Nucleoproteins are compounds containing nucleic acid and protein, especially, protamines and histones. These are usually the salt-like compounds of proteins since the two components have opposite charges and are bound to each other by electrostatic forces. They are present in nuclear substances as well as in the cytoplasm. These may be considered as the sites for the synthesis of proteins and enzymes.

There are two kinds of nucleoproteins: 1) *Deoxyribonucleoproteins* (DNP) – *deoxyribonucleic acid* (DNA) is prosthetic group; 2) *Ribonucleoproteins* (RNP) – *ribonucleic acid* (RNA) is prosthetic group.

Nucleoproteins are of central importance in the *storage*, *transmission*, and *expression* of genetic information.

Nucleotides and their derivatives are biologically ubiquitous substances that participate in nearly all biochemical processes:

1. They form the monomeric units of nucleic acids and thereby play central roles in both the storage and the expression of genetic information.

2. **Nucleoside triphosphates**, most conspicuously ATP, are the "energy-rich" end products of the majority of energy-releasing pathways and the substances whose utilization drives most energy-requiring processes.

3. Most metabolic pathways are regulated, at least in part, by the levels of nucleotides such as ATP and ADP. Moreover, certain nucleotides function as intracellular signals that regulate the activities of numerous metabolic processes.

4. Nucleotide derivatives, such as **nicotinamide adenine dinucleotide**, **flavin adenine dinucleotide**, and **coenzyme A**, are required participants in many enzymatic reactions.

5. As components of the enzymelike nucleic acids known as **ribozymes**, nucleotides have important catalytic activities themselves.

Nucleotides, Nucleosides, and Bases

Nucleotides are phosphate esters of a five-carbon sugar (which is known as a **pentose**) in which a nitrogenous base is covalently linked to C1′ of the sugar residue.

In **ribonucleotides** (Fig. 1), the monomeric units of RNA, the pentose is **D-ribose**, whereas in **deoxyribonucleotides** (or just deoxynucleotides; Fig. 1), the monomeric units of DNA, the pentose is 2′-deoxy-D-ribose (note that the "primed" numbers refer to the atoms of the ribose residue; "unprimed" numbers refer to atoms of the nitrogenous base). The phosphate group may be bonded to C5′ of the pentose to form a 5′-nucleotide (Fig. 1) or to its C3′ to form a 3′-nucleotide. If the phosphate group is absent, the compound is known as a nucleoside. A 5′-nucleotide, for example, may therefore be referred to as a nucleoside-5′-phosphate. In all naturally occurring nucleotides and nucleosides, the bond linking the nitrogenous base to the pentose C1′ atom (which is called a glycosidic bond) extends from the same side of the ribose ring as does the C4′-C5′ bond (the so-called β configuration) rather than from the opposite side (the α configuration). Note that nucleotide phosphate groups are doubly ionized at physiological pH's; that is, *nucleotides are moderately strong acids*.

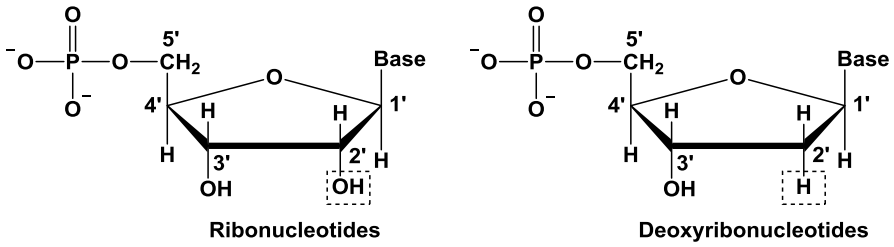
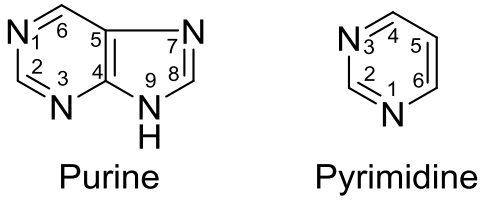


Figure 1. Chemical Structures of Ribonucleotides and Deoxyribonucleotides.

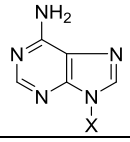
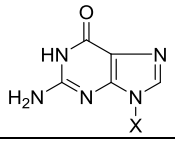
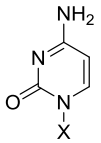
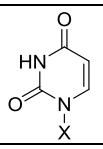
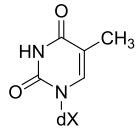
The nitrogenous bases are planar, aromatic, heterocyclic molecules which, for the most part, are derivatives of either **purine** or **pyrimidine**.



The structures, names, and abbreviations of the common bases, nucleosides, and nucleotides are given in Table 1. The major purine components of nucleic acids are adenine and guanine residues; the major pyrimidine residues are those of cytosine, uracil (which occurs mainly in RNA), and thymine (5-methyluracil, which occurs mainly in DNA). The purines form glycosidic bonds to ribose via their N9 atoms,

whereas pyrimidines do so through their N1 atoms (note that purines and pyrimidines have dissimilar atom numbering schemes).

Table 1. Names and Abbreviations of Nucleic Acid Bases, Nucleosides, and Nucleotides.

Base Formula	Base (X = H)	Nucleoside (X = ribose*)	Nucleotide (X = ribose phosphate*)
	Adenine Ade A	Adenosine Ado A	Adenylic acid Adenosine monophosphate AMP
	Guanine Gua G	Guanosine Guo G	Guanylic acid Guanosine monophosphate GMP
	Cytosine Cyt C	Cytidine Cyd C	Cytidylic acid Cytidine monophosphate CMP
	Uracil Ura U	Uridine Urd U	Uridylic acid Uridine monophosphate UMP
	Thymine Thy T	Deoxythymidine dThd dT	Deoxythymidylic acid Deoxythymidine monophosphate dTMP

* The presence of a 2′-deoxyribose unit in place of ribose, as occurs in DNA, is implied by the prefixes “deoxy” or “d.” For example, the deoxynucleoside of adenine is deoxyadenosine or dA. However, for thymine-containing residues, which rarely occur in RNA, the prefix is redundant and may be dropped. The presence of a ribose unit may be explicitly implied by the prefixes “ribo” or “r.” Thus the ribonucleotide of thymine is ribothymidine or rT.

