
150 k calmodulin binding protein from chicken gizzard	Caldesmon
α1(VI)α2(VI)α3(VI)	Collagen type VI
α1(IX)	Collagen type IX
$\alpha 1(X)]_3$	Collagen type X
α1α2α3	Collagen type XI
α1-Achy	Alpha-1-antichymotrypsin
α1-ACT	Alpha-1-antichymotrypsin
α_1-antitrypsin	Alpha-1-proteinase inhibitor
α_1-AT	Alpha-1-proteinase inhibitor
α_1-m	Alpha-1-microglobulin
α_1m	Alpha-1-microglobulin
α1m	Alpha-1-microglobulin
α-1-microglobulin	Alpha-1-microglobulin
α_1-microglycoprotein	Alpha-1-microglobulin
α_1-PEG	Insulin-like growth factor binding protein-1
α-1-proteinase inhibitor	Alpha-1-proteinase inhibitor
α_1-PI	Alpha-1-proteinase inhibitor
α2 (IX)	Collagen type IX
α_2-AP	Alpha-2-plasmin inhibitior
α-2-antiplasmin	Alpha-2-plasmin inhibitor
α_2-HS	Alpha2-HS glycoprotein
α_2-HS glycoprotein	
α_2M	Alpha-2-macroglobulin
α_2MR	Alpha-2-macroglobulin receptor
α_2-macroglobulin	$Alpha_2$-macroglobulin
α-2-macroglobulin receptor	Alpha-2-macroglobulin receptor
α_2-microglobulin-related protein	Neutrophil lipocalin
α_2-neuramino-glycoprotein	C1-inhibitor
α_2-PI	Alpha-2-plasmin inhibitor
α_2-plasmin inhibitor	Alpha-2-plasmin inhibitor
α3 (IX)	Collagen type IX

α-amylase	Alpha-amylase
α-cysteine proteinase inhibitor	Kininogens
α-glucan phosphorylase	Glycogen phosphorylase
α-L-idopyranosyluronic acid 2-sulphate sulphohydrolase	Iduronate-2-sulphatase
α-L-iduronidase	α-L-iduronidase
α-L-iduronide iduronohydrolase	Kininogens
α-thiol proteinase inhibitor	Tryptase
α, β-tryptase	Amyloid β-protein precursor
β-amyloid precursor protein	Amyloid β-protein precursor
β-amyloid protein precursor	Beta 1,4-galactosyltransferase
β-1,4-galactosyltransferase	Beta 1,4-galactosyltransferase
β1,4-gal-T	Hemopexin
β_1-glycoprotein	C3 complement protein
β1c-globulin	C3 complement protein
β1c-protein	C4 complement protein
β_1 E globulin	Factor H
β_1 H	Immunoglobulin A
β_{2A}-globulin	Beta-2-microglobulin
β_2-m	Beta-2-microglobulin
β_2m	Beta-2-microglobulin
β2m	Beta-2-microglobulin
beta-2-m	Apolipoprotein B-100
β-apolipoprotein	Glucocerebrosidase
β-Glc	Lactase-phlorizin hydrolase
β-glycosidase complex	Plasminogen activator inhibitor type-1
β-migrating plasminogen activator inhibitor	β-N-acetyl-D-hexosaminidase
β-N-acetyl-D-hexosaminidase	Nerve growth factor
β-N-acetylglucosaminidase	Nerve growth factor
β-nerve growth factor	Amyloid β-protein precursor
β-NGF	Platelet basic protein
βPP	Platelet basic protein
β-TG (β-thromboglobulin)	Immunoglobulin G
β-thromboglobulin (β-TG)	Immunoglobulin A
γ_1-globulin	Immunoglobulin G
γ_{1A}-globulin	Immunoglobulin A
γ_2-globulin	Aldehyde dehydrogenase, E3, liver cytoplasm
γA	Cystatin C
γ-aminobutyraldehyde dehydrogenase	Cystatin C
γ c globulin	Immunoglobulin G
γCSF	Immunoglobulin G
γG	Vitamin K-dependent carboxylase
γ-globulin	Aldehyde dehydrogenase, E3, liver cytoplasm
γ-glutamyl-carboxylase	Immunoglobulin M
γ-guanidinobutyraldehyde dehydrogenase	Immunoglobulin M
γM	Prostate specific antigen
γ-macroglobulin	Cystatin C
γ-seminoprotein	Cystatin C
γ-trace	Alpha2-HS Glycoprotein
γaT	Cytochrome C oxidase
A2HS.human (Swiss Prot)	Aspartate aminotransferase
aa_3	Aspartate aminotransferase
AAT	
AAtase	

A-B collagen	Collagen type V
Abnormal protein kinase C	Protein kinase C
Accelerator globulin	Factor V
ACE	Angiotensin converting enzyme
Acetylcholinesterase	
AChE	Acetylcholinesterase
Acid β-glucosidase	Glucocerebrosidase
Acid cysteine proteinase inhibitor	Cystatin A
Acid stable subunit of the 140kDa IGF complex	Insulin-like growth factor binding protein-3
ACP1	Erythrocyte acid phosphatase
ACPI	Cystatin A
Acyl-CoA: (acceptor) oxidoreductase	Acyl-CoA dehydrogenase
Acyl-CoA dehydrogenase	
ADA-bp	CD 26
ADA-cp	CD 26
Adenosine 5'-trisphosphate-creatine phosphotransferase	Creatine kinase
Adenosine deaminase binding protein	CD 26
Adenosine-deaminase complexing protein	CD 26
AD	Acyl-CoA dehydrogenase
ADH	Alcohol dehydrogenase
Adipogenesis inhibitory factor	Interleukin 11
Adipsin	Factor D
ADR	Aldose reductase
AE1	Band 3 protein (AE1)
AFBP (amniotic fluid binding protein)	Insulin-like growth factor binding protein-1
AHF	Factor VIII
AHG	Factor VIII
AHSG	Alpha2-HS Glycoprotein
AKR1A1	Aldehyde reductase
AKR1B1	Aldose reductase
Alanine aminopeptidase	Aminopeptidase N
Alanyl aminopeptidase	Aminopeptidase N
Alb	Albumin
Albumin	
Alcohol dehydrogenase	
Aldehyde dehydrogenase, E1, liver cytoplasm	
Aldehyde dehydrogenase, E2, liver mitochondria	
Aldehyde dehydrogenase, E3, liver cytoplasm	
Aldehyde dehydrogenase, stomach cytoplasm	
Aldehyde oxidase	
Aldehyde reductase	
ALDH 1	Aldehyde dehydrogenase, E1, liver cytoplasm
ALDH 2	Aldehyde dehydrogenase,E2, liver mitochondria
ALDH 3	Aldehyde dehydrogenase, stomach cytoplasm
ALDH I	Aldehyde dehydrogenase, E2, liver mitochondria
ALDH II	Aldehyde dehydrogenase, E1, liver cytoplasm
ALDH III	Aldehyde dehydrogenase, stomach cytoplasm
ALDH IV	Glutamic-γ-semialdehyde dehydrogenase, liver mitochondria
Aldolase	Fructose-1,6-biphosphate aldolase
Aldose reductase	
Alpha-1-antichymotrypsin	

Alpha-1-antitrypsin	Alpha-1-proteinase inhibitor
Alpha-1-AT	Alpha-1-proteinase inhibitor
Alpha-1-m	Alpha-1-microglobulin
Alpha-1-microglobulin	
Alpha-1-microglycoprotein	Alpha-1-microglobulin
$Alpha_1$-PI	Alpha-1-proteinase inhibitor
Alpha-1-proteinase inhibitor	
Alpha-2-antiplasmin	Alpha-2-plasmin inhibitor
$Alpha_2$-AP	Alpha-2-plasmin inhibitior
Alpha2-HS glycoprotein	
Alpha-2-macroglobulin	
Alpha-2-macroglobulin receptor	
Alpha-2-neuramino-glycoprotein	C1-inhibitor
$Alpha_2$-PI	Alpha-2-plasmin inhibitor
Alpha-2-plasmin inhibitor	
Alpha-aminoacyl-peptide hydrolase	Aminopeptidase N
Alpha-amylase	
Alpha,beta-tryptase	Tryptase
Alpha-L-idopyranosyluronic acid 2-sulphate sulphohydrolase	Induronate-2-sulphatase
Alpha-L-iduronidase	α-L-iduronidase
Alpha-L-iduronide iduronohyrolase	α-L-iduronidase
ALR1	Aldehyde reductase
ALR2	Aldose reductase
Aminoacyl-prolyl-peptide hydrolase	Aminopeptidase P
Amino-oligopeptidase	Aminopeptidase N
Aminopeptidase M	Aminopeptidase N
Aminopeptidase N	
Aminopeptidase P	
Amniotic fluid binding protein	Insulin-like growth factor binding protein-1
AMPEPN	Aminopeptidase N
AmP	Aminopeptidase P
AMPP	Aminopeptidase P
Amy	Alpha-amylase
amy1	Alpha-amylase
amy2	Alpha-amylase
amy2A	Alpha-amylase
amy2B	Alpha-amylase
amy3	Alpha-amylase
Amyloid β-protein precursor	
Amyloid precursor protein	Amyloid β-protein precursor
Amyloid protein precursor	Amyloid β-protein precursor
Ang	Angiogenin
Anaphylatoxin inhibitor	Carboxypeptidase N
Angiogenin	
Angiotensin converting enzyme	
Anion exchange protein	Band 3 protein (AE1)
Anionic trypsin(ogen)	Trypsin(ogen) 2
Ank1	Ankyrin
Ankyrin	
Ankyrin 1	Ankyrin
$Ankyrin_R$	Ankyrin
Annexin I	Lipocortin I
Anticonvertin	Tissue factor pathway inhibitor
Anti-gelatin factor	Fibronectin
Antihemophilic factor	Factor VIII
Antihemophilic factor B	Factor IX
Antihemophilic globulin	Factor VIII

Antithrombin	
Antithrombin BM	Heparin cofactor II
Antithromboplastin	Tissue factor pathway inhibitor
Antitrypsin	Alpha-1-proteinase inhibitor
Apase M	Aminopeptidase N
aPKC	Protein kinase C
APN	Aminopeptidase N
Apo(a)	Apolipoprotein(a)
ApoA-I	Apolipoprotein A-I
ApoA-II	Apolipoprotein A-II
ApoA-IV	Apolipoprotein A-IV
ApoB-48	Apolipoprotein B-48
ApoB-100	Apolipoprotein B-100
ApoB mRNA Editase	Apolipoprotein B messenger RNA editing protein
ApoCI	Apolipoprotein CI
ApoCII	Apolipoprotein CII
ApoCIII	Apolipoprotein CIII
ApoD	Apolipoprotein D
apo-Gln-I	Apolipoprotein A-I
apo-Gln-II	Apolipoprotein A-II
apo-LP-A-I	Apolipoprotein A-I
apo-LP-A-II	Apolipoprotein A-II
ApoE	Apolipoprotein E
apoJ	Clusterin
ApoLDL	Apolipoprotein B-100
ApoLf (iron-free)	Lactoferrin
apoLp-Ala	Apolipoprotein CIII
apoLp-Glu	Apolipoprotein CII
apoLp-Ser	Apolipoprotein CI
apopec1 (apoB mRNA editing component 1)	Apolipoprotein B messenger RNA editing protein
apoSAA	Serum amyloid A
apoTf (iron-free)	Transferrin
Apoferritin (iron-free)	Ferritin
Apolipoprotein(a)	
Apolipoprotein alanine (apoLp-Ala)	Apolipoprotein CIII
Apolipoprotein A-I	
Apolipoprotein A-II	
Apolipoprotein A-IV	
Apolipoprotein B-48	
Apolipoprotein B-100	
Apolipoprotein B messenger RNA editing protein	
Apolipoprotein Bs	Apolipoprotein B-48
Apolipoprotein CI	
Apolipoprotein CII	
Apolipoprotein CIII	
Apolipoprtoein CIV	
Apolipoprotein D	
Apolipoprotein E	
Apolipoprotein-Gln-I	Apolipoprotein A-I
Apolipoprotein-Gln-II	Apolipoprotein A-II
Apolipoprotein glutamine (apoLp-Gln)	Apolipoprotein CII
Apolipoprotein J	Clusterin
Apolipoprotein serine (apoLp-Ser)	Apolipoprotein CI and Apolipoprotein CIV
APP	Aminopeptidase P
APP	Amyloid β-protein precursor
AQP1	Aquaporin-1

Aquaporin-1	
Arginine carboxypeptidase	Carboxypeptidase N
Arginine procarboxypeptidase	Procarboxypeptidase U
Arginine-rich apoprotein	Apolipoprotein E
AR1	Aldehyde reductase
ARP	Apolipoprotein E
Arr	Arrestin
Arrestin	
Arylamine N-acetyltransferase	
Arylsulphatase B	N-Acetylgalactosamine-4-sulphatase
ASAT	Aspartate aminotransferase
ASB	N-Acetylgalactosamine-4-sulphatase
Aspartate aminotransferase	
Aspartate transaminase	Aspartate aminotransferase
AspAT	Aspartate aminotransferase
AST	Aspartate aminotransferase
AT-III	Antithrombin
Atmungsferment	Cytochrome C oxidase
Atypical protein kinase C	Protein kinase C
Autoprothrombin II-A	Protein C
Autoprothrombin III	Factor X
Autorosette inhibition factor	Histidine-rich glycoprotein
B	Factor B
BAF (B cell activating factor)	Interleukin 1
Band 3 protein (AE1)	
BCDF-γ	Interleukin 4
B cell activating factor (BAF)	Interleukin 1
B cell differentiation factor for IgG1	Interleukin 4
B cell growth factor II (murine)	Interleukin 5
B cell stimulatory factor (BSF-2)	Interleukin 6
B cell stimulatory factor-1 (BSF-1)	Interleukin 4
BCHE	Butyrylcholinesterase
BChE	Butyrylcholinesterase
Beta-1-glycoprotein	Hemopexin
Beta-1c-globulin	C3 complement protein
Beta-1c-protein	C3 complement protein
Beta-1E globulin	C4 complement protein
Beta1,4-gal-T	Beta 1,4-galactosyltransferase
Beta 1,4-galactosyltransferase	
beta-2-m	Beta-2-microglobulin
Beta-2-microglobulin	
Beta-apolipoprotein	Apolipoprotein B-100
Beta-Glc	Glucocerebrosidase
Beta-glycosidase complex	Lactase-phlorizin hydrolase
Beta-migrating plasminogen activator inhibitor	Plasminogen activator inhibitor type 1
Beta-N-acetyl-D-hexosaminidase	β-N-acetyl-D-hexosaminidase
Beta-N-acetylglucosaminidase	β-N-acetyl-D-hexosaminidase
Beta-nerve growth factor	Nerve growth factor
Beta-NGF	Nerve growth factor
Beta-thromboglobulin	Platelet basic protein
BHCP (human C-peptide, synthetic)	Insulin and proinsulin
Binding protein 24	Insulin-like growth factor binding protein-4
Bile salt-acitvated lipase	Carboxyl ester lipase
Bile salt-stimulated lipase	Carboxyl ester lipase
Biliverdin reductase	
BP-24	Insulin-like growth factor binding protein-4
BP-25	Insulin-like growth factor binding protein-1

BP-28	Insulin-like growth factor binding protein-1
BP-29	Insulin-like growth factor binding protein-3
BP-53	Insulin-like growth factor binding protein-3
Blood coagulation factor XIII	Plasma factor XIII
Blood plasminogen activator	Tissue-type plasminogen activator
BM-40 (basement membrane)	Osteonectin
bNOS	Nitric oxide synthase
BPA (Burst Promoting Activity)	Interleukin 3
BSF-1 (B cell stimulatory factor-1)	Interleukin 4
Brain-type facilitative glucose transport protein	Glucose transport protein 3
BSF-2 (B cell stimulatory factor-2)	Interleukin 6
BSSL	Carboxyl ester lipase
BuChE	Butyrylcholinesterase
Burst promoting activity (BPA)	Interleukin 3
Butyrylcholinesterase	
BVR	Biliverdin reductase
C1-bar-esterase inhibitor	C1-inhibitor
C1(bar)-INH	C1-inhibitor
C1(bar)-inhibitor	C1-inhibitor
C1-INA	C1-inhibitor
C(bar)1-INH	C1-inhibitor
C1-INH	C1-inhibitor
C1-inhibitor	
C1q	C1q complement protein
C1q complement protein	
C1qR	Collectin receptor
C1q receptor	Collectin receptor
C1r (activated form)	C1r complement protein
C1r (proenzyme form)	C1r complement protein
C1r complement protein	
C1s(bar)-inhibitor	C1-inhibitor
C1s-inhibitor (C1-inactivator)	C1-inhibitor
C3	C3 complement protein
C3b-inactivator	Factor I
C3b-inactivator accelerator	Factor H
C3b/C4b inactivator	Factor I
C3b/C4b receptor	Complement receptor type 1
C3 complement protein	
C3PA	Factor B
C3 proactivator	Factor B
C3 proactivator convertase	Factor D
C4	C4 complement protein
C'4	C4 complement protein
C4-binding protein	C4b-binding protein
C4b-binding protein	
C4b-INA cofactor	C4b-binding protein
C4BP	C4b-binding protein
C4 complement protein	
C5	C5 complement protein
C5 complement protein	
C6	C6 complement protein
C6 complement protein	
C7	C7 complement protein
C7 complement protein	
C8	C8 complement protein
C8 complement protein	
C9	C9 complement protein

C9 complement protein	
CA I	Carbonic anhydrase I
CA II	Carbonic anhydrase II
CA III	Carbonic anhydrase III
CA IV	Carbonic anhydrase IV
$Ca^{2+}$and phospholipid-dependent protein kinase	Protein kinase C
Ca^{2+} ATPase	Calcium adenosinetriphosphatase (plasma membrane)
Ca^{2+}/calmodulin-dependent protein kinase II	
Ca^{2+} pump	Calcium adenosinetriphosphatase (plasma membrane)
CA B	Carbonic anhydrase I
CaD (h-CaD, 1-CaD)	Caldesmon
Cachectin	Tumor necrosis factor
CAF (calcium activated factor)	Calpain
Cal_l	Caldesmon
Cal_h	Caldesmon
Calcitonin	Calcitonin and procalcitonin
Calcitonin and procalcitonin	
Calcitonin gene-related peptide	
Calcium activated factor (CAF)	Calpain
Calcium activated neutral protease (CANP)	Calpain
Calcium activated sarcoplasmic factor (CASF)	Calpain
Calcium adenosinetriphosphatase (plasma membrane)	
Calcium-dependent muscle regulatory protein	Troponin C
Calcium dependent protease (CDP)	Calpain
Calcium-dependent protein kinase C	Protein kinase C
Calcium-dependent regulatory protein	Calmodulin
Calcium protease	Calpain
Caldesmon	
CALLA	Endopeptidase-24.11
Calmodulin	
Calpactin II	Lipocortin I
Calpain	
CaM	Calmodulin
CaM kinase II	Ca^{2+}/calmodulin-dependent protein kinase II
CaM-PK II	Ca^{2+}/calmodulin-dependent protein kinase II
CANP (calcium activated neutral protease)	Calpain
Capnophorin	Band 3 protein (AE1)
Carbonate dehydratase (isoenzyme I)	Carbonic anhydrase I
Carbonate dehydratase (isoenzyme II)	Carbonic anhydrase II
Carbonate dehydratase (isoenzyme III)	Carbonic anhydrase III
Carbonate dehydratase (isoenzyme IV)	Carbonic anhydrase IV
Carbonate hydrase IV	Carbonic anhydrase IV
Carbonate hydro-lyase (isoenzyme I)	Carbonic anhydrase I
Carbonic anhydrase I	
Carbonic anhydrase II	
Carbonic anhydrase III	
Carbonic anhydrase IV	
Carbonic anhydrase B	Carbonic anhydrase I
Carbonyl reductase	
Carboxylase	Vitamin K-dependent carboxylase
Carboxyl ester hydrolase	Carboxyl ester lipase
Carboxyl ester lipase	
Carboxypeptidase N	
Cartilage oligomeric matrix protein	

CASF (calcium activated sarcoplasmic factor)	Calpain
Catabolin	Interleukin 1
Catalase	
Catechol-O-methyltransferase	
Cat G	Leukocyte cathepsin G
Cationic trypsin(ogen)	Trypsin(ogen) 1
CBG	Cortisol binding globulin
Cbr (15-OH-PGDH)	Carbonyl reductase
CCK-RP (cholecystokinin releasing peptide)	Pancreatic secretory trypsin inhibitor
CD	Caldesmon
CD-10	Endopeptidase-24.11
CD 13	Aminopeptidase N
CD 23	
CD 26	
CD 35	Complement receptor type 1
CD42a	Platelet membrane glycoprotein IX (GPIX)
CD42b	Platelet membrane glycoprotein Ibα (GPIbα)
CD42c	Platelet membrane glycoprotein Ibβ (GPIbβ)
CD42d	Platelet membrane glycoprotein V (GPV)
CD46	Membrane cofactor protein (CD46)
CD 55	Decay accelerating factor
CD 59	
CDF (cytotoxic T cell differentiation factor)	Interleukin 6
CDM	Caldesmon
CDP (calcium dependent protease)	Calpain
CEH	Carboxyl ester lipase
CEL	Carboxyl ester lipase
Cerebrospinal fluid IGFBP-22	Insulin-like growth factor binding protein-5
Cerebrospinal fluid IGFBP-32	Insulin-like growth factor binding protein-6
Ceredase	Glucocerebrosidase
CG	Choriogonadotropin
CgA	Chromogranin A
CgB	Chromogranin B
CgC	Secretogranin II
cGPx	Glutathione peroxidase, cellular
cGSHPx	Glutathione peroxidase, cellular
CGRP	Calcitonin gene-related peptide
CHE	Butyrylcholinesterase
ChE	Butyrylcholinesterase
CHIP28 (Channel-forming integral protein of 28kD)	Aquaporin-1
Cholecystokinin releasing peptide (CCK-RP)	Pancreatic secretory trypsin inhibitor
Cholesterol esterase	Carboxyl ester lipase
Cholinergic neuronal differentiation factor (CNDF)	Leukemia inhibitory factor (LIF)
Choriogonadotropin	
Chorionic gonadotropin	Choriogonadotropin
Christmas factor	Factor IX
Chromogranin A	
Chromogranin A:CgA	Chromogranin A
Chromogranin B	
Chromogranin C	Secretogranin II
Chymase	
Chymotrypsin-like neutral protease	Leukocyte cathepsin G
CK	Creatine kinase
C kinase	Protein kinase C
C-kinase	Protein kinase C

Class 1	Aldehyde dehydrogenase, E1, liver cytoplasm
Class 2	Aldehyde dehydrogenase, E2, liver mitochondria
Classical protein kinase C	Protein kinase C
Clearing factor	Lipoprotein lipase
Clearing-factor lipase	Lipoprotein lipase
CLG1	Neutrophil collagenase
CLI	Clusterin
CLMF	Interleukin 12
Cluster differentiation antigen 10	Endopeptidase-24.11
Clusterin	
CM	Aldehyde oxidase
CM	Xanthine oxidoreductase
CNDF (cholinergic neuronal differentiation factor)	Leukemia inhibitory factor (LIF)
Coagulation factor II	Prothrombin
Coagulation factor III	Tissue factor
Coagulation factor V	Factor V
Coagulation factor VII	Factor VII
Coagulation factor VIII	Factor VIII
Coagulation factor IX	Factor IX
Coagulation factor X	Factor X
Coagulation factor XI	Factor XI
Coagulation factor XII	Factor XII
COL13A1	Collagen type XIII
COL15A1	Collagen type XV
COL18A1	Collagen type XVIII
Cold-insoluble globulin	Fibronectin
Collagen I	Collagen type I
Collagen II	Collagen type II
Collagen V	Collagen type V
Collagen VI	Collagen type VI
Collagen VII	Collagen type VII
Collagen IX	Collagen type IX
Collagen X	Collagen type X
Collagen XI	Collagen Type XI
Collagen XII	Collagen type XII
Collagen XIII	Collagen type XIII
Collagen XIV	Collagen type XIV
Collagen XI	Collagen type XI
Collagen type I	
Collagen type II	
Collagen type V	
Collagen type VI	
Collagen type VII	
Collagen type IX	
Collagen type X	
Collagen type XI	
Collagen type XII	
Collagen type XIII	
Collagen type XIV	
Collagen type XV	
Collagen type XVIII	
Collectin receptor	
Colony-stimulating factor-1	Macrophage colony-stimulating factor (M-CSF)
Colony-stimulating factor-2	Granulocyte-macrophage colony-stimulating factor GM-CSF

Colony-stimulating factor-α	Granulocyte-macrophage colony-stimulating factor GM-CSF
Colony-stimulating factor-β	Granulocyte colony-stimulating factor (G-CSF)
Common acute lymphoblastic leukaemia antigen	Endopeptidase-24.11
COMP	Cartilage oligomeric matrix protein
Complement C3	C3 complement protein
Complement C4	C4 complement protein
Complement C5	C5 complement protein
Complement C6	C6 complement protein
Complement C8	C8 complement protein
Complement factor D	Factor D
Complement factor H	Factor H
Complement factor I	Factor I
Complement lysis inhibitor	Clusterin
Complement protein C1q	C1q complement protein
Complement protein C1r	C1r complement protein
Complement protein C3	C3 complement protein
Complement protein C4	C4 complement protein
Complement protein C5	C5 complement protein
Complement protein C6	C6 complement protein
Complement protein C7	C7 complement protein
Complement protein C8	C8 complement protein
Complement protein C9	C9 complement protein
Complement receptor type 1	
Complement S-protein	Vitronectin
COMT	Catechol-O-methyltransferase
Conglutinin activating factor (KAF)	Factor I
Connective tissue activating peptide-III	Platelet basic protein
Conventional protein kinase C	Protein kinase C
Convertase	Furin
Core-specific lectin	Mannose-binding lectin
Cortisol binding globulin	
COX	Cytochrome C oxidase
CP4	Surfactant protein D
CP	Insulin and proinsulin
CPI-A	Cystatin A
CPI-B	Cystatin B
CPK	Creatine kinase
cPKC	Protein kinase C
$cPLA_2$, cytosol	Phospholipase A_2
CPN	Carboxypeptidase N
CPR	Procarboxypeptidase U
CPU	Procarboxypeptidase U
CR1	Complement receptor type 1
C-reactive protein	
Creatine kinase	
Creatine kinase conversion factor	Carboxypeptidase N
Creatine phosphokinase	Creatine kinase
Creatine phosphoryltransferase	Creatine kinase
Cromer blood group antigen	Decay accelerating factor
CRP	C-reactive protein
CRP	Cysteine-rich protein
CSF-1	Macrophage colony-stimulating factor (M-CSF)
CSF-2	Granulocyte-macrophage colony-stimulating factor GM-CSF
CSF-2α	Interleukin 3

CSF-2β	Interleukin 3
CSF-α	Granulocyte-macrophage colony-stimulating factor GM-CSF
CSF-β Ppo	Granulocyte colony-stimulating factor (G-CSF)
CSF-Eo	Interleukin 5
CSIF	Interleukin 10
CSRP	Cysteine-rich protein
CT	Calcitonin and procalcitonin
CTAP-III (connective tissue activating peptide-III)	Platelet basic protein
CxRP (rabbit)	C-reactive protein
Cylcophilin A	Cyclophilin-18
Cyclophilin B	Cyclophilin-20
Cyclophilin-1	Cyclophilin-18
Cyclophilin-18	
Cyclophilin-20	
Cyclophilin-40	
CyP-A	Cyclophilin-18
CyPB	Cyclophilin-20
CPH-B	Cyclophilin-20
CyP-1	Cyclophilin-18
CyP-18	Cyclophilin-18
CyP-20	Cyclophilin-20
CyP-40	Cyclophilin-40
Cystatin A	
Cystatin B	
Cystatin C	
Cystatin D	
Cystatin S,SN,SA	
Cysteine-rich protein	
Cytochrome aa_3	Cytochrome C oxidase
Cytochrome C	
Cytochrome C oxidase	
Cytochrome c:O_2 oxidoreductase	Cytochrome C oxidase
Cytokine synthesis inhibitory factor	Interleukin-10
Cytosolic-transglutaminase	Tissue transglutaminase
Cytotactin	Tenascin
Cytotoxic lymphocyte maturation factor	Interleukin 12
Cytotoxic T cell differentiation factor (CDF)	Interleukin 6
DAF	Decay accelerating factor
DAP IV	CD 26
DβH	Dopamine beta-hydroxylase
DBH	Dopamine beta-hydroxylase
DBP	Vitamin D binding protein
DEAE	Aldehyde oxidase
DEAE	Xanthine oxidoreductase
Decay accelerating factor	
delta a T	Cystatin C
Dermatan sulfate cofactor	Heparin cofactor II
D-factor	Leukemia inhibitory factor (LIF)
D-fructose-1,6-biphosphate D-glyceraldehyde-3-phosphate-lyase	Fructose-1,6-biphosphate aldolase
D-glyceraldehyde-3-phosphate dehydrogenase	Glyceraldehyde-3-phosphate dehydrogenase
DIA (differentiation inhibitory activity)	Leukemia inhibitory factor (LIF)
Diastase	Alpha-amylase
DIF (differentiation inducing factor)	Leukemia inhibitory factor (LIF)
Differentiation inducing factor (DIF)	Leukemia inhibitory factor (LIF)
Differentiation inhibitory activity (DIA)	Leukemia inhibitory factor (LIF)

Differentiation retarding factor (DRF)	Leukemia inhibitory factor (LIF)
Dipeptidyl aminopeptidase IV	CD 26
Dipeptidyl carboxypeptidase	Angiotensin converting enzyme
Dipeptidyl-peptidase IV	CD 26
Dopamine β-hydroxylase	Dopamine beta-hydroxylase
Dopamine beta-hydroxylase	
Dopamine beta-monoxygenase	Dopamine beta-hydroxylase
DP IV	CD 26
DPP IV	CD 26
DRF (differentiation retarding factor)	Leukemia inhibitory factor (LIF)
E1 isozyme	Aldehyde dehydrogenase, E1, liver cytoplasm
E2 isozyme	Aldehyde dehydrogenase, E2, liver mitochondria
E3 isozyme	Aldehyde dehydrogenase, E3, liver cytoplasm
E-24.11	Endopeptidase-24.11
EAP	Erythrocyte acid phosphatase
ecNOS	Nitric oxide synthase
EDF (eosinophil differentiation factor)	Interleukin 5
eGPx	Glutathione peroxidase, extracellular
eGSHPx	Glutathione peroxidase, extracellular
Endogenous pyrogen (EP)	Interleukin 1
Endopeptidase-24.11	
Endoplasmic reticulum glucose transport protein	Glucose transport protein 7
Endothelial cell type plasminogen activator inhibitor	Plasminogen activator inhibitor type-1
Endothelial-transglutaminase	Tissue transglutaminase
Enkephalinase	Endopeptidase-24.11
eNOS	Nitric oxide synthase
Eosinophil colony stimulating factor	Interleukin 5
Eosinophil differentiation factor	Interleukin 5
EP	Erythropoietin
EP (endogenous pyrogen)	Interleukin 1
Epidermal SH-Proteinase inhibitor	Cystatin A
EPI (extrinsic pathway inhibitor)	Tissue factor pathway inhibitor
EPO	Erythropoietin
EPO-R	Erythropoietin-receptor
EpoR	Erythropoietin-receptor
Erythrocyte acid phosphatase	
Erythrocyte ankyrin	Ankyrin
Erythrocyte-transglutaminase	Tissue transglutaminase
Erythrocyte-type facilitative glucose transport protein	Glucose transport protein 1
Erythropoietin	
Erythropoietin receptor	
Extrinsic pathway inhibitor	Tissue factor pathway inhibitor
F II	Prothrombin
F IIa (activated)	Prothromin
F V (procofactor)	Factor V
F Va (cofactor)	Factor V
F VII	Factor VII
F VIII:C	Factor VIII
F IX	Factor IX
F IXα	Factor IX
F IXaα	Factor IX
F IXa (activated)	Factor IX
F IXaβ	Factor IX

F X	Factor X
F XI	Factor XI
F XIa (activated)	Factor XI
F XII	Factor XII
F XIIa (activated)	Factor XII
F XIII	Plasma factor XIII
F XIIIa (activated)	Plasma factor XIII
Factor I (Roman One)	Fibrinogen
Factor I-cofactor	Factor H
Factor II (Roman Two)	Prothrombin
Factor V	
Factor VII	
Factor VIII	
Factor VIII-related protein	Von Willebrand factor
Factor IX	
Factor X	
Factor XI	
Factor XII	
Factor XIII	Plasma factor XIII
Factor B	
Factor D	
Factor H	
Factor I (letter)	
FAD	Aldehyde oxidase
FAD	Xanthine oxidoreductase
Fast acting inhibitor	Plasminogen activator inhibitor type-1
Fast (F) chain	J chain
Fat/muscle plasma membrane glucose transport protein	Glucose transport protein 4
FBP aldolase	Fructose-1,6-biphosphate aldolase
FcεRII	CD 23
Fcε receptor II	CD 23
FD	Factor D
Fe_2Lf (iron-saturated)	Lactoferrin
Fe_2Tf (iron-saturated)	Transferrin
Fe_CTf	Transferrin
Fe_NTf	Transferrin
Ferritin	
Fetomodulin	Thrombomodulin
Fetuin	Alpha2-HS glycoprotein
Fg	Fibrinogen
Fgb	Fibrinogen
Fibrinase	Plasma factor XIII
Fibrinogen	
Fibrinoligase	Plasma factor XIII
Fibrin polymerase	Plasma factor XIII
Fibrin stabilizing factor	Plasma factor XIII
Fibroblast surface glycoprotein	Aminopeptidase N
Fibronectin	
Fletcher factor	Plasma prekallikrein
FN	Fibronectin
Fructose-1,6-biphosphate aldolase	
Fructose-1,6-biphosphate-triosephosphate-lyase	Fructose-1,6-biphosphate aldolase
FSG	Aminopeptidase N
Furin	
(5-glutamyl)-peptide:amino acid 5-glutamyltransferase	Gamma-glutamyltransferase
G3PD	Glyceraldehyde-3-phosphate dehydrogenase
G3PDH	Glyceraldehyde-3-phosphate dehydrogenase
G-6-Pase	Glucose-6-phosphatase

G6PD	Glucose-6-phosphate dehydrogenase
G6PDH	Glucose-6-phosphate dehydrogenase
G6S	N-Acetylgalactosamine-6-sulphatase
GaI-6-S	N-Acetylgalactosamine-6-sulphatase
GaINAc-6-S	N-Acetylgalactosamine-6-sulphatase
GALNS	N-Acetylgalactosamine-6-sulphatase
Galactose-6-sulphate sulphatase	N-Acetylgalactosamine-6-sulphatase
Galactose-6-sulphatase	N-Acetylgalactosamine-6-sulphatase
gal-T	Beta 1,4-galactosyltransferase
Gamma-aminobutyraldehyde dehydrogenase	Aldehyde dehydrogenase, E3, liver cytoplasm
Gamma c globulin	Cystatin C
gamma-CSF	Cystatin C
Gamma-globulin	Immunoglobulin G
Gammaglutamylcarboxylase	Vitamin K-dependent carboxylase
Gamma-glutamyltransferase	
Gamma-glutamyltransferase	Gamma-glutamyltransferase
Gamma-glutamyltranspeptidase	Gamma-glutamyltransferase
Gamma-guanidinobutyraldehyde dehydrogenase	Aldehyde dehydrogenase, E3, liver cytoplasm
Gamma-M	Immunglobulin M
Gamma-macroglobulin	Immunglobulin M
Gamma-trace	Cystatin C
gamma-GT	*Gamma-glutamyltransferase*
GAPD	Glyceraldehyde-3-phosphate dehydrogenase
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GAS	Glucuronate-2-sulphatase
GBG	Factor B
GC	Vitamin D binding protein
GC'ase	Glucocerebrosidase
GCDFP-24 (gross cystic disease fluid protein of 24 kDa)	Apolipoprotein D
GCDFP-44 (gross cystic disease fluid protein of 44 kDa)	Zn alpha-2-glycoprotein
GC-globulin	Vitamin D binding protein
GCP	Interleukin 8
G-CSF	Granulocyte colony-stimulating factor (G-CSF)
Gelatin (after denaturation)	Collagen type I
GGT	Gamma-glutamyltransferase
GH	Somatotropin
GHBP	Somatotropin binding protein
GHR	Somatotropin receptor
GLO (gene)	Glyoxalase I
GLO 1-1	Glyoxalase I
GLO 1-2 (allozyme)	Glyoxalase I
GLO 2-2 (allozyme)	Glyoxalase I
GloI (protein)	Glyoxalase I
GloI-A	Glyoxalase I
GloI-E	Glyoxalase I
GloII (protein)	Glyoxalase I
Globin	Glyoxalase II
Glucocerebrosidase	Hemoglobin
Glucogenase	Alpha-amylase
Glucose-6-phosphatase	
Glucose-6-phosphate dehydrogenase	
Glucose-6-phosphatase system transport protein	Glucose transport protein 7
Glucose transport protein 1	
Glucose transport protein 2	
Glucose transport protein 3	

Glucose transport protein 4	Glucocerebrosidase
Glucose transport protein 5	Aldehyde reductase
Glucose transport protein 7	
Glucosylceramidase	Glucuronate-2-sulphatase
Glucuronate reductase	Glucose transport protein 1
Glucuronate-2-sulphatase	Glucose transport protein 2
Glucuronate-2-sulphate sulphatase	Glucose transport protein 3
GLUT 1	Glucose transport protein 4
GLUT 2	Glucose transport protein 5
GLUT 3	Glucose transport protein 7
GLUT 4	Aspartate aminotransferase
GLUT 5	Aspartate aminotransferase
GLUT 7	
Glutamate-α-ketoglutarate transaminase	
Glutamate-oxalacetate transaminase	
Glutamic-γ-semialdehyde dehydrogenase, liver mitochondria	
Glutamic-aspartic transaminase	Aspartate aminotransferase
Glutamic-oxaloacetic transaminase	Aspartate aminotransferase
Glutamine:D-glutamyl-peptide 5-glutamyltransferase	Gamma-glutamyltransferase
Glutaredoxin	Thioltransferase
Glutathione peroxidase, cellular	
Glutathione peroxidase, cytosolic	Glutathione peroxidase, cellular
Glutathione peroxidase, classical	Glutathione peroxidase, cellular
Glutathione peroxidase-1	Glutathione peroxidase, cellular
Glutathione peroxidase, extracellular	
Glutathione peroxidase plasma	Glutathione peroxidase, extracellular
Glyal-mesenchymal extracellular matrix protein (GMEM)	Tenascin
Glyceraldehyde-3-phosphate dehydrogenase	
Glycyl-prolyl-β-naphthylamidase	CD 26
Glycine-rich beta-globulin	Factor B
Glycogen phosphorylase	
Glycocalicin (extracellular fragment)	Platelet membrane glycoprotein Ibα (GPIbα)
Glycoprotein 4-beta-galactosyltransferase	Beta 1,4-galactosyltransferase
Glyoxalase I	
Glyoxalase II	
GlxI	Glyoxalase I
GlxII	Glyoxalase II
GM-CSF	Granulocyte-macrophage colony-stimulating factor GM-CSF
GMEM (Glyal-mesenchymal extracellular matrix protein)	Tenascin
Goblin	Ankyrin
GOT	Aspartate aminotransferase
gp-80	Clusterin
gp150	Aminopeptidase N
gp III	Clusterin
GP	Glycogen phosphorylase
GPDH	Glyceraldehyde-3-phosphate dehydrogenase
GPIX	Platelet membrane glycoprotein IX (GPIX)
GPV	Platelet membrane glycoprotein V (GPV)
GPIbα	Platelet membrane glycoprotein Ibα (GPIbα)
GPIbβ	Platelet membrane glycoprotein Ibβ (GPIbβ)
G protein-coupled receptor kinase 1	Rhodopsin kinase
Granulocyte chemotactic protein	Interleukin 8
Granulocyte collagenase	Neutrophil collagenase
Granulocyte colony-stimulating factor (G-CSF)	
Granulocyte-macrophage colony-stimulating factor (GM-CSF)	

GRK1	Rhodopsin kinase
Group-specific component of serum	Vitamin D binding protein
Growth hormone	Somatotropin
Growth hormone dependent IGF-binding protein	Insulin-like growth factor binding protein-3
Growth hormone binding protein	Somatotropin binding protein
Growth hormone receptor	Somatotropin receptor
GRx	Thioltransferase
GSHPx-1	Glutathione peroxidase, cellular
GT	Beta 1,4-galactosyltransferase
GT	Gamma-glutamyltransferase
H	Factor H
Haemoglobin	Hemoglobin
Haemosiderin	
Hageman factor	Factor XII
HAGH (gene)	Glyoxalase II
hAR	Aldose reductase
Hb	Hemoglobin
HCII	Heparin cofactor II
HCA I	Carbonic anhydrase I
HCA II	Carbonic anhydrase II
HCA III	Carbonic anhydrase III
HCA IV	Carbonic anhydrase IV
HCA B	Carbonic anhydrase I
hCG	Choriogonadotropin
HCSF (histamine-producing cell-stimulatory factor)	Interleukin 3
HCPTP-A	Erythrocyte acid phosphatase
HCPTP-B	Erythrocyte acid phosphatase
hCyP2	Cyclophilin-20
Heat shock protein 32	Heme oxygenases
Heme oxygenases	
Hemoglobin	
Hemopexin	
Hemopoietin-1	Interleukin 1
Hemopoietin-2	Interleukin 3
Heparin cofactor	Antithrombin
Heparin cofactor II	
Heparin neutralizing factor	Platelet factor 4
Heparin-releasable lipase	Lipoprotein lipase
Hapatic lipase	
Hepatic triacylglycerol lipase	Hepatic lipase
Hepatocyte stimulating factor (HSF)	Interleukin 6
Hepatocyte stimulating factor III (HSF III)	Leukemia inhibitory factor (LIF)
HEPR	Apolipoprotein B messenger RNA editing protein
Hex	β-N-acetyl-D-hexosaminidase
Hexabrachion	Tenascin
HF	Factor XII
Hg	Hemoglobin
Hgb	Hemoglobin
hGHBP	Somatotropin binding protein
hGH	Somatotropin
hGHR	Somatotropin receptor
High affinity platelet factor 4	Platelet factor 4
High Km aldehyde reductase	Aldehyde reductase
HILDA (human interleukin for DA cells)	Leukemia inhibitory factor (LIF)
Histamine-producing cell-stimulatory factor (HCSF)	Interleukin 3
Histidine-rich glycoprotein	

HK	Kininogens
hK3	Prostate specific antigen
HKBP	Kallistatin
H-kininogen	Kininogens
HL	Hepatic lipase
HLE	Leukocyte elastase
HNC	Neutrophil collagenase
HNE	Leukocyte elastase
HNL	Neutrophil lipocalin
HO	Heme oxygenases
HO-1	Heme oxygenases
HO-2	Heme oxygenases
Hormone-sensitive lipase	
HPGF (hybridoma/plasmacytoma growth factor)	Interleukin 6
HPI	Insulin and proinsulin
HPRO (proinsulin)	Insulin and proinsulin
HSL	Hormone-sensitive lipase
Hpx	Hemopexin
HRF20 (20 kDA homologous restriction factor)	CD 59
HRG	Histidine-rich glycoprotein
HRGP	Histidine-rich glycoprotein
HSA	Albumin
HSF (hepatocyte stimulating factor)	Interleukin 6
HSFIII (hepatocyte stimulating factor III)	Leukemia inhibitory factor (LIF)
HSP32 (Heat shock protein 32)	Heme oxygenases
HTGL	Hepatic lipase
HUK	Tissue kallikrein
Human cytoplasmic phosphotyrosyl phosphatase	Erythrocyte acid phosphatase
Human bone-derived IGFBP	Insulin-like growth factor binding protein-5
Human glandular kallikrein	Tissue kallikrein
Human glandular kallikrein 3	Prostate specific antigen
Human "high Km" aldehyde dehydrogenase	Glutamic-γ-semialdehyde dehydrogenase, liver mitochondria
Human interleukin for DA cells (HILDA)	Leukemia inhibitory factor (LIF)
Human red milk protein	Lactoferrin
Human tissue kallikrein	Tissue kallikrein
HuMb	Myoglobin
HUMTFRR	Transferrin receptor
Hx	Hemopexin
HxB	Tenascin
Hybridoma/plasmacytoma growth factor (HPGF)	Interleukin 6
I	Factor I
I2S	Iduronate-2-sulphatase
IαI	Inter-alpha-inhibitor
IaI	Inter-alpha-inhibitor
IATI	Inter-alpha-inhibitor
IBP-1	Insulin-like growth factor binding protein-1
IDS	Iduronate-2-sulphatase
IDUA	α-L-iduronidase
Iduronate sulphatase	Iduronate-2-sulphatase
Iduronate-2-sulphatase	
IFN	Interferon gamma
IFN-αA	Interferon alpha
IFN α-2a	Interferon alpha
IFN alfa-2a	Interferon alpha
IFN-γ	Interferon gamma

IFNγ-R	Interferon gamma receptor
IFN-g	Interferon gamma
IFNg-R	Interferon gamma receptor
IFN-R	Interferon gamma receptor
IgA	Immunoglobulin A
IgD	Immunoglobulin D
IgE	Immunoglobulin E
IGF-I	Insulin-like growth factor I
IGF-II (rat IGF-II: MSA)	Insulin-like growth factor II
IGFBP-1	Insulin-like growth factor binding protein-1
IGFBP-2	Insulin-like growth factor binding protein-2
IGFBP-3	Insulin-like growth factor binding protein-3
IGFBP-4	Insulin-like growth factor binding protein-4
IGFBP-5	Insulin-like growth factor binding protein-5
IGFBP-6	Insulin-like growth factor binding protein-6
IgG	Immunoglobulin G
IgG1	Immunoglobulin G
IgG2	Immunglobulin G
IgG3	Immunglobulin G
IgG4	Immunglobulin G
IgM	Immunoglobulin M
IL-1	Interleukin 1
IL-1Ra	Interleukin-1 receptor antagonist
IL-1ra	Interleukin-1 receptor antagonist
IL-2	Interleukin-2
IL-3	Interleukin 3
IL-4	Interleukin 4
IL-5	Interleukin 5
IL-6	Interleukin 6
IL-8	Interleukin 8
IL-10	Interleukin 10
IL-11	Interleukin-11
IL-12	Interleukin 12
IL12	Interleukin 12
Immune adherance receptor	Complement receptor type 1
Immune interferon	Interferon gamma
Immunoglobulin A	
Immunoglobulin D	
Immunoglobulin E	
Immunoglobulin G	
Immunoglobulin M	
Inducible carboxypeptidase activity	Procarboxypeptidase U
Inhibitory IGF-binding protein	Insulin-like growth factor binding protein-4
In-IGFBP	Insulin-like growth factor binding protein-4
iNOS	Nitric oxide synthase
Insulin	Insulin and proinsulin
Insulin and proinsulin	
Insulin-like growth factor binding protein-1	
Insulin-like growth factor binding protein-2	
Insulin-like growth factor binding protein-3	
Insulin-like growth factor binding protein-4	
Insulin-like growth factor binding protein-5	
Insulin-like growth factor binding protein-6	
Insulin-like growth factor I	
Insulin-like growth factor II	
Insulin responsive glucose transport protein	Glucose transport protein 4

Inter-alpha-inhibitor	
Inter-alpha-trypsin-inhibitor	Inter-alpha-inhibitor
Interferon alpha	
Interferon β_2 (IFNβ_2)	Interleukin 6
Interferon gamma	
Interferon γ receptor	Interferon gamma receptor
Interferon gamma receptor	
Interleukin 1	
Interleukin-1 receptor antagonist	
Interleukin 2	
Interleukin 3	
Interleukin 4	
Interleukin 5	
Interleukin 6	
Interleukin 8	
Interleukin 10	
Interleukin 11	
Interleukin 12	
Intima collagen	Collagen type VI
IRAP	Interleukin-1 receptor antagonist
Isoniazid transacetylase	Arylamine N-acetyltransferase
ITI	Inter-alpha-trypsin-inhibitor
IVD (isovaleryl-CoA-dehydrogenase)	Acyl-CoA-dehydrogenase
J1	Tenascin
"J1" antigen	Tenascin
J chain	
Joining chain	J-chain
KAF (Conglutin activating factor)	Factor I
KAF (kinase activating factor)	Calpain
Kalinin	Laminin
Kallikrein-binding protein	Kallistatin
Kallistatin	
Katacalcin (PDN-21)	Calcitonin and procalcitonin
KBP	Kallistatin
Kidney brush border neutral peptidase	Endopeptidase-24.11
Kinase activating factor (KAF)	Calpain
Kininase I	Carboxypeptidase N
Kininase II	Angiotensin converting enzyme
Kininogens	
k-laminin	Laminin
KO-reductase	Vitamin K epoxide reductase
Labile factor	Factor V
LACI (lipoprotein-associated coagulation inhibitor)	Tissue factor pathway inhibitor
Lactase-glycosylceramidase	Lactase-phlorizin hydrolase
Lactase-phlorizin hydrolase	
Lactate dehydrogenase	
Lactoferrin	
Lactogenic hormone	Prolactin
Lactogen(ic) receptor	Prolactin receptor
Lactose synthase A protein	Beta 1,4-galactosyltransferase
Lactotransferrin	Lactoferrin
LAF (Lymphocyte-activating factor)	Interleukin 1
Laki-Lorand factor	Plasma factor XIII
Laminin	
LA PF4 (low affinity platelet factor 4)	Platelet basic protein

Large external transformation-sensitive(LETS) protein	Fibronectin
L-aspartate: 2-oxoglutarate aminotransferase	Aspartate aminotransferase
LC-I	Lipocortin I
LCAT	Lecithin-cholesterol acyltransferase
LCAD (long chain acyl-CoA-dehydrogenase)	Acyl-CoA-dehydrogenase
LD	Lactate dehydrogenase
LDH	Lactate dehydrogenase
Lecithin cholesterol acyltransferase	
LEM (Leukocyte endogenous mediator)	Interleukin 1
Leucyl-beta-naphtylamidase	Aminopeptidase N
Leukemia inhibitory factor (LIF)	
Leukocyte cathepsin G	
Leukocyte elastase	
Leukocyte endogenous mediator (LEM)	Interleukin 1
Leukocyte interferon A	Interferon alpha
Leukocyte interferon alpha-2	Interferon alpha
Leuserpin 2	Heparin cofactor II
Lf	Lactoferrin
L-hexonate dehydrogenase	Aldehyde reductase
L-idurono-sulphate sulphatase	Iduronate-2-sulphatase
LIF	Leukemia inhibitory factor (LIF)
Lipid-soluble vitamin ester hydrolase	Carboxyl ester lipase
Lipocalin 2	Neutrophil lipocalin
Lipocortin I	
Lipoprotein lipase	
Lipoprotein-associated coagulation inhibitor	Tissue factor pathway inhibitor
Liver plasma membrane glucose transport protein	Glucose transport protein 2
Liver transglutaminase	Tissue transglutaminase
Liver-type glucose transport protein	Glucose transport protein 2
LK	Kininogens
L-kininogen	Kininogens
L(+)lactate:NAD^+ oxidoreductase	Lactate dehydrogenase
LM	Laminin
Ln	Laminin
Low affinity Fcε receptor	CD 23
Low affinity IgE receptor	CD 23
Low affinity platelet factor 4 (LA PF4)	Platelet basic protein
Low density lipoprotein receptor related protein	Alpha-2-macroglobulin receptor
Low-KM aldehyde reductase	Aldose reductase
Low molecular weight phosphotyrosyl protein phosphatase	Erythrocyte acid phosphatase
Low M_rPTPase	Erythrocyte acid phosphatase
LPH	Lactase-phlorizin hydrolase
L-phenylalanine 4-monooxygenase	Phenylalanine hydroxylase
LPL	Lipoprotein lipase
LRP	Alpha-2-macroglobulin receptor
Lung fibroblasts-derived IGFBP	Insulin-like growth factor binding protein-6
Luteotropic hormone (LTH)	Prolactin
Lymphocyte-activating factor (LAF)	Interleukin 1
Lymphocyte-derived neutrophil activating peptide	Interleukin 8
LYNAP	Interleukin 8
Lysophospholipase	Carboxyl ester lipase
Lysozyme	
Lysine carboxypeptidase	Carboxypeptidase N
mAAP	Aminopeptidase N
MACIF (Membrane attack complex inhibitory factor)	CD 59
MacNOS	Nitric oxide synthase

Macrophage and granulocyte inducer-M	Macrophage colony-stimulating factor (M-CSF)
Macrophage colony-stimulating factor (M-CSF)	
Macrophage granulocyte inducer-1G	Granulocyte colony-stimulating factor (G-CSF)
Macrophage-granulocyte inducer-1 GM	Granulocyte-macrophage colony-stimulating factor
Macrophage granulocyte inducing protein 2 (MGI-2)	Interleukin 6
Mannan-binding lectin	Mannose-binding lectin
Mannan-binding protein	Mannose-binding lectin
Mannan binding protein receptor	*Collectin receptor*
Mannose-binding lectin	
Matrix metalloproteinase-8	Neutrophil collagenase
Mast cell chymase	Chymase
Mast cell growth factor	Interleukin 3
Mast cell growth factor-2 (MCGF-2)	Interleukin 4
Mast cell tryptase	Tryptase
Matrix metalloproteinase 14 (MMP-14)	Membrane-type 1 matrix metalloproteinase
MBL	Mannose-binding lectin
MBP	Mannose-binding lectin
MCAD (medium chain acyl-CoA-dehydrogenase)	Acyl-CoA dehydrogenase
MCF (mononuclear cell factor)	Interleukin 1
MCGF-2 (mast cell growth factor-2)	Interleukin 4
MCH class I light chain	Beta-2-microglobulin
MCP	Membrane cofactor protein (CD46)
M-CSF	Macrophage colony-stimulating factor (M-CSF)
MDR1 protein	P-glycoprotein
MDNCF	Interleukin 8
Medullasin	Leukocyte elastase
Megakaryocyte colony-stimulating factor (Mk-CSF)	Interleukin 3
Melanoma-derived lipoprotein lipase inhibitor (MLPLI)	Leukemia inhibitory factor (LIF)
MEM-43 antigen	CD 59
Membrane attack complex inhibitory factor (MACIF)	CD 59
Membrane attack complex inhibitory protein (MIP)	CD 59
Membrane carbonic anhydrase	Carbonic anhydrase IV
Membrane cofactor protein (CD46)	
Membrane inhibitor of reactive lysis (MIRL)	CD 59
Membrane-type 1 matrix metalloproteinase	
Merosin	Laminin
Mesosecrin	Plasminogen activator inhibitor type 1
Metserpin	*Alpha-1-proteinase inhibitor*
MGI-1G	Granulocyte colony-stimulating factor (G-CSF)
MGI-1GM	Granulocyte-macrophage colony-stimulating factor GM-CSF
MGI-2 (macrophage granulocyte inducing protein 2)	Interleukin 6
MGI-M	Macrophage colony-stimulating factor (M-CSF)
MHC class I β-chain	Beta-2-microglobulin
Microsomal aminopeptidase	Aminopeptidase N
Microsomal proline aminopeptidase	Aminopeptidase P
Microvillar aminopeptidase	Aminopeptidase N
Milk derived growth factor	Transforming growth factor - beta 3 (1,2)
Minactivin	Plasminogen activator inhibitor type-2
MIP (Membrane attack complex inhibitory protein)	CD 59
MIRL (Membrane inhibitor of reactive lysis)	CD 59
Mk-CSF (megakaryocyte colony stimulating factor)	Interleukin 3

MLPLI (melanoma-derived lipoprotein lipase inhibitor)	Leukemia inhibitory factor (LIF)
MMP-8	Neutrophil collagenase
MMP-14	Membrane-type 1 matrix metalloproteinase
Modulator protein	Calmodulin
MONAP	Interleukin 8
Monoacylglycerol acyltransferase	Hepatic lipase
Monocyte-derived neutrophil activating peptide	Interleukin 8
Monocyte-derived neutrophil chemotactic factor	Interleukin 8
Mononuclear cell factor (MCF)	Interleukin 1
Monotor peptide	Pancreatic secretory trypsin inhibitor
mPAP	Aminopeptidase P
MPO	Myeloperoxidase
MSA (multiplication stimulating activity)	Insulin-like growth factor II
MT1-MMP	Membrane-type 1 matrix metalloproteinase
MT-MMP	Membrane-type 1 matrix metalloproteinase
MT-MMP-1	Membrane-type 1 matrix metalloproteinase
Mucopeptide glycohydrolase	Lysozyme
Mucopeptide N-acetylmuramoylhydrolase	Lysozyme
Multi-colony stimulating factor	Interleukin 3
Multidrug resistance protein	P-glycoprotein
Multidrug transporter	P-glycoprotein
Multifunctional Ca^{2+}/calmodulin-dependent protein kinase	Ca^{2+}/calmodulin-dependent protein kinase II
Multimerin	
Multiplication stimulating activity (MSA)	Insulin-like growth factor binding protein-2
Muramidase	Lysozyme
Myeloperoxidase	
Myoglobin	
Myotendinous antigen	Tenascin
N-Acetylgalactosamine-4-sulphatase	
N-Acetylgalactosamine-4-sulphate sulphatase	N-Acetylgalactosamine-4-sulphatase
N-Acetylgalactosamine-6-sulphatase	
N-Acetylgalactosamine-6-sulphate sulphatase	N-Acetylgalactosamine-6-sulphatase
NAD^+	Aldehyde oxidase
NAD^+	Xanthine oxidoreductase
NAD-linked aldehyde dehydrogenase	Aldehyde dehydrogenase, E1, liver cytoplasm
NAD-linked aldehyde dehydrogenase	Aldehyde dehydrogenase, E2, liver mitochondria
NADP-dependent 15-hydroxyprostaglandin dehydrogenase	Carbonyl reductase
NAF (neutrophil activating factor)	Interleukin 8
NAG	ß-N-acetyl-D-hexosaminidase
NAP-1 (neutrophil activating peptide 1)	Interleukin 8
NAP-2 (neutrophil activating peptide 2)	Platelet basic protein
NAT	Arylamine N-acetyltransferase
Natural killer cell stimulatory factor	Interleukin 12
NCF	Interleukin 8
ncNOS	Nitric oxide synthase
NCPI	Cystatin B
NE	Leukocyte elastase
Nerve growth factor	
Neumega®	Interleukin 11
Neutral cysteine proteinase inibitor	Cystatin B
Neutral endopeptidase	Endopeptidase-24.11
Neutrophil activating factor	Interleukin 8
Neutrophil activating peptide-1 (NAP-1)	Interleukin 8
Neutrophil activating peptide-2 (NAP-2)	Platelet basic protein

Neutrophil chemotactic factor	Interleukin 8
Neutrophil collagenase	
Neutrophil elastase	Leukocyte elastase
Neutrophil gelatinase associated lipocalin	Neutrophil lipocalin
Neutrophil lipocalin	
NGAL	Neutrophil lipocalin
NGF	Nerve growth factor
Nicein	Laminin
Nitric oxide synthase	
NKSF	Interleukin 12
nNOS	Nitric oxide synthase
N,O-acetyltransferase	Arylamine N-acetyltransferase
N,O-AT	Arylamine N-acetyltransferase
NOS	Nitric oxide synthase
Non-conventional protein kinase C	Protein kinase C
Nonspecific cholinesterase	Butyrylcholinesterase
Non-specific esterase	Carboxyl ester lipase
Non-specific lipase	Carboxyl ester lipase
Novel protein kinase C	Protein kinase C
nPKC	Protein kinase C
N-procalcitonin (PAS-57)	Calcitonin and procalcitonin
O-acetyltransferase	Arylamine N-acetyltransferase
OAF (Osteoclast-activating factor)	Interleukin 1
OAT	Arylamine N-acetyltransferase
ON	Osteonectin
Opsin	
Ornithine carbamoyltransferase	
Ornithine transcarbamylase	Ornithine carbamoyltransferase
Osteoclast-activating factor (OAF)	Interleukin 1
Osteonectin	
OTC (OCT)	Ornithine carbamoyltransferase
Oxidase	Cytochrome C oxidase
P	Properdin
P18	CD 59
p30	Prostate specific antigen
p35	Lipocortin I
p146	Aminopeptidase N
p155	Multimerin
PA	Transthyretin
PACE (Paired basic Amino acid Cleaving Enzyme)	Furin
PAH	Phenylalanine hydroxylase
PAI-1	Plasminogen activator inhibitor type-1
PAI-2	Plasminogen activator inhibitor type-2
PAI-3	Protein C inhibitor
p-Amy	Alpha-amylase
Pancreatic amylase	Alpha-amylase
Pancreatic kallikrein	Tissue kallikrein
Pancreatic secretory trypsin inhibitor	
Parathormone	Parathyroid hormone
Parathyrin	Parathyroid hormone
Parathyroid hormone	
Parathyroid hormone-like peptide	Parathyroid hormone-related protein
Parathyroid hormone-related protein	
Particle-bound aminopeptidase	Aminopeptidase N
PAS-57 (N-procalcitonin)	Calcitonin and procalcitonin
Pathway I reductase	Vitamin K epoxide reductase

PBP	Platelet basic protein
P-cell stimulating factor	Interleukin 3
PCI	Protein C inhibitor
P component	Serum amyloid P component
pCPB	Procarboxypeptidase U
PDGF	Platelet-derived growth factor
PDN-21 (Katacalcin)	Calcitonin and procalcitonin
PE	Prolyl oligopeptidase
PEP	Prolyl oligopeptidase
PEPP	Aminopeptidase P
Pepsinogen	
Peptidase P	Aminopeptidase P
Peptidyl-dipeptidase A	Angiotensin converting enzyme
Peptidyldipeptide hydrolase	Angiotensin converting enzyme
Peptidyl- prolyl *cis-trans* isomerase B	Cyclophilin-20
PF4	Platelet factor 4
PG	Pepsinogen
Pg	Pepsinogen
Pg	Plasminogen
P-glycoprotein	
PG-M (chick homologue)	Versican
P-gp	P-glycoprotein
pGPx	Glutathione peroxidase, extracellular
pGSHPx	Glutathione peroxidase, extracellular
PH	Phenylalanine hydroxylase
Phenylalanine hydroxylase	
p-HMW-collagen	Collagen type IX
Phorbol ester receptor	Protein kinase C
Phosphatide 2-acylhydrolase	Phospholipase A_2
Phosphatidylcholine cholinephosphohydrolase	Phospholipase C
Phosphatidylcholine phosphodiesterase	Phospholipase C
Phosphatidylcholine:sterol O-acyltransferase	Lecithin cholesterol acyltransferase
Phosphatidylcholine-sterol acyltransferase	Lecithin cholesterol acyltransferase
Phosphatidylinositol biphosphate phosphodiesterase	Phospholipase C
Phosphatidylinositol inositolphosphohydrolase	Phospholipase C
Phosphatidylinositol phosphodiesterase	Phospholipase C
Phosphodiesterase activator	Calmodulin
Phospholipase A-1	Hepatic lipase
Phospholipase A-2	
Phospholipase C	
Phospholipid-cholesterol acyltransferase	Lecithin cholesterol acyltransferase
Phosphorylase	Glycogen phosphorylase
PI	Alpha-2-plasmin inhibitor
pIgR	Secretory component
PK	Plasma prekallikrein
PKC	Protein kinase C
PLA-1	Hepatic lipase
PLA_2	Phospholipase A_2
Placental plasminogen activator inhibitor	Plasminogen activator inhibitor type 2
Placental protein 12	Insulin-like growth factor binding protein-1
Plasma albumin	Albumin
Plasma cholinesterase	Butyrylcholinesterase
Plasma carboxypeptidase B	Carboxypeptidase N
Plasma factor XIII	
Plasma membrane facilitative glucose transport protein 1	Glucose transport protein 1
Plasma membrane facilitative glucose transport protein 3	Glucose transport protein 3
Plasma membrane facilitative glucose transport protein 5	Glucose transport protein 5
Plasma plasmin inhibitor	Alpha-2-plasmin inhibitor

Plasma prekallikrein	
Plasma procarboxypeptidase B	Procarboxypeptidase U
Plasma thromboplastin antecedent (PTA)	Factor XI
Plasma thromboplastin component (PTC)	Factor IX
Plasma transglutaminase	Plasma factor XIII
Plasmin inhibitor	Alpha-2-plasmin inhibitor
Plasminogen	
Plasminogen activator inhibitor type-1	
Plasminogen activator inhibitor type-2	
Plasminogen activator inhibitor type-3	Protein C inhibitor
Platelet antiheparin activity	Platelet factor 4
Platelet basic protein	
Platelet-derived growth factor	
Platelet factor 4	
Platelet glycoprotein G	Thrombospondin-1
Platelet membrane glycoprotein Ibα (GPIbα)	
Platelet membrane glycoprotein Ibβ (GPIbβ)	
Platelet membrane glycoprotein V (GPV)	
Platelet membrane glycoprotein IX (GPIX)	
PLAUR (gene)	Urokinase plasminogen activator receptor
PLC	Phospholipase C
Plg (gene name Plg)	Plasminogen
Pluripoietin	Granulocyte colony-stimulating factor (G-CSF)
PMNL-collagenase	Neutrophil collagenase
PO	Prolyl oligopeptidase
Poly-Ig receptor	Secretory component
Polymeric immunoglobulin receptor	Secretory component
POPase	Prolyl oligopeptidase
post-γ-globulin	Cystatin C
post-γ-protein	Cystatin C
Post-proline cleaving enzyme	Prolyl oligopeptidase
Post-heparin lipase	Hepatic lipase
Post-proline dipeptidyl aminopeptidase	CD 26
PP12	Insulin-like growth factor binding protein-1
PPCE	Prolyl oligopeptidase
P-PI	Alpha-2-plasmin inhibitor
PPIB	Cyclophilin-20
Ppo	Granulocyte colony-stimulating factor (G-CSF)
Prealbumin	Transthyretin
Precursor of serum prothrombin conversion accelerator	Factor VII
Pregnancy-associated endometrial α_1-globulin	Insulin-like growth factor binding protein-1
Prekallikrein (plasma)	
PRL	Prolactin
PRLR	Prolactin receptor
Proaccelerin	Factor V
Procalcitonin	Calcitonin and procalcitonin
Procarboxypeptidase R	Procarboxypeptidase U
Procarboxypeptidase U	
Processing endoprotease	Furin
Proconvertin	Factor VII
ProCPU	Procarboxypeptidase U
Profibrinolysin	Plasminogen
Proinsulin	Insulin and proinsulin
Prolactin	
Prolactin receptor	
Proline-rich protein	C4b-binding protein

Proline specific endopeptidase	Prolyl oligopeptidase
Prolyl endopeptidase	Prolyl oligopeptidase
Prolyl oligopeptidase	
pro-pCPB	Procarboxypeptidase U
Properdin	
Proserozyme (archaic)	Prothrombin
Prostaglandin 9-keto reductase	Carbonyl reductase
Prostasin	
Prostate specific antigen	
Protectin	CD 59
Protein I	Synapsin I, Synapsin II
Protein III	Synapsin I, Synapsin II
Protein 2.1	Ankyrin
Protein C	
Protein C inhibitor	
Protein HC (human complex-forming)	Alpha-1-microglobulin
Protein kinase C	
Protein S	
Proteoglycan-LT	Collagen type IX
Prothrombin	
Prothrombokinase	Factor X
Protirelin receptor	TRH receptor
Prower factor	Factor X
PSA	Prostate specific antigen
PSAP	Surfactant protein A
PSE	Prolyl oligopeptidase
Pseudocholinesterase	Butyrylcholinesterase
PSTI	Pancreatic secretory trypsin inhibitor
PTA (plasma thromboplastin antecedent)	Factor XI
PTC (plasma thromboplastin component)	Factor IX
PTH	Parathyroid hormone
PTHrP	Parathyroid hormone-related protein
Ptyalin	Alpha-amylase
Pulmonary surfactant apoprotein	Surfactant protein A
R	Opsin
RBP	Retinol-binding protein
RCAP	Erythrocyte acid phosphatase
Receptor for immune or type II interferon	Interferon gamma receptor
Receptor transforming factor (RTF)	Calpain
Red cell acid phosphatase	Erythrocyte acid phosphatase
Red milk protein	Lactoferrin
Renal kallikrein	Tissue kallikrein
Retinol-binding protein	
Rho	Opsin
Rhodopsin kinase	
RK	Rhodopsin kinase
rHTF (recombinant)	Tissue factor
rHuEPO (recombinant)	Erythropoietin
RTF (receptor transforming factor)	Calpain
4-S	N-Acetylgalactosamine-4-sulphatase
4-sulphatase	N-Acetylgalactosamine-4-sulphatase
4-sulphate sulphatase	N-Acetylgalactosamine-4-sulphatase
4-sulpho-N-acetylgalactosamine sulphatase	N-Acetylgalactosamine-4-sulphatase
S-2-hydroxyacylglutathione hydrolase	Glyoxalase II
SA	Albumin
SAA	Serum amyloid A

Salivary cysteine proteinase inhibitor	Cystatin S, SN, SA
Salivary amylase	Alpha-amylase
Salt-resistant lipase	Hepatic lipase
s-Amy	Alpha-amlyase
S-antigen	Arrestin
SAP	Serum amyloid P component
SC	Secretory Component
SCAD (short chain acyl-CoA-dehydrogenase)	Acyl-CoA dehydrogenase
S-cyclophilin	Cyclophilin-20
SCYLP	Cyclophilin-20
S-D-lactoylglutathione methylglyoxal lyase (isomerising)	Glyoxalase I
Secreted cyclophilin -like protein	Cyclophilin-20
Secreted protein, acid and rich in cysteine (SPARC)	Osteonectin
Secretogranin II	
Secretogranin I	Chromogranin B
Secretory component	
Secretory piece	Secretory component
Secretory protein I	Chromogranin A
Semenogelin I	
Semenogelin II	
Serum albumin	Albumin
Serum amyloid A	
Serum amyloid P component	
Serum carboxypeptidase B	Carboxypeptidase N
Serum cholinesterase	Butyrylcholinesterase
Serum-spreading factor	Vitronectin
Sg I	Semenogelin I
SgI	Chromogranin B
SgII	Secretogranin II
Sg II	Semenogelin II
SGP-2	Clusterin
Short chain collagen	Collagen type VI
SI	Sucrase-isomaltase
S-I	Sucrase-isomaltase
Siderophilin	Transferrin
Sixth component of the complement system	C6 complement protein
Skeletal growth factor	Insulin-like growth factor II
Skin chymotrypsin-like serine protease	Chymase
s-laminin	Laminin
SMA	Insulin-like growth factor I
SMC	Insulin-like growth factor I
Small intestine facilitative glucose transport protein	Glucose transport protein 5
Small intestine sugar transport protein	Glucose transport protein 5
SOD	Superoxide dismutase
Somatomedin A	Insulin-like growth factor I
Somatomedin C	Insulin-like growth factor I
Somatotrophin	Somatotropin
Somatotropic hormone	Somatotropin
Somatotropin	
Somatotropin receptor & somatotropin binding protein	
SP-40,40	Clusterin
SP-A	Surfactant protein A
SP-A receptor	Collectin receptor
SPARC (secreted protein, acid and rich in cysteine)	Osteonectin
SP-D	Surfactant protein D
Spectrin	
Sperm mobility inhibitor precursor	Semenogelin I
SP-I	Chromogranin A

$sPLA_2$, secretory	Phospholipase A_2
SPMIP	Semenogelin I
Stefin A	Cystatin A
Stefin B	Cystatin B
STH	Somatotropin
Stuart factor	Factor X
Sucrase	Sucrase-isomaltase
Sucrase-isomaltase	
Sucrase-isomaltase complex	Sucrase-isomaltase
Sulfated glycoprotein-2	Clusterin
Sulphoiduronate sulphatase	Iduronate-2-sulphatase
Superoxide dismutase	
Superoxide: superoxide oxidoreductase	Superoxide dismutase
Surface fibroblast (SF) antigen	Fibronectin
Surfactant protein A	
Surfactant protein D	
Synapsin I	
Synapsin II	
Syndein	Ankyrin
T1	Trypsin(ogen) 2
T3	Trypsin(ogen) 1
TAFI	Procarboxypeptidase U
Tamm-Horsfall glycoprotein	Tamm-Horsfall protein
Tamm-Horsfall protein	
TATI (tumor-associated trypsin inhibitor)	Pancreatic secretory trypsin inhibitor
Tb	Tubulin
TBG	Thyroxine-binding globulin
T-cell growth factor	Interleukin 2
T cell growth factor-2	Interleukin 4
T cell replacing factor (murine)	Interleukin 5
T cell stimulating factor	Interleukin 12
TCGF	Interleukin 2
TCGF-2	Interleukin 4
T-chain	Secretory component
T component	Secretory component
Tenascin	
Testosterone-repressed prostatic messenger 2	Clusterin
Tetranectin	
TF	Tissue factor
Tf	Transferrin
TfnR	Transferrin receptor
TfR	Transferrin receptor
TFPI	Tissue factor pathway inhibitor
Tg	Thyroglobulin
Tg1	Trypsin(ogen) 1
Tg2	Trypsin(ogen) 2
TG_C	Tissue transglutaminase
TGF-β3 (1,2)	Transforming growth factor - beta 3 (1,2)
TH	Tyrosine hydroxylase
Thioltransferase	
THP	Tamm-Horsfall protein
Thrombin-activable fibrinolysis inhibitor	Procarboxypeptidase U
Thrombin-sensitive protein	Thrombospondin-1
Thrombinogen (archaic)	Prothrombin
Thrombomodulin	
Thrombospondin-1	
Thrombospondin-2	

Thrombospondin-3	
Thrombospondin-4	
Thromboxane synthase	
Thy-1 inducing factor	Interleukin 3
Thyrocalcitonin	Calcitonin and procalcitonin
Thyroglobulin	
Thyroid galactosyltransferase	Beta 1,4-galactosyltransferase
Thyroid hormone-binding globulin	Thyroxine-binding globulin
Thyroid prohormone	Thyroglobulin
Thyroid-stimulating hormone	Thyrotropin
Thyroliberin receptor	TRH receptor
Thyropexin	Thyroxine-binding globulin
Thyrotropin	
Thyrotropin receptor	
Thyrotropin releasing factor (TRF) receptor	TRH receptor
Thyrotropin releasing hormone (TRH) receptor	TRH receptor
Thyroxine-binding globulin	
Ti2	Trypsin(ogen) 2
Ti1	Trypsin(ogen) 1
Tissue factor	
Tissue factor inhibitor	Tissue factor pathway inhibitor
Tissue factor pathway inhibitor	
Tissue kallikrein	
Tissue plasminogen activator	Tissue-type plasminogen activator
Tissue thromboplastin	Tissue factor
Tissue transglutaminase	
Tissue-type plasminogen activator	
TK	Tissue kallikrein
TM	Glucocerebrosidase
Tm	Thrombomodulin
TN	Tenascin
TN	Tetranectin
TnC	Troponin C
TNF	Tumor necrosis factor
TNF-α	Tumor necrosis factor
t-PA	Tissue-type plasminogen activator
t-PA inhibitor	Plasminogen activator inhibitor type 1
TPH	Tryptophan hydroxylase
TPOH	Tryptophan hydroxylase
TPOHase	Tryptophan hydroxylase
TPP II	Tripeptidyl-peptidase II
TR	Transferrin receptor
Transcortin	Cortisol binding globulin
Transferrin	
Transferrin binding protein	Transferrin receptor
Transferrin receptor	
Transforming growth factor - beta 3 (1,2)	
Transglutaminase type II	Tissue transglutaminase
Transport piece	Secretory component
Transthyretin	
TRHR	TRH receptor
TRH receptor	
Triose phosphate-dehydrogenase	Glyceraldehyde-3-phosphate dehydrogenase
Tripeptidyl aminopeptidase	Tripeptidyl-peptidase II
Tripeptidyl-peptidase II	
Triphosphoinositide inositoltriphosphohydrolase	Phospholipase C
Troponin A	Troponin C
Troponin C	

TRPOHase	Tryptophan hyrodxylase
TRY I	Trypsin(ogen) 1
TRY II	Trypsin(ogen) 2
Trypsin(ogen) 1	
Trypsin(ogen) 2	
Tryptase	Tryptase
Tryptase I-III	Tryptase
Tryptophan-5-hydroxylase	Tryptophan hydroxylase
Tryptophan-5-monooxygenase	Tryptophan hydroxylase
Tryptophan hydroxylase	
TS	Thrombospondin-1
TSF	Interleukin 12
TSH	Thyrotropin
TSHR	Thyrotropin receptor
TSH receptor	Thyrotropin receptor
TSP	Thrombospondin-1
TSP-1	Thrombospondin-1
TSP-2	Thrombospondin-2
TSP-3	Thrombospondin-3
TSP-4	Thrombospondin-4
TTase	Thioltransferase
tTG	Tissue transglutaminase
TTR	Transthyretin
Tub	Tubulin
Tubulin	
Tumor-associated trypsin inhibitor (TATI)	Pancreatic secretory trypsin inhibitor
Tumor necrosis factor	
TXS	Thromboxane synthase
Type I	Collagen type I
Type II	Collagen type II
Type II Ca^{2+}/calmodulin-dependent protein kinase	Ca^{2+}/calmodulin-dependent protein kinase II
Type II IFN	Interferon gamma
Type II IFN-R	Interferon gamma receptor
Type II interferon	Interferon gamma
Type I NOS	Nitric oxide synthase
Type II NOS	Nitric oxide synthase
Type III NOS	Nitric oxide synthase
Type V	Collagen type V
Type VI	Collagen type VI
Type VII	Collagen type VII
Type IX	Collagen type IX
Type X	Collagen type X
Type XII	Collagen type XII
Type XII A	Collagen type XII
Type XII B	Collagen type XII
Type XIII	Collagen type XIII
Type XIV	Collagen type XIV
Type M-collagen	Collagen type IX
TyrOHase	Tyrosine hydroxylase
Tyrosine 3-hydroxylase	Tyrosine hydroxylase
Tyrosine 3-monooxygenase	Tyrosine hydroxylase
Tyrosine hydroxylase	
UDPgalactose-glycoprotein: galactosyltransferase	Beta 1,4-galactosyltransferase
UDP-glucuronosyltransferases	
UDP-glucuronyltransferase	UDP-glucuronosyltransferases
UDPGTs	UDP-glucuronosyltransferases

UGTs	UDP-glucuronosyltransferases
Undulin	Collagen type XIV
uPAR	Urokinase plasminogen activator receptor
Urinary kallikrein	Tissue kallikrein
Urokinase plasminogen activator receptor	
Uromodulin	Tamm-Horsfall protein
Uromucoid	Tamm-Horsfall protein
Vascular plasminogen activator	Tissue-type plasminogen activator
VC	Versican
Verdoperoxidase	Myeloperoxidase
Versican	
Visual purple	Opsin
Vitamin D binding protein	
Vitamin K-dependent carboxylase	
Vitamin K-epoxidase	Vitamin K-dependent carboxylase
Vitamin K epoxide reductase	
Vitronectin	
VN	Vitronectin
Vn	Vitronectin
Von Willebrand factor	
vWF	Von Willebrand factor
WEHI-3 growth factor	Interleukin 3
Xanthine dehydrogenase	Xanthine oxidoreductase
Xanthine oxidase	Xanthine oxidoreductase
Xanthine oxidoreductase	
XPNPEP	Aminopeptidase P
X-Pro aminopeptidase	Aminopeptidase P
X-Pro-dipeptidyl aminopeptidase	CD 26
Zn-α_2 gp	Zn alpha-2-glycoprotein
Zn alpha-2-glycoprotein	
Zwischenferment	Glucose-6-phosphate dehydrogenase
Zymohexase	Fructose-1,6-biphosphate aldolase

General Abbreviations

aa	Amino acid
bp	Base pair
Da	Dalton
DFP	Diisopropylfluorophosphate
EM	Electron microscopy
GuHCl	Guanidine hydrochloride
hr/hrs	Hour/hours
IEF	Isoelectric focusing
kb	Kilo base
L	Liter
M	Moles/L, molar
mM	Millimoles/L, millimolar
mw	molecular weight
nm	Nanometer
PAGE	Polyacrylamide gel electrophoresis
S	Svedberg unit (10^{-13} s)
SDS	Sodium dodecyl sulfate (e.g. SDS-PAGE)
Sia	Sialic acid
S.D.	Standard deviation
Å	Angström = 10^{-10} meters or 0.1 nm
ε	Molar extinction coefficient
µM	Micromoles/L, micromolar
µm	Mikrometer

Abbreviations for Amino Acids

A	Ala	Alanine
B	Asx	Asparagine/Aspartate
C	Cys	Cysteine
D	Asp	Aspartate
E	Glu	Glutamate
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
Z	Glx	Glutamine/Glutamate

β-N-Acetyl-D-hexosaminidase

Don J. Mahuran and Roderick B. C. Tse

Synonyms	β-N-Acetylglucosaminidase; NAG (EC 3.2.1.30)
Abbreviations	Hex
Classifications	EC 3.2.1.52
Description	Hex is a lysosomal hydrolase. The Hex isozymes are composed of all possible dimeric combinations of an α and/or a β subunit (60% aa identity). There are two major isozymes in normal human tissues, Hex A and B. A minor, unstable isozyme of Hex termed Hex S is found at low levels in samples from patients with Sandhoff disease (β-subunit deficiencies). Hence, Hex S is not considered a physiologically functional form of Hex. The heat stability and pI of these isozymes are in a decreasing order such that Hex B (ββ) > Hex A (αβ) >> Hex S (αα). Hex A (or Hex S) in serum can be specifically denatured by heating at 59°C for 15 min (see Concentration).
Structure	Both subunits undergo extensive modification during their biosynthesis. The mature subunits have complex polypeptide chain structures due to proteolytic processing in the lysosome. The pro-α-polypeptide chain (64 kDa) is converted to a small non-glycosylated α_p (7 kDa) and major glycosylated α_m (56 kDa) chain held together by a disulfide bond in the mature α-subunit. The common pro-β chain (67 kDa) is cleaved into three mature glycosylated polypeptide chains, a small β_p (12 kDa) and two major β_b (24 kDa) and β_a (28 kDa) chains, held together by disulfide bonds. Association of the inactive subunits (likely in the ER), produces active pro-isozymes. Hence, the structures of the precursor->mature forms of isozymes are: Hex A, αβ- > ($\alpha_p\alpha m$) ($\beta_p\beta_b\beta_a$); Hex B, ββ- > 2($\beta_p\beta_b\beta_a$). Brackets indicate disulfide bonded polypeptide chain. For simplicity the precursor structures are normally given.
Molecular Weight	Hex A and Hex B isozymes have a molecular mass of about 120 (± 10) kDa
Sedimentation Coeff.	5.8S (native enzyme)
Isoelectric Point	4.8 (Hex A), 6.9 (Hex B), ≤ 3.5 (Hex S)
Extinction Coeff.	11.11 (280nm, 1cm, 1%)
Enzyme Activity	6.5 nmol of MU/hr/ng of pure enzyme (assayed with 1.6 mM 4-MU-GlcNAc.
Coenzymes/Cofactors	Hydrolysis of G_{M2} ganglioside by Hex A requires the presence of a 22 kDa, heat stable, cofactor known as the G_{M2} activator protein (activator). The hydrophilic moiety of G_{M2} is recognized and bound by a specific region on the activator. The ceramide residue of G_{M2} is then pulled out of the lipid membrane bilayer and folded into a hydrophobic groove of the activator. The resulting activator-lipid complex (1:1) is fully water soluble. The activator then binds to a specific recognition site of Hex A in such a way that the terminal β-linked GalNAc residue of G_{M2} is correctly positioned in the active center of the enzyme (see Substrates) [1,2].

Substrates

Hex is specific for GlcNAc and GalNAc residues in β-anomeric linkage. In addition to the natural substrates (see Biological Function), the enzyme also recognizes several artificial substrates consisting of GlcNAc or GalNAc residues β-linked to molecules that fluoresce or have chromogenic properties after hydrolysis. The artificial substrates are commercially available and are hydrolyzed at a greater rate than the natural substrates, *i.e.* the G_{M2} (with saturated [activator]): 4-MU-GlcNAc-6-SO_4 hydrolysis ratio = 1:1300), but they lack part of the specificity of the natural substrates [3,4].

Natural substrates that can be hydrolyzed by Hex A (and possibly Hex S) but not Hex B, generally contain negatively charged group(s) on or near the terminal non-reducing sugar residue; this suggests the presence of a positively charged binding pocket in the α-subunit that is not found in the β-subunit. All Hex isozymes catabolize neutral artificial substrates. However, *in vivo* only Hex A in combination with the G_{M2} activator protein, can catabolize negatively charged G_{M2}-ganglioside. *In vitro* assays demonstrate that detergents can be substituted for the activator; however under these conditions Hex S, as well as Hex A, but not Hex B can efficiently hydrolyze G_{M2} ganglioside. Interestingly Hex B can hydrolyse G_{A2}, the neutral, asialo derivative of G_{M2} in the presence of detergent, but not in the presence of activator alone. As well the activator, even in the absence of G_{M2}, can inhibit the hydrolysis of 4-MU-GlcNAc-6-SO_4 by both Hex A and Hex S. These data indicate that the binding site for the complex is also located in the α-subunit, but that elements of the β-subunit are necessary to somehow correctly orientate the complex and allow hydrolysis of the ganglioside. The most common artificial substrate used to assay total Hex protein is the fluorescent compound 4-methylumbelliferyl-N-acetyl-β-D-glucosamine (4-MU-GlcNAc). Hex A (and Hex S) can be assayed specifically using a negatively charged 4-MU derivative containing β-linked GlcNAc-6-SO_4 (4-MU-GlcNAc-6-SO_4). Ratios of 4-MU-GlcNAc:4-MU-GlcNAc-6-SO_4 hydrolysis using 1.6mM substrates are: Hex A, 4:1; Hex S, 1:1; and Hex B, 300:1 [1,3].

The colorimetric substrate, p-nitrophenyl-β-D-GlcNAc and β-D-GlcNAc-SO_4 are also widely used but are less sensitive and can be subjected to some interference from colored materials in crude tissue extracts or in body fluids [2,4].

Inhibitors

The following are competitive inhibitors, with the first being an inhibitory ion and should be avoided in buffers, whereas the latter two are specific inhibitors of the enzmye.

Acetate ions; N-actyl-hexosaminolactone; (±) -6-Acetamido-1,2-anhydro-6-deoxy-myo-inositol (N-acetylconduramine B trans-epoxide, also a "pseudo-substrate")[5].

Biological Functions

Lysosomal Hex cleaves glycosidic linkages of non-reducing, terminal β-D-N-acetylglucosamine (GlcNAc) or β-D-N-acetylgalactosamine (GalNAc) residues on glycolipids, glycoproteins, steroids, and glycosaminoglycans. *In vivo*, these enzymes are active on a variety of glycoconjugates, such as glycolipids (globoside) and certain mucopolysaccharides. However, only Hex A, in combination with a specific activator protein, can hydrolyze the terminal N-acetylgalactosamine residue from the G_{M2} ganglioside (glycolipid) [1,2,4].

Physiology/Pathology

Most naturally occurring deleterious mutations affecting the α-subunit result in a total loss of Hex A activity and mature α-CRM (Hex B levels are normal or increased), and are associated with the most common form of G_{M2}-gangliosidosis, i.e. infantile (classical) Tay-Sachs disease. In the rare B1 variant form of Tay-Sachs disease normal levels of mature α- and

β-CRM and Hex A and Hex B activities are found when samples are assayed with neutral substrates, e.g., 4-MU-GlcNAc. However, the B1-variant Hex A is inactive toward G_{M2} and α-specific GlcNAc-6-sulfate containing substrates, suggesting an active site mutation (see Genetics/ Abnormalities). Mutations affecting the β-subunit result in the deficiency of both Hex A and B activity (1-5% residual activity is from Hex S) and mature β-CRM, and give rise to another form of G_{M2}-gangliosidosis, Sandhoff disease. The third form of G_{M2}-gangliosidosis is caused by mutations in the *GM2A* gene encoding the activator [2].

The G_{M2}-gangliosidoses show extreme variability in clinical phenotype. Patients are generally classified according to the age at onset of clinical symptoms and the age at death. There are three such categories: infantile, subacute, and chronic. Patients with the common infantile form display the most consistent phenotype. There is evidence of neurological disease by 6 months of age with more profound symptoms occurring later in the first year of life. Thereafter motor and mental deterioration progress rapidly. Death usually occurs by the age of 5 years. Biochemically, these patients lack any residual activity towards G_{M2}-ganglioside. Sandhoff has estimated that 10% of normal Hex A activity can be compatible with normal life. Hence, small variations in residual activity, e.g., 1 - 5% may account for the large clinical variations found within the less severe phenotpyes [2].

Degradation

Intralysosomal

Genetics/Abnormalities

The α and β subunits are encoded by the evolutionarily related *HEXA* and *HEXB* genes mapped on chromosome 15q23-q24 and 5q13, respectively [2]. Mutations in the *HEXA* gene are responsible for Tay-Sachs disease (TSD) which occurs with high incidence (1/3900) in the Ashkenazi Jewish population (classical infantile form). The major mutation (81%) in this population is a 4 bp insertion in exon 11 (early stop codon and unstable mRNA, also found in the general population). A "G to C" transversion in the 5'end of intron 12 (abnormal splicing and unstable mRNA) accounts for another 20% of the Jewish Tay-Sachs alleles. The "G to A" substitution in exon 7 (Gly-269 → Ser, chronic variant, also present in the general population) makes up 3% of the mutant Jewish alleles. In addition, a 7.6 kb deletion in the 5'-end of the *HEXA* gene (no mRNA synthesized), is commonly found in TSD patients of French Canadian heritage. A rare enzymatic variant of TSD disease, the B1 variant, is due to one of two mutations which change α-Arg-178 to either His or Cys ("G to A" or "C to T", respectively). The Arg-178 → His substitution is the most common and is found in patients from different geographic and ethnic origins. Recently α-Arg-178 and its β homolog, β-Arg-211, have been shown to be part of the catalytic sites [6]. Over 60 mutations in the *HEXA* gene have been identified [2].

Mutations in the *HEXB* gene can result in Sandhoff disease with a predicted incidence of about 1 in 300,000 in the general population. A surprisingly high frequency (8 out 30 Sandhoff alleles examined) 16 kb deletion mutation (removing the promotor and exon 1-5), was found in patients from at least three ethnic backgrounds, suggesting that it may have been subjected to genetic drift. Almost 20 *HEXB* mutations have been identified [2, 7].

Half-life

Long lived

Concentration

This can be given by Hex A activity[a]/total activity[b] times 100%

Genotype	Total activity	% HexA	Hex A Activity
"Noncarriers"	700 ± 200	63 ± 6.0	470 ± 80
Tay-Sachs disease	900 ± 100	4.0 ± 2.0	40 ± 20
Sandhoff disease	40 ± 20	100 (Hex S)	
TSD-Heterozygotes	600 ± 100	40 ± 5	280 ± 70
SD-Heterozygotes	600 ± 100	80 ± 4.0	500 ± 90

a. After heating, obtained by subtracting the residual from the total activity.

b. Activity is expressed as nmoles of 4-methylumbelliferone produced/ml of serum/hour ± SD (unpublished data).

Isolation Method

Most convenient source human placenta:

Concanavalin-A-Sepharose affinity column chromatography separates glycoproteins (including Hex) from the total soluble proteins.

2-acetamido-N-(ε-aminocaproyl)-2-deoxy-β-D-glucopyranosylamine coupled to Sephacryl S-200 is used as a specific Hex affinity column.

The anionic exchanger DEAE coupled to various supports, e.g. Sepharose 6B-Cl, is used to finally separate the Hex isozymes [8,9].

Amino Acid Sequence

α-subunit-[α_p]{α_m}

MTSSRLWFSL LLAAAFAGRA TA[LWPWPQNF QTSDQRYLYP
NNFQFQYDVS SAAQPGCSVL DEAFQRYRDL LFG]SGSWPRP
YLTGKRH{TLE KNVLVVSVVT PGCNQLPTLE SVENYTLTIN
DDQCLLLSET VWGALRGLET FSQLVWKSAE GTFFINKTEI
EDFPRFPHRG LLLDTSRHYL PLSSILDTLD VMAYNKLNVF
HWHLVDDPSF PYESFTFPEL MRKGSYNPVT HIYTADQVKE
VIEYARLRGI RVLAEFDTPG HTLSWGPGIP GLLTPCYSGS
EPSGTFGPVN PSLNNTYEFM STFFLEVSSV FPDFYLHLGG
DEVDFTCWKS NPEIQDFMRK KGFGEDFKQL ESFYIQTLLD
IVSSYGKGYV VWQEVFDNKV KIQPDTIIQV WREDIPVNYM
KELELVTKAG FRALLSAPWY LNRISYGPDW KDFYVVEPLA
FEGTPEQKAL VIGGEACMWG EYVDNTNLVP RLWPRAGAVA
ERLWSNKLTS DLTFAYERLS HFRCELLRRG VQAQPLNVGF
CEQEFEQT}*

β-subunit [β_p]{β_b}(β_a)

MELCGLGLPR PPMLLALLLA TLLAAMLALL TQVALVVQVA
EAARAPSVSA [KPGPALWPLP LSVKMTPNLL HLAPENFYIS
HSPNSTAGPS CTLLEEAFRR YHGYIFG]FY KWHHEPAEFQ
AK{TQVQQLLV SITLQSECDA FPNISSDESY TLLVKEPVAV
LKANRVWGAL RGLETFSQLV YQDSYGTFTI NESTIIDSPR
FSHRGILIDT SRHYLPVKII LKTLDMAFNK FNVLHWHIVD
DQSFPYQSIT FPELSNKGSY SLSHVYTPND VRMVIEYARL
RGIRVLPEFD TPGHTLSWGK GQKDLLTPCY S}RQN{KLDS
FGPINPTLNT TYSFLTTFFK EISEVFPDQF IHLGGDEVEF
KCWESNPKIQ DFMRQKGFGT DFKKLESFYI QKVLDIIATI
NKGSIVWQEV FDDKAKLAPG TIVEVWKDSA YPEELSRVTA
SGFPVILSAP WYLDLISYGQ DWRKYYKVEP LDFGGTQKQK
QLFIGGEACL WGEYVDATNL TPRLWPRASA VGERLWSSKD
VRDMDDAYDR LTRHRCRMVE RGIAAQPLYA GYCNHENM)*

*The N-terminal signal peptide sequence is underlined. Sequences of each mature polypeptide are in brackets. Each polypeptide is linked by disulfide bonds in the mature subunit. The aa residues between brackets are cleaved off during maturation processes. Double underlined residues represent used glycosylation sites [2]. The most conserved domain within 15 Hex related sequences from various species including bacteria is between βIle-194/αIle-161 and βSer-259/αSer-226 (dotted underline). Two of the invariant residues in this domain have been shown to be involved in substrate catalysis, βArg-211/αArg-178 and βAsp-196/αAsp-163 (underlined)[10].

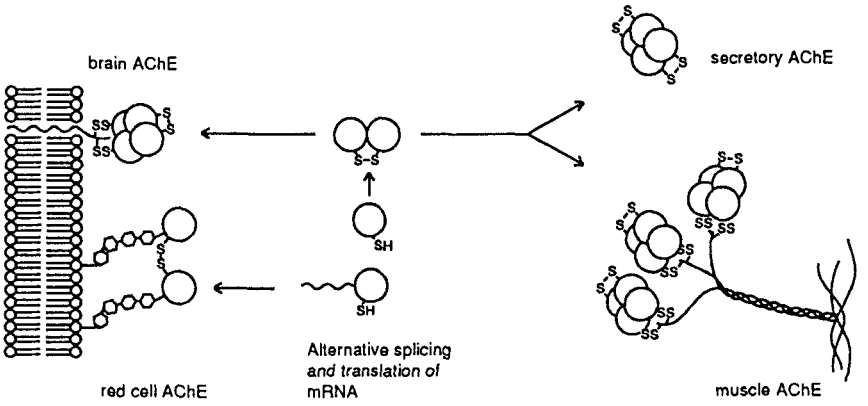
Disulfides/SH-Groups	Hex A -(native)-at least two disulfide bonds. Hex B -(native)-at least two disulfide bonds.
General References	1.) Furst, W. and Sandhoff, K. *Biochim. Biophys. Acta* 1992, **1126**:1-16. 2.) Gravel, R.A. et al. The G_{M2} gangliosidoses. In: *The Metabolic Basis of Inherited Disease*. Scriver, C.V., Beaudet, A.L., Sly, W.S., Valle, D. (eds). McGraw-Hill, New York 1995, pp.2839-2879. 3.) Hou, Y. , Tse, R., Mahuran, D.J. *Biochemistry* 1996, **35**:3963-3969. 4.) Mahuran, D., Novak, A., Lowden, J.A. Isozymes. *Curr. Top. Biol. Med. Res.* 1985, 12:229-288. 5.) Legler, G. and Bollhagen, R. *Carbohydr. Res.* 1992, **233**:113-123. 6.) Brown, C.A. and Mahuran, D.J. *J. Biol. Chem.* 1991, **266**:15855-15862. 7.) Mahuran, D.J. *Biochim. Biophys. Acta* 1991, **1096**:87-94. 8.) Mahuran, D.J. and Lowden, J.A. *Can. J. Biochem.* 1980, **58**:287-294. 9.) Geiger, B. and Arnon, R. *Biochemistry* 1976, **15**:3483-3493. 10.)Tse, R., Wu, Y.J., Vavougios, G., Hou, Y., Hinek, A., Mahuran, D.J. *Biochemistry* 1996, **35**:10894-10903.
Ref. for DNA/AA Sequences	The GSDB (http://www.ncgr.org:80(gsdb/) accession numbers of the nucleotide and deduced sequences of the two human subunits are as follows: α-chain = M13520 β-chain = M13519.

Acetylcholinesterase

Urs Brodbeck

Synonyms	None
Abbreviations	AChE
Classifications	EC 3.1.1.7
Description	A membrane bound enzyme functionally occuring in multiple molecular forms which differ in the number of catalytically active subunits, in the primary structure, in glycosylation as well as in the mode of anchoring to cell membranes.
Structure	AChE from red cell membranes (G2-form) consists of two disulfide linked catalytic subunits to which a glycosyl-phosphatidylinositol membrane anchor is attached via amide linkage to the C-terminal amino acid. In brain, AChE is composed of 4 catalytic subunits (G4-form) connected by disulfide bonds to a scarcely defined hydrophobic peptide of approximately 20 kDa which serves as membrane anchor. In motor end plates, AChE is a dodecameric, asymmetric structure in which the 12 catalytic subunits are linked by disulfide bonds to 3 collagen-like molecules (A 12-form). Minor amounts of AChE exist as intracellular or secretory soluble monomeric G1 and G4-forms.
Molecular Weight	G2-AChE from red cells in presence of Triton X-100: 151 kDa (sedimentation equilibrium); subunit molecular weight: 71–73 kDa, after deglycosylation 60 kDa (SDS-PAGE). G4-AChE from human brain 66 kDa before and 59 kDa after deglycosylation (SDS-PAGE)
Sedimentation Coeff.	G2-AChE: 6.5 S; G4-AChE: 10.0 S
Isoelectric Point	G2-AChE: pH 4.5–5.2 (5 isoforms).
Extinction Coeff.	Unknown
Enzyme Activity	Hydrolyzes acylcholine-esters with a high preference fo acetylcholine
Coenzymes/Cofactors	None
Substrates	Acetylcholine, acetylthiocholine > propionylcholine > butyrylcholine
Inhibitors	See Physiology/Pathology
Biological Functions	Hydrolysis of the neurotransmitter acetylcholine in cholinergic synapses and in motor end plates; occurs in many non-cholinergic cells in which function is unknown.
Physiology/Pathology	Specific inhibition of AChE by organophosphorous compounds, carbamates, and quaternary nitrogen containing compounds, leads to endogenous accumulation of acetylcholine; results in hypersalivation, miosis, watery nasal discharge, bronchoconstriction, increased bronchial secretion, nausea, abdominal cramps, involuntary defecation and urination, bradycardia, hypotension, muscle twitching and fasciculation, eventually severe muscle weakness and paralysis. Death may be caused by respiratory and/or circulatory failure.

Degradation	Unknown
Genetics/Abnormalities	All forms of AChE are thought to be derived from one gene by multiple splicing. Within the coding region, there are two splicing sites which generate three coding exons. In all cases known so far, exons 1 and 2 do not vary while exon 3 leads to the multiplicity of molecular forms differing in the total number of amino acid residues. Brain G4-AChE contains 583 amino acids (data for mouse brain AChE) while the glycosyl-phosphatidylinositol membrane anchored G2-form in erythrocytes is 38 amino acids shorter. All potential N-glycosylation sites reside within coding region 1 and 2.
Half-life	Unknown
Concentration	Minor membrane constituent. In red cell membranes, estimated number of AChE molecules is 700–800 per single cell.
Isolation Method	By affinity chromatography: G2-form of AChE from human red cell membranes, G4-form from human caudate nucleus.
Amino Acid Sequence	AChE belongs to the class of serine esterases. Highly homologous to cholinesterase (EC 3.1.1.8) and structurally similar to other esterases as well as to thyroglobulin (in the C-terminal domain). Nucleotide sequence and primary structure of human AChE elaborated by H. Soreq and collaborators.
Disulfides/SH-Groups	7 Cys residues, 6 involved in intrachain disulfide bonding (conserved in all forms of AChE as well as in thyroglobulin), one in interchain disulfide bonding (near C-terminus, in variable region of G2 and G4-form).
General References	Brodbeck, U. "Amphiphilic acetylcholinesterase: properties and interactions with lipids and detergents". In: *Progress in Protein-Lipid Interactions 2*, Watts, A. and De Pont, J. J. H. H. M. (eds.) Elsevier, Amsterdam, 1986; pp. 303–338. Massoulié, J. and Toutant, J. P. "Vertebrate Cholinesterases: Structures and Types of Interaction". In: *Handbook of Experimental Pharmacology 86*, Whittaker, V. P. (ed.) Springer, Berlin, 1988; pp. 167–224. Taylor, P. "The Cholinesterases". *J. Biol. Chem.* 1991, **266:** 4025–4028.
Ref. for DNA/AA Sequences	Soreq, H. and Zakut, H. "Cholinesterase Genes: Multileveled Regulation". In: *Monographs in Human Genetics*, Sparkes, R. S. (ed.) 1990, Vol. **13**, Karger Basel. Soreq, H. et al. Molecular cloning and construction of the coding region for human acetylcholinesterase reveals a G + C-rich attenuating structure. *Proc. Natl. Acad. Sci. USA* 1990, **87:** 9688–9692.



Molecular Model adopted from A. Chatonnet and O. Lockridge (Biochem. J. **260**, 1989; pp. 625–634). Open circles designate catalytic subunits. Intersubunit disulfide bonds are indicated by S-S. The G2-amphiphilic forms of red cell AChE has a glycolipid anchor, the G4 bain form is disulfide bonded to a 20 kD peptide anchor while the A12 form of muscle AChE has three G4 heads linked to a collagen tail via disulfide bonds.

Acyl-CoA dehydrogenase

Kay Tanaka and Yasuyuki Ikeda

Synonyms	Acyl-CoA:(acceptor) oxidoreductase
Abbreviations	AD preceded by abbr for substr size.
Classifications	EC 1.3.99.2(SCAD); 1.3.99.3(MCAD);1.3.99.0(IVD)
Description	A group of 5 similar enzymes in a gene family, each having distinct substrate specificity [short chain (SCAD), medium chain (MCAD), long chain (LCAD), 2-methyl-branched (2-meBC) ADs and isovaleryl-CoA dehydrogenase (IVD). Human SCAD, MCAD and IVD have also been purified, but rat enzymes are particularly well studied, providing most of the following data. Human enzyme share similar properties with their rat counterparts. All ADs are ubiquitously distributed, with high concentrations in heart muscle, liver, kidney and skeletal muscle. All are homotetrameric, located in mitochondrial matrix and contain one mol FAD per subunit. Each AD is encoded in a nuclear gene, synthesized in cytosol as a precursor with a 2.8 - 3.2 kDa leader peptide at the N-terminal, imported into mitochondria and processed. Recently, another enzyme, very long chain acyl-CoA dehydrogenase (VLCAD) was identified. This enzyme differs considerably from other ACDs for its larger subunit size and for its dimeric structure.
Structure	The monomer peptide chain folded into 3 domains: N- and C-terminal domains mainly consist of antiparallele alpha-helices; middle domain is packed with 2 beta-sheets. FAD is burried in a crevice, that is formed between the N-terminal and middle domains. The C-terminal domain forms the core of the tetramer. In addition to these three domains, VLCAD contains a long extra domain (180 aa) at the C-terminus.
Molecular Weight	(Subunit Size) SCAD: 42.2 kDa; MCAD: 43.7 kDa; LCAD: 44.7 kDa; IVD: 43.1 kDa; 2-meBCAD: ≅ 42 kDa. VLCAD: 71 kDa
Sedimentation Coeff.	Unknown
Isoelectric Point	SCAD, MCAD, LCAD 4.8 - 4.9; IVD, 2-meBCAD 5.5 - 5.6
Extinction Coeff.	SCAD 12.8; MCAD: 12.6; LCAD: 22.6; IVD: 12.3; 2-meBCAD: 12.5 (all 280 nm, 1 %, 1 cm, rat enzymes).
Enzyme Activity	Abstracts one hydrogen each from the à and á positions of acyl-CoA, and transfer electrons to electron transfer flavoprotein (ETF). Each AD has a distinct substrtate specificity (see under Substrates). When AD interacts with a substrate, a base (γ-carboxyl of glutamate-376 in the case of MCAD) in the protein abstracts an αR-proton, producing a carbanion. A βR-hydrogen is then abstracted by N5 of flavin as a hydride ion. This sequence can be completed only in the presence of an appropriate electron acceptor. In the absence of electron acceptor, the original β-hydrogen returns to flavin. The α-proton exchanges with protons in the media before returning to the substrate. Thus, substrate remains chemically unchanged, but with a new α-proton.

1. Installment 1992 (revised 1997)

Coenzymes/Cofactors	All ADs contain FAD as a prosthetic group. In the dehydrogenation reaction, each AD requires ETF as an obligatory electron acceptor. In in vitro experiments, ETF can be replaced with phenazine methosulphate.
Substrates	Each AD has a distinct substrate specificity: SCAD: C4-C6; MCAD: C4-C12; LCAD: C8-C20; VLCAD: C14-C24 (all straight chain acyl-CoAs); IVD: isoC5; 2-meBCAD: isoC4, 2-meC4, 2-propyl-C5 (valproyl-CoA).
Inhibitors	(Methylelecyclopropyl)acetyl-CoA (MCPA-CoA), a metabolite of hypoglycin, selectively inhibits some ADs very severely. Hypoglycin is a toxin that is contained in unripe ackee, a Jamaican local fruit. MCPA-CoA is a enzyme-activated inhibitor. MCPA-CoA inhibits MCAD most severely, followed by SCAD and IVD. 2-meBCAD is inhibited only slowly and mildly. LCAD is not inhibited at all, obviously because the molecular size of MCPA-CoA is too small as a substrate analog for LCAD. When MCPA-CoA interacts with AD, its α-proton abstraction leads to opening of the cyclopropyl ring and its covalent addition to the flavin. Likewise, 2-alkyl- and 3-alkyl-acyl-CoA inhibit ADs in a similar manner.
Biological Functions	VLCAD, LCAD, MCAD and SCAD catalyze the first reaction in a sequence of mitochondrial β-oxidation cycles for fatty acids with varying chain length, converting acyl-CoA to the corresponding 2-enoyl-CoA. VLCAD appears to be the rate-limiting enzyme. IVD catalyzes the third reaction in the leucine pathway, converting isovaleryl-CoA to 3-methyl-crotonyl-CoA. 2-meBCAD catalyzes the third reaction in the isoleucine and valine pathways, converting 2-methylbutyryl-CoA and isobutyryl-CoA to tiglyl- and methacrylyl-CoAs, respectively.
Physiology/Pathology	Hypoglycin ingestion in the form of unripe ackee fruit is the cause for Jamaican vomiting sickness, an acute violent and highly lethal metabolic disease characterized by repeated vomiting, severe hypoglycemia and absence of ketone bodies. Patietns with this disease excrete various abnormal metabolites produced by the inhibition of various ADs and glutaryl-CoA dehydrogenase. These include various short and medium chain- monocarboxylic acids and their glycine conjugates, and dicarboxylic acids.
Degradation	Degradation has been not studied.
Genetics/Abnormalities	cDNAs for all six ACDs both from human and rat sources have been sequenced. The assignment in human are: LCAD, chromosome 2, q34-q35; MCAD, chromosome 1, P-31; SCAD, chromosome 12, q22-qter; IVD, chromosome 15, q14-qter. Genetic deficiencies of IVD (isovaleric acidemia), MCAD, LCAD and SCAD have been identified. Isovaleric acidemia is clinically characterized with episodic vomiting and ketoacidosis accompanied by the accumulation of isovaleric acid in body fluids. Five different IVD mutations have been identified. The disorder, which was previously conceived LCAD deficiency, have recently been shown, in fact, to be VLCAD deficiency. MCAD and VLCAD-deficiencies are characterized by hypoketotic hypoglycemic episodes, could be fatal without treatment. The incidence of MCAD-deficiency (1:10,000) is one of the highest among genetic metabolic disorders. MCAD-deficiency occurs mostly in northern European population. 90% of the variant MCAD genes carry A985G transition, resulting Glu substitution of Lys-304. This mutation originated from a single person.
Half-life	SCAD, MCAD, IVD: 4 hrs; LCAD: > 4 hrs. (all rat)

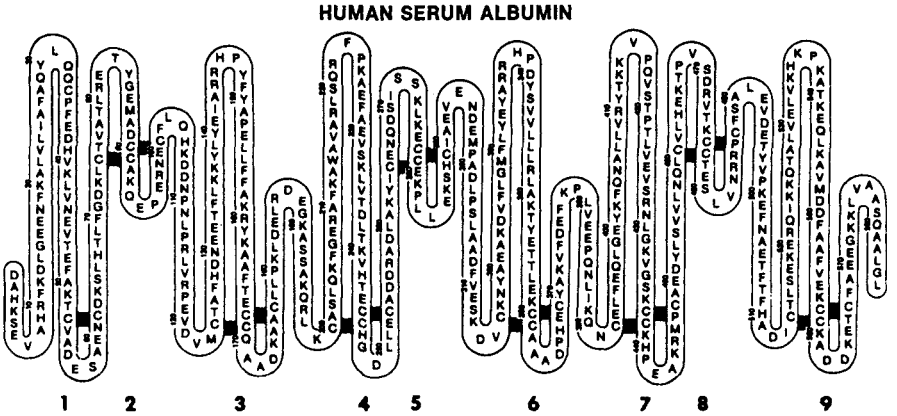
Concentration	Only guess can be obtained from purification data. In order to achieve homogeneity, rat ADs were purified as follows: SCAD, 131 fold; MCAD, 149 fold; LCAD, 78 fold; VLCAD 178 fold; IVD, 957 fold, all in liver mitochondria. 2-meBCAD, considerably lower than IVD.
Isolation Method	Before FPLC application, the preferred system was a sequence of mitochondria isolation/sonication, 40 - 80% ammoniumsulfate precipitation, DEAE-Sephadex (1), hydroxyapatite (2), Matrex Gel Blue A (with NaCl/FAD double gradient) (3), Agarose-hexane-CoA and Bio-Gel A-0.5 columns. (1) partially isolated IVD from 4 others. (2) was effective separating SCAD + 2-meBCAD, MCAD and LCAD from each other. (3) removed remaining IVD from SCAD, and also residual SCAD from 2-meBCAD preparations. (4) is effective isolating LCAD from MCAD and SCAD. The isolation method using FPLC has been applied only to isolating MCAD, SCAD and LCAD that are expressed in bacteria: these data suggest that the use of FPLC would greatly simplify the isolation of various Ads.
Amino Acid Sequence	The sequences of all human and rat ACDs, 6 from either source, have been deduced from the respective cDNA sequences. The size of the mature enzymes and leader peptides are: SCAD, 388 + 26; MCAD, 396 + 25; LCAD, 400 + 30; and IVD 393 + 30 (rat), 29 (human). They share 30-35% sequence identity, thus belonging to a gene family. The sequence similarity in the leader peptide and in the N-terminal region is lower. Otherwise, their sequence similarity extends from the N- to the C-terminals. The γ-carboxyl of Glu-376 is shown to be the α-proton abstracting base in MCAD, based on the use of mechanism based inactivator and also on crystallographic study. However, this Glu residue is not conserved in LCAD and IVD. A number of mutations in human variant ADs (IVD, MCAD and SCAD) has been identified: see above.
Disulfides/SH-Groups	Unknown
General References	Ikeda, Y. and Tanaka, K. *J. Biol. Chem.* 1983, **258**:1077-1085. Ikeda, Y. and Tanaka, K. *J. Biol. Chem.* 1983, **258**:9477-9487. Ikeda, Y. et al. *J. Biol. Chem.* 1985, **260**:1311-1325. Ikeda, Y. et al. *Arch. Biochem. Biophys.* 1987, **252**:662-674. Ikeda, Y. and Tanaka, K. *Biochim. Biophys. Acta* 1990, **1038**:216-221. Finocchiaro, G., et al. *J. Biol. Chem.* 1987, **262**:7982-7989. Kim, J.P. and Wu, J. *Proc. Natl. Acad. Sci. USA* 1988, **85**:6677-6681. Powell, P.J. and Thorpe, C. *Biochemistry* 1988, **21**:6685-6695. Vockley, J. et al. *Am. J. Human Genet.* 1991, **49**:147-157. Yokota, Y. et al. *Am. J. Human Genet.* 1991, **49**:1281-1290. Izai, K. et al. *J. Biol. Chem.* 1992, **267**:1027-1033.
Ref. for DNA/AA Sequences	The enzyme name is followed by the EMBL/GenBank Data Libraries accession number and reference. Rat SCAD (J05030), LCAD (J05029) and IVD (05031): Matsubara, Y. et al. *J. Biol. Chem.* 1989, **264**:16321-1331. Human SCAD: Naito, E. et al. *J. Clin. Invest.* 1989, **83**:1605-1613. Human IVD: Matsubara, Y. et al. *J. Clin. Invest.* 1990, **85**:1058-1064. Human LCAD (M74096): Indo, Y. et al. *Genomics* 1991, **11**:609-620. Rat MCAD (J02791): Matsubara, Y. et al. *J. Biol. Chem.* 1987, **262**: 10104-10108. Human MCAD: Kelly, D. et al. *Proc. Natl. Acad. Sci. USA* 1987, **84**:4068-4072. Human VLCAD: Aoyama, T. et al. *Am. J. Hum. Genet.* 1995, **57**:273-283. Human 2-meBCAD: Rosen, R. et al. *Genomics* 1994, **24**:280-287.

Albumin

Theodore Peters, Jr.

Synonyms	Serum albumin; Plasma albumin
Abbreviations	SA; HSA; Alb
Classifications	Soluble in water, 0.5 sat'd. ammon.sulfate
Description	The major circulating plasma protein, synthesized in the liver and present in all body fluids. A monomeric, single-chain molecule, with no prosthetic groups or carbohydrates attached. Homologous in structure to α-fetoprotein, α-albumin (afamin), and vitamin D-binding protein (Gc-component).
Structure	Single peptide chain divided into 9 SS-bonded loops grouped into 3 homologous domains (see figure). Hydrodynamic studies suggest an ellipsoid with axes 40 X 140 Å units. X-ray crystallography shows as heart-shaped, ≈ 80 X 80 X 30 Å units, 67% helical. Two ligand sites symmetrically placed in subdomains IIA and IIIA.
Molecular Weight	66,438 (calculated from aa composition).
Sedimentation Coeff.	4.4 S
Isoelectric Point	5.6 defatted; 4.6 with 6 fatty acids
Extinction Coeff.	5.31 (280 nm, 1%, 1cm); molar absorbance 35,279.
Enzyme Activity	None, but can slowly catalyze hydrolysis of S-fatty acyl and nitrophenyl esters. Influences the metabolism of the eicosanoids; affects activity of lipases by binding long-chain fatty acids.
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Acts as transport agent for long-chain fatty acids, prostaglandins, hematin, thyroxine, copper(II), cystine, nitric oxide and pyridoxal phosphate. Single thiol has antioxidant function. Binds bilirubin and many drugs (e.g. warfarin) in IIA site, other drugs (e.g. benzodiazepines) and short-chain fatty acids in IIIA site. Binds more weakly to calcium, steroid hormones, and tryptophan.
Physiology/Pathology	Provides 80% of colloid osmotic pressure and 100% of protein contribution to acid/base balance of plasma. Its concentration is a marker for good nutrition and longevity. A negative acute phase protein, production declines in trauma, infection, and illness; hypoalbuminemia presages increased morbidity or mortality. Values less than 20 g/L are usually accompanied by edema. A major pharmaceutical product, administered parenterally in circulatory failure.
Degradation	Normally degraded intracellularly in many tissues to free amino acids. About 10% believed to be lost into the gut.

Genetics/Abnormalities	A single gene situated at proximal end of long arm of chromsome 4, band q11 - 22; order of albumin superfamily genes is: 5'-albumin, α-fetoprotein, α-albumin, vitamin D-binding protein-centromere-3'. Codominant expression. More than 50 variants exhibiting electrophoretic differences identified as "bisalbuminemia". Of these, the site of mutation has been identified in 54. Three of these carry carbohydrate as a result of generation of a glycosylation sequence. Five are variant proalbumins, carrying a basic aminoterminal hexapeptide. "Analbuminemia", reported in 28 cases, is caused by mutations at mRNA intron splice junctions; actually 0.1-1 g/L albumin is present in serum due to "leaky" nature of the mutations.
Half-life	19 days (blood circulation)
Concentration	Serum or plasma: 42 g/L (range 35 - 50 g/L)
Isolation Method	Isolated from serum or plasma by precipitating globulins at 0.5 sat'd. ammon. sulfate, then reducing pH to 4.9; also by affinity chromatography on Cibacron Blue columns and by conventional ion exchange and molecular exclusion chromatography. Commercially produced by low temperature-ethanol "Cohn" procedures developed during World War II. Recombinant production is in progress by several firms.
Amino Acid Sequence	See figure below. Homology with α-fetoprotein 40%, with α-albumin 36%, and with vitamin D-binding protein 19%. High in charged aa residues, giving high total charge and solubility; a single Trp residue at position 214.
Disulfides/SH-Groups	Contains 17 disulfide bonds situated sequentially. A single cysteine SH is at residue 34.
General References	Peters, T., Jr. All About Albumin: *Biochemistry, Genetics and Medical Applications*. San Diego: Academic Press, 1996, pp. 432 Carter, D.C. and Ho, J.X. Structure of serum albumin. *Adv. Prot. Chem.* 1994, **45**:153-204. Kragh-Hansen, U. Structure and ligand binding properties of human serum albumin. *Dan. Med. Bull.* 1990, **37**:57-84. Rothschild, M.A., Oratz, M. and Schreiber, S.S. Serum albumin. *Hepatology* (Review) 1988, **8**:385-401. Rochu D. Human albumin: structure, synthesis and functions (Fre). *Rev. Fr. Transfus. Immuno-Hematol.* 1986, **29**:13-33. Peters, T., Jr. Serum albumin. *Adv. Prot. Chem.* 1985, **37**:161-245.
Ref. for DNA Sequences	Nishio, H., Heiskanen, M., Palotie, A. et al. Tandem arrangement of the human serum albumin multigene family in the sub-centromeric region of 4q: Evolution and chromosomal direction of transcription. *J. Mol. Biol.* 1996, **259**:113-119. Bélanger, L., Roy, S. and Allard, D. New albumin gene 3' adjacent to the α1-fetoprotein locus. *J. Biol. Chem.* 1994, **269**:5481-5484. Lichenstein, H.S., Lyons, D.E., Wurfel, M.M. et al. Afamin is a new member of the albumin, alpha-fetoprotein, and vitamin D-binding protein gene family. *J. Biol. Chem.* 1994, **269**:18149-18154. Minghetti, P.P., Ruffner, D.E., Kuang, W.J. et al. Molecular structure of the human albumin gene is revealed by nucleotide sequence within q11-22 of chromosome 4. *J. Biol. Chem.* 1986, **261**:6747-6757. Lawn, R.M., Adelman, J., Bock, S.C. et al. The sequence of human serum albumin cDNA and its expression in E. coli. *Nucleic Acids Res.* 1981, **9**:6103-6114.



"Model in one-letter amino acid code created by Drs. Richard C. Feldhoff and David J. Ledden, School of Medicine, University of Louisville, Kentucky, USA.; used with kind permission of Dr. Feldhoff. Dark bands represent disulfide bonds."

Alcohol dehydrogenase

Bendicht Wermuth

Synonyms	None
Abbreviations	ADH
Classifications	EC 1.1.1.1
Description	Cytosolic, polymorphic enzymes derived from at least 7 genes with distinct but overlapping patterns of expression. ADH1 (α subunit) ADH2 (β) and ADH3 (γ) constitute Class 1 ADH and are predominantly expressed in the liver. ADH4/Class 2 ADH (π subunit) is found in liver and to a lesser extent in the stomach. ADH5/Class 3 ADH (χ subunit) is expressed in all tissues, but to different extents. ADH6 mRNA occurs in liver and stomach, but the corresponding protein has not convincingly been demonstrated in humans. ADH7/Class 4 ADH (σ or μ subunit) predominates in stomach and oesophagus.
Structure	Dimeric, globular protein of 368-391 aa/subunit. Each subunit is composed of a coenzyme binding and a catalytic domain and contains two Zn atoms essential for catalytic activity and structural integrity. Polymorphisms exist within classes. The 3D structures of ADH 1, 2, 5 and 7 have been determined at 2.5, 2.2, 2.7 and 3.0 Å resolution, respectively.
Molecular Weight	80,000 approx. (aa sequence). Exact values depend on the subunit composition
Sedimentation Coeff.	5.1 S (Class 1)
Isoelectric Point	Calculated IEPs are 7.9-8.3 for Class 1 ADH, and 7.8, 7.3 and 7.9 for Class 2, 3 and 4 ADH, respectively. However, Class I isozymes migrate towards the cathode upon starch gel electrophoresis at pH 8.5, whereas class 2 ADH migrates towards the anode.
Extinction Coeff.	Calculated values are between 0.46 and 0.85 (280 nm, 1%, 1cm) dependent on aa composition. Exact values for individual isozymes have not been determined.
Enzyme Activity	Oxidoreductase acting on the alcohol group of donors with NAD^+ as acceptor.
Coenzymes/Cofactors	NAD^+, acceptor of hydride ion. Zn, tetrahedrally coordinated with one coordination position bonding to the OH-group of alcohols.
Substrates	Primary and secondary alcohols of great diversity, including ethanol, medium chain-length alcohols, ω-hydroxyfatty acids, retinol, steroids and alcohols derived from biogenic amines. Class 3 ADH also acts as formaldehyde dehydrogenase using glutathione as cosubstrate. Marked differences exist between isozymes in substrate specificity and kinetic constants (Km, V).
Inhibitors	Pyrazole and 4-alkyl derivatives act primarily on class 1 isozymes.

1. Installment 1992 (revised 1997)

Biological Functions	Oxidation of toxic alcohols, including ethanol. Physiological alcohols probably include steroids (Class 1 ADH), biogenic alcohols (Class 2 ADH) and retinol (Class 4 ADH).
Physiology/Pathology	Largely unknown. Class 4 ADH may play a role in embryogenesis by providing retinal for retinoic acid synhesis.
Degradation	Unknown
Genetics/Abnormalities	The loci for all ADH genes are clustered on chromosome 4 between q21 and q25. Phylogenetic studies suggest that ADH5 (Class 3 ADH) is the parent enzyme from which the other classes evolved by gene duplication.
Half-life	Unknown
Concentration	≈ 1% of soluble proteins of hepatocytes; not usually detectable in plasma and other body fluids.
Isolation Method	Class 1 ADH: Affinity chromatography on immobilized pyrazole. Class 2, 3 and 4 ADHs do not bind to pyrazole and are purified by ion exchange and nucleotide affinity chromatography. Identification of isozymes is best achieved by starch gel electrophoresis.
Amino Acid Sequence	Sequences of all ADH isozymes have been determined and are deposited in the Swiss Prot database (ADHA_HUMAN, ADHB_HUMAN, ADHG _HUMAN, ADHP_HUMAN, ADHX_HUMAN, ADH6_HUMAN, ADH7 _HUMAN). Sequence homologies range from ca. 60 % positional identiy between the different ADH classes and 93 % between the Class 1 isoymes.
Disulfides/SH-Groups	No disulfides
General References	Jörnvall, H. and Höög, J.O. Nomenclature of alcohol dehydrogenases. *Alcohol.* 1995, **30**:153-161.
Ref. for DNA/AA Sequences	Sequences are deposited in the EMBL/Genbank databases.

Aldehyde dehydrogenase, E1, liver cytoplasm

Regina Pietruszko

Synonyms	NAD-linked aldehyde dehydrogenase
Abbreviations	E1 isozyme, ALDH 1, ALDH II, Class 1.
Classifications	EC 1.2.1.3
Description	A cytoplasmic enzyme, found in the liver, also in other organs, including red blood corpuscles. Homotetramer of 500 aa residue polypeptide chains. Polymorphisms with diminished enzyme activity and one with changed electrophoretic mobility are known. Absence of this enzyme has never been reported.
Structure	A tetrameric protein of identical subunits, acetylated at the N-terminal end, sensitive to atmospheric oxygen. Primary structure is given below; NMR or X-ray crystallography not available.
Molecular Weight	218,840 Da for the intact protein; 54,710 Da for the subunits (summation of component aa residues determined via the aa sequence analysis).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.3
Extinction Coeff.	9.6 (280 nm, 1%, 1 cm)
Enzyme Activity	Catalyses irreversible, NAD-linked dehydrogenation of aldehydes; also catalyses hydrolysis of esters.
Coenzymes/Cofactors	NAD, serves as an acceptor of the hydride ion during dehydrogenation of the aldehyde. p-Nitrophenyl acetate hydrolysis is stimulated by NAD but can also proceed without NAD.
Substrates	Dehydrogenase reaction: a large variety of aldehydes including straight chain and branched aliphatic aldehydes, aromatic aldehydes (eg. benzaldehyde, phenylacetaldehyde, 3,4-dihydroxyphenylacetaldehyde, cinnamaldehyde), hydroxyaldehydes (eg. glycolaldehyde, lactaldehyde, glyceraldehyde), steroid aldehydes, halogen-substituted aldehydes (eg. monochloroacetaldehyde). Formaldehyde and γ-aminobutyraldehyde are poor substrates. Esterase reaction is assayed employing p-nitrophenyl acetate as substrate.
Inhibitors	Biological inhibitors have not been identified. The following are known reversible (R) and irreversible (I) inhibitors: acetophenone (R), bromoacetophenone (I), chloral (R), p-chloromercuribenzoate (I), diethylaminobenzaldehyde (R), disulfiram (Ir), Ellmans reagent (I), iodoacetamide (I), isosorbide dinitrate (Ir), metal ions (R), methylglyoxal (R), propiolaldehyde (I). Inactivation by some irreversible inhibitors can be reversed by reduction with 2-mercaptoethanol (Ir).
Biological Functions	Dehydrogenation of retinal to retinoic acid has been proposed (*Biochem. Pharmacol.* 1992, **43:** 2453–2469). Metabolism of aldehydes ingested in food stuffs, metabolism of aldehydes arising from serotonin, dopamine, norepinephrine via monoamine oxidase, probably in metabolism of corticosteroid aldehydes. It has been also reported to be an androgen-binding protein (*Biochem. Biophys. Res. Commun.* 1991, **175:** 831–838) and to be a positional marker in the retina (*Development* 1991, **112:** 693–702).

Physiology/Pathology	Decreased concentration causes aversive reaction to ingested alcohol.
Degradation	Unknown
Genetics/Abnormalities	The gene is on chromosome 9; absence of this enzyme has never been reported; only cases of decreased enzyme activity and of an electrophoretic variant. Individuals with low enzyme activity show aversion to alcohol.
Half-life	Unknown
Concentration	$\simeq$ 1 g/L (range 0.2–2 g/L) in the average Caucassian liver.
Isolation Method	Chromatography of centrifuged and dialysed human liver homgenate on CM-Sephadex, DEAE-Sephadex and 5'AMP-Sepharose 4B; the homogeneous enzyme is obtained from 5'AMP-Sepharose 4B step. All operations under nitrogen.
Amino Acid Sequence	Acetyl-SSS GTPDLPVLLT DLKIQYTKIF INNEWHDSVS GKKFPVFNPA TEEELCQVEE GDKEDVDKAV KAARQAFQIG SPWRTMDASE RGRLLYKLAD LIERDRLLLA TMESMNGGKL YSNAYLNDLA GCIKTLRYCA GWADKIQGRT IPIDGNFFTY TRHEPIGVCG QIIPWNFPLV MLIWKIGPAL SCGNTVVVKP AEQTPLTALH VASLIKEAGF PPGVVNIVPG YGPTAGAAIS SHMDIDKVAF TGSTEVGKLI KEAAGKSNLK RVTLELGGKS PCIVLADADL DNAVEFAHHG VFYHQGQCCI AASRIFVEES IYDEFVRRSV ERAKKYILGN PLTPGVTQGP QIDKEQYDKI LDLIESGKKE GAKLECGGGP WGNKGYFVQP TVFSNVTDEM RIAKEEIFGP VQQIMKFKSL DDVIKRANNT FYGLSAGVFT KDIDKAITIS SALQAGTVWV NCYGVVSAQC PFGGFKMSGN GRELGEYGFH EYTEVKTVTV KISQKNS. Its positional identity with E2 isozyme is 68% and E3 isozyme is only 40.6%. Regions that show conservation in a large number of other aldehyde dehydrogenases are residues 265–274 containing Glu-268; also Cys-302 (a catalytic residue) can be identified in all sequenced enzymes.
Disulfides/SH-Groups	No disulfides known; 11 SH-groups per subunit of 54,710 Da molecular weight.
General References	Greenfield, N. J. and Pietruszko, R. *Biochim. Biophys. Acta* 1977, **483:** 35–45. Hempel, J. et al. *Eur. J. Biochem.* 1984, **141:** 21–35. Pietruszko, R. *Biochem. Physiol. Subst. Abuse.* **Vol 1**, CRC Press, Inc., 1989, pp. 89–127. Goedde, W. H. et al. *Isozymes: Current Topics in Biol. Med. Res.* 1983, **8:** 175–193. Yoshida et al. *Prog. Nucleic Acid Res. Mol. Biol.* 1991, **40:** 255–287.
Ref. for DNA/AA Sequences	Hsu, L. C. et al. *Proc. Natl. Acad. Sci. USA* 1985, **82:** 3771–3775.

Aldehyde dehydrogenase, E2, liver mitochondria

Regina Pietruszko

Synonyms	NAD-linked aldehyde dehydrogenase
Abbreviations	E2 isozyme, ALDH-2, ALDH I, Class 2.
Classifications	EC 1.2.1.3
Description	A mitochondrial protein; mature form consists of four identical polypeptide chains of ca. 500 aa residues. The subunits in the mature enzyme are ragged at the N-terminus with heterogeneous starting positions. The enzyme is synthesised on ribosomes then is transported into mitochondria. The immature enzyme subunit consists of 517 aa residues and includes a 17 aa signal peptide which directs the enzyme into the mitochondria. This peptide is at the amino terminal end of the peptide chain and is removed by postranslational modification. The structure of the signal peptide is: MLRAAARFGPRLGRRLL.
Structure	Primary structure is known, but tertiary structure has not been determined.
Molecular Weight	216,620 Da consisting of four identical subunits of 54,155 Da (mw determined from aa sequence by summation of molecular weights of component aa).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.1, single protein band.
Extinction Coeff.	10.5 (280 nm, 1%, 1 cm).
Enzyme Activity	Aldehyhyde: NAD Oxidoreductase. Catalyses irreversible, NAD-linked dehydrogenation of aldehydes; also catalyses hydrolysis of esters.
Coenzymes/Cofactors	NAD functions as an acceptor of hydride during aldehyde dehydrogenation; NAD is not required for ester hydrolysis but its presence affects the velocity of ester hydrolysis.
Substrates	Acetaldehyde (Km less than 1 μM), other aliphatic branched and straight chain aldehydes, aromatic aldehydes, α-hydroxyaldehydes, monohaloaldehydes, 3,4-dihydroxyphenylacetaldehyde, 5-hydroxyindoleacetaldehyde, 3,4-dihydroxyphenylglycolaldehyde, imidazoleacetaldehyde. Formaldehyde and γ-aminobutyraldehyde are poor substrates. p-Nitrophenyl acetate and p-nitrophenyl derivatives of acyl groups from 2–8 carbon atoms are substrates for esterase reaction.
Inhibitors	The following are known reversible (R) and irreversible (I) inhibitors: acetophenone (R), bromoacetophenone (I), chloral (R), p-chloromercuribenzoate (I), diethylaminobenzaldehyde (R), N-ethylmaleimide (I), iodoacetamide (I), isosorbide dinitrate (I, but activity can be regained by treatment with 2-mercaptoethanol), propiolaldehyde (I). Disulfiram is a poor inhibitor of this enzyme in vitro.
Biological Functions	Dehydrogenation of aldehydes formed from monoamines via monoamine oxidase. Metabolizes acetaldehyde formed from ingested ethanol.

Physiology/Pathology	Although its major physiological role appears to be in monoamine metabolism, absence of this enzyme in ca. 50% of Oriental individuals does not produce any known pathological symptoms. Individuals devoid of this enzyme exhibit alcohol aversion manifested as the pronounced facial flush "alcohol flush reaction" after consumption of only small quantities of alcohol. Overproduction of this enzyme appears to make alcohol drinking enjoyable.
Degradation	Unknown
Genetics/Abnormalities	Abnormal form (pI = 5.3) occurs in ca. 50% of the Oriental population. In the abnormal form only one substitution is known: Glu-487 has been replaced by Lys-487. The catalytic activity of this abnormal form is greatly diminished (claimed to be totally absent in some cases) compared to the normal enzyme.
Half-life	Unknown
Concentration	$\simeq$ 1 g/L (range 0–2 g/L) in the liver of average Caucassian.
Isolation Method	Liver is the best source. Centrifuged, dialysed liver homogenate is applied on CM-Sephadex, followed by chromatography on DEAE-Sephadex followed by chromatography on 5'AMP Sepharose 4B. The homogeneous protein is obtained from 5'AMP Sepharose which also separates E2 from E1 and E3 aldehyde dehydrogenases.
Amino Acid Sequence	MLRAAAAWPA WAPPLVSRRH PGRAAPNQQP EVFCNQIFIN NEWHDAVSRK TFPTVNPSTG EVICQVAEGD KEDVDKAREG RPGAFQLGSP WRRMDASHRG RLLNRLADLI ERDRTYLAAL ETLDNGKPYV ISYLVDLDMV LKCLRYYAGW ADKYHGKTIP IDGDFFSYTR HEPVGVCGQI IPWNFPLLMQ AWKLGPALAT GNVVVMKVAE QTPLTALYVA NLIKEAGFPP GVVNIVPGFG PTAGAAIASH EDVDKVAFTG STEIGRVIQV AAGSSNLKRV TLELGGKSPN IIMSDADMDW AVEQAHFALF FNQGQCCCAG SRTFVQEDIY DEFVVRSVAR AKSRVVGNPF DSKTEQGPQV DETQFKKILG YINTGKQEGA KLLCGGGIAA DRGYFIQPTV FGDVQDGMTI AKEEIFGPVM QILKFKTIEE VVGRANNSTY GLAAAVFTKD LDKANYLSQA LQAGTVWVNC YDVFGAQSPF GGYKMSGSGR ELGEYGLQAY TEVKTVTVKV PQKNS. Its structure has 68% of positional identity with that of cytoplasmic E1 isozyme and 41.6% of positional identity with that of the cytoplasmic E3 isozyme. Regions that show conservation in other aldehyde dehydrogenases are residues 265–274 containing Glu-268. Cys-302 (a catalytic residue) can be identified in all aldehyde dehydrogenases whose structure is known.
Disulfides/SH-Groups	No disulfides known; 9 SH-groups per subunit, 36 per molecule.
General References	Greenfield, N. J. and Pietruszko, R. *Biochem. Biophys. Acta* 1977, **483:** 35–45. Hempel, J. et al. *Eur. J. Biochem.* 1985, **153:** 13–28. Pietruszko, R. *Biochem. Physiol. Subst. Abuse* CRC Press Inc. 1989, **Vol 1**, pp.89–127. Agarwal, D. P. and Goedde, H. W. *Alcohol Metabolism, Alcohol Intolerance, and Alcoholism.* Springer-Verlag, Berlin, Heidelberg, New York, 1990. Yoshida et al. *Prog. Nucleic Acid Res. Mol. Biol.* 1991, **40:** 255–287.
Ref. for DNA/AA Sequences	Hsu, L. C. et al. *Proc. Natl. Acad. Sci. USA* 1985, **82:** 3771–3775.

Aldehyde dehydrogenase, E3, liver cytoplasm

Regina Pietruszko and Gloria Kurys

Synonyms	Probably previously referred to as either γ-Aminobutyraldehyde dehydrogenase or γ-Guanidinobutyraldehyde dehydrogenase or Betaine aldehyde dehydrogenase.
Abbreviations	E3 isozyme
Classifications	EC 1.2.1.3
Description	A cytoplasmic protein consisting of four polypeptide chains of ca. 500 aa residues. Polypeptide chains are blocked at the N-terminal end.
Structure	Primary structure, shown below, is incomplete at the N-terminal end of the molecule. Tertiary structure is unknown.
Molecular Weight	230,000 Da (gradient gel electrophoresis) consisting of subunits of 54,000 Da (SDS-PAGE).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.3 (major), 5.5 (minor).
Extinction Coeff.	10.0 (280nm, 1%, 1cm).
Enzyme Activity	Aldehyde: NAD Oxidoreductase. Catalyses irreversible, NAD-linked dehydrogenation of aldehydes; also catalyses ester hydrolysis.
Coenzymes/Cofactors	NAD functions as an acceptor of hydride during aldehyde dehydrogenation; NAD is not required for ester hydrolysis but its presence alters the velocity of ester hydrolysis.
Substrates	Amino aldehydes derived from diamines and polyamines are dehydrogenated at low micromolar concentrations. Substrates also include straight chain saturated aldehydes, acrolein, imidazoleacetaldehyde, 3,4-dihydroxyphenyl acetaldehyde, 5-hydroxyindole acetaldehyde and betaine aldehyde.
Inhibitors	Chloral and H2 receptor antagonists (reversible), disulfiram poor inhibitor, iodoacetamide (irreversible).
Biological Functions	Probably a major metabolic role in putrescine: γ-aminobutyric acid pathway; oxidation of aldehydes derived from diamines and polyamines via diamine and polyamine oxidases choline matabolism.
Physiology/Pathology	Unknown
Degradation	Unknown
Genetics/Abnormalities	The gene is on chromosome 1 (McPherson, J.D. et al. *Human Genetics,* 1994, **93**:211-212.
Half-life	Unknown
Concentration	Approx. 0.2g/L in liver (no variation observed to date).

Isolation Method

Liver is the best known source of this protein. The only procedure known involves use of 6 chromatographic steps: CM-Sephadex, DEAE-Sephadex, 5'AMP Sepharose 4B, Decyl Agarose, Aminodecyl Agarose and NAD-Agarose with an 8 carbon spacer. All chromatographic steps are carried out under nitrogen or argon because the enzyme is extremely sensitive to atmospheric oxygen.

Amino Acid Sequence

The aa sequence starts at residue 41 of the E1 and E2 aldehyde dehydrogenases:

EPATGRVIAT FTCSGEKEVN LAVQNAKAAF KIWSQKSGME
RCRILLEAAR IIREREDEIA TMECINNGKS IFEARLDIDI
SWQSLEYYAG LAASMAGEHI QLPGGSFGYT RREPLGVCVG
IGAWNYPFQI ASWKSAPALA CGNAMVFKPS PFTPVSALLL
AEIYSEAGVP PGLFNVVQGG AATGQFLCQH PDVAKVSFTG
SVPTGMKIME MSAKGIKPVT LELGGKSPLI IFSDCDMNNA
VKGALMANFL TQGQVCCNGT RVFVQKEILD KFTEEVVKQT
QRIKIGDPLL EDTRMGPLIN RPHLERVLGF VKVAKEQGAK
VLCGGDIYVP EDPKLKDGYY MRPCVLTNCR DDMTCVKEEI
FGPVMSILSF DTEAEVLERA NDTTFGLAAG VFTRDIQRAH
RVVAELQAGT CFINNYNVSP VELPFGGYKK SGFGRENGRV
TIEYYSQLKT VCVEMGDVES AF.

Regions that show conservation in other aldehyde dehydrogenases are residues 265-274 containing Glu-268. Cys-302 (a catalytic residue) can be identified in here as in all sequenced enzymes. Its positional identity with the E1 isozyme is 40.6% and with the E2 isozyme is 41.6%. The greatest positional identity was seen with E. coli betaine aldehyde dehydrogenase (52.7%).

Disulfides/SH-Groups

15 SH groups per subunit or 60 SH groups per molecule in the available sequence (see above). This probably accounts for extreme sensitivity of this enzyme to atmospheric oxygen.

General References

Kurys, G. et al. *J. Biol. Chem.* 1989, **264**:4715-4721.
Ambroziak, W. and Pietruszko, R. *J. Biol. Chem.* 1991, **266**:13011-13018.
Chern, M.K. and Pietruszko, R. *Biochem. Biophys. Res. Commun.* 1995, **213**:561-568.
Ambroziak, W. et al. *Comp. Biochem. Physiol.* 1991, **100B**:321-327.

Ref. for DNA/AA Sequences

Kurys, G. et al. *Eur. J. Biochem.* 1993, **218**:311-320.

Aldehyde dehydrogenase, stomach cytoplasm

Regina Pietruszko

Synonyms	
Abbreviations	ALDH3, ALDH III
Classifications	EC 1.2.1.5
Description	A cytoplasmic protein occurring in the human stomach wall, oesophagus and lung, sometimes also present in the liver. A homodimer consisting of polypeptide chains of 453 aa residues.
Structure	Primary structure has been established via cloning; tertiary structure has not been determined.
Molecular Weight	130,000 Da (gradient gel elctrophoresis) consisting of subunits of 55,000 Da (SDS-PAGE). The subunit molecular weight computed from the primary structure is: 50,335 Da.
Sedimentation Coeff.	Unknwon
Isoelectric Point	6.5 $\simeq$ (several bands)
Extinction Coeff.	Molar extinction coefficient computed from the aa sequence is 46,220.
Enzyme Activity	Aldehyde: NAD(P) Oxidoreductase. Catalyzes dehydrogenation of a variety of aldehydes with hexanaldehyde, heptanaldehyde, furfuraldehyde and benzaldehyde as the best substrates.
Coenzymes/Cofactors	Both NAD and NADP are coenzymes but NAD is better than NADP; in aldehyde dehydrogenation both NAD and NADP function as acceptors of the hydride ion.
Substrates	Long chain aliphatic aldehydes, furfuraldehyde and benzaldehyde; short chain aliphatic aldehydes are poor substrates.
Inhibitors	Unknown
Biological Functions	Unknown
Physiology/Pathology	Unknown
Degradation	Unknwon
Genetics/Abnormalities	The gene is on human chromosome 17.
Half-life	Unknown
Concentration	Unknown
Isolation Method	The enzyme was purified from human stomach by a two step procedure employing CM-Sephadex and affinity chromatography on 5'AMP Sepharose (*Ann. Hum. Genet.* 1985, **49**: 87–100).

Amino Acid Sequence	MSKISEAVKR APAAFSSGRT RPLQFRIQQL EALQRLIQEQ EQELVGALAA DLHKNEWNAY YEEVVYVLEE IEYMIQKLPE WAADEPVEKT PQTQQDELYI HSEPLGVVLV IGTWNYPFNL TIQPMVGAIA AGNSVVLKPS ELSENMASLL ATIIPQYLDK DLYPVINGGV PETTELLKER FDHILYTGST GVGKIIMTAA AKHLTPVTLE LGGKSPCYVD KNCDLDVACR RIAWGKFMNS GQTCVAPDYI LCDPSIQNQI VEKLKKSLKE FYGEDAKKSR DYGRIISARH FQRVMGLIEG QKVAYGGTGD AATRYIAPTI LTDVDPQSPV MQEEIFGPVL PIVCVRSLEE AIQFINQREK PLALYMFSSN DKVIKKMIAE TSSGGVAAND VIVHITLHSL PFGGVGNSGM GSYHGKKSFE TFSHRRSCLV RPLMNDEGLK VRYPPSPAKM TQH
Disulfides/SH-Groups	7 SH groups/subunit, 14 per molecule. No information about disulfides.
General References	Santisteban, L. et al. *Ann. Hum. Genet.* 1985, **49:** 87–100. Duley, J. A. et al. *Alcohol. Cli. Expl. Res.* 1985, **9:** 263–271. Yoshida, A. et al. *Prog. Nucleic Acid Res. Mol. Biol.* 1991, **40:** 225–287.
Ref. for DNA/AA Sequences	Hsu, L. C. et al. *J. Biol. Chem.* 1992, **267:** 3030–3037.

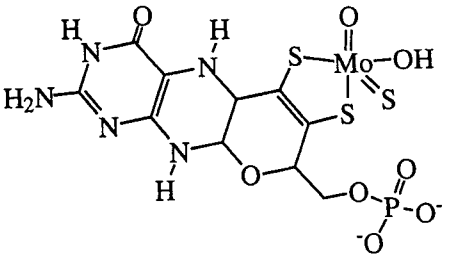
Aldehyde Oxidase

Russ Hille

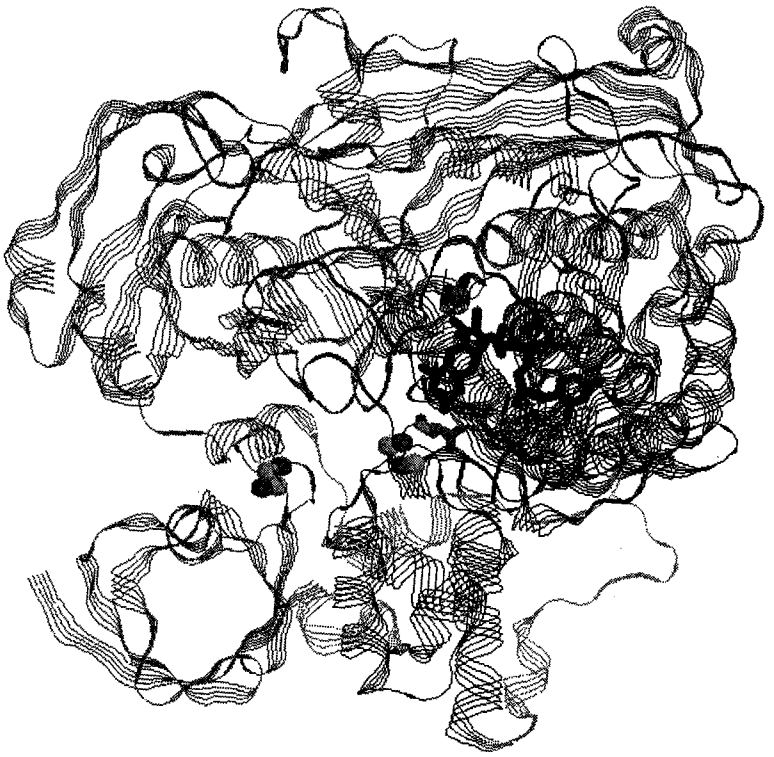
Synonyms	None
Abbreviations	FAD; NAD^+; DEAE; CM
Classifications	EC 1.2.3.2; a molybdenum-containing hydroxylase
Description	Aldehyde oxidase is a complex metalloflavoprotein possessing a molybdenum center, two 2Fe/2S iron-sulfur centers of the spinach ferredoxin variety and FAD, with each redox-active center residing in a discrete domain of the protein structure. The molybdenum center consists of the metal coordinated to an unusual pterin cofactor via a dithiolene side chain of the latter, and is the site of substrate hydroxylation. The enzyme is predominantly localized in the liver. Reducing equivalents obtained by the enzyme in the course of substrate hydroxylation are passed on to O_2; aldehyde oxidase is unable to utilize NAD^+ as oxidizing substrate.
Structure	The structure of human aldehyde oxidase is at present unknown. However, the x-ray crystal structure of a closely related aldehyde oxidoreductase from *Desulfovibrio gigas* has recently been reported, and on the basis of the extensive sequence homologies between the two proteins, it is expected that the human protein will exhibit the same general structure as that from *D. gigas*. This protein possesses a pair of iron-sulfur centers and a molybdenum center, but lacks a flavin. The two iron-sulfur centers are found in separate well-defined protein folding domains at the N-terminus of the protein. The first of these possesses a protein fold very similar to that of spinach ferredoxin, but the second possesses a unique fold, consisting of two long α helices in parallel, with two shorter helices set at an oblique angle to these; the 2Fe/2S cluster lies at one end of the pair of long helices. This second iron-sulfur domain is connected to the molybdenum-binding portion of the protein by an extended meander over the surface of the protein, and on the basis of the sequence homology between the aldehyde and xanthine oxidoreductases, the flavin domain of the human protein is expected to be inserted at some point along this meander. The molybdenum-binding portion of the protein consists of two large domains that lay across one another, with the metal center at their interface; there are extensive contacts between the polypeptide and the pterin cofactor. In addition, the 2-amino group of the pterin, located distal to the dithiolene moiety that binds the metal, is hydrogen-bonded to one of the cysteine residues that coordinates the second of the two iron-sulfur centers. This interaction clearly plays a role in facilitating electron transfer out of the molybdenum center to the other redox-active sites of the enzyme. Access to the active site is provided by a 30Å-long channel from the surface of the protein, with access to the metal center opposite the coordination position of the pterin cofactor.
Molecular Weight	298,800, homodimer
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	35,000 M^{-1} cm^{-1} (450 nm) per subunit.

Coenzymes/Cofactors	Aldehyde oxidase possesses a molybdenum center (consisting of the metal plus pterin cofactor), two 2Fe/2S centers and flavin adenine dinucleotide.
Substrates	Aldehyde oxidase is able to hydroxylate a wide range aldehydes to the corresponding carboxylic acid, as well as a variety of aromatic heterocycles. With regard to reducing substrate, it is distinct from xanthine oxidoreductase in being relatively unreactive toward xanthine, while effectively hydroxylating 1-methylnicotinamide. Aldehyde oxidase is unable to utilize NAD^+ as oxidizing substrate.
Inhibitors	Xanthine dehydrogenase is inhibited by arsenite, cyanide and methanol.
Biological Functions	Aldehyde oxidase in liver catalyzes the oxidation of a variety of aldehydes to carboxylic acids *in vitro*; its physiological substrates may include retinaldehyde.
Physiology/Pathology	Unknown
Degradation	Unknown
Genetics/Abnormalities	Genetic variants in aldehyde oxidase leading to clinical manifestations are not known.
Half-life	Unknown
Concentration	Unknown
Isolation Method	The enzyme is most conveniently isolated from rabbit liver. Published isolation procedures typically involve ammonium sulfate fractionation, followed by column chromatography on hydroxyapatite, DEAE, size exclusion materials or a combination of these.
Amino Acid Sequence	Aldehyde oxidase exhibits extensive sequence homologies to human xanthine oxidoreductase, and is a member of an extended family of molybdenum-containing hydroxylases that is phylogenetically widely dispersed, being found most eukaryotic, bacterial and archaeal sources. Both human and bovine aldehyde oxidases possess several loci where their consensus sequence diverges from that of the xanthine oxidoreductases, however, and these include a distinct aa sequence FLxKCP that is conserved in enzymes from a variety of sources, and is distinct from the corresponding FFxxYR sequence of the aldehyde oxidases. This sequence maps to the flavin domain of the protein.
Disulfides/SH-Groups	The number of disulfide bonds in human aldehyde oxidase is not known.
General References	Hille, R. *Chem. Rev.* 1996, **96**:2757-2816. Branzoli, U. and Massey, V. *J. Biol. Chem.* 1974, **249**:4339-4345. Romão, M.J., Archer, M., Moura, I. et al. *Science* 1995, **270**:1170-1176. Huber, R., Hof, P., Moura, J.J.G. et al. *Proc. Natl. Acad. Sci. USA* 1996, **93**:8846-8851.
Ref. for DNA/AA Sequences	Wright, R.M., Vaitaitis, G.M., Wilson, C.M. et al. *Proc. Natl. Acad. Sci. USA* 1993, **90**:10690.

Molecular models



The active site of aldehyde oxidase



The protein structure of aldehyde oxidoreductase from *Desulfovibrio gigas* (see Romão *et al.*, 1995) (after Hille, 1996, with permission).

Aldehyde Reductase

Oleg A. Barski and Kurt M. Bohren

Synonyms	High Km aldehyde reductase; Glucuronate reductase; L-hexonate dehydrogenase
Abbreviations	ALR1; AR1; AKR1A1
Classifications	EC 1.1.1.2; Aldo-keto reductase AKR1A1
Description	A cytosolic NADP-dependent monomeric oxidoreductase, consisting of 324 aa with no carbohydrate chain. The enzyme is a member of the aldo-keto reductase superfamily, and catalyzes the reduction of a wide variety of aldehydes to their corresponding alcohols. Ubiquitous tissue distribution with highest levels in kidney cortex and liver.
Structure	β/α barrel with an active site located at the C-terminus of the barrel.
Molecular Weight	36,442
Sedimentation Coeff.	2.9S
Isoelectric Point:	5.3
Extinction Coeff.	15.8 (280nm, 1%, 1cm); 57,750 M^{-1} cm^{-1} (280nm)
Enzyme Activity	Catalyzes the conversion of aldehydes to alcohols and vice versa. Follows an ordered bi-bi mechanism with coenzyme binding first and leaving last. Oxidation of alcohols is catalyzed at a rate of less than 2% of the rate measured with p-nitrobenzaldehyde in the reduction reaction at pH 7.0. Dehydrogenase by IUB classification.
Coenzymes/Cofactors	NADPH/$NADP^+$ In the reduction reaction the pro-4R hydride (A-side stereospecific) is transferred from NADPH to the aldehyde substrate. In the oxidation reaction a hydride is transferred from the substrate to the A-side of $NADP^+$.
Substrates	Isocorticosteroids, D-glucuronate, DL-glyceraldehyde, succinic semialdehyde, substituted benzaldehydes, etc.
Inhibitors	Barbiturates, flavonoids, dicarboxylic acids (dimethylsuccinic, tetramethyleneglutaric acid), and Aldose Reductase Inhibitors (ARI): AL1576 is 10 times more specific for aldehyde than for aldose reductase. All inhibitors bind to the enzyme-$NADP^+$ complex, resulting in a competitive pattern in the oxidation reaction and an uncompetitive or non-competitive pattern in the reduction reaction.
Biological Function	Catalyzes the reduction of aldehyde compounds of various structure. May be important in steroid metabolism, glucuronate metabolism, detoxification of methylglyoxal, osones and xenobiotic aldehydes.
Physiology/Pathology	Wide substrate specificity and presence in many tissues suggests a housekeeping role in detoxification of toxic or carcinogenic aldehydes.
Degradation	Unknown

Genetics/Abnormalities	Unknown
Half-life	Unknown
Concentration	Unknown
Isolation Method	Extraction by homogenization of brain, liver, placenta, or kidney (preferably cortex) tissue followed by ion exchange chromatography on DEAE-Sephadex. Subsequent affinity chromatography on Matrex Orange A yields at least 95% homogenous protein. Recombinant protein is available.
Amino Acid Sequence	Important active site residues: D43, K79, and Y49 form a catalytic triad with Y49 being the proton donor. H112 is important for NADPH and substrate orientation. R311 binds distal carboxyl group of substrate. Hydrophobic residues W21, W81, W113, Y115, F117, F124, Y209, W219, I48, I298, P300, L302 line the active site pocket. R268, K22, K262 form salt bridges with NADPH. Homologous to the members of the aldo-keto reductase superfamily with 65% identity to aldose reductase.
Disulfides/SH-Groups	No disulfides. Free SH groups at positions 4, 45, 133, 186, 199, 259.
General References	Barski, O.A. et al. *Biochemistry* 1995, **34**:11264-11275. El-Kabbani, O. et al. *Nature Struct. Biol.* 1995, **2**:687-692. Barski, O.A. et al. Aldehyde reductase. Catalytic mechanism and substrate recognition. In: *Enzymology and Molecular Biology of Carbonyl Metabolism 6*, Weiner, H. et al. (eds.), Plenum Press, 1997, pp 443-451. Wermuth, B. Aldo-keto reductases. In: *Enzymology of Carbonyl Metabolism* 2, Flynn G.T. and Weiner H. (eds.), Alan R. Liss, Inc., 1985, pp 209-230. Maser, E. *Biochem. Pharmacol.* 1995, **49**:421-440. Aldo-Keto Reductase home page: http://pharme26.med.upenn.edu.
Ref. for DNA/AA Sequences	Bohren K.M. et al. *J. Biol. Chem.* 1989, **264**:9547-9551. (GB accession number of aldehyde reductase cDNA: J04794).

Aldose Reductase

Kurt M. Bohren and Oleg A. Barski

Synonyms	Low-Km aldehyde reductase
Abbreviations	ADR; ALR2; hAR; AKR1B1
Classifications	EC 1.1.1.21; Aldo-keto reductase AKR1B1
Description	A cytosolic NADP-dependent monomeric oxidoreductase, consisting of 315 aa with no carbohydrate chains. The enzyme is a member of the aldo-keto reductase superfamily, and catalyzes the reduction of a wide variety of carbonyl containing compounds to their corresponding alcohols. The enzyme follows an ordered bi bi mechanism with the coenzyme binding first and leaving last. Ubiquitous tissue distribution with highest levels in kidney medulla.
Structure	The atomic structure of the holoenzyme reveals a parallel ß8/α8 barrel motif with an unprecedented motif for NADP-binding oxidoreductases. The nucleotide enfolding loop (Gly-213-Leu-227) is the basis for the isomerization in the kinetic mechanism.
Molecular Weight	35,722
Sedimentation Coeff.	Unknown
Isolectric Point	5.9
Extinction Coeff.	13.6 (280nm, 1%, 1 cm); 48,640 M^{-1} cm^{-1} (280nm)
Enzyme Activity	NADPH-dependent reduction of sugar aldehydes, aliphatic and aromatic aldehydes. The efficiency of the dehydrogenase activity (oxidation of alcohols) is three orders of magnitude lower than the one of the reduction activity at pH 7.0. Dehydrogenase by IUB classification.
Coenzyme/Cofactors	NADPH/$NADP^+$. In the reduction reaction the pro-4R hydride (A-side stereospecific) is transferred from NADPH to the aldehyde substrate. In the oxidation reaction a hydride is transferred from the substrate to the A-side of $NADP^+$.
Substrates	D-Glucose, methylglyoxal, isocorticosteroids, DL-Glyceraldehyde, Nitro-benzaldehydes, etc.
Inhibitors	Thiolation of Cys-298 by physiological disulfides such as GSSG and cystine. Zopolrestat, Sorbinil, and many other negatively charged pharmacological inhibitors (-COOH and spirohydantoin-type) bind to the positively charged anion well formed by $Asp\text{-}43^-$/$Lys\text{-}77^+$/$Tyr\text{-}48^0$/$NADP^+$ in the active site of the enzyme (binding to the enzyme:$NADP^+$ complex).
Biological Function	Aldose reductase together with sorbitol dehydrogenase form the sorbitol or polyol pathway that converts glucose to fructose via sorbitol in extrahepatic tissues. Sorbitol is the source of fructose in seminal vesicles, and in renal medulla sorbitol is an osmotically active organic solute produced by transcriptionally regulated aldose reductase.

Physiology/Pathology	Participates in osmotic homeostasis in renal medullary cells. Wide substrate specificity and presence in many tissues suggests a role in detoxification of reactive aldehydes. During hyperglycemia of diabetes an increased flux through the sorbitol pathway plays an important role in the pathogenesis of several diabetic complications, such as cataracts and neuropathy.
Degradation	Unknown
Genetics/Abnormalities	A putative functional aldose reductase gene has been mapped to chromosome 7 at 7q35, and a pseudogene to chromosome 3. Sequences identified by in situ hybridization on chromosomes 1,9,11,13,14 may represent other active genes, non-aldose reductase homologous sequences, or pseudogenes. The gene extends over approximately 20 kilobases and consists of 10 exons giving rise to 1.4 kilobase mRNA. No protein variants are known.
Half-life	6 days (continuous lines of epithelial cells from rabbit renal inner medulla).
Concentration	>10% of soluble cell protein in cells mentioned above grown in medium made hyperosmotic to 600 mosmol/kg H_2O by addition of NaCl.
Isolation Method	Extraction by homogenization of brain, placenta, muscle, or kidney (preferably medulla) tissue or by using a French Press on *E.coli* that contain recombinant human protein, followed by ion exchange chromatography on DEAE-Sephadex. Subsequent affinity chromatography on Matrex Orange A yields at least 95% homogenous protein.
Amino Acid Sequence	The protein contains one acid labile Asp-Pro (D_{230}-P_{231}) bond. The extremely tight binding of the coenzyme (K_d = 5 nM) is mainly mediated by salt links of K_{262} and R_{268} with the pyrophosphate and 2'-phosphate of the nucleotide, and a loop (G_{213}SPDRPWAKPEDPSL$_{227}$) enfolding the nucleotide. Important active site residues are H_{110} (orientation of substrates), Y_{48} (proton donor during aldehyde reduction), K_{77} and D_{43} (lowering of pKa of Y_{48}), and C_{298} (thiol mediated activity changes). Homologue to 7 other human proteins (notably aldehyde reductase) and over 40 proteins if all eukaryotes are included.
Disulfides/SH-Groups	No disulfides, 7 free sulfhydrils at residues 44, 80, 92, 186, 199, 298, 303
General References	Petrash, J.M. et al. *Diabetes* 1994, **43**:955-959. Harrison, D.H. et al. *Biochemistry* 1994, **33**:2011-2020. Bohren, K.M. et al. *Biochemistry* 1994, **33**:2021-2032. Lee, A.Y.W. et al. *Proc.Natl.Acad. Sci.USA* 1995, **92**:2780-2784. Ko, B.C.B. et al. *J.Biol.Chem.* 1997, **272**:16431-16437. Aldo-Keto Reductase home page: http://pharme26.med.upenn.edu.
Ref. for DNA/AA Sequences	Bohren, K.M. et al. *J. Biol. Chem.* 1989, **264**:9547-9551 (GB accession number of human aldose reductase cDNA: J04795). Graham, A. et al. *J.Biol.Chem.* 1991, **266**:6872-6877 (GB accession number of hAR gene segment1: M59856; segment2: M59783; complete gene: L14440).

β_8/α_8barrel structure: Cartoon of human aldose reductase based on crystallographic data. The view is down the COOH-terminal end of the barrel with the coenzyme drawn in stick model. The eight α helices and the eight β strands are indicated as well as two extra helices, H_1 and H_2. Note the unusual sequestering of the coenzyme across the lip of the barrel with a movable loop folding over, and the center location of the nicotinamide (Courtesy of David H.T. Harrison, Medical College of Wisconsin)

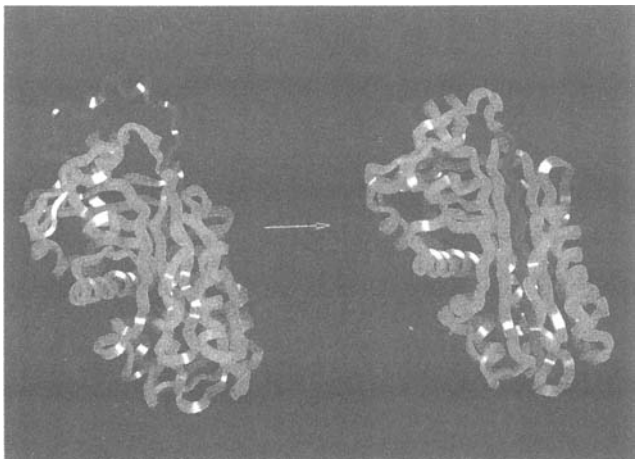
Alpha-1-Antichymotrypsin

Harvey Rubin and Michael Plotnick

Synonyms	None
Abbreviations	α1-ACT; α1-Achy
Classifications	Electrophoretic mob.: α1 region
Description	ACT is a plasma protein synthesized predominantly in the liver, but also in prostate cancers with low Gleason score, histiocytes, mast cells (immunocytochemical evidence), choroid plexus epithelium, skeletal muscle, astrocytes and astrocytomas, breast and intestinal epithelial cells, epithelial, stromal and endothelial layers of the human cornea, chorionic villi obtained from first trimester and term placenta, trophoblast cells in primary culture. ACT is found in amyloid deposits in Alzheimer's disease, Down's syndrome, normal aging, and hereditary cerebral hemorrhage with amyloidosis of Dutch origin. The protein is a single polypeptide chain of 398 aa starting from an N-terminal Asn-Ser-Pro. There are four glycosylation sites linked to Asn-68, Asn-102, Asn-161, Asn-205 (counting starting at N-terminal Asn-Ser-Pro). The reactive site is found at Leu-358-Ser-359.
Structure	A crystal structure of a modified form of ACT cleaved in the reactive center between Leu-358 and Ser-359 shows the large rearrangement of the reactive loop strand s4A into beta sheet-A, separating the P1 residue from the P1' residue by approximately 70A and the prime side residues associating with the beta sheet-C as strand s1C. The crystal structure of an intact, uncleaved variant of ACT has been solved. In this ACT variant, six residues containing the reactive site, TLLSAL, have been replaced by six residues around the reactive site of alpha-1-antitrypsin, IPMSIP. There are three additional replacements: A,A→G,T at the P10, P9 positions and V→T at the P10' position. The intact reactive loop of this structure is a distorted helix and strand s4A is not preinserted into beta sheet A.
Molecular Weight	68,000 (25% carbohydrate)
Sedimentation Coefficient	Unknown
Isoelectric Point	Microheterogeneity with pI range of 5.0 to 5.25
Extinction Coefficient	6.2 (1% w/v solution measured at 280nm where protein concentration was determined using interference optics).
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Can be cleaved in the reactive loop by a number of proteases thereby inactivating enzyme inhibitory activity.
Biological Functions	Serine protease inhibitor. Target enzymes include chymotrypsin, cathepsin G, prostate specific antigen.

Physiology/Pathology	Found in complex with prostate specific antigen (PSA) in circulation in diseases of the prostate. Also found in amyloid plaques in a variety of diseases. Inhibits NADPH oxidase dependent free radical production when in complex with target enzyme.
Degradation	ACT in complex with target enzyme cleared by certain cell surface receptors.
Genetics/Abnormalities	The gene encoding ACT is located at chromosomal position 14q32.1 where it is found approximately 20kB from the telomere and is the last of a cluster of other serpins in the order: corticosteroid binding globulin-alpha-1-protease inhibitor-like pseudogene-alpha-1-protease inhibitor-protein C inhibitor-ACT. The gene is organized in 5 exons and 4 introns spanning approximately 12 kilobases. The first intron is located in the 5' untranslated region of the gene. Exon 2 contains the initiation methionine codon. ACT is a DNA binding protein and the binding sites involve Lys residues at positions 210, 211, 212, 391 and 396. A common aa polymorphism, Thr and Ala, occurs in the signal peptide. ACT-Isehara-1: point mutation resulting in Met-89→Val. Clinical phenotype-occlusive cerebrovascular disease. ACT-Isehara-2: deletion of two bases in codon 391 resulting in frame shift downstream adding 10 aa residues. Normal clinical phenotype but approximately 40% of normally circulating ACT. ACT-Bochum-1 point mutation resulting in Leu-3→Pro. Clinical phenotype-chronic obstructive lung disease (COPD), 80% normal circulating levels of ACT. ACT-Bonn-1 point mutation resulting in Pro-227→Ala. Clinical phenotype-COPD, 60% normal circulating levels of ACT. Deletion of (14) (q24.3q32.1) resulting in normal alpha-1-protease inhibitor levels but 43% ACT levels. Clinical phenotype described in a 4 year old girl with delayed psychomotor development and characteristic facial features.
Half life	Unknown for the native, glycosylated protein.
Concentration	Plasma: 30-60 mg/100ml in adults, male = female, in newborn infants plasma level is 10mg/100ml.
Isolation Method	From pooled plasma, 50% ammonium sulfate fraction followed by two Cibacron Blue Sepharose chromatography steps. DNA-cellulose chromatography can be incorporated in the purification scheme to take advantage of the affinity of ACT for DNA.
Amino Acid Sequence	(HP)NSPLDEEN LTQENQDRGT HVDLGLASAN VDFAFSLYKQ LVLKAPDKNV IFSPLSISTA LAFLSLGAHN TTLTEILKGL KFNLTETSEA EIHQSFQHLL RTLNQSSDEL QLSMGNAMFV KEQLSLLDRF TEDAKRLYGS EAFATDFQDS AAAKKLINDY VKNGTRGKIT DLIKDLDSQT MMVLVNYIFF KAKWEMPFDP QDTHQSRFYL SKKKWVMVPM MSLHHLTIPY FRDEELSCTV VELKYTGNAS ALFILPDQDK MEEVEAMLLP ETLKRWRDSL EFREIGELYL PKFSISRDYN LNDILLQLGI EEAFTSKADL SGITGARNLA VSQVVHKAVL DVFEEGTEAS AATAVKITLL SALVETRTIV RFNRPFLMII VPTDTQNIFF MSKVTNPKQA
Disulfides/SH Groups	Single Cys at position 236
General References	Wei, A. et al. *Nature*, Structural Biology, 1994, **1**: 250-257. *SERPINS: Structure, Function and Biology*. Gettins, PGW, Patston, PA & Olson, ST. R.G. Landes Publishing Company. 1996

Ref. for DNA Sequence Rubin, H. et al. *J. Biol. Chem.* 1990, **265**:1199-1207.



Ribbon models of uncleaved rACT-P3P3' and cleaved antichymotrypsin illustrate the serpin conformational transition; coordinates of the latter structure were obtained from the Brookhaven Protein Data Bank.

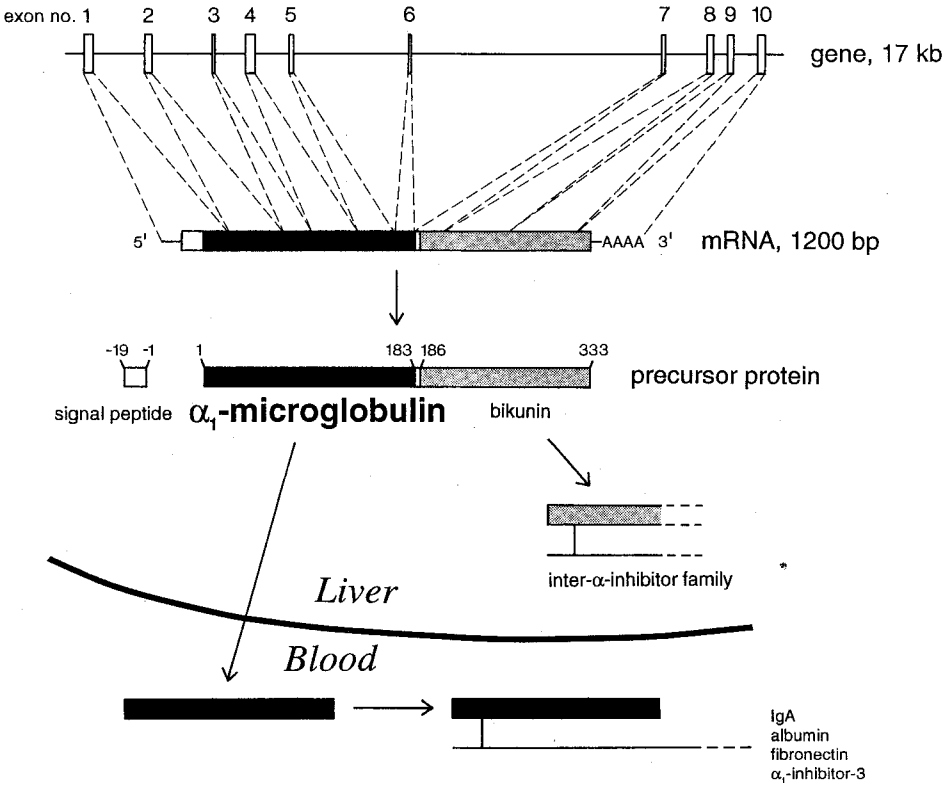
Alpha-1-microglobulin

Bo Åkerström

Synonyms	Protein HC (Human Complex-forming; Heterogeneous in Charge); α_1-Microglycoprotein
Abbreviations	α_1-m; α_1m; α1m
Classifications	Electrophoretic mobility: α_1
Description	A plasma protein with a strongly attached yellow-brown chromophore of unknown structure. Synthesized in liver, where it is translated from the same mRNA as bikunin, the light chain of the plasma protein inter-α-trypsin inhibitor. This diprotein is cleaved before secretion of α_1-m to the circulation. It is found in plasma in free form and covalently linked to monomeric IgA and albumin in man, and to α_1-inhibitor-3 and fibronectin in rats. Involved in lymphocyte regulation. By aa sequence homology found to be member of a protein superfamily consisting mostly of lipophilic carrier proteins, the lipocalins. α_1-m has been found in several mammalian species and in two fish species.
Structure	183 aa peptide folded to a globular molecule with a Stokes' radius of 2.85 nm and the frictional ratio 1.45. Glycoprotein with two N-linked complex type oligosaccharides at aa position 17 and 96, and one O-linked oligosaccharide at position 5. Crystallized members of the lipocalin protein superfamily, for example RBP and β-lactoglobulin, have a similar shape: two β-sheets, composed of eight or nine β-strands, form a hollow cone with a hydrophobic interior. α_1-m has not yet been crystallized.
Molecular Weight	26,700 (sedimentation equilibrium centrifugation); 31,000 (SDS polyacrylamide gel electrophoresis).
Sedimentation Coeff.	2.35 S
Isoelectric Point	4.3-4.8, electrophoretically heterogeneous
Extinction Coeff.	17.7-18.3 (280nm, 1%, 1cm) or 4.72×10^4 ($M^{-1} cm^{-1}$).
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Unknown biological function. Several indications that it is involved in immunoregulation: inhibition of antigen-stimulation of lymphocytes, inhibition of migration and chemotaxis of granulocytes, and a direct mito-genic effect on lymphocytes in in vitro assays. Association with proteinase inhibition: translation in liver as a diprotein with bikunin, the active part of the proteinase inhibitor inter-α-trypsin inhibitor, and in rat serum covalently linked to the proteinase inhibitor α_1-inhibitor 3, one of the three α_2-macroglobulin homologues.

1. Installment 1992 (revised 1997)

Physiology/Pathology	Elevated levels in plasma in patients with renal disorders, during third trimester of pregnancy and in cases with elevated plasma IgA. Subnormal plasma levels associated with HIV-infection, some liver dysfunctions, and with decreased plasma IgA. No other known deviations from normal plasma concentrations. Normally excreted in urine, and elevated urinary levels indicate renal tubular disorders.
Degradation	Eliminated from circulation by glomerular filtration in the kidneys.
Genetics/Abnormalities	A gene, which maps to 9q22.3-q33, is transcribed into a single mRNA, which codes for both α_1-m and bikunin. No genetic variants of α_1-m have been described.
Half-life	Unknown
Concentration	Plasma: 20-100 mg/L. Urine: 2-10 mg/24 hr volume.
Isolation Method	Isolated from pathological urine by size exclusion chromatography, preparative electrophoresis and ion-exchange chromatography. Isolated from plasma or serum by affinity chromatography on antibody columns. Expressed by insect cells infected by baculovirus carrying the α_1-m-bikunin gene. Isolated by immunoaffinity chromatography.
Amino Acid Sequence	Belongs to the protein superfamily lipocalins, and the sequence motifs GXWY and TD/NY, located at position 23-26 and 106-108, respectively, in human α1-m, are highly conserved in all members of the superfamily.
Disulfides/SH-Groups	One disulfide bond (between position 72 and 169) and one free sulfhydryl group (at position 34), suspected to be partially engaged in S-S binding with unknown substance.
General References	Ekström, B. et al. *Biochem. Biophys. Res. Commun.* 1975, **65**:1427-1433. Grubb, A. et al.*J. Biol. Chem.* 1986, **261:**14313-14320. Itoh, Y. and Kawai, T. *J. Clin. Lab. Anal.* 1990, **4:**376-384. Åkerström, B. and Lögdberg, L. *Trends Biochem. Sci.* 1990, **15**:240-243. Bratt, T. and Åkerström, B. *Protein Expr. Purif.* 1995, **6**:431-438.
Ref. for DNA Sequences	Lopez Otin, C. et al. *Arch. Biochem. Biophys.* 1984, **228:**544-554 (aa sequence). Kaumeyer, J.F. et al. *Nucleic Acids Res.* 1986, **14:**7839-7850 (cDNA sequence). Diarra-Mehrpour, M. et al. *Eur. J. Biochem.* 1990, **191:**131-139 (gene structure).



Synthesis of α_1-microglobulin. A ten-exon gene codes for α_1-microglobulin and the protease inhibitor bikunin. A precursor, α_1-m-bikunin, is translated in the liver and cleaved in the Golgi apparatus at a tetrabasic site RVRR. Free α_1-m is secreted into the bloodstream, where part of it forms covalent complexes with IgA and albumin (man) or α_1-inhibitor-3 and fibronectin (rat). Bikunin is linked to at least four other polypeptides, forming the plasma protein inter-α-inhibitor and related protein.

Alpha-1-proteinase inhibitor

Maurice C. Owen and Robin W. Carrell

Synonyms

Alpha-1-Antitrypsin, Antitrypsin, Metserpin

Abbreviations

α_1-PI, α_1-AT.

Classifications

Electrophoretic Mob.: α_1 region at pH 8.6

Description

A plasma protein, synthesized in the liver, comprising a single polypeptide chain of 394 amino acids with three carbohydrate side chains, N-linked to Asn-46, Asn-83 & Asn-247. These have a common core sequence of Asn-$(NAcGlc)_2$-$(Man)_3$- with either 2 or 3 antennae of NAcGlc-Gal-Sia. There are 3 common subtypes M1, M2 & M3 that arise from polymorphisms at 3 positions:

	213	376	101
M1	Val/Ala	Glu	Arg
M3	Val	Asp	Arg
M2	Val	Asp	His

The Protein contains a single reactive site at a Met-Ser sequence at position 358–359.

Structure

The tertiary structure is not yet determined for the native form, however the crystal structure of the protein cleaved at the reactive site and the structure of the homologous protein ovalbumin are known. There are 8 helices (A–H) and 3 large β-sheets (A–C).

Molecular Weight

51,000 (13% carbohydrate)

Sedimentation Coeff.

3.4 S

Isoelectric Point

Mean pI 4.8. IEF gives 3 main bands.

Extinction Coeff.

5.0 (280 nm,1%,1 cm)

Enzyme Activity

None

Coenzymes/Cofactors

None

Substrates

None

Inhibitors

Suscptible to inactivation by catalytic cleavage at or near reactive centre by a number of proteases (e.g. papain). Oxidation of Met-358 results in inactivation.

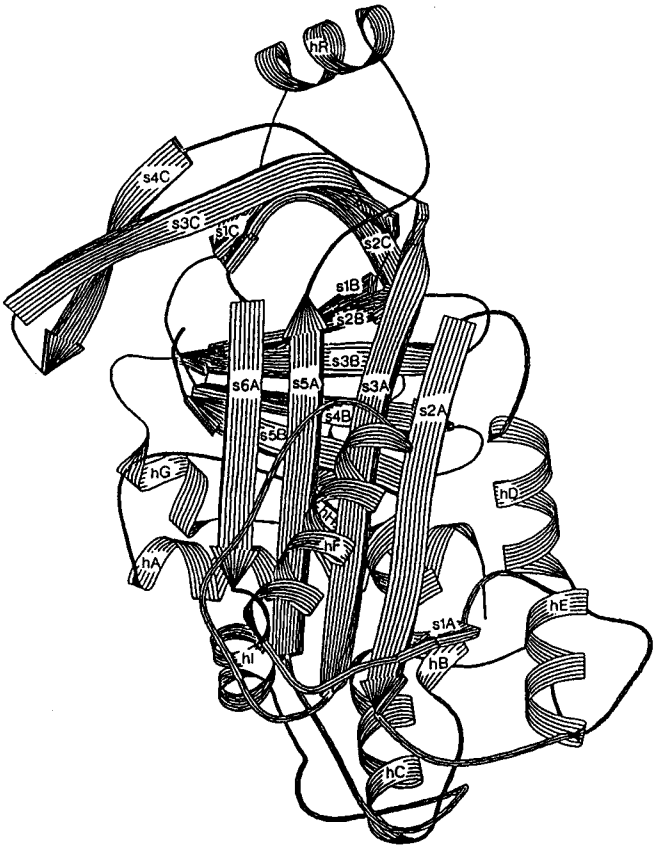
Biological Functions

Serine proteinase inhibitor. Elastase, with other hydrolytic enzymes (such as Cathepsin G), released on degranulation of neutrophils can degrade connective tissue proteins such as elastin, collagen and proteoglycan. The prime role of α_1-PI is as an inhibitor of neutrophil elastase, and it thus protects the susceptible lung tissue.

Physiology/Pathology

Essential to protect the lung tissue from proteolytic damage. Low plasma levels (less than 35% of normal) are associated with an increased risk of emphysema. Two common deficiency variants: (i) Z variant, serious deficiency with 15% normal plasma levels. Mutation of Glu-342 to Lys results

	in a secretion defect with accumulation within the hepatocyte. Also associated with liver disease – neonatal cholestasis in infants and cirrhosis and primary liver cancer in older adults. (ii) S variant, mild deficiency with 60% normal plasma levels. Increased risk of lung damage only in mixed SZ heterozygote.
Degradation	Complexes formed with target enzymes are cleared by the liver.
Genetics/Abnormalities	The gene is on chromosome 14 at q31–32.3, and is 12.2 kb long with 7 exons, the last four of which contain all information coding for the α_1-PI protein including the 24 amino acid signal sequence. More than 75 variants have been reported, the majority of which function normally. There are a number of rare null variants in which the protein is essentially absent from the plasma. The Z variant occurs with the highest frequencies in the northern Europeans (esp. Scandinavia), and the S is most common in southern Europe (Spain, Portugal).
Half-life	4–5 days (blood circulation)
Concentration	Plasma: 1.3 g/L, with the level increasing up to 4-fold during acute phase reactions.
Isolation Method	(a) 50–75% sat. ammonium sulphate fraction of plasma followed by affinity chromatography on gluathione-Sepharose, and ion exchange chromatography on DEAE-Sephadex. (b) Precipitation with polyethylene glycol from Cohn Fraction IV-1 followed by ion exchange chromatography on DEAE-Sepharose CL-6B. (c) Chromatography on Cibacron Blue Sepharose, followed by DEAE-Cellulose.
Amino Acid Sequence	EDPQGD AAQKTDT SHHDQD HPTFNKI TPNLAEF AFSL YRQLAH QSNSTN IFFSPVS IATAFAML SLGTKADT HDEI LEGLN FN LTQIPEA QIHEGFQ ELLRTLN QPDSQL QLTTGNG LFLSEG LKLVD KFLEDVK KLYHSE AFTVN FGDTE EAKKQI NDYVEKG TQGKIVD L VKELDRD TVFALVNY IFFKG KWE RPFE VKDTE EEDFHVDQ VTTVKVP MMKRL GMFNIQ HCKK LSSW VLL MKY LGN ATAIFF LPDEGK LQHLEN ELTHDII TKFLENE DRRSAS LHLPKL SITGTY DLKSVLG QLGITK VFSN GADLS GVTE EAPL KLSKA VHKAV LTIDEK GTEAAG AMFLEA IPMSI PPEVKF NKPFV FLMI EQNTK SPLFM GKVVN PTQK A member of a family of homologous inhibitors (the serpins).
Disulfides/SH-Groups	No disulfides, 1 free sulfhydryl at residue 232.
General References	Carrell, R. W. et al. *Nature* 1982, **298:** 329–334. Brantly, M. et al. *Am. J. Med.* 1988, **86** (6A)13–31. Heidtmann, H. and Travis, J. "Human $_1$-Proteinase Inhibitor". In: *Proteinase Inhibitors*, Battett, A. J. and Salvesen, G. (eds.) Elsevier Science Publishers, 1986; pp 441–456.
Ref. for DNA/AA Sequences	Long, G. L. et al. *Biochemistry* 1984, **23:** 4828–4837.



Serpin structure: The known crystallographic structure of modified α-1-proteinase inhibitor matches well that of the only intact serpin structure ovalbumin – shown here. The reactive centre is on helix R though in the inhibitor the loop is likely to be partially refolded into the A sheet.

Alpha2-HS glycoprotein

Willi Jahnen-Dechent and Werner Müller-Esterl

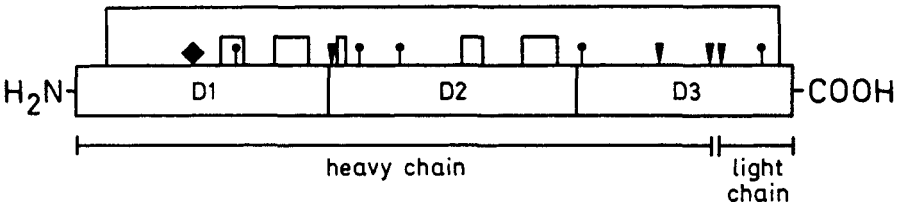
Synonyms	Fetuin
Abbreviations	α_2-HS (HS for discoverers, Heremans, Schmid); AHSG (genetic locus); A2HS_human (Swiss Prot)
Classifications	Family 4 of the cystatin superfamily; electrophoretic mobility α2-globulin fraction
Description	α2-HS is synthesized by the liver in a preform of 367 residues including a transient signal peptide of 18 residues. Posttranslational modifications include glycosylation (2 N- and 3 O-glycosylation sites), serine phosphorylation and limited proteolysis to generate the two-chain form of α_2-HS. The circulating form of α_2-HS consits of a heavy chain of 321 residues (positions 1 to 321 of the mature protein) linked to a light chain of 27 residues (positions 323 to 349) via a single disulfide bridge. Arg-322 is absent indicating that endoproteolytic cleavage of native α_2-HS is followed by exoproteolytic removal of this residue. Bulk isolation procedures have recovered a proteolytically trimmed two-chain protein lacking the 39 C-terminal residues ("connecting peptide region") of the heavy chain (this shortened version was designated "A-chain").
Structure	Multi-domain molecule with 3 modules: 2 cystatin-like units (domains D1 to D2) followed by the terminal domain D3 which is unique to α_2-HS/ fetuins. The proposed domain boundaries are given below. An X-ray structure of α_2-HS is not available.

protein	residues
α_2-HS glycoprotein	1-349
domain D1	1-118
domain D2	119-234
domain D3	235-349
heavy chain	1-321
light chain	323-349
A chain	1-282
connecting peptide	283-322

Molecular Weight	58,000 (apparent M_r by non-reducing SDS electrophoresis).
Sedimentation Coeff.	3.6 S
Isoelectric Point	4.1-4.7
Extinction Coeff.	5.6 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None

1. Installment 1992 (revised 1997)

Biological Functions	α_2-HS and the species homologue fetuins occur in high serum concentration during fetal life. They accumulate in bone and teeth as a major fraction of non-collagenous bone proteins suggesting a role in bone formation and resorption. α_2-HS/fetuins interfere with calcium salt precipitation and could therefore regulate bone mineral deposition and inhibit phase separation in serum. Rat fetuin (synonymous with pp63, BSP59) and human α_2-HS inhibit the tyrosine kinase associated with the insulin receptor.
Physiology/Pathology	α_2-HS is among the few negative acute phase proteins in man. The plasma concentrations of α_2-HS are elevated in the perinatal period. Immunoreactive α_2-HS has been demonstrated in the cortical plate of the developing brain and in calcified tissues.
Degradation	The pathways of α_2-HS degradation are presently unknown.
Genetics/Abnormalities	α_2-HS is encoded by a single gene(AHSG) located on chromosome 3q21-q29 clustering with kininogen (KNG) and histidine-rich glycoprotein (HRG), two more members of the cystatin superfamily. The corresponding rat gene is of 8 kBp and contains 7 exons. The mode of inheritance is autosomal codominant in man. At least 15 distinct allotypes of α_2-HS have been identified suggesting that multiple alleles of the α_2-HS locus exist. This genetic polymorphism is exploited in forensic hemogenetics where α_2-HS is widely used as a genetic marker.
Half-life	1.4 days (in rabbits)
Concentration	Mean plasma concentration 630 (range 400 - 850) mg/L (10.9 µmol/L).
Isolation Method	Classical isolation schemes from Cohn fraction V by a combination of precipitation methods and gel filtration plasma. Alternatively, selective precipitation from serum with calcium phosphate, zinc chelate affinity chromatography or triazine dyes are applied.
Amino Acid Sequence	Domains D1 to D2 share significant sequence homology with mammalian cystatins, kininogens and histidine-rich glycoprotein. Domain D3 seems to be unique for α_2-HS/fetuins.
Disulfides/SH-Groups	12 cysteine residues; 5 intra-chain disulfide loops of the heavy chain, 1 inter-chain loop spanning the heavy and light chain portions.
General References	Arnaud, P. et al. α_2-HS glycoprotein. *Methods Enzymol.* 1988, **163**:431-441. Jahnen-Dechent, W. et al. *Eur. J. Biochem.* 1994, **226**:59-69. Dziegielewska, K.M. and Brown, W.M. Fetuin, Springer-Verlag, Heidelberg, 1995.
Ref. for DNA/AA Sequences	Lee, C.C. et al. *Proc. Natl. Acad. Sci. USA* 1987, **84**:4403-4407. Kellermann, J. et al. *J. Biol. Chem.* 1989, **264**:14121-14128.



Gross structure of human α_2-HS

⊓ , disulfide loop; ┬ , carbohydrate side chain; ▼ , proteolytic cleavage site; ◆, potential Ca^{2+} binding site (modified after J. Biol. Chem. 1989, *264*:14121)

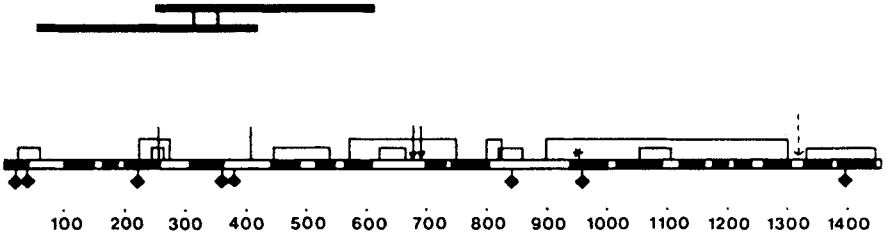
Alpha-2-macroglobulin

Lars Sottrup-Jensen

Synonyms	None
Abbreviations	α_2M
Classifications	Electrical mobility: α_2-fraction
Description	A circulating plasma protein, primarily synthesized in liver. A tetrameric molecule: two subunits are disulfide bridged forming dimers; two dimers are noncovalently associated into tetrameric structure. A glycoprotein containing appr. 10% carbohydrate attached to Asn-residues no. 32, 47, 224, 373, 387, 846, 968, 1401. An internal β-Cys-γ-Glu thiol ester is present in each subunit. Two strong binding sites for Zn^{2+} are present in addition to many weaker sites.
Structure	Native protein is open and flexible. Molecule with cleaved thiol esters forms well-defined H-like objects in electron micrographs. Complex with trypsin crystallized; no X-ray diffraction structure is available.
Molecular Weight	718,000 (diffusion and sedimentation). Subunit peptide: 160,795 (aa sequence, 1451 residues).
Sedimentation Coeff.	17.3-19.2 S dependent on conformational state.
Isoelectric Point	5.0-5.5
Extinction Coeff.	8.93 (280nm, 1%, 1cm)
Enzyme Activity	Complex with proteinase shows proteolytic activity against low molecular weight substrates. Molecule provides protection of the bound proteinase against high molecular weight proteinase inhibitors.
Coenzymes/Cofactors	None
Substrates	Proteinase complex cleaves small proteinase substrates (esters and anilides). Specificity of bound proteinase is unchanged.
Inhibitors	Inactivated by nucleophiles capable of cleaving the internal thiol esters.
Biological Functions	Scavenger for proteinases: Following specific limited proteolysis in 'bait' region extensive conformational changes result in deposition of proteinase within a large central cavity of molecule. Proteinase is still active but largely prevented from reaction with high molecular weight substrates and inhibitors. In a 'nascent' state activated thiol esters effect cross-linking of proteinase primarily through stable ε-Lys-γ-Glu bonds. Low molecular weight nucleophiles and other proteins including peptide hormones, growth factors, and lymphokines may also bind covalently. Through exposure of specific recognition sites on α_2M-proteinase the complex is rapidly cleared from circulation by receptor mediated endocytosis. Receptor is related to LDL-receptor, identical to a recently cloned lipoprotein-receptor-related protein (LRP) (α-chain, 515 kDa, β-chain, 85 kDa). Receptors are present in many cells and tissues, including fibroblasts, macrophages, liver and placenta.

1. Installment 1992 (revised 1997)

Physiology/Pathology	Controls together with other proteinase inhibitors activity of proteinases in circulation and tissues. No pathological states are known; not an acute phase reactant.
Degradation	Bait region cleaved protein (proteinase complex) is rapidly eliminated from circulation ($t_{1/2}$ appr. 2-5 min) and degraded in lysosomes. Protein is very stable to proteolytic degradation at pH > 5.
Genetics/Abnormalities	Gene located on chromosome 12. A few genetic variants but no deficiency states are known.
Half-life	Native protein appr. 10 days (circulation).
Concentration	Adults: appr. 1.2 g/L; infants: two to three-fold higher. Levels have earlier been estimated at 1.8-3.3 g/L.
Isolation Method	Polyethylene glycol prepicitation (4-12%), Zn^{2+} chelate affinity chromatography, Sephacryl S-300 gel chromatography. Protein is stable for 1-2 months when stored in solution at 0°C.
Amino Acid Sequence	Protein sequence: Sottrup-Jensen, L. et al. *J. Biol. Chem.* 1984,**259**: 8318-8327. Bait region: residues 666-706; internal thiol ester: residues 949 and 952; receptor recognition domain: residues 1314-1451. Plasma contains another α-macroglobulin, pregnancy zone protein (PZP), sequence identity with α_2-macroglobulin appr. 70 %. Proteins are homologous with complement proteins C3, C4, and C5, overall sequence identity appr. 25%.
Disulfides/SH-Groups	24 halfcystines; no free sulfhydryls; 11 intrachain disulfides, 2 interchain disulfides in dimer (antiparallel alignment of subunits). One free sulfhydryl after cleavage of thiol ester in subunit.
General References	Andersen, G. et al. Molecular model: 10-Å X-ray structure. *J. Biol. Chem.* 1995, **270**:25133-25141. Roberts, R. C. Alpha-2-Macroglobulin. *Reviews in Hematology.* 1986, Vol. **2**, pp. 129-224. PJD Publications Limited, Westbury, NY. Sottrup-Jensen, L.: α_2-Macroglobulin and Related Thiol Ester Plasma Proteins. In: *The Plasma Proteins,* Vol.**5** 1987, Putnam, F.W.(ed.), Academic Press, Orlando, FL, pp. 191-291. Sottrup-Jensen, L. α-Macroglobulins: Structure, Shape, and Mechanism of Proteinase Complex Formation. *J. Biol. Chem.* 1989, **264**:11539-11542. Delain, E. et al. Ultrastructure of alpha 2-macroglobulins. *Electron. Microsc. Rev.* 1992, vol. 5, pp. 231-281.
Ref. for DNA/AA Sequences	Kan, C.-C. et al. *Proc. Natl. Acad. Sci. USA* 1985, **82**:2282-2286. Sottrup-Jensen, L. et al. *J. Biol. Chem.* 1984, **259**:8318-8327.



Schematic representation of the 180 kDa subunit of human α_2-macroglobulin.
Filled segments represent stretches of > 20% sequence identity among α_2-macroglobulin and complement proteins C3, C4, and C5. The bait region is indicated by the pair of arrows, and the positions of carbohydrate groups are indicated by filled diamonds. The broken line indicates the cleavage point which leads to separation of the receptor recognition domain from the rest of the molecule. The position of the internal thiol ester is indicated by an asterisk. Inset shows the antiparallel alignment of subunits in the dimer.

Alpha-2-macroglobulin receptor

Dudley K. Strickland

Synonyms	Low Density Lipoprotein Receptor Related Protein
Abbreviations	LRP, α_2MR
Classifications	Cell surface receptor
Description	Cell surface receptor that is synthesized as a single chain precursor, and cleaved in the trans Golgi to a heavy and light chain. The heavy chain contains 3,923 aa, and has been demonstrated to bind most of the ligands. The light chain has 603 aa, and contains the transmembrane domain. The subunits are noncovalently associated on the cell surface. This receptor is widely distributed and is located in liver, lung and brain tissue, and on many cell types.
Structure	The three dimensional structure has not been determined.
Molecular Weight	502,724: receptor (calculated from sequence, and excluding carbohydrate); 436,658: heavy chain (calculated from sequence, and excluding carbohydrate); 66,066: light chain (calculated from sequence, and excluding carbohydrate).
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	1.36: receptor; 1.39: heavy chain; 1.16: light chain (all 280 nm, 1%, 1 cm)
Enzyme Activity	None known
Coenzymes/Cofactors	None known
Substrates	Ligands that bind to this receptor include: activated alpha-2-macroglobulin, tissue-type plasminogen activator (t-PA), urokinase-type plasminogen activator (u-PA), complexes of u-PA and t-PA with plasminogen activator inhibitor type I, apolipoprotein E-enriched lipoprotein particles, lipoprotein lipase, lactoferrin, and Pseudomonas exotoxin A.
Inhibitors	A 39 kDa receptor associated protein (RAP) co-purifies with the alpha-2-macroglobulin receptor. While the physiological function of RAP is not known, RAP blocks binding, internalization, and degradation of alpha-2-macroglobulin receptor ligands.
Biological Functions	Cell surface receptor that mediates the cellular internalization of various proteins leading to their degradation. Internalizes proteinase-inhibitor complexes such as alpha-2-Macroglobulin : proteinase complexes, PAI-1 complexes of uPA and tPA, and proteinases, such as uPA and tPA. Also may function in the uptake of apoE-enriched and lipoprotein lipase enriched lipoprotein particles, such as chylomicron remnants and beta-VLDL.
Physiology/Pathology	Functions in the hepatic clearance of certain proteinases and proteinase inhibitor complexes, and may function as the chylomicron remnant receptor. Is essential for successful embryo development. Disruption of the gene in mice blocks development of embryos around day 13.

Degradation	No studies available
Genetics/Abnormalities	The gene is on chromosome 12 in the segment q13-q14. No variants have yet been described.
Half-life	Unknown
Concentration	Many cell types, including hepatocytes, fibroblasts, monocytes, and macrophages express the alpha-2-macroglobulin receptor. Functional levels of receptor are decreased upon transformation.
Isolation Method	Human placenta is extracted with 50 mM octyl-beta,D-glucopyranoside. The receptor is purified from the extract by ligand affinity chromatography over immobilized activated alpha-2-macroglobulin. The 39 kDa receptor associated protein co-purifies, and can be removed by anion exchange chromatography over a Mono Q column.
Amino Acid Sequence	The alpha-2-macroglobulin receptor is a member of the LDL receptor family, which contains three other cell surface proteins, the LDL receptor, the VLDL receptor and glycoprotein 330. These four proteins are structurally analogous in that they each have several copies of cysteine-rich repeats. The alpha-2-macroglobulin receptor contains multiple copies of three types of repeating sequences: 24 epidermal growth factor like repeats, 31 LDL receptor ligand binding repeats, and numerous YWTD repeats.
Disulfides/SH-Groups	The number of disulfide bonds is not known.
General References	Strickland, D. K. et al. *J. Biol. Chem.* 1990, **265:** 17401–17404. Williams, S. E. et al. *J. Biol. Chem.* 1992, **267:** 9035–9040. Nykjaer, A. et al. *J. Biol. Chem.* 1992, **267:** 14543–14546. Herz, J. et al. *Cell* 1992, **71:** 411–421.
Ref. for DNA/AA Sequences	Herz, J. et al. *EMBO J.* 1988, **7:** 4119–4127.

Alpha-2-plasmin inhibitor

Nobuo Aoki

Synonyms	Alpha-2-antiplasmin; Plasma plasmin inhibitor; Plasmin inhibitor
Abbreviations	α_2-PI; α_2-AP; PI; P-PI
Classifications	Glycoprotein. Electrical mob.: α_2-fraction
Description	A circulating plasma glycoprotein, serine protease inhibitor, synthesized in the liver. A single chain molecule. Carbohydrate content 12 - 14%. Mature α_2PI with the N-terminal Met (Met-form) loses its N-terminal 12 residues during circulation and is converted to the protein with the N-terminal Asn (Asn-form). Both forms are present in plasma in a ratio of about 6 (Asn-form) to 4 (Met-form). In addition to the reactive site, α_2 PI contains affinity binding sites for plasmin(ogen) that make the inhibitor most effective and specific inhibitor of plasmin. In addition to these sites, the Asn-form has a functional factor XIII-catalyzed crosslinking site for fibrin. The crosslinking of this protein to fibrin increases the stability of fibrin clot by endowing the clot with the resistance against plasmin-mediated fibrinolysis.
Structure	Not yet crystallized.
Molecular Weight	68,400 for Met-form and 58,700 for Asn-form (aa sequence and 14% carbohydrate); 65,000-70,000 (reduced SDS-PAGE); 63,000 (sedimentation equilibrium)
Sedimentation Coeff.	3.43 S
Isoelectric Point	Unknown
Extinction Coeff.	7.03 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	No biological inhibitor to this protein.
Biological Functions	α_2-PI is one of the major serine protease inhibitors, in plasma mainly as an inhibitor of plasmin because of its high affinity for plasmin(ogen). Two step reaction: First, the inhibitor rapidly forms a reversible complex with plasmin through noncovalent binding between amino terminal kringle structures of plasmin and their complementary site (plasminogen binding site) on the carboxyl terminal region of α_2-PI. Second, an enzymatically inactive stable complex is formed through covalent binding between plasmin active site serine and reactive site of α_2-PI. α_2-PI possesses two additional characteristic functions: it interferes with the binding of plasminogen to fibrin and it undergoes factor XIII-catalyzed crosslinking with fibrin during clotting. The former function retards initiation of fibrinolysis and the latter protects fibrin clots from plasmin. Binding of plasminogen to fibrin is one of the initial steps of naturally occurring

	endogenous fibrinolysis, and α_2-PI hinders the binding by cempeting with fibrin for its binding to plasminogen, thereby retarding an initiation of fibrinolysis. When blood clots, approx. 30% of α_2PI, mostly Asn-form, present in plasma are rapidly crosslinked to fibrin α-chain by activated coagulation factor XIII. α_2-PI crosslinked to fibrin becomes condensed in the clot while platelet-mediated clot retraction is taking place. The cross-linked inhibitor immediately inhibits in situ plasmin generated on fibrin molecules. These two properties in addition to rapid inhibition of plasmin are characteristic to α_2-PI and render it a specific and most effective inhibitor of fibrinolysis.
Physiology/Pathology	When fibrin is formed, plasminogen and plasminogen activator are bound to fibrin. Fibrin-bound plasminogen is activated by fibrin-bound plasminogen activator, and plasmin thus formed on fibrin degrades fibrin. This naturally occurring endogenous fibrinolysis is efficiently inhibited by α_2-PI. Three functional properties specific to α_2-PI are responsible. Particulary factor XIII catalyzed binding of the inhibitor to fibrin stabilizes crosslinked fibrin clots by protecting the clots from endogenously occurring fibrinolysis. Failure of this protection causes spontaneous dissolution of hemostatic plugs before restoration of injured vessels, resulting in a hemorrhagic tendency such as seen in a patient with deficiency of α_2-PI or factor XIII. Congenital deficiency of α_2-PI exhibits a life-long severe hemorrhagic tendency. Hemorrhagic episodes include umbilical bleeding, prolonged bleeding from a wound, hematoma, hemarthroses, and bleeding into the central nervous system.
Degradation	The C-terminal region of approximately 20 aa containing plasminogen binding region is gradually lost in circulation by a proteolytic cleavage. This partially degraded form, non-plasminogen binding form, constitutes approximately 30% of the protein in circulation.
Genetics/Abnormalities	The genetic symbol, PLI; 10 exons and 9 introns, 16-kb, chromosome 17 at p13. A restriction fragment length polymorphism, generated by a deletion in intron 8. More than 10 families with inherited deficiency of this protein in plasma. One family with circulating abnormal dysfunctional protein. Two families with the deficiency were found to have mutations leading to deficient intracellular transport of protein from endoplasmic reticulum to Golgi's apparatus; a trinucleotide deletion in exon 7 resulting in deletion of Glu-149 (revised numbering), and another family having a frameshift mutation leading to replacement and elongation of the carboxyl terminal sequence. Dysfunctional circulating abnormal protein; alanine insertion near the reactive site.
Half-life	2.6-3.3 days (blood circulation)
Concentration	Plasma: 69 ± 6 mg/L
Isolation Method	Isolated from plasminogen-depleted plasma (plasminogen depleted by lysine-Sepharose) by affinity chromatography on plasminogen-Sepharose followed by ion-exchange chromatography, or by immunoaffinity chromatography using monoclonal antibody-coupled Sepharose.
Amino Acid Sequence	A single chain of 464 aa. N-terminal: MEPLGRQLTSGP$^{\downarrow}$NQ* EQVSPLTLLK- ($^{\downarrow}$cleavage site for yielding Asn-form. * crosslinking site for fibrin). C-terminal: -PDKLK*LVPPMEEDYPQFGSPK* (Plasminogen binding region. * involved in the binding). Reactive site for plasmin: R^{376} - M^{377} (revised numbering). 4 potential asparagine-linked glycosylation sites. Sequence homology: 23-28% with plasma serine protease inhibitors.

Disulfides/SH-Groups	4 SH-groups.
General References	Aoki, N. and Harpel, P.C.: Inhibitors of the fibrinolytic enzyme system. *Semin. Thromb. Hemost.* 1984, **10**:24-41. Aoki, N. et al. *Methods Enzymol.* 1993, **223**:185-197. Aoki, N.: Fibrinolysis. Its initiation and regulation. *J. Protein Chem.* 1986, **5**:269-277. Aoki, N. Genetic abnormalities of the fibrinolytic system. *Semin. Thromb. Hemost.* 1984, **10**:42-50. Moroi, M. and Aoki, N. Isolation and characterization of α_2-plasmin inhibitor from human plasma. *J. Biol. Chem.* 1976, **251**:5956-5965. Wiman, B., and Collen, D. On the kinetics of the reaction between human antiplasmin and plasmin. *Eur. J. Biochem.* 1978, **84**:573-578. Wiman, B., and Collen, D. On the mechanism of the reaction between human antiplasmin and plasmin. *J. Biol. Chem.* 1979, **254**:9291-9297. Tamaki, T., and Aoki, N. Cross-linking of α_2-plasmin inhibitor to fibrin catalyzed by activated fibrin-stabilizing factor. *J. Biol. Chem.* 1982, **257**: 14767-14772.
Ref. for DNA/AA Sequences	Sumi, Y et al. cDNA: *J. Biochem.* 1986, **100**:1339-1402. Holmes, W.E. et al. *J. Biol. Chem.* 1987, **262**:1659-1664 (Gene Bank/EMBL J02654). Tone, M. et al. *J. Biochem.* 1987, **102**:1033-1041. Hirosawa, S. et al. Genomic DNA. *Proc. Natl. Acad, Sci. USA* 1988, **85**: 6836-6840 (Gene Bank/EMBL J03830). Miura, O. et al. *Biochemistry* 1989, **28**:4934-4938 (Gene Bank M27147 allele A5', M27148 allele B, M27149 allele A3'). Kato, A. et al. Gene Mapping. *Cytogenet. Cell Genet.* 1993, **62**:190-191. Holmes, W.E. et al. Abnormal Genes. *Science* 1987, **238**:209-211. Miura, O. et al. *J. Clin. Invest.* 1989, **83**:1598-1604. Miura, O. et al. *J. Biol. Chem.* 1989, **264**:18213-18219. Sumi, Y. et al. AA Sequences *J. Biochem.* 1989, **106**:703-707. Bangert, K. et al. *Biochem. J.* 1993, **291**:623-625. Enghild, J.J. et al. *Biochem. J.* 1993, **291**:933-938. Koyama, T. et al. *Biochem. Biophys Res. Commun.* 1994, **200**:417-422.

Alpha-amylase

Michio Ogawa

Synonyms	Diastase, Ptyalin, Glucogenase pancreatic amylase, salivary amylase
Abbreviations	Amy, p-Amy, s-Amy, amy1, amy2, amy3, amy2A, amy2B
Classifications	1,4-α-D-glucan 4-glucanohydrolase (E C 3.2.1.1)
Description	One of digestive enzymes secreted by pancreas and salivary gland. Both produced by the expression of different genes. Those two isozymes differ in molecular size, isoelectric point, and antigenic properties. Recent cloning studies revealed the third amylase gene, and the product amy2B detected in a lung carcinoid tissue, pancreas, liver and thyroid gland. But the main producing organ and enzymatic properties are unknown. A part of secreted amylase enters into blood stream, and the determination of serum amylase used for the diagnosis of pancreatic diseases.
Structure	A single polypeptide chain with (family A of salivary amylase) or without (family B and pancreatic amylase) sugar moiety. Two possible N-glycosylation sites in three amylase isozymes. Family A of salivary amylase contains 6 mol/mol glucosamine, 3 mol/mol fucose, 2 mol/mol mannose and 2 mol/mol galactose.
Molecular Weight	Pancreatic 54,000; salivary-family A 62,000; family B 56,000 (diffusion and sedimentation); pancreatic (amy2A) 56,172; salivary (amy1) 56,502 (nucleotide sequence)
Sedimentation Coeff.	4.2S (pancreatic)
Isoelectric Point	amy2A 7.5; amy1 6.5; amy2B (expr in yeast) 7.4
Extinction Coeff.	21.5 (280 nm, 1%, 1 cm), pH 6.5, pancreatic.
Enzyme Activity	Km(M): G4–1.7×10^{-3}, G5–1.7×10^{-2}. G6–7.5×10^{-4} (pancreatic); G6–7.5×10^{-4} (salivary)
Coenzymes/Cofactors	None
Substrates	Starch, glycogen
Inhibitors	Inhibited by Ca-chelator. Salivary amylase inhibited by wheat germ amylase inhibitor.
Biological Functions	A digestive enzyme which hydrolyzes starch and glycogen in alimentary canal. Serum amylase level kept in a certain constant range, but no biological function in serum is known. Postoperative transient salivary-type hyperamylasemia is occasionally reported. Also expressed in several malignant tumors (lung, ovarian, gastric, breast or thyroid cancer). Most of them produce salivary-type amylase. Biological function of cancer-produced amylase is unrevealed.
Physiology/Pathology	Measurement of serum amylase is an established diagnostic test for pancreatic disease (especially for acute and chronic relapsing pancreatitis). Measurement of amylase content in pancreatic juice is used to estimate pancreatic exocrine dysfunction.

Degradation	Normally eliminated from circulation by the kidney. Family A of salivary amylase (glycoprotein) may be eliminated by the liver.
Genetics/Abnormalities	Encoded by different, but closely related three genes. Located within a limited region of chromosome 1 (1s21).
Half-life	Unknown (80 min in baboon)
Concentration	Serum: total 122.1 µg/L (55–250), pancreatic 61.7 µg/L (15.7–107.7)
Isolation Method	Efficient purification achieved by cornstarch affinity when the starting material is pancreatic juice or saliva. Also purified by repeated ion-exchange chromatography and gel filtration.
Amino Acid Sequence	Consisted of 496 aa. Sequence homologies: salivary vs. pancreatic 98.0%, salivary vs. amy2B 97.5%, pancreatic vs. amy2B 98.8%.
Disulfides/SH-Groups	12 cysteine residues in salivary and pancreatic amylase. In amy2B, 11 are conserved and one (aa position 118) is substituted by a Ser residue.
General References	Ogawa, M. et al. *Methods Enzymol.* 1981, **74:** 290–298. Tomita, N. et al. *Cancer Res.* 1988, **48:** 3292–3296. Tomita, N. et al. *Gene* 1989, **76:** 11–18. Shiosaki, K. et al. *Gene* 1990, **89:** 253–258. Groot, P. C. et al. *Genomics* 1990, **8:** 97–105. Higashiyama, M. et al. *J. Clin. Pathol.* 1991, **44:** 144–146.
Ref. for DNA/AA Sequences	Nakamura, Y. et al. *Gene* 1984, **28:** 263–270. Wise, R. J. et al. *Mol. Biol. Med.* 1984, **2:** 307–322. Nishide, T. et al. *Gene* 1986, **50:** 371–372. Gumucio, D. L. et al. *Mol. Cell Biol.* 1988, **8:** 1197–1205. Emi, M. et al. *Gene* 1988, **62:** 229–235. Samuelson, L. C. et al. *Nucleic Acids Res.* 1988, **16:** 8261–8276. Yokouchi, H. et al. *Gene* 1990, **90:** 281–286.

α-L-iduronidase

Peter R. Clements and John J. Hopwood

Synonyms	α-L-iduronide iduronohydrolase
Abbreviations	IDUA
Classifications	EC 3.2.1.76
Description	A lysosomal protein, α-L-iduronidase is synthesized in the endoplasmic reticulum as a 653 aa precursor and modified by removal of a 26 aa signal peptide. N-linked glycosylation also occurs during synthesis in the ER with high mannose oligosaccharides being transferred to up to six potential sites. Glycosyl trimming and mannose-6-phosphorylation occurs in the Golgi prior to targetting to a mannose-6-phosphate receptor in the trans Golgi network. Vesicular budding of the receptor-mediated TGN regions occurs to create prelysosomal compartments and finally lysosomes, when fused with incoming substrate containing vesicles. Alternative glycosyl modification to form complex type oligosaccharides occurs in some tissues, particularly liver. Further proteolytic cleavage occurs specific to different tissues resulting in "mature" forms which may have up to 7 polypeptides derived from the original gene product all forming a catalytic entity.
Structure	Unknown
Molecular Weight	65,000 Da: native form of liver (by gel permeation either in high ionic strength, 0.5 M NaCl, or at pH 5.0). Associates with itself and with other proteins and surfaces at low ionic strength to give multiple active species during gel filtration or by isoelectric focussing. 74 kDa: precursor form (SDS-PAGE). Polypeptide components: 65, 60, 49, 44, 18, 13 kDa: polypeptide components (SDS-PAGE). Tissue specific differences in the relative amounts of these components occur.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown, elutes at pH 7.5 on chromatofocussing
Extinction Coeff.	Unknown
Enzyme Activity	Active towards α glycosidic bonds between α-L-iduronic acid residues on the non-reducing terminus and a range of aglycone moieties. The enzyme is an exohydrolase working only on α-L-iduronic acid when it is in the non-reducing terminal position on the polysaccharide chain. The bond cleaved by this hydrolase is between the C1 carbon of α-L-iduronic acid and oxygen of the glycosidic bond. Specific activity for purified liver α-L-iduronidase = 100 units/mg.
Coenzymes/Cofactors	None known
Substrates	Fluorogenic and chromogenic substrates: 4-methylumbelliferyl-α-L-iduronide (two Km values 37 μM and 1.82 mM), phenyl-α-L-iduronide. Radiolabelled substrates derived from the natural substrates heparan sulphate and dermatan sulphate:

	1. from heparan sulphate O-(α-L-iduronic acid)-(1 → 4)-anhydromannitol 6-sulphate (Km = 2.2 μM); O-(α-L-iduronic acid)-(1 → 4)-anhydromannitol (Km = 20 μM); O-(α-L-iduronic acid)-(1 → 4)-D-O-(α-2-sulphaminoglucosamine 6-sulphate)-L-O-(α-iduronic acid 2-sulphate)-O-D-2,5-anhydromannitol 6-sulphate (Km = 1.0 μM); O-(α-iduronic acid)-(1 → 4)-D-O-(α-2-N-acetylglucosamine 6-sulphate)-L-O-(α-iduronic acid 2-sulphate)-O-D-2,5-anhydromannitol 6-sulphate (Km = 1.0 μM). 2. from dermatan sulphate O-(α-L-iduronic acid)-(1 → 3)-anhydrotalitol 4-sulphate (Km = 76 μM).
Inhibitors	Cu^{2+} (Ki = 5 μM, non-competitive); SO_4^{2-} (Ki = 3 mM, competitive); Cl^- (Ki = 64 mM, competitive); 2,5 anhydromannitol 6-sulphate (Ki = 25 μM); 2,5 anhydromannitol (Ki = 6 μM); N-acetylglucosamine 6-sulphate (Ki = 23 μM); N-acetylglucosamine (Ki = 67 μM); O-(α-L-idose)-(1 → 4)-2,5 anhydromannitol 6-sulphate (Ki = 2μM). Inhibition has been observed with the above range of substrates.
Biological Functions	Degrades by exohydrolase action in the lysosome, endocytosed heparan sulphate and dermatan sulphate fragments which have been trimmed by endosomal endohydrolase action to fragments of approximately 5 kDa and then transferred to the lysosomal compartment. Part of the pathway of turnover of cell surface proteoglycans occuring in every cell type, with turnover times of one to four hours. Products of this process are iduronic acid, glucuronic acid, N-acetylgalactosamine, N-acetylglucosamine and sulphate ions.
Physiology/Pathology	A deficiency of α-L-iduronidase in humans results in the disease Mucopolysaccharidosis type I which results in the accumulation of undegraded fragments of heparan sulphate and dermatan sulphate in the lysosome. Clinically this manifests as a spectrum of symptoms ranging from severe with skeletal deformities, mental retardation, coarse hirsute facies, corneal clouding, hepatosplenomegaly, umbilical hernia and early death, to mild, with mild skeletal deformities, corneal clouding but normal lifespan and intelligence. Medically this condition is known as Hurler syndrome (severe phenotype) and Scheie syndrome (mild phenotype).
Degradation	α-L-Iduronidase exhibits a relatively long lifetime in the hydrolytic environment of the lysosome. No information is yet available on its ultimate fate.
Genetics/Abnormalities	Gene location is on human chromosome 4p16.3. Inheritance of the α-L-iduronidase deficiency is autosomal and recessive.
Half-life	A 69 kDa form stable 5 days in skin fibroblasts
Concentration	Liver 0.0001 units/mg protein or 9 units/kg wet weight.
Isolation Method	Purification from human liver by a series of: 1. Salt extraction/lipid extraction/centrifugation; 2. Concanavalin A-Sepharose chromatography combined with Blue A-Agarose chromatography; 3. CM Sepharose chromatography; 4. Cu^{2+} chelating Sepharose chromatography. Purification from lung, kidney or urine was also possible. Current availability of tissues make placenta or leucocytes (from whole blood) the best possible choices. Isolation by monoclonal antibody affinity chronatography is more effective, but the antibodies are not generally available.

Amino Acid Sequence	A 20% identity with β-xylosidase from C. saccharolyticum with further 32.8% similarity was the only homology found in the database.
Disulfides/SH-Groups	Unknown
General References	Clements, P. R. et al. *Eur. J. Biochem.* 1985, **152:** 29–34. Clements, P. R. et al. *Biochem. J.* 1989, **259:** 199–208. Scott, H. S. et al. *Am. J. Hum. Genetics* 1990, **47:** 802–807. Scott, H. S. et al. *Proc. Natl. Acad. Sci. USA* 1991, **88:** 9695–9699. Taylor, J. A. et al. *Biochem. J.* 1991, **274:** 263–268. Freeman, C. and Hopwood, J. J. *Biochem. J.* 1992, **282:** 899–908. Scott, H. S. et al. *Genomics* 1992, **13:** 1311–1313.
Ref. for DNA/AA Sequences	Scott, H. S. et al. *Proc. Natl. Acad. Sci. USA* 1991, **88:** 9695–9699. Scott, H. S. et al. *Genomics* 1992, **13:** 1311–1313.

Aminopeptidase N

Hans Sjöström and Ove Norén

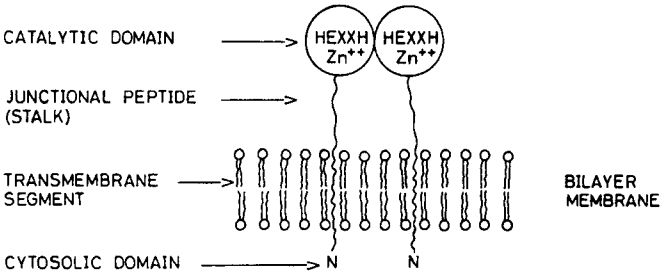
Synonyms	Alpha-aminoacyl-peptide hydrolase, aminopeptidase M, microsomal aminopeptidase, microvillar aminopeptidase, particle-bound aminopeptidase, amino-oligopeptidase, alanine aminopeptidase, alanyl aminopeptidase, leucyl-beta-naphtylamidase, fibroblast surface glycoprotein.
Abbreviations	APN, AMPEPN, mAAP, CD13, gp150, p146, FSG
Classifications	EC 3.4.11.2; integral membrane protein
Description	Most abundant in kidney proximal tubule cells and in intestinal eptihelial cells. In these cells the enzyme is localized to the brush border of the apical part of the cell. The enzyme is as well present in the canalicular membrane of the hepatocytes. On myeloid cells (CD13/gp150) and on alveolar epithelial cells (p146) it was first described as a differentiation antigen. The enzyme is present on the plasma membrane of most cells studied, for example on the plasma membrane of the fibroblasts. It can be released from the membrane by proteolytic enzymes (hydrophilic form) and by non-ionic detergents (amphiphilic form). There is only one gene for aminopeptidase N and aminopeptidase N sequenced from different tissues shows the same structure indicating that there is only one type of gene product. However, the degree of glycosylation (N-linked and O-linked) vary between cell types giving rise to differences in electrophoretic mobility. The pig enzyme is tyrosine-sulphated.
Structure	The structure of the human enzyme has not been studied by any physicochemical technique. Electron-microscopy of the purified pig enzyme reconstituted into liposomes shows it to be a dimeric structure of 5.5×13.5 nm at 5 nm distance out from the membrane. This distance is spanned by a peptide-segment (junctional peptide or the stalk) sensitive to limited proteolysis.
Molecular Weight	140,000 Da (SDS-PAGE under reducing and non-reducing conditions); 280,000 Da (gel filtration under non-denaturing conditions); the enzyme thus is a homodimer held together by non-covalent forces.
Sedimentation Coeff.	
Isoelectric Point	4.0–4.7 (PHASTGelIEF 3–9, Pharmacia)
Extinction Coeff.	Unknown. 15.6 (280 nm, 1 %, 1 cm) for hydrophilic pig intestinal enzyme.
Enzyme Activity	Releases preferentially amino acids with aliphatic side chains from polypeptides. Can as well release acidic and basic amino acids. Amino acids penultimately to proline and proline itself are not released.
Coenzymes/Cofactors	One zinc-ion is bound per 140,000 Da polypeptide.
Substrates	Natural: polypeptides with unblocked N-terminal amino acids other than proline and aminoacyl groups linked to proline. The enzyme can also hydrolyze dipeptides. Chromogenic substrates: L-alanine or L-leucine p-nitroanilides; L-alanine or L-leucine beta-naphtylamides; L-alanine or L-leucine 4-methoxy-beta-naphtylamides.

Inhibitors	Amastatin and bestatin. Zinc-ion chelators like EDTA and O-phenantroline. Bile acids.
Biological Functions	In the small intestine the enzyme has a role in the final phase of protein digestion. On the surface of the brush border of the intestinal epithelial cells it hydrolyzes peptides generated in the lumen by the combined action of pancreatic endopeptidases and carboxypeptidases. In other cells the biological function is unknown. It has been suggested that the enzyme is protecting cells against unwanted peptide signals, or that the enzyme terminates important peptide signals. A role in cell nutrition has also been suggested. The enzyme serves as a receptor for a corona virus. Aminopeptidase N is synthesized on the RER. The primary translation product is 110,000 Da. This is co-translationally high-mannose glycosylated to 140,000 Da form which in the Golgi apparatus undergo complex glycosylation and O-glycosylation to the final form which has a mw of $\simeq$ 150,000 Da. This form is transported by smooth vesicles to the microvillar membrane.
Physiology/Pathology	Decreased levels in the intestinal epithelium may impair protein assimilation.
Degradation	The enzyme is released into the intestinal lumen from the villi together with the desquamated epithelial cells. It may as well be released by the pancreatic proteases from intact or desquamated cells.
Genetics/Abnormalities	Localized to chromosome 15q13-qter. The gene shows DraI polymorphism with allele specific fragments and frequencies of 5.3 kb (0.75) and 6.5 kb (0.25).
Half-life	Unknown
Concentration	About 5% of the microvillar membrane proteins.
Isolation Method	Several monoclonal antibodies are available for CD13. These can be used for the immuno-purification of the enzyme. The richest sources of the enzyme is kidney cortex and small intestinal epithelium.
Amino Acid Sequence	The enzyme has zinc-ion binding motif HEXXH. It also has a 23 aa long hydrophobic stretch between position 10–32 which function as an RER-translocation signal which is not cleaved off. Instead it forms the membrane spanning part. Sequence 33–68 forms the junctional peptide/stalk and is very rich in serine and threonine. It is thought to be O-glycosylated. There are 11 potential N-linked glycosylation sites. There are two potential sulphation sites.
Disulfides/SH-Groups	Unknown. There are 7 Cys and some of them might be involved in disulfide formation.
General References	Danielsen, E. M., et al. *Biochem. J.* 1984, **221:** 1–14. Look, T. A., et al. *J. Clin. Invest.* 1985, **75:** 569–579. Nakanishi, M., et al. *J. Biochem.* 1989, **106:** 818–825. Hansen, G. H., et al. *Eur. J. Cell Biol.* 1987, **43:** 253–259. Norén, O., et al. The enzymes of the enterocyte plasma membrane. In: *Molecular and Cellular Basis of Digestion.* Desnuelle, P., et al. (eds.) Elsevier, Amsterdam 1986, pp. 335–365. Semenza, G. *Annu. Rev. Cell Biol.* 1986, **2:** 255–313.

Ref. for DNA/AA Sequences

Olsen, J., et al. *FEBS Lett.* 1988, **238:** 307–314.
Accession no. X13276, Entry name: HSAMPEPN in EMBL Data Library.
Look, T., et al. *J. Clin. Invest.* 1989, **83:** 1299–1307.
Accession no. M22324, Entry name HSAMIPO1 in Gene Bank.

Molecular Model



Aminopeptidase P

Greet C. Vanhoof and Filip J. Goossens

Synonyms	X-Pro aminopeptidase; Peptidase P; Microsomal proline aminopeptidase; Aminoacyl-prolyl-peptide hydrolase
Abbreviations	APP; AmP; AMPP; mPAP; PEPP; XPNPEP
Classifications	EC 3.4.11.9. Aminopeptidases; peptidase family M24
Description	APP has been located in a variety of human tissues, with high specific activity in testis, placenta, spleen and kidney. A glycosyl-phosphatidylinositol anchored membrane bound form and a soluble form are distinguished, which differ in their pH optimum and temperature stability. In kidney, intestine, and lung, the membrane bound form of the enzyme is predominant. APP is present on the surface of human endothelial cells and on erythrocytes. In erythrocytes the soluble form is most abundant, while it is the only form in platelets and leukocytes. APP activity increases after stimulation of lymphocytes. Preliminary data reveal that APP activity is decreased in malignant tissues.
Structure	Dimer (leukocytes, erythrocytes) or trimer (platelets) of identical subunits. Membrane-bound form is glycosyl-phosphatidylinositol anchored and glycosylated. Soluble form of platelets and leukocytes is not glycosylated. APP lacks the typical HEXXH Zn^{2+} binding sequence of metalloproteases.
Molecular Weight	Native: soluble form: platelets: 223 kDa; leukocytes: 140 kDa; erythrocytes: 155 kDa (gel filtration) Membrane bound form: lung: 188 kDa (gel filtration) Subunit: platelets and leukocytes: 71 kDa (SDS PAGE)
Sedimentation Coeff.	Unknown
Isoelectric Point	4.9 (platelets and lymphocytes) (IEF; RF3)
Extinction Coeff.	Unknown
Enzyme Activity	Unusual restricted specificity: removes N-terminal aa from peptides and proteins provided that the penultimate residue is a proline. A free N-terminus is obligatory. The cleavable bond must be in *trans* configuration. APP can hydrolyze dipeptides although at much lower rates than tripeptides and longer peptides.
Coenzymes/Cofactors	Mn^{2+}-dependent (for part of the substrates), Zn-containing metalloprotease (one atom Zn per polypeptide chain)
Substrates	Peptides and proteins with unblocked N-terminus and penultimate proline residue. Natural substrates: APP has a high affinity for bradykinin, cleaves substance P, neuropeptide Y, peptide YY, morphiceptin, kentsin, the biologically active Lys-Pro-Gln-Leu-Trp-Pro (C-terminal fragment of calcitonin-related peptide), and interleukin-6. Synthetic substrates: Arg-Pro-Pro, Gly-Pro-Hyp, Gly-Pro-*p*-nitroanilide, Arg-Pro-Pro (^{3}H)benzylamide, and the intramolecularly quenched fluorogenic substrate Lys(2,4-dinitrophenyl)Pro-Pro-NH-CH_2-CH_2-NH-2-aminobenzoyl (DLPPEA)
Inhibitors	Apstatin [N-(2S,3R)-3-amino-2 hydroxy-4 phenylbutanoyl) L-prolyl-L-prolyl-L-alaninamide] Ki 3 µM; linear mixed-type Carbobenzoxy-Phe Ki

	100 μM; non-competitive Carbobenzoxy-Pro-Pro-Ala Ki 200 μM; competitive Metal ion chelators e.g. EDTA and phenanthrolin.
Biological Functions	The exact biological significance of APP is at present unknown. The specificity of the enzyme suggests a participation in the final catabolism of collagen and in the resorption processes of collagen breakdown fragments. Participation of APP in bradykinin, neuropeptide Y and peptide YY metabolism results either in inactivation, or in regulation of relative concentrations of different receptor-selective forms. APP on vascular endothelium, smooth muscle cells and in platelets may especially be important in bradykinin-regulated cardiovascular and pulmonary functions. The nature of the substrates of APP (interleukins, hormones, neuropeptides, tachykinins, kinins), and the activation of APP upon lymphocyte stimulation, suggest an involvement in immunological processes as inflammation and wound repair.
Physiology/Pathology	Abscence of APP in small intestine may lead to excretion of massive quantities of imino-oligopeptides, especially Gly-Pro-Hyp-Gly in the urine.
Degradation	Unknown
Genetics/Abnormalities	Aminopeptidase P deficiency is described in two brothers, both excreting large quantities of imino-oligopeptides in the urine.
Half-life	Unknown
Concentration	Unknown
Isolation Method	Human kidney, placenta, platelets and leukocytes are rich sources of APP. Membrane bound APP can be released from the membrane using phosphatidylinositol specific phospholipase C or plasma phospholipase D. The solubilized enzyme can be purified by immunoaffinity purification using anti-APP antibodies. The cytosolic APP can be purified using a multistep chromatography procedure including anion exchange, hydroxylapatite chromatography and gel filtration.
Amino Acid Sequence	Unknown. Only the prokaryotic, yeast, and pig APP aa sequence is known. The C-terminal part of bacterial APP shares the same pita bread fold common to creatinase (EC 3.5.3.3) and methionine aminopeptidase (EC 3.4.11.18).
Disulfides/SH-Groups	Unknown
General References	Blau, N. et al. *J. Inher. Metab. Dis.* 1988, **11**:240-242. Hooper, N.M. and Turner, A.J. *FEBS Letters* 1988, **229**:340-344. Prechel, M.M. et al. *J. Pharm. Exp. Ther.* 1995, **275**:1136-1142. Vanhoof, G. et al. *Biochem. Pharmacol.* 1992, **44**: 479-487. Vanhoof, G. et al. *FASEB J* 1995,**9**: 736-744. Yaron, A. and Naider, F. *Crit. Rev. Biochem. Molec. Biol.* 1993, **28**:31-81.
Ref. for DNA/AA Sequences	Pig APP amino acid sequence: Romero V.C. et al. *Eur. J. Biochem.* 1995, **229**:262-269; accession number gi 994862. Pig cDNA sequence: Hyde, R.J. et al. *Biochem. J.* 1996, **319**:197-201. EMBL accession number U55039 Human lymphocyte XPNPEPL sequence: Vanhoof, G. et al. *Cytogenet. Cell Genet.* 1997, in press. Embl accession number X95762.

Amyloid β-Protein Precursor

Wilma Wasco

Synonyms	Amyloid precursor protein; Amyloid protein precursor; β-amyloid precursor protein; β-amyloid protein precursor. Secreted forms of the molecule are identical to protease nexin II.
Abbreviations	APP; βPP
Classification	Electrophoretic mobility is 100-120 kDa. APP runs anonymously on SDS/PAGE due to high electronegative aa content.
Description	APP is a type 1 integral membrane protein that is present in all tissues and cell types examined to date. It is a member of a conserved gene family that includes two mammalian amyloid precursor-like proteins, APLP1 and APLP 2 as well as APP-like proteins in Drosophila and C. Elegans. Three of the 18 exons that encode APP can be alternatively spliced to give rise to a number of transcripts. The alternatively spliced exons encode 1) a 56 aa Kunitz protease inhibitor (KPI) domain, 2) a 19 aa exon of unknown function that has homology to the MRC Ox-2 antigen, and 3) a 15 aa exon whose presence prevents the addition of chondroitin sulfate glycosaminoglycan moieties. The extracellular domain undergoes O- and N-linked glycosylation and tyrosine sulfation as well as phosphorylation. This domain of the molecule also contains zinc- and heparin-binding sites. Cleavage of APP near the membrane-extracellular junction results in the secretion of this extracellular domain. Processing also gives rise to a 39-43 aa peptide termed the amyloid β peptide, or Aβ, which is the primary component of the senile plaques that are characteristically present in the brains of patients with Alzheimer's disease.
Structure	Unknown
Molecular Weight	86,942 (calculated from the aa composition of the form of the protein that contains all three of the alternatively spliced exons).
Sedimentation Coeff.	Unknown
Isoelectric Point	4.5 (calculated); 4.6 (experimentally determined).
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Forms of APP that contain the Kunitz protease inhibitor domain (KPI) are presumed to function as secreted protease inhibitors. These forms of the protein, which have been shown to have the ability to inhibit trypsin, chymotrypsin and other serine proteases in vitro, are identical to protease nexin II, a molecule that was previously identified based on its tight association with proteases, and to factor XIa inhibitor, which is released from the alpha-granules of platelets and inhibits activated factor XIa at the

	late stages of the cascade. The function of forms of APP that lack a KPI domain remains unclear. As a cell surface molecule, it may be involved in cell-cell interactions. Localization to the synapse suggest a fundamental role in the maintenance, structural integrity or function of synapses. Secreted forms of the molecule have been reported to have growth promoting activity. The potential interaction of APP with the extracellular matrix suggests that it may be influential in the guidance of neurites in the developing nervous system and during the regeneration of neurites following injury.
Physiology/Pathology	Mutations in APP have been linked to specific forms of early-onset familial Alzheimer's disease (FAD). All of these mutations result in aa substitutions within or adjacent to the Aβ domain, and the majority of these mutations have been directly demonstrated to cause either quantitative or qualitative changes in the secretion of Aβ. Although the precise mechanism by which these Aβ. alterations exert their effect remains unclear, a variety of data indicate that aggregated Aβ may be neurotoxic. Overproduction of APP has been linked to Alzheimer's disease in Down's syndrome patients, who due to the trisomy of chromosome 21, have 1.5 fold more APP than normal individuals.
Degradation	Binding of the secreted, KPI-containing forms of APP to the LDL-receptor related protein (LRP) mediate internalization and degradation. The C-terminal domain of APP contains an NPXY sequence that targets C-terminal membrane bound fragments and some percentage of full length APP for re-internalization via clathrin coated pits and subsequent degradation in the endosomal-lysosomal pathway. Normal processing of APP leads to the generation of limited amounts of the 39-43 aa Aβ peptide; Alzheimer's disease-associated alterations in processing lead to an increase in the production of Aβ.
Genetics/Abnormalities	There is a single copy of the APP gene located on chromosome 21(1q21). Six AD-associated APP point mutations have been identified. In addition, there are a number of non-pathogenic polymorphism present in the gene. Overproduction of APP due to trisomy of chromosome 21 (Down's Syndrome) results in the pathological and behavioral hallmarks of Alzheimer's disease. Mice deficient for the amyloid precursor protein gene do not show overt abnormalities.
Half-life	~ 1.5 hours in cultured cells.
Concentration	The APP content of rat brain has been reported to be ~ 46 micrograms/gram tissue. In human brain the concentration has been estimated to be ~ 190 micrograms/gram tissue for membrane-bound APP and ~ 140 micrograms/gram tissue for secreted APP. The concentration in cerebral spinal fluid and plasma is ~ 100 nM.
Isolation Method	The best methods are based on ion exchange, heparin sepharose and either dye ligand or phenyl sepharose chromatography. The best source for isolation of membrane associated APP is mammalian brain (yield 0.7 micrograms/gram of tissue, assuming a ~ 15% recovery). The best source for isolation of secreted APP is the media of stably transfected cells (~ 20 micrograms/liter).
Amino Acid Sequence	APP is 68% identical to amyloid precursor like protein (APLP2), and 56% identical to amyloid precursor-like protein 1 (APLP1). It is 43% identical to a Drosophila homologue (APPL) and 46% identical to a C. elegans homologue (APL-1).

Disulfides/SH Groups	There are six cysteines and three disulfide bridges within the Kunitz protease inhibitor domain.
General References	Selkoe, D.J. *Ann. Rev. Neurosci.* **17**:489-517, 1994. Masters, C.L. and Beyreuther, K. In: *Amyloid precursor protein in development, aging and Alzheimer's disease*. Masters, C.L. et al. (eds.), Springer-Verlag Press, 1994, pp. 1-9. Howlett, D.R. et al. In: *Neurobiology of Alzheimer's disease*, (Dawbarn, D. and Allen, S.J. (eds.), *BIOS Scientific*, 1995, pp. 9-50. Selkoe, D.J. *J. Biol. Chem.* **271**:18295-18298, 1996. Wasco, W. and Tanzi, R.E. In: *Molecular Mechanisms of Dementia.* Wasco, W. and Tanzi, R.E. (eds.), Humana Press, 1997, pp. 1-20.
Ref. for DNA/AA Sequences	Daigle, I. and Li, C. *Proc. Natl. Acad. Sci. USA*. **90**:12045-12049, 1993. Goldgaber, D. et al. *Science* **235**:877-880, 1987. Kang, J. et al. *Nature* **325**:733-736, 1987. Luo, L.Q. et al. *J. Neurosci.* **10**:3849-3861, 1990. Tanzi, R.E. et al. *Science* **235**:880-884, 1987. Wasco, W. et al. *Proc. Natl. Acad. Sci .USA*. **89**:10758-10762, 1992. Wasco, W. et al. *Nat. Genet.* **5**:95-100.1993.

Angiogenin

Bert L. Vallee

Synonyms	None
Abbreviations	Ang
Classifications	Member of ribonuclease superfamily of proteins
Description	A plasma protein originally isolated from the conditioned medium of a human colon carcinoma cell line based on its capacity to induce neovascularization on the chicken embryo chorioallantoic membrane and in the rabbit cornea. It was subsequently shown to be a ribonuclease whose enzymatic activity is necessary for its angiogenic activity. A monomeric, single-chain molecule of 123 aa with no prosthetic groups or sugars attached.
Structure	The x-ray structure, determined at 2.4Å, is remarkably similar to that of pancreatic ribonuclease. It differs in the active center, particularly in the pyrimidine binding site, which is blocked by the side-chain of glutamine-117 and thus explains the low in vitro ribonucleolytic activity of Ang. It also differs in the putative receptor binding site, which like the catalytic site is critically involved in biological function.
Molecular Weight	14,124
Sedimentation Coeff.	Unknown
Isoelectric Point	> 9.5
Extinction Coeff.	8.85 (280nm, 1%, 1cm); molar absorptivity 12,500 $M^{-1} cm^{-1}$
Enzyme Activity	Endonucleolytic cleavage of rRNA but at 10^{-5}x the rate of pancreatic ribonucleases and generating products containing from 100 to 500 nucleotides. Also active against dinucleotides (CpA > CpG > UpA > UpG). Potent inhibitor of cell-free protein synthesis owing to specific cleavage of ribosomal RNA. Angiogenic activity may involve cleavage of RNA-like molecule.
Coenzymes/Cofactors	None
Substrates	Physiological substrate is unknown but mutagenesis studies suggest that it is a RNA-like molecule.
Inhibitors	Placental ribonuclease inhibitor (present in many tissues) a 50 kDa leucine-rich protein composed of 460 aa in seven tandem internal repeat units of 57 aa. It inhibits both the ribonucleolytic and angiogenic activities of angiogenin ($K_i = 7.1 \times 10^{-16}$ M).
Biological Functions	The function of vascular Ang is unknown. It is a potent inducer of blood vessel formation in model systems, appears to be required for xenograft tumor growth in nude mice, and induces second messenger responses (diacylglycerol, prostacyclin) when added to endothelial cells in culture. It also inhibits the degranulation of polymorphonuclear leukocytes and acts as an immune suppressor by preventing the proliferation of lymphocytes stimulated by phytohemagglutinin or concanavalin A.

Physiology/Pathology	*In vitro* studies have shown that it binds to an endothelial cell-surface actin-like angiogenin binding protein. It is also endocytosed by endothelial cells and is translocated to the nucleus where it accumulates in the nucleolus. Angiogenic activity requires an intact cell binding site; a separate, catalytically competent active site and a nuclear localization sequence. It has been shown to act as an adhesion molecule for endothelial cells when coated on plastic culture dishes. Its complex with actin promotes the activation of plasminogen by tPA and stimulates endothelial cell invasiveness. A monoclonal antibody that neutralizes the in vivo and in vitro activities of Ang prevents or delays the appearance of human tumors in athymic mice in a statistically significant, dose-dependent manner. A sensitive ELISA method has revealed marked elevations of plasma Ang in a wide range of diesases.
Degradation	Unknown
Genetics/Abnormalities	The gene is present as a single copy on chromosome 14 at q11, proximal to the T cell receptor α/δ locus. It has no introns in its coding region. An AvaII RFLP is present in an unselected collection of human placental DNAs at a frequency of 29%. It results from a single, silent transversion in the coding region of the gene.
Half-life	Unknown
Concentration	388 ± 115 μg/L in plasma.
Isolation Method	Human plasma (3-10L) adjusted to pH 6.6 with HCl is chromatographed on CM-52 cellulose. After washing with 100 mM P_i, pH 6.6, Ang is eluted with 1 M NaCl and further purified by Mono S or C18 HPLC.
Amino Acid Sequence	
Disulfides/SH-Groups	Disulfides link C26-C81, C39-C92, and C57-C107; no SH-groups.
General References	Fett, J.W. et al. *Biochemistry* 1985, **24**:5480-5486. Shapiro, R. et al. *Biochemistry* 1986, **25**:3527-3532. Palmer, K.A. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:1965-1969. Hallahan, T.W. et al. *Proc. Natl. Acad. Sci. USA* 1991, **88**:2222-2226.
Ref. for DNA/AA Sequences	Strydom, D.J. et al. *Biochemistry* 1985, **24**:5486-5494. Kurachi, K. et al. *Biochemistry* 1985, **24**:5494-5499. Bond, M.D. et al. *Biochim. Biophys Acta* 1993, **1162**:177-186.

Angiotensin converting enzyme

James F. Riordan

Synonyms	Kininase II; peptidyl-dipeptidase A; dipeptidyl carboxypeptidase; peptidyl-dipeptide hydrolase
Abbreviations	ACE
Classifications	EC 3.4.15.1
Description	Somatic ACE is a class I ectoenzyme associated with the luminal membrane of endothelial cells and most abundant in the pulmonary vascular bed. Also present in kidney, intestine, choroid plexus, prostate and epididymis bound to epithelial cell membranes. A monomeric glycoprotein with 26% carbohydrate. A form of the protein circulates in plasma and is derived from endothelial ACE by proteolysis. Another form is present in mature testis.
Structure	Not yet crystallized.
Molecular Weight	Human endothelial cell (somatic) ACE contains 1277 aa with an inferred polypeptide mw 147,000. Human testis ACE contains 701 residues with mw 80,000.
Sedimentation Coeff.	Unknown
Isoelectric Point	4.3-5.2
Extinction Coeff.	290,000 M^{-1} cm^{-1} (280nm) for somatic ACE
Enzyme Activity	A peptidyldipeptide hydrolase that principally catalyzes the hydrolytic cleavage of dipeptides from the C-terminus of unblocked oligopeptides. An endopeptidase activity has been noted with some C-terminal amidated oligopeptides.
Coenzymes/Cofactors	ACE is a chloride-dependent, zinc metalloenzyme.
Substrates	Angiotensin I, bradykinin and other oligopeptides such as hippuryl-His-Leu and furanacryloyl-Phe-Gly-Gly.
Inhibitors	Captopril; lisinopril; enalaprilat; ramipril; EDTA; 1,10-phenanthroline.
Biological Functions	Vascular ACE participates in blood pressure regulation by converting the decapeptide angiotensin I to the potent vasopressive octapeptide angiotensin II. Inactivates the vasodilator bradykinin. Tissue ACE function unknown.
Physiology/Pathology	Decreased renal perfusion pressure leads to the release of renin, an aspartic protease that acts on the amino terminus of plasma angiotensinogen (mw 50,000) to generate angiotensin I which is converted to angiotensin II by ACE. Angiotensin II is a vasoconstrictor and also stimulates adrenal glomerulosa cells to secrete aldosterone which increases renal sodium retention and blood volume. The resulting increased renal perfusion pressure decreases renin release. ACE inhibitors are anti-hypertensive agents. Plasma ACE activity is elevated in sarcoidosis and occasionally in

	leprosy, primary biliary cirrhosis, beryllium disease, silicosis, diabetes mellitus and hyperthyroidism.
Degradation	Unknown
Genetics/Abnormalities	The mRNAs encoding the two ACE isoforms are transcribed from a single gene by a tissue-specific, hormone-induced choice of alternative transcription initiation sites. A polymorphism consisting of the presence or absence of a 250-bp DNA fragment within the human ACE gene has been detected.
Half-life	Unknown
Concentration	Serum ACE measured by direct RIA are 408.5 ± 98.2 µg/L for males and 417.1 ± 103.5 µg/L for females.
Isolation Method	From human kidney homogenate by detergent extraction and affinity chromatography on lisinopril coupled to Sepharose via a 28 Å spacer.
Amino Acid Sequence	Somatic ACE contains 1277 aa in a tandem repeat with two HEMGH sequences indicative of zinc binding sites. Testicular ACE contains 701 aa; residues 36-701 are identical to the C-terminal half of somatic ACE.
Disulfides/SH-Groups	Somatic ACE has 14 Cys and at least 6 disulfides; testicular ACE has 7 Cys and 3 disulfides.
General References	Ehlers, M.R.W. and Riordan, J.F.: Angiotensin convering enzyme: New concepts concerning its biological role. *Biochemistry* 1989, **28:**5311-5317. Ehlers, M.R.W. and Riordan, J.F.:"Angiotensin converting enzyme biochemistry and molecular biology". In: *Hypertension: Pathology, Diagnosis, and Management*, Laragh, J.H., and Brenner, B.M. (eds.). Raven Press 1990; pp. 1217-1231. Erdos, E.G.: Angiotensin I converting enzyme and the changes in our concepts through the years. *Hypertension* 1990, **16**: 363-370. Rigat, B. et al. *J. Clin. Invest.* 1990, **86**:1343-1346.
Ref. for DNA/AA Sequences	Soubrier, F. et al. *Proc. Natl. Acad. Sci. USA* 1988, **85**:9386-9390. Ehlers, M.R.W. et al. *Proc. Natl. Acad. Sci. USA* 1989, **86**:7741-7745. Kumar, R.S. et al. *J. Biol. Chem.* 1989, **264**:16754-16758. Bernstein, K.E. et al. *J. Biol. Chem.* 1989, **264**:11945-11951. Lattion, A.L. et al. *FEBS Lett.* 1989, **252**:99-104. Kumar, R.S. et al. *J. Biol. Chem.* 1991, **266**:3854-3862.

Ankyrin

William T. Tse and Samuel E. Lux

Synonyms	Erythrocyte ankyrin; Ankyrin 1; Ankyrin$_R$; Protein 2.1; Syndein; Goblin
Abbreviations	Ank1
Classifications	Cellular structural protein
Description	Ankyrin 1, a 1,881 aa membrane skeleton protein originally identified in red blood cells, contains 3 principal domains: an N-terminal 89 kDa membrane-binding domain, a 62 kDa spectrin-binding domain, and a 55 kDa C-terminal putative regulatory domain. The 89 kDa domain is almost entirely composed of 24 tandem 33-aa repeats, organized in groups of six, with the signature consensus sequence -G-TPLH-AA--GH---V/A--LL--GA--N/D----. A similar repeat motif has also been found in over 200 proteins of diverse function, including proteins involved in transcriptional regulation, cell-cycle control, and cell differentiation. The regulatory domain of ankyrin contains a death domain motif, commonly found in proteins involved in programmed cell death; however, its function in ankyrin is unknown. There are two related proteins, ankyrin 2 and ankyrin 3, which have a similar structure but are encoded by distinct genes. The tissue distribution patterns are different for the three ankyrins. Ankyrin 1 is found mainly in erythroid tissues, cerebellum and skeletal muscle. Ankyrin 2 is found predominantly in the brain and ankyrin 3, in epithelial tissues, muscles, and neuronal axons.
Structure	The tertiary structure is not yet determined for the native protein. The crystal structure of the p53 binding protein 53BP2, which contains four ankyrin repeats, has been determined. In this protein, each ankyrin motif forms two antiparallel alpha-helices and one hairpin loop, the latter joining with those of the adjacent ankyrin motifs to form a beta-sheet.
Molecular Weight	206,275 (calculated); 210,000 (SDS PAGE)
Sedimentation Coeff.	6.9S
Isoelectric Point	pI 5.95 (calculated)
Extinction Coeff.	5.63 (280nm, 1%, 1cm, calculated)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	By anchoring the erythrocyte membrane to the underlying membrane skeleton, ankyrin 1 helps maintain the structural integrity of the red blood cell.
Physiology/Pathology	Ankyrin 1 functions as a linker protein attaching the spectrin-based membrane skeleton to band 3, the anion-exchanger protein (AE1), in the erythrocyte plasma membrane. Phosphorylation down-regulates its

interactions with spectrin and band 3. Protein 4.2, another peripheral membrane protein, probably also binds to ankyrin. Ligands of ankyrin in non-erythroid tissues include the Na^+/K^+-ATPase, the voltage-dependent and amiloride-sensitive sodium channels, the Na^+/Ca^{2+} exchanger, the L1 family of neural cell adhesion molecules, the IP_3 receptor, and the hyaluronic acid receptor, CD44. Ankyrin appears to play a role in maintaining the polarized distribution of these proteins in these tissues.

Degradation

Unknown

Genetics/Abnormalities

The ankyrin 1 gene is on chromosome 8 at q11.2, and is 160 kb long with 42 exons. Multiple different transcripts of the gene exist, alternatively spliced near the 3' end. Ankyrin 1 mutations are a major cause of dominant and recessive hereditary spherocytosis (35-65%). Frameshift and nonsense null mutations are common in dominant hereditary spherocytosis, whereas missense mutations and a mutation in the promoter prevail in the recessive form. Most are unique mutations. The ankyrin 2 and ankyrin 3 genes are on chromosome 4 at q25-q27 and chromosome 10 at q21, respectively. No known human disorders are associated with mutations in ankyrin 2 or ankyrin 3.

Half-life

Unknown

Concentration

124,500 ± 11,000 copies per red blood cell

Isolation Method

(a) Extraction from spectrin-depleted erythrocyte ghost membranes or Triton X-100-insoluble membrane skeletons by 1M KCl. (b) Chromatography on DEAE-cellulose column. (c) Preparative rate zonal sedimentation on sucrose gradients.

Amino Acid Sequence

Ankyrin repeats (residue 9):

```
DAATSFLRAA RSGNLDKALD HLRNGVDINT CNQ
NGLNGLHLAS KEGHVKMVVE LLHKEIILET TTK
KGNTALHIAA LAGQDEVVRE LVNYGANVNA QSQ
KGFTPLYMAA QENHLEVVKF LLENGANQNV ATE
DGFTPLAVAL QQGHENVVAH LINYG....T KGK
VRLPALHIAA RNDDTRTAAV LLQNDPNPDV LSK
TGFTPLHIAA HYENLNVAQL LLNRGASVNF TPQ
NGITPLHIAS RRGNVIMVRL LLDRGAQIET KTK
DELTPLHCAA RNGHVRISEI LLDHGAPIQA KTK
NGLSPIHMAA QGDHLDCVRL LLQYDAEIDD ITL
DHLTPLHVAA HCGHHRVAKV LLDKGAKPNS RAL
NGFTPLHIAC KKNHVRVMEL LLKTGASIDA VTE
SGLTPLHVAS FMGHLPIVKN LLQRGASPNV SNV
KVETPLHMAA RAGHTEVAKY LLQNKAKVNA KAK
DDQTPLHCAA RIGHTNMVKL LLENNANPNL ATT
AGHTPLHIAA REGHVETVLA LLEKEASQAC MTK
KGFTPLHVAA KYGKVRVAEL LLERDAHPNA AGK
NGLTPLHVAV HHNNLDIVKL LLPRGGSPHS PAW
NGYTPLHIAA KQNQVEVARS LLQYGGSANA ESV
QGVTPLHLAA QEGHAEMVAL LLSKQANGNL GNK
SGLTPLHLVA QEGHVPVADV LIKHGVMVDA TTR
MGYTPLHVAS HYGNIKLVKF LLQHQADVNA KTK
LGYSPLHQAA QQGHTDIVTL LLKNGASPNE VSS
DGTTPLAIAK RLGYISVTDV LKVVTDETSF VLV
```

Consensus:

```
-G-TPLH-AA --GH---V-- LL--GA--N- ---
                   A           D
```

Death domain (residue 1394):

PGSLSGTEQA EMKMAVISEH LGLSWAELAR ELQFSVEDIN
RIRVENPNSL LEQSVALLNL WVIREGQNAN MENLYTALQS
IDRGEIVNML EGSGRQSRNL KPDR

Disulfides/SH-Groups

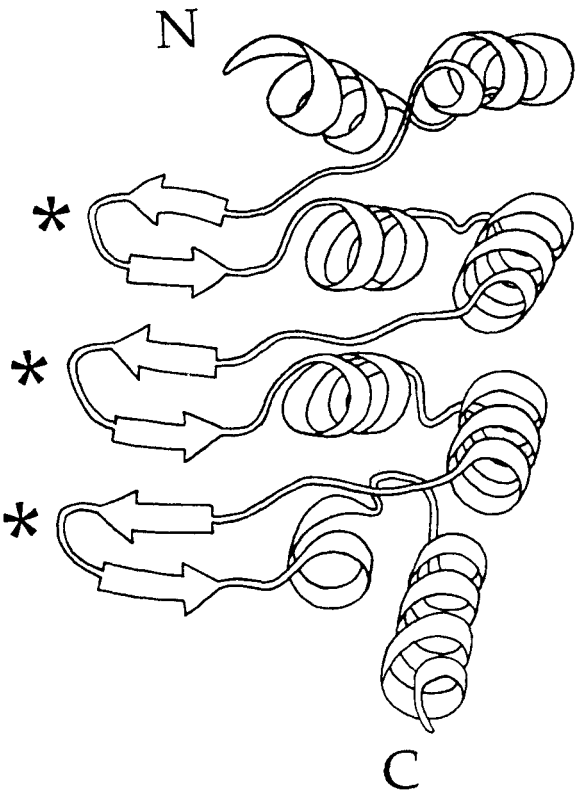
Unknown

General References

Bennet, V. and Stenbuck, P.J. *J. Biol. Chem.* 1980, **255**:2540-2548.
Eber, S.W. et al. *Nature Genet.* 1996, **13**:214-218.
Gorina, S. and Pavletich, N.P. *Science* 1996, **274**:1001-1005.
Peters, L.L. and Lux, S.E. *Seminars Haematol.* 1993, **30**:85-118.

Ref. for DNA/AA Sequences

Lux, S.E. et al. *Nature* 1990, **344**:36-43.
Lambert, S. et al. *Proc. Natl. Acad. Sci. U.S.A.* 1990, **87**:1730-1734.



Crystal structure of the four ankyrin repeats in 53BP2: Each ankyrin repeat forms an L-shaped structure that consists of a beta hairpin and two alpha helices. Neighboring repeats pack via their hairpins to form an antiparallel beta sheet, and via their helix pairs to form helix bundles. The asterisks indicate the junction between adjacent repeats.

Antithrombin

Jui-Yoa Chang

Synonyms	Heparin cofactor
Abbreviations	AT-III
Classifications	Serine proteinase inhibitor (serpins)
Description	A plasma glycoprotein synthesized in liver. The most important inhibitor that regulates blood coagulation by forming an irreversible complex with thrombin. It is a single chain polypeptide with three disulfide linkages and four glycosylation sites. AT-III is an unique serine proteinase inhibitor because it contains in addition to the reactive site, a heparin binding site that modulates the heparin enhanced AT-III/thrombin interaction. AT-III alone inactivates thrombin (progressive inhibitory activity). In the presence of heparin, the inhibitory activity is enhanced by about 1000 fold (heparin cofactor activity).
Structure	Slightly elongated than typical globular proteins. The 3-D structure of AT-III has been determined at 3.2 Å resolution. The overall topology of AT-III resembles that of alpha-l-antitrypsin with helices mainly located in the N-terminal domain. Intactness of the three disulfide bond is essential for the activity. The four carbohydrate chains have minimal importance for the functional properties of AT-III. The reactive site is Arg-393 (P1 residue). The heparin binding site of AT-III is comprised of clusters of basic amino acids, including Arg-47, Lys-125, Lys-136, Arg-129 and Arg-145.
Molecular Weight	58,200 (calculated from aa sequence and the carbohydrate structure data).
Sedimentation Coeff.	Unknown
Isoelectric Point	4.9–5.3
Extinction Coeff.	6.0 (280 nm, 1%, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	Heparin
Substrates	None
Inhibitors	None
Biological Functions	The major plasma inhibitor that regulates blood coagulation. It inhibits thrombin, factors IXa, Xa and XIa. The inhibitory activity is greatly accelerated in the presence of heparin.
Physiology/Pathology	Reduced AT-III concentration or functional deficiency (congenital variants) are associated with deep-vein thrombosis and plumonary embolism.
Degradation	Degraded in human body with a half-life of about 3 days. If human subjects are treated with heparin, the degradation rate is faster with a half-life of about 2.1 days.
Genetics/Abnormalities	AT-III gene contains six exons and five introns. They distribute over 19 kb of DNA. Congenital variants with either defective reactive site or impaired heparin binding site were reported.

Half-life	About 3 days
Concentration	115–160 mg/L (2–2.7 μM)
Isolation Method	Isolated from citrated plasma first by dextran sulfate precipitation. AT-III which is contained in the supernatant is then isolated by affinity chromatography using a heparin Sepharoase CL-6B column.
Amino Acid Sequence	A single polypeptide chain of 432 amino acids. 85% sequence homology with bovine AT-III. It is also homologous with other serine proteinase inhibitors (alpha-1 proteinase inhibitor, alpha-1 antichymotrypsin and heparin cofactor II et al.).
Disulfides/SH-Groups	3 disulfide bonds: Cys-8/Cys-128, Cys-21/Cys-95 and Cys-247/Cys-430). No free sulfhydryl group.
General References	Rosenberg, R. D. and Damus, P. S. *J. Biol. Chem.* 1973, **248:** 6490–6505. Bjork, I. et al. In: *Heparin*, Lane, D. A. and Lindahl, U. (eds.) 1989, pp 229–255, printed by Edward Arnold.
Ref. for DNA/AA Sequences	Chandra, T. et al. *Proc. Natl. Acad. Sci. USA* 1983, **80:** 1845–1848.

Apolipoprotein(a)

Gunther M. Fless and Angelo M. Scanu

Synonyms	None
Abbreviations	Apo(a)
Classifications	Kringle-containing proteins
Description	Apolipoprotein(a), apo(a), is a large glycoprotein (23% carbohydrate by weight) containing from 12 to over 51 kringles (K) that are homologous to KIV of plasminogen. In addition, apo(a) has a single KV and one protease region that are highly homologous to their plasminogen counterpart. Apo(a) is synthesized in the liver; up to 98% circulates in plasma linked to apoB100 as the protein moiety of the cholesteryl ester-rich lipoprotein Lp(a). Minor quantities of apo(a) are also found associated with hepatic-derived triglyceride-rich particles. The linkage between apo(a) and apoB100 probably occurs at the hepatocyte surface via a two-step process, the first involving weak non-covalent interactions and the second the formation of a single disulfide bond.
Structure	Currently there is no information on the hydrodynamic properties of free native apo(a). The studies on a recombinant 17K apo(a) (r-apo(a)), expressed and secreted in human embryonic kidney cells (293s cells), indicated a large Stokes radius of 94 Å due to a highly asymmetric and flexible structure. These data are consistent with the circular dichroic spectrum of minimally reduced apo(a) that gave 0-2% α-helix, 66% β structure, and 32% random coil. In physiological buffers, apo(a) assumes a compact structure when bound to apoB100 on the Lp(a) surface, probably due to its interaction with lysine, proline or perhaps other aa residues of apoB100. In the presence of 6-amino hexanoic acid, apo(a) undergoes a large conformational change possibly due to disruption of weak lysine binding sites located in KIV 5-8, and assumes a highly asymmetric structure similar to free apo(a). The KIV domains consist of 78 aa that are constrained into a triple-loop structure by three highly conserved disulfide bonds. Apo(a) kringles are connected by aa linker regions variable in length and sequence each containing up to six potential O-linked glycosylation sites. Moreover, potential N-linked glycosylation sites are present in KIV-1 and in the linker between KIV-7 to KIV-8. Neither KV nor the protease domain contain glycosylation sites. The KIV domains can be organized into ten different types. KIV-9 has an extra cysteine (Cys-4057) that is involved in disulfide linkage with Cys-4326 of apoB100. The protease domain has 88% aa identity with that of plasminogen, and has the same catalytic triad, Ser-His-Asp. However, contrary to plasminogen, apo(a) has no plasmin-like activity because Ser has replaced Arg at the equivalent acitvation site.
Molecular Weight	Apo(a) mass is directly proportional to the number of KIV domains. Molecular weights range from 243,000 to 876,000 for apo(a)s having 12 to 51 KIV domains and 23% by weight carbohydrate.
Sedimentation Coeff.	$s^{\circ}_{20,w}$ of a 17 KIV r-apo(a) is 9.30 S; $D_{20,w}$ is 2.29 ficks. Both parameters will vary with the number of KIVs.
Isoelectric Point	The isoelectric point of r-apo(a) is 4.3

Extinction Coeff.	19.4 for r-apo(a) as calculated from its sequence (280nm, 1%, 1cm)
Enzyme Activity	Weak serine protease activity with a specificity for arginine.
Coenzymes/Cofactors	Unknown
Substrates	Fibronectin and LDL may be biological substrates for the proteolytic activity of Lp(a). However, these isolated observations need to be corroborated. Useful chromogenic substrates are N-α-Benzoyl-Arg-pNA(L-BAPA) and N-β-Cbo-Arg-Gly-Arg-pNA(S-2765).
Inhibitors	10-20 mM phenylmethyl sulfonyl fluoride inhibits Lp(a)-induced fragmentation of fibronectin.
Biological Functions	These are still poorly defined. Lp(a) may participate in the regulation of fibrinolysis by interacting with fibrin and/or cellular binding sites (endothelial cells, mononuclear cells, and other blood cells) in competition with plasminogen. These interactions are mediated in part by the lysine binding capacity of apo(a). Lp(a) has also an affinity for fibronectin, tetranectin, glycosaminoglycans, proteoglycans and extracellular matrix components. Apo(a) may also direct the cholesterol-rich (Lp(a) particle to sites of tissue damage. There Lp(a) may cause cell proliferation and active membrane synthesis by inhibiting the activation of transforming growth factor-β by plasmin, thereby promoting wound healing. Lp(a) may also be an acute phase reactant.
Physiology/Pathology	Lp(a) by interfering with fibrinolysis as a competitive inhibitor of plasminogen activation, as a regulator of plasmin activity, and as an inhibitor of the interaction between tissue plasminogen activator and plasminogen activator inhibitor, may be thrombogenic. Moreover, Lp(a) because of its cholesterol content may have a pro-atherogenic role. Upon entrapment in the subendothelial intima by fibrin and/or subendothelial matrix components, Lp(a) may undergo oxidative modification leading to particle uptake by macrophages and formation of foam cells. In the subendothelial extracellular matrix of the arterial wall, Lp(a) may also bind to collagen and elastin due to the affinity of apo(a) for proline and hydroxyproline.
Degradation	The primary site of synthesis of apo(a) is the hepatocyte, from which it is secreted either free or linked to apolipoprotein B100. In the plasma, apo(a) is mostly affiliated with cholesteryl ester-rich Lp(a) particles; however, it is also present in triglyceride-rich lipoproteins. The catabolic fate of Lp(a)/apo(a) is still undefined. Because of the presence of apoB100, some studies have suggested that Lp(a), like LDL, may be taken up and degraded via the LDL receptor. However, later studies have provided no support for this proposal and drugs known to affect LDL receptor expression do not affect plasma Lp(a) levels.
Genetics/Abnormalities	Apo(a) is under the control of a single gene located in chromosome 6, band q 26-27, adjacent to the plasminogen gene. The apo(a) gene locus specifies for at least 34 alleles, each coding for apo(a) size isoforms differing in the number of KIV-2 repeats. To date up to 51 potential apo(a) isoforms have been proposed. Human subjects usually have two apo(a) isoforms. In rare instances, no apo(a) may be detected in the plasma either because of a null allele or post-translational modifications preventing the secretion of apo(a) from the hepatic cell. Apo(a) size polymorphism accounts for about 40% of the plasma Lp(a) levels. The remainder may be due structural variations in the promoter region of the apo(a) gene and to a lesser extent, environmental factors. Plasma Lp(a) levels vary among

different populations. Caucasian and Orientals have highly skewed distributions with most individuals exhibiting plasma Lp(a) concentrations lower than Black Americans and Sudanese. One of the properties of apo(a) is to bind to immobilized lysine (lys+). However, there are human mutants in whom apo(a) is lys- due to the fact that in the lysine binding site (LBS) of KIV-10, Trp-72 has been replaced by Arg. In this respect, it is interesting to note that the apo(a) of rhesus monkeys is lys- and like the human mutant, has Arg-72 in the LBS of KIV-10.

