**NUCLEOTIDES AND**

**NUCLEIC ACIDS**

**Nucleic Acid Structure**

The discovery of the structure of DNA by Watson and Crick in 1953 was a momentous event in science, an event that gave rise to entirely new disciplines and influenced the course of many established ones. Our present understanding of the storage and utilization of a cell’s genetic information is based on work made possible by this discovery, and an outline of how genetic information is processed by the cell is now a prerequisite for the discussion of any area of biochemistry. Here, we concern ourselves with DNA structure itself, the event that led to its discovery, and more recent refinements in our understanding. RNA structure is also introduced.

As in the case of protein structure, it is sometimes useful to describe nucleic acid structure in terms of hierarchical levels of complexity (primary, secondary, tertiary). The primary structure of a nucleic acid is its covalent structure and nucleotide sequence. Any regular, stable structure taken up by some or all of the nucleotides in a nucleic acid can be referred to as secondary structure. All structures considered in the remainder of this chapter fall under the heading of secondary structure. The complex folding of large chromosomes within eukaryotic chromatin and bacterial nucleoids is generally considered tertiary structure.

When Watson and Crick constructed their model, they had to decide at the outset whether the strands of DNA should be **parallel** or **antiparallel**—whether their 5',3'-phosphodiester bonds should run in the same or opposite directions. An antiparallel orientation produced the most convincing model, and later work with DNA polymerases provided experimental evidence that the strands are indeed antiparallel, a finding ultimately confirmed by x-ray analysis.

To account for the periodicities observed in the x- ray diffraction patterns of DNA fibers, Watson and Crick manipulated molecular models to arrive at a structure

**Watson - Crick Model for the structure of DNA.** The original model proposed by Watson and Crick had 10 base pairs, or 34 Å (3.4 nm), per turn of the helix; subsequent measurements revealed 10.5 base pairs, or 36 Å (3.6 nm), per turn. **(a)** Schematic representation, showing dimensions of the helix. **(b)** Stick representation showing the backbone and stacking of the bases. **(c)** Space-filling model.

In which the vertically stacked bases inside the double helix would be 3.4 Å apart; the secondary repeat distance of about 34 Å was accounted for by the presence of 10 base pairs in each complete turn of the double helix. In aqueous solution the structure differs slightly from that in fibers, having 10.5 base pairs per helical turn.

The two antiparallel polynucleotide chains of double-helical DNA are not identical in either base sequence or composition. Instead they are **complementary** to each other. Wherever adenine occurs in one chain, thymine is found in the other; similarly, wherever guanine occurs in one chain, cytosine is found in the other.

The DNA double helix, or duplex, is held together by two forces, as described earlier: hydrogen bonding between complementary base pairs and base-stacking interactions. The complementarity between the DNA strands is attributable to the hydrogen bonding between base pairs. The base-stacking interactions, which are largely nonspecific with respect to the identity of the stacked bases, make the major contribution to the stability of the double helix.

The important features of the double-helical model of DNA structure are supported by much chemical and

**Complementarity of strands in the DNA double helix.** The complementary antiparallel strands of DNA follow the pairing rules proposed by Watson and Crick. The base-paired antiparallel strands differ in base composition: the left strand has the composition A3 T2 G1 C3; the right, A2 T3 G3 C1. They also differ in sequence when each chain is read in the 5' n 3' direction. Note the base equivalences: A = T and G = C in the duplex.

**Replication of DNA as suggested by Watson and Crick.**

The preexisting or “parent” strands become separated, and each is the template for biosynthesis of a complementary “daughter” strand (in red).

Biological evidence. Moreover, the model immediately suggested a mechanism for the transmission of genetic information. The essential feature of the model is the complementarity of the two DNA strands. As Watson and Crick were able to see, well before confirmatory data became available, this structure could logically be replicated by (1) separating the two strands and (2) synthesizing a complementary strand for each. Because nucleotides in each new strand are joined in a sequence specified by the base-pairing rules stated above, each preexisting strand functions as a template to guide the synthesis of one complementary strand. These expectations were experimentally confirmed, inaugurating a revolution in our understanding of biological inheritance.

*SUMMARY* Nucleic Acid Structure

■ Many lines of evidence show that DNA bears genetic information. In particular, the Avery- MacLeod-McCarty experiment showed that DNA isolated from one bacterial strain can enter and transform the cells of another strain, endowing it with some of the inheritable characteristics of the donor. The Hershey-Chase experiment showed that the DNA of a bacterial virus, but not its protein coat, carries the genetic message for replication of the virus in a host cell.

