Organic Chemistry III

Mohammad Jafarzadeh

Faculty of Chemistry, Razi University

Organic Chemistry, *Structure and Function* **(7th edition)**

By *P. Vollhardt* and *N. Schore*, Elsevier, 2014

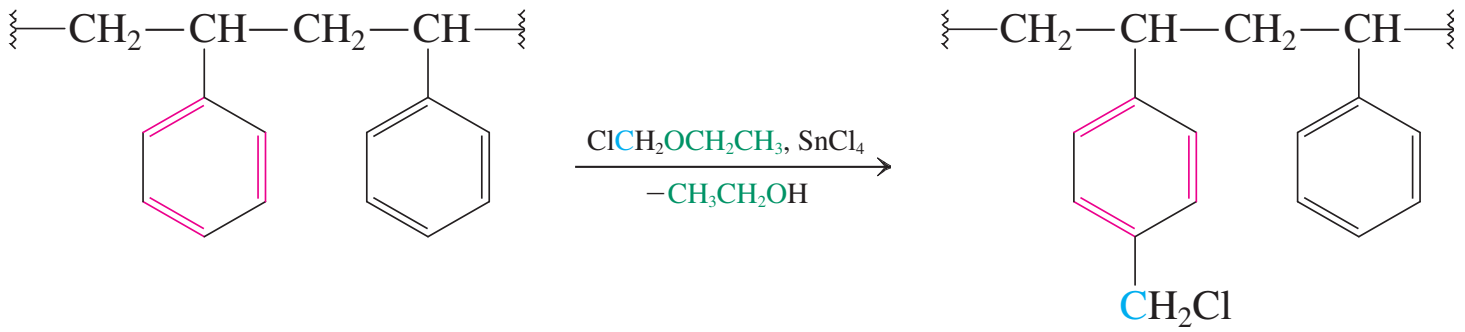
26-7 MERRIFIELD SOLID-PHASE PEPTIDE SYNTHESIS

Polypeptide synthesis has been automated. This ingenious method, known as the **Merrifield solid-phase peptide synthesis,** uses a solid support of polystyrene to anchor a peptide chain.

Although beads of polystyrene are insoluble and rigid when dry, they swell considerably in certain organic solvents, such as dichloromethane. The swollen material allows reagents to move in and out of the polymer matrix easily. Thus, its phenyl groups may be functionalized by electrophilic aromatic substitution.

For peptide synthesis, a form of Friedel-Crafts alkylation is used to chloromethylate a few percent of the phenyl rings in the polymer.

Electrophilic Chloromethylation of Polystyrene

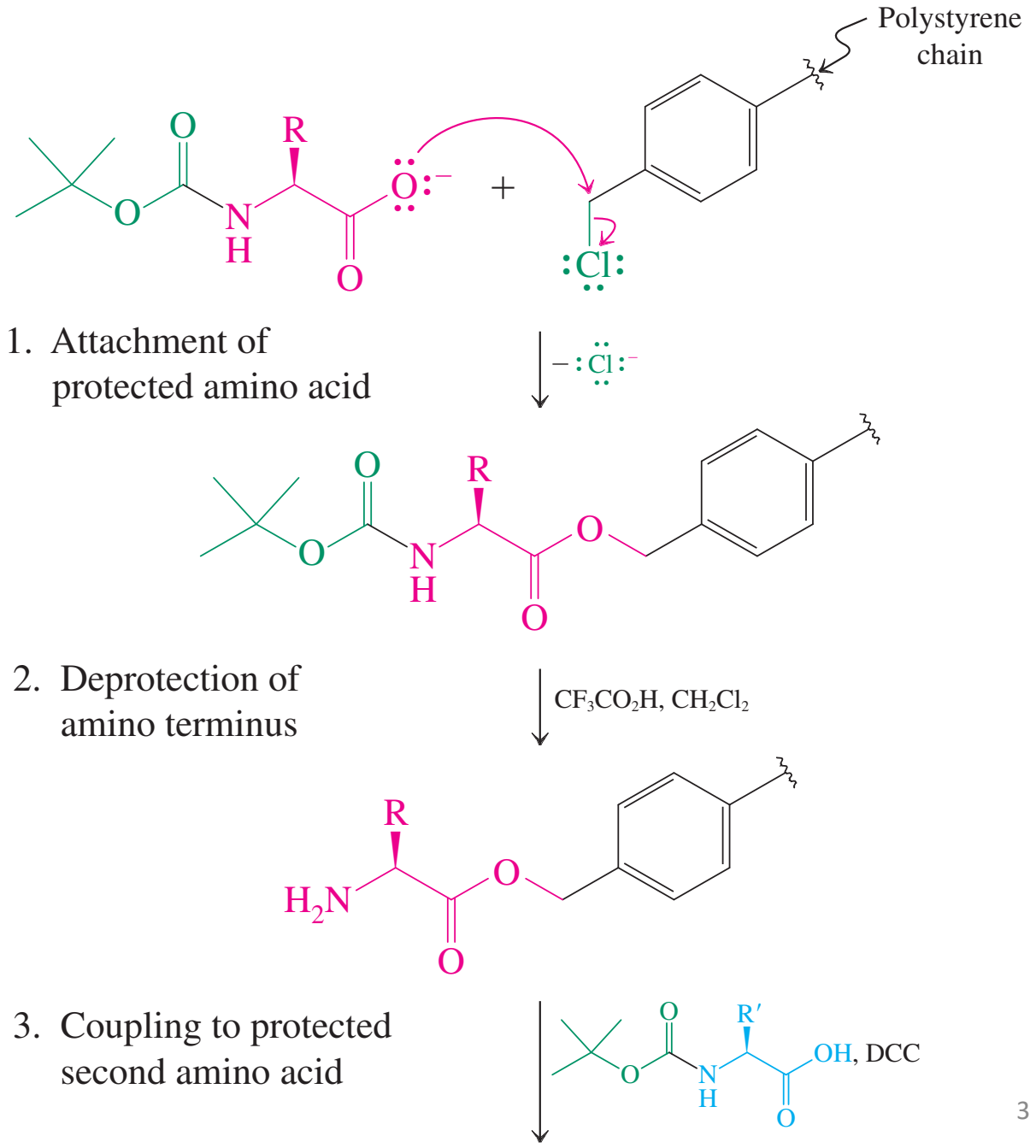


Polystyrene **Functionalized polystyrene**

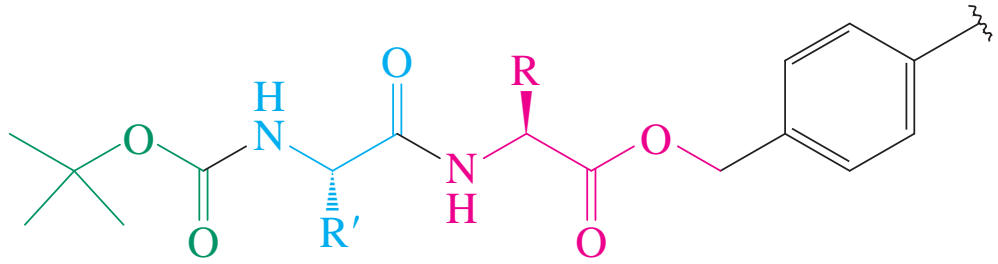
First an amino-protected amino acid is anchored on the polystyrene by nucleophilic substitution of the benzylic chloride by carboxylate.