Half-life

Fract.catab. rate of Lp(a) in man is 0.26 days

Concentration

Mean plasma Lp(a) concentration in Caucasians is 0.1-0.2 g/L when expressed as lipoprotein, or 0.025-0.050 g/L when expressed as Lp(a) protein. However, concentrations vary more than a thousand-fold from less than 0.001 g/L to over 1 g/L lipoprotein or 0.00025-0.25 g/L Lp(a) protein.

Isolation Method

Lp(a) is commonly isolated from plasma by centrifugation in high salt media followed by lysine-Sepharose chromatography. A free functional apo(a) is obtained by subjecting Lp(a) to mild reductive conditions in the presence of ε-aminocaproic acid, followed by ultracentrifugation and recovery of the free apo(a) from the bottom of the centrifuge tube.

Amino Acid Sequence

Based on the analysis of one apo(a) cDNA the deduced protein sequence indicates that apo(a) is 4529 aa in long. The region between aa 1-4207 contains the 10 classes of 37 kringles homologous to plasminogen KIV. The region between residues 4208 and 4308 represents KV and that between 4309 and 4529 the protease domain. Residue 4057 in KIV-9 is the free Cys involved in the linkage of apo(a) to apoB100. In apo(a) residue 4308 is ser contrary to arg in plasminogen. Because of this critical substitution at the activation site sensitive to the action of either tPA or urokinase, apo(a) fails to generate an active protease. Residues 4350 (His), 4393 (Asp) and 4479 (Ser) represent the potential catalytic triad in plasmin(ogen) and apo(a).

Disulfides/SH-Groups

Each kringle domain contains 3 disulfides bonds; KIV-9 contains an additional free Cys. The protease domain has six disulfides.

General References

Scanu, A. M. and Fless, G. M. *J. Clin. Invest.* 1990, **85**:1709-1715.
Loscalzo, *J. Arteriosclerosis* 1990, **10**:672-679.
Utermann, G. *Science* 1989, **246**:904-910.
Gaubatz, J. W. et al. *J. Biol. Chem.* 1986, **129**:167-186.
Lackner, C. et al. *J. Clin. Invest.* 1991, **87**:2153-2161.
Klezovitch, O. and Scanu, A.M. *Current Opinion Lipidology* 1995, **6**:223-228.
Scanu, A.M. and Edelstein, C. *Biochim. Biophys. Acta* 1995, **1256**:1-12.
Edelstein, C. et al. *Biochemistry* 1995, **34**:16483-16492.
Fless, G.M. et al. *Biochemistry* 1996, **35**:2289-2298.

Ref. for DNA/AA Sequences

McLean, J. W. et al. *Nature* 1987, **330**:3224-3228.
Kratzin, H. et al. *Biol. Chem. Hoppe-Seyler* 1987, **368**:1533-1544.

Apolipoprotein A-I

Lawrence Chan

Synonyms	Apolipoprotein-Gln-I
Abbreviations	ApoA-I, apo-LP-A-I, apo-Gln-I
Classifications	Major high density lipoprotein apoprotein.
Description	ApoA-I is the major apolipoprotein in high density lipoproteins. It is synthesized primarily by the liver and small intestine. It exists in four different isoforms, designated A-I_{-2}, A-I_{-1}, A-I_0 and A-I_{+1}. There is another form, A-I_{+2} which corresponds to proapoA-I.
Structure	ApoA-I contains multiple 11- and 22-amino acid residue degenerate repeats. Some of these are alpha-amphipathic sequences.
Molecular Weight	28,000
Sedimentation Coeff.	Unknown
Isoelectric Point	5.62, 5.53, 5.45 & 5.36 for different isoforms
Extinction Coeff.	Unknown
Enzyme Activity	Unknown
Coenzymes/Cofactors	ApoA-I increase the enzymatic activity of lecithin : cholesterol acyltransferase, which catalyzes the esterification of plasma cholesterol.
Substrates	None
Inhibitors	None
Biological Functions	As a major protein in high density lipoproteins, apoA-I may be protective against the development of atherosclerosis. By increasing the activity of lecithin : cholesterol acyltransferase, apoA-I is important in "reverse cholesterol transport."
Physiology/Pathology	Plasma concentrations of apoA-I are negatively correlated with atherosclerosis. Tangier disease is associated with extremly low plasma apoA-I. However, the primary defect resides outside the apoA-I gene in this syndrome.
Degradation	Mainly in the liver. Products are presumably inactive.
Genetics/Abnormalities	The apoA-I gene has been localized to chromosome 11 at 11g13 – 11gter. A large number of genetic variants of apoA-I have been described most of which are not associated with known biological consequences. The apoA-I gene is immediately 5' to the apoC-III and apoA-IV genes. The A-I and A-IV genes are in the same orientation whereas the C-III gene in between them has the opposite orientation. Familial apoA-I/C-III deficiency has been reported due to rearrangement of this region of the gene complex.
Half-life	Plasma residence time 3 – 5 days
Concentration	Serum apoA-I concentration is 1.27 g/L (S. D. 0.19) for men and 1.46 g/L (S. D. 0.27) for women.

Isolation Method	ApoA-I is best isolated from high density lipoproteins. The latter are isolated from plasma at density 1.063–1.21 g/ml by sequential ultracentrifugation. After delipidation apoA-I is isolated by gel permeation chromatography on a Sephacryl S-200 column. Alternative methods of purification include high pressure liquid chromatography, preparative sodium dodecyl sulfate-gel electrophoresis, DEAE-cellulose chromatography or affinity chromatography using monospecific polyclonal or monoclonal apoA-I antibodies.
Amino Acid Sequence	ApoA-I contains 243 amino acids residues. The protein is initially synthesized with an 18 amino acid signal peptide and a 6 amino acid propeptide, with the sequence RHFWQQ.
Disulfides/SH-Groups	None
General References	Cheung, P. & Chan, L. *Nucleic Acids Res.* 1983, **11:** 3703–3715. Cheung, P., Kao, F. T., Law, M. L., Jones, C., Puck, T. T. & Chan, L. *Proc. Natl. Acad. Sci. USA.* 1984, **81:** 508–511. Brewer, H. B., Ronan, R., Meng, M. & Bishop, C. *Methods Enzymol.* 1986, **128:** 223–235. Mills, G. L., Lane, P. A. & Weech, P. K. *A guidebook to lipoprotein technique.* Elsevier 1984 pp 394–403. Li, W-H, Tanimura, M., Luo, C-C., Datta, S., & Chan, L. *J. Lipid Res.* 1988, **29:** 245–271.
Ref. for DNA/AA Sequences	GeneBank accession numbers: J00098, J00099, J00100, J03222, K01518, M10372, X00566, X00567.

Apolipoprotein A-II

Lawrence Chan

Synonyms	Apolipoprotein-Gln II
Abbreviations	ApoA-II, apo-LP-A-II, apo-Gln-II
Classifications	Major high density lipoprotein apoprotein.
Description	ApoA-II exists in human plasma as a dimer of two identical chains of 77 amino acid residues linked by a disulfide bridge at position 6. The protein is synthesized primarily in the liver.
Structure	ApoA-II contains multiple 11- and 22-amino acid degenerate repeats. Some of these are alpha-amphipathic sequences.
Molecular Weight	17,400
Sedimentation Coeff.	Unknown
Isoelectric Point	5.10
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	At high concentration, apoA-II increases hepatic triglyceride lipase activity in vitro. The physiological significance of this observation is unclear.
Substrates	None
Inhibitors	None
Biological Functions	Other than lipid transport, the precise biological function of apoA-II is unknown. Isolated deficiency of apoA-II has been described, recently but is not associated with any lipid or lipoprotein abnormalities. (Deeb, S. S. et al. *Am. J. Hum. Genet.* 1990, **46**: 822–827.)
Physiology/Pathology	ApoA-II plasma concentrations, like apoA-I, have also been reported to have a negative correlation with the prevalence of myocardial infarction and peripheral vascular disease.
Degradation	Little is known concerning the specific pathway of degradation.
Genetics/Abnormalities	The apoA-II gene has been mapped to human chrosome 1. ApoA-II also contains 11 and 22 amino acid internal repeats that have been identified in other soluble apolipoproteins.
Half-life	Plasma residence time 3–6 days.
Concentration	Serum apoA-II concentration is 0.65 g/L (S. D. 0.11) for men and 0.74 g/L (S. D. 0.16) for women.
Isolation Method	ApoA-II is best isolated from high density lipoproteins. The latter are isolated from plasma at density 1.063–1.21 g/ml by sequential ultracentrifugation. After delipidation, ApoA-II is isolated by gel permeation chromatography on Sephacryl S-200. Alternative methods of purification include high-pressure liquid chromatography, preparative sodium dodecyl sulfate-gel electrophoresis, DEAE-cellulose chromatography or affinity chromatography using monospecific polyclonal or monoclonal antibodies.

Amino Acid Sequence	ApoA-II contains 77 amino acid residues. The protein is initially synthesized with an 18 amino acid signal peptide and a 5-amino acid propeptide with the sequence ALVRR.
Disulfides/SH-Groups	There is a single Cys residue at position 6. The protein normally exists as a dimer with a single intermolecular disulfide bridge at this position.
General References	Moore, M. N., et al.. *Biochem. Biophys. Res. Commun.* 1984, **123:** 1–7. Brewer, H. B., et al. *Methods Enzymol.* 1986, **128:** 235–240. Mills, G. L., Lane, P. A. & Weech, P. K. *A guidebook to lipoprotein technique.* Elsevier 1984, p.403. Li, W-H, et al. *J. Lipid Res.* 1988, **29:** 245–271.
Ref. for DNA/AA Sequences	GenBank accession numbers: X00927, K01686, K02216, X00569.

Apolipoprotein A-IV

Richard B. Weinberg

Synonyms	None
Abbreviations	APO A-IV
Classifications	Apolipoprotein; α-2 mobility
Description	A serum glycoprotein synthesized in the enterocytes of the small intestine. Synthesis stimulated by triglyceride absorption and chylomicron assembly. Apo A-IV is initially incorporated into nascent lymph chylomicrons, but dissociates from their surface following their entry into circulation. The affinity of apo A-IV for lipid surfaces is the weakest of all human apolipoproteins; consequently it circulates mainly as a free protein, unassociated with lipoproteins.
Structure	Secondary structure is 54% α-helical, the remainder random coil. α-Helical domains are amphiphatic. α-Helicity increases to 83% when combined with lipid. In aqueous solution adopts a globular conformation with a hydrophobic interior, and is easily denatured by urea or guanidine. Self-associates to form dimers even at low monomer concentrations with a K_a of 2.9×10^5 L/mol. 6% carbohydrate, probably O-linked.
Molecular Weight	46,000 (analytic ultracentrifugation and PAGE).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.05 - 5.20, major band at 5.12 (8M urea)
Extinction Coeff.	16,700 $Mol^{-1}cm^{-1}$ (280nm, 100mM TRIS, 100mM NaCl)
Enzyme Activity	Cofactor for activation of lecithin-cholesterol acyltransferase
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Weak activator of lecithin-cholesterol acyl transferase. Facilitates activity of cholesterol ester transfer protein. Implicated in absorption of dietary lipid, lipolysis of chylomicrons, efflux of lipid from peripheral cells, lymphatic cholesterol transport, intravascular remodeling of high density lipoproteins, and as a post-prandial satiety factor.
Physiology/Pathology	Specific role in human lipid metabolism not established. No known isolated deficiency states. Combined apo A-I/apo C-III/apo A-IV deficiency associated with fat malabsorption. Serum levels decrease rapidly with fasting, and increase with fat feeding. Overexpression of human apo A-IV in transgenic mice causes mild hypertriglyceridemia.
Degradation	Rapid plasma clearance, probably by hepatic uptake. May also be degraded by glomeluar filtration and uptake by renal tubular cells. Interacts with a receptor on endothelial cells, fibroblasts, and hepatocytes.

1. Installment 1992 (revised 1997)

Genetics/Abnormalities	Common variant allele, apo A-IV-2, codes for Gln-360 → His-360, resulting in a basic isoform. Population frequency of apo A-IV-2 allele is 0.07 - 0.09 in Western Europe; rare in other parts of the world. Apo A-IV-1/2 heterozygotes have increased serum HDL levels and lower serum triglyceride levels than normals and demonstrate an attenuated hypercholesterolemic response to dietary cholesterol. Other rare mutants have been described.
Half-life	16 - 20 hrs.(I^{125}-APO A-IV turnover)
Concentration	0.12 - 0.15 g/L in serum or plasma. Levels increase both acutely and chronically by dietary fat. Has been detected in peripheral lymph and bile.
Isolation Method	Adsorption to phospholipid-triglyceride emulsion particles from lipoprotein-depleted serum, followed by delipidation and low pressure chromatography or HPLC (anion exchange, gel filtration, or reverse phase).
Amino Acid Sequence	376 aa. No Cys, a single Trp at residue 12, four Met at residues 11, 144, 225, 302. Fourteen 22 aa repeats. No N-glycosylation sites. C-Terminal: EQQQEQQQEQQQEQVQLAPLES. High homology to apolipoproteins A-I and E.
Disulfides/S-H-Groups	None
General References	Lefevre, M. and Roheim, P.S. *J. Lipid Res.*, 1984, **25:**1603-16102. Weinberg, R.B. and Spector, M.S. *J. Biol. Chem.*, 1985, **260:**4914-4921. Weinberg, R.B. and Spector, M.S. *J. Biol. Chem.*, 1985, **260:**14270-14286. Weinberg, R.B. and Jordan, M.K. *J. Biol. Chem.*, 1990, **265**:8081-8086.
Ref. for DNA/AA Sequences	Elshourbagy, N.A. et al. *J. Biol. Chem.*, 1986, **257**:1998-2002. Karathanasis, S.K. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:8457-8461. Yang, C.Y. et al. *Biochim. Biophys. Acta.*, 1989, **1002**:231-237.

Apolipoprotein B-48

Lawrence Chan

Synonyms	Apolipoprotein Bs
Abbreviations	ApoB-48
Classifications	Major apoprotein of chylomicrons and chylomicron remnants.
Description	ApoB-48 is the major lipid-transport protein in chylomicrons. It is synthesized in the small intestine of all mammals, and is required for dietary fat absorption. In some mammals, notably rats and mice, but not humans, apoB-48 is also synthesized by the liver. In the small intestine, apoB-48 is an integral part of intestinal chylomicrons and their metabolic products, chylomicron remnants, and represents a marker for these lipoprotein particles. In the rat, hepatic apoB-48 is secreted as part of very low density lipoproteins (VLDL). ApoB-48-containing VLDL, in contrast to apoB-100-containing VLDL, are not metabolized to low density lipoprotein (LDL) in this mammalian species.
Structure	ApoB-48 is a monomeric glycoprotein. Determination of tryptic peptides of purified apoB-48 indicates that it is colinear with the N-terminal 48% of apoB-100. There has been no structural analysis of purified apoB-48.
Molecular Weight	264,000 (SDS-PAGE).
Sedimentation Coeff.	ApoB-48 is an integral part of chylomicrons which have a density of 0.93 g/ml. The sedimentation coefficient of purified apoB-48 is unknown.
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	There is no known enzyme activity.
Coenzymes/Cofactors	None
Substrates	Unknown
Inhibitors	None
Biological Functions	ApoB-48 is synthesized exclusively in the small intestine in man and is required for fat absorption. Since the protein corresponds to the N-terminal 48% of apoB-100, it misses the C-terminal half of the latter protein that is required for LDL receptor binding and biogenesis of lipoprotein (a). ApoB-48 contains lipid-associating peptides that allow it to become an integral part of intestinal chylomicrons.
Physiology/Pathology	ApoB-48 is required for absorption of dietary fat. Chylomicron formation is defective in patients with the syndrome of abetalipoproteinemia in which they fail to secrete either apoB-100 or apoB-48. Because of its role in fat absorption, apoB-48 is also important in facilitating the absorption of fat-soluble vitamins.
Degradation	ApoB-48 exists in the circulation tightly associated with chylomicrons. Chylomicrons are rapidly converted to chylomicron remnants by

lipoprotein lipase. Chylomicron remnants are taken up by hepatic chylomicrons remnant receptors which are heterogeneous in nature and include low density lipoprotein (LDL) receptors and LDL receptor-related protein. The apoB-48 taken up with the remnant particles is degraded intracellularly.

Genetics/Abnormalities

ApoB-48 and apoB-100 are encoded by the same gene, which is located on the short arm of chromosome 2 (2p23-p24). ApoB-48 mRNA is identical in structure to apoB-100 mRNA except for a single uridine (U) for cytidine (C) substitution involving nucleotide 6666, the first base of the codon CAA for Gln-2153, changing it to UAA, a stop codon. This C → U change occurs within exon 26, a 7.5 kilobase pair exon and is the result of a posttranscriptional process known as RNA editing. Translation of apoB-48 mRNA produces a truncated apoB protein because of the presence of the in-frame premature stop codon. ApoB mRNA editing is the first example of RNA editing described in mammals.

ApoB-48 secretion is essentially eliminated in patients with abetalipoproteinemia. Abetalipoproteinemia is an autosomal recessive disorder resulting from mutations in the gene encoding the large subunit of microsomal triglyceride transfer protein. Microsomal triglyceride transfer protein is required for the assembly and secretion of apoB-100- and apoB-48-containing lipoproteins. Patients with abetalipoproteinemia fail to produce chylomicrons or very low density lipoproteins or their metabolic products. ApoB-48 production is also affected in some cases of familial hypobetalipoproteinemia. These patients have mutations in the apoB gene causing premature termination of apoB translation. If the mutant stop codon occurs 5' to the normal stop codon position for apoB-48, the synthesis of the latter protein will be affected.

Half-life

Chylomicrons have a half life of 5-10 minutes.

Concentration

Concentration of apoB-48 in plasma is extremely low in the postabsorptive state (<1mg/dl). It goes up with a fatty meal. Therefore, the exact plasma concentration is controversial and has not been established. In isolated triglyceride-rich lipoproteins, apoB-48 can be accurately measured and in postabsorption state, it is ~ 2.5 ± 1.2 mg/L.

Isolation Method

ApoB-48 is associated with triglyceride-rich lipoproteins (TRL). TRL can be prepared from plasma by ultracentrifugal flotation. ApoB-48 will be one of a number of apoproteins present in the TRL. The detailed methods are provided by Bergeron et al., 1996. If large amounts of apoB-48 are needed (e.g. for peptide mapping and sequence analysis), the TRL can be isolated from hypertriglyceridemic individuals or from chylous ascitic fluid as was reported by Chen et al., 1987.

Amino Acid Sequence

ApoB-48 contains 2152 aa residues. It has an N-terminal glutamic acid and a C-terminal isoleucine. ApoB-48 sequence is colinear with the N-terminal 48% of apoB-100.

Disulfides/SH-Groups

The disulfide distribution in apoB-48 has not been determined. If we assume that it is similar to that of apoB-100, the following description will apply. There are 18 cysteine residues in apoB-48. Except for the 4 cysteine residues in the C-terminal half of apoB-48, all the other cysteine residues are projected to be involved in disulfide bridge formation with their immediately adjacent cysteines.

General References

Chen, S.H. et al. *Science* 1987, **238**:363-366.
Powell, L.M. et al. *Cell* 1987, 50:**831**-840.

Kotite, L. et al. *J. Lipid Res.* 1995, **36**:890-900.
Bergeron, N. et al. *Methods Enzymol.* 1996, **263**: 82-94.
Chan, L. and Seeburg, P.H. *Science and Medicine* 1995, **2**:68-77.
Chan, L. *J. Biol. Chem.* 1992, **267**:25621-25624.

Ref. for DNA/AA Sequence

Chen, S.H. et al. *Science* 1987, **238**:363-366.
Powell, L.M. et al. *Cell* 1987, **50**:831-840.

Apolipoprotein B-100

Lawrence Chan

Synonyms	ApoB-100, ApoLDL, Beta-apolipoprotein
Abbreviations	ApoB-100
Classifications	Beta lipoprotein apoprotein
Description	ApoB-100 is a major protein in very low density, intermediate density and low density lipoproteins. It is also an important component in lipoprotein(a). It is a physiological ligand for the LDL (apoB, E) receptor and is synthesized mainly in the liver. A related protein, ApoB-48, is synthesized in the intestine. ApoB-48 contains the N-terminal 2152 residues of apoB-100.
Structure	Native apoB-100 in low density lipoprotein contains approximately 43% α-helix, 21% β-sheet, 20% random structure and 16% β-turns.
Molecular Weight	The native molecular weight is 549,000 (SDS-PAGE). It is a glycoprotein. The deduced molecular weight from its aa sequence is 512,937.
Sedimentation Coeff.	Low density lipoprotein has flotation Sf 0–20
Isoelectric Point	5.5 (for low density lipoprotein)
Extinction Coeff.	Unknown
Enzyme Activity	There is preliminary evidence that apoB-100 has phospholipase A_2 activity.
Coenzymes/Cofactors	None
Substrates	Natural substrate unknown.
Inhibitors	Phospholipase A_2 activity in low density lipoprotein (which contains apoB-100) can be inhibited by p-bromophenacyl bromide.
Biological Functions	Apolipoprotein B-100 is a physiological ligand for the low density lipoprotein (apoB, E) receptor.
Physiology/Pathology	Elevated plasma apoB-100 is associated with premature atherosclerosis. ApoB-100 is a major protein component of low density lipoprotein and lipoprotein(a). Elevated levels of these two lipoproteins are strongly correlated with premature atherosclerosis.
Degradation	ApoB-100 is degraded mainly in the liver, but also in peripheral tissues. Low density lipoprotein and intermediate density lipoprotein apoB is taken up by the cell via high affinity cell-surface receptor and non-high affinity receptor pathways. Approximately 70% of the total body uptake of low density lipoprotein takes place in the liver by the high affinity receptor pathway.
Genetics/Abnormalities	The apoB gene is localized to the short arm of chromosome 2(2p23–p24). It is a highly polymorphic protein. The only apoB-100 mutation that is associated with elevated serum cholesterol results in an Arg-3500→Gln substitution. Familial hypobetalipoproteinemia, a rare syndrome, has

	been reported to be associated with truncated apoB-100 resulting from mutations that cause premature termination. Such mutations differ in different families and occur throughout the length of apoB-100 (Young, S. G. et al. *J. Clin. Invest.* 1990, **85:** 933–942)
Half-life	About 36 hrs (for low density lipoprotein)
Concentration	Normal plasma apoB concentration is approximately 800 mg/L. It consists of almost exclusively apoB-100.
Isolation Method	The best way is to isolate low density lipoprotein by ultracentrifugal flotation of human plasma in the presence of appropriate protease inhibitors.
Amino Acid Sequence	ApoB-100 is one of the largest proteins known, containing 4536 amino acid residues. Two regions, designated domain A and B, have been postulated to be important for receptor binding. Domain A (residues 3147–3157) has the sequence KAQYKKNKHRH, Domain B (residues 3359–3367) has the sequence RLTRKRGLK. Sequences within and flanking domain B but not domain A have significant similarity to apo E receptor binding domain.
Disulfides/SH-Groups	There are 25 cysteine residues in apoB-100 with a concentration in the N-terminal region. Most of these are in disulfide linkage with immediately neighboring cysteine residues. At least 2 are free, residues 3734 and 4190.
General References	Yang, C. Y., et al. *Nature* 1986, **323:** 738–742. Yang, C. Y., et al. *Arteriosclerosis* 1989, **9:** 96–108. Chen, S. H., et al. *Science* 1987, **238:** 363–366. Li, W. H., et al. *J. Lipid Res.* 1988, **29:** 245–271. Kane, J. P. & Havel, R. J. "Disorders of the biogenesis and secretion of lipoproteins containing the B apolipoproteins." In: *The Metabolic Basis of Inherited Disease*, Scriver, C. R. et al. (eds.) McGraw-Hill 6th Edition. pp 1139–1164.
Ref. for DNA/AA Sequences	Chen, S. H., et al., *J. Biol. Chem.* 1986, **261:** 12918–12921. Knott, T. J., et al., *Nature* 1986, **323:** 734–738. Law, S. W., et al., *Proc. Natl. Acad. Sci. USA* 1986, **83:** 8142–8146. Cladaras, C., et al., *EMBL J.* 1986, **5:** 3495–3507. GenBank™/EMBL Data Bank accession no. J02610.

Apolipoprotein B Messenger RNA Editing Protein

Lawrence Chan

Synonyms	ApoB mRNA Editase
Abbreviations	HEPR, apobec1 (apoB mRNA editing component 1)
Classifications	Apolipoprotein B mRNA-Specific Cytidine Deaminase
Description	Apolipoprotein (apo) B mRNA editing protein is the catalytic subunit of an apoB mRNA editing enzyme complex. It is synthesized in the small intestine. By itself, the editing protein is inactive against apoB mRNA. It requires the other component(s) of the enzyme complex for activity. These other component(s) have not been isolated. As assayed by their ability to complement the editing protein in editing apoB mRNA in vitro, these component(s) have been detected in extracts from small intestine and liver as well as tissues that do not synthesize apoB. ApoB mRNA editing consists in a C → U conversion involving the first base of the codon CAA encoding glutamine-2153 to UAA, a stop codon. It is an intranuclear process that occurs posttranscriptionally coincident with mRNA splicing and polyadenylation (Lau et al., *J. Biol. Chem.* 1990, **266**: 20550–20554). The unedited mRNA codes for apoB-100 in the liver. ApoB-100 is a major protein in very low density lipoproteins, low density lipoproteins, and lipoprotein(a). It contains 4536 aa residues. Edited apoB mRNA codes for apoB-48 in the small intestine. ApoB-48 is an obligate component of chylomicrons and chylomicron remnants. It contains 2152 aa residues. ApoB mRNA editing was first described in 1987 (Powell et al., *Cell* 1987, **5**: 831–840; Chen et al., *Science* 1987, **238**: 363–366). To date, it is the only example of C → U editing in vertebrates (Chan, L., *BioEssays* 1993, **15**: 33–41).
Structure	ApoB mRNA editing protein exists as a dimer. Dimerization occurs spontaneously and is not dependent on disulfide bridge formation. Although the protein contains a consensus sequence for N-linked glycosylation (residues 57–59), it is not a glycoprotein. There is a leucine zipper-like sequence motif spanning residues 170–210. There is evidence that this sequence is important for dimerization. There is no experimental data on the secondary or tertiary structure of the apoB mRNA editing protein. ApoB mRNA editing proteins synthesized in rabbit reticulocyte lysate or in insect cells tends to aggregate *in vitro*.
Molecular Weight	28,177 (calculated from aa sequence, also consistent with SDS-polyacrylamide gel migration position).
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	The apoB mRNA editing protein shows limited sequence homology to mammalian and bacteriophage cytidine deaminases and deoxycytidylate deaminases. It has intrinsic cytidine deaminase activity. The protein does not itself edit apolipoprotein B mRNA. In the presence of extracts from various mammalian tissues or chicken small intestine, the protein will edit synthetic apoB mRNA *in vitro*, changing the C in codon-2153 (for Gln) from CAA to UAA, a stop codon. The active site comprises a His and two

Cys residues and the protein requires a zinc ion for catalysis. Dimerization appears to be required for activity.

Coenzymes/Cofactors

The apoB mRNA editing protein requires other protein(s) for its editing function. A number of other proteins have been identified that bind to apoB mRNA in the vicinity of the editing site (e.g., Lau et al., *Nucleic Acids Res.* 1990, 18: 5817–5821; Harris et al., *J. Biol. Chem.* 1993, **268**: 7382–7392). It is not clear, however, if they are the accessory proteins required for activity.

Substrates

The natural substrate is apoB-100 mRNA. The C → U conversion involves nucleotide 6666, the first nucleotide of codon-2153 (CAA) which is changed to a stop codon (UAA) upon editing. Other minor sites in apoB mRNA exist but do not seem to be of physiological significance. It is not clear if other naturally occurring mRNAs are edited by the apoB mRNA editing protein.

Inhibitors

Since the apoB mRNA editing protein requires zinc for activity, any zinc-specific chelating agent such as o-phenanthroline seems to inhibit activity. Cytidine deaminase-specific inhibitors have not been tested.

Biological Functions

Editing of apoB-100 mRNA by the editing protein produces apoB-48 mRNA. Therefore, apoB mRNA editing is a cellular mechanism for producing apoB-48 in place of apoB-100. ApoB mRNA editing occurs in the small intestine of all mammals examined. In most mammals, the editing protein and activity are absent in the liver. In rats and mice, substantial editing activity is detected in the liver. There is no apoB mRNA editing activity in the liver or small intestine in the chicken. The mRNA for apoB mRNA editing protein is detectable in the small intestine but not in the liver or other tissues in humans. In contrast, in rats and mice, the mRNA is detected in most tissues examined. It is not clear if the apoB mRNA editing protein produced in these other tissues functions as an editing enzyme for other (non-apoB) mRNAs.

Physiology/Pathology

The apoB mRNA editing protein produces apoB-48 mRNA from apoB-100 mRNA. The protein products of the edited and unedited apoB mRNA are apoB-48 and apoB-100, respectively. The structure of apoB-48 is identical to the N-terminal 48% (2152 aa residues) of apoB-100, which contains 4536 residues (see Apolipoprotein B-100 by L. Chan, **In** *Human Protein Data*, edited by André Haeberli, 1st Installment, 1992, VCH). The two proteins have very different properties. ApoB-100 is a major protein and an essential component in very low density lipoproteins, intermediate density lipoproteins, low density lipoproteins (LDL) and lipoprotein(a). It is also a ligand for the LDL receptor. LDL apoB-100 has a residence time of approximately 36 hours. ApoB-48 is an essential component of chylomicrons and chylomicron remnants. It is required for intestinal fat absorption. The LDL receptor-binding domain and the attachment site to apolipoprotein(a) in lipoprotein(a) are located in the C-terminal half of apoB-100 which is missing in apoB-48. Therefore, apoB-48 does not bind to the LDL receptor and does not participate in lipoprotein(a) production because it cannot form a complex with apo(a). The residence time of chylomicrons is 5–10 minutes. Therefore, the production of apoB-48 in place of apoB-100 (as a result of apoB mRNA editing) has major physiological significance in terms of lipoprotein metabolism and lipid homeostasis. Recently, Teng et al. *(J. Biol. Chem.* 1994, **269**:29395–29404) showed that somatic gene transfer of the apoB mRNA editing protein gene in mice virtually eliminates apoB – 100 and LDL production, suggesting that this is a potentially promising approach to the treatment of hyperlipidemia involving LDL.

Degradation	Unknown
Genetics/Abnormalities	The gene is localized on chromosome band 12p13.1–13.2. To date, there is no known genetic abnormality of the editing protein.
Half-life	Unknown
Concentration	Unknown
Isolation Method	The apoB mRNA editing protein has never been purified from its natural tissue of production. Information on the structure and function of this protein is deduced from partially purified tissue extracts or impure preparations of the protein expressed *in vitro*.
Amino Acid Sequence	The editing protein exists as a homodimer. Each monomer contains 236 aa residues and is 7 residues longer than the corresponding rat editing protein. It shares 69% sequence identity (excluding the 7-residue C-terminal extension in the human protein) with the latter. Three oligopeptide regions (residues 60–67, 89–104, 154–182) are homologous to highly conserved sequence blocks found in cytidine and deoxycytidylate deaminases from mammals and bacteriophages. Residues His-61, Cys-93 and Cys-96 correspond to the equivalent residues in T4 bacteriophage dCMP deaminase that are thought to be important for interaction with zinc at the active site. There are consensus sequences that potentially serve as phosphorylation sites for cAMP-dependent kinase (residue 71), protein kinase C (residues 13, 47, 54, 72 and 196) and casein kinase (residue 7). There is a tripeptide (residues 57–59) that signals N-linked glycosylation at residue 57. However, the apoB mRNA editing protein is not glycosylated *in vitro* and it is likely that this signal is not recognized *in vivo*. There is a leucine-zipper-like sequence spanning residues 173–210. It differs from classical leucine zippers in that there are proline residues breaking its coiled coil structure. This region may be important for dimerization and interaction with complementation factors.
Disulfides/SH-Groups	There are 8 Cys residues (including the 2 putative active-site Cys residues). There is no intermolecular disulfide bond. The oxidation state and intramolecular disulfide linkage of the individual Cys residues is unknown.
General References	Chan, L. *J. Biol. Chem.* 1992, **267**:25621–25624. Powell, L. M. et al. *Cell* 1987, **50**:831–840. Chen, S.-H. et al. *Science* 1987, **238**:363–366. Chan, L. *BioEssays* 1993, **15**:33–41. Teng, B. B. et al. *Science* 1993, **264**:1816–1819. Lau, P. P. et al. *Proc. Natl. Acad. Sci. USA* 1994, **91**:8522–8526. Chan, L. and Seeburg, P. H. *Science and Medicine* 1995, **2**:68–77.
Ref. for DNA/AA Sequences	Hadjiagapiou, C. et al. *Nucl. Acids Res.* 1994, **22**:1874–1879. Lau, P. P. et al. *Proc. Natl. Acad. Sci. USA* 1994, **91**:8522–8526. GenBank Accession Numbers: L25877, L26234.

Apolipoprotein CI and Apolipoprotein CIV

Vassilis I. Zannis and Dimitris Kardassis

Synonyms	Apolipoprotein serine (apoLp-ser)
Abbreviations	ApoCI
Classifications	A circulating plasma apolipoprotein member of the apolipoprotein family, it resides mostly in chylomicrons, very low density lipoprotein (VLDL), and high density lipoprotein (HDL).
Description	ApoCI is a circulating plasma apolipoprotein with defined protein and gene sequences (1-3). Plasma apoCI consists of 57 aa and it is a major component of very low density lipoprotein (VLDL) and a minor component of high density lipoprotein (HDL). The major sites of apoCI synthesis is the liver and the adrenal and monocyte macrophages and a minor site is the intestine. The apoCI mRNA encodes a protein of 83 aa including a 26 aa signal peptide which is cleaved cotranslationally in the rough endoplasmic reticulum. In plasma, lipoprotein particles have been described which contain apoB, apoCI, apoCII and apoCIII (LpB:C) or apoB, apoCI, apoCII, apoCIII and apoE (LpB:C:E) (4). The physiological significance as well as the sites of assembly of these particles is not known.
Structure	Comparison of the nucleotide and aa sequences of the apolipoproteins showed that they contain similar repeated units with sequence homologies. ApoCI contains a 33 and an 11 residue unit between nucleotides 7 to 31 and 40 to 50, respectively (5). Analysis of the secondary structure of apoCIII according to Chou and Fasman predicted α-helices between residues 7 to 14, 18 to 20, and 33 to 53 (6). Association of apoCI with egg yolk phosphatidylcholine increases its α-helical content from 56% to 73% (6). The phospholipid binding domain of apoCI is between residues 1 to 38. The association of apoCI with egg yolk lecithin forms discoidal particles with a minor axis of 4 nm and a major axis of 20 nm (7). The initial binding studies of apoCI to egg yolk lecithin lead to the hypothesis of the amphipathic α-helix. According to the model proposed by Segrest, the amphipathic α-helices have two phases, one hydrophobic which can associate with the hydrophobic fatty acyl chains of the phospholipids and another hydrophilic which can associate with the phospholipid head groups and the aqueous phase (7).
Molecular Weight	6,605
Sedimentation Coeff.	Unknown
Isoelectric Point	The most basic apolipoprotein estimated pI ~ 7.5
Extinction Coeff.	Unknown
Enzyme Activity	Activates moderately the enzyme lecithin:cholesterol acyl transferase (LCAT) (3). Inhibits the activity of cellular phospholipase A2 (Poensger, *Biochim. Biophys. Acta* 1990, **1042**:188).
Coenzymes/Cofactors	None
Substrates	None

Inhibitors	None
Biological Functions	ApoCI activates moderately LCAT. This activation is 25% of that achieved with apoA-I (3). The region between residues 17 to 57 activates LCAT as effectively as the intact apoCI (6). There are conflicting reports on the role of apoCI in the regulation of lipoprotein lipase (3). It was shown previously that apoCI inhibits the catabolism of apoE-containing lipoproteins by the perfused liver (3). Recently, it has been shown that apoCI and to a lesser extent apoCII (but not apoCIII) inhibits the binding of apoE-containing lipoproteins such as β-VLDL to the LDL receptor-related protein (LRP) as well as to the LDL receptor (Weisgraber et al. *J. Biol. Chem.* 1990, **265**:22453). Transgenic animals expressing apoCI have increased triglyceride levels, suggesting a physiological role for apoCI in the catabolism of triglyceride-rich lipoproteins (Simonet et al. *J. Biol. Chem.* 1991, **266**:8651). Recently it has been shown that apoCI binds to phospholipids and cell membranes and inhibits cellular phospholipase A2 possibly by preventing the access of the enzyme to its phospholipid substrate.
Physiology/Pathology	Inhibits the binding of apoE-containing β-VLDL to the LDL receptor-related protein (LRP) and the LDL receptor.
Degradation	A portion of apoCI is taken up by the liver as a component of VLDL and IDL. The uptake is mediated by LDL and LRP receptors and requires a high E:CI ratio. A certain percentage of apoCI must also be removed from plasma following proteolytic degradation.
Genetics/Abnormalities	The cDNA and gene sequences of the human apoCI gene have been reported (9,10) (Figure 1A). The human apoE, apoCI apoCIV, and apoCII genes are closely linked. The cluster of the four genes maps on the long arm of chromosome 19 and spans a 45 Kb region. The apoCI gene is located 5.5 Kb 3' of the apoE gene, and the apoCI' pseudogene is located 7.5 Kb 3' to the apoCI gene. The human apoCII gene is found 20 Kb 3' of the apoCI' pseudogene and the human apoCIV gene is 0.55 Kb 5' of the apoCII gene. The LDL receptor gene is closely linked with the apoE/ apoCI/apoCII/apoCIV cluster (Figure 1B). There are no reports of genetic alterations involving the human apoCI gene. The apoCIV gene has been identified recently (Allan et al. *Genomics* 1995, **28**:291). The human gene encodes for a 127aa long protein including a 25 residue signal peptide. The apoCIV gene has been highly conserved in mammalian species and has homologies to upstream regions of the apoCII gene. In rabbit plasma, apoCIV is a component of VLDL and HDL. The protein is post-translational modified with carbohydrate chains containing sialic adic. It contains two putative amphipathic α-helices and can form discoidal particles with DMPC (Zhang et al. *J. Biol. Chem.* 1996, **271**:1776). Gene regulation: The region in the intergenic sequence between apoCI gene and apoCI' pseudogene was originally shown to contain elements that are required for the hepatic expression of the human apoE and apoCI genes and was designated hepatic control region 1 (HCR-1) (Figure 1B). HCR-1 was mapped within 319 bp approximately 15 Kb downstream of the apoE gene. This sequence is sufficient to direct liver specific expression of the apoE gene in a copy dependent manner (Dang et al. *J. Biol. Chem.* 1995, **270**:22577). A second hepatic control region designated HCR-2 was identified 27 Kb 3' to the apoE gene in the middle of the intergenic sequence between the apoCI' pseudogene and the newly discovered apoCIV HCR-2 gene which has 85% sequence identity to the HCR-1 (Figure 1B). A 632 bp sequence containing HCR-2 can by itself direct hepatic transcription of the apoE gene (Allan et al. *J. Biol. Chem.* 1995,

260:26278). A variety of regulatory elements extending from 5 kb 5' of the apoE gene to 1 kb 3' of the apoCI' pseudogene also control positively or negatively the expression of the apoCI gene in different tissues. Hepatic expression requires sequences extending 3.1 kb 5' of apoCI and the intergenic sequences between the apoCI gene and apoCI' pseudogene that contains the HCR-1. Deletions or additions in the 5' and 3' sequences cause different expression patterns in various tissues studied.

Half-life

Residence time 3.24 ± 0.27 days. Catabolism depends on the lipoprotein particle it is found (Malmendier et al. *Atherosclerosis* 1986, **62**:167).

Concentration

40 - 60 mg/L

Isolation Method

VLDL is isolated from plasma by isopycnic ultracentrifugation at 100,000xg for 28 hrs at 0°C. VLDL in 0.15M NaCl, 1mM EDTA, 5mM NH_4HCO_3 is delipidated by dropwise addition with vortexing of 7.5 volumes of methanol or ethanol at 0°C followed by mixing with 15 volumes of diethyl ether. The precipitate is collected by low speed centrifugation (3000 rpm for 4 min.). ApoVLDL is solubilized in buffers containing 0.1 to 0.2M sodium decyl or sodium dodecyl sulfate and is fractionated by column chromatography in buffers containing 5M guanidine-HCl or 2mM SDS. Fractionation of apoVLDL on Sephadex G-200 results in the isolation of a fraction containing all the apoC peptides (apoCI, apoCII, apoCIII). Fractionation on Biogel-A 0.5M in 0.1M NH_4HCO_3, 2mM SDS, pH 7.9, results in separation of apoCI and most of the apoCIII from apoCII. Fractionation of apoVLDL on Biogel P-100 in 0.1M NH_4HCO_3, 5M guanidine-HCl separates apoCII and apoCIII from apoCI (8). The mixtures of apoC peptides can be further fractionated by DEAE-cellulose chromatography, high pressure liquid chromatography (Weisgraber et al. *J. Biol. Chem.* 1990, **265**:22453), and isoelectric focusing. Purified apoCII is solubilized in 0.1M NH_4OH, apoCIII in 0.1M NH_4HCO_3, and apoCI in 1M acetic acid (8).

Amino Acid Sequence

Plasma apoCI consists of 57 aa of known sequence.

Disulfides/SH-Groups

None

General References

1. Herbert, P.N. et al. Familial lipoprotein deficiency: Abetalipoproeteinemia, hypobetalipoproteinemia and Tangier disease. In: *The Metabolic Basis of Inherited Disease.* Stanbury, J.B. et al. (eds.) McGraw-Hill, New York 1982, **5th ed.** pp.589-651.
2. Karathanasis, S.K. et al. Nucleotide and corresponding amino acid sequences of human apoA-I, apoA-II, apoCII, apoCIII and apoE cDNA clones. In: *Biochemistry and Biology of Plasma Proteins.* Scanu, A. and Spector, A. (eds.) 1985, pp. 475-493.
3. Zannis, V.I., Breslow, J.L. Genetic mutation affecting human lipoprotein metabolism. *Adv. Hum. Genet.* 1985, **14**:125-215, and **14**: 383-386, and Zannis, V.I., Kardassis, D., Zanni, E.E. Genetic mutations affecting human lipoproteins, their receptors, and their enzymes. In: *Adv. Hum. Genetics.* Harris, H. and Hirschhorn K (eds.) Plenum Press, NY. 1993, **Vol 21,** pp. 145-319.
4. Alaupovic, P. et al. Lipoprotein particles in hypertriglyceridemic states. In: *Advances in Experimental Medicine and Biology.* Malmendier, C.L. and Alaupovic, P. (eds.) 1988, **Vol. 243**, pp. 289-297.
5. Li, W.H. et al. *J. Lipid Res.* 1988, **29**:245-271.
6. Sparrow, J.T. and Gotto, A.M., Jr. *CRC Critical Reviews in Biochemistry* 1982, **13**:87-107.

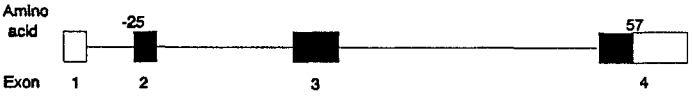
7. Atkinson, D. and Small, D.M. *Ann. Rev. Biophys. Chem.* 1986, **15**:403-456.
8. Herbert, P.N. et al. Apolipoprotein quantitation. In: *The Lipoprotein Molecule*. Peeters, H. (ed.) Plenum Publishing Corp. 1978, pp. 35-56.

Ref. for DNA/AA Sequences

9. Knott, T.J. et al. *Nucleic Acids Res.* 1984, **12**:3909-3915.
10. Lauer, S.J. et al. *J. Biol. Chem.* 1988, **263**:7277-7286.
11. Lusis, A.J. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:3929-3933.
12. Allan, C.M. et al. *Genomics* 1995, **28**:291-300.
13. Dang, Q. et al. *J. Biol. Chem.* 1995, **270**:22577-22585.
14. Allan, C. et al. *J. Biol. Chem.* 1995, **270**:26278-26281.
15. Zhang, L.H. et al. *J. Biol. Chem.* 1986, **271**:1776-1783.

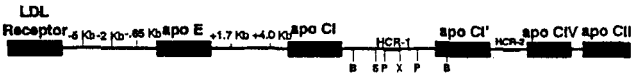
A.

The human apo CI gene [a]



B.

Linkage of human apoE, apoCI, apoCII, apoCIV and LDL receptor genes [b]



b: Lauer et al. (1988), Lusis et al. (1986), Dang et al. (1995), Allan et al. (1995)

Apolipoprotein CII

Eleni E. Zanni and Vassilis I. Zannis

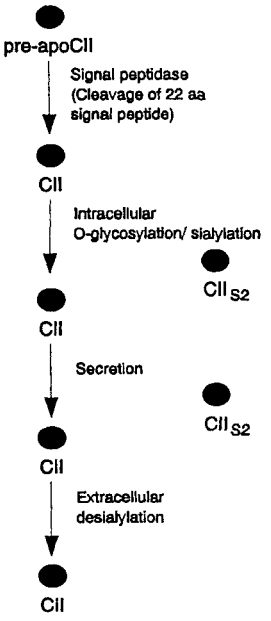
Synonyms	Apolipoprotein glutamine (apoLp-Gln)
Abbreviations	ApoCII
Classifications	A circulating plasma apolipoprotein, a member of the apolipoprotein family, it resides mostly in chylomicrons, very low density lipoprotein (VLDL) and high density lipoprotein (HDL).
Description	ApoCII is a small exchangeable apolipoprotein with defined protein and genes sequences (1-3). It is a major component of chylomicrons and very low density lipoprotein (VLDL) and a minor component of high density lipoprotein (HDL). The major site of apoCII mRNA and protein synthesis is the liver, and a minor site is the intestine (3). The apoCII mRNA encodes a 101 aa long protein which contains a 22 residue long signal peptide. Following cotranslational cleavage of the signal peptide the 79 residue long protein is partially modified in the ER/Golgi by O-glycosylation and sialylation. The glycosylation is inefficient and the protein is secreted in the disialylated (CII_{s2}) and nonsialylated (CII) forms at 1:1 ratio. $ApoCII_{s2}$ is subsequently desialylated in plasma to generate the asialo apoCII form (Figure 1A). The physiological importance of the modification of apoCII is not known. In plasma lipoprotein particles Lp have been described which contain apoB, apoCI, apoCII, and apoCIII (LpB:C) or apoB, apoCI, apoCII, apoCIII and apoE (LpB:C:E) (4). The sites of assembly and the physiological significance of these particles are not known.
Structure	Comparison of the nucleotide and aa sequences of the apolipoproteins showed that they contain similar repeated units with sequence homologies. ApoCII contains a 33 and 22 residue unit between nucleotides 18 to 50 and 51 to 72, respectively (5). Analysis of the secondary structure of apoCII according to Chou and Fasman predicted amphipathic α-helices between residues 13 to 22, 28 to 39, and 42 to 51, and β-turns in residues 9 to 12, 23 to 26, and 52 to 55, and β-sheet between residues 60 to 74 (6) (Figure 1B). Association of apoCII with phospholipids increases its α-helical content from 35% to 59% (6). Association of apoCII with egg lecithin forms discoidal particles with a minor axis of 4 nm and a major axis of 20 nm (7). The dissociation constant Kd of apoCII from model lipid emulsions (microemulsions and large emulsions containing egg yolk lecithin and triolein and cholesterol) ranges from 0.45 to 1.07 μm and the saturation number N (number of apolipoprotein molecules/1000 phospholipid molecules) ranges from 8.3 to 11.8 (7).
Molecular Weight	8824 (calculated on the basis of aa sequence).
Sedimentation Coeff.	Unknown
Isoelectric Point	CII: 4.82; CII_{S2}: 4.50
Extinction Coeff.	None
Enzyme Activity	None
Coenzymes/Cofactors	None

Substrates	None
Inhibitors	None
Biological Functions	ApoCII is a potent activator of lipoprotein lipase but not hepatic lipase (3). Maximal stimulation of LPL is achieved at a molar ratio of apoCII to enzyme of 1:1. The reported dissociation constant of the complex is 10^{-8}M to 10^{-10}M (6). The complex dissociates at high NaCl concentrations. The region 43 to 79 which can bind to VLDL and phospholipid vesicles is sufficient for maximum activation of the lipoprotein lipase and the C-terminal region 55 to 79 retains 90% of the activation capacity (6). The physiological importance of apoCII in activating LPL has been established by the finding of individuals with inherited apoCII deficiency who have great difficulty in clearing triglyceride-rich lipoprotein particles from their plasma (3). Natural mutations in the C-terminal apoCII region or mutations generated by *in vitro* mutagenesis (Holtfreter and Stoffel, *Biol. Chem. Hoppe-Seyler* 1988, **369:**1045) abolish the ability of the mutant protein to activate lipoprotein lipase. Recently it has been shown that apoCI and to a lesser extent apoCII (but not apoCIII) inhibit the binding of apoE containing lipoproteins such as β-VLDL to the LDL receptor-related protein (LRP) as well as to the LDL receptor (Weisgraber et al. *J. Biol. Chem.* 1990, **265**:22453).
Physiology/Pathology	Activates lipoprotein lipase, the enzyme which catalyzes the hydrolysis of triglycerides of chylomicrons and VLDL. Mutations in apoCII are associated with familial type I hypertriglyceridemia. Overexpression of the apoCII gene in transgenic mice causes severe hypertriglyceridemia which is exacerbated by high fat diet and delayed clearance of VLDL (Schachter et al. *J. Clin. Invest.* 1994, **93**:16883).
Degradation	A portion of apoCII is taken up by the liver as a component of VLDL and IDL and HDL subfraction. The uptake is mediated by LDL and LRP receptors and requires a high E:CII ratio. A certain percentage of apoCII must also be removed from plasma following proteolytic degradation.
Genetics/Abnormalities	The cDNA and gene sequences encoding the human apoCII have been reported (9-13). The human apoCII gene is closely linked to the human apoE and apoCI genes. The gene cluster maps on the long arm of chromosome 19 (14) (Figure 1C). Initial protein analysis of patients with apoCII deficiency by two dimensional polyacrylamide gel electrophoresis showed the presence of small quantities of variant protein forms. Subsequent nucleotide sequence analysis of the patients' genes showed a variety of mutations which are associated with apoCII deficiencies and type I hypertriglyceridemia. These mutations are shown in Figure 1D. Gene Regulation: The -550 to + 18 apoCII promoter region contains five regulatory elements: A (-74/-44), B (-102/-81), C (-159/-116), D (-288/-265) and E (-497/-462) and is sufficient for the hepatic and intestinal expression of the apoCII gene *in vitro*. Linkage of the -550/+18 apoCII promoter with HCR-1 increased the strength of the proximal promoter 2.5 to 3-fold in HepG2 cells (Kardassis et al. *Circulation* 1996, in press).
Half-life	Residence time 2.90 ± 0.24 days (Malmendier et al. *Adv. Exp. Med. and Biol.* 1988, **243**:299). Catabolism depends on the lipoprotein particles it is found.
Concentration	30-50 mg/L

Isolation Method	VLDL is isolated from plasma by isopycnic ultracentrifugation at 100,000xg for 28 hrs at 0°C. VLDL in 0.15M NaCl, 1mM EDTA, 5mM NH_4HCO_3 is delipidated by dropwise addition with vortexing of 7.5 volumes of methanol or ethanol at 0°C followed by mixing with 15 volumes of diethyl ether. The precipitate is collected by low speed centrifugation (3000 rpm for 4 min.). ApoVLDL is solubilized in buffers containing 0.1 to 0.2M sodium decyl or sodium dodecyl sulfate and is fractionated by column chromatography in buffers containing 5M guanidine-HCl or 2mM SDS. Fractionation of apoVLDL on Sephadex G-200 results in the isolation of a fraction containing all the apoC peptides (apoCI, apoCII, apoCIII). Fractionation on Biogel-A 0.5M in 0.1M NH_4HCO_3, 2mM SDS, pH 7.9, results in separation of apoCI and most of the apoCIII from apoCII. Fractionation of apoVLDL on Biogel P-100 in 0.1M NH_4HCO_3, 5M guanidine-HCl separates apoCII and apoCIII from apoCI (8). The mixtures of apoC peptides can be further fractionated by DEAE-cellulose chromatography, high pressure liquid chromatography (Weisgraber et al. *J. Biol. Chem.* 1990, **265**:22453), and isoelectric focusing. Purified apoCII is solubilized in 0.1M NH_4OH, apoCIII in 0.1M NH_4HCO_3, and apoCI in 1M acetic acid (8).
Amino Acid Sequence	Plasma apoCII consists of 79 aa of known sequence.
Disulfides/SH-Groups	None
General References	1). Herbert, P.N. et al. "Familial lipoprotein deficiency: Abetalipoproteinemia, hypobetalipoproteinemia and Tangier disease". In: *The Metabolic Basis of Inherited Disease.* Stanbury, J.B. et al. (eds.) McGraw-Hill, New York 1982, **5th ed.**, pp. 589-651. 2). Karathanasis, S.K. et al. Nucleotide and corresponding amino acid sequences of human apoA-I, apoA-II, apoCI, apoCII, apoCIII and apoE cDNA clones. In: *Biochemistry and Biology of Plasma Proteins,* Scanu, A. and Spector, A. (eds.) 1985, pp. 475-493. 3). Zannis, V.I. and Breslow, J.L. Genetic mutation affecting human lipoprotein metabolism. *Adv. Hum. Genet.* 1985, **14**:125-215, and **14**:383-386 and Zannis, V.I., Kardassis, D., Zanni, E.E. Genetic mutations affecting human lipoproteins, their receptors and their enzymes. In: *Adv. Hum. Genetics.* Harris, H. and Hirschorn, K. (eds.) Plenum Press, NY. 1993, **Vol 21,** pp. 145-319. 4). Alaupovic, P. et al. Lipoprotein particles in hypertriglyceridemic states. In: *Advances in Experimental Medicine and Biology,* Malmendier C.L. and Alaupovic, P. (eds.) 1988, Vol. **243**, pp. 289-297. 5). Li, W.H. et al. *J. Lipid Res.* 1988, **29**:245-271. 6). Sparrow, J.T., Gotto, A.M., Jr. *CRC Critical Reviews in Biochemistry* 1982, **13**:87-107. 7). Atkinson, D., Small, D.M. *Ann. Rev. Biophys. Chem.* 1986, **15**:403-456. 8). Herbert, P.N. et al. Apolipoprotein quantitation. In: *The Lipoprotein Molecule,* Peeters, H. (ed.) Plenum Publishing Corp. 1978, pp.35-56.
Ref. for DNA/AA Sequences	9). Sharpe, C.R. et al. *Nucleic Acid Res.* 1984, **12**:3917-3932. 10). Fojo, S.S. et al. *Proc. Natl. Acad. Sci. USA* 1984, **81**:6354-6357. 11). Myklebost, O. et al. *J. Biol. Chem.* 1984, **259**:4401-4404. 12). Jackson, C.L. et al. *Proc. Natl. Acad. Sci. USA* 1984, **81**:2945-4949. 13). Wei, C.F. et al. *J. Biol. Chem.* 1985, **260**:15211-15221. 14). Lusis, A.J. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:3929-3933

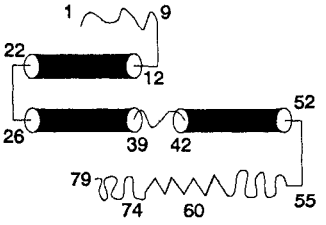
A.

Intra and extracellular modifications of human apo CII [a]



B.

Predicted secondary structure of human apo CII [b]



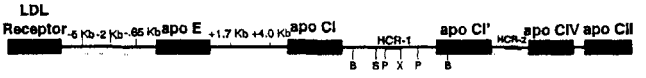
Symbols:

amphipathic a- helix * | β- turn | β- sheet | random

* The orientation of the helices is not known

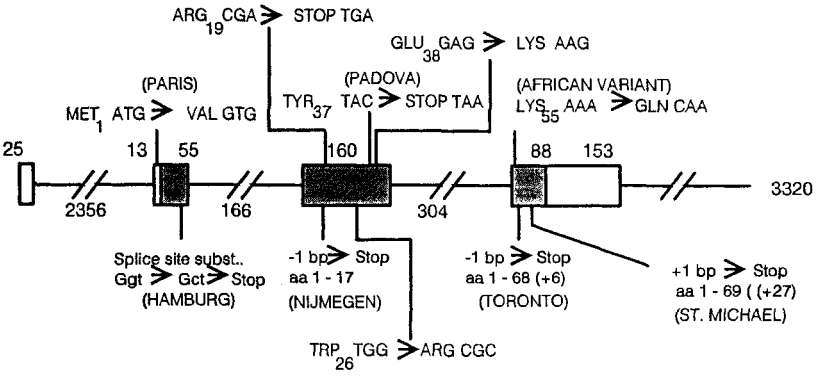
C.

Linkage of human apoE, apoCI and apoCII genes [c]



D.

Human apoCII gene structure and genetic variation [d]



a: Hussain and Zannis, Biochemistry 1990; 29: 209, b: Reviewed in Sparrow and Gotto (1982), c: Lusis et al (1986) d: Conelly et al, J. Clin. Invest. 1987; 80:1597; Cox et al, J. Med. Genetics 1988; 25:649; Fojo et al J. Clin. Invest. 1988; 82:1489; Fojo et al. J. Biol. Chem. 1988; 263: 17913; Menzel et al J. Clin. Invest. 1986; 77:595; Reina et al. J. Lipid Res. 1992; 33:1823; Inadera et al. BBRC 1993; 1:74; Pullinger et al, Hum. Mol. Genet.1993; 2:69.

Apolipoprotein CIII

Dimitris Kardassis and Eleni E. Zanni

Synonyms

Apolipoprotein alanine (apoLp-Ala)

Abbreviations

ApoCIII

Classifications

A circulating plasma apolipoprotein member of the apolipoprotein family, it resides mostly in chylomicrons, very low density lipoprotein (VLDL) and high density lipoprotein (HDL).

Description

ApoCIII is a small exchangeable apolipoprotein with defined protein and gene sequences (1, 2). It is a major component of chylomicrons and very low density lipoprotein (VLDL), and a minor component of high density lipoprotein (HDL). Small quantities of apoCIII are also found in the intermediate density lipoprotein (IDL) and low density lipoprotein (LDL) (3). ApoCIII protein and mRNA synthesis occurs predominantly in the liver and to a lesser extent in the intestine, and it is developmentally regulated (3). The apoCIII mRNA encodes a 99 aa long protein which contains a 20 residue long signal peptide. Following cotranslational cleavage of the signal peptide the 79 residue long protein is modified in the ER/Golgi by O-glycosylation and sialylation and is secreted in the disialylated form ($CIII_{s2}$). The protein is subsequently desialylated in plasma to generate a mixture consisting of 14% of the asialo form (CIII-0), 59% of the monosialo form ($CIII_{s1}$), and 27% of the disialo form ($CIII_{s2}$) (Figure 1A). The physiological importance of the modifications of apoCIII is not known. In tissue culture experiments, 90% of secreted apoCIII is found in lipoprotein free fraction, and it can associate with exogenously added VLDL and HDL and to a lesser extent with LDL. Secreted apoCIII is found as a component of particles with defined lipid and apolipoprotein composition designated LpB:CIII, LpB:E:CIII, LpB:A-II:CIII, LpAI:AII:CIII as well as complexes containing apoCI and apoCII. Similar particles containing LpB:C or LpB:C:E have been isolated from plasma (4). The physiological significance as well as the sites of assembly of these particles is not known.

Structure

Comparison of the nucleotide and aa sequences of the apolipoproteins showed that they contain similar repeated units with sequence homologies. ApoCIII contains a 33, 11, and 22 residue unit between nucleotides 8 to 40, 41 to 51, and 52 to 72, respectively (5). Analysis of the secondary structure of apoCIII according to Chou and Fasman predicted α-helices between residues 1 to 39 and 54 to 69, β-turns at residues 39 to 42 and 72 to 75, and an amphipathic α-helix between 40 to 67 (Figure 1B) (6). Association of apoCIII with phospholipids increases its α-helical content from 26% to 70% (6). Based on phospholipid binding properties of apoCIII peptides, it has been proposed that the phospholipid binding domain of apoCIII is localized in the 39 C-terminal aa residues 41 to 79 (7). This region contains the 11 and 22 residue repeated unit. Association of apoCIII with egg lecithin forms discoidal particles with a minor axis of 4nm and a major axis of 20nm (7). Small angle X-ray scattering studies showed the particles of DMPC-apoCIII complexes have ellipsoidal structure 17x5nm and a 1nm thick shell representing the apolipoprotein on the surface of the hydrophobic chains of the phospholipids (7). The dissociation constants Kd of apoCIII from model lipid emulsions (microemulsions and large emulsions containing egg yolk lecithin and triolein) range from 0.53 to 1.07 and the saturation numbers N (number of apolipoprotein molecules/ 1000 phospholipid molecules) range from 8.2 to 13.2.

Molecular Weight	8,750
Sedimentation Coeff.	Unknown
Isoelectric Point	CIII-0 = 5; $CIII_{s1}$ = 4.8; $CIII_{s2}$ = 4.65 (3)
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	ApoCIII inhibits *in vitro* the function of lipoprotein lipase, the enzyme that hydrolyzes the triglyceride moieties of chylomicrons and VLDL (3). ApoCIII also inhibits the binding of apoE-containing lipoproteins to the LDL receptor but not to the LDL receptor-related protein (LRP) (Weisgraber et al. *J. Biol. Chem.* 1990, **265**:22413). The receptor mediated binding and catabolism of lipoproteins is enhanced by apoE and is inhibited by apoCIII (3). These observations suggested a role of apoCIII in the catabolism of triglyceride rich lipoprotein. Recent experiments showed that overexpression of the human apoCIII gene in transgenic mice results in severe hypertriglyceridemia and accumulation in plasma of apoB-48 containing remnants. In addition, inactivation of the apoCIII gene by homologous recombination increased the clearance of chylomicrons and conferred resistance to post-prandial hypertriglyceridemia (Maeda et al. *J. Biol. Chem.* 1994, **269**:23610). These findings reinforce the notion of a potential role of apoCIII in the catabolism of triglyceride-rich lipoproteins *in vivo*.
Physiology/Pathology	Catabolism of triglyceride rich lipoproteins. Excessive plasma apoCIII levels are associated with hypertriglyceridemia.
Degradation	A portion of apoCIII is taken up by the liver as a component of VLDL and IDL and by HDL subfraction. The uptake is mediated by LDL and LRP receptors and requires a high E:CIII ratio. A certain percentage of apoCIII must also be removed from plasma following proteolytic degradation.
Genetics/Abnormalities	The cDNA and gene sequences for human apoCIII have been reported (9-11). The apoCIII gene is 3.5 kb long and contains four exons and three introns (Figure 1C), and it is closely linked to the human apoA-I and apoA-IV genes. The apoCIII gene is localized 2.5 kb downstream of the apoA-I gene and 5 kb upstream of the apoA-IV gene (Figure 1D). The direction of transcription of the apoCIII gene is opposite to that of apoA-I and apoA-IV genes. The cluster of the three genes is mapped on the long arm of chromosome 11, in the region 11g13→9ter (5). Two human conditions characterized by severe deficiencies of plasma apoCIII and HDL are associated with premature atherosclerosis. One condition has resulted from deletion of apoA-I, apoCIII, apoA-IV locus (Ordovas et al. *J. Biol. Chem.* 1989, **264**:16369) and the other from an inversion of the apoA-I, apoCIII locus. The break points of this inversion are in the third exon of the apoA-I gene and the first intron of the apoCIII gene (Karathanasis et al. *Proc. Natl. Acad. Sci. USA* 1987, **84**:7198) Patients with type III, type IV, and type V hyperlipoproteinemia have elevated plasma apoCIII levels. Variations in the relative concentration of apoCIII isoproteins have also

been observed which deviate from normal values (3). For instance, patients with type V hyperlipoproteinemia are reported to have statistically lower relative apoCIII-0 concentrations (3). Similarly, absence of sialylated apoCIII forms ($CIII_{s1}$, and $CIII_{s2}$) has been reported due to a genetic alteration of Thr-74 (ACT) to Ala-74 (GCT) (Maeda et al. *J. Lipid Res.* 1987, **28**:1405).

Gene Regulation: The -792 to +24 apoCIII promoter region contains a set of four proximal (A through D) and six distal (E through J) regulatory elements which can direct hepatic and intestinal transcription *in vitro* (Ogami et al. *J. Biol. Chem.* 1990, **265**:9808). Hormone nuclear receptors which bind to the regulatory element B and I and SP1 which binds to the regulatory elements F, H and I, are important for the transcriptional regulation of the apoCIII gene (Figure 1E). The regulatory elements F to J of the apoCIII gene act as an enhancer of the apoA-I, apo CIII and apoA-IV genes in vitro (Kardassis et al. *Hypertension* 1996, **27**:980). It appears that the transcriptional activation of these genes involves synergistic interactions between nuclear hormone recpetors which bind to proximal promoter and distal enhancer sites. These interactions are facilitated by SP1 which binds to three sites on the apoCIII enhancer (Figure 1E).

Half-life

Residence time 2.45 ± 0.33 days. Catabolism depends on the lipoprotein particle it is found (Malmendier et al. *Atherosclerosis* 1988, **69**:51).

Concentration

120 - 140 mg/L

Isolation Method

VLDL is isolated from plasma by isopycnic ultracentrifugation at 100,000xg for 28 hrs at 0°C. VLDL in 0.15M NaCl, 1mM EDTA, 5mM NH_4HCO_3 is delipidated by dropwise additiow ith vortexing of 7.5 vol umes of methanol or ethanol at 0°C followed by mixing with 15 volumes of diethyl ether. The precipitate is collected by low speed centrifugation (3000 rpm for 4 min.). ApoVLDL is solubilized in buffers containing 0.1 to 0.2M sodium decyl or sodium dodecyl sulfate and is fractionated by column chromatography in buffers containing 5M guanidine-HCl or 2mM SDS. Fractionation of apoVLDL on Sephadex G-200 results in the isolation of a fraction containing all the apoC peptides (apoCI, apoCII, apoCIII). Fractionation on Biogel-A 0.5M in 0.1M NH_4HCO_3, 2mM SDS, pH 7.9, results in separation of apoCI and most of the apoCIII from apoCII. Fractionation of apoVLDL on Biogel P-100 in 0.1M NH_4HCO_3, 5M guanidine-HCl separates apoCII and apoCIII from apoCI (8). The mixtures of apoC peptides can be further fractionated by DEAE-cellulose chromatography, high pressure liquid chromatography (Weisgraber et al. *J. Biol. Chem.* 1990, **265**:22453), and isoelectric focusing. Purified apoCII is solubilized in 0.1M NH_4OH, apoCIII in 0.1M NH_4HCO_3, and apoCI in 1M acetic acid (8).

Amino Acid Sequence

Plasma apoCIII consists of 79 aa of known sequence.

Disulfides/SH-Groups

None

General References

1. Herbert, P.N. et al. Familial lipoprotein deficiency: Abetalipoproteinemia, hypobetalipoproteinemia and Tangier disease. In: *The Metabolic Basis of Inherited Disease,* Stanbury, J.B. et al. (eds.) McGraw-Hill, New York, 1982, **5th ed.** pp. 589-651.

2. Karathanasis, S.K. et al. Nucleotide and corresponding amino acid sequences of human apoA-I, apoA-II, apoCI, apoCII, apoCIII and apoE cDNA clones. In: *Biochemistry and Biology of Plasma Proteins,* Scanu, A. and Spector, A. (eds.) 1985, pp. 475-493.

3. Zannis, V.I. and Breslow, J.L. Genetic mutation affecting human lipoprotein metabolism. *Adv. Hum. Genet.* 1985, **14**:125-215 and **14**:383-386 and
Zannis, V.I., Kardassis, D. Zanni, E.E. Genetic mutations affecting human lipoproteins, their receptors, and their enzymes. *Adv. Hum. Genetics* Harris, H. and Hirschhorn, K. (eds.) Plenum Press, NY. 1993, **Vol. 21**, pp. 145-319.
4. Alaupovic, P. et al. Lipoprotein particles in hypertriglyceridemic states. In: *Advances in Experimental Medicine and Biology*, Malmendier, C.L. and Alaupovic, P. (eds.) 1988, vol. **243**, pp. 289-297.
5. Li, W.H. et al. *J. Lipid Res.* 1988, **29**:245-271.
6. Sparrow, J.T. and Gotto, A.M., Jr. *CRC Critical Reviews in Biochemistry* 1982, **13**:87-107.
7. Atkinson, D. and Small, D.M. *Ann. Rev. Biophys. Chem.* 1986, **15**:403-456.
8. Herbert, P.N. et al. Apolipoprotein quantitation. In: *The Lipoprotein Molecule*, Peeters, H. (ed.) Plenum Publishing Corp. 1978, pp. 35-56.

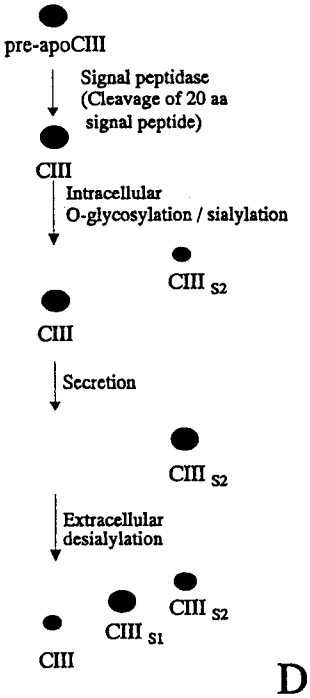
Ref. for DNA/AA Sequences

9. Sharpe, C.R. et al. *Nucleic Acids Res.* 1984, **12**:3917-3932.
10. Karathanasis, S.K. et al. *J. Lipid Res.* 1985, **2**:451-456.
11. Protter, A.A. et al. *DNA* 1984, **3**:449-456.
12. Ogami, K. et al. *J. Biol. Chem.* 1990, **265**:9808-9815.
13. Kardassis, D. et al. *Hypertension* 1996, **27**:980-1008.

Molecular Model according to Doolittle, R. F. (*Annu. Rev. Biochem.*, 1984)

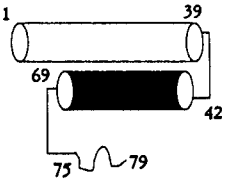
A

Intra and extracellular modifications of human apo CIII [a]

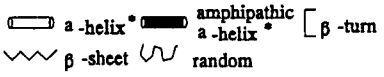


B

Predicted secondary structure of human apo CIII [b]



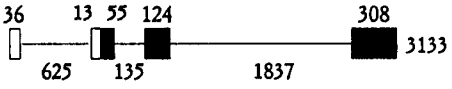
Symbols:



* The orientation of the helices is not known

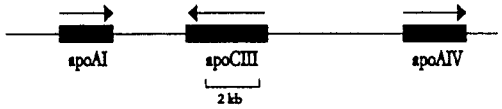
C

The human apo CIII gene [c]



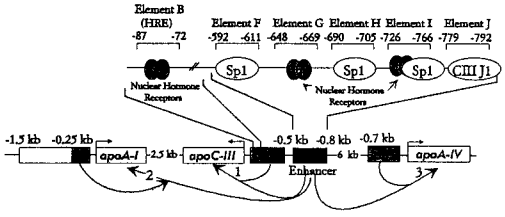
D

Linkage of human apoAI, apoCIII and apoAIV genes [d]



E.

Regulatory elements and factors involved in the transcriptional regulation of the human apoCIII gene: Role of the apoCIII enhancer in the regulation of the closely linked apoA-I and apoA-IV genes [e]



A: Hussain and Zannis, *Biochemistry* 1990, **29**:209; B: reviewed in Sparrow and Gotto (1982); C: Protter et al. (1984); D: reviewed in Li et al. (1988); E: Ogami et al. (1990); Kardassis et al. (1996)

Apolipoprotein D

Carlos López-Otín

Synonyms	GCDFP-24 (gross cystic disease fluid protein of 24 kDa)
Abbreviations	Apo D
Classification	Lipocalin
Description	Apo D is a glycoprotein originally isolated from the high density lipoprotein fraction of human plasma but whose precise role in lipid metabolism remains unclear. Unlike other apolipoproteins which are expressed predominantly in the liver and the intestinal epithelium, apo D has been also found in other tissues like adrenal glands, pancreas, kidney, testes, placenta, spleen, and brain. It is also the major protein component present in cyst fluid from women with gross cystic disease of the breast.
Structure	Apo D consists of a single polypeptide chain of 169 aa with two carbohydrate side chains, N-linked to Asn-45, and Asn-78. These carbohydrate chains are of the complex type and possess sialylated triantennary and fucosylated sialylated biantennary structure, respectively. Apo D lacks in its structure the amphipathic α-helices which mediate lipid binding in other members of the apolipoprotein family. The tertiary structure has not yet been determined, although on the basis of molecular modeling, it has been suggested that the polypeptide chain is folded into an eight-stranded β-barrel that is open at the top and closed at the bottom.
Molecular Weight	24,000-32,000 (SDS-PAGE); 81,860 (sedimentation equilibrium); 18% carbohydrate.
Sedimentation Coeff.	4.3 S
Isoelectric Point	4.7-5.2 (4 major bands)
Extinction Coeff.	9.5 (280nm, 1%, 1cm, pH 7.0)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Apo D belongs to the lipocalin family of proteins whose common function is to bind and transport small hydrophobic substances in the plasma. Proposed ligands for apoD include cholesteryl esters, progesterone, arachidonic acid, heme-related compounds, or odorant substances. Apo D has been also suggested to play a local paracrine role in the distribution of cholesterol between cells and HDL or in the steroid hormone metabolism in the mammary tissue. Finally, a potential apoD participation in neuronal repair mechanisms has been proposed.
Physiology/Pathology	Increased concentrations have been reported in mammary tissue from women with benign and malignant breast diseases, in prostate carcinomas and during regeneration of peripheral nervous tissue. Apo D expression in

	breast cancer cells is induced by retinoids, interleukin-1a, androgens and glucocorticoids, and down-regulated by estrogens and IL-6. Stimulation of apo D secretion by retinoids and steroids is concomitant with growth arrest and cell differentiation in breast cancer cells. Plasma concentration of apo D is significantly reduced in individuals with familial LCAT deficiency.
Degradation	Unknown
Genetics/Abnormalities	The gene is on chromosome 3 at the p14.2-qter region, and is 12 kb long with six exons, the last five of which contain all information coding for the apo D protein, including the 20 aa signal sequence. Two common DNA polymorphisms have been reported at the apoD locus. A rare variant of apoD phenotyped by isoelectric focusing and immunoblotting has been described in the black population.
Half-life	Unknown
Concentration	Serum: 60-120mg/L; breast cyst fluid: 6-25g/L
Isolation Method	(a) combination of hydroxylapatite and Sephadex G-100 chromatography from HDL fraction of pooled plasma. (b) size exclusion HPLC from breast cyst fluid.
Amino Acid Sequence	Apo D has no significant aa sequence similarity with other apolipoproteins, but exhibits homology to members of the lipocalin family of proteins. The two structural motifs that are conserved in lipocalins are present in the apo D aa sequence (Gly-X-Trp-Tyr and Thr-Asp-Tyr-Asp/Glu-X-Tyr). The N-terminal of the protein purified from plasma and breast cyst fluid is blocked by a pyroglutamic acid residue.
Disulfides/SH-Groups	Apo D contains five half-cystine residues which form two intramolecular disulfide bonds, one between residues 8 and 114, and the other between residues 41 and 165. Cys residue at position 116 forms an intermolecular disulfide bridge with Cys-6 of apoA-II.
General References	McConathy, W.J. and Alaupovic, P. *Methods Enzymol.* 1986, **128**:297-304. Balbín, M. et al. *Biochem. J.* 1990, **271**:803-807. Milne, R.W. et al. *Curr. Op. Lipidol.* 1993, **4**:100-106. Díez-Itza, I. et al. *Am. J. Pathol.* 1994, **144**:310-320. López-Boado, Y.S. et al. *J. Biol. Chem.* 1994, **269**:26871-26878.
Ref. for DNA/AA Sequences	Drayna, D. et al. *J. Biol. Chem.* 1986, **261**:16535-16539. Drayna, D. et al. *DNA* 1987, **6**:199-204.

Apolipoprotein E

Stanley C. Rall, Jr.

Synonyms	Arginine-rich apoprotein
Abbreviations	ApoE; ARP
Classifications	On lipoproteins, has α, β, or pre-β mobility
Description	Circulates in plasma as a component of various lipoproteins (chylomicron remnants, VLDL and their remnants, certain subclasses of HDL). Synthesized in a variety of tissues by specific cell types: hepatic synthesis accounts for 80-90% of plasma apoE; other major sites of synthesis include macrophages, astrocytes of the brain, smooth muscle cells. A single-chain protein of 299 aa, with 3 major polymorphic isoforms that differ at 2 sites, residues 112 and 158; apoE2 has Cys/Cys, apoE3 Cys/Arg, and apoE4 Arg/Arg at the 2 sites, respectively. These cysteines are the only ones in the protein. A portion of plasma apoE is glycosylated at Thr-194. Sialic acid content varies, mostly 1 or 2 residues per mole apoE.
Structure	Two distinct and independently folded structural domains: a globular N-terminal domain with high free energy of stabilization (ΔG = 8-12 kcal/mol) that contains the receptor binding site; an elongated C-terminal domain with low free energy of stabilization (ΔG = 4 kcal/mol) that contains the major lipid binding site. In the absence of lipid, self-associates (at most concentrations) through its C-terminal domain to form non-covalent stable tetramers. The monomeric N-terminal domain (residues 1-191) has been crystallized (in the absence of lipid) and its structure determined to 1.8 Å resolution; exists as a 4-helical bundle (residues 24-42, 54-81, 87-122, 130-164) with no β-sheet structures.
Molecular Weight	34,200 (aa sequence); 136,000 ± 8700 (for the stable, lipid-free tetramer, determined by sedimentation and diffusion).
Sedimentation Coeff.	5.3 S (for the stable, lipid-free tetramer).
Isoelectric Point	5.4 - 6.1
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	One of 2 ligands for the LDL receptor (the other being apoB). Also a ligand for the LDL receptor-related protein (LRP). Has a high affinity heparin binding site that is coincident with the receptor binding site (residues 130-160). Binds to lipid and lipoprotein surfaces, mainly through amphipathic α-helices in the C-terminal domain.

1. Installment 1992 (revised 1997)

Physiology/Pathology	Functions in lipid (especially cholesterol) transport/redistribution among tissues. Determinant for receptor-mediated catabolism of lipoproteins (chylomicron and VLDL remnants, certain HDL subclasses). Also serves a lipid transport function in the brain and functions in neurobiology and Alzheimer's Disease.
Degradation	Endocytosed and degraded intracellularly along with other components of lipoprotein particles, mainly by the liver.
Genetics/Abnormalities	Polymorphic protein resulting from multiple alleles at a single gene locus (chromosome 19, part of an apolipoprotein gene complex that includes apoCI and apoCII). The 3 major alleles for apoE4 (15%), apoE3 (77%), and apoE2 (8%) give rise to 6 phenotypes, 3 homozygous and 3 heterozygous. The different apoE polymorphs have significant effects on normal variation of plasma lipid levels. ApoE2 is defective in receptor binding (1-2% of normal), leading to accumulation of lipoproteins in circulation. ApoE2 and other rare receptor binding-defective variants are the causal molecular defect for the lipid disorder type III hyperlipoproteinemia (familial dysbetalipoproteinemia), which is characterized by accumulation of cholesterol-enriched chylomicron and VLDL remnants (β-VLDL) and a predisposition to accelerated cardiovascular disease in affected individuals. ApoE4 is linked to the occurrence of late-onset and sporadic Alzheimer's Disease. Appears to function as a susceptibility factor. ApoE4 may affect amyloid plaque formation and/or neurofibrillary tangle formation (the latter by altering the cytoskeletal apparatus). ApoE2 appears to be protective against Alzheimer's Disease.
Half-life	8-12 hrs (blood circulation, normal conditions)
Concentration	0.05 g/L (normal range 0.02-0.08 g/L); in type III hyperlipoproteinemic subjects, can reach 0.5 g/L or higher.
Isolation Method	Isolated from plasma by ultracentrifugation of d < 1.02 g/ml lipoproteins, followed by gel filtration chromatography (S-300) of delipidated lipoproteins and heparin affinity chromatography.
Amino Acid Sequence	Receptor binding site (residues 130-160) contains critical basic residues necessary for normal interaction with the LDL receptor: TEELRVRLASHLRKLRKRLLRDADDLQKRLA
Disulfides/SH-Groups	Intrachain disulfides (possible only for apoE2) have not been described. ApoE3 or apoE2 is found in plasma with up to 40% of the total as homodimers and/or heterodimers (with apoAII). Interchain disulfide oligomers (possible only for apoE2) have not been definitively demonstrated.
General References	Mahley, R.W. *Science* 1988, **240**:622-630. Davignon, J. et al. *Arteriosclerosis* 1988, **8**:1-21. Corder, E.H. et al. *Science* 1993, **261**:921-923. Mahley, R.W. and Rall, S.C., Jr. Type III hyperlipoproteinemia (dysbetalipoproteinemia): The role of apolipoprotein E in normal and abnormal lipoprotein metabolism. In: *The Metabolic and Molecular Basis of Inherited Disease*, 7th edition Scriver, C.R., et al. (eds.). McGraw-Hill, New York 1995; pp. 1953-1980.
Ref. for DNA/AA Sequences	Rall, S.C.,Jr. et al. *J. Biol. Chem.* 1982, **257**:4171-4178 (aa sequence). Paik, Y.-K. et al. *Proc. Natl. Acad. Sci. USA* 1985, **82**:3445-3449 (DNA sequence).

Aquaporin-1

Landon S. King and Peter Agre

Synonyms	CHIP28 (Channel-forming Integral Protein of 28 kD)
Abbreviations	AQP1
Classifications	None
Description	Aquaporin-1, the first identified molecular water channel, is abundantly expressed in red blood cells, renal proximal tubule epithelium, lung vascular endothelium, choroid plexus, anterior ciliary body of the eye, and biliary epithelium. Discovery of AQP1 made possible the subsequent identification of a growing family of water channel proteins: eight mammalian aquaporins are now known, and homologs have also been identified in plants, bacteria, and yeast.
Structure	The AQP1 monomer has 6 membrane spanning domains, with intracellular N- and C-termini (Fig. 1). The N- and C-halves of the molecule are highly related, but on opposite sides of the cell membrane. Loops B and E of the monomer fold into the bilayer to create the aqueous pore (Fig. 1). AQP1 is present as a tetramer in the membrane; one of the four subunits is glycosylated at Asn-42. All of the known aquaporins, including plant and bacterial homologs, have an Asn-Pro-Ala (or NPA) motif in both the N- and C-halves of the molecule (Fig. 1), with highly conserved aa flanking the NPA. Recognition of these highly conserved regions has formed the basis for identification of additional family members by degenerate cloning.
Molecular Weight	AQP1 has a molecular weight of 28 kD (SDS-PAGE). Glycosylated AQP1 has a molecular weight from 40-60 kD (SDS-PAGE).
Sedimentation Coefficient	5.7S
Isoelectric Point	Unknown
Extinction Coefficient	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Organic (PCMBS, per-chloromercuribenzene sulfonate) and inorganic mercurial compounds block the water permeability of AQP1 by reversibly binding to a cysteine at position 189.
Biological Functions	Studies of water permeability in renal proximal tubule and red blood cells strongly suggested the existence of water specific transmembrane channels. Injection of AQP1 complementary RNA into Xenopus oocytes (or reconstitution of purified protein into proteoliposomes) produces a 30-fold increase in osmotic water permeability, with an activation energy of approximately 3 kcal/mol, consistent with channel-mediated water movement. AQP1 is not permeated by protons, ions, or other small molecules such as urea or glycerol. AQP1 is present as a tetramer in the

membrane, but complementation studies suggest that each monomer conducts water independently.

Physiology/Pathology

The proposed physiologic roles for AQP1 arise from studies of distribution. Quantitative analysis of AQP1 expression in the renal proximal tubule, coupled with studies of the unit water conductance for AQP1, suggest that AQP1 may account for all of the water permeability of the proximal tubule. AQP1 in the choroid plexus of the brain and anterior ciliary body of the eye participates in generation of cerebrospinal fluid and aqueous humor, respectively. In lung, AQP1 is expressed from late gestation in endothelial cells of the vascular plexus surrounding airways, as well as the visceral pleura; expression is sustained at high levels in adults. AQP1 may participate in generation or reabsorption of pulmonary edema and pleural effusions, as well as perinatal lung water clearance. Each of the other mammalian aquaporins has a distinct distribution which suggests specific functions. The reader is referred to recent reviews for further details.

Degradation

The degradation pathway for AQP1 is unknown. There are no known biologically active degradation products.

Genetics/Abnormalities

AQP1 is located at human chromosome 7p14. Identification of the Colton blood group system on AQP1 allowed discovery of rare AQP1 null humans (6 kindreds in the world to date), who appear to be phenotypically normal. Recent demonstration that AQP1 knockout mice have severe urinary concentrating defects supports the idea that the AQP1 null humans have as yet unidentified forms of compensation. In contrast to mutations in AQP1, mutations in AQP2, the water channel expressed in the collecting duct of the kidney, lead to congenital nephrogenic diabetes insipidus. Recently it has been shown that mutations in MIP (AQP0) lead to congenital cataracts in mice.

Half-life

Not applicable

Concentration

AQP1 is an integral membrane protein, with no free or circulating form. Human red blood cells express approximately 200,000 AQP1 molecules per cell. AQP1 makes up one-fourth of the membrane protein in the renal proximal tubule epithelium. AQP1 has not been quantitated at other sites.

Isolation Methods

Red blood cells provide an abundant source of AQP1. KI stripped red cell membrane vesicles can be extracted with 1% N-lauroylsarcosine, and solubilized in 4% Triton X-100. The solubilized material is filtered and loaded onto a POROS Q/F column by FPLC. The column is washed, and eluted with a NaCl gradient. Fractions from the column can be examined for purified AQP1.

Amino Acid Sequence

```
MASEFKKKLF WRAVVAEFLA TTLFVFISIG SALGFKYPVG
NNQTAVQDNV KVSLAFGLSI ATLAQSVGHI SGAHLNPAVT
LGLLLSCQIS IFRALMYIIA QCVGAIVATA ILSGITSSLT
GNSLGRNDLA DGVNSGQGLG IEIIGTLQLV LCVLATTDRR
RRDLGGSAPL AIGLSVALGH LLAIDYTGCG INPARSFGSA
VITHNFSNHW IFWVGPFIGG ALAVLIYDFI LAPRSSDLTD
RVKVWTSGQV EEYDLDADDI NSRVEMKPK
```

Disulfides/SH-Groups

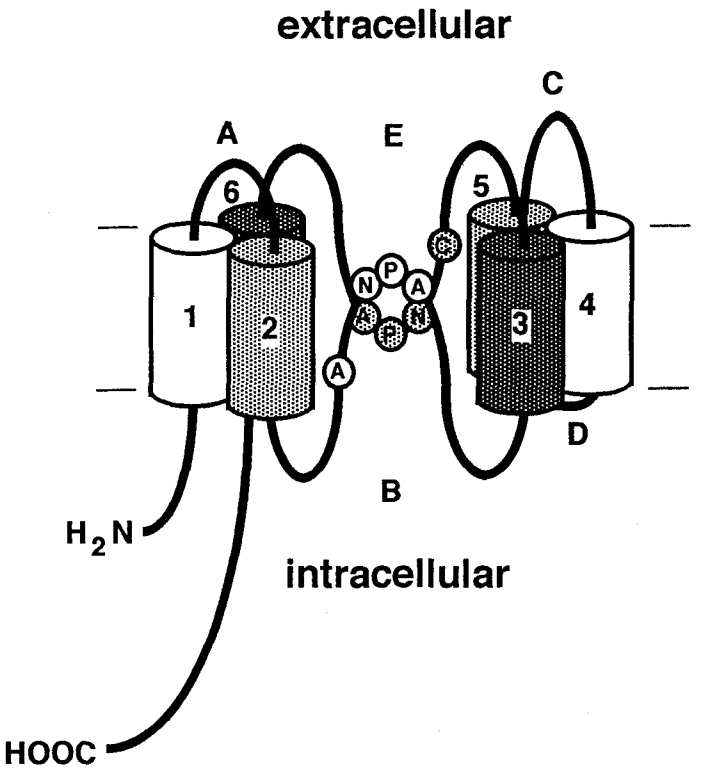
AQP1 contains four cysteines, in positions 87, 102, 152, and 189. AQP1-mediated water permeability is inhibited by interaction of mercurial compounds with Cys-189. There are no disulfide bonds between subunits, however the presence of intrasubunit disulfides in unknown.

General References

Preston, G.M. et al. *Science* 1992, **256**:385-387.
Smith, B.L. and Agre, P. *J. Biol. Chem.* 1991, **266**:6407-15.
King, L.S. and Agre, P. *Ann. Rev. Physiol.* 1996, **58**:619-648.
Lee, M.D. et al. *Medicine.* 1997, 76:141-56.
Agre, P. et al. *Methods Enzymology.* 1998, in press.
Walz, T. et al. *Nature* 1997, **387**:624-27.

Ref. for DNA/AA Sequences

The human AQP1 sequence is present in the GenBank data base, accession number M77829.



Arrestin

Krzysztof Palczewski and Larry A. Donoso

Synonyms	S-antigen; 48 kDa protein
Abbreviations	Arr; 48K
Classifications	Soluble protein in retinal extracts
Description	Arrestin is a soluble protein expressed predominantly in photoreceptor cells in the retina and in the pineal gland. The mRNA coding for this protein has been detected in brain, lens and peripheral blood leucocytes. Its homologs (β-arrestins) have been identified in many different tissues. Several splice variants of arrestin were identified.
Structure	Arrestin is a monomeric protein. Analysis of CD spectra revealed that arrestin has virtually no α-helical structure, approximately 40% β-structure, about 18% β-turns and 40% other structure. Based upon biochemical studies, it has been proposed that the C-terminal region of arrestin (20-30 aa) is exposed, especially upon binding to light-activated and phosphorylated rhodopsin. Due to internal aa sequences homology and the pattern of limited proteolysis, it has been proposed that arrestin is composed of two similar domains. The molecule has not yet been crystallized in its native form.
Molecular Weight	45,045 Da: human, based on aa sequence; predicted mass for bovine arrestin obtained from its sequence: 45,275 Da. Based on a mass spectrometry analysis, the molecular weight of the protein was verified to be 45,317 Da (due to acetylation of the N-terminal Met residue).
Sedimentation Coeff.	3.6
Isoelectric Point	5 - 5.9, at least seven different forms
Extinction Coeff.	6.38 (278nm, 1%, 1cm), absorption maximum for arrestin is at 278nm.
Enzyme Activity	None, exhibits light-dependent binding to phosphorylated rhodopsin.
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Inhibitors of phosphorylated photolyzed rhodopsin/arrestin interaction: heparin, inositol hexakisphosphate.
Biological Functions	The binding of arrestin to photolyzed and phosphorylated rhodopsin results in blocking (or arresting) G-protein (transducin) activation. Arrestin also acts in the phototransduction process by blocking rapid dephosphorylation of rhodopsin until photolyzed rhodopsin decays. The dissociation of arrestin from the photoreceptor protein occurs when photolyzed rhodopsin is deactivated by removal of the photoisomarized chromophore all-*trans*-retinal.
Physiology/Pathology	Arrestin is a highly pathogenic retinal autoantigen. Injection of microgram quantities of purified protein into susceptible animal strains induces a severe CD4 + T-cell mediated inflammatory response in the retina and

	pineal gland. This disease process mimics several conditions of ocular inflammation observed in man.
Degradation	Proteolysis at the C-terminus of arrestin leads to tight association with the disk membranes of the rod outer segment. Hypothetically, this could lead to transport of arrestin with shed rod outer segment discs to pigment epithelium cells and its subsequent degradation.
Genetics/Abnormalities	The gene corresponding to human arrestin is located on chromosome 2 at 2q24-q37 and is approximately 50 kb long. The gene contains 16 exons and 15 introns.
Half-life	Unknown
Concentration	100 - 500 μM in rod cells.
Isolation Method	Arrestin can be isolated from retina using one of the three general methods: (1) conventional chromotography on DEAE-Cellulose followed by hydroxyapatite, Phenyl-Sepharose and Mono Q chromatographies; (2) by taking advantage of the light dependent binding of arrestin to phosphorylated photolyzed rhodopsin followed by Mono Q chromatography and (3) affinity chromatography on Heparin-Sepharose using inositol hexakisphosphate for elution.
Amino Acid Sequence	Proposed binding site on arrestin to phosphorylated C-terminal region of rhodopsin 167-KIPKKSSVRYLIRSVQHAPL. Major pathogenic sites in arrestin: Retinal arrestins, including human, contain at least four major pathogenic sites. Two sites are located in the mid-portion of the molecule at aa residues 286 to 297 and 303 to 314, and two sites are located at the C-terminus of the molecule at 333 to 352 and 352 to 364.
Disulfides/SH-Groups	Three free SH-groups.
General References	Wacker et al. *J. Immunol.* 1977, **119**:1949-1958. Shinohara et al. *Progress in Retinal Res.* 1989, **8**:51-66. Wilden et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:1174-1178. Hofmann et al. *J. Biol. Chem.* 1992, **267**:15701-15706. Palczewski et al. *Neuron* 1992, **8**: 117-126.
Ref. for DNA/AA Sequences	Shinohara et al. *Proc. Natl. Acad. Sci. USA* 1987, **84**:6975-6979. Yamaki et al. *FEBS Lett.* 1988, **234**:39-43. Lohse et al. *Science* 1990, **248**:1547-1550. Attramadal et al. *J. Biol. Chem.* 1992, **267**:17882-17890. Craft et al. *J. Biol. Chem.* 1994, **269**:4613-4619.

Arylamine N-acetyltransferase

Denis M. Grant and Urs A. Meyer

Synonyms	Isoniazid transacetylase, O-acetyltransferase, N,O-acetyltransferase
Abbreviations	NAT, OAT, N,O-AT
Classifications	EC 2.3.1.5
Description	A family of three soluble cytoplasmic proteins, NAT1, NAT2A and NAT2B, each composed of a single polypeptide chain of 290 aa. NAT1 and NAT2A are the primary products of the *NAT1* and *NAT2* gene loci, respectively, and share 81% aa sequence identity; NAT2B is thought to arise from NAT2A by as yet unidentified post-translational event(s) which alter charge (elution profile on anion exchange columns) but not electrophoretic mobility on SDS-PAGE. The NAT2A/B isoforms show tissue-selective expression in liver hepatocytes and duodenal mucosa, while NAT1 is likely expressed in most tissues.
Structure	Unknown
Molecular Weight	33,898: NAT1 (deduced); 33,000 (SDS-PAGE) 33,542: NAT2 (deduced); 31,000 (SDS-PAGE)
Sedimentation Coeff.	Unknown
Isoelectric Point	4.8–4.9: NAT2; NAT1 unknown
Extinction Coeff.	Unknown
Enzyme Activity	Each enzyme catalyses the transfer of acetate from acetyl CoA (CoAS-$COCH_3$) to nitrogen on aromatic amine or oxygen on hydroxylamine xenobiotic substrates within the chemical classes listed below. The reaction follows a classical two-step, substituted-enzyme catalytic mechanism as illustrated below for the primary aromatic amine Ar-NH_2: 1) NAT + CoAS-$COCH_3$ → NAT-S-$COCH_3$ + CoASH 2) NAT-S-$COCH_3$ + Ar-NH_2 → NAT + Ar-NH-$COCH_3$ These enzymes may also mediate intramolecular N,O-acetyltransfer, converting hydroxamic acids to acetoxy esters.
Coenzymes/Cofactors	Acetyl CoA is the major physiologic acetyl donor; acetylthiocholine and N-diacetylcysteamine are much less efficient.
Substrates	1) primary carbocyclic arylamines, e.g. sulfamethazine, p-aminobenzoic acid 2) primary heterocyclic arylamines, e.g. Glu-P-2 (2-aminodipyridol [1,2-a:3',2'-d]imidazole) 3) primary aromatic hydrazines, e.g. isoniazid (isonicotinic acid hydrazide) 4) N-hydroxyarylamines, e.g. N-hydroxy-2-aminofluorene.
Inhibitors	CoASH (product inhibition) Sulfhydryl group inhibitors: N-ethylmaleimide, p-chloromercuribenzoate Active-site directed covalent inhibitors: iodoacetic acid, bromoacetanilide.

Biological Functions

Enzymes of xenobiotic biotransformation, converting lipophilic drugs and environmental chemicals (including several carcinogens) into more water-soluble metabolites for subsequent renal excretion. Since no endogenous substrates have yet been identified, it is possible that the NAT proteins have evolved specifically for the purpose of protecting the organism from chemical insults in the environment.

Physiology/Pathology

Important in determining the elimination rates of many therapeutically important drugs and aromatic amine carcinogens. A genetic polymorphism of NAT2 may result in increased incidence of dose-related side-effects from drugs such as isoniazid, procainamide, hydralazine and salicylazosulfapyridine in either 'slow' or 'rapid' acetylator subjects. Slow acetylators are also at higher risk for the occurrence of bladder cancer due to environmental exposure to benzidine, 4-aminobiphenyl and β-naphthylamine. In addition, NAT enzymes may directly contribute to the carcinogenicity of such chemicals due to their ability to 'metabolically activate' (via O-acetylation or N,O-acetyltransfer) hydroxylamine and hydroxamic acid metabolites to unstable acetoxy esters that decompose to electrophilic, DNA-binding arylnitrenium ions.

Degradation

Unknown

Genetics/Abnormalities

NAT1 and *NAT2* genes are both located on chromosome 8 (pter-q11), but separated by at least 25kb. Both genes have 870 bp intronless protein coding regions. The entire transcript for NAT1 is encoded on a single exon; that for NAT2 arises from a 5' non-coding exon located 8 kb upstream from a second exon containing coding and 3' regions. A related pseudogene (*NATP*) contains multiple internal termination signals. A genetic polymorphism at the *NAT2* locus segregates individuals as 'rapid' and 'slow' acetylators, and is caused by point mutations producing 8 presently known variant alleles. The 4 most commonly occurring alleles produce aa variation at the following positions:

	114	197	268	286
R1	Ile	Arg	Lys	Gly
S1A	Thr	Arg	Arg	Gly
S2	Ile	Gln	Lys	Gly
S3	Ile	Arg	Lys	Glu

The Arg-197 → Gln change in variant S2 produces a protein with reduced stability, while Ile-114 → Thr in S1A does not affect stability or catalytic activity. The effect of other aa changes has not been determined.

Half-life

3–4 hr: NAT1; 30–40 hr: NAT2 (in vitro, 37 °C)

Concentration

The NAT2A/B isoforms account for less than 0.01 % of soluble intracellular protein in hepatocytes and intestinal epithelium from genetically rapid acetylators. NAT1 is expressed at 10–20-fold lower levels than NAT2A/B in rapid acetylator liver but probably has a much wider tissue distribution.

Isolation Method

NAT2: Ammonium sulfate fractionation (0–55% cut) of human liver cytosol, pH-shift precipitation (pH 5.5 supernatant), followed by anion exchange chromatography on DEAE-Sephacel, hydroxylapatite chromatography, gel filtration on Sephacryl S-200 and preparative SDS-PAGE to isolate denatured immunogenic NAT2 protein. NAT1 is not sufficiently stable to survive extensive purification procedures.

Amino Acid Sequence

All mammalian NAT proteins share 65–80% aa sequence identity; chicken liver NAT is 45–50% identical to the mammalian enzymes; and the N-terminal 170 aa of S.typhimurium O-acetyltransferase (OAT) shares 25–50% identity with the mammalian enzymes. Site-directed mutagenesis studies with human NAT2 and S.typhimurium OAT have identified a conserved Cys residue at position 68 that directly participates in the catalytic mechanism of acetyl transfer from acetyl CoA to acceptor amines. The region surrounding this residue (shown in bold) is shown below for all acetyltransferases so far cloned and characterized.

	61 * 75
Human NAT1	RRNRGGW**C**LQVNHLL
Human NAT2	RRNRGGW**C**LQVNQLL
Rabbit NAT1	RRNRGGW**C**LQVNYLL
Rabbit NAT2	RRNRGGW**C**LQVNYLL
Mouse NAT1	RKKRGGW**C**LQVNHLL
Mouse NAT2	RKKRGGW**C**LQVNHLL
Hamster-M	RKKRGGW**C**LQVNHLL
Chicken NAT3	RKKRGGW**C**MENNHLL
Chicken NAT10	HKKRGGW**C**MENNQLL
Chicken NATL	KKKRGGW**C**METNYLL
S.typhimurium OAT	YARRGGY**C**FELNGLF

Disulfides/SH-Groups

None; 5 free SH-residues, one of which (Cys^{68}) is contained in the active site and participates in acetyl transfer.

General References

Grant, D. M. *Pharmacogenetics* 1993, **3:** 45–50.
Grant, D. M. et al. *Xenobiotica* 1992, **22:** 1073–1081.
Grant, D. M. and Spielberg, S. P. Genetic regulation of drug metabolism. In: *Fetal and Neonatal Physiology*, Polin, R. A. and Fox, W. W. (eds.), W. B. Saunders Co., Philadelphia 1992, pp. 141–148.
Evans, D. A. P. *Pharmacol. Ther.* 1989, **42:** 157–234.
Weber, W. W. *The acetylator Genes and Drug Response*, Oxford University Press, New York, 1987.

Ref. for DNA/AA Sequences

Grant, D. M. et al. *Nucl. Acids Res.* 1989, **17:** 3978.
Blum, M. et al. *DNA Cell Biol.* 1990, **9:** 193–203.
Ohsako, S. and Deguchi, T. *J. Biol. Chem.* 1990, **265:** 4630–4634.
Deguchi, T. et al. *J. Biol. Chem.* 1990, **265:** 12757–12760.
Ebisawa, T. and Deguchi, T. *Biochem. Biophys. Res. Comm.* 1991, **177:** 1252–1257.
Blum, M. et al. *Proc. Natl. Acad. Sci. USA* 1991, **88:** 5237–5241.
Hickman, D. and Sim, E. *Biochem. Pharmacol.* 1991, **42:** 1007–1014.
Vatsis, K. P. et al. *Proc. Natl. Acad. Sci. USA* 1991, **88:** 6333–6337.

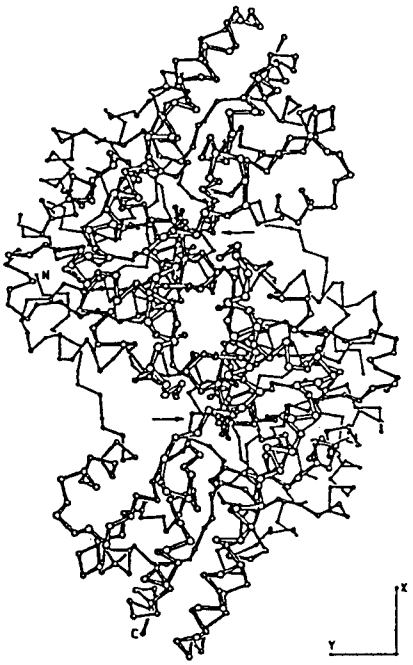
Aspartate aminotransferase

Erika Sandmeier and Philipp Christen

Synonyms	Aspartate transaminase; Glutamate-oxalacetate transaminase; Glutamic-oxaloacetic transaminase; Glutamic-aspartic transaminase; L-Aspartate: 2-oxoglutarate aminotransferase; Glutamate-α-ketoglutarate transaminase.
Abbreviations	AspAT; AST; ASAT; GOT; AATase; AAT
Classifications	EC 2.6.1.1
Description	An ubiquitous pyridoxal-5'-phosphate (vitamin-B_6)-dependent enzyme catalyzing transamination of dicarboxylic amino and oxo acids. Two genetically independent isoenzymes in eucaryotic cells: cytosolic AspAT and mitochondrial AspAT (in mitochondrial matrix).
Structure	A dimer of identical subunits, each consisting of a large pyridoxal-5'-phosphate binding domain, a small domain and an N-terminal segment that extends in front of the active site cleft and forms an extra contact with the adjacent subunit. Spatial structure known for AspAT of E. coli, as well as cytosolic and mitochondrial AspAT of chicken and pig.
Molecular Weight	2 x 45 kDa (cytosolic: 2 x 412; mitochondrial:2 x 401 aa residues).
Sedimentation Coeff.	Cytosolic 5.4 S; mitochondrial: 4.7 S (bovine)
Isoelectric Point	Cytosolic 5.2-5.6; mitochondrial: 9.6
Extinction Coeff.	14.0 (280nm, 1%, 1cm); 7.0 x 10^4 $M^{-1}cm^{-1}$ (280nm, subunit concentration). Values for pig isoenzymes but assumed to apply also to isoenzymes from other sources. (See also Coenzyme).
Enzyme Activity	Catalyzes transfer of amino group from dicarboxylic or aromatic aa to corresponding oxo acid (transamination): L-Aspartate + 2-oxoglutarate Δ oxalacetate + L-glutamate. Specific activity (30°C): cytosolic: 198 U/mg; mitochondrial: 208 U/mg
Coenzymes/Cofactors	Pyridoxal 5'-phosphate, a derivative of vitamin B_6, covalent Schiff base linkage to active-site lysine residue. Transamination follows double displacement mechanism with coenzyme shuttling between the pyridoxal (γ_{max} approx. 360 nm) and the pyridoxamine form (γ_{max} approx. 330 nm).
Substrates	Aspartate, glutamate, cysteine sulfinate, phenylalanine, tryptophan, tyrosine, and the corresponding 2-oxo acids.
Inhibitors	Dicarboxylic acids as competitive inhibitors, e.g. maleate, glutarate, adipate, o-phthalate. Affinity labels, e.g. 3-bromopropionate, iodoacetate. Mechanism-based inhibitors, e.g. β-chloro-L-alanine, propargylglycine, β-cyano-L-alanine, vinylglycine, L-serine-O-sulfate, gostatin. Carbonyl reagents, e.g. hydroxylamine, hydrazine, isoniazid, inhibit by reacting with pyridoxal-5'-phosphate.
Biological Functions	Important role in aa metabolism, linking it with the urea and tricarboxylic acid cycles. In addition, the two isoenzymes are integrated into the maleate-aspartate shuttle, a pathway transporting reducing equivalents from cystol into mitochondria.

1. Installment 1992 (revised 1997)

Physiology/Pathology	No inherited AspAT deficiencies known. Determination of AspAT activity in human serum is a valuable diagnostic aid in diseases of liver and biliary tract as well as heart and muscle; in particular it is used for assessing the severity of liver disorders.
Degradation	Unknown
Genetics/Abnormalities	Both cystosolic AspAT and the higher molecular mass precursor of mitochondrial AspAT are coded for by nuclear genes and synthesized on cytosolic ribosomes. The gene for cytosolic AspAT is situated on chromosome 10; mitochondrial AspAT is coded for by a multigene family (functional gene on chromosome 16).
Half-life	2.7 days for both isoenzymes (chicken liver).
Concentration	Serum, male: < 18 U/L (96% cytosolic AspAT), female: < 15 U/L. Liver 96 U (25°C), heart 52, skeletal muscle 36 (ratio cytosolic/mitochondrial AspAT 0.25 to 4 depending on tissue), erythrocytes 0.8 (cytosolic AspAT only), all U/g wet weight.
Isolation Method	Purification of cytosolic and mitochondrial AspAT from human liver by heat treatment, ammonium sulfate precipitation, anion and cation exchange chromatographies, affinity chromatography, and isoelectric focusing (Leung, F.Y. and Henderson, A.R., *Clin. Chem.* 1981, **27**:232-238).
Amino Acid Sequence	49% identity between cytosolic and mitochondrial isoenzyme of human. Homologous with all other aminotransferases.
Disulfides/SH-Groups	None
General References	Christen, P. and Metzler, D.E. (eds.) Transaminases. John Wiley and Sons, New York, 1985.
Ref. for DNA/AA Sequences	Bousquet-Lemercier, B. et al. *Biochemistry* 1990, **29**:5293-5399 AC: M37400. Doyle, J.M. et al. *Biochem. J.* 1990, **270**:651-657. AC:P17174. Pol, S. et al. *Biochem. Biophys. Res. Commun.* 1988, **157**:1309-1315. AC: M22632. Martini, F. et al. *Biochim. Biophys. Acta* 1985, **832**:46-51. AC P0050.



α-Carbon chain model of mitochondrical AspAT of chicken. The dimeric enzyme is viewed along its molecular twofold symmetry axis. The lower subunit is depicted in the open conformation, the upper subunit in the closed conformation which is assumed upon binding of substrates or inhibitors. Arrows point to the active sites. In the lower subunit, N- and C-terminus are indicated. (Courtesy of J. N. Jansonius, Biocenter, University of Basel.)

Band 3 protein (AE1)

Hermann Passow

Synonyms	Anion exchange protein, capnophorin
Abbreviations	AE1
Classifications	
Description	AE1 is the most abundant membrane protein in red blood cells. On SDS-PAGE it forms the third major band from the top, hence its name. Each monomer (MW ≃ 96,000) consists of N-terminal hydrophilic and C-terminal hydrophobic domain. The former protrudes into the cytosol, the latter is associated with lipid bilayer. The hydrophilic domain is N-glycosylated at Asn-642 with a branched aminoglycan which, in the adult red cell, acts as the I antigen. Different band 3 molecules may carry different numbers of repeating N-acetyl-lactoseamine units. This gives rise to the typical diffuse appeareance of band 3 on PAGE. In the plasma membrane the molecule exists in the form of monomers, dimers and tetramers. In solutions of nonionic detergents, the three forms are at equilibrium with each other. However, under certain conditions stable dimers may also occur. AE proteins (see below) are not only present in red cells but also in kidney, testis, heart, lung, liver, brain, gastric cells, placenta and other types of cells.
Structure	The human band 3 protein consists of 911 aa residues. Both the C- and N-terminal reside at the inner membrane surface. The tertiary structure is still unknown. Hydrophobicity plots and studies with proteolytic enzymes, antibodies, and non penetrating agents suggest that the hydrophilic domain comprises the first 400 N-terminal aa residues and that the hydrophobic domain of the peptide chain crosses the lipid bilayer 10–14 times. Two-dimensional crystals with a resolution of about 20 Å have been obtained from band 3 as a whole and from its hydrophobic domain. The analysis of the structural information yielded somewhat contradictory results.
Molecular Weight	96 kDa ≃ : monomer; 5.5 kDa carbohydrate side chain; ≃ 42 kDa: hydrophilic domain; ≃ 55 kDa: hydrophobic domain.
Sedimentation Coeff.	7 S (non-ionic det.); 4–12 S (monomer-tetramer)
Isoelectric Point	≃ 5.5
Extinction Coeff.	$\Delta A = OD_{280} - 1.95 \times OD_{320} = 7.7$, where ΔA represents the extinction coefficient as measured in a 1 % solution of band 3 in non-ionic detergent as corrected for light scattering. OD_{280}, OD_{320} refer to the optical densities measured at 280 and 320 nm.
Enzyme Activity	None
Coenzymes/Cofactors	Not required
Substrates	Physiological substrates: chloride and bicarbonate. The protein also transports a large variety of other inorganic and organic aninon species. Half-saturation concentration for chloride (equilibrium exchange at equal chloride concentration inside and outside the cells): 38.5 mM, 37 °C. Divalent anions are cotransported with the proton, which leads to pH equilibration between cells and medium.

Inhibitors	Chemical reagents that react more or less specifically with Lys, Arg, His and Glu or Asp residues. The most specific inhibitors include non-penetrating stilbene disulfonate derivatives like the covalently binding DIDS (4,4'-diisothiocyanostilbene-2,2'-disulfonate), H2DIDS (4,4'-diisothiocyano dihydrostilbene-2,2'-disulfonate), SITS (4-isothiocyano-4'-acetamido stilbene-2,2'-disulfonate) and the non-covalently binding DNDS (4,4'-dinitro stilbene-2,2'-disulfonate) or DBDS (4,4'-dibenzoyl stilbene-2,2'-disulfonate). The latter is fluorescent and undergoes considerable fluorescence enhancement upon binding to the transporter. The site of action of the following inhibitors has been identified: Eosine maleimide (Lys-430), H_2DIDS, 1-fluoro-2,4-dinitrobenzene (Lys-539), pyridoxal-5'-phosphate (Lys-539 and Lys-851), phenylglyoxal (Arg-490, Arg-730), Woodwards reagent K (Glu-681), diethylpyrocarbonate (His-734). Other His required for normal activity: His-703, His-819, His-834. Lys-340; Glu-681 and Lys-851 seem to be accesible from either membrane surface.
Biological Functions	1. The hydrophobic domain mediates an electrically neutral anion exchange across the plasma membrane of the red blood cell. 2. It is stipulated that aggregation or partial degradation of band 3 plays a decisive role in the recognition and elimination of senescent red blood cells in the circulating blood. 3. The hydrophilic domain serves as an anchor (a) for constituents of the membrane skeleton, notably ankyrin, band 4.1 and 4.2, (b) for certain glycolytic enzymes (phosphofructokinase, glyceraldehyde-phosphate dehydrogenase, aldolase) and (c) hemoglobin, hemichromes and catalase. Binding is pH-dependent and, in the case of the glycolytic enzymes, dependent on the phosphorylation of Tyr-8, and to a minor extent on the phosphorylation of other tyrosine residues in the hydrophilic domain. Regardless of its state of association, each protomer acts as an independently operating unit. Turnover number for transport of chloride at Donnan equilibrium is 5×10^4/band 3 monomer/sec at pH 7.4, 38 °C. Activation enthalpy: 30 Kcal/mole or 20 Kcal/mole for the temperature ranges 0–15 °C and 15–38 °C, respectively.
Physiology/Pathology	There exist naturally occuring mutants. The Memphis variants 1 and 2 occur in 6–7% of randomly collected blood samples. Variant 1 represents the single point mutation K56E. No pathological consequences are known. In Southeast Asian Hereditary Ovalocytosis, in addition to the mutation K56E, aa residues 400–408 are deleted. This is asssociated with changes of mechanical properties of the red cells, loss of anion transport function and increased resistance against infection with Plasmodium falciparum and other parasites.
Degradation	Degradation by phagocytosis after removal of the senescent red blood cells from the circulation. Perhaps some degradation of AE1 already occurs while the red blood cells are still in the circulating blood.
Genetics/Abnormalities	The erythroid band 3 protein AE1 is the product of one gene of a family of three closely related genes which encode three distinct classes of anion exchange proteins (AE1 – AE3). AE1 and AE2 are products of single copy genes located in chromosomes 17q21-qter and 7q35–7q36, respectively. The entire AE1 gene of man and mouse has been cloned and sequenced. The structure of the genes in the two species is essentially similar (17 kb; 20 exons). The products of the other genes (called AE2 and AE3) all possess the capacity of transporting anions and of participating in pH equilibration. Their hydrophobic domains show a rather high sequence homology with the hydrophobic domains of AE1 ($\simeq$ 60%). The hydrophilic domains have higher molecular weights than those of AE1. They show a lower sequence homology amongst each other and with AE1. In the basolateral membranes of subset of collecting duct cells of the kidney, an AE1 species exists which lacks the N-terminal 65 aa residues, corresponding the lack of exons 1–3.

Half-life	See Degradation
Concentration	$\simeq 1 \times 10^6$ protomers per human red blood cell (AE1).
Isolation Method	Native band 3, suitable for reconstitution in lipid bilayers can be obtained by solublization and chromatographic purification in solutions of non-ionic detergents (Triton X-100, $C_{12}E_9$).
Amino Acid Sequence	The aa sequence of AE1 from the red cells of man, mouse, rat, chicken and trout are known. the sequence homologies in the hydrophilic domains amount to about 40%, in the hydrophobic domains to about 65%. The lengths of the peptide chains range from 848 to 929 aa residues for AE1 and from 1227 to 1237 residues for AE2 and AE3. To date at least six species of AE2 and two species of AE3 have been cloned and sequenced.
Disulfides/SH-Groups	Crosslinking of AE1 monomers to dimers occurs by oxidation of two out of three SH groups (Cys-201, Cys-317, Cys-479) in the hydrophilic domain.
General References	Alper, S. L. *Annu. Rev. Physiol.* 1991, **53:** 549–564. Bamberg, E. and Passow, H. (eds.) The band 3 proteins: Anion transporters, binding proteins and senescent antigens. *Progress in Cell Research* 1992, **2:** 1–358. Cabantchik, Z. I. et al. *Biochim. Biophys. Acta* 1978, **515:** 239–302. Fukuda, M. et al. *J. Biol. Chem.* 1984, **259:** 8260–8273. Jennings, M. L. *Annu. Rev. Biophys. Biophys. Chem.* 1989, **18:** 397–430. Jennings, M. L. In: *The Kidney: Physiology and Pathology*, Seldin, D. W. and Giebisch, G. (eds.) Raven Press 1992, pp. 503–535. Knauf, P. A. In: *The Red Cell Membrane*, Raess, B. U. and Tunnicliff, G. (eds.), Humana Pres, Clifton, NJ 1989. Passow, H. *Rev. Physiol. Biochem. Pharmacol.* 1986, **103:** 62–223. Salhany, J. M. *Erythrocyte Band 3 Protein*, CRC Press, Boca Raton, 1990. Tanner, M. J. A. *Seminars in Hematology* 1993, **30:** 34–57. Dolder, M. et al. *J. Mol. Biol.* 1993, **231:** 119–132 (crystal structure). Wang, D. N. et al. *EMBO J.* 1993, **12:** 2233–2239.
Ref. for DNA/AA Sequences	AE1, AE2 and AE3: Wood, P. G. *Progress in Cell Research* 1992, **2:** 325–352.

Beta 1,4-galactosyltransferase

Eric G. Berger and Martine Malissard

Synonyms	Glycoprotein 4-beta-galactosyltransferase; Thyroid galactosyltransferase; UDPgalactose-glycoprotein:galactosyltransferase; Lactose synthase A protein
Abbreviations	gal-T; GT; beta1,4-gal-T
Classifications	EC 2.4.1.38; part of EC 2.4.1.22
Description	Gal-T is a Golgi-associated glycosyltransferase located in trans Golgi cisternae and possibly on cell surfaces. The enzyme is a type II membrane protein consisting of a single polypeptide with 397 (placenta) or 398 (HeLa cells) aa. Gal-T is solubilized in vivo and released from cells in an enzymatically active form truncated between Leu-38 and Ala-44 (milk) or at Arg-77 (placenta). The peptide contains a single N-glycosylation site at Asn-113 and various O-glycosylation sites. Glycans bear blood group genetic markers.
Structure	Unknown Amino acids of the binding sites: The aa residues involved in donor or acceptor binding sites have been identified by experiments using chemical labelling mehtods or site directed mutagenesis. Thus Tyr-284, Tyr-309 and Trp-310 are involved in the binding of N-acetylglucosamine. The following residues have been shown to be involved in the binding of UDP-Galactose; Tyr-309, Cys-340, Asn-307, Asn-308 and the tetrapeptide Asp-Lys-Lys-Asn (348-352).
Molecular Weight	43,8 kDa (full length, non-glycosylated, as predicted from the cDNA by converting 100 aa to 11 kDa); 45/47 kDa (full length, N-glycosylated forms as determined in HeLa cells by metabolic labeling after 10 min pulse, followed by immunoprecipitation/SDS-PAGE/fluorography; 49 kDa (soluble protein isolated from serum, determined by SDS-PAGE); 55 kDa (soluble protein isolated from milk, determined by SDS-PAGE); 106 kDa (dimeric form of the soluble protein isolated from milk, determined by light scattering experiments).
Sedimentation Coeff.	Unknown
Isoelectric Point	3.9 - 7.2, 4 major forms: 4.90, 5.18, 5.52, 6.44
Extinction Coeff.	Unknown
Enzyme Activity	UDPgalactose + N-acetyl-D-glucosaminyl-glycopeptide (1) or glucose (2) → UDP + 4-beta-G-galactosyl-N-acetyl-D-glucosaminyl-glycopeptide (1) or lactose (2)•glycopeptides, glycolipids or proteoglycans (keratan) bearing terminal N-acetylglucosamine residues may serve as acceptor substrates. Glucosyl-ceramide is not an acceptor substrate.
Coenzymes/Cofactors	Mn^{2+} (involved in binding of donor substrate UDPgalactose). Alpha-lactalbumin or B protein of the lactose synthase complex: promotes reaction with substrate 2 and inhibits reaction with substrate 1; Alpha-lactalbumin is only expressed in the lactating mammary gland and has been designated

	"modifier" protein; in its presence the K_m for glucose is lowered from the molar to the millimolar range.
Substrates	Donor substrate: UDPgalactose Acceptor substrate: N-acetyl-D-glucosaminyl-R (1) or glucose (2).
Inhibitors	Competitive inhibitors (Ki in parenthesis, in mmol/L): MnUTP (0.57), MnUDP (0.043), UMP (0.82), uridine (13.8).
Biological Functions	Gal-T is involved in the biosynthesis of complex type N-glycans and in elongation of core 2 O-glycans of glycoproteins; it also participates in the biosynthesis of repeating galactose beta 1,4 N-acetyl-D-glucosaminyl structures found in lactosaminoglycans and keratan. In lactating mammary gland, alpha-lactalbumin is specifically expressed: in its presence the acceptor substrate specificity is shifted from N-acetyl-D-glucosaminyl-R to free glucose. In this case, gal-T is designated the A protein, alpha-lactalbumin the B protein of lactose synthase. The enzyme has also been claimed to be involved in a variety of cell surface recognition phenomena. Such a role has not been demonstrated in human biology.
Physiology/Pathology	The physiological role of gal-T is given by its enzymatic activity (see above). The enzyme is constitutively expressed in all human cells. Absence of its expression or overproduction are not known. In a case of congenital dyserythropoietic anemia type II, gal-T activity of a microsomal preparation of peripheral blood mononuclear cells was 24% of a reference value. Cancer-associated "isoenzymes" have been reported; their structural and genetic basis have not been determined.
Degradation	Gal-T is completely released from the cells after proteolytic cleavage. Soluble gal-T is thus present in all body fluids including urine; renal excretion may be the major route of elimination.
Genetics/Abnormalities	The human gene is located on chromosome 9, band p13. The human locus contains six exons spanning over 50 kb of germline genomic DNA. Heterologous expression: Human full length gal-T has been expressed in E. coli as soluble active enzyme or as fusion protein with beta-galactosidase; the enzyme has also been expressed in Saccharomyces cerevisiae in a membrane-bound form, in a soluble form and in a soluble form devoid of the N-glycosylation site. The last form of gal-T is more homogeneous than the human enzyme (2 isoforms on isoelectric focusing instead of the 13 observed for the human gal-T).
Half-life	19 h, approx. (cell culture); *in vivo*: nd.
Concentration	0.18 mg/L in plasma; 10 mg/L in breast milk.
Isolation Method	From milk: Initial steps involve defattening by centrifugation; removal of casein by acidification at pH 4.6 followed by centrifugation. From plasma or serum: Initial (optional) steps involve 25 - 40% $(NH_4)_2SO_4$ cuts and dialysis against water. Specific purification methods make use of affinity chromatography on N-acetyl-D-glucosamine-p-aminophenyl-Sepharose and/or alpha-lactalbumin-Sepharose. UDP-hexanolamine-Sepharose and hydrophobic chromatography have also been reported.
Amino Acid Sequence	Available from GenBank Accession M22921
Disulfides/SH-Groups	A free thiol group may be located close to the UDPgalactose binding site.

General References

Hill, R.L. and Brew, K. *Adv. Enzymol.* 1975, **43**:411-490.
Gerber, A.C. et al. *Eur. J. Biochem.* 1979, **93**:453-460.
Ram, B.P. and Munjal, D.D. CRC *Crit. Rev. Biochem.* 1985, **17**:257-311.
Strous, G.J. CRC *Crit. Rev. Biochem.* 1986, **21**:119-151.
Lowe, J.B. *Seminars in Cell Biol.* 1991, **2**:289-307.
Aoki, D., Appert, H.E., Johnson, D., Wong, S.S., Fukuda, M.N. *EMBO J.* 1990, **9**:3171-3178.
Wang, Y., Wong, S.S., Fukuda, M.N., Zu, H., Liu, Z., Tang, Q., Appert, H.E. *Biochem. Biophys. Res. Com.* 1994, **204**:701-709.
Zu, H., Fukuda, M.N., Wong, S.S., Wang, Y., Liu, Z., Tang, Q., Appert, H.E. *Biochem. Biophys. Res. Com.* 1995, **206**:362-369.
Malissard, M., Borsig, L., Di Marco, S., Grutter, M.G., Kragl, U., Wandrey, C., Berger, E.G. *Eur. J. Biochem.* 1996, **239**:340-348.

Ref. for DNA/AA Sequences

Masri, K.A. et al. *Biochem. Biophys. Res. Commun.* 1988, **157**:657-663.
Watzele, G. and Berger, E.G. *Nucl. Acid. Res.* 1990, **18**:7174.

Beta-2-Microglobulin

Lennart E. Lögdberg

Synonyms	MHC class I β-chain, MHC class I light chain
Abbreviations	β_2-m, β_2m, β2m, beta-2-m
Classifications	Electrophoretic mobility: β_2
Description	A cell surface and plasma protein consisting of a single immunoglobulin (Ig)-like domain. Synthesized by all nucleated cells. Expressed on cell surfaces as the invariant light chain of major histocompatibility (MHC) class I antigen heterodimers and related molecules, and thus involved in antigen presentation to T lymphocytes. Occurs in all biological fluids, both as monomers and in higher molecular weight complexes. Amino Acid Sequence was found to be homologous to Ig-domains, which led to the establishment of the Ig protein superfamily.
Structure	A non-glycosylated protein of 99 aa, folded as a sandwich of two β-sheets, with 3 and 4 antiparallel β-strands, respectively, into a compact globular molecule with a Stokes' radius of 1.6 nm and a frictional ratio of 1.16. The two β-sheets are stabilized by an intrachain disulfide bond between C25 and C80. This structure is characteristic of the immunoglobulin fold, found in Ig-domains and Ig-like domains of other members of the Ig-superfamily, such as T lymphocyte antigen receptors or MHC antigens. Crystallized in free form and also as part of several MHC class I antigens.
Molecular Weight	11,600 (sedimentation equilibrium ultracentrifugation); 11,730 (calculated from aa sequence).
Sedimentation Coeff.	1.6 S
Isoelectric Point	5.8 (at 5 °C)
Extinction Coeff.	16.9 (280 nm,1 %,1 cm) or 1.98×10^4 (M^{-1} cm^{-1}).
Enzyme Activity	Unknown
Coenzymes/Cofactors	Unknown
Substrates	Unknown
Inhibitors	Unknown
Biological Functions	Important role in the immune system, particularly with respect to the role of MHC class I antigens in antigen presentation. Thus, the protein is necessary for proper cell surface expression of MHC class I antigens. It stabilizes the conformation of the MHC class I antigen and appears to promote optimal peptide binding to form ternary complexes. These complexes serve as ligands for the antigen specific receptors on one of the major subpopulations of T lymphocytes (CD8+ cells).
Physiology/Pathology	In the absence of β_2-microglobulin (targeted gene disruption in mice), MHC class I expression is dramatically reduced. Such mice exhibit a profound deficiency in the development of mature CD8+ T lymphocytes and display defective cell-mediated antigen-specific cytotoxicity. The mice also demonstrate less well understood deficiencies in other T lymphocyte subpopulations and have abnormal natural killer cell function.

	β_2-Microglobulin has served as a model protein for the study of kidney function. Decreased glomerular filtration correlates with raised serum levels of the protein. In hemodialysis patients, the dramatically increased levels are associated with amyloid osteoarthropathy, where the amyloid fibrils are actually formed from β_2-microglobulin. Raised urinary levels of the protein is a sensitive indicator of renal proximal tubular damage. In patients with intact kidney function, raised serum levels are recorded in inflammatory and malignant conditions. Particularly, β_2-microglobulin is an important prognostic indicator in hematological malignancies.
Degradation	Elimination from circulation predominantly by glomerular filtration, followed by uptake and degradation in the proximal tubules of the kidneys.
Genetics/Abnormalities	Single gene located on chromosome 15 (15q21-q22.2), containing 4 exons separated by three introns. The bulk of the protein encoded by the second exon. While different murine alleles have been described, only a single human sequence has been identified, with some evidence for allelism from RFLP-analyses. A human electrophoretic variant appears to result from proteolytic processing.
Half-life	1.2–2.8 h (blood circulation)
Concentration	Plasma: 1–2 mg/L. Urine: About 0.1 mg/24 hr volume.
Isolation Method	Conveniently purified from the urine of patients with tubular proteinuria, using a combination of charge and size dependent preparative techniques, *e.g.*, ultrafiltration followed by sequential gel filtration chromatography, zone electrophoresis, and ion-exchange chromatography. Isolation from serum or plasma by immunoaffinity chromatography. Can be produced by overexperssion in *E. coli.*
Amino Acid Sequence	IQRTPKIQVY SRHPAENGKS NFLNCYVSGF HPSDIEVDLL KDGERIEKVE HSDLSFSKDW SFYLLYYTEF TPTEKDEYAC RVNHVTLSQP KIVKWDRDM A member of the Ig-superfamily.
Disulfides/SH-Groups	One disulfide bond (between position 25 and 80). No free sulfhydryl groups.
General References	Berggård, B. et al. *Scand. J. Clin. Lab. Invest.* 1980, **40 (Suppl. 154):** 13–25. Karlsson, F. A. et al. *Scand. J. Clin. Lab. Invest.* 1980, **40 (Suppl. 154):** 27–37. Raulet, D. H. *Adv. Immunol.* 1994, **55:** 381–421. Williams, A. F. and Barclay, A. N. *Ann. Rev. Immunol.* 1988, **6:** 381–405.
Ref. for DNA/AA Sequences	Cunningham, B. A. et al. *Biochemistry* 1973, **12:** 4811–4822 (aa sequence). Güssow, D. et al. *J. Immunol.* 1987, **139:**3132–3138 (gene structure). Suggs, S. V. et al. *Proc. Natl. Acad. Sci. U. S. A.* 1981, **78:** 6613–6617 (cDNA sequence).

Biliverdin Reductase

Mahin D. Maines

Synonyms	None
Abbreviation	BVR
Classification	EC 1.3 1.24
Description	Cytoplasmic protein expressed to varying degrees in all cell types examined. In mammalian species, BVR appears to be the product of a single gene encoding a transcript of ~1.3–1.5 Kb. While not stress-inducible to the degree of heat-shock proteins, BVR enzyme activity appears to be thermostable in brains of heat-shocked rats, and is co-expressed in regions of the brain expressing constitutive or inducible forms of heme oxygenase isozymes.
Structure	Overall structure is hydrophobic with hydrophilic amino terminus.
Molecular Weight	39,000–42,000; rodent 33,000–36,000 (SDS-PAGE)
Sedimentation Coeff.	Unknown
Isoelectric Point	6.03, 5.83, 5.68, 5.55, 4 distinct zones.
Extinction Coeff.	Unknown
Enzyme Activity	Catalyzes reduction of the γ-carbon bridge of biliverdin IX to produce bilirubin IX. Exhibits 2 pH optima, 6.75 and 8.7; utilizes a different cofactor at each optimum.
Coenzymes/Cofactors	Utilizes reducing potential of NADPH (optimum pH 8.7) or NADH (optimum pH 6.75) K_M for NADPH (~3 M) is 100x lower than that for NADH.
Substrates	Biliverdin IX ; also active on the mono-methyl ester of biliverdin IX , and with 50x higher K_M on biliverdin IX.
Inhibitors	SH reagents (5,5′-dithiobis (2-nitrobenzoic acid) [DTNB], N-ethylmaleimide, p-chloromercuribenzoic acid, iodoacetamide) $HgCl_2$, hematin, non-physiological metalloporphyrins (Zn- and Sn-). Substrate inhibition at 5 μM.
Biological Functions	The heme moiety of senescent/degraded hemoproteins is converted to biliverdin by the action of two isoenzymes of heme oxygenase. In mammalian species, biliverdin is subsequently converted to bilirubin by biliverdin reductase in nearly all organs. Bilirubin, primarily bound to albumin, is transported to the liver, where after conjugation with glucoronic acid, it is excreted in the bile. Both biliverdin and bilirubin bound to albumin display potent antioxidant activity.
Physiology/Pathology	Increased production of bilirubin, combined with decreased rate of bile pigment excretion cause jaundice. Decreased activity may contribute to green jaundice.
Degradation	Unknown

Genetics/Abnormalities	The human gene encoding BVR has not been cloned, however, the rat gene is about 12 Kb in length and is organized into five exons. Occurrence of genetic abnormalities is not known.
Half-life	Unknown
Concentration	Concentration varies in different organs. Highest levels are found in tissues associated with heme catabolism such as spleen and liver, and also in kidney and brain. Thymus and testes also have detectable levels of mRNA.
Isolation Method	Cytosol from tissue homogenates is ammonium sulfate and potassium chloride fractionated, then subjected to NADP-agarose affinity columm chromatography.
Amino Acid Sequence	The two N terminal aa of rat BVR are apparently removed during processing of the protein. The N terminal sequence of human BVR suggests that is is processed in the same way. Microsequencing of HPLC purified human peptide fragments indicates both conserved regions (80–90% identity) and peptides unique to the human form.
Disulfides/SH-Groups	Human BVR contains 4–5 titratable SH-groups. In rat BVR, there are 3 cysteine residues, all of which are titratable. Of these, one is absolutely essential for activity, while the others are individually dispensable, although mutation of both abolishes activity.
General References	Lemberg, R. and Legge, J. W. **In** *Hematin Compounds and Bile Pigments*, Interscience 1949, pp. 114–123. Singleton, J. W. and Laster, L. *J. Biol. Chem.* 1965, **240**: 4780–4789. Colleran, E. and O'Carra, P. **In** *Chemistry and Physiology of the Bile Pigments*, Fogerty International Center 1977, pp. 69–80. Kutty, R. K. and Maines, M. D. *J. Biol. Chem.* 1981, **256**: 3956–3962. Huang, T. J. et al. *J. Biol. Chem.* 1989, **264**: 7844–7849. Maines, M. D. **In** *Heme Oxygenase-Clinical Applications and Functions*, CRC Press 1992, pp. 43–61. Ewing, J. F. et al. *J. Neurochem.* 1993, **61**: 1015–1023. McCoubrey, W. K., Jr. and Maines, M. D. *Eur. J. Biochem.* 1994, **222**: 597–603.
Ref. for DNA/AA Sequences	Fakhrai, H. and Maines, M. D. *J. Biol. Chem.* 1992, **267**: 4023–4029. Maines, M. D. and Trakshel, G. M. *Arch. Biochem. Biophys.* 1993, **300**: 320–366. McCoubrey, W. K., Jr. et al. *Gene* 1995, in press.

Butyrylcholinesterase

Oksana Lockridge

Synonyms	Serum cholinesterase; Plasma cholinesterase; Pseudocholinesterase; Non-specific cholinesterase
Abbreviations	BChE; BuChE; CHE; ChE. BCHE is the official abbreviation for the gene.
Classifications	EC 3.1.1.8, acylcholine acylhydrolase
Description	Found in human serum or plasma where it is a soluble glycoprotein synthesized in the liver. Also present in most other tissues including lung, brain, heart, muscle, lining of the blood capillaries, intestinal mucosa. Human red blood cells contain membrane-bound acetylcholinesterase (AChE), an enzyme with very similar properties to BChE. Belongs to family of serine esterases that have an active site serine and are irreversibly inhibited by organophosphate esters.
Structure	The enzyme in human serum is a tetramer of four identical subunits arranged as a dimer of dimers. The interchain disulfide bond (Cys-571) is important for stability but unimportant for tetrameric structure as this bond can be selectively reduced and alkylated without changing the molecular weight. Noncovalent, hydrophobic interactions hold the four subunits together. Membrane-bound forms are found in muscle, intestinal mucosa, capillaries, brain, but the identity of the membrane anchor is not yet known.
Molecular Weight	Molecular weight of BChE in human serum is 340,000 to 348,000 (ultracentrifugation, Sephadex gel chromatography). Subunit weight of the 574 aa (65,092) plus 9 carbohydrate chains (23.9%) is approximately 85,500. The value is approximate because the carbohydrate weight is not exact.
Sedimentation Coeff.	10.7 S
Isoelectric Point	3.99
Extinction Coeff.	18 (280nm, 1%, 1cm)
Enzyme Activity	Hydrolyzes benzoylcholine, butyrylthiocholine, acetylthiocholine, succinyldicholine, as well as noncholine esters heroin, aspirin, alpha-naphthyl acetate, o-nitrophenyl butyrate, phenyl acetate, p-nitrophenyl acetate. Also hydrolyzes o-nitroacetanilide.
Coenzymes/Cofactors	None
Substrates	Clinically important substrates are the muscle relaxants succinyldicholine, and mivacurium, which are hydrolyzed by people with usual BChE but not hydrolyzed by people with rare genetic variants of BChE. Cocaine is hydrolyzed by BChE to form methylecgonine, an inactive product. Metabolism by BChE is the major determinant of the fate of cocaine in humans. A substrate useful for measuring enzyme activity is benzoylcholine (0.05 mM benzoylcholine, 0.067 M Na/K phosphate buffer pH 7.4, 25°C; difference in absorbance at 240 nm between substrate and product is 6,700 M^{-1} cm^{-1}). Butyrylthiocholine (1.0 mM butyrylthiocholine, 0.3 mM DTNB, 0.1 M Na phosphate pH 8.0, 25°C; extinction coefficient of prod-

uct is 13,600 M^{-1} cm^{-1} at 412 nm) and propionylthiocholine are also widely used.

Inhibitors

Naturally occurring inhibitors are eserine (physostigmine), from calabar beans, and solanine from potato peel. Synthetic inhibitors include: organophosphate insecticides, chemical warfare nerve agents, carbamates used as pesticides, antiasthmatic bronchodilator (bambuterol), and drugs to treat myasthenia gravis (neostigmine and pyridostigmine) and psychosis (chlorpromazine). Inhibitors that selectively inhibit BChE include iso-OMPA, ethopropazine, and bambuterol. Inhibition by eserine at 10^{-5}M (after 30 min of preincubation) defines an esterase as a cholinesterase. Inhibitors for classifying genetic variants are dibucaine, NaF, and RO2-0683. The mechanism of inhibition by organophosphates is irreversible alkylation of the active site serine (Ser-198). Carbamates also alkylate Ser-198 but inhibition is reversible.

Biological Functions

Biological function is unknown. Role in cell proliferation and differentiation is suggested by its highly specific distribution in developing chicken retina and monkey visual pathway. Protective role against poisons that are eaten or inhaled is suggested by its high concentration in liver and lungs, tissues known as the principal detoxication sites.

Physiology/Pathology

Clinically important for diagnosis of poisoning by carbamate and organophosphate insecticides. These poisons are toxic because they inhibit AChE at the nerve muscle junction, not because they inhibit BChE. However, serum BChE activity indicates extent of inhibition of AChE. Decreased concentration of BChE in human serum accompanies severe liver disorders, such as cancer and cirrhosis, reflecting the diminished capacity of hepatocytes to synthesize proteins. When serum BChE activity falls below 0.2 U/ml (normal is 1 U/ml) the patient will experience prolonged apnea after receiving a single dose of succinyldicholine.

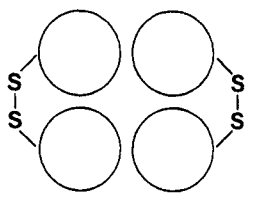
Degradation

Proteolysis cleaves up to 40 aa from the C-terminus without destroying activity. Monomers and dimers result from proteolysis. The tetramer contains 36 Asn-linked carbohydrates of the complex type terminating in sialic acid which suggests that clearance of aged BChE occurs via the galactose receptor liver.

Genetics/Abnormalities

One out of 3000 Caucasians is homozygous for the atypical variant (Asp 70 Gly), in which Asp-70 has been replaced by Gly. Patients with atypical BChE are unable to breathe for 2 hours after a normal dose of succinyldicholine; this dose produces apnea of 3 to 5 min duration in most people. Other genetic variants that respond with prolonged apnea are 21 silent BChE: frameshift at Ile-6, Tyr-33→Cys, Pro-37→Ser, frameshift at Gly-117, Gln-119 stop, Leu-125→Phe, Asp-170→Glu, Ser-198→Gly, Ala-201 →Thr, Thr-250→Pro, Glu-271 stop, frameshift at Thr-315, insert at Pro-359, Gly-365→Arg, Phe-418→Ser, Arg-465 stop, Trp-471→Arg, Tyr-500 stop, altered splicing in intron 2, Arg-515→Cys, Gln-518→Leu. The silent homozygote occurs with a frequency of 1 in 100,000; no health abnormalities have been noted in people with silent BChE. The H variant, Val-142 →Met, has 10% of normal activity. The J variant, Glu-497→Val, has 40% of normal activity. The Fluoride-1 variant, Thr-243→Met, the Fluoride-2 variant, Gly-390→Val, and the Ala-184→Val variant also respond with prolonged apnea. The K allele (Ala-539→Thr), carried by 12% of Caucasians and Japanese, has 70% of normal activity but has no problems with succinyldicholine. The Cynthiana variant is resistant to succinyldicholine because of 2 to 3 fold elevated BChE activity. There is a single gene for

	BCHE in man as well as in monkey, cow, sheep, pig, rabbit, dog, rat, mouse, guinea pig and chicken. The human BCHE gene is on the long arm of chromosome 3 at q26.2.
Half-life	11 days in serum
Concentration	Normal concentration in human serum is 0.005 g/L.
Isolation Method	Purification of serum BChE by ion exchange chromatography at pH 4.0 purifies BChE about 800 fold. The 2nd step is affinity chromatography on procainamide-Sepharose. The 3rd step is ion exchange at pH 7.0. Purified enzyme is stored in sterile phosphate buffer at neutral pH at 4°C. Partially purified preparations are unstable because proteases copurify; the preparation is stable for years after affinity column chromatography. Specific activity of highly purified human BChE is 200 micromoles/min/mg (0.05 mM benzoylcholine, 0.067 M Na/K phosphate pH 7.4, 25°C), or 700 micromoles/min/mg (1mM butyrylthiocholine, 0.1 M Na phosphate pH 7.4, 25°C). kcat/Km = 50 x 10^6 $M^{-1}s^{-1}$ for benzoylcholine; 15 x 10^6 $M^{-1}s^{-1}$ for butyrylthiocholine.
Amino Acid Sequence	Mature BChE in serum has 574 aa per subunit. The catalytic triad consists of Ser-198, His-438, Glu-325. The atypical variant (Asp-70→Gly) has reduced affinity for all positively charged substrates and inhibitors; Asp-70 is a component of the peripheral anionic site.
Disulfides/SH-Groups	Each subunit contains 3 intrachain disulfide bonds at Cys 65-92, 252-263, and 400-519. Two subunits are covalently linked through a disulfide bond at Cys-571. One free but inaccessible sulfhydryl is at Cys-66.
General References	Lockridge, O. Genetic variants of human serum butyrylcholinesterase influence the metabolism of the muscle relaxant succinylcholine. In: *Pharmacogenetics of Drug Metabolism.* Kalow W. (ed.) Pergamon Press, New York, 1992, pp.15-50. Kalow, W. and Grant, D.M. Pharmacogenetics. In: *The Metabolic and Molecular Bases of Inherited Disease.* Scriver C.R., Beaudet, A.L., Sly, W.S., Valle, D. (eds.), McGraw-Hill Inc., N.Y. 1995, **Vol. 1,** pp 303-326. Whittaker, M Cholinesterase. *Monographs in Human Genetics.* Beckman, L. (ed.), Karger, Basel, Switzerland 1986, **Vol. 11.**
Ref. for DNA/AA Sequences	Lockridge, O. et al. Complete amino acid sequence of human serum cholinesterase. *J.Biol.Chem.* 1987, **262**:549-557. McTiernan, C. et al. Brain cDNA clone for human cholinesterase. *Proc. Natl. Acad. Sci. USA* 1987, **84**:6682-6686. Arpagaus, M. et al. Structure of the gene for human butyrylcholinesterase. Evidence for a single copy. *Biochemistry* 1990, **29**:124-131.



Schematic model of serum cholinesterase.

Human Protein Data
Edited by André Haeberli

C1-inhibitor

Peter J. Späth

Synonyms	α_2-Neuramino-glycoprotein; C1-esterase inhibitor; C1s-inhibitor; $\overline{C1s}$-inhibitor (C1-inactivator)
Abbreviations	C1-INH; $\overline{C1}$-INH; (C1-INA)
Classifications	Electr. mob.: α_2 (pH 8.6)
Description	A plasma protease inhibitor that is essential and has a pivotal role for the regulation of early events in precipitation of the contact activation pathways (coagulation and kinin system) and the classical pathway of the complement system. C1-INH is a member of the SERPIN 'superfamily' (***ser***in ***p***rotease ***in***hibitors). C1-INH is an acute-phase protein.
Structure	C1-INH is a compact, single polypeptide chain, two domain molecule with approximate N-terminal 113 aa rod-like moiety and C-terminal 365 aa globular domain. The globular domain includes the reactive site of the inhibitor. The native molecule has an unusually high carbohydrate content with a particularly large amount of galactose and N-acetylgalactosamine. With at least 26% carbohydrate content it is one of the most heavily glycosylated plasma proteins. The elongated N-terminal domain contains tandem repeat sequences and multiple ($\equiv$ 17) carbohydrate attachement sites. 14 of the 20 probable oligosaccharides are galactosamine based units while the remeining 6 are asparagine-linked glucosamine based oligosaccharides. Removal of a large portion of the N-terminal carbohydrate has no effect on C1-INH function. The marginal role of the N-terminal domain for functional activity is underscored by retained protease inhibitor activity of truncated C1-INH molecules and the lack of sequence conservation at the N-terminus of C1-INH from different species.
Molecular Weight	104,000 to 105,000: Native, glycosylated molecule (by analytical ultracentrifugation and SDS-PAGE); 71,100 when calculated from protein backbone and carbohydrate content. Neutron scattering of the native molecule indicates an M_r of 76,000. Protein backbone: 52,869 to 52,880 (478 aa according to cDNA sequencing).
Sedimentation Coeff.	3.7-4.2 S
Isoelectric Point	2.7-2.8
Extinction Coeff.	3.6 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None; heparin enhances inhibitory activity towards enzymatically active complement subcomponent C1r and C1s.
Substrates	Enzymatically active complement subcomponents C1r and C1s; coagulation factors αFXIIa (HFa, FXIIa), βFXIIa (HFf, FXIIb), and kallikrein. *In vitro* (in addition): mouse nerve growth factor, coagulation factor XIa, plasmin.

Inhibitors

As C1-INH has no enzymatic activity, inhibitors have to be considered as molecules interfering with recognition and/or cleavage of C1-INH by its target enzymes: e.g. anti-C1-INH autoantibodies. Activity of C1-INH is destroyed by flufenemic acid, EtOH (cold ethanol fractionation of plasma), ether and heating at 60° C for 10 min. in absence of stabilisators.

Biological Functions

The primary site of biosynthesis *in vivo* are probably the hepatic parenchymal cells. Main extrahepatic sites of biosynthesis are: skin fibroblasts and monocytes/macrophages. In unstimulated cell cultures the secretion rates of C1-INH are in the order of monocytes >> fibroblasts (skin, synovial). > HepG2 >HUVEC >> chondrocytes. Interferons α, β, and γ have a positive effect on the biosynthesis by human monocytes and skin fibroblasts. Interferons α and β apparently can act synergistically with INF-γ to stimulate biosynthesis. When cell cultures are stimulated by IFN-γ (100-200 IU/ml) the increase in synthesis rate is most pronounced in monocytes >> chrondrocytes > HepG2 > HUVEC > fibroblasts. IFN-γ is also able to increase C1-INH mRNA expression in hepatoma cell line Hep 3B. Apparently, IFN-γ has a marked effect on the stability of monocyte C1-INH mRNA but not on the stability of mRNA in other cells. IL-6 (recombinant or monocyte conditioned media) induce in Hep G2 cell cultures the *de novo* C1-INH synthesis dose dependently and this effect can be abrogated by anti-IL-6 antibodies. INF-γ synergistically increases C1-INH synthesis from IL-6 stimulated HepG2 cells. M-CSF apparently also can enhance C1-INH synthesis by human monocytes.

C1-INH acts as a competitive substrate that mimics the natural substrate of its target enzymes and is split by them. The reactive site which becomes cleaved by the target protease, is apparently located in a loop protruding the surface of the molecule. Cleavage site is between the specific reactive P1 (Arg 444) and the C-terminal standing P1' residue (Thr 445). The cleavage apparently induces a transition of a native C1-INH molecule under energetic tension to a complexed C1-INH of much improved stability. The stabilization apparently requires the partial insertion of the reactive centre peptide (residues P1-P14 of the cleaved loop) into the β-sheet A of the molecule. Upon cleavage C1-INH covalently (acyl bond?) and stoichometrically combines with its target enzyme and forms denaturation resistant inactive complexes. A 34 residue peptide is released. Formation of the complexes abrogates the activity of the target enzymes. The inhibition by C1-INH is irreversible and C1-INH is irreversibly consumed. The regulation by C1-INH of active C1 comprises covalent binding of the target proteases C1r and C1s and the subsequent disintegration of the pentamolecular complex of C1, composed of C1q and the active enzymes $C1r_2$ and $C1s_2$. The complexes resulting are C1q, which remains bound to the activator of the classical pathway of complement and two macromolecules composed of C1-INH•C1s•C1r•C1-INH.

C1-INH apparently helps to eliminate activated first component of complement, C1. Inactivation and elimination of C1 occurs in several steps: complex formation and disintegration of enymatically active C1. The inhibitor-enzyme complexes are removed very rapidly from the circulation. The receptors and cells involved are not well known. Based on a specific pentapeptide sequence highly conserved in SERPINs, a proposed serpin-enzyme-complex (SEC) receptor, most likely on hepatic cells, could be operative.

C1-INH is the only inhibitor with high efficacy for regulation of the complement classical pathway activation. The cationic protein Factor J is, like C1-INH, able to inhibit enzymatically active C1 by dissociating it into its subunits. An enzyme known to inhibit enzymatically active C1s is nexin I which has low efficacy. αFXIIa, βFXIIa, and kallikrein are also inactivated by other plasma proteinase inhibitors. There is evidence for C1-INH

being the only efficient inhibitor of the permanent autoactivation of coagulation factor XII.

Physiology/Pathology

Functional deficiency of C1-INH result in continuous consumption of C4 and C2, the disregulation of the kinin pathway and from time to time the local accumulation of bradykinin. The clinical manifestation of functional C1-INH deficiency is angioedema (AE). AE results from bradykinin-induced vascular leakage reaction of post-capillary venules. Leakage leads to an influx of plasma into the intercellular space. Transient, well circumscribed, cold, white, hard, painless, and not itching swellings of the deep cutaneous, and subcutaneous tissue of the extremities, the genitals or the face are the consequence. In case of a gastrointestinal leakage reaction of submucous tissue colic-like pain is suffered. Larynx edema, due to acute airway obstruction, might be lethal.

Functional deficiency in C1-INH might be hereditary or acquired. With one known exception C1-INH deficiency is inherited as an autosomal dominant trait (heterozygous individuals). In type I or common form of hereditary angio(neurotic) edema (HAE) a normal and a null allele are present in the patients (one exception). An insufficient synthesis and elevated turnover of a normal C1-INH protein results in C1-INH levels of 5-30% of normal. From skin fibroblast cultures evidence was obtained, that in some of the type I of HAE patients an increased turnover and not a synthesis reduced to levels below 50%, nor a reduced secretion of the normal protein is the apparent background for the low level of functional C1-INH.

In type II or the variant form of HAE various functionally inactive C1-INH proteins can be detected at normal or elevated levels. Type II fibroblasts contain two forms of C1-INH, the normal and the abnormal form, each of them accounting for ≈ 50% of the synthesized C1-INH. Due to elevated turnover the concentration in the circulation of the product of the normal allele can be below 50%.

In acquired angioedema (AAE) type I C1-INH and C1 (C1q) are diminished. In many cases the underlying disease is a lymphoproliferative disorder. Diminution of C1-INH may precede manifestation of lymphoproliferative disease by years. In type II of AAE an anti-C1-INH autoantibody apparently enables cleavage of C1-INH by its target enzyme however interferes with complex formation between enzyme and inhibitor. An inactive inhibitor of apparent M_r of 95,000 results (SDS-PAGE).

Attenuated androgens like Danazol® or Stanazol® can induce a 20% increase of C1-INH levels in normal individuals. These attenuated androgens are used for long-term therapy in patients deficient in C1-INH function. C1-INH concentrate is available for replacement therapy (see Isolation Method).

C1-INH concentrate was used to attenuate inflammation, e.g. in septic shock; vascular leak syndrome after bone marrow transplantation or in connection with high doses of interleukin-2 therapies; and endotoxin-induced pulmonary dysfunction. *In vitro* C1-INH prevented complement-mediated activation by xenogeneic endothelial cells.

Degradation

Unknown

Genetics/Abnormalities

The inhibitor protein gene locus is on chromosome 11, p11.2-q13.1. The EMBL sequence accession number of the single gene is X54486. The gene spans approximately a 17 kilobases long DNA sequence. It consists of 8 exons and 7 introns. Repetitive Alu elements and some possible regulatory elements were identified within the gene, among others an androgen response element (ARE), an IL-6 response element (IL-6 RE), an interferon response element (ISRE). The number of Alu elements is 19 whereof 17 lie

within the introns. The mature C1-INH mRNA is approximately 1.85 kb in length and has 60 untranslated and 1437 coding bases.
On a molecular level a distinction between type I and type II HAE is not possible. The study of the first and yet single homozygous defect of the C1-INH gene revealed that the defect apparently occurs in the gene control region and that it might have evolved due to multiple consanguinity. The homozygous defect apparently does not completely shut off the regulatory capacity of this part of the two genomes because the patient shows a response to therapy with attenuated androgens.
Studies on the background of (heterozygous) HAE revealed in about 20% of the families a gross alteration of the C1-INH gene. These patient-specific alterations can involve deletion of exon 4, 7 and 8, deletion of exons 1 through 3, exons 1 through 4, exons 1 through 6, exons 4 through 6, and exons 5 through 8, as well as duplication of exon 4. Clusters of intragenic Alu repeats predispose the C1-INH gene to deletions and duplications.
In about 80% of the patients point mutations or mini duplications or deletions are the backgroud for functional C1-INH deficiency:
Consequences of single base changes: Gly-407→Arg, Val-432→Glu, Val-433→Met, Ala-434→Glu, Ala-436→Thr or Val, Ile-440→Val, Arg-444 →Cys, His, Leu, or Ser, Val-451→Met, Gln-452→Glu, Phe-455→Ser, Leu-459→Arg or Pro, Pro-467→Arg, Arg-472STOP, Pro-476→Ser
Consequences of mini deletions: Asn-250, Lys-251, Val-334→Met with generation of a stop codon 40 nucleotides downstream of the frameshift, Ser-400STOP, Leu-414STOP
Consequences of mini duplications or insertions: Tyr401STOP, Gly-431 →Val-Trp
Polymorphism: Val-458→Met, Gln-452→Glu

Half-life

67 to 72 hrs

Concentration

0.2 g/L (mean value in plasma); detectable by radial immunodiffusion or nephelometry (routine diagnostic laboratory). Normal range of concentration: 0.1 - 0.26 g/L
Function can be assessed by two principles: a) tests based on inhibition by endogenous C1-INH of amidolyc or esterolytic cleavage by exogenous, enzymatically active C1s of small substrate molecules, and b) tests based on detection of complexes formed between C1-INH and C1r or C1s following activation of C1.

Isolation Method

Step 1: Plasma fractionation by polyethylenglycol; step 2: Ion-exchange chromatography on DEAE-cellulose or -Sephacel; step 3: (Immuno) Affinity chromatography on Cibachrome-Blue-Sepharose or anti-C1-INH-Sepharose. Industry scale purification of C1-INH is performed by Immuno Vienna, Austria, and the Behring-Werke Marburg, FRG. The Netherlands Red Cross apparently has some facilities to produce a C1-INH concentrate. For therapeutic use C1-INH concentrate can be made virus safe by pasteurization.

Amino Acid Sequence

By sequence homology a member of the SERPIN ' superfamiliy' with proteins such as: α1-proteinase inhibitor (α-1-antitrypsin or 3.5 S-α-1-glycoprotein) antithrombin III, α1-antichymotrypsin, α2-antiplasmin, heparin cofactor II, protease nexin I plasminogen activator inhibitor, elastase inhibitor, angiotensinogen, corticosteroid binding globulin, ovalbumin, mouse contrapsin, chicken gene Y. C1-INH is synthesized with an N-terminal 22 residue signal peptide with a unique cleavage site between Ser(-1) and Asn(+1). The molecule which is composed of 478 aa, has a 21 to 27% sequence homology to other SERPINs, however homology does

not exist at the reactive site region. In spite of this homology, C1-INH is unique among the SERPINs, due to its N-terminal portion which contains a highly repetitive region (positions 43 through 97) comprising 14 tandem repeats of the tetrapeptide Gln-Pro-Thr-Thr and variations thereof in a large hydrophilic segment of the molecule.

Disulfides/SH-Groups

2 Interchain disulfide bonds (Cys108-Cys183 and Cys101-Cys406); no free S-H-groups.

General References

Aulak, K.S. et al. C1-inhibitor structure/function and biologic role. *Behring Inst. Res. Commun.* 1993, **93**:204-213.

Harrison, R.A. Human C1 inhibitor: Improved isolation and preliminary structural characterization. *Biochemistry* 1983, **22**:5001-5007.

Shoemaker, L.R. et al. Hereditary Angioneurotic Oedema - Characterization of Plasma Kinin and Vascular Permeability-Enhancing Activities. *Clin. Exp. Immunol.* 1994, **95**:22-28.

Späth, P.J. and Wüthrich, B. Angioedema. A review on the acquired, allergic or non-allergic, and the hereditary forms. *Rec. Prog. Med.* 1990, **81**:513-531.

Stoppa-Lyonnet, D. et al. Cluster of intragenic Alu repeats predispose the human C1 inhibitor locus to deleterious rearrangements. *Proc. Natl. Acad. Sci. USA* 1990, **87**:1551-1555.

Travis, J. and Salvesen, G.S. Human plasma proteinase inhibitors. *Annu. Rev. Biochem.* 1983, **52**: 655-709.

Verpy, E. et al. Crucial residues in the carboxy-terminal end of C1 inhibitor revealed by pathogenetic mutations impaired in secretion or function. *J. Clin. Invest.* 1994, **95**:350-359.

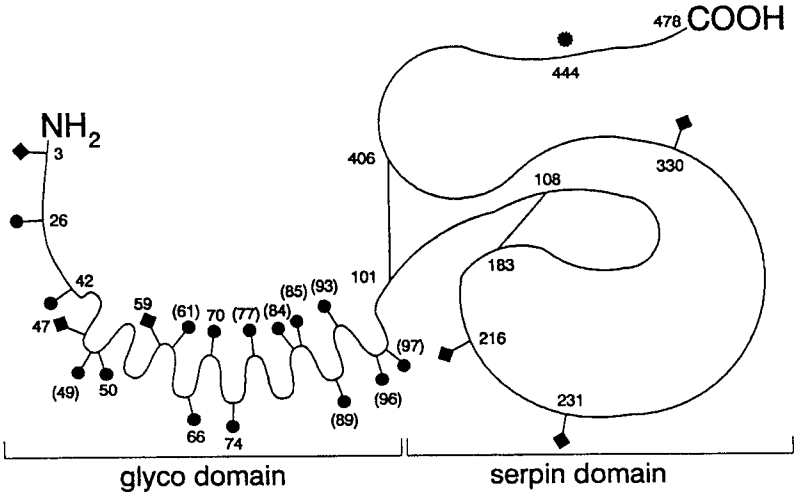
Verpy, E. et al. Efficient detection of point mutations on color-coded strands of target DNA. *Proc. Natl. Acad. Sci. USA* 1994, **91**:1873-1877.

Ref. for DNA/AA Sequences

Bock, S.C. et al. C1-inhibitor: Primary structure, cDNA cloning and chromosomal localization. *Biochemistry* 1986, **25**: 4292-4301.

Carter, P.E. et al.: Genomic and cDNA cloning of the human C1-ii bitor. Intron-exon junctions and comparison with other SERPINs. *Eur. J. Biochem.* 1988, **173**: 163-169.

Schematic Structure of Human C1 Inhibitor



● O-glycosidic
◆ N-glycosidic
carbohydrate attachment sites; ()=tentative
● reactive site (P1)

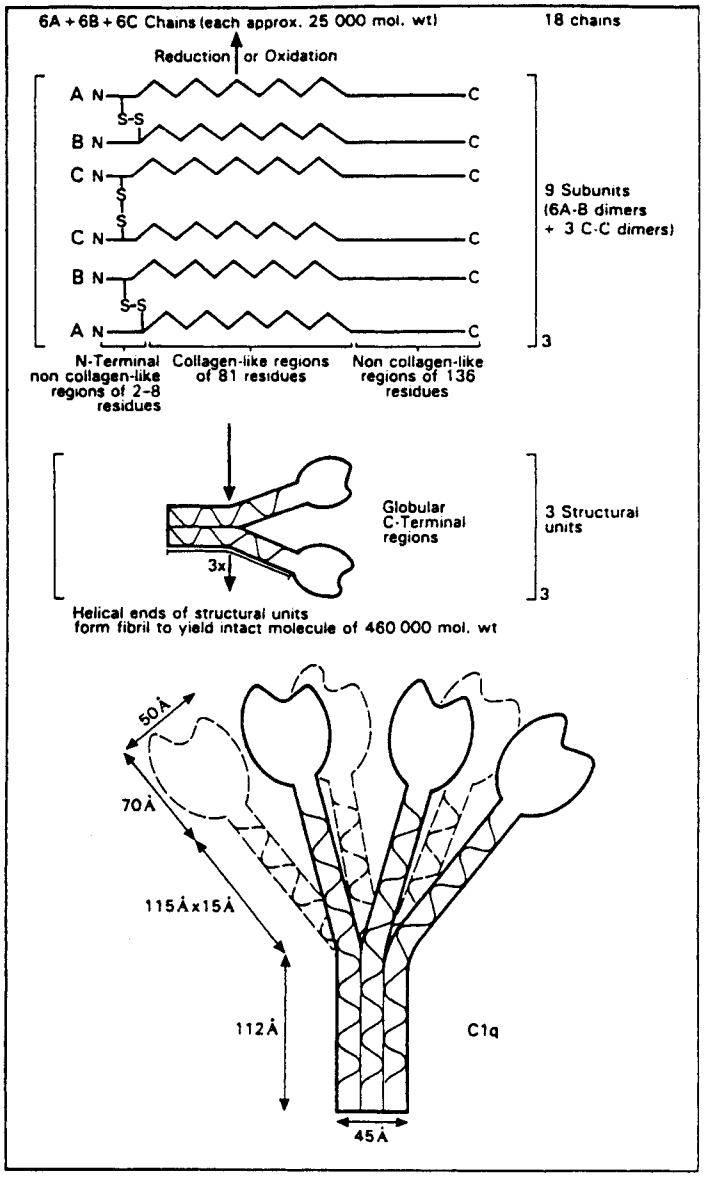
C1q complement protein

Michael Loos and Franz Petry

Synonyms	11S-Protein
Abbreviations	C1q
Classifications	Electrophoretic mobility: gamma2-globulin
Description	Subcomponent of the first component of the complement system. One of the most cationic proteins in human serum. A member of a family of soluble collagen-like proteins such as mannan binging lectin, conglutinin, lung surfactant protein A and D (see also collectin receptor; this volume). Six A-B subunits and three C-C subunits form macromolecular C1q. Carbohydrate content 8% (69% of total carbohydrate are O-linked to certain hydroxylysine residues of all three chains, 31% N-linked to Asn-124 in A-chain).
Structure	Appearance in the electron micrograph as a 'bunch of tulips' with six globular heads connected by six collagen-like stalks forming a central fibril-like stem.
Molecular Weight	459,300 (based on aa sequences and total carbohydrate content). Subunit A-B dimer: 52,750; subunit C-C dimer: 47,600. Monomers A-, B- and C-chain 27,550, 25,200 and 23,800, respectively.
Sedimentation Coeff.	11.1 S
Isoelectric Point	9.3
Extinction Coeff.	6.82 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	C1 esterase C1r and C1s; Ca^{2+}
Substrates	None
Inhibitors	None
Biological Functions	C1q binds to IgG and IgM bearing immune complexes (IC) via the globular heads. This leads to the activation of the subcomponents C1r and C1s of the first component of complement, that are bound to the collagen-like region of C1q. Therefore C1q initiates the classical pathway of complement activation. After activation of C1, the macromolecular complex is dissociated by C1 esterase inhibitor (C1INH) forming a covalent complex with C1r and C1s and leaving C1q bearing IC. C1q-IC bind to C1q receptors on various cell types, incl. B- and T-lymphocytes, neutrophils, platelets and fibroblasts. Antibody-independent activation of the classical pathway has been demonstrated for certain viruses, gram-neg. bacteria, DNA and polyanions. The binding sites for a number of molecules like endotoxin, CRP, serum amyloid A, DNA and heparin has been located on the A chain of C1q.
Physiology/Pathology	C1q mediates solubilization and clearance of IC from circulation. C1q deficiency (32 cases reported) leads to lupus erythematosus-like symptoms

	as chronic vasculitis, skin lessions and glomerulonephritis in 90% of the cases. Severe recurrent viral and bacterial infections are commonly seen in C1q deficient patients.
Degradation	The globular regions are sensitive to pepsin, the collagen-like regions are sensitive to collagenase.
Genetics/Abnormalities	The individual genes that code for the A-; B-and C-chain are located on chromosome 1p34.1-1p36.3. The molecular basis of several types of C1q deficiency has been investigated. In every case single base mutations have been found leading to premature termination codons, frame shifts or changes in the aa sequence.
Half-life	12.4 hrs (in blood circulation of rats)
Concentration	70 - 180 mg/L in serum
Isolation Method	The cationic charge and low solubility at reduced ionic strength, plus the ability to bind to aggregated IgG and to polyanions facilitates the isolation. C1q can be isolated from serum by euglobulin precipitation followed by gel filtration and cation-exchange chromatography or affinity chromatography on IgG-Sepharose.
Amino Acid Sequence	The three distinct chains consist of 223, 226 and 217 residues (A-, B-, C-chain, respectively). The N-terminal region has a collagen-like structure (Gly-X-Y repeats). The C-terminal regions with approx. 140 residues form the globular heads.
Disulfides/SH-Groups	12 free sulfhydryl residues per C1q molecule. 12 intrachain disulfides. 9 interchain disulfide bonds (6 A-B dimers, 3 C-C dimers).
General References	Reid, K.B.M. The complement system. In: *Molecular Immunology*, Hames, B.D. and Glover, D.M. (eds.) Oxford: IRL Press 1988; pp 189-241. Reid, K.B.M. *Behring Inst. Mitt.* 1989, **84**:8-19. Trinder, P.K. et al. *Behring Inst. Mitt.* 1993, **93**:180-188. Petry, F. et al. *J. Immunol.* 1995, **155**:4734-4738.
Ref. for DNA/AA Sequences	Reid, K.B.M. *Biochem. J.* 1985, **231**:729-735. Sellar, G.C., Blake, D.J. and Reid, K.B.M. *Biochem. J.* 1991, **274**:481-490. The nucleotide and aa sequence data are available from the EMBL, GenBank and Swiss Prot Databases under the accession numbers P02745; K03430; P02746; P02747. Further sequence data have been compiled by Petry, F. *Behring Inst. Mitt.* 1993, **93**:321-328 and are available through the author.

Molecular model of human C1q (*Reid*, 1988)



The wavy line represents the tripel-helical, collagen-like region of the molecule. In the lower part of the figure, the 11S C1q molecule is shown. The bold line indicates portions of the molecule pointing towards the reader and the interrupted line indicates portions pointing away from the reader.

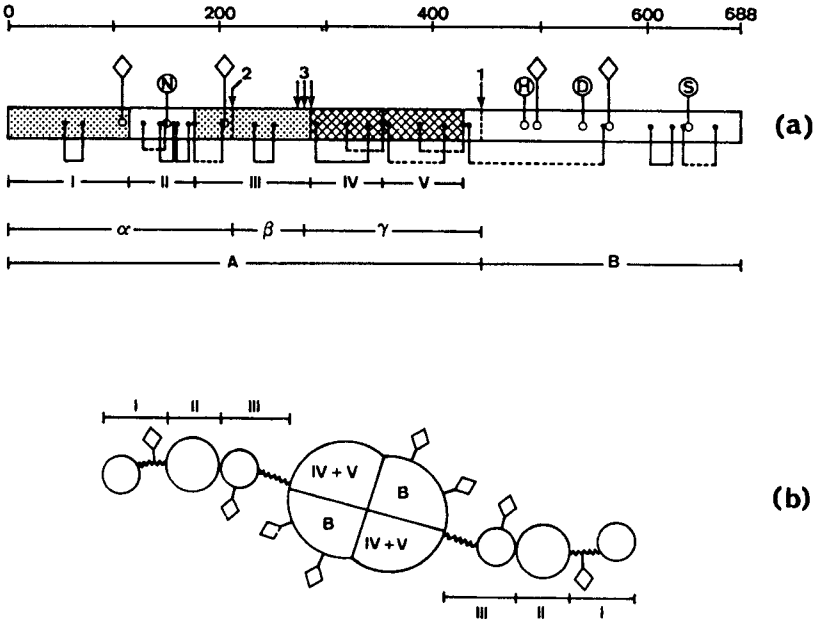
C1r complement protein

Gérard J. Arlaud

Synonyms	None
Abbreviations	C1r (proenzyme form); $C\overline{1}r$ (activated form)
Classifications	EC 3.4.21.41; electrophoretic mobility: β
Description	One of the two serine proteases of C1, the first component of complement, a Ca^{2+}-dependent pentameric protease ($C1_q$, $C1r_2$, $C1s_2$) circulating in plasma. Synthesized by various cells (intestinal epithelium, hepatocytes, monocytes ...). A non covalent homodimer. Each monomer (688 aa) is single-chain in the proenzyme form, and is cleaved on activation into two disulfide-linked chains A (N-terminal, 446 aa) and B (C-terminal, 242 aa). A glycoprotein with 9.3% (w/w) carbohydrate, in the form of Asn-linked complex-type oligosaccharides located in both the A chain (Asn-108 and Asn-204) and the B chain (Asn-51 and Asn-118). The serine active site is located in the B chain and the A chain contains one high-affinity calcium binding site.
Structure	An elongated molecule with a multi-domain structure. Each monomer contains at least four domains, distributed in two functionally distinct regions. The catalytic γ-B regions, located in the centre of the C1r-C1r dimer, mediate intermonomer interactions. The α interaction regions, at each end of the dimer, are responsible for Ca^{2+}-dependent C1r-C1s interactions within C1.
Molecular Weight	172,600. A chain: 55,200; B chain: 31,100 (aa sequence and carbohydrate composition).
Sedimentation Coeff.	7.1 S
Isoelectric Point	4.9
Extinction Coeff.	12.4 (280nm, 1%, 1cm); 214,000 M^{-1}, cm^{-1} (280nm)
Enzyme Activity	Serine protease with trypsin-like specificity. Cleaves Arg-Ile bonds.
Coenzymes/Cofactors	C1q, C1s, Ca^{2+} ions.
Substrates	Itself (autoactivating protease); C1s; chromogenic substrate: N-α-acetyl-glycyl-L-lysine methyl ester.
Inhibitors	C1 inhibitor: regulates C1r activation in C1; blocks $C\overline{1}r$ (and $C\overline{1}s$) activity in $C\overline{1}$ through conversion of the activated $C\overline{1}s$-$C\overline{1}r$-$C\overline{1}r$-$C\overline{1}s$ tetramer into two C1 inhibitor- $C\overline{1}s$-$C\overline{1}r$-C1 inhibitor complexes. Synthetic inhibitors: DFP (di-isopropylphosphorofluoridate); NPGB (p-nitrophenyl-p'-guanidinobenzoate).
Biological Functions	Interacts with C1s to form a Ca^{2+}-dependent tetramer (C1s-C1r-C1r-C1s) which in turn interacts with C1q to form C1. Responsible for C1 activation, a two-step process that is triggered by C1 binding to an activator and involves: (i) C1r self-activation; (ii) activation of C1s by $C\overline{1}r$.

1. Installment 1992 (revised 1997)

Physiology/Pathology	Essential factor for activation of the classical pathway of complement. Known cases of partial or complete genetic deficiency (usually associated with a C1s deficiency) are correlated with lupus erythematosus and renal disease.
Degradation	Physiological pathway of degradation is not known. Degraded in vitro by plasmin and other proteases.
Genetics/Abnormalities	A single polymorphic site (serine/leucine) identified at position 135 of the A chain. Two common alleles described. No known abnormal variants. The C1r and C1s genes are located in region 12p13 of chromosome 12, and lie in a close tail-to-tail orientation.
Half-life	Unknown
Concentration	Plasma: 34 mg/L (0.2 μmol/L)
Isolation Method	Purified from serum by methods involving: (i) isolation of the whole C1 complex by affinity-based approaches utilizing IgG-Sepharose or insoluble immune complexes; (ii) separation of C1 subcomponents by ion-exchange and gel filtration chromatography. Isolation of proenzyme C1r is carried out in the presence of serine protease inhibitors.
Amino Acid Sequence	The A chain is subdivided into five structural modules (I-V). Modules I and III belong to the CUB family (**C**omplement C1r/C1s, **U**egf, **B**one morphogenetic protein). Module II contains erythro-β-hydroxyasparagine (pos. 150) and belongs to a sub-family of EGF (**E**pidermal **G**rowth **F**actor)-like modules known to participate in calcium binding. Modules IV and V belong to the CCP family (**C**omplement **C**ontrol **P**rotein). The B chain is homologous to type I serine proteases, but lacks the "histidine loop" disulfide bridge.
Disulfides/SH-Groups	13 disulfide bridges in each monomer: 1 interchain and 12 intrachain (10 in the A chain, 2 in the B chain). No free sulfhydryls.
General References	Arlaud, G.J. et al. *Biosci. Rep.* 1985, **5**:831-837. Cooper, N.R. *Adv. Immunol.* 1985, **37**:151-216. Arlaud, G.J. et al. *Immunol. Today* 1987, **8**:106-111. Arlaud, G.J. and Thielens, N.M. *Methods Enzymol.* 1993, **223**:61-82. Arlaud, G.J. et al. *Behring Inst. Mitt.* 1993, **93**:189-195.
Ref. for DNA/AA Sequences	Arlaud, G.J. and Gagnon, J. *Biochemistry* 1983, **22**:1758-1764. Gagnon, J. and Arlaud, G.J. *Biochem. J.* 1985, **225**:135-142. Leytus, S.P. et al. *Biochemistry* 1986, **25**: 4855-4863. Journet, A. and Tosi, M. *Biochem. J.* 1986, **240**:783-787. Arlaud, G.J. et al. *Biochem. J.* 1987, **241**:711-720.



Molecular model according to Arlaud et al. (Immunol. Today, 1987). (a) linear representation of the monomer; (b) model of the domain structure of the C1r-C1r dimer.

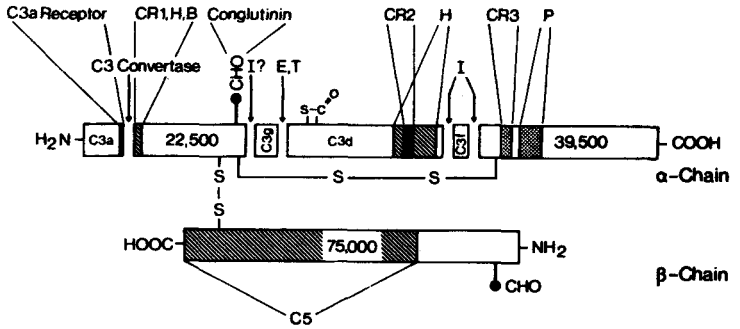
C3 complement protein

Reinhard Burger

Synonyms	β1c-Protein, β1C-Globulin
Abbreviations	C3
Classifications	Electrical mobility: β_2-fraction
Description	A circulating plasma glycoprotein representing a key protein within the complement system with marked functional versatility. Liver main site of synthesis; synthesized as single polypeptide chain (prepro-C3) by hepatocytes, monocytes, macrophages, fibroblasts and endothelial cells. Two-chain molecule, α- and β-chains disulfide linked. N-linked carbohydrate (1.7%) at residues 63 and 917. Reactive β-cysteinyl-γ-glutamyl thiolester bond in α-chain (Cys-988 and Glu-990) critical for biological functions and for covalent attachment to other soluble molecules or to receptive surfaces via ester or amide bond. Activation of circulating protein by proteolytic cleavage results in generation of biologically active fragments.
Structure	Two-chain molecule, chains linked by one disulfide bond and by noncovalent forces. Little information on tertiary structure; two-domain shape; flat ellipsoid with a second smaller flat domain. Structure of the cristalline fragment C3a known.
Molecular Weight	185,000 (SDS-PAGE); α-chain: 115,000, β-chain: 75,000.
Sedimentation Coeff.	9.5 S
Isoelectric Point	5.9
Extinction Coeff.	9.7 (280 nm, 1%, 1 cm); 1.82×10^5 mole^{-1} cm^{-1} (280 nm).
Enzyme Activity	Fragment C3b has substrate modulating function for substrate C5. C3b is component of C5-convertase of the classical pathway and of the C3-convertase of the alternative pathway of complement activation (C3bBb). Enzyme activity stabilized by properdin or autoantibodies, inactivation by several regulatory proteins.
Coenzymes/Cofactors	Association with fragments of other complement proteins required for enzyme activity.
Substrates	C5 and C3 substrates of C3b-containing convertases. No efficient chromogenic substrates available.
Inhibitors	No efficient inhibitors known. C3 depletion in vivo achieved by cobra venom factor induced C3-cleavage. Thiolester of native C3 cleaved by strong nucleophiles like methylamin.
Biological Functions	C3-activation by cleavage through C3-convertases or other proteases into the active fragments C3a (anaphylatoxin) and C3b. Further cleavage of C3b into iC3b, C3c, C3dg, C3d, C3e, C3g, C3f. Feedback amplification reaction via C3-convertase of the alternative pathway. Interaction of various C3-fragments with soluble complement proteins and regulatory active proteins (C5, factors B, H, I, properdin, conglutinin), with membrane-associated regulatory proteins and with complement receptors CR 1 – CR 5 and C3a-receptor. Binding to complement receptors modulates cellular

	functions and cellular interaction. Responsive cells include macrophages, monocytes, B- and also T-lymphocytes, mast cells, platelets and granulocytes. Formation of membrane attack complex C5b – C9 triggered by C3b. C3 protein important in host defence reactions (e.g. opsonization), as mediator in inflammatory reactions, in the handling of immune complexes and as immunoregulatory active mediator.
Physiology/Pathology	Central function within the complement system. C3-fragments affect a variety of cellular functions after binding to specific receptors. Acute-phase protein. Absence of C3 or hypercatabolism result in severe, frequently fatal recurrent infections. Importance as diagnostic parameter to detect complement activation in the circulation or in the tissue.
Degradation	Fragments C3a and C3b as products of proteolytic activation. Inactivation of C3a by serum-carboxypeptidase N through removal of C-terminal Arg 77. Inactivation of C3b by stepwise cleavage, initially through factor I, requiring one of the plasma- or membrane-associated cofactors H, CR1, CR2, C4bp, DAF, or MCP, resulting in the formation of the cleavage fragments iC3b (inactive), C3c (containing β-chain and N-(22500) and C-terminal (39500) α-chain segments), C3dg as central segment of the α-chain and C3f (2000). C3dg further cleaved by tryptic enzymes into C3d and C3g. Elimination of C3 fragments by phagocytic cells in liver.
Genetics/Abnormalities	26 allelic variants with two common electrophoretic variants C3F (fast) and C3S (slow). C3 gene localized on chromosome 19. α-chain encoded by more than 24 exons. Less than 20 patients with genetic C3-deficiency known (predominantly females), suffering from increased susceptibilty to infections.
Half-life	2.5 days; fract. catabolic rate 1.7%/hr.
Concentration	Plasma: 1.3 g/L (range 1–2 g/L).
Isolation Method	Isolated from fresh frozen plasma through PEG precipitation, and, after removal of plasminogen by L-lysin-Sepharose, by ion exchange chromatography.
Amino Acid Sequence	Deduced from cDNA-sequence or by partial sequencing of the protein. Homology with C5 and C4. Thiolester region homologous with the corresponding region of C4 and α2-macroglobulin. Sequence homologies of C3a with the other anaphylatoxins (C4a and C5a).
Disulfides/SH-Groups	Linkage of α- and β-chain by one disulfide, N- and C-terminal α-chain segments disulfide-linked. Thiolester group in α-chain (Cys-Glu), additional intra-chain disulfides, predominantly in α-chain.
General References	*The third component of complement – chemistry and biology*, Lambris, J. D. (ed.) Springer, Berlin, 1990. Janatova, J. C3, C5 components and C3a, C4a and C5a fragments of the complement system. *Meth. Enzymol*, 1988, **162**, 579–625. Lambris, J. D. The multifunctional role of C3, the third component of complement. *Immunol. Today*, 1988, **9**, 387–393.
Ref. for DNA/AA Sequences	DeBruijn, M. H. L., and Fey, G. H. *Proc. Natl. Acad. Sci.* 1985, **82:** 708–712.

Molecular model according to J. D. Lambris (*Immunol. Today*, 1988).



Schematic representation: Sites of C3 cleavage by C3-convertases, factor I (I), elastase (E) and trypsin (T) and the C3-binding sites for complement receptors (CR1, CR2, CR3, C3a-receptor), C5, properdin (P), H and conglutinin are indicated.

C4 complement protein

R. Duncan Campbell and S. K. Alex Law

Synonyms	The fourth component of human complement; β_1E globulin
Abbreviations	C4; C'4
Classifications	Electrophoretic mobility: β_1 globulin
Description	A circulating plasma protein. Major synthesis site is the liver, but also synthesized by macrophages. Synthesised as a single chain (pro-C4) subsequently processed into three chains in the order of β-α-γ involving the removal of 4 residues (RKKR) between the β-α junction and the removal of a 26 residue peptide (.....RRRR) between the α-γ junction. The C-terminal of the γ chain is not modified. All three chains are interconnected by disulfide bonds. Carbohydrate content about 7% at 4 N-linked sites: one in the α-chain of high mannose type, and three in the β-chain with complex fucosylated oligosaccharides. Up to 3 sulphates per molecule at three tyrosine residues at the C-terminal end of the α-chain. Sulphated C4 is about 3-fold higher in haemolytic activity. Contains an internal thiolester between a cysteine and glutamine residue within the tetrapeptide CGEQ.
Structure	Globular molecule by sedimentation. Also globular but with irregular substructures by EM. Appears as a flat disc by X-ray and neutron scattering. Crystals of C4, or any of its fragments, have not been obtained.
Molecular Weight	202,000 Da (non-reduced); α-chain: 95,000, β-chain: 75,000, γ-chain: 33,000 (reduced) by SDS-PAGE). The values from the derived aa sequence of C4 are 71,651, 82,129, and 33,122 for the α, β, and γ chains respectively disregarding contributions from carbohydrates and other modifications.
Sedimentation Coeff.	10 S
Isoelectric Point	7.5 about, broad range spanning ± 0.5 pH
Extinction Coeff.	8.3 (280nm, 1%, 1cm)
Enzyme Activity	None (however, its covalent binding activity may be considered a pseudo-enzymatic activity)
Coenzymes/Cofactors	None
Substrates	None (covalent binding to hydroxyl and amino groups to form ester and amide linked complexes)
Inhibitors	Covalent binding activity may be inactivated by small amines such as ammonia, hydroxylamine, and methylamine. Hydroxamate may also act as efficient competitive inhibitor to covalent binding reaction.
Biological Functions	C4 is activated by C1 and serves as a cofactor for the activation of C2, C3 and C5 in the complement system. Activation involves the cleavage of a peptide bond and the release of a 77 aa peptide (C4a) together with a conformational change in the larger part (C4b). The thiolester becomes exposed and allows C4b to bind to the activation surface. Subsequent binding of C2 to C4b allows its efficient cleavage by C1 resulting in a

C4b2a complex. With C2a as the enzyme and C4b the cofactor, the C4b2a complex cleaves and activates C3, which, like C4, also has an internal thiolester and binds covalently on the cell surface. Some C3b molecules bind to the C4b covalently thus generating the C4b2a3b complex. Again, C2a acts as the enzyme with the C4b3b complex as cofactor to cleave and activate C5.

The covalent binding properties of the two C4 isotypes (see Genetics) are different. C4A binds exclusively with hydroxyl groups to form amide linked complexes and C4B reacts with amino and hydroxyl groups with apparently similar efficiency. This difference is caused by the change in the residue at position 1106. In C4B, the residue is a histidine, which attacks the thioester upon activation to form an acyl-imidazole intramolecular bond. The released thiol group from the thioester then acts as a base to catalyse the binding of compounds with hydroxyl groups to the acyl function of the acyl-imidazole intermediate. C4A has an aspartic acid at position 1106 which cannot mediate the above reaction. Molecules with amino groups react directly with the thioester to form amide linked complexes. This reaction is apparently very efficient since hydrolysis of the thioester, which may be considered a competing reaction, is not catalysed.

C4b has affinity for the complement receptor type 1 (CR1).

Physiology/Pathology

Cells coated with C4b can be phagocytosed by activated monocytes and neutrophils via CR1.

C4b also binds to immune complexes, which also bear C3b. The C4b and C3b complexes bind to CR1 on erythrocytes and it is thought that this is the major route in the transport of immune complexes to the liver and spleen for their ultimate clearance in primates.

Patients deficient in C4 usually suffer from immune complex disorders (see Genetics)

Degradation

C4b is limited in its active live-span by the control proteins of the complement system. CR1, which is a receptor for C4b, also serves as a cofactor for factor I, a serine protease that cleaves the α' chain of C4b at two sites to yield C4d, a small fragment of 46,000 Da which contains the thiolester binding site, and C4c, which is comprised of two fragments of the α' chain disulfide linked to the intact β and γ chains. Two other proteins which also serve as cofactor for this degradation are the C4b binding protein (C4BP) and the membrane cofactor protein (MCP). C4BP is a plasma protein, whereas MCP is surface bound and has a wide distribution on all cell surfaces.

Genetics/Abnormalities

The two isotypes, C4A and C4B, are coded for by tandem genes approx. 10 kb apart in the Class III region of the Major Histocompatibility Complex on chromosome 6. Each gene contains 41 exons. In general the gene at the first (or C4A) locus is approx. 22 kb in length. The gene at the second (or C4B) locus can be either 22 kb or 16 kb in length which is accounted for by the variation in size of intron 9. Variation in the number of genes present has also been observed. Thus in addition to the common situation of having two C4 loci per chromosome, the presence of one, three or even four C4 loci is not uncommon. The two isotypes are highly polymorphic with at least four C4A allotypes and three C4B allotypes occurring at over 1% frequency. The consensus structural difference between C4A and C4B lies in 4 residues in a stretch of 6: PCPVLD for C4A and LSPVIH for C4B. It is likely that C4A is more important in binding to immune complexes and C4B to cellular targets. About 10 % of the population are deficient in either C4A or C4B. Immune complex disorders are associated with C4A deficiencies.

	Because of the high level of polymorphism, C4A and C4B are common markers for MHC typing especially in extended haplotype analysis. C4A6: the only known allotype with no haemolytic activity. R to W substitution in β-chain destroys the binding of C5 to the C4b3b complex. All other activities of C4A6 are normal.
Half-life	Unknown
Concentration	200 - 600 mg/L, large range due to the variable number of genes present.
Isolation Method	Plasma proteins by ion exchange chromatography on QAE-Sepharose followed by Mono Q in FPLC system. Further purification may be achieved by affinity chromatography using monoclonal antibodies.
Amino Acid Sequence	Pro-C4 has 1725 residues exclusive of leader peptide. Mature molecule in three chains: β (656 residues), α (748 residues) and γ (291 residues). Does not contain any known structurally defined motifs. Homologous in its entirety to C3, C5, α_2-macroglobulin, pregnancy zone protein and related protease inhibitors. Contains the sequence CGEQ at which the thiolester is formed. The sequence also found in all homologues except C5, which has SAEA and does not have a thiolester.
Disulfides/SH-Groups	One between β and α; two between α and γ. Maximum number of intrachain disulfide bonds: β (2), α (3), γ (5). The extra C in C4A is not disulfide bonded and is not detected to be a free sulfhdryl even after denaturation.
General References	Law, S.K.A. and Reid, K.B.M. Complement (second edition). In: *Focus Series*, Male, D. (ed.), IRL Press, Oxford, 1995. Campbell, R.D. et al. *Exp. Clin. Immunogenet.* 1990, **7**:69-84. Aguado, B. et al. In: *HLA and MHC: genes, molecules and function.* Browning, M. and McMichael, A. (eds.), Bios Scientific Publishers, Oxford, 1996, pp 39-75. WHO-IUIS Nomenclature Sub-Committee. *J. Immunol. Methods* 1993, **163**:3-7. Yu, C.Y. et al. *EMBO J.* 1986, **5**:2873-2881. Dodds, A.W. et al. *Nature* 1996, **379**:177-179. Law, S.K.A. and Dodds, A.W. *Immunol. Today* 1996, **107**:105.
Ref. for DNA/AA Sequences	Belt, K.T. et al. *Cell* 1984, **36**:907-914. Belt, K.T. et al. *Immunogenetics* 1985, **21**:173-180. Yu, C.Y. *J. Immunol.* 1991, **146**:1057-1066.

C4b-binding protein

Andreas Hillarp and Björn Dahlbäck

Synonyms	C4-binding protein; C4b-INA cofactor; Proline-rich protein
Abbreviations	C4BP
Classifications	Electrical mobility: β_2-fraction
Description	Multimeric plasma glycoprotein, synthesized in the liver. The major form is composed of seven identical α-chains (549 aa) and one β-chain (235 aa); all chains linked by disulfide bonds. Other isoforms in plasma are composed of six α- and one β-chain or seven α- and no β-chain. The α- and β-chains are structurally similar and contain internally homologous repeats of approx. 60 aa residues in length. The repeat, denoted short consensus repeat or complement control protein repeat or Sushi domain, is found in a group of proteins classified as the superfamily of C3b/C4b-binding proteins. In addition to its C4b binding activity, C4BP circulates in non-covalent complexes with serum amyloid P-component and the vitamin K-dependent coagulation protein, protein S. Only the β-chain containing isoforms of C4BP bind protein S as the binding site is located on the β-chain.
Structure	An assymetric molecule as indicated by Stokes radii (115Å), calculation of a high frictional ratio (2.1) and high resolution electron microscopy. Images of C4BP revealed an octopus- or spider-like structure with seven thin, elongated tentacles emanating from a ringlike central core region. Each tentacle being identified to as an α-chain. The binding site for C4b appeared to be located on the peripheral end of each tentacle, whereas the single binding site for protein S appeared to be placed on a small distinct subunit (now known to as the β-chain).
Molecular Weight	570,000 (sedimentation equilibrium ultracentrifugation). α-chain: 70,000 (reduced SDS-PAGE), 61,500 (aa sequence). β-chain: 45,000 (reduced SDS-PAGE), 26,400 (aa sequence).
Sedimentation Coeff.	9.1 S
Isoelectric Point	6.60-6.75 (neuraminidase treated protein).
Extinction Coeff.	14.1 (calculated from α-chain); Intact protein: 9.3 (280nm, 1%, 1cm).
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Regulates the classical pathway of complement: By virtue of its binding capacity for C4b, it can inhibit the formation of C3-convertases (C4bC2a) and accelerate the decay of existing convertases, thereby leading to a down-regulation of an activated complement. C4BP also serves as a cofactor to Factor I-mediated proteolysis of C4b. Binding of protein S has no direct effect on complement function (i.e., binding of C4b and cofactor

activity). However, protein S bound to C4BP is not functionally active in the blood coagulation which might indicate that C4BP has regulatory functions in the coagulation system. In vitro experiments have shown that the interaction with serum amyloid P component inhibits the factor I-C4BP mediated degradation of C4b.

Physiology/Pathology

One case with a heterozygous deficiency associated with Behçets syndrom known. Acquired deficiency can be seen in hypocomplementic patients with systemic lupus erythematosus. An acute phase protein. Patients with an abnormal complexformation between C4BP and protein S have a selective deficiency of free protein S and suffer an increased risk for thromboembolic diseases. The plasma concentration of free protein S is probably the result of an equimolar relationship betweeen β-chain containing isoforms of C4BP and protein S.

Degradation

Unknown

Genetics/Abnormalities

Genes for the α- and β-chains are closely linked in a head-to-tail arrangement, separated by a short intergenic region of approx. 4 kb. The genes are linked to other genes encoding structurally and functionally related molecules in a gene cluster, termed regulator of complement activation gene cluster, located on the long arm of chromosome 1, band q32. Transcripts of the α- and β-chain genes yield messages of 2.5 and 1.0 kb resp. Three allelic variants termed C4BP*1, C4BP*2 and C4BP*3 with gene frequencies of 0.986, 0.01 and 0.004 resp.

Half-life

Unknown

Concentration

150 mg/L, range 105 - 305 (plasma).

Isolation Method

Isolated from plasma either by polyethylene glycol or ammonium sulphate precipitation followed by ionexchange chromatography, affinity chromatography on heparin-Sepharose and gel filtration chromatography. Residual protein S bound to C4BP can be removed by gel filtration chromatography in strong denaturant (e.g., >3M guanidinium hydrochloride).

Amino Acid Sequence

α-chain: The 491 N-terminal aa residues can be divided into eight internal homologous repeats, each of approx. 60 residues. The sequence is ended with a 58 residues long C-terminal nonrepeat region.
β-chain: The 175 N-terminal aa residues contains three repeats, homologous to those found in the α-chain. The repeats are followed by a nonrepeat C-terminal region, similar to the corresponding region in the α-chain.

Disulfides/SH-Groups

The major form of C4BP ($\alpha_7\beta_1$) contains 126 disulfides, no free sulfhydryl. The α-chain contains 16 intrachain and the β-chain contains 6 intrachain linkages.

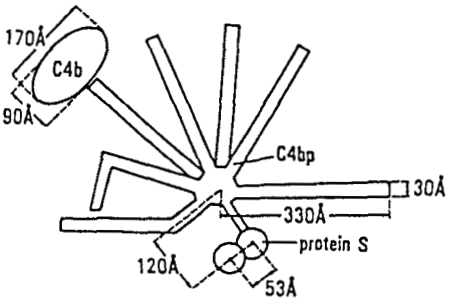
General References

Hillarp, P. and Dahlbäck, B. *Proc. Natl. Acad. Sci. USA* 1990, **87**:1183-1187.
Dahlbäck, B. *Thromb. Haemost.* 1991, **66**:49-61.

Ref. for DNA/AA Sequences

Chung, L.P. et al. *Biochem. J.* 1985, **230**:133-141.
Lintin, S.J. et al. *FEBS Lett.* 1988, **232**:328-332.
Hillarp, A. and Dahlbäck, B. *Proc. Natl. Acad. Sci. USA* 1990, **87**:1183-1187.

Molecular model according to B. Dahlbäck et al. *Proc. Natl. Acad. Sci. USA* 1983, **80:** 3461–3465

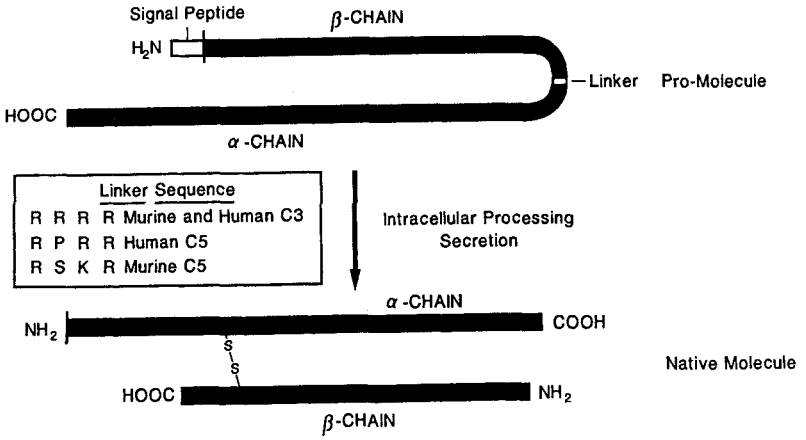


C5 complement protein

Rick A. Wetsel

Synonyms	None
Abbreviations	C5
Classifications	Serum glycoprotein
Description	C5 is biosynthesized in hepatocytes as an intracellular single chain precursor, pro-C5. The pro-molecule is processed, glycosylated, and secreted as a Mr 190,000 glycoprotein comprised of disulfide-linked α- and β-polypeptides with molecular weights of 115,000 and 75,000, respectively. C5 is cleaved by limited proteolysis by complement convertase enzymes into two biologically active fragments C5a (Mr 11,000) and C5b (Mr 179,000). Apporximately 3% of the mass of C5 is carbohydrate. Galactose (0.3%), mannose (0.6%), N-acetylglucosamine (1.5%), and N-acetylneuraminate (0.6%) have been detected.
Structure	Electron Microscopy: multilobal, irregular ultrastructure with dimensions of 104 x 140 x 168 Å. X-ray Scattering: R_G (nm) 5.3; R_{XS} (nm) 2.3 Neutron Scattering: $R_{G\text{-}C}$ (nm) 4.9; $R_{XS\text{-}C}$ (nm) 2.4; αG (x 10^{-5}) 13; αXS (x 10^{-5}) 13 CD Spectra: 17% α-helix; 20% β-Sheet f/f_0 1.45, indicating an asymmetric structure.
Molecular Weight	190,000 (SDS-PAGE, non-reduced) α-chain 115,000 (SDS-PAGE, reduced); β-chain 75,000 (SDS-PAGE, reduced).
Sedimentation Coeff.	7.9 S
Isoelectric Point	4.7-5.5
Extinction Coeff.	10.8 (C5); 10.9 (C5b); 3.5(C5a), all 280nm, 1%, 1cm.
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Plasma enzyme carboxypeptidase N (E.C. 3.412.7) removes the c-terminal arginine from the C5a activation peptide forming the C5 des-Arg derivative. On a molar basis, human C5a des-Arg expresses only 1% as much anaphylatoxic activity and 1% as much PMN chemotactic activity as does undigested C5a.
Biological Functions	C5 is cleaved during activation of either complement pathway into C5a and C5b fragments at a site (Arg74-Leu75) in the α-chain. After activation, the 11,000 Mr C5a peptide derived from the amino-terminus of the α-chain is released into the surrounding plasma and serves as a potent anaphylatoxin causing smooth muscle contraction, increased vascular permeability, basophil and mast cell degranulation and lysosomal enzyme release. In

	addition to its anaphylatoxic properties, C5a stimulates directed migration of neutrophils, eosinophils, basophils, and monocytes. Recent studies suggest that C5a also modulates the hepatic acute phase response. The C5b fragment initiates the assembly of the membrane attack complex (MAC) that mediates cytolysis of viral and bacterial pathogens.
Physiology/Pathology	May play direct or contributing role(s) in acute lung injury, septic shock, and immune complex induced glomerular lesions.
Degradation	In vivo: C5a and C5b In vitro: studies suggest a protease sensitive site in the α-chain that when cleaved results in the generation of 25,000 Mr and 90,000 Mr α-chain fragments that remain disulphide bonded to the parent molecule.
Genetics/Abnormalities	79 kb single copy gene of 41 separate exons on chromosome 9 at band positions q32-34. Several C5-deficient kindred have been described. Homozygous C5 deficient individuals display a propensity for severe, recurrent infections particularly to Neisserial species, including meningitis and extragenital gonorreha. The molecular genetic defects causing human C5 deficiency can be partly explained in the African-American population by nonsense mutations in exons 1 and 36. Murine C5 deficiency is caused by a 2-bp deletion in exon 7.
Half-life	Catabolic rate 1.9 % /hr.
Concentration	Serum concentration 0.075 g/L.
Isolation Method	Three main procedures are employed presently to isolate biologically active, highly pure C5 from serum. Wetsel, R.A., et al. *J. Immunol. Meth.* 1980, **35**:319-335. Kunkel, S.L., et al. *J. Immunol. Meth.* 1980, **35**:337-351. Hammer, C.H.. et al. *J. Biol. Chem.* 1981, **256**:2995-4006.
Amino Acid Sequence	C5 is a family member of structurally homologous proteins that include two other complement proteins, C3 and C4, as well as α_2-macroglobulin and the pregnancy zone proteins. C5 is the only member of this family that does not contain a thiolester bond.
Disulfides/SH-Groups	29 cysteine residues are contained in the C5 primary structure. Although no conclusive data has been obtained, it is suspected that 28 of these cysteines are involved in forming intra- and inter-chain disulfides. One known free cysteine is present in the C5a peptide.
General References	Lambris, J.D., Sahu, A. and Wetsel, R.A. The Chemistry and Biology of C3, C4, and C5. In: *Human Complement System in Health and Disease.* Frank, M. and Volanakis, J. (eds.) 1997, in press. Rother, K. and Till, G.O. The complement System. In: *The complement system,* Rother, K. and Till, G.O. (eds.) Springer-Verlag, Berlin, Germany, 1988, pp. 1-535. Haviland, D.L. et al. *J. Immunol.* 1991, **146**:362-368. Carney, D.F. et al. *J. Biol. Chem.* 1991, **266**:18786-18791. Wetsel, R.A. et al. *Biochemistry* 1988, **27**:1474-1482. DiScipio, R.G. et al. *J. Biol. Chem.* 1983, **258**:10629-10636. EMBL/Genbank acccession number: human M57729; murine J05234.
Ref. for DNA/AA Sequences	EMBL/GenBank accession number: human M57729; murine J05234.



C5 protein processing and subunit structure. This model diagrams the biosynthesis and intracellular processing of pro-C5 yielding the native circulating C5 molecule (190,000 Mr) with disulphide linked α- and β-chain polypeptides. The α- and β-chains have Mr of 115,000 and 90,000, respectively.

C6 complement protein

Dieter E. Jenne

Synonyms	Complement C6; the sixth component of the complement system
Abbreviations	C6
Classifications	Electr. mobility: β_2-fraction (pH 8.6)
Description	A circulating acute phase plasma protein, synthesized in hepatocytes, macrophages, fibroblasts, present in platelets and polymorphonuclear leukocytes. More than 90% are produced by the liver. A monomeric molecule. A glycoprotein with approximately 5% carbohydrate probably in the form of two biantennary oligosaccharides located at Asn-303 and Asn-834. Sialic acid content: 6 moles/mole of protein.
Structure	An elongated molecule of 18 ± 2 nm length and 6 ± 1 nm diameter in the central region. Not yet crystallized. Contents of secondary structural elements as estimated by circular dichroism spectroscopy: 12% α-helix, 29% β-sheet, 21% β-turns.
Molecular Weight	104,800 ± 5,700 (sedimentation equilibrium), 102,297 (aa sequence, mature 913 residue polypeptide chain)
Sedimentation Coeff.	6.0 S
Isoelectric Point	6.15 - 6.55
Extinction Coeff.	9.3 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	Unknown
Substrates	None
Inhibitors	The cytolytic potential of nascent terminal complement complexes is inhibited in the fluid phase by vitronectin (S-protein), clusterin (complement cytolysis inhibitor, SP-40,40, apolipoprotein J, sulfated glycoprotein 2), plasma lipoproteins, complement C8 and by the apolipoproteins A-I and A-II.
Biological Functions	One of the terminal complement proteins that participate in the terminal reaction of the complement cascade. The classical and alternative pathways enter into the common terminal pathway which is initiated by the cleavage of C5 after residue 74. The large C5b fragment then acquires a metastable binding site for C6 in the C-terminal region which contains two short consensus repeats and two Factor I modules. The water soluble C5b-6 complex subsequently reacts with C7, C8 and C9 in this order. On the surface of cellular plasma membranes, the cytolytic C5b-9 complex (M_r of 1.7 x 10^6) is formed and generates a hydrophilic channel across the lipid bilayer. Apart from lipid membranes, in the plasma or interstitial fluid, a cytolytically inactive, water-soluble complex is assembled which contains at least three additional plasma proteins: vitronectin (S-protein), clusterin (synonyms:complement cytolysis inhibitor, apolipoprotein J, SP-40,40, sulfated glycoprotein 2) and proteolytically cleaved antithrombin III.

1. Installment 1992 (revised 1997)

Physiology/Pathology	Essential for the terminal complement reaction. Two groups of C6-deficient individuals are distinguished, C6Q0 individuals show a complete deficiency of C6 (C6 levels in plasma are quantitatively zero) whereas C6SD (C6 subtotal deficiency) subjects have a C6 concentration of about 1-2% of the normal mean. C6SD subjects do not appear to be more susceptible to Neiserial infections than the population at large. C6Q0 individuals, however, show an increased risk for recurrent meningococcal and gonococcal infections, in particular with Neisseria meningitis. Complete C6 deficiencies are found at a frequency of 1:40, 000 among healthy blood donors in Osaka, Japan. Sera lacking C6 completely do not show any haemolytic activity towards sensitized erythrocytes and bactericidal activity towards complement-sensitive bacteria.
Degradation	Rapidly eliminated after formation of soluble terminal complement complexes (sTCC) by phagocytic cells of the liver and spleen.
Genetics/Abnormalities	Synthesized from a 3.5 kb mRNA. C6 is represented by a single copy gene in the human genome and consists of 18 exons disseminated over a DNA segment of 85 kb. Exon-intron organization is almost identical to that of the C7 gene except for the first exon. The C6 gene has been mapped to human chromosome 5p12-14 and is closely linked to the genes coding for C7 and C9. The C6 and C7 genes are separated by only 160 kbp and are orientated in a tail-to-tail manner. The physical distance between the C9 gene and the C6-C7 gene cluster is most likely more than 2.5 Mbp. C6 has a common charge polymorphism with similar allele frequencies of 0.6 for C6 A and 0.35 for C6 B in all major racial groups. The "acidic" allotype C6 A has a glutamic acid residue, the more basic allotype C6 B an Ala at position 77 (numbers refer to the mature protein). Individuals with subtotal complement C6 deficiency possess a truncated molecule which has lost 140 aa residues at the C-terminal end. Serum levels are only 1 to 2% of normal values. In two families with subtotal complement deficiency a change of the splice donor site of intron 15 from AG/gt to AG/gc was observed. The unspliced intron contains an in-frame stop 17 codons downstream from the exon-intron boundary which is consistent with the shortening of the size of the molecule.
Half-life	Soluble terminal complement complexes (sTCC) of rabbit plasma are cleared from the circulation within 30 to 50 min in rabbits.
Concentration	45 ± 16 mg/L (mean ± standard deviation)
Isolation Method	Biological source: whole human serum. Anti-C6 IgG affinity chromatography followed by DEAE Sephacel chromatography. Alternatively: polyethylene glycol 4000 precipitation from serum, plasminogen depletion on lysine-Sepharose followed by DEAE-Sephadex and dextrane sulfate-Sepharose chromatography.
Amino Acid Sequence	C6 is a secretory protein synthesized with a 21 residue long typical signal peptide. Its mature form consists of 913 aa residues. Nine cysteine-rich structural units (modules) have been identified: two thrombospondin type I repeats (residues 1 to 59 and residues 60 to 116), one LDL-receptor class A repeat (residues 117 to 156), a LDL-receptor class B (epidermal growth factor-like) repeat (residues 501 to 535), a third thrombospondin type I repeat (residues 541 to 590), two short consensus repeats (residues 621 to 680 and 681 to 744) and two complement control factor I modules (residues 745 to 819 and 838 to 913). The 34 kDa C-terminal fragment consisting of two short consensus repeats and two factor I modules mediates binding of the C5b fragment. C6 shares 33.5%, 30.1%, 28.4,

26.6% and 20% identical residues with C7, C8α, C8β, C9 and perforin, respectively, and is most similar to C7 except for an additional thrombospondin type I repeat at the N-terminus.

Disulfides/SH-Groups

32 disulfides, no interchain disulfides; no free sulfhydryls.

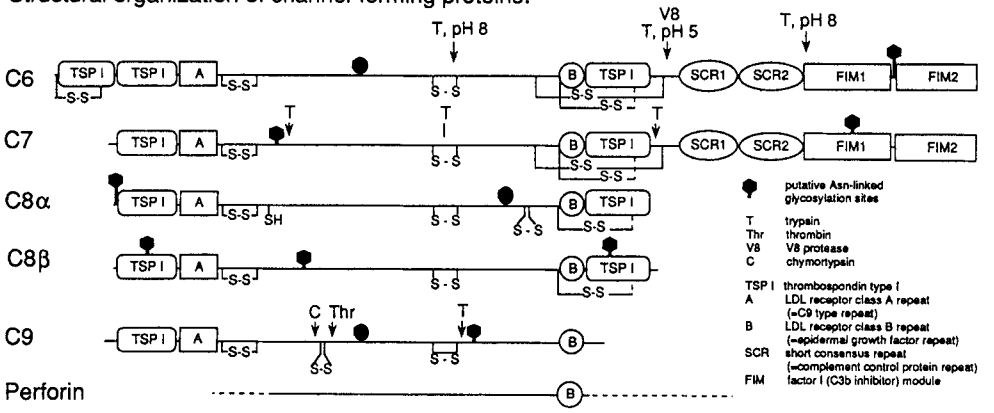
General References

Müller-Eberhard, H.J. *Ann. Rev. Biochem.* 1988, **57**:321-347.
Reid, K.B.M. and Day, A. J. *Immunol. Today* 1989, **10**:177-180.
Bhakdi, S. and Tranum-Jensen, J. *Rev. Physiol. Biochem. Pharmacol.* 1987, **107**:147-223.
Shin, M.L. and Carney, D.F. Cytotoxic action and other metabolic consequences of terminal complement proteins. In: *Progress in Allergy,* Vol. 40, Cytotoxic Mediators of Inflammation and Host Defense, Shin, M.L. (ed.), S.Karger AG, Basel 1988; pp. 44-81.
Rother, U. Component Deficiencies: 6. The Sixth Component. In: *Progress in Allergy,* Vol. 39, Hereditary and acquired complement deficiencies in animals and man, (Rother, K., and Rother, U. (eds.), S. Karger AG, Heidelberg 1986; pp. 283-288.

Ref. for DNA/AA Sequences

Haefliger, J.A. et al. *J. Biol. Chem.* 1989, **264**:18041-18051.
DiScipio, R.G. and Hugli, T.E. *J. Biol. Chem.* 1989, **264**:16197-16206.
Chakravarti, D.N. et al. *Proc. Natl. Acad. Sci. USA*. 1989, **86**:2799-2803.
Genbank/EMBL database: HUMC6A, HUMMHCC6A.*Swissprot database: CO6$HUMAN.
Hobart, M.J. et al. *Biochemistry* 1993, **32**:6198-6205.
Würzner, R. et al. *J. Clin. Invest.* 1995, **95**:1877-1883.

Structural organization of channel-forming proteins:



C7 complement protein

Richard G. DiScipio

Synonyms	None
Abbreviations	C7
Classifications	None
Description	C7 is a single chain plasma glycoprotein of 821 aa that is synthesized predominantly by mononuclear phagocytic cells, e.g. Kupffer cells. Carbohydrate accounts for about 5% of the molecular mass of C7, and the two asparaginyl-linked glycosylation sites are at aa positions 180 and 732.
Structure	Transmission electron microscopy has visualized C7 as an irregular ellipsoid of length 151 Å with one bulbous end having a diameter of 59 Å and the other of 42 Å.
Molecular Weight	97,000 (based on aa sequence and carbohydrate content)
Sedimentation Coeff.	5.6 S
Isoelectric Point	6.05; 6.17; 6.28; 6.40
Extinction Coeff.	12.4 OD units/cm/ml/dg (calculated from content of tryptophan, tyrosine and cysteine); 9.9 (refractive index measurements); 19.2 (aa analysis).
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Inhibitors unique to C7 are not reported; however, several plasma and cellular proteins are known to inhibit the complement lytic pathway. The plasma regulatory factors are vitronectin (S-protein) and SP 40,40 (Clusterin), and a cellular control factor is CD59.
Biological Functions	C7 participates in the lytic phase of the complement system. It combines with the soluble complex C5b,6 to form C5b-7, which can attach to target cell membranes. Subsequent addition of C8 and C9 results in the formation of the membrane attack complex, which is a protein walled transmembrane tubule.
Physiology/Pathology	C7 is essential for the lytic action of complement. Deficiences in C7 are usually associated with a predisposition to meningococcal and gonococcal infections. A combined deficiency for both C6 and C7 is also reported.
Degradation	C7 is stable and not substantially degraded in plasma.
Genetics/Abnormalities	The gene for C7 is ≈ 80kbp in length and is comprised of 17 exons. This gene is located on chromosomal region 5p13 and is separated from the C6 gene by ≈ 160 kbp. The genes for C7 and C6 are arranged in a 3' to 3' orientation. Deficiencies for C7 are homozygous autosomal recessive traits.
Half-life	61 hrs (blood circulation)

1. Installment 1992 (revised 1997)

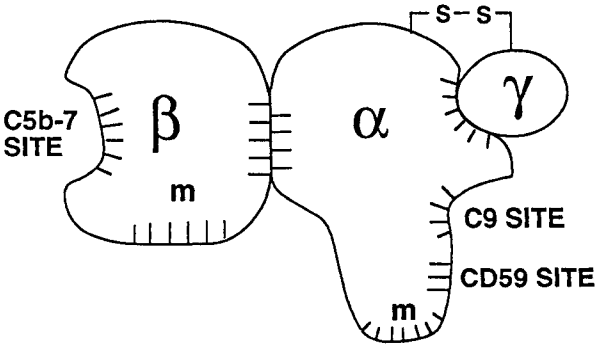
Concentration	Plasma: 55 mg/L (range 40-70 mg/L)
Isolation Method	C7 can be purified from citrated plasma after removal of the vitamin K dependent proteins with barium citrate absorption; 4.5-12 % polyethylene glycol precipitation; DEAE-Sephadex column chromatography; CM-Sephadex column chromatography; gel filtration on Sephacryl S300; and benzamidine Sepharose column chromatography.
Amino Acid Sequence	C7 consists of 821 aa and its precursor has a hydrophobic leader peptide of 22 aa. Modular homology: discrete autonomously folding cysteine rich segments (35-77 aa) are homologous to those found in thrombospondin, LDL-receptor, epidermal growth factor precursor, and complement factors H and I. Domain homology: the N-terminal two thirds of C7 is homologous to the corresponding regions from C6, C8, C9 and perforin. The C-terminal third of C7 consists of two Sushi modules and two FIM modules. This C-terminal domain of C7 has a counterpart in C6.
Disulfides/SH-Groups	C7 has 28 disulfide bonds; and no free sulfhydryl groups. All but 3 disulfide bonds are contained within autonomously folding modules.
General References	Podack, E.R. et al. *J. Immunol.* 1979, **123**:1071-1077. Muller-Eberhard, H.J. *Annu. Rev. Immunol.* 1986, **4**:503-528.
Ref. for DNA/AA Sequences	DiScipio, R.G. et al. *J. Biol. Chem.* 1987, **263**:549-560. Hobart et al. *J. Immunol.* 1995, **154**:5188-5193.

C8 complement protein

James M. Sodetz

Synonyms	None
Abbreviations	C8
Classifications	Electrophoretic mobility: γ_1
Description	A circulating serum protein of the complement system. Synthesized in liver, monocytes and fibroblasts. Consists of three non-identical subunits (α, β, γ) arranged as a disulfide-linked α-γ dimer that is noncovalently associated with β. The α and β chains contain N-linked carbohydrate (1-2 chains). The γ-chain has no carbohydrate.
Structure	Not yet crystallized.
Molecular Weight	151,000 (sedimentation equilibrium). α-chain: 64,000; β-chain: 64,000; γ-chain: 22,000 (all SDS-PAGE).
Sedimentation Coeff.	8.2 ± 0.4
Isoelectric Point	6.2-7.5 (estimation, not determined)
Extinction Coeff.	C8: 14.9; α - γ dimer: 11.8; β-chain: 13.7 (all 280nm, 1%, 1cm).
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Human C8 along with C9 is recognized by human CD59, a GPI-anchored membrane protein expressed on a wide variety of cell types. CD59 protects host cells from complement-mediated lysis by inhibiting the formation and function of C5b-9. Inhibition occurs as a consequence of CD59 binding to C8 and/or C9 within the nascent C5b-9 complex. Within C8, CD59 has been shown to recognize a site that is internal to residues 320-415 of the α-chain.
Biological Functions	One of five complement components (C5b, C6, C7, C8, C9) that associate in a sequential manner on target cell membranes to form C5b-9, the membrane attack complex (MAC) of complement which disrupts membrane permeability and contributes to cell lysis and destruction. Specific functions have been assigned to the C8 subunits.
Physiology/Pathology	C8 deficiencies are characterized by an absence of circulating α - γ or β but not both. Deficients generally exhibit increased susceptibility to Neisserial infections.
Degradation	Unknown
Genetics/Abnormalities	The α, β and γ chains are encoded in separate genes. The α and β genes are located in a 3'-3' orientation on chromosome 1p32. The intergenic distance estimated from analysis of YAC clones is <23kb. The γ gene is located on 9q22.3-q32. The genomic organization of α and β has been

	determined. Exon number, boundaries and phases are similar for α, β and C9 and the corresponding homologous regions of C6 and C7, thus confirming that all five proteins are ancestrally related. The γ gene has been completely sequenced and its organization resembles that of the α-2u-globulin family. Protein polymorphisms have been identified for α - γ and β and some have been correlated with polymorphisms at the DNA level.
Half-life	Unknown
Concentration	≈ 80 mg/L in serum (range 70-90)
Isolation Method	Isolated from serum or plasma Fraction III by ammonium sulfate precipitation, CM and QAE-Sephadex ion-exchange chromatography. The α - γ and β subunits can be separated by gel filtration in 1.5 M NaCl.
Amino Acid Sequence	The α and β chains exhibit significant sequence similarity to each other and to C6, C7, and C9. All contain structural molecules found in thrombospondin, LDL receptor and EGF precursor. These modules probably have a role in mediating protein-protein interactions during C5b-9 assembly. Based on similarities in aa sequence and genomic structure, γ has been assigened to the α-2u-globulin (or lipocalin) family of widely distributed proteins that bind small lipophilic ligands.
Disulfides/SH-Groups	The α chain contains 29 cysteines which are all involved in intrachain disulfides except one that is linked to γ. The β chain contains 28 cysteines of which all are in intrachain disulfides. The γ chain contains 3 cysteines, of which 2 are in intrachain disulfide and one is linked to α.
General References	Sodetz, J.M. *Curr. Top. Microbiol. Immunol.* 1988, **140**:19-31. Kaufmann, T. et al. *Hum. Genet.* 1993, **92**:69-75. Kaufmann, K.M. and Sodetz, J.M. *Biochemistry* 1994, **33**:5162-5166. Michelotti, G.A. et al. *Hum. Genet.* 1995, **95**:513-518. Lockert, D.H. et al. *J. Biol. Chem.* 1995, **270**:19723-19728.
Ref. for DNA/AA Sequences	Rao, A.G., et al. *Biochemistry* 1987, **26**:3556-3564. Howard, O.M.Z., et al. *Biochemistry* 1987, **26**:3565-3570. Ng, S.C., et al. *Biochemistry* 1987, **26**:5229-5233. Genbank sequence accession numbers are α (M16974); β (M16973); γ (M17263).



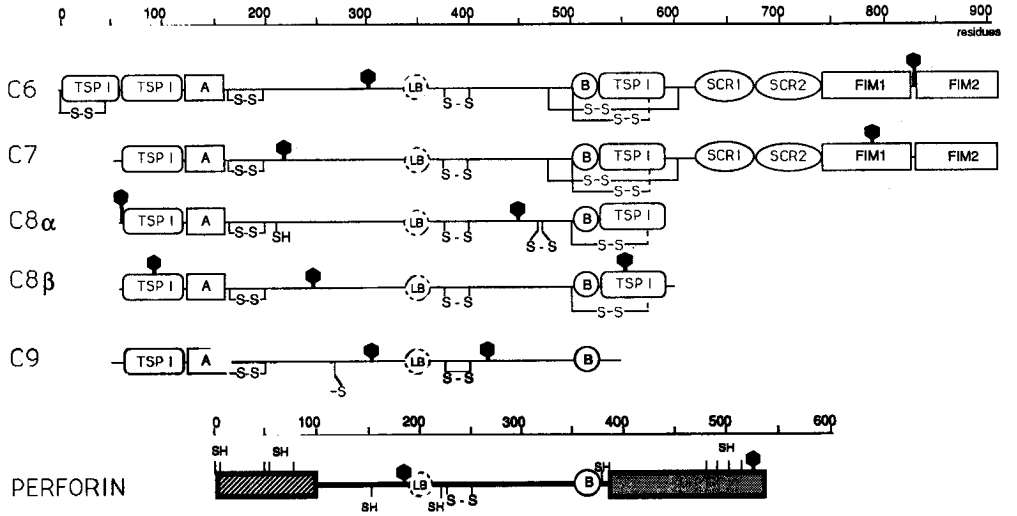
Functional sites within the C8 subunits. The β chain contains the site recognized by C5b-7 during assembly of the nascent C5b-9 complex. It also contains sites for interaction with the target menbrane (m) and α. The α chain contains sites for interaction with β, γ, the target membrane, the regulatory protein CD59 and C9. Locations of these sites within α and β are unknown. The γ chain retains its affinity for α after cleavage of the interchain disulfide, however it is not essential for C8 lytic activity. A functional role for γ has not yet been indentified.

C9 complement protein

Jürg Tschopp

Synonyms	None
Abbreviations	C9
Classifications	Electrical mobility: β-fraction
Description	A circulating plasma protein, synthesized in the liver and in monocytes. Present in plasma as monomeric, one chain polypeptide. Polymerizes into cylindrical structures upon interaction with nascent C5b-8 assembled on a complement-activating cell or in vitro in the presence of 10^{-4} M Zn^{2+}. This cylindrical structure, designated poly C9, is formed by 12–18 C9 molecules. C9 is a glycoprotein with approximately 8–15% carbohydrate content: two N-linked glycosylation sites (Asn-256, Asn-393); O-linked carbohydrates in the amino-terminal fragment (residues 1–244). One binding site for calcium.
Structure	Monomeric form: globular protein (4 × 8 nm) Poly C9: cylindrical polymerheight: 16 nm; outer diameter: 21 nm; inner diameter: 10 nm.
Molecular Weight	77,000 (sedimentation equilibrium), 71,000 (SDS-PAGE), 60,700 (aa sequence).
Sedimentation Coeff.	4.5 S
Isoelectric Point	4.95
Extinction Coeff.	9.1 (280 nm, 1%, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	Membrane-bound C5b-8 is required as cofactor for C9's lytic activity.
Substrates	None
Inhibitors	The lytic activity of the nascent C5b-9 complex formed in the fluid phase is inhibited by S-protein/vitronectin and clusterin. The lysis of homologous cells by C5b-9 is inhibited by membrane-bound CD59 and HRF/C8bp.
Biological Functions	Participates in the lytic event of the complement cascade: C9 is part of the lytic membrane attack complex (MAC). MAC assembly is initiated by proteolytic cleavage of C5. The fragment C5b interacts with C6, C7 and C8 forming the amphiphilic C5b-8 complex which becomes stably inserted into the target membrane. Membrane bound C5b-8 induces C9 binding and subsequent circular polymerization of C9 leads to transmembrane channel formation and cell death. If complement activation occurs in the absence of membranes, C9 becomes incorporated into the soluble terminal complex (sTTC) consisting of clusterin, S-protein/vitronectin, C5b, C6, C7, C8 and C9.
Physiology/Pathology	Essential for optimal lytic activity of the complement system.
Degradation	Eliminated from circulation by the liver, in particular after formation of the soluble terminal complement complex.

Genetics/Abnormalities	No abnormal variants of C9 described. High frequency of inherited C9 deficiencies. Gene localized on chromosome 5p13.
Half-life	Unknown
Concentration	Plasma: 60 mg/L
Isolation Method	Isolated from EDTA-plasma by polyethylenglycol precipitation and subsequent DEAE-Sephacel and hydroxyl-apatite chromatography.
Amino Acid Sequence	Overall aa sequence homology with other pore forming proteins of the complement cascade, i.e. C6, C7, C8 and perforin of cytolytic T cells. C9 contains 3 distinct, characteristic cysteine-rich motives (see figure).
Disulfides/SH-Groups	12 disulfide bridges clustered in cysteine-rich modules, no free sulfhydryls.
General References	Low, S. K. A. and Reid, K. B. M. *Complement, In focus*, IRL Press, Oxford 1988. Podacek, E. R. and Tschopp, J. Membrane Attack by complement. *J. Immunol.* 1984, **21**: 589–603. Müller-Eberhard, H. J. *Ann. Rev. Biochem.* 1988, **57**: 321–347.
Ref. for DNA/AA Sequences	Stanley, K. K., et al. *EMBO J.* 1985, **4**: 375–382. DiScipio, R. G., et al. *Proc. Natl. Acad. Sci. USA* 1985, **81**: 7298–7302. Accession number: EMBL K02766.



Structural organization of human perforin and late complement components. The drawing shows the various structural motives present in perforin, C6, C7, C8a, C8b, and C9: TSP 1, type I thrombospondin module; B, LDL receptor class B module (EGF module); SCR, short consensus repeat; FIM, factor I module; LB, candidate for lipid binding domain. Only cysteines outside motives or differing from invariant cysteines within motives are indicated. By analogy to the complement proteins, Cys 236 and 258 of perforin are likely to be disulfide-bonded. Putative asparagine-linked glycosylation sites are shown as black hexagons. Regions in perforin which show no detectable homology to complement proteins are boxed in with thick solid lines.

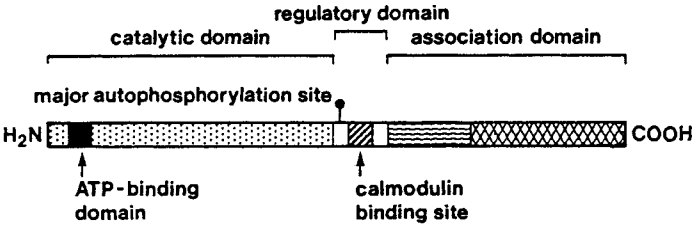
Ca^{2+}/calmodulin-dependent protein kinase II

Yoko Yamagata and Gerald Thiel

Synonyms	Multifunctional Ca^{2+}/calmodulin-dependent protein kinase, type II Ca^{2+}/calmodulin-dependent protein kinase
Abbreviations	CaM kinase II, CaM-PK II
Classifications	EC 2.7.10
Description	CaM kinase II is a multifunctional protein kinase that is regulated by Ca^{2+} and calmodulin. An increase in the cytosolic Ca^{2+}-concentration activates the enzyme leading to the phosphorylation of substrate proteins. CaM kinase II is one of the most abundant protein kinases in the brain but it is also present in other tissues like heart, muscle, liver and pancreas. In the brain only half of the enzyme is cytosolic, and the rest is associated with particulate structures. The holoenzyme is a large multimeric complex composed of 10 to 12 α or β/β' subunits. The ratio between α and β/β' subunits varies between different brain regions. The mRNAs for α, β and β' subunits are most enriched in the brain whereas the mRNAs of two additional subunits, γ and δ, are present in different tissues in the same amount as in brain. The β and β'subunits are encoded by a single gene through alternative splicing whereas the α, γ, and δ subunits are each encoded by distint genes. All subunits are highly homologous, consisting of catalytic, regulatory and association domains.
Structure	Brain CaM kinase II holoenzyme has been considered as a decamer or dodecamer composed of different subunits based on hydrodynamic studies. However, a recent electron microscopic study indicates that the holoenzyme is a homopolymer, either an octamer or a decamer, consisting of 8 or 10 catalytic domains in the peripheral particles and a central large particle formed by gathered association domains. The α subunits preferentially form decamers, the β subunits octamers.
Molecular Weight	500,000–700,000, depending on the purification procedure (gel filtration and sucrose density gradient centrifugation); Molecular masses deduced from rat cDNA clones: α-subunit: 54,111; β-subunit: 60,333; β'-subunit: 58,705; γ-subunit: 59,038; δ-subunit: 60,080
Sedimentation Coeff.	14.0–16.4 (varies with purification procedure).
Isoelectric Point	6.7–7.2, for both α and β subunits
Extinction Coeff.	Unknown
Enzyme Activity	CaM kinase II is a Ca^{2+} and calmodulin regulated protein kinase that phosphorylates serine or threonine residues. The preferred phosphorylation site consensus sequence is R-X-X-S/T, but numerous exceptions have been found. A synthetic peptide based on the phosphorylation site 2 of glycogen synthase, syntide 2 (sequence P-L-A-R-T-L-S-V-A-G-L-P-G-K-K) is frequently used for the measurement of kinase activity. CaM kinase II undergoes autophosphorylation at several serine and threonine residues. Autophosphorylation at Thr-286 (α subunit)/ Thr-287 (β/β'subunit) generates a Ca^{2+}-independent form of the kinase. The location of this autophosphorylation site close to the calmodulin-binding site suggests that autophosphorylation mimics the effect of calmodulin binding by stabilizing the active forms of the enzyme and prevents the refolding into an inactive conformation as a result of reduced Ca^{2+}-concentration. Classification according to the rules of the I. U. B.: Phosphotransferases with a protein alcohol group as acceptor called protein-serine/threonine kinases.

Coenzymes/Cofactors	Ca^{2+} and calmodulin
Substrates	CaM kinase II phosphorylates a variety of substrates in different cells underlining the "multifunctional" nature of this enzyme. Selected substrates are synapsin I, tyrosine hydroxylase, microtubule-associated protein 2 (MAP-2), smooth muscle myosin light chain, phospholamban, glycogen synthase, 6-Phosphofructo-1-kinase, ribosomal protein S6, cyclic AMP response element binding protein (CREB).
Inhibitors	A variety of calmodulin antagonists (trifluoperazine, calmidazolium, W-7) that bind to the calmodulin-binding site, and general protein kinase inhibitors (staurosporine, K-252a, K-252b, H-7) that interact with the ATP-binding site, are effective but not specific for CaM kinase II. The recently developed drug KN-62 (1-[N,O-bis (1,5-isoquinolinesulfonyl)-N-methyl-L-tyrosyl]-4-phenylpiperazine) seems to be a selective inhibitor that interacts with the calmodulin-binding site and blocks the activation by calmodulin. Synthetic peptides based on the autoinhibitory sequence of CaM kinase II (e.g. aa 281–309, 273–302, 281–302 of the α subunit) are relatively selective inhibitors. They interact with the ATP-binding site and/or substrate binding site.
Biological Functions	In the nervous system CaM kinase II seems to be involved in the regulation of basic neurological functions such as neurosecretion, neurotransmitter synthesis and microtubule assembly/disassembly. Furthermore, it has been suggested that CaM kinase II is involved in several aspects of synaptic plasticity, e.g. long-term potentiation, memory, and learning. Formation of Ca^{2+}-independent activity through autophosphorylation implicates that CaM kinase II may act as a molecular switch. The autophosphorylated CaM kinase II retains information about prior activating Ca^{2+} signals. The enzyme would be the "third messenger" that passes the stored information through phosphorylation of functionally significant substrate molecules. However, the proposed information storage function of autophosphorylated CaM kinase II has not been shown so far. Furthermore, the substrate proteins that are phosphorylated by a Ca^{2+}-independent form of the kinase have to be identified. Nevertheless the involvement of CaM kinase II in certain types of learning and memory seems clear since recent gene targeting experiments of CaM kinase II α-subunit revealed that transgenic mice lacking this subunit are deficient in the generation of long-term potentiation and exhibit impairments in spatial learning. In addition, CaM kinase II regulates a variety of metabolic processes in other cell types. Enzymes involved in lipid and phospholipid metabolism like diacylglycerol kinase, diacylglycerol acyltransferase, lysophosphatidate acyltransferase, and 1-O-alkylglycero-3-phosphocholine acetyltransferase, are strongly activated by CaM kinase II. Finally CaM kinase II seems also to be involved in transcriptional regulation by phosphorylating cyclic AMP-response element binding protein (CREB).
Physiology/Pathology	CaM kinase II has been postulated to be involved in several pathological conditions, e.g. epilepsy, cerebral ischemia and organophosphorus toxicity.
Degradation	Unknown
Genetics/Abnormalities	Unknown
Half-life	Unknown
Concentration	1% of total protein in the cerebral cortex, 2% of total hippocampal protein, 20–30% of total protein in postsynaptic densities.

Isolation Method	Fresh rat brains are homogenized under hypotonic conditions and centrifuged. CaM kinase II is isolated from the soluble fraction by DEAE-cellulose chromatography, hydroxyapatite chromatography, ammonium sulfate precipitation, gel filtration chromatography and finally calmodulin affinity chromatography.
Amino Acid Sequence	The aa sequence of the catalytic domain is highly homologous to that of other serine/threonine protein kinases, in particular to that of Ca^{2+}/calmodulin-dependent protein kinases. Within the regulatory domain, the major autophosphorylation site Thr-286 (α subunit) resides on the preferred CaM kinase II consensus sequence R-X-X-S/T. The calmodulin-binding site (aa 290–314 of the α subunit) shows high homology to the calmodulin-binding sites of other Ca^{2+} /calmodulin-dependent protein kinases, e.g. smooth muscle and skeletal muscle myosin light chain kinases and phosphorylase kinase. The aa sequence corresponding to the residues 281–309 is referred to autoinhibitory sequence, and seems to play a role in maintaining the kinase in the inactive form in the absence of Ca^{2+}/calmodulin.
Disulfides/SH-Groups	Unknown
General References	Colbran, R. J. et al. *Biochem. J.* 1989, **258:** 313–325. Hanson, P. I. and Schulman, H. *Annu. Rev. Biochem.* 1992, **61:** 559–601. Kennedy, M. B. et al. *Cold Spring Harbor Symposia on Quantitative Biology* 1990, **55:** 101–110.
Ref. for DNA/AA Sequences	Lin, C. R. et al. *Proc. Natl. Acad. Sci. USA* 1987, **84:** 5962–5966. Bennett, M. K. and Kennedy, M. B. *Proc. Natl. Acad. Sci. USA* 1987, **84:** 1794–1798. Tobimatsu, T. et al. *J. Biol. Chem.* 1988, **263:** 16082–16086. Tobimatsu, T. and Fujisawa, H. *J. Biol. Chem.* 1989, **264:** 17907–17912. Sunyer, T. and Sahyoun, N. *Proc. Natl. Acad. Sci. USA* 1990, **87:** 278–282. The nucleotide sequences for rat are deposited to the EMBL/GenBank data base (rat α-subunit: J02942; rat β-subunit: M16112; rat γ-subunit: J04063; rat δ-subunit: J05072)



Calcitonin and Procalcitonin

Jan A. Fischer and Walter Born

Synonyms	Thyrocalcitonin; N-Procalcitonin (PAS-57); Katacalcin (PDN-21)
Abbreviations	CT
Classifications	Polypeptide hormone
Description	Expression of the calcitonin/calcitonin gene-related peptide (CGRP) gene reveals procalcitonin or proCGRP through tissue-specific splicing of the initial gene transcript (Rosenfeld, 1983). In thyroid C-cells human procalcitonin is synthesized and cleaved into PAS-57, calcitonin and PDN-21 (see fig.). Calcitonin is the most important hypocalcaemic hormone. The biological activities of PAS-57 and PDN-21 remain to be elucidated.
Structure	All known calcitonins from fish to mammals have in common 32 aa with a 7 aa N-terminal ring structure linked by a disulfide bridge and C-terminal prolineamide. Calcitonin is a polypeptide with a hydrophobic N-terminal ring structure and hydrophilic middle and C-terminal regions.
Molecular Weight	3418: calcitonin; 6221: PAS-57; 2437: PDN-21 (human)
Sedimentation Coeff.	Unknown
Isoelectric Point	7.5, approx.
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	Unknown
Substrates	None
Inhibitors	N-terminal truncation of salmon calcitonin leads to calcitonin antagonists (Feyen 1992). The most potent antagonist is salmon calcitonin-(8-32). Serum antibodies to salmon calcitonin in humans (50% aa homology) may inhibit the action of the hormone.
Biological Functions	The secretion of calcitonin, PAS-57 and PDN-21 is stimulated by the intravenous administration of calcium (Born, 1991). But the levels are within the reference range during chronic hypercalcaemia. Calcitonin lowers serum calcium levels through inhibition of bone resorption and stimulation of the urinary calcium excretion. Inhibition of bone resorption by calcitonin is obtained through suppression of osteoclastic activity mediated by stimulation of 3',5'-cyclic AMP and raised cytosolic calcium. In the kidney, calcitonin stimulates the excretion of calcium, phosphate, sodium and chloride and enhances 1,25-dihydroxyvitamin D production (Bijvoet, 1971; Horiuchi, 1979). Calcitonin receptors have seven transmembrane domains and are linked through G proteins to adenylate cyclase activation and to intracellular calcium mobilization (Lin, 1991). Structural and functional properties of calcitonin receptor isoforms, generated by alternative splicing of gene transcripts, have been described in man (Albrandt, 1995; Gorn, 1995; Nussenzveig, 1995), pig (Zolnierowicz, 1994), rat (Sexton, 1993) and mice (Yamin, 1994). In the brain the presence of calcitonin and of its receptors has been revealed in the circumventricular region (Fischer,

1981). Central actions of calcitonin include antinociceptive properties as well as inhibition of gastric secretion and inhibition of food intake (Pecile, 1975; Morley, 1981; Fargeas, 1984).

Physiology/Pathology

The physiological relevance of calcitonin is difficult to delineate since massively raised levels recognized in patients with medullary carcinoma of the thyroid and undetectable or low levels seen in thyroidectomized patients do not bring about changes of serum calcium and bone mineral content. Raised serum levels of calcitonin are encountered in patients with sporadic or familial C-cell hyperplasia and medullary thyroid carcinoma. There they are used as tumor markers together with PAS-57 and PDN-21. In patients with malignant tumors and ectopic calcitonin production as well as severe illness and septicaemia the levels of procalcitonin unlike those of intact calcitonin are elevated (Assicot, 1993).

Degradation

Calcitonin is eliminated from the circulation primarily by the kidney (glomerular filtration and degradation in proximal renal tubule). Negligible proteolytic degradation occurs in the general circulation.

Genetics/Abnormalities

The human calcitonin gene is located on chromosome 11.

Half-life

8 ml/kg/min (metabolic clearance rate)

Concentration

Circulating levels of calcitonin are of the order of 20 ng/L monomeric calcitonin (5 pMole/L). Highest concentrations of calcitonin have been identified in thyroid glands, smaller amounts are present in the hypothalamus, the pituitary gland and in tumors of the lung.

Isolation Method

Isolated from thyroid glands through extraction by C_{18}-cartridges, and sequential high performance liquid chromatography.

Amino Acid Sequence

See structure

Disulfides/SH-Groups

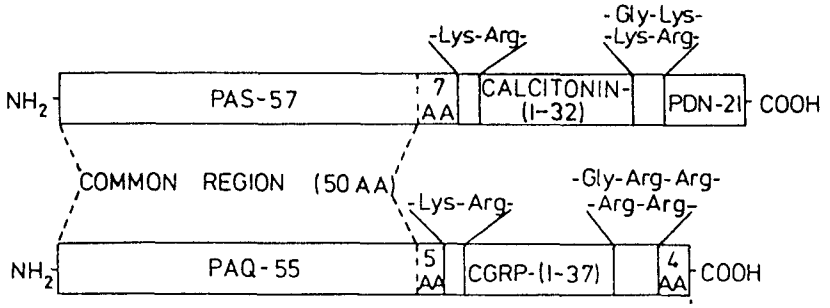
See structure

General References

Albrandt, K. et al. *Endocrinology* 1995, **136**:5377-5384.
Assicot, M. et al. *Lancet* 1993, **341**:515-518.
Bijvoet, O.L.M. et al. *N. Engl. J. Med.* 1971, **284**:681-688.
Born, W. et al. *Regul. Pept.* 1991, **32**:311-319.
Born, W. and Fischer, J.A. In: *Handbook of Experimental Pharmacology*, Mundy, G.R., Martin, T.J. (eds.) Springer, New York, 1993, **107**:569-616.
Deftos, L.J. and Roos, B.A. Bone *Miner. Res.* 1989, **6**:267-316.
Fargeas, M.J. et al. *Science* 1984, **225**:1050-1052.
Feyen, J.H.M. et al. *Biochem. Biophys. Res.Commun.* 1992, **187**:8-13.
Fischer, J.A. et al. *Proc. Natl. Acad. Sci. USA* 1981, **78**:7801-7805.
Gorn, A.H. et al. *J. Clin. Invest.* 1995, **95**:2680-2691.
Horiuchi, N. et al. *Biochem. J.* 1979, **184**:269-275.
Huwyler, R. et al. *Am. J. Physiol.* 1979, **236**:E15-E19.
Lin, H.Y. et al. *Science* 1991, **254**:1022-1024.
Morley, J.E. et al. *Science* 1981, **214**:671-673.
Nussenzveig, D.R. et al. *Endocrinology* 1995, **136**:2047-2051.
Pecile, A. et al. *Experientia* 1975, **31**:332-333.
Sexton, P.M. et al. *Mol. Endocrinol.* 1993, **7**:815-821.
Yamin, M. et al. *Endocrinology* 1994, **135**:2635-2643.
Zolnierowicz, S. et al. *J. Biol. Chem.* 1994, **269**:19530-19538.

Ref. for DNA/AA Sequences

Rosenfeld, M.G. et al. *Nature* 1983, **304**:129-135.
Steenbergh, P.H. et al. *FEBS Lett.* 1986, **209**:97-103.



Schematic structures of human procalcitonin and proCGRP.

Calcitonin gene-related peptide

Jan A. Fischer and Walter Born

Synonyms	None
Abbreviations	CGRP
Classifications	Neuropeptide
Description	In humans and rats two genes (I,II or a,b) encode calcitonin and calcitonin gene-related peptide (CGRP) (Rosenfeld, 1983; Steenberg, 1986) (fig.). CGRP and calcitonin are synthesized in the nervous system and in thyroid C-cells, resprectively, and cleaved from CGRP and calcitonin precursor proteins (procalcitonin, proCGRP). In the nervous system so far unknown regulation mechanisms direct splicing to the deletion of the calcitonin encoding exon 4, and formation of proCGRP mRNA. CGRP I and II are encoded by different genes on the same chromosome, and differ in three aa in humans and one aa in rats. In the human CGRP II gene, translation stopcodons within a calcitonin-like coding sequence prevent its expression. CGRP is the most potent vasodilator known. Other biological effects are discussed below.
Structure	All known CGRP have in common 37 aa with a 6 aa ring structure between aa 2 and 7 linked by a disulfide bridge and C-terminal phenylalanineamide. CGRP is a neuropeptide with a hydrophobic N-terminal ring structure and hydrophilic middle and C-terminal regions.
Molecular Weight	3786: CGRP I (human)
Sedimentation Coeff.	Unknown
Isoelectric Point	> 10
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	Unknown
Substrates	None
Inhibitors	N-terminal truncation of human CGRP leads to CGRP antagonists. The most potent antagonist is human CGRP(8-37).
Biological Functions	CGRP is released through stimulation of voltage-dependent calcium channels, as well as through treatment with capsaicin from afferent nerve fibres. In small amounts, CGRP is released in calcium-dependent manner from thyroid C-cells. But acute infusions of calcium do not raise serum levels of intact CGRP in humans (Born, 1991). CGRP is the most potent vasodilator acting directly on vascular smooth muscle through activation of adenylate cyclase. 3',5'-cyclic AMP production is antagonized by CGRP(8-37). Furthermore, CGRP brings about relaxation of the rat spleen precontracted with norepinephrine, and of the uterus and the ureter (Sigrist, 1986; Born, 1993). In the heart, CGRP has positive chronotropic and inotropic effects not suppressed by labetalol (Gennari, 1985). Tissue specific N-glycosylation of rat CGRP receptors, but indistinguishable protein molecular mass

have been identified in the cerebellum, brainstem, spinal cord, liver, and spleen. Activation of adenylate cyclase by CGRP in peripheral organs alone provide evidence for CGRP receptor subtypes (Stangl, 1993). Besides, stimulation of 3',5'-cyclic AMP production by the linear analog, [acetamidomethyl-$Cys^{2,7}$] a-human CGRP, in the liver and not in the spleen is consistent with $CGRP_2$ and $CGRP_1$ receptor subtypes according to the classification of Dennis et al. (1989).
Co-localization of CGRP and acetylcholine in motor endplates and stimulation of the acetylcholine receptor à subunit give a rational for longterm effects of a neuropeptide (Fontaine, 1986). While human CGRP I is a potent vasodilator of skin and carotid arteries, human CGRP II inhibits gastric secretion (Beglinger, 1991).

Physiology/Pathology

The stimulation of renal blood flow by CGRP is compatible with a physiological function of the peptide (Kurtz, 1989). In view of the widespread distribution of CGRP in central and peripheral nerve fibres, and the paracrine mode of action, it is impossible to eliminate the sources of CGRP to define physiological actions. Progress may be expected from targeted inactivation of the calcitonin/CGRP genes by homologous recombination.

Degradation

CGRP is eliminated from synapses and muscular endplates through proteolytic degradation.

Genetics/Abnormalities

The human calcitonin genes I and II are located on chromosome 11.

Half-life

10 - 20 min.

Concentration

Circulating levles of CGRP are of the order of 10 ng/L (2 pMole/L). CGRP is found throughout the central and peripheral nervous systems. The highest concentrations are recognized in the spinal cord as well as in thyroid C-cells. Human CGRP I and II have, moreover, been identified in the thalamus, pituitary and medullary thyroid carcinoma (Petermann, 1987).

Isolation Method

Isolated from spinal cord and thyroid glands through extraction by C_{18} cartridges, and sequential high performance liquid chromatography steps.

Amino Acid Sequence

See structure

Disulfides/SH-Groups

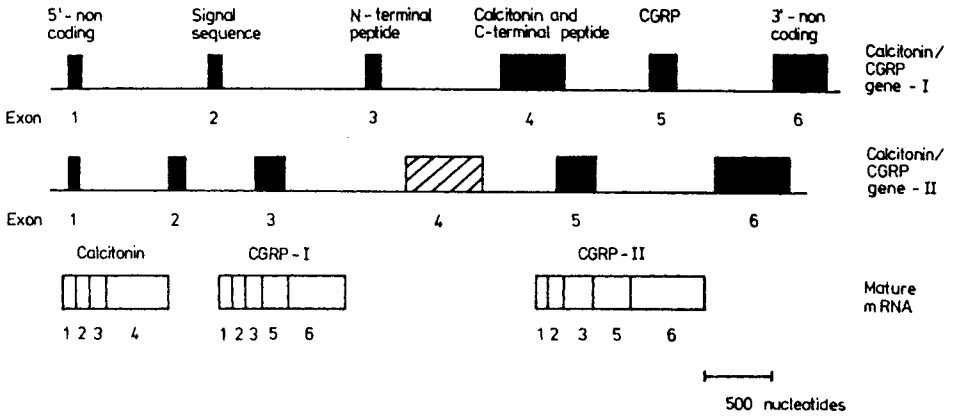
See structure

General References

Beglinger, C. et al. *Peptides* 1991, **12**:1347-1351.
Born, W., et al. *Regul. Pept.* 1991, **32**:311-319.
Born, W. and Fischer, J.A. In: *Handbook of Experimental Pharmacology*, Mundy, G.R. and Martin, T.J. (eds.) Springer, New York 1993, **107**:569-616.
Deftos, L.J. and Roos, B.A. *Bone Miner. Res.* 1989, **6**:267-316.
Dennis, T. et al. *J. Pharmacol. Exp. Ther.* 1989, **251**:718-725.
Fontaine, B. et al. *Neurosci. Lett.* 1986, **71**:59-65.
Gennari, C. and Fischer, J.A. *Calcif. Tissue Int.* 1985, **37**:581-584.
Kurtz, A. et al. *Kidney Int.* 1989, **36**:222-227.
Petermann, J.B. et al. *J. Biol. Chem.* 1987, **262**:542-545.
Sigrist, S. et al. *Endocrinology* 1986, **119**:381-389.
Stangl, D. et al. *Endocrinology* 1993, **132**:744-750.

Ref. for DNA/AA Sequences

Rosenfeld, M.G. et al. *Nature* 1983, **304**:129-135.
Steenbergh, P.H. et al. *FEBS Lett.* 1986, **209**:97-103.



Schematic structures of human CGRP I and II genes and mRNA products.

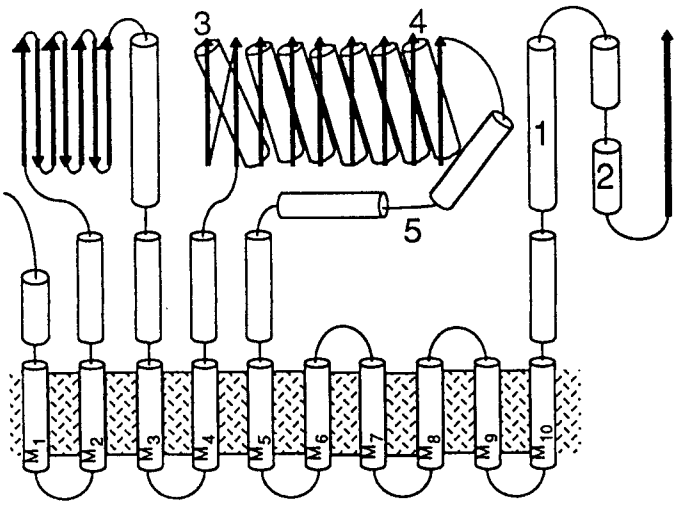
Calcium adenosinetriphosphatase (plasma membrane)

Ernesto Carafoli, Thomas Vorherr and Danilo Guerini

Synonyms	Ca^{2+} pump
Abbreviations	Ca^{2+} ATPase
Classifications	Membrane protein
Description	Protein consisting of a single polypeptide chain which is present in all eucaryotic plasma membranes. DNA cloning and protein sequencing have revealed several isoforms. Some are the product of different genes, some arise from the alternative splicing of mRNA. Isoforms show differences in the affinity to calmodulin and to ATP. The protein forms an aspartyl phosphate during the reaction cycle and has a Kd for Ca^{2+} of 0.2 μM in its high affinity state. The protein dimerizes (oligomerizes) in vitro. It exchanges Ca^{2+} for H^{+}. It is claimed to be localized in caveolae.
Structure	About 20% of the molecule (8-10 putative transmembrane helices, still undefined) is embedded in the plasma membrane. About 80% of the protein, carrying all functional domains, protrudes into the cytosolic space. Not yet crystallized.
Molecular Weight	138 kDa (SDS-PAGE), about 134 kDa (calculated from the cDNA)
Sedimentation Coeff.	Unknown
Isoelectric Point	5.81, 6.44, calculated for two isoforms
Extinction Coeff.	Unknown
Enzyme Activity	ATP driven Ca^{2+} transport: Average initial rate 1.86 μMoles of Ca^{2+} per mg of protein per min for the reconstituted enzyme. 0.93 μMoles ATP split per mg protein in 0.4% Triton, 0.05% phosphatidylcholine, 3.75 μMoles in the presence of calmodulin.
Coenzymes/Cofactors	None (however, calmodulin binds to the pump and stimulates it)
Substrates	Ca^{2+}, ATP
Inhibitors	Vanadate inhibits the ATPase activity and the associated Ca^{2+} transport reaction. Endogenous protein factors have been claimed to inhibit.
Biological Functions	Transport of Ca^{2+} from the cytosolic space to the extracellular space against the concentration gradient to keep a low intracellular Ca^{2+} concentration. Essential to the secondary messenger function of Ca^{2+}. Calmodulin, cAMP dependent phosphorylation, dimerization (oligomerization), acidic phospholipids and long chain polyunsaturated fatty acids stimulate the enzyme. Protein kinase C has also been claimed to activate.
Physiology/Pathology	Alterations of the Ca^{2+} ATPase have been described in the hypertensive state and patients with cystic fibrosis, Duchenne's muscular dystrophy and sickle cell disease. Modulation of the ATPases by hormones, e.g. insulin and oxytocin, has also been desribed. Inhibition of the activity in uncontrolled diabetics as a result of nonenzymatic glycosylation.

1. Installment 1992 (revised 1997)

Degradation	Partially proteolyzed by calpain with irreversible stimulation of the enzyme. Could be important in the process of cell aging.
Genetics/Abnormalities	Unknown. Chromosomal location determined for the 4 human genes: PMCA1 on chromosome 12 (q21-q23), PMCA4 on chromosome 1 (q25-q37), PMCA2 on chromosome 3 (3p26-p25), PMCA3 on chromosome X (Xq28).
Half-life	Unknown
Concentration	< 0.1% of the protein content of the plasma membrane.
Isolation Method	Isolated from human erythrocytes by solubilization of the membranes with Triton X-100 or other detergents and subsequent calmodulin affinity chromatography. Also isolated from other eucaryotic plasma membranes, e.g., heart, skeletal and smooth muscle. Some of the isoforms have been isolated from cultured cells after overexpression.
Amino Acid Sequence	Functionally interesting sites found in the hPMCA1b isoform (one of the two isoforms expressed in human erythrocytes). Domains homologous to calmodulin EEIPEEELAEDVEEIDHAERE (1079 - 1099), IHNFMTHPERFRIEDSEPHIPLIDDTDAEDD (1141 - 1170), GVKNSLKEANHDGDFGITLAELRALM (13 - 38). Site of aspartyl phosphate formation CSDKT (D475, 473 - 477). Fluorescein isothiocyanate binding site (putative ATP binding site), RIFSKGAS (K601, 597 - 603). Calmodulin binding site LRRGQILWFRGLNRIQTQIRVVNAFRSS (1100 - 1127). cAMP dependent phosphorylation site RNSS (S1178, 1175 - 1178). Protein kinase C dependent phosphorylation site QTQ (T 1116).
Disulfides/SH-Groups	Unknown
General References	Schatzmann, H.J. In: *Membrane Transport of Calcium,* Carafoli, E. (ed.), Academic Press, London, 1982: pp 41-108. Carafoli, E. *Physiological Reviews* 1991, **71**:129-153. Carafoli, E. *FASEB J.* 1994, **8**:993-1002.
Ref. for DNA/AA Sequences	Shull, G.E. and Greeb, J. *J. Biol. Chem.* 1988, **263**:8646-8657. Verma, A. K. et al. *J. Biol. Chem.* 1988, **263**:14152-14159. Strehler, E. et al. *Proc. Natl. Acad. Sci. USA* 1989, **86**:6908-6912. Shull, G.E. and Greeb, J. *J. Biol. Chem.* 1989, **264**:18569-18576. Strehler, E. et al. *J. Biol. Chem.* 1990, **265**:2835-2842.



The model, showing ten transmembrane domains, is based on hydrophathy plots and computer predictions. The number of transmembrane domains is provisional, and may have to be reduced downwards as direct experimental information on the membrane topology of the pump will become available. 1) Calmodulin binding domain. 2) c-AMP-dependent phosphorylation site. 3) Site of aspartyl-phosphate formation. 4) Site of FITC finding. 5) Flexible hinge, which permits sites 3 and 4 to come close to each other during the reaction cycle.

Caldesmon

Anindita Sen and Joseph M. Chalovich

Synonyms	150 k calmodulin binding protein from chicken gizzard
Abbreviations	CD; CaD (h-CaD, l-CaD); CDM; Cal_h; Cal_l
Classifications	High M_r / muscle and low M_r /non-muscle isoforms.
Description	Caldesmon is a heat-stable, Ca^{2+}- calmodulin binding protein which has been implicated in the regulation of smooth muscle contraction. This protein was originally isolated from contractile smooth muscle (chicken gizzard) and was consequently found to be ubiquitous in smooth muscle. Another low-molecular weight isoform of caldesmon has been identified in many non-muscle avian and mammalian tissues. The cDNAs for both h- and l-CaD are identical in the C- and N-terminal regions except for the insertion of Ala-508 in l-CaD. The central helical region is absent from the l-CaD cDNA. Skeletal muscle is devoid of caldesmon. In smooth muscle cells, caldesmon is associated with those actin filaments which contain myosin and tropomyosin and which are devoid of intermediate filament proteins.
Structure	Caldesmon is a highly extended, flexible protein. Hydrodynamic measurements have revealed a length of 74 nm for smooth muscle and 53 nm for non-muscle caldesmon. The secondary structure for caldesmon evaluated by CD spectra revealed that it contains 51% α-helix, 9% β-strands and 40% remaining structures. The caldesmon molecule is made up of two functional regions, which in muscle caldesmon, are separated by a helical region. The N-terminal domain which extends from aa residue 1 to residue 200 (using the numbering of Bryan, J. et al. *J. Biol. Chem.* 1989) has been shown to have an α-helical content of 43%. This domain encompasses the major myosin binding region. The second domain (201-439) is often referred to as the central helical region and is absent from the non-muscle isoform. This region is high in α-helix (59%), and contains 10 repeats of a 13-aa residue motif which is rich in charged side-chains arranged in an α-helical structure. The helix is stabilized by salt bridges between oppositely charged residues at positions of i and i + 4. The C-terminal domain which extends from residues 440-756 is made up of two regions. The C1 region (440-579) consists primarily of a long α-helix stabilized by intrahelical salt bridges. This region has low actin affinity and contains the Troponin-T like sequence (508-565). The C2 domain is high in Gly and Pro, has little periodic structure and has high actin affinity. Caldesmon can bend in half so that the actin-binding and myosin-binding regions can come into close contact with each other but the extended form appears to be more stable. No crystal structure has yet been determined for caldesmon.
Molecular Weight	*Smooth muscle isoform* (chicken gizzard): 120-150 kDa (gel electrophoresis), 86-89 kDa (sequence derived), 93 ± 4 kDa (equilibrium sedimentation). The difference between the derived molecular weight and that estimated from gel electrophoresis has been attributed to the high content of acidic aa. Caldesmon is monomeric under reducing conditions. The molecular weight of turkey gizzard caldesmon is 90 kDa (equilibrium sedimentation).

	Non-muscle isoform: 70-80 kDa (gel electrophoresis) and 59-60 kDa (sequence derived). (Rabbit liver caldesmon has a molecular weight of 66 kDa (equilibrium sedimentation)).
Sedimentation Coeff.	2.6S (avian gizzard) and 2.49S (rabbit liver).
Isoelectric Point	Heterogeneous; pI = 5.07 - 5.75 for muscle caldesmon.
Extinction Coeff.	3.3 (280nm, 1%, 1cm) for chicken gizzard caldesmon. The calculated molar extinction coefficient for caldesmon is 33,270 M^{-1} cm^{-1} (280nm).
Enzyme Activity	Unknown
Coenzymes/Cofactors	Unknown
Substrates	Binds to actin, actin-tropomyosin and myosin.
Inhibitors	Ca^{2+}-calmodulin reverses many of the effects of caldesmon including the binding to actin, the inhibition of ATPase activity of myosin and the binding of caldesmon to myosin. Caldesmon binds to Ca^{2+}-calmodulin with a stoichiometry of 1:1 and an affinity constant between 3 X 10^5 M^{-1} and 1.3 X 10^7 M^{-1}. The main calmodulin binding site is on the C-terminal region of caldesmon. Other Ca^{2+} binding proteins, such as caltropin, S100 and calcimedin, also attenuate the inhibition of ATPase activity by caldesmon. It is unclear whether calmodulin or any other single factor is present in sufficient quantities to control caldesmon interactions with actin and myosin in smooth muscle. Caldesmon may also be phosphorylated at several sites. Non-muscle caldesmon is phosphorylated during mitosis by cdc2 kinase and subsequently dissociates from actin. MAP kinase has been shown to phosphorylate caldesmon in smooth muscle cells but there is no proof that phosphorylation by MAP kinase has a role in contraction.
Biological Functions	The C-terminal region binds to actin with an affinity of about 10^7 M^{-1}. Reported stoichiometries in solution range from 1 caldesmon per 7 actin monomers to 1:20. The C-terminal region inhibits both the stimulation of myosin catalyzed ATP hydrolysis by actin and the binding of actin to myosin. Inhibitory activity is maximal when tropomyosin is bound to actin. The N-terminal region of caldesmon binds to myosin with an affinity between 10^5 and 10^7 M^{-1}. Caldesmon may cross-link actin and myosin together under conditions where the binding of myosin to actin is weak as in the presence of ATP (K = 10^4 M^{-1}). The binding of caldesmon to both myosin and to actin are attenuated by phosphorylation of caldesmon and by the binding of Ca^{2+}-calmodulin to caldesmon. Caldesmon also reportedly binds microtubules and to phospholipids suggesting an organizational role.
Physiology/Pathology	The physiological role has not been proven but appears to be a regulator of contraction and possibly involved in organization of contractile apparatus. Exogenously added caldesmon causes an increase in the level of myosin phosphorylation required for maximum tension production in smooth muscle fibers. An inhibitor of caldesmon binding to actin causes activation of smooth muscle contraction. Addition of caldesmon to skeletal muscle fibers (lacking caldesmon) results in an inhibition of both force and the amount of myosin bound to actin. Caldesmon may facilitate some motile processes such as receptor capping and granule movement perhaps by crosslinking myosin and actin. Stimulation of motility could occur if part of the actin filament remained active (i.e. partial decoration by caldesmon and no long range negative cooperative interactions) and if there is

sufficient slippage between caldesmon-actin or caldesmon-myosin bond so that actin can slide past the myosin. The ability of caldesmon to bind to myosin and to actin means that caldesmon may also have a function in organizing the contractile apparatus of non-striated muscle cells.

