Nucleotides and Nucleic Acids

Nucleotides
- Energy Currency in Metabolic Transactions
- Essential Chemical Links in Response of Cells to Hormones and Extracellular Stimuli
- Structural Component Some Enzyme Cofactors and Metabolic Intermediate
- Constituents of Nucleic Acids: DNA & RNA

Basics about Nucleotides

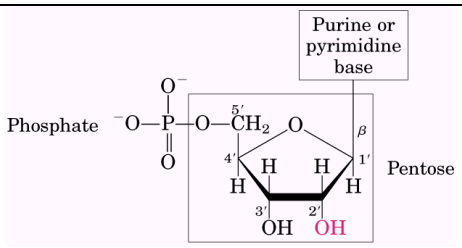
1. Term

Gene: 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.

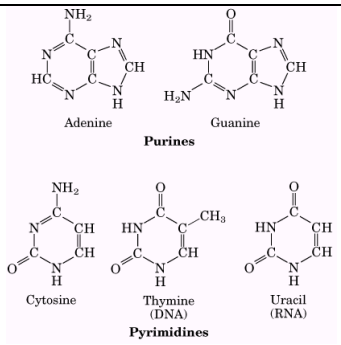
Nucleotides: Nucleotides have three characteristic components: (1) a nitrogenous (nitrogen-containing) base, (2) a pentose, and (3) a phosphate. The molecule without the phosphate groups is called a nucleoside.

Oligonucleotide: A short nucleic acid is referred to as an oligonucleotide, usually contains 50 or fewer nucleotides.

Polynucleotide: Polymers containing more than 50 nucleotides is usually referred to as polynucleotide.



General structure of nucleotide, including a phosphate group, a pentose and a base unit (either Purine or Pyrimidine).



Major purine and Pyrimidine bases of nucleic acid

The roles of RNA and DNA

DNA: a) Biological Information Storage, b) Biological Information Transmission

RNA: a) Structural components of ribosomes and carry out the synthesis of proteins (Ribosomal RNAs: rRNA); b) Intermediaries, carry genetic information from gene to ribosomes (Messenger RNAs: mRNA); c) Adapter molecules that translate the information in mRNA to proteins (Transfer RNAs: tRNA); and a variety of RNAs with other special functions.

table 10–1

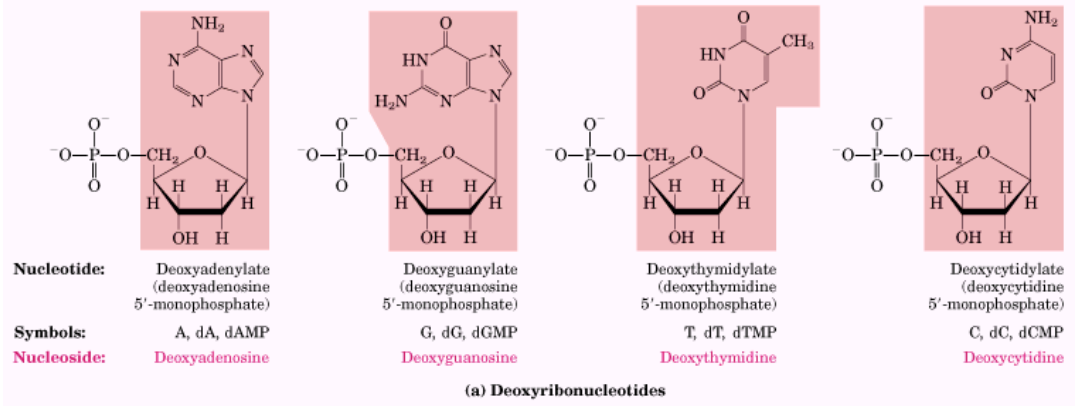
Nucleotide and Nucleic Acid Nomenclature

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

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

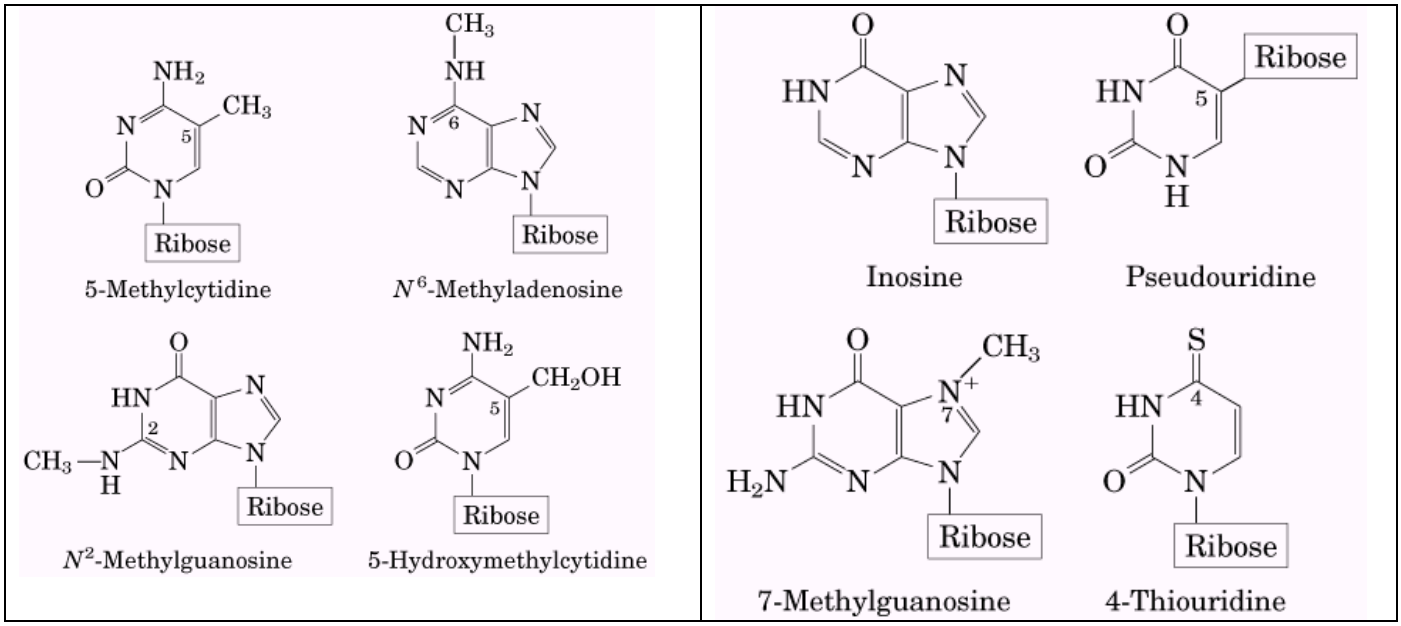
Both DNA and RNA contain two major **purine** bases, adenine (A) and guanine (G), and two major **pyrimidines**. In both DNA and RNA, one of the Pyrimidine is cytosine (C), but the second major pyrimidine is thymine (T) in DNA and uracil (U) in RNA. The names of these bases in nucleosides and nucleotides are shown in table 10-1.