Deprotection is then followed by coupling with a second amino-protected amino acid.

Solid-Phase Synthesis of a Dipeptide



Renewed deprotection and final removal of the dipeptide by treatment with hydrogen fluoride complete the sequence.

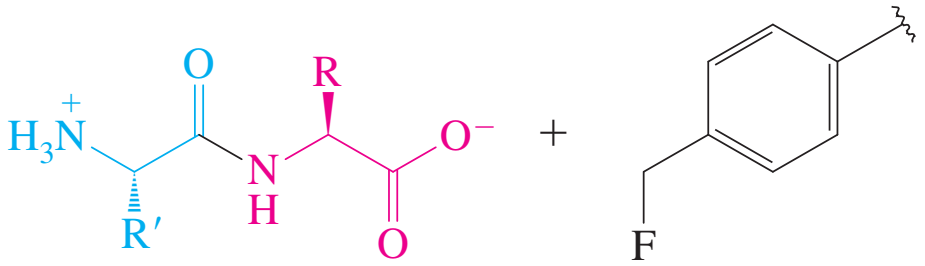


4. Deprotection of amino terminus

CF_3CO_2H, CH_2Cl_2

5. Disconnection of dipeptide from polymer

HF



The great advantage of solid-phase synthesis is the ease with which products can be isolated. Because all the intermediates are immobilized on the polymer, the products can be purified by simple filtration and washing.

It is not necessary to stop at the dipeptide stage. Repetition of the deprotection–coupling sequence leads to larger and larger peptides.

The first total synthesis of the protein insulin was accomplished by more than 5000 separate operations to assemble the 51 amino acids in the two separate chains; thanks to the automated procedure, this took only several days.

Automated protein synthesis has opened up exciting possibilities. First, it is used to confirm the structure of polypeptides that have been analyzed by chain degradation and sequencing.

Second, it can be used to construct synthetic proteins that might be more active and more specific than natural ones. Such proteins could be invaluable in the treatment of disease or in understanding biological function and activity.

26-8 POLYPEPTIDES IN NATURE: OXYGEN TRANSPORT BY THE PROTEINS MYOGLOBIN AND GEMOGLOBIN

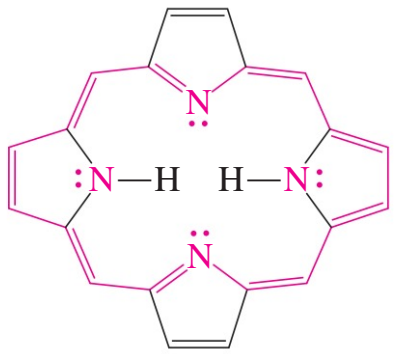
Two natural polypeptides function as oxygen carriers in vertebrates: the proteins myoglobin and hemoglobin.

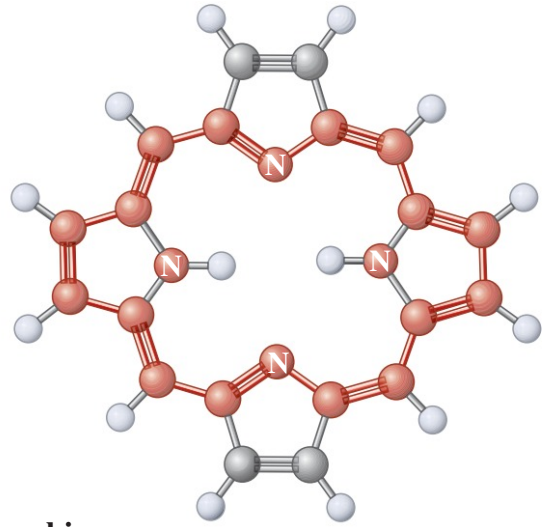
Myoglobin is active in the muscle, where it stores oxygen and releases it when needed. Hemoglobin is contained in red blood cells and facilitates oxygen transport. Without its presence, blood would be able to absorb only a fraction (about 2%) of the oxygen needed by the body.

How is the oxygen bound in these proteins?

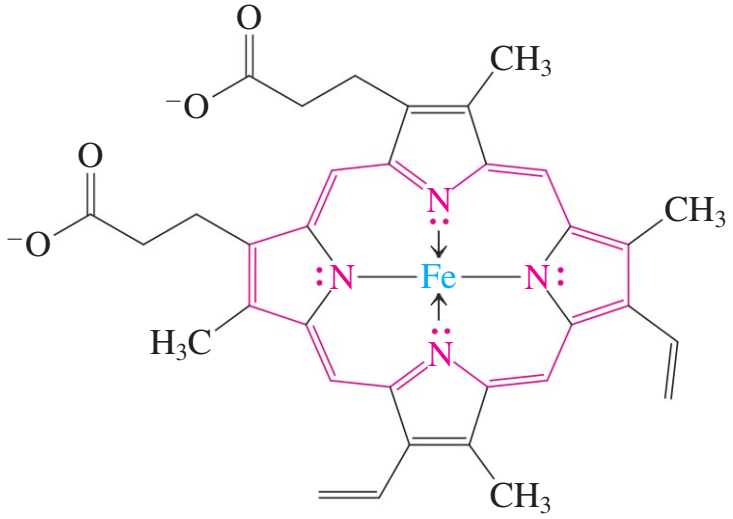
The secret of the oxygen-carrying ability of myoglobin and hemoglobin is a special nonpolypeptide unit, called a **heme group,** attached to the protein.

Heme is a cyclic organic molecule (called **porphyrin**) made of four linked, substituted pyrrole units surrounding an iron atom. The complex is red, giving blood its characteristic color.