Degradation

Unknown

Genetics/Abnormalities

A single gene (100-150 kb long) encoding chicken caldesmon and consisting of 17 exons has been identified and subsequently cloned. A single splicing pattern has been noted for all smooth muscle cDNAs sequenced thus far. The non-muscle isoform (in human, 14 exons) arises by alternative splicing. The regulation of h- and l-CaD expression depends on selection of two splice sites within exon 3. The human CaD gene has been mapped to a single locus 7q33-q34. The two transcripts encoding the smooth and non-muscle isoforms are 4.8 and 4.1 kb in length, respectively. Genomic analyses have revealed two promoters for caldesmon. The gizzard-type promoter has higher activity than the brain-type promoter. The gizzard-type promoter displays cell-specificity which is dependent on an essential cis-element CArG1. Recently, molecular cloning of a complex translocation in a Burkitt lymphoma cell line identified a new gene (BCL7A), the protein product of which has homology with caldesmon in the helical central region and in the C-terminal region.

Half-life

Unknown

Concentration

The molar ratio of actin to caldesmon in phasic (visceral) smooth muscle was 20-50 while that in tonic (vascular) smooth muscle it was 200. Some laboratories report a constant molar ratio of ~30 regardless of the muscle type.

Isolation Method

Purified from heat-treated extracts of smooth muscle myofibrils using cibracon affinity, calmodulin affinity and DEAE column chromatography.

Amino Acid Sequence

The aa sequence of the 20 kDa chymotryptic fragment of caldesmon from Lys-579 to the C-terminal Pro-756 is shown below. This fragment has the actin and calmodulin binding properties and also inhibits the actin-activated ATPase activity of myosin.

K_{579}CFSPKGS SLKIEERAEF LNKSAQKSGM KPAHTTAVVS
KIDSRLEQYT SAVVGNKAAK PAKPAASDLP VPAEGVRNIK
SMWEKGNVFS SPGGTGTPNK ETAGLKVGVS SRINEWLTKT
PEGNKSPAPK PSDLRPGDVS GKRNLWEKQS VLKPAASSSK
VTATGKKSET NGLRQFEKEP$_{756}$

The region on caldesmon which contains the Troponin-T like sequence (508-565) is shown below.

E_{508}LDELKKR REERRKILEE EEQKKKQEEA ERKIREEEEK
KRMKEEIERR RAEAAEKRQK$_{565}$

Disulfides/SH-Groups

Two Cys-residues at positions 153 and 580. These are believed to be reduced *in vivo* but they can be cross-linked together by oxidation.

General References

Chalovich, J.M. and Pfitzer, G. Structure and function of the thin filament proteins of smooth muscle. In: *Cellular aspects of smooth muscle function.* Kao, C.Y. and Carsten, M.E. (eds.) Cambridge Univ. Press, New York, in press.

Chalovich, J.M. et al. *Annals of the New York Academy of Sciences* 1990, **599**:85-99.

Chalovich, J.M. *Pharmac. Ther.* 1992, **55**:95-148.

Marston, S.B. and Huber P.A.J. Caldesmon. In: *Biochemistry of smooth muscle contraction.* Barany, M. (ed.) Academic Press, Inc., San Diego, 1996, pp 77-88.

Ref. for DNA/AA Sequences

Bryan, J. et al. *J. Biol. Chem.* 1989, **264**:13873-13879.
Humphrey, M. J. et al. *Gene* 1992, **112**:197- 204.
Hayashi, K. et al. *Proc. Natl. Acad. Sci. USA* 1992, **89**:12122-12126.
Haruna, M. et al. *Biochem. Biophys. Res. Commun.* 1993, **197**:145-153.

Calmodulin

Wai Yiu Cheung and Dennis L. Merat

Synonyms	Phosphodiesterase activator; modulator protein; calcium-dependent regulatory protein
Abbreviations	CaM
Classifications	Ca^{2+}-binding protein
Description	A small, highly conserved, heat-stable protein present in all eukaryotic cells. Single polypeptide chain of 148 amino acids. Four Ca^{2+}-binding sites: two high affinity (sites III and IV, $K_d = 1$ to 3×10^{-6} M) and two low affinity (sites I and II, $K_d = 1$ to 2×10^{-5} M). Covalent modifications include acetylation (amino terminus) and methylation (on lysine 115).
Structure	The molecule consists of two globular lobes separated by an exposed seven-turn alpha-helix; each lobe has two Ca^{2+}-binding domains. The length of the molecule is 65 Å and each lobe has dimensions of 25x20x20 Å (crystallography, 2.2 Å resolution; Ca^{2+}-loaded; bovine brain). Binding of Ca^{2+} results in increased alpha-helicity and surface hydrophobicity.
Molecular Weight	16,790 (human aa sequence); 17,000 ($CaCl_2$, SDS-PAGE, bovine brain); 19,000 (EGTA, SDS-PAGE)
Sedimentation Coeff.	1.85–2.0 S ($CaCl_2$); 1.83 S (EGTA, EDTA)
Isoelectric Point	3.9–4.3
Extinction Coeff.	1.8 ($CaCl_2$; 276 nm, 1%, 1 cm); 2.0 (EGTA, EDTA; 276 nm, 1%, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	Ca^{2+} binding to calmodulin required for activation of target enzymes.
Substrates	None
Inhibitors	Drugs including phenothiazines (e.g. trifluoperazine), naphthalenesulfonamides (W compounds) and calcium channel blockers (e.g. felodipine); insect venom peptides (melittin and mastoparan); calmidazolium. Inhibitors appear to bind to hydrophobic regions on calmodulin that are presumably involved in interactions with the target enzymes or proteins.
Biological Functions	Mediates many of the intracellular effects of Ca^{2+}. When the intracellular Ca^{2+} concentration is increased to 10^{-6} or greater in response to stimuli, CaM binds Ca^{2+} and assumes a conformation favorable for interaction with target enzymes or proteins. When the Ca^{2+} concentration returns to steady state levels, CaM dissociates, and the enzymes/proteins return to their basal activity levels. CaM has been shown to stimulate the activities of a number of enzymes including CaM-dependent kinases I, II, and III, glycogen synthase kinase, phosphorylase kinase, myosin light chain kinase, calcineurin, inositol 1,4,5-trisphosphate kinase, Ca^{2+}-adenosine triphosphatase, adenylate cyclase, guanylate cyclase, Ca^{2+}-dependent phosphodiesterase, NAD kinase, and nitric oxide synthase. CaM is the subunit of phosphorylase kinase. CaM may be involved in many cellular processes including: neurotransmission, muscle contraction, mitosis, and the regulation of cytoskeletal changes.

Physiology/Pathology	Elevated CaM levels have been reported in many diseased states, including psoriasis, hepatomas, and certain hematological disorders and in certain transformed cell lines. Overexpression of CaM results in changes in the cell cycle (due to reduction of G1 phase) and in the cytoskeleton. Some of the toxic effects of heavy metals such as cadmium, chromium, and lead may be due to activation of calmodulin by these metals, thus upsetting the normal regulation of CaM by the cellular flux of Ca^{2+}.
Degradation	May involve ubiquitination prior to proteolysis.
Genetics/Abnormalities	At least three distinct rat CaM genes are transcribed in a tissue-specific manner; pseudo-CaM genes also found. Human CaM is encoded by at least two or three genes.
Half-life	25 hrs, 3T3 cells; 18 hrs, transformed 3T3 cells
Concentration	Cerebral cortex (600 mg/kg tissue); testis (480); cerebellum (440); lung (150); adrenal gland (110); prostate (110); liver (110); kidney (90); and spleen (80). (rat tissue); Human erythrocyte concentration is 7 μM.
Isolation Method	Isolated from bovine brain by isoelectric precipitation and heat treatment, followed by phenyl-Sepharose chromatography. Some procedures have incorporated one or more of the following: salt fractionation, chromatography on anion exchange, gel filtration, phenothiazine-Sepharose, Affi-Gel 501, and/or melittin-Sepharose columns. Gram quantities have been isolated by ammonium sulfate fractionation, DEAE-cellulose, and then hydroxyapatite chromatography. Good tissue sources are bovine brain or testis.
Amino Acid Sequence	The amino acid sequences of the proposed Ca^{2+}-binding loops for rat calmodulin (slight species variation) are shown. These stuctures are part of helix-loop-helix (EF hand) structures common to many Ca^{2+}-binding proteins. (Ca^{2+}-binding residues are indicated by an asterick.) * * * * * * Loop I (20–31) DKDGDGTITTKE Loop II (56–67) DADGNGTIDFPE Loop III (93–104) DKDGNGYISAAE Loop IV (129–140) DIDGDGQVNYEE
Disulfides/SH-Groups	None (mammals); One sulfhydryl (some plants)
General References	Babu, Y. S. et al. *J. Mol. Biol.* 1988, **204:** 191–204. Manalan, A. S. and Klee, C. B. "Calmodulin". In: *Advances in Cyclic Nucleotide and Protein Phosphorylation Research* 1984, **18:** 227–278. Means, A. R. "Molecular Mechanisms of action of calmodulin". In: *Recent Progress in Hormone Research* 1988, **44:** 223–262. Winkler, M. A. et al. *Hypertension* 1987, **9:** 217–223.
Ref. for DNA/AA Sequences	Fischer, R. et al. *J Biol. Chem.* 1988, **263:** 17055–17062 (accession # J04046, cDNA, human). Sasagawa, T. et al. *Biochemistry* 1982, **21:** 2565–2569 (aa, human). SenGupta, B. et al. *J. Biol. Chem.* 1987, **262:** 16663–16670 (Accession #'s M19311, J03468, cDNA; A28479, aa, human).

Calpain

Koichi Suzuki and Hiroshi Kawasaki

Synonyms	Calcium Activated Neutral Protease (CANP); Calcium protease; Calcium dependent protease (CDP); Calcium activated factor (CAF); Calcium activated sarcoplasmic factor (CASF); Kinase activating factor (KAF); Receptor transforming factor (RTF)
Abbreviations	None
Classifications	EC 3.4.22.17
Description	Cytosolic cysteine proteinase requiring Ca^{2+} for activity. Heterodimer composed of a catalytic 80 kDa and a regulatory 30 kDa subunit. Two isozymes are known; μ-calpain and m-calpain with high and low Ca-sensitivities. Calpain dissociates into subunits in the presence of Ca^{2+} and the dissociated 80kDa subunit is the active species. Their catalytic subunits are distinct but the regulatory subunits are identical. Their contents in cells vary significantly and erythrocytes contain only μ-calpain. Several tissue-specific calpains have been identified. These include skeletal muscle specific calpain (p94) which interacts with titin/connectin, stomach specific calpain, smooth muscle specific calpain, etc.
Structure	Not yet crystallized. No physicochemical measurements.
Molecular Weight	100,000 - 110,000 (gel filtration), 81,888 (μ-calpain large subunit, aa sequence), 80,019 (m-calpain large subunit, aa sequence), 28,315 (small subunit, aa sequence)
Sedimentation Coeff.	Unknown
Isoelectric Point	4.5 (mobilities on cellulose acetate strip)
Extinction Coeff.	14.2 (280nm, 1%, 1cm) (rabbit m-calpain)
Enzyme Activity	Hydrolase acting on peptide bonds (peptide hydrolase)
Coenzymes/Cofactors	Ca^{2+} is essential for protease activity. Ba^{2+}, Sr^{2+} and some other metal ions can substitute for Ca^{2+}.
Substrates	Cytoskeletal and membrane proteins (fodrin, neurofilaments, etc), enzymes (protein kinase C, phosphorylase kinase, etc.), muscle proteins (troponin, tropomyosin, etc.) are presumed endogenous substrates. Casein is mostly used for assays. Succinyl-Val-Leu-Lys-methylcoumarylamide (MCA), Suc-Leu-Met-MCA, Suc-Leu-Tyr-MCA can be used as substrates but small peptides are generally poor substrates.
Inhibitors	Calpastatin, proteinaceous endogenous inhibitor (mw: 100,000) specific to calpain, contains four tandemly repeated domains of ca.140 residues, each of which inhibits one mole of calpain competitively. Calpastatin forms a complex with calpain only in the presence of Ca^{2+}. Kininogens, but not other family members of cystatins, inhibit calpain. E64, leupeptin, and carbobenzyloxy-Leu-norleucinal (calpeptin) are strong inhibitors.
Biological Functions	Precise biological function is unknown. Calpain has a restricted proteolytic activity. Even the susceptible proteins are cleaved only to large fragments

	and not to small peptides or amino acids. Presumed functions are; initiates degradation of intracellular proteins, especially muscle proteins and short-lived proteins, functions as processing protease to give active proteins or domains, affects cellular signal transduction mediated by Ca^{2+} mainly by hydrolyzing protein kinase C.
Physiology/Pathology	Clear pathological states due to unusual calpain levels are unknown. Calpain may be responsible for muscle atrophy in muscular dystrophy. The calpastatin level is decreased to 1/10 of control in Milan hypertensive rats and μ-calpain level is less than 50% in Montreal platelet syndrome, although direct effect of the calpain level on the phathological states is not clear.
Degradation	Unknown
Genetics/Abnormalities	The p94 gene has been identified as responsible for limb girdle muscular dystrophy type 2A. Various nonsense, frameshift, splice site mutations have been identified together with point mutations.
Half-life	Unknown
Concentration	Unknown
Isolation Method	Both isozymes are isolated from muscle, kidney, lung and hearts by chromatographies on DEAE-cellulose, Ultrogel and phenyl-Sepharose. Erythrocytes are good source for μ-calpain.
Amino Acid Sequence	The catalytic subunit contains a cysteine proteinase domain homologous to other cysteine proteinases like papain, cathepsins B, L, and H. The sequences around the active site cysteine and histidine residues are CQGALGDCWLLAAI and LVKGHAYSVT, respectively for both μ- and m-calpain large subunits. Large and small subunits contain a calmodulin-like domain with 4 consecutive EF-hand structures at the C-terminal region.
Disulfides/SH-Groups	Presumably no S-S bonds.
General References	Mellgren, R.L. and Murachi, T. *Intracellular calcium-dependent proteolysis*. CRC Press, Boca Raton, 1990. Wang, K.K.W. and Yuen, P.W. *Trends Pharmacol. Sci.* 1994, **15**:412-419. Saido, T.C. et al. *FASEB J.* 1994, **8**:814-822. Sorimachi, H. et al. *FEBS Lett.* 1994, **343**:1-5. Suzuki, K. et al. *Biol. Chem. Hoppe-Seyler* 1995, **376**:523-529. Sorimachi, H. et al. *J. Biol. Chem.* 1995, **270**:31158-31162. Richard, I. et al. *Cell* 1995, **81**:27-40.
Ref. for DNA/AA Sequences	Aoki, K. et al. *FEBS Lett.* 1986, **205**:313-317. Ohno, S. et al. *Nucleic Acids Res.* 1986, **14**:5559. Imajoh, S. et al. *Biochemistry* 1988, **27**:8122-8128.

Carbonic anhydrase I

Sven Lindskog

Synonyms	Carbonic anhydrase B; Carbonate dehydratase (isoenzyme I); Carbonate hydro-lyase (isoenzyme I).
Abbreviations	CA I; HCA I; CA B; HCA B.
Classifications	EC 4.2.1.1
Description	A zinc metalloenzyme present mainly in erythrocytes but also in the cytosol of some other cells, for example, in vascular endothelium and corneal epithelium as well as in certain surface epithelial cells of the ileum and colon. The molecule is monomeric with a single polypeptide chain of 260 aa residues. The N-terminus is acetylated.
Structure	The crystal structure has been determined to a resolution of 2 Å. The molecule has an approximately ellipsoidal shape with the dimensions 41 Å x 41 Å x 47 Å. The predominant secondary structure is a 10-stranded, twisted β-sheet going through the center of the molecule. All strand pairs except 6-7 and 9-10 are antiparallel. About 12% of the aa residues are found in 5 short helical segments, all of which are located on the molecular surface. The active site is a cone shaped cavity reaching almost to the center of the molecule. The zinc ion is located near the bottom of the cavity and tetrahedrally coordinated to His-94, His-96 (both from β-strand 4), His-119 (from β-strand 5) and a catalytically functional water molecule or hydroxide ion.
Molecular Weight	28,850 (calculated from aa sequence plus one zinc ion and an acetyl group).
Sedimentation Coeff.	3.2 S (water, 20°C, infinite dilution)
Isoelectric Point	6.6
Extinction Coeff.	16.3 (280nm, 1%, 1cm); 47,000 $M^{-1}cm^{-1}$ (280nm).
Enzyme Activity	The enzyme catalyzes the reversible hydration of carbon dioxide: $CO_2 + H_2O \rightarrow HCO_3^- + H^+$. Kinetic parameters for CO_2 hydration: $k_{cat} = 1.5 \times 10^5$ s^{-1}, K_m = 5 mM (50 mM 1,2-dimethylimidazole buffer, pH 9, 25°C). However, the activity depends on pH, buffer concentration and the buffer system used. The enzyme also catalyzes the hydrolysis of certain esters. The esterase pH-rate profile corresponds to a titration curve with pK_a 7.1 and maximal activity at high pH. The maximal value of k_{cat}/K_m for the substrate 4-nitrophenyl acetate is about 900 $M^{-1} s^{-1}$ at 25°C.
Coenzymes/Cofactors	One firmly bound zinc ion, which can be removed by dialysis against certain chelating agents, for example, 1,10-phenanthroline or 2,6-pyridinedicarboxylate. The cobalt(II)-substituted enzyme is catalytically active. Its spectroscopic properties have been widely used in studies of the metal center.
Substrates	Carbon dioxide and bicarbonate are the only substrates of known physiological significance. A frequently used chromogenic substrate is 4-nitrophenyl acetate.

Inhibitors	The most powerful inhibitors are certain aromatic and heterocyclic sulfonamides, which coordinate to the zinc ion via the N-atom as $R\text{-}SO_2NH^-$ anions. A frequently used sulfonamide inhibitor is 5-acetylamido-1,3,4-thiadiazole-2-sulfonamide (acetazolamide, Diamox). Another group of inhibitors are monovalent anions, the most potent of which are cyanide, sulfide (SH^-) and cyanate. The anions bind at the metal ion or in its vicinity. Imidazole is the only compound that has been shown to act as a competitive inhibitor with respect to carbon dioxide. It is a moderately weak inhibitor. Iodoacetate and bromopyruvate are active-site directed inactivators which react covalently with His-200 to produce derivatives with substantially reduced, but not abolished, catalytic activities.
Biological Functions	The specific biological functions of this particular carbonic anhydrase isoenzyme are unknown.
Physiology/Pathology	A red cell CA I deficiency has been reported. The lesion involves a R-246 →H mutation, probably resulting in an unstable enzyme. Individuals with this deficiency show no clinical symptoms. A low-activity form of CA I, probably having glutathione disulfide-linked to Cys-212, is found in red cells of patients with primary aldosteronism. The expression of the CA I gene might be under the control of thyroid hormones. The concentration of red cell CA I is markedly decreased in patients with thyrotoxicosis.
Degradation	After a single intravenous injection of purified, radiolabelled CA I in rats, the elimination from plasma was found to be biexponential with half-times of 7 min and 122 min. The rapid phase is probably due to filtration of unbound enzyme at the glomeruli and subsequent degradation by the proximal tubules of the kidney. The slow phase seems to be due to binding to a plasma protein yielding a complex having a molecular mass of 115 kDa.
Genetics/Abnormalities	The gene has a total size of about 50 kb. It has been assigned to the long arm of chromosome 8 (q22) and is closely linked to the genes for isoenzymes II and III. The gene order is CA1, CA3, CA2. The coding regions of the CA1 gene are found in 7 exons. Two short noncoding exons (1a and 1b) are located 36.5 kb and 9.5 kb, respectively, upstream of the first coding exon (1c). However, exon 1b is absent in the majority of transcripts in erythroid cells. There are two, tissue-specific promoters. An erythroid-specific promoter is located upstream of exon 1a, whereas another promoter, used in colon epithelial cells, is located upstream of exon 1c. About 20 genetic variants of CA I have been detected. In nine cases, the amino acid substitutions have been identified. These are D8G, R76Q, D86G, T100K, E102K, Q225K/R, D236V, G253R, T255R.
Half-life	See degradation
Concentration	Normal red cells contain 12 ± 3 (mean ± SD) mg CA I per g hemoglobin.
Isolation Method	Erythrocytes are the best source. Packed and saline-washed cells are lyzed by the addition of an equal volume of water. The hemoglobin is precipitated by the addition of 800 ml of 40% ethanol and 400 ml of chloroform per liter of hemolysate. The aqueous phase, obtained after centrifugation, contains CA I and CA II. These isoenzymes can be separated by affinity chromatography on a material containing sulfonamide groups (for example, 4-aminobenzenesulfonamide carbodiimide-coupled to a carboxyl-based cation exchanger). The final purification of CA I can be achieved by chromatography at pH 8.7 on an anion exchanger, such as DEAE cellulose.

Amino Acid Sequence	CA I belongs to a family of homologous isoenzymes with seven known members. The degree of sequence identity with CA II is 59%, with CA III 55%, with CA IV 32%, with CA V 47%, with CA VI 32% and with CA VII 52%. The active site of CA I contains a few conserved, isoenzyme-specific residues. These are Val-62, His/Gln-67 and His-200. The isoenzyme-specific functional properties seem to depend to a considerable extent on His-200.
Disulfides/SH-Groups	A single Cys residue at sequence position 212 reacts with 2-chloromercuri-4-nitrophenol but does not react with iodoacetate.
General References	Deutsch, H.F. *Int. J. Biochem.* 1987, **19**:101-113. Dodgson, S.J. et al. (eds.) *The Carbonic Anhydrases.* Plenum Press, New York 1991. Kannan, K.K. et al. *Proc. Natl. Acad. Sci. USA.* 1975, **72**:51-55. Ren, X. and Lindskog, S. *Biochim. Biophys. Acta* 1992, **1120**:81-86. Silverman, D.N. and Lindskog, S. *Acc. Chem. Res.* 1987, **21**:30-36. Engstrand, C. et al. *Eur. J. Biochem.* 1995, **229**:696-702. Kumar, V. and Kannan, K.K. *J. Mol. Biol.* 1994, **241**:226-232.
Ref. for DNA/AA Sequences	Amino acid sequence: Andersson, B. et al. *Biochem. Biophys. Res. Commun.* 1972, **48**:670-677. Lin, K.-T.D. and Deutsch, H.F. *J. Biol. Chem.* 1973, **248**:1885-1893 (SWISSPROT accession number, P00915). cDNA sequence: Barlow, J.H. et al. *Nucleic Acids Res.* 1987, **15**:2386 (EMBL accession number, X05014). Genomic sequences: Lowe, N. et al. *Gene* 1990, **93**:277-283 (EMBL accession number, M33987).

Carbonic anhydrase II

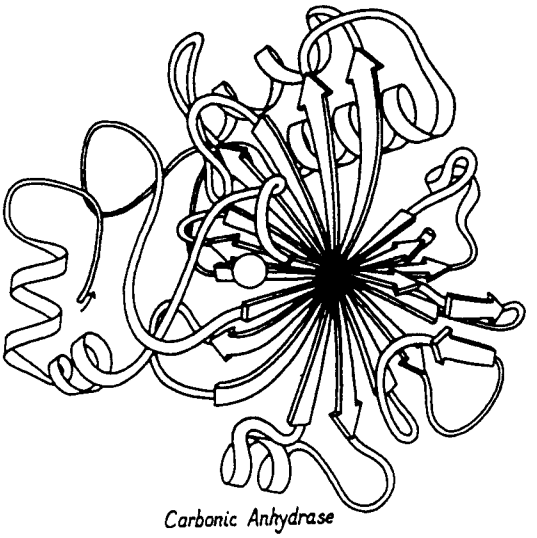
Teaster T. Baird, Jr., Eric D. Roush and Carol A. Fierke

Synonyms	Carbonate dehydratase
Abbreviations	HCA II; CA II
Classifications	EC 4.2.1.1; fastest known isozyme
Description	CA II is a cytosolic protein found in most tissues, although present in highest concentration in erythrocytes. It is a single polypeptide chain of 259 aa, with a single active site located at the bottom of a 15 Å deep, 15 Å wide cavity. The active site is formed by a Zn^{2+} ion binding to residues His-94, His-96, and His-119. A hydroxide molecule forms a fourth zinc ligand and acts as a nucleophile in catalysis. CA II has the highest turnover number (k_{cat} = 1.4 x 10^6 s^{-1}) of any CA isozyme and is one of the fastest enzymes extant. It can also be distinguished from other isozymes of carbonic anhydrase by decreased sensitivity to iodide inhibition, by basicity in isoelectric focusing, or by immunochemical methods.
Structure	The crystallographic structure of CA II has been refined to 1.54 Å. The primary structural characteristics is a 10-stranded twisted ß-sheet which encloses the active site pocket.
Molecular Weight	29,300 Da
Sedimentation Coeff.	3.3 S
Isoelectric Point	7.6
Extinction Coeff.	54,000 M^{-1} cm^{-1} (280nm)
Enzyme Activity	Catalyses the reversible hydration of CO_2 to bicarbonate and a proton. Also can catalyse the hydrolysis of esters and hydration of aldehydes.
Coenzymes/Cofactors	One Zn^{2+} ion is required for activity. Co^{2+} may substitute for Zn^{2+} with little loss of activity.
Substrates	CO_2, HCO_3^-, p-nitrophenyl acetate
Inhibitors	Molecules bearing a sulfonamide ($R\text{-}SO_2NH_2$) moiety, where R is an aromatic or heteroaromatic ring, or a halo-aliphatic group, are noncompetitive inhibitors of CO_2 hydration with binding constants in the micromolar to nanomolar range. Monovalent anions (including Cl^-, I^-, SCN^-, OCN^-, HS^-, etc.) inhibit with dissociation constants in the millimolar to micromolar range. Cu^{2+} and Hg^{2+} inhibit by binding to His-64 and interfering with the rate-limiting proton transfer reaction. A 79 kDa glycoprotein purified from porcine plasma , pICA, inhibits with a K_i of 1 nM at pH 8.0.
Biological Functions	The main biological function is to catalyze the interconversion of CO_2 and $HCO_3$and thereby: i) facilitate CO_2 transport; ii) provide H^+ and HCO_3^- to serve as counterions, as in the H^+/Na^+ exchange mechanism for recycling Na^+ from urine; iii) regulate the density of bone matrix. Also, the H^+ and HCO_3^- ions that are the products of CO_2 hydration are used to affect extracellular pH, for gastric secretions, and to stimulate secretion of bodily fluids such as aqueous humor.

Physiology/Pathology	Activity is found in almost every tissue type. Absence of CA II can lead to osteopetrosis, renal tubular acidosis, bvrain calcification, and mild mental retardation. Sulfonamide inhibitors of CA II are used in the treatment of glaucoma and altitude sickness. Renal tubular acidosis is a side effect of sulfonamide therapy.
Degradation	Unknown
Genetics/Abnormalities	The gene is approx. 9.2 kb long and contains 7 exons. All exons are required for coding of the complete enzyme. The gene has been linked to the long arm of chromosome 8 at q22, in conjunction with the genes for human carbonic anhydrases I and III. Complete absence of CA II (CA II deficiency syndrome) leads to osteopetrosis, renal tubular acidosis, mild mental retaradation, and cerebral calcification. There are several mutations in the CA II gene which cause CA II deficiency syndrome, including point mutations at His-107→Tyr, Tyr-10→Stop, the intron 2 donor site, and the intron 5 acceptor site. Other mutations known to lead to the CA II deficiency syndrome phenotype include single base deletion leading to a frameshift at Lys-228 and a 15-20 kb deletion including exon 1 or 2. Point mutations at Lys-18→Glu (CA II Jogjakarta) and Pro-237→His (CA II Melbourne), occur in CA II without apparent biological effect. So far, the only major change attributable to these mutations is a decrease in heat stability. In addition, there is a known allelic variant CA II_2 (Asn-253→Asp).
Half-life	4.2 days (in vitro, 48°C)
Concentration	Erythrocyte 20 μM, other tissues 5-10 μM – determined by inhibition characteristics and/or immunochemistry.
Isolation Method	CA II is commonly purified using sulfonamide affinity chromatography and elution with NaN_3. Ca II can also be purified by sequential ion exchange chromatography using DEAE and S-Sepharose resins. CA II can be eluted from S-Sepharose via a positive linear ammonium sulfate gradient. The most suitable source for large scale purification of CA II is the erythrocyte. CA II has also been cloned and overexpressed in E. coli.

Amino Acid Sequence	A member of carbonic anhydrase family.
	`--SHHWGYGK- HNGPEHWHKD FPIAKGERQS- PVDIDTHTAK`
	`YDPSLKPLSV SYDQATSLR--I LNNGHAFNVE FDDSQDKAVL`
	`KGGPLDGTYR LIQFHFHWG--S LDGQGSEHTV DKKKYAAELH`
	`LVHWNT-KYG DFGKAVQQPD GLAVLGIFLK VGSAKPG--LQK`
	`VVDVLDSIKT KGKSADFTN-F DPRGLLPESL DYWTYPGSLT`
	`TPPLLECVTW IVLKEPISVS SEQVLKFRKL NFNGEGEPEE`
	`LMVDNWRPAQQ PLKNRQIKAS FK`
	(– inserted to preserve alignment with other carbonic anhydrases).

Disulfides/SH-Groups	None
General References	Botrè, F. et al. *Carbonic Anhydrase from Biochemistry and Genetics to Physiology and Clinical Medicine*, VCH, Weinheim, 1991. Dodgson, S.J. et al. *The Carbonic Anhydrases: Cellular Physiology and Molecular Genetics*, Plenum Press, 1991. Tashian, R.E. *Adv. Genetics* 1992, **30**:321-356. Christianson, D.W. and Fierke, C.A. *Acc. Chem. Res.* 1996 (in press). Lindskog, S. and Liljas, A. *Curr. Opin. Struct. Biol.* 1993, **3**:915-920. Sly, W.S. and Hu, P.Y. *Annu. Rev. Biochem.* 1995, **64**:375-401.
Ref. for DNA/AA Sequences	Lin, K.T.D. and Deutsch, H.F. *J. Biol. Chem.* 1974, **249**:2329-2337. Montgomery, J.C. et al. *Nucl. Acids Res.* 1987, **15**:4678.



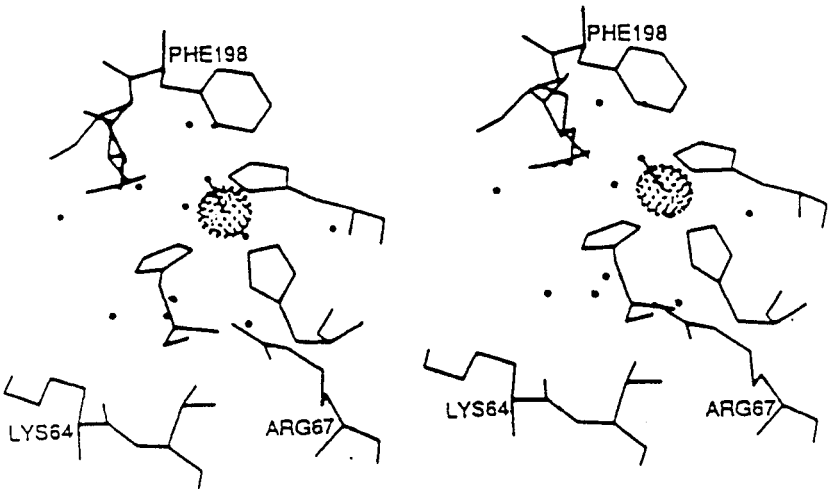
Ribbon drawing of HCA II, courtesy of Jane Richardson. Note the location of the Zn^{2+} ion deep within the active site pocket formed by the twisted β-sheet backbone.

Carbonic anhydrase III

David N. Silverman

Synonyms	Carbonate dehydratase III
Abbreviations	CA III
Classifications	EC 4.2.1.1
Description	A monomeric zinc metalloenzyme that catalyzes the reversible hydration of carbon dioxide to form bicarbonate and a proton.
Structure	The crystal structure of bovine CA III determined at 2.0 Å resolution is very similar to that of human red cell CA II. The foundation of this structure is a 10 stranded twisted β-sheet. The zinc is coordinated to three histidine residues and a water molecule and is located at the bottom of a funnel shaped active-site cavity with hydrophobic and hydrophilic sides.
Molecular Weight	29,300
Sedimentation Coeff.	2.8 S
Isoelectric Point	8.5 (bovine CA III)
Extinction Coeff.	20.7 (280nm, 1%, 1cm); molar extinction coefficient: 6.1 x 10^4 M^{-1} cm^{-1} (280nm).
Enzyme Activity	k_{cat} = 2 x 10^3 s^{-1} and k_{cat}/K_m = 3 x 10^5 M^{-1} s^{-1} for the hydration of CO_2 at pH 7.4 catalyzed by human CA III. This is the least efficient of the known carbonic anhydrase isozymes with activity about 100 times less than red cell CA II. CA III has very weak catalytic activity in the hydrolysis of 4-nitrophenyl phosphate and negligible activity in the hydrolysis of 4-nitrophenyl acetate.
Coenzymes/Cofactors	One zinc atom per molecule of CA III. The zinc-bound hydroxide is the catalytic unit in the hydration of CO_2.
Substrates	Carbon dioxide; bicarbonate.
Inhibitors	CA III is weakly inhibited by the classic carbonic anhydrase inhibitors. Acetazolamide (K_i = 300μM), chlorzolamide (0.3μM), and cyanate (K_i = 0.5μM) are inhibitors of feline CA III, each of which blocks catalysis by binding to the zinc.
Biological Functions	The function is probably related to the reversible hydration of CO_2 to form HCO_3^- and a proton but is uncertain because of difficulty in inhibiting CA III in tissues.
Physiology/Pathology	CA III is the predominant cytosolic protein in fibers of red skeletal muscle (Type I, slow oxidative) and in rat adipocytes. It is also found in the liver of adult male rats. CA III causes the enhancement of CO_2 diffusion in muscle cells and there is a possible role in acid-base homeostasis in the liver. However, a specific physiological function of CA III is uncertain; the low catalytic activity of CA III in hydration of CO_2 suggests another, yet unknown role for this enzyme.

Degradation	Unknown
Genetics/Abnormalities	The human gene is located on chromosome 8q22. The genes for CA III, CA II, and CA I each contain seven exons and six introns and all are closely linked on chromosome 8. No abnormal forms are known.
Half-life	Unknown
Concentration	CA III is the predominant cytosolic protein in skeletal muscle, with concentration as high as 5×10^{-4} mol L^{-1} in rat soleus. CA III comprises up to 25% of cytosolic protein in rat adipocytes.
Isolation Method	A convenient source is fresh skeletal muscle from bovine (e.g. soleus). Purification usually involves gel filtration followed by ion-exchange chromatography. An affinity gel utilizing paraaminobenzolamide provides partial purification which must be accompanied by another chromatographic step to achieve greater purity.
Amino Acid Sequence	There is a 56% identity in the aa sequences of human CA III and human CA II, which includes the three histidine ligands of zinc. In the active-site cavity, residues unique to CA III are Lys-64 (or Arg-64), Cys-66, Arg-67, Arg-91, Phe-198. This active-site cavity has more positive charge and is more sterically constrained than in CA II.
Disulfides/SH-Groups	No disulfides/5 SH-groups in human CA III.
General References	Engberg, P. et al. *Arch. Biochem. Biophys.* 1985, **241**:628-638. Eriksson, A.E. and Liljas, A. *Proteins: Struct. Func. Genet.* 1993, **16**:29-42. Gros, G. and Dodgson, S.J. *Ann. Rev. Physiol.* 1988, **50**:669-694. Lynch, C.J. et al. *Amer. J. Physiol.* 1993.**264**:E621-630. Tashian, R.E. *BioEssays* 1989, **10**:186-192. Tu, C.K. et al. *J. Biol. Chem.* 1983, **258**:8867-8871.
Ref. for DNA/AA Sequences	Lloyd, J. et al. *Gene* 1986, **41**:233-239 EMBL accession number M22658.



Stereo diagram of residues near the zinc (dotted sphere) in bovine carbonic anhydrase III from the crystal structure of A. E. Eriksson and A. Liljas (1993). Water molecules are shown as individual dots and the zinc-water bond is represented as a line connecting the two.

Carbonic anhydrase IV

Abdul Waheed and William S. Sly

Synonyms	Membrane carbonic anhydrase; Carbonate hydrase IV
Abbreviations	CA IV
Classifications	EC 4.2.1.1
Description	An extracellular carbonic anhydrase expressed on the apical and basolateral surfaces of proximal tubules, thick ascending limb of Henle, and some cortical collecting tubule cells in kidney. Also on the luminal surface of endothelial cells in selected capillary beds, including brain, choriocapillaris of the eye, muscle, heart, and lung microvasculature. It is synthesized as 312 aa precursor containing an N-terminal signal sequence, and a 28 aa C-terminal hydrophobic sequence that is cleaved in the rough surface endoplasmic reticulum to allow transfer of the CA to the ethanolamine group of the phosphatidyl inositol glycan anchor.
Structure	Crystal structure resolved to 2.8 Å
Molecular Weight	35 kDa (SDS-PAGE); 29,800 polypeptide without signal sequence and C-terminal anchor.
Sedimentation Coeff.	Unknown
Isoelectric Point	6.1 (4 bands)
Extinction Coeff.	13.0 (280nm 1%, 1cm)
Enzyme Activity	Catalyzes reversible dehydration of HCO_3^- from glomerular filtrate to capillary in kidney and hydration of CO_2 in some microcapillary beds (e.g., brain).
Coenzymes/Cofactors	Bound zinc is essential for catalysis.
Substrates	CO_2, HCO_3^-, H^+
Inhibitors	It is more sensitive than CA II to inhibition by halide ions, and less sensitive to sulfonamide inhibitors. It is unique among CAs in its resistance to inhibition by 0.2% SDS that is conferred by its S-S bonds.
Biological Functions	Facilitates CO_2 exchange and HCO_3^- transport.
Physiology/Pathology	CA IV mediates the reabsorption of HCO_3^- in kidney and reversible hydration of CO_2 in capillary beds facilitates CO_2 and HCO_3^- transport.
Degradation	Unknown
Genetics/Abnormalities	The gene for CA IV spans 9.5 kb on chromosome 17q23, contains 8 exons (1a, 1b, and 2-7). No known abnormal genes or gene products are identified.
Half-life	Unknown
Concentration	Unknown

Isolation Method	CA IV is purified from lung or kidney, by solubilizing membranes in 1-5% SDS, absorbing to sulfonamide-inhibitor affinity column, and eluting in 0.1% Triton containing buffer (see Zhu, X.L. and Sly, W.S., 1990).
Amino Acid Sequence	Amino acid sequence of CA IV and sequence homology among seven CA isozymes are shown in ref. Okuyama, T. et al., 1992.
Disulfides/SH-Groups	Mature CA IV has 4 cysteines which form two disulfide bridges Cys6-Cys18 and Cys28-Cys211.
General References	Zhu, X.L. and Sly, W.S. *J. Biol. Chem.* 1990, **265**:8795-8801. Sato, S. et al. *Proc. Natl. Acad. Sci. USA* 1990, **87**:6073-6076. Brown, D. et al. *Proc. Natl. Acad. Sci. USA* 1990, **87**:7457-7461. Hageman, G.S. et al. *Proc. Natl. Acad. Sci. USA* 1991, **88**:2716-2720. Waheed, A. et al. *Arch. Biochem. Biophys.* 1992, **294**:550-556. Waheed, A. et al. *J. Biol. Chem.* 1992, **267**:3308-3311. Ghandour, M.S. et al. *Proc. Natl. Acad. Sci. USA* 1992, **89**:6823-6827. Parkkila, S. et al. *J. Histochem. Cytochem.* 1993, **41**:751-757. Fleming, R.E. et al. *Am. J. Physiol.* 1993, **265**:L627-L635. Sender, S. et al. *J. Histochem. Cytochem.* 1994, **42**:1229-1236. Okuyama, T. et al. *Arch. Biochem. Biophys.* 1995, **320**:315-322. Kaunisto, K. et al. *Biol. Reprod.* 1995, **52**:1350-1357. Fleming, R.E. et al. *J. Clin. Invest.* 1995, **96**:2907-2913. Tamai, S. et al. *Biochem. Genet.* 1996, **34**:31-43. Parkkila, S, et al. *Hepatology* 1996, **24**:1104-1108. Waheed, A. et al. *Arch. Biochem. Biophys.* 1996, **333**:432-438. Waheed, A. et al. *Protein Expr. Purif.*, in press. Tamai, S., Waheed, A., Cody, L.B., Sly, W.S. *Proc. Natl. Acad. Sci. USA* 1996, **93**:13647-13652. Stams, T., Nair, S.K., Okuyama, T. et al. *Proc. Natl. Acad. Sci. USA* 1996, **93**:13589-13594. Baird, T. et al. *Biochemistry*, submitted
Ref. for DNA/AA Sequences	Okuyama, T. et al. *Proc. Natl. Acad. Sci. USA* 1992, **89**:1315-1319. Okuyama, T. et al. *Genomics* 1993, **16**:678-684.

Carbonyl Reductase

Bendicht Wermuth

Synonyms	NADP-dependent 15-hydroxyprostaglandin dehydrogenase; Prostaglandin 9-keto reductase
Abbreviations	Cbr (15-OH-PGDH)
Classifications	EC 1.1.1.184 (probably identical with EC 1.1.1.189 and 1.1.1.196)
Description	A cytosolic, monomeric protein consisting of 276 aa. There are multiple molecular forms due to autocatalytic reductive alkylation of Lys-238 by pyruvate, 2-oxoglutarate and other 2-oxocarboxylic acids. It is constitutively expressed in essentially all tissues.
Structure	It is a member of the short-chain dehydrogenase/reductase (SDR) superfamily and contains the classical Rossmann-fold secondary structure for coenzyme binding. The tertiary structure is not yet determined, however the sequence has been modelled to the known structure of 3α,20ß-hydroxysteroid dehydrogenase from *Streptomyces hydrogenans*, another member of the SDR superfamily. The active site contains a Ser-**Tyr**-Lys triad.
Molecular Weight	30,268
Sedimentation Coeff.	Unknown
Isoelectric Point	8.5 (pyruvate-modified enzyme, 8; 2-oxoglutarate-modified enzyme, 7)
Extinction Coeff.	7.1 (280nm, 1%, 1cm; calculated from aa composition)
Enzyme Activity	Oxidoreductase: Catalyzes the reduction of a broad range of carbonyl compounds. Shows very low dehydrogenase activity with the exception of the oxidation of 15-hydroxyprostaglandins. Specific activity with menadione as substrate: ca.12 µmol/min per mg protein (0.1 M Na-phosphate, pH 7, 30 °C)
Coenzymes/Cofactors	NADPH
Substrates	The best substrates are quinones (ubiquinone-1, menadione and quinones of polycyclic aromatic hydrocarbons), followed by dicarbonyls (2,3-butanedione, phenylglyoxal) and aromatic aldehydes and ketones (2-nitrobenzaldehyde, pyridine-4-carboxaldehyde, 4-benzoylpyridine). Also acts on the 9-oxo group of prostaglandins E and the 3,17-oxo groups of androstanones and etiocholanones.
Inhibitors	Susceptible to inactivation by organomercurials and dicarbonyls. Reversible inhibition by flavonoids (quercetine, rutin; $IC_{50} < 1$ µM)
Biological Functions	Reduction of reactive carbonyl compunds. No specific physiological substrate known. (Detoxification?)
Physiology/Pathology	Unknown
Degradation	Unknown

Genetics/Abnormalities	The gene is on chromosome 21 at q22.2, and contains 3150 bases with 3 exons. No mutations/polymorphisms of the coding sequence are known.
Half-life	Unknown
Concentration	1-2 ‰ of cytosolic proteins in liver, kidney and brain. Present in red and white blood cells but not detectable in plasma or serum.
Isolation Method	Ion exchange chromatography on DEAE-cellulose, gel filtration on Sephadex G-100 and affinity chromatography on 2'-ADPR-Sepharose or Matrex Orange. Multiple forms can be separated by DEAE-Sephadex chromatography.
Amino Acid Sequence	SSGIHVALVT GGNKGIGLAI VRDLCRLFSG DVVLTARDVT RGQAAVQQLQ AEGLSPRFHQ LDIDDLQSIR ALRDFLRKEY GGLDVLVNNA GIAFKVADPT PFHIQAEVTM KTNFFGTRDV CTELLPLIKP QGRVVNVSSI MSVRALKSCS PELQQKFRSE TITEEELVGL MNKFVEDTKK GVHQKEGWPS SAYGVTKIGV TVLSRIHARK LSEQRKGDKI LLNACCPGWV RTDMAGPKAT KSPEEGAETP VYLALLPPDA EGPHGQFVSE KRVEQW
Disulfides/SH-Groups	No disulfides, 5 free sulfhydryls. Cys-226 reacts with stoichiometric concentrations of organomercurials.
General references	Bohren, K.M. et al. *J. Mol. Biol.* 1994, **244:**659-664. Krook, M. et. al. *Proc. Natl. Acad. Sci. USA* 1993, **90:**502-506. Wermuth, B. *J. Biol. Chem.* 1981, **256:**1206-1213. Wirth, H. and Wermuth, B. *J. Histochem. Cytochem.* 1992, **40:**1857-1863.
Ref. for DNA/AA Sequences	Forrest, G.L. et al. *Mol. Pharmacol.* 1991, **40:**502-507. Wermuth, B. et al. *J. Biol. Chem.* 1988, **263:**16185-16188.

Carboxyl Ester Lipase

Karen Reue

Synonyms	Carboxyl ester hydrolase, cholesterol esterase, non-specific lipase, non-specific esterase, lysophospholipase, lipid-soluble vitamin ester hydrolase, bile salt-stimulated lipase, bile salt-activated lipase
Abbreviations	CEL, CEH, BSSL
Classifications	EC 3.1.1.13; lipase
Description	CEL is a glycoprotein containing approximately 20% carbohydrate. This enzyme is synthesized in the acinar cells of the pancreas, stored in zymogen granules, and secreted into the intestinal lumen. A protein having identical aa sequence but slightly different molecular weight is synthesized in human mammary gland and secreted in human milk (known as bile salt-stimulated lipase). A cholesterol esterase with a substrate specificity similar to CEL has also been detected in rat liver, although mRNA for this protein could not be detected in human liver.
Structure	The crystal structure of CEL has not been determined, however sedimentation studies demonstrate a high frictional ratio (1.768) indicating that the CEL molecule resembles an ellipsoid with an axial ratio between 15 and 20.
Molecular Weight	100 kDa (SDS-PAGE) apparent molecular weight, 20% of which is due to carbohydrate. Interestingly, the bile salt-stimulating lipase expressed in human milk has an identical aa sequence to CEL as determined by cDNA sequence, but in direct comparison with pancreatic CEL demonstrates a higher molecular weight on SDS-PAGE (107 kDa-112 kDa). The nature of this difference in molecular weight is not known but may be due to a difference in glycosylation or other post-translational modification of the enzyme in pancreas compared to mammary gland.
Sedimentation Coeff.	4.46
Isoelectric Point	4.2
Extinction Coeff.	Unknown
Enzyme Activity	The two substrates that have primarily been utilized to monitor CEL activity are cholesteryl oleate and p-nitrophenylacetate. CEL activity in human pancreatic juice has been determined to be 2400 μmol/hr/mg (p-nitrophenylacetate as substrate) or 220 μmol/hr/mg (cholesteryl oleate as substrate).
Coenzymes/Cofactors	CEL does not requre any cofactor, but is activated by low concentrations of bile salts. Primary bile salts containing the 7-α-hydroxyl group are effective activators, whereas secondary bile salts, such as deoxycholate are not. The concentration of bile salt required for maximal activity is substrate dependent.
Substrates	CEL has a broad substrate range including cholesteryl esters, fat-soluble vitamin esters (A, D_3, and E), phospholipids, and tri-, di-, and monoglycerides. Activity against triglycerides requires higher concentrations of bile salts than for other ester substrates.

Inhibitors	*In vitro*, CEL is completely inhibited by eserine and by diisopropylfluorophosphate, suggesting that it is a serine esterase. Compete inhibition also occurs with the carboxyl group reagent N-ethyl-5-phenylisoxazolium-3'-sulfonate.
Biological Functions	CEL is one of two major lipolytic enzymes produced by the pancreas for the hydrolysis of dietary lipids. Unlike the other enzyme, pancreatic lipase, CEL is present at high concentration in neonatal rat pancreas, and may therefore play a role in assimilation of dietary lipid during neonatal development. CEL is also unique among pancreatic enzymes in its ability to hydrolyze cholesteryl esters and may therefore influence absorption of dietary cholesterol. Absorption of fat-soluble vitamins may also be mediated by CEL.
Physiology/Pathology	The physiological function of CEL is not clear. It has been suggested that CEL plays a critical role in absorption of dietary cholesterol as removal of CEL from pancreatic juice by immunoprecipitation or surgical diversion of pancreatic juice produces an 80% reduction in dietary cholesterol absorption. However, since the majority of dietary cholesterol is free rather than esterified, the importance of CEL for the absorption of dietary cholesterol remains unproven.
Degradation	Unknown
Genetics/Abnormalities	The gene for human CEL has been cloned and mapped to chromosome 9qter. The CEL locus appears to be highly polymorphic as determined by restriction fragment length polymorphism analysis, and at least 4 allelic forms of the CEL gene have been identified. These data are consistent with the occurence of a hypervariable region associated with the CEL gene. Humans with CEL deficiency have not been documented.
Half-life	Unknown
Concentration	In humans, CEL constitutes 4% of human pancreatic juice protein.
Isolation Method	CEL has been prepared from both whole pancreas and pancreatic juice using standard chromatographic techniques. More recently, immunoaffinity chromatography has been employed.
Amino Acid Sequence	The primary aa sequence has been deduced from cDNA clones and shown to have a high degree of similarity to acetylcholinesterase and cholinesterase. CEL is a member of the serine esterase family and contains the sequence Gly-X-Ser-X-Gly, where X is any aa. The serine residue in this motif has been identified as the fatty acid acyl acceptor within the active site of CEL. The enzyme also has a series of proline-rich tandem repeat units occuring within the carboxyl region that vary in number among species.
Disulfides/SH-Groups	There are four Cys residues and no free -SH groups in the mature CEL protein. Disulfide bridges occur at Cys64-Cys80 and Cys246-Cys257.
General References	Rudd, E. A. and Brockman, H. L. **In** *Lipases* 1984 (B. Borgström and H. L. Brockman, eds.), Elsevier Science, Amsterdam, pp. 185–204. Abouakil, N. et al. *Biochim. Biophys. Acta* 1988, **961**: 299–308. Guy, O. et al. *Eur J. Biochem.* 1981, **117**: 457–460.
Ref. for DNA/AA Sequences	Reue, K. et al. *J. Lipid Res.* 1991, **32**: 267–276. (CEL cDNA sequence) Baba, T. et al. *Biochemistry* 1991, **30**: 500–510. (BSSL cDNA sequence) Taylor, A. K. et al. *Genomics* 1991, **10**: 425–431 (CEL DNA polymorphisms).

Carboxypeptidase N

Dirk F. Hendriks and Katinka A. Schatteman

Synonyms	Lysine carboxypeptidase; Arginine carboxypeptidase; Serum carboxypeptidase B; Plasma carboxypeptidase B; Anaphylatoxin inhibitor; Kininase I; Creatine kinase conversion factor
Abbreviations	CPN
Classifications	EC 3.4.17.3 Carboxypeptidases; Peptidase family M14
Description	A tetrameric plasma protein, synthesized in the liver. It is composed of two high molecular weight subunits (83 kDa) which are heavily glycosilated but have no enzymatic activity and two lower molecular weight subunits (55 kDa) which lack carbohydrate but contain the catalytic center. The 83 kDa subunit functions to stabilize the active subunit at body temperature and keep it in the circulation.
Structure	The 3D structure of carboxypeptidase N has not been elucidated to date. The 55 kDa active subunit of CPN exhibits high sequence similarity with carboxypeptidases M (41%) and H (49%), and a more distant similarity with pancreatic carboxypeptidases A (19%) and B (17%). The active site of these carboxypeptidases is much better conserved: CPN active site residues within 8 Å from the center of the active site are 43% identical to pancreatic carboxypeptidase A. The general fold of the 55 kDa active subunit of CPN is presumed to be similar to the well known carboxypeptidase A structure. It is a zinc metalloenzyme having a central mixed parallel/antiparallel eight-strand β-sheet over which eight α-helices pack on both sides to form a globular molecule. The most exceptional feature of the inactive subunit of CPN is that it contains 12 leucine-rich tandem repeats.
Molecular Weight	280 kDa (tetrameric enzyme); 83 kDa (high molecular weight subunit); 55 kDa (low molecular weight subunit).
Sedimentation Coeff.	Unknown
Isoelectric Point	3.8 - 4.3 (IEF)
Extinction Coeff.	Unknown
Enzyme Activity	Removes C-terminal basic aa arginine and lysine from peptides and proteins.
Coenzymes/Cofactors	Zn-containing metalloprotease (one atom Zn per polypeptide chain)
Substrates	Peptides and proteins with unblocked C-terminal arginine or lysine. Possible natural substrates include: kinins (bradykinin and kallidin), anaphylatoxins C_{3a}, C_{4a} and C_{5a}, enkephalin hexapeptides, fibrinopeptides 6A and 6D, atriopeptin II, protamine, creatine kinase MM and MB, α-enolase, albumin propeptide, calmodulin. Synthetic substrates: hippuryl-L-arginine, hippuryl-L-lysine, Fa-Ala-Arg, Fa-Ala-Lys, Bz-Ala-Arg, Bz-Ala-Lys, hippuryl-L-argininic acid (esterase activity of the enzyme).

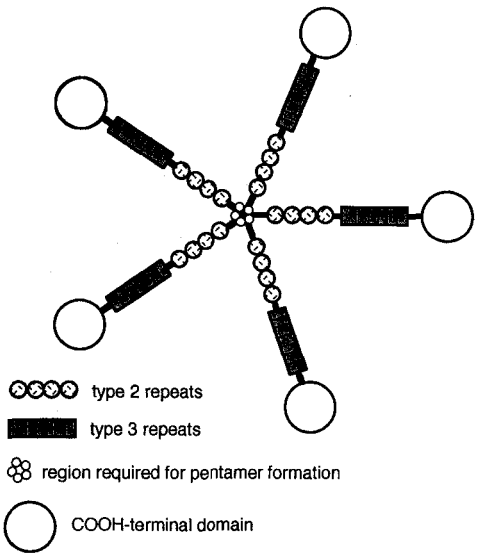
Inhibitors	GEMSA (guanidonoethylmercaptosuccinic acid), MERGEPTA (mercaptomethylguanidinoethylthiopropanoic acid). Metal ion chelators EDTA and phenanthrolin; Cd^{2+}.
Biological Functions	Since CPN circulates as an active enzyme in human blood in a relative high concentration, its most likely physiological function is believed to be the protection of the organism agains the actions of potent peptides which may escape from circulation or be released into the circulation.
Physiology/Pathology	Treatment of guinea pigs with cobra venom factor, causing massive release of anaphylatoxins, showed a nonlethal reaction because the generated anaphylatoxins are rapidly converted to their desArg forms, presumably by the action of CPN. However, when these animals were first treated with a specific inhibitor of CPN and then challenged with cobra venom factor, they died from acute respiratory distress. Genetically determined low blood levels of CPN (about 20% of normal levels) were associated with repeated attacks of angioneurotic oedema.
Degradation	CPN can be completely dissociated into its constituent subunits in 3 M guanidine, suggesting that ionic forces hold the protein together. Intact CPN is stable at 37° C, but the separate active subunit is unstable. Proteolytic cleavage of the active subunit of CPN results in the generation of smaller fragments which have increased enzymatic activities, but decreased stability.
Genetics/Abnormalities	The gene for the CPN high-molecular-weight-subunit is located on chromosome 8 region p22-23.
Half-life	Intravenous injection of ^{125}I-CPN resulted in a plasma half-life of approximately 24 hours.
Concentration	In plasma: 30 mg/L
Isolation Method	Isolation of CPN from human serum is performed by affinity chromatography on arginine-Sepharose, using specific elution with GEMSA (a specific inhibitor of basic carboxypeptidases).
Amino Acid Sequence	A cDNA clone containing the entire coding sequence of the active subunit of CPN has been isolated from a human liver cDNA library. It encodes a signal sequence of 20 aa and the 438 aa of the mature subunit. The cDNA for the 83 kDa subunit of CPN encodes a 59 kDa protein with no sequence similarity to the active subunit of CPN or any other carboxypeptidase.
Disulfides/SH-Groups	Unknown
General References	Levin, Y. et al. *Proc. Natl. Acad. Sci. USA* 1982, **798**:4618-4622. Hendriks, D. et al. *Clin. Chem.* 1985, **31**:1936-1939. Skidgel, R. In: *Zinc metalloproteases in health and disease.* Hooper, N. (ed.), Taylor & Francis Ltd., London. 1996, pp 241-283.
Ref. for DNA/AA Sequences	Gebhard, W. et al. *Eur. J. Biochem.* 1989, **178**:603-607. Tan, F. et al. *J. Biol. Chem.* 1990, **265**:13-19.
Molecular Models	Hendriks, D. et al. *Biol. Chem. Hoppe-Seyler* 1993, **374**:843-849.

Cartilage Oligomeric Matrix Protein

Jack Lawler

Synonyms	None
Abbreviations	COMP
Classifications	Extracellular calcium-binding protein
Description	COMP is a 524,000-Da glycoprotein component of the extracellular matrix of cartilage. It is synthesized and secreted by chondrocytes from various types of cartilage and by Swarm rat chondrosarcoma cells. COMP reportedly localizes to the territorial matrix on chondrocytes. The COMP mRNA is also present in aortic tissue.
Structure	COMP is composed of five copies of a 100,000–120,000-Da polypeptide. Structurally, COMP is similar to the members of the thrombospondin gene family. The five subunits are linked by disulfide bonds that are near the NH_2-terminal. Electron microscopy reveals the presence of a cylindrical structure at the point where the subunits are connected. This structure is reportedly composed of five parallel a-helices that are formed by amino acid residues 20–83. This region seems to be sufficient to direct pentamer assembly. Thin-flexible regions of polypeptide that originate at the point where the subunits are connected and terminate in globular domains comprise the remainder of the molecule. The thin-flexible regions have been reported to be 28 nm long.
Molecular Weight	524,000: intact molecule by sedimentation equilibrium; 100,000–120,000: subunit by SDS-PAGE; 82,842: subunit based on amino acid sequence.
Sedimentation Coeff.	Not determined
Isoelectric Point	4.1 (calculated value from aa composition)
Extinction Coeff.	Not determined
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	The biological function of COMP is unknown at this time. The fact that the protein is extracted in buffers containing EDTA suggests that the binding of COMP to a component of connective tissue depends on calcium. Since the intact protein may bind 60 calcium ions, COMP may function to sequester calcium in a readily mobilizable form.
Physiology/Pathology	Unknown
Degradation	Unknown
Genetics/Abnormalities	The human COMP gene is located on chromosome 19p13.1 and the mouse gene is located in the center of chromosome 8.

Half-life	Unknown
Concentration	Unknown
Isolation Method	COMP has been purified from bovine fetlock joint articular cartilage and Swarm rat chondrosarcoma tissue. It is extracted from either source with buffer containing 4M guanidine HCl, 10 to 50 mM EDTA and protease inhibitors. Proteoglycans can be removed by CsCl-density gradient centrifugation. Because of its high molecular weight, COMP is eluted in the high molecular weight pools of Sephadex G-200 or Sepharose CL-4B. Ion exchange chromatography on DEAE-Sepharose is also used as a step in the purification of COMP.
Amino Acid Sequence	The open reading frames for rat and human COMP are 755 and 757 aa long, respectively. The molecular architecture of COMP is similar to that of thrombospondin-3 and -4, except that COMP lacks an approximately 200 amino acid NH_2-terminal domain. COMP contains four type 2 or epidermal growth factor-like repeats and seven type 3 repeats. The type 3 repeats appear to represent a contiguous set of calcium-binding sites. The human and bovine sequences contain an RGD sequence in the third type 3 repeat, suggesting a possible interaction with integrins. This sequence is not conserved in the rat. The COOH-terminal domain of approximately 200 amino acids shows the greatest similarity to the other members of the thrombospondin gene family.
Disulfides/SH-Groups	Cys-68 and Cys-71 (numbering from the initiating methionine of the human sequence) participate in the formation of the interchain disulfide bonds. Since the mature peptide appears to have 45 cysteine residues, it probably has at least one free sulfhydryl group.
General References	Hedbom, E. et al. *J. Biol. Chem.* 1992, **267**:6132–6136. Morgelin, M. et al. *J. Biol. Chem.* 1992, **267**:6137–6141. Efimov, V. P. et al. *J. Biol. Chem.* 1994, **341**:54–58.
Ref. for DNA/AA Sequences	Oldberg, A. et al. *J. Biol. Chem.* 1992, **267**:22346–22350. GenBank Accession numbers X74326 and X72914. Newton, G. et al. *Genomics* 1994, **24**:435–439. GenBank Accession number L32137.



Catalase

Gian Franco Gaetani and Anna-Maria Ferraris

Synonyms	None
Abbreviations	None
Classifications	EC 1.11.1.6.
Description	Catalase is one of the most active catalysts produced by nature. The enzyme is widely distributed in plant and animal tissues. In mammalian tissues there is considerable variation, catalase concentration being highest in liver and kidney and lowest in connective tissue. In the liver cell it is mainly localized in peroxisomes and mithocondria. In the mature red cell catalase is essentially free, only traces being bound to stroma proteins.
Structure	The active enzyme is composed of four protein subunits, with four ferriprotoporphyrin (heme) groups corresponding to 1.1 % protohaemin, and 0.09 % iron, and four molecules of NADPH. The heme group of catalase is buried within a 20 Å-wide hydrophobic channel, while NADPH, which has a dissociation constant for catalase less than 1×10^{-8} M, is located near the C-terminal end of two α-helices and lies within 14 Å of the heme group. The four monomers form a tetrahedrally symmetric ellipsoid molecule, with a diameter of 70-90 Å.
Molecular Weight	240 kDa: tetramer; 60 kDa: each subunit.
Sedimentation Coeff.	11.3 S
Isoelectric Point	5.5
Extinction Coeff.	380-400 $mM^{-1}cm^{-1}$ (405 nm) for the tetramer; 100 $mM^{-1}cm^{-1}$ (405nm) per heme group. The specific activity and the ratio E_{405}/E_{280} can be used as an index of purity of catalase.
Enzyme Activity	The decomposition of H_2O_2 can be followed directly by the decrease in absorbance at 240 nm (E_{240}= 0.00394 ± 0.0002 L $mMol^{-1}$ mm^{-1}). The difference in absorbance ($Delta_{240}$) per unit time is a measure of catalase activity. The decomposition of H_2O_2 initially (from 0-30 sec) follows that of the first order reaction with H_2O_2 concentration between 10 and 50 mM. The rate constant (k) for the reaction is: $k = 1/t \times \ln S_1/S_2$ where t = measured time interval (t_2-t_1) S_1 and S_2 = H_2O_2 concentrations measured at time t_1 and t_2. The constant k can be used as a direct measure of the concentration of catalase. To avoid inactivation of the enzyne during assay (usually 30 sec) or formation of bubbles in the cuvette due to liberation of O_2, it is necessary to use relatively low H_2O_2 concentration (10 mM). Dependence on temperature is negligible. The pH activity curve relative to V_0 has a broad pH optimum (pH 6.8-7.5).
Coenzymes/Cofactors	Each tetrameric molecule of human or bovine catalase contains four molecules of tightly bound NADPH with a dissociation constant less than 1×10^{-8} M. The affinity of human catalase for dinucleotides follows the order: NADPH > NADH > NADP > NAD. The function of NADPH is to prevent or reverse the formation of Compound II, the inactive form of catalase.

Substrates	The specifity of catalase for peroxides is high. Only hydrogen, methyl and ethyl hydroperoxides give appreciable activity; t-butylperoxide does not react with catalase, but is a substrate of glutathione peroxidase. Hydrogen donors for catalase include a sequence common to aliphatic alcohols, with high activities for methyl and ethyl and low activities for butyl and higher homologues. Phenols and formic acid are also suitable hydrogen donors.
Inhibitors	3-amino-1,2,4,-triazole (AT). The inhibition depends upon the presence of peroxides. Only the subunits with intact prosthetic groups are modified and the inhibited product contains approximately one equivalent AT per hematin bound at His-74. The derivative retains both the ferric and the oligomeric structure of the native enzyme. This inhibitor may be used as a measurement of catalase activity in intact cells. Semicarbazide. This agent, reacting with Compound I, gives origin to nitrogen and N-carbamyl derivate. Cyanogen bromide (BrCN). Selective modifications of the active site of catalase are not limited to reactions involving Compound I. The inhibition results from the incorporation of one C moiety of BrCN into the apo-enzyme of each subunit of catalase. Sodium azide (NaN_3). Phenols. Sulfides. Sodium fluoride (NaF).
Biological Functions	$2\ H_2O_2 \rightarrow 2\ H_2O + O_2$ (1) $ROOH + AH_2 \rightarrow H_2O + ROH + A$ (2) Catalase appears to be the only enzyme having a dual function: decomposition of H_2O_2 to H_2O and O_2 (catalase activity) (1) and oxidation of H donors, e.g. methanol, ethanol, formic acid, phenols, with the consumption of 1 mol of peroxide (peroxidase activity) (2). When $[H_2O_2]$ is low, the peroxidatic pathway will predominate, when $[H_2O_2]$ is high the catalatic activity predominates.
Physiology/Pathology	Catalase is widely distributed and present in almost all aerobic organisms. Together with superoxide dismutase and glutathione peroxidase, catalase plays a central role in oxidant defense. The fact that it is missing in most anaerobic microorganisms and it is abundantly present in radioresistant species has stimulated speculations about its physiological role. The role of catalase is probably different in blood and tissues. In the latter, its function is prevalently peroxidatic, at least where the peroxisome fraction is concerned. These organelles contain high concentration of H_2O_2-producing enzymes (uricase, D-aminoacid oxidase) and catalase. In the red cell catalase, together with glutathione peroxidase, is capable of protecting hemoglobin and other cell components against oxidizing agents. The contribution of catalase toward protection of erythrocytes has been controversial. Some authors believe that under physiological conditions catalase does not play any role in this respect, while others consider it to be the main protecting enzyme. A unifying view at this regard has been the finding that both catalase and glutathione peroxidase depend on NADPH generation, and that these two enzymes are active in the red cell with a prevalence of catalase. The existence of individuals with practically no catalase activity was first described by Takahara et al. in 1947 in a Japanese girl suffering from ulcers and gangrene in the oral and nasal cavity. In reviewing the reported cases, it was found that only half of acatalasemic individuals had peculiar progressive gangrene of the mouth.
Degradation	Unknown
Genetics/Abnormalities	The gene for human catalase has been mapped to chromosome 11, band p13. The gene is 34kb long and is split into 13 exons (coding region 1581 base pair). A guanine-to-adenine substitution at the fifth position of intron

4 (a splice-site mutation) seems to be responsible for the defective enzyme synthesis in some cases of acatalasemia. Acatalasemia is an autosomal recessive disease, with heterozygotes having intermediate levels of red cell catalase activity. An average gene frequency of 0.006 has been calculated, with a frequency for heterozygotes of about 0.01. A syndrome of oral gangrene and ulcerations is associated with Japanese acatalasemia in about half the cases. The reasons for the variable expression of these symptoms are not known. Expression of the syndrome seems age-related: patients are rarely affected after puberty. In contrast with Japanese patients, none of the Swiss acatalasemics had oral gangrene or other health problems associated with the deficiency. Swiss acatalasemics have detectable amounts of catalase in other tissues. Cultured fibroblasts from Swiss acatalasemic patients have about 15 % of normal catalase activity, and those from Japanese patients only 2-4 %. It may then be argued that the tissues of Japanese patients are without protection from H_2O_2, thereby explaining their predisposition to oral gangrene.

Half-life

Unknown

Concentration

2.5-3.5 μmol L^{-1} of red blood cells (immunoprecipitation).

Isolation Method

Scheme of preparation of human erythrocytic catalase: 1. Preparation of stroma-free hemolysate; 2. Dialysis; 3. Chromatography on DEAE-cellulose and elution with Na-K phosphate buffer, pH 6.8; 4. Chromatography on CM-Cellulose.

Amino Acid Sequence

The complete aa sequence of bovine liver and erythrocyte catalase is known. Partial sequence of human erythrocyte catalase has also been obtained. The monomer of human erythrocyte catalase comprises 526 aa residues. In Penicillum vitale catalase has an additional "flavodoxin-like" domain at the C-terminus, whereas in mammalian catalase the hinge region is occupied by an NADPH molecule in a closed conformation.

Disulfides/SH-Groups

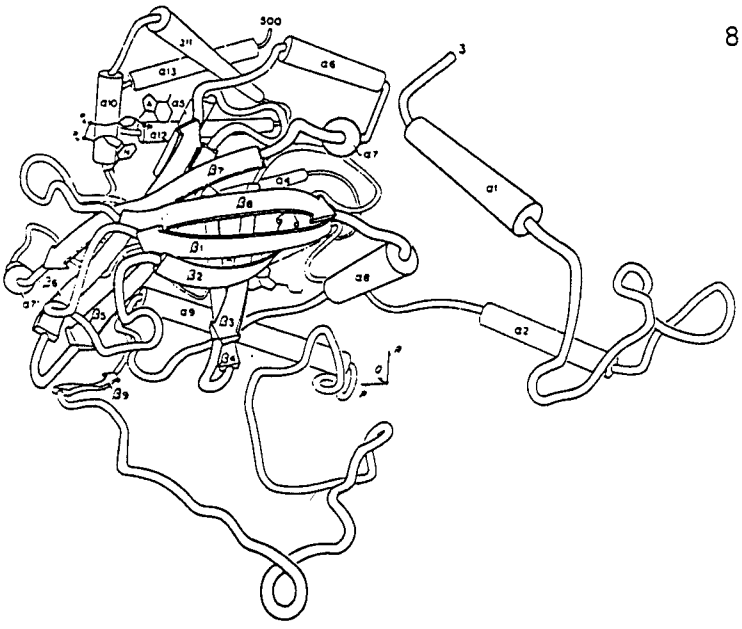
Human catalase prepared under aerobic conditions contains no disulfide bridge. Native human catalase contains 16 sulfhydryl groups.

General References

Takahara, S. In: *Acatalasemia in Japan. Hereditary Disorders of Erythrocyte Metabolism.* Beutler E.(ed.) Grune & Stratton, 1968 pp. 21-40.
Aebi, H. and Suter, H. Acatalasemia. *Adv. Hum. Genet.* 1971, **2**:143-199.
Kirkman, H.N. and Gaetani, G.F. *Proc. Natl. Acad. Sci. USA* 1984, **81**:4343-4347.
Kirkman, H.N. et al. *J. Biol. Chem.* 1987, **262**:660-655.
Eaton, J.W. and Muchou, M. Acatalasemia. In: *The Metabolic Basis of Inherited Disease.* Scriver, C.R. et al. (eds.), Mc Graw-Hill, 1995, pp. 2371-2383.
Ogata, M. *Hum. Genet.* 1991, **86**:331-340.
Gaetani, G.F. et al. *Blood* 1996, **87**:1595-1599.

Ref. for DNA/AA Sequences

Schroeder, W.A. et al. *Arch. Biochem. Biophys.* 1982, **214**:422-424.
Fita, I. and Rossmann, M.G. *Proc. Natl. Acad. Sci. USA* 1985, **82**:1604-1608.
Quan, F. et al. *Nucl. Acid Res.* 1986, **14**:5321-5335.
Wen, J.K. et al. *J. Mol. Biol.* 1990, **72**:383-393.
Human sequences are indexed under "HUMCATF" in GenBank listing (Accession number X04085).



Diagrammatic view of one subunit of catalase showing the bound heme and NADPH groups (from Fita, I. and Rossmann, M. G., *J. Mol. Biol.* 1985, **185**: 21–37).

Catechol-O-Methyltransferase

Barbara Bertocci and Mose' Da Prada

Synonyms	None
Abbreviations	COMT
Classification	EC 2.1.1.6
Description	COMT catalyses the O-methylation of a wide variety of compounds carrying a catechol group, including catecholamine neurotransmitters, catechol hormones and several xenobiotic catechols. The enzyme is widely distributed within the brain and the peripheral organs. Two forms of COMT exist, a membrane-bound form (MB) and a soluble one (S), whose relative abundance differs in various tissues. In the CNS, MB-COMT seems to be the prevalent form, whereas the opposite is true in peripheral organs. The difference between MB- and S-COMT consists exclusively in an N-terminal stretch of 50 aa in the MB-form which contains the membrane anchor domain. The MB-COMT is thought to be mainly associated to the endoplasmic reticulum with most of its mass exposed to the cytosol. A single gene encodes both forms, which are most likely derived from initiation of transcription at alternative sites, but the molecular mechanism regulating the production of the MB- and S-COMT is not yet fully elucidated.
Structure	COMT is a monomeric protein. The fine structure of the human enzyme has not yet been reported. On the other hand, the crystal structure of rat S-COMT (complexed with S-adenosyl-L-methionine, Mg^{2+} and the inhibitor 3,5-nitrocatechol) has been solved at 2.0 Å resolution. The enzyme consists of eight α-helices and seven β-strands with a typical α/β-fold.
Molecular Weight	30,053 Da: MB-COMT; 24,481 Da: S-COMT(values calculated from the deduced aa sequence)
Sedimentation Coeff.	3.61 S
Isoelectric Point	5.23: MB-COMT; 5.05: S-COMT (predicted from the deduced aa sequence)
Extinction Coeff.	Unknown
Enzyme Activity	COMT catalyzes the transfer of a methyl group from S-adenosyl-L-methionine (AdoMet) to a catecholic hydroxyl group of the substrate. The presence of Mg^{2+} is required for the enzymatic activity. Other divalent cations (e.g. Cd^{2+}, Hg^{2+}) also promote the methylation, whereas Ca^{2+} non competitively inhibits the reaction. *In vivo*, COMT O-methylates catecholic substrates mainly in the *meta* position. *In vitro*, both *meta* and *para* positions are O-methylated, depending on the nature of the substrate and the pH of the reaction. COMT exhibits a pH optimum between 7.5–8.0. The reaction proceeds via a sequential ordered mechanism with Mg^{2+} binding in a rapid equilibrium sequence prior to AdoMet. The kinetic propeties of MB- and S-COMT are very similar, except that MB-COMT displays a higher affinity toward substrates (e.g. dopamine) than the S-form.
Coenzymes/Cofactors	S-adenosyl-L-methionine (AdoMet) is the natural cofactor serving as methyl donor. The affinity of AdoMet for COMT is in the mid μM range. The coenzyme binding domain of COMT shares a similar structure to an

	AdoMet-dependent DNA methylase. Mg^{2+} is required for the catalytic function of COMT. No direct interaction exists between AdoMet and Mg^{2+}, but AdoMet binding occurs only when the enzyme is complexed with magnesium.
Substrates	COMT has a broad substrate specificity. The endogenous substrates of COMT include catecholamines neurotransmitters (dopamine, adrenaline and noradrenaline), their metabolites, L-dopa, catechol oestrogens and ascorbic acid. Several drugs such as apomorphine, benserazide, carbidopa, dobutamide, isoprenaline, isoproterenol and a wide variety of xenobiotic catechols are O-methylated by COMT.
Inhibitors	The inhibitors of COMT known so far are substrate analogs in that they all contain a catechol moiety. The first generation of COMT inhibitors (e.g. pyrogallol, tropolone) were rather weak compounds active only *in vitro*. Recently, potent, selective and orally active COMT inhibitors have been developed. These inhibitors include tolcapone (RO 40–7592), RO 40–0960, nitecapone, entacapone, all sharing a common 5-nitro catecholic chemical structure. Tolcapone and RO 40–0960 inhibit the enzyme both in the CNS and in the periphery, whereas nitecapone and entacapone are mainly peripheral inhibitors. The therapeutic interest of COMT inhibitors is mainly in the treatment of patients affected by Parkinson's disease as adjuvant to L-dopa therapy. COMT inhibitors, by preventing the conversion of L-dopa to 3-O-methyldopa, improve the bioavailability and the duration of action of L-dopa. Tolcapone and entacapone are currently under clinical trials. CGP 28014 is a pyridine derivative that, albeit devoid of COMT-inhibitory activity *in vitro*, mimics the effects of COMT inhibitors *in vivo*. Whether the effects of this compound are related to COMT inhibition remains to be established.
Biological Functions	COMT plays a major role in the inactivation of catecholamines both in CNS and periphery. A general function of COMT is the inactivation of biologically active or toxic catechol compounds (e.g. mutagenic flavonoids).
Physiology/Pathology	Variations in COMT activity have been described in several pathologies. Low COMT activity was found in the erythrocytes of patients with major, recurrent and bipolar affective disorders. A significantly increase of COMT activity has been observed in children with Down's syndrome. Decreased placental enzymatic activity has been detected in pregnancies associated with toxaemia and chronic hypertension. It has also been suggested that COMT plays a regulatory role in melanin synthesis.
Degradation	Unknown
Genetics/Abnormalities	The COMT gene is assigned to the band q11.2 on chromosome 22. Studies of COMT activity in erythrocytes have suggested that individual variation are genetically determined. Data from segregation analysis have provided evidence for the control of COMT by a major autosomal locus with two alleles. These biochemical studies were confirmed by the detection of a relative frequent two alleles COMT gene RFLP in human DNA's digested with Bgl II.
Half-life	Unknown
Concentration	Unknown

Isolation Method	The best sources of human COMT are liver and placenta. The purification steps include ammonium sulphate fractionation, gel filtration and ion-exchange chromatography.
Amino Acid Sequence	cDNA clone encoding human COMT have been isolated from an hepatoma cell line (Hep G2) and placenta. The nucleotide sequence predicts for a protein of 271 aa (MB-COMT) and 221 aa (S-COMT). The MB-COMT contains at its N terminus a stretch of 21–24 hydrophobic aa which are responsible for the insertion of the protein into endoplasmic reticulum membranes. No potential N-linked glycosylation sites are present in the deduced aa sequence. Human COMT shares a high degree of similarity (80%) with the rat and porcine enzyme. The regulation of the COMT gene expression in the production of different amounts of S- and MB-COMT in various tissues is unknown and seems to have a control system at different levels, transcription initiation (two promoter regions), mRNA splicing (two RNA transcripts 1.3 and 1.5 kb) and translation initiation (two ATG codon).
Disulfides/SH-Groups	Human COMT has seven cysteine residues which are all present as free thiols. Cys-68 and Cys-94 have been suggested to belong to regions at or near the binding site of AdoMet. However, X-ray crystallographic studies of rat S-COMT have revealed that no cysteine residue appears to be directly involved in cofactor binding and in the catalytic mechanism of COMT.
General References	Männistö, P. T. et al. Characteristics of catechol O-methyltransferase COMT) and proprieties of selective COMT inhibitors. *Prog. Drug Res.* 1992, **39**: 293–342. Roth, J. A. Membrane-bound catechol-O-methyltransferase. A revaluation of its role in the O-methylation of catecholamine neurotransmitters. Physiol. *Rev. Biochem. Pharmacol.* 1992, **120**: 1–29. Vidgren, J. et al. Crystal structure of catechol-O-methyltransferase. *Nature* 1994, **368**: 354–357.
Ref. for DNA/AA Sequence	Bertocci, B. et al. *Proc. Natl. Acad. Sci. USA* 1991, **88**: 1416–1420. GeneBank accession number M58525 Lundstrom, K. et al. *DNA and Cell Biol.* 1991, **10**: 181–189.

CD23

Erich Kilchherr and Christoph H. Heusser

Synonyms	Low affinity IgE receptor, low affinity Fcε receptor, Fcε receptor II.
Abbreviations	CD23, FcεRII
Classifications	C-type lectin; type II membrane protein
Description	A membrane protein with inverted orientation expressed on a variety of cells, mainly on B cells and monocytes. Expression on B cells is restricted to IgM^+/IgD^+ cells, prior to heavy chain class switching. Of the two forms A and B, the A form is expressed only on B cells. This form is constitutively expressed and can be upregulated by IL-4; expression of the B form is induced by IL-4 on B cells and monocytes. The cytoplasmic domain consists of 22(B) to 23(A) amino acids. The two forms differ in the 6 to 7 C-terminal (cytoplasmic) amino acids. The extracellular domain contains 4 disulfides and one potential N-glycosylation site. Several soluble forms (37,000; 33,000; 29,000; 25,000; 16,000) are released by a proteolytic process. These soluble CD23 (sCD23) molecules still retain the binding activity to IgE and are therefore called IgE binding factors (IgE-bf). The 25,000 fragment is the most stable and most abundant. CD23 binds IgE with $K_A = 5 \times 10^7\ M^{-1}$. The affinity of the 25,000 fragment is approximately 100-fold lower. The receptor form and several fragments have been cloned and expressed in mammalian and insect cells.
Structure	Little information is available on the three-dimensional conformation. The receptor probably forms a homodimer, possibly including a coiled-coil α-helical structure. A high proportion of β-sheet has been reported for the 37,000 fragment. The intact receptor is N-glycosylated, but the site is lost upon fragmentation. O-glycosylation has been reported for some fragments.
Molecular Weight	45,000 (SDS-PAGE), 36,455 (sum of amino acids). Fragments: 37,000 (SDS-PAGE), 27,159 (sum of amino acids); 33,000, 29,000 (SDS-PAGE); 25,000 (SDS-PAGE), 19,345 (sum of amino acids); 16,000 (SDS-PAGE).
Sedimentation Coeff.	Unknown
Isoelectric Point	4.4–5.2 (IEF); 5.37 (aa composition)
Extinction Coeff.	17.7 (278 nm, 1%, 1 cm); 18.2 (278 nm; calculated from aa composition (aa 148–321)), $\varepsilon = 44250\ M^{-1}\ cm^{-1}$, all for 25,000 recombinant fragment expressed in mammalian cells.
Enzyme Activity	Autoproteolysis leading to the soluble fragments has been reported in two independent reports. No information on the protease class available.
Coenzymes/Cofactors	Unknown
Substrates	Unknown
Inhibitors	Unknown

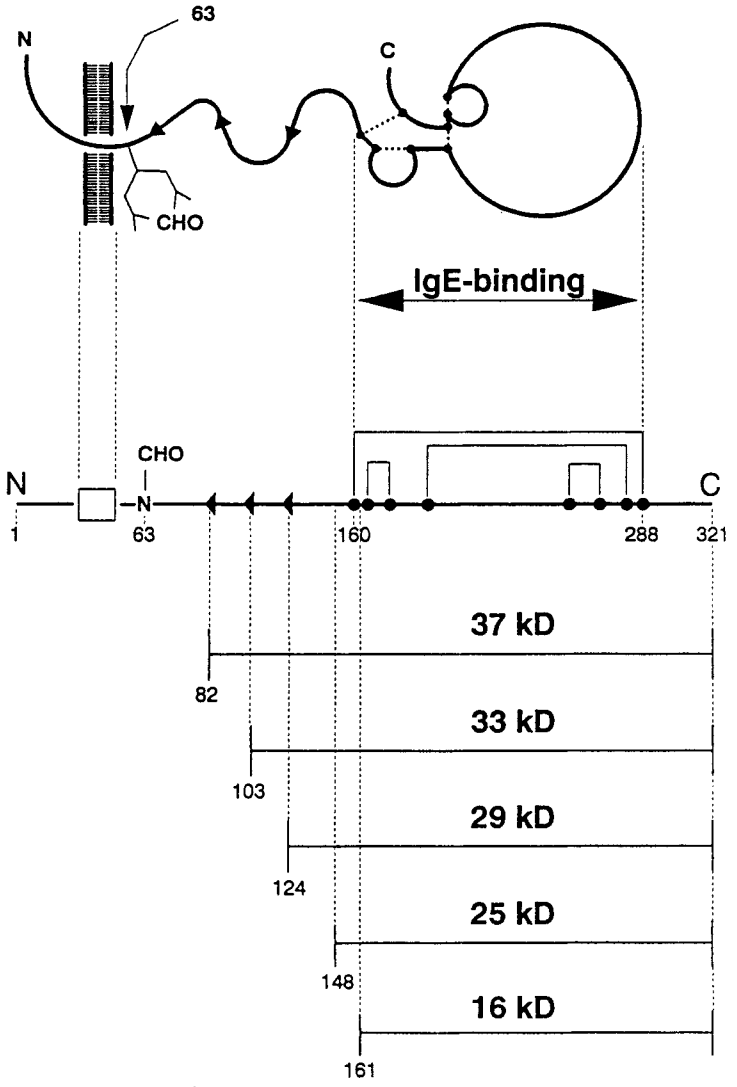
Biological Functions	A wide variety of functions have been described for CD23 and its fragments. Some of the most important are: Thymocyte differentiation; sCD23 in the presence of IL-1 induces $CD7^+CD2^-CD3^-CD4^-CD8^-$ T cell precursors to differentiate in vitro into mature $CD4^+$ or $CD8^+$ T cells with concomitant deletion of the TCR-δ locus. Myeloid cell proliferation; in the presence of IL-1, sCD23 promotes the proliferation of $CD34^+$ myeloid precursors. Rescue from apoptosis of germinal center B cells; germinal center centrocytes have been described to be rescued from apoptosis by 25,000 sCD23 in the presence of IL-1. For these latter three functions the presence of IL-1 is absolutely essential. IgE mediated antigen presentation; CD23 has been shown in murine and human systems to present antigen in an IgE dependent but antigen non-specific way to T cells, thereby inducing T cell activation. Inhibition of IgE production: 16,000 sCD23 has been reported to suppress IgE production by peripheral blood mononuclear cells from chronic lymphocytic leukemia patients. Not all of these functions are related to the IgE binding ability. Binding of CD23 to CD21, the complement C3d receptor CR2 (which is also the receptor for EBV) and association of CD23 with HLA-DR have also been described.
Physiology/Pathology	The relative importance of the listed biological functions and thus the precise physiological role of CD23 are not known. However, the expression of CD23 on thymic epithelial cells and in the light zone of germinal centers suggests that induction of thymocyte differentiation and/or prevention of centrocyte apoptosis are relevant physiological functions of CD23. Chronic lymphocytic leukemia patients have greatly increased serum levels of sCD23 (25,000). The levels relate directly to the progression of the disease.
Degradation	Degradation of intact receptor to the 25,000 IgE-BF occurs via the instable intermediates with molecular sizes of 37,000, 33,000, and 29,000. This process involves most probably an autoproteolytic activity of the CD23 itself. The 16,000 species may be obtained by treatment of the B lymphoblastoid line RPMI 8866 with tunicamycin and iodoacetamide.
Genetics/Abnormalities	Unknown
Half-life	Unknown
Concentration	0.15–1.5 μg/L: serum of normal individuals; 5–200 μg/L: chronic lymphocytic leukemia patients.
Isolation Method	The receptor is most easily isolated from NP-40 solubilized RPMI 8866 cells by affinity chromatography with an anti-CD23 monoclonal antibody. Further purification by anion exchange or reversed phase HPLC. Supernatants of RPMI 8866 cells are the source for the various fragments that are isolated similarly but omitting the detergent.
Amino Acid Sequence	FcεRII: 321 residues; cytoplasmic domain: res. 1–23; transmembrane domain: res. 24–44; N-glycosylation site: res. 63. Fragments: 37,000: 82–321; 33,000: presumably 103–321; 29,000: presumably 124–321; 25,000: 148–321; 16,000: presumably 161–321.
Disulfides/SH-Groups	Four disulfides in the extracellular domain. Bonding pattern partially determined: Cys191-Cys282; Cys259-Cys273. The remaining 2 disulfides are likely to be arranged as Cys160-Cys288 and Cys163-Cys174 for experimental reasons and due to homology to the C-type lectins.

General References

Lynch, R. G. et al. *Immunobiol.* 1992, **185:** 235–267.
Delespesse, G. et al. *Adv. Immunol.* 1991, **49:** 149–191.
Conrad, D. H. *Annu. Rev. Immunol.* 1990, **8:** 623–645.
Delespesse, G. et al. Human FcεRII. Molecular, Biological and Clinical Aspects. In: *Structures and Functions of Low Affinity Fc Receptors.* Fridman, W. H. (ed.), Chem. Immunol., Karger, Basel, 1989, Vol.**47**, pp. 79–105.

Ref. for DNA/AA Sequences

Lüdin, C. et al. *EMBO J.* 1987, **6:** 109–114.
Kikutani, H. et al. *Cell* 1986, **47:** 657–665.
Ikuta, K. et al. *Proc. Natl. Acad. Sci. USA* 1987, **84:** 819–823.
EMBL nucleotide sequence database, accession number: X04772.



CD23 and its soluble fragments
The model shows the orientation of CD23 in the membrane, the N-glycosylation site and the probable connectivity of the disulfides. IgE binding is restricted to the domain confined by the disulfide bonds. The triangles mark amplified exons. The N-termini of some of the fragments are putative, as well as the C-terminal of the 16 kDa fragment.

CD26

Ingrid A. De Meester and Anne-Marie Lambeir

Synonyms	Dipeptidyl-peptidase IV; Dipeptidyl aminopeptidase IV; Post-proline dipeptidyl aminopeptidase; X-Pro-dipeptidyl aminopeptidase; Glycyl-prolyl-ß-naphthylamidase; Adenosine deaminase binding protein; Adenosine deaminase complexing protein
Abbreviations	CD26; DPP IV; DP IV; DAP IV; ADA-bp; ADA-cp
Classifications	EC 3.4.14.5. Prolyl oligopeptidase family
Description	CD26 was first described as a T cell activation antigen expressed on a subset of CD4 and CD8 + T cells. The unique population of mononuclear cells expressing high levels of CD26 is necessary for the response upon recall antigens, and CD26+ T lymphocytes are shown to behave as Th1 like cells producing high amounts of IL2 and IFNγ. The CD26 antigen was identified as dipeptidyl-peptidase IV. This ecto-enzyme is widely distributed, but specific activities differ largely between tissues. CD26/DPP IV is abundant on kidney proximal tubular cells and in seminal plasma. In the hematopoetic system, enzymatic and immunostaining largely confined DPP IV/CD26 to the T cell subset. Recent data descibe low level expression on certain NK cells, B cells and monocytic cells. As on T cells, expression is increased after activation. CD26/DPP IV was found identical to the adenosine deaminase binding protein. In addition CD26 was reported to interact with CD45, with extracellular matrix proteins (collagen, fibronectin) and with the HIV-1 Tat protein. Some proteins with DPP IV like enzymatic activity have been found which differ from CD26 notably in molecular weight and binding to adenosine deaminase. On the cell-surface, CD26/DPP IV occurs as a type II sialoglycoprotein; a soluble form circulates in plasma. Part of the heterogeneity is due to differences in glycosylation.
Structure	The 3D structure has not been determined to date. The membrane-bound as well as the soluble form occur as dimers of 2 identical subunits; no interchain disulfide bonds are found. There is structural relationship with the family of α/β hydrolases and secondary structure predictions based on sequence alignments point to a similar organization of the DPP IV/CD26 catalytic region in alternating α helices and β sheets. The remainder part of the molecule contains a significant proportion of β sheet. On cultured fibroblasts and stimulated melanocytes, heteromeric cell surface complexes are formed with Fibroblast Activation Protein α (FAPα), a protein with 48% aa sequence identity to CD26.
Molecular Weight	260 000 - 280 000: dimer (gelfiltration) 103 000 - 120 000: monomer (SDS PAGE) 104 000 - 105 000: monomer (mass spectrometry) 88 300: monomer (sum of amino acids)
Sedimentation Coeff.	Unknown
Isoelectric Point	3.5-6.0: IEF
Extinction Coeff.	Molar extinction coefficient (280nm): 171 600 $M^{-1}cm^{-1}$
Enzyme Activity	Removes N-terminal dipeptides from peptides possessing a proline or alanine at the penultimate position, provided that the third aa is not proline

	or hydroxyproline. A free N-terminus is obligatory. The cleavable bond must be in *trans* configuration.
Coenzymes/Cofactors	Unknown
Substrates	Peptides with unblocked N-terminus and penultimate proline or alanine residue. The removal of N-terminal dipeptides by CD26/DPP IV was demonstrated for following peptides: substance P, neuropeptide Y, peptide YY, gastric inhibitory peptide, growth hormone releasing hormone, enterostatin, glucagon-like peptide-1 (7-36) amide, glucagon-like peptide-2. Chromogenic substrate: glycyl-prolyl-p-nitroanilide. Fluorogenic substrates: glycyl-prolyl 4-methoxy-2-naphthylamide; glycyl-prolyl-7-amino-4-methyl-coumarine
Inhibitors	No endogenous inhibitors have been found. The following groups of compounds behave as reversible (R) and irreversible (I) inhibitors of DPP IV: oligo- and dipeptides with N-terminal X-Pro (R) e.g. Diprotin A; N-amino acyl pyrrolidides and thiazolidides (R) e.g. Ile-Thiazolidide; peptidyl boronic acids (R) e.g. Pro-boroPro dipeptide nitriles (R); N-peptidyl-O-(4-nitrobenzoyl)hydroxylamines (I); dipeptide phosphonates (I) e.g. Pro-Pro-diphenylphosphonate. Most of the potent inhibitors suffer from toxicity or chemical instability which prohibit their use in biological sytems *in vitro* and *in vivo*. Promising compounds emerge within the group of dipeptide nitriles and dipeptide phosphonates.
Biological Functions	Participates in the activation process of T lymphocytes by providing a secondary signal for T cell receptor/CD3 mediated activation. There is *in vitro* and *in vivo* evidence for a role of the peptidase activity in the immune regulation; precise pathways however remain to be elucidated. Removes N-terminal dipeptides from peptides with penultimate Pro or Ala. In human serum, DPP IV initiates the metabolization of growth hormone releasing hormone, gastric inhibitory peptide, and glucagon-like peptide-1(7-36)amide by the removal of N-terminal Tyr-Ala and His-Ala respectively. There is evidence that the enzyme plays a significant role for the *in vivo* processing of substance P and enterostatin. DPP IV cleaves Tyr-Pro from neuropeptide Y and peptide YY and is as such a candidate enzyme for the *in vivo* generation of peptide YY(3-36), which has altered receptor specificity. Plays a major role in the final hydrolysis and subsequent intestinal absorption of peptides containing proline at the penultimate position. Is involved in the renal handling of proline and hydroxyproline containing peptides. Interacts with collagen and fibronectin and as such a role in cell to extracellullar matrix interactions is postulated.
Physiology/Pathology	Decreased numbers and proportions of circulating CD26+ cells are seen in patients with immunodeficiency (HIV infection, exogenous immunosuppression). In immune pathologies associated with increased Th1 like reactions (autoimmune and granulomatous diseases), an accumulation of CD26+ cells at the disease site was observed.
Degradation	Unknown
Genetics/Abnormalities	The gene is located at the long arm of chromosome 2(2q24.3). It spans approximately 70 kilobases and contains 26 exons. The nucleotides that encode the consensus motif around the active site serine are split between 2 exons. A 300 bp region extremely rich in C and G contains potential binding sites for transcriptional factors. The CD26/DPP IV promotor possesses the ability to initiate transcription in a tissue-specific fashion in spite of having the sequence characteristics of a housekeeping gene

	promotor. EcoRI identifies a two allele polymorphism (A1: 16kb, A2: 13kb).
Half-Life	Unknown
Concentration	Unknown for the CD26 protein levels. Normal values for DPP IV enzymatic activity in plasma have been determined.
Isolation Method	Purification of the native, enzymatically active protein from human seminal plasma can be carried out conveniently by using DEAE ion exchange chromatography followed by affinity chromatography onto immobilized adenosine deaminase (bovine, commercially available). Placenta has also been reported as a suitable source of DPP IV.
Amino Acid Sequence	The isolation and characterization of the cDNA encoding the human CD26 predicted a protein of 766 aa. The primary structures of the CD26 homologues in different species revealed to be highly conserved. The transmembrane protein has a very short intracytoplasmatic tail (6 aa), a transmembrane region, and a large extracellullar part, putatively organized in 3 domains: (1) a highly glycosylated domain, (2) a cysteine-rich domain and (3) the catalytic region at the C-terminal end. The monomer has 10 potential N-glycosylation sites and 12 cysteine residues. This exopeptidase belongs to the serine-type protease group, with the consensus aa around the catalytic serine (Gly-Xaa-Ser-Xaa-Gly). The active site residues for the human DPP IV/CD26 are Ser-624, Asp-702, and His-734. This characteristic sequential arrangement of the catalytic triad is common for families diverged from the α/β hydrolase subfamily. The CD26 molecule shows a high degree of structural homolgy with Fibroblast Activation Protein α (FAPα) and with another group of proteins having restricted expression in brain tissue, the DPXs.
Disulfides/SH-Groups	There are 12 cysteine residues present and no free sulfhydryl groups could be detected; no evidence for interchain disulfides.
General References	Fleisher, B. (ed.) Dipeptidyl peptidase IV (CD26) in metabolism and the immune response. Springer-Verlag, Heidelberg, 1995. Yaron, A. and Naider, F. *Crit. Rev. Biochem. Molec. Biol.* 1993, **28**:31-81. Morrison, M. et al. *J. Exp. Med.* 1993, **177**:1135-1143. De Meester, I. et al. *J. Immunol. Meth.* 1996, **189**:99-105. Lambeir, A.M. et al. *Biochim. Biophys. Acta* 1997, **1340**:215-226.
Ref. for DNA/AA Sequences	Tanaka, T. et al. *J. Immunol.* 1992, **149**: 481-486. Abbott, C. et al. *Immunogenetics* 1994, **40**:331-338. Böhm, S. et al. *Biochem. J.* 1995, **311**:835-843. Sequence available from GENBANK under accession number M74777.

CD59

Peter J. Sims

Synonyms	P18; MEM-43 antigen; 20 kDa homologous restriction factor (HRF20); Membrane inhibitor of reactive lysis (MIRL); Membrane attack complex inhibitory protein (MIP); Membrane attack complex inhibitory factor (MACIF); Protectin.
Abbreviations	CD59; P18; HRF20; MIRL; MACIF
Classifications	Cell surface glycoprotein; CD59 antigen
Description	A glycosyl-phosphatidylinositol(GPI)-anchored plasma membrane glycoprotein that expresses the CD59 antigen. This protein has been shown to be an inhibitor of the human complement system that protects cells from lysis by interrupting assembly of the membrane attack complex of complement. CD59 is also implicated as an accessory ligand mediating lymphocyte adherence, and may play a role in T-cell activation by antigen presenting cells.
Structure	The mature protein is comprised of a single polypeptide chain consisting of 77 aa that are looped through 5 disulfide bonds. Disulfide bonding pattern and tertiary structure are analogous to that of erabutoxin b and related snake venoms. Other post-translational modifications include N-linked glycosylation at residue Asn-18, and attachment to glycosyl-phosphatidylinositol moiety at residue Asn-77. There is no detectable 0-linked glycosylation. Presence of CD59 in urine, plasma, cerebrospinal fluid and other body fluids suggests that CD59 may also be synthesized as a soluble protein lacking the GPI-anchor, or degraded by hydrolysis of the phospholipid. NMR data are available (MMDB #496-498, MMDB, #890, PDB #1ERH, PDB #1ERG, PDB #1CDQ, PDB #1CDR, PDB #1CDS).
Molecular Weight	18-25 kDa diffuse band by SDS-PAGE. After N-glycanase digest, migrates as discrete 12-14 kDa band. Precursor polypeptide of 128 aa residues (14,177 Da) including 25 residue signal peptide.
Sedimentation Coeff.	Unknown
Isoelectric Point	5.0-6.5 (urea)
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	When extracted from the plasma membrane in non-ionic detergent, CD59 has been reported to noncovalently associate with numerous cellular components, including membrane glycolipids, glycoproteins and cellular tyrosine kinases, including p56lck. The significance of these associations has not been established.
Substrates	None
Inhibitors	Several monoclonal antibodies with inhibitory activity have been reported.
Biological Functions	A membrane inhibitor of the complement system. Prevents assembly of the complement pore by binding to functionally-important domains of comple-

ment proteins C8 and C9 upon their incorporation into the membrane C5b-9 complex. Through these interactions, CD59 inhibits assembly of the membrane-embedded C9 homopolymer. The complement-inhibitory function of CD59 exhibits species-selectivity, and is most pronounced when complement C8 and C9 are derived from human serum. This relative species-restriction in the activity of CD59 accounts for the differential sensitivity of human blood cells to lysis by complement derived from primate versus non-primate sera. In addition to its complement-inhibitory function, there is evidence that CD59 may contribute to T-lymphocyte adherence to antigen presenting cells. CD59 has also been implicated as an accessory ligand for T-cell stimulation, through its affinity for cell-surface CD2.

Physiology/Pathology

CD59 is thought to protect human blood cells, vascular endothelium, and other cells that are normally exposed to complement in blood plasma and other body fluids from the pore-forming and cytolytic activities of the C5b-9 complex. It has been proposed that CD59 on spermatozoa may protect from immune rejection by antibody and complement in the female genital tract. In the acquired hematopoietic stem cell disorder, Paroxysmal Nocturnal Hemoglobinuria (PNH), GPI-anchored membrane proteins are deleted from surface of various blood cells due to a defect in biosynthesis of the glycan moiety of the GPI anchor. The deletion of CD59 from the membranes of PNH erythrocytes renders these cells exquisitely sensitive to lysis by complement, accounting for the intravascular hemolysis seen in this disorder. Decreased cell-surface CD59 has also been proposed to exacerbate tissue damage in a variety of immune and inflammatory disorders. Therapeutic application of exogenous CD59 to protect xenografts from hyperacute rejection by the human complement system has been proposed. Recent evidence suggests that certain pathogenic microorganisms utilize cell-surface proteins that mimic the properties of CD59, in order to evade complement lysis.

Degradation

Protein is resistant to proteolysis in native (non-reduced) form. The presence of soluble CD59 antigen detected in plasma and other body fluids suggests that it is released from the plasma membrane, possibly by hydrolysis of the GPI-anchor. CD59 antigen is also shed from the cell surface during vesiculation of the plasma membrane.

Genetics/Abnormalities

Gene assigned to region p14-p13 of the short arm of chromosome 11. The gene is distributed over 27 kb, and consists of one 5'-untranslated exon and three coding exons. A single patient identified with homozygous gene defect resulting in complete absence of CD59 has been described. This individual is reported to exhibit clinical syndrome of Paroxysmal Nocturnal Hemoglobinuria.

Half-life

Unknown

Concentration

All estimates of membrane surface concentration derive from binding studies performed with monoclonal antibodies: human RBC 10,000-40,000 molecules/cell; human umbilical vein endothelium 150,000-200,000 molecules/cell; Results suggest a plasma membrane surface concentration of approximately 200 molecules CD59 per micron^2.

Isolation Method

Immunoaffinity chromatography of detergent- extracts prepared from wash human erythrocyte ghost membranes.alternatively: butanol/water extraction from human erythrocyte ghost membranes, followed by anion exchange chromatography. Similar methods have been utilized to isolate CD59 from other tissues.The protein is resistant to denaturation in both

ionic and non-ionic detergents, and can be purified by SDS-PAGE under non-reducing conditions without loss of function.

Amino Acid Sequence

Gene codes for a 128 aa polypeptide. Mature protein contains 77 aa residues, including 10 cysteines, all disulfide bonded. Homologies: murine ly6-E antigen; receptor for human urokinase-type plasminogen activator; Herpes saimiri virus (HSV)-15.

Disulfides/SH-Groups

5 intrachain disulfide bonds.

General References

Walsh, L.A. et al. *Tissue Antigens* 1992, **40**:213-220.
Lachmann, P.J. *Immunology Today* 1991, **12**:312-315.
Hahn, W.B. et al. *Science* 1992, **256**:1805-1807.
Ninomiya, H. and Sims, P.J. *J. Biol. Chem.* 1992, **267**:13675-13680.
Meri, S. et al. *Immunol.* 1990, **71**:1-9.
Sugita, Y. et al. *J. Biochem. (Tokyo)* 1988, **104**:633-637.
Hüsler, T. et al. *J. Biol. Chem.* 1995, **270**:3483-3486.
Lockert, D.H. et al. *J. Biol. Chem.* 1995, **270**:19723-19728.
Fletcher, C.M. et al. *Structure* 1994, **2**:185-199 (NMR Structure).
Kieffer, B. et al. *Biochemistry* 1994, **33**:4471-4482 (NMR Structure).

Ref. for DNA/AA Sequences

Sugita, Y. et al. *J. Biochem (Tokyo)* 1989, **106**:555-557.
Davies, A. et al. *J. Exp. Med.* 1989, **170**:637-654.
Philbrick, W.M. et al. *Eur. J. Immunol.* 1990, **20**:87-92.
Sawada, R. et al. *DNA and Cell Biol.* 1990, **9**:213-220.
Tone, M. et al. *J. Molec. Biol.* 1992, **27**:971-976.
Petranka, J.G. et al. *Proc. Natl. Acad. Sci USA* 1992, **89**:7876-7879.

Choriogonadotropin

David Puett

Synonyms	Chorionic Gonadotropin
Abbreviations	CG; hCG
Classifications	Glycoprotein hormone (acidic pI)
Description	A plasma glycoprotein hormone, synthesized by syncytiotrophoblasts, that reaches maximal levels in the first trimester of pregnancy. The hormone is a heterodimer consisting of an alpha subunit of 92 aa residues with two carbohydrate chains N-linked to Asn-52 and Asn-78 and a beta subunit of 145 aa residues with six carbohydrate chains, two linked to Asn-13 and Asn-30 and four O-linked to Ser-121, Ser-127, Ser-132, and Ser-138.
Structure	The crystal structure of HF-treated hCG is known. The alpha and beta subunits do not exhibit aa sequence homology, but their folding patterns are similar. Each subunit contains a cystine knot motif, present in a number of growth factors, consisting of three disulfides, two of which link adjacent extended antiparallel strands and form a ring that is penetrated by the third disulfide. A unique feature of the structure is a segment of the beta subunit that wraps around the alpha subunit and is covalently linked by disulfide 26-110. The molecule is elongated (75 x 35 x 30 Å), and the two subunits are highly intertwined. The subunit interface is extensive, as is the purported receptor interface.
Molecular Weight	38,633: whole molecule (33% carbohydrate) 14,612: alpha-subunit (30% carbohydrate) 24,021: beta-subunit (35% carbohydrate) [determined from chemical composition using average carbohydrate structures].
Sedimentation Coeff.	3.4 S
Isoelectric Point	3.8-5.4 (heterogenous)
Extinction Coeff.	2.1 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Susceptible to inactivation by a number of endo- and exo-proteases.
Biological Functions	Maintains functionality of the corpus luteum in early pregnancy and stimulates testosterone secretion in 46,XY fetal Leydig cells.
Physiology/Pathology	Essential to maintain maternal progesterone levels and thus ensure the continuation of pregnancy until the placenta is sufficiently large to meet the progesterone requirements. Essential for normal male sexual differentiation in utero. Low levels in the first nine weeks of gestation would lead to spontaneous abortion of the fetus. Used clinically to treat various disorders including anovulation, delayed puberty, and hypogonadism. Also

	produced by certain tumors; in some cases plasma levels can indicate rate of tumor growth or responsiveness to therapy.
Degradation	Significant quantities of intact holoprotein are excreted into the urine, which is the principal source for purification. An inactive fragment of the beta subunit (beta core) has been described that results from multiple proteolytic cleavages.
Genetics/Abnormalities	The gene for alpha is present as a single copy on chromosome 6q21.1-23. There are six closely spaced genes for beta located on chromosome 19q13.3, all of which may be expressed in varying degrees. No abnormal forms have been described.
Half-life	6 hrs (rapid component), 36 hrs (slow component)
Concentration	Detectable in maternal serum and urine throughout pregnancy; the maximal concentration occurs in the first trimester, 400-600 nmol/L.
Isolation Method	First trimester pregnancy urine is used as a starting material. Activity is precipitated by kaolin or a benzoic acid-acetone mixture, then purification is achieved by gel filtration and anion exchange chromatography.
Amino Acid Sequence	Homologous with three pituitary glycoprotein hormones, lutropin (LH), follitropin (FSH), and thyrotropin (TSH). All share a common alpha subunit, and the beta subunit confers hormonal specificity. CG and LH are sufficiently similar that they act via the same G protein-coupled receptor. When aligned to maximize homology, the beta subunits of CG and LH contain an additional six or seven aa residues not present on the beta subunits of FSH and TSH; CG beta is also characterized by an extension at the C-terminus of nearly 30 aa residues, containing the four O-linked oligosaccharides, that is not present on the other beta subunits. The highly conserved CAGYC region (aa residues 34-38) on beta has been suggested to be important in alpha binding, and the Keutmann and Ward loops (aa residues 38-57 and 93-100, respectively) are believed to be involved in receptor binding. The five C-terminal aa residues of alpha (YYHKS) are highly conserved in mammals, and this region seems important in receptor binding.
Disulfides/SH-Groups	Eleven (11) intrachain disulfide bonds, no interchain disulfide bonds, and no free sulfhydryl groups: alpha (5 disulfides) and beta (6 disulfides).
General References	Keutmann, H.T. *Mol. Cell. Endocrinology* 1992, **86**:C1-C6. Wehmann, R.E. and Nisula, B.C. *J. Clin.Invest.* 1981, **68**:184-194. Grotjan, Jr., H.E. and Cole, L.C. Human chorionic gonadotropin microheterogeneity. In: *Microheterogeneity of Glycoprotein Hormones.* Keel, B.A. and Grotjan, Jr., H.E. (eds.), CRC Press, Inc., Boca Raton 1989, pp. 219-237. Ryan, R.J. et al. *FASEB J.* 1988, **2**:2661-2669. References for Crystallography Lapthorn, A.J. et al. *Nature* 1994, **369**:455-461. Wu, H. et al. *Structure* 1994, **2**:545-558.
Ref. for DNA/AA Sequences	Mise, T. and Bahl, O.P. *J. Biol. Chem.* 1981, **256**:6587-6592. Fiddes, J.C. and Talmadge, K. *Rec. Progr. Horm. Res.* 1984, **40**:43-74. Jameson, J.L. and Hollenberg, A.N. *Endocrine Rev.* 1993, **14**:203-221. Policastro, P.F. et al. *J. Biol. Chem.* 1986, **261**:5907-5916.

Human Protein Data
Edited by André Haeberli

Chromogranin A

Hans-Hermann Gerdes and Wieland B. Huttner

Synonyms	Secretory protein I, see nomenclature proposal in *Neuroscience* (1987), **21**: 1019–1021
Abbreviations	Chromogranin A: CgA; secretory protein I: SP-I
Classifications	Member of the chromogranins/secretogranins, referred to in abbreviated form as granins.
Descriptions	Highly acidic secretory protein found in secretory granules of a wide variety of endocrine cells and neurons. Together with chromogranin B and secretogranin II, two other members of the granin family, it can be considered as the most widespread marker of the matrix of neuroendocrine secretory granules.
Structure	The mature protein is a single polypeptide of 439 aa. It contains 3–5% O-linked carbohydrate, binds calcium with moderate affinity at multiple sites and aggregates in the presence of this divalent cation (*). These properties probably reflect both the excess of negative charge and the secondary structure of this protein which largely consists of helices and turns as determined by circular dichroism measurements (*). NMR and X-ray diffraction studies are not available.
Molecular Weight	48,918 Da: deduced from cDNA sequence without signal peptide and posttranslational modifications. 75,000 M_r: from SDS-PAGE under reducing conditions.
Sedimentation Coeff.	Unknown
Isoelectric Point	4.9
Extinction Coeff.	Unknown
Enzyme Activity	Unknown
Coenzymes/Cofactors	Unknown
Substrates	Unknown
Inhibitors	Unknown
Biological Function	Various functions have been suggested from observations obtained under *in vitro* conditions. A proposed extracellular function is that the intact protein or proteolytic fragments derived from it exert biological activities on target cells. In this context it was shown that CgA is the precursor of pancreastatin, a peptide which partially inhibits glucose-induced insulin release from isolated pancreas. A proposed intracellular function is that of a helper protein in packaging and/or processing of certain peptide hormones and neuropeptides.
Physiology/Pathology	CgA and its proteolytic products are released from cells by regulated exocytosis into the extracellular space and have been detected in blood plasma of humans. The adrenal medulla has been shown to be the major physiological source. Patients with neuroendocrine tumors have elevated plasma levels of CgA. Thus, estimation of CgA concentrations in blood plasma is useful in the diagnosis of neuroendocrine tumors.

Degradation	In the *trans* Golgi network, and during its storage in neuroendocrine granules, CgA undergoes partial proteolytic processing. This process involves specific endoproteases that cleave the precursor protein at dibasic and monobasic sites. The amount of proteolytic products varies with tissue. Several products have been identified: betagranin and pancreastatin from islet β-cells, corresponding to rat CgA_{1-128} and porcine $CgA_{240-288}$, respectively; chromostatin and vasostatin from adrenal medulla, corresponding to bovine $CgA_{124-143}$ and bovine CgA_{1-76}, respectively; and parastatin from parathyroid, corresponding to porcine $CgA_{347-419}$.
Genetics/Abnormalities	The gene (presumably a single copy gene) was located to the q32.2-q32.3 region of chromosome 14, is (without the promoter region) 13.2 kb long and consists of eight exons. The exon organisation of the gene reveals a striking correspondence to the predicted structural domains of the protein. In particular, the disulfide-bonded loop near the N terminus and the C-terminal region, where CgA is highly homologous to chromogranin B, are encoded by individual exons, i. e. exon 3 and exon 8, respectively.
Half-Life	In blood plasma ~16 min.
Concentration	In blood plasma ~40 ng/mL.
Isolation Methods	a) Preparation of chromaffin granules from pheochromocytoma by sucrose density gradient centrifugation, followed by isolation of soluble matrix proteins, Con A-Sepharose affinity chromatography and gel filtration chromatography on Sephacryl S-300. b) Isolation of chromaffin granules from adrenal medulla by sucrose density gradient centrifugation, followed by the isolation of soluble matrix proteins, Con A-Sepharose, DE-52, Phenyl-Sepharose and calmodulin-Sepharose 4B affinity chromatograhy (*). c) Due to its heat stability a highly enriched soluble fraction of CgA can be obtained from chromaffin cells of adrenal medulla after homogenisation and 5 min boiling at pH 7.4 and 150 mM NaCl followed by centrifugation.
Amino Acid Sequence	Characteristic features are: high content of acidic residues (Glu + Asp = 24.6%) and of Pro (6.6%); N-terminal disulfide-bonded loop structure (residues 17–38) and carboxyterminal sequence (residues 415–439) which are highly homologous to the corresponding domains in chromogranin B, pancreastatin sequence at residues 250–301, potential N-glycosylation site at Asn 92, potential phosphorylation sites at Ser and Tyr, RGD-sequence at residues 43–45, nine potential dibasic cleavage sites.
Disulfides/SH-Groups	One intramolecular disulfide bond between Cys 17 and Cys 38.
General References	Wiedenmann, B. and Huttner, W. B. *Virchows Archiv B Cell Pathol.* 1989, **58:** 95–121. Simon, J.-P. and Aunis, D. *Biochem. J.* 1989, **262**: 1–13. Huttner, W. B. et al. *Trends Biochem. Sci.* 1991, **16:** 27–30. Winkler, H. and Fischer-Colbrie, R. *Neuroscience* 1992, **49**: 497–528. Rosa, P. and Gerdes, H.-H. J., *Endocrinol. Invest.* 1994, **17:** 207–225.
Ref. for DNA/AA Sequences	Konecki, D. S. et al., *J. Biol. Chem.* 1987, **262:** 17026–17030 (aa sequence). Wu, H.-J. et al., *J. Biol. Chem.* 1991, **266:** 13130–13134. Iacangelo, A. L. et al., *Mol. Endocrinol.* 1991, **5:** 1651–1660. Mouland, A. J. et al., *J. Biol. Chem.* 1994, **269:** 6918–6926. Modi, W. S. et al., *Am. J. Genet.* 1989, **45:** 814–818.
	(*) data were obtained for CgA of other species. It can be assumed that they are also valid for human CgA.

Chromogranin B

Hans-Hermann Gerdes and Wieland B. Huttner

Synonyms	Previously also called secretogranin I, see nomenclature proposal in *Neuroscience* (1987), **21**: 1019–1021
Abbreviations	Chromogranin B: CgB; secretogranin I: SgI
Classifications	Member of the chromogranins/secretogranins, referred to in abbreviated form as granins.
Descriptions	Highly acidic secretory protein found in secretory granules of a wide variety of endocrine cells and neurons. Together with chromogranin A and secretogranin II, two other members of the granin family, it can be considered as the most widespread marker of the matrix of neuroendocrine secretory granules.
Structure	The mature protein is a single polypeptide of 657 aa. It contains 3–5% mainly O-linked carbohydrate, binds calcium with moderate affinity at multiple sites and aggregates in the presence of this divalent cation (*). These properties probably reflect both the excess of negative charge and the presumptive secondary structure of this protein, which largely consists of helices and turns. NMR and X-ray diffraction studies are not available.
Molecular Weight	76,295 Da: deduced from the cDNA sequence without signal peptide and posttranslational modifications. 120,000 M_r: from SDS-PAGE under reducing conditions.
Sedimentation Coeff.	Unknown
Isoelectric Point	5.1
Extinction Coeff.	Unknown
Enzyme Activity	Unknown
Coenzymes/Cofactors	Unknown
Substrates	Unknown
Inhibitors	Unknown
Biological Function	Various functions have been suggested from observations obtained under *in vitro* conditions. A proposed extracellular function is that the intact protein or proteolytic fragments derived from it exert biological activities on target cells. A proposed intracellular function is that of a helper protein in packaging and/or processing of certain peptide hormones and neuropeptides.
Physiology/Pathology	CgB and its proteolytic products are physiologically released by regulated exocytosis and may be detected in serum. So far no reliable data on the concentration of hCgB in serum are available to be used as a tool in the diagnosis of neuroendocrine tumors.
Degradation	In the *trans* Golgi network, and during its storage in neuroendocrine granules, hCgB undergoes partial proteolytic processing. This process involves specific endoproteases that cleave the precursor protein at dibasic

and monobasic sites. The amount of proteolytic products varies with tissue. Three proteolytic peptides have been characterized: GAWK and CCB from pituitary, corresponding to $CgB_{420-493}$ and $CgB_{597-653}$, respectively; and BAM-1745 from adrenal medulla, corresponding to bovine $CgB_{547-560}$.

Genetics/Abnormalities

The gene (presumably a single copy gene) was located to the pter-p12 region of chromosome 20. It is (without the promoter region) 12.1 kb long and consists of five exons (*). The exon organisation of the gene reveals a striking correspondence to the predicted structural domains of the protein. In particular, the disulfide-bonded loop near the N terminus, and the C-terminal region, where CgB is highly homologous to chromogranin A, are encoded by individual exons, i. e. exon 3 and exon 5, respectively (*).

Half-Life

Unknown

Concentration

Unknown

Isolation Methods

a) Preparation of chromaffin granules from pheochromocytoma by sucrose density gradient centrifugation, followed by isolation of soluble matrix proteins and HPLC reverse phase chromatography.
b) Due to its heat stability a highly enriched soluble fraction of CgB can be obtained from pheochromocytoma after homogenization and 5 min boiling at pH 7.4 in 150 mM NaCl followed by centrifugation.

Amino Acid Sequence

Characteristic features are: high content of acidic residues, (Glu + Asp = 23.6% and of Pro (5%); N-terminal disulfide-bonded loop structure (residues 16–37) and of carboxyterminal sequence (residues 635–657) which are highly homologous to the corresponding domains in chromogranin A; an identified tyrosine sulfation site at Tyr 321, a potential tyrosine sulfation site at Tyr 153; a potential N-glycosylation site at Asn 295; potential phosphorylation sites at Ser and Tyr, ten potential dibasic cleavage sites, internal homology to the proteolytic products GAWK and CCB at residues 420–493 and 597–653, respectively.

Disulfides/SH-Groups

One intracellular disulfide bond between Cys 16 and Cys 37.

General References

Wiedenmann, B. and Huttner, W. B. *Virchows Archiv B Cell Pathol.* 1989, **58:** 95–121.
Huttner, W. B. et al., *Trends Biochem. Sci.* 1991, **16:** 27–30.
Winkler, H. and Fischer-Colbrie, R. *Neuroscience* 1992, **49**: 497–528.
Rosa, P. and Gerdes, H.-H. *J. Endocrinol. Invest.* 1994, **17:** 207–225.

Ref. for DNA/AA Sequences

Benedum, U. M. et al., *EMBO J.* 1987, **6:** 1203–1211 (aa sequence).
Craig, S. P. et al., *Cytogenet. Cell Genet. (abstract)* 1987, **46:** 600.
Pohl, T. M. et al., *FEBS Lett.* 1990, **262:** 219–224.

(*) data were obtained for CgB of other species. It can be assumed that they are also valid for human CgB.

Chymase

George H. Caughey

Synonyms	Mast cell chymase, skin chymotrypsin-like serine protease
Abbreviations	None
Classifications	EC 3.4.21.39
Description	A chymotrypsin-like serine protease synthesized and secreted by the MC_{TC} subset of mast cells. It is stored in secretory granules along with proteoglycans and other preformed mast cell mediators, e.g., histamine, tryptase, mast cell carboxypeptidase, and cathepsin G. Although chymase is the major protein in murine mast cell granules, it is a relatively minor component (compared to tryptase) of human mast cells. The cloned chymase cDNA and gene predicts an active enzyme of 226 aa modified by N-glycosylation at two sites (Asn-59 and Asn-82). Like several leukocyte and mast cell granule-associated serine proteases, chymase is predicted to be activated in a presecretory intracellular compartment by dipeptidyl peptidase I-mediated removal of an acidic activation dipeptide. Chymase's closest known human relative is cathepsin G, whose substrate preferences are similar but not identical. Mast cells are the main storage site; cardiac endothelium may be a second site of synthesis.
Structure	Tertiary structure not yet determined. Diffraction of crystallized rat chymase 2 (rat mast cell proteinase-II), whose catalytic domain amino acid sequence is 59% identical to that of human chymase but is not glycosylated, predicts an active site and substrate binding region formed by the cleft between twin β-barrel domains, like other trypsin-family serine proteases. The most distinct feature of chymase tertiary structure is a hydrophobic protuberance that looms over the P' side of the substrate binding cleft.
Molecular Weight	29,000–31,000 (by SDS-PAGE and by gel filtration in high ionic strength buffer); released from mast cells with other granule components as a complex with a molecular weight of 400,000–560,000.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown: predicted to be alkaline.
Extinction Coeff.	9.56 (280 nm, 1%, 1 cm), estimated from catalytic domain's Trp and Tyr content.
Enzyme Activity	Chymotrypsin-like endoprotease and peptidase with optimal activity at $\approx$ pH 8. Preferentially cleaves at the carbonyl end of aromatic aa, with Phe > Tyr $\gg$ Trp, in order of preference. Also is an esterase, and appears to be the principal enzyme responsible for the strong hydrolysis of chloroacetyl esters of naphthols by mast cells in enzyme histochemical reactions.
Coenzymes/Cofactors	Secreted in a complex with mast cell carboxypeptidase and sulfated proteoglycans; the latter modulate (but are not required for) chymase activity.
Substrates	Hydrolyzes benzoyl-L-Tyr-ethyl ester (BTEE) at a rate similar to that of chymotrypsin and much faster than that of cathepsin G. 4-Nitroanilide and thiobenzyl ester derivatives of succinyl-L-Val-Pro-Phe are useful for

assays and are very efficiently cleaved. The corresponding derivatives of succinyl-L-Ala-Ala-Pro-Phe and succinyl-L-Phe-Pro-Phe are good available alternatives. The best reported natural substrate is angiotensin I, which is cleaved by chymase between Phe-8 and His-9 to generate angiotensin II, with k_{cat}/K_m at least as high as for angiotensin converting enzyme itself. Human chymase also degrades connective tissue proteins, including components of the dermal-epidermal junction, and activates other proteins, such as interleukin-1β. Purified, highly concentrated chymase undergoes autolytic hydrolysis between Phe-114 and Asn-115, without inactivation.

Inhibitors

In the secretory granule, catalytic activity may be limited by low pH and by steric hindrance from associated proteoglycans. Following secretion from mast cells, chymase is inactivated by extracellular serpins (primarily α_1-antichymotrypsin and α_1-proteinase inhibitor). Rate constants for association with these inhibitors are orders of magnitude lower than for neutrophil cathepsin G and elastase associating with the same inhibitors. Furthermore, inhibition stoichiometry is higher than unity for chymase because of competing catalytic inacitvation of the inhibitors by chymase. Therefore, chymase released from human mast cells may have more opportunity to interact with extracellular targets than do secreted neutrophil serine proteases. α_2-Macroglobulin is a minor inhibitor of chymase incubated with plasma. Nonphysiological inactivators of chymase include soybean and lima bean trypsin inhibitors, eglin C, chymostatin, and various chloromethyl ketone and sulfonyl fluoride derivatives of aromatic peptides and amino acids. Chymase is unusual among serine proteinases in resisting inactivation by aprotinin (bovine pancreatic trypsin inhibitor).

Biological Functions

Unknown; main site of action is thought to be extracellular, although also may have an intracellular role. Data gathered primarily using non-human chymases suggest that the enzymes may facilitate regress of inflammatory cells and plasma into tissues by degrading connective tissue matrix proteins. Chymases also may enhance lipoprotein catabolism and histamine-induced increases in vascular permeability, stimulate airway submucosal gland cell secretion and the extravascular generation of angiotensin, and control complement-mediated and neurogenic inflammation by degrading C3a and inflammatory neuropeptides.

Physiology/Pathology

No proven roles in humans; speculated to be involved in airway hypersecretion in asthma and bronchitis, in inflammation and wheal formation associated with cutaneous immediate hypersensitivity responses, and in host defense against invasion by metazoan parasites.

Degradation

Unknown

Genetics/Abnormalities

The chymase gene maps to cluster of genes encoding serine proteases, including cathepsin G and granzyme B, located on a fragment of chromosome 14q11.2. Although several chymase genes are present in mice and rats, only one has been identified in humans. The intron-exon organization of the gene is similar to that of several other leukocyte and mast cell serine protease genes, including cathepsin G, granzymes, neutrophil elastase and proteinase-3.

Half-life

Unknown, expected to be inactivated soon after release.

Concentration

$\approx$4.5 pg per MC_{TC} mast cell; undetectable in MC_T cells, the other major human mast cell subset. Tissue chymase levels will vary according to local numbers and subtypes of mast cells. In dermis, where MC_{TC} mast cells are most abundant, with concentrations ranging from 1000 to 10,000 cells/

mm^3, tissue chymase concentration may range from 4.5–45 μg per gram of tissue.

Isolation Methods

Several schemes have been used, typically involving extraction of tissue homogenates in high ionic strength buffers, followed by heparin and soybean trypsin inhibitor affinity and hydrophobic interaction chromatography.

Amino Acid Sequence

IIGGTECKPH SRPYMAYLEI VTSNGPSKFC GGFLIRRNFV
LTAAHCAGRS ITVTLGAHNI TEEEDTWQKL EVIKQFRHPK
YNTSTLHHDI MLLKLKEKAS LTLAVGTLPF PSQFNFVPPG
RMCRVAGWGR TGVLKPGSDT LQEVKLRLMD PQACSHFRDF
DHNLQLCVGN PRKTKSAFKG DSGGPLLCAG VAQGIVSYGR
SDAKPPAVFT RISHYRPWIN QILQAN

Sequence is that of the predicted mature catalytic domain, showing "catalytic triad" residues (in bold) common to the active site of serine proteinases.

Disulfides/SH-Groups

3 intrachain Cys-Cys pairs, by analog to rat chymase 2; the human enzyme also appears to have an unpaired Cys (Cys-7 of the catalytic domain), which does not seem to participate in the formation of intermolecular Cys-Cys linkages except as an artifact of isolation.

General References

Schechter, N. M. *Monogr. Allergy.* 1990, **27**:114–131.
Caughey, G. H. *Am. J. Respir. Cell. Mol. Biol.* 1991, **4**:387–394.

Ref. for DNA/AA Sequences

Caughey, G. H. et al. *J. Biol. Chem.* 1991, **226**:12956–12963.
Urata, H. et al. *J. Biol. Chem.* 1991, **226**:17173–17179.
Caughey, G. H. et al. *Genomics* 1993, **15**:614–620.
Schechter, N. M. et al. *J. Immunol.* 1994, **152**:4062–4069.

GenBank accession numbers: M64269 & M69137 (gene); M69136 (cDNA).

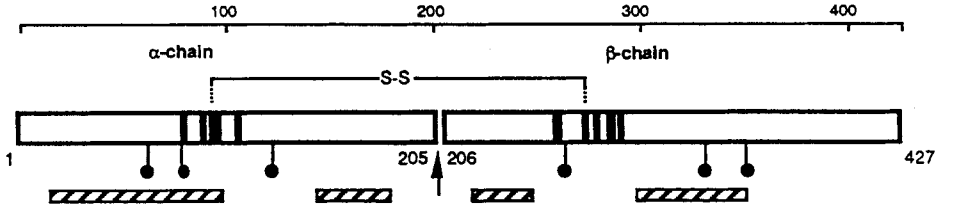
Clusterin

Arcadio Chonn and Jürg Tschopp

Synonyms	SP-40,40; Complement lysis inhibitor; Apolipoprotein J; sulfated glycoprotein-2; Testerone-repressed prostatic messenger 2; gp III; gp-80.
Abbreviations	CLI; apoJ; SGP-2
Classifications	Electr. mob.: α_2 (lipoprotein complex)
Description	First purified from ram rete testis fluid as a protein that elicits aggregation or clustering of Sertoli cells or erythrocytes. In humans, plasma glycoprotein associated with complement membrane attack complexes, immunoglobulin complexes and with a subpopulation of high density lipoproteins often containing apolipoprotein A-1, cholesteryl ester transfer protein and approximately 22% (w/w) lipids. Also, a major glycoprotein found in human seminal fluid. Widely expressed with highest levels detected in brain, ovary, testis and liver; especially in epithelial cells. Expressed in platelets, but not in lymphocytes.
Structure	Mature protein is a heterodimer consisting of one α and one β subunit, which are derived from a single precursor. Glycoprotein with approximately 30% N-linked sulfated carbohydrates. Four heparin binding domains and 3–4 amphiphathic helix domains predicted from the linear protein sequence.
Molecular Weight	70,000–80,000 (nonreducing SDS-PAGE) 35,000–40,000: α and β subunits, human (reducing SDS-PAGE) 24,000: deglycosylated α subunit (reducing SDS-PAGE) 27,000: deglycosylated β subunit (reducing SDS-PAGE).
Sedimentation Coeff.	Unknown
Isoelectric Point	4.9–5.4 α and β subunits
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Chloroquine blocks the proteolytic maturation of clusterin.
Biological Functions	Capable of binding to hydrophobic complexes such as the C5b-7 complex of the terminal membrane attack complex, neutral lipids and high density lipoproteins. Ability of clusterin to aggregate cells in vitro suggests a role in cell/cell interactions, notably in fertilization. The membrane receptors for clusterin have yet to be identified.
Physiology/Pathology	Message is highly induced in majority of tissues undergoing programmed cell death (apoptosis); for example in androgen-repressed prostate, in vitamin A-deprived seminal vesicle epithelia, and in kidney injured by increased hydrostatic pressure. Enhanced expression in Alzheimer's disease or Pick's disease brain tissue. Increased expression in retinitis pigmentosa

	retinas. In situ hybridization studies of developing mouse embryos suggest a role in tissue remodelling. High density lipoprotein association suggests a potential role in cholesterol transport. Clusterin level is positively correlated with total plasma triglyceride and cholesterol levels; however, no significant relationships are apparent between plasma apoJ levels and high density lipoprotein cholesterol or apolipoprotein A-1.
Degradation	Unknown
Genetics/Abnormalities	Gene locus mapped to chromosome 8p21. Three isoforms have been identified as apoJ-1, apoJ-2 and apoJ-3.
Half-life	Unknown
Concentration	35–105 mg/L in human serum; 2–15 g/L in human seminal fluid.
Isolation Method	Immunoaffinity chromatography, sources are human plasma and semen.
Amino Acid Sequence	Signal peptide of 22 residues; the 427 aa precursor protein is post-translationally cleaved into the two subunits after Arg-205.
Disulfides/SH-Groups	Ten disulfides (5/subunit) involved in interchain disulfide bond of α and β subunits.
General References	Jenne, D. E. and Tschopp, J. *Trends Biochem. Sci.* 1992, **17:** 154–159. Jordan-Starck et al. *Current Opinion Lipidology* 1992, **3:** 75–85. Jenne, D. E. and Tschopp, J. *Proc. Natl. Acad. Sci. USA* 1989, **86:** 7123–7127. de Silva, H. V. et al. *J. Biol. Chem.* 1990, **265:** 14292–14297. O'Brian, M. K. et al. *J. Clin. Invest.* 1990, **85:** 1477–1486.
Ref. for DNA/AA Sequences	Jenne, D. E. and Tschopp, J. *Proc. Natl. Acad. Sci. USA* 1989, **86:** 7123–7127, human: SwissProt P10909. Collard M. W. and Griswold M. D. *Biochemistry* 1987, **26:** 3297–3303, rat: SwissProt P05371; dog: SwissProt P25473; bovine: SwissProt P17697; mouse: GenBank L08235.

Molecular Model



Schematic structural organization of human clusterin. Clusterin is a disulfide-linked heterodimer with an α and β chain of approximately equal size. The 6 putative N-linked sulfated carbohydrate side chains (⏚) are evenly distributed between the two chains. The two subunits each have 5 clustered cysteines (dotted boxes) forming interchain disulfide bridrges. Dashed rectangles represent predicted amphipathic helices.

Introduction to Collagens

Michel van der Rest

Definition

A collagen is currently defined as a structural protein from the extracellular matrix that contains one or more domain in the characteristic triple helical conformation.

Description of the triple helix

The collagen triple helix (see fig.) is made of three polypeptide chains each forming a left handed helix of the proline type. The three helical chains intertwine to form a right handed superhelix. Every third residue of each chain occupies the centre of the helix and is displaced by 30° from the preceeding central residue. Only glycyl residues can occupy this position without interference with the triple helix structure. The primary structure of the triple helical domains is thus $(Gly\text{-}Xaa\text{-}Yaa)_n$. About 30% of the Xaa positions have to be occupied by prolyl residues and about 30% of Yaa positions by hydroxyprolyl residues for the triple helix to be stable at physiological temperature.

Properties of the triple helix

Triple helical domains form relatively rigid molecular rods (approximately 0.29 nm long/amino acyl residue and 1.5 nm in diameter) and have 2/3 of their residues with their side-chains pointing outward of the helix, favoring the formation of lateral interactions between adjacent triple helices. Parallel, staggered and antiparallel interactions have been observed in the supramolecular aggregates formed by collagen molecules.

Protein nomenclature

Collagen types have been numbered (in roman numbers) according to their order of discovery. Individual polypeptide chains are called α chains and are numbered within each type in arabic numbers. For example, the α1 chain of type I collagen is called α1(I). It is assumed that molecules can only form from chains of the same type. Recent data indicate however that this is not always true.

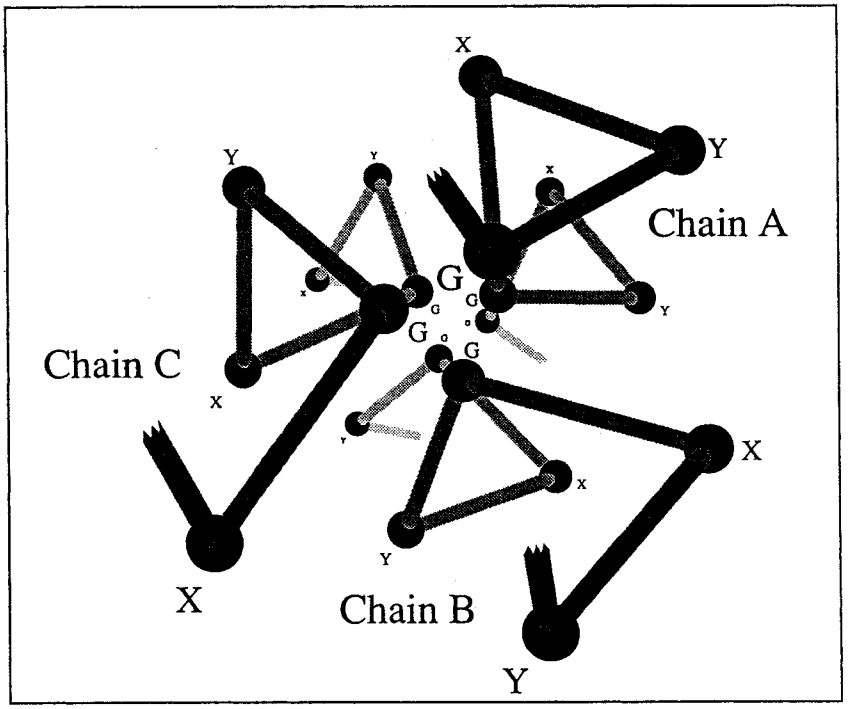
Gene nomenclature

The gene coding for a given collagen chain is designated by COL(type number in arabic number)A(chain number). For example, the gene coding for the α2(XI) chain is called COL11A2.

General References

Vuorio, E. and de Crombrugghe, B. The family of collagen genes. *Annu. Rev. Biochem.* 1990, **59**:837-872.
van der Rest, M. and Garrone, R. Collagen family of proteins. *FASEB J.* 1991, **5**:2814-2823.
Prockop, D.J. and Kivirikko, K.I. Collagens: Molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.* 1995, **64**:403-434.

Ref. for X-ray structure of triple helix
Bella, J., Eaton, M., Brodsky, B., Berman, H.M. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* 1994, **266**:75-81.



Schematic representation of a cross-section through a collagen triple helix. The balls represent the α carbons of the amino-acyl residues and the sticks represent the peptide bonds. The side chains of the residues are omitted. (Reprinted from van der Rest and Garrone, *Biochimie*, **72:**473–484, 1990, with permission).

Collagen type I

Michel van der Rest

Synonyms	Collagen I; Gelatin (after denaturation)
Abbreviations	None
Classifications	Fibrous protein
Description	Collagen I, the most abundant protein in vertebrates, is expressed in most connective tissues. It assembles, together with other collagens, into striated fibrils that confer tensile strength to these tissues. It undergoes during its biosynthesis extensive and specific posttranslational modifications that include hydroxylation of prolyl residues in the Yaa position of the triplets Gly-Xaa-Yaa, hydroxylation of specific lysyl residues, glycosylation of hydroxylysyl residues, cleavage of propeptides, and oxidation of the ε-NH_2 of specific lysyl or hydroxylysyl residues into aldehydes. These aldehydes further react in a complex way with other amino-acyl residues to form covalent crosslinks that stabilize the supramolecular assembly.
Structure	A heterotrimeric molecule made of two distinct polypeptide chains, α1(I) and α2(I) in a 2:1 stoichiometry. It is secreted as procollagen (see fig. 1). The N- and C-propeptides are cleaved in the extracellular space. The processed molecule is 294 nm in length and 1.5 nm in diameter. A homotrimeric $[\alpha 1(I)]_3$ molecule has been described as a minor constituent of some tissues, in tissue culture and in pathological conditions.
Molecular Weight	284 kDa for the processed molecule
Sedimentation Coeff.	3.0 S
Isoelectric Point	6.6
Extinction Coeff.	The processed collagen I molecule has no tryptophan and very few tyrosines. It is best detected by absorbance of the peptide bond at 230nm or below. No extinction coefficient has been reported.
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	The collagen I molecules assemble in a quarter-staggered array and are the major structural constituents of the striated fibrils of connective tissues and of the organic matrix of bone (fig. 2). Stabilized by the formation of intra- and intermolecular lysine-derived crosslinks, these fibrils are responsible for the tensile strength of connective tissues. Gaps between collagen I molecules in the fibrils are the initial sites of mineral deposition in bone matrix. The cleaved N-propeptide has a negative feed-back on collagen I synthesis. Trimer formation occurs intracellularly and is initiated by the C-propeptide. Triple helix folding occurs then by a "zipper" mechanism. Collagen I is a ligand for integrins $\alpha 1\beta 1$, $\alpha 2\beta 1$ and $\alpha 3\beta 1$.

Physiology/Pathology	Collagen I expression is tightly regulated. Oversynthesis of collagen I is the hallmark of the various fibrotic processes, such as lung or liver fibrosis.
Degradation	Only collagenases can cleave the native triple helical domain of type I collagen. Interstitial collagenase (matrix metalloproteinase-1, MMP-1)) makes a single clip between Gly-775 and Ile-776 (Leu in the α2(I) chain). The resulting 3/4 and 1/4 fragments denature at physiological temperature and can then be degraded by general proteolysis. Crosslinked fibrillar collagen I molecules can be solubilized by several enzymes (elastase, pepsin) that cleave the telopeptides that are the sites of crosslinks.
Genetics/Abnormalities	The genes encoding the proα1(I) chain (COL1A1) and the proα2(I) chain (COL1A2) are located on chromosome 17 at q21.3-q22 and on chromosome 7 at q21.3-q22.1 respectively. Over 300 different point mutations have been described in either COL1A1 or COL1A2 genes. Most of these mutations are dominant and result in the phenotype of Osteogenesis Imperfecta (OI). Many lethal cases of OI are due to the substitution of a Gly residue occupying the center of the helix by another amino acid, resulting in alteration of helix stability. The severity varies however with the position of the mutation and with the nature of the substituting aminoacyl residue. COL1A2 mutations appear less severe than similar COL1A1 mutations. Splicing mutations affecting exon 6 of either COL1A1 or COL1A2, which contain the sequence coding for the N-protease cleavage site, result, because of an impaired N-terminal processing, in the dominant Ehlers-Danlos sydrome type VII which is characterized by extreme joint hypermobility.
Half-life	Unknown (variations from and within tissues)
Concentration	Type I collagen is the most abundant protein in vertebrates and represents about 25% of the proteins of the human body.
Isolation Method	Small quantities of native human type I collagen can be extracted from minced skin by extraction with 0.5 M acetic acid. Larger quantities can be obtained if the tissue is digested with pepsin but the telopeptides are cleaved off (see above). Collagen I can be purified by salt precipitation, first in 0.5 M acetic acid, where it precipitates in 0.7 M NaCl together with type III collagen. The precipitate is then redisolved in 0.1 M acetic acid and dialysed against 50 mM Tris-HCl, pH 7.5, 0.52 M NaCl. Increasing the NaCl concentration to 1.71 M precipitates collagen III. Collagen I can then be precipitated at 2.56 M NaCl. Human type I procollagen can be obtained in very small quantity from cultured human fibroblasts. Labelling with radioactive proline for up to 48 h in serum-free medium is advisable. It is necessary to include 50μg/ml ascorbate in the culture medium to obtain high secretion level of procollagen. Addition of 50μg/ml β-aminopropionitrile to prevent crosslinking will increase the yield of collagen extraction from the cell layer. Extraction of procollagen must be performed in the presence of a cocktail of protease inhibitors (25 mM EDTA, 0.2 mM phenylmethanesulfonyl fluoride, 10 mM N-ethylmaleimide). Procollagen can be precipitated with ammonium sulfate at 30% saturation or the medium can be dialyzed against 1 mM ammonium bicarbonate, 0.1 mM phenylmethanesulfonyl fluoride, 0.5 mM N-ethylmaleimide and lyophilized. The major labelled constituent obtained by this procedure is type I procollagen, but many other extracellular constituents are also present in minor amounts.

Amino Acid Sequence	Residues in α1(I) chain	Residues in α2(I)chain

Signal peptide	22	22
N-propeptide (non-triple helical)	86	11
N-propeptide (short triple-helix)	45	45
N-telopeptide (with N-protease cleavage site)	25	3
Main triple helix	1014	1014
C-telopeptide (with C-protease cleavage site)	26	15
C-propeptide	246	245

Disulfides/SH-Groups

10 cysteines in the non-triple helical N-propeptide region and 8 cysteines in the C-propeptide of α1(I) chain; 7 cysteines in the C-propeptide of the α2(I) chain.
Interchain disulfide bonds are formed in the C-propeptide.

General References

Engel, J. and Prockop, D.J. The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu. Rev. Biophys. Biophys. Chem.* 1991, **20**:137-152.
Eyre, D. Collagen cross-linking amino acids. *Methods. Enzymol.* 1987, **144**:115-139.
Fietzek, P.P. and Kühn, K. The primary structure of collagen. *Int. Rev. Connect. Tissue Res.* 1976, **7**:1-60.
Prockop, D.J., et al. The biosynthesis of collagen and its disorders (first of two parts) *N. Engl. J. Med.* 1979, **301**:13-23.
Prockop, D.J. et al. The biosynthesis of collagen and its disorders (second of two parts) *N. Engl. J. Med.* 1979, **301**:77-85.
Prockop, D.J. et al. Type I procollagen: The gene-protein system that harbors most of the mutations causing osteogenesis imperfecta and probably more common heritable disorders of connective tissue. *Am. J. Med. Genet.* 1989, **34**:60-67.
van der Rest, M. The collagens of bone. In *Bone: A treatise. Bone matrix and bone-specific products.* Vol.3, Hall, B.K. (ed.) The Tellford Press, Caldwell, NJ, pp. 187-237.
van der Rest, M. and Garrone, R. Collagen family of proteins. *FASEB J.* 1991, **5**:2814-2823.
Prockop, D.J. and Kivirikko, K.I. Collagens: Molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.* 1995, **64**:403-434.

Ref. for DNA/AA Sequences

For the proα1(I) chain:
Chu, M.L. et al. *Nucleic Acid Res.* 1982, **10**:5925-5934.
Bernard, M.P. et al. *Biochemistry* 1983, **22**:5213-5223.
For the proα2(I) chain:
DeWet, W. et al. *J. Biol. Chem.* 1987, **262**:16032-16036.

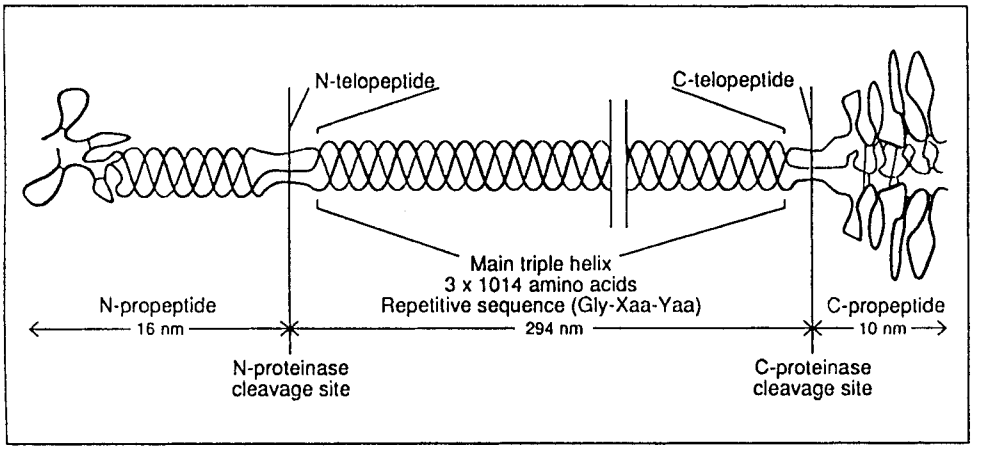


Figure 1. Schematic representation of a procollagen I molecule. (Reprinted from van der Rest, M., 1990, with permission).

Collagen type II

Michel van der Rest

Synonyms	Collagen II
Abbreviations	None
Classifications	Fibrous protein
Description	Collagen II is the predominant protein of cartilages. It assembles, together with collagens IX and XI, into striated fibrils. Like type I collagen, it undergoes during its biosynthesis extensive and specific posttranslational modifications that include hydroxylation of prolyl residues in the Yaa position of the triplets Gly-Xaa-Yaa, hydroxylation of specific lysyl residues, glycosylation of hydroxylysyl residues, cleavage of propeptides, and oxidation of the ε-NH_2 of specific lysyl or hydroxylysyl residues into aldehydes. These aldehydes further react in a complex way with other amino-acyl residues to form covalent crosslinks that stabilize the supramolecular assembly. The extent of lysyl hydroxylation and glycosylation is more extensive in type II collagen than in type I collagen. The major crosslinks are of trivalent fluorescent pyridinoline type.
Structure	Collagen II is homotrimeric molecule, $[\alpha 1(II)]_3$. It is secreted as procollagen (see fig. 1). The N- and C-propeptides are cleaved in the extracellular space. Two splicing variants exist, type IIA and type IIB. In type IIA, the globular region of the N-propeptide is 76 residues long. In type IIB, it is shorter by 69 residues and does not contain any cysteine. The $\alpha 1(II)$ chain (probably IIB) can participate in a heterotrimeric molecular assembly with the $\alpha 1(XI)$ and $\alpha 2(XI)$ chains and is then designated $\alpha 3(XI)$.
Molecular Weight	284 kDa for the processed molecule
Sedimentation Coeff.	3.0 S (analogy with type I collagen)
Isoelectric Point	Unknown
Extinction Coeff.	The processed collagen II molecule has no tryptophan and very few tyrosines. It is best detected by absorbance of the peptide bond at 230nm or below. No extinction coefficient has been reported.
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	The collagen II molecules assemble in a quarter-staggered array and are the major structural constituents of the striated fibrils of cartilage and vitreous humour. The tensile strength of these fibrils oppose the swelling pressure generated by proteoglycans. These fibrils are stabilized by the formation of intra- and intermolecular lysine-derived crosslinks. Trimer formation occurs intracellularly and is initiated by the C-propeptide. Triple helix folding occurs then by a "zipper" mechanism.

Collagen II is a ligand for anchorin CII (a member of the annexin family), a peripheral membrane protein found in chondrocytes.

Physiology/Pathology

Collagen II degradation is the hallmark of degenerative disease of cartilage. Autoantibodies to type II collagen have been observed in some cases of rhumatoid arthritis.

Degradation

Only collagenases can cleave the native triple helical domain of type II collagen. Interstitial collagenase (matrix metalloproteinase-1, MMP-1) makes a single clip between Gly-775 and Leu-776. The resulting 3/4 and 1/4 fragments denature at physiological temperature and can then be degraded by general proteolysis. Crosslinked fibrillar collagen II molecules can be solubilized by several enzymes (elastase, pepsin) that cleave the telopeptides that are the sites of the crosslinks.

Genetics/Abnormalities

The gene encoding the proα1(II) chain (COL2A1) is located on chromosome 12 at q14.3. Several diseases involving cartilage have been shown to be caused by mutations in COL2A1. Spondyloepiphyseal dysplasia congenita (SEDc) has been linked with the gene for type II collagen. The mutations described include a 36 residue deletion in exon 48, a 45 bp duplication in the same exon, a RNA-splicing mutation, several glycine substitutions and recurrent Arg789→Cys substitutions. A Arg519→Cys substitution has been described in a familial osteoarthrosis associated with mild SEDc. Several premature termination mutations have been found in Stickler syndrome. Gly→Ser substitutions at positions 574 and 805 have been observed in hypochondrogenesis. Gly-943→Ser and Gly310→Asp were demonstrated in achondrogenesis II (or IB-hypochondrogenesis). A case of Kniest dysplasia with Gly103→Asp has also been reported.

Half-life

Unknown

Concentration

Type II collagen represents about 60% of the dry weight of cartilage and 90% of the collagen of this tissue.

Isolation Method

Collagen II can be solubilized from cartilage if the tissue is digested with pepsin. The telopeptides are cleaved off (see above). Collagen II can be purified by salt precipitation in 0.5 M acteic acid where it precipitates at 0.7 M NaCl.

Amino Acid Sequence

	Residue in α1(II) chain
Signal peptide	25
N-propeptide (non-triple helical)	76(IIA) or 7(IIB)
N-propeptide (short triple helix)	79
N-telopeptide (with N-protease cleavage site)	24
Main triple helix	1014
C-telopeptide (with C-protease cleavage site)	27
C-propeptide	246

Disulfides/SH-Groups

10 cysteines in the non-triple helical N-propeptide region of IIA (none in IIB) and 8 cysteines in the C-propeptide. Interchain disulfide bonds are only formed in the C-propeptide.

General References

Ahmad, N.N. et al. Stop codon in the procollagen II gene (*COL2A1*) in a family with the Stickler syndrome (arthro-ophtalmopathy). *Proc. Natl. Acad. Sci. USA* 1991, **88**:6624-6627.

Anderson, I.J., et al. Spondyloepiphyseal dysplasia congenita: Genetic linkage to type II collagen (COL2A1). *Am. J. Hum. Genet.* 1990, **46**:896-901.

Bonaventura, J. et al. Substitution of aspartic acid for glycine at position 310 in type II collagen produces achondrogenesis II, and substitution of serine at position 805 produces hypochondrogenesis: Analysis of genotype-phenotype relationships. *Biochem. J.* 1995, **307**:823-830.
Dodge, G.R. and Poole, A.R. Immunohistochemical detection and immunochemical analysis of type II collagen degradation in human normal, rheumatoid, and osteoarthritic articular cartilages and in explants of bovine articular cartilage cultured with Interleukin 1. *J. Clin. Invest.* 1989, **83**: 647-661.
Eyre, D. Collagen cross-linking amino acids. *Meth. Enzymol.* 1987, **144**: 115-139.
Prockop, D.J. and Kivirikko, K.I. Collagens: Molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.* 1995, **64**:403-434.
Ritvaniemi, P. et al. Identification of COL2A1 gene mutations in patients with chondrodysplasias and familial osteoarthritis. *Arthritis Rheum.* 1995, **38**:999-1004.
van der Rest, M. and Garrone, R. Collagen family of proteins. *FASEB J.* 1991, **5**:2814-2823.

Ref. for DNA/AA Sequences

Su, M.W. et al. *Nucleic Acid Res.* 1989,**17**:9473.
Baldwin, C.T. et al. *Biochem. J.* 1989,**262**:521-528.

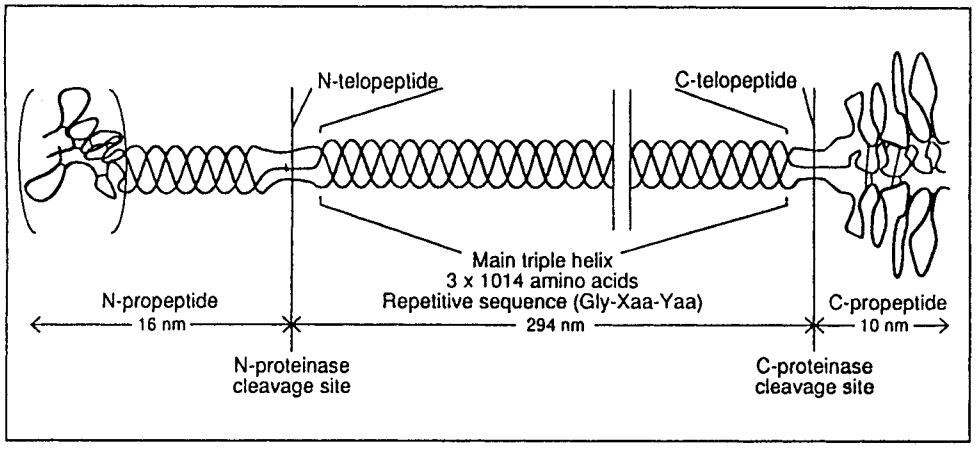


Figure 1. Schematic representation of a procollagen II molecule. The alternatively spliced region in the N-propeptide is placed in parentheses.

Collagen type V

Mahnaz Moradi-Améli and Michel van der Rest

Synonyms	Collagen V; A-B collagen
Abbreviations	None
Classifications	Fibrous protein
Description	Collagen V is a fibril forming, or fibrillar, collagen which represents a minor percentage of the fibrous components in extracellular matrix. Its overall structure is similar to that of collagen I. It is coexpressed with type I collagen in many connective tissues. It is capable, at least in vitro, of forming thin homotypic fibrils but has been observed in vivo in combination with collagen I and collagen III when present to form heterotypic interstitial fibers. Type V collagen is masked in such heterotypic fibrils since immunostaining can only be achieved after partial disruption of the fibril structure.
Structure	Heterotrimeric molecules can be made from three distinct polypeptide chains, α1(V), α2(V) and α3(V) in $[\alpha1(V)]_2\alpha2(V)$ or α1(V)α2(V)α3(V) stoichiometries, the former one being the more abundant. A homotrimeric form $[\alpha1(V)]_3$ has been reported in cell culture. A heterotypic form of collagens V and XI in $[\alpha1(XI)]_2\alpha2(V)$ form also exists in some human cell line and in vivo. Collagen V chains are synthesized and secreted as procollagen in the extracellular space. They undergo a partial processing which involves removal of the C-propeptide. In addition, the α1(V)chain is processed at the N-terminus. The main cleavage occurs at a Pro-Ala site 8 residues away from the short triple helix toward the N-terminus. There is also some evidence for a complete processing of some α1(V) chains, possibly corresponding to the $[\alpha1(V)]_3$ homotrimer. The α2(V) chain is not processed at its N-terminus. Nothing is known about the processing of the α3(V) chain.
Molecular Weight	444 kDa for the intact form of the molecule.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	It is best detected by absorbance of the peptide bond at 230nm or below. No extinction coefficient has been reported.
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Collagen V combines with collagen I to form heterotypic interstitial fibrils and in this way it may play a role in fibrillogenesis by regulating the growth and orientation of type I collagen fibrils. Collagen V binds specifically thrombospondin, heparan sulfate proteoglycans, heparin, insulin, osteonectin, and macrophage colony-stimulating factor. Although collagen

V is reported to mediate cell adhesion its anti-adhesive activity has been reported as well in some tissues.

Physiology/Pathology

Collagen V production is increased in inflammatory and fibrotic conditions, in some forms of cancer and in atherosclerotic tissues.

Degradation

A neutral metalloproteinase, termed gelatinase (MMP-9), specifically degrades collagen V. It is otherwise resistant to interstitial collagenase (MMP-1). Clostridiopeptidase A (EC 3.4.24.3) can cleave the native triple helical domains. The denatured molecule can be cleaved by general proteolysis. Crosslinked molecules can be solubilized by enzymes such as pepsin that cleaves the telopeptides and the non-helical extension remaining on the molecules.

Genetics/Abnormalities

The genes encoding the proα1(V) chain (COL5A1) and the proα2(V) chain (COL5A2) are located on chromosome 9 at q34.2-q34.3 and chromosome 2 at q24.3-q31, respectively. The COL5A1 has 66 exons and is shown to be considerably diverged from the conserved structure of the genes for major fibrillar collagen types I-III. The COL5A2 seems to be evalutionary related to the latter collagen genes. The COL5A1 gene have been shown to be linked to Ehlers-Danlos Syndrome (EDS) types I and II which are characterized by joint hypermobility and skin fragility. A targeted mutation in the COL5A2 gene in transgenic mice (exon 6 deletion) resulted in a severe recessive condition with spinal deformities and skin and eye abnormalities.

Half-life

Unknown

Concentration

Type V collagen was identified in many tissues such as blood vessel wall, placenta, bone, skin, tendon, skeletal muscle, corneal stroma and lung in a minor amount. Its concentration does not exceed 5% of collagen I.

Isolation Method

Very small amount of native type V collagen can be extracted from human placenta or from demineralized matrix of human fetal bone by extraction with 0.5 M acetic acid. Larger quantities can be obtained if the tissue is digested with pepsin but the non-helical extensions and the telopeptides are cleaved. Collagen V can be purified by differential salt precipitation between 0.9 M and 1.2 M NaCl in 0.5 M acetic acid.

Amino Acid Sequence

	Residues in α1(V)chain	Residues in α2(V)chain
Signal peptide	36	26
N-propeptide (non-triple helical)	407	83
N-propeptide (short triple helix)	95	79
N-telopeptide	20	24
Main triple helix	1014	1014
C-telopeptide	22	24
C-propeptide	244	246

Disulfides/SH-Groups

4 cysteines in the non-triple helical N-propeptide region and 8 cysteines in the C-propeptide of the α1(V) chain. 10 cysteines in the N-propeptide and 7 cysteines in the C-propeptide of the α2(V) chain. Interchain disulfide bonds may only be formed in the C-propeptide.

General References

Andrikopoulos, K. et al. Targeted mutation in the *col5a2* gene reveals a regulatory role for type V collagen during matrix assembly. *Nature Genet.* 1995, **9**:31-36.

Fessler, J.H. and Fessler, L.I. Type V Collagen. In: *Structure and function of collagen types,* Mayne, R. and Burgeson, R.E. (eds.), 1987, Academic Press, pp. 81-103.
Moradi-Améli, M. et al. Diversity in the processing events at the N-terminus of type-V collagen. *Eur. J. Biochem.* 1994, **221**:987-995.
Niyibizi, C. and Eyre, D.R. Structural analysis of the extension peptides on matrix forms of type V collagen in fetal calf bone and skin. *Biochim. Biophys. Acta* 1993, **1203**:304-309.
van der Rest, M. and Garrone, R. Collagen family of proteins. *FASEB J.* 1991, **5**:2814-2823.
Fichard, A., Kleman, J.P. and Ruggiero, F. Another look at collagen V and XI molecules. *Matrix Biol.* 1995, **14**:515-531.

Ref. for DNA/AA Sequences

For the proα1(V) chain:
Takahara, K. et al. *J. Biol. Chem.* 1991, **266**:13124-13129.
Greenspan, D.S. et al. *J. Biol. Chem.* 1991, **266**:24727-24733.
For the proα2(V) chain:
Myers, J.C. et al. *J. Biol. Chem.* 1985, **260**:11216-11222.
Weil, D. et al. *Nucleic Acids Res.* 1987, **15**:181-198.
Woodbury, et al. *J. Biol. Chem.* 1989, **264**:2735-2738.

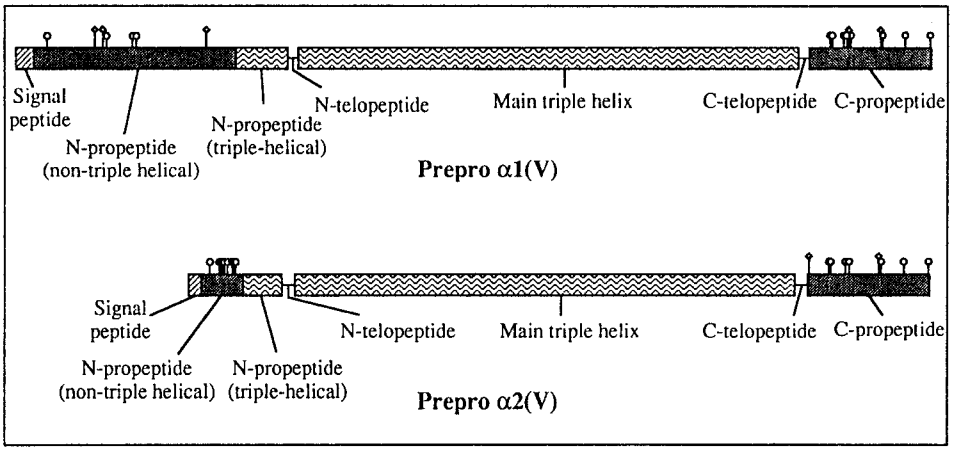


Figure 1. Schematic representation of the structures of the prepro-type V collagen chains. The triple helical domains are represented by (▨), the globular domains by (▩), the signal peptide by (▨) and the junction regions containing demonstrated or putative cleavage sites for the N- and C-proteinases by a solid bar. The positions of cysteine residues (♀) and of demonstrated or putative N-glycosylation sites (♀) are indicated.

Collagen type VI

Beat Trueb

Synonyms	Collagen VI; Intima collagen; Short chain collagen
Abbreviations	Type VI; [α1(VI)α2(VI)α3(VI)]
Classifications	Fibrous protein
Description	Collagen VI is a ubiquitous glycoprotein with a central triple helix of 105 nm flanked at each end by globular domains. It is composed of three different polypeptide chains, α1(VI), α2(VI) and α3(VI), which form a heterotrimeric molecule (see Fig. 1). Collagen VI assembles into microfibrils of 5 nm diameter that encircle various biological structures, including fibroblasts, chondrocytes, interstitial collagen fibers, blood vessels and nerves.
Structure	Triple stranded monomers align in an anti-parallel fashion with a stagger of 30 nm to form disulfide-linked dimers. The triple helices of such a dimer are intertwined in a superhelix. Dimers associate laterally to form disulfide-bonded tetramers, and tetramers aggregate end-to-end to form filamentous structures (see Fig. 2).
Molecular Weight	α1(VI): 140 kDa; α2(VI): 140 kDa; α3(VI): 260-350 kDa (SDS-PAGE)
Sedimentation Coeff.	Unknown
Isoelectric Point	α1(VI): 5.1; α2(VI): 6.0; α3(VI): 6.6 (predicted from aa sequence)
Extinction Coeff.	216,300 $cm^{-1}M^{-1}$ (predicted)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Collagen VI promotes attachment and spreading of fibroblasts in vitro. It interacts with integrins of the β1 family, suggesting an important function in the anchorage of cells to the extracellular matrix. Since it specifically binds to von Willebrand factor it might also play a role during haemostasis. Additional interactions have been described between collagen VI and interstitial collagens, proteoglycans and glycosaminoglycans.
Physiology/Pathology	Expression of collagen VI is increased by TGF-β and decreased by interferon γ. Synthesis of the α3(VI) chain appears to be the rate-limiting step in the assembly of stable heterotrimers. Synthesis and/or turnover of collagen VI is increased in fibrotic and regenerating tissues.
Degradation	The globular domains are degraded by various proteases, in particular by serine proteases. The triple helix is resistant to proteolytic attack.
Genetics/Abnormalities	COL6A1 and COL6A2 are both located on chromosome 21 in band q22.3. COL6A3 lies on chromosome 2 in the region q37.3. Mutations in the genes for collagen VI cause Bethlem myopathy, an autosomal dominant myopathy with contractures.
Half-life	Unknown

Concentration

Collagen VI is found in most connective tissues in amounts similar to or higher than collagen V. In the cornea, it appears to be a major constituent contributing up to 25% of the total protein.

Isolation Method

A fragment of collagen VI representing essentially the central triple helix is obtained by pepsin digestion of placenta or other tissues, followed by fractionated salt precipitation. A more intact form of collagen VI can be extracted with denaturing solvents such as 6 M guanidinium chloride and purified by gel filtration in the presence of detergents. More gentle solubilization is achieved by digestion of fetal tissues with bacterial collagenase or hyaluronidase.

Amino Acid Sequence

	Residues in α1(VI) chain	Residues in α2(VI) chain	Residues in α3(VI) chain
Signal peptide	1-19	1-20	1-25
N-globular domain	20-256	21-254	26-2036
Triple Helix	257-592	255-589	2037-2372
C-globular domain	593-1028	590-1018	2373-3175

The globular domains are composed of multiple repeats that are related to the A domains of von Willebrand factor. Three of these VWFA motifs are found in each the α1(VI) and the α2(VI) chains, 12 are found in the α3(VI) chain. Several of these repeats are subject to alternative splicing events. The C-terminal repeat of the α2(VI) chain can be deleted or replaced by a fibronectin type III repeat, giving rise to three mutually exclusive isoforms. Multiple isoforms have also been described for the α3(VI) chain, where at least four VWFA motifs are subject to alternative splicing. The C-terminal end of the α3(VI) chain contains three additional domains, a threonine-rich segment, a fibronectin type III repeat and a motif related to the Kunitz-type of protease inhibitors (see Fig. 1).

Disulfides/SH-Groups

Each of the three chains contains a single cysteine residue within the collagenous sequence. These residues are involved in the formation of disulfide-linked dimers and tetramers (see Fig. 2). Clusters of cysteines are found adjacent to the collagenous sequence within the globular domains. Some of these residues participate in the formation of disulfide bonds between the three chains in a type VI collagen heterotrimer.

General References

Timpl, R. and Chu, M.-L. Microfibrillar collagen type VI. In: *Extracellular Matrix Assembly and Structure.* Yurchenco, P.D., Birk, D.E. and Mecham, R.P. (eds.), Academic Press, San Diego 1994, pp. 207-242.

Timpl, R. and Engel, J. Type VI Collagen. In: *Structure and Function of Collagen Types.* Mayne, R. and Burgeson, R.E. (eds.), Academic Press, Orlando 1987, pp. 105-143.

Ref. for DNA/AA Sequences

Chu, M.-L. et al. Amino acid sequence of the triple-helical domain of human collagen type VI. *J. Biol. Chem.* 1988, **263:**18601-18606.

Chu, M.-L. et al. Sequence analysis of α1(VI) and α2(VI) chains of human type VI collagen reveals internal triplication of globular domains similar to the A domains of von Willebrand factor and two α2(VI) chain variants that differ in the carboxy terminus. *EMBO J.* 1989, **8:**1939-1946.

Chu, M.-L. et al. Mosaic structure of globular domains in the human type VI collagen α3 chain: similarity to von Willebrand factor, fibronectin, actin, salivary proteins and aprotinin type protease inhibitors. *EMBO J.* 1990, **9:**385-393.

Koller, E. et al. The globular domains of type VI collagen are related to the collagen-binding domains of cartilage matrix protein and von Willebrand factor. *EMBO J.* 1989, **8:**1073-1077.

Saitta, B. et al. Alternative splicing of the human α2(VI) collagen gene generates multiple mRNA transcripts which predict three protein variants with distinct carboxyl termini. *J. Biol. Chem.* 1990, **265**:6473-6480.
Zanussi, S. et al. The human type VI collagen gene; mRNA and protein variants of the α3 chain generated by alternative splicing of an additional 5-end exon. *J. Biol. Chem.* 1992, **267**:24082-24089.

Accession numbers: α1(VI) S05377; α2(VI) S05378, S09646, C35243; α3(VI) S13679, S28776.

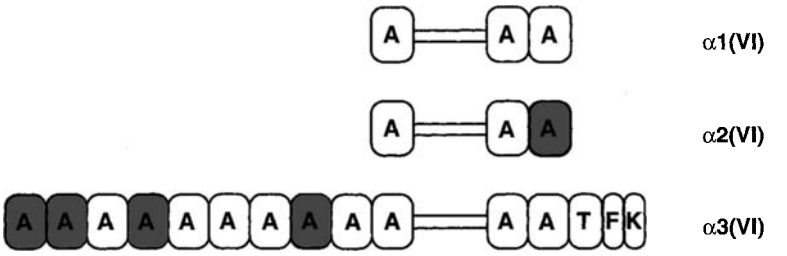


Figure 1. Domain structure of the three type VI collagen subunits. The collagenous domains are shown as thin rods, the repeats related to von Willebrand factor are indicated by A, the threonine-rich domain by T, the domain related to fibronectin by F and the domain related to the Kunitz protease inhibitors by K. Domains that are subject to alternative splicing events are shaded.

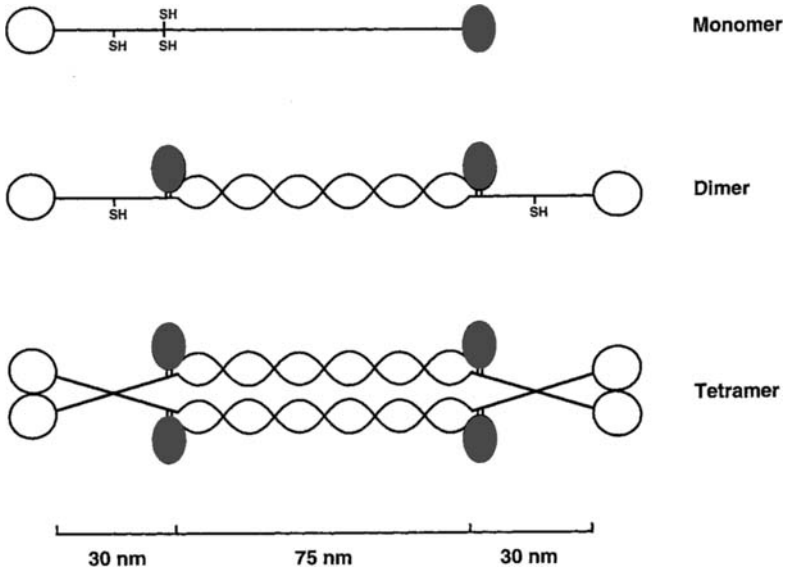


Figure 2. Model of the assembly of type VI collagen monomers into dimers and tetramers. The C-terminal globular domain is shaded.

Collagen type VII

Louise M. Rosenbaum

Synonyms	None
Abbreviations	Type VII
Classifications	Fibrous protein
Description	One of the largest collagens identified to date, type VII is presumably a homotrimer. It is the major constituent of anchoring fibrils, an ultrastructural feature of particular epithelial basement membranes.
Structure	The C-terminal non-collagenous domain (NC-2) appears as a terminal globule. The large N-terminal non-collagenous domain (NC-1) has a trident-like appearance in rotatory-shadow images, with terminal globules, and a large globule at the transition region between the NC-1 domain and the triple helical region. The higher order structure of type VII colagen is an anti-parallel dimer, in which the C-terminal portions of the triple-helix overlap by 60 nm, placing the trident-like N-terminal domains at each end of the dimer (see fig.).
Molecular Weight	1,000 kDa $\simeq$ for the homotrimer, with about 330 kDa for each α-chain. 30 kDa for NC-2, 168 kDa for NC-1, $\simeq$, based on globular standards. 180 kDa $\simeq$ for the collagenous domain, based on collagen α-chain standards.
Sedimentation Coeff.	3.9 (helical domain); 6.5 (NC-1, monomer)
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Collagen VII is the major structural component of anchoring fibrils. The NC-1 domain binds to epithelial basement membranes (BM) or to anchoring plaques, small structures made of the same constituents as BM. The network made by these interactive elements entraps collagen fibrils, thereby contributing to the adherence of the epithelial BM to the underlying stroma.
Physiology/Pathology	Type VII collagen is implicated in particular forms of epidermal bullous disease. The NC-1 domain is the antigen in the auto-immune disease, epidermolysis bullosa acquisita (EBA).
Degradation	The triple-helix is sensitive to bacterial collagenase, and houses a single pepsin-sensitive site. The NC-1 domain can be digested into smaller peptides with numerous proteases, including elastase, thermolysin, chymotrypsin, trypsin and V8 protease.

Genetics/Abnormalities	Localized to 3p21. The incomplete characterization of the gene reveals numerous small exons encoding the NC-1 domain, and characteristically small exons for the collagenous domain, but with unusually small introns, rendering a rather compact gene by collagen standards. The NC-1 domain is not detectable in the tissues of most patients with recessive dystrophic epidermolysis bullosa (RDEB), whereas several other major proteins in affected tissues are present, presumably at normal levels. Genetic linkage studies of families with dominant dystrophic epidermolysis bullosa (DDEB) history display a correlation with the COL7A1 gene.
Half-life	Unknown
Concentration	Unknown
Isolation Method	Isolation from cultures of particular epidermal cell lines with salt precipitation and DEAE chromatography, and amniotic membranes extracted with guanidine hydrochloride, followed by DEAE chromatography.
Amino Acid Sequence	Incomplete, but the current aa sequence deduced from cDNA sequences reveals several fibronectin type III repeats, and other regions with homology to the A domain of von Willebrand factor and cartilage matrix protein (CMP) in the NC-1 domain. The NC-2 domain is rich in cysteines and carries a domain similar to the Kunitz module found in α3(VI) collagen chain, among other proteins. Several NC-1 sequences were corroborated with direct protein sequences of peptides. The collagenous domain contains 22 imperfections in the Gly-X-Y pattern, and a 39-aa non-collagenous segment, presumably forming the pepsin-sensitive "hinge" region.
Disulfides/SH-Groups	Interchain disulfide bonds occur in the "transition" region, between NC-1 and the collagenous domain. There is evidence for intrachain disulfide bonds at the N-terminus of the α-chain, which may contribute to the globular appearance and, by conjecture and the localization of several cysteines to this domain by the cDNA sequence, intrachain disulfide bonds may also occur in the NC-2 domain, to give rise to its globular appearance. The antiparallel dimer is stabilized by intermolecular disulfide bonds, apparently between the C-termini.
General References	Burgeson, R. E., et al. The structure of type VII collagen. *Ann. N. Y. Acad. Sci.* 1986, **460:** 47–57. Burgeson, R. E. Type VII collagen. In: *Structure and function in collagen types*, Mayne, R. and Burgeson, R. E., eds., Academic Press, Orlando, FL, 1987, pp. 145–172. Burgeson, R. E. et al. The Structure and function of type VII collagen. *Ann. N. Y. Acad. Sci.* 1990, **580:** 32–43.
Ref. for DNA/AA Sequences	Parente, M. G. et al. Human type VII collagen: cDNA cloning and chromosomal mapping of the gene. *Proc. Natl. Acad. Sci. USA* 1991, **88:** 6931–6935. Chung-Honet, L. C. et al. Partial characterization of the gene for human type VII collagen (COL7A1). *Matrix* 1993, **13:** 8 (abst). Christiano, A. M. et al. Structure of the human COL7A1 gene and characterization of rodent α1(VII) cDNAs. *Matrix* 1993, **13:** 4 (abst).

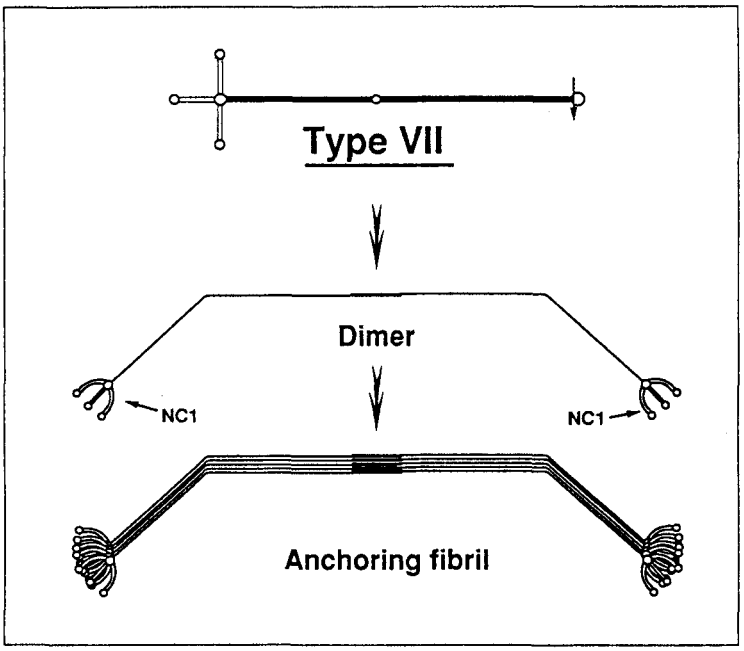


Figure 1. Molecular structure and supramolecular assembly of collagen VII. Type VII collagen molecules have a N-terminal domain (NC1) with three fingers. After processing of the C-terminal globular domain (NC2), they form dimers by antiparallel, partial overlapping of the carboxyl end of their triple helical region. Several dimers aggregate laterally in register to form anchoring fibrils ended by a large bunch of N-terminal domains. These N-terminal domains interact with epithelial basement membranes.

Collagen type IX

Peter Bruckner

Synonyms	Proteoglycan-LT, type M-collagen, p-HMW-collagen
Abbreviations	α1(IX)α2(IX)α3(IX)
Classifications	Fibrous protein
Description	Collagen IX is expressed mostly in cartilage and in vitreous humor as a less abundant collagen. Together with collagens II and XI, it assembles into striated fibrils that confer tensile strength to the tissue. During its biosynthesis, collagen IX undergoes extensive and specific posttranslational modifications. The most prominent, but not obligatory modification is the attachment of a chondroitin sulfate chain of variable length to a non-helical portion of the protein. Other posttranslational modifications include hydroxylation of prolyl residues in the Yaa position of the triplets Gly-Xaa-Yaa, hydroxylation of specific lysyl residues, glycosylation of hydroxylylsyl residues, and oxidation of the ε-NH_2 of specific lysyl or hydroxylysyl residues into aldehydes. These aldehydes react with other amino-acyl residues to form covalent crosslinks that stabilize the fibrils. These crosslinks occur either between two molecules of collagen IX (homotypic) or between collagen II and collagen IX molecules (heterotypic). Crosslinks between collagen type IX and collagen XI have not been described.
Structure	Collagen IX is a heterotrimeric molecule consisting of three distinct polypeptide chains, α1(IX), α2(IX) and α3(IX) in a 1 : 1 : 1 a stoichiometry. The molecule is 190 nm in length and 1.5 nm in diameter. In contrast to other collagens, collagen IX is synthesized in its final form and not as a precursor molecule which is proteolytically processed in the extracellular space. The protein contains three triple helical domains (COL1, COL2, and COL3) and non-helical domains at the end of the molecules and between the triple helical domains (NC1, NC2, NC3, and NC4). The NC4 domain of the α1-chain is much larger that that of the other chains and is missing in some forms of the protein due to the utilization of an alternative promoter. The chondroitin sulfate chain is attached to the NC3-domain of the α2(IX)-chain.
Molecular Weight	240,000 ≃ (without chondroitin sulfate). The α1-chain with the large NC4-domain has an Mr 94,000, all other chains an Mr 69,000 upon SDS-PAGE.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Collagen IX has only a few tryptophan and tyrosine residues. It is best detected by absorbance of the peptide bond at 230 nm or below. No extinction coefficient has been reported.
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None

Biological Functions

Collagen IX is incorporated into weakly banded fibrils with uniform diameter of 20 nm responsible for tissue coherence. The protein is a major element of the diameter-control. At least the domains COL3 and NC4 at the N-terminus of the protein, as well as the chondroitin sulfate chains are located at the surface of the fibrils where they may mediate contacts with other fibrils or the extrafibrillar matrix.

Physiology/Pathology

Collagen IX expression is coordinately regulated with that of collagens II and XI. The protein may be one of the first targets of degradation in degenerative cartilage diseases resulting in a loss of cohesiveness of the fibrilar network in the tissues.

Degradation

Collagen IX is not cleaved by MMP-1 (collagenase) but is susceptible to degradation by gelatinase. The protein contains two major non-helical domains between adjacent helical regions which may be the target of many tissue proteinases with wide specificities.

Genetics/Abnormalities

Collagen IX is a good candidate for mutations to be found in several forms of cartilage diseases, including osteoarthritis or chondrodysplasias. No such mutation has yet been identified.

Half-life

Unknown

Concentration

The protein represents about 10% of the collagens in cartilage fibrils.

Isolation Method

Very small quantities of native human collagen IX can be extracted with neutral buffers contianing 1 M NaCl from fetal or infant, but not from adult cartilage. Better yields are obtained from cultures of chondrocytes in agarose gels kept in the presence of the lysyl oxidase inhibitor 2-aminopropionitrile fumarate. The protein can then be strongly enriched by ion exchange chromatography on DEAE-cellulose. Larger quantities of proteolytic, triple helical fragments HMW and LMW can be obtained if the tissue is digested with pepsin. HMW and LMW can be purified by salt precipitation in 0.5 M acetic acid, where they precipitate in 2 M NaCl after separation of collagen II at 0.9 M NaCl and collagen XI at 1.2 M NaCl.

Amino Acid Sequence

The sequence of the human protein is only partially known and highly homologous to the chicken protein. There the numbers of aa in the domains starting from the N-terminus are as follows:

Residues in the	α1 (IX) upstream promoter	α1 (IX) downstream	α2 (IX)	α3 (IX)
Signal peptides	23	23	21	21
NC4	243	2	3	3
COL3	137	137	137	137
NC3	12	12	17	15
COL2	339	339	339	339
NC2	30	30	30	31
COL1	115	115	115	112
NC1	21	21	15	17
Total	930	679	677	675

Disulfides/SH-Groups

Pairs of disulfide bonds occur at the junctions of NC3 and COL2 and of COL1 and NC1 which are used for interchain disulfide bonds.

General References

Bruckner, P: et al. The structure of human collagen type IX and its organization in fetal and infant cartialge. *J. Biol. Chem.* 1988, **261:** 16911–16917.

van der Rest, M. and Mayne, R. Type IX collagen. In: *Structure and Function of Collagen Types*, Mayne, R. and Buregeson, R. E. (eds.) Academic Press, New York, pp. 195–221.

Shimokomaki, M. et al. Identification of a new disulfide-bonded collagen from cartilage. *FEBS Lett.* 1980, **121:** 51–54.

Noro, A. et al. Isolation and characterization of a third proteoglycan (Pg-Lt) from chick embryo cartilage which contains disulfide-bonded collagnous polypeptide. *J. Biol. Chem.* 1983, **258:** 9323–9331.

Vaughan, L. et al. Proteoglycan Lt from chicken embryo sternum identified as type IX collagen. *J. Biol. Chem.* 1985, **260:** 4758–4763.

Eikenberry, E. F. et al. Fibrillar organization in cartilage. In: *Articular Cartilage and Osteoarthritis*, Kuettner, K. E. et al. (eds.) Raven Press, New York, pp. 133–149.

Vaughan, L. et al. D-periodic distribution of collagen IX along cartilage fibrils. *J. Cell Biol.* 1988, **106:** 991–997.

Ref. for DNA/AA Sequences

Brewton, R. G. et al. Cloning fo the chicken α3(IX) collagen chain completes the primary structure of type IX collagen. *Eur. J. Biochem.* 1992, **205:** 443–449.

Collagen type X

J. Terrig Thomas and Raymond P. Boot-Handford

Synonyms	None
Abbreviations	Type X; $[\alpha 1(X)]_3$
Classifications	Matrix protein
Description	A short chain, developmentally regulated, homotrimeric collagen ($\alpha 1(X)$ Mr = 59,000) transiently expressed by hypertrophic chondrocytes at sites of endochondral bone formation. It consists of a small N-terminal domain, triple-helical domain of 132 nm (approx. half the length of that of fibrillar collagens), and a large C-terminal globular domain. It has a relatively high hydroxylysine content corresponding to a carbohydrate content of approx. 5%.
Structure	Studies in vitro have shown that type X collagen forms a hexagonal lattice-like structure similar to that observed in vivo for the genetically-related type VIII collagen. The aggregation appears to be initiated by interactions between C-terminal globular domains followed by interactions of the juxtaposed triple-helical domains.
Molecular Weight	180 kDa approx.; 59 kDa for individual α-chains (SDS-PAGE).
Sedimentation Coeff.	Unknown
Isoelectric Point	10.52 (predicted from aa sequence)
Extinction Coeff.	38,320 $cm^{-1}M^{-1}$ (predicted value).
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	During development, longitudinal bone growth takes place via endochondral ossification, a process typified by the replacement of the predominantly type II collagen-rich cartilage matrix by the predominantly type I collagen-rich matrix of bone. Type X collagen is the major structural protein secreted by hypertrophic chondrocytes at the cartilage/bone transition. Although no biological function has yet been assigned to type X collagen, its specific and transient expression at sites of new bone formation suggest it may play an important role in this process.
Physiology/Pathology	Mutations in the C-terminal non-collagenous domain of type X collagen cause metaphyseal chondrodysplasia type Schmid. No mutations in the collagenous or N-terminal non-collagenous domains have been identified in any human chondrodysplasia to date. However, transgenic mice expressing the chick collagen X gene containing a deletion in the collagenous domain exhibit a phenotype that resembles spondylometaphyseal dysplasia in humans.

Degradation	Has two mammalian collagenase cleavage sites within the triple-helical domain resulting in a Mr 32,000 collagenase-resistant fragment which is stable at 37°C.
Genetics/Abnormalities	COL10A1 localized to chromosome 6q21-22.3. A number of polymorphisms have been found within the coding region which are not linked to any disease phenotype.
Half-life	Unknown
Concentration	Unknown
Isolation Method	Isolation of intact form from medium of culture hypertrophic chondrocytes by precipitation with 30% ammonium sulfate. Can be separated from other collagens by its solubility in acid salt solutions (soluble in 0.5 M acetic acid containing 1.2 M NaCl) or by gel filtration on Sephacryl S-500. Isolation of helical domain from cartilage by proteolytic digestion.
Amino Acid Sequence	Consists of three identical α chains each comprising of a signal peptide (18 aa), NH_2-non-collagenous domain (38 aa), triple-helical domain (463 aa), and COOH-non-collagenous domain (161 aa). Triple-helix contains three imperfections of the type Gly-Xaa-Yaa-Xaa-Yaa-Gly of which two are sites recognized by mammalian collagenase and five of the type Gly-Xaa-Gly.
Disulfides/SH-Groups	Bovine type X collagen has disulfide bonds within the triple-helical domain whereas human, mouse, rabbit, rat and chicken do not. In all species studied there is one cysteine residue within the COOH-non-collagenous domain of each α1(X) chain, which may be involved in dimer formation or intermolecular disulfide bonding.
General References	Schmid, T.M. and Linsenmayer, T.F. Type X collagen. In: *Structure and Function of Collagen Types,* Mayne, R. and Burgeson, R.E. (eds.) Academic Press 1987, pp. 223-259. Hoyland, J.A. et al. Distribution of type X collagen mRNA in normal and osteoarthritic cartilage. *Bone and Mineral* 1991, **15**:151-164. Kwan, A.P.I. et al. Macromolecular organization of chicken type X collagen in vitro. *J. Cell Biol.* 1991, **114**:597-604. Jacenko, O., LuValle, P.A., Olsen, B.R. Spondylometaphyseal dysplasia in mice carrying a dominant negative mutation in a matrix protein specific for cartilage-to-bone transition. *Nature* 1993, **365**:56-61. Warman, M.L. et al. A mutation in the human type X collagen gene in a family with Schmid metaphyseal chondrodysplasia. *Nature Genet.* 1993, **5**:79-82. Wallsi, G.A. et al. Mutations within the gene encoding the a1 (X) chain of type X collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but not several other forms of metaphyseal chondrodysplasia. *J. Med. Genet.* 1996, **33**:450-457.
Ref. for DNA Sequences	Apic, S.S. et al. Cloning of human α1(X)a collagen DNA and localization of the COL10A1 gene to the q21-q22 region of human chromosome 6. *FEBS Lett.* 1991, **282**:393-396. Reichenberger, E. et al. Genomic organization and full-length cDNA sequence of human collagen X. *FEBS Lett.* 1992, **311**:305-310. Thomas, J.T. et al. The human collagen X gene: complete primary translated sequence and chromosomal localization. *Biochem J.* 1991, **280**: 617-623.

Collagen type XI

Mahnaz Moradi-Améli and Michel van der Rest

Synonyms	Collagen XI, $\alpha1\alpha2\alpha3$
Abbreviations	None
Classifications	Fibrous protein
Description	Collagen XI is a fibril forming, or fibrillar, collagen which is distributed in minor amounts in the extracellular matrix of cartilaginous tissues and of the vitreous. It is coexpressed with collagen II, the major collagen constituent in these tissues. As other fibrillar collagens, it is composed of a non interrupted triple helix with Gly-Xaa-Yaa sequence where the Yaa is occupied by a posttranslationaly hydroxylated proline. Other posttranslational modification like hydroxylation of specific lysyl residues, glycosylation of hydroxylysyl residues and oxidation of the ε-NH_2 of specific lysyl or hydroxylysyl residues into aldehydes are known to occur. These aldehydes further react in a complex way with other amino-acyl residues to form covalent crosslinks that stabilize the supramolecular assembly. Collagen XI is capable, at least *in vitro*, of forming thin homotypic fibrils but has been observed *in vivo* in combination with fibrillar collagen II and non-fibrillar collagen IX to form heterotypic fibrils. Type XI collagen is not exposed on the surface of such heterotypic fibrils since an immunostaining can only be realized after partial disruption of the fibrils.
Structure	Collagen XI is a heterotrimeric molecule consisting of three distinct polypeptide chains, $\alpha1$(XI), $\alpha2$(XI), and $\alpha3$(XI) which can be assembled in $\alpha1$(XI)$\alpha2$(XI)$\alpha3$(XI) stoichiometry. A heterotypic form of collagens XI and V in $[\alpha1(XI)]_{2\alpha}2(V)$ form also exists in some human cell line and probably *in vivo*. The $\alpha3$(XI) chain appears to be the same gene product as $\alpha1$ chain of collagen II but both chains differ posttranslationally in their degree of lysine hydroxylation and glycosylation. Although $\alpha1$(XI) and $\alpha2$(XI) chains present very high sequence similarity with the $\alpha1$(V) collagen chain, they are each distinct gene products. Collagen XI chains are synthesized by chondrocytes and secreted as procollagen in extracellular space. Unlike the major collagens which undergo full proteolysis of their non collagenic extension, collagen XI seems to be only partially processed when extracted from tissues (intact form).
Molecular Weight	$\approx$480 kDa: estimated by gel electrophoresis for the intact form of the processed molecule.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	It is best detected by absorbance of the peptide bond at 230 nm or below. No extinction coefficient has been reported.
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None

Biological Functions

Collagen XI combines with collagen II and IX to form heterotypic fibrils and in this way it may play a role in fibrillogenesis by regulating the growth of type II collagen fibrils in cartilage. Collagen XI binds heparin and seems to bind cartilage proteoglycans.

Physiology/Pathology

Collagen XI seems to be involved in Stickler syndrome.

Degradation

The neutral metalloproteinase, 92 kDa gelatinase, (MMP-9) specifically degrades collagen XI. It is otherwise resistant to interstitial collagenase (MMP-1). Clostridiopeptidase A (EC 3.4.24.3) can cleave the native triple helical domain. The denatured molecule can be cleaved by general proteolysis. Crosslinked molecules can be solubilized by enzymes such as pepsin that cleaves the telopeptides and the non-helical extension remaining on the intact form of the molecules.

Genetics/Abnormalities

The genes encoding the pro α1(XI) chain (COL11A1) and the pro α2(XI) chain (COL11A2) are located on chromosome 1 at p21 and on chromosome 6 at p21.2 respectively. The gene encoding the pro α3(XI) is the COL2A1 localized on chromosome 12 at q13–14. The possible involvement of type XI collagen genes in human disease has not yet been demonstrated.

Half-life

Unknown

Concentration

Type XI collagen was identified in minor amount in hyaline cartilages (epiphyseal growth, articular and costosternal) and in fibrocartilage (intervertebral disc). Its concentration does not exceed 3% of collagen II.

Isolation Method

Very little amount of native type XI collagen can be extracted from fetal and neonatal cartilage by 4M guanidine HCl extraction. Larger quantities can be obtained if the tissue is digested by pepsin followed by differential salt precipitation between 0.9M and 1.2M NaCl in 0.5M acetic acid. However, the pepsin-digest material is devoid of the non-helical extension and the telopeptides.

Amino Acid Sequence

	Residues in α1(XI) chain	Residues in α2(XI) chain	Residues in α3(XI) chain
Signal peptide	36	–	25
N-propeptide			
(non-triple helical)	383	(263)	7
(short triple-helix)	89	90	79
N-telopeptide	20	20	24
Main triple helix	1014	1014	1014
C-telopeptide	21	24	27
C-propeptide	243	(166)	246

Disulfides/SH-Groups

4 cysteines in the non-triple helical N-propeptide region and 7 cysteines in the C-propeptide of the α1(XI) chain. 4 cysteines in the non-triple helical N-propeptide of α2(XI) chain and 7 cysteines in the uncompleted sequence of its C-propeptide. The N-propeptide region of the α3(XI) chain corresponds to the form IIB of the α1(II) chain and is devoid of cysteines, its C-propeptide contains 8 cysteines. Interchain disulfide bonds may only be formed in the C-propeptide.

General References

Eyre, D. R. and Wu, J. J. Type XI or 1a 2a 3a collagen. **In** *Structure and function of collagen types*, Mayne R. and Burgeson R. E. (eds) 1987, Academic Press, pp. 261–281.

van der Rest, M. and Garrone, R. Collagen family of protein *FASEB J.* 1991, **5**:2814–2823.
Eikenberry, M. et al., Fibrillar organization in cartilage. **In** *Articular cartilage and osteoarthritis*, Kuettner K. et al. (eds) 1992, Raven Press, pp. 133–149.
van der Rest, M. et al. Collagen: a family of proteins with many facets. **In** *Advances in molecular and cell biology*, Bittar, E. E. and Kleinman, H. K. (eds) 1993, Jai Press Inc. Vol. 6, pp. 1–68.

Ref. DNA/AA Sequences

For the proα1(XI) chain:
Bernard, M. et al. *J. Biol. Chem.* 1988, **263**: 17159–17166.
Yoshioka, H. and Ramirez, F. *J. Biol. Chem.* 1990, **265**: 6423–6426

For the proα2(XI) chain:
Kimura, T. et al. *J. Biol. Chem.* 1989, **264**: 13910–13916.
Neame, P. J. et al. *J. Biol. Chem.* 1990, **265**: 20401–20408.
Zhidkova, N. I. et al. *FEBS Lett.* 1993, **326**: 25–28.

For the proα3(XI) chain:
Su, M. W. et al. *Nucleic Acid Res.* 1989, **17**: 9473.
Baldin, C. T. et al. *Biochem. J.* 1989, **262**: 521–528.
Ryan, M. C. et al. *Genomics* 1990, **8**: 41–48.

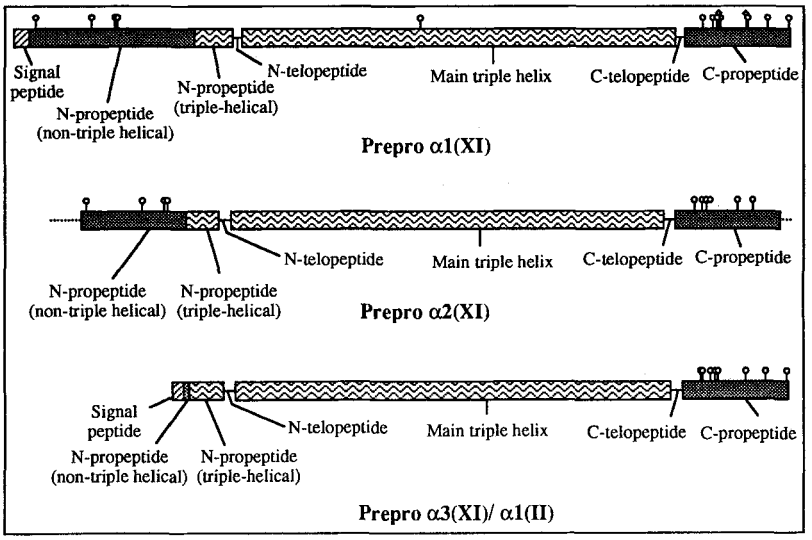


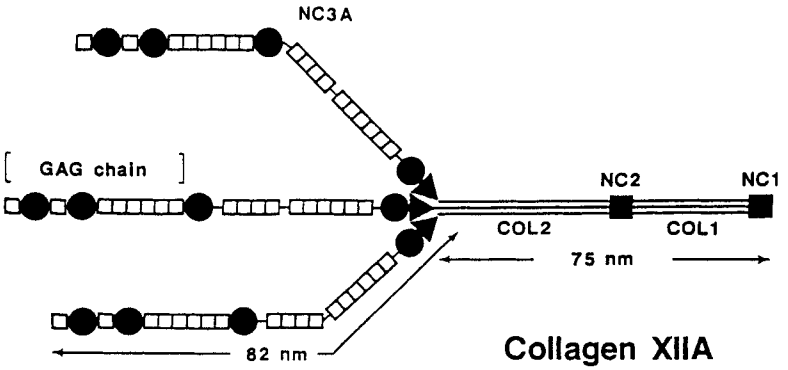
Figure 1. Schematic representation of the structure of prepro-type XI collagen chains. The positions of cysteine residues (Ŷ) and of demonstrated or putative N-glycosylation sites (Ŷ) are indicated. Regions that have not been studied are shown as dotted line.

Collagen type XII

Gregory P. Lunstrum

Synonyms	None
Abbreviations	Type XII, XIIA, XIIB
Classifications	Matrix protein, extracellular
Description	A high molecular weight glycoprotein present in many different connective tissues. It is a homotrimeric assembly of α1A chains comprised of two relatively short collagen domains. COL1 (103 aa) and COL2 (152 aa), bonded by three non-collagen domains, the C-terminal NC1 (76 aa), NC2 (43 aa) and the large N-terminal NC3A (2750 aa). Two alternatively expressed forms have been identified as XIIA and XIIB. The smaller α1B chain is missing the N-terminal one third (1189 aa) of the NC3A domain. It is not known whether α1A and α1B participate in a heterotrimeric assembly. Type XIIA may also be a proteoglycan, modified in the distal portion of NC3A.
Structure	A four armed structure of a short collagen helix (75 nm) and three extended globular NC3 domains. These globular arms measure 82 nm in length for XIIA and 55 nm for XIIB. Distinctive flexible regions in the structure appear at the NC2 domain and within NC3 at both the central von Willebrand factor A domain and a hinge region separating the 12th and 13th fibronectin type III repeats.
Molecular Weight	1,000 kDa ($\simeq$ 330 kDa/α chain) for type XIIA 700 kDa ($\simeq$ 225 kDa/α chain) for type XIIB.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Structural protein proposed to interact with the surface of interstitial collagen fibers through COL and von Willebrand factor A domains. While an RGD adhesion sequence is present in the COL2 domain, type XII collagen is a relatively poor substrate for cell attachment when compared to other matrix components.
Physiology/Pathology	Although collagens XII and XIV generally have similar distributions, some tissues, including skin, blood vessels, and articular cartilage, exhibit unique and often complementary localization of these two components.
Degradation	Partial degradation by bacterial collagenase. Some chymotrypsin sites on native type XIIA have been mapped, however native material is relatively protease insensitive.

Genetics/Abnormalities	Although two variants have been described it is unknown whether they are the result of alternate gene expression or mRNA splicing.
Half-life	Unknown
Concentration	Unknown
Isolation Method	Native XIIA is prepared from cell culture and partially purified from tissue using lectin affinity and standard chromatography techniques. Pepsin resistant collagen domains and the collagenase resistant NC3B have been purified from fetal bovine and chick tissue extracts.
Amino Acid Sequence	Types XII and IX share homologies between COL1 domains (50% identity) and a portion of NC3(XII) and NC4(IX) are 35% identical. Comparison of XII and XIV from chick show 58% identity overall. Three interruptions in the Gly-x-y triplet sequences of α1(IX), α1(XII) and α1(XIV) are located in homologous positions. The NC3A domain is composed of repeated motifs, 4 von Willebrand factor A motifs and 18 fibroncetin type III motifs, and a domain homologous to NC4 of type IX collagen. NC3B is lacking 2 factor A motifs and 8 type III motifs from the N-terminus. Several predicted sites for both N- and O-glycosylation and some twenty Ser-Gly seqeunces which could serve as sites for the addition of glycosaminoglycan side chains have been identified.
Disulfides/SH-Groups	Several interchain (within collagen domains and at the C-terminal end of NC3) and intrachain disulfide bonds are predicted.
General References	Dublet, B., et al. The structure of avian type XII collagen. *J. Biol. Chem.* 1989, **264:** 13150–13156. Shaw, L. M. and Olsen, B. R. FACIT collagens: diverse molecular bridges in extracellular matrices. *Trends Biochem. Sci.* 1991, **16:** 191–194. Lunstrum, G. P. et al. Identification and partial characterization of two type XII-like collagen molecules. *J. Cell Biol.* 1991, **113:** 963–969. Keene, D. R. et al. Two type XII-like collagens localize to the surface of banded collagen fibrils. *J. Cell Biol.* 1991, **113:** 971–978. Lunstrum, G. P. et al. Identification and partial purification of a large, variant form of type XII collagen. *J. Biol. Chem.* 1992, **267:** 20087–20092. Watt, S. L. et al. Characterization of collagen Types XII and XIV from fetal bovine cartilage. *J. Biol. Chem.* 1992, **267:** 20093–20099. Koch, M. et al. A major oligomeric fibroblast proteoglycan identified as a novel large form of type XII collagen. *Eur. J. Biochem.* 1992, **207:** 847–856.
Ref. for DNA/AA Sequences	Dublet, B. and van der Rest, M. Type XII collagen is expressed in embryonic chick tendons: isolation of pepsin derive fragments. *J. Biol. Chem.* 1987, **262:** 17724–17727. Gordon, M. K. et al. Type XII collagen: A large multidomain molecule with partial homology to type IX collagen. *J. Biol. Chem.* 1989, **264:** 19772–19778. Yamagata, M. et al. The complete primary structure of type XII collagen shows a chimeric molecule with reiterrated fibronectin type III motifs, von Willebrand factor A motifs, a domain homologous to a noncollagenous region of type IX collagen, and short collagenous domains with a Arg-Gly-Asp site. *J. Cell Biol.* 1991, **115:** 209–221.

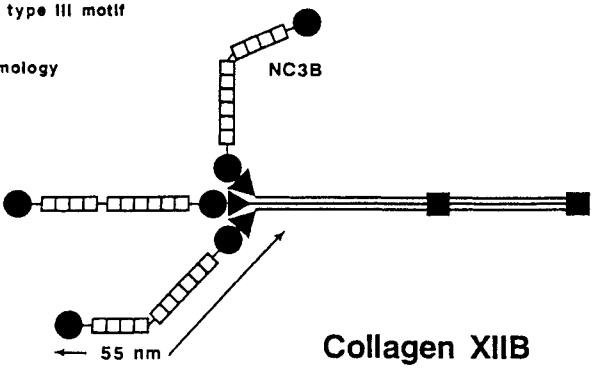


Collagen XIIA

●-von Willebrand Factor A motif

□-Fibronectin type III motif

▶-NC4(IX) Homology



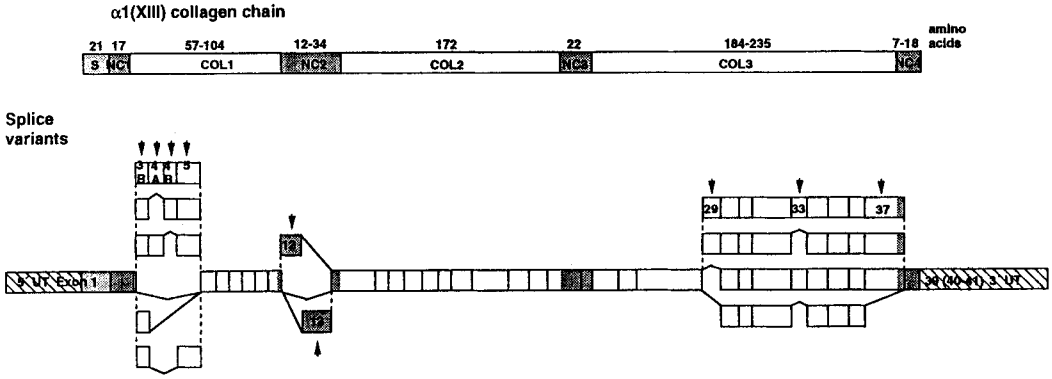
Collagen XIIB

Collagen type XIII

Taina Pihlajaniemi

Synonyms	COL13A1
Abbreviations	None
Classifications	Matrix protein, extracellular
Description	A short-chain, non-fibrillar collagen that is expressed in bone, cartilage, intestine, placenta, skin and striated muscle. Type XIII collagen has been characterized via cDNA and genomic clones. The predicted polypeptide consists of three collagenous domains, COL1–3, and four non-collagenous domains, NC1–4. Type XIII collagen RNA transcripts undergo complex alternative splicing involving both triple-helical and non-collagenous segments of the resulting polypeptides. Sequences corresponding to nine exons of the gene are subject to alternative splicing, namely exons 3B, 4A, 4B and 5 encoding part of COL1, exons 12 and 13 encoding most of NC2, exons 29 and 33 encoding part of COL3 and exon 37 encoding the COL3/NC4 junction. At least 12 mRNA species exist through the alternations of exons 3B-12 and distinct differences in the proportions of the variant mRNA have been found in RNAs from various sources. Characterization of cDNA clones shows that four combinations of exon 29, 33 and 37 exist but additional ones are likely to be found.
Structure	Type XIII collagen is probably synthesized as a homotrimer consisting of three α1(XIII) chains that possibly represent more than one splice variant. The protein has not yet been isolated.
Molecular Weight	60,123 (calculated for the longest splice variant) and 48,150 (calculated for the shortest splice variant).
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown, but predicted high based on cDNA
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Unknown
Physiology/Pathology	Unknown
Degradation	Unknown
Genetics/Abnormalities	The gene is located on human chromosome 10, band q22. The gene is over 130 kb in size and consists of 40–41 exons. The exons coding for translated sequences vary in size between 24 (possibly 8) and 153 bp. Eight of the 29 exons solely encoding reapeating Gly-X-Y sequences are 27 bp and four are 54 bp. All but one exon begin with a complete codon for an amino acid. Each of the nine alternatively spliced sequences corresponds precisely to one exon of the gene.

Half-life	Unknown
Concentration	Unknown
Isolation Method	The protein has not yet been isolated.
Amino Acid Sequence	The polypeptide consists of 623 to 495 aa residues depending on splice variations. The predicted length of the COL1 domain may vary between 57 and 104 residues, the NC2 domain is either 12, 31 or 34 residues, the COL3 domain may vary between 184 and 235 residues and the NC4 is either 7 or 18 residues in length.
Disulfides/SH-Groups	Two cystein residues in the NC1, one in the COL1, one in an alternate NC2 variant, one in the invariant portion of NC4, and one in the NC4 region affected by alternative splicing. Occurence of disulfide bonds not known.
General References	van der Rest, M. and Garrone, M. Collagen family of proteins. *FASEB J.* 1991, **5:** 2814–2823. Sandberg, M. et al. Expression of mRNAs coding for the α1 chain of type XIII collagen in human fetal tissues: comparison with expression of mRNAs for collagen types I, II and III. *J. Cell Biol.* 1989, **109:** 1371–1379. Juvonen, M. and Pihlajaniemi, T. Characterization of the spectrum of alternative splicing of α1(XIII) collagen transcripts in HT-1080 cells and calvarial tissue resulted in identification of two previously unidentified alternatively spliced sequences, one previously unidentified exon, and nine new mRNA variants. *J. Biol. Chem.* 1992, **267:** 24693–24699. Juvonen, M. et al. Patterns of expression of the six alternatively spliced exons affecting the structures of COL1 and NC2 domains of the α1(XIII) collagen chain in human tissue and cell lines. *J. Biol. Chem.* 1992, **267:** 24700–24707.
Ref. for DNA/AA Sequences	Pihlajaniemi, T. et al. Partial characterization of a low molecular weight human collagen that undergoes alternative splicing. *Proc. Natl. Acad. Sci. USA* 1987, **84:** 940–944. Pihlajaniemi, T. and Tamminen, J. The α1 chain of type XIII collagen consists of three collagenous and four noncollagenous domains, and its primary transcript undergoes complex alternative splicing. *J. Biol. Chem.* 1990, **265:** 16922–16928. Tikka, L. et al. Gene structure for the α1 chain of a human short-chain collagen (type XIII) with alternatively spliced transcripts and translation termination codon at the 5'end of the last exon. *Proc. Natl. Acad. Sci. USA* 1988, **85:** 7491–7495. Tikka, L. et al. Human α1(XIII) collagen gene. Multiple forms of the gene transcripts are generated through complex alternative splicing of several short exons. *J. Biol. Chem.* 1991, **266:** 17713–17719. The nucleotide sequences are available from GenBank/EMBL Data Bank with accession numbers J04085, J05580, and M68974-M69010.



Top, schematic diagram of the cDNA-deduced α1 (XIII) polypeptide structure. COL1–3 indicate collagenous domains, NC1–4 noncollagenous domains, and S the putative signal peptide. Bottom, the corresponding mRNA with exons encoding the polypeptide shown as boxes. Arrows indicate the alternatively spliced exons, and the numbers of these exons are given. A recently defined exon is named as 4B, and the previously characterized exon 3 is believed to consist of two exons with the 3′ one being marked here as 3B. Thus although the last exon is numbered as 39 the gene consists of 40 or 41 exons.

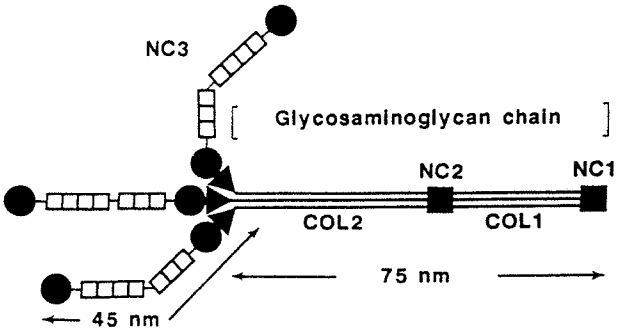
Collagen type XIV

Gregory P. Lunstrum

Synonyms	Undulin
Abbreviations	None
Classifications	Extracellular matrix protein , FACIT family
Description	A high molecular weight glycoprotein in many different connective tissues. It is a homotrimeric assembly of α chains comprised of two relatively short collagen domains, COL1 (106 aa) and COL2 ($\simeq$ 150 aa), bounded by three noncollagen domains, C-terminal NC1 ($\simeq$ 75 aa), NC2 (43 aa) and the large N-terminal NC3 ($\simeq$ 190 kDa). An alternately expressed form of human collagen XIV, undulin, is truncated within NC3 and does not contain a portion of this domain and the C-terminal collagen and noncollagen domains. Complete N-terminal sequences for both molecules remain to be determined. A proteoglycan form of collagen XIV, modified within the collagenase sensitive domain, has been identified.
Structure	A four armed structure comprised of a short collagen helix (75 nm) and three extended globular NC3 domains 45 nm in length. Flexible regions appear at NC2 and within NC3, between modules of fibronectin type III repeats. While similar 45 nm arms are present on images of undulin, the nature and assembly of the observed 75 nm noncollagen tail is unclear.
Molecular Weight	700 kDa ($\simeq$ 225 kDa/α chain).
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Structural protein proposed to interact with the surface of interstitial collagen fibers through COL and von Willebrand factor A domains. Affinities for heparin sulfate proteoglycan and type VI collagen have been demonstrated. While an RGD adhesion sequence is present in the COL2 domain, type XIV collagen is a relatively poor substrate for cell attachment.
Physiology/Pathology	Although collagens XII and XIV generally have similar distributions, some tissues, including skin, blood vessels and articular cartilage, exhibit unique and often complementary localization of these two components.
Degradation	Partial degradation by collagenase, however native material is relatively protease insensitive.

Genetics/Abnormalities	Although variants have been described it is unknown whether they are the result of alternate gene expression or mRNA processing.
Half-life	Unknown
Concentration	Unknown
Isolation Method	Native type XIV has been prepared from chick and bovine tissues and human placenta using lectin affinity and standard chromatography techniques. Pepsin resistant collagen domains and the collagenase resistent NC3 domain have been purified from chick and bovine tissue extracts.
Amino Acid Sequence	As with type XII, type XIV shares homologies with type IX in the COL1 domains and between a portion of NC3(XIV) and NC4(IX). Comparison of types XII and XIV from chick show 58% identity overall. Three interruptions in the Gly-X-Y triplet sequences of α1(XIV), α1(XII) and α1(IX) are located in homologous positions. The incompletely sequenced NC3 domain is composed of repeated modules, 2 von Willebrand factor A motifs and 7 fibronectin type III motifs, and a region homologous to NC4 of type IX collagen. Some potential glycosylation sites have been identified. The variant, undulin, is truncated within NC3 and lacks the C-terminal von Willebrand A, NC4(IX), COL2, NC2, COL1, and NC1 domains.
Disulfides/SH-Groups	Several interchain (within collagen domains and the C-terminal of NC3) and intrachain disulfide bonds are predicted. In undulin, where these domains are absent, Cys residues unique to the C-terminal fibronectin type III repeat may participate in interchain assembly.
General References	Dublet, B. and van der Rest, M. Type XIV collagen, a new homotrimeric molecule extracted from fetal bovine skin and tendon, with a triple helical disulfide bonded domain homologous to type IX and type XII collagen. *J. Biol. Chem.* 1991, **266:** 6853–6858. Shaw, L. M. and Olsen, B. R. FACIT collagens; diverse molecular bridges in extracellular matrices. *Trends Biochem. Sci.* 1991, **16:** 191–194. Keene, D. R. et al. Two type XII-like collagens localize to the surface of banded collagen fibers. *J. Cell Biol.* 1991, **113:** 971–978. Aubert-Foucher, E. et al. Purification and characterization of native type XIV collagen. *J. Biol. Chem.* 1992, **267:** 15759–15764 Watt, S. L. et al. Characterization of collagen types XII and XIV from fetal bovine cartilage. *J. Biol. Chem.* 1992, **267:** 20093–20099. Brown, J. C. et al. Structure and binding properties of collagen type XIV isolated from human placenta. *J. Cell Biol.* 1993, **120:** 557–567.
Ref. for DNA/AA Sequences	Gordon, M. K. et al. Cloning of a cDNA for a new member of the class of fibril-associated collagens with interrupted triple helices. *Eur. J. Biochem.* 1991, **201:** 333–338. Just, M. et al. Undulin is a novel member of the fibronectin-tenascin family of extracellular matrix glycoproteins. *J. Biol. Chem.* 1991, **266:** 1726–1732. Trueb, J. and Trueb, B. Type XIV collagen is a variant of undulin. *Eur. J. Biochem.* 1992, **207:** 549–557.

Collagen XIV



●-von Willebrand Factor A motif

□-Fibronectin type III motif

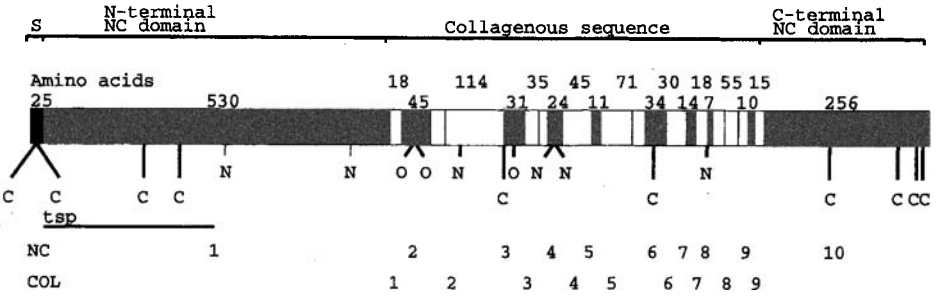
▶-NC4(IX) Homology

Collagen type XV

Taina Pihlajaniemi

Synonyms	COL15A1
Abbreviations	None
Classification	Matrix protein, extracellular
Description	A collagen with extensive interruptions in the collagenous sequence that is expressed in fibroblasts in a wide range of tissues. The complete primary structure of human type XV collagen has been characterized via cDNA clones and the exon-intron structure of the 3′ portion of the human gene has been elucidated. The polypeptide contains a central mainly collagenous portion, and NH_2 (NC1) and COOH-terminal (NC10) noncollagenous domains. Due to the presence of a classical signal peptide sequence the molecule is predicted to be secreted into the extracellular matrix. The NC1 domain contains an approximately 200-residue sequence that is homologous with a large NH_2-terminal segment of thrombospondin-1. The polypeptide contains several potential sites for serine-linked glycosaminoglycans and asparagine-linked oligosaccharides. Type XV collagen is homologous with the newly-described type XVIII collagen and the two collagens form a new subgroup in the collagen family. The most striking homology between collagens XV and XVIII is found in the COOH-terminal noncollagenous domains, which are uniquely characterized by four conserved cysteine residues. The two collagens share also homology in their collagenous sequences the structural differences being nevertheless sufficient to preclude their existence in the same molecule.
Structure	Type XV collagen may be synthesized as a homotrimer consisting of three α1(XV) chains since extensive analyses of cDNA clones have not indicated existence of additional type XV chains. The protein has not yet been isolated.
Molecular Weight	141,000 (calculated based on cDNA-derived aa sequence). Use of antipeptide antibodies results in detection of a 125,000 Da polypeptide by Western blotting of HeLa cell homogenates.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Unknown
Physiology/Pathology	Unknown
Degradation	Unknown

Genetics/Abnormalities	The gene is located in human chromosome 9, region q21 → q22. The exon-intron organization of the seven 3′ exons corresponding to the extreme COOH-terminal 2/3 + 275 residues has been determined. The NC10 domain is encoded by these seven exons. Five of these begin with a complete codon for an aa.
Half-life	Unknown
Concentration	Unknown
Isolation Method	The protein has not yet been isolated.
Amino Acid Sequence	The human type XV collagen chain consists of 1388 aa residues with the following domains: a 25-residue putative signal peptide; a 530-residue NH_2-terminal noncollagenous domain (NC1); a 577-residue collagenous sequence; and a 256-residue COOH-terminal noncollagenous domain (NC10). The collagenous sequence consists of nine collagenous domains, COL1-COL9, varying in length between 15 and 114 residues and separated by eight noncollagenous sequences, NC2-NC9, which vary in length between 7 and 45 residues.
Disulfides/SH-Groups	Two cysteine residues in the signal peptide, two in the thrombospondin sequence motif present in NC1, one in NC3, one in NC6, and four in NC10. Occurrence of disulfide bonds not known. The two NC1 cysteine residues are conserved in proteins containing the thrombospondin motif, and they are known to form a disulfide bond in the α2 chains of type XI collagen.
General References	Pihlajaniemi, T. and Rehn, M. Two new collagen subgroups: membrane-associated collagens and types XV and XVIII. *Prog. Nucl. Acid Res.* 1995, **50**: 225–262. Prockop, D. J. and Kivirikko, K. I. Collagens: Molecular Biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.* 1995, **64**: 403–434. Huebner, K., et al. Chromosomal assignment of a gene encoding a new collagen type (COL15A1) to 9q21 → q22. *Genomics* 1992, **14**: 220–224.
Ref. for DNA/AA Sequences	Myers, J. C. et al. Identification of a previously unknown human collagen chain, α1(XV), characterized by extensive interruptions in the triple-helical region. *Proc. Natl. Acad. Sci. U. S. A.* 1992, **89**: 10144–10148. Kivirikko, S. et al. Primary structure of the α1 chain of human type XV collagen and exon-intron organization in the 3′ region of the corresponding gene. *J. Biol. Chem.* 1994, **269**: 4773–4779. Muragaki, Y. et al. The human α1(XV) collagen chain contains a large amino-terminal non-triple helical domain with a tandem repeat structure and homology to α1(XVIII) collagen. *J. Biol. Chem.* 1994, **269**: 4042–4046.



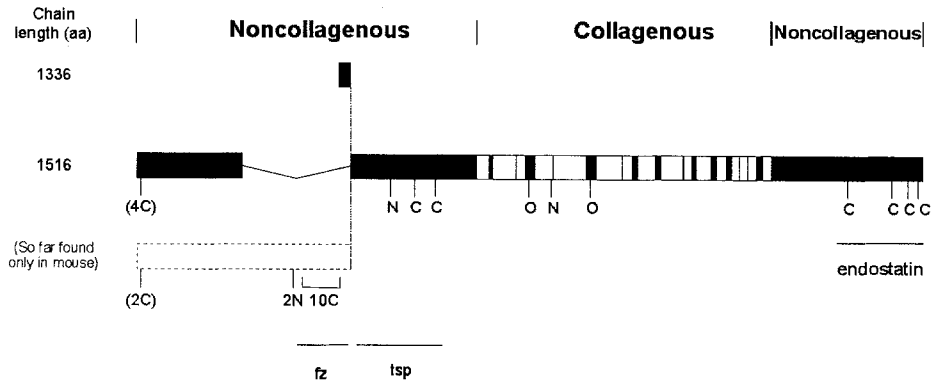
Schematic diagram of the cDNA-deduced type XV collagen polypeptide. The numbering of the noncollagenous (NC) and collagenous (COL) domains is shown below the polypeptide, and the lengths of these domains are given above the polypeptide in terms of their numbers of aas. S and **black box**, signal peptide; **grey boxes**, noncollagenous sequences; **white boxes**, collagenous sequences; **white boxes** with **black bars**, bars indicate 2–3 residue imperfections in the collagenous sequence; **C**, cysteine residues; **N**, potential Asn-linked glycosylation sites; **O**, potential glycosaminoglycan attachment sites; **tsp**, thrombospondin homology.

Collagen type XVIII

Marko Rehn

Synonyms	COL18A1
Abbreviation	None
Classification	Extracellular matrix protein
Description	A non-fibrillar collagen located in many basement membrane zones. High expression levels observed in the human liver, kidney, heart, placenta, pancreas and ovary. Type XVIII collagen was initially characterized by cDNA cloning. The predicted polypeptide has a central highly interrupted collagenous sequence (COL1-9, interrupted by NC2-10) flanked by the N-terminal non-collagenous NC1 and the C-terminal NC11 domains. Long and short N-terminal forms of collagen XVIII have been identified in the human with their NC1 domains of 493 and 303 residues synthesized from two separate promoters. The third variant has also been identified in the mouse with the NC1-domain of 764 residues resulted from alternative splicing of the long form mentioned above. This form probably exists also in the human, but clearly at lower levels than in the mouse. Type XVIII is homologous to the type XV collagen, and they are therefore suggested to form the subgroup of MULTIPLEXINs (multiple triple-helix domains with interruptions) among the family of collagens.
Structure	Type XVIII collagen is at least synthesized as homotrimers with respect to its N-terminal forms. It is not known, whether the variant N-terminal forms can be synthesized as heterotrimers.
Molecular Weight	Calculated molecular weights of 136 and 154 kD for the short and long form, respectively.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Present in many basement membrane zones with unknown functions.
Physiology/Pathology	Type XVIII collagen knock-out mice are without any obvious phenotype.
Degradation	A 20-kD proteolytic fragment called endostatin is released from the C-terminus of type XVIII collagen capable of inhibiting endothelial cell proliferation, angiogenesis, and tumor growth.

Genetics/Abnormalities	The gene is located on the human locus of 21q22.3. No mutations in the human gene have yet been identified.
Half-life	Unknown
Concentration	Unknown
Isolation Methods	The protein has not been isolated.
Amino Acid Sequence	Type XVIII collagen has a modular structure containing sequence homologies to other proteins. The N-terminal common region of NC1 domain contains a sequence homology to thrombospondin-1 found also in other collagens. The strongest homology between type XVIII and XV collagens comprises the endostatin region. The longest N-terminal variant found in the mouse but not yet in the human contains ten closely separated cysteine residues similar to frizzled receptors involved in the *wingless* pathway.
Disulfides/SH-Groups	Two and four cysteine residues are located in NC1 and NC11 domains, respectively. The sequence homology to frizzled receptors contains ten conserved cysteine residues. The nature of the potential disulfide bonds is not known.
General References	Pihlajaniemi, T. and Rehn, M. *Prog. Nucleic Acid Res. Mol. Biol.* 1995, **50**:225-262. Prockop, D.J. and Kivirikko, K.I. *Annu. Rev. Biochem.* 1995, **64**:403-434. Rehn, M. and Pihlajaniemi, T. *Seminars Cell Dev. Biol.* 1996, 7:673-679. O'Reilly, M.S. et al. *Cell.* 1997, **88**:277-285.
Ref. for DNA/AA Sequences	Abe, N. et al. *Biochem. Biophys. Res. Commun.* 1992, **196**:576-582. Muragaki, Y. et al. *Proc Natl Acad Sci USA.* 1995, **92**:8763-8767. Oh, S.P. et al. *Proc Natl Acad Sci USA.* 1994, **91**:4229-4233. Oh, S.P. et al. *Genomics.* 1994, **19**:494-499. Rehn, M. and Pihlajaniemi, T. *Proc. Natl. Acad. Sci. USA.* 1994, **91**:4234-4238. Rehn, M. and Pihlajaniemi, T. *J. Biol. Chem.* 1995, **270**:4705-4711. Saarela, J. et al. *Matrix Biol.* 1997, in press.



Schematic structures of the human α1(XVIII) collagen chains. Collagenous sequences are shown in white, non-collagenous domains in black. The longest N-terminal variant found so far only in the mouse is shown with broken lines. The lengths of the polypeptides are given. The location of endostatin sequences is also shown. C, cysteine residue(s); potential N-glycosylation site; O, potential glycosaminoglycan attachment site; fz and tsp, frizzled and thrombospondin sequence motifs, respectively. Cysteine residues in parentheses are located in the signal sequences.

Collectin receptor

Rajneesh Malhotra and Robert B. Sim

Synonyms	C1q receptor, SP-A receptor, mannan binding protein receptor
Abbreviations	C1qR
Classifications	Acidic cell surface glycoprotein
Description	Collectin receptor is an acidic glycoprotein containing about $17 \pm 3\%$ carbohydrate. Collectin receptor is present on B-lymphocytes, monocytes, polymorphonuclear leucocytes, endothelial cells, fibroblasts, U937 cells and platelets and specialized epithelial cells such as amnion and alveolar cells.
Structure	Physico-chemical measurements, i.e. Stokes radius, sedimentation coefficient values, indicate that detergent solubilized collectin receptor is an elongated dimer of non-covalently linked monomers. It is not known whether it is dimeric or monomeric on the cell surface.
Molecular Weight	56–70 kDa (SDS-PAGE, depending on the acrylamide content of the gel). Laser-desorbtion mass spectrometry indicates that the mw of collectin receptor monomer is $\simeq$ 47 kDa. 115 ± 7 kDa (non-denatured dimer in solution).
Sedimentation Coeff.	4.8 ± 0.2 S
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None. Binds Ca^{2+} ions, but these are not required for ligand binding.
Substrates	None
Inhibitors	Unknown
Biological Functions	Collectin receptor binds to a family of proteins called collectins (defined as molecules which contain collagen-like sequences associated with non-collagenous sequences). The proteins in this group include lung surfactant protein A (SP-A), mannan binding protein (MBP), C1q and conglutinin. C1q consists of collagenous triple helical domains and globular IgG binding domains, while the other named ligands have collagenous triple helical domains and globular C-type lectin domains. Each of these ligands binds to the receptor with similar affinity via charged residues on the collagenous region of the ligands.
Physiology/Pathology	Several physiological functions have been reported to be mediated by interaction of C1q or other collectins with the receptor, e.g. C1q-mediated cytolytic activity of eosinophils, internalisation of particles bearing collectins by polymorphonuclear cells, macrophages and fibroblasts, trapping of C1q-bearing immune complexes by endothelial cells. SP-A has been shown to regulate the secretion of surfactant and to induce the phagocytosis of target particles bearing SP-A. Mannan binding protein has been

	shown to induce the internalisation of bacteria by monocytes and polymorphonuclear cells. In general the collectins appear to function as opsonins, and bind to microorganisms or particulate materials via lectin or charge/hydrophobic interactions. Phagocytic uptake is then mediated via collectin receptor.
Degradation	Unknown
Genetics/Abnormalities	Unknown
Half-life	Unknown
Concentration	Unknown
Isolation Method	Isolation is conveniently done using, e.g. U937 cells or B-lymphoblastoid cells, or tonsil lymphocytes. Cells are lysed in non-ionic detergent, and the solubilised proteins subjected either to affinity chromatography on ligand-Sepharose (e.g. C1q-Sepharose) or the receptor is purified by conventional ion-exchange chromatography and gel filtration. Affinity chromatography alone produces material which is about 50% pure. Conventional chromatography, taking advantage of the high negative charge of the receptor, is more efficient. Reported yields by ion exchange and gel filtration are about 500 μg from 10^{11} tonsil lymphocytes.
Amino Acid Sequence	The N-terminal aa sequence of isolated collectin receptor and the sequence of peptides obtained by V8/trypsin digestion show a high degree of similiarity to the cDNA-derived aa sequence of a human protein reported to be a component of RoSSA autoantigen, or as calreticulin: other similar proteins include Onchocerca volvulus antigen (RAL-1), Aplysia p407 and rat p425, and B50 murine melanoma antigen.
Disulfides/SH-Groups	Unknown
General References	Erdei, A. and Reid, K. B. M. *Mol. Immunol.* 1988, **25:** 1067–1073 (affinity isolation). Malhotra, R. and Sim, R. B. *Biochem. J.* 1989, **262:** 625–631 (isolation and anomalous migration, SDS-PAGE). Malhotra, R., et al. *J. Exp. Med.* 1990, **172:** 955–959 (multiple ligands). Malhotra, R. et al. *Immunology* 1993, **78:** 341–348 (protein sequence). Peerschke, E. I. B. et al. *J. Leukoc. Biol.* 1993, **53:** 179–184 (anomalous migration, SDS-PAGE). Malhotra, R. et al. *Biochem. J.* 1993, **293:** 15–19 (ligand binding site).
Ref. for DNA/AA Sequences	None

Complement receptor type 1

Alison L. Gibb and Edith Sim

Synonyms	CD35; C3b/C4b receptor; Immune adherance receptor.
Abbreviations	CR1
Classifications	Receptor, membrane glycoprotein
Description	CR1 is a single polypeptide chain integral membrane glycoprotein. It has been identified on many different cell types – erythrocytes, neutrophils, monocytes, B-lymphocytes, some T-lymphocytes, eosinophils, mast cells, follicular dendritic cells and glomerular podocytes. A soluble form of the receptor has been identified in plasma. Four polymorphic variants of CR1 have been identified that differ in Mr but appear not to differ in function.
Structure	CR1 is a mosaic protein containing 6-9% complex N-linked carbohydrate. It is composed almost entirely of modules known as short concensus repeats (SCRs), also known as complement control protein modules (CCPs). These modules each contain 60 to 70 aa and every 8th SCR is a highly homologous repeat (65-100% identical). Thus seven SCRs make up what is known as a long homologous repeat (LHR). SCRs occur over 140 times in more than 20 extracellular mosaic proteins including 12 proteins of the complement system. The polymorphic forms of CR1 differ in the number of SCRs they contain. The coding sequence of the most common allotype of CR1 (A) is composed of 2039 aa: a 41-residue signal peptide (removed in the mature protein), an extracellular domain of 1930 residues (30 SCRs, 4 LHRs), a 25-residue transmembrane domain and a 43-residue cytoplasmic domain. In electron microscopy, the molecule has an elongated appearance. Recently, the 2- and 3- dimensional structures of an SCR from the homologous complement protein Factor H have been described. The globular module is approx. 3.8 nm long and is based on a β-sandwich arrangement. One face of the module is made up of 3β strands, hydrogen bonded to form a triple-stranded region at its centre, and the other face is formed from 2 separate β strands. Both faces contribute highly conserved hydrophobic side chains to the compact core. The C- and N-termini of the module are at opposite ends of the module.
Molecular Weight	190 kDa: allotype A 220 kDa: allotype B 160 kDa: allotype C 250 kDa: allotype D All determined by SDS-PAGE, non-reducing conditions. The mw of each allotype appears to be 30-50 kDa greater under reducing conditions. The mw of the A allotype is calculated to be 223,589 Da from sequence and carbohydrate data.
Sedimentation Coeff.	Unknown
Isoelectric Point	6.6: A; 6.2: B; 6.7: C; D - unknown
Extinction Coeff.	15 (280nm, 1%, 1cm)
Enzyme Activity	None

Coenzymes/Cofactors	CR1 is itself a cofactor for the serine protease, Factor I, which digests the complement protein fragments C3b (to iC3b), C4b (to C4c and C4d) and probably also iC3b (to C3c and C3dg). In order to become susceptible to cleavage by Factor I, C3b or C4b has to form a complex with the CR1 molecule. Once cleavage has occurred, CR1 dissociates, apparently unaltered.
Substrates	None
Inhibitors	The receptor can be actively blocked in vitro by the use of antibodies directed against the receptor. Zn^{2+} inhibits the cleavage of C3b by Factor I in the presence of CR1.
Biological Functions	CR1 has a variety of biological functions, all dependent on its ability to bind the complement protein fragments C3b, C4b and iC3b. CR1 has been shown to bind immune complexes with C3b or C4b bound to their surfaces, showing a preference for the C4A isotype of C4 rather than the C4B isotype. CR1 acts as a cofactor for the breakdown of C3b and C4b by Factor I. CR1 has also been shown to promote decay of the alternative and classical pathway C3 and C5 convertases, essential enzymes within the complement cascade. Cells bearing CR1 can be identified by their ability to form rosettes with erythrocytes coated with C3b or C4b.
Physiology/Pathology	Since CR1 acts as a cofactor for the breakdown of the complement proteins above, and promotes the decay of convertases essential for the continuation of the complement system once activated, CR1 plays a key regulatory role in the complement cascade. The importance of CR1 as a regulator has been emphasised by reports of the effectiveness of a recombinant, soluble form of CR1 (sCR1) in the reduction of complement-mediated tissue damage, e.g. after myocardial ischemia and after both xeno- and allotransplantation. CR1 is mainly involved in the clearance of immune complexes. CR1 binds reversibly the C3b and C4b fragments of C3 and C4 which have, as a result of complement activation, become covalently attached to the surface of immune complexes and other complement activating particles. Since erythrocytes predominate numerically in blood, they are the main cell type for immune complex binding via CR1. C3b/C4b coated immune complexes are bound to erythrocytes and are removed from the cells during transit through hepatic circulation. CR1 on circulating phagoytic cells promotes removal of immune complexes by endocytosis and phagocytosis. CR1 on lymphocytes may participate in the regulation of immunoglobulin secretion, whereas CR1 on follicular dendritic cells may be involved in the generation of B memory cells. Reduced erythrocyte CR1 levels have been detected in disorders associated with complement activation including cold agglutinin disease, autoimmune haemolytic anaemia, paroxysomal nocturnal haemoglobinuria, Sjogren's syndrome, lepromatous leprosy, juvenile rheumatoid arthritis and systemic lupus erythematosus (SLE). Reduced CR1 levels result in inefficient clearance of immune complexes, leading to the deposition of large, insoluble immune complexes at inappropriate sites within the body, e.g. in immune complex diseases such as SLE. The reduction of CR1 on erythrocytes in SLE patients could well be due to degradation of CR1 during the disease process (see below).
Degradation	In a study of the catabolism of CR1 on erythrocytes, the decay of CR1 was exponential with a half life which differed between individuals. Removal of CR1 from erythrocytes is thought to occur by proteolysis in both physiological and pathological conditions. Removal of CR1 may occur during the passage of erythrocytes bearing immune complexes through the liver. In vitro, mild treatment of erythrocytes with trypsin, plasmin or thrombin

results in the release of Factor I cofactor activity into the supernatant. The enzyme responsible for the proteolysis of CR1 in vivo has not yet been identified.

Genetics/Abnormalities

CR1 is located within the RCA (Regulation of Complement Activation) locus on chromosome 1 (band q32). The B allele extends approximately 160 kb and contains at least 42 exons. This gene contains a region of 20 kb that does not hybridise to cDNA probes specific for the 4 LHRs in the A allele, thus the B allele has an additional segment that appears to encode an additional LHR. The CR1 gene is linked to the genes for the functionally and structurally similar molecules, CR2, DAF, MCP and C4bp. The CR1 gene lies between the genes for MCP and CR2, followed by the genes for DAF and C4bp, all in a 1500 kb region of DNA. CR1 displays an inherited polymorphism of numerical expressions on erythrocytes. The level of expression correlates with 2 autosomal codominant alleles which display a Hind III restriction fragment length polymorphism. The level of expression of different CR1 forms on erythrocytes may vary markedly between individuals but is constant in any particular healthy subject. The gene frequencies of the 4 allotypes of CR1 are A - 0.83, B - 0.16, C - 0.01, D - 0.001. There appears to be no correlation between a particular CR1 allotype and disease.

Half-life

11-32 days

Concentration

This varies with both individual, clinical condition and cell type. The mean level of CR1 on erythrocytes is approx. 550 per cell. Resting monocytes and neutrophils express approx. 5000 CR1 on each cell surface, the remaining 80-90% of the cell's CR1 exists in an intracellular compartment from which it can be mobilised. The level of CR1 on neutrophils can increase 20-fold after neutrophil activation. B-lymphocytes have approx. 20,000-40,000 CR1 per cell. The level of soluble CR1 in plasma is between 13 and 81 $\mu g\ L^{-1}$. Erythrocyte CR1 levels are decreased when circulating immune complexes are increased e.g. in active SLE.

Isolation Method

Erythrocytes are the most convenient source of CR1. The first method described for the purification of CR1 from detergent lysates of erythrocyte membranes involved purification on a BioRex-70 column, followed by a Sepharose C3 column. This was followed by purification on a Bio-Gel A-5m column finishing with affinity chromatography with Sepharose-lentil-lecitin. Various modifications have since been made to this method, and schemes have been described using monoclonal antibody affinity and heparin-Sepharose columns.

Amino Acid Sequence

There are many sequence homologies between CR1 and the functionally similar proteins Factor H, C4bp, CR2, MCP and DAF, all of which are C3/C4 binding proteins and contain SCR modules. There are 4 invariant cysteines in each SCR in CR1, along with 17 other aa which are conserved in 40 - 100% of SCRs. Major determinants of C4b binding have been localised to SCR-1 and SCR-2, whereas those for C3b binding are in SCR-8, SCR-9 and in the nearly identical SCRs, SCR-15 and SCR-16. Mutagenesis has been used to study which aa within each SCR are important for the preferential binding of C4b and C3b to particular SCR modules. The following residues (bold) have so far been reported to be involved in ligand binding in CR1:

C4b-SCR-1 28-LNYEC**R**P**G**YS-37
C4b-SCR-2 1-KSC**RN**PPDP-9..28-IKYSCTKGYR-37
C3b-SCR-9 45-CILSGNAAHW**STKP**PICQR-63

The invariant cysteines are shown in italics.

Disulfides/SH-Groups	All cysteines are involved in disulfide bridges. Within each SCR, the first and third cysteine residues are disulfide bridged, as are the second and fourth cysteines.
General References	Fearon, D. *PNAS* 1979, **76**:5867-5871 (isolation). Ahearn, J.M. and Fearon, D.T. *Adv. Immunol.* 1989, **46**:183-219 (review). Weisman, H.F. et al. *Science* 1990, **249**:146-151 (sCR1). Norman, D.G. et al. *J. Mol. Biol.* 1991, **219**:717-725 (SCR tertiary structure). Krych, M. et al. *Curr. Opin. Immunol.* 1992, **4**:8-13 (review including mutagenesis studies). Birmingham, D.J. *Critical Rev. in Immunology* 1995, **15**:133-154 (review of physiological role - comparison with other sepcies). Molina, H. et al. *J. Immunol.* 1995, **154**:5426-5435 (Model of a pair of SCR structures). Ryan, U.S. *Nature Medicine* 1995, **1**:967-968 (review of potential role in xenotransplantation).
Ref. for DNA/AA Sequences	Wong, W.W. et al. *Proc. Natl. Acad. Sci. USA* 1985, **82**:7711-7715. Klickstein, L.B. et al. *J. Exp. Med.* 1987, **165**:1095-1112. Hourcade, D. et al. *J. Exp. Med.* 1988, **168**:1255-1270. Klickstein, L.B. et al. *J. Exp. Med.* 1988, **168**:1699-1717. AA Sequence - SwissProt p17927 DNA Sequence - GenEmbl x05309, x14362, y00812.

Cortisol Binding Globulin

Paul F. Edgar and Robin W. Carrell

Synonyms	Transcortin
Abbreviations	CBG
Classification	Non-inhibitory member of the serpin family of serine protease inhibitors.
Description	A circulating plasma glycoprotein synthesized in the liver comprising a single polypeptide chain of 383 amino acids with carbohydrate side chains at 5 of 6 consensus sequences for N-linked glycosylation. Oligosaccharides attached to Asn^{9}, Asn^{238}, Asn^{308} and Asn^{347} are predominantly biantennary while more branching oligosaccharides are preferentially linked to Asn^{74} and Asn^{154}. There is one steroid binding site per molecule which interacts preferentially with biologically active glucocorticoids.
Structure	By homology will predictably have typical serpin structure (see α-1-proteinase inhibitor).
Molecular weight	50–60,000 (SDS-PAGE); 42,646 (aa sequence)
Sedimentation	3.79 S
Isoelectric Point	3.7–4.3 5.5 (desialylated)
Extinction Coeff.	6.45 (280 nm,1 %,1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	Binds cortisol/progesterone, K_a for cortisol; $71 \pm 7 \times 10^7\ M^{-1}$, progesterone; $59 \pm 9 \times 10^7\ M^{-1}$ (in the presence of 20 mM 2-mercaptoethanol, 4 °C, pH 7.4).
Substrates	None
Inhibitors	Cleavage of CBG by human neutrophil elastase leads to a ten fold reduction in affinity for cortisol.
Biological Function	The regulation of steady state free glucocorticoid concentrations in plasma and the delivery of bound cortisol to sites of inflammation where it is released by neutrophil elastase cleavage.
Physiology/Pathology	No known disease association with deficiency or low affinity variants. No null variants have been reported.
Degradation	Unknown
Genetics/Abnormalities	The gene is on chromosome 14 at q31–32.1 and consists of 5 exons distributed over 19kb. The last four exons contain all information coding for the CBG protein including the 22 amino acid signal sequence. One reduced cortisol affinity variant has been discovered and characterised: transcortin Leuvin, $Leu \rightarrow His^{93}$.
Half-life	Unknown

Concentration	Plasma: 15–25 mg/L . A negative acute phase protein with levels decreasing to zero in septic shock. During late pregnancy concentration increases 2 fold.
Isolation Method	Steroid affinity chromatography of plasma followed by hydroxylapatite chromatography.
Amino Acid Sequence	Shares greater than 40% direct amino acid homology with a-1-proteinase inhibitor and a_1-antichymotrypsin. Post-translational N-linked glycosylation at Asn^{238} during biosynthesis is necessary for high affinity cortisol binding. Of the two cysteines, at positions 60 and 238, one is present in the binding site for cortisol as is one of the four tryptophan residues.
Disulfides/SH-groups	No intramolecular disulfide bonds. Two free sulfhydryl groups.
General References	Hammond, G. L. *Endocrine Rev.* 1990, **11**: 65–79. Rosner, W. *Endocrine Rev.* 1990, **11**: 80–91. Hammond, G. L., Smith, C. L., Underhill, D. A. *J. Steroid. Molec. Biol.* 1991, **40**: 755–762.
Ref. for DNA/AA Sequences	Hammond, G. L. et al. *Proc. Natl. Acad. Sci. U. S. A.* 1987, **84**: 5153–5157.

C-reactive protein

Richard F. Mortensen

Synonyms	None
Abbreviations	CRP; CxRP (rabbit)
Classifications	Gamma-globulin in serum
Description	CRP is an acute phase serum protein synthesized by liver hepatocytes in response to inflammatory cytokines.
Structure	Pentraxin based on electron microscopy and centrifugation analysis. Structure of each protomer subunit defined by X-ray diffraction as a flattened β-jelly roll topology similar to lectins such as concanavalin A. Crystal structure of both serum amyloid P-component (SAP) and CRP recently published.
Molecular Weight	118,000 (sedimentation equilibrium). 120,000 (aa sequence). Five identical subunits of 24,000. (SDS-PAGE and aa sequence).
Sedimentation Coeff.	6.5 S
Isoelectric Point	6.4
Extinction Coeff.	19.5 (280nm, 1%, 1cm.)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	Phosphorylcholine (10^{-5} M). Primary phosphate monoesters with lower affinity. Fibronectin, laminin, chromatin, histones, snRNP, polycations. Binding requires Ca^{2+}-ions to alter conformation of CRP.
Inhibitors	None
Biological Functions	CRP's precise biological function is unknown. CRP displays biological activities such as complement activation, opsonization, and activation of leukocyte antimicrobial activities consistent with a nonspecific host defense function. CRP, especially its peptides derived by degradation, also displays a few anti-inflammatory activities.
Physiology/Pathology	Acute phase reactant and thus its blood level is a gauge of the extent of tissue damage and infection. May increase in serum concentration by up to 1,000-fold within 24 hrs.
Degradation	Peptides generated by neutrophil enzymatic degradation decrease neutrophil activities, but increase monocyte phagocytosis, chemotactic activity, superoxide production, and cytokine production.
Genetics/Abnormalities	No known deficiency. Single copy gene on long arm of human chromosome 1.
Half-life	13 - 16 hrs (based on animal studies).

1. Installment 1992 (revised 1997)

Concentration	Normal serum con centration is 100-500 $\mu g\ L^{-1}$. Acute phase levels are as high as 200-500 mg L^{-1}. Doubling time is approximately 8 hrs.
Isolation Method	Affinity chromatography on phosphorylcholine-agarose with elution by Ca^{2+} chelation. Source of protein is ascitic fluids or acute phase serum.
Amino Acid Sequence	Each subunit has 206 aa. Phosphorylcholine binding region is conserved: (47)RGYSIFSYATKRQDNE(62);
Disulfides/SH-Groups	1 intrachain disulfide. No interchain disulfide linkage. Subunits non-covalently linked.
General References	Gewurz, H., Zhang, X.H. and Lint, T.F. Structure and Function of the Pentraxins. *Current Opinion in Immunol.* 1995, **7**:54-64. Steel, D.M. and Whitehead, A.S. *Immunology Today* 1994, **15**:81-88. Srinivasan, N., White, H.E., Emsley, J. et al. *Structure* 1994, **2**:1017-1027. Shrive, A.K., Cheetham, G.M.T., Holden, D. et al. *Nature Struct. Biol.* 1996, **3**:346-354.
Ref. for DNA Sequences	Lei, K.J. et al. *J. Biol. Chem.* 1985, **260**:13377-13383. Woo, P. et al. *J. Biol. Chem.* 1985, **260**:13384-13388.

Creatine kinase

Lucia Sacchetti and Giuliana Fortunato

Synonyms	Adenosine 5'-trisphosphate-creatine phosphotransferase; Creatine phosphokinase; Creatine phosphoryltransferase.
Abbreviations	CK; CPK
Classifications	EC 2.7.3.2
Description	Creatine kinase contributes to restoring the cellular energy reservoir catalyzing the reversible reaction between phosphocreatine (PCr) and ADP. Five CK isoenzymes have been described in mammalian tissues, three forms are cytosolic and two are mitochondrial. Cytosolic CK is a dimeric molecule consisting of two polypeptide subunits, namely MM-CK, MB-CK and BB-CK. Mammalian CK-MM is mainly located in mature skeletal muscle and myocardium; CK-BB in brain, neural tissues, embryonic skeletal and cardiac muscle; and CK-MB in adult heart tissue. Two forms of mitochondrial CK (Mi-CK) have recently been reported, the ubiquitous (uMt-CK) and the sarcomeric (sMt-CK). The latter is expressed only in striated muscle including myocardium. Both occur within the mitochondrial intermembrane space.
Structure	Each M and B subunit of the cytosolic enzyme consists of a single polypeptide chain without disulfide bridges and contains one catalytic site and one reactive SH group. Optical rotatory dispersion studies indicate a compact globular structure containing 25-30% α-helix and less than 15% β-pleated sheet. Within the molecule the subunits are organized as two cigar-shaped structures lying side-by-side. In vitro, Mi-CK forms dimeric or octameric molecules. With electron microscopy, the latter appear as cube-like molecules with a side-length of 10 nm, and show a four-fold symmetry and a central stain-filled cavity.
Molecular Weight	81 kDa: cytosolic CK (sedimentation) 80-85 kDa: dimeric Mi-CK 350 kDa: octameric Mi-CK (gel permeation chromatography and analytical ultracentrifugation).
Sedimentation Coeff.	5.0 S (dimeric heart CK)
Isoelectric Point	5.1 - 6.8: cytosolic CK; 6.8 - 7.0 (dimeric Mi-CK)
Extinction Coeff.	8.8: muscle CK (280nm, 1%, 1cm), pH 8.0.
Enzyme Activity	CK isoenzymes catalyze the reversible transphosphorylation reaction between PCr and ADP: $PCr^{2-} + MgADP^{-} + (x)H^{+} \leftrightarrow MgATP^{2-} + Cr$ where (x) is about 1 at pH $\geq$ 6.5.
Coenzymes/Cofactors	CK requires bivalent metal ions (Mg^{2+}, Mn^{2+}, Co^{2+}, Sr^{2+} and Ba^{2+}) as activators. The enzyme combines with ATP or ADP to form a metal nucleotide-complex.
Substrates	PCr, MgATP and the corresponding dephosphorylated compounds, creatine and MgADP; all are ionized as indicated above.

Inhibitors	A number of anions are non competitive inhibitors. In decreasing order of efficiency these are: I^-, SO_4^{2-}, NO_2^-, Br^-, SO_3^{2-}, Cl^- and F^-. Ca^{2+} competes with Mg^{2+} as do Fe^{3+} and Cu^{2+}. The CK inactivation is caused by the oxidation of the sulfide groups of the enzyme; thiols, including N-acetyl-cysteine, monothioglycerol, β-mercaptoethanol, dithiothreitol, and glutathione are reactivators.
Biological Functions	The major concentration of CK is found in high-energy tissues such as brain, heart and skeletal muscle, where it supplies the high energy phosphate compounds (ATP and phosphocreatine) pool and ensures effective energy transport from the sites of production (mitochondria) to the sites of utilization (such as the muscle myofibrillar M-band). This high-energy phosphate shuttle is made possible by the presence of both mitochondrial and cytosolic CK, PCr being readily diffusable. In addition, mRNAs for the cytosolic isoforms of CK have a differential localization in muscle cell lines with respect to non-muscle cell lines. In muscle cells, the M form is localized at the cell periphery and the B form in the perinuclear region, whereas in non-muscle cells, both the M and B forms are distributed uniformly over the cytoplasm.
Physiology/Pathology	Decrease of MM-CK has been observed in dystrophic skeletal muscle, as well as in cardio-myopathic hearts; in serum CK activity is highly increased. Measurement of CK activity in human serum is useful in the diagnosis of heart (CK-MB), muscle (CK-MM) and brain (CK-BB) diseases. A diagnostic role has not yet been established for Mi-CK. However, recent studies have indicated the presence of Mi-CK in serum as a sensitive biochemical marker of irreversible cell damage.
Degradation	In vivo, serum carboxypeptidase N hydrolyzes the M subunit of native CK (i.e., MM_3-CK) removing the positively charged C-terminal lysine. Thus, there are three isoforms for MM-CK (MM_3, MM_2 and MM_1), and two isoforms for MB-CK (MB_2 and MB_1). Also these forms are used for the diagnosis of acute myocardial infarction.
Genetics/Abnormalities	Four nuclear genes encode the known CK monomers: M-CK, B-CK, ubiquitous and sarcomeric mitochondrial CK. Loci for the M and B subunits have been localized on chromosome 19q13 and 14q32.2, respectively. The genes for ubiquitous and sarcomeric Mi-CK differ considerably from those of cytosolic CK isoenzymes. They span from 5.5 to 37 kb, contain 9 and 11 exons respectively and are located on chromosomes 15 and 5. The identical length of the coding region exons and the locations of the exon junctions of the human isoenzymes suggest that a first duplication of a common primordial CK gene gave rise to the two ancestral mitochondrial and cytosolic CK genes. In addition, in contrast to M and B CK mRNAs, Mi-CK mRNAs code for an amino-terminal target peptide of 38 aa which is presumably essential for import into the mitochondria.
Half-life	Serum: CK-MM ≈ 15 h, CK-MB ≈ 12 h, CK-BB ≈ 3 h.
Concentration	Normal serum CK levels (at 37°C) reach up to 160 U L^{-1} in adult males and up to 130 U L^{-1} in adult females (one CK unit catalyzes the cleavage of 1 μmole of PCr in one minute).
Isolation Method	Human muscle and brain CK have been purified by ammonium sulphate fractionation followed by chromatography on DEAE-Sephadex. The purification of human heart Mi-CK, after extraction with phosphate, was carried out by ammonium sulphate precipitation, chromatofocusing ion exchange, affinity and hydrophobic interaction chromatography.

Amino Acid Sequence	The aa sequences of CK subunits were obtained by translating the cDNA sequences. A high degree of conservation was observed in the CK gene family and was confirmed by the high homology observed between the human B-CK polypeptide and B-CK from other sources (90-98%) as well as between human B-CK and M-CK polypeptides (80%). In particular, the amino acid sequence around the active site (275-291 aa) is highly conserved in the M and B subunits. The Asp-335 residue is conserved in all Mi-CK and in the majority of cytosolic CK; its involvement in Mg^{2+} binding has been suggested. Less conserved regions might be involved in binding to cellular membranes or to the myofibrillar M-band.
Disulfides/SH-Groups	One SH-group per subunit.
General References	Wyss, M. et al. *Biochim. Biophys. Acta* 1992, **1102**:119-166. Lang, H. (ed.) *Creatine Kinase Isoenzymes.* Springer-Verlag, Berlin, Heidelberg, New York, 1981. Lang, H. and Würzburg, U. *Clin. Chem.* 1982, **28**:1439-1447. Wilson, I.A. et al. *Biochem. J.* 1995, **308**:599-605.
Ref. for DNA/AA Sequences	Haas, R.C. et al. *J. Biol. Chem.* 1989, **264**:2890-2897. Klein, S.C. et al. *J. Biol. Chem.* 1991, **266**:18058-18065. Mariman, E.C.M. et al. *Genomics* 1987, **1**:126-137. Perryman, M.B. et al. *Biochim. Biophys. Res. Commun.* 1986, **140**:981-989. The nucleotide sequences have been deposited in the EMBL GENBANK data libraries under the following accession numbers: ubiquitous Mi-CK (J04469); sarcomeric Mi-CK (M72981); B-CK (J03036).

Cyclophilin-18

Wei Li and Robert E. Handschumacher

Synonyms	Cyclophilin A, Cyclophilin-1
Abbreviations	CyP-18, CyP-A, CyP-1
Classifications	
Description	An abundant, highly conserved, ubiquitous cytosolic protein found in all human, bovine and rat tissues tested. It is identical to the pig peptidyl-prolyl cis-trans isomerase and binds to the immunosuppressive drug, cyclosporin A. This is the most abundant member of the cyclophilin family which includes cyclophilin-40, the other cytosolic form, as well as cyclophilins B, C and 3 that are membrane associated.
Structure	The overall structure is a β-barrel, consisting of eight antiparallel β-strands wrapping around the barrel surface and two α-helices sitting on the top and the bottom closing the barrel. Inside the barrel, seven aromatic and other hydrophobic residues form a compact hydrophobic core.
Molecular Weight	17,737 Da (human aa sequence)
Sedimentation Coeff.	Unknown
Isoelectric Point	9.1 and 8.5, two major species
Extinction Coeff.	Unknown
Enzyme Activity	Peptidyl-prolyl *cis-trans* isomerase
Coenzymes/Cofactors	Unknown
Substrates	*In vivo* substrate unknown. *In vitro* substrate used to determine biological activity: succinyl-Ala-Ala-Pro-Phe-*para*-nitroaniline.
Inhibitors	Immunosuppressive drug, cyclosporin A. Endogenous inhibitor unknown.
Biological Functions	*In vitro*, cyclophilin-18 complexed to CsA inhibits the phosphatase activity of calcineurin and the assembly of functional NF-AT (nuclear factor of activated T-cells). Cyclophilin-18 also causes chemotaxis of neutrophils and eosinophils and has been shown to bind to the Gag polyprotein (Pr55)of the HIV-1 virus.
Physiology/Pathology	Unknown
Degradation	Unknown
Genetics/Abnormalities	Human cyclophilin-18 gene contains five exons and four introns. Six members of the middle repetitive alu gene family are present in one or other orientation in the non-coding regions of the cyclophilin gene. There is only single copy in the human haploid genome.
Half-life	About 40 hr in Jurkat cells
Concentration	Approx. 0.1–0.3% of soluble protein in many tissues.

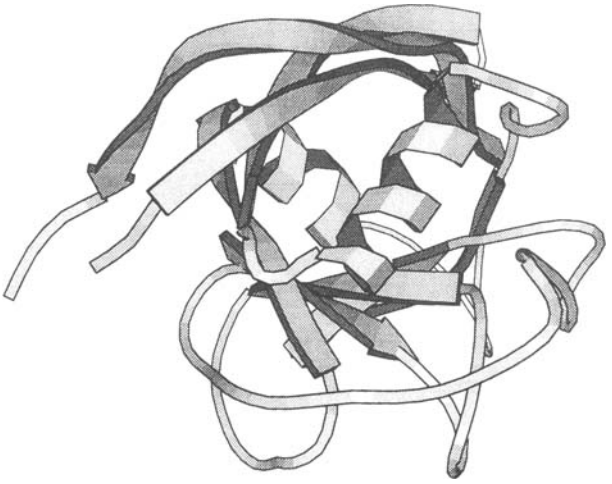
Isolation Method	Isolated from calf thymus 100,000 g supernatant using filtration with sulfone exclusion membrane (Gelman), Matrex blue A column, preparative isoelectric focusing column and CM-300 weak cation exchange HPLC column.
Amino Acid Sequence	Human cyclophilin-18 consists of 163 aa residues. Bovine cyclophilin-18 differs from human cyclophilin-18 by two aa residues. Human CyP-18 shares 65%, 76% and 61% identity with human cyclophilin 2, 3 and 40. CyP-18 homologs have been found in *Saccharomyces cerevisiae, Neurospora crassa, Escherichia coli,* tomato, maize, and *Brassica napus.* The Trp 120 is involved in high affinity binding of cyclosporin A to cyclophilin-18.
Disulfides/SH-Groups	4 cysteines but no disulfides.
General References	Handschumacher, R. E. et al., *Science* 1984, **226**: 544–547; Harding, M. W. et al., *J. Biol. Chem.* 1986, **261**: 8547–8555.
Ref. for DNA/AA Sequences	Haendler, B. et al., *EMBO J.* 1987, **6**: 947–950; Haendler, B. and Hofer, E. *Eur. J. Biochem.* 1990, **190**: 477–482.

