CHAPTER THREE

DNA, RNA and Protein

Nucleic Acid Molecules Carry Genetic Information

Chemical Structure of Nucleic Acids

DNA and RNA Each Have Four Bases

Nucleosides Are Bases Plus Sugars; Nucleotides Are Nucleosides Plus Phosphate

Double Stranded DNA Forms a Double Helix

Base Pairs are Held Together by Hydrogen Bonds

Complementary Strands Reveal the Secret of Heredity

Constituents of Chromosomes

The Central Dogma Outlines the Flow of Genetic Information

Ribosomes Read the Genetic Code

The Genetic Code Dictates the Amino Acid Sequence of Proteins

Various Classes of RNA Have Different Functions

Proteins, Made of Amino Acids, Carry Out Many Cell Functions

The Structure of Proteins Has Four Levels of Organization

Proteins Vary in Their Biological Roles

Nucleic Acid Molecules Carry Genetic Information

Chapter 1 discussed how the fundamentals of modern genetics were laid when Mendel found that hereditary information consists of discrete fundamental units now called genes. Each gene is responsible for a single inherited property or characteristic of the organism. Just as the discovery that atoms are made of subatomic particles ushered in the nuclear age, so the realization that genes are made up of **DNA** molecules opened the way both to a deeper understanding of life and to its artificial alteration by genetic engineering.

Genetic information is encoded by molecules named **nucleic acids** because they were originally isolated from the nucleus of eukaryotic cells. There are two related types of nucleic acid, **deoxyribonucleic acid** (**DNA**) and **ribonucleic acid** (**RNA**). The master copy of each cell's genome is stored on long molecules of DNA, which may each contain many thousands of genes. Each gene is thus a linear segment of a long DNA molecule. In contrast, RNA molecules are much shorter, are used to transmit the genetic information to the cell machinery, and carry only one or a few genes. [Certain viruses use RNA to encode their genomes as well as transmitting genetic information to the cell machinery. These RNA viruses have short genomes, rarely more than a dozen genes, as opposed to the hundreds or thousands of genes carried on the DNA genomes of cells.]

Genetic information is carried on long linear polymers, the nucleic acids. Two classes of nucleic acid, DNA and RNA, divide up the responsibility of storing and deploying the genetic information.

Chemical Structure of Nucleic Acids

DNA and RNA are linear polymers made of subunits known as **nucleotides**. The information in each gene is determined by the order of the different nucleotides, just as the information in this sentence is due to the order of the 26 possible letters of the alphabet. There are four different nucleotides in each type of nucleic acid and their order determines the genetic information (Fig. 3.01).

Each nucleotide has three components: a **phosphate group**, a five-carbon sugar, and a nitrogen-containing **base** (Fig. 3.02). The phosphate groups and the sugars form the backbone of each strand of DNA or RNA. The bases are joined to the sugars and stick out sideways.

In DNA, the sugar is always **deoxyribose**; whereas, in RNA, it is **ribose**. Both sugars are **pentoses**, or five-carbon sugars. Deoxyribose has one less oxygen than ribose (Fig. 3.03). It is this chemical difference that gave rise to the names deoxyribonucleic acid and ribonucleic acid. Both sugars have five-membered rings consisting of four carbon atoms and an oxygen. The fifth carbon forms a side chain to the ring. The five carbon atoms of the sugar are numbered 1′, 2′, 3′, 4′ and 5′ as shown in Fig. 3.02. By convention, in nucleic acids, numbers with prime marks refer to the sugars and numbers without prime marks refer to the positions around the rings of the bases.

