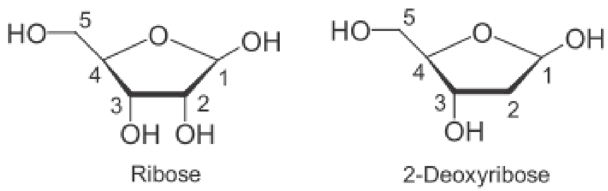
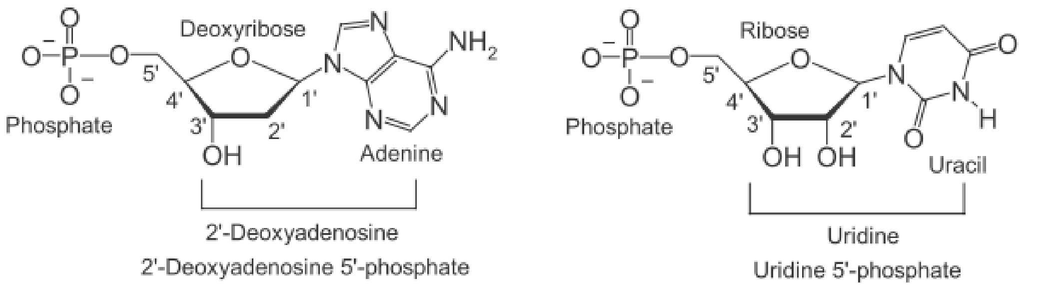
Nucleic acids

The nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), are the chemical carriers of a cell's genetic information. Nucleic acids are biopolymers made of *nucleotides* joined together to form a long chain. These biopolymers are often found associated with proteins, and in this form they are called *nucleoproteins*. Each nucleotide comprises a *nucleoside* bonded to a phosphate group, and each nucleoside is composed of an aldopentose sugar, ribose or 2-deoxyribose, linked to a heterocyclic purine or pyrimidine base (see Section 4.7).



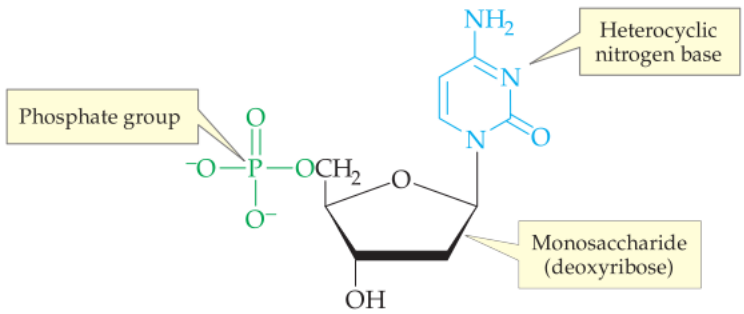
The sugar component in RNA is *ribose*, whereas in DNA it is *2-dexoyribose*. In deoxyribonucleotides, the heterocyclic bases are purine bases, adenine and guanine, and pyrimidine bases, cytosine and thymine. In ribonucleotides, adenine, guanine and cytosine are present, but not thymine, which is replaced by uracil, another pyrimidine base.

In the nucleotides, while the heterocyclic base is linked to C-1 of the sugar through an *N*-glycosidic β-linkage, the phosphoric acid is bonded by a phosphate ester linkage to C-5. When the sugar is a part of a nucleoside, the numbering of sugars starts with 1′, i.e. C-1 becomes C-1′, for example 2′-deoxyadenosine 5′-phosphate and uridine 5′-phosphate.

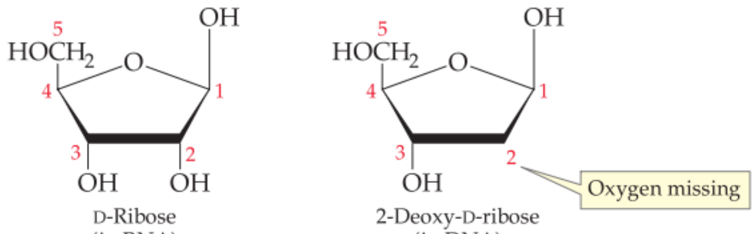


Despite being structurally similar, DNA and RNA differ in size and in their functions within a cell. The molecular weights of DNA, found in the nucleus of cells, can be up to 150 billion and lengths up to 12 cm, whereas the molecular weight of RNA, found outside the cell nucleus, can only be up to 35 000.