The Chemical Structures of DNA and RNA. The chemical structures of the nucleic acids were elucidated by the early 1950s largely through the efforts of Phoebus Levene, followed by the work of Alexander Todd. Nucleic acids are, with few exceptions, linear polymers of nucleotides whose phosphate groups bridge the 3′ and 5′ positions of successive sugar residues (e.g., Fig. 2). The phosphates of these **polynucleotides**, the **phosphodiester** groups, are acidic, so that, at physiological pH’s, nucleic acids are polyanions. Polynucleotides have directionality, that is, each

has a **3′ end** (the end whose C3′ atom is not linked to a neighboring nucleotide) and a **5′ end** (the end whose C5′ atom is not linked to a neighboring nucleotide).

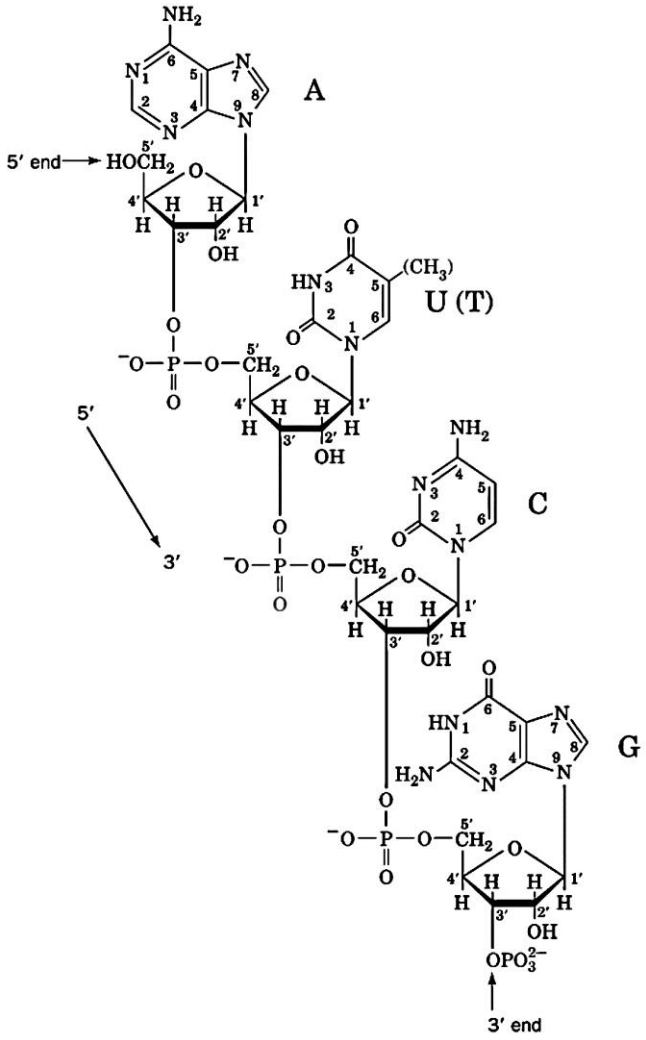


Figure 2. The Tetranucleotide Adenyl-3′,5′-uridyl-3′,5′-cytidyl-3′,5′-guanylyl-3′-phosphate.

DNA has equal numbers of adenine and thymine residues (A = T) and equal numbers of guanine and cytosine residues (G = C). These relationships, known as **Chargaff's rules**, were discovered in the late 1940s by Erwin Chargaff, who first devised reliable quantitative methods for the separation and analysis of DNA hydrolysates. Chargaff also found that the base composition of DNA from a given organism is characteristic of that organism; that is, it is independent of the tissue from which the DNA is taken as well as the organism's age, its nutritional state, or any other environmental factor. The structural basis for Chargaff's rules is that in double-

stranded DNA, G is always hydrogen bonded (forms a **base pair**) with C, whereas A always forms a base pair with T (Fig. 3).

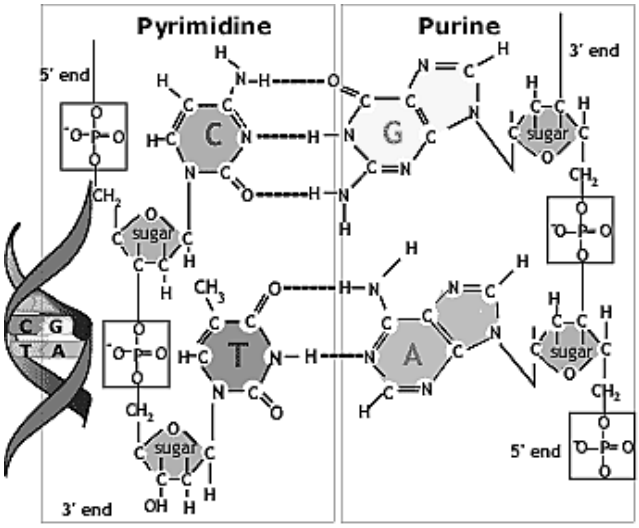
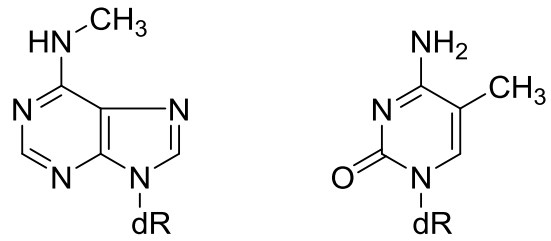


Figure 3. Base Pairs in DNA Structure.

DNA's base composition varies widely among different organisms. It ranges from ~25% to 75% G+C in different species of bacteria. It is, however, more or less constant among related species; for example, in mammals G+C ranges from 39% to 46%.

RNA, which usually occurs as single-stranded molecules, has no apparent constraints on its base composition. However, double-stranded RNA, which comprises the genetic material of certain viruses, also obeys Chargaff's rules (here A base pairs with U in the same way it does with T in DNA). Conversely, single-stranded DNA, which occurs in certain viruses, does not obey Chargaff's rules. On entering its host organism, however, such DNA is replicated to form a double-stranded molecule, which then obeys Chargaff's rules.

Some DNAs contain bases that are chemical derivatives of the standard set. For example, dA and dC in the DNAs of many organisms are partially replaced by **N^6-methyl-dA** and **5-methyl-dC**, respectively.



N^6-Methyl-dA 5-Methyl-dC

The altered bases are generated by the sequence-specific enzymatic modification of normal DNA. The modified DNAs obey Chargaff's rules if the derivatized bases are taken as equivalent to their parent bases. Likewise, many bases in RNAs and, in particular, those in **transfer RNAs** (**tRNAs**) are derivatized.