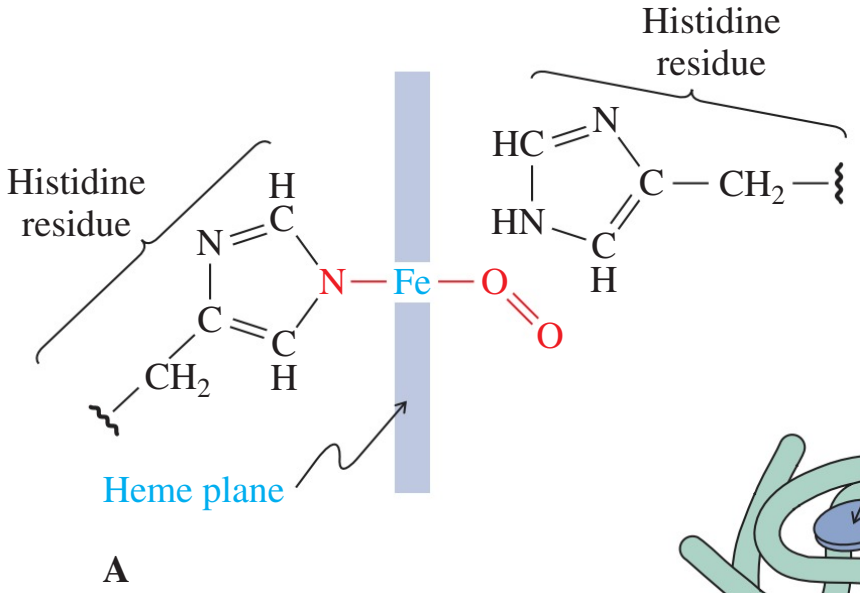
Porphine

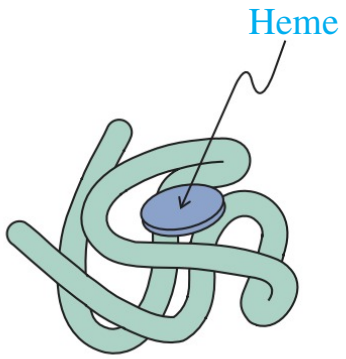


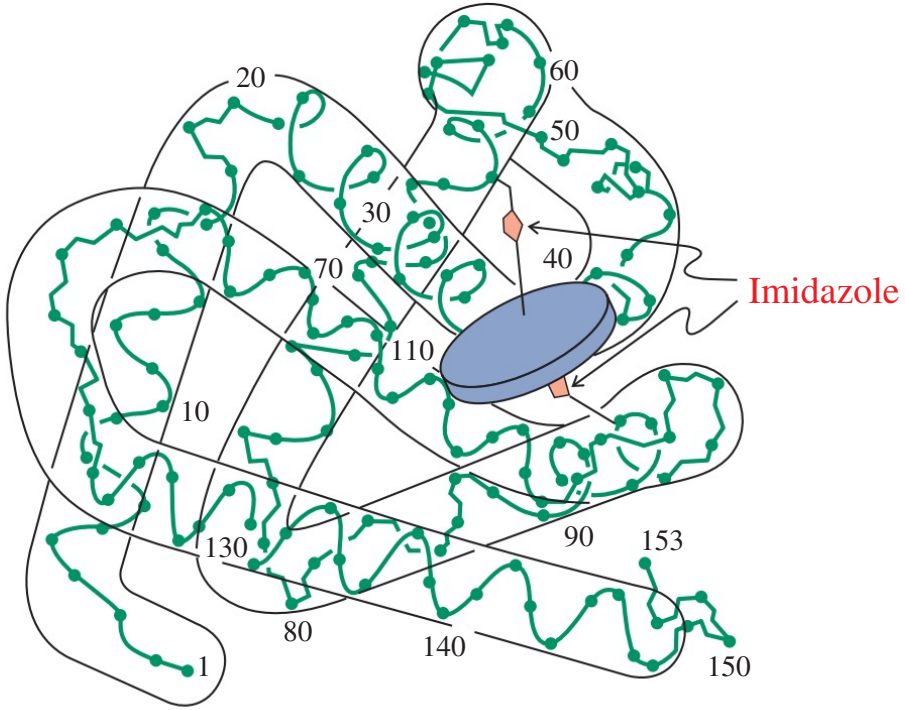
Heme

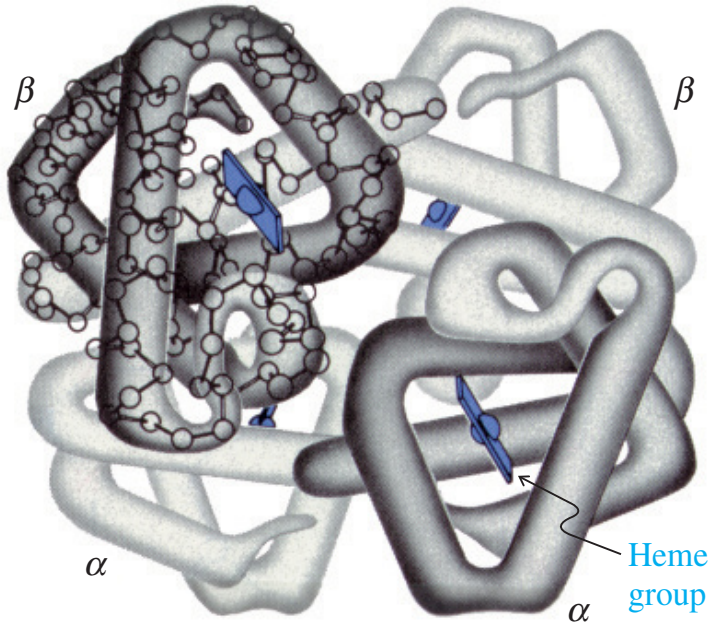
The iron in the heme is attached to four nitrogens but can accommodate two additional groups above and below the plane of the porphyrin ring. In myoglobin, one of these groups is the imidazole ring of a histidine unit attached to one of the α-helical segments of the protein.

The other is most important for the protein's function—bound oxygen. Close to the oxygen-binding site is a second imidazole of a histidine unit, which protects this side of the heme by steric hindrance.


Histidine residue
Histidine residue
HC
N
C — CH2 —
HN
C
H
N — Fe — O
O
Heme plane
A


Heme


20
60
50
30
40
70
Imidazole
110
10
90
153
130
80
140
1
150


β
β
α
α
Heme group

For example, carbon monoxide, which also binds to the iron in the heme group, and thus blocks oxygen transport, is prevented from binding as strongly as it normally would because of the presence of the second imidazole group. Consequently, CO poisoning can be reversed by administering oxygen to a person who has been exposed to the gas.

The two imidazole substituents in the neighborhood of the iron atom in the heme group are brought into close proximity by the unique folding pattern of the protein.

The rest of the polypeptide chain serves as a mantle, shielding and protecting the active site from unwanted intruders and controlling the kinetics of its action.

Myoglobin and hemoglobin offer excellent examples of the four structural levels in proteins. The primary structure of myoglobin consists of 153 amino acid residues of known sequence.