A nucleotide

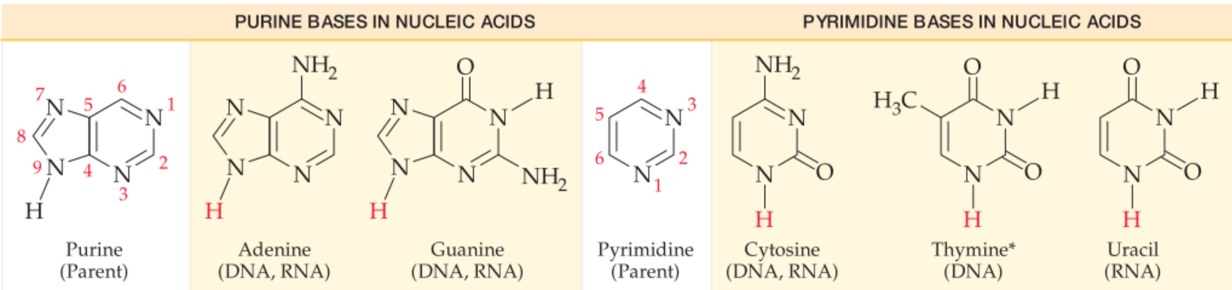


The Sugars



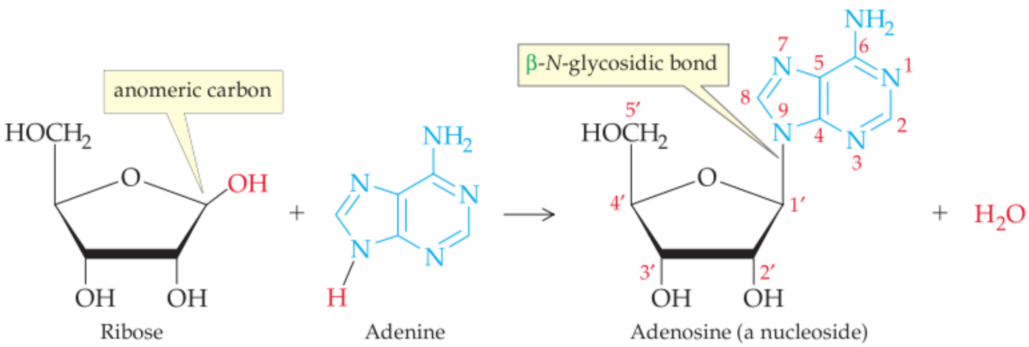
D-Ribose (in RNA)

2-Deoxy-D-ribose (in DNA)

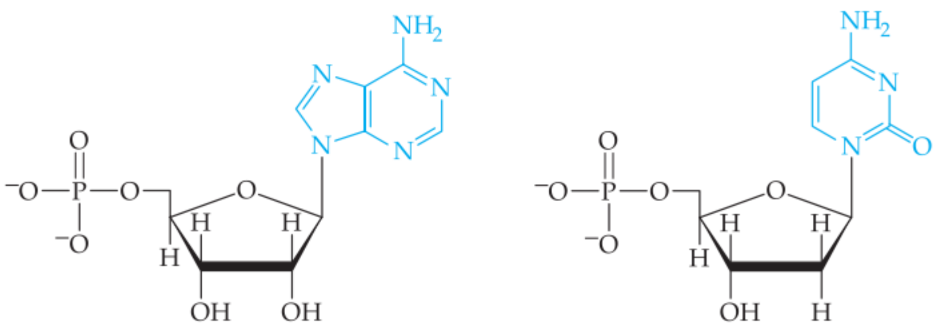


*Thymine occurs in a few cases of RNA.

Sugar + Base = Nucleoside



Nucleoside + Phosphate = Nucleotide



Adenosine 5′-monophosphate (AMP) (a ribonucleotide)

Deoxycytidine 5′-monophosphate (dCMP) (a deoxyribonucleotide)

Deoxyribonucleic acid (DNA)	
Name of the nucleotide	*Composition*
2′-Deoxyadenosine 5′-phosphate	Adenine + deoxyribose + phosphate Nucleoside is 2′-deoxyadenosine, composed of adenine and deoxyribose
2′-Deoxyguanosine 5′-phosphate	Guanine + deoxyribose + phosphate Nucleoside is 2′-deoxyguanosine, composed of guanine and deoxyribose
2′-Deoxycytidine 5′-phosphate	Cytosine + deoxyribose + phosphate Nucleoside is 2′-deoxycytidine, composed of cytosine and deoxyribose
2′-Deoxythymidine 5′-phosphate	Thymine + deoxyribose + phosphate Nucleoside is 2′-deoxythymidine, composed of thymine and deoxyribose

