CONTENTS

CHAPTER

15

Nucleic Acids

Model of DNA built by James D. Watson and Francis H.C. Crick at Cambridge University, 1953.

[*Courtesy: Science and Society Picture Library, Science Museum, London.*]

INTRODUCTION

The final class of biomolecules to be considered, the nucleotides plus molecules derived from them (*i.e.*, nucleic acids), represent a clear case in which last is not least. "Nucleotides themselves participate in a plethora of crucial supporting roles in cell metabolism, and the nucleic acids provide the script for everything that occurs in a cell" (Lehninger, Nelson and Cox, 1993).

Nucleotides are energy-rich compounds that drive metabolic processes (*esp*., biosynthetic) in all cells. They also serve as chemical signals, key links in cellular systems that respond to hormones and other extracellular stimuli, and are structural components of a number of enzyme cofactors and metabolic intermediates.

The **nucleic acids** (DNA and RNA) are the molecular repositories for genetic information and are jointly referred to as the 'molecules of heredity'. The structure of every protein, and ultimately of every cell constituent, is a product of information programmed into the nucleotide sequence of a cell's nucleic acids.

HISTORICAL RESUME

The nucleic acids have been the subject of biochemical investigations since 1869 when **Friedrich Miescher**, a 25-year old Swiss chemist, isolated nuclei from pus cells (white blood corpuscles) and found that they contained a hitherto unknown phosphate-rich substance, which he named *nuclein*. This substance was quite different from the carbohydrates, proteins and fats. He continued his

studies on salmon sperm, a prime source of nuclein and isolated it as a nucleoprotein complex in 1871, when he prophetically wrote :

> "It seems to me that a whole family of such phosphorus-containing substances, differing somewhat from each other, will emerge, as a group of nuclein substances, which perhaps will deserve equal consideration with the proteins."

FRIEDRICH MIESCHER (LT, 1844-1895), a Swiss and a student of the eminent German chemist, Felix Hoppe-Seyler, became interested in the chemistry of the nucleus in 1860s. He chose to work with white blood corpuscles (WBCs) because they contained a large and easily observable nucleus. A hospital at Tubingen supplied him with surgical bandages that had been peeled off purulent wounds. From the pus on the bandages, he obtained the WBCs, which he then treated with gastric juice (we now know that gastric juice contains an enzyme, pepsin, that digests protein). He observed microscopically that only the shrunken nuclei were left after the pepsin treatment; the remainder of the cell had dissolved. These nuclei were then analyzed and found to have different composition from any cellular component then known. It was called **nuclein.** Miescher continued his studies when he returned to his native city, Basel, in Switzerland. He soon found that a more convenient source of this material was salmon sperm. With this favourable material, he further purified nuclein. When all the proteins was completely removed, it became clear that the new material was an acid. It was then referred to as **nucleic acid.**

It was not until 1874, that he (Miescher) isolated pure nucleic acid from the DNA-protamine complex in salmon sperm nuclei. Later, it was found that the nuclein had acid properties and hence **Altmann**, in 1899, introduced the term *nucleic acid* to replace nuclein.

However, in 1880s, **Fischer** discovered purine and pyrimidine bases in nucleic acids. In 1881, **Zacharis** identified nuclein with chromatin. In 1882, **Sachs** stated that nucleins of sperm and egg are different and in 1884, **Hertwig** claimed that nuclein is responsible for the transmission of hereditary characteristics. The same year (*i.e.*, in 1894), Geheimrat Albrecht Kossel, of the University of Heidelberg, Germany, recognized that histones and protamines are associated with nucleic acids and also found that the histones were basic proteins. He was honoured with Nobel Prize for Physiology or Medicine in 1910 for demonstrating the presence of two purine and two pyrimidine bases in nucleic acids. In 1894, cytosine was identified and in 1909, uracil was isolated. **Levene**, a Russian-born biochemist, recognized the 5-carbon ribose sugar in 1910 and later also discovered deoxyribose in nucleic acids. In 1914, **Robert Feulgen**, another German chemist, demonstrated a colour test known as *Feulgen test* for the deoxyribonucleic acid.

P.A. Levine (1931) stressed that there are 2 types of nucleic acids, *viz.*, deoxyribonucleic acid and ribonucleic acid. In 1941, **Caspersson** and **Brachet**, independently, related that nucleic acids were connected to protein synthesis. **Oswald T. Avery, Colin M. MacLeod** and **Maclyn McCarty**, in 1944, first of all demonstrated that DNA is directly involved in inheritance. **Alfred D. Hershey** and **Martha J. Chase** of Cold Spring Harbor Lab., New York, in 1952, demonstrated that only the DNA of T4 bacteriophage enters the host, the bacterium *Escherichia coli*, whereas the protein (*i.e.*, capsid) remains behind. They, thus, confirmed that DNA is the genetic material of most living organisms. **Matthew S. Meselson** and **Franklin H. Stahl** (1957), at California Institute of Technology, presented evidence that nucleic acid forms the genetic material.

In 1953, **James D. Watson** and **Francis H.C. Crick** constructed the double helical model for the DNA molecule which could successfully explain DNA replication. In 1957, Arthur Kornberg proved the Watson-Crick model in the cell-free system. In 1967, he also synthesized a molecule of DNA from the 6,000 nucleotides.

DEFINITION

Although the name nucleic acid suggests their location in the nuclei of cells, certain of them are, however, also present in the cytoplasm. The nucleic acids are the hereditary determinants of

living orgainsms.They are the macromolecules present in most living cells either in the free state or bound to proteins as nucleoproteins. Like the proteins, *the nucleic acids are biopolymers of high molecular weight with mononucleotide as their repeating units*, just as amino acids are the repeating units of proteins. As regards their elemental composition, the nucleic acids contain carbon, hydrogen, oxygen, nitrogen and, strangely enough, phosphorus ; the percentage of the last two elements being about 15 and 10, respectively.

TYPES

There are two kinds of nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Both types of nucleic acids are present in all plants and animals. Viruses also contain nucleic acids; however, unlike a plant or animal, a virus has either RNA or DNA, but not both. The previously held view, that the DNA occurred only in animals and the RNA only in plants, is now known to be incorrect. DNA is found mainly in the chromatin of the cell nucleus whereas most of the RNA (90%) is present in the cell cytoplasm and a little (10%) in the nucleous. It may be added that extranuclear DNA also exists; it occurs, for example, in mitochondria and chloroplasts. Upon hydrolysis, under different set of conditions (Fig. 15–1), the two nucleic acids yield 3 components : phosphoric acid, a pentose sugar and nitrogenous bases. Table 15–1. summarizes the structural components of the two types of nucleic aicds, DNA and RNA.

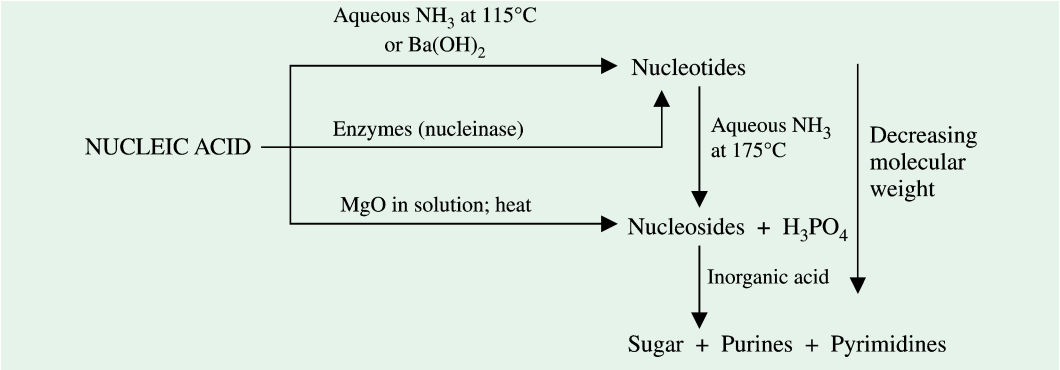


Fig. 15–1. Hydrolytic products of nucleic acid

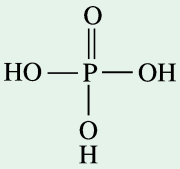
Table 15–1. Structural components of RNA and DNA

Components	*Ribonucleic acid*	*Deoxyribonucleic acid*
Acid	Phosphoric acid	Phosphoric acid
Pentose sugar	D-ribose	D-2-deoxyribose
Nitrogenous bases		
Purines	Adenine	Adenine
	Guanine	Guanine
Pyrimidines	Cytosine	Cytosine
	Uracil	Thymine

Each of the 3 components of nucleic acids is discussed sepatately.

Phosphoric Acid

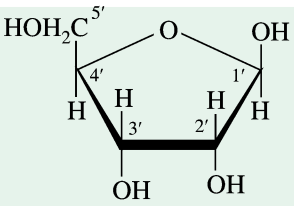
The molecular formula of phosphoric acid is H_3PO_4. It contains 3 monovalent hydroxyl groups and a divalent oxygen atom, all linked to the pentavalent phosphorus atom



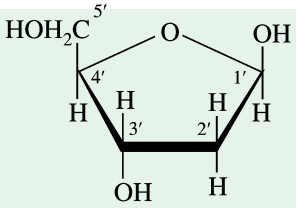
Phosphoric acid

Pentose Sugar

The two types of nucleic acids are distinguished primarily on the basis of the 5-carbon keto sugar or pentose which they possess. One possesses D-2-deoxyribose, hence the name deoxyribose nucleic acid or deoxyribonucleic acid, while the other contains D-ribose, hence the name ribose nucleic acid or ribouncleic acid. *Both these sugars in nucleic acids are present in the furanose form and are of* β *configuration.*



D-ribose
(β-D-ribofuranose)



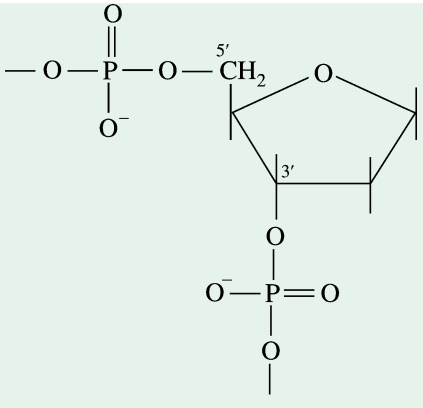
D-2-deoxyribose
(β-D-2-deoxyribofuranose)

[Note that in chemical nomenclature, the carbon atoms of sugars are designated by primed numbers, *i.e.*, C-1′, C-2′, C-3′ etc., while the various atoms in the bases lack the prime (′) sign and are designated by the cardinal numbers, *i.e.*, 1, 2, 3 etc.]

A perusal of the structure of the two types of sugars reveals that D-ribose is the parent sugar while D-2-deoxyribose is a derivative in which OH group on C2 has been replaced by an H atom.

The two sugars may be differentiated by means of specific colour reactions. Ribose reacts with orcinol in hydrochloric acid solution containing ferric chloride. Deoxyribose reacts with diphenylamine in acid solution.

An important property of the pentoses is their capacity to form esters with phosphoric acid. In this reaction the OH groups of the pentose, especially those at C3 and C5, are involved forming a 3′, 5′- phosphodiester bond between adjacent pentose residues. This bond, in fact, is an integral part of the structure of nucleic acids.



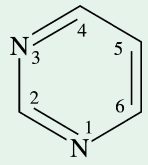
3′, 5′-phosphodiester bond

Nitrogenous Bases

Two types of nitrogenous bases are found in all nucleic acids. The base is linked to the sugar moiety by the same carbon (C1) used in sugar-sugar bonds. The nitrogenous bases are derivatives of pyrimidine and purine. Owing to their π electron clouds, both the pyrimidine and purine bases are planar molecules.

Pyrimidine Derivatives

These are all derived from their parent heterocyclic compound *pyrimidine*, which contains a six-membered ring with two-nitrogen atoms and three double bonds. It has a melting point of 22°C and a boiling point of 123.5°C.

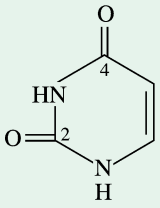


Pyrimidine

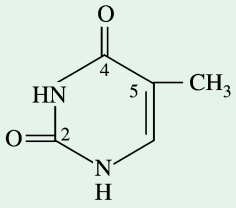
[For pyrimidine, the Chemical Abstracts use the numbering system, as depicted in the formula. This system is proposed by IUPAC. However, an older convention used a system, analogous to that indicated for the pyrimidine portion of purine, whose structural formula is given on the next page.]

The common pyrimidine derivatives found in nucleic acids are *uracil*, *thymine* and *cytosine* (Fig. 15–2).

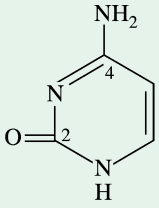
Uracil ($C_4H_4O_2N_2$), found in RNA molecules only, is a white, crystalline pyrimidine base with MW = 112.10 daltons and a m.p. 338°C. Only rarely does uracil occur in DNA.



Uracil
(2,4-dioxypyrimidine
or pyrimidinedione)



Thymine
(5-methyl-2, 4-dioxypyrimidine
or 5-methyluracil)



Cytosine
(2-oxy-4-aminopyrimidine
or aminopyrimidone)

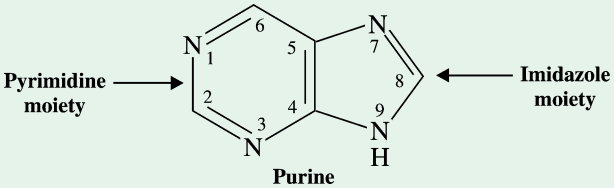
Fig. 15–2. Major pyrimidine derivatives

Thymine ($C_5H_6O_2N_2$), found in DNA molecules only, has MW = 126.13 daltons. It was first isolated from thymus, hence so named. Only rarely does thymine occcur in RNA.

Cytosine ($C_4H_5ON_3$), found in both RNA and DNA, is a white crystalline substance, with MW = 111.12 daltons and a m.p. 320-325°C.

Purine Derivatives

These are all derived from their parent compound *purine*, which contains a six-membered pyrimidine ring fused to the five-membered imidazole ring and is related to uric acid. It has a melting point of 216°C.



Purine

The prevalent purine derivatives found in nucleic acids are *adenine* and guanine (Fig. 15–3).

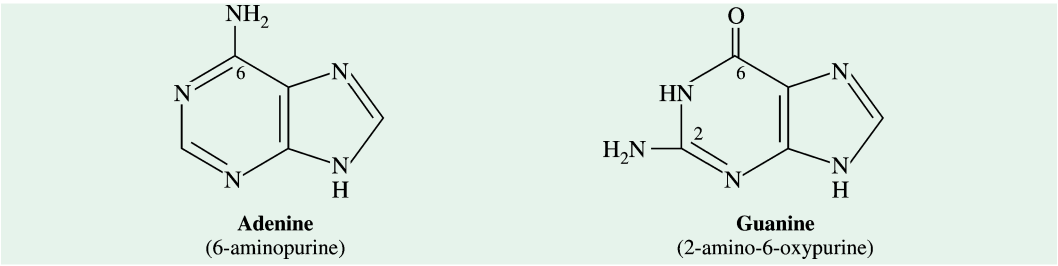


Fig. 15–3. Major purine derivatives

Adenine ($C_5H_5N_5$), found in both RNA and DNA, is a white crystalline purine base, with MW = 135.15 daltons and a m.p. 360-365°C.

Guanine ($C_5H_5ON_5$), also found in both RNA and DNA, is a colourless, insoluble crystalline substance, with MW = 151.15 daltons. It was first isolated from guano (bird manure), hence so named.

Other purine derivatives are also found in plants. For example, *caffeine* (1,3,7-trimethylxanthine) is present in coffee and tea and *theobromine* (3, 7-dimethylxanthine) is found in tea and cocoa. Both caffeine and theobromine have important pharmacological properties.

Modified Nitrogenous Bases

The above 5 nitrogenous bases were once believed to account for the total base composition of animal and plant nucleic acids and were hence designated as **major nitrogenous bases**. It is now known that other minor bases called **modified nitrogenous bases** also occur in polynucleotide structures. Some naturally-occurring forms of **modified pyrimidines** are shown in Fig. 15–4. The 5-methylcytosine (MC) is a common component of plant and animal DNA ; in fact, up to 25% of the cytosyl residues of plant genome are methylated. The DNA of T-even bacteriophages of *Escherichia coli* has no cytosine but instead has 5-*hydroxylmethylcytosine* and its glucoside derivatives.

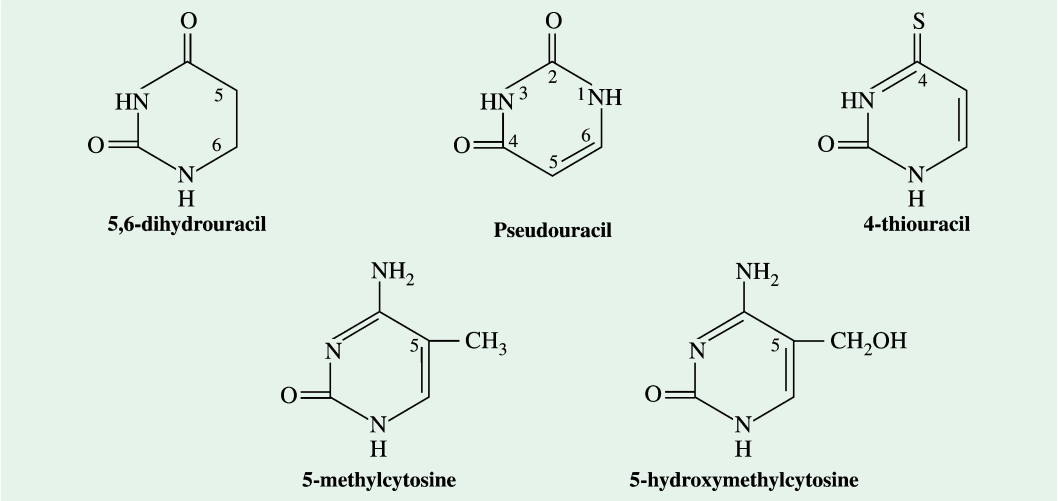


Fig. 15–4. Some modified (or minor) pyrimidines

Note that pseudouracil is identical to uracil ; the distinction is that point of attachment to the ribose-uracil is attached through N-1, the normal attachment point for pyrimidine derivatives, and pseudouracil is attached through C-5.

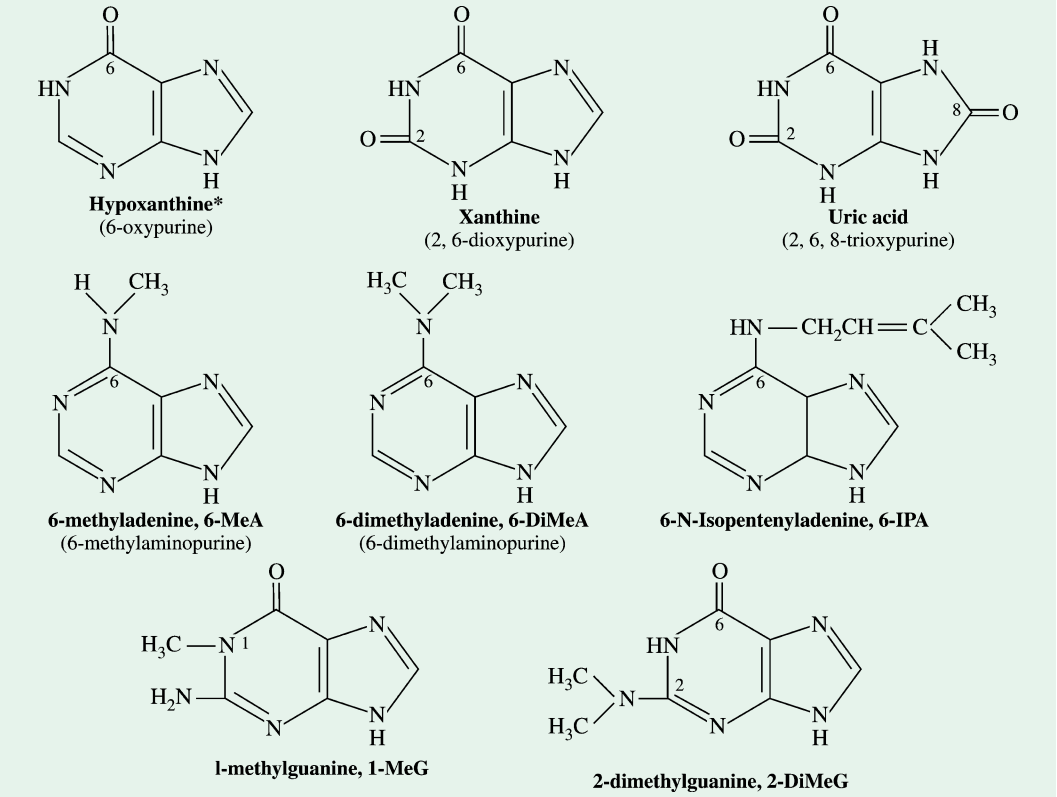


Fig. 15–5. Some modified (or minor) purines

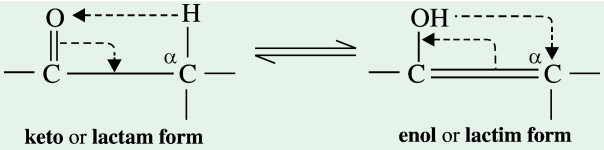
Among the **modified purines** (Fig. 15–5), some of them are found in transfer RNAs (tRNAs), which are a class of nucleic acids. *Methylation is the most common form of purine modification*. Methylation of purines (particularly of adenine) in DNA is now known to occur in the genetic material of microorganisms, and it is believed that plant genomes will also be shown to have methylated purines. The 6-methyladenine is found in bacterial DNA.

While names of nucleosides and nucleotides are generally derived from the corresponding bases, we have here an exception to this rule : the base corresponding to the nucleoside called inosine (and the derived nucleotides) is called **hypoxanthine**.

Note that pseudouracil is identical to uracil; the distinction is the point of attachment to the ribose: uracil is attached through N–1, the normal attachment point for pyrimidine derivatives, and pseudouracil is attached through C–5

Tautomerism in Nitrogenous Bases

Compounds that exist in 2 structural isomeric forms which are mutually interconvertible and exist in dynamic equilibrium are called **tautomers** and the phenomenon is termed **tautomerism**. The term tautomer (*tauto*^{G} = same ; *meros*^{G} = part) was first coined by Laar in 1885. *Aldehydes, ketones and other carbonyl compounds such as esters exhibit tautomerism*. Tautomerism, which expresses wandering tendency of an H atom, involves migration of a proton (H^+) from α-carbon to carbonyl oxygen by the following mechanism :



The tautomer containing the carbonyl group (= CO) is designated as the *keto* or *lactam* form and the other one having a hydroxy group (—OH) attached to a doubly-bonded carbon is referred to as the *enol* (alkENe + alcohOL) or *lactim* form. This kind of tautomerism is called *keto-enol* or more appropriately **lactam-lactim tautomerism.**

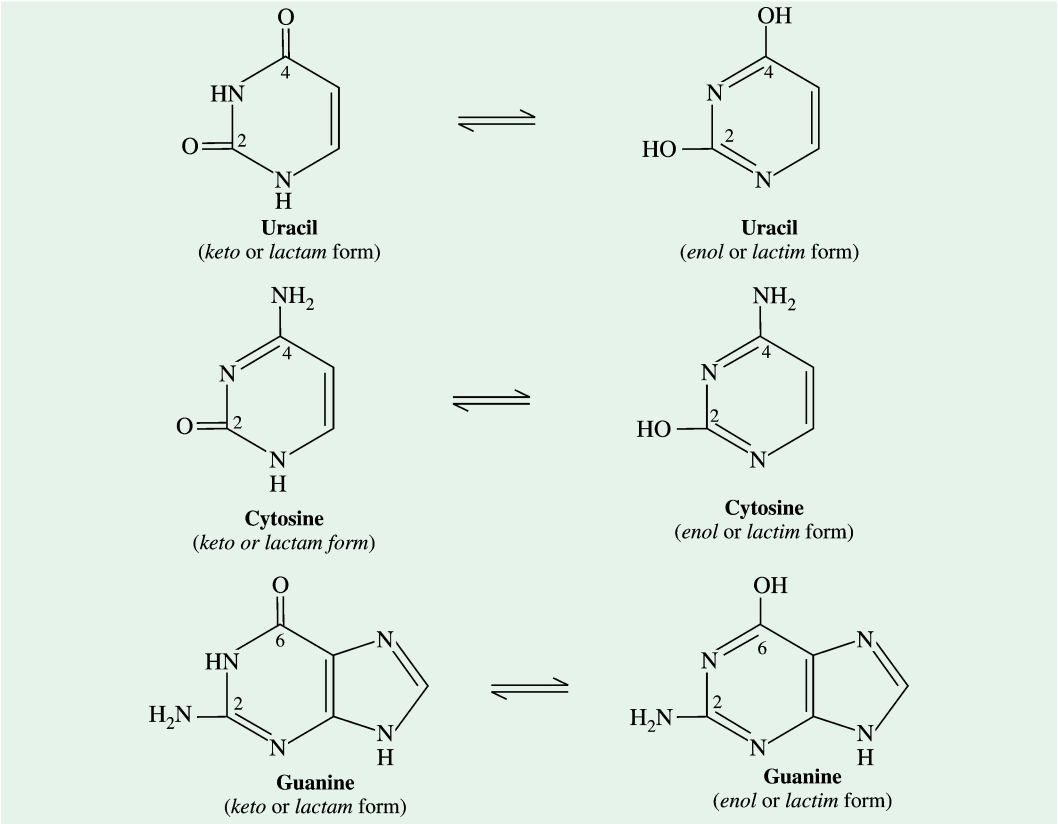


Fig. 15–6. Tautomeric forms of uracil, cytosine and guanine

The oxygen-containing nitrogenous bases, for example, uracil, thymine and cytosine among the pyrimidine derivatives and guanine, hypoxanthine, xanthine, uric acid, 1-methylgunanine, 2-dimethylguanine etc., among the purine derivatives exist in keto-enol (= lactam-lactim) forms. Three such examples are given in Fig. 15–6 and may be written in either form. However, *it is the keto (= lactam) form which predominates at neutral and acid pH values which are of physiological importance* ; the enol (= lactim) form becomes more prominent as pH decreases (*i.e.*, at acidic values).