Myoglobin has eight α-helical segments that constitute its secondary structure, the longest having 23 residues. The tertiary structure has the bends that give myoglobin its three-dimensional shape.

Hemoglobin contains four protein chains: two α chains of 141 residues each, and two β chains of 146 residues each. Each chain has its own heme group and a tertiary structure similar to that of myoglobin.

There are many contacts between the chains; in particular, α_1 is closely attached to β_1, as is α_2 to β_2. These interactions give hemoglobin its quaternary structure.

The folding of the hemoglobin and myoglobin of several living species is strikingly similar even though the amino acid sequences differ. This finding implies that this specific tertiary structure is an optimal configuration around the heme group.

The folding allows the heme to absorb oxygen as it is introduced through the lung, hang on to it as long as necessary for transport, and release it when required.

26-9 BIOSYNTHESIS OF PROTEINS: NUCLEIC ACIDS

How does nature assemble proteins? The nature and workings of the genetic code.

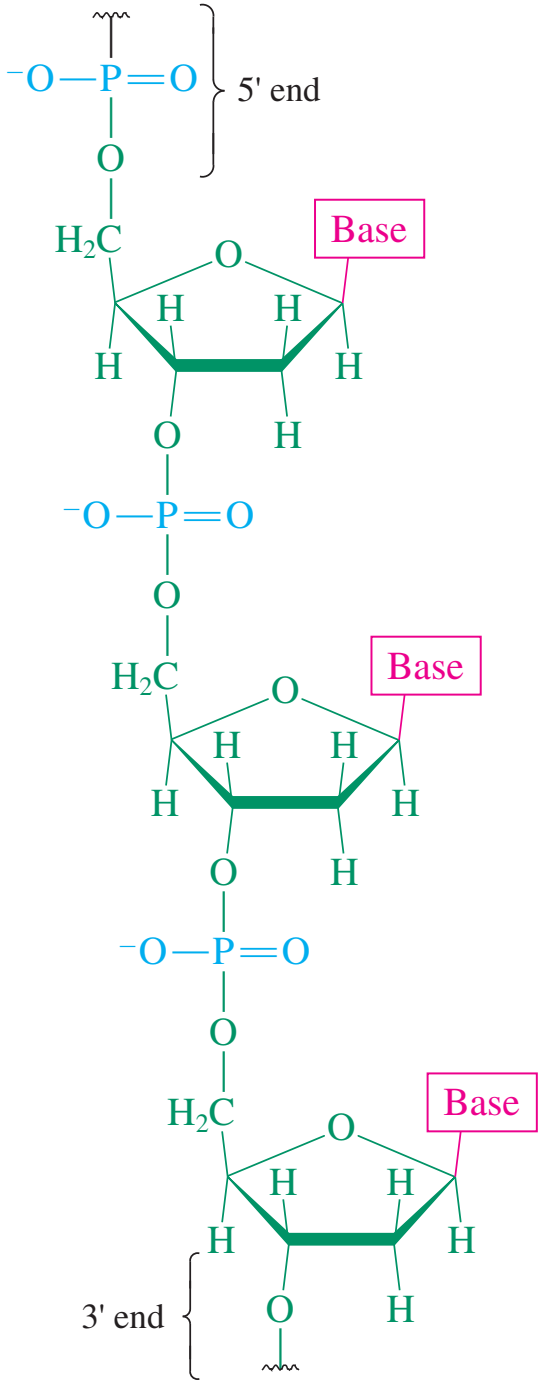
All hereditary information is embedded in the **deoxyribonucleic acids (DNA).** The expression of this information in the synthesis of all proteins, including the many enzymes necessary for cell function, is carried out by the **ribonucleic acids (RNA).**

After the carbohydrates and polypeptides, the nucleic acids are the third major type of biological polymer.

Four heterocycles define the structure of nucleic acids

The structures of DNA and RNA are simple. All their components, called **nucleotides,** are polyfunctional.

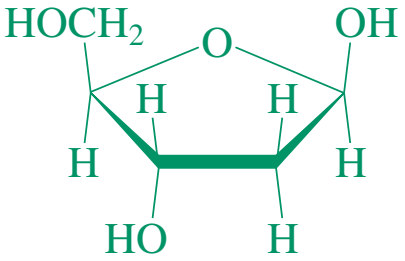
Nucleic acids are polymers in which phosphate units link sugars, which bear various heterocyclic nitrogen **bases**.



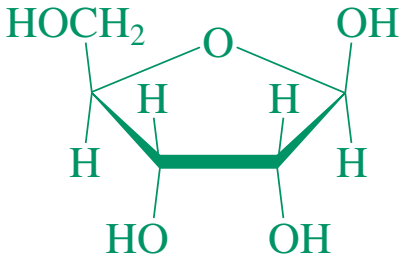
In DNA, the sugar units are 2-deoxyriboses, and only four bases are present: **cytosine (C), thymine (T), adenine (A),** and **guanine (G).**

In RNA, the sugar characteristic is ribose, and there are four bases, but the nucleic acid incorporates **uracil (U)** instead of thymine.

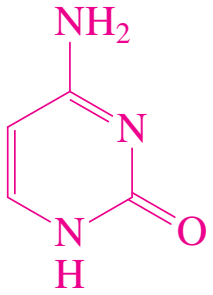
Nucleic Acid Sugars and Bases



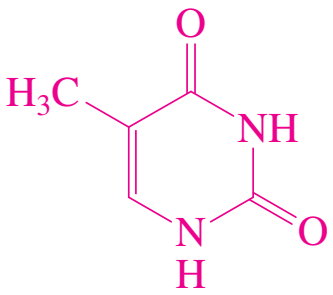
2-Deoxyribose



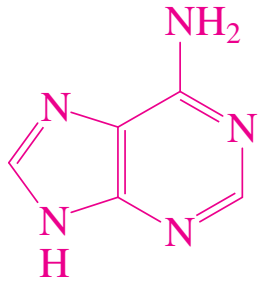
Ribose



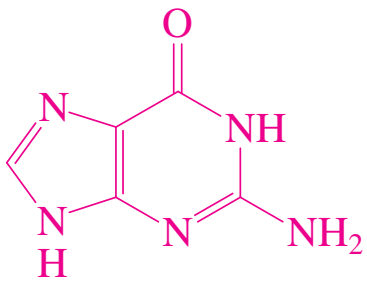
Cytosine (C)



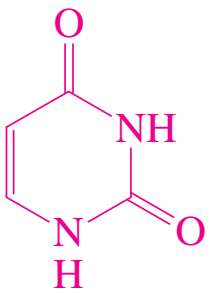
Thymine (T)



Adenine (A)



Guanine (G)

