# NUCLEOTIDES AND
# NUCLEIC ACIDS

Nucleotides have a variety of roles in cellular metabolism. They are the energy currency in metabolic transactions, the essential chemical links in the response of cells to hormones and other extracellular stimuli, and the structural components of an array of enzyme cofactors and metabolic intermediates. And, last but certainly not least, they are the constituents of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), the molecular repositories of genetic information. The structure of every protein, and ultimately of every biomolecule and cellular component, is a product of information programmed into the nucleotide sequence of a cell's nucleic acids. The ability to store and transmit genetic information from one gener ation to the next is a fundamental condition for life.
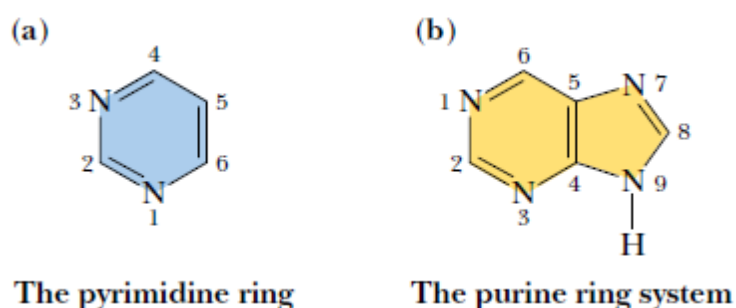
## Some Basics

**Nucleotides, Building Blocks of Nucleic Acids:** The amino acid sequence of every protein in a cell, and the nucleotide sequence of every RNA, is specified by a nucleotide sequence in the cell's DNA. A segment of a DNA molecule that contains the information required for the synthesis of a functional biological product, whether protein or RNA, is referred to as a **gene.** A cell typically has many thousands of genes, and DNA molecules, not surprisingly, tend to be very large. The storage and transmission of biological information are the only known functions of DNA.
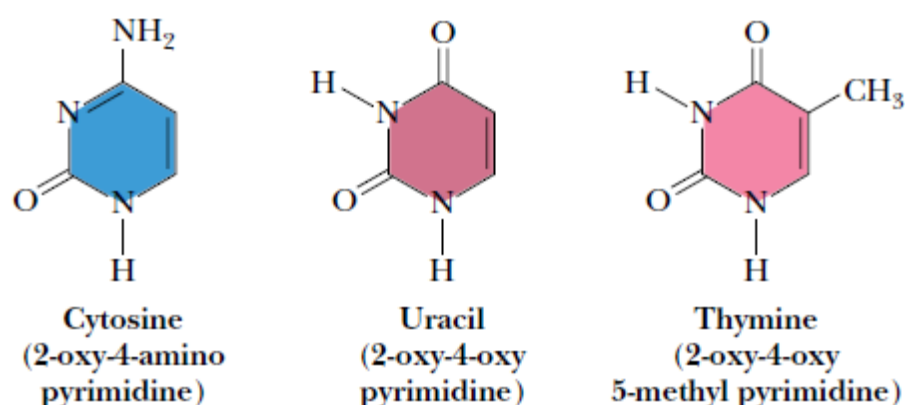
RNAs have a broader range of functions, and several classes are found in cells. **Ribosomal RNAs (rRNAs)** are components of ribosomes, the complexes that carry out the synthesis of proteins. **Messenger RNAs (mRNAs)** are intermediaries, carrying genetic information from one or a few genes to a ribosome, where the corresponding proteins can be synthesized. **Transfer RNAs (tRNAs)** are adapter molecules that faithfully translate the information in mRNA into a specific sequence of amino acids.

## Nucleotides and Nucleic Acids Have Characteristic Bases and Pentoses

**Nucleotides** have three characteristic components: (1) a nitrogenous (nitrogen-containing) base, (2) a pentose, and (3) a phosphate. The molecule without the phosphate group is called a **nucleoside.** The nitrogenous bases are derivatives of two parent com- pounds, **pyrimidine** and **purine.** The bases and pentoses of the common nucleotides are heterocyclic compounds. The carbon and nitrogen atoms in the parent structures are conventionally numbered to facilitate the naming and identification of the many derivative compounds.
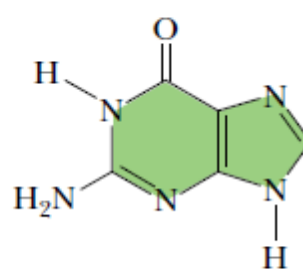


(a) The pyrimidine ring    (b) The purine ring system

Structure of nucleotides.



Cytosine
(2-oxy-4-amino pyrimidine)

Uracil
(2-oxy-4-oxy pyrimidine)

Thymine
(2-oxy-4-oxy 5-methyl pyrimidine)

**The common pyrimidine bases—cytosine, uracil, and thymine—in the tautomeric forms predominant at pH 7.**
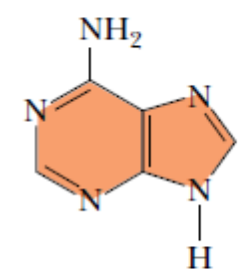
**The common purine bases—adenine and guanine—in the tautomeric forms predominant at pH 7.**

## The Properties of Pyrimidines and Purines Can Be Traced to Their Electron-Rich Nature
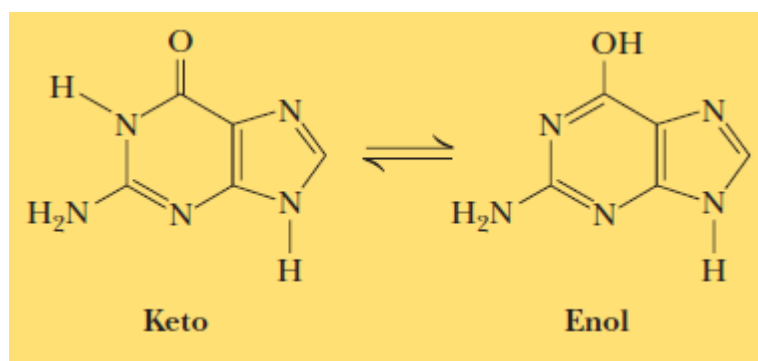
The aromaticity of the pyrimidine and purine ring systems and the electron-rich nature of their carbonyl and ring nitrogen substituents endow them with the capacity to undergo **keto–enol tautomeric shifts.** That is, pyrimidines and purines exist as tautomeric pairs.

The keto tautomers of uracil, thymine, and guanine vastly predominate at neutral pH. In other words, p$K$a values for ring nitrogen atoms 1 and 3 in uracil are greater than 8 (the p$K$a value for N-3 is 9.5). In contrast, the enol form of cytosine predominates at pH 7 and the p$K$a value for N-3 in this pyrimidine is 4.5. Similarly, for guanine, the p$K$a value is 9.4 for N-1 and less than 5 for N-3. These p$K$a values specify whether protons are associated with the various ring nitrogens at neutral pH. As such, they are important in determining whether these nitrogens serve as H-bond donors or acceptors. Hydrogen bonding between purine and pyrimidine bases is fundamental to the biological functions of nucleic acids, as in the formation of the double-helix structure of DNA.