Nucleic Acids

DNA and RNA are long linear polymers, called nucleic acids, that carry information in a form that can be passed from one generation to the next. These macromolecules consist of a large number of linked nucleotides, each composed of a sugar, a phosphate, and a base. Sugars linked by phosphates form a common backbone, whereas the bases vary among four kinds. Genetic information is stored in the sequence of bases along a nucleic acid chain. The bases have an additional special property: they form specific pairs with one another that are stabilized by hydrogen bonds. The base pairing results in the formation of a double helix, a helical structure consisting of two strands. These base pairs provide a mechanism for copying the genetic information in an existing nucleic acid chain to form a new chain. Although RNA probably functioned as the genetic material very early in evolutionary history, the genes of all modern cells and many viruses are made of DNA. DNA is replicated by the action of DNA polymerase enzymes. These exquisitely specific enzymes copy sequences from nucleic acid templates with an error rate of less than 1 in 100 million nucleotides.

DNA. The determination of the structure of DNA by Watson and Crick in 1953 is often said to mark the birth of modern molecular biology. The **Watson-Crick structure** of DNA is of such importance because, in addition to providing the structure of what is arguably the central molecule of life, it suggested the molecular

mechanism of heredity. Watson and Crick's accomplishment, which is ranked as one of science's major intellectual achievements, tied together the less than universally accepted results of several diverse studies:

1. Chargaff's rules. At the time, the relationships A=T and G=C were quite obscure because their significance was not apparent. In fact, even Chargaff did not emphasize them.

2. Correct tautomeric forms of the bases. X-ray, nuclear magnetic resonance (NMR), and spectroscopic investigations have firmly established that the nucleic acid bases are overwhelmingly in the keto tautomeric forms shown in Table 1. In 1953, however, this was not generally appreciated. Indeed, guanine and thymine were widely believed to be in their enol forms (Fig. 4) because it was thought that the resonance stability of these aromatic molecules would thereby be maximized. Knowledge of the dominant tautomeric forms, which was prerequisite for the prediction of the correct hydrogen bonding associations of the bases, was provided by Jerry Donohue, an office mate of Watson and Crick and an expert on the X-ray structures of small organic molecules.

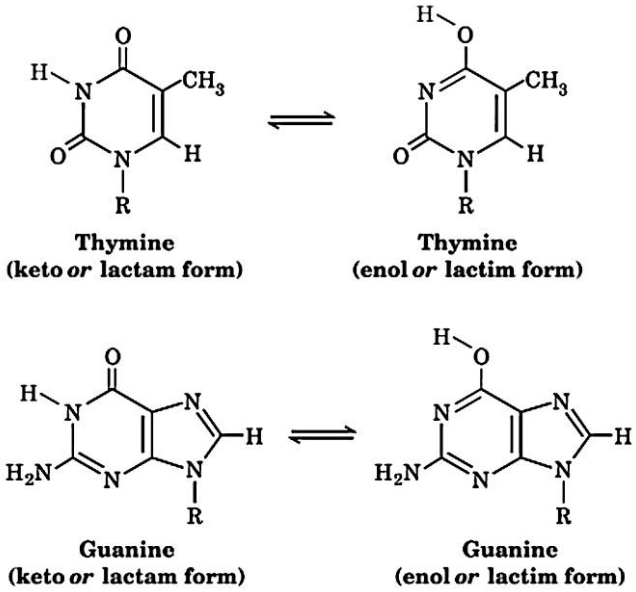


Figure 4. Some Possible Tautomeric Conversions for Bases.

3. Information that DNA is a helical molecule. This was provided by an X-ray diffraction photograph of a DNA fiber taken by Rosalind Franklin. This photograph

enabled Crick, an X-ray crystallographer by training who had earlier derived the equations describing diffraction by helical molecules, to deduce (a) that DNA is a helical molecule and (b) that its planar aromatic bases form a stack of parallel rings which is parallel to the fiber axis.

This information only provided a few crude landmarks that guided the elucidation of the DNA structure. It mostly sprang from Watson and Crick's imaginations through model building studies. Once the Watson-Crick model had been published, however, its basic simplicity combined with its obvious biological relevance led to its rapid acceptance. Later investigations have confirmed the essential correctness of the Watson-Crick model, although its details have been modified.

The Watson-Crick Structure: B-DNA. Fibers of DNA assume the so-called **B conformation**, as indicated by their X-ray diffraction patterns, when the counterion is an alkali metal such as Na^+ and the relative humidity is >92%. **B-DNA** is regarded as the **native** (biologically functional) form of DNA because, for example, its X-ray pattern resembles that of the DNA in intact sperm heads.

The Watson-Crick structure of B-DNA has the following major features:

1. It consists of two polynucleotide strands that wind about a common axis with a right-handed twist to form an ~20-Å-diameter double helix (Fig. 5). The two strands are antiparallel (run in opposite directions) and wrap around each other such that they cannot be separated without unwinding the helix. The bases occupy the core of the helix and the sugar-phosphate chains are coiled about its periphery, thereby minimizing the repulsions between charged phosphate groups.

2. The planes of the bases are nearly perpendicular to the helix axis. Each base is hydrogen bonded to a base on the opposite strand to form a planar base pair (Fig. 5). It is these hydrogen bonding interactions, a phenomenon known as **complementary base pairing**, that result in the specific association of the two chains of the double helix.

3. The "ideal" B-DNA helix has 10 base pairs (**bp**) per turn (a helical twist of 36° per bp) and, since the aromatic bases have van der Waals thicknesses of 3.4 Å

and are partially stacked on each other (**base stacking**, Fig. 5), the helix has a **pitch** (rise per turn) of 34 Å.

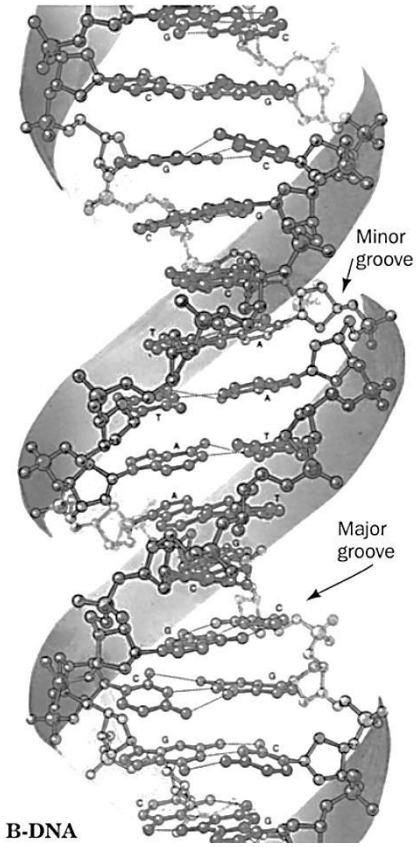


Figure 5. Three-dimensional Structure of B-DNA.