Cyclophilin-20

Fabrice Allain and Geneviève Spik

Synonyms	Cyclophilin B, S-cyclophilin, peptidyl- prolyl *cis-trans* isomerase B, Secreted Cyclophilin-Like Protein
Abbreviations	CyP-20, CyPB, CPH-B, hCyP2, PPIB, SCYLP
Classifications	Immunophilin, Cyclosporin binding protein, Peptidyl prolyl *cis-trans* isomerase (EC 5.2.1.8)
Description	Cyclophilin is a family member of structurally homologous proteins that include CyPA (CyP-18), CyPC (CyP-23), and CyPD. Cyclophilins have been shown to bind stereospecifically one molecule of the immunosuppressive drug cyclosporin A (CsA) used to prevent graft rejection after organ transplantation. CyPB, CyPA and CyPC are supposed to recognize the same site of CsA, however the relative affinity of CyPB for CsA is ten-fold higher than that of CyPA and twenty-fold higher than that of CyPC. CyPB is an abundant and highly conserved protein, exhibiting wide tissue distribution. It is targeted to the secretory pathway via a signal sequence and occurs in the intracellular compartments where it colocalizes with Ca^{2+}-binding proteins. The secreted form found in biological fluids is not glycosylated.
Structure	Crystallographic structure of a complex between recombinant human CyPB and a cyclosporin A analog has been determined at 1.85 Å resolution and compared to that of cyclophilin A (CyPA). The overall structures of the two proteins are similar, consisting in a globular protein type, with high β-barrel content, however, significant differences occur in two loops and at the N- and C-termini. The CsA- binding pocket is constituted by a loop and four β-strands and has the same structure as in CypA.
Molecular Weight	20.4 kDa (calculated) for the recombinant form expressed in *E. coli*, 21 kDa (SDS-PAGE) for the secreted form isolated from human milk.
Sedimentation Coeff.	Unknown
Isoelectric Point	9.5
Extinction Coeff.	15.5 (280 nm,1 %,1 cm)
Enzyme Activity	Peptidyl- prolyl *cis-trans* isomerase. CyPB catalyzes rate- limiting proline isomerization steps of small proteins and peptides, *in vitro*.
Coenzymes/Cofactors	None
Substrates	Unknown
Inhibitors	Cyclosporin A and analogs inhibit the enzymatic properties of the protein, but the peptidyl-prolyl isomerase activity and the immunosuppression are probably unrelated.
Biological functions	Not yet elucidated. CyPB is thought to play a role in protein folding by catalyzing isomerization of X-Pro imide bonds, and may act as chaperone for protein trafficking and macromolecular assembly. CyPB specifically interacts with a transmembrane protein located in vesicular-like structures. This protein, termed calcium-signal modulating cyclophilin ligand (CaML)

	appears to be a new participant in calcium-signal transduction pathway by causing an influx of calcium. The complex CyPB-CsA has a four-fold higher efficiency to inhibit *in vitro*, the calcineurin phosphatase activity than the complex CyPA-CsA. Both CsA- complexed and free CyPB interact with a T-cell membrane receptor and are internalized into the cell.
Physiology/Pathology	Ubiquitous protein mainly present in cellular content and biological fluids as milk and plasma at various concentrations. Inactivation of calcineurin by the CsA-CyPB complex results in the inhibition of transcription of the IL-2 gene in activated T-cell. High levels of CyPB are observed in CsA-resistant cells. The drug mobilizes the protein from the endoplasmic reticulum and promotes its secretion in the medium.
Degradation	CyPB is internalized and degraded by T-cells. The major pathway for its degradation is not yet known.
Genetics/Abnormalities	Gene of CyPB is localized on chromosome 15.
Half-life	Unknown
Concentration	Mature milk: 120 μg/L; plasma: 150 μg/L; leukocyte content: 250 μg/g of cell protein.
Isolation Method	Secretory CyPB was isolated for the first time from human milk. Recombinant material has been produced in a number of different expression systems. Both native and recombinant CyPB are usually isolated by gel filtration and cation- exchange chromatography, using the FPLC system on a Mono-S column.
Amino Acid Sequence	Single polypeptide chain containing 184 aa for the mature protein and 216 aa for the preprotein with signal peptide. CyPB shares about 65% identical residues with CyPA and CyPC, and more than 80% in the central core of the proteins. Significant differences occur between CyPB and CyPA in two loops (Ile_{20}-Glu_{23} and Thr_{153}-Lys_{163}) and at the C- and N- terminal extensions. CyPB contains an endoplasmic reticulum retention signal (Val_{175}-Glu_{184}) and a putative N-glycosylation site (Asn_{116}-Gly-Ser).
Disulfide/SH-Groups	No disulfide. A single free sulfhydryl in the C-terminus.
General References	Bergsma, D. J. et al. *J. Biol. Chem.* 1991, **266:** 23204- 23214. Walsh, C. T. et al. *J. Biol. Chem.* 1992, **267:** 13115–13118. Galat, A. *Eur. J. Biochem.* 1993, **216:** 689–707. Ryffel, B. *Biochem. Pharmacol.* 1993, **46:** 1–12. Mikol, V. et al. *Proc. Natl. Acad. Sci. USA* 1994, **91:** 5183–5186. Allain, F. et al. *J. Biol. Chem.* 1994, **269:** 16537–16540. Bram, R. J. and Crabtree G. R. *Nature* 1994; **371**:355–358.
Ref. for DNA/AA Sequences	Spik, G. et al. *J. Biol. Chem.* 1991, **266:** 10735–10738. EMBL accession No M63573. Price, E. R. et al. *Proc. Natl. Acad. Sci. U. S. A.* 1991, **88:** 1903–1907. EMBL accession No M60857.



Molecular structure of recombinant human cyclophilin-20, according to Mikol V. et al. *Proc. Natl. Acad. Sci. USA* 1994, **91:** 5183–5186.

Cyclophilin-40

Lynda J. Kieffer and Robert E. Handschumacher

Synonyms	None
Abbreviations	CyP-40
Classifications	
Description	Present in all rat and human tissues tested. N-terminal domain is homologous to cyclophilin-18, an ubiquitous, abundant peptidyl-prolyl cis-trans isomerase which binds the immunosuppressive drug, cyclosporin A. C-terminal domain is homologous to P59, a member of the steroid receptor complex.
Structure	Unknown
Molecular Weight	40,633 Da: human (calculated from aa sequence) 40 kDa: bovine (SDS-PAGE) 40,027 Da: bovine (electrospray mass spectrometry).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.3 and 5.5 (2 major species)
Extinction Coeff.	Unknown
Enzyme Activity	Peptidyl-prolyl cis-trans isomerase
Coenzymes/Cofactors	Unknown
Substrates	In vivo substrate unknown. In vitro substrate used to determine biological activity: succinyl-Ala-Ala-Pro-Phe-para-nitroaniline.
Inhibitors	Immunosuppressive drug, cyclosporin A. Endogenous inhibitor unknown.
Biological Functions	In vitro, cyclophilin-40 complexed to Cyclosporin A inhibits the phosphatase activity of calcineurin. However, not thought to be biologically significant because a complex of Cyclosporin A with cyclophilin-18, more abundant than cyclophilin-40, is a better inhibitor of calcineurin. Cyclophilin-40, like cyclophilin-18, causes chemotaxis of neutrophils.
Physiology/Pathology	Unknown
Degradation	Unknown
Genetics/Abnormalities	Synthesized from one mRNA which has two poly-adenylation signals. Only the larger mRNAspecies (1800 bp) is found in tissues.
Half-life	Unknown
Concentration	$\simeq$ 2mg/kg tissue.
Isolation Method	Isolated from calf brain 100,000 × g supernatant using cyclosporin A affinity column followed by a molecular sieving column.

Amino Acid Sequence	Human cyclophilin-40 consists of 370 aa residues. Amino acids 15–185 are 60.6% identical to cyclophilin-18. Amino acids 220–369 are 30.7% identical to P59. Cyclophilin-40 contains a His at position 141, which corresponds to Trp-120 of cyclophilin-18. This Trp is involved in high affinity binding of cyclosporin A to cyclophilin-18. The lack of Trp at this position in cyclophilin-40 may explain the 15-fold lower affinity of cyclophilin-40 for cyclosporin A compared to the affinity of cyclophilin-18 for cyclosporin A. Cyclophilin-40 has a consensus sequence for nitrogenases at positions 175–189, LCVIAECGELKEGDD. Biological significance is unknown.
Disulfides/SH-Groups	7 cycteines but no disulfides are suspected since this protein is believed to be cytosolic.
General References	Kieffer, L. J. et al. *J. Biol. Chem.* 1992, **267:** 5503–5507. Kieffer, L. J. et al. *J. Biol. Chem.* 1993, **268:** in press.
Ref. for DNA/AA Sequences	Kieffer, L. J. et al. *J. Biol. Chem.* 1993, **268:** in press.

Cystatin A

Ari Rinne

Synonyms	Epidermal SH-Proteinase Inhibitor, Acid cysteine proteinase inhibitor, Stefin A
Abbreviations	ACPI, CPI-A
Classifications	None
Description	It is a non-glycosylated protein comprising a single polypeptide chain of 98 aa residues. The polypeptide chain does not contain any disulfide loops. The most highly conserved part of aa sequence is QVVAG at the position 46–50. The other part which is constant in cystatins is Gly-4. The presence of Gly-4 is essential for the inhibitory activity. Cystatin A is stable at elevated temperatures (100 °C, 10 min.) and its inhibitory activity is not reduced after acid or alkali treatment (pH 2–11). It is present in squamous epithelia, dendritic reticulum cells and granulocytes. It is also found in human saliva, tears, seminal plasma/amniotic fluid and in the urine of uremic patients. Cystatin A belongs to family I of the cystatin superfamily of proteins. They all are inhibitors of cysteine proteinases of papain type, e.g. cathepsin B, H, L and S. They do not inhibit serine proteinases or aspartyl proteinases.
Structure	The tertiary structure has not been determined.
Molecular Weight	12,600 Da (gel electrophoresis) 11,006 Da (aa sequence)
Sedimentation Coeff.	Unknown
Isoelectric Point	4.6, 4.8 and 4.9
Extinction Coeff.	Unknown
Enzyme Activity	Unknown
Coenzymes/Cofactors	Unknown
Substrates	Unknown
Inhibitors	Unknown
Biological Functions	The biological function is not clear. Cystatin A is assumed to be the intracellular inhibitor of cysteine proteinases of papain type. The inhibition is expected to be tight-binding, reversible type. the enzyme-inhibitor complex has a molar ratio 1:1 for the interaction with both papain and cathepsin H. The K_i values (nM) for papain, human cathepsin B, human cathepsin H, human cathepsin L, bovine cathepsin S and bovine dipeptidyl peptidase I are 0.019, 8.2, 0.31, 1.3, 0.04–0.06 and 33, respectively. It is not capable to inhibit calpain.
Physiology/Pathology	Cystatin A may take part in squamous-epithelial differentiation and/or prevention of microbial invasion. Intracellular location: Cytosol of suprabasal squamous epithelial cells, dendritic reticulum cells, granulocytes and Hodgkin cells. The expression of cystatin A decreases in malign transformations.

Degradation	Cystatin A is susceptible to the proteolytical cleavage by cathepsin D, which id demonstrated to be cable to destroy this protein completely.
Genetics/Abnormalities	The gene of cystatin A is mapped to human chromosome 3 cen-q21. The cDNA contains 297 bp and shows a PstI polymorphism. Until now no pathological conditions have been described in which cystatin A is responsible.
Half-life	Unknown
Concentration	Epidermis 200 µg/g of tissue Tongue epithelium 110 Mouth mucosae 85 Bone marrow 8 Liver 5 Spleen 3 Lung 0.3
Isolation Method	By affinity chromotography using carboxymethylpapain and isoelectric focusing.
Amino Acid Sequence	Met-Ile-Pro-Gly-Gly-Leu-Ser-Glu-Ala-Lys-Pro-Ala-Thr-Pro-Glu-Ile-Gln-Glu-Ile-Val-Asp-Lys-Val-Lys-Pro-Gln-Leu-Glu-Glu-Lys-Thr-Asn-Glu-Thr-Tyr-Gly-Lys-Leu-Glu-Ala-Val-Gln-Tyr-Lys-Thr-Gln-Val-Val-Ala-Gly-Thr-Asn-Tyr-Tyr-Ile-Lys-Val-Arg-Ala-Gly-Asp-Asn-Lys-Tyr-Met-His-Leu-Lys-Val-Phe-Lys-Ser-Leu-Pro-Gly-Gln-Asn-Glu-Asp-Leu-Val-Leu-Thr-Gly-Tyr-Gln-Val-Asp-Lys-Asn-Lys-Asp-Asp-Glu-Leu-Thr-Gly-Phe.
Disulfides/SH-Groups	Contains no disulfides or SH-groups.
General References	Järvinen, M. *J. Invest. Dermatol.*, 1978, **71**: 114–118. Järvinen, M. and Rinne, A. *Biochim. Biophys. Acta* 1982, **708**: 210–217. Hopsu-Havu, V. K. et al. *Eur. Rev. Med. Pharmacol. Sci.* 1983, **V**: 1–4. Green, G. D. J. et al. *Biochem J.* 1984, **218**: 939–946. Abrahamson, M. et al. *J. Biol. Chem.* 1986, **261**: 11282–11287. Rinne, A. et al. *Anat. Anz. (Jena)* 1986, **161**: 215–230. Lenarcic, B. et al. *Biochem. Biophys. Res. Comm.* 1988, **154**: 765–722. Hsieh, W.-T. et al. *Nucleic Acids Res.* 1991, **19**: 1722. Brömme, D. et al. *Biomed. Biochim. Acta* 1991, **4–6**: 631–635. Eide, T. J. et al. *Acta histochem. (Jena)* 1992, **93**: 241–248.
Ref. for DNA/AA Sequences	Machleidt, W. et al. *Hoppe-Seyler's Physiol. Chem.* 1983, **364**: 1481–1486. Kartasova, T. et al. *Nucleic Acids Res.* 1987, **15**: 5945–5962. Rinne, R. et al. Upper Respiratory Tract Mucosa – Tromso, August 1994. Abstract p. 37. The nucleotide sequences data are available from EMBL Nucleotide Sequences Databases under the accession number X05978.

Cystatin B

Ari Rinne, Riitta Rinne and Mikko Järvinen

Synonyms	Neutral cysteine proteinase inhibitor, Stefin B
Abbreviations	NCPI, CPI-B
Classifications	None
Description	It is a non-glycosylated protein comprising a single polypeptide chain of 98 aa residues with acetylated amino terminus. The most highly conserved part of aa sequence is QVVAG at the position 46–50. The other part which is constant in all cystatins is Gly-4. Cystatin B tolerates pH values 2–11 but the activity is destroyed at temperatures over 70 °C. Cystatin B belongs to family I of the cystatin superfamily of proteins. They all are inhibitors of cysteine proteinases of papain type, e.g. cathepsin B, H, L and S, but the inhibitory capacity of cystatin B against cathepsin B is poor. The cystatins do not inhibit serine proteinases or aspartyl proteinases.
Structure	The crystal structure of cystatin B-papain complex has been elucidated at 2.4 Å resolution. Removal of a 7 aa aminoterminal peptide does not significantly reduce the inhibitory activity, while changing of Val 48 of the conserved QVVAG sequence to Asp causes a 240 fold lower affinity for papain.
Molecular Weight	12,000 Da (by gel electrophoresis), 11,175 Da (calculated from aa sequence).
Sedimentation Coeff.	Unknown
Isoelectric Point	6.0 and 6.5
Extinction Coeff.	Unknown
Enzyme Activity	Unknown
Coenzymes/Cofactors	Unknown
Substrates	Unknown
Inhibitors	Unknown
Biological Functions	The biological function is not clear but cystatin B is assumed to be an intracellular inhibitor of cysteine proteinases of papain type. The inhibition is expected to be tight-binding, reversible type. The K_i values (nM) for papain, human cathepsin B, human cathepsin H, human cathepsin L, bovine cathepsin S and bovine dipeptidyl peptidase I are 0.012, 73, 0.58, 0.23, 0.07 and 0.23, respectively.
Physiology/Pathology	Location: Cytosol of various cell types, especially of the histiocytic reticulum cells, macrophages and chondrocytes.
Degradation	Cystatin B is susceptile to the proteolytical cleavage by cathepsin D, which is demonstrated to be capable to destroy this protein completely.
Genetics/Abnormalities	A synthetic gene has been prepared and used for expression in *E. coli*.

Half-life	Unknown	
Concentration	Seminal plasma	1.0 mg/L
	Blood plasma	0.58 mg/L
	Other body fluids	<0.1 mg/L
Isolation Method	By affinity chromatography using carboxylmethylpapain, and isoelectric focusing from striated muscle (contains no cystatin A), spleen and liver.	
Amino Acid Sequence	MMCGA PSATQ PATAE TQHIA DQVRS QLEEK YNKKF PVFKA VSFKS QVVAG TNYFI KVHVG DEDFV HLRVF QSLPH ENKPL TLSNY QTNKA KHDEL TYF	
Disulfides/SH-Groups	It does not contain any disulfides. It contains one cysteine residue and is capable to form a dimer by disulfide formation.	

General References

Järvinen, M. and Rinne, A. *Biochim. Biophys. Acta* 1982, **708**: 210–217.
Green, G. D. J. et al. *Biochem J.* 1984, **218**. 939–946.
Abrahamson, M. et al. *J. Biol. Chem.* 1986, **261**: 11282–11287.
Rinne, A. et al. *Anat. Anz. (Jena)*, 1986, **161**: 215–230.
Lenarcic, B. et al. *Biochem. Biophys. Res. Comm.* 1988, **154**: 765–772.
Stubles, M. T. et al. *EMBO J.* 1990, **9**: 1939–1947.
Brömme, D. et al. *Biomed. Biochim. Acta* 1991, **4–6**: 631–635.
Machleidt, W. et al. *Biomed. Biochim. Acta* 1991, **4–6**: 613–620.

Ref. for DNA/AA Sequences

Ritonja, A. et al. *Biochem. Biophys. Res. Comm.* 1985, **131:** 1187–1192.
Thiele, U. et al. *Biol. Chem. Hoppe-Seyler* 1988, **369:** 1167–1178.

The nucleotide sequences data are available from EMBL Nucleotide Sequences Databases under accession number L 03558.

Cystatin C

Anders O. Grubb

Synonyms	γ-trace; Post-γ-globulin; Post-γ-protein; γ CSF: γaT; γc globulin
Abbreviations	None
Classifications	Electr. mob. γ_3 (pH 8.6)
Description	A non-glycosylated protein comprising a single polypeptide chain of 120 aa residues. It is present in all investigated body fluids with the highest concentration in seminal plasma and cerebrospinal fluid. It is synthesized by all investigated human cell lines and its mRNA encodes a hydrophobic leader sequence of 26 aa residues which is cleaved off at the secretion of the mature protein. Cystatin C belongs to family 2 of the cystatin superfamily of proteins. All proteins of this superfamily display sequence homologies and are inhibitors of cysteine proteinases of the papain type e.g. cathepsin B, cathepsin L, cathepsin S.
Structure	The tertiary structure is not yet determined but resembles most probably that of chicken egg white cystatin. The proline residue at position 3 of cystatin C is partly hydroxylated.
Molecular Weight	13,343 Da: non-hydroxylated form 13,359 Da: hydroxylated form (both deduced from aa sequences and verified by electrospray ionization mass spectrometry).
Sedimentation Coeff.	1.6 S
Isoelectric Point	9.3
Extinction Coeff.	9.1 (280nm, 1 %, 1cm) or 1.22×10^4 (mol^{-1} L cm^{-1}).
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Susceptible to inactivation by proteolytic removal of the N-terminal dekapeptide by neutrophil elastase. Inactivated by reduction of the two intrachain disulfide bridges.
Biological Functions	Inhibitor of cysteine proteinases of the papain type. Both mammalian, protozoan, viral and plant cysteine proteinases are inhibited. The equilibrium constants for dissociation of the cystatin C complexes with the human cysteine proteinases cathepsins B, H and L are 0.25, 0.28 and < 0.005 nmol L^{-1}, respectively. The association rate constants for the interactions between cystatin C and cathepsins B, H and L are 3.1×10^6, 24×10^6 and 140×10^6 mol^{-1} L s^{-1}, respectively.
Physiology/Pathology	The production of cystatin C does not increase in inflammatory conditions and the protein is thus not an acute phase protein. Cystatin C is mainly catabolized by renal glomerular filtration followed by tubular reabsorption and degradation of the protein. The plasma concentration of cystatin C is therefore almost exclusively determined by the glomerular filtration rate

for which it is an excellent indicator. Dysfunction of the proximal renal tubules results in impaired reabsorption of the cystatin C in the glomerular ultrafiltrate and the urinary concentration of cystatin C may increase more than hundredfold and is thus a sensitive indicator for the status of the renal tubules. Hereditary cystatin C amyloid angiopathy (HCCAA), also called "Hereditary cerebral hemorrhage with amyloidosis, Icelandic", is a dominantly inherited disease characterized by deposition of cystatin C amyloid in most tissues and particularly in the cerebral vasculture. The most serious consequence of the condition is the occurrence of multiple cerebral hemorrhages resulting in paralysis, dementia and death of affected family members before forty years of age. The disease is caused by a T→A mutation in the cystatin C gene resulting in a L→Q substitution in position 68 of the cystatin C polypeptide chain. Patients and carriers of the HCCAA-trait can be identified by their low cystatin C level in cerebrospinal fluid and by direct demonstration of the T→A mutation by a rapid polymerase chain reaction based analysis.

Degradation

Cystatin C may be completely proteolytically degraded in samples of urine and cerebrospinal fluid unless protected by the addition to the samples of suitable proteinase inhibitors e.g. benzamidinium chloride and EDTA.

Genetics/Abnormalities

The gene is on chromosome 20 at p11.2, spans about 4.3 kb and comprises 3 exons and 2 introns. The introns with 2252 and 1254 bp, respectively, are positioned between the nucleotide triplets encoding aa residues 55-56 and 93-94 of the mature protein. One T→A mutation causes the dominantly inherited disease hereditary cystatin C amyloid angiopathy. Several mutations in the promotor region have been described but do not seem to be associated with disease.

Half-life

2 min. (human cystatin C in the rat)

Concentration

Concentrations in body fluids of healthy adult persons:

	mean (mg L^{-1})	range (mg L^{-1})
blood plasma	1.1	0.6 - 1.5
cerebrospinal fluid	5.8	3.2 - 12.5
urine	0.095	0.033 - 0.29
saliva	1.8	0.36 - 4.8
seminal plasma	51.0	41.2 - 61.8
amniotic fluid	1.0	0.8 - 1.4
tears	2.4	1.3 - 7.4
milk	3.4	2.2 - 3.9

Isolation Method

(a) By ion-exchange and gel chromatography from urine of patients with tubular dysfunction or, for recombinant cystatin C, from *E. coli* periplasm.
(b) By affinity chromatography using carboxymethylpapain or antibody columns.

Amino Acid Sequence

SSPGK PPRLV GGPMD ASVEE EGVRR ALDFA VGEYN KASND
MYHSR ALQVV RARKQ IVAGV NYFLD VELGR TTCTK TQPNL
DNCPF HDQPH LKRKA FCSFQ IYAVP WQGTM TLSKS TCQDA

The primary structure has been established both at the protein, the cDNA and the genomic level. The following segments are involved in the inhibitory centre of cystatin C: 8-11 (RLVG), 55-59 (QIVAG) and 105-106 (PW). The cystatin C variant causing hereditary cystatin C amyloid angiopathy differs from wildtype cystatin C by carrying a glutamine residue at position 68.

Disulfides/SH-Groups	Cystatin C contains two intrachain disulfide bridges, one connecting Cys-73 with Cys-83 and the other connecting Cys-97 with Cys-117. No free sulfhydryl groups.
General References	Löfberg, H. and Grubb, A.O. *Scand. J. Clin. Lab. Invest.* 1979, **39**:619-626. Barrett, A.J. et al. *Biochem. Biophys. Res. Commun.* 1984, **120**:631-636. Grubb, A. et al. *N. Engl. J. Med.* 1984, **311**:1547-1549. Abrahamson, M. et al. *J. Biol. Chem.* 1986, **261**:11282-11289. Ghiso, J. et al. *Proc. Natl. Acad. Sci. USA* 1987, **83**:2974-2978. Abrahamson, M. et al. *J. Biol. Chem.* 1987, **262**: 9688-9694. Abrahamson, M. et al. *Hum. Genet.* 1992, **89**:377-380. Abrahamson, M., and Grubb, A. *Proc. Natl. Acad. Sci. USA* 1994, **91**:1416-1420. Kyhse-Andersen, J. et al. *Clin. Chem.* 1994, **40**:1921-1926.
Ref. for DNA/AA Sequences	Grubb, A. and Löfberg, H. *Proc. Natl. Acad. Sci. USA* 1982, **79**:3024-3027. Abrahamson, M. et al. *FEBS Lett.* 1987, **216**:229-233. Abrahamson, M. et al. *Biochem. J.* 1990, **268**:287-294. The nucleotide sequences data are available from the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X52255.

Cystatin D

Milagros Balbín and Anders O. Grubb

Synonyms	None
Abbreviations	None
Classifications	Electrophoretic mobility at pH 8.6: γ1
Description	A non-glycosylated protein comprising a single polypeptide chain of 118 aa residues. It is present in significant concentrations in saliva and tears but not in blood plasma, milk, seminal plasma or cerebrospinal fluid. Its mRNA encodes a preprotein of 142 aa residues containing a hydrophobic leader sequence. Two variants of the protein differing in one aa residue (Cys/Arg) have been identified. Cystatin D belongs to family 2 of the cystatin superfamily of proteins. All proteins of this superfamily display sequence homologies and are inhibitors of cysteine proteinases of the papain type e.g. cathepsin S, cathepsin H, cathepsin L.
Structure	The tertiary structure is not yet determined but resembles most probably that of chicken egg white cystatin.
Molecular Weight	13,536 (deduced from the aa sequence of the most extended form of the cysteine variant of mature cystatin D).
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coefficient	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Unknown
Biological Functions	Inhibitor of cysteine proteinases of the papain type. Both mammalian and plant cysteine proteinases are inhibited. The equilibrium constants for dissociation of the cystatin D complexes with the mammalian cysteine proteinases cathepsins S, H and L are 0.24, 8.5 and 25 nmol/L, respectively. In contrast to other family 2 cystatins, cystatin D does not seem to be an efficient inhibitor of cathepsin B (equilibrium constant for dissociation >1000 nmol/L).
Physiology/Pathology	Cystatin D is the first described member of a third subfamily of family 2 cystatins with the other two subfamilies being constituted by cystatin C and cystatins S, SA and SN, respectively. Cystatin D has a much more restricted tissue distribution than the other family 2 cystatins and an inhibitor spectrum unique among the cystatins. No pathophysiological condition connected to abnormalities in cystatin D structure or production has been described so far.
Degradation	Cystatin D may be completely proteolytically degraded in samples of saliva unless protected by the addition to the samples of suitable proteinase

inhibitors e.g. benzamidinium chloride and EDTA. Cystatin D isolated from saliva, even in the presence of proteinase inhibitors, displays a ragged N-terminus.

Genetics/Abnormalities

The gene spans about 4.3 kilobases and comprises 3 exons and 2 introns. The introns with 1826 and 1234 bp, respectively, are positioned between the nucleotide triplets encoding aa residues 53–54 and 91–92 of the most extended form of the mature protein. One T→C transition in exon 1 causes an aa variation, Cys/Arg, at the protein level, which does not seem to be connected to any variation in inhibition properties; nor to any pathophysiological process.

Half-life

Unknown

Concentrations

Concentrations in body fluids of healthy adult persons:

	mean (mg/L)	range (mg/L)
saliva	3.8	1.6–5.1
tears	0.5	<0.1–1.5
blood plasma	<0.1	
seminal plasma	<0.1	
milk	<0.1	
cerebrospinal fluid	<0.1	

Isolation Method

By immunosorption followed by ion-exchange chromatography from saliva or, for recombinant cystatin D, from *E. coli* periplasm.

Amino Acid Sequence

AQSRT LAGGI HATDL NDKSV QCALD FAISE YNKVI NKDEY YSRPL QVMAA YQQIV GGVNY YFNVK FGRTT CTKSQ PNLDN CPFND QPKLK EEEFC SFQIN EVPWE DKISI LNYKC RKV

The sequence given represents the most extended form of the cysteine variant of mature cystatin D. The primary structure has been established at the genomic level and, partly, at the protein level. The following segments are involved in the inhibitory center of cystatin D: 5–8 (TLAG), 53–57 (QIVGG) and 103–104 (PW). The arginine variant of cystatin D has an arginine residue at position 22.

Disulfides/SH-Groups

Cystatin D contains two intrachain disulfide bridges, one connecting the cysteine residue at position 71 with that at position 81 and the other connecting the cysteine residue at position 95 with that at position 115. The cysteine variant of cystatin D has probably a free sulfhydryl group at position 22.

General References

Freije, J. P. et al. *J. Biol. Chem.* 1991, **266**: 20538–20543.
Balbin, M. et al. *Hum. Genet.* 1993, **90**: 668–669.
Freije, J. P. et al. *J. Biol. Chem.* 1993, **268**: 15737–15744.
Balbin, M. et al. *J. Biol. Chem.* 1994, **269**: 23156–23162.

Ref. for DNA/AA Sequences

Freije, J. P. et al. *J. Biol. Chem.* 1991, **266**: 20538–20543.
Freije, J. P. et al. *J. Biol. Chem.* 1993, **268**: 15737–15744.
The nucleotide sequence data are available from the GenBankTM/EMBL Data Bank under the accession number X59964.

Cystatin S, SN, SA

Satoko Isemura

Synonyms

Salivary cysteine proteinase inhibitor

Abbreviations

Classifications

Classified as family II cystatin on the basis of homology with egg white cystatin and having a low molecular weight (around 14,000) and two disulfide bridges.

Description

Salivary proteins. Cysteine proteinase inhibitor. Having high degree of sequence homology among them. Each protein comprises single polypeptide chain of 121 aa residues with no carbohydrate side chain. Cystatin S is mostly phosphorylated at Ser-3 or both of Ser-1 and Ser-3.

Structure

None of the tertiary structures of cystatin S, SA and SN have been reported yet. However, the structure of a homologous protein, chicken cystatin is known. It has a tripartie wedge slotting into the active site cleft of papain, a cysteine proteinase.

Molecular Weight

14,184 : Cystatin S (non-phosphorylated form)
14,345 . Cystatin SA
14,311 : Cystatin SN

Sedimentation Coeff.

Isoelectric Point

	full-sized form	N-terminally truncated form
Cystatin S		4.7 (8)
Cystatin SA		4.3 (4)
Cystatin SN	6.8	7.5 (8)

(Numbers in parentheses indicate the residues truncated.)

Extinction Coeff.

Cystatin S (8)	2.0×10^4
Cystatin SA (4)	1.4×10^4
Cystatin SN (8)	1.8×10^4

(280 nm, 1M, 1 cm. Measured for the forms N-terminally truncated by the number of residues indicated in parentheses).

Enzyme Activity

None

Coenzymes/Cofactors

None

Substrates

None

Inhibitors

None

Biological Functions

Cysteine proteinase inhibitor. All three proteins inhibit papain, ficin and dipeptidyl peptidase (cathepsin C). None of them inhibit cathepsin B.

Physiology/Pathology

Unknown
Speculative function : protection of oral tissues from attack of proteinases of endogenous/exogenous origin. Anti-bacteria activity . Anti-virus activity. Stability (unpublished) pH and thermostable. Retains full inhibitory activity at room temperature in a wide range of pH (2 to 11) for more than a week. Retains full inhibitory activity after heating at 70 C for 30 min in phosphate buffer (pH 7).

Degradation	Unknown
Genetics/Abnormalities	The genes of cystatin S, cystatin SA and cystatin SN are members of a cystatin gene family located on chromosome 20. They are 3.5 kb long with three exons. Exons 1, 2 and 3 of each gene encode 76, 38 and 27 amino acids residues, respectively. Signal sequences are composed of 20 amino acids. A variant of cystatin SA has been reported.
Half-life	
Concentration	Whole saliva : cystatin S >6.7 mg/L cystatin SA >1.1 mg/L cystatin SN >4.4 mg/L Distribution: saliva, tear, seminal plasma.
Isolation Method	80% methanol fractionation (supernatant), followed by chromatography on DE32, DEAE Sepharose CL-6B and Sephacryl S200. N-teminal several residues may be cleaved off occasionaly during purification.
Amino Acid Sequence	cystatin S: SSSKEENRII PGGIYDADLN DEWVQRALHF AISEYNKATE DEYYRRPLQV LRAREQTFGG VNYFFDVEVG RTICTKSQPN LDTCAFHEQP ELQKKQLCSF EIYEVPWEDR MSLVNSRCQE A cystatin SA: WSPQEEDRII EGGIYDADLN DERVQRALHF VISEYNKATE DEYYRRLLRV LRAREQIVGG VNYFFDIEVG RTICTKSQPN LDTCAFHEQP ELQKKQLCSF QIYEVPWEDR MSLVNSRCQE A cystatin SN: WSPKEEDRII PGGIYNADLN DEWVQRALHF AISEYNKATK DDYYRRPLRV LRARQQTVGG VNYFFDVEVG RTICTKSQPN LDTCAFHEQP ELQKKQLCSF EIYEVPWENR RSLVKSRCQE S
Disulfides/SH-Groups	Cys(74)-Cys(84), Cys(98)-Cys(118); No free SH group.
General References	Isemura, S. et al. *J. Biochem.* 1991, **110**:648–654. Saitoh, E. et al. *Gene* 1987, **61**: 329–338.
Ref. for DNA/AA Sequences	

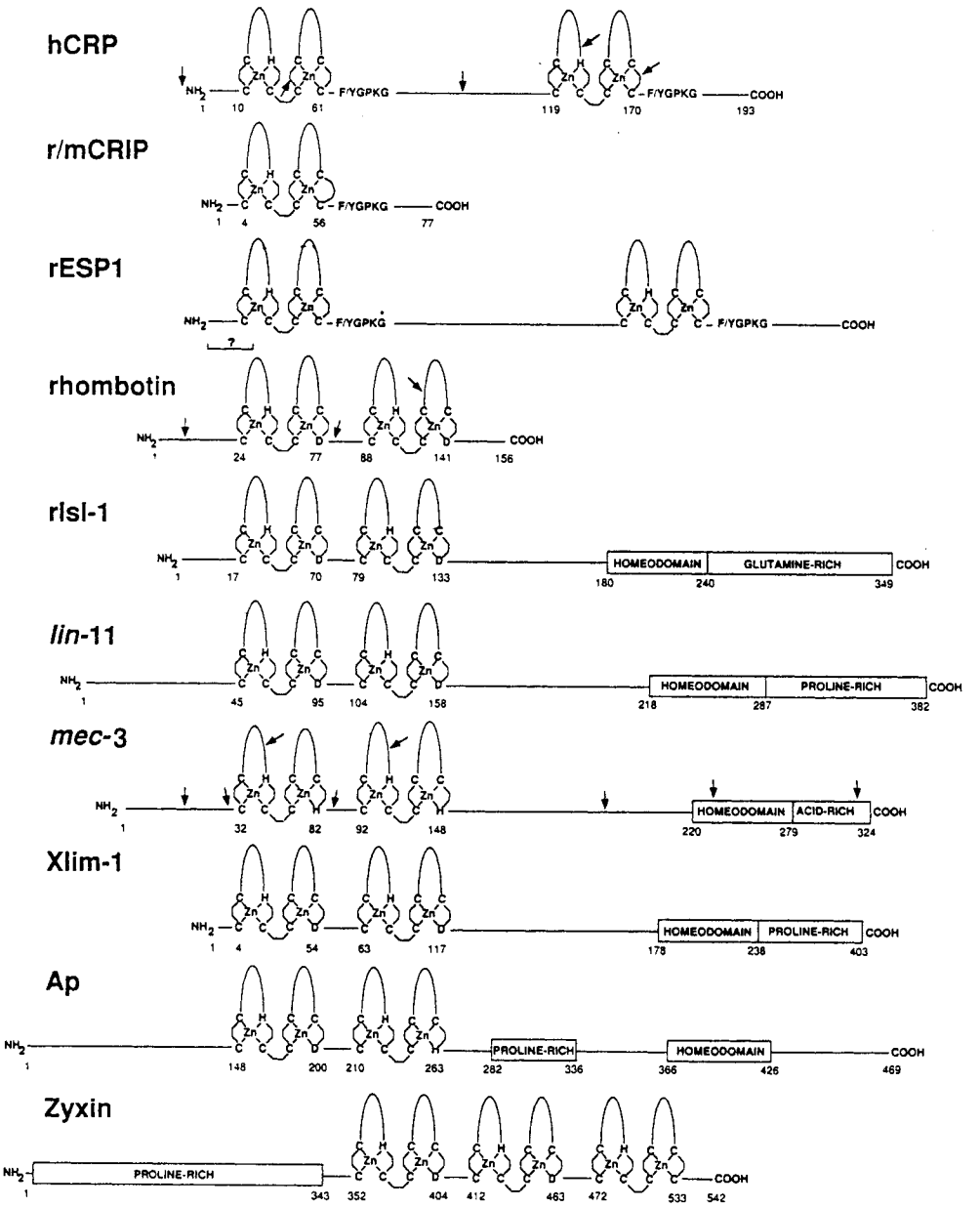
Cysteine-rich protein

Nancy E. Cooke and Stephen A. Liebhaber

Synonyms	None
Abbreviations	CSRP, CRP
Classifications	LIM/double zinc finger protein
Description	An intracellular, single chain protein synthesized in a wide variety of tissues. It contains 193 aa and chelates four moles of zinc with two sets of a double zinc finger motif. The first and third fingers are of the C_4 class, while the second and fourth are of the C_2HC class. These four finger domains appear to have evolved from duplication of a preexisting two finger unit.
Structure	Unknown
Molecular Weight	20,457 (calculated from aa sequence) 23,400 (estimated on SDS-PAGE)
Sedimentation Coeff.	Unknown
Isoelectric Point	10.38 (calculated from aa sequence)
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Induced as an immediate-early response protein in G0 cells stimulated with serum. The kinetics of its serum induction parallels that of the oncogene *c-myc*. The chick CRP homologue was purified from adhesion plaques of myocytes, and it is postulated that CRP may play a role in cell-cell interactions.
Physiology/Pathology	Unknown
Degradation	Unknown
Genetics/Abnormalities	The CRP gene is on chromosome 1 at q24-q32. The gene is 23.2 kb in length and contains 6 exons encoding an mRNA of 1.9 kb. RFLPs have been identified. The gene is highly conserved through vertebrate evolution, for example, containing 91 % sequence identity to the chick homologue. It is a member of an extended family of LIM domain containing proteins which includes developmentally critical proteins in *C. elegans*, a transactivator of insulin gene expression, and human rhombotin, a gene isolated from the breakpoint of a T cell leukemia, and others (see figure).
Half-life	Unknown
Concentration	Unknown

Isolation Method	The most highly enriched sources are: lung, brain, colon, spleen, thymus, placenta, and ovary. Co-purifies with the adhesion plaque protein zyxin.
Amino Acid Sequence	Consensus sequence for the LIM domain is as follows: first and third domains Cys-$(X)_2$-Cys-$(X)_{1\,7}$-His-$(X)_2$-Cys; second and fourth domains Cys-$(X)_2$-Cys-$(X)_{1\,7}$-Cys-$(X)_2$-Cys.
Disulfides/SH-Groups	Unknown
General References	Liebhaber, S. A. et al.*Nucl. Acids Res.* 1990, **18:** 3871–3879. Sadler, I. et al. *J. Cell Biol.* 1992, **119:** 1573–1587. Wang, X. et al. *J. Biol. Chem.* 1992, **267:** 9176–9184. Wang, X. et al.*Genomics* 1992, **14:** 391–397.
Ref. for DNA/AA Sequences	Liebhaber, S. A. et al.*Nucl. Acids Res.* 1990, **18:** 3871–3879. Wang, X. et al.*J. Biol. Chem.* 1992, **267:** 9176–9184. Genbank Accession Numbers: cDNA – M33146; gene – M76375-M76378

The LIM/Double Zinc Finger Family of Genes



Cytochrome C

Teresa J. T. Pinheiro and Anthony Watts

Synonyms	None
Abbreviations	None
Classifications	Electrophoretic Mob.: 9.52 $cm^2 s^{-1}$ V $\times 10^{-5}$ at 0 °C, and pH 6.9; Diffusion Coeff.: 13.0 cm^2 s^{-1} at 20 °C in water (equine protein).
Description	A heme *c*-containing protein found in mitochondria, comprising a single polypeptide chain of 104 aa. This soluble protein is folded into an almost globular shape around its heme prosthetic group. The heme is covalently attached to the polypeptide chain *via* thioether bonds to Cys-14 and Cys-17. His-18 and Met-80 form the axial ligands to the heme iron.
Structure	The tertiary structure is not determined for human cytochrome *c*, however the protein is highly homologous to the equine heart protein, for which X-ray crystal and NMR solution structures are available. Overall, it contains three major and two minor α-helix elements involving 47 residues (6–14, 49–54, 60–69, and 87–102), 2β-turns (type III, 3_{10}) (residues 14–17, and 67–70), and 4 β-turns (type II) (residues 21–24, 32–35, 35–38, and 75–78) interconnected by strands of polypeptide chain.
Molecular Weight	~12,300
Sedimentation Coeff.	1.83 $\times$ 10^{-13} s^{-1} ($S_{20,w}$) for the equine protein.
Isoelectric Point	pI 10.04 $\pm$ 0.04 (20 °C); net charge at pH 7.0, + 9.5 for ferric cytochrome *c* (Fe^{3+}).
Extinction Coeff.	29.5 mM cm^{-1} at 550 nm for the ferrous protein (Fe^{2+}).
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Cytochrome *c* functions in the mitochondrial electron transfer chain which leads to energy production in the form of ATP. It carries one electron from cytochrome bc_1 complex to cytochrome oxidase, by switching from heme Fe^{3+} to Fe^{2+} (271 $\pm$ 6 mV (E'_m) for free-NH_2 protein).
Physiology/Pathology	Cytochrome *c* is known to protect the ischemic myocardium during acute coronary occlusion.
Degradation	Unknown
Genetics/Abnormalities	The chromosome location of human cytochrome *c* gene has not been identified. The nucleotide sequence of a partial *c*DNA and three complete pseudogenes have been determined. Construction of the gene and its functional expression in *Sacchromyces cerevisiae* has been achieved. Heterogeneities in amino acid sequence of the protein extracted from various individuals indicated the possibility of genetic variance.

Half-life	Unknown
Concentration	~ 0.3 nmol per mg of mitochondrial protein.
Isolation Method	(a) Human hearts, obtained at autopsy, are kept frozen until used. (b) Thawed heart tissue extracted with weak acetic acid. (c) Purification by chromatography on Amberlite XE-64 resin.
Amino Acid Sequence	GDVEKG KKIFIM KCSQCH TVEKGG KHKTGP NLHGLF GRKTGQ APGYSY TAANKN KGIIWG EDTLME YLENPK KYIPGT KMIFVG IKKKEE RADLIAY LKKATNE
Disulfides/SH-Groups	No disulfide bonds and no free sulfhydryls. The two conservative cysteines in the whole aa sequence (Cys-14 and Cys-17) are involved in the thioether bonds to the heme group.
General References	Wallace, J. A. and Tanaka, Y. *J. Biochem. Tokyo* 1994, **115:** 693–700. Moore, G. R. and Pettigrew, G. W. *Cytochromes c: Evolutionary, Structural and Physiochemical Aspects*, Rich, A. (ed.) Springer-Verlag, Berlin, 1990.
Ref. for DNA/AA Sequences	Tanaka, Y. et al. *J. Biochem. Tokyo* 1988, **103:** 954–961. Matsubara, H. and Smith, E. L. *J. Biol. Chem.* 1963, **238:** 2732–2753.

Cytochrome *c* Structure: X-ray crystal of the equine protein, which is expected to match well human cytochrome *c*. The equine and human proteins are strikingly similar in many primary structural features: (a) *N*-acetylation and amino-terminal sequence, (b) position and attachment of the heme by thioether bridges, (c) distribution of basic and hydrophobic residues, and (d) complete homology of many long sequences. The two proteins, of identical residue length, differ in 12 residues which involves some replacements with similar residues and some with dissimilar residues. In the latter cases, it is likely that such residues play no major role in function or in theh protein conformation. The heme atoms are represented by balls, and the residues involved in its attachment to the polypeptide chain are shown dark. [From PDB *Equus caballus* heart cytochrome *c*]

Cytochrome C oxidase

Angelo Azzi

Synonyms	Atmungsferment, cytochrome c: O_2 oxidoreductase, cytochrome aa_3
Abbreviations	oxidase, aa_3, COX
Classifications	EC 1.9.3.1
Description	It is a inner membrane bound mitochondrial protein, which brings about the rapid oxidation of reduced cytochrome c by molecular oxygen. It is the terminal oxidase of the electron transport chain.
Structure	The enzyme from bovine heart (but not the human) has been crystallized in two dimension. Its shape is described as lop-sided Y with a length of about 110 Å, having the separated domains 40 Å apart and 55 Å long.
Molecular Weight	$\simeq$ 200,000 (from the sum of the single subunits)
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	$E_{(reduced-oxidized)} = 24\ mM^{-1}cm^{-1}$ (605 nm)
Enzyme Activity	$80\ s^{-1}$ (in presence of 0.1 % laurylmaltoside, 25 mM MOPS, 1 mM EDTA, pH 7.8). It should be noticed that in all species the activity of isolated cytochrome c oxidase depends on the type and amount of detergent employed.
Coenzymes/Cofactors	2 hemes a and 2 Cu ions are responsible for the electron transfer through the protein and the reduction of oxygen. A third copper may be involved in dimerization of the protein and a Mg ion is important for its structure
Substrates	Reduced cytochrome c from most pro- and eukaryotic sources.
Inhibitors	Cyanide, sulfite, azide and carbon monoxide are poisons able to tightly bind to the enzyme site responsible for the interaction with oxygen. Polylysine, spermine, spermidine and other polycations compete with cytochrome c binding.
Biological Functions	The protein is responsible for the transfer of the electrons from the penultimate element of the respiratory chain, cytochrome c, to molecular oxygen. Part of the free energy produced in this reaction is conserved as a transmembrane protonmotive force (electric potential and proton gradient) created by the activity of the proton pump associated with electron transfer.
Physiology/Pathology	The major goal of the protein is that of catalyzing the energy coupled flow of electrons derived from reduced cell substrates to molecular oxygen, with conservation of energy used for the synthesis of ATP. Cytochrome oxidase deficiencies are associated with myopathies, Leigh syndrome, Myoclonus epilepsy with ragged red fibers (MERRF), progressive external ophthalmoplegia (PEO), Kearns-Sayre syndrome (KSS)
Degradation	Mechanism and pathways are not known

1. Installment 1992

Genetics/Abnormalities	Some genetic defects are autosomal recessive and tissue specific. They are due to alterations of some of the 10 nuclear genes coding for the subunits 4 to 13. Kearns-Sayre syndrome is instead associated with deletions of mitochondrial DNA coding for subunits 1 to 3. Cases are reported showing tissue specificity of the abnormalities.
Half-life	Unknown
Concentration	The concentration of the enzyme varies from tissue to tissue being of the order of 0.2 nmoles/mg of mitochondrial protein.
Isolation Method	Detergent (cholate, deoxycholate, Triton X-100)/salt fractionation procedures are commonly used. Affinity chromatography on cytochrome c matrices and ion exchange HPLC (Mono-Q, Pharmacia) are also employed especially for small scale preparations. The best sources for the purification are heart and liver.
Amino Acid Sequence	The 13 subunits of the human enzyme are homologous with other eukaryotes. The mitochondrially coded subunits (called 1–3 or I–III) are also homologous in prokaryotes. Tissue specificity exists for some of the nuclear coded subunits (called 4–13 or IV–VIII with Va,b; VIa,b,c; VIIa,b,c)
Disulfides/SH-Groups	
General References	Azzi, A. and Müller, M., *Arch. Biochem. Biophys.* 1990, **280:** 242–251. Sinjorgo, K. M. C., et al. *Biochim. Biophys. Acta* 1987, **850:** 144–150. Wikström, M. et al. *Cytochrome Oxidase.* Academic Press, London, New York, Toronto, San Francisco, 1981.
Ref. for DNA/AA Sequences	Subunit sequence accession numbers in NBRF(*) and genEMBL libraries and references: I = *A00463 Anderson et al. *Nature* 1981, **290:** 457 (also for subunits II and III); II = *A00472: Power et al. *Nucleic Acid Res.* 1989, **17:** 6734. X15759, J01416 III = *A00482; IV = M21575: Zeviani et al. *Gene* 1989, **55:** 205. M34600: Lomax et al (1990) Gene 86, 209. Va = *JT0342: Rizzuto et al. *Gene* 1988, **69:** 245. Vb = *JT0324,*A28817, M19961: Zeviani et al. *Gene* 1988, **65:** 1. VIa = *S05304: Fabrizi et al. *Nucleic Acid Res.* 1989, **17:** 6409. X15341; VIb = *S03287: Taanman et al. *Nucleic Acid Res.* 1989, **17:** 1766; X13923; VIc = *S01960: Otsuka et al. *Nucleic Acid Res.* 1988, **16:** 10916; X13238; VIIa = *S06897: Fabrizi et al. *Nucleic Acid Res.* 1989, **17:** 7107; X15822; VIIb = not available; VIIc = *S08217: Koga et al (1990) Nucleic Acid Res. 18, 684; X16560; VIII = *A34103: Rizzuto et al. (1989) *J. Biol. Chem.* 1989, **264:** 10595; S01785: van Kuilenburg et al. *FEBS Lett.* 1988, **240:** 127; J04823

Decay accelerating factor

Douglas M. Lublin

Synonyms	CD55; Cromer blood group antigen
Abbreviations	DAF
Classifications	Regulators of Complement Activation-RCA-Family
Description	A single-chain polypeptide membrane protein that is widely distributed on all hematopoietic cells and on epithelial and endothelial cells. The bulk of the protein consists of 4 repeats of approx. 60 aa termed short consensus repeats (SCR), characterized by four conserved cysteines and an overall homology of approx. 30%. There are several important posttranslational modifications: (i) one complex, N-linked oligosaccháride unit attached to Asn-61; (ii) multiple O-linked oligosaccharide units attached to the Ser/Thr-rich domain between aa 253-319; and (iii) a glycosyl-phosphatidyl-inositol (GPI) anchor attached to Ser-319 (with the removal of the C-terminal aa 320-347). The GPI anchor provides the only membrane attachment. Minor species of DAF include a high molecular weight form that appears to be a covalent DAF dimer, as well as low concentrations of a soluble form of DAF in plasma and other body fluids.
Structure	The tertiary structure has not been determined, but NMR analysis of the solution structure of homologous SCR domains from the complement protein factor H demonstrates that an SCR unit consists of 5 β-strands separated by turns, with hydrophobic side-chains of the strands contributing to a compact core. The overall length of one SCR unit is 3.8 nm.
Molecular Weight	74,000 (SDS-PAGE), 41% carbohydrate; tissue-specific variations in molecular weight between 70,000 and 84,000, likely due to differences in glycosylation.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	No established physiological inhibitors. Several enzymes can remove DAF from the plasma membranes of cells, including papain and PI-specific phospholipase C.
Biological Functions	Inhibitor of the complement cascade. DAF regulates the central enzyme of the complement system, C3 convertase, which is a non-covalent, bimolecular complex. DAF accelerates the temperature-dependent dissociation (decay) of the two components of the C3 convertase enzyme (specifically, dissociating C2a from C4b2a or Bb from C3bBb). By blocking the complement cascade at the C3 step, DAF protects cells from damage caused by the downstream effector arms of the complement system, including genera-

tion of anaphylatoxins, chemotaxins, and the membrane attack complex of complement that can lyse cells.

Physiology/Pathology

DAF protects cells and tissues of the body from destruction brought about by the complement system. This is necessary to prevent complement-mediated damage to bystander tissues during the constant low-level activation of the complement system. Rare individuals with no DAF expression show a moderate increase in the sensitivity of their cells to complement-mediated cytotoxicity in vitro, although no clear in vivo pathology has been demonstrated. This is probably because of the existence of several other complement regulatory proteins that provide overlapping protection to cells and tissues. The potential therapeutic use of recombinant DAF to prevent inflammation has been demonstrated by recent experiment results. Xenotransplants (transplants of organs between different species) that would normally be destroyed within hours by a complement-mediated hyperacute rejection process have prolonged survival in transgenic animals expressing human DAF.

Degradation

DAF levels on erythrocytes decrease only modestly during the 100 day life-span of these cells. A low concentration of DAF in plasma and other body fluids might arise by an uncharacterized pathway of degradation from the cell membrane.

Genetics/Abnormalities

The 40 kb gene is on the long arm of chromosome one, band q3.2, part of the Regulators of Complement Activation gene family. The DAF gene comprises 11 exons, with the major DAF transcript that encodes membrane DAF composed of ten exons. (An additional, alternatively spliced exon arises in a minor class of DAF transcript, causing a shift in the reading frame that results in a non-membrane DAF, but it is uncertain whether this form of DAF is secreted or physiologically relevant.) There are common restriction fragment length polymorphisms involving the enzymes HindIII and BamHI, and these map to introns of the DAF gene. Several rare genetic variants of DAF arise from single point mutations in exons, and all produce functional protein except the Inab phenotype in which a mutation introduces a new stop codon and leads to a failure to express any surface DAF.

Half-life

The half-life of DAF in tissue culture cells is approximately 7 hours. The half-life of DAF in nucleated cells in vivo is not known.

Concentration

Copies per cell: erythrocytes, 3,000; granulocytes, 85,000; monocytes, 68,000; B cells, 54,000; T cells, 9,000-17,000; platelets, 2,000; HeLa epithelial cell line, 250,000. Concentrations of soluble DAF in body fluids range from 40-400 $\mu g\ L^{-1}$.

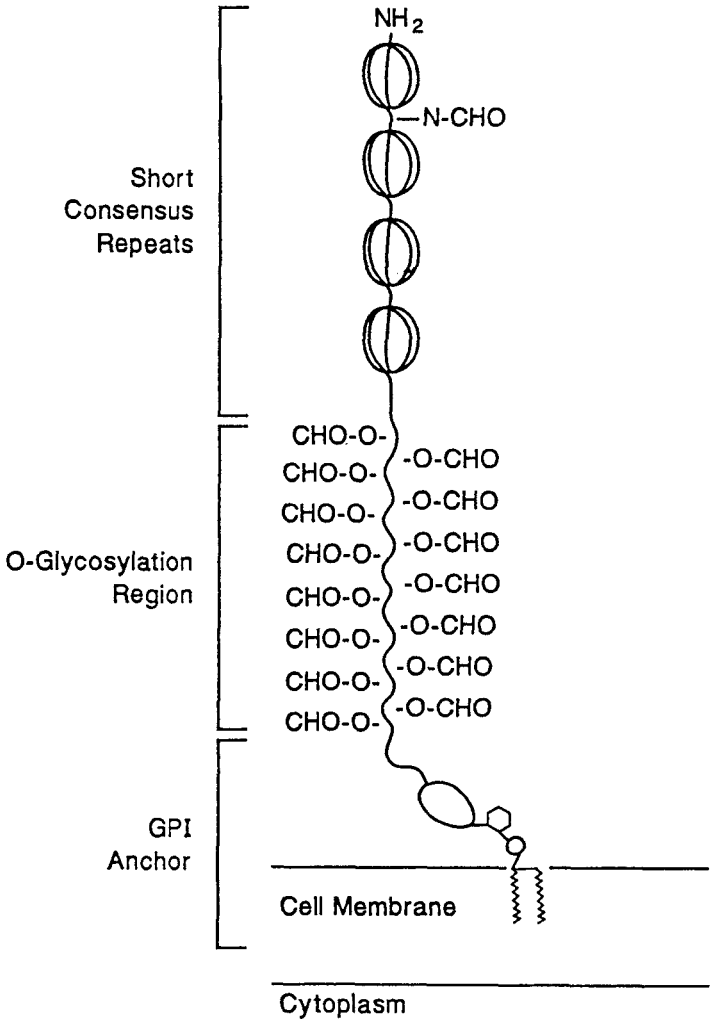
Isolation Method

(i) Butanol extraction of erythrocyte stroma followed by sequential chromatography on DEAE-Sephacel, hydroxylapatite, phenyl-Sepharose, and trypan blue Sepharose.
(ii) Non-ionic detergent lysate of cells isolated by affinity chromatography on anti-DAF monoclonal antibody column.

Amino Acid Sequence

SCR-1 of DAF, aa 1-62: DCGLPPDVPN AQPALEGRTS FPEDTVITYK CEESFVKIPG EKDSVICLKG SQWSDIEEFC NR (SCR domains with approx. 30% homology are found in all members of the Regulators of Complement Activation gene family as well as in non-complement proteins such as the IL-2 receptor and selectins. Signal for GPI anchor attachment, aa 319-347: SGTTRLLSGH TCFTLTGLLG TLVTMGLLT.

Disulfides/SH-Groups	There are 8 intrachain disulfide bonds that utilize the 4 invariant Cys residues in each of the 4 SCR domains. The specific bonds are formed as Cys-1-Cys-3 and Cys-2-Cys-4.
General References	Lublin, D.M. and Atkinson, J.P. *Ann. Rev. Immunol.* 1989, **7**:35-58. Nicholson-Weller, A. *Current Topics in Microbiol. Immunol.* 1992, **178**:7-30.
Ref. for DNA/AA Sequences	Medof, M.E. et al. *Proc. Natl. Acad. Sci. USA* 1987, **84**:2007-2011. Caras, I.W. et al. *Nature* 1987, **325**:545-549.



Model of decay accelerating factor (DAF). This model highlights the main features of the molecule: 4 short consensus repeat (SCR) domains with a β-strand structure, one attached N-linked oligosaccharide (N-CHO), a heavily O-glycosylated (O-CHO) region with a linear rod-like shape, and a glycosyl-phosphatidylinositol (GPI) membrane anchor.

Dopamine beta-hydroxylase

Annie Lamouroux and Jacques Mallet

Synonyms	Dopamine beta-monoxygenase.
Abbreviations	DBH, DβH
Classifications	EC 1.14.17.1
Description	The dopamine beta-hydroxylase catalyses the conversion of dopamine to norepinephrine within the catecholamine secreting vesicles (chromaffin granules) of the adrenal medulla and the large densed cored synaptic vesicles of the sympathetic nervous system. In the central nervous system DBH is expressed in the locus coeruleus. The enzyme exists in similar amounts both as a soluble and a membrane-bound protein.
Structure	DBH is a tetrameric copper-containing glycoprotein. Two subunits are joined by disulfide bonds to form a dimer, and two dimers are associated by noncovalent interactions to form the tetramer.
Molecular Weight	Tetramer: 300,000 Da (Sucrose gradient ultracentrifugation and analytical gel filtration chromatography). Dimer: 123,000 ($\pm$ 14,000) (radiation inactivation). Monomer: 75,000 (SDS-PAGE). 67,527 (calculated from the deduced nucleotidic sequence).
Sedimentation Coeff.	10 S
Isoelectric Point	5.4–6.8 (experimental range)
Extinction Coeff.	Unknown
Enzyme Activity	DBH is a monooxygenase enzyme that catalyzes hydroxylation of a phenylethylamine (dopamine in vivo) at the beta-C of the side chain to a phenylethanolamine (norepinephrine in vivo) using molecular oxygen and ascorbic acid.
Coenzymes/Cofactors	Ascorbate serves as an electron donor. Cu^{2+} is a bound cofactor (electron accepting group).
Substrates	The biological substrate is dopamine. In vitro both dopamine and tyramine are used.
Inhibitors	Sulfhydril compounds (cysteine, glutathion and coenzyme A) inhibit DBH as a result of their chelating action on copper. Fusaric acid is one of the most potent and specific inhibitors. It is a copper chelator, but it also inhibits DBH competitively with ascorbate. It has a hypotensive activity. Substrate analogs, such as benzyloxyanine or benzylhydrazine, are competitive inhibitors with the dopamine substrate. Benzyl cyanide analogs are suicide inactivators. Hydrogen peroxyde (H_2O_2) inhibits DBH.
Biological Functions	DBH is an important enzyme in the catecholamine pathway. It converts dopamine into noradrenaline. DBH is localized in catecholamine storage vesicles in nerve endings of noradrenergic and adrenergic neurons and in adrenal medulla chromaffin cells, and is secreted, together with noradrenaline and/or adrenalin, by a process of exocytosis to finally appear in the cerebrospinal fluid (CSF) from the brain and in the serum from the

peripheral sympathetic nerves and the adrenal medulla. DBH is considered to be a biochemical index of the activity of catecholamine containing cells.

Physiology/Pathology

Several cases of DBH deficiency in humans have been reported where DBH is undetectable in plasma and cerebrospinal fluid. This autonomic disorder is characterized by lifelong severe orthostatic hypotension, ptosis, nasal stuffiness, hyperextensible joints and retrograde ejaculation. DBH level is increased in pheochromocytoma tumors, as well as tyrosine hydroxylase level.

Degradation

Unknown

Genetics/Abnormalities

The DBH gene is located on chromosome 9q34. Several restriction fragment length polymorphism (RFLP) have been described using a DBH cDNA (Xba I, Taq I, BamH 1). They are presently investigated in a number of various neuropsychiatric disorders such as schizophrenia and manic-depression. The DBH gene is also a good candidate gene for being implicated in idiopathic torsion dystonia, a neurological disorder recently located on chromosome 9 q 32–34.

Half-life

Unknown

Concentration

The DBH concentration in human serum is about 30 mg/L.

Isolation Method

The best source of DBH is pheochromocytoma tumors or adrenal glands. The principle of purification is based on the isolation of pure chromaffin granules from fresch tissue using a sucrose density gradient and subsequent solubilization and purification by DEAE-cellulose and Sephadex – G 200 chromatographies. Use of concanavalin A – sepharose column or affinity chromatography using tyramine-Sepharose is also an effective purification procedure. Human plasma DBH can be purified to homogeneity by con A-Sepharose, Ocytyl-Sepharose, Biogel A-0.5m and isoelectric focusing.

Amino Acid Sequence

The polypeptide chain of DBH comprises 578 aa and is preceeded by a signal peptide of 38 aa (deduced from the genomic analysis of DBH gene). The DBH exists both as a soluble and membrane-bound form. The hydropathy plot reveals no obvious hydrophobic segment, except the signal peptide. The present hypothesis is that DBH is anchored to the membrane by its signal peptide. DBH is homologous to the peptide alpha-amidating enzyme (PAM), a copper containing monooxygenase that uses ascorbate as a cofactor. This homology is located in the central part of the DBH (aa 190–485) and the N-terminal region of the PAM. In particular, two histidine clusters are conserved at position 250 and 398 in human DBH providing some indications about the fixation sites of Cu^{2+} and the location of the active site.

Disulfides/SH-Groups

Interchain: two disulfide bounds/tetramer. Intrachain: unknown.

General References

Nagatsu, T. In: *Neuromethod 5: Neurotransmitter enzymes*. Humana Press, New Jersey, 1986, pp. 79–115.
Stewart, L. C., et al. *Ann. Rev. Biochem.* 1988, **57:** 551–592.

Ref. for DNA/AA Sequences

Lamouroux, A., et al. *EMBO J.* 1987, **6:** 3931–3937.
Kobayashi, K., et al. *Nucleic Acids Res.* 1989, **17:** 1089–1102.
Craig, S. P., et al. *Cytogenet. Cell. Genet.* 1988, **48:** 48–50.
The nucleotide sequences are accessible in the EMBL Data Bank under the mnemonic HSDBHRA and HSDBHRB.

Endopeptidase-24.11

A. John Kenny

Synonyms	Neutral endopeptidase, Kidney brush border neutral peptidase, Enkephalinase, Common acute lymphoblastic leukaemia antigen, Cluster differentiation antigen 10
Abbreviations	E-24.11, CALLA, CD-10
Classifications	EC 3.4.24.11
Description	A cell surface peptidase present on many different cell types. Its topology is typical of other ectoenzymes, i.e. a glycoprotein comprising a globular head, a stalk, a hydrophobic peptide anchor and a short cytoplasmic tail. For this enzyme, the stalk is short, the anchor is the uncleaved signal sequence and the N-terminus is at the cytoplasmic surface of the membrane. The protein has 15%-25% carbohydrate, the quantity and pattern of glycosylation varies with the tissue source (e.g. kidney, intestine, brain, leukaemic cells). This endopeptidase has a very wide distribution on the surface of many cell types. It is, however, not ubiquitous, being absent, for example from vascular endothelial cells and having a very distinct location within the CNS. The main sites are: Epithelia (brush borders): renal proximal tubule, enterocyte, syncytial trophoblast, choroid plexus. Epithelia (other types): cells lining reproductive tract (male and female). Myoepithelial cells: mammary gland, mucous glands, etc. Steroid endocrine cells: Leydig cells (testis), interstitial cells (ovary), zona glomerulosa (adrenal): Other endocrine cells: gonadotrophs (adenohypophysis). Connective tissues: fibroblasts (human skin in culture), chondrocytes (articular cartilage). Central nervous system: some neurons in corpus striatum. Peripheral nervous system: Schwann cells. Blood: pre B cells (bone marrow), $CALLA^+$ cells (common acute lymphocytic leukemia), polymorphonuclear leukocytes.
Structure	Homodimeric glycoprotein (in most species). Subunits are noncovalently linked. May be isolated, after solubilization by detergent, as an amphipathic protein or after proteinase treatment as a hydrophilic protein.
Molecular Weight	Subunits 89,000–110,000 (SDS PAGE, range relating to differences in glycosylation). Native enzyme by gel filtration, 216,000 (when released from membrane by trypsin), 320,000 (when released by Triton X-100 and including a detergent micelle).
Sedimentation Coeff.	5.77 S (rabbit enzyme)
Isoelectric Point	4.9–5.45
Extinction Coeff.	Unknown
Enzyme Activity	A metallo-endopeptidase containing one zinc atom/subunit. It hydrolyses a wide range of peptides where the P_1' residue is hydrophobic. An optimum sequence has been suggested: Phe-Phe-Ala-Phe-Leu-Ala- (with cleavage at the Ala-Phe bond).
Coenzymes/Cofactors	Zn^{2+} is bound at the active site, but there is no requirement for additional Zn^{2+}, unless the activity has been reduced by treatment with metal-chelating agents.
Substrates	Natural substrates are neuropeptides, with tachykinins being among the best substrates (e.g. substance P, $k_{cat}/Km = 159\ min^{-1}\ \mu M^{-1}$, or

	enkephalins (e.g. [Met]enkephalin, $k_{cat}/Km = 42\ min^{-1}\ \mu M^{-1}$. It may be assayed radiometrically with [^{125}I]insulin β chain or [Tyrosyl-3,5-^{3}H-Leu^{5}]-enkephalin. Fluorometric assays use either an internally quenched peptide (e.g. AAGLAN, 2-aminobenzoyl-Ala-Gly-Leu-Ala-nitrobenzylamide) or, in two steps, with glutaryl-Ala-Ala-Phe-2-naphthylamide as substrate; aminopeptidase being required to hydrolyse the Phe-2-naphthylamide. So far the most efficiently hydrolysed substrates are simple neuropeptides and peptide hormones. Although the chains of insulin are degraded, intact insulin is wholly resistant to hydrolysis. Its roles and its target molecules at various sites are not at present fully defined, though in general it may act (a) to terminate peptide signals (e.g brain), (b) to protect cells by clearance of unwanted peptide signals (e.g. choroid plexus), and (c) scavenging for nutritional purposes (e.g intestine).
Inhibitors	Phosphoramidon (a natural product from Streptomyces tanashiensis) and a number of synthetic compounds, e.g. thiorphan and retrothiorphan. All have Ki values in the range 2–10 nM and are reversible, competitive inhibitors.
Biological Functions	The enzyme is present on a diverse range of cell surfaces and it is clear that the physiological targets will vary with the location. In the central nervous system, neuropeptides such as tachykinins and enkephalins are probable targets, the endopeptidase having the role of terminating peptide signals at synapses. Atrial and brain natriuretic peptides are also probable targets. The possibility that larger molecules, e.g. cytokines, might be targets, has evoked recent interest. In vitro, interleukin-1β can be hydrolysed slowly, but the physiological significance of this observation has yet to be evaluated. Transforming growth factor alpha is resistant to hydrolysis by this enzyme as is IL-1α and tumour necrosis factor.
Physiology/Pathology	Since the precise target peptides in the nervous system have yet to be defined, the physiological role in the brain remains somewhat speculative. There is some evidence that endopeptidase-24.11 may have a pathophysiological role in the circulation by inactivating natriuretic peptides thus tending to increase blood pressure and the retention of NaCl. Hence inhibitors of the enzyme are currently on trial in hypertension and congestive heart-failure. Serum levels of endopeptidase-24.11 have been reported to be elevated in two pathological states: sarcoidosis and the adult respiratory distress syndrome (ARDS) and in some cases of acute lymphoblastic leukaemia.
Degradation	Unknown
Genetics/Abnormalities	No genetic abnormality reported.
Half-life	Unknown
Concentration	In renal brush border membranes, the protein may be 4% of the total protein. In striatal synaptic membranes, it is two orders of magnitude less. It is detectable in plasma.
Isolation Method	Purification (e.g. from kidney or small intestine) requires solubilization from the membrane. Proteinases are relatively ineffective and a non-ionic detergent, such as Triton X-100, is a necessary first step in treating the membrane fraction. Purification may be completed by conventional chromatographic steps, but immunoaffinity purification using a monoclonal antibody offers major advantages.

Amino Acid Sequence	749 aa residues in sequence deduced from cDNA cloning. 6 potential N-glycosylation sites exist. A consensus Zn-binding motif -H-E-X-X-H- in residues 583–587 is present, in common with other metallopeptidases.
Disulfides/SH-Groups	12 half-cystines in predicted sequence.
General References	Kerr, M. A., and Kenny, A. *J. Biochem. J.* 1974, **137:** 477–495. Kenny, A. J. et al. *Biochem. J.* 1983, **211:** 755–762. Matsas, R. et al. *Biochem. J.* 1984, **223:** 435–440. Kenny, A. J., and Stephenson, S. L. *FEBS Lett.* 1988,**232:** 1–8. Barnes, K. et al. *Neuroscience* 1988, **27:** 799–817.
Ref. for DNA/AA Sequences	Malfroy, B. et al. *Biochem. Biophys. Res. Commun.* 1987, **144:** 59–66. Devault, A. et al. *EMBO J.* 1987, **6:** 1317–1322. Shipp et al. *Proc. Natl. Acad. Sci. USA* 1988, **85:** 4819–4822. Letarte et al. *J. Exp. Med.* 1988, **168:** 1247–1254.

Erythrocyte Acid Phosphatase

Robert L. Van Etten

Synonyms	Red cell acid phosphatase; Low molecular weight phosphotyrosyl protein phosphatase; Human cytoplasmic phosphotyrosyl phosphatase
Abbreviations	RCAP; EAP; ACP1; Low M ,PTPase; HCPTP-A and -B
Classifications	EC 3.1.3.2; phosphomonoesterase, "acid" optimum
Description	A cytoplasmic phosphomonoesterase enzyme. As isolated from tissues, the major isoenzymes consist of 157 residues. The N-terminal residue is alanine, which is blocked by acetylation as the initiating methionine is removed during post-translational processing. The protein contains no other covalent modification, and no metal ion, coenzyme or cofactor is needed for enzymatic activity. Although known historically as red cell acid phosphatase, recent mRNA hybridization experiments make it clear that this enzyme is synthesized in all human tissues. Numerous ACP1 phenotypes can be separated and roughly quantitated by electrophoretic methods, and this has led to a major role for this enzyme as a genetic marker in population genetics and in forensic studies.
Structure	The structure of the highly homologous bovine enzyme (93% sequence identity) has been solved by high resolution X-ray crystallography and, in solution, by multidimensional, multinuclear NMR spectroscopy. Unpublished data from the author's laboratory confirms that the human isoenzyme structures are virtually identical. The protein consists of a four-stranded central parallel β sheet with α-helices flanking both sides. The overall fold resembles a classic dinucleotide binding or Rossmann fold, with two right-handed βαβ motifs. The phosphate binding site is formed from main chain NH-groups of residues 13-18, as well as the positively charged arginine side chain of Arg-18. Residue Cys-12 acts as a nucleophile, with the intermediate formation of a cysteinyl phosphate covalent intermediate, while Asp-129 acts as a proton donor to protonate the leaving alcohol or phenol group of the substrate.
Molecular Weight	For HCPTP-B (ACP1-slow), 17,888 ± 8 (electrospray mass spectroscopy), 17,889 (calculated from sequence); for HCPTP-A, 17,953 (calculated from sequence).
Sedimentation Coeff.	Unknown
Isoelectric Point	HCPTP-A (fast) 5.9; HCPTP-B(slow) 6.1
Extinction Coeff.	For HCPTP-A: 0.818 (280nm, 1%, 1cm) and ε = 14,650; for HCPTP-B: 0.996 (280nm, 1%, 1cm) and ε = 17,780.
Enzyme Activity	The enzyme exhibits phosphomonoester hydrolase activty against a number of unhindered phosphate esters, including ones with activated leaving groups such a *p*-nitrophenyl phosphate (pNPP) and tyrosine phosphate, and ones with a phosphate ester function bound through a spacer arm to a lipophilic molecule, such as flavin mononucleotide (FMN). At low, fixed substrate concentrations the activity exhibits a prominent acidic pH optimum centered around pH 4-5, but this is largely the result of a decreases in K_m. When high concentrations of pNPP are used to assay the enzyme, a broad flat plateau extending from pH 3 to 7 is seen for V_{max}

Coenzymes/Cofactors	None
Substrates	Routine assays of the enzyme in the laboratory are most conveniently done with the chromogenic substrate pNPP at pH 5, followed by quenching of the reaction mixture in strong base. Phosphate release assays can be conducted by a variety of means including malachite green colorimetric procedures.
Inhibitors	The enzyme is subject to strong (micromolar K_i) competitive inhibition by vanadate ion, which forms a trigonal bipyramidal, transition state analog at the active site. Other common competitive inhibitors include phosphate, arsenate and arsenite. Importantly, the enzyme is not significantly inhibited by L-(+)-tartrate, which is a strong inhibitor of the distinct lysosomal and prostatic acid phosphatases. The enzyme is inhibited and inactivated by phenylarsine oxide, in a slow, tight-binding reaction. It is also sensitive to sulfhydryl reagents such as iodoacetate and N-ethyl maleimide, which block the essential Cys-12 nucleophile.
Biological Functions	At this time, the biological substrate(s) for this enzyme has not been definitively established. FMN is a possible cellular substrate, and the enzyme may serve to regulate cytoplasmic flavin levels. The enzyme has also been shown to dephosphorylate phosphotyrosyl residues of the cytoplasmic domain of red cell band 3. The enzyme can also associate with and dephosphorylate certain growth factor receptors, at least *in vitro*, but the fact that homologous enzymes are found in yeast and even bacteria raises some question about the generality of such a role.
Physiology/Pathology	Elevated levels have been associated with megaloblastic anemia. The incidence and severity of favism in glucose 6-phosphate dehydrogenase-deficient individuals may be associated with ACP1 isoenzyme variations.
Degradation	Unknown
Genetics/Abnormalities	Because of the ease with which certain electrophoretic phenotypes could be identified in blood samples, ACP1 gene products have been the subject of extensive population surveys. Maps detailing the world distribution of some of the gene polymorphisms are readily available (see reference list). A link to favism has been suggested, and this may involve the role of the enzyme in dephosphorylating red cell band 3 and consequent effects on glycolysis.
Concentration	The yield of homogeneous enzyme on purifcation from human placenta is 2 mg/kg of tissue, while purification of the enzyme from red cells yields approximately 1 mg/L of packed cells. From mRNA hybridization experiments, the enzyme content of all tissues, including adult and fetal organs, may be expected to be roughly comparable.
Isolation Method	The two major human isoenzyme forms HCPTP-A and -B (corresponding to ACP1-fast and -slow), have been cloned, expressed and purified from bacterial expression systems in high yield, so this is clearly the method of choice for *in vitro* studies. The cellular enzyme, which differs only in that it has a blocked N-terminal, has been isolated from red cells, placenta and liver by typical procedures involving $(NH_4)_2SO_4$ precipitation, one or two ion exchange chromatography steps, gel permeation chromatography, and finally a somewhat selective affinity chromatography step involving Red Procion-H3B dye or benzylphosphonic acid agarose.

Amino Acid Sequence

An important feature is the presence of two major isoenzymes (electrophoretic fast and slow forms), which differ in the segment containing aa 38-76 of the 157 aa mature protein. (Note that this numbering ignores the N-terminal Met, which is removed in processing to form the mature enzyme.) These residues are coded for by distinct exons in the ACP1 gene, and the primary mRNA transcript is evidently subject to exclusive, alternative splicing.

```
HCPTP-A (ACP1-fast)³⁸NWRVDSAATSGYEIGNPPDYRGQSCMKRHGIPMSHVARQ⁷⁶
                     || || | |   | || |  ||  |||  | |||
HCPTP-B (ACP1-slow)³⁸NWVIDSGAVSDWNVGRSPDPRAVSCLRNHGIHTAHKARQ⁷⁶
```

Disulfides/SH-Groups

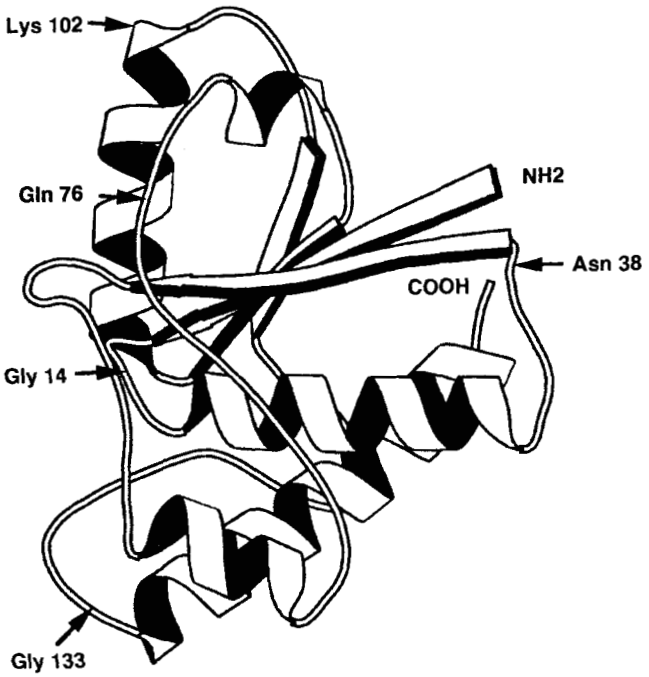
The enzyme contain 8 cysteines, all of which are present as free SH groups.

General References

Zhang et al. *Adv. Protein Phosphatases* 1995, **9**:1-23.
Bryson, G. M. et al. *Genomics* 1995, **30**:133-140.
Kane, M. *J. Chromatog.* 1991, **569**:297-321.
Bottini, E. et al. *Human Genet.* 1995, **96**:629-637.

Ref. for DNA/AA Sequences

Wo, Y.-Y. P. et al. *J. Biol. Chem.* 1992, **267**:10856-10865.
Protein sequences currently found as NCBI entries 585633 and 585634 are incorrect; residue #32 should be T, not W.



The structure of human low molecular weight phosphotyrosyl protein phosphatase (red cell acid phosphatase) closely resembles that of the homologous bovine enzyme (Zhang et. al., *Biochemistry* 1994, **33**:11097-11105), and consists of four-stranded central parallel *beta*-sheet with *alpha* helices flanking both sides. The phosphate binding loop consists primarily of residues 12-18 (CLGNICR); the position of Gly-14 is indicated on the figure. Note that the phosphate binding loop is located near the N-terminal end of an extended helix, and this may contribute to the binding of the negatively charged substrate. Each of the amino acid positions indicated on the structure diagram corresponds to the location of intron/exon junctions in the human gene. The protein segment that differs in the electrophoretic fast and slow form, and which results from alternative splicing of the primary mRNA transcript, extends from Asn-38 to Gln-76.

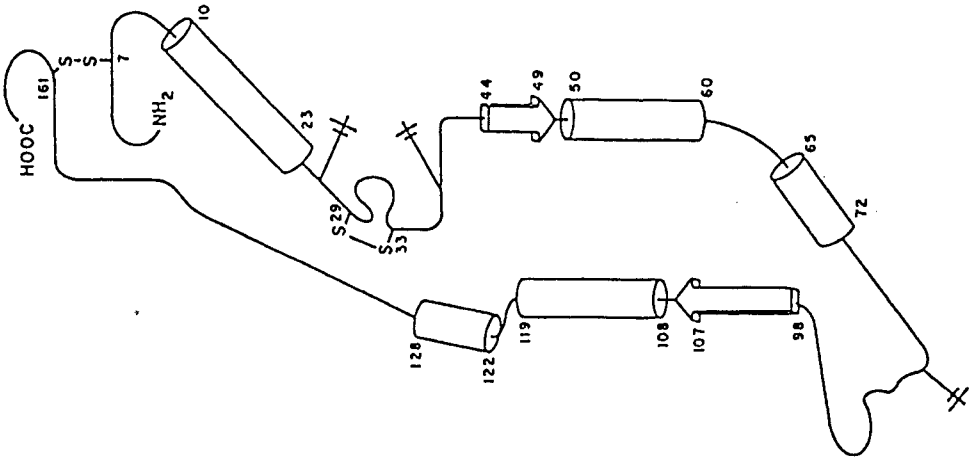
Erythropoietin

Jerry L. Spivak

Synonyms	None
Abbreviations	EPO; rHuEPO; EP
Classifications	Bioassay; radioimmunoassay
Description	A heavily glycosylated single chain polypeptide (165 aa) with two internal disulfide bonds which is produced primarily in the kidneys but also to a small extent in the liver and circulates in the plasma.
Structure	A globular protein (predicted radius 20.2 Å) with an observed Stokes radius of 32 Å due to its carbohydrate residues, constituted as four anti-parallel α helical bundles with two long and one short intervening loops and several stretches of β sheets; a structure which is characteristic of hematopoietic growth factors in general even though they share little aa sequence homology. The protein appears to have two independent binding sites for its receptor which differ in their binding affinity.
Molecular Weight	30,400 (sedimentation equilibrium)
Sedimentation Coeff.	Unknown
Isoelectric Point	4.2-4.6 (glycosylated), 9.2 for the aglycone
Extinction Coeff.	8.51 (278nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	A growth factor: erythropoietin interacts with specific receptors on the surface of erythroid progenitor cells to promote their proliferation and maintain their viability during erythroid differentiation. Receptor-ligand binding is rapid and of high affinity (Kd = 100-500 pM) and results in the formation of erythropoietin receptor dimers. Receptor dimerization leads to the apposition of two JAK2 molecules resulting in JAK2 autophosphorylation, and posphorylation of the erythropoietin receptor and a number of protein involved in intracellular signal transduction including certain STAT proteins, and activation of $p21^{Ras}$ and several tyrosine phosphatases.
Physiology/Pathology	Erythropoietin which is constitutively produced in the kidneys is an essential growth factor for erythroid progenitor cells and serves to maintain the red cell mass at a constant level by facilitating erythrocyte production in the bone marrow. Tissue hypoxia whether due to anemia or a change in ambient oxygen tension stimulates further erythropoietin production for the recruitment of additional erythroid progenitor cells to proliferate and differentiate. Decreased production of erythropoietin leads to anemia while overproduction results in an expansion of the red cell mass above the normal range. Since erythropoietin is normally constitutively produced, it

	is never absent from the circulation even in the anephric state as it is also produced in the liver.
Degradation	Erythropoietin is internalized and degraded by erythroid progenitor cells but this does not constitute the major pathway for its degradation which is as yet unknown.
Genetics/Abnormalities	Erythropoietin is highly conserved with over an 80% homology between the human, primate and rodent genes. No abnormal forms of the protein have been identified. The human erythropoietin gene is located on chromosome 7.
Half-life	7.5 hrs (range 4 - 12 hrs).
Concentration	Plasma, 0.1 $\mu g\ L^{-1}$ (range 0.5-2.5 $\mu g\ L^{-1}$).
Isolation Method	Conventional isolation techniques using human urine as the source of erythropoietin have been replaced by molecular cloning and stable transfection of mammalian cell lines for the production of recombinant erythropoietin.
Amino Acid Sequence	Erythropoietin has a 20% homology with the N-terminus of thrombopoietin, an hematopoietic growth factor also produced in the liver and kidneys. Within the erythropoietin molecule two specific domains comprised by aa residues on the A, C and D helices appear to be important with respect to its biological activity as are its internal disulfide bonds.
Disulfides/SH-Groups	Two intrachain disulfide bonds; no free sulfhydryl residues.
General References	Miyake, T. et al: Purification of human erythropoietin. *J. Biol. Chem.* 1977, **252**:5558-5564. Sasaki, H. et al. Carbohydrate structure of erythropoietin expressed in Chinese hamster ovary cells by a human erythropoietin cDNA. *J. Biol. Chem.* 1987, **262**:12059-12076. Davis, J. M. et al. Characterization of recombinant human erythropoietin produced in Chinese hamster ovary cells. *Biochem.* 1987, **26**:2633-2638. Philo, J. et al. Dimerization of the extracellular domain of the erythropoietin (EPO) receptor by EPO - One high-affinity and one low-affinity interaction. *Biochem.* 1996, **35**:1681-1691. Wen, D. et al. Erythropoietin structure-function relationships: Indentification of functionally important domains. *J. Biol. Chem.* 1994, **269**: 22839-22846. Krantz, S. Erythropoietin. *Blood* 1991, **77**:419-434.
Ref. for DNA/AA Sequences	Jacobs, K. et al. *Nature* 1985, **313**:806-809. Lin, F-K. et al. *Proc. Natl. Acad. Sci. USA* 1985, **82**:7580-7584. Recny, M.A. et al. *J. Biol. Chem.* 1987, **262**:17156-17163. Lai, P-H. et al. *J. Biol. Chem.* 1986, **261**:3116-3121.

Molecular model according to McDonald et al. (*Mol Cell Biol* **3:** *842*, 1986).



Erythropoietin Receptor

Lena Avedissian and Jerry L. Spivak

Synonyms	None
Abbreviations	EPO-R, EpoR
Classifications	Type I hematopoietic growth factor receptor
Description	A glycosylated, single chain, membrane-spanning polypeptide (507 amino acids in mice, 508 amino acids in humans) with a 24 amino acid signal peptide, a 223 amino acid extracytoplasmic ligand binding domain (224 in humans), a 24 amino acid transmembrane domain and a 236 amino acid cytoplasmic domain, expressed primarily by erythroid progenitor cells but also expressed by rodent placenta, embryonal stem cells, multipotent hematopoietic progenitor cells, endothelial cells and neural cells. The receptor is expressed not only as a full length protein in erythroid progenitor cells but also in soluble and truncated forms. The erythropoietin receptor is a member of the hematopoietic growth factor receptor superfamily, sharing in common with other members of this family four positionally conserved cysteines as well as a Tryptophan-Serine-X-Tryptophan-Serine motif in its extracellular domain and lack of a kinase motif in its cytoplasmic domain. However, the cytoplasmic domains of these receptors do contain homologous regions in their membrane proximal regions responsible for mitogenesis.
Structure	Not yet crystallized but by analogy with other members of its receptor family, the extracellular domain of the erythropoietin receptor is thought to have a secondary structure consisting of two 100 amino acid domains, each containing seven beta strands whose tertiary structure creates a docking site for the cognate ligand.
Molecular Weight	55,000 (based on amino acid content) 66,000 (electrophoretic mobility)
Sedimentation Coeff.	Unknown
Isoelectric Point	5.6
Extinction Coeff.	20.8 (280 nm,1 %,1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	Erythropoietin is the cognate ligand for the receptor, and JAK2 is the tyrosine kinase which associates with the receptor and is activated by receptor-ligand binding and oligomerization.
Substrates	Tyrosine phosphorylation of the receptor permits proteins with SH2 domains such as Shc, and PI-3 kinase to associate with it.
Inhibitors	The truncated form of the receptor through dimerization with the full length receptor may prevent signal transduction. The carboxyl terminal portion of the receptor also exerts a negative regulatory effect on receptor function.
Biological Function	The erythropoietin receptor, following binding of its ligand erythropoietin, self-association and association with other partner proteins, initiates signal

transduction through activation of the tyrosine kinase, JAK2. The signal transduction pathway involves activation of p21 Ras, Raf-1 and protein kinase C as well as phosphorylation of nuclear transcription factors. The effects of erythropoietin receptor-mediated signal transduction vary with the maturation stage of the erythroid progenitor cell. Early erythroid progenitor cells which are largely quiescent are triggered into cell cycle while signal transduction in late stage erythroid progenitor cells, which are largely in cell cycle (mainly in S phase), serves to maintain cell viability (prevention of apoptosis) while these cells mature. Additionally, in late stage erythroid progenitor cells, receptor-ligand binding triggers differentiation events.

Physiology/Pathology

The erythropoietin receptor is responsible for the proliferation, survival and maturation of erythroid progenitor cells. It functions as both a competence and progression factor. While other hematopoietic growth factor receptors such as those for IL-3 and stem cell factor are expressed by erythroid progenitor cells which respond to their cognate ligands with respect to proliferation at certain stages of maturation, erythropoietin receptor-mediated signaling is required for complete maturation of erythroid progenitor cells.

Degradation

Ligand binding results in rapid down regulation of erythropoietin receptors expressed at the cell surface, increasing their turnover half time from 3 hours to 15 minutes. Degradation of the mature receptor appears to occur in lysosomes.

Genetics/Abnormalities

The murine erythropoietin receptor is expressed on chromosome 9 and the human receptor on chromosome 19. Deletion of the terminal 70 a.a. of the receptor cytoplasmic domain due to a point mutation was associated with dominantly-inherited erythrocytosis in a large Finnish kindred. Deletions of the carboxyl terminal domain experimentally have been shown to increase the responsiveness of the erythropoietin receptor to its ligand.

Half-life

The half-life of erythropoietin receptor mRNA is 75–150 minutes, while the half life of the mature unbound receptor is 60–180 minutes. Ligand binding accelerates the downregulation and metabolism of surface receptors.

Concentration

The number of surface receptors varies according to the state of erythroid progenitor cell maturation with receptor expression being lowest in the most primitive erythroid progenitors. On average, late erythroid progenitor cells express 300–1000 surface receptors with an affinity for erythropoietin of 100–600 pM by equilibrium binding. Kinetic studies of ligand binding by purified full lenth erythropoietin receptors in solution revealed a binding affinity of 3pM.

Isolation Method

The most effecient method of isolation is expression in Sf9 insect cells using a recombinant baculovirus vector containing the receptor cDNA. Biologically active receptor can be purified from infected Sf9 cells by lectin affinity chromatography.

Amino Acid Sequence

Murine and human receptors are 82% homologous.

Disulfides/SH-Groups

No disulfides; 11 SH-groups

General References

Youssoufian, H. et al. Structure, function and activation of the erythropoietin receptor. *Blood* 1993, **81**:2223–2236.
Mayeux, P. et al. Structure of the murine erythropoietin receptor complex. *J. Biol. Chem.* 1991, **266**: 23380–23385.

Bazan, J. F. Structural design and molecular evolution of a cytokine receptor superfamily. *Proc. Natl. Acad. Sci.* 1990, **87**:6934–6938.
Witthuhn, B. A. et al. JAK2 associates with the erythropoietin receptor and is tyrosine phosphorylated and activated following stimulation with erythropoietin. *Cell* 1993, **74**:227–236.
de la Chapelle, A. et al. Truncated erythropoietin receptor causes dominantly inherited benign human erythrocytosis. *Proc. Natl. Acad Sci* 1993, **90**:4495–4499.

Ref. for DNA/AA Sequences

D'Andrea, A. D. et al. Expression cloning of the murine erythropoietin receptor. *Cell* 1989, **57**: 277–285.
Winkelmann, J. C. et al. The gene for the human erythropoietin receptor: analysis fo the coding sequence and assignment to chromosome 19p. *Blood* 1990, **76**: 24–30.

Factor I

Michael K. Pangburn

Synonyms	C3b Inactivator; C3b/C4b Inactivator; Conglutinin activating factor (KAF)
Abbreviations	I
Classifications	EC 3.4.21.45
Description	Factor I is a serine protease highly specific for complement proteins. It circulates in active form and functions as a regulator of complement activation.
Structure	Typical serine protease domain (38,000 Da chain), two domains homologous with the Low-Density Lipoprotein Receptor and one 60 - 70 residue domain with eight Cys and one Trp as in complement proteins C6 and C7.
Molecular Weight	88,000 Da (SDS-PAGE). Two disulfide-bonded chains of 50,000 and 38,000 Da.
Sedimentation Coeff.	5.5 S
Isoelectric Point	5.7-6.1
Extinction Coeff.	14.3 (280nm, 1%, 1cm)
Enzyme Activity	Serine protease, cleaves at carboxyl side of arginine residues. Highly specific for proteins of the complement system: C3b, iC3b, iC3, iC4 and C4b. Requires a cofactor.
Coenzymes/Cofactors	Absolute requirement for protein cofactors. Complement factor H, CR1 or MCP for cleavage of C3-derived products. Requires C4bp or CR1 to cleave C4-derived products. A cofactor must be bound to the substrate protein for factor I to cleave it.
Substrates	No synthetic substrates are known.
Inhibitors	No natural inhibitors in blood. Not inhibited by standard trypsin inhibitors (soybean trypsin inhibitor, leupeptin). Poorly inhibited by DFP and PMSF. Synthetic inhibitors: FUT-175, Boroarginine-containing peptides.
Biological Functions	Factor I cleaves and inactivates two proteins of the complement system thereby stopping activation of complement. It circulates in blood in active form and only cleaves the activated forms of C3 and C4 (C3b and C4b) or their breakdown products when they are bound to cofactor proteins.
Physiology/Pathology	Complete deficiencies of factor I are characterized by partial deficiency of C3 and complete loss of functional factor B due to uncontrolled activation of the alternative pathway of complement. These individuals have no functional alternative pathway of complement and suffer recurrent infections. Heterozygotes exhibiting low factor I concentrations in blood are not detectably at risk.
Degradation	Unknown
Genetics/Abnormalities	Gene located on chromosome 4.

Half-life	Approx. 3 days in humans.
Concentration	34 mg/L, (0.39 μM) in plasma.
Isolation Method	Isolated from plasma by antibody affinity chromatography or conventional chromatography (*Methods Enzymol.* 1982, **80**: 112 and 1988, **162**: 639).
Amino Acid Sequence	Known (see Ref. for DNA sequence below).
Disulfides/SH-Groups	Unknown
General References	Reid, K.B.M. and Porter, R.R. *Ann.Rev. Biochem.* 1981, **50**:433-464. Muller-Eberhard, H.J. *Ann. Rev. Biochem.* 1988, **57**:321-347. Pangburn, M.,K. and Muller-Eberhard, H.J. *Springer Semin. Immunopathol.* 1984, **7**:163-192. Rother, K. and Till, G.O.(eds.) *The Complement System.* 1988, Springer-Verlag, Berlin.
Ref. for DNA/AA Sequences	Goldberger, G. et al. *J. Biol. Chem.* 1987, **262**:10065-10071.

Factor V

Richard J. Jenny

Synonyms	Accelerator globulin; Proaccelerin; Labile factor
Abbreviations	FV (procofactor); FVa (cofactor)
Classifications	Electrical mobility: β-fraction
Description	A soluble plasma protein essential to the blood coagulation cascade. Circulates as a single chain procofactor (Mr = 330,000) which is proteolytically converted by thrombin or factor Xa to the active cofactor, factor Va. Factor Va is a heterodimer composed of an N-terminal derived heavy chain and a C-terminal derived light chain which remain non-covalently associated in the presence of divalent metal ions (ie; Ca^{2+}). Factor Va is a cofactor for the serine protease factor Xa. Factor Va and factor Xa combine on membrane surfaces to form the prothrombinase complex. This enzyme complex catalyzes the rapid conversion of the zymogen prothrombin to the enzyme thrombin.
Structure	Factor V is a large single-chain plasma glycoprotein. Sedimentation velocity and sedimentation equilibrium studies indicate that factor V has a molecular weight of 330,000 and that the molecule is assymetric with a rod-like structure. Electron microscopy data also indicate that factor V is assymetric, appearing as an irregular, multi-domain molecule. The domains are presumed to represent subunits which ultimately form factor Va and related activation peptides. The cDNA encoding factor V has been isolated and sequenced. The translated aa sequence consists of 2224 aa which is in good agreement with the random coil chain length of 2630 aa deduced from sedimentation studies.
Molecular Weight	330,000 (sedimentation equilibrium); Factor Va 168,000; Factor Va heavy chain 94,000; Factor Va light chain 74,000; thrombin generated activation peptides 120,000 and 71,000 (SDS-PAGE).
Sedimentation Coeff.	9.19 S
Isoelectric Point	Unknown
Extinction Coeff.	9.6: factor V; 17.4: factor Va; 22.3: factor Va heavy chain; 12.4: factor Va light chain; (all 280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	Single chain factor V is a procofactor. Selective proteolysis by either thrombin or factor Xa converts factor V to the cofactor, factor Va. Factor Va combines with the serine protease factor Xa on lipid membranes, in a calcium dependent manner, to form the enzyme complex prothrombinase. Prothrombinase rapidly converts the zymogen prothrombin to the active enzyme thrombin. The rate of thrombin generation by prothrombinase is 300,000 fold greater than the rate with factor Xa alone.
Substrates	None

Inhibitors	Factor Va is proteolytically inactivated by activated protein C. Activated protein C cleaves the heavy chain region of human factor Va at three sites (Arg-306, Arg-506, and Arg-679). Cleavage at Arg-306 appears to be the event which ultimately renders the cofactor inactive. The resulting inactive cofactor, factor Vai, no longer supports assembly of the prothrombinase complex.
Biological Functions	Factor Va is an essential cofactor of the prothrombinase enzyme complex. As a cofactor, it exhibits two distinct functions: 1) it effectively serves as a receptor for factor Xa on membrane surfaces; and 2) it binds the substrate prothrombin in a manner which facilitates rapid proteolysis of prothrombin by the enzyme factor Xa. Thus, factor Va is required for assembly of the prothrombinase enzyme complex as well as for efficient binding of the subsrate prothrombin.
Physiology/Pathology	The majority of factor V in whole blood circulates in the plasma, but a significant portion (nearly 20%) is contained within the blood platelets. A deficiency of factor V is often accompanied by a significant bleeding diathesis, the severity of which is determined by the nature of the deficiency. Evidence indicates that a deficiency of plasma factor V is of little clinical significance if concentrations of functional platelet factor V are near normal. Furthermore, the platelet store of factor V is relatively inaccessible to plasma proteins, which allows for normal hemostasis to occur even when acquired factor V inhibitors or inactivators may be present in plasma. Thus, the platelet store of factor V is thought to be essential for normal hemostasis.
Degradation	Factor V and factor Va are proteolytically degraded by both plasmin and activated protein C. Plasmin cleavage of factor V generates a transiently active cofactor species which is ultimately inactivated by further plasmin degradation. Activated protein C proteolytically cleaves both factor V and factor Va, but unlike plasmin, no transiently active species is generated.
Genetics/Abnormalities	A deficiency of factor V is commonly called parahemophilia and is a rare occurrence. A form of thrombophilia has been documented which is caused by a single nucleotide change in the factor V gene which leads to an aa substitution, whereby Arg-506 becomes a Glu. This form of factor V (Factor V Leiden) is characterized by a resistance to inactivation by activated protein C.
Half-life	Unknown
Concentration	Platelet Poor Plasma (0.004 - 0.010 g/L); Platelets (0.002 - 0.003 g/L of whole blood)
Isolation Method	Isolated from freshly prepared plasma treated with a variety of serine protease inhibitors to prevent degredation of factor V. Isolation steps often include barium citrate adsorbtion to remove the plasma vitamin K-dependent proteins, and additional fractionation with polyethylene glycol. Final purification is accomplished either by conventional chromatography which includes ion exchange and gel filtration, or by affinity chromatography employing immobilized antibody.
Amino Acid Sequence	The cDNA encoding factor V has been isolated and sequenced. The 6672 base pair cDNA predicts an amino acid chain length of 2224 aa including a 28 aa leader peptide. The aa sequence is composed of homologous domains which include a triplicated "A" domain, a duplicated "C" domain, and a large central "B" domain, arranged in the order A1-A2-B-A3-C1-C2.

A comparison of the factor V sequence to that of other plasma proteins reveals homologies between factor V, factor VIII and Ceruloplasmin. Each of these proteins posses triplicated "A" domains which are nearly 40% identical on both an inter- and intra-molecular basis.

Disulfides/SH-Groups

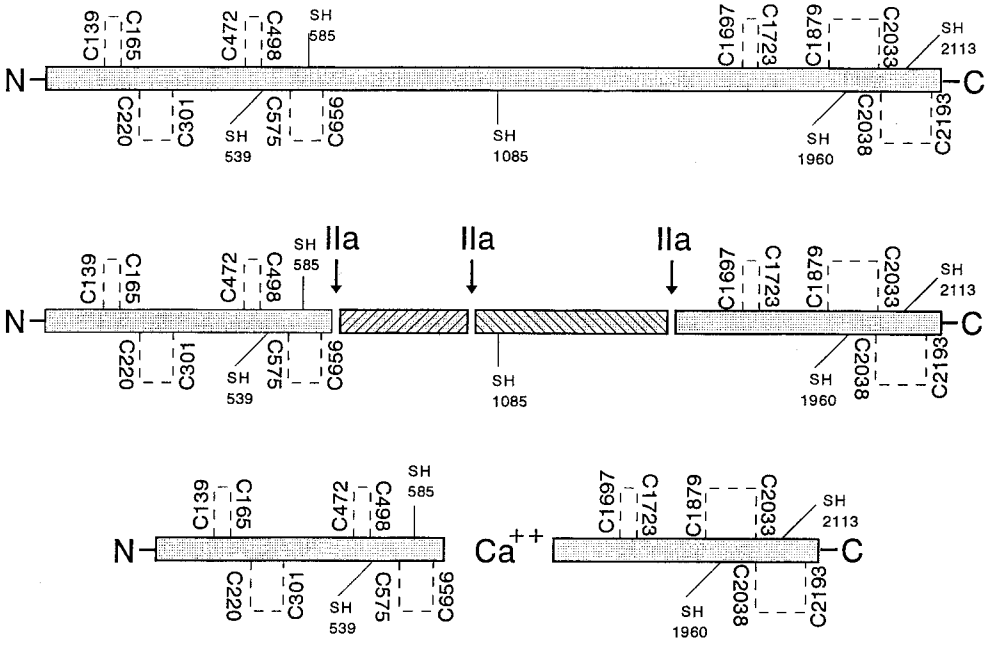
The location of disulfide bonds and free sulfhydryl groups in human factor V have been inferred from analyses completed on the bovine molecule (see Figure). The deduced aa sequence indicates a total of 19 cysteine residues. Titration of free sulfhydryl groups with DTNB [5,5'-dithiobis-(2-nitro-benzoic acid)] indicates the presence of two free sulfhydryl groups in the heavy chain region and two in the light chain region. A free sulfhydryl group is also present in the 120,000 activation peptide.

General References

Mann, K.G. et al. *Ann. Rev. Biochem.* 1988, **57**:915-956.
Jenny, R.J. and Mann, K.G. "Factor V: a prototype pro-cofactor for vitamin K-dependent enzyme complexes in blood clotting". In: *Bailliere's Clinical Haematology.* Bailliere Tindall Limited, London, 1989; pp. 919-944.
Kane, W.H. et al. *Blood* 1988, **71**:539-555.
Bertina, R.M. et al. *Nature* 1994, **369**:64-67.
Kalafatis, M. et al. *J. Biol. Chem.* 1995, **270**:4053-4057.
Xue, J. et al. *Biochemistry* 1993, **32**:5917-5923.
Xue, J. et al. *Biochemistry* 1994, **33**:13109-13116.

Ref. for DNA/AA Sequences

Jenny, R.J. et al. *Proc. Natl. Acad. Sci. USA* 1987, **84**:4846-4850.
Kane, W.H. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:6800-6804.
Kane, W.H. et al. *Biochemistry* 1987, **26**:6508-6514.



Thrombin cleavage sites and location of paired cysteine residues in human factor V/Va. Thrombin cleaves single chain human factor V at three sites (Arg 709, Arg 1018, Arg 1545) to yield two activaton peptides (residues 710-1018 and 1019-1545), and the heavy and light chains of human factor Va (residues 1-709 and 1546-2196 respectively). The heavy and light chains remain non-covalently associated by virtue of a divalent metal ion-dependent interaction. The location of disulfide bonded cysteine residues and free sulfhydryl groups in human factor V and Va (as inferred from analyses of the bovine molecule) are illustrated. All amino acid numbers refer to the intact human factor V molecule.

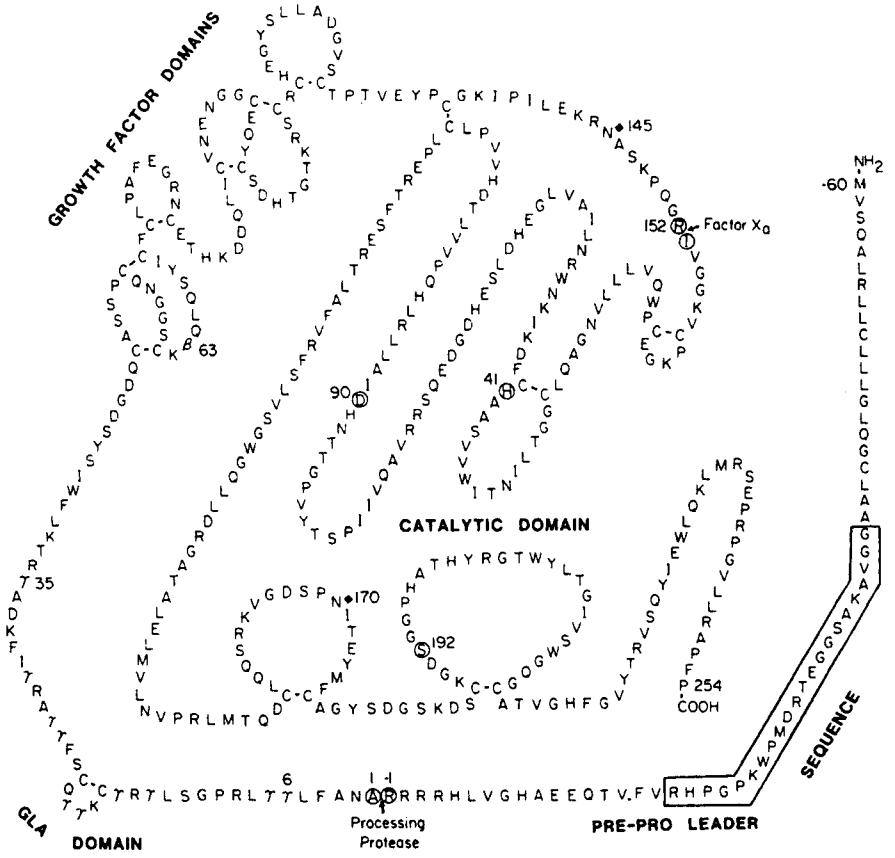
Factor VII

Walter Kisiel

Synonyms	Proconvertin, Precursor of Serum Prothrombin Conversion Accelerator
Abbreviations	VII
Classifications	Zymogen of a serine protease
Description	A multi-domain, single-chain plasma glycoprotein synthesized in the liver and secreted as a zymogen of a serine protease, factor VIIa. It contains 10 γ-carboxyglutamic acid residues located in the amino-terminal region of the protein, and these residues required vitamin-K for their biosynthesis. Factor VII also contains one β-hydroxyaspartic acid residue at position 63, two N-glycosidically linked carbohydrate chains at Asn-145 and Asn-322, and a disaccharide (or trisaccharide) O-glycosidically linked to Ser-52. Single-chain factor VII is converted to two-chain factor VIIa by cleavage of an internal peptide bond located at Arg-152-Ile-153.
Structure	A globular molecule with five distinct structural domains: an amino-terminal γ-carboxyglutamic acid-rich domain (residues 1–49), two epidermal growth factor-like domains (residues 50–130), a connecting region (residues 131–152), and a serine protease catalytic domain (residues 153–406).
Molecular Weight	50,000 (SDS-PAGE) 45,512 (aa sequence of CHO-free protein)
Sedimentation Coeff.	Unknown
Isoelectric Point	5.0–5.3
Extinction Coeff.	13.9 (280 nm, 1%, 1 cm)
Enzyme Activity	Zymogen of a serine protease (factor VIIa)
Coenzymes/Cofactors	Tissue factor apoprotein (TFA). TFA is a cell-surface, transmembrane glycoprotein that forms a 1 : 1 molecular complex with factor VII or VIIa. Complex formation with factor VIIa augments factor VIIa proteolytic activity several orders of magnitude.
Substrates	Blood coagulation factor X. Blood coagulation factor IX
Inhibitors	Extrinsic pathway inhibitor (also called lipoprotein associated coagulation inhibitor, or LACI). This plasma inhibitor (Mr $\simeq$ 38,000) consists of three Kunitz-type domains which recognize and neutralize a ternary complex of factor VIIa-tissue factor-factor Xa in the presence of calcium ions.
Biological Functions	Participates in the extrinsic pathway of blood coagulation following conversion to factor VIIa by other blood coagulation serine proteases (either thrombin, factor Xa, factor IXa or factor XIIa). In the presence of tissue factor and calcium ions, factor VIIa converts factor X to factor Xa by limited proteolysis, and the latter enzyme converts prothrombin to thrombin in the presence of factor Va, calcium and phospholipids. Factor VIIa also converts factor IX to IXa in the presence of tissue factor and calcium. The relative physiological importance of factor X versus factor IX activation by factor VIIa has not been established thus far.

Physiology/Pathology	Essential for blood coagulation; deficiencies result in a bleeding diathesis, the clinical severity of which is inversely related to its plasma antigenic or functional level.
Degradation	Unknown
Genetics/Abnormalities	Synthesis from a single mRNA. Deficiencies in factor VII activity are rare, affecting roughly 300 individuals in the United States. Localization: Chromosome 8, band 13q34.
Half-life	2.0–2.5 hrs (blood circulation)
Concentration	Plasma: 500 μg/L (range 400–600 μg/L)
Isolation Method	Isolated from citrated plasma by a combination of barium citrate adsorption and elution, DEAE-Sepharose column chromatography, and immunoaffinity chromatography using murine anti-factor VII IgG monoclonal antibodies coupled to agarose.
Amino Acid Sequence	Factor VII consists of 406 residues, and exhibits extensive sequence homology and identity with other plasma vitamin K-dependent coagulation factors. Catalytic domain is also highly homologous to that of the mammalian pancreatic serine proteases.
Disulfides/SH-Groups	12 disulfides, all intrachain in factor VII. No free sulfhydryls. In two-chain factor VIIa, a single disulfide pair joins the light and heavy chains of this molecule.
General References	Broze, G. J. and Majerus, P. W. *J. Biol. Chem.* 1980, **255:** 1242–1247. Bajaj, S. P. et al. *J. Biol. Chem.* 1981, **256:** 253–259. Davie, E. W. "The blood coagulation factors: Their cDNAs, genes, and expression." In: *Hemostasis and Thrombosis: Basic Principles and Clinical Practice* (Colman, R. W. et al., eds.) J. B. Lippincott Co. 1987, pp. 242–267.
Ref. for DNA/AA Sequences	Hagen, F. S. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83:** 2412–2416. O'Hara, P. et al. *Proc. Natl. Acad. Sci. USA* 1987, **84:** 5158–5162.

Molecular model according to Hagen et al. (*Proc. Natl. Acad. Sci. USA*, 1986)



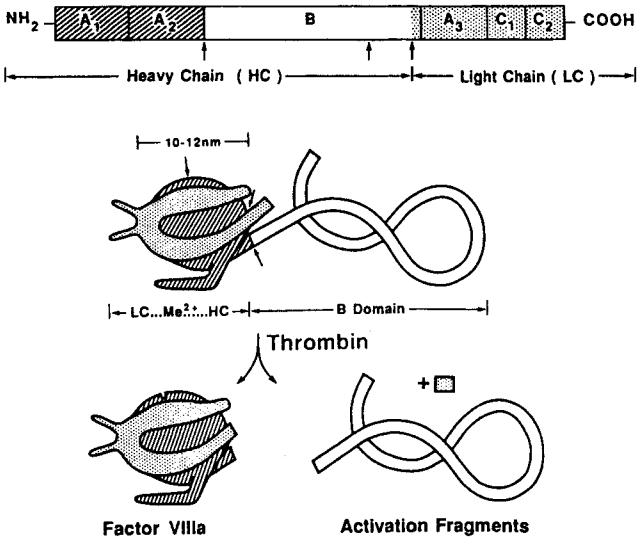
Factor VIII

Leon W. Hoyer

Synonyms	Antihemophilic factor; Antihemophilic globulin
Abbreviations	FVIII:C; AHF; AHG
Classifications	Not available
Description	A plasma protein, synthesized in the liver and elsewhere, that circulates in the blood in a non-covalent complex with von Willebrand factor. In plasma, the factor VIII molecule consists of an 80 kDa light chain associated with a heavy chain of variable length (90 - 220 kDa); the 2 chains are linked by a divalent cation. The heavy chain consists of 2 A domains; there is a connecting (B) domain; and the light chain includes a third A domain and 2 C domains. Factor VIII is synthesized as a single chain precursor polypeptide of 2,351 aa that includes a 19 aa hydrophobic signal peptide that is removed during secretion. Factor VIII contains 25 potential N-linked oligosaccharide attachment sites, most within the B domain.
Structure	By scanning transmission electron microscopy, factor VIII consists of a globular 10 - 12 nm core that contains both heavy and light chains, and stalk-like extensions (B domain structure). Examination of rotary-shadowed samples also shows a globular region and associated "tails". Not yet crystalized in its native form.
Molecular Weight	285,000 (sedimentation and gel diffusion). Light chain: 80,000; heavy chain: 90,000. (SDS-PAGE)
Sedimentation Coeff.	8.2 S
Isoelectric Point	Unknown
Extinction Coeff.	12 (280nm, 1%, 1cm) for porcine factor VIIIa
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Antibodies to factor VIII develop in transfused hemophilic patients, and occasionally are also recognized as autoantibodies.
Biological Functions	Essential cofactor in factor X cleavage by factor IXa. Cofactor function requires prior activation of factor VIII by thrombin cleavage at Arg-372 - Ser-373 and Arg-1689-Ser-1690. Factor VIII circulates in plasma in a non-covalent complex with von Willebrand factor. The interaction with von Willbrand factor is necessary for release of factor VIII from cells, for factor VIII stability in plasma, and for effective concentration of factor VIII at sites of hemostasis.
Physiology/Pathology	Essential for normal blood coagulation. Concentrations below 20% of the normal level cause a bleeding disorder designated hemophilia A. When less than 1% of normal factor VIII is present in the plasma, a severe

	bleeding disorder is present, with spontaneous joint bleeding being the most common symptom.
Degradation	Factor VIII is very susceptible to inactivation by protease cleavage. Factor Xa inactivates factor VIIIa through cleavage at residues 336 and 1721, and activated protein C inactivates factor VIIIa by cleavage at residue 336. In addition, thrombin activation of factor VIII is followed by non-proteolytic inactivation due to subunit dissociation and insolubility of the free A2 subunit. None of the inactive products have a known biologic activity.
Genetics/Abnormalities	Factor VIII is the product of a large (186 kb pair) gene at the tip of the long arm of the X chromosome. Hemophilia A, a heterogeneous disorder, is a consequence of factor VIII gene abnormalities including deletions, inversions, missense mutations, and nonsense mutations. In 5% of hemophilic plasmas, a dysfunctional factor VIII-like protein is present. To date, identified mutations responsible for loss of factor VIII activity include substitutions that abolish thrombin-cleavage sites and mutations that introduce new N-glycosylation sites.
Half-life	8 - 12 hrs
Concentration	200 μg/L (range 100 - 400 μg/L)
Isolation Method	Factor VIII is present in cryoprecipitate prepared from plasma. Further purification can be obtained by a variety of precipitation techniques or by immunoaffinity chromatography using antibodies to von Willebrand Factor or to factor VIII.
Amino Acid Sequence	Factor VIII has important sequence homologies with factor V and ceruloplasmin (A domains) and discoidin 1 (C domains). The mature protein has 2,332 aa.
Disulfides/SH-Groups	There are 22 cysteine residues, clustered in the A1, A2, and light chain regions. There are no interchain disulfide bridges. Cys-528 (domain A1) and Cys-1858 (domain A3) are free.
General References	Kane, W.H. and Davey, E.W. *Blood* 1988, **71**:539-555. White, G.C. and Shumaker, C.B. *Blood* 1989, **73**:1-12. Foster, P.A. and Zimmerman, T.S. *Blood Reviews* 1989, **3**:180-191. Eaton, D.L. and Vehar, G.A. *Prog. Hemost. Thromb.*1986, 8:47-70. Lollar, P., Fay, P.J. and Fass, D. *Meth. Enzymol.* 1993, **222**:128-143.
Ref. for DNA/AA Sequences	Gitschier, J. et al. *Nature* 1984, **312**:326-330. Toole, J.J. et al. *Nature* 1984, **312**:342-347.

Molecular model according to M. W. Mosesson (*J. Clin. Invest.* 1984, 85; 1983–1990.



Schematic drawing of the factor VIII domains and a model based on electron microscope images.

Factor IX

Syed S. Ahmad and Peter N. Walsh

Synonyms	Plasma thromboplastin component (PTC); Christmas factor; Antihemophilic factor B
Abbreviations	FIX; FIXα; FIXaα; FIXaβ; FIXa
Classifications	Zymogen of a serine protease
Description	Factor IX (FIX) is a plasma glycoprotein synthesized by liver that participates in the middle phase of intrinsic blood coagulation. Vitamin K is required for the biosynthesis of FIX as well as FVII, FX, prothrombin, protein S and protein C, all of which share considerable sequence homology, suggesting that these vitamin-K-dependent proteins have evolved from a common ancestral gene.
Structure	Human FIX is a single chain glycoprotein, containing 18% carbohydrate (including hexose, N-acetylhexosamine and N-acetylneuraminic acid) and consisting of 415 aa. The FIX gene (approx. 34 kb) contains seven introns and eight exons that code for distinct structural domains that are highly conserved among other homologous vitamin-K-dependent plasma coagulation proteins. Exon I encodes the signal peptide and exon II and exon III code for the propeptide (which is cleaved from the mature protein prior to secretion through the Golgi apparatus) and the γ-carboxyglutamic acid (Gla) domain (comprising residues 1 - 46 including 12 glutamic acid residues that are post-translationally modified to the dicarboxylic Gla form by a vitamin-K-dependent carboxylase). Exon IV and exon V code for two similar epidermal growth factor (EGF)-like domains (residues 47 - 145), each of which contains six cysteines; the specific functions of the EGF domains are unknown but are postulated to include binding to platelet receptors (EGF-2) and to FVIII (EGF-2). The sixth exon codes for an activation peptide (residues 146 - 180), containing two carboyhdrate binding sites), formed when FIX is converted to FIXa by FXIa or by FVIIa in the presence of tissue factor (TF). Exon VII and exon VIII (and the 3' portion of exon VI) code for residues 181 - 415, which comprises the trypsin-like heavy chain domain of FIXa.
Molecular Weight	The Mr of human FIX is 54,000 - 57,000 Da. Activated FIX (Mr approx. 45,000) results from cleavage of two peptide bonds and the formation of an activation peptide (Mr approx. 11,000) plus the two-chain FIXa, consisting of disulfide-linked heavy (Mr 28,000) and light (Mr 18,000) chains.
Sedimentation Coeff.	Mr 55,000 (sedimentation equilibrium)
Isoelectric Point	3.85 - 6.04, main components 4.1 - 4.5
Extinction Coeff.	14.9 (280nm, 1%, 1cm)
Enzyme Activity	FIX exists in normal human plasma as a zymogen that requires proteolytic activation to develop serine protease activity. Normal activation can be achieved either by FIXa or by the FVIIa-tissue factor complex in the presence of Ca^{2+} ions by an initial cleavage at the Arg-145-Ala-146 bond to give rise to an inactive intermediate, FIXα, which is subsequently cleaved at the Arg-180-Val-181 bond to generate the fully active FIXαβ. In vitro,

an enzyme from the venom of the Russell's viper can cleave the Arg-180-Val-181 bond first give rise to the active intermediate FIXaα. FIX can also be activated by FXa at about 20% of the rate obtained with FXIa. The trypsin-like enzymatic activity of FIXa arises as a consequence of proteolytic cleavage at Arg-180-Val-181 which generates a free amino group at Val-181 that facilitates ion pair formation with the free carboxyl group of Asp-364. This in turn makes the hydroxyl group of the active center Ser-365 more labile. The catalytic components of FIXa include the active site (Ser-365, His-221, and Asp-269) and the substrate binding pocket (Asp-359). FIXa activates FX by cleaving a specific peptide bond (Arg-52-Ile-53) in the N-terminal region of the FX heavy chain.

Coenzymes/Cofactors

Although FIXa can slowly activate its normal plasma substrate, FX, this highly inefficient process is greatly facilitated by the presence of Ca^{2+} ions, activated platelets and a non-enzymatic protein cofactor, FVIII, that has previously been activated to FVIIIa by either thrombin or FXa. Human platelets, activated with thrombin, collagen or the Ca^{2+} ionophore A23187, can specifically and tightly (Kd approx. 0.5 nM) bind FIXa in the presence of FVIIIa and can promote F-X activation in part by exposing membrane aminophospholipids. Activated platelets or phosphatidylserine-containing phospholipid vesicles act by decreasing the Michaelis constant (Km) for F-X activation and permitting FVIIIa to increase the turnover number (kcat) resulting in (17×10^6)-fold increase in catalytic efficiency (kcat/Km).

Substrates

FIXa is the most specific and least reactive of the coagulation serine proteases. It recognizes FX as its normal macromolecular substrate, and can also cleave and activate FVIII but less effectively than thrombin or FXa. Although several synthetic dipeptide and tripeptide isobutyl thioesters and a few amide substrates have been described that are hydrolyzed by FIXa, none is sufficiently reactive and specific for FIXa to be in general use.

Inhibitors

The best documented physiological inhibitor of FIXa is antithrombin III which forms ae quimolar complex with the enzyme, a reaction that is accelerated several hundred-fold in the presence of heparin. FIXa can also be inhibited by forming an equimolar complex with the leech saliva protein, hirudin. An oligopeptide active site inhibitor of FIXa is dansyl-glutamyl-glycyl-arginyl chloromethyl ketone (DEGR-CK). However, FIXa, unlike other coagulation serine proteases, is resistant to inhibition by diisopropylphosphofluoridate (DFP). A variety of p-amidino esters, including the fluorescent compound, p-aminobenzamidine, have been used not only to inhibit FIXa by binding irreversibly to its active site, but also to measure FIX activation by monitoring their intrinsic fluorescence.

Biological Functions

After conversion from zymogen to active enzyme either by FXIa or by FVIIa-tissue factor, FIXa binds to a high-affinity, specific receptor on platelets (or on endothelial cells), consisting in part of aminophospholipids including phosphatidylserine, and forms a complex with the nonenzymatic protein cofactor, FVIIIa. This complex activates FX via the intrinsic coagulation pathway by a kinetic mechanism similar to that involving the extrinsic (FVII-tissue factor) pathway.

Physiology/Pathology

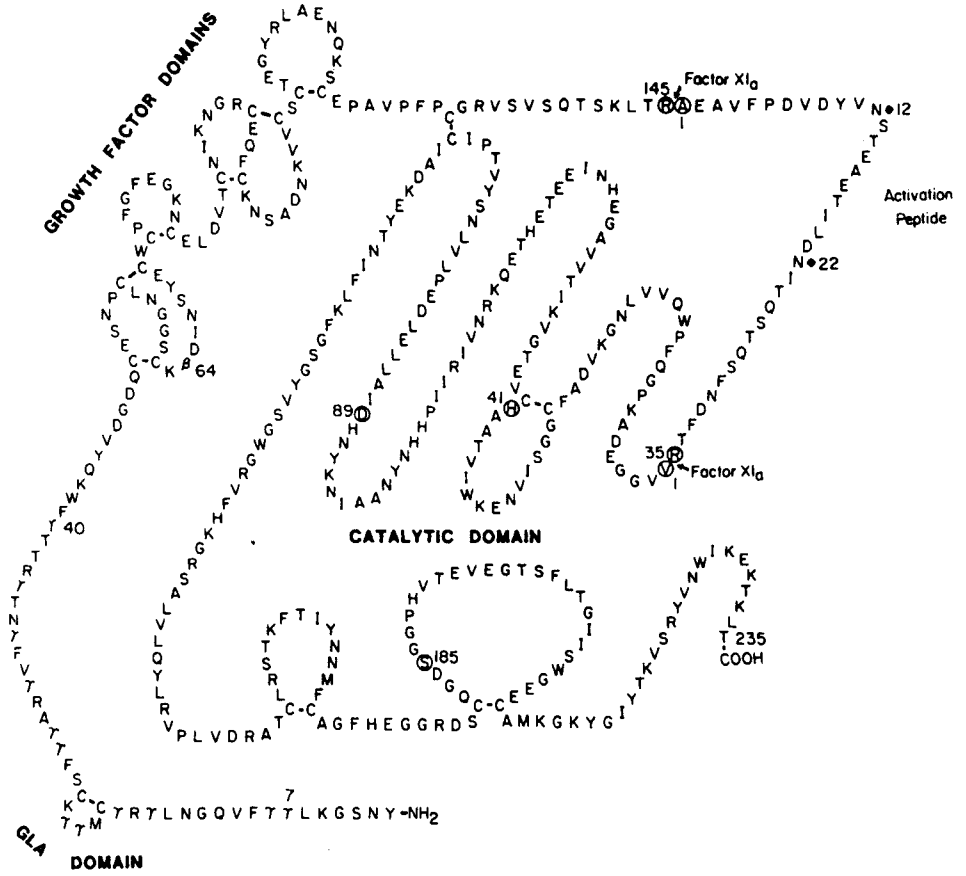
FIX, a zymogen of a serine protease is synthesized in the liver and is essential for normal hemostasis. It circulates in plasma as a zymogen and is activated by limited proteolysis by either FXIa (intrinsic pathway) or the FVIIa-tissue factor complex (extrinsic pathway). Hemophilia B is a coagulation disorder characterized by a deficiency of or defect in plasma FIX. Patients with hemophilia B exhibit considerable genetic heterogeneity: some have normal levels of antigen (cross-reacting material positive,

	CRM^+); in others the amount of FIX antigen is reduced (CRM^{red}); and in some patients little or no FIX is synthesized (CRM^-). The clinical symptoms in hemophilia B include: 1) lack of excessive hemorrhage from minor cuts or abrasions, since platelet function is normal; 2) joint and muscle hemorrhages leading to disabling long term sequelae; 3) prolonged and potentially fatal post-operative hemorrhage; and 4) a variety of social, psychological, vocational and economic problems.
Degradation	The principal known mechanism of inactivation of FIXa in vivo involves complex formation with the plasma serine protease inhibitor, antithrombin III. The rates of biosynthesis and metabolism of FIX are equal to maintain a steady-state concentration in plasma, but the specific mechanisms of degradation of the zymogen are largely unknown.
Genetics/Abnormalities	FIX deficiency or hemophilia B is a sex-linked recessive bleeding disorder, the incidence of which is approx. 1 in 50,000 in the population, occurring almost exclusively in males. Female hemophilia is extremely rare. Immunologic and cDNA sequencing data indicate that there are a wide variety of genetic defects which account for hemophilia B in different families. The FIX gene, located near the tip of the long arm of the x-chromosome, contains 34 kb of nucleotide sequence, 4% of which comprises eight separate exons that code for FIX protein. Over 400 separate genetic defects have been shown to cause hemophilia B; these include deletions (partial or complete) and a variety of point mutations affecting the catalytic, activation, EGF-like, Gla, propeptide and 5' flanking regions of the protein, as well as a number of mutations causing premature termination codons, or mutations at exon-intron junctions. A number of missence mutations (aa substitutions) have been identified in unrelated patients with mild to moderately severe hemophilia b, including $FIX_{Chapel\ Hill}$ (Arg-145 →His substitution) and $FIX_{Alabama}$ (Asp-47→Gly substitution).
Half-life	18 to 24 hrs (plasma)
Concentration	Normal plasma concentration is 3 - 5 mg/L (≈ 70 nM). One unit is defined as the amount of FIX in 1 ml of normal human plasma. The specific activity of FIX is 250 - 350 units per mg protein.
Isolation Method	FIX has been purified from human plasma by conventional multistep methods which include barium salt adsorption, ion exchange chromatography, gel filtration and affinity chromatography. More recent isolation techniques employ monoclonal antibodies or conformation specific antibodies which facilitate rapid and simple immunoaffinity purification of FIX with high specific activity.
Amino Acid Sequence	Human FIX consists of 415 aa, the sequence of which are shown in the molecular model below. Human FIX shows considerable sequence homology with bovine FIX and with other vitamin-K-dependent coagulation proteins such as FVII, FX, prothrombin and protein C.
Disulfides/SH-Groups	The disulfide linkages are shown in the molecular model below. The growth factor-like domains contain 12 of the 22 1/2-cysteine residues in FIX.
General References	Thompson, A.R. *Blood* 1986, **67**:565-572. Hedner, U. and Davie, E.W. "Factor IX" In: *Hemostasis and Thrombosis* (Colman, R.W. et al., eds.) J. B. Lippincott Company, 1987, pp. 97-111. Ahmad, S.S. et al. *J. Biol. Chem.* 1989, **264**:3244-3251. Rawala-Sheikh, R. et al. *Biochemistry* 1990, 29:2606-2610.

Thompson, A.R. "Molecular Biology of the Hemophilias" In: *Progress in Hemostasis and Thrombosis* (Coller, B.S., ed.) Vol. 10. W. B. Saunders Company, 1991, pp. 175-214.
Ahmad, S.S. and Walsh, P.N. *Trends in Cardiovascular Medicine* 1994, 4:271.
High, K.A. and Roberts, H.R. "Factor IX". In: *Molecular Basis of Thrombosis and Hemostasis*. High, K.A. and Roberts, H.R. (eds.) Marcel Dekker, Inc., New York, 1995, pp. 215-237.

Ref. for DNA/AA Sequences

Kurachi, K. et al. *Proc. Natl. Acad. Sci. USA* 1982, **79**:6461-6464.
Davie, E.W. "The blood Coagulation Factors: Their cDNAs, Genes and Expression" In: *Hemostasis and Thrombosis* (Colman. R.W. et al., eds.) J.B. Lippincott Company, 1987, pp. 242-267.



The single chain structure of FIX is shown with disulfide linkages. The arrows indicate the cleavage sites for FXIa and the diamonds designate N-linked glycosylation sites. Reprinted with permission from E. W. Davie, In: *Hemostasis and Thrombosis*, Colman, R. W. et al. (eds.), J. B. Lippincott Company, 1987.

Factor X

Craig M. Jackson

Synonyms	Stuart Factor, Prower Factor, Prothrombokinase, Autoprothrombin III
Abbreviations	F-X, X
Classifications	Zymogen/proenzyme to factor Xa (EC 3.4.21.6)
Description	Precursor of activated factor X (factor Xa), the proteinase that cleaves two peptide bonds in prothrombin to convert prothrombin into thrombin. Factor X is synthesized in the liver, requires vitamin-K for synthesis in fully, biologically active form, is post-translationally modified in reactions requiring vitamin K to produce 11 (12, bovine) γ-carboxyglutamate residues from glutamate. In a second post-translational modification one β-hydroxy-aspartate residue is formed from aspartate. Electrophoretic mobility between albumin and prealbumin.
Structure	Factor X contains 445 aa residues and is comprised of two polypeptide chains. A light chain of 139 aa residues contains a γ-carboxyglutamate-containing (Gla) domain and two growth factor-like domains. A heavy chain of 306 aa residues is homologous to thrombin and members of pancreatic serine proteinase family. Based on aa sequence homology with thrombin, the heavy chain is predicted to have a globular structure similar to that determined by x-ray crystallography for thrombin.
Molecular Weight	58,900 (sedimentation equilibrium); 16,200 light chain (aa sequence); 42,000 heavy chain (aa sequence).
Sedimentation Coeff.	3.6 S (bovine)
Isoelectric Point	4.7 (bovine)
Extinction Coeff.	11.6 (280 nm, 1%, 1 cm) human; 12.4 (280 nm, 1%, 1 cm) bovine.
Enzyme Activity	Factor X: None; Factor Xa: Proteinase (EC 3.4.21.6), hydrolyzes Arg-X bonds.
Coenzymes/Cofactors	For Factor Xa: Factor Va (in the activation of prothrombin to thrombin), membrane phospholipids (requires negatively charged phospholipids), Ca^{2+} ions. Factor Va catalyzes prothrombin activation by binding Factor Xa and prothrombin; all components bind to phospholipid membrane surfaces in the presence of Ca^{2+} ions.
Substrates	For Factor Xa: Naturally Occurring Proteins: prothrombin, Factor VII, Factor V, Factor VIII. Peptide Chromogenic Substrates: Methoxycarbonyl cyclohexylglycyl-Gly-Arg-pNA (Spectrozyme Xa), Methanesulfonyl-D-Leu-Gly-Arg-pNA (CBS 31.39), Bz-Ile-Glu (50 percent OME), Gly-Arg-pNA (S-2222), Bz-Ile-Glu (piperidine amide) Gly-Arg-pNA (S2337).
Inhibitors	For Factor Xa: Naturally Occurring Protein Inhibitors: Antithrombin III, α-1 proteinase inhibitor, soybean trypsin inhibitor (reversible), tissue factor pathway inhibitor (previously lipoprotein associated coagulation inhibitor or extrinsic pathway inhibitor), antistasin;

	Low Molecular Weight Inhibitors: Dansyl-Ile-Glu-Gly-Arg-CH_2Cl, Diisopropylphosphorofluoridate, Phenylmethanesulfonylfluoride (requires high concentrations). Antithrombin III irreversibly inactivates Factor Xa. In the presence of heparin, more specifically heparin with a high affinity for Antithrombin III, the inactivation rate for Factor Xa is increased as many as 1,000 times. α-1-proteinase inhibitor irreversibly inactivates Factor Xa albeit relatively inefficiently.
Biological Functions	Factor X is the circulating precursor to the proteinase (Factor Xa) that converts prothrombin to thrombin. Factor Xa also activates Factor VII and can convert Factor V to Factor Va and Factor VIII to Factor VIIIa. Factor X and Factor Xa bind to negatively charged membrane phospholipids via the γ-carboxyglutamate residues of the protein light chain; this requires Ca^{2+} ions.
Physiology/Pathology	Factor X is a necessary component in hemostatic blood coagulation. Decreased concentration in heterozygous deficiency can result in increased risk of bleeding.
Degradation	Cleavage of Factor X at Arg-52 by Factor VII or VIIa or by Factor IXa produces Factor Xa – the active proteinase. In vitro an autolytic cleavage at Arg-274 (numbered with respect to the heavy chain) produces β-Factor Xa, without alteration in proteolytic activity. Other proteolytic cleavages occur in vitro in both the heavy and light chains to produce forms of Factor Xa with markedly reduced enzymatic activity. A proteinase from Vipera russellii venom, which also cleaves at Arg-52, is widely used to activate Factor X in vitro for laboratory assay.
Genetics/Abnormalities	The gene for Factor X is found on chromosome 13q34; inheritance is autosomal recessive. Approximately 25 examples of hereditary Factor X deficiency have been reported. Amino acid substitutions have been reported in the Gla domain and in the heavy chain.
Half-life	40 hrs
Concentration	In blood, 15 nmol/L, 7–10 mg/L.
Isolation Method	Isolation of Factor X from human plasma is readily achieved by precipitation with barium chloride from plasma collected into sodium citrate anticoagulant, chromatography on heparin-Sepharose or sulfated dextran beads and anion exchange chromatography. Blue-Sepharose is also effective in separating Factor X from the other vitamin-K dependent proteins. Bovine Factor X isolation does not require chromatography on heparin-Sepharose or sulfated dextran beads.
Amino Acid Sequence	The Factor X molecule consists of 445 aa residues. The heavy chain N-terminal amino acid is serine and the C-terminal amino acid is lysine. The light chain N-terminal amino acid is alanine and the carboxyl terminal is arginine. The 11 γ-carboxyglutamate residues are located at positions 6, 7, 14, 16, 19, 20, 25, 26, 29, 32, 39 in the light chain. Aspartic acid 63 in the light chain is β-hydroxylated. In bovine Factor X, one tyrosine residue at position 18 in the activation peptide is O-sulfated. Approximately 40% of the circulating bovine Factor X has this additional post-translationally modified tyrosine. Two oligosaccharide chains are found in the activation peptide region.
Disulfides/SH-Groups	Twelve disulfide bonds are found in Factor X. Seven (7) intra-chain bonds are in the light chain, four (4) in the heavy chain and one (1) interchain bond links the light and heavy chains.

General References

Jackson, C. M. In: *Progress in Haemostasis and Thrombosis* 1984, **7:** 55–109.
DiScipio, R. G., et al. *Biochemistry* 1977, **16:** 698–706.
Miletich, J., et al. *Meth. Enzymol.* 1981, **80:** 221–228.

Ref. for DNA/AA Sequences

Fung, M. R., et al. *Proc. Natl. Acad. Sci. USA* 1985, **82:** 3591–3595.
Leytus, S. P., et al. *Biochemistry* 1986, **25:** 5098–5102.

Factor XI

Frank A. Baglia and Peter N. Walsh

Synonyms	Plasma thromboplastin antecedent (PTA)
Abbreviations	FXI
Classifications	Serine protease

Description

FXI is a glycoprotein present in plasma as a zymogen that, when converted by limited proteolysis to an active serine protease, participates in the contact phase of blood coagulation. It is unique among plasma coagulation enzymes since it exists as a homodimer composed of two identical polypeptide chains linked by disulfide bonds. FXI circulates in plasma noncovalently bound to high molecular weight (Mr) kininogen, a nonenzymatic cofactor that promotes FXI activation by FXIIa, a process that is accelerated when the three proteins are bound in a ternary complex to negatively charged surfaces, such as glass, kaolin, celite, or sulfitides in vitro or activated platelets in vivo. Therefore, FXI is referred to as a "contact protein" together with FXII, prekallikrein and high Mr kininogen, all of which participate in the surface-mediated contact phase reactions of coagulation. FXI shares considerable sequence homology with prekallikrein, suggesting they have evolved from a common ancestral gene. A platelet form of FXI exists in platelet membranes that is functionally and antigenically similar to plasma FXI but different from plasma FXI in mw and subunit composition.

Structure

The zymogen FXI (Mr approx. 143,000) is a disulfide-linked homodimer containing 5% carbohydrate including 0.6% hexose, 2.7% N-acetylhexosamine and 1.7% N-acetylneuraminic acid. Each of the two identical polypeptide chains contains 607 aa, each of which can be proteolytically cleaved by FXIIa at an internal Arg-369-Ile-370 bond during limited proteolytic activation to yield a heavy chain (369 aa) and a light chain (238 aa) that contain thN-terminal sequence Ile-Val-Gly-Gly, typical of serine proteases. Each heavy chain contains four tandem repeats of 90-91 aa, each of which contains 6 (or 7) Cys that form three internal disulfide bonds, thus forming four distinct domains within the heavy chain region. These four tandem repeats contain aa sequences that are 23-24% identical. The first tandem repeat (Glu-1-Ser-90) contains a domain (Val-59-Lys-83) identified as the high Mr binding site. A thrombin binding site exists in the A1 domain spanning residues Ala-45-Arg-70 that is contiguous with but separate and distinct from the high Mr kininogen binding site. A FXIa substrate-binding site consists of a sequence of aa (Ala-134-Leu-172) in the A2 domain of the heavy chain that contains three antiparallel β-strands connected by β-turns, which together comprise a continuous surface utilized for the binding of FIX. However, the sequence of aa, Asn-235-Arg-268, of the A3 domain of FXI comprises a contact surface for interaction with a platelet receptor. Finally, the sequence of aa from Ala-317-Gly-350 of the heavy chain of the A4 domain of FXI contain three peptide structures, possibly consisting of three antiparallel β-strands that together comprise a contact surface for interacting with FXIIa. The light chain of FXI contains the catalytic traid, including His-44 (or His-413), Asp-93 (or Asp-462), and Ser-188 (or Ser-557), and the presence of Asp-182 at the bottom of the substrate binding pocket indentifies FXIa as a typical trypsin-like serine protease. Potential N-linked glycosylation sites are present at Asn-72, Asn-108, Asn-335, Asn-432 and Asn-473.

Molecular Weight	175,000 (gel filtration), 160,000 (SDS-PAGE, nonreduced); 135,979 (cDNA seqeuncing) resulting in a total Mr of 143,000 including carbohydrate. After reduction, activated FXI appears in two chains of 50,000 and 30,000 daltons (SDS-PAGE).
Sedimentation Coeff.	6.96 S (sucrose gradient)
Isoelectric Point	8.9 - 9.1
Extinction Coeff.	13.4 (280nm, 1%, 1cm)
Enzyme Activity	FXI exists in normal human plasma as a zymogen that requires proteolytic activation by FXIIa at Arg-369-Ile-370 to develop serine protease activity with the exposure of two active sites, one in each of two identical polypeptide chains. Although trypsin can also effect this cleavage it also cleaves FXI at other sites with resultant loss of enzymatic activity. The catalytic components of FXIa include the active site (His-413, Asp-462 and Ser-557) and a typical trypsin-like substrate binding pocket. FXIa activates FIX by exposing a substrate binding site present in one of the four tandem repeats of the heavy chain and cleaving two internal polypeptide bonds (Arg-145-Ala-146 and Arg-180-Val-181) in FIX in the presence of Ca^{2+} ions. One unit of FXI is defined as the amount of activity present in 1.0 ml of normal human plasma.
Coenzymes/Cofactors	The requirement for kallikrein in FXIa formation is indirect, through activation of FXII by kallikrein. The requirement of high Mr kininogen for the activation of FXI is two-fold: it functions as a cofactor that accelerates the reciprocal proteolytic activation of FXII by kallikrein and prekallikrein by FXIIa and is required for activation of FXI by FXIIa on anionic surfaces in vitro or on the surface of activated platelets, where FXI, high Mr kininogen and FXIa have been shown to bind specifically to high-affinity, saturable receptors.
Substrates	Activated FXI recognizes FIX as its normal substrate. FXIa can be assayed using the amidolytic assay which utilizes the oligopeptide chromogenic substrate, pyro-Glu-Pro-Arg-pNA. The specific activity of FXIa against this substrate is 1.05 μMole of substrate hydrolyzed per min. per mg of enzyme. In addition, FXIa recognizes a variety of other amino acid, dipeptide and longer peptide thioester substrates.
Inhibitors	The best documented and most potent physiological plasma inhibitors of FXIa are α_1-protease inhibitor and antithrombin III-heparin complex whereas C1-inhibitor, α_2-plasmin inhibitor and α_2-macroglobulin are less potent inhibitors of FXIa. Two potent and specific FXIa inhibitors are present in and secreted from platelets, one of which is protease nexin II (Mr approx. 112,000) whereas the other is an approx. 8,500 Mr peptide referred to as the Platelet Inhibitor of FXIa (PIXI). Other inhibitors of FXIa include the serine protease active site inhibitor, diisopropylphosphofluoridate (DFP), soybean trypsin inhibitor and trasylol, whereas lima bean trypsin inhibitor and ovomucoid trypsin inhibitor do not inhibit FXIa.
Biological Functions	FXI participates in the early or contact phase of blood coagulation. Four proteins, FXII, prekallikrein, FXI and high Mr kininogen, have been shown to be the major factors required for the activation of surface-mediated pathways. FXI circulates in plasma in a noncovalent complex with high Mr kininogen which binds to specific, saturable, high-affinity sites on activated platelets in the presence of Zn^{2+} ions or can be adsorbed to anionic surfaces such as glass, celite, kaolin or sulfatides. A surface-

associated ternary complex of FXI, high Mr kininogen and FXIIa is formed and results in the proteolytic activation of FXI with formation of the light-chain associated active site of FXIa which remains disulfide-linked to the surface through high Mr kininogen. FXIa has also been shown to bind to activated platelets in the presence of high Mr kininogen and Zn^{2+} ions, where it activates FIX.

Physiology/Pathology

FXI is synthesized in the liver and secreted into the plasma as a zymogen that circulates in a complex with high Mr kininogen. Whereas hereditary deficiencies have been described for all four of the "contact factors" (FXI, FXII, high Mr kininogen and prekallikrein) resulting in defects of plasma coagulation in vitro, only FXI-deficiency is associated with a bleeding state. About half the patients with FXI deficiency experience post-operative or post-traumatic (but usually not spontaneous) hemorrhage, the other half appear to be hemostatically normal. Thus, FXI deficiency is remarkable for the extreme variation in bleeding tendencies seen in homozygotes with no correlation between the severity of bleeding experienced by the patient and the level of FXI activity or FXI antigen in the plasma. The kinetics of FIX activation by FXIa are similar to those for FIX activation by FVIIa-tissue factor, suggesting that both pathways are physiologically important and that FXI is an essential coagulation protein required for normal hemostasis.

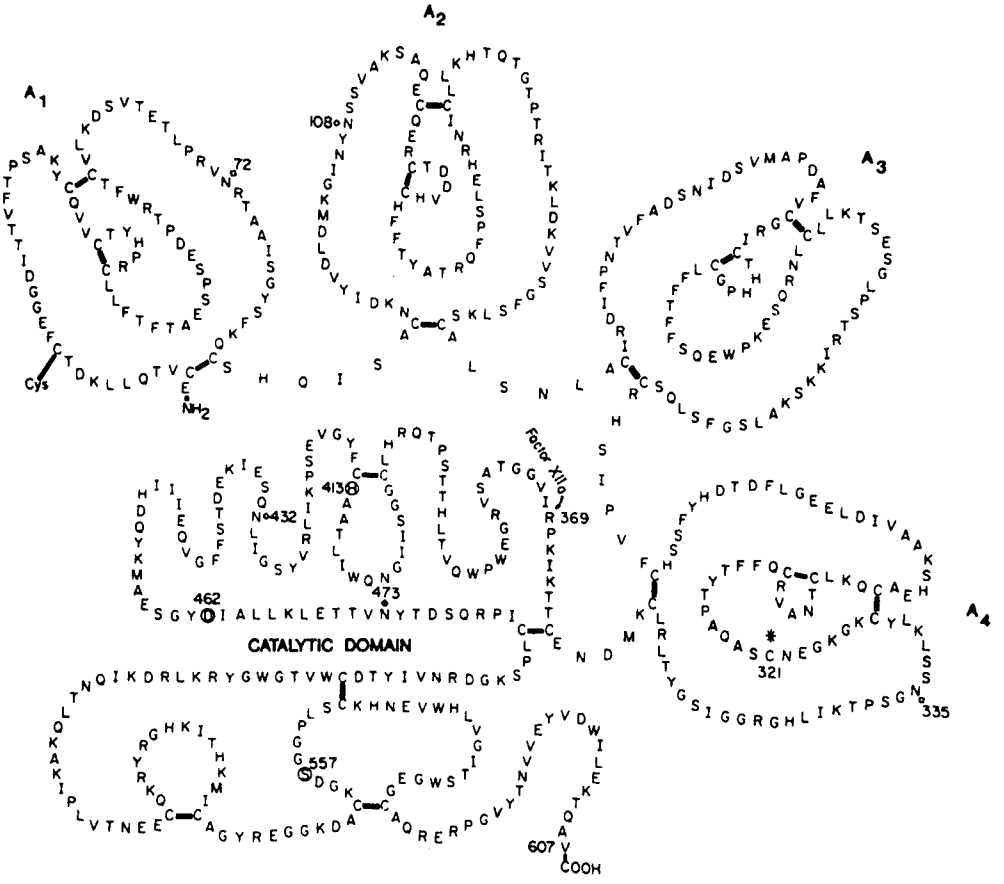
Degradation

The principal known mechanism of FXIa inactivation in vivo probably involves complex formation with the plasma serine protease inhibitors, α_1-protease inhibitor and antithrombin III, although two secreted platelet inhibitors (see above) may be important for regulating FXIa activity. The rate of biosynthesis and metabolism of FXI are equal to maintain a steady-state concentration in plasma, but the specific mechanisms of degradation of the zymogen are largely unknown.

Genetics/Abnormalities

FXI deficiency is inherited as an autosomal recessive trait characterized by low levels (0 - 10%) of plasma FXI antigen in homozygotes and intermediate levels in heterozygotes. Characteristically FXI antigen and activity levels correlate well with each other and FXI deficiency with cross-reacting material (CRM+) is rare. FXI deficiency has been estimated to occur with a frequency of 1:100,000 in the general population and constitute 7% of hereditary coagulation defects. However, the gene frequency in Ashkenazi Jews is very high (1 in 90 individuals in Israel), and very low in non-Jewish ethnic groups (1:1,000,000). Recent studies have estimated the risk for severe FXI deficiency in Ashkenazi Jews as 0.22% and Iraqi Jews 0.03% and that the estimated risk of heterozygosity in Ashkenazi Jews is 9.0% and Iraqi Jews 3.3%. The FXI gene, located on the distal end of the long arm of chromosome 4 (4q35), contains 23 kb of nucleotide sequence with 15 exons and 14 introns. Exon I codes for the 5' untranslated region, exon II for the signal peptide, exons III-X for the four tandem repeats of 90 - 91 aa and exons XI - XV for the catalytic domain. Analysis of cDNA sequencing data has identified several independent point mutations in the FXI gene including a mutation at an intron-exon boundary resulting in abnormal mRNA splicing (Type I), a nonsense mutation converting Glu-117 (GAA) to stop codon (TAA) and resulting in premature termination of translation (Type II) and a missence mutation resulting in the substitution of Phe-283 (TCC) by Leu (TCT) in the protein (Type III). Recently two novel mutations in non-Jewish FXI deficiency patients have been described. In Case 1, a T to G transition in exon 12 results in the substitution of Phe-442 by Val. In Case 2, a C to A transition in exon 5 results in the substitution of Cys-128 by a nonsense codon. Another study identified six other types of mutations that cause FXI deficiency. Two are point muta-

	tions that interfere with normal splicing of exons in the mRNA and four are point mutations that result in aa substitutions. One is a substitution of His for Asp-16 (Asp-16→His) near the N-terminal end of the protein. The other three aa substitutions (Leu-302→Pro, Thr-304→ile, Glu-323→Lys) are in the fourth Apple domain.
Half-life	60 - 80 hrs (infused FXI)
Concentration	4 - 6 mg/L (approx. 30 nM) in plasma. One unit is defined as the amount of FXI in 1 ml of normal human plasma. The specific activity of FXI is 200 - 300 units per mg protein.
Isolation Method	FXI has been purified from human plasma by conventional multistep procedures including ion exchange chromatography, gel filtration and affinity chromatography. More recent isolation techniques employ monoclonal antibodies to facilitate rapid and simple immunoaffinity purification of FXI with high specific activity.
Amino Acid Sequence	There are 607 aa present in each of the two chains of the mature protein, the sequence of which is shown in the molecular model below. Each heavy chain of FXIa (369 aa) contains four tandem repeats of 90 (or 91) aa plus a short connecting peptide. Each repeat forms a separate domain. The light chain of FXIa (each 238 aa) contains the catalytic portion of the enzyme with sequences that are typical of the trypsin family of serine proteases. The aa sequence of FXI shows 58% identity with human plasma kallikrein.
Disulfides/SH-Groups	Thirty-six 1/2-Cys residues are present in each FXI monomer. In the heavy chain of FXIa, the first and fourth repeats have 7 1/2-Cys residues, while 6 1/2-Cys residues are present in the second and third repeats. The dislufide linkages are shown in the molecular model below.
General References	Kurachi, K. et al. *Biochemistry* 1977, **16**:5831-5839. Sinha, D. et al. *Biochemistry* 1987, **26**:3768-3775. Davie, E.W. "The Blood Coagulation Factors: Their cDNAs, Genes and Expression" In: *Hemostasis and Thrombosis*, Colman, R.W. et al. (eds.). J. B. Lippincott Co., 1987, pp. 242-267. Asakai, R. et al. *Proc. Natl. Acad. Sci.* 1989, **86**:7667-7671. Baglia, F.A. et al. *J. Biol. Chem.* 1990, **265**:4149-4154. Baglia, F.A. et al. *J. Biol. Chem.* 1991, **266**:24190-24197. Baglia, F.A. et al. *J. Biol. Chem.* 1992, **267**:4247-4252. Baglia, F.A. et al. *J. Biol. Chem.* 1993, **268**:3838-3844. Baglia, F.A. et al. *J. Biol. Chem.* 1995, **270**:6734-6740. Baglia, F.A. et al. *J. Biol. Chem.* 1996, **271**:3652-3658. Shpilberg, O. et al. *Blood* 1995, **85**:429-432.
Ref. for DNA/AA Sequences	Fujikawa, K. et al. *Biochemistry* 1986, **25**:2417-2424. Asakai, R. et al. *Biochemistry* 1987, **26**:7221-7228. Imanaka, Y. *Brit. J. Haemat.* 1995, **90**:916-920. Pugh, R.E. et al. *Blood* 1995, **85**:1509-1516.



One of the two subunits is shown with the asterix on Cys 321 indicating the disulfide bond between the two subunits. The arrow indicates the cleavage site of FXIIa. The circled amino acids designate the catalytic triad. The diamonds indicate potential (open) or proven (closed) N-linked glycosylation sites. Reprinted with permission from B. A. McMullen, K. Fujikawa and E. W. Davie. *Biochemistry*, 1991, **30**: 2056–2060.

Factor XII

Bernhard Lämmle and Walter A. Wuillemin

Synonyms	Hageman factor
Abbreviations	FXII; HF
Classifications	EC 3.4.21.38; β-globulin
Description	A circulating plasma glycoprotein consisting of a single polypeptide chain (596 aa) and 16.8% carbohydrate (2 N-linked and up to 6 proposed O-linked saccharide side chains). An inactive zymogen to a serine protease.
Structure	Not yet crystallized. A model for the tertiary structure of the enzyme β-Factor XIIa has been proposed (based on the tertiary structure of bovine trypsin as determined by X-ray diffraction techniques). For primary structures of active enzymes, α-Factor XIIa and β-Factor XIIa, see "Biological functions".
Molecular Weight	65,722 (calculated Mr of polypeptide from aa sequence); 78,000 (calculated MW for Factor XII including 16.8% carbohydrate).
Sedimentation Coeff.	4.5 S (sucrose density gradient, 20°C)
Isoelectric Point	6.5 - 7.1
Extinction Coeff.	14.2 (280nm, 1%, 1cm)
Enzyme Activity	α-Factor XIIa proteolytically activates plasma prekallikrein and Factor XI in the presence of anionic surfaces (and high molecular weight kininogen). Kinetic parameters for prekallikrein activation by α-Factor XIIa (5 μg/ml dextran sulfate, pH 8.0, 37°C): k_{cat} 3.6 s^{-1}, Km 0.091 μM. β-Factor XIIa has enzymatic activity towards plasma prekallikrein in the fluid phase.
Coenzymes/Cofactors	Anionic surfaces (such as kaolin, ellagic acid, dextran sulfate, sulfatides...) and high-molecular weight kininogen required for optimal activation of prekallikrein and Factor XI by α-Factor XIIa.
Substrates	Plasma prekallikrein, Factor XI as probable physiological substrates for α-Factor XIIa; plasma prekallikrein as probable physiological substrate for β-Factor XIIa. Factor VII, plasminogen, C1 of the complement system may also be substrates of activated Factor XII. Chromogenic substrates H-D-Pro-Phe-Arg-pNA (S-2302) or Bz-Ile-Glu-Gly-Arg-pNA (S-2222) can be used for in vitro assay of activated Factor XII.
Inhibitors	C1-inhibitor is the main plasma inhibitor for α- and β-Factor XIIa. α_2-antiplasmin, antithrombin III, α_2-macroglobulin are probably of minor importance. Inhibition of α- and β-Factor XIIa by C1-inhibitor involves 1: 1 stoichiometric complex formation. Corn trypsin inhibitor is a rather specific inhibitor of activated Factor XII for in vitro use. Surface-dependent activation of Factor XII has been reported to be inhibited by platelet factor 4, β2-glycoprotein I, complement subcomponent C1q, and eosinophilic granulocytes.

1. Installment 1992 (revised 1997)

Biological Functions

The serine protease zymogens, Factor XII, plasma prekallikrein, Factor XI, and the non-enzymatic cofactor, high molecular weight kininogen, are referred to as contact activation factors. In the presence of anionic surfaces, Factor XII binds to the surface and becomes susceptible to proteolytic cleavage (activation) by plasma kallikrein. Alternatively, surface-dependent Factor XII autoactivation may generate initial amounts of α-Factor XIIa. Reciprocal activation of Factor XII and prekallikrein, non covalently complexed with high molecular weight kininogen, on the surface results in formation of plasma kallikrein and α-Factor XIIa, consisting of N-terminal heavy chain (Ile-1-Arg-353), disulfide-linked to C-terminal light-chain (Val-354-Ser-596). α-Factor XIIa-heavy chain mediates surface binding, α-Factor XIIa-light chain contains catalytic active site. Subsequent cleavage of α-Factor XIIa by kallikrein at Arg-334-Asn-335 and at Arg-343-Leu-344 results in formation of the protease β-Factor XIIa, consisting of a heavy chain (identical to α-Factor XIIa light chain), disulfide linked to a nonapeptide light chain (Asn-335-Arg-343). β-Factor XIIa dissociates from the surface. In purified systems surface-dependent Factor XII activation by plasma kallikrein and Factor XII autoactivation are enhanced by physiological concentrations of Zn^{2+}, and Zn^{2+} may even allow slow activation of Factor XII in the absence of anionic surface, probably by inducing a conformational change in Factor XII facilitating its proteolytic cleavage.

On the surface, α-Factor XIIa initiates the intrinsic coagulation cascade by activating F XI (bound to the surface as a bimolecular complex with high molecular weight kininogen). α-Factor XIIa activates prekallikrein to plasma kallikrein, thereby initiating bradykinin generation and intrinsic fibrinolysis. β-Factor XIIa dissociates from the surface and may activate prekallikrein in the fluid phase.

Factor XII may assemble together with high molecular weight kininogen, prekallikrein and Factor XI on the surface of neutrophilic granulocytes which may provide a circulating platform for the contact activation factors. Binding of Factor XII (and of the other contact activation factors) to endothelial cells has also been reported.

Physiology/Pathology

Decreased concentration of Factor XII prolongs plasma clotting time.

Severe hereditary Factor XII deficiency does not lead to hemorrhagic diathesis. Thus, the in vivo contribution of Factor XII to the initiation of coagulation is probably of minor importance. Several reports of thromboembolic complications including myocardial infarction in severely Factor XII deficient subjects led to the hypothesis that Factor XII deficiency may predispose to thromboembolic disease. Controlled studies suggest, however, that mild Factor XII deficiency is not a major risk factor for thrombosis.

Factor XII level is elevated in pregnancy and by estrogen use.

Contact system activation may occur in several disease states, such as hereditary angioedema attacks, bacterial sepsis, e.g. meningococcal septic shock in children, and adult respiratory distress syndrome.

Degradation

Physiologic catabolism unknown; fast clearance from plasma of complexes between activated Factor XII and C1-inhibitor.

Genetics/Abnormalities

The gene for Factor XII is located on chromosome 5q33-qter and is approximately 12 kb pairs in size, composed of 13 introns and 14 exons. Intron/exon organization is similar to that in the genes of the plasminogen activators, urokinase-plasminogen activator and tissue-plasminogen activator. Promoter contains an estrogen-responsive element.

Additional Taq I restriction site in intron B (due to point mutation exon 3: -224T→C) found in many unrelated congenitally Factor XII deficient

subjects. This mutation was found to be frequently associated with another point mutation in the 5' regulatory region (exon 1: -8G→C). An acceptor splice site mutation (nucleotide 11397 G→A) has been found in other Factor XII deficient subjects.
A few subjects with dysfunctional Factor XII molecules (cross-reacting-material-positive Factor XII deficiency) have been reported. The defect in Factor XII Washington D.C. is due to aa substitution Cys-571→Ser, whereas in Factor XII Locarno a kallikrein cleavage site is abolished by aa substitution Arg-353→Pro.

Half-life

50 - 70 hrs (plasma half-life)

Concentration

Pooled normal citrated plasma contains 24 mg/L (approx. 0.3 μM).

Isolation Method

Isolated from citrated plasma by ion exchange chromatography steps or affinity purification on immobilized monoclonal antibody.

Amino Acid Sequence

See Molecular Model below

Disulfides/SH-Groups

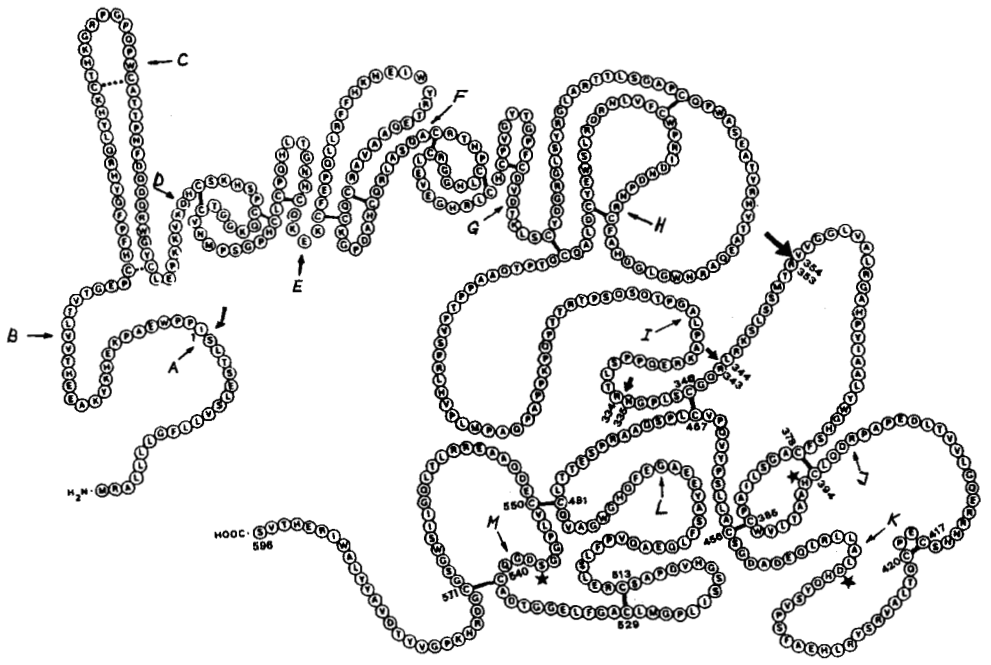
20 intrachain disulfide bonds (most of them tentatively located based on homologies with other proteins).

General References

Cochrane, C.G. and Griffin, J.H. *Adv. Immunol.* 1982, **33**:241-306.
Kaplan, A.P. and Silverberg, M. *Blood* 1987, **70**:1-15.
Tans, G. and Rosing, J. *Sem. Thromb. Hemostas.* 1987, **13**:1-14.
Bernardi, F. et al. *Hum. Genet.* 1988, **80**:149-151.
Lämmle, B. et al. *Thromb. Haemostas.* 1991, **65**:117-121.
Schousboe, I. *Blood* 1985, **66**:1086-1091.
Nuijens, J.H. et al. *Blood* 1988, **72**:1841-1848.
von Känel, R. et al. *Blood Coag. Fibrinol.* 1992, **3**:555-561.
Wachtfogel, Y.T. et al. *Thromb. Res.* 1993, **72**:1-21.
Bernardo, M.M. et al. *J. Biol. Chem.* 1993, **268**:12468-12476, 12477-12483.
Pixley, R.A. and Colman, R.W. *Meth. Enzymol.* 1993, **222**:51-65.
Henderson, L.M. et al. *Blood* 1994, **84**:474-482.
Koster, T. et al. *Br. J. Haematol.* 1994, **87**:422-424.
Farsetti, A. et al. *Endocrinology* 1995, **136**:5076-5083.
Wuillemin, W.A. et al. *Thromb. Haemostas.* 1995, **74**:1436-1441.

Ref. for DNA/AA Sequences

Fujikawa, K. and McMullen, B.A. *J. Biol. Chem.* 1983, **258**:10924-10933.
McMullen, B.A. and Fujikawa, K. *J. Biol. Chem.* 1985, **260**:5328-5341.
Cool, D.E. et al. *J. Biol. Chem.* 1985, **260**:13666-13676.
Cool, D.E. and MacGillivray, R.T.A. *J. Biol. Chem.* 1987, **262**:13662-13673.
Miyata, T. et al. *Proc. Natl. Acad. Sci. USA* 1989, **86**:8319-8322.
Kremer Hovinga, J. et al. *Blood* 1994, **84**:1173-1181.
Schloesser, M. et al. *Hum. Mol. Genet.* 1995, **4**:1235-1237.
Hofferbert, S. et al. *Hum. Genet.* 1996, **97**:838-841.



Primary sequence, position of introns, proposed domain structure, and cleavage sites in human factor XII (from Pixley and Colman 1993, modified from Cool and MacGillivray 1987)

One-letter code for amino acids (open circles), numbers refer to AA position. Solid bars indicate (tentative) disulfide bonds based on homologies with other proteins. Arrows A-M indicate positions of introns. Domains: fibronectin type II homology (B-D), epidermal growth factor region (D-E), fibronectin type I homology (E-F), epidermal growth factor region (F-G), kringle domain and proline rich connecting region (G-I), catalytic domain (I-COOH) with active site triad His-393, Asp-442, Ser-544 marked by asterisks. Curved arrow denotes signal peptidase cleavage site. Thick straight arrows are plasma kallikrein cleavage sites: Cleavage at Arg-353-Val-354 results in α-Factor XIIa, subsequent cleavages at Arg-334-Asn-335 and at Arg-343-Leu-344 generate β-Factor XIIa (see „Biological Functions", above). Solid squares are carbohydrate moieties attached to the polypeptide chain.

Factor B

Michael K. Pangburn

Synonyms	C3 proactivator; C3PA; Glycine-rich beta-globulin; GBG
Abbreviations	B
Classifications	Beta-1-globulin
Description	Factor B is a serine protease zymogen of 93,000 Da. Upon activation in the complement cascade it forms the enzyme which activates both C3 and C5 resulting in either direct killing of potential pathogens by complement or clearance by phagocytosis.
Structure	Factor B is a globular protein with between 7 and 10% carbohydrate. It has beta-1 electrophoretic mobility, a diffusion coefficient of 5.4×10^{-7}, and a frictional ratio of 1.28. In the electron microscope it appears as a three domain tightly packed molecule in the shape of a triangle. After activation one domain is removed and Bb appears as a dumbbell-shaped molecule of 60,000 Da.
Molecular Weight	93,000
Sedimentation Coeff.	5.9 S
Isoelectric Point	6.6
Extinction Coeff.	12.7 (280nm, 1%, 1cm).
Enzyme Activity	Factor B is a zymogen of the serine protease C3b,Bb (EC 3.4.21.47).
Coenzymes/Cofactors	For factor B to become an active enzyme it must first bind to the activated complement protein C3b. In this magnesium ion-dependent complex (written C3b,B) factor B is cleaved by factor D (another serine protease) and a 33kD fragment (Ba) is released. The remaining complex (C3b,Bb) is an active serine protease which is highly specific for converting C3 to C3a and C3b.
Substrates	In activated form in complex with C3b, the proteolytic subunit Bb cleaves complement proteins C3 and C5.
Inhibitors	None of the protease inhibitors in plasma inhibit the proteolytically active fragment Bb while it is bound to C3b, nor do they bind to the active site of the released Bb fragment. The biological ability of C3b,Bb to cleave C3 and C5 is inhibited by complement factor H. Factor H inhibits this protease by accelerating the release of Bb from the C3b,Bb complex.
Biological Functions	Forms central enzyme of the alternative pathway of complement activation. Cleavage of C3 and C5 releases the anaphylatoxins C3a and C5a and deposits C3b on the surface of the complement activating particle. Released C5b combines with C6, C7, C8 and C9 to form the cytolytic complex which disrupts biological membranes.
Physiology/Pathology	Factor B is important for humoral defense against bacterial and fungal infections and in the processing of immune complexes. No factor B deficiencies have been found leading to the conclusion that these are fatal.

Degradation	Factor B is degraded by a single proteolytic cleavage during activation of complement. Once the activated fragment decays from its complex with C3b it is no longer active. No known further degradation occurs.
Genetics/Abnormalities	Gene located in human chromosome 6 in the MHC class III region. It contains 18 small exons. Factor B is polymorphic, but only a small number of alleles have been found and all are functional.
Half-life	Approx. 3 days in humans.
Concentration	0.21 g/L. Normal range 0.17 to 0.26 g/L.
Isolation Method	Factor B is isolated from human plasma or serum by conventional chromatographic methods. See *Methods in Enzymology* 1981, **80**:102.
Amino Acid Sequence	One domain of the protein is homologous with other serine proteases. Three domains of approx. 60 residues each are homologous with the CCP domains found in most "complement control proteins".
Disulfides/SH-Groups	Unknown
General References	Reid, K.B.M. and Porter, R.R. *Annu. Rev. Biochem.* 1981, **50**:433-464. Muller-Eberhard, H.J. *Annu. Rev. Biochem.* 1988, **57**:321-347. Pangburn, M.K. and Muller-Eberhard, H.J. *Springer Semin. Immunopathol.* 1984, **7**:163-192. The Complement System. Rother, K. and Till, G.O. (eds.) Springer-Verlag, 1988.
Ref. for DNA/AA Sequences	Mole, J.E. et al. *J. Biol. Chem.* 1984, **259**:3407-3414. Horiuchi, T. et al. *Mol. Immunol.* 1993, **30**:1587-1592.

Factor D

Manuel Pascual and Jürg A. Schifferli

Synonyms	Adipsin, C3 proactivator convertase
Abbreviations	FD
Classifications	Electrical mobility: α_1-α_2-fraction
Description	A circulating plasma serine protease that catalyzes an essential step of the alternative pathway of complement. Factor D occurs in plasma in enzymatically active form (absence of a zymogen in plasma). Molecular studies have shown that factor D is identical to adipsin, a recently described serine protease found in normal adipose tissue. Factor D/Adipsin mRNA is expressed at high levels in human adipose tissue and in human lung. Factor D is synthetized in vitro by U937 and HepG2 cells. It is composed of a single polypeptide chain, with 40% homology with plasmin, 35% with pancreatic enzymes kallikrein, trypsin, chymotrypsin, and elastase and 30% with thrombin and factor X. The sequence of factor D shows a high proline content, and absence of phenylalanine.
Structure	A single polypeptidic chain consisting of 222 aa residues. The N-terminus of native factor D is buried inside the molecule, unavailable for reaction with monoclonal antibodies of predetermined specificity for that region of the enzyme. Crystallization and preliminary X-ray investigation of human factor D have been obtained.
Molecular Weight	23,748 (calculated from aa sequencing).
Sedimentation Coeff.	None
Isoelectric Point	7.4
Extinction Coeff.	16.0 (280 nm, 1%, 1 cm) (arbitrary value which corresponds to values determined by aa analysis).
Enzyme Activity	Serine protease that catalyzes the cleavage of C3b-bound factor B, thus completing the assembly of the C3 convertase of the alternative pathway of complement, C3bBb. No natural inhibitor of the enzymatic activity has been found. Irreversible inhibition of the enzymatic activity is produced by DFP. Factor D is a highly specific serine protease, hydrolysing a single arginyl-lysil bond in factor B, only when B is complexed with C3b. Using various substrates, it has been shown that its catalytic efficiency as a serine protease is low. The low esterolytic efficiency of purified D is compatible with the absence of a zymogen form. Binding of D to the C3bB complex or to the resulting C3bBb convertase has not been demonstrated.
Coenzymes/Cofactors	None
Substrates	C3b-bound factor B, with magnesium.
Inhibitors	None
Biological Functions	Essential for the activation sequence of the alternative pathway of complement: thus it participates to inflammatory reactions, and it plays a role in the host defense against infections and in the elimination of immune complexes. Its role in lipid metabolism is not clear: factor D/adipsin synthesis is deficient in several animal models of obesity, suggesting a role in normal adipocyte differentiation.

Physiology/Pathology	Factor D has the lowest serum concentration of all complement proteins: 2 mg/L. The extravascular/intravascular (EV/IV) ratio under normal circumstances is 8.5, indicating that factor D is predominantly extravascular (as expected from its synthesis by adipocytes). In normal serum, factor D is the rate-limiting enzyme of the alternative pathway of complement. Patients with end-stage renal failure (ESRF) have an approximately 10-fold increase in their serum factor D levels, that is responsible for an enhanced alternative pathway activation: the function of factor D is not modified by renal failure. In chronic renal failure, serum factor D correlates with serum creatinine. Accumulation of factor D in renal failure is due to an impaired normal renal catabolism, and it occurs manily intravascular (EV/IV ratio is diminished to 2.4 in patients with ESRF).
Degradation	Factor D is normally eliminated via glomerular filtration and it is catabolized after reabsorption in the proximal tubule. In patients with tubular dysfunction (Fanconi's syndrome), the urinary concentration of factor D is increased. The fractional metabolic rate (FMR) of factor D in normal individuals is very rapid (59,6% per hour) and it is essentially renal. In patients with chronic renal failure, the FMR is reduced, and it correlates well with creatinine clearance. Patients with ESRF have a ten-fold reduction in their FMR (5,7% per hour).
Genetics/Abnormalities	Only one case of complete (antigenic and functional) factor D deficiency has been reported, in a young man suffering from recurrent Neisseria infections. No abnormal variant has been described.
Half-life	One hour
Concentration	1.8 ± 0.4 mg/L
Isolation Method	Factor D can be isolated from plasma with ion exchange chromatography and gel filtration procedures. It is possible to purify the protein from the urine of patients with Fanconi's syndrome, with ion exchange chromatography, hydroxyapatite HPLC, and reverse phase HPLC. Another excellent source of factor D is represented by peritoneal fluid of patients on chronic ambulatory peritoneal dialysis: the purification is done in 3 steps, by ion exchange chromatography, heparin-Sepharose and MonoS HPLC.
Amino Acid Sequence	Unknown
Disulfides/SH-Groups	4 disulfide bonds; no free sulfhydryls.
General References	Catana, E. and Schifferli, J. A. Purification of human complement factor D from the peritoneal fluid of patients on chronic ambulatory peritoneal dialysis. *J. Immunol. Methods* 1991, **138:** 265–271. Lesavre, P. H. and Müller-Eberhard, H. J. Mechanism of action of factor D of the alternative complement pathway. *J. Exp. Med.* 1978, **148:** 1498–1509. Narayana, S. V. et al. Crystallization and preliminary X-ray investigation of factor D of human complement. *J. Mol. Biol.* 1991, **219:** 1–3. Pascual, M. et al. A. Metabolism of complement factor D in renal failure. *Kidney Int.* 1988, **34:** 529–536. Pascual, M. et al. Complement activation by the alternative pathway is modified in renal failure: the role of factor D. *Clin. Nephrol.* 1989, **32:** 185–193. Rosen, B. S. et al. Adipsin and complement factor D activity: an immune-related defect in obesity. *Science* 1989, **244:** 1483–1487. Volanakis, J. E. et al. Renal filtration and catabolism of complement protein D. *N. Engl. J. Med.* 1985, **312:** 395–399. Volanakis, J. E. Participation of C3 and its ligands in complement activation. In: *Current topics in microbiology and immunology* 1989, **153:** 1–21 (The third component of complement), Lambris, J. D. (ed.), Springer Verlag, Berlin.
Ref. for DNA/AA Sequences	White, T. et al. Isolation and characterization of cDNA clones for human complement factor D. *Complement Inflamm* 1989, **6:** 415 (abstract).

Factor H

Robert B. Sim

Synonyms	Complement Factor H; H; β_1H; Factor I-cofactor; C3b-inactivator accelerator
Abbreviations	
Classifications	Electr. mob.: β_1 fraction. Pseudoglobulin
Description	A single polypeptide chain plasma glycoprotein, synthesised mainly in liver but also in macrophage, endothelium. Present in platelets, and also as a membrane bound form on some leukocytes. Total mw 155,000 containing about 12% carbohydrate in the form of 6 Asn-linked oligosaccharides. Two variant forms, phi 1 and phi 2 reported to exist, probably differing in post-synthetic modification. Allelic variants, termed FH1 to FH5 have been identified by IEF. Two further "truncated" forms exist, which are translation products of different mRNA species arising by alternative splicing. One truncated form is a 49 kDa non-glycosylated species, the other is a glycoprotein of 39–43 kDa.
Structure	The single polypeptide chain of H is 1213 aa long and is divided into 20 homologous domains or modules, termed CCPs (complement control protein repeats: alternative name SCRs or short concensus repeats). Each CCP is about 60 aa long, and is folded mainly in β-sheet. The independently-folding CCPs are joined together as "beads on a string" forming an elongated thread-like structure up to 80 nm long.
Molecular Weight	Major form is 155,000 mw from aa sequence and carbohydrate composition. Variant forms phi 1 and phi 2 are of same apparent mw. Truncated forms are 49,000 and 39–43,000 mw from aa sequence and carbohydrate composition.
Sedimentation Coeff.	5.5 s for 155 kDa form
Isoelectric Point	5.1–5.7: each allele with a 5 to 7-band pattern
Extinction Coeff.	14.2 (experimentally determined) to 16.2 (calculated) (280 nm, 1%, 1 cm)
Enzyme Activity	Unknown
Coenzymes/Cofactors	None
Substrates	Binds to the C3b fragment of complement protein C3.
Inhibitors	Zn^{2+}ions
Biological Functions	Major regulatory protein of the complement system. When the complement system is activated by, e.g. micro-organisms, cell debris or immune complexes, the most abundant protein of the system, C3, is cleaved by complement proteases to form the fragment C3b, which can become covalently attached to the complement activator, or may remain free in solution. C3b can then bind factor B, which is in turn cleaved by factor D, to form the complex protease C3bBb. C3bBb itself cleaves more C3. To prevent complete consumption of C3, several regulatory proteins control the formation and stability of C3bBb, and the supply of C3b. Factor H is the major soluble protein regulating this process. Membrane-bound regu-

latory proteins, namely CR1 (complement receptor type 1), DAF (decay-accelerating factor) and MCP (membrane cofactor protein), which are all homologous to factor H, also regulate C3 consumption, by mechanisms similar to factor H. Factor H binds to C3b, and displaces Bb or prevents factor B from binding. Once factor H is bound to C3b, C3b is cleaved at two sites by the protease factor I, forming iC3b. Factor H is released unmodified after cleavage of C3b. iC3b formed from C3b covalently bound to the complement activator is the ligand for CR3 (complement receptor type 3), which mediates phagocytosis of the activator by macrophage. The C3b binding site in factor H involves 2 or more of CCP modules 3–6. The phi 2 form of factor H binds to a receptor on B-lymphoblastoid cells. A different factor H receptor may exist on myelomonocytic cells. Factor H is also involved in distinguishing between activators and non-activators of the complement alternative pathway by binding to sites of high charge density.

Physiology/Pathology

Rare inherited deficiencies of factor H occur. Absence of functional factor H leads to uncontrolled turnover of C3 and factor B and to secondary deficiency of C3. This in turn leads to susceptibility to bacterial infection and failure to clear immune complexes. Damage to tissues, particularly kidney, may arise from immune complex deposition.

Degradation

H is not normally degraded during complement activation and regulation. Plasmin cleaves H, to a minor extent, to a two chain form (38 kDa + 120 kDa fragments, disulfide-linked), which has slightly altered activity.

Genetics/Abnormalities

Human H is encoded by a single gene about 120 kb long on chromosome 1q31–32. This is part of the Regulation of Complement Activation (RCA) gene cluster in which are encoded other complement control proteins, CR1, CR2, DAF, MCP and C4bp, which are made up of the same structural domains (CCPs) as factor H. With few exceptions, each CCP is encoded in a single exon. Three different mRNA species arise from the gene by alternative splicing: 4.3, 1.8 and 1.4 kb. These are equally abundant in liver, and encode respectively the 155, 49 and 39–43 kDa proteins.

Half-life

4–5 days

Concentration

plasma: 155 kDa form is in the range 200–600 mg/L; 49 and 39–43 kDa forms are 1–5 mg/L

Isolation Method

Isolated from citrated plasma by polyethylene glycol precipitation, followed by ion-exchange, gel filtration and hydroxyl apatite chromatography. A more rapid method is by affinity chromatography on immobilised monoclonal antibody MRC OX23.

Amino Acid Sequence

The 155 kDa form of H contains 1213 aa, arranged into 20 homologous domains (CCPs) each about 60 aa long. Two C-terminal basic residues are trimmed off in circulation. Homologous CCP domains occur in many other complement proteins (CR1, CR2, MCP, DAF, C2, Factor B, C6, C7, C4bp, C1r and C1s) and in non-complement proteins such as coagulation factor XIII B chain, IL-2 receptor, β_2I. The 49 kDa form of H contains 7 CCPs, which are identical to the first 7 CCPs of the 155 kDa form of H, followed by a unique 4-residue C-terminal, SFTL. The 39–43 kDa form of H contains 5 CCPs, of which 3 are identical to CCPs 18–20 of the 155 kDa form.

Disulfides/SH-Groups

Each CCP contains 2 disulfide bridges: each 155 kDa monomer therefore has 40 disulfide bridges and no free SH groups.

General References

Whaley, K., and Ruddy, S. Modulation of the alternative complement pathway by β_1H globulin. *J. Exp. Med.* 1976, **144:** 1147–1167.
Fontaine, M. et al. Truncated forms of human complement factor H. *Biochem. J.* 1989, **258:** 927–930.
Ripoche, J. et al. Two populations of complement factor H differ in their ability to bind to cell surfaces. *Biochem J.* 1988, **253:** 475–480.
Perkins, S. J. et al. A study of the structure of human complement factor H by Fourier transform infrared spectroscopy and secondary structure averaging methods. *Biochemistry* 1988, **27:** 4004–4012.
Reid, K. B. M. et al. Complement system proteins which interact with C3b or C4b: a superfamily of structurally related proteins. *Immunology Today* 1986, **7:** 230–234.
Sim, R. B., and DiScipio, R. G. Purification and structural studies on the complement system control protein β_1H (Factor H). *Biochem. J.* 1982, **205:** 285–293.
Norman, D. G. et al. The three-dimensional structure of a complement control protein module in solution. *J. Mol. Biol.* 1991, **219:** 717–725.

Ref. for DNA/AA Sequences

Ripoche, J. et al. *Biochem. J.* 1988, **249:** 593–602 (4.3 kb and 1.8 kb mRNAs).
Estaller, C. et al. *J. Immunol.* 1991, **146:** 3190–3196 (1.4 kb mRNA).

Ferritin

Robert R. Crichton and Roberta J. Ward

Synonyms	Apoferritin (this refers to the ferritin protein without any iron present), 'native apoferritin' refers to that part of the ferritin in a preparation which is naturally without iron.
Abbreviations	None
Classifications	None
Description	An iron storage protein which is synthesized by all human cells. The highest concentration is found in liver parenchymal cells, with smaller amounts in reticuloendothelial cells, particularly in spleen, bone marrow and skeletal muscle. Apoferritin consists of 24 subunits of two classes – H and L. The protein shell is thus a hetero-polymer of variable subunit content ranging, in principle from $H_{24}L_0$ to H_0L_{24} with L-rich polymers predominant in liver and spleen and H-rich polymers in heart. Ferritins of variable H and L contents are found in other tissues. The protein shell, 12 nm in diameter, encloses a roughly spherical cavity of diameter 8 nm in which a variable amount of iron is deposited. The micellar iron core contains from zero to about 4,500 atoms iron, essentially as a ferric-oxy-hydroxide, similar in structure to the mineral phase ferrihydrite, $(9\,Fe_2O_3 \cdot 9\,H_2O)$. The iron mineral phase also contains significant amounts of phosphate.
Structure	The three-dimensional structures of recombinant human H and L apoferritins have been determined (Lawson et al., 1991). Structural features of the apoferritin molecule are presented in Fig. 1. In the centre the symmetrical quaternary structure of twenty-four equivalent subunits is presented; the N-terminus and the position of the C-terminal E helix are marked as N and E respectively. A ribbon diagram of one subunit is in the lower right of Fig. 1. It is an anti-parallel helix bundle of four helices A–D, which constitute the core of the molecule. Two of these helices, A and C, run in the same direction with their outer faces on the exterior of the protein shell, while the two others, B and D, run in the opposite direction, forming the internal surface of the apoferritin oligomer. The helices A–D interact along much of their length to form the essentially hydrophobic core of the subunit. The antiparallel helix dimers A–B and C–D are connected by the loop L, which together with its 2-fold related counterpart in a second subunit forms a section of antiparallel β-sheet within the dimer at the external surface of the apoferritin protein shell (Fig. 1, upper left). In addition to the helix bundle and the external loop L, there are two N- and C-terminal appendages. The N-terminal sequence of eight aa residues in human L ferritin projects towards the outside of the multimeric protein shell (it becomes twelve in H ferritin) whereas the C-terminal sequence forms a short helix E (residues 158–172) which lies almost perpendicular to the principal helix bundle, and which terminates in a dipeptide (hexapeptide in H subunits) oriented towards the internal cavity of the protein shell (reviewed in Crichton, 1991). Access to the interior of the protein may be possible by hydrophobic channels running down the 4-fold symmetry axes (lower left) or down the hydrophilic and negatively charged channels down the 3-fold symmetry axes.
Molecular Weight	21,099 for H apoferritin subunit; 19,766 for L apoferritin subunit. Hence, the mw of apoferritin heteropolymers present in different tissues can vary from about 506,000 to 474,000. The mw of human ferritins with their

	mineral phase of variable iron content, can vary from around 500,000 for iron-poor ferritins to values in excess of 900,000 for iron-rich ferritins.
Sedimentation Coeff.	18 (apoferritin); > 18 (with iron); 2 (subunit)
Isoelectric Point	4.6 (H-rich ferritins) – 5.7 (L-rich ferritins)
Extinction Coeff.	Not known for ferritin (function of iron content). Apoferritins would be expected to have an extinction coefficient around 10 (280 nm, 1%, 1 cm).
Enzyme Activity	Ferroxidase activity has been reported – H subunits appear to be more active than L.
Coenzymes/Cofactors	None
Substrates	Fe^{2+}
Inhibitors	Other transition metal ions such as Zn^{2+}, Tb^{3+}, Cr^{3+}, VO^{2+}.
Biological Functions	Storage of intracellular iron in a non-toxic, soluble and bioavailable form. Probably has a key role in intracellular iron homeostasis (see Physiology/Pathology), and serves as a source of available iron within cells. Can be used by reticuloendothelial cells, at least in vitro, as a means of exporting iron into the extracellular compartment where it can either transfer its iron to apotransferrin, or alternatively be taken up, for example by liver hepatocytes via plasma membrane ferritin receptors; its iron is then released within the cell to mitochondria and cytosol ferritin. Serum ferritin levels reflect body iron stores, except in acute phase response; little is known about its origin or its regulation.
Physiology/Pathology	Ferritin is an essential buffer compartment in which iron can be stored in non-toxic, yet bioavailable form. It plays an essential role in cellular iron homeostasis. Ferritin synthesis is regulated essentially at the translational level. There is a pool of ferritin mRNA within the cell which can be rapidly mobilized in conditions of iron excess for ferritin synthesis. This pool of mRNA is under the control of an IRE (iron regulatory element), a 29 residue stem loop in the 5′ untranslated region of all ferritin mRNA's analysed to date, which interacts with a specific protein, the IRE-BP (IRE-binding protein). When iron is available within the cell the IRE-BP is in a low affinity form which binds weakly to the ferritin mRNA and allows its translation. When intracellular iron levels are low, the IRE-BP is in a high affinity conformation, binds to the IRE of ferritin mRNA and prevents its expression. A reciprocal process functions with regard to the mRNA of the transferrin receptor where equivalent stem loops (IRE's) bind the IRE-BP in its high affinity form when intracellular iron concentrations are low, thus ensuring its protection against nuclease degradation, and assuring intracellular iron uptake. When iron is present within the cell at adequate levels, the IRE-BP switches to its low affinity form, allows transferrin mRNA to be degraded, and releases ferritin mRNA to be translated. This seems to be the basis of cellular iron homeostasis. In iron overload both serum and tissue ferritin levels increase. Serum ferritin levels are increased when acute phase response occurs, eg. in infection and inflammation. In iron deficiency they decrease.
Degradation	After its formation into clusters, ferritin is engulfed by lysosomes where the protein shell is degraded forming a water-insoluble intralysosomal iron-protein complex known as haemosiderin (Crichton and Ward, 1992).

Genetics/Abnormalities	H gene localised to chromosome 11, near 11q23, L to chromosome 19 near 19q13.3–13.14; the presence of many pseudogenes has been reported on many different chromosomes (at least 14 H-like and 2 L-like).
Half-life	Unknown (3–4 days in tissue, animal)
Concentration	Tissue ferritin levels are poorly established since one does not take biopsy samples from healthy patients – could be in the range of 1–2 μg/mg protein in liver, and at least 10-fold higher in patients with iron overload. Serum ferritin levels are considered normal in the range 10–250 μg/L. Higher values indicate iron overload, lower iron deficiency.
Isolation Method	Homogenize the tissue, treat at 75–80 °C for 10 min, precipitate ferritin at 55% saturation with $(NH_4)_2SO_4$, reprecipitation in the same conditions. After a third precipitation with $(NH_4)_2SO_4$, precipitate with $CdSO_4$(5%). Repeat twice. Pass on a Sepharose-6B column (ICSH Iron Panel, 1985). Ferritin rich human tissues include liver and spleen.
Amino Acid Sequence	See Fig. 2.
Disulfides/SH-Groups	One free cysteine Cys-126 (H + L), one free cysteine Cys-86 (H).
General References	Crichton, R. R. *Adv. Prot. Chem.* 1990, **40:** 281–36. Crichton, R. R. *Inorganic Biochemistry of Iron Metabolism.* Ellis Horwood, Chichester, 1991, pp. 131–172. Crichton, R. R. and Ward, R. J. In: *Iron and Human Disease.* Lauffer, R. B. (ed.) CRC Press, Boca Raton, 1992 in press. International Committee for Standardization in Haematology (Expert Panel on Iron) *Brit. J. Haematol.* 1985, **61:** 61–63. Lawson, D. M., et al. *Nature* 1991, **349:** 541–544.
Ref. for DNA/AA Sequences	Boyd, D., et al. *J. Biol. Chem.* 1985, **260:** 11755–11761.

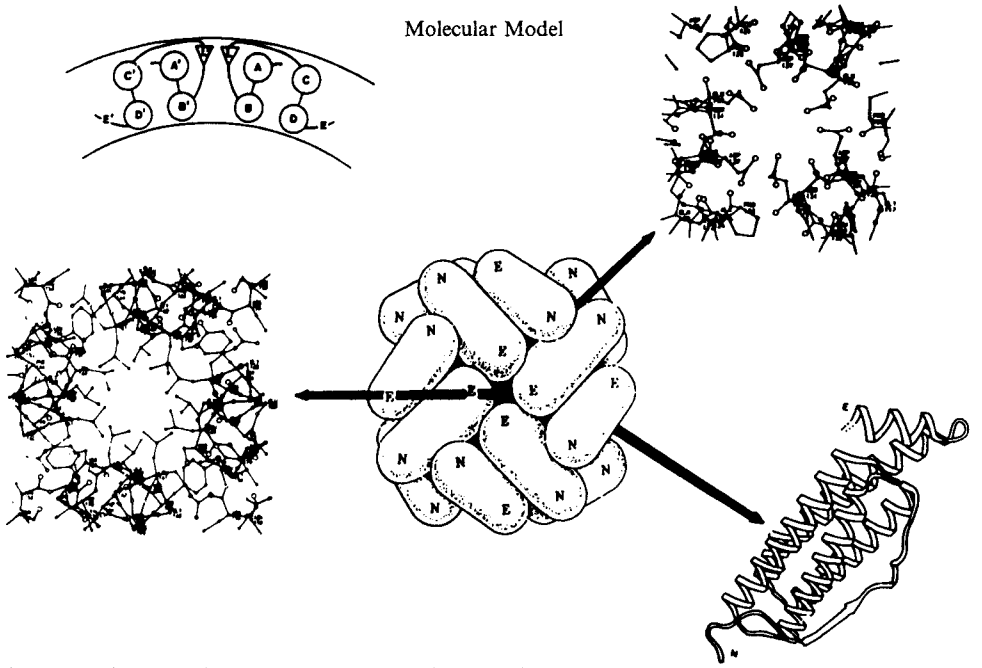


Figure 1 Molecular model according to P. M. Harrison et al (in *Iron Transport in Microbes, Plants and Animals*, ed. Winkelmann, G., van der Helm, D. & Neilands, J. B., VCH Publishers, Weinheim, 1987, pp. 445–475).
The central figure is the symmetrical quaternary structure of the apoferritin molecule composed of 24 equivalent subunits. Details are shown as; lower left a ribbon diagram of one subunit; upper right, region around 3-fold channel; lower left, region around 4-fold channel; upper left, end-on view of interface between 2-fold axis related subunit pairs.

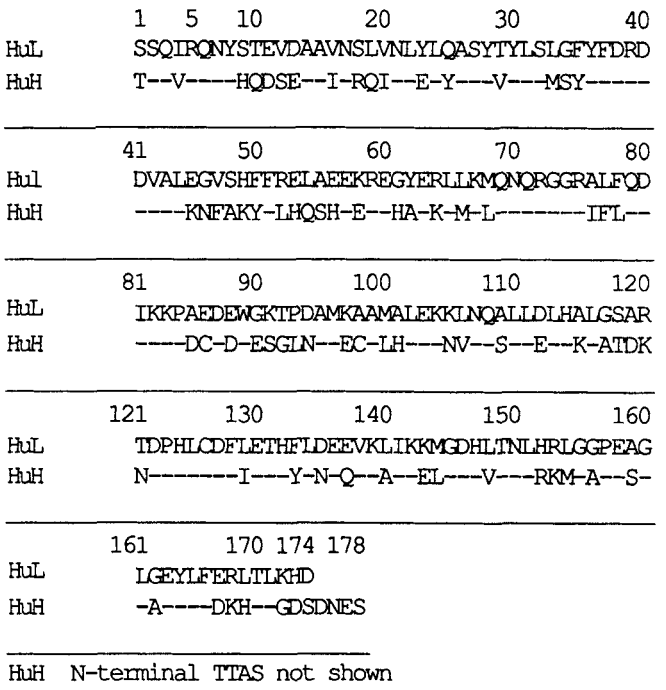


Figure 2 Amino acid sequences of human L (HuL) and H (HuH) subunits

Fibrinogen

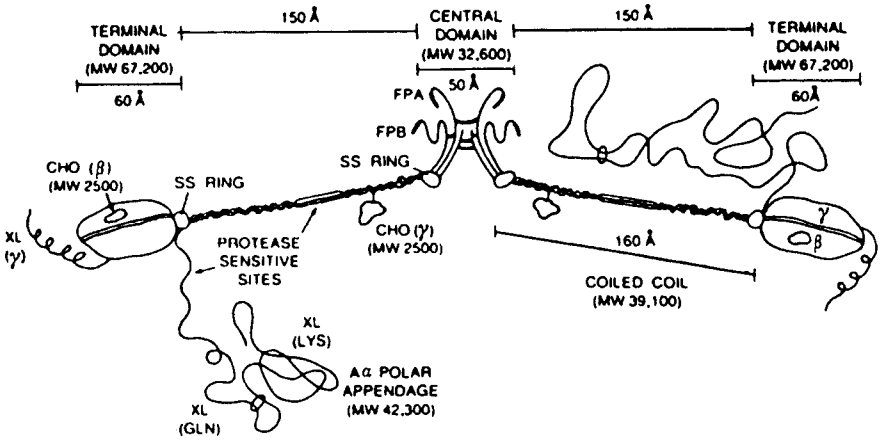
Miha Furlan

Synonyms	Factor I
Abbreviations	Fg; Fbg
Classifications	Electrical mobility: β_2-fraction
Description	A circulating plasma protein, syntesized in the liver and present in platelets (probably identical to plasma fibrinogen). A dimeric molecule: each of its two half-molecules contains the same three polypeptide chains (Aα- Bβ-, and γ-chains); all chains interconnected by disulfide bonds. A phosphorylated glycoprotein with approximately 3% carbohydrate in the form of 4 biantennary oligosaccharides located one on each Bβ- and γ-chain: (Bβ-chain: Asn-364, γ-chain: Asn-52). Sialic acid content heterogeneous (4.5 - 8 moles/mole of protein). Three high-affinity binding sites for calcium.
Structure	An elongated molecule with three nodules: 2 terminal D- and 1 central E-domain. Not yet crystallized in its native form.
Molecular Weight	340,000 (diffusion and sedimentation). Aα-chain: 66,100; Bβ-chain: 52,300; γ-chain: 46,500 (aa sequence); the actual mw of Bβ- and γ-chain should be incremented by 2200 for each fully sialylated carbohydrate side chain.
Sedimentation Coeff.	7.95 S
Isoelectric Point	5.1 - 6.3
Extinction Coeff.	15.1 - 15.5; Aα-chain: 11.8; Bβ-chain: 17.4; γ-chain: 20.4 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Participates in the blood coagulation process: Two fibrinopeptide A fragments (FPA mw 1600) from the N-terminal parts of the Aα-chains and two fibrinopeptide B fragments (FPB mw 1400) from the N-terminal parts of the Bβ-chains cleaved by thrombin - the cleavage transforms fibrinogen to a fibrin monomer. Removal of fibrinopeptides uncovers polymerization sites in the N-terminal domain that spontaneously bind to complementary sites on the C-terminal domains of other fibrin(ogen) molecules. The resulting protfibrils interact laterally to multistranded fibres. Crosslinking (γ-γ and α-α covalent crosslinks) mediated by activated factor XIII, stabilizes the fibrin clot. Has affinity for platelets (glycoproteins IIb-IIIa); heparin; some receptor-mediated bacteria; and proteins such as thrombin, fibronectin, factor XIII, plasminogen activators (t-PA, urokinase), α_2-antiplasmin, thrombospondin, von Willebrand factor, histidine-rich glycoprotein, collagen.

1. Installment 1992 (revised 1997)

Physiology/Pathology	Essential for blood coagulation; concentration below 0.5 g/L in blood may cause severe coagulation disorders. An acute phase protein.
Degradation	Normaly eliminated from circulation by the liver. May be degraded by plasmin, elastases and other proteases. Degradation products usually found in circulation arise from proteolysis of fibrin by plasmin. Major degradation products: D and E fragments from non-crosslinked fibrin, D-dimer from crosslinked fibrin.
Genetics/Abnormalities	Synthesized from three different mRNAs. Localization: distal third of the long arm of chromosome 4, band q23-q32. More than 250 inherited abnormal variants of fibrinogen described (July 1993). In about 90 cases, the molecular defect has been identified. the most frequent aa substitutions were Aα16 Arg → His and Aα16 Arg → Cys, interfering with the cleavage of fibrinopeptide A by thrombin. Genetic polymorphisms within the coding region: Aα312 Thr/Ala, Bβ448 Arg/Lys.
Half-life	3.5 - 4.5 days (blood circulation)
Concentration	Plasma: 2.6 g/L (range 2.0 - 4.0 g/L)
Isolation Method	Isolated from citrated plasma either by ammonium sulfate or ethanol precipitation, or by affinity chromatography on fibrin-Sepharose, Gly-Pro-Arg-Pro-Sepharose, or ristocetin-Sepharose.
Amino Acid Sequence	Human fibrinogen consists of 2964 aa residues (610, 461 and 411 in each of the two Aα-, Bβ- and γ-chains, respectively). Aα-chains contain each two RGD-sequences (95-97, 572-574), γ-chain shows partial homologies with kappa-casein and β-thromboglobulin. Heterogeneity: splicing of Aα- and γ-chains, phosphorylation (Aα-chain), sulfation (Bβ- and γ-chain), proline hydroxylation (Bβ-chain), glycosylation (Bβ- and γ-chain) and proteolytic degradation (Aα-chain).
Disulfides/SH-Groups	29 disulfides: 17 interchain (3 linking the two half-units) and 12 intrachain, 2x (1 Aα, 3 Bβ, 2 γ); no free sulfhydryls.
General References	Doolittle, R.F. A detailed consideration of a principal domain of vertebrate fibrinogen and its relatives. *Protein Sci.*1992, **1**:1563-1577. Henschen, A.H. Human Fibrinogen - structural variants and functional sites. *Thromb. Haemost.* 1993, **70**:42-47. Furlan M. "Structure of fibrinogen and fibrin." In: *Fibrinogen, Fibrin Stabilization, and Fibrinolysis. Clinical, Biochemical and Laboratory Aspects.* Francis, J.L. (ed.), VCH Verlagsgesellschaft, Weinheim, 1988; pp. 17 - 64. Shafer, J.A., and Higgins, D.L. Human fibrinogen. *CRC Crit. Rev. Clin. Lab. Sci.* 1988, **26**:1 - 41.
Ref. for DNA/AA Sequences	Chung, D.W. et al. *Biochemistry* 1983, **22**:3244-3250; 3250-3256. Rixon, M.W. et al. *Biochemistry* 1983, **22**:3237-3244. Fu, Y. et al. *Biochemistry* 1992, **31**:11968-11972.

Molecular model according to R. F. Doolittle, *Annu. Rev. Biochem.* 1984, **53**: 159–229.

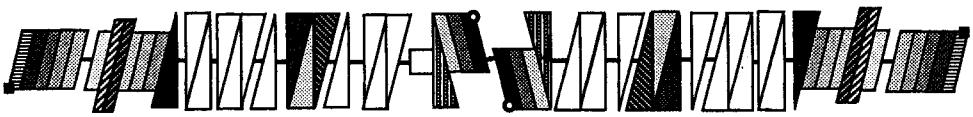


Fibronectin

Deane F. Mosher

Synonyms	Cold-insoluble globulin, Large external transformation-sensitive (LETS) protein, Surface fibroblast (SF) antigen, Anti-gelatin factor
Abbreviations	FN
Classifications	Electr. mob.: Slow α_2-globulin
Description	Circulating and extracellular matrix glycoprotein. FN is a disulfide-bonded dimer of subunits 2135–2446 aa in length. The subunits are comprised of three types of modules: 12 type I modules, 2 type II modules, and (depending on splicing of nascent mRNA) 15–17 type III modules. FN is synthesized as a prepropolypeptide and subjected to cleavage by signal and processing proteases, N- and O-glycosylation, phosphorylation, and sulfation in the secretory apparatus of the cell. Plasma FN is synthesized by hepatocytes and is a heterodimer of subunits with 15 type III modules plus or minus the alternatively spliced "V" (variable) region. Many other cell types synthesize and secrete FN. Subunits of this "cellular" FN may contain one or both of the extra type III modules and several variations of sequence from the "V" region. Thus, over 20 different types of heterodimers are possible.
Structure	Elongated structure 120 nm in length and 2.5 nm in diameter. The modules are thought to be arranged like "beads on a string", and the two subunits are held together by disulfides at their carboxyl-termini. At physiological pH and ionic strength, the "beaded string" folds up to form a more compact structure. NMR analyses of type I and II modules demonstrate β-sheet structure stabilized by hydrophobic cores of conserved cystines and aromatic residues.
Molecular Weight	460,000 to 500,000 (calculated by aa sequence and carbohydrate content and consistent with various experimental estimates).
Sedimentation Coeff.	12 S (physiological conditions)
Isoelectric Point	Approximately 5.0
Extinction Coeff.	12.8 (280 nm, 1%, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	FN forms thin aperiodic fibrils in tissues that are thought to serve as a "scaffolding" upon which cells adhere, move, and differentiate during embryogenesis, organogenesis, orderly tissue turnover, and wound repair. FN interacts with other components of extracellular matrix, including fibrin, collagens, and proteoglycans, and a number of integrin cell adhesion receptors, including $\alpha_5\beta_1$, $\alpha_4\beta_1$, $\alpha_3\beta_1$, $\alpha_v\beta_1$, $\alpha_v\beta_3$, and $\alpha_{IIb}\beta_3$. Although plasma FN binds many interesting substances and has been considered to be a non-specific opsonin, the major role of circulating FN is

	probably as a precursor to tissue fibronectin. Insolubilization of FN takes place at specialized sites on surfaces of cells and is accomplished by thiol-disulfide exchange to form disulfide bonded multimers or by covalent cross-linking catalyzed by transglutaminase.
Physiology/Pathology	A deficiency state has not been well defined, perhaps because FN is an absolutely essential glycoprotein. Plasma concentration tends to rise during inflammation and fall during starvation.
Degradation	Pathway(s) not defined. Although FN is susceptible to a number of proteases, literature on degradative fragments in vivo is sparse.
Genetics/Abnormalities	Single large (50–70 kb) gene on human chromosome 2 or mouse chromosome 1. Transcriptional control is complex and probably tissue specific. Nascent mRNA is differentially spliced.
Half-life	1–2 days (blood circulation); unknown (tissues)
Concentration	Plasma: 0.3 g/L (rises with age, greater in males than in females)
Isolation Method	Binds specifically to gelatin and can be isolated to near purity from a variety of sources by affinity chromatography on gelatin-agarose.
Amino Acid Sequence	The type I homology is found in blood coagulation Factor XII and tissue plasminogen activator. The type II homology is found in Factor XII, several bovine seminal proteins, and several metalloproteases. The type III homology is found in a large number of extracellular and intracellular proteins. FN contains an important RGDS cell adhesive sequence and several strong heparin-binding sequences.
Disulfides/SH-Groups	60 disulfides per heterodimer (58 intramodular and 2 intersubunit); 2 free sulfhydryls only accessible after denaturation
General References	Mosher, D. F. (ed.) *Fibronectin*. Academic Press, San Diego, 1989. Hynes, R. O. *Fibronectins*. Springer-Verlag, New York, 1990. Baron, M. et al. *Nature*, 1990, **345**: 642–646. Constantine, K. L. et al. *Biochemistry*, 1991, **30**: 1663–1672.
Ref. for DNA/AA Sequences	See general references. Human sequences are indexed under "Humfn" in GenBank listing.



Schematic model of plasma FN in its elongated conformation. The type I, II, and III modules are depicted as small parallelograms, larger parallelograms, and triangles, respectively. The alternatively spliced "V" region is depicted as a box. The two subunits of the heterodimer are joined near their carboxyl-termini.

Fructose-1,6-bisphosphate aldolase

Francesco Salvatore and Paola Costanzo

Synonyms	D-fructose-1,6-bisphosphate D-glyceraldehyde-3-phosphate-lyase; Fructose-1,6-bisphosphate-triosephosphate-lyase; Zymohexase; Aldolase
Abbreviations	FBP aldolase
Classifications	EC 4.1.2.13
Description	Aldolase is a glycolytic enzyme. Class I aldolase is found in plants and higher animals. In mammals there are three isozymes: aldolase A in skeletal muscle and most other adult tissues; aldolase B in adult liver; and aldolase C in adult brain. The native enzymes are a tetrameric combination of monomers that are homologous in isozymes A and B (four A monomers in aldolase A and four B monomers in aldolase B) and essentially heterologous in the C isozyme (A_3C, A_2C_2, AC_3, C_4). An AB hybrid set (A_4, A_3B, A_2B_2, AB_3) occurs in kidney. The monomers differ in amino acid composition and electrophoretic mobility. The A and B forms migrate towards the cathode, C migrates towards the anode, while the hybrid species have intermediate mobility. Class II aldolase, found mostly in moulds and bacteria, is composed of two subunits.
Structure	The three-dimensional structure of fructose-1,6-bisphosphate aldolase from rabbit and human muscle (aldolase A) has been determined by X-ray crystallography. The active protein is a tetramer of four identical subunits (A monomers), each composed of an eight-stranded alpha/beta-barrel structure. Close contacts between tetramer subunits are virtually all between regions of hydrophobic residues. The lysine residue responsible for Schiff-base formation with the substrate is located near the center of the barrel in the middle of the sixth beta strand. Also Asp-33 and Lys-146 seem to play a critical role in the reaction mechanism of aldolase. Site-directed mutagenesis of the evolutionarily conserved Asp-33 drastically reduces the activity of the enzyme, while apparently leaving the structure unchanged. Substitution of Lys-146 by various aa can affect multiple steps in the reaction pathway. Biochemical and crystallographic data suggest that the C-terminal region of the mammalian aldolase covers the active site pocket from the C-side of the beta-barrel and mediates access to the active site. For the aldolase A isozyme the C-terminal Tyr-363 is required to maintain a high level of catalytic activity towards fructose-1-,6-bisphosphate. Crystallography of the human liver aldolase B has revealed two main regions of differences between aldolase A and B: the C-terminus, which in aldolase B folds back into the barrel and the 6-phosphate sugar binding site, which is much more sterically hindered in the B isozyme.
Molecular Weight	Aldolase A 160,000, subunit: 40,000; aldolase B: 158,000, subunit: 39,000; aldolase C: 148,000, subunit: 37,000; class II aldolase: 80,000, subunit: 40,000 (sedimentation equilibrium).
Sedimentation Coeff.	7.5 S
Isoelectric Point	Unknown
Extinction Coeff.	Aldolase A: 9.2 in water (280nm, 1%, 1cm) Aldolase B: 8.5 in water (280nm, 1%, 1cm)

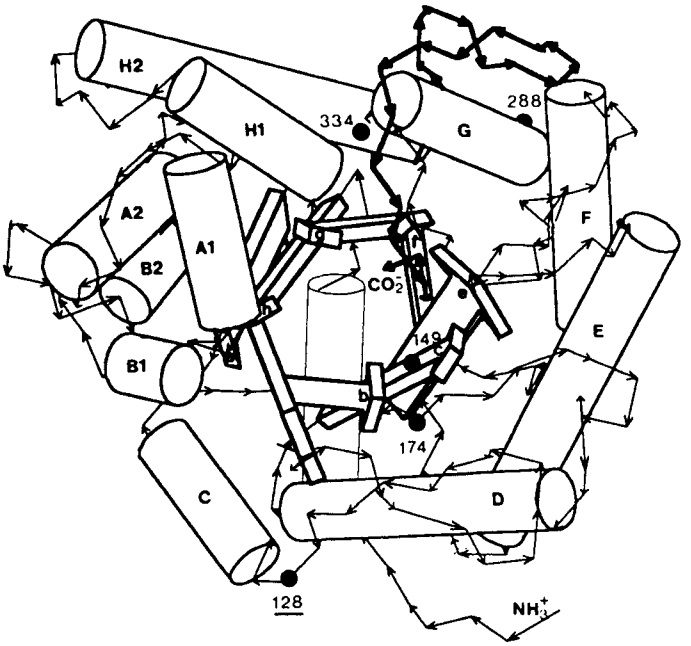
Enzyme Activity	Aldolase A catalyzes the reversible aldol cleavage of fructose-1,6-bisphosphate (FBP) into the triose phosphates D-glyceraldehyde-3-phosphate and dihydroxyacetone-phosphate. By splitting preferentially fructose-1-phosphate, aldolase B produces glyceraldehyde (not glyceraldehyde-3-phosphate). Class I aldolases form a Schiff-base intermediate with the substrate and contain a specific lysine residue (aa 229) at the active site that reacts with the carbonyl group of the substrate. The reaction mechanism of different isozymes of class I is identical, although the rate limiting step can vary. The slow step for the B isozyme is protonation of the carbanion intermediate, whereas the slow step for the A isozyme is the release of the second product dihydroxyacetone-phosphate. The aldolase A monomer is catalytically active and has a slightly lower Michaelis constant, clearly indicating that the quaternary structure is not required for catalysis. Class II aldolases require a metal cofactor such as Zn^{2+}. Arg-331 has been identified as a critical residue in class II FBP-aldolases. This residue is involved in the binding of the C-6 phosphate of the substrate to the enzyme.
Coenzymes/Cofactors	Class II aldolases require a metal ion cofactor. The metal ion appears to play a role in the polarization of the C-2 carbonyl group of the substrate, and also in the orientation of the C-1 phosphoryl group.
Substrates	Class I: D-glyceraldehyde-3-phosphate and dihydroxyacetonephosphate in gluconeogenesis; fructose-1,6-bisphosphate and fructose-1-phosphate in glycolysis. The specificity of the three isozymes towards the substrates is different. The activity ratio towards the two substrates (fructose-1,6-bisphosphate/fructose-1-phosphate) is 50, 25 and 1 for aldolases A, C and B, respectively. Class II aldolases show little or no activity with fructose-1-phosphate, and very high activity with FBP.
Inhibitors	Aldolase A is inhibited by heavy metals. A competitive inhibition is possible by glucose, fructose and fructose-6 phosphate. EDTA, o-phenanthroline and, at high concentrations, even cysteine for class II aldolase.
Biological Functions	Aldolase A is involved in glycolysis and gluconeogenesis. It has long been known that several glycolytic enzymes can interact with cytoskeletal proteins and it has been proposed that some may play structural and/or regulatory roles in the cytoplasm, in addition to their catalytic role. Aldolase A has one of the highest bound fractions to myofibrils, stress fibers and F-actin among the glycolytic enzymes. The catalytic site overlaps the actin-binding site topologically as well as functionally. A Ni^{2+} binding protein present in *Xenopus laevis* oocytes and embryos was identified as the monomer of fructose-1,6-bisphosphate aldolase A and raises the possibility that aldolase A is a target enzyme for metal toxicity. In a murine sarcoma cell line (SEWA) aldolase A is present in nuclei and interacts with DNA sequences. This feature suggests a novel role for this metabolic enzyme. Aldolase B degrades liver fructose-1-phosphate, the main source of which is dietary fructose. Aldolase B also shows higher affinity for triose phosphates with respect to fructose-bisphosphate, consistent with the role of liver in gluconeogenesis. The specific function of aldolase C in brain is unknown. Aldolase C is organized in cerebellar Purkinje cells into alternating bands with a high and low concentration of enzyme (zebrin). Aldolase C is abundant in a cytoskeletal fraction of porcine tracheal smooth muscle and it binds and releases inositol 1,4,5-trisphosphate. The large content of aldolase C in porcine tracheal smooth muscle supports an additional non glycolytic role of this protein.

Physiology/Pathology	Decreased aldolase B activity causes hereditary fructose intolerance (HFI). In HFI, dietary fructose accumulates and provokes liver damage that can lead to cirrhosis and malfunctioning of the gastrointestinal tract. Aldolase A deficiency in red cells is associated with haemolytic anaemia.
Degradation	A and B aldolases are, in vivo, a likely substrate for limited proteolysis by lysosomal cathepsins, particularly cathepsins B, M and D.
Genetics/Abnormalities	The three isozymes for aldolases A, B and C are encoded by three evolutionarily related but unlinked genes that are localized on chromosomes 16, 9 and 17, respectively. From the aldolase A gene are transcribed at least three mRNAs with different 5' non coding sequences that are expressed in a tissue-specific manner. For aldolases B and C, only one mRNA has been precisely characterized so far. Hereditary fructose intolerance is a recessive autosomic disease caused by a deficit of aldolase B activity in the liver. At gene level the disease is heterogeneous. Deletions in the gene or point mutations cause single aa changes or frameshift mutation. The alteration in the aldolase A gene that is associated with haemolytic anaemia is a point mutation in the codon for aa 128 (Asp→Gly) that renders the enzyme thermolabile.
Half-life	In serum: 21 ± 2 hrs. In liver: 2.8 days.
Concentration	In serum the aldolase activity is: 8 ± 4 U/L for adult males and children, 4.7 ± 3.2 U/L for females. (One unit of aldolase activity is the amount of enzyme that catalyzes the aldol cleavage of 1 μmole of D-fructose-1,6-bisphosphate in one minute). Aldolase is also found in several tissues in different quantities.
Isolation Method	The enzyme is easily purified and crystallized from mammalian muscle and liver by alcohol fractionation, ammonium sulfate precipitation and affinity chromatography. In man, aldolase A has been purified from skeletal muscle, aldolase B from liver and aldolase C from brain.
Amino Acid Sequence	The amino acid sequences of aldolases A, B and C are known from man and other mammals, obtained by direct protein sequencing or by translating the cDNA sequences. The sequences of the three aldolase isozymes are all related (homology within the same species, e.g. man, is around 70%). The homology is spread along the whole protein molecule. Around the active site (aa 215-241) there is a greater conservation of the sequence, the major differences are observed in the carboxyterminus (58% homology between A and C; 42% between B and C). The homology is much higher than 70% when each aldolase (A or B or C) is compared among various animal species.
Disulfides/SH-Groups	None
General References	Salvatore, F., Izzo, P., Paolella, G.: Aldolase gene and protein families: structure, expression and molecular pathophysiology. In: *Horizons in Biochemistry and Biophysics.* Blasi, F. (ed.), John Wiley, Chichester 1986, Vol. **8**: pp. 611-665. Gamblin, S.J. et al. *J.Mol.Biol.* 1991, **219**:573-576. Cross, N.C.P. and Cox, T.M. Int. *J. Biochem.* 1990, **22**:685-689. Takasaki, Y. et al. *J. Biochem.* 1990, **108**:153-157. Morris, A.J. et al. *J. Biol. Chem.* 1993, **268**:1095-1100. Littlechild, J.A. et al. *TIBS* 1993, **18**:36-39. Dalby, A. et al. *Protein and Peptide Letters* 1996, in press. Beernink, P.T. et al. *Proc. Natl. Acad. Sci. USA* 1996, **93**:5374-5379.

Qamar, S. et al. *Protein Science* 1996, **5**:154-161.
Wang, J. et al. *J. Biol. Chem.* 1996, **271**:6861-6865.
Antonijczuk, K. et al. *Biochem. Biophys. Acta* 1995, **1247**:81-89.
Ronai Z. et al. *Biochem. Biophys. Acta* 1992, **1130**:20-28.
Ahn, A.H. et al. *Development* 1994, **120**:2081-2090.
Baron, C.B. et al. *J. Biol. Chem.* 1995, **270**:20459-20465.

Ref. for DNA/AA Sequences

Tolan, D.R. and Penhoet, E.E. *Mol. Biol. Med.* 1986, **3**:245-264.
Maire, P. et al. *J. Mol. Biol.* 1987, **197**:425-438.
Izzo, P. et al. *Eur. J. Biochem.* 1987, **164**:9-13.
Izzo, P. et al. *Eur. J. Biochem.* 1988, **174**:569-578.
Buono, P. et al. *Eur. J. Biochem.* 1990, **192**:805-811.
The nucleotide sequence has been deposited with the EMBL Data Library (Accession number XO7292).



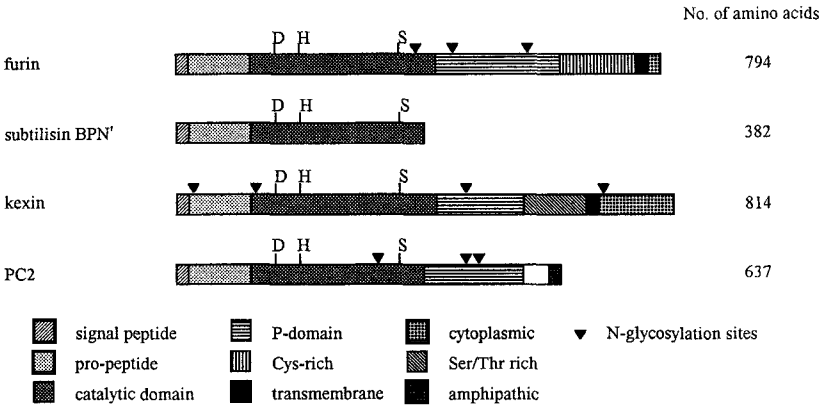
Schematic illustration of the polypeptide chain structure from human aldolase A. The helices are shown as barrels, the strands of β-sheet as arrows and the positions of hereditary mutations (aa 128 for haemolytic anaemia, aldolase A; aa 149, 174, 288 and 334 for HFI, aldolase B) with filled circles. (Reproduced by courtesy of Dr. J. Littlechild; see also Gamblin et al., J. Mol. Biol. 1991, **219:** 573–576.)

Furin

Elizabeth C. Ledgerwood and Stephen O. Brennan

Synonyms	PACE (Paired basic Amino acid Cleaving Enzyme), convertase, processing endoprotease
Abbreviations	None
Classifications	Serine protease (mammalian subtilisin homologue)
Description	A microsomal membrane bound endoprotease located primarily in Golgi/secretory vesicles. Ubiquitously expressed but highest levels are found in the liver and kidney. Involved in the processing of propeptides through proteolytic cleavage after paired and multi-basic sequences. Probably identical to the Ca^{2+}-dependent proalbumin convertase.
Structure	Unknown
Molecular Weight	Predicted translation product 86.7kDa not including possible CHO side chains. Predicted mature protein approximately 90kDa.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	A Ca^{2+} dependent subtilisin-like serine protease with the characteristic Asp, His, Ser catalytic triad. Cleaves polypeptide chains in the secretory pathway after an RXXR or RR sequence. There is no cleavage if the P_2 residue is a His or if the P_1 residue is basic or a large alkyl. pH optimum 5.5–6.0.
Coenzymes/Cofactors	Ca^{2+} absolute requirement for activity (mM). Two putative Ca^{2+} binding sites.
Substrates	Plasma proteins (e.g. proalbumin), vitamin K-dependent coagulation factors, receptors (e.g. insulin receptor). Putative role in the processing of some retroviral envelope glycoproteins (e.g. HIV-1 gp160) and cellular toxins (e.g. anthrax toxin protective antigen).
Inhibitors	Only naturally occurring inhibitor is α_1 antitrypsin Pittsburgh. Inhibited by EDTA but not by classical serine protease inhibitors such as PMSF or DIPFP.
Biological Functions	Removal of the pro-region from proteins as they transit the secretory pathway.
Physiology/Pathology	No documented human case lacking furin. A patient with antitrypsin Pittsburgh, an *in vivo* furin inhibitor, was found to have circulating proalbumin. The LoVo cell line has a mutation in the P domain of furin and fails to process proproteins. There exist a number of cases where a mutation at the cleavage site results in lack of processing by furin and secretion of the proprotein, sometimes with serious effects. eg Factor IX Oxford and factor IX Cambridge (causing hemophilia), unprocessed insulin pro-receptor (causing severe insulin resistance). Also a number of circulating proalbumin variants that escape furin processing have been isolated. The mutation

	in fibrinogen Canterbury (Aα 20 Val→Asp) introduces a new furin cleavable sequence of – R G P R D –. This results in the removal of the activation peptide and the primary polymerisation site and causes a mild bleeding problem.
Degradation	Unknown
Genetics/Abnormalities	cDNA's for human, rat and mouse furin have been sequenced. In human, the gene assignment is 15q25–26; a single copy gene located upstream of the *fes/fps* proto-oncogene. Gene has 17 exons, two non-coding. A 4.5kb mRNA transcript is detected in most tissues with highest levels in liver and kidney.
Half-Life	Unknown
Concentration	Unknown
Isolation Method	Partially purified from Triton extracts of hepatic Golgi/secretory vesicles. A purified truncated form, lacking the transmembrane domain, has been obtained by expression in CHO cells.
Amino Acid Sequence	794 aa deduced from the human cDNA sequence. Rat and mouse furin cDNA sequences also known. Catalytic domain homologous to the subtilisin family of serine proteases. Other homologous eukaryote members of this family include PACE4, PC1/3, PC2, PC4, PC5/6A, PC6B (all mammalian) and kexin (*Saccharomyces cerevisiae*).
Disulfides/SH-Groups	Has a cysteine-rich domain of unknown function, not essential for activity.
General References	Bloomquist, B. T. and Mains, R. E. The eukaryotic prohormone-processing endoproteases. *Cell. Physiol. Biochem.* 1993, **3**:197–212. Brennan, S. O. and Peach, R. J. Calcium-dependent Kex2- like protease found in hepatic secretory vesicles converts proalbumin to albumin. *FEBS Lett.* 1988, **229**:167–170. Smeekens, S. P. Processing of protein precursors by a novel family of subtilisin-related mammalian endoproteases. *Bio/Technol.* 1993, **11**: 182–186. Steiner, D. F., Smeekens, S. P., Ohagi, S. and Chan, S. J. The new enzymology of precursor processing endoproteases. *J. Biol. Chem.* 1992, **267**: 23435–23438.
Ref. for DNA/AA Sequences	Van den Ouweland, A. M. W. et al. Nucleotide sequence of the human *fur* gene. *Nucl. Acids Res.* 1989, **17**:7101–7102. Barr, P. J. et al. cDNA and gene structure for a human subtilisin-like protease with cleavage specificity for paired basic amino acid residues. *DNA and Cell Biol.* 1991, **10**:319–328.



Schematic representation of furin, subtilisin BPN' and two related members of this family, kexin and PC2. Active site aspartic acid, histidine and serine are shown.

Human Protein Data
Edited by André Haeberli

Gamma-Glutamyltransferase

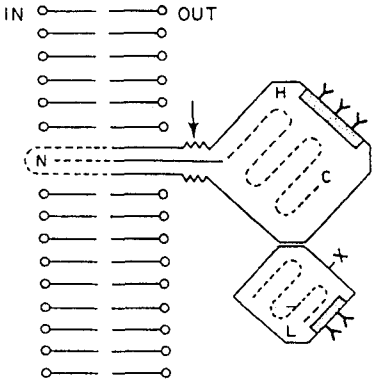
Giuseppe Castaldo and Lucia Sacchetti

Synonyms	gamma-glutamyltransferase, gamma-glutamyltranspeptidase, glutamine: D-glutamyl-peptide 5-glutamyltransferase, (5-glutamyl)-peptide:amino acid 5-glutamyltransferase
Abbreviations	GGT, gamma-GT, GT
Classifications	EC 2.3.2.2
Description	GGT plays a fundamental role in the "gamma-glutamyl cycle", which has been implicated in the amino acid transport across cell membrane. The enzyme is localized on the external cellular membrane in tissues with secretory or absorptive functions where a high aa uptake occurs, e.g. kidney (mainly proximal tubules), pancreas, epididymis, crypt cells of the intestinal brush border, liver, brain, mammary gland, the ciliary and retinal pigment epithelium of the eye and choroid plexus. Fetal liver, like other fetal tissues, contains more GGT than the adult liver; during carcinogenesis, large amounts of the enzyme are reexpressed in liver cells. The enzyme is present in such body fluids as urine, bile, seminal fluid, amniotic fluid, colostrum, and, in small amounts, in blood. Nearly all GGT in blood originates from the liver.
Structure	GGT is a dimeric glycoprotein and consists of a light subunit (catalytically active) and a heavy subunit that is bound, via a short hydrophobic N-terminal sequence (29 aa), to the membrane. The two subunits are non-covalently bound and are synthesized simultaneously from a single mRNA as a single-chain precursor, cotranslationally inserted in the cell membrane. In the heavy subunit, six possible N-glycosylation sites have been described, while one is present in the light subunit. The rate and quality of GGT glycosylation is organ-specific; in addition, differences in the saccharidic moieties have been described in GGT forms within the same tissue, and in neoplastic cells as compared to the normal counterparts.
Molecular Weight	Light subunit: 20,000 Da (non glycosylated) Heavy subunit: 38,336 Da (non glycosylated)
Sedimentation Coeff.	Unknown
Isoelectric Point	3.85 (human liver enzyme). The I. P. differs greatly among the different forms of GGT extracted from the same or different tissues (mainly due to the different glycosylation); in addition, the I. P. of GGT forms extracted from neoplastic cells is different from that of the normal counterparts; this is also due to the different sialylation of the enzyme in neoplastic cells.
Extinction Coeff.	Unknown
Enzyme Activity	GGT acts with a "ping-pong bi-bi" mechanism; it is the only known enzyme that catalyzes the first reaction of the breakdown of glutathione; in this reaction, the gamma-glutamyl moiety of glutathione is transferred to an acceptor that can be an aa, a dipeptide or glutathione itself. The acceptor can also be water, in which case a hydrolytic reaction occurs. The three general types of these reactions are: 1) Transpeptidation: Glutathione + L-amino acid ⇆ L-gamma-glutamyl-L-amino acid + L-Cys-Gly

	2) Autotranspeptidation: 2 Glutathione ⇆ L-gamma-glutamyl-glutathione + L-Cys-Gly 3) Hydrolysis: Glutathione + H_2O → L-glutamate + L-Cys-Gly Monovalent cations activate GGT (when gamma-glutamyl-p-nitroanilide or gamma-glutamylnaphthylamide are donor substrates).
Coenzymes/Cofactors	None
Substrates	(donors or acceptors): glutathione (also disulfide and S- conjugates); aas, dipeptides and tripeptides; gamma-glutamyl-p-nitroanilide; poly-gamma-glutamyl derivatives; glutamine; ophthalmic acid; leukotriene C4, D4 and E4; hydroxilamine; homoarginine; L-isoglutamine; glycinamide; L-methionine methyl ester; mesolanthionine; L-cystathionine; DL-allocystationine; DL-methionine; S-methyl-L-cysteine; L-gamma-glutamyl-7-amino-4-methyl coumarin.
Inhibitors	Reversible inhibitors: Sulfophthalein derivatives; sulfobromophtalein; bromocresol green; acetazolamide; L- and D-serine combined with borate; gamma-glutamylhydrazones; gamma-glutamylphenylhydrazides. Irreversible inhibitors: Iodoacetamide; phenobarbital; thiobarbituric acid; glutamine antagonist (L-azaserine, DON, AT 125, etc.); phenylmethane-sulfonyl fluoride.
Biological Functions	The enzyme is involved in the "gamma-glutamyl cycle" by which aa transport across the cell membrane is regulated. The enzyme is also involved in the interorgan glutathione transport by which cysteine is distributed in all tissues. GGT is a key enzyme in the processes of cellular detoxication; it is involved in mercapturic acid biosynthesis, in the metabolism of estrogens, peptidoleukotrienes (conversion of leukotrienes C4 in D4) and prostaglandins.
Physiology/Pathology	Increased serum levels of GGT are mainly observed in hepatobiliary diseases. The highest values are found in obstructive diseases. High values are also observed in chronic liver diseases. In addition, several drugs (alcohol, steroids, aflatoxins, barbiturates and caffeine) are strong inducers of GGT synthesis, mainly in liver cells, thus the enzyme induction seems to be tissue-specific. Also useful is GGT evaluation in other biological fluids. e.g., in urine, for the monitoring of renal transplants and for the diagnosis of several infectious kidney diseases, and in amniotic fluid for the prediction of neural tube defects. GGT isoenzymes described so far differ for: 1) Genetic variability, due to the possible presence, in serum, of GGT forms with slight differences in aa sequence; these forms have not yet been studied for diagnostic purposes. 2) Glycidic variability, due to the presence of GGT forms with differences in glycidic moieties. These forms have different electrophoretic mobility and so they can be easily evaluated in serum for diagnostic purposes. 3) Variability, due to the presence in serum of complexes between the enzyme and lipoproteins having different electrophoretic mobility and molecular weight (up to 1,000,000 Da). These different GGT forms are analyzed by the selective precipitation of the various classes of lipoproteins. Complexes between GGT and LDL or VLDL (resulting in isoforms with a beta-globulin or gamma-globulin electrophoretic co-migration) are typically evident in serum from patients with obstructive hepatobiliary diseases, while complexes between the enzyme and HDL are more typically associated to chronic non obstructive hepatobiliary diseases. In normal subjects a hydrophilic enzyme, not bound to lipids, is usually found.

Degradation	Serum GGT mainly derives from liver; it is cleared from blood by the liver sugar receptor, and eliminated through the biliary fluid. Thus GGT clearance is strongly dependent on the amount and type of glycosylation of the enzyme. In addition, the complex between GGT and lipoproteins, usually present in serum from hepatobiliary patients, shows a clearance rate much slower than the hydrophilic form present in normal sera. Small amounts of the enzyme are catabolized by the kidney.
Genetics/Abnormalities	In humans, seven potential, closely related GGT genes have been identified. Of these, three map on chromosome 22, two being located between the centromere and the BCR gene (a gene involved, after a chromosomal translocation, in some forms of human leukemias), the third one is telomeric to the BCR gene. The others have been mapped on chromosomes 18, 19 and 20. At least five of the seven GGT potential genes are expressed. One of those mapping on chromosome 22 is ubiquitously expressed in humans; the other four are tissue-specifically expressed also depending on a tissue specific enhancer of the gene. The remaining two genes could be pseudogenes. A single case has been described in which a congenital deficiency of the enzyme was postulated. The patient had an increased urinary concentration of glutathione.
Half-life:	9 h (hydrophilic form); 20 h (GGT complexed with lipoprotein).
Concentration	In human serum from normal adult subjects, GGT activity is (at 37 °C) less than 40 U/L in males and less than 35 U/L in females (one GGT unit catalyzes the cleavage of 1 micromole of gamma-glutamyl-p-nitroanilide in one minute).
Isolation Method	Various procedures are used to extract and purify GGT from human cell membranes; the use of detergents yields a hydrophobic form containing the short hydrophobic N-terminal sequence of the heavy subunit; subsequent treatment with proteases, cleaving the hydrophobic sequence yields a hydrophilic enzyme. One procedure involves the extraction of the enzyme from cell membrane with Lubrol then with acetone purification, followed by bromelain treatment and two purification steps by DEAE-cellulose and Sephadex G-150 chromatography. This procedure yields GGT with a very high specific activity (8800 units/mg).
Amino Acid Sequence	The heavy and light subunits are composed of 351 and 189 aa respectively. The sequence of human and rat GGT is very similar except for a very dissimilar group of amino acids (67 to 135); the dissimilarity is due to a few single base deletions and insertions, thus supporting the close evolutionary relation between the two enzymes.
Disulfides/SH-Groups	None
General References	Tate S. S. et al. *Mol. Cell. Biochem.* 1981, **39**:357. Nemesanszky E. et al. *Clin. Chem.* 1985, **31**:797. Sacchetti L. et al. *Electrophoresis* 1989, **10**:619. Figlewicz D. A. et al. *Genomics* 1993, **7**:299.
Ref. for DNA/AA Sequences	Coloma J. et al. *Nucl. Acid Res.* 1986,**14**:1393. Sakamuro D. et al. *Gene* 1988,**73**:1. Rajpert-De Meyts E. et al. *Proc. Natl. Acad. Sci. U. S. A.* 1988,**85**:8840. Courtay C. et al. *Biochem. Pharmacol.* 1992, **43**: 2527. Courtay C. et al. *Biochem. J.* 1994, **297**:503.

The nucleotide sequence data reported by Courtay C. et al. (1994) will appear in the EMBL Genebank and DDBJ Nucleotide sequence databases under the following accession numbers: GGT1, L 10394; GGT2, L 10395; GGT3, L 10396; GGT11, L 10397; GGT12, L 10398; GGT13, L 10399.



Schematic representation of the topology of gamma-glutamyltransferase in microvillus membranes. H and L refers to heavy and light subunits; X refers to the active site and the shaded areas are the glycosylated regions. The arrows indicates the cleavage site by papain. (Taken from Tate, S. and Meister A., *Mol. Cell Biochem.* 1981, **39**: 357).

Glucocerebrosidase

Gregory A. Grabowski

Synonyms	Glucosylceramidase; Acid β-glucosidase; Ceredase; TM
Abbreviations	GC'ase; β-Glc
Classifications	EC 3.2.1.45; glucohydrolase
Description	A membrane associated lysosomal hydrolase, synthesized in all human tissues; The activity is very low or not present in plasma. Active as a monomer in purified preparations, but may exist as a homodimer in tissues. A glycoprotein with about 8 - 15% carbohydrate depending on the source. Multiple forms are present in different tissues due to differential post-translational glycosidic modification. No proteolytic processing occurs after entering the ER lumen. Four of five N-glycosylation sequons are occupied by triantennary complex- and hybrid-type oligosaccharides on the placental and fibroblast enzyme. Requires negatively-charged bile acids or phospholipids for maximal activity in vitro. Negatively charged sphingolipids or phospholipids and a naturally occurring acidic glycopeptide activator, termed saposin C, is required in vivo. Binding sites for saposin C, lipid activators and substrates are present. Needs glycosylation at N19 for development and maintenance of active conformation.
Structure	No physico-chemical measurements available.
Molecular Weight	55,575 (calculated from complete aa sequence, mature, glycosylated), approx. 56,000 (SDS-PAGE). Cellular glycosylated forms of 57,000 to 69,000 (SDS-PAGE, Western blots).
Sedimentation Coeff.	Unknown
Isoelectric Point	4.5 - 7.3 (granular bed gels)
Extinction Coeff.	Unknown
Enzyme Activity	N-acyl-sphingosyl-1-O-β-glucoside: glucohydrolase. Hydrolyzes the β-glucosidic linkage from sphingosyl-glucosides in the lysosomes of all tissues.
Coenzymes/Cofactors	A protein activator, saposin C, identified by in vitro assays and is necessary in vivo. Saposin C's effect requires the presence of negatively charged phospholipids, in vitro, to increase the catalytic rate constant; Km unaffected. Mechanisms of negatively charged phospholipids or saposin C effect are not clear. Negatively charged bile salts or phospholipids (phosphatidylserine) required for activity in vitro.
Substrates	Major substrates are glucosylceramides (N-acyl-sphingosyl-1-O-β-glucosides). Varying fatty acid acyl and sphingosyl chain lengths are used. Glucosylsphingosine, sphingosyl-1-O-β-glucoside, is hydrolyzed at 10-100 fold lower rates than glucosylceramides. 4-alkyl-umbelliferyl-1-O-β-glucosides are excellent fluorogenic substrates; the 4-methyl derivative is commonly used.
Inhibitors	No demonstrated in vivo inhibitors. Glucosylsphingosine likely candidate. 5-imino-5-deoxyglucosides and their N-alkyl dervitives (nojirimycins),

	competitive; N-alkyl-β-glucosylamines (chain length 14-18 carbons) slow, tight binding competitive (Ki < 1nM), probably transition state analogues; 2-N-alkyl-glucosylsphingosines, mixed type inhibitors; conduritol B epoxide derivatives covalently inhibit. 2-deoxy-2-fluoro-glucosidefluorides are mechanism-based, active-site directed covalent inhibitors.
Biological Functions	Final hydrolytic step in the lysosomal catabolism of complex glycosphingolipids. Has no known functions outside of the acidic lysosomal environment. Major natural source for the glucosylceramide substrate is the turnover of leukocyte membranes by the reticuloendothelial system.
Physiology/Pathology	Essential for complex glycosphingolipid catabolism. Deficiency states result in the lysosomal storage disease, Gaucher disease.
Degradation	Degraded by lysosomal proteases, $t_{1/2}$ approx. 60 hrs. Synthesized in the ER, transported through golgi apparatus and arrived in lysosome in about 2-3 hrs.
Genetics/Abnormalities	No known natural polymorphisms. From DNA analysis, the human, canine and murine proteins are about 85% identical. N-glycosylation sequon position preserved. More than 50 mutant human alleles have been found in the variants of Gaucher disease (see MMBID review). The most frequent alleles are N370S and L444P which when found homoallelically are associated with the absence and presence of neuronopathic Gaucher disease, respectively. In humans the gene maps to 1q21-23. In humans a highly homologous nonprocessed pseudogene is within 32 kb of the structural gene. The pseudogene may play an important role in creating mutations in Gaucher disease.
Half-life	≈ 60 hrs. in skin fibroblasts and Hep G2 cells
Concentration	≈ 10 ng/mg cellular protein in normal cultured skin fibroblasts. About 10-fold less in peripheral blood lymphocytes and monocytes.
Isolation Method	Isolated from human placenta (richest source) by aqueous extraction in cholate, ammonium sulfate fractionation, citric acid precipitation, butanol extraction, hydrophobic chromatography, and, either, dodecyl-deoxynojirimycin affinity or immunoaffinity chromatography. Secreted when overexpressed in insect, CHO and C2C12 cells.
Amino Acid Sequence	Complete sequence deduced from cDNA sequence, 497 aa. No superfamily membership. No transmembrane domains; peripheral membrane protein. Asn-443 binding site for conduritol B epoxide. Gln-340 catalytic nucleophile.
Disulfides/SH-Groups	Seven cysteines in the mature protein at aa 4, 16, 18, 23, 126, 248 and 342. Three intrachain disulfides including: C4-C16, C23-C342 and one more involving C248 and C18 or C126: Assignment not conclusive.
General References	Beutler, E. and Grabowski, G.A. Glucosylceramide Lipidoses: Gaucher Disease. In: *The Metabolic and Molecular Basis of Inherited Disease.* Scriver, C.R. Beaudet, A.L., Sly, W.S. and Valle, D. (eds.) McGraw-Hill, New York, 1995, pp. 2641-2670. Miao, S., McCarter, J., Tull, D. et al. Identification of Glu340 as the active site nucleophile in human glucocerebrosidase by use of electrospray tandem mass spectrometry. *J. Biol. Chem.* 1994, **269**:10975-10978.

Qi, X., Qin, W., Sun, Y. et al. Functional organization of saposin C: Definition of the neurotrophic and acid β-glucosidase activation regions. *J. Biol. Chem.* 1996, **271**:6874-6880.

Ref. for DNA/AA Sequences

Horowitz, M. et al. The human glucocerebrosidase gene and pseudogene: Structure and evolution. Genomics 1989, **4**:87 (genomic DNA sequence-gene and pseudogene).

Tsuji, S., Choudary, P.V., Martin, B.M. et al. Nucleotide sequence of cDNA containing the complete coding sequence for human lysosomal glucocerebrosidase. *J. Biol. Chem.* 1986, **261**:50-53.

Glucose-6-phosphatase

Ann Burchell

Synonyms	
Abbreviations	G-6-Pase
Classifications	EC 3.1.3.9
Description	The liver endoplasmic reticulum is the major site of expression. The enzyme is also present in the endoplasmic reticulum in kidney, pancreatic islets, in some (but not all) brain astrocytes and in intestinal and gall bladder mucosa. During early human development it is also found in embryonic and fetal red blood cells, the adrenal fetal zone and in fetal testis. The enzyme is an integral membrane protein and its active site is inside the lumen of the endoplasmic reticulum. For normal enzyme activity in *vivo* and *in vitro* transport proteins are needed to transport the substrate glucose-6-phosphate and the products phosphate and glucose across the endoplasmic reticulum membrane, these are termed respectively T_1, T_2 and T_3. In addition a regulatory Ca^{2+} binding protein is associated with the enzyme.
Structure	Not yet known. None of the proteins have been crystallised.
Molecular Weight	36,500 (SDS PAGE); 22,000 for the stabilising protein; T_1 unknown; T_2 37,000; T_3 52,000.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	The enzyme is a phosphohydrolase and can function as a phosphotransferase.
Coenzymes/Cofactors	None
Substrates	The most important known biological substrate is glucose-6-phosphate. Other substrates most commonly used in vitro are pyrophosphate, mannose-6-phosphate and carbamylphosphate. The enzyme can use a wide range of substrates when it functions as a phosphotransferase.
Inhibitors	The best inhibitors of the enzyme are vanadate and aluminium tetrafluoride.
Biological Functions	Glucose-6-phosphatase catalyses the terminal step of the two pathways of glucose production (gluconeogenesis and glycogenolysis). In liver the major role of the enzyme is to release glucose into the bloodstream for use by other tissues which cannot make glucose (eg. brain and muscle). The liver enzyme plays a major role in the homeostatic regulation of blood glucose levels. The enzyme's role in pancreatic islets is less clear but is has been proposed that it plays roles in both the regulation of cytoplasmic Ca^{2+} levels and in the regulation of glucose stimulated insulin release from β-cells.

1. Installment 1992 (revised 1997)

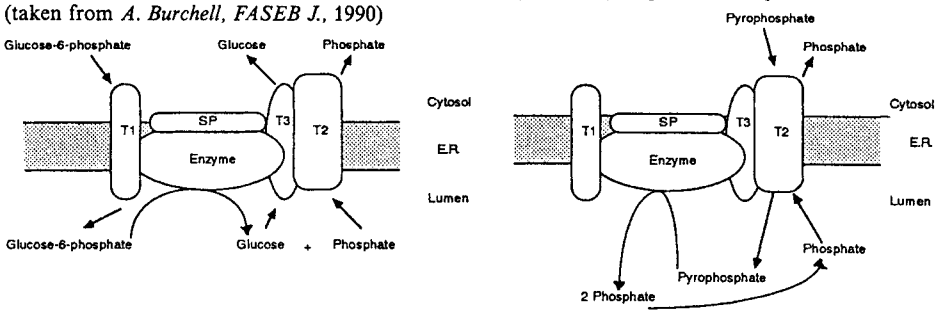
Physiology/Pathology	Essential for normal blood glucose regulation. Complete deficiencies of enzyme activity are severe metabolic disorders. Deficiencies which result in low and/or abnormal activity are usually milder. Genetic deficiencies of the enzyme, the stabilising protein, T_1, T_2 and T_3 are termed respectively type 1a, 1aSP, 1b, 1c and 1d glycogen storage disease.
Degradation	Unknown
Genetics/Abnormalities	The gene is located on chromosome 17. It has 5 exons. Genetic deficiencies of any of the glucose-6-phosphatase enzyme system proteins can cause impaired liver glucose-6-phosphatase activity. In addition to total deficiencies of individual proteins a large number of different partial deficiencies have been described, based on enzyme activity and immunoblotting data. To date 19 different mutations have been described in the glucose-6-phosphatase enzyme gene. Mutations have been found in all 5 exons and in some introns. However, no mutations have been found in the glucose-6-phosphatase enzyme gene in a significant number of patients (approx. 20%) with glycogen storage disease type 1a.
Half-life	Unknown
Concentration	Unknown
Isolation Method	Isolated from liver microsomes. Enzyme isolation very difficult as it is an integral membrane protein that is difficult to solubilise. Once solubilised, inhibited by detergents and unstable activity. Enzyme inactive in the absence of the stabilising protein. Shortest method requires five different affinity columns, eg. pentylamine sepharose, a stabilising protein antibody immunoaffinity column, a hexylamine pyrophosphate sepharose column, a hexylamine phosphate sepharose column and a histidyl diazobenzyl phosphoric acid agarose column. Stabilising protein can be isolated by Lubrol solubilisation of microsomes followed by chromatography on Fractogel TSK DEAE-650 (S). T_2 solubilisation in 8% Triton X-100 followed by a Sephadex G-25M column and an immunoaffinity column. T_3 detergent sucrose gradient centrifugation, a Sephadex G-25M column followed by immunoaffinity chromatography. T_1 not yet purified.
Amino Acid Sequence	The human liver endoplasmic reticulum glucose-6-phosphatase is a single chain protein containing 357 aa. It is a very hydrophobic glycoprotein with several potential membrane spanning regions but its topology in the endoplasmic reticulum membrane has not been unequivocally established. There are two lysines at positions 3 and 4 from the C-terminus of the protein, proteins with this motif are maintained in the endoplasmic reticulum by retrieval from both the Golgi and inermediate compartments (Jackson et al. *J. Cell Biol.* 1993, **121**:317-333). The reaction mechanism involves the formation of a phosphohistidine intermediate.
Disulfides/SH-Groups	Unknown
General References	Burchell, A. The molecular basis of the type 1 glycogen storage diseases. *Bioessays* 1992, **14**:395-400. Burchell, A., Allan, B.B. and Hume, R. Glucose-6-phosphatase proteins of the endoplasmic reticulum. *Mol. Membr. Biol.* 1994, **11**:217-227. Burchell, A. and Hume, R. The glucose-6-phosphatase system in human development. *Histol. and Histopath.* 1995, **10**:979-993. Chen, Y.T. and Burchell, A. Glycogen Storage Diseases. In: *The Metabolic and Molecular Basis of Inherited Disease.* Scriver, C.R., Beaudet,

A.L., Sly, W.S. and Valle, D. (eds.), McGraw-Hill 1995, **7th** Ed, Chapter 24, pp. 935-965.
Nordlie, R.C. and Sukalski, K.A. Multifunctional glucose-6-phosphatase a critical review. In: *The Enzymes in Biological Membranes*. Martonosi, A.N. (ed.), Plenum Press, New York, 1985, pp. 349-398.

Ref. for DNA/AA Sequences

Lei, K.J., Shelly, L.L., Pan, C.J. et al. Mutations in the glucose-6-phosphatase gene that causes glycogen storage disease type 1a. *Science* 1993, **262**:580-583.
Schmoll, D., Allan, B.B. and Burchell, A. Cloning and sequencing of the 5' region of the human glucose-6-phosphatase gene: transcriptional regulation by cAMP, insulin and glucocorticoids in H4IIE hepatoma cells. *FEBS Lett.* 1996, **383**:63-66.

A schematic working model of the hepatic microsomal glucose-6-phosphatase complex.
(taken from *A. Burchell, FASEB J.*, 1990)



Glucose-6-phosphate dehydrogenase

Gian Franco Gaetani and Anna Maria Ferraris

Synonyms	Zwischenferment
Abbreviations	G6PD; G6PDH
Classifications	D-G6P:NADP oxidoreductase, E.C. 1.1.1.49
Description	An ubiquitous cytosolic enzyme identical in all cells and expressed at different levels. The enzyme has been purified in homogenous form from different species and tissues. Biochemical data mainly refer to erythrocytic G6PD.
Structure	The active enzyme is a dimeric protein, with two molecules of tightly bound $NADP^+$ per molecule of dimer. Relationship between structural $NADP^+$ and substrate $NADP^+$ is not yet clear. Recently it has been reported a detailed description of the three dimensional structure of G6PD from *Leuconostoc mesenteroides*. The enzyme from this species has significant homology to human G6PD, and since cristallographic studies on human G6PD protein, for technical difficulties is not at hand, the structure of the bacterial enzyme has been exploited to model a human G6PD three-dimensional structure and to predict with some confidence, the position and role of both normal and mutated residues in the tertiary structures of human G6PD (Naylor, C.E. et al. *Blood* 1996:**87**, 1974-1982).
Molecular Weight	118,512 Da: dimer (active); 59,256 Da (515 aa): monomer (inactive); 237,029 Da: tetramer (partially active). At neutral pH and ionic strength 0.1 the enzyme is in rapid equilibrium between tetramer and dimer; when the enzyme is stripped of its $NADP^+$ and allowed to stand at low enzyme concentrations, or when it is in presence of urea, it dissociates in monomers. The discrete tetrameric species is observed at pH 6.0 and ionic strength 0.55.
Sedimentation Coeff.	9.0 S (pH 6.0,I = 0.55); 5.6 S (pH 7 - 8,I = 0.1)
Isoelectric Point	5.8 - 6.0 (in the presence of $NADP^+$)
Extinction Coeff.	12.2 ± 0.4 (280nm, 1%, 1cm)
Enzyme Activity	Glucose-6-phosphate + $NADP^+$ → 6-phosphogluconolactone + NADPH + H^+. Specific activity: 160 - 180 IU/mg of protein. Activity is measured in excess of G6P and $NADP^+$ by the increase in absorbance at 340 nm due to NADPH generation. Definition of Unit: a IU of enzyme activity is the amount of enzyme which reduces 1 µmole of $NADP^+$ per min at 25°C. The activity is increased by $MgCl_2$, NaCl and KCl. pH optimum is 8.0. The intracellular activity of G6PD is regulated by the availability of G6P and free $NADP^+$.
Coenzymes/Cofactors	Physiologic coenzyme: Nicotine Adenine Dinucleotide Phosphate ($NADP^+$); Km 3.5 - 5.5 µM. G6PD obeys classical hyperbolic saturation kinetics, based on Michaelis-Menten equation. Artificial coenzyme: deamino-NADP (50 - 70% of the maximal activity).

Substrates	Physiologic substrate: Glucose-6-Phosphate (G6P). Km 50 - 70 μM. Analogous substrates: 2-deoxy-glucose-6-phosphate < 5% of the maximal activity; galactose-6-phosphate < 5% of the maximal activity. Chromogenic substrate: MTT [2-(4,5-dimethylthiazolyl-2)-2,5 diphenyltetrazolium bromide]
Inhibitors	Physiologic inhibitors: NADPH, competitive with $NADP^+$ (Ki= 7-10 μM); ATP, competitive with G6P. Artificial inhibitors: p-chloromercuribenzoate.
Biological Functions	G6PD is the first and rate-limiting enzyme of the pentose-phosphate pathway, or hexosemonophosphate shunt. Its primary purpose in most cells is to generate reducing power in the extramitochondrial cytoplasm in the form of NADPH. This function is especially prominent in tissues, e.g., liver, mammary gland and adrenal cortex, that actively carry out the reductive synthesis of fatty acids and steroids from acetyl-CoA. A second special function of the pathway is to convert hexoses into pentoses, particularly D-ribose-5-phosphate, required in the synthesis of nucleic acids.
Physiology/Pathology	NADPH is required for regeneration of reduced glutathione (GSH), and for catalase stability. Since glutathione and catalase play an essential role in the detoxification of hydrogen peroxide, the defense of cells against this oxidizing agent depends on G6PD. This is particularly true for human erythrocytes, where deficiency of G6PD makes the cells highly susceptible to oxidative hemolysis (hemoglobin denaturation, Heinz bodies).
Degradation	No evidence of degradation product of G6PD in normal cells. However, it is evident that G6PD deficiency often results from enhanced rate of destruction in vivo, presumably as a consequence of a mutation affecting the stability of the enzyme.
Genetics/Abnormalities	The gene encoding G6PD maps to the long arm of chromosome X (band Xq 28). One of the two G6PD alleles undergoes inactivation in somatic cells of females. G6PD deficiency is the most prevalent enzyme defect (it affects about 400 million people worldwide). Over 300 different variants have been identified through biochemical methods, but at molecular level only 100 mutations have been reported sofar. G6PD-deficient subjects are particularly prone to develop acute hemolytic anemia after exposure to certain oxidizing agents.
Half-life	58 - 62 days in human erythrocytes (in vivo).
Concentration	Variable concentration and enzyme activity in different tissues: Erythrocytes 8.5 IU x 10^{-3}/mg protein Granulocytes 851 Fibroblasts 174 Muscle 3.3 Liver 7.2 Brain (fetal) 85
Isolation Method	Isolated from different tissues by chromatography and ammonium sulphate precipitation or by affinity chromatography on 2',5'-ADP-Sepharose.
Amino Acid Sequence	The complete primary sequence of 515 aa has been determined from the cDNA sequence and matches the previously known sequence of several peptides. The deduced aa composition is similar to that from Saccharomyces cerevisiae, suggesting considerable sequence conservation throughout eukaryotic evolution. This is made evident from inspection of a yeast

peptide which differs in only two residues from the homologous human peptide; it contains a lysine residue (essential for catalytic activity in the yeast enzyme), which may be near the G6P binding site and has been found to be reactive with pyridoxal phosphate in the human enzyme. Homologies have been detected with several other dehydrogenases.

Disulfides/SH-Groups

The total sulfydryl content of G6PD is 17 x 10^{-5} mole per gram. The fully active enzyme has only cysteine residues and no inter or intra -S-S- bridges in the molecule. In the absence of urea, the -SH content is estimated to be $7x10^{-5}$ per gram, indicating that approximately 60% of the sulfydryl of the enzyme is not exposed to react with -SH reagent in the absence of urea.

General References

Luzzatto, L. and Mehta, A. Glucose-6-phosphate dehydrogenase deficiency. In: *The Metabolic and Molecular Basis of Inherited Disease.* Scriver, C.R. et al. (eds.), Mc Graw-Hill, 1995, pp. 3367-3398.
Gaetani, G.F. et al. *Blood* 1989, **73**:334-338.
Cohen, P. and Rosemeyer, M.A. Glucose-6-phosphate dehydrogenase from human erythrocytes. *Meth. Enzymol.* 1975, **XLI**:208-214.
Luzzatto, L. and Battistuzzi, G. Glucose-6-phosphate dehydrogenase. In: *Adv. Hum. Genet.* Harris, H. and Hirschorn, K. (eds.), Plenum Press, 1985, Vol. **14**, pp. 217-329.
Kirkman, H.N. and Gaetani, G.F. *J. Biol. Chem.* 1986, **261**:4033-4038.
Rowland, P. et al. *Structure* 1994, **2**:1073-1087.

Ref. for DNA/AA Sequences

Persico, M.G. et al. *Nucl. Acids Res.* 1986, **14**:2511-2522.
Takizawa, T. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:4157-4161.

Glucose Transport Protein 1

Ann Burchell

Synonyms	Plasma membrane facilitative glucose transport protein 1, erythrocyte-type facilitative glucose transport protein , GLUT 1.
Abbreviations	GLUT 1
Classifications	Member of the facilitative glucose transport protein family.
Description	A brain, placenta and red blood cell plasma membrane facilitative glucose transport protein; lower levels in many tissues including liver and also expressed in transformed cells and tissue culture lines.
Structure	The family of mammalian plasma membrane facilitative glucose transport proteins are predicted to contain 12 membrane spanning helices, with intracellular N- and C-termini. There are relatively large loops between helices 1 and 2 and between 6 and 7, the latter divides the structure into two halves. The length of the remaining loops at the cytoplasmic surface are very short :- approx. eight residues.
Molecular Weight	53,000 Da (7.5% SDS-PAGE apparent molecular weight), 56,500 Da (10% SDS-PAGE apparent molecular weight), 60,000 Da (12.5% SDS-PAGE apparent molecular weight), the transporter is a glycoprotein containing about 15% carbohydrate by weight.
Sedimentation Coeff.	Unknown
Isolectric point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	Unknown
Substrates	Glucose and galactose (K_m for galactose 17 mM); kinetic asymmmetric transporter, the K_m for net influx into the cell is lower than the K_m for net efflux.
Inhibitors	Cytochalasin B, forskolin, 2-N-4-(1-azi-2,2,2-trifluoroethyl)-benzoyl-1,3-bis-(D-mannos-4-yloxy)-2-propylamine (ATB-BMPA). Inhibited by intracellular ATP and allosterically regulated by binding of intracellular metabolites.
Biological Functions	Transports glucose from the blood into many cell types.
Physiology/Pathology	Functions effectively as a unidirectional transporter (into the cell) under conditions where extracellular glucose is low and intracellular demand for glucose is high. Mitogens stimulate transcription and starvation stimulates expression.
Degradation	Unknown
Genetics / Abnormalities	Chromosomal location of gene 1p35 → 31.3
Half-life	Unknown

Concentration	Unknown
Isolation Method	Purification from human erythrocytes by pH 12 treatment detergent, solubilisation with octyl glucoside followed by ultracentrifugation and DEAE-cellulose chromatography (Baldwin, S. A., Baldwin, J. A. and Lienhard, G. E. *Biochemistry* 1982, **21**: 3836–3842).
Amino Acid Sequence	A single chain protein containing 492 amino acids. There is considerable homology among all the facilitative glucose transport protein family of proteins.
Disulfides/SH-Groups	Disulfide bridges have not been described.
General References	Bell, G. I. et al. *Diabetes Care* 1990, **13**: 198–208. Devaskar, S. U. and Mueckler, M. *Pediatric Res.* 1992, **31**: 1–13. Mueckler, M. *Eur. J. Biochem.* 1994, **219**: 713–725.
Ref. for DNA/AA Sequences	Mueckler, M. et al. *Science* 1985, **229**: 941–945. Birnbaum, M. J. et al. *Proc. Natl. Acad. Sci. U. S. A.* 1985, **83**: 5784–5788.

Glucose Transport Protein 2

Ann Burchell

Synonyms	Liver plasma membrane glucose transport protein, the liver-type glucose transport protein , GLUT 2.
Abbreviations	GLUT 2
Classifications	Member of the facilitative glucose transport protein family.
Description	The major liver plasma membrane facilitative glucose transport protein also present in pancreatic islet b cells, kidney proximal tubules and in the basolateral membrane of the small intestine.
Structure	The family of mammalian plasma membrane facilitative glucose transport proteins are predicted to contain 12 membrane spanning helices, with intracellular N-and C-termini. There are relatively large loops between helices 1 and 2 and between 6 and 7, the latter divides the structure into two halves. The length of the remaining loops at the cytoplasmic surface are very short :- approx. eight residues.
Molecular Weight	55,000 Da, SDS-PAGE apparent molecular weight (Ferrer, G. et al. *Diabetes* 1993, **42**: 1273–1280).
Sedimentation Coeff.	Unknown
Isolectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	Unknown
Substrates	Glucose and fructose (the K_m for fructose is very high approximately 66 mM)
Inhibitors	Cytochalasin B, forskolin, 2-N-4-(1-azi-2,2,2-trifluoroethyl)-benzoyl-1,3-bis-(D-mannos-4-yloxy)-2-propylamine (ATB-BMPA).
Biological Functions	Transports glucose from the blood into and out of cells.
Physiology/Pathology	Highest levels in cells which are capable of producing glucose. A high K_m for glucose which is advantageous for rapid efflux of glucose into the blood following gluconeogenesis. Glucose flux through GLUT 2 at physiological glucose concentrations should change in an almost linear way depending on the intracellular/extracellular glucose concentrations.
Degradation	Unknown
Genetics/Abnormalities	Chromosomal location of gene 3q26
Half-life	Unknown
Concentration	Unknown
Isolation Method	GLUT 1 cDNA was used as a probe to isolate GLUT 2 cDNA.

Amino Acid Sequence	A single chain protein containing 524 aa. There is considerable homology among all the facilitative glucose transport protein family of proteins.
Disulfides/SH-Groups	Disulfide bridges have not been described.
General References	Bell, G. I. et al. *Diabetes Care* 1990, **13**: 198–208. Devaskar, S. U. and Mueckler, M. *Pediatric Res.* 1992, **31**: 1–13.
Ref. for DNA/AA Sequences	Fukumoto, H. et al. *Proc. Natl. Acad. Sci. U. S. A.* 1988, **85**:5434–5438. Thorens, B. et al. *Cell* 1988, **55**: 281–290.

Glucose Transport Protein 3

Ann Burchell

Synonyms	Plasma membrane facilitative glucose transport protein 3, the brain-type facilitative glucose transport protein, GLUT 3.
Abbreviations	GLUT 3
Classifications	Member of the facilitative glucose transport protein family.
Description	A brain and nerve facilitative glucose transport protein, lower levels have been reported in placenta, kidney, liver and heart.
Structure	The family of mammalian plasma membrane facilitative glucose transport proteins are predicted to contain 12 membrane spanning helices, with intracellular N-and C-termini. There are relatively large loops between helices 1 and 2 and between 6 and 7, the latter divides the structure into two halves. The length of the remaining loops at the cytoplasmic surface are very short:- approx. eight residues.
Molecular Weight	47,000–50,000 daltons; SDS PAGE apparent molecular weight (Mantych, G. J. et al. *Endocrinology* 1992, **131**:1270–1278).
Sedimentation Coeff.	Unknown
Isolectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	Unknown
Substrates	Glucose and galactose (K_m for galactose 8.5 mM).
Inhibitors	Cytochalasin B, forskolin, 2-N-4-(1-azi-2,2,2-trifluoroethyl)-benzoyl-1,3-bis-(D-mannos-4-yloxy)-2-propylamine (ATB-BMPA).
Biological Functions	Transports glucose from the blood into brain and nerves.
Physiology/Pathology	Transports glucose in tissues with high energy demands, may work in tandem with GLUT 1 which is also present in brain. Low K_m for hexoses which may help brain transport of glucose at times of hypoglycaemia or high glucose demand.
Degradation	Unknown
Genetics/Abnormalities	Chromosomal location of gene 12p3
Half-life	Unknown
Concentration	Unknown
Isolation Method	GLUT 1 cDNA was used as a probe to isolate GLUT 3 cDNA.
Amino Acid Sequence	A single chain protein containing 496 aa. There is considerable homology among all the facilitative glucose transport protein family of proteins.

Disulfides/SH-Groups	Disulfide bridges have not been described.
General References	Devaskar, S. U. and Mueckler, M. *Pediatric Res.* 1992, **31**: 1–13. Gould, G. W. and Holman, G. D. *Biochem. J.* 1993, **295**: 329–341. Kayano, T. et al. *J. Biol. Chem.* 1988, **263**: 15245–15248
Ref. for DNA/AA Sequences	Kayano, T. et al. *J. Biol. Chem.* 1988, **263**: 15245–15248.

Glucose Transport Protein 4

Ann Burchell

Synonyms	Fat/muscle plasma membrane glucose transport protein, insulin responsive glucose transport protein, GLUT 4.
Abbreviations	GLUT 4
Classifications	Member of the facilitative glucose transport protein family.
Description	A major plasma membrane facilitative glucose transport protein present in adipose tissue, skeletal muscle and heart.
Structure	The family of mammalian plasma membrane facilitative glucose transport proteins are predicted to contain 12 membrane spanning helices, with intracellular N-and C-termini. There are relatively large loops between helices 1 and 2 and between 6 and 7, the latter divides the structure into two halves. The length of the remaining loops at the cytoplasmic surface are very short :-approx. eight residues.
Molecular Weight	46,000 Da (SDS-PAGE apparent molecular weight).
Sedimentation Coeff.	Unknown
Isolectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	Unknown
Substrates	Glucose
Inhibitors	Cytochalasin B, forskolin, 2-N-4-(1-azi-2,2,2-trifluoroethyl)-benzoyl-1,3-bis-(D-mannos-4-yloxy)-2-propylamine (ATB-BMPA).
Biological Functions	Transports glucose from the blood into adipose tissue and muscle.
Physiology/Pathology	An insulin regulated glucose transport protein which translocates from intracellular stores to the plasma membrane in response to insulin. Relatively low K_m transporter which should work close to its V_{max} at physiological blood glucose concentrations.
Degradation	Unknown
Genetics/Abnormalities	Chromosomal location of gene 17p13
Half-life	Unknown
Concentration	Unknown
Isolation Method	GLUT 1 cDNA was used as a probe to isolate GLUT 4 cDNA.
Amino Acid Sequence	A single chain protein containing 509 aa. There is considerable homology among all the facilitative glucose transport protein family of proteins.

