UNIT (12) MOLECULES OF LIFE: NUCLEIC ACIDS

Nucleic acids are extremely large molecules that were first isolated from the nuclei of cells. Two kinds of nucleic acids are found in cells:

- RNA (ribonucleic acid) is found mainly in the cytoplasm of living cells.
- DNA (deoxyribonucleic acid) is found primarily in the nucleus of cells.

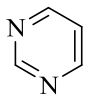
Both RNA and DNA are large polymers containing repeating structural units, or monomers, called *nucleotides*.

12.1 Components of Nucleic Acids

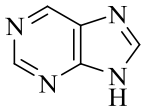
A nucleotide is composed of three units: an **organic base,** a **sugar**, and a **phosphate**.

A) Organic Bases

The organic bases found in nucleic acids are derivatives of **pyrimidine** or **purine**.
Pyrimidine is a six-membered **heterocyclic ring**. A heterocyclic ring is a ring compound containing atoms that are not all identical.
Purine is a fused-ring compound containing a six-membered ring connected to a five-membered ring.

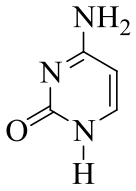


Pyrimidine

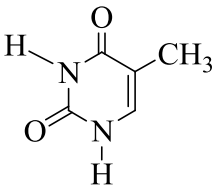


Purine

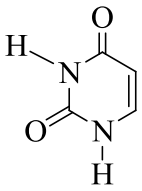
The three pyrimidine derivatives found in nucleic acids are cytosine (C), thymine (T), and uracil (U). They are commonly identified using the first letter in their name which is always capitalized.



Cytosine (C)
(DNA and RNA)

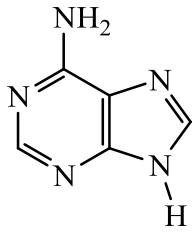


Thymine (T)
(DNA only)

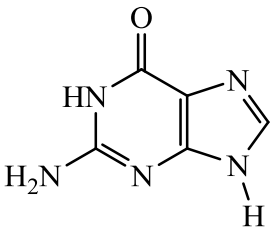


Uracil (U)
(RNA only)

The two purine derivatives found in nucleic acids are adenine (A) and guanine (G).



Adenine (A)
(DNA and RNA)

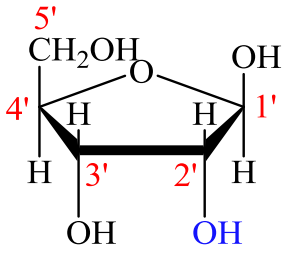


Guanine (G)
(DNA and RNA)

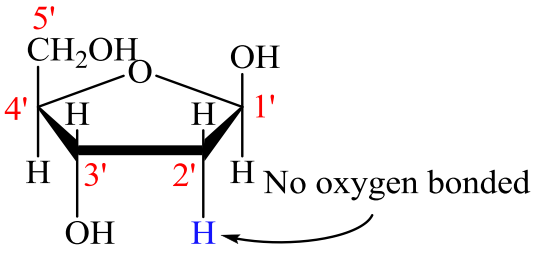
Adenine, guanine, and cytosine are found in both DNA and RNA. Thymine is found only in DNA, while uracil is found only in RNA. Thymine and uracil are often used to differentiate DNA from RNA.

B) Sugars

The five-carbon sugar in nucleic acids is ribose or a ribose derivative. In RNA the sugar is *ribose*, in DNA it is 2′-*deoxyribose*.
The *only* difference between these two sugars is found at the 2′-carbon of the ribose ring. Ribose has a hydroxyl group (-OH) bound to this carbon, while deoxyribose has a hydrogen atom ("deoxy" means no oxygen).



Ribose in RNA



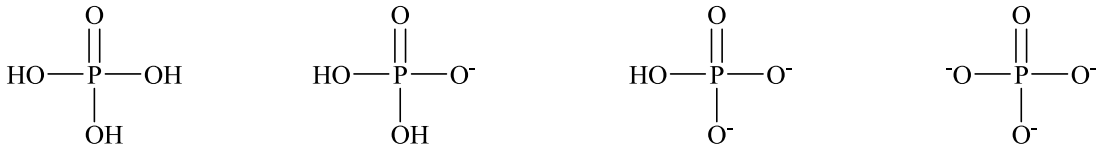
Deoxyribose in DNA

Notice that the carbon atoms in five-carbon sugars are numbered with primes (1′, 2′, 3′, 4′, and 5′). This is done to differentiate them from the atoms in the nitrogenous bases (purines and pyrimidines).