Nucleic acids have two kinds of **pentoses**. The recurring deoxyribonucleotide units of DNA contain 2'-deoxy-D-ribose, and the ribonucleotide units of RNA contain D-ribose. This is the actual fact the name of DNA and RNA come from.

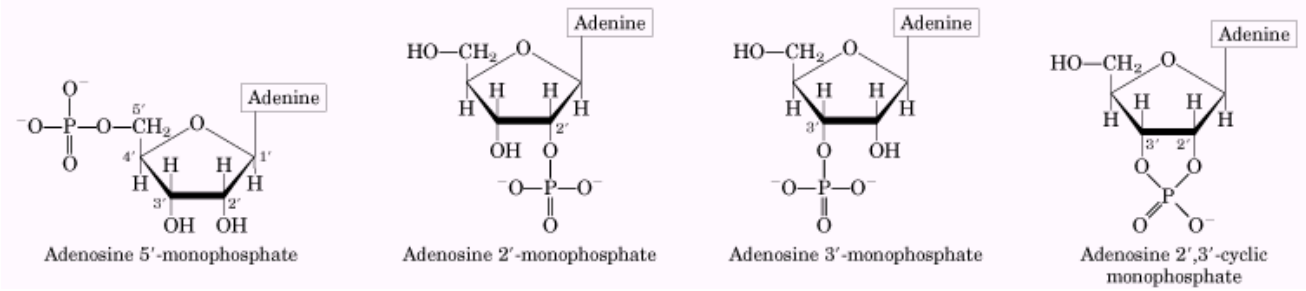


Deoxyribonucleotides and **ribonucleotides** of nucleic acids. All nucleotides are shown in their free form at pH 7.0. The nucleotide units of DNA (a) are usually symbolized as **A**, **G**, **T**, and **C**, sometimes as dA, dG, dT, and dC; those of RNA (b) as **A**, **G**, **U**, and **C**. in their free form the deoxyribonucleotides are commonly abbreviated damp, dGMP, dTMP, and dCMP; the ribonucleotides, **AMP**, **GMP**, **UMP**, and **CMP**. For each nucleotide, the more common names is followed by the complete name in parentheses. All abbreviations assume that the phosphate group is at the 5' posotion. The nucleotide portion of each molecule is shade.

Besides the major base mentioned previously, other minor base units also occur in RNA or DNA. When these minor bases occur in DNA, they may regulate or protect the genetic information; many of these minor bases also appear in RNA, especially in **tRNA**.
Remember in nucleotides, two sets of numbering systems are used, the first is located on the base units, the second one is used for pentose unit, and the number with prime' to differentiate from the first set. When substituents appear on the base unit ring, as shown in those minor base units, on either C or N atom, it is not necessary to point out these atoms; however, when the substituents occur at the atom outside the ring of base unit, the atom with substituents must be pointed out, as shown in the structure below.



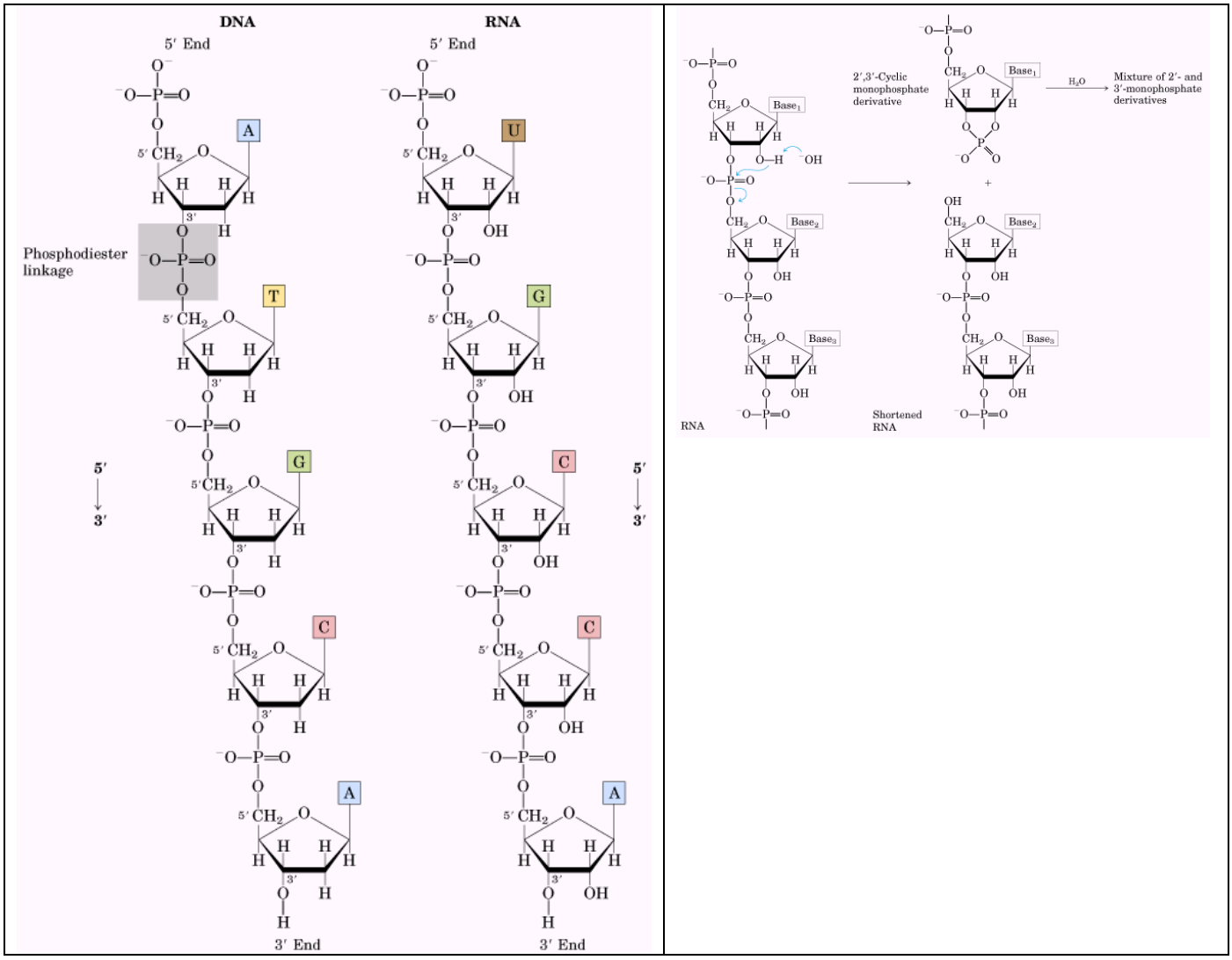
The variation for the structures of nucleotide not only happens on the base units, but also appears on the sugar segment. For example, Ribonucleoside 2',3'-cyclicmonophosphates are isolatable intermediates and ribonucleoside 3'-monophosphates are end products of the hydrolysis of RNA by certain **ribonucleases**. Examples are shown below.