NUCLEOSIDES

The nucleosides are compounds in which nitrogenous bases (purines and pyrimidines) are conjugated to the pentose sugars (*ribose or deoxyribose*) *by a* β*-glycosidic linkage.* This is, in fact, the configuration in the polymeric nucleic acids. The β-glycosidic linkage involves the C-1′ of sugar and the hydrogen atom of N-9 (in the case of purines) or N-1 (in the case of pyrimidines), thus eliminating a molecule of water. Therefore, the purine nucleosides are N-9 glycosides and the pyrimidine nucleosides are N-1 glycosides. Like the O-glycosides, the nucleosides are stable in alkali. Purine

The present author proposes the abbreviation '**ns**' for the nucleoside(s). This is in conformity with the abbreviation 'nt' meant for the nucleotide(s).

nucleosides are readily hydrolyzed by acid whereas pyrimidine nucleosides are hydrolyzed only after prolonged treatment with concentrated acid.

The nucleosides are generally named for the particular purine or pyrimidine present. Nucleosides containing ribose are called *ribonucleosides*, while those possessing deoxyribose as *deoxyribonucleosides*. Table 15–2 lists the trival names of the purine and pyrimidine nucleosides which are related to the bases that occur in RNA and DNA.

Table 15–2. The nucleosides

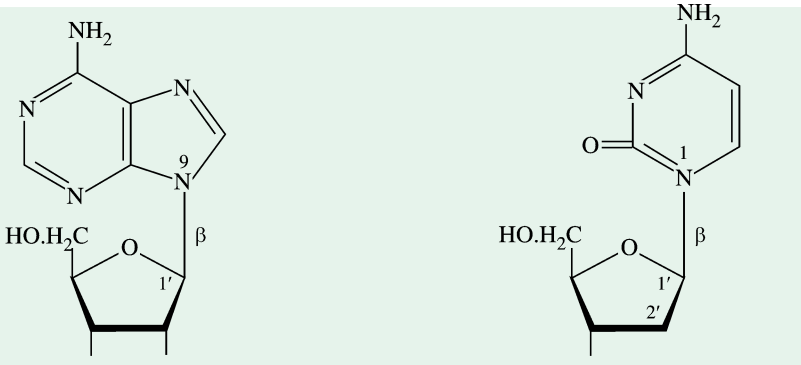
Base	*Sugar*	*Nucleoside*	*Trivial name**	*Abbreviation*
Ribonucleosides				
Adenine	Ribose	Adenine ribonucleoside	Adenosine	AR
Guanine	Ribose	Guanine ribonucleoside	Guanosine	GR
Cytosine	Ribose	Cytosine ribonucleoside	Cytidine	CR
Thymine	Ribose	Thymine ribonucleoside	Thymidine	TR
Uracil	Ribose	Uracil ribonucleoside	Uridine	UR
Deoxyribonucleosides				
Adenine	Deoxyribose	Adenine deoxyribonucleoside	Deoxyadenosine	AdR
Guanine	Deoxyribose	Guanine deoxyribonuceloside	Deoxyguanosine	GdR
Cytosine	Deoxyribose	Cytosine deoxyribonucleoside	Deoxycytidine	CdR
Thymine	Deoxyribose	Thymine deoxyribonucleoside	Deoxythymidine**	TdR
Uracil	Deoxyribose	Uracil deoxyribonucleoside	Deoxyuridine	UdR

* Note that the trivial names for the pyrimidine nucleosides end with the suffix *-dine*, whereas those of purine nucleosides end with the suffix *-sine*.

** The deoxyribonucleoside of thymine is usually termed as thymidine instead of deoxythymidine since this pyrimidine is primarily found in DNA. However, as thymine also occurs in one type of RNA termed transfer RNA (tRNA), a growing tendency amongst the biochemists is to call this deoxyribonucleoside as deoxythymidine in conformity to the nomenclature and to restrict the use of the term thymidine to the ribonucleoside of thymine only.

The structure of two nucleosides, one ribonucleoside (also called riboside) and another deoxyribonucleoside (or deoxyriboside) is given in Fig. 15–7.

Pseudouridine (ψU), an unusual nucleoside, is present as a constituent of the transfer RNAs. In it, the β-glycoside linkage occurs between the C1 of ribose and C5 (instead of N3) of uracil (Fig. 15–8). The uracil moiety at physiological pH has one oxy group in the *keto* form and the other in the *enol* form.



Adenosine
(9-β-D-ribofuranosyladenine)

Deoxycytidine
(1-β-2′-deoxy-D-ribofuranosylcytosine)

Fig. 15–7. A ribo- and a deoxyribonucleoside

Table 15–3. The common nucleotides

*Nucleotide**	*Trivial name*	*Abbreviations†*	
Ribouncleotides			
Adenosine-5′-monophosphate	Adenylic acid	AMP	Ado-5′-P
Guanosine-5′-monophosphate	Guanylic acid	GMP	Guo-5′-P
Cytidine-5′-monophosphate	Cytidylic acid	CMP	Cyd-5′-P
Uridine-5′-monophosphate	Uridylic acid	UMP	Urd-5′-P
2′-deoxyriobonucleotides			
Deoxyadenosine-5′-monophosphate	Deoxyadenylic acid	dAMP	dAdo-5′-P
Deoxyguanosine-5′-monophosphate	Deoxyguanylic acid	dGMP	dGuo-5′-P
Deoxyeytidine-5′-monophosphate	Deoxycytidylic acid	dCMP	dCyd-5′-P
Deoxythymidine-5′-monophosphate	Deoxythymidylic acid	dTMP	dThd-5′-P

* The table lists only the monophosphate derivatives at C5′ position of both ribonucleotides and 2′-deoxyribonucleotides. Among the other monophosphates of **ribonucleotides,** those at C2′ position are called as adenosine-2′-monophosphate (abbreviated as 2′-AMP or Ado-2′-P), cytidine-2′-monophosphate (2′-CMP or Cyd-2′-P) etc., and those at C3′ position as adenosine-3′-monophosphate (3′-AMP or Ado-3′-P), cytidine-3′-monophosphate (3′-CMP or Cyd-3′-P) etc. Similarly, in the case of **2′-deoxyribonulceotides,** the 3′-monophosphate derivatives are called as adenosine- 3′-monophosphate (3′-dAMP or dAdo-3′P), cytosine-3′-monophosphate (3′-dCMP or dCyd-3′-P) etc.

† In each case, the first abbreviation is more generally used. Note that in the case of 5′-monophosphate derivatives, the primed number 5′ is not written in the first system of abbreviation as these derivatives (as well as the 5′-diphosphates and 5′-triphosphates) are more important physiologically and are very frequently used in the literature. While adopting the second sytem of abbreviation, the diphosphates are written as Ado-5′-PP, Cyd-5′-PP etc., and the triphosphates as Ado-5′-PPP, Cyd-5′-PPP etc.

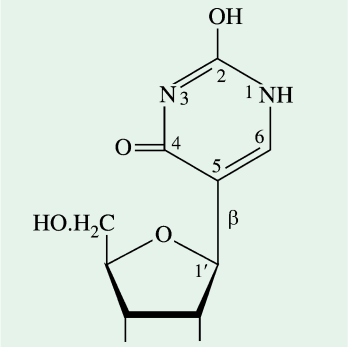


Fig. 15–8. Pesudouridine, ψU (5-ribosyluracil)
Note the absence of a true nucleosidic bond.

Nucleoside Analogues as Drugs

Recently two nucleoside analogues (Fig. 15–9), 3′-*azidodeoxythymidine* (AZT) and 2′, 3′-*dideoxycytidine* (DDC), have been therapeutically used for the treatment of acquired immune deficiency syndrome (AIDS) patients. The disease is caused by the human immunodeficiency viurs (HIV), which is an RNA virus that requires a specific enzyme, *an RNA-dependent DNA polymerase*, for its replication. When given to AIDS patients, AZT and DDC are converted into their triphosphate forms, which can then compete with dTTP and dCTP, respectively, as substrates for DNA synthesis. When incorporated, the analogues terminate DNA synthesis because the absence of a 3′-OH group in them prevents continued elongation of the DNA molecule being synthesized. As the DNA-synthesizing enzyme of HIV is much more sensitive to AZT and DDC inhibition than are the analogous enzymes of its host cell, these nucleoside analogues offer some hope as an effective treatment, if not a cure, for AIDS patients.

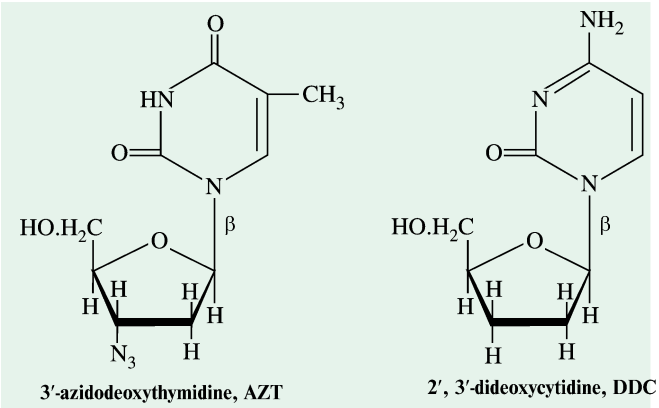


Fig. 15–9. Structure of two nucleoside analogues

NUCLEOTIDES

Nucleotides are the phosphoric acid esters of nucleosides. These occur either in the free form or as subunits in nucleic acids. As mentioned earlier, the phosphate is always esterified to the sugar moiety. The component units of nucleotides are shown in Fig 15-10.

In the ribose moiety of a **ribonucleoside**, phosphorylation is possible only at *three* positions (C2′, C3′, C5′) since C1′ and C4′ are involved in the furanose ring formation. In other words, the

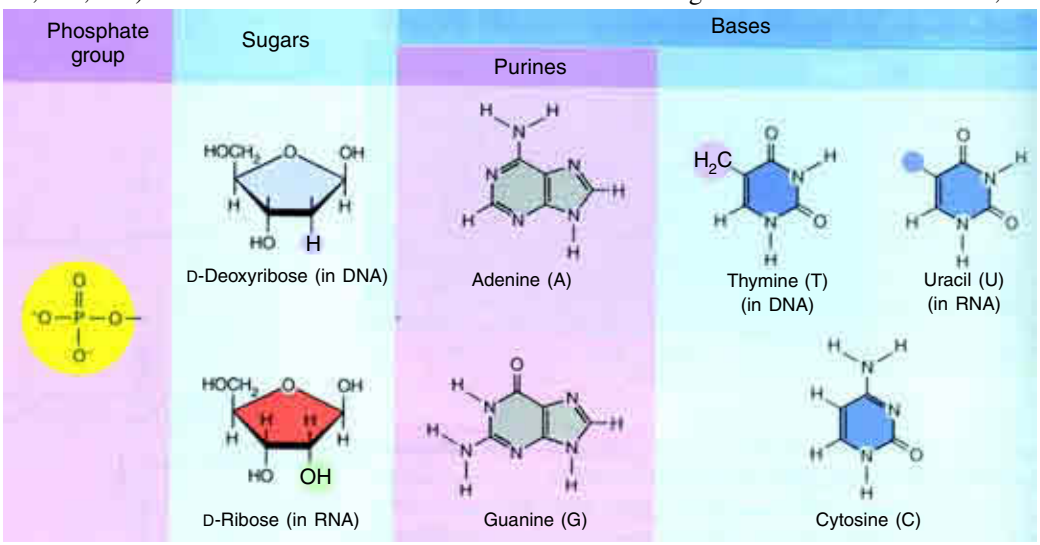
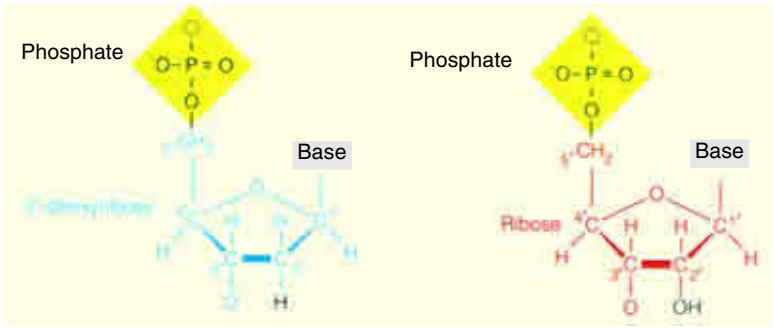


Fig. 15-10. The components of nucleotides

The three building blocks of a nucleotide are one or more phosphate groups, a sugar, and a base. The bases are categorized as purines (adenine and guanine) and pyrimidines (thymine, uracil, and cytosine).

phosphate group could be esterified only at these three places. On the contrary, in the deoxyribose component of a **2′-deoxyribonucleoside**, only *two* positions (C3′, C5′) are available for phosphorylation, since in this sugar C1′ and C4′ are involved in the furanose ring and C2′ does not bear a hydroxyl gorup. Accordingly, hydrolysis of the two nucleic acids, RNA and DNA, by various methods and under different set of conditions gives rise to isomeric nucleotides of 3 types and 2 types respectively. Table 15–3 lists some important ribonucleotides (also called ribotides) and deoxyribonucleotides (or deoxyribotides). While names of nucleosides and nucleotides are generally derived from the corresponding bases, there is one exception to



(*a*) Repeating unit of deoxyribonucleic and (DNA)

(*b*) Repeating unit of ribonucleic acid (RNA)

Fig. 15-11. Structure of nucleotides found in (*a*) DNA and (*b*) RNA

DNA contains deoxyribose as its sugar, and the bases A, T, G, and C RNA contains ribose as its sugar, and the bases A, U, G, and C.

In a DNA or RNA strand, the oxygen on the 3′ carbon is linked to the phosphorus atom of phosphate in the adjacent nucleotide. This bond formation removes the hydrogen attached to the 3′-oxygen atom, and it removes an oxygen (shown with a dashed line) attached to phosphorus.

this rule : the base corresponding to the nucleoside called inosine (and the derived nucleotides) is called hypoxanthine. Fig. 15-11 depicts the structure of nucleotides found in DNA and RNA both.

Functions of Nucleotides

In addition to their roles as the subunits of nucleic acids, nucleotides perform some other functions. These are enumerated below :

1. As carriers of chemical energy. Nucleotides may have one, two or three phosphate groups covalently linked at 5′-OH of ribose. These are referred to as **nucleoside mono-**, **di-** and **triphos-phates** and abbreviated as NMPs, NDPs and NTPs, respectively. The 3 phosphate groups are generally labelled as α, β and γ, *starting from the ribose*. NTPs are used as a source of chemical energy to drive many biochemical reactions. Adenosine triphosphate (ATP) is, by far, the most widely used (Fig. 15–12). Others such as uridine triphosphate (UTP), guanosine triphosphate (GTP) and cytidine triphosphate (CTP) are used in specific reactions.

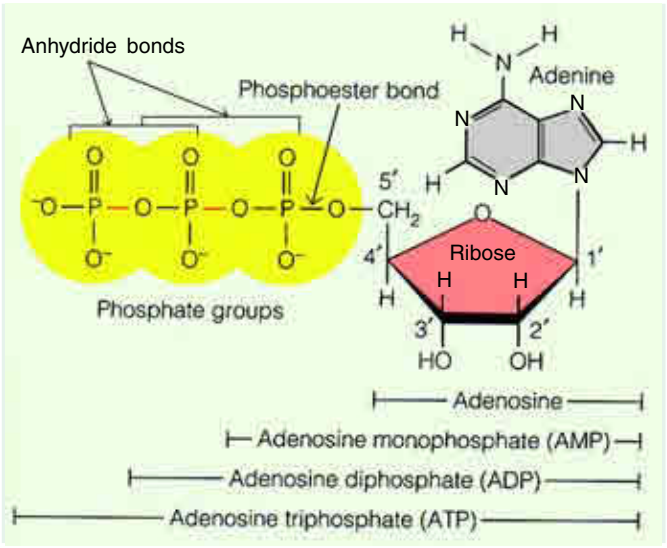


Fig. 15–12. The phosphate ester and phosphoric acid anhydride bonds of ATP
Hydrolysis of an anhydride bond yields more energy than hydrolysis of the ester.

The hydrolysis of ATP and other nucleoside triphosphates is an exergonic reaction. The bond between the ribose and the α-phosphate is an ester linkage. The α-β and β-γ linkages are phosphoric acid anhydrides. Hydrolysis of the ester linkage yields about 14 kJ/mol, whereas hydrolysis of each of the anhydride bond yields about 30 kJ/mol. In biosynthesis, ATP hydrolysis often drives less favourable metabolic reactions (*i.e.*, those with $\Delta G^{\circ\prime} > 0$). When coupled to a reaction with a positive free-energy change, ATP hydrolysis shifts the equilibrium of the overall process to favour product formation.

2. As components of enzyme factors. Many enzyme cofactors and coenzymes (such as coenzyme A, NAD^+ and FAD) contain adenosine as part of their structure (Fig. 15–13). They differ from each other except for the presence of adenosine. In these cofactors, adenosine does not participate directly, but removal of adenosine from these cofactors usually results in drastic reduction of their activities. For instance, removal of adenosine nucleotide from acetoacetyl-CoA reduces its reactivity as a substrate for β-ketoacyl-CoA transferase, an enzyme of lipid metabolism, by a factor of 10^6.

3. As chemical messengers. The cells respond to their environment by taking cues from hormones or other chemical signals in the surrounding medium. The interaction of these chemical signals (first messengers) with receptors on the cell surface often leads to the formation of **second messengers** (Fig. 15–13) inside the cell, which in turn lead to adaptive changes inside the cell. *Often, the second messenger is a nucleotide*.

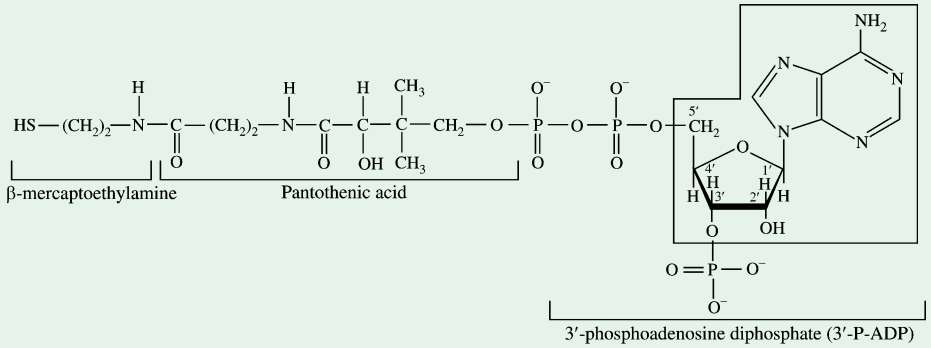
One of the most common second messengers is the nucleotide **adenosine 3′, 5′- cyclic monophosphate** (**cyclic AMP or cAMP**), formed from ATP in a reaction catalyzed by *adenylate cyclase*, associated with the inner face of the plasma membrane. Cyclic AMP serves regulatory functions in virtually every cell outside the plant world. **Guanosine 3′, 5′- cyclic monophosphate** (**cGMP**) occurs in many cells and also has regulatory functions.

DEOXYRIBONUCLEIC ACID

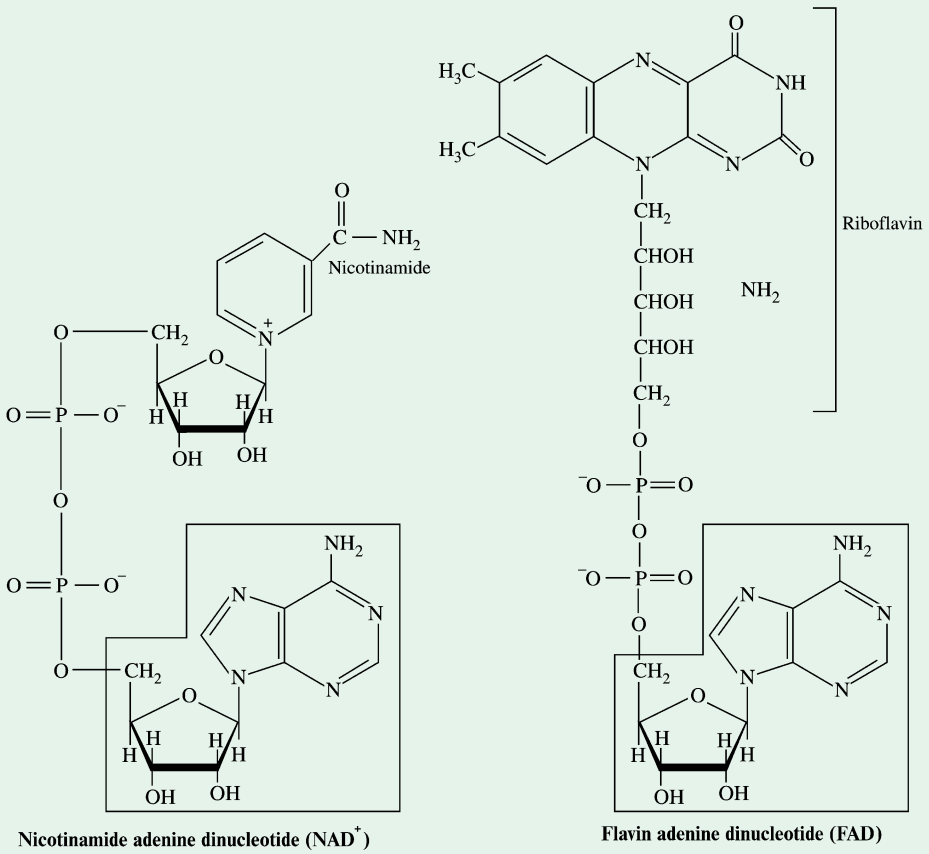
A nucleic acid may be visualized as a polymer of a nucleotide monomer. In other words, it may be considered as a polynucleotide.

DNA as .... $(\text{base-deoxyribose-phosphate})_n$
RNA as .... $(\text{base-ribose-phosphate})_n$

The name **'polymer'** was coined by Jöns Jacob Berzelius in 1827.



Coenzyme A (Co–A)



Nicotinamide adenine dinucleotide (NAD^+) **Flavin adenine dinucleotide (FAD)**

Fig. 15–13. Enzyme cofactors and coenzymes containing adenosine as their component
The adenosine portion is shown in an enclosure in each case.

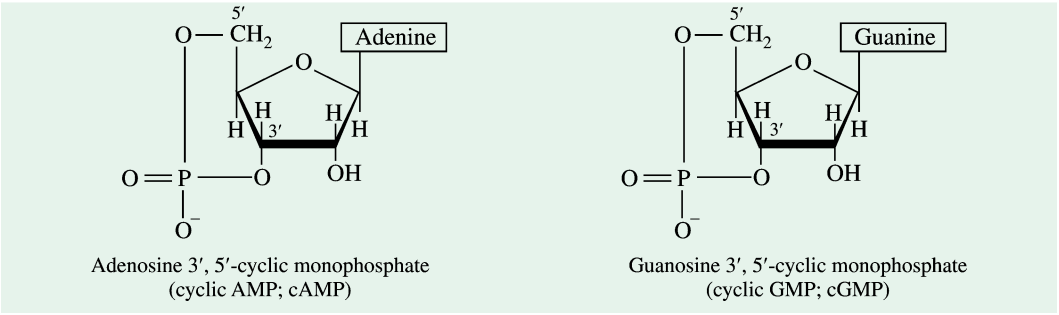


Fig. 15–14. Structures of cyclic AMP and cyclic GMP

The constituent units are coupled, as already stated, by means of 3′, 5′-phosphodiester bonds. The nature, properties and function of the two nucleic acids (DNA and RNA) depend on the exact order of the purine and pyrimidine bases in the molecule. This sequence of specific bases is termed the **primary structure.** *The purine and pyrimidine bases of DNA carry genetic information whereas the sugar and phosphate groups perform a structural role.* Interestingly enough, the human body contains about 0.5 gm of DNA!