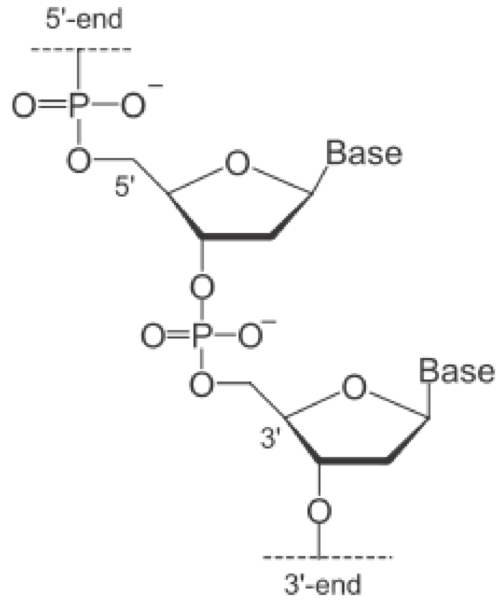
Ribonucleic acid (RNA)	
Adenosine 5′-phosphate	Adenine + ribose + phosphate Nucleoside is adenosine, composed of adenine and ribose
Guanosine 5′-phosphate	Guanine + ribose + phosphate Nucleoside is guanosine, composed of guanine and ribose
Cytidine 5′-phosphate	Cytosine + ribose + phosphate Nucleoside is cytidine, composed of cytosine and ribose
Uridine 5′-phosphate	Uracil + ribose + phosphate Nucleoside is uridine, composed of uracil and ribose

Structure of nucleic acids

Primary structure

Nucleotides join together in DNA and RNA by forming a phosphate ester bond between the 5′-phosphate group on one nucleotide and the 3′-hydroxyl group on the sugar (ribose or 2′-deoxyribose) of another nucleotide. In the nucleic acids, these phosphate ester links provide the nucleic acids with a long unbranched chain with a ‘backbone’ of sugar and phosphate units with heterocyclic bases sticking out from the chain at regular intervals. One end of the nucleic acid polymer has a free hydroxyl at C-3′ (the *3′-end*), and the other end has a phosphate at C-5′ (the *5′-end*).

The structure of nucleic acids depends on the sequence of individual nucleotides. The actual base sequences for many nucleic acids from various species are available to date. Instead of writing the full name of each nucleotide, abbreviations are used, e.g. A for adenine, T for thymidine, G for guanosine and C for cytidine. Thus, a typical DNA sequence might be presented as TAGGCT.