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' hydroxyl group of one nucleotide unit is joined to the 3'-hydroxyl group of the next nucleotide by a **phosphodiester** linkage. Thus the backbone of nucleic acid consist of the alternating phosphate and pentose residues, and nitrogenous base are side groups joined to the backbone at regular intervals. The phosphate groups, with a pKa near 0, are completely ionized. (Both DNA and RNA are hydrophilic, why?)

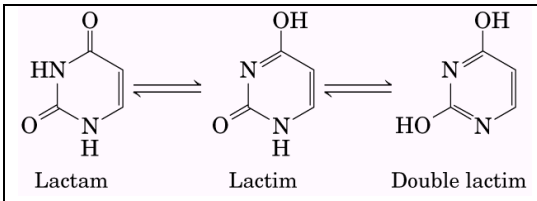
When draw the structure of DNA or RNA fragment, it is always to start from 5'-nucleotide on the left to 3'-nucleotide on the right, for example, the name of the DNA fragment shown below is pA-T-G-C-A_{OH}, p means on the 5'- end this DNA fragment has a phosphate group, A_{OH} stands for this DNA fragment ends at nucleotide A with free 3' hydroxyl group. This DNA fragment sequence can be written as: pApTpGpCpA or simply as pATGCA. (compare with the convention during writing the amino acid sequence in peptide or protein!).

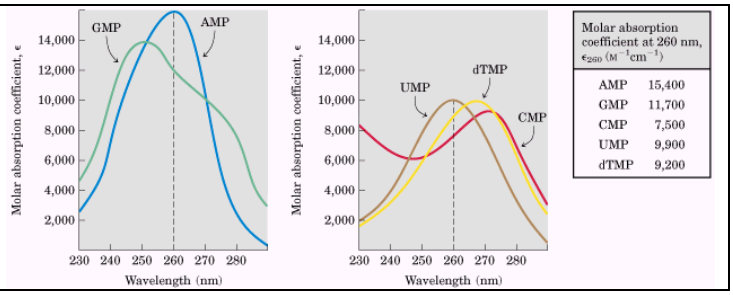


Under normal conditions, the phosphodiester linkages in both DNA and RNA are relatively stable without enzymatic catalysis. Under basic condition, the backbone of DNA is more stable than RNA, because the 2'-OH group plays a critical role in the process of nucleic acid backbone cleavage. (Why?) A sample is shown above.

The properties of nucleotide bases affect the three-dimensional structure of nucleic acids

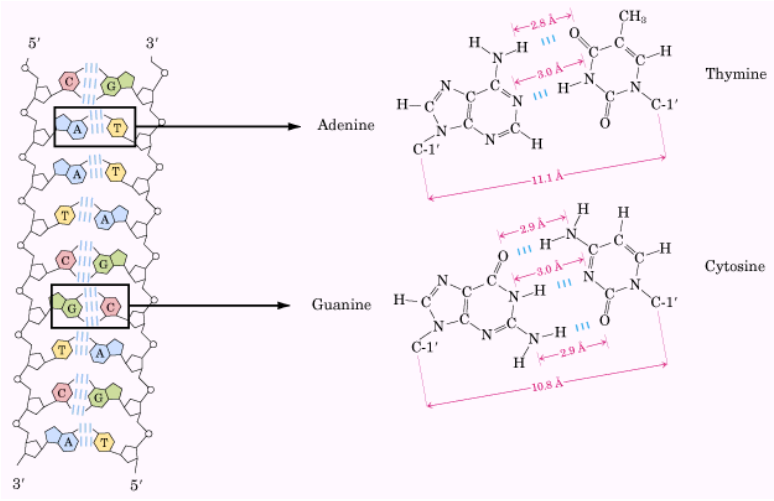
Free purines and pyrimidines are weakly basic (why?) compounds and are thus called bases. They are highly conjugated molecules, consequently, they exist as a variety of resonance forms, and the resonance among atoms in the ring gives most of the bonds partial double-bond character. Shown bellow is the example of uracil, in the form of **lactam**, **lactim** and double **lactim**. Because of this kind of conjugation, each base shows strong absorption at wavelength of 260 nm (What is the difference from some amino acids that also have strong UV absorption?). The UV absorption spectra are shown below.





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 are one of two important modes of interaction between bases in nucleic acids. (what is the other one?) The stacking also involves a combination of van der Waals and dipole-dipole interactions between the bases (what is the role of this kind of stacking?).

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. A sample proposed by Watson and Crick is shown below. Remember these bases such as **A-T**, **G-C** are the predominated form of the possible resonance structure in these conjugated base rings.



Nucleic Acid Structure

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**.
The complex folding of large chromosomes within eukaryotic chromatin and bacterial nucleoids is generally considered **tertiary structure**.