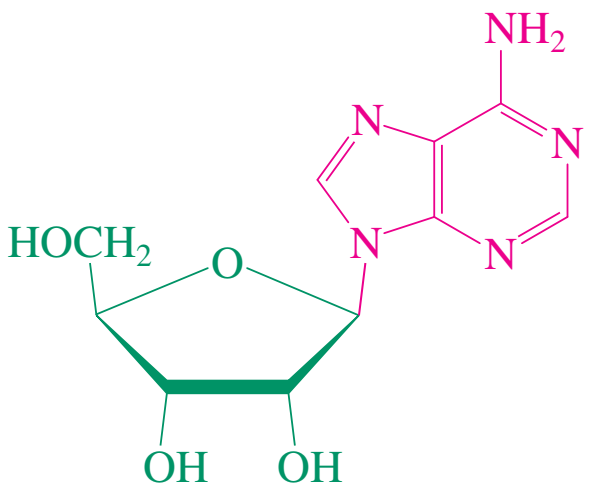


Uracil (U)

We construct a nucleotide from three components. First, we replace the hydroxy group at C1 in the sugar with one of the base nitrogens. This combination is called a **nucleoside.**

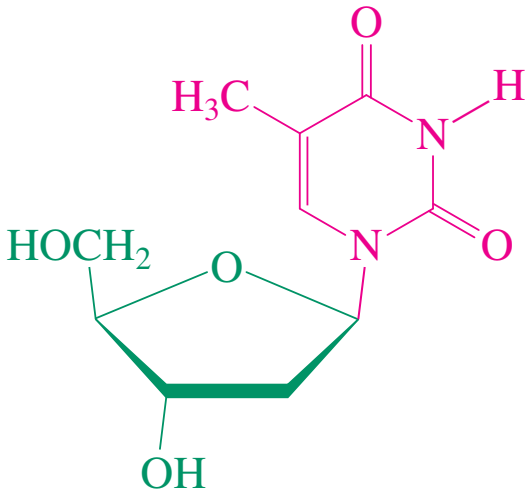
Second, a phosphate substituent is introduced at C5. In this way, we obtain the four nucleotides of both DNA and RNA.

The positions on the sugars in nucleosides and nucleotides are designated 1', 2', and so forth, to distinguish them from the carbon atoms in the nitrogen heterocycles.



Adenosine

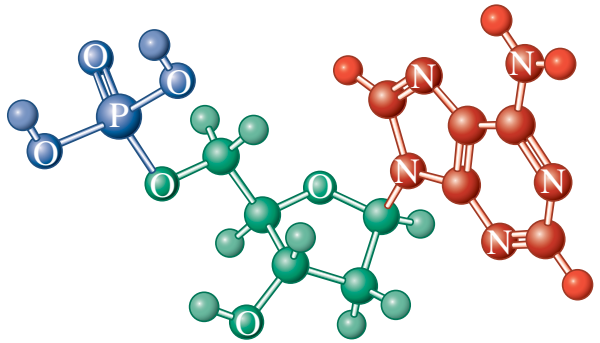
(A nucleoside)

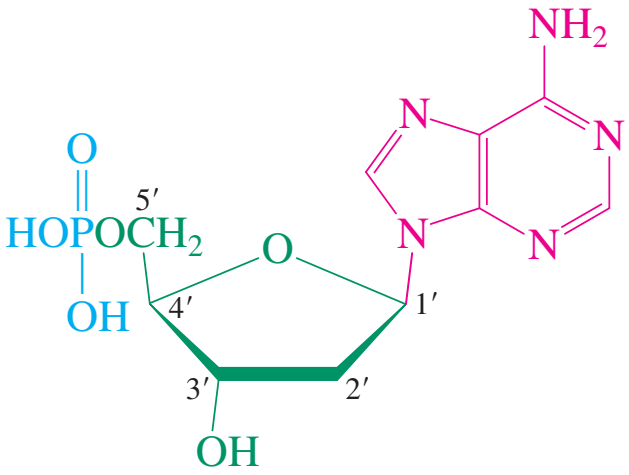


Thymidine

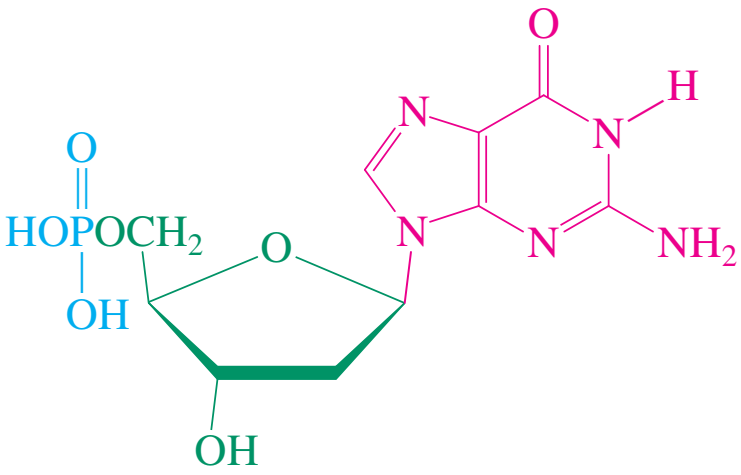
(A nucleoside)