The important functional groups participating in H-bond formation are the amino groups of cytosine, adenine, and guanine; the ring nitrogens at position 3 of pyrimidines and position 1 of purines; and the strongly electronegative oxygen atoms attached at position 4 of uracil and thymine, position 2 of cytosine, and position 6 of guanine.
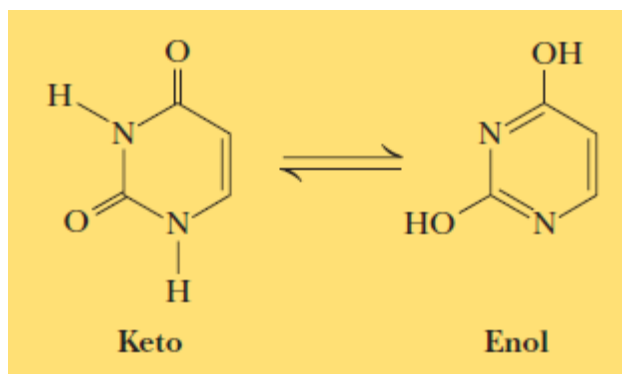


**The keto–enol tautomerization of uracil.**



**The tautomerization of the purine guanine.**

Another property of pyrimidines and purines is their strong absorbance of ultraviolet (UV) light, which is also a consequence of the aromaticity of their heterocyclic ring structures. Characteristic absorption spectra of several of the common bases of nucleic acids—adenine, uracil, cytosine, and guanine—in their nucleotide forms: AMP, UMP, CMP, and GMP. This property is particularly useful in quantitative and qualitative analysis of nucleotides and nucleic acids.
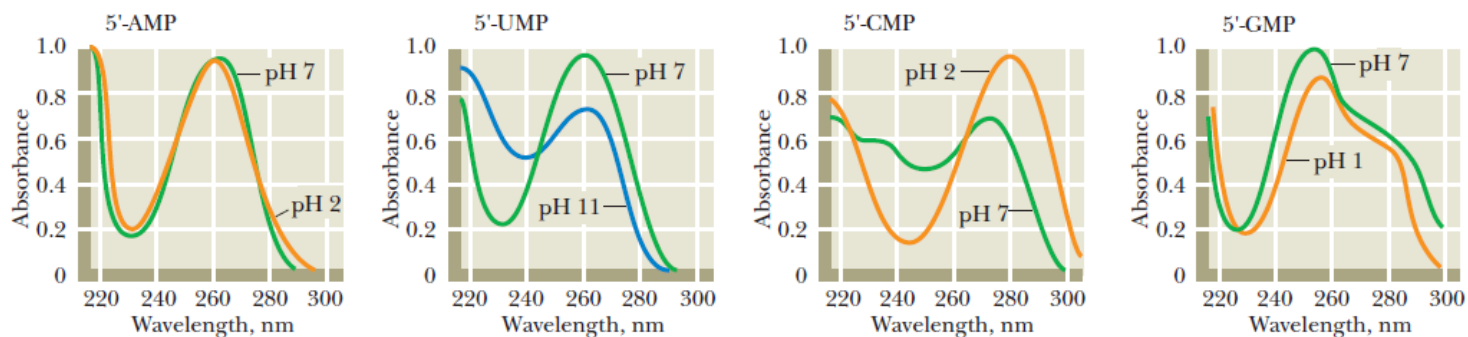
**The UV absorption spectra of the common ribonucleotides.**

**Major purine and pyrimidine bases of nucleic acids.** Some of the common names of these bases reflect the circumstances of their discovery. Guanine, for example, was first isolated from guano (bird manure), and thymine was first isolated from thymus tissue.

And nucleosides the carbon numbers are given a prime (') designation to distinguish them from the numbered atoms of the nitrogenous bases.

The base of a nucleotide is joined covalently (at N-1 of pyrimidines and N-9 of purines) in an *N-β-glycosyl* bond to the 1' carbon of the pentose, and the phosphate is esterified to the 5' carbon. The *N-β-glycosyl* bond is formed by removal of the elements of water (a hydroxyl group from the pentose and hydrogen from the base), as in *O*-glycosidic bond formation.

Both DNA and RNA contain two major purine bases, **adenine** (A) and **guanine** (G), and two major pyrim- idines. In both DNA and RNA one of the pyrimidines is **cytosine** (C), but the second major pyrimidine is not the same in both: it is **thymine** (T) in DNA and **uracil** (U) in RNA. Only rarely does thymine occur in RNA or uracil in DNA. The structures of the five major bases are shown in, and the nomenclature of their corresponding nucleotides and nucleosides is summa- rized.

Nucleic acids have two kinds of pentoses. The re- curring deoxyribonucleotide units of DNA contain 2'- deoxy-D-ribose, and the ribonucleotide units of RNA contain D-ribose. In nucleotides, both types of pentoses are in their *β*-furanose (closed five-membered ring) form. The pentose ring is not planar but occurs in one of a variety of conformations generally described as "puckered."

The structures and names of the four major **deoxyribonucleotides** (deoxyribonucleo- side 5'-monophosphates), the structural units of DNAs, and the four major **ribonucleotides** (ribonucleoside 5'- monophosphates), the structural units of RNAs. Specific long sequences of A, T, G, and C nucleotides in DNA are the repository of genetic information. Although nucleotides bearing the major purines and pyrimidines are most common, both DNA and RNA also.
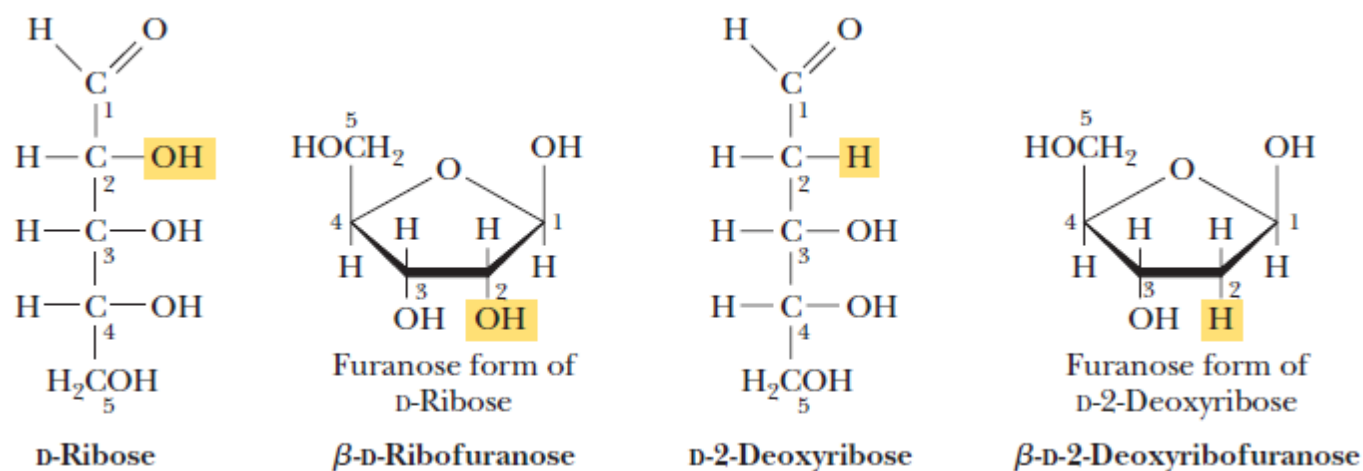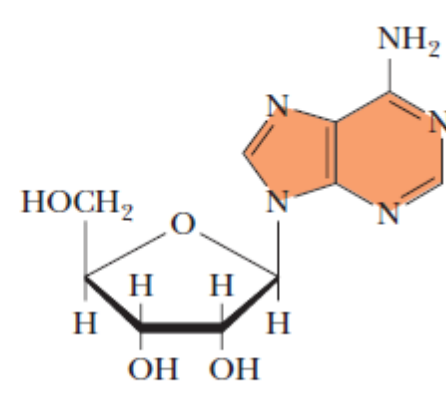