C) Phosphate Group

The third component of a nucleotide is derived from phosphoric acid (H_3PO_4). Phosphoric acid contains three hydrogen atoms and it can exist in one of the following four different forms depending on the pH of the solution.

present at low pH | present at physiological pH | present at high pH

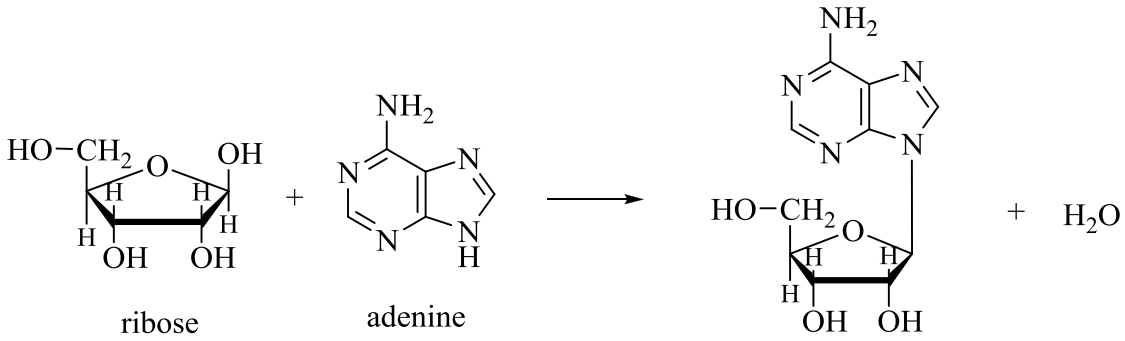


12.2 Nucleosides and Nucleotides

When ribose or 2′-deoxyribose is combined with a purine or pyrimidine base, a **nucleoside** is formed. A nucleoside is basically a nucleotide that is missing the phosphate portion.

Sugar + Base ⟶ Nucleoside

The formation of a nucleoside (in this case adenosine) could be shown as:



ribose + adenine ⟶ adenosine + H_2O

Note the name of adenine changes to adenosine when it is used to form a nucleoside. These subtle changes must be recognized because they identify different structures. Other nitrogenous bases (purines and pyrimidines) also have subtle name changes when used to form nucleosides. The table below lists the names.

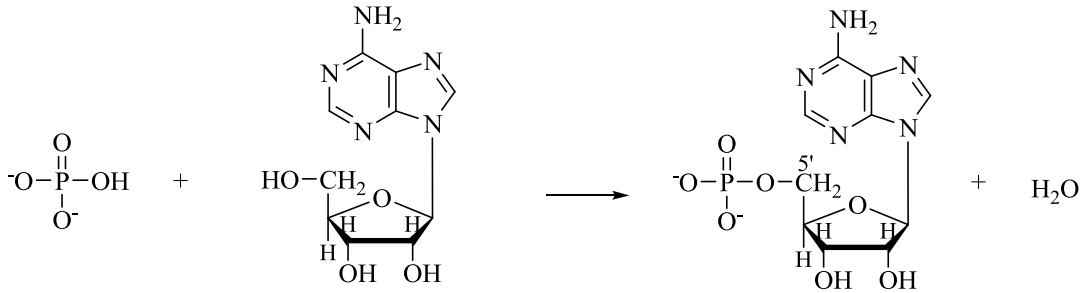
Names of Nucleosides

	Base	Nucleosides
RNA		
	Adenine	Adenosine
	Guanine	Guanosine
	Cytosine	Cytidine
	Uracil	Uridine
DNA		
	Adenine	Deoxyadenosine
	Guanine	Deoxyguanosine
	Cytosine	Deoxycytidine
	Thymine	Deoxythymidine

Phosphate ion reacts with the –OH groups on the sugar residue of a nucleoside to form a phosphate monoester and a **nucleotide** is produced. This commonly occurs at the OH attached at the 5′ carbon.

Phosphate(s) + Nucleoside ⟶ Nucleotide

The formation of a nucleotide from a nucleoside and a phosphate is shown below:



Names of Nucleotides

Base	Nucleosides	Nucleotides
RNA		
Adenine	Adenosine	Adenosine-5′-monophosphate (AMP)
Guanine	Guanosine	Guanosine-5′-monophosphate (GMP)
Cytosine	Cytidine	Cytidine-5′-monophosphate (CMP)
Uracil	Uridine	Uridine-5′-monophosphate (UMP)
DNA		
Adenine	Deoxyadenosine	Deoxyadenosine-5′-monophosphate (dAMP)
Guanine	Deoxyguanosine	Deoxyguanosine-5′-monophosphate (dGMP)
Cytosine	Deoxycytidine	Deoxycytidine-5′-monophosphate (dCMP)
Thymine	Deoxythymidine	Deoxythymidine-5′-monophosphate (dTMP)

12.3 Polynucleotides

A **polynucleotide** chain is formed by connecting several nucleotides in succession.
RNA is a polynucleotide that, upon hydrolysis, yields D-ribose, phosphoric acid, and the four bases adenine, guanine, cytosine, and uracil.
DNA is a polynucleotide that yields D-2′-deoxyribose, phosphoric acid, and the four bases adenine, guanine, cytosine, and thymine.
Nucleotides can be connected to one another to form **oligonucleotides** (2 to 10 nucleotide residues) and polynucleotides (more than 10 nucleotides).