(Compare these with the structure of proteins)

DNA store genetic information

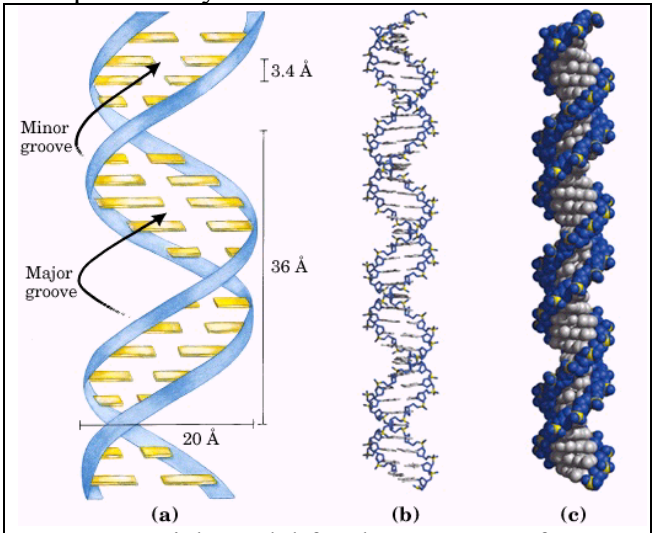
Understand the Avery-MacLeod-McCarty experiment(1944) and Hershey-Chase Experiment(1952)
DNA molecules have distinctive base compositions

Erwin Chargaff rules:

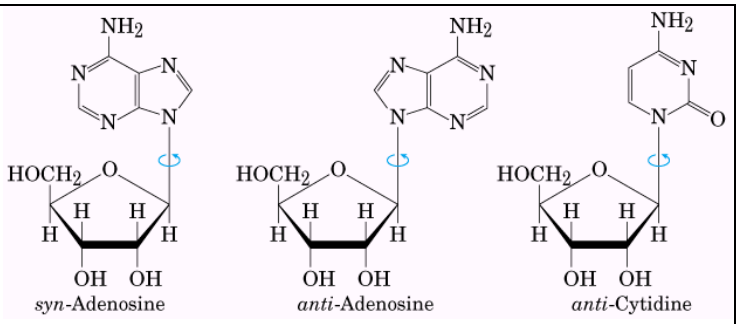
a. The base composition of DNA generally varies from one species to another.
b. DNA specimens isolated from different tissues of the same species have the same base composition.
c. The base composition of DNA in a given species does not change with an organism's age, nutritional state, or changing environment.
d. In all cellular DNAs, regardless of the species, the number of adenosine residues is equal to the number of thymidine residues (that is, A = T), and the number of guanosine residues is equal to the number of cytidine residues (G = C). from these relationships it follows that the sum of the purine residues equals the sume of the Pyrimidine residues; that is, A + G = T + C.

DNA is a double helix

As shown in the following figure, two helical DNA chains wound around the same axis to form a right-handed double helix. The hydrophilic backbones of alternating Deoxyribose and phosphate groups are on the outside of the double helix, facing the surrounding water. The furanose ring of each Deoxyribose is in the C-2' endo conformation. The purine and pyrimidine bases of both strands are stacked inside the double helix, with their hydrophobic and nearly planar ring structures very close together and perpendicular to the long axis. The offset pairing of the two strands creates a major groove and minor groove on the surface of the duplex. Each nucleotide base of one strand is paired in the same plane with a base of the other strand. It is important to note that three hydrogen bonds can form between G and C, symbolized G≡C, but only two can form between A and T, symbolized A=T. This is one reason for the finding that separation of paired DNA strand is more difficult the higher the ratio of G≡C to A=T base pairs. Remember that the strands of DNA are antiparallel, and 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.



Watson-Crick model for the structure of DNA (a) Schematic representation; (b) stick representation; and (c) space-filling model.

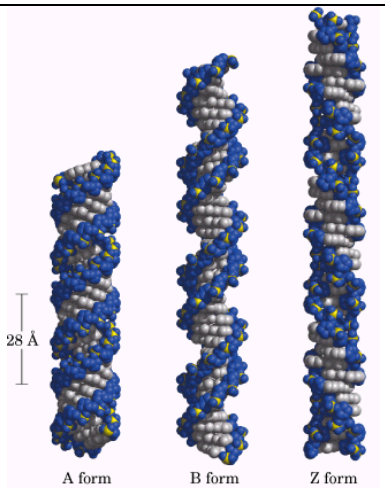


DNA can occur in different three-dimensional forms

DNA is a remarkably flexible molecule. Considerable rotation is possible around a number of bonds in the sugar-phosphate backbone, and thermal fluctuation can produce bending, stretching and unpairing of the strands. However, these structural variation generally do not affect the key properties of DNA, such as strand complementarity, antiparallel strands, requirement of A=T and G≡C base pairs.

Structure variation in DNA come from (a) the different possible conformation of the Deoxyribose, (b) rotation about the contiguous bonds that make up the phosphodeoxyribose backbone, and (c) free rotation about the C-1'-N-glycosyl bond. (Compare with the conformation restriction on proteins).

The Watson-Crick structure of DNA is also referred to as B-form DNA, or B-DNA. The B-form is the most stable structure for a random-sequence DNA molecule under physiological conditions and is therefore the standard point of reference. Apart from B-DNA, another two well-characterized DNA structure are A and Z forms. The comparison of these three forms are listed below.



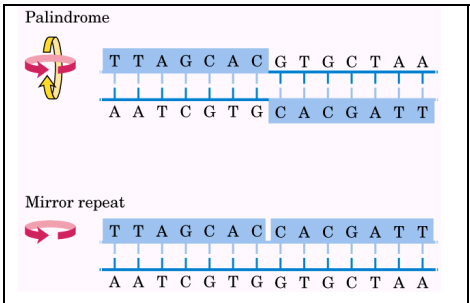
Comparison of forms A, B, and Z

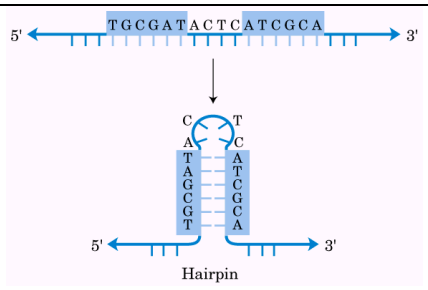
	A form	B form	Z form
Helical sense	Right handed	Right handed	Left handed
Diameter	~26 Å	~20 Å	~18 Å
Base pairs per helical turn	11	10.5	12
Helix rise per base pair	2.6 Å	3.4 Å	3.7 Å
Base tilt normal to the helix axis	20°	6°	7°
Sugar pucker conformation	C-3′ endo	C-2′ endo	C-2′ endo for pyrimidines; C-3′ endo for purines
Glycosyl bond conformation	Anti	Anti	Anti for pyrimidines; syn for purines

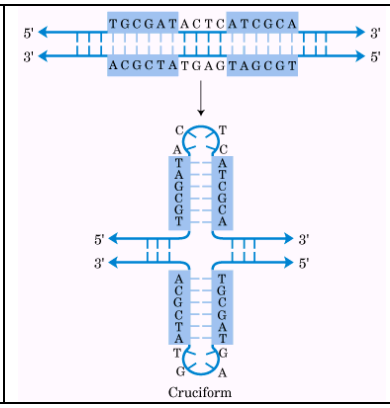
comparison of DNA form A, B, and Z.