Disulfides/SH-Groups	Disulfide bridges have not been described.
General References	Devaskar, S. U. and Mueckler, M. *Pediatric Res.* 1992, **31**: 1–13. Kayano, T. et al. *J. Biol. Chem.* 1990, **265**: 13276–13282. Bell, G. I. et al. *Diabetes Care* 1990, **13**: 198–208.
Ref. for DNA/AA Sequences	James, D. E. et al. *Nature* (London) 1989, **338**: 83–87. Birmbaum, M. J. *Cell* 1989, **57**: 305–315. Charron, M. J. et al. *Proc. Natl. Acad. Sci. U. S. A.* 1989, **86**: 2535 2539. Kaestner, K. H. et al. *Proc. Natl. Acad. Sci. U. S. A.* 1989, **86**: 3150–3154. Fukumoto, H. et al. *J. Biol. Chem.* 1989, **264**: 7776–7779.

Glucose Transport Protein 5

Ann Burchell

Synonyms	Plasma membrane facilitative glucose transport protein 5, small intestine facilitative glucose transport protein, the small intestine sugar transport protein, GLUT 5.
Abbreviations	GLUT 5
Classifications	Member of the facilitative glucose transport protein family.
Description	A small intestine (apical membrane) fructose transport protein, lower levels in muscle and brain, also present in adipose tissue.
Structure	The family of mammalian plasma membrane facilitative glucose transport proteins are predicted to contain 12 membrane spanning helices, with intracellular N-and C-termini. There are relatively large loops between helices 1 and 2 and between 6 and 7, the latter divides the structure into two halves. The length of the remaining loops at the cytoplasmic surface are very short :- approx. eight residues.
Molecular Weight	49,000 Da (SDS-PAGE apparent molecular weight).
Sedimentation Coeff.	Unknown
Isolectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	Unknown
Substrates	Fructose
Inhibitors	Cytochalasin B, forskolin, 2-N-4-(1-azi-2,2,2-trifluoroethyl)-benzoyl-1,3-bis-(D-mannos-4-yloxy)-2-propylamine (ATB-BMPA).
Biological Functions	Fructose transport across plasma membranes. It is the major small intestine fructose transport protein. It was originally mistakenly reported to be a glucose transport protein but is not now believed to function as a glucose transport protein *in vivo*.
Physiology/Pathology	Major role is fructose transport/absorption from the small intestine.
Degradation	Unknown
Genetics/Abnormalities	Chromosomal location of gene 1p31
Half-life	Unknown
Concentration	Unknown
Isolation Method	GLUT 1 cDNA was used as a probe to isolate GLUT 5 cDNA.
Amino Acid Sequence	A single chain protein containing 501 aa. There is considerable homology among all the facilitative glucose transport protein family of proteins.

Disulfides/SH-Groups	Disulfide bridges have not been described.
General References	Devaskar, S. U. and Mueckler, M. *Pediatric Res.* 1992, **31**: 1–13. Gould, G. W. and Holman, G. D. *Biochem. J.* 1993, **295**: 329–341. Kayano, T. et al. *J. Biol. Chem.* 1990, **265**: 13276–13282.
Ref. for DNA/AA Sequences	Kayano, T. et al. *J. Biol. Chem.* 1988, **263**: 15245–15248.

Glucose Transport Protein 7

Ann Burchell

Synonyms	Endoplasmic reticulum glucose transport protein, the glucose-6-phosphatase system transport protein, GLUT 7.
Abbreviations	GLUT 7
Classifications	Member of the facilitative glucose transport protein family.
Description	A liver endoplasmic reticulum facilitative glucose transport protein. A probably identical endoplasmic reticulum transport protein is also present in all tissues which contain glucose-6-phosphatase i.e. kidney, pancreatic islets, intestine, gallbladder, human fetal adrenal, some (but not all) brain astrocytes and testis.
Structure	Not yet known
Molecular Weight	52,000 Da (SDS-PAGE apparent molecular weight).
Sedimentation Coeff.	Unknown
Isolectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	Glucose and other hexoses
Inhibitors	None known
Biological Functions	The glucose-6-phosphatase enzyme catalyzes the terminal step of liver glucose production. The enzyme is situated with its active site inside the lumen of the endoplasmic reticulum where it produces glucose by the hydrolysis of glucose-6-phosphate. Endoplasmic reticulum glucose transport protein(s) are needed to allow the glucose produced in the lumen to reach the cytoplasm of the cell. It is not yet known whether there are several different endoplasmic reticulum glucose transport proteins or if GLUT 7 is the only one.
Physiology/Pathology	The liver protein is essential for normal hepatic glucose production and normal blood glucose homeostasis.
Degradation	Unknown
Genetics/Abnormalities	The genetic deficiency of liver endoplasmic reticulum glucose transport (glycogen storage disease type 1 d) is a severe metabolic disorder which is clinically similar to a deficiency of the glucose-6-phosphatase enzyme, as both result in impaired hepatic glucose production.
Half-life	Unknown
Concentration	Unknown

Isolation Method	Detergent sucrose gradient centrifugation, Sephadex G25M chromatography, affinity chromatography on anti-(human) erythrocyte glucose transport protein) antibody-Sepharose.
Amino Acid Sequence	The sequence shows homology with the family of plasma membrane glucose transport proteins.
Disulfides/SH-Groups	Disulfide bridges have not been described.
General References	Burchell, A. Glucose transport across hepatic membranes. **In** *Transport in the liver* (Keppler, D. and Jungermann, K., eds.) Kluwer Academic Press, London 1994, pp. 59–72. Burchell, A. Hepatic microsomal glucose transport. *Biochem. Soc. Trans.* 1994, **22**: 658–663. Chen, Y-T. and Burchell, A. Glycogen storage diseases. **In** *The metabolic basis of inherited disease.* (Scriver, C. R., Beaudet, A. L., Sly, W. S. and Valle, D., eds.) 7th edn 1995 chapter 24.
Ref. for DNA/AA Sequences	Waddell, I. D. et al. Cloning and expression of a hepatic microsomal glucose transport protein. *Biochem. J.* 1992, **286**:173–177.

Glucuronate-2-Sulphatase

Julie Bielicki, Craig Freeman and John J. Hopwood

Synonyms	Glucuronate-2-sulphate sulphatase
Abbreviations	GAS
Classifications	EC 3.1.6. X
Description	GAS is a lysosomal enzyme synthesized in the endoplasmic reticulum and is present in all cells. The mature form, isolated from liver, comprises two polypeptide chains which are not S-S linked and have a molecular size of 47 kDa and 19.5 kDa.
Structure	Unknown
Molecular Weight	Native M_r of human liver mature GAS is 63 kDa (gel permeation in 50 mM Na acetate, 0.5 M NaCl, 10% (v/v) glycerol, 0.1 mM-dithioerythritol, pH 5.6).
Sedimentation Coeff.	Unknown
Isoelectric Point	>7.5 (chromatofocussing)
Extinction Coeff.	Unknown
Enzyme Activity	GAS hydrolyzes the C2-sulphate ester bond from the non-reducing terminal glucuronic acid residues of the glycosaminoglycans heparan sulphate (HS) and chondroitin 6-sulphate (C6S).
Coenzymes/Cofactors	Unknown
Substrates	Biological substrates are heparan sulphate and chondroitin 6-sulphate. Diagnostic substrates are derived from these biological substrates and are 0-(β-glucuronic acid 2-sulphate)-(1 → 4)-0-(2,5)-anhydro [1-^{3}H] mannitol 6-sulphate (GSMS) from HS and 0-(β-glucuronic acid 2-sulphate)-(1 → 3)-0-(2,5)-anhydro [1-^{3}H] talitol 6-sulphate (GSTS) from C6S.

	GSMS	GSTS
K_m	0.3 μM	0.6 μM
pH opt	3.2	3.0

Inhibitors

Concentration of substances which give 50% inhibition of GAS activity towards GSMS are as follows:

Inhibitor:	Concentration:
Na_2SO_4	80 μM
Na_2HPO_4	250 μM
EDTA	2 mM
NaCl	75 mM
GMS	1 μM
ISMS	6 μM

Biological Functions	Human liver GAS is one of a number of enzymes involved in the degradation of glycosaminoglycans.
Physiology/Pathology	There is no known lysosomal storage disease resulting from a deficiency of GAS activity.

Degradation	Unknown
Genetics/Abnormalities	Chromosome location unknown.
Half-life	Unknown
Concentration	Tissue Source — pmol/min/per mg of protein liver — 3.3 lung — 2.2 kindey — 7.1 skin fibroblasts (n=2) — 3.1, 7.3
Isolation Method	4 step, 6 column procedure Step 1: Concanavalin A Sepharose-Blue A-Agarose Step 2: DEAE-Sephacel-Octyl Sepharose Step 3: CM-Sepharose Step 4: TSK-G3000SW chromatography
Amino Acid Sequence	Unknown
Disulfides/SH-Groups	Unknown
General References	Freeman, C. and Hopwood, J. J. *Biochem. J.* 1989, **259**: 209–216.
Ref. for DNA/AA Sequences	

Glutamic-γ-semialdehyde dehydrogenase, liver mitochondria

Regina Pietruszko

Synonyms	Human "high Km" aldehyde dehydrogenase; 1-pyrroline-5-carboxulate dehydrogenase.
Abbreviations	ALDH IV
Classifications	EC 1.5.1.12
Description	A mitochondrial protein, a dimer of polypeptide chains of approximately 70,000 Da each.
Structure	The cDNA has been cloned and aa sequence determined by Hu, C.A. et al. *J. Biol. Chem.* 1996, **271**:9795-9800.
Molecular Weight	118,000 Da (from aa sequence) consisting of two subunits of 59,000 Da.
Sedimentation Coeff.	Unknown
Isoelectric Point	6.8 and 6.9 (2 components)
Extinction Coeff.	11.0 (280nm, 1%, 1cm).
Enzyme Activity	Glutamic-γ-semialdehyde:NAD Oxidoreductase. Catalyses irreversible dehydrogenation of glutamic semialdehyde and some other aldehydes; also catalyses the hydrolysis of esters.
Coenzymes/Cofactors	NAD is the coenzyme, it functions as hydride acceptor during aldehyde dehydrogenation. NAD is not essential for the esterase reaction but it affects its velocity.
Substrates	Glutamic-γ-semialdehyde is the most important substrate. Other substrates are: acetaldehyde, propionaldehyde, succinic semialdehyde, glutaric semialdehyde, adipic semialdehyde. p-Nitrophenyl acetate is used to assay the esterase.
Inhibitors	3-Hydroxybenzaldehyde, 4-hydroxybenzaldehyde, proline, γ-aminobutyric acid, Tris. The enzyme is irreversibly inhibited by iodoacetamide.
Biological Functions	The enzyme is metabolically important, since its substrate, glutamic-γ-semialdehyde (1-pyrroline-5-carboxylate) is a common intermediate in the degradative and biosynthetic pathways of amino acids: arginine, citrulline, ornithine and proline.
Physiology/Pathology	Type I hyperprolinaemia, which results in elevated levels of proline in blood and in high urinary excretion of proline, hydroxyproline, glycine, 1-pyrroline-5-carboxylate and 3-hydroxy-1-pyrroline-5-carboxylate is associated with the complete absence of glutamic-γ-semialdehyde dehydrogenase. Deafness and slight mental retardation are among symptoms of the disease.
Degradation	Unknown

Genetics/Abnormalities	The gene has been localized on chromosome 1 (Hu, C.A. et al. *J. Biol. Chem.* 1996, **271**:9795-9800).
Half-life	Unknown
Concentration	approx. 500 mg L^{-1} in liver.
Isolation Method	The enzyme was purified from human liver by chromatography of dialyzed, centrifuged homogenate on CM-Sephadex, followed by ammonium sulphate fractionation, chromatography on Blue Sepharose CL 6B (Pharmacia) and 5'AMP-Separose 4B.
Amino Acid Sequence	MLLPAPALRR ALLSRPWTGA GLRWKHTSSL KVANEPVLAF TQGSPERDAL QKALKDLKGR MEAIPCVMGD EEVWTSDVQY QVSPFNHGHK VAKFCYADKS LLNKAIEAAL AARKEWDLKP IADRAQIFLK AADMLSGPRR AEILAKTMVG QGKTVIQAEI DAAAELIDFF RFNAKYAVEL EGQQPISVPP STNSTVYRGL EGFVAAISPF NFTAIGGNLA GAPALMGNVV LWKPSDTAML ASYAVYRILR EAGLPPNIIQ FVPADGPLFG DTVTSSEHLC GINFTGSVPT FKHLWKQVAQ NLDRFHTFPR LAGECGGKNF HFVHRSADVE SVVSGTLRSA FEYGGQKCSA CSRLYVPHSL WPQIKGRLLE EHSRIKVGDP AEDFGTFFSA VIDAKSFARI KKWLEHARSS PSLTILAGGK CDDSVGYFVE PCIVESKDPQ EPIMKEEIFG PVLSVYVYPD DKYKETLQLV DSTTSYGLTG AVFSQDKDVV QEATKVLRNA AGNFYINDKS TGSIVGQQPF GGARASGTND KPGGPHYILR WTSPQVIKET HKPLGDWSYA YMQ
Disulfides/SH-Groups	8 Cysteine residues per subunit
General References	Forte-McRobbie, C.M., and Pietruszko, R. *J. Biol. Chem.* 1986, **261**:2154-2163. Forte-McRobbie, C.M. and Pietruszko, R. *Biochem. J.* 1989, **261**:935-943. Hu, C.A. et al. *J. Biol. Chem.* 1996, **271**: 9795-9800.
Ref. for DNA/AA Sequences	EMBL Data Bank with accession numbers: U24266 and U24267.

Glutathione Peroxidase, Cellular

Kazuhiko Takahashi and Ikuko Saito

Synonyms	Glutathione peroxidase, cytosolic; Glutathione peroxidase, classical; Glutathione peroxidase-1
Abbreviations	cGPx, cGSHPx, GSHPx-1
Classifications	EC 1.11.1.9
Description	A cytosolic protein, found in virtually all cells, comprising four identical 23,000 subunits. Each subunit contains one atom of selenium in the active site in the form of selenocysteine residue.
Structure	The tertiary structure is not yet determined, however the structure of bovine cGPx is known. The rigid core of the subunit is composed primarily of the four β-strands and the first and fourth α-helix, whereas the more remote α_2- and α_3-helices are involved in subunit interactions.
Molecular Weight	90,000 (gel filtration); 23,000 (SDS-PAGE)
Sedimentation	Unknown
Isoelectric Point	pI 5.8
Extinction Coeff.	Unknown
Enzyme Activity	Glutathione:hydrogen-peroxide oxidoreductase $ROOH + 2\,GSH \rightleftarrows GSSG + ROH + H_2O$
Coenzymes/Cofactors	None
Substrates	Electron donor: glutathione. Electron acceptor: hydrogen peroxide, t-butyl hydroperoxide, cumene hydroperoxide.
Inhibitors	Mercaptosuccinate, the active site selenocysteine is trapped by the rapid binding of the inhibitor in competition with glutathione.
Biological Function	Using glutathione as the reducing equivalent, cGPx reduces hydrogen peroxide to water and organic hydroperoxides to the corresponding alcohols. cGPx functions in the protection of cells against oxidative damage.
Physiology/Pathology	Essential to regulate intracellular peroxide concentrations. The expression of cGPx depends on the availability of selenium. Selenium deficiency results in a decrease in cGPx activity in tissues and blood cells.
Degradation	Unknown
Genetics/Abnormalities	Selenocysteine at its active site is coded by a UGA stop codon in mRNA. The gene is on chromosome 3 at q11–13.1, and has 2 exons and 1 intron.
Half-life	Unknown
Concentration	Varies greatly between tissues.

Isolation Methods	25–50% saturated ammonium sulphate fraction of 15% hemolysate of erythrocytes, followed by CM cellulose CM-52, DEAE cellulose DE-52, Sephadex G-200, and DEAE-Sephadex A-25.
Amino Acid Sequence	Selenocysteine at its active site. The amino acid sequence exhibits 44% homology with that of extracellular glutathione peroxidase.
Disufides/SH-Groups	Unknown
General References	Awasthi, Y. C. et al. *J. Biol. Chem.* 1975, **250:** 5114–5149. Flohe, L. The selenoprotein glutathione peroxidase. **In** *Glutathione: chemical, biochemical and medical aspects.* Dolphin, D., Poulson, R., Avamovic, O. (eds.), Wiley, New York 1989, pp. 644–731.
Ref. for DNA/AA Sequences	Sukenaga, Y. et al. *Nucl. Acids Res.* 1987, **15:** 7178. Y00433.

Glutathione Peroxidase, Extracellular

Kazuhiko Takahashi and Ikuko Saito

Synonyms	Glutathione peroxidase, plasma
Abbreviations	eGPx, eGSHPx, pGPx, pGSHPx
Classifications	EC 1.11.1.9
Description	A plasma protein, synthesized mainly in the kidney proximal tubules, comprising four identical 23,000 subunits. eGPx is also detectable in milk, urine, and sperm plasma. Each subunit contains one atom of selenium in the active site in the form of selenocysteine residue.
Structure	The tertiary structure is not yet determined, however the structure of the homologous protein (bovine cellular GPx) is known. The rigid core of the subunit is composed primarily of the four β-strands and the first and fourth α-helix, whereas the more remote α_2- and α_3-helices are involved in subunit interactions.
Molecular Weight	90,000 (gel filtration): 23,000 (SDS-PAGE)
Sedimentation	Not determined
Isoelectric Point	pI 5.2
Extinction Coeff.	Not determined
Enzyme Activity	Glutathione:hydrogen-peroxide oxidoreductase $ROOH + 2\ GSH \leftrightarrows GSSG + ROH + H_2O$
Coenzymes/Cofactors	None
Substrates	Electron donor: glutathione. Electron acceptor: hydrogen peroxide, phosphatidylcholine hydroperoxide, t-butyl hydroperoxide, cumene hydroperoxide.
Inhibitors	Unknown
Biological Function	Using glutathione as the reducing equivalent, eGPx reduces hydrogen peroxide to water and organic hydroperoxides to the corresponding alcohols. eGPx protects extracellular fluid components and cell surfaces against peroxide-mediated damage.
Physiology/Pathology	Essential to protect extracellular fluid components and cell surfaces against peroxide-mediated damage. The expression of eGPx depends on the availability of selenium. Selenium deficiency results in a decrease in eGPx activity in plasma/serum. The plasma levels of eGPx have been used as a clinical marker of selenium deficiency. eGPx in the plasma of patients with renal failure on dialysis was 32%.
Degradation	Unknown
Genetics/Abnormalities	Selenocysteine at its active site is coded by a UGA stop codon in mRNA. The gene is on chromosome 5 at q32, and is approximately 10 kb long with 5 exons.

Half-life	Unkown
Concentration	Plasma: 17.8 ± 2.2 mg/L
Isolation Methods	(a) 0–40% saturated ammonium sulphate fraction of serum, followed by DEAE cellulose DE-52 (batchwise and column chromatography), hydroxyapatite, Sephadex G-200 (1st and 2nd), and DEAE-Sephadex A-25. (b) Chromatography of plasma on DEAE-cellulose 650M, followed by Phenyl-Sepharose, Q-Sepharose, Mono-Q, and Phenyl-Superose.
Amino Acid Sequence	Selenocysteine at its active site. The amino acid sequence exhibits 44% homology with that of cellular glutathione peroxidase.
Disufides/SH-Groups	Not determined
General References	Takahashi, K., Cohen, H. J. *Blood* 1986, **68**:640–645. Takahashi, K. et al. *Arch. Biochem. Biophys.* 1987, **256**:677–686. Yamamoto, Y., Takahashi, K. *Arch. Biochem. Biophys.* 1993, **305**:541–545. Avissar, N. et al. *Am. J. Physiol.* 1994, **266**: C378-C375.
Ref. for DNA/AA Sequences	Takahashi, K. et al. *J. Biochem.* 1990, **108**:145–148. D00632.

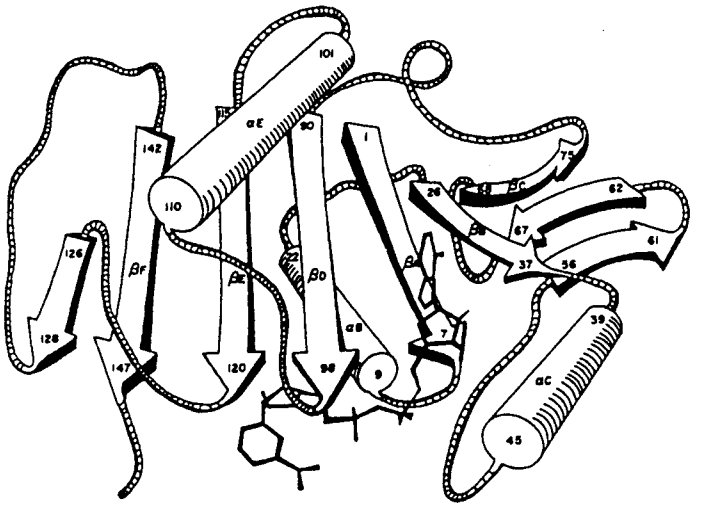
Glyceraldehyde-3-phosphate dehydrogenase

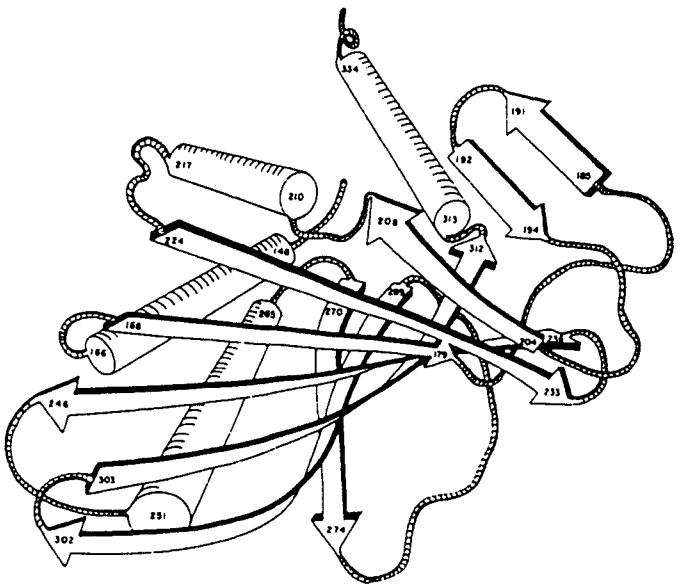
Francesco Salvatore and Lisa de Conciliis

Synonyms	Triose phosphate-dehydrogenase; D-glyceraldehyde-3-phosphate dehydrogenase.
Abbreviations	GAPDH; G3PD; GPDH; G3PDH; GAPD
Classifications	EC 1.2.1.12. phosphorylating oxidoreductase
Description	D-glyceraldehyde-3-phosphate dehydrogenase: NAD^+ oxidoreductase is a crucial enzyme of carbohydrate metabolism and it occurs widely and abundantly throughout the whole evolutionary scale. It has been found in such diverse organisms as man, pig, lobster, *Drosophila melanogaster*, *Escherichia coli*, etc. where it is strongly conserved, at the level of both primary and tertiary structure. It occurs in all mammalian tissues including red cells. Multiple secondary isozymes have been claimed to occur in human tissues, but this has yet to be substantiated.
Structure	The active enzyme is a tetramer of chemically identical but crystallographically asymmetric chains tightly associated with one NAD^+ per chain. Each subunit of the enzyme consists of two domains: the coenzyme binding domain (1 - 148) and the catalytic domain (149 - 334). The architecture of the NAD^+ binding domain is based on a six-stranded parallel β-plated sheet, with α-helices providing the return. The catalytic domain consists of an antiparallel β-plated sheet, that forms a subunit-subunit contact and three α-helices that connect different strands of this sheet. The main residues involved in catalysis are Cys-149 and His-176. The tridimensional structure of GAPD has been determined by X-ray crystallography.
Molecular Weight	37,000 monomer (cDNA sequence analysis). 142,000 tetramer (sedimentation).
Sedimentation Coeff.	7.5 S
Isoelectric Point	7.8-8.4
Extinction Coeff.	9.30; in a solution of 0.05M Tris-HCl, pH 7.4 (280nm, 1%, 1cm)
Enzyme Activity	GAPD is involved in glycolysis as well as in gluconeogenesis. The enzyme reversibly catalyses oxidation and phosphorylation of D-glyceraldehyde-3-phosphate to 1,3 diphosphoglycerate in the presence of inorganic phosphate and NAD^+, forming a stable acyl intermediate as a thioester with Cys-149. His-176 is the proton donor and acceptor during the redox step. GAPD is a B class dehydrogenase.
Coenzymes/Cofactors	The enzymatic activity requires the presence of NAD^+ or NADH in the glycolytic and gluconeogenic reaction, respectively. NAD^+ binding is associated with changes in the configurational state of the protein that influence the reactivity of its sites. A complex sequential binding mechanism, characterized by negative cooperativity has been identified in the muscle enzyme. In the glycolytic pathway inorganic phosphate is required for phosphorolysis of linkage between the acyl group and the enzyme.
Substrates	D-glyceraldehyde-3-phosphate, 1,3-diphosphoglycerate.

Inhibitors	NAD^+ is a competitive inhibitor of NADH, glyceraldehyde-3-phosphate is a competitive inhibitor of 1,3 diphosphoglycerate. Inhibitory effects are exerted by such alkylating agents as iodoacetate. GAPD is also inhibited by high concentrations of ATP at low temperature.
Biological Functions	GAPD is common to both the glycolytic and gluconeogenic pathways. Even though the enzymes isolated from various human tissues are similar in structure and properties, they show different behaviour depending on the cellular environments and requirements. GAPD is primed for glycolysis or gluconeogenesis depending on the NAD^+/NADH and ATP/ADP+Pi ratios in the cell in conjunction with the appropriate substrate concentration. In addition to its well known glycolytic activity, GAPDH has recently been reported to have numerous non glycolytic activities: the 37-kDA monomer may function as a uracil DNA glycosylase (UDG). Consistent with a role in DNA repair, GAPDH is found not only in the cytosol, where it is an abundant protein, but also in the nucleus. Moreover, GAPDH binds to transfer RNA in a sequence-specific fashion and to AU-rich mRNA and may play a role in the export of nuclear tRNA and mRNA. GAPDH binds also single-stranded DNA. GAPDH has also been observed to associate with the plasma membrane, filamentous actin, tubulin and microtubules.
Physiology/Pathology	GAPD shows different maximal activity and concentrations in different human tissues. In thromboasthenia there is a deficiency of the enzyme. Increased levels of GAPD protein and mRNA are found in pancreatic adenocarcinoma and in human lung cancer. High mRNA levels are also found in hepatocarcinoma and in renal carcinoma. High enzyme activity levels have been reported in serum of patients with acute hepatitis, infectious mononucleosis, tumors and after myocardial infection.
Degradation	As most proteins, the mechanism of proteolytic degradation is used by the cell catabolic machinery.
Genetics/Abnormalities	In the human genome only one functional gene has been isolated coding for a single polypeptide chain. Multiple GAPD-related sequences, due to the presence of an elevated number of pseudogenes, have also been found. The functional locus has been localized on chromosome 12, p13. No genetic abnormalities have been detected in the human gene and this may be related to the essential metabolic role of the enzyme.
Half-life	3.1 days in liver.
Concentration	23 - 41 $\mu mol\ min^{-1}\ L^{-1}$ (U/L) or 384 - 685 $nmol\ s^{-1}\ L^{-1}$ (nKat/L).
Isolation Method	Pure crystalline GAPD has been isolated from a number of different sources. Methods of purification are based upon the solubility of the enzyme-NAD complex at high concentrations of ammonium sulfate. The pure enzyme can be obtained from liver, kidney, heart and skeletal muscle by ammonium sulfate fractionation followed by chromatography on DEAE Sephadex G-50, molecular sieve chromatography and crystallization from 65% to 70% ammonium sulfate. GAPD can be isolated also from erythrocyte ghosts by ammonium sulfate precipitation followed by affinity chromatography on immobilized NAD^+.
Amino Acid Sequence	The coenzyme binding site is similar in tertiary structure to those found in a number of other dehydrogenases. The GAPD protein sequence is highly conserved, particularly the active site and the aa residues of the NAD binding site.

Disulfides/SH-Groups	Three SH groups per subunit.
General References	Wang, C.S., Alaupovic P. *Arch. Biochem. Biophys.* 1980, **205**:136-145. Martinelli, R. et al. Expression of human glyceraldehyde-3-phosphate dehydrogenase in normal and neoplastic tissues. In: *Pathology of gene expression.* Frati, L. and Aaronson, S.A. (eds.), Raven Press/Serono Symposia 1989, pp. 35-60. Mansur, N.R. et al. *Nucl. Acids Res.* 1993, **21**:993-998. Singh, R. et al. *Science* 1993, **259**:365-368. Nagy, E. et al. *J. Biol. Chem.* 1995, **270**:2755-2763.
Ref. for DNA/AA Sequences	Arcari, P. et al. *Nucl. Acids Res.* 1984, **12**:9179-9189. Ercolani, L. et al. *J. Biol. Chem.* 1988, **263**:15335-15341. Martinelli, R. et al. *Biochem. Genetics* 1989, **27**:439-450.





Diagrammatic representation of the GAPD domains. A: The NAD^+ binding domain. B: The catalytic domain. (Reproduced by courtesy of Buehner et al., JMC (1974, 90:25–49)

Glycogen phosphorylase

Christopher B. Newgard

Synonyms	α-glucan phosphorylase, 1,4-D-glucan : orthophosphate D-glcosyltransferase
Abbreviations	Phosphorylase, GP
Classifications	EC 2.4.1.1
Description	Glycogen phosphorylase is the rate limiting enzyme of glycogenolysis or glycogen breakdown. It exists in an active phosphorylated form (glycogen phosphorylase a or GPa) and an inactive dephosphorylated form (glycogen phosphorylase b or GPb). The activites of GP in both the b and a forms can also be regulated by a variety of allosteric ligands. There are three distinct genes encoding GP isozymes in humans that are named muscle (M), liver (L) and brain (B) after the tissue in which they are preferentially expressed. The primary aa sequences of the three human enzymes are related as follows: M compared to L = 80% identical; M compared to B = 83% identical; B compared to L = 80% identical.
Structure	The primary sequences of the three human phosphorylase isozymes are known. No crystal structures are yet available for these proteins. The crystal structure of rabbit muscle GPa is known at 2.1 Å resolution and that of rabbit muscle GPb is known at 1.9 Å resolution. High resolution structures are also available for the rabbit muscle enzyme in the presence of a variety of allosteric ligands. Structures for the liver or brain phosphorylase isozymes from any species are not available.
Molecular Weight	GP exists mainly as a homodimer. Subunit mw for the human isoforms are: liver GP 97,228 Da (cDNA cloning); brain GP 99,069 Da (cDNA cloning); muscle GP 96,965 (genomic sequencing).
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	GP (α-1,4 glucan : orthophosphate glycosyl transferase) catalyzes the first (and rate limiting) step in the intracellular degradation of glycogen: $(\alpha\text{-1,4-glucoside})_n + Pi \leftrightarrow (\alpha\text{-1,4-glucoside}_{n-1} + \alpha\text{-D-glucose-1-P})$. The cleavage of glycogen occurs by a phosphorolytic mechanism and utilizes orthophosphate to generate glucose-1-phosphate, without utilization of metabolic energy in the form of nucleotide triphosphate. The equilibrium constant of this reaction for rabbit muscle phosphorylase at pH 6.8 is 0.28.
Coenzymes/Cofactors	GPs have an absolute requirement for pyridoxal phosphate. This cofactor is found in the active site and is covalently linked via a Schiff base to lysine 680 in the liver, muscle and brain isoforms of human phosphorylase. This cofactor aids in the phosphorolytic cleavage of glycogen, probably by serving as a proton donor and acceptor (general acid-base catalyst). Phosphorolytic cleavage requires exclusion of water from the active site, hence the critical role of the pyridoxal phosphate cofactor.
Substrates	The single biological substrate of GP is glycogen. The enzyme removes glucose units in α-1,4 linkage until a branchpoint (α-1,6 linked residue) is

encountered, at which point a separate enzyme known as the debranching enzyme is required to remove the α-1,6 linked sugars. Although GP also can catalyze the reverse reaction (incorporation of glucose-1-phosphate into glycogen) in vitro, the enzyme functions solely in the direction of glycogen degradation in vivo because the ratio of P_i to glucose-1-phosphate concentration greatly exceeds the equilibrium constant.

Inhibitors

GP activity is inhibited by a number of allosteric ligands. Glucose is a potent biological inhibitor that binds at the active site of the enzyme. It exerts its effect by causing a conformational change from the "R" (relaxed or active) conformation to the "T" (tight or inactive) conformation. In the case of phosphorylase a, this conformational change renders the enzyme a better substrate for the phosphatase that removes the phosphate covalentely linked to Ser-14, thus leading to a rapid conversion of the enzyme to the b or inactive form. The enzmye is also inhibited by ATP, ADP, and glucose-6-phosphate. These ligands bind at some distance from the active site, using components of the AMP activator site, and cause a similar R to T conformational change as does glucose. AMP is a potent activator of the muscle and brain phosphorlyase b enzymes at concentrations below 5 mM, although it does not activate liver phosphorylase b. At higher concentrations, this nucleotide can bind to the so-called "purine inhibitor site" which is comprised of a slot formed by parallel "stacking" of the hydrophobic side chains of Phe-285 and Tyr-613, leading to inhibition of activity. The non-biologic ligand caffeine also binds at the purine inhibitor site and can act synergistically with glucose to inhibit the enzyme.

Biological Functions

GP degrades glycogen, the storage form of glucose. The product of the reaction, glucose-1-phosphate is rapidly converted to glucose-6-phosphate in a reaction catalyzed by phosphoglucomutase. In this way, the glucose building blocks of glyocogen are made avialable for cellular metabolism, since glucose-6-phosphate can enter the glycolytic sequence directly, or in the case of liver, be converted to glucose for release into the bloodstream and transfer to other tissues.

Physiology/Pathology

The physiological role of GP varies depending of the isoform and site of expression. Thus, the primary function of liver GP is to participate in the control of blood glucose levels. This enzyme is activated by a high glucagon : insulin ratio in the blood stream, a milieu typically encountered during the transition from the fed to the fasted state. The combined action of GP, which breaks down glycogen, the storage form of glucose, and glucose-6-phosphatase, the enzyme which converts glucose-6-phosphate to glucose, allows the liver to function as an organ of glucose production. Glucose production during periods when exogenous nutrients are not entering the system is a critical component of glucose homeostasis, protecting against hypoglycemia and resultant nervous system dysfunction. Muscle GP, in contrast, serves primarily to provide fuel for muscle metabolism and ensures that the energy required for bursts of muscle activity is available. Channeling of glucose-6-phosphate generated from glycogenolysis and the action of phosphoglucomutase into glycolysis is ensured by the absence of glucose-6-phosphatase in msucle. Glycogenolysis in muscle in response to stressfull circumstances (the so-called "fight or flight" response) is mediated by catecholamine-induced conversion of GPb to GPa. Brain GP appears to function primarily to protect the brain from transient periods of ischemia, anoxia or hypoglycemia. A limited role for this enzyme is implied by the fact that the brain stores far less glycogen than either liver or muscle.

Absence or underproduction of muscle and liver GP have been described. Genetic deficiency in muscle GP is known as McArdle's disease or glyco-

	gen storage disease V. Patients with this disorder have reduced muscle function, and are susceptible to severe cramping and discomfort in response to even moderate exercise. Genetic deficiency in liver GP is known as Hers' disease or glycogen storage disease VI. Patients with this disorder have enlarged livers, and in severe cases, cirrhosis and liver dysfunction. They also tend to exhibit mild to moderate fasting hypoglycemia. A genetic deficiency of the brain isozyme of phosphorylase has not been described in humans.
Degradation	Unknown
Genetics/Abnormalities	See Physiology/Pathology
Half-life	Unknown
Concentration	The concentration of human GP isozmes in tissues is not known. The concentration of liver GP in crude liver extracts of mouse has been estimated to be in the range of 5–9 μg/mg total protein by immunochemical techniques (Roesler, W. J. et al. *Arch. Biochem. Biophys.* 1986, **244:** 397–407).
Isolation Method	GP may be purified by combinations of: 1) Precipitation of the enzyme by addition of saturating glycogen or with ethanol followed by ammonium sulfate, 2) DEAE-cellulose chromatography (elution with NaCl, 0–0.5 M), 3) finally size exclusion (Sephadex G-25) or affinity chromatography (AMP-Sepharose). Liver and muscle tissues are rich sources of GP enzyme. Recently human liver GP has been expressed in bacteria (Coats, W. S. et al., *J. Biol. Chem.* 1991, **266:** 16113–16119). The purification involves addition of rabbit liver glycogen to the crude bacterial supernatant and centrifugation at 35,000 rpm to pellet the glycogen/enzyme complex. This material is then resuspended and bound to a column containing 6-aminohexyl-Sepharose and eluted with a gradient of 0–50 mM Na_2SO_4, according to a published method (Livanova, N. B. et al. *FEBS Lett.* 1976, **69:** 95–98).
Amino Acid Sequence	The aa sequences of the human liver, muscle, and brain GP are known.
Disulfides/SH-Groups	The GP enzymes do not contain disulfide bridges.
General References	Newgard, C. B. et al. *CRC Critical Reviews in Biochemistry and Molecular Biology* 1989, **24:** 69–99. Newgard, C. B. et al. *J. Biol. Chem.* 1988, **265:** 3850–3857. Johnson, L. N. and Barford, D. *J. Biol. Chem.* 1990, **265:** 2409–2412. Graves, D. J. and Wang, J. H. *Enzymes* 1972, **7:** 435–482. Flettrick, R. J. and Madsen, N. B. *Ann. Rev. Biochem.* 1980, **49:** 31–61.
Ref. for DNA/AA Sequences	Newgard, C. B. et al. *CRC Critical Reviews in Biochemistry and Molecular Biology* 1989, **24:** 69–99. Newgard, C. B. et al. *J. Biol. Chem.* 1988, **263:** 3850–3857.

Glyoxalase I

Paul J. Thornalley

Synonyms	S-D-Lactoylglutathione methylglyoxal lyase (isomerising)
Abbreviations	GlxI; GloI (protein); *GLO* (gene); GLO^1 and GLO^2 (alleles); GloI-A and GloI-E (expression products of alleles GLO^1 and GLO^2, respectively); GLO 1-1; GLO 1-2 and GLO 2-2 (allozymes).
Classifications	EC 4.4.1.5
Description	A dimeric, zinc-containing metalloenzyme, found in the cytosol of all cells (its presence in organelles is less well-established). It consists of two identical subunits in homodimers and two similar subunits in the heterodimer. Each subunit consists of 183 aa residues and differ in aa sequence only at position 111: in subunit GloI-A there is an alanine residue, and in subunit GloI-E there is a glutamic acid residue.
Structure	There is a zinc-binding motif formed by Glu-100, His-127 and one of Asp-121, His-127, Cys-139 or Glu-143. There is also an active site tyrosine residue. There is a proteolytic cleavage site at Ala-92/Leu-93 and sequence homology of aa 100-111 with the "catalytic loop" of triose-phosphate isomerase. Mechanistic similarities of the enzymatic reactions of glyoxalase I and triosephosphate isomerase suggested that Glu-100 of glyoxalase I may be the active site base involved in the catalytic mechanism. Crystals of human glyoxalase I have been obtained (unit cell dimensions $a = b = 67$ Å and $c = 171$ Å, space group $P4_1$ or $P4_3$) but the crystal structure has not yet been reported.
Molecular Weight	46kDa (gel filtration), 42kDa (sequence)
Sedimentation Coeff.	4.0 S
Isoelectric Point	4.8
Extinction Coeff.	Unknown
Enzyme Activity	Hemithioacetal(methylglyoxal-reduced glutathione)-S-D-lactoylglutathione isomerase. It catalyses the isomerisation of the hemithioacetal, formed non-enzymatically from methylglyoxal and reduced glutathione, to S-D-lactoylglutathione. The reaction mechanism proceeds via an ene-diolate intermediate. For the methylglyoxal-glutathione hemithioacetal, the K_M value is 71-130 µM and the k_{cat} value is 7-11 x 10^4 min^{-1}. The glutathionyl hemithioacetal of other α-oxoaldehydes are similarly converted to the corresponding S-2-hydroxyacylglutathione; usually the (R)-isomer is the major stereoisomer.
Coenzymes/Cofactors	Reduced glutathione. It forms the hemithioacetal substrate with methylglyoxal in a non-enzymatic pre-equilibrium.
Substrates	Physiological substrates: methylglyoxal and glyoxal. Synthetic substrates: aliphatic and aromatic α-oxoaldehydes (e.g. hydroxypyruvaldehyde, phenylglyoxal).
Inhibitors	S-p-Bromobenzylglutathione: a substrate analogue, competitive inhibitor (Ki = 83 nM). S-(N-p-Bromphenyl-N-hydroxycarbamoyl)glutathione: an

analogue of the ene-diolate mechanistic intermediate, a competitive inhibitor (Ki = 14 nM).

Biological Functions

The conversion of methylglyoxal and glyoxal to S-D-lactoylglutathione, and S-glycolylglutathione, respectively. This is the first step in the glyoxalase pathway which catalyses the conversion of methylglyoxal and glyoxal to D-lactate and glycolate, respectively. Other acylic α-oxoaldehydes are detoxified similarly to the corresponding aldonic acid via an S-2-hydroxyacylglutathione intermediate.

Physiology/Pathology

Detoxification of methylglyoxal and glyoxal to prevent the modification of nucleotides and proteins by methylglyoxal and glyoxal that would otherwise lead to mutagenesis, apoptosis, protein degradation and induction of a cytokine-mediated immune response. Failure of glyoxalase I to prevent accumulation of methylglyoxal and glyoxal in diabetes mellitus is linked to associated chronic clinical complications. Pharmacological inhibition of glyoxalase I leads to methylglyoxal accumulation and apoptosis. Inhibitor prodrugs of glyoxalase I have antiproliferative activity and are under development as antitumour and antimalarial agents.

Degradation

Unknown

Genetics/Abnormalities

The gene *GLO* is a two-allelic system, GLO^1 and GLO^2, located on chromosome 6, near the HLA complex. The locus of *GLO* is 6p, 6 cM centromeric to the HLA-DR locus. The alleles are inherited in a simple co-dominant manner. Expression of GLO^1 and GLO^2 produces GloI-A and GloI-E subunits, giving rise to one of two phenotypes GLO 1-1 or GLO 2-2 in homozygotes and three phenotypes GLO 1-1, GLO 1-2 and GLO 2-2 in heterozygotes. The relative electrophoretic mobilities of the allozymes are: GLO 1-1 - slow, GLO 1-2 heterodimer - intermediate, and GLO 2-2-fast. GLO^1 allele frequency is highest (0.7 - 0.9) for indigenous tribes in Alaska, and decreases east and south (0.4 - 0.5 in Europe), to low values (0 - 0.1) for indigenous tribes in Oceania.

Half-life

Unknown

Concentration

Tissue concentration (μg/mg protein): adipose tissue 0.04, adrenal gland 0.15, brain 0.20, heart 0.14, kidney 0.13, liver 0.12, lung 0.10, muscle 0.14, small intestine 0.16, spleen 0.17. Specific activity is higher in fetal than adult tissue.

Isolation Method

Butanol-chloroform solvent precipitation, ammonium sulphate fractionation, S-hexylglutathione affinity chromatography, gel filtration; blue dextran chromatography may also be used (erythrocytes).

Amino Acid Sequence

The translation product contains 184 aa. The N-terminal Met is removed in post-translational processing and the N-terminal Ala blocked by an as-yet, unknown modification. There are at least 4 possible phosphorylation sites. The sequence identity of human glyoxalase I with the bacterial enzyme (*Pseudomonas putida*) is 55% and with the yeast enzyme between residues 1-182 and 183-326 (*Saccharomyces cerevisiae*) is 47%, suggesting glyoxalase I of different origins may have arisen by divergent evolution from a common ancestor.

Disulfides/SH-Groups

Three disulfides.

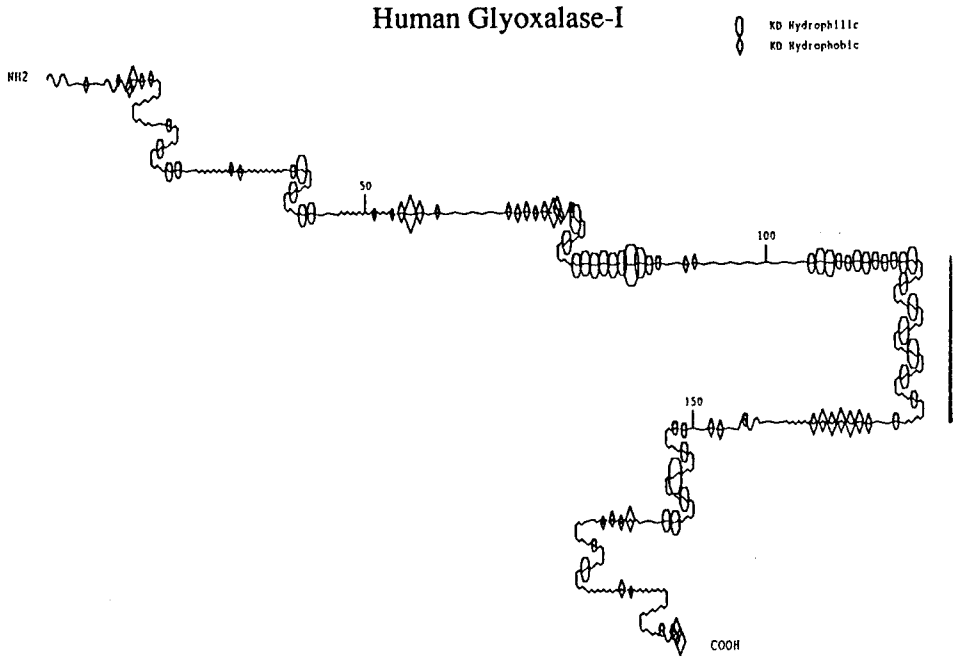
General References

Thornalley, P.J. *Mol. Aspects of Med.* 1993, **14**:287-371.
Thornalley, P.J. *Amino Acids* 1994, **6**:15-23.

Thornalley, P.J. *Crit. Rev. Oncol. Hematol.* 1995, **20**:99-128.
Lan, Y. et al. *J. Biol. Chem.* 1995, **270**:12957-12960.

Ref. for DNA/AA Sequences

Ranganathan, S. et al. *J. Biol. Chem.* 1993, **268**:5661-5667.
Kim, N.S. et al. *J. Biochem.* 1995, **117**:359-361.
Ridderström, M. and Mannervik, B. *Biochem. J.* 1996, **314**:453-467.



Composite secondary structure plot of human glyoxalase I (reproduced with permission from Ranganathan, S. et al., *J. Biol. Chem.* 268:5661-5667).

Glyoxalase II

Paul J. Thornalley

Synonyms	S-2-Hydroxyacylglutathione hydrolase
Abbreviations	GlxII; GloII (protein); *HAGH* (gene); $HAGH^1$ and $HAGH^2$ (alleles)
Classifications	EC 3.1.2.6
Description	A monomeric protein, present in the cytosol of all cells and in mitochondria (its presence in other organelles is less-well established).
Structure	Active site arginine, histidine and lysine
Molecular Weight	29 kDa (SDS PAGE), 29 kDa (sequence)
Sedimentation Coeff.	Unknown
Isoelectric Point	8.3 - 8.5
Extinction Coeff.	Unknown
Enzyme Activity	Thiolester hydrolase. It catalyses the hydrolysis of S-2-hydroxyacylglutathiones to reduced glutathione and the corresponding aldonic acid. The catalytic mechanism is thought to involve acylation of an active site histidine residue.
Coenzymes/Cofactors	None
Substrates	Physiological substrates: S-D-lactoylglutathione and S-glycolylglutathione. Other S-2-hydroxyacylglutathione derivatives (S-L-glyceroylglutathione and S-mandelylglutathione) are used as synthetic substrates. For S-D-lactoylglutathione, the K_M value is 172μM and the k_{cat} value is 755 s^{-1}.
Inhibitors	Substrate analogue, competitive inhibitors: S-p-nitrobenzoxycarbonylglutathione (K_i = 1.2μM) and S-(N-p-Bromophenyl-N-hydroxycarbamoyl)glutathione. Also, methylglyoxal-reduced glutathione hemithioacetal, the physiological substrate of glyoxalase I (K_i = 834 μM).
Biological Functions	It catalyses the final step in the glyoxalase pathway which converts S-D-lactoylglutathione to D-lactate and reduced glutathione, and S-glycolylglutathione to glycolate and reduced glutathione.
Physiology/Pathology	It regenerates reduced glutathione from the products of the glyoxalase I-catalysed reaction and thereby minimises the steady state consumption of reduced glutathione by the glyoxalase pathway.
Degradation	Unknown
Genetics/Abnormalities	Genetic locus: HAGH, location 16p13.3 Usually only one phenotype is expressed, although a rare second form (allele frequency 0.016) was found in a Micronesian population.
Half-life	Unknown

1. Installment 1992 (revised 1997)

Concentration	≈ 20μg/g of haemoglobin cells in erythrocytes and 0.2μg/mg of protein in liver and brain.
Isolation Method	Affi-gel blue chromatography, S-benzoxycarbonylglutathione affinity chromatography (erythrocytes).
Amino Acid Sequence	It is a single polypeptide chain of 260 aa residues.
Disulfides/SH-Groups	Unknown (there are 6 cysteine residues in the aa sequence deduced from the cDNA sequence).
General References	Thornalley, P.J. *Mol. Aspects of Med.* 1993, **14**:287-371. Thornalley, P.J. *Crit. Rev. Oncol. Hematol.* 1995, **20**:99-128. Allen, E. et al. *Eur. J. Biochem.* 1993, **213**:1261-1267.
Ref. for DNA/AA Sequences	Ridderström, M. et al. *J. Biol. Chem.* 1996, **271**:319-323.

Granulocyte colony-stimulating factor (G-CSF)

Nicos A. Nicola

Synonyms	Colony-stimulating factor-β; Pluripoietin; Macrophage granulocyte inducer-1G
Abbreviations	G-CSF; CSF-β Ppo; MGI-1G
Classifications	Growth and differentiation factor
Description	A circulating plasma protein synthesized by a variety of tissue cells (macrophages, endothelial cells, fibroblasts) in response to bacterial infection. A monomeric molecule containing two internal disulfide bonds (36 - 42, 64 - 74) and a single sulfhydryl group (17) contains a single site of O-glycosylation (THR 133) containing the sequence NeuAcα2-3Galβ1-3(+ NeuAcα2-6) GalNAcol.
Structure	A monomeric glycoprotein containing two internal disulfide bonds. The X-ray crystallographic structure of human G-CSF shows that it has the conformation of a long-chain 4-α-helical bundle similar to that of growth hormone, interferon-α, erythropoietin and leukemia inhibitory factor. The topology of the four helices (A-D) is up-up-down-down with two long overhand loops connecting the A-B and C-D helices with an additional short helical segment in the A-B loop.
Molecular Weight	19,000 (SDS-PAGE); 18,600 (aa sequence, not including carbohydrate or the 30 aa leader sequence).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.5 - 6.1, dependent on sialic acid heterogeneity
Extinction Coeff.	9.9 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Participates in the production and degree of functional activity of neutrophilic granulocytes. In vitro G-CSF stimulates the survival, proliferation and differentiation of precursors of neutrophilic granulocytes and activates neutrophils for respiratory burst activity, phagocytosis and killing of bacteria and antibody-dependent cell cytotoxicity. In patients, injected G-CSF controls the levels of circulating neutrophils, enhances netrophil recovery after cancer chemotherapy and bone marrow transplantation and reduces infections. It also dramatically elevates the circulating levels of various hemopoietic stem and progenitor cells.
Physiology/Pathology	Circulating levels of G-CSF determine the number of circulating neutrophilic granulocytes and their level of functional activity. Induced by bacterial infection. Excess G-CSF levels in mice lead to neutrophil accumulation in liver and lungs with little pathological consequences.

1. Installment 1992 (revised 1997)

Degradation	Degraded by the liver and kidneys and by circulating neutrophilic granulocytes. Degraded products appear in the blood and urine.
Genetics/Abnormalities	Synthesized from a single gene as two different mRNAs differing in the inclusion or omission of codons for three amino acids between positions 35 - 36 of the mature sequence. No other variants known. Localization: chromosome 17q11.2 - 21.
Half-life	β-phase 1.4 - 7.2 hr.dose dependent,human blood
Concentration	25 ± 20 ng/L in normal patient sera; elevated to 730 ± 900 ng/L in acute stages of infection, cyclic neutropenia and chemotherapy. Also elevated in some cases of anemia and leukemia (up to 900 ng/L). In aplastic anemia, serum G-CSF levels were inversely proportional to the neutrophil counts.
Isolation Method	Usually isolated from pathological sources (e.g., secreting tumors) by a variety of chromatographic procedures including hydrophobic, salting out, gel-filtration and reverse-phase chromatography.
Amino Acid Sequence	A single chain protein containing two internal disulfide bonds (36 - 42, 64 -74) and a single sulfhydryl group (17). A single site of 0-glycosylation (Thr 133). Weak aa homology to interleukin-6. A leader sequence of 30 aa. Two alternative transcripts differ in the inclusion or omission of the sequence Val-Ser-Glu between aa 35 and 36 of the mature sequence.
Disulfides/SH-Groups	Contains 5 cysteine residues positions (17, 36, 42, 64, 74 of the mature sequence) 4 of which are involved in internal disulfide bonds (36-42,64-74).
General References	Nicola, N.A. Granulocyte Colony-Stimulating Factor. In: *Colony-Stimulating Factors: Molecular and Cellular Biology.* Dexter, T.M. et al. (eds.), Marcel Dekker Inc.,N.Y. 1990, pp.77-109. Metcalf, D. and Nicola, N.A. *The Hemopoietic Colony Stimulating Factors: From Biology to Clinical Applications.* Cambridge University Press, Cambridge, U.K. 1995. Nicola, N.A. *Ann. Rev. Biochem.* 1989, **58**:45-77.
Ref. for DNA/AA Sequences	Nagata, S. et. al. *Nature* 1986, **319**:415-418. Souza, L.M. et. al. *Science* 1986, **232**:61-65. Nagata, S. et. al. *EMBO J.* 1986, **5**:575-581.

Granulocyte-macrophage colony-stimulating factor (GM-CSF)

Nicos A. Nicola

Synonyms	Colony-stimulating factor α; Macrophage-granulocyte inducer-1 GM; Colony-stimulating factor-2.
Abbreviations	GM-CSF; CSF-α; MGI-1GM; CSF-2.
Classifications	Growth and differentiation factor
Description	Usually present in plasma only at very low levels. Synthesized by a variety of tissue cells (endothelial cells and fibroblasts) in response to bacterial infection and by T-lymphocytes after immune activation. A monomeric molecule containing two internal disulfide bonds (54 - 96, 88 - 121) and two sites of N-glycosylation (29, 39) (complex acidic type) as well as O-glycosylation sites (3-4 near the N-terminus).
Structure	A single chain glycoprotein with internal disulfide bonds. X-ray crystallographic studies have shown that human GM-CSF has a short-chain 4-α-helical bundle structure with the helices having an up-up-down-down topology. This results in two long overhand loops connecting the A-B and C-D helices, each loop contributing a short strand of anti-parallel β-sheet. No significant aa homologies to other proteins but structurally related to interferon-γ, interleukins 2, 4 and 5 and macrophage colony-stimulating factor.
Molecular Weight	20,000 - 30,000 (SDS-PAGE) and 14,000 after treatment with N-glycanase which removes asparagine linked complex carbohydrates. 14,700 from aa sequence - not including carbohydrate or the 17 aa leader sequence.
Sedimentation Coeff.	Unknown
Isoelectric Point	4 - 5 depending on sialic acid heterogeneity
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	GM-CSF stimulates the survival, proliferation and differentiation in vitro of precursors of neutrophils, eosinophils, macrophages, megakaryocytes and to some extent erythrocytes. It also enhances the functional activities of mature neutrophils, macrophages and eosinophils associated with host defence against bacterial and parasitic infections. Together with TNF-α it stimulates the production of some types of dendritic cell involved in antigen presentation.

Physiology/Pathology	Injected GM-CSF elevates circulating levels of neutrophils, macrophages and eosinophils and enhances hemopoietic recovery after cancer chemotherapy and bone marrow transplantation. It also dramatically elevates the circulating levels of various hemopoietic stem and progenitor cells. In the periphery it may be produced, and act, locally in response to infections. Animal models of GM-CSF overproduction and clinical side effects of injected GM-CSF suggest a variety of pathologies primarily associated with activated macrophages in muscle, liver and lung causing tissue damage and enhanced autoimmune reactions. Gene deletion studies in mice have shown that GM-CSF plays an essential role in the maintenance of the surfactant-clearing capacity of lung macrophages and that its absence produces a disease similar to human alveolar proteinosis.
Degradation	Thought to be degraded by the liver and kidneys. Possibly also utilized by circulating or tissue derived granulocytes and macrophages.
Genetics/Abnormalities	Synthesized from a single gene of 2.5 Kbp with 3 introns which produces an mRNA of approx. 780 bases. The mRNA has a 3' untranslated A/U rich sequence thought to confer instability. The gene expression is inducible in various cell types by phorbol esters and Ca^{2+} (lymphocytes) or by bacterial lipopolysaccharides, interleukin-1 and tumor necrosis factors in endothelial cells and fibroblasts. Localization: Chromosome 5q21 - 31.
Half-life	α-phase, 5min.; β-phase, 0.24 - 1.18 hrs human.
Concentration	Not normally detectable (< 1 μg/L) in plasma.
Isolation Method	Isolated from phytohemagglutinin stimulated peripheral blood lymphocyte-conditioned medium or various tumor cell-conditioned media by a series of chromatographic steps.
Amino Acid Sequence	An acidic protein with a 17 aa hydrophobic leader sequence cleaved during secretion, two sites of potential N-glycosylation and two internal disulfide bonds. No significant sequence homology to other known proteins.
Disulfides/SH-Groups	The mature protein contains four cysteine residues forming two internal disulfide bonds (54 - 96, 88 - 121).
General References	Gough, N.M. and Nicola, N.A.: In: *Colony-Stimulating Factors: Molecular and Cellular Biology.* Dexter, T.M. et al. (eds.) Marcel Dekker Inc., N.Y. 1990, pp. 111-153. Metcalf, D. and Nicola, N.A. *The Hemopoietic Colony-Stimulating Factors: From Biology to Clinical Applications.* Cambridge University Press, Cambridge, U.K. 1995. Nicola, N.A. *Ann. Rev. Biochem.* 1989, **58**:45-77. Gasson, J.C. *Blood* 1991, **77**:1131-1145.
Ref. for DNA/AA Sequences	Wong G.G. et. al. *Science* 1985, **228**:810-815. Lee F. et al. *Proc. Natl. Acad. Sci. USA* 1985, **82**:4360-4364. Kaushanksky K. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:3101-3105. Miyatake T. et al. *EMBO J.* 1985, **4**:2561-2568.

Haemosiderin

Roberta J. Ward

Synonyms	None (evel Hs)
Abbreviations	None
Classifications	None
Description	An iron storage protein, found in excessive amounts in iron loaded tissues. It is found primarily within lysosomes in which it is formed by the lysosomal uptake of ferritin clusters from the cytosol. It is unclear whether excess iron entering the cell by diffusion, non-transferrin bound iron, can move directly to lysosomes without being incorporated into ferritin. The liver and spleen contains the highest concentration of haemosiderin in genetic haemochromatosis and thalassaemia respectively. The iron core has been shown to be heterogeneous with regard to its mineralisation, mixtures of amorphous ferric oxide, ferrioxyhydrite and goethite have been identified by biophysical techniques. Haemosiderin isolated from the tissues of treated genetic haemochromatosis patients tends to have a higher percentage of amorphous ferric oxide present. The core size is less than ferritin, between 5 and 6.5 nm.
Structure	Aggregated lumps of mineralization products with associated protein.
Molecular Weight	>4,000,000, variable
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coefficient	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	Unknown
Inhibitors	Unknown
Biological Functions	Plays an important role in limiting the pro-oxidant effect of iron. The iron is less available to participate in the formation of free-radicals by Fenton chemistry, especially when present as goethite or ferrihydrite.
Physiology/Pathology	In the initial formation of haemosiderin the protein will be contained within the lysosome. However with increasing loading of the cell with iron, the lysosomes will become enlarged and eventually may lyse.
Degradation	Not possible within biological tissues although the iron can be mobilised by venesection or chelators.
Genetics/Abnormalities	Unknown
Half-life	Concentration remains unchanged unless there is chemical intervention, i.e. chelation, or a rapid loss of blood when iron will be mobilised. The haemosiderin content of the tissues will diminish in both of these situations.

Concentration	Approximately 80% of iron present in iron loaded tissues will be present as haemosiderin.
Isolation Method	Haemosiderin is prepared from iron-loaded tissues after the removal of soluble ferritin. The resulting insoluble protein is mixed with 5.1 M potassium iodide and crude haemosiderin is isolated as a precipitate after ultracentrifugation for 2 h at 100,000 × g. The haemosiderin is resuspended in 4.4 M KI, recentrifuged at 100,000 × g for a further 2 h. After washing with water for 30 minutes combined with ultracentrifugation the protein is lyophilised.
Amino Acid Sequence	Gives characteristic bands at 20 kD after SDS/PAGE in all haemosiderin preparations. Haemosiderins isolated from thalassaemic patients show an additional band at 14.5 kD.
Disulfides/SH-Groups	Unknown
General References	Weir, M. P., Gibson, J. F. and Peters, T. J. *Biochem. J.* 1984, **223**:31–38. Mann, S. et al. *FEBS Lett.* 1988, **234**: 69–72. Crichton, R. R. and Ward, R. J. *Biochemistry* 1992, **30**:11255–11264.
Ref. for DNA/AA Sequences	

Heme oxygenases

Mahin D. Maines

Synonyms	HSP32 (Heat shock protein 32) is synonym for one form of heme oxygenase; HO-1.
Abbreviations	HO; HO-1; HO-2
Classifications	EC 1.14.99.3
Description	Endoplasmic reticulum (ER) bound proteins found in all cell types. The 2 forms, HO-1 and HO-2, are products of 2 different genes and differ in tissue expression and regulation. Human HO-1 is product of a single transcript of approx. 1.8 kb and is made of 288 aa; human HO-2 has two transcripts of approx. 1.3 and 1.7 kb and is made of 315 residues. The overall identity of the isozymes is only 42% at aa level. HO-1 gene expression is inducible by heme, metals, hormones and numerous other stimuli. HO-2 is only induced by adrenal glucocorticoids and its expression is developmentally regulated. Under normal conditions HO-2 is the prevalent form in most tissues.
Structure	The isozymes have a hydrophobic region at the C-terminus that serves as membrane anchor. A conserved hydrophobic region in the middle is predicted to be substrate (heme) binding site.
Molecular Weight	31,000 - 33,000: mammalian HO-1; 36,000: mammalian HO-2's; 33,500: avian HO.
Sedimentation Coeff.	Unknown
Isoelectric Point	5.4 - 5.8: HO-1; 6.4: HO-2, major of 3 bands
Extinction Coeff.	In complex with ferric heme shows absorption maximum at 405, 497 and 631nm; extinction coefficient at 405nm is about 140 mM^{-1} cm^{-1}.
Enzyme Activity	Catalyses specific cleavage of heme b molecule at α-methene carbon bridge with formation of stoichiometric amounts of biliverdin 1Xα, carbon monoxide and iron. It cleaves heme c at α, β, γ, δ methene bridges, with formation of one major product (54%) and three minor components (20%, 17% and 9%).
Coenzymes/Cofactors	Utilizes the reducing potential of NADPH, thecofactor, to convert chelated heme iron to the ferrous (Fe^{2+}) state. Flavoprotein, NADPH-cytochrome P450 reductase, directly catalyzes electron transfer from NADPH.
Substrates	Heme b (Fe-protoporphyrin IX, hemin, protoheme) free or lossely bound to protein e.g. methemoglobin, methemalbumin, isolated hemoglobin chains. Heme c derivates: hematoheme, dimethyl hematoheme, heme undecapeptide of cytochrome c, dicysteinyl hematoheme.
Inhibitors	Several nonphysiological metalloporphyrins competitively inhibit oxidation of Fe-protoporphyrin IX. The list includes: Zn-, Sn-, Mn-, Cu-, Ni-, Cd-, Cr- and Co- protoporphyrin IX, and metallocomplexes of protoporphyrin 1X derivates.

Biological Functions	Degrades heme moiety of senescent/denatured hemoglobin, myoglobin and tissue hemoproteins to biliverdin. In mammals biliverdin is converted to bilirubin in nearly all organs by the cytosolic enzyme biliverdin reductase. Bilirubin is carried in plasma bound mainly to albumin, and after conjugation, which mainly takes place in the liver, with 2 moles of glucuronic acid, it is excreted in the bile. Products of heme degradation have biological functions: biliverdin and bilirubin bound to albumin display potent antioxidant activity; iron regulates ferritin and transferrin synthesis; and carbon monoxide may act as a second messenger and neurotransmitter by binding to heme of guanylate cyclase to stimulate formation of cGMP.
Physiology/Pathology	Enhanced activity of heme oxygenase and increased production of bilirubin, in combination with a decreased rate of elimination of the bile pigment causes jaundice. Physiology/pathology of a decreased activity of heme oxygenase is not known.
Degradation	Unknown
Genetics/Abnormalities	The gene for human HO-1 is about 14 kb long, and is organized into 5 exons all of which contain protein coding region (exons 1 and 5 also contain an untranslated region). Human HO-2 gene has not been cloned, but rat HO-2 gene has been cloned and is about 13 kb organized into 5 exons and 4 introns. Occurrence of genetic aberration(s) is not known.
Half-life	18 hrs approx: HO-1; HO-2: unknown
Concentration	Concentration of heme oxygenases greatly varies in different organs and cell types. Under normal conditions spleen has the highest level of HO-1, while testes and brain have the highest HO-2 levels. In a normal liver HO-1 concentration is about 40 pmol enzyme/mg microsomal protein, the level can increase up to 20-fold after exposure to various chemicals. HO-2 concentration is about 130 pmol/mg microsomal protein in the testes. The proteins are essentially absent from the plasma.
Isolation Method	HO-1 and HO-2 are isolated by 2 different methods. Both are solubilized from ER with Triton X-100 and sodium cholate and precipitated with ammonium sulfate (0-40% for HO-1 and 35-60% for HO-2). Isozymes are isolated by a combination of a series of ion exchange and hydroxylapatite columns.
Amino Acid Sequence	A domain of 24 aa, conserved among all HO-1 and HO-2 isozymes and the avian HO, is postulated to be the heme binding site/catalytic site and has the following sequence: PELLVAHY(M/L*)GDLSGGQVLKK *M, in all mammalian HO-1's and avian HO; L, in all mammalian HO-2's.
Disulfides/SH-Groups	Whether 2-3 cysteines in mammalian HO-2's form a disulfide bond is unknown. Avian HO has 1 cysteine, mammalian HO-1 has none.
General References	Maines, M.D. *CRC Crit. Rev. Tox.* 1984, **12**:241-314. Schacter, B.A. *Seminars Hemat.* 1988, **25**:349-369. Shibahara, S. *Seminars Hemat.* 1988, **25**:370-375. Maines, M.D. *FASEB J.* 1988, **2**:2557-2568. Maines, M.D. *Heme Oxygenase - Clinical Applications and Functions.* CRC Press 1992, pp. 1-276. Weber, C.M. et al. *J. Neurochem.* 1994, **63**:953-962. Ewing, J.F. and Maines, M.D. *Endocrinology* 1995, **136**:2294-2302.

Ref. for DNA/AA Sequences

Yoshida, T. et al. *Eur. J. Biochem.* 1988, **171**:457-461.
Cruse, I. and Maines, M.D. *J. Biol. Chem.* 1988, **263**:3348-3353.
Shibahara, S. et al. *Eur. J. Biochem.* 1989, **179**:557-563.
Rotenberg, M.O. and Maines, M.D. *Arch. Biochem. Biophys.* 1991, **290**: 336-344.
Evans, C.O. et al. *Biochem. J.* 1991, **273**:659-666.
McCoubrey, W.K. et al. *Arch. Biochem. Biophys.* 1992, **295**:13-20.
McCoubrey, W.K. et al. *Biol. Reprod.* 1995, **53**:1330-1338.

Hemoglobin

H. Franklin Bunn

Synonyms	Haemoglobin; Globin
Abbreviations	Hb (preferred); Hg and Hgb also used.
Classifications	Heme protein
Description	Hemoglobin is a heme protein composed of two pairs of unlike globin polypeptide subunits, each of which binds to heme (iron-protoporphyrin IX). Hemoglobin tetramers are generally given the designation $\alpha_2\beta_2$. Other globin peptides that arise during ontogeny are designated by additional Greek letters. Thus, human embryonic hemoglobins include $\zeta_2\gamma_2$, $\zeta_2\varepsilon_2$ and $\alpha 2\varepsilon_2$ while human fetal hemoglobin is designated $\alpha_2\gamma_2$. Hemoglobin subunits are produced specifically and exclusively in erythroid precursor cells. Hemoglobin comprises approximately 95% of the total protein in mature circulating red blood cells and is responsible for the transport and unloading of oxygen.
Structure	The primary aa sequence has been determined on globins from over 200 species including approx. 120 mammals and approx. 50 other vertebrates as well as smaller numbers of invertebrates, plants and bacteria. In addition, over 500 human hemoglobin variants have been structurally characterized. The human α globin subunit contains 141 residues while the β globin subunit has 146 residues. The hemoglobin tetramer ($\alpha_2\ \beta_2$) dissociates readily into $\alpha\beta$ dimers under physiologic conditions of pH and ionic strength. Native hemoglobin has about 75% α helical secondary structure, no β pleated sheet and no disulfide bonds. The three-dimensional structure of human hemoglobin has been established by x-ray crystallography at high (1.6 Å) resolution. The structures of a number of human hemoglobin variants as well as other animal hemoglobins have also been determined by x-ray diffraction. The molecule is a globular tetramer with a single dyad axis of symmetry. The heme groups are inserted into hydrophobic clefts on the surface of each of the four polypeptide subunits. During deoxygenation, there is considerable movement of the globin subunits along intersubunit surfaces.
Molecular Weight	64,650 (tetramer).
Sedimentation Coeff. Isoelectric Point	7.0 (human hemoglobin A).
Extinction Coeff.	540nm: e_{oxy} 14.2: $e_{cyanmet}$ = 11.0 (mM, heme).
Enzyme Activity	Hemoglobin binds oxygen reversibly.
Coenzymes/Cofactors	The oxygen affinity of hemoglobin is influenced by pH, pCO_2, and organic phosphates. In comparison to oxygenated hemoglobin, deoxyhemoglobin has a higher affinity for protons at certain sites, particularly immidazole groups on histidines. The resulting pH dependency of oxygen binding is commonly known as the Bohr effect. In like manner, the most abundant organic phosphate in human and mammalian red cells, 2, 3 bisphosphoglycerate, binds preferentially to deoxyhemoglobin. Therefore, the concentration of 2,3 BPG or other organic phosphates in red cells is an important determinant of oxygen affinity. Carbon dioxide can bind reversibly to the

N-terminal amino groups of globin subunits by formation of carbamino adducts ($\alpha NH_2 + CO_2$ 1 $\alpha NHCOOH$; $\beta NH_2 + CO_2$ 1 $\beta NHCOOH$).

Substrates

Oxygen

Inhibitors

Carbon monoxide, nitric oxide and other heme ligands.

Biological Functions

Hemoglobin is responsible for the transport of oxygen in the blood and the unloading of oxygen to tissues. The presence of two pairs of unlike subunits endows the molecule with cooperative behavior, resulting in a sigmoid oxygen binding curve. This phenomenon, comonly also known as heme-heme interaction, permits the unloading of an enhanced amount of oxygen over a narrow range of oxygen tension, thereby facilitating the release of oxygen to respiring tissues. The pH dependency on oxygen binding is also physiologically appropriate, enabling optimal uptake of oxygen in the lungs, where pH is increased owing to CO_2 expulsion, and optimal release of oxygen in the capillary circulation owing to a decrease in pH as the result of influx of CO_2 from respiring tissue. Hemoglobin also serves as an important blood buffer, allowing the transport of large amounts of CO_2 from tissue to lungs with only a modest decrease in blood pH. In addition it transports a small amount of nitric oxide bound to a reactive sulfhydryl group.

Physiology/Pathology

Normal cooperative behavior and normal oxygen affinity is essential for the physiologic transport of oxygen from the lungs to the tissues. Mutant hemoglobins having abnormally increased oxygen affinity impose impaired delivery of oxygen to tissues, resulting in secondary erythrocytosis. Conversely, if the oxygen affinity is abnormally low, there may be cyanosis owing to decreased oxygen saturation of the arterial blood. Cyanosis also may be observed in mutant hemoglobins in which there is a rapid oxidation of Fe^{2+} in heme to Fe^{3+} (methemoglobin). A common and important disorder found in man, sickle cell disease, is due to a globin mutation (β^6Glu→Val).When Hb S is deoxygenated, it readily forms polymers resulting in distortion of the red cell into a sickle shape and occlusion of blood flow in the microcirculation.

Degradation

The senescent red cell is destroyed in the monocyte-phagocyte system. The globin subunits are proteolyzed and the heme group is degraded by heme oxygenase, resulting in the conversion of the porphyrin ring to a tetrapyrrole, designated bilirubin, and the transfer of the iron into a storage protein, ferritin.

Genetics/Abnormalities

Over 500 human hemoglobin variants have been characterized, many of which have provided important insights into structure-function relations of normal human hemoglobin. The great majority of variants arose because of a single base substitution within the coding portion of a globin gene, resulting in a single aa substitution. Other variants have arisen because of deletions, frame shifts, or mutations in the termination codon, leading to either extended chains or premature termination of translation. Structural human hemoglobin variants can give rise to secondary polycythemia, cyanosis, hemolysis, or sickling (see above). In addition, a number of globin gene mutations result in decreased synthesis of structurally normal globin, resulting in a marked imbalance between α and β globin subunit production. These disorders, commonly known as the thalassemias, result in the formation of red cells with decreased hemoglobin synthesis and, in many cases, decreased red cell survival owing to the membrane damage imposed by unbalanced globin chain production.

Half-life	120 days (human hemoglobin).
Concentration	Whole blood 150g/L; erythrocyte: 330g/L.
Isolation Method	Hemoglobin can be readily isolated from anticoagulated blood by removal of plasma and washing of red cells with an isotonic buffer, followed by hypotonic lysis, resulting in a solution that is greater than 95% hemoglobin. Hemoglobin can be further purified by a number of techniques utilizing high performance liquid chromatography.
Amino Acid Sequence	There is considerable homology among globin subunits from different species, particularly in domains of functional importance, such as a) heme binding crevices, b)subunit interfaces, c)residues responsible for the binding of organic phosphate and Bohr protons.
Disulfides/SH-Groups	None; β subunit has 2 SH groups; α subunit has one.
General References	Bunn, H.F. and Forget, B.G. *Hemoglobin: Molecular, Genetic and Clinical Aspects.* W.B. Saunders Co., Philadelphia 1986. Bunn, H.F. Sickle cell hemoglobin and other hemoglobin variants. In: *The Molecular Basis of Blood Diseases.* Stamatoy-annaopoulos, G. et al. (eds.), W. B. Saunders, Philadelphia, 1992. Dickerson, R.E. and Geis, I. *Hemoglobin: Structure-function, Evolution, Pathology.* Benjamin/Cummings, Menlo Park 1983. Edelstein, S.J. An allosteric theory for hemoglobin incorporating asymmetric states to test the putative molecular code for cooperativity. *J. Mol. Biol.* 1996, **297**:737-744. Honig, G.R. and Adams, J.G. *Human Hemoglobin Genetics.* Springer-Verlag, Wien 1986. Jia, L., Bonaventura, C., Bonaventura, J. and Stamler, J.S. S-nitrosohaemoglobin: a dynamic activity of blood involved in vascular control. *Nature* 1996, **380**:221-228. Kleinschmidt, T. and Sgouros, J.G. Hemoglobin sequences. *Biol. Chem. Hoppe-Seyler* 1987, **368**:579-615. Perutz, M.F. Molecular anatomy, physiology and pathology of hemoglobin. In: *The Molecular Basis of Blood Diseases.* G. Stamatoyannopoulos et al. (eds.), W. B. Saunders, Philadelphia, 1987.
Ref. for DNA/AA Sequences	

Hemopexin

Ursula Muller-Eberhard and Stephan Immenschuh

Synonyms	β_1-glycoprotein
Abbreviations	Hx; Hpx
Classifications	Electr. Mob.: β_1 region at pH 8.6
Description	A plasma heme binding glycoprotein synthesized mainly in the liver. The mature human Hx consists of a single polypeptide chain of 439 aa residues with six carbohydrate side chains, O-linked Thr-1, and N-linked Asn-41, Asn-164, Asn-217, Asn-223, and Asn-430. The human Hx precursor protein contains a leader sequence of 23 aa residues deduced from the gene sequence. One high affinity binding site for heme (Kd < 1 pM), and lesser affinities for other metalloporphyrins and porphyrins.
Structure	The tertiary structure of Hx is unknown. The secondary structure, predicted from aa sequence and CD measurements, contains largely random chain, some ß-structures and little α-helix.
Molecular Weight	63,000 (20% carbohydrate), calculated; 58,000 (SDS-PAGE)
Sedimentation Coeff.	4.33-4.80 S
Isoelectric Point	5.8±0.05
Extinction Coeff.	19.7: apo-Hx (280nm, 1%, 1cm); 21.8: heme-Hx (280nm, 1%, 1cm); 19.2: heme-Hx (414nm, 1%, 1cm)
Enzyme Activity	None
Coenzyme/ Cofactors	None
Substrates	None
Inhibitors	There are no known inhibitors of heme binding.
Biological Functions	Major intravascular heme transporter. Hx is the heme binding protein with the highest affinity for heme of any intra- or extracellular protein described. Hx is a scavenger of circulating heme. When heme is bound by Hx, in a molar ratio of heme: protein of < 1.0, it does not participate in the formation of radicals which oxidize lipids and proteins.
Physiology/ Pathology	Hx besides conserving iron is a powerful antioxidant preventing heme-catalyzed tissue damage. Therefore, Hx is one of the positive acute-phase reactants which are transport proteins with antioxidant activity. A type-2 interleukin-6 response element has been identified in the proximal promoter region of the human and rat Hx genes. A liver preference enhancer element has been identified in the proximal promoter region of the rat Hx gene.
Degradation	Hx is degraded in all organs, but preferably in hepatocytes when saturated with heme. Details of the degradation process of Hx are unknown.

Genetics/ Abnormalities	The human Hx gene is on chromosome 11 and spans about 12 kb with 10 exons. Southern blot analysis reveals only one copy in the haploid genome. The gene consists of 10 internal repeats which correspond directly to the position of the exons, suggesting that the gene has been evolved through intron mediated duplications of a primordial sequence. Disease caused by Hx abnormalities have not been reported.
Half-life	7 days: apo-Hx in plasma
Concentration	0.5-1.0 g/L in plasma
Isolation Method	Affinity chromatography on heme-agarose followed by ion-exchange chromatography on SP-Sephadex. Further purification can be achieved by wheat germ lectin chromatography.
Amino Acid Sequence	Homology: extensive with vitronectin, some with stromelysin. Many extracellular and cytosolic proteins share hemopexin-type repeats.
Disulfides/ SH-Groups	Six disulfides, no titrable free sulfhydryl.
General References	Muller-Eberhard, U. *Methods Enzymol.* 1988, **163**:536-565. Muller-Eberhard, U. and Morgan, W.T. *Ann. N.Y. Acad. Sci.* 1975, **244**: 624-650. Muller-Eberhard, U. and Fraig, M. *Am. J. Hematol.* 1993, **42**:59-62. Swerts, J.-P. et al. *J. Biol. Chem.* 1992, **267**:10596-10600. Nagae, Y. et al. *Biochem. Biophys. Res. Comm.* 1992, **185**:420-429. Potter, D. et al. *Arch. Biochem. Biophys.* 1993, **300**:98-104. Satoh, T. et al. *Proc. Natl. Acad. Sci. USA* 1994, **91**:8423-8427. Satoh, H. et al. *J. Biol. Chem.* 1994, **269**:6851-6858. Immenschuh, S. et al. *J. Biol. Chem.* 1994, **269**:12654-12661. Kietzmann, T. et al. *Biochem. Biophys. Res. Comm.* 1995, **213**:397-403.
Ref. for DNA/AA Sequences	Altruda, F. et al. *Nucl. Acid Res.* 1985, **13**:3841-3859. Nikkilä, H. et al. *Biochemistry* 1991, **30**: 823-829.

Heparin cofactor II

Douglas M. Tollefsen

Synonyms	Leuserpin 2
Abbreviations	HCII
Classifications	Electr. mob.: α_2-globulin (trailing edge)
Description	A single-chain protein containing 480 aa. Synthesized in the liver. Circulates in plasma as a monomer. Contains approx. 10% carbohydrate by weight, including 10 moles mannose, 8.1 moles galactose, 8.5 moles N-acetylglucosamine, and 5.9 moles sialic acid per mole protein. Contains 2 tyrosine-O-sulfate groups per molecule.
Structure	Not yet crystallized.
Molecular Weight	65,600 (sedimentation equilibrium)
Sedimentation Coeff.	4.31 S
Isoelectric Point	4.9 - 5.3 (5 major bands)
Extinction Coeff.	11.7 (280nm, 1%, 1cm) determined by refractometry in the ultracentrifuge.
Enzyme Activity	None
Coenzymes/Cofactors	Activated by dermatan sulfate, heparin, and certain other polyanions. Binds non-specifically to heparin oligosaccharides 84 sugars in length. Preferentially binds to dermatan sulfate oligosaccharides containing the structure (iduronic acid 2-sulfate Δ N-acetylgalactosamine 4-sulfate)$_3$.
Substrates	None
Inhibitors	None
Biological Functions	Inhibits the blood coagulation protease thrombin. Thrombin attacks a specific peptide bond (termed the reactive site) near the C-terminal end of HCII and becomes trapped as a stable enzyme-substrate complex. The rate of complex formation is increased approx. 1000-fold in the presence of heparin or dermatan sulfate. Accelerated inhibition requires interaction of the N-terminal acidic domain of HCII with anion-binding exosite I of thrombin. Does not inhibit other proteases involved in coagulation or fibrinolysis. Inhibits chymotrypsin and cathepsin G, but the rate of inhibition of these proteases is not increased by heparin or dermatan sulfate.
Physiology/Pathology	Plasma concentrations of HCII < 50% of normal occur in a few patients with thrombosis and in approx. 1% of healthy individuals.
Degradation	Thrombin-HCII complexes are cleared from the circulation by the low density lipoprotein receptor-related protein on hepatocytes. Partial degradation of HCII by neutrophil elastase or cathepsin G produces peptides with potent chemotactic activity for neutrophils and monocytes.
Genetics/Abnormalities	Gene localized to human chromosomal band 22q11. Restriction fragment length polymorphisms with *Bam*HI, *Msp*I and *Hind*III. Protein synthesized

	from a 2.3 kb mRNA in human liver. A variant form ($HCII_{Oslo}$: Arg-189→His) has a selective defect in binding of dermatan sulfate but binds to heparin normally. A frameshift mutation in the codon for Asp-88 ($HCII_{Awaji}$) results in reduced levels of circulating HCII.
Half-life	2.5 days (blood circulation)
Concentration	Plasma: 1.2 ± 0.2 μmol/L (mean ± SD) in adults, 0.5 ± 0.3 μmol/L (mean ± SD) in neonates.
Isolation Method	Isolated from citrated plasma by chromatography on heparin-agarose, QAE-Sephadex, and Sephacryl S-300.
Amino Acid Sequence	Homologous to α_1-antitrypsin and other members of the serpin family. Tyrosine-O-sulfation sites: EDDDY^{60}LD and EDDDY^{73}ID. Potential N-linked glycosylation sites: N^{30}, N^{169} and N^{368}. Reactive site (thrombin cleavage site): $L^{444}S^{445}$. Heparin-binding site: includes K^{173} and K^{185}. Dermatan sulfate-binding site: includes K^{173} and K^{185}. Dermatan sulfate-binding site: includes K^{185}, R^{189}, R^{192} and R^{193}. Chemotactic peptide: D^{49}-I^{66}.
Disulfides/SH-Groups	None; 3 free sulfhydryl groups.
General References	Tollefsen, D.M. *Thromb. Haemost.* 1995, **74**:1209-1214. Church, F.C. and Hoffman, M.R. *Trends Cardiovasc. Med.* 1994, **4**:140-146. Maimone, M.M. and Tollefsen, D.M. *J. Biol. Chem.* 1990, **265**:18263-18271. Van Deerlin, V.M.D. and Tollefsen, D.M. *J. Biol. Chem.* 1991, **266**: 20223-20231. Phillips, J.E. et al. *J. Biol. Chem.* 1993, **268**:3321-3327. Sheehan, J.P. et al. *J. Biol. Chem.* 1994, **269**:32747-32751. Kounnas, M.Z. et al. *J. Biol. Chem.* 1996, **271**:6523-6529.
Ref. for DNA/AA Sequences	Blinder, M.A. et al. *Biochemistry* 1988, **27**:752-759. Ragg, H. and Preibisch, G.J. *J. Biol. Chem.* 1988, **263**:12129-12134. Herzog, R. et al. *Biochemistry* 1991, **30**:1350-1357. GenBank: M12849, M19241, X03498, J05309.

Hepatic Lipase

John S. Hill and Howard Wong

Synonyms	Hepatic triacylglycerol lipase, salt-resistant lipase, monoacylglycerol acyltransferase, post-heparin lipase, phospholipase A-1.
Abbreviations	HL, HTGL, PLA-1
Classifications	EC 3.1.1.4; lipase
Description	HL is a glycoprotein containing about 18% carbohydrate. This lipase is synthesized and secreted exclusively by hepatocytes, although enzyme activity has been detected within adrenal and ovary tissues. The physiologically functional enzyme is bound to the luminal surface of hepatic endothelial cells, presumably through ionic interaction with heparan sulfate. There is some evidence that HL exists as a dimer in solution; however, this remains to be verified.
Structure	The crystal structure of HL has not been determined, however the crystal structure of a highly related protein, pancreatic lipase, suggests that the monomer is composed of two globular domains which correspond to the catalytic and substrate binding functions of HL.
Molecular Weight	65,000 for the monomeric form as determined by SDS-PAGE. The predicted molecular weight of the polypeptide chain as determined by cDNA cloning is 53,400.
Sedimentation Coeff.	Unknown
Isoelectric Point	9 (urea); 4.9 (glycerol).
Extinction Coeff.	Unknown
Enzyme Activity	HL activity is commonly measured in post-heparin plasma at pH 8.0 with emulsified triolein substrate. In normal subjects, the post-heparin activity is about 200 nmol free fatty acid hydrolyzed/min/mL. Purified HL has a specific activity of approximately 15,000 μmol/h/mg.
Coenzymes/Cofactors	HL requires no polypeptide cofactor for catalytic activity. However, several apolipoproteins have been shown to modify the hydrolytic activity of the lipase, *in vitro*, namely apolipoprotein A-I, A-II, C-II, C-III, and E.
Substrates	*In vitro*, HL hydrolyzes ester bonds at the Sn 1 position of long and short chain triacylglycerols, monoacylglycerols, phosphatidylcholine and phosphatidylethanolamine. In addition, this lipase hydrolyzes water-soluble, monodisperse esters and is capable of transacylation reactions.
Inhibitors	*In vitro*, HL is inhibited by diisopropylfluorophosphate and boronic acid derivates, but not by sulfhydryl reacting reagents, suggesting the catalytic center of the enzyme contains a serine residue. Recent site-directed mutagenesis experiments support this view.
Biological Functions	The function of HL *in vivo* has not been fully characterized. Several studies have suggested that this lipase is involved in the metabolism of chylomicron remnants, intermediate density lipoproteins and high density lipoproteins. When HL hydrolyzing activity is inhibited *in vivo*, the concentration of low, intermediate and high density lipoproteins increase.

Physiology/Pathology	A small number of families have been documented with hepatic lipase deficiency. Subjects with complete HL deficiency have elevated plasma cholesterol and triglyceride levels, and some have premature atherosclerosis.
Degradation	Unknown
Genetics/Abnormalities	The HL gene is located on the long arm of chromosome 15 (q21) and is about 35 kb in length consisting of 9 exons interrupted by 8 introns. Both the size and organization of the HL gene are very similar to the LPL gene as they are members of the same gene family. There have been six reported DNA sequence variants of the HL gene. However, only two appear to be functionally relevant, S267F and T383M.
Half-life	Unknown
Concentration	Normally not present in plasma. However, in post-heparin plasma HL concentration ranges from 120–250 ng/ml.
Isolation Method	Employing heparin-Sepharose affinity chromatography as a component of a 4-step procedure, HL has been purified from post-heparin plasma to homogeneity.
Amino Acid Sequence	HL aa sequence has been determined from cDNA sequencing. The newly synthesized enzyme is 499 aa, including a 23 aa leader sequence. The cDNA sequence has four potential N-linked glycosylation sites, which are all utilized in the mature secreted protein. The central region of HL is highly homologous with regions of lipoprotein and pancreatic lipases, suggesting conserved function is localized to this region. The putative active-site region of these lipases share a glycine-X-serine-X-glycine motif (where X=any aa); serine 145 is believed to be the acylatable site of the HL.
Disulfides/SH-Groups	HL contains 10 cysteine residues; based on homology to other members of the lipase gene family, it is believed all are disulfide linked.
General References	Davis, R. C. et al. *J. Biol. Chem.* 1992, **263**: 21499–21504. Ameis, D. et al. *Sem. Liver Dis.* 1992, **12**: 397–402. Olivecrona, T. and Bengtsson-Olivecrona, G. *Curr. Opin. Lipidol.* 1993, **4**: 187–196. Hegele, R. A. et al. *Arterioscler. Thromb.* 1993, **13**: 720–728.
Ref. for DNA/AA Sequences	Stahnke, G. et al. *Differentiation* 1987, **35**: 45–52. (cDNA sequence) Ameis, D. et al. *J. Biol. Chem.* 1990, **265**: 6552–6555. (gene sequence)

Histidine-rich glycoprotein

H. Roger Lijnen

Synonyms	Autorosette inhibition factor
Abbreviations	HRG; HRGP
Classifications	Electrical mobility: α_2-fraction
Description	Single chain glycoprotein of 507 aa, unusually rich in histidine (13%) and proline (12.8%). The protein contains 14% carbohydrate, attached to Asn-45, Asn-107, Asn-326 and/or Asn-327. It is synthesized in the liver and present in plasma and platelets.
Structure	Secondary structure consists of 8% α-helix, 15% β-sheet, 46% β-turn and 32% random coil.
Molecular Weight	75,000 (SDS-PAGE); 57,646 (aa composition); 67,000 (aa composition including carbohydrate) A M_r 77,000 form (SDS-PAGE) with additional glycosylation at Asn-184 was observed as a result of a Pro/Ser polymorphism at position 186 in exon 5.
Sedimentation Coeff.	3.8 S
Isoelectric Point	5.6 - 6.2
Extinction Coeff.	5.85 (278nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	No coenzymes. Divalent metal ions may act as cofactors for the interaction with heparin.
Substrates	None
Inhibitors	None
Biological Functions	Known functions of HRG include interactions with divalent metal ions (role in transport of metals), with plasminogen (antifibrinolytic role), with rosette formation between erythrocytes and lymphocytes, with fibrinogen and fibrin (role in fibrin polymerization ?), with platelet thrombospondin and with heparin (role in neutralization of its anticoagulant activity). Its primary physiological function may be unknown.
Physiology/Pathology	Increased levels may be associated with a reduction of the concentration of plasminogen available for binding to fibrin and thus with an antifibrinolytic effect. This may result in thrombotic complications, but increased levels of HRG as an isolated cause of thrombosis has been reported in only a few patients.
Degradation	A proteolytic derivative with M_r about 60,000 (SDS-PAGE) can be obtained by degradation with plasmin.

Genetics/Abnormalities	An association was reported between familial elevated HRG levels and a specific allele of a dinucleotide repeat polymorphism located between the last two exons of the 10 kb HRG gene (located on chromosome 3q28-q29).
Half-life	2.9 days
Concentration	In plasma 0.1 g/L (range 0.06 - 0.14 g/L).
Isolation Method	Ion exchange chromatography on CM-cellulose. Affinity chromatography on plasminogen-Sepharose. Affinity chromatography on heparin-Sepharose.
Amino Acid Sequence	The NH_2-terminal region is homologous with antithrombin III and contains two cystatin-like sequences in tandem. More than half of the molecule consists of 5 different types of internal repeats. The histidine-rich region (Asp-330 to His-389) shows 50% homology with HMW-kininogen, and contains 12 tandem repetitions of a 5-aa segment with consensus sequence Gly-His-His-Pro-His. Prior to and following the histidine-rich region, there are two proline-rich regions (Pro-271 to Pro-303 and Pro-398 to Pro-425).
Disulfides/SH-Groups	Not determined. There are 16 half Cys residues.
General References	Lijnen, H.R. and Collen, D. *Ann. N. Y. Acad. Sci.* 1989, **556**:181-185. Lane, D.A. et al. *J. Biol. Chem.* 1986, **261**: 3980-3986.
Ref. for DNA/AA Sequences	Koide, T. et al. *Biochemistry* 1986, **25**:2220-2225.

Hormone-sensitive Lipase

Mandeep Dhadly and Karen Reue

Synonyms	None
Abbreviation	HSL
Classification	EC 3.1.1.3; lipase
Description	HSL is a lipolytic enzyme with a broad tissue distribution. Found principally in adipose tissue, it catalyzes the rate-limiting step in the lipolysis of stored triacylglycerols to diacylglycerols, and the subsequent hydrolysis to monoacylglycerols. HSL has three unique features among lipases: It is the only triacylglycerol lipase controlled through phosphorylation; the phosphorylation, and therefore activity, is under acute hormonal and neural control; and the enzyme exhibits catalytic activity of a similar magnitude towards both triacylglycerol and cholesteryl ester substrates.
Structure	The crystal structure of HSL has not been determined. By molecular sieve chromatography, in a detergent-solubilized form, the rat enzyme appears to be a dimer composed of identical subunits. Similar studies of the human enzyme have not been reported.
Molecular Weight	85,500: predicted molecular mass from DNA sequence in the absence of post-translational modification; 88,000: observed molecular weight on SDS-PAGE.
Sedimentation Coeff.	4.5 S (rat HSL)
Isoelectric Point	6.7–6.8 (rat HSL)
Extinction Coeff.	Unknown
Enzyme Activity	Human adipose HSL, in cellular extracts, has a specific activity of 96 nmol/min/mg with diacylglycerol substrate. pH optimum is fairly broad around pH 7.0, distinguishing HSL from other lipases including lipoprotein lipase, monoacylglycerol lipase, and lysosomal lipase.
Coenzymes/Cofactors	No specific cofactor required. HSL activity is regulated by fast-acting lipolytic hormones, i.e. catecholamines, through cAMP-dependent phosphorylation. Conversely, insulin exerts an anti-lipolytic effect by decreasing the extent of phosphorylation at the regulatory site of HSL.
Substrates	HSL from most species studied hydrolyzes long chain tri-, di-, and monoacylglycerols and cholesteryl esters at the relative rates of 1 : 10 : 4 : 1.5. As the enzyme exhibits maximal activity against diacylglycerol, this substrate is usually used to assay HSL activity. HSL also exhibits positional specificity, with a 4-fold greater reactivity at 1(3)-ester bonds than at 2-ester bonds. Phosphorylation of the regulatory site enhances activity against triacylglycerols approximately 3-fold in cellular extracts (50-fold in the intact adipocyte) but does not appreciably affect activity against diacylglycerols or monoacylglycerols.
Inhibitors	HSL activity is inhibited by diisopropyl fluorophosphate indicating that a reactive serine residue is involved in its catalytic function. Other inhibitors include mercury chloride and sodium fluoride.

Biological Functions	The major energy depot in mammals is white adipose tissue, where lipids are stored during times of caloric excess. HSL is responsible for the hydrolysis of stored adipocyte lipids to free fatty acids which are then mobilized and utilized to maintain energy homeostasis. In brown adipose tissue, HSL may play an important role in thermogenesis by hydrolysing triacylglycerols to supply fatty acids for β-oxidation. A cholesteryl ester hydrolase activity in steroid producing tissues (adrenal cortex and corpus luteum) has also been attributed to HSL; it has been proposed that HSL hydrolyzes cytosolic cholesteryl esters to provide free cholesterol for biosynthesis of steroid hormones. A neutral cholesteryl ester hydrolase which can be activated by cAMP-dependent protein kinase has also been described in macrophages, cells which are known to accumulate cholesteryl esters during the development of atherosclerosis. Attempts to detect HSL mRNA in macrophages have shown very low levels in peritoneal macrophages from mouse, but no detectable HSL mRNA in human monocyte-derived macrophages.
Physiology/Pathology	Regulation of HSL activity appears to be mediated principally through phosphorylation of the regulatory site. Transcriptional regulation has been demonstrated with fasting and during pregnancy and hibernation. Currently, the link between abnormalities of HSL activity and specific disease processes is unclear.
Degradation	Factors affecting degradation are not known. More importantly, deactivation is due to the action of protein phosphatases, principally 2A but also 1 and 2C.
Genetics/Abnormalities	The human HSL gene has been mapped to chromosome 19 cent – q13.3.
Half-life	Unknown
Concentration	Although HSL protein concentration has not been determined, HSL activity has been quantitated as 0.91 μmol/min per gram of tissue wet weight for human adipose tissue.
Isolation Methods	Difficult to purify because of low abundance, hydrophobic character, and general lability. HSL has been prepared from rat adipose tissue using detergent solubilization, ion exchange, and hydrophobic interaction chromatography.
Amino Acid Sequence	The 786 aa sequence has been deduced from DNA clones and contains the putative catalytic site motif common to serine esterases (Gly-X-Ser-X-Gly; X is any aa). HSL does not share homology with other members of the mammalian lipase gene family including pancreatic lipase, hepatic lipase, and lipoprotein lipase, but shows similarity to 5 bacterial proteins, including a lipase from an antarctic psychrotrophic bacterium, an esterase from *Acinetobacter*, and bialaphos acetylhydrolases from *Streptomyces*. Phosphorylation sites occur at Ser^{551} and Ser^{553}; a putative lipid-binding domain is located in the COOH-terminal region of the protein.
Disulfides/SH-Groups	HSL requires the presence of sulfhydryl-protecting groups to retain full enzymatic activity, and has been shown to be inhibited by sulfhydryl-directed reagents, e.g. Hg ions. Although there are many Cys residues no specific disulfide bonds have been identified.
General References	Holm, C., Belfrage, P., and Fredrikson, G., *Biochim. Biophys. Acta.* 1989, **1006**: 193–197 Belfrage, P. et al. **In** *Lipases* 1984 (B. Borgstrom and H. L. Brockman, eds.), Elsevier Science Publishers, Amsterdam, pp. 365–416.

Strålfors, P. et al. **In** *The Enzymes* (P. D. Boyer and E. G. Krebs, eds.), vol. XVIII, Academic Press, New York, pp. 147–177.

Ref. for DNA/AA Sequences

Langin, D. et al., *Proc. Natl. Acad. Sci. USA* 1993, **90**: 4897–4901(Human HSL gene structure and organization)
Holm, C. et al., *Science* 1988, **241**: 1503–1506 (Chromosomal localization)
Langin, D. and Holm, C. *Trends Biochem. Sci.* 1993, **18**: 466–467.

Iduronate-2-sulphatase

Julie Bielicki and John J. Hopwood

Synonyms	Iduronate sulphatase, Iduronate-2-sulphatase, L-idurono-sulphate sulphatase, α-L-idopyranosyluronic acid 2-sulphate sulphohydrolase, 2-sulpho-L-iduronate-2-sulphatase, sulphoiduronate sulphatase.
Abbreviations	I2S,IDS
Classifications	EC 3.1.6.13
Description	A lysosomal enzyme present in all cells, plasma and sweat. The mature form isolated from liver, kidney, lung and placenta comprises two polypeptide chains which are not S-S linked and have molecular size of 43 kDa and 14.4 kDa. I2S is synthesized in the endoplasmatic reticulum as a 550 aa precursor and modified by removal of a 25 aa signal peptide before a further eight N-terminal aa are removed. Further proteolysis occurs to produce a 43 kDa polypeptide and a 95 aa C-terminal polypeptide which has a molecular size of 14.4 kDa on SDS/PAGE. This polypeptide does not contain cysteine residues and contains two potential N-glycosylation sites.
Structure	Not known. Subunit size as determined by SDS/PAGE under reducing conditions is 43 kDa and 14.4 kDa. Precursor recombinant I2S isolated from CHO cells has a native Mr of 90 kDa (FPLC TSK G3000SW ultrapac) and subunit size of 80–92 kDa (SDS/PAGE) of which 27–36% is carbohydrate.
Molecular Weight	Native Mr of human liver mature I2S is variable depending on gel and buffer conditions used. 42–65 kDa in high salt (0.5 M) and low pH (6.0) and 90–130 kDa in low salt (0.01 M) and high pH (7.6).
Sedimentation Coeff.	Unknown
Isoelectric Point	4.5–4.8: mature I2S; Precursor recomb. I2S < 4.0
Extinction Coeff.	Unknown
Enzyme Activity	I2S hydrolyzes the C2-sulphate ester bond from the non-reducing terminal iduronic acid residues of the glycosaminoglycans heparan sulphate and dermatan sulphate.
Coenzymes/Cofactors	None known.
Substrates	Biological substrates are heparan sulphate and dermatan sulphate. Diagnostic substrate is O-(α-L-iduronic acid 2-sulphate)-(1- > 4)-D-O-2,5-anhydro[1-^{3}H]mannitol 6-sulphate (IdoA2S-anM6S). Km is 5 μM and pH optimum 4.5.
Inhibitors	Mature I2S activity towards IdoA2S-anM6S

Inhibitor	Ki (μM)
IdoA-anM6S	1.7
IdoA-anM	11.7
anM6S	0.25
SO_4^{2-}	64
Ido2S-anM6S	1.0
GlcA2S-anM6S	1.0

Others inhibitors: 50% inhibition of activity with 40 mM NaCl, 30 μM Na_2HPO_4, 15 mM cupric acetate.
Precursor recombinant form: 50% inhibition of activity with 160 mM NaCl, 115 μM Na_2SO_4, 35 μM Na_2HPO_4 and 8 mM cupric acetate.

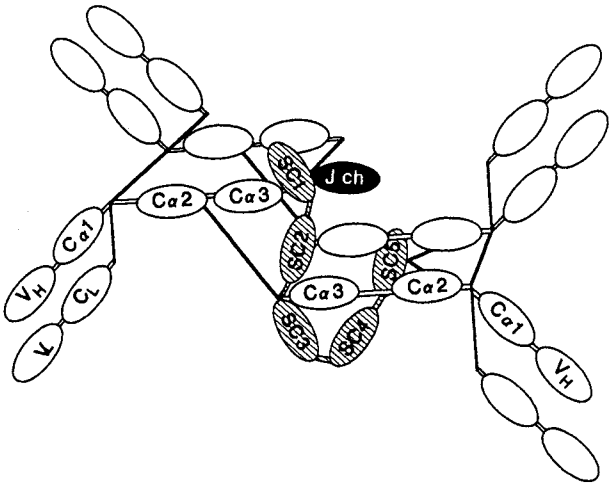
Biological Functions	I2S is one of 13 highly specific exo enzyme activities involved in the degradation of the glycosaminoglycans heparan sulphate and dermatan sulphate.
Physiology/Pathology	A deficiency of I2S activity results in storage in tissues and excretion in urine of fragments of heparan sulphate and dermatan sulphate with non-reducing termini 2-sulphated iduronic acid residues. The disorder resulting from this deficiency is known as Mucopolysaccharidosis type II or Hunter syndrome. It is inherited as an X-chromosome linked recessive trait. There is a broad spectrum of clinical phenotypes which range from very mild to extremely severe and they include such symptoms as skeletal dysmorphism, coarse facies, hepatosoplenomegaly and in the severe form there is neurological involvement.
Degradation	Unknown
Genetics/Abnormalities	Chromosome location Xq28. There is no common mutation. Genetic abnormalities are variable and include complete deletion of the gene, point mutations, partial deletions, stop mutations, frame shifts, insertions.
Half-life	Unknown
Concentration	$\simeq$ 140 μg/kg human liver.
Isolation Method	Mature form: 5 step, 6 column procedure: Step 1: Concanavalin A Sepharose-Blue A-Agarose; Step 2: Chromatofocussing chromatography (PBE94); Step 3: TSK HW50S Fractogel; Step 4: Phenyl Sepharose CL-4B; Step 5: TSK G3000SW chromatography. Precursor recombinant form: 3 step procedure: Step 1: Chromatofocussing (PBE94); Step 2: Blue A-Agarose; Step 3: TSK G3000SW chromatography.
Amino Acid Sequence	I2S sequence has strong sequence homology with other sulphatases, e.g. sea urchin arylsulphatase, human arylsulphatase A,B and C, human glucosamine 6-sulphatase and human galactose-6-sulphatase.
Disulfides/SH-Groups	None known.
General References	Bielicki, J. et al. *Biochem. J.* 1990, **271:** 75–76. Wilson, P. J. et al. *Proc. Natl. Acad. Sci. USA* 1990, **87:** 8531–8535. Wilson, P. J. et al. *Hum. Genetics* 1992, **86:** 505–508. Bielicki, J. et al. *Biochem. J.* 1993, **29:** 241–246.
Ref. for DNA/AA Sequences	Wilson, P. J. et al. *Proc. Natl. Acad. Sci. USA* 1990, **87:** 8531–8535.

Immunoglobulin A

Jiri Mestecky and Michael W. Russell

Synonyms	β_{2A}-, γ_{1A}-globulin
Abbreviations	IgA (preferred), γA
Classifications	Antibody molecule, electr. mob. β2-γ1
Description	A glycoprotein found in plasma, lymph, interstitial and cerebrospinal fluids, and external secretions (tears, saliva, milk, urine, sweat, genital, gastrointestinal and respiratory fluids); synthesized by plasma cells distributed in the mucosal tissues, secretory glands, bone marrow, spleen, and lymph nodes; daily production ($\simeq$ 66 mg/kg/day) greater than for all other immunoglobulin isotypes combined. 85–95% of plasma IgA is monomeric consisting of two heavy (α) and two light (L; κ or λ) chains linked by disulfide bridges; the remainder polymeric consisting of 2–4 disulfide-linked monomers and joining (J) chain. Secretory IgA (S-IgA) composed of 2–4 disulfide-linked monomers, J chain, and secretory component (SC), predominates in external secretions (> 95%). The α and J chains, and SC are glycosylated with various N-linked oligosaccharides. Two IgA subclasses (IgA1 and IgA2) differ in their α chain sequences, carbohydrates (IgA1 also has short O-linked GalNAc oligosaccharides), and body fluid distribution (serum IgA is mostly IgA1, external secretions contain up to 50% IgA2).
Structure	The monomeric IgA molecule is Y-shaped with 70–80 Å arms. In dimeric molecules of serum and secretory IgA two monomers are linked as depicted below. All component chains display Ig-domain structures (α-chain: 4 domains; L-chains: 2 domains; J-chain: 1 domain; SC: 5 domains).
Molecular Weight	150,000 (monomeric serum IgA); 390,000–690,000 (dimeric and tetrameric S-IgA); α-chain: $\simeq$ 54,000 (depending on subclass and glycosylation); κ- or λ-chain: 22,500; J-chain: 16,000; SC: 70,000.
Sedimentation Coeff.	7 S (serum); 11S, 15.5S (dimeric, tetrameric S-IgA)
Isoelectric Point	Variable for IgA from different sources
Extinction Coeff.	13.4 (280 nm, 1%, 1 cm); SC: 12.66 (280 nm); J chain: 6.5 (275 nm)
Enzyme Activity	None; IgA may be complexed with enzymes such as amylase, lactate dehydrogenase, galactosyl transferase, trypsin, chymotrypsin, alkaline phosphatase.
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Serum and secretory IgA antibodies neutralize biologically active antigens (viruses, enzymes, toxins). Does not activate and may inhibit complement (classical pathway) but may activate alternative pathway when aggregated or denatured (controversial); may promote phagocytosis (conflicting reports). S-IgA inhibits uptake of antigens from mucosal surfaces, and interferes with adherence of microorganisms to mucosal epithelial cells. Epithe-

	lial cells, hepatocytes, macrophages, monocytes, neutrophils, eosinophils, natural killer cells, B and T lymphocytes that transport, catabolize, or respond to IgA have various surface receptors for IgA.
Physiology/Pathology	Antibody activity in serum and external secretions. Deficiency of IgA is the most frequent immunodeficiency disorder in humans (1 case per 500–2000 individuals). IgA-containing immune complexes are frequently present in plasma of patients with IgA nephropathy, Henoch-Schönlein purpura, dermatitis herpetiformis and AIDS; complexes may be deposited in the kidney glomerular mesangium or skin.
Degradation	More than half of IgA produced is selectively transported into external secretions as S-IgA. Serum IgA is degraded mainly in the liver (most by hepatocytes, less by Kupffer cells). S-IgA is relatively resistant to gastrointestinal proteases, but many bacterial species produce IgA1-specific proteases that yield Fab and Fc fragments.
Genetics/Abnormalities	Monomeric IgA is synthesized from two different mRNA (for α- and for κ- or λ-chains). Genes coding for both α1- and α2-chains are present on chromosome 14, for L-chains on chromosome 2 (κ-chain) and 22 (λ-chain), for J-chain on chromosome 4, and for SC on chromosome 1. Two allotypic forms of IgA2, A2m(1) and A2m(2), differ in sequence, antigenic determinants, and racial distribution. Several mutant forms of IgA have been described in patients with immunoproliferative diseases.
Half-life	4.5–5.9 days (blood circulation)
Concentration	Plasma: 2.5 g/L (normal range 0.5–3.5 g/L); external secretions: < 0.01 g/L (urine) to > 12 g/L (colostrum).
Isolation Method	IgA from serum and secretions can be isolated by ammonium sulfate precipitation followed by molecular sieve, ionexchange, or hydrophobic interaction chromatography, or block electrophoresis. IgA1 can also be purified by affinity chromatography on a lectin (jacalin).
Amino Acid Sequence	Primary structures of the IgA component chains (α, L, J chains and SC) have been determined (see DNA/AA references).
Disulfides/SH-Groups	The α-chain contains 17 Cys residues (6 intra- and 5 inter-chain disulfide bonds); L-chain contains 5 Cys residues (2 intra- and 1 inter-chain disulfide bonds); J chain contains 8 Cys residues (3 intra- and 1 inter-chain disulfide bonds); SC contains 20 Cys residues (9 intra- and 1 inter-chain disulfide bonds).
General References	Heremans, J. F. Immunoglobulin A. In: *The Antigens*, Volume **II**, Sela, M. (ed.) Academic Press, New York, 1974, pp. 365–522. Kilian, M., et al. Defense mechanisms involving Fc-dependent functions of immunoglobulin A and their subversion by bacterial immunoglobulin proteases. *Microbiol. Rev.* 1988; **52:** 296–303. Mestecky, J. and Kilian, M. Immunoglobulin A (IgA). *Methods Enzymol.* 1985; **116:** 37–75. Mestecky, J. and McGhee, J. R. Immunoglobulin A (IgA): Molecular and cellular interactions involved in IgA biosynthesis and immune response. *Adv. Immunol.* 1987; **40:** 153–245. Mestecky, J. and Russell, M. W. IgA subclasses. *Monogr. Allergy* 1986, **19:** 277–301.
Ref. for DNA/AA Sequences	Eifert, H. et al.: *Hoppe-Seyler's Z. Physiol. Chem.* 1984; **365:** 1489–1495. Kabat, E. A. et al. *Sequences of proteins of immunological interest.* Fourth edition, U. S. Department of Health and Human Services, 1987. Koshland, M. E., *Ann. Rev. Immunol.* 1985; **3:** 425–453. Putnam, F. W. *Protides Biol. Fluids* 1989; **36:** 27–37.



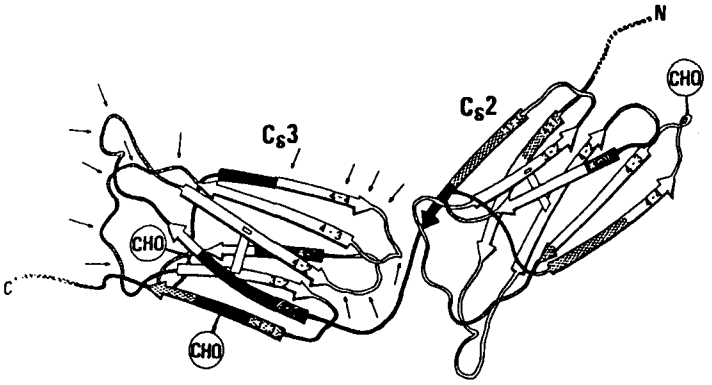
Molecular model of a human dimeric S-IgA molecule. Each elliptical area represents an immunoglobulin (V_LC_L; V_H Cα1–3) or immunoglobulin-like (SC 1–5, J chain) domain of H, L, and J chains and SC. Bold lines indicate the probable position of interchain disulfide bridges that connect the component chains.

Immunoglobulin D

Philip W. Tucker

Synonyms	None
Abbreviations	IgD
Classifications	Secreted and membrane-bound immunoglobulin
Description	Tetramer of 2 δ (H) heavy chains (66 kDa each) and two (L) light chains (22 kDa each). Structure stabilized by intra-chain (H and L) and inter-chain (H-H) disulfide bonds.
Structure	Typical immunoglobulin organization with δ H chain of one variable and three constant (C) domains. An extended hinge segment separates C_H1 and C_H2 domains. Based on homologies to other Ig, structure is primary β-barrel. Highly glycosylated.
Molecular Weight	δ chain is 66 kDa (estimated from SDS-PAGE); L chain is $\simeq$ 22 kDa; total tetramer 176 kDa. H chain contains 512 aa (56,213), 3 Glc N glycans ($\simeq$ 2,000) and 5 Gal N glycans ($\simeq$ 9,000). Each domain is $\simeq$ 110 aa; the hinge is 58 aa.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	The variable domains bind antigens (can be proteins, carbohydrates, or small organic molecules).
Inhibitors	None
Biological Functions	At low concentrations in serum, IgD functions as a typical antibody. Its most important role is as a membrane antigen receptor. Binding of antigen to membrane IgD results in endocytosis and cell activation.
Physiology/Pathology	Physiologic role is discussed above. Consequences of over-expression or loss of IgD are not known.
Degradation	Eliminated from circulation by proteolysis. Hinge particularly susceptible, yielding $F(ab)'_2$ and Fc fragments. $F(ab)'_2$ retains binding ability but no effector function.
Genetics/Abnormalities	δ H chain is in two forms: secreted (s) and membrane-bound (m). A single δ gene is alternatively processed into s and m mRNAs by use of alternative poly(A) sites. M form is 100 fold more abundant, thus the reason for paucity of IgD in serum. No genetic variants have been described and the gene has average level of polymorphism.
Half-life	2.8 days

Concentration	0.03 g/L in serum.
Isolation Method	From plasma by column chromatography on DEAE-sephadex A-50, DEAE-sephadex CL-6B, Bio-gel A-5.
Amino Acid Sequence	C-terminal portion of the hinge segment (58 aa) is highly charged (E and R): C_H3 domain has unusual P-P-P not found in other Igs.
Disulfides/SH-Groups	12 intrachain (2 in each L; 4 in each H) and 3 interchain (2 H-L, 1 H-H).
General References	Putnam, F. W., et al. The last of the immunoglobulins: Complete amino acid sequence of human IgD. In: Immunoglobulin D: Structure and Function. Thorbecke, G. J. and Leslie, G. A. (eds.) *N. Y. Acad. Sci.* 1982, **399:** 41–65.
Ref. for DNA/AA Sequences	Word, C. J., et al. *Int. Immunol.* 1989, **1:** 296–309. Kuziel, W. A., et al. *Int. Immunol.* 1989, **1:** 310–319. White, M. B., et al. *Science* 1985, **228:** 733–737.



Spatial model of the IgD Fc region The shading on the polypeptide backbone indicates the extent of sequence homology with the four other human heavy chains: open, highly conserved; cross-hatched, high divergence; solid, high homology among all except the δ chain. Arrows indicate the clustering of proline residues in Cδ3. (From Putnam et al. N.Y. Acad. Sci 1982).

Immunoglobulin E

Kimishige Ishizaka

Synonyms	
Abbreviations	IgE
Classifications	Immunoglobulin; electr. mob. $\gamma 1$
Description	IgE represents one of the five major isotypes (classes) of immunoglobulins and is identified by virtue of specific antigenic determinants which are not present in the other immunoglobulin isotypes. The IgE antibodies cause immediate type allergic reactions, which were classically called reaginic hypersensitivity. Among all isotypes of human immunoglobulins, only IgE causes the allergic reactions.
Structure	The protein is composed of two light chains and two heavy chains, which are called γ chains. Light chains are shared with the other immunoglobulin isotypes. Epsilon chains are unique for IgE, and composed of 5 homologous regions (domains) (V, $C_\varepsilon 1 - C_\varepsilon 4$). The unique antigenic determinants characteristic for IgE are associated with ε chains. IgE is cleaved into one Fc and two Fab fragments by partial hydrolysis with papain. The Fc fragment consists of one light chain and amino terminal 2/5 (Fd portion) of one ε chain. The Fd portion and Fc portion of ε chain consists of two (V and $C_\varepsilon 1$) domains and three ($C_\varepsilon 2$, $C_\varepsilon 3$, $C_\varepsilon 4$) domains, respectively. A large fragment, $F(ab')_2$, is obtained by pepsin digestion. The fragment contains V, $C_\varepsilon 1$ and $C_\varepsilon 2$ domains of two ε chains and two light chains. Complex form N-linked oligosaccharides, associated with ε chains, account for 12% of the molecular weight.
Molecular Weight	The molecular weight of one E myeloma protein (ND protein), calculated from the sedimentation coefficient and diffusion constant, was 193,000 and that determined by sedimentation equilibrium was 188,000 $\pm$ 3000. Molecular weight of another myeloma protein (PS protein) determined by sedimentation equilibrium was 184,500 $\pm$ 1500. Molecular weight of ε chain in ND and PS proteins was 69,000 $\pm$ 800 and 72,500 $\pm$ 2400, respectively.
Sedimentation Coeff.	7.92–8.20 (ND protein); 7.86 (PS protein)
Isoelectric Point	Unknown
Extinction Coeff.	15.33 (280 nm, 1%, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Since IgE has high affinity for its receptors on mast cell and basophil granulocytes, the IgE antibodies against allergens bind to these cells and the reactions of allergen to cell-bound IgE antibodies induce the release of a variety of chemical mediators, which cause allergic symptoms. An intracutaneous injection of a minute quantity of IgE antibodies, followed by challenge of the skin site with allergen induces erythema-wheal reactions.

Similarly, incubation of lung cell suspension with IgE antibodies results in selective binding of the antibodies to tissue mast cells, and incubation of the "sensitized cells" with allergen results in the release of histamine, leukotriens and prostaglandins that cause allergic inflammation.

Physiology/Pathology

Physiological role of this protein is unknown, except that IgE antibodies against certain parasites are believed to be involved in immunity against the parasite infection. Some healthy individuals have decreased concentrations of IgE to an undetectable level. An increased concentration of serum IgE is frequently observed in atopic patients. Particularly, some patients with atopic dermatitis have high levels of serum IgE. Nematode infection also cause a great increase in IgE synthesis.

Degradation

Site of degradation has not been studied. However, IgE molecules, that bound to receptors on mast cells, remain associated with the cells for several days to 2 weeks.

Genetics/Abnormalities

Since IgE represents an immunoglobulin, N-terminal domains of ε chains and light chains represent variable portion, which account for the specificity of the antibody molecules. No genetic abnormality in the constant region of the molecules has been described.

Half-life

2.3 days in serum of normal individuals

Concentration

The average concentration of IgE in normal human serum is in the order of 100 to 400 μg/L. Cord blood serum also contains IgE at the level of 12 to 100 μg/L. The serum IgE concentration in atopic individuals frequently reaches 1 mg/L to 5 mg/L.

Isolation Method

E myeloma protein can be isolated by combination of precipitation with ammonium sulfate, DEAE cellulose column chromatography and gel filtration. IgE is precipitated between 38 and 55% saturation of ammonium sulfate at pH 7.0. The precipitate fraction is dialyzed against 0.05M Tris-phosphate buffer, pH 8.0, and applied to a DEAE cellulose column equilibrated with the buffer. Proteins passed through the column are then applied to either Sephadex G 200, ACA34 or Sephacryl S-200 column to remove IgG and transferin. The major protein peak represents IgE. Since IgE is a minor component of serum proteins in normal and atopic individuals, the method described above is not sufficient to isolate normal IgE. Further purification can be made by removing IgG with Protein A-coupled Sepharose. Proteins passed through the column is then applied to anti-IgE-coupled Sepharose, and IgE bound to the immunosorbent is recovered by elution with 0.1M glycine-HCI buffer, pH. 3.0.

Amino Acid Sequence

Epsilon chain of the E myeloma protein, ND, is composed of 547 aa. Since $C_\varepsilon 2$-$C_\varepsilon 3$ domains contain unique structures involved in biologic functions of IgE, aa sequence in the regions is shown below. Numbering of the aa followed the scheme of Dorrington and Bennich. The Cys designated 225, 241 and 328 are equivalent to the Cys 237, 248 and 337, respectively, in EU scheme of Kabat and Wu.

```
224       230       240       250       260
VCSRDFTPPTVKILZSSCBGLGHFPPTIZLCLVSGYTPGT
          270       280       290       300
INITWLZBGZVMDVDLSTASTESQGELASTESQLTLSQKH
          310       320       330       340
WLSDATYTCQVTYQGHTFQDSTKKCADSNPRGVSAYLSR
          350       360       370       380
PSPFDLFIRKSPTITCLVVDLAPSKGTVNLTWSRASGKPVN
          390       400       410
HSTRKEQKQRNGTLTVTSTLPVGTRDWIEGETYQC
  420       430
RVTHPHLPRALMRSTTKTSG
```

Disulfides/SH-Groups

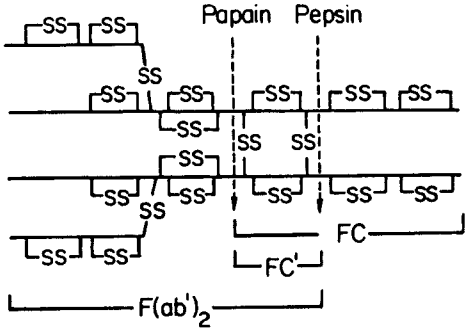
IgE molecules contain 40 half-cystine residues per molecule, 30 of which are accounted for by the two ε chains and 10 by the light chains. Ten half-cystine residues in a ε chain participate in the formation of intra-chain disulfide bridges; one in each of the 5 domains. One inter-heavy-light chain bond involves Cys-138 and two inter-heavy chain bonds are located at Cys-241 and Cys-328, respectively. The two remaining half-cystines – Cys-139 and Cys-225 – are engaged in the formation of an additional intra-ε chain bond in the $C_\varepsilon 1$ domain. Each light chain has two intra-chain disulfide bonds, and the C-terminal half-cystine is engaged in the inter-heavy-light chain bond.

General References

Bennich, H. H. and Johansson, S. G. O. *Advances in Immunology* 1971, **13:** 1.
Dorrington, K. J. and Bennich, H. H. *Immunological Reviews* 1978, **41:** 1.
Ishizaka, K. *Methods in Enzymology* 1985, **16:** 76.

Ref. for DNA/AA Sequences

Kenten, J. H., et al. *Proc. Natl. Acad. Sci. USA* 1982, **79:** 6661.
Kurokawa et al. *Nucleic Acids Research* 1983, **11:** 3077.
Liu, F.-T. et al. *Proc. Natl. Acad. Sci. USA* 1984, **81:** 5369.



Immunoglobulin G

Frantisek V. Skvaril

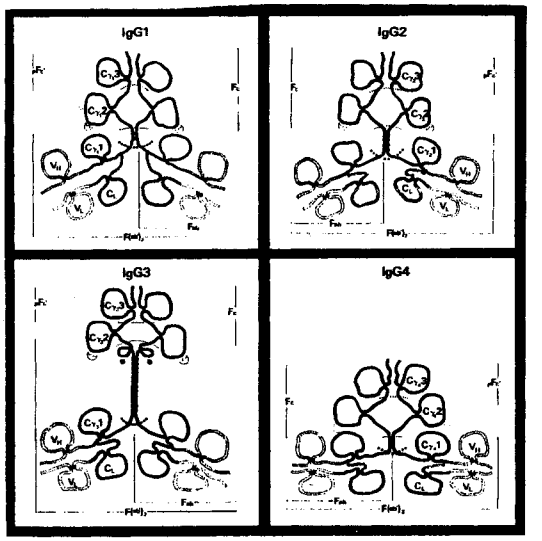
Synonyms	Gamma-globulin, γ-globulin, γ_1-globulin, γ_2-globulin.
Abbreviations	IgG, γG. Subclasses: IgG1, IgG2, IgG3, IgG4
Classifications	Electrophoretic mobility: γ_1–γ_2
Description	A circulating protein found in serum, plasma, interstitial and cerebrospinal fluids and in external secretions. Synthesized by B plasma cells in bone marrow, lymph nodes and spleen. Four subclasses exist with different antigenic, structural and biological properties residing in the heavy chains. They differ in antibody characteristics: antibodies to polysaccharide antigens belong mainly to IgG2, antiviral antibodies to IgG1 and IgG3 and reaginic type antibodies to IgG4. The relative distribution of subclasses in serum is about 60% IgG1, 30% IgG2, 4% IgG3 and 6% IgG4. Some differences include the restricted anodal electrophoretic mobility of IgG4 and cathodal mobility of IgG3. IgG3 is also characterized by a large number of interheavy disulfide bonds.
Structure	Each molecule consists of two gamma heavy chains (γ) and two identical κ or λ light chains. The heavy chains (Hc) consist of about 420, the light chains (Lc) of about 210 aa residues. The Hc are linked together and to the Lc by inter-chain disulfide bridges. Other disulfide bridges (interchain) within the Hc and Lc effect the formation of the loops (domains). The Hc consist in respect to their aa composition of one "variable" (V_H) and three "constant" (C_H1, C_H2 and C_H3) domains, the Lc of one "variable" (V_L) and one "constant" (C_L) domain. The IgG molecule can be split by papain, plasmin, etc., into two Fab fragments retaining the antigen binding site (ab) and one crystallizable (c) Fc fragment. Pepsin splits IgG into a larger $F(ab')_2$ and a smaller pFc_1 fragment.
Molecular Weight	150,000 (IgG1, IgG2, IgG4), 170,000 (IgG3). 50,000: papain Fab; 52,000: papain Fc; 60,000: papain Fc/IgG3; 50,000: Hc; 60,000: Hc/IgG3; 25,000: Lc.
Sedimentation Coeff.	7 S
Isoelectric Point	7.2–4.6 (whole IgG)
Extinction Coeff.	13.8 (280 nm, 1%, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	IgG is the most important serum antibody. It neutralizes (or clump) viruses, bacterial antigens (polysaccharides and toxins), bacteria, fungi, yeasts, etc. The binding site is localized in the N-terminal of the Fab part. IgG3, IgG1 and IgG2 activate complement (classical pathway), IgG4 does not. IgG binds to mononuclear cells, neutrophils, mast cells and basophils (differences in subclass activities).

Physiology/Pathology	Deficiency of IgG is relatively frequent and results in chronic or reccurent infections. For replacement therapy IgG preparations for i. m. or i. v. use have been developed. Sometimes, the deficiency concerns only one or two IgG subclasses: most frequent is the IgG2 deficiency which can be associated with IgA and IgG4 deficiency. – In patients with multiple myeloma, IgG bearing structural properties of one subclass and one Lc type only appears in the serum in extreme high concentration: the synthesis of other immunoglobulin classes and/or subclasses is drastically depressed. – Antigen-antibody complexes or aggregated IgG are capable to mediate tissue injury (serum sickness, glomerulonephritis, SLE, vasculitis, polyarthritis, rheumatoid arthritis, etc.).
Degradation	Proceeds probably after a conformational change in the Fc region. After pinocytosis, interchain disulfide bonds are split by disulfide reductase and cathepsins to yield small fragments.
Genetics/Abnormalities	Genes coding for Hc are present on chromosome 14, for Lc κ and λ on chromosomes 2 and 22, respectively. Three constant domains and hinge region of the Hc are encoded by separate exons; similar the two Lc domains. In the C_H gene locus the $\gamma 3$ and $\gamma 1$ are encoded sequentially as are $\gamma 2$ and $\gamma 4$ chains.
Half-life	IgG1, IgG2, IgG4: 21 days; IgG3: 7 days
Concentration	IgG (whole): 5.0–14.0 g/L. IgG1: 4.2–13.0 g/L; IgG2: 1.2–7.5 g/L; IgG3: 0.40–1.30 g/L; IgG4: 0.01–2.90 g/L (normal ranges). Concentrations in newborns and children are different.
Isolation Method	IgG (whole) from serum can be isolated by ammoniumsulfate- or Rivanol- or cold ethanol-precipitation. All subclasses can be isolated from serum by Streptococcus protein G chromatography; IgG1 + IgG2 + IgG3 can be separated from the whole IgG by anion-exchange chromatography (removes the IgG4); IgG3 can be separated from IgG1 + IgG2 + IgG4 by Staphylococcus protein A chromatography.
Amino Acid Sequence	Constant domains of the four subclasses have very similar sequences. The greatest differences are found in the hinge region (region of interchain bonds) in terms of the number of residues involved, IgG1 : 15; IgG2 : 12; IgG3 : 62; IgG4 : 12.
Disulfides/SH-Groups	IgG is organized into 12 homology regions each possesing an internal disulfide bond. The number of interchain bonds is different in IgG subclasses: IgG1 and IgG4 have two, IgG2 four and IgG3 probably 13.
General References	Porter, R. R. The hydrolysis of rabbit γ-globulin and antibodies with crystalline papain. *Biochem. J.* 1959, **73:** 119. Stanworth, D. R. and Turner, M. W. Immunochemical analysis of immunoglobulins and their subunits. In: *Handbook of Experimental Immunology*, Weir, D. M. (ed.) Blackwell, Oxford 1978, **Vol. 1**, pp. 6.1.-6.102. Burton, D. R., et al. Aspects of the molecular structure of IgG subclasses. In: *Basic and Clinical Aspects of IgG Subclasses, Monogr. in Allergy*, Shakib, F. (ed.) Karger, Basel 1986, **Vol. 19**, pp. 7–35. Skvaril, F. IgG subclasses in viral infections. In: *Basic and Clinical Aspects of IgG Subclasses, Monogr. in Allergy*, Shakib, F. (ed.) Karger, Basel 1986, **Vol. 19**, pp. 134–143. Hammarström, L. and Smith, C. I. E. IgG subclasses in bacterial infections. In: *Basic and Clinical Aspects of IgG Subclasses, Monogr. In Allergy*, Shakib, F. (ed.) Karger, Basel 1986, **Vol. 19**, pp. 122–133.

Ref. for DNA/AA Sequences

Kabat, E. A., et al. *Sequences of proteins of immunological interest.* 4th edition, US Department of Health and Human Services, 1987.
Tamamashi, N., et al. *Cell* 1982, **29**: 671–679.

Molecular model according to M. W. Turner (*Immunology Today*, Vol. 1., No. 1, July 1980)



Immunoglobulin M

Stephen J. Perkins

Synonyms	19S γ-globulin, γ-macroglobulin
Abbreviations	IgM (preferred), γM
Classifications	Secreted and membrane-bound immunoglobulin
Description	IgM is a pentameric plasma protein with a theoretical total of up to 10 antigen sites per antibody molecule, and accounts for about 7% to 10% of the immunoglobulin pool in plasma. It occurs also as a membrane-bound monomeric form. It is of relatively low affinity as measured against single determinants, but because of its high valency it binds with quite respectable avidity to antigens with multiple epitopes. IgM is a very effective agglutinator and cytolytic agent, and is produced early in the immune response against foreign pathogens (before IgG), thus making it an effective first-line defence against bacteraemia. It activates the classical pathway of complement through the presence of C1q sites on IgM which are exposed only after the multiple binding of IgM to an antigenic surface.
Structure	IgM occurs predominantly as a pentameric molecule constructed from 10 disulphide-linked μ chains and 10 disulphide-linked L chains, where each monomer is represented as μ_2L_2 and has a Y-shaped structure (see Figure). Each μ chain contains one variable-type (V) immunoglobulin (Ig) fold domain and four constant-type (C) Ig domains. Each L chain contains one variable and one constant Ig domain. The pentamer also contains an Ig-fold-like J (joining) chain which is possibly incorporated in the Fc_5 disc. Hexameric and pentameric forms of IgM that lack the J chain have also been identified. Membrane-bound IgM differs at the C-terminus of the μ chain in which Cys-575 and the fifth glycosylation site are replaced by a hydrophobic transmembrane segment. The pentamer has a planar structure in solution with a moderate degree of flexibility between the central Fc_5 disk (Cμ3 and Cμ4 domains) and the five Fab'$_2$ arms (L chain; V_H, Cμ1, Cμ2 domains). The diameter of IgM would be 37 nm if a planar IgM structure were fitted into a circle. When bound to antigen, the Fab'$_2$ arms are able to bend out of the plane of IgM and this results in a "staple-like" structure in which the Fab'$_2$ arms are in close contact with antigen. This exposes a complement binding site for C1q close to Asp-432 in the Cμ3 domain.
Molecular Weight	966,000 = M_r, when calculated from a typical sequence (IgM OU); this includes 10% carbohydrate. 190,000 = M_r of IgM-S monomer μ_2L_2.
Sedimentation Coeff.	17.7 S: IgM; 7.4 S: IgM-S; 11.4 S: Fc_5
Isoelectric Point	Unknown
Extinction Coeff.	12.5: human IgM, when calculated from a typical sequence (human IgM OU); 8.8 for the Fc_5 disk; 12.7 for the IgM-S monomer; 13.7 for the Fab'$_2$ arm; 15.3 for Fab (all at 280 nm, 1%, 1 cm).
Enzyme Activity	None
Coenzymes/Cofactors	None

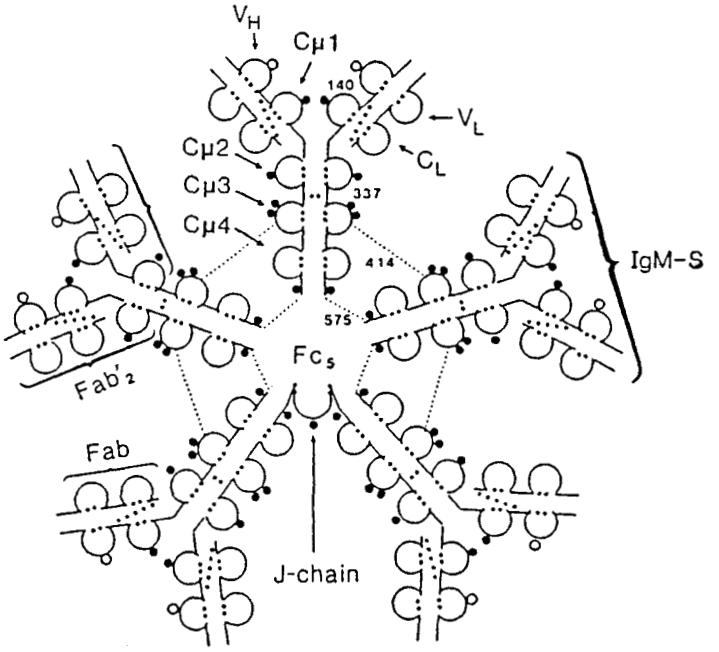
Substrates	Antigens (bind to the Fab variable domains at the hypervariable loops); C1q of complement (binds to the Cμ3 domains).
Inhibitors	None
Biological Functions	The major effector function of IgM is to bind to antigen. Once bound, it activates complement and macrophages. It appears to be particularly equipped to deal with particulate antigens in the bloodstream.
Physiology/Pathology	IgM is found in higher concentrations in patients with various immunological disorders (especially Waldenström's macroglobulinaemia, systemic lupus erythematosus, and ataxia telangiectasia). Levels of IgM are also known to be high following continuous and direct exposure to particulate antigens in blood, e.g. in malaria and trypanosomiasus (parasites with an intravascular phase), and also toxoplasmosis, congenital syphilis and congenital rubella. Autoantibodies of the IgM class (rheumatoid factors) are usually found in the serum of patients with rheumatoid arthritis, and these are directed against determinants on the Fc domain of IgG. Monomeric IgM-S is known to occur in normal human serum at low concentrations, and at higher concentrations in patients with various immunological disorders.
Degradation	Unknown
Genetics/Abnormalities	Unknown
Half-life	5 days in serum
Concentration	0.5–2 g/L in plasma
Isolation Method	(1) A raw immunoglobulin fraction which contains all the immunoglobulins (IgG, IgM, IgA, IgD, IgE) is obtained by precipitation of serum at 45% ammonium sulphate saturation, pH 7.0. This solution is not especially stable, and unstable proteins can be removed with Aerosil at pH 5.5. The separation of IgM can then be achieved by dialysis against water. The euglobulin precipitate is then dissolved in Longsworth buffer (diethylbarbituric acid 3.7 g; sodium diethylbarbiturate 16.5 g; distilled water to 1 L, pH 8.6, ionic strength 0.1), and fractionated by preparative zone electrophoresis. The final separation step on Sephadex G-200 results in purified IgM. (2) A procedure used to purify Waldenstrom's IgM is to dilute the serum of the patient 10-fold with water to give a preciptate of euglobulin. After centrifugation, the precipitate is dissolved in a minimal volume of 0.2M NaCl and reprecipitated four times with water to yield purified IgM. (3) Fragments of IgM: The IgM subunits can be readily prepared by mild reduction of IgM using mercaptoethylamine. The Fc_5 disc can be prepared by tryptic cleavage of IgM. The Fab and Fab'_2 fragments are prepared by pepsin cleavage of IgM.
Amino Acid Sequence	Each Ig fold domain contains about 110 aa and these are very similar. These lead to a seven-stranded or nine-stranded β-sheet sandwich structure for the constant- or variable-type domains respectively. There are five oligosaccharide sites on the Cμ1, Cμ2 and Cμ3 domains and on the C-terminal tailpiece of the μ chain.
Disulfides/SH-Groups	There is one intradomain disulphide bridge in each of the 70 Ig folds per pentamer. Interdomain bridges occur also, one connecting each L chain to a μ chain (Cys_L214-Cys_H140), one connecting each pair of μ chains per monomer (Cys337-Cys337), and two connecting adjacent IgM-S monomers (Cys414-Cys414 and Cys575-Cys575). The J chain contains 8 Cys residues.

General References

Beale, D. and Feinstein, A. *Biochem J.* 1969, **118:** 187–194.
Stanworth, D. R. and Turner, M. W. Immunochemical analysis of immunoglobulins and their subunits. In: *Handbook of experimental immunology*, Weir, D. M.(ed.) Blackwell, Oxford. 1978, **Vol 1**, pp 6.1–6.102.
Feinstein, A. et al. *Immunol. Today* 1986, **7:** 169–174.
Davis, A. C. and Shulman, M. J. *Immunol. Today* 1989, **10:** 118–128.
Perkins, S. J. et al. *J. Mol. Biol.* 1991, **221:** 1345–1366.

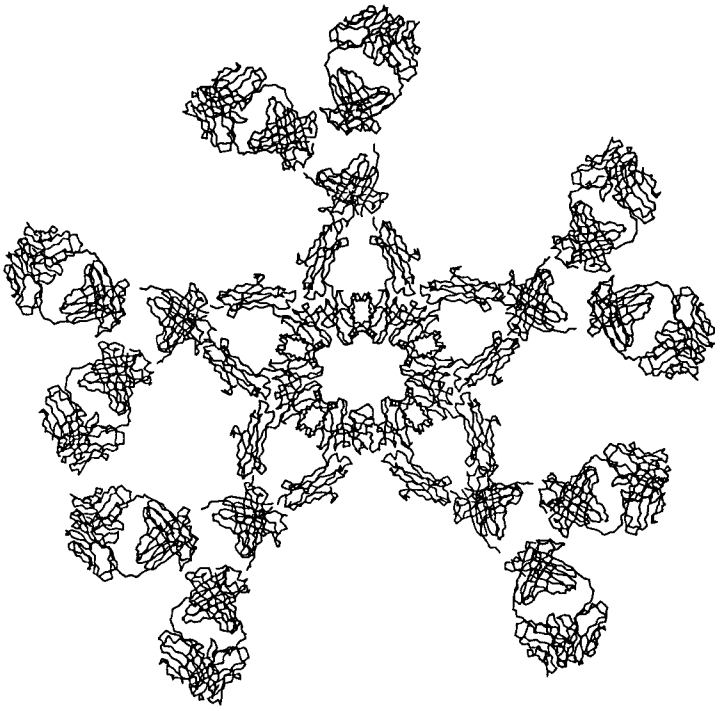
Ref. for DNA Sequences

The complete sequences for the μ chains from two Waldenstrom IgM OU and Gal are known. Sequences for the Ig fold domains are listed in Kabat, E. A. et al. (1987) *Sequences of proteins of immunological interest*, 4th edition, National Institute of Health, Bethseda, MD.



Molecular models

(1) Schematic diagram of the 71 Ig folds in the pentameric structure of IgM. Each μ chain contains 5 domains, while each L chain contains 2 domains. Disulphide bridges are denoted by dotted lines. The oligosaccharide sites are denoted by filled circles. The fragments of IgM are denoted by brackets. Taken from Perkins et al. (1991) *J. Mol. Biol.* **221:** 1345–1366



(2) Solution scattering model of the domain structure of IgM.
The α-carbon skeleton of intact IgM used to fit X-ray scattering data is shown. Two of the Fab'_2 arms are rotated sideways by 45° to yield an asymmetric structure; the IgM structure is best viewed as a planar family of structures in which any one Fab'_2 arm is rotated to one side or the other by up to 45° at any instant. Taken from Perkins et al. (1991) *J. Mol. Biol.* **221:** 1345–1366.

Insulin and proinsulin

John A. Galloway and Ronald E. Chance

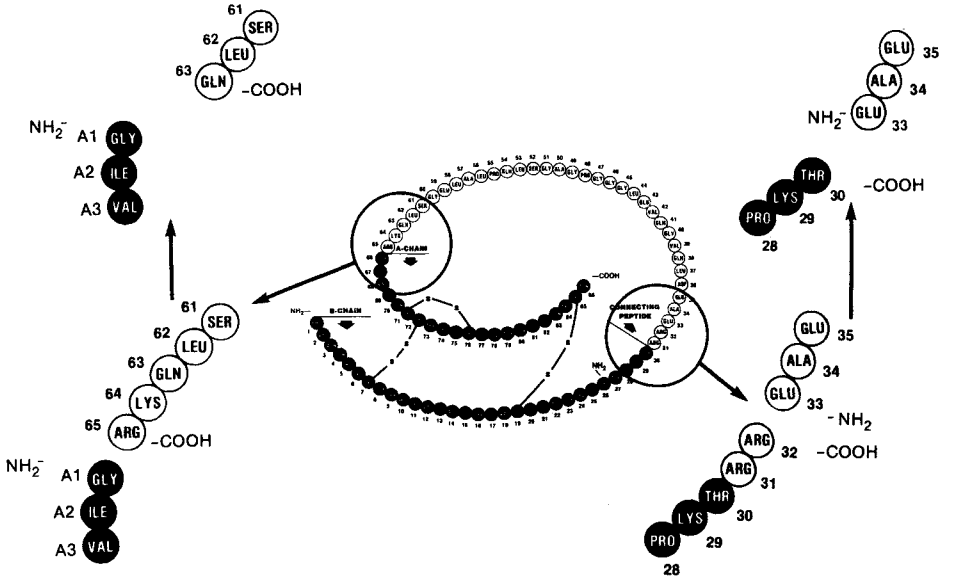
Synonyms	None
Abbreviations	HPI, HPRO (proinsulin); CP, BHCP (C-peptide)
Classifications	Protein hormones
Description	A protein hormone synthesized in the beta cells of the pancreas and secreted into the portal vein. Biosynthesis involves formation from the precursor molecules (pre-proinsulin, molecular weight about 12,000) and HPI. Single chain HPI is converted to two-chain insulin by prohormone converting enzymes which result in the secretion of equimolar quantities of insulin and CP. Small quantities of HPI and conversion intermediates are also secreted into the portal circulation (see Fig.).
Structure	Insulin is a globular structure comprised of a 21 aa A-chain and a 30 aa B-chain linked by two disulfide bonds. It is a flexible and dynamic structure containing three alpha helices. The X-ray structure is known and is essentially identical to that of pork insulin, which has been the traditional model for such studies. HPI is an 86 aa single chain protein in which the A and B chains are linked by a connecting peptide consisting of 35 aa (positions 31 to 65 in Fig.). In vivo processing of the connecting peptide consists of removal of the two terminal dibasic aa sequences to form CP (positions 33 to 63).
Molecular Weight	5807.6 for insulin; 9388.6 for HPI; 3020.3 for CP (calculated from aa sequences).
Sedimentation Coeff.	3.2S for insulin; 3.4S for HPI (based on pork material, 0.1 mM Zn).
Isoelectric Point	5.3 for insulin and HPI; 3.2 for CP
Extinction Coeff.	Insulin: 1.05; HPI: 0.65 (276 nm, 0.1 %, 1 cm); CP 0.00 (between 240 and 300 nm, 0.1 %, 1 cm).
Enzyme Activity	None established for insulin, HPI or CP.
Coenzymes/Cofactors	Unknown
Substrates	None
Inhibitors	Serum antibodies to insulin may inhibit the action of this hormone and HPI. Antibodies to the insulin receptor may also impede the action of these hormones.
Biological Functions	The principal biological function of insulin is to promote normal carbohydrate and secondarily protein and fat metabolism by binding to specific receptors located chiefly on the plasma membranes of muscle, fat and liver cells. The binding process leads to activation of transmembrane and intracellular kinases which in turn results in a cascade of biochemical reactions that activate glucose transporters as well as a variety of enzymes. This complex sequence of events brings about both non-oxidative (glycogen deposition) and oxidative disposal of glucose.

Physiology/Pathology	Absence or insufficiency of insulin leads to diabetes mellitus. When there is total destruction of beta cells by auto-immune or other inflammatory phenomena as in insulin-dependent diabetes or pancreatitis failure to provide exogenous hormone (therapeutic insulin) will result in polydipsia, polyuria, polyphagia, dehydration, ketoacidosis, coma and death. Conditions associated with resistance to the effects of insulin, such as non-insulin-dependent diabetes mellitus, usually result in chronic hyperglycemia and occasionally hyperosmolarity of the serum. Although coma may result from hyperosmolarity, insulin resistant syndromes are associated with ketoacidosis only when the toxic effects of hyperglycemia on the beta cells totally compromise endogenous insulin secretion. Tumors (benign and malignant adenomas) and hyperplasia of the beta cells may result in excessive secretion of insulin and HPI, leading to spontaneous hypoglycemia. Hypoglycemia is also a frequent complication of treatment with therapeutic insulin. The effects of endogenous or exogenous CP have not been adequately described.
Degradation	The degradation of insulin, which is the result of insulin protease, which convention now calls "insulin-degrading enzyme" or "IDE", begins with binding of insulin to its receptor and continues as insulin is internalized into the cell in endosomes and prelysosomes. While other enzymes have been implicated in the degradation process, IDE is now thought be responsible for the degradation of virtually all of the hormone presented to cells.
Genetics/Abnormalities	Three insulin mutants have been described: (1) Insulin Chicago: Leu-B25; (2) Insulin Los Angeles: Ser-B24; and (3) Insulin Wakayama: Leu-A3. In addition, two HPI mutants have been described: (1) Proinsulin Tokyo/Boston: His-65, and (2) Proinsulin Providence: Asp-10.
Half-life	Insulin: 5–8 min; HPI: 26 min.; CP: 41 min.
Concentration	Insulin: 30–900 pmol/L; HPI (endogenous): 10–150 pmol/L; CP: 300–2000 pmol/L.
Isolation Method	Pancreatic human insulin is purified from acid-alcohol extracts using isoelectric precipitation, column chromatography, and zinc crystallization. Human insulin is also made by a trypsin-catalyzed transpeptidation method starting with pork insulin in which Ala-B30 in pork insulin is replaced by Thr. Human insulin rDNA is prepared by recombinant DNA technology using microbial fermentation followed by chromatographic purification and zinc crystallization.
Amino Acid Sequence	See Fig.
Disulfides/SH-Groups	See Fig.
General References	Kitabchi, A. E., et al. Insulin synthesis, proinsulin and c peptides. In: *Diabetes Mellitus. Theory and Practice.* Rifkin, H. and Porte, D. Jr. (eds.) Elsevier, New York, 1990, pp. 71–88. Hutton, J. C. The insulin secretory granule. *Diabetologia* 1989, **32:** 271–281. Harrison, S. A., et al. Mechanism of action of insulin. In: *Diabetes Mellitus. Theory and Practice.* Rifkin, H. and Porte, D. Jr. (eds.) Elsevier, New York, 1990, pp. 61–70. Galloway, J. A., et al. Human proinsulin: review of chemistry, in vitro and in vivo receptor binding, animal and human pharmacology studies, and clinical trial experience. *Diabetes Care* 1992, **15:** 666–692.

Baker, E. N., et al. The structure of 2 ZN pig insulin crystals at I.S.A. resolution. *Phil. Trans. R. Soc. Lond.* 1988, **B319:** 369–456.
Duckworth, W. C. Insulin degrading enzyme. In: *Insulin.* Cuatrecasas, P. and Jacobs, S. (eds.) Springer, New York, 1990, pp. 143–165.
Grant, P. T., et al. Differences in the nature of the interaction of insulin and proinsulin with zinc. *Biochem. J.* 1972, **126:** 433–440.

Ref. for DNA/AA Sequences

Nicol, D. S. H. W. and Smith, L. F. Amino Acid Sequence of human insulin. *Nature* 1960, **187:** 483–485.
Frank, B. H. and Chance, R. E. Two routes for producing human insulin utilizing recombinant DNA technology. *Münch. med. Wschr.* 1983, **125 (Suppl. 1):** 14–20.
Barfoed, H. C. Insulin production technology. *Chem. Eng. Prog.* 1987, **83:** 49–54.
Bell, G. I., et al. Sequence of the human insulin gene. *Nature* 1980, **284:** 26–34.
Steiner, D. F. The biosynthesis of insulin. In: *Insulin.* Cuatrecasas, P. and Jacobs, S. (eds.) Springer, New York, 1990, pp. 67–92.
Tager, H. S. Mutant human insulins and insulin structure-function relationships. In: Insulin. Cuatrecasas, P. and Jacobs, S. (eds.) Springer, New York, 1990, pp. 41–64.



The primary structures HPI, CP and insulin. Disulfide linkages occur between 7 and 72, 19 and 85, and 71 and 76. CP consists of the sequence 33–63. This figure also demonstrates intermediates in the conversion between HPI and insulin: Split 65, 66 and des (64, 65) HPI and split 32, 33 and des (31, 32) HPI. With permission of the American Diabetes Association (*Diabetes Care* 1992, **15:** 666–692).

Insulin-like growth factor binding protein-1

Robert C. Baxter and Janet L. Martin

Synonyms	Placental protein-12; Amniotic fluid binding protein; Pregnancy-associated endometrial α_1-globulin
Abbreviations	IGFBP-1; PP12; AFBP; α_1-PEG; IBP-1; BP-28; BP-25
Classifications	IGF binding protein family; IGFBP-1 to IGFBP-6
Description	One of at least six structurally-related IGF-binding proteins, single-chain polypeptides with relatively conserved N- and C-terminal domains, and structurally distinct central domains. IGFBP-1 contains 234 aa, and may be phosphorylated at several serine residues in the central domain, with phosphorylation reported to increase IGF-binding affinity. There are no *N*-glycosylation sites, but 4% of *O*-linked glycosylation has been reported.
Structure	No crystallographic, NMR or other physical structure reported.
Molecular Weight	25.3 kDa, calculated from deduced aa sequence. Appears approx. 28 kDa on nonreduced SDS-PAGE and approx. 34 kDa on reduced SDS-PAGE.
Sedimentation Coeff.	Unknown
Isoelectric Point	Major form: 4.9, minor form: 4.8
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	hIGFBP-1 has high binding affinity ($K_a > 10^9$ L/mol) for both IGF-I and IGF-II. Its function in amniotic fluid, where it is found in the highest concentrations, is unknown. In serum, it is believed to act as a short-lived carrier of IGFs. Specifically, it has been postulated to have a glucoregulatory role, inhibiting the insulin-like activity of circulating IGFs that are not bound in a stable complex with IGFBP-3 and the acid-labile subunit. Consistent with this view, injection of IGFBP-1 in bolus form causes transient hyperglycemia in rats, and animals overexpressing the IGFBP-1 gene have fasting hyperglycemia. At the cellular level, dephospho-IGFBP-1 is able to enhance the mitogenic activity of IGF-I, whereas phosphorylated IGFBP-1 is inhibitory.
Physiology/Pathology	The predominant source of serum IGFBP-1 is the liver. The major negative regulators of IGFBP-1 is insulin, while corticosteroids generally appear stimulatory. Hepatic IGFBP-1 production is also regulated by carbohydrate independently of insulin, with substrate deprivation stimulating production by a cyclic AMP-dependent mechanism. Fasting, prolonged exercise, and insulin-induced hypoglycemia are all powerful stimuli to IGFBP-1 secretion in humans. No diseases have been reported in which

IGFBP-1 abnormalities are the primary cause. In pregnancy, IGFBP-1 is secreted by the decidua.

Degradation

IGFBP-1 appears relatively resistant to proteases than can degrade other IGFBPs. A 21-kDa N-terminal fragment with IGF-binding activity has been isolated from placenta.

Genetics/Abnormalities

The human IGFBP-1 gene, located in the 7p12-p13 region of the short arm of chromosome 7, contains 4 exons spanning approx. 5.2 kb, and encodes a single mRNA transcript of 1.55 kb. The gene has regulatory elements in common with the phosphoenolpyruvate carboxykinase gene, which shows similar regulation by insulin and glucocorticoids. No genetic abnormalities of IGFBP-1 have been reported.

Half-Life

≈ 12 min (human protein, in rats)

Concentration

Serum concentrations determined by immunoassay depend on the analytical method, probably reflecting specificity of different antibodies for different phospho-forms. Typical levels in overnight-fasted adults can vary by different techniques from approx. 10 to 100 μg/L. Levels are higher in children than adults, higher in women than men, and higher still in pregnancy. Food intake suppresses IGFBP-1, and the reported diurnal fluctuation in serum levels probably reflects varying nutritional status throughout the day. Levels up to 50 mg/L are reported in third-trimester amniotic fluid.

Isolation Method

Purification from amniotic fluid, the most abundant source, can be achieved by ammonium sulfate precipitation, hydrophobic interaction chromatography and anion-exchange, or by a simple two-step method of IGF-I affinity chromatography and reverse-phase HPLC.

Amino Acid Sequence

The three domains span residues 1-79, 80-144, and 145-234. 29% of N-terminal domain residues, and 14% of C-terminal domain residues, are conserved in IGFBP-1 to -6. The occurrence of the PEST motif in the central domain suggests rapid turnover. An RGD motif in the C-terminal domain (residues 221-223) may be involved in cell or matrix interactions.

Disulfides/SH-Groups

IGFBP-1, like IGFBP-2 to -5, has 12 disulfide-bonded cysteines within the N-terminal domain and 6 disulfide-bonded cysteines within the C-terminal domain.

General References

Lee, P.D.K., Conover, C.A. and Powell, D.R. Regulation and function of insulin-like growth factor-binding protein-1. *Proc. Soc. Exptl. Biol. Med.* 1993, **204**:4-29.

Baxter, R.C. Insulin-like growth factor binding proteins as glucoregulators. *Metabolism* 1995, **44** (Suppl 4):12-17.

Koistinen, R., Huhtala, M.L., Stenman, U.H. and Seppälä, M. Purification of placental protein PP12 from human amniotic fluid and its comparison with PP12 from placenta by immunological, physicochemical and somatomedin-binding properties. *Clin. Chim. Acta* 1987, **164**:293-303.

Busby, W.H., Klapper, D.G. and Clemmons, D.R. Purification of a 31,000-dalton insulin-like growth factor binding protein from human amniotic fluid. *J. Biol. Chem.* 1988, **263**:14203-14210.

Rutanen, E.M., Koistinen, R., Wahlström, T. et al. Synthesis of placental protein 12 by human decidua. *Endocrinology* 1985, **116**:1304-1309.

Póvoa, G., Enberg, G., Jörnvall, H. and Hall, K. Isolation and characterization of a somatomedin-binding protein from mid-term human amniotic fluid. *Eur. J. Biochem.* 1984, **144**:199-204.

Ref. for DNA/AA Sequences

SwissProt locus IBP1_HUMAN, accession P08833.
GenBank accession M20841.
Lee, Y.L., Hintz, R.L., James, P.M. et al. *Mol. Endocrinol.* 1988, **2**:404-411.
Brinkman, A., Groffen, C., Kortleve, D.J. et al. *EMBO J.* 1988, **7**:2417-2423.
Julkunen, M., Koistinen, R., Aalto-Setälä, K. et al. *FEBS Lett.* 1988, **236**: 295-302.

Insulin-like growth factor binding protein-2

Robert C. Baxter and Janet L. Martin

Synonyms	Multiplication stimulating activity (MSA) carrier protein is rat IGFBP-2.
Abbreviations	IGFBP-2
Classifications	IGF binding protein family; IGFBP-1 to IGFBP-6
Description	One of at least six structurally-related IGF-binding proteins, single-chain polypeptides with relatively conserved N- and C-terminal domains, and structurally distinct central domains. IGFBP-2 contains 289 aa, with no *N*-glycosylation sites, and no reported *O*-glycosylation.
Structure	No crystallographic, NMR or other physical structure reported.
Molecular Weight	31.3 kDa, calculated from deduced aa sequence.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	hIGFBP-2 has high binding affinity ($K_a > 10^{10}$ L/mol) for IGF-II, slightly lower for IGF-I. IGFBP-2 inhibits, and in some circumstances may enhance, IGF actions on target cells. It is expressed by many fetal tissues and may have a role in embryonic development; however, mice bearing a disrupted IGFBP-2 gene are viable.
Physiology/Pathology	In the adult circulation IGFBP-2 is postulated to act as an alternative IGF-carrier to IGFBP-3, particularly in situations where total IGFs exceed the capacity of available IGFBP-3, such as growth hormone deficiency or in the presence of IGF-II-secreting tumors. Also elevated in renal failure and prostate carcinoma. No diseases have been reported in which IGFBP-2 abnormalities are the primary cause.
Degradation	Degradation sites or pathway not delineated. C-terminal fragments of 12-14 kDa retain some IGF-binding activity.
Genetics/Abnormalities	The human IGFBP-2 gene, located in the q33-q34 region of chromosome 2, contains four exons spanning over 32 kb, mainly due to the first intron of 27 kb. The single mRNA transcript is 1.43 kb. No genetic abnormalities of IGFBP-2 have been reported.
Half-Life	≈ 14 min (bovine IGFBP-2 in rats)

Concentration	Adult human serum: 0.07-0.45 mg/L, higher in neonates. High levels in seminal plasma (11±5 mg/L) and milk (2.0±0.2 mg/L).
Isolation Method	Hydrophobic or IGF-affinity chromatography followed by reverse-phase HPLC.
Amino Acid Sequence	The three domains span residues 1-95, 96-189, and 190-289. 29% of N-terminal domain residues, and 14% of C-terminal domain residues, are conserved in IGFBP-1 to -6. The role of an RGD sequence at residues 265-267 is not yet established.
Disulfides/SH-Groups	IGFBP-2, like IGFBP-1 to -5, has 12 disulfide-bonded cysteines within the N-terminal domain and 6 disulfide-bonded cysteines within the C-terminal domain.
General References	Blum, W.F., Horn, N., Kratzsch, J. et al. Clinical studies of IGFBP-2 by radioimmunoassay. *Growth Reg.* 1993, **3**:100-104. Clemmons, D.R., Snyder, D.K., and Busby, W.H. Variables controlling the secretion of insulin-like growth factor binding protein-2 in normal human subjects. *J. Clin. Endocrinol. Metab.* 1991, **73**:727-733. Zapf, J., Kiefer, M., Merryweather, J. et al. Isolation from adult human serum of four insulin-like growth factor (IGF) binding proteins and molecular cloning of one of them that is increased by IGF-I administration and in extrapancreatic tumor hypoglycemia. *J. Biol. Chem.* 1990, **265**: 14892-14898. Baxter, R.C. Circulating binding proteins for the insulinlike growth factors. *Trends Endocrinol. Metab.* 1993, **4**:91-96.
Ref. for DNA/AA Sequences	SwissProt locus IBP2_HUMAN, accession P18065 GenBank accession M69237 Ehrenborg, E., Vilhelmsdotter, S., Bajalica, S. et al. *Biochem. Biophys. Res. Commun.* 1991, **176**:1250-1255. Binkert, C., Landwehr, J., Mary, J.L. et al. *EMBO J.* 1989, **8**:2497-2502.

Insulin-like growth factor binding protein-3

Janet L. Martin and Robert C. Baxter

Synonyms	Growth hormone dependent IGF-binding protein; Acid-stable subunit of the 150 kDa ternary complex
Abbreviations	IGFBP-3; BP-53; BP-29
Classifications	IGF binding protein family; IGFBP-1 to IGFBP-6
Description	One of at least six structurally-related IGF-binding proteins, single-chain polypeptides with relatively conserved N- and C-terminal domains, and structurally distinct central domains. IGFBP-3 has 264 aa and is subject to extensive posttranslational modification by glycosylation, phosphorylation and limited proteolysis. Neither phosphorylation nor glycosylation is believed to affect IGF-binding affinity.
Structure	No crystallographic, NMR or other physical structure reported.
Molecular Weight	28.7 kDa. Appears as a doublet of 40-45 kDa by non-reduced SDS-PAGE, due to extensive glycosylation.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	Weak disulfide isomerase activity has been reported
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	hIGFBP-3 has high binding affinity (K_a>10^{10} L/mol) for IGF-I and IGF-II. IGF-bound IGFBP-3 undergoes a second, lower affinity association (K_a of ~2.5 x 10^8 L/mol for IGF-I–IGFBP-3, and 5.8 x 10^7 L/mol for IGF-II–IGFBP-3) with a third protein, the 85 kDa acid-labile subunit, to form a stable ternary complex of ~150 kDa. The formation of this complex, which carries >90% of IGFs in the circulation, results in inhibition of IGF bioactivity, and minimization of the insulin-like potential of the IGFs. At the cellular level, IGFBP-3 can both enhance and inhibit IGF-stimulated DNA synthesis and cell proliferation. IGFBP-3 may also have cell growth inhibitory effects unrelated to its IGF-binding activity.
Physiology/Pathology	The predominant source of circulating IGFBP-3 is believed to be the liver. Plasma concentrations of IGFBP-3 are growth hormone regulated, being reduced by up to 80% in hypopituitarism, and elevated 2-3-fold in acromegaly. Age-related changes in IGFBP-3 parallel changes seen in IGF-I; levels are very low in young children, increase throughout childhood to reach a maximum in mid-puberty, and decline gradually throughout adulthood. IGFBP-3 is elevated in pregnancy and in chronic renal failure, and decreased in liver cirrhosis, poorly controlled diabetes, and untreated Cushing's disease. IGFBP-3 is also synthesized and secreted by a wide

	variety of cell types, including fibroblasts, chondroblasts, osteoblasts, and many tumor tissues.
Degradation	IGFBP-3 serves as a substrate for a wide variety of proteases, including prostate specific antigen, matrix metalloproteinases, plasmin and cathepsin D. Increased degradation of IGFBP-3 in serum is also apparent during pregnancy and various catabolic states. Fragments generated range in size from 10.8 to 30 kDa, depending on the protease involved, and the analytical method used to detect protease activity. Most of these fragments have reduced IGF-binding activity; some may have intrinsic bioactivity.
Genetics/Abnormalities	The human IGFBP-3 gene, located in the 7p14-p12 region of chromosome 7, contains 4 exons and spans 8.9 kb. The single mRNA transcript is 2.5 kb. No genetic abnormalities of IGFBP-3 have been reported.
Half-Life	Free: < 30 min; 150 kDa complex : ~15 h
Concentration	Adult serum: 2.2-4.6 mg/L; similar in amniotic fluid.
Isolation Method	Purification from Cohn Fraction IV of plasma, or cell-conditioned medium by IGF-affinity chromatography followed by reverse-phase HPLC.
Amino Acid Sequence	The three domains span residues 1-87, 88-184 and 185-264. 29% of the N-terminal residues, and 14% of the C-terminal domain are conserved in IGFBP-1 to -6. An 18-residue basic region in the C-terminal domain (residues 215-232), and additional residues in the central domain, interact with polysulfated glycosaminoglycans on cell membranes or extracellular matrix. Three N-glycosylation sites (N^{89}, N^{109} and N^{172}) are all used, and major phosphorylation sites have been reported at S^{111} and S^{113}.
Disulfides/SH-Groups	IGFBP-3, like IGFBP-1 to -5, has 12 disulfide-bonded cysteines within the N-terminal domain and 6 disulfide-bonded cysteines within the C-terminal domain.
General References	Martin, J.L. and Baxter, R.C. Insulin-like growth factor binding protein from human plasma: Purification and characterization. *J. Biol. Chem.* 1986 **261**:8754-8760 Martin, J.L. and Baxter, R.C. Insulin-like growth factor binding protein-3: biochemistry and physiology. *Growth Reg.* 1992, **2**:88-99. Lewitt, M.S., Saunders, H. and Baxter, R.C. Bioavailability of insulin-like growth factors (IGFs) in rats determined by the molecular distribution of human IGF binding protein-3. *Endocrinology* 1993, **133**:1797-1802. Kelley, K.M., Oh, Y., Gargosky, S.E. et al. Insulin-like growth factor (IGF)-binding proteins and their regulatory dynamics. *Int. J. Biochem.* 1996, **28**:619-637. Jones, J.I. and Clemmons, D.R. Insulin-like growth factors and their binding proteins: Biological actions. *Endocrine Revs.* 1995, **16**:3-34. Zapf, J. Physiological role of the insulin-like growth factor binding proteins. *Eur. J. Endocrinol.* 1995, **132**:645-654.
Ref. for DNA/AA Sequences	SwissProt locus IBP3_HUMAN, accession P17936. GenBank accession M31159. Wood, W.I., Cachianes, G., Henzel, W.J. et al. *Molec. Endocrinol.* 1988, **2**:1176-1185 Cubbage, M.L., Suwanichkul, A. and Powell, D.R. *J. Biol. Chem.* 1990, **265**:12642-12649.

Insulin-like growth factor binding protein-4

Janet L. Martin and Robert C. Baxter

Synonyms	Binding protein 24; Inhibitory IGF-binding protein
Abbreviations	IGFBP-4; BP-24; In-IGFBP
Classifications	IGF binding protein family; IGFBP-1 to IGFBP-6
Description	One of at least six structurally-related IGF-binding proteins, single-chain polypeptides with relatively conserved N- and C-terminal domains, and structurally distinct central domains. IGFBP-4 contains 237 aa, with a potential *N*-glycosylation site in the central domain. Note incorrect designation as IGFBP-5 in some publications – see aa data below.
Structure	No crystallographic, NMR or other physical structure reported.
Molecular Weight	Core protein of 26.3 kDa (from deduced aa sequence). Appears 24-26 kDa (non-glycosylated) and 28-30 kDa (glycosylated) by non-reducing SDS-PAGE.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	hIGFBP-4 has high binding affinity (K_a>10^9 L/mol) for IGF-I and IGF-II, and may carry some IGFs in the circulation. At the cellular level IGFBP-4 inhibits IGF bioactivity, and a role for IGFBP-4 in ovarian follicular development has been postulated.
Physiology/Pathology	Little is known about the physiology of circulating IGFBP-4, and pathological states in which circulating levels of the protein are altered have not been described. Serum levels are reported to be increased by growth hormone in Turner's syndrome, after administration of 1,25-dihydroxy-vitamin D3, and in elderly women with fractures. No diseases have been reported in which IGFBP-4 abnormalities are the primary cause.
Degradation	Cation-dependent proteolysis of circulating IGFBP-4 may occur during pregnancy, but the protease(s) involved have not been identified. Proteolysis of IGFBP-4 by medium conditioned by fibroblasts, osteoblasts, and by follicular fluid yields products of between 14 and 22 kDa, many of which show reduced IGF-binding activity. The activity of some of these proteases is increased by IGFs. The central, non-conserved domain of IGFBP-4 appears to contain the sites targetted by these enzymes.

Genetics/Abnormalities	The human IGFBP-4 gene is located on chromosome 17q12-21.1, close to the hereditary breast cancer gene BRCA1. Regulatory elements contained within the rat gene include cyclic AMP response elements, AP-1 binding sites, and a progesterone receptor binding site. The mRNA for IGFBP-4 is ~2.6 kb. No genetic abnormalities for IGFBP-4 have been reported.
Half-Life	Unknown
Concentration	Unknown
Isolation Method	Hydrophobic or IGF-affinity chromatography of serum or cell-conditioned medium, and reverse-phase HPLC.
Amino Acid Sequence	The three domains span residues 1-79, 80-151, and 152-237. 29% of the N-terminal residues, and 14% of the C-terminal domain are conserved in IGFBP-1 to -6. IGFBP-4 has two cysteine residues in the central domain in addition to the 18 conserved residues common to IGFBP-1 to -5. Note: The N-terminal sequence of IGFBP-4 is DEAIHCPPCSEEKLA. Incorrectly designated IGFBP-5 in some early reports.
Disulfides/SH-Groups	IGFBP-4, like IGFBP-1 to -5, has 12 disulfide-bonded cysteines within the N-terminal domain and 6 disulfide-bonded cysteines within the C-terminal domain.
General References	Jones, J.I. and Clemmons, D.R. Insulin-like growth factors and their binding proteins: Biological actions. *Endocr. Rev.* 1995, **16**:3-34. Rosen, C., Donahue, L.R., Hunter, S. et al. The 24/25-kDa serum insulin-like growth factor-binding protein is increased in elderly women with hip and spine fractures. *J. Clin. Endocrinol. Metab.* 1992, **74**:24-27. Conover, C.A., Kiefer, M.C. and Zapf, J. Posttranslational regulation of insulin-like growth factor binding protein-4 in normal and transformed human fibroblasts. *J. Clin. Invest.* 1992, **91**:1129-1137.
Ref. for DNA/AA Sequences	SwissProt locus IBP4_HUMAN, accession P22692. GenBank accession M62403. La Tour, D., Mohan, S., Linkhart, T.A. et al. *Mol. Endocrinol.* 1990, **4**: 1806-1814. Shimasaki, S., Uchiyama, F., Shimonaka, M. and Ling, N. *Mol. Endocrinol.* 1990, **4**:1451-1458. Kiefer, M. C., Masiarz, F. R., Bauer, D. M. and Zapf, J. *J. Biol. Chem.* 1991 **266**:9043-9049.

Insulin-like growth factor binding protein-5

Janet L. Martin and Robert C. Baxter

Synonyms	Human bone-derived IGFBP; Cerebrospinal fluid IGFBP-22
Abbreviations	IGFBP-5
Classifications	IGF binding protein family, IGFBP-1 to IGFBP-6
Description	One of at least six structurally-related IGF-binding proteins, single-chain polypeptides with relatively conserved N- and C-terminal domains, and structurally distinct central domains. IGFBP-5 contains 252 aa, with O-glycosylation but no N-glycosylation sites. IGFBP-5 may occur as a phosphoprotein. Note incorrect designation as IGFBP-6 in some publications – see aa data below.
Structure	No crystallographic, NMR or other physical structure reported.
Molecular Weight	28.5 kDa (from deduced aa sequence). Appears as a diffuse doublet or triplet of 30-32 kDa by non-reducing SDS-PAGE, probably as the result of some *O*-glycosylation.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	hIGFBP-5 has high binding affinity (>10^9 L/mol) for IGF-I and IGF-II, with approximately 5-fold higher affinity for IGF-II than IGF-I. In the circulation, IGFBP-5 may serve as an alternative carrier of IGFs. At the cellular level, matrix bound IGFBP-5 can potentiate IGF action in skin fibroblasts, while free IGFBP-5 inhibits IGF action. IGF-independent stimulation of osteoblast proliferation by a carboxyl-truncated form of IGFBP-5 has also been reported.
Physiology/Pathology	Not well understood. Believed to form a reservoir of IGFs, particularly IGF-II, in bone, and may be involved in ovarian follicle development.
Degradation	IGFBP-5 may be degraded by an unidentified cation-dependent protease in pregnancy serum. Other proteases secreted by fibroblasts, osteoblasts and granulosa cells degrade IGFBP-5 to yield fragments of between 17 and 23 kDa with reduced IGF-binding activity. In some cases, proteolysis can be partially inhibited by IGFs and heparin.
Genetics/Abnormalities	The gene for human IGFBP-5 is located on chromosome 2q33-34, and in a total length of 33 kb, contains a long intron of 25 kb. The mRNA for IGFBP-5 is 6 kb. No genetic abnormalities for IGFBP-5 have been reported.
Half-Life	Unknown

Concentration	500-600 µg/L in adults, higher pre-pubertally, lower in the elderly.
Isolation Method	Hydrophobic and ligand affinity chromatography and reverse-phase HPLC.
Amino Acid Sequence	The three domains span residues 1-80, 81-170 and 171-252. 29% of the N-terminal residues, and 14% of the C-terminal domain are conserved in IGFBP-1 to -6. IGFBP-5 has two highly basic sequences, one in the central, non-conserved domain, and the other, very similar to the 18 residue sequence found in the same region of IGFBP-3, in the C-terminal domain. One or both of these may be involved in association of IGFBP-5 with acidic elements on the cell surface or matrix. The N-terminus of IGFBP-5 has the sequence LGSFVHCEPCDEKAL. Note: incorrectly designated IGFBP-6 in some early reports.
Disulfides/SH-Groups	IGFBP-5, like IGFBP-1 to -4, has 12 disulfide-bonded cysteines within the N-terminal domain and 6 disulfide-bonded cysteines within the C-terminal domain.
General References	Jones, J.I. and Clemmons, D.R. Insulin-like growth factors and their binding proteins : Biological actions. *Endocrine Revs.* 1995, **16**:3-34. Jones, J.I., Gockerman, A., Busby, W.H. et al. Extracellular matrix contains insulin-like growth factor binding protein-5: Potentiation of the effects of IGF-I. *J. Cell Biol.* 1993, **121**:679-687. Andress, D.L. and Birnbaum, R.S. Human osteoblast-derived insulin-like growth factor (IGF) binding protein protein-5 stimulates osteoblast mitogenesis and potentiates IGF action. *J. Biol. Chem.* 1992, **267**:22467-22472. Mohan, S., Libanati, C., Dony, C. et al. Development, validation, and application of a radioimmunoassay for insulin-like growth factor binding protein-5 in human serum and other biological fluids. *J. Clin. Endocrinol. Metab.* 1995, **80**: 2638-2645.
Ref. for DNA/AA Sequences	SwissProt locus IBP5_HUMAN, accession P24593. GenBank accession M65062. Kiefer, M.C., Ioh, R.S., Bauer, D.M. and Zapf, J. *Biocehm. Biophys. Res. Commun.* 1991, **176**:219-225. Shimasaki, S., Shimonaka, M., Zhang, H.P. and Ling, N. *J. Biol. Chem.*, 1991, **266**:10646-10653. Allander, S.V., Larsson, C., Ehrenborg, E. et al. *J. Biol. Chem.* 1994, **269**: 10891-10898.

Insulin-like growth factor binding protein-6

Janet L. Martin and Robert C. Baxter

Synonyms	Cerebrospinal fluid IGFBP-32; Lung fibroblasts-derived IGFBP
Abbreviations	IGFBP-6
Classifications	IGF binding protein family; IGFBP-1 to IGFBP-6
Description	One of at least six structurally-related IGF-binding proteins, single-chain polypeptides with relatively conserved N- and C-terminal domains, and structurally distinct central domains. IGFBP-6 contains 216 aa, and is O-glycosylated. There is one consensus N-glycosylation site, which is not used, and phosphorylation of the protein has not been reported. Note incorrect designation as IGFBP-4 in some publications – see aa data below.
Structure	No crystallographic, NMR or other physical structure reported.
Molecular Weight	22.8 kDa (from deduced aa sequence). Appears as a diffuse species of 32-34 kDa by non-reducing SDS-PAGE as the result of extensive *O*-glycosylation.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	IGFBP-6 binds IGF-II with an estimated Ka $>10^{11}$ L/mol, approximately 30- to 100-fold higher than its affinity for IGF-I. May serve as a carrier for IGFs in the circulation. High concentrations in cerebrospinal fluid suggest a role for IGFBP-6 in regulating IGF bioactivity in the nervous system. IGFBP-6 inhibits IGF-I and IGF-II bioactivity in a wide variety of cell types.
Physiology/Pathology	Not well understood. Source of circulating IGFBP-6 not known. Serum levels are slightly higher in men than women, and are slightly elevated in pregnancy.
Degradation	Circulating IGFBP-6 protease activity has not been reported. Acid activated protease activity in cell conditioned medium may involve a cascade of serine, aspartic and cysteine proteases.
Genetics/Abnormalities	Located on chromosome 12.
Half-Life	Unknown

Concentration	Adult human serum: 0.22 ± 0.11 mg/L; cerebrospinal fluid: ~ 0.152±0.049 mg/L
Isolation Method	Hydrophobic or affinity chromatography followed by reverse-phase HPLC. Best sources are serum, cerebrospinal fluid, and cell-conditioned medium
Amino Acid Sequence	The three domains span residues 1-80, 81-137 and 138-216. 29% of the amino terminal residues, and 14% of the carboxy-terminal domain are conserved in IGFBP-1 to -6. Potential heparin-binding motif in the carboxy-terminal domain. The amino terminal sequence for IGFBP-6 is ALARCPGCGQGVQAG. Note: incorrectly designated IGFBP-4 in some early reports.
Disulfides/SH-Groups	Human IGFBP-6 has 10 disulfide-bonded cysteines within the N-terminal domain and 6 disulfide-bonded cysteines within the C-terminal domain, i.e. it lacks 2 of the N-terminal domain cysteines of IGFBP-1 to -5.
General References	Roghani M., Hossenlopp, P., Lepage, P. et al. Isolation from human cerebrospinal fluid of a new insulin-like growth factor-binding protein with selective affinity for IGF-II. *FEBS Lett.* 1989, **255**:253-258. Martin, J.L., Willetts, K.E. and Baxter, R.C. Purification and properties of a novel insulin-like growth factor-II binding protein from transformed human fibroblasts. *J. Biol. Chem.* 1990, **265**:4124-4130. Bach L.A., Thotakura, N.R. and Rechler, M.M. Human insulin-like growth factor binding protein-6 is *O*-glycosylated. *Biochem. Biophys. Res. Commun.* 1992, **186**:301-307. Baxter, R.C. and Saunders, H. Radioimmunoassay of insulin-like growth factor-binding protein-6 in human serum and other body fluids. *J. Endocrinol.* 1992, **134**:133-139.
Ref. for DNA/AA Sequences	SwissProt locus IBP6_HUMAN, accession P24592. GenBank accession M62402. Shimasaki, S., Gao, L., Shimonaka, M. and Ling, N. *Mol. Endocrinol.* 1991, **5**:938-948. Shimasaki, S., Shimonaka, M., Zhang, H.P. and Ling, N. 1991, In: *Modern Concepts of Insulin-like Growth Factors* Spencer, E.M. (ed.), Elsevier, New York pp343-358.

Insulin-like growth factor-I

Robert C. Baxter and Janet L. Martin

Synonyms	Somatomedin C; Somatomedin A
Abbreviations	IGF-I; SMC; SMA
Classifications	Insulin/IGF family
Description	A member of a peptide family including IGF-I, IGF-II, and proinsulin, IGF-I is an anabolic and mitogenic peptide secreted by many cells of the body. Its activity is modulated by a family of at least six binding proteins. Although a variety of IGF-I mRNA species have been described, resulting from both alternative splicing and alternative polyadenylation, only one form of the mature peptide is known, a single chain polypeptide of 70 aa. Its structure resembles that of proinsulin, with the N-terminal B domain (IGF-I[1-29]) joined to the A domain [42-62] by a connecting C-domain which remains in the mature peptide [30-41]. Unlike proinsulin, IGF-I has a C-terminal D domain of 8 residues [63-70]. Two major precursor (proIGF-I) forms have been described, IGF-Ia extended C-terminally by 35 residues, and IGF-Ib extended by 77 residues.
Structure	Solution structure by NMR indicates one a-helix in the B-domain, followed by a turn, and two helices in the A-domain.
Molecular Weight	7649 (calculated from primary sequence)
Sedimentation Coeff.	Unknown
Isoelectric Point	8.0-8.5
Extinction Coeff.	Unknown
Enzyme Activity	Weak disulfide isomerase activity has been reported.
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	A family of six IGF-binding proteins (IGFBPs) has been described, which inhibit the biological actions of IGF-I in many systems, apparently by blocking IGF-I binding to its receptor.
Bioogical Function s	IGF-I has both mitogenic and anabolic activities, and stimulates differentiation of some cell types. It acts both in an endocrine manner, transported through the bloodstream in association with a family of IGF binding proteins, and in an autocrine or paracrine manner, acting on cells near the site of its production. The naturally occurring des(1-3) form has a greatly reduced affinity for the binding proteins, with near-normal receptor affinity.
Physiology/Pathology	Circulating IGF-I is believed to be derived largely from the liver, where its synthesis is closely regulated by growth hormone (GH). IGF-I deficiency in childhood, whether due to GH deficiency or GH insensitivity, leads to severe growth retardation. IGF-I overproduction, most commonly the result of a pituitary tumor hypersecreting GH, leads to gigantism in children, and acromegaly when it occurs in adults. IGF-I production by tumors

	has been described *in vitro* but is not well documented to occur *in vivo*. Since a number of tumor types are driven by activation of the IGF-I receptor, it is likely that specific tumor growth could occur as a result of local IGF-I overproduction. Mice with a null mutation of the IGF-I gene are severely growth retarded at birth and most die perinatally.
Degradation	No biologically active degradation products have been described, except for des(1-3)IGF-I.
Genetics/Abnormalities	The human IGF-I gene, located on the long arm of chromosome 12, contains six exons and spans 90-100 kb of genomic DNA. Of a variety of IGF-I mRNA species, those of 7.6, 1.3 and 1.1 kb are most prominent.
Half-Life	12 min (free peptide); ~15 h (ternary complex)
Concentration	Serum IGF-I levels in adults range from approx. 120-350 μg/L (16-45 nmol/L), with levels decreasing with increasing age. Levels are low at birth, and increase throughout childhood, peaking in mid-puberty at values 2- to 3-times the adult level.
Isolation Method	The most successful natural source for IGF-I is human plasma (Cohn Fraction IV). Typical purification involves cation-exchange chromatography, preparative isoelectric focusing, and reverse-phase HPLC.
Amino Acid Sequence	IGF-I is most closely related to IGF-II, with which it shares about 70% sequence similarity in the A and B domains. A naturally-occurring variant, in which residues 1-3 are deleted, arises post-translationally; the truncated tripeptide (GPE) may also have biological activity in the brain.
Disulfides/SH-Groups	C^{6}-C^{48}, C^{18}-C^{61} (joining B and A domains); C^{47}-C^{52} (within the A domain).
General References	Baxter, R.C., Gluckman, P.D. and Rosenfeld, R.G. (eds.), The Insulin-like Growth Factors and their Regulatory Proteins. *Excerpta Medica*, Amsterdam 1994. Humbel, R.E. Insulin-like growth factors I and II. *Eur. J. Biochem.* 1990, **190**:445-462. Sato, A., Nishimura, S., Ohkubo, T. et al. 1H-NMR assignment and secondary structure of human insulin-like growth factor-I (IGF-I) in solution. *J. Biochem.* 1992, **111**:529-536. Guler, H.P., Zapf, J., Schmid, C. and Froesch, E.R. Insulin-like growth factors I and II in healthy man. Estimation of half-lives and production rates. *Acta Endocrinol.* 1989, **121**:753-758.
Ref. for DNA/AA Sequences	IGF-IA: SwissProt locus IGFA_HUMAN, accession P01343. GenBank accession M14156. IGF-IB: SwissProt locus IGFB_HUMAN, accession P05019. GenBank accession M11568. Rinderknecht, E. and Humbel, R.E. *J. Biol. Chem.* 1978, **253**:2769-2776. Jansen, M., Van Schaik, F.M.A., Ricker, A.T. et al. *Nature* 1983, **306**:609-611. De Pagter-Holthuizen, P., Van Schaik, F.M.A., Verduijn, G.M. et al. *FEBS Lett.* 1986, **195**:179-184.

Insulin-like growth factor-II

Robert C. Baxter and Janet L. Martin

Synonyms	Rat IGF-II was originally known as multiplication-stimulating activity.
Abbreviations	IGF-II (rat IGF-II : MSA)
Classifications	Insulin/IGF family
Description	A member of a peptide family including IGF-I, IGF-II, and proinsulin, IGF-II is an anabolic and mitogenic peptide secreted by many cells of the body. Its activity is modulated by a family of at least six binding proteins. A variety of IGF-II mRNA species have been described, resulting from both alternative splicing and alternative polyadenylation. IGF-II is a single chain polypeptide of 67 aa. Its structure resembles that of proinsulin, with the N-terminal B domain (IGF-II[1-28]) joined to the A domain [41-61] by a connecting C-domain which remains in the mature peptide [29-40]. Unlike proinsulin, IGF-II has a C-terminal D domain of 6 residues [62-67]. Naturally-occurring variants, in which the tetrapeptide RLPG occurs in place of S^{29}, or the tripeptide CGD in place of S^{33}, have been described. In proIGF-II the peptide is extended C-terminally by 89 residues.
Structure	Solution structure by NMR indicates one helical segment in the B-domain (A^{8}-C^{18}) and two in the A-domain (G^{42}-C^{48}, L^{54}-C^{61}); beta-turn in the region G^{19}-G^{22}.
Molecular Weight	7471 (calculated from primary sequence)
Sedimentation Coeff.	Unknown
Isoelectric Point	6.2
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	A family of six IGF-binding proteins (IGFBPs) has been described, which inhibit the biological actions of IGF-II in many systems, apparently by blocking IGF-II binding to cell-surface receptors.
Biological Functions	IGF-II has both mitogenic and anabolic activities, and stimulates differentiation of some cell types. Although it may act in an endocrine manner, transported through the bloodstream in association with a family of IGF binding proteins, its best defined functions are of an autocrine or paracrine nature, acting on cells near the site of its production. It binds with high affinity to two cell-surface receptors, the heterotetrameric type I receptor, which also binds IGF-I and is believed to mediate most IGF-II functions, and the single-chain type II receptor which also binds glycoproteins containing mannose 6-phosphate. The role of this receptor in IGF-II signalling is still controversial.

Physiology/Pathology	IGF-II circulates at 2-3 times the concentration of IGF-I in the adult human circulation, but under relatively weak growth hormone (GH) regulation, probably conferred by its binding to the GH-dependent protein, IGF binding protein-3. IGF-II has been implicated in fetal rather than postnatal growth, and experimental deletion of the IGF-II gene leads to severe fetal growth restriction. IGF-II is overexpressed by some mesenchymal and other tumors leading to hypoglycemia, and by a variety of other tumor types.
Degradation	Sites and mechanism of degradation in man not documented.
Genetics/Abnormalities	The human IGF-II gene spans approx. 30 kb on the 11p15 region of the short arm of chromosome 11, near the insulin gene. It consist of 9 exons, of which exons 7, 8, and part of 9 encode pro-IGF-II. The human IGF-II gene is maternally imprinted, i.e. only the paternal allele is expressed. Bi-allelic expression (loss of imprinting) is thought to be involved in many Wilms' and other tumors.
Half-Life	10-12 min (free); ~15 h (ternary complex)
Concentration	Adult human serum: Approx. 50-100 nmol/L (375-750 µg/L), lower in neonates. No metabolic or diurnal variation.
Isolation Method	Typically by ion-exchange and reverse-phase HPLC. The preferred human source is Cohn Fraction IV-I of human plasma.
Amino Acid Sequence	IGF-II is most closely related to IGF-I, with which it shares about 70% sequence similarity in the A and B domains.
Disulfides/SH-Groups	C^{9}-C^{47}, C^{21}-C^{60} (joining B and A domains); C^{46}-C^{51} (within the A domain).
General References	Humbel, R.E. Insulin-like growth factors I and II. *Eur. J. Biochem.* 1990, **190**:445-462. Terasawa, H., Kohda, D., Hatanaka, H. et al. Solution structure of insulin-like growth factor II; recognition sites for receptors and binding proteins. *EMBO J.* 1994, **13**:5590-5597. Baxter, R.C., Gluckman, P.D., and Rosenfeld, R.G. (eds.), The Insulin-like Growth Factors and their Regulatory Proteins. *Excerpta Medica*, Amsterdam 1994. Jones, J.I. and Clemmons, D.R. Insulin-like growth factors and their binding proteins: Biological actions. *Endocr. Rev.* 1995, **16**:3-34.
Ref. for DNA/AA Sequences	SwissProt locus IGF2_HUMAN, accession P01344. GenBank accession M14118. Rinderknecht, E. and Humbel, R.E. *FEBS Lett.* 1878, **89**:283-286. De Pagter-Holthuizen, P., Van Schaik, F.M.A., Verduijn, G.M. et al. *FEBS Lett.* 1986, **195**:179-184.

Inter-alpha-inhibitor

Wolfgang Gebhard

Synonyms	Inter-alpha-trypsin inhibitor
Abbreviations	IαI; IaI; ITI; IATI
Classifications	Electrical mobility α_1 - α_2
Description	A circulating plasma glycoprotein complex composed of the protein subunits bikunin, HCl and HC2. The small inhibitor subunit or light chain bikunin (formerly HI30) is an acid-stable double-headed Kunitz-type proteinase inhibitor with oligosaccharides at Ser-10 and Asn-45. Its 16 kDa polypeptide sequence is identical to that of urinary trypsin inhibitor (UTI, EDC1) and acid stable serum trypsin (or proteinase) inhibitor (STI, ASTI or ASPI). The inhibitor subunit is unrelated to the large glycosylated subunits or heavy chains HC1 and HC2 which are similar to one another. Inter-α-like inhibitor (IαLI) and pre-α-inhibitor (PαI; electrical mobility: pre α) are IαI-related serum glycoprotein complexes consisting of bikunin/HC2 and bikunin/HC3, respectively. HC3 is related to HC1 and HC2. Biosynthesis of bikunin, HC1, HC2, and HC3 is restricted to the liver and perinatally triggered. A further relative of the heavy chain proteins that does probably not bind to bikunin is the inter-α-trypsin inhibitor family heavy chain-related protein IHRP, also known as the plasma kallikrein-sensitive glycoprotein PK-120.
Structure	Not yet crystallized. The protein-glycosaminoglycan-protein cross-links between bikunin and HC2 (in IαLI) and bikunin and HC3 (in PαI) are identical and probably a characteristic of the bikunin proteins. The cross-links are mediated by a chondroitin-4-sulfate chain that originates from a typical O-glycosidic link to Ser-10 of bikunin. The C-terminal Asp-648 residue of HC2 (Asp-618 residue of HC3) is esterified via the alpha-carbon to C-6 of an internal N-acetylgalactosamine of the chondroitin-4-sulfate chain. For IαLI a glycosaminoglycan link is discussed between two HC2 chains each linked to a bikunin chain by a chondroitin sulfate bridge.
Molecular Weight	180,000 - 240,000 (SDS-PAGE); bikunin: 45,000 (SDS-PAGE), 60-70,000 (gel filtration); large subunits: 78,000 - 86,000 and 85,000 - 96,000 (SDS-PAGE).
Sedimentation Coeff.	6.4 S
Isoelectric Point	Unknown
Extinction Coeff.	7.1 (280nm, 1%, 1cm)
Enzyme Activity	Unknown
Coenzymes/Cofactors	Unknown
Substrates	Human serine proteinases (K_i-values): trypsin (6.4×10^{-8} M), chymotrypsin (8.0×10^{-7} M), plasmin (1.3×10^{-7} M), neutrophil elastase (1.5×10^{-7} M), neutrophil cathepsin G (5.9×10^{-6} M).
Inhibitors	None

1. Installment 1992 (revised 1997)

Biological Functions	Bikunin binds to proteases less avidly than other plasma protease inhibitors, making its role in the blood unclear. The complex is as active as the inhibitor subunit alone. IαI and related proteins possibly have a general function as stabilizers of hyaluronan-containing pericellular coats through a direct binding of their heavy chains with hyaluronic acid.
Physiology/Pathology	Concentration in blood about 0.5 g/L. Patients with disseminated neoplastic disease excrete bikunin into the urine in amounts of 100 mg - 500 mg/day compared to a normal rate of less than 1 mg/day. Increased serum concentrations of bikunin under various inflammatory conditions, rheumatoid arthritis, uremia and after trauma. Bikunin and HC2 (and their corresponding complexes IαI and IαLI) are discussed as negatively regulated acute phase proteins, HC3 (and its corresponding complex PαI) as a positively regulated acute phase protein. IαLI forms a stable complex with glycoprotein TSG-6, an IL-1-and TNF-inducible arthritis-associated hyaluronan binding protein.
Degradation	Normally eliminated from circulation by the liver, especially the proteinase-proteinase inhibitor complexes. During degradation by plasmin, elastases and other proteases bikunin is released. H2C of Iα seems to be particularly sensitive to elastase, PK-120 to plasma kallikrein.
Genetics/Abnormalities	The genes are located on chromosomes 9q32-q33 (bikunin); 3p211-p212 (H1 and H3); 10P15 (H2). A growing list of IαI polymorphisms provides a new and informative phenotypic marker system that can be used in the field of forensic haematogenetics.
Half-life	Unknown
Concentration	Plasma: 0.5 g/L
Isolation Method	Isolated from citrated plasma by ammonium sulfate or ethylene glycol 8000 precipitation, anion exchange chromatography and Blue-Sepharose, and gel filtration.
Amino Acid Sequence	Determined by amino-acid sequencing as well as deduced from cDNA sequences. All three mature subunits are proteolytically processed from significantly larger precursor proteins. Bikunin: released - together with α_1-microglobulin - as the C-terminal 147 residues from a common precursor of 352 residues. Amino acid sequences (positions 33-39) around the reactive centres of the inhibitor domains of bikunin: N-terminal domain: GPCMGMT, C-terminal domain: GPCRAFI.
Disulfides/SH-Groups	Characteristic Kunitz-type disulfide backbone in each bikunin domain (3 disulfides). No interchain disulfides. Intrachain disulfides (2) might exist in both large subunits.
General References	Gebhard, W. et al. Structure of Inter-α-Inhibitor (Inter-α-trypsin inhibitor) and pre-α-inhibitor: current state and proposition of a new terminology. *Biol. Chem. Hoppe-Seyler* 1990, **371**:13-22. Salier, J.-P. Inter-α-trypsin-inhibitor: emergence of a family within the Kunitz-type protease inhibitor superfamily. *Trends Biochem. Sci.* 1990, **15**: 435-439.
Ref. for DNA/AA Sequences	Bikunin cDNA: Kaumeyer, J.F. et al. *Nucleic Acids Res.* 1986, **14**:7839-7849. GenBank, accession number: X04494.

Bikunin genomic DNA: Vetr, H. and Gebhard, W. *Biol. Chem. Hoppe-Seyler* 1990, **371**:1185-1196. EMBL Data Library, accession numbers X54816, X54817, X54818.
Diarra-Mehrpour, M. et al. *Eur. J. Biochem.* 1990, **191**:131-139. GENBANK/M27410 M27411, M27412, M27413, M27414, M27415, M27416, M27417.
Rouet, P. et al. *J. Biol. Chem.* 1992, **267**:20765-20773.
GENBANK/X67082, M76532, M76533, M76534, M76535, D12491, D12492, D12493, M96014, M96015.
Gebhard, W. et al Large subunits, cDNA: *FEBS Lett.* 1988, **229**:63-67. EMBL Data Library, accession number X07173.
Gebhard, W. et al. *Eur. J. Biochem.* 1989, **181**:571-576. EMBL Data Library, accession number X16260.
Diarra-Mehrpour, M. et al. *Biochim. Biophys. Acta* 1992, **1132**:114-118.
GENBANK/X63652,/M76728, M86808, Z11902, M80530, M80531, M80532, M80533, M91429, M91430.
Bost, F. et al. Large subunits, genomic DNA. *Eur. J. Biochem.* 1993, **218**: 283-291. GENBANK/X69532, X69533, X69534, X69535, X69536, X69537, X69538, X69539, X69540, X69541, X69542, X69543, X69544, X69545, X69546, X69547.
Bourguignon, J. et al. PαI cDNA, large subunit H3. *Eur. J. Biochem.* 1993, **212**:771-776. GENBANK/X67055, X70793, X70999, X71000, X57950, L09679, L09680, L09681, L09682, L09683.
Diarra-Mehrpour, M. et al. *Biochim. Biophys Acta* 1994, **1219**:551-554. GENBANK/X75318.
Saguchi, K. et al. IHRP/PK-120 cDNA. *J. Biochem.* 1995, **117**:14-18. GENBANK/D38595.
Nishimura, H. et al. *FEBS Lett.* 1995, **357**:207-211. GENBANK/D38535.

Interferon alpha

Denis O'Shaughnessy and Erich Hochuli

Synonyms	Leukocyte interferon alpha-2, leukocyte interferon A
Abbreviations	IFN alfa-2a, IFN α-2a, IFN-αA
Classifications	alpha or leukocyte interferon
Description	The alpha interferons are a family of proteins characterized by their potent ability to confer a virus-resistant state on their target cells and are classed as being members of the cytokine family. At leas twenty four subtypes of alpha interferon have been sequenced. Most of these proteins are found to be comprised of 165 to 166 aa, with four conserved cyteine residues forming disulfide bridges beween residues 1 and 98 or 99, and between 29 and 138 or 139. There are five commercially available preparations of interferon for clinical use. Two are produced by viral stimulation of white cells: Human lymphoblastoid interferon (IFN alfa-n1, Wellcome) and natural leukocyte interferon (Cantell, Finish Red Cross Blood Center). The other three are genetically engineered recombinant proteins: rIFN alfa-2a (Roferon, A. Hoffmann-La Roche), rIFN alfa-2b (Intron, A. Schering) and rIFN alfa-2c (Boehringer). The lymphoblastoid and natural leukocyte preparations are not so well characterized and contain many, if not all, of the alpha interferon subtypes. Recombinant IFN alfa-2a is identical to the human interferon alpha-2 isolated from induced lymphocytes.
Structure	Interferon alfa-2a is a monomeric, single chain non-glycosylated protein with 165 aa and two disulfide bridges. The X-ray structure is not yet resolved, but the predicted secondary structure is one comprising a large α-helical component and essentially no β-sheet.
Molecular Weight	IFN alfa-2a: 19,600 Da (SDS-PAGE); 19,238.6 Da (electrospray mass spectrometry).
Sedimentation Coeff.	Unknown
Isoelectric Point	IFN alfa-2a: $\simeq 6.3$
Extinction Coeff.	IFN alfa-2a: 1.0 (280 nm, 1%, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	In vitro studies reveal that neutralizing antibodies, which on rare occasions occur naturally and, in some cases, in response to prolonged exogenous treatment, can inhibit the bioactivity of alpha IFN's.
Biological Functions	Alpha IFN's can be induced by a variety of infectious agents from most cells. In particular, viral nucleic acids will induce release of alpha IFN's from macrophages, B cells, and non-B and non-T lymphocytes. IFN alfa-2a (in common with other alpha IFN's) exerts antiviral activity via induction of 2',5' oligoadenylate synthetase and protein kinase, which are thought to be involved in the specific inhibition of viral RNA replication. Alpha IFN's are also known to induce expression of major histocom-

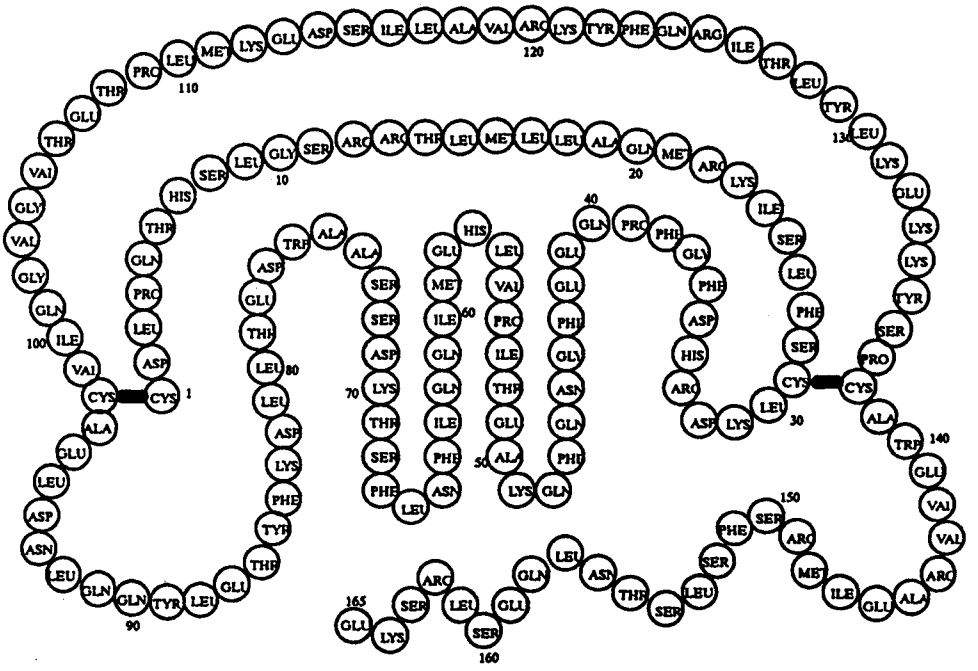
	patibility Class I antigens, augment natural killer cell activity, activate cytotoxic T-cells and enhance expression of tumor associated antigens. These cytokines also possess antiproliferative properties, are involved in cell differentiation and are known to inhibit angiogenesis.
Physiology/Pathology	Alpha IFN's are known to interact with growth factors, oncogenes and other cytokines. Alpha IFN's and other interferons are involved in various stages of the immuno-modulatory cascade, but their exact physiological role has yet to be clarified. The role of alpha IFN's in pathology is also poorly understood, although the therapeutic benefits that result from their administration in viral and oncological disorders suggest that either IFN itself, or other cytokines which are modulated by the IFN's, are involved in these pathologies.
Degradation	Neither the specific catabolic pathways nor degradation products of alpha IFN's are known.
Genetics/Abnormalities	Alpha IFN's are encoded by chromosome 9. There are no known genetic abnormalities which result in either overproduction, or a failure in production, of interferons.
Half-life	5.1 hrs. (3.7–8.5 hrs, i.v. rIFN alfa-2a)
Concentration	Endogenous interferons are produced in small amounts and probably act in a paracrine fashion, without entering the circulation. Although viral infections and exogenous administration can introduce significant amounts of interferons into the circulation, there are no established ranges for circulating levels of endogenous interferons.
Isolation Method	Recombinant IFN alfa-2a is expressed in E. coli as host organism. To engineer the producing organism, the mRNA of interferon alpha-2 was isolated from Sendai virus-infected leukocytes. The mRNA was used to prepare recombinant DNA clones. The rDNA clones were isolated and identified, so that the entire gene for the active protein could be traced. The interferon alfa-2a gene was introduced in the plasmid BR322 and expressed in E. coli. IFN alfa-2a is extracted, refolded and purified by several chromatographic steps.
Amino Acid Sequence	The subtypes of alpha interferon show in excess of 80% homology in their aa sequences. IFN alfa's- 2a, 2b and 2c differ from each other by only 1 or 2 aa at either postions 23 and/or 34. The aa residues at position 23 and 34 are: Lys and His for IFN alfa-2a, Arg and His for IFN alfa-2b and Arg and Arg for IFN alfa-2c respectively. The full aa sequences for these proteins are known.
Disulfides/SH-Groups	Two disulfides (Cys1–Cys98 and Cys29–Cys138); no free SH-groups.
General References	Maeda, S., et al. Construction and identification of bacterial plasmids containing nucleotide sequence for human leukocyte interferon. *Proc. Natl. Acad. Sci. USA* 1981, **77**: 7010–7013. Goeddel, D. V., et al. Human leukocyte interferon produced by E. coli is biologically active. *Nature* 1980, **287**: 411–416. Pestka, S. The human interferons. From protein purification and sequence to cloning and expression in bacteria: Before, between, and beyond. *Arch. Biochem. Biophys.* 1983, **221**: 1–37. Biologic Therapy of Cancer, de Vita, V. T. and Rosenberg, S. A. (eds.) J. B. Lippincott Co. 1991. Hochuli, E. Large scale recovery of interferon alpha-2a synthesized in bacteria. *Chimie* 1990, **40**: 408–412.

Ref. for DNA/AA Sequences

Levy, W. P., et al. Amino Acid Sequence of a human leukocyte interferon. *Proc. Natl. Acad. Sci. USA* 1981, **78:** 6188–6190.

Kontsek, P., et al. Distinct effect of pH 2 on common antigenic structure found in human interferons-α1and -α2 in the region of 30–35. *J. Ifn. Res.* 1991, **11**: 327–332.

Structural Formula of Interferon alfa-2a (Ro 22-8181)



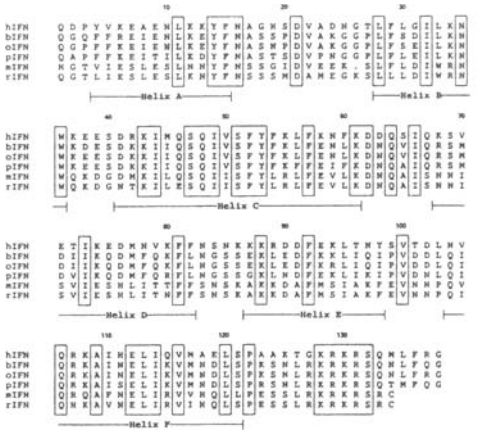
Interferon gamma

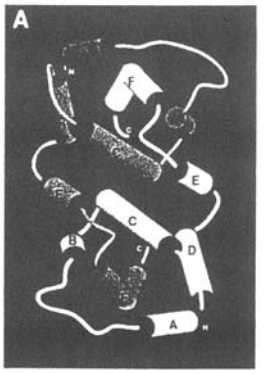
Şefik Ş. Alkan

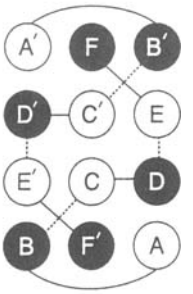
Synonyms	Immune Interferon, Type II Interferon.
Abbreviations	IFN-γ, IFN-g, Type II IFN.
Classifications	Cytokines
Description	It is a glycoprotein produced naturally only by T lymphocytes upon stimulation with antigens or mitogens and also by natural killer cells. It exhibits anti-viral, anti-proliferative and immunomodulatory activities. It is acid (pH 2) labile and sensitive to heat. It shows strict species specificity. Antigenically dissimilar to type I interferons (IFN-alpha and IFN-beta). IFN-gamma is known to synergize with several cytokines. Specific activity: 2×10^7 units/mg protein.
Structure	There are two active species, both glycosylated, 20 kDa and 25 kDa. These forms have identical primary aa sequence but differ in their degree of glycosylation. 25 kDa species glycosylated at both Asn-25 and Asn-97 while the 20 kDa is glycosylated only at Asn-97. Both forms exhibit a great deal of heterogeneity at C-terminus. Discrepancy on the first 3 N-terminal residues is now cleared up. IFN-g lacks Cys-Tyr-Cys residues and has a blocked N-terminus pyroglutamate. Mature IFN-gamma has no Cys residues hence no disulfide bonds. Natural state of IFN-gamma is suggested to be a dimer. rIFN-gamma has been crystallyzed.
Molecular Weight	Natural, glycoylated forms: 20,000 and 25,000. Natural random dimer: 45,000. Recombinant Monomer, non-glycosylated: 17,000 (SDS-PAGE).
Sedimentation Coeff.	Unknown
Isoelectric Point	8.5–8.7
Extinction Coeff.	0.7 (280 nm, 0.1%, 1 cm)
Enzyme Activity	Unknown
Coenzymes/Cofactors	Unknown
Substrates	Unknown
Inhibitors	Unknown
Biological Functions	It is a modulator of the immune system: 1. Essential element in the activation of macrophages (the main macrophage-activating factor, MAF). 2. Powerful inducer of major histocompatibility complexes (MHC) class II antigens and Fc receptors. 3. An important factor in the production of antibodies. Mainly produced by $CD4^+$ T-helper 1 type "inflammatory" cells. It suppresses interleukin-4 induced allergic antibody (IgE) production while enhancing IgG2a production. Important for fighting against viral and parasitic diseases. May play a pivotal role in the onset of inflammatory reactions. IFN-gamma is an inducer of differentiation and activation of myelomonocytic cells. Like type I IFNs it also has anti-viral and anti-proliferative activity.

Physiology/Pathology	It is a pleiotropic cytokine with multiple effects. Lack of IFN-gamma may cause reduced resistance to viral and intracellular parasitic infections. Too high levels may cause autoimmune diseases.
Degradation	It is thought to be eliminated from circulation by the kidneys.
Genetics/Abnormalities	IFN-gamma gene, which exists only as a single copy gene is located on the long arm of chromosome 12 (q24.1). The gene contains 4 exons and 3 introns. Some minor heterogeneity has been reported in the cDNA sequence. recIFN-gamma also undergoes processing of its C-terminus when expressed in E.coli. Abnormal forms have not been described.
Half-life	$\simeq$ 20 min.
Concentration	Not normally found in blood or only at very low levels (1 unit/ml).
Isolation Method	Producer cells: T lymphocytes and large granular lymphocytes (LGL). Stimulators: wide variety of compounds, e.g. antigens, lectins, enterotoxins, antibodies, cytokines, ionophores, phorbol esters. Isolation: chromatography on controlled-pore glass, ultrafiltration and high performance cation exchange chromatography. Alternatively, monoclonal antibody affinity chromatography.
Amino Acid Sequence	Consists of 143 aa. Remarkably high content of basic residues possibly related to its acid lability. Model according to Ealick, S. E., et al. *Science* 1991, **252:** 698 (see below)
Disulfides/SH-Groups	None
General References	Miyata, K., et al. *J. Biochem.* 1986, **99:** 1681–1688. Rubinstein, M., et al. Purification of biologically active human immune interferon. In: US patent No. 4,617,378, Patent date Oct. 14, 1986. Zoon, K. Human interferon: Structure and function. In: *Interferon.* Gresser, I. (ed.) Acad. Press, 1987, **Vol. 9**, pp. 1–12. Dijkmans, P. and Billiau, A. *Current Opinion in Immunology.* 1988, **1:** 269–274. Hochkeppel, H. K. and Alkan, S. S. Immunomodulatory action of interferons. In: *Clinical aspects of Interferons.* Revel, M. (ed). Kluwer Acad. Pub., Boston, 1988, pp. 341–351. Ealick, S. E., et al. *Science* 1991, **252:** 698–702.
Ref. for DNA/AA Sequences	Gray, P. W. and Goeddel, D. V. *Nature* 1982, **298:** 859–863. Derynck, R. More about interferon cloning. In: *Interferon.* Gresser, I. (ed) Acad. Press, 1983, **Vol. 5**, pp. 183–205. De Maeyer, E. and De Maeyer-Guignard. The interferon gene family. In: *Interferons and other regulatory cytokines.* Wiley, 1988, pp. 6–38.

Model according to Ealick et al.







Interferon gamma receptor

Gianni Garotta

Synonyms	Receptor for Immune or Type II Interferon
Abbreviations	IFNγ-R, IFNg-R, Type II IFN-R
Classifications	IFN Type II Receptor
Description	Present on different cell types (from 500 receptors on T lymphocytes to $> 10^4$ receptors on epithelial cells). Only high-affinity binding sites with a calculated Kd of 0.05–0.3 mM. Each IFNγ-R binds at least a dimer IFNγ molecule. The IFNγ-R differs from the receptors of other cytokines. However, the computer analysis of sequence predicts a similar domain size and organization of the extracellular domain of IFNα/β-R, IFNγ-R and other cytokine receptors.
Structure	The IFNγ-R is a N-glycosylated protein. O-linked glycosylation was never detected. The extracellular domain of the human IFNγ-R include the 228 aa between the N-terminus and the transmembrane region with 5 N-glycosylation sites and 8 cysteines forming 4 disulfide bridges. The whole extracellular domain and the S-S bridges need for the full binding capacity. N-linked sugars do not affect IFNγ binding. The residues 229–249 forms the transmembrane region. The residues 250–472 are intracellular. In this domain, a membrane proximal region (residues 256–303) is important for internalization and degradation of the receptor protein and the C-terminal residues 434–472 are required for induction of cellular response. The IFNγ-R protein has a signal peptide of 17 residues.
Molecular Weight	90 kDa (SDS-PAGE). 54 kDa predicted from the aa sequences. The N-linked glycosylation contributes 20 kDa.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	Unknown
Coenzymes/Cofactors	Unknown
Substrates	Unknown
Inhibitors	Inhibitors of IFNγ or IFNγ-R are not known. After virosis, autoantibodies that neutralize IFNγ can be produced. The extracellular domain of IFNγ-R was engineered as recombinant soluble form of the IFNγ-R. It retains the full capacity to bind IFNγ and could be used as inhibitors of IFNγ activity.
Biological Functions	IFNγ exerts antiviral, antiproliferative, immunoregulatory, and proinflammatory activities. These effects are mediated through the cell IFNγ-R. The signal transduction through the human IFNγ-R needs species specific accessory protein(s). This protein seems to interact outside the cell with the complex IFNγ/IFNγ-R. The intracellular domain of the IFNγ-R plays an obligatory role in the signal transduction. After binding, the IFNγ-R is phosphorylated on serine and threonine residues by a specific protein

kinase and rapidly internalized. It may be that IFNγ is translocated to some kind of nuclear membrane of receptors.

Physiology/Pathology

The inhibition of the IFNγ-R may produce less resistance to virus and intracellular parasites or immunodepression. Chronic stimulation of the IFNγ-R may cause autoimmune disease or sensitization for septic shock.

Degradation

After the interaction with IFNγ, the IFNγ-R is internalized and degraded. The intracellular domain shows a cluster of sites for proteases and the trypsin digestion of the isolated IFNγ-R generates a 50 kDa fragment. Up to now, a blood soluble form of the IFNγ-R was never detected. It is likely that the 40 kDa fragment of the IFNγ-R that was identified in the urine is a degradation product of IFNγ-R released by dead cells. The IFNγ-R expressed on living cells is resistant to proteolytic enzymes.

Genetics/Abnormalities

The human IFNγ-R is coded by a gene on the long arm of chromosome 6 (q23–q24). The gene contains 7 exons. The first one comprises the 5′ untranslated region and encodes for the signal peptide. Exons 2 through 5 code for the extracellular domain. The sixth exon codes for the transmembrane region and the seventh for the intracellar domain (Dembic et al., submitted).
The species specific accessory protein(s) involved in the signal transduction of human IFNγ-R is/are encoded by gene(s) that spans 800 kb from the gene(s) for the receptor of human type I IFNs (IFNα and IFNβ). The gene(s) for the IFNα-R is/are in the long arm of chromosome 21 (q21-qter).

Half-life

rIFNγ-R (35 kDa) 1–3 hrs. (in mouse, blood)

Concentration

Up to now, a natural form of soluble IFNγ-R was not found in blood or physiological fluids.

Isolation Method

Extracts from human cell lines or placenta in PBS with proteases inhibitors and 1% Triton X-100 then centrifuged 100,000 × g. The supernatant on IFNγ-Affigel column. The eluate on reversed phase liquid chromatography developed with a linear gradient of acetonitrile.

Amino Acid Sequence

The human IFNγ-R protein consists of 472 aa. (According to Aguet M. et al. (1988) *Cell* **55**:273–280].S1S171Extracellular Domain

```
S1                S17    1 Extracellular Domain
MALLFLLPLVMQGVSRA    EMGTADLGPSSVPTPTNVTI  ESYNMNPIVYWEYQIMPQVP   40

VFTVEVKNYGVKNSEWIDAC INISHHYCNISDHVGDPSNS  LWVRVKARVGQKESAYAKSE  100

EFAVCRDGKIGPPKLDIRKE EKQIMIDIFHPSVFVNGDEQ  EVDYDPETTCIYRVYNVYVR  160

MNGSEIQYKILTQKEDDCDE IQCQLAIPVSSLNSQYCVSA  EGVLHVWGVTTEKSKEVClT  220
                  Transmembrane Region      Intracellular Domain
IFNSSIKGS LWIPVvAALLLFLVLSLVFICF YIKKINPLKEKSIILPKSLISVVRSATLE  280

TKPESKYVSLITSYQPFSLE KEVVCEEPLSPATVPGMHTE  DNPGKVEHTEELSSITEVVT  340

TEENIPDVVPGSHLTPIERE SSSPLSSNQSEPGSIALNSY  HSRNCSESDHSRNGFDTDSS  400

CLESHSSLSDSEFPPNNKGE IKTEGQELITVIKAPTSFGY  DKPHVLVDLLVDDSGKESLI  460

GYRPTEDSKEFS  472
```

Disulfides/SH-Groups	12 cysteines at least 4 S-S bridges in the extracellular domain.
General References	Aguet, M. and Merlin, G. *J. Exp. Med.* 1987, **165:** 988–999. Basu, M., et al. *Proc. Natl. Acad. Sci. USA* 1988, **85:** 6282–6286. Bazan, J. F. *Proc. Natl. Sci. USA* 1990, **87:** 6934–6938. Calderon, J., et al. *Proc. Natl. Acad. Sci. USA* 1988, **85:** 4837–4841. Farrar, M. A., et al. *J. Biol. Chem.* 1991 **266:** 19626–19635. Fountoulakis, M., et al. *J. Immunol.* 1989, **143:** 3266–3276. Fountoulakis, M., et al. *J. Biol. Chem.* 1990, **265:** 13268–13275. Fountoulakis, M., et al. *J. Biol. Chem.* 1990, **265:** 19758–19767. Fountoulakis, M., et al. *Eur. J. Biochem.* 1991, **198:** 441–450. Fountoulakis, M., et al. *J. Biol. Chem.* 1991, **266:** 14970–14977. Garotta, G., et al. *J. Biol. Chem.* 1990, **265:** 6908–6915. Khurana Hershey, G. K. *J. Biol. Chem.* 1990, **265:** 17868–17875. Landolfo, S. and Garotta, G. *J. Immunol. Res.* 1991, **3:** 81–94. Langer, J. A. and Pestka, S. *Immunol. Today* 1988, **9:** 393–400. van Loon, A. P. G. M., et al. *J. Leukoc. Biol.* 1991, **49:** 462–473. Sheehan, K. C. F., et al. *J. Immunol.* 1988, **140:** 4231–4237.
Ref. for DNA/AA Sequences	Aguet, M., et al. *Cell* 1988, **55:** 273–280. Cofano, F. et al. *J. Biol. Chem.* 1990, **265:** 4064–4071. Gray, P. W., et al. *Proc. Natl. Acad. Sci. USA* 1989, **86:** 8497–8501. Hemmi, S., et al. *Proc. Natl. Acad. Sci. USA* 1989, **86:** 9901–9905. Jung, V., et al. *Proc. Natl. Acad. Sci. USA* 1987, **84:** 4151–4155. Kumar, C. S., et al. *J. Biol. Chem.* 1989, **264:** 17939–17946. Langer, J. A., et al. *Somat. Cell. Molec. Genet.* 1990, **16:** 231–240. Munro, S. and Maniatis, T. *Proc. Natl. Acad. Sci. USA* 1989, **86:** 9248–9252. Rashidbaigi et al. *J. Biol. Chem.* 1985, **260:** 8514–8519. Rashidbaigi et al. *Proc. Natl. Acad. Sci. USA* 1986, **83:** 384–388.

Interleukin 1

Teresa Krakauer and Joost J. Oppenheim

Synonyms	Lymphocyte-activating factor (LAF); Endogenous pyrogen (EP); Leukocyte endogenous mediator (LEM); Catabolin; B cell-activating factor (BAF); Mononuclear cell factor (MCF); Osteoclast-activating factor (OAF); Hemopoietin-1
Abbreviations	IL-1
Classifications	None
Description	IL-1 is a pleiotropic mediator of the host response to infection and injury. Two distinct forms of IL-1 encoded by two different genes have been identified, IL-1α and IL-1β. IL-1 is produced by most nucleated cells; mononuclear phagocytes are a major cell source. Regulation of IL-1 production is primarily at the transcriptional level, upon induction by antigens, mitogens, toxins, inflammatory agents, complement, and clotting components; as well as other cytokines, such as tumor necrosis factor α (TNFα) and IL-1 itself. IL-1 acts locally to enhance many cell activities in an autocrine/paracrine manner and systematically to promote hematopoiesis, to amplify immunological reactions, to induce fever, hypoglycemia, hypotension, and to stimulate the hepatic acute-phase response.
Structure	Crystallography indicates that IL-1β is a tetrahedron with the triangular faces formed by 12 β-strands held together by hydrogen bonds. The hydrophobic core consists of side chains contributed from multiple positions along its primary sequence. The folding of IL-1α is probably very similar to that of IL-1β, as they have the same Stokes radius (1.738 for IL-1α, 1.711 for IL-1β), and a similar frictional ratio (1.12 for IL-1α and 1.09 for IL-1β). Both N- and C-termini are exposed and participate in a tertiary structure which binds to two distinct types of cell surface receptors.
Molecular Weight	17,500 for both IL-1α and IL-1β (SDS-PAGE).
Sedimentation Coeff.	IL-1α: 2.17 S; IL-1β: 2.30 S
Isoelectric Point	IL-1α: 5.4; IL-1β: 6.8
Extinction Coeff.	IL-1α: 20.45; IL-1β: 10.61 (280nm, 1mM, 1cm).
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	A variety of naturally occuring inhibitors of IL-1 are present in serum, urine, and conditioned media. IL-1 is the only cytokine for which a natural endogenous inhibitor with specificity exists, IL-1 receptor antagonist (IL-1ra), to counteract its biological effects. IL-1ra is structurally similar to IL-1 and inhibits the action of IL-1 by competitively binding to IL-1 receptors without transducing a signal. In addition, type II IL-1 rceptors which lack the intracellular signal transducing domain, act as "decoy" receptors for IL-1 and reduce the level of IL-1. Shed soluble extracellular portions of the type I and type II receptors for IL-1 also act as specific binding inhibitors

	of IL-1. Prostaglandin E suppresses the synthesis and release of IL-1. Corticosteroids suppress expression of IL-1 mRNA. Corticotropin-releasing hormone inhibits IL-1-induced anorexia, sleep and prostaglandin synthesis. Transforming growth factor β, as well as interleukin 4 (IL-4), IL-10 and IL-13, suppress IL-1 production. β-melanocyte-stimulating hormone inhibits the synthesis and several activities of IL-1.
Biological Functions	Both IL-1α and IL-1β exhibit identical biological activities, and act at low concentrations (nM to pM). IL-1 augments immunologically mediated responses to antigens, and is a key mediator of fever, the acute-phase response, and hematopoiesis. The broad spectrum of activities of IL-1 overlaps considerably with those of TNFα and IL-6. Many of the IL-1 effects are mediated by enhancing the production of cytokines, such as TNFα, colony-stimulating factors (CSF), IL-1 itself, IL-2, IL-6, IL-8, and homologous members of the chemokine family. In addition, the effects of IL-1 are amplified through the ability of IL-1 to upregulate receptor expression for itself, for the IL-2 receptor α chain, and for the receptors for IFNγ, IL-3 and GM-CSF. IL-1 also synergizes with most of these cytokines in promoting growth and differentiation of many cell types.
Physiology/Pathology	IL-1 is not produced by cells under normal conditions. Perturbation by infection and injury induces IL-1. IL-1 has been implicated in vivo in the pathogenesis of a variety of acute and chronic inflammatory diseases such as septic shock, rheumatoid arthritis, and atherosclerosis. Synovial fluids contain IL-1, which acts as a potent stimulator of cartilage resorption. Synovial tissue macrophages also produce high levels of IL-1β followed by IL-1ra. Endogenously produced IL-1ra levels are normally a later response that limits the proinflammatory, pathogenic effects of IL-1.
Degradation	Subsequent to binding to cell surface receptors, IL-1 is internalized and found in part, in the nuclear compartment. Low molecular weight products appear extracellularly, presumably as products of degradation.
Genetics/Abnormalities	All three genes (IL-1α, IL-1β, and IL-1ra) are located on the long arm of chromosome 2, band q13-21. Homology of the IL-1α and IL-1β gene sequences is 45%.
Half-life	Both IL-1α and IL-1β have a short half-life of 10 min. in serum.
Concentration	Plasma IL-1α and IL-1β levels are undetectable in healthy volunteers (<40 ng/L). In patients with sepsis, plasma IL-1β is elevated to 120-260 ng/L but IL-1α remains undetectable. Normal serum IL-1ra is 200-300 ng/L but in patients with endotoxemia, IL-1ra rises to 6000-200,000 ng/L.
Isolation Method	IL-1α and IL-1β are produced by stimulating human monocytes with lipopolysaccharide in vitro, yielding culture supernatants with a low concentration of IL-1. Purification can be achieved by hydrophobic chromatography (phenyl-Sepharose), size fractionation (Superose 12 FPLC or AcA54), and isoelectrofocusing to separate IL-1α and IL-1β. IL-1ra is produced by stimulating human monocytes on adherent IgG and purified by using a combination of anion exchange, gel filtration and reverse-phase HPLC. However, expression of all three molecules in E. coli by using cloned IL-1 cDNA from a human monocyte library is the method of choice for large-scale preparation.
Amino Acid Sequence	There is only 26% homology in the aa sequences of IL-1α and IL-1β. Sequence homology of IL-1ra to IL-1α and IL-1β is 18% and 26%,

respectively. Both IL-1α and IL-1β molecules are initially synthesized as 31 kDa precursors (271 aa for IL-1α, 269 aa for IL-1β) and lack a classical hydrophobic signal sequence. An intracellular cysteine protease, named IL-1β converting enzyme, specifically cleaves the inactive pro-IL-1β (31K) between Asn-116 and Ala-117 to generate the mature, biologically acitve IL-1β (17.5 kDa, 153 aa). Pro-IL-1α is cleaved by calpain and other extracellular nonspecific proteases between Arg-112 and Ser-113 to the mature form (17.5 kDa, 159 aa). Neither IL-1α nor IL-1β is glycosylated. The precursor protein of IL-1ra is a single, glycosylated peptide of 177 aa with a 25 aa leader sequence. IL-1ra is secreted extracellularly by the classical secretory pathway and three structural variants of 18-22 kDa exist depending on the extent of glycosylation.

Disulfides/SH-Groups

None

General References

Arend, W.P. *J. Clin. Invest.* 1991, **88**:1445-1451.
Dinarello, C.A. *Blood* 1996, **87**:2095-2147.
Oppenheim, J.J. et al. *Immunol. Today* 1986, **7**:45-56.

Ref. for DNA/AA Sequences

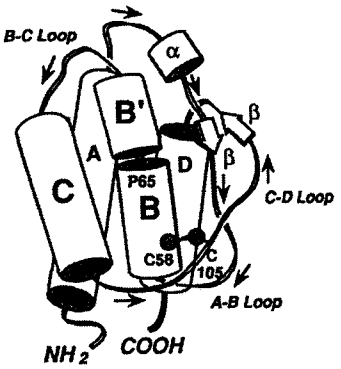
Clark, B.D. et al. *Nucleic Acid Res.* 1986, **14**:7897-7914.
Eisenberg, S.P. *Nature* 1990, **343**:341-346.
March, C.J. et al. *Nature* 1985, **315**:641-646.

Interleukin-2

Thomas L. Ciardelli

Synonyms	T-Cell Growth Factor
Abbreviations	IL-2, TCGF
Classification	No standard classification
Description	A variably glycosylated serum protein secreted primarily by a subset of antigen activated T-lymphocytes. IL-2 is comprised of a single polypeptide chain of 133 aa glycosylated near the N-terminal region on Thr at position three. There are no common subtypes.
Molecular Weight	14,000–17,500 (SDS-PAGE, 15% gel): Normal IL-2 is secreted primarily as three species. 15,420: recombinant IL-2 (from the cDNA).
Sedimentation Coeff.	1.8 : recombinant IL-2
Isoelectric Point	6.6, 7.2, 7.9 (IEF gives three bands), 7.7: recombinant IL-2
Extinction Coeff.	$9.58 \times 103\ M^{-1}\ cm^{-1}$ (280 nm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Other than anti-IL-2 and anti-IL-2 receptor antibodies, no known biological compounds directly inhibit bioactivity by interrupting the interaction of IL-2 with its receptors. Some IL-2 mutant analogs display partial antagonist activity.
Biological Function	Interleukin-2 serves as a growth factor for antigen activated T-lymphocytes and some B-lymphocytes and non-activated Natural Killer cells. IL-2 also potentiates the cytotoxic activity of Natural Killer cells. Recombinant IL-2 exhibits $\sim 1.8 \times 10^7$ IU/mg activity.
Physiology/Pathology	Interleukin-2 is a multifunctional lymphocytotrophic hormone that participates in the manifestation of an antigen specific immune response. IL-2 mediates the clonal expansion, activation, and/or differentiation of selected immune cell types. Some immune deficiencies associated with T-cell dysfunction respond favorably to exogenous IL-2 administration. Experiments with genetically deficient ("IL-2 knockout") mice indicate only minor immunodeficiency and suggest that loss of the ability to produce IL-2 can be compensated for, developmentally. Severe X-linked combined immunodeficiency has been associated with mutations in the IL-2 receptor gamma subunit, a receptor protein shared with several other lymphokine receptors.
Degradation	Metabolized and eliminated by the kidney.
Genetics/Abnormalities	Gene location: chromosome band 4q26, no common abnormalities.
Half-life	After I. V. admin.: $\alpha\ t_{1/2} = 13$ min.; $\beta\ t_{1/2} = 1.5–4$ hrs.

Concentration	Not measurable in systemic circulation under normal physiologic conditions.
Isolation Method:	1) From supernatants of stimulated normal peripheral blood lymphocytes or the JURKAT cell line: sequential column chromatography on SP-Sephadex C-25, DEAE-sephacel, Ultragel AcA 54 and DEAE-cellulose (DE-52) followed by C-18 reverse phase HPLC. Purification monitored by bioassay or IL-2 ELISA. 2) Immunoaffinity purification over an IL-2 antibody column (commercially available).
Amino Acid Sequence:	APTSSSTKKT QLQLEHLLLD LQMILNGINN YKNPKLTRML TFKFYMPKKA TELKHLQCLE EELKPLEEVL NLAQSKNFHL RPRDLISNIN VIVLELKGSE TTFMCEYADE TATIVEFLNR WITFCQSIIS TLT Bold indicates probable receptor contact residues. No sequence homologies with any other proteins. Several cytokines retain a **D** residue at or near position 20.
Disulfides/SH-Groups	One intrachain disulfide between Cys-58 and Cys-105. One free SH and Cys 125 (can be replaced by Ala without loss of function).
General References	Smith, K. A., *Science* 1988, **240**:1169–1176. *Interleukin-2*. Smith, K. A. (ed.), Academic Press, San Diego 1988. *Interleukin-2*. Waxman, J. and Balkwill, F. (eds.) Blackwell Scientific Publications, Oxford 1992. Bazan, J. F., *Science* 1992, **257**: 410–412.
Ref. for DNA/AA Sequence	Taniguchi, T. et al., *Nature* 1983, **302**: 305–308.



IL-2 (model)

From Bazan, J. F. *Science* 1992, **257**:410–412

Interleukin 3

James N. Ihle

Synonyms	Multi-colony stimulating factor, P-cell stimulating factor, Thy-1 inducing factor, Mast cell growth factor, WEHI-3 Growth Factor, CSF-2α, CSF-2β, Burst Promoting Activity (BPA), Histamine-producing cell-stimulatory factor (HCSF), Hemopoeitin 2, Megakaryocyte colony stimulating factor (Mk-CSF)
Abbreviations	IL-3
Classifications	None
Description	IL-3 is primarily produced by activated T-cells and is a growth factor for a early hematopoietic progenitor cells as well as cells committed to a variety of the myeloid lineages.
Structure	Unknown
Molecular Weight	15–30 kDa (SDS-PAGE). Heterogenity is due to variable extents of glycosylation of the secreted protein.
Sedimentation Coeff.	Unknown
Isoelectric Point	5.0–8.0 (heterogenous)
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	IL-3 is growth factor that can support the proliferation and continued differention of hematopoietic cells including early hematopoietic multi-lineage progenitors and cells committed to differentiation to mast cells, eosinophiles, granulocytes, macrophages,megakaryocytes and erythroid lineage progenitors.
Physiology/Pathology	The normal function of IL-3 is hypothesized to be the increased production of hematopoietic lineage cells during physiological conditions resulting in the activation of T-cells. Overproduction of IL-3 results in the increased production of cells of a variety of the hematopoietic lineages. Characteristically the overproduction of IL-3 is associated with increased numbers of eosinophiles and mast cells.
Degradation	Unknown
Genetics/Abnormalities	Chromosome 5q23–q31. Physically located 9 kb 3′ of the gene for GM-CSF. Also closely linked genetically to the genes for IL-5 and IL-4.
Half-life	Half-life of minutes in serum.
Concentration	IL-3 is not normally detected in serum or various tissues. The growth factor is thought to function locally at sites of T-cell activation.

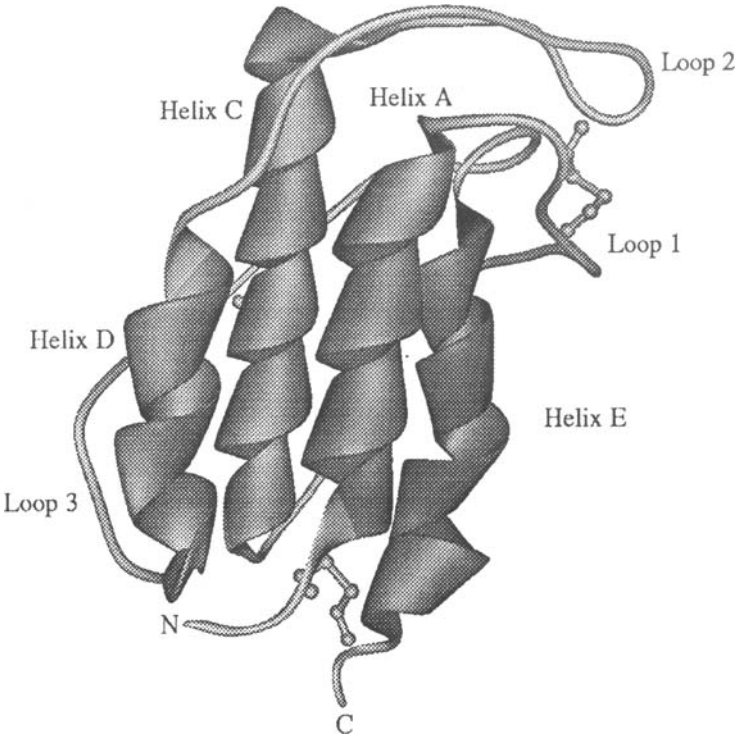
Isolation Method	IL-3 is produced by activated T-cells although the levels are extremely low and make purificaton difficult. Generally IL-3 is purified from conditioned media of cells that have been transfected with a eucaryotic expression vector containing the IL-3 cDNA. Purification procedures involve ammonium sulfate fractionation, DEAE ion exchange chromatography and HPLC purification methods employing C18 columns.
Amino Acid Sequence	The amino acid sequence of IL-3 has deduced from the sequence of cDNA and can be obtained from the reference given for the cDNA.
Disulfides/SH-Groups	IL-3 has an essential intramolecular disulfide bond between cysteine residues 35 and 103.
General References	Lymphokines, Volume 15 Interleukin 3: The Panspecific Hemopoietin. Schrader, J. W. (ed.) Academic Press Inc. 1988 Ihle, J. N. Interleukin-3. In: *Handbook of Experimental Pharmacology.* **Vol. 95/I:** Peptide Growth Factors and Their Receptors I. Sporn M. B. and Roberts, A. B. (eds.), Springer-Verlag 1990; pp. 541–575.
Ref. for DNA/AA Sequences	Yang,Y-C, et al. *Cell* 1986, **47:** 3–10. GenBank Accession Number M14743.

Interleukin 4

Erich Kilchherr and Christoph H. Heusser

Synonyms	B cell stimulatory factor-1, T cell growth factor-2, mast cell growth factor-2, B cell differentiation factor for IgG1
Abbreviations	IL-4, BSF-1, TCGF-2, MCGF-2, BCDF-Υ
Classifications	None
Description	A lymphokine produced by activated T cells and mast cells. cDNA isolated by homology to mouse interleukin-4. The serum concentration is below detection limit. A monomeric protein with 3 intramolecular disulfide bonds and 2 potential glycosylation sites. Glycosylation of natural protein unknown. Probably one site used in recombinant protein expressed in mammalian cells. No difference of specific activity of recombinant protein if protein is unglycosylated (expressed in E. coli) or glycosylated at one or both possible sites.
Structure	Globular protein, high α-helix content. α-helix: 55–75%; β-sheet: 5–12%; β-turn: 21–25%.
Molecular Weight	14,963 (sum of amino acids) for the recombinant protein, unglycosylated. 15,000; 18,000; 19,000 (SDS-PAGE) for the recombinant protein expressed in mammalian cells.
Sedimentation Coeff.	1.5 S
Isoelectric Point	> 10 (IEF); 9.42 (aa composition)
Extinction Coeff.	6.25 (278 nm; 1%; 1 cm); $\varepsilon = 9370\ M^{-1}\ cm^{-1}$ (278 nm) for IL-4 expressed in E. coli.
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Soluble IL-4 receptor: receptor binding inhibitor. Interferon-γ (IFN-γ): inhibits all IL-4 effects on B cells. Interferon-α (IFN-α): inhibits IL-4 induced IgE production but this effect may not be IL-4 specific. Inhibits IL-4 induced soluble CD23 release from purified B cells.
Biological Functions	Multiple functions shown. The most important are: Induction of immunoglobulin class switch from IgM to certain IgG subclasses (human IgG4, mouse IgG1) and/or IgE in presence of T cells. Enhances HLA-DR and HLA-DQ expression on B cells and monocytes. Costimulatory factor for B cell growth with immobilized anti-IgM or anti-CD40 monoclonal antibodies. Induction of CD23 (Fcε-receptor-II) on B cells and monocytes. All effects on B cells inhibited by IFN-γ. Inhibitory effect on LAK (lymphokine activated killer) cell induction by IL-2. Growth factor for thymocytes and peripheral T cells. T cell differentiation factor for development of T cells towards a Th_2 phenotype. Inhibits expression of Fcγ-receptors I, II, III on monocytes, increases expression of CR3 and CR4 on monocytes. Mast cell growth factor.

Physiology/Pathology	IL-4 may be overexpressed in parasitic and presumably in allergic responses. In mice monoclonal anti-IL-4 strongly suppressed an in vivo IgE response induced e.g. by parasites. The IgG1 production was not affected. IL-4 deficient transgenic mice are specifically incapable to mount an IgE response, but show normal B and T cell development. This indicates that IL-4 is required for IgE production and that other functions may be secured by a rescue pathway. Anti-IL-4 antibodies administered along with a lethal dose of Leishmania major into susceptible Balb/c mice led to complete resolution of the disease and protective immunity. Overexpression of murine IL-4 in vivo resulted in a "complex inflammatory reaction resembling that observed in human allergic diseases". IL-4 may exhibit a strong anti-tumor activity. In in vivo studies, IL-4 transformed plamacytoma cells were completely growth inhibited and suppressed completely the growth of other murine carcinoma lines injected concomitantly to the same site.
Degradation	Unknown
Genetics/Abnormalities	Interleukin-4 gene on chromosome 5, band q23-q31, spans approximately 10 kb, with 4 exons and 3 introns. Abnormalities unknown since human protein has not been isolated and characterized from natural sources.
Half-life	rhIL-4 (glyc.) injected in scid mice: 12 min.
Concentration	Below detection limit (10 ng/L) in body fluids.
Isolation Method	Isolated from transformed mammalian cell culture supernatants by cation exchange chromatography and reversed phase HPLC on C4 columns. E. coli expressed interleukin-4 solubilized and denatured by chaotrop (e.g. guanidine. HCl) in presence of glutathione (red./ox.: 10/1), renatured by dilution and dialysis. Further purification by cation exchange and/or gel filtration chromatography.
Amino Acid Sequence	Coding sequence: 153 residues; mature protein: 129 residues. Coding regions 1–90 and 129–149 share 50% homology to mouse interleukin 4 coding sequence at aa level.
Disulfides/SH-Groups	3 disulfides, no free sulfhydryls in mature protein. Disulfides link the following cysteins: C(3)–C(127); C(24)–C(65); C(46)–C(99).
General References	Finkelman, F. D., et al. *Ann. Rev. Immunol.* 1990, **8:** 303–333. Snapper, C. M., et al. *Immunol. Rev.* 1988, **102:** 51–75. Yokota, T., et al. *Immunol. Rev.* 1988, **102:** 137–188. Tepper, R. I., et al. *Cell* 1989, **57:** 503–512. Tepper, R. I., et al. *Cell* 1990, **62:** 457–467. Kühn, R., et al. *Science* 1991, **254:** 707–710.
Ref. for DNA/AA Sequences	Yokota, T., et al. *Proc. Natl. Acad. Sci. USA* 1986, **83:** 5894–5898. Arai, N., et al. *J. Immunol.* 1989, **142:** 274–282.



Ribbon diagram of the complete predicted structure of human IL-4 (from Curtis, B. M., et al. *Proteins: Struct. Funct. Genet.* 1991, **11**: 111–119.
Molecular graphics from UCSF Computer Graphics Laboratory using UCSF **MidasPlus** software. Copyright 1990, Regents, University of California.

Interleukin 5

Colin J. Sanderson

Synonyms	Eosinophil differentiation factor; Eosinophil colony stimulating factor. This molecule was briefly known as interleukin-4. The analogous murine molecule also has activity on B cells, and has a number of synonyms: T cell replacing factor; B cell growth factor II.
Abbreviations	IL-5; EDF; CSF-Eo
Classifications	Haemopoietic growth factor
Description	A glycoprotein produced by T cells (lymphokine or cytokine) in response to antigen or polyclonal stimulation.
Structure	A disulfide linked homodimer of two polypeptides of 115 aa. The molecule is made up of two helical bundles each of three helices (A, B, C,) and a fourth (D) intercalated from the other polypeptide. Each of the bundles has an up-up, down-down configuration similar to monomeric members of structurally related cytokines. There is an N-linked carbohydrate chain on residue Asn-55, and an O-linked on residue Thr-3.
Molecular Weight	40,000 (gel permeation); 26,300 (aa composition).
Sedimentation Coeff.	Unknown
Isoelectric Point	4 - 6
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None known
Biological Functions	Induces eosinophil production from bone marrow cultures in vitro, and at least in the mouse appears to be the main cytokine controlling the production of eosinophils in vivo. Although the murine analogue has well defined activities on B cells, human IL5 has no activity in analogous human assays.
Physiology/Pathology	Appears to be produced as a result of activation of cellular immunity to certain metazoan parasites and allergens. Not detectable in normal individuals. Detectable in the serum of patients with parasite induced eosinophilia, and in the sputum of asmatics with lung eosinophilia.
Degradation	Unknown
Genetics/Abnormalities	The gene is located on chromosome 5q31 in a cluster with the genes encoding other cytokines: interleukin-13, interleukin-4, interlukin-5, interleukin-3, granulocyte-macrophage colony stimulating factor.
Half-life	Unknown

Concentration	Unknown
Isolation Method	The native material has not been characterized, although mRNA is detectable in activated T cells. Recombinant protein has been produced in a number of different expression systems.
Amino Acid Sequence	The aa sequence shows no significant homology with any other human protein, although there is an area of apparent similarity with IL-3, IL-4 and GM-CSF near the C-terminus.
Disulfides/SH-Groups	Two disulfide bonds are present. The cysteine at position 44 of one peptide forms a disulphide with cysteine at position 86 of the other peptide.
General References	Sanderson, C.J. Interleukin-5 and the regulation of eosinophil production. In: *Immunopharmacology of Eosinophils*. Smith, H. and Cook, R.M. (eds.), Academic Press Ltd. London 1993, pp 11-24. Sanderson, C.J. Interleukin-5. In: *The Cytokine Handbook*. Thomson, A.W. (ed.), Edition 2, Academic Press, London 1994, pp 127-143. Sanderson, C.J. Interleukin-5. In *Guidebook to Cytokines and their Receptors*. N. Nicola (ed.), Sambrook and Tooze, Oxford 1994, pp 49-52.
Ref. for DNA/AA Sequences	Azuma, C. et al. Cloning of cDNA for human T-cell replacing factor (interleukin-5) and comparison with the murine homologue. *Nucleic Acids Research* 1986, **14**:9149-9158. Campbell, H.D. et al. Molecular cloning, nucleotide sequence, and expression of the gene encoding human eosinophil differentiation factor (interleukin-5). *Proc. Natl. Acad. Sci. USA* 1987, **84**:6629-6633.

Interleukin 6

Shizuo Akira and Tadamitsu Kishimoto

Synonyms	B-cell stimulatory factor (BSF-2); Interferon β_2 (IFNβ_2); 26kDa protein; Hybridoma/plasmacytoma growth factor (HPGF); Hepatocyte stimulating factor (HSF); Macrophage granulocyte inducing protein 2 (MGI-2); Cytotoxic T cell differentiation factor (CDF)
Abbreviations	IL-6
Classifications	Cytokine
Description	A multi-functional cytokine acting on a wide variety of cells, produced by a wide variety of cells such as T cells, B cells, monocytes/macrophages, fibroblasts, keratinocytes, glia cells, etc. Usually not produced constitutively by normal cells, but induced by viral infections, lipopolysaccharide, a variety of cytokines including IL-1 and TNF. Glucocorticoids inhibit its production.
Structure	Four α-helical bundle structure (predicted from analyses of the crystal structures of growth hormone and IL-2).
Molecular Weight	21 - 28 kDa (mature), approx. 20 kDa (core). Modifications: 2 N-glycosylation sites, several O-glycosylation sites, phosphorylations on serine residues.
Sedimentation Coeff.	Unknown
Isoelectric Point	5.0 - 5.1
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Participates in the immune responses, hemopoiesis and inflammatory responses: Induction of cell proliferation (myeloma/hybridoma/plasmacytoma cells, T cells, EBV-transformed B cells, multipotential hematopoietic progenitor cells, mesangial cells, keratinocytes, Kaposi sarcoma-derived cells); Inhibition of cell proliferation (myeloid leukemic cell lines, breast carcinoma cell lines); Induction of differentiation or specific gene expression (neural differentiation of PC12 cells, cytotoxic T cell differentiation, megakaryocyte maturation, macrophage differentiation of myeloid leukemic cell lines, induction of osteoclast formation, immunoglobulin production in B cells, production of acute phase proteins in hepatocytes).
Physiology/Pathology	Induces antibody production. Induces acute phase reaction. Induces fever. Increases plasma level of ACTH. Abnormal production of IL-6 causes polyclonal B cell activation or autoimmune symptoms in cardiac myxoma, uterine cervical carcinoma, rheumatoid arthritis, Castleman's disease and AIDS; generation of lymphoid malignancies (multiple myeloma, Lennert's

	T cell lymphoma); generation of mesangial proliferative glomerulonephritis; development of AIDS Kaposi sarcoma. IL-6 knockout mice demonstrate a critical in vivo role for IL-6 in stimulation of osteoclast formation and activity, development of IgA antibody production in the lung, induction of acute phase reaction, and defense against bacteria (*Listeria monocytogenes*) and virus (*vesticular stomatitis virus*).
Degradation	^{125}I-rhIL-6 injected intravenously is taken up by the liver within a few minutes, subsequently released, then accumulates in the skin with maximal levels between 4 and 8 h, and is degraded there. The radioactivity appears in urine (from experiments in rats).
Genetics/Abnormalities	Size of mRNA: 1.3kb. Size of gene: approx. 5kb. Number of exons and introns: 5 and 4, respectively. Localization: chromosome 7 (7p21).
Half-life	Biphasic clearance pattern from plasma.
Concentration	Plasma: not detectable approx. minimal (< 100 ng/L); however elevated in some pathologic conditions, ex. 50 to > 500 µg/L in patients with acute bacterial infection.
Isolation Method	Harvest serum-free culture supernatant of IL-6 producing cells, concentrate it by ultrafiltration, run gel-permeation chromatography, chromatofocusing and reverse-phase HPLC. Bioassay: Immunoglobulin production assay or hybridoma growth assay.
Amino Acid Sequence	Significant homology with granulocyte colony stimulating factor (G-CSF).
Disulfides/SH-Groups	2 intrachain disulfides.
General References	Kishimoto, T. and Hirano, T. *Ann. Rev. Immunol.* 1988, 6:485-512. Kishimoto, T. *Blood* 1989, **74**:1-10. Hirano, T. and Kishimoto, T. Interleukin 6. In: *Handbook of Experimental Pharmacology* **Vol 95/I**, Peptide Growth Factors and Their Receptors, Sporn, M.B. (ed.), Springer-Verlag, Berlin 1989; pp. 633-665. Akira, S., Taga, T. and Kishimoto, T. Adv. *Immunol.* 1993, **54**:1-78.
Ref. for DNA/AA Sequences	Hirano, T. et al. *Nature* 1986, **324**:73-76.

Interleukin 8

Alfred Walz and Marco Baggiolini

Synonyms	Neutrophil activating peptide 1, Neutrophil activating factor, monocyte-derived neutrophil chemotactic factor, monocyte-derived neutrophil activating peptide, lymphocyte-derived neutrophil activating peptide, granulocyte chemotactic protein, neutrophil chemotactic factor
Abbreviations	IL-8, NAP-1, NAF, MDNCF, MONAP, LYNAP, GCP, NCF
Classifications	Chemotactic cytokine, proinflammatory cytokine
Description	IL-8 is produced and secreted by monocytes, macrophages and endothelial cells upon stimulation with endotoxin, TNF or IL-1, and by tissue cells (fibroblasts, keratinocytes, various types of epithelial cells, synovial cells, etc.) upon stimulation with TNF or IL-1. IL-8 stimulates primarily neutrophils inducing chemotaxis and release responses, and has, therefore, proinflammatory activity.
Structure	IL-8 consists of a single, non-glycosylated peptide chain. It is produced as a precursor of 99 aa. Its major form consists of 72 aa. Several amino-terminal variants with biological activity have been isolated (77, 70, 69 aa). Endothelial cells secrete predominantly the 77 amino acid form, also called 'endothelial IL-8'. The mature form contains four cysteines and 14 basic aa. Nuclear magnetic resonance spectroscopy shows that in solution IL-8 forms a dimer through six backbone hydrophobic bonds. IL-8 has been crystallized.
Molecular Weight	8,383 (calculated from aa sequence of the 72 aa form).
Sedimentation Coeff.	Unknown
Isoelectric Point	8.0–8.5
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Activates human neutrophils, inducing shape change, chemotaxis in vitro and in vivo, rise of intracellular free calcium, release of azurophil and specific granule content, upregulation of CR1 and CR3 receptors, and respiratory burst. Half-maximal effects on chemotaxis and intracellular calcium rise in human neutrophils were obtained at 0.3 and 3 nM, respectively. Has low chemotactic activity for human basophils, and induces histamine release from basophils treated with IL-3. Human IL-8 acts on mouse, rat, guinea-pig and rabbit neutrophils. Chemotactic activity for lymphocytes has been reported.
Physiology/Pathology	High levels of IL-8 are found in psoriasis scales, synovial fluids of rheumatoid arthritis (RA) patiens, and bronchoalveolar lavage fluids in correlation with high neutrophil counts. Mononuclear cells from the blood

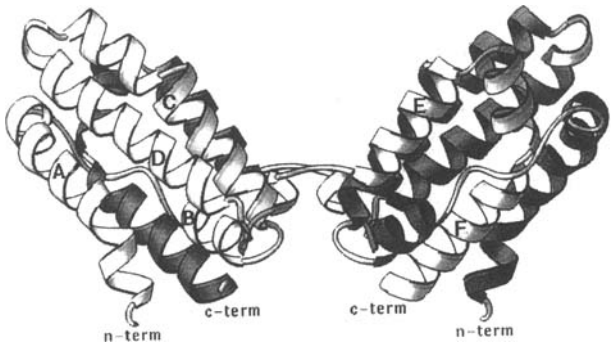
	and synovial fluids of RA patients produce higher amounts of IL-8 than cells from healthy individuals and patients with other arthritic diseases. IL-8 is considered a main mediator of neutrophil recruitment into diseased tissues.
Degradation	IL-8 binding to receptors on neutrophils is internalized and degraded.
Genetics/Abnormalities	Unknown
Half-life	Unknown
Concentration	Plasma: not detectable in healthy individuals (less than 10 ng/L).
Isolation Method	Isolated from supernatants of stimulated mononuclear cells by cation-exchange chromatography on phoshocellulose or Mono-P, hydroxylapatite chromatography, reverse-phase HPLC on C4, C18 and CN-propyl columns.
Amino Acid Sequence	SAKELRCQCIK TYSKPFHP KFIKELRV IESGPHCA NTEIIVK LS-DGRELCLDPKENWVQRVVEKFLKRAENS. The sequence of the 72 aa form is similar to that of NAP-2, gro/MGSA and platelet factor 4 (47%, 43% and 39% identity, respectively). The 77 aa form contains an additional N-terminal pentapeptide, AVLPR. The IL-8 genomic DNA consists of 4 exons and 3 introns. The gene maps to chromosome 4 (q12–21). This locus also contains related genes, such as gro/MGSA, PF4 and gamma-IP 10.
Disulfides/SH-Groups	Two intrachain disulfide-bridges, Cys7-Cys34 and Cys9-Cys50.
General References	Baggiolini, M. et. al. Novel neutrophil-activating peptides and their role in inflammation. In: *Progress in Immunology*, Melchers, F. (ed.) Vol. VII, 1989, pp. 765–771, Springer Verlag, Berlin. Baggiolini, M. et al. *J. Clin. Invest.* 1989, **84:** 1045–1049. Matsushima, K. et al. *Cytokine* 1989, **1:** 2–13. Walz, A. et al. *Biochem. Biophys. Res. Commun.* 1987, **149:** 755–761. Westwick, J. et al. *Immunol. Today* 1989, **10:** 146–147.
Ref. for DNA/AA Sequences	Schmid, J. and Weissmann, C. *J. Immunol.* 1987, **139:** 250–256. Lindley, I. et al. *Proc. Natl. Acad. Sci. USA* 1988, **85:** 9199–9203. Mukaida, N. et al. *J. Immunol.* 1989, **143:** 1366–1371.

Interleukin-10

Kristopher Josephson and Mark R. Walter

Synonyms	Cytokine synthesis inhibitory factor
Abbreviations	IL-10; CSIF
Classifications	
Description	Interleukin 10 (IL-10) is an α-helical cytokine synthesized by T_H2 helper T-cells, B cells, monocoytes, and macrophages. The mature polypeptide consists of 160 aa which associates to form a homodimer. The human IL-10 dimer is not glycosylated.
Structure	The crystal stucture of IL-10 reveals a dimer with overall dimensions of 30 x 40 x 70Å. Each peptide chain consists of six helices (A-F) which tightly associate to form an intertwined dimer. The dimer contains two distinct structural domains which are oriented at right angles to one another. Each domain is composed of helices A, B, C, and D from one chain and E and F from the other. The tertiary fold of IL-10 is most similar to interferon-γ.
Molecular Weight	37,294 = homodimer; 18,647 = monomer (calculated)
Sedimentation Coeff.	Unknown
Isoelectric Point	7.81 (calculated)
Extinction Coeff.	$\varepsilon^{0.1\%} = 0.47\ mg^{-1}\ cm^2$ or $8740\ M^{-1}\ cm^{-1}$ ($\lambda = 280$)
Enzyme Activity	None
Coenzymes/Cofactors	Unknown
Inhibitors	Unknown
Biological Functions	IL-10 was originally isolated for its ability to inhibit the production of the pro-inflammatory cytokines from activated T_{H1} helper T-cells including IL-2, interferon-γ, and tumor necrosis factor β. Many additional activities have been identified including differentiation of activated B cells into antibody secreting cells. These activities result in the IL-10 induced suppression of cell mediated immunity while stimulating humoral immunity. The pleotropic activities of IL-10 are conveyed to cells by high affinity cell surface receptors. Cell signaling is believed to occur by ligand induced receptor oligomerization.
Physiology/Pathology	IL-10 deficient mice develop chronic inflammatory bowel disease, or enterocholitis. The anti-inflammatory affects of IL-10 are seen upon administation of recombinant IL-10 through inhibition of IL-6, IL-8, and interferon-γ production.
Degradation	Unknown
Genetics/Abnormalities	Unknown
Half-life	Unknown

Concentration	< 5 pg/ml in human serum of healthy individuals
Isolation Methods	
Amino Acid Sequence	Human IL-10 shares 84% sequence homology wiht an open reading frame of the Epstein Barr virus (BCRF1). A second viral homolog has been identified in equine herpes virus type II. IL-10 is structurally related to the α-helical cytokines which bind to receptors in the cytokine receptor superfamily. Cytokines in this family include growth hormone, the interferons, and the interleukins.
Disulfides/SH-Groups	Interleukin-10 has two disulfide bonds which link Cys-12 to Cys-108 and Cys-62 to Cys-114.
General References	Mosman, T.R. *Advances in Immunology* 1994, **56**:1-26. Windsor, W.T. *Biochemistry* 1993, **32**:8807-8815. Walter, M.R. *Biochemistry* 1995, **34**:12118-12125. Eskdale, J. *Immunogenetics* 1997, **46**:120-128.
Ref. for DNA/AA Sequences	Viera, P. et al. *Proc. Natl. Acad. Sci. USA* 1991, **88**:1172-1176.



Ribbon diagram of IL-10. Helices from one chain of IL-10 are labeled A-F.

Interleukin-11

David A. Williams

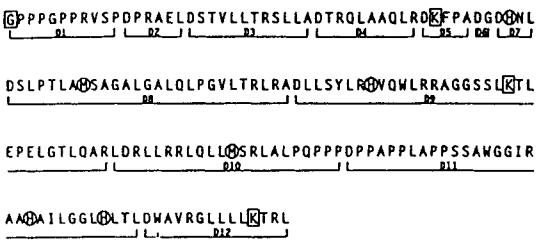
Synonyms	Adipogenesis inhibitory factor; Neumega ®
Abbreviations	IL-11
Classifications	None
Description	A cytokine expressed by most cells, but at high levels in hippocampus and testis (of mice). Precursor protein consists of 199 aa, including a 21 aa leader sequence. Rich in proline residues (12%), lacks cysteine (ie, lacks potential disulfide bonds). IL-11 is highly helical (57±1%) and is thermally stable (Tm=90^{O}C). The C-terminal region of IL-11 is predicted to be helical and to be involved in primary recepter binding.
Structure	The tertiary structure of IL-11 has not been determined. According to one structural model proposed, IL-11 contains a four-helix bundle topology. Based on chemical modifications and deletion analysis, IL-11 is predicted to contain two recepter-binding sites. Met-58 and the C-terminus are involved in primary recepter binding, whereas Lys-41 and Lys-98, as well as, positively charged arginine residues, which are focused on the exposed face of helix C, may be involved in GP130 recepter binding.
Molecular Weight	19,144
Sedimentation Coeff.	Unknown
Isoelectric Point	11.7
Extinction Coeff.	NA
Enzyme Activity	None
Coenzymes/Cofactors	None
Inhibitors	Chemical modification (alkylation or site-directed mutagenesis) of Met-58 results in a 25-fold decrease in bioactivity. Elimination of 8 or more C-terminal residues completely abolish activity.
Biological Functions	Pleiotropic effects on multiple tissues. Characterized initially by thrombopoietic activity. Shown to have stimulatory activity on hippocampal cells, small intestinal crypt cells and inhibitory effects on the differentiation of adipocytes.
Physiology/Pathology	Stimulates recovery of hematopoiesis, small intestine and spermatogenesis after injury from cytotoxic agents and irradiation. Appears to reduce proinflammatory cytokine expression in some cells, particularly the release of tumor necrosis factor-α (TNF-α) by monocyte/macrophages. Asthma-inducing viruses, such as respiratory syncytial virus (RSV), parainfluenza virus type 3, and rhinovirus stimulate the expression of IL-11 by respiratory epithial cells. IL-11 induces synthesis of the tissue inhibitor of metalloproteinase 1, which may have a protective effect on connective tissue.
Degradation	Unknown

Genetics/Abnormalities	Gene located on 19 q13.3- q13.4. This region of chromosome 19 contains several finger genes. IL-11 gene not implicated in known human genetic disease. Gene spans 7 kb and consists of 5 exons and 4 introns.
Half-life	Mean resonance time-12 hrs.
Concentration	Undetectable in serum from normal individuals. Plasma levels increased to 400-500 pg/ml during thrombocytopenia following bone marrow transplantation.
Isolation Methods	Isolation of recombinant protein expressed in *E.coli*
Amino Acid Sequence	
Disulfides/SH-Groups	No disulfides; no cysteines.
General References	Du, X. and Williams, D.A. *Blood* 1997, **89**:3897-3908. Elias, J.A. et al. *J. Biol. Chem.* 1994, **269**:22261-22268. Nandurkar, H.H. et al. *Oncogene* 1996, **12**:585-593.
Ref. for DNA/AA Sequences	Paul, S.R. et al. *Proc. Natl. Acad. Sci. USA* 1990, **87**:7512-7516.

Appendix 1

Marta J. Czupryn. John M. McCoy, and Hubert A. Scoble. Structure -Function Relationships in Human Interleukin-11. *J. Biol. Chem.* 1995, **270**:978-985.

Published by:
American Society for Biochemistry and Molecular Biology
9650 Rockville Pike
Bethesda, Maryland 20814-3996
(301) 530-7145 Fax: (301) 571-1824

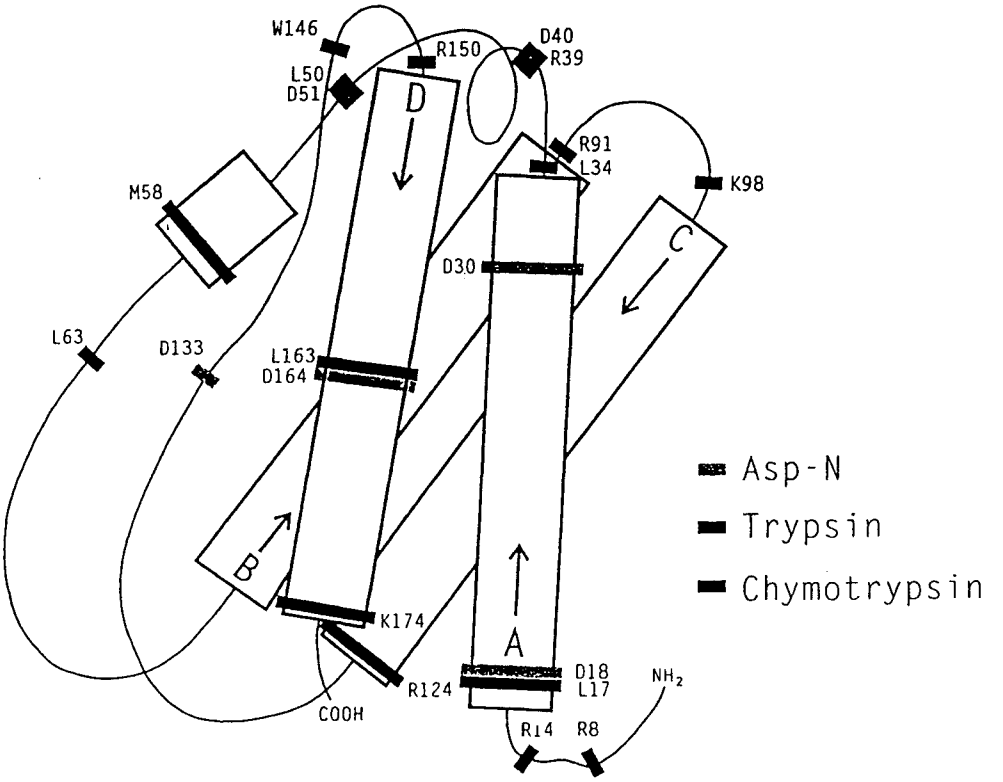


Amino acid sequence of rhIL-11. Potential alkylation sites are denoted by *circles*, potential succinylation sites by *squares*. Peptides expected from endoproteinase Asp-N cleavage are *underlined* and *numbered D1 through D12*.