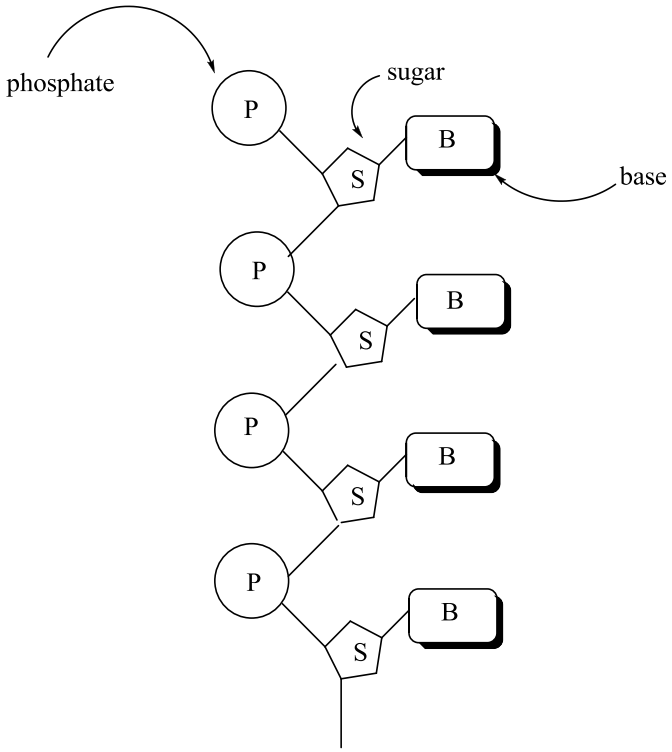
12.4 The Structure of DNA

Each cell in a particular living organism contains the exact same DNA. In plant and animal cells, most of the DNA is found in the cell nucleolus. The size of the DNA polymer is directly related to the complexity of the organism; more complex organisms tend to have larger molecules of DNA, while less complex organisms have smaller. The DNA in simple bacteria contains about 8 million nucleotides, whereas human DNA contains up to 500 million nucleotides.

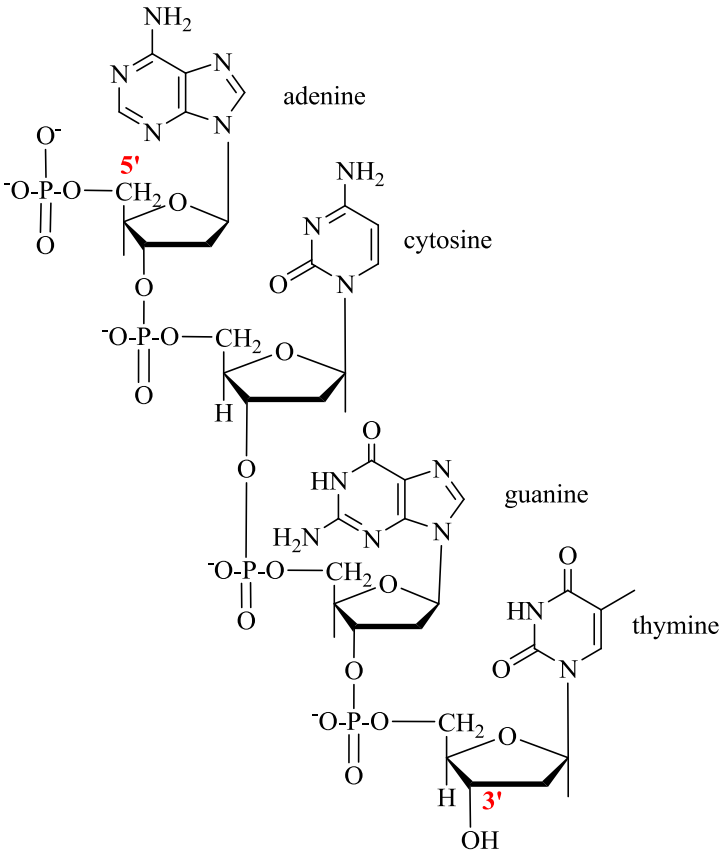
In *unit 11*, we learned that proteins have primary, secondary, and higher structures. Nucleic acids are also chains of monomeric units that have primary, secondary, and higher structures.