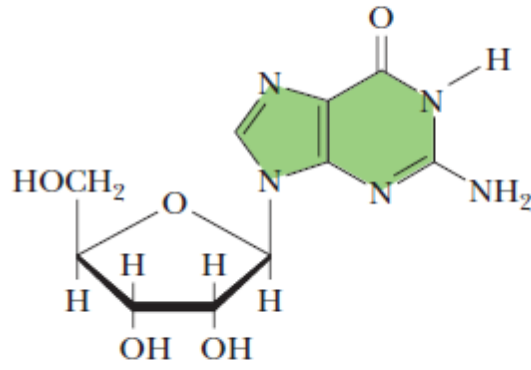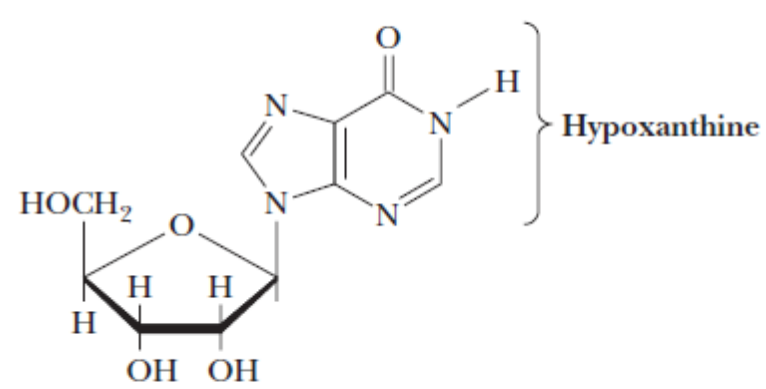**Furanose structures—ribose and deoxyribose.**
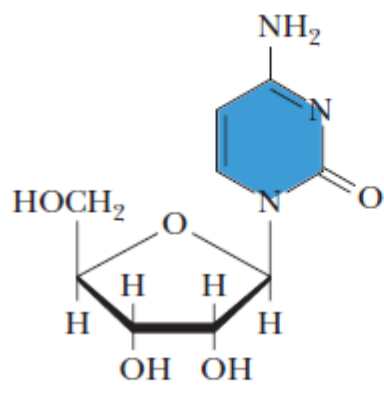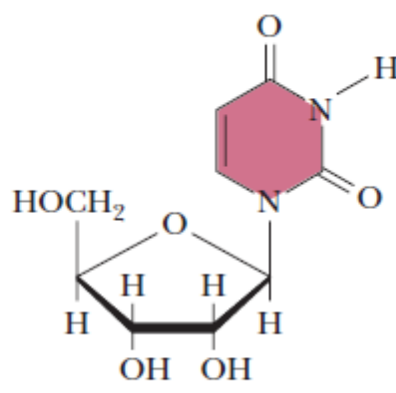
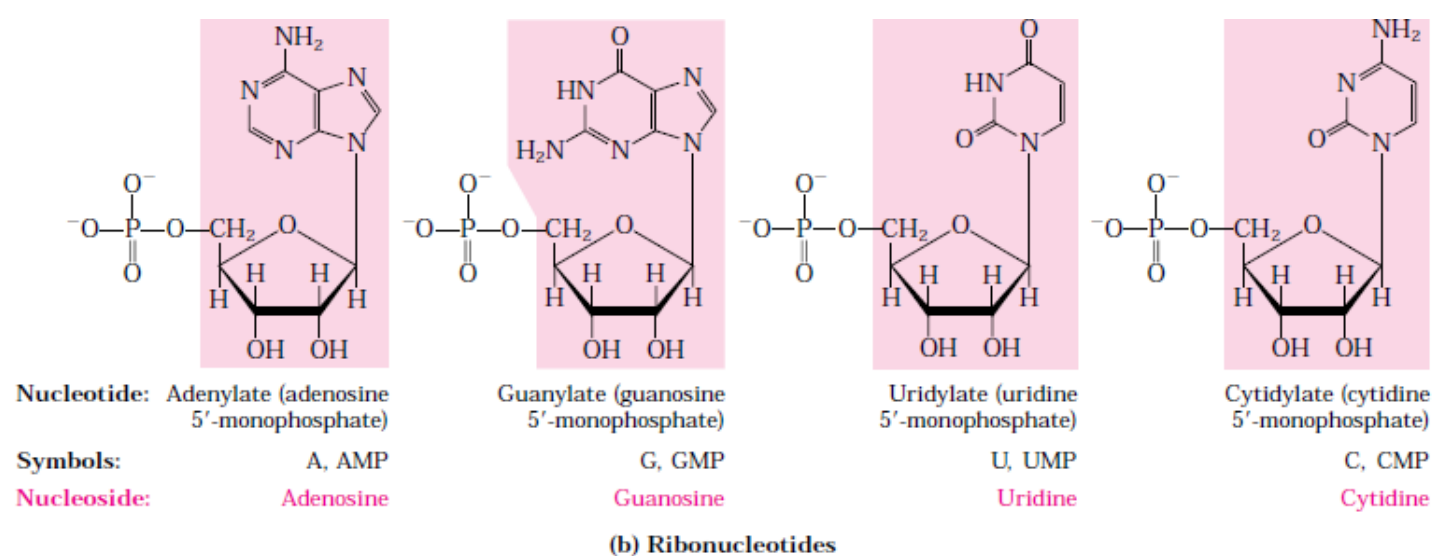**Cytidine**     **Uridine**     **Adenosine**

**Guanosine**     **Inosine, a less common nucleoside**

The common ribonucleosides—cytidine, uridine, adenosine, and guanosine—and the lesscommon inosine. (Purine nucleosides and nucleotides usually adopt the anti-conformation, where the purine ring is not above the ribose, as it would be in the syn conformation. Pyrimidines are always anti, never syn, because the 2-O atom of pyrimidines sterically hinders the ring from a position above the ribose).



| Nucleotide: | Deoxyadenylate (deoxyadenosine 5'-monophosphate) | Deoxyguanylate (deoxyguanosine 5'-monophosphate) | Deoxythymidylate (deoxythymidine 5'-monophosphate) | Deoxycytidylate (deoxycytidine 5'-monophosphate) |
|---|---|---|---|---|
| Symbols: | A, dA, dAMP | G, dG, dGMP | T, dT, dTMP | C, dC, dCMP |
| Nucleoside: | Deoxyadenosine | Deoxyguanosine | Deoxythymidine | Deoxycytidine |

**(a) Deoxyribonucleotides**

| Nucleotide: | Adenylate (adenosine 5'-monophosphate) | Guanylate (guanosine 5'-monophosphate) | Uridylate (uridine 5'-monophosphate) | Cytidylate (cytidine 5'-monophosphate) |
|---|---|---|---|---|
| Symbols: | A, AMP | G, GMP | U, UMP | C, CMP |
| Nucleoside: | Adenosine | Guanosine | Uridine | Cytidine |