Appendix 2

Marta J. Czupryn. John M. McCoy, and Hubert A. Scoble. Structure -Function Relationships in Human Interleukin-11. *J. Biol. Chem.* 1995, **270**:978-985.

Published by:
American Society for Biochemistry and Molecular Biology
9650 Rockville Pike
Bethesda, Maryland 20814-3996
(301) 530-7145 Fax: (301) 571-1824



Limited proteolysis of rhIL-11 with endoproteinase Asp-N, trypsin, and chymotrypsin. Location of protease-sensitive sites in the proposed four-helix bundle of rhIL-11.

Interleukin 12

David H. Presky and Alvin S. Stern

Synonyms	Cytotoxic Lymphocyte Maturation Factor, Natural Killer Cell Stimulatory Factor, T Cell Stimulating Factor
Abbreviations	IL-12, IL12, CLMF, NKSF, TSF
Classifications	Interleukin
Description	IL-12 is a heterodimeric cytokine that was identified independently by investigators at Genetics Institute with the Wistar Institute and at Hoffmann-La Roche. IL-12 has been shown *in vitro* to exert a number of immunomodulating effects on the activities of NK/LAK cells, T lymphocytes, and hematopoietic progenitor cells. The biological activities of IL-12 suggest that it may have therapeutic utility as an immunoenhancing antitumor, antiparasite, or antiallergy agent when used alone or in combination with other interleukins, such as IL-2 or IL-4.
Structure	The IL-12 protein consists of two disulfide-bonded subunits having approximate molecular masses of 40 kDa and 35 kDa, respectively. The 40-kDa subunit cDNA has been found to encode a protein of 328 aa, including a 22 aa signal peptide. The mature protein of 306 aa (calculated $M_r = 34{,}699$) contains 10 cysteine residues and four potential N-linked glycosylation sites. The 35-kDa subunit cDNA was found to code for mature-protein of 197 aa (calculated $M_r = 22{,}513$) preceded by a 22 aa signal peptide. The predicted aa sequence of the mature 35-kDa subunit contains seven cysteine residues and three potential N-linked glycosylation sites. The IL-12 heterodimer 40- and 35-kDa subunits contain 10 and 20% carbohydrate, respectively. The two subunits of IL-12 are encoded by two distinct genes and are unrelated to each other. The 35-kDa subunit of IL-12 shares homology with IL-6, G-CSF, and chicken myelomonocytic growth factor, and has, like most other cytokines, an α-helix rich structure. The 40-kDa subunit is not homologous to other cytokines, but belongs to the hemopoietin receptor family and most resembles the IL-6 receptor and the ciliary neurotrophic factor receptor.
Molecular Weight	The protein backbone has a molecular size of 57.2 kDa although the glycosylated protein is approximately 70 kDa as determined by analytical ultracentrifugation.
Sedimentation Coeff.	Unknown
Isoelectric Point	4.5–5.3
Extinction Coeff.	1.34 (280 nm, 1%, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Neutralizing antibodies raised against both human IL-12 and murine IL-12 are potent inhibitors of IL-12 bioactivity. Recent reports suggest that the murine 40-kDa subunit has inhibitory activity against murine IL-12.

Biological Functions	IL-12 has been shown to have numerous biological activities including: augments the cytolytic activity of NK and LAK cells; induces IFN-γ production by T and NK cells; stimulates proliferation of activated T and NK cells; inhibits production of IgE; promotes the development of T_h1 cells; synergizes with various hematopoietic growth factors to enhance hematopoietic stem cell proliferation; induces the expression of various cell surface proteins including adhesion molecules.
Physiology/Pathology	While the exact physiological role of IL-12 is still being clarified, IL-12 has multiple immunomodulatory properties, many of which may be at least partially mediated indirectly by the stimulation of release of IFN-γ and other cytokines. IL-12 is secreted from activated monocyte/macrophages and B lymphocytes and has potent *in vitro* and *in vivo* effects on stimulating NK and LAK cell cytolytic activity, inducing CTL responses to tumor cells, and inhibiting IgE production. IL-12 also plays a major role in promoting the development of T_h1 over T_h2-type responses. IL-12 exhibits *in vivo* activity against murine *Leishmania* and *Toxoplasma* infection, and antitumor activity in a number of murine tumor models. Mice treated with anti-IL-12 exhibit decreased resistance to some intracellular pathogens such as *Leishmania* and *Listeria*. In addition, PBMC from HIV-infected individuals show a marked decrease in IL-12 production in response to bacterial antigens.
Degradation	Neither the specific catabolic pathways nor degradation products of IL-12 are known.
Genetics/Abnormalities	The 40-kDa subunit of human IL-12 is encoded by chromosome 5 (5q31-q33), and the 35-kDa subunit by chromosome 3 (3p12–3q13.2). No genetic abnormalities in IL-12 production or responsiveness have been identified.
Half-life	2–3 hours in rats (i.v. bolus administration, human IL-12) 17 hours in rhesus monkeys (40 minutes i.v. infusion, human IL-12)
Concentration	IL-12 is secreted by activated mononuclear cells and acts locally in a paracrine fashion. Local concentrations generated by IL-12 secretion are not known.
Isolation Method	The recombinant protein is purified from the medium of cultures of CHO cells transfected with IL-12 cDNAs. Natural sources of the protein include activated monocyte/macrophages and B lymphocytes. The secreted IL-12 is purified by several chromatographic steps including ion exchange, reversed phase HPLC and/or immunoaffinity chromatography.
Amino Acid Sequence	The full aa sequences of both subunits of human and mouse IL-12 are known. The cloning of the mouse IL-12 cDNAs revealed that IL-12 is relatively well conserved between mouse and human. The predicted aa sequence of the mouse 40-kDa subunit is 70% identical to that of man, whereas the human and mouse 35-kDa IL-12 subunits display 60% aa sequence identity.
Disulfides/SH-Groups	Although it is known that the two subunits of IL-12 are disulfide-bonded, it is not known which of the subunits' cysteines are intra- or interchain bridged. There is at least one free SH-group.
General References	Kobayashi, M. et al. Identification and purification of natural killer cell stimulatory factor (NKSF), a cytokine with multiple biological effects on human lymphocytes. *J. Exp. Med.* 1989, **170**: 827–845.

Stern, A. S. et al. Purification to homogeneity and partial characterization of cytotoxic lymphocyte maturation factor from human B-lymphoblastoid cells. *Proc. Natl. Acad. Sci. USA* 1990, **87**: 6808–6812.
Gately, M. K. et al. Regulation of human lymphocyte proliferation by a heterodimeric cytokine, IL-12 (cytotoxic lymphocyte maturation factor). *J. Immunol.* 1991, **147**: 874–882.
Perussia, B. et al. Natural killer (NK) cell stimulatory factor or IL-12 has differential effects on the proliferation of TCR-$\alpha\beta^+$,TCR-γd^+ T lymphocytes, and NK cells. *J. Immunol.* 1992, **149**: 3495–3502.
Manetti, R. et al. Natural killer cell stimulatory factor (IL-12) induces T helper type 1 (Th1)-specific immune responses and inhibits the development of IL-4-producing Th cells. *J. Exp. Med.* 1993, **177**: 1199–1204.

Ref. for DNA/AA Sequences

Gubler, U. et al. Coexpression of two distinct genes is required to generate secreted, bioactive cytotoxic lymphocyte maturation factor. *Proc. Natl. Acad. Sci. USA* 1991, **88**: 4143–4147.
Wolf, S. F. et al. Cloning of cDNA for natural killer cell stimulatory factor, a heterodimeric cytokine with multiple biologic effects on T and natural killer cells. *J. Immunol.* 1991, **146**: 3074–3081.
Schoenhaut, D. S. et al. Cloning and expression of murine IL-12. *J. Immunol.* 1992, **148**: 3433–3440.

Interleukin-1 Receptor Antagonist

Michael F. Smith, Jr.

Synonyms	None
Abbreviations	IL-1Ra; IL-1ra; IRAP
Classifications	
Description	A cytokine produced by monocytes/macrophages, neutrophils, keratinocytes, hepatocytes, and some fibroblasts and epithelial cells. Two isoforms of the protein have been identified, one secreted (sIL-1Ra) and one intracellular (icIL-1Ra). sIL-1Ra is translated as a 177 aa precursor protein which includes a N-terminal 25 aa hydrophobic leader sequence which is cleaved during secretion. icIL-1Ra is produced as a result of the use of an alternative first exon which is spliced into the region encoding the hydrophobic leader sequence resulting in a 159 aa protein in which the N-terminal 21 aa of the sIL-1Ra leader sequence are replaced by MAL.
Structure	The structure of sIL-1Ra has been solved by 3-D NMR spectroscopy and x-ray crystalography. It has the same overall β-trefoil structure as IL-1α and IL-1β, characterized by six β strands forming a β-barrel which is closed at one end by another six β-strands.
Molecular Weight	sIL-1Ra 17.1 kDa (unglycosylated) 22-26kDa, glycosylated; icIL-1Ra 18 kDa
Sedimentation Coeff.	Unknown
Isoelectric Point	4.9 to 5.5
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Binds to the cell surface receptors for Interleukin 1 (IL-1) and inhibits the binding of IL-1α and IL-1β. IL-1Ra does not induce receptor internalization and has no known agonist activity although there has been a report that icIL-1Ra can decrease mRNA stability for some genes without effecting IL-1β-induced signaling.
Physiology/Pathology	Serves as an anti-inflammatory molecule to modulate the effects of interleukin-1. IL-1Ra levels are found to be elevated in acute and chronic inflammatory disorders including sepsis, rheumatoid arthritis, inflammatory bowel disease (IBD), psoriasis and Hodgkin's disease. Exogenous administration of IL-1Ra has demonstrated beneficial effects in human and animal models of rheumatoid arthritis, IBD, graft versus host disease, and chronic myelogenous leukemia. Approximately 100-1000-fold more IL-1Ra than IL-1 is required to completely inhibit the effects of IL-1

	due to the "spare receptor effect". Thus the ratio of IL-1:IL-1Ra may be the critical factor in determining the outcome of disease.
Degradation	Uncharacterized. Plasma clearance rate of 2.0 ± 0.3 mg/min/kg.
Genetics/Abnormalities	The IL-1Ra gene (IL-1RN) is on chromosome 2q14.2 in the same region as IL-1α and IL-1β. The entire IL-1Ra gene encompasses 16.5 kb with 4 exons. The intron between the icIL-1Ra exon 1 and sIL-1Ra exon 1 spans approximately 9.5 kb and may contain exons for other splice variants of which the mRNA for one has been reported. A genetic polymorphism consisting of a variable number (2 to 6) tandem 86-bp repeats (VNTR) is located in intron 2. The most common allele (allele 1) has 4 repeats. Allele 2, which has 2 repeats has been associated with psoriasis, ulcerative colitis, SLE, and alopecia areata. No association of allelles containing 3,5, or 6 repeats with disease has been reported. Additionally, a translational silent, single base pair polymorphism has been found in exon 2 that is in 100% linkage with VNTR allele 2.
Half-life	Initial half life of 21 minutes, terminal half life of 108 minutes
Concentration	Plasma: 0.2-0.4 ng/L increasing to as much as 54 μg/L in acutely ill patients
Isolation Method	a) Chromatography on Mono-Q superose followed by gel filtration over Superose 12, and C4-reverse phase HPLC. b) Superose 12 FPLC, followed by multiple TSK Biol-Sil 125 HPLC, and finally C4 reverse phase HPLC. c) recombinant protein purified by sequential S-sepharose and Mono-Q ion exchange chromatography.
Amino Acid Sequence	sIL-1Ra leader: MEICRGLRSHLITLLLFLFHSETIC mature sIL-1Ra N-terminus RPSGRKKSKM icIL-1Ra N-terminus: MALETICRPSGRKKSKM Homologous to IL-1α and IL-1β -C-terminus
Disulfides/SH Groups	None
General References	Arend, W.P. *Adv. Immunol.* 1993, **54**:167-227. Lennard, A.C. *Crit. Rev. Immunol.* 1995, **15**:77-105. Arend, W.P. et al. Interleukin-1 Receptor Antagonist. In: *Human Cytokines: Handbook for Basic and Clinical Research.*, Aggarwal, B.B. and Gutterman, J.U. (eds.), Blackwell Scientific Publishers, Boston 1996, **Vol. II.**
Ref. for DNA/AA Sequences	Eisenberg, S.P. et al. *Nature* **343**:341-346. Haskill, S. et al. *Proc. Natl. Acad. Sci. (USA)* **88**:3681-3685.

Human Protein Data
Edited by André Haeberli

J Chain

Jiri Mestecky and Michael W. Russell

Synonyms	Joining chain, Fast (F) chain
Abbreviations	J chain
Classifications	Component chain of polymeric immunoglobulins, fast electrophoretic mobility (albumin – α1 globulin region).
Description	J chain is a glycoprotein covalently linked by disulfide bridges to the penultimate cys residue of the α and μ chains of human IgA and IgM; found in association with polymeric IgA and IgM in serum and external secretions (saliva, tears, colostrum and milk, intestinal, respiratory and genital secretions, bile); found inconsistently and only in trace amounts in urine, in free form; produced by plasma cells and their precursors at various stages of differentiation (in human pro-B, pre-B, and B cells); almost all Ig-producing plasma cells in mucosal tissues express J chain irrespective of the Ig isotype; Ig-producing cells in the bone marrow are mostly J chain-negative (except for multiple myeloma see Physiology/Pathology).
Structure	Member of the Ig superfamily, single polypeptide chain with the secondary structure of one Ig-like domain; 137 aa residues, single N-linked complex carbohydrate side-chain (7–8% of mol. wt.).
Molecular Weight	15,600 ± 200
Sedimentation Coeff.	1.28 S
Isoelectric Point	Unknown
Extinction Coeff.	6.53 (275 nm [max], 1%, 1 cm)
Enzyme Activity	Unknown
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	May participate, directly or indirectly, in the intracellular assembly of polymeric Ig; binding of Secretory Component (SC) by polymeric Ig during the transepithelial transport depends on the content of J chain in Ig polymers.
Physiology/Pathology	Plasma cells from bone marrow of patients with multiple myeloma are usually highly positive for J chain, irrespective of Ig isotype or L chain type produced; 6 cases of J chain disease have been described (bone marrow contains variable numbers of J chain-positive but Ig-negative plasma cells with other clinical and laboratory sings of multiple myeloma).
Degradation	J chain not linked to Ig is degraded intracellularly.
Genetics/Abnormalities	The gene encoding J chain is located on chromosome 4; a processed pseudogene is located on chromosome 8. J chain deficiency has not been described.

Half-life	Unknown
Concentration	Not found in a free form in serum or external secretions; therefore, its level depends on the concentration of polymeric Ig in these fluids.
Isolation Method	Free J chain can be isolated from purified polymeric Ig (after cleavager of disulfide bonds) by a combination of electrophoretic and chromatographic procedures.
Amino Acid Sequence	Primary structure has been determined by protein and DNA sequencing.
Disulfides/SH-Groups	J chain contains 8 cys residues (6 in 3 intra chain disulfide bridges, 2 in bridges with penultimate cys residues of the α or μ chains of polymeric Ig).
General References	Koshland, M. E. The coming of age of the immunoglobulin J chain. *Annu. Rev. Immunol.* 1985, **3**:425–453. Mestecky, J. J chain. **In** *Encyclopedia of Immunology*, Roitt, I. M. and Delves, P. J. (eds.) Academic Press, San Diego, 1992, pp. 943–944. Brandtzaeg, P. Role of J chain and secretory component in receptor-mediated glandular and hepatic transport of immunoglobulins in man. *Scand. J. Immunol.* 1985, **22**:111–146.
Ref. for DNA/AA Sequences	Max, E. E. and Korsmeyer, S. J. *J. Exp. Med.* 1985, **181**:832–849. Max, E. E. et al. *Mol. Immunol.* 1994, **31**:1029–1036.

Kallistatin

Julie Chao, Robert Q. Miao and Lee Chao

Synonyms	Kallikrein-binding protein
Abbreviations	KBP; HKBP
Classification	Electrophoretic Mobility: α_1 region at pH 8.6
Description	Kallistatin is a new member of the serine proteinase inhibitor (serpin) superfamily with a unique reactive site P_2P_1-P_1' of Phe-Phe-Ser. Unlike many other serpins which are only present in the plasma, kallistatin is found in various tissues, cells and bodily fluids. The major site of kallistatin synthesis is the liver, and to a lesser extent the pancreas, kidney, lung, heart, colon and salivary gland. A Phe-Ser peptide bond in kallistatin acts as the reactive site and is cleaved by human tissue kallikrein following complex formation, resulting in the release of the C-terminal fragment from the kallistatin molecule.
Structure	The tertiary structure is not yet determined for the native form, however molecular modeling of the protein and the crystal structure of the homologous protein α1-AT are known. There are 8 helices and 3 large β-sheets. Kallistatin has a unique reactive site P_1-P_1' of Phe-Ser. The P_1 Phe is a crucial determinant of specificity and the P_2 Phe is essential for creating the hydrophobic environment for kallikrein-kallistatin interaction. The Ala residues at the hinge region (P_{10} and P_{12}) play a pivotal role in maintaining the structure and conformation of the kallistatin molecule.
Molecular Weight	58,000 Da
Sedimentation Coeff.	Unknown
Isoelectric Point	4.6 to 5.2, IEF gives 4 main bands.
Extinction Coeff.	8.4 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Susceptible to inactivation by catalytic cleavage at the reactive center by tissue kallikrein, chymotrypsin, pancreatic elastase
Biological Functions	Serine proteinase inhibitor. It is capable of inhibiting tissue kallikrein' kininogenase and amidolytic activity in vitro. Kallistatin may have multiple functions: maintenance of kallikrein's bioavailability, vasodilation, regulation of cell growth, protection in inflammation and blood coagulation.
Physiology/Pathology	Kallistatin is present in the circulation and in tissues, blood cells and endothelial cells involved in cardiovascular function. Kallistatin plays a potential role in blood pressure regulation and has potent vasodilating activity independent of kallistatin's interaction with tissue kallikrein. Low plasma levels are associated with an increased risk of hypertension. The plasma kallistatin levels in patients with liver disease and sepsis were significantly reduced as compared with normal controls. Urinary kallistatin

levels were higher in patients with pregnancy-induced hypertension with proteinuria.

Degradation

Complexes formed with target enzymes are cleared by the liver.

Genetics/Abnormalities

The gene is on chromosome 14q31-32.1, and is 9.618 kb in length with the typical five exon-four intron gene structure. The structure and organization of the kallistatin gene are similar to those of the genes encoding α1-ACT (antichymotrypsin), PCI (protein C inhibitor), and α1-AT (antitrypsin). The human kallistatin gene is a single-copy gene. The kallistatin cDNA is 1.284 kb in length encoding 427 aa residues. It shares 44-46% homology with other serpins such as human α1-ACT, PCI, α1-AT and thyroxin-binding globulin. Several polymorphisms were identified in the promoter, coding and intron regions (eg. GCG type and Deletion type) among different ethnic groups.

Half-life

Unknown

Concentration

Plasma: 21.2 ± 2.0 μg/ml; seminal fluid: 8.4 ± 3.4 μg/ml; amniotic fluid: 0.75 ± 0.08 μg/ml

Isolation Method

80 mM barium citrate adsorption and 6-16% polyethylene glycol fractionation of human plasma followed by affinity chromatography on heparin-agarose, ion exchange chromatography on DEAE-Sepharose CL-6B and hydroxylapatite chromatography, with further purification by phenyl-superose FPLC.

Amino Acid Sequence

A member of a family of homologous inhibitors (the serpins). The alignments of the reactive center regions of selected serpins:

		P1	P1'
HKS	AGTEAAAATTFAIKFF	SAQTN.	RHIL
RKS	VGTKAAAATGSFATFF	SAQPK.	KRYL
PCI	SGTRAAAATGTIFTFR	SARLNSQ	RLV
ACT	EGTEASAATAVKITLL	SALVETR	TIV
AT	KGTEAAGAMFLEAIPM	SIPPE....	V
HCII	EGTQATTVTTVGFMPL	STQVR....	F
PN-1	DGTKASAATTAILIAR	S...SPP	WFI
PAI-1	SGTVASSSTAVIVSAR	M...APE	EII
ATIII	EGSEAAASTAVVIAGR	SLNPNRV	TFK
PAI-2	EGTEAAAGTGGVMTGR	T.GHGGP	QFV
C1	TGVEAAAASAISVA.R	TLLVF	EVQQ

HKS: human kallistatin; RKS: rat kallistatin; PCI: protein C inhibitor; ACT: α1-antichymotrypsin; AT: α1-antitrypsin; HCII: heparin cofactor II; PN-1: protease nexin I; PAI: plasminogen activator inhibitor-I; ATIII: antithrombin III; PAI-2: plasminogen activator inhibitor-II; C1: C1 inhibitor.

Disulfides/SH-Groups

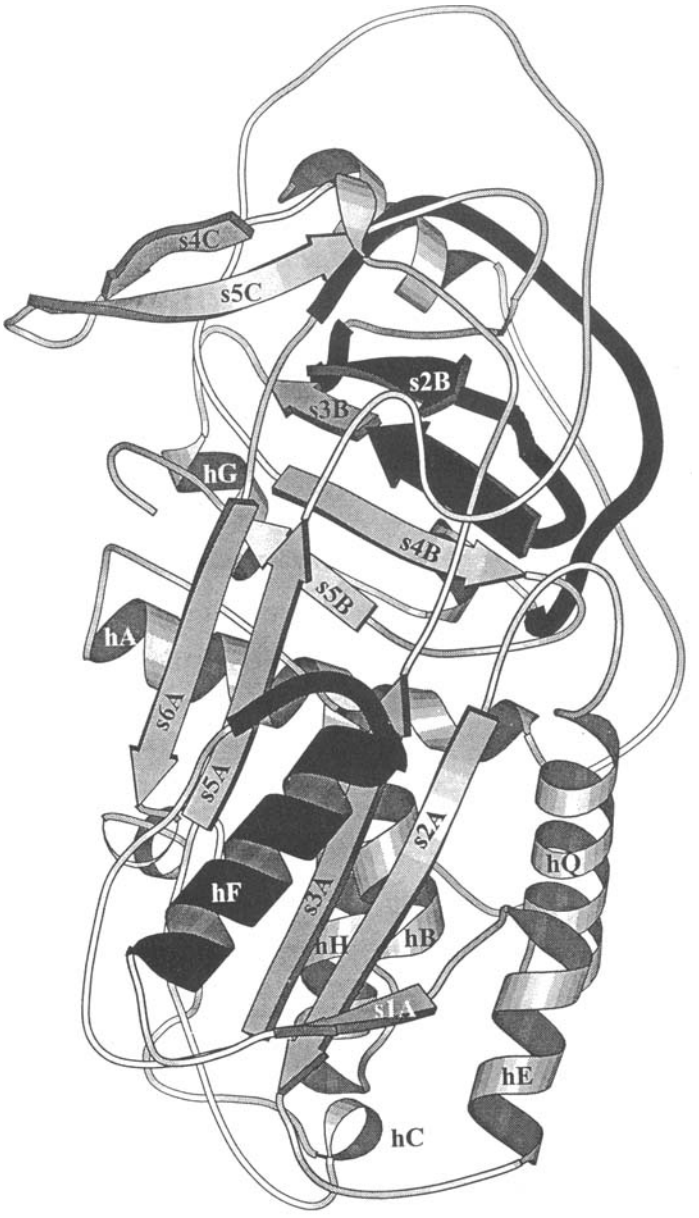
3 sulfhydryl residues at C31, C94 and C263, respectively

General References

Zhou, G.X. et al. *J. Biol. Chem.* 1992, **267**:25873-25880.
Chai, K.X. et al. *Genomics* 1994, **23**:370-378.
Chao, J. and Chao, L. *Biol. Chem. Hoppe-Seyler* 1995, **376**:705-713.
Chao, J. et al. *Immunopharmacology* 1996, **32**:67-72.
Chao, J. et al. *J. Lab. Clin. Med.* 1996, **127**:612-620.
Chao, J. et al. *Clin. Invest.* 1997, **100**:11-17.

Ref. for DNA/AA Sequences

Chai, K. X. et at. *J. Biol. Chem.* 1993, **268**:24498-24505.



Homology model of human kallistatin

Kininogens

Werner Müller-Esterl

Synonyms

α-Cysteine proteinase inhibitor; α-Thiol proteinase inhibitor

Abbreviations

HK; H-kininogen; LK; L-kininogen

Classifications

Family 3 of the cystatin superfamily

Description

The bulk of kininogens is synthesized by the liver and secreted into the plasma. Two types exist in man, i.e. H-kininogen and L-kininogen; a third type, T-kininogen is present in the rat. Kininogens are secretory proteins endowed with transient signal peptides of 18 residues. The mature proteins consist of 626 residues (HK) and 409 residues (LK); note that residues 1 through 383 are shared by the two types of kininogens. The kininogens undergo post-translational modification by glycosylation at 3 N- and 9 O-linkage sites (HK) or 3 N-linkage sites (LK). Both kininogens have a hydroxylated proline residue at position 365. Kininogens are present in plasma in the single-chain form. Release of kinin moiety from the kininogens results in the formation of two-chain molecules with the heavy and light chain portions interconnected via disulfide-bridge.

Structure

Multi-domain molecules with 5 (LK) or 6 modules (HK): 3 cystatin-like units (domains D1 to D3), the kinin-bearing domain (D4), followed by a single C-terminal domain ($D5_L$) in LK or two distinct domains ($D5_H$ and $D6_H$) in HK. The proposed domain boundaries are given below. An X-ray structure of the kininogens are not available.

Protein	Residues	Protein	Residues
L-kininogen	1-409	H-kininogen	1-626
D1	1-112	D1	1-112
D2	113-234	D2	113-234
D3	235-357	D3	235-357
D4	358-383	D4	358-383
$D5_L$	384-409	$D5_H$	384-492
		$D6_H$	493-626
heavy chain	1-362	heavy chain	1-362
kinin segment	363-371	kinin segment	363-371
light chain	372-409	light chain	372-626

Molecular Weight

(apparent M_r by non-reducing SDS-PAGE unless otherwise indicated)
HK 120,000
HK 83,000 (by sedimentation analysis)
HK heavy chain 63,000
HK light chain 5,000

LK 68,000
LK heavy chain 63,000
LK light chain 5,000

Sedimentation Coeff.

Unknown

Isoelectric Point

HK: 4.45 to 4.65; LK: 4.45 to 4.90

Extinction Coeff.

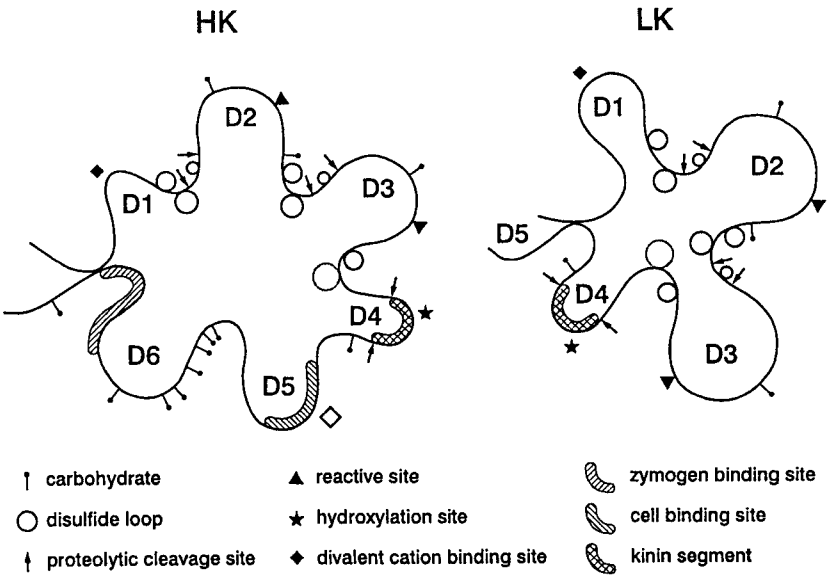
HK: 7.3; LK: 7.8 (280nm, 1%, 1cm)

Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Kininogens are multifunctional proteins which serve biological roles as kinin precursors in blood pressure regulation, as cofactors of the endogenous blood coagulation pathway, and as proteinase inhibitors in inflammatory processes.
Physiology/Pathology	Kininogens are the precursor molecules of the vasoactive kinins. Limited proteolysis of kininogens by kallikreins results in the release of bradykinin (plasma kallikrein, trypsin) or kallidin (tissue kallikrein) from domain D4. Kininogens dock to specific binding proteins (e.g. gC1qR) exposed on the surface of endothelial cells; the cell binding is mediated by kininogen domains D3 and $D5_H$. HK (but not LK) participates in the contact phase activation of the endogenous blood coagulation cascade. HK exposes a unique binding site of 31 residues on $D6_H$ which forms tight complexes with the plasma proteins prekallikrein or factor XI. Following exposure of subendothelial surfaces, the circulating HK complexes bind to the newly exposed surfaces via a region of 20 residues located in $D5_H$. Binding of factor XII (Hageman factor) next to HK results in the local assembly of the initiator factors (FXII, FXI, prekallikrein, HK) thereby triggering contact phase activation. Circumscribed release of kinin from HK by activated plasma kallikrein regulates local blood flow and promotes inflammatory reactions at the site of injury, and contraction of smooth muscle cells. Kininogens are the major cysteine proteinase inhibitors of the human plasma. They inhibit papain (D2, D3) and calpain (D2 only) with K_i's in the picomolar range at a stoichiometry of 1:2 (kininogen over enzyme). Cystatin-like domain D1 which does not inhibit cysteine proteinases is a weak binder of divalent cations (Ca^{2+}).
Degradation	Kininogens follow the normal degradation pathways of plasma glycoproteins. Degradation is typically initiated by kinin release; the HK light chain is then further trimmed by plasma kallikrein. Limited proteolysis of the heavy chain by trypsin and chymotrypsin-like enzyme releases single domains (D2, D3) effective in cysteine proteinase inhibition.
Genetics/Abnormalities	Kininogens are encoded by a single gene of approx. 27 kBp located on chromosome 3q-26qter; it contains 11 exons. The two distinct mRNAs are generated by alternative splicing: the mRNA for HK contains exons 1 through exon 10. The LK mRNA comprises exons 1 through 9, the 5' part of exon 10 (exon 10a), and exon 11. Accordingly the translation products of the kininogen mRNAs share their N-terminal portions (residues 1 to 383 encoded by exons 1 to 10a) and differ in their C-terminal regions (residues 384 on, encoded by exons 10b and 11, respectively). At least two inherited deficiencies are known: Fitzgerald trait (plasma concentration of HK $\leq$ 1% of normal; LK $\leq$ 48 %) and Williams trait (HK and LK $\leq$ 0.1%). Other deficiencies (Flaujeac, Read, Dayton, Fujiwara) are characterized by the absence of immunoreactive HK and the presence of varying amounts of LK. Deficiency of kininogens is not associated with bleeding disorders indicating that the contact phase activation system can be bypassed. Kininogen deficiency does not result in hypertension suggesting that back-up systems for blood pressure regulation must exist. Kininogen plasma levels are largely unchanged in the acute phase. An

	animal model of kininogen deficiency has been described for Brown Norway rats ("catholic" strain) where HK and LK, but not T-kininogen, are absent from plasma (≤ of the normal plasma concentration).
Half-life	Half-life of kininogens in plasma is approximately 2 to 3 days
Concentration	Plasma: HK 74 ± 18.5 mg/L (0.9 μmol/L); LK 135 ± 25.9 mg/L (2.7 μmol/L). Ascites: HK 20.9 mg/L; LK 86.0 mg/L. Urine: HK 0.05 mg/L; LK 2.8 mg/L. Seminal plasma: HK 0.25 mg/L; LK 0.9 mg/L. Liquor: HK 0.07 mg/L; LK 0.4 mg/L. Saliva: HK ≤ 0.001 mg/L; LK 0.085 mg/L.
Isolation Method	Isolation from plasma by affinity chromatography on CM-papain-Sepharose, followed by ion exchange chromatography.
Amino Acid Sequence	Domains D1 to D3 share significant sequence homology with the mammalian cystatins; D2 and D3 contain the prominent sequence motif of QVVAG forming part of the active site of cysteine proteinase inhibitors. Furthermore, D1 to D3 have significant sequence identities with human α_2-HS glycoprotein (fetuin). Domain D3 contains a cell binding site which has been precisely mapped to 27 residues. Domain D4 bears the kinin peptide (RPPGFSPFR) including the tripeptide PPG, a canonical acceptor site for proline hydroxylases. Domain $D5_H$ contains a region rich in histidine, glycine and lysine with multiple copies of the dipeptide Gly-His; $D5_H$ is structurally related to the histidine-rich glycoprotein of the human plasma. The cell binding site has been mapped to 20 residues of $D5_H$; it contains a Zn^{2+} binding site of the general structure (HEXXH) and overlaps the artificial surface binding site. The zymogen binding site have been mapped to 25 residues of $D6_H$.
Disulfides/SH-Groups	18 cysteine residues; 8 intra-chain disulfide loops of the heavy chain, and a single inter-chain loop spanning the heavy and light chain portions.
General References	Müller-Esterl, W. et al. Kininogens revisited. *Trends Biochem. Sci.* 1986, **11**:336-339. Müller-Esterl, W. et al. Human kininogens. *Methods Enzymol.* 1988, **163**: 240-256. De la Cadena, R.A. and Colman, R.W. Structure and functions of human kininogens. *Trends Pharmacol. Sci.* 1991, **12**:272-275.
Ref. for DNA/AA Sequences	Takagaki, Y. et al. Cloning and sequence analysis of cDNAs for human high molecular weight and low molecular weight prekininogens. *J. Biol. Chem.* 1985, **260**:8601-8609. Kitamura, N. et al. Structural organization of the human kininogen gene and a model for its evolution. *J. Biol. Chem.* 1985, **260**:8610-8617. Kellermann, J. et al. Completion of the primary structure of human high-molecular-mass kininogen. The amino acid sequence of the entire heavy chain and evidence for its evolution by gene triplication. *Eur. J. Biochem.* 1986, **154**:471-478. Kellermann, J. et al. Arrangement of disulphide loops in human low-M_r kininogen. *Eur. J. Biochem.* 1987, **247**:15-21.

Domain model of the mammalian kininogens

(not drawn to scale)



Lactase-phlorizin hydrolase

Giorgio Semenza and Ned Mantei

Synonyms	Lactase-glycosylceramidase; Beta-glycosidase complex
Abbreviations	LPH
Classifications	EC 3.2.1.23; EC 3.2.1.45; EC 3.2.1.62
Description	A glycoprotein intrinsic to the microvillar (brush border) membrane of small intestinal epithelial cells (enterocytes) of (young) mammals. The enzyme is not found in other cells. LPH is synthesized as a large precursor (pre-pro-LPH, 1927 aa residues), which is processed first by signal peptidase (after pos. 19) and then by at least another intracellular protease after Arg-868. The polypeptide chain from pos. 869 to the C-terminus (pos. 1927), i.e., LPH, is transferred to the Golgi and then to the brush border membrane to which it is anchored (N_{out}-C_{in}) via a hydrophobic stretch between approx. pos. 1883 and 1901, as an "ectoenzyme". During its passage it is polymannosylated (in the ER) and then complex glycosylated (in the Golgi). LPH is both N- and O-glycosylated. The polypeptide chain between pos. 20 and 868 of pro-pro-LPH (the pro-sequence(s)) is rather hydrophilic and has no detectable enzymatic activity.
Structure	The complete aa sequence of the precursor of LPH, i. e., of pre-pro-LPH, is known. The 1927-residue-long polypeptide comprises five domains: (i) a putative signal sequence of 19 aa at the N-terminus, (ii) a long pro region of 849 aa, (iii) mature LPH, containing both lactase and phlorizin hydrolase domains, (iv) a hydrophobic membrane anchor near the C-terminus, and (v) a short, hydrophilic segment at the very C-terminus.
Molecular Weight	Pre-pro-LPH: 216,393 (unglycosylated, calculated from the inferred aa sequence), or approx. 240,000 (high mannose intermediate). LPH: 120,068 (unglycosylated; calculated); approx. 160,000 (fully glycosylated form).
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	Beta-glycosidases
Coenzymes/Cofactors	None
Substrates	The single polypeptide LPH carries two independent catalytic sites: "lactase" splits lactose, cellobiose, and, in general, beta-glycosides having a hydrophilic (but, in some cases, a hydrophobic) aglycone; "phlorizin hydrolase" (identical with glycosylceramidase) splits aryl- and alkyl-beta glycosides.
Inhibitors	Competitive inhibition by tris (hydroxymethyl)-aminomethane, 1-deoxynorjirimycin and phlorizin (the latter for lactase; it is a substrate for phlorizin hydrolase).
Biological Functions	Hydrolysis of lactose (for lactase) and of glycosylceramides (for phlorizin hydrolase).

Physiology/Pathology	In all mammals LPH begins to appear in fetal life (at approx. the 15th week in man), and develops rapidly shortly before birth. At weaning it declines (in most mammals) to the low values of adulthood (1/5 to 1/20 of that at birth). In man also it can decline in childhood or early youth in many Black and Asian races ("adult-type hypolactasia", affecting more than one half of mankind, transmitted as an autosomal recessive trait LPH-mRNA may stay at high levels, or, alternatively, declines parallel to LPH); in most Caucasians LPH remains high in adulthood and declines only slightly in old age. LPH can be absent at birth ("congenital lactose intolerance", an autosomal recessive trait). In a large number of acute and chronic intestinal diseases, LPH is reduced to low levels; among the disaccharidases it is the first to decline and the last to recover.
Degradation	Normally LPH is eliminated through loss with desquamated enterocytes from the villus tip. It is possible that in adult-type hypolactasia LPH is degraded intracellularly.
Genetics/Abnormalities	The gene encoding LPH has been localized to chromosome 2. The chromosomal gene is about 55 kb in length. The complete intron-exon organisation (including the sequences of the 17 exons and the borders of all introns, as well as the sequence of about 1,000 bp of 5'-flanking region) is known.
Half-life	Unknown
Concentration	5% of total microvillar protein (in babies).
Isolation Method	Solubilisation from the brush border membrane with detergents such as Triton X-100 followed by various chromatographic steps and immunoadsorption.
Amino Acid Sequence	The sequence of the 1927 aa residues in pre-pro-LPH and of the 1069 residues in LPH is known. The N-terminal sequence is: AFTFPSEV...
Disulfides/SH-Groups	Unknown
General References	Semenza G. and Auricchio, S. Small-intestinal disaccharidases. In: *The Metabolic and Molecular Basis of Inherited Disease,* Scriver, C., et al. (eds.) McGraw-Hill, New York 1995, pp. 4451-4480. Skovbjerg, H., et al. *Eur. J. Biochem.* 1981, **114**:653-661. Sterchi, E., et al. *J. Clin. Invest.* 1990, **86**:1329-1337. Sebastio, G., et al. *Am. J. Hum. Gen.* 1989, **45**:489-497. Lloyd, M., et al. *J. Clin. Invest.* 1992, **89**:524-529. Montgomery, R.K., et al. *FASEB J.* 1991, **5**:2824-2832.
Ref. for DNA/AA Sequences	Boll, W., et al. *Amer. J. Hum. Genet.* 1991, **48**:889-902. Mantei, N., et al. *EMBO J.* 1988, **7**:2705-2713.

Lactate dehydrogenase

Lucia Sacchetti and Giuliana Fortunato

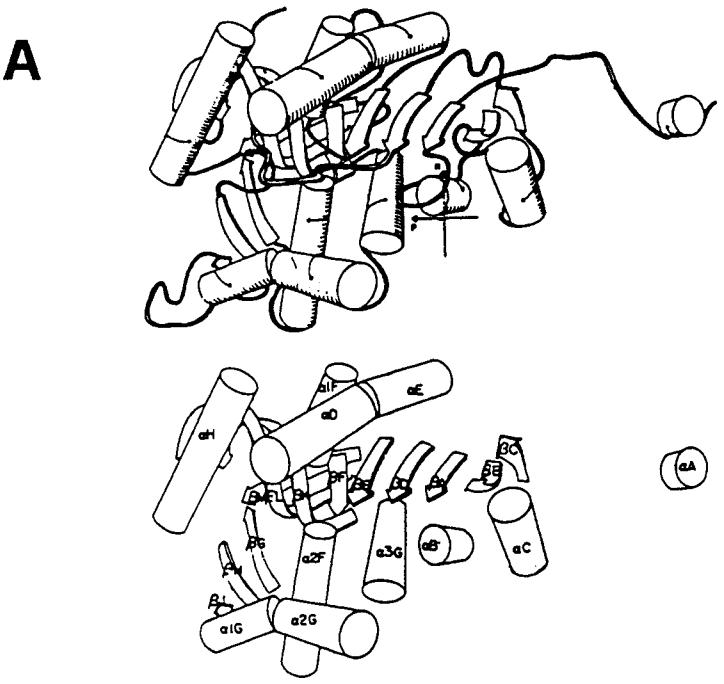
Synonyms	L(+)lactate: NAD^+ oxidoreductase
Abbreviations	LDH; LD
Classifications	EC 1.1.1.27
Description	LDH plays a relevant role in the glycolytic pathway. The enzyme is ubiquitous, being widely distributed in most tissues as a cytosolic enzyme. It has also been found in the external surface of the inner mitochondrial membrane in liver cells, but its metabolic function is unknown. LDH is a tetramer consisting of two different subunits, A or M (muscle), and B or H (heart). Five different isoenzymes: LDH_1 (H_4), LDH_2 (H_3M), LDH_3 (H_2M_2), LDH_4 (HM_3), LDH_5 (M_4) according to their electrophoretic mobility, result from the combination of the two different subunits. Another LDH, the homotetrameric C, or X, lactate dehydrogenase, LDH (C_4) is found only in mature testis and spermatozoa.
Structure	Each subunit of tetrameric native LDH has an independent binding site for NAD^+ or NADH. X-ray crystallography of the dog-fish LDH_5 (M_4) showed that each monomer is globular with an arm of 20 aa at the N-terminal which links it to other monomers of the tetramer. About 45% of the aa chain has an α-helix conformation, and an additional 20-25% has a β-structure in the form of a parallel pleated sheet and antiparallel pleated ribbon. The main residues involved in catalysis are Arg-171 and His-195.
Molecular Weight	135 kDa: tetramer (diffusion and sedimentation); 37 kDa, approx: LDH(M) subunit; 36 kDa, approx: LDH(H) subunit (translation from cDNA sequence analysis).
Sedimentation Coeff.	6.48 S
Isoelectric Point	4.5 LDH(M); 4.8 LDH(C)
Extinction Coeff.	16.4 (280nm, 1%, 1cm) LDH(H).
Enzyme Activity	LDH catalyzes the reversible oxidation of L-lactate to pyruvate in the presence of NAD^+, with the release of a proton. Pyruvate + NADH + H^+ 1 L(+)lactate + NAD^+. A salt bridge is formed between the carboxylate group of the substrate and the side-chain of Arg-171. The lactate hydroxyl group forms a hydrogen bond with the unprotonated imidazole ring of His-195, while the carbonyl group of the pyruvate can form a hydrogen bond with the protonated form. His-195 acts to orient the substrate and as an acid-base catalyst, removing the proton from lactate during oxidation (see Figure).
Coenzymes/Cofactors	The enzymatic activity requires the presence of NAD^+. Crystallographic studies have shown that binding of NAD^+ to the enzyme changes the configurational state of the protein so that residues from 98 to 114 of the chain acquire the conformation required for catalysis.

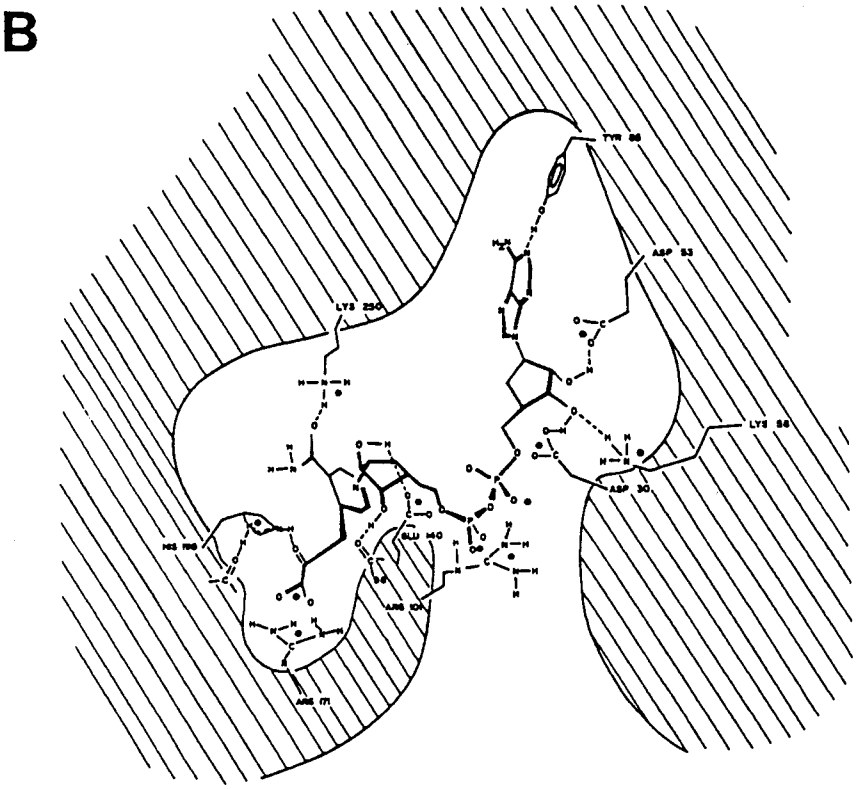
Substrates	Pyruvate, Km: 1.2 x 10^{-4} M for LDH(H) and 4.6 x 10^{-4} M for LDH(M); L-lactate; NAD^+, NADH. Other substrates: α-ketobutyrate, α-ketovalerate, L-α-hydroxybutyrate.
Inhibitors	Competitive inhibitors are oxamate and oxalate for pyruvate and lactate, respectively.
Biological Functions	LDH plays a multiple role in the metabolism of carbohydrates: it promotes the breakdown of glucose to lactate in anaerobic glycolysis; in aerobiosis, LDH catalyzes the oxidation of lactate to pyruvate that, rough the citric acid cycle generates NADH and ATP; it is also involved in liver reverse glycolysis converting lactate to glycogen. The different kinetic properties of the two homotetramers, M-type and H-type, allows the action of LDH in both aerobic and anaerobic metabolism. The M-type is abundant in tissues with anaerobic metabolism, whereas in aerobic tissues the H-type enzyme prevails.
Physiology/Pathology	Because LDH is widely distributed in mammalian tissue, elevated serum enzyme levels are observed in many conditions: intravascular hemolysis, acute myocardial infarction, hepatitis, infectious mononucleosis, and cancer of various tissues and organs. More specific signals of disease are LDH isoenzyme variations, for instance in myocardial infarction LDH_1 increases more than LDH_2 (flip: $LDH_1/LDH_2 > 1$) and in hepatic diseases LDH_5 is increased; the LDH_4/LDH_5 ratio has also been proposed for the discrimination between primary and secondary hepatic neoplasias.
Degradation	Mainly within the reticulo-endothelial system.
Genetics/Abnormalities	In humans, the LDH-M, H and C polypeptides are encoded by three different genes, most likely originated from one ancestral gene during the course of evolution. The LDH-M and LDH-C genes are closely linked on chromosome 11, while the LDH-H gene is located on chromosome 12. The coding sequence is interrupted by six introns at homologous positions. Genetic abnormalities of LDH-M and H affect the expression of the protein and cause enzyme deficiency. LDH-M deficiency has been associated to a form of hereditary myopathy, characterized by acute renal failure caused by myoglobinuria after severe exercise. In these patients the disease severity may be related to the degree of muscle oxidative capacity.
Half-life	≈116 h: LDH; ≈ 113 h: LDH_1; ≈ 14h: LDH_5
Concentration	In human serum, LDH activity is (at 37°C) 227-450 U L^{-1} in both adult males and females (one LDH unit catalyzes the cleavage of 1 μmole of pyruvate in one minute).
Isolation Method	LDH has been isolated from different muscles as well as from yeast and bacteria, the main methods used being ammonium sulphate fractionation, ion exchange chromatography and affinity chromatography.
Amino Acid Sequence	LDH-H consists of 333 aa residues, whereas LDH-M consists of 332 aa residues. The deduced aa sequence from cDNA of human LDH-H showed a 75% homology with human LDH-M, whereas there was a 76% homology between human LDH-M and LDH-C.
Disulfides/SH-Groups	One SH group per subunit.
General References	Vassault, A. Lactate dehydrogenase. In: *Methods of enzymatic analysis* Bergmeyer, H.V. (ed.), VCH, Weinheim, 1983, pp.118-126.

Castaldo, G. et al. *Clin. Chem.* 1991, **37**:1419-1423.
Miyamjima, H. et al. *Intern. Med.* 1995, **34**:502-506.

Ref. for DNA/AA Sequences

Takeno, T. and Li, S.S.-L. *Biochem. J.* 1989, **257**:921-924.
Maekawa, M. et al. *Hum. Genet.* 1991, **88**:34-38.
Chung, F.Z. et al. *Biochem. J.* 1985, **231**:537-541.
Millan, J.L. et al. *Proc. Natl. Acad. Sci. USA* 1987, **84**:5311-5315.
The nucleotide sequences have been deposited in the EMBL/GENBANK data libraries under the following accession numbers: LDH-H (Y00711); LDH-C (J02938).





A. Schematic representation of a subunit of dogfish A_4 lactate dehydrogenase.
(Taken from: Holbrook, J. J., Liljas, A., Steindel, S. J., and Rossman, M. G. Lactate dehydrogenase, 11:210, in: The enzyme, 3rd ed., Boyer, P. D. (ed.). New York, San Francisco, London: Academic Press, 1975).

B. The binding sites of lactate dehydrogenase for the coenzyme, NADH, and substrate, pyruvate.
(Taken from: Holbrook, J. J., Liljas, A., Steindel, S. J., and Rossman, M. G. Lactate dehydrogenase, 11:240, in: The enzyme, 3rd ed., Boyer, P. D. (ed.). New York, San Francisco, London: Academic Press, 1975).

Lactoferrin

Geneviève Spik and Jean Montreuil

Synonyms	Lactotransferrin; Human red milk protein
Abbreviations	Lf; Fe_2Lf (iron-saturated); ApoLf (iron-free)
Classifications	Electrical mobility: β_1-fraction
Description	Protein of the transferrin family which binds reversibly 2 ferric and 2 carbonate ions. Synthesized by different secretory epithelial cells and early cells of the neutrophilic blood series. First isolated from human milk and found latter in almost all biological fluids, on mucosal surfaces, in brain and in the secondary granules of polymorphonuclear leukocytes. A glycoprotein containing 6% carbohydrate in the form of 2 biantennary glycans of the N-acetyllactosamine type. The presence of three potential glycosylation sites (Asn-138; Asn-479; Asn-624) leads to the existence of 3 types of glycovariants. The major one (85%) is glycosylated at Asn-138 and Asn-479 while a minor glycovariant (9%) is glycosylated at the 3 glycosylation sites and a second one (5%) at only one site (Asn-479). The glycan located at Asn-479 is highly conserved in lactoferrins from different species suggesting that it should play a crucial role in the biological functions of lactoferrins. Sialic acid and fucose content heterogeneous (2 to 4 moles/mole of protein). Iron affinity higher (300-fold and dissociation pH lower (1.5 versus 4.5) compared to serum transferrin. Binds iron, chrome, manganese, cobalt, copper and gallium and indium used in nuclear medicine.
Structure	A bilobed molecule with 4 globular domains (N1 and N2 in the N-lobe and C1 and C2 in the C-lobe). Location of one iron binding site in the inter-domain cleft of each lobe. Binding of iron in the N-lobe induces conformational changes. Crystallized in native (iron-saturated and iron-free) and deglycosylated (iron-free) forms. Presents conformational similarity with bacterial proteins binding sulfate, L-arabinose and D-galactose.
Molecular Weight	80,800 (based on structural calculation); peptide chain: 76,300 (from aa sequence); carbohydrate moiety: 4,500 (from glycans structure).
Sedimentation Coeff.	5.10 S
Isoelectric Point	7.8 - 8.3
Extinction Coeff.	ApoLf: 11.0; Fe_2Lf: 14.3 (280nm, 1%, 1cm); Fe_2Lf: 0.55 (465nm, 1%, 1cm).
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Protein of iron transport. Multifunctional molecule whose activity depends mostly to its ability to tightly bind iron. Both iron-free and iron-saturated forms interact with membrane specific receptors which have been charac-

terized in human intestinal cells, hepatocytes, activated lymphocytes, bacteria and various cancerous cells. The binding site of lactoferrin to the mammalian receptor is located in the N-terminal part of the molceule covering aa residues 4 to 52. Is internalized and released in a degraded form. Does not bind to transferrin receptors. Has affinity for DNA, heparan sulfate proteoglycans, LDL receptor - related protein lipid A, lysozyme and sialylated milk glycopeptides. Modulates cell growth; protects bactericidal and anti-viral activities; plays a role in protection against enteric infection in the newborn infant.

Physiology/Pathology

Lactoferrin is involved in numerous inflammatory events and immune response functions such as regulation of granulocyte monocyte stimulating factor synthesis with suppression of myelopoiesis by inhibition of the production of IL-1, regulation of interleukin synthesis, natural killer cell activation and anti-tumor effects and maturation of T and B cells. Regulates iron kinetics in inflammation. Plasma Lf is an indicator of bone marrow activity. Regulates iron absorption.

Degradation

Eliminated from circulation by the liver. Two degradation products generated by partial proteolysis (Nt fragment: 30 kDa; Ct fragment: 50 kDa) characterized in milk, urine and feces.

Genetics/Abnormalities

Gene localized on chromosome 3, band 3q21 - q23. DNA extracted from malignant cells are less methylated than DNA extracted from normal breast cells.

Half-life

Very rapid clearance from the plasma

Concentration

Colostrum: 6.6 g/L (range 2.0 - 20.0); mature milk: 2.6 g/L (range 0.8 - 7.5); plasma: 120µg/L (range 20 - 2500).

Isolation Method

Isolated from human milk either by cation exchange chromatography or by affinity chromatography on metal-chelates, heparin-Sepharose or single stranded DNA-Agarose.

Amino Acid Sequence

Single polypeptide chain containing 692 aa. Significant sequence homologies between N- and C-lobes (37%), with serum transferrin (59%) and with the membrane associated melanotransferrin from melanocytes (39%). Partial homology with bovine-lactoglobulin and the translated sequence of Blym-1 gene. Contains KRDS sequence.

Disulfides/SH-Groups

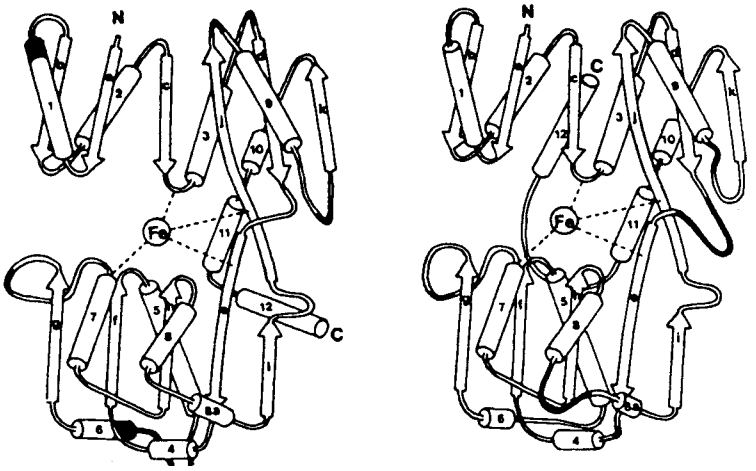
16 disulfides (6 in the N-lobe; 10 in the C-lobe). No free sulfhydryls.

General References

Montreuil, J. et al. *Biochim. Biophys. Acta.* 1960, **45**:413-421.
Metz-Boutigue, M.H. et al. *Eur. J. Biochem.* 1984, **145**:659-676.
Montreuil, J. et al. Human lactotransferrin: structure and function. In: *Proteins of Iron Storage and Transport,* Spik, G., et al. (eds.)., Elsevier Science Publishers, 1985, pp.25-38.
Németh, K. and Simonovitz, I. The biological role of lactoferrin. *Haematologia* 1985, **18**:3-12.
Fletcher, J. Iron, the Iron-binding Proteins and Bone Marrow Cell Differentiation. In: *Iron in Immunity, Cancer and Inflammation,* Souza, M. and Brock, J.H. (eds.), John Wiley and Sons Ed., 1989, pp. 223-244.
Hutchens, T.W., Rumball, S.V. and Lonnerdal, B. Lactoferrin, Structure and Function. *Adv. Exper. Med. Biol.* Plenum Press New York and London 1994, Vol **357**.

Ref. for DNA/AA Sequences

Rey, M.W. et al. *Nucleic Acids Res.* 1990, **17**:5288. EMBL accession No X53961.
Powell, M.J. and Ogden, J.E. *Nucleic Acids Res.* 1990, **18**:4013. EMBL accession No X52941.



Three dimensional structure of N-lobe (left) and C-lobe (right) human lactotransferrin determined by X-ray diffraction according to Anderson, B. F., Baker, H. M., Norris, G. E., Rice, R. W., Baker, E. N., J. Mol. Biol. 1989, 209 : 711 – 734.

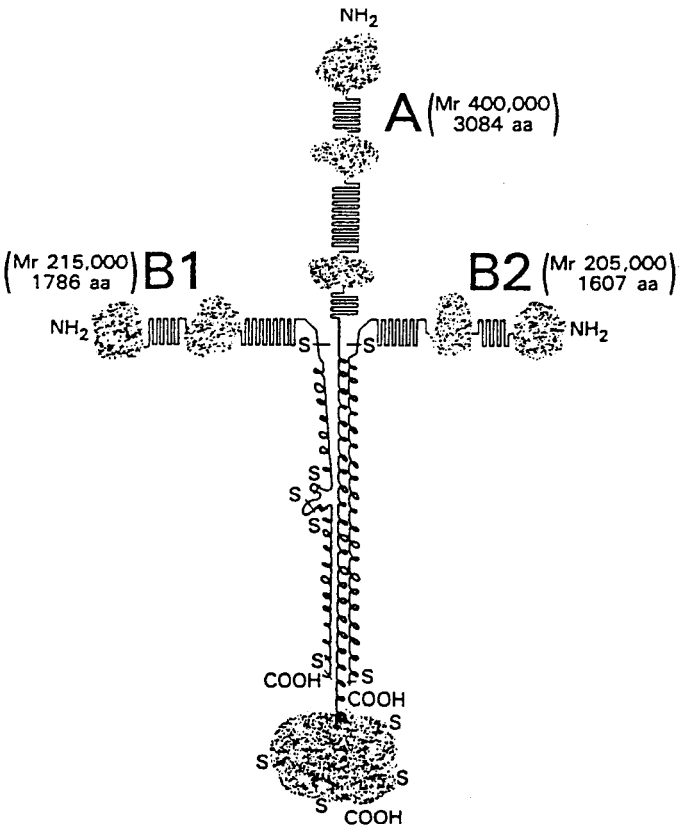
Laminin

Hynda K. Kleinman

Synonyms	Family members include EHS laminin, merosin, s-laminin, kalinin, nicein, k-laminin
Abbreviations	LM; Ln
Classifications	Extracellular matrix protein
Description	A family of high molecular weight glycoproteins (Mr=900,000) present in basement membrane, the thin extracellular matrix underlying epithelial and endothelial cells. Laminins are synthesized by many cultured epithelial and tumor cells. Contain approximately 15% carbohydrate. Have binding sites for itself, type IV collagen, heparin, perlecan and entactin. A laminin chain is the first matrix protein synthesized in development: it is detected at the two cell stage. Cell surface receptors include integrins, amyloid precursor protein, a 32/67 Kd protein, sulfatides, gangliosides, and galactosyltransferases.
Structure	Cross shaped molecules with seven globules on the short arms and one large globule at the end of the long arm are generally observed. The laminins are composed of three chains designated α, β and γ. Various homologues of each chain have been described. Seven different laminin molecules have been described , each with different chain compositions and others likely exist. Laminin-1: α1β1γ1; laminin-2 (merosin): α2β1γ1: laminin-3 (s-laminin): α1β2γ1; laminin-4: α2β2γ1; laminin-5 (kalinin/nicein): α3β3γ2; laminin-6 (k-laminin): α3β1γ1; laminin-7 (k-laminin): α3β2γ1.
Molecular Weight	900,000; α-chain: 400,000; β-chain: 210,000 (also some truncated forms); γ-chain: 200,000
Sedimentation Coeff.	11.5 S
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Promotes adhesion, migration, growth, differentiation, cytotoxic drug resistance (tumor cells), phagocytosis, neurite outgrowth (and regeneration in vivo), tumor metastases. Various active domains have been defined by synthetic peptides. Examples of active domains on the α1-chain: SIKVAV (aa 2099-2104), and ALRGD (aa 1120-1124), and LQVQLSIR (aa 2623-2630), on the β1-chain: LGTIPG (aa 442-447), RYVVLPR (aa 641-647), YIGSR (aa 929-933) and on the β2-chain: LRE.

1. Installment 1992 (revised 1997)

Physiology/Pathology	Active at 0.5-5.0 μg ml^{-1}. Not present in serum. Several autoimmune diseases have been described including Chagas' disease, African sleeping sickness (and other parasitic infections), mercury poisoning, and eclampsia. The α2-chain is absent from neuromuscular junctions in congenital dystrophy (CMD).
Degradation	Very protease sensitive, degraded by pepsin, elastase, thrombin, trypsin, cathepsin, etc.
Genetics/Abnormalities	Synthesized from three different mRNA's. Localization: β1 on chromosome 7, band q 22.
Half-life	Hours - days
Concentration	None
Isolation Method	Isolation from EHS (Engelbreth-Holm-Swarm) tumor and from murine PYS (Parietal yolk sac) carcinoma cells by salt extraction, ammonium sulfate precipitation, and Sepharose 2B and/or heparin Sepharose chromatography. After limited pepsin digestion of placenta, a similar method can be employed to obtain partially degraded laminin.
Amino Acid Sequence	Contains 7 EGF-like repeat domains, 3 on the α-chain and 2 on each of the β- and 2 on the γ-chains. Approximately 25% homology between Drosophila and mouse and about 90% homology between mouse and human.
Disulfides/SH-Groups	Many inter and intra disulfide bonds exist.
General References	Beck, K., et al. Structure and function of laminin: Anatomy of a multidomain glycoprotein. *FASEB J.* 1990, **4**:148-160. Kleinmann, H.K. et al. The laminins: a family of basement membrane glycoproteins important in cell differentiation and tumor metastases. *Vitamins Hormones* 1993, **47**:161-186. Burgeson, B. et al. A new nomenclature for the laminins. *Matrix Biol.* 1994, **14**:209-211. Timpl, R. and Brown, J.C. Supramolecular assembely of basement membranes. *BioEssays* 1996, **18**:123-132.
Ref. for DNA/AA Sequences	References for the murine α1, β1 and γ1 are given below. Human and Drosophila sequences have been determined as well as the other chain variants. Sasaki, M., et al. Laminin, a multidomain protein. The A chain has a unique globular domain and homology with the basement membrane proteoglycan and laminin. *J. Biol. Cham.* 1988, **263**:16536-16544. Sasaki, M., and Yamada, Y. The laminin B2 chain has a multidomain structure homologous to the B1 chain. *J. Biol. Chem.* 1987, **262**:17111-17117. Sasaki, M., et al. Sequence of the cDNA encoding the laminin B1 chain reveals a multidomain protein containing cysteine-rich repeats. *Proc. Natl. Acad. Sci. USA* 1987, **84**:935-939.


NH2
A (Mr 400,000 3084 aa)
(Mr 215,000 1786 aa) B1
B2 (Mr 205,000 1607 aa)
NH2
NH2
S
S
S
S
S
S
S
S
S
S
S
S
S
S
S
S
COOH
COOH
COOH

Lecithin Cholesterol Acyltransferase

Ana Jonas

Synonyms	Phosphatidylcholine-sterol acyltransferase; Phospholipid-cholesterol acyltransferase; Phosphatidylcholine:Sterol O-acyltransferase
Abbreviations	LCAT
Classification	EC 2.3.1.43
Description	A plasma enzyme synthesized principally in the liver, and at much lower levels in the brain, it consists of a single polypeptide chain of 416 aa, four N-linked carbohydrate chains at N-20, N-84, N-272, N-384, and two O-linked carbohydrate chains at T-407 and S-409. The N-linked carbohydrate chains consist of sialylated triantennary and/or biantennary complex structures with compositions of: Hex5/6 HexNAc4/5/6 NeuAc1/2/3. LCAT contains active site S-181 and H-377 residues which, together with an unidentified D or E residue, form a catalytic triad in the transesterification of a fatty acyl chain from phosphatidylcholine to cholesterol on the surface of high density lipoproteins (HDL).
Structure	The tertiary structure of LCAT has not been determined; however, because of the functional similarity and presence of comparable functional regions, human pancreatic lipase, whose structure is known, may be a reasonable model.
Molecular Weight	59,000 (20% carbohydrate) (average from composition, mass spectrometry, and sedimentation equilibrium).
Sedimentation Coeff.	3.9 S
Isoelectric Point	3.9-4.8 (at least 9 bands)
Extinction Coeff.	20.0 (280 nm, 1%, 1 cm) (for polypeptide chain only); $9.4 \times 10^4\ M^{-1}\ cm^{-1}$ (280nm) (for polypeptide + carbohydrate).
Enzyme Activity	Acyltransferase (EC 2.3.1.43) Phosphatidylcholine + cholesterol → l-lysophosphatidylcholine + cholesteryl ester. Other acyl acceptors include various sterols, lysophosphatidylcholines, and water.
Coenzymes/Cofactors	Activated by apolipoprotein A-I, the major protein component of HDL.
Substrates	Molecular substrates are phosphatidylcholines and cholesterol. The reaction occurs predominantly on the surface of HDL, which are the preferred particulate substrates because LCAT binds to HDL with higher affinity and is activated by apolipoprotein A-I.
Inhibitors	Covalent synthetic inhibitors of active site residues: diisopropylfluorophosphate, diethyl (p-nitrophenyl) phosphate, diethylpyrocarbonate; 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) and other sulfhydryl reagents inhibit by modifying cysteine residues adjacent to the active site. Noncovalent inhibitors: fatty acids, lysophosphatidylcholines, sphingomyelin, ether analogs of phosphatidylcholines.
Biological Functions	LCAT has a key role in the reverse cholesterol transport by HDL from peripheral tissues to the liver. By converting cholesterol to cholesteryl

esters on HDL, LCAT creates a gradient for cholesterol flux from cell membranes to HDL. LCAT participates in the maturation of HDL from nascent discoidal forms to the mature spherical species, containing a core of cholesteryl esters. Indirectly, LCAT affects the lipid composition and morphology of other lipoproteins and cell membranes and prevents cholesterol accumulation in tissues.

Physiology/Pathology

Primary LCAT deficiencies are relatively rare. In familial LCAT deficiency (FLD) LCAT activity and/or mass are absent in plasma; cholesteryl ester contents of lipoproteins are markedly reduced, lipoprotein patterns are grossly abnormal, and erythrocytes have abnormal structures and elevated cholesterol contents. The major clinical findings include corneal opacification, anemia, preteinuria, and renal disease. In fish eye disease (FED) LCAT activity in plasma is reduced, particularly with HDL substrates; cholesteryl ester contents in plasma are almost normal due to the low LCAT activity on LDL. In contrast to FLD, the partial LCAT deficiency in FED presents no clinical signs except for corneal opacification. Interestingly, primary LCAT deficiencies are rarely associated with premature coronary heart disease. Secondary LCAT deficiencies occur in obstructive liver disease.

Degradation

Has not been studied; but because LCAT is a plasma glycoprotein, clearance by the liver via the asialoglycoprotein receptor is likely.

Genetics/Abnormalities

The LCAT gene is located in the 16q22 region of chromosome 16, it is 4.5 kb long and contains six exons. Exon 1 encodes the 24 aa hydrophobic leader peptide and about 5% of the N-terminus of the mature protein. Exons 2-4 encode 25% of the mature protein including the lid-region that covers the active site in solution and forms part of the interfacial binding surface. Exon 5 encodes the 15% of the mature protein sequence that contains the active site serine (S-181) and regions that bind phospholipid substrates. Exon 6 encodes the remaining 50% of the mature protein and a short 3'-untranslated region of the message. To date, over 35 mutations of the LCAT gene have been described, all appear to directly affect LCAT function and cause FLD or FED. These functional mutations are dispersed over the entire LCAT gene. No genetic polymorphisms have yet been described for the LCAT gene.

Half-life

Unknown

Concentration

Plasma: 5.8 ± 1.0 mg/L. Cerebrospinal fluid: 0.12 ± 0.05 mg/L.

Isolation Method

(a) Ultracentrifugation of plasma at $d = 1.25$ g/ml yields a middle fraction enriched in LCAT and depleted in lipoproteins and plasma proteins. (b) Affigel-blue and DEAE-Sepharose column chromatography removes remaining albumin and additional proteins. (c) The final purification step consists of hydroxylapatite or Phenyl Sepharose column chromatography.

Amino Acid Sequence

The sequence from C-50 to C-74 includes a surface structural feature that is essential for binding to and reaction with lipoprotein substrates. It is functionally homologous to the lid region of pancreatic lipase and lipoprotein lipase. The sequence between residues 151-168 encompasses a predicted amphipathic helical sequence, similar to amphipathic helices of apolipoproteins, which are known to bind to phospholipids. The sequence IGHSLG-183 includes the active site serine S-181 and is a conserved sequence among various lipases. The H-377 residue has been identified as the active site histidine. Removal of the C-terminal sequence, rich in proline residues, to G-401 has no major functional consequences.

Disulfides/SH-Groups	LCAT contains two intrachain disulfide bridges, between C-50 and C-74 and C-313 and C-356. Two free cysteine residues at C-31 and C-184 are located adjacent to the active site and must be in the reduced state for optimal enzyme activity.
General References	Jonas, A. Lecithin cholesterol acyltransferase. In: *Plasma Lipoproteins*, Gotto, M.A. Jr. (ed.), Elsevier, Amsterdam. 1987, pp 299-333. Fielding, C.J. Lecithin: cholesterol acyltransferase. In: A*dvances in Cholesterol Research*, Esfahani, M. and Swaney, J.B. (eds.), The Telford Press, New Jersey. 1990, pp 271-314. Jonas, A. *Biochim. Biophys. Acta* 1991, **1084**:205-220. Kuivenhoven, J.A. et al. *J. Lipid Res.* 1997, **38**:191-205.
Ref. for DNA/AA Sequences	McLean, J. et al. *Nucleic Acids Res.* 1986, **14**:9397-9406. McLean, J. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:2335-2339. Yang, C.Y. et al. *J. Biol. Chem.* 1987, **262**:3086-3091.

Leukemia inhibitory factor (LIF)

Nicos A. Nicola

Synonyms	D-factor; Differentiation inducing factor(DIF); Differentiation inhibitory activty (DIA); Differentiation retarding factor (DRF); Human interleukin for DA cells (HILDA); Cholinergic neuronal differentiation factor (CNDF); Hepatocyte stimulating factor III (HSF III); Melanoma-derived lipoprotein lipase inhibitor (MLPLI).
Abbreviations	LIF; DIF; DIA; DRF; CNDF; HSFIII; MLPLI.
Classifications	Pleitropic Growth and Differentiation Factor.
Description	A single chain, disulfide containing glycoprotein with pleitropic actions on cells of widely different origin. May act locally where production by fibroblasts, macrophages and lymphocytes is regulated by infection or other stimuli.
Structure	A heavily glycosylated basic protein with three internal disulfide bonds. Non-glycosylated recombinant forms are fully biologically active. The structure of mouse LIF has been determined by X-ray crystallography and show to have the conformation of a long-chain 4-α-helical bundle similar to that of growth hormone, interferon-α, G-CSF and erythropoietin. The topology of the four α-helices is up-up-down-down with two long overhand loops connecting the A-B and C-D helices and a short helical segment in the A-B loop.
Molecular Weight	Protein core molecular weight 22,000 predicted from cDNA sequence. Native glycosylated form (e.g., as produced by bladder carcinoma 5637 cells) is 32,000 - 45,000 (SDS-PAGE). Hyperglycosylated recombinant forms are produced in yeast with apparent molecular weight in excess of 200,000 (SDS-PAGE).
Sedimentation Coeff.	Unknown
Isoelectric Point	8.5 - 9.0 (depending on glycosylation)
Extinction Coeff.	5.5 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	LIF is a highly pleitropic molecule. It induces the differentiation and inhibits the proliferation of M1 myeloid leukemia cells. It inhibits the differentiation of embryonic stem cells and is required for blastocyst implantation in the uterus. It induces adrenergic to cholinergic neurotransmitter switching in some neurones and is a neurotrophic and myoblast-stimulating factor. It induces the release of acute phase proteins from hepatocytes. It inhibits the lipoprotein lipase activity of adipocytes. It synergistically stimulates megakaryocyte colony formation in vitro with interleukin-3 and

	stimulates platelet formation in vivo. It acts on osteoblasts stimulating bone turnover and, in some cases, causes net bone formation.
Physiology/Pathology	LIF injected into mice causes platelet increases in the blood, increases blood calcium levels, causes the release of acute phase proteins and results in weight loss due to loss of subcutaneous and abdominal fat. It also results in elevated megakaryocytes in the spleen and bone marrow, and increased numbers of granulocyte/macrophage and eosinophil progenitors in the bone marrow and spleen. At high doses it results in thymus atrophy and pylorospasm.
Degradation	Injected LIF accumulates in several organs the most striking of which are the liver (hepatocytes) and kidney (especially the glomerulus). Degradation products accumulate rapidly in the blood and the urine.
Genetics/Abnormalities	Synthesized from a single gene (6.3 kb) containing 2 introns. Two different mRNAs (4.2 - 4.3 kb) produced by alternate splicing at the exon1/exon2 boundary produce 2 LIF proteins differing only in the leader sequence. Nevertheless, one form is thought to be secreted into the medium while the other becomes associated with the extracellular matrix. Localization: human chromosome 22q12.
Half-life	α-phase 6 - 8 min, β-phase 6 - 7 hrs (in mice)
Concentration	Plasma: normally less than 1 μg/L.
Isolation Method	Isolation from bladder carcinoma 5637 conditioned medium. Purification steps include ion-exchange chromatography (DEAE-Sepharose, does not bind in 10mM Tris HCL pH8), lentil lectin chromatography (LIF binds and is eluted by α-methyl-D-mannopyrannoside), ion-exchange chromatography (CM-Sepharose, LIF binds and is eluted during a salt gradient at pH 5.0) and reverse-phase high performance liquid chromatography (Phenyl-silica column; LIF is eluted at 41% acetonitrile in 0.1% trifluoroacetic acid).
Amino Acid Sequence	A basic protein with a 22 aa leader sequence and seven sites of potential N-glycosylation. Six cysteines probably form three internal disulfide bonds. Weak homology (approx. 30% sequence identity) with Oncostatin M.
Disulfides/SH-Groups	Six cysteine residues at position 12, 18, 60, 130, 134, 163 of the mature protein. The likely 3 disulfides are 12-134, 18-130, 60-163.
General References	Metcalf, D. et al. Leukemia Inhibitory Factor. In: *Human Cytokines: Handbook for Basic and Clinical Researchers.* Aggarwal, B.B. and Gutterman, J. (eds.) CRC Press, N.Y., in press. Gough, N.M. and Williams, R.L. *Cancer Cells* 1989, **1**:77-80. Nicola, N.A. and Hilton, D.J. Leukemia inhibitory factor and its receptor. In: *Cytokines in Health and Disease.* LeRoith, D. and Bandy, C. (eds.) JAI Press, N.Y. pp 605-660, in press.
Ref. for DNA/AA Sequences	Gough, N.M. et al. *Proc. Natl. Acad. Sci. USA* 1988, **85**:2623-2627. Stahl, J. et al. *J. Biol. Chem.* 1990, **265**:8833-8841. Lowe, D.G. et al. *DNA* 1989, **8**:351-359. Rathjen, P.D. et al. *Cell* 1990, **62**:1105-1114.

Leukocyte cathepsin G

David L. Farley

Synonyms	Chymotrypsin-like neutral protease
Abbreviations	Cat G
Classifications	EC 3.4.21.20
Description	A neutral serine proteinase synthesized in myelomonocytic precursor cells and stored in azurophil granules of neutrophils; also present to a lesser extent in monocytes. A small amount of cathepsin G is expressed on the cell surface of neutrophils which may facilitate migration through tissue during chemotactic responses. A basic (16% arginine residues) glycoprotein consisting of a single polypeptide chain which exists in at least four forms as demonstrated by electrophoresis. Contains only one potential N-linked glycosylation site; slight differences in the carbohydrate moiety believed responsible for the isoforms.
Structure	A member of the chymotrypsinogen superfamily. Amino acid sequence comparison suggests an overall structure similar to chymotrypsin. The crystal structure of cathepsin G complexed with the peptidyl phosphonate inhibitor Suc-Val-Pro-PheP-(OPh)2 has recently been determined. The protein is folded into two six stranded b-barrels held together by three trans domain segments forming a structure closely related to rat mast cell proteinase II. Unique to cathepsin G is the presence of a glutamate residue located at the bottom of the S1 specificity pocket, which divides the pocket into two compartments and allows the cleavage of synthetic peptides after lysine residues. Arginine residues are clustered in several positive charged surface patches and the single carbohydrate linkage site is distant and facing away from the active site cleft.
Molecular Weight	28,000 (SDS-PAGE), 24,000 (aa composition and equilibrium studies)
Sedimentation Coeff.	Unknown
Isoelectric Point	> 12
Extinction Coeff.	6.64 (280nm, 1%, 1cm)
Enzyme Activity	A serine proteinase with chymotrypsin-like specificity; primarily cleaves bonds adjacent to bulky aliphatic or aromatic residues such as phenylalanine or leucine. Can also cleave synthetic peptides after methionine and tryptophan. pH optimum 7.5.
Coenzymes/Cofactors	None
Substrates	Natural: glycoproteins and collagen. Specific: MeO-Suc-Ala-Ala-Pro-Phe-4-NA
Inhibitors	Alpha-1-antichymotrypsin, alpha-1-proteinase inhibitor, alpha-2-macroglobulin, DFP, chymostatin and a host of small molceular weight synthetic inhibitors.
Biological Functions	Located within the azurophil granules and demonstrating chymotryptic and bacteriocidal activity suggests that cathepsin G, together with leukocyte

	elastase and proteinase 3, participates in the killing and digestion of pathogens engulfed during phagocytosis. The actual in vivo function is unknown. Cleaves connective tissue but at a much slower rate than elastase. Cathepsin G, along with elastase, bound to neutrophil surface may facilitate directed migration from the vasculature through tissue during inflammation. Can function as a angiotension II converting enzyme and a bradykinin inactivating enzyme. Exhibits bacteriostatic and bacteriocidal properties which are independent of proteolytic activity and probably due to the cationic nature of the protein. Cathepsin G has also been reported to be involved in regulation of inflammatory responses by activation of complement C3, processing of IL-8 and in stimulation of human lymphocytes.
Physiology/Pathology	Degrades connective tissue, glycoproteins, and collagen and therefore may cause damage to tissue during an inflammatory response. Elastase activity is enhanced in the presence of cathepsin G. No disease development has been directly linked to this proteinase. However, in the absence of adequate proteinase inhibitors, uncontrolled digestion of healthy tissue could lead to the development of connective tissue diseases such as emphysema.
Degradation	Eliminated by normal neutrophil turnover; Free cathepsin G forms complexes with plasma proteinase inhibitors which are rapidly removed from circulation by specific hepatic receptors.
Genetics/Abnormalities	cDNA sequence predicts an N- and C-terminal extension not found in the mature peptide. The amino extension consists of a signal sequence followed by a dipeptide whose removal is believed necessary for zymogen activation. The role of the C-terminal extension is unknown; apparently not required for trafficking to azurophil granules. The gene contains 5 exons localized on the long arm of chromosome 14 at band q11.2 in close proximity to the human genes for granzyme B, granzyme H and mast cell chymase.
Half-life	12 min (complexed to α_1-antichymotrypsin)
Concentration	2-4 pg/neutrophil, 10-12 mM within azurophil granules.
Isolation Method	Isolated from leukocytes using a two step procedure of affinity chromatography on aprotinin-sepharose followed by ion exchange chromatography on CM-cellulose.
Amino Acid Sequence	Contains the His-57, Asp-102, and Ser-195 (chymotrypsinogen numbering) which make up the catalytic triad common to all serine proteinase. Two peptides IIGGR and HPQYNQR (residues 1-5 and 77-83 in cathepsin G, respectively) were found to exert antimicrobial activity.
Disulfides/SH-Groups	Three intramolecular disulfide bonds, no free sulfhydryls.
General References	Watorek, W. et al. Neutrophil elastase and cathepsin G: Structure, function, and biological control. *Adv. Exp. Med. Biol.* 1988, **240**:23-31. Travis, J. Structure, function, and control of neutrophil proteinases. *Am. J. Med.* 1988, **84**:37-42. Barrett, A. Cathepsin G. *Methods Enzymol.* **80**:561-565.
Ref. for DNA/AA Sequences	Salvesen, G. et al. *Biochemistry* 1987, **26**:2289-2293. GenEMBL accession number: M16117. Holn, P.A. et al. *J. Biol. Chem.* 1989, **264**:13412-13419. GenEMBL accession number: J04990

Leukocyte elastase

David L. Farley

Synonyms	Neutrophil elastase; Medullasin
Abbreviations	HLE; HNE; NE
Classifications	EC 3.4.21.37
Description	A glycoprotein composed of a single polypeptide chain consisting of 218 aa residues which is synthesized in the bone marrow by myelomonocytic lineage precursor cells. Elastase, a serine proteinase, is stored in azurophil granules of polymorphonuclear neutrophils and found to a lesser extent in monocytes. A small amount of the enzyme is expressed in a catalytically active form bound to the neutrophil cell surface. Basic protein due to the large number of arginyl residues and stabilized by four disulfide bridges. At least three isoforms exist due to minor differences in two N-linked carbohydrate side chains. The major form contains about 22% carbohydrate.
Structure	Member of the chymotrypsinogen superfamily whose overall structure is similar to porcine pancreatic elastase especially around the substrate binding and catalytic sites. An asymmetric grouping of arginine residues on one side of the enzyme may play a role in proteoglycan association within the granule. The two carbohydrate attachment sites are on opposite sides of the molecule away from the reactive site. Crystal structure of the enzyme in complex with turkey ovomucoid inhibitor and low molecular weight chloromethyl ketone inhibitors have been reported.
Molecular Weight	29,500 (SDS-PAGE and sedimentation equilibrium).
Sedimentation Coeff.	Unknown
Isoelectric Point	10 - 11
Extinction Coeff.	9.85 (280nm, 1%, 1cm)
Enzyme Activity	Elastase is a serine proteinase which preferentially cleaves after small aliphatic aa residues such as valine or alanine. Can also cleave on the C-terminal side of methionine residues, such as the Met-Ser peptide bond which resides in the active site of alpha-1 proteinase inhibitor, the physiological inhibitor of elastase. pH optimum is about 8.3.
Coenzymes/Cofactors	None
Substrates	Natural: elastin, proteoglycans, collagen. Synthetic: MeO-Suc-Ala-Ala-Pro-Val-4-NA.
Inhibitors	Naturally occurring human protein inhibitors: alpha-1-proteinase inhibitor, alpha-2-macroglobulin, secretory leukocyte proteinase inhibitor and SKALP/elafin. Other inhibitors: soybean trypsin inhibitor, eglin C, DFP, PMSF and a host of small molecular weight synthetic inhibitors.
Biological Functions	Probably involved in the hydrolysis of protein substrates taken up by neutrophils during the process of phagocytosis. This function is deduced from the location of elastase, along with cathepsin G and proteinase 3, within azurophil granules and due to their ability to digest bacterial pro-

	teins. Readily degrades elastin, collagen, proteoglycans and several plasma proteins suggesting a possible extracellular function such as connective tissue remodeling, wound debridement, neutrophil movement through tissue, and possibly complement activation. Cell surface bound elastase may facilitate the migration of neutrophils from the vasculature through tissue barriers into sites of inflammation.
Physiology/Pathology	Release of elastase from neutrophils as a result of cell death or leakage during phagocytosis can result in abnormal degradation of healthy tissue. In the absence of controlling inhibitors, elastase rapidly degrades connective tissue proteins and has been implicated in the development of pulmonary emphasema, rheumatoid arthritis, and adult respiratory disease syndrome.
Degradation	Mainly eliminated by normal turnover of neutrophils. Released proteinase is inactivated by plasma proteinase inhibitors (alpha-1-proteinase inhibitor) and quickly removed from blood circulation by a receptor mediated mechanism in liver.
Genetics/Abnormalities	cDNA sequence predicts that elastase is synthesized as a preproenzyme with N- and C-terminal peptide extensions not found in mature protein. The N-terminal extension contains a rough endoplasmic reticulum signal peptide followed by a dipeptide, removal of which is necessary for zymogen activation. The gene contains 5 exons and is located on the short arm of chromosone 19 (19p13.3) closely clustered to genes encoding proteinase 3 and azurocidine, two other proteins found within azurophil granules.
Half-life	~ 5 min (In complex with alpha-1-proteinase inhibitor).
Concentration	2 - 4 pg/neutrophil; 10 - 12 mM within granules.
Isolation Method	Isolated from leukocytes using affinity chromatography on aprotinin-sepharose followed by ion exchange chromatography on CM-cellulose.
Amino Acid Sequence	Contains His-57, Asp-102, and Ser-195 (chymotrypsinogen numbering) which make up the catalytic triad common to all serine proteinases.
Disulfides/SH-Groups	Four intramolecular disulfide bonds; no free sulfhydryls.
General References	Watorek, W. et al. Neutrophil elastase and cathepsin G: Structure, function, and biological control. *Adv. Exp. Med. Biol.* 1988, **240**:23-31. Barrett, A.J. Leukocyte elastase. *Methods Enzymol.* 1981, **80**:581-588. Bode, W. et al. *EMBO J.* 1986, **5**:2453-2458. Sinha, S. et al. *Proc. Natl. Acad. Sci. USA.* 1987, **84**:2228-2232. Bode, W. The three dimensional structure of human neutrophil elastase. In: *Alpha-1-antitrypsin Deficiency.* Crystal, R.G. (ed). Marcel Dekker Inc. NY, Basel, Hong Kong, 1996, pp 97-117.
Ref. for DNA/AA Sequences	Takahashi, R. et al. *J. Biol. Chem.* 1988, **263**:14739-14747. GenEMBO accession number: M20199-M20203. Farley, D. et al. *Biol. Chem. Hoppe-Seyler.* 1989, **370**:737-744.

Lipocortin I

Joyce A. Eldering and Brigitte M. Frey

Synonyms	Calpactin II; Annexin I; p35
Abbreviations	LC-I
Classifications	Calcium- and phospholipid-binding protein
Description	A monomer (in placenta: 20% as a covalently cross-linked homodimer) belonging to the lipocortin family of calcium- and phospholipid-binding proteins. Predominantly an intracellular protein but also found in extracellular fluids such as peritoneal fluid and lung lavage fluid. Expressed differentially in a wide variety of tissues and cell lines up to 0.2% of total protein.
Structure	The N-terminus of natural, mature LC I is acetyl-Ala-1-Met-2 indicated by mass spectroscopy analysis. The crystal structure has been determined by molecular replacement at 3.0 Å resolution.
Molecular Weight	38 kDa (determined by computer analysis of aa sequence).
Sedimentation Coeff.	Unknown
Isoelectric Point	≈ 7.9
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	Calcium required for binding to liposomes containing phospholipids and to cell membranes.
Substrates	None
Inhibitors	None
Biological Functions	A coherent universally-accepted function not yet determined. Postulated roles include: element of signal transduction, inhibitor of phospholipase-A2, -C, and -D, mediator of the anti-inflammatory properties of glucocorticoid, components of exocytotic machinery, calcium channels, cytoskeletal elements, coagulation, immune response and collagen binding.
Physiology/Pathology	Unknown
Degradation	Endogenous pathways unknown.
Genetics/Abnormalities	A one copy gene mapped to chromosome region 9q11-9q22. Abnormal forms are unknown.
Half-life	15 to 35 hrs in cultured fibroblasts.
Concentration	Unknown
Isolation Method	Placenta contains relatively high amounts of LC I (approximately 0.2% of total protein). Convential extraction procedures are used with buffer con-

	taining EDTA. Chromatography using DEAE-cellulose with subsequent fractionation by gel filtration followed by FPLC on a Mono S column.
Amino Acid Sequence	Approximately 50% aa sequence homology with other distinct proteins of the lipocortin family. Contains four copies of a conserved consesus region composed of 70 aa residues. This conserved region comprises the calcium- and phospholipid-binding domain. The N-terminus contains a unique 30 aa segment sensitive to endogenous phosphorylation (Tyr-21) by epidermal growth factor receptor tyrosine kinase and also contains a potential glycosylation site. There is no signal sequence.
Disulfides/SH-Groups	None in native isolated form.
General References	Flower, R.J. *Br J Pharmacol.* 1988, **94**:987-1015. Davidson, F.F. and Dennis, E.A. *Biochemical Pharmacology* 1989, **38**: 3645-3651. Melli, M. and Parente, L.: *Cytokines and Lipocortins in Inflammation and Differentiation - Progress in Clinical and Biological Research.* **Vol. 349**. New York: Wiley-Liss, New York, 1990. Moss, S.E. The Annexins Portland Press Research Monograph.
Ref. for DNAAA Sequences	Pepinsky, R.B., et al. *J. Biol. Chem.* 1988, **263**:10799-10811.

Lipoprotein lipase

Howard Wong and Michael C. Schotz

Synonyms	Clearing factor; Clearing-factor lipase; Heparin-releasable lipase.
Abbreviations	LPL
Classifications	EC 3.1.1.34; lipase
Description	LPL is a glycoprotein, containing 8% carbohydrate, synthesized and secreted by parenchymal cells of various extra-hepatic tissues, principally adipose and muscle. The functional enzyme appears to be a homodimer bound to the luminal surface of capillaries.
Structure	LPL crystal structure has not been obtained, however the structure of a highly related protein, pancreatic lipase suggests the monomer is composed of two independent globular domains, consisting of a catalytic and substrate binding domain.
Molecular Weight	58,000 to 60,000 Da for the monomer (SDS-PAGE), whereas the predicted molecular mass by DNA sequence is 55,000 Da. The active form of the enzyme is a dimer with a molecular mass of 120,000 Da as determined by radiation inactivation analyses.
Sedimentation Coeff.	Unknown (human enzyme), similar to 4.3 (bovine)
Isoelectric Point	Unknown
Extinction Coeff.	90,386 M^{-1} cm^{-1} (280nm)
Enzyme Activity	LPL activity is commonly measured in post-heparin plasma, or adipose tissue biopsies, at pH 8.5, with emulsified triolein substrate. Post-heparin plasma activity levels are about 220 nmol free fatty acid hydrolyzed/min/ml. Purified LPL specific activity has been reported to be approx. 35,000 nmol/h/mg.
Coenzymes/Cofactors	LPL requires a cofactor, apolipoprotein C-II, for maximal activity. The cofactor increases V_{max} of the reaction, without an effect on K_m. In genetic disorders in which the cofactor is absent, massive hypertriglyceridemia is observed.
Substrates	In the presence of apolipoprotein C-II, the major substrates are chylomicrons and very low density lipoprotein (VLDL) triglycerides, however in vitro assays have been conducted with a variety of triolein emulsions, as well as water-soluble substrates, such as p-nitrophenylacetate.
Inhibitors	Apolipoprotein C-III may be an important biological inhibitor of LPL. In vitro, LPL is inhibited by DFP and boronic acid derivatives, but not by sulfhydryl reacting reagents, suggesting the active center of the enzyme contains a serine residue. Site-directed mutagenesis studies of LPL also supports this conclusion.
Biological Functions	The major functions of LPL are to hydrolyze chylomicron and VLDL triglycerides to fatty acids for tissue utilization. LPL reaction products also contribute to the formation of LDL and HDL particles. Thus, LPL functions in tissue energy balance and lipoprotein transformations. The enzyme

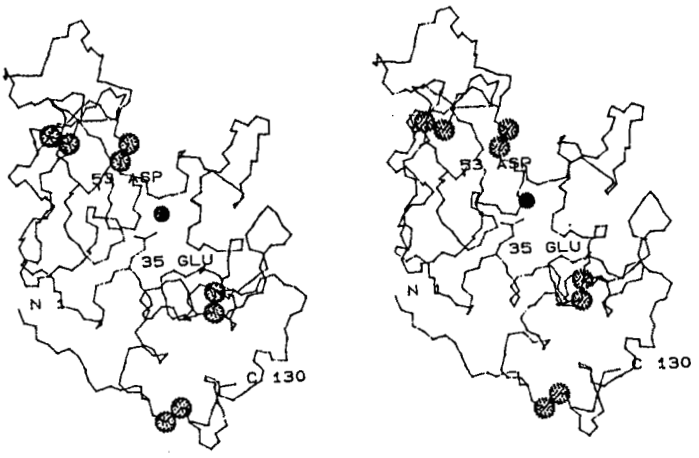
	may also bridge lipoproteins to cells and contribute to cellular lipid accumulation.
Physiology/Pathology	LPL's ability to utilize chylomicrons and VLDL as substrates indicates it is a key enzyme in lipoprotein metabolism. LPL initiates the metabolic processing of dietary lipids within the plasma compartment. Overproduction of LPL has not been observed, however the underproduction or absence of the enzyme leads to Type 1 hypertriglyceridemia. LPL deficiency is an extremely rare autosomal disorder, with an estimated frequency of one in a million.
Degradation	Release of LPL from endothelium to the plasma, followed by hepatic uptake is an important route of LPL catabolism. In addition, local turnover, i.e. uptake and degradation in either parenchymal or endothelial cells of the tissue, may also be involved. Turnover of LPL is influenced by nutritional status, exercise and pathological conditions, such as diabetes and hyperlipidemias.
Genetics/Abnormalities	The LPL gene is about 30 Kb and has been mapped to the short arm of chromosome 8 (p22). The gene consists of 10 exons and 9 introns. Exons 3 - 6 bear a high degree of homology to the gene structure of hepatic and pancreatic lipases. Numerous LPL deficient mutations have been reported, ranging from gross rearrangements of the gene to single base changes resulting in missense and nonsense mutations.
Half-life	30 - 40 min (estimation, on vascular endothelium)
Concentration	LPL post-heparin plasma concentration ranges from 150 - 300 µg/L (mean value: 210 µg/L).
Isolation Method	The enzyme has been isolated from post-heparin plasma utilizing affinity, ion-exchange, hydrophobic and gel permeation chromatography steps. Heparin-Sepharose affinity columns afford the primary means of purification.
Amino Acid Sequence	LPL aa sequence has been determined from cDNA sequencing. The overall sequence shows high degrees of homology with hepatic lipase, pancreatic lipase and vitellegenin. A central region of LPL is particularly homologous with the aforementioned proteins, suggesting that this highly conserved region is important for the function of this enzyme. The conserved region contain the active-site of these lipases and share a Gly-X-Ser-X-Gly motif (where X = any aa).
Disulfides/SH-Groups	LPL contains 10 Cys residues which are all disulfide linked. Reduction and alkylation of the disulfide bridges does not affect catalytic activity.
General References	Wong, H., Davis, R.C., Thuren, T., et al. Lipoprotein lipase domain function. *J. Biol. Chem.* 1994, **269**:10319-10323. Olivecrona, T. and Bengtsson-Olivecrona, G. Lipoprotein lipase and hepatic lipase. *Current Opinion in Lipidology* 1990, **1**:222-230. Eckel, R. H. Lipoprotein lipase; A multifunctional enzyme relevant to common metabolic diseases. *N. Engl. J. Med.* 1989, **320**:1060-1068. *Lipoprotein Lipase*, Borensztajn, J. (ed.) Evener Publishers, Chicago 1987.
Ref. for DNA/AA Sequences	Wion, K. L., et al. Human lipoprotein lipase complementary DNA sequence. *Science* 1987, **235**:1638-1641. Kirchgessner, T.G., et al. Organization of the human lipoprotein lipase gene and evolution of the lipase gene family. *Proc. Natl. Acad. Sci. USA* 1989, **86**:9647-9651.

Lysozyme

Masakazu Kikuchi and Masaaki Matsushima

Synonyms	Muramidase; Mucopeptide glycohydrolase; Mucopeptide N-acetyl-muramoylhydrolase
Abbreviations	None
Classifications	EC 3.2.1.17; Hydrolase
Description	Lysozyme is a highly cationic and simple protein which is widely distributed in several human tissues and secretions including milk, tears and saliva. It is composed of 130 aa residues which is structurally homologous to hen egg white lysozyme and has been demonstrated to hydrolyze β1-4 linkages between N-acetylglucosamines.
Structure	The number of each aa residues are 14 Ala, 14 Arg, 8 Asp, 10 Asn, 8 Cys, 3 Glu, 6 Gln, 11 Gly, 1 His, 5 Ile, 8 Leu, 5 Lys, 2 Met, 2 Phe, 2 Pro, 6 Ser, 5 Thr, 5 Trp, 6 Tyr and 9 Val. Secondary structures: α-helices Arg-5-Arg-14, Leu-25-Glu-35, Ala-90-Val-99, Val-110-Cys-116; antiparallel β-sheets A: Leu-1-Phe-3, Thr-40-Tyr-38; B: Ala-42-Asn-46, Gly55-Ser-51, Ile-59-Ser-61. Tertiary structure: The lysozyme molecule is composed of two distinct domains. The larger domain, which is rich of α-helices, consists of the residues of 1-38, 56-57 and 86-130. The other is made of a β-sheet and loops. The residues in the domain are 41-55 and 58-81. The disulfide bonds of Cys-6-Cys-128 and Cys-30-Cys-116 are in the large domain, the bond of Cys-65-Cys-81 in the minor domain and that of Cys-77-Cys-95 connects these domains. The reactive aa are Glu-35 and Asp-53. The side chain of Glu-35 is interacted with the main chain NH-group of Ala-111 via a water molecule. The side chain carboxylate oxygen atoms of Asp-53 are interacted with the side chains of Asn-46 and Asn-60. The root of mean squares deviations between the main chain atoms (N, Cα, C, O) of human lysozyme and hen egg-white lysozyne are 0.8 Å in the triclinic and tetragonal crystal forms.
Molecular Weight	14,340 Da, calculated from the formula (no carbohydrate).
Sedimentation Coeff.	2.2 S
Isoelectric Point	10.5 -11.0
Extinction Coeff.	25.7 (280nm, 1%, 1cm)
Enzyme Activity	Hydrolysis of β-(1-4) glycoside bond, kcat = 1.2 x 10^{-3} sec^{-1} at pH 5.1, 40°C.
Coenzymes/Cofactors	None
Substrates	Mucopolysaccharide, mucopeptide and synthetic substrates, such as 3,4-dinitrophenyl -β-$(GlcNAc)_4$.
Inhibitors	Tri-N-acetylchitotriose; $(GlcNAc)_3$
Biological Functions	Lysozyme dissolves living cells of *Micrococcus lysodeikticus*. Human lysozyme but not hen egg-white one is a potent inhibitor of chemotaxis and

	of the production of toxic oxygen radicals by stimulated polymorphonuclear leukocytes.
Physiology/Pathology	It has been observed that lysozyme reduces pain when administrated to cancer patients or those infected with herpes zoster. Lysozyme may also serve some function in calcification process. The level of lysozyme has been reported to change strikingly in several disease states, but the role and function of lysozyme are largely unknown.
Degradation	Unknown
Genetics/Abnormalities	Elevated serum and urine lysozyme levels have been demonstrated in monocytic and myelomonocytic leukemia. Several colon adenocarcinoma cell lines have been shown to synthesize and secrete remarkably large amounts of lysozyme. Gene mutations cause hereditary systemic amyloidosis.
Half-life	Unknown
Concentration	Normal serum level of lysozyme is 9.8 ±2.90 mg L^{-1} and lysozyme levels in tears, gastric juice and milk have been reported to be 1267 ± 58 mg L^{-1}, 74.2 ± 31.4 mg L^{-1} and 65.0 ± 10.20 mg L^{-1}, respectively. Urine levels and serum or plasma levels in patients with acute myelomonocytic and acute monocytic leukemia are 65-5250 mg L^{-1} and 5-230 mg L^{-1}, respectively.
Isolation Method	Isolation is achieved by sequential steps of acidification, cation-exchange chromatography, and reversed-phase high-performance liquid chromatography.
Amino Acid Sequence	KVFERCELAR TLKRLGMDGY RGISLANWMC LAKWESGYNT RATNYNAGDR STDYGIFQIN SRYWCNDGKT PGAVNACHLS CSALLQDNIA DAVACAKRVV RDPQGIRAWV AWRNRCQNRD VRQYVQGCGV
Disulfides/SH-Groups	4 disulfide bonds (Cys-6-Cys-128, Cys-30-Cys-116, Cys-65-Cys-81, and Cys-77-Cys-95).
General References	Fett, J.W. et al. *Biochemistry* 1985, **24**:965-975. Artymiuk, P.J. and Blake, C.C.F. *J. Mol. Biol.* 1981, **152**:737-762. Gordon, L.I. et al. *J. Clin. Invest.* 1979, **64**:226-232. Hankiewicz, J. and Sweierczek, E. *Clinica Chimica Acta* 1974, **57**:205-209. *Lysozyme*, Osserman et al. (ed.), Academic Press, New York and London 1974.
Ref. for DNA/AA Sequences	Yoshimura, K. et al. *Biochem. Biophys. Res. Commun.* 1988, **150**:794-801.



Main Chain (N, Cα and C) Model of Human Lysozyme.
The sulfur atoms of the disulfide bonds are shown by large doted spheres and a small sphere is a water molecule which is bound to the side chain of Glu35 and the main chain NH group of Ala111. The amino and carboxyl terminals are indicated by N1 and C130, respectively. The reactive amino acid residues of Glu35 and Asp53 are also shown with the labels.

Human Protein Data
Edited by André Haeberli

Macrophage colony-stimulating factor (M-CSF)

Nicos A. Nicola

Synonyms	Colony-stimulating factor-1; Macrophage and granulocyte inducer-M.
Abbreviations	M-CSF; CSF-1; MGI-M
Classifications	Growth and differentiation factor
Description	A homodimeric, disulfide bonded glycoprotein that can exist in membrane-bound as well as proteolytically released forms.
Structure	Three different proteins are possible. Each is a disulfide-linked homodimer in the cell membrane (trans-membrane glycoprotein) and each can be proteolytically released into the medium (see figure). Some forms of M-CSF have glycosaminoglycan attachment at SER277 to form a high molecular weight proteoglycan. Potential N-glycosylation sites also occur at ASNs 122, 140, 349 and 383. The X-ray crystallographic structure of soluble M-CSF shows that it is a dimer with each subunit having the conformation of a short-chain 4-α-helical bundle similar in structure to that of GM-CSF, interleukins 2, 4 and 5 and interferon-γ. However, for M-CSF the two subunits form independent 4-α-helical bundles with a head-to head arrangement. The two long overhand loops (connecting the A-B and C-D helices) form the major interface between the subunits.
Molecular Weight	Three different primary transcripts of 256 (α), 554 (β), and 438 (γ) aa (each containing a 32 aa leader sequence and a 23 aa transmembrane region) are possible and can be expressed as homodimers on fibroblast membranes (see figure). Proteolytic cleavage of each can occur to result in the releases of disulfide-linked homodimeric glycoproteins of molecular weight 88,000 (β or γ) or 56,000 (α) (SDS-PAGE) each of which is biologically active.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	M-CSF has a major role in the survival, proliferation and differentiation of monocytes and macrophages and, to a lesser extent, neutrophils. It also regulates the function of mature macrophages (tumoricidal and antibacterial activity, chemotaxis, cytokine release and inhibtion of the respiratory burst). It may also have a function on placental trophoblasts.
Physiology/Pathology	Circulating M-CSF levels are raised upon infection suggesting a role in host defence. Injected M-CSF in mice and rats leads to a peripheral monocytosis after a transient monocytopenia probably reflecting margination.

M-CSF levels are highly elevated in pregnancy and the presence of receptors on trophoblasts suggests a role in pregnancy. The op/op mouse characterized by monocytopenia and osteopetrosis (due to a lack of osteoclasts) suggests a role for M-CSF in the generation of osteoclasts.

Degradation

Biologically active M-CSF appears in the urine of patients suggesting an involvement of the kidneys in clearance. In addition activated macrophages rapidly degrade M-CSF by a receptor-mediated mechanism suggesting that macrophages may regulate the levels of M-CSF by a feedback mechanism.

Genetics/Abnormalities

A single gene produces at least three alternate m-RNAs (1.5-4.4 kb) by alternate splicing from 10 exons (see figure). In mice, the op/op mouse (osteopetrosis) has a mutation in the M-CSF gene such that no functional M-CSF is produced. In humans, the M-CSF gene was thought to be on chromosome 5q33 but more recent data suggest it is on chromosome 1.

Half-life

Unknown

Concentration

Normal plasma levels are approx. 2 μg/L but, can be elevated up to 40 μg/L in burns patients, infected patients or patients with ovarian or endometrial tumors. About 30 ng of M-CSF can be recovered from peritoneal washouts.

Isolation Method

M-CSF can be purified from human urine or from the culture supernatants of cell lines such as MIAPaCa. Current methods generally utilize calcium phosphate absorption, lectin affinity chromatography and immunoaffinity chromatography. Recombinant M-CSF, produced as a monomer in E. coli, can be recovered from inclusion bodies and refolded to produce active dimers.

Amino Acid Sequence

The longest form of M-CSF precursor (β) contains 4 N-glycosylation sites at positions 122, 140, 349 and 383 of the mature sequence and additional O-glycosylation sites have been detected. All three forms of M-CSF have a 32 aa leader sequence (cleaved during membrane expression) and a 23 aa transmembrane region (at positions 464 to 486 of the mature sequence). Only the first 150 aa of the mature sequence seem to be required for biological activity and this region contains 2 of the 4 potential N-glycosylation sites and 7 of the 10 cysteine residues. Each form of M-CSF forms a membrane-bound disulfide-linked homodimer that can be proteolytically released into the medium. Distant sequence similarity to Stem Cell Factor and flk2/flt3 ligand.

Disulfides/SH-Groups

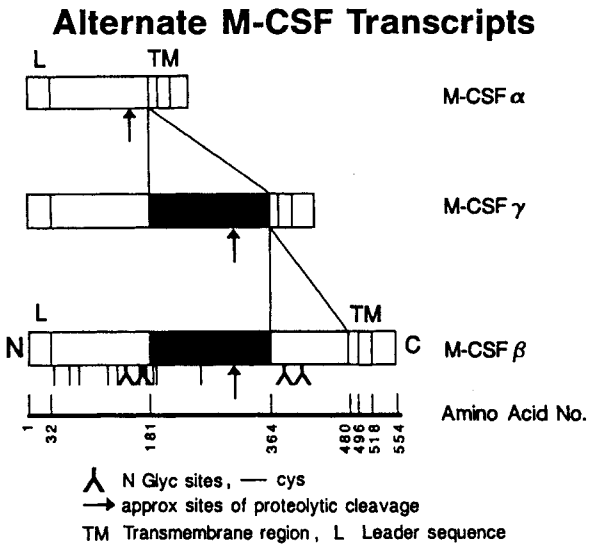
Homodimers in both membrane-bound and proteolytically released forms are known to possess inter-subunit disulfide bonds essential for biological activity. A single cysteine in each monomer (Cys-31) forms an intermolecular disulfide bond while the other six cysteines form intramolecular disulfide bonds (Cys-7-90, Cys-48-139, Cys-102-146).

General References

Ceretti, D.P., et al. *Mol. Immunol.* 1988, **25**:761-770.
Halenbeck, R., et al. *Biotechnology* 1989, **7**:710-715.
Clark, S. and Kamen, R. *Science* 1987, **230**:1229-1236.
Ralph, P., et al. In: *Molecular Basis of Lymphokine Action*, Webb, D.R., et al. (eds.) Humana Press, N.Y. 1987; pp. 295-311.
Metcalf, D. and Nicola, N.A. *The Hemopoietic Colony Stimulating Factors: From Biology to Clinical Applications*. Cambridge University Press, Cambridge, U.K., 1995.

Ref. for DNA/AA Sequences

Kawasaki, E.S., et al. *Science* 1985, **230**:291-296.
Wong, G.G., et al. *Science* 1987, **235**:1504-1508.
Ceretti, D.P., et al. *Mol. Immunol.* 1988, **25**:761-770.
Ladner, M.B., et al. *EMBO J.* 1987, **6**:2693-2698.



"Alternative protein transcripts of M-CSF according to Ceretti et al. (1988). The N-terminal side is the extracellular side, the leader sequence (L) is cleaved and the transmembrane region (TM) is shown. Each form is a disulphide-linked homodimer in both membrane-bound and secreted forms."

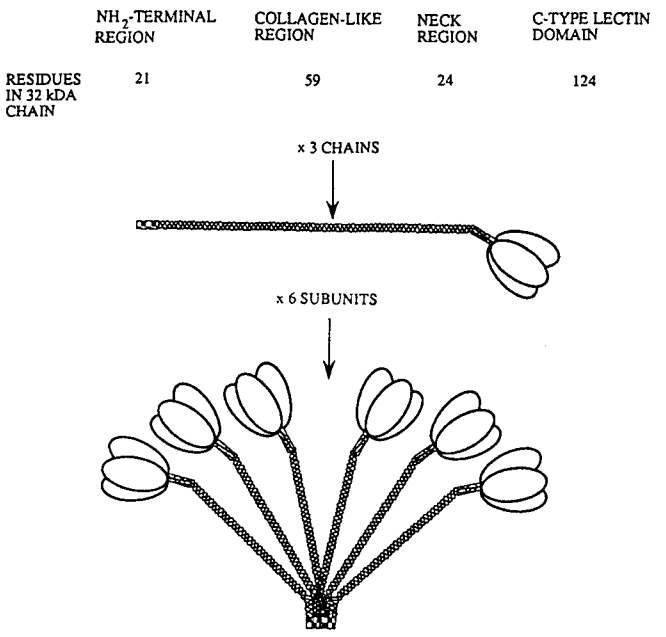
Mannose-Binding Lectin

Uday Kishore and Kenneth B.M. Reid

Synonyms	Mannan-binding lectin; Mannan-binding protein; Core-specific lectin
Abbreviations	MBL; MBP
Classifications	Group III (the Collectins) of the C-type lectins.
Description	MBL, a Ca^{2+} lectin synthesized primarily in the liver, is an acute phase reactant which shows a modest (1.5 to 3 fold) rise in serum levels during stress or infection. MBL binds to carbohydrate structures found on a wide range of pathogenic organisms. On binding to these carbohydrate structures, MBL can bring about activation of the serum complement system thus allowing recruitment of a variety of inflammatory, killing and clearance mechanisms, in an antibody-independent manner. MBL is therefore considered to play an important role in innate immunity - especially in very young children or immunodeficient individuals. MBL is also found in amniotic fluid, nasal secretions, middle ear fluid, saliva and inflamed sites - such as rheumatic joint fluid.
Structure	MBL is composed of multimers of identical polypeptide chains of 32 kDa. Three 32 kDa chains combine to make a structural subunit of 92 kDa and oligomeric forms of MBL are composed of 2 to 6 of the 92 kDa subunits. Each 32 kDa chain is composed of: an N-terminal region containing cysteine residues involved in inter-chain and inter-subunit disulfide bonding; a region of Gly-Xaa-Yaa repeating triplets which is involved in the formation of collagen-like triple helical structure; an α-helical neck region of approximately 34 residues; a C-terminal domain which contains the 14 invariant aa residues characteristic of the C-type lectin carbohydrate recognition domain (CRD). In the electron microscope the highest oligomeric form of the MBL, the hexamer of the 92 kDa structural unit, appears as a bouquet-like structure with six globular 'heads' each connected by collagen-like strands to a central core. Each globular 'head', seen in the electron microscope, contains three C-type lectin domains held together by a parallel triple-stranded coiled-coil of α-helices derived from the neck region. The trimeric CRD structure is stabilized primarily through the α-helical coiled-coils with additional interactions between a CRD and certain residues on a neighbouring chain's α-helical neck peptide as judged by X-ray crystallography of a fragment of MBL. The arrangement of the CRDs places the carbohydrate binding sites 45 Å apart, which may be important in determining how MBL trimers recognize regular arrays of branched oligosaccharides on the surfaces of microorganisms.
Molecular Weight	The largest oligomeric form of serum MBL, which consists of 18 identical 32 kDa polypeptide chains arranged as a hexamer of trimers of these chains, has a molecular weight of 576 kDa. MBL isolated from liver appears to be composed of 6 identical 32 kDa chains has a molecular weight of 192 kDa.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	5.90 (280nm, 1%, 1cm)

Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	Carbohydrate binding specificity: GlcNAc > Fucose, Mannose, ManNAc > Maltose > Glucose >> Galactose, GalNAc.
Inhibitors	None
Biological Function	The principal biological functions of MBL are: (i) recognition of pathogens by binding of the MBL CRDs to carbohydrate structures on the pathogens; (ii) activation of the serum complement system in order to recruit its inflammatory, opsonization and killing mechanisms. In the serum MBL circulates with a serine protease, designated MBL-associated serine protease (MASP), which shows approximately 40% aa sequence identity to the complement enzymes C1r and C1s. MASP preparations show functional similarity to C1s since they activate the C4 and C2 components of complement. There are two closely related (45% aa sequence identity) forms of MASP, designated MASP-1 and MASP-2. These enzymes may serve similar functions to C1r and C1s in the form of an MBL:MASP-1:MASP-2 complex similar in overall structure and function to the C1q:$C1r_2$:$C1s_2$ complex. MBL, via its CRDs, binds to a wide range of pathogens, which include Gram negative and Gram positive bacteria, yeasts, viruses, mycobacteria and various parasites. Complement activation, brought about by the MBL-MASP complex interacting with the pathogen surface, results in coating of the target pathogen with large amounts of activated C4 and C3 which leads to opsonization.
Physiology/Pathology	One of MBL's main functions is the enhancement of the killing and clearance of pathogens by bringing about antibody-independent activation of the complement system, thus zero, or low, levels of MBL may greatly increase risk to certain infections in young children and immunodeficient individuals.
Degradation	Unknown
Genetics/Abnormalities	The 7 kb gene encoding human MBL is located on the long arm of chromosome 10, within a gene cluster, at 10q11.2-q23, which also includes the genes for SP-A, SP-D and a pseudogene of SP-A. Four exons encode the four distinct regions seen in the mature 32 kDa polypeptide chain of MBL, that is, the N-terminal cysteine rich region (exon 1), a collagenous region (exons 1 and 2) and an α-helical neck region (exon 3), followed by a C-type carbohydrate recognition domain (exon 4). There are four allelic forms of the MBL gene which provide structural variants of the MBL polypeptide chain. These allelic forms are designated A, B, C and D. A, being the most common, is taken as the normal, or wild-type, form. In the B and C alleles, one glycine residue is replaced by aspartic acid (allele B) or glutamic acid (allele C) within the collagenous region. In the D allele, an arginine residue in the collagenous region is replaced by a cysteine residue. These three substitutions probably affect the formation of a stable collagen-like triple-helix. The homozygotes with respect to the B, C, or D alleles have undetectable or trace amounts of MBL. The heterozygotes (A/B, A/C or A/D) have lower levels of MBL (approximately 15 % of A/A homozygous individuals). A/A homozygous individuals can also have low levels of MBL since there are variants within the promotor region of the gene which influence serum levels.
Half-life	Approximately five days

Concentration	An average value of 1 mg/L is found in the sera of Caucasians, but there is a wide (0-5 mg/L) variation in individual values, due to the B, C and D alleles and variants in the promoter region.
Isolation Method	Application of human serum to a mannan-Sepharose column, in the presence of calcium, followed by elution with mannose, or EDTA. Further purification involves removal of anti-carbohydrate antibodies (by affinity methods), ion-exchange chromatography and size fractionation.
Amino Acid Sequence	MBL is a member of the C-type lectins and therefore contains the 14 invariant and 18 highly conserved residues seen within the approximately 120-residue-CRD found in all C-type lectins. The MBL CRD is located C-terminal to a short (34 residue-long) α-helical neck region and a collagen-like sequence composed of Gly-Xaa-Yaa repeating triplets. The presence of collagen-like sequence is a feature shared with several other C-type lectins such as lung surfactant protein A, lung surfactant protein D, and the bovine serum proteins conglutinin and Collectin-43.
Disulfides/SH-Groups	There are 7 cysteine residues at positions 5, 12, 18, 135, 202, 216 and 224 (based on numbering of the 32 kDa chain of the mature protein). The disulfide bond arrangement has not been determined, but the residues at positions 5,12, and 18 are probably involved in interchain and inter-subunit disulfide bonds. The remaining 4 residues are expected to form two intra-chain disulfide bonds (135 to 224 and 202 to 216) characteristic of that observed in other C-type lectin domains.
General References	Turner, M. *Immunology Today* 1996, **17**:532-540. Ezekowitz, R.A.B., Sastry, K. and Reid, K.B.M. (eds.) *Collectins and Innate Immunity (whole volume) Molecular Biology Intelligence Unit*, R.G. Landes Co. Austin, Texas, USA 1996. Holmskov, U. et al. *Immunology Today* 1994, **15**:67-74. Sherrif, S. et al. *Nat. Struct. Biol.* 1994, **11**:789-794. Hoppe, H.J. and Reid, K.B.M. *Structure* 1994, **2**:1129-1133. Madsen, H.O. et al. *J. Immunol.* 1995, **155**:3013-3020. Thiel, S. et al. *Nature* 1997, **386**:506-510.
Ref. for DNA/AA Sequences	Taylor, M.E. et al. *Biochem. J.* 1989, **262**:763-771. Ezekowitz, R.A.B. et al. *J.Exp. Med.* 1988, **167**:1034-1046. The sequence published by Taylor et al. (1989) is available from the EMBL/Gen Bank Data Libraries under the accession numbers X15954, X15955, X15956, X15957.



Hexameric form of MBL

Membrane Cofactor Protein (CD46)

M. Kathryn Liszewski and John P. Atkinson

Synonyms	None
Abbreviations	MCP; CD46
Classifications	
Description	MCP is a widely expressed complement regulatory protein that also serves as the measles virus receptor. MCP protects host tissue from complement activation by acting as a cofactor, along with plasma serine protease factor I, to inactivate C3b and C4b that deposit on autologous cells. Most cells express variable quantities of four isoforms that arise by alternative splicing.
Structure	MCP consists of a signal peptide (34 aa) followed by four of the repeating motifs called short consensus repeats or complement control protein (CCP) modules. The latter include 10 to 18 highly conserved residues within an ~56 to 70 aa segment. The CCP are the sites of interaction of C3b and C4b. This domain is followed by an alternatively spliced "STP" region (i.e. enriched in serines, threonines, and prolines) that is a site for O-glycosylation. The gene contains three STP exons termed A, B and C. The most common isoforms contain B + C (29 aa) or C (14 aa). Bordering this region is an area of 12 aa of unknown functional significance. A transmembrane domain, intracytoplasmic anchor and one of two distinct cytoplasmic tails (of 16 or 23 aa termed CYT-1 or CYT-2) complete the C-terminus. Designations of the four common isoforms are MCP-BC1, MCP-BC2, MCP-C1 and MCP-C2 (named for the alternatively spliced STP and tail regions).
Molecular Weight	MCP possesses an unusual electrophoretic pattern on SDS-PAGE. It migrates as two variably expressed broad bands of 51-58 kDa ("lower" band) and 59-68 kDa ("upper" band). Isoforms MCP-BC1 and MCP-BC2 constitute the upper band (bearing more O-linked sugars) and MCP-C1 and MCP-C2 constitute the lower band. The diffuseness of the bands is accounted for by the N-linked sugars.
Isoelectric Point	3.9 -5.8; higher molecular weight species are more acidic
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	MCP serves as a cofactor for the plasma serine protease factor I to proteolytically cleave and inactivate C3b and C4b deposited on self-tissue.
Substrates	C3b and C4b
Inhibitors	Unknown
Biological Functions	Host cells must be protected from "accidental" complement activation on their surfaces as may occur from proximity to an inflammatory locus or due to the physiologic, low grade turnover of C3b (and possibly C4b). MCP is a member of a family of genetically, structurally and functionally related proteins termed the "Regulators of Complement Activation" that

	are designed to provide this autologous protection. MCP is well-positioned for this role since it is expressed on most cells (except RBC) of the body and serves as a cofactor for the factor I-mediated degradation (and thus inactivation) of C3b and C4b deposited on host cells. Both a human and simian cell line showed MCP on the basolateral cell surface. MCP is an intrinsic regulator in that it protects only the cell on which it is anchored and not neighboring cells. MCP also is the receptor for measles virus (MV) and an adherence factor for group A streptococcus (*Streptococcus pyogenes*). MCP is expressed abundantly in reproductive tissues and as a nonglycosylated form on the inner acrosomal membrane of human sperm. Some evidence suggests that alterations in MCP function or expression may have a bearing on habitual abortion or infertility. Also, MCP may be part of the egg-sperm attachment mechanism and/or protect against C3b deposition during penetration by sperm.
Physiology/Pathology	No cases of MCP deficiencies have been detected although several studies have addressed MCP expression in disease states. MCP levels are increased in certain hematologic malignancies and on most epithelial-derived tumor cell lines. MCP levels are up-regulated in glomerular capillary walls and mesangial regions of diseased kidney tissues and in astrocytes following cytomegalovirus infection. SV40 transformation of fetal fibroblast lines produces a 5 to 10-fold increase in MCP expression and, interestingly, in the preferential expression of MCP-C isoforms.
Degradation	Unknown
Genetics/Abnormalities	The gene for MCP lies on a 900 kb fragment within the RCA locus on chromosome 1 at position q3.2. It consists of 14 exons and 13 introns for a minimum length of 43 kb. A Hind III RFLP has been found that correlates with the phenotypic polymorphism of MCP. Other RFLPs include Pvu II and Bgl II. There is an MCP-like genetic element that includes sequences homologous to the 5'end of MCP including the signal peptide and CCP 1-3. This MCP-like element is 93% and 84% homologous to MCP at the nucleotide and aa level, respectively. It is unknown if this partial duplication produces a protein.
Half-life	The half-life of the BC isoforms is ~ 12 hours in transfected cells (Chinese hamster ovary and NIH-3T3 mouse embryo fibroblast), while the half-life of C isoforms is ~ 10 hours. The precursor (pro-MCP) is processed differentially depending upon which tail it possesses. Isoforms with tail 1 (i.e., MCP-BC1 and MCP-C1) have a precursor half-life of approximately 13 minutes, while those bearing tail 2 (i.e., MCP-BC2 and MCP-C2) have a half-life of approximately 40 minutes.
Concentration	Radioassays have suggested the expression levels on peripheral blood mononuclear cells and polymorphonuclear, cells are ~ 10,000/cell; hemapoietic cell lines 20,000-50,000/cell; human epithelial carcinoma cells lines, HeLa and Hep-2 100,000/cell and 250,000 cell, respectively; and variable quantities on other tumor cell lines.
Isolation Method	MCP may be purified by affinity chromatography using available monoclonal antibodies or natural ligands (C3b or C4b). The ligand interaction is enhanced in low ionic strength buffer.

Amino Acid Sequence

A). MCP contains four of the ~ 60 aa repeating motifs of RCA proteins called complement control protein (CCP) modules. A CCP consensus sequence is as follows with invariant residues asterisked:

```
      I  G I       Y   E I Y     Y       I     G
-C--PP-L-NS-L-------F--GD-L-F-C--GF---G---L-C---S-W----P-C---
 *    V  A V            V   *       *   V *   A *      *
```

B). The alternatively spliced STP region is enriched in serines, threonines and prolines. This is a region that may be extensively O-glycosylated. The common higher molecular weight forms of MCP contain B + C, while lower forms contain C:

```
B:VSTSSTTKSPASSAS
C:GPRPTYKPPVSNYP
```

C) MCP isoforms have one of two alternatively spliced tails with a common anchor:

```
CYT-1 RYLQRRKKKG→TYLTDETHREVKFTSL          (16 aa)
CYT-2 RYLQRRKKKG→KADGGAEYATYQTKSTTPAEQRG   (23 aa)
```

Disulfides/SH-Groups

A) Intrachain disulfides: Each of the four CCP modules contains four invariant cysteine residues (see above) that form disulfide bridges between C1 + C3, and C2 + C4. This forms a double-loop structure for each independent CP.

General References

Liszewski, M.K. et al. *Adv. Immunol.* 1996, **61**:201-283.
Liszewski, M.K. et al. *Annu.Rev. Immunol.* 1991, **9**:431-455.
Dorig, R.E. et al. *Trends in Microbiol.* 1994, **2**:312-318.

Ref. for DNA/AA Sequences

Lublin, D.M. et al. *J. Exp. Med.* 1988, **168**:181-194.
Post, T.W. et al. *J. Exp. Med.* 1991, **174**:93-102.
Russell, S.M. et al. *Eur. J. Immunol.* 1992, **22**:1513-1518.

Database accession numbers:

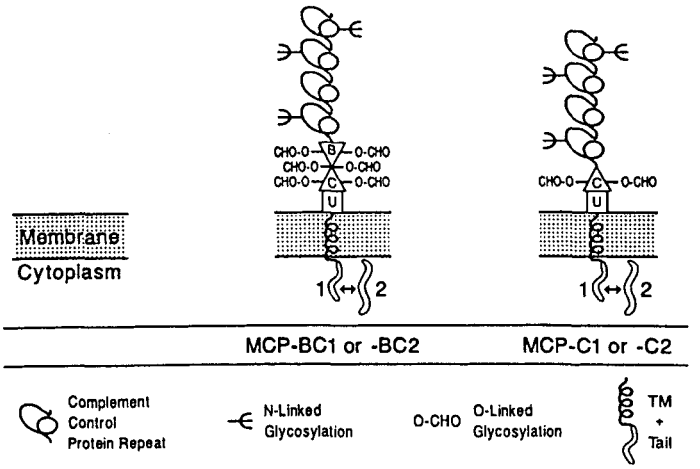
MCP-BC2	Y00651	HSMCP
MCP-BC1(full length 3')	X59405	HSMCP05
MCP-C1	X59406	HSMCP06
MCP-C2	X59407	HSMCP07
MCP-BC2 (full length 3')	X59408	HSMCP08
MCP-ABC2	X59409	HSMCP09
MCP-ABC1	X59410	HSMCP10

Ser/Thr/Pro-Enriched (STP) Domain

B: - VSTSSTTKSPASSAS -
C: - GPRPTYKPPVSNYP -

Cytoplasmic Domain

RYLQRRKKKG → TYLTDETHREVKFTSL (CYT-1)
RYLQRRKKKG → KADGGAEYATYQTKSTTPAEQRG (CYT-2)



Complement Control Protein Repeat
N-Linked Glycosylation
O-CHO O-Linked Glycosylation
TM + Tail

Membrane-type 1 Matrix Metalloproteinase

Yoshifumi Itoh and Motoharu Seiki

Synonyms	MT-MMP; MT-MMP-1; Matrix Metalloproteinase 14 (MMP-14)
Abbreviations	MT1-MMP; MMP-14
Classification	EC 3.4.24.xx (undesignated yet).
Description	MT1-MMP is a member of matrix metalloproteinase (MMP) family (Matrixins). MMPs can be classified into two subgroups. One is for the soluble type enzymes and another is for the membrane type enzymes (MT-MMPs). Up to date, four enzymes including MT1-MMP are reported for the membrane type subgroup and they are characterized by a transmembrane domain at the C-terminus and a conserved basic aa stretch (furin-motif) at the processing site to generate mature enzymes. MT1-MMP is translated as a pre-pro enzyme and activated via processing by furin or related enzymes intracellularly. MT1-MMP expressed on the cell surface is thought to be responsible for the localized degradation of the extracellular matrix at periphery of the cells during cell proliferation, migration and invasion. In addition to the direct activity against the components of extracellular matrix, MT1-MMP acts as a physiological activator of proMMP-2 (pro-gelatinase A) and also activates proMMP-13 (pro-collagenase 3). Thus, MT1-MMP triggers activation of other proMMPs on the cell surface.
Structure	Unknown
Molecular Weight	Based on aa sequence, the zymogen form is 63.5 kDa, and the active form is 53.9 kDa. However, the molecular masses of the latent and activated forms of the enzyme are reported to be 63 and 60 kDa, respectively from the mobility on a SDS-containing polyacrylamide gel.
Sedimentation Coeff.	Unknown
Isoelectric Point	6.97 based on aa composition of the zymogen.
Extinction Coeff.	17.2 (280nm, 1%, 1cm) for zymogen; 18.5 for active form (both calculated).
Enzyme Activity	MT1-MMP introduces a cleavage between the Asn-37-Leu-38 residues in the propeptide of proMMP-2 and subsequently induces autocatalytic activation of proMMP-2. MT1-MMP also cleaves type I collagen generating 3/4 and 1/4 fragments with same specificity as MMP-1 (interstitial collagenase) and MMP-8 (neutrophil collagenase) (cleaves Gly-775-Leu-776 for α1 and Gly-781-Ile-782 for α2 chain), though the specific activity is 1/10 of MMP-1. The k_{cat}/K_m of MT1-MMP against type I collagen is reported to be 2.4 ($\mu M^{-1}h^{-1}$) compared to 17.1 ($\mu M^{-1}h^{-1}$) for MMP-1.
Coenzymes/Cofactors	The enzyme requires Zn^{2+} and Ca^{2+} for the activity and is inhibited by chelating agents. Small amount of tissue inhibitor of metalloproteinases (TIMP-2) is essential for efficient activation of proMMP-2 by MT1-MMP expressing cells. TIMP-2 has dual functions on the cell-mediated activation of proMMP-2. TIMP-2 acts as an inhibitor against MT1-MMP in one hand, TIMP-2 that inhibited MT1-MMP then acts as a binding site for proMMP-2 in the other hand. Thus, TIMP-2/MT1-MMP complex on the cell surface is thought to be a machinery to concentrate proMMP-2 in

	solution to the cell surface and to present the substrate to the neighboring free MT1-MMP.
Substrates	ProMMP-2 and proMMP-13. Collagen I, II, III, gelatin, fibronectin, vitronectin, and laminin-1. As a chromogenic substrate, fluorecence quenching synthetic substrate of Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH_2 can be used to detect enzymic activity.
Inhibitors	Endogenous inhibitor, TIMP-2 and -3 have an inhibitory activity by forming stoichiometric complex but not TIMP-1. Association constant (k_{on}) of these to recombinant MT1-MMP catalytic domain has been reported as 4.2 for TIMP-2 and 3.07 for TIMP-3 compared to ~ 0.000053 ($M \cdot S^{-1} \times 10^{-6}$) for TIMP-1.
Biological Functions	Although the biological function of MT1-MMP has not been well characterized yet, it is thought to be important for cell migration, invasion and proliferation since it is expressed on the cell surface and degrades the extracellular matrix at the cell periphery together with other MMPs activated by MT1-MMP. MT1-MMP was found to localize to pseudopodial structures of invasive tumor cells called as invadopodia where many proteolytic enzymes are reported to localize. Enforced expression of MT1-MMP in several tumor cell lines by transfection of the gene enhances cell's ability to invade reconstituted basement membrane and gelatin film. Transfected tumor cells also showed higher metastatic potential by experimental metastasis assay.
Physiology/Pathology	Expression of MT1-MMP is increased during mouse embryogenesis and decreased according to maturation. Predominant expression was observed in the cells of mesenchymal tissues such as muscle, bone, cartilage and vasculature together with TIMP-2 and MMP-2. These results indicate that MT1-MMP and MMP-2 are important for the formation of the tissues in embryo. Expression is also induced in fibroblasts during rat skin wound healing process suggesting importance of the enzyme in tissue remodeling. In human cancers, both cancer cells and surrounding fibroblasts express MT1-MMP. Although MMP-2 is mainly produced by the fibroblasts in the vicinity of tumors, it bounds to and is activated by MT1-MMP expressing cancer cells. The rate of active form MMP-2 against latent form in cancer tissues well correlated to the expression levels of MT1-MMP and to tumor spread. MT1-MMP was also detected in macrophages infiltrating the tumor tissues and endothelial cells of newly formed vessels in the tissue.
Degradation	Since the soluble recombinant enzyme is easy to degrade by autolysis, the enzyme may be degraded by itself under the concentrated situation on the cell surface. However, conclusive data is not available at moment.
Genetics/Abnormalities	Unknown
Half-life	Unknown
Concentration	Unknown
Isolation Method	When human fibroblasts or cancer cells such as HT-1080 are treated with concanavalin A or phorbol ester, MT1-MMP is expressed on the cell surface of these cells. Membrane-bound MT1-MMP can be extracted from the plasma membrane fraction of these cells using neutral detergent solution such as 1% Triton X-100, NP-40, or polydocanol (Lubrol) in appropriate buffer. Enzyme in the solution can be then isolated using specific antibody immobilized-column. However, some portion of MT1-MMP is likely to present as a complex with TIMP-2 depending on its amount produced from the cells expressing MT1-MMP.

Amino Acid Sequence	MT1-MMPs has characteristic common sequences for all MMPs that are 91-PRCGVPD motif in propeptide whose cystein coordinates with catalytic zinc atom to maintain its latency and catalytic zinc binding motif of 239-HELGHALGLEH in the catalytic domain. In addition, it has a furin cleavage motif of 108-RRKR right before Y-112 where catalytic domain starts. MT1-MMP also has a transmembrane domain and a short cytoplasmic tail at its C-terminus that are AAAVVLPVLL LLLVLAVGLA VFFFRRHGTP RRLLYCQRSL LDKV (underlined for transmembrane domain).
Disulfides/SH-Groups	It has a disulfide bond between C-319 and C-508 forming hemopexin-like domain. Other cysteins are as follows; one in the propeptide as described above, one in the catalytic domain (C-127), and one in the cytoplasmic tail as described above. It is not known whether C-127 is free or forming disulfide bond with other molecule.
General References	Sato, H., Takino, T.,Okada, Y., Cao, J., Shinagawa, A., Yamamoto E. and Seiki, M. *Nature* 1994, **370**:61-65. Seiki, M. *Curr. Topics Microbiol. Immunol.* 1996, **213**:23-30. Coussens, L.M. and Werb, Z. *Chem. Biol.* 1996, **3**:895-904. Nagase, H. *Biol. Chem.* 1997, **378**:151-160.
Ref. for DNA/AA Sequences	Accession Numbers for MT-MMPs. Human MT1-MMP: D26512, X83535. Rat MT1-MMP: X83537. Mouse MT1-MMP: X83536. Human MT2-MMP: Z48482. Human MT3-MMP: D50477, D85511. Human MT4-MMP: X89576.

Multimerin

Catherine P. M. Hayward

Synonyms	p155
Abbreviations	None
Classification	None
Description	A variably-sized, soluble, disulfide-linked homomultimeric protein, synthesized by megakaryocytes and vascular endothelial cells. Multimerin is the major binding protein for coagulation factor V in platelets. Multimerin is also found in vascular endothelium and in the fibrillary extracellular matrix of cultured endothelium. Multimerin was initially designated as p-155, based on its reduced mobility in platelets, however, the term p-155 is no longer in use.
Structure	The tertiary organization and structure of multimerin is not yet known.
Molecular Weight	Multimerin is variable in its nonreduced molecular mass. It is composed of homomultimers that range in size from a 400 kDa trimer to massive multimers, that are approximately 5-10 million daltons (nonreduced, SDS-agarose gels). Its multimers are assembled from a single, 1228 aa precursor protein, prepromultimerin. Removal of the 19 aa signal peptide from prepromultimerin generates promultimerin. The reduced mobility of mature multimerin is variable, due to differences in the extent of proteolytic processing of promultimerin, but the sites of proteolytic cleavage are not yet known. The predominant form of multimerin stored in platelets has a reduced mobility of 155 kDa (SDS-PAGE). Platelets also contain lesser amounts of a 170 kDa subunit and trace amounts of promultimerin (196 kDa). The multimerin constitutively secreted by cultured endothelial cells and megakaryocyte cells is less fully processed and contains mainly promultimerin. Fully glycosylated promultimerin has a reduced mobility of 186-196 kDa (SDS-PAGE), with minor differences in the extent of its N-glycosylation by different cells. N-linked carbohydrate accounts for approximately 33% of its mature, reduced molecular mass (SDS-PAGE).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.4 - 6.75 (reducing conditions, with urea).
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Unknown
Biological Functions	Multimerin is a specific binding protein for coagulation factor V. In resting platelets, but not in plasma, all of the biologically active factor V is complexed with multimerin. When platelets are activated by thrombin, multimerin-factor V complexes dissociate, suggesting multimerin may be a

	carrier protein for factor V during storage. Following secretion, multimerin binds to external membrane of activated platelets and endothelial cells, but its function on the membrane of activated platelets and endothelial cells is not yet known. Multimerin may also function as a fibrillary, extracellular matrix protein in vascular endothelium.
Physiology/Pathology	Although multimerin is a soluble protein, it is not present in detectable amounts in normal plasma. In response to secretagogues, stored multimerin is released from platelets and endothelial cells, and it binds to their external membranes. Small amounts of multimerin are detectable in platelet releasates. The platelet stores of multimerin are abnormal in the Quebec platelet disorder, a congenital bleeding disorder associated with pathological proteolysis of proteins stored in platelet alpha-granules.
Degradation	Unknown
Genetics/Abnormalities	Genetic disease due to defects in the multimerin gene have not been reported.
Half-life	Unknown
Concentration	Multimerin is not detectable in normal plasma. Cultured endothelial cells contain much lower levels of multimerin antigen than platelets (approx. 1% of the platelet level/mg cell protein).
Isolation Method	Monoclonal antibody affinity purification, from the hydrophilic protein fraction of Triton X-114 platelet protein lysates.
Amino Acid Sequence	Multimerin does not belong to any of the recognized protein families. Prepromultimerin contains a 19 aa signal peptide and several domains that may be important for its structure and function. These include an RGDS site (aa 186-189), partial and complete EGF-like domains (aa 269-280, 1065-1076), and a central region that contains probable coiled-coil structures (aa 317-873). Its has sites for tyrosine sulfation (aa 1038), asparagine hydroxylation (aa 1058), and 23 potential N-glycosylation sites. Its C-terminal region (aa 1117-1228) is similar to the C-terminal domains in complement C1q and collagens type VIII and X. The similar C-terminal domains in these other proteins form a trimeric, globular structure important for protein binding.
Disulfides/SH-Groups	Each prepromultimerin subunit contains 21 cysteine residues (aa 211, 238, 245, 271, 272, 280, 315, 474, 592, 599, 765, 931, 984, 985, 1045, 1050, 1056, 1065, 1067, 1076, 1183). The complete and partial EGF-like domains are likely sites for intrachain disulfide bonds.
General References	Hayward, C.P.M. et al. *Blood* 1991, **77**:2556-2560. Hayward, C.P.M. et al. *J. Clin. Invest.* 1993, **91**:2630-2639. Hayward, C.P.M. et al. *J. Biol. Chem.*1995, **270**:18246-18251. Hayward, C.P.M. et al. *J. Biol. Chem.* 1995, **270**:19217-19224. Hayward, C.P.M. *Clin. Invest. Med.* 1997, **20**:176-187. Hayward, C.P.M. et al. *Blood* 1998, **91**:1304-1317.
Ref. for DNA/AA Sequences	Hayward, C.P.M. et al. *J. Biol. Chem.* 1995, **270**:18246-18251. GenBank/EMBL databank accession number U27109.

Myeloperoxidase

William M. Nauseef

Synonyms	Verdoperoxidase
Abbreviations	MPO
Classifications	Donor: H_2O_2 oxidoreductase; EC 1.11.1.7
Description	MPO is a iron-containing protein located in the azurophilic granules of phagocytic cells of the myeloid lineage, predominantly human neutrophils (PMNs) and monocytes and heterophils isolated from a variety of non-human species (except goose and chicken which are MPO-deficient). MPO is structurally distinct from eosinophil peroxidase, another peroxidase present in cirulating myeloid cells.
Structure	MPO consists of two large subunits (Mr 59,000) and two small subunits (Mr 13,500) and behaves as a symmetric molecule, linked by a single disulfide bridge between the two heavy chains. Cleavage of this disulfide bond in native MPO results in the generation of hemi-MPO with the same specific activity as that of the native enzyme. Each molecule of MPO has two tightly bound calcium ions and two iron molecules. The heme in MPO is a novel derivative of protoporphyrin IX wherein substituents on three of the rings form covalent bonds with aa in the protein (Glu-408, Met-409 and Asp-260). A 2.28 Å X-ray crystal structure of human MPO has been published and the properties of the heme group better defined.
Molecular Weight	≈ 150,000 (native MPO), composed of a pair of identical (or nearly identical) heavy-light protomers.
Sedimentation Coeff.	8 S (sedimentation), axial ratio %:1
Isoelectric Point	> 10 (estimation)
Extinction Coeff.	75 ($nM^{-1}\,cm^{-1}$) at 472nm. In reduced MPO the Soret band is at 475nm and a distinct band appears at 637nm. Pure MPO has A_{430}/A_{280} of 0.81-0.83 at 1 mg/ml.
Enzyme Activity	The most important physiologic activity of MPO is the catalysis of chloride oxidation by H_2O_2 to generate the highly reactive hypochlorous acid. In vivo, HOCl participates in a wide variety of nonenzymatic reactions, including the generation of chloramines from NH_3 and amines.
Coenzymes/Cofactors	In the presence of halide ions ($I^- > Br^- > Cl^-$) or the pseudohalide SCN^- and H_2O_2, MPO catalyzes the production of HOCl. At high H_2O_2 concentrations, MPO acts catalatically to decompose H_2O_2 and is inactivated.
Substrates	Physiologically, the most important substrates are Cl^- and H_2O_2, as noted above. The MPO-H_2O_2-halide system interacts with a wide variety of mammalian cells, multicellular organisms, humoral substances, PMN granule and other proteins.
Inhibitors	The activity of MPO can be inhibited by azide, cyanide, and hydroxylamine. In addition, excess H_2O_2 can inhibit MPO activity, since the enzyme acts catalatically under these conditions and is concomitantly

inactivated. The microbicidal activity of normal human PMNs (but not MPO-deficient cells) is inhibited by azide, cyanide, and sulfonamides.

Biological Functions

Optimal microbicidal activity of human PMNs depends on two events which occur nearly simultaneously: the generation of reactive species from molecular oxygen (a process known as the respiratory burst) and e release of granule products into the phagolysosome (a process known as degranulation). These two features of PMN activation result in the production of an extremely hostile milieu for the ingested microorganism or tumor cell. MPO, a major component of the azurophilic granules in PMNs, is released into the phagolysosome during degranulation. In the presence of the H_2O_2 generated by the respiratory burst, the MPO released by degranulation catalyzes the oxidation of Cl^- to form HOCl and indirectly generate various chloramines which are microbicidal and tumoricidal. The precise mechanism(s) responsible for the lethal event is(are) not known, but may include cleavage of critical peptide bonds in the target bacteria, oxidative decarboxylation and deamination of important bacterial proteins, generation of toxic aldehydes, or the oxidation of components of the bacterial respiratory chain.

Physiology/Pathology

The major physiologic function of MPO is in optimal activity of the oxygen-dependent microbicidal system of human PMNs best illustrated by abnormalities in MPO-deficient PMNs. When stimulated, MPO-deficient PMNs generate greater amounts of superoxide and hydrogen peroxide than do normal PMNs. However, the rate of bactericidal activity is retarded, such that the MPO-deficient PMN may take 60 minutes to achieve the same degree of microbicidal activity as effected by normal PMNs in 20 minutes. MPO-deficient PMNs are unable to kill a variety of fungi in vitro. This defect is most striking for the clinically significant species of Candida. Clinically significant sequelae of MPO deficiency have been limited to frequent or severe candidal infections in MPO-deficient individuals who were also diabetics.

Degradation

MPO is inactivated by high concentrations of H_2O_2 and may undergo proteolysis by the other proteases released into the phagolysosome during agonist-elicited degranulation.

Genetics/Abnormalities

MPO is synthesized only during the promyelocytic stage of myeloid maturation; circulating PMNs have no detectable mRNA for MPO. The gene is organized into 12 exons and 11 introns and has been localized to 17q22-q23 in humans. Hereditary MPO deficiency is common (1 in 4'000) and it is likely that many genotypes exist. Although, most patients have immunochemical evidence of MPO precursors in their cells, suggesting a post-translational defect, evidence supports the existence as well of pre-translational defects. A missense mutation at nucleotide 8.089 in the MPO gene is a relatively common cause of MPO deficiency and this results in substitution of W for R at codon 569. Transfected cells expressing MPO with the R569W mutation demonstrate an arrest in the normal biosynthetic maturation of enzymatically active MPO.

Half-life

Unknown

Concentration

Nearly all of MPO is localized to the azurophilic granule of PMNs. There is a small amount of immunoreactive material in serum, not all of which is enzymatically active. The biological significance of this circulating material is not known. Estimates of the amount of MPO in human PMNs range from 1 - 2 % to greater than 5% of the dry weight of the cell.

Isolation Method	Numerous methods for the isolation of MPO have been described, using canine pyometrial pus, bovine spleen, circulating myeloid cells from patients with chronic myelogenous leukemia, and abnormal peripheral blood neutrophils as a source for the enzyme. Unbroken cells and nuclei are removed (1800 x g, 10 min, 4°C) and the supernatant centrifuged (12,000 x g, 30 min, 4°C) to sediment granules. Granules are freeze-thawed and resuspended in 20 mM K-phosphate (pH 7.4) containing 0.5% cetyltrimethylammonium bromide and sonicated on ice. Insoluble material is removed by centrifugation and the solublized granule proteins are dialyzed overnight at 4°C against 20 mM K-phosphate (pH 7.4). The white precipitate which forms during dialysis is removed by centrifugation and the bright green supernatant is lyophilized. The lyophilized material is disolved in 0.2 - 0.5 ml of column buffer (130 mM NaCl, 20 mM sodium acetate, 0.5% 2-mercaptoethanol, pH 5.0) and applied to a Sephadex G-150 column and eluted with column buffer. Green fractions are assayed for purity by determining the A_{430}/A_{280}. Fractions with the highest ratio are pooled, lyophilized, and applied to the same column for further purification.
Amino Acid Sequence	Mature MPO contains four N-linked oligosaccharide side chains, all high mannose groups and all on the heavy subunit. MPO has sequence homology with eosinophil peroxidase, thyroid peroxidase and lactoperoxidase (LPO). The ester linkages to the heme may be shared by these homologous peroxidases. It is likely that MPO, EPO, and TPO are related members of the same gene family. In the region near proximal histidines thought to be important for linkup of iron in the enzymes, the homology between MPO and TPO is 74%. Three complete predicted aa sequences for preproMPO have been published and predict an 83,750 Da proteicontaining 745 aa. The putative signal sequence is followed by a 125 aa pro region of unknown significance. Both cDNA and genomic sequence for human and murine MPO have been published.
Disulfides/SH-Groups	Native MPO contains an intermolecular disulfide bond which joins the two heavy subunits.
General References	Klebanoff, S.J. Myeloperoxidase: occurrence and biological function. In: *Peroxidases in Chemistry and Biology.* **Vol.1**, Everse, J. et al. (eds.) 1991, pp. 1-36. Hurst, J.K. Myeloperoxidase: Active site structure and catalytic mechanisms. In: *Peroxidases in Chemistry and Biology.* **Vol.1**, Everse, J. et al. (eds.) 1991, pp. 37-62. Johnson, K.R. and Nauseef, W.M. Molecular biology of myeloperoxidase. In: *Peroxidases in Chemistry and Biology.* **Vol.1**, Everse, J. et al. (eds.) 1991, pp.63-82. Forehand, R., Nauseef, W.M., Curnutte, J.T., Johnston, R.B. Inherited disorders of phagocytic killing. In: *The Metabolic and Molecular Basis of Inherited Disease.* Scriver, C.R., et al. (eds.) 7th Edition, **Vol. III**, Chapter 133, 1995 pp.3995-4026. Nauseef, W.M., Brigham, S., Cogley, M. Hereditary myeloperoxidase deficiency due to a missense mutation of arginine 569 to tryptophan. *J. Biol. Chem.* 1994, **269**:1212-1216. Nauseef, W.M., McCormick, S., Cogley, M. Effect of the R569W missense mutation on the biosynthesis of myeloperoxidase. *J. Biol. Chem.* 1996, **271**:9546-9549. Fenna, R., Zeng, J., Davey, C. Structure of the green heme in myeloperoxidase. *Arch. Biochem. Biophys.* 1995, **316**:653-656.

Taylor, K.L., Strobel, F., Yue, K.T., et al. Isolation and identification of a protoheme IX derivative released during autolytic cleavage of human myeloperoxidase. *Arch. Biochem. Biophys.* 1995, **316**:635-642.

Ref. for DNA/AA Sequences

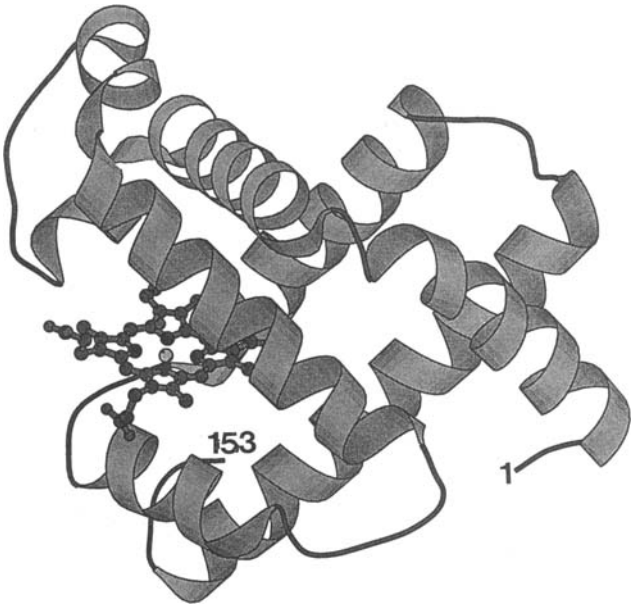
Johnson, K.R., et al. Isolation of cDNA clones for human myeloperoxidase: predicted amino acid sequence and evidence for two mRNA species. *Nucleic Acids Res.* 1987, **15**:2013-2028.
Morishita, K., et al. Molecular cloning and characterization of cDNA for human myeloperoxidase. *J. Biol. Chem.* 1987, **262**:3844-3851.
Johnson, K.R., et al. Complete nucleotide sequence of the human myeloperoxidase gene. *Nucleic Acids Res.* 1989, **17**:7985-7986.
EMBL accession number X15377.

Myoglobin

Chandramowli Ganesh and Raghavan Varadarajan

Synonyms	None
Abbreviations	HuMb
Classification	Heme protein
Description	An oxygen storage protein found primarily in muscle tissue. Mb is a single polypeptide, 153 aa long. Contains a non-covalently associated heme prosthetic group. Member of globin superfamily.
Structure	Crystal structure determined by X-ray crystallography, of a mutant (K45R, C110A) at 2.8 Å. All α-helical protein with 8 helices, A-H. The heme iron is coordinated to histidine residue 93, located at 8th position in helix F. Oxygen binds reversibly to the ferrous iron in heme, without changing the oxidation status of the bound ferrous atom.
Molecular Weight	17.7 kDa
Sedimentation Coeff.	1.815 S; diffusion coefficient ($D_{20,w}$) 9.6x10^{-7} cm^{2}sec^{-1} and partial specific volume 0.743 ml g^{-1}
Isoelectric Point	7.2 (*metaquo* form). Redox potential: 58.9 mV ($E^{o'}$, pH 7, 25°C)
Extinction Coeff.	160 mM^{-1}cm^{-1} (409nm, heme absorption)
Enzyme Activity	None. Binds oxygen reversibly.
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Carbon monoxide, cyanide
Biological Functions	Oxygen storage in muscle tissue. Myoglobin binds oxygen with a higher affinity than Hemoglobin which is the oxygen carrier protein in red blood cells. Myoglobin binds oxygen in the tissue decreasing the free oxygen concentration. This leads to diffusive transport of oxygen from blood capilliares to the surrounding tissue.
Physiology/Pathology	Myoglobinuria is observed in the cases of rhabdomyoloses, lactate transporter defect or deficiency of phosphoglycerate mutase. These are characterized by muscle degeneration and subsequent excretion of myoglobin.
Degradation	Resistant to proteolytic cleavage by Trypsin and Factor X_a. Resistant to alkali.
Genetics/Abnormalities	The gene for Mb is present in the q arm of chromosome 22 and the gene map locus of the gene is 22q11.2-q 13. The gene has been cloned and sequenced. 3 exons coding 31, 74 and 49 aa are present, interspersed with intron sequences and Alu repeats. Naturally occuring variants are E54K, K133N, R139Q, R139W.

Half-life	80-90 days, as measured in skeletal muscles of rats.
Concentration	6 mg per gram tissue (limb muscles)
Isolation Methods	E. coli system available. HuMb expressed as a fusion product with a fragment of phage λ cII protein. Cells lysed and fusion protein containing inclusion bodies solubilized in 8M urea. Solubilized denatured protein dialyzed to remove the denaturant and reconstituted with required amount of heme. Product subjected to Trypsin cleavage to remove the fusion partner. HuMb further purified on DE-52 and S-200 chromatography columns.
Amino Acid Sequence	GLSDGEWQLV LNVWGKVEAD IPGHGQEVLI RLFKGHPETL EKFDKFKHLK SEDEMKASED LKKHGATVLT ALGGILKKKG HHEAEIKPLA QSHATKHKIP VKYLEFISEC IIQVLQSKHP GDFGADAQGA MNKALELFRK DMASNYKELG FQG
Disulfides/SH-Groups	No disulfides. Single, free sulfhydryl at residue 110.
General References	Herrera, A.E. and Lehmann, H. *Nature New Biol.* 1971, **232**: 149-152. Weller, P. et al. *EMBO J.* 1984, **3**: 439-446. Akaboshi, E. *Gene* 1985, **33**: 241-249. Varadarajan, R. et al. *Proc. Natl. Acad. Sci.USA* 1985 **82**: 5681-5684. Hubbard, S.R. et al. *J. Mol. Biol.* 1990, **213**: 215-218.



Human Myoglobin structure, taken from the crystal structure of the aquomet form, PDB ID 2MM1.

N-Acetylgalactosamine-4-Sulphatase

Julie Bielicki and John J. Hopwood

Synonyms	Arylsulphatase B, N-Acetylgalactosamine 4-sulphate sulphatase, 4-sulphate sulphatase, 4-sulpho-N-acetylgalactosamine sulphatase
Abbreviations	ASB, 4-S, 4-sulphatase
Classifications	EC 3.1.6.1
Description	ASB is a lysosomal hydrolase which is synthesized in the endoplasmic reticulum as a 533 aa precursor and subsequently modified by removal of a 39 aa signal peptide which results in a 66 kDa single polypeptide subunit precursor. This is further modified by proteolysis to produce the mature form of the enzyme which comprises three disulphide-linked polypeptides of 43 kDa, 7 kDa and 8 kDa. In total there are 6 potential asparagine-linked glycosylation sites, 4 on the 43 kDa polypeptide and 2 on the 7 kDa polypeptide.
Structure	Unknown
Molecular Weight	Native M_r of human liver mature ASB is 58 kDa and that of the precursor recombinant ASB is 66 kDa.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	ASB catalyses the hydrolysis of the C4-sulphate ester bond of N-acetylgalactosamine 4-sulphate residues at the non-reducing termini of the glycosaminoglycans dermatan sulphate (DS) and chondroitin-4-sulphate (CS).
Coenzymes/Cofactors	Unknown
Substrates	Biological substrates are DS and C4S. Diagnostic substrate derived from C4S is 0-(β-N-acetylgalactosamine 4-sulphate)-(1 → 4)-0-D-(β-glucuronic acid)-(1 → 3)-0-D-N-acetyl [1-^{3}H] galactosaminitol 4-sulphate (K_m is 5.7 µM, pH opt. 3.5). Flurogenic substrate is 4- methylumbelliferyl sulphate (K_m is 4.8 mM, pH opt. 5.3).
Inhibitors	Unknown
Biological Functions	ASB is one of several highly specific exo-enzyme activities involved in the degradation of the glycosaminoglycans DS and C4S.
Physiology/Pathology	A deficiency of ASB is the cause of the lysosomal storage disease mucopolysaccharidosis type VI (MPS VI) otherwise known as Maroteaux-Lamy syndrome. Patients with MPS VI syndrome store and excrete abnormal amounts of DS. Depending on the severity of the disease, a variety of clinical phenotypes is observed. Symptoms of MPS VI include growth retardation, corneal clouding, skeletal deformities and hepatosplenomegaly. However, even in the severe form of the disease, neurological development is generally normal.

Degradation	Unknown
Genetics/Abnormalities	ASB maps to chromosome 5q13-q14. There is no common mutation which causes MPS VI syndrome and this is consistent with the broad spectrum of clinical phenotypes that is observed.
Half-life	9–17 days for endogenous ASB in normal skin fibroblasts 3.7 and 3.3 days for endocytosed recombinant ASB and ASB immunopurified from skin fibroblasts respectively.
Concentration	1 mg/kg of human liver.
Isolation Method	1. 5 step, 6 column procedure Step 1: Concanavalin-A-Sepharose-Blue A-Agarose Step 2: Blue A-Agarose Step 3: Hydroxyapatite Step 4: Blue B-Agarose Step 5: Cu^{2+}-chelating Sepharose 2. One step immunopurification with Mab 4.1 linked Affigel.
Amino Acid Sequence	ASB has strong sequence homology with other arylsulphatases, namely sea urchin arylsulphatase and human arylsulphatase A and C. It also has sequence homology with non-arylsulphatases such as human iduronate-2-sulphatase, human glucosamine-6-sulphatase and human N-acetylgalactosamine-6-sulphatase.
Disulfides/SH-Groups	The 3 polypeptides which comprise the mature form of ASB are disulphide-linked.
General References	Hopwood, J. J., et al. *Biochem. J.* 1986, **234**: 507–514. Gisbon, G. J., et al. *Biochem. J.* 1987, **248**: 755–764. Taylor, J. A., et al. *Biochem. J.* 1990, **268**: 379–386. Anson, D. S., et al. *Biochem. J.* 1992, **284**: 789–794. Kobayashi, T., et al. *Biochim. Biophys. Acta* 1992, **1159**: 243–247.
Ref. for DNA/AA Sequences	Peters, C., et al. *J. Biol. Chem.* 1990, **265**: 3374–3381. Schuchman, E. H., et al. *Genomics* 1990, **6**: 149–158. Litjens, T., et al. *Biochem. Int.* 1991, **24**: 209–215.

N-Acetylgalactosamine-6-Sulphatase

Julie Bielicki and John J. Hopwood

Synonyms	N-Acetylgalactosamine-6-sulphate sulphatase, Galactose-6-sulphate sulphatase, Galactose-6-sulphatase
Abbreviations	G6S, GalNAc-6-S, Gal-6-S, GALNS
Classifications	EC 3.1.6.4
Description	G6S is a lysosomal exohydrolase which is synthesized in the endoplasmic reticulum as a 522 aa precursor and subsequently modified by removal of a 26 aa N-terminal signal peptide. The mature form, purified from liver, comprises two disulphide-linked polypeptides of 39 kDa and 19 kDa and has two potential asparagine-linked glycosylation sites.
Structure	Unknown
Molecular Weight	Precursor form 58 kDa (SDS-PAGE). Mature form from liver, 55–58 kDa, made up of two disulphide-linked polypeptides of 39 kDa and 19 kDa. Recombinant human precursor G6S 58–60 kDa (SDS-PAGE and gel filtration).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.7 (chromatofocussing)
Extinction Coeff.	Unknown
Enzyme Activity	G6S catalyzes the hydrolysis of the C6-sulphate ester bond present at the non-reducing terminal residue of chondroitin-6-sulphate (C6S) oligosaccharides, namely N-acetylgalactosamine-6-sulphate. It also acts on galactose-6-sulphate linkages in keratan sulphate (KS).
Coenzymes/Cofactors	Unknown
Substrates	The biological substrates are C6S and KS. The diagnostic substrates are: 1. the trisaccharide N-acetylgalactosamine-6-sulphate-glucuronic acid-N-acetylgalactosaminitol-6-sulphate derived from C6S (K_m is 12 µM and pH opt 3.5 for liver enzyme; K_m is 15 µM and pH opt 4.0 for recombinant enzyme). 2. the disaccharide substrate galactose-6-sulphate-anhydrominnitol-6-sulphate from KS (K_m is 50 µM, pH opt 4.0 for liver enzyme and 96 µM and 4.0 for recombinant enzyme).
Inhibitors	Concentration of substances which give 50% inhibition of liver and recombinant G6S activity towards the trisaccharide substrate.

	Liver G6S	Recombinant G6S
NaCl (mM)	24	48
Na_2SO_4 (µM)	50	40
Na_2HPO_4 (µM)	325	450
Cu acetate (µM)	not determined	78

Biological Functions	G6S is one of the lysosomal enzymes involved in the degradation of the glycosaminoglycans KS and C6S.

Physiology/Pathology	A deficiency of G6S results in the lysosomal storage disorder mucopolysaccharidosis type IV A (Morquio syndrome A) which is inherited as as autosomal recessive trait. The clinical presentation of this disorder is characterized by normal intelligence, dwarfism, corneal clouding, spondyloepiphysial dysplasia and dental abnormalities. There is storage of partially degraded KS and C6S fragments in tissues, as well as excessive urinary excretion of these fragments. As with other MPS disorders, phenotypic variability exists, consistent with the occurrence of different mutant alleles at the G6S locus.
Degradation	Unknown
Genetics/Abnormalities	G6S maps to chromosome 16q24.3 Mutations are predicted to be heterozygous based on the broad spectrum of clinical phenotypes that are observed. There is no common mutation to date two single-base alterations, one two base pair deletion and several gross structural gene rearrangements have been identified.
Half-life	Unknown
Concentration	510 pmole/hr per mg (liver)
Isolation Method	Mature form: 7 step, 8 column procedure. Step 1: Concanavalin A Sepharose-Blue A-Agarose Step 2: Chromatofocussing (PBE94) Step 3: TSK HW 50S Fractogel Step 4: Cu^{2+}-chelation Sepharose Step 5: Phenyl Sepharose Step 6: TSK G-3000SW Step 7: Cu^{2+}-chelation Sepharose Precursor form: 2 column precedure Step 1: DEAE-Sephacel Step 2: CM-Sepharose
Amino Acid Sequence	Human G6S has strong sequence homology with other sulphatases, e.g. human iduronate-2-sulphatase, human glucosamine-6-sulphatase, sea urchin arylsulphatase and human arylsulphatase A, B and C.
Disulfides/SH-Groups	There are several cysteine residues in G6S. The two polypeptide subunits in the mature form are disulphide-linked.
General References	Bielicki, J. and Hopwood, J. J. *Biochem.* J. 1991, **279**: 515–520. Masue, M. et al. *J. Biochem.* 1991, **110**: 965–970. Bielicki, J. et al. *Biochem. J.* 1995, submitted.
Ref. for DNA/AA Sequences	Tomatsu, S. et al. *Biochem. Biophys. Res. Commun.* 1991, **181**: 677–683. Morris, C. P. et al. *Genomics* 1994, **22**:652–654.

Nerve growth factor

Gerhard Heinrich

Synonyms	β-Nerve growth factor
Abbreviations	NGF; β-NGF; 2.5S NGF
Classifications	None
Description	Synthesized from a precursor and secreted constitutively by fibroblasts, Schwann cells, and certain central neurons and glia. Secreted via a storage-type regulated pathway by granulated convoluted tubular cells of mouse salivary gland. Forms stable, non-covalent homodimers (Kd ranges from 10^{-9} M to 10^{-12} M). NGF isolated from mouse salivary gland may be heterogeneous due to proteolytic removal of an N-terminal octapeptide and/or a C-terminal arginine during isolation, and glycosylation (about 15%). Truncated forms dimerize and are biologically active as are glycosylated and dimerized NGF. The NGF dimer forms a stable complex with homodimers of an enzymatically active (γ-NGF) and homodimers of an enzymatically inactive kallikrein (α-NGF). The hexameric complex ($\alpha_2\beta_2\gamma_2$) is known as 7S NGF and may exist exclusively in mouse salivary gland.
Structure	Globular protein. Crystals are bipyramidal hexagonal with a hexagonal unit cell, space group $P6_122$. There are six NGF dimers/cell, and each dimer has a 2-fold axis of symmetry. Crystal structure has been solved.
Molecular Weight	29,000 (equilibrium sedimentation); 16,200 (equilibrium sedimentation in 6 M GuHCl). Recombinant human NGF: 13,000 (gel permeation chromatography).
Sedimentation Coeff.	2.5 S
Isoelectric Point	9.3
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Required for the survival of sympathetic and sensory neurons perinatally in rodents. Supported by gene knock-out experiments. Regulates neurotransmitter synthesis throughout life. Stimulates choline acetyltransferase activity in basal forebrain cholinergic neurons. Possible role in spermatogenesis and immune system. Enhances memory function in aged rats. Recombinant human NGF is under clinical evaluation for prevention and treatment of peripheral neuropathies and neurodegenerative diseases. Biological functions are mediated by high affinity receptors.
Physiology/Pathology	Present at exceedingly low levels in targets of NGF-responsive neurons (< 10 ng g^{-1} wet weight). Found in human semen (2 ng mL^{-1}). Stored and secreted by mouse salivary gland (1.5 mg g^{-1} wet weight, adult male), bull

seminal vesicle, guinea pig prostate and snake venom gland. Production is regulated developmentally, by steroid hormones, phorbol ester, cAMP, and interleukin-1. Production increases rapidly in the sciatic nerve upon transfection, in the iris upon explantation, and in the hippocampus after seizures.

Degradation

Removed by receptor-mediated uptake into neurons followed by retrograde axonal transport and intracellular degradation. Mode of disposal of NGF secreted via regulated pathway is unknown.

Genetics/Abnormalities

Unique gene on human chromosome 1 (1p22) and mouse chromosome 3. The transcription unit spans >60 kb and has three promoters resulting in three distinct 5' leading exons, 1A, 1B and 3A. There are a total of six exons. The largest message has 4 exons. All mature transcripts are about 1.3 kb long. Exon 4 is common to all transcripts and encodes a complete precursor of 243 aa. In the mouse, exon 2 is alternatively spliced in/out and has an inframe initiation codon that generates a larger precursor of 309 aa.. An additional 2.4 kb mRNA is present in most rat tissues. The relationship to the shorter alternate forms of 1.3 kb is unknown.

Half-life

Unknown

Concentration

Undetectable in human blood. Transiently increases in the blood of some mouse strains after fighting (peak levels 300 ng mL^{-1} serum).

Isolation Method

Isolated from adult male mouse submandibular gland by sequential homogenization in distilled water, gel filtration of the aqueous extract, acidification and cation exchange chromatography on CM-cellulose. Preparative amounts of human recombinant NGF, BDNF, NT-3 and NT-4 are obtained by bacterial expression followed by denaturation and refolding in vitro. Refolded neurotrophins are purified by methods similar to isolation of NGF from mouse salivary gland.

Amino Acid Sequence

The human NGF (H-NGF) sequence deduced from the human gene is shown and compared with the mouse (M) NGF sequence. Hippocampus-derived neurotrophic factor (HDNF) is a neurotrophic factor closely related to NGF. Mouse (M) and rat (R) HDNF sequences are identical. Brain-derived neurotrophic factor from mouse (M-BDNF) is a third NGF-related protein. Dashes indicate residues identical to those in human (H) NGF. Stars denote gaps introduced to maximize similarity. The C-terminal dipeptide is removed from mouse NGF during processing of the precursor. The C-terminus of H-NGF is unknown.

```
H-NGF:      SSSHPIFHRG  EFSVCDSVSV  WV**GDKTTATD  IKGKEVMVLG
M-NGF:       --T--V--M-  ----------  --**--------  ------T--A
M/R-HDNF:   -YAEHKS---  -Y-----E-L  --**T--SS-I-  -R-HQ-T---
M-BDNF:     --HSDPAR--  -L-----I-E  --TAA--K--V-  MS-GT-T--E

H-NGF:      EVNINNSVFK  QYFFETKCRD  PNPVDSGCRG  IDSKHWNSYC
M-NGF:      ---------R  ---------A  S---E-----  ----------
M/R-HDNF:   -IKTG--PV-  ---Y--R-KE  AR--KN----  --D-----Q-
M-BDNF:     K-PVSKGQL-  ---Y----NP  MGYTKE----  --KR----Q-

H-NGF:      TTTHTFVKAL  TMDG*KQAAWR  FIRIDTACVC  VLSRKAVRrg
M-NGF:      ----------  -T-E*------  ----------  ------T--a
M/R-HDNF:   K-SQ-Y-R--  -SENN-LVG--  W-----S---  AL---IG--t
BDNF:       R--QSY-R--  ---SK-RIG--  ------S---  T-TI-**RGR
```

Disulfides/SH-Groups	3 disulfides, no free sulfhydryls.
General References	Server, A.C. and Shooter, E.M. Nerve growth factor. *Ann. Rev. Prot. Chem.* 1977, **31**:339-409. Levi-Montalcini, R. The nerve growth factor 35 years later. *Science* 1987, **237**:1154-1162. Thoenen, H. Nerve growth factor: cellular localization and regulation of synthesis. *Cell. Mol. Neurobiol.* 1988, **8**:35-40. Hefti, F. Function of neurotrophic factors in the adult and aging brain and their possible use in the treatment of neurodegenerative diseases. *Neurobiol. Aging* 1989, **10**:515-533. McDonald, N.Q., et al. New protein fold revealed by a 2.3 Å resolution crystal structure of nerve growth factor. *Nature* 1991, **354**:411-415. Snider, W.D. Functions of the neurotrophins during nervous system development: what the knockouts are teaching us. *Cell* 1994, **77**:627-638. Thoenen, H. Neurotrophins and neuronal plasticity. *Science* 1995, **270**: 593-598. Ip, N.Y. and Yancopoulos, G.D. Neurotrophic factors and their receptors. *Progress in Brain Research* 1995, **105**:189-195.
Ref. for DNA/AA Sequences	Selby, M.J. et al. Mouse nerve growth factor gene: structure and expression. *Mol. Cell. Biol.* 1987, **7**:3057-3064. Zheng, M. and Heinrich, G. Structural and functional analysis of the promotor region of the nerve growth factor gene. *Mol. Brain Res.* 1988, **3**: 133-140. Hohn, A. et al. Identification and characterization of a novel member of the nerve growth factor/brain-derived neutrophic factor family. *Nature* 1990, **344**:339-341. Human and mouse DNA sequences have been deposited in Genbank and EMBO.

Neutrophil Collagenase

Michael Pieper and Harald Tschesche

Synonyms	Granulocyte collagenase; PMNL-collagenase; Matrix metalloproteinase-8
Abbreviations	HNC; CLG1; MMP-8
Classification	EC 3.4.24.34
Description	Neutrophil collagenase is stored intracellular as a latent proenzyme in the specific granules of polymorphonuclear leukocytes. Its single polypeptide chain is organized in four domains: A signal peptide, not present in the mature proenzyme, a prodomain containing a sequence motif responsible for maintaining the enzyme in the latent state, a catalytic domain and a hemopexin-like domain likely responsible for the capability to cleave native triple helical collagens type I, II and III.
Structure	No three-dimensional structure of the full-length enzyme is available so far. However, the recombinant catalytic domain has been cocrystallized with several substrate analogue inhibitors. The structure displays a spherical molecule divided by the active site cleft into an 'upper' main body and a smaller 'lower' part, Fig 1. The former consists mainly of a central, highly twisted five-stranded β-pleated sheet and two long α-helices, whereas the 'lower' part exhibits multiple turns and finally ends in a long α-helix.
Molecular Weight	85,000 Da (~ 40% carbohydrate); by SDS-PAGE
Sedimentation Coeff.	Unknown
Isoelectric Point	6.05 (calculated)
Extinction Coeff.	76,360 $M^{-1}cm^{-1}$ (calculated)
Enzyme Activity	Specific activity against soluble type I collagen: 3,300 units/mg ($HgCl_2$-activated); 11,000 units/mg (Stromelysin-activated).
Coenzymes/Cofactors	The catalytic activity depends on a zinc ion at the active site. This metal polarizes a water molecule, which then is capable of attacking the carbonyl group of the scissile bond. A second zinc ion coordinated in a similar manner has only structural significance. Two calcium ions within the catalytic domain and likely another one within the hemopexin-like domain fulfil also structural functions.
Substrates	The enzyme degrades components of the extracellular matrix such as collagens type I, II and III, gelatin, fibronectin, proteoglycans and aggrecan. It also cleaves and inactivates serpins. In vitro the synthetic fluorogenic peptide substrate Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH_2 is valuable to determine the enzyme activity.
Inhibitors	Enzymatic activity is inhibited specifically by the family of TIMPs (tissue inhibitors of metalloproteinases) and nonspecifically by α_2-macroglobulin. Besides tetracyclins, especially doxycyclins, and chelating agents like 1,10-phenanthrolin, a broad spectrum of small synthetic inhibitors were developed. They are based on short substrate analogue peptide sequences linked to chelating moieties such as hydroxamate, thiol, carboxylate or

phosphinic groups. Marimastat and $HSCH_2CH(CH_2Ph)CO$-L-Ala-Gly-NH_2 are two examples.

Biological Function

Human polymorphonuclear leukocytes (PMNLs) play a key role in the defence system. Their main function is to protect the body against invading microorganisms by phagocytosis. During the chemotactically induced process of migrating to the site of infection the cells must overcome barriers and obstacles build up by the extracellular matrix. PMNLs have the ability to enzymatically degrade this matrix in a locally limited manner. Since neutrophil collagenase is the only PMNL-enzyme capable of degrading collagens type I-III, it is, amongst others, hold responsible for the cell movement through tissue boundaries.

Physiology/Pathology

The enyme is important for PMNL function. Dysregulation of neutrophil collagenase activity leads to an excessive degradation of the extracellular matrix. The enzyme participitates in pathological processes mediated by PMNLs e.g. inflammation, rheumatoide arthritis or osteoarthritis.

Degradation

The enzyme rapidly cleaves off its hemopexin-like domain by autoproteolysis, which is accompanied by the loss of tripelhelical cleavage activity against collagens type I-III.

Genetics/Abnormalities

The gene for human neutrophil collagenase is located on the long arm of chromosom 11 and mapped to 11q21-q22 by in situ hybridization. Four variants have been reported, all of them function normally.

Half-life

Unknown

Concentration

Plasma: 24.5µg/L; Serum: 70.6 µg/L; Normal tissue:29.5µg/g protein; Coloncarcinoma: 78.0 µg/g protein; Adenocarcinoma: 82.9µg/g protein

Isolation Method

Human neutrophil collagenase was isolated from leukocytes obtained from the buffy coat layer of citrated blood. After homogenization and centrifugation the crude enzyme preparation was applied to an affinity chromatography on zinc chelate Sepharose. This was followed by ion exchange chromatography on Q-Sepharose fast flow, affinity chromatography on orange Sepharose and finally a gel-permeation step on Sephacryl S-300.

Amino Acid Sequence

MFSLKTLPFL LLLHVQISKA FPVSSKEKNT KTVQDYLEKF
YQLPSNQYQS TRKNGTNVIV EKLKEMQRFF GLNVTGKPNE
ETLDMMKK**PR** **CGVPD**SGGFM LTPGNPKWER TNLTYRIRNY
TPQLSEAEVE RAIKDAFELW SVASPLIFTR ISQGEADINI
AFYQRDHGDN SPFDGPNGIL AHAFQPGQGI GGDAHFDAEE
TWTNTSANYN LFLVAA**HEFG** **HSLGLAH**SSD PG**ALMY**PNYA
FRETSNYSLP QDDIDGIQAI YGLSSNPIQP TGPSTPKPCD
PSLTFDAITT LRGEILFFKD RYFWRRHPQL QRVEMNFISL
FWPSLPTGIQ AAYEDFDRDL IFLFKGNQYW ALSGYDILQG
YPKDISNYGF PSSVQAIDAA VFYRSKTYFF VNDQFWRYDN
QRQFMEPGYP KSISGAFPGI ESKVDAVFQQ EHFFHVFSGP
RYYAFDLIAQ RVTRVARGNK WLNCRYG

Disulfides/SH-Groups

One disulfide bond connects the beginning of the hemopexin-like domain with its end (C-259-C-464). In the latent state the catalytic zinc ion is complexed by a free cysteine residue at position 71, which is cleaved off during activation.

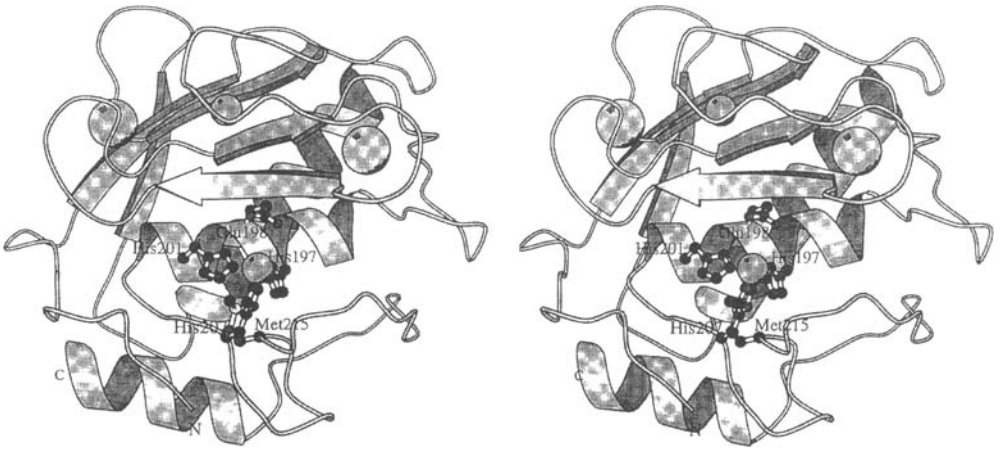
General References

Bode, W. et al. *EMBO J.* 1994, **13**:1263-1269.
Nagase, H. *Biol. Chem.* 1997, **378**:151-160.

Netzel-Arnett, S. et al. *J. Biol. Chem.* 1991, **266**:6747-6755.
Stöcker, W. et al. *Prot. Sci.* 1995, 4:823-840.
Tschesche, H. et al. *Biomed. Biochim.* Acta 1994, **50**:755-761.
Tschesche, H. *Methods Enzymol.* 1995, **248**:431-449.

Ref. for DNA/AA Sequences

Hasty, K.A. et al. *J. Biol. Chem.* 1990, **265**:11421-11424.
Sang, Q.A. and Douglas, D.A. *J. Prot. Chem.* 1996, **15**:137-160.



Structure of the Phe79 catalytic domain of human neutrophil collagenase displayed as a ribbon plot (MOLSCRIPT, Kraulis, 1991). Zinc- and calcium ions are shown as small and big spheres, respectively.

Neutrophil Lipocalin

Volker Zölzer and Harald Tschesche

Synonyms	Neutrophil gelatinase associated lipocalin; Lipocalin 2; α_2-microglobulin-related protein
Abbreviations	HNL; NGAL
Classifications	The HNL belongs to the family of lipocalins, that is a large group of proteins characterized by the ability to bind small molecules.
Description	The HNL is a member of the lipocalin familiy. The 25 kDa protein is mainly synthesized and stored in the specific granules of neutrophils until it is secreted upon stimulation of the cells. The HNL is stored in its monomeric form but due to its impair number of Cys residues the HNL is also found as a homodimer and a heterodimer covalently linked with gelatinase B via the free Cys-87. There are hints that low amounts of HNL are present in mast cells, too. The mature protein is a 178 aa residue glycoprotein with a backbone of calculated 20.5 kDa. An oligosaccharide is N-linked with Asn-65.
Structure	In spite of a low degree in aa sequence identity (20 - 30%) most of the lipocalin family members share a high similarity in their tertiary structures: three highly conserved sequence motifs in the HNL-related lipocalins form eight strands of anti-parallel β-sheets which are twisted against each other and form a funnel-like β-barrel. This structure encloses a hydrophobic pocket that functions as internal ligand binding site. The β-sheets are connected by 7 β-hairpin loops. Near the C-terminus of the protein the structure of all known lipocalins possesses one conserved α-helix.
Molecular Weight	24,500 Dalton glycosylated (16.3% carbohydrate); 20,500 Dalton unglycosylated
Sedimentation Coeff.	Unknown
Isoelectric Point	8.8 (measured); 9.02 (calculated)
Extinction Coeff.	24,375 M^{-1} cm^{-1}; 12.0 (280nm, 1%, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Several features of the HNL have been investigated but the main function of the protein remains unknown. The HNL is found to bind retinol in a 1:1-stoichiometry and therefore might serve as an transport protein. The chemotactic active peptide FMLP also binds to the protein, that indicates a possible regulative role for the physiological stimulation of neutrophils. This functional property mirrors in the effect on degranulation and chemotaxis of neutrophils that is reduced by small concentrations of HNL. Further it was demonstrated that HNL binds to the cell surface of neutrophils via a receptor that probably mediates the cellular effects. The

fact that the HNL is partially secreted in association with gelatinase B indicates a putative function in the activation process of the MMPs. There are hints that the activation of latent Pro-MMPs can be supported by HNL.

Physiology/Pathology

The HNL is produced and stored in the specific granules of human neutrophils. It is secreted upon stimulation of the cells and accumulates in the surrounding tissue. Thus the HNL can serve as an inflammatory marker protein as it is shown by high concentrations of the protein in synovial fluids of patients suffering from rheumatoid arthritis.

Degradation

The lipocalins are a group of very compact proteins that is stable towards degradation by proteolytic enzymes or cleaving agents. The HNL is resistant to digestion by trypsin and cannot be cleaved by CNBr.

Genetics/Abnormalities

The gene for the HNL is located on the chromosome 9 and was mapped by in situ hybridisation to 9q34.

Half-life

The HNL is not a circulating protein, but once secreted into the plasma it is assumed to have a long half life considering its complex structure and its proteolytic resistance.

Concentration

The avarage concentration of HNL found in human plasma is about 100 µg/L. The level is significantly increased in synovial fluids of patients suffering from rheumatoid arthritis and reaches about 1.7 mg/L.

Isolation Method

Best source of native HNL are leukocytes prepared from human buffy coats. After homogenisation with an Ultraturrax and centrifugation the supernatant is subjected to DEAE-Sepharose. The monomeric HNL is not retarded by the resin and can thus be seperated from the gelatinase-bound fraction. After dialysis the solution is applied to Cibacron Blue Sepharose. The HNL is eluted by a NaCl gradient and finally purified by gelfiltration with a Sephacryl S-100 HR column.

Amino Acid Sequence

The N-terminal aa of the mature protein is a pyroglutamate and therefore resistant towards Edman-degradation, even after reduction and carboxymethylation.

QDSTSDLIPA PPLSKVPLQQ **NFQDNQFQGK WYVVGLA**GNA
ILREDKDPQK MYATIYELKE DKSYNVTSVL FRKKKCDYWI
RTFVPGCQPG EFTLGNIKSY **PGLTSYLVRV VSTNYNQHAM**
VFFKKVSQNR **EYFKITLYGR TKELTSELKE** NFIRFSKSLG
LPENHIVFPV PIDQCIDG

The marked parts of the protein are sequences that form structure conserved regions. These motifs are responsible for the high structural similarity of different members of the lipocalins. The three Cys-residues are marked as well.

Disulfides/SH-Groups

The HNL has three Cys-residues, the C-76 and C-175 form an intramolecular disulfide-bridge, the C-87 is free and can form an intermolecular disulfide-bond towards other proteins.

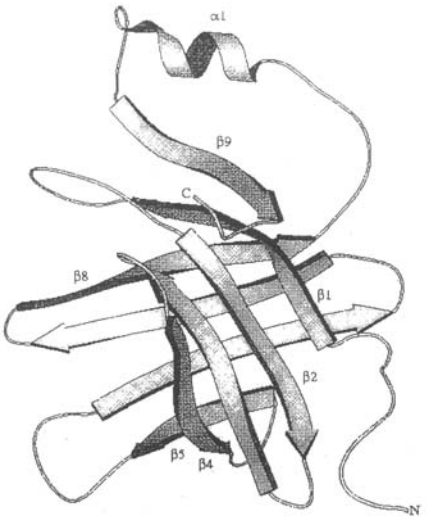
General References

Triebel, S. et al. *FEBS Lett.* 1992, **314**:386-388.
Kjeldsen, L. et al. *J. of Biol. Chem.* 1993, **268**:10425-10432.
Flower, D.R. *FEBS Lett.* 1994, **354**:7-11.
Bläser, J. et al. *Clin. Chim. Acta* 1995, **235**:137-145.

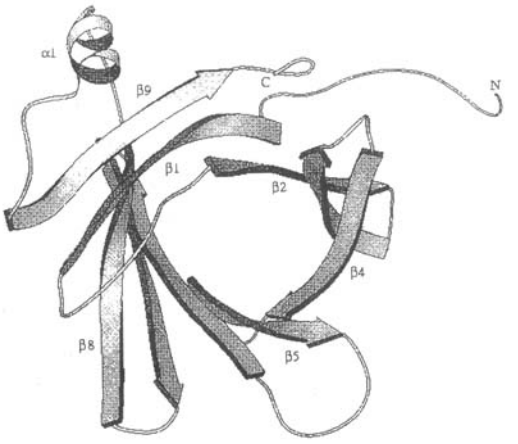
Flower, D.R. *Biochem. J.* 1996, **318**:1-14.
Kjeldsen, L. *Eur. J. of Haematol. Suppl. 56* 1995, **54**:1-30.

Ref. for DNA/AA Sequences

Chan, P. et al. *Genomics* 1994, **23**:145-150.



side view of HNL



end view of HNL

These are low resolution NMR structures of HNL that show some typical features of the lipocalin family: eight β-sheets form an internal ligand binding site, near the C-terminus one α-helix is formed.

Nitric Oxide Synthase

Benjamin Hemmens and Bernd Mayer

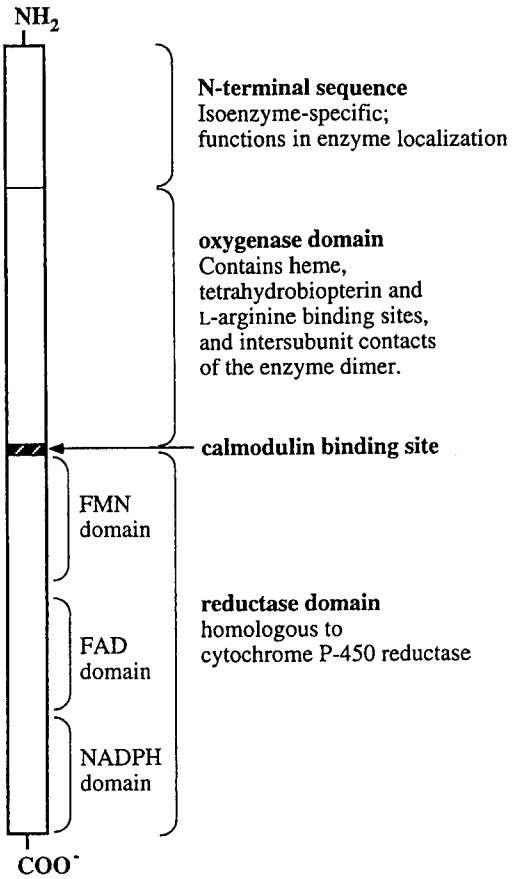
Synonyms	None
Abbreviations	NOS A variety of abbreviations are used for the three isoenzymes: neuronal NOS: nNOS; bNOS; ncNOS; Type I NOS endothelial NOS: eNOS; ecNOS; Type III NOS inducible NOS: NOS; macNOS; Type II NOS
Classification	EC 1.14.13.39
Description	Intracellular enzyme containing bound flavin, heme and tetrahydrobiopterin cofactors. The isoenzymes specified above have the following distinct properties: nNOS: expressed in neuronal cells and skeletal muscle; has an N-terminal PDZ-domain extension that is implicated in binding of nNOS to membrane bound proteins. Enzyme activity requires Ca^{2+}-dependent binding of calmodulin. eNOS: expressed in vascular endothelial cells, epithelial cells, and cardiomyocytes; has an N-terminal extension with myristoylation and palmitoylation sites, which may cause association of eNOS to membranes. Enzyme activity requires Ca^{2+}-dependent binding of calmodulin. iNOS: expressed in macrophages, hepatocytes, astroglia, smooth muscle; binds calmodulin Ca^{2+}-independently and is therefore not regulated by changes in Ca^{2+} concentration.
Structure	The active form of the enzyme is the homodimer. Subunits contain an N-terminal isoenzyme-specific sequence; a heme (Fe(III) protoporphyrin IX)- and tetrahydrobiopterin-binding oxygenase domain, a calmodulin-binding sequence, and a C-terminal reductase domain containing one binding site each for FMN, FAD and NADPH. The crystal structure of the oxygenase domain shows that the heme-binding structure is composed mainly of β-sheet, unlike any other known heme-binding fold.
Molecular Weight	nNOS: 320,000 (sequence); subunit 160,000 eNOS: 268,000 (sequence); subunit 134,000 iNOS: 260,000 (sequence); subunit 130,000
Sedimentation Coeff.	For pig brain enzyme, 7.9 ± 0.2 S; the human enzyme comigrates on size exclusion chromatography.
Isoelectric Point	Unknown
Extinction Coeff.	Absorbance at 280 nm should not be used to measure the enzyme concentration, because the content of flavins can vary. Heme is thiolate-ligated and therefore when reduced, its CO complex gives a typical P-450 type spectrum with a maximum at 446 nm; no extinction coefficient is available for the human enzyme (rat nNOS, 121 mM^{-1}).
Enzyme Activity	Catalyses the production of nitric oxide free radical and L-citrulline from L-arginine and 2 molecules of oxygen with the oxidation of 1.5 molecules of NADPH. Under some circumstances may also catalyse the following part reactions:

	1) oxidation of NADPH with reduction of cytochrome *c* or other electron acceptors 2) oxidation of NADPH with release of reduced oxygen species such as superoxide or peroxide from the heme center
Coenzymes/Cofactors	FMN and FAD: involved in shuttling reducing equivalents from NADPH to the heme Tetrahydrobiopterin ((6*R*)-5,6,7,8-tetrahydro-L-biopterin): allosteric activator; stabilizes dimeric structure; necessary for coupling of NADPH oxidation to NO synthesis; nature of chemical involvement in catalysis is not clear. Heme: site of reductive oxygen activation and conversion of L-arginine to N^G-hydroxy-L-arginine and then to L-citrulline with the release of NO.
Substrates	L-arginine, O_2, NADPH
Inhibitors	N^G-monomethyl-L-arginine, N^G-nitro-L-arginine (please note: N^G-nitro-L-arginine methyl ester is *not* an inhibitor), imidazole, S-ethylisothiourea, 2-methylthioacetamidine, 2-thienylcarbamidine, 2-iminopiperidine, 2-iminohomopiperidine, 2-amino-5,6-dihydro-6-methyl-4H-1,3-thiazine, S-methyl-L-thiocitrulline, S-ethyl-L-thiocitrulline, 2,4-diamino-5,6,7,8-tetrahydro-6-(L-erythro-1,2-dihydroxypropyl)pteridine (the 4-amino derivative of tetrahydrobiopterin)
Biological Functions	Synthesis of NO. NO has the following major functions: 1) as the endothelium-derived relaxing factor, reduces blood pressure and inhibits platelet activation 2) as a modulator of neurotransmission in the brain, proposed to be important for hippocampal long-term potentiation and thus for learning and memory 3) as one of several cytotoxic compounds released by activated macrophages, to kill pathogenic organisms in the inflammatory response
Physiology/Pathology	1) nNOS may contribute to ischemia/reperfusion injury in brain following stroke 2) excessive activation of iNOS is implicated as a cause of the hypotension in septic shock 3) iNOS may be responsible for tissue damage in rheumatoid arthritis and other inflammatory diseases 4) supplementation of endogenous eNOS by NO-releasing drugs (organic nitrates, sydnonimines, nitroprusside) is used to treat a variety of cardiovascular disease states
Genetics/Abnormalities	The genes for the three isoenzymes are located on the following chromosomes: nNOS, chromosome 11; eNOS, chromosome 7, iNOS, chromosome 14. Initial studies suggest there may be functionally significant polymorphisms in some populations.
Isolation Method	Affinity chromatography on ADP-Sepharose and calmodulin-Sepharose; immunoaffinity methods.
Amino Acid Sequence	The flavin and NADPH-binding domains are homologous to other nucleotide-binding proteins, and the calmodulin binding sequence was also recognizable by sequence similarity. The N-terminal half of the protein, including the oxygenase and the N-terminal extension, has no known homologues outside the NOS family.
Disulfides/SH-Groups	

General References

Mayer, B. and Hemmens, B. *Trends Biochem. Sci.* 1997, **22**:477-481.
Hemmens, B. and Mayer, B. Enzymology of Nitric Oxide Synthases. In: *Methods in Molecular Biology Vol. 100. Nitric Oxide Protocols.* Titheradge, M.A. (Ed.) Humana Press Inc., Totowa, N.J. 1997, pp. 1-32.
Stuehr, D.J. *Annu. Rev. Pharmacol. Toxicol.* 1997, **37**:339-359.
Crane, B.R. et al., *Science* 1997, **278**:425-431.
Förstermann, U. and Kleinert, H. *Naunyn-Schmiedebergs Arch. Pharmacol.* 1995, **352**:351-364.
NO home page: http://www.apnet.com/no.

GenBank accession numbers for cDNA sequences:
nNOS: L02881, D16408, U17327, U17326, U31466.
eNOS: M93718, M95296, D26607, L26914, L23210.
iNOS: L02910, D26525, U20141, U05810, U31511, L24553.



Domain structure of nitric oxide synthase

Opsin

Paul A. Hargrave and J. Hugh McDowell

Synonyms	Visual purple
Abbreviations	R; Rho
Classifications	Glycoprotein; lipoprotein; retinoylprotein; phosphoprotein; intrinsic membrane protein; photoreceptor protein; G-protein-coupled receptor protein.
Description	Rhodopsin is a photoreceptor protein found in the eye, in outer segments of rod cells in the vertebrate retina. The protein is called opsin, whereas rhodopsin is the term for rod opsin that has bound 11-*cis* retinal. Following synthesis in the rod cell inner segment, rhodopsin is transported to the outer segment where it is incorporated into its plasma membrane. The outer segment plasma membrane evaginates at its base, eventually pinching off to form a disk whose membrane is not continuous with the plasma membrane. This process is continually repeated, forming a stack of intracellular disk membranes of which rhodopsin is the principal protein component.
Structure	X-ray and neutron diffraction data show (bovine) rhodopsin to be prolate in shape and approximately 60-65Å in length. CD measurements suggest the molecule contains 50-60% α-helix, with the helices being oriented perpendicular to the plane of the membrane. Cryo-electron microscopy shows rhodopsin to be a helical bundle composed of seven transmembrane helices. The helical bundle forms an enclosed pocket for the chromophore, 11-*cis* retinal.
Molecular Weight	The protein molecular weight of human rod opsin is 38,892, calculated from its aa sequence. Its total molecular weight is 41,831, assuming acetylation at its N-terminus, palmitoylation at two cysteines, retinal attachment at one lysine, and glycosylation with $Man_3GlcNAc_3$ at two asparagines.
Sedimentation Coeff.	Unknown
Isoelectric Point	6.62 (calc); the isoelectric point of bovine rhodopsin is 6.0 (6.23 calc).
Extinction Coeff.	Molar extinction coefficient 40,600 at 498 nm (bovine); human rhodopsin maximal absorption at 496 nm, molar extinction coefficient not determined.
Enzyme Activity	The photoactivated form of rhodopsin causes activation of the heterotrimeric G-protein, transducin.
Coenzymes/Cofactors	11-*cis* retinal is covalently attached to the epsilon amino group of K296 via a Schiff base. Absorption of the energy of a photon of light causes the isomerization of 11-*cis* retinal to all-*trans* retinal, leading to a conformational change in the protein.
Substrates	Light-activated rhodopsin acts to convert the GDP-bound form of the G-protein alpha subunit to its GTP-bound (activated) form.

Inhibitors	Any compound that interacts directly with rhodopsin, or indirectly at any step in the visual cycle, to prevent association of 11-*cis* retinal with the protein opsin, may be considered to be an inhibitor. Compounds have been identified that influence steps such as rhodopsin regeneration and activation of transducin and rhodopsin kinase. This would include the anti-Schistosomal agent diaminophenoxypentane (DAPP) and related retinotoxic primary aromatic amines. DAPP markedly decreases the rate of rhodopsin regeneration and eventually causes blindness. It acts to deplete the stores of 11-*cis* retinoids in the eye by catalyzing their isomerization to all-*trans*, thus short-circuiting the visual cycle.
Biological Functions	Rhodopsin is the photosensitive pigment that mediates human (and animal) dim-light and black-and-white vision. A photon of light strikes a molecule of rhodopsin and causes rhodopsin's bound 11-*cis* retinal to isomerize to the all-*trans* configuration. This causes the protein to change its conformation to a light-activated form, which activates the heterotrimeric G-protein transducin. Transducin in turn activates a cGMP-phosphodiesterase which then lowers the rod cell concentration of cGMP. This drop in cGMP causes a plasma membrane cation channel to close, causing the rod cell to signal a second-order neuron, leading to transmission of visual information via the optic nerve to the brain.
Physiology/Pathology	The physiological role of rhodopsin is to initiate the pathway of visual transduction. Decreased amounts of rhodopsin in the retina leads to decreased light sensitivity, and complete absence of the protein leads to blindness. Overproduction of the protein (in transgenic mice) can lead to rod cell death and eventual blindness.
Degradation	Rhodopsin is incorporated into the lipid bilayer of newly synthesized disk membranes at the base of the outer segment of retinal rod cells. These newly synthesized intracellular membranes join a stack of disk membranes that become displaced from the base to the tip of the outer segment by newly added disks. Packets of disk membranes are periodically removed from the tips of the rod cell outer segments by a surrounding pigment epithelium cell.
Genetics/Abnormalities	The gene for human rod opsin is located at 3q21-q24. Mutations in rhodopsin are one of the causes of dominant retinitis pigmentosa, recessive retinitis pigmentosa and dominant congenital stationary night blindness. As much as 30% of autosomal dominant retinitis pigmentosa is due to mutations in rhodopsin. More than 70 such disease-causing mutations in rhodopsin have been identified. The mutations appear to affect properties or functions of rhodopsin such as its ability to fold properly and bind retinal, its glycosylation, or its ability to be transported to its proper location in the rod cell outer segment.
Half-life	The average lifetime of disks (and therefore opsin) in human rod cells, from synthesis to removal by phagocytosis, is approximately 10 days.
Concentration	Opsin constitutes greater than 95% of the membrane protein in the disk membranes of rod cell outer segments. Its concentration is high; 3.3 mmol/L (in frog).
Isolation Method	Methods have been developed for preparation of rhodopsin from cattle, and may be applied to human and other species. Dark-adapted retinas are homogenized in a sucrose and buffered salts solution and submitted to centrifugation to float rod cell outer segments that contain rhodopsin. Crude rod outer segments are further purified by sucrose density gradient

centrifugation. Rhodopsin may be solubilized by dissolving rod outer segment membranes in a mild detergent (octyl glucoside, dodecylmaltoside). Purification of rhodopsin may be achieved by chromatography on concanavalin A-agarose or hydroxyapatite.

Amino Acid Sequence

Characteristic elements of the rod opsin sequence include N-acetyl M1, N2 and N15 to which $Man_3GlcNAc_3$ are attached, a C110-C187 disulfide bridge, ERY136 sequence for activation of G-protein, K296 binding site for retinal, C322 and C323 palmitoylation sites, and a C-terminal sequence A333SATVSKTETS343 enriched in S and T sites of phosphorylation by rhodopsin kinase (principal phosphorylation sites are S334, S338 and S343).

Disulfides/SH-Groups

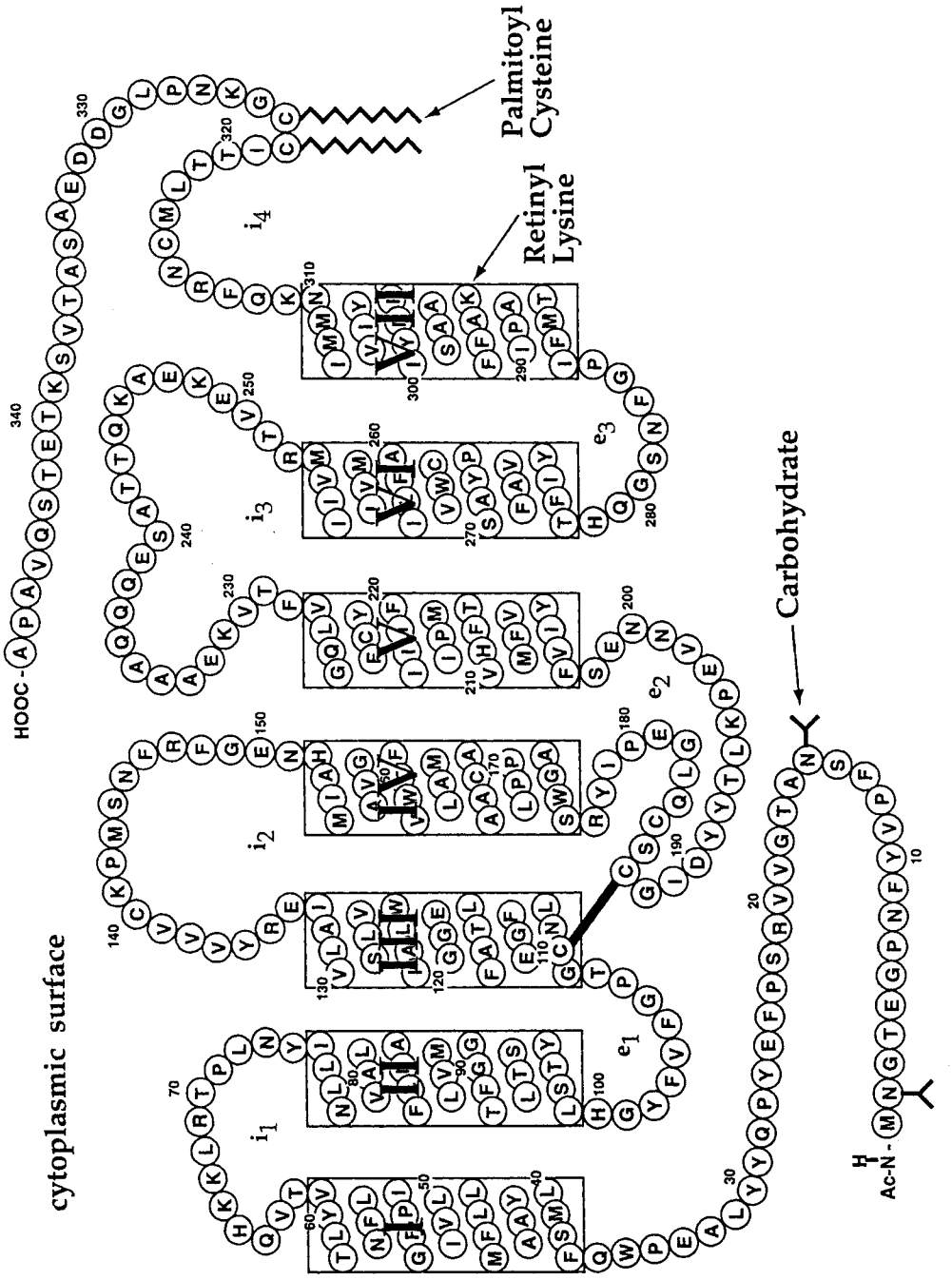
There are 10 cysteines in the sequence. An intrachain disulfide bridge links C-110 to C-187.

General References

Hargrave, P.A. and McDowell, J.H. *International Review of Cytology* 1992, **137B**:49-97.
Khorana, H.G. *J Biol Chem.* 1992, **267**:1-4.
Nathans, J. *Biochemistry* 1992, **31**:4923-4931.
Rando, R.R. The Bioorganic Chemistry of Vision. In: *Chemistry and Biology of Synthetic Retinoids.* Dawson, M.I. and Akamura, W.H. (eds.), CRC Press, Boca Raton, FL. 1990, pp. 1-26.
Daiger, S.P. et al. *Behavioral and Brain Sciences* 1995,**18**:452-467.

Ref. for DNA/AA Sequences

Nathans, J. and Hogness, D.S. *Proc. Natl. Acad. Sci. USA* 1984, **81**:4851-4855. Genbank HSU49742.



Molecular model: A topographic model for human rod opsin in the disk membrane. Opsin's polypeptide chain traverses the membrane lipid bilayer seven times, shown by the transmembrane helices I-VII. Helix-connecting loops i_1-i_4 and the C-terminal region are located facing the rod cell cytoplasmic surface. The N-terminal region and loops e_1-e_3 are on the rod cell extracellular surface for molecules of opsin located in the rod outer segment plasma membrane, and on the intradiskal surface of opsin molecules in the disk membranes. Oligosaccharide chains are attached to N2 and N15. A disulfide bridge links C110 and C187. C322 and C323 are palmitoylated. Retinal is linked to K296.

Ornithine Carbamoyltransferase

Bendicht Wermuth

Synonyms	Ornithine Transcarbamylase
Abbreviations	OTC (OCT)
Classifications	EC 2.1.3.3
Description	A homotrimeric, nuclear encoded, mitochondrial matrix protein occurring in the liver and small intestine. Subunits are synthesized as precursor peptide chains consisting of 354 aa. After import into the mitochondrion a leader peptide of 32 aa is cleaved off and three subunits combine in a chaperonin-mediated folding process to form the active enzyme.
Structure	The tertiary structure is not yet determined, however the sequence has been modelled to the known structures of *Pseudomonas aeruginosa* OTC and *Escherichia coli* aspartate transcarbamylase.
Molecular Weight	108,000 (subunit, 36,112; precursor, 39,901)
Sedimentation Coeff.	6.5 (porcine liver)
Isoelectric Point	6.8, 7.3, 7.95 (dependent on reference)
Extinction Coeff.	12 (280nm, 1%, 1cm; calculated from aa composition)
Enzyme Activity	Transferase: Catalyzes the formation of citrulline and inorganic phosphate from carbamyl phosphate and ornithine. Specific activity ca. 200 µmol citrulline formed/min per mg protein (0.2 M triethanolamine-HCl, pH 7.7, 37°C).
Coenzyme/Cofactors	None
Substrates	Carbamyl phosphate (K_m = 0.2 mM), L-ornithine (K_m = 0.5 mM)
Inhibitors	δ-N-(phosphonoacetyl)-L-ornithine (PALO, K_i = 0.27 µM, rat enzyme); several aa, e.g. norvaline (IC_{50} = 0.24 mM).
Biological Functions	Second enzyme of the urea cycle of ureotelic animals, involved in ammonia detoxification.
Physiology/Pathology	Content of OTC and other urea cycle enzymes is positively correlated with protein intake. OTC deficiency (< 160 mIU/mg liver protein) results in hyperammonemia and orotic aciduria, often with fatal outcome.
Degradation	Unknown
Genetics/Abnormalities	The gene is on the X-chromosome at p 21.1. It spans at least 73 kb and contains 10 exons. More than 130 mutations, the majority of them missense mutations, associated with clinically manifest OTC deficiency have been described. About 80 % of the mutations in girls and 20 % in boys have arisen de novo.
Half-life	Unknown

Concentration	~ 2 ‰ of soluble proteins in liver homogenate (after disruption of mitochondria). Not usually present in plasma or serum.
Isolation Method	(a) Ion exchange chromatography on DEAE-cellulose, followed by chromatography on hydroxylapatite and gel filtration on Sephadex G-200. (b) Ammonium sulfate precipitation followed by gel filtration on Sephacryl S-200 and PALO-affinity chromatography.
Amino Acid Sequence	MLFNLRILLN NAAFRNGHNF MVRNFRCGQP LQNKVQLKGR DLLTLKNFTG EEIKYMLWLS ADLKFRIKQK GEYLPLLQGK SLGMIFEKRS TRTRLSTETG LALLGGHPCF LTTQDIHLGV NESLTDTARV LSSMADAVLA RVYKQSDLDT LAKEASIPII NGLSDLYHPI QILADYLTLQ EHYSSLKGLT LSWIGDGNNI LHSIMMSAAK FGMHLQAATP KGYEPDASVT KLAEQYAKEN GTKLLLTNDP LEAAHGGNVL ITDTWISMGQ EEEKKKRLQA FQGYQVTMKT AKVAASDWTF LHCLPRKPEE VDDEVFYSPR SLVFPEAENR KWTIMAVMVS LLTDYSPQLQ KPKF Precursor protein; the mature protein starts at Asn-33
Disulfides/SH-Groups	No disulfides, 3 free sulfhydryls, one of them in the leader peptide.
General References	Brusilow, S.W.and Horwich A.L. "Urea Cycle Enzymes". In: *The Metabolic and Molecular Bases of Inherited Disease, 7th ed.*, Scriver C.R. et al. (eds.) Mc Graw Hill, 1995; pp 1187-1232. Tuchman, M. *J. Lab. Clin. Med.* 1993, **120**: 836-850. Tuchman, M. et al. *J. Med. Genet.* 1995, **32:**680-688.
Ref. for DNA/RNA Sequences	Hata, A. et al. *J. Biochem.* 1988, **103:**302-308. Horwich, A.L. et al. *Science* 1984, **224:**1068-1074.

Osteonectin

Georg L. Long and Kenneth G. Mann

Synonyms	SPARC (Secreted Protein; Acidic and Rich in Cysteine); 43K Protein; BM-40 (Basement Membrane)
Abbreviations	ON
Classifications	None
Description	An extracellular, single polypeptide, glycoprotein (approx. 5% carbohydrate), most abundant in bone tissue. Also found in many other cell types including platelets, endothelial cells, fibroblasts, parietal yolk sac cells and basement membrane tumor. High affinity binding to Ca^{2+}, hydroxypatite and type I collagen suggests a possible role in mineralization. Association with a number of highly proliferative, mobile cell types and binding to thrombospondin, plasminogen and tissue plasminogen activator (tPA) suggests that the protein may play an important role in extracellular matrix formation and regulation. Secreted by platelets upon activation. Of the two potential Asn-X-Thr/Ser N-glycosylation sites, apparently only Asn-99 is modified. The lack of carbohydrate at this position increases the affinity for type V collagen. Glycosylation pattern and resulting molecular weight are different for secreted versus non-secreted protein. The protein contains a segment homologous to avian ovomucoid protease inhibitors, and two potential troponin C-like "EF-hands" for Ca^{2+} binding. The N-terminal 52 residue domain is extremely acidic (15 Glu/3 Asp, no Lys or Arg). Elements within the first 17 aa are responsible for its binding to type V collagen.
Structure	Tertiary structure not determined. Chou-Fasman analysis predicts 38% α-helix, 14% β-strand, 48% random-coil or β-turn.
Molecular Weight	29 kDa: bone (sedimentation equilibrium) 36 kDa: cell lysates, bone (SDS-PAGE, reduced) 39 kDa: secreted, platelets (SDS-PAGE, reduced)
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown; 5.5 for bovine bone.
Extinction Coeff.	Unknown; 3.6 (280nm, 1%, 1cm) for bovine bone.
Enzyme Activity	Unknown
Coenzymes/Cofactors	Unknown
Substrates	Specific binding to thrombospondin, types I, IV and V collagen, plasminogen, tissue plasminogen activator.
Inhibitors	Unknown
Biological Functions	No known function. See "Description" for possible function.
Physiology/Pathology	Unknown

Degradation	Slow fragmentation during isolation and storage.
Genetics/Abnormalities	Located on chromosome 5q31-q32. Believed to be a single copy gene. No known mutations.
Half-life	Unknown
Concentration	Varies considerably depending on cell and tissue type. Highest concentrations associated with active extra-cellular matrix formation and bone tissue. Estimated approximately 2 mg osteonectin/g whole fetal bone; 1.9 μg/ 2 x 10^8 platelets.
Isolation Method	Extraction of milled bone (most abundant source) with 5 mM EDTA, protease inhibitor-containing buffer; Sephadex G-100 fractionation; monoclonal antibody immunoaffinity chromatography.
Amino Acid Sequence	Nascent protein consists of a 17 aa signal peptide (in italics below) followed by the 286 residue mature protein. *MRAWIFF LLCLAGRALA* APQQEALPDE TEVVEETVAE VTEVSVGANP VQVEVGEFDD GAEETEEEVV AENPCQNHHC KHGKVCELDE NNTPMCVCQD PTSCPAPIGE FEKVCSNDNK TFDSSCHFFA TKCTLEGTKK GHKLHLDYIG PCKYIPPCLD SELTEFPLRM RDWLKNVLVT LYERDEDMNL LTEKQKLRVK KIHENEKRLE AGDHPVELLA RDFEKNYNMY IFPVHWQFGQ LDQHPIDGYL SHTELHPLRA PLIPMEHCTT RFFETCDLDN DKYIALDEWA GCFGIKQKDI DKDLVI Human osteonectin is a member of a family of closely related proteins described for other species including murine (SPARC or BM-40) and bovine (43K-protein); and a protein from human high endothelial venules, hevin.
Disulfides/SH-Groups	Increased mobility of the non-reduced form by SDS-PAGE suggests extensive intrachain disulfide bonding. Alkylation of reduced versus non-reduced protein suggests that most of the 14 Cys residues exist with disulfide bonds. Disulfide bond patterns are established for Cys 137-247 and 255-271 in mouse. Others are tentatively assigned only by homology with related proteins.
General References	Romberg, R.W. et al. *J. Biol. Chem.* 1985, **260**:2728-2736. Fisher, L.W. et al. *J. Biol. Chem.* 1987, **262**:9702-9708. Swaroop, A. et al. *Genomics* 1988, **2**:37-47. Kelm, R.J. et al. *Blood* 1992, **80**:3112-3119. Kelm, R.J. et al. *J. Biol. Chem.* 1994, **269**:30147-30153. Xie, R.L. and Long, G.L. *J. Biol. Chem.* 1995, **270**:23212-23217. Xie, R.L. and Long, G.L. *J. Biol. Chem.* 1996, **271**:8121-8125. Lane, T.F. and Sage, H. *FASEB J.* 1994, **8**:163-173 (Review). Tracy, R.P. et al. *Int. J. Biochem.* 1988, **7**:653-660 (Review).
Ref. for DNA/AA Sequences	Villareal, X.C. et al. *Biochemistry* 1989, **28**:6483-6491. *Gen Bank Association Numbers* J02863; Y00755; Y03040.

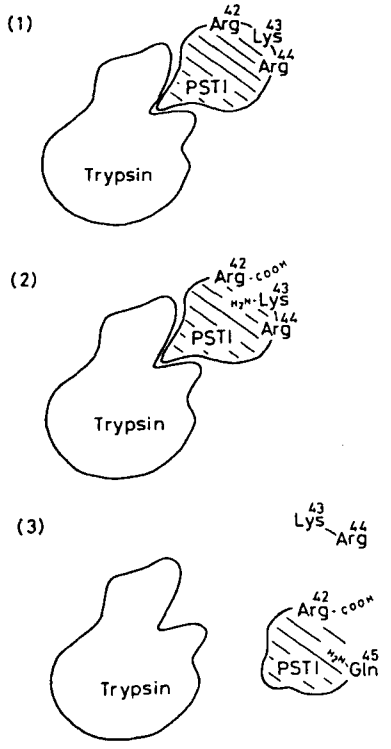
Human Protein Data
Edited by André Haeberli

Pancreatic secretory trypsin inhibitor

Michio Ogawa

Synonyms	Tumour-associated trypsin inhibitor (TATI), monotor peptide, cholecystokinin releasing peptide (CCK-RP)
Abbreviations	PSTI, TATI, CCK-RP
Classifications	Kazal-type trypsin inhibitor (temporary inhib.)
Description	A specific inhibitor thought to exist solely in the pancreas and pancreatic juice. Also produced by various cancer cells. A trace amount enters the circulating blood-stream.
Structure	A single polypeptide chain without sugar moiety. At least 4 forms in human pancreatic juice, one lacking 5 N-terminal aa residues of native form and two being deamidated forms.
Molecular Weight	6,242 (aa sequence)
Sedimentation Coeff.	Unknown
Isoelectric Point	5.8 (multiple forms due to deamidation)
Extinction Coeff.	8.4 (280 nm, 1%, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	A specific trypsin inhibitor that prevents autoactivation of trypsinogen in the pancreas and pancreatic juice. Stimulates DNA synthesis in various fibroblasts and growth of endothelial cells. Elicit cholecystokinin (CCK) release when injected into the duodenum.
Physiology/Pathology	Pancreatic: protects the pancreas from autodigestion by inhibiting activated trypsin and prevents acute pancreatitis. Cancer-produced: supposed to be a growth stimulating factor. Serum: an acute phase protein for the host defence or the repair of destructed tissues.
Degradation	Intact form eliminated from the circulation by the kidney. After complex formation with trypsin, inactivation of PSTI caused by deletion of Lys-43-Arg-44 and disruption of the conformation (see figure).
Genetics/Abnormalities	No genetic abnormalities reported. PSTI gene located on chromosome 5.
Half-life	6 min (blood circulation)
Concentration	11.5 μg/L (6.0–16.3)
Isolation Method	Trypsin-affinity chromatography and reversed-phase HPLC (or gel filtration).

Amino Acid Sequence	Consisted of 55 aa. The aa sequences (and also the nucleotide sequences) of PSTI and epidermal growth factor (EGF) are highly homologous, but the exon-intron junctions are not related at all, indicating that both do not have evolved from a common ancestor.
Disulfides/SH-Groups	6 cystein residues, no free sulfhydryls.
General References	Huhtala, M.-L., et al. *J. Biol. Chem.* 1982, **257:** 13713–13716. Iwai, K., et al. *J. Biol. Chem.* 1987, **262:** 8956–8959. Ogawa, M., et al. *Adv. Exper. Med. Biol.* 1988, **240:** 547–553. Ogawa, M. *Clin. Biochem.* 1988, **21:** 19–25. Ogawa, M., et al. Biological function of pancreatic secretory trypsin inhibitor expressed in various cancer tissues. In: *Intracellular Proteolysis – Mechanisms and Regulations.* Katunuma, N. and Kominami, E. (eds.) Japan Scientific Societies Press, Tokyo, 1989; pp. 519–526. Niinobu, T., et al. *J. Exper. Med.* 1990, **172:** 1133–1142. Tomita, N., et al. *Cancer* 1990, **66:** 2144–2149.
Ref. for DNA/AA Sequences	Yamamoto, T., et al. *Biochem. Biophys. Res. Commun.* 1985, **132:** 605–612. Horii, A., et al. *Biochem. Biophys. Res. Commun.* 1987, **149:** 635–641. Tomita, N., et al. *FEBS Lett.* 1987, **225:** 113–119.



Proposed mechanism of temporary inhibition of human PSTI. Human PSTI forms a complex with human trypsin and inhibits enzymatic activity under physiological conditions, (1). Next, the Arg^{42}-Lys^{43} bond of PSTI in the complex is cleaved by trypsin, but PSTI retains its inhibitory activity and binds to trypsin, (2). Further cleavage of the Arg^{44}-Gln^{45} bond of PSTI causes disruption of the conformation of PSTI and leads to dissociation of the complex, (3).
(From: Kikuchi, N., et al., *J. Biochem.*, 1989, **106:** 1059–1063.

Parathyroid hormone

Roman Muff and Jan A. Fischer

Synonyms	Parathormone; Parathyrin
Abbreviations	PTH
Classifications	Polypeptide hormone
Description	Intact PTH(1-84) is cleaved in parathyroid cells from larger molecular weight precursor proteins which include preproPTH and proPT. Transcription of the PTH gene is suppressed by raised extracellular calcium while stimulation by low calcium may be controlled by post-transcriptional events (Moran, 1981; Russell, 1983; Hawa, 1993). 1,25-dihydroxy vitamin D_3 also suppresses PTH gene transcription in vivo and in vitro (Silver, 1986). The secretion of PTH(1-84) is inversely related to the extracellular calcium concentration. While 1,25-dihydroxyvitamin D_3 suppresses PTH gene transcription, possible direct effects on the secretion of PTH not related to raised extracellular calcium remain to be substantiated (Fischer, 1982). Besides, β-adrenergic catecholamines and dopamine also stimulate the secretion of PTH. Activation of protein kinase C stimulates PTH secretion while changes of the cytosolic calcium, which parallel extracellular calcium concentrations, and of cyclic AMP production are not clearly related to the calcium regulated PTH secretion (Muff, 1986). PTH is the most important hypercalcaemic hormone. The biological activity is restricted to the N-terminal parts of the intact PTH(1-84) molecule.
Structure	All known PTH are single chain polypeptides with 84 aa. Human proPTH has 90 aa and preproPTH 115 aa.
Molecular Weight	9425: human PTH(1-84)
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	Unknown
Substrates	None
Inhibitors	N-terminal truncation of PTH(1-34) leads to PTH antagonists. The most potent antagonist while retaining some agonist properties is PTH(3-34). A pure antagonist is PTH(7-34).
Biological Functions	PTH raises serum calcium levels through stimulation of renal tubular reabsorption of calcium, and of bone resorption through interaction with osteoblasts. Probably indirectly, through activation of the renal 1α-vitamin D hydroxylase PTH enhances the intestinal calcium absorption. In the kidney, PTH also stimulates the tubular reabsorption of phosphate thus lowering serum phosphate concentrations. Recently, the structure of human, rat, mouse and American opossum G protein-coupled PTH-receptors with seven transmembrane domains linked to adenylyl cyclase activation and to intracellular calcium mobilization has been revealed through cloning of

complementary and genomic DNA (Jüppner, 1991; Abou-Samra, 1992; Schipani, 1993; McCuaig, 1994). High homology exists between the rat, mouse and human PTH-receptor genes (Kong, 1994) which have been assigned to the chromosomes 3, 8 and 9 in man, rat and mouse, respectively (Pausova, 1994). PTH-related peptide (PTHrP), associated with humoral hypercalcaemia of malignancy, interacts indistinguishably with these PTH receptors. The proximal tubular effects of PTH and PTHrP include inhibition of brush border transport systems, Na/P_i-cotransport and Na/H-exchange (Muff, 1992). More recently a PTH2-receptor with 70% sequence homology to the PTH/PTHrP receptor which is selectively activated by PTH, but not by PTHrP, has been identified in brain and pancreas (Usdin, 1995).

Physiology/Pathology

Developmental absence or surgical removal of parathyroid glands results in life threatening hypocalcaemia which requires treatment with calcium and D-vitamines. Hypercalcaemia, on the other hand, is the leading symptom of unregulated hypersecretion of PTH from parathyroid tumors seen in primary hyperparathyroidism, or results from the administration of pharmacological amounts of PTH.

Degradation

PTH(1-84) is eliminated from the circulation through cleavage in the liver into N-terminal fragments not clearly recognized in the peripheral circulation as well as into C-terminal PTH fragments accumulating following nephrectomy and in chronic renal failure. The C-terminal fragments are removed by the kidney through glomerular filtration and degradation in renal tubules. Negligible proteolytic dagradation occurs in the general circulation.

Genetics/Abnormalities

The human PTH gene is located on chromosome 11p15. A rare cause of autosomal recessive hypoparathyroidism is a donor splice site mutation in the PTH gene (Parkinson, 1992).
Pseudohypoparathyroidism is a form of clinical hypoparathyroidism and resistance of target organs to PTH. In some patients $G_s\alpha$ deficiency may be pathogenetically linked (Loveridge, 1986).

Half-life

9.1-22.9 mL min^{-1} kg^{-1} metabolic clearance, bovine

Concentration

Circulating levels of human PTH(1-84) in normal subjects are of the order of 10 ng L^{-1} (1 pMole L^{-1}) as measured after gel permeation chromatography of normal human serum as well as in cytochemical bioassay (Fischer, 1993). PTH has been isolated from parathyroid glands. mRNA encoding PTH has only exceptionally been recognized in carcinoma of the ovary and the lung (Nussbaum, 1990). In the majority of patients with humoral hypercalcaemia of malignancy unrelated to the parathyroid glands, the levels of circulating PTH are low or undectable, but those of PTHrP are recognizable or elevated (see there).

Isolation Method

Isolated from parathyroid glands through extraction by C_{18}-cartridges, and sequential high performance liquid chromatography steps.

Amino Acid Sequence

See structure

Disulfides/SH-Groups

None

General References

Abou-Samra, A.B. et al. *Proc. Natl. Acad. Sci. USA* 1992, **89**:2732-2736.
Fischer, J.A. *Clin. Investig.* 1993, **71**:505-518.
Fischer, J.A. Calcif. *Tissue Int.* 1982, **34**:313-316.
Hawa, N.S. et al. *J. Mol. Endocrinol.* 1993, **10**:43-49.

Hendy, G.N. et al. *Proc. Natl. Acad. Sci. USA* 1981, **78**:7365-7369.
Hunziker, W.H. et al. *Pflügers Arch.* 1977, **371**:185-192.
Jüppner, H. et al. *Science* 1991, **254**:1024-1026.
Kong, X.-F. et al. *Biochem. Biophys. Res. Commun.* 1994, **200**:1290-1299.
Loveridge, N. et al. *Biochim. Biophys. Acta* 1986, **889**:117-122.
McCuaig, K.A. et al. *Proc. Natl. Acad. Sci.* 1994, **91**:5051-5055.
Moran, J.R. et al. *Endocrinology* 1991, **108**:2264-2268.
Muff, R. et al. *Annu. Rev. Physiol.* 1992, **54**:67-79.
Nussbaum, S.R. et al. *N. Engl. J. Med.* 1990, **323**:1324-1328.
Parkinson, D.B. and Thakker, R.V. *Nature Genetics* 1992, **1**:149-152.
Pausova, Z. et al. *Genomics* 1994, **20**:20-26.
Russell, J. et al. *J. Clin. Invest.* 1983, **72**:1851-1855.
Schipani, E. et al. *Endocrinology* 1993, **132**:2157-2165.
Silver, J. et al. *J. Clin. Invest.* 1986, **78**:1296-1301.
Usdin, T.B. et al. *J. Biol. Chem.* 1995, **270**:15455-15458.

Ref. for DNA/AA Sequences

Hendy, G.N. et al. *Proc. Natl. Acad. Sci. USA* 1981, **78**:7365-7369.

Parathyroid hormone-related protein

Jan A. Fischer and Roman Muff

Synonyms	Parathyroid hormone-like peptide
Abbreviations	PTHrP
Classifications	Protein factor
Description	In the majority of patients with humoral hypercalcaemia of malignancy, PTHrP produced by the tumors raises serum levels of calcium through actions on bone and kidney. The hypercalcaemic effects cannot be distinguished from those of parathyroid hormone. This can be explained by high structural homology of N-terminal aa in PTHrP and parathyroid hormone interacting with the same receptors (Muff, 1992), see fig. Specific physiological actions different from parathyroid hormone include stimulation of placental calcium transport and conservation of fetal calcium (Rodda, 1992) (see below).Human PTHrP has been isolated from carcinoma such as the medium of the BEN lung cancer cell line, and the N-terminal sequence used for cloning (Suva, 1987). Up to 7 mRNA species are capable of coding for three forms of the mature protein comprised of 139, 141, 173 aa, respectively (Suva, 1987; Mangin et al., 1988,1989; Thiede, 1988; Yasuda, 1989). The regulation of transcription, and of the release or secretion of PTHrP remains to be explored in detail (Hendy, 1992). So far, glucocorticoids and 1,25$(OH)_2$ -vitamin D_3 have been shown to suppress PTHrP mRNA through cis-acting elements in tumor cell lines and keratinocytes (Lu, 1989; Ikeda, 1989; Kremer, 1991). Stimulation of PTHrP by transforming growth factor β implies paracrine action (Kiriyama, 1993).
Structure	Human PTHrP are single chain proteins with 139, 141 and 173 aa deduced from the respective cDNA. Analysis of the sequence of aa has not been performed.
Molecular Weight	16,043: human PTHrP(1-141)
Sedimentation Coeff.	Unknown
Isoelectric Point	> 8.7
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	Unknown
Substrates	None
Inhibitors	Truncated parathyroid hormone (7-34) suppressed biological effects of PTHrP (Horiuchi, 1990).
Biological Functions	PTHrP raises serum calcium levels through stimulation of renal tubular reabsorption of calcium and of bone resorption. The latter is stimulated in endocrine manner with PTHrP reaching bone osteoblasts through the circulation and in paracrine manner produced locally from bone metastases. Different native forms of PTHrP that all include an intact 1-36 N-terminal sequence and PTH share G protein-coupled PTH/PTHrP-

receptors with seven transmembrane domains linked to adenylyl cyclase activation and to intracellular calcium mobilization, recently identified through DNA cloning in man, rat, mouse and American opossum (Jüppner, 1991; Abou-Samra, 1992; Schipani, 1993; McCuaig, 1994). Receptor interactions are the same as those of parathyroid hormone (Muff, 1992).
Vasodilation is also shared with parathyroid hormone (Musso, 1989). Stretch induced stimulation of PTHrP mRNA in the urinary bladder and in the uterus evokes a paracrine mechanism of action (Yamamoto, 1992; Daifotis, 1992). The latter my be important for relaxation of the uterus through extension by the foetus. After rupture of foetal membranes, lowered PTHrP mRNA may play a role in the onset of labor (Ferguson, 1992).
Besides the similar effects of PTHrP and of parathyroid hormone on kidney and bone, PTHrP has specific actions not shared by parathyroid hormone. These include stimulation of placental calcium transport and retention of calcium in the foetus (Rodda, 1992; Thiede, 1992). To this end, PTHrP is expressed in foetal parathyroid glands. Moreover, suckling enhances PTHrP mRNA in the lactating rat breast as a result of prolactin secretion. PTHrP identified in large amount in the milk may reach the newborn immediately after birth and there raise serum levels of calcium and lower those of parathyroid hormone (Grill, 1992).

Physiology/Pathology

Beside endocrine functions of PTHrP mainly during fetal life and the smooth muscle relaxant activity shared with PTH (see above), PTHrP acts as a key developmental factor regulating growth, differentiation and development in auto- and paracrine manner (for a review see Philbrick, 1996). Furthermore from all humoral factors producing hypercalcaemia of malignancy which include prostaglandins, cytokines and transforming growth factors, PTHrP is the most studied and probably the most frequent. To this end, antibodies to PTHrP have been shown to lower serum calcium in mice with implanted human tumors causing hypercalcaemia (Kukreja, 1988).

Degradation

The metabolism of the presumed intact 139, 141 and 173 aa proteins has not been studied, and little is known about circulating components and possible breakdown products.

Genetics/Abnormalities

The human PTHrP gene is located on the short arm of chromosome 12 (Mangin, 1988).

Half-life

See degradation

Concentration

Circulating levels of human PTHrP in normal subjects are of the order of 1 pMole L^{-1}. mRNA encoding PTHrP has been identified in multiple organs including the brain, and in some immunohistochemical analyses and peptide purification have been carried out. In the majority of patients with humoral hypercalcaemia of malignancy unrelated to the parathyroid glands, the levels of circulating PTHrP are recognizable or elevated. Poor discrimination of serum values between normal subjects and patients with humoral hypercalcaemia of malignancy may be technical, but also explained by the localized release and action (paracrine) of PTHrP with the proteins not reaching the peripheral circulation in significant quantities (Soifer, 1992).

Isolation Method

Isolated from tumor cell lines and medium thereof through extraction by ammonium sulfate, and sequential high performance liquid chromatography steps.

Amino Acid Sequence

Deduced from cDNA (see there).

Disulfides/SH-Groups

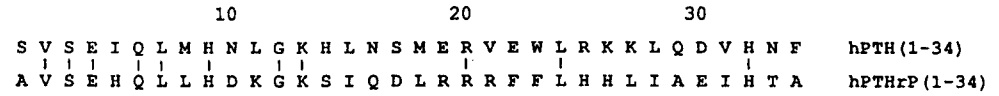
None

General References

Abou-Samra, A.B. et al. *Proc. Natl. Acad. Sci. USA* 1992, **89**:2732-2736.
Daifotis, A.G. et al. *J. Biol. Chem.* 1992, **267**:23455-23458.
Ferguson II, J.E. et al. *Proc. Natl. Acad. Sci. USA* 1992, **89**:8384-8388.
Grill, V. et al. *Clin. Endocrinol.* 1992, **37**:405-410.
Hendy, G.N. and Goltzman, D. In:*Parathyroid hormone-related protein: Normal physiology and its role in cancer.* Halloran, B.P. and Nissenson, R.A. (eds.), CRC Press, Boca Raton 1992, pp. 25-55.
Horiuchi, N. et al. *J. Bone Miner. Res.* 1990, **5**:541-545.
Ikeda, K. et al. *J. Biol. Chem.* 1989, **264**:15743-15746.
Jüppner, H. et al. *Science* 1991, **254**:1024-1026.
Kiriyama, T. et al. *Mol. Cell. Endocrinol.* 1992, **92**:55-62.
Kong, X.F. et al. *Biochem. Biophys. Res. Commun.* 1994, **200**:1290-1299.
Kremer, R. et al. *J. Clin. Invest.* 1991, **87**:884-893.
Kukreja, S.C. et al. *J. Clin. Invest.* 1988, **82**:1798-1802.
Lu, C. et al. *Mol. Endocrinol.* 1989, **3**:2034-2040.
Mangin, M. et al. *Proc. Natl. Acad. Sci. USA* 1988, **85**:597-601.
Mangin, M. et al. *Proc. Natl. Acad. Sci. USA* 1989, **86**:2408-2412.
McCuaig, K.A. et al. *Proc. Natl. Acad. Sci.* 1994, **91**:5051-5055.
Muff, R. et al. *Annu. Rev. Physiol.* 1992, **54**:67-79.
Musso, M.J. et al. *Europ. J. Pharmacol.* 1989, **174**:139-151.
Pausova, Z. et al. *Genomics* 1994, **20**:20-26.
Philbrick, W.M. et al. *Physiol. Rev.* 1996, **76**:127-173.
Rodda, C.P. et al. In: *Parathyroid hormone-related protein: Normal physiology and its role in cancer.* Halloran, B.P. and Nissenson, R.A. (eds.), CRC Press, Boca Raton 1992, pp.169-196.
Schipani, E. et al. *Endocrinology* 1993, **132**:2157-2165.
Soifer, N.E. and Stewart, A.F. In: Parathyroid hormone-related protein: Normal Physiolgy and its role in cancer. Halloran, B.P. and Nissenson, R.A. (eds.), CRC Press, Boca Raton 1992, pp.93-143.
Suva, L.J. et al. *Science* 1987, **237**:893-896.
Thiede, M.A. et al. *Proc. Natl. Acad. Sci. USA* 1988, **85**:4605-4609.
Thiede, M.A. In: *Parathyroid hormone-related protein: Normal physiology and its role in cancer.* Halloran, B.P. and Nissenson, R.A. (eds.), CRC Press, Boca Raton, 1992, pp.57-91.
Yamamoto, M. et al. *Proc. Natl. Acad. Sci.USA* 1992, **89**:5326-5330.
Yasuda, T. et al. *J .Biol. Chem.* 1989, **264**:7720-7725.

Ref. for DNA/AA Sequences

Suva, L.J. et al. *Science* 1987, **237**:893-896.
Mangin, M. et al. *Proc. Natl. Acad. Sci. USA* 1989, **86**:2408-2412.
Thiede, M.A. et al. *Proc. Natl. Acad. Sci. USA* 1988, **85**:4605-4609.



Amino-terminal sequences of human parathyroid hormone (PTH) and human PTHrP

Pepsinogen

Kenji Takahashi

Synonyms	None
Abbreviations	PG, Pg
Classifications	Pepsin A: EC 3.4.23.1; pepsin C. EC 3.4.23.3
Description	The zymogen of pepsin, an aspartic proteinase, synthesized in the gastric mucosa. There are two types, pepsinogen A (PGA) and pepsinogen C (PGC) (or progastricsin), which are autocatalytically activated to pepsin A and pepsin C (or gastricsin), respectively, under acidic conditions by liberation of the NH_2-terminal 49-residue segment each. There are at least 5 PGA and 2 PGC isozymogens separable by chromatography or electrophoresis. PGC is also synthesized in the prostate (known as seminal pepsinogen).
Structure	The 3D-structure not determined.
Molecular Weight	40,300 (PGA) and 40,600 (PGC) (aa sequence).
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	Under acidic conditions, it becomes active and cleaves its own NH_2-terminal pro-segment intramolecularly to become pepsin. The resulting pepsin is a typical aspartic proteinase with an optimum pH at about 2, with a rather broad specificity, preferentially attacking peptide bonds involving amino acids with aromatic or bulky aliphatic side chains.
Coenzymes/Cofactors	None
Substrates	Pepsinogen itself (autocatalytic proteolysis under acidic conditions to produce pepsin, which hydrolyzes various proteins and peptides).
Inhibitors	Pepstatin. Competitive inhibition.
Biological Functions	The precursor form of pepsin which is responsible for digestion of food proteins in the stomach. The function of seminal pepsinogen is unknown.
Physiology/Pathology	Important for food protein digestion in the stomach; its low concentration may cause incomplete digestion of food proteins.
Degradation	Degraded by intestinal proteolytic enzymes in the digestive canal.
Genetics/Abnormalities	Complex inter-individual heterogeneity in PGA phenotypes exists, due to genetic polymorphism resulting from chromosome haplotypes containing different numbers of genes. Gene localization: PGA, chromosome 11, band q13; PGC, chromosome 6.
Half-life	Unknown
Concentration	Unknown

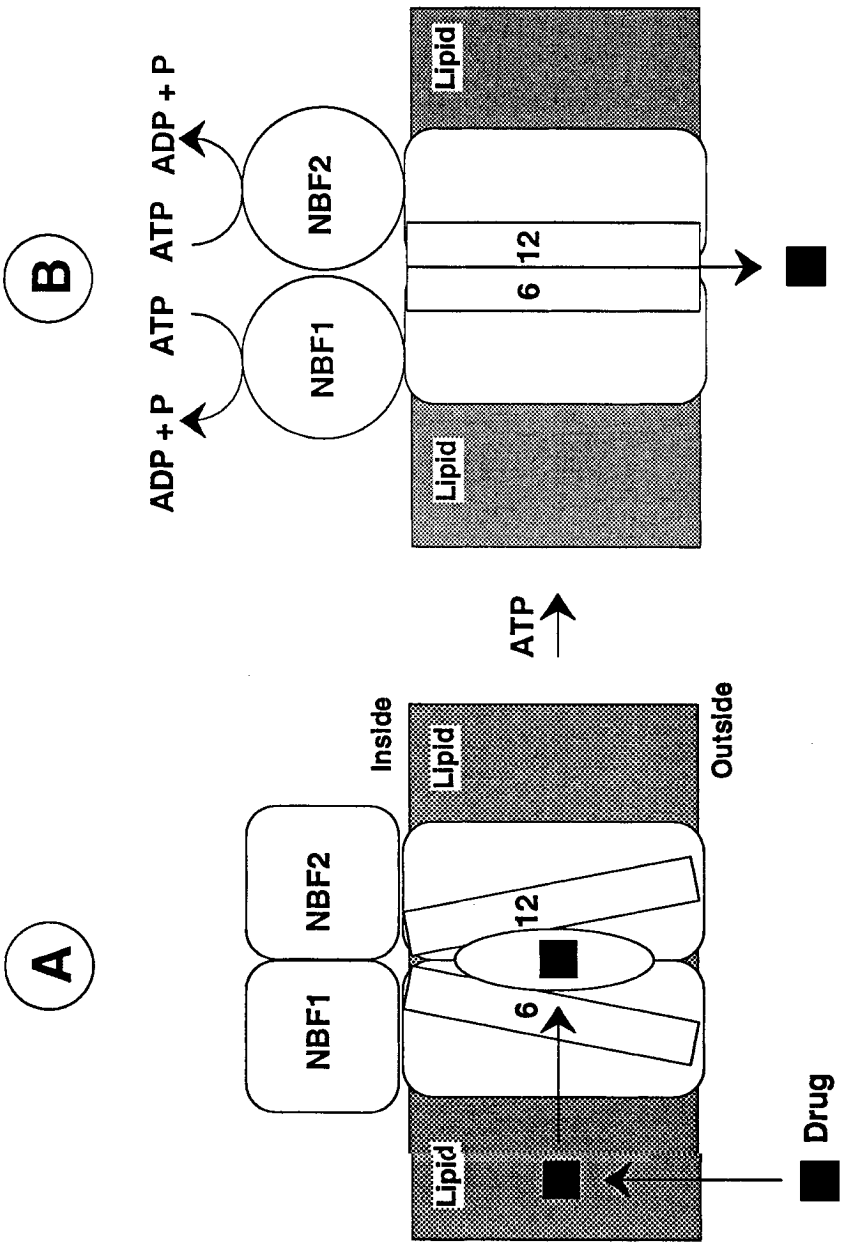
Isolation Method	Isolated from the mucosal extract by successive chromatographies on DEAE-cellulose and Sephadex G-100 (or Sephacryl S-200). Both types of pepsinogens can be further resolved into isozymogens by Mono Q chromatography.
Amino Acid Sequence	2 DTG sequences are present involving the 2 active site aspartic acid residues. Homologous with other aspartic proteinases, including renin, chymosin, cathepsins D and E and several microbial acid proteinases.
Disulfides/SH-Groups	3 intrachain disulfide bonds.
General References	Samloff, I. M. *Gastroenterol.* 1971, **60:** 586–604. Foltmann, B. *Essays in Biochem.* 1981, **17:** 52–84. Athauda, S. B. P. et al. *J. Biochen.* (Tokyo) 1989, **106:** 920–927.
Ref. for DNA/AA Sequences	Sogawa, K. et al. *J. Biol. Chem.* 1983, **258:** 5306–5311. (GenBank/EMBL, J00279-J00286). Hayano, T. et al. *J. Biol. Chem.* 1988, **263:** 1382–1385. (GenBank/EMBL, J03508). Taggart, R. T. et al. *J. Biol. Chem.* 1989, **264:** 375–379. (GenBank/EMBL, J04443).

P-Glycoprotein

Tip W. Loo and David M. Clarke

Synonyms	Multidrug transporter; Multidrug resistance protein.
Abbreviations	P-gp; MDR1 protein.
Classifications	Member of the ABC (ATP-Binding Cassette) family of transport proteins.
Description	A plasma membrane protein, comprising of a single polypeptide chain of 1280 aa with three carbohydrate side chains, N-linked to Asn-91, Asn-94 and Asn-99. The protein is expressed on the luminal surfaces of epithelial cells in kidney, small and large intestines, biliary hepatocytes, capillary endothelial cells of the brain, testes and placenta. Also found in high levels in the adrenal cortex, and in smaller amounts in the bone marrow stem cells and T-cells.
Structure	The tertiary structure is not yet determined. Structural predictions based on its 1280 aa sequence shows two tandem repeats of 610 aa joined by a linker region of 60 aa. Each repeat consists of an N-terminal hydrophobic domain containing six transmembrane-spanning helices followed by a hydrophilic domain that contains an ATP-binding site. This model is supported by topology studies. Both predicted ATP-binding sites are required for activity. The three glycosylated sites are on the first extracellular loop between transmembrane segments TM1 and TM2.
Molecular Weight	141,429 (calculated). On SDS-PAGE, the unglycosylated, core-glycosylated and mature forms of the enzyme migrate with apparent masses of approximately 140,000, 150,000 and 170,000, respectively.
Enzyme Activity	Mediates ATP-dependent transport of hydrophobic compounds. In the presence of verapamil and lipids, purified P-gp exhibits ATPase activity in the range of 1 to 5 μmoles Pi/min/mg enzyme.
Coenzymes/Cofactors	Requires lipids for stimulation of basal ATPase activity by hydrophobic compounds.
Substrates	A broad range of hydrophobic compounds with diverse structures. Examples include anticancer drugs such as actinomycin D, doxorubicin, taxol, vinblastine and vincristine; cytotoxic agents such as colchicine, puromycin and ethidium bromide; peptides such as gramicidin; calcium channel blocker, verapamil; immunosuppressants such as cyclosporin and steroid hormones. Natural substrate is unknown.
Inhibitors	Reversibly inactivated by vanadate through formation of Mg.ADP.vanadate complex at either ATP-binding site. Irreversible inactivation with thiol-reactive agents such as N-ethylmaleimide. Competitive inhibition of drug binding or transport by using high-affinity substrates. For example, cyclosporin A can be used as a competitive inhibitor in the transport of vinblastine or vincristine due to its higher affinity for P-gp.
Biological Functions	Unknown, but may act to protect the body from endogenous and exogenous cytotoxic agents. P-glycoprotein "knock-out mice" appear normal, but accumulate more cytotoxic compounds. May be involved in transporting steroids.

Physiology/Pathology	Overexpression in certain cancers (especially pediatric cancers) responsible for the phenomenon of multidrug resistance during chemotherapy. Studies on knock-out mice suggest P-gp is not essential, but organs such as the brain will accumulate high levels of some hydrophobic drugs, and kidney and liver show delayed excretion of some drugs.
Genetics/Abnormalities	The gene is on the long arm of chromosome 7, near 7q21.1, and consists of 28 exons with a total span of more than 100 kbp.
Half-life	2-3 days in studies with cultured cells.
Isolation Method	The enzyme can be purified by modifying its cDNA such that a polyhistidine tag is added to the C-terminal of P-gp, followed by expression in mammalian cells and nickel-chelate chromatography.
Amino Acid Sequence	MDLEGDRNGG AKKKNFFKLN NKSEKDKKEK KPTVSVFSMF RYSNWLDKLY MVVGTLAAII HGAGLPLMML VFGEMTDIFA NAGNLEDLMS NITNRSDIND TGFFMNLEED MTRYAYYYSG IGAGVLVAAY IQVSFWCLAA GRQIHKIRKQ FFHAIMRQEI GWFDVHDVGE LNTRLTDDVS KINEGIGDKI GMFFQSMATF FTGFIVGFTR GWKLTLVILA ISPVLGLSAA VWAKILSSFT DKELLAYAKA GAVAEEVLAA IRTVIAFGGQ KKELERYNKN LEEAKRIGIK KAITANISIG AAFLLIYASY ALAFWYGTTL VLSGEYSIGQ VLTVFFSVLI GAFSVGQASP SIEAFANARG AAYEIFKIID NKPSIDSYSK SGHKPDNIKG NLEFRNVHFS YPSRKEVKIL KGLNLKVQSG QTVALVGNSG CGKSTTVQLM QRLYDPTEGM VSVDGQDIRT INVRFLREII GVVSQEPVLF ATTIAENIRY GRENVTMDEI EKAVKEANAY DFIMKLPHKF DTLVGERGAQ LSGGQKQRIA IARALVRNPK ILLLDEATSA LDTESEAVVQ VALDKARKGR TTIVIAHRLS TVRNADVIAG FDDGVIVEKG NHDELMKEKG IYFKLVTMQT AGNEVELENA ADESKSEIDA LEMSSNDSRS SLIRKRSTRR SVRGSQAQDR KLSTKEALDE SIPPVSFWRI MKLNLTEWPY FVVGVFCAII NGGLQPAFAI IFSKIIGVFT RIDDPETKRQ NSNLFSLLFL ALGIISFITF FLQGFTFGKA GEILTKRLRY MVFRSMLRQD VSWFDDPKNT TGALTTRLAN DAAQVKGAIG SRLAVITQNI ANLGTGIIIS FIYGWQLTLL LLAIVPIIAI AGVVEMKMLS GQALKDKKEL EGSGKIATEA IENFRTVVSL TQEQKFEHMY AQSLQVPYRN SLRKAHIFGI TFSFTQAMMY FSYAGCFRFG AYLVAHKLMS FEDVLLVFSA VVFGAMAVGQ VSSFAPDYAK AKISAAHIIM IIEKTPLIDS YSTEGLMPNT LEGNVTFGEV VFNYPTRPDI PVLQGLSLEV KKGQTLALVG SSGCGKSTVV QLLERFYDPL AGKVLLDGKE IKRLNVQWLR AHLGIVSQEP ILFDCSIAEN IAYGDNSRVV SQEEIVRAAK EANIHAFIES LPNKYSTKVG DKGTQLSGGQ KQRIAIARAL VRQPHILLLD EATSALDTES EKVVQEALDK AREGRTCIVI AHRLSTIQNA DLIVVFQNGR VKEHGTHQQL LAQKGIYFSM VSVQAGTKRQ
Disulfides/SH-Groups	Does not appear to have disulfide bonds. Cys-less version of P-gp is still active. Modification of either Cys-431 or Cys-1074 by NEM inactivates P-gp. Other cysteine residues at positions 137, 717, 956, 1125 and 1227.
General References	Gottesman, M.M. et al. *Annu. Rev. Genet.* 1995, **29**:607-649. Ling, V. *Amer. J. Med.* 1995, **99**:315-345. Schinkel, A.H. et al. *Cell* 1994, **77**: 491-502. Urbatsch, I.L. et al. *J. Biol Chem.* 1995, **270:** 19383-19390. Loo, T.W. and Clarke, D.M. *J. Biol. Chem.* 1995, **270**:843-848. Loo, T.W. and Clarke, D.M. *J. Biol. Chem.* 1995, **270**:21449-21452. Loo, T.W. and Clarke, D.M. *J. Biol. Chem.* 1995, **270**:22957-22961.
Ref. for DNA/AA Sequences	Chen, C.J. et al. Cell 1986, **47**:381-389.



Proposed Model for P-glycoprotein-Mediated Drug Efflux. A. The hydrophobic substrate enters the lipid bilayer and interacts with residues in the transmembrane domain that form a drug-binding domain (oval). Recent evidence suggests that some of the substrate-binding residues reside in transmembrane segments 6 and 12. B. Upon ATP hydrolysis, there is a conformational change in P-gp that leads to drug efflux.

Phenylalanine hydroxylase

Randy C. Eisensmith and Savio L.C. Woo

Synonyms	L-Phenylalanine 4-monooxygenase
Abbreviations	PAH; PH
Classifications	EC 1.14.16.1
Description	PAH is a cytosolic ferroprotein. Despite initial reports of its presence in both liver and kidney, more recent evidence suggests that it is produced exclusively in the liver in man. Human PAH is a homopolymeric protein, although it is still unclear whether it is a homodimer or homotrimer in its native state.
Structure	Detailed structural studies of the native human protein have not yet been reported.
Molecular Weight	107,000 to 275,000 have been reported for the native protein, with most recent estimates clustered between 150,000 Mr values ranging from 49,000 to 54,000 have been reported for a single subunit of PAH, in general agreement with a total molecular weight of 51,868 as predicted from the cDNA sequence. Heavy (H) and light (L) subunits differ in electrophoretic mobility due to the incorporation of phosphate into the H subunit by cAMP-dependent protein kinase.
Sedimentation Coeff.	5.45 S
Isoelectric Point	5.0 - 5.6
Extinction Coeff.	Unknown
Enzyme Activity	Phenylalanine hydroxylase is a mixed-function oxygenase or monooxygenase enzyme that catalyzes the hydroxylation of L-phenylalanine to L-tyrosine.
Coenzymes/Cofactors	L-erythro-tetrahydrobiopterin (BH4)
Substrates	Phenylalanine is the primary substrate. Some reports have suggested that the rat enzyme can also hydroxylate the aromatic amino acid L-tryptophan, albeit to a lesser extent. It is not yet clear whether the human enzyme can utilize other substrates.
Inhibitors	Para-chlorophenylalanine (PCPA) is the most commonly used inhibitor of PAH. PCPA acts as a classical competitive inhibitor. Other para-halogenated phenylalanine derivatives may also act as inhibitors at high concentrations.
Biological Functions	PAH is the rate-limiting enzyme in the catabolicpathway that leads to the complete oxidation of L-phenylalanine to CO_2 and H_2O. Metabolic intermediates of this pathway, such as fumarate or acetoacetate, can enter the gluconeogenic pathway, the tricarboxylic acid cycle, or fatty acid biosynthetic pathway. The initial intermediate in this pathway is the non-essential aa L-tyrosine. In the absence of PAH activity, L-tyrosine can no longer be synthesized and thus becomes an essential component of the diet.

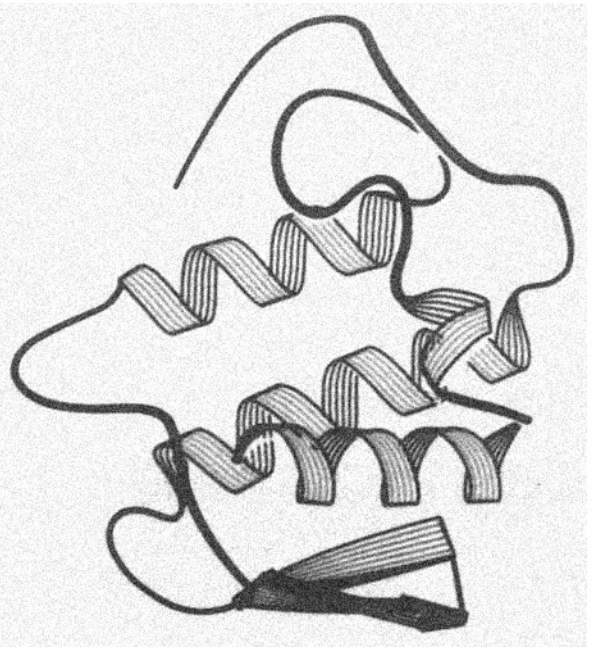
Physiology/Pathology	Decreased levels of PAH activity are the principle cause of the genetic disorder phenylketonuria (PKU). PKU patients can be detected initially by their significantly elevated serum phenylalanine levels, and later through their urinary excretion of large amounts of phenylpyruvate and other related compounds. The primary symptom of PKU is profound and irreversible mental retardation that develops early in life. The precise mechanisms responsible for the pathological effects of PAH deficiency are not yet known.
Degradation	PAH is degraded intracellularly in the liver. There are no physiologically important breakdown products of the PAH protein.
Genetics/Abnormalities	The gene encoding PAH is found on chromosome 12 (12q22-q24.1). To date, over 250 genetic variants of the PAH protein have been reported. Many of these mutant proteins are devoid of enzymatic activity and are usually present in individuals with severe forms of PKU. Other mutant proteins exhibit significant amounts of residual enzyme activity and are observed in individuals with mild disease phenotypes.
Half-life	Unknown
Concentration	Unknown
Isolation Method	The first complete purification of PAH from human liver tissue was performed using a combination of ammonium sulfate fractionation, DEAE-cellulose chromatography, and hydroxyapatite adsorption chromatography. More recent reports have described the purification of the human enzyme by substrate-induced hydrophobic chromatography using phenyl-Sepharose or by affinity chromatography using immobilized monoclonal antibodies.
Amino Acid Sequence	The aa sequence of PAH, as predicted from the primary sequence of its cDNA, shows significant homology to that of the closely related monooxygenase proteins, tryptophan hydroxylase and tyrosine hydroxylase, especially in the regions of the protein encoded by exons 6-11. It has been suggested that functions common to all three of these enzymes may lie within this region and that functions unique to each individual enzyme lie outside this region.
Disulfides/SH-Groups	The number of free sulfhydryl groups and disulfide bridges has not been determined. Modification of a single sulfhydryl group has been shown to produce significant activation of the rat enzyme. Similar studies on the human protein have not been reported.
General References	Abita, J.-P. et al. Phenylalanine 4-monooxygenase from human liver. *Meth. Enzymol.* 1987, **142**:27-35. Scriver, C.R. et al. The hyperphenylalaninemias. In: *The Metabolic and Molecular Bases of Inherited Disease.* Scriver, C.R. et al. (eds.), **Vol. I,** 7th ed. Mac-Graw-Hill, New York 1995, pp. 1015-1075. Eisensmith, R.C. and Woo, S.L.C. Molecular genetics of phenylketonuria: from molecular anthropology to gene therapy. In: *Advances in Genetics.* Hall, J.C. et al. (eds.), **Vol. 32** Academic Press, San Diego 1995, pp. 199-272.
Ref. for DNA/AA Sequences	Kwok, S.C.M. et al. *Biochemistry* 1985, **24**: 556-561. GenBank accession number K03020 (HUMPHH).

Phospholipase A_2

Fritz Märki

Synonyms	Phosphatide 2-acylhydrolase
Abbreviations	PLA_2; $sPLA_2$, secretory; $cPLA_2$, cytosol
Classifications	EC 3.1.1.4
Description	Family of ubiquitous enzymes occurring as soluble and/or membrane-bound proteins. $sPLA_2$ (14 kDa) classified into two groups based on disulfide pattern: group I (e.g. pancreas) contains Cys^{11}; group II (e.g. platelets) lacks Cys^{11}, contains instead half-cystine at carboxy terminus. Remarkably resistant against denaturation by heat and acid. $cPLA_2$ (50–110 kDa) translocates from cytosol to membrane in response to physiological increase of free Ca^{2+}. Acid-labile.
Structure	$sPLA_2$: Single polypeptide chain with disulfide crosslinks. Active site contains Ca^{2+} essential for enzymatic activity. X-ray structure of synovial fluid PLA_2 basically similar to that of snake venom and bovine pancreatic PLA_2.
Molecular Weight	$sPLA_2$: 13,903 (synovial fluid); 14,003 (pancreas); $cPLA_2$ 85.2 kDa (U 937 cells), all calculated from aa sequence. $sPLA_2$: 14–17 kDa (synovial fluid, platelets, pancreas); $cPLA_2$ 90 kDa (platelets); (56), 100, 110 kDa (U 937 cells), all determined by SDS-PAGE and gel filtration.
Sedimentation Coeff.	Unknown
Isoelectric Point	5.1 (U937); 8.7,9.2 (pancr.); > 10.5 (syn.fluid)
Extinction Coeff.	Unknown
Enzyme Activity	Ca^{2+}-dependent selective hydrolysis of 2-acyl ester bond of glycerophospholipids to produce free fatty acid and lysophospholipid.
Coenzymes/Cofactors	Ca^{2+}. Optimal concentration (in vitro assay): $sPLA_2$ 1–10 mM; $cPLA_2$ 0.1–1 μM.
Substrates	Phosphatidylethanolamine and phosphatidylcholine containing long chain fatty acids. Acyl ester in sn-1 may be replaced by O-alkyl or alkyl. $cPLA_2$ – in contrast to $sPLA_2$ – preferentially hydrolyzes substrates with arachidonic acid in sn-2. Reaction rates are markedly influenced by the physical form (vesicles, micelles) of the largely water-insoluble substrate. Group I $sPLA_2$ also accepts substrate in detergent (Triton X-100, deoxycholate) mixed micelles.
Inhibitors	Many lipophilic compounds inhibit $sPLA_2$ activity in vitro; the majority acts indirectly by physico-chemical interaction with the substrate, e.g. unsaturated long chain fatty acids, retinoids, phenothiazines, local anesthetics, mepacrine, lipocortin, duramycins, detergents. Manoalide and p-bromophenacyl bromide inhibit irreversibly by covalent modification of specific amino acids at the active site of the enzyme. Chelating agents (EGTA, EDTA) block the essential Ca^{2+}. Substrate analogues (e.g. acylamino phospholipids) and transition state mimics (e.g. phosphonate-containing phospholipids) may effectively compete with the substrate.

Biological Functions	Phospholipid Turnover: As component of deacylation-reacylation cycle involved in turnover of cellular phospholipids to remodel and maintain acyl composition (cell membrane homeostasis). Mediator Regulation: Provides the rate-limiting precursor for eicosanoid lipid mediator generation by releasing arachidonic acid from membrane phospholipids. Specialized Functions: Pancreatic PLA_2 contributes to food digestion by hydrolyzing dietary phospholipds. PMN PLA_2 participates in host defense by attacking membrane phospholipids of invading bacteria.
Physiology/Pathology	Essential for turnover and homeostasis of phospholipids and for regulation of eicosanoid mediator formation. Deficiency not known. Pathological increase in serum of patients with rheumatoid arthritis, pancreatitis, septic shock, uremia, malaria; in synovial fluid and peripheral blood PMN leukocytes in rheumatoid arthritis; in epidermis in psoriasis. PLA_2 activity sometimes correlates with severity of disease.
Degradation	Unknown
Genetics/Abnormalities	Pancreatic and synovial fluid PLA_2 contain four exons encoding 148 and 144 aa proteins, respectively (mature protein of 124 aa preceded by signal sequence starting with initiating methionine). $cPLA_2$ (U 937 cells) present as single copy gene; amino-terminal part encodes a Ca^{2+}-dependent phospholipid-binding domain homologous with corresponding domain in protein kinase C, phospholipase C-γ_1, protein p65 and GTP-activating protein.
Half-life	Unknown
Concentration	Pancreatic PLA_2 in serum (determined by fluoroimmunoassay): healthy control subjects 5.5 μg/L (range 1.8–9.2; N = 58); patients: rheumatoid arthritis 0.3–15 μg/L; acute pancreatitis 42.6 μg/L (SD 29.5); pancreatic cancer 29.2 μg/L (SD 21.3).
Isolation Method	Purified to near homogeneity from pancreas by salt fractionation and CM-Sephadex chromatography; from platelets, arthritic synovial fluid and U 937 cells by combination of chromatographic steps including HPLC.
Amino Acid Sequence	Determined: pancreas 1–125; platelets and synovial fluid 1–19 (identical); placenta 1–13 (identical with platelets 1–13). Deduced from DNA: lung, synovial fluid 1–126; U 937 cells 1–749. Partial homology among $sPLA_2$; no homology between $sPLA_2$ and $cPLA_2$.
Disulfides/SH-Groups	$sPLA_2$: 7 disulfides (different cross-linking for group I and II). $cPLA_2$ (U 937 cells): 9 cysteines.
General References	Van den Bosch, H. Intracellular Phospholipases A. *Biochim. Biophys. Acta* 1980, **604:** 191–246. Waite, M. The Phospholipases. In: *Handbook of Lipid Research*, Hanahan, D. J. (ed.) Plenum Press, New York 1987, pp. 1–332. Pruzanski, W. and Vadas, P. *Immunology Today* 1991, **12:** 143–146. Kaiser, E., et al. *Clin. Biochem.* 1990, **23:** 349–370.
Ref. for DNA/AA Sequences	Seilhamer, J. J., et al. *DNA* 1986, **5:** 519–527. Seilhamer, J. J., et al. *J. Biol. Chem.* 1989, **264:** 5335–5338. (GenBank/EMBL Data Bank, Acc. No. J04704). Kramer, R. M., et al. *J. Biol. Chem.* 1989, **264:** 5768–5775 (GenBank/EMBL Data Bank, Acc. No. J04705). Clark, J. D., et al., *Cell* 1991, **65:** 1043–1051.



Molecular structure of recombinant human rheumatoid synovial fluid phospholipase A_2 according to J.-P. Wery et al., Nature 1991, 352:79–82.