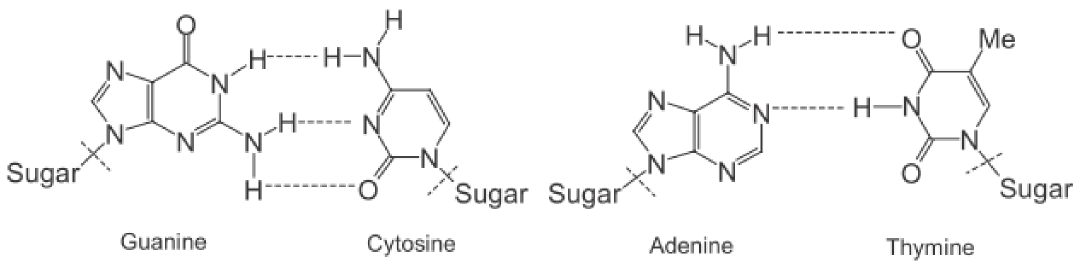
Generalized structure of DNA

Secondary structure: base pairing

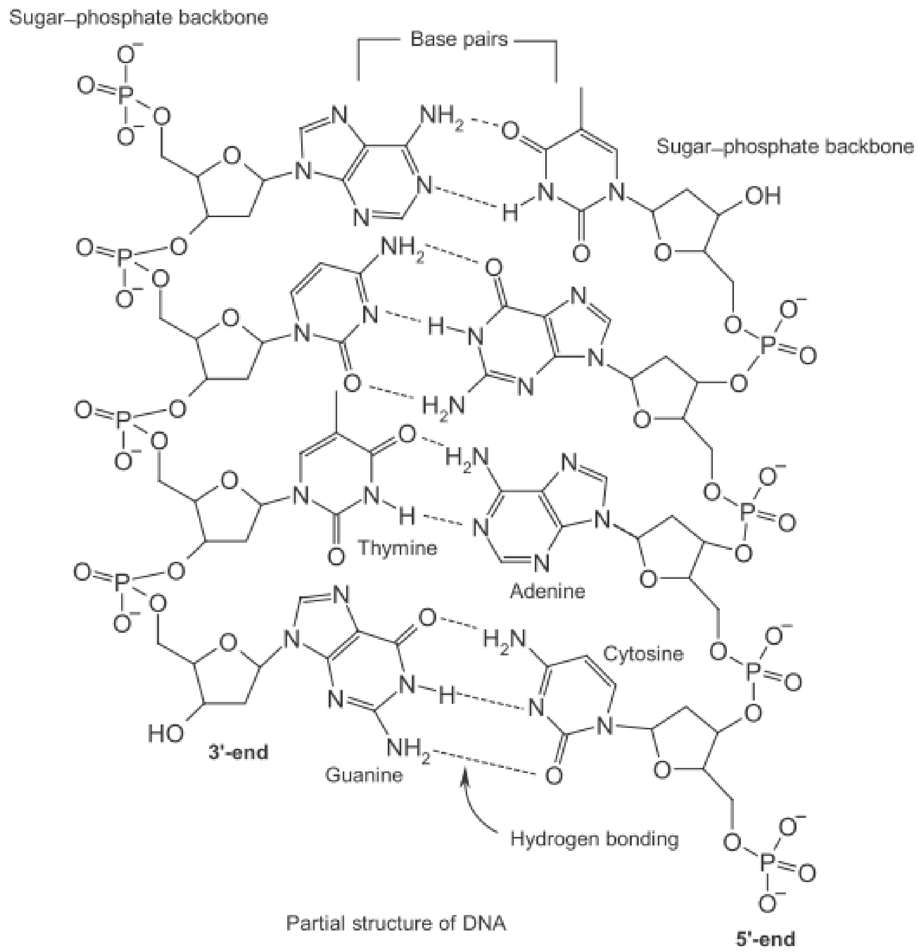
The base sequence along the chain of a DNA contains the genetic information. Samples of DNA isolated from different tissues of the same species have the same proportions of heterocyclic bases, but the samples from different species often have different proportions of bases. For example, human thymus DNA comprises 30.9% adenine, 29.4% thymine, 19.9% guanine and 19.8% cytosine, while the bacterium *Staphylococcus aureus* contains 30.8% adenine, 29.2% thymine, 21% guanine and 19% cytosine. In these examples, it is clear that the bases in DNA occur in pairs. Adenine and thymine are usually present in equal amounts; so are cytosine and guanine. In the late 1940s, E. Chargaff pointed out these regularities and summarized as follows.

(a) The total mole percentage of purines is approximately equal to that of the pyrimidines; i.e., $(\%G + \%A)/(\%C + \%T) \cong 1$.

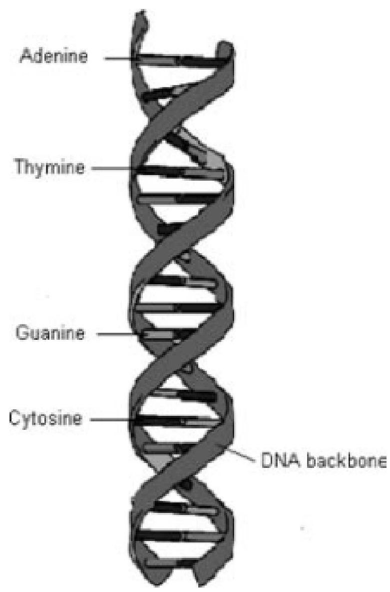
(b) The mole percentage of adenine is nearly equal to that of thymine, i.e $\%A/\%T \cong 1$, and same is true for guanine and cytosine, i.e. $\%G/\%C \cong 1$.



Hydrogen bonding between base pairs of the DNA double helix



While the sugar–phosphate backbone of DNA is completely regular, the sequence of heterocyclic base pairs along the backbone can be of different permutations. It is the precise sequence of base pairs that carries the genetic information.



Three fundamental processes are involved in the transfer of this stored genetic information.

- *Replication.* This process creates the identical copies of DNA, so that information can be preserved and handed down to offspring.