The most remarkable feature of the Watson-Crick structure is that it can accommodate only two types of base pairs: Each adenine residue must pair with a thymine residue and vice versa, and each guanine residue must pair with a cytosine residue and vice versa. The geometries of these A-T and G-C base pairs, the so-called Watson-Crick base pairs, are shown in Figure 6. It can be seen that both of these base pairs are interchangeable in that they can replace each other in the double helix without altering the positions of the sugar-phosphate backbone's C1′ atoms. Likewise, the double helix is undisturbed by exchanging the partners of a Watson-Crick base pair, that is, by changing a G-C to a C-G or an A-T to a T-A. In contrast, any other combination of bases (e.g., A-G or A-C) would significantly distort the double helix since the formation of a non-Watson-Crick base pair would require considerable reorientation of the sugar-phosphate chain.

B-DNA has two deep exterior grooves that wind between its sugar-phosphate chains as a consequence of the helix axis passing through the approximate center of each base pair. However, the grooves are of unequal size (Fig. 5) because (1) the top edge of each base pair, as drawn in Figure 6, is structurally distinct from the bottom edge; and (2) the deoxyribose residues are asymmetric. The minor groove exposes that edge of a base pair from which its C1′ atoms extend, whereas the major groove exposes the opposite edge of each base pair.

Although B-DNA is, by far, the most prevalent form of DNA in the cell, double helical DNAs and RNAs can assume several distinct structures.

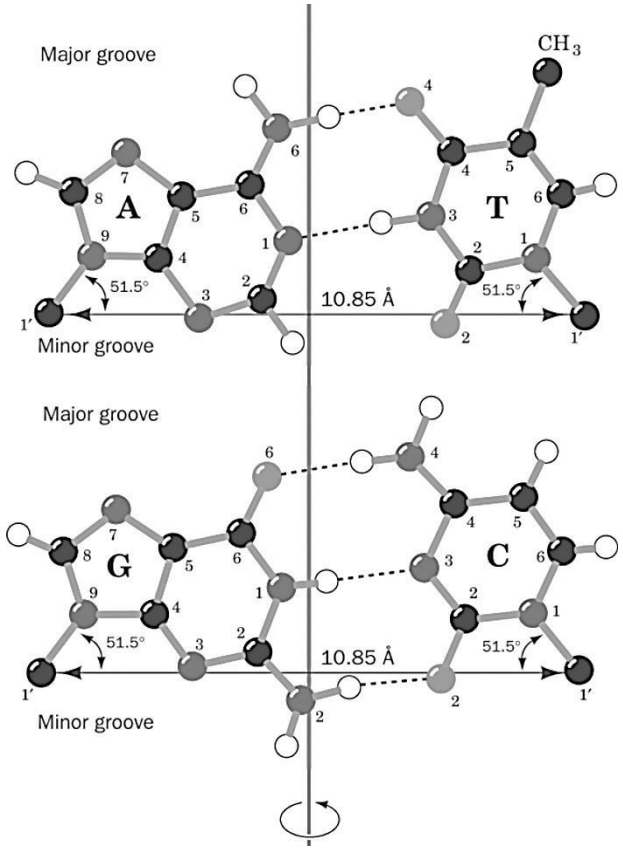


Figure 6. Watson-Crick Base Pairs.

Other Nucleic Acid Helices. X-ray fiber diffraction studies, revealed that nucleic acids are conformationally variable molecules. Indeed, double helical DNA and RNA can assume several distinct structures that vary with such factors as the humidity and the identities of the cations present, as well as with base sequence. For example, fibers of B-DNA form in the presence of alkali metal ions such as Na^+ when the relative humidity is 92%.

When the relative humidity is reduced to 75%, B-DNA undergoes a reversible conformational change to the so-called A form. Fiber X-ray studies indicate that A-DNA forms a wider and flatter right-handed helix than does B-DNA (Table 2). A-DNA has 11.6 bp per turn and a pitch of 34 Å, which gives A-DNA an axial hole (Fig. 7). A-DNA's most striking feature, however, is that the planes of its base pairs are tilted 20° with respect to the helix axis. Since its helix axis passes "above" the major groove side of the base pairs rather than through them as in B-DNA, A-DNA has a deep major groove and a very shallow minor groove; it can be described as a flat ribbon wound around a 6-Å-diameter cylindrical hole. Most self-complementary oligonucleotides of <10 base pairs, for example, d(GGCCGGCC) and d(GGTATACC), crystallize in the A-DNA conformation. Like B-DNA, these molecules exhibit considerable sequence-specific conformational variation although the degree of variation is less than that in B-DNA.

Table 2. Structural Features of Ideal A-, B-, and Z-DNA.

	A-DNA	**B-DNA**	**Z-DNA**
Helical sense	Right-handed	Right-handed	Left-handed
Diameter	~26 Å	~20 Å	~18 Å
Base pairs per helical turn	11.6	10	12 (6 dimers)
Helical twist per base pair	31°	36°	9° for pyrimidine-purine steps; 51° for purine-pyrimidine steps
Helix pitch (rise per turn)	34 Å	34 Å	44 Å
Helix rise per base pair	2.9 Å	3.4 Å	7.4 Å per dimer
Base tilt normal to the helix axis	20°	6°	7°
Major groove	Narrow and deep	Wide and deep	Flat
Minor groove	Wide and shallow	Narrow and deep	Narrow and deep

A-DNA has, so far, been observed in only three biological contexts: at the cleavage center of topoisomerase II, at the active site of DNA polymerase, and in certain Gram-positive bacteria that have undergone sporulation (the formation, under environmental stress, of resistant although dormant cell types known as **spores**; a sort of biological lifeboat). Such spores contain a high proportion (20%) of **small acid-soluble spore proteins** (**SASPs**). Some of these SASPs induce B-DNA to assume the A form, at least in vitro. The DNA in bacterial spores exhibits a resistance to UV-

induced damage that is abolished in mutants that lack these SASPs. This occurs because the B→A conformation change inhibits the UV-induced covalent cross-linking of pyrimidine bases, in part by increasing the distance between successive pyrimidines.