Certain DNA sequences adopt unusual structures

A very common type of DNA sequence is a palindrome. A **palindrome** is a word, phrase, or sentence that is spelled identically reading forward or backward. The term is applied to regions of DNA with inverted repeats of base sequence having twofold symmetry over two strands of DNA. Such sequences are self-complementary within each strand and therefore have the potential to form **hairpin** or **cruciform** structures. When the inverted repeat occurs within each individual strand of the DNA, the sequence is called a mirror repeat. Mirror repeats cannot form hairpin or cruciform structures (why?)

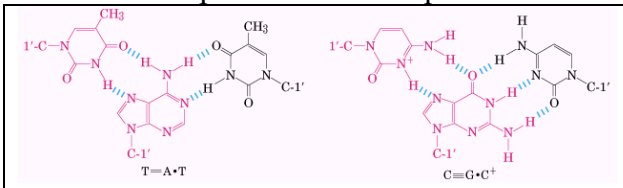




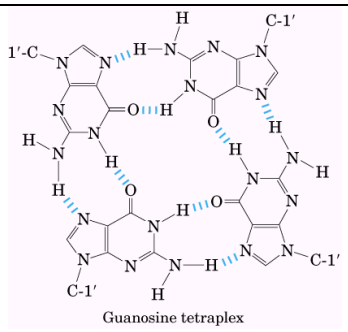


Besides the above conformation, more complicated conformations have been detected. For example, a cytidine residue (if protonated) can pair with the guanosine residue of a G≡C nucleotide pair, and a thymidine can pair with the adenosine of an A=T pair. The N-7, O6 of purines, the atoms that participate in the hydrogen bonding of triplex DNA, are often referred to as **Hoogsteen** positions, and the non-Watson-Crick pairing is called Hoogsteen pairing. In addition, four DNA strands can also pair to form a **tetraplex**, but this occurs readily only for DNA sequences with a very high proportion of guanosine residues (why?), and this guanosine or G tetraplex is quite stable over a wide range of conditions.

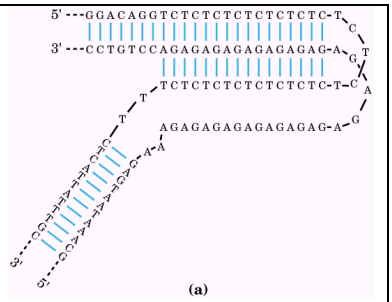
A particularly exotic DNA structure, known as H-DNA, is found in polypyrimidine or polypurine tracts that also incorporate a mirror repeat.



Base-pairing patterns in triplex DNA



Guanosine tetraplex



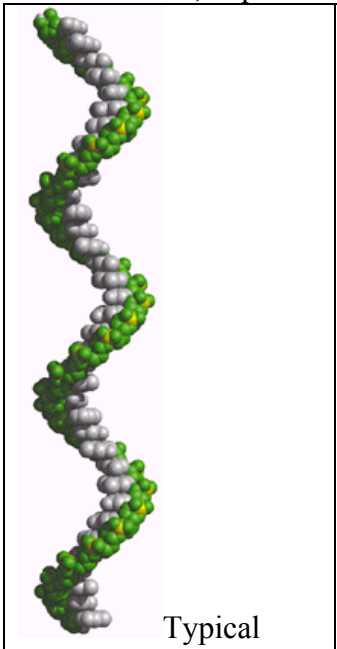
H-DNA

Messenger RNAs code for polypeptide chains

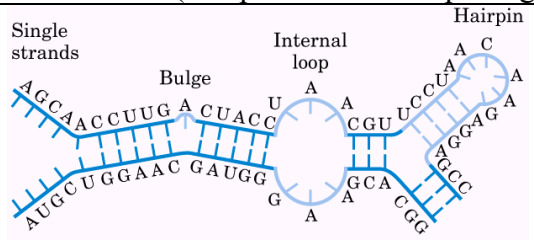
RNA is the molecule between DNA and protein, as evidenced by the following facts: RNA is found in both the nucleus and the cytoplasm, and an increase in protein synthesis is accompanied by an increase in the amount of cytoplasmic RNA and an increase in its rate of turnover. In prokaryotes a single mRNA molecule may code for one or several polypeptide chains. If it carries the code for only one polypeptide, the mRNA is **monocistronic**; if it codes for two or more different polypeptides, the mRNA is polycistronic. (cistron refers to a gene). In eukaryotes, most mRNAs are monocistronic (why?).

Many RNAs have more complex three-dimensional structures

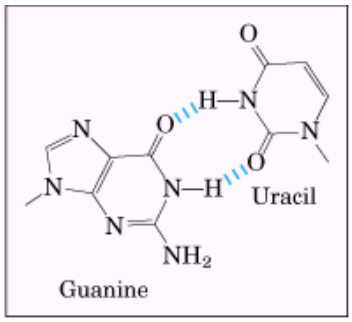
The product of transcription of DNA is always single-stranded RNA. The single strand tends to assume a right-handed helical conformation dominated by base-stacking interactions, which are stronger between two purines than between a purine and pyrimidine or between two pyrimidines (why?). The purine-purine Interaction is so strong that a Pyrimidine separating two purines is often displaced from the stacking pattern so that the purines can interact. Any self-complementary sequences in the molecule produce more complex structures. RNA can base-pair with complementary regions of either RNA or DNA. In RNA, the nucleotide G pairs with C, but A with U (remember RNA does not contains any T nucleotide). However, in some cases, G pairs with U in RNA (compare with DNA pairing).