Phospholipase C

Fritz Märki

Synonyms	Phosphatidylcholine cholinephosphohydrolase, phosphatidylcholine phosphodiesterase / Phosphatidylinositol inositolphosphohydrolase, phosphatidylinositol phosphodiesterase / Triphosphoinositide inositoltrisphosphohydrolase, phosphatidylinositol bisphosphate phosphodiesterase.
Abbreviations	PLC
Classifications	EC 3.1.4.3 / EC 3.1.4.10 / EC 3.1.4.11
Description	Family of ubiquitous enzymes occurring as soluble and membrane-bound proteins. Ca^{2+}-dependent translocation from cytosol to membrane. Isozymes with similar catalytic properties, yet different structure and molecular size, classified as PLC-α, -β, -γ, -δ according to chromatographic behaviour and immunoreactivity.
Structure	PLC-α, -β, -γ, -δ are separate gene products with little sequence homology and with different molecular size.
Molecular Weight	Heterogeneous, especially when proteolysis not prevented during isolation. Platelet cytosol: 38, 45, 57, 98, 140–150 kDa (gel filtration, SDS-PAGE); oligomers of 146 kDa: 290, 440 kDa. Platelet membranes: 61–69 kDa (gel filtration, SDS-PAGE).
Sedimentation Coeff.	Unknown
Isoelectric Point	Platelet cytosol: 6.1, 7.5–8.0
Extinction Coeff.	Unknown
Enzyme Activity	Ca^{2+}-dependent hydrolysis of phosphatidylcholine to produce 1,2-diacylglycerol and choline phosphate. Ca^{2+}-dependent hydrolysis of phosphatidylinositol [PI] and/or its 4-mono- and 4,5-bisphosphate [PIP, PIP_2] to produce 1,2-diacylglycerol and inositol mono-, bis- or trisphosphate [IP, IP_2, IP_3], respectively, as well as small amounts of inositol 1,2-cyclic phosphate, inositol 1,2-cyclic 4-bisphosphate and inositol 1,2-cyclic 4,5-trisphosphate. Hydrolysis of PIP_2 by (some) membrane-bound PLCs is stimulated by GTP and GTP-γ-S.
Coenzymes/Cofactors	Ca^{2+}. Optimal concentration (in vitro assay) 1–100 µM free Ca^{2+}, depending on substrate.
Substrates	EC 3.1.4.3: phosphatidylcholine; EC 3.1.4.10: phosphatidylinositol [PI]; EC 3.1.4.11: phosphatidylinositol-4,5-bisphosphate (triphosphoinositide) [PIP_2]; phosphatidylinositol-4-phosphate (diphosphoinositide) [PIP].
Inhibitors	Mercurials (Hg^{2+}, 4-chloromercuribenzoic acid); gold complexes (e.g. [triethylphosphine] gold chloride and aurothiomalate); heparin; compound 48/80; neomycin.
Biological Functions	Participates in agonist-stimulated signal transduction across cell membranes by producing two second messengers, diacylglycerol (activates protein kinase C) and IP_3 (raises intracellular free Ca^{2+}). In some tissues activation of PLC by agonists is modulated by GTP-binding protein(s).

	Special function: Release of proteins covalently attached to the cell membrane via a glycophosphatidylinositol anchor, by selective cleavage of the anchor.
Physiology/Pathology	Essential link in agonist-induced transmembrane signalling. Pathological incrase of PLC activity observed in blood monocytes and PMNs in rheumatoid arthritis and in affected epidermis in psoriasis.
Degradation	Unknown
Genetics/Abnormalities	Unknown
Half-life	Unknown
Concentration	Unknown
Isolation Method	Some isozymes from platelet cytosol and membranes purified to near homogeneity by combinations of several chromatographic steps.
Amino Acid Sequence	So far only sequences of several isozymes from animal sources (e.g. bovine brain) determined; extensive sequence similarity between bovine brain and Drosophila.
Disulfides/SH-Groups	Free thiol group essential for enzymatic activity.
General References	Rhee, S. G., et al. Studies of inositol phospholipid-specific phospholipase C. *Science* 1989, **244:** 546–550. Fain, J. N. Regulation of phosphoinositide-specific phospholipase C. *Biochim. Biophys. Acta* 1990, **1053:** 81–88. Abdel-Latif, A. A. Calcium-mobilizing receptors, polyphosphoinositides, and the generation of second messengers. *Pharmacol. Reviews* 1986, **38:** 227–272. Rana, R. S. and Hokin, L. E. Role of phosphoinositides in transmembrane signalling. *Physiol. Reviews* 1990, **70:** 115–164.
Ref. for DNA/AA Sequences	Unknown

Plasma factor XIII

Thung-Shenq Lai, Komandoor E. Achyuthan, and Charles S. Greenberg

Synonyms	Blood coagulation factor XIII; Fibrin stabilizing factor; Plasma transglutaminase; Fibrinase; Fibrinoligase; Fibrin polymerase; Laki-Lorand Factor
Abbreviations	F XIII; F XIIIa
Classification	EC 2.32.13
Description	Plasma factor XIII is a zymogen composed of two A-chains and two B-chains that are associated noncovalently in the plasma. The protein is activated to a transglutaminase (factor XIIIa) by limited thrombin proteolysis of the A-chains, followed by calcium induced dissociation of the B-chains. Transglutaminases are calcium dependent enzymes that catalyze the formation of an intermolecular isopeptide bond involving the carboxamide group of specific peptide bound glutamine residues and the ε-amino group of a lysine in another protein or peptide. Primary amines (i.e. polyamines) can substitute for the ε-amino group of lysine. Factor XIII A-chain is predominantly found in platelets and accounts for 50% of the total factor XIIIa activity in whole blood. The A-chain is also present in monocytes, macrophages, placenta, megakaryocytes and hepatocytes. The hepatocytes synthesize the B-chains and secret the glycosylated protein into the plasma where it binds to the A-chains.
Structure	Plasma factor XIII is a tetramer composed of two A-chains and two B-chains noncovalently associated. Factor XIII A-chains were expressed in yeast and crystallized. Electron microscopy revealed that the A chains existed as globular particles, each 6 x 9 nm in size. The A2 dimer was elongated, 6 x 18 nm in size. In contrast, the B-chains appeared as flexible strands 2-3 nm in diameter and 30 nm long. A model for plasma factor XIII has been proposed where the A2-dimer forms the core and the two B-chains wrap around the A-chains. The X-ray crystal structure has revealed that the factor XIII A-chain is composed of four sequential domains; a b-sandwich (residue 43-184), a catalytic core region (residue 185-515), and two b-barrels (residues 516-628 and 629-727). The activation peptide (residues 1-37) of each A-subunit crosses the dimer interface and partially occludes the opening of the catalytic cavity in the second subunit.
Molecular Weight	326,000 (tetramer); A-chain 83,005-83,150 (aa composition and sequence data); B-chain 79,700 (aa composition of the unglycosylated proteins).
Sedimentation Coeff.	7.3S (A-chain); 3.7S (B-chain); 9.6S (A2B2)
Isoelectric Point	5.2-5.8 (A-chain); 5.1-5.6 (B-chain); 5.2 (A2B2)
Extinction Coeff.	13.8 (280 nm, 1% 1cm) for A2B2
Enzyme Activity	The enzyme commission on nomenclature defines the reaction as R-glutaminyl: Peptide-amine γ-glutamyl transferase. Alternate terms used include transglutaminase, transamidase and isopeptide bond forming activity. Transglutaminases catalyze the post-translational modifications of glutamine residues in specific proteins forming an εN-(γ-glutamyl)lysine bond between proteins. A calcium dependent acyl transfer reaction is catalyzed between the γ-carboxamide group of a peptide bound glutamine residue

and the ε-amino group in a peptide bound lysine or the primary amino group of a polyamine. Calcium ions induce a conformational change in the enzyme exposing the active center cysteine residue. A γ-glutamyl thioester is the acyl-enzyme intermediate between the glutamine substrate and the active center cysteine in the enzyme. The catalytic triad is composed of Cys-0314, His-373, and Asp-396 residues with a three-dimensional structure similar to the calcium-dependent proteinase calpain.

Coenzymes/Cofactors

Coenzymes or cofactors of the "classical" type are not known for plasma factor XIII. Calcium ions dissociate the noncatalytic B-chains from thrombin-cleaved A-chains and then unmask the thiol group in the active site cysteine which is required for enzymatic activity. Fibrin lowers the calcium concentration required for dissociation of the B-chains. Fibrin polymers also function to accelerate thrombin cleavage of plasma factor XIII. The surface of activated platelets bind factor XIIIa and also accelerate fibrin crosslinking. Fibrin polymerization also promotes the expression of the glutamine residue that participates in the intermolecular crosslinking of fibrin polymers.

Substrates

Fibrin is the major physiological substrate for Factor XIIIa. Gln-398, Gln-399, and Lys-406 in the γ-chain and Gln-328 and 366 and Lys-508, 556 and 562 in the α-chains were identified as crosslinking sites for Factor XIIIa resulting in the formation of γ-γ dimers, α-γ and α-chain polymers. Recent studies demonstrate that either Gln-398 or Gln-399 can serve as intermolecular crosslinking sites. In addition, fibrinogen, fibronectin, α_2-plasmin inhibitor, von Willebrand Factor, Factor V, thrombospondin, vinculin, actin, myosin, lipoprotein(a), plasminogen activator type II and vitronectin are also Factor XIIIa substrates.

Inhibitors

There are no known natural factor XIIIa inhibitors in plasma or tissues. The binding of factor XIIIa to fibrin inhibits factor XIIIa crosslinking of plasma proteins by localizing it to sites of blood clotting. Primary amines including putrescine, spermine, spermidine, histamine, and glycine ethyl ester compete for lysyl crosslinking sites in proteins and inhibit intermolecular protein-protein crosslinking in vitro. EDTA inhibits factor XIIIa activity by chelating essential Ca^{2+} ions. Iodoacetic acid alkylates the active site thiol of cysteine residue and inhibits catalytic activity. Several patients with bleeding disorders were reported to acquire antibodies to factor XIII A-chains or the binding site on fibrin resulting in inhibition of enzymatic activity.

Biological Functions

Plasma factor XIIIa is primarily responsible for crosslinking fibrin during blood coagulation. Factor XIIIa introduces intermolecular isopeptide bonds resulting in the formation of γ-γ dimers, α-chain polymers, α-γ heterodimer, and γ-chain multimers of fibrin. Crosslinking of fibrin by factor XIIIa results in a mechanically and chemically stable clot. Factor XIIIa also crosslinks α_2-plasmin inhibitor to the α-chain of fibrin. Crosslinked α_2-plasmin inhibitor protects fibrin from digestion by plasmin. Furthermore, Factor XIIIa also crosslinks fibronectin to itself and to fibrin and collagen. These reactions are also physiologically important and aid attachment of fibrin to the vessel wall and facilitate tissue repair and clot retraction.

Physiology/Pathology

Intracranial bleeding, umbilical stump bleeding, abnormal wound healing, unstable clot formation, and spontaneous abortion in untreated pregnant women are the chief problems of the factor XIII deficient patients. Intracranial bleeding is often fatal. Congenital factor XIII deficiency is treatable using either plasma, cyroprecipitate, or a placental derived concentrate of

	the A chains. Recombinant factor XIII A-chains may become available for replacement therapy in the near future.
Degradation	The in vivo fate of factor XIII is not known. Complexation of factor XIII A-chains with fibrinogen and clearance of the complex via the reticulo-endothelial system of the liver are the proposed degradation pathways. Factor XIII A-chains can be degraded by thrombin with loss of enzymatic activity. Elastase and trypsin are reported to degrade factor XIII A-chains leading to inactivation. There exists a secondary serine protease cleavage site located at Lys513-Ser514 that must be cleaved to get complete degradation of factor XIIIa.
Genetics/Abnormalities	Factor XIII deficiency is inherited as an autosomal recessive disorder. Congenital factor XIII deficiency is rare. Heterozygous individuals have 50 % levels of the A-chain protein and 50 % of normal plasma factor XIIIa activity with near normal (80%) levels of the B-chains. Individuals who are homozygously deficient (less than 1% of Factor XIIIa levels) have a severe bleeding disorder. Factor XIII A-chain gene was localized to bands p24 - 25 on chromosome 6. Factor XIII B-chain gene was located to chromosome 1 bands q3 1 - q32.1. Polymorphism was detected with both A- and B-chains of factor XIII based upon protein sequencing and isoelectric focusing. The factor XIII deficiency are mostly reported on the A chains and few cases on the B-chains and have been defined at the molecular level. In factor XIII A-chain deficiency, the defects are either due to a point mutation that results in premature termination of translation, defective splicing, aa substitution, or minor deletion that results in frame-shift mutation. B chain deficiency is reportedly caused by either a deletion of nucleotide at the splice junction of intron A, an aa substitution at Cys430-Phe, or an insertion of triplet (AAC) in exon III that leads to premature termination.
Half-life	12 days (range 10-14 days)
Concentration	A2B2 = 0.022 g/L; A-chain = 0.011 g/L; B-chain = 0.021 g/L
Isolation Method	The steps involved in the purification of plasma factor XIII include ammonium sulfate precipitation, DEAE- and gel-filtration chromatography and affinity chromatography on organomercurial resins. Purification of the B-chains is achieved by thrombin proteolysis of plasma factor XIII in presence of calcium ions followed by gel filtration to separate the A- and B-chains. Platelets and placenta are the best sources for purifying the A-chains. In addition, factor XIII A chains can also be expressed and purified from yeast and E. coli.
Amino Acid Sequence	The factor XIII A-chain consists of 731 aa residues with sequence homology to the keratinocyte transglutaminase, endothelial or tissue transglutaminase from guinea pig liver and the erythrocyte membrane protein, band 4.2. The most homologous areas include the N-terminal aa residues 31- 39, active site region residues 275 - 305 and the calcium binding site residues 427 - 455. The B-chain is composed of 641 aa residues with sequence homology to other complement proteins that reside on chromosome 1 and have tandem repeats containing 60 aa and two disulfide bonds. These repeats were first detected in b2-glycoprotein I and were subsequently found among others, in complement proteins, Clr, Cls, C2, C6, C7, C4 binding protein, protein H and decay accelerating factor.
Disulfides/SH-Groups	The A-chains have 9 cysteine residues but no intra- or inter-chain disulfide bonds. The active site of factor XIII has a cysteine residue at position 314.

The B-chains contain 40 cysteine residues crosslinked by 17-20 intra-chain disulfide bonds with no free thiol group in the protein.

General References

Lorand, L. Activation of Blood Coagulation Factor XIII. *Ann. N. Y. Acad. Sci.* 1986, **485**:144-158.

McDonagh, J. Structure and Function of Factor XIII. In*: Hemostasis and Thrombosis.* Colman, RW. et al. (eds.) J. B. Lippincott Company, Philadelphia, 1987, pp. 289-300.

Ichinose, A. and Davie, E.W. Structure of Transglutaminases. *J. Biol. Chem.* 1990, **265**:13411-13414.

Lorand, L. et al. Human Factor XIII: Fibrin Stabilizing Factor. *Progr. Hemost. Thromb.* 1980, **5**:245-290.

Curtis, C.G. Plasma Faclor XIII. In: *Hemostasis and Thrombosis.* Bloom, A.L. and Thomas, D.P. (eds.), Churchill Livingstone, New York, 1987, pp. 216-222.

Greenberg, C.S., Birckbichler, P.J., Rice, R.H. Transglutaminases: Multi-functional cross-linking enzymes that stabilize tissues. *FASEB J.* 1991, **5**:3071-3077.

Board, P.G., Losowsky, M.S., Miloszewski, K.J. Factor XIII: Inherited and acquired deficiency. *Blood Rev.* 1993, **7**:229-242.

Yee, V.C., Pedersen, L.C., Trong, I.L., Bishop, P.D., Stenkamp, R.E., Teller, D.C. Three dimensional structure of a transglutaminase: Human blood coagulation factor XIII. *Proc. Natl. Acad. Sci.* 1994, **91**:7296-7300.

Ref. for DNA/AA Sequences

c-DNA code for Factor XIII A chain in GenBank: M14354

Plasminogen

Johann Schaller and Egon E. Rickli

Synonyms	Profibrinolysin
Abbreviations	Pg (gene name Plg)
Classifications	Electrophoretic mobility: β-fraction.
Description	A circulating plasma protein, synthesized mainly in the liver. A monomeric glycoprotein with 1.8% carbohydrate. Variant 1 (approx. 50% of the molecules) contains a diantennary N-acetyllactosamine-type carbohydrate structure (partially sialylated) at Asn-289 and a mucin-type chain at Thr-346 (monosialo-tri- and disialo-tetrasaccharides). Variant 2 is only O-glycosylated. Traces of plasmin convert native plasminogen with N-terminal Glu (Glu-Pg) to Lys-Pg by elimination of an N-terminal 77 residue fragment ("pre-activation peptide"). Plasminogen has affinity for fibrin and omega-aminocarboxylic acids. Binding sites are located in kringle structures.
Structure	A globular, single-chain protein (791 residues) with 5 homologous kringle structures in N-terminal region. Conformations: α-helix 3%, β-sheet 35%, β-turns 28%, others 34%. Prolate ellipsoid 147Å x 57Å (small angle neutron scattering). Not yet crystallized.
Molecular Weight	88,384 Da (Glu-Pg); 79,612 Da (Lys-Pg) calculated from aa sequence (without carbohydrate).
Sedimentation Coeff.	5.7 S
Isoelectric Point	6.2 - 6.6
Extinction Coeff.	16.1 - 16.9 (280nm, 1%, 1cm).
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Participates in fibrinolysis: activation by activators (tissue-plasminogen activator, urokinase, streptokinase) converts plasminogen to plasmin (EC 3.4.21.7) by cleavage of the Arg-561-Val-562 bond and concomitant release of the "preactivation peptide". Plasmin consists of a heavy (A-) chain, containing 5 kringles and a light (B-)chain with the active site (trypsin-type catalytic chain). The chains are connected by 2 disulfide bonds. Natural substrate: fibrin. Attacks in vitro also other proteins, predominantly at Arg-bonds. Synthetic substrates: basic amino acid esters, Val-Leu-Lys-p-nitroanilide (specific). Natural inhibitor: α_2-antiplasmin. Synthetic inhibitors: serine protease inhibitors, omega-aminocarboxylic acids with variable efficiency.
Physiology/Pathology	As precursor of plasmin essential for fibrinolysis. Relatively constant level in adults, although somewhat higher in Africans and black Americans than

	in white Europeans and Americans. Decreased fibrinolytic potential of congenital abnormal plasminogens (active site deficiency, defects in the activation site) may lead to recurring thromboses, thrombophlebitis, pulmonary embolism.
Degradation	Unknown
Genetics/Abnormalities	Congenital abnormalities: Tochigi I + II, Chicago I, II + III, Tokyo I, Paris I, Nagoya, Frankfurt I. Thrombophilia.
Half-life	2.2 days (blood circulation)
Concentration	70-200 mg L^{-1} (plasma)
Isolation Method	Affinity chromatography with lysine-substituted Sepharose or Bio Gel.
Amino Acid Sequence	Sequence homology between the 5 kringles and also with kringle structures of other proteins (prothrombin, tissue-plasminogen activator, urokinase, factor XII, apo-lipoprotein A, Pg-related growth factors). C-terminal region of the polypeptide chain (plasmin B-chain) is homologous with other serine proteases. Active site residues: Ser-741, His-603, Asp-646 (plasminogen sequence).
Disulfides/SH-Groups	24 disulfides; no free sulfhydryls.
General References	Robbins, K.C., Summaria, L. *Meth. Enzymol.* 1976, **45**:257-273. Castellino, F.J., Powell. J.R. *Meth. Enzymol.* 1981, **80**:365-378. Miyashita, C. et al. *Haemostasis* 1988, **18**: Suppl 1, 7-13. Lorand, L. and Mann, K.G. *Meth.. Enzymol.* 1993, **223**:145-288. Section II Fibrinolysis.
Ref. for DNA/AA Sequences	Forsgren, M. et al. FEBS Lett. 1987, **213**:254-260 (DNA sequence). Petersen, T.E. et al. *J. Biol. Chem.* 1990, **265**:6104-6111 (gene structure). Sottrup-Jensen, L. et al. The primary structure of human plasminogen: Isolation of two lysine-binding fragments and one "mini"-plasminogen (MW 38,000) by elastase-catalyzes-specific limited proteolysis". In: *Progress in Chemical Fibrinolysis and Thrombolysis.* Davidson, J.F. et al. (eds.), Raven Press, New York 1978, **Vol. 3**, pp. 191-209. Wiman, B. Eur. J. Biochem. 1977, **76**:129-137.

Plasminogen activator inhibitor type-1

Daniel A. Lawrence and David J. Loskutoff

Synonyms	Fast acting inhibitor; Endothelial cell type plasminogen activator inhibitor; β–Migrating plasminogen activator inhibitor
Abbreviations	PAI-1
Classifications	serpin (**ser**ine **p**roteinase **in**hibitor), Electrophoretic mobility: β
Description	A single-chain glycoprotein (approx. 13% carbohydrate) produced as 402 aa precursor that is processed to a 379 or 381 aa (N-terminal heterogeneity) mature form that lacks cysteines. Present in plasma as both an active inhibitor bound to vitronectin and as an inactive complex with tissue-type plasminogen activator. Also present in the α-granules of platelets and in the extracellular matrix of a variety of cultured cells. Site of synthesis in man unknown. However, in normal mouse tissue, synthesis is detected primarily in vascular smooth muscle cells. PAI-1 is also produced in high concentrations by endothelial cells and hepatocytes in culture, and expression can be induced in these cells *in vivo* by systemic administration of endotoxin or TNF. Synthesis is highly regulated (e.g., stimulated by growth factors, cytokines, endotoxin, hormones, and proteinases).
Structure	The structure of PAI-1 is similar to that of other serpins, and like other serpins, active PAI-1 appears to be in a stressed conformation that is sensitive to thermal denaturation. Free PAI-1 is also biologically unstable in solution, and decays into an inactive (latent) form with a $T_{½}$ 1-2 hours at 37°C. This transition from the active to the latent form is associated with a large conformational change in the molecule that involves insertion of the exposed reactive center loop into β-sheet A. This converts β-sheet A from a 5 stranded to a 6 stranded primarily anti-parallel β-sheet. The stokes radius of active PAI-1 is reported to be 2.77 nm, and the latent conformation 2.40 nm.
Molecular Weight	46,000 - 50,000 (SDS-PAGE)
Sedimentation Coeff.	Unknown
Isoelectric Point	4.5-5.0: bovPAI-1; 7.0: recombinant (*E. coli*) hPAI-1
Extinction Coeff.	0.93 (active) and 0.62 (latent) (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	In plasma and the subcellular matrix active PAI-1 is bound to vitronectin. This interaction stabilizes the PAI-1 active conformation, increasing the $T_{½}$ for conversion to the latent conformation approximately 2-fold in solution and up to 10-fold on the matrix. This interaction also enhances PAI-1 reactivity with thrombin accelerating the second-order rate constant approximately 300-fold. Heparin also binds to PAI-1 and enhances its reactivity with thrombin approximately 10-fold.
Substrates	Tissue-type plasminogen activator; urokinase-type plasminogen activator. PAI-1 also inhibits activated protein C, thrombin, plasmin, and trypsin. However, these latter interactions are not as efficient as the reaction of PAI-1 with plasminogen activators, and their biological significance

remains to be determined. Inhibition is associated with the formation of SDS-stable-enzyme-inhibitor complexes.

Inhibitors

Both neutrophil elastase and cathepsin G inactivate PAI-1 through catalytic cleavage of the reactive center loop. Three synthetic peptides related to the PAI-1 reactive center loop have been reported to inactivate PAI-1. Two of these insert into β-sheet A and prevent the native reactive center loop from inserting during the inhibition reaction and thus convert PAI-1 into a proteinase substrate. The third peptide appears to accelerate PAI-1's conversion to the latent conformation. The mechanism of this interaction is not known. Three small molecule inhibitors of PAI-1 all based on diketopiperazine have also been described. Their mechanism of action is not known.

Biological Functions

PAI-1 is the primary inhibitor of tissue-type and urokinase-type plasminogen activators. As such it regulates fibrinolysis and may influence other processes employing plasminogen activation (e.g., ovulation, tumor cell invasion, angiogenesis, tissue remodeling, etc.). PAI-1 released from platelets protects platelet-rich thrombi from premature lysis, and fibrin associated PAI-1 may also prevent premature clot lysis. In the extracellular matrix PAI-1 bound to vitronectin prevents the binding of integrins and uPAR to vitronectin, and inhibits cell adhesion and migration on vitronectin. Matrix associated PAI-1 may also protect the cell substrate from degradation by cellular proteinases. PAI-1 circulates in blood complexed to vitronectin, and is also reported to bind to fibrin. Vitronectin bound PAI-1 is an efficient inhibitor of thrombin and can accelerate thrombin clearance via members of the low density lipoprotein receptor family such as LRP.

Physiology/Pathology

Increased levels of PAI-1 in the circulation are associated with thrombotic disease, including myocardial infarction and deep vein thrombosis. In addition, elevated preoperative levels of PAI-1 have been correlated with postoperative deep vein thrombosis. Reduced postoperative fibrinolytic activity has also been correlated with an increase in PAI-1 activity immediately following surgery. PAI-1 deficiency is associated with abnormal bleeding. PAI-1 probably functions in tissues to regulate and limit plasminogen activation during various invasive and/or degradative processes.

Degradation

PAI-1 in complex with a proteinase is cleared very efficiently via the low density lipoprotein receptor related protein (LRP) on many different cell types. The interaction of latent or uncomplexed active PAI-1 with LRP is not clear. The site of clearance *in vivo* may be the liver.

Genetics/Abnormalities

PAI-1 is produced by a single gene of approximately 12.3 kbp in length (8 introns) located on chromosome 5. Human cells produce two PAI-1 mRNA transcripts (2.3 and 3.2 kbp) suggesting that PAI-1 may be synthesized from two different mRNAs. Four cases of either partial or complete PAI-1 deficiency have been reported in humans, all associated with abnormal bleeding.

Half-life

The reported half-life is approximately 6 to 7 minutes for human PAI-1 infused into rabbits. In other studies comparing the clearance of active and latent PAI-1 the active form was cleared in a biphasic manner with half-lives of 6 and 25 minutes, while latent PAI-1 was cleared with a half-life of only 1.7 minutes.

Concentration	Highly variable plasma levels. Antigen ranges between 6μg/L to 86μg/L with geometric mean at 24μg/L.
Isolation Method	Isolated from serum-free conditioned media of endothelial cells, HAT-1080 human fibrosarcoma cells, melanoma cells, hepatoma cells, WI 38VA13 cells, as well as from porcine platelets. Purification approaches in general employ concanavalin A and/or heparin affinity chromatography followed by either chromatography on hydroxylapatite, gel filtration, or HPLC. Also purified utilizing specific monoclonal antibodies. Recombinant PAI-1 has been isolated from *E. coli* by heparin affinity chromatography followed by hydrophobic interaction chromatography on phenyl-sepharose.
Amino Acid Sequence	PAI-1 shows greater than 30% homology to several other members of the serpin family. The reactive center loop containing the P_1-$P_{1'}$ residues (Arg_{346}-Met_{347}) is located in the C-terminal region of the molecule.
Disulfides/SH-Groups	None
General References	Loskutoff, D.J. et al. Type 1 plasminogen activator inhibitors. In: *Progress in Hemostasis and Thrombosis*, Coller, B. (ed.) 1989, **Vol 9** p. 87-115. Andreasen, P.A. et al. *Mol. Cell. Endocrinol.* 1990, **68**:1-19. Lawrence, D.A. and Ginsburg, D. Gene Expression and Function of Plasminogen Activator Inhibitor-1 In: *Fibrinolysis in Disease: Molecular and Hemovascular Aspects of Fibrinolysis*. Glas-Greenwalt, P. (ed.), CRC Press Inc., Boca Raton FL., pp 21-29, 1995. van Meijer, M. and Pannekoek, H. *Fibrinolysis* 1995, **9**:263-276.
Ref. for DNA/AA Sequences	GenBank accession number: M16006. Ny, T. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:6776-6780. Pannekoek, H. et al. *EMBO J.* 1986, **5**:2539-2544. Andreasen, P.A. et al. *Mol. Cell Biol.* 1987, **7**:3021-3025. Ginsburg, D. et al. *J. Clin. Invest.* 1986, **78**:1673-1680. Loskutoff, D.J. et al. Biochem. 1987, **26**:3763-3768. Bosma, P.J. et al. *J. Biol. Chem.* 1988, **263**:9129-9141. Strandberg, L. et al. *Eur. J. Biochem.* 1988, **176**:609-616.

Plasminogen activator inhibitor type-2

Egbert K.O.Kruithof

Synonyms	Placental plasminogen activator inhibitor; Minactivin
Abbreviations	PAI-2
Classifications	None
Description	PAI-2 is an efficient inhibitor of urokinase and to a lesser extent of two chain tissue type plasminogen activator.
Structure	Unknown
Molecular Weight	There are two forms of PAI-2 a 60 kDa glycosylated, secreted form and a 47 kDa nonglycosylated intracellular form.
Sedimentation Coeff.	Unknown
Isoelectric Point	Glycosylated ≈ 4.5; nonglycosylated ≈ 5.0
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	PAI-2 is an efficient inhibitor of free and cell-surface bound urokinase. Since urokinase plays a key role in extracellular proteolysis in processes such as tissue remodelling, cell migration, inflammation, trophoblast implantation, fibrinolysis and cancer invasion and metastasis, it is likely that PAI-2 constitutes one of the regulators of these processes.
Physiology/Pathology	The major physiological role of secreted PAI-2 may be the inhibition of urokinase. The function of the intracellular form of PAI-2 is not yet known; several reports suggested that intracellular PAI-2 may protect from apoptosis. In human plasma PAI-2 normally is undetectable (< 10 µg/L) but during pregnancy PAI-2 levels may be as high as 300 µg/L. Low PAI-2 levels during pregnancy seem to be associated with a poor placental function and a low birth weight. In acute (myelomonocytic and monoblastic) leukemias as well as chronic myelomonocytic leukemias high plasma levels of PAI-2 are encountered.
Degradation	Unknown
Genetics/Abnormalities	Unknown
Half-life	Unknown
Concentration	In human plasma below 10 µg/L. In pregnancy PAI-2 may be as high as 300 µg/L and in myelomonocytic and monoblastic leukemias as high as 150 µg/L. After stimulation of monocyte-like cells, fibroblast-like cells and

	endothelial cells with phorbol myristate acetate intracellular PAI-2 levels may be over 1 μg per 10^6 cells and secreted PAI-2 over 50 ng per 10^6 cells per 24h.
Isolation Method	From phorbol ester stimulated human U-937 monocyte-like cells. PAI-2 has also been produced by recombinant DNA technology from yeast (Delta Biotechnology, Nottingham, Great Britain) or from E.coli (Biotechnology Australia Ltd, Sidney, Australia).
Amino Acid Sequence	PAI-2 has been partially sequenced and the complete sequence has been derived from its cDNA sequence. PAI-2 is a member of the serpin family of proteins.
Disulfides/SH-Groups	Probably two intrachain disulfides.
General References	Kruithof, E.K.O. *J. Biol. Chem.* 1986, **261**:11207-11213. Kruithof, E.K.O. *Enzyme* 1988, **40**:113-121. Belin, D, et al. *EMBO J.* 1989, **8**:3287-3294. Kruithof, E.K.O. et al. *Blood* 1995, **86**:4007-4024.
Ref. for DNA/AA Sequences	Schleuning, W.D. et al. *Mol. Cell. Biol.* 1987, **7**:4564-4567. Ye, R.D. et al. *J. Biol. Chem.* 1987, **262**:3718-3725. Webb, A.C. et al. *J. Exp. Med.* 1987, **166**:77-94. Kruithof, E.K.O., Cousin, E. *Biochem. Biophys. Res. Commun.* 1988, **156**: 383-388. Swissprot: human P01520, mouse P12388; EMBL: HSPAI2B.

Platelet basic protein

Kenneth J. Clemetson

Synonyms	(all these are now known to be fragments) connective tissue activating peptide-III (CTAP-III), low affinity platelet factor 4 (LA PF4), β-thromboglobulin (β-TG), neutrophil activating peptide-2 (NAP-2).
Abbreviations	PBP, CTAP-III, β-TG, NAP-2
Classifications	CTAP-III, mitogen; NAP-2, cytokine
Description	Synthesized in megakaryocytes, stored in platelet α-granules, and released on platelet activation. Already largely degraded to CTAP-III in platelet α-granules and rapidly degraded to β-TG in the blood stream.
Structure	PBP is a single, non-glycosylated peptide chain. Produced as a 128 aa precursor containing a 34 aa leader sequence. Largely globular molecule with a domain containing two disulfide bonds very similar to other molecules in this superfamily. Not yet crystallized but other closely related members of the superfamily (platelet factor 4, neutrophil activating peptide-1) have been.
Molecular Weight	PBP: 10,414 Da; CTAP-III: 9278 Da; β-TG: 8851 Da; NAP-2: 7416 Da (all from sequence data).
Sedimentation Coeff.	Unknown
Isoelectric Point	PBP, 10.0; CTAP-III, 8.0; β-TG, 6.7.
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Function of PBP not known. CTAP-III has synovial cell stimulating activity. β-Thromboglobulin is inactive or its function is not yet known. NAP-2 has strong neutrophil activating properties which may be important in local inflammatory responses.
Physiology/Pathology	Raised levels of β-thromboglobulin in plasma are extensively used as an indicator that platelet activation and granule release have occured and as such are a good indicator of various pathological conditions.
Degradation	Degradation to CTAP-III already starts in the platelet α-granules and after release, depending upon the proteases encountered can give rise to various biologically active or inactive peptide forms, before final clearance through the kidneys. Cathepsin G produces NAP-2, while elastase causes rapid degradation.
Genetics/Abnormalities	There exists evidence for at least one variant gene as has been found with platelet factor 4. The function of this variant is not yet established. No mutants are known.

Half-life	7.6 min; 56.6 min (two clearance pathways)
Concentration	Plasma 30 ± 10 ng/L (measured as β-TG antigen)
Isolation Method	Isolated from the supernatant of activated platelets (by thrombin or ionophore A23187) by fractionation on CM-Sephadex, heparin-agarose and Sephadex G-75. Reverse phase HPLC is an effective analytical method to separate the various species and can also be used for small scale preparations.
Amino Acid Sequence	SSTKGQT KR@NL AK#GK E ESLDSD LY*AELRC MCIKTTS-GIHPKN IQSLEVIGK GTHCNQVEVI ATLKDGRKIC LDPDAPR IKKIVQK KLAGDES AD N-terminus from @, CTAP-III; #, β-TG; *, NAP-2. The sequence is homologous to those of platelet factor 4, gro/MGSA and NAP-1 and a growing number of other peptides in this superfamily. Both PBP and PF4 have unusually long (34 for PBP) leader sequences that may be involved in the targeting to the α-granules. The PBP genomic DNA consists of 4 exons and 3 introns. The gene maps to a locus on chromosome 4q12–13 containing a cluster of the related genes mentioned above.
Disulfides/SH-Groups	Two intrachain disulfide bridges, Cys29-Cys55 and Cys31-Cys71.
General References	Holt, J. C. et al. Characterisation of human platelet basic protein, a precursor form of low affinity platelet factor 4 and β-thromboglobulin. *Biochemistry* 1986, **25:** 1988–1996. Holt, J. C. and Niewiarowski, S. Platelet basic protein, low affinity platelet factor 4, and β-thromboglobulin: Purification and identification. In: *Meth. Enzymol.* (ed. Hawiger, J.) 1989, **169**, pp.224–232, Academic Press Inc., San Diego. Walz, A. and Baggiolini, M. A novel cleavage product of β-thromboglobulin formed in cultures of stimulated mononuclear cells activates human neutrophils. *Biochem. Biophys. Res. Commun.* 1989, **159**:969–975.
Ref. for DNA/AA Sequences	Wenger, R. H., et al. (1989) Cloning of cDNA coding for connective tissue activating peptide-III from a human platelet-derived λ gt11 expression library. *Blood* 1989, **73:** 1498–1503. Majumdar, S., et al. Isolation and characterisation of the human β-thromboglobulin gene: Comparison with the gene for platelet factor 4. *Blood* 1990, **76:** 465a.

Platelet-derived growth factor

Carl-Henrik Heldin

Synonyms	None
Abbreviations	PDGF
Classifications	None
Description	Originally identified in platelet α-granules. Normally not circulating freely in plasma.
Structure	Disulfide-bonded dimers of homologous A- and B-polypeptide chains. All possible dimers, -AA, -AB and -BB have been identified. Synthesized as precursor molecules that undergo processing in the N-terminus and, in the case of the B-chain, also in the C-terminus. The A-chain occurs as two variants due to differential splicing of the gene product; the three carboxy-terminal aa residues in the short variant are replaced in the long variant by 18 different aa.
Molecular Weight	≈ 30,000 for the native dimeric protein (SDS-PAGE). The short version of the A-chain precursor and mature form contain 196 and 110 aa, respectively; the long version of the A-chain precursor and mature form 211 and 125 aa, respectively; the B-chain precursor and mature form 241 and 109 aa, respectively.
Sedimentation Coeff.	Unknown
Isoelectric Point	≈ 10
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Unknown
Biological Functions	Stimulates cell proliferation of connective tissue cells and glial cells through interaction with two distinct cell surface receptors (α- and β-receptors), both endowed with protein tyrosine kinase activities. PDGF induces dimerization of its receptors; the A-subunit binds α-receptors whereas the B-subunit binds both α- and β-receptors.
Physiology/Pathology	Stimulates growth of the placenta and embryo during the development, and wound healing in the adult. May also be involved in adverse reactions involving cell proliferation, such as atherosclerosis, fibrosis and in malignancies. The B-chain of PDGF is the normal counterpart of the Sis oncogene product.
Degradation	Unknown
Genetics/Abnormalities	Unknown

1. Installment 1992 (revised 1997)

Half-life	Cleared within min. from the blood circulation.
Concentration	Unknown
Isolation Method	Isolated from platelets or conditioned media from cultured cells by chromatographies on ion exchangers, hydrophobic matrices, sizing gels and by reversed-phase HPLC.
Amino Acid Sequence	Homologous to that of vascular endothelial cell growth factor/vascular permeability factor.
Disulfides/SH-Groups	Each polypeptide chain contains 8 cysteine residues. Two of the cysteine residues form interchain disulfide bonds, and the remaining six form three disulfides in a tight knot-like structure.
General References	Raines, E.W. et al. Platelet-derived growth factor. In: *Handbook of Experimental Pharmacology,* **Vol. 95**, part I, Sporn, M.B. and Roberts, A.B. (eds.) Springer-Verlag, Heidelberg, 1990, pp. 173-262. Heldin, C.-H. and Westermark, B. Platelet-derived growth factor: mechanism of action and possible in vivo function. *Cell Regulation* 1990, **1**:555-566.
Ref. for DNA/AA Sequences	Betsholtz, C. et al. cDNA sequence and chromosomal localization of human platelet-derived growth factor A-chain and its expression in tumour cell lines. *Nature* 1986, **320**:695-699. Josephs, S.F. et al. Transforming potential of human c-sis nucleotide sequences encoding platelet-derived growth factor. *Science* 1984, **225**:636-639.

Platelet factor 4

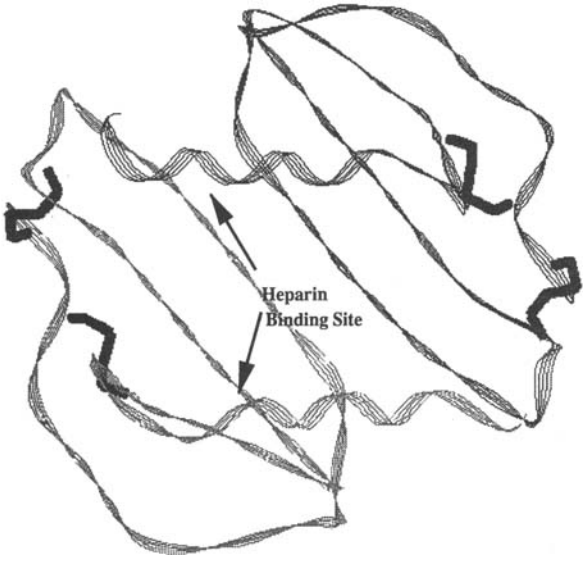
Stefan Niewiarowski and Bradford A. Jameson

Synonyms	High affinity platelet factor 4; Platelet antiheparin activity; Heparin neutralizing factor
Abbreviations	PF4
Classifications	C x C cytokine family
Description	Specific platelet protein synthesized in megakaryocytes and stored in platelet α-granules. It is released from activated platelets in a complex with proteoglycan carrier and cleared through the liver.
Structure	PF4 tetramer contains two extended six stranded β sheets, each formed by two subunits.
Molecular Weight	7,800 Da (aa sequencing, PF4 monomer). PF4 exists as a tetramer under physiological conditions .
Sedimentation Coeff.	7.1 S (0.15 M NaCl); 2.4 S (0.75 M NaCl)
Isoelectric Point	7.6
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Heparin
Biological Functions	Neutralization of heparin anticoagulant activity, reversal of immunosuppression in mice, inhibition of megakaryocyte maturation, inhibition of angiogenesis, chemotactic activity vs neutrophils by PF4 derivatives.
Physiology/Pathology	Stimulates fibroblasts attachment to substrata, inhibits tumor growth in nude mice, inhibits endothelial cell proliferation, enhances affinity of thrombomodulin-thrombin complex for protein C.
Degradation	Normaly eliminated from circulation by the liver. Uptake and degradation by endothelial cells and hepatocytes.
Genetics/Abnormalities	Human PF4 gene is mapped to 4q12-q21 (is proximal to the breakpoint on chromosome 4). Structure of human and rat PF4 genes have been determined. PF4 mutant (PF4 arl) has been identified. Significance unknown.
Half-life	t 1/2 fast = 1-2 min; t1/2 slow = 20-80 min
Concentration	1.8-16 μg L^{-1} in platelet poor plasma; 11.2-12.4 μg/10^9 platelets.
Isolation Method	Heparin-agarose chromatography of supernatant from outdated platelet concentrates followed by reverse phase HPLC.

1. Installment 1992 (revised 1997)

Amino Acid Sequence	Composed of 70 aa. C-terminal sequence PLYKKIIKKLLES critical for heparin binding.
Disulfides/SH-Groups	Two disulfide bridges (Cys-10-Cys-36 and Cys-12-Cys-52) are critical for high affinity heparin binding.
General References	Holt, J.C., Niewiarowski, S. *Semin. Hemat.* 1985, **22**:151-163. Kawahara, S., Deuel, T.F. *J. Biol. Chem.* 1989, **264**:679-682. St.Charles, R. *J. Biol. Chem.* 1989, **264**:2092-2099. Zucker, M.B. and Katz, I.R. *Proc. Soc. Exp. Biol. Med.* 1991, **198**:693-702.
Ref. for DNA/AA Sequences	Poncz, M. et al. *Blood* 1987, **69**: 219-223. Doi T. et al. *Mol. Cell Biol.* 1987, **7**:898-904. Shigeta, O. et al. *Thromb. Res.* 1991, **64**:509-520.

Human Platelet Factor 4 Dimer



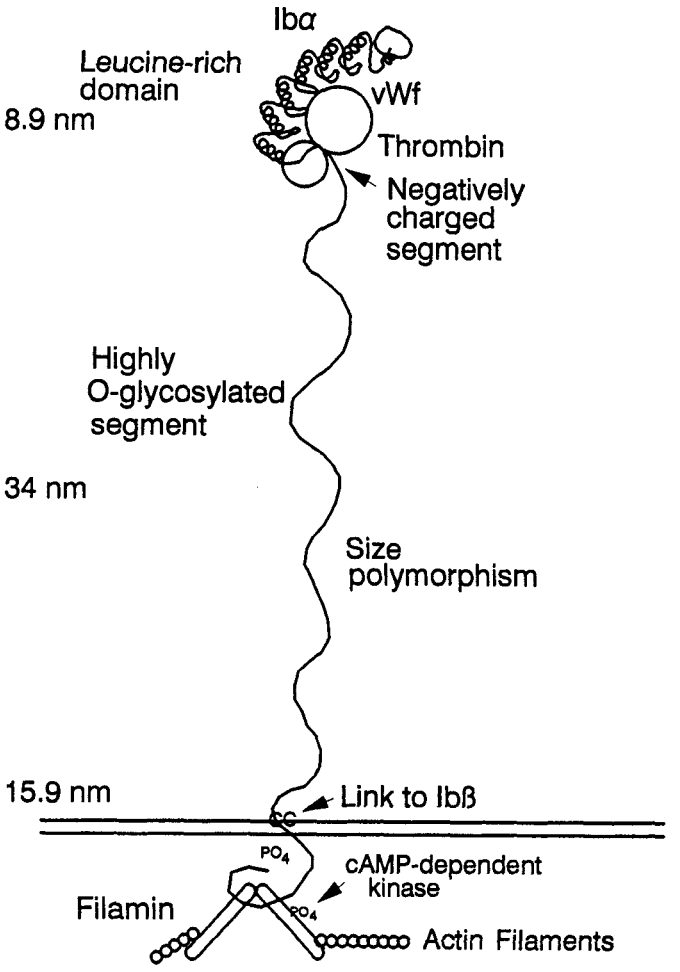
Platelet Membrane Glycoprotein Ibα (GPIbα)

Kenneth J. Clemetson

Synonyms	CD42b, Glycocalicin (extracellular fragment, see below)
Abbreviations	GPIbα
Classifications	Leucine-rich repeat family
Description	GPIbα is part of a complex consisting of GPIb, GPV and GPIX (referred to as GPIb-V-IX) which is one of the major glycoprotein complexes on the platelet surface (about 25–50,000 copies per platelet) and carries most of the sialic acid and therefore the charge. All of these subunits are of the type 1 membrane protein class with a major N-terminal extracellular domains, a single transmembrane region and a C-terminal cytoplasmic domain. The cytoplasmic domains are linked to the membrane-associated platelet cytoskeleton via actin binding protein (filamin). GPIb-V-IX expression is restricted to platelets, megakaryocytes (where it is a marker of maturation) and to vascular and tonsillar endothelial cells that have been stimulated with TNF and/or IFNγ.
Structure	Human GPIbα consists of 610 aa residues. The N-terminal domain contains a disulfide loop and is similar to other leucine-rich repeat protein N-terminal flanking regions. The following approximately 150 aa residues make up 6 leucine-rich repeats, 24 aa long, arranged in an arc, followed by two loops of 40 and 20 aa held together by disulfide bridges formed by cysteines 209–248 and 211–264, respectively. This is followed by a highly negatively charged linear segment about 23 aa long, containing 3 sulphated tyrosine residues, followed by a short positively charged sequence. This linear region contains the primary cleavage site on GPIbα for several physiologically important proteases, such as elastase, cathepsin G and plasmin and a secondary site for calpain. The next region of about 150 aa contains five (in the smallest polymorphism) 8 aa repeats and is rich in serine and threonine residues many of which are O-glycosylated. The next 40 aa, just above the membrane, also contain proteolytic cleavage sites, in particular the primary cleavage site for calpain which liberates the major water-soluble fragment, glycocalicin. Just on the extracellular side of the membrane are two cysteine residues, one of which (probably the outer) forms the disulfide link to GPIbβ. The transmembrane domain contains 29 aa and is followed by the cytoplasmic C-terminal domain containing 95 aa.
Molecular Weight	145 kDa (reduced GPIbα chain), 67 kDa (peptide alone); 170 kDa (GPIb containing GPIbβ); 130 kDa (glycocalicin).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.7–7.0 approx. (Heterogeneous and dependent on conditions used)
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	The other subunits are required for efficient expression and for biological activity.
Substrates	None

Inhibitors	Biological functions are inhibited by specific antibodies and by peptides from the region of the binding sites (aa 240–289) as well as peptides from the region of von Willebrand factor which binds to GPIb. The snake C-type lectin echicetin inhibits function.
Biological Functions	1. The N-terminal 45 kDa domain including the leucine-rich repeats and the double-loop as well as the negatively charged segment contains a binding site for von Willebrand factor which is critical for platelet binding to subendothelium in primary haemostasis. The binding site is not expressed in resting platelets and becomes available due to shear stress. Resting platelets can be aggregated by von Willebrand factor plus the antibiotic ristocetin or the snake peptide botrocetin, or by bovine von Willebrand factor or the snake peptide alboaggregin alone. Activation of platelets via von Willebrand factor binding induces tyrosine phosphorylation of cytoplasmatic components in a signal cascade involving Ca^{2+} influx. 2. The double-loop/negatively charged segment contains a binding site for thrombin that modulates the platelet response to that agonist. Although not essential for the platelet response (the 7-transmembrane, G-protein linked, moderate affinity, thrombin receptor is a prerequisite), if the GPIb site is blocked or removed about 10 times more thrombin is necessary to give the same response. This high affinity receptor must contain GPIb plus additional components, or be in a special conformation since there are only 50–100 copies per platelet. The mechanism is still unknown. The bulk of the GPIbα functions as a low affinity thrombin binding site.
Physiology/Pathology	1. GPIbα and/or the other subunits are absent or deficient in the rare, recessive, inherited bleeding disorder, Bernard-Soulier syndrome. Many different mutations and deletions have been described leading either to virtually complete non-expression of GPIbα or to expression at lower levels of an (often) non-functioning molecule due apparently to misfolding, particularly in the leucine-rich repeat domains. Since at least the three subunits, GPIbα, GPIbβ and GPIX need to be expressed for an efficient expression of the complex, defects in any of these will may affect overall expression. The role of GPV remains unknown. 2. Platelet-type von Willebrand's disease is caused by rare mutations in the double-loop region of GPIbα giving rise to molecules that bind von Willebrand factor under lower shear conditions than normal. This causes platelet aggregation *in vivo* and leads to thrombocytopenia and depletion of higher multimers of von Willebrand factor. This is a dominant, inherited bleeding disorder.
Degradation	As noted above GPIbα is readily cleaved by a variety of physiologically important proteases and degradation products (generally classed as glycocalicin) can be detected in plasma. Glycocalicin in plasma is elevated in diseases such as cirrhosis. Increased platelet destruction in immune thrombocytopenia may also lead to increased release of calpain and degradation of GPIbα. However, the significance is still poorly understood. Proteolytic degradation and loss of sialic acid residues during the average platelet lifetime of 10 days may contribute to the signals leading to platelet withdrawal from the circulation.
Genetics/Abnormalities	The complex GPIb-V-IX is the product of four genes on three separate chromosomes. The single copy gene for GPIbα is on chromosome 17p12-ter. The entire human gene has been cloned and sequenced and contains only a single short intron just upstream of the initiation site. Like other platelet and erythroid specific genes GATA-1 and ets transcription factor binding sites appear critical for regulating transcription. As well as the mutations leading to bleeding disorders GPIbα contains an important size polymorphism within the O-glycosylated region where a 13 aa segment

	containing 5 possible O-glycosylation sites is present in 1, 2 or 3 copies, called D, C, or B. A further category leading to a still higher mass was called A but the molecular origin has not yet been established. A dimorphism in position 145 with Thr (89%) and Met (11%) can lead to alloantibody formation.
Half-life	Platelet lifetime averages 10 days.
Concentration	approx. 25,000 molecules per platelet present on the surface-exposed plasma membrane surface. Equivalent amounts are thought to be present on the surface connected canalicular system and smaller amounts on the surface of α-granules.
Isolation Method	Several isolation procedures have been described. One involves platelet solubization in non-ionic detergent and affinity chromatography on a monoclonal antibody column. The other uses isolation of platelet membranes, solubilization in Triton X-114 and phase-separation followed by affinity chromatography of the aqueous phase on wheat germ agglutinin and thrombin columns. GPIb is separated from GPIX by SDS treatment and gel filtration and GPIbα from GPIbβ by reduction, alkylation, SDS treatment and gel filtration. Glycocalicin is prepared by treating platelets with calpain followed by affinity chromatography of the supernatant on wheat germ agglutinin and ion exchange chromatography.
Amino Acid Sequence	The aa sequence of the human protein has been determined partly directly and partly from the cDNA sequence. Direct sequencing showed that the 3 putative N-glycosylation sites in the N-terminal domain are in fact occupied.
Disulfides/SH-Groups	GPIbα contains 8 cysteines grouped in 3 intrachain disulfides (4–17; 209–248; 211–264), 1 interchain (to GPIbβ), probably 484) and 1 free, probably 485. Molecular model see Figure.
General References	Berndt, M. C. et al. *Eur. J. Biochem.* 1985, **151**:637–649. Clemetson, K. J. and Clemetson, J. M. *Current Opinion in Hematology* 1994, **1**:388–393. Clemetson, K. J. and Clemetson, J. M. *Semin. Thromb. Hemostas.* 1995, **21**:134–140. Kobe, B. and Deisenhofer, J. *Trends Biochem. Sci.* 1994, **19**:415–421. López, J. A. *Blood Coag. Fibrinol.* 1994, **5**:97–119. Roth, G. J. *Blood* 1991, **77**:5–19. Ruggeri, Z. M. *Prog. Hemost. Thromb.* 1991, **10**:35–68. Wicki, A. N. et al. **In** *Methods in Enzymology-Platelets: Receptors, Adhesion, Secretion* Hawiger, J. (ed.), Vol. **215**, pp. 276–288, Academic Press, NY 1992.
Ref. for DNA/AA Sequences	Handa, M. et al. *J. Biol. Chem.* 1986, **261**:12579–12585. López, J. A. et al. *Proc. Natl. Acad. Sci. USA* 1987, **84**:5615–5619. Wenger, R. et al. *Biochem. Biophys. Res. Commun.* 1988, **156**:389–395. Wenger, R. H. et al. *Gene*, 1989, **85**:519–524. Database accession numbers: PIR A27O75; SWISSPROT PO7359; EMBL/GenBank JO294O, gene M224O3.


Ibα
Leucine-rich domain
8.9 nm
vWf
Thrombin
Negatively charged segment
Highly O-glycosylated segment
34 nm
Size polymorphism
15.9 nm
Link to Ibß
cAMP-dependent kinase
Filamin
Actin Filaments

Platelet Membrane Glycoprotein Ibβ (GPIbβ)

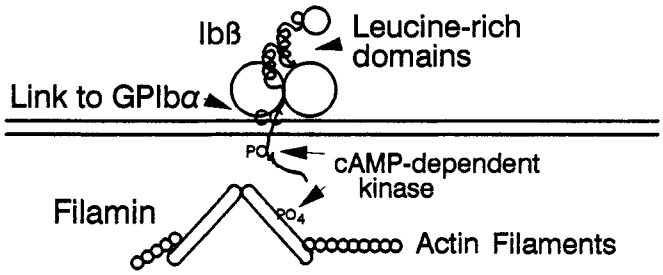
Kenneth J. Clemetson

Synonyms	CD42c
Abbreviations	GPIbβ
Classifications	Leucine-rich repeat family
Description	GPIbβ is part of a complex consisting of GPIb, GPV and GPIX (referred to as GPIb-V-IX) which is one of the major glycoprotein complexes on the platelet surface (about 25–50,000 copies per platelet). GPIbβ is linked to GPIbα by a disulfide bridge. All of these subunits are of the type 1 membrane protein class with a major N-terminal extracellular domains, a single transmembrane region and a C-terminal cytoplasmic domain. The cytoplasmic domains are linked to the membrane-associated platelet cytoskeleton via actin binding protein (filamin). GPIb-V-IX is restricted to platelets, megakaryocytes (where it is a marker of maturation) and to vascular and tonsillar endothelial cells that have been stimulated with TNF and/or IFNγ.
Structure	Human platelet GPIbβ consists of 181 aa residues. The N-terminal domain contains 2 disulfide loops (probably 1–14 and 5–7) and is similar to other leucine-rich repeat protein N-terminal flanking regions. The following approximately 50 aa residues make up 2 leucine-rich repeats, 24 aa long followed by two loops held together by disulfide bridges (probably cysteines 68–93 and 70–116, respectively). A short segment just above the membrane contains cysteine 122 which forms the link to GPIbα. The transmembrane domain contains 24 aa and is followed by the cytoplasmic C-terminal domain containing 33 aa among which is a cysteine (148) just under the membrane. Ser166 is phosphorylated by cAMP dependent protein kinase when adenylate kinase levels are raised in platelets. There is a single N-glycosylation site at Asp 41 and no evidence of any O-glycosylation. Human endothelial GPIbβ is 411 aa long with the additional 230 aa sequence all at the N-terminus. It is not known what the structural effects are but the extra sequence contains 6 cysteines implying additional loops. A baboon GPIbβ has also been cloned and sequenced and showed 94% similarity to the human platelet sequence.
Molecular Weight	27 kDa (reduced GPIbβ chain), 19 kDa (peptide alone); 170 kDa (GPIb containing GPIbα); 45 kDa in endothelial cells.
Sedimentation Coeff.	Unknown
Isoelectric Point	5.7–6.0 approx.
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	The other subunits are required for efficient expression and for biological activity.
Substrates	None
Inhibitors	None yet described.

Biological Functions	Probably critical in signal transduction from the binding sites on the a-subunit to the platelet interior. The ser166 phosphorylation site has been implicated in the control of actin polymerization.
Physiology/Pathology	No mutation in GPIbβ have been detected so far but it is likely that Bernard-Soulier syndrome could also be caused by mutations in this subunit.
Degradation	No evidence for any.
Genetics/Abnormalities	The complex GPIb-V-IX is the product of four genes on three separate chromosomes. The single copy gene for GPIbβ is on chromosome 22pter-22q11.2. The entire human gene has been cloned and sequenced and contains only a single short intron. Like other platelet and erythroid specific genes GATA-1 and ets transcription factor binding sites appear critical for regulating transcription. No polymorphism or mutants have been detected. The longer transcript obtained in endothelial cells and megakaryocyte-like cell-lines is the result of alternative splicing of the gene.
Half-life	Platelet lifetime averages 10 days.
Concentration	approx. 25,000 molecules per platelet present on the surface-exposed plasma membrane surface. Equivalent amounts are thought to be present on the surface connected canalicular system and smaller amounts on the surface of α-granules.
Isolation Method	Several isolation procedures have been described. One involves platelet solubization in non-ionic detergent and affinity chromatography on a monoclonal antibody column. The other uses isolation of platelet membranes, solubilization in Triton X-114 and phase-separation followed by affinity chromatography of the aqueous phase on wheat germ agglutinin and thrombin columns. GPIb is separated from GPIX by SDS treatment and gel filtration and GPIbβ from GPIbα by reduction, SDS treatment and gel filtration.
Amino Acid Sequence	The aa sequence of the human protein has been determined mostly from the cDNA sequence. The differences between the platelet and endothelial cell sequences are still unexplained.
Disulfides/SH-Groups	GPIbβ contains 10 cysteines grouped in 4 intrachain disulfides (1–14, 5–7, 68–93 and 70–116), 1 interchain (122) to GPIbβ and 1 free (148) that is partly palmitoylated. Molecular model see Figure.
General References	Clemetson, K. J. and Clemetson, J. M. *Semin. Thromb. Hemostas.* 1995, **21**:134–140. Fox, J. E. and Berndt, M. C. *J. Biol. Chem.* 1989, **264**:9520–9526. Hayzer, D. J. et al. *Gene* 1993, **127**:271–272. López, J. A. *Blood Coag. Fibrinol.* 1994, **5**:97–119. Rajagopalan, V. et al. *Blood* 1992, **80**:153–161. Roth, G. J. *Blood* 1991, **77**:5–19. Ruggeri, Z. M. *Prog. Hemost. Thromb.* 1991, **10**:35–68. Wardell, M. R. et al. *J. Biol. Chem.* 1989, **264**:15656–15661. Wicki, A. N. et al. **In** *Methods in Enzymology-Platelets: Receptors, Adhesion, Secretion* Hawiger, J. (ed.), Vol. **215**, pp. 276–288, Academic Press, NY 1992.

Ref. for DNA/AA Sequences

Kelly, M. D. et al. *J. Clin. Invest.* 1994, **93**:2417–2424.
López, J. A. et al. *Proc. Natl. Acad. Sci. USA* 1988, **85**:2135–2139.
Yagi, M. et al. *J. Biol. Chem.* 1994, **269**:17425–17427.
Database accession numbers: PIR A31929; SWISSPROT P13224; EMBL/GenBank J03259, gene U07983; endothelial L20860; baboon L05927.



Platelet Membrane Glycoprotein V (GPV)

Kenneth J. Clemetson

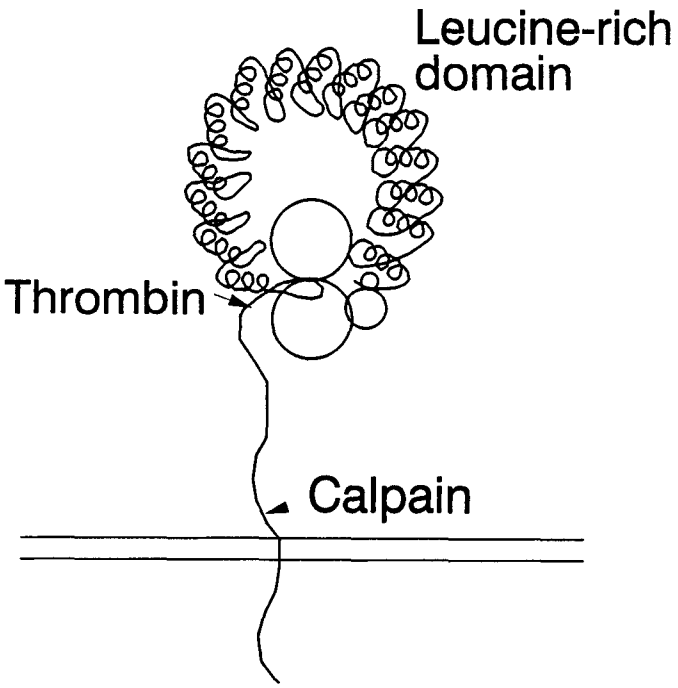
Synonyms	CD42d
Abbreviations	GPV
Classifications	Leucine-rich repeat family
Description	GPV is part of a complex consisting of GPIb, GPV and GPIX (referred to as GPIb-V-IX) which is one of the major glycoprotein complexes on the platelet surface (about 25–50,000 copies per platelet). GPV is present in about half this number of copies (12–25,000) and is therefore thought to link 2 GPIb-IX molecules. GPV is linked to GPIb-IX by weak non-covalent interactions that are disrupted by some non-ionic detergents and may therefore be hydrophobic. All of these subunits are of the type 1 membrane protein class with a major N-terminal extracellular domains, a single transmembrane region and a C-terminal cytoplasmic domain. The cytoplasmic domains are linked to the membrane-associated platelet cytoskeleton via actin binding protein (filamin). GPIb-V-IX is restricted to platelets, megakaryocytes (where it is a marker of maturation) and to vascular and tonsillar endothelial cells that have been stimulated with TNF and/or IFNγ.
Structure	Human GPV consists of 544 aa residues. The N-terminal domain contains 2 disulfide loops (probably 5–19 and 9–11) and is similar to other leucine-rich repeat protein N-terminal flanking regions. The following approximately 380 aa residues make up 16 leucine-rich repeats, 24 aa long, arranged in a horseshoe. followed by two loops held together by disulfide bridges probably formed by cysteines 409–435 and 411–436, respectively, followed by a short linear sequence containing a thrombin cleavage site (and also susceptible to other proteases). The following 40 aa till the membrane surface contain 1 N-glycosylation site and several O-glycosylation sites as well as a cleavage site for calpain. There is a 24 aa transmembrane segment and a 16 aa cytoplasmic domain. The leucine-rich repeats and N-terminal flanking domain contain 7 further putative N-glycosylation sites of which 4 (35, 165, 282 and 369) have been shown to be occupied.
Molecular Weight	83 kDa (non-reduced and reduced); 59 kDa (peptide alone); 81 kDa (calpain fragment); 69 kDa (thrombin fragment).
Sedimentation Coeff.	Unknown
Isoelectric Point	7.0–8.0 approx. (Heterogeneous and dependent on conditions used)
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	The other subunits are thought to be required for efficient expression and for biological activity.
Substrates	None
Inhibitors	There is some tentative evidence that specific antibodies affect platelet aggregation by von Willebrand factor.

Biological Functions	It is not yet clear whether GPV has a critical role in the function of the GPIb-V-IX complex. The fact that GPV is readily cleaved by both thrombin and calpain suggests that this may influence function but blockage of the cleavage site had no effect on platelet activation by thrombin.
Physiology/Pathology	GPV and/or the other subunits are absent or deficient in the rare, recessive, inherited bleeding disorder, Bernard-Soulier syndrome but no mutations have yet been described in the GPV gene. A case of Bernard-Soulier syndrome is known where alloantibodies against GPV developed that inhibited von Willebrand factor induced aggregation in normal platelets suggesting a critical function.
Degradation	GPV is readily cleaved by several physiologically important proteases including thrombin and calpain.
Genetics/Abnormalities	The complex GPIb-V-IX is the product of four genes on three separate chromosomes. The single copy gene for GPV is on chromosome 17p12-ter. The entire human gene has been cloned and sequenced and contains only a single short intron just upstream of the initiation site. Like other platelet and erythroid specific genes GATA-1 and ets transcription factor binding sites appear critical for regulating transcription but unlike the other members the gene for GPV contains a canonical TATA box.
Half-life	Platelet lifetime averages 10 days.
Concentration	approx. 12,500 molecules per platelet present on the surface-exposed plasma membrane surface. Equivalent amounts are thought to be present on the surface connected canalicular system and smaller amounts on the surface of α-granules.
Isolation Method	There are two general isolation methods depending whether the calpain fragment or the intact molecule are sought. The calpain fragment is only slightly smaller but lacks the transmembrane domain and was the first to be purified. Platelets are treated with calpain and then GPV is purified from the supernatant by wheat germ affinity chromatography followed by ion exchange chromatography. Intact GPV is obtained from platelet membranes by Triton X-114 phase separation. The Triton phase is then separated by wheat germ affinity chromatography followed by ion exchange chromatography.
Amino Acid Sequence	The aa sequence of the human protein has been determined partly directly and partly from the cDNA sequence. Direct sequencing showed that at least 4 of the 8 putative N-glycosylation sites in the N-terminal domain are occupied.
Disulfides/SH-Groups	GPV contains 8 cysteines in 2 groups of 2 intrachain disulfides (probably 5–19 and 9–11 and 409–435 and 411–436). Molecular model see Figure.
General References	Berndt, M. C. and Phillips, D. R. *J. Biol. Chem.* 1981, **256**:59–65. Bienz, D. et al. *Blood* 1986, **68**:720–725. Clemetson, K. J. and Clemetson, J. M. *Current Opinion in Hematology* 1994, **1**:388–393. Clemetson, K. J. and Clemetson, J. M. *Semin. Thromb. Hemostas.* 1995, **21**:134–140. Kobe, B. and Deisenhofer, J. *Trends Biochem. Sci.* 1994, **19**:415–421. López, J. A. *Blood Coag. Fibrinol.* 1994, **5**:97–119. Roth, G. J. *Blood* 1991, **77**:5–19. Ruggeri, Z. M. *Prog. Hemost. Thromb.* 1991, **10**:35–68.

Ref. for DNA/AA Sequences

Shimomura, T. et al., *Blood* 1990, **75**:2349–2356.
Roth, G. J. et al. *Biochem. Biophys. Res. Commun.* 1990, **170**:153–161.
Hickey, M. J. et al. *Proc. Natl. Acad. Sci. USA* 1993, **90**:8327–8331.
Lanza, F. et al. *J. Biol. Chem.* 1993, **268**:20801–20807.
Databank accession numbers: PIR S34329, A60164, C35483; SWISS-PROT P40197; EMBL/GenBank L11238, Z23091.

Glycoprotein V (CD42d)

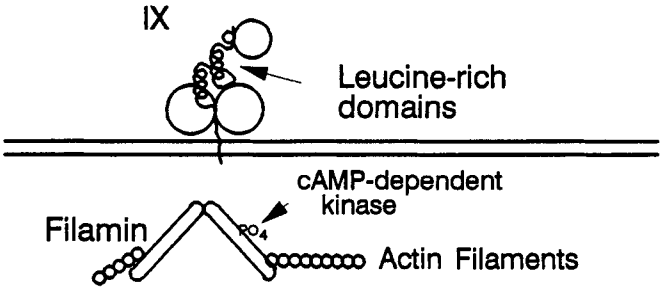


Platelet Membrane Glycoprotein IX (GPIX)

Kenneth J. Clemetson

Synonyms	CD42a
Abbreviations	GPIX
Classifications	Leucine-rich repeat family
Description	GPIX is part of a complex consisting of GPIb, GPV and GPIX (referred to as GPIb-V-IX) which is one of the major glycoprotein complexes on the platelet surface (about 25–50,000 copies per platelet) and carries most of the sialic acid and therefore the charge. All of these subunits are of the type 1 membrane protein class with a major N-terminal extracellular domains, a single transmembrane region and a C-terminal cytoplasmic domain. The cytoplasmic domains are linked to the membrane-associated platelet cytoskeleton via actin binding protein (filamin). GPIX and GPIb are linked by strong non-covalent interactions. GPIb-V-IX is restricted to platelets, megakaryocytes (where it is a marker of maturation) and to vascular and tonsillar endothelial cells that have been stimulated with TNF and/or IFNγ.
Structure	Human platelet GPIX consists of 160 aa residues. The N-terminal domain contains 2 disulfide loops (probably 4–22 and 8–10) and is similar to other leucine-rich repeat protein N-terminal flanking regions. The following approximately 50 aa residues make up 2 leucine-rich repeats, 24 aa long followed by two loops held together by disulfide bridges (probably cysteines 63–97 and 65–118, respectively). After a short segment of 11 aa comes the transmembrane domain containing 24 aa followed by the short cytoplasmic C-terminal domain containing 6 aa. There is a cysteine (153) just inside the membrane on the cytoplasmic side. There is a single N-glycosylation site at Asp 44 and no evidence of any O-glycosylation.
Molecular Weight	22 kDa; 17 kDa (peptide alone).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.2–5.8 approx.
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	The other subunits are required for efficient expression and for biological activity.
Substrates	None
Inhibitors	None yet described.
Biological Functions	Probably necessary for correct conformation of other subunits and in signal transduction from the binding sites on the α-subunit to the platelet interior.
Physiology/Pathology	Several mutations in GPIX have been detected leading to reduced expression of this and the other subunits and hence Bernard-Soulier syndrome.

Genetics/Abnormalities	The complex GPIb-V-IX is the product of four genes on three separate chromosomes. The single copy gene for GPIX like GPV is on chromosome 3. The entire human gene has been cloned and sequenced and contains only a single short intron. Like other platelet and erythroid specific genes GATA-1 and its transcription factor binding sites appear critical for regulating transcription. No polymorphisms have been detected.
Half-life	Unknown
Concentration	approx. 25,000 molecules per platelet present on the surface-exposed plasma membrane surface. Equivalent amounts are thought to be present on the surface connected canalicular system and smaller amounts on the surface of α-granules.
Isolation Method	Several isolation procedures have been described. One involves platelet solubization in non-ionic detergent and affinity chromatography on a monoclonal antibody column. The other uses isolation of platelet membranes, solubilization in Triton X-114 and phase-separation followed by affinity chromatography of the aqueous phase on wheat germ agglutinin and thrombin columns. GPIb is separated from GPIX by SDS treatment and gel filtration.
Amino Acid Sequence	The aa sequence of the human protein has been determined mostly from the cDNA sequence.
Disulfides/SH-Groups	GPIX contains 9 cysteines grouped in 4 intrachain disulfides (4–22, 8–10, 63–97 and 65–118), 1 free (153) that is partly palmitoylated.
General References	Clemetson, K. J. and Clemetson, J. M. *Semin. Thromb. Hemostas.* 1995, **21**:134–140. López, J. A. *Blood Coag. Fibrinol.* 1994, **5**:97–119. Roth, G. J. *Blood* 1991, **77**:5–19. Ruggeri, Z. M. *Prog. Hemost. Thromb.* 1991, **10**:35–68. Wicki, A. N. et al. **In** *Methods in Enzymology-Platelets: Receptors, Adhesion, Secretion* Hawiger, J. (ed.), Vol. **215**, pp. 276–288, Academic Press, NY 1992.
Ref. for DNA/AA Sequences	Hickey, M. J. et al. *Proc. Natl. Acad. Sci. USA* 1989, **86**:6773–6777. Hickey, M. J. et al. *FEBS Lett.* 1990, **274**:189–192. Hickey, M. J. and Roth, G. J. *J. Biol. Chem.* 1993, **268**:3438–3443. Database accession numbers: PIR A33731; SWISSPROT P14770; EMBL/GenBank x52997, gene M80478.

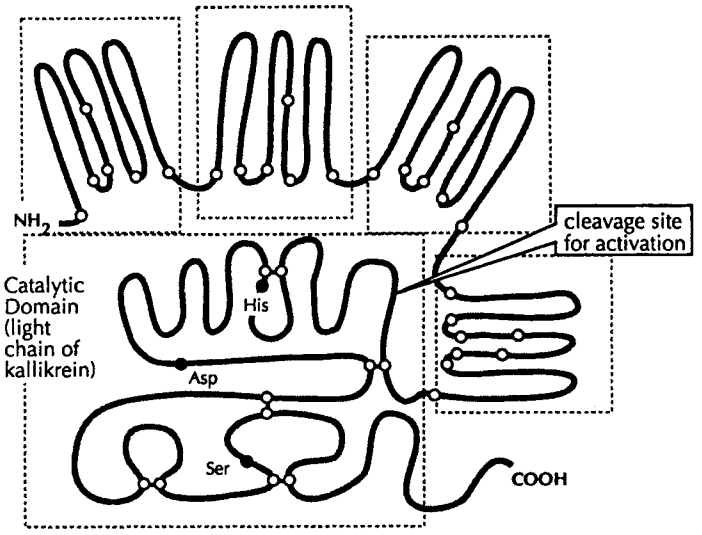


Plasma prekallikrein

Michael Silverberg

Synonyms	Fletcher Factor
Abbreviations	PK
Classifications	EC 3.4.21.8 (Kallikrein, as Tissue Kallikrein)
Description	Plasma Prekallikrein is a circulating plasma glycoprotein synthesized in the liver; it contains 15% carbohydrate. Prekallikrein is the zymogen form of the enzyme plasma kallikrein, which is distinct from tissue kallikrein. Kallikrein is formed from prekallikrein by cleavage at a single site yielding a two chain disulfide linked protein. The active site is on the light chain. Two asparagine residues in the heavy chain and three in the light chain are attachment sites for carbohydrate.
Structure	A repeating structure of 3 disulfide bonded loops repeated four times has been suggested for the heavy chain. The light chain is homologous with trypsin, chymotrypsin etc.
Molecular Weight	79,545 (aa sequence).
Sedimentation Coeff.	5.2 S
Isoelectric Point	8.7
Extinction Coeff.	11.7 (280 nm, 1%, 1 cm)
Enzyme Activity	Plasma kallikrein is a trypsin-like serine protease.
Coenzymes/Cofactors	Prekallikrein circulates in plasma as a non-covalent complex with high molecular weight kininogen. In the absence of the cofactor, surface mediated activation of prekallikrein does not occur. Kallikrein is active against synthetic substrates in the absence of high molecular weight kininogen.
Substrates	Factor XII (Hageman Factor), high molecular weight kininogen. Synthetic chromogenic substrates include Bz-Pro-Phe-Arg-p-Nitro-Anilide (chromozym PK) and H-D-Pro-Phe-Arg-p-Nitro-Anilide (S-2302).
Inhibitors	C1-Inhibitor, α_2-macroglobulin in plasma. Synthetic inhibitors include diisopropyl fluorophosphate, p-amidino phenylmethylsulfonyl fluoride, Phe-Phe-Arg-chloromethyl ketone.
Biological Functions	Plasma prekallikrein interacts with factor XII and high molecular weight kininogen in the presence of a negatively charged surface to initiate the intrinsic pathway of blood coagulation as well as the generation of the vasoactive peptide bradykinin. Activated factor XII converts prekallikrein to kallikrein and kallikrein reciprocally activates factor XII by proteolytic cleavage. Kallikrein also cleaves high molecular weight kininogen to liberate bradykinin. Kallikrein is also active in fibrinolysis by activating plasminogen directly and activating urokinase. Other activities have been noted.
Physiology/Pathology	Fletcher Trait is the syndrome resulting from the absence of prekallikrein. Subjects' plasma shows delayed clotting in the activated partial thromboplastin time; a correction of the clotting time is observed if the first (activation) step is prolonged. Fletcher trait is not associated with any

	observable defect in vivo. Several pathological states appear to be associated with consumption of prekallikrein.
Degradation	Kallikrein formed in plasma is rapidly inactivated by plasma inhibitors. The half-time of clearance of kallikrein C1-INH complexes may be about 1 hr; that for kallikrein α^2-macroglobulin complexes significantly longer.
Genetics/Abnormalities	Prekallikrein deficiency is known in about 30 families. Non-functional antigenic species have been found in several cases. One such protein was cleaved by activated factor XII at $< 5\%$ the normal rate.
Half-life	Unknown
Concentration	35–50 mg/L in plasma = 534 nM.
Isolation Method	Prepared from human plasma by ion exchange chromatography or by affinity chromatography on a monoclonal antibody column.
Amino Acid Sequence	The entire aa sequence of the protein has been determined by a combination of direct protein sequencing and aa sequence prediction from CDNA's isolated from a λgt-11 expression library. The structure of the gene has not been reported. The signal peptide of 19 residues is followed by the sequence of the mature plasma prekallikrein molecule which has 619 aa. Five asparagine residues have been identified as attachment sites for carbohydrate, three in the light chain and two in the heavy chain; all are Asparagine residues. The site of the cleavage that generates kallikrein from prekallikrein is at Arg^{371} -Ile^{372}, generating a light chain of 248 aa residues with the new amino-terminal sequence Ile-Val-Gly. The catalytic triad is comprised of His^{415}, Asp^{464} and Ser^{559}. The aa sequence of the heavy chain is unusual and has little homology with other serine proteases of the coagulation cascade with the exception of factor XI. The heavy chain sequence of prekallikrein comprises four tandem repeats each of which contains approximately 90–91 aa. The two carbohydrate sites found in the heavy chain are located in the second and fourth repeats. It is postulated that a gene segment coding for the ancestor of the repeat sequence duplicated and then the entire segment duplicated again to give the present structure.
Disulfides/SH-Groups	18 disulfides; 13 in the heavy chain of kallikrein, 4 in the light chain and 1 between the heavy and light chains. The presence of 6 conserved $^1/_2$ cystines per repeat suggests a repeating structure with three disulfide loops in each repeat.
General References	Silverberg, M., et al. The Contact System and it's Disorders. In: *Blood: Principles and Practice of Hematology*, Handin, R. I., et al. (eds.) Lippincott, Philadelphia (in Press). Kaplan, A. P., and Silverberg, M. The coagulation-kinin pathway of human plasma. *Blood* 1987, **70:** 1–15.
Ref. for DNA/AA Sequences	Chung, D. W., et al. *Biochemistry* 1986, **25:** 2410–2417.



Schematic diagram showing the structural domains inferred from sequence homologies. The catalytic triad residues are shown by filled circles, $\frac{1}{2}$ cystines by open circles and the cleavage site for activation by activated factor XII by an open arrow. Based on the illustration of Fujikawa et al.. Biochemistry, 1986, 25 :2417–2424.

Procarboxypeptidase U

Dirk F. Hendriks and Katinka A. Schatteman

Synonyms	Procarboxypeptidase R; Arginine procarboxypeptidase; Plasma procarboxypeptidase B; Thrombin-activable fibrinolysis inhibitor; Inducible carboxypeptidase activity
Abbreviations	Proenzyme: proCPU; pro-pCPB; TAFI Active enzyme: CPU; CPR; pCPB
Classifications	EC 3.4.17.x Carboxypeptidases; peptidase family M14
Description	A plasma protein, synthesized in the liver, comprising a single polypeptide chain of 401 aa. Proteolytic cleavage of the 92 aa activation peptide results in the formation of the 309-aa catalytic domain designated carboxypeptidase U.
Structure	The active enzyme carboxypeptidase U exhibits high homology with the pancreatic carboxypeptidases A and B and thus is presumed to have a similar fold as the well known carboxypeptidase A structure. It is a zinc metalloenzyme having a central mixed parallel/antiparallel eight-strand β-sheet over which eight α-helices pack on both sides to form a globular molecule.
Molecular Weight	Procarboxypeptidase U: 60 kDa (SDS-PAGE); 48,442 (aa sequence) Carboxypeptidase U: 36 kDa (SDS-PAGE) Activation peptide is heavily glycosylated; active enzyme is not glycosylated
Sedimentation Coeff.	Unknown
Isoelectric Point	5.0 for CPU (IEF)
Extinction Coeff.	Unknown
Enzyme Activity	Removes C-terminal basic aa arginine and lysine from peptides and proteins
Coenzymes/Cofactors	Zn-containing metalloprotease (one atom Zn per polypeptide chain)
Substrates	Peptides and proteins with unblocked C-terminal arginine or lysine. Natural substrates: partially degraded fibrin, bradykinin. Synthetic substrates: hippuryl-L-arginine, hippuryl-L-lysine, Fa-Ala-Arg, Fa-Ala-Lys
Inhibitors	GEMSA (guanidonoethylmercaptosuccinic acid), MERGETPA (2-mercaptomethyl-3guanidinoethyl-propanoic acid). Potato carboxypeptidase inhibitor Metal ion chelators EDTA and phenanthrolin
Biological Functions	It was postulated that CPU, generated during coagulation, could play a role in the fibrinolytic system. This hypothesis was substantiated by the fact that proCPU exhibits a high affinity for plasminogen. Moreover, the circulating zymogen proCPU is activated by either thrombin (in conjunction with thrombomodulin, an integral endothelial cell

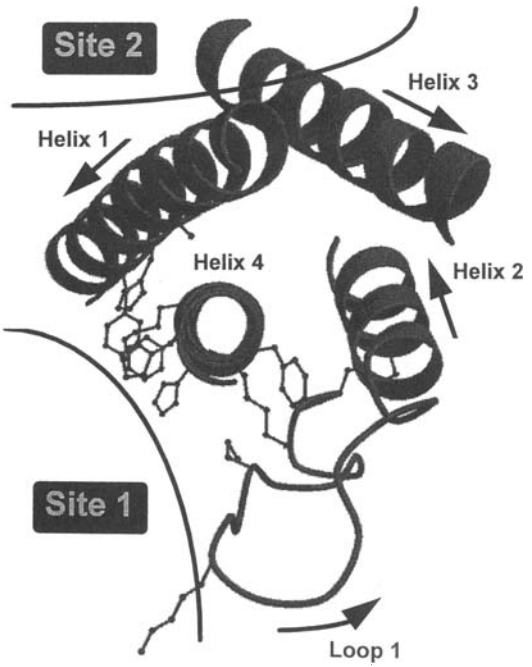
	membrane protein) or by plasmin. Plasmin degradation of fibrin exposes C-terminal lysine residues which are essential for the high affinity binding of plasminogen to fibrin. Thus CPU could control the rate of fibrinolysis by cleaving off part of these C-terminal lysine residues. Recent data indicate that CPU indeed could play a role in plasminogen activation since it is able to delay tPA-induced clot lysis in vitro.
Physiology/Pathology	Unknown
Degradation	Proteolytic cleavage of procarboxypeptidase U by thrombin or plasmin after arginine-92 produces catalytically active carboxypeptidase U. Further cleavage at arginine-330 inactivates CPU.
Genetics/Abnormalities	The gene for procarboxypeptidase U is located on chromosome 13 region q14.11.
Half-Life	Unknown
Concentration	2 - 5 mg/L in plasma
Isolation Method	Isolation of procarboxypeptidase U from human plasma collected into sodium citrate anticoagulant is obtained by precipitation with $BaCl_2$ and ammonium sulphate, anion exchange chromatography and affinity chromatography on plasminogen-Sepharose
Amino Acid Sequence	The primary structure has been deduced from a cDNA sequence. The primary translation product consists of 423 aa containing a 22 aa signal peptide, a 92 aa activation peptide, and a 309 aa catalytic domain. ProCPU exhibits high sequence similarity to both pancreatic carboxypeptidases A and B. Amino acids believed to be important in the catalytic mechanism of these carboxypeptidases are also conserved in proCPU.
Disulfides/SH-Groups	Unknown
General References	Hendriks, D. et al. *Biochim. Biophys. Acta* 1990, **1034**:86-92. Eaton, D. et al. *J. Biol. Chem.* 1991, **266**:21833-21838. Wang, W. et al. *J. Biol. Chem.* 1994, **269**:15937-15944. Bajzar, L. et al. *J. Biol. Chem.* 1995, **270**:14477-14484. Redlitz, A. et al. *J. Clin. Invest.* 1995, **96**:2534-2538.
Ref. for DNA/AA Sequences	Eaton, D. et al. *J. Biol. Chem.* 1991, **266**:21833-21838. EMBL accession number M75106.

Prolactin

Vincent Goffin and Joseph A. Martial

Synonyms	Luteotropic hormone (LTH); Lactogenic hormone
Abbreviations	PRL
Classification	Polypeptide hormone
Descriptions	PRL is synthesized mainly in anterior pituitary; peripheric organs (mammary gland, placenta) and lymphoid cells also produce low amounts of PRL. Belongs to a hormonal family including growth hormone and placental lactogen.
Structure	All α-helix protein (CD analysis). No experimental 3D structure available; hPRL is predicted to fold in a four-helix bundle (up-up-down-down connectivity).
Molecular Weight	Theoretical: 22,898. Electrophoretic mobility: 23,000. Post-translational modifications (glycosylation, phosphorylation,...) can increase MW. Dimers or multimers can be observed. The monomeric, oxydized, non-glycosylated hPRL is the major form.
Sedimentation Coefficient	Unknown
Isoelectric Point	6.2 (recombinant hPRL - I.E.F.).
Extinction Coefficient	20,700
Enzymatic Activity	None
Coenzyme/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Involved in lactation (best known effect, from which prolactin name was given to the hormone) and reproduction. PRL also plays a role in osmoregulation, growth, morphogenesis and immunomodulation. PRL bioactivities are mediated by specific membrane receptors.
Physiology/Pathology	The main target tissue is mammary gland, on which PRL has mammotrophic (proliferative) and lactogenic (induction of milk protein synthesis) actions. No known pathology directly linked to PRL defect or overproduction. Hyperprolactinemia (excess of PRL secretion) is treated by bromocriptin (analog of dopamine, natural down-regulator of pituitary PRL secretion).
Degradation	Several fragments of varying size resulting from proteolysis of hPRL. The N-terminal fragment, so-called 16K hPRL, is anti-angiogenic (intact hPRL is angiogenic).
Genetics/Abnormalities	hPRL gene is ~10 kb and contains 5 exons.

	Gene localization: Chromosome 6. mRNA: ~1 kb. Tissue-specific usage of two different PRL gene promoters (pituitary *versus* non-pituitary). No reported hPRL gene abnormalities.
Half-life	Less than 1 hour
Concentration	From less than 20 ng/ml to more than 200 ng/ml (late pregnancy) in blood.
Isolation Method	From anterior pituitary. Today, from recombinant organisms
Amino Acid Sequence	199 aa (excluding the signal peptide of 28 aa) Glycosylation site: Asn 31 Two receptor binding sites (sites 1 and 2), leading to receptor homodimerization Binding site 1 involves a dozen of residues In binding site 2, helices 1 and 3 create a cleft around Gly-129 for receptor docking. Any mutation of Gly-129 is detrimental to receptor dimerization (i.e. activation)
Disulfide/SH-Groups	3 intramolecular disulfides: Cys-4-Cys-11, Cys-58-Cys-174, Cys-191-Cys-199
General References	Cooke, N.E. et al. *J. Biol Chem.* 1981, **256**:4007-4016 (genetics). Truong, A.T. et al. *EMBO J.* 1984, **3**,429-437 (genetics). Sinha, Y.N. *Endocr. Rev.* 1995, **16**:354-369 (isoforms). Goffin, V. et al. *Endocr. Rev.* 1996, **17**:385-410 (structure/binding). Bole-Feysot et al. This review should appear in the last 1997 issue of *Endocrine Review.*
Ref. for DNA/AA Sequences	Genebank: J00299



Molecular model: 3D model of hPRL based on the homolog porcine growth hormone structure (Goffin et al, *Prot. Eng.* 1995, **8**:1215-1231). Receptor binding sites 1 and 2 are indicated.

Prolactin Receptor

Vincent Goffin and Paul A. Kelly

Synonyms	Lactogen(ic) receptor
Abbreviations	PRLR
Classifications	Single transmembrane receptor: N-terminus extracellular, C-terminus cytoplasmic.
Description	Found in a wide spectrum of cell types (mammary gland, liver, brain, bone, etc.). Binds to all lactogenic hormones (prolactin, growth hormone, placental lactogen). Activated by ligand-induced homodimerization. Belongs to class-1 cytokine receptor superfamily. Extracellular domain can be found as soluble receptor (resulting from membrane receptor proteolysis and/or alternative splicing).
Structure	Extracellular: 2 β-sheet sandwiches of 7 β-strands (X-ray). Intracellular: unknown
Molecular Weight	Theoretical: 66,905. Electrophoretic mobility: 90,000-95,000, due to post-translational modifications (glycosylation, phosphorylation).
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzymatic Activity	None (PRLR is a non-tyrosine kinase receptor)
Coenzyme/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	PRLR mediates biological activities of lactogenic hormones, which concern lactation and reproduction (best known effects of these hormones), osmoregulation, growth, morphogenesis and immunomodulation.
Physiology/Pathology	The main target tissue is mammary gland, on which PRL has mammotrophic (proliferative) and lactogenic (induction of milk protein synthesis) actions. No known pathology directly linked to PRLR defect or overproduction.
Degradation	On SDS-PAGE, degradation products presumably resulting from C-terminal tail proteolysis are observed.
Genetics/Abnormalities	hPRLR gene is > 100 kb and contains at least 10 exons. 3 mRNA identified: 2.8, 3.5 and 7.3 kb. Gene localization: Chromosome 5, p13→14. No genetic aberration reported.

Half-life	Ligand-bound PRLR is internalized within 30-60 minutes and degraded in lysosomes
Concentration	10-2000 fmol/mg membrane protein
Isolation Method	Immunoprecipitation from tissues rich in PRLR (e.g. mammary gland). Today, from recombinant organisms
Amino Acid Sequence	598 aa (excluding the signal peptide of 24 aa) Extracellular domain: 210 aa 3 glycosylation sites: Asn-35, Asn-80, Asn-209 2 pairs of disulfide cysteines (required for ligand binding, folding and expression at plasma membrane). WS motif (Trp-191-Ser-192-any aa-Trp-194-Ser-195), conserved in all class-1 cytokine receptors. Trp-72 and Trp-139 involved in ligand binding. Transmembrane domain: 24 aa Cytoplasmic domain: 364 aa Box 1: proline-rich region (aa 245-250) conserved in all class-1 cytokine receptors Some tyrosine residues that bind SH2-containing transducers when phosphorylated.
Disulfides/SH-Groups	2 intramolecular disulfides: Cys-12-Cys-22, Cys-51-Cys-62
General References	Boutin, J.M. et al. *Mol. Endocrinol.* 1989, **3**:1455-1461 (genetics). Arden, K.C. et al. *Cytogenet. Cell Genet.* 1990, **53**:161-165 (genetics). Kelly, P.A. et al. *Endocr. Rev.* 1991, **12**:235-251 (genetics). Bole-Feysot et al. This review should appear in the last 1997 issue of *Endocrine Review*. Somers, W. et al. *Nature* 1994, **372**:478-481 (structure).
Ref. for DNA/AA Sequences	Genebank: M31661.

Molecular model: Schematic representation of hPRLR interacting with one molecule of human growth hormone (Somers et al, 1994). The seven β-strands of each β-sheet (A to F and A' to F') are represented. H1 to H4 indicate the four helices of the ligand.

Prolyl Oligopeptidase

Filip J. Goossens and Greet C. Vanhoof

Synonyms	Post-proline cleaving enzyme; Proline specific endopeptidase; Prolyl endopeptidase
Abbreviations	PO; PPCE; PSE; PE; PEP; POPase
Classifications	EC 3.4.21.26
Description	Prolyl oligopeptidase is a proline specific endopeptidase with serine type mechanism. The enzyme is distributed throughout the whole body. Prolyl oligopeptidase is primarily a cytosolic enzyme, displaying a much higher tissue activity compared to the body fluids. The precise cellular localization is yet unknown. Prolyl oligopeptidase is not glycosylated and contains 16 cysteine residues. There are indications that apart from the cytosolic form there exists also a membrane-bound form.
Structure	Prolyl oligopeptidase is a monomeric protein consisting of 710 aa. The structure of the enzyme is not resolved. It is a member of a new class of serine peptidases, the oligopeptidases. This group of enzymes has a Ser-Asp-His linear arrangement of the active site. Prediction of the secondary structure did indicate that the protease domain of the enzyme could adopt the α/β hydrolase fold, and the overall fold was predicted as a mixed class protein. Limited proteolysis by trypsin suggests a multiple domain structure.
Molecular Weight	70-75 kDa: (SDS-PAGE, reduced). 65-79 kDa: (gel permeation chromatography) 80,745 and 80,750 Da (calculated from its aa sequence from two different clones)
Sedimentation Coeff.	Unknown
Isoelectric Point	4.7 - 4.9
Extinction Coeff.	Unknown
Enzyme Activity	The enzyme specifically cleaves the peptide bond at the carboxyl side of proline residues. The enzyme only cleaves peptide bonds bearing natural amino acids. The bond preceding the scissile bond must have the *trans* configuration. The prolyl oligopeptidase activity is restricted towards small peptides.
Coenzymes/Cofactors	Unknown
Substrates	The natural substrate for this enzyme is still unknown, but *in vitro* studies indicate that bradykinin is one of the best substrates. The synthetical substrate Z-Gly-Pro-AMC has a K_m of 120 μM and a neutral pH optimum.
Inhibitors	Specific inhibitors for this enzyme have been designed based on the Z-prolyl-proline skeleton. Aldehydes of this peptide interact with the active site serine yielding irreversible inhibitors with a K_i in the nanomolar range. Derivatives bearing no reactive C-terminal moiety have been designed and these interact as competitive inhibitors. It has been shown in animal models that inhibitors for prolyl oligopeptidase do prevent the amnesic

	effect induced by scopolamine. Inhibitors for prolyl oligopeptidase were also able to restore the TRH modulated acetylcholine liberation.
Biological Functions	Unknown biological function. With respect to its high substrate selectivity, the enzyme was thought to be involved in the processing of neuropeptides and hormones. It is implicated in the process of memory retrieval (cfr. Inhibitors). The enzyme is also investigated for its potential role in the renin-angiotensin system as an angiotensin I and II processing enzyme.
Physiology/ Pathology	No deficiency of prolyl oligopeptidase is reported. There are indications for a possible involvement of the enzyme in psychological disorders. The serum activity of prolyl oligopeptidase is correlated with the different stages of depression. Treatment of mania and depression with current drug therapy revealed that the enzyme activity returns to normal values. The activity of prolyl oligopeptidase tends to be enhanced in tumour tissues with respect to the controls.
Degradation	Slow degradation upon storage even at -80°C. Purity is crucial. Trypsin treatment resulted in a specific cleavage of the enzyme at the Lys-196-Ser-197 peptide bond, without affecting the enzyme activity. No information is known concerning its *in vivo* pathways of degradation
Genetics/Abnormalities	The gene for the human enzyme is located on chromosome 6 region q 22.1. No known mutations.
Half-life	Unknown
Concentration	Unknown
Isolation Method	A combination of precipitation and chromatographic procedures. Following a freeze thaw cycle, a precipitation with both aceton and afterwards ammonium sulphate is performed. The precipitate is recovered and the separation is completed by two ion exchange columns intersected by hydrophobic interaction chromatography on phenyl Sepharose.
Amino Acid Sequence	The sequence of the human enzyme has been reported indepedently by two different groups. The sequence was checked by aa sequencing by only one group. The enzyme consisits of 710 aa. Seven mutations are identified between both human cDNA clones. The active site residues are Ser-554, Asp-641 and His-680. The sequence displays high homology with CD26 (EC 3.4.14.5) and acylaminoacyl peptidase (EC 3.4.19.1). Although reports indicate the presence of a different form of prolyl oligopeptidase only two, highly homologous sequences (2 clones), are known to date.
Disulfides/SH-Groups	16 cysteines are present in the sequence. The content of disulfides is unknown. Cys-25 an Cys-57 are thought to be located in close proximity of the active site.
General References	Goossens, F. et al. *Eur. J. Biochem.* 1995, **233**:432-441. Goossens, F. et al. *Eur. J. Clin. Chem. Clin. Biochem.* 1996, **34**:17-22. Kalwant, S. and Porter, G. *Biochem. J* . 1991, **276**:237-244. Maes, M. et al. *Psychiatry Res.* 1995, **58**:217-225. Welches, W. et al. *Life Sci.* 1993, **52**:1461-1480. Yoshimoto, T. et al. *J. Pharmacobio-Dyn.* 1987 **10**:730-735.
Ref. for DNA/AA Sequences	Vanhoof, G. et al. *Gene* 1994, **149**:363-366. (X74496). Shirasawa, Y. et al. *J. Biochem.* (Tokyo) 1994 **115**:724-729. (D21102).

Goossens, F. et al. *Cytogenet. Cell Genet.* 1996, **74**:99-101. (PREP, GDB362664).
Genbank Acc. No. X74496 (SP P48147), D21102 (PIR JC2257), Gene GDB362664

Properdin

Kenneth B.M. Reid

Synonyms	None
Abbreviations	P
Classifications	Electrophoretic mobility: γ_2-region
Description	A circulating plasma glycoprotein which acts as a positive regulator of the alternative pathway of the complement system. Spleen, peritoneal macrophage and lung (but not liver) have been shown to be sites of synthesis of properdin mRNA in the mouse, by Northern blot analysis. Cultured human monocytes and the human cell lines U937, HL60 and HepG2 have been shown to synthesise properdin as judged by radioimmunobinding assays. Properdin is present in plasma as a polydisperse mixture of cyclic polymers, mainly dimers, trimers and tetramers, in the ratio 20:54:26, formed by the (presumably head to tail) interactions of a single, 53 kDa, asymmetric monomer. Conversion of these polymeric forms into large amorphous aggregates yields material which has been designated 'activated' properdin. This process is not a physiological event associated with complement activation in serum, it is simply an artifact of the isolation and handling of purifird properdin oligomers.
Structure	By electron microscopy studies, the monomeric units, within the cyclic polymers, appear as flexible rods (approx. 26 nm in length) possessing a reproducible sharp bend near the middle and thickening at the ends. It is not known how the monomers interact, but the fact that the monomeric form has not been detected under physiological conditions, and the observation that there is apparently no redistribution of the individual oligomeric forms, indicates that the postulated N- to C- association between monomers is strong. Structure prediction methods and Fourier Transform Infrared Spectroscopy indicate a high, 56 - 67% β-turn structure and 19 - 38% β-sheet structure which is consistent with Gly, Pro, Cys and Ser being the four most abundant residues in properdin.
Molecular Weight	The molecular weight of the glycosylated single chain 'monomer' of human properdin is 53,300 (this value includes 9.8% (w/w) carbohydrate). It should be noted that the monomeric form of properdin is only seen in dissociating conditions and that normally dimers, trimers and tetramers of this chain are seen in physiological conditions.
Sedimentation Coeff.	5.0 - 5.3 S (polydisperse nature)
Isoelectric Point	8.2 - 8.7
Extinction Coeff.	23.4 (280nm, 1%, 1cm). This extremely high value is due to a tryptophan content of 4.5% in human properdin.
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None

Biological Functions	Properdin binds to and stabilises the inherently labile C3 and C5 convertase enzyme complexes (C3bBb and $C3b_nBb$) which split the C3 and C5 components, respectively during activation of the alternative pathway of complement. Since the alternative pathway provides an antibody independent route for the destruction of bacteria, neutralisation of viruses and lysis of certain mammalian cells, properdin plays an important role in complement mediated defence mechanisms.
Physiology/Pathology	Approx. 50% of the individuals with a complete deficiency of properdin suffer from severe, fulminant, pyrogenic bacterial infections with meningococci being frequently involved and with a fatality rate approaching 75%. These observations emphasise the important role properdin plays in mediating efficient activation of the complement system with respect to killing and clearance of bacteria.
Degradation	Unknown
Genetics/Abnormalities	Several families with one, or more members, completely deficient in properdin function, have been described. The deficiency is transmitted as an X-linked recessive trait which is consistent with the mapping of the gene to Xp11.23-Xp11.3. In most cases the molecule is antigenically as well as functionally absent from the plasma.
Half-life	Unknown
Concentration	Serum: approx. 6.5 mg/L (range 4.7-13.7 mg/L). Early estimates of the serum concentration were at least two fold higher due to use of an extintion coefficient which was approx. half the true value of 23.4 (280nm, 1%, 1cm).
Isolation Method	Precipitation from serum by 5% polyethylene glycol (Mr4000), chromatography on QAE-Sephadex, removal of IgG contaminants by antihuman IgG Sepharose. Better yields can be obtained by affinity chromatography of the PEG precipitate on anti-properdin Sepharose, elution of the protein at low pH followed by QAE-Sephadex.
Amino Acid Sequence	The derived sequence of the mature protein (442 aa) shows that the single chain "monomer" contains a 49 aa long N-terminal region of no known homology, followed by six non-identical repeats of approximately 60 aa which are each considered to form independently folded modules. The intron-exon structure of the gene reflects this organisation very closely, since the N-terminal region and the first five repeats are encoded by individual exons whereas the DNA encoding the sixth repeat is interrupted by a single exon. These repeats, of the type DxGWxxWSxWxxCSxTCGx GVxxxRxRxCNxxxxxxxxCxGxxxExxxCxxQxxCxx, were first identified in the cell adhesion molecule thrombospondin and are named thrombospondin type I repeats (TSRs). Similar repeats are found in the terminal components of complement (C6, C7, C8a, C8b and C9), the neural protein F-spondin, UNC-S of *Caenorhabditis elegans*, proteins related to connective tissue growth factor (including CEF-10, CYR-61, FISP-12 and NOV) and in various parasite proteins such as the circumsporozoite and thrombospondin-related anonymous proteins of *Plasmodium* species. The derived sequences for mouse and pig properdins show 76% identity to the human properdin sequence.
Disulfides/SH-Groups	There are 44 cysteine residues per 442 residue-long chain of human properdin and since no free sulfydryl groups have been detected it is

presumed there are 22 disulfides per molecule. No information is available regarding the arrangement of the disulfide bridges.

General References

Lepow, I.H. Louis Pillemer, Properdin and scientific controversy. *J. Immunol.* 1980, **125**:471-478.

Smith, C.A. et al. Molecular architecture of human properdin a positive regulator of the alternative pathway of complement. *J. Biol. Chem.* 1984, **259**:4582-4588.

Farries, T.C. et al. Resolution and analysis of "native" and "activated" properdin. *Biochem. J.* 1987, **243**:507-517.

Figueroa, J.E. et al. Infectious diseases associated with complement deficiencies. *Clin. Microb. Rev.* 1991, **4**:359-395.

Pangburn, M.K. Analysis of the natural polymeric forms of human properdin and their functions in complement activation. *J. Immunol.* 1989, **142**:202-207.

Higgins, J.M.G. et al. Characterization of mutant forms of recombinant human properdin lacking single thrombospondin type I repeats - Identification of modules important for function. *J. Immunol.* 1995, **155**:5777-5785.

Ref. for DNA/AA Sequences

Reid, K.B.M., Gagnon, J. Amino acid sequence studies of human properdin-N-terminal sequence analysis and alignment of the fragments produced by limited proteolysis with trypsin and the peptides produced by cyanogen bromide treatment. *Mol. Immunol.* 1981, **18**:949-959.

Goundis, D. and Reid, K.B.M. Properdin, the terminal complement components, thrombospondin and the circumsporozoite protein of malaria parasites contain similar sequence motifs. *Nature* 1988, **335**:82-85.

Nolan, K.F. et al. Molecular cloning of the cDNA coding for properdin, a positive regulator of the alternative pathway of human complement. *Eur. J. Immunol.* 1991, **21**:771-776.

Nolan, K.F. et al. Characterization of the human properdin gene. *Biochem. J.* 1992, **287**:291-297.

Maves, K.K. et al. Cloning and characterization of the cDNA encoding guinea-pig properdin: a comparison of properdin from three species. *Immunology* 1995, **86**:475-479.

Prostasin

Jack X. Yu, Lee Chao and Julie Chao

Synonyms	None
Abbreviations	None
Classification	None
Description	Prostasin is a newly identified human kallikrein-like serine proteinase with unique structure and properties. Prostasin is synthesized as a preproprostasin, which is converted into proprostasin by removal of the signal peptide. Active prostasin with two chains arises when the peptide bond between Arg-12 and Ile-13 is cleaved. Prostasin is a membrane-anchored proteinase through its C-terminal hydrophobic portion and it is released after a cleavage at Arg-290-Pro-291. Prostasin is present in the prostate, colon, lung, kidney, pancreas, salivary gland, liver, bronchus; but not in the brain, muscle, testis, ventricle, atrium and aorta. The prostate has the highest level of prostasin, 20-fold higher than any other gland.
Structure	The active prostasin consists of a 12-aa light chain and a 299-aa heavy chain. They are associated through a disulfide bond. Prostasin has a transmembrane domain at its C-terminus and a N-glycosylation site at Asn-127. The catalytic triad is His-53, Asp-102, and Ser-206. In addition, an Asp residue is present at position 200, six residues before the active Ser-206. The Asp-200 is believed to be located at the bottom of the substrate-binding pocket and determines prostasin's specific arginine or lysine amidolytic activity in peptidyl substrates. At the present time, the tertiary structure of prostasin has not been determined.
Molecular Weight	40,000 (SDS-PAGE); heavy chain: 32,186, and light chain:1,169 (based on the aa sequence translated from the cDNA)
Sedimentation Coeff.	Unknown
Isoelectric Point	4.5 to 4.8, IEF displays five major bands
Extinction Coeff.	16.3 (280nm, 1%, 1cm)
Enzyme Activity	Prostasin displays arginine/lysine amidolytic activities on synthetic peptidyl substrates, such as D-Pro-Phe-Arg-7-amino-4-methyl-coumarin (AMC), D-Phe-Phe-Arg-AMC, and Boc-Val-Leu-Lys-AMC. It has the highest activities at pH 9.0.
Coenzymes/Cofactors	None
Substrates	Prostasin can cleave the following synthetic peptidyl substrates: D-Pro-Phe-Arg-AMC, D-Phe-Phe-Arg-AMC, D-Val-Leu-Arg-AMC, Z-Gly-Pro-Arg-7-amino-4-trifluoromethyl coumarin (AFC), Suc-Ala-Phe-Lys-AMC, and Boc-Val-Leu-Lys-AMC. The biological substrate has not been determined.
Inhibitors	Prostasin is strongly inhibited by aprotinin (IC_{50}: 1.8 x 10^{-9}M), and moderately by antipain (IC_{50}: 6.4 x 10^{-6}M), leupeptin (IC_{50}: 1.0 x 10^{-4}M), and benzamidine (IC_{50}: 8.6 x 10^{-4}M). It is also sensitive to diisopropyl-fluorophosphate while soybean trypsin inhibitor has no effect on its activity.

Biological Functions

Prostasin is a kallikrein-like, membrane-anchored proteinase, indicating it could play important roles in biological processes occurring on cell surfaces, such as processing of protein hormones and growth factors, and activation of precursors of proteinases. Channel-activating protease 1 from Xenopus, which shares the highest identity with prostasin, has been shown to enhance the activity of the amiloride-sensitive epithelial sodium channel. This finding suggests that prostasin may be involved in the regulation of sodium balance and blood pressure. The precise functions of prostasin still remain to be established.

Physiology/Pathology

Prostasin has a unique structure and is present at extremely high levels in the prostate and seminal fluid, suggesting it may play a role in the development of the prostate gland and in the process of reproduction. Whether abnormal expression of prostasin is associated with prostatic diseases and/or infertility has yet to be investigated.

Degradation

The first step of degradation is probably the release from the cellular membrane by a cleavage at Arg-290-Pro-291.

Genetics/Abnormalities

The gene spans over 4.4 kb, and consists of six exons and five introns. It is a single-copy gene localized on chromosome 16p11.2. The gene has a TATA-less promoter and its transcription starts at a transcription initiator element. There are no known abnormalities in the prostasin gene.

Half-life

Unknown

Concentration

Seminal fluid, 8.61 ± 0.42 μg/ml; urine, 0.201 ± 0.029 μg/ml; prostate gland, 143.7 ± 15.9 ng/mg protein; kidney, 2.64 ± 0.50 ng/mg protein; plasma: not detectable.

Isolation Method

Human seminal fluid was centrifuged at 12,000 x g for 30 min at 4°C and the supernatant was applied to a DEAE-Sepharose CL-6B column. The column was eluted with a linear gradient of 0-0.4 M NaCl in 25 mM TrisHCl (pH 7.6), and the fractions, eluted with 0.15-0.25 M NaCl, were pooled and concentrated by PEG compound. It was then loaded on an aprotinin-agarose column and prostasin was eluted with 0.1 N glycine-HCl, pH 3.0.

Amino Acid Sequence

Prostasin shares 34-42% identity in aa sequence with the human serine proteinases listed below. The similarity is especially high in the sequences around the active sites.

PROSTASIN	WVLSAAHCF	DIALL	DACQGDSGGPL
P. KALLI.	WVLTAAHCF	DIALI	DACKGDSGGPL
COAG. XI	WILTAAHCF	DIALL	DACKGDSGGPL
β-TRYPSIN	WVLTAAHCV	DIALL	DSCQGDSGGPL
HEPSIN	WVLTAAHCF	DIALV	DACQGDSGGPF
PLASMIN.	WVLTAAHCL	DIALL	DSCQGDSGGPL
ACROSIN	WVLTAAHCF	DIALV	DTCQGDSGGPL
CAP1	YILTAAHCF	DVALA	DACQGDSGGPL
	*	*	*

P. KALLI: plasma kallikrein; COAG. XI: coagulation factor XI; PLASMIN: plasminogen; CAP1: channel-activating protease 1.
* indicates the active sites.

Disulfides/SH-Groups

There are 12 cysteine residues in active prostasin. Nine of them are highly conserved when compared with other serine proteinases whose disulfide bond arrangement has been determined. Thus the intrachain disulfide bonds in prostasin are predicted at cysteine pairs 38/54, 136/212, 169/191, and 202/230. An interchain disulfide is expected at cysteine pair 5/122, which hold the two chains together. In addition, prostasin has two

unique cysteine residues at positions 171 and 274. It is not known whether they form an intrachain disulfide bond.

General References

Vallet, V. et al. *Nature* 1997, **389**:607-610.
Yu, J. X. et al. *J. Biol. Chem.* 1994, **269**:18843-18848.
Yu, J. X. et al. *J. Biol. Chem.* 1995, **270**:13483-13489.
Yu, J. X.et al. *Genomics* 1996, **32**:334-340.

Ref. for DNA/AA Sequences

Yu, J. X. et al. *J. Biol. Chem.* 1995, **270**:13483-13489.

Prostate Specific Antigen

Anders Peter, Johan Malm and Hans Lilja

Synonyms	Human glandular kallikrein 3; γ-seminoprotein; p30
Abbreviations	PSA; hK3
Classifications	EC 3.4.21.77
Description	PSA is a kallikrein-like serine protease. It was first found to be produced under androgen control in the epithelial cells of the prostate, from where it is secreted to semen at ejaculation. However, the protein is not entirely prostate-specific; it is also present at low concentrations in the periurethral glands and other body fluids. The inactive prepro-precursor (261 aa) contains a signal peptide (17 aa) and a propeptide (7 aa). The propeptide is cleaved off by another glandular kallikrein, hK2, which converts the inactive proPSA (zymogen) to the enzymatically active mature form (237 aa). Measurements of PSA in serum is used for detection and monitoring of prostate cancer. In most patients with prostate cancer the level of PSA in serum is elevated, though both cancer and hyperplasia may give rise to moderately elevated levels of PSA. Therefore the ratio of free PSA to PSA in complex with inhibitors may be used to further discriminate between these conditions.
Structure	The predicted structure of PSA is very similar to that of porcine pancreatic kallikrein, with identical or very similar disulphide bridges, folding, and catalytic triad.
Molecular Weight	The calculated molecular weight of the mature protein is 26.1×10^3. Mass spectometry gives a molecular weight of 28.5×10^3 due to glycosylation (a single oligosacharide at Asn-45). SDS-PAGE of the non-reduced protein gives an apparent molecular weight of 28×10^3, while the reduced single chain protein gives an apparent molecular weight of 33×10^3.
Sedimentation Coeff.	Unknown
Isoelectric Point	6.8-7.5
Extinction Coefficient	1.84 and 0.99 (280nm, 1 mg/mL, 1cm) for free PSA and PSA in complex with α_1-antichymotrypsin, respectively.
Enzyme Activity	A chymotrypsin-like serine protease activity.
Coenzymes/Cofactors	Unknown
Substrates	The main substrates are the major gel-forming proteins in semen (semenogelin I and II and to a smaller extent fibronectin, all secreted from the seminal vesicles). In vitro, PSA has also been shown to digest insulin-like growth factor-binding protein 3 (IGFBP-3) and parathyroid hormone related protein (PTHrP). Chromogenic or fluorogenic peptide substrates may be used to determine enzyme activity in vitro.
Inhibitors	Serine protease inhibitors. In vivo, Protein C inhibitor (PCI) in semen and α_1-antichymotrypsin, α_2-macroglobulin and pregnancy-zone protein in serum inhibits PSA. Since the inhibition of PSA with these proteins is

	quite slow, PSA activity in vitro is best inhibited by diisopropyl-phosphofluoridate (DFP) or phenylmethanesulphonyl fluoride (PMSF).
Biological Functions	PSA cleaves the major gel-forming proteins (semenogelin I, semenogelin II, and fibronectin) in human semen into small soluble peptides. This results in liquefaction of semen, with a concomitant release of progressively motile spermatozoa.
Physiology/Pathology	PSA deficiency results in impaired liquefaction of semen. The consequences of this is unknown, but it may affect fertility.
Degradation	Cleavages are known to take place at Lys145-Lys146 and Lys182-Ser183, resulting in inactive forms of PSA.
Genetics/Abnormalities	The gene, which contains five exons, is located on the long arm of chromosome 19 in a region of 65-70 kb also including the genes for hK1 and hK2. The promoter region includes an androgen responsive element.
Half-life	Free PSA has a half-life of about 12-14 hours in serum. However, the predominant form in serum, PSA in complex with α_1-antichymotrypsin, is cleared much slower from the circulation (approximately 0.8 ng/mL per day).
Concentration	0.2-5 mg/mL in seminal fluid and normally less than 3 µg/L in serum.
Isolation Method	PSA can be isolated from both seminal plasma and prostate tissue. Heparin Sepharose, eluted with a salt gradient, has proven to be useful in both cases. Affinity chromatography with antibodies towards PSA can also be used. However, due to the structural similarity between PSA and hK2, several epitopes are shared between the proteins. Thus, some monoclonal antibodies crossreact.
Amino Acid Sequence	The primary structure of PSA has 62% sequence identity with hK1 and 80% sequence identity with hK2. PSA has an N-linked oligosaccharide attached to Asn-45.
Disulfides/SH-Groups	Five disulfide bridges has been suggested, based on the existence of ten cystein residues and similarity to other serine proteases (for example porcine glandular kallikrein).
General References	Chen, Z. et al. *Clin. Chem.* 1995, **41(9)**:1273-1282. Denmeade, S. et al. *Cancer Res.* 1997, **57**:4924-4930. Diamandis, E.P. and Yu, H. *Urol. Clin. North Am.* 1997, **24(2)**:275-282. Lilja, H. *J. Clin. Invest.* 1985, **76**:1899-1903. McCormack, R. et al. *Urology* 1995, **45(5)**:729-744. Villoutreix, B.O. et al. *Protein Sci.* 1996, **5(5)**:836-851
Ref. for DNA/AA Sequences	Lundwall, Å. and Lilja, H. *FEBS Lett.* 1987, **214(2)**:317-322.

Protein C

Lei Shen and Björn Dahlbäck

Synonyms	Autoprothrombin II-A
Abbreviations	None
Classifications	Electr. mobility influenced by Ca^{2+}
Description	A multidomain plasma glycoprotein. A precursor of a serine protease. Belongs to a larger family of vitamin K-dependent proteins. Mainly synthesized in the liver. It is also synthesized in Leydig cells of testis, in the excretory epithelium of epididymis and in some of the epithelial cells of the prostate glands. Activated by thrombin in conjunction with an endothelial cell membrane protein, thrombomodulin. The activated protein participates in the regulation of blood coagulation by selectively inactivating the two important procoagulant cofactors, factors Va and VIIIa. The complete nucleotide sequence of the gene has been determined. The different introns seem to separate exons that code for functional domains. In the presence of Ca^{2+} an N-terminal γ-carboxyglutamic acid (Gla) rich region is involved in binding of the protein to negatively charged phospholipid surfaces, where the protein exerts its biological function. The Gla region is followed by two epidermal growth factor-like domains. The N-terminal of these domains contains one β-hydroxyaspartic acid residue. This region contains one Ca^{2+}-binding site of functional importance. The serine protease part is located in the heavy chain.
Structure	Not yet crystallized
Molecular Weight	47,456 (apoprotein); 61,600 (including 23% carbohydrate). 21,000 (light chain, including carbohydrate); 41,000 (heavy chain, including carbohydrate).
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	14.5 (280nm, 1%, 1cm)
Enzyme Activity	Proenzyme of a serine protease.
Coenzymes/Cofactors	Protein S and intact factor V.
Substrates	Coagulation factor Va and factor VIIIa; S-2366 and Spectrozyme PCa (synthetic chromogenic substrate).
Inhibitors	1. Protein C inhibitor (PCI = plasminogen activator inhibitor 3 = PAI-3), a member of the serine protease inhibitor (serpin) superfamily. PCI inhibits protein C by forming an equimolar complex with the enzyme. 2. α_1-antitrypsin (α_1- AT). 3. α_2-macroglobulin.
Biological Functions	The circulating zymogen form is activated by thrombin in conjunction with an integral endothelial cell membrane protein, thrombomodulin. The activated form (APC) functions as an anticoagulant enzyme by selective degradation of the procoagulant cofactors, factor Va and VIIIa. This

degradation requires Ca^{2+}, phospholipid and an intact active site in the enzyme. Protein S and factor V function as synergistic cofactors in this reaction. A recently described protein C receptor presents on the surface of endothelial cells increases the rate of protein C activation. The protein C receptor may be involved in modulating the role of protein C in regulation the inflammatory response.

Physiology/Pathology

Inherited deficiency delineates an increased risk for thromboembolic disease. The homozygous form of the disease results in a severe deficiency (< 5% of normal plasma level) and the occurence of thromboses in the microcirculation in early life. A severe deficiency can also be found in the starting phase of oral anticoagulant therapy in heterozygotes. A relationship between heterozygous deficiency and venous thromboembolism has also been established. Here the first symptoms usually occur in early adulthood. Influence of other factor(s) on the clinical expression of thrombotic disease in patients with heterozygous deficiency has been suggested. The prevalence of heterozygous deficiency has been estimated to be 1 in 200 to 300 individuals. APC-resistance, which is a recently elucidated major cause of familial thrombophilia is associated with a single point mutation in the factor V gene, changing Arg^{506} in the APC cleavage site to a Gln. As a result, the APC-mediated cleavage of mutated factor Va is impaired as compared to normal factor Va. The impaired APC-degradation of mutated factor Va is presumably the most important mechansim by which APC-resistance yields a hypercoagulable state and an increased risk of thrombosis. The factor V mutation is confined to the Caucasian population and prevalence up to 15% has been reported.

Degradation

APC is slowly inhibited by its plasma inhibitors. No data are available concerning the elimination of the complexes from the circulation.

Genetics/Abnormalities

The gene is located on chromosome 2, spans 11 kb of DNA and is made up of 9 exons and 8 introns. A few abnormalities have been described at the genetic level, resulting in antigen and/or functional deficiency of protein C.

Half-life

6 - 8 hours

Concentration

In plasma 3 - 5 mg/L.

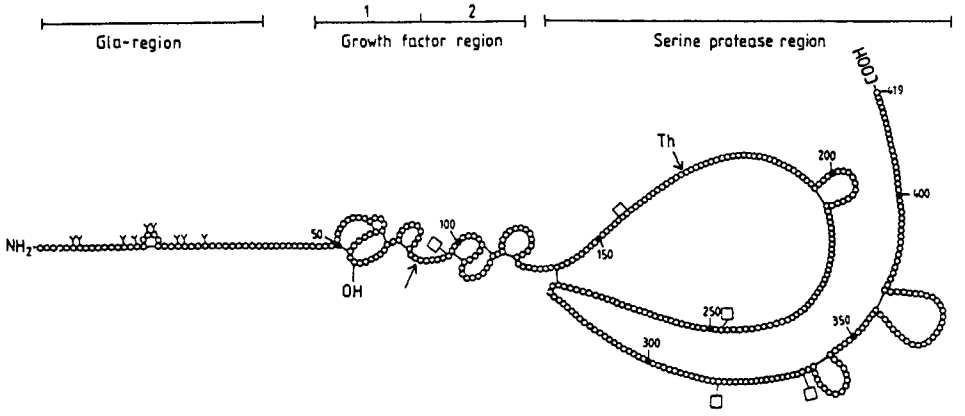
Isolation Method

Isolated from citrated blood plasma by $BaCl_2$ precipitation, ammonium sulphate precipitation, ion-exchange chromatography and preparative electrophoresis. Alternatively, protein C can be isolated from citrated plasma directly by ion-exchange chromatography followed by immunoaffinity chromatography using monoclonal antibodies.

Amino Acid Sequence

The primary structure has been deduced from a cDNA sequence. Bovine protein C has been sequenced by conventional protein chemistry methods. The protein is synthesized as a single chain polypeptide, with a 42 aa long leader sequence, consisting of an N-terminal signal peptide that is followed by a propeptide. The mature protein consists of two peptide chains that are interconnected by one disulfide bond; hence it must undergo postsecretion processing. The details of this processing are not well understood. The protein contains two postribosomally modified aa: 9 γ-carboxyglutamic acid residues and 1 β-hydroxyaspartic residue. It contains 3 potentially N-linked carbohydrate binding sites.

Disulfides/SH-Groups	12 disulfide bonds; 8 intrachain in the light chain and 3 in the heavy chain. One of the disulfide bonds connects the two chains. No free sulfhydryl groups.
General References	Stenflo, J. *J. Biol. Chem.* 1976, **251**:355-363. Kisiel, W. *J. Clin. Invest.* 1983, **64**:761-769. Dahlbäck, B. *Thromb. Res.* 1995, **77**:1-43. Esmon, C.T. and Fukudome, K. *Semin. Cell Biol.* 1995, **6**:259-268. Esmon, C.T. *FASEB. J.* 1995, **9**:946-955. He, X., Shen, L., Bjartell, A., et al. *J. Histochem. Cytochem.* 1995, **43**:563-570.
Ref. for DNA/AA Sequences	Foster, D. and Davie, E.W. *Proc. Natl. Acad. Sci. USA* 1984, **81**:4766-4770. Foster, D.C. et al. *Proc. Natl. Acad. Sci. USA* 1985, **82**:4673-4677. Plutzky, J. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:546-555. Long, G.L. et al. *Somatic Cell and Molecular Genetics* 1988, **14**:93-98.



Schematic molecular model of human protein C according to Öhlin, A.-K., and Stenflo, J. *(J. Biol. Chem.* 1987, **262**, 13798–13804)

Protein C inhibitor

Koji Suzuki

Synonyms	Plasminogen activator inhibitor-3
Abbreviations	PCI; PAI-3
Classifications	Electr. mob. late-α1-globulin
Description	A circulating plasma protein, synthesized in the liver, kidney, seminal vesicles, megakaryocytes and neutrophils, and also present in platelets, urine, Graaf follicle, synovial fluid and seminal fluid. A single chain glycoprotein containing 3 Asn- and 2 Thr/Ser-linked carbohydrate binding sites. A member of plasma serine protease inhibitor (serpin) family proteins involving antithrombin III, heparin cofactor II, α_1-antitrypsin, α_1-antichymotrypsin, α_2-plasmin inhibitor, PAI-1 and PAI-2.
Structure	A globular molecule composed of a single chain glycoprotein consisting of 387 aa. A reactive site is located at Arg-354 to Ser-355.
Molecular Weight	57,000 (SDS-PAGE)
Sedimentation Coeff.	Unknown
Isoelectric Point	4.5-6.0 (microheterogeneity)
Extinction Coeff.	14.1 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	A major inhibitor for activated protein C (APC), and also inhibits plasma kallikrein, thrombin, factor Xa, factor XIa and urokinase (u-PA), but not plasmin and factor XIIa. Serine protease, acrosin that is existed in the sperm head, is also inhibited by PCI. During inhibition of enzyme, a reactive site of PCI is cleaved by enzyme and an acyl-bond complex between PCI and enzyme is formed. A part of PCI is cleaved by enzyme as a substrate. A part of PCI in seminal plasma is complexed with prostate-specific antigen (PSA) and seminal coagulation protein, semenogelin. PCI also inhibits human sperm-egg binding. PCI has affinity for heparin and negatively charged dextran sulfate, and inhibition rate of PCI is accelerated by heparin dextran sulfate and phospholipids, but not by chondroitin sulfate A and C, dermatan sulfate, heparan sulfate nor hyaluronic acid.
Physiology/Pathology	PCI decreases in plasma of patients with disseminated intravascular coagulation and liver diseases. A complex with APC increases in patients with disseminated intravascular coagulation, thrombotic disease, lupus anticoagulant or diabetes mellitus. PCI also decreases in seminal plasma of patients with nonfunctional seminal vesicles.

1. Installment 1992 (revised 1997)

Degradation	Degraded form with Mr 54,000 is produced by cleavage of peptide bonds at the reactive site by plasma kallikrein and other proteases. Urinary PCI is almost all degraded.
Genetics/Abnormalities	Synthesized from liver mRNA and may be also from seminal vesicle, kidney and megakaryocyte mRNA. Gene is 13 kbp in size and consists of 5 exons and 4 introns. No abnormality have been reported. Localization: chromosome 14q31-32.3. Regulation of Transcription: Sp1-binding site is a major promoter region on expression of PCI gene in the liver.
Half-life	23.4 hrs (PCI); 19.6 min (APC-PCI complex)
Concentration	Plasma: 5.3 mg L^{-1} (range 3.3-6.8 mg L^{-1}) Urine: 0.75 mg L^{-1} (range 0.5-1.0 mg L^{-1}) Seminal plasma: appro. 220mg L^{-1}
Isolation Method	Isolated from citrated plasma by barium citrate precipitation followed by dextran sulfate-agarose chromatography, and either by DEAE-Sepharose chromatography and heparin-agarose chromatography or affinity chromatography with monoclonal anti-PCI-IgG-agarose. Isolated from seminal plasma by affinity chromatography with monoclonal anti-PCI-IgG-agarose.
Amino Acid Sequence	The reactive site for APC is at Arg-354/Ser-355.
Disulfides/SH-Groups	One Cys residue; no disulfide bonds.
General References	Suzuki, K. et al. *J. Biol. Chem.* 1983, **258**:163-168. Suzuki, K. et al. *J. Biochem.*(Tokyo) 1984, **95**:187-195. Heeb, M.J. et al. *J. Biol. Chem.* 1987, **262**:15813-15816. Meijers, J.C.M. et al. *Biochemistry* 1988, **27**:4231-4237. Suzuki, K. et al. *Thromb. Haemostas.* 1989, **61**:337-342. Laurell, M., Stenflo, J. *Thromb. Haemostas.* 1989, **62**:885-891. Espana, F. et al. *Thromb. Haemostas.* 1991, **65**:46-51. Zheng, X.L. et al. *Am. J. Physiol.* 1994, **C**466-472. Hermans, J.M. et al. *Biochemistry* 1994, **33**:5440-5444. Kise, H. et al. *Eur. J. Biochem.* 1996, **238**:88-96.
Ref. for DNA/AA Sequences	Suzuki, K. et al. *J. Biol. Chem.* 1987, **262**:611-616. Meijers, J.C.M. and Chung, D.C. *J. Biol. Chem.* 1991, **266**:15028-15034. Hayashi, T., Suzuki, K. *Int. J. Hematol.* 1993, **58**:213-224.

Protein kinase C

Doriano Fabbro and Christoph Borner

Synonyms	Ca^{2+} and phospholipid-dependent protein kinase; C kinase; Phorbol ester receptor; Conventional, classical or calcium-dependent PKC; non-conventional or novel PKC; atypical or abnormal PKC.
Abbreviations	PKC; C-kinase; cPKC; nPKC; aPKC
Classifications	EC 2.7.1.-; Serine-threonine specific protein kinase.
Description	PKC is a multifunctional S- and T-specific protein kinase which utilizes agonist-induced "second messengers" like diacylglycerol (DAG) and/or other lipids in combination with or without calcium to modulate various cellular functions through phosphorylation of specific cellular protein substrates. PKC is the major cellular receptor for a number of tumor promoting and non-tumor promoting agents of different chemical structures such as phorbol esters, indole alkaloids (teleocidins), polyacetates (aplysiatoxin) or macro cyclic lactones (bryostatin). Distribution and localization of PKC in tissues and cells depends on the subtype and cell type. PKC consists at present of a multi-gene family of at least 10 enzymes which can be classified into 3 groups: conventional calcium-dependent and DAG-dependent PKC subtypes (cPKCs: α-, βI-, βII- and γ-PKCs), non-conventional calcium-independent but DAG-dependent PKC subtypes (nPKCs: δ-, ε-, η-, θ- and γPKCs) the atypical calcium-independent and DAG-unresponsive PKC subtypes (aPKCs: ζ- and λ-PKCs). The latter class has structural similarities to the raf kinase family.
Structure	Unknown
Molecular Weight	67.2 - 102 kDa, calculated molecular weight depend on the subtype. 74 - 115 kDa, apparent molecular size (SDS-PAGE).
Sedimentation Coeff.	Unknown for individual subtypes
Isoelectric Point	Unknown for individual subtypes
Extinction Coeff.	Unknown for individual subtypes
Enzyme Activity	PKC is a protein kinase which catalyzes the transfers of the γ-phosphate of ATP onto S or T of substrate proteins. PKC binds DAG and phorbol esters with high affinity in a phospholipid-dependent manner with the exception of aPKCs.
Coenzymes/Cofactors	For full activity the combination of anionic phospholipids (phosphatidyl-L-serine) with DAG or phorbol esters is obligatory. Calcium enhances the activity only of the cPKCs. Other potential lipid cofactors capable of partially substituting and/or synergizing with DAG or phospholipids are unsaturated free fatty acids (FFA), lysophosphatidylcholine (Lyso-PC), phosphatidylinositol 4,5-bisphosphate (PIP_2), phosphatidylinositol 3,4,5-trisphosphate (PIP_3). No coenzymes known.
Substrates	Cellular substrates: A variety of potential cellular substrates for PKC have been reported including cytoskeletal proteins (PLECKSTRIN, MARCKS, talin, vinculin, neuromodulin, lamin B, etc.), transcription factors (CREBP, myc, IκB, etc.), enzymes (raf, GSK-3, topoisomerase, etc.), receptors

(epidermal growth factor receptor, transferrin receptor, etc.) calcium channels and gp170/mdr-1. PKC has a broad substrate specificity in vitro for basic proteins like histones, protamine sulfate and myelin basic protein. All PKCs efficiently phosphorylate protamine sulfate in co-factor-independent manner. Synthetic peptides derived from the pseudosubstrate region (ps) of PKC are used as "PKC specific substrates". The consensus phosphorylation sequence of PKC is (S/T)-X-(R/K), where S or T is the phosphorylation site.

Inhibitors

Various naturally ocurring polypetide inhibitors have been isolated from tissues but are poorly characterized with respect to the mechanism of inhibition and subtype specificity. A variety of synthetic or semi-synthetic inhibitors are either directed towards the regulatory (DAG/phorbol ester binding site) or catalytic site of PKC. Regulatory site directed inhibitors (displacement of phorbol ester binding concomitant with inhibition of catalytic activity) include calphostin C, sphingosine and related lipoidal amines. Catalytic site directed inhibitors against the substrate binding site are peptides based on the pseudosubstrate sequence (ps) of PKC. Catalytic site directed inhibitors directed against the ATP-binding site are ATP analogs (i.e. 5'-p-fluorosulfonylbenzoyladenosine), the indolecarbazole staurosporine and its derivatives as well as isoquinoline derivatives (H7). All these inhibitors are poorly characterized with respect to their selectivity against the various PKC subtypes as well as against other protein kinase families.

Biological Functions

Reversible protein phosphorylation plays a key role in the regulation of a variety of cellular functions. As a member of the protein kinases, PKC phosphorylates a set of specific intracellular target proteins on S or T to modulate various cellular functions. PKC is the principal transducer for DAG generated upon activation of cell surface receptors by neurotransmitter, hormones or growth factors. DAG can be derived from the hydrolysis of PIP_2, phosphatidylcholine (PC) or phosphatidylinositolglycans (PIG). Hydrolysis of these phospholipids are initiated by the activation of phospholipase C (PLC) or phospholipase D (PLD) by either receptor activated tyrosine phosphorylation or by receptor coupling to G proteins. In contrast phorbol esters directly activate PKC, bypassing the agonist-induced generation of DAG. The various PKC subtypes show distinct enzymological properties, differential tissue expression and subcellular localization, display different modes of activation and have different oncogenic potentials. Since most mammalian cells co-express at least one member of each class it is assumed that the three classes of PKC may play specific roles in the processing of various physiological and pathological responses to extracellular stimuli. Accordingly, PKC, which plays a central role in the early processing of a variety of external signals which is configured as a sequentially ordered protein kinase cascade, is involved in a variety of agonist-induced cellular responses like exocytosis, gene expression, proliferation, differentiation, tumor promotion and tumor progression.

Physiology/Pathology

The multi-gene family of PKC has been implicated in a variety of cellular disorders where signal transduction pathways are deregulated. However, the functions and contributions of this enzyme family to a particular pathophysiology has remained elusive. PKC has been implicated in tumorigenesis due to the fact that this enzyme family represents the major receptor for tumor promoting phorbol esters. The multistage carcinogenic process (initiation, promotion, progression) involves a close cooperation between oncoproteins or anti-oncogenes with PKC in regulating cellular growth and neoplastic transformation. A variety of oncogenes increase

DAG levels and permanently activate PKC. Alternatively, activation of PKC by tumor promoters or overexpression of a specific PKC isoform increases the transforming potential of oncogenes. In addition when normal PKC subtypes are overexpressed in fibroblasts they display different oncogenic potentials. However, PKC seems not to be an oncogene in its own right. Accordingly, an oncogenic version of PKC has not been described so far. The PKC oncogene may be found within the family of the raf protein kinases which share similarity with the class of aPKCs. The complexity of synergistic interaction between PKC and oncogenes is compounded by the fact that PKC is a multigene family. Accordingly, different types of tumors have variable expression levels and subcellular localization of PKC subtypes. In addition, the deregulated PKC activity in tumors may contribute to tumor progression by continuous phosphorylation of the mdr-1/gp170, a membrane-spanning protein product of the mdr-1 gene (gp170) which functions as a multi-drug transporter by depleting cells from cytotoxic agents. Once overexpressed, this multi-drug transporter confers resistance to tumor cells against clinically used chemotherapeutic agents like anthracyclins and vinca alkaloids.

Degradation

Sustained activation of PKC by either DAG or tumor promoting agents results in a proteolytic degradation of PKC, a process referred to as the down regulation of PKC. Proteolytic degradation of PKC is initiated by cellular proteases (calpains) on the exposed hinge or linker region of the activated enzyme generating a N-terminal regulatory domain and a C-terminal catalytic fragment. Both fragments have short half-lives and are further degraded. The rate of degradation depends on the subtype.

Genetics/Abnormalities

None so far described

Half-life

>48 h: resting cells; <2-3 h: activated cells (down regulation)

Concentration

Depends on tissue and cell type. Highest expression of most PKC subtypes is found in brain.

Isolation Method

Conventional chromatographic methods yield usually a mixture of PKC subtypes from various tissues or cells which can be resolved into individual subtypes by additional laborious chromatographic steps. PKC expressed in bacteria bind phorbol esters but are inactive with respect to protein kinase activity. Expression of PKC in yeast has not been successful. Isolation of PKC subtypes expressed in mammalian cell systems (COS and CHO cells), are hampered by the presence of endogenous PKC subtypes. Among the recombinant techniques the baculovirus expression system has allowed reliable purification of individual PKC subtypes in reasonable quantities by conventional chromatography. The advantage of this expression system is the fact that insect cells have undetectable levels of co-factor dependent PKC activity.

Amino Acid Sequence

The PKC molecule contains conserved sequences (C1-C4) and variable subtype specific stretches (V1-V5). Interspecies differences within one particular subtype are lower than between the subtypes. Homologies in conserved sequences range between 45 to 85%. If PKC is minimally defined as a S/T-specific protein kinase containing at least one cysteine-rich repeat the members of the raf protein kinase family may be included into the PKC family.

A flexible hinge region (V3 or V2/V3) links the C-terminal catalytic protein kinase domain (C3V4C4V5; V4 is only present in the cPKC-γ) containing highly conserved residues common to all S/T-specific protein

kinases to the N-terminal regulatory domain (V1/V0C1V2/V3 or V1C1V2C2V3) which is responsible for binding phospholipids, DAG, alterrnative lipids, phorbol esters and/or Ca^{2+}. The very N-terminus is variable in sequence and length (V1 or V1/V0) being longer (V1/V0) in the nPKCs and aPKCs than in the cPKCs. Responsible for the phospholipid-dependent phorbol ester binding or DAG is a single (aPKCs and raf) or duplicated (nPKCs and cPKCs) cysteine-rich motif (C_6H_2) which binds zinc with high affinity (2 moles of tightly bound zinc per C_2H_6) and which is also found in n-chimerin, DAG-kinase, unc-13 and the vav-oncogene. The aPKCs as well as the raf kinases which are unable to bind phorbol esters/DAG carry only one C_6H_2 motif. N-terminal to the first cysteine-rich repeat lies the autoinhibitory pseudosubstrate (ps) (not present in the raf kinases) which inhibit PKC activity in the resting state. The C2 region is only present in the cPKCs and presumably functions as a calcium-dependent phospholipid binding domain. This C2 domain is conserved in proteins other than cPKCs like the PLC-γ, ras-GAP, synaptogamins, unc-13 and cytoplasmic form of the phospholipase A2.

Disulfides/SH-Groups

Unknown

General References

Asaoka, Y. et al. *Trends Biochem. Sci.* 1992, **17**:414-418.

Protein kinase C: Current concepts and future perspectives. Lester, D. and Epand, R. (eds.). Ellis Horwood Series in Biochemistry and Biotechnology. Chichester, West Sussex, England, 1992.

Houslay, M.D. *Eur. J. Biochem.* 1991, **195**:9-27.

Matter, A. et al. *The therapeutic implications of the molecular biology of breast cancer.* PEZCOLLER Foundation Symposia, Vol. 2. Lippman, M. and Mihich, E. (eds.). John Libbey CIC, Rome, 1992, pp. 227-247.

Protein Phosphorylation Part A: Protein kinases: Assays, Purification, Antibodies, Functional analysis, Cloning, and Expression. In: *Methods in Enzymology*. Hunter, T. and Sefton, M.B. (eds.). Academic Press Inc. 1991, **Vol. 200.**

Protein Phosphorylation Part B: Analysis of protein phosphorylation, Protein kinase inhibitors, and Protein phosphatases In: *Methods in Enzymology*. Hunter, T. and Sefton, M.B. (eds.). Academic Press Inc. 1991, **Vol. 201**.

Li, P. et al. *Cell*, 1991, **64**:479-482.

Morrison, D.K. *Cancer Cell* 1990, **2**:377-382.

Ref. for DNA/AA Sequences

When available references, identification and accession numbers of the human PKC subtypes are given.

human cPKC-α: Finzenkeller G. et al. *Nucleic Acid Res.* 1990, **18**:2183. SwissProt:KPCA-HUMAN (P17252); EMBL: HSPKCA1 (X52479).

human cPKC-β1: Kubo, K. et al. *FEBS Lett.* 1987, **223**:138-142. SwissProt: KPC1-HUMAN (P05771); EMBL: HSPKCB1A (X06318).

human cPKC-β2: Kubo, K. et al. *FEBS Lett.* 1987, **223**:138-142. SwissProt: KPC2-HUMAN (P05127); EMBL: HSPKCB2A (X07109).

human cPKC-γ: SwissProt: KPCG-HUMAN (P05129); EMBL: HSPKCA1 (X52479).

human cPKC-δ: SwissProt: KPCD-HUMAN (Q05655); EMBL: HSPKCD13X (L07860).

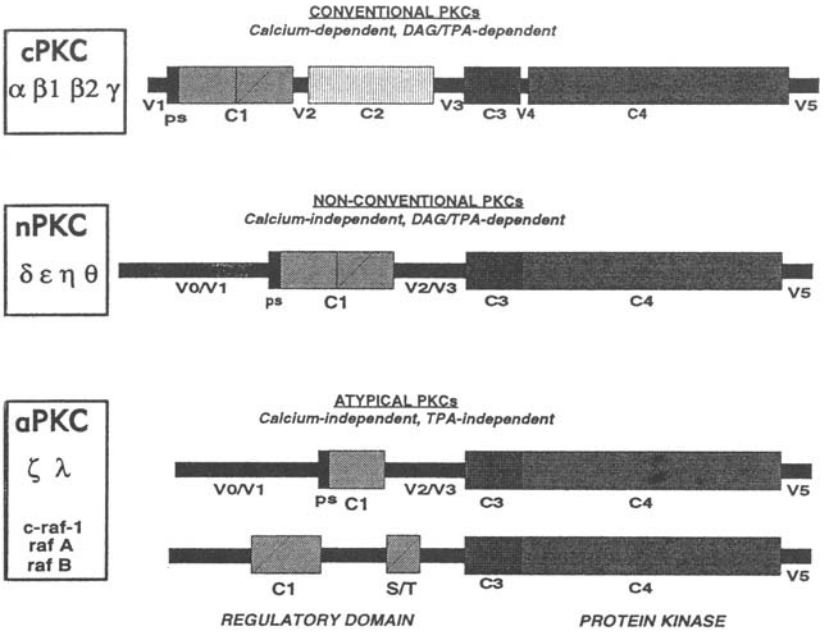
human cPKC-ε: Swiss Prot: none; EMBL: HSPKCE (X65293).

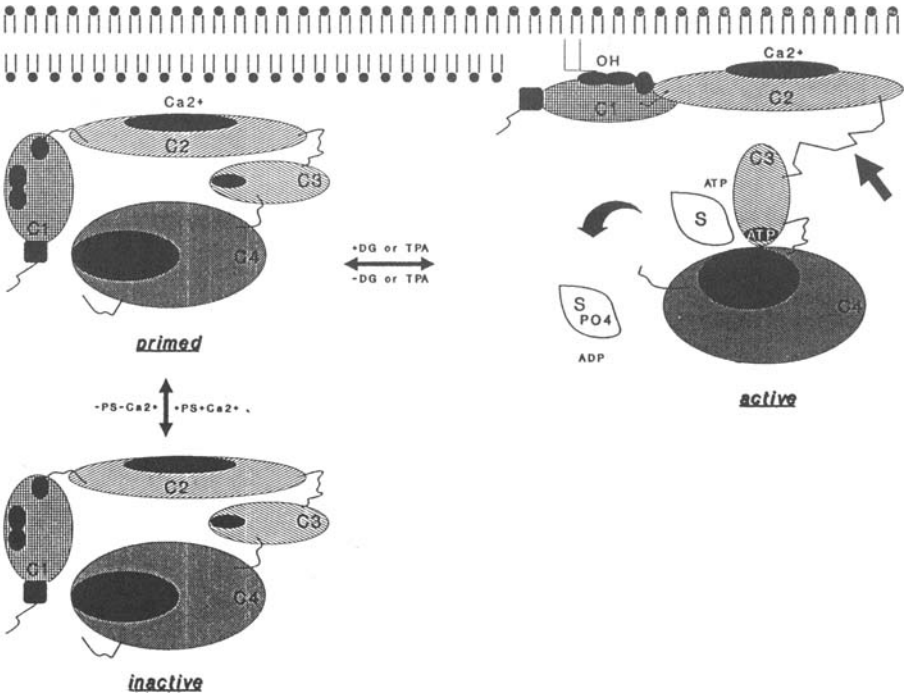
rat cPKC-γ. Ono, Y. et al. *Nucleic Acid Res.* 1988, **16**:5199-5200. SwissProt: KPCG-RAT (P05697); EMBL: HSPKCA1 (X52479).

rat nPKC-δ: Ono, Y. et al. *J. Biol. Chem.* 1988, **263**:6927-6932. SwissProt: KPCD-RAT (P09215); EMBL: RSPKCB (M18330).

rat nPKC-ε: Ono, Y. et al. *J. Biol. Chem.* 1988, **263**:6927-6932. SwissProt: KPCE-RAT (P09216); EMBL: RSPKCE (M18331).

mouse nPKC-η: Osada, S.I. et al. *J. Biol. Chem.* 1990, **265**:22434-22440. SwissProt: KPCL-MOUSE (P23298); EMBL: MMNPKCET (M62980).
mouse nPKC-θ: Osada, S.I. et al. *Mol. Cell. Biol.* 1992, **12**:3930-3938. SwissProt: not filed; EMBL: not filed; GenBank S111692 (D11091).
rat aPKC-ζ: Ono, Y. et al. *Proc. Natl. Acad. Sci. USA* 1989, **86**:3099-3103. SwissProt: KPCZ-RAT (P09217); EMBL: RSPKCZ (M18332).
rat aPKC-λ: Asaoka, Y. et al. *Trends Biochem. Sci.* 1992, **17**:414-418. Sequence not yet published.





Molecular Model

Proposed model for activation of PKC:
1) inactive: In this form the PKC is not membrane bound and its activity is presumably blocked by the interaction of ps (filled square) in C1 with the peptide binding site in C4 (filled ellipse).
2) primed: The C2 domain binds anionic phospholipids (PS) in a calcium-dependent manner. In this form the enzyme is still inactive but bound to membranes.
3) active: Phorbol esters (TPA) or DAG bind to the cysteine-rich clusters (filled) in C1 leading to a conformational change which releases the ps from the peptide binding site. This allows the interaction of ATP and substrate (S) with the catalytic domain (C3C4) and the transfer of the γ-PO4 of ATP on S/T of the protein substrate (S). This conformational change exposes the hinge region V3 which eventually becomes accessible to proteases (arrow, down regulation).
The above model illustrates the steps leading to the activation of cPKCs. In the absence of a C2 domain (nCPKs) the activation depends on the presence of DAG, TPA or other activating lipids in the membrane (omission of step 1). Activation of aPKCs and raf kinases may occur through a similar lipid-dependent mechanism although they do not bind DAG nor phorbol esters. In analogy to the alternative modes of raf kinase activation, the aPKC may be activated by homologous and/or heterologous phosphorylation.

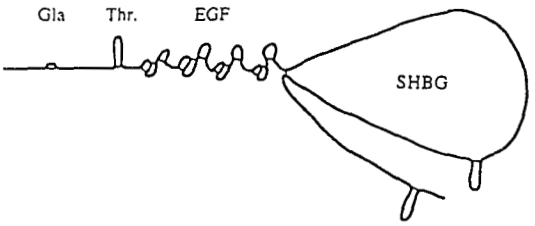
Protein S

Andreas Hillarp and Björn Dahlbäck

Synonyms	None
Abbreviations	None
Classifications	EM: α_1 (- Ca2+), β_2 (+ Ca^{2+})
Description	A circulating plasma protein, synthesized in the liver and present in platelets. It is also synthesized by megakaryocytes, by endothelial cells, in the reproductive tissues and in the brain. The different forms of protein S are probably identical. A single-chain (635 aa) molecule composed of discrete regions. An N-terminal γ-carboxyglutamic acid (Gla) rich contains several Gla-dependent Ca^{2+}-binding sites (K_d approx. 0.5 mM). In the presence of Ca^{2+} this region is responsible for the binding to negatively charged phospholipids. The Gla-region is followed by the thrombin-sensitive region, containing two peptide bonds sensitive to cleavage by thrombin. The third region in the protein S molecule is the EGF-region composed of four domains homologous to the epidermal growth factor precursor protein. The most N-terminal EGF-domain contains a β-hydroxyaspartic acid residue (Hya) and the following three a β-hydroxyasparagine residue (Hyn) each. The EGF region contains high affinity Ca^{2+}-binding sites (K_d approx. 20 μM and > 10^{-8}) and is probably responsible for the interaction with activated protein C. The C-terminal half of the molecule is homologous to sex hormone binding globulin (SHBG). Human protein S is a glycoprotein and contains three potential N-glycosylation sites. Protein S circulates both as a free protein and bound to C4b-binding protein, a regulator of the complement system. In normal plasma, approximately 40% of protein S occurs in free form, and only in this free form is protein S active as a cofactor to activated protein C.
Structure	More rod-like than spherical (indicated by frictional coefficient and axial ratio).
Molecular Weight	75,000 (SDS-PAGE, nonreduced), 70,690 (aa sequence).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.0 - 5.5
Extinction Coeff.	9.0 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Participates in the regulation of blood coagulation. Protein S is a cofactor to activated protein C (APC) in the degradation of the activated coagulation factors V and VIII, thereby inhibiting coagulation. Protein S exerts its APC cofactor activity by enhancing the binding of APC to phospholipid surfaces and by abrogating factor Xa- and factor IXa-

	dependent protection against APC of factor Va and factor VIIIa, respectively. Protein S is also reported to express APC-independent anti-coagulant activity, by inhibiting activation of prothrombin through direct interactions with factor Va and factor Xa.
Physiology/Pathology	Patients with congenital heterozygous protein S deficiency suffer an increased risk for thromboembolic disease, the first episodes often occur in early adulthood. Two forms of protein S deficiency are seen. In one form the total concentration of protein S is decreased while the ratio between free protein S and protein S in complex with C4b-binding protein is normal. In the other form of protein S deficiency the total concentration of protein S is normal or only mildly decreased while the concentration of free protein S is very low, i.e. most of the protein S is complexed to C4b-binding protein.
Degradation	No data available concerning the elimination from the circulation. Protein S can be cleaved by thrombin after which the cofactor activity is lost. Not known whether this occurs in vivo.
Genetics/Abnormalities	There are two protein S genes, one active gene (PSα or PROS1) and one silent pseudogene (PSβ), in the human genome. The protein S locus is located on chromosome 3q11.2. PSα span over 80 kb of DNA and comprises 15 exons. Several genomic abnormalities, involving deletions and deleterious point mutations, responsible for protein S deficiencies have been identified.
Half-life	Unknown
Concentration	Total protein S concentration: 25 mg/L (0.3 μM) (normal range 17 - 33 mg/L), concentration of free protein S: 10 mg/L (normal range 5 - 15 mg/L).
Isolation Method	Isolated from citrated plasma by $BaCl_2$-precipitation followed by ion-exchange chromatography or affinity chromatography.
Amino Acid Sequence	The respective regions mentioned above (see 'description') are homologous to the corresponding regions in the other vitamin K-dependent coagulation proteins.
Disulfides/SH-Groups	17 disulfides (all intrachain); no free sulfhydryls.
General References	Dahlbäck, B. *Thromb. Haemost.* 1991, **66**:49-61. Dahlbäck, B. and Stenflo, J. In: *The molecular basis of blood diseases.* W.B. Saunders Co. Philadelphia, PA. 1994, pp. 599-628.
Ref. for DNA/AA Sequences	Lundwall, Å. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:6716-6720. Hoskins, J. et al. *Proc. Natl. Acad. Sci. USA* 1987, **84**:349-353. Ploos van Amstel, H.K. et al. *FEBS Lett.* 1987, **222**:186-190. Schmidel, D.K. et al. *Biochemistry* 1990, **29**:7845-7852. Ploos van Amstel, H.K. et al. *Biochemistry* 1990, **29**:7853-7861. Edenbrandt, C.G. et al. *Biochemistry* 1990, **29**:7861-7868.

Molecular model according to B. Dahlbäck et al. (*Proc. Natl. Acad. Sci. USA* 1986, **83:** 4199–4203)



Prothrombin

Craig M. Jackson

Synonyms	Factor II, Thrombinogen (archaic), Proserozyme (archaic)
Abbreviations	F-II, II
Classifications	α_2-globulin, zymogen/proenzyme of thrombin
Description	Single polypeptide chain precursor of the blood clotting proteinase thrombin. Prothrombin is synthesized in the liver, requires vitamin-K for synthesis in fully biologically active form, post-translationally modified in reactions requiring vitamin-K to produce 10 γ-carboxyglutamate residues from 10 glutamate residues.
Structure	Prothrombin contains 579 aa residues and is comprised of several distinct domains: 1) γ-carboxyglutamate-containing (Gla domain); 2) kringle 1, which together with the Gla domain comprise the Fragment 1 domain (residues 1–155); 3) kringle 2 or the Fragment 2 domain, (residues 156–271); and 4) the proteinase domain (residues 272–579). Three dimensional structures of the Fragment 1 domain and the thrombin domain are known from X-ray crystallography.
Molecular Weight	Prothrombin: 72,500 (amino acid sequence), 72,000 ± 3,000 (sedimentation equilibrium). Proteolysis products generated during activation: Fragment 1 : 22,300 (aa sequence), 26,000 (sedimentation diffusion); Fragment 2 : 12,866 (aa sequence), 16,600 (sedimentation diffusion); Fragment 1–2: 35,200 (aa sequence); Prethrombin 1: 50,200 (aa sequence), 51,000 (sedimentation equilibrium, bovine); Prethrombin 2: 37,400 (aa sequence); Thrombin: 37,400 (aa sequence).
Sedimentation Coeff.	4.6S (Prothrombin); 3.8S (Prethrombin 1, bovine); 2.3S (Fragment 1, bovine)
Isoelectric Point	4.2–4.3(prothrombin); 7.5(thrombin) all bovine
Extinction Coeff.	Prothrombin: 14.7, 13.8 (human) 14.4 (bovine); Fragment 1: 11.9 (human); Fragment 2: 13.8 (bovine); Prethrombin 2: 17.3 (human); Prethrombin 1: 17.8 (human); Thrombin: 17.9 (human). All 280 nm, 1%, 1 cm.
Enzyme Activity	Prothrombin: None. Thrombin: Proteinase (EC 3.4.21.5) hydrolyzes Arg-X and Lys-X bonds.
Coenzymes/Cofactors	For Factor Xa – Factor Va (in activation of prothrobin), membrane phospholipids (negatively charged phospholipids are required). The cofactor catalyzes prothrombin activation by binding it and Factor Xa to membrane phospholipid surfaces. Activation velocity increases are from 350 times (Factor Va and Factor Xa with prothrombin) to 300,000 times (Factor Va, membrane phospholipids, Factor Xa and prothrombin), all in the presence of Ca^{2+} ions. Thrombomodulin is a cofactor in the thrombin-catalyzed activation of Protein C.

Substrates	For Thrombin – Naturally Occurring Proteins: Fibrinogen, Factor V, Factor VIII, Factor XIII, thrombospondin, Protein C. For Thrombin – Peptide Chromogenic Substrates: D-Phe-Pip-Arg-pNA (S-2238, AB Kabi),Tos-Gly-Pro-Arg-pNA (Chromozym TH, Pentapharm, Boehringer-Mannheim), and H-D-Hexahydrotyrosyl-Ala-Arg-pNA (Spectrozyme TH, American Diagnostica).
Inhibitors	For Thrombin – Naturally Occurring Protein Inhibitors: Antithrombin III (Heparin cofactor), Heparin cofactor II, α-1 proteinase inhibitor, α_2-macroglobulin, Hirudin. For Thrombin – Low Molecular Weight Inhibitors: D-Phe-Pro-Arg-CH_2Cl, Di-isopropyl phosphorofluoridate, Phenylmethane sulfonyl-fluoride, Dansyl-Arginine-4-ethylpiperidine amide (DAPA). Antithrombin III irreversibly inactivates thrombin. In the presence of heparin, more specifically heparin with a high affinity for Antithrombin III, the inactivation rate for thrombin is increased as many as 10,000 times. Heparin cofactor II similarly irreversibly inactivates thrombin. Many polyanions, e.g., glycosaminoglycans such as heparan sulfate, dermatan sulfate, heparin (independent of any special affinity for Antithrombin III) increase the inactivation rate by as much as 10,000 times. α-1 Proteinase inhibitor irreversibly inactivates thrombin albeit relatively inefficiently. A mutant, α-1 proteinase inhibitor Pittsburgh, in which the reactive site methionine is replaced by arginine is an extremly efficient inhibitor of thrombin. α_2-macroglobulin inhibits thrombin by capturing it within the cage-like inhibitor sturcture. The active site of thrombin is blocked from access by large protein substrates but is accessible to low molecular weight peptide substrates such as the chromogenic substrates listed above. Hirudin inhibits thrombin (Kd $\simeq 10^{-14}$ M) by binding at an exosite on thrombin and at the active site. The crystal structure of the hirudin-thrombin complex has been determined. The peptide chloromethyl ketones react by alkylating the active site histidine residue. the phosphofluoridates and sulfonyl fluorides inactivate thrombin by forming esters with the active site serine.
Biological Functions	Prothrombin is the circulating precursor of thrombin, the proteinase responsible for the clotting of fibrinogen. Prothrombin binds to phospholipids from ruputured cell membranes and platelets via the Gla domain of the Fragment 1 region; a process that localizes the prothrombin activation to the site of injury. The Fragment 2 region of prothrombin binds noncovalently to the immediate precursor of thrombin and also mediates the binding of prothrombin to the cofactor, Factor Va.
Physiology/Pathology	Primary physiological role is as a precursor of thrombin in the hemostatic blood coagulation. Decreased concentration leads to increased risk of bleeding. A decreased concentration of prothrombin molecules with the full complement of γ-carboxyglutamate residues, a situation achieved by administration of vitamin-K antagonist-type anticoagulants, diminishes the rate of prothrombin activation. This alteration in the activation rate is the basis for oral (vitamin-K antagonist) anticoagulant effect.
Degradation	Proteolytic cleavage of prothrombin by Factor Xa at two peptide bonds, Arg-271 and Arg-320, produces catalytically active thrombin. If cleavage of Arg-271 precedes Arg-320, thrombin is designated α-thrombin; if cleavage of Arg-320 precedes Arg-271 the transiently formed species is designated meizothrombin. α-thrombin is the normal, final product of prothrombin activation and possesses activity toward all substrates listed above; meizothrombin activity is as high as α-thrombin toward low molec-

ular weight substrates but not high molecular weight substrates. Factor Va, phospholipids and Ca^{2+} ions increase the rates at which Factor Xa cleaves these peptide bonds during the activation related degradation of prothrombin. The activation fragments from prothrombin are required for the formation of the complex which forms among Factor Xa, Factor Va, membrane phospholipids and Ca^{2+} ions. This complex is the primary physiological activator of prothrombin and is called prothrombinase by some workers.
Prothrombin (human) can be cleaved in vitro by thrombin at Arg-155 to produce Fragment 1 and Prethrombin 1; and at Arg-284 to produce Prethrombin 2′ (a form of Prethrombin 2 to which the first 3 residues of the A-chain of thrombin are attached).
In vitro thrombin can undergo further proteolytic cleavage to a variety of forms which result in altered specificity with respect to protein and low molecular weight substrates and altered reaction with inhibitors.

Genetics/Abnormalities

The gene for prothrombin is located on chromosome 11, p11–q12; inheritance is autosomal and recessive. More than 15 dysprothrombinemias are known. Amino acid substitutions in these altered prothrombins have been reported in the Gla domain and in the thrombin-forming region as well as at the sites of cleavage that are required for prothrombin activation.

Half-life

60 hrs

Concentration

1.5 μM/L, 110 mg/L.

Isolation Method

Isolation of prothrombin from human plasma is readily achieved by precipitation with $BaCl_2$ from plasma collected into sodium citrate anticoagulant, chromatography on heparin-Sepharose or sulfated dextran beads, and anion exchange chromatography. Bovine prothrombin isolation does not require chromatography on heparin-Sepharose or sulfated dextran beads.

Amino Acid Sequence

The single chain in the human prothrombin molecule contains 579 aa residues with N-terminal (Ala) and C-terminal (Glu). The locations for the 10 γ-carboxyglutamate residues in prothrombin are positions 6, 7, 14, 16, 19, 20, 25, 26, 29, and 32. The positions and numbers of the γ-carboxyglutamic acid residues are similar within the entire family of vitamin-K dependent plasma proteins. The kringle domains of Fragment 1 amd Fragment 2 are structural motifs that are found also in plasminogen, tissue plasminogen activator, and in apolipoprotein Lp(a). Thrombin is a classical serine proteinase and is highly homologous with trypsin, chymotrypsin and elastase. Structural differences between thrombin and the pancreatic proteinases are primarily insertions at the ends of the β-strands and lie on or near the surface of the thrombin molecule. The location of the residues of the serine proteinase catalytic triad, His, Ser and Asp are conserved in the thrombin tertiary structure.

Disulfides/SH-Groups

Prothrombin contains 13 intra-chain disulfides. Two (2) are in the Gla domain and three (3) in each of the kringle domains and five (5) in the proteinase-forming domain. After cleavage of Arg-320 one disulfide is an inter-chain bond that links the A- and B-chains of thrombin.

General References

Suttie, J. and Jackson C. M. *Physiological Reviews* 1977, **57:** 1–55.
Miletich, J., et al. *Methods Enzymol.* 1981, **80:** 221–228.
Coleman, R. W., et al. *Hemostasis and Thrombosis* 1987, Ed.2: pp. 1–1507.
Tulinsky, A., et al. *Biochemistry* 1989, **29:** 6805–6810.
Huber, R., et al. *EMBO J.* 1989, **8:** 3467–3475.

Ref. for DNA/AA Sequences

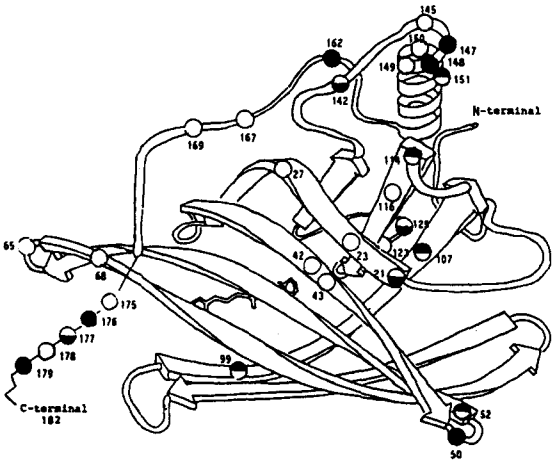
Degen, S. J. F. and Davie, E. W. *Biochemistry* 1987, **26:** 6165–6177.
DiScipio, R. G., et al. *Biochemistry* 1977, **16:** 698–706.

Retinol-binding protein

Lars Rask

Synonyms	None
Abbreviations	RBP
Classifications	Electrophoretic mobility: α_2
Description	A non-glycosylated single-chain plasma protein binding a single molecule of retinol (vitamin A). RBP is in plasma non-covalently bound to transthyretin with a dissociation constant of about 2×10^{-7} M. Under physiological conditions the stoichiometry of the complex is 1:1. The stability of the protein complex is highly dependent on the ionic strength. At an ionic strength of 0.001 at pH 8.0, no binding of RBP to transthyretin can be detected. The association of apo-RBP to transthyretin is at least three-fold weaker than that of holo-RBP. RBP belongs to the lipocalin protein superfamily, consisting mostly of lipophilic carrier proteins.
Structure	182 aa peptide folded to a globular molecule with a Stokes'radius of 2.05 nm and a frictional ratio of 1.28. The molecule is an eight-stranded up-and-down β-barrel cone that encapsulates the retinol molecule. On top of the β-barrel a single α-helix is located.
Molecular Weight	21,200 (sedimentation equilibrium ultracentrifugation); 21,210 (calculated from sequence data).
Sedimentation Coeff.	2.3 S
Isoelectric Point	4.4–4.8, electrophoretically heterogeneous
Extinction Coeff.	18.4 (280 nm, 1%, 1 cm) or 3.91×10^4 ($M^{-1} cm^{-1}$)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Transports vitamin A in the form of retinol from its storage site in the liver to target cells like epithelial cells, where the vitamin is utilized. RBP might also transfer retinol between hepatocytes and stellate cells in the liver. RBP interacts with a cell surface receptor and thereby delivers the retinol molecule to the cell. RBP is mainly synthesized in the liver, but RBP mRNA has also been detected in other tissues like the kidney.
Physiology/Pathology	The release of RBP from the hepatocyte is dependent upon vitamin A. In vitamin A-deficiency, RBP is trapped mainly in the Golgi. Accordingly, in vitamin A-deficiency the plasma concentration of RBP is decreased.
Degradation	Eliminated from circulation by glomerular filtration in the kidney. RBP has also been reported to be degraded in the liver.
Genetics/Abnormalities	No genetic variants have been found.

Half-life	12 hrs for RBP with affinity for transthyretin
Concentration	Plasma: 46 mg/L (range 37–56 mg/L). The plasma concentration is age dependent; Urine: 0.11 mg/24h volume (range 0.04–0.22 mg/24h volume); Cerebrospinal fluid: 0.35 mg/L (range 0.31–0.4 mg/L)
Isolation Method	Isolated from plasma either by ion exchange and gel chromatograhies or affinity chromatography on immobilized transthyretin. Large amounts of RBP can conveniently be isolated from pathological urine by gel chromatography, preparative zone electrophoresis and ion exchange chromatography.
Amino Acid Sequence	Belongs to the protein superfamily of lipocalins.
Disulfides/SH-Groups	Contains three intra-chain disulfide bonds (between Cys4-Cys160, Cys70-Cys174 and Cys120-Cys129).
General References	Rask, L., et al. *Scand. J. clin. Lab. Invest.* 1980, **40, Suppl. 154:** 45–61. Goodman, D. S. Plasma Retinol-Binding Protein. In: *The Retinoids*, Sporn, M. B., et al. (eds.) Academic Press, Orlando 1984, **2:** 42–82. Rask, L., et al. *Upsala J. Med. Sci.* 1987, **92:** 115–146. Newcomer, M., et al. *EMBO J.* 1984, **3:** 1451–1454.
Ref. for DNA/AA Sequences	Colantuoni, V., et al. *Nucleic Acids Res.* 1983, **11:** 7769–7776. D'Onofrio, D., et al. *EMBO J.* 1985, **4:** 1981–1989.



Retinol-RBP schematic drawing, together with mutations in the human, rat and rabbit sequences. The symbol ⊖ respresents a change in human sequence from the consensus sequence, ○ in the rabbit, ⊜ in rat and ● represents all three differing.

Rhodopsin Kinase

Krzysztof Palczewski

Synonyms	G protein-coupled receptor kinase 1
Abbreviations	RK; GRK1
Classifications	A Ser/Thr protein kinase; a member of G protein-coupled receptor kinases subfamily
Description	RK is a soluble protein kinase expressed exclusively in retinal photoreceptor cells and in the pineal gland. RK phosphorylates photoactivated rhodopsin during phototransduction.
Structure	RK is a monomeric protein with an overall structure related to other protein kinases, and in particular to the G protein-coupled receptor kinases. The catalytic domain is located in the mid-region of the RK sequence (position 185–455) and the N-terminal region was implicated in the recognition of photoactivated rhodopsin. RK is post-translationally modified: (1) at the C-terminus by farnesylation, proteolysis, and α-carboxylmethylation; (2) by autophosphorylation at residues Ser-21, Ser-488, and Thr-489; and (3) at the N-terminus is blocked by an unidentified yet modification. RK has not yet been crystallized in its native form.
Molecular Weight	62,934 Da for bovine RK based on amino acid composition, and post-translational modification were not included.
Sedimentation Coeff.	Not determined.
Isoelectric Point	~6.0 (calculated). Several forms are predicted due to multiple autophosphorylation.
Extinction Coeff.	1.0212 (278 nm, 0.1% solution (g/L) in 6 M guanidine hydrochloride at pH 6.5) (calculated).
Enzyme Activity	RK phosphorylates photolyzed rhodopsin at several Ser and Thr residues. The phosphorylation reaction is sequential with the first phosphate transferred preferentially to Ser-338 of rhodopsin, and subsequent phosphorylations are at Ser-343 and Thr-336.
Coenzymes/Cofactors	RK utilizes Mg-ATP complex as a phosphate donor and requires an extra Mg^{2+} for maximum activity.
Substrates	RK is a highly specific protein kinase that phosphorylates photoactivated rhodopsin. It catalyzes also phosphorylation of topologically related receptors such as photolyzed iodopsin (chicken red cone pigment) and agonist-occupied β_2-adrenergic receptor. RK phosphorylates peptide substrates only marginally and presumably by different mechanisms than that of receptors. Mg-ATP is a better substrate than Mg-GTP.
Inhibitors	Inhibitors of RK are: polyanions, peptides derived from the cytosolic surface of rhodopsin, and nucleoside sangivamycin.
Biological Functions	In retinal photoreceptor cells, during transition of photolyzed rhodopsin to inactivated opsin, the receptor assumes three relatively stable conformations metarhodopsin I, II, and III. Metarhodopsin II binds to and activates a photoreceptor-specific G protein, and initiates the signal-amplifying

	cascade of reactions. This activated state of rhodopsin is phosphorylated by RK. The dissociation of the kinase is followed by the binding of arrestin, thereby preventing continuous G protein activation.
Physiology/Pathology	Unknown
Degradation	*In vitro*, RK is susceptible to degradation by various proteolytic enzymes. Presumably, RK is shed with rod outer segments discs to pigment epithelium.
Genetics/Abnormalities	Unknown
Half-life	The estimated half-life is 30 hours, and the protein is considered to be stable.
Concentration	10–50 μM in rod cells
Isolation Method	RK is isolated from photoreceptor outer segments using affinity chromatography on Heparin-Sepharose. RK is prepared in its dephosphorylated form and bound to Heparin-Sepharose. An addition of ATP leads to autophosphorylation of RK and elution from the column.
Amino Acid Sequence	The catalytic domain of RK contains essential motifs of protein kinases: G-x-G-x-x-G implicated in a fold that is responsible for ATP binding at position 193; invariable Asp-332 and Lys-216 involved in catalysis.
Disulfides/SH-Groups	Ten Cys residues identified in the sequences, from which an unknown number of disulfide bridges may be formed.
General References	Kühn, H. *Progress in Retinal Res.* 1989, **3:** 1123–1156. Palczewski, K. et al. *J. Biol. Chem.* 1988, **263:**14067–14073. Palczewski, K. et al. *J. Biol. Chem.* 1991, **266:**12949–12955. Palczewski, K. et al. *Neuron.* 1992, **8:**117–126.
Ref. for DNA/AA Sequences	Lorenz, W. et al. *Proc. Natl. Acad. Sci. USA.* 1991, **88:**8715–8719.

Secretogranin II

Hans-Hermann Gerdes and Wieland B. Huttner

Synonyms	Chromogranin C, see nomenclature proposal in *Neuroscience* (1987), **21:** 1019–1021
Abbreviations	Secretogranin II: SgII; chromogranin C: CgC
Classifications	Member of the chromogranins/secretogranins, referred to in abbreviated form as granins.
Descriptions	Highly acidic secretory protein found in secretory granules of a wide variety of endocrine cells and neurons. Together with chromogranin A and chromogranin B, two other members of the granin family, it can be considered as the most widespread marker of the matrix of neuroendocrine secretory granules.
Structure	The mature protein is a single polypeptide of 587 aa. It contains 3–5% mainly O-linked carbohydrate, binds calcium with moderate affinity at multiple sites and aggregates in the presence of this divalent cation (*). These properties probably reflect both the excess of negative charge and the presumptive secondary structure of this protein, which largely consists of helices and turns. NMR and X-ray diffraction studies are not available.
Molecular Weight	67,757 Da: deduced from cDNA sequence without the signal peptide and posttranslational modifications. 85,000: from SDS-PAGE under reducing conditions.
Sedimentation Coeff.	Unknown
Isoelectric Point	5.0
Extinction Coeff.	Unknown
Enzyme Activity	Unknown
Coenzymes/Cofactors	Unknown
Substrates	Unknown
Inhibitors	Unknown
Biological Function	Various functions have been suggested from observations obtained under *in vitro* conditions. A proposed extracellular function is that the intact protein or proteolytic fragments derived from it exert biological activities on target cells. In this context it was shown that SgII is the precursor of secretoneurin, a peptide which induces dopamine release from rat caudate-putamen slices in a dose-dependent manner. A proposed intracellular function is that of a helper protein in packaging and/or processing of certain peptide hormones and neuropeptides.
Physiology/Pathology	SgII and its proteolytic products are released from cells by regulated exocytosis into the extracellular space and may be also detectable in blood plasma.
Degradation	In the *trans* Golgi network, and during its storage in neuroendocrine granules, SgII undergoes partial proteolytic processing. This process in-

	volves specific endoproteases that cleave the precursor protein at dibasic and monobasic sites. The amount of proteolytic products varies with tissue. Two products have been identified: secretoneurin from brain, corresponding to bovine $SgII_{154-186}$ and AMENM from adrenal medulla, corresponding to bovine $SgII_{582-586}$.
Genetics/Abnormalities	The gene (presumably a single copy gene) was located to the q35-q36 region of chromosome 2 (O'Connor, D. personal communication). It is (without the promoter region) 5.5 kb long and consits of 2 exons, of which exon 2 encodes for the entire SgII protein (*).
Half-Life	Unknown
Concentration	Unknown
Isolation Methods	a) Preparation of chromaffin granules from adrenal medulla by sucrose density gradient centrifugation, followed by the isolation of soluble matrix proteins and sequential application of HPLC anion exchange, gel filtration and reversed phase chromatography and preparative SDS-PAGE (*). b) Due to its heat stability a highly enriched soluble fraction of SgII can be obtained from pituitary adenomas after homogenization and 5 min boiling at pH 7.4 and 150 mM NaCl followed by centrifugation.
Amino Acid Sequence	Characteristic features are: high content of acidic residues (Glu + Asp = 19.9%) and of Pro (7%), secretoneurin sequence at residues 152–184, potential N-glycosylation site at Asn 346, potential phosporylation sites at Ser and Tyr, putative tyrosine sulfation site at Tyr 121 and nine potential dibasic cleavage sites. Notably, SgII lacks the N-terminal disulfide-bonded loop structure and the C-terminal domain which are highly conserved in the two other classical granins, chromogranin A and chromogranin B.
Disulfides/SH-Groups	None
General References	Wiedenmann, B. and Huttner, W. B. *Virchows Archiv B Cell Pathol.* 1989, **58:** 95–121. Huttner, W. B. et al. *Trends Biochem. Sci.* 1991, **16:** 27–30. Rosa, P. and Gerdes, H.-H. *J. Endocrinol. Invest.* 1994, **17**: 207–225.
Ref. for DNA/AA Sequences	Gerdes, H.-H. et al. *J. Biol. Chem.* 1989, **264**: 12009–12015 (aa sequence). Schimmel, A. et al. *FEBS Lett.* 1992, **314**: 375–380
	(*) data are obtained for SgII of other species. It can be assumed that they are also valid for human SgII.

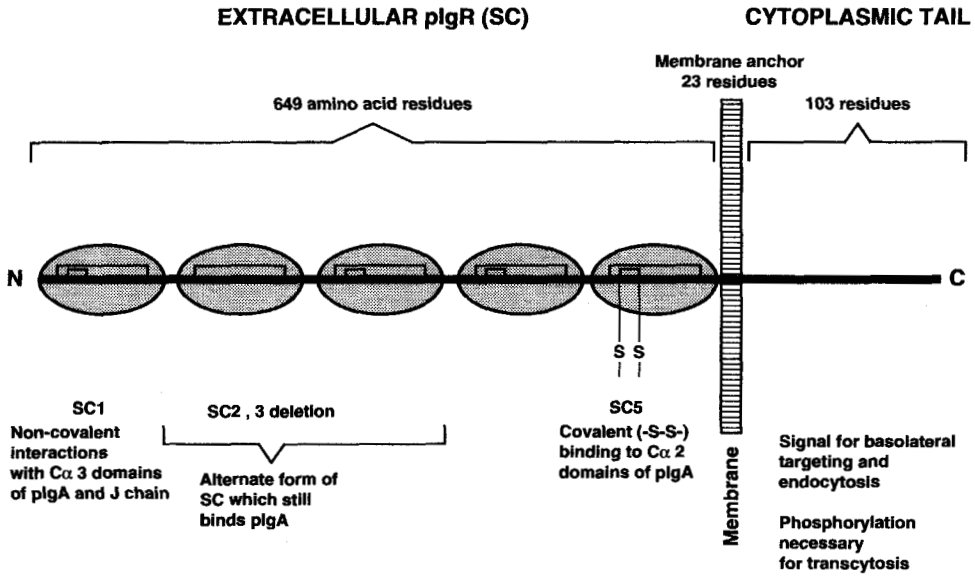
Secretory Component

Michael W. Russell and Jiri Mestecky

Synonyms	Secretory piece, transport piece, T-chain, T component, polymeric immunoglobulin receptor, poly-Ig receptor.
Abbreviations	SC, pIgR
Classifications	Membrane receptor, component chain of secretory immunoglobulins, electrophoretic mobility β.
Description	Secretory component (SC) is a glycoprotein found covalently associated with secretory immunoglobulins, S-IgA and S-IgM, in external secretions (tears, saliva, milk, urine, sweat, genital, gastrointestinal and respiratory fluids); synthesized by many types of mucosal epithelial and glandular cells (but not multilayered keratinized epithelia or human hepatocytes) in a precursor form as a transmembrane receptor expressed on basolateral surfaces (polymeric Ig receptor; pIgR); also found in Hassel's corpuscles of the thymus. SC, the extracellular part of pIgR, also occurs in free form in secretions; small amounts of SC are found in the circulation associated with IgM or polymeric IgA.
Structure	pIgR is a member of the Ig superfamily, containing 5 Ig-like extracellular domains corresponding to SC (residues 1 to approx. 559), a single transmembrane segment (residues 621–643), and a cytoplasmic segment (residues 644–746); Molecular modelling and mutagenesis studies of (rabbit) pIgR suggest that domain I possesses 3 exposed loops analogous to the complementarity-determining regions of Ig variable domains, which are involved in binding to polymeric IgA. SC is glycosylated with 5–7 complex N-linked oligosaccharides (22% carbohydrate by wt).
Molecular Weight	$\approx$100,000–120,000 (pIgR); 66,000–86,000 (SC)
Sedimentation Coeff.	4.3–5 S
Isoelectric Point	4.75–5.25
Extinction Coeff.	12.66 (280 nm,1%,1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	pIgR on the basolateral surface of epithelial cells binds J chain-containing polymeric IgA and pentameric IgM that are secreted by plasma cells in the lamina propria of mucosae or interstitium of exocrine glands. pIgR with bound Ig is translocated by vesicular transcytosis to the apical surface, during which disulfide bonds are formed between SC and Ig heavy chains, and the extracellular domains (SC) are proteolytically cleaved from the transmembrane segment of pIgR; secretory (S-) IgA or S-IgM is then released at the apical surface. pIgR lacking bound Ig is also transcytosed and free SC released at the apical surface. SC enhances resistance of secretory IgA to proteolysis; confers hydrophilic and electronegative

	charge properties on secretory Ig. SC has been found to be identical to gravidin, an inhibitor of phospholipase A_2 present in amniotic fluid. The cytoplasmic segment of pIgR is reported to bind calmodulin.
Physiology/Pathology	Expression is enhanced by certain cytokines and hormones. Circulating SC complexed with polymeric IgA or IgM is elevated in obstructive liver disease, hepatic necrosis, pregnancy, lactation, and breast carcinoma. Deficiency is extremely rare (2 cases reported).
Degradation	Circulating SC (bound to polymeric Ig) is rapidly removed from the circulation and catabolized mainly in the liver (most by hepatocytes, less by Kupffer cells). Free SC is highly susceptible to gastrointestinal proteases, but not when bound to polymeric Ig.
Genetics/Abnormalities	The gene encoding pIgR is located on chromosome 1, q31-q41. Two-allelic restriction-fragment-length polymorphism reported.
Half-life	Rapidly removed from the circulation.
Concentration	External secretions: bound SC as S-IgA <0.01 g/L (urine) to >12 g/L (colostrum); free SC 0.012 g/L (stimulated parotid saliva) to 2.09 g/L (colostrum), higher in cases of IgA deficiency. Normal serum: bound SC as S-IgA 0–0.034 g/L.
Isolation Method	Free SC can be isolated from secretions (best from IgA-deficient milk or colostrum) by ammonium sulfate precipitation followed by molecular sieve and ion-exchange chromatography, or block electrophoresis; or by affinity chromatography on a matrix containing polymeric IgA (myeloma) or IgM (Waldenström's macroglobulinemia) proteins.
Amino Acid Sequence	Primary structure of SC has been determined by peptide sequencing of proteolytic fragments (see DNA/AA references); confirmed by cDNA sequencing of pIgR.
Disulfides/SH-Groups	SC contains 20 Cys residues (9 intra-chain disulfide bonds; 1 or 2 disulfide bonds with heavy chain(s) of IgA in S-IgA).
General References	Heremans, J. F. Immunoglobulin A. **In** *The Antigens*, Volume II, Sela, M. (ed.) Academic Press, New York, 1974, pp. 365–522. Mestecky, J. and Kilian, M. Immunoglobulin A. *Methods Enzymol.* 1985, **116**: 37–75. Apodoca, G. et al. The polymeric immunoglobulin receptor. A model to study transcytosis. *J. Clin. Invest.* 1991, **87**: 1877–1882. Brandtzaeg, P. et al. Epithelial and hepatobiliary transport of polymeric immunoglobulins. **In** *Handbook of Mucosal Immunology*, Ogra, P. L. et al. (eds.) Academic Press, San Diego, 1994, pp. 113–126. Coyne, R. S. et al. Mutational analysis of polymeric immunoglobulin receptor/ligand interactions. Evidence for the involvement of multiple complementarity-determining region (CDR)-like loops in receptor domain I. *J. Biol. Chem.* 1994, **269**: 31620–31625.
Ref. for DNA/AA Sequences	Eiffert, H. et al. *Biol. Chem. Hoppe-Seyler* 1991, **372**: 119–128. Krajci, P. et al. *Eur. J. Immunol.* 1992, **22**: 2309–2315.

DOMAIN STRUCTURE OF SECRETORY COMPONENT



Structure of pIgR (SC). Each Ig-like domain (shaded elliptical area) consists of 100–110 amino acid residues, and contains one or two intra-chain disulfide bridges as shown. The probable functions are listed beneath each domain.

Semenogelin I

Anders Peter, Johan Malm and Hans Lilja

Synonyms	Sperm motility inhibitor precursor
Abbreviations	Sg I; SPMIP
Classifications	
Description	Semenogelin I occurs together with semenogelin II in disulfide-linked, high molecular mass aggregates in freshly ejaculated semen. These proteins constitute the major gel proteins in human semen. Semenogelin I is produced exclusively in the seminal vesicles. The mature secreted protein consists of 439 aa with a blocked N-terminal (pyroglutamate). Semenogelin I has a high degree of identity in primary structure to semenogelin II (~80%). Immunocytochemistry has shown that monoclonal anti-SgI/SgII IgGs (which crossreact with the two proteins) identify epitopes on the posterior head, midpiece and tail of spermatozoa, both from the epididymal cauda and after ejaculation. Semenogelin I and II have recently been shown to be substrates for transglutaminase. However, the physiological relevance of these findings is not clear.
Structure	Unknown
Molecular Weight	The calculated molecular weight of the mature protein is $49.6x10^3$. Measurements with mass spectometry gives a molecular weight of $49.9x10^3$, whereas SDS-PAGE with reduced proteins gives an apparent molecular weight of $52x10^3$.
Sedimentation Coeff.	Unknown
Isoelectric Point	>9.5 (theoretical value: 10.1)
Extinction Coefficient	5.5 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Semenogelin I interacts with semenogelin II to form the network giving human semen the gel-like appearance, which entraps the spermatozoa at ejaculation. The proteolytic digestion of these proteins by prostate specific antigen (PSA) into soluble fragments dissolves the gel, and progressively motile spermatozoa are released. Semenogelin I is now known to be identical to the sperm motility inhibitor precursor (SPMIP), which has been shown to inhibit sperm motility in a dose-dependent manner. The mechanism for this is unknown.
Physiology/Pathology	Deficiency of semenogelin I and II, due to aplasia of the seminal vesicles, results in defective gel formation. However, the physiological consequences of this deficiency are not clear.

Degradation	Semenogelin I is degraded by PSA within a few minutes after ejaculation. At least 13 fragments, which are more soluble than the intact protein, are formed. Cleavage by PSA mostly occurs carboxy-terminal of certain Tyr and Leu residues. It is not clear whether the fragments generated from the proteolysis by PSA have any biological function. *In vitro*, semenogelin I can be cleaved also by hK2, a kallikrein secreted by the prostate. Whether this has any significance *in vivo* is unclear.
Genetics/Abnormalities	The gene coding for semenogelin I, consisting of three exons, is located on the long arm of chromosome 20. The gene belongs to a family of genes designated REST (rapidly evolving substrates for transglutaminase). These genes have conserved 5´ and 3´-nontranslated regions, whereas the second exon undergoes a rapid evolution. Based on the sequence similarity between semenogelin I and II and the close distance between their genes on the chromosome, semenogelin I and II are believed to have evolved by gene duplication.
Half-life	See Degradation
Concentration	No method has been described for accurate estimation of the concentration of semenogelin I, but it is a very abundant protein in semen. Estimated values generally give concentrations higher or much higher than 10 mg/mL.
Isolation Method	To prevent the rapid degradation of semenogelin I, semen may be collected in a buffer containing 4 M urea, 30 mM DTT, 25 mM EDTA and 3 mM benzamidine. Heparin Sepharose can then be used to isolate the intact protein from the non-liquefied seminal plasma. Both semenogelin I and II are eluted with a salt gradient. Gel filtration may be used to further separate semenogelin I from contaminants.
Amino Acid Sequence	Semenogelin I consists of 439 aa, including six 60-aa repeats with extensive structural similarity. The protein is rich in Gln, Lys and Ser residues (13.2, 9.1 and 11.9%, respectively).
Disulfides/SH-Groups	The aa sequence deduced from cDNA contains one cystein residue.
General References	Bjartell, A. et al. *J. Androl.* 1996, **17**:17-26. Lilja, H. et al. *Scand. J. clin. Lab. Invest.* 1984, **44**:439-446. Malm, J. et al. *Eur. J. Biochem.* 1996, **238**:48-53. Robert, M. and Gagnon, *C. Biol. Reprod.* 1996, **55(4)**:813-821.
Ref. for DNA/AA Sequences	Lilja, H. et al. *J. Biol. Chem.* 1989, **264**:1894-1900. Lundwall, Å. and Lazure, C. *FEBS Lett.* 1995, **374**:53-56. Ulvsbäck, M. et al. *J. Biol. Chem.* 1992, **25**:18080-18084.

Semenogelin II

Anders Peter, Johan Malm and Hans Lilja

Synonyms	None
Abbreviations	Sg II
Classifications	
Description	Semenogelin II occurs together with semenogelin I in disulfide-linked, high molecular mass aggregates in freshly ejaculated semen. These proteins constitute the major gel proteins in human semen. Semenogelin II is produced mainly in the seminal vesicles, but also in the caudal region of the epididymis. The mature secreted protein consists of 559 aa with a blocked N-terminal (pyroglutamate). Semenogelin II has a high degree of identity in primary structure to semenogelin I (~80%). Immunocytochemistry has shown that monoclonal anti-SgI/SgII IgGs (which crossreact with the two proteins) identify epitopes on the posterior head, midpiece and tail of spermatozoa, both from the epididymal cauda and after ejaculation. Semenogeiin I and II have recently been shown to be substrates for transglutaminase. However, the physiological relevance of these findings is not clear.
Structure	Unknown
Molecular Weight	The calculated molecular weight of the mature protein is $62.9x10^3$. Measurements with mass spectrometry gives a molecular weight of $63.6x10^3$, whereas SDS-PAGE with reduced proteins gives an apparent molecular weight of approximately $70\text{-}76x10^3$ due to the occurrence of one glycosylated and one non-glycosylated form.
Sedimentation Coeff.	Unknown
Isoelectric Point	9.8
Extinction Coefficient	5.4 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Semenogelin II interacts with semenogelin I to form the network giving the human semen the gel-like appearance, which entraps the spermatozoa at ejaculation. The proteolytic digestion of these proteins by prostate specific antigen (PSA) into soluble fragments dissolves the gel, and progressively motile spermatozoa are released.
Physiology/Pathology	Deficiency of semenogelin I and II, due to aplasia of the seminal vesicles, results in defective gel-formation. However, the physiological consequences of this deficiency are not clear.
Degradation	Semenogelin II is degraded by PSA within a few minutes after ejaculation. At least 15 fragments, which are more soluble than the intact protein, are formed. Cleavage by PSA mostly occurs C-terminal of certain Tyr and Leu

	residues. It is not clear whether the fragments generated from the proteolysis by PSA have any biological function. *In vitro*, semenogelin II can be cleaved also by hK2, a kallikrein secreted by the prostate. Whether this has any significance *in vivo* is unclear.
Genetics/Abnormalities	The gene coding for semenogelin II, consisting of three exons, is located on the long arm of chromosome 20. The gene belongs to a family of genes designated REST (rapidly evolving substrates for transglutaminase). These genes have conserved 5´ and 3´-nontranslated regions, whereas the second exon undergoes a rapid evolution. Based on the sequence similarity between semenogelin I and II and the close distance between their genes on the chromosome, semenogelin I and II are believed to have evolved by gene duplication.
Half-life	See Degradation
Concentration	No method has been described for accurate estimation of the concentration of semenogelin II, but it is a very abundant protein in semen. Estimated values generally give concentrations higher or much higher than 10 mg/mL.
Isolation Method	To prevent the rapid degradation of semenogelin II, semen may be collected in a buffer containing 4 M urea, 30 mM DTT, 25 mM EDTA and 3 mM benzamidine. Heparin Sepharose can then be used to isolate the intact protein from the non-liquefied seminal plasma. Both semenogelin I and II are eluted with a salt gradient. Gel filtration may be used to further separate semenogelin II from contaminants.
Amino Acid Sequence	Semenogelin II consists of 559 aa, including eight 60-aa repeats with extensive structural similarity. A comparison with semenogelin I, which contains 120 aa less, reveals that semenoglin II is extended by two 60-aa repeats. The protein is rich in Gln, Lys and Ser (13.1, 10.0 and 13.2%, respectively). Unlike semenogelin I, semenogelin II can be glycosylated at Asn-249.
Disulfides/SH-Groups	The aa sequence deduced from cDNA contains two cystein residues, but no intra-molecular disulfide bond formation has been reported.
General References	Bjartell, A. et al. *J. Androl.* 1996, **17**:17-26. Kise, H. et al *Eur. J. Biochem.* 1996, **238**:88-96. Lilja, H. *Scand. J. Clin. Invest.* 1984, **44**:447-452. Malm, J. et al. *Eur. J. Biochem.* 1996, **238**:48-53.
Ref. for DNA/AA Sequences	Lilja, H. and Lundwall, Å. *Proc. Natl. Acad. Sci.* 1992, **88**:4559-4563. Lundwall, Å. and Lazure, C. *FEBS Lett.* 1995, **374**:53-56. Ulvsbäck, M. et al. *J. Biol. Chem.* 1992, **25**:18080-18084.

Serum amyloid A

Earl P. Benditt and Nils Eriksen

Synonyms	None
Abbreviations	SAA; apoSAA
Classifications	Apolipoprotein
Description	A family of plasma lipoproteins, associated mainly with the high-density fraction, HDL-3 in particular, and synthesized principally in the liver. Five isoforms with a high degree of sequence similarity have been described (SAA1α, -β, -γ and SAA2α, -β); a corresponding desArg form, lacking the N-terminal Arg residue, has been found for each of these isoforms. Variants differing only by substitution at a single aa position have been described for SAA1α and SAA1β and also for SAA2α and SAA2β. Another isoform, C-SAA, a constitutive apolipoprotein having structural features readily distinguishing it from the other isoforms, is known. Additional isoforms of SAA may exist.
Structure	Single-chain proteins, ≈ 100 residues, not yet obtained in crystalline form. The structure of SAA1α from an analysis of the aa sequence indicates the presence of two amphipathic helical domains and a putative calcium-binding loop in the molecule. The N-terminal sequence of the molecule is considered to be a lipid-binding transmembrane region, of interest because it could add lipid-binding properties to those already present in the molecule as hydrophobic faces of the amphipathic helices.
Molecular Weight	11,685 (apoSAA1α, from aa sequence)
Sedimentation Coeff.	Unknown
Isoelectric Point	SAA1α, -2α, -2β: pH 6.4, 7.5, 8.0; desArg: pH 6.0, 7.0, 7.4
Extinction Coeff.	26 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	At present unknown. The high degree of homology and the conservation of the protein among different vertebrate species argue for some important function, possibly in relation to responses to injury. Proposed functions include an autocrine-like induction of collagenase activity, inhibition of platelet aggregation, and participation in cholesterol transport, particularly a direct interaction between cholesterol and apoSAA resulting in modulation of cholesterol flux.
Physiology/Pathology	An acute-phase protein family. Plasma level rises in response to a cytokine-mediated increase in transcription of SAA genes. Overproduction may result in a condition characterized by the deposition, in tissues, of

	SAA and a truncated form called amyloid A (AA), which lacks the *circa* 3-kDa C-terminal portion of SAA.
Degradation	Normal degradative pathway unknown. Proteases of the elastase type may be involved in the degradation and, along with lysosomal cathepsin B, in the generation of amyloid fibrils from precursor SAAs.
Genetics/Abnormalities	Three human SAA genes, one pseudogene and several polymorphisms are known. The genes have been shown to be localized to a 90-kb region on the short arm of chromosome 11. This situation is similar to that found in the mouse, where 3 highly homologous genes and a pseudogene are present in a 79-kb region on chromosome 7; the latter has linkage groups similar to those found in human chromosome 11. The existence of a fith mouse SAA gene, encoding the recently described mouse SAA5 protein, is presumed.
Half-life	Man: unknown; rabbit, mouse: ≈ 1 hr (plasma).
Concentration	Plasma: 0.2 μM (normal range 0.05 - 1.0 μM). In an acute-phase response concentration may exceed 100 μM.
Isolation Method	Best source: acute-phase plasma high-density lipoproteins, particularly HDL-3, separated by centrifugation of density-adjusted plasma. Flotation fractions concentrated by ultrafiltration and subjected to size-exclusion chromatography in 6 M guanidine-HCl. Chromtographic fractions immunoreactive with anti-SAA or anti-AA further purified by ion-exchange and/or high-performance liquid chromatography.
Amino Acid Sequence	No physiologically important parts known. The sequence DKYFHAR GNYDAA (residues 33 to 45, human numbering system) occurs without substitution, except in human C-SAA and mouse SAA5, in the SAA or AA of all twelve species studied. This constancy suggests a functionally important role for the sequence. Also noteworthy is a generally highly conserved 6-residue sequence invariably flanked, again with the exception of C-SAA and mouse SAA5, by methionine residues (positions 17 and 24) in all twelve species studied. An interesting situation prevails in human SAA at positions 52 and 57, where 1α has Val/Ala, 1β, 2α and 2β have Ala/Val, 1γ has Ala/Ala, C-SAA has Val/Leu, and the predicted gene product of pseudogene SAA3 would have Val/Val.
Disulfides/SH-Groups	None
General References	Eriksen, N. and Benditt, E.P. *Methods Enzymol.* 1986, **128**:311-320. Turnell, W.G. et al. *Mol. Biol. Med.* 1986. **3**:387-407; 409-424. Strachan, A.F. et al. *J. Biol. Chem.* 1989, **264**:18368-18373. Malle, E. et al. *Atherosclerosis* 1993, **102**:131-146. Marhaug, G. and Dowton, S.B. *Baillière's Clin. Rheumatol.* 1994, **8**:553-573. Banka, C.L. et al. *J. Lipid Res.* 1995, **36**:1058-1065. Uhlar, C.M. et al. *Scand. J. Immunol.* 1996, **43**:271-276.
Ref. for DNA/AA Sequences	Kluve-Beckerman, B. et al. *J. Clin. Invest.* 1988, **82**:1670-1675. Sack, G.H. Jr. et al. *Scand. J. Immunol.* 1989, **29**:113-119. Steinkasserer, A. et al. *Biochem. J.* 1990, **268**:187-193 Betts, J.C. et al. *Scand. J. Immunol.* 1991, **34**:471-482. Sellar, G.C., Uhlar, C.M. et al. *Genomics* 1994, **19**:221-227;228-235. Steel, D.M. and Whitehead, A.S. *Immunol. Today* 1994, **15**:81-88.

Serum amyloid P component

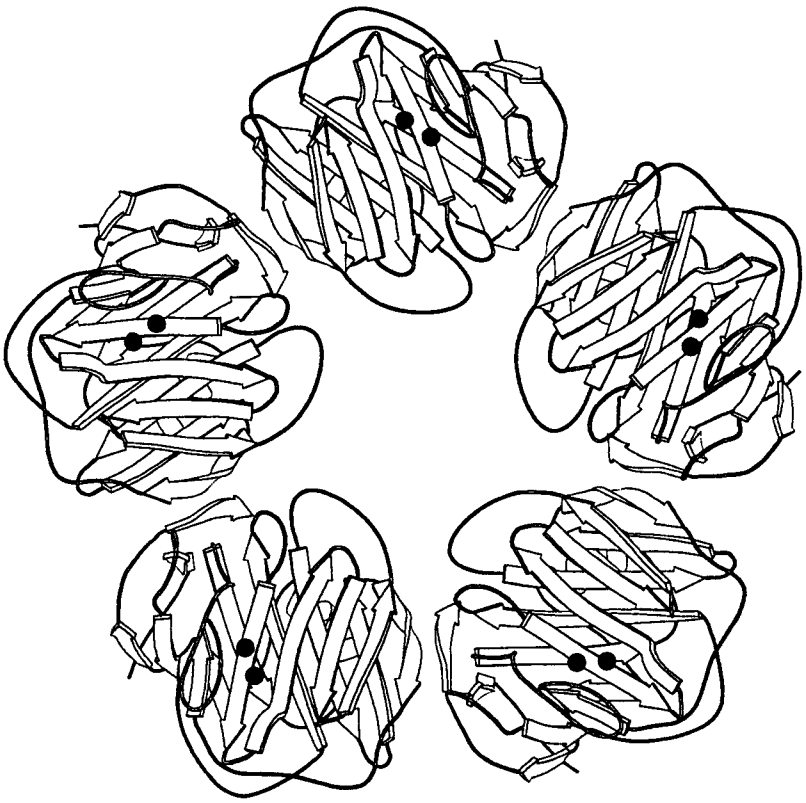
Mark B. Pepys

Synonyms	P component; 9.5S α_1-glycoprotein
Abbreviations	SAP
Classifications	None
Description	SAP is a circulating plasma glycoprotein and a member of the pentraxin family that includes C-reactive protein. It is synthesized exclusively by hepatocytes. An antigenically identical protein, the site of synthesis of which is not known, is a normal tissue constituent found in the lamina rara interna of the glomerular basement membrane, where it is covalently associated with collagen and/or other matrix glycoproteins, and in non-covalent association with the microfibrils of elastic fibres throughout the body. SAP has the capacity for Ca^{2+}-dependent ligand binding, by virtue of which it binds to amyloid fibrils in the pathological condition of amyloidosis, and thereby accumulates in all known forms of amyloid deposit. In this situation it is called amyloid P component (AP) but this is structurally identical to circulating SAP.
Structure	SAP is composed of identical non-covalently associated subunits arranged in a disc with cyclic pentameric symmetry. In the absence of calcium isolated pure SAP adopts a decameric form composed of two pentameric discs interacting face-to-face, but in whole serum it is probably a single pentamer. Pure SAP rapidly precipitates in the presence of calcium. Wide angle solution scattering curves from decameric SAP in EDTA-containing buffer, measured both with X-rays and neutrons, give a ring diameter of 7.4 nm. In the crystal structure of pentameic SAP at 2Å each subunit consists of a flattened β-jelly roll, a single α-helix; the loops joining the β-strands are short and compact to the body of the molecule. The tertiary fold, which is shared by human CRP, is also found in a number of plant, animal and bacterial lectins and other carbohydrate-binding proteins, now recognized to be members of the "lectin fold" superfamily. There is a single typical complex biantennary oligosaccharide attached to residue Asn-32 in each SAP protomer, and its structure is essentially the same regardless of the source of SAP; indeed it is one of the most invariant oligosaccharides of any known glycoprotein.
Molecular Weight	254,620; each subunit being 25,462, as calculated from the complete covalent structure of both protein and glycan, and confirmed directly by electrospray mass spectrometry
Sedimentation Coeff.	9.5S (for decameric form)
Isoelectric Point	≈ 5.2 (native state), 6.0 (8 M urea)
Extinction Coeff.	17.1 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None

1. Installment 1992 (revised 1997)

Inhibitors	Methyl 4,6-O-(1-carboxyethylidene)-β-D-galactopyranoside and closely related sugars inhibit all the known calcium-dependent ligand binding interactions of SAP, presumably by competing for the ligand binding sites on the protein.
Biological Functions	Many properties and interactions of SAP have been described, but its biological function is not known. In the presence of physiological calcium concentrations, it is remarkably resistant to proteolysis. The stable evolutionary conservation of its lectin-like calcium-dependent binding activity for specific sugars, for certain polyanions and for particular proteins suggests that this property is likely to be significant, although it is not clear precisely what effects it may have in vivo. Mice with targeted delition of the SAP gene and therefore no SAP, develop apparently normally and are fertile, but subtle phenotypic abnormalities have not yet been investigated.
Physiology/Pathology	No deficiency and no polymorphism of either SAP protein or glycan have been reported in man suggesting an important physiological role, even though absence of SAP has so far not caused overt problems in SAP knockout mice. The plasma concentration is tightly regulated and is normal even when there is a huge extravascular SAP pool in systemic amyloidosis. Subnormal levels are seen only with hepatocellular failure. Raised levels have been reported in patients with macroglobulinaemia and in end stage renal failure. Although aggregated SAP exposed to whole serum binds selectively to fibronectin and C4-binding protein, native SAP in the plasma is not complexed with these or any other macromolecular ligands. Aggregated SAP can also bind to fixed C3 and to cell membranes but it is not clear whether this is of physiological significance. Native SAP binds to heparin, heparan and dermatan sulfates and it is also the single plasma protein which undergoes specific calcium-dependent binding to DNA and to chromatin in vitro and in vivo. In binding to chromatin, SAP selectively displaces H1-type histones and solubilises long chromatin which is otherwise highly insoluble under physiological conditions. SAP may thus participate in the handling in vivo of chromatin released from cells. SAP undergoes calcium-dependent binding to all types of amyloid fibrils and is present in all amyloid deposits. Binding of SAP to ex vivo amyloid fibrils or to synthetic fibrils, protects them from proteolysis in vitro. Deposition of experimentally induced AA amyloidosis is delayed in SAP knockout mice compared to SAP-sufficient controls, indicating that SAP contributes significantly to the pathogenesis of amyloidosis. SAP is also a normal connective tissue constituent, present from early childhood onwards in the glomerular basement membrane and in the microfibrillar mantle of elastic fibres throughout the body, and may protect against proteolysis.
Degradation	Catabolised by hepatocytes and can be taken up via the asialoglycoprotein receptor. No particular breakdown products recognized.
Genetics/Abnormalities	There is a single copy of the SAP gene, located on chromosome 1 (1q12- > q23), and there is no known polymorphism. No variant or abnormal forms of SAP have yet been described.
Half-life	24.5 hrs (plasma), fract. cat. Rate 68%
Concentration	The serum concentration of SAP in 500 healthy adult subjects (274 women, 226 men) measured by electroimmunoassay calibrated with standards of highly purified isolated SAP was mean (SD, range), as follows: women 24 mg/L (8, 8 - 55); men 32 mg/L (7, 12 - 50) ($P < 0.001$).

Isolation Method	The best methods are based on calcium-dependent affinity chromatography of serum on either pyruvate-rich plain agarose beads or phosphoethanolamine immobilized on agarose beads. Further purification by reversed affinity chromatography to remove traces of albumin and immunoglobulin, followed by gel filtration, yields isolated SAP.
Amino Acid Sequence	SAP shows 52% sequence identity with C-reactive protein (CRP), the other human pentraxin and high levels of homology with its SAP counterparts in other vertebrate species.
Disulfides/SH-Groups	There is a single intrachain disulfide bond in each SAP subunit, between the cysteine residues at positions 36 and 95.
General References	Pepys, M.B. et al. *Adv. Immunol.* 1983, **34**:141-212. Kushner, I. et al. (ed.), C-reactive protein and the plasma protein response to tissue injury. *Ann. NY Acad. Sci.*, **Vol. 389,** 1982. Pepys, M.B. Amyloidosis. In: *Samter's Immunological Diseases,* Frank, M.M. et al. (eds.), Fifth Edition, Boston: Little Brown & Co. 1994; pp. 637-655. Hawkins, P.N. et al. *N. Engl. J. Med.* 1990, **323**:508-513. Hawkins, P.N. et al. *J. Clin. Invest.* 1990, **86**:1862-1869. Butler, P.J.G. et al. *J. Exp. Med.* 1990, **172**:13-18. Emsley, J. et al. *Nature* 1994, **367**:338-345. Pepys, M.B. et al. *Proc. Natl. Acad. Sci. USA* 1994, **91**:5602-5606. Hutchinson, W.L. et al. *J. Clin. Invest.* 1994, **94**:1390-1396. Tennent, G.A. et al. *Proc. Natl. Acad. Sci. USA* 1995, **92**: 4299-4303. Srinivasan, N. et al. *Chemtracts-Biochem. Mol. Biol.* 1996, **6**:149-164. Botto, M. et al. Submitted for publication.
Ref. for DNA/AA Sequences	Prelli, F. et al. *J. Biol. Chem.* 1985, **260**:12895-12898. Mantzouranis, E.C. et al. *J. Biol. Chem.* 1985, 280:7752-7756. Ohnishi, S. et al. *J. Biochem.* 1986, **100**:849-858.
	PDB Entry Code 1 SAC



Cartoon representation of the SAP pentamer structure viewed down the five-fold symmetry axis. The SAP subunit is composed of a sandwich of two large β-sheets with a single α-helix on one face and the double calcium ligand binding site on the other face. The calcium ions are shown as black spheres in this figure. Made with MOLSCRIPT (Kraulis, P.J. *J. Appl. Cryst.* 1991, **24**:946-950.

Somatotropin

Gerhard Baumann

Synonyms	Growth Hormone; Somatotrophin; Somatotropic hormone
Abbreviations	GH; hGH; STH
Classifications	Polypeptide hormone
Description	A protein synthesized and secreted by the pituitary gland (and by the placenta). Several molecular variants exist. The principal pituitary form (22 kDa form) is a 191 aa, single chain, simple protein (mw approx. 22 kDa). A shorter variant (20 kDa form) has an internal deletion (aa 32-46); it comprises 5-10% of the pituitary GH and arises from the same gene as the 22 kDa form (hGH-N gene) by alternative mRNA splicing. The placental form is also a 191 aa protein which differs from the pituitary form in 13 aa positions. This form is the product of a separate gene (hGH-V gene). A glycosylated (presumably at Asn-140) form of placental GH also exists. Several posttranslational modifications of pituitary GH result in a family of proteins (N_{α}-acylated, deamidated and a number of oligomeric forms).
Structure	A globular protein with four antiparallel α-helices, packed in a twisted bundle, and two mini-helices.
Molecular Weight	22,124 for pituitary GH, 20,269 for the 20 kDa variant, and 22,320 for placental GH (all from aa composition). Occurs physiologically as an oligomeric series up to at least pentameric (100-110,000 kDa). Glycosylated placental GH has an approximate mw of 25,000 (from SDS-PAGE).
Sedimentation Coeff.	2.55 S
Isoelectric Point	5.15 (22 kDa form); 5.5 (20 kDa variant)
Extinction Coeff.	22 kDa form: 6.8-8.2 (280nm, 1%, 1cm) or 1.5-1.8 x 10^4 (molar). 20 kDa form: 5.4-6.8 (280nm, 1%, 1cm) or 1.1-1.4 x 10^4 (molar).
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	An anabolic hormone essential for postnatal growth and development. Exerts many of its effects through stimulation of insulin-like growth factor I (IGF-I) production in target tissues. IGF-I in turn acts as a local mitogen in a paracrine/autocrine manner. GH also acts directly as a differentiation/cellular commitment factor. In addition, it has a number of metabolic activities, including effects on cellular aa and glucose uptake, renal phosphate and sodium retention, and insulin-antagonizing properties.
Physiology/Pathology	Secreted from the pituitary in a pulsatile pattern, at approximately 2-3 hour intervals. A major secretory surge occurs during early sleep. Vigorous

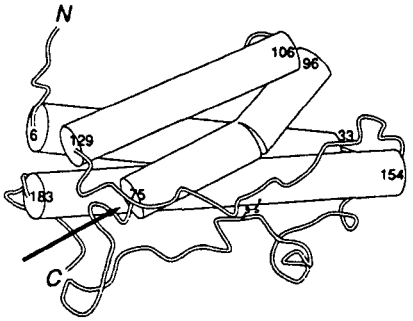
	exercise and hypoglycemia also stimulate GH secretion. GH secretion is high and unregulated in the fetus, falls postnatally to the pattern described, is amplified during puberty, and progressively declines after age 25. Placental GH supplants pituitary GH in maternal blood during the second half of pregnancy. Lack or underproduction of GH leads to growth retardation and a tendency to hypoglycemia. In adults, lack of GH leads to altered body composition (increased fat, [especially visceral fat], decreased lean body mass, and possibly bone loss). Overproduction (pituitary tumor or rarely hyperplasia) produces excessive growth (acromegaly or gigantism) and is accompanied by insulin resistance with or without frank diabetes, sodium retention and hypertension, bony overgrowth, arthritis, and several other complications resulting from exposure to excessive amounts of GH and IGF-I.
Degradation	Eliminated from circulation primarily by the kidney (glomerular filtration and degradation in proximal renal tubule). Secondary degradation is in liver (and probably other organs) by receptor-mediated cell entry and intracellular degradation. Little or no proteolysis occurs in the circulation.
Genetics/Abnormalities	The GH gene cluster (including hGH-N, hGH-V and placental lactogen genes) resides on chromosome 17. Deletions of the hGH-N gene result in complete absence of GH and severe growth retardation. Deletion of the hGH-V gene produces no clinical abnormality. Structurally abnormal GHs have not yet been identified.
Half-life	15-25 min. (blood circulation). The half-life is a composite of the half-lives of free GH ($t_{½}$ 3-5 min) and GH bound to the GH binding protein ($t_{½}$ ~ 30 min).
Concentration	Plasma: ranges from 20 ng L^{-1} to 70 µg L^{-1} (1 pM-3 nM) (wide concentration range due to pulsatile pituitary secretion and short plasma half life). Pituitary gland: 5 g L^{-1} or greater.
Isolation Method	Isolated from human pituitary glands by neutral saline extraction, ammonium sulfate precipitation, and sequential gel filtration and ion exchange chromatography steps. Recombinant hGH is commercially available.
Amino Acid Sequence	Two conformational epitopes for receptor binding (Sites 1 & 2) exist and have been mapped in great detail. Binding of two receptors occurs sequentially, with Site 1 (primarily W104, W169, I103, I105, P106, I165, R43, E44, D126, E127, D164) binding one receptor molecule first, followed by recruitment of another receptor molecule to Site 2 (primarily composed of 14, F1, D116, E119, R8). This GH-induced receptor dimerization is important for signal transduction and biological function. 85% sequence homology with placental lactogen/chorionic somatomammotropin; 35% homology with prolactin. Internal sequence (residues 32 - 46) deleted in 20 kDa variant.
Disulfides/SH-Groups	2 intra-chain disulfides (53-165 and 182-189), no free sulfhydryls. Oligomers linked by inter-molecular disulfide bridges exist.
General References	Baumann, G. Growth hormone heterogeneity: Genes, isohormones, variants and binding proteins. *Endocr. Rev.* 1991, **12**:424-449. Cunningham, B.C., Ultsch, M., DeVos, A.M. et al. Dimerization of the extracellular domain of the human growth hormone receptor by a single hormone molecule. *Science* 1991, **254**:821-825.

DeVos, A.M. Ultsch, M., Kossiakoff, A.A. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* 1992, **255**:306-312.
Ultsch, M.H. Somers, W., Kossiakoff, A.A., DeVos, A.M. The crystal structure of affinity-matured human growth hormone a 2 A resolution. *J. Mol. Biol.* 1994, **236**:286-299.
Clackson, T., Wells, J.A. A hot spot of binding energy in a hormone-receptor interface. *Science* 1995, **267**:383-386.
Parks, J.S. Molecular Biology of Growth Hormone. *Acta Paediatr. Scand.* 1989, **349** (Suppl) 127-135.
Isaksson, O.G. et al. Mechanisms of the Stimulatory Effect of Growth Hormone on Longitudinal Bone Growth. *Endocr. Rev.* 1987, **8**:426-438.

Ref. for DNA/AA Sequences

Niall, H.D. *Nature N. Biol.* 1971, **230**:90-91.
Li, C.H. and Dixon, J.S. *Arch. Biochem. Biophys.* 1971, **146**:233-236.
Martial, J.A. et al. *Science* 1979, **205**:602-607.
Roskam, W.J. and Rougeon, F. *Nucleic Acids Res.* 1979, **7**:305-320.
Goeddel, D.V. et al. *Nature* 1979, **281**:544-548.
GenBank accession codes J00148 K00612.

Molecular model of hGH based on the tertiary structure of porcine GH according to Cunningham et al. (*Science* 1989; 243: 1330–1336)



Somatotropin Receptor & Somatotropin Binding Protein

Gerhard Baumann

Synonyms	Growth Hormone Receptor & Growth Hormone Binding Protein
Abbreviations	GHR, hGHR, GHBP, hGHBP
Classifications	Membrane-anchored polypeptide hormone receptor & soluble extracellular domain of same.
Description	A receptor which binds growth hormone (GH), expressed ubiquitously, though its abundance differs among tissues. Exists as a plasma membrane-anchored receptor (GHR) and as a soluble ectodomain of the GHR in the form of a circulating GH binding protein (GHBP). The GHR is a single chain, single transmembrane domain, heavily glycosylated protein. Two molecular variants are known: a full-length, 620 aa polypeptide and a splicing variant with an internal 22 aa deletion (encoded by exon 3) near the amino terminus. The extracellular domain contains 246 aa, the transmembrane domain 24 aa, and the intracellular domain 350 aa. The GHR is a member of the cytokine/ hematopoietic growth factor superfamily. It shares structural homology with the other members of the superfamily in the extracellular domain, but little similarity in the intracellular domain. Important features are an extracellular YGEFS motif (homologous to the WSxWS motif in other members of the cytokine receptor family) and a proline-rich region ("Box 1") in the intracellular part near the transmembrane domain. The latter region interacts with the tyrosine kinase JAK2 which initiates the phosphorylation cascade responsible for signalling. In humans, the GHBP is proteolytically cleaved from the GHR near its transmembrane portion by an unknown protease. The exact site of cleavage is not known.
Structure	Crystal structure only available for the (non-glycosylated) extracellular portion (i.e., the GHBP). A compact protein with two domains, each composed of two antiparallel β-sheets of four and three strands, respectively. The amino-terminal domain contains 3 small disulfide loops and is involved in GH binding; the carboxy-terminal domain makes contact with a second GHBP to form a dimeric complex. GHR dimerization is induced by ligand binding, with two GHR molecules binding to two separate sites on GH. The GHR dimer is then further stabilized by receptor-receptor interaction in the carboxyterminal part of the extracellular domain.
Molecular Weight	$\approx$120 kDa: GHR (69.6 kDa contributed by protein backbone). Exon 3 deleted variant 67.3 kDa for the protein backbone; glycosylation state or native form not yet known. $\approx$60 kDa: GHBP (protein backbone 28.4 kDa); exon 3 deleted form (protein backbone 26.1 kDa) not yet demonstrated.
Sedimentation Coeff.	Unknown
Isoelectric Point	5.0 for GHBP when complexed with GH
Extinction Coeff.	Unknown for native GHBP
Enzyme Activity	None intrinsic. Associates with JAK2 tyrosine kinase as the initial step of the signalling pathway.
Coenzymes/Cofactors	None

Substrates	None
Inhibitors	None
Biological Functions	Receptor: Recognizes and binds GH on cell surfaces, transmits hormone signal intracellularly, thereby initiating GH action. Ligand-induced GHR dimerization is crucial for signalling and biological action. GHBP: Provides a circulating GH reservoir by complexing GH in the circulation and in interstitial fluid. Prolongs the biological persistence of GH by delaying its clearance. Modulates GH binding to receptors and hence GH action by competing for ligand.
Physiology/Pathology	In man, GHR levels in tissues are difficult to measure because of a) relatively low abundance in most tissues and b) inaccessibility. Therefore, most physiological data are based on GHBP measurements in plasma, which are assumed to reflect GHR status. This is a reasonable assumption based on correlative data between GHBP and GHR in animals. The organ with the highest abundance of GHRs is the liver. It is probably also the principal source of the GHBP. The GHR is expressed in virtually all tissues, and hence the GHBP is probably ubiquitously produced. Plasma GHBP levels (and probably GHR levels) are low in the fetus and neonate, undergo a developmental upregulation during childhood, and stay constant during adult life. There is no diurnal variation. Women have slightly higher levels than men. GHR and GHBP are absent or dysfunctional in Laron dwarfism (GH resistance syndrome) because of mutations in the GHR gene. GHBP is also low in a variety of congenital and acquired conditions associated with GH resistance, probably due to altered regulation of gene expression. Examples are pygmy dwarfism, malnutrition, type I diabetes, liver cirrhosis, renal failure and hypothyroidism. Obesity, a condition with enhanced GH action, is associated with high GHBP levels. It is assumed that GHBP is positively correlated with GH responsivity by a) reflecting GHR levels, and b) its effect on GH bioavailability through prolongation of GH half-life. At pathologically high GHBP levels, the inhibitory effect on GH action may become dominant through sequestration of GH in the circulation.
Degradation	GH-GHR complexes are internalized into cells via clathrin-coated pits, with subsequent degradation of the majority of receptor (and ligand) in lysosomes. A minor fraction of the receptor population is recycled to the plasma membrane. Little is known about the degradative pathways for GHBP.
Genetics/Abnormalities	Single copy gene with 10 exons spanning ~87 kb on chromosome 5p13-p12. Loss or dysfunction of both alleles results in complete GH resistance and severe growth abnormalities (Laron dwarfism). Two partial deletions and 19 point mutations (nonsense and missense mutations) have been described, all in the extracellular domain.
Half-life	GHR: Residence half-time in plasma membrane ranges from 30 min to several hours, depending on cell type. GHBP: Plasma half-life 17 min for recombinant, non-glycosylated GHBP (in the guinea pig). Not yet determined in man. The plasma half-life of the GH-GHBP *complex* in man is about 30 min.
Concentration	GHBP in plasma: 0.1–2 nM (wide individual variability and developmental variation). GHR in tissues: <500 to ~30,000 receptors per cell, depending on cell type.
Isolation Method	Isolated from solubilized liver membranes (GHR) or from plasma (GHBP) by affinity chromatography on GH columns, followed by gel filtration.

Amino Acid Sequence	Biologically important regions involve residues 42–44, 101–106, 125–132, 164–171 and 218 for GH binding, 143–154 and 200–201 for dimer stabilization, and 280–286 for JAK2 binding/signalling. GHR has 30% overall homology with the prolactin receptor, but certain regions of the extracellular domain are 70% homologous. Lesser degrees of homology with the other members of the cytokine receptor superfamily.
Disulfides/SH-Groups	Extracellular domain: 3 intra-chain, in-sequence disulfide loops (38–48, 83–94, and 108–122). Free cysteine at position 241. Intracellular domain: 9 cysteine residues.
General References	Kelly, P. A. et al. The prolactin/growth hormone receptor family. *Endocrine Rev.* 1991, **12**: 235–251. Bass, S. H. et al. A systematic mutational analysis of hormone-binding determinants in the human growth hormone receptor. *Proc. Natl. Acad. Sci. USA* 1991, **88**: 4498–1502. De Vos et al. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* 1992, **255**: 306–312. Mercado, M. et al. Distribution of growth hormone receptor messenger ribonucleic acid containing and lacking exon 3 in human tissues. *J. Clin. Endocrinol. Metab.* 1994, **78**: 731–735. Roupas, P. and Herington, A. C. Cellular mechanisms in the processing of growth hormone and its receptor. *Mol. Cell Endocrinol.* 1989, **61**: 1–12. Baumann, G. et al. A specific growth hormone-binding protein in human plasma: Initial characterization. *J. Clin. Endocrinol. Metab.* 1986, **62**: 134–141. Baumann, G. Growth hormone binding protein: State of the Art. *J. Endocrinol.* 1994, **141**: 1–6. Amselem, S. et al. Spectrum of growth hormone receptor mutations and associated haplotypes in Laron syndrome. *Human Mol. Genetics* 1993, **2**: 355–359. Berg, M. A. et al. Diverse growth hormone receptor gene mutations in Laron syndrome. *Am. J. Hum. Genet.* 1993, **52**: 998–1005. Argetsinger, L. S. et al. Identification of JAK2 as a growth hormone receptor-associated tyrosine kinase. *Cell* 1993, **74**: 237–244.
Ref. for DNA/AA Sequences:	Leung, D. W. et al. *Nature* 1987, **330**: 537–543. Godowski, P. J. et al. Characterization of the human growth hormone receptor gene and demonstration of a partial gene deletion in two patients with Laron-type dwarfism. *Proc. Natl. Acad. Sci. USA* 1989, **86**: 8083–8087. GenBank accession code X06562

Spectrin

Jean Delaunay

Synonyms	None
Abbreviations	None
Classifications	Actin-binding protein
Description	Spectrin is the most abundant protein of the red cell membrane skeleton. It refers to a $\alpha\beta$ dimer. The α- and β-chains are antiparallel. Dimers self-associate at one extremity (α-chain N-terminal region, β-chain C-terminal region), constituting the oligomerization site, to form tetramers and higher order oligomers. At the other extremity, the dimer, through its β-chain, interacts with actin and protein 4.1. Near this extremity, the nucleation site delineates the region where a α- and a β-chain initiate assemblage into a dimer. Erythroid spectrin may occur in nonerythroid tissues, with possible alterations of the chains: in skeletal muscle, the β-chain has an extended C-terminal segment (due to alternative splicing). Nonerythroid spectrin (fodrin) has a similar structure, yet its chains arise from distinct genes. Alpha-actinin and dystrophin are more remote proteins. All proteins have a recurrent organization and bind actin.
Structure	The extended spectrin dimer is 100 nm long. The α-chain (2429 residues) and the β-chain (2137 residues) are comprised of repeating segments, numbered $\alpha 1 - \alpha 22$ and $\beta 1 - \beta 17$ (N$\rightarrow$C dirction), respectively, and containing about 106 residues. Position 45 is invariant (Trp). In space, the repeating segments fold into conformational units in such a way that residues 26 of the former become residues 1 of the latter. Typically, a conformational unit contains a β-strand, a α-helix (helix 1), a omega loop, a α-helix (helix 2), antiparallel to helix 1, a reverse turn and a α-helix (helix 3). Helix 3 is parallel to helix 1 but belongs, linearly, to the following repeating segment. Some repeating segments are atypical ($\alpha 10$, $\alpha 22$, $\beta 15$). Repeating segment $\alpha 10$ harbours a *src* homology 3 sequence. The C-terminal nonrepeat portion of α-chain contains sequences homologous to Ca^{2+}-binding EF hands structures. Repeating segment $\beta 15$ represents the binding site for ankyrin. In the β-chain, the 17-repeat stretch is preceded by a nonrepeat segment (272 residues), that carries the binding site(s) for actin and protein 4.1, and followed by another nonrepeat segment (52 residues), that bears phosphorylation sites and participates in the oligomerization process. The role of phosphorylation is unknown.
Molecular Weight	280 kDa (α-chain); 246 kDa (β-chain)
Sedimentation Coeff.	12.5 S (tetramer); 9.5 S(dimer);4.4 S (monomer)
Isoelectric Point	5.5 (α-chain); β-chain unknown
Extinction Coeff.	10.7 (280 nm, 1%,1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None

Biological Functions	Spectrin constitutes a submembrane network. It provides the red cell with its mechanical properties. It also stabilizes the membrane lipid bilyayer. The meshwork contains two types of crossroads: the oligomerization site, at which dimers self-associate into tetramers, hexamers (Y-shaped structures) and octamers (X-shaped structures), and the junctional complex, at which spectrin (through its β-chain) interacts with actin and protein 4.1 and which contains many other proteins (adducin, dematin, tropomyosin,...). The spectrin network is also connected, yet indirectly, to transmembrane components. Ankyrin mediates the attachment of spectrin (through its β-chain) to the cytoplasmic domain of the anion exchanger. The junctional complex is linked to several transmembrane proteins, including glycophorin C. In *situ*, the spectrin dimer would be twisted, having a length of 29–37 nm only.
Physiology/Pathology	Spectrin accounts for red cell resistance, elasticity and shape. Abnormalities result in erythrocyte fragility (hyperhemolysis, anemia) associated with morphological alterations (spherocytosis, elliptocytosis, poikilocytosis). Reduction in spectrin, due to defective attachment to ankyrin, causes a loss in membrane bilayer and yields spherocytosis. Mutations in or near the oligomerization site generates a defect in the oligomerization process, and yields elliptocytosis and/or poikilocytosis.
Degradation	During erythropoiesis, the β-chain and even more the α-chain of spectrin are synthesized in large excess. The chains that do not undergo incorporation in the skeleton are bound for rapid degradation. On the contrary, duely incorporated oligomers are stable, especially since the mature erythrocyte no longer performs protein synthesis.
Genetics/Abnormalities	Spectrin α- and β-genes map to 1q22-q23 and 14q23-q24.2, respectively. Alpha and β-spectrin mRNAs are $\simeq$ 8 and 7.5 kb in size, respectively. Both exist as one major splice form (in the erythroid cells). Polymorphisms exist in the α-chain at positions 701 (R/H), 809 (I/V), 853 (T/R) and 1857 (L/V). Intronic polymorphisms are known in the α-gene (*Xba*I, *PvuI*, *Msp*I) and in the β-gene (*Hind*III, *Pvu*II, *Pst*I, *Stu*I, *Taq*I). Low expression allele α^{LELY}, that is very common, contains valine (instead of leucine) at position 1857, as mentioned, T (instead of C) at position – 12 of intron 45, and is associated with partial skipping of exon 46. About 20 and 10 pathogenic mutations are known in spectrin α- and β-chains, respectively. They stand in or near the oligomerization site (repeating segments β17, α1 – α5, α8 and α9), and as such cause elliptocytosis and/or poikilocytosis, depending on the mutation or the association of two mutations in *trans* to one another. Allele α^{LELY} sensitizes, sometimes dramatically, the effects of α-spectrin mutations occuring in *trans*.
Half-life	$\simeq$ 120 days (as the red cell)
Concentration	20 to 30% of total red cell membrane protein.
Isolation Method	Ghosts are prepared from fresh erythrocytes by osmotic lysis using dilute phosphate buffer (5 mMole/L, pH 8.0; solution : RBC 20 : 1). Hydrosoluble proteins are prepared from fresh or frozen ghosts by incubation (37 °C, 30 min.) in a buffer containing: 0.3 Mole/L phosphate buffer, pH 8.0, 0.1 mMole/L EDTA, 0.1 mMole/L PMSF and 0.1 mMole/L β-mercaptoethanol. Following centrifugation (20,000 rpm, 45 min.), pure spectrin may be obtained from the supernatant by gel filtration.
Amino Acid Sequence	Repeating segment α1 (α-chain): IQERRQEVLT RYQSFKERVA ERGQKLEDSY HLQVFKRDAD DLGKWIMEKV NILTDKSY – EDPTNIQGKY QKHQSLEAEV QTKSRLMSEL EKTREERFTM GHSAHEE

Repeating segment $\beta 17$ (β-chain):
IREKLQQVMS RRKEMNEKWE ARWERLRMLL EVCQFSRDAS VAEAWLIAQE PYLASGDFGH TVDSVEKLIK RHEAFEKSTA SWAERFAALE KPTTLELKER QIAERP

Disulfides/SH-Groups

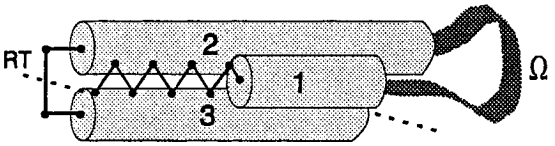
None

General References

Delaunay, J. and Dhermy, D. In: *Cellular and Molecular Biology of the Red Cell Membrane Proteins and its Disorders. II. Their Genes and Mutations.*
Palek, J. (ed.) *Semin. Hematol.* 1993, **30:** 21–33.
Dodge, J. T et al. *Arch. Biochem. Biophys.* 1963, **100:** 119–130.
Hsu, C. J. et al. *J. Supramol. Struct.* 1979, **10:** 227–239.
Kam, Z. et al. *Biochemistry* 1977, **16:** 5568–5572.
Liu, S. C. et al. *J. Clin. Invest.* 1981, **68:** 597–605.
Marchesi, V. T. and Steers, E. Jr. *Science* 1968, **159:** 203–204.
Speicher, D. W. et al. *J. Biol. Chem.* 1993, **268:** 4227–4235.
Speicher, D. W. et al. *J. Biol. Chem.* 1982, **257:** 9093–9101.

Ref. for DNA/AA Sequences

Sahr, K. E. et al. *J. Biol. Chem.* 1990, **265:** 4434–4443.
Winkelmann, J. C. et al. *J. Biol. Chem.* 1990, **265:** 11827–11832.



Three dimensional structure of a conformational unit. Alpha-helical segments 1, 2 and 3 are indicated (Speicher et al. *J. Biol. Chem.* 1993, 268: 4227–4235).
~: β-strand. Ω: omega loop. RT: reverse turn.

Sucrase-isomaltase

Erwin E. Sterchi

Synonyms	Sucrase; Sucrase-isomaltase complex; Sucrase-isomaltase
Abbreviations	SI; S-I
Classifications	Glycosidase; EC 3.2.1.48-10
Description	A glycoprotein localized in the microvillus (brush border) membrane of small intestinal epithelial cells (enterocytes). The enzyme is not found in kidney microvillus membranes. SI is synthesized as a single chain precursor (pro-sucrase-isomaltase) in the ER of enterocytes, is post-translationally glycosylated and inserted into the luminal membrane. SI contains both O- and N-linked carbohydrates. Both the sucrase and isomaltase subunits contain eight N-linked glycan units, at least one of which is of the high mannose type and found on the sucrase. In addition, sucrase displays at least four populations varying in their content of O-linked glycans. After insertion into the luminal membrane, the pro-sucrase-isomaltase is proteolytically converted to its subunits sucrase and isomaltase by trypsin.
Structure	The complete 6 kb cDNA of human SI has been cloned. This cDNA contains a single ORF 5481 nt long which encodes a precursor polypeptide 1827 aa long. Amino acid 1-12, cytoplasmic domain; aa 13-32, signal-anchor (type II membrane protein; aa 33-1827, luminal domain; aa 43-60 Ser/Thr-rich stalk-region, probably a site of O-glycosylation; aa 62-109, P trefoil-like domain; aa 110-1007, isomaltase subunit; aa 1008-1827, sucrase subunit. The two subunits show 37.7% idendity and an additional 34% similarity and they both contain the active site sequence Trp-Ile-Asp-Met-Asn-Glu. The aa sequence of the human enzyme shows 75% identity with the rat SI; on the cDNA level, human SI is 83% identical to the rabbit enzyme. There is also homology to a yeast glucoamylase.
Molecular Weight	Human single chain precursor: 210,000 (high mannose intermediate), 212,000 (core glycosylated) and 245,000 (mature, i.e. fully glycosylated). Sucrase subunit: 130,000. Isomaltase subunit: 145,000. (All values determined by SDS-PAGE under reducing conditions). 209,403 as determined from the aa sequence.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	Glycosidase, α-glucopyranosidase
Coenzymes/Cofactors	None
Substrates	Sucrose, maltose, isomaltose, maltotriose, p-nitrophenyl-α-glucopyranoside, p-chlorophenyl-α-glucopyranoside, 4-methylumbelliferyl-α-glucopyranoside.
Inhibitors	Competitive inhibition by nojirimycin, 1-deoxynojirimycin, acarbose, and tris- (hydroxymethyl)-aminomethane (Tris).

Biological Functions	Hydrolysis of sucrose, maltose, isomaltose, and small dextrins in the lumen of the small intestine. Sucrase-isomaltase accounts for up to 80% of the maltase, most of the isomaltase and all of the sucrase activity in the microvillus membrane.
Physiology/Pathology	SI is responsible for the final steps in dietary sucrose, starch, and glycogen digestion in the small intestine. A decrease or total absence of sucrase-isomaltase activity causes sucrose intolerance, i.e. the development of fermentative and osmotic diarrhoe upon ingestion of sucrose. A deficiency in sucrase-isomaltase may be due to a primary defect, congenital sucrase-isomaltase deficiency (CSID), an autosomal recessive disease, or represent a secondary lesion due to a damaged intestinal mucosa (e.g. in coeliac disease or microvillus atrophy).
Degradation	Normally SI is eliminated through loss with desquamated enterocytes from villus tips. In addition it may be degraded intracellularly, particularly in cases where a faulty protein is synthesized (e.g. CSID).
Genetics/Abnormalities	The gene encoding for SI has been assigned to chromosome 3q25-26. Different molecular defects lead to a congenital sucrase-isomaltase deficiency (CSID). Using pulse-labelling techniques in organ cultured small intestinal biopsies and immuno-electronmicroscopy at least five distinct molecular phenotypes have been found. Phenotype I is characterized by an over-expression of a core-glycosylated pro-SI, which is not converted to the mature, i.e. complex glycosylated pro-SI efficiently, and fails to reach the luminal microvillus membrane. Phenotype II shows the synthesis of electrophoretically normal pro-SI (high mannose) at practically normal levels, but no conversion to mature pro-SI. No immunoreactive SI can be detected at the cell surface. In phenotype III, precursor-SI is synthesized with apparently normal molecular weight, transported to the luminal membrane and cleaved to the two subunits. The mutation in this phenotype thus affects the catalytic activity of the sucrase subunit only, without interfering with the intracellular transport. In a very recent study (Fransen et al.) two more variants have been described. In one of these, phenotype IV, a high-mannose form of SI accumulates in the ER and is partially missorted to the basolateral membrane, as shown by immuno-electron microscopy. In the second, phenotype V, SI-precursor is prematurely cleaved, probably intracellularly, the free sucrase subunit is degraded while the isomaltase subunit appears to be transported to the microvillus membrane correctly, as demonstrated by immuno-electron microscopy using isomaltase-specific antibodies. Recently, a glutamine to proline substitution has been shown to lead to a transport block of SI in a pre-Golgi compartment.
Half-life	Unknown
Concentration	8 - 10 % of total microvillus membrane protein (in adults). The normal value of enzyme (sucrase) activity: > 2 IU/g mucosa.
Isolation Method	Isolation from small intestinal mucosa: solubilization from microvillus membrane vesicles with detergents such as Triton X-100, Nonidet P-40/ sodium deoxycholate or by proteolytic enzymes, usually papain, followed by immunopurification using specific monoclonal and polyclonal antibodies.
Amino Acid Sequence	The SI precursor comprises 1827 aa. SWISS-PROT, accession number: P14410
Disulfides/SH-Groups	The subunits are not joined by disulfide bridges.

General References

Naim, H.Y. et al. Biosynthesis of the human sucrase-isomaltase complex. Differential O-glycosylation of the sucrase subunit correlates with its position within the enzyme complex. *J. Biol. Chem.* 1988, **263**:7242-7253.

Naim, H.Y. et al. Sucrase-isomaltase deficiency in humans. Different mutations disrupt intracellular transport, processing, and function of an intestinal brush border enzyme. *J. Clin. Invest.* 1988, **82**:667-679.

Semenza, G. Glycosidases. In: *Mammalian ectoenzymes.* Kenny, A.J. and Turner, A.J. (eds.), Elsevier Science Publishers B. V. (Biomedical Division) 1987, pp.265-287.

Sterchi, E.E. et al. Molecular aspects of disaccharidase deficiencies. In: *Molecular biology and its impact on gastroenterology,* Ciclitira, P.J. (ed.), Baillière's Clinical Gastroenterology. Baillière Tindall, 1990, pp. 79-96.

Fransen, J.A.M. et al. Naturally occuring mutations in intestinal sucrase-isomaltase provide evidence for the existence of an intracellular sorting signal in the isomaltase subunit. *J. Cell Biol.* 1991, **115**:45-57.

West, L.F. et al. Regional assignment of the gene coding for human sucrase-isomaltase (SI) to chromosome 3q25-26. *Ann. Hum. Genet.* 1988, **52**:57-61.

Ouwendijk, J., Moolenaar, C.E.C., Peters, W.J. et al. Congenital sucrase-isomaltase deficiency - identification of a glutamine to proline substitution that leads to a transport block of sucrase-isomaltase in a pre-golgi compartment. *J. Clin. Invest.* 1996, **97**:633-641.

Ref. for DNA/AA Sequences

Chantret, I., Lacasa, M., Chevalier, G. et al. Sequence of the complete cDNA and the 5' structure of the human sucrase-isomaltase gene. Biochem. J. 1992, 285:915-923.

SWISS-PROT, accession no: P14410

GenBank accession: X63597, S41833, S41836.

Superoxide Dismutase

Lucia Sacchetti and Giuliana Fortunato

Synonyms	Superoxide: superoxide oxidoreductase
Abbreviations	SOD
Classifications	EC 1.15.1.1
Description	The superoxide dismutases are a family of enzymes that catalyze the dismutation of variety of superoxide radicals; they own the function to protect living organisms from oxygen toxicity. Three different types of SOD have been described in mammalian tissues: Cu,Zn-SOD, Mn-SOD and extracellular SOD (EC-SOD). Another type of SOD, Fe-SOD that is structurally homologous to Mn-SOD, has been only found in bacteria. Cu,Zn-SOD and EC-SOD, are widely distributed in the cytosol and in extracellular fluids respectively, whereas Mn-SOD is mainly located in the matrix of mitochondria.
Structure	The three dimensional structures of SODs have been determined by X-ray crystal structural analysis. Cu, Zn-SOD is a dimeric molecule comprised of two identical subunits related by a non-crystallographic 2-fold axis. Each subunit is a cylindrical barrel made up of eight anti-parallel beta strands plus three external loops, and it contains one Cu and one Zn ion at the active site. Cu and Zn are bridged by His-63. Cu is also coordinated to His-48, His-46 and His-120 in a square planar geometry, whereas Zn also interacts with His-71, His-80 and Asp-83 (see figure). Most of the surface of Cu, Zn-SOD is negatively charged to repel O_2^-, however Cu ion at the active-site is partially buried at the base of a channel that contains several positively charged residues (e.g. Thr 137 and Arg 143) thought to be involved in the electrostatic guidance of the superoxide to the active site. A similar arrangement probably exists in Mn-SOD. Most mammalian Mn-SODs are tetrameric enzymes and contain manganese at their active site, where Mn binds His-26, His-81, His-167 and Asp-163. EC-SOD is a tetrameric copper- and zinc-containing glycoprotein.
Molecular Weight	32,000 Da (dimer Cu,Zn-SOD) 80,000 Da (tetramer Mn-SOD) 135,000 Da (tetramer EC-SOD)
Sedimentation Coeff.	3.02 (Cu,Zn-SOD from human erythrocytes)
Isoelectric Point	4.75 (Cu,Zn-SOD from human erythrocytes)
Extinction Coeff.	5.5 (265 nm,1 %,1 cm) Cu,Zn-SOD from human blood.
Enzyme Activity	Superoxide dismutases catalyze the conversion of two superoxide anions into hydrogen peroxide and molecular oxygen: $O_2^- + O_2^- + 2H^+ \rightarrow H_2O_2 + O_2$ The Cu ion dismutes O_2^- by a cyclical mechanism: it first attaches to an electron from one O_2^- to produce O_2 and then it donates an electron to a second O_2^- to produce, together with two protons, H_2O_2.
Coenzymes/Cofactors	Cu, Zn-SOD and EC-SOD require one Cu^{2+} and one Zn^{2+} per subunit for their activity. The Cu ion appears to function in the enzymatic reaction, whereas the Zn^{2+} ion may be replaced by Co^{2+}, Hg^{2+}, or Cd^{2+} without loss of activity. Mn-SOD contains one atom of manganese per subunit; its replacement results in loss of activity.

Substrates	The substrate of SOD is "O_2^-", as it is an unstable free radical that readily undergoes disproportionation or dismutation reaction. Enzyme activity is often measured by an indirect assay that combines a reaction generating "O_2^-" and an indicating scavenger for this radical (i.e. xanthine/xanthine oxidase, as generators of O_2^- and cytochrome c as indicating scavenger with which SOD competes for the flux of O_2^-).
Inhibitors	Cu, Zn-SOD and EC-SOD are reversibly inhibited by cyanide, they are also inactivated by diethyldithiocarbamate and hydrogen peroxide, whereas Mn-SOD is resistent to these reagents but is inhibited by azide and destroyed by chloroform plus ethanol.
Biological Functions	SOD is a major defense system against active oxygen species that are generated within all respiring organisms by many cytotoxic spontaneous and enzymatic oxidations.
Physiology/Pathology	The physiological roles of EC-SOD are still obscure. Mn-SOD is present in high concentrations in liver, heart, kidney, adrenal gland and brain grey matter; its presence in plasma, in healthy subjects, is probably the result of passive leakage from the intracellular space. Increases of serum Mn-SOD are reported in acute myocardial infarction (peak at 108 h) and in malignant conditions; in epithelial ovarium cancer, the serum levels of Mn-SOD correlate with the progression of disease. Leukocyte Mn-SOD activity is decreased in rheumatoid arthritis. Cu,Zn-SOD besides in erythrocytes, is also widely distributed in mammalian tissues and is particularly abundant in liver. Its activity is increased in erythrocytes from patients with Down's syndrome, uremia, renal failure and liver diseases. Cu,Zn-SOD activity is decreased in erythrocytes of diabetics, where the glycated enzyme (non-enzymatic glycosylation often causes inactivation of the protein) increases. Decreases are also reported in anemias and in Duchenne muscular dystrophy.
Degradation	Cu, Zn-SOD rapidly disappears from the circulation by renal filtration. The degradations of EC-SOD and Mn-SOD are unknown.
Genetics/Abnormalities	In humans, Cu, Zn-SOD and Mn-SOD are encoded by two different genes, located on chromosome 21q22.1 and 6q21 respectively. The Cu, Zn-SOD gene is 11 Kb long with 5 exons and 4 introns; the Mn-SOD gene is 15 Kb long and is composed of 6 exons and 5 introns. Mutations of the human Cu, Zn-SOD gene have been associated with familial amyotrophic lateral sclerosis, where they alter the conserved interactions critical to the beta-barrel fold and dimer contact of the enzyme, rather than catalysis.
Half-life	6–10 min: Cu,Zn-SOD; 5–6 h: Mn-SOD
Concentration	Blood serum Mn-SOD: male (99.8 ± 24.8 μg/L); female (88.8 ± 20.8 μg/L). Human erythrocyte Cu,Zn-SOD (0.5–0.75 μg/g Hb).
Isolation Method	Cu, Zn-SOD has been obtained in purified form from a variety of materials (e.g. blood, brain, liver, heart, pea, spinach, leaves, wheat germ, fungi) from animals, plants and microorganisms. The most widely used purification methods are chloroform-ethanol fractionation in connection with ammonium sulfate precipitation, ion exchange chromatography and affinity chromatography. After tissue extraction (e.g. 50 mM sodium acetate pH 5.5) EC-SOD can be purified by a lenghty series of ion-exchange and affinity chromatography.
Amino Acid Sequence	The aa sequence of the Mn-SOD subunit was obtained by translating the cDNA sequence isolated from a placental cDNA library. The protein

contains 198 aa and has an N-terminal leader sequence of 24 aa, essential for translocation into mitochondria. The aa sequence of the Cu, Zn-SOD human subunit contains 153 aa, it shows no homology with Mn-SOD aa sequence, suggesting that the two isoenzymes (Cu, Zn-SOD and Mn-SOD) have evolved from different ancestors.

Disulfides/SH-Groups

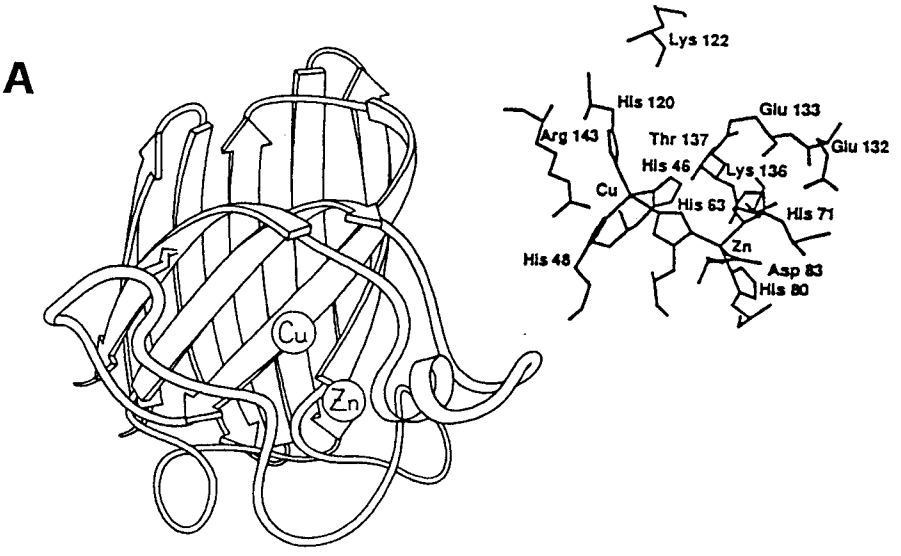
Cu, Zn-SOD: one intrachain disulfide bridge and two sulfhydryls per subunit. Mn-SOD: two sulfhydryls per subunit.

General References

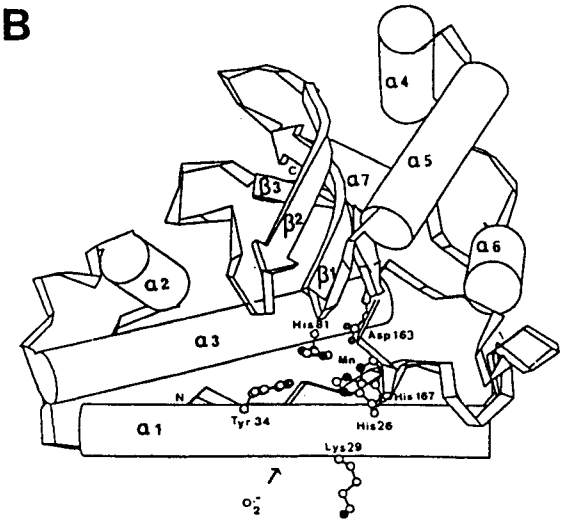
Taniguchi, N. *Adv. Clin. Chem.* 1992, **29**: 1–47.
Banci, L. et al. *Biochemistry* 1993, **32**: 4384–4388.
Fridovich, I. *Adv. Enzymol. Relat. Areas Mol. Biol.* 1986, **58**: 61–97.
Deng, H. X. et al. *Science* 1993, **261**: 1047–1051.
Rosen, D. R. et al. *Nature* 1993, **362**: 59–62.
Gurney, M. E. et al. *Science* 1994, **264**: 1772–1774.

Ref. for DNA/AA Sequences

Lenon, D. et al. *EMBO J.* 1985, **4**: 77–84.
Beck, Y. et al. *Nucleic Acid Res.* 1987, **15**: 9076.
GEN-EMBL Data Bank Accession No. Y00472 (human cDNA Mn-SOD).
Heckl, K. *Nucleic Acid Res.* 1988, **16**:6224.
GEN-EMBL Data Bank Accession No. X07834 (human placental cDNA Mn-SOD).



A – Diagrammatic representation of the Cu,Zn-SOD subunit (left) and interactions between the ions and the aminoacids at the active site (right). (Taken from: Tainer J. A., Getzoff E. D., Beem K. M., Richardson J. S. and Richardson D. C. *J. Mol. Biol.* 1982, **160**: 181–217 and Banci L., Bertini I., Bauer D., Hallewell R. A. and Viezzoli M. S. *Biochemistry* 1993, **32**: 4384–4388).



B – Diagrammatic representation of the Mn-SOD subunit and interactions between Mn and aminoacids at the active site. (Taken from: Parker M.W. and Blake C. C. F. *J. Mol. Biol.* 1988, **199**: 649–661).

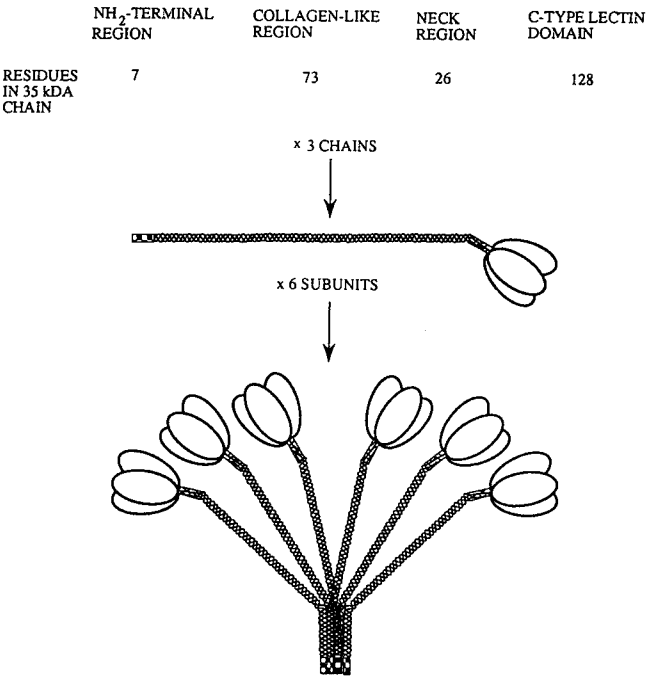
Surfactant Protein A

Uday Kishore and Kenneth B.M. Reid

Synonyms	Pulmonary surfactant apoprotein
Abbreviations	SP-A; PSAP
Classifications	Group III (the Collectins) of the C-type lectins.
Description	SP-A accounts for 5% of the weight of the surface active mixture of phospholipids and proteins which form the pulmonary surfactant in the lung. Nearly all the SP-A is tightly associated with lipids, such as dipalmitoylphosphatidylcholine and sphingomylein. It is considered to play an important role in surfactant secretion and uptake by type II alveolar cells, and it is also involved in the organization of tubular myelin. SP-A can also bind to carbohydrate structures, a property also shown by SP-D which has an overall structural similarity to SP-A. Both SP-A and SP-D appear to provide innate immunity against lung pathogens by binding to carbohydrates on the pathogens, agglutinating them and triggering effector mechanisms which kill and opsonize the pathogens. SP-A is produced by alveolar type II cells, nonciliated bronchiolar alveolar cells and the serous glands of the proximal trachea.
Structure	SP-A has a hexameric structure in which six structural subunits of 105 kDa associate to yield a molecule of 630 kDa. Each structural subunit is composed of three 35 kDa polypeptide chains which are held together by: disulfide bonds located in the N-terminal halves of the chains; a triple-helical collagen-like region formed by a 73-residues long Gly-Xaa-Yaa sequence in each chain; an α-helical coiled-coil region located between the collagen-like sequence and, C-terminal, C-type lectin domain present in each chain. The overall shape of SP-A is very similar to that of the complement protein C1q, both molecules appearing in the electron microscope as a bouquet-like structure with six globular heads linked by collagen-like strands to a fibril-like central core. In SP-A each of the globular heads is composed of three C-type lectin domains.
Molecular Weight	The hexameric SP-A molecule has an apparent molecular weight of approximately 630 kDa in non-dissociating conditions. Eighteen chains of 35 kDa are found in one molecule. Human SP-A contains two types of chain (α_2 and α_3) and it has been proposed that each trimeric subunit has one α_2 chain and two α_3 chains but this has not been formally proved.
Sedimentation Coeff.	Unknown
Isoelectric Point	Range of 4.8-5.5 shown by isoforms
Extinction Coeff.	8.83 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	Carbohydrate binding specificty: N-acetylmannosamine > L-fucose maltose > glucose > mannose >> galactose, GlcNac.
Inhibitors	None

Biological Functions	SP-A forms a part of the mixture of phospholipid and protein which lines the alveolar space and acts to reduce surface tension forces and prevent atelectasis during expiration. It can function as an inhibitor of phospholipid secretion by alveolar type II cells, via interaction with a high affinity receptor. SP-A binds to dipalmitoylphosphatidylcholine (DPCC) and galactosylceramide (GalCer) and preferentially enhances DPCC uptake by type II cells, as well as the incorporation of this lipid into lamellar bodies. Since DPCC is the principal component responsible for the biophysical properties of lung surfactant, SP-A may play an important role in phospholipid homeostasis in the alveolar space. SP-A can bind to carbohydrate structures on the surfaces of a wide variety of lung pathogens which include bacteria, viruses and fungi. This binding is mediated via the groups of trimeric C-type lectin domains which are located at the ends of the six collagen-like regions in each molecule. SP-A can also interact with macrophages and increase their chemotactic, phagocytic and chemiluminescent properties. It is therefore considered that SP-A may play an important role in the rapid recognition and clearance of pathogens, especially in immunodeficient individuals.
Physiology/Pathology	SP-A is important for the formation of tubular myelin and other surfactant aggregates and in concert with surfactant protein B facilitates rapid adsorption and spreading of surface active phospholipids at the air-liquid interface of alveoli. SP-A level in amniotic fluid has been used to assess the maturity of fetal lung. Its level in bronchoalveolar lavage fluid (BAL) is used for diagnosis and assessment of the activity of interstitial lung diseases. A rise in the level of SP-A has been reported in alveolar proteinosis patients. SP-A has also been detected in serum and increased SP-A levels in the serum from patients with idiopathic pulmonary fibrosis (5-fold higher than normals) and with alveolar proteinosis (6-fold increase). SP-A-deficient mice appear to have normal lung functions but are susceptible to bacterial infections.
Degradation	Unknown
Genetics/Abnormalities	Two transcribed SP-A genes (SP-AI and SP-AII) and one pseudogene have been localized to chromosome 10q21-24 and have been shown to lie within a cluster that includes the SP-D and mannose-binding lectin genes. The sequences and genomic organizations of the SP-A I and SP-A II genes are very similar with each probably being composed of seven exons. The expressed proteins are each encoded with four exons (I-IV): I covering part of the 5' untranslated region, the leader peptide, the N-terminal region and part of the collagen-like sequence; II covering the remainder of the collagen-like sequence; III the α-helical neck sequence; IV the C-type lectin domain plus the 3' untranslated sequence. The few differences in aa sequence seen between the SPA-1 and SPA-II mature polypeptide chains are located within the collagen-like region. Allelic variants of each gene are generated by splicing variability in the 5' untranslated regions and by sequence variability in the 3' untranslated regions. No complete genetic deficiencies of human SP-A have been described, however SP-A deficient mice have been produced and they do not appear to have any major respiratory defects or malfunction of regulation of surfactant phospholipid levels.

Concentration	Accurate measurements of SP-A concentrations in lung surfactant are difficult to make. SP-A is found at levels of approximately 5 mg/L in amniotic fluid at term.
Isolation Method	Commonly prepared from bronchoalveolar lavage by butanol extraction of sedimented surfactant and then further purified by affinity chromatography on a mannose agarose column and also by size fractionation.
Amino Acid Sequence	The mature forms of the two SP-A polypeptide chains designated α_2 (product of the SP-A II gene) and α_3 (product of the SP-A I gene) are both composed of 248 aa residues which includes: an N-terminal segment of 7 residues followed by a collagen-like region of 73 residues; an α-helical neck region of 26 residues and a C-terminal C-type lectin domain of 123 residues. There is one N-linked glycosylation site at Asn-198 in each chain.
Disulfides/SH-Groups	There are seven cysteine residues, in the mature α_3 polypeptide encoded by the SP-A I gene and six cysteine residues in the α_2 chain encoded by the SP-A II gene. The mature α_2 chain contains cysteine residues at positons 6 and 48 while the α_3 chain contains cysteine residues at positions 6, 48 and 65. The remaining 4 cysteine residues in each chain are the conserved intra-chain cysteine residues found within the C-type lectin domain. The disulfide bridges have not been mapped but the cysteines at positions 6, 48 and 65 are considered to form interchain disulfide bonds.
General References	Korfhagen, T.R. et al. *Proc. Natl. Acad. Sci USA* 1996, **93**:9594-9599. Hawgood, S. and Clements, J.A. *J. Clin. Invest.* 1990, **86**:1-6. Floros, J. and Karinch, A.M. *Am. J. Physiol.* 1995, **268** (*Lung Cell Mol. Physiol. 12*):L162-L165. McCormack, F.X. and Whitsett, J.A. In: *Collectins and innate immunity*. Ezekowitz, R.A.B., Sastry, K. and Reid, K.B.M. (eds.) R.G. Landes Co., Texas 1996, pp 9-50. Voss, T. et al. *Am. J. Respir. Cell Mol. Biol.* 1991, **4**:88-94. La Vine, A.M. et al. *J. Immunol.* 1997, **158**:4336-4340.
Ref. for DNA/AA Sequences	White, R.T. et al. *Nature* 1985, **317**:361-363. Floros, J. et al. *J. Biol. Chem.* 1986, **261**:9029-9033. Katyl, S.L. et al. *Am. J. Respir. Cell Mol. Biol.* 1992, **6**:446-452.



Surfactant protein A

Surfactant Protein D

Uday Kishore and Kenneth B.M. Reid

Synonyms	CP4
Abbreviations	SP-D
Classifications	Group III (the Collectins) of the C-type lectins.
Description	SP-D is one of the proteins found in the airspace lining material in the lungs. Although it does bind to specific phospholipids in the lung surfactant it shows quite different properties to those of the hydrophobic peptides, SP-B and SP-C, which are strongly associated with lipids. SP-D has an overall structural similarity to SP-A, the major surfactant protein, and both appear to be involved with providing innate immunity against lung pathogens. Like SP-A, SP-D is a lectin which recognizes and binds to carbohydrate structures on pathogens, thus triggering effector mechanisms which kill and opsonize the pathogen. Pulmonary SP-D is produced by alveolar type II cells and nonciliated bronchiolar alveolar cells. However, SP-D may not be a lung specific protein since low levels of material antigenically similar to SP-D are found in normal human serum and, animal studies indicate the presence of SP-D, or SP-D like proteins, in gastric mucosa, tracheobronchial, lacrymal and salivary glands.
Structure	SP-D is composed of oligomers of a 130 kDa subunit formed from three identical, 43 kDa, polypeptide chains. As judged by electron microscopy, human SP-D is assembled into a 520 kDa tetrameric structure with four of the 130 kDa, homotrimeric, subunits linked via their N-terminal regions, but trimers, dimers and monomers of the 130 kDa subunit are also seen in SP-D preparations. The triple-helical arms in each 130 kDa subunit are approximately 46 nm in length and although appearing flexible show no sharp bends, or distribution, which is consistent with there being no interruptions to the Gly-Xaa-Yaa repeating sequence in SP-D. Clusters of three C-type lectin domains are held together by the α-helical coiled-coil region found at the C-terminal end of the collagen-like triple helix present in each 130 kDa subunit. Up to eight of the 520 kDa tetrameric structures can undergo further oligomerization to give SP-D multimers having a large array, of up to 96 (8 x 12), C-type lectin domains.
Molecular Weight	The molecular weight of a single polypeptide chain of SP-D is 43 kDa, (which includes an N-linked oligosaccharide structure at Asn-70). The molecular weight of the homotrimeric subunit is 130 kDa, although it has an apparent molecular weight of between 120-180 kDa when examined by different SDS-PAGE methods under non-reducing conditions. The tetrameric form of native SP-D has a molecular weight of 540 kDa in non-dissociating conditions.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	4.0 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None

Substrates	Carbohydrate binding specificity: Maltose > Mannose, Glucose > Lactose, Galactose > GlcNAc > Fucose.
Inhibitors	None
Biological Functions	SP-D binds to carbohydrates/lipopolysaccarides on the surfaces of a variety of pathogens such as influenza A virus; Gram-negative bacteria (*E. coli, Salmonella, Pseudomonas aeruginosa, Klebsiella pneumoniae*) fungal organisms (*Cryptococcus neoformans, Pneunocystis carinii, Aspergillus fumigatus*). The binding of SP-D to these organisms leads to reduction in infectivity, agglutination and enhanced killing. SP-D appears to achieve some of these protective effects by triggering phagocytic cells such as alveolar macrophages and neutrophils, but the putative receptors on these cells for SP-D, or SP-D-pathogen complexes, have not yet been characterized.
Physiology/Pathology	There are no reports of complete SP-D deficiency. Raised bronchoalveolar lavage levels of SP-D are seen in alveolar proteinosis and preliminary studies have indicated a lowering of levels in asthmatics.
Degradation	Unknown
Genetics/Abnormalities	Human SP-D is encoded by a single gene with 7 exons spanning >11 kb of DNA on the long arm of chromosome 10, at a locus on 10q22.2-23.1, which also includes the SP-A genes. The first protein-encoding exon (exon 2) includes sequences corresponding to the signal peptide, N-terminal region and first seven Gly-Xaa-Yaa triplets. The remainder of the collagen region (the total of 59 Gly-Xaa-Yaa repeats), is encoded by three exons of the same 117 bp size. The sixth exon encodes the coiled-coil neck region, the linking sequence between collagenous region and the carbohydrate recognition domain (CRD). The seventh exon codes for the CRD.
Half-life	Unknown
Concentration	Accurate measurements of SP-D concentrations in lung surfactant are difficult to make. The concentrations of SP-D in the bronchoalveolar lavage of healthy individuals is approximately 800 μg/L. Material showing immunochemical cross-reactivity with SP-D is found in the serum of normal adults at a level of 70 μg/L. SP-D is found at levels of approximately 700 μg/L in amniotic fluid at term.
Isolation Method	Affinity chromatography of bronchoalveolar lavage or amniotic fluid on maltose-agarose followed by elution with maltose or EDTA.
Amino Acid Sequence	Each chain contains four distinct regions: a 25 residue-long N-terminal distinct region which contains cysteine residues involved in interchain disulfide bonding; a 177-residue long collagen-like region formed of Gly-Xaa-Yaa repeating triplets; a 28-residue long α-helical neck region; a 125 residue long C-type lectin domain. The 130 kDa homotrimeric subunit is held together by means of association of the collagen-like regions into a triple-helix, and the neck regions into an α-helical coiled-coil. This trimeric subunit structure is characteristic of other members of group III of the C-type lectins i.e. SP-A, mannose-binding lectin, conglutinin and collectin-43, which all contain collagen-like regions linked to C-type lectin domains. SP-D shows a strong overall similarity to the two bovine serum collectins, conglutinin and collectin-43. An important structural feature is the repeating heptad pattern of hydrophobic residues, in the a and d positions,

seen in the neck region of the SP-D polypeptide chain. The alignment of these hydrophobic residues allows the formation of a self-associating triple-stranded parallel α-helical bundle which determines the trimeric orientation of the C-type lectin domains and possibly also acts as a nucleation point for triple-helix formation

VASLRQQVEALQGQVQHLQAAFSQY
a d a d a d a d

Disulfides/SH-Groups

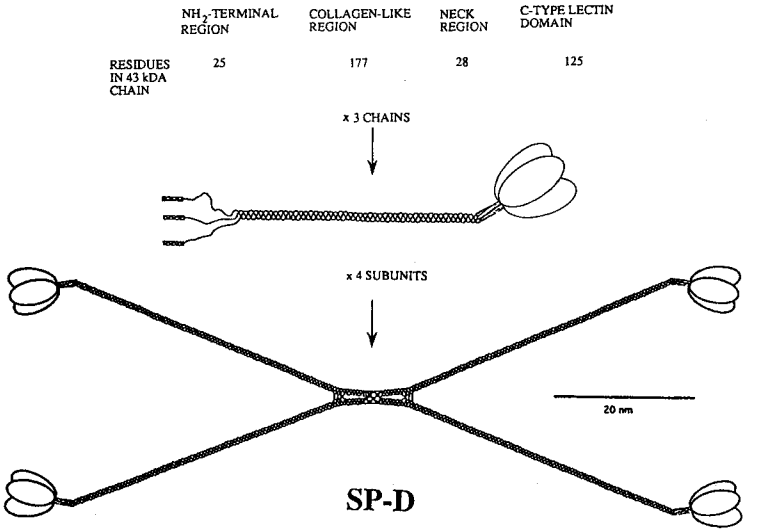
The two cysteines at positions 15 and 20 (mature protein numbering of the 43 kDa chain are considered to be involved in the interchain disulfide bonding while the four within the C-type lectin domain, are considered to be involved in intrachain disulfide bonding (261-353, 331-345).