Nucleotides are joined by linking the phosphate on the 5′-carbon of the (deoxy) ribose of one nucleotide to the 3′-position of the next as shown in Fig. 3.04. The phosphate group is joined to the sugar on either side by ester linkages, and the overall structure is therefore a **phosphodiester** linkage. The phosphate group linking the sugars has a negative charge.

base Alkaline chemical substance, in molecular biology especially refers to the cyclic nitrogen compounds found in DNA and RNA
deoxyribonucleic acid (DNA) Nucleic acid polymer of which the genes are made
deoxyribose The sugar with five carbon atoms that is found in DNA
DNA Deoxyribonucleic acid, nucleic acid polymer of which the genes are made
nucleic acid Class of polymer molecule consisting of nucleotides that carries genetic information
nucleotide Monomer or subunit of a nucleic acid, consisting of a pentose sugar plus a base plus a phosphate group
pentose A five carbon sugar, such as ribose or deoxyribose
phosphate group Group of four oxygen atoms surrounding a central phosphorus atom found in the backbone of DNA and RNA
phosphodiester The linkage between nucleotides in a nucleic acid that consists of a central phosphate group esterified to sugar hydroxyl groups on either side
ribonucleic acid (RNA) Nucleic acid that differs from DNA in having ribose in place of deoxyribose
ribose The 5-carbon sugar found in RNA

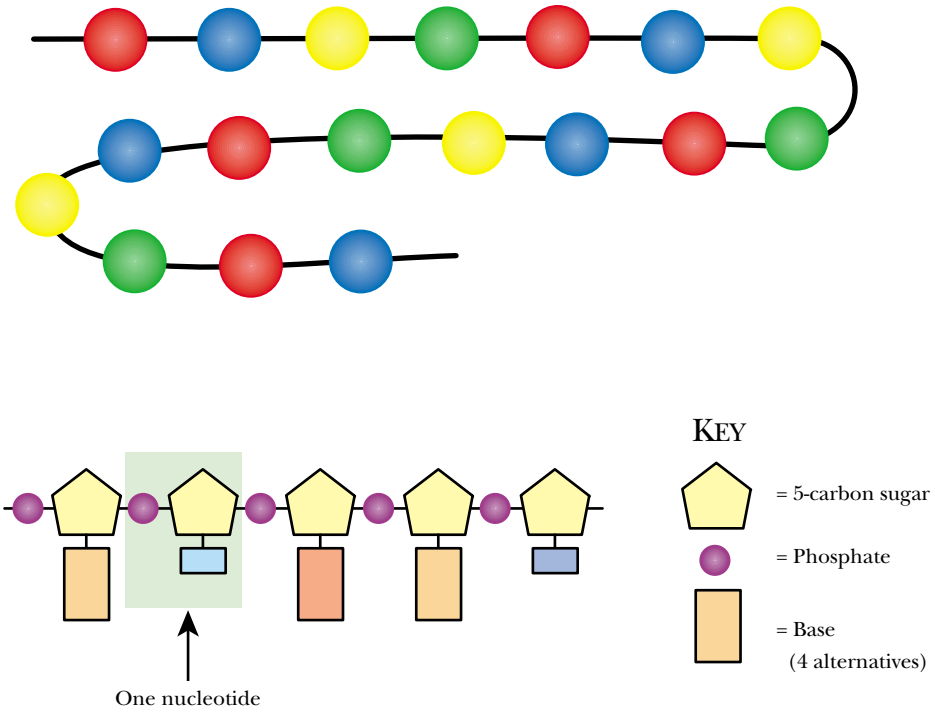


FIGURE 3.01 ***The Order of the Nucleotides Encodes the Genetic Information***

Nucleotides are ordered along a string of DNA or RNA. It is the ordering of the different nucleotides that dictates the nature of the information within the nucleic acid.