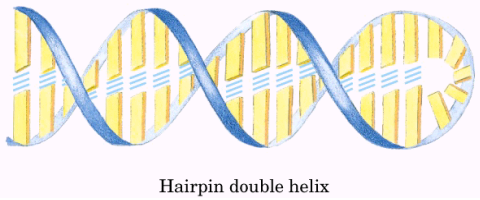


Typical right-handed stacking pattern of single-stranded RNA



Secondary structure of RNA. (a) Bulge, internal loop, hairpin loop.





Hairpin double helix

RNA has no simple, regular secondary structure that serves as a reference point, as does the double helix for DNA. The three-dimensional structure of many RNAs, like those of proteins, are complex and unique. Weak interactions, especially base-stacking interactions, play a major role in stabilizing RNA structures, just as they do in DNA. Where complementary sequence are present, the predominant double-stranded structure is an A-form right-handed double helix. Breaks in the regular A-form helix caused by mismatched or unmatched bases in one or both strands are common and result in bulge or internal loops. Hairpin loops form between nearby self-complementary sequences. Specific short base sequences are often found at the ends of RNA hairpins and are known to form particularly tight and stable loops. Such sequences may act as starting points for the folding of an RNA molecule into its precise three-dimensional structure.

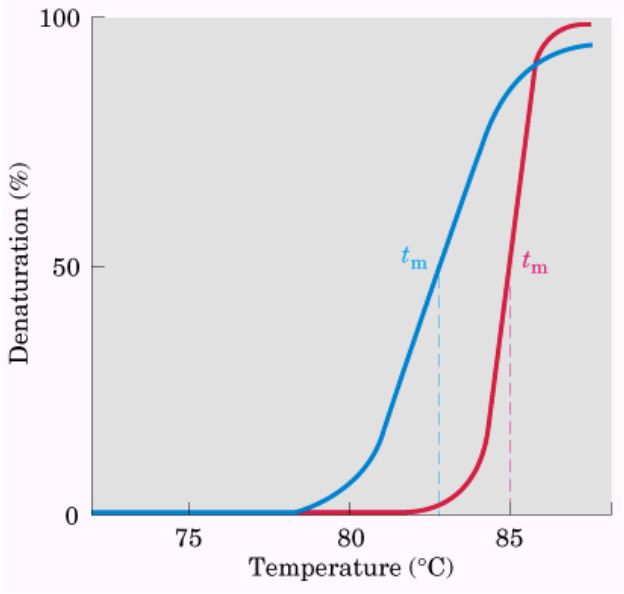
Nucleic Acid Chemistry

Double-helical DNA and RNA can be denatured

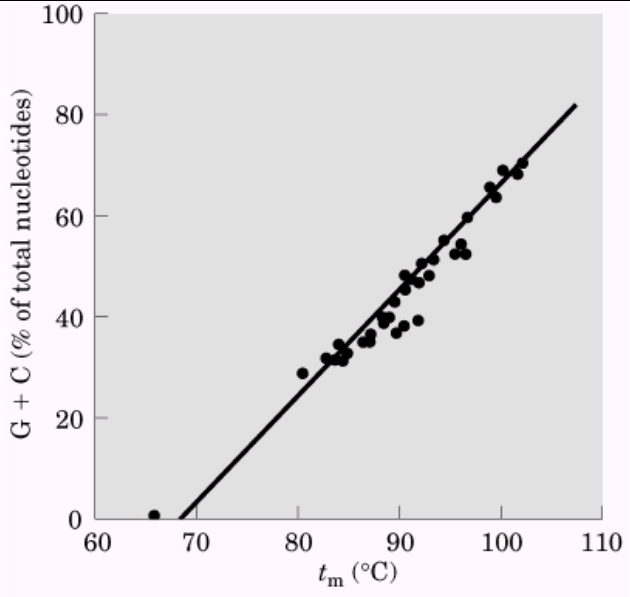
Just like in protein, the double-helical DNA and RNA can be denatured also. During the denaturation of DNA, the hydrogen bonds between paired bases and the base stacking are disrupted to unwind the double helix and form two single strands completely or partially. Renaturation of a DNA molecule is a rapid 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.

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 are paired. This is called the **hypochromic** effect. Denaturation of a double-stranded nucleic acid produces the opposite result, an increase in absorption called the **hyperchromic** effect. The transition from double-stranded DNA to the single-stranded, denatured form can thus be detected by monitoring the absorption of UV light.

Each species of DNA has a characteristic denaturation temperature or melting point (tm): the higher its content of G≡C base pairs, the higher the melting point of the DNA (why?). Thus careful determination of the melting point of a DNA specimen, under fixed conditions of pH and ionic strength, can yield an estimate of its base composition. If denaturation conditions are carefully controlled, regions that are rich in A=T base pairs will specifically denature while most of the DNA remains double-stranded. Such denatured regions (called bubbles) can be visualized with electron microscopy.
Duplexes of two RNA strands or one RNA strand and one DNA strand (RNA-DNA hybrids) can also be denatured. Notably, RNA duplexes are more stable than DNA duplexes. At neutral pH, denaturation of a double-helical RNA often requires temperatures 20°C or more higher than those required for denaturation of a DNA molecule with a comparable sequence. The temperature effect and the composition of base pairs on the melting point of DNA are shown in the following figures.



Different DNA has its intrinsic melting point



Relationship between tm and G≡C content of DNA

Nucleic acids from different species can form hybrids

When the DNA molecules from different species are denatured and mixed, when they are anneal, the mixed DNA double-strands will form, where the individual DNA strand in the double-strands are from different species, and this kind of double-strands are called hybrid duplexes.

Nucleotides and nucleic acids undergo nonenzymatic transformations

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 and the processes of aging and carcinogenesis.