General References

Crouch, E. and Hartshorn, K. Surfactant Protein D. In: *Collectins and innate immunity*. Ezekowitz, R.A.B., Sastry, K. and Reid, K.B.M. (eds.) R.G. Landes Co., Texas 1996. pp 113-131.
Crouch, E. et al. *J. Biol. Chem.* 1994, **269**:17311-17319.
Brown-Augsburger, P. et al. *J. Biol. Chem.* 1996, **271**:18912-18919.

Ref. for DNA/AA Sequences

Lu, J. et al. *Biochem. J.* 1994, **284**:795-802.
Rust, K. et al. *Arch. Biochem. Biophys.* 1991, **290**:116-126.
Crouch, E. et al. *J. Biol. Chem.* 1993, **268**:2976-2983.
The nucleotide sequence data is in the EMBL, GenBank and DDBJ databases under the accession number X65018.



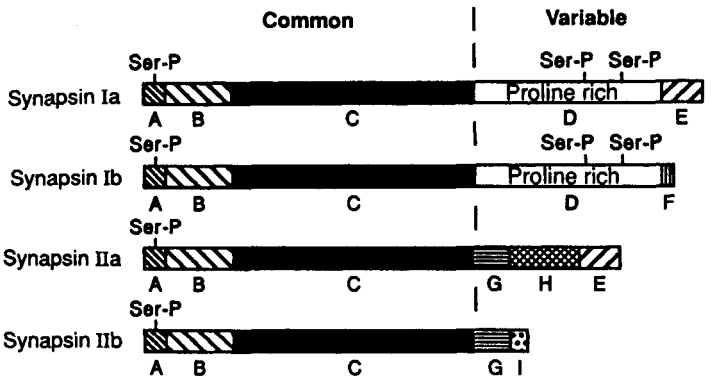
Surfactant protein D

Synapsin I, Synapsin II

Gerald Thiel

Synonyms	Protein I, Protein III
Abbreviations	None
Classifications	None
Description	The synapsins are a family of proteins which are highly specific to synapses of the central and peripheral nervous system. They include four homologous proteins, synapsin Ia and synapsin Ib (collectively referred to as synapsin I) and synapsin IIa and IIb (collectively referred to as synapsin II), which are derived from two genes by differential splicing. All four synapsins are localized in the nerve terminal, where they coat the cytoplasmic face of small synaptic vesicles, which contain the classical neurotransmitters. The synapsins are phosphorylated upon excitation of the nerve terminal. They serve in vivo and in vitro as substrates for the cAMP-dependent protein kinase and the Ca^{2+}/calmodulin-dependent protein kinase I. Synapsin I can also be phosphorylated on two sites by Ca^{2+}/calmodulin-dependent protein kinase II. In vitro synapsin I binds to actin, tubulin, spectrin and calmodulin. It bundles actin filaments, suggesting the presence of several actin binding sites.
Structure	Synapsin I is an elongated protein, consisting of a globular head domain and an elongated tail domain. Synapsin II is less elongated, specially synapsin IIb, which contains mainly the globular head domain. Not yet crystallized in its native form.
Molecular Weight	86,000 and 80,000 for synapsins Ia and Ib (SDS-PAGE); 74,000 and 55,000 for synapsins IIa and IIb (SDS-PAGE). Molecular weights deduced from rat cDNA clones: 73,997 and 69,918 (synapsins Ia and Ib), 63,449 and 52,315 (synapsins IIa and IIb).
Sedimentation Coeff.	2.9 S for synapsins I and II
Isoelectric Point	Ia and Ib: > 10.5; IIa: 6.8–7.1; IIb: 6.6–6.9
Extinction Coeff.	6.74 (277 nm, 1%, 1 cm)
Enzyme Activity	Unknown
Coenzymes/Cofactors	Unknown
Substrates	Unknown
Inhibitors	Unknown
Biological Functions	The synapsins are believed to function in the regulation of neurotransmitter release by controlling the number of vesicles available for exocytosis. In the presynaptic nerve terminal synapsin I links synaptic vesicles to actin filaments. These interactions are controlled by reversible phosphorylation of synapsin I through various signal transduction pathways. Nerve impulses increase the intracellular calcium concentration in the nerve terminal, which leads to activation of Ca^{2+}/calmodulin-dependent protein kinase II. The kinase phosphorylates synapsin I, leading to a decrease in binding of synapsin I to synaptic vesicles and cytoskeletal elements. Thus, vesicles are

	released from the cytoskeleton and are available for exocytosis. Synapsin II is highly homologous to synapsin I suggesting that some of the functional properties of synapsin I are also shared by synapsin II. However, synapsin II lacks the phosphorylation sites for Ca^{2+}/calmodulin-dependent protein kinase II, which are believed to be responsible for the phosphorylation-dependent binding of synapsin I to synaptic vesicles and F-actin.
Physiology/Pathology	Unknown
Degradation	Unknown
Genetics/Abnormalities	In humans, the gene for synapsin I is localized on chromosome X and the gene for synapsin II on chromosome 3. The synapsin I locus may be affected in X-linked genetic diseases involving the nervous system, such as Rett's syndrome, but currently only the chromosomal localization is known.
Half-life	Unknown
Concentration	Synapsin I: 0.4% of total brain protein, 6–7% of total vesicle protein. Synapsin II: 0.2% of total brain protein.
Isolation Method	Synapsins Ia and Ib: Rat or bovine brains are homogenized under isotonic conditions and centrifuged. Synapsin I is extracted from the particulate fraction with Nonidet-P40 (NP-40) in the presence of salt (150 mM NaCl), followed by cation-exchange, hydroxyapatite and gel filtration chromatography. Synapsin IIb can be purified from human, rat and bovine brain by immunoaffinity chromatography of the NP-40 extract using a monoclonal antibody specific for synapsin II. Synapsin IIa can be purified in a similar way from Spodoptera frugiperda (Sf9) cells infected with a recombinant baculovirus.
Amino Acid Sequence	The aa sequence surrounding phosphorylation site 1 (RRLS(P)) in both synapsins I and II conforms to the consensus sequence RRXS(P) found in many substrates for cAMP-dependent protein kinase. The sequences around phosphorylation sites 2 and 3 in synapsin I (RQTS(P) and RQAS(P)) conform to the consensus sequence RXXS(P)/T(P) for Ca^{2+}/calmodulin-dependent protein kinase II. No statistically significant homology has been observed with other actin binding proteins (villin, profilin, gelsolin, protein 4.1).
Disulfides/SH-Groups	Possible intermolecular S-S bridges in synapsin I.
General References	Südhof, T. C., et al. *Science* 1989, **254:** 1474–1480. Bähler, M., et al. *BioEssays* 1990, **12:** 259–263. DeCamilli, P., et al. *Annu. Rev. Cell Biol.* 1990, **6:** 433–460.
Ref. for DNA/AA Sequences	Südhof, T. C., et al. *Science* 1989, **254:** 1474–1480. Südhof, T. C. *J. Biol. Chem.* 1990, **265:** 7849–7852. The nucleotide sequences are deposited to the GenBank database (bovine synapsin Ia, M27810; bovine synapsin Ib, M27811; rat synapsin Ia, M27812; rat synapsin Ib, M27924; rat synapsin IIa, M27925; rat synapsin IIb, M27926; human synapsin I gene, J05431).


Common
Variable
Ser-P
Ser-P
Ser-P
Synapsin Ia
Proline rich
A
B
C
D
E
Ser-P
Ser-P
Ser-P
Synapsin Ib
Proline rich
A
B
C
D
F
Ser-P
Synapsin IIa
A
B
C
G
H
E
Ser-P
Synapsin IIb
A
B
C
G
I

Tamm-Horsfall Protein

Krishna Prasadan and Satish Kumar

Synonyms	Tamm-Horsfall glycoprotein; Uromodulin; Uromucoid
Abbreviations	THP
Classifications	Electrophoretic Mob.: Slightly faster than albumin at pH 8.6.
Description	An abundant urinary protein, synthesized exclusively in kidney tubular cells of the thick ascending limb of the loop of Henle. It is released into the urine in a soluble form but can gel to form cylindrical casts of renal tubules. The tendency to gel is increased *in vitro* by increasing concentrations of calcium, sodium and hydrogen ions.
Structure	Crystal structure of THP is not known. Circular dichoric spectroscopy suggests a β-structure. Electron microscopic studies of THP show a "fishing net" appearance made of filaments of 15 - 45nM in diameter and a pore size of 0.1 - 1.0nM.
Molecular Weight	80,000 - 100,000 Da (30% carbohydrate). Native protein occurs as a polymer of 1 - 10 million Da molecular weight.
Sedimentation Coeff.	29.5 S
Isoelectric Point	pI = 4.5
Extinction Coeff.	10.8 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Not known. THP has been shown to bind specifically to a number of substances *in vitro*. These include cells (neutrophils, lymphocytes and monocytes), cytokines (interleukins I and II, and tumor necrosis factor), calcium, immunoglobulins, calcium oxalate crystals and bacteria.
Physiology/Pathology	Not known. Physiological role for THP has been postulated in water and ion transport in kidney tubules. Pathological role for THP has been postulated in acute renal failure, myeloma kidney, nephrolithiasis and in urinary tract infection.
Degradation	THP is expressed on the luminal cell surface of renal tubular cells in the thick ascending limb of Henle's loop where it is attached to the cell surface membran0e by a glycosyl-phosphatidyl-inositol (GPI) anchor. It is released into the urine, presumably by action of a GPI-specific phospholipase.
Genetics/Abnormalities	The gene is on chromosome 16 at p12.3 - 13.11. No definite phenotype of THP deficiency or abnormality has been established in humans or animals.
Half-life	16 hours

Concentration	50 - 200 mg/day is excreted in urine.
Isolation Method	Add 0.58 M NaCl to urine and allow to settle for 24 hrs. Centrifuge and discard the supernatant. Wash the pellet 3 times with 0.58 M NaCl. Dissolve the pellet in H_2O in 10% of the original volume, add 0.58 M NaCl and stir overnight at 4°C. Centrifuge, dissolve the pellet in minimal volume of H_2O and dialyze overnight against H_2O. Centrifuge and discard the pellet. Lyophilize the sample.
Amino Acid Sequence	640 aa including a signal peptide of 24 aa. THP cDNA sequence shows several modular domains: - ZP domain 486 YVGTMLDGGD LSRFALLMTN CYATPSSNAT DPLKYFIIQD RCPHTRDSTI QVVENGESSQ GR 547 This domain is contained in the zona pellucida proteins ZP2 and ZP3, in pancreatic protein GP2 and in betaglycan. - Epidermal growth factor (EGF) like repeats C X_{2-7} C X_{4-5} C X_8 C X C X_{10-17} C THP has 22 cysteine residues arranged in four EGF like repeats, three in tandem and the fourth in reverse orientation. The second and third EGF-repeats in THP show a consensus sequence that is seen in sub-type of EGF-repeats associated with post-translational hydroxylation of the Asp or Asn residue between the third and fourth cysteine residues. D X D E C X_7 C X_{4-5} C V N X_4 Y/F X C X C X_{10-11} C The hydroxylated Asp/Asn site has been shown to be a site for calcium binding in other proteins. - RGD Sequence The third EGF-repeat contains an arginine-glycine-aspartate (RGD) sequence in human and rat THP, but this sequence is absent in the mouse. GSYLCVCPAG **YRGD**GWHCEC SPGSCGPGLD CVPEGDALVC ADPCQAHRTL - Glycosyl-phosphatidyl-inositol (GPI) anchor. At the C-terminal is a consensus sequence for attachment of a GPI anchor. It has been shown that recombinant THP can be released from the surface of cultured cells transfected with the THP cDNA by GPI-specific phospholipase C. THP in human urine contains remnants of the GPI anchor. - Glycosylation There are ten consensus sites for asparagine-linked glycosylation. Nine of the ten N-glycosylation sites are conserved between mouse and rat and eight sites are conserved between mouse and man. Analytical data in human THP suggests that five glycosylation sites are utilized. The oligosaccharides are mostly complex type. A small number of high-mannose oligosaccharides are present. No O-linked oligosaccharides are found.
Disulfides/SH-Groups	48 cysteine residues, all involved in intrachain disulfide bonds. No free SH-groups.
General References	Tamm, I. and Horsfall, F.L. *Proc. Soc. Exp. Biol. Med.* 1950, **74**: 108-114. Hoyer, J.R. and Seiler, M.W. *Kidney Int.* 1979, **16**:279-289. Kumar, S. and Muchmore, A. *Kidney Int.* 1990, **37**:1395-1401. Kumar, S. and Stein, J.H. Tubular cast formation and Tamm-Horsfall glycoprotein. In: *Acute Renal Failure.* Goligorsky, M.S. (ed.), Churchill Livingstone, New York. 1995, pp 267-286.
Ref. for DNA/AA Sequences	Pennica, D. et al. *Science* 1987, **236**:83-88. Hession, C. et al. *Science* 1987, **237**:1479-1484. Fukuoka, S. et al. *Proc. Natl. Acad. Sci. (USA).* 1992, **89**:1189-1193. Prasadan, K. et al. *Biochim. Biophys. Acta* 1995, **1260**:328-332.

Tenascin

Annalisa Siri and Luciano Zardi

Synonyms	Cytotactin, hexabrachion, myotendinous antigen, glyal-mesenchymal extracellular matrix protein (GMEM), "J1" antigen.
Abbreviations	TN, GMEM, J1, HxB
Classifications	Extracellular matrix glycoprotein
Description	A polymorphic high molecular mass (about 1.9×10^6 Da) extracellular matrix glycoprotein composed of six similar subunits joined together at their N-terminal by disulfide bonds. It has binding sites for heparin, fibronectin and chondroitin sulfate proteoglycans. TN presence has been documented in the central nervous system and, in a diffuse layer, beneath the basement membrane of different epithelia of ectodermic and endodermic origin: along the dermal-epidermal junction, the laryngo-tracheobronchial tract, the esophagus, the doudenum and the colon-rectum. The liver displays homogeneous decoration of the sinusoids. TN is also detectable at the interface between some epithelia of mesodermal origin and their stroma such as kidney. TN has also been found in some tissues of mesenchymal origin including vascular walls, perinerurium, the perichondrium of hyaline cartilage, smooth-muscle and tendons. The sequence of cDNA clones codifying for the complete human TN molecule has been reported. The deduced aa sequence shows that human TN is mainly made up of 14 and one half epidermal growth factor like repeats, of 16 units similar to fibronectin type III homology repeats and, at the C-terminal, of a sequence with homology to the globular domain of the β- and γ-chains of fibrinogen. Recently two TN related proteins have been described. Both show the same general structure but include a lower (restrictin and J1-160/180) or a higher (TN-X and HTN-like) number of repeats. Different terminology systems have been suggested for the TN protein family but no one has been so far, widely accepted.
Structure	TN micrographs obtained by shadowing electron microscopy show its six armed structure; one of its names, hexabrachion, stems from this characteristic feature. Each arm is formed of a terminal knob, a thick distal segment and a thin proximal segment. Three arms are joined in a T junction and two trimers are linked together in a central knob. The average arm length of human TN purified from conditioned medium of U-251 MG glioma cell line is 87 nm. The far-UV circular dichroism spectrum indicates a predominance of β-structure and a lack of collagen-like or α-helical structure.
Molecular Weight	1,900 kDa (native, non-reduced, 0.75% agarose gel). From sedimentation equilibrium studies the M_r was calculated to be $1.9 \pm 0.5 \times 10^6$. Cultured human fibroblasts produce two different major subunits (generated by alternative splicing of the pre-mRNA) with molecular masses of 280 and 190 kDa (SDS-PAGE, reducing conditions). Cultured melanoma cell lines (SK-MEL-28) produce a single TN subunit of about 330 kDa (SDS-PAGE, reducing conditions). Transformed or tumor derived human cells produce TN that is more sialylated compared to TN from normal cell lines.
Sedimentation Coeff.	14 S
Isoelectric Point	Unknown

Extinction Coeff.	9.7 (277 nm, 1%, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	During development, TN displays a time and a space dependent tissue distribution with morphogenically significant boundaries and has been proposed as a possible modulator of epithelial-mesenchymal and neuronal-glial cell interaction during morphogenesis. There are a number of reports suggesting that TN plays a role in a variety of processes including cell spreading, adhesion, anti-adhesion and proliferation. However, there are also contradicting data due most likely to different isoform compositions of the TN preparations used in the various reports and to objective difficulties in obtaining truly pure TN. Many concerns about the functional role of TN have been raised by a recent paper reporting that homozygous mice, obtained from embryonic stem cells in which the TN gene was nully mutated by homologous recombination, develop normally and do not display any phenotypic alteration. Two different hypotheses have been raised to explain this observation. The first is based on the concept of compensation by functionally related genes [two other molecules with structure similar to TN have been recently reported]. The second explanation advances the notion of the redundancy of gene products. This last hypothesis suggests the expression of a gene product even in anatomical regions where it is not needed, thus refuting the assumption that a gene is expressed only where the protein is necessary.
Physiology/Pathology	Neoexpression or markedly increased expression of TN has been documented during wound healing and in a variety of tumors. This last observation suggests that this glycoprotein may play a role in tumor-cell stroma interdependence and that its high level in cancer has potential diagnostic and therapeutic implications. Monoclonal antibodies to human TN have been used in the therapy of glioblastoma. Moreover, it has been documented that tumor stroma cells in vivo as well as tumor derived cells in vitro preferentially produce the larger TN forms. Some differences in biological function between TN isoforms have recently been reported: in particular focal adhesion integrity is down regulated by the alternatively spliced domain of human TN and its expression seems to facilitate cell migration.
Degradation	The in vivo degradation of TN is unknown. It is sensitive to pepsin, pronase, α-chymotrypsin and thermolysin.
Genetics/Abnormalities	Synthesized from at least 8 different mRNA derived from alternative splicing patterns of a single primary transcript. The TN gene has been localized in q32/34 region of human chromosome 9.
Half-life	Unknown
Concentration	Concentration of TN in human plasma of healthy blood donors has been reported to be about 150 μg/L. Several human fibroblasts, glial cells, glioma, melanoma, as well as some epithelial cell lines produce TN and accumulate about 1–20 $\mu g/10^6$ cells 6 days after plating. It has been shown that 60–90% of the produced TN is accumulated in the extracellular matrix of normal human fibroblasts; on the contrary, transformed or tumor derived cell lines release more than 95% into culture media.

Isolation Method

Most of the methods developed for TN purification both from culture media or tissues are based on specific immunoadsorbents. Some authors have used single-step procedures of affinity chromatography on monoclonal antibodies, while others have used this technique associated with anion-exchange chromatography, gel filtration or glycerol-gradient sedimentation. Recently three methods excluding the use of specific immunoadsorbents have been described: two of them are based on gel filtration and ion-exchange chromatography and the third on hydroxyapathite chromatography. The best sources for human TN purification are the conditioned media from the U-251 MG glioma and the SK-MEL-28 melanoma cell lines. The media from these lines contain 5–20 μg/ml of TN. Both these cell lines mainly produce the high molecular mass TN isoform.

Amino Acid Sequence

The longer human TN reading frame deduced from cDNA sequence is 2201 aa to which we must add an additional 92 aa of a novel type III repeat (AD1) recently reported. The first 22 aa correspond to a typical signal peptide as expected for a secreted protein; these are followed by a strongly basic sequence of 10 residues. The adjacent domain (34–173) contains 8 Cys residues and a stretch of α-helix from 119–147. This α-helix could form a triple stranded coil that contributes to the association of subunits in trimers and the Cys residues present in this domain could stabilize this structure. The following sequence is composed of 14 and a half EGF-like repeats containing 31 aa each. Adjacent to these follow 16 repeats similar to the type III homology repeat of fibronectin. The first 5 and the last 3 are present in all cDNA clones sequenced and in all protein isoforms described. On the contrary the repeats from A1 to D have been shown to be included or ommitted according to a complex alternative splicing pattern of the TN pre-mRNA. Recently, a novel alternatively spliced type III repeat (AD1), has been reported. The great majority of glycosylation sites of TN are contained in these repeats. The repeats from A1 to A4 have extremly high homology among them, and with the repeat A in chicken thus supporting the hypothesis of recent reduplication of these repeats. The C-terminal domain contains a sequence of about 200 aa similar to β- and γ-chains of fibrinogen; this sequence shows a high degree of homology with the similar domain in chicken, mouse and pig, thus suggesting functional importance.

Disulfides/SH-Groups

In the N-terminal sequence, the most proximal Cys could connect two trimers together in the hexamer formation, while the other 7 contained in this sequence could link three subunits into a trimer. The EGF-like repeats contain 6 Cys each (87 total) that define the secondary structure of these sequences through the formation of intrachain bonds. 4 Cys residues are also contained in the fibrinogen-like sequence.

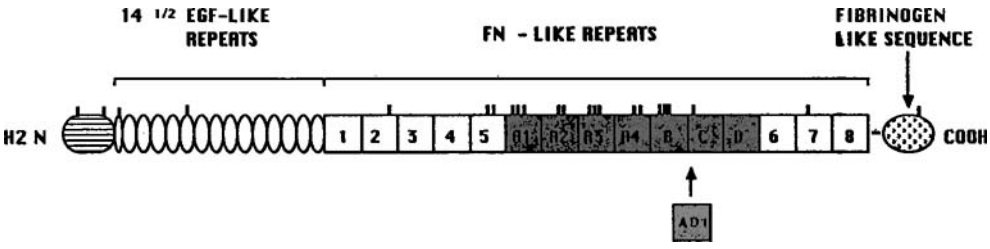
General References

Erickson, H. P. and Bourdon, M. A. *Ann. Rev. Cell Biol.* 1989, :71–92.
Taylor, H. C. et al. *J. Cell Biochem.* 1989, **41:** 71–90.
Chiquet-Ehrismann, R. *FASEB J.* 1990, **4:** 2598–2604.
Koukoulis, G. K. et al. *Hum. Pathol.* 1991, **22:** 636–643.
Edelman, G. M. and Jones, F. S. *Trends Biochem. Sci.* 1992, **17:** 228–232.

Ref. for DNA/AA Sequences

Nies, D. E. et al. *J. Biol. Chem.* 1991, **266:** 2818–2823. EMBL Data Bank Accession No. M55618 (human TN).
Siri, A. et al. *Nucleic Acids Res.* 1991, **19:** 525–531. EMBL Data Bank Accession No. X56160 (human TN).
Gulcher, J. R. et al. *Proc. Natl. Acad. Sci. USA* 1991, **88:** 9438–9442 (human TN).
Sriramaro, P. and Bourdon, M. A. *Nucleic Acids Res.* 1993, **21:** 163–168. GenBank Accesion No. M67026 (human TN).

References for TN parologues (related genes):
Norenberg, U. et al. *Neuron* 1992, **8:** 849–863. EMBL Data Bank Accession No. X64649.
Fuss, B. et al. *J. Cell Biol.* 1993, **120:** 1237–1249. EMBL Data Bank Accession No. Z18630.
Matsumoto, K. et al. *Genomics* 1991, **12:** 485–491. EMBL/GenBank/DDBJ Data Bases Accession No. X60189 (TNC).
Matsumoto, K. et al. *Immunogenetics* 1992, **36:** 400–403. GenBank Accession No. X62489.
Bristow, J. et al. *J. Cell Biol.* 1993, **122:** 265–278. EMBL Data Bank Accession No. X71923–X71938.



Model of the domain structure of human tenascin subunit. The EGF-like, FN-like repeats as well as fibrinogen like sequences are indicated. The FN-like repeats from A1 to D, whose expression is regulated by the alternative splicing of the pre-mRNA, are shaded. AD1 is the novel type III repeat recently reported. The potential N-linked glycosylation site are indicated by small dashes.

Tetranectin

Inge Clemmensen

Synonyms	
Abbreviations	TN
Classifications	Electrical mobility: β1-fraction
Description	A circulating plasma protein composed of four identical peptide chains, non-covalent linked, containing 181 aa each. The protein is not glycosylated.
Structure	Unknown
Molecular Weight	80,400 (from aa sequence, 20,100 × 4).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.8
Extinction Coeff.	12.5 (280 nm, 1 %, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	The protein binds to kringle 4 in plasminogen in a lysine dependent manner, Ca^{2+} and sulphated polysacchrides. It is present in the cytoplasma of all epithelial and mesenchyma cells as well as in endocrine cells. All cells with a function of secretion and or storage contains tetranectin in the cytoplasma. It is absent in normal connective tissue.
Physiology/Pathology	The biological function is still unknown. It is secreted from neutrophils upon stimulation by PMA or FMPL. Embryonal lung fibroblasts in vitro secrete and deposit tetranectin into the extracellular matrix. In some breast carcinomas where the surrounding fibroblasts are activated, tetranectin is found in connective tissue as well as in the newly formed connective tissue in tissue repair. It has been proposed that the protein might have a function in tissue repair and remodeling. Recently tetranectin present in extracellular matrix around tumor cells was found associated with a poor prognosis in ovarian carcinoma. Also low concentration of the protein in plasma has been found associated with a poor prognosis in both breast and ovarian carcinoma.
Degradation	The protein is sensitive to degradation by plasmin during its purification from plasma.
Genetics/Abnormalities	Unknown
Half-life	Unknown
Concentration	The concentration in plasma varies with age and sex. In adult 10–12 mg/L, newborn 5–6 mg/L, increacing to 10 mg/L in the age of 12–13 years. The concentration is very low in cancer diseases especially when disseminated.

Isolation Method	From plasma: Ammoniumsulphate precipitation, affinity chromatograpy on plasminogen or plasminogen kringle 4-Sepharose, ion-exchange chromatography and gelfiltration.
Amino Acid Sequence	Sequence homology with the C-terminal part of different lectins e.g human hepatic asialoglycoprotein receptor and C-terminal part of proteoglycan core protein from chondrosarcoma.
Disulfides/SH-Groups	12 disulfides, 3 intrachain in each monomer, no inter-chain and no free sulfhydryls.
General References	Clemmensen, I., et al. *Eur. J. Biochem.* 1986, **156:** 327–333. Fuhlendorff, J. et al. *Biochemistry* 1987, **26:** 6757–6764. Jensen, B. A., et al. *J. Lab. Clin. Med.* 1987, **110:** 612–617. Jensen, B. A. and Clemmensen, I. *Cancer* 1988, **62:** 869–872. Christensen, L. and Clemmensen, I. *Histochemistry* 1989, **92:** 29–35. Borregård, N., et al. *J. Clin. Invest.* 1990, **85:** 408–416. Christensen, L. and Clemmensen, I. *Histochemistry* 1991, **95:** 427–433. Clemmensen, I., et al. *Eur. J. Biochem.* 1991, **195:** 735–741.
Ref. for DNA/AA Sequences	

Thioltransferase

John J. Mieyal

Synonyms	Glutaredoxin
Abbreviations	TTase; Grx
Classifications	EC 1.8.4.2; Thiol:disulfide oxidoreductase
Description	A cytosolic enzyme isolated from human red blood cells and expected to be a ubiquitous protein found in most cells. Human thioltransferase also has been produced as a recombinant enzyme in E. Coli and shown to have properties identical to the natural enzyme. According to its properties, this human thioltransferase is classified as a thiol-disulfide oxidoreductase. Thus, the enzyme has a dithiol (Cys-Pro-Tyr-Cys) active site and catalyzes thiol-disulfide exchange reactions. Thioltransferase displays a rather strict substrate specificity for glutathione-containing mixed disulfides and its turnover is coupled to glutathione (GSH) and glutathione disulfide reductase, in contrast to the functionally related enzyme thioredoxin.
Structure	The tertiary structure has not been determined for the human enzyme; however, by analogy to pig liver thioltransferase, E. coli glutaredoxin, E. coli. thioredoxin and human thioredoxin, human thioltransferase is expected to be a globular protein comprised of helical and beta-sheet domains. Although prediction of structure from the linear aa sequence is consistent with this assessment, actual structure determination awaits NMR and/or x-ray analysis analogous to pig liver thioltransferase.
Molecular Weight	11,688 Da (according to aa sequence). SDS-PAGE and gel filtration analyses are consistent with this molecular weight value and indicate that the native protein is monomeric.
Sedimentation Coeff.	Unknown
Isoelectric Point	7.1: native protein, reduced; 7.5: oxidized
Extinction Coeff.	Unknown
Enzyme Activity	Catalysis of glutathione-dependent reduction of disulfides, including protein mixed disulfides. Recent evidence indicates a strict preference for glutathione-containing mixed disulfides. Two substrate kinetic studies are consistent with a catalytic mechanism that involves a TTase-S-S-Glutathione intermediate.
Coenzymes/Cofactors	Enzymatic turnover typically is coupled to glutathione disulfide (GSSG) reductase and NADPH. GSH-dependent, TTase-catalyzed reduction of disulfides produces GSSG as the accompanying product; the GSSG is reduced by GSSG reductase with stoichiometric oxidation of NADPH.
Substrates	Glutathione is the typical reduced substrate, and glutathione-containing mixed disulfides like cysteine-SSG and albumin-SSG are typical oxidized substrates. Formation of GSSG from GSH-dependent reduction of non-glutathione-containing mixed disulfides is catalyzed by TTase after non-enzymatic reaction of GSH with the disulfide to form a GS-containing disulfide substrate for the enzymatic reaction. Thus, the enzyme activity can be assayed conveniently by using the prototype "pro-substrate"

	hydroxyethyldisulfide or preferably cysteine-SSG, along with GSH, GSSG reductase, and NADPH, and monitoring the time dependent change in A_{340nm}. An alternative assay of the TTase activity involves measuring the GSH- and TTase-dependent release of trichloroacetic acid soluble radioactivity from albumin-SSG [^{35}S].
Inhibitors	Thioltransferase is inactivated in a concentration and time dependent manner by iodoacetamide. The pH dependence and stoichiometry of incorporation of radioactivity from [^{14}C]-iodoacetamide indicate that inactivation is associated with covalent modification of one of the active site cysteine residues with an apparent pKa of 3.5.
Biological Functions	Because of its disulfide reduction activity, TTase likely contributes to maintenance of the reduced status of sulfhydryl moeities on other proteins and small molecules. Because TTase catalyzes reversible transfer of the glutathione moiety, it is conceivable that TTase could function as a mediator of signal transduction by modulating the activity of redox-sensitive sulfhydryl proteins via glutathionylation.
Physiology/Pathology	Thioltransferase is expected to play a role in homeostatic maintenance of the sulfhydryl moeities of enzymes, receptors, and membrane proteins during oxidative stress.
Degradation	Unknown
Genetics/Abnormalities	Unknown
Half-life	Unknown
Concentration	Estimated to be about 1 μM in human erythrocytes.
Isolation Method	I. Natural Enzyme: (a) Lysis of red blood cells and differential solvent precipitation to remove hemoglobin selectively. (b) Gel filtration, then Phenyl-Sepharose reverse phase chromatography involving ammonium sulfate and ethylene glycol gradients. (c) Gel filtration, then Hydroxylapatite chromatography. (d) SDS PAGE and HPLC analysis of purity. II. Recombinant Enzyme: (a) PCR cloning of TTase from human brain cDNA. (b) Lysis of E. coli cells containing hTTase expressed at high levels. (c) Conventional gel filtration chromatography. (d) Gel filtration HPLC.
Amino Acid Sequence	105 aa starting with N-acetyl-Ala like other mammalian thioltransferases (glutaredoxins). The sequence of hRBC TTase is highly homologous (69-82%) to those of the other mammalian TTases that have been sequenced. Remarkable differences include the absence of any Met residues in hRBC TTase and the presence of a fifth Cys residue in the N-terminus preceding the active site Cys-Pro-Tyr-Cys sequence. DNA sequence: Analysis of the human thioltransferase cDNA insert in plasmid PCC-TT2 gave a sequence of codons that corresponded exactly to the aa sequence that was determined from pure hRBC TTase.
Disulfides/SH-Groups	Five cysteine residues (Cys 7, 21, 25, 78, 82). Cys-21 and Cys-25 can be cycled between thiol and intramolecular disulfide forms as indicated by reactivity with iodoacetamide. The redox behavior of the other Cys residues has not been determined definitively.

General References	Mieyal, J.J. et al. *Biochemistry* 1991, **30**:6082-6097. Mieyal, J.J. et al. *Biochemistry* 1991, **30**:8883-8891. Gravina, S.A. and Mieyal, J.J. *Biochemistry* 1993, **32**:3368-3376. Mieyal, J.J. et al. Glutathionyl specificity of thioltransferases: mechanistic and physiological implications. In: *Biothiols in Health and Disease.* Packer, L. and Cadenas, E. (eds.) Marcel Dekker, Inc. New York, 1995, **Chapter 14**, pp 305-372.
Ref. for DNA/AA Sequences	For aa sequence: Papov, V.V. et al. *Protein Science* 1994, **3**:428-434. for recombinant enzyme, and cDNA sequence: Chrestensen, C.A., Eckman, C.B., Starke, D.W. and Mieyal, J.J. *FEBS Lett.* 1995, **134**:25-28.

Thrombomodulin

William A. Dittman

Synonyms	Fetomodulin
Abbreviations	Tm
Classifications	
Description	Thrombomodulin is a cell surface trans-membrane glycoprotein, initially described on endothelial cells, but also present on platelets. A smaller, functionally active form of thrombomodulin circulates in the plasma and is also found in urine.
Structure	Thrombomodulin consists of an amino-terminal domain with weak lectin-like homology, six epidermal growth factor-like repeats, a short serine and theronine-rich region, a presumed transmembrane domain, and a C-terminal short cytoplasmic tail.
Molecular Weight	Tissue: 105,000 (SDS-PAGE, reduced); 75,000 (SDS-PAGE, unreduced). Plasma: 85,000 (SDS-PAGE, reduced).
Sedimentation Coeff.	Unknown
Isoelectric Point	4.2
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	Thrombin (Kd 0.5 nM), protein C (Km 5 μM). When bound to thrombomodulin thrombin can much more efficiently activate protein C.
Substrates	The thrombin-thrombomodulin complex proteolytically activates protein C. Thrombomodulin may be assayed by determination of activation of protein C with chromgenic substrates including S-2238, S-2266.
Inhibitors	Tumor necrosis factor, Interleukin-1, endotoxin, and phorbol myristate acetate will decrease cell surface thrombomodulin activity.
Biological Functions	Thrombomodulin functions as a natural anticoagulant by accelarating the activation of protein C by thrombin. Protein C, in turn, inhibits the coagulation cascade by inactivating factors Va and VIIIa. At the same time thrombin bound to thrombomodulin is less efficient as a procoagulant, with decreased ability to activate platelets, factors V and VIII, factor XIII, and cleave fibrinogen to fibrin.
Physiology/Pathology	There are no known abnormalities of thrombomodulin, although abnormalities of protein C and protein S result in hypercoagulable states, and presumably so would thrombomodulin deficiency.
Degradation	After endocytosis of cell surface thrombomodulin, in response to thrombin, or TNF, thrombomodulin is transported to the lysozymes, where some is degraded. This endocytosis in response to thrombin is inhibited by protein C but not activated protein C. Since the gene for thrombomodulin is single and intronless, the circulating form presumably is a degraded form of the membrane molecule, with loss of the trans-

	membrane domain and cytoplasmic tail. The fate of the circulating form is unknown.
Genetics/Abnormalities	Thrombomodulin is an intronless gene found on human chromosome 20p12-cen. A single species of mRNA is known, with a 5' untranslated region of approximately 158 bases, followed by 1725 bases of coding sequence, encoding 575 amino acids, and 1779 bases of 3' untranslated sequence. There are no known genetic abnormalities.
Half-life	Unknown
Concentration	There are approximately 100,000 TM molecules per endothelial cell. In the murine system the protein has a half-life of approxiamtely 18 hours. The plasma form circulates at approximately 20 μg/L. Its half-life is unknown.
Isolation Method	Human thrombomodulin has been isolated from lung, plasma, platelets and placenta, although the placenta is the most accesible rich source of TM. A detergent solubilized extract of membranes can be applied to an inactivated-thrombin affinity column, the thrombomodulin eluted, applied to an immunoaffinity column of polyclonal rabbit anti-human thrombomodulin IgG. The eluted protein is primarily thrombomodulin. Anion exchange chromatography has also been used in addition to, or in place of immunoaffinity chromatography.
Amino Acid Sequence	The amino-terminal 21 aa of thrombomodulin are hydrophobic, with characteristics of a typical signal peptide. The following 223 aa have weak homology to lectin-like molecules, but has no knwon lectin activity. This region is followed by six epidermal growth factor-like repeats; the fifth and sixth are sufficient for thrombin binding, while the fourth is also required for protein C activation. There are five potential N-glycosylation sites (N-X-S/T). A region of potential O-linked glycosylation follows the EGF-like region. This is followed by a hydrophobic presumed transmembrane region of 23 aa, followed by a 38 aa cytoplasmic tail.
Disulfides/SH-Groups	Thrombomodulin contains 49 cysteines; the secondary structure is not known, but the EGF-like repeats presumably form typical EGF disulfides.
General References	Dittman, W. A., and Majerus, P. W. *Blood* 1990, **75:** 329–336. Esmon, C. T. *J. Biol. Chem.* 1989, **264:** 4743–4746. Esmon, C. T. *Science* 1987, **235:** 1348–1352.
Ref. for DNA/AA Sequences	Wen, D. et al. *Biochemistry* 1987, **26:** 4350–4357. Suzuki, K. et al. *EMBO J.* 1987, **6:** 1891–1897.

Thrombospondin-1

Jack Lawler

Synonyms	Thrombin-sensitive protein; Platelet glycoprotein G
Abbreviations	TSP-1; TSP; TS
Classifications	Adhesive glycoprotein
Description	Thrombospondin-1 is a 140,000 Da polypeptide. It assembles into a trimer with a molecular weight of 420,000. The protein is synthesized and secreted by many cells in culture. It represents approximately 1% of the total proteins in human blood platelets. In terms of the structural and functional properties of thrombospondin-1, it is similiar to the adhesive glycoproteins and extracellular matrix components. It is unique in that it can bind large amounts of calcium ions.
Structure	Thrombospondin-1 and 2 form disulfide bonded homo- and heterotrimers. In electron microscopic images, each subunit is approximately 54 nm long and contains globular domains at the N- and C-termini that are 7 and 11.8 nm in diameter, respectively. The globular domains are connected by a region of protein that appears thin and flexible. The three subunits are connected to each other near the N-terminal domain. The electron microscopic appearance of thrombospondin-1 is affected by the removal of calcium. The globular regions at the C-termini decrease in size to 8 nm and the thin, connecting regions increase in length by approximately 32%.
Molecular Weight	420,000: intact protein (sedimentation equilibrium); 136,000: subunit, reduced (sedimentation equilibrium). The reduced subunit migrates anomalously on SDS-PAGE using the discontinous system of Laemmli giving a value of 185,000 Da.
Sedimentation Coeff.	9.7 S (in calcium); 8.6 (in EDTA)
Isoelectric Point	4.7
Extinction Coeff.	9.5 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Thrombospondin-1 has the ability to bind to heparin, calcium, plasmin, other matrix components and cell surfaces. Proteoglycans, sulfatides, integrins, CD36, and other proteins have been reported to function as cell surface receptors for thrombospondin-1. Through these various interactions, thrombospondin-1 modulates cell shape, migration and proliferation. Thrombospondin-1 may also serve to sequester calcium in a readily mobilizable form at the cell surface.
Physiology/Pathology	The physiological role of thrombospondin-1 is unknown. It has been proposed to be involved in platelet aggregation, tumor cell metastasis,

neutrophil migration and tissue genesis and remodeling. Thrombospondin-1 and a naturally occurring proteolytic fragment inhibit angiogenesis.

Degradation

The N-terminal domain (25,000-30,000 daltons) can be cleaved from the molecule by trypsin, chymotrypsin, thrombin and plasmin. This degradation can occur to a variable degree in platelets and during purification. The molecule becomes much more labile to proteolysis when calcium is removed. In the absence of calcium, a 210,000-dalton fragment (that is composed of three 70,000-dalton polypeptides derived from the center of each subunit) is a major fragment that is produced.

Genetics/Abnormalities

The location of the thrombospondin-1, 2 and 3 genes in the human and the mouse are indicated in the table below. No abnormal forms of these genes have been reported to date. A dinucleotide (CT) repeat polymorphism at the human thrombospondin-1 gene has been reported (*Nucleic Acids Research* **18**:7467).

Gene Location	THBS1	THBS2	THBS3
Human	15q15	6q27	1q21-24
Mouse	2 band F	17 band A	3 band E_3 - F_1

Half-life

Unknown

Concentration

Blood platelets: 25-89 ng/10^6 platelets
Human plasma: 97-163 $\mu g\ L^{-1}$
Human serum: 15-18 $mg\ L^{-1}$
Human colostrum: 70-88 $mg\ L^{-1}$
Human milk: 4 $mg\ L^{-1}$
Human umbilical vein endothelial cells in culture: 9-49 $\mu g/10^6$ cells/24 hrs

Isolation Method

Thrombospondin-1 can be purified from the supernatant that is produced by treating human blood platelets with thrombin or the ionophor A23187. Due to the protein's high molecular weight, gel filtration chromatography on Sepharose 4B, FPLC or sucrose density gradient centrifugation can be used to separate thrombospondin from many of the other secreted proteins. Affinity chromatography on heparin-Sepharose is frequently used. Thrombospondin-1 contains a free sulfhydryl group that is protected by including 1 to 2 mM calcium during the purification. If the calcium concentration becomes too low, the sulfhydryl group will react with disulfide bonds within thrombospondin-1 and other molecules. These changes are not reversed by adding calcium back to the purified protein. Thus, unless the application does not require the preservation of native structure, methods that use EDTA elution should be avoided. These include barium citrate and integrin beta-1 sbunit cytoplasmic domain peptide chromatography.

Amino Acid Sequence

Thrombospondin-1 contains sequences of approximately 200 aa at the N- and C-termini that do not have strong homology to other proteins and do not have internal repeating motifs. The sequences RKGSGRR and MKKTRG in the N-terminal domain have been reported to be involved in heparin binding. The center portion of the protein is composed of repeated sequence motifs that are homologous to other proteins. Amino acids 263 - 360 are homologous to procollagen. Amino acids 361 - 530 comprise three copies of the type 1 repeat. These repeats are found in proteins from the malaria parasite, the complement factors and the axon guidance proteins UNC-5 and F-spondin. These proteins, along with thrombospondin-1 and 2 comprise the thrombospondin type-1 repeat (TSR) supergene family. The type 1 repeats contain the cell binding sequence VTCG as well as unusual heparin-binding motifs. Amino acids 531-673 comprise three copies of the EGF repeat found in many cell adhesion molecules and adhesive glycopro-

teins. Amino acids 698-925 represent seven copies of the thrombospondin type 3 repeat. These sequence motifs are calcium binding sites. The last type 3 repeat of thrombospondin-1 contains the cell adhesion sequence RGDA.

Disulfides/SH-Groups

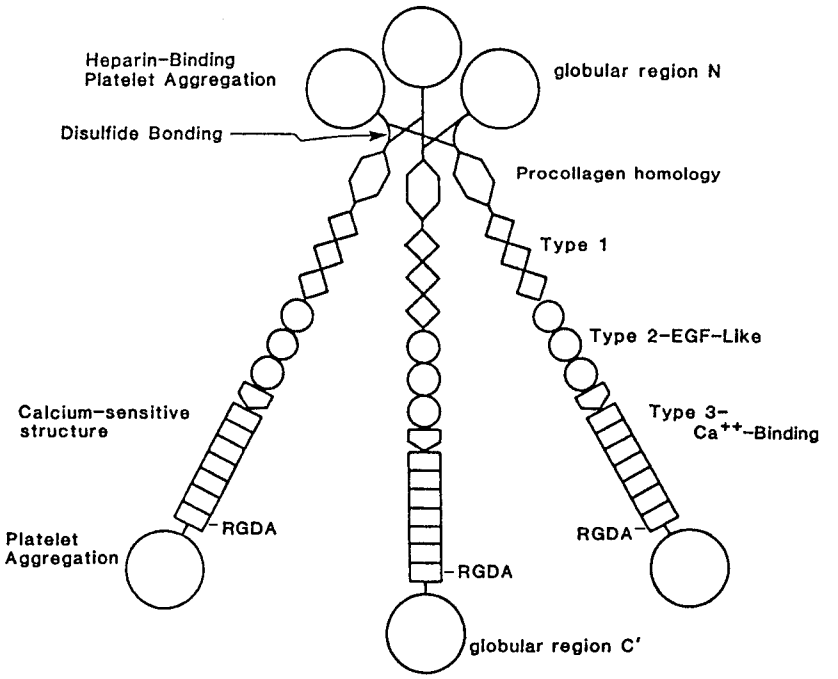
Cys-252 and Cys-256 are the interchain disulfides. Most of the other Cys residues form disulfide bonds; however, the specific pattern of disulfide bonds is not known. Thrombospondin-1 has 69 Cys residues indicating that one of these residues exists as a free sulfhydryl. The protein has been reported to undergo a conformational change that results in the rearrangement of the free sulfhydryl group to several different positions.

General References

Adams, J. and Lawler, J. *Current Biology* 1993, **3**:188-190.
Bornstein, P. *FASEB J.* 1992, **6**:3290-3299.
Frazier, W.A. *Curr. Op. Cell Biol.* 1991, **3**:792-799.
Adams, J.C. Tucker, R.P., Lawler, J. The thrombospondin gene family. R.G. Landes, Austin, 1995.

Ref. for DNA/AA Sequences

Lawler, J. and Hynes, R.O. *J. Cell Biol.* 1986, **103**:1635-1648 (GenBank Accession number X04665, human thrombospondin-1).
Lawler, J. et al. *Genomics* 1991, **11**:587-600 (GenBank Accession numbers M62461-M62470, mouse thrombospondin-1 gene).
Laherty, C.D. et al. *J. Biol. Chem.* 1989, **264**:11222-11227. (GenBank Accession number J04835, human thrombspondin-1 promoter).
Donoviel, D.B. et al. *J. Biol. Chem.* 1988, **263**:18590-18593 (GenBank Accession number J04447, human thrombospondin-1 promoter).

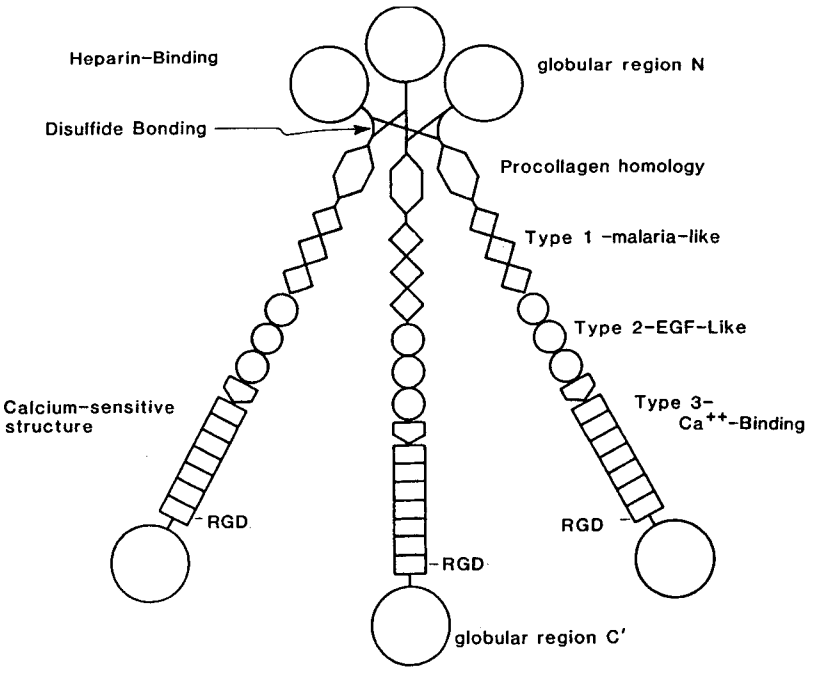


Thrombospondin-2

Jack Lawler

Synonyms	None
Abbreviations	TSP-2
Classifications	Extracellular calcium-binding protein
Description	Thrombospondin-2 is a high molecular weight secreted glycoprotein that binds heparin, calcium and cell surfaces. It is expressed in many tissues during mouse and chicken development including: perichondrium, blood vessels, embryonic mesenchyme, bone, skin, skeletal and cardiac muscle, brain, kidney, lung and gut. The presence of RGD and VTCG cell-binding sequence motifs suggests that thrombospondin-2 may interact with the same cell surface receptors as thrombospondin-1.
Structure	The electrophoretic mobility of the reduced subunit of thrombospondin-2 is slightly less than that of thrombospondin-1. Disulfide bonded homo- and heterotrimers are formed from the thrombospondin-1 and -2 polypeptides. Whereas the thrombospondin-2 homotrimers have not been visualized by electron microscopy, the similarity of the aa sequences of thrombospondin-1 and -2 suggests that their tertiary structures are similar.
Molecular Weight	450,000: intact protein by SDS-PAGE; 185,000–190,000: subunit by SDS-PAGE; 129,954: subunit based on aa sequence.
Sedimentation Coeff.	Unknown
Isoelectric Point	4.5 (calculated value from aa composition)
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Thrombospondin-2 has the ability to bind to heparin and calcium. The thrombospondin-2 homo-trimer is eluted from heparin-Sepharose with 0.35–0.5 M NaCl suggesting that thrombospondin-2 has a lower affinity for heparin than does thrombospondin-1. Peptides that are based on the NH_2-terminal heparin-binding domains of thrombospondin-1 and -2 have the ability to disrupt focal adhesion contacts formed by endothelial cells and fibroblasts. The presence of an RGD sequence in human, mouse and chicken thrombospondin-2 suggests that it can interact with integrin receptors on cell surfaces. The potential to bind 36 calcium ions suggests that thrombospondin-2 functions as a source of readily mobilizable calcium at the cell surface.
Physiology/Pathology	Unknown
Degradation	Unknown

Genetics/Abnormalities	The human thrombospondin-2 gene is located on chromosome 6q27 and the mouse gene is located on chromosome 17 band A. The thrombospondin-1 and -2 genes are very similar in terms of sizes of the exons and the patterns of interruption of the reading frames by introns.
Half-life	Unknown
Concentration	Unknown
Isolation Method	Unknown
Amino Acid Sequence	The human thrombospondin-2 sequence has an open reading frame that encodes a 1172 aa peptide. The first 18 aa comprise a signal sequence consistent with the fact that thrombospondin-2 is secreted from cultured cells. Human thrombospondin-1 and -2 are 54% identical on an aa level. The greatest number of differences exist in the NH_2-terminal heparin-binding domain. Like thrombospondin-1, thrombospondin-2 has regions of aa sequence that are approximately 200 aa long at the NH_2- and COOH-terminals. The center of the polypeptides contains a region of homology with procollagen, three type 1 repeats, three type 2 or epidermal growth factor-like repeats, and seven type 3 calcium-binding repeats. The type 1 repeats include the VTCG cell-binding motif.
Disulfides/SH-Groups	Cys-266 and Cys-270 (numbering from the initiating methionine of the human sequence) have been identified as forming the interchain disulfide bonds based on homology to thrombospondin-1 and the fact that thrombospondin-2 can form heterotrimers with thrombospondin-1.
General References	Bornstein, P. et al. *Proc. Natl. Acad. Sci. U. S. A.* 1991, **88**:8636–8640. Chen, M. et al. *J. Biol. Chem.* 1994, **269**:32226–32232. Tucker, R. P. *Development* 1993, **117**:347–358. Iruela-Arispe, M. L. et al. *Developmental Dynamics* 1993, **197**:40–56. O'Rourke K. M. et al. *J. Biol. Chem.* 1992, **267**:24921–24924. Murphy-Ullrich, J. E. et al. *J. Biol. Chem.* 1992, **268**:26784–26789.
Ref. for DNA/AA Sequences	Bornstein, P. et al. *J. Biol. Chem.* 1991, **266**:12821–12824. GenBank Accession number M64866, Mouse TSP-2 sequence. Laherty, C. D. et al. *J. Biol. Chem.* 1992, **267**:3274–3281. GenBank Accession number A42587, mouse TSP-2 sequence. Shigu, T. and Bronstein, P. *Genomics* 1993, **16**:78–94. GenBank Accession number L06421 and L06422, mouse genomic TSP-2 sequence. Lawler, J. et al. *J. Biol. Chem.* 1991, **266**:8039–8043. GenBank Accession number M60853, chicken TSP-2 sequence. LaBell, T. L. and Byers, P. H. *Genomics* 1993, **17**:225–229. GenBank Accession number L12350, human TSP-2 sequence.


Heparin-Binding
globular region N
Disulfide Bonding
Procollagen homology
Type 1 -malaria-like
Type 2-EGF-Like
Calcium-sensitive structure
Type 3- Ca++-Binding
RGD
RGD
RGD
globular region C'

Thrombospondin-3

Jack Lawler

Synonyms	None
Abbreviations	TSP-3
Classifications	Extracellular calcium-binding protein
Description	Thrombospondin-3 was identified as a gene that is (1) immediately upstream of the epithelial mucin *Muc*1 (episialin) gene and (2) homologous to thrombospondin-1 and 2. The structure of thrombospondin-3 suggests that it is a cell surface and matrix-associated calcium-binding protein. Thrombospondin-3 is expressed in the lung, skeletal muscle, forming bone, intestine, spinal cord and brain of developing mouse embryo. The highest levels of expression are observed late in murine development (day 16–19).
Structure	The naturally occurring form of thromospondin-3 has not been purified. Electron microscopy and SDS-PAGE of the recombinant protein as it is expressed in the human renal epithelial cell line 293T indicates that thrombospondin-3 is a pentamer. Thrombospondin-3 is indistinguishable from thrombospondin-4 by electron microscopy. Globular domains are present at the ends of thin-connecting regions. These globular domains decrease in size and the thin-connecting regions increase in length when calcium is removed from the molecule. A globular domain is also present near the site where the five subunits are connected to each other.
Molecular Weight	140,000–150,000: subunit by SDS-PAGE; 103,972: subunits based on aa sequence.
Sedimentation Coeff.	Unknown
Isoelectric Point	4.2 (calculated value from aa composition)
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Recombinant thrombospondin-3 binds heparin and calcium. Thrombospondin-3 can be eluted from heparin-Sepharose with buffer containing 0.3–0.35 M NaCl suggesting that it has a lower affinity for heparin than thrombospondin-1. Since each molecule may be able to bind 60 calcium ion, the function of thrombospondin-3 may be to sequester extracellular calcium.
Physiology/Pathology	Unknown
Degradation	Unknown
Genetics/Abnormalities	The human thrombospondin-3 gene is located on chromosome 1q21–24. The mouse gene is located on mouse chromosome 3 band E3-F1. In both

genomes the thrombospondin-3 gene is immediately upstream of the gene for the epithelial mucin *Muc*1.

Half-life

Unknown

Concentration

Unknown

Isolation Method

Unknown

Amino Acid Sequence

The open reading frame for mouse thrombospondin-3 is 956 aa long. The first 21–23 aa appear to comprise a signal sequence. Like thrombospondin-4 and cartilage oligomeric matrix protein, thrombospondin-3 has four type 2 or EGF-like repeats and lacks the region of homology with procollagen and the type 1 repeats that are found in thrombospondin-1 and -2. Thrombospondin-3 contains seven type 3 repeats, which are highly acidic and homologous to the calcium-binding pockets of calmodulin and parvalbumin. The NH_2- and COOH-terminal regions are each approximately 200 aa long and do not display internal repeating sequence motifs or significant homology to proteins that are not members of the thrombospondin gene family.

Disulfides/SH-Groups

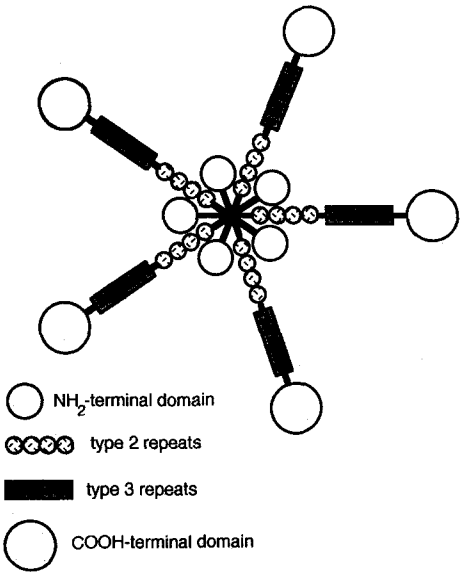
The electron microscopic appearance of thrombospondin-3 and its similarity to thrombospondin-1 suggest that Cys-266 and Cys-269 (numbered from the initiating methionine of the mouse sequence) participate in the formation of the interchain disulfide bonds, Since thrombospondin-3 has an odd number of cysteine residues, it probably has at least one free sulfhydryl group.

General References

Iruela-Aripse, M. L. et al. *Developmental Dynamics* 1993, **197**:40–56.

Ref. for DNA/AA Sequences

Vos, H. L. et al. *J. Biol. Chem.* 1992, **267**:12192–12196. (GenBank Accession numbers M86611-M86620, mouse sequence).
Bornstein, P. et al. *Genomics* 1993, **15**:607–613. (GenBank Accession number L04302, mouse sequence).
Qabar, A. N. *J. Biol. Chem.* 1994, **269**:1262–1269. (GenBank Accession number L24434, mouse sequence).

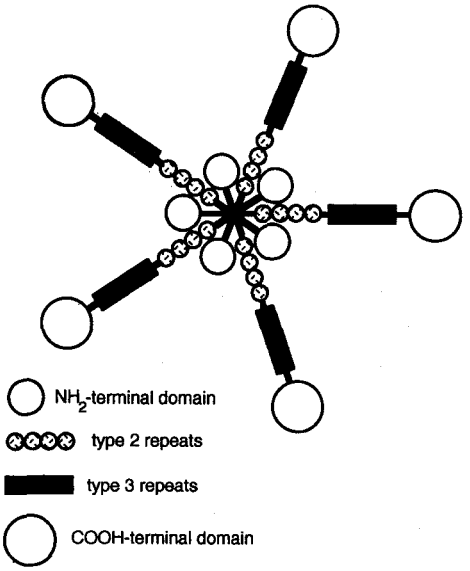


Thrombospondin-4

Jack Lawler

Synonyms	None
Abbreviations	TSP-4
Classifications	Extracellular calcium-binding protein
Description	Thrombospondin-4 was identified as a cDNA with sequence similarity to thrombospondin-1. The structure of thronbospondin-4 suggests that it is a cell surface and matrix-associated calcium-binding protein. It is expressed in cardiac and skeletal muscle in adult human tissue and in the frog embryo. The mRNA is also detected in developing bone of chicken embryos.
Structure	The natrually occurring form of thrombospondin-4 has not been purified. Electron microscopy and SDS-PAGE of the recombinant protein as it is expressed in NIH 3T3 cells indicates that it is a pentamer. Like thrombospondin-1, globular structures are present at both ends of each subunit and the conformation of the molecule is altered when calcium is removed. The region of polypeptide that connects the globular domains is approximately 28 nm long, after treatment with EDTA.
Molecular Weight	520,000: intact protein by SDS-PAGE; 140,000: subunit by SDS-PAGE; 106,326: subunit based on aa sequence
Sedimentation Coeff.	Unknown
Isoelectric Point	4.3 (calculated value from aa composition)
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Recombinant thrombospondin-4 binds heparin and calcium. A bacterial fusion protein that includes the COOH-terminal (residues 713–940) has been shown to support attachment of the C2C12 skeletal myoblast cell line. Whereas the human and mouse thrombospondin-4 sequences include the RGD cell adhesion sequence, the frog sequence does not. Since each subunit appears to have 12 calcium-binding sites, the intact pentamer may bind 60 calcium ions. The function of thrombospondin-4 may be to sequester calcium in a readily mobilizable form at the cell surface.
Physiology/Pathology	Unknown
Degradation	Unknown
Genetics/Abnormalities	The human thrombospondin-4 gene is located on chromosome 5q13. The mouse gene is located on chromosome 13,52 cM from the centromere.

Half-life	Unknown
Concentration	Unknown
Isolation Method	Unknown
Amino Acid Sequence	The open reading frame for human thrombospondin-4 is 961 aa long. The first 21 aa appear to comprise a signal sequence. The NH_2- and COOH-terminal regions are each composed of approximately 200 aa that do not show homology to other proteins (besides other thrombospondins) and do not have repeated sequences. Two types of repeating sequences, designated the thrombospondin type 2 and type 3 repeats can be identified. There are four adjacent type 2 or epidermal growth factor-like repeats. The consensus sequence and overall organization of the type 3 repeats is highly conserved among the members of the thrombospondin gene family, with the second and fourth type 3 repeats being truncated after the second cysteine. The type 3 repeats are homologous to the calcium-binding pockets of calmodulin and parvalbumin. The RGD sequence is present in the third type 3 repeat of the human and mouse sequences.
Disulfides/SH-Groups	The electron microscopic appearance of thrombospondin-4 and its similarity to thrombospondin-3 suggests that Cys-258 and Cys-261 (numbering from the initiating methionine of the human sequence) participate in the formation of the interchain disulfide bonds. Since thrombospondin-4 has an odd number of cysteine residues, it probably has at least one free sulfhydryl group.
General References	Adams, J. C. and Lawler, J. *Mol. Biol. Cell* 1994, **5**:423–437. Adams, J. C. and Lawler, J. *Current Biology* 1993, **3**:188–190.
Ref. for DNA/AA Sequences	Lawler, J. et al. *J. Cell Biol.* 1993, **120**:1059–1067 (GenBank Accession number Z19091, *Xenopus laevis* sequence). Lawler, J. et al. *J. Mol. Evol.* 1993, **36**:509–516 (GenBank Accession number Z19585, human sequence). Lawler, J. et al. *J. Biol. Chem.* 1995, **270**:2809–2814 (GenBank Accession number Z19585, complete human sequence).



Thromboxane synthase

Lee-Ho Wang

Synonyms	None
Abbreviations	TXS
Classifications	EC 5.3.99.5
Description	TXS is characterized by absorption spectrum and DNA sequence to be a cytochrome P-450 enzyme with one heme per polypeptide. TXS catalyzes an isomerization rather than mono-oxygenation reaction. TXS undergoes suicide inactivation during catalysis. Txs is found in many tissues including lung, kidney, spleen, brain, gastrointestinal tracts and cell types such as platelets, lymphocytes, leukocytes and macrophages. TXS is a membrane-bound protein located at endoplasmic reticulum.
Structure	Unknown
Molecular Weight	60,684 (calculated from cDNA); 58.8 kDa (determined from SDS-PAGE)
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	TXS converts prostaglandin H_2 to thromboxane A_2. In addition, the enzyme can also convert prostaglandin H_2 to malondialdehyde and 12-L-hydroxy-5,8,10-heptadecatrienoic acid.
Coenzymes/Cofactors	TXS contains heme as the cofactor. The heme involves in the substrate binding.
Substrates	Prostaglandin H_2 (15-hydroperoxy-9,11-peroxidoprosta-5,13-dienoic acid)
Inhibitors	15(S)-hydroxy-9,11-epoxymethanoprosta-5,13(Z,E)-dienoic acid, sodium 3-(4-(1-imidazolylmethyl)phenyl)-2-propenoate
Biological Functions	Thromboxane A_2, the natural product of TXS, is a potent mediator of platelet aggregation and vasoconstriction.
Physiology/Pathology	TXA_2 is a crucial factor contributing to the pathophysiology of a variety of disease processes such as thrombosis, stroke and atherosclerosis.
Degradation	The biosynthesis of thromboxane A_2 is initiated by phospholipase which liberates arachidonic acid from membrane phospholipids. Arachidonic acid is then converted to prostaglandin H_2 by prostaglandin endoperoxide synthase and subsequently to thromboxane A_2 by TXS. Thromboxane A_2 is unstable (half-life is 30 sec at physiological conditions) and hydrolysed to biologically inactive thromboxane B_2.
Genetics/Abnormalities	Unknown
Half-life	26 hrs in human monocytes

Concentration	The concentration of TXS in human platelets, monocytes and lung are 2,187, 1,548 and 765 ng per mg microsomal protein.
Isolation Method	Mono Q high performance liquid chromatography; Immunoaffinity chromatography.
Amino Acid Sequence	The aa sequence of cysteine pocket in the TXS, FGAGPRSCLGV, is highly conserved on P-450s. The cysteine is thought to be the proximal ligand for the heme iron.
Disulfides/SH-Groups	Unknown
General References	Hecker, M. and Ullrich, V. *J. Biol. Chem.* 1989, **264:** 141–150. Jones, D. A. and Fitzpatrick, F. A. *J. Biol. Chem.* 1991, **266:** 23510–23514. Shen, R.-F. and Thai, H.-H. *J. Biol. Chem.* 1986, **261:** 11592–11599. Holman, R. T. et al. *Prog. Lipid Res.* 1986, **28:** 273–301.
Ref. for DNA Sequences	Ohashi, K. et al. *J. Biol. Chem.* 1992, **267:** 789–793. GenBank accession numbers: M80646 and M80647.

Thyroglobulin

John T. Dunn and Ann D. Dunn

Synonyms	Thyroid prohormone
Abbreviations	Tg
Classifications	None
Description	A large glycoprotein synthesized in the thyroid cell as a 330 kDa monomer of 2767 residues (including 19 residue leader). Further maturation includes glycosylation, iodination, dimerization to 660 kDa, and secretion into thyroid's follicular lumen to constitute bulk of colloid. About 10% carbohydrate in two major types of N-linked unit: a polymannose, MW $\simeq$ 1,800, 7–8 units/mol Tg; and a complex, MW 2,100–3,300, about 22 per mol Tg. Human Tg also contains an O-linked unit of MW 2,000–3,000 containing galactosamine, and a chondrotin sulfate, MW 5750. Most unique feature is iodine, 0.2–1% or 10–50 atoms per 660 kDa, of which about 30% is thyroxine, (T_4) or triiodothyronine (T_3), rest is 3-iodotyrosine (MIT), and 3,5-diiodotyrosine (DIT). Typical number residues of iodoamino acids for 660 kDa Tg, 0.5% iodine (25 atoms per molecule): T_4, 2; T_3, 0.3; DIT, 5; MIT, 6. Also contains phosphate, some as phosphoserine.
Structure	Ovoid, 300 × 150 Å, 2 symmetric halves (= monomers).
Molecular Weight	660,000 for dimer, the usual mature form (by sedimentation and aa composition).
Sedimentation Coeff.	19S (dimer), 12S (monomer), 27S, 33S (polymers)
Isoelectric Point	4.4–4.7
Extinction Coeff.	10 (280 nm, 1%, 1 cm), varies with iodine content.
Enzyme Activity	Unknown
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Provides the matrix for formation of thyroid hormones and their subsequent storage. Tg is first iodinated to form MIT and DIT at certain tyrosyls. T_4 forms by coupling of the diiodophenyl ring of a donor tyrosyl to an acceptor diiodotyrosyl. Of the 67 tyrosyls per 330 kDa chain, 4 (sites A–D) are major hormonogenic (acceptor) sites; minor hormonogenic sites also exist. The donor iodotyrosyls are not established, but tyrosyls 130, 1447, and 2553 have been suggested. T_3 formation is favored at site C. TSH, iodine, and animal species affect priority of utilization among hormonogenic sites. Thyroperoxidase catalyzes iodination of tyrosyls and coupling to form hormones. Hormone-containing Tg is stored in follicular lumen, principally as a 19S dimer.
Physiology/Pathology	Essential for thyroid hormone synthesis. Goiter and hypothyroidism can develop from decreased or absent Tg production, or from altered chemical structure. Serum Tg useful as marker for differentiated thyroid cancer.

Degradation	Digested in thyroidal lysosomes by cathepsins D, B, and L, each with different cleavage sites. Further degradation by exopeptidases, particularly lysosomal dipeptidase I. T_4 and T_3 released into circulation, but DIT and MIT deiodinated and their iodine recycled within thyroid.
Genetics/Abnormalities	Gene on long arm of 8q24 distal to c-myc oncogene linked with genes for carbonic anhydrase II and MOS protooncogene, consists of more than 260 kb, but less than 10% are in exons. Most mRNA is 8.5 kb, but small amount of $\simeq$ 1 kb also present. Goiters are associated with abnormal Tg structure, including alterations in tertiary structure, iodotyrosyl coupling, glycosylation, immunoreactivity, and composition of iodine, carbohydrates, and amino acids. A splicing error causing omission of exon 4 has recently been described in a familial goiter (J. Clin. Invest. 1991, **88:** 1901–1905).
Half-life	$\simeq$ 30 hrs in serum
Concentration	5 μg/L (serum); 75 mg/ml (thyroid colloid).
Isolation Method	Isolated from supernatant of centrifuged thyroid hormonogenate by gel filtration (e.g., BioGel A-5m, $K_{av} = 0.5$), by salt fractionation (precipitates at 38% saturated $(NH_4)_2SO_4$), or by density gradient ultracentrifugation (19S).
Amino Acid Sequence	cDNA sequence known. Hormonogenic sites: A (Y = residue #5), NIFEYQVDAQ; B (Y = 2553), HSTDDYASFS; C (Y = 2746), QEPGSKTYSK; D (Y = 1290), MCSADYAGLL. Consensus sequences for early iodination: (1) E/D-Y; (2) S/T-Y-S; (3) E-X-Y. First 1177 residues have internal homology with C-W/Y-C-V-D motif repeated 10 times. C-terminal 540 residues homologous (28%) with acetylcholinesterase.
Disulfides/SH-Groups	4.5% of residues are C, of which only 1–6 are free sulfhydryls, number decreasing inversely with iodine content. Positions of disulfides and intrachain and interchain distribution not known.
General References	Dunn, J. T. Thyroglobulin: chemistry and biosynthesis. In: *The Thyroid*, Braverman, L. E. and Utiger, R. D. (eds.) 6th edition. J. B. Lippincott Co., Philadelphia 1991 pp. 98–110. *Thyroglobuin – The Prothyroid Hormone*. Eggo, M. C. and Burrow, G. N. (eds.) Raven Press, New York 1985. Malthiery, Y., et al. *Eur. J. Biochem.* 1989, **71:** 195–210.
Ref. for DNA/AA Sequences	Malthiery, Y. and Lissitzky, S. *Eur. J. Biochem.* 1987, **165:** 491–498.

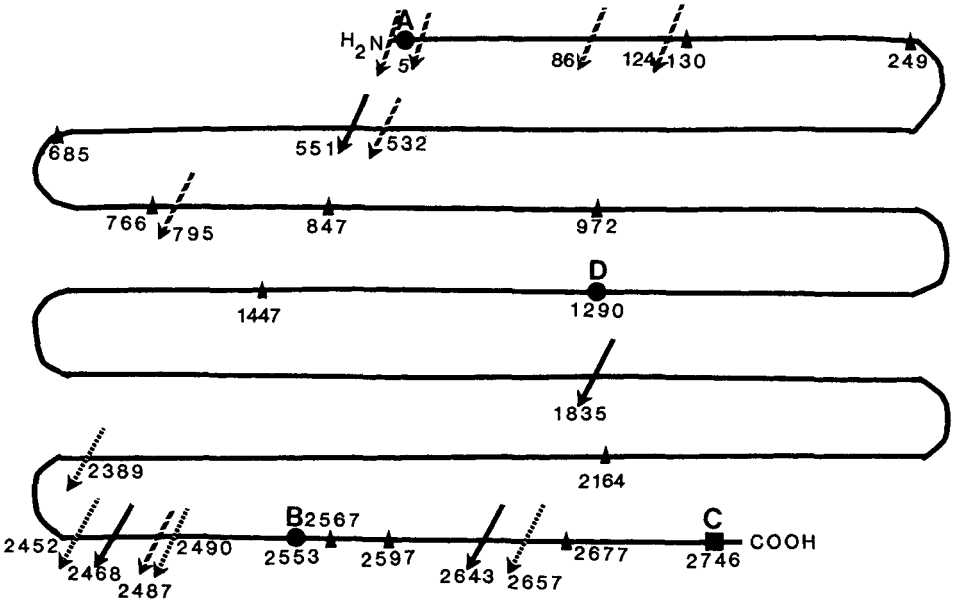


Diagram of Tg polypeptide chain showing major T_4-(●) and T_3-(■) forming sites (A–D), early iodination sites (▲), and cleavage points for cathepsins L (↙), B (↙), and D (↙).

Thyrotropin

James Magner

Synonyms	Thyroid-stimulating hormone
Abbreviations	TSH
Classifications	None
Description	A circulating glycoprotein hormone synthesized in the anterior pituitary that binds to a specific receptor on thyroid follicular cells stimulating their function. TSH is a heterodimer composed of two noncovalently linked subunits, α and β, which are approximately 21% and 12% carbohydrate by weight, respectively.
Structure	The α-subunit has 92 aa and two asparagine-linked oligosaccharides. The β-subunit has 112 aa and one asparagine-linked oligosaccharide. There is substantial heterogeneity of the oligosaccharides; two to four antennae on each oligosaccharide may terminate with galactose-sialic acid or N-acetyl-galactosamine-sulfate.
Molecular Weight	Heterodimer: 28,000; α-subunit: 14,000; β-subunit: 14,000.
Sedimentation Coeff.	2.63 S
Isoelectric Point	8.6–6.0 (7 isoforms); α: 8.8–5.4; β: 8.7–5.8
Extinction Coeff.	9.9 (292 nm, 1%, 1 cm, 0.1N NaOH)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	TSH with altered oligosaccharides may act as a competitive inhibitor to TSH receptor binding, with increased binding affinity for the TSH receptor but lower biological potency as measured by cAMP generation.
Biological Functions	TSH stimulates thyroid function, causing serum levels of thyroxine (T_4) and triiodothyroine (T_3) to rise. Levels of T_3 within the pituitary then reduce the transcription of TSH subunits – a classic negative feedback endocrine system. The individual TSH subunits have little biological activity. TSH heterodimers activate the adenylate cyclase-cAMP system. Other second messengers may play lesser roles. Thyroid cells are stimulated to take up colloid; thyroglobulin is degraded and T_4 and T_3 are secreted. TSH also has dramatic effects on thyroid morphology and growth, and the metabolism of nucleic acids, proteins, carbohydrates, phospholipids and iodine.
Physiology/Pathology	Essential for the proper control of thyroid function. When primary thyroid disease occurs, serum levels of TSH rise. Rarely, pituitary diseases may cause a state of secondary hypothyroidism with low serum levels of thyroid hormone without substantial TSH elevation. Prolonged fasting or critical illness may induce a state in which TSH fails to rise in spite of low serum thyroid hormones ("euthyroid sick syndrome"). TSH may be elevated during the first 24 hours after birth, but then should normalize. Very rare

	pituitary tumors may secrete excess TSH. Thyrotropin-releasing hormone (TRH), a hypothalamic tripeptide, stimulates the release of TSH from the pituitary. There is a slight diurnal variation in TSH serum levels, with higher levels during early sleep.
Degradation	Eliminated from the circulation by the kidney and liver.
Genetics/Abnormalities	The gene for the α-subunit is located on chromosome 6, is 9.4 kb in length and contains three introns. The gene for the β-subunit is located on chromosome 1, is 4.9 kb in length and contains two introns. The DNA sequence suggests that the β-subunit has 118 aa, but only a 112 aa form has been found in man, perhaps due to proteolytic cleavage. Two Japanese kindreds with recessively inherited hypothyroidism were found to have a point mutation in exon 2 of the β-subunit that changed a glycine to an arginine. Two Greek families had mutations in the 12th amino acid of the β-subunit that altered glutamic acid to a stop codon.
Half-life	Mean $\simeq$ 30 min.
Concentration	Serum: 0.5–6.0 mU/L
Isolation Method	Extract from acetone-dried pituitaries with 2% NaCl, pH 7.0. Ion-exchange and/or immune affinity chromatography.
Amino Acid Sequence	The α-chains are identical for hTSH, LH, FSH and CG; the TSH β-chain confers biological and immunological specificity.
Disulfides/SH-Groups	The α-subunit has 10 and the β-subunit has 12 half-cystine residues; all form internal disulfide bonds.
General References	Magner, J. A., Thyroid-stimulating hormone: biosynthesis, cell biology and bioactivity. *Endocrine Reviews* 1990, **11:** 354–385. Shupnik, M. A., et al. Molecular biology of thyrotropin. *Endocrine Reviews* 1989, **10:** 459–475. Wondisford, F. E., et al. Chemistry and biosynthesis of thyrotropin. In: *The Thyroid*, Braverman, L. E. and Utiger, R. D. (eds.), Sixth Ed. Lippincott, 1991, p. 257–276.
Ref. for DNA/AA Sequences	Fiddes, J. and Goodman, H. *J. Mol. Appl. Genet.* 1981, **1:** 3–18. Tatsumi, K. et al. *Gene* 1988, **73:** 489–497. Guidon, P. T. et al. *DNA* 1988, **7:** 691–699. Wondisford, F. E., et al. *J. Biol. Chem.* 1988, **262:** 12538–12542.

Thyrotropin Receptor

Basil Rapoport, Gregorio D. Chazenbalk and Sandra M. McLachlan

Synonyms	None
Abbreviations	TSH receptor; TSHR
Classifications	Member of the superfamily of G protein coupled receptors with seven membrane spanning segments.
Description	A membrane glycoprotein synthesized in thyroid follicular cells as a single polypeptide chain of 764 aa, including a 21 residue signal peptide. There are six potential N-linked glycosylation sites in the TSHR ectodomain. Although glycosylation at individual sites has not been confirmed directly, the very large carbohydrate content of the TSHR (25-30 kDa) suggests that all are glycosylated.
Structure	The mature receptor on the cell surface exists in two forms; one with a single polypeptide chain and another in which intramolecular cleavage creates a heterodimer of two subunits (A and B) linked by disulfide bonds. Recent evidence suggests that intramolecular cleavage occurs at two sites in the ectodomain, with the excision of a ~ 5 kDa C peptide. TSH binds with similar high affinity to both single chain and two subunit forms of the receptor and cleavage into two subunits does not appear to be necessary for hormone action. The ectodomain (residues 22-418) consists of three regions; N- and C-terminal regions with low homology to other glycoprotein hormone receptors and a more conserved mid-regions (residues 54-254) with 9 leucine-rich repeats. The serpentine membrane spanning region (residues 419-764) traverses the membrane 7 times, ending in an ~81 residue cytoplasmic tail.
Molecular Weight	The deduced molecular weight of the single chain polypeptide (less the signal peptide) is 84,405. The mature, glycosylated TSHR is ~ 110-115 kDa.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	Unknown
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Function	TSH binds to and activates the TSHR by inducing coupling to Gs with activation of adenylate cyclase and increased intracellular cAMP synthesis. Cyclic AMP increases thyroid hormone synthesis and secretion, as well as thyrocyte proliferation. Although signalling through Gs is dominant, the TSHR also signals through the other three forms of G protein families (Gq/11, Gi and G12). Signalling via Gq/11 occurs at supraphysiologic TSH concentrations.

Physiology/Pathology	TSH activation of the TSHR is one of the two major mechanisms that influence thyroid gland function in maintaining thyroid hormone homeostasis and euthyroidism (the second being iodine autoregulation). TSH secretion by the anterior pituitary gland is subject to a negative feedback servo-regulatory mechanism by circulating thyroid hormones. In Graves' disease, a common organ-specific autoimmune disease, autoantibodies to the ectodomain of the TSHR mimic the action of TSH and activate the receptor, leading to autonomous thyroid function with goiter and clinical thyrotoxicosis. More rarely, occupancy of the TSHR with non-stimulatory autoantibodies can prevent the action of TSH and cause hypothyroidism. Somatic point mutations, primarily in the membrane spanning and cytoplasmic regions of the TSHR can cause constitutive activation thereby leading to autonomously functioning thyroid adenomata with clinical thyrotoxicosis. The frequency of TSHR mutations as the basis for this disease is quite variable in studies from different countries (3-75%). Functional mutations have been reported in the third cytoplasmic loop (residues 619, 623), the second cytoplasmic loop (residue 528), the transmembrane segments (residues 453, 505, 509, 597, 629, 631, 632, 633, 670, 672), the exo-loops (residues 486, 568, 650) and the ectodomain (residues 281, 276). Very rarely germline mutations are the cause of non-autoimmune, familial thyrotoxicosis, generally with involvement of the same residues as in toxic adenomata. Even more rarely, homozygous or complex heterozygous germline mutations or deletions that cause reduced or absent responsiveness to TSH have been identified in patients with familial and congenital hypothyroidism. Involvement of residues 109, 162, 167, 309 and 546 has been reported, as well as deletion of residues 406-412.
Degradation	Unknown. Presumed internalization and degradation by the endosomal pathway.
Genetics/Abnormalities	The gene is on chromosome 14 at q31 and is > 60 kb long with 10 exons. The ectodomain is encoded by exons 1-9 and part of exon 10. The transmembrane and cytoplasmic regions are all encoded by exon 10, which is the prototypic exon in the superfamily of receptors. As discussed above, a wide variety of somatic and germline point mutations are the cause of thyroid dysfunction in some individuals and families. A non-functional polymorphism at residue 52 exists in ~10% of the normal population.
Half-life	Unknown
Concentration	Estimated to be 2000-5000 receptors per thyroid cell. Up to 2 x 10^6 receptors expressed on the surface of Chinese hamster ovary cells.
Isolation Method	The mammalian, native TSHR has not been purified to homogeneity. A truncated, secreted ectodomain variant epitope tagged with 6 His residues has been partially purified by affinity chromatography.
Amino Acid Sequence	MRPADLLQLV LLLDLPRDLG GMGCSSPPCE CHQEEDFRVT CKDIQRIPSL PPSTQTLKLI ETHLRTIPSH AFSNLPNISR IYVSIDVTLQ QLESHSFYNL SKVTHIEIRN TRNLTYIDPD ALKELPLLKF LGIFNTGLKM FPDLTKVYST DIFFILEITD NPYMTSIPVN AFQGLCNETL TLKLYNNGFT SVQGYAFNGT KLDAVYLNKN KYLTVIDKDA FGGVYSGPSL LDVSQTSVTA LPSKGLEHLK ELIARNTWTL KKLPLSLSFL HLTRADLSYP SHCCAFKNQK KIRGILESLM CNESSMQSLR QRKSVNALNS PLHQEYEENL GDSIVGYKEK SKFQDTHNNA HYYVFFEEQE DEIIGFGQEL KNPQEETLQA FDSHYDYTIC GDSEDMVCTP

KSDEFNPCED IMGYKFLRIV VWFVSLLALL GNVFVLLILL
TSHYKLNVPR FLMCNLAFAD FCMGMYLLLI ASVDLYTHSE
YYNHAIDWQT GPGCNTAGFF TVFASELSVY TLTVITLERW
YAITFAMRLD RKIRLRHACA IMVGGWVCCF LLALLPLVGI
SSYAKVSICL PMDTETPLAL AYIVFVLTLN IVAFVIVCCC
HVKIYITVRN PQYNPGDKDT KIAKRMAVLI FTDFICMAPI
SFYALSAILN KPLITVSNSK ILLVLFYPLN SCANPFLYAI
FTKAFQRDVF ILLSKFGICK RQAQAYRGQR VPPKNSTDIQ
VQKVTHDMRQ GLHNMEDVYE LIENSHLTPK KQGQISEEYM
QTVL

Disulfides/SH Groups

The pairing of cysteines to form disulfide bonds has not been determined. The disulfide(s) of greatest interest will be that which links the A and B subunits in the heterodimeric form of the receptor.

General References

Nagayama, Y. and Rapoport, B. *Mol. Endocrinol.* 1992, **6**:145-156.
Vassart, G. and Dumont, J.E. *Endocr. Rev.* 1992, **13**:596-611.
Van Sande, J. et al. *J. Clin. Endocrinol. Metab.* 1995, **80**:2577-2585.
Russo, D. et al. *J. Endocrinol. Invest.* 1997, **20**:36-47.
Chazenbalk, G.D. et al. *Endocrinology* 1997, **138**:2893-2899.

Ref. for DNA/AA Sequences

Nagayama, Y. et al. *Bichem. Biophys. Res. Comm.* 1989, **165**:1184-1190.

Thyroxine-binding globulin

Yoshitaka Hayashi and Samuel Refetoff

Synonyms	Thyroid hormone-binding globulin; Thyropexin
Abbreviations	TBG
Classifications	Electr. Mob.:α1 - α2 globulin at pH 8.6
Description	A plasma protein synthesized in the liver, comprising a single polypeptide chain of 395 aa with four carbohydrate side chains, N-linked to Asn-16, Asn-79, Asn-145, Asn-233. There are three relatively common variants: 1) TBG-Poly: Phe-283, found with variable frequency in individuals of different races. The molecule is indistinguishable from the wild (common) type TBG (TBG-C) by functional or electrophoretic analyses. 2) TBG-Slow: Asn-283, with an allele frequency of 5-16% in Black populations and 2-11% in Pacific Islanders. It is functionally indistinguishable from TBG-C but has slower electrophoretic mobility on SDS-PAGE and cathodal shift on isoelectric focusing (IEF) due to a loss of a negative charge. 3) TBG-A: Thr-191, has a 51% allele frequency in Australian Aborigines. It binds thyroid hormones with approximately 50% reduced affinity and has reduced stability.
Structure	The tertiary structure has not yet been determined due to the difficulty in crystalizing the molecule. Carbohydrate residues are essential for proper folding of the molecule during intracellular processing. However, removal of the carbohydrate from the mature protein does not cause major alteration in conformation or thyroid hormone binding.
Molecular Weight	54,000 (18-20% carbohydrate)
Sedimentation Coeff.	3.9 S
Isoelectric Point	4.3-4.5 (three major bands, IEF)
Extinction Coeff.	7 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Susceptible to irreversible denaturation by heat, acid, and reduction as well by proteolytic cleavage.
Biological Functions	Monomeric molecule with single iodothyronine-binding site, carrying 75% of serum thyroxine (T_4) and 70% of triiodothyronine (T_3). Ka for T_4 is $1.1 \times 10^{10}\ M^{-1}$ and Ka for T_3 is $1.4 \times 10^{9}\ M^{-1}$.
Physiology/Pathology	Transports iodothyronines, in serum and contributes to the maintenace of the large extrathyroidal pool of thyroid hormone, prolonging the half life of T_4 to 8 days rather than 2 days in the absence of TBG. Several agents are known to alter the serum concentration of TBG. A number of variants have been described causing complete or incomplete TBG deficiency as

well as TBG excess, or molecule with decreased Ka for thyroid hormones. In such individuals, serum thyroid hormone levels are high or low while the subjects maintain an euthyroid state. Measurement of serum thyrotropin or free hormone levels is helpful in the exclusion of hyper- or hypothyroidism.

Degradation

Cleared by liver probably following partial desialylation.

Genetics/Abnormalities

The gene is on the long arm of x-chromosome. 5.6kb long with 5 exons, the last four containing all coding information for the TGB protein including a 20 aa signal sequence. Gene duplication and triplication were detected in several families with TBG-excess. In addition to the three variant TBG's occurring with high frequency, 7 other mutations have been so far characterized. Three cause complete deficiency (CD) of TBG in serum: [i] CDJ, a truncated protein with 373 aa, resulting from a frame shift mutation at Leu-352; [ii] CD-5, Pro-227 producing a retention of the mutant TBG in the endoplasmatic reticulum; and [iii] CD-6 a truncated protein with 167 aa, resulting from frame shift mutation at Val-165. Other variants due to single nucleotide substitutions mutations are:

Type	Abbreviation	Mutation (amino acid)	Affinity (Ka mutant) (Ka normal)	Stability	IEF shift
Common	TBG-C		1	normal	normal
Polymorphic	TBG-Poly	Phe-283	1	normal	normal
Slow	TBG-S	Asn-171	1	minmally reduced	cathodal
Aborigine	TBG-A	Thr-191	0.47	slightly reduced	normal
San Diego	TBG-SD	Thr-23	0.44	slightly reduced	normal
Gary	TBG-G	Asn-96	<0.1	markedly reduced*	anodal
Montreal	TBG-M	Pro-113	0.73	moderately reduced*	cathodal
Quebec	TBG-Q	Tyr-331	0.70	moderately reduced*	cathodal
PD Japan	TBG PDJ	Leu-363	-	slightly reduced	normal
Chicago	TBG-Cgo	Phe-309	1	increased	normal

* Increased concentration of denatured TBG in serum

Half-life

5.3 ± 0.4 days

Concentration

9-21 mg L^{-1} (serum)

Isolation Method

Bioselective absorption chromatography, followed by ion exchange chromatography and affinity chromatography on Con-A-Speharose or fractionation by electrophoresis.

Amino Acid Sequence

```
ASPEGKVTAC HSSQPNATLY KMSSIFNLYR RFTVETPDKN
IFFSPVSISA ALVMLSFGAC CSTQTEIVET LGRNLTDTPM
VEIQHGFQHL ICSLNFPKKE LELQIGNALF IGKHLKPLAK
FLNDVKTLYE TEVFSTDFSN ISAAKQEINS HVEMQTKGKV
VGLIQDLKPN TIMVLVNYIH FKAQWANPFD PSKTEDSSSF
LIDKTTTVQV PMMHQMEQYY HLVDMELNCT VLQMDYSKNA
LALFVLPKEG QMESVEAAMS SKTLKKWNRL LQKGWVDLFV
EAAMSSKTLK KWNRLLQKGW VDLFVPKFSI SATYDLGATL
LKMGIQHAYS ENADFSGLTE DNGLKLSNAA HKAVLHIGEK
GTEAAAVPEV ELSDQPENTF LHPIIQIDRS FMLLILERST
RSILFLGKVV NPTEA
```

Belongs to serine protease inhibitor (SERPIN) superfamily.

Disulfides/SH-Groups

Not well characterized. Five cysteine residues at positions 10, 65, 66, 97 and 234.

General References

Refetoff, S. *Endocrine Reviews* 1989, **10**:275-293.
Bartalena, L. *Endocrine Reviews* 1990, **11**:47-64.
Refetoff, S. et al. *Horm. Res.* 1996, **45**:128-138.

Ref. for DNA/AA Sequences

Flink, L. et al. *Proc. Natl. Acad. Sci. USA* 1986, **83**:7708-7712.
Hayashi, Y. et al. *Mol. Endocrinol.* 1993, **7**:1049-1060.

Tissue factor

Arabinda Guha and Yale Nemerson

Synonyms	Tissue Factor, Coagulation Factor III, Tissue Thromboplastin.
Abbreviations	TF, rHTF (recombinant human tissue factor).
Classifications	None
Description	Membrane-bound glycoprotein: not normally present in circulation or in contact with it. A specific protein present in plasma membranes of many cells and is very rich in carbohydrates. Purified to homogeneity from bovine and human sources having mw of $\simeq$ 43kDa and $\simeq$ 46kDa, respectively. Requires insertion into lipid vesicles for its cofactor activity. Complete primary sequence inferred from cDNA and aa sequencing. The latter covered 72% of the molecule. The mature single-chain protein consists of an extracellular domain (residues 1–219), a membrane-spanning hydrophobic domain (residues 220–242) and a cytoplasmic tail (residues 243–263); four concensus N-linked carbohydrate attachment sites (Asn-Xaa-Ser/Thr) occur in the sequence: three of them contained in the extracellular domain on the cell surface and the fourth site exists in cytoplasmic domain.
Structure	2.2: extracellular domain.
Molecular Weight	46,000 (native TF from human brain, SDS-PAGE); 29,593 (rHTF, from aa composition and cDNA sequence). Carbohydrate content: 7,500–8,000 obtained from chemical and enzymatic deglycosylation.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	TF is insoluble. The extracellular domain, residues 1–219, is soluble and has an E_{280} of 4.13×10^4.
Enzyme Activity	None
Coenzymes/Cofactors	Obligatory co-factor for factor VII mediated initiation of extrinsic pathway of blood coagulation.
Substrates	No known substrate. TF : FVII/VIIa catalytic complex activates factor X and factors X_a and IX_a, thereby initiating coagulation in presence of Ca^{2+}.
Inhibitors	Tissue Factor Pathway Inhibitor purified from HepG2 cell culture media and plasma is a potent inhibitor of TF : VIIa complex; however, TFPI requires the presence of factor Xa to form a stable quaternary complex that results in inhibited TF.
Biological Functions	Essential cofactor for the initiation of coagulation. In TF-initiated coagulation, factor X_a is generated directly by the action of TF : VIIa complex and indirectly by the activation of factor IX to IX_a. The latter then converts factor X to Xa in the presence of factor VIII.
Physiology/Pathology	See Biological Functions. Human, rabbit, cow, sheep, horse and pig Tf's, accelerate one-stage clotting time of homologous and heterologous plas-

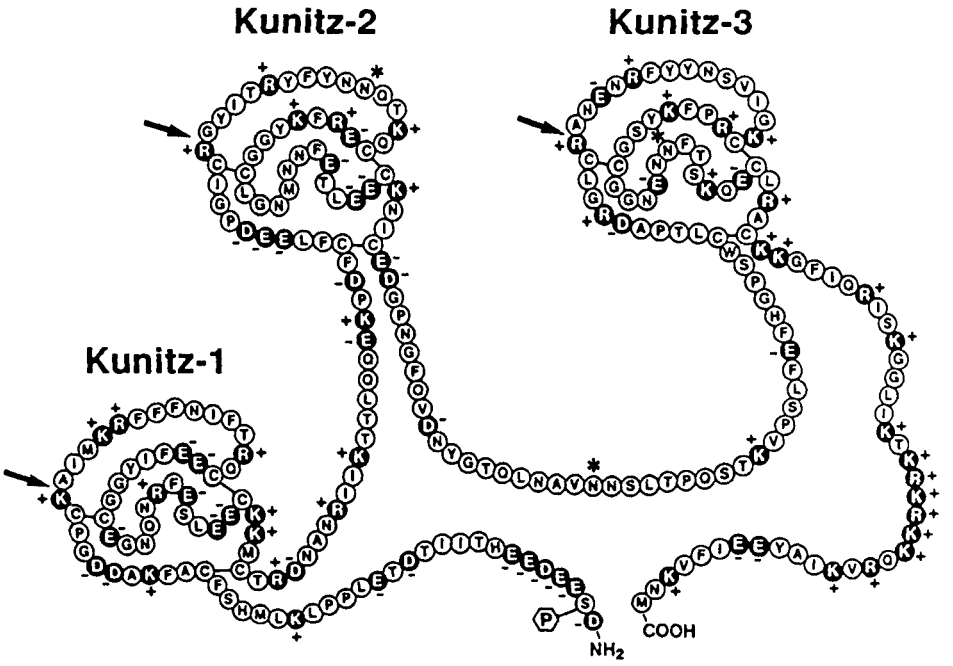
	ma, whereas Tf from mouse, rat, hamster and guinea pig have very little or no clot-promoting activity in heterologous plasmas from the larger animals.
Degradation	No known physiological degradation occurs. Tryptic digestion removes a C-terminal peptide from bovine TF without loss of activity. Chemical and enzymatic deglycosylation does not affect its activity. Subtilisin degrades TF with concomittent loss of activity. Effect of other proteolytic enzymes are not known.
Genetics/Abnormalities	The gene was mapped to the short arm of human chromosome 1 (pter-1p21), interpreted as evidence for a single TF gene in human genome. Synthesized by a single mRNA.
Half-life	Unknown
Concentration	Unknown
Isolation Method	Purified from human brain: extracted from acetone-dried brain power using 0.1% Triton X-100; chromatographed on monoclonal antibody column. Immunoaffinity purified TF from human brain, placenta and rHTF have identical amino-acid composition, N-terminal sequence and specific procoagulant activity.
Amino Acid Sequence	When examined for homology against 4668 sequences in the protein data base of National Biomedical Research Foundation, as well as several procoagulant and anticoagulant proteins, no significant homologies were observed. Partial aa sequence of 188 residues of placental TF from the N-terminus of mature protein confirms the protein sequence deduced from cDNA.
Disulfides/SH-Groups	Contains 5 half-cysteine residues: four half-cysteine residues in extracellular domains form two disulfide bonds Cys-49 to Cys-57 and Cys-186 to cys-209. The fifth cysteine residue in the cytoplasmic domain, Cys-245, is thioesterified to palmitate and stearate.
General References	Nemerson, Y. Tissue factor and hemostasis. *Blood* 1988, **70**:1–8. Bach, R. R. Initiation of hemostasis by tissue factor. *CRC Critical Review in Biochemistry* 1988, **23**: 339–368. Broze, G. J., Jr. and Miletich, J. P. Isolation of tissue factor inhibitor produced by HepG2 hepatoma cells. *Proc. Natl. Acad. Sci. USA* 1987, **84**: 1886–1890.
Ref. for DNA/AA Sequences	Spicer, E. K., et al. Isolation of cDNA clones coding for human tissue factor: Primary structure of the protein and cDNA. *Proc. Natl. Acad. Sci. USA 1978, **84**: 5148–5152.* Scarpati, E. M., et al. Human tissue factor: cDNA sequence and chromosome localization of the gene. *Biochemistry* 1987, **26**: 5234–5238. Morrissey, J. H., et al. Molecular cloning of the cDNA for tissue factor, the cellular receptor for the initiation of the coagulation cascade. *Cell* 1987, **50**: 129–135.

Tissue factor pathway inhibitor

George J. Broze, Jr.

Synonyms	Antithromboplastin; Anticonvertin; Tissue factor inhibitor; Extrinsic pathway inhibitor; Lipoprotein-associated coagulation inhibitor
Abbreviations	TFPI; EPI; LACI
Classifications	Multivalent Kunitz-type coagulation inhibitor
Description	Circulates in plasma in both lipoprotein-associated and free forms and is secreted by stimulated platelets. Additional TFPI, possibly derived from the endothelium, is released into plasma following heparin infusion. Molecule contains an acidic N-terminal region followed by three tandem Kunitz-type proteinase inhibitor domains and a basic C-terminus. Contains three potential sites for N-linked glycosylation. TFPI produced by cultured endothelial and kidney cells contains sulfated N-linked oligosaccharides.
Structure	Three dimensional structure has not been determined.
Molecular Weight	32,000 (as protein composition, predicted by cDNA sequencing); 43,000 (SDS-PAGE).
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	Factor Xa required for efficient inhibition of the factor VIIa/tissue factor enzymatic complex by TFPI. Heparin and other glycosaminoglycans enhance Factor Xa inhibition by TFPI.
Substrates	None
Inhibitors	None
Biological Functions	Directly inhibits Factor Xa, and in a Factor Xa dependent fashion produces feedback inhibition of the factor VIIa/Tissue Factor catalytic complex. Second Kunitz-type domain is responsible for factor Xa inhibition; first Kunitz-type domain required for inhibition of factor VIIa in a putative quaternary inhibitory complex containing factor Xa-TFPI-factor VIIa/ Tissue Factor; function of third Kunitz-type domain is not known.
Physiology/Pathology	Regulation of tissue factor-induced blood coagulation. Human deficiency has not been reported; TFPI gene deletion in mice produces intrauterin mortality.
Degradation	Carboxy-terminal truncated forms of TFPI circulate in plasma. Mechanism for the apparent proteolytic degradation is not known.
Genetics/Abnormalities	Location 2q31-32. No variants described.

Half-life	Very rapid alpha phase potentially related to endothelial binding; beta phase 30-60 minutes.
Concentration	Plasma: 2.5 nmol/L
Isolation Method	Initial purification from conditioned media of cultured human hepatoma cells (HepG2) in multistep procedure including affinity chromatography on factor Xa-agarose. Subsequent purifications from plasma and conditioned media rely on immunoaffinity chromatography.
Amino Acid Sequence	Substantial homology with other proteinase inhibitors of the Kunitz-type (e.g. bovine pancreatic trypsin inhibitor).
Disulfides/SH-Groups	9; 3 in each Kunitz-type domain.
General References	Broze, G.J., Jr. et al. Regulation of coagulation by a multivalent Kunitz-type inhibitor. *Biochemistry* 1990, **29**:7539-7546. Rapaport, S.I. The extrinsic pathway inhibitor: A regulator of tissue factor-dependent blood coagulation. *Thromb. Haemostas.* 1991, **66**:6-15. Novotny, W.F. et al. Plasma antigen levels of the lipoprotein-associated coagulation inhibitor in patient samples. *Blood* 1991, **78**:387-393.
Ref. for DNA/AA Sequences	Wun, T.C. et al. Cloning and characterization of a cDNA coding for the LACI shows that it consists of three tandem Kunitz-type inhibitory domains. *J. Biol. Chem.* 1988, **263**:6001-6004. Girard, T.J. et al. Structure of the human lipoprotein-associated coagulation inhibitor gene: Intron/exon gene organization and localization of the gene to chromosome 2. *J. Biol. Chem.* 1991, **266**:5036-5041. GenBank/EMBL Data Bank accession numbers: cDNA - J03225, Gene - M59491-M59499.



→ $P_1 - P_1'$ sites of Kunitz domain * Potential N-linked glycosylation Ⓟ Phosphorylation

Tissue Kallikrein

Cindy Wang, Lee Chao and Julie Chao

Synonyms	Human glandular kallikrein; Human tissue kallikrein; Pancreatic kallikrein; Renal kallikrein; Urinary kallikrein
Abbreviations	HUK; TK
Classifications	None
Description	A serine proteinase, synthesized in the pancreas, kidney and salivary glands. The active human tissue kallikrein consists of a single polypeptide chain held together by intrachain disulfide bonds. The human tissue kallikrein is synthesized as a zymogen (prokallikrein) with an attached 17 aa signal peptide preceding a 7 aa activation sequence that must be cleaved to activate the enzyme. Human prokallikrein can be activated by thermolysin, trypsin and human plasma kallikrein *in vitro*, however, the *in vivo* activating enzyme is unknown. Human tissue kallikrein is an acidic glycoprotein, variably and extensively (~20% molecular weight) glycosylated.
Structure	The tertiary structure has not been determined for human tissue kallikrein, however human tissue kallikrein shares high sequence identity to porcine kallikrein and tonin.
Molecular Weight	40,000
Sedimentation Coeff.	None
Isoelectric Point	3.9-4.0. IEF gives 4 main bands
Extinction Coeff.	15 (280nm, 1%, 1cm)
Enzyme Activity	Human tissue kallikrein releases lysyl-bradykinin while rat and mouse tissue kallikreins release bradykinin from kininogens. For synthetic peptide substrates, Arg is favored in position P1. One cause of the narrow specificity of tissue kallikrein is its pronounced secondary specificity for a bulky, hydrophobic aa residue in the P2 position. The pH optimum is about 8.5 to 9.0. Assays are most conveniently made with the alpha-N-tosyl-L-arginine methyl ester (Tos-Arg-OMe) by measuring the arginine esterase activity, and with chromogenic substrates (Bz-Pro-Phe-Arg-*p*-nitroanilide and Val-Leu-Arg-*p*-nitroanilide), and fluorogenic substrates (Pro-Phe-Arg-7-amino-4-methylcoumarin, Phe-Phe-Arg-7-amino-4-methylcoumarin and Val-Leu-Arg-7-amino-4-methylcoumarin). The measurement of tissue kallikrein activity can also be made by measuring kininogenase activity with a kinin radioimmunoassay.
Coenzymes/Cofactors	None
Substrates	Chromogenic substrates: Bz-Pro-Phe-Arg-*p*-nitroanilide and Val-Leu-Arg-*p*-nitroanilide. Synthetic fluorogenic substrates: Pro-Phe-Arg-7-amino-4-methylcoumarin, Phe-Phe-Arg-7-amino-4-methylcoumarin and Val-Leu-Arg-7-amino-4-methylcoumarin.
Inhibitors	The activity of tissue kallikrein is inhibited by a number of serine proteinase inhibitors (serpins) including kallistatin, protein C inhibitor and alpha 1-antitrypsin. The activity of tissue kallikrein is inhibited by aprotinin, phenylmethylsulfonyl fluoride, benzamidine and leupeptin, but

is not affected by soybean trypsin inhibitor, limabean trypsin inhibitor and ovomucoid trypsin inhibitor.

Biological Functions

The tissue kallikrein-kinin system has been implicated in the pathogenesis of hypertension. Kallikrein excretion in urine was decreased in patients with essential hypertension and increased by sodium depletion. A family pedigree study indicated that a dominant allele expressed as high urinary kallikrein excretion may be associated with a decreased risk of essential hypertension. By employing molecular genetic approaches, a direct link between alteration of tissue kallikrein gene expression and blood pressure regulation has been demonstrated. Transgenic mice overexpressing human tissue kallikrein under the control of the metallothionein metal response element (MRE) or albumin gene enhancer/promoter are hypotensive throughout their life span. In adult spontaneously hypertensive rats (SHR), somatic gene delivery of human tissue kallikrein by systemic and targeted approaches caused a sustained reduction of blood pressure for up to 8 weeks. These findings support the notion that tissue kallikrein plays an important physiological role in maintaining low blood pressure.

Physiology/Pathology

Tissue kallikrein is present in cells of most tissues, but is particularly abundant in pancreas, kidney, urine, and saliva. It is localized in the epithelial or secretory cells of various ducts, including salivary, sweat, pancreatic, prostatic, intestinal, and the distal nephron. In addition, it has been found in neutrophils, colonic mucous cells, the trachea, nasal mucosa, and the anterior pituitary. The best-known physiological function of tissue kallikrein is the cleavage of low molecular weight kininogen substrate to release the vasoactive kinin peptide. Kinins bind to their receptors at target organs and exerts a broad array of biological activities including vasodilation, blood pressure reduction, smooth muscle relaxation and contraction, pain induction and inflammation.

Degradation

The plasma clearance rate of human tissue kallikrein is much shorter than for the kallikrein-kallistatin complex. The liver and kidney are the major sites for circulatory clearance and catabolism of human tissue kallikrein.

Genetics/Abnormalities

The gene is localized on chromosome 19q13.3-13.4 and consists of five exons and four introns. Human tissue kallikrein gene polymorphisms were identified in the promoter region between 121 and 133. Ten alleles with length and nucleotide sequence variations were identified. Promoter analysis revealed that alleles D and H had significantly lower promoter activities than the other alleles.

Half-life

8 minutes in plasma.

Concentration

Normal serum, 3.8 ± 0.7 ng/ml; urine, 36.8 ± 9.3 ng/ml; saliva, 1.86 ± 0.34 μg/ml.

Isolation Method

35-80% sat. ammonium sulphate fractionation of concentrated urine followed by ion-exchange chromatography on DE-52 cellulose and affinity chromatography on aprotinin-agarose.

Amino Acid Sequence

Human tissue kallikrein shares 60-66% identity in amino acid sequence with other members of human tissue kallikrein family listed below. The similarity is especially high in the sequences around the active sites.

```
Kallikrein   LTAAHCIDYS HDLMLLRLKD TCVGDSGGPLM
PSA          LTAAHCIDSS HDLMLLRLKS TCSGDSGGPLV
hGK-1        LTAAHCLDSS HDLMLLRLKD TCGGDSGGPLV
                  *       *            *
```

Kallikrein: tissue kallikrein; PSA: prostate specific antigen; hGK-1: human glandular kallikrein-1. * indicates the active sites.

Disulfides/SH-Groups

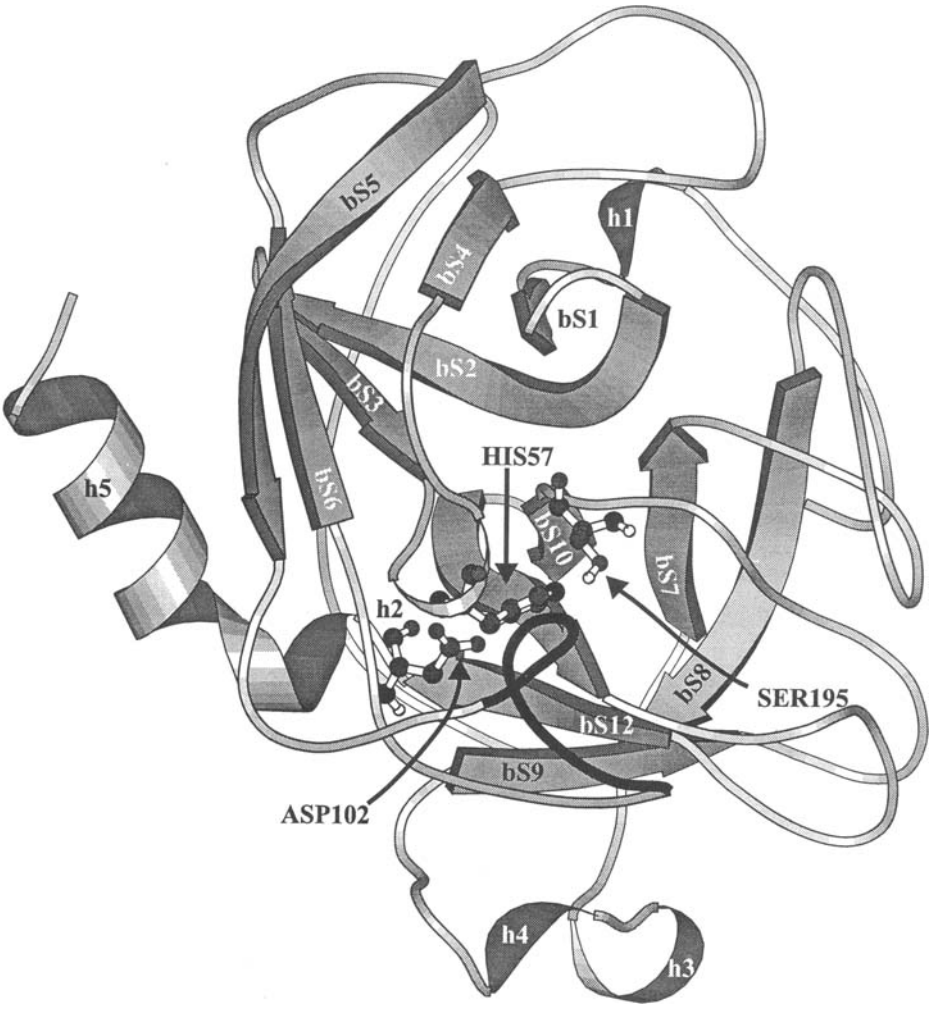
There are 5 intrachain disulfide bonds in human tissue kallikrein, which are highly conserved in tissue kallikrein family members.

General References

Berry,T.D. et al. *Hypertension* 1989, **13**:3-8.
Bhoola, K.D. et al. *Pharmacol. Rev.* 1992, **44**:1-80.
Chao, J. et al. *Hypertension* 1996, **27**:491-494.
Fujimoto, Y. et al. *Biochem. Med. Metab. Biol.* 1990, **44**:218-227.
Geiger, R. et al. *Hoppe-Seylers Z. Physiol. Chem.* 1981, **362**:317-325.
Kato, H. et al. *J. Biochem.* 1987, **102**:1389-1404.
Margolius, H.S. et al. *Lancet* 1971, **2**:1063-1065.
Shimamoto, K. et al. *J. Clin. Endocrinol. Metab.* 1980, **51**:840-848.
Zhou, X.G. et al. *J. Biol. Chem.* 1992, **267**:25873-25880.

Ref. for DNA/AA Sequences

Fukushima, D. et al. *Biochemistry* 1985, **24**:8037-8043.
Lin, F.K. et al. *Biochim. Biophys. Acta* 1993, **1173**:325-328.



Homology model of human tissue kallikrein

Tissue Transglutaminase

Daniel Aeschlimann

Synonyms	cytosolic-, endothelial-, erythrocyte-, liver transglutaminase, trans-glutaminase type II
Abbreviations	TG_C, tTG
Classifications	E. C. 2.3.2.13 pI < 6.4 and $M_r \sim 77{,}000$ Da in human keratinocyte 2-D gel protein database (Celis, J. E. et al. *Electrophoresis* 1993, **14**:1091–1198).
Description	Tissue transglutaminase is a Ca^{2+}-dependent enzyme that catalyzes the formation of intra- and intermolecular isopeptide bonds in intra- and extracellular proteins. The cross-linking reaction involves the γ-carboxamide group of specific peptide-bound glutamine residues and the ε-amino group of lysine residues or of low molecular weight amines structurally related to lysine. The enzyme is widespread in different organs and cell types in the vertebrate body. It is predominantly present in liver, spleen, lung, intestine, testes, heart, smooth muscle cells, endothelial cells, macrophages/monocytes, erythrocytes, differentiating chondrocytes, a sub-population of fibroblast-like cells and basal keratinocytes. The cross-linking reaction may serve multiple purposes, but the predominant expression of tissue transglutaminase in tissues lining the body surfaces and tissue interfaces, and the upregulation of the enzyme in remodelling processes in development and after injury, indicates a function in providing mechanical strength and chemical stability in order to maintain the integrity of the tissue.
Structure	Tissue transglutaminase is made up by a single polypeptide chain which folds into a globular particle with a size of $\sim 7 \times 9.5$ nm as revealed by electron microscopy. The frictional ratio, $f/f_{min.}$, of 1.4 obtained from sedimentation equilibrium centrifugation is consistent with a somewhat assymetric shape ($R_{Stokes} \sim 4$ nm). The monomeric enzyme is the predominant form but in addition it may form homoaggregates of ~400,000 Da (pentamer) as detected by sedimentation equilibrium centrifugation and gel filtration. The three-dimensional structure of the α-subunit of factor XIII, which is highly homologous to tissue transglutaminase, has recently been solved by X-ray crystallography (Yee, V. C. et al. *Proc. Natl. Acad. Sci. U. S. A.* 1994, **91**:7296–7300).
Molecular Weight	76,400–77,700 Da [calculated from amino acid sequence data, including man (77,253 Da), mouse, rat, guinea pig, cow, and chicken]; 74,000 ± 3,000 Da (guinea pig enzyme; sedimentation equilibrium centrifugation); ~77,000 Da (SDS-PAGE under reducing conditions)
Sedimentation Coeff.	4.4 S (guinea pig enzyme; physiological conditions)
Isoelectric Point	6.4 (human epidermal keratinocytes), 5.6 (human erythrocytes), 5.35 (rabbit liver)
Extinction Coefficient	13.0 ± 0.90 (guinea pig enzyme; 280 nm,1 %,1 cm)
Enzyme Activity	Transglutaminase (or transamidase) is referred to as R-glutaminyl-peptide: amine-γ-glutamyl-transferase by the Enzyme Commission on Nomenclature. It catalyzes the Ca^{2+}-dependent acyl-transfer reaction which re-

sults in the formation of new γ-amide bonds between γ-carboxamide groups of peptide-bound glutamine residues and various primary amines. The reaction proceeds in different steps, a first one in which the enzyme active site cysteine reacts with the glutaminyl residue to form the acyl-enzyme intermediate and ammonia. In a second step, the intermediate reacts with a primary amine to form an isopeptide bond and liberates the reactivated enzyme. The most common amine-donors are peptide-bound lysine residues and naturally occurring polyamines such as putrescine or spermidine. The reaction with lysyl residues results in intra- or intermolecular γ-glutamyl-ε-lysine cross-links and serves to form or covalently stabilize protein aggregates. The reaction with polyamines results in protein modifications possibly affecting the biological activity or turnover of the protein, but one main function is to compete for protein cross-linking.

Coenzymes/Cofactors

Transglutaminases are activated through a conformational change induced by binding of Ca^{2+} ions. Tissue transglutaminase was found to bind 3–4 Ca^{2+} ions and shows half maximal activation at a Ca^{2+} concentration of 0.1–0.5 mM.

Substrates

The number of proteins acting as glutaminyl substrates is restricted since there are strict requirements on primary structure and conformation in the regions adjacent to the reactive glutamine residues. However, no consensus sequence could be derived from known substrate sites. Extracellular amine acceptor substrates are β-casein, collagens type II, type III (Gln^{14}, in *N*-propeptide), and type V (α1-chain), fibrinogen, fibronectin (Gln^{3}, others unknown), lipoprotein (a), nidogen/entactin (Gln^{726}), osteonectin/BM-40/SPARC (Gln^{3} and Gln^{4}), osteopontin/SPP-1 (Gln^{34} and Gln^{36}), phospholipase A_2 type 1 (Gln^{4}), plasminogen activator inhibitor type 2 (Gln^{83}, Gln^{84} and Gln^{86}), and vitronectin. Intracellular amine acceptor substrates include actin (Gln^{41}), erythrocyte band 3 (Gln^{30}), β-crystallins (A3: Gln^{6}, Gln^{7}; B_P: Gln^{7}; B3: Gln^{3}; in *N*-terminal extensions), lipocortin/annexin 1(Gln^{19} and/or Gln^{23}), and micro-tubules associated protein τ. Amino acid residues adjacent to the lysine residue in amine donor substrates do not have a pronounced influence on the cross-linking, and it appears that the structure of the protein assembly rather than the primary structure of the protein determines the location of the cross-linking. A single lysine in αB-crystallin has been identified as a major amine donor in lens. For detection of transglutaminase activity *in vitro*, the most frequently used compounds are carbobenzoxy-Gln-Gly as amine acceptor and monodansylcadaverine or radioactively labeled histamine, methylamine, putrescine, spermidine, or spermine as amine donor.

Inhibitors

Divalent cations:
Several heavy metal ions such as Fe^{2+}, Cu^{2+}, Hg^{2+} etc. strongly inhibit transglutaminases by oxidation of free sulfhydryl groups to intramolecular disulfide bonds and thereby induced conformational changes. Zn^{2+} ions influence transglutaminase activity by a different molecular mechanism and act in a competitive manner in presence of Ca^{2+} with a $K_i \sim 10^{-7}$ M. Zn^{2+} might be an important negative regulator intracellularly.
Guanosine nucleotides:
Tissue transglutaminase, in contrast to the other members of the transglutaminase family, is inhibited by GTP and GDP, which bind to the enzyme in a reversible noncompetitive manner in presence of Ca^{2+}. Half-maximal inhibition is obtained with GTP at ~100 μM and with GDP at ~400 μM. The inhibition is due to a conformational change and does not depend upon hydrolysis of the nucleotide even though the enzyme contains GTPase activity. GMP and cyclic GMP show no inhibitory effect.

Artificial Inhibitors:
Examples of potent amine-donor inhibitors are phenylthiourea derivatives [i.e. 1-(5-aminopentyl)-3-PTU] ($K_M \sim 10^{-6}$ M, maximal nontoxic dose in cell culture: 0.5–1 mM), monodansylcadaverine ($K_M \sim 10^{-5}$ M, maximal nontoxic dose in cell culture: 0.1–0.2 mM). Recently, transglutaminase active site-directed inhibitors were developed: 2-[(2-oxopropyl)thio]-imidazolium derivatives (L683685, L683696, L682777), 2-(1-acetonylthio)-5-methylthiazolo[2,3]-1,3,4-thiadiazolium perchlorate (L722151), 3-halo-4,5-dihydroisoxazoles, and sulfonium methylketones [i.e. *N*-carbobenzoxy-phenyl-γ-aminobutyl sulfonium methylketone).

Biological Functions

1) Tissue transglutaminase-catalyzed cross-linking has been demonstrated for extracellular proteins such as fibrinogen and fibronectin in the pericellular matrix of hepatocytes and endothelial cells, for laminin-nidogen complexes in basement membranes, and for osteonectin and collagen type II in differentiating cartilages.
2) Tissue transglutaminase accumulates in the cytoplasm of many cells undergoing programmed cell death (apoptosis). Its function might be to form a cross-linked envelope when the cells loose their membrane integrity to prevent leaking of intracellular components. Cross-linked cell envelopes, so-called apoptotic bodies, have been described in numerous tissues including liver and related to tissue transglutaminase activity. Aged erythrocytes undergo an irreversible shape change upon tissue transglutaminase-catalyzed cross-linking of the cytoskeletal network underlying the plasma membrane.

Physiology / Pathology

No case of tissue transglutaminase deficiency has been reported. However, abnormal high levels of tissue transglutaminase cross-linking have been associated with different pathological situations, i.e. loss of eye lens transparency (senile cataract), formation of artherosclerotic plaques, promotion of inflammatory diseases such as rheumatoid arthritis, and accelerated ageing of erythrocytes in Hb-Koln disease.

Degradation

physiological degradation pathways unknown

Genetics / Abnormalities

single copy gene located on human chromosome 20q12; promoter region contains consensus motifs of estrogen receptor binding-site (which may provide the basis for retinoic acid regulation) and responsive elements for interleukin-6; no mutations known.

Half-life

unknown

Concentration

Varies considerably between cell types and tissues. Estimated concentration of human erythrocyte and guinea pig liver transglutaminase is approximately 10 mg/L of blood and 0.5–1 mg/g wet tissue, respectively. Tissue transglutaminase expression is strongly induced by retinoic acid and interleukin-6

Isolation Method

Tissue transglutaminase has classically been purified from two sources, human erythrocytes (Signorini, M. et al. *Biol. Chem. Hoppe-Seyler* 1988, **369**: 276–281) and rodent (e.g. guinea pig, rat, rabbit) liver (Connellan, J. M. et al. *J. Biol. Chem.* 1971, **246**: 1093–1098). Purification steps include DEAE-chromatography and gel filtration. The liver enzyme is precipitated from the DEAE-eluate with protamine sulfate prior to gel filtration. Tissue transglutaminase binds to heparin affinity matrices. Chromatography on strong anion-exchangers such as Mono Q (FPLC) or Q-Sepharose might be applied as a final step to obtain high purity.

Amino Acid Sequence	cDNA-cloning of tissue transglutaminase from human and bovine endothelial cell -, mouse macrophage -, guinea pig and rat liver -, and chicken red blood cell libraries revealed polypeptide chains of 685–691 amino acids with an overall sequence identity of 65–88%. Tissue transglutaminase lacks a hydrophobic leader sequence found in conventionally secreted proteins and the NH_2-terminus of the mature protein is blocked by removal of the terminal methionine residue and acetylation of the penultimate residue. Tissue transglutaminase is not glycosylated despite containing numerous potential sites for *N*-linked glycosylation. The proteins most closely related to tissue transglutaminase are the other vertebrate members of the transglutaminase family, i.e. plasma factor XIII, keratinocyte -, epidermal- and prostate transglutaminase, and the catalytically inactive erythrocyte membrane protein band 4.2. The most prominent sequence homology is found around the transglutaminase active site cysteine residue (consensus: PV-YGQC*WVFAGV--T-LR-LGIP-R-VTN--SAHDT). Regions rich in negatively charged residues might function as Ca^{2+}-binding sites since the sequence contains no typical Ca^{2+}-binding motifs such as the EF-hand structure.
Disulfides/SH-Groups	Human tissue transglutaminase contains 20 cysteine residues, but no disulfide bonds. The sequence position of 13 cysteine residues is completely conserved between species. The active site involves cysteine^{277} as determined for the guinea pig enzyme.
General References	Aeschlimann, D. and Paulsson, M. *Thromb. Haemostasis* 1994, **71**: 402–415. Fésus, L., Davies, P. J. A., Piacentini, M. *Eur. J. Cell Biol.* 1991, **56**: 170–177. Folk, J. E. and Finlayson, J. S. *Adv. Protein Chem.* 1977, **31**: 1–133. Lorand, L. and Conrad, S. M. *Mol. Cell. Biochem.* 1984, **58**: 9–35.
Ref. for DNA/AA Sequences	Gentile, V. et al. *J. Biol. Chem.* 1991, **266**: 478–483. Nucleotide sequences in GenBank/EMBL data bank have the accession numbers: M55153 (human), M55154 (mouse), M19646 (guinea pig), X60686 (bovine) and L02270 (chicken).

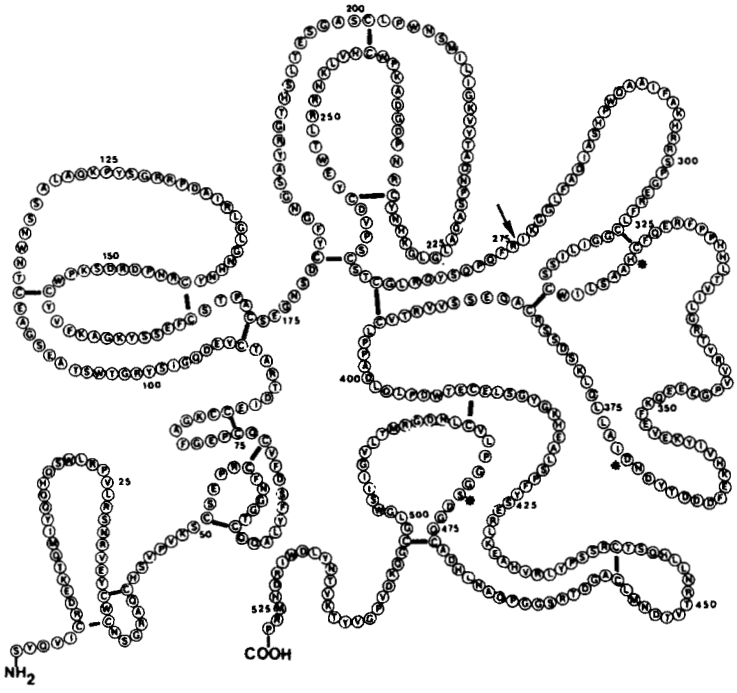
Tissue-type plasminogen activator

Désiré J. Collen

Synonyms	Blood plasminogen activator, Vascular plasminogen activator, Tissue plasminogen activator
Abbreviations	t-PA
Classifications	None
Description	The plasminogen activator t-PA is a serine proteinase that is synthesized and secreted by endothelial cells. It is a single chain glycoprotein of 527 aa that is converted to a two-chain form by proteolytic cleavage of the Arg-275-Ile-276 peptide bond. The native molecule may contain an NH_2-terminal extension of 3 aa (Gly-Ala-Arg). t-PA isolated from melanoma cells exists in 2 variants that differ in carbohydrate; type I contains N-linked glycans at residues Asn-117, Asn-184 and Asn-448, whereas type II lacks glycosylation at Asn-184.
Structure	Unknown
Molecular Weight	70,000 (SDS-PAGE); 59,000 (aa composition without carbohydrate).
Sedimentation Coeff.	Unknown
Isoelectric Point	7–8
Extinction Coeff.	Unknown
Enzyme Activity	500,000 IU/mg
Coenzymes/Cofactors	Fibrin acts as a cofactor for plasminogen activation.
Substrates	Plasminogen.
Inhibitors	Plasminogen activator inhibitor-1 (PAI-1) is the physiological inhibitor. It forms irreversible inactive stoichiometric complexes with t-PA with a very high second-order rate constant ($> 10^7 M^{-1}s^{-1}$). Useful synthetic inhibitors include tripeptide chloromethylketones (i.e. D-Phe-Pro-Arg-CH_2Cl).
Biological Functions	Dissolution of fibrin clots in circulating blood, via activation of plasminogen.
Physiology/Pathology	Preservation of a normal hemostatic balance. Increased levels may lead to bleeding complications whereas decreased levels may be associated with thrombotic complications.
Degradation	Conversion to a two chain molecule, which has a similar activity towards plasminogen in the presence of fibrin as the single chain molecule.
Genetics/Abnormalities	The gene contains 14 exons and is localized on chromosome 8, bands 8.p.12 → q.11.2. Abnormal forms have not been reported.
Half-life	$t^1/_2\alpha$: 6 min in blood.
Concentration	5–10 µg/L

Isolation Method	Immunoadsorption, Zinc-chelate chromatography, Lysine-Sepharose chromatography (mainly from cell cultures)
Amino Acid Sequence	The NH_2-terminal region is composed of several domains: a finger-like domain (aa 4 to 50) homologous to the finger-domains in fibronectin, an epidermal growth factor domain (aa 50 to 87) homologous to human epidermal growth factor and two kringle domains (aa 87 to 176 and 176 to 262) homologous to similar domains in other proteins such as plasminogen, urokinase, prothrombin ... The COOH-terminal region (aa 276 to 527) contains the serine proteinase domain homologous to other serine proteinases. The catalytic site is composed of His-322, Asp-371 and Ser-478. The highly positively charged aa sequence Lys-296 to Arg-304 is involved in interaction with PAI-1.
Disulfides/SH-Groups	17 disulfide bonds and one free Cys at position 83.
General References	Collen, D. et al. *Drugs* 1989, **38:** 346–388.
Ref. for DNA/AA Sequences	Pennica, D. et al. *Nature* 1983, **301:** 214–221. Ny, T. et al. *Proc. Natl. Acad. Sci. USA* 1984, **81:** 5355–5359. Degen, S. F. et al. *J. Biol. Chem.* 1986, **261:** 6972–6985.

Structure of t-PA, modified from Pennica et al. *Nature* 1983, **301:** 214–221.



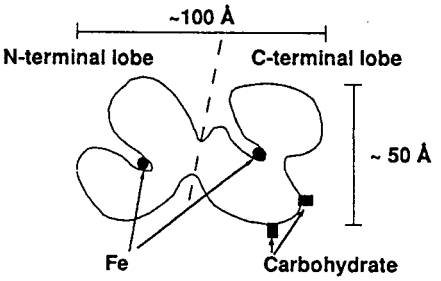
Transferrin

Evan H. Morgan

Synonyms	Siderophilin
Abbreviations	Tf; apoTf; Fe_NTf; Fe_CTf; Fe_2Tf
Classifications	Electrophoretic mobility: β1-globulin
Description	A circulating plasma protein, member of a family of iron-binding glycoproteins which also includes lactoferrin (milk and neutrophil granulocytes), melanotransferrin (melanoma antigen, p97) and ovotransferrin (egg white conalbumin). Circulating transferrin is synthesized in the liver; synthesis for local use also occurs in brain, mammary gland, testis and avian oviduct. Consists of a single polypeptide chain of 679 aa organized as two domains of equal size, each containing one high affinity metal binding site which binds Fe^{3+} reversibly (stability constant in plasma approx. 10^{22} M^{-1}). The protein may exist in the iron-free (apo), diferric (Fe_2Tf) or two monoferric forms (Fe_NTf, Fe_CTf). The binding sites can react with many other metal ions, less firmly than with Fe^{3+}. Carbohydrate (approx. 6%) is attached as two chains to the C-terminal domain at Asn-415 and Asn-618 (mostly as bi-antennary, some tri-antennary and tetra-antennary chains).
Structure	Crystallographic structures of rabbit plasma transferrin, human lactoferrin and hen ovotransferrin, have been solved to 3.2 Å, 2.8 Å and 2.4 Å respectively. Structures are very similar. The molecules are bilobal; each lobe consists of one of the domains and is further organized into two subdomains of approx. 160 aa separated by a cleft which contains the metal binding site where iron is bound by interaction with 2 tyrosine, 1 histidine, 1 aspartic acid residue and 1 bicarbonate (or carbonate) ion.
Molecular Weight	79,570 (aa and carbohydrate composition).
Sedimentation Coeff.	5.2 ± 0.2 S
Isoelectric Point	ApoTf: 5.6 - 6.0; Fe_2Tf: 5.0 - 5.4
Extinction Coeff.	ApoTf: 11.1; Fe_2Tf: 14.3 (280nm, 1%, 1cm) Fe_2Tf: 0.58 (465nm, 1%, 1cm) ApoTf: 88.200; Fe_2Tf: 114,000 (280nm, $M^{-1}cm^{-1}$) Fe_2Tf: 4,620 (465nm, M^{-1} cm^{-1})
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Transports iron in blood and interstitial fluid from sites of intestinal absorption, storage or erythrocyte destruction to erythropoietic, proliferating and iron storage cells and to the placenta. Iron uptake by cells usually occurs by receptor-mediated endocytosis of Fe_2Tf or FeTf, followed by recycling of the apoTf and receptor to the cell membrane and release of the apoTf from the cell. The transferrin and receptors are not degraded during

1. Installment 1992 (revised 1997)

	the process.In many types of cells proliferation requires the delivery of iron by Tf, hence Tf-iron functions as an essential growth factor. ApoTf may also function as a bacteriostatic agent by depriving micro-organisms of iron.
Physiology/Pathology	Essential for extracellular iron transport and haemoglobin synthesis in vivo and for cell division in many types of cells in culture. Atransferrinaemia (rare inherited disorder) leads to severe hypochromic anaemia and recurrent infections. Hepatic synthesis and plasma concentration rise in iron deficiency, pregnancy and oestrogen administration, and fall in iron overload, malnutrition, infections, trauma, inflammatory conditions and malignancies.
Degradation	Sites and mechanisms of degradation poorly defined; probably include liver and gut.
Genetics/Abnormalities	Transferrin gene located at q21-25 on chromosome 3, consists of 33.5 kb and contains 17 exons and 16 introns; thought to arise by duplication of an ancestral gene. More than 20 genetic variants recognized, all with normal function; most common variant is transferrin C; of the others only 4 have frequencies greater than 1% in various populations.
Half-life	8 - 10 days (blood circulation)
Concentration	Plasma: mean 2.3 g/L (29 µmole/L), range 1.8 - 2.7 g/L (23 - 34 µmole/L). Normally approx. 30% saturated with iron.
Isolation Method	Isolated from plasma by ammonium sulphate precipitation, ion-exchange chromatography and gel filtration; may also require affinity chromatography or preparative electrophoresis to eliminate hemopexin.
Amino Acid Sequence	Polypeptide chain consists of 2 domains formed from the N-terminal (aa 1 - 336) and C-terminal (aa 337 - 679) halves of the molecule, with approx. 40% homology in aa sequence between the domains. High degrees of sequence homology (40 - 70%) are found with human lactoferrin and melanotransferrin, and with bovine and murine lactotransferrins and chicken ovotransferrin.
Disulfides/SH-Groups	19 disulfide bonds, 8 in the N-terminal domain and 11 in the C-terminal domain, none between the domains.
General References	Brock, J.H. Transferrins. In: *Metalloproteins*, Harrison, P. (ed.) Macmillan, London 1985; 183-261. Harris, D.C., Aisen, P. Physical Biochemistry of the Transferrins. In: *Iron Carriers and Iron Proteins*, Loehr, T.M. (ed.) VCH Publishers, New York 1989, pp. 239-371. Huebers, H.A., Finch, C.A. *Physiol. Rev.* 1987, **67**:520-582. Morgan, E.H. *Molec. Aspects Med.* 1981, **4**:1-123. Williams, J. *Trends Biochem. Sci.* 1982, **7**:394-397.
Ref. for DNA/AA Sequences	Anderson, B.F. et al. *J. Mol. Biol.* 1989, **209**:711-734. Bailey S. et al. *Biochemistry* 1988, **27**:5804-5812. Kurokawa, H. et al. *J. Mol. Biol.* 1995, **254**:196-207. MacGillivray, R.T.A. et al. *J. Biol. Chem.* 1983, **258**:3543-3549 Schaeffer, E. et al. *Gene* 1987, **56**:109-116. Yang, F. et al. *Proc. Nat. Acad. Sci. USA* 1984, **81**:2752-2756.

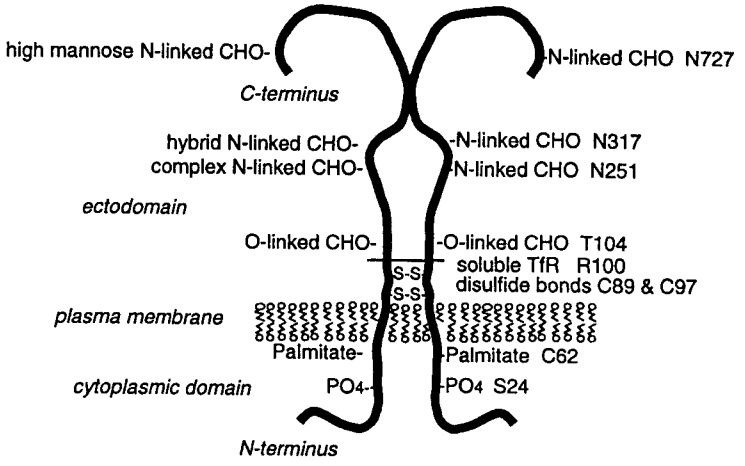

~100 Å
N-terminal lobe
C-terminal lobe
~ 50 Å
Fe
Carbohydrate

Transferrin Receptor

Caroline A. Enns and Robin A. Warren

Synonyms	Transferrin binding protein
Abbreviations	TfR; TR; TfnR; HUMTFRR
Classification	Plasma membrane receptor
Description	A type-II transmembrane protein consisting of 760aa with two intermolecular disulfide bonds and four carbohydrate side chains: One O-linked at Thr-104 and three N-Linked at: Asn-251, Asn-317, and Asn-727. The N-terminal 62 aa contain the sequence YTRF, necessary for targeting for internalization.
Structure	The transferrin receptor has not been determined. Its tertiary structure remains unknown. FTIR and CD of the ectodomain indicate 56% α-helix, 19% β-sheet, 14% turns and 10% other structure (Hadden, J.M. et al. *FEBS Letters* 1994 **350**:235-239).
Molecular Weight	94,000 (SDS-PAGE)
Sedimentation Coeff.	9.8S (Enns, C.A. and Sussman H.H. *J. Biol. Chem.* 1981 **256**:9820-9823)
Isoelectric Point	6.6 (Wada, H.G. et al. *J. Biol. Chem.* 1979 **254**:12629-12635)
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Inhibitors	None
Biological Functions	Binds Transferrin. Expressed in every proliferating cell and cells which have a differentiated function for iron transport or metabolism, especially rich in the placenta. The transferrin receptor is responsible for the internalization of transferrin, which is the major iron-transport protein in the blood.
Physiology/Pathology	Elevated levels of the soluble transferrin receptor are a marker of iron deficiency anemia, hemolytic anemia and polycythemia (Flowers et al. *J. Lab. Clin. Med.* 1989 **114**:368-377). Since the TfR is a marker of proliferating cells, it is found in cancerous cells.
Degradation	Unknown
Genetics/Abnormalities	The TfR gene, located on human chromosome 3 at q26.2->qter, is 31 Kb in length and contains at least 19 introns.
Half life	24-30 hrs (Rutledge, E.A. et al. *Blood* 1994 **83**:580-586)
Concentration	A soluble form of the transferrin receptor is found in the blood ~5 mg/L in blood. It is the result of cleavage at Arg 100 (Shih, Y.J. et al. *J. Biol. Chem.* 1990 **265**:19077-19081).

Isolation Method	Affinity chromatography on transferrin-Sepharose columns (Turkewitz, A.P. et al. *J. Biol. Chem.* 1988 **254**:8318-8325)
Amino Acid Sequence	MMDQARSAFS NLFGGEPLSY TRFSLARQVD GDNSHVEMKL AVDEEENADN NTKANVTKPK RCSGSICYGT IAVIVFFLIG FMIGYLGYCK GVEPKTECER LAGTESPVRE EPGEDFPAAR RLYWDDLKRK LSEKLDSTDF TSTIKLLNEN SYVPREAGSQ KDENLALYVE NQFREFKLSK VWRDQHFVKI QVKDSAQNSV IIVDKNGRLV YLVENPGGYV AYSKAATVTG KLVHANFGTK KDFEDLYTPV NGSIVIVRAG KITFAEKVAN AESLNAIGVL IYMDQTKFPI VNAELSFFGH AHLGTGDPYT PGFPSFNHTQ FPPSRSSGLP NIPVQTISRA AAEKLFGNME GDCPSDWKTD STCRMVTSES KNVKLTVSNV LKEIKILNIF GVIKGFVEPD HYVVVGAQRD AWGPGAAKSG VGTALLLKLA QMFSDMVLKD GFQPSRSIIF ASWSAGDFGS VGATEWLEGY LSSLHLKAFT YINLDKAVLG TSNFKVSASP LLYTLIEKTM QNVKHPVTGQ FLYQDSNWAS KVEKLTLDNA AFPFLAYSGI PAVSFCFCED TDYPYLGTTM DTYKELIERI PELNKVARAA AEVAGQFVIK LTHDVELNLD YERYNSQLLS FVRDLNQYRA DIKEMGLSLQ WLYSARGDFF RATSRLTTDF GNAEKTDRFV MKKLNDRVMR VEYHFLSPYV SPKESPFRHV FWGSGSHTLP ALLENLKLRK QNNGAFNETL FRNQLALATW TIQGAANALS GDVWDIDNEF
Disulfides/SH-Groups	Transferrin receptor monomers are connected by intermolecular disulfide bonds at Cys-89 and Cys-98, though it is not required for dimer formation.
General References	Enns, C.A. et al. *Biomembranes* 1996, **4**:255-287. Testa U. et al. *Crit. Rev. Oncog.* 1993 **4**:241-276. Kemp, J.D. *Histol. Histopathol.* 1997 **12**:291-296.
Ref. for DNA/AA Sequences	McClelland, A. et al. *Cell* 1984, **39**:267-274. Schneider, C. et al. *Nature* 1984, **311**:675-678. GenBank Accession Numbers: Human Acc# M11507 Rat (Partial cds) Acc# M58040 Chicken Acc# X55348 Mouse Acc# X57349 Hamster Acc# L19142



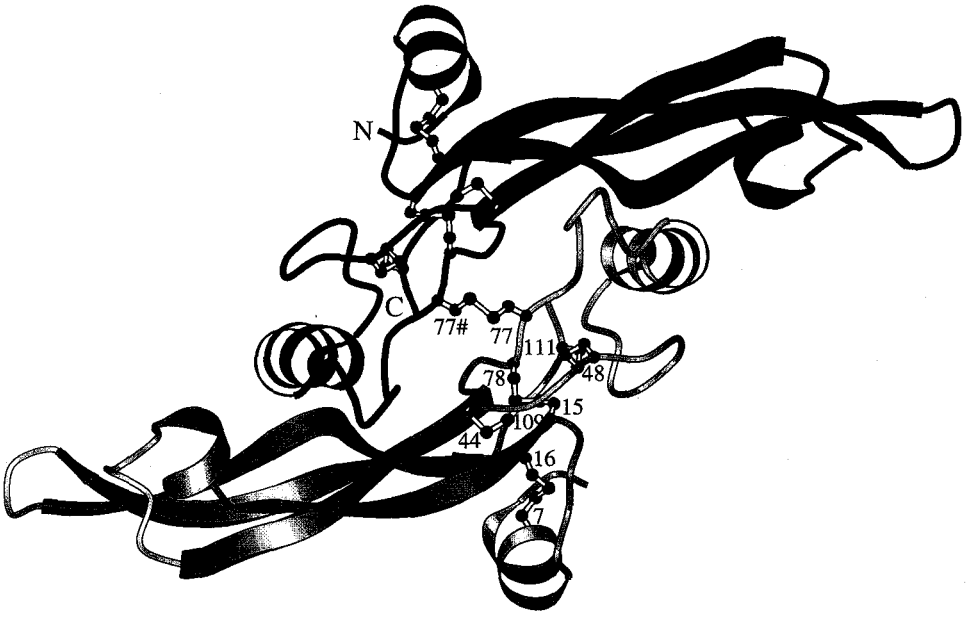
Schematic of the structure of the human transferrin receptor

Transforming Growth Factor - beta 3 (1, 2)

Peer R.E. Mittl and Markus G. Grütter

Synonyms	Milk derived growth factor
Abbreviations	TGF-β3 (1, 2)
Classification	None
Description	A group of multifunctional cytokines that are synthesized as a precursor of 412 aa, which is activated by proteolytic cleavage at the RKKR site. The mature cytokine comprises a disulfide-linked dimer with 112 aa per chain. There are 3 isoforms that share approximately 70-80% sequence identity.
Structure	The overall dimensions of the dimer are 60 x 20 x 15 Å. There are 3 α-helices and 4 distorted β-sheets. TGF-βs belong to the family of disulfide-knot proteins. Disulfide-bridge Cys-15-Cys-78 passes through a ring that is formed by Cys-44-Cys-109 and Cys-48-Cys-111 from adjacent β-sheets. The disulfide bridge Cys-77-Cys-77# links the subunits of the dimer and Cys-7-Cys-16 connects the N-terminal α-helix to the central disulfide-knot.
Molecular Weight	25,000 Da (mature dimer)
Sedimentation Coeff.	Unknown
Isoelectric Point	6.2 [TGF-β3]
Extinction Coeff.	45,150 M^{-1} cm^{-1} (280nm) [TGF-β3]
Enzyme Activity	None
Coenzymes	None
Substrates	None
Inhibitors	TGF-β activity is modulated by decorin and biglycan.
Biological Functions	TGF-β signaling involves three membrane bound receptor types. Recruitment of TGF-β to the type II receptor enables the phosphorylation of the cytoplasmic domain of the type I receptor and subsequent signal propagation. The type III receptor, which is a proteoglycan, has no cytoplasmic domain, but effects signal transduction by its ability to present TGF-β to the type II receptor.
Physiology/Pathology	TGF-βs are involved in differentiation, cell proliferation and tissue morphogenesis. They are differently expresed during embriogenesis. During later stages of development they have been detected in most tissues. The three isoforms are often coexpressed although the temporal and spatial patterns are distinct. Overproduction of TGF-β is a problem in many fibrotic diseases, liver cirrhosis and scaring.
Degradation	In the serum TGF-β is bound to macroglobulin and cleared by the liver. TGF-β / receptor-complexes are internalized and degraded by lysosomal enzymes.

Genetics/Abnormalities	The genes for TGF-β1, -2 and -3 map to chromosomes 19q13, 1q41 and 14q24, repectively. Each of these genes comprises 7 exons. The intron-exon splice junctions, with the exception of the first intron, are localized at exactly corresponding positions although the genes are located on different chromosomes. Mice which lack a functional TGF-β3 gene die shortly after birth.
Half-life	90 min. [TGF-β1], 30 min. [TGF-β3]
Concentration	Bone: 0.3 mg/kg; soft tissue (placenta, kidney): 0.004 mg/kg.
Isolation Method	Cation-exchange chromatography of diluted bovine milk, low-pressure hydrophobic interaction chromatography on phenyl-Sepharose, reversed phase HPLC first on butyl- and second on an octyl-polyol resin followed by size-exclusion chromatography.
Amino Acid Sequence	TGF-β1 ALDTNYCFSS TEKNCCVRQL YIDFRKDLGW KWIHEPKGYH ANFCLGPCPY IWSLDTQYSK VLALYNQHNP GASAAPCCVP QALEPLPIVY YVGRKPKVEQ LSNMIVRSCK CS TGF-β2 ALDAAYCFRN VQDNCCLRPL YIDFKRDLGW KWIHEPKGYN ANFCAGACPY LWSSDTQHSR VLSLYNTINP EASASPCCVS QDLEPLTILY YIGKTPKIEQ LSNMIVKSCK CS TGF-β3 ALDTNYCFRN LEENCCVRPL YIDFRQDLGW KWVHEPKGYY ANFCSGPCPY LRSADTTHST VLGLYNTLNP EASASPCCVP QDLEPLTILY YVGRTPKVEQ LSNMVVKSCK CS The TGF-β superfamily also contains the activins/inhibins, bone morphogenic proteins and Mullerian inhibiting substance.
Disulfides/SH-Groups	5 Disulfides: Cys-7-Cys-16, Cys-15-Cys-78, Cys-44-Cys-109, Cys-48-Cys-111, Cys-77-Cys-77# (linkage between subunits); no free sulfhydryl groups.
General References	Border W.A. and Ruoslahti E. *J. Clin. Invest.* 1992, **90**:1-7. Cox, D. *Cell Biol. Int.* 1995, **19**:357-371. Lawrence, D.A. *Eur. Cytokine Netw.* 1996, **7**:363-374. Massagué, J. *Annu. Rev. Cell. Biol.* 1990, **6**:597-641. Massagué, J. *Cell* 1996, **85**:947-950. Roberts A.B. and Sporn, M.B. In: *Peptide Growth Factors and Their Receptors I.* Sporn, M.B. and Roberts, A.B. (eds.), Springer-Verlag, 1990, pp. 419-472.
Ref. for DNA/AA Sequences	de Martin, R. et al. *EMBO J.* 1987, **6**:3673-3677. Derynck, R. et al. *Nature* 1985, **316**:701-705. Derynck, R. et al. *EMBO J.* 1988, **7**:3737-3743.



Two subunits of mature TGF-β3 homodimer are shown in light and dark gray. The disulfide bridges and the N- and C-termini are labeled. The tertiary structures of TGF-β1 (PDB code: 1KLC), TGF-β2 (1TFG, 2TGJ) and TGF-β3 (1TGJ, 1TGK) possess a very similar topology but differences exist in the N-terminal α-helix and in the orientation of the β-sheet loops.

Transthyretin

Maria J. Saraiva

Synonyms	Prealbumin
Abbreviations	TTR; PA
Classifications	Prealbumin mobility in electrophoresis pH 8.6
Description	A protein synthesized in the liver, choroid plexuses of brain and the eye, comprised by four identical subunits of 127 aa each, with no carbohydrate, and having binding sites for thyroxine (T_4), retinol binding protein (RBP) and polyhalogenated aromatic hydrocarbons (PAH).
Structure	Each monomer organize as two four-stranded betasheets. Antiparallel beta-sheet interactions between monomers results in two stable dimers, which assemble to form the tetramer. The TTR quaternary structure presents a central channel that runs through the tetramer in which are located two symmetry-related high affinity binding sites for iodothyronine molecules and flavonoids.
Molecular Weight	54,980 (aa analysis)
Sedimentation Coeff.	3.75
Isoelectric Point	5.7: major monomeric form (urea)
Extinction Coeff.	14.06 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	A transport protein. Approximately 15% of the total thyroxine in plasma circulates bound to TTR. In the cerebrospinal fluid (CSF), TTR is the major carrier for T_4. TTR is also an important component of the transport complex for retinol, through its association with RBP. RBP circulates in plasma mainly as a 1:1 molar protein-protein complex with TTR. The RBP so transported, serves to mobilize retinol from its stores in the liver and deliver retinol to tissues throughout the body. The formation of the TTR-RBP complex serves to protect RBP by preventing its glomerular filtration and subsequent renal catabolism. The formation of this complex also increases the affinity and stability of the binding of retinol to RBP.
Physiology/Pathology	High levels of TTR in the plasma or an increased affinity for T_4 can lead to euthyroid hyperthyroxinemia, characterized by high total T_4 concentrations. Binding of polyhalogenated aromatic hydrocarbons to TTR occurs and is thought both to facilitate toxicant transport to their tissues sites of action and to disrupt normal retinoid and thyroid hormone homeostasis, contributing to the pathogenic effects of the toxicants. By still unknown mechanisms, TTR can deposit extracellularly in several organs and tissues as amyloid. A number of different point mutations in TTR have been

	identified in patients with familial amyloidosis, associated either with neuropathies (FAP) or cardiopathies (FAC).
Degradation	The protein is thought to be mainly cleared by liver.
Genetics/Abnormalities	The gene is localized on chromosome 18, is 7 kb long and has 4 exons. About 40 different single base substitutions have been identified; a small number are non-pathogenic or associated with hyperthyroxinemia. The vast majority are from patients with FAP and FAC.
Half-life	2 day
Concentration	0.25 g/L in plasma, with the level decreasing during acute phase reactions. In cerebrospinal fluid, TTR is highly concentrated relative to total protein.
Isolation Method	(a) ion exchange chromatography on DEAE-cellulose; (b) chromatography on Cibacron Blue sepharose; (c) gel filtration
Amino Acid Sequence	See Kanda, Y. et al. *J. Biol. Chem.* 1974, **249**:6796-6805.
Disulfides/SH-Groups	No disulfides, 1 sulfhydryl at residue 10 of each subunit.
General References	Robbins, J. et al. *Rec. Prog. Horm. Res.* 1976, **34**:477-519. Blake, C.C.F. et al. *J. Mol. Biol.* 1978, **121**:339-356. Goodman, D.S. *The Harvey Lectures* 1986, **81**:111-132. Saraiva, M.J.M. et al. "Transthyretin and familial amyloidotic polyneuropathy". In: *The molecular and genetic basis of neurological disease.* Rosenberg, R.N. et al. (eds.) Butterworths, 1993, pp 889-894. Saraiva, M.J.M. Transthyretin mutations in health and disease. *Human Mutation* 1995, **5**:191-196.
Ref. for DNA/AA Sequences	Kanda, Y. et al. *J. Biol. Chem.* 1974, **249**:6796-6805. Sasaki, H. et al. *Gene* 1985, **37**:191-197.

TRH Receptor

Patricia M. Hinkle

Synonyms	Thyrotropin releasing hormone (TRH) receptor; Thyrotropin releasing factor (TRF) receptor; Thyroliberin receptor; Protirelin receptor
Abbreviations	TRHR
Classifications	None
Description	A glycosylated, transmembrane protein, a member of the G protein-coupled receptor family, synthesized in thyrotrophs and lactotrophs of the anterior pituitary gland and in other tissues. When activated by TRH, the TRHR stimulates phospholipase C, mobilizing intracellular calcium, with signal transduction via heterotrimeric G proteins in the Gq class.
Structure	The structure of the TRHR has not been determined. The receptor is a member of the rhodopsin-like receptor family characterized by an extracellular, N-glycosylated N-terminus, seven transmembrane regions, and an intracellular C-terminal tail with potential palmitoylation and phosphorylation sites. Sequences of rat, human, mouse and sheep TRHRs are highly conserved.
Molecular Weight	Calculated protein molecular weight of human TRHR is 45,085, but molecular weight of the glycosylated, posttranslationally modified receptor is unknown.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown, calculated pI of the unmodified protein is 8.62
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Chlordiazepoxide acts as a weak competitive antagonist or inverse agonist.
Biological Functions	The pituitary TRHR responds to the hypothalamic tripeptide TRH. TRH acts on TRHR on thyrotrophs and lactotrophs in the anterior pituitary gland, causing the release and synthesis of thyrotropin and prolactin. The function of the TRHR in other tissues is less well understood.
Physiology/Pathology	TRHR regulates secretion of the pituitary hormone thyrotropin, which controls the function of the thyroid gland. In hypothyroidism, TRH rises and activates the TRHR; this increases circulating thyrotropin, which in turn activates thyroid hormone production. In hyperthyroidism, the opposite occurs. Absence of functional TRHR causes hypothyroidism.
Degradation	Unknown
Genetics/Abnormalities	TRHR gene located on human chromosome 8q23. Central hypothyroidism was seen in a child who had one completely nonfunctional TRHR and one apparently inactive TRHR.
Half-life	Unknown

Concentration	Unknown, but a minor membrane protein
Isolation Method	Not yet purified
Amino Acid Sequence	MENETVSELN QTQLQPRAVV ALEYQVVTIL LVLIICGLGI VGNIMVVLVV MRTKHMRTPT NCYLVSLAVA DLMVLVAAGL PNITDSIYGS WVYGYVGCLC ITYLQYLGIN ASSCSITAFT IERYIAICHP IKAQFLCTFS RAKKIIIFVW AFTSLYCMLW FFLLDLNIST YKDAIVISCG YKISRNYYSP IYLMDFGVFY VVPMILATVL YGFIARILFL NPIPSDPKEN SKTWKNDSTH QNTNLNVNTS NRCFNSTVSS RKQVTKMLAV VVILFALLWM PYRTLVVVNS FLSSPFQENW FLLFCRICIY LNSAINPVIY NLMSQKFRAA FRKLCNCKQK PTEKPANYSV ALNYSVIKES DHFSTELDDI TVTDTYLSAT KVSFDDTCLA SEVSFSQS
Disulfides/SH-Groups	Unknown, but a disulfide between cysteines in the first and second extracellular loops is predicted based on homology with the rat TRHR.
General References	Jackson, I.M. *N. Eng. J. Med.* 1982 **306**:145-155. Gershengorn, M.C. and Osman, R. *Physiol. Rev.* 1996, **76:**175-191. Hinkle, P.M. et al. *Trends Endocrinol. Metab.* 1996, **7:**370-374. Collu, R. et al. *J. Clin. Endocrinol. Metab.* 1997, **82**:1361-1364. Straub, R.E. et al. *Proc. Natl. Acad. Sci. U.S.A.* 1990, **87**:9514-9518.
Ref. for DNA/AA Sequences	Matre, V. et al. *Biochem. Biophys. Res. Commun.* 1993 **195**:179-185. Hinuna, S. et al. *Biochim. Biophys. Acta* 1994, **1219**:251-259.

Tripeptidyl-peptidase II

Birgitta Tomkinson

Synonyms	Tripeptidyl aminopeptidase
Abbreviations	TPP II
Classifications	EC 3.4.14.10; exopeptidase
Description	A high molecular weight cytosolic exopeptidase, present in a number of different tissues. It is oligomeric, and is built up by identical subunits. The oligomeric structure seems to be essential for full enzymatic activity, since the activity is lost in parallel with dissociation. The enzyme is classified as a serine peptidase and has an active site of the subtilisin-type. It is, however, also dependent on at least one cysteine residue for activity and/or stability. Recently, a membrane-associated form of the enzyme was identified.
Structure	Electron-microscopy on samples negatively stained by ammonium molybdate show oligomeric structures (20x50 nm), built up by 14-18 smaller units (3x10 nm). It is not known if the smaller units correspond to the subunit or a dimer of subunits.
Molecular Weight	Native enzyme > 10^6 (gel filtration). Subunit 138,300 (aa sequence).
Sedimentation Coeff.	Unknown
Isoelectric Point	6.2 (9M urea); 6.28 (aa sequence)
Extinction Coeff.	Unknown
Enzyme Activity	Peptide hydrolases; tripeptidylpeptide hydrolases. Removes tripeptides, with little apparent similarity, from the free N-terminus of longer peptides. Proline is not accepted in the P_1- or the P_3-position. Neutral pH-optimum.
Coenzymes/Cofactors	None
Substrates	The standard substrates used for measurements of enzymatic activity are the ^{32}P-labelled, phosphorylated hexapeptide Arg-Arg-Ala-Ser(^{32}P)-Val-Ala and the chromogenic substrates Ala-Ala-Phe-aminomethyl coumarin and Ala-Ala-Phe-*para*-nitroanilide.
Inhibitors	Natural inhibitors have not been found. TPP II is inhibited by serine peptidase inhibitors, such as diisopropyl fluorophosphate and phenylmethanesulfonyl fluoride. Some thiol-reactive compounds, e.g. N-ethylmaleimide, p-chloromercuribenzoate and 2,2'-dithiodipyridine, are efficient inhibitors.
Biological Functions	The broad substrate specificity and widespread distribution of the cytosolic enzyme would indicate a role in the general intracellular proteolysis. The formation of tripeptides is likely to facilitate the action of other exopeptidases, thereby providing an efficient degradation of longer peptides into amino acids. The membrane-associated variant has been suggested to be involved in inactivation of cholecystokinin (CCK-8).
Physiology/Pathology	No variation in intracellular concentration has been detected.

1. Installment 1992 (revised 1997)

Degradation	Unknown
Genetics/Abnormalities	The enzyme is represented by two mRNAs with different length, which probably result from the utilization of two different polyadenylation sites. The gene has been localized to 13q32-q33.
Half-life	Unknown
Concentration	The subunit concentration is about 0.07 μmol L^{-1} in human red blood cells and approximately four times higher in liver cells.
Isolation Method	Routinely prepared from lysates of outdated human erythrocytes by chromatography on two succesive ion exchangers (DEAE-cellulose), Sepharose CL-4B, and hydroxylapatite.
Amino Acid Sequence	The protein subunit consists of 1249 aa residues. The N-terminal part shows a high degree of similarity to the bacterial endopeptidase subtilisin. In particular, the residues surrounding the active site catalytic triad (Ser, His, Asp) appears to be conserved. In addition, a splicing variant containing 39 bp extra, encoding 13 aa in the C-terminal part of the enzyme, has been identified.
Disulfides/S-H-Groups	Unknown
General References	Bålöw, R.-M. et al. *J. Biol. Chem.* 1986, **261**:2409-2417. Tomkinson, B.et al. *Proc. Natl. Acad. Sci. USA* 1987, **84**:7508-7512. Macpherson, E. et al. *Biochem. J.* 1987, **248**:259-263. Tomkinson, B., Zetterqvist, Ö. *Biochem. J.* 1990, 267:149-154. Martinsson, T. et al. *Genomics* 1993, **17**:493-495. Wilson, C. et al. *Neurochem. Res.* 1993, **18**:743-749. Tomkinson, B. *Biochem. J.* 1994, **304**:517-523. Rose, C. et al. *Nature* 1996, **380**:403-409.
Ref. for DNA/AA Sequences	Tomkinson, B. and Jonsson, A.K. *Biochemistry* 1991, **30**:168-174. Tomkinson, B. *Biomed. Biochim. Acta* 1991, **50**:727-729. GenBank accession number: J05299.

Troponin C

Monica X. Li and Brian D. Sykes

Synonyms	Calcium-dependent muscle regulatory protein; Troponin A
Abbreviations	TnC
Classifications	Ca^{2+}-binding protein
Description	There are two isoforms of TnC; fast skeletal TnC (sTnC), which is expressed exclusively in fast skeletal muscle cells, and slow/cardiac TnC (cTnC), which is expressed in both slow skeletal and cardiac muscle cells. Both isoforms are small, highly conserved, and heat-stable proteins. sTnC consists a single polypeptide chain of 159 aa and cTnC consists a single polypeptide chain of 161 aa. Both isoforms contain two high affinity Ca^{2+}-binding sites (III/IV) (K_d — 10^{-7}M) that also bind Mg^{2+} (K_d —10^{-3} M) and have been called the Ca^{2+}/Mg^{2+} sites. sTnC also contains two sites (I/II) of lower Ca^{2+} affinity (K_d — 10^{-5}M) which have been called the Ca^{2+}-specific sites. cTnC contains only one Ca^{2+}-specific site (II). The N-termini in both isoforms are acetylated.
Structure	The structures of turkey and chicken sTnC were solved at 2.0Å resolution by X-ray crystallography. Both molecules consist of two globular domains separated by an exposed 31-residue α-helix; the length of the molecule is 70Å. The domains have mean radii of 17Å and their centers are separated by 44Å. In both structures, the C-terminal sites (III/IV) are occupied by Ca^{2+}-ions, whereas the N-terminal sites (I/II) have no bound Ca^{2+}. NMR solution structures of the N-terminal domain of chicken sTnC reveal that Ca^{2+}-binding induces the exposure of a large hydrophobic surface. The NMR solution structure of chicken cTnC shows two globular domains with a flexible central linker; both domains are in Ca^{2+}-bound state. NMR solution structures of the N-domain human cTnC show that Ca^{2+}-binding to site II does not induce the exposure of a large hydrophobic surface, unlike the skeletal isoform.
Molecular Weight	sTnC: 17,991 (aa sequence); cTnC: 18,416 (aa sequence)
Sedimentation Coeff.	sTnC: 2.5 S ($CaCl_2$); 1.84 S (EGTA, EDTA) cTnC: 2.04 S ($CaCl_2$); 1.87 S (EGTA, EDTA)
Isoelectric Point	sTnC: 4.2; cTnC: 4.2.
Extinction Coeff.	sTnC: 1.93 (280nm, 1%, 1cm); cTnC: 2.3 (276nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	Ca^{2+}-binding to TnC required for activation of muscle contraction.
Substrates	None
Inhibitors	None
Biological Functions	TnC is the Ca^{2+}-binding member of the troponin complex and plays a key role in the Ca^{2+} regulation of contraction/relaxation in skeletal and cardiac muscles. Conformational changes in TnC induced by Ca^{2+}-binding and dissociation are transmitted through the other members of the troponin

complex, troponin I and troponin T, to tropomyosin-actin, resulting in activation/inhibition of actomyosin ATPase and muscle contraction/relaxation.

Physiology/Pathology

To date, TnC has not been shown to play a role in the pathogenesis of human cardiac disease. However, studies have shown that inappropriate expression of sTnC in the heart might be of clinical relevance since sTnC is unable to mediate the Starling properties normally subserved by cTnC. Although no evidence suggests that sTnC is expressed in the heart, cTnC has been shown to be expressed in adult rabbit fast skeletal muscle after chronic low-frequency electrical stimulation in vivo. Since cTnC is a relatively small protein, it might leak from damaged cardiac myocytes after an ischemic insult. cTnC is a potential target for therapeutic approaches aimed at increasing myocardial performance in congestive heart failure, where the diseased myocradium is "desensitized" to Ca^{2+}. A novel group of positive inotropic agents known as "calcium sensitizers" is known to increase the affinity of cTnC for Ca^{2+}.

Degradation

Susceptible to a number of proteases (e.g. trypsin).

Genetics/Abnormalities

There are two TnC genes in the human genome, one encoding fast skeletal muscle and the other encoding slow/cardiac TnC. The expression of these two genes is developmentally regulated and controlled at the transcriptional and/or post-transcriptional level. The expression of the fast TnC gene is highly fiber type-specific while mRNA for the slow/cardiac isoform is abundant in slow-twitch skeletal and cardiac muscle but also present at lower levels in some non-muscle cells. The cTnC gene is on human chromosome 3 at p21.3 - p14.3, and is 3.4kb long, with 6 exons and 5 introns. The sTnC gene is 2.6 kb long, with 6 exons and 5 introns.

Half-life

Unknown

Concentration

~1% (by weight) in both cardiac and skeletal muscles.

Isolation Method

Troponin (I, C, T complex) is isolated by either muscle powder method or LiCl extraction from muscle mince, followed by ammonium sulfate fractionation. To isolate TnC from troponin involves ion exchange chromatography in urea and ammonium sulfate precipitation in the presence of SDS. Good tissue sources are bovine heart for cTnC and rabbit leg/back for sTnC.

Amino Acid Sequence

The aa sequences of the proposed Ca^{2+}-binding loops for sTnC and cTnC are shown. These structures are part of helix-loop-helix (EF hand) structures common to many Ca^{2+}-binding proteins. (Ca^{2+}-binding residues are indicated by an asterisk.

sTnC:

			* * * * * *
Loop	I	(27-38)	DADGGGDISVKE
Loop	II	(63-74)	DEDGSGTIDFEE
Loop	III	(103-114)	DRNADGYIDPEE
Loop	IV	(139-150)	DKNNDGRIDFDE

cTnC:

			* * * * * *
Loop	I	(28-40)	VLGAEDGCISTKE
Loop	II	(65-76)	DEDGSGTVDFDE
Loop	III	(105-116)	DKNADGYIDLDE
Loop	IV	(141-152)	DKNNDGRIDYDE

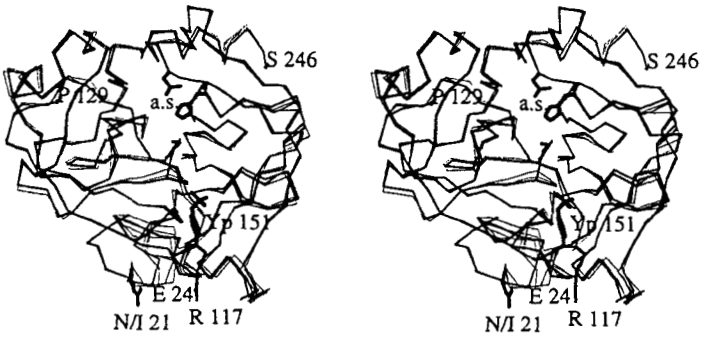
Disulfides/SH-Groups	sTnC: One SH-Group (residue 98); cTnC: Two SH-Groups (residues 35, 84).
General References	Gagné, S.M. et al. *Nat. Struct. Biol.* 1995, **2**:784-789. Leavis, P.C. and Gergely, J. *CRC Crit. Rev. Biochem.* 1984, **16**:235-305. Sia, S.K. et al. *J. Biol. Chem.* 1997, **272**:18216-18221. Slupsky, C.M. and Sykes, B.D. *Biochemistry* 1995, **34**:15953-15964. Spyracopoulos, L. et al. *Biochemistry* 1997, **36**:12138-12146. Strynadka, N.C.J. and James, M.N.G. *Annu. Rev. Biochem.* 1989, **58**:951-998. Tobacman, L.S. *Annu. Rev. Physiol.* 1996, **58**: 447-481.
Ref. for DNA/AA Sequences	Gahlmann, R. et al. *J. Mol. Biol.* 1988, **201**:379-391. Gahlmann, R. and Kedes L. *J Biol. Chem.* 1990, **265**:12520-12528 (accession # 191039, cDNA (sTnC, human). Parmacek, M.S. and Leiden, J. M. *Circulation* 1991, **84**:991-1003. Schreier, T. et al. *J Biol. Chem.* 1990, **265**:21247-21253 (accession # 191040, cDNA (cTnC, human). Song, W.J. et al. *Cytogenet. Cell. Genet.* 1996, **75**:36-37.

Trypsin(ogen) 1 and Trypsin(ogen) 2

Odette M. Guy-Crotte and Catherine G. Figarella

Synonyms	Cationic Trypsin(ogen) for Trypsin(ogen) 1; Anionic Trypsin(ogen) for Trypsin(ogen) 2.
Abbreviations	Tg1 or (T3,TRYI);Ti1 Tg2 or(T1,TRYII);Ti2
Classifications	EC 3.4.21.4
Description	Trypsinogens are proteins of pancreatic external secretion synthesized by pancreas as inactive precursors of zymogens. Trypsinogens are transformed into active trypsins by hydrolysis of the 8^{th} peptide bond and liberation of an octapeptide. This activation is physiologically performed by enterokinase "in vivo" and/or by trypsin. Trypsinogens can also be activated by cathepsin B at acidic pH (3.8 - 5.0).
Structure	One polypeptide chain with 5 disulfide bridges for trypsinogen 1 (determined by aa composition and cDNA sequence) and 4 disulfide bridges (determined by cDNA sequence) or 5 (determined by aa composition) for trypsinogen 2. Trypsins are serine proteinases. The tertiary structure has been determined by X-ray crystal structure analysis of diisopropyl phosphofluoridate (DFP) -inhibited human Trypsin 1. Crystals belong to the space group *P*4, with two independent molecules in the asymmetric unit packing as crystallographic tetramers. Trypsin 1 fold is very similar to those of the bovine rat and porcine trypsins. The unexpected feature of human trypsin 1 structure is the phosphorylated state of tyrosine residue 151.
Molecular Weight	26,000 ± 1,000 (SDS-PAGE); 25,002 (cDNA sequence) for Tg1. 26,000 ± 1,000 (SDS-PAGE); 24,930 (cDNA sequence) for Tg2.
Sedimentation Coeff.	Unknown
Isoelectric Point	6.4 (Tg1); 4.4 (Tg2)
Extinction Coeff.	15.2: (Tg1); 15.1: (Tg2)
Enzyme Activity	Trypsinogens are devoid of enzyme activity. Trypsins hydrolyse proteins, peptides, amides and esters specifically at the carboxyl groups of the basic amino acids L-arginine and L-lysine. Optimum pH 8.0 with 20 mM calcium.
Coenzymes/Cofactors	None
Substrates	N-Benzol-L-arginine ethyl ester, N-Tosyl-L-arginine methyl ester, N-Benzoyl-L-arginine-4-nitroanilide, CBZ-L-Arginine-7-amido-4-methyl coumarin.
Inhibitors	Protease inhibitors: Kunitz pancreatic inhibitor (or aprotinin); human and porcine pancreatic secretory trypsin inhibitors, soybean trypsin inhibitor (40% of inhibition for trypsin 1 and 80% for trypsin 2); ovomucoid (no inhibition for trypsin 1 and 50% of inhibition for trypsin 2); serum trypsin inhibitors: α_1-antitrypsin and α_2-macroglobulin. Synthetic inhibitors: DFP, PMSF, TLCK.

Biological Functions	Trypsins are responsible for the activation of all pancreatic zymogens: chymotrypsinogens, proelastase and proprotease E, prekallikrein, procarboxypeptidases A and B, prophospholipase A_2 and procolipase. Trypsins are endopeptidases that function as proteolytic enzymes in food digestion.
Physiology/Pathology	Trypsins are essential for the activation of pancreatic zymogens. Their absence due to an enterokinase deficiency can lead to severe disturbance in digestion. Serum trypsinogen level is significantly increased in newborns with cystic fibrosis. In acute pancreatitis a significant elevation of serum trypsinogen level is observed with the presence of trypsin bound to α_1-antitrypsin and α_2-macroglobulin.
Degradation	Trypsins may be degraded by autolysis, giving inert proteins in the absence of calcium or different forms of proteolyzed trypsin in the presence of calcium at pH 8.0. In physiological conditions (duodenal juice) trypsins are inactivated after 4-6 hours.
Genetics/Abnormalities	Trypsin 1 is located on the long arm of chromosome 7 (q22-qter).
Half-life	Unknown
Concentration	Tg1 concentration (determined by immunoenzymatic assay): in serum: 22 ± 7 µg/L in adults, 32 ± 13 µg/L in newborns. In pancreatic juice: 18% of secretory proteins. Tg concentrations in pancreatic juice (determined by activity): 12% of secretory proteins for Tg1 and 6% for Tg2.
Isolation Method	Trypsinogens are isolated from human pancreatic juice by filtration on Sephadex G-100 and chromatography on DEAE-Trisacryl at pH 8.0 in the presence of benzamidine and lima bean inhibitor. Trypsins are isolated from activated trypsinogens by affinity chromatography on benzamidine-Sepharose or pancreatic trypsin inhibitor-Sepharose.
Amino Acid Sequence	N-terminal of trypsinogens 1 and 2: APFDDDDKIVG-. Peptide of the active site (Ser 185) of trypsins 1 and 2: CQGDSGGPVVC.
Disulfides/SH-Groups	5 disulfide bridges for trypsinogen 1, 4 or 5 disulfide bridges for trypsinogen 2. No free sulfhydryls.
General References	Figarella, C. et al. *Eur. J. Biochem.* 1975, **53**:457-463. Guy, O. et al. *Biochemistry* 1978, **17**:1669-1675. Scheele, G. et al. *Gastroenterology* 1981, **80**:461-473. Figarella, C. et al. *Biol. Chem. Hoppe Seyler* 1988, **369**:293-298.
Ref. for DNA/AA Sequences	Emi, M. et al. *Gene* 1986, **4**:305-310.



Superposition of the C^{α} atoms of human trypsin1 (thick line) with bovine, porcine, and rat (thinner lines) trypsins. From C. Gaboriaud, L. Serre, O. Guy-Crotte, E. Forest and J.-C. Fontecilla-Camps, *J. Mol. Biol.* 1996, **259**:995-1010. Glu24, Pro129 and the additional C-terminal Ser246 are the most significant divergent positions. The catalytic triad (Ser195, His57, Asp102) is fully displayed at the active site (a.s.)

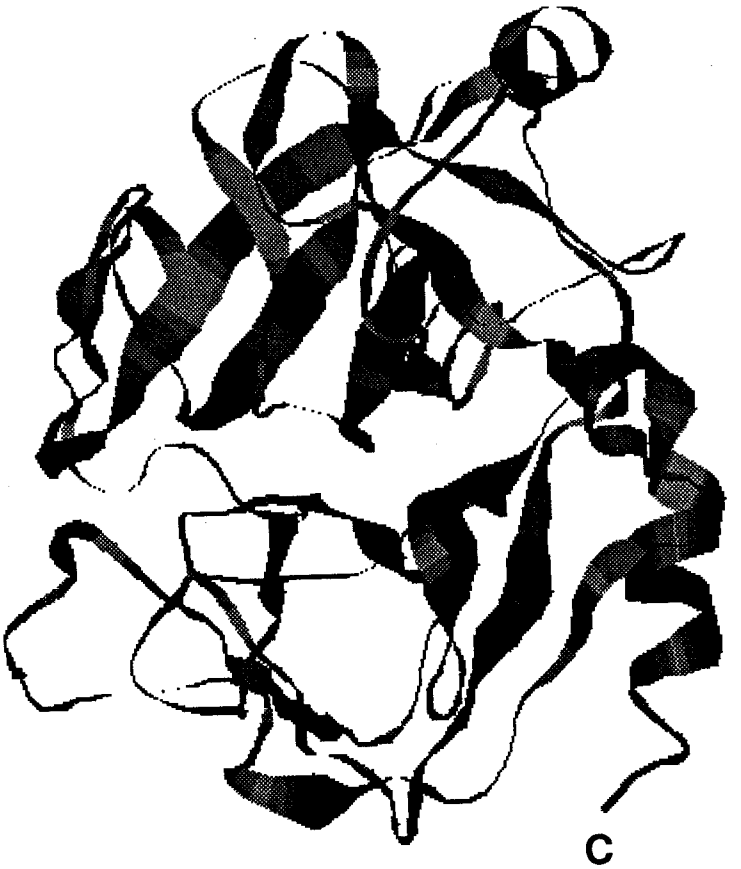
Tryptase

George H. Caughey

Synonyms	Mast cell tryptase; Tryptase I-III; α, β-tryptase
Abbreviations	
Classifications	EC 3.4.21.59
Description	An oligomeric, trypsin-like serine proteinase. It is the major protein of human mast cell secretory granules. Small amounts also reside in basophils. Highest tissue levels are observed in the lung. The enzyme is released from mast cells in association with heparin proteoglycan, which stabilizes the noncovalent tetrameric association of subunits, each of which contains an active site. Tryptase monomers are inactive. Subunits are single polypeptide chains of approx. 245 aa and are N-glycosylated at one or two sites (Asn-102 and Asn-203). Subunit molecular weight varies with extent of glycosylation. Purified enzyme typically resolves to two main bands on SDS-PAGE. Unusual features include strong resistance to circulating antiproteinases and extracellular release of the enzyme in a fully active form, rather than as a zymogen.
Structure	Tertiary and quaternary structures are not yet determined. Modeling based on the structure of other serine proteases predicts an active site and substrate binding region formed by the cleft between twin β-barrel domains. Circular dichroic spectra of the native enzyme suggest the presence of 60% β-strand and 40% random coil.
Molecular Weight	120,000-150,000: native tetramer (gel filtration); 30,000-37,000: denatured monomers (SDS-PAGE).
Sedimentation Coeff.	Unknown
Isoelectric Point	5.6-7.2 (heterogeneous bands)
Extinction Coeff.	28.1 (280nm, 1%, 1cm), each monomer has 9 Trp residues.
Enzyme Activity	Trypsin-like endoprotease and esterase with optimal activity at pH 7-8. Preferentially cleaves at the carbonyl end of basic residues, Arg > Lys >> His.
Coenzymes/Cofactors	Heparin glycosaminoglycan or proteoglycan, which stabilizes the activity and the tetrameric configuration of subunits.
Substrates	Tosyl-L-Arg-methyl ester (TAME) and thiobenzyl esters or aminomethylcoumarin amides of peptides ending in a dibasic pair of aa, e.g. Lys-Arg, are hydrolyzed extremely efficiently. The best 4-nitroanilide substrate is benzyloxycarbonyl-L-Lys-Gly-L-Arg-4-nitroanilide. A weaker but useful (and commercially available) substrate is tosyl-L-Gly-Pro-Arg-4-nitroanilide. Notwithstanding the results with small synthetic substrates, tryptase does not favor hydrolysis at the carbonyl end of a pair of basic residues in natural peptides examined to date. The best reported natural peptide substrate of tryptase is calcitonin gene-related peptide. Tryptase has limited activity towards most potential protein targets, but in the presence of heparin it destroys high molecular weight kininogen, C3 and C3a, and fibrinogen. Tryptase also activates prostromelysin and pro-urokinase in vitro.

Inhibitors	No potent extracellular, physiological inhibitors of tryptase are known. Within the secretory granule, catalytic activity may be limited by low pH and by high local concentrations of histamine and heparin. Unlike most trypsin-like serine proteases, tryptase resists inactivation by α_1-proteinase inhibitor, soybean trypsin inhibitor, and aprotinin. Tryptase also is not inhibited by α_2-macroglobulin, probably because of the large size of the heparin-associated tetramer. However, tryptase is inactivated by several low molecular weight inhibitors, including leupeptin, benzamidine, diisopropylfluorophosphate, various chloromethyl and fluoroalkyl ketone derivatives of basic aa by APC-366 and by leech-derived tryptase inhibitor. The most potent competitive inhibitors of tryptase are bis(5-amidino-2-benzimidazolyl)methane (BABIM) and related compounds ($K_i \approx 1$ nM). Tryptase activity also is reversibly inhibited by high concentrations of various salts, including NaCl and $CaCl_2$.
Biological Functions	Unknown; probably plays an extracellular role though an intracellular role has not been disproved. In vitro data suggest that tryptase may modulate neurogenic inflammation by degrading inflammatory neuropeptides and, by promoting local anticoagulation, may facilitate the entry of inflammatory cells and proteins into tissues following release of mast cell inflammatory mediators.
Physiology/Pathology	Roles unknown; possibly increases bronchoconstriction in asthma by destroying bronchodilating peptides, may promote wound healing and fibrosis by stimulating the growth of fibrolasts, and may participate in tissue inflammation and remodeling by activating prostromelysin and prourokinase. Tryptase inhibitors reduce bronchoconstriction in a sheep model of asthma. Appearance of tryptase in the bloodstream following severe immediate hypersensitivity reactions is a clinical and forensic marker of anaphylaxis. Tryptase levels in tears, skin blisters, or in lung or nasal lavage fluids reflect extent of local mast cell activation in response to antigen and other stimuli.
Degradation	Unknown
Genetics/Abnormalities	Tryptase monomers are products of a multigene family mapping to chromosome 16. The total number of tryptase genes is not known. The differences between several of the cloned forms of human tryptase are small; some may be allelic variants. Four different tryptase cDNA's (tryptase α, I, II/β, and III) and one complete tryptase gene (encoding tryptase I) have been reported. The organization of the gene differs from that of other serine proteases, particular in the placement of its first intron. The sequences predict the presence of propeptide ending in Gly, a finding that is unique among known serine protease proenzymes. Propeptide removal appears to require the sequential action of tryptase itself and of dipeptidyl peptidase I. α-Tryptase may be catalytically inactive because of mutations in the propeptide. Whether there are tissue- or mast cell subset-specific differences in tryptase gene expression is not known.
Half-life	$\approx$ 2 h (circulating immunoreactive tryptase)
Concentration	10-35 pg per mast cell; local tissue levels, which vary according to local numbers and subtypes of mast cells, can exceed 1 g per kg of tissue. Hundreds of μg per L of immunoreactive tryptase can appear in the serum in the first 2 h following systemic anaphylaxis, especially after insect stings. Normal serum levels are $< 10\ \mu g\ L^{-1}$. The half-life of active enzyme following mast cells degranulation in vivo is not known.

Isolation Method	Several schemes have been used, typically involving a) extraction of tissue homogenates in high ionic strength buffers; b) fractionation in ammonium sulphate; and c) chromotography on hydrophobic interaction and heparin affinity media.
Amino Acid Sequence	IVGGQEAPRS KWPWQVSLRV HGPYWMHFCG GSLIHPQWVL TAA**H**CVGPDV KDLAALRVQL REQHLYYQDQ LLPVSRIIVH PQFYTAQIGA **D**IALLELEEP VNVSSHVHTV TLPPASETFP PGMPCWVTGW GDVDNDERLP PPFPLKQVKV PIMENHICDA KYHLGAYTGD DVRIVRDDML CAGNTRRDSC QGD**S**GGPLVC KVNGTWLQAG VVSWGEGCAQ PNRPGIYTRV TYYLDWIHHY VPKKP Sequence is that of mature tryptase I, showing "Catalytic triad" residues (in bold) common to the active site of serine proteinases.
Disulfides/SH-Groups	4 intrachain Cys-Cys pairs, by analogy to trypsin.
General References	Neutral Proteases of Mast Cells, Schwartz, L.B. (vol. ed.), Vol. 27 of *Monographs in Allergy*, Hanson, L.A. and Shakib, F. (series eds.), Karger, Basel, 1990. Mast Cell Proteases in Immunoloy and Biology, Caughey, G.H. (vol. ed.), **Vol. 6** of *Clinical Allergy and Immunology*, Kaliner, M.A. (series ed.), Marcel Dekker, New York, 1995. Johnson, D.A. and Barton G.J. *Protein Science* 1992, **1**:370-377.
Ref. for DNA/AA Sequences	1.) α-Tryptase cDNA: GenBank accession M30038. Miller, J.S. et al. *J. Clin. Invest.* 1989, **84**:1188-1195. 2.) Tryptase I, II and III cDNA and Tryptase I gene: GenBank accession M33491, M33492, M33493 and M33494. Vanderslice, P. et al. *Proc. Natl. Sci. USA* 1990, **87**:3811-3815. 3.) β-Tryptase cDNA (same as Tryptase II): GeneBank accession M37488. Miller, J.S. et al. *J. Clin. Invest.* 1990,**86**:864-870.


C

Tryptophan hydroxylase

Sylviane Boularand and Jacques Mallet

Synonyms	Tryptophan-5-monooxygenase, Tryptophan-5-hydroxylase.
Abbreviations	TPH, TPOH, TPOHase, TRPOHase
Classifications	EC 1.14.16.4
Description	Tryptophan hydroxylase catalyses the first step of the synthesis of serotonine in neurons of the central and peripheral nervous system, in enterochromaffin cells and in carcinoid tumor. In the pineal gland, serotonine is further converted into melatonine.
Structure	There exists only one report concerning the purification of human tryptophan hydroxylase, from carcinoid tumor, so few biophysical parameters are available on human protein. In rat, TPH is an homotetramer.
Molecular Weight	The molecular weight of the subunit is 50,952 Da, according to the nucleotide sequence translated into protein.
Sedimentation Coeff.	Unknown
Isoelectric Point	7.47 (calculated from the aa sequence)
Extinction Coeff.	Unknown
Enzyme Activity	TPH catalyses the formation of L-5-hydroxytryptophan from L-tryptophan, the initial step in the biosynthesis of serotonine and melatonine.
Coenzymes/Cofactors	Fe^{2+} and tetrahydrobiopterin.
Substrates	The hydroxylation of L-Tryptophan is coupled to the oxydation of tetrahydrobiopterin with molecular oxygen. This reaction produces L-5-hydroxytryptophan, dihydrobiopterin and water.
Inhibitors	O-phenanthroline is a potent inhibitor when assayed on human TPH purified from carcinoid tumor. And parachlorophenylalanine is a competitif inhibitor of rat TPH in vitro.
Biological Functions	TPH activity controls the synthesis of serotonine and is also a marker for serotoninergic neurons. In the pineal gland, TPH catalyses the first step of the synthesis of melatonine.
Physiology/Pathology	Lower activity of TPH has been described in the brain of Parkinsonian patients, and in the brain of patients with Senile Dementia of Alzheimer Type. Over production of serotonine and high TPH activity have been found in carcinoid tumors. In the pineal gland melatonine is synthesized during the night, and TPH activity exhibits a circadian cycle.
Degradation	Unknown
Genetics/Abnormalities	Chromosomal localization: 11p15 → p14. Two RFLPs of human TPH have been used to study the genetic implication of TPH in Manic Depressive Illness by linkage and association analyses. This first study suggests that TPH does not play a major role in the susceptibility of Manic Depressive Illness.

Half-life	Unknown
Concentration	Unknown
Isolation Method	A six-fold purification was achieved with differential ammonium sulfate precipitation from human carcinoid tumor.
Amino Acid Sequence	The nucleotide sequence encodes a protein of 444 aa. Three putative phosphorylation sites are conserved between rat and human TPH. Ser-58 may be phosphorylated by cAMP-dependent protein kinase, and Ser-260 and Ser-443 may be phosphorylated by Ca^{2+}-calmodulin-dependent protein kinase. This protein exhibits high homologies with other aromatic aa hydoxylases such as phenylalanine hydroxylase and tyrosine hydroxylase.
Disulfides/SH-Groups	Unknown
General References	Hosoda, et al. *Biochim. Biophys. Acta* 1977, **482:** 27–34. Yamaguchi, et al. *Biochem. Int.* 1981, **2:** 295–303. Sawada et al. *J. Neural. Trans.* 1985, **62:** 107–115. Sawada et al. *J. Neurochem.* 1987, **48:** 760–764.
Ref. for DNA/AA Sequences	Boularand, S., et al. *Nucl. Acids Res.* 1990, **18:** 4257. Human TPH is accessible in EMBL Data Bank under the mnemonic HSWHYDR.

Tubulin

José Manuel Andreu

Synonyms	
Abbreviations	Tub; Tb
Classifications	None
Description	A cytoskeletal GTP-binding protein present in most types of eukaryotic cells. The αβ-tubulin heterodimer (noncovalently associated homologous subunits) assembles forming microtubules. γ-tubulin is required for microtubule nucleation at human centrosomes and microtubule organizing centres of other species. Each subunit consists of different isotypes, which are highly conserved in evolution. Additional microheterogeneity results from post-translational modification: glutamylation, glyosylation, α-tubulin detryosination/tyrosylation and acetylation, β-tubulin posphorylation. There are about 6 α- and 12 β-tubulin isoforms in human brain. Gene sequences and the limited biochemical data available on human tubulin/microtubules (from brain, platelets, leucocytes, sperm, cultured fibroblasts and human cell lines) are complemented by data from homologous mammalian/vertebrate tubulins. The only known relative of tubulin is the bacterial cel division protein FtsZ.
Structure	Each tubulin monomer has a roughly ellipsoidal, irregular shape, and apparently consists of two or three domains. The high resolution structure of tubulin is unknown at the time of this revision. Circular dichroism, infrared spectroscopy and secondary structure prediction indicate approximately 33% α-helix and 24% β-sheet structure in the tubulin dimer.
Molecular Weight	αβ-tubulin dimer: 100,000. Each subunit: 50,000 (aa sequence); 110,000 and 55,000 respectively from sedimentation equilibirum. The two chains can be separated by SDS-PAGE, where their apparent molecular weights are 56,000 to 51,000.
Sedimentation Coeff.	5.8 ± 0.2 S
Isoelectric Point	5.3 - 5.8 in IEF (8M urea, 2% Nonidet P40)
Extinction Coeff.	11.6: native protein in neutral aqueous buffer, light scattering corrected (276nm, 1%, 1cm) 10.9: in 6M GuHCL (275nm, 1%, 1cm) 10.7: in 1% SDS (275nm, 1%, 1cm) All values are given for calf brain tubulin.
Enzyme Activity	As a result of assembly one molecule of bound GTP per tubulin dimer is hydrolysed to GDP, which can only be exchanged for GTP in the disassembled protein. A slow GTPase activity of unassembled tubulin is induced by binding of colchicine and other compounds.
Coenzymes/Cofactors	GTP bound to the α subunit is non-exchangeable; the exchangeable GTP binding site is at the β subunit. The γ-phosphate group and Mg^{2+} at the exchangeable site, and at lower affinity cation binding sites, are required for microtubule assembly, which is inhibited by Ca^{2+}.
Substrates	GTP

Inhibitors	Assembly inhibitors: tubulin is the cellular target of antitumour and other drugs which inhibit cell proliferation by impairing the mitotic spindle. Main inhibitors of microtubule assembly are the vinblastine, colchicine, and nocodazole families of compounds. Taxol is a powerful inducer of microtubule assembly, which also results in mitotic arrest. A few endogenous microtubule inhibitory proteins are known.
Biological Functions	Assembly of microtubules, long hollow structures of 24 nm diameter, which are stabilized by the binding of associated proteins. Microtubules are essential for chromosome segregation, flagellar motility, cellular architecture, cytoplasmic organelle transport and membrane traffic. GTP binding and hydrolysis coupled to polymerization confer microtubules dynamic assembly-disassembly properties, which are the basis for their cellular regulation and organization by means of the associated proteins, microtubule organizing centres and the activation of intracellular signals.
Physiology/Pathology	An essential cellular protein. The unassembled tubulin pool was reported to decrease in aging human cerebral cortex. The microtubule associated protein tau forms part of the paired helical filaments characteristic of Alzheimer's disease. There are also other microtubule-related pathologies not involving tubulin itself.
Degradation	Unknown
Genetics/Abnormalities	15 - 20 disperse genomic sequences, many of them pseudogenes, coding for each tubulin subunit; 6 human β-tubulin sequences were demonstrated to be functional genes. Tubulin genes of vertebrates group in six isotypic classes per subunit, several of them expressed in different tissues. Genetic abnormalities are unknown in humans.
Half-life	Fast: 5 days; slow: 15 days (rat brain)
Concentration	Cytoplasmic concentration strongly dependent of the cell type, from roughly 5% of protein in brain extracts to 0.1% or less in other organs; microtubules are absent from mature mammalian erythrocytes.
Isolation Method	Typically prepared from vertebrate brain by two types of procedures: i) cycles of in vitro microtubule assembly and disassembly, which gives microtubule protein (a mixture of tubulin and microtubule associated proteins), followed by purification by phosphocellulose chromatography (PC-tubulin) and ii) directly, by ammonium sulfate fractionation, batchwise DEAE-Sephadex ion exchange and Mg^{2+}-induced precipitation (modified Weisenberg procedure, W-tubulin). Tubulin can be prepared from human platelets. Tubulin can be also prepared with the aid of taxol, particularly from sources were it is less abundant.
Amino Acid Sequence	More than 200 tubulin gene sequences from diverse organisms are available. Highly conserved parts of the ca. 450 residue sequences include putative nucleotide binding motifs DLEP near position 70 (guanine), (KR)GXXXXG at positions 104-110 (GA)GGTGSG at positions 140-147 (phosphate), and the zones comprised between residues 390 and 430 approximately. The main variable sequence parts which distinguish the isotypes of each chain are the 10 to 20 C-terminal residues, which can be removed by controlled subtilisin cleavage. These highly acidic C-terminal zones of tubulin are exposed on the surface of microtubules and have been implied in interactions with microtubule associated proteins.

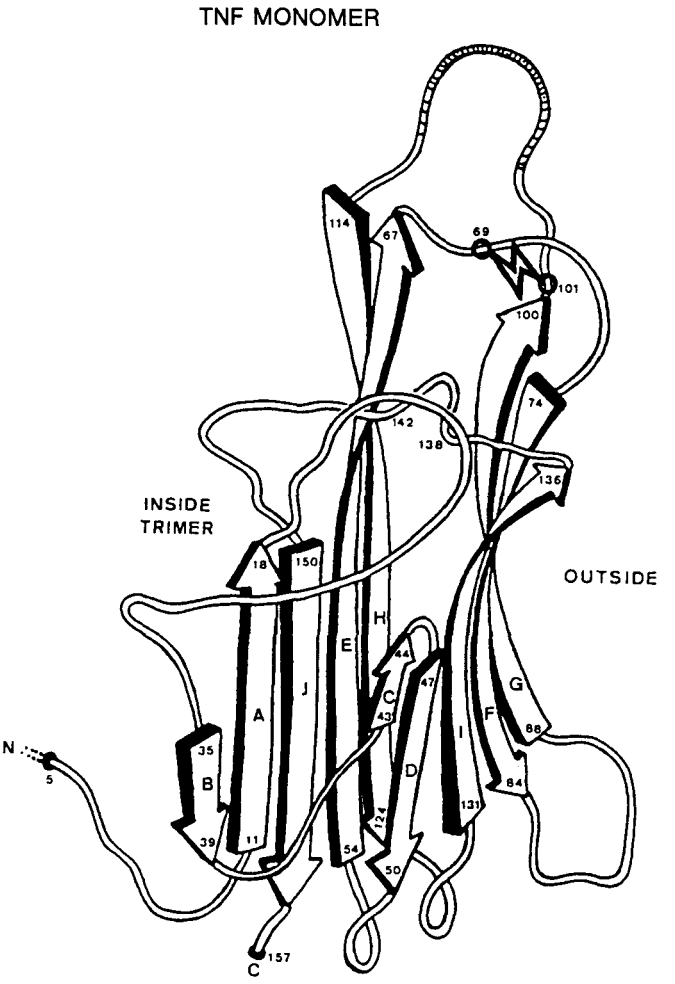
Disulfides/SH-Groups	Apparently one intrachain disulfide bond per monomer (calf brain tubulin); a total of 20 cysteines per tubulin dimer (pig brain tubulin sequence); some of the sulfhydryls are essential for assembly.
General References	Williams, R.C. and Lee, J.C. The preparation of brain tubulin. *Meth. Enzymol.* 1982, **85**:376-385. Hyams, J.S. and Lloyd, C.W. (eds.), *Microtubules.* Wyley Liss, New York 1994. Mandelkow, E. and Mandelkow, E.M. Microtubules and microtubule-associated proteins. *Curr. Opin. Cell Biol.* 1995, **7**:72-81. Monasterio, O., Andreu, J.M., Lagos, R. Tubulin structure and function. *Comments Mol. Cell. Biophys.* 1995, **8**:273-306. Erickson, H.P. Cell 1995, **80**:367-370. Amos, L. Structure of microtubules. In: *Microtubules.* Roberts, K., Hyams, J.S. (eds.), Academic Press, New York 1979, pp 1-64. Beese, L. et al. *J. Mol. Biol.* 1987, **194**:257-264. Kirchner, K. and Mandelkow, E.M. *EMBO J.* 1985, **4**:2397-2402. de la Vina, S. et al. *Biochemistry* 1988, **27**:5352-5365. Arévalo, M.A. et al. *J. Mol. Biol.* 1990, **214**:105-120. Ward, L. et al. *Biochemistry* 1994, **33**:11900-11908. Nogales, E. et al. *Nature* 1995, **375**:424-427.
Ref. for DNA/AA Sequences	Valenzuela, P. et al. *Nature* 1981, **289**:650-655. Ponstingl, H. et al. *Proc. Natl. Acad. Sci. USA* 1981, **78**:2757-2761. Krauhs, E. et al. *Proc. Natl. Acad. Sci. USA* 1981, **78**:4156-4160. Sullivan, K. Structure and utilization of tubulin isotypes. *Ann. Rev. Cell Biol.* 1988, **4**:687-716. Little, M. and Seehaus, T. Comparative analysis of tubulin sequences. *Comp. Biochem. Physiol.* 1988, **90B**:655-670. Oakley, B.R. et al. *Cell* 1990, **61**:1289-1301.

Tumor necrosis factor

Nigel P. C. Walker

Synonyms	Cachectin
Abbreviations	TNF, TNF-α
Classifications	Cytokine
Description	A circulating plasma protein secreted by macrophages. A trimeric molecule of identical non-covalently bound subunits. The pre-protein is expressed with a 76 aa signal sequence, containing a trans-membrane section, which is cleaved to give the mature 157 aa monomer. The protein is neither glycosylated nor phosphorylated.
Structure	X-ray crystallographic analysis of recombinant mature TNF has shown the trimer to be conical in shape with all the disulphide bonds located at the apex. The polypeptide chain of each monomer consists almost entirely of β-pleated sheet which forms a β-sandwich similar in topology to the 'jelly-roll' folding motif found in viral coat proteins.
Molecular Weight	52,000 (17,350 per monomer).
Sedimentation Coeff.	4.1 S
Isoelectric Point	6.7
Extinction Coeff.	11.2 (280 nm, 1%, 1 cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	Unknown
Biological Functions	It is secreted by macrophages in response to certain invasive stimuli, either infectious or neoplastic. A particularly potent stimulus is endotoxin, the lipopolysaccharide of gram-negative bacterial cell walls. Endogenous TNF is a major mediator of many immune and inflammatory reponses, including fever and shock. TNF exhibits a multiplicity of actions since a) TNF receptors are present on virtually all cell types, and b) TNF action occurs through multiple signal transduction.
Physiology/Pathology	During endotoxin induced septic shock, plasma levels are dramatically raised above their normal levels. A similar increase results from graft-versus-host disease, a major complication following bone marrow or organ transplantation.
Degradation	Proteolytic degradation in plasma.
Genetics/Abnormalities	Localization: short arm of chromosome 6 within the major histocompatibility complex (MHC) region.
Half-life	$\simeq$ 15 min (blood circulation)

Concentration	Plasma: normal level ca. 10 ng/L, pathological levels are > 100 ng/L; in acute endotoxin induced shock, plasma levels may reach as much as 600 ng/L.
Isolation Method	Recombinant expression in *E.coli*, ammonium sulphate fractionation, ion-exchange chromatography.
Amino Acid Sequence	Homology with lymphotoxin, also referred to as TNF-β. Sequence identity is only 32%, however very strong structural homology.
Disulfides/SH-Groups	One disulfide per monomer subunit (aa 69–101) and no free sulfhydryls; no inter-subunit disulfides. Reduced protein retains activity.
General References	Beutler, B. and Cerami, A. *Nature* 1986, **320:** 584–588. Eck, M. J. and Sprang, S. R. *J. Biol. Chem.* 1989, **264:** 17595–17605. Jones, E. Y., et al. *Nature* 1989, **338:** 225–228. Old, L. J. *Scientific American* 1988, **258**(5): 41–49.
Ref. for DNA/AA Sequences	Pennica et al. *Nature* 1984, **312:** 724–728. NBRF code: QWHUN.



Tyrosine hydroxylase

Annie Lamouroux and Jacques Mallet

Synonyms	Tyrosine 3-monooxygenase, Tyrosine 3-hydroxylase.
Abbreviations	TH, TyrOHase
Classifications	EC 1-14-16-2
Description	TH catalyzes the initial rate-limiting step in the biosynthesis of catecholamines (dopamine, norepinephrine, epinephrine) in the peripheral and central catecholaminergic neurons and in chromaffin cells of the adrenal medulla. TH is an iron cofactor enzyme that hydroxylates L-tyrosine to L-dopa. The catalytic site of the enzyme is located in the C-terminal moiety. TH activity is short-term regulated by reversible phosphorylation reactions. The N-terminal domain is a regulatory domain that can be phosphorylated by a variety of protein kinases, thus affecting the rate of catecholamine biosynthesis. Human TH is encoded by a single gene that generates, by alternative splicing, four forms of TH monomers differing in their N-terminal extremities.
Structure	Tetramer.
Molecular Weight	Tetramer: 255,000 Da (gel filtration experiment) Monomer: 60,000 Da (SDS-PAGE) TH1: 55,535 (calculated from the deduced aa sequence); TH2: 55,941; TH3: 58,083; TH4: 58,524.
Sedimentation Coeff.	9.085 S
Isoelectric Point	6.35(TH1), 6.46(TH2), 6.46(TH3), 6.57(TH4) calc.
Extinction Coeff.	Unknown
Enzyme Activity	TH is a mixed function oxidase or monooxygenase that catalyzes the hydroxylation of L-tyrosine to form 3,4-dihydroxy-phenylalanine (Dopa) by using molecular oxygen and the cofactor tetrahydrobiopterin (BH4). The hydroxylation of tyrosine involves the transient formation of a charged quaternary complex intermediate. BH4 is oxidized to BH2 during the course of this reaction, and the regeneration of the reduced cofactor is catalyzed by dihydropterine-reductase.
Coenzymes/Cofactors	6 R-L-erythro-tetrahydropterin is the natural cofactor, and serves as an electron donor. 6-methyl and 6-7 dimethyl tetrahydropterin are the preferred synthetic cofactors. Fe^{2+} is required in the catalytic function of the enzyme.
Substrates	The biological substrate is L-tyrosine. L-phenylalanine is also hydroxylated by the purified enzyme. In this case, the rate of phenylalanine hydroxylation is equal to the rate of tyrosine hydroxylation.
Inhibitors	TH is submitted to feed-back inhibition by high concentrations of catechols (dopa, dopamine, norepinephrine and epinephrine). The inhibition is competitive with the pterin cofactor. Iron chelating agents, such as α,α'-dipyridyl and O-phenanthroline, also inhibits the enzyme.

Biological Functions	TH is the rate limiting enzyme in the synthesis of catecholamines. It therefore plays a key role in the physiology of adrenergic neurons. TH levels vary both during development and as a function of the activity of the nervous system. The activity of TH is subjected to short and long term regulation by extra and intracellular signals. This provides mechanisms by which the amount of dopamine synthesized and available for secretion can be modulated by physiological requirments.
Physiology/Pathology	The role of TH is to synthesize catecholamine neurotransmitters. A decrease of TH activity is described in Parkinson's disease, especially in the nigrostriatal system. However, the homospecific activity of residual TH in parkinsonian brain is increased. In pheochromocytoma tumors, the level of TH is increased.
Degradation	Unknown
Genetics/Abnormalities	The TH gene is located on chromosome 11p15, contiguous to the insulin gene. A genetic component of manic-depressive psychosis has been shown to reside in this region suggesting that the TH gene may be implicated. Several restriction fragment length polymorphisms (RFLP) are descibed for the TH gene and are used in the analysis of various neuropsychiatric disorders where catecholamines synthesis is affected.
Half-life	Unknown
Concentration	Unknown
Isolation Method	The best source of human TH is pheochromocytoma tumors or adrenal glands. The purification steps include ammonium sulfate fractionnation, DEAE chromatography, gel permeation chromatography and/or heparin sepharose affinity chromatography.
Amino Acid Sequence	Four types of TH proteins are generated by alternative splicing, diverging in the N-terminal extremity. MPTPDATTP QAKGFRRAV SELDAKQ AEAIMvrg qGAPGP SLTGSPW PGTTAA PAASYT PTPRspr. The N-terminal extremity of TH4 form is represented; in TH1, two exons are missing: aa 31 to 34,and aa 35 to 62. In TH2, aa 35 to 62 are missing; in TH3, the aa 31 to 34 are missing. TH1, TH2, TH3 and TH4 are 497, 501, 524 and 528 aa long respectively. TH1 is 89% homologous to rat TH. The catalytic site, which has been shown to be in the carboxy-terminal domain, is thus conserved in all four proteins. Interestingly, the addition of V-R-G-Q in TH2 endows the adjacent serine (31) with the possibility of being phosphorylated by a calmodulin-dependant protein kinase II (CaM-PK II). TH1 and TH2 have been produced in E. Coli and phosphorylation sites determined. In both isoforms, Ser 40 was found to be phosphorylated by a cAMP-dependent protein kinase (PKA), Ser 19 and Ser 40 are phosphorylated by a CaM-PKII, and Ser 31 (the putative phosphorylation site generated by alternative splicing in TH2) was specifically phosphorylated by a CaM-PKII in THII only and not in TH1. Thus, alternative splicing of human TH provides a supplementary means of regulating catecholamine levels in normal and pathological neurons. TH belongs to a family of hydroxylases that include phenylalanine hydroxylase and tryptophan hydroxylase, the rate limiting enzyme in the biosynthesis of serotonin. The three enzymes are mixed function oxidases that share many characteristic biochemical and immunologic properties. The comparison of the three human sequences reveals that they share a high degree of homology in the central (over 75%identity) and carboxy termini of the proteins. This is taken as further evidence that the catalytic site lies in the carboxy end of the protein and that these three enzymes have evolved from a common ancestor.

Disulfides/SH-Groups	Interchain: no disulfide bonds. Intrachain: unknown.
General References	Mallet, J., et al. *Progress in Catecholamine Research,* Part A. Alan R. Liss, Inc. 1988, pp. 21–27. Masserano, J. M., et al. *Handbook of Experimental Pharmacology,* Vol **90/II**. Springer-Verlag, Berlin, Heidelberg 1989, pp. 428–469. Zigmond, R. E., et al. *Ann. Rev. Neurosci.* 1989, **12:** 415–461. Nagatsu, T. *Cellular and Molecular Neurobiology* 1989, **9:** 313–321.
Ref. for DNA/AA Sequences	Grima, B., et al. *Nature* 1987, **326:** 707–711. Kobayashi, K., et al. *J. Biochem.* 1988, **103:** 907–912. Le Bourdelles B., et al. *J. Biol. Chem.* 1991, In press. The nucleotidic sequences are accessible in the EMBL data bank under the mnemonic HSTHA (TH1), HSTHBA (TH2), HSTHR (TH3) and HSTHX (TH4).

Human Protein Data
Edited by André Haeberli

UDP-Glucuronosyltransferases

Brian Burchell

Synonyms	UDP-Glucuronyltransferase
Abbreviations	UGTs, UDPGTs
Classifications	EC 2.4.1.17
Description	A family of microsomal isoenzymes predominantly located in liver and kidney, but also different isoenzyme profiles are present in many tissues including intestine, olfactory epithelium and brain. The enzymes are integral membrane proteins and the active sites are believed to be inside the lumen of the endoplasmic reticulum. Several transporter proteins may be required to permit substrate entry and product release from the ER lumen.
Structure	None of the proteins have been crystallised.
Molecular Weight	50–57 kDa (SDS-PAGE) dependent on the individual isoenzyme.
Sedimentation Coeff.	Unknown
Isoelectric Point	UDPGT: 6.2 and 7.4 (isoenzyme dependent)
Extinction Coeff.	Unknown
Enzyme Activity	The enzymes are sugar tranferases and conjugases which can also act slowly in reverse as β-glucuronidases.
Coenzymes/Cofactors	None
Substrates	UDP-glucuronic acid as donor substrate and thousands of drugs, pesticides, carcinogens, xenobiotics and endobiotics as acceptor substrates.
Inhibitors	Novobiocin inhibits bilirubin UGT. There are no good general inhibitors.
Biological Functions	UGTs have evolved as part of the chemical defence system to facilitate elimination of potentially toxic endogenous compounds and xenobiotics as water soluble derivates. One major role in liver is the diglucuronidation of bilirubin as the terminal biotransformation step prior to elimination of this toxic heme breakdown product. The UGTs are also involved in glucuronidation of excess 6α-hydroxy bile acids and hormonally active steroids to facilitate inactivation, transport or excretion.
Physiology/Pathology	Essential for liver detoxicatory function. Loss of bilirubin UGT activities results in severe rare and common mild unconjugated hyperbilirubinaemias.
Degradation	Unknown
Genetics/Abnormalities	Multigene family. UGT subfamily 1 appears to be a large single gene which can be differentially spliced to produce at least four different isoenzymes. UGT subfamily 2 consists of several different genes. Genetic deficiencies of at least two bilirubin UGTs are known in Crigler-Najjar syndromes and Gilbert's disease. The protein defects are heterozygous and DNA analysis of these disorders has not been completed.

Half-life	Unknown
Concentration	Unknown
Isolation Method	Isolated from liver microsomes. Solubilize microsomes with non-ionic detergent, separate and purify isoenzymes by using ion-exchange chromatography, chromatofocusing and UDP-hexanolamine Sepharose 4B affinity chromatography. Solubilisation and maintenance of enzyme activity during purification of these membrane bound lipid-dependent enzymes is very difficult. Purification work has to some extent been superseded by cDNA cloning and stable expression of the UGT clones in tissue culture cells.
Amino Acid Sequence	Available in most protein databases.
Disulfides/SH-Groups	The functional role of SH-groups is unknown although enzyme activities can be inhibited by SH-reagents. The formation of disulfide bonds is unknown.
General References	Burchell, B. and Coughtrie, M. W. H. UDP-Glucuronosyltransferases. *Pharmac. Therap.* 1989, **43:** 261–289. Burchell, B., et al. The UDP-Glucuronosyltransferase gene Superfamily DNA. *Cell Biol.* 1991, **10:** 487–494. Tephly, T. R. and Burchell, B. UDP-Glucuronosyltransferases: a family of detoxifying enzymes. *Trends in Pharmacol. Sci.* 1990, **11:** 276–280. Irshaid, Y. M. and Tephly, T. R. Isolation and Purification of two human liver UDP-Glucuronosyltransferases. *Mol. Pharmacol.* 1987, **31:** 27–34.
Ref. for DNA/AA Sequences	In most sequence databases (also see Jackson et al. *Biochem. J.* 1987, **242:** 581–588).

Urokinase plasminogen activator receptor

Francesco Blasi

Synonyms	None
Abbreviations	uPAR; PLAUR (gene)
Classifications	None
Description	A plasma membrane protein present at rather low level of expression on most cells. In particular, present on monocytes, B-lymphocytes, granulocytes. Absent on red blood cells. Increased in activated macrophages and in tumors. It is attached at the plasma membrane via a glycophosphatidylinositol anchor, and therefore released by PI-specific phospholipase C. Contains 2 moles/mole phosphoethanolamine. Highly glycosylated (about 40% of weight is due to carbohydrates), contains glycosamine (31%) and sialic acid.
Structure	Formed by three internally homologous domains, sharing the cysteine spacing and disulfide-bonds. No tertiary structure data. However, each domain has homology to members of the Ly-6 gene family, coding for single domain, GPI-linked cell surface proteins.
Molecular Weight	32,000 (SDS-PAGE and aa sequence determined on cDNA). Domain 1 contains the first 87 aa residues.
Sedimentation Coeff.	Unknown
Isoelectric Point	5.5-6.2 (receptor-ligand complex)
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	Focalizes urokinase activity on the cell surface, accelerates activation of pro-urokinase to urokinase, participates in the overall fibrinolytic balance. Mediates the internalization and degradation of urokinase-inhibitor complexes, through the interaction with members of the LDL-receptor family (LRP/α2-macroglobulin receptor, gp330, VLDL-receptor).
Physiology/Pathology	Binds to N-terminal sequence of urokinase (residues 20-32). K_d= 0.2 nM. Function required for cell migration and for the invasive behaviour of tumor cells. Mediates cell migration in inflammatory responses and in metastasis. Its function in fibrinolysis is not yet known. Receptor is missing in blood cells of patients with paroxysmal nocturnal hemoglobinuria.
Degradation	Unknown

1. Installment 1992 (revised 1997)

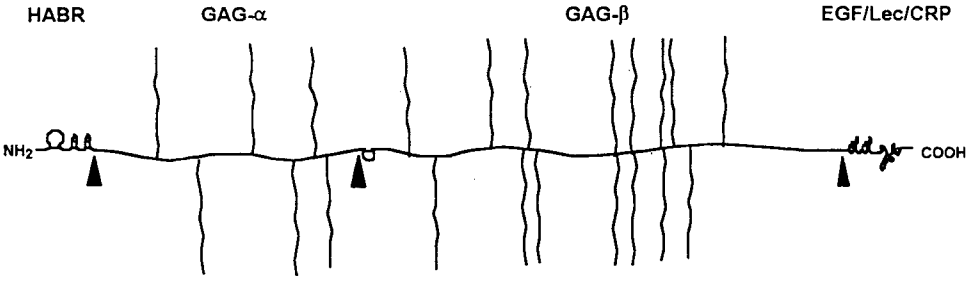
Genetics/Abnormalities	Synthesized from a 1.4 kb mRNA. In mouse an alternatively spliced form has been identified, which appears to be a secreted, ligand-binding variant. Localization: chromosome 19q13.1-q13.2.
Half-life	In cell culture, it has an half-life of 8h in monocytic cell lines.
Concentration	The number of receptors per cell is variable. About 50,000 in monocytes, about 5,000 in B-lymphocytes. Number of receptor in vitro can be regulated by phorbol esters, inflammatory agents, growth factors.
Isolation Method	Isolated from cultured cells by detergent fractionation (Triton X114) and affinity chromatography on pro-urokinase Sepharose or on Sepharose-monoclonal antibody column.
Amino Acid Sequence	A single chain molecule formed by three domains probably arisen by gene duplication. The N-terminal domain contains the ligand-binding activity. The nascent protein is made of 313 residues. The C-terminal 32 aa are processed during synthesis and substituted by the glycosylphosphatidylinositol anchor. The aa sequence bears homology (based on cysteines and gap spacing) to the Ly6 proteins.
Disulfides/SH-Groups	Contains 28 cysteines/mole, probably all in disulfide bonds. The N-terminal domain contains 8 cysteines, the other two domains 10 each.
General References	Conese, M. and Blasi F. et al. *Biological Chemistry Hoppe Seyler* 1995, **376**:143-155. Ploug, M. and Ellis, V. *FEBS Lett.* 1994, **349**:163-168. Blasi, F. et al. *J. Cell Biol.* 1987, **104**:801-804.
Ref. for DNA/AA Sequences	Roldan, A. et al. *EMBO J.* 1990, **9**:467-474.

Versican

Dieter R. Zimmermann and María T. Dours-Zimmermann

Synonyms	PG-M (chick homologue)
Abbreviation	VC
Classification	Large aggregating proteoglycan
Description	Versican is an extracellular matrix proteoglycan consisting of a multi-domain core protein substituted with chondroitin sulfate side-chains, N- and O-linked oligosaccharides. The newly synthesized versican polypeptide is preceded by a 20 aa-long secretory signal sequence, which is removed co-translationally. The currently described three isoforms result from alternative splicing of two exons (7 and 8) coding for the central glycosaminoglycan carrying domains. The largest splice variant, versican V0, includes 3376 aa after signal peptide cleavage. The widely expessed versican V1-variant is 2389 aa long, whereas the V2-isoform comprises 1622 aa. Versican is expressed by primary fibroblasts of various origins, by epidermal cells, by arterial smooth muscle cells as well as by the osteosarcoma cell line MG63 and by U251MG glioma cells. Versican has a wide tissue distribution including skin, blood vessel walls, nervous tissues, placenta and more. The N-terminal sequences of versican are virtually identical to the 68 kD hyaluronectin and the 60–70 kD glial hyaluronate-binding protein GHAP, suggesting that these proteins are derived from the same gene.
Structure	A large aorta proteoglycan, most likely being identical to versican, appears on rotary shadowing electron micrographs as long extended structures with one globular domain at either end of the core protein. The more than 200 nm long molecule carries about 20–25 glycosaminoglycan side chains, each approximately 60 nm in length. The proteoglycan forms large aggregates with hyaluronan and cartilage link protein, *in vitro*.
Molecular Weight	$\geq$1,000,000: intact versican V0 and V1; protein size of versican V2 currently unknown. Chondroitinase ABC digestions give rise to core proteins with relative migration of about 500 kDa (V0) and 400 kDa (V1) in SDS-PAGE. Calculated mass of polypeptide chains are 370,493 (V0), 262,726 (V1) and 179,739 (V2).
Sedimentation Coeff.	Unknown
Isoelectric Point	$\sim$4: calculated from aa sequence.
Extinction Coeff.	Unknown
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	The expression of versican is linked to cell proliferation of dermal fibroblasts and keratinocytes *in vitro*, and probably also *in vivo*. In the dermis, versican appears to be associated with microfibrils of the elastic network.

	There is growing evidence that versican inhibits fibronectin-, laminin- and collagen-mediated cell-matrix interactions and thus may modulate cell adhesion and migration.
Physiology/Pathology	No deficiency state identified at present.
Degradation	Hyaluronectin and glial hyaluronate-binding protein GHAP, that are most likely proteolytic fragments of versican containing the hyaluronan-binding domain, have been isolated from nervous tissues. It is currently not known, whether these putative degradation products of 60–70 kD size are generated by a physiological process.
Genetics/Abnormalities	The human gene on chromosome 5q12–14 extends over more than 90 kb and is subdivided into 15 exons. The gene gives rise to at least 3 alternatively spliced mRNAs, which are in addition heterogenous at their 3′ untranslated ends due to different polyadenylation sites. The expression of the different versican splice-variants seems to be controlled in a tissue-specific manner. No genetic abnormalities associated with the versican gene have been identified so far.
Half-life	Unknown
Concentration	Unknown, rather low abundant matrix protein.
Isolation Method	Versican can be isolated from 4M Gu-HCl extracts of various tissues (e.g. placenta, aorta) or directly from culture medium of fibroblast or glioma cell lines. The isolation protocol usually includes CsCl-density-gradient centrifugation followed by anion-exchange chromatography and gel filtration.
Amino Acid Sequence	The primary structure of versican has been deduced from cDNA data. The mature core proteins of the three isoforms contain at the N-terminal end a hyaluronan-binding region (HABR) followed by one or two glycosaminoglycan attachment domains (versican V0: GAG-α and GAG-β; versican V1: only GAG-β; versican V2: only GAG-α). The C-terminal portion of the versican splice-variants consists of a set of two EGF-like repeats (EGF), a lectin (Lec) domain and a complement regulatory protein-like element (CRP). The central GAG-α and GAG-β domains include 5 to 8 and 12 to 15 sequences, respectively, that are likely to carry chondroitin sulfate side chains.
Disulfides/SH-Groups	The different structural elements include the following number of intramodular disulfides: HABR five; GAG-α none; GAG-β one; EGF three; Lec three; CRP two.
General References	Zimmermann, D. R. and Ruoslahti, E. *EMBO J.* 1989, **8**: 2975–2981. Mörgelin, M. et al. *J. Biol. Chem.* 1989, **264**: 12080–12090. Dours-Zimmermann, M. T. and Zimmermann, D. R. *J. Biol. Chem.* 1994, **269**: 32992–32998. Naso, F. M. et al. *J. Biol. Chem.* 1994, **269**: 32999–33008. Yamagata, M. and Kimata, K. *J. Cell Sci.* 1994, **107**: 2581–2590.
Ref. for DNA/AA Sequence	See general references. EMBL/Genbank accession numbers: X15998 (versican V1), U16306 (versican V0).



Schematic model of the V0 splice variant of versican. The wavy lines depict glycosaminoglycan side chains. Splice junctions subjected to alternative splicing are indicated by arrowheads. The shorter versican V1 and versican V2 isoforms lack the GAG-α or the GAG-β domain, respectively. HABR: hyaluronan-binding region; GAG: glycosaminoglycan attachment domain; EGF: epidermal growth factor-like repeats; Lec: lectin-like domain; CRP: complement regulatory protein-like element.

Vitamin D binding protein

Nancy E. Cooke and John G. Haddad

Synonyms	GC-globulin; Group-specific component of serum
Abbreviations	DBP; GC
Classifications	Inter α-globulin
Description	DBP is a single-chain plasma protein synthesized predominantly in liver, but an identical protein is likely to be synthesized from transcripts detected in kidney, placenta, and testis. Less than 5% of DBP is occupied by vitamin D sterols. Human DBP contains a 16 aa signal peptide and 458 aa in the mature protein. There are 3 common alleles, Gc1F, Gc1S, and Gc2, that can be distinguished electrophoretically. Gc1F and Gc1S each have two bands: Gc1a with pI 4.85 containing a single N-acetylneuraminic acid residue and a nonglycosylated Gc1c with pI 4.95. Both Gc1F proteins migrate faster than the Gc1S proteins; the molecular basis for this difference remains unknown. Gc2 migrates as a single band and lacks a Thr residue at 420 postulated to be the site of O-linked glycosylation in the Gc1 forms. DBP forms a linked multigene family with albumin, α-feto-protein and α-albumin, sharing significant structural similarities.
Structure	The tertiary structure of DBP has not been determined, although crystals have been obtained. DBP contains 23% aa identity with albumin, including 28 identically positioned Cys residues. There is an intramolecular triplication. The crystal structure of the related albumin has been resolved to 6.0 Å, revealing predominantly α-helical protein.
Molecular Weight	58,000 by gel electrophoresis; 51,335 calculated from aa structure (non-glycosylated).
Sedimentation Coeff.	3.5 S
Isoelectric Point	4.85 - 5.10
Extinction Coeff.	9.1 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	DBP is the main plasma carrier of 25-OHD_3 and provides a circulating reservoir of that sterol. DBP binds monomeric actin with high affinity and along with gelsolin constitutes the plasma actin scavenger system. DBP binds to complement component C5a and C5a-des-Arg acting as a co-chemotaxin for neutrophils. DBP binds and presumably transports fatty acids. DBP has been detected on the surface of lymphocytes, monocytes, and placental cytotrophoblasts, although its function on the cell surface remains unknown.
Physiology/Pathology	DBP is not known to be etiological in any disease states. However, DBP levels are increased during pregnancy and estrogen therapy, decreased in

	nephrotic syndrome, malnutrition, and liver disease. Despite extensive worldwide screening a human DBP^0 homozygote has not been identified. An experimental mouse DBP null homozygote has been found to be viable however.
Degradation	Very small degradation fragments have been detected in urine. No large fragments are seen in plasma, suggesting its complete proteolysis intracellularly.
Genetics/Abnormalities	The DBP gene is located on chromosome 4q11-q13. DBP, albumin, and α-fetoprotein are clustered in the 4q11-q22 region. The rat DBP gene has been characterized and contains 13 exons spanning 35 kb. DBP mRNA is 1.7 kb in length. At least 125 electrophoretic variants and several RFLPs have been identified.
Half-life	2.5 days approx. in plasma
Concentration	0.4 g/L approx. in plasma, constituting about 6% of plasma α-globulins.
Isolation Method	Best sources for isolation are plasma or plasma α-globulin fractions. Various isolation methods have been used including: ion-exchange chromatography; gel filtration; affinity chromatography on G-actin, organic anion or anti-DBP columns; preparative gel electrophoresis.
Amino Acid Sequence	The vitamin D sterol binding domain is located between aa 35-49, while the actin binding domain lies between 350-403. Nucleotide (NT) and aa identity respectively, within the multigene family are as follows: DBP/ALB 39%, 23%; DBP/AFP 37%, 20%; DBP/α-ALB 41%, 20%; ALB/AFP 52%, 40%. This has suggested that divergence of the DBP gene predated the divergence of albumin and α-fetoprotein.
Disulfides/SH-Groups	14 putative intrachain disulfide bonds defining an intramolecular triplication. This is based upon homology to albumin's disulfide structure. DBP is the only member of the albumin, α-fetoprotein, DBP family to retain a potential disulfide bond anchoring the first loop of the first internal domain.
General References	Cooke, N.E. and Haddad, J.G. Vitamin D Binding Protein. In: *Endocrine Reviews Monographs: 4. Hormonal Regulation of Bone and Mineral Metabolism.* Bikle, D.D. and Negro-Vilar, A. (eds.), Endocrine Society, Bethesda, MD, 1995, pp. 111-124. Cooke, N.E. and Haddad, J.G. *Endocr. Rev.* 1989, **10**:294-307. Haddad, J.G. Clinical aspects of measurements of plasma vitamin D sterols and the vitamin D binding protein. In: *Disorders of Bone and Mineral Metabolism.* Coe, R.L. and Favus, M.J. (eds.), Raven Press, New York, 1992, pp. 195-216. Lee, W.M. and Galbraith, R.M. N. Engl. J. Med. 1992, **326**:1335-1341. Haddad, J.G. et al. *Biochemistry* 1992, **31**:7174-7181. Cooke, N.E. J. *Steroid Biochem. Molec. Biol.* 1991, **40**:4-6.
Ref. for DNA/AA Sequences	Cooke, N.E. and David, E. V.*J. Clin. Invest.* 1985, **76**:2420-2423. Sequence submitted to Genbank. Yang, F. et al.*Proc. Natl. Acad. Sci. USA* 1985, **82**:7994-7998. Schoentgen et al. *Biochim. Biophys. Acta* 1986, **871**:189-191. Lichenstein, H.S. et al. *J. Biol. Chem.* 1994, **269**:18149-18154.

Vitamin K-dependent carboxylase

CeesVermeer

Synonyms	Gammaglutamylcarboxylase;Vitamin K-epoxidase
Abbreviations	Carboxylase
Classifications	None
Description	A hydrophobic, integral membrane protein present at the luminal side of the Endoplasmic Reticulum of many different types of cells and organs. Identity of the various forms of carboxylase seems likely but has not been demonstrated.
Structure	Unknown
Molecular Weight	Carboxylase from bovine liver has a molecular weight of 94,000 (SDS-PAGE, reducing conditions). No other mammalian carboxylases have been purified as yet.
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	The enzyme catalyses the posttranslational conversion of peptide- and protein-bound glutamate residues into gammacarboxyglutamate (Gla). Gla-residues are calcium-binding groups in proteins, which are essential for their biological activity.
Coenzymes/Cofactors	Vitamin K-hydroquinone, O_2, CO_2. The oxydation of vitamin K-hydroquinone into vitamin K-epoxide provides the energy required for the carboxylation reaction. Carbon dioxide rather than bicarbonate is used to form the extra carboxyl group present in Gla.
Substrates	In vivo substrates are nascent polypeptide chains entering the endoplasmic lumen, and containing the 'pro'-sequence, an 18 aa residues sequence directly preceding the structure of the mature protein. In vitro substrates are peptides and proteins with a high similarity to one of the Gla-containing proteins, but in which Gla has been replaced by Glu. These substrates may be obtained by a) peptide synthesis, b) thermal decarboxylation of a suitable Gla-protein or c) cloning and expression of recombinant peptides derived from various Gla-proteins in prokariotic cell system. Substrates containing the 'pro'-sequence are superior to any other ones. The pro-sequence serves as a recognition site for carboxylase. Endogenous substrates are precursor proteins which accumulate in the ER during vitamin K deficiency.
Inhibitors	Direct inhibitors are cyanide, 2,3,5,6-tetrachloro-4-pyridinol and the structural vitamin K-analogue 2-chloro-3-phytyl-1,4-naphtoquinone (chloro-K). Coumarin derivatives (phenprocoumon, acenocoumarol, warfarin, brodifacoum) and aspirin act as indirect inhibitors of the carboxylation reaction because they block the enzyme KO-reductase, thus preventing the recycling of vitamin K. In humans coumarin derivatives are used as oral anticoagulants.

1. Installment 1992 (revised 1997)

Biological Functions	Participates in the posttranslational modification of proteins. Products are Gla-containing proteins, which play a role in blood coagulation (prothrombin, factors VII, IX and X), blood coagulation inhibition (proteins C and S) and in bone metabolism (osteocalcin and matrix Gla-protein).
Physiology/Pathology	Essential for blood coagulation; genetic defects may lead to combined functional deficiency of the various Gla-proteins. Essential for bone development of the fetus; administration of coumarin derivatives to pregnant women may result in severe deformatons of the fetal bones, known as 'fetal warfarin syndrome' or 'chondrodysplasia punctata'.
Degradation	Unkown
Genetics/Abnormalities	Various combined coagulation factor deficiencies have been reported, which are probably due to a genetic mutation of carboxylase.
Half-life	Unknown
Concentration	Less than 0.01% of the total liver protein content, in other tissues probably even less.
Isolation Method	Isolated from salt-washed, detergent-solubilized liver microsomes by ammonium sulphate precipitation and affinity chromatography using the 59-amino acid peptide FIXQ/S (residues -18 to 41 of factor IX, with mutations at positions -4 and -1). A critical step in the purification procedure is the size reduction of the protein/detergent micelles by sonication.
Amino Acid Sequence	The enzyme has 758 aa residues as derived from the c-DNA for human γ-glutamyl carboxylase.
Disulfides/SH-Groups	Disulfides: unknown; Sulfydryl reagents inhibit carboxylase activity, demonstraing the presence of one or more essential SH-groups.
General References	Wu, S.M. et al. *Proc. Natl Acad. Sci. USA* 1991, **88**:2236-2240. Vermeer, C. *Biochem. J.* 1990, **266**:625-636. Olson, R.E. *Annu. Rev. Nutr.* 1984, **4**:281-337. Suttie, J.W. *Hepatology* 1987, **7**:367-376. Morris, D.P. et al. *J. Biol. Chem.* 1993, **268**:8735-8742.
Ref. for DNA/AA Sequences	Wu, S.M. et al. *Science* 1991, **254**:1634-1636.

Vitamin K epoxide reductase

Cees Vermeer

Synonyms	KO-reductase, Pathway I reductase.
Abbreviations	None
Classifications	None
Description	A hydrophobic membrane protein present at the luminal side of the Endoplasmic Reticulum of many different types of cells and organs. Identity of the various forms of KO-reductase seems likely, but has not been proven.
Structure	Unknown
Molecular Weight	Unknown
Sedimentation Coeff.	Unknown
Isoelectric Point	Unknown
Extinction Coeff.	Unknown
Enzyme Activity	Vitamin K epoxide (KO) is formed during the posttranslational carboxylation of glutamate into gammacarboxyglutamate (GLA) residues. KO-reductase is the principal enzyme involved in the reduction of KO to vitamin K hydroquinone, which is the active coenzyme in the carboxylation reaction. By the action of KO-reductase vitamin K is recycled several thousand fold before being degraded into inactive metabolites.
Coenzymes/Cofactors	Dithiols serve as hydrogen-donating cofactors for KO-reductase. In vitro both synthetic dithiols like dithiothreitol as well as the protein dithiol thioredoxin are active. The physiological cofactor has not yet been identified.
Substrates	KO, which is converted into vitamin K quinone;Probably also vitamin K quinone, which is converted into K hydroquinone.
Inhibitors	The most important inhibitors are coumarin-derivatives, which form tight complexes with KO-reductase. The molecular mechanism of their inhibitory activity is not clear, as a result of their action the plasma level of the GLA-containing blood coagulation factors will decrease, resulting in prolonged coagulation times. First generation coumarin-derivatives are characterized by biological half-life times of 12–48 h. Examples are phenprocoumon, acenocoumarol and warfarin. All three are used for oral anticoagulant therapy, warfarin is also widely used as a rodenticide. Second generation coumarin-derivatives have biological half-life times of several months to more than half a year. Examples are: brodifacoum, difenacoum and flocoumafen. They are exclusively used as rodenticides. Besides coumarin drugs, also aspirin has a weak inhibitory activity.
Biological Functions	The biological function of KO-reductase is to convert vitamin K-epoxide into more reduced forms of the vitamin. Because the dietary intake of vitamin K is about 1000-fold lower than the amount required for the synthesis of the various GLA-containing proteins, the recycling of vitamin K is crucial for an adequate production of these proteins.

Physiology/Pathology	GLA-containing proteins play a crucial role in blood coagulation and in bone metabolism. Genetic defects, inhibition or absence of KO-reductase will lead to the production of GLA-deficient or 'descarboxy'-proteins, which are generally inactive. This may lead to a combined functional deficiency of the GLA-proteins resulting in severe bleeding disorders, or (in fetuses) in bone deformations known as 'chondrodysplasia punctata'. High dosages of vitamin K may be used to bypass the effect. In that case vitamin K quinone is reduced by a less specific enzyme, DT-diaphorase.
Degradation	Unknown
Genetics/Abnormalities	Prolonged use of coumarin derivatives results in resistance in rats within 2–3 years. The resistance is probably caused by point mutations altering the affinity of the enzyme for the drug. In humans no genetic abnormalities have been identified.
Half-life	Unknown
Concentration	Less than 0.01% of the total liver protein content, in other tissues probably even less.
Isolation Method	Present in salt-washed detergent-solubilized liver microsomes. Further purification steps may include anion exchange or affinity chromatography, but have not lead to a homogenous enzyme preparation.
Amino Acid Sequence	Unknown
Disulfides/SH-Groups	One or more disulfides are present and their reduction/oxidation is essential for the enzyme activity.
General References	Zaidi, I. M., et al. In: *Current advances in vitamin K research*. Elsevier, New York 1988, pp 419–428. Gardill, S. L., and Suttie, J. W. *Biochem. Pharmacol.* 1990, **40:** 1055–1061. Fasco, M. J., and Principe, L. M. *J. Biol. Chem.* 1982, **257:** 4894–4901. Thijssen, H. H. W., et al. *Thromb. Haemostas.* 1988, **60:** 35–38.
Ref. for DNA/AA Sequences	

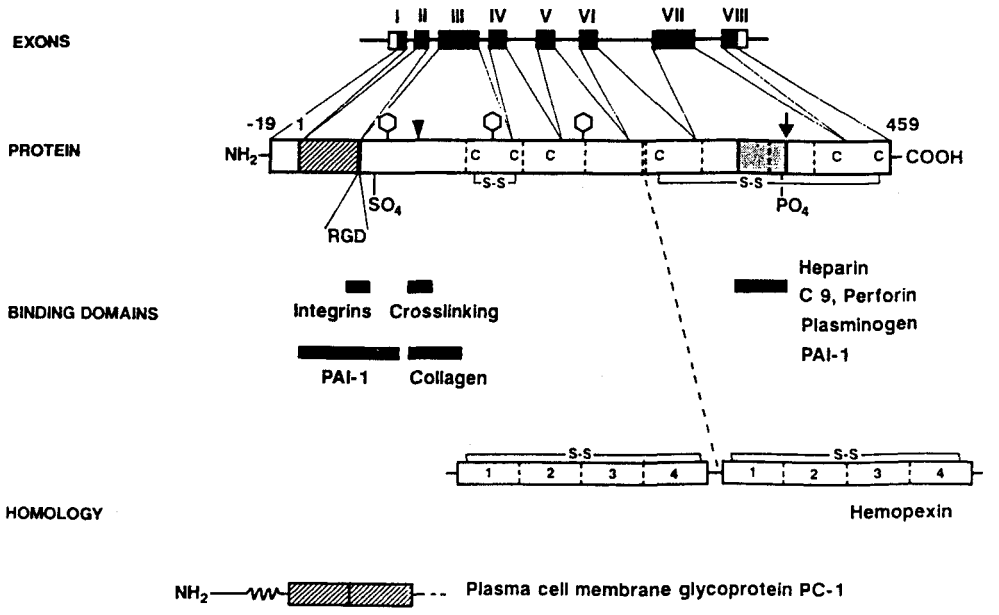
Vitronectin

Klaus T. Preissner

Synonyms	Complement S-protein; Serum-spreading factor
Abbreviations	VN; Vn
Classifications	Electrphoretic mobility: α-fraction
Description	Multifunctional plasma protein mainly synthezised in the liver, present in platelets, monocytes and certain cells in culture (probably identical to plasma vitronectin); possible matrix-associated component of the subendothelium. Circulates in monomeric, dimeric and possibly higher oligomeric forms, associated with plasminogen activator inhibitor-1 in a binary complex, bound to thrombin-antithrombin III in a ternary complex and is a component of the macromolecular SC5b-9 complex of complement.
Structure	Globular molecule (Stokes radius 3.7 nm, frictional ratio 1.37) which undergoes conformational changes into an extended form with the concomitant exposure of binding domains for heparin and other ligands upon reduction/carboxymethylation or denaturation by urea. Cell attachment domain and possible collagen binding sites are located in the amino-terminal portion and a unique glycosaminoglycan-binding domain (348-379) with 40% basic aa residues, followed by a proteolysis-sensitive peptide bond R-A (379-380) are at the C-terminal region. The monomeric molecule contains 10-15% (w/w) carbohydrate at three potential N-glycosylation sites, has at least one free sulfhydryl group which is involved in dimerization.
Molecular Weight	78,000 - 83,000 (SDS-PAGE under denaturing conditions or sedimentation equilibrium ultracentrifugation, respectively); 52,203 (aa content); 160,000 in dimeric form. Due to natural limited proteolysis to a varying degree, two-chain form with 65,000 and 10,000 (detectable after reduction and denaturation).
Sedimentation Coeff.	4.6 S
Isoelectric Point	4.75 - 5.25 (8M urea)
Extinction Coeff.	14.0 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	Cofactor for inhibition of serine protease(s) (thrombin) by plasminogen activator inhibitor-1.
Substrates	None
Inhibitors	None
Biological Functions	Adhesive protein, major ligand for the vitronectin receptor (αV/β3) or αIIb/β3 integrin on adhesive cells, binds to bacteria and supports their adhesion to cells. Additional ligand for the urokinase receptor and the C1q-binding protein, gC1qR. Through provision of different structural domains, vitronectin serves as functional, molecular link between cell adhesion and regulation of defense mechanisms: Heparin-binding and -

	neutralizing protein; complement lysis inhibitor and "scavenger" protein for macromolecular products of the complement and haemostasis cascade systems. Extracellular binding protein for plasminogen activator inhibitor-1 and plasminogen, thereby stabilizing this inhibitor and influencing tissue plasminogen activator-mediated plasmin formation. Binding protein for β-endorphin. "Serpin" cofactor and association with thrombin-"serpin" complexes. Crosslinking substrate for transglutaminase (factor XIIIa) and affinity for collagens; substrate for cAMP-dependent protein kinase and tyrosyl protein sulfotransferase.
Physiology/Pathology	Decreased plasma level (acquired deficiency in liver diseases; isolated hereditary deficiency not reported so far). Accumulation of vitronectin in diseased vessels and association with arteriosclerotic lesions.
Degradation	Eliminated from the circulation possibly by the liver and the vessel wall. Trypsin-like proteinases which initially generate the two-chain form of vitronectin may inactivate some of its functions upon further degradation. Cleavage by acid and thrombin generates two major degradation fragments.
Genetics/Abnormalities	Gene (3 kb) is localized on chromosome 17 (centromeric region of 17q11) and organized in 8 exons and 7 introns. Vitronectin is synthezised as a single-chain polypeptide in polymorphic forms. Gene inactivation in mice has no obvious phenotype.
Half-life	Unknown
Concentration	0.35 g/L (range 0.25 - 0.5 g/L) in plasma.
Isolation Method	Native protein: Isolation from plasma by poly(ethylene)glycol precipitation, ion-exchange-chromatography, gel filtration. Denatured protein: Isolation from plasma/serum by adsorption on glass beads, concanavalin-agarose or by immunoadsorption, heparin-Sepharose; alternatively by heparin-Sepharose under denaturing conditions.
Amino Acid Sequence	Amino-terminus (1 - 45) identical to peptide "somatomedin B"; attachment domain RGD (45 - 47); proteolysis-sensitive peptide bond R-A (379 - 380). Middle portion exhibits weak homology to hemopexin, interstitial collagenase, stromelysin. At S (378) is a potential phosphorylation site, Y (56,59) are potential sulfation sites.
Disulfides/SH-Groups	Monomer: seven intrachain disulfides, one free sulfhydryl. Dimer: one additional inter-chain disulfide, no free sulfhydryl.
General References	Ruoslahti, E. et al. Purification and characterization of vitronectin. *Methods Enzymol.* 1987, **144**:430-437. Tomasini, B.R. and Mosher, D.F. Vitronectin. *Prog. Hemost. Thromb.* 1990, **10**:269-305. Preissner, K.T. Structure and biological role of vitronectin. *Annu. Rev. Cell Biol.* 1991, **10**:275-310. Felding-Habermann, B. and Cheresh, D.A. Vitronectin and its receptors. *Curr. Opin. Cell Biol.* 1993, **5**:864-868. Stockman A. et al. Multimeric vitronectin, *J. Biol. Chem.* 1993, **268**: 22874-22882.
Ref. for DNA/AA Sequences	Jenne, D. and Stanley, K.K.: Molecular cloning of S-protein, a link between complement, coagulation and cell-substrate adhesion. *EMBO J.* 1985, **4**:3153-3157.

Jenne, D. and Stanley, K.K. Nucleotide sequence and organization of the human S-protein gene: repeating peptide motifs in the "pexin" family and a model for their evolution. *Biochemistry* 1987, **26**:6735-6742.
Suzuki, S. et al. Complete amino acid sequence of human vitronectin deduced from cDNA. Similarity of cell attachment sites in vitronectin and fibronectin. *EMBO J.* 1985, **4**:2519-2524.
Sato, R. et al. Monoclonal antibody EMR 1a/212D recognizing site of deposition of extracellular lipid in atherosclerosis. Isolation and charcterization of cDNA clone for the antigen. *J. Biol. Chem.* 1990, **265**: 21232-21236.
Preissner, K.T. et al. (eds.), Biology of Vitronectins and Their Receptors. *Excerpta Medica*, International Congress Series 1042, Amsterdam 1993.



Von Willebrand factor

Miha Furlan

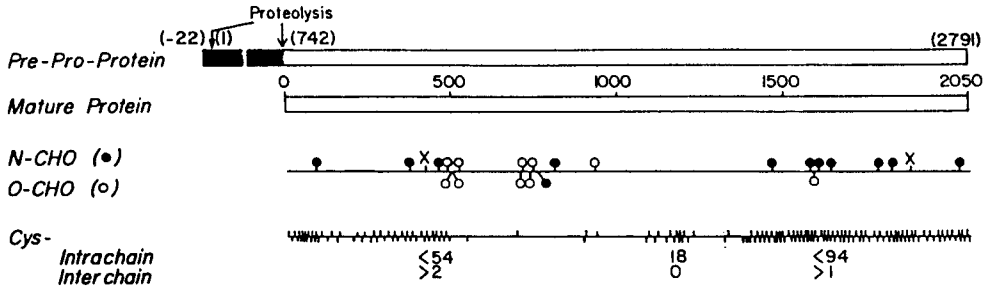
Synonyms	Factor VIII-related protein
Abbreviations	vWF
Classifications	Electr. mob.: between α_2 and β_1 globulins
Description	vWF is a glycoprotein synthesized in endothelial cells and megakaryocytes; circulates in blood associated in a noncovalent complex with factor VIII procoagulant protein. Native vWF consists of a series of multimers arising from polymerization of subunits into long filaments.
Structure	The mature vWF subunit is composed of 2050 aa residues, with a calculated molecular mass of 225,843. Approximately 18.7% of the vWF molecule consists of carbohydrate. Thus, the estimated mw of a subunit is about 270 kDa. The smallest circulating vWF is a dimer held together by disulfide bonds between the C-termini of two subunits. Interdimeric disulfide bridges linking the N-terminal domains on each side of the dimer, lead to linear multimers composed of up to 50 dimers. vWF multimers seen in the electron microscope after rotary shadowing appear as unbranched thin strands. The repeating disulfide-linked dimer has a length of 120 nm. The longest filaments are 3 μm long, a size approximately equal to the platelet diameter, and are the largest known plasma proteins. Only a small fraction of vWF molecules appear to be extended, most of them are thought to be in a "ball of yearn" form coiled upon themselves. Circular dichroism studies demonstrated regions rich in α-helix and β-pleated sheet structure, as well as a high percentage of random coil conformation.
Molecular Weight	The molecular weights of the circulating vWF are in the range between 540 kDa (dimer) and 20,000 kDa (largest polymers), as determined by SDS-PAGE in agarose gels. In a normal human plasma, the majority of vWF molecules have molecular weights between 2,000 and 5,000 kDa.
Sedimentation Coeff.	Up to 20 S
Isoelectric Point	5.7 - 5.9 (7 M urea)
Extinction Coeff.	12.3 (280nm, 1%, 1cm)
Enzyme Activity	None
Coenzymes/Cofactors	Binding to platelets in the presence of ristocetin or botrocetin.
Substrates	None
Inhibitors	None
Biological Functions	vWF has two functions: 1. vWF serves as a carrier for factor VIII in plasma protecting it from proteolytic degradation and inactivation. The binding site for factor VIII is situated in the N-terminal part, between aa 1 and 272 of the vWF subunit. Although each subunit of vWF is capable of binding one factor VIII molecule (mw 280 kDa), the relative ratio of both components in the complex is about 100 to 1 (w/w).

1. Installment 1992 (revised 1997)

	2. vWF is required for normal interactions of platelets with the subendothelium of the injured vessel wall. Several structural domains, binding to platelet glycoproteins GpIb (aa 474 - 488 and 694 - 708) and GpIIb/IIIa (RGDS-sequence 1744 - 1747), and to collagen (542 - 622 and 948 - 998) are responsible for this function. Amino acid residues 1 - 298 and 512 - 673 are involved in vWF binding to heparin. Due to multiple interactions, the binding affinity for platelets and collagen is strongly increased in large polymeric forms of vWF, which represent the hemostatically active fraction of the circulating vWF.
Physiology/Pathology	vWF is essential for primary hemostasis. Decreased concentration or size of vWF result in severe bleeding disorders.
Degradation	Endothelial cells secrete high molecular weight vWF which are slowly degraded in normal plasma by a specific plasma protease (cleavage site 842-Tyr-843-Met).
Genetics/Abnormalities	The primary translation product (short arm of chromosome 12) is the pre pro polypeptide of 2813 aa. The signal peptide of 22 aa and a propeptide (vWFAgII, 741 aa) are cleaved from the precursor vWF, resulting in the mature vWF subunit of 2050 aa. Several abnormalities of the gene coding for vWF have been described, leading to a quantitative deficiency or functional abnormality of vWF. von Willebrand disease is the commonest inherited bleeding disorder (prevalence about 1%). Classification of von Willebrand disease: Type 1: partial quantitative deficiency of vWF Type 2: qualitative deficiency of vWF Type 3: virtually complete deficiency of vWF
Half-life	8 - 12 hrs (blood circulation)
Concentration	About 10 mg/L plasma and about 5 $\mu g/10^9$ platelets (α-granules).
Isolation Method	Cryoprecipitation of plasma, gel filtration on large-pore Sepharose (CL-2B), immunoaffinity chromatography.
Amino Acid Sequence	The sequence has five sets of internally homologous regions. One of these repeated regions (A domains) appears to be homologous to domains found in complement factors B and C2, cartilage matrix protein, collagen type VI and in a number of leucocyte adhesion molecules. Another repeated region (domain D) showed homology to vitellogenins. The single RGDS-sequence (1744 - 1747), common in several other cell adhesive proteins, is involved in vWF binding to platelets.
Disulfides/SH-Groups	Human vWF is a cystine-rich protein in which 169 half-cystine residues are engaged in intrachain or interchain disulfide bridges. Half-cystines are clustered in N-terminal and C-terminal regions, where one third of these residues are in Cys-X-Cys or in Cys-Cys sequences. In contrast, there is only a single Cys residue between aa 516 and 905.
General References	Titani, K. et al. Primary structure of human von Willebrand factor. In: *Coagulation and Bleeding Disorders. The Role of Factor VIII and von Willebrand Factor.* Zimmerman, T.S. and Ruggeri, Z.M. (eds.), Marcel Dekker, New York 1989, pp. 99-116. Fujimura, Y., et al. Structure and Function of Human von Willebrand Factor. In: *Coagulation and Bleeding Disorders. The Role of Factor VIII and von Willebrand Factor.* Zimmerman, T.S. and Ruggeri, Z.M. (eds.), Marcel Dekker, New York 1989, pp. 77-97.

Meyer, D., Girma, J.P. Von Willebrand factor: structure and function. *Thromb. Haemost.* 1993, **70**:99-104.
Sadler, J.E. et al. Molecular mechanism and classification of von Willebrand disease. *Thromb. Haemost.* 1995, **74**:161-166.
Montgomery, R.R., Coller, B.S. Von Willebrand disease. In: *Hemostasis and Thrombosis: Basic Principles and Clinical Practice.* 3rd Edition. Colman R.W. et al. (eds.), Lippincott, Philadelphia 1994, pp. 134-168.

Ref. for DNA/AA Sequences Shelton-Inloes, B.B., et al. *Biochemistry* 1986, **25**:3164-3171.



Structural features of human vWF according to Titani et al. *Biochemistry* 1986, **25**: 3171–3184.

Xanthine Oxidoreductase

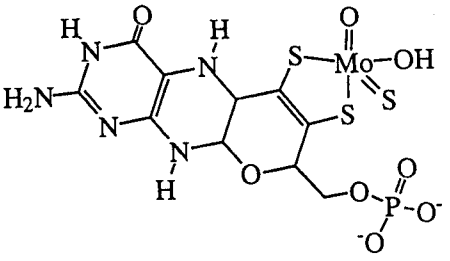
Russ Hille

Synonyms	Xanthine dehydrogenase; Xanthine oxidase
Abbreviations	FAD; NAD^+; DEAE; CM
Classifications	EC 1.2.1.37 (alternatively, 1.2.3.2); a molybdenum-containing hydroxylase
Description	Xanthine oxidoreductase is a complex metalloflavoprotein possessing a molybdenum center, two 2Fe/2S iron-sulfur centers of the spinach ferredoxin variety and FAD, with each redox-active center residing in a discrete domain of the protein structure. The molybdenum center consists of the metal coordinated to an unusual pterin cofactor via a dithiolene side chain of the latter, and is the site of substrate hydroxylation. The enzyme is predominantly localized in the liver, where it catalyzes the final two steps in purine degradation: the oxidation of hypoxanthine to xanthine, and xanthine to uric acid. The reducing equivalents obtained by the enzyme in the course of these substrate hydroxylations are passed on to NAD^+, a reaction that takes place at the flavin center of the enzyme. The enzyme is also found in relatively high amounts in milk, where it is present principally as an oxidase that utilizes O_2 rather than NAD^+ as oxidizing substrate.
Structure	The structure of human xanthine oxidoreductase is at present unknown. However, the x-ray crystal structure of a closely related aldehyde oxidoreductase from *Desulfovibrio gigas* has recently been reported, and on the basis of the extensive sequence homologies between the two proteins, it is expected that the human protein will exhibit the same general structure as that from *D. gigas*. This protein possesses a pair of iron-sulfur centers and a molybdenum center, but lacks a flavin. The two iron-sulfur centers are found in separate well-defined protein folding domains at the N-terminus of the protein. The first of these possesses a protein fold very similar to that of spinach ferredoxin, but the second possesses a unique fold, consisting of two long α helices in parallel, with two shorter helices set at an oblique angle to these; the 2Fe/2S cluster lies at one end of the pair of long helices. This second iron-sulfur domain is connected to the molybdenum-binding portion of the protein by an extended meander over the surface of the protein, and on the basis of the sequence homology between the aldehyde and xanthine oxidoreductases, the flavin domain of the human protein is expected to be inserted at some point along this meander. The molybdenum-binding portion of the protein consists of two large domains that lay across one another, with the metal center at their interface; there are extensive contacts between the polypeptide and the pterin cofactor. In addition, the 2-amino group of the pterin, located distal to the dithiolene moiety that binds the metal, is hydrogen-bonded to one of the cysteine residues that coordinates the second of the two iron-sulfur centers. This interaction clearly plays a role in facilitating electron transfer out of the molybdenum center to the other redox-active sites of the enzyme. Access to the active site is provided by a 30Å-long channel from the surface of the protein, with access to the metal center opposite the coordination position of the pterin cofactor.
Molecular Weight	296,444 (M_r, homodimer)
Sedimentation Coeff.	Unknown

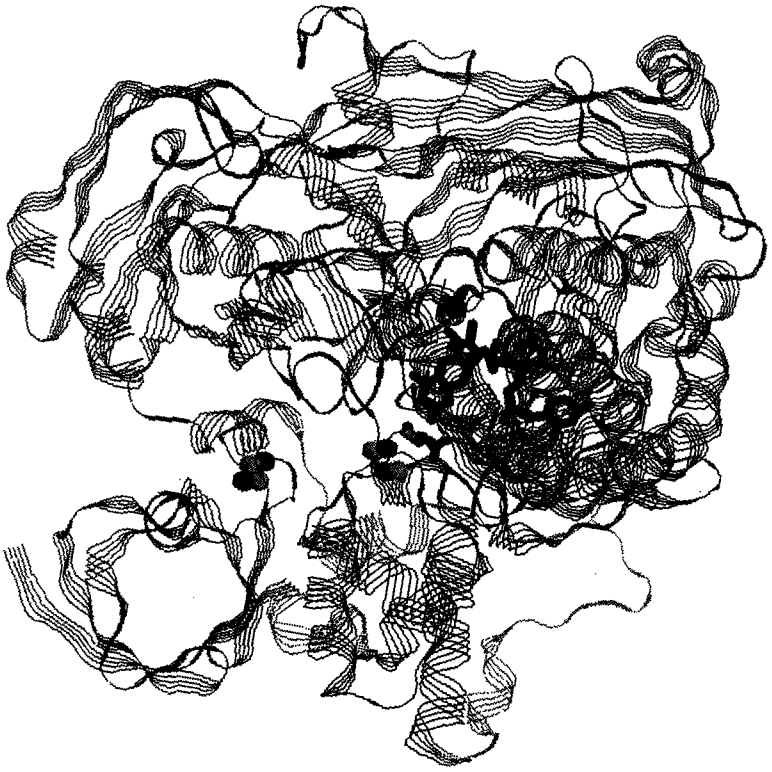
Isoelectric Point	Unknown
Extinction Coeff.	37,800 $M^{-1}cm^{-1}$ (450 nm) per subunit.
Coenzymes/Cofactors	Xanthine oxidoreductase possesses a molybdenum center (consisting of the metal plus pterin cofactor), two 2Fe/2S centers and flavin adenine dinucleotide.
Substrates	Xanthine oxidoreductase is able to hydroxylate a wide range of aromatic heterocycles, and also to hydroxylate a variety of aldehydes to the corresponding carboxylic acid. Depending on whether NAD^+ or O_2 is used in assays as oxidizing substrate, the enzyme is classified as either a dehydrogenase or oxidase.
Inhibitors	Xanthine oxidoreductase is inhibited by arsenite, cyanide, methanol and allopurinol (after enzymic oxidation to alloxanthine). This last inhibitor is available pharmaceutically as the antihyperuricemic drug Xyloprem.
Biological Functions	Xanthine oxidoreductase in liver catalyzes the final two steps of purine degradation, hydroxylating hypoxanthine to xanthine and xanthine to uric acid. The role of the enzyme in milk is unknown.
Physiology/Pathology	Surprisingly, xanthine oxidoreductase activity appears not to be critical in humans: the inhibitor Xyloprem is generally very well tolerated in the treatment of hyperuricemia, and the effects of genetic lesions due to a failure to synthesize the pterin cofactor of the molybdenum center appear to be due to loss of sulfite oxidase activity (another molybdenum-containing enzyme) rather than xanthine oxidoreductase. Xanthine oxidoreductase has been implicated in ischemia/reperfusion injury, associated with the conversion of the NAD^+- to the O_2-utilizing form of the enzyme, but its precise role in this phenomenon remains controversial.
Degradation	Unknown
Genetics/Abnormalities	Genetic lesions in the structural gene for xanthine oxidoreductase give rise to Type I xanthinuria, a rare autosomal recessive disorder characterized by urinary tract calculi or myositis, and occasionally renal failure. The severity of these conditions is variable, however, and some patients with homozygous xanthinuria are asymptomatic.
Half-life	Unknown
Concentration	Unknown
Isolation Method	The enzyme is most conveniently isolated from chicken liver as the NAD^+-utilizing dehydrogenase form, and from cow's milk as the O_2-utilizing oxidase form. Published isolation procedures typically involve ammonium sulfate fractionation, followed by column chromatography on hydroxyapatite, DEAE, CM or a combination of these.
Amino Acid Sequence	Xanthine oxidoreductase exhibits extensive sequence homologies to human aldehyde oxidase, and is a member of an extended family of molybdenum-containing hydroxylases that is phylogenetically widely dispersed, being found most eukaryotic, bacterial and archaeal sources. It possesses a distinct aa sequence FFxxYR that is conserved in enzymes from a variety of sources, and is distinct from the corresponding FLxKCP sequence of the aldehyde oxidases. This sequence corresponds to residues 400-405 in the human xanthine oxidoreductase and maps to the flavin

	domain of the protein. Covalent modification of the tyrosine residue present in this peptide results in loss of NAD^+ binding ability.
Disulfides/SH-Groups	The number of disulfide bonds in human xanthine oxidoreductase is not known. It is known, however, that conversion of the dehydrogenase to the oxidase form of the enzyme in some species can be reversibly brought about by oxidation of cysteine residues in the flavin domain of the protein to disulfides.
General References	Hille, R. *Chem. Rev.* 1996, **96**: 2757-2816. Hille, R., and Nishino, T. *FASEB J.* 1995, **9**: 995-1003. Romão, M.J., Archer, M., Moura, I., Moura, J.J.G., LeGall, J., Engh, R., Schneider, M., Hof, P., Huber, R. *Science* 1995, **270**: 1170-1176. Huber, R., Hof, P., Moura, J.J.G. et al. *Proc. Natl. Acad. Sci. USA* 1996, **93**:8846-8851.
Ref. for DNA/AA Sequences	Ichida, K., Amaya, Y., Noda, K. et al. *Gene* 1993, **133**:279-284.

Molecular models



The molybdenum center of xanthine oxidoreductase



The model above shows the protein structure of aldehyde oxidoreductase from *Desulfovibrio gigas* and is also valid for xanthine oxidoreductase (see Romão *et al.*, 1995) (after Hille, 1996, with permission).

Zn alpha-2-glycoprotein

Carlos López-Otin

Synonyms	GCDFP-44 (**G**ross **C**ystic **D**isease **F**luid **P**rotein of 44 kDa)
Abbreviations	Zn-α_2-gp
Classifications	Electr. mobility: α_2 region
Description	Zn-α_2-gp is a protein originally isolated from plasma and of as yet unknown biological function. Its name derives from its ability to be precipitated by zinc acetate, its electrophoretical mobility in the α_2-region and its high carbohydrate content. It is also present in other body fluids including sweat, saliva, seminal plasma and cyst fluid from women with gross cystic disease of the breast.
Structure	Zn-α_2-gp consists of a single polypeptide chain of 278 aa with three identical carbohydrate side chains, N-linked to Asn-92, Asn-108, and Asn-239. These carbohydrate chains are of the complex type and possess the typical biantennary structure. Calculation of the secondary structure reveals 23% α-helix, 27% β-sheet and 22% β-turn. The tertiary structure has not yet been determined.
Molecular Weight	40,000 - 44,000 (SDS-PAGE); 18% carbohydrate
Sedimentation Coeff.	3.2 S
Isoelectric Point	3.8-3.9 (0.1 M acetate-NaCl); isoionic 4.4
Extinction Coeff.	18.0 (278nm, 1%, 1cm, pH 7.0)
Enzyme Activity	None
Coenzymes/Cofactors	None
Substrates	None
Inhibitors	None
Biological Functions	The aa sequence and domain structure of Zn-α_2-gp reveal significant similarity to class I molecules of the major histocompatibility complex, although it lacks the transmembrane and cytoplasmatic domains characteristic of these transplantation antigens. On this basis, Zn-α_2-gp has been suggested to be involved in the immune response as a soluble MHC-like molecule. Other proposed roles for Zn-α_2-gp include transport of the nephritogenic renal glycoprotein, binding and transport of nonpolymorphic substances in plasma, and participation in cell adhesion processes.
Physiology/Pathology	Increased concentrations have been reported in mammary tissue from women with benign and malignant breast diseases, in some prostate carcinomas and in renal failure. Zn-α_2-gp expression in breast cancer cells is induced by androgens and glucocorticoids. Below normal concenctrations are apparently associated with chronic liver disease.
Degradation	Unknown

1. Installment 1992 (revised 1997)

Genetics/Abnormalities	The Zn-α_2-gp gene (AZGP1) is on chromosome 7q22, and is 9.7kb long with four exons which contain all information coding for the Zn-α_2-gp protein, including the 20 aa signal sequence. There are at least two Zn-α2-gp pseudogenes also located at 7q22. Studies on the aa sequence gave no clear evidence of polymorphism. A rare variant of asialo-Zn-α2-gp has been detected in a black individual. Nine rare variants phenotyped by isoelectric focusing and immunoblotting have been reported in Japanese, Korean, Chinese, and Papua New Guinean population.
Half-life	Unknown
Concentration	Serum: 40 - 75 mg/L; seminal fluid: 320 mg/L; breast cyst fluid: 800 - 900 mg/L.
Isolation Method	(a) Supernatant of Cohn fraction V from pooled normal human plasma, concentrated by ultrafiltration, followed by ammonium sulfate fractionation, zone electrophoresis and gel filtration. (b) size exclusion HPLC from breast cyst fluid.
Amino Acid Sequence	Pyroglutamic acid was found to be the N-terminal of the Zn-α_2-gp purified form plasma and from breast cyst fluid, whereas the protein from seminal plasma lacks this residue. Serine is the C-terminal aa. The content of Trp (8 residues) and Tyr (18 residues) is very high. Zn-α_2-gp aa sequence shows about 35% identity with that of the three extracellular domains (α_1, α_2, and α_3) of class I MHC-like molecules.
Disulfides/SH-Groups	Zn-α_2-gp contains four half-cystine residues which form two disulfide bonds, one between residues 103 and 166, and the other between residues 205 and 260.
General References	Bürgi, W. and Schmid, K. *J. Biol. Chem.* 1961, **236**:1066-1074. Bürgi W. et al. *Clin. Chem.* 1989, **35**:1649-1650. Sanchez, L.M. et al. *Cancer Res.* 1992, **52**:95-100. Takagaki, M. et al. *Biochem. Biophys. Res. Commun.* 1994, **201**:1339-1347. López-Boado, Y.S. et al. *Breast Cancer Res. Treatm.* 1994, **29**:247-258.
Ref. for DNA/AA Sequences	Araki, T. et al. *Proc. Natl. Acad. Sci. USA* 1988, **85**:679-683. Frejie, J.P. et al. *Genomics* 1993, 18:575-587.