The terms, primary, secondary and tertiary structures for nucleic acids have much the same meaning as they do for proteins Fig. 15-15. The *primary structure* is the order of the nucleotides in the chain; the term *secondary structure* relates to regions of regular conformation of the chain, stabilized by regular, repeating interactions (*e.g.*, the double helix of DNA) ; and the *tertiary structure* is the overall conformation of the chain.

Both DNA and protein are composed of a linear sequence of subunits. *The two sequences are colinear, i.e.*, the nucleotides in DNA are arranged in an order of the amino acids in the protein they specify. It is, thus, apparent that the DNA sequence contains a coded specification of the protein structure.

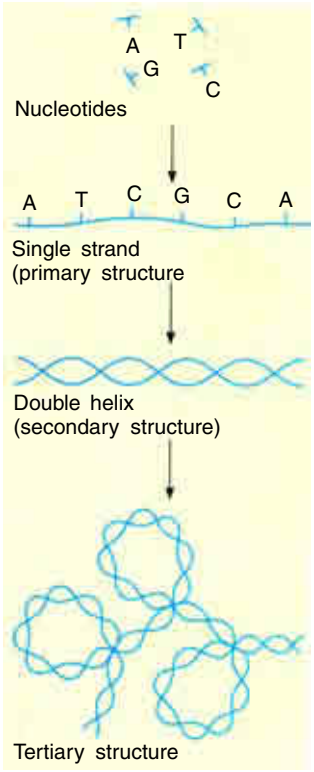
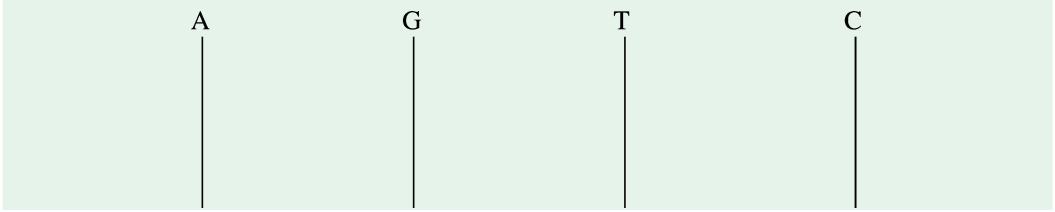


Fig. 15-15. Levels of nucleic acid structure

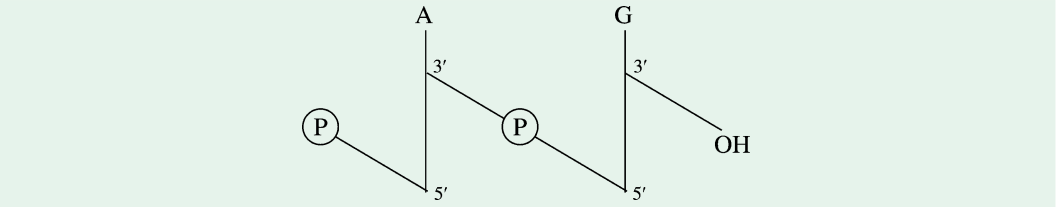
Internucleotide Linkages

The *backbone* of DNA (Fig. 15–15), which is constant throughout the molecule, consists of deoxyriboses linked by phosphodiester bridges. The 3′-hydroxyl of the adjacent sugar moiety of one deoxyribonucleotide is joined to the 5′-hydroxyl of the adjacent sugar by an internucleotide linkage called a phosphodiester bond. *The variable part of DNA is its sequence of bases.*

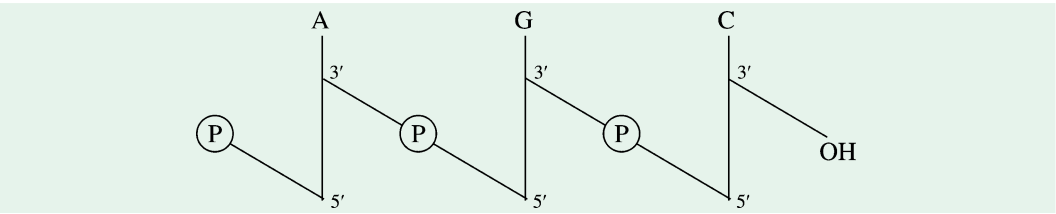
Representation of Nucleic Acid Backbones. The writing of the molecular structure of DNA chain, as depicted in Fig. 15–15, is tedious and time-taking. Henceforth, a system of schematic representation has been evolved. According to this system, the symbols for the 4 principal deoxyribonucleosides are :



The bold line refers to the sugar and the letters A, G, T and C represent the bases. The encircled P within the diagonal line in the diagram (given below) represents a phosphodiester bond. This diagonal line joins the middle of one bold line and the end of another. These junctions represent the 3′-OH and 5′-OH respectively. In this example, the symbol (P) indicates that deoxyadenylate is linked to the deoxyguanosine unit by a phosphodiester bond. The 3′-OH of deoxyadenylate is linked to the 5′-OH of deoxyguanosine by a phosphate group.



In the above dinucleotide, suppose deoxycytidine unit now becomes linked to the deoxyguanosine unit. The resulting trinucleotide can be schematically represented as follows :



For writing long sequences of polynucleotides, a **shorthand system** is followed. The letters A, G, T, C and U represent the nucleosides as in Table 15.3. The phosphate group is represented by *p*. When *p* is placed to the left of the nucleoside symbol, esterification is at C-5′ ; when placed to the right, esterification is at C-3′. Thus, ApGp is a dinucleotide with a monoester at C-3′ of a guanosine and a phosphodiester bond between C-5′ of G and C-3′ of A. In order to specify the type of sugar involved, the letter *d* (for deoxyribose) is prefixed to the above notation if all the nucleosides contain deoxyribose. Thus, the perfect notation for this dinucleotide shall be d-ApGp.

A short nucleic acid is referred to as an **oligonucleotide**. The definition of "short" is somewhat arbitrary, but the term oligonucleotide is often used for polymers containing 50 or fewer nucleotides. A longer nucleic acid is called a **polynucleotide**.

Base Composition of DNA

The base composition of DNA from a number of sources has been worked out by a number

Erwin Chargaff

Chargaff, an Austrian refugee biochemist, was born in 1905 in Czernowitz, Austria (now Chernovtsy, the Ukraine). He received his Ph.D. from the University of Vienna in 1928, and then after periods of Yale, Berlin and Paris, joined Columbia University, New York city, in 1935. He has worked in many fields of biochemistry and his research papers include studies on subjects as diverse as lipid metabolism and blood coagulation. However, his most influential contribution was the demonstration during 1945-50 that the base ratios in DNA are constant. Using the simple but sensitive technique of paper chromatography, he along with his collaborators at Columbia University analyzed the base composition of DNA from various sources. Although this was one of the seminal discoveries that led to the structure of DNA, Chargaff has never claimed any allegiance to molecular biology, and indeed has become one of the sternest critics. He is the source of many pungent comments about contemporary science and scientist. He described Watson and Crick as two pitchmen in the search of a helix and explains his objection to molecular biology as 'by its claim to be able to explain everything, is acutally inpedes the flow of scientific explanation'. His autobiography is ***Heraclitean Fire: Sketches From a Life Before Nature***, Rockefeller University Press, New York, 1978. He believed in the dictum, 'we can only seek what we have already found' Chargaff is also famous for his following quote:

"What counts..... in science is to be not so much the first as the last."

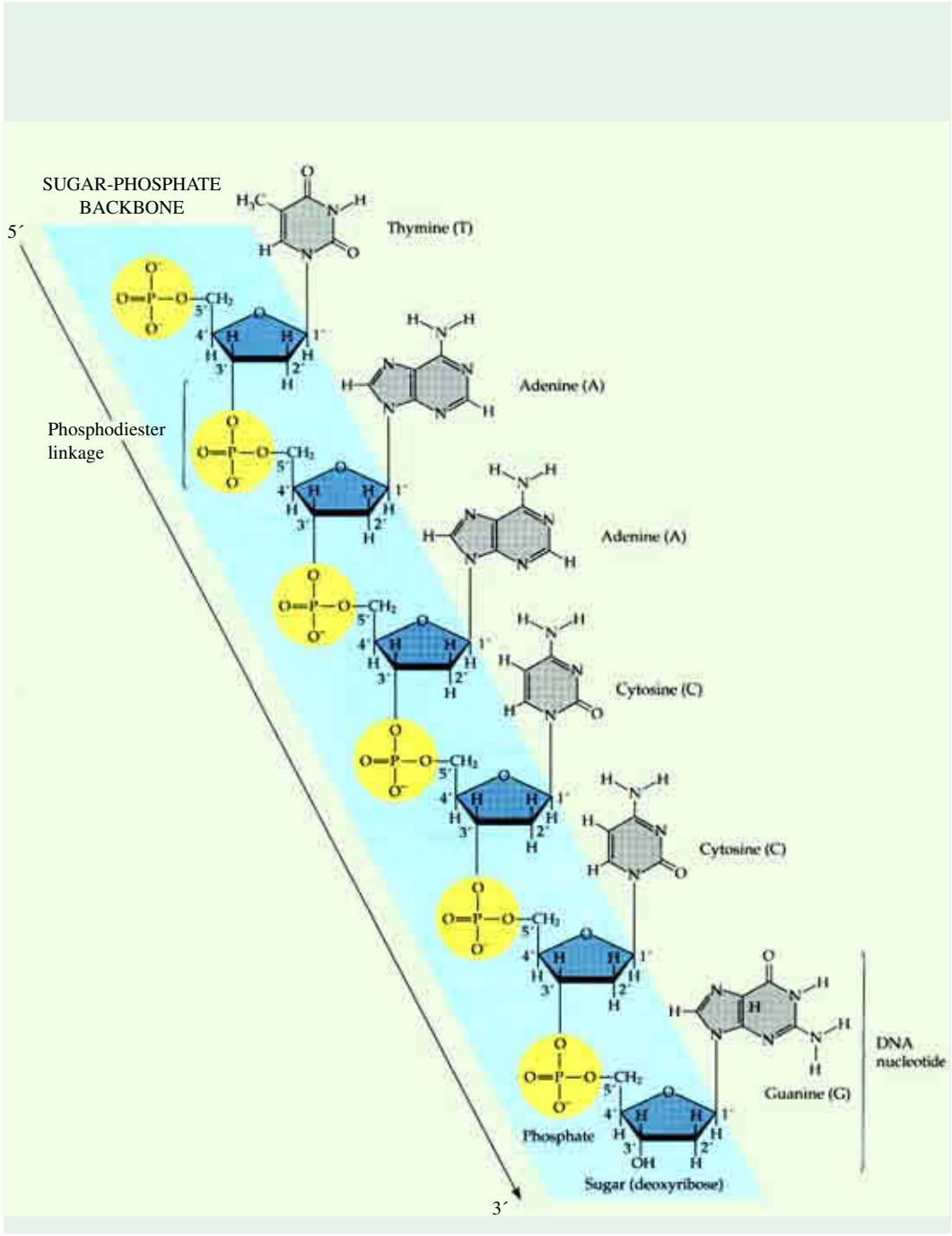


Fig. 15–16. A short strand of DNA containing six nucleotides

Nucleotides are covalently linked together to form a strand of DNA. Note the position of the internucleotide linkage between C3′ and C5′.

Table 15–4. **DNA composition of various species**

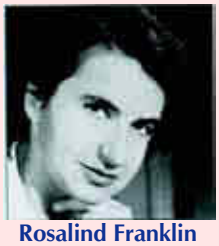
Species	Base proportions, moles%*					Pu/Py ratio $\frac{A+G}{T+C(+MC)}$	$\frac{A}{T}$	$\frac{G}{C(+MC)}$	Dissymmetry ratio $\frac{A+T}{G+C\,(+MC)}$
	Adenine A	*Guanine G*	*Thymine T*	*Cytosine C*	*Methyl-cytosine MC*				
1. *Sarcina lutea*	13.4	37.1	12.4	37.1	—	1.02	1.08	1.00	0.35
2. *Acaligenes faecalis*	16.5	33.9	16.8	32.8	—	1.02	0.98	1.03	0.50
3. *Brucella abortus*	21.0	20.0	21.1	28.9	—	1.00	1.00	1.00	0.73
4. *Escherichia coli K 12*	26.0	24.9	23.9	25.2	—	1.08	1.09	0.99	1.00
5. *Salmonella paratyphi A*	24.8	24.9	25.3	25.0	—	0.99	0.98	1.00	1.00
6. Wheat germ	28.1	21.8	27.4	16.8	5.9	0.99	1.02	0.96	1.25
7. Rat bone marrow	28.6	21.4	28.4	20.4	1.1	1.00	1.00	0.99	1.33
8. Ox	29.0	21.2	28.7	19.9	1.3	1.00	1.01	1.00	1.36
9. *Staphyllococcus aureus*	30.8	21.0	29.2	19.0	—	1.07	1.05	1.11	1.50
10. Human thymus	30.9	19.9	29.4	19.8	—	1.03	1.05	1.01	1.52
11. Human liver	30.3	19.5	30.8	19.9	—	0.99	1.00	0.98	1.54
12. *Saccharomyces cerevisiae*	31.7	18.3	32.6	17.4	—	1.00	0.97	1.05	1.80
13. *Paracentrus lividus*	32.8	17.1	32.1	16.2	1.1	1.01	1.02	0.98	1.85
14. *Pasteurella tularensis*	32.4	17.6	32.9	17.1	—	1.00	0.98	1.03	1.88
15. *Clostridium perfringens*	36.9	14.0	36.3	12.8	—	1.04	1.02	1.09	2.70

* The mole per cent is the number of molecules of one base divided by the total number of molecules of all 4 bases, multiplied by 100. The data are accurate to ± 1 per cent.

Contents

Maurice Wilkins (born in 1916 at Pongaroa, New Zealand) and **Rosalind Franklin** (LT, 1920-1958)

Maurice Wilkins



Rosalind Franklin

Maurice Wilkins began as a physicist (BA Cambridge 1938, PhD Birmingham 1940) but his experiences on the atomic bomb project during the War turned him away from physical science and towards biology. He joined King's College, London, in 1947 and began studying the structure of chromosomes and genes using physical methods. In May 1950 he was given by Rudolf Signer of Bern a sample of what was probably the purest DNA available at the time, and from it obtained extremely detailed X-ray diffraction patterns. In January 1951 Rosalind Franklin joined King's from Paris and took over analysis of the Signer DNA, producing pictures that showed clearly that DNA is a helix. Unfortunately there was a personality clash between Franklin and Wilkins and the intellectual discussion needed to solve the structure of DNA (which Watson and Crick enjoyed) never took place at King's. Neverthless, Franklin was very close to a double helix structure when Watson and Crick announed their results.

Franklin moved to Birkbeck College to work on virus structure in 1953 and remained there until she died of breast cancer - at the premature age of 38. Most scientists agree that she had played a crucial but under-appreciated role in elucidation of the structure of DNA. Her portrayal in *The Double Helix* as 'Rosie' is now considered a fiction and it is clear from accounts of her contemporaries that, had she lived, she would probably have become one of the most eminent British scientist of her day (see *Rosalind Fanklin and DNA* by A. Sayre in 'Reading'). Wilkins continued with DNA analysis for a number of years and shared the 1962 Nobel Prize with Watson and Crick. In 1969 he became the founding President of the British Society for Social Responsibility in Science.

of investigators. Table 15–4 lists the DNA composition of various species. These analyses show a characteristic regularity of pattern as first noted by Erwin Chargaff and his coworkers in 1950. The important conclusions drawn by him are :

1. The sum of purines (Pu) is equal to the sum of pyrimidines (Py), *i.e.*, Pu/Py = 1. In other words, A + G = T + C (+ MC) where MC stands for methylcytosine, where it occurs.
2. The ratio of adenine to thymine is also one, *i.e.*, A/T = 1.
3. The ratio of guanine to cytosine (plus methylcytosine where it occurs) is also one, *i.e.*, G/C(+ MC) = 1.
4. Bases with 6-amino groups are equal to bases with 6-keto (hydroxyl) groups, *i.e.*, A + C (+ MC) = G + T.
5. The ratio of A + T/G + C(+ MC), known as *dissymmetry ratio*, varies greatly from one species of DNA to the other and is characteristic of that species. When the dissymmetry ratio exceeds one, such a DNA is called *AT type* ; when the value is less than one, such a DNA is designated as *GC type*. In bacteria, both AT and GC types of DNA are found. However, in the higher organisms, the range of dissymmetry ratio is more limited : in most animals the value ranges from 1.3 to 2.2 and in higher plants from 1.1 to 1.7. The value of dissymmetry ratio in human beings is 1.4 and that in *Mycobacterium tuberculosis* is 0.60.

Chargaff's data suggest that A is always paired with T and G is always paired with C. A-paired bases occur in any order :

A...T	T...A	C...G	T...A
T...A	C...G	A...T	A...T
C...G	A...T	T...A	C...G
1	2	3	4

The variability that can be observed is overwhelming. For example, it has been calculated that human chromosome contains on the average about 140 million base pairs. Since any of the 4 possible nucleotides can be present at each nucleotide position, the total number of possible nucleotide sequences is $4^{140} \times 10^6$ or $4^{14,00,00,000}$. No wonder each species has its own base percentages!

Only the 2 purines (adenine and guanine) and the 2 pyrimidines (thymine and cytosine) have been found in DNA molecule of most microorganisms such as bacteria, actinomycetes, algae, fungi and protozoa. However, the presence of methylcytosine, a pyrimidine, is a characteristic feature of DNA of higher plants and animals ; *the DNA of plants being richer in methylcytosine than the DNA of animals*. The richest source of methylcytosine, found as yet, is wheat germ which contains 6 moles of methylcytosine per 100 moles of bases (refer Table 15–4).

Evolution of Watson-Crick Model

W. T. Astbury was the first person to give any thought to the 3-'D' structure of DNA. By his x-ray crystallographic studies on DNA molecule, conducted in 1940s, he concluded that because DNA has a high density, so its polynucleotide was a stack of flat nucleotides, each of which was oriented perpendicularly to the long axis of the molecule and was placed 3.4 Å apart from each other, a concept given the name a **"3.4 Å repeat".** The x-ray crystallographic studies of Astbury were continued in early 1950s by **Maurice Wilkins** and his student **Rosalind Franklin*,** both of King's College, London, who obtained a superior x-ray diffraction photograph of a DNA fibre (Fig. 15-17) which confirmed Astbury's earlier inference of 3.4 Å internucleotide distance and suggested a helical configuration for DNA molecule. They also proposed a **"34 Å repeat"**, *i.e.*, the helix is folded into many turns and each turn causes a vertical rise of 34 Å. Furthermore their findings indicated a helical structure containing two or more strands of polydeoxyribonucleotides. **Erwin Chargaff**, in late 1940s, independently had provided the crucial observation that, in DNA obtained from a wide variety of organisms, the molar ratio of adenine to thymine and that of guanine to cytosine were very close to unity. These results indicated that a specific relationship must exist between the two bases within each of the ratios. These and some other generalizations (discussed earlier) were collectively termed as **"Chargaff's equivalence rules".** Later, the results of titration studies (= analytical studies) also suggested that the long polynucleotide chains were held together by hydrogen bonding between the base residues.

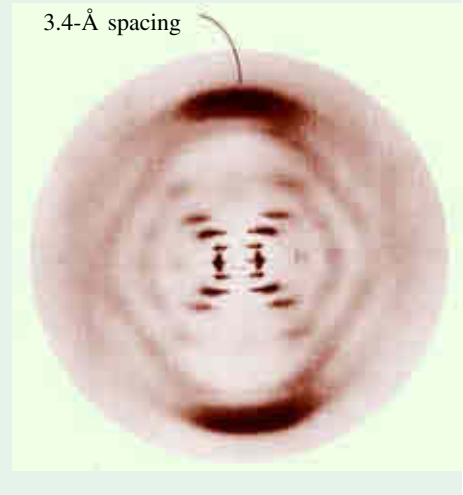


Fig. 15–17. X-ray diffraction photograph of a hydrated DNA fibre

The spots forming a cross in the centre denote a helical structure. The strong arcs at the top and bottom correspond to the stack of base pairs, which are 3.4 Å apart.

Until the 1950s, tetranucleotides with the four different bases in varying arrangements were believed to be the basic repeating unit of RNA structure, and the analogous deoxytetranucleotides the basic repeating unit of DNA. This so-called "**tetranucleotide hypothesis**" conferred nucleic acids a monotonous structure of repeating tetranucleotides which, however, could not account for the idea of DNA as genetic material as was first proposed by **Avery et al** in 1944. Later findings that DNA bases are not arranged in repetitive tetranucleotide sequences, dispelled doubts about the hereditary role of the nucleic acids. **Linus Pauling** suggested that the nitrogenous bases projected outwards from a central core formed by the backbone phosphate group. But Franklin felt she had

evidence that phosphate pointed towards outside and that the bases were in the centre.

This was the situation around 1951 when **James Dewey Watson**, a 22-year-old American postdoctoral research fellow arrived in Cambridge and met **Francis Harry Compton Crick,** a physicist working for his Ph.D. degree in Biophysics. Although they were to work on different problems, they decided to collaborate on the study of DNA. Using intelligently all the available information, Watson and Crick, in April, 1953, published an article proposing the double helical structure for DNA molecule in the same issue of journal 'Nature', in which Wilkins *et al* presented the x-ray evidence for that structure. Thus, we see that Watson-Crick structural model of DNA is the outcome of 3 types of studies (x-ray diffraction, base-pairing and analytical), admirably reasoned, interwoven and adjusted into a hypothetical double helical structure. In fact, their discovery of the double helix represents an elegant case of scientific sleuthing in the literature, thinking and master model building. This brilliant accomplishment ranks as one of the most significant in the history of biology because it led the way to an understanding of gene function in molecular terms. For this epoch-making discovery, Watson and Crick were awarded the prestigious Nobel Prize in medicine or physiology in 1962. In fact, their model of DNA structure was radically different from that suggested by Linus Pauling and Robert Corey in February, 1953 which was a triple helix. The latter model, however, failed to explain the process by which a DNA molecule copies itself (*i.e.*, DNA replication) whereas Watson and Crick's model admirably explained this phenomenon.

In fact, Watson and Crick shared the Nobel Prize with Wilkins, although elucidation of the double helical structure of DNA was a concerted effort on the part of 5 scientists as **Watson** had, truly, once said, "actually it was a matter of 5 persons: Maurice Wilkins, Rosalind Franklin, Linus Pauling, F.H.C. Crick and me."

Double Helical Structure of DNA (Watson-Crick Model)

The *salient features* of the **Watson-Crick model** (Fig. 15-18) for the commonly-found DNA (known as the B-DNA) are :

1. DNA molecule consists of two helical polynucleotide chains which are coiled around (or wrapped about) a common axis in the form of a right-handed double helix (refer Fig. 15–19). The two helices are wound in such a way so as to produce 2 interchain spacings or grooves, a major or wide groove (width 12 Å, depth 8.5 Å) and a minor or

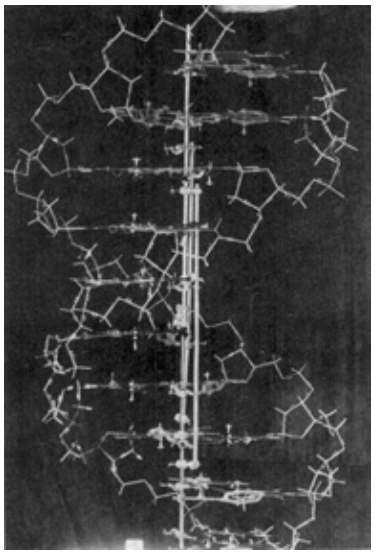


Fig. 15–18. The original molecular model of DNA double helix, proposed by Watson and Crick
The model shows a hydrogen-bonded to T and G hydrogen bonded with C.

Sometimes, one chain or strand of DNA is called the Watson strand (W strand) and its complement the **Crick strand** (C strand) as a gesture to the propounders of this model.

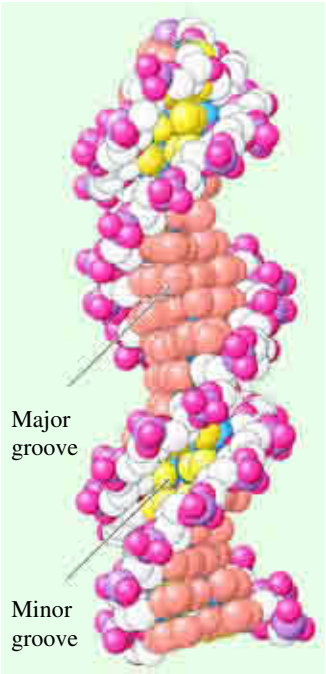


Fig. 15-19. Space-filing model of B-form DNA showing major and minor grooves
The major groove is depicted in orange, and the minor groove is depicted in yellow. The carbon atoms of the backbone are shown in white. Note that the bases fill the space between the two backbones. Atoms are shown at a size that includes their Van der Waals radii.

narrow groove (width 6 Å, depth 7.5 Å). The two grooves arise because the glycosidic bonds of a base pair are not diametrically opposite each other. The minor groove contains the pyrimidine O-2 and the purine N-3 of the base pair, and the major groove is on the opposite side of the pair. Each groove is lined by potential hydrogen bond donor and acceptor atoms. It is noteworthy that the major groove displays more distinctive features than does the minor groove. Also, the larger groove is more accessible to proteins interacting with specific sequences of DNA.