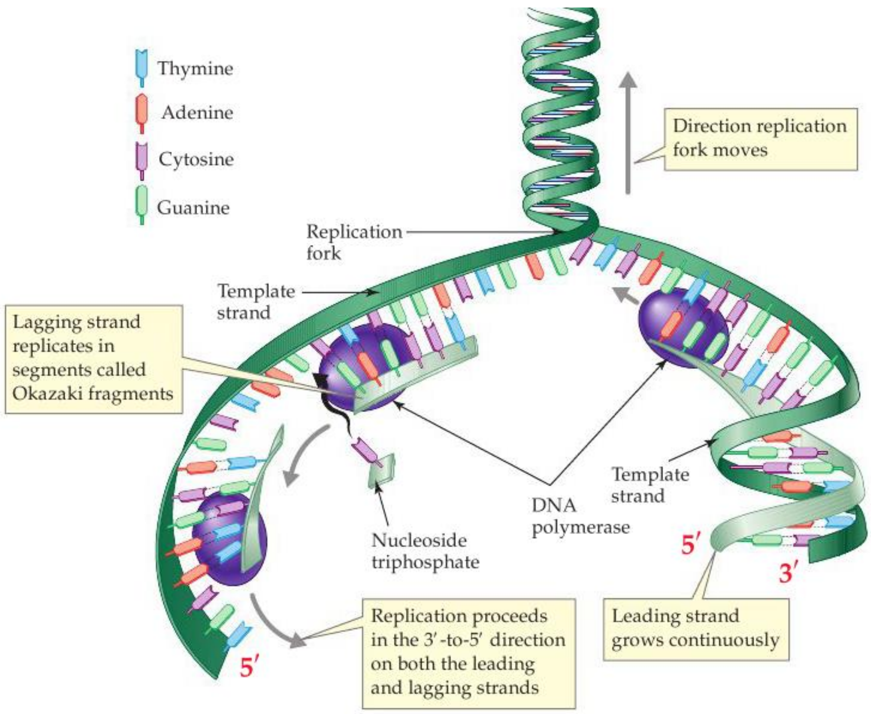
- *Transcription.* This process reads the stored genetic information and brings it out of the nucleus to ribosomes, where protein synthesis occurs.

- *Translation.* In this process, the genetic messages are decoded and used to build proteins.

Replication of DNA

Replication of DNA is an enzymatic process that starts with the partial unwinding of the double helix. Just before the cell division, the double strand begins to unwind. As the strands separate and bases are exposed, new nucleotides line up on each strand in a complementary fashion, A to T, and C to G. Two new strands now begin to grow, which are complementary to their old template strands. Two new identical DNA double helices are produced in this way, and these two new molecules can then be passed on, one to each daughter cell. As each of the new DNA molecules contains one strand of old DNA, and one new, the process is called *semiconservative replication*.

Addition of new nucleotide units to the growing chain occurs in the 5′ to C′ direction, and is catalysed by the enzyme DNA polymerase. The most important step is the addition of a 5′-mononucleoside triphosphate to the free 3′-hydroxyl group of the growing chain as the 3′-hydroxyl attacks the triphosphate and expels a diphosphate leaving group.



Transcription: synthesis of RNA

Transcription starts with the process by which the genetic information is transcribed onto a form of RNA, called mRNA. Ribonucleic acid, RNA, is structurally similar to DNA with the exceptions that its nucleotides contain ribose, instead of a 2′-deoxyribose, and the base thymine is replaced by uracil. There are three major types of RNA depending on their specific functions. However, all three types of RNA are much smaller than DNA and they are single stranded, rather than double stranded.

(a) *Messenger RNA (mRNA)* carries genetic information from DNA to ribosomes where protein synthesis occurs.

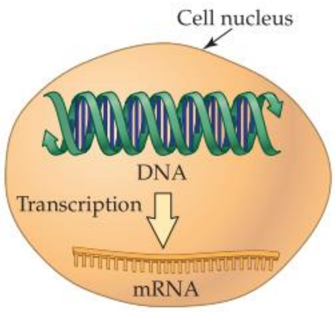
(b) *Ribosomal RNA (rRNA),* complexed with proteins (nucleoproteins), provides the physical make up of ribosomes.