Nucleotides of DNA

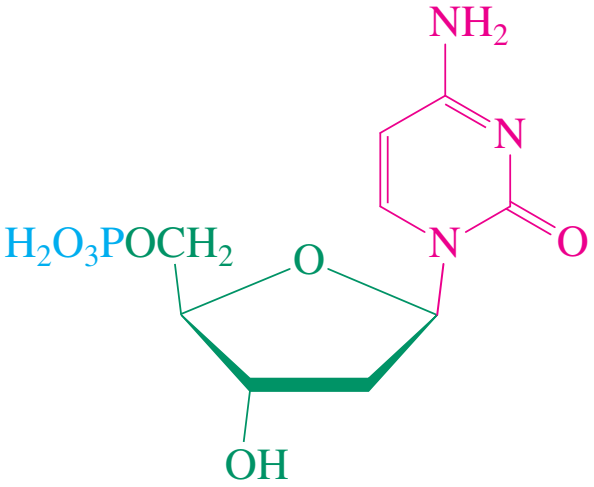




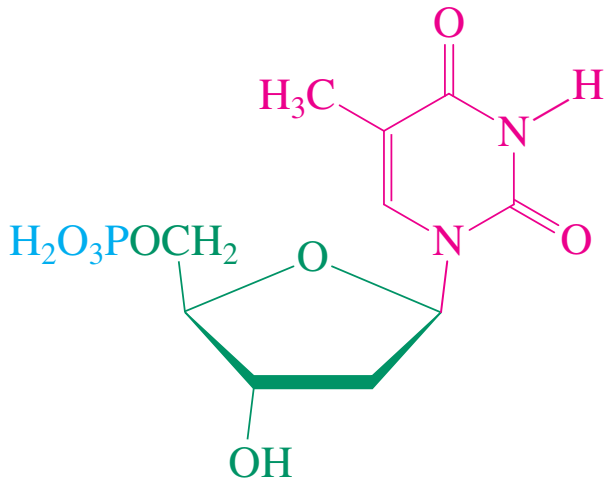
2′-Deoxyadenylic acid



2′-Deoxyguanylic acid

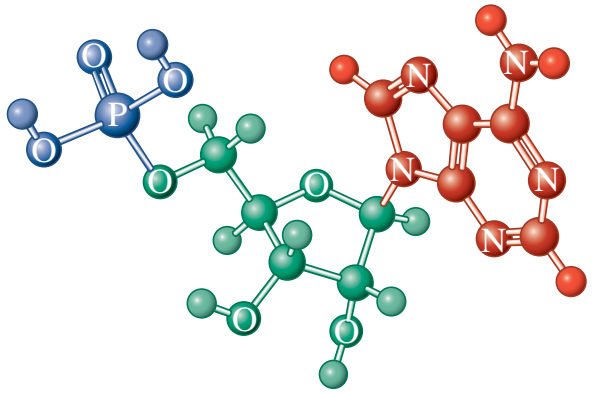


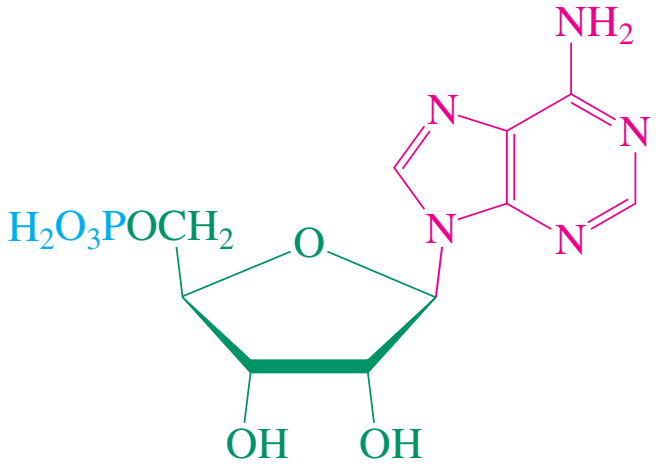
2′-Deoxycytidylic acid



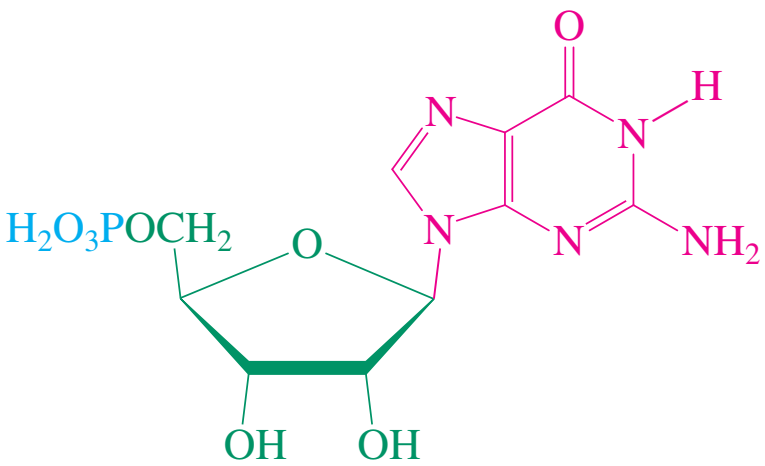
Thymidylic acid

Nucleotides of RNA

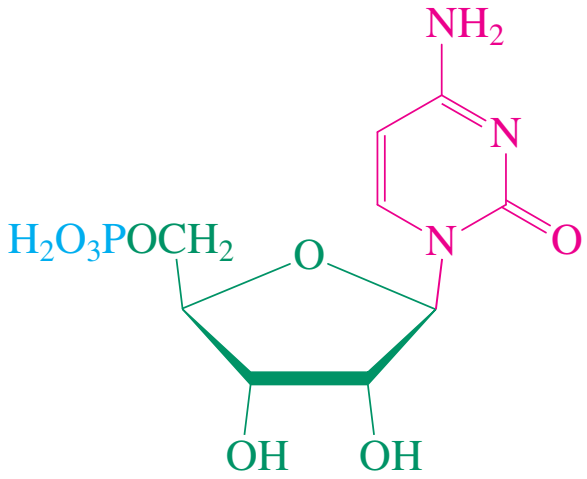




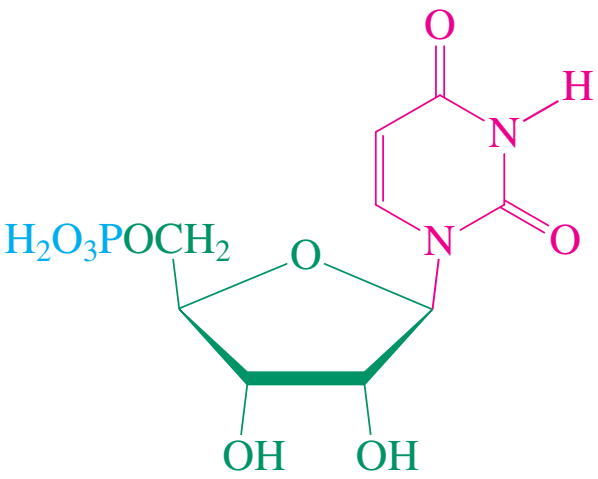
Adenylic acid



Guanylic acid



Cytidylic acid



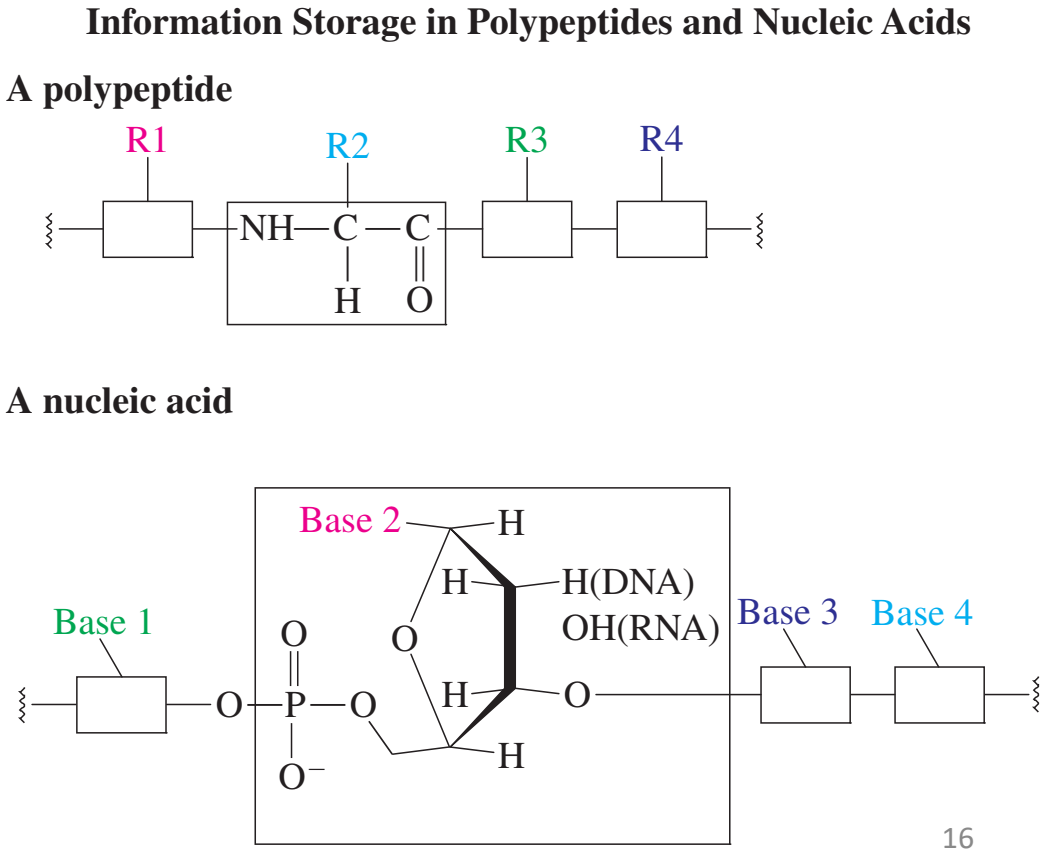
Uridylic acid

The polymeric chain is then readily derived by repeatedly forming a phosphate ester bridge from C5’ (called the **5’ end**) of the sugar unit of one nucleotide to C3’ (the **3’ end**) of another.

In this polymer, the bases adopt the same role as that of the 2-substituent in the amino acids of a polypeptide: Their sequence varies from one nucleic acid to another and determines the fundamental biological properties of the system.

Viewed from the perspective of storing information, polypeptides do so by using an amide polymer backbone along which differing side chains are attached, like the letters in a word.

In polypeptides there are 20 such letters for the 20 natural amino acids. In nucleic acids, through a sugar-phosphate polymeric array, exist a sequence of only four amine bases, with corresponding four letters: C, T, A, G for DNA and C, U, A, G for RNA.