The two helices wind along the molecule parallel to the phosphodiester backbones. These may be visualized as to have formed when two wires are wound about a pencil or a common axis. In these grooves, specific proteins interact with DNA molecules. Such double helices cannot be pulled apart and can be separated only by an unwinding process. They are called as *plectonemic coils*, *i.e.*, coils which are interlocked about the same axis. The two chains run in opposite direction, *i.e.*, they are *antiparallel* which means that in one strand, the ring oxygen of sugar moieties face upwards while in the other strand, the ring oxygen faces downwards. Also they are not identical but, because of base pairing, are *complementary* to each other.

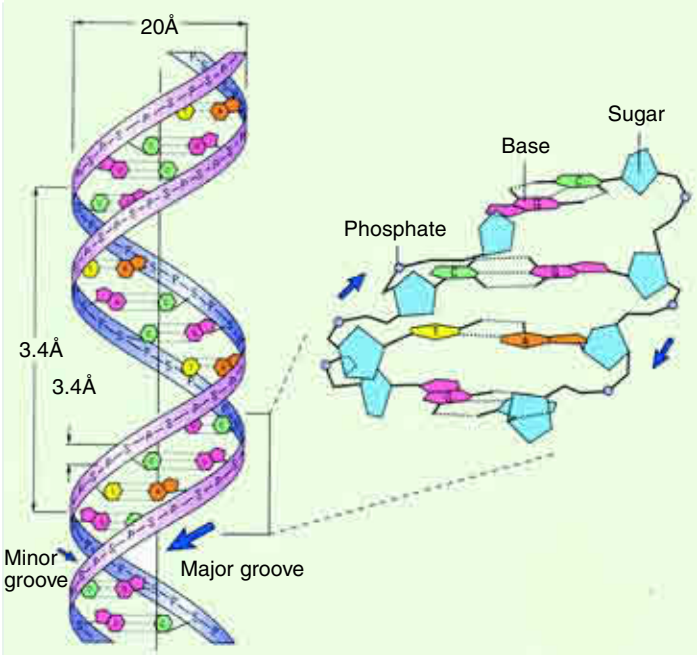


Fig. 15–20. Schematic representation of the Watson-Crick ribbon model of the B-DNA double helix

[P = phosphate; S = deoxyribose sugar ; A = adenine; G = guanine ; T = thymine ; C = cytosine]

The two ribbons represent the phosphate-sugar chains and the horizontal rods represent the bonding between the pair of bases. Two kinds of grooves, a major and a minor, are evident. The horizontal parallel lines symbolize hydrogen bonding between complementary bases. The vertical line represents the fibre axis. Note that the structure repeats at intervals of 34 Å, which corresponds to 10 residues on each chain.

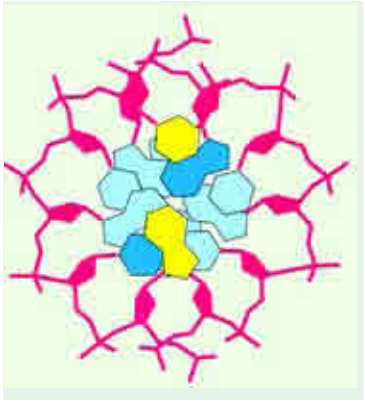


Fig. 15–21. Axidhal view of a one of the strands of a DNA double helix

(as viewed down the helix axis)

Base pairs are stacked nearly one on top of another in the double helix. Note the position of the sugar-phosphate backbone and bases which is outward and inward respectively. The 10-fold symmetry is evident.

(*Adapted from Lubert Stryer, 2000*)

2. The phosphate and deoxyribose units are found on the periphery of the helix, whereas the purine and pyrimidine bases occur in the centre (refer Fig. 15–21). The planes of the bases are perpendicular to the helix axis. The planes of the sugars are almost at right angles to those of the bases.

3. The diameter of the helix is 20 Å. The bases are 3.4 Å apart along the helix axis and are related by a rotation of 36 degrees. Therefore, the helical structure repeats after 10 residues on each chain, *i.e.*, at intervals of 34 Å. In other words, each turn of the helix contains 10 nucleotide residues.

JAMES DEWEY WATSON (born in 1928 at Chicago, USA) &
FRANCIS HARRY COMPTON CRICK (born in 1916 at Northampton, USA)

James D. Watson

Francis H.C. Crick

James Watson was a child prodigy who entered the University of Chicago at the age of 15. He graduated 1in 1947 and obtained his PhD from the University of Indiana in 1950. He was then awarded a fellowship to study in Copenhagen but after a year there he moved to the Cavendish Laboratory, Cambridge, with the specific objective of studying the gene. There he met Francis Crick, who was working on protein structure for his PhD after starting as a physicist but becoming diverted to biology after the War. The famous fusion of minds that led Watson and Crick to deduce that DNA is a double helix has been described many times, notably by Watson in *The Double Helix*. The two shared the 1962 Nobel Prize with Maurice Wilkins. The molecular details of DNA that continue to unfold are reminiscent of a statement once made by the great Albert Einstein (Nobel laureate in Physics 1921) that in science, for every milestone reached there is a signpost pointing to yet another.

After the double helix, Watson and Crick went separate ways. Watson returned to the USA in 1953 and eventually took up a professorship at Harvard before moving in 1976 to the Cold Spring Harbor Laboratory on Long Island, where he became Director. He has worked on RNA synthesis, protein synthesis and the role of viruses in cancer. Recently he has been one of the prime motivators behind the Human Genome Project, the ambitious research programme that has as its goal the complete nucleotide sequence of the human genome (Section 16.3). Crick stayed in Cambridge until moving to the Salk Institute in Southern California in 1977. During the 1950s and 1960s he was the major influence in molecular biology and many of the great advances in understanding genes and gene expression are due directly to him. His autobiography is *What Mad Pursuit,* Weidenfield and Nicolson, London 1989.

Crick named his house at Cambridge the '*Golden Helix*'. He had hung on the wall behind his desk in the Cavendish Lab the following motto : "Reading Rots the Mind."

He often says :

"It is very difficult to explain something scientifically when you cannot explain what it is."

4. The two chains are held together by hydrogen bonds between pairs of bases. Adenine always pairs with thymine by 2 hydrogen bonds and guanine with cytosine with 3 hydrogen bonds. This specific positioning of the bases is called *base complementarity*. The individual hydrogen bonds are weak in nature but, as in the case of proteins, a large number of them involved in the DNA molecule confer stability to it. It is now thought that the stability of the DNA molecule is primarily a consequence of van der Waals forces between the planes of stacked bases.

5. The sequence of bases along a polynucleotide chain is not restricted in any way. *The precise sequence of bases carries the genetic information.*

As a corollary, the entire structure of a DNA molecule resembles a *winding staircase*, with the sugar and phosphate molecules of the nucleotides forming the railings and the linked nitrogen base pairs (A-T and G-C) forming the steps.

Antiparallelity of the polynucleotide chain. The procedure of *nearest neighbour analysis* has shown that the two complementary chains do not run in the same direction with respect to their internucleotide linkages, but rather are *antiparallel*, *i.e.*, one strand runs in the 5′ → 3′ direction and the other in the 3′ → 5′ direction (refer Fig. 15–22). Thus, of the two chains, one ascends and the other descends. They have reverse polarity, that is to say the sequence of the atoms in the two chains is opposite to each other. This is analogous to 2 parellel streets, each running one way but carrying traffic in opposite directions.

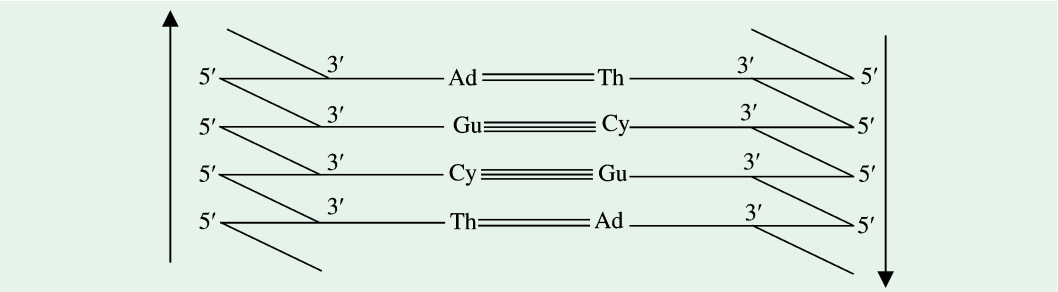


Fig. 15–22. Antiparallelity of the two polynucleotide chains of DNA molecule

Base complementarity of the polynucleotide chain. An important feature of the double helix is the specificity of the pairing of bases. *Pairing always occurs between adenine and thymine and between guanine and cytosine* (Fig. 15-21). This means that a 6-aminopurine will always bond with a 6-ketopyrimidine, and a 6-ketopurine with a 6-aminopyrimidine. Consequently, the two helices of a DNA molecule are complementary to each other and not identical Fig 15-23. Base-pairing is due to steric and hydrogen-bonding factors.

Complementarity is a term introduced into quantum theory by a physicist Niels Bohr (LT, 1885–1962), implying that evidence relating to atomic systems, that has been obtained under different experimental conditions, cannot necessarily be comprehended by one single model. Thus, for example, the wave model of the electron is **complementary** to the particle model.

(A) *Steric factor.* The steric restriction is imposed by the regular helical nature of the sugar-phosphate backbone of each polynucleotide chain. The geometry of the base pair has some special consequences in that the distances between the glycosidic bonds are the same for both the base pairs and also because the bond angle between the glycosidic bond direction and the line joining the C 1 atoms is the same in each pair. With the result, the glycosidic bonds of all nucleotides are arranged in an identical manner in relation to the axis of the helix, despite the differences in the bases.

The glycosidic bonds that are attached to a base pair are always 10.85 Å apart. A purine-pyrimidine base pair fits perfectly

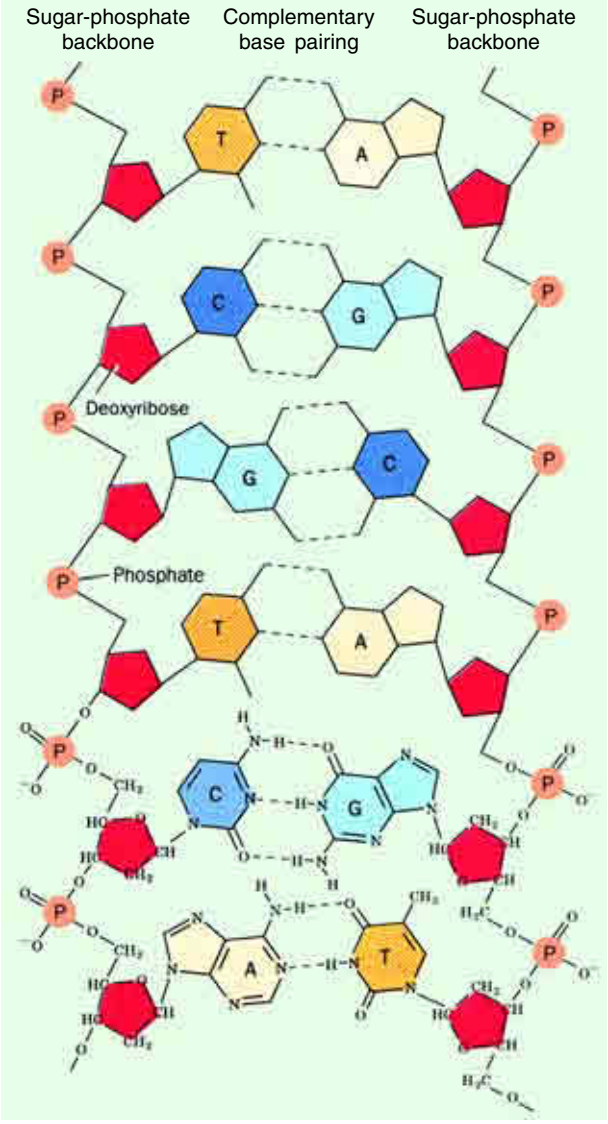


Fig. 15-23. Double-stranded DNA
The two polynucleotide chains associate through complementary hydrogen bonding.

in this space. In contrary to this, there is insufficient space for two purines, whereas there is more than enough space for two pyrimidines so that they would be far apart to form hydrogen bonds. Hence, one member of a base pair in a DNA helix must invariably be a purine and the other a complementary pyrimidine because of steric reasons. Obviously, if the base order in one strand is known, the sequence of bases in the other strand can be predicted. For instance, if one strand has a portion which has arrangement, *adenine*, *guanine*, *thymine*, *cytosine*, the corresponding region in the complementary strand will have *thymine, cytosine, adenine, guanine*.

(B) *Hydrogen-bonding factor*. The base pairing is further restricted by hydrogen-bonding requirements. The hydrogen atoms in purine and pyrimidine bases have well-defined positions. Adenine cannot pair with cytosine because there would be two hydrogen atoms near one of the bonding positions and none at the other. Similarly, guanine cannot pair with thymine. In contrast, *adenine forms 2 hydrogen bonds with thymine whereas guanine forms 3 with cytosine* (refer Fig. 15–24). Thus, *the G-C bond is stronger by 50% than the A-T bond*. The higher the G-C content of a DNA moelcule, the greater is its buoyant density.

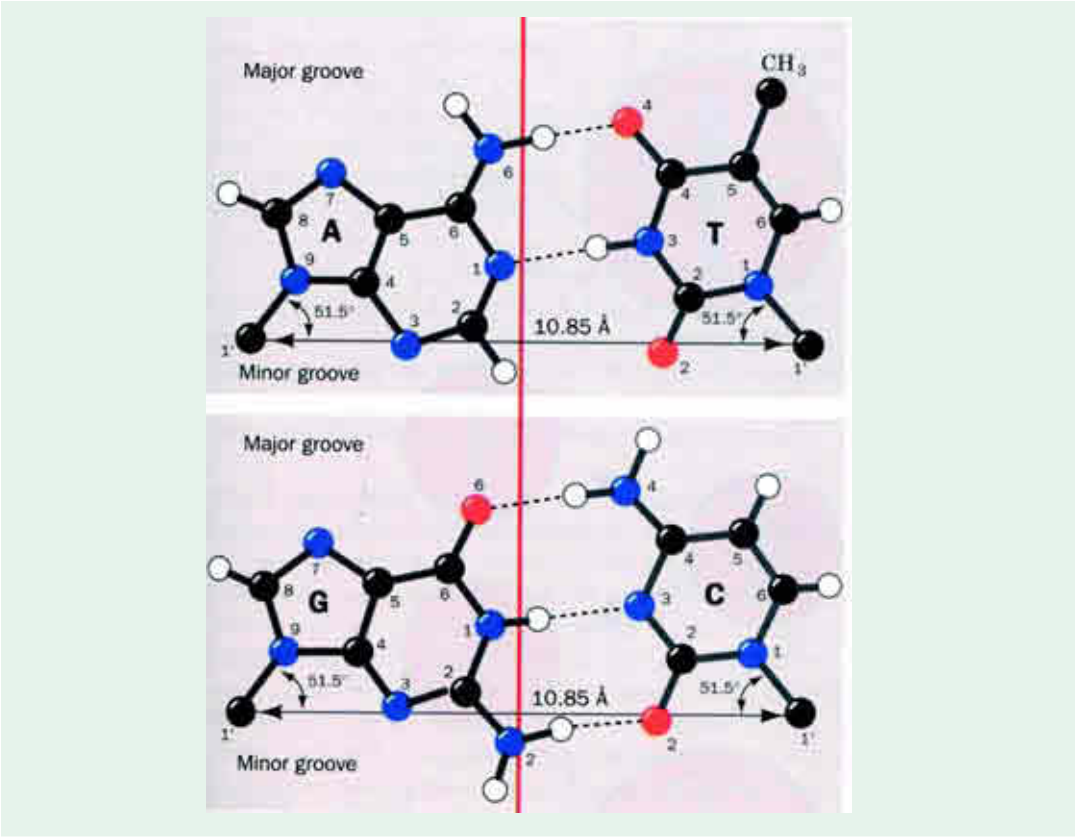


Fig. 15–24. Structure of the two Watson Crick base pairs

The line joining the C1 atoms is the same length in both base pairs and makes equal angles with the glycosidic bonds to the bases. This gives DNA a series of pseudo-twofold symmetry axes (often referred to as **dyad axes)** that pass through the centre of each base pair (*red line*) and are perpendicular to the helix axis. Note that A. T base pairs associate *via* two hydrogen bonds, whereas C. G base pairs are joined by three hydrogen bonds.

(*After Arnott, S., Dover, S.D. and Wonacott, A.J., 1969*)

Denaturation and Renaturation of DNA Helix

Denaturation of DNA is a loss of biologic activity and is accompanied by cleavage of hydrogen bonds holding the complementary sequences of nucleotides together. This results in a separation of the double helix into the two constituent polynucleotide chains. In it, the firm, helical, two-stranded native structure of DNA is converted to a flexible, single-stranded `denatured' state. The splitting of DNA molecule into its two strands or chains, during denaturation, is obvious because of the fact that the hydrogen bonds holding the bases are weaker than the bonds holding these bases to the sugar-phosphate groups. The transition from native to a denatured form is usually very abrupt and is accelerated by reagents such as urea and formamide, which enhance the aqueous solubility of the purine and pyrimidine groups. Denaturation involves the following changes :

1. *Increase in absorption of ultraviolet light* (= *Hyperchromic effect*). As a result of resonance, all of the bases in nucleic acids absorb ultraviolet light. And all nucleic acids are characterized by a maximum absorption of UV light at wavelengths near 260 nm. When the native DNA (which has base pairs stacked similar to a stack of coins) is denatured, there occurs a marked increase in optical absorbancy of UV light by pyrimidine and purine bases, an effect called **hyperchromicity** or **hyperchromism** whch is due to unstacking of the base pairs. This change reflects a decrease in hydrogen-bonding. Hyperchromicity is observed not only with DNA but with other nucleic acids and with many synthetic polynucleotides which also possess a hydrogen-bonded structure.

2. *Decrease in specific optical rotation.* Native DNA exhibits a strong positive rotation which is highly decreased upon denaturation. This change is analogous to the change in rotation observed when the proteins are denatured.

3. *Decrease in viscosity*. The solutions of native DNA possess a high viscosity because of the relatively rigid double helical structure and long, rodlike character of DNA. Disruption of the hydrogen bonds causes a marked decrease in viscosity.

Effect of pH on denaturation. Denaturation of DNA helix also occurs at acidic and alkaline pH values at which ionic changes of the substituents on the purine and pyrimidine bases can occur. In acid solutions near pH 2 to 3, at which amino groups bind protons, the DNA helix is disrupted. Similarly, in alkaline solutions near pH 12, the enolic hydroxyl groups ionize, thus preventing the keto-amino group hydrogen bonding.

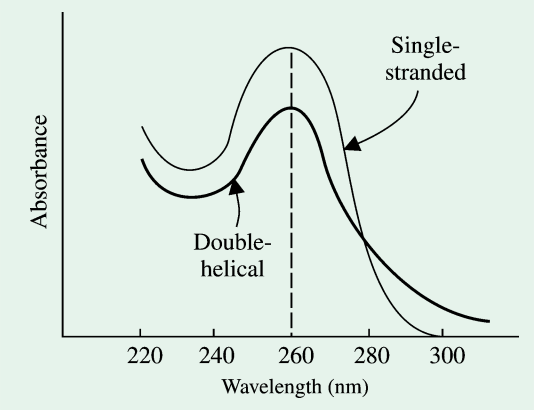


Fig. 15–25. The absorbance spectra of a DNA solution at 260 nm and pH 7.0

When mixtures of mucleotides are present, the wavelength at 260 nm (dashed vertical line is used for measurement.

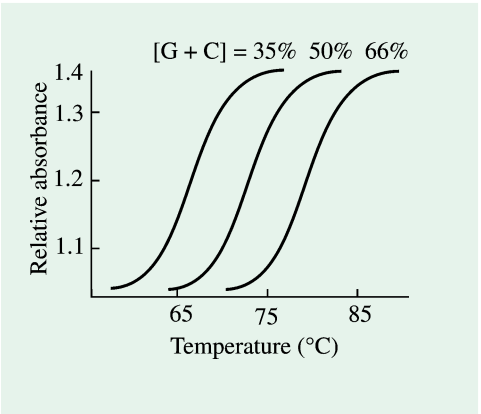


Fig. 15–26. DNA melting curves

The absorbance relative to that at 25°C is plotted against temperature. The wavelength of the incident light was 260 nm. The Tm is about 72°C for *Escherichia coli* DNA (50% G-C pairs) and 79°C for the bacterium, *Pseudomonas aeruginosa* DNA (66% G-C pairs).

Effect of temperature on denaturation. The DNA double helix, although stabilized by hydrogen bonding, can be denatured by heat by adding acid or alkali to ionize its bases. The unwinding of the double helix is called *melting* because it occurs abruptly at a certain characteristic temperature called *denaturaiton temperature* or *melting temperature* (T_m). The melting temperature is defined as the temperature at which half the helical structure is lost. The abruptness of the transition indicates that *the DNA double helix is highly cooperative structure*, held together by many reinforcing bonds ; it is stabilized by the stacking of bases as well as by pairing. The melting of DNA is readily monitored by measuring its absorbance of light at wavelength near 260 nm. As already stated, the unstacking of the base pairs results in increased absorbance, a phenomenon termed hyperchromicity (Fig. 15–25). T_m is analogous to the melting point of a crystal. It can be lowered by the addition of urea which disrupts hydrogen bonds. In 8M urea, T_m is decreased by nearly 20°C. DNA can be completely denatured (*i.e.*, separated into a single-stranded structure) by 95% formamide at room temperature only.

Since the G-C base pair has 3 hydrogen bonds as compared to 2 for A-T, it follows that DNAs with high concentrations of G and C might be more stable and have a higher T_m than those with high concentrations of A and T (Fig. 15–26). In other words, the DNA molecules containing less G-C bond denature first as G-C bond has higher thermal stability. In fact, the Tm of DNA from many species varies linearly with G-C content, rising from 77 to 100°C as the fraction of G-C pairs increases from 20% to 78%. This measurement may, hence forth, be used as an index of heterogeneity of nucleic acid molecules. For instance, viral DNA has a much sharper thermal transition than that prepared from animal sources.

Complete rupture of the two-stranded helix by heating is not a readily reversible process. However, if a solution of denatured DNA, prepared by heating, is cooled slowly to room temperature, some amount of DNA is renatured. Maximum reversibility (50-60%) is usually attained by *annealing* (= slow cooling) the denatured DNA, *i.e.*, holding the solution at a temperature about 25°C below T_m and above a concentration of 0.4 M Na^+ for several hours. Fast cooling will not reverse denaturation, but if the cooled solution is again heated and then cooled slowly, renaturation takes place. The restoration occurs because the complementary bases reunite by hydrogen bonds and the double helix again forms.

Renaturation of complementary single strands to produce fully double-stranded moelcule requires 2 separate reactions (Fig. 15–27) :

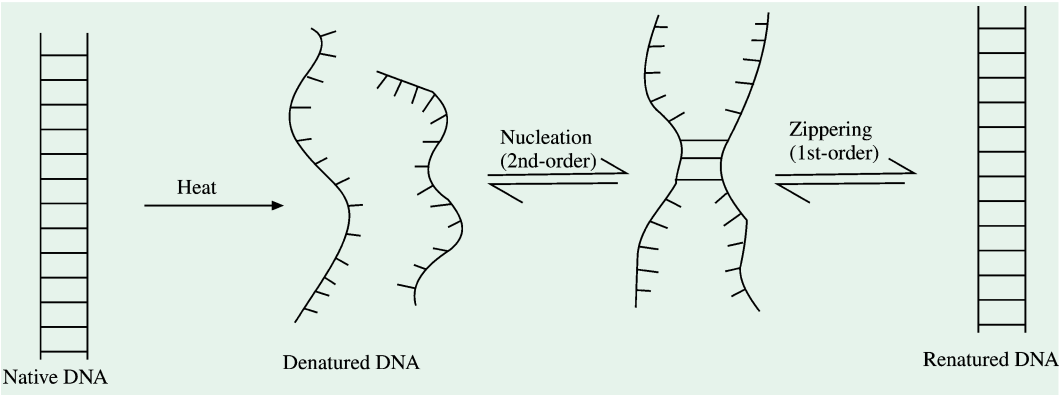


Fig. 15–27. Steps in the denaturation and renaturation of DNA fragments

I. **Nucleation reaction.** In this hydrogen bonds form between two complementary single strands ; this is a bimolecular, second-order reaction.

II. **Zippering reaction.** In this hydrogen bonds form between all the bases in the complementary strands ; this is a unimolecular, first-order reaction.