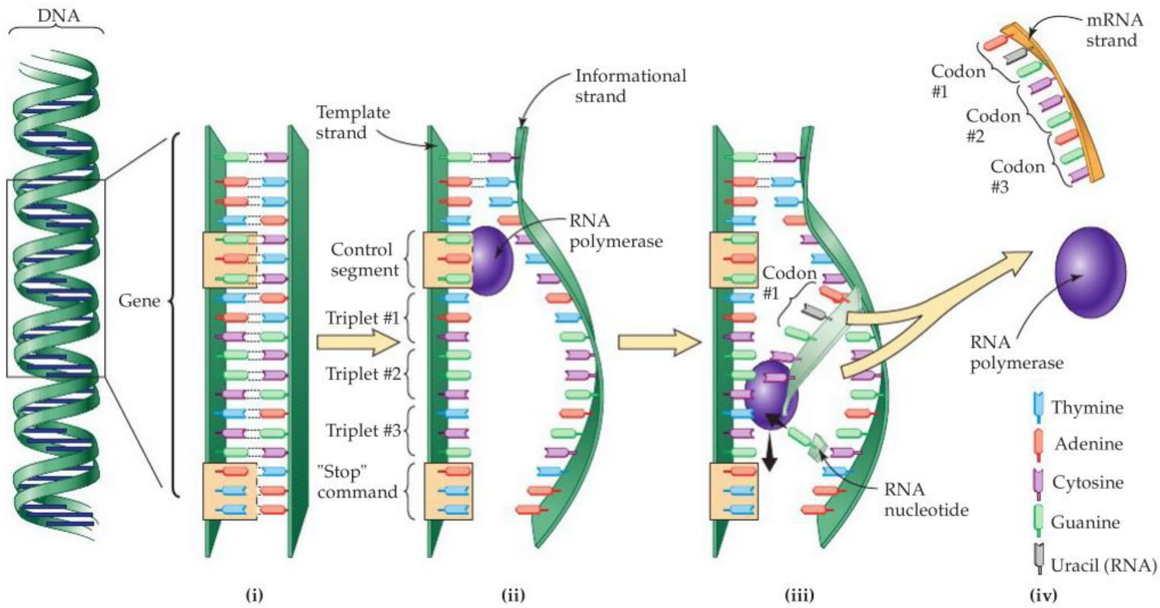
(c) *Transfer RNA (tRNA)* transports amino acids to the ribosomes for protein synthesis.

Protein synthesis takes place in the cell nucleus with the synthesis of mRNA. Part of the DNA double helix unwinds adequately to expose on a single chain a portion corresponding to at least one *gene*. Ribonucleotides, present in the cell nucleus, assemble along the exposed DNA chain by pairing with the bases of DNA in a similar fashion that is observed in DNA base pairing. However, in RNA uracil replaces thymine. The ribonucleotide units of mRNA are joined into a chain by the enzyme *RNA polymerase*. Once the mRNA is synthesized, it moves into the cytoplasm, where it acts as a template for protein synthesis. Unlike what is seen in DNA replication, where both strands are copied, only one of the two DNA strands is transcribed into mRNA. The strand that contains the gene is called the *coding strand* or *sense strand*. The strand that gets transcribed is known as the *template strand* or *antisense strand*. As the template strand and the coding strand are complementary, and as the template strand and the RNA molecule are also complementary, the RNA molecule produced during transcription is a copy of the coding strand, with the only exception that the RNA molecule contains a U everywhere the DNA coding strand has a T.

Ribosomes are small granular bodies scattered throughout the cytoplasm, and this is the place where protein synthesis starts. rRNA itself does not directly govern protein synthesis. A number of ribosomes get attached to a chain of mRNA and form a *polysome*, along which, with mRNA acting as the template, protein synthesis occurs. One of the major functions of rRNA is to bind the ribosomes to the mRNA chain.

tRNA is the smallest of all three types of RNA mentioned above, and consequently much more soluble than mRNA and rRNA. This is why tRNA is also sometimes called *soluble RNA*. tRNA transports amino acids, building blocks of protein synthesis, to specific areas of the mRNA of the polysome. tRNAs are composed of a small number of nucleotide units (70–90 units) folded into several loops or arms through base pairing along the chain.