■ Putting together much published data, Watson and Crick postulated that native DNA consists of two antiparallel chains in a right-handed double-helical arrangement. Complementary base pairs, A=T and G=C, are formed by hydrogen bonding within the helix. The base pairs are stacked perpendicular to the long axis of the double helix, 3.4 Å apart, with 10.5 base pairs per turn.

■ DNA can exist in several structural forms. Two variations of the Watson-Crick form, or B-DNA, are A- and Z-DNA. Some sequence-dependent structural variations cause bends in the DNA molecule. DNA strands with appropriate sequences can form hairpin/cruciform structures or triplex or tetraplex DNA.

■ Messenger RNA transfers genetic information from DNA to ribosomes for protein synthesis. Transfer RNA and ribosomal RNA are also involved in protein synthesis. RNA can be structurally complex; single RNA strands can be folded into hairpins, double-stranded regions, or complex loops.

**The Information in DNA Is Encoded in Digital Form:**

In this digital age, we are accustomed to electronic information encoded in the form of extremely long arrays of just two digits: ones (1s) and zeros (0s). DNA uses four digits to encode biological information: A, C, G, and T. A significant feature of the DNA double helix is that virtually any base sequence (encoded information) is possible: Other than the base-pairing rules, no structural constraints operate to limit the potential sequence of bases in DNA.

DNA contains two kinds of information:

1. The base sequences of genes that encode the amino acid sequences of proteins and the nucleotide sequences of functional RNA molecules such as rRNA and tRNA.

2. The gene regulatory networks that control the expression of protein-encoding (and functional RNA-encoding) genes.

**The Chemical differences between DNA and RNA have biological significance:**

Two fundamental chemical differences distinguish DNA from RNA:

1. DNA contains 2-deoxyribose instead of ribose.

2. DNA contains thymine instead of uracil.

**Nucleases Differ in Their Specificity for Different Forms of Nucleic Acid:**

Nucleases play an indispensable role in the cellular breakdown of nucleic acids and salvage of their constituent parts. Nucleases also participate in many other cellular functions, including (1) aspects of DNA metabolism, such as replication and repair; (2) aspects of RNA metabolism, such as splicing of the primary gene transcript, processing of mRNA, and RNAi; (3) rearrangements of genetic material, such as recombination and transposition; (4) host defense mechanisms against foreign nucleic acid molecules; and (5) the immune response, through assembly of immunoglobulin genes. Some nucleases are not even proteins but instead are catalytic RNA molecules. Like most enzymes, nucleases exhibit selectivity or specificity for the nature of the substance on which they act. That is, some nucleases act only on DNA **(DNases),** whereas others are specific for RNA (the **RNases**). Still others are nonspecific and are referred to simply as **nucleases**. Nucleases may also show specificity for only single-stranded nucleic acids or may act only on double helices. Some display a decided preference for acting only at certain bases in a polynucleotide, or as we shall see for restriction endonucleases, act only at a particular nucleotide sequence four to eight nucleotides (or more) in length. To the molecular biologist, nucleases are the surgical tools for the dissection and manipulation of nucleic acids in the laboratory.

**Nucleic Acid Chemistry**

To understand how nucleic acids function, we must understand their chemical properties as well as their structures. The role of DNA as a repository of genetic information depends in part on its inherent stability. The chemical transformations that do occur are generally very slow in the absence of an enzyme catalyst. The long-term storage of information without alteration is so important to a cell, however, that even very slow reactions that alter DNA structure can be physiologically significant. Processes such as carcinogenesis and aging may be intimately linked to slowly accumulating, irreversible alterations of DNA. Other, nondestructive alterations also occur and are essential to function, such as the strand separation that must precede DNA replication or transcription. In addition to providing insights into physiological processes, our understanding of nucleic acid chemistry has given us a powerful array of technologies that have applications in molecular biology, medicine, and forensic science. We now examine the chemical properties of DNA and some of these technologies.