Molecular Weight of DNA

DNA molecules are among the largest known. It is difficult to isolate DNA without fragmentation. The entire DNA of bacteria such as *Escherichia coli* has a molecular weight of 2.6 $\times 10^9$. Viral DNAs range in size from about 1 to 350 $\times 10^6$. The residue weight of a single nucleotide is 300 to 350. Thus, there are about 3,000 nucleotides per million molecular weight of DNA.

Length of DNA Molecule

The amount of DNA is usually measured by the microunit of the weight known as the **picogram, pg** (1 pg = 10^{-12} g). The amount of DNA has been found to be constant from cell to cell and species. *A piece of double helical DNA, 31 cm in length, weighs one picogram.* So by taking the weight of DNA amount in the nucleus, the length of the DNA molecule can be calculated easily. By employing this method, Dupraw and Bahr (1968) have calculated that the human diploid cells contain 5.6 pg DNA amount which is equivalent to (5.6 × 31) or 173.6 cm-long DNA molecule. Similarly, they also calculated that a diploid cell of *Trillium* has dsDNA of 37.2 m (or 3,720 cm) length, which is equivalent to (3,720/31) or 120 pg DNA by weight. Also, the polytenic chromosomes of *Drosophila*, with 293 pg DNA contents, have (293 × 31) or 90.83 m-long DNA molecule.

Shape and Size of DNA Molecules

The DNA molecules are characteristically highly **elongate** structures. The *Escherichia coli* chromosome is a single molecule of double helical DNA, containing about 4 million base pairs. This DNA is highly asymmetric. Its contour length is 14 $\times 10^6$ Å and its diameter is 20 Å. Whereas its contour length (1.4 mm) corresponds to a macroscopic dimension, its width (20 Å) is on the atomic scale. However, the largest chromosome of *Drosophila melanogaster* contains a single DNA molecule with 6.2 $\times 10^7$ base pairs. These have a contour length of 2.1 cm. Such highly asymmetric DNA molecules are very susceptible to fragmentation by shearing forces.

Table 15–5 lists the dimensions of the various viral, bacterial and eukaryotic DNA molecules. A perusal of the table indicates that *even the smallest DNA molecules are highly elongate*. For instance, the DNA from polyoma virus contains 5,100 base pairs and has a contour length of 1.7 µm (or 17,000 Å). Furthermore, tropocollagen, the basic structural unit of collagen, has the shape of a rod which is 3,000 Å in length and 15 Å in diameter. *Tropocollagen* is the longest known protein.

Table 15–5. Dimensions of certain DNA molecules

Organism	*Number of base pairs (in thousands or kb)*	*Length* (in µm)*	*Molecular weight*
Viruses			
Polyoma or SV40	5.1	1.7	3.1×10^6
λ phage	48.6	17	31×10^6
T 2 phage	166	56	122×10^6
Vaccinia	190	65	157×10^6
Bacteria			
Mycoplasma	760	260	504×10^6
Escherichia coli	4,000	1,360	$2{,}320 \times 10^6$
Eukaryotes			
Yeast	13,500	4,600	—
Drosophila	1,65,000	56,000	—
Human	2,900,000	9,90,000	—

* 1 µm of double helix = 2.94×10^3 base pairs = 1.94×10^6 daltons.

(*Modified from A. Kornberg and T.A. Baker, 1992*)

The existence of a **circular** form of DNA was first demonstrated by John Cairns (1963) in *Escherichia coli* which was grown in the presence of tritium-labelled thymidine. The chromosomal material was then visualized by radioautography. This revealed a tangled circle that, when fully extended, had a circumference of 1,100 to 1,400 μ (refer Fig. 15–28). In fact, the finding that *E. coli* has a circular chromosome was anticipated by genetic studies that revealed that the gene-linkage map of this bacterium is circular. *The term 'circular' refers to the continuity of the DNA chain, and not to its geometric form.*

DNA molecule from the T 7 bacteriophage is, however, **linear** in shape. Interestingly, the DNA molecules of some viruses such as λ bacteriophage interconvert between linear and circular forms. The linear form is present inside the virus particle whereas the circular form is present in the host cell.

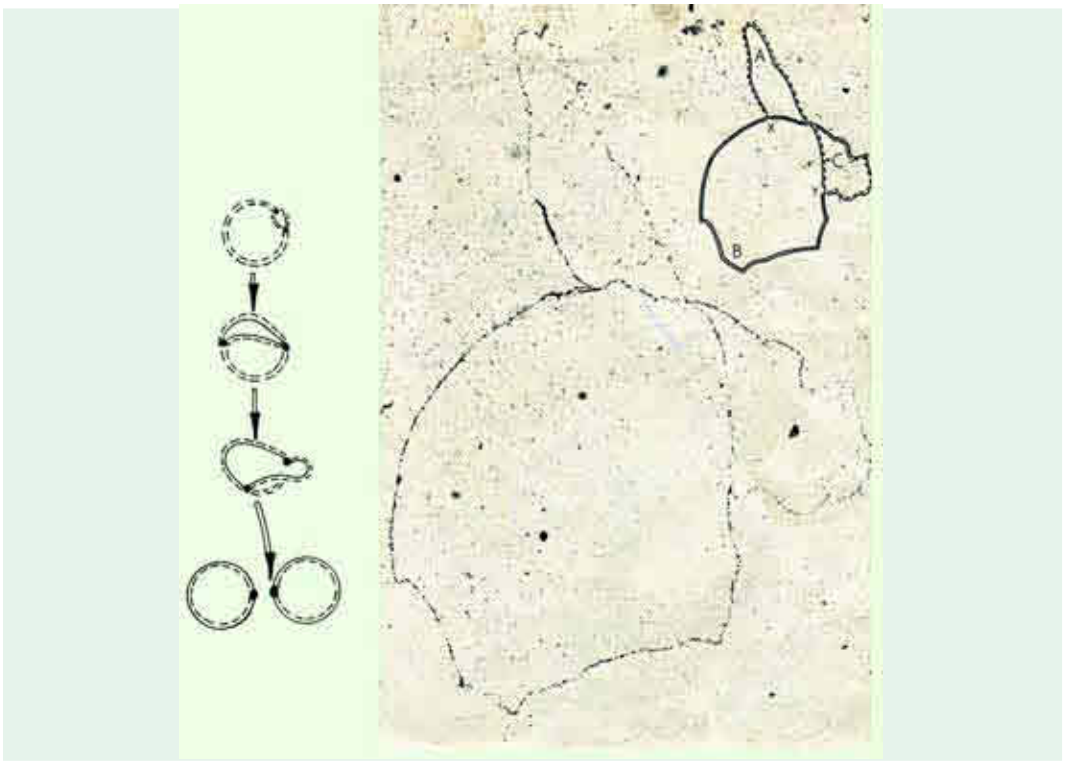


Fig. 15–28. Autoradiograph of the chromosome of *Escherihia coli*

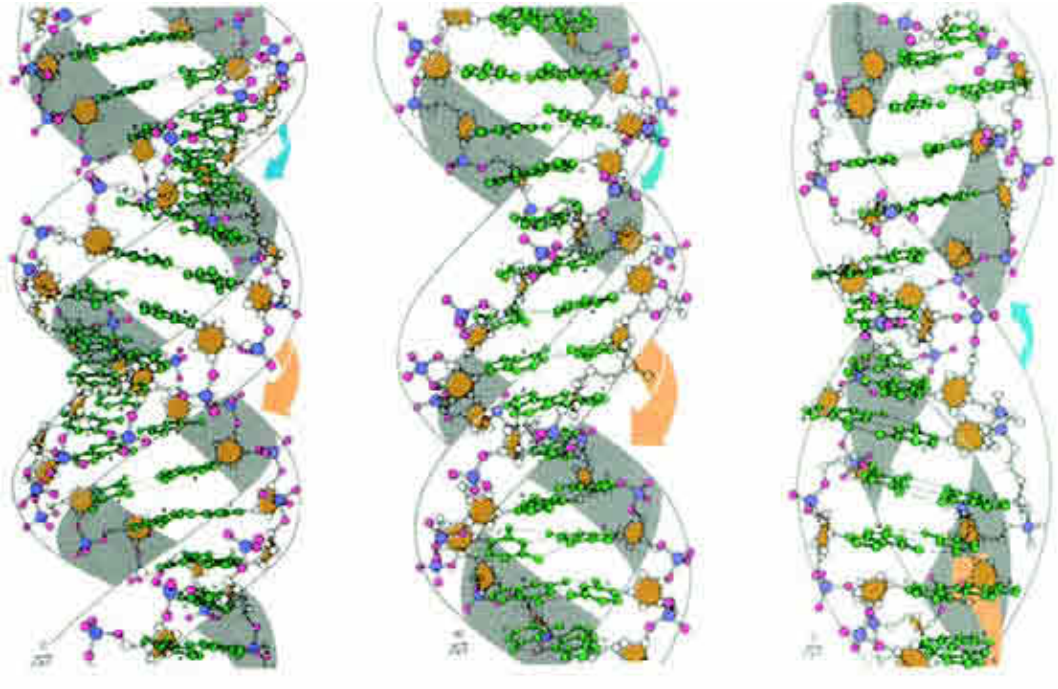
Right : *E. coli DNA in the process of replication*. The inset is a schematic representation of the autoradiogram. The dashed line in regions A and C represents one parent strand, one-fourth tritiated (solid portion) from the previous replication. The other parent strand and the new daughter strands (solid lines) are fully tritiated. X and Y are the "replication forks."

Left : The replication scheme suggested by autoradiography. Here the addition of tritium is synchronized with the beginning of replication.

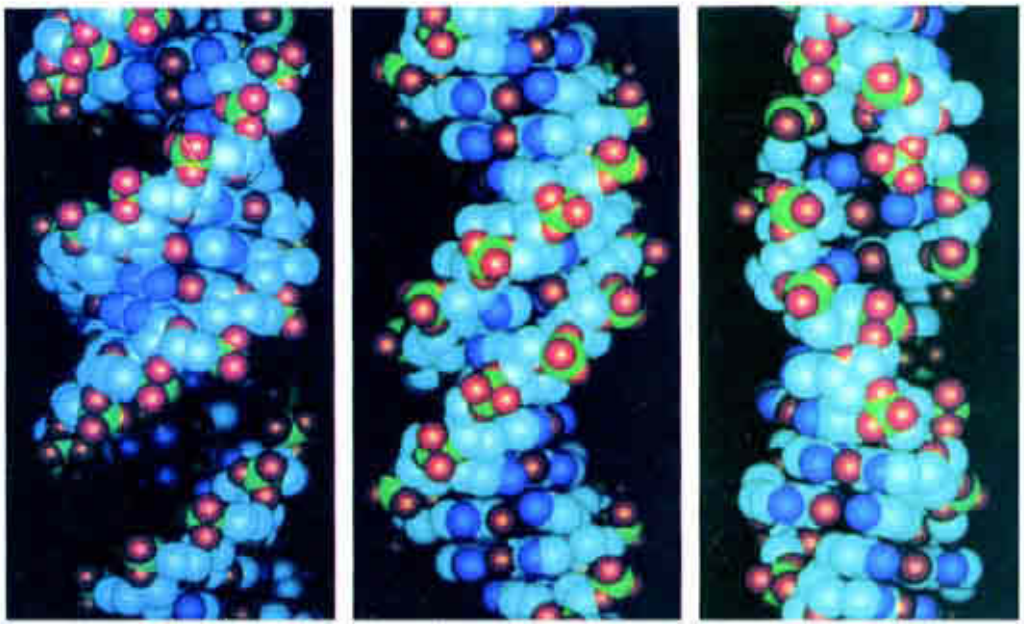
[*Courtesy of John Cairns, from Cold Spring Harbor Symp. Quant. Biol.,* **28 :** *43 (1963).*]

Variants of Double Helical DNA

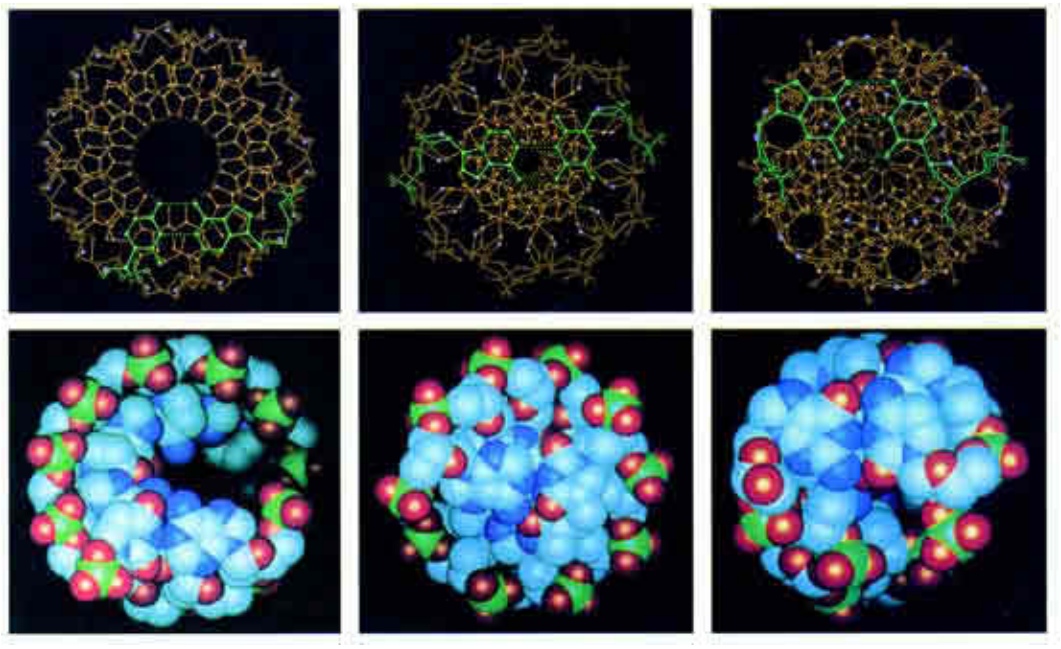
Under physiologic conditions, most of the DNA in a bacterial or eukaryotic genome is in the classic Watson-Crick form which is called the **B-form DNA** or simply **B-DNA**. The B-form is the most stable structure for a random-sequence DNA molecule, and is therefore the standard point of reference in any study of the properties of DNA. X-ray analyses of DNA crystals at atomic resolution have revealed that DNA exhibits much more structural diversity than formerly envisaged. Two such variations are :



A-DNA B-DNA Z-DNA



Part (a)



Part (b)

Fig. 15-29. A comparison of the structures of 3 important forms of DNA– A-DNA, B-DNA and Z-DNA
Part (a). Views perpendicular to the helix axis (*i.e.*, axial or side views) of all the 3 forms.
Part (B). Views down the helix axis (*i.e.*, radial or top views) of all the 3 forms.

The top half of both parts shows ball-and-stick drawings; the bottom half of both parts include computer-generated space-filling models. The H atoms have been omitted for the sake of clarity in both drawings. The highly detailed structures shown here were deduced by X-ray crystallography performed on short segments of DNA, rather than the less detailed structures obtained from DNA wet fibres. The diffraction pattern obtained from the crystallization of short segments of DNA provides much greater details concerning the exact placement of atoms within a double helical structure. Alexander Rich at MIT, Richard Dickerson then at the California institute of Technology, and their colleagues were the first reseachers to crystallize a short piece of DNA.

In the **A-DNA**, the based pairs have a marked propeller twist with respect to the helix axis. Note that the base pairs are inclined to the helix axis and that the helix has a hollow core. In the **B-DNA**, the base pairs lie in a plane that is close to perpendicular to the helix axis. Note that the helix axis passes through the base pairs so that the helix has a solid core. In the **Z-DNA ,** the helix is left-handed and in this respect differs from A-DNA and B-DNA, both of which are right-handed. Note that in Z-DNA, the sugar, phosphate chains follow a zigzag course (alternate ribose residues lie at different radii in part *b*) indicating that Z-DNA's repeating motif is a dinucleotide. Under physiologic conditions, all DNA molecules are in B form.

(*a*) A DNA chain can be rotated about 6 bonds in each monomer—the glycosidic bond between the base and sugar, the $C_{4'}$ — $C_{5'}$, bond of the sugar and four bonds in the phosphodiester bridge joining $C_{3'}$ of one sugar and $C_{5'}$ of the next one — compared with just two for polypeptide chains.

(*b*) The DNA helix can be smoothly bent into an arc or it can be supercoiled with rather little change in local structure. This ease of deformation enables circular DNA to be formed and allows DNA to be wrapped around proteins. Deformability of DNA also allows compacting of DNA into a much smaller volume. *DNA can also be kinked, i.e.*, bent at discrete sites. Kinking can be induced by specific base sequences, such as a segment of at least 4 adenine residues or by the binding of a protein.

It may, thus, be inferred that "DNA is a remarkably flexible molecule. Considerable rotation is possible around a number of bonds in the sugar-phosphate backbone and thermal fluctuations can produce bending, stretching and unpairing (melting) in the structure." (Lehninger, Nelson and Cox, 1993).

However, these structural variations do not affect the key properties of the DNA molecule as defined by Watson and Crick : strand complementarity, antiparallel nature of strands and base pairing specificity of A combining with T (A = T) and G combining with C (G ≡ C).

The DNA can, in fact, adopt several 3-'D' conformations and *can change reversibly from one form to the other*. These alternative forms of DNA differ in features like nature of the helix, the number of residues per turn ('n'), the spacing of the residues along the helix axis ('h') etc. Some significant alternative forms of DNA are described below (Fig. 15–29) :

A-DNA : A-DNA appears when the DNA fibre (B-DNA) is dehydrated, *i.e.*, relative humidity is reduced from 92 to 75% and Na^+, K^+ and Cs^+ ions are present in the medium. In other words, *in solution, DNA assumes the B form and under conditions of dehydration, the A form*. This is because the phosphate groups in the A-DNA bind fewer water molecules than do phosphates in B-DNA.

Like B-DNA, A-DNA is a right-handed double helix (Fig. 15–28) made of antiparallel strands held together by Watson-Crick base pairing. But the vertical rise per base pair (*bp*) is 2.3 Å and the number of base pairs per helical turn is 11, relative to the 3.4 Å rise and 10.4 base pairs per turn found in B-DNA. As such, the helix in A-DNA moves a total distance of (2.3 × 11) or 25.30 Å per turn which value is called the *helix pitch* (In B-DNA, the value of helix pitch is (3.4 × 10.4) or 35.36 Å. Obviously, the rotation per base pair in A-DNA is 360/11 or 32.72° while in B-DNA the value of rotation is 360/10.4 or 34.61° (Recall that during one turn, the helix rotates by an angle of 360°). The diameter of A-DNA helix is 25.5 Å whereas that of B-DNA helix is 23.7 Å. The helix of A-DNA is, therefore, wider and shorter than that of B-DNA.

The base pairs are tilted rather than normal to the helix axis. As far as puckering (wrinkles or small folds) of the ribose units is concerned, in A-DNA, $C_{3'}$ is out of the plane (called **C3′-endo**) formed by the other 4 atoms of the furanose ring of sugar. By contrast, in B-DNA, $C_{2'}$ is out of plane (called **C2′-endo**). In A-DNA, the base pairs are inclined at 19° with the axis of the helix but in B-DNA, the base pairs lie almost perpendicular (1° tilt) to the helix axis. Moreover, the minor groove is practically nonexistent in A-DNA, whereas it is quite deep in B-DNA.

C-DNA : C-DNA is formed at 66% relative humidity in the presence of Li^+ ions. This form of DNA is also right-handed, with an axial rise of 3.32 Å per base pair. There are 9.33 base pair per turn of the helix ; the value of helix pitch is, therefore, 3.32 × 9.33 Å or 30.97 Å. The rotation per base pair in C-DNA is 360/9.33 or 38.58°. The C-helix has a diameter of 19 Å, smaller than that of both B- and A-helix. The tilt of the base pairs is 7.8°.

D-DNA : D-DNA is an extremely rare variant with only 8 base pairs per helical turn. This form of DNA is found in some DNA molecules devoid of guanine. By contrast, *A-, B- and C-forms of DNA are found in all DNA molecules, irrespective of their base sequence.* There is an axial rise of 3.03 Å per base pair, with a tilting of 16.7° from the axis of the helix.

Z-DNA : Z-DNA is the more radical departure from B-DNA and is characterized by a *left-handed helical rotation*. It was discovered by Rich, Nordheim and Wang in 1984. They found that a hexanucleotide, CGCGCG, forms a duplex of antiparallel strands held together by Watson-Crick base pairing, as expected. Surprisingly, they found that this double helix was left-handed and the phosphates in the DNA backbone were in a *zigzag* manner ; hence, they termed this new form as Z-DNA. By contrast, other forms of DNA are right-handed, with a regular phosphate backbone. Another remarkable characteristic of Z-DNA is that in it the adjacent sugar residues have *alternating* orientation and it is because of this reason that in Z-DNA, the repeating unit is

a *dinucleotide* as against the B-DNA, where the adjacent sugar residues have same orientation so that the repeating unit in B-DNA is a mononucleotide (Fig. 15–30). Unlike A- and B-DNAs, Z-DNA contains only one deep helical groove. There are 12 base pairs (or six repeating dinucleotide units) per helical turn, with an axial rise of 3.8 Å per base pair ; the bases are inclined at 9° with the axis of the helix. Because 12 base pairs are accomodated in one helix in Z-DNA, as against 10.4 in B-DNA, the angle of twist per repeating unit (dinucleotide) is (360/12 × 2) or 60° as against 34.61° per nucleotide in B-DNA. One complete helix is 45.60 Å in length in contrast to 35.36 Å in B-DNA. Since the bases get more length to spread out in Z-DNA and since the angle of tilt is 60°, they are more closer to the axis ; hence, the diameter of Z-DNA molecule is 18.4 Å hereas it is 23.7 Å in B-DNA.

The Z-DNA form occurs in short oligonucleotides that have sequences of alternating pyrimidine and purine bases. Methylation of C-5 of cytosyl residues in alternating CG sequences (as for example in the sequence CGCGCG) facilitates the transition of B-DNA to Z-DNA, because the added hydrophobic methyl groups stabilize the Z-DNA structure. The alternance of purines and pyrimidines permits at the glycosidic bonds an alternance of conformations, *syn* (base and sugar are near one another on the same side) and *anti* (base and sugar are distant, because they are on opposite sides). While in A- and B-DNA, all glycosidic bonds are *anti*, in Z-DNA, the purines are *syn* and the pyrimidines are *anti*.

Fig. 15–30. Different orientation patterns of adjacent sugar residues in B-DNA and Z-DNA

Note that in B-DNA, the orientaion of adjacent sugar residues is same, while it is opposite in Z-DNA. The result is that in B-DNA, the repeating unit is a mononucleotide, while in Z-DNA, the repeating unit is a dinucleotide.

General characteristics of various forms of DNA are summarized in Table 15–6.

Table 15–6. Comparison of different forms of DNA

Characteristics	*A-DNA*	*B-DNA*	*C-DNA*	*Z-DNA*
Conditions	75% relative humidity ; Na^+, K^+, Cs^+ ions	92% relative humidity ; Low ion strength	60% relative humidity ; Li^+ ions	Very high salt concentration
Shape	Broadest	Intermediate	Narrow	Narrowest
Helix sense	Right-handed	Right-handed	Right-handed	Left-handed
Helix diameter	25.5 Å	23.7 Å	19.0 Å	18.4 Å
Rise per base pair (‘h’)	2.3 Å	3.4 Ã	3.32 Å	3.8 Å
Base pairs per turn of helix (‘n’)	11	10.4	9.33	12 (= 6 dimers)
Helix pitch (h × n)	25.30 Å	35.36 Å	30.97 Å	45.60 Å

Rotation per base pair	+ 32.72°	+ 34.61°	+ 38.58°	–60° (per dimer)
Base pair tilt	19°	1°	7.8°	9°
Glycosidic bond	*anti*	*anti*	—	*anti* for C, T *syn* for A, G
Major groove	Narrow and very deep	Wide and quite deep	—	Flat
Minor groove	Very broad and shallow	Narrow and quite deep	—	Very narrow and deep

DNAs with Unusual Structures

A number of other sequence-dependent structural variations in DNA have been detected that may serve locally important functions in DNA metabolism. These are :

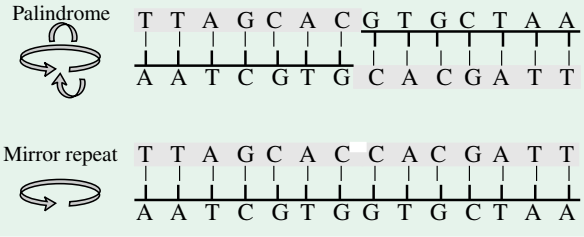


Fig. 15–31. Palindromic DNAs and mirror repeats

A. Palindromic DNA : It has sequences with twofold symmetry. In order to superimpose one repeat (shaded sequence) on the other (unshaded sequence), it must be rotated 180° around the horizontal axis and then again about the vertical axis, as shown by arrows.

B. Mirror repeat : It has a symmetric sequence on each strand. Superimposing one repeat on the other requires only a single 180º rotation about the vertical axis.