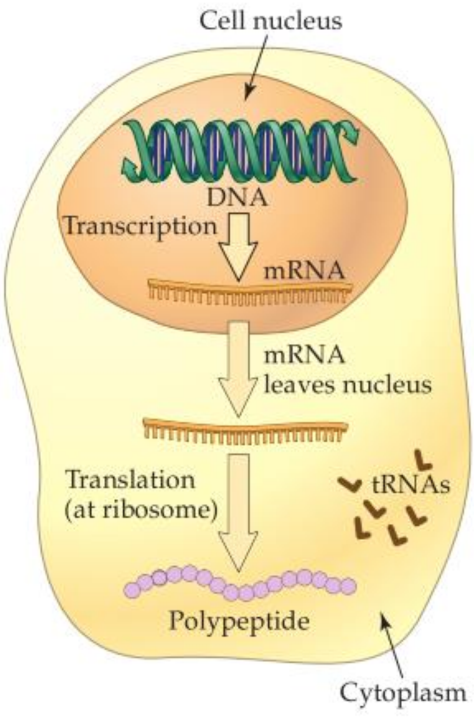


Transcription of DNA to produce mRNA. The transcription shown here produces a hypothetical three-codon mRNA. From left to right, (i) the DNA unwinds; (ii) the RNA polymerase connects with the control, or start, segment on the template strand; (iii) the mRNA is assembled as the polymerase moves along the template strand; and (iv) transcription ends when the polymerase reaches the stop command, releasing both the new mRNA strand and the RNA polymerase.

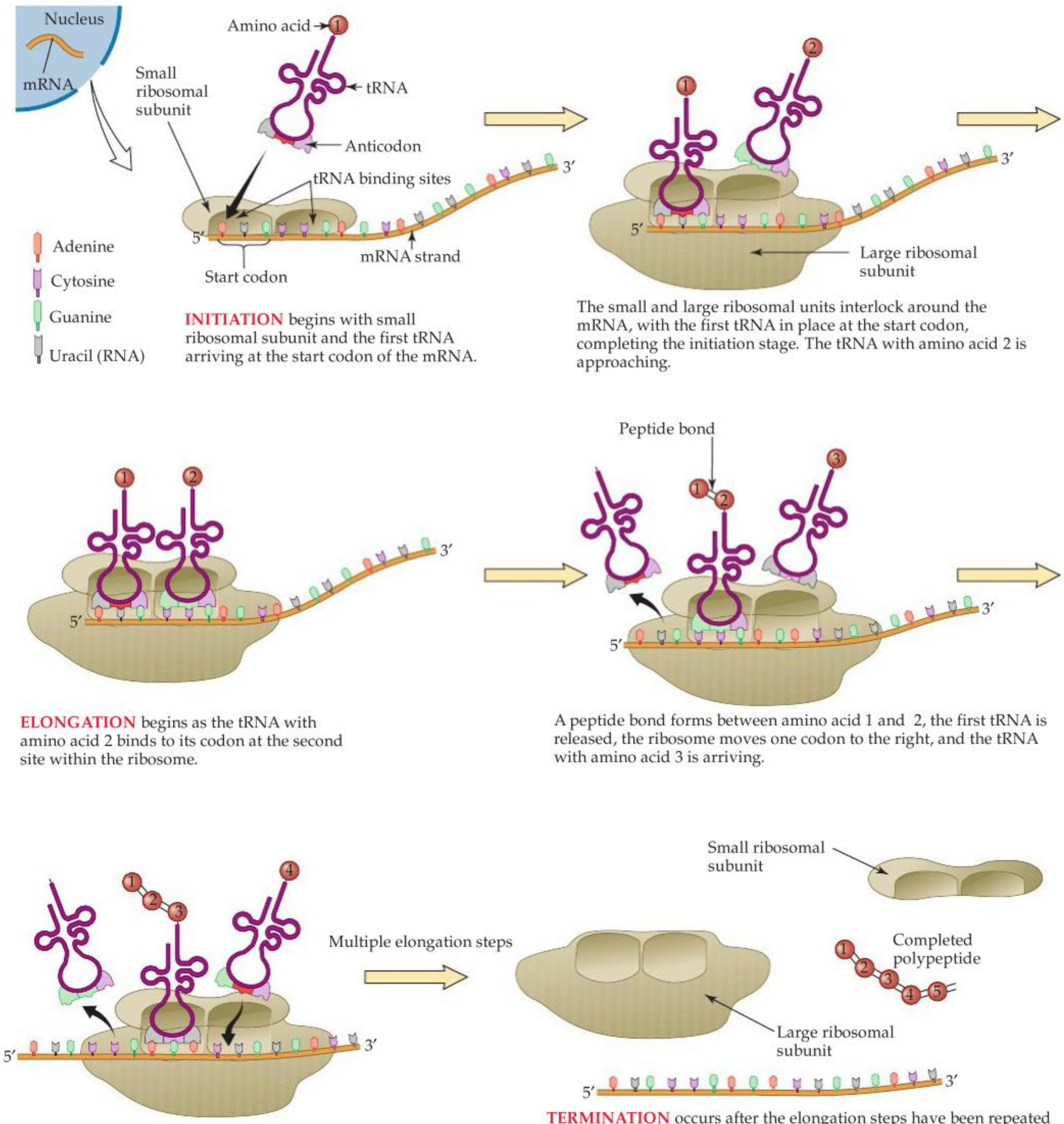
Translation: RNA and protein biosynthesis