**Double-Helical DNA and RNA Can Be Denatured**

Solutions of carefully isolated, native DNA are highly viscous at pH 7.0 and room temperature (25 °C). When such a solution is subjected to extremes of pH or to temperatures above 80 °C, its viscosity decreases sharply, indicating that the DNA has undergone a physical change. Just as heat and extremes of pH denature globular proteins, they also cause denaturation, or melting, of double-helical DNA. Disruption of the hydrogen bonds between paired bases and of base stacking causes unwinding of the double helix to form two single strands, completely separate from each other along the entire length or part of the length ( partial denaturation) of the molecule. No covalent bonds in the DNA are broken.

Renaturation of a DNA molecule is a rapid one-step process, as long as a double-helical segment of a dozen or more residues still unites the two strands. When the temperature or pH is returned to the range in which most organisms live, the unwound segments of the two strands spontaneously rewind, or anneal, to yield the intact duplex. However, if the two strands are completely separated, renaturation occurs in two steps. In the first, relatively slow step, the two strands “find” each other by random collisions and form a short segment of complementary double helix. The second step is much faster: the remaining unpaired bases successively come into register as base pairs, and the two strands “zipper” themselves together to form the double helix.

The close interaction between stacked bases in a nucleic acid has the effect of decreasing its absorption of UV light relative to that of a solution with the same concentration of free nucleotides, and the absorption is decreased further when two complementary nucleic acids strands are paired. This is called the hypochromic effect. Denaturation of a double-stranded nucleic acid produces the opposite result: an increase in absorption.

**Nucleotides and Nucleic Acids Undergo Non-enzymatic Transformations**

Purines and pyrimidines, along with the nucleotides of which they are a part, undergo a number of spontaneous alterations in their covalent structure. The rate of these reactions is generally very slow, but they are physiologically significant because of the cell’s very low tolerance for alterations in its genetic information. Alterations in DNA structure that produce permanent changes in the genetic information encoded therein are called mutations, and much evidence suggests an intimate link between the accumulation of mutations in an individual organism and the processes of aging and carcinogenesis.

Several nucleotide bases undergo spontaneous loss of their exocyclic amino groups (deamination). For example, under typical cellular conditions, deamination of cytosine (in DNA) to uracil occurs in about one of every 107 cytidine residues in 24 hours. This corresponds to about 100 spontaneous events per day, on average, in a mammalian cell. Deamination of adenine and guanine occurs at about 1/100th this rate.

The slow cytosine deamination reaction seems innocuous enough, but is almost certainly the reason why DNA contains thymine rather than uracil. The product of cytosine deamination (uracil) is readily recognized as foreign in DNA and is removed by a repair system. If DNA normally contained uracil, recognition of uracils resulting from cytosine deamination would be more difficult, and unrepaired uracils would lead to permanent sequence changes as they were paired with adenines during replication. Cytosine deamination would gradually lead to a decrease in G≡C base pairs and an increase in A=U base pairs in the DNA of all cells. Over the millennia, cytosine deamination could eliminate G≡C base pairs and the genetic code that depends on them. Establishing thymine as one of the four bases in DNA may well have been one of the crucial turning points in evolution, making the long-term storage of genetic information possible.

Another important reaction in deoxy ribonucleotides is the hydrolysis of the N-{3-glycosyl bond be- tween the base and the pentose. This occurs at a higher rate for purines than for pyrimidines. As many as one in 105 purines (10,000 per mammalian cell) are lost from DNA every 24 hours under typical cellular conditions. Depurination of ribonucleotides and RNA is much slower and generally is not considered physiologically significant. In the test tube, loss of purines can be accelerated by dilute acid. Incubation of DNA at pH 3 causes selective removal of the purine bases, resulting in a derivative called a purinic acid.



**Some well-characterized non-enzymatic reactions of nucleotides. (a) Deamination reactions. Only the base is shown.** **b) Depurination, in which a purine is lost by hydrolysis of the N-{3- glycosyl bond. The deoxyribose remaining after depurination is readily converted from the {3-furanose to the aldehyde form.**

Other reactions are promoted by radiation. UV light induces the condensation of two ethylene groups to form a cyclobutane ring. In the cell, the same reaction between adjacent pyrimidine bases in nucleic acids forms cyclobutane pyrimidine dimers. This happens most frequently between adjacent thymidine residues on the same DNA strand. A second type of pyrimidine dimer, called a 6-4 photoproduct, is also formed during UV irradiation. Ionizing radiation (x-rays and gamma rays) can cause ring opening and fragmentation of bases as well as breaks in the covalent back- bone of nucleic acids.