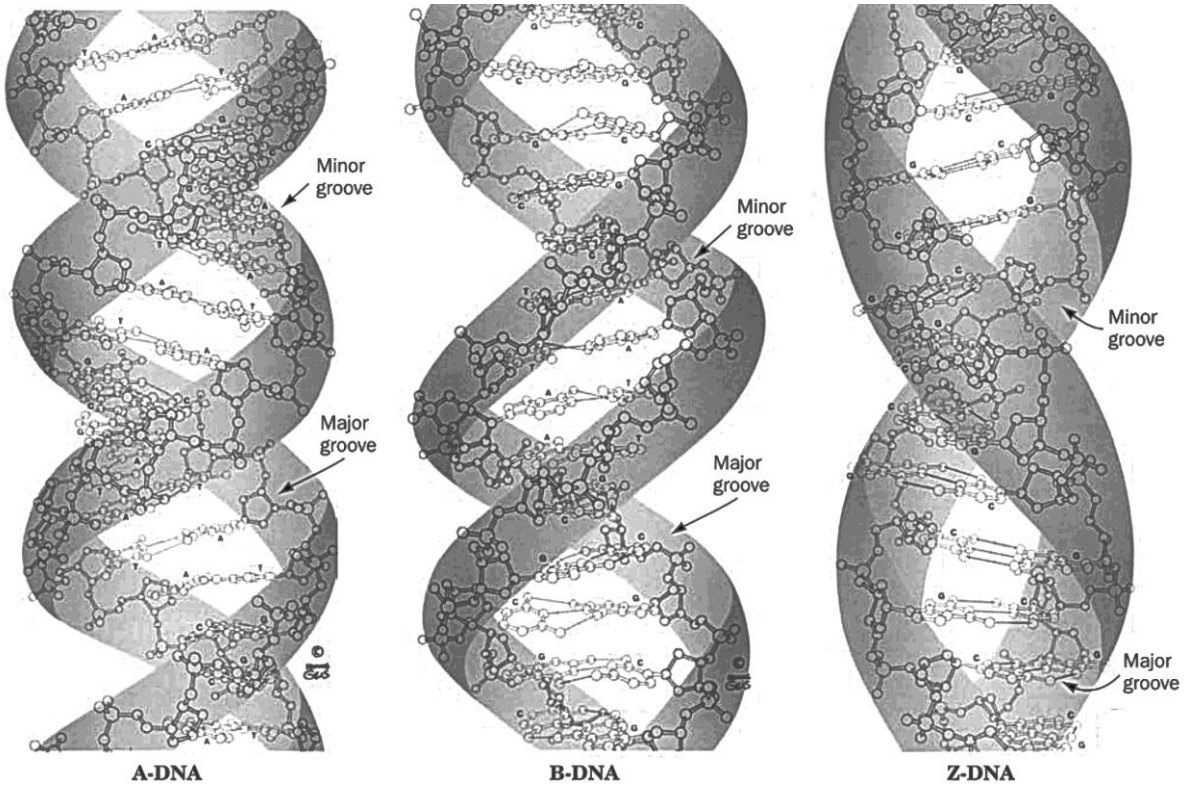


Figure 7. Structures of A-, B-, and Z-DNAs.

Occasionally, a seemingly well-understood or at least familiar system exhibits quite unexpected properties. Over 25 years after the discovery of the Watson-Crick structure, the crystal structure determination of the self-complementary hexanucleotide d(CGCGCG) by Andrew Wang and Alexander Rich revealed, quite surprisingly, a *left-handed double helix* (Fig. 7, Table 2). A similar helix is formed by d(CGCATGCG). This helix, which has been dubbed **Z-DNA**, *has 12 Watson-Crick base pairs per turn, a pitch of 44 Å*, and, in contrast to A-DNA, *a deep minor groove and no discernible major groove* (its helix axis passes "below" the minor groove side of its base pairs). Z-DNA therefore resembles a left-handed drill bit in appearance. The line joining successive phosphorus atoms on a polynucleotide strand of Z-DNA therefore follows a zigzag path around the helix (hence the name Z-DNA) rather than a smooth curve as it does in A- and B-DNAs.

A high salt concentration stabilizes Z-DNA relative to B-DNA by reducing the otherwise increased electrostatic repulsions between closest approaching phosphate groups on opposite strands (8 Å in Z-DNA vs 12 Å in B-DNA). The methylation of cytosine residues at C5, a common biological modification, also promotes Z-DNA formation since a hydrophobic methyl group in this position is less exposed to solvent in Z-DNA than it is in B-DNA.

Does Z-DNA have any biological function? Rich has proposed that the reversible conversion of specific segments of B-DNA to Z-DNA under appropriate circumstances acts as a kind of switch in regulating genetic expression, and there are indications that it transiently forms behind actively transcribing RNA polymerase. It was nevertheless surprisingly difficult to prove the *in vivo* existence of Z-DNA. A major difficulty was demon-strating that a particular probe for detecting Z-DNA, for example, a Z-DNA-specific antibody, does not in itself cause what would otherwise be B-DNA to assume the Z conformation – a kind of biological uncertainty principle (the act of measurement inevitably disturbs the system being measured). However, Rich has discovered several proteins that specifically bind Z-DNA, including a family of Z-DNA-binding protein domains named Zα. The existence of these proteins strongly suggests that Z-DNA does, in fact, exist *in vivo*.

The DNA molecules in human chromosomes are linear. However, electron microscopic and other studies have shown that intact DNA molecules from some other organisms are circular (Fig. 8). The term circular refers to the continuity of the DNA chains, not to their geometric form. DNA molecules inside cells necessarily have a very compact shape. Note that the E. coli chromosome, fully extended, would be about 1000 times as long as the greatest diameter of the bacterium.

A closed DNA molecule has a property unique to circular DNA. The axis of the double helix can itself be twisted or supercoiled into a superhelix (Fig. 9). Supercoiling is biologically important for two reasons. First, a supercoiled DNA molecule has a more compact shape than does its relaxed counterpart. Second, supercoiling may hinder or favor the capacity of the double helix to unwind and thereby affect the interactions between DNA and other molecules.