Translation is the process by which mRNA directs protein synthesis. In this process, the message carried by mRNA is read by tRNA. Each mRNA is divided into codons, ribonucleotide triplets that are recognized by small amino-acid-carrying molecules of tRNA, which deliver the appropriate amino acids needed for protein synthesis.

RNA directs biosynthesis of various peptides and proteins essential for any living organisms. Protein biosynthesis seems to be catalysed by mRNA rather than protein-based enzymes and occur on the ribosome. On the ribosome, the mRNA acts as a template to pass on the genetic information that it has transcribed from the DNA. The specific ribonucleotide sequence in mRNA forms an '*instruction*' or *codon* that determines the order in which different amino acid residues are to be joined. Each 'instruction' or codon along the mRNA chain comprises a sequence of three ribonucleotides that is specific for a given amino acid. For example, the codon U–U–C on mRNA directs incorporation of the amino acid phenylalanine into the growing protein.



▲ **Overview of protein synthesis.** The codons of mature mRNA are translated in the ribosomes, where tRNAs deliver amino acids to be assembled into proteins (polypeptides).



INITIATION begins with small ribosomal subunit and the first tRNA arriving at the start codon of the mRNA.

The small and large ribosomal units interlock around the mRNA, with the first tRNA in place at the start codon, completing the initiation stage. The tRNA with amino acid 2 is approaching.

ELONGATION begins as the tRNA with amino acid 2 binds to its codon at the second site within the ribosome.

A peptide bond forms between amino acid 1 and 2, the first tRNA is released, the ribosome moves one codon to the right, and the tRNA with amino acid 3 is arriving.

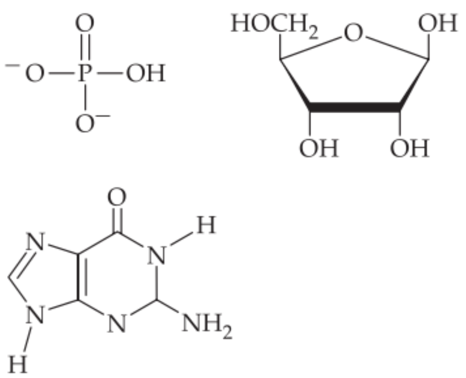
Elongation continues with three amino acids in the growing chain and the fourth one arriving with its tRNA.

TERMINATION occurs after the elongation steps have been repeated until the stop codon is reached. The ribosomal units, the mRNA, and the polypeptide separate.

PROBLEMS

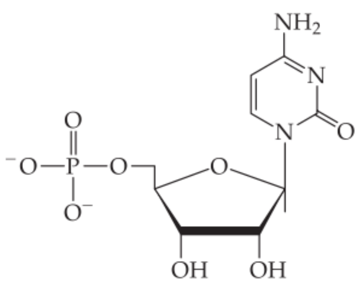
1.

Combine the structures below to create a ribonucleotide. Show where water is removed to form an *N*-glycosidic linkage and where water is removed to form a phosphate ester. Draw the resulting ribonucleotide structure, and name it.



2.

For the following molecule:



(a) Label the three nucleic acid building blocks it contains.

(b) Draw a box around the nucleoside portion of the molecule.

(c) Draw a circle around the nucleotide portion of the molecule.

3.

What amino acids are specified by the following codons?

(a) C-C-U
(b) G-C-A
(c) A-U-U

Literature:

1. Sarker, Satyajit D. Chemistry for pharmacy students: general, organic, and natural product chemistry / Satyajit D. Sarker, Lutfun Nahar. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England
2. Fundamentals of general, organic, and biological chemistry/John McMurry, Mary E. Castellion, David S. Ballantine. 6th ed. Prentice Hall: New York, Boston, San Francisco, London, Toronto, Sydney, Tokyo, Singapore, Madrid, Mexico City, Munich, Paris, Cape Town, Hong Kong, Montreal.- 2007. - 901 p.