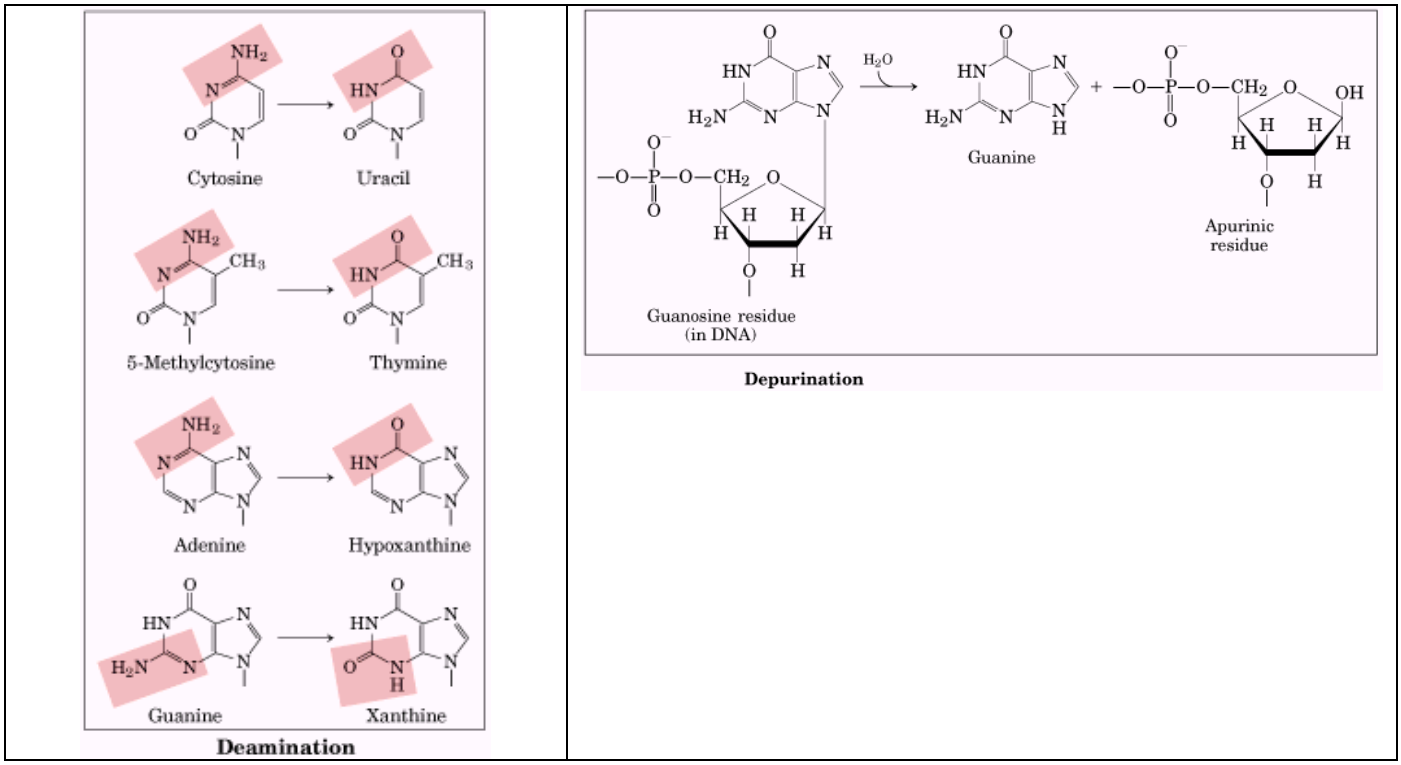
Demaination is a process when bases lose their exocyclic amino groups. The slow cytosine deamination reactions is the reason why DNA contains thymine rather than uracil.

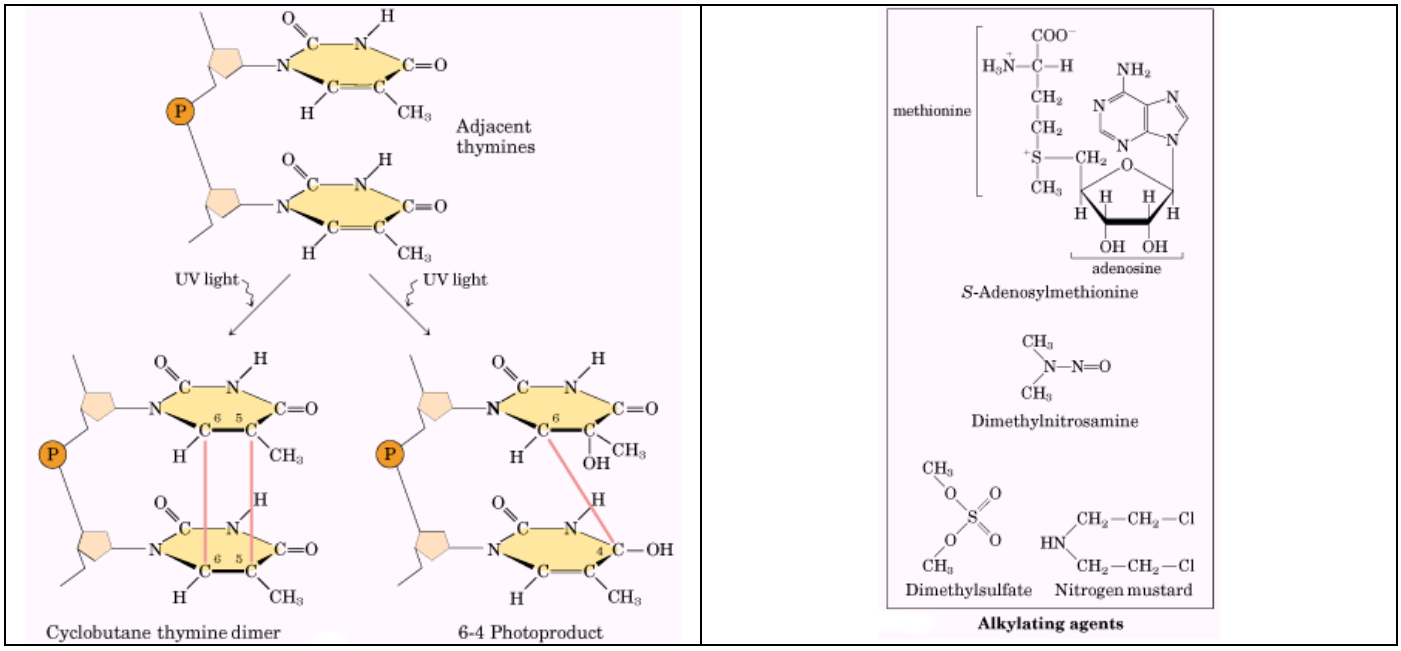
Another important reaction in deoxyribonucleotides is the **hydrolysis** of the N-β-glycosyl bond between the base and the pentose, as shown in below figures. This process occurs at a higher rate for purines than for pyrimidines.

Other reactions are promoted by radiation, such as from UV lignt, x-ray and gamma rays of cosmic rays, or other kind of radiation sources.