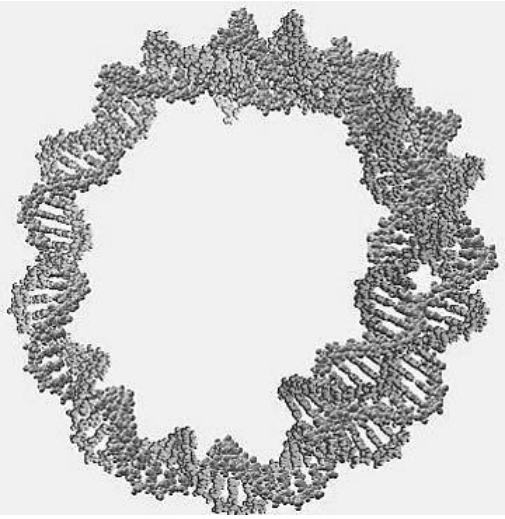


Figure 8. Circular DNA.

In general, adenine base containing DNA tracts are rigid and straight. **Bent** conformation of DNA occurs when A-tracts are replaced by other bases or a collapse of the helix into the minor groove of A-tract. Bending in DNA structure has also been reported due to photochemical damage or mispairing of bases Certain antitumor drugs (e.g. cisplatin) produce bent structure in DNA. Such changed structure can take up proteins that damage the DNA.

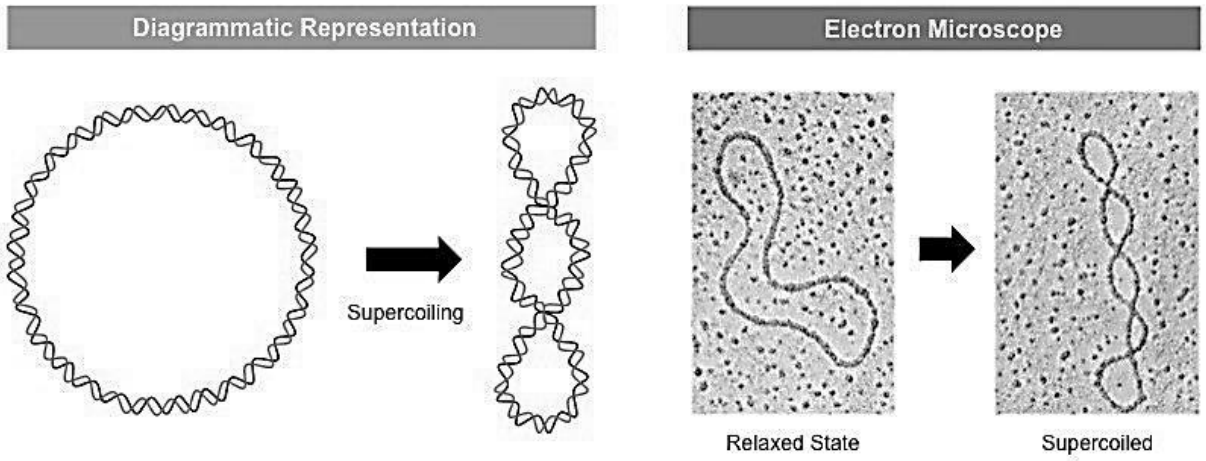


Figure 9. Supercoiling in a Prokaryotic Cell.

Triple-stranded DNA formation may occur due to additional hydrogen bonds between the bases. Thus, a thymine can selectively form two **Hoogsteen hydrogen bonds** to the adenine of A-T pair to form **T-A-T**. Likewise, a protonated cytosine can also form two hydrogen bonds with guanine of C-C pairs that results in C-G-C.

Triple-helical structure is less stable than double helix. This is due to the fact that the three negatively charged backbone strands in triple helix results in an increased electrostatic repulsion.

Polynucleotides with very high contents of guanine can form a novel tetrameric structure called **G-quartets**. These structures are planar and are connected by Hoogsteen hydrogen bonds. Antiparallel four-stranded DNA structures, referred to as **G-tetraplexes** have also been reported. The ends of eukaryotic chromosomes namely telomeres are rich in guanine, and therefore form G-tetraplexes. In recent years, telomeres have become the targets for anticancer chemotherapies. G-tetraplexes have been implicated in the recombination of immunoglobulin genes, and in dimerization of double-stranded genomic RNA of the human immunodeficiency virus (HlV).

RNA. RNA molecules are synthesized in a process referred to as transcription. During transcription, new RNA molecules are produced by a mechanism similar to DNA synthesis, that is, through complementary base pair formation. The sequence of bases in RNA is therefore specified by the base sequence in one of the two strands in DNA. For example, the DNA sequence 5′-CCGATTACG-3′ is transcribed into the RNA sequence 3′-GGCUAAUGC-5′. Complementary DNA and RNA sequences are antiparallel. RNA molecules differ from DNA in the following ways:

1. The sugar moiety of RNA is ribose instead of deoxyribose in DNA.

2. The nitrogenous bases in RNA differ somewhat from those observed in DNA. Instead of thymine, RNA molecules use uracil. In addition, the bases in some RNA molecules are modified by a variety of enzymes (e.g., methylases, thiolases, and deaminases).

3. In contrast to the double helix of DNA, RNA exists as a single strand. For this reason, RNA can coil back on itself and form unique and often quite complex three-dimensional structures (Fig. 10). The shape of these structures is determined by complementary base pairing by specific RNA sequences, as well as by base stacking. In addition, the 2′-OH of ribose can form hydrogen bonds with nearby molecular groups. Because RNA is single stranded, Chargaff's rules do not apply. An RNA molecule’s contents of A and U, as well as C and G, are usually not equal.

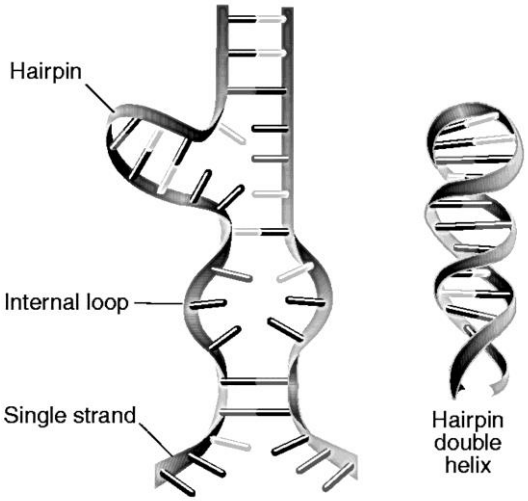
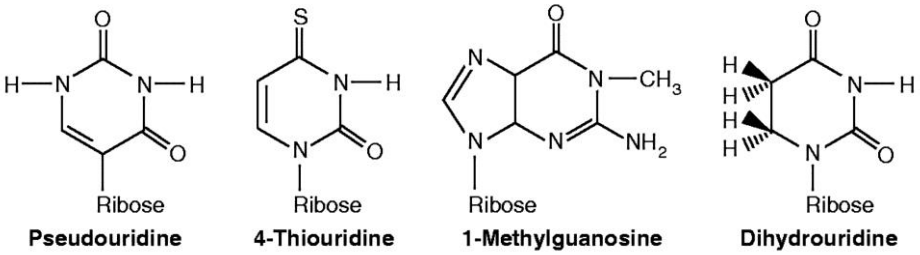


Figure 10. Secondary Structure of RNA.