Virtually all forms of life are exposed to energy-rich radiation capable of causing chemical changes in DNA. Near-UV radiation (with wavelengths of 200 to 400 nm), which makes up a significant portion of the solar spectrum, is known to cause pyrimidine dimer formation and other chemical changes in the DNA of bacteria and of human skin cells. We are subject to a constant field of ionizing radiation in the form of cosmic rays, which can penetrate deep into the earth, as well as radiation emitted from radioactive elements, such as radium, plutonium, uranium, radon, 14C, and 3H. X-rays used in medical and dental examinations and in radiation therapy of cancer and other diseases are another form of ionizing radiation. It is estimated that UV and ionizing radiations are responsible for about 10% of all DNA damage caused by environmental agents.

DNA also may be damaged by reactive chemicals introduced into the environment as products of industrial activity. Such products may not be injurious per se but may be metabolized by cells into forms that are. Two prominent classes of such agents are (1) deaminating agents, particularly nitrous acid (HNO2) or compounds that can be metabolized to nitrous acid or nitrites, and (2) alkylating agents.

Nitrous acid, formed from organic precursors such as nitrosamines and from nitrite and nitrate salts, is a potent accelerator of the deamination of bases. Bisulfite has similar effects. Both agents are used as preservatives in processed foods to prevent the growth of toxic bacteria. They do not appear to increase cancer risks significantly when used in this way, perhaps because they are used in small amounts and make only a minor contribution to the overall levels of DNA damage. (The potential health risk from food spoilage if these preservatives were not used is much greater.)

*SUMMARY* Nucleic Acid Chemistry

■ Native DNA undergoes reversible unwinding and separation of strands (melting) on heating or at extremes of pH DNAs rich in G≡C pairs have higher melting points than DNAs rich in A=T pairs.

■ Denatured single-stranded DNAs from two species can form a hybrid duplex, the degree of hybridization depending on the extent of sequence similarity. Hybridization is the basis for important techniques used to study and isolate specific genes and RNAs.

■ DNA is a relatively stable polymer. Spontaneous reactions such as deamination of certain bases, hydrolysis of base-sugar N-glycosyl bonds, radiation-induced formation of pyrimidine dimers, and oxidative damage occur at very low rates, yet are important because of cells’ very low tolerance for changes in genetic material.

■ DNA sequences can be determined and DNA polymers synthesized with simple, automated protocols involving chemical and enzymatic methods.

**Other Functions of Nucleotides**

In addition to their roles as the subunits of nucleic acids, nucleotides have a variety of other functions in every cell: as energy carriers, components of enzyme cofactors, and chemical messengers.

**Nucleotides Carry Chemical Energy in Cells**

The phosphate group covalently linked at the 5' hydroxyl of a ribonucleotide may have one or two additional phosphates attached. The resulting molecules are referred to as nucleoside mono-, di-, and triphosphates. Starting from the ribose, the three phosphates are generally labeled α, {3, and y. Hydrolysis of nucleoside triphosphates provides the chemical energy to drive a wide variety of cellular reactions. Adenosine 5'-triphosphate, ATP, is by far the most widely used for this purpose, but UTP, GTP, and CTP are also used in some reactions. Nucleoside triphosphates also serve as the activated precursors of DNA and RNA synthesis.

The energy released by hydrolysis of ATP and the other nucleoside triphosphates is accounted for by the structure of the triphosphate group. The bond between the ribose and the α-phosphate is an ester linkage. The α, β and β, γ linkages are phosphoanhydrides. Hydrolysis of the ester linkage yields about 14 kJ-1 mol under standard conditions, whereas hydrolysis of each anhydride bond yields about 30 kJ-1 mol. ATP hydrolysis often plays an important thermodynamic role in biosynthesis. When coupled to a reaction with a positive free energy change, ATP hydrolysis shifts the equilibrium of the overall process to favor product formation.