**(b) Ribonucleotides**

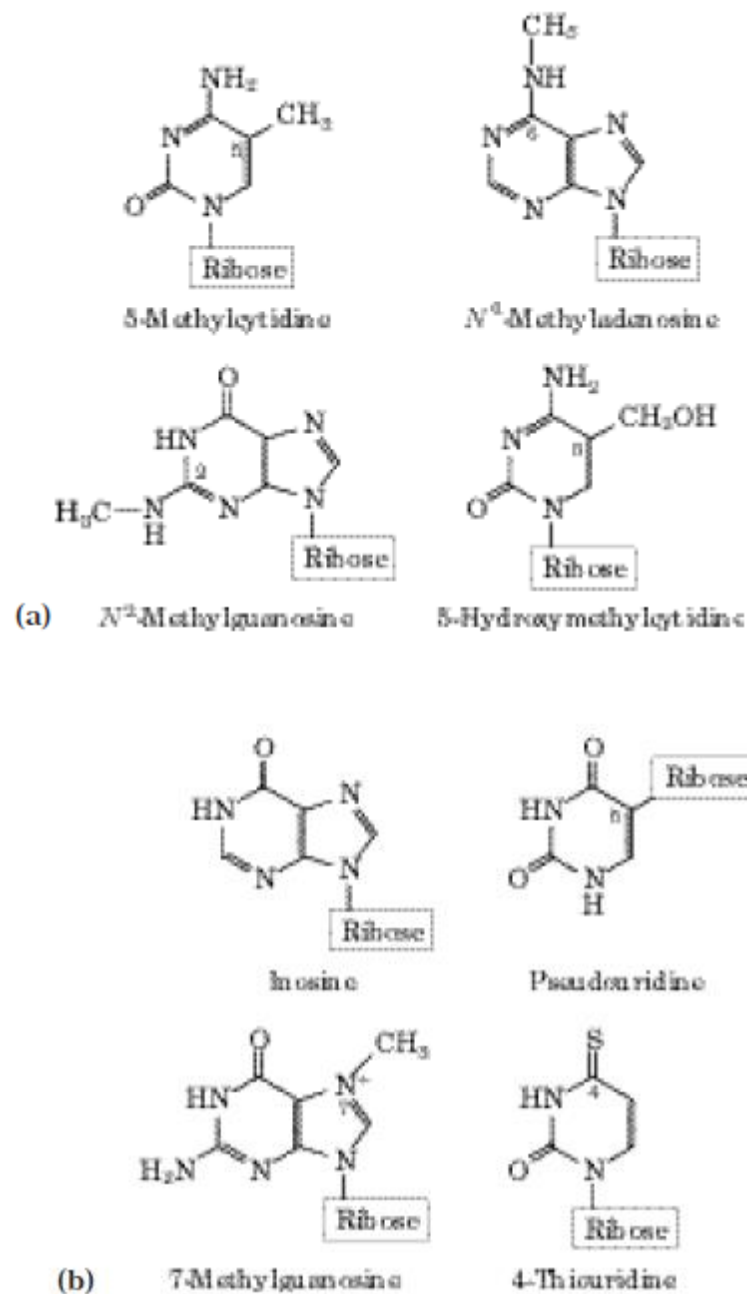**Deoxyribonucleotides and ribonucleotides of nucleic acids.** All nucleotides are shown in their free form at pH 7.0. The nu- cleotide units of DNA **(a)** are usually symbolized as A, G, T, and C, sometimes as dA, dG, dT, and dC; those of RNA **(b)** as A, G, U, and C. In their free form the deoxyribonucleotides are commonly abbre- viated dAMP, dGMP, dTMP, and dCMP; the ribonucleotides, AMP, GMP, UMP, and CMP. For each nucleotide, the more common name is followed by the complete name in parentheses. All abbreviations assume that the phosphate group is at the 5' position. The nucleoside portion of each molecule is shaded in light red. In this and the fol- lowing illustrations, the ring carbons are not shown.

| Base | Nucleoside | Nucleotide | Nucleic acid |
|------|-----------|-----------|--------------|
| **Purines** | | | |
| Adenine | Adenosine | Adenylate | RNA |
| | Deoxyadenosine | Deoxyadenylate | DNA |
| Guanine | Guanosine | Guanylate | RNA |
| | Deoxyguanosine | Deoxyguanylate | DNA |
| **Pyrimidines** | | | |
| Cytosine | Cytidine | Cytidylate | RNA |
| | Deoxycytidine | Deoxycytidylate | DNA |
| Thymine | Thymidine or deoxythymidine | Thymidylate or deoxythymidylate | DNA |
| Uracil | Uridine | Uridylate | RNA |

Note: "Nucleoside" and "nucleotide" are generic terms that include both ribo- and deoxyribo- forms. Also, ribonucleosides and ribonucleotides are here designated simply as nucleosides and nucleotides (e.g., ribo- adenosine as adenosine), and deoxyribo- nucleosides and deoxyribonucleotides as deoxynucleosides and deoxynucleotides (e.g., deoxyriboadenosine as deoxyadeno- sine). Both forms of naming are acceptable, but the shortened names are more commonly used. Thymine is an exception; "ribothymidine" is used to describe its unusual occurrence in RNA.

$NH_2$  $CH_3$  $CH_3$  $NH$

| Ribose | | Ribose |

5-Methylcytidine    $N^6$-Methyladenosine

$H_3C-N$  $O$  $HN$  $N$  $NH_2$  $CH_2OH$

| Ribose | | Ribose |

(a)    $N^2$-Methylguanosine    5-Hydroxymethylcytidine

$O$  $HN$  $N$  $O$  $HN$  | Ribose |

| Ribose |

Inosine    Pseudouridine

$O$  $CH_3$  $HN$  $N^+$  $H_2N$  $N$  $S$  $HN$  $O$  $N$

| Ribose | | Ribose |

(b)    7-Methylguanosine    4-Thiouridine

**Some minor purine and pyrimidine bases, shown as the nucleosides. (a)** Minor bases of DNA. 5-Methylcytidine occurs in the DNA of animals and higher plants, $N^6$-methyladenosine in bacterial DNA, and 5-hydroxymethylcytidine in the DNA of bacteria infected with certain bacteriophages. **(b)** Some minor bases of tRNAs. Inosine contains the base hypoxanthine. Note that pseudouridine, like uridine, contains uracil; they are distinct in the point of attachment to the ribose in uridine, uracil is attached through N-1, the usual attachment point for pyrimidines; in pseudouridine, through C-5

contain some minor bases. In DNA the most common of these are methylated forms of the major bases; in some viral DNAs, certain bases may be hy- droxymethylated or glucosylated. Altered or unusual bases in DNA molecules often have roles in regulating or protecting the genetic information. Minor bases of many types are also found in RNAs, especially in tRNAs.

The nomenclature for the minor bases can be con- fusing. Like the major bases, many have common names hypoxanthine, for example, shown as its nucleoside inosine. When an atom in the purine or pyrimidine ring is substituted, the usual convention (used

here) is simply to indicate the ring position of the substituent by its number for example, 5-methylcytosine, 7-methylguanine, and 5-hydroxymethylcytosine (shown as the nucleosides in The element to which the substituent is attached (N, C, O) is not identified. The convention changes when the substituted atom is exocyclic (not within the ring structure), in which case the type of atom is identified and the ring position to which it is attached is denoted with a superscript. The amino nitrogen attached to C-6 of adenine is $N^6$; similarly, the carbonyl oxygen and amino nitrogen at C-6 and C-2 of guanine are $O^6$ and $N^2$, respectively. Examples of this nomenclature are $N^6$-methyladenosine and $N^2$- methylguanosine.

Cells also contain nucleotides with phosphate groups in positions other than on the 5' carbon.

**Ribonucleoside 2',3'-cyclic monophosphates** are isolatable intermediates, and **ribonucleoside 3'-monophosphates** are end products of the hydrolysis of RNA by certain ribonucleases. Other variations are adenosine 3',5'-cyclic monophosphate (cAMP) and guanosine 3',5'-cyclic monophosphate (cGMP).

## Phosphodiester Bonds Link Successive Nucleotides in Nucleic Acids

The successive nucleotides of both DNA and RNA are covalently linked through phosphate-group "bridges," in which the 5'-phosphate group of one nucleotide unit is