1. Bent DNA. Some sequences cause bends in the DNA helix. Bends are produced whenever 4 or more adenine residues appear sequentially in one of the two strands. Six adenines in a row produce a bend of about 18°. Bending may be important in the binding of some proteins to DNA.

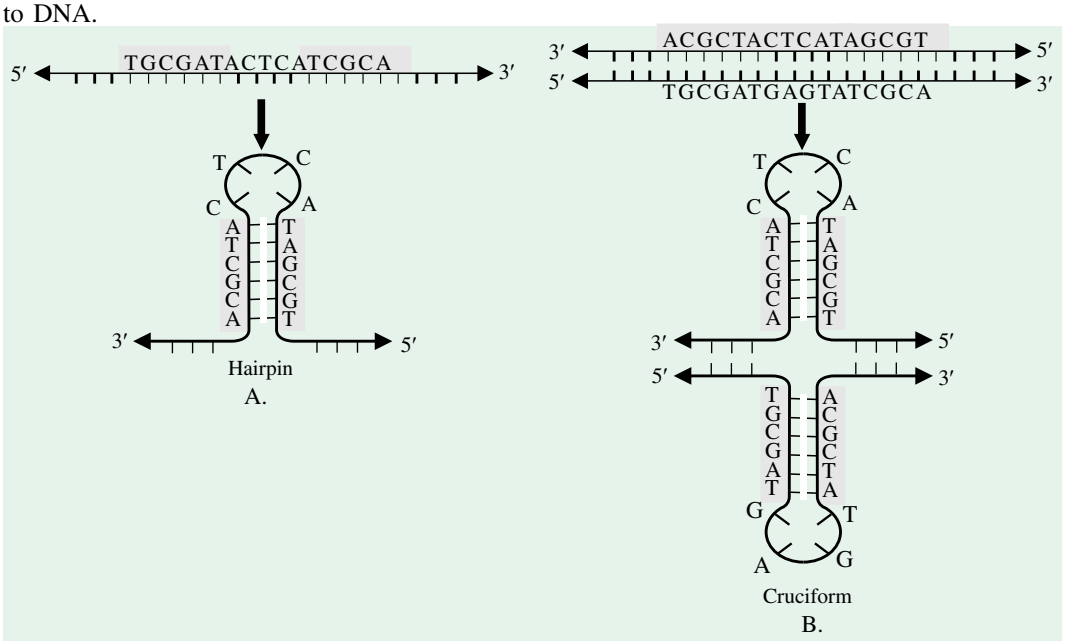


Fig. 15–32. Hairpins and cruciforms

Bases (enclosed in rectangle) are the asymmetric sequences that can pair alternatively with a complementary sequence in the same or opposite strand.

A. Hairpin : When only a single strand of palindromic DNA (or RNA) is involved, a hairpin is formed.

B. Cruciform : When both the strands of a double helical DNA are involved, a cruciform is formed.

2. Palindromic DNA. The term palindromic DNA is applied to regions of DNA in which there are *inverted repetitions* of base sequence with twofold symmetry occuring over two strands (Fig. 15–31A). Such sequences are self-complementary within each of the strands and therefore have the potential to form hairpin (when only a single strand is involved) or cruciform (when both strands of a duplex DNA are involved) structures, involving intrastrand base pairing (Fig. 15–32).

A **palindrome** (*palindromos*^{G} = running back again) is a name, word, phrase, sentence, or verse that reads the same in either direction; for examples names such as NITIN, words such as LEVEL and ROTATOR, and sentences such as SLAP NO GAG ON PALS and ABLE WAS I ERE I SAW ELBA. A spectacular, but less pure, example that ignores spaces and punctuation has been devised by Alastair Reid : "T. Eliot, top bard, notes putrid tang emanating, is sad. I'd assign it a name: gnat dirt upset on drab pot toilet." [Quoted by Brendan Gill in *Here at the New Yorker*, Random House (1975).]

When the inverted sequence occurs within each individual strand of the DNA, the sequence is called a **mirror repeat** (Fig. 15–31B). Mirror repeats do not have complementary sequences within the same strand and cannot form hairpin or cruciform structures. Sequences of these types are found in virtually every large DNA molecule and can involve a few or up to thousands of base pairs.

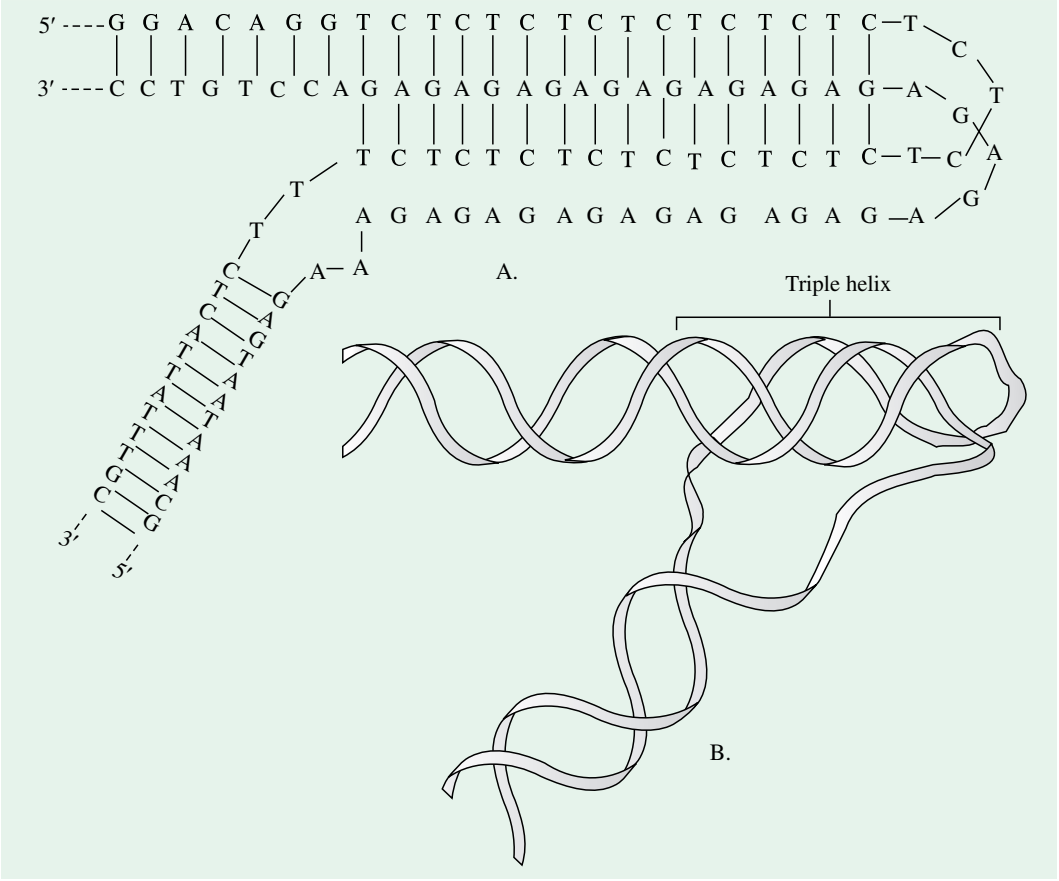


Fig. 15–33. H-DNA

A. A sequence of alternating T and C residues can be considered a mirror repeat centred about one of the central T or C residues.

B. The sequences of alternating T and C residues form an unusual structure in which the strands in one half of the mirror repeat are separated and the pyrimidine-containing strand folds back on the other half of the repeat to form a triple helix.

3. H-DNA. H-DNA is usually found in polypyrimidine or polypurine segments that contain within themselves a mirror repeat. One simple example is a long stretch of alternating T and C residues (Fig. 15–33). A striking feature of H-DNA is the pairing and interwinding of 3 strands of DNA to form a triple helix. Triple-helical DNA is produced spontaneously only within long sequences containing only pyrimidines (or purines) in one strand. Two of the three strands in the H-DNA triple helix contain pyrimidines and the third contains purines.

Single-stranded DNA

The DNA molecules are not always a double-helical structure. Robert Sinsheimer, in 1959, discovered that the DNA in ϕ X174, a small virus that infects Escherichia coli, is single-stranded. His finding was based on the following evidences :

1. The base ratios (A/T and G/C) of ϕ X174 DNA do not approach unity.
2. The ϕ X174 DNA behaves as a randomly coiled polymer whereas the DNA double helix behaves as a rigid rod.
3. The amino groups of the bases of ϕ X174 DNA react readily with formaldehyde. In contrast, the bases in the double-helical DNA do not react with this reagent.

The discovery of this single-stranded DNA raised doubts as to whether the semiconservative scheme of replication as proposed by Watson and Crick, is universally true. However, it was soon found that ϕ X174 DNA is single-stranded for only a part of its life span and the infected *E. coli* cells contain a double helical form of ϕ X174 DNA. This double helical DNA is called *replicative form*, since it acts as a template for the synthesis of DNA of the progeny virus. Viruses that contain single-stranded RNA also replicate *via* a double-stranded replicative form.

Table 15–7. Differences between double-stranded and single-stranded DNAs

dsDNA	*ssDNA*
1. It is linear or filamentous types.	1. It is somewhat star-shaped (stellate).
2. DNA double helix behaves as a rigid rod.	2. ϕ X 174 DNA behaves as a randomly coiled polymer.
3. A/T ratio approaches unity.	3. A/T ratio is 0.77.
4. G/C ratio also approaches unity.	4. G/C ratio is 1.3.
5. It is resistance to the action of formaldehyde.	5. The amino groups of the bases of ϕ X 174 DNA react readily with formaldehyde.
6. Ultraviolet absorption increases from 0 to 80°C	6. Ultraviolet absorption increases from 20 to 90°C.

RIBONUCLEIC ACID

Ribonucleic acid (RNA), like DNA, is a long, unbranched macromolecule consisting of nucleotides joined by 3′ → 5′ phosphodiester bonds (refer Fig. 15–34). The number of ribonucleotides in RNA ranges from as few as 75 to many thousands.

Differences with DNA

Although sharing many features with DNA, the RNA molecules possess several sepcific differences :

1. As apparent from its name, the sugar moiety in RNA, to which the phosphate and the nitrogen bases are attached, is *ribose* rather than 2′- deoxyribose of DNA. Ribose contains a 2′-hydroxyl group not present in deoxyribose.

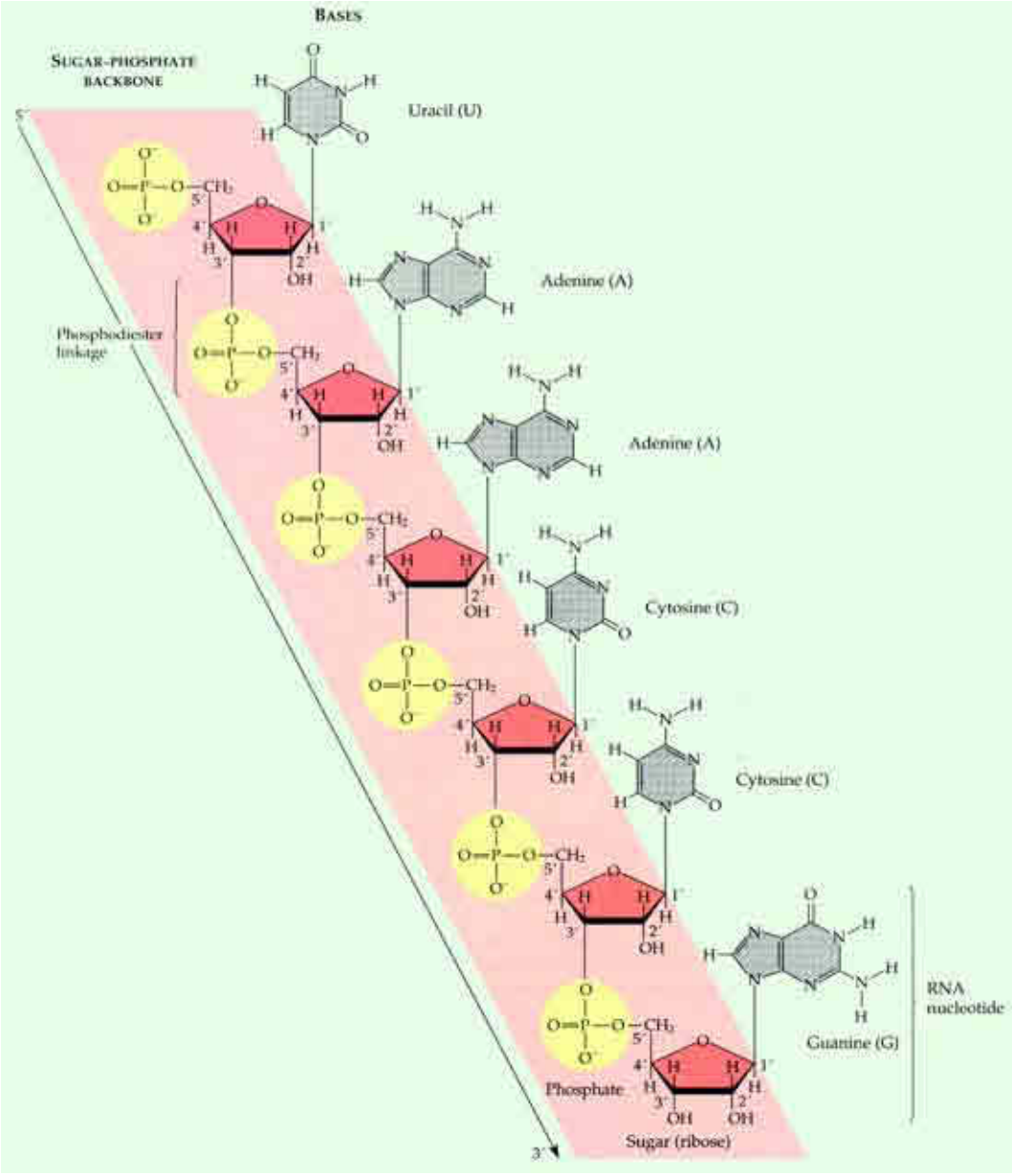


Fig. 15–34. A strand of RNA

This structure is very similar to a DNA strand (see Figure 9.10), except that the sugar is ribose instead of deoxyribose and uracil is substituted for thymine.

2. RNA contains the pyrimidine *uracil* (U) in place of thymine which is characteristic of DNA molecule. Uracil, like thymine, can form a base pair with adenine by 2 hydrogen bonds. However, it lacks the methyl group present in thymine. It may be noted that in one case, however, RNA possesses thymine.

3. The native RNA is *single-stranded* rather than a double-stranded helical structure characteristic of DNA. However, given the complementary base sequence with opposite polarity, the single strand of RNA may fold back on itself like a hairpin and thus acquire the double-stranded pattern (refer Fig. 15–35). In the region of hairpin loops, A pairs with U and G pairs with C. Guanine can also form a base pair with uracil but it is less strong than the G-C base pair. The base pairing in RNA hairpins is frequently imperfect. Some of the apposing bases may not be complementary and one or more bases along a single strand may be looped out to facilitate the pairing of the others. The proportion of helical regions in various types of RNA varies over a wide range, although a value of 50% is typical.

4. Since the RNA molecule is single-stranded and complementary to only one of the two strands of a gene, *it need not have complementary base ratios*. In other words, its adenine content does not necessarily equal its uracil content, nor does its guanine content necessarily equal its cytosine content.

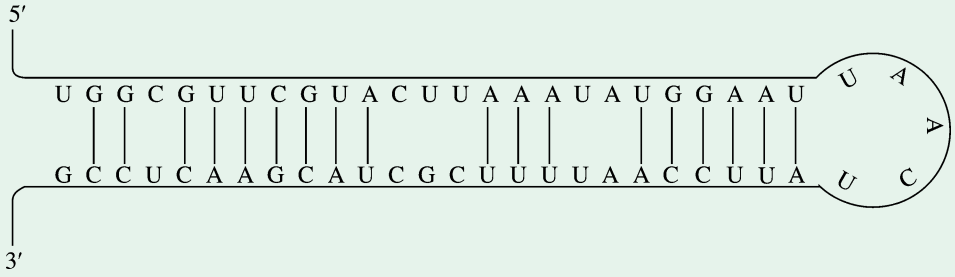
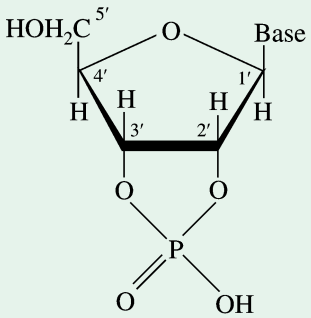


Fig. 15–35. Secondary structure of an RNA molecule

[Note the formation of a 'hairpin' which is dependent upon the intramolecular base pairing.]



A 2′, 3′-cyclic monophosphate nucleotide

5. RNA can be hydrolyzed by weak alkali (pH 9 at 100°C) to 2′, 3′- cyclic diesters of the mononucleotides (shown on page 263) *via* an intermediate compound called 2′, 3′, 5′-triester. This intermediate, however, cannot be formed in alkali-treated DNA because of the absence of a 2′-hydroxyl group in its molecule. Thus, *RNA is alkali-labile whereas DNA is alkali-stable*.

Table 15–8. Comparison between DNA and RNA

DNA	*RNA*
1. Found mainly in the chromatin of the cell nucleus	1. Most of RNA (90%) is present in the cell cytoplasm and a little (10%) in the nucleolus.
2. Never present in free state in cytoplasm	2. May be present in free state.
3. Normally double-stranded and rarely single-stranded.	3. Normally single-stranded and rarely double-stranded.
4. DNA has both 'sense' and 'antisense' strands.	4. The sequence of an RNA molecule is the same as that of the 'antisense' strand.
5. Sugar moiety in DNA is 2′-deoxyribose (hence the nomenclature) which contains an H atom at C-2.	5. Sugar moiety in RNA is ribose (hence the nomenclature) which contains a 2′-hydroxyl group.
6. Sugars in DNA are in the $C_{2'}$ - endo form.	6. Sugars in RNA are in the $C_{3'}$ - endo form.
7. The common nitrogenous bases are adenine, guanine, cytosine and thymine (but not uracil).	7. The common nitrogenous bases are adenine, guanine, cytosine and uracil (but not thymine)
8. Base pairing is inevitable during which adenine pairs with thymine and guanine with cytosine.	8. In case pairing takes place, adenine pairs with uracil and guanine with cytosine.
9. The base ratios (A/T and G/C) are necessarily around one.	9. It need not have complementary base ratios.
10. Base pairing involves the entire length of DNA molecule.	10. Base pairing takes place in only the helical regions of RNA molecule, which amount to roughly half (50%) of the entire RNA molecule.
11. DNA contains only few unusual bases.	11. RNA contains comparatively more unusual bases.
12. DNA is of 3 types : filamentous (double helical or duplex), circular or single-stranded.	12. RNA is of 5 tyes : viral RNA, rRNA, tRNA, mRNA and double-stranded RNA.
13. It consists of a large number of nucleotides (up to 3-4 million) and has, therefore, high molecular weight.	13. It consists of fewer nucleotides (up to 12,000) and has, therefore, low molecular weight.
14. DNA is alkali-stable.	14. RNA is alkali-labile.
15. DNA stains blue with azureph thalate.	15. RNA stains red with azureph thalate.
16. DNA acts as a template for its synthesis.	16. RNA does not act as a template for its synthesis.
	17. Usually RNA does not replicate or transcribe.
17. DNA on replication forms DNA and on transcription forms RNA.	18. During biosynthesis, exonuclease is not needed.
18. During replicatin, exonuclease is needed.	19. RNA exhibits complete and practically instantaneous reversibility of the process of melting.
19. DNA is partially reversible only under certain conditions of slow cooling (= annealing).	
20. DNA undergoes mutation.	20. RNA does not undergo mutation.
21. DNA is the usual genetic material.	21. RNA is the genetic material of some viruses only.
22. DNA is stained green with a dye, pyronin.	22. RNA is stained red with pyronin.

Types of RNA

In all procaryotic and eucaryotic organisms, 3 general types of RNAs are found : *ribosomal*, *transfer* and *messenger RNAs*. Each of these polymeric forms serves as extremely important informational links between DNA, the master carrier of information and proteins. The 3 types of RNA molecules differ from each other by size, function and general stability. Their properties are summarized in Table 15–9.

Table 15–9. RNA molecules in Escherichia coli

Type	*Relative amount*	*Sedimentation coefficient*	*Molecular weight*	*Number of nucleotides*
Ribosomal RNA (rRNA)	80%	23 s	1.2×10^6	3,700
		16 s	0.55×10^6	1,700
		5 s	3.6×10^4	120
Transfer RNA (tRNA)	15%	4 s	2.5×10^4	75
Messenger RNA (mRNA)	5%	—	Heterogeneous	—

Ribosomal RNA (rRNA) or Insoluble RNA

It is the most stable form of RNA and is found in ribosomes. It has the highest molecular weight and is sedimented when a cell homogenate containing 10^{-2} M of Mg^{2+} is centrifuged at high speed (100,000 gravity for 120 minutes). In the bacterium, *Escherichia coli*, there are 3 kinds of RNA called 23 s, 16 s, and 5 s RNA because of sedimentation behaviour. These have molecular weights of 1,200,000, 550,000 and 36,000 respectively. One molecule of each of these 3 types of rRNA is present in each ribosome. *Ribosomal RNA is most abundant of all types of RNAs* and makes up about 80% of the total RNA of a cell. Ribosomal RNA represents about 40-60% of the total weight of ribosomes.

Table 15–10. Ribosomes and their RNAs

Ribosomes	*rRNA*
Procaryotic ribosomes	
30 s	16 s
50 s	5 s, 23 s
Eucaryotic ribosomes	
40 s	18 s
60 s	5 s, 28 s

The ribosomes of procaryotic and eucaryotic cells possess different species of RNA which have been summarized in Table 15–10. The ribosomes of procaryotes and those found in the plastids and mitochondria of eucaryotes contain 3 different types of RNA : 23 s RNA and 5 s RNA in the larger (50 s) subunit and 16 s RNA in the smaller (30 s) subunit. The ribosomes of the eucaryotic cells also contain 3 kinds of RNA : 28 s RNA and 5 s RNA in the larger (60 s) subunit and 18 s RNA in the smaller (40 s) subunit.

The mammalian ribosomes deserve special mention (refer Table 15–11). They also contain 2 major nucleoprotein subunits : a larger one of 2.7 megadaltons (60 s) and a smaller one of 1.3 megadaltons (40 s). The 60 s subunit contains a 5 s rRNA, a 5.8 s rRNA (formerly called 7 s rRNA) and a 28 s rRNA. There are also probably more than 50 specific polypeptides. The 40 s subunit contains a single 18 s rRNA and about 30 polypeptides chains. All types of rRNAs, except the 5 s rRNA, are processed from a single 45 s precursor RNA molecule in the nucleolus. The 5 s rRNA apparently has its own precursor which is independently transcribed.

Table 15–11. RNA components of mammalian ribosomes

Subunit		*RNA*	
Size	Mol. wt.	Size	Mol. wt.
60 s	2.7×10^6	5 s	3.5×10^4
(> 50 polypeptides)		5.8 s	4.5×10^4
		28 s	1.5×10^6
40 s	1.3×10^6	18 s	7.5×10^5
(> 30 polypeptides)			

It is remarkable to note that *rRNA from all sources has G-C contents more than 50%.* The rRNA molecule appears as a single unbranched polynucleotide strand (= *primary structure*). At low ionic strength, the molecule shows a compact rod with random coiling. But at high ionic strength, the molecule reveals the presence of compact helical regions with complementary base pairing and looped outer region (= *secondary structure*). The helical structure (Fig. 15–36) results from a folding back of a single-stranded polymer at areas where hydrogen bonding is possible because of short lengths of complementary structures. The double helical secondary structures in RNA can form within a single RNA molecule or between 2 separate RNA molecules (Fig. 15-37). RNAs can often assume even more complex shapes as in bacteria (Fig. 15-38). The function of rRNA molecules in the ribosomes is not yet fully

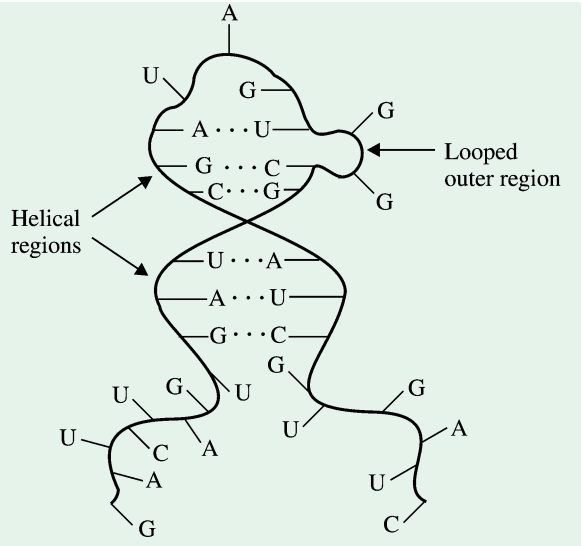
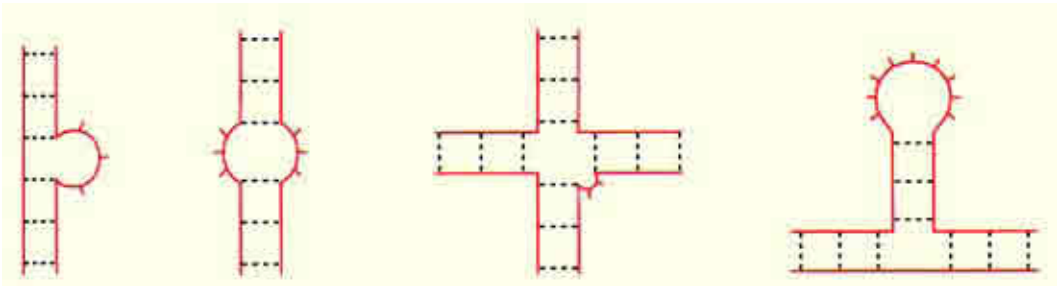


Fig. 15–36. Molecular strucutre of rRNA

Note the helical regions with complementary base pairs and a looped outer region.