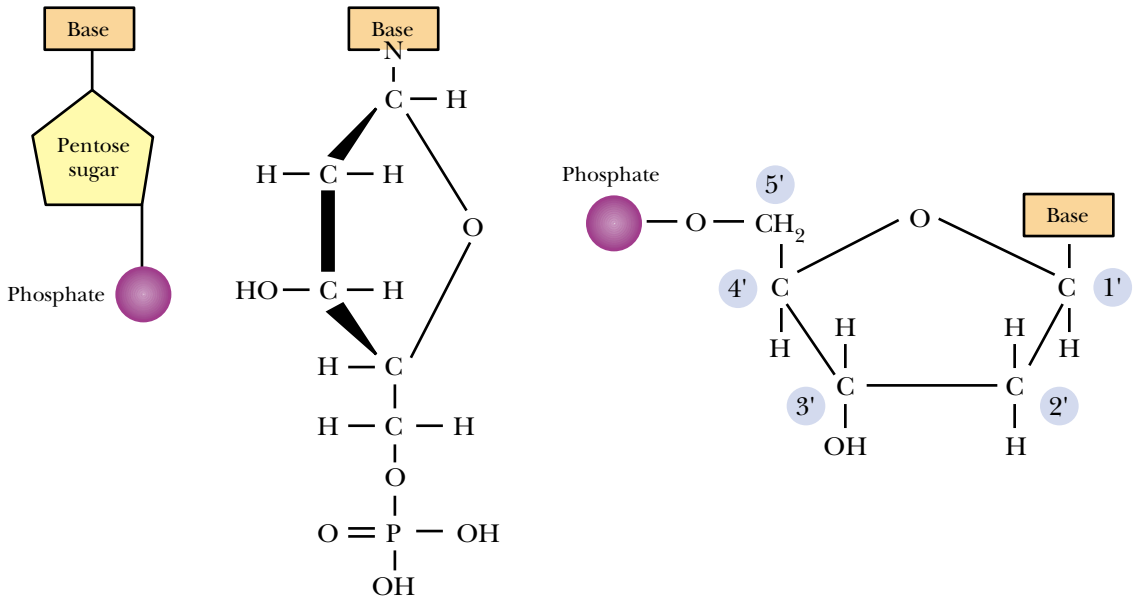
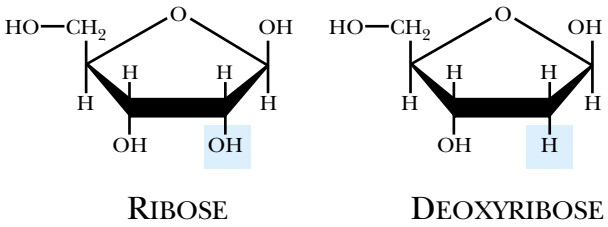


FIGURE 3.02 ***Three Views of a Nucleotide***

The three components of a nucleotide are shown to the left. The structures on the right show the pentose sugar (deoxyribose) connected to the phosphate and the base.

FIGURE 3.03 ***The Sugars Composing RNA and DNA***

Ribose is the five-carbon sugar (pentose) found in RNA. Deoxyribose is the pentose of DNA. It has one less oxygen than ribose as it has a hydrogen in place of the hydroxyl group on position 2′ of the ribose ring.