Adenosine 5'-monophosphate     Adenosine 2'-monophosphate

Adenosine 3'-monophosphate     Adenosine 2',3'-cyclic monophosphate
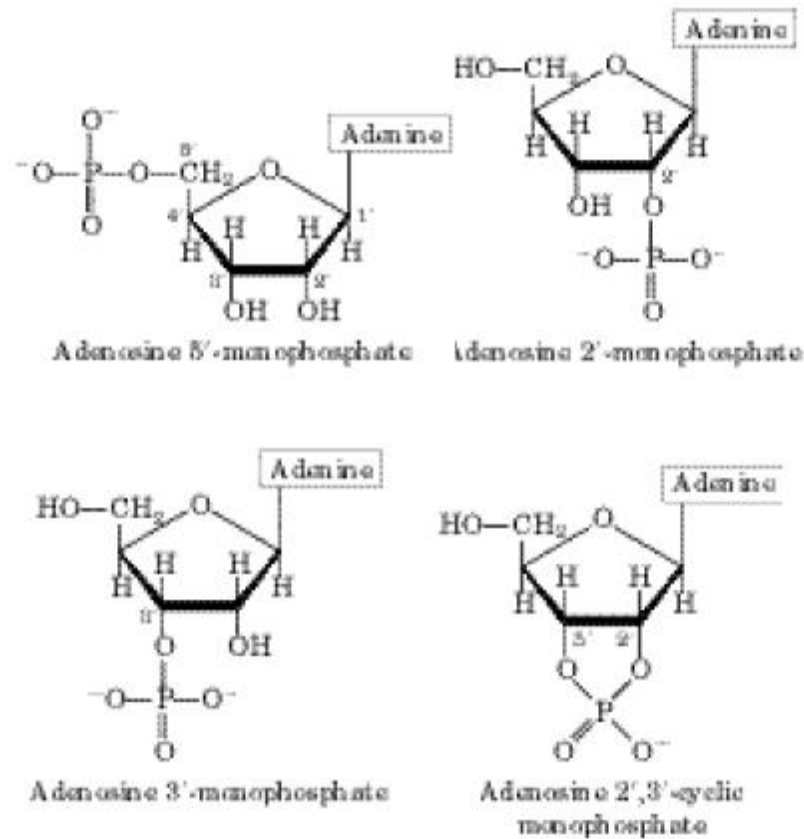
**Some adenosine monophosphates.** Adenosine 2'-monophosphate, 3'-monophosphate, and 2',3'-cyclic monophosphate are formed by enzymatic and alkaline hydrolysis of RNA.

joined to the 3'-hydroxyl group of the next nucleotide, creating a **phosphodiester linkage**. Thus the covalent backbones of nucleic acids consist of al- ternating phosphate and pentose residues, and the ni- trogenous bases may be regarded as side groups joined to the backbone at regular intervals. The backbones of both DNA and RNA are hydrophilic. The hydroxyl groups of the sugar residues form hydrogen bonds with water

The phosphate groups, with a p$K_a$ near 0, are completely ionized and negatively charged at pH 7.

The negative charges are generally neutralized by ionic interactions with positive charges on proteins, metal ions, and polyamines.
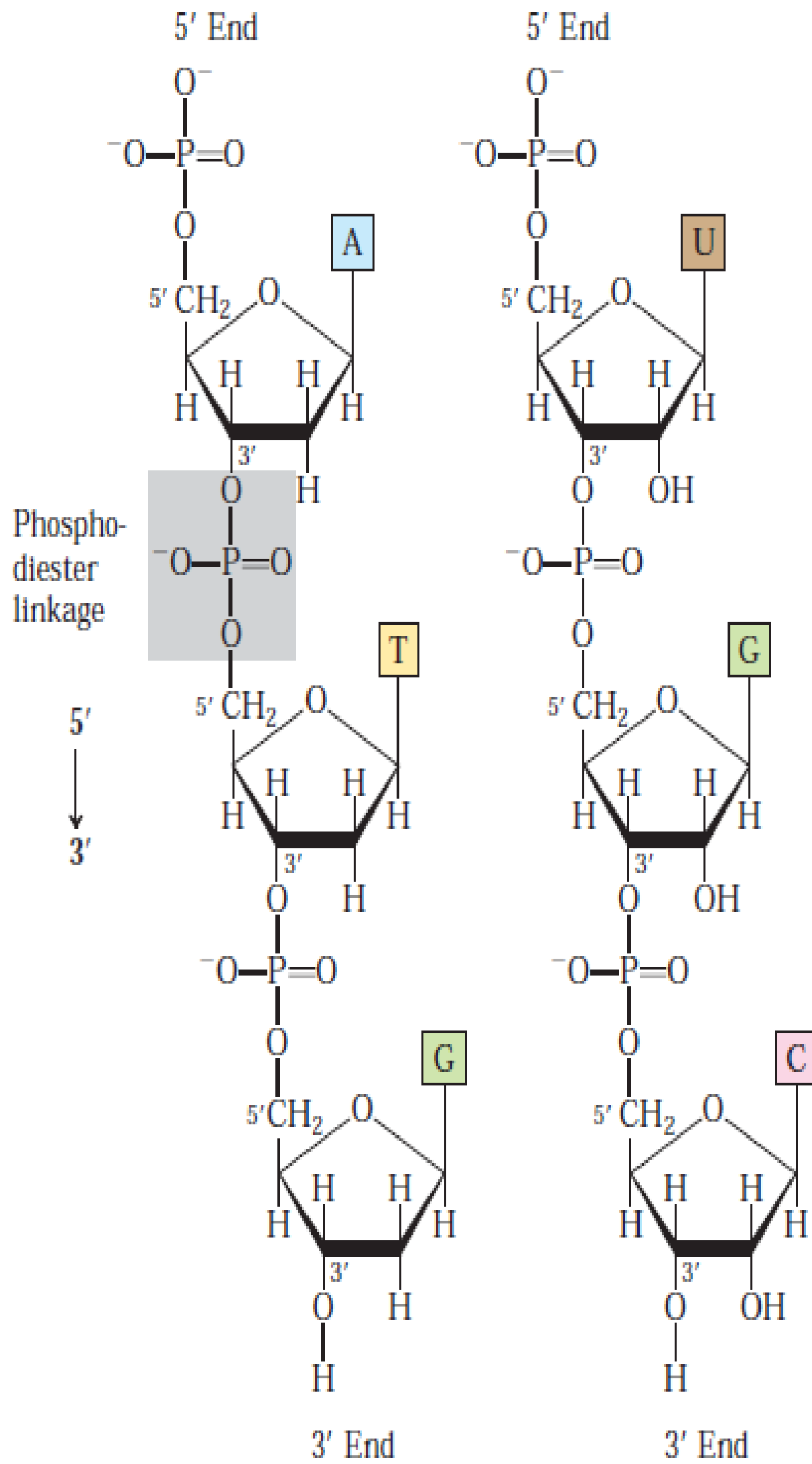
All the phosphodiester linkages have the same orientation along the chain giving each linear nucleic acid strand a specific polarity and distinct 5' and
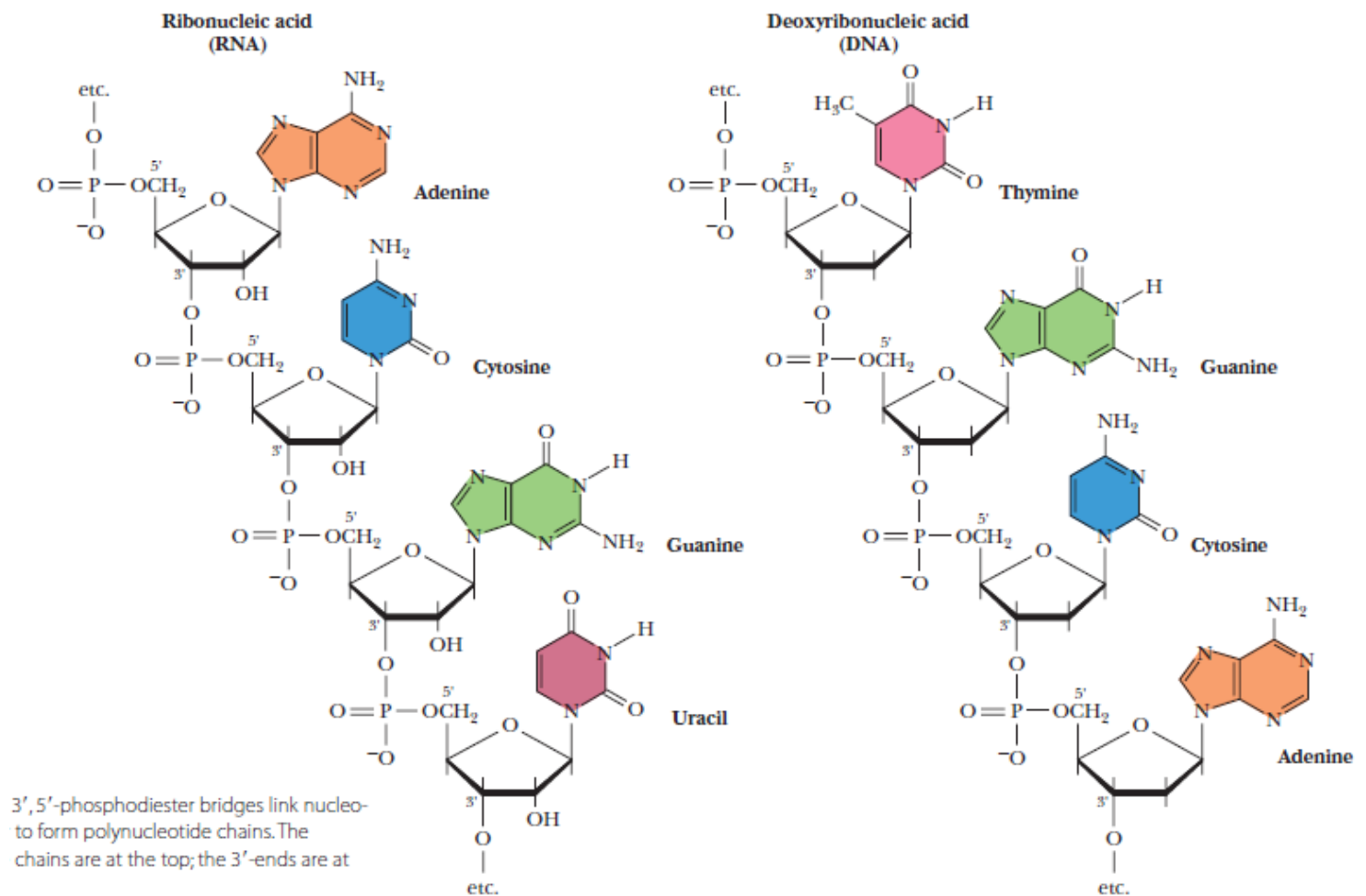
3' ends. By definition, the **5' end** lacks a nucleotide at the 5' position and the **3' end** lacks a nucleotide at the 3' position. Other groups (most often one or more phos- phates) may be present on one or both ends.