Primary Structure of DNA

The primary structure of DNA is simply the sequence of nucleotides. The sugar-phosphate chain is called the DNA *backbone*, and it is constant throughout the entire DNA molecule. The variable portion of DNA is the *sequence of nitrogenous bases*. A diagram of a nucleic acid is shown below



The phosphate groups link the 3′ carbon of one sugar (of deoxyribose or ribose) to the 5′ carbon of the next sugar (of deoxyribose or ribose).
The following illustrates the structure of ACGT (a tetranucleotide). It represents an example of the structural formula of a partial DNA molecule (note presence of thymine, therefore DNA). A strand of DNA has two distinct terminals or ends, one will be a 5′-phosphate end and the other will be a 3′-hydroxyl end. By convention, a nucleic acid sequence is always read in the 5′ to 3′ direction, that is, from the sugar with the free 5′-phosphate to the sugar with 3′-hydroxyl group. The order of nucleotides is generally written using the capitalized first letter of the name of base. As stated above, the following structure is written ACGT (in the 5′ to 3′ direction).



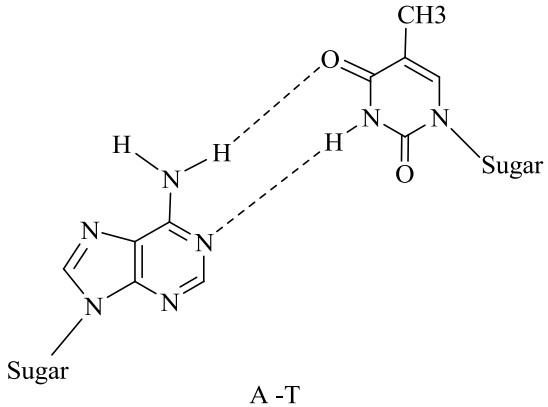
Secondary Structure of DNA: The DNA Double Helix

The secondary structure of DNA was proposed by James Watson and Francis Crick in 1953. This was perhaps the greatest discovery of modern biology and one of the most remarkable and profound events in the history of science.

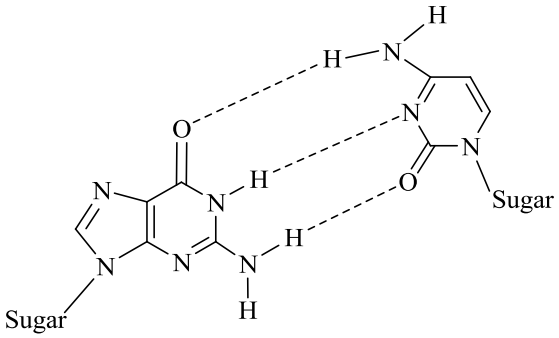
Watson and Crick concluded that DNA is a **double helix** containing two polynucleotide strands wound as if around a central axis. A good analogy would be to think of a rope ladder fixed at one end to the top of a pole, and subsequently wound downward around it without twisting the ladder. The two polynucleotide strands are connected by hydrogen bonds formed between a purine on one strand and a pyrimidine on the other. In DNA, adenine is always paired with thymine and guanine is always paired with cytosine. The pairs A-T and G-C are called **complementary base pairs**. Revisiting our rope ladder analogy, the two pieces of rope (two polynucleotide strands) are connected by the rungs of the ladder (hydrogen bonding between complementary base-pairs). According to base-pairing rules discovered by Watson and Crick, each A is bound to T and each G is bound to C. Therefore, the total number of A's in any molecule of DNA must be equal to total number of T's (the same is true of G and C). Thus, the % of A in DNA must equal the % of T (the same is true of G and C). The total percent of A, T, G, and C must, of course, equal 100.

Always %A = %T and %C = %G

Human DNA contains 30% adenine, 30% thymine, 20% cytosine, and 20% guanine.