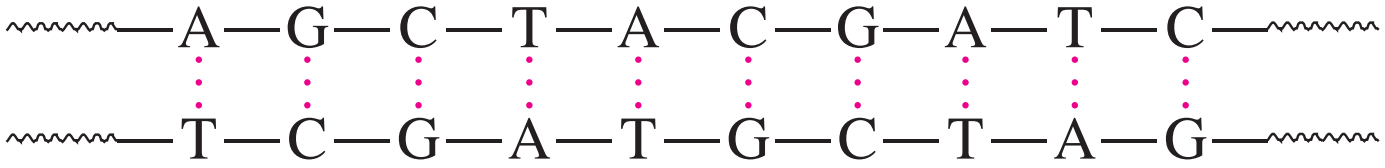


Nucleic acids form a double helix

Nucleic acids, especially DNA, can form extraordinarily long chains (as long as several centimeters) with molecular weights of as much as 150 billion. Like proteins, they adopt secondary and tertiary structures.

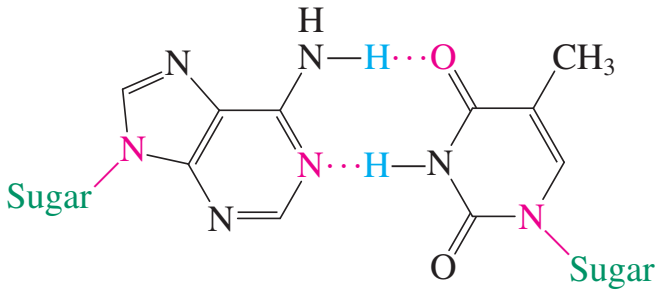
In 1953, Watson and Crick made their ingenious proposal that DNA is a double helix composed of two strands with complementary base sequences. In the DNA of various species, the ratio of adenine to thymine and guanine to cytosine is always one to one. It is concluded that two chains are held together by hydrogen bonding in such a way that adenine and guanine, respectively, in one chain always face thymine and cytosine in the other.

Thus, if a piece of DNA in one strand has the sequence -A-G-C-T-A-C-G-A-T-C-, this entire segment is hydrogen bonded to a complementary strand running in the opposite direction, -T-C-G-A-T-G-C-T-A-G-.

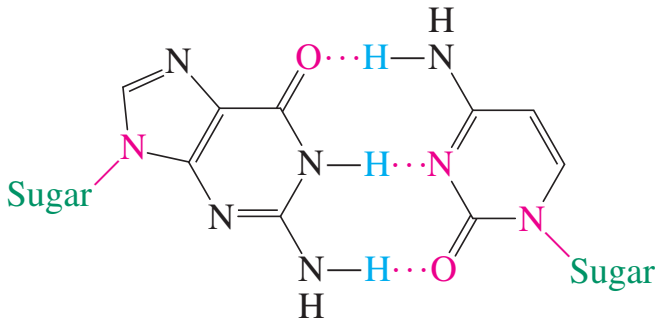


Because of other structural constraints, the arrangement that maximizes hydrogen bonding and minimizes steric repulsion is the **double helix**.

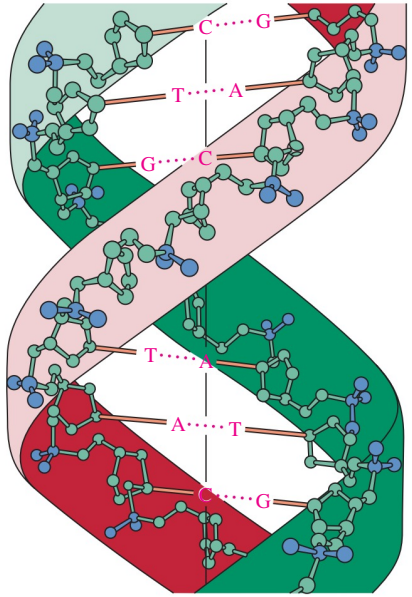
The cumulative hydrogen-bonding energies involved are substantial considering that just single base pairing is favorable by 5.8 kcal mol^{-1} (24 kJ mol^{-1}) for G–C and 4.3 kcal mol^{-1} (18 kJ mol^{-1}) for A–T.