The covalent backbone of DNA and RNA is subject to slow, nonenzymatic hydrolysis of the phosphodiester bonds. In the test tube, RNA is hydrolyzed rapidly under alkaline conditions, but DNA is not; the 2'-hydroxyl groups in RNA (absent in DNA) are directly involved in the process. Cyclic 2',3'-monophosphate nucleotides are the first products of the action of alkali on RNA and are rapidly hydrolyzed further to yield a mixture of 2' and 3'-nucleoside monophosphates.

DNA       RNA

**Ribonucleic acid (RNA)**
**Deoxyribonucleic acid (DNA)**

3′, 5′-phosphodiester bridges link nucleo-
to form polynucleotide chains. The
chains are at the top; the 3′-ends are at

**3′, 5′-phosphodiester bridges link nucleotides together to form polynucleotide chains. The 5′-ends of the chains are at the top; the 3′-ends are at the bottom.**

**Phosphodiester linkages in the covalent backbone of DNA and RNA.** The phosphodiester bonds (one of which is shaded in the DNA) link successive nucleotide units. The backbone of alternating pentose and phosphate groups in both types of nucleic acid is highly polar. The 5' end of the macromolecule lacks a nucleotide at the 5' position, and the 3' end lacks a nucleotide at the 3' position.

The nucleotide sequences of nucleic acids can be represented schematically, as illustrated on the following page by a segment of DNA with five nucleotide units. The phosphate groups are symbolized by P, and each deoxyribose is symbolized by a vertical line, from C-1' at the top to C-5' at the bottom (but keep in mind that

2',3'-Cyclic monophosphate derivative
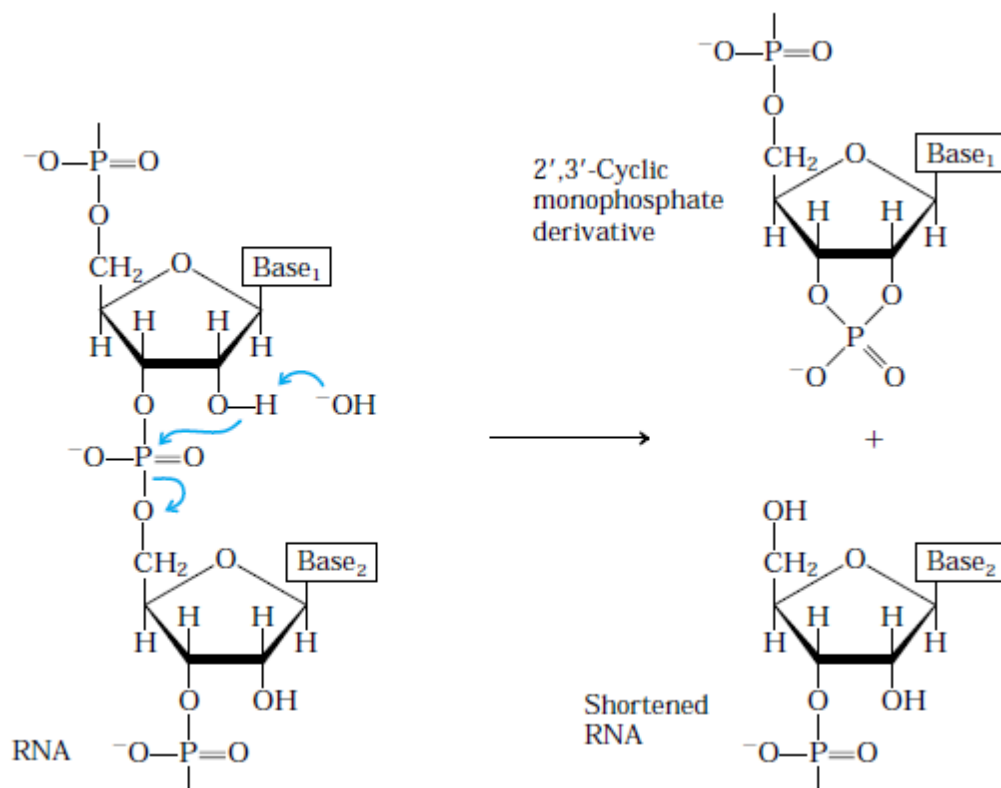


**Mixture of 2'- and 3'-monophosphate derivatives**

**Hydrolysis of RNA under alkaline conditions.** The 2' hydroxyl acts as a nucleophile in an intramolecular displacement. The 2',3'-cyclic monophosphate derivative is further hydrolyzed to a mixture of 2'- and 3'-monophosphates. DNA, which lacks 2' hydroxyls, is stable under similar conditions.