RIBOSE DEOXYRIBOSE

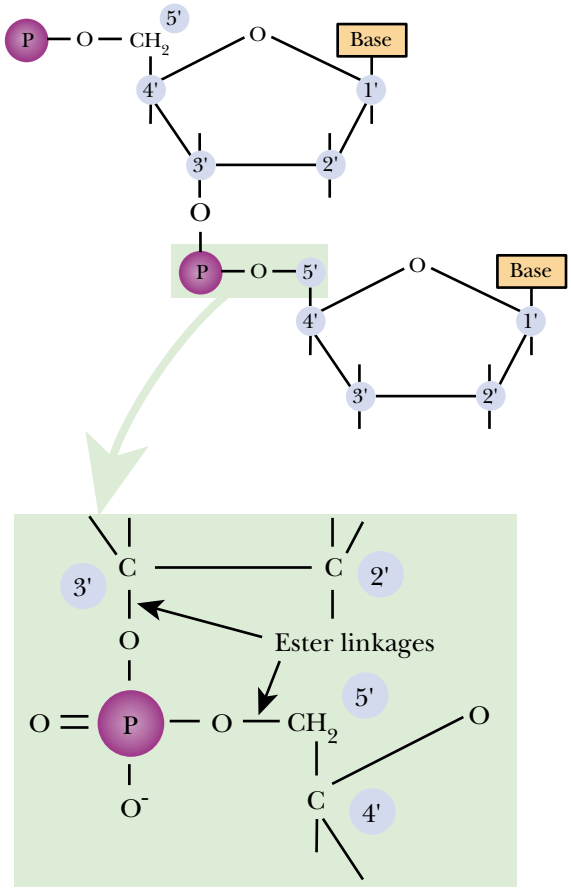


FIGURE 3.04 ***Nucleotides Are Joined by Phosphodiester Linkages***

The nucleotides that form the backbone of DNA and RNA are joined together by linkages involving their phosphate groups. One nucleotide is linked via its 5′-carbon to the oxygen of the phosphate group and another nucleotide is linked via its 3′-carbon to the other side of the central phosphate. These linkages are termed phosphodiester groups.

DNA and RNA Each Have Four Bases

There are five different types of nitrogenous bases associated with nucleotides. DNA contains the bases **adenine**, **guanine**, **cytosine** and **thymine**. These are often abbreviated to A, G, C and T, respectively. RNA contains A, G and C, but T is replaced by **uracil** (**U**). From the viewpoint of genetic information, T in DNA and U in RNA are equivalent.

The bases found in nucleic acids are of two types, **pyrimidines** and **purines**. The smaller pyrimidine bases contain a single ring whereas the purines have a fused double ring. Adenine and guanine are purines; and thymine, uracil and cytosine are pyrimidines. The purine and pyrimidine ring systems and their derivatives are shown in Figure 3.05.

adenine (A) A purine base that pairs with thymine, found in DNA or RNA
cytosine (C) One of the pyrimidine bases found in DNA or RNA and which pairs with guanine
guanine (G) A purine base found in DNA or RNA that pairs with cytosine
purine Type of nitrogenous base with a double ring found in DNA and RNA
pyrimidine Type of nitrogenous base with a single ring found in DNA and RNA
thymine (T) A pyrimidine base found in DNA that pairs with adenine
uracil (U) A pyrimidine base found in RNA that may pair with adenine