(*After Davidson JN, 1972*)



(*a*) Bulge loop (*b*) Internal loop (*c*) Multibranched loop or junction (*d*) Stem-loop

Fig. 15-37. Possible secondary structures of RNA molecues

The double-stranded regions are depicted by connecting hydrogen bonds. Loops are noncomplementary regions that are not hydrogen bonded with complementary bases. Double-stranded RNA structures can form within a single RNA molecule or between two separate RNA molecules.

(*Modified from Jaeger et al, 1993*)

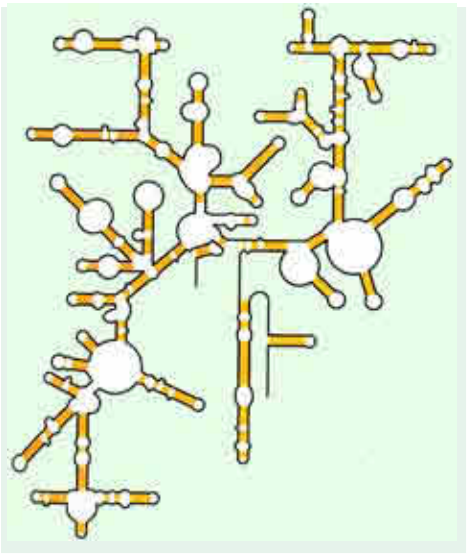


Fig. 15–38. An unusual complex shape of RNA This ribosomal RNA is an integral component of the small ribosomal subunit of a bacterium. The RNA strand is folded back on itself in a highly ordered pattern so that most of the molecule is double stranded.

understood, but they are necessary for ribosomal assembly and seem to provide a specific sequence to which the messenger RNA moelcule can bind in order to be translated.

Transfer RNA (tRNA) or Soluble RNA (sRNA)

Transfer RNA is the smallest polymeric form of RNA. These molecules seem to be generated by the nuclear processing of a precursor molecule. In abundance, the tRNA comes next to rRNA and amounts to about 15% of the total RNA of the cell. The tRNA remains dissolved in solution after centrifuging a broken cell suspension at 100,000 X gravity for several hours (It is for this reason that previously it used to be called *soluble RNA*). The tRNA molecules serve a number of functions, the most important of which is to act as specific carriers of activated amino acids to specific sites on the protein- synthesizing templates. As the function of tRNA is to bind the specific amino acids, one might think that there are 20 types of tRNAs (*i.e.*, as many as the constituent amino acids of proteins). Since the code is **degenerate** (*i.e.*, there is more than one codon for an amino acid), there may also be more than one tRNA for a specific amino acid. In fact, their total number far exceeds than 20. In a bacterial cell, there are more than 70 tRNAs and in eukaryotic cells, this number is even greater, because there are tRNAs specific of mitochondria and chloroplasts (which usually differ from the corresponding cytoplasmic tRNAs). Therefore, there are generally several tRNAs specific of the same amino acid (sometimes up to 4 or 5) ; they are called **isoacceptor tRNAs.** These various tRNAs, capable of binding the same amino acid, differ in their nucleotide sequence; they can either have the same anticodon and therefore recognize the same codon or have different anticodons and thus permit the incorporation of the amino acid in response to multiple codons specifying the same amino acid.

Hydroxyproline (Hyp) is a constituent amino acid of some proteins, but it is formed by a modification of proline after its incorporation into the polypeptide chain, so that there is no tRNA specific of hydroxyproline. There is no tRNA specific of **cystine** either, because the latter is formed by the union of 2 cysteine residues present in the same chain at places sometimes rather distant from one another or even in different polypeptide chains. But there are tRNAs specific of **asparagine (Asn)** and **glutamine (Gln)**, different from the tRNAs specific of the corresponding dicarboxylic amino acids, *i.e.*, aspartic acid (Asp) and glutamic acid (Glu).

As mentioned, each tRNA is specific of an amino aicd, *i.e.*, it can bind (or "accept") only that particular amino acid. Thus, $tRNA^{Ala}$ denotes a tRNA specific of alanine, capable of binding then transferring alanine ; $tRNA^{Ser}$ denotes a tRNA specific of serine, etc. In fact, $tRNA^{Ala}$ denotes a tRNA specific of alanine, but devoid of the amino acid (the terms "discharged" or "deacylated" are also used). To designate the ester bond formed by the binding of an amino acid to a specific tRNA, one must write alanyl-$tRNA^{Ala}$ or simply alanyl-tRNA.

Because of its small size, tRNA has lent itself more readily to isolation procedures and to studies of nucleotide sequence. *Although tRNAs are quite stable in procaryotes, they are somewhat less stable in eucaryotic organisms.*

Primary structure (or base sequence) of a tRNA molecule. The base sequence of a

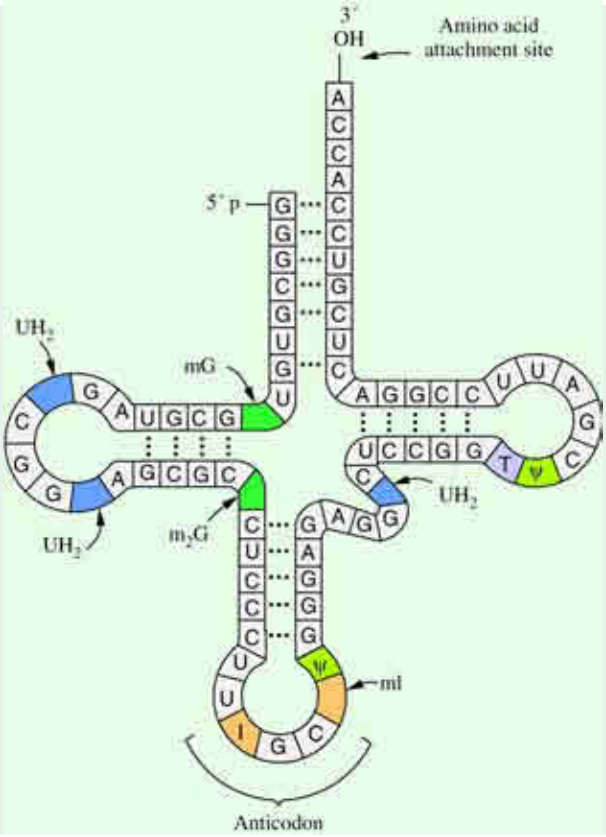


Fig. 15–39. Secondary structure, showing the base sequence of yeast alanine-specific tRNA, drawn in the cloverleaf form.

The abbreviations used for modified nucleosides stand for I = inosine, mI = methylinosine, T = ribothymidine, UH_2 = dihydrouridine, ψ = pseudouridine, mG = methylguanosine and m_2G = dimethylguanosine. Note that a molecule of $tRNA^{Ala}$ contains a total of 77 amino acid residues, with the break-up : 8A + 11U + 25G + 23C + 10 Modified bases.

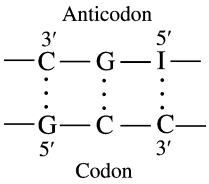
(*Adapted from Lubert Stryer, 1995*)

tRNA molecule specific for the amino acid alanine (refer Fig. 15–40) was first determined by Robert W. Holley and coworkers in 1965, the culminaion of seven years of effort. In fact, this was the first determination of the complete sequence of a nucleic acid. Holley first suggested the '*cloverleaf*' model (*i.e.*, consisting of 3 folds) which is based on a secondary structure which would contain a maximum of intramolecular hydrogen bonding.

The alanyl-tRNA of yeast consists of an unbranched chain of 77 ribonucleotides including 8A, 11U, 25G, 23C and 10 unusual nucleosides which are inosine, ribothymidine, dihydrouridine, pseudouridine and methylated derivatives of inosine and guanosine. The 77 residues represent a molecualr weight of 26,600 as the sodium salt. The 5′ terminus is phosphorylated whereas the 3′ terminus has a free hydroxyl group. While writing down the nucleotide sequence, the usual conventions are followed in that the *5′-phosphate is shown at the left hand and the 3′-hydroxyl on the right side.* The attachment site for the amino acid alanine is 3′-hydroxyl group of the adenosine residue at 3′ terminus of the molecule. The sequence *inosine-guanine-cytosine* in the middle of the molecule is the anticodon. It is complementary to *guanine-cytosine-cytosine*, one of the codons for alanine.

ROBERT W. HOLLEY (LT, 1922-1993)

After obtaining his Bachelor's degree from the University of Illinois, Robert Holley went to Cornell University at Ithaca, New York, gaining his PhD in 1947. From 1944 to 1946, he worked at Cornell University Medical College in New York City on penicillin synthesis. He started working with nucleic, acids at the California Institute of Technology (CALTECH) in 1955 and continued on this area when he returned to Cornell a year later. He chose to study alanine tRNA because it was relatively easy to purify; nevertheless, it took 3 years to work out new procedures and isolate 1 g of tRNA from 140 kg of yeast. He developed new methods for the isolation and sequencing of large oligonucleotides in order to work out the nucleotide sequence of the tRNA, finally completing the work in March 1965. He states that the cloverleaf structure was not his own idea but was discovered by two of his associates, Betty Keller and John Penswick. In fact the cloverleaf was only one of several possible two-dimensional structures and was not confirmed until other tRNAs were found to fit the same pattern. In 1968 Holley received the Nobel Prize for determing the complete sequence of yeast alamyl-tRNA. He also became Professor of Molecular Biology at the Salk Institute in San Diego, the same year. He died in 1993 at Los Gators, California.



Common structural features of tRNAs. Soon after the determination of amino acid sequence of alanyl-tRNA by Holley *et al* in 1965, the sequences of amino acids of many other tRNA

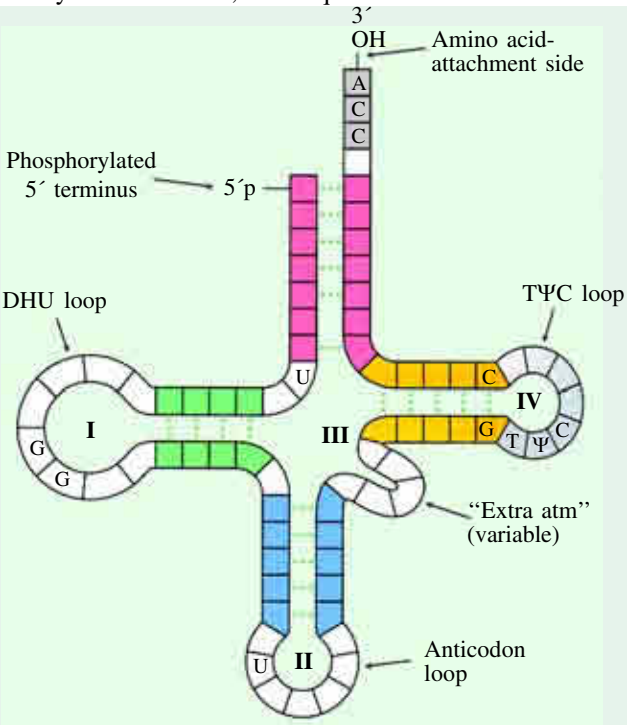


Fig. 15–40. Common features of tRNA molecules

Comparison of the base sequences of many tRNAs reveals a number of conserved features. Note the presence of an amino acid attachment site, the 4 'loops' and the 4 'arms' or stems.

molecules were determined by many workers. As a result, about 50 sequences are now known. These have been found to possess certain *common structural features* (refer Fig. 15–40). These are enumerated below :

1. All tRNA molecules have a common design and consist of 3 folds giving it a shape of the cloverleaf with four arms (refer Fig. 15–41) ; the longer tRNAs have a short fifth or extra arm. The actual 3-dimensional structure of a tRNA looks more like a twisted L than a cloverleaf (Fig. 15–42).

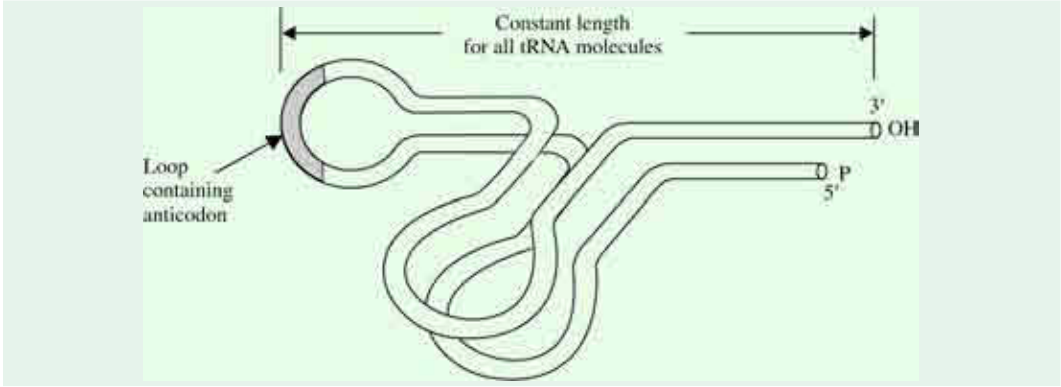


Fig. 15–41. Cloverleaf model of tRNA

2. All tRNA molecules are unbranched chains containing from 73 to 93 ribonucleotide residues, corresponding to molecular weights between 24,000 and 31,000 (Mitochondria contain distinct RNAs that are somewhat smaller). For example, $tRNA^{Ala}$ contains 77 nucleotides (Holley *et al*, 1965), $tRNA^{Tyr}$ contains 78 nucleotides (Medison *et al*, 1967) $tRNA^{Ser}$ contains 85 nucleotides (Zachan *et, al,* 1967).
3. They contain from 7 to 15 unusual modified bases. Many of these unusual bases are methylated or dimethylated derivatives of A, U, G and C. These include nucleotides of pseudouridine, various methylated adenines and guanines, methylated pyrimidines such as thymine and 5-methylcytosine and others. Not all these are present in any one source of tRNA but *pseudouridine* (ϕU) *is the most abundant and universally distributed*. Although, the role of these bases is uncertain, yet two roles seem certain :
   a. Methylation prevents the formation of certain base pairs so that some of the bases become accessible for other interactions.
   b. Methylation imparts hydrophobic character to some portions of tRNA molecules which may be important for their interaction with the *synthetases* and with ribosomal proteins.

   In fact, the modified bases of tRNAs prompted Crick to remark :

   "It almost appears as if tRNAs were nature's attempt to make an RNA molecule play the role of a protein."

   What he meant by this statement is that the unusual bases are probably needed to stabilize a nucleic acid molecule that has the intricate folding of the 3-'D' structure of a globular protein.
4. The 5′ end of tRNAs is phosphorylated. The 5′ terminal residue is usually guanylate (pG).
5. The base sequence at the 3′ end of all tRNAs is CCA. All amino acids bind to this terminal adenosine *via* the 3′-OH group of its ribose.

   (5′ end) G.....tRNA.....CCA (3′ end)
6. About 50% of the nucleotides in tRNAs are base-paired to form double helices.

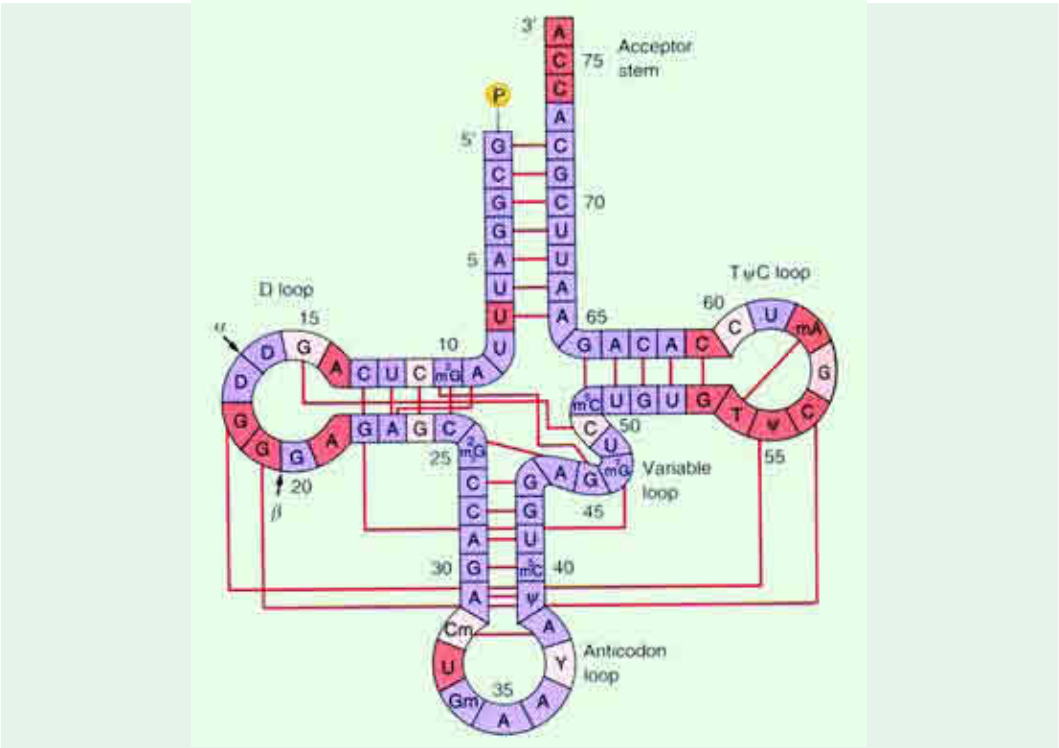


Fig. 15-42. Secondary or two-dimensional structure of yeast phenylalanine tRNA in the form of a cloverleaf

The conventional numbering of nucleotides begins at the 5′ end and proceeds toward the 3′ end. In all tRNAs, the nucleotides at the 3′ end contains the sequence CCA. Certain locations can have additional nucleotides that are not found in all tRNA molecules. The entire nucleotide sequence of yeast tRNA is shown. Red lines connecting nucleotides indicate hydrogen-bonded bases. Rose squares indicate constant nucleotides; tan squares indicate a constant purine or pyrimidine. Insertion of nucleotides in the D loop occurs at positions α and β for different tRNAs.

(*Courtesy: Quigley GJ and Rich A, 1976*)

7. There are, however, 5 groups of bases which are not base-paired. These 5 groups, of which 4 form 'loops', are :
   a. the 3′ CCA terminal region,
   b. the ribothymine-pseudouracil-cytosine (= T ϕ C) loop,
   c. the 'extra arm' or little loop, which contains a variable number of residues,
   d. the dihydrouracil (= DHU) loop, which contains several dihydrouracil residues, and
   e. the anticodon loop, which consists of 7 bases with the sequence,
   5′ —pyrimidine—pyrimidine—X—Y—Z—modified purine—variable base — 3′
   This loop contains a triplet of bases which allows the tRNA to hydrogen-bond to a complementary sequence on mRNA attached to a ribosome.
8. The 4 loops are *recognition sites*. Each tRNA must have at least two such recognition sites : one for the activated amino acid-enzyme complex with which it must react to form the aminoacyl-tRNA and another for the site on a messenger RNA molecule which contains the code (codon) for that particular amino acid. It is interesting to note that the former involves recognition by bases of amino acid residues (either of the activated amino acid or of a site on the enzyme molecule) whereas the latter involves recognition by bases of bases (hydrogen bonding).
9. A unique similarity among all tRNA molecules is that *the overall distance from CCA at*

one end to the anticodon at the other end is constant (refer Fig. 15–33). The difference in nucleotide numbers in various tRNA molecules is, in fact, compensated for by the size of the "extra arm", which is located between the anticodon loop and T ϕ C loop.

Tertiary structure (or three-dimensional structure) of a tRNA molecule. Alexander Rich and Aaron Klug (1960s), on the basis of their x-ray crystallographic studies, have elucidated the 3-'D' structure of the phenylalanine-accepting tRNA from yeast (refer Figs. 15–42 and 15-43). This

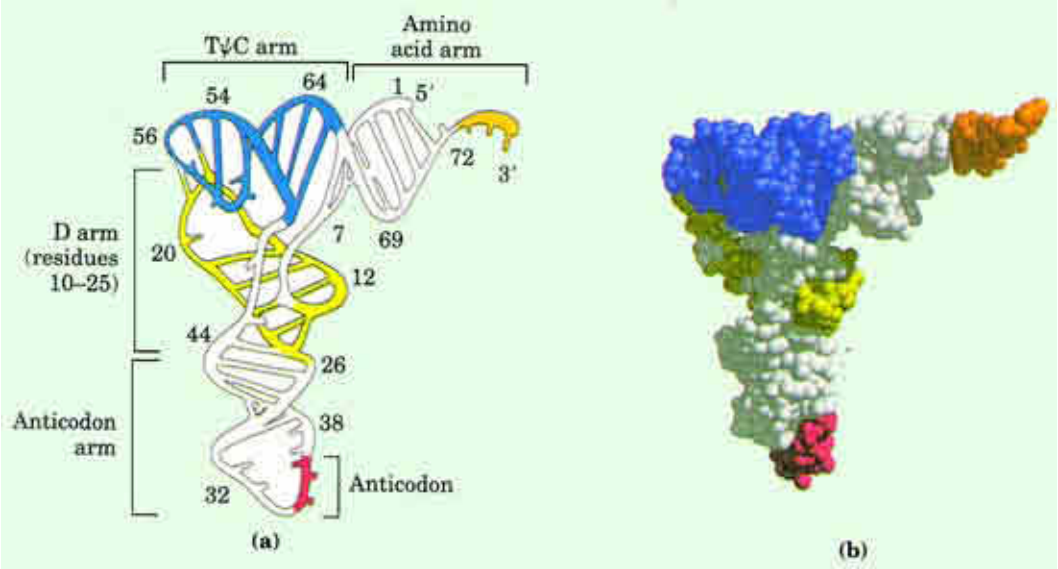


Fig. 15–43. The 3-dimensional tertiary structure of yeast phenylalanine tRNA, as deducted from diffraction analysis

The specified amino acid is attached at the CCA 3′ terminus. The amino acid (or amino acceptor) arm and TΨC) arm form a continuous double helix, and the anticodon (AC) arm and the DHU (D) arm form the other partially continuous double helix. These 2 helical columns meet to form a twisted L-shaped molecule.

is probably closer to reality than the postulated cloverleaf structure. The important features observed by them are :

1. The molecule is L-shaped.
2. There are 2 segments of the double helix. Each of these helices contains about 10 base pairs which correspond to one turn of the helix. These helical segments are perpendicular to each other, thus forming an L.
3. The CCA terminus containing the attachment site for the amino acid is at one end of the L. The other end of the L is occupied by the anticodon loop. The DHU and TϕC loops form the corner of the L.
4. The CCA terminus and the adjacent helical region do not interact strongly with the rest of the molecule. This part of the molecule may change conformation during amino acid activation and also during protein synthesis on the ribosome.

Messenger RNA (mRNA) or Template RNA

The abundance of RNA in the cytoplasm and its role in protein synthesis suggested that the genetic information of nuclear DNA is transmitted to an RNA which functions at the sites of protein synthesis. In 1961, the two Nobel laureates, Francois Jacob and Jacques Monod postulated that control of protein formation, at least in certain microorganisms, is determined by the rate of synthesis of templates. This requires that the templates do not accumulate in contrast of the constant presence of DNA, rRNA and tRNA. They, therefore, suggested the transient existence

of an RNA, which they called "the messenger" RNA (mRNA). Later, they proposed that "the messenger" should have the following properties :

1. The messenger should be a polynucleotide.
2. The base composition of the messenger should reflect the base composition of the DNA that specifies it.
3. The mRNA should be very heterogeneous in size because genes or groups of genes vary in length. They also correctly assumed that 3 nucleotides code for one amino acid and that the molecular weight of an mRNA should be at least a half million.
4. The messenger should be, for a short period, associated with ribosomes.
5. The messenger should be synthesized and degraded very rapidly.

All these properties are nowadays ascribed to the messenger RNA because the other 2 types of RNAs (rRNA and tRNA)are homogeneous and also their base composition is similar in species that have very different DNA base ratios.