In addition, DNA also may be damaged by reactive chemicals introduced into the environment as products of industrial activity. Such as nitrous acid and alkylating agents. And possibly the most important source of mutagenic alteration in DNA is oxidative damage from radicals and hydrogen peroxide or superoxide radicals.

When the bases is changed, the genetic information will be changed also, and this is the reason for carcinogenesis.





Some bases of DNA are methylated

Certain nucleotide bases in DNA molecules are enzymatically methylated. Adenine and cytosine are methylated more often than guanine and thymine. Methylation is generally confined to certain sequences or regions of a DNA molecules. The methylation of nucleotide can serves as part of a defense mechanism that helps the cell to distinguish its DNA from foreign DNA by marking its own DNA with methyl groups and destroying (foreign) DNA without the methyl group (restriction-modification system). The other system methylates adenosine residues within the sequence (5')gatc(3') TO N^6-methyladenosine, this is

mediated by the Dam (DNA adenine methylation) methylase, a component of a system that repairs mismatched base pairs.

The sequence of long DNA strands can be determined

Similar to the sequence of proteins, the sequence of DNA can be determined also. There are two sequencing methods, Maxam-Gilbert and Sanger methods. The techniques depend on an improved understanding of nucleotide chemistry and DNA metabolism, and on electrophoretic methods for separating DNA strands differing in size by only one nucleotide. Polyacrylamide is often used as the gel matrix when working with short DNA molecules (< a few hundred nucleotides), agarose is generally used for longer pieces of DNA.

For Sanger's method makes use of the mechanism of DNA synthesis by DNA **polymerases**. DNA polymerase requires both a primer (a short oligonucleotide strand), to which nucleotides are added, and a template strand to guide selection of each new nucleotide. In cells, the 3'-hydroxyl group of the primer reacts with an incoming deoxynucleotide triphosphate (dNTP) to form a new phosphodiester bond. However, when dideoxynucleoside triphosphate (ddNTP with different fluorescent tag) is added, and connected to the growing DNA strands, the growth of DNA strand is stopped because no 3'-hydroxyl group is available. So, during the DNA synthesis, all dideoxynucleotides of four deoxynucleotides are added into the solution, and a variety of DNA strand will be synthesized. And these strands will be separated by electrophoresis, and can be differentiated by only one nucleotide. Similarly, the smaller the DNA strand, the longer the strand moves on the electrophoretic gel. The bottom spot is corresponding to the first nucleotide on the 5' end, and the whole sequence will be read out by automation. Because the synthesis is based on the real DNA template, the sequence obtained directly from machine reading is the complementarity of real DNA sequence, so this result will be converted into the real DNA sequence in the end.

The chemical synthesis of DNA has been automated

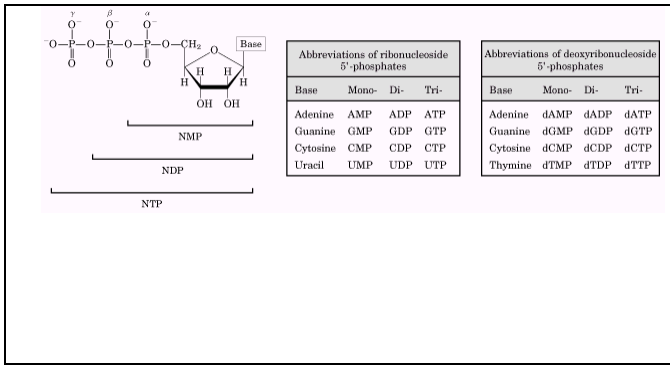
Similarly, the DNA can be synthesized automatically as the automatic synthesis of proteins.

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 triphosphate. Starting from the ribose, the three phosphates are generally labeled α, β, and γ. Hydrolysis of nucleoside triphosphates provides the chemical energy to drive a wide variety of biochemical reactions. These nucleotides are listed below, of them ATP is the most popular one to use as energy carrier.

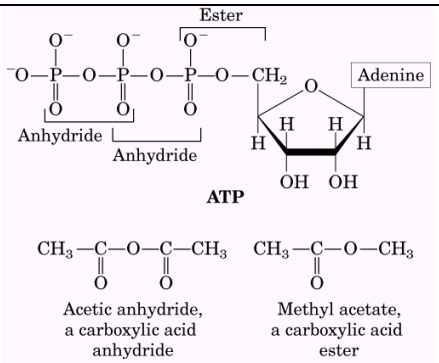


Abbreviations of ribonucleoside 5'-phosphates

Base	Mono-	Di-	Tri-
Adenine	AMP	ADP	ATP
Guanine	GMP	GDP	GTP
Cytosine	CMP	CDP	CTP
Uracil	UMP	UDP	UTP

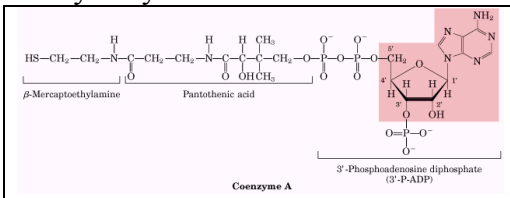
Abbreviations of deoxyribonucleoside 5'-phosphates

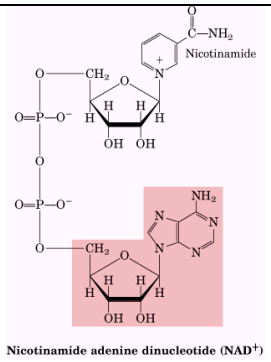
Base	Mono-	Di-	Tri-
Adenine	dAMP	dADP	dATP
Guanine	dGMP	dGDP	dGTP
Cytosine	dCMP	dCDP	dCTP
Thymine	dTMP	dTDP	dTTP

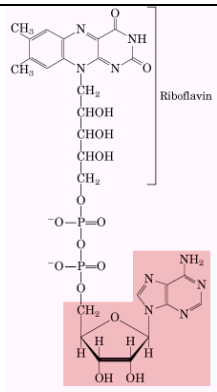


Adenine nucleotides are components of many enzyme cofactors

A variety of enzyme cofactors serving a wide range of chemical functions include adenosine as part of their structure. Some of them are listed in figures shown below. A 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 regulatory 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 lead 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 adenylate cyclase, an enzyme associated with the inner face of the plasma membrane. Other nucleotides are **cGMP** and **ppGpp**, as shown below.