A -T

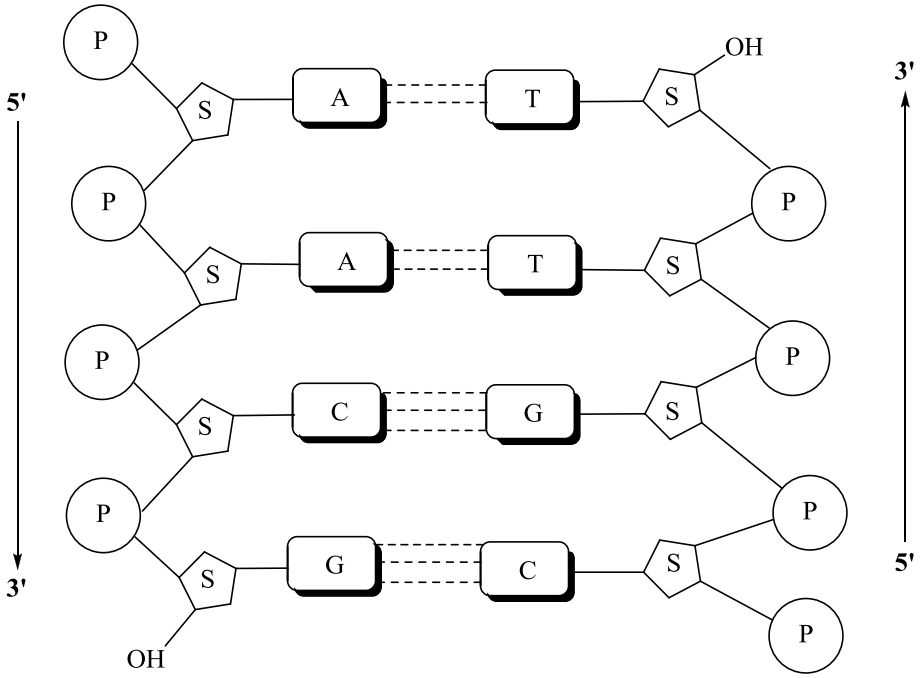


G - C

Base Pairing: Hydrogen bonding between the complementary base pairs: adenine/ thymine and cytosine/guanine.

Notice that A-T pairing has two hydrogen bonds (AT is a two letter word) and G-C pairing has three hydrogen bonds.

One important feature of the DNA double helix is that the two strands run *antiparallel* to one another, that is, the two strands run in opposite directions-one in the 5’ to 3’ direction, the other in the 3’ to 5’ direction. Therefore, both ends of the double helix contain the 5’ end of one strand (5’ phosphate) and the 3’ end of the other (3’ OH).



DNA is responsible for the storage and transmission of hereditary information.

A human cell normally contains 46 chromosomes.
Each chromosome contains one molecule of DNA bound to a group of proteins called histones.
A **gene** is a segment of DNA that carries a single, specific command, for example, "make a globin molecule".

Practice 12-1

Write the complementary strand of DNA to the following sequence.

5′ A-C-T-C-G-G-T-A-A 3′

Answer

12.5 DNA Replication

When a cell divides, each of the resulting "daughter cells" receives a copy of DNA that is nearly identical to the DNA of the parent cell. Replication is a biological process that duplicates the DNA molecule. In DNA replication, the double helix (parent strand) unzips forming two separate strands called templates. These templates provide the base sequences used to synthesize new DNA (daughter) strands.
Replication is a very complicated enzyme-catalyzed process. Enzymes are needed to unwind the DNA prior to replication and repackage the DNA after synthesis.

12.6 Ribonucleic Acid (RNA)

One of the main functions of DNA is to direct the synthesis of RNA molecules.
There are four major differences between RNA molecules and DNA molecules.

1) RNA contains ribose sugar units rather than deoxyribose.
2) RNA contains the base uracil instead of thymine.
3) RNA is single stranded, except in some viruses.
4) RNA molecules are much smaller than DNA molecules.

Types of RNA Molecules

There are three classes of RNA:
Messenger RNA (**mRNA**) carries genetic information from DNA to the ribosomes and serves as a template for protein synthesis.
Transfer RNA (**tRNA**) delivers individual amino acids to the site of protein synthesis.
Ribosomal RNA **(rRNA**) combines with a series of proteins to form ribosomes, the physical site of active protein synthesis.

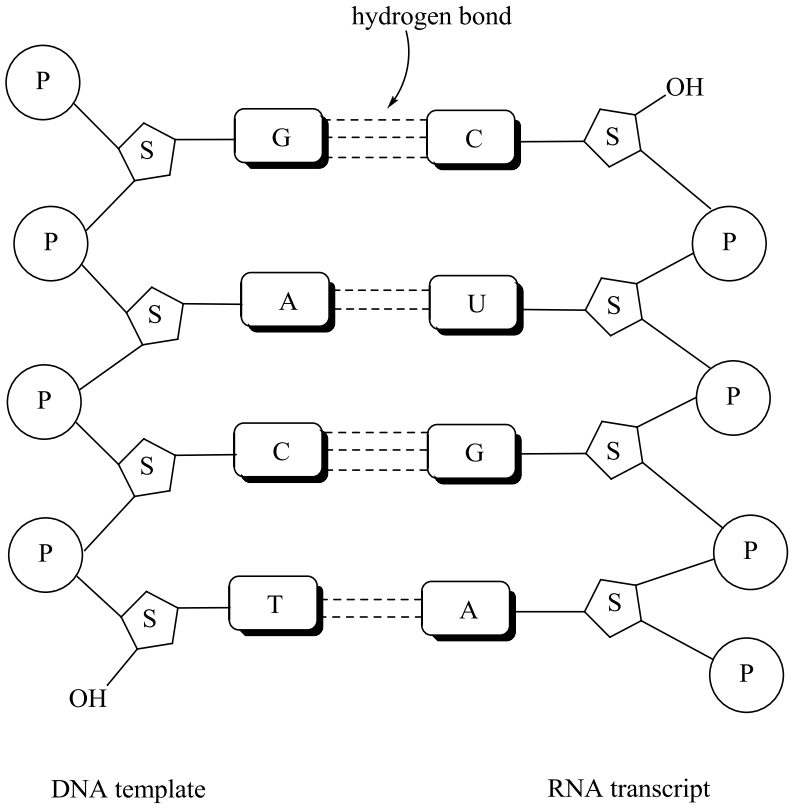
12.7 Gene Expression and Protein Synthesis