The most prominent types of RNA are transfer RNA, ribosomal RNA, and messenger RNA.

Transfer RNA. Transfer RNA (tRNA) molecules transport amino acids to ribosomes for assembly into proteins. Comprising about 15 % of cellular RNA the average length of a tRNA molecule is 75 nucleotides. Because each tRNA molecule becomes bound to a specific amino acid, cells possess at least one type of tRNA for each of the 20 amino acids commonly found in protein. The three-dimensional structure of tRNA molecules, which resembles a warped cloverleaf (Fig. 11), results primarily from extensive intrachain base pairing. tRNA molecules contain a variety of modified bases. Examples include pseudouridine, 4-thiouridine, 1-methylguanosine, and dihydrouridine:



The structure of tRNA allows it to perform two critical functions involving the most important structural components: the 3′-terminus and the anticodon loop.

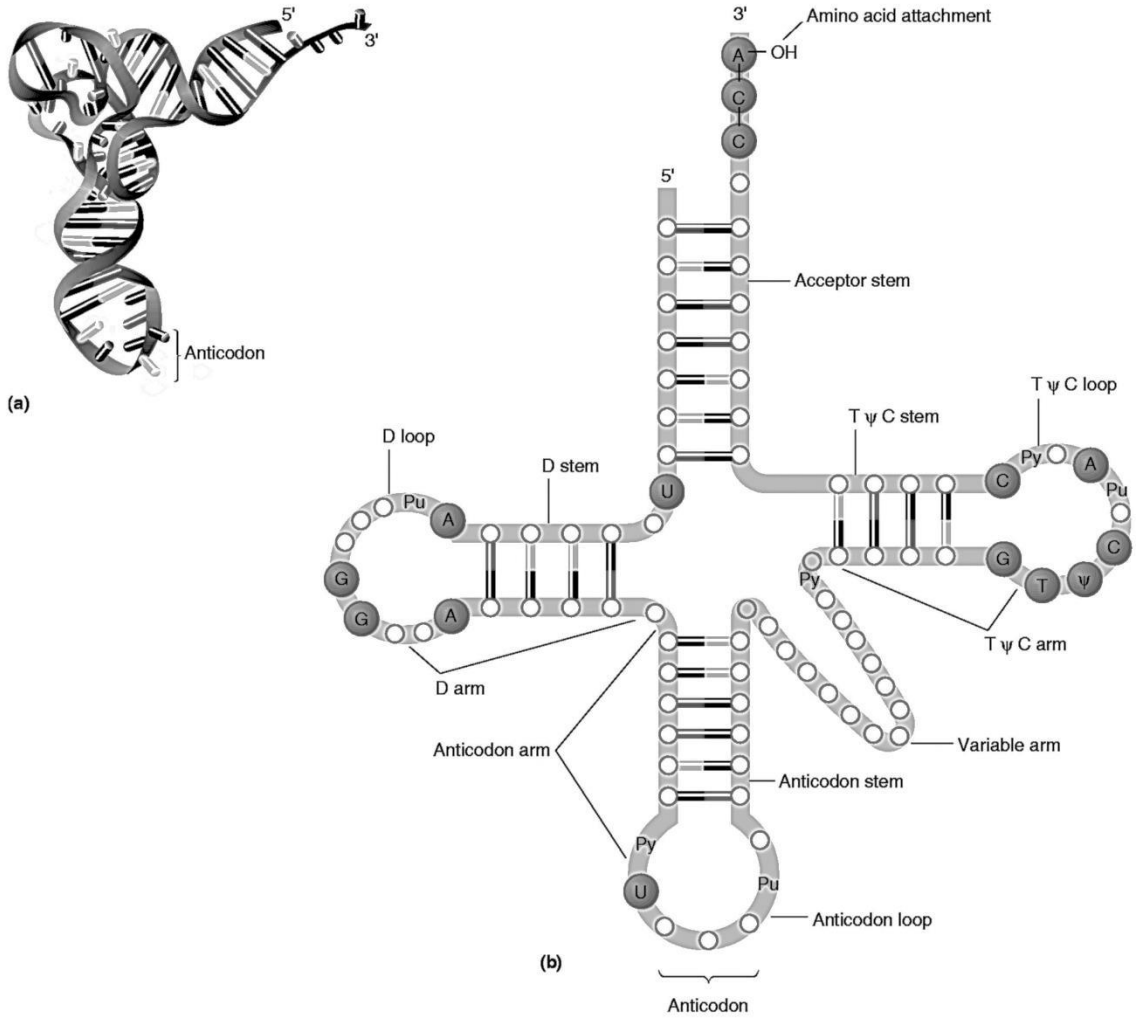


Figure 11. Transfer RNA. (a) Three-dimensional structure of a tRNA molecule. (b) A schematic view of a tRNA molecule.

The *3′-terminus* forms a covalent bond to a specific amino acid. The *anticodon loop* contains a three-base-pair sequence that is complementary to the DNA triplet code for the specific amino acid. The conformational relationship between the 3′-terminus and the anticodon loop allows the tRNA to align its attached amino acid properly during protein synthesis. tRNAs also possess three other prominent structural features, referred to as the *D loop*, the *TψC loop* (ψ is an abbreviation for the modified base pseudouridine), and the variable loop. The function of these structures is unknown, but they are presumably related to the alignment of tRNA within the ribosome and/or the binding of a tRNA to the enzyme that catalyzes the attachment of the appropriate amino acid. The D loop is so named because it contains dihydrouridine. Similarly, the TψC loop contains the base sequence thymine,

pseudouridine, and cytosine. tRNAs can be classified on the basis of the length of their *variable loop*. The majority (approximately 80 %) of tRNAs have variable loops with four to five nucleotides, whereas the others have variable loops with as many as 20 nucleotides.