**The phosphate ester and phosphoanhydride bonds of ATP.** 

**Nucleoside phosphates. General structure of the nucleoside 5'-mono-, di-, and triphosphates (NMPs, NDPs, and NTPs) and their standard abbreviations. In the deoxy ribonucleoside phosphates (dNMPs, dNDPs, and dNTPs), the pentose is 2'-deoxy-D-ribose.**

**Adenine Nucleotides Are Components of Many Enzyme Cofactors**

A variety of enzyme cofactors serving a wide range of chemical functions include adenosine as part of their structure. They are unrelated structurally except for the presence of adenosine. In none of these cofactors does the adenosine portion participate directly in the primary function, but removal of adenosine generally results in a drastic reduction of cofactor activities. For example, removal of the adenine nucleotide (3'-phosphoadenosine diphosphate) from acetoacetyl CoA, the coenzyme A derivative of acetoacetate, reduces its reactivity as a substrate for {3-ketoacyl-CoA transferase (an enzyme of lipid metabolism) by a factor of 106. Although this requirement for adenosine has not been investigated in detail, it must involve the binding energy between enzyme and substrate (or cofactor) that is used both in catalysis and in stabilizing the initial enzyme-substrate complex. In the case of {3-ketoacyl-CoA transferase, the nucleotide moiety of coenzyme A appears to be a binding “handle” that helps to pull the substrate (acetoacetyl-CoA) into the active site. Similar roles may be found for the nucleoside portion of other nucleotide cofactors.

Why is adenosine, rather than some other large molecule, used in these structures? The answer here may involve a form of evolutionary economy. Adenosine is certainly not unique in the amount of potential binding energy it can contribute. The importance of adenosine probably lies not so much in some special chemical characteristic as in the evolutionary advantage of using one compound for multiple roles. Once ATP became the universal source of chemical energy, systems developed to synthesize ATP in greater abundance than the other nucleotides; because it is abundant, it becomes the logical choice for incorporation into a wide variety of structures.

The economy extends to protein structure. A single protein domain that binds adenosine can be used in a wide variety of enzymes. Such a domain, called a **nucleotide-binding fold,** is found in many enzymes that bind ATP and nucleotide cofactors.



**Some coenzymes containing adenosine. The adenosine portion is shaded in light red. Coenzyme A (CoA) functions in acyl group transfer reactions; the acyl group (such as the acetyl or acetoacetyl group) is attached to the CoA through a thioester linkage to the {3-mercaptoethylamine moiety. NAD+ functions in hydride transfers, and FAD, the active form of vitamin B2 (riboflavin), in electron transfers. Another coenzyme incorporating adenosine is 5'-deoxyadenosylcobalamin, the active form of vita- min B12, which participates in intramolecular** **three regulatory nucleotides**

Certainly not unique in the amount of potential binding energy it can contribute. The importance of adenosine probably lies not so much in some special chemical characteristic as in the evolutionary advantage of using one compound for multiple roles. Once ATP became the universal source of chemical energy, systems developed to synthesize ATP in greater abundance than the other nucleotides; because it is abundant, it becomes the logical choice for incorporation into a wide variety of structures. The economy extends to protein structure. A single protein domain that binds adenosine can be used in a wide variety of enzymes. Such a domain, called a **nucleotide-binding fold,** is found in many enzymes that bind ATP and nucleotide cofactors.

**Some Nucleotides Are Regulator y Molecules**

Cells respond to their environment by taking cues from hormones or other external chemical signals. The interaction of these extracellular chemical signals (“**first messengers**”) with receptors on the cell surface often leads to the production of **second messengers** inside the cell, which in turn leads to adaptive changes in the cell interior. Often, the second messenger is a nucleotide. One of the most common is **adenosine 3**'**,5**'**-cyclic monophosphate** **(cyclic AMP,** or **cAMP),** formed from ATP in a reaction catalyzed by adenylyl cyclase, an enzyme associated with the inner face of the plasma membrane. Cyclic AMP serves regulatory functions in virtually every cell outside the plant kingdom. Guanosine 3',5'-cyclic mono- phosphate (cGMP) occurs in many cells and also has regulatory functions.

Another regulatory nucleotide, ppGpp, is produced in bacteria in response to a slowdown in protein synthesis during amino acid starvation. This nucleotide inhibits the synthesis of the rRNA and tRNA molecules needed for protein synthesis, preventing the unnecessary production of nucleic acids.

*SUMMARY* Other Functions of Nucleotides

■ ATP is the central carrier of chemical energy in cells. The presence of an adenosine moiety in a variety of enzyme cofactors may be related to binding-energy requirements.

■ Cyclic AMP, formed from ATP in a reaction catalyzed by adenylyl cyclase, is a common second messenger produced in response to hormones and other chemical signals.

**References:**

* Lehninger Principles of Biochemistry (Nelson W. H. Freeman. 4th Ed, 2004).
* Biochemistry (Reginald H. Garrett and Charles M. Grisham, University of Virginia 4th Ed, 2010).