The central dogma (something held as an established opinion) of molecular biology states that the information contained in DNA molecules is transferred to RNA molecules which is subsequently expressed in the structure of proteins. More simply stated; DNA produces RNA which produces proteins.
Gene expression is the activation (turning on) of a gene to produce a specific protein. Two steps are involved in the flow of genetic information: *transcription* and *translation*.

Transcription: Synthesis of mRNA

Transcription is the process of mRNA synthesis from a single stranded DNA template. The enzyme that catalyzes transcription is called RNA polymerase.
Transcription begins when a portion of the DNA double helix unwinds near the gene to be expressed. Ribonucleotides assemble along the unwound DNA strand according to complementary base pairing. There is no change in G-C base pairing, G or C on DNA pairs with C and G on mRNA. There is a significant point of difference with A-T base pairing; T on DNA pairs with A on mRNA, but A on DNA pairs with U on mRNA. Recall that RNA contains no thymine (T), it has uracil (U) instead. Remember NEVER write T in mRNA (see worked example 12-1).

When RNA polymerase reaches the termination site, transcription ends and the newly formed mRNA is released. The unwound portion of the DNA returns to its double helix configuration.



Worked Example 12-1

Write the mRNA produced from the following DNA template.

3′ G-A-A-C-T 5′.

Solution

The bases on the DNA template are paired with their complementary bases to form mRNA. Remember C with G, G with C, T (on DNA) with A (on RNA), and A (on DNA) with U (on RNA). There are two ways to approach this problem. Some find it easier to simply memorize the base-pairings above and apply them to mRNA synthesis. Others chose to associate mRNA synthesis with the procedure used previously to write complementary strands of DNA and simply replace all T's with U's.

Applying memorized base-pairings:

DNA template	3′ G-A-A-C-T 5′.
Complementary bases in mRNA:	5′ C-U-U-G-A 3′.

Associate with DNA synthesis:

DNA template	3′ G-A-A-C-T 5′.
Complementary bases in DNA:	5′ C-T-T-G-A 3′.
Change all T's to U (no T in RNA):	5'C-U-U-G-A 3'.

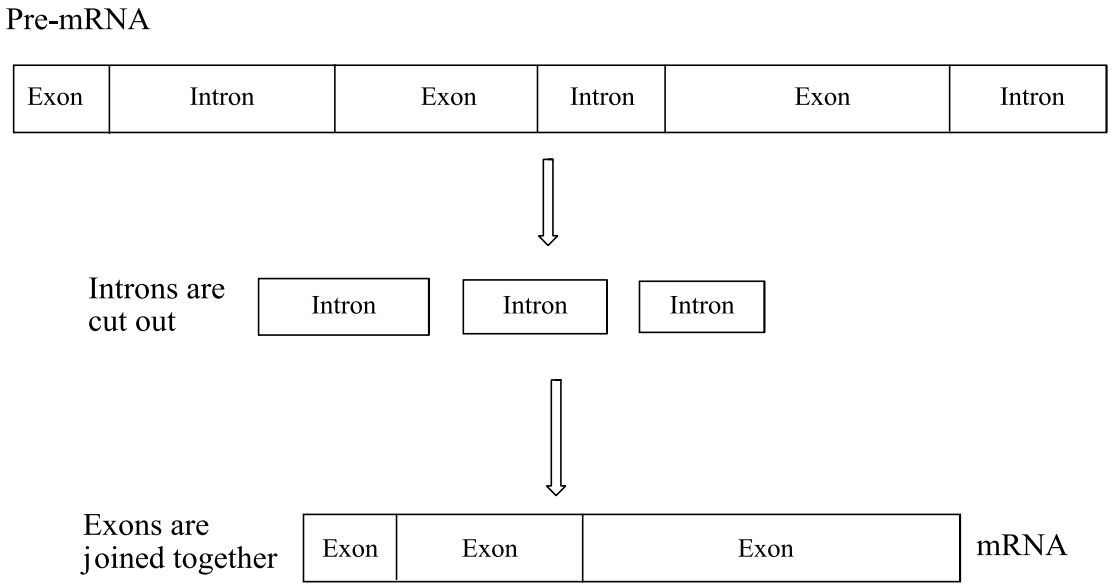
Practice 12-2

What is the DNA template that codes for the mRNA segment with the nucleotide sequence of sequence of 5′ G-C-U-A-G-U 3′?