The sugar is always in its closed-ring *β*-furanose form in nucleic acids). The connecting lines between nucleotides (which pass through Đ are drawn diagonally from the middle (C-3') of the deoxyribose of one nucleotide to the bottom (C-5') of the next.
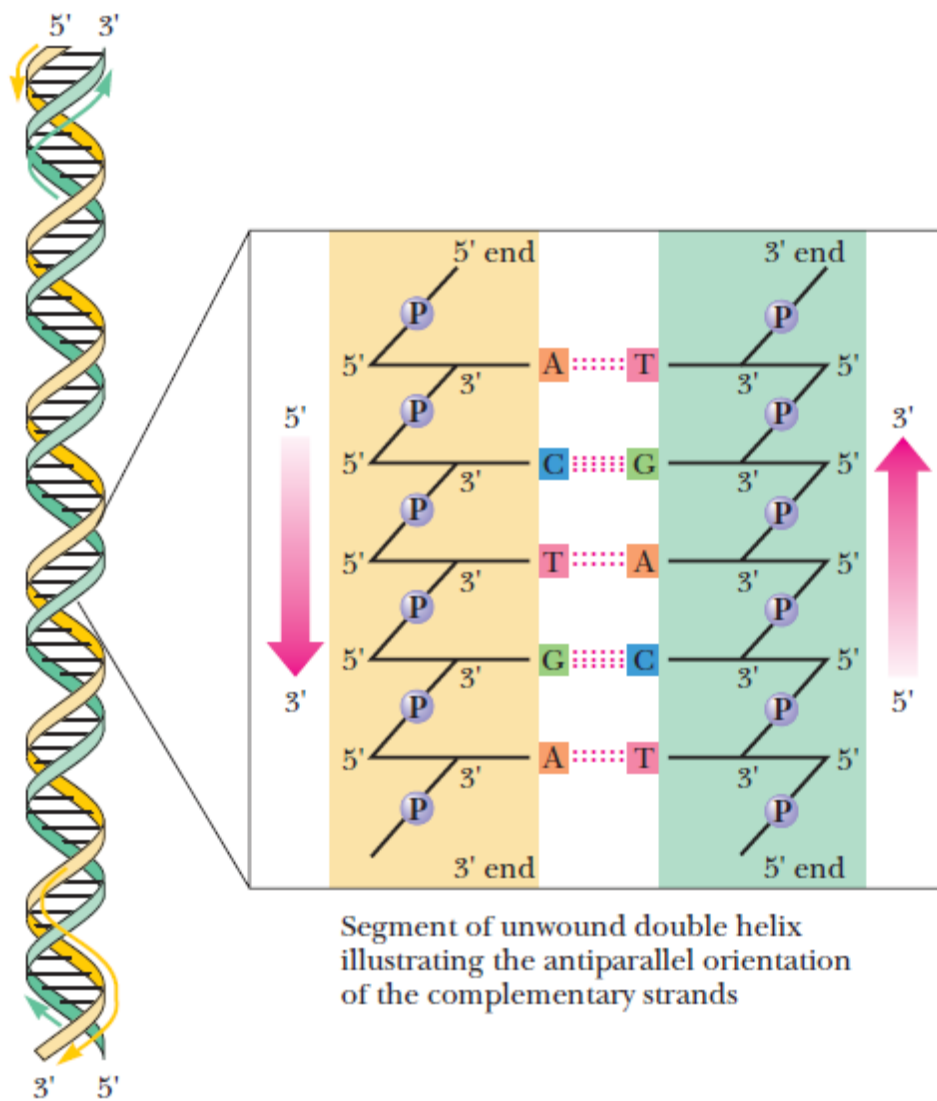


**The antiparallel nature of the DNA double helix.**

By convention, the structure of a single strand of nucleic acid is always written with the 5' end at the left and the 3' end at the right—that is, in the 5' ⟶ 3' di- rection. Some simpler representations of this pentadeoxy ribonucleotide are pA-C-G-T-AOH, pApCpGpTpA, and pACGTA.

A short nucleic acid is referred to as an **oligonucleotide.** The definition of "short" is somewhat arbi- trary, but polymers containing 50 or fewer nucleotides are generally called oligonucleotides. A longer nucleic acid is called a **polynucleotide.**

## The Properties of Nucleotide Bases Affect the Three-Dimensional Structure of Nucleic Acids:

Free pyrimidines and purines are weakly basic com- pounds and are thus called bases. They have a variety of chemical properties that affect the structure, and ultimately the function, of nucleic acids. The purines and pyrimidines common in DNA and RNA are highly conjugated molecules (Fig. 8–2), a property with im- portant consequences for the structure, electron distri- bution, and light absorption of nucleic acids. Resonance among atoms in the ring gives most of the bonds par- tial double-bond character. One result is that pyrim- idines are planar molecules; purines are very nearly



Lactam      Lactim      Double lactim
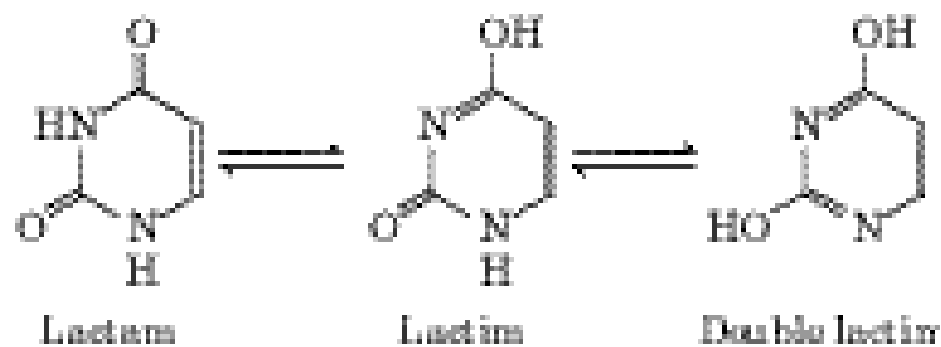
**Uracil**

**Tautomeric forms of uracil.** The lactam form predominates at pH 7.0; the other forms become more prominent as pH decreases. The other free pyrimidines and the free purines also have tautomeric forms, but they are more rarely encountered.

Planar, with a slight pucker. Free pyrimidine and purine bases may exist in two or more tautomeric forms depending on the pH. Uracil, for example, occurs in lactam, lactim, and double lactim forms. The structures shown in are the tautomers that predominate at pH 7.0. As a result of resonance, all nucleotide bases absorb UV light, and nucleic acids are characterized by a strong absorption at wavelengths near 260 nm.

The purine and pyrimidine bases are hydrophobic and relatively insoluble in water at the near-neutral pH of the cell. At acidic or alkaline pH the bases become charged and their solubility in water increases. Hydrophobic stacking interactions in which two or more bases are positioned with the planes of their rings parallel (like a stack of coins) are one of two important modes of interaction between bases in nucleic acids. The stacking also involves a combination of van der Waals and dipole-dipole interactions between the bases. Base stacking helps to minimize contact of the bases with water, and base-stacking interactions are very important in stabilizing the three-dimensional structure of nucleic acids, as described later.
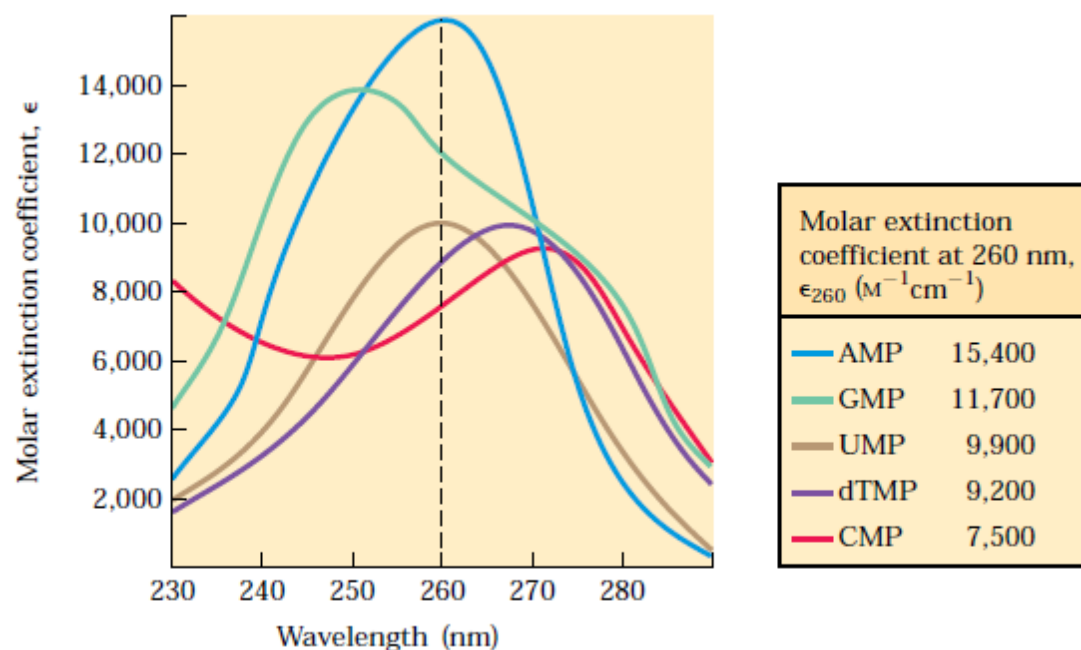


**Absorption spectra of the common nucleotides.** The spectra are shown as the variation in molar extinction coefficient with wavelength. The molar extinction coefficients at 260 nm and pH 7.0 ($c_{260}$) are listed in the table. The spectra of corresponding ribonucleotides and deoxyribonucleotides, as well as the nucleosides, are essentially identical. For mixtures of nucleotides, a wavelength of 260 nm (dashed vertical line) is used for absorption measurements. **Hydrogen-bonding patterns in the base pairs defined by Watson and Crick.** Here as elsewhere, hydrogen bonds are represented by three blue lines.