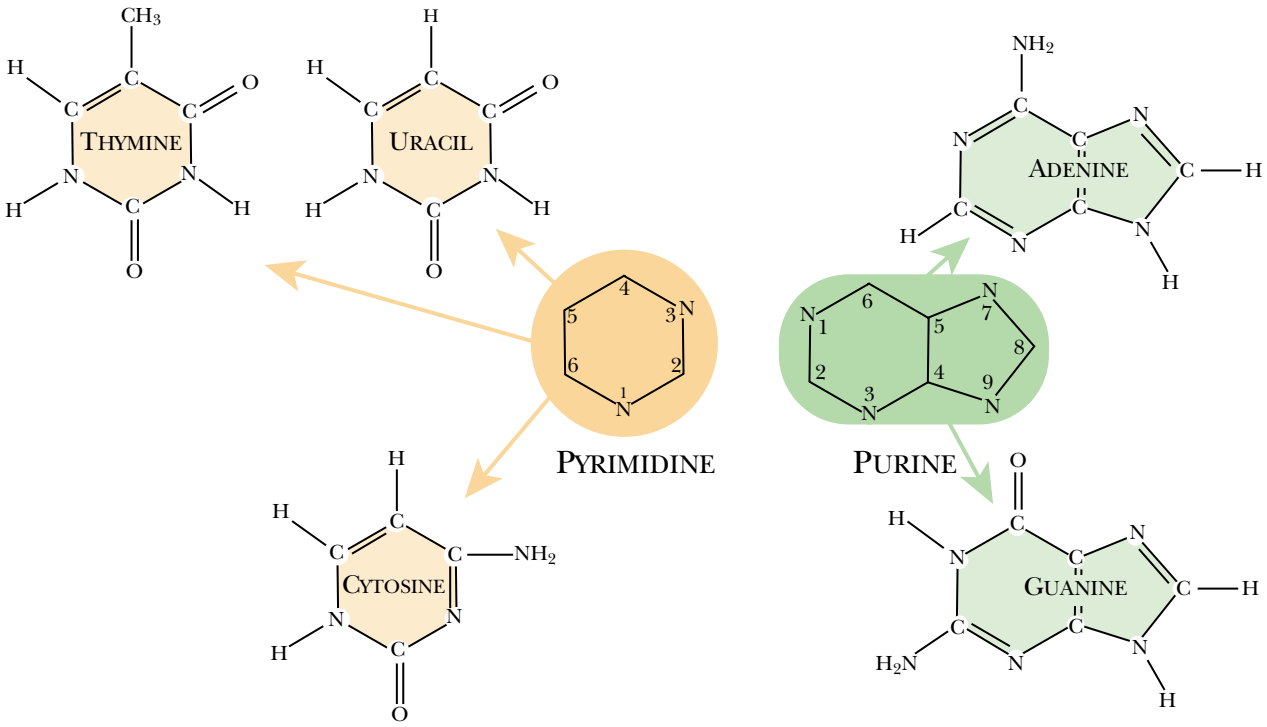


FIGURE 3.05 ***The Bases of the Nucleic Acids***

The four bases of DNA are adenine, guanine, cytosine and thymine. In RNA, uracil replaces thymine. Pyrimidine bases contain one-ring structures, whereas purine bases contain two-ring structures.

TABLE 3.01 Naming Bases, Nucleosides and Nucleotides

Base	Abbreviations		Nucleoside	Nucleotide
Adenine	ade	A	adenosine	adenosine monophosphate (AMP)
Guanine	gua	G	guanosine	guanosine monophosphate (GMP)
Cytosine	cyt	C	cytidine	cytidine monophosphate (CMP)
Thymine	thy	T	thymidine	thymidine monophosphate (TMP)
Uracil	ura	U	uridine	uridine monophosphate (UMP)

Nucleosides Are Bases Plus Sugars; Nucleotides Are Nucleosides Plus Phosphate

A base plus a sugar is known as a **nucleoside**. A base plus a sugar plus phosphate is known as a **nucleotide**. If necessary, one may distinguish between **deoxynucleosides** or **deoxynucleotides** where the sugar is deoxyribose, and **ribonucleosides** or **ribonucleotides** that contain ribose. The names of the nucleosides are similar to the names of the corresponding bases (see Table 3.01). The nucleotides do not have names of their own but are referred to as phosphate derivatives of the corresponding nucleoside. For example, the nucleotide of adenine is **adenosine** monophosphate, or AMP.

Three-letter abbreviations for the bases such as ade, gua, etc., are sometimes used when writing biochemical pathways or for the names of genes involved in nucleotide metabolism. When writing the sequence of a nucleic acid, the single letter abbrevia-

adenosine The nucleoside consisting of adenine plus (deoxy)ribose
deoxynucleoside A nucleoside containing deoxyribose as the sugar
deoxynucleotide A nucleotide containing deoxyribose as the sugar
nucleoside The union of a purine or pyrimidine base with a pentose sugar
nucleotide Monomer or subunit of a nucleic acid, consisting of a pentose sugar plus a base plus a phosphate group
ribonucleoside A nucleoside whose sugar is ribose (not deoxyribose)
ribonucleotide A nucleotide whose sugar is ribose (not deoxyribose)