Answer

Post-Transcription

The RNA produced from gene activation in transcription is a *pre*-mRNA.
The *pre*-mRNA contains two segments: one is coded for amino acids (exon) and the other carries no codes for amino acids (intron).
An **exon** is a gene segment that conveys (codes for) genetic information.
An **intron** is a gene segment that does not convey (code for) genetic information.
Splicing is the process of removing the introns from the *pre*-mRNA molecule and joining (splicing) the remaining exons together to form a mRNA molecule.



12.8 The Genetic Code

The information carried on the mRNA will be used to produce proteins. The mRNA sequence is read **three bases (triplet)** at a time and each segment of three bases is called a **codon**. Each codon specifies a particular amino acid in the primary structure of the protein (its sequence of amino acids).
There are 64 different codons used to specify amino acids and each could possibly appear on the mRNA molecule. A triplet arrangement of adenine (A), guanine (G), cytosine (C), or uracil (U) results in a total of 64 different combinations (64 different sets of 3 bases).
It has been found that 61 of the 64 codons identify specific amino acids; the other three combinations are termination codons ("stop" signals) for protein synthesis.
Codons have been determined for all 20 amino acids.
The **genetic code** is the assignment of the 64 mRNA codons to specific amino acids (or stop signals). One important characteristic of the genetic code is that it is almost universal. With minor exceptions, the triplet codons represent the same amino acids in every organism.
Another interesting feature of the genetic code is that it is highly degenerative. Many amino acids are designated by more than one codon. This allows for slight mutations in the code without changing the amino acid, i.e.; glycine is represented by four codons.

The 64 possible codons for mRNA are given in tables 1 and 2. It should be noted that the codons are always read in the 5' to 3' direction on the mRNA strand.

The concepts are consistent in the two tables. The first table is used if an amino acid is given and the triplet code is asked. The second table is used if the triplet code is given and the amino acid is asked. You will NOT be required to memorize the tables.

Table (1) For a given amino acid find the triplet codon.

Amino Acids	Codons	Number of codons
Alanine	GCA, GCC, GCG, GCU	4
Arginine	AGA, AGG, CGA, CGC, CGG, CGU	6
Asparagine	AAC, AAU	2
Aspartic acid	GAC, GAU	2
Cysteine	UGC, UGU	2
Glutamic acid	GAA, GAG	2
Glutamine	CAA, CAG	2
Glycine	GGA, GGC, GGG, GGU	4
Histidine	CAC, CAU	2
Isoleucine	AUA, AUC, AUU	3
Leucine	CUA, CUC, CUG, CUU, UUA, UUG	6
Lysine	AAA, AAG	2
Methionine, **initiation**	AUG	1
Phenylalanine	UUC, UUU	2
Proline	CCA, CCC, CCG, CCU	4
Serine	UCA, UCC, UCG, UCU, AGC, AGU	6
Threonine	ACA, ACC, ACG, ACU	4
Tryptophan	UGG	1
Tyrosine	UAC, UAU	2
Valine	GUA, GUC, GUG, GUU	4
Stop signals	UAG, UAA, UGA	3
Total number of codons		64

Table (2) Triplet codes to assigned amino acids

First Base	Second Base	Third Base			
		U	C	A	G
U	U	Phe	Phe	Leu	Leu
	C	Ser	Ser	Ser	Ser
	A	Tyr	Tyr	Stop	Stop
	G	Cys	Cys	Stop	Trp
C	U	Leu	Leu	Leu	Leu
	C	Pro	Pro	Pro	Pro
	A	His	His	Gln	Gln
	G	Arg	Arg	Arg	Arg
A	U	Ile	Ile	Ile	Met
	C	Thr	Thr	Thr	Thr
	A	Asn	Asn	Lys	Lys
	G	Ser	Ser	Arg	Arg
G	U	Val	Val	Val	Val
	C	Ala	Ala	Ala	Ala
	A	Asp	Asp	Glu	Glu
	G	Gly	Gly	Gly	Gly

Practice 12-3

Answer the following:

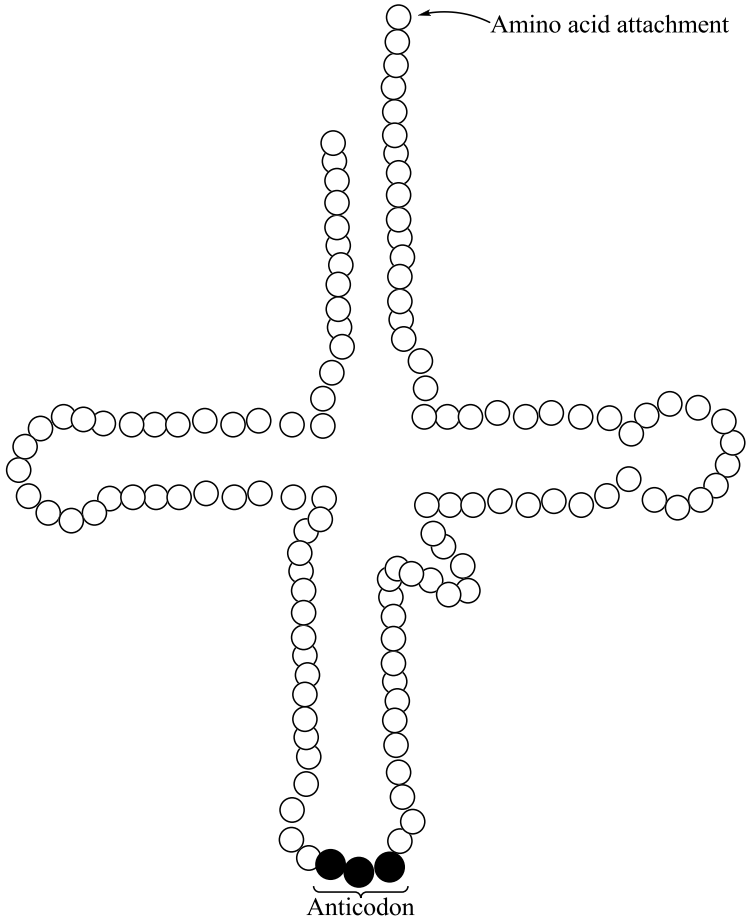
a) What codons specify tyrosine?

b) What amino acid is coded by CCG?

Answer

Translation: Protein Synthesis

The process of protein synthesis from mRNA is called **translation**. Proteins contain amino acids and mRNA contains nucleotides, we think of these as different "languages" so we *translate* mRNA into proteins. To direct the synthesis of a particular protein, the mRNA migrates out of the nucleus and into the cytoplasm where it binds to structures called *ribosomes*. The *transfer* RNAs (tRNAs) deliver individual amino acids to the mRNA as each codon is read. There are 61 different tRNAs, one for each of the 61 codons that specify an amino acid. A typical tRNA is roughly the shape of a cloverleaf as shown below.



Each tRNA molecule carries a three-base sequence called an **anticodon** that specifies which amino acid it will deliver.

Anticodon: A sequence of three nucleotides on tRNA, complementary to the codon on mRNA.

For example, the codon sequence UGG on a mRNA is read by a tRNA having the complementary anticodon sequence ACC and carrying a tryptophan.

Successive codons on the mRNA are read and the appropriate tRNA's bring the correct amino acid into position for enzyme-mediated transfer to the growing peptide. When synthesis of the proper protein is complete, a "stop" codon signals the end of translation and the protein is released from the ribosome.

Homework Problems

12.1 Draw the structures of the following nucleosides:
a. uridine
b. deoxythymidine

12.2 Draw the structure of the dinucleotide CG that would be in RNA.

12.3 Draw a structure showing the hydrogen bonding between uracil and adenine, and compare it with that of adenine and thymine.

12.4 Write the base sequence in a new DNA segment if the original segment has the following base sequence:
a. 5′ C – T–G – T – A – T – A – C – G – T – T – A 3′
b. 5′ A – G–T – C – C – A – G – G – T 3′

12.5 What is the difference between a *codon* and an *anticodon*?

12.6 A segment of a DNA strand consists of GCTTAGACCTGA.
a. What is the nucleotide order in the complementary mRNA?
b. What is the anticodon order in the tRNA?
c. What is the sequence of amino acids coded by the DNA?

12.7 Consider the following portion of mRNA produced by a normal order of DNA nucleotides: 5′-ACC – AGU- AGG – GUU – 3′
a. What is the amino acid order produced for normal DNA?
b. What is the amino acid order if a mutation changes AGU to ACA?
c. What is the amino acid order if a mutation changes AGG to GGG?
d. What happens to protein synthesis if a mutation changes AGU to UGA?