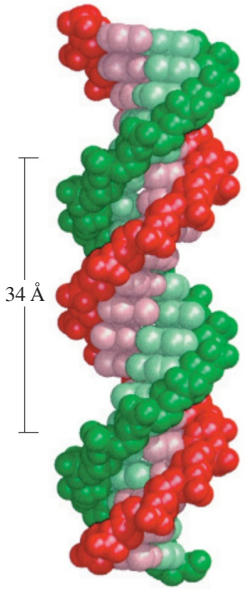


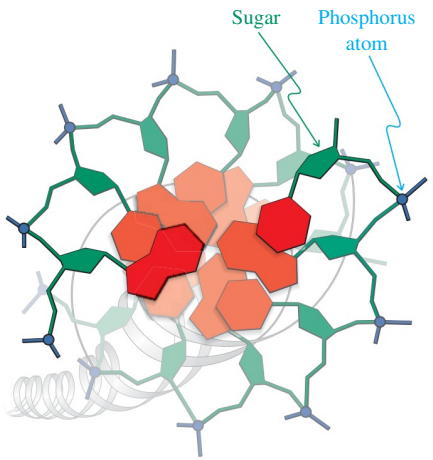
Adenine–thymine



Guanine–cytosine







DNA replicates by unwinding and assembling new complementary strands

There is no restriction on the variety of sequences of the bases in the nucleic acids.

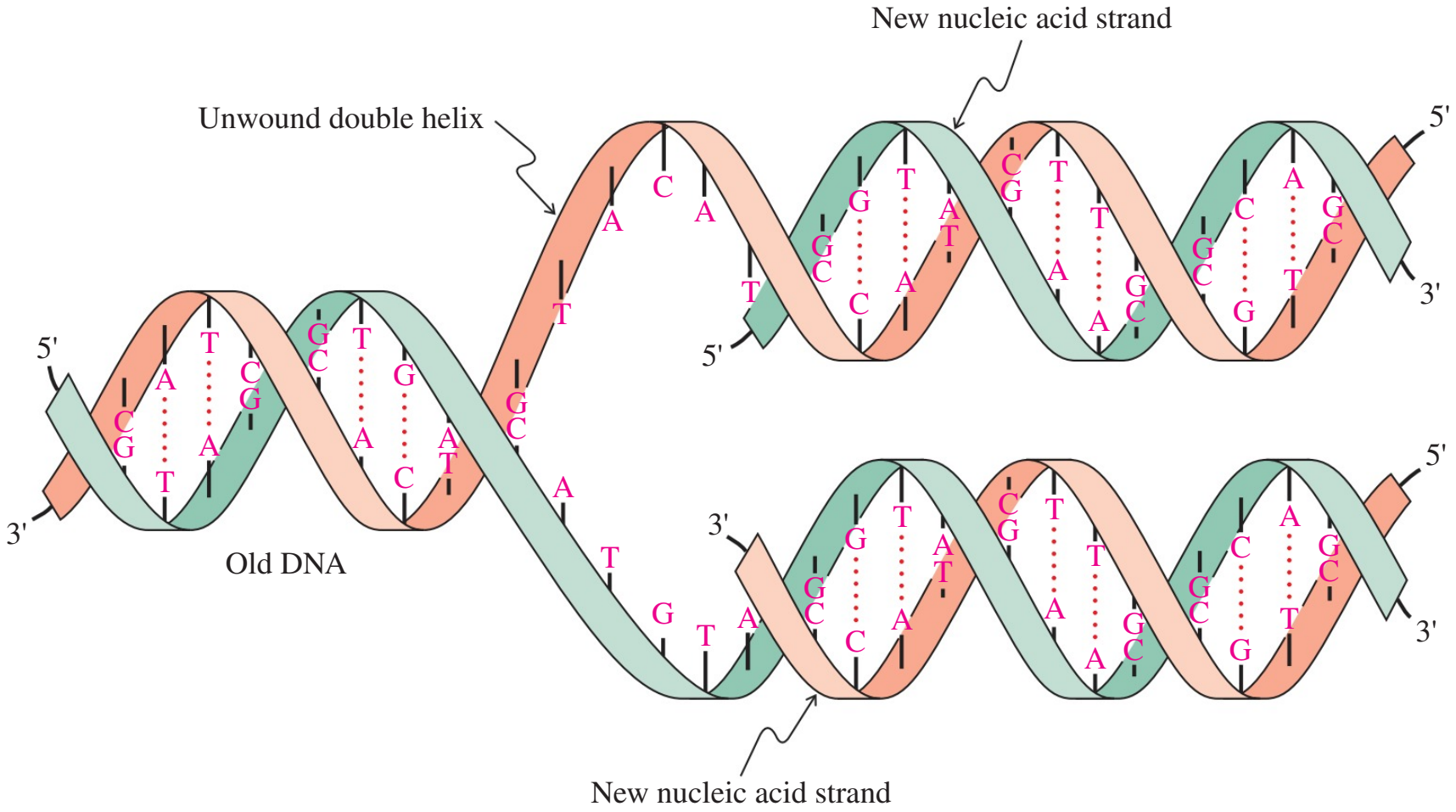
Watson and Crick proposed that the specific base sequence of a particular DNA contained all genetic information necessary for the duplication of a cell and, indeed, the growth and development of the organism as a whole.

The double-helical structure suggested a way in which DNA might **replicate**—make exact copies of itself—and so pass on the genetic code. In this mechanism, each of the two strands of DNA functions as a template.

The double helix partly unwinds, and enzymes called DNA polymerases then begin to assemble the new DNA by coupling nucleotides to one another in a sequence complementary to that in the template, always juxtaposing C to G and A to T.

Eventually, two complete double helices are produced from the original.

This process is at work throughout the entire human genetic material, or **genome**—some 2.9 billion base pairs—with an error frequency of less than 1 in 10 billion base pairs.



26-10 PROTEIN SYNTHESIS THROUGH RNA

The mechanism of duplicating the entire nucleotide sequence in DNA replication is used by nature and by chemists to obtain partial copies of the genetic code for various purposes.

In nature, the most important application is the assembly of RNA, called **transcription,** which transcribes the parts of the DNA that contain the information (the **genes**) necessary to synthesize proteins in the cell.

The process by which this transcribed information is decoded and used to construct proteins is called **translation.**

The three key players in protein synthesis are the "DNA transcript" **messenger RNA (mRNA),** the "delivery unit" for the specific amino acids to be connected by peptide bonds, **transfer RNA (tRNA),** and the catalyst that enables amide bond formation—the **ribosome.**

Mutations in the base sequence of DNA can be caused by physical (radiation) or chemical (carcinogens) interference. Mutations can either replace one base with another or can add or delete one base or more.