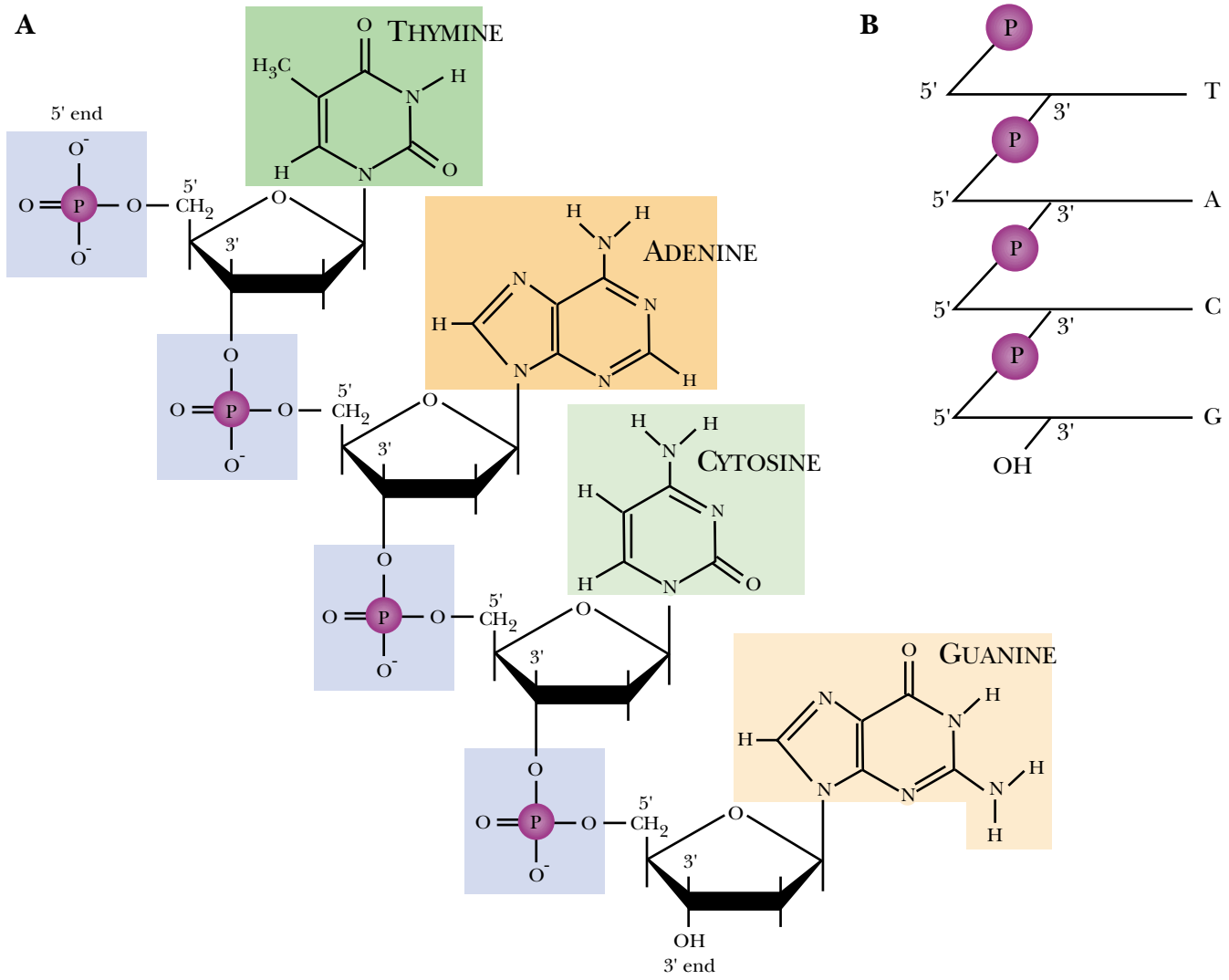


FIGURE 3.06 ***Some Variations in the Ways Nucleic Acids are Represented***

(A) More elaborate drawings show the chemical structures of the nucleic acid components, including the pentose sugar, phosphate groups and bases. (B) Simple line drawings may be used to summarize the linkage of sugars by the 5′ and 3′ phosphodiester bonds. Here, the protruding bases have been abbreviated to a single letter.

tions are used (A, T, G and C for DNA or A, U, G and C for RNA). The letter N is often used to refer to an unspecified base.

Double Stranded DNA Forms a Double Helix

A strand of nucleic acid may be represented in various ways, either in full or abbreviated to illustrate the linkages (Fig. 3.06). As remarked above, nucleotides are linked by joining the 5′- phosphate of one to the 3′-hydroxyl group of the next. Typically, there is a free phosphate group at the 5′-end of the chain and a free hydroxyl group at the 3′-end of a nucleic acid strand. Consequently, a strand of nucleic acid has polarity and it matters in which direction the bases are read off. The 5′-end is regarded as the beginning of a DNA or RNA strand. This is because genetic information is read starting at the 5′-end. [In addition, when genes are replicated, nucleic acids are synthesized starting at the 5′-end as described in Ch. 5.]