The most important functional groups of pyrim- idines and purines are ring nitrogens, carbonyl groups, and exocyclic amino groups. Hydrogen bonds involving the amino and carbonyl groups are the second important mode of interaction between bases in nucleic acid molecules. Hydrogen bonds between bases permit a complementary association of two (and occasionally three or four) strands of nucleic acid. The most important hydrogen-bonding patterns are those defined by James D. Watson and Francis Crick in 1953, in which A bonds specifically to T (or U) and G bonds to C. These two types of **base pairs** predominate in double-stranded DNA and RNA, and the tautomers shown in Figure 8–2 are responsible for these patterns. It is this specific pairing of bases that permits the du- plication of genetic information, as we shall discuss later in this chapter.
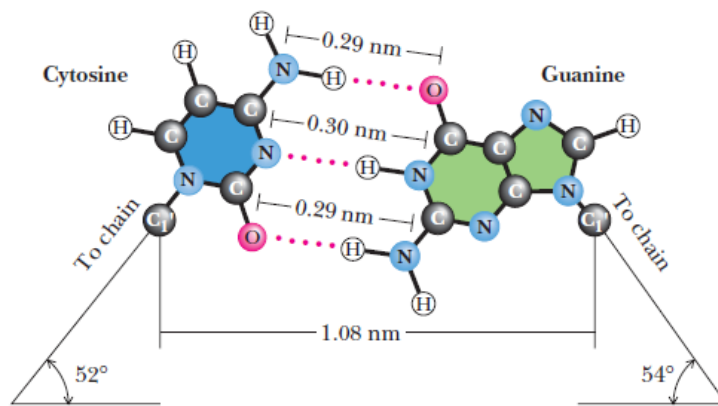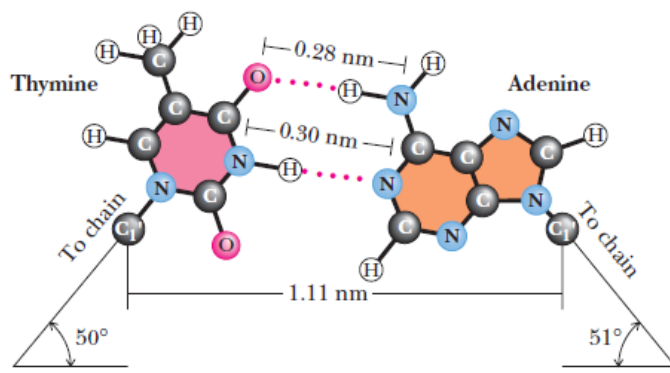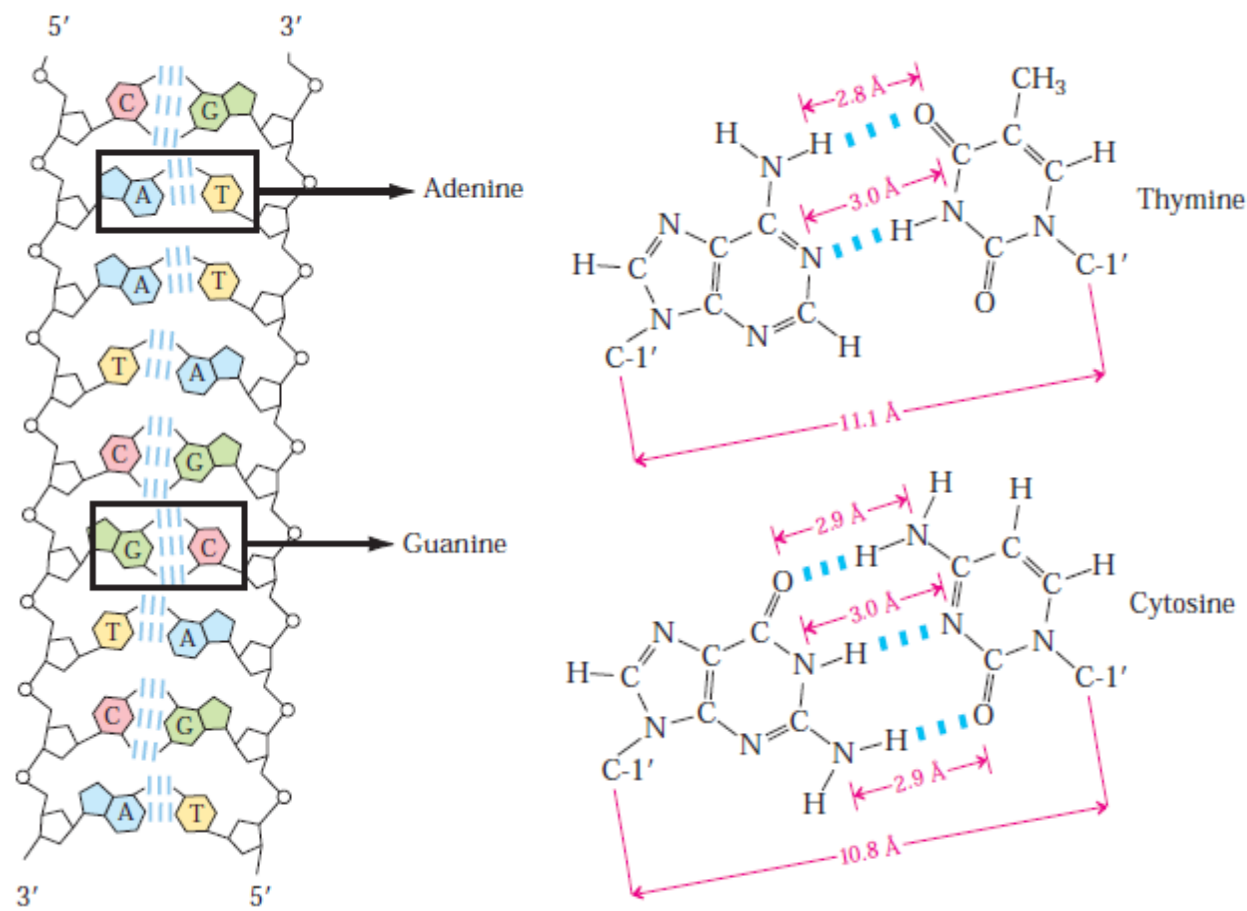
## SUMMARY

■ A nucleotide consists of a nitrogenous base (purine or pyrimidine), a pentose sugar, and one or more phosphate groups. Nucleic acids are polymers of nucleotides, joined together by phosphodiester linkages between the 5'- hydroxyl group of one pentose and the 3'- hydroxyl group of the next.

■ There are two types of nucleic acid: RNA and DNA. The nucleotides in RNA contain ribose, and the common pyrimidine bases are uracil and cytosine. In DNA, the nucleotides contain 2'-deoxyribose, and the common pyrimidine bases are thymine and cytosine. The primary purines are adenine and guanine in both RNA and DNA.

**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).