Ribosomal RNA. Ribosomal RNA (rRNA) is the most abundant form of RNA in living cells. In most cells, rRNA constitutes approximately 80% of the total RNA. The secondary structure of rRNA is extraordinarily complex. Although there are species differences in the primary nucleotide sequences of rRNA, the overall three-dimensional structure of this class of molecules is conserved. As its name suggests, rRNA is a component of ribosomes.

Ribosomes are cytoplasmic structures that synthesize proteins. Because they are composed of both protein and rRNA, the ribosomes are sometimes described as ribonucleoprotein bodies. The ribosomes of prokaryotes and eukaryotes are similar in shape and function, although they differ in size and their chemical composition. Several different kinds of rRNA and protein are found in each type of ribosomal subunit. The large ribosomal subunit of E. coli, for example, contains 5 S and 23 S rRNAs and 34 polypeptides. The small ribosomal subunit of E. coli contains a 16 S rRNA and 21 polypeptides. A typical large eukaryotic ribosomal subunit contains three rRNAs (5 S, 5.8 S, and 28 S) and 49 polypeptides; the small subunit contains an 18 S rRNA and approximately 30 polypeptides. The functions of the rRNA and polypeptides in ribosomes are poorly understood and are being investigated.

Messenger RNA. As its name suggests, messenger RNA (mRNA) is the carrier of genetic information from DNA for the synthesis of protein. mRNA molecules, which typically constitute approximately 5 % of cellular RNA, vary considerably in size. For example, mRNA from E. coli varies from 500 to 6000 nucleotides.

Prokaryotic mRNA and eukaryotic mRNA differ in several respects. First, many prokaryotic mRNAs are polycistronic, that is, they contain coding information for several polypeptide chains. In contrast, eukaryotic mRNA typically codes for a single polypeptide and is therefore referred to as monocistronic. A **cistron** is a DNA

sequence that contains the coding information for a polypeptide and several signals that are required for ribosome function. Second, prokaryotic and eukaryotic mRNAs are processed differently. In contrast to prokaryotic mRNAs, which are translated into protein by ribosomes during or immediately after they are synthesized, eukaryotic mRNAs are modified extensively. These modifications include capping (linkage of 7-methylguanosine to the 5′-terminal residue), splicing (removal of introns), and the attachment of an adenylate polymer referred to as a poly A tail.

Heterogeneous RNA and small nuclear RNA. Heterogeneous RNA and small nuclear RNA play complementary roles in eukaryotic cells. Heterogeneous nuclear RNA (hnRNA) molecules are the primary transcripts of DNA and are the precursors of mRNA. HnRNA is processed by splicing and modifications to form mRNA. Splicing is the enzymatic removal of the introns from the primary transcripts. A class of small nuclear RNA (snRNA) molecules (containing between 90 and 300 nucleotides), which are complexed with several proteins to form small nuclear ribonucleoprotein particles (snRNP or snurps), are involved in splicing activities and other forms of RNA processing.

Chromatin Organization in Nucleus

The fact that DNA in eukaryotic chromosomes is not bare. Instead, eukaryotic DNA is tightly bound to a group of small basic proteins called *histones*. Histones constitute half the mass of a eukaryotic chromosome. The entire complex of a cell's DNA and associated protein is called *chromatin*. Five major histones are present in chromatin: four histones, called *H2A*, *H2B*, *H3*, and *H4*, associate with one another; the other histone is called *H1*. Histones have strikingly basic properties because a quarter of the residues in each histone are either arginine or lysine.

Chromatin is made up of repeating units, each containing 200 bp of DNA and two copies each of H2A, H2B, H3, and H4, called the *histone octamer*. These repeating units are known as nucleosomes. Strong support for this model comes from the results of a variety of experiments, including observations of appropriately prepared samples of chromatin viewed by electron microscopy. Chromatin viewed

with the electron microscope has the appearance of beads on a string; each bead has a diameter of approximately 100 Å. Partial digestion of chromatin with DNase yields the isolated beads. These particles consist of fragments of DNA about 200 bp in length bound to the eight histones. More-extensive digestion yields a shorter DNA fragment of 145 bp bound to the histone octamer. The smaller complex formed by the histone octamer and the 145-bp DNA fragment is the *nucleosome core particle*. The DNA connecting core particles in undigested chromatin is called *linker DNA*. Histone Hl binds, in part, to the linker DNA (Fig. 12).

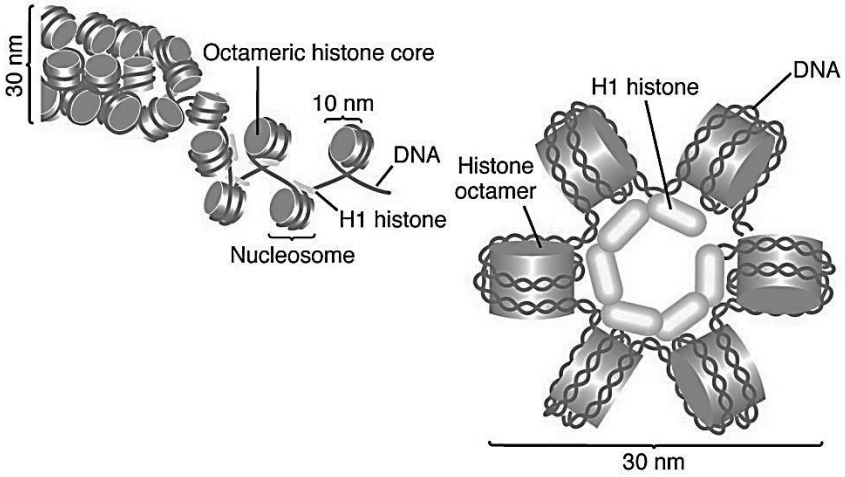


Figure 12. Chromatin organization.

The overall structure of the nucleosome was revealed through electron microscopic and x-ray crystallographic studies pioneered by Aaron Klug and his colleagues. More recently, the three-dimensional structure of a reconstituted nucleosome core was determined to higher resolution by x-ray diffraction methods. As was shown by Evangelos Moudrianakis, the four types of histone that make up the protein core are homologous and similar in structure. The eight histones in the core are arranged into a $(H3)_2(H4)_2$ tetramer and a pair of H2A-H2B dimers. The tetramer and dimers come together to form a left-handed superhelical ramp around which the DNA wraps. In addition, each histone has an amino-terminal tail that extends out from the core structure. These tails are flexible and contain a number of lysine and arginine residues. As we shall see, covalent modifications of these tails play an essential role in modulating the affinity of the histones for DNA and other properties.