The structure of the DNA double helix is critical to replication of the genes, as described in more detail in Chapter 5.

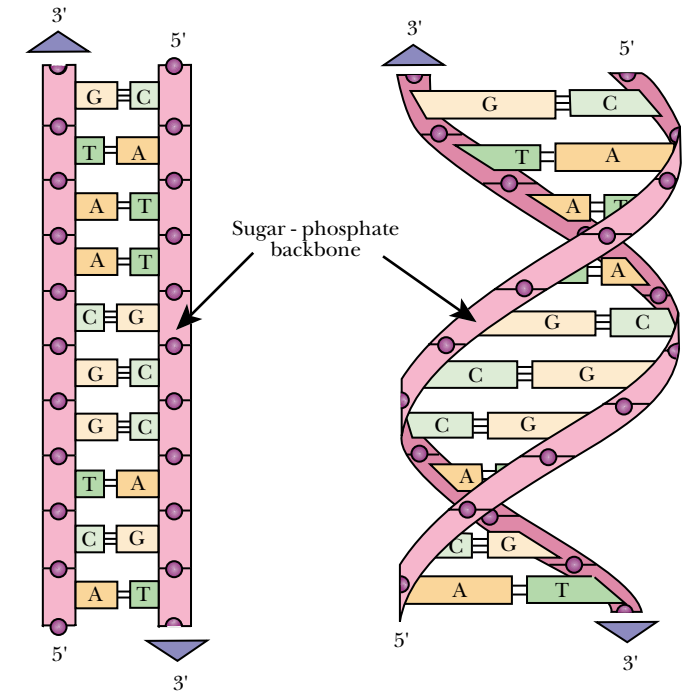


FIGURE 3.07
Representations of Double Stranded DNA

On the left DNA is represented as a double line consisting of two complementary strands. Actually DNA forms a double helix, as shown to the right.

RNA is normally found as a single-stranded molecule, whereas DNA is double-stranded. Note that the two strands of a DNA molecule are **antiparallel**, as they point in opposite directions. This means that the 5′-end of one strand is opposite the 3′-end of the other strand (Fig. 3.07). Not only is DNA double-stranded, but the two separate strands are wound around each other in a helical arrangement. This is the famous **double helix** first proposed by Francis Crick and James Watson in 1953 (Fig. 3.08). The DNA double helix is stabilized both by hydrogen bonds between the bases (see below) and by stacking of the aromatic rings of the bases in the center of the helix.