Messenger RNA is most heterogeneous in size and stability among all the types of RNAs. It has large molecular weight approaching 2×10^6 and amounts to about 5% of the total RNA of a cell. It is synthesized on the surface of DNA template. Thus, it has base sequence complementary to DNA and carries genetic information or 'message' (hence its nomenclature) for the assembly of amino acids from DNA to ribosomes, the site of protein synthesis. In procaryotic cells, mRNA is metabolically unstable with a high turnover rate whereas it is rather stable in eucaryotes (cf tRNA). It is synthesized by *DNA-dependent RNA polymerase*.

On account of its heterogeneity, mRNA varies greatly in chain length. Since few proteins contain less than 100 amino acids, the mRNA coding for these proteins must have at least 100×3 or 300 nucleotide residues. In *E. coli*, the average size of mRNA is 900 to 1,500 nucleotide units.

If mRNA carries the codes for the synthesis of simple protein molecule, it is called *monocistronic type* and if it codes for more than one kind of protein, it is known as *polycistronic type* as in *Escherichia coli*.

The mRNAs are unstable in the bacterial systems with a half-life from a few seconds to about 2 minutes. In mammalian systems, however, mRNA molecules are more stable with a half-life ranging from a few hours to one day.

Half-life is the period of time after which half is degraded.

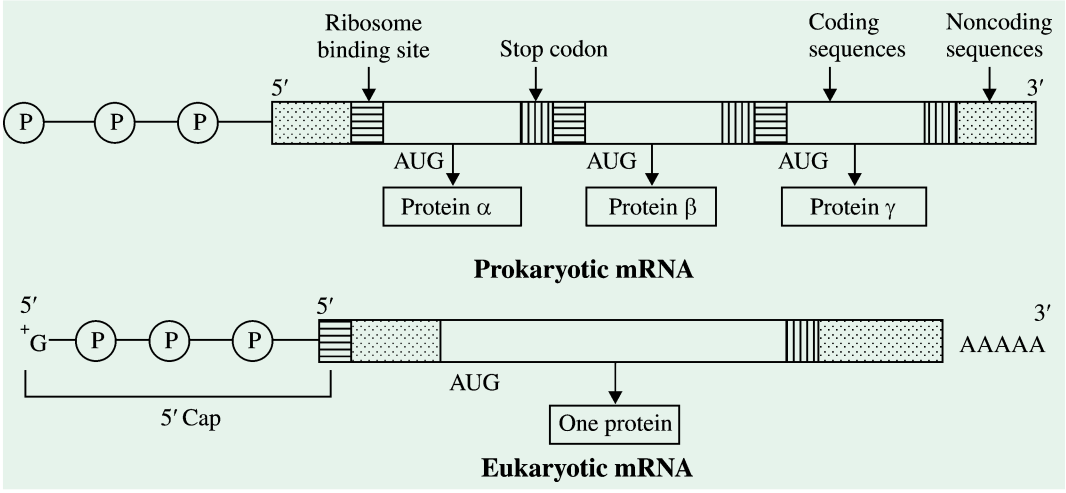


Fig. 15–44. A comparison of the structures of prokaryotic and eukaryotic messenger RNA molecules

The mRNAs are single-stranded and complementary to the sense strand of their respective structural genes. Although both types of mRNA molecules (prokaryotic and eukaryotic) are synthesized with a triphosphate group at the 5′ end, there is a basic difference between the two, as depicted in Fig. 15–44. In eukaryotes, the mRNA molecule immediately acquires a 5′ cap, which is part of the structure recognized by the small ribosomal subunit. Protein synthesis, therefore, begins at the start codon near the 5′ end of the mRNA. On the contrary, in prokaryotes the 5′end has no special significance, and there can be many ribosome-binding sites (called shine-Dalgarno sequences, see page 620) in the interior of an mRNA chain, each resulting in the synthesis of a different protein.

As stated above, the eukaryotic mRNA molecules, especially those of mammals, have some peculiar characteristics. The 5′ end of mRNA is ‘capped’ by a 7-methylguanosine triphosphate which is linked to an adjacent 2′-O-methylribonucleoside at its 5′-hydroxyl through the 3 phosphates (Fig. 15–45). Although the function of this capping of mRNA is not well understood, the cap is probably involved in the recognition of mRNA by the translating machinery. The translation of mRNA into proteins begins at the capped 5′ end. The other end of most mRNA molecules, the 3′ hydroxyl end, has attached a polymer of adenylate residues, 20–250 nucleotides in length. The specificity of the poly A “tail” at the 3′ hydroxyl end of mRNA is not understood. It probably serves to maintain the intracellular stability of the specific mRNA.

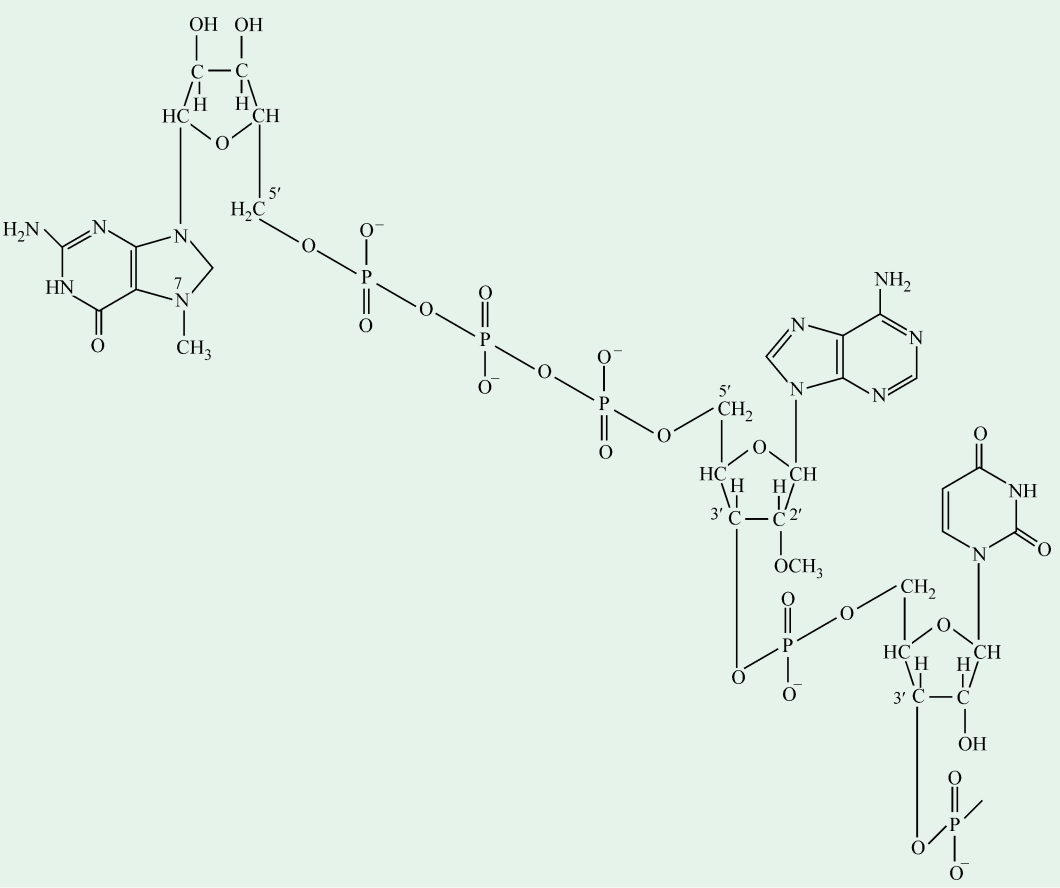


Fig. 15–45. The ‘cap’ structure attached to the 5′ end of most mRNA molecules

Note that a 7-methylguanosine triphosphate is attached at the 5′ terminus of the mRNA, which usually contains a 2′-O-methylpurine nucleotide.

Heterogeneous Nuclear RNA(hnRNA)

In mammalian cells including those of human beings, a precursor RNA is first synthesized in the nucleoplasm by *DNA-dependent RNA polymerase*. This precursor is then degraded by a *nuclear nuclease* to mRNA that is then translocated to the cytoplasm where it becomes associated to the ribosomal system. This precursor RNA constitutes the fourth class of RNA molecules and is designated as heterogeneous nuclear RNA (*hn*RNA). The *hn*RNA molecules may have molecular weights exceeding 10^7 daltons whereas the mRNA molecules are generally smaller than 2×10^6 daltons. Most mammalian mRNA molecules are 400–4,000 nucleotides in length whereas a *hn*RNA molecule possesses 5,000–50,000 nucleotides. Some uncertainty still exists concerning the precursor-product relationship between *hn*RNA and mRNA, the former being 10—100 times longer than the latter. Thus, the hnRNA molecules appear to be processed to generate the mRNA molecules which then enter the cytoplasm to serve as templates for protein synthesis.

The painstaking researches conducted by the two 1989 Nobel Laureates in chemistry, Thomas Cech from the University of Colorado and Sidney Altman from the Yale University, have recently shown that RNA not only carries the blueprint message from the imperious DNA to the factories that churn out proteins but it could catalyze the reactions in the same way that protein enzymes do. The particular type of RNA acting as a biological catalyst has, hence, been called by them as *riboenzyme* or simply *ribozyme* (Fig. 15-46). This has, in fact, demolished the firm impression to date that proteins alone can act as enzymes, the biological catalysts. They have, independently of each other, shown that pieces of RNA can slip themselves out of larger chains and then splice together the severed ends. Thus, their work has strengthened the claim of RNA as being the progenitor molecule of life in this world. And the 'plain-looking' RNA has been transformed by them into a fairy princess. They, in fact, showed that RNA is the Cinderella among the nucleic acids. In the words of Thomas Cech, the 'ribozymes have the ability to act as a sort of molecular scissors.'

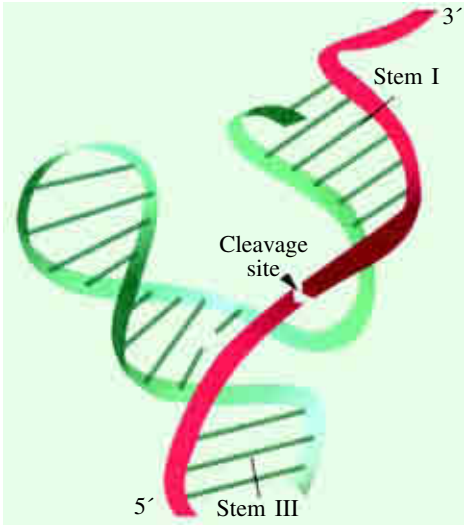


Fig. 15-46. Custom-designed ribozymes The ribozyme shown here in green, called a hammerhead ribozyme, bind to RNAs having a complementary nucleotide sequence and catalyzes a reaction that cleaves the back bone of the substrate RNA (show in Red).

THOMAS CECH

Thomas Cech was born in 1947 at Chicago, USA. He received his PhD from the University of California, Berkeley, and then moved to the Massachusetts Institute of Technology, (MIT) in 1975. In 1978, he joined the faculty of the University of Colorado at Boulder and began the research that led to the discovery of self-splicing RNA. The first suggestion that the *Tetrahymena* intron might mediate its own splicing was published in 1981, and was followed the next year by the demonstration that the nitron RNA excises itself from the adjoining exons in the complete absence of any proteins. In 1986, Cech and his group engineered the intron RNA into an RNA enzyme, capable of cleaving other RNA molecules without being altered in the process. As well as the importance of these discoveries in understanding introns and how they are spliced. Cech's work has implications regarding the origins of life, as ribozymes may have served as both templates and catalysts to produce a primordial, self-replicating biochemical system. Cech shared the 1989 Nobel Prize for Chemistry with Sydney Altman of Yale University.

In fact, the discovery of ribozymes has led to the concept of an **"RNA world",** in which the promordial biomolecule was RNA. Later researches have revealed that RNA cannot only catalyze excision of introns and splicing of exons, it can even break bonds between amino acids and nucleic acids. Cech has suggested that if RNA can break the bonds, it may also make them. And evidences indicate that RNA-driven protein synthesis may occur even now in certain bacterial systems. Nevertheless, an intricate problem with the notion of an RNA world is the origin of RNA in the first place. RNA is a complex molecule, notoriously difficult to synthesize under ideal conditions in the laboratory. How could it have arisen spontaneously under conditions prevailing in the prebiotic soup ?

Informosome

Spirin, Beltisina and Lerman (1965) have investigated that in certain eucaryotic cells, the mRNA does not enter the cytoplasm as a naked RNA strand but often remains encased by certain proteins. Spirin has proposed the term informosome for this mRNA-protein complex. The proteins bound to informosomes provide stability to mRNA molecule and also protect it from degradation by the enzyme *ribonuclease*.

REFERENCES

1. **Adams RLP, Burdon RH, Cambell AM, Leader DP, Smellie RMS :** The Biochemistry of the Nucleic Acids. *9th ed., Chapman and Hall, London. 1981.*
2. **Bloomfield VA, Crothers DM, Tinoco I :** Physical Chemistry of Nucelic Acids. *Harper and Row, New York. 1974.*
3. **Borst P :** Structure and function of mitochondrial DNA. *Trends Biochem. Sci.* **2 :** *31, 1977.*
4. **Brawerman G :** Eukaryotic messenger RNA. *Ann. Rev. Biochem.*, **43 :** *621, 1974.*
5. **Cantor CR, Schimmel PR :** Biophysical Chemistry. 3 vols. *W.H.Freeman. 1980.*
6. **Chargaff E, Davidson JN (editors) :** The Nucleic Acids : Chemistry and Biology. *3 vols. Academic Press, Inc., New York. 1955, 1955, 1960.*
7. **Crick FHC :** The structure of the hereditary material. *Sci. Amer.* **191 (4) :** *54-61, 1954.*
8. **Crick FHC :** Nucleic acids. *Sci. Amer.* **197(3) :** *188, 1957.*
9. **Crick FHC, Klug A :** Kinky Helix. *Nature* **255 :** *530, 1975.*
10. **Darnel JE Jr :** RNA. *Sci. Amer.* **253 (4) :** *68-78, 1985.*
11. **Davidson JN :** The Biochemistry of the Nucleic Acids. *8th ed., Chapman and Hall, London. 1976.*
12. **Dickerson RE :** The DNA helix and how it is read. *Sci. Amer.* **249 (6) :** *94-111, 1983.*
13. **Dickerson RE :** DNA structure from A to Z. *Methods Enzymol.* **211 :** *67-111, 1992.*
14. **Doctor et al :** Nucleotide sequence in *E. coli* tyrosine-tRNA. *Science*. **163 :** *693, 1969.*
15. **Falsenfeld G :** DNA. *Sci. Amer.* **253 (4) :** *58-67, 1985.*
16. **Felsenfeld G, Miles HT :** The physical and chemical properties of nucleic acids. *Ann. Rev. Biochem.*, **36 :** *407-448, 1967.*
17. **Fraenkel-Conrat H :** The Chemistry and Biology of Viruses. *Academic Press, Inc., New York. 1969.*
18. **Friedberg EC :** DNA Repair. *W.H. Freeman and Company, New York. 1985.*
19. **Gesteland RF, Atkins JF (editors) :** The RNA World. *Cold Spring Harbour, New York. 1993.*
20. **Grivell LA :** Mitochondrial DNA. *Sci. Amer.* **248 (3) :** *60-73, 1983.*
21. **Hartman SC :** Purines and Pyrimidines. *In Greenberg DM (editor), Metabolic Pathways. (3rd editon) vol. IV, 1-68, Academic Press, Inc., New York. 1970.*

22. **Henderson JF, Paterson ARP :** Nucleotide Metabolism : An Introduction. *Academic Press, Inc., New York. 1973.*
23. **Holley RW :** The nucleotide sequence of a nucleic acid. *Sci. Amer.* **214 (2) :** *30-39, 1966.*
24. **Holley RW et al :** Structure of a ribonucleic acid. *Science.* **147 :** *1462-1465, 1965; also Science.* **153 :** *531-534, 1966.*
25. **Htun H, Dalhberg JE :** Topology and formation of triple-stranded H-DNA. *Science.* **243 :** *1571-1576, 1989.*
26. **Hurst DT :** An Introduction to the Chemistry and Biochemistry of Pyrimidines, Purines, and Pteridines. *Wiley, New York. 1980.*
27. **Jordan :** The Chemistry of Nucleic Acids. *Butterworth, 1960.*
28. **Kornberg A :** Biologic synthesis of deoxyribonucleic acid. *Science.* **131 :** *1503-1508, 1960.*
29. **Kornberg A :** DNA Synthesis. *W.H. Freeman and Company, New York. 1974.*
30. **Kornberg A, Baker TA :** DNA Replication. 2nd ed., *W.H. Freeman and Company, New York. 1991.*
31. **Mainwaring WIP, Parish JH, Pickering JD, Manna NH :** Nucleic Acid Biochemistry and Molecular Biology. *Blackwell Scientific Publications, Boston. 1982.*
32. **McCarty M :** The Transforming Principle : Discovering that Genes Are Made of DNA. *Norton. 1985.*
33. **McClain WH :** Transfer RNA identity. *FASEB J.* **7 (1) :** *72-78, 1993.*
34. **Michelson AM :** The Chemistry of Nucleosides and Nucleotides. *Academic Press, Inc., New York. 1963.*
35. **Missky AE :** The discovery of DNA. *Sci. Amer.* **218 (6) :** *78, 1968.*
36. **Olby R :** The Path to the Double Helix. *University of Washington Press, Seattle. 1974.*
37. **Orgell LE, Crick FHC :** Selfish DNA : the ultimate parasite. *Nature.* **284 :** *604-607, 1980.*
38. **Perutz MF :** Proteins and Nucleic Acids. *Elsevier Publishing Co., New York. 1962.*
39. **Portugal FH, Cohen JS :** A Century of DNA : A History of the Discovery of the Structure and Function of the Genetic Substance. *MIT Press. 1977.*
40. **Rich A, Kim H :** The three-dimensional structure of transfer RNA. *Sci. Amer. January, 1978.*
41. **Rich A, Nordheim A, Wang AH-J :** The chemistry and biology of left-handed Z-DNA. *Ann. Rev. Biochem.* **53 :** *791-846, 1984.*
42. **Rich A, Raj Bhandari UL :** Transfer RNA : Molecular structure, sequence and properties. *Ann. Rev. Biochem.*, **45 :** *805, 1976.*
43. **Robertus JD et al :** Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature.* **250 :** *546-551, 1974.*
44. **Saenger W :** Principles of Nucleic Acid Structure. *Springer-Verlag, New York. 1984.*
45. **Sayre A :** Rosalind Franklin and DNA. *W.W. Norton & Co., Inc., New York. 1978.*
46. **Scheit KH :** Nucleotide Analogs : Synthesis and Biological Function. *John Wiley and Sons, New York. 1980.*
47. **Schimmel PRD, Soll D, Abelson JN (editors) :** Transfer RNA : Structure, Properties and Recognition. *Cold Spring Harbour, New York. 1979.*
48. **Singer MF, Leder P :** Messenger RNA : An evaluation. *Ann. Rev. Biochem.,* **35 :** *195-230, 1966.*
49. **Smith MM :** Histone structure and function. *Curr. Opin. Cell Biol.* **3 :** *429-437, 1991.*
50. **Spencer JH :** The Physics and Chemistry of DNA and RNA. *W.B. Saunders, Philadelphia. 1972.*

51. **Stewart PR, Letham DS (editors) :** The Ribonucleic Acids. *2nd ed., Springer-Verlag, New York 1977.*
52. **Watson JD :** The Double Helix. *Atheneum. 1968.*
53. **Watson JD, Crick FHC :** Molecular structure of nucleic acid. A structure for deoxyribose nucleic acid. *Nature.* **171 :** *737-738, 1953 a.*
54. **Watson JD, Crick FHC :** Genetic implications of the structure of deoxyribonucleic acid. *Nature*, **171 :** *964-967, 1953 b.*
55. **Watson JD, Hopkins NH, Roberts JW, Steitz JA, Weiner AM :** Molecular Biology of the Gene. *4th ed., The Benjamin/Cummings Publishing Company, Menlo Park, CA. 1987.*
56. **Wells RD, Harvey SC (editors) :** Unusual DNA Structures. *Springer-Verlag, New York. 1988.*
57. **Wilkins M :** Molecular Configuration of Nucleic Acids. *Science.* **140 :** *941, 1963.*
58. **Younghusband HB, Inman RB :** The electron microscopy of DNA. *Ann. Rev. Biochem.* **43 :** *605, 1974.*

PROBLEMS

1. Write the complementary sequence (in the standard 5′ → 3′ notation) for **(*a*)** GATCAA, **(*b*)** TCGAAC, **(*c*)** ACGCGT, and **(*d*)** TACCAT.
2. The composition (in mole-fraction units) of one of the strands of a double-helical DNA molecule is [A] = 0.30 and [G] = 0.24. **(*a*)** What can you say about [T] and [C] for the same strand ? **(*b*)** What can you say about [A], [G], [T], and [C] of the complementary strand ?
3. **(*a*)** Write the sequence of the mRNA molecule synthesized from a DNA template strand having the sequence :

   5′-ATCGTACCGTTA-3′

   (*b*) What amino acid sequence is encoded by the following base sequence of an mRNA molecule ? Assume that the reading frame starts at 5′ end.

   5′-UUGCCUAGUGAUUGGAUG-3′

   (*c*) What is the sequence of the polypeptide formed on addition of poly(UUAC) to a cell-free protein-synthesizing system ?
4. RNA is readily hydrolyzed by alkali, whereas DNA is not. Why ?
5. The transition from B-DNA to Z-DNA occurs over a small change in the superhelix density, which shows that the transition is highly cooperative.

   (*a*) Consider a DNA molecule at the midpoint of this transition. Are B- and Z-DNA regions frequently intermingled or are there long stretches of each ?

   (*b*) What does the finding reveal reveal about the energetics of forming a junction between the two kinds of helices ?

   (*c*) Would you expect the transition from B– to A-DNA to be more or less cooperative than the one from B– to Z-DNA? Why ?
6. The sequence of part of an mRNA is

   5′-AUGGGGAACAGCAAGAGUGGGGCCCUGUCCAAGGAG-3′

   What is the sequence of the DNA coding strand ? Of the DNA template strand ?
7. Why is RNA synthesis not as carefully monitored for errors as is DNA synthesis ?
8. Why is it advantageous for DNA synthesis to be more rapid than RNA synthesis ?
9. A viral DNA is analyzed and found to have the following base composition, in mole present : A = 32, G = 16, T = 40, C = 12.

(*a*) What can you immediately conclude about this DNA ?

(*b*) What kind of secondary structure do you think it would have ?

10. Given the following sequence for one strand of a double-strand oligonucleotide : $^{5'}$ACCGTAAGGCTTTAG$^{3'}$

(***a***) Write the sequence for the complementary DNA strand.

(***b***) Suppose you knew that the strand shown above had phosphate on both ends. Using an accepted nomenclature, write the sequence so as to show this.

(***c***) Write the sequence of the RNA complementary to the strand shown above.

11. What positions in a purine ring have the potential to form hydrogen bonds, but are not involved in the hydrogen bonds of Watson–Crick base pairs ?

12. Write the base sequence of the complementary strand of double-helical DNA in which one strand has the sequence (5′)ATGCCCGTATGCATTC(3′).

13. Calculate the weight in grams of a double-helical DNA molecule stretching from the earth to the moon (~320,000 km). The DNA double helix weighs about 1×10^{-18} g per 1,000 nucleotide pairs; each base pair extends 0.34 nm. For an interesting comparison, your body contains about 0.5 g of DNA !

14. Hairpins may form at palindromic sequences in single strands of either RNA or DNA. How is the helical structure of a hairpin in RNA different from that of a hairpin in DNA ?

15. In the cells of many eukaryotic organisms, there are highly specialized systems that specifically repair G–T mismatches in DNA. The mismatch is repaired to form a G≡C base pair (not A=T). This G–T mismatch repair system occurs in addition to a more general system that repairs virtually all mismatches. Can you think of a reason why cells require a specialized system to repair G–T mismatches ?

16. Explain why there is an increase in the absorption of UV light (hyperchromic effect) when double-stranded DNA is denatured.

17. In samples of DNA isolated from two unidentified species of bacteria, adenine makes up 32 and 17%, respectively, of the total bases. What relative proportions of adenine, guanine, thymine, and cytosine would you expect to find in the two DNA samples ? What assumptions have you made ? One of these bacteria was isolated from a hot spring (64 ºC). Which DNA came from this thermophilic bacterium ? What is the basis for your answer ?

18. In the coding strand of DNA for the alpha gene of normal hemoglobin (HbA), the three bases that correspond to position 142 of the mRNA synthesized are TAA and the alpha chain has 141 amino acids. In the coding strand of the gene for the alpha chain of hemoglobin Constant Spring, the three bases in the same position as above are CAA and the alpha chain produced contains 172 amino acids. Explain the mutation that has occurred. Is the mutation a frameshift or a point mutation ?

19. How would you measure the turnover of DNA ?

20. Write the complementary sequence (in the standard 5′ → 3′ notation) for :

(***a***) GATCAA

(***b***) TCGAAC

(***c***) ACGCGT

(***d***) TACCAT

21. Write the sequence of the mRNA molecule synthesized from a DNA template strand having the sequence :

5′-ATCGTACCGTTA-3′