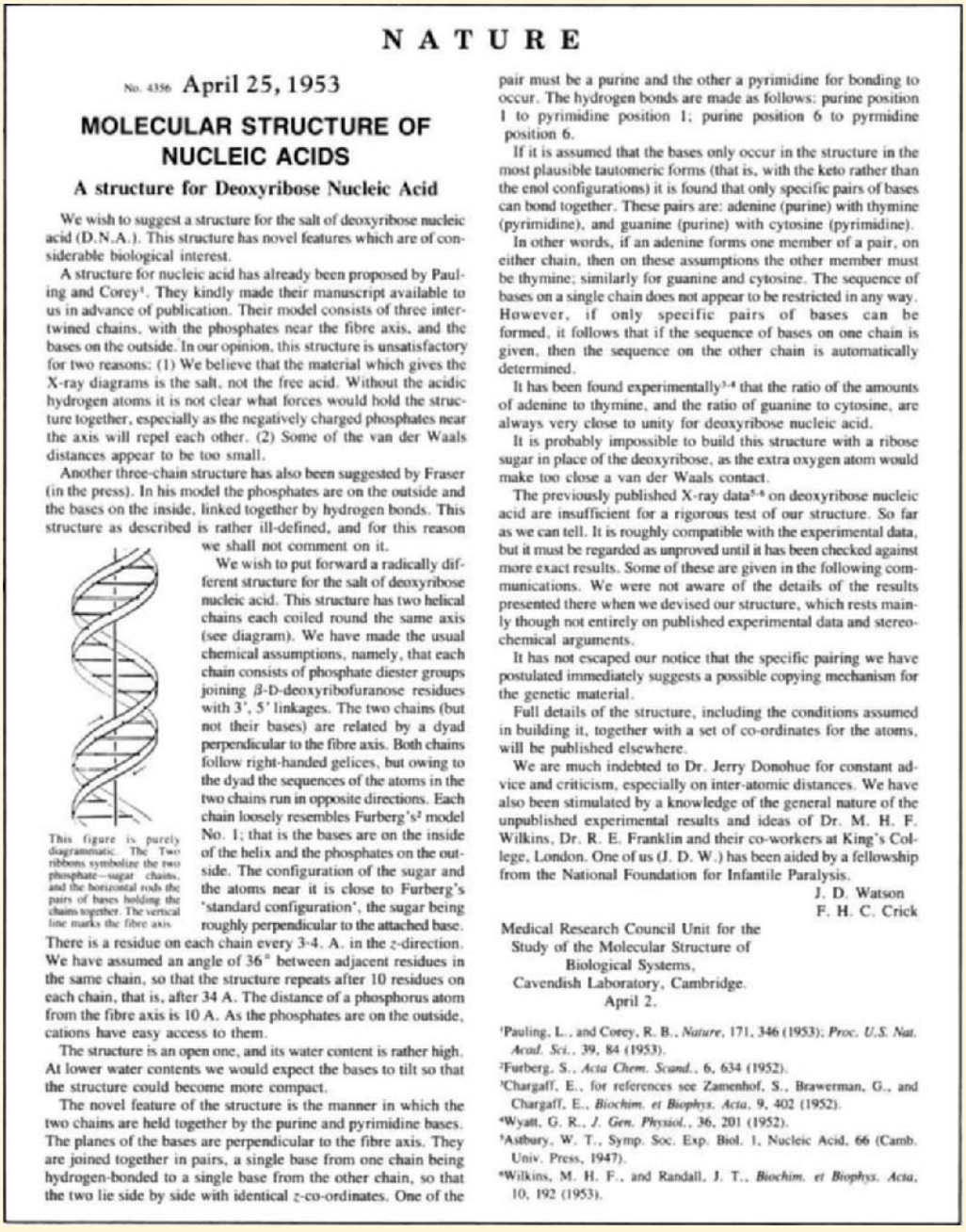
DNA forms a **right-handed** double helix. To tell a right-handed helix from a left-handed helix, the observer must look down the helix axis (in either direction). In a right-handed helix, each strand turns clockwise as it moves away from the observer (in a left-handed helix it would turn counterclockwise).

Base Pairs are Held Together by Hydrogen Bonds

In double stranded DNA, the bases on each strand protrude into the center of the double helix where they are paired with the bases in the other strand by means of **hydrogen bonds**. Adenine (A) in one strand is always paired with thymine (T) in the other, and guanine (G) is always paired with cytosine (C) (Fig. 3.10). Consequently, the number of adenines in DNA is equal to the number of thymines, and similarly the numbers of guanine and cytosine are equal. Note that the nucleic acid bases have amino or oxygen side-groups attached to the ring. It is these chemical groups, along with the nitrogen atoms that are part of the rings themselves, that allow the formation of hydrogen bonds. The hydrogen bonding in DNA **base pairs** involves either oxygen

antiparallel Parallel, but running in opposite directions
base pair Two bases held together by hydrogen bonds
double helix Structure formed by twisting two strands of DNA spirally around each other
hydrogen bond Bond resulting from the attraction of a positive hydrogen atom to both of two other atoms with negative charges
right-handed helix In a right-handed helix, as the observer looks down the helix axis (in either direction), each strand turns clockwise as it moves away from the observer

Working in Cambridge, England, James Watson and Francis Crick based their model of the double helix partly on the interpretation of data from X-ray crystallography by Rosalind Franklin and Maurice Wilkins, which suggested a helical molecule. Chemical analysis by Erwin Chargaff showed that DNA contained equimolar amounts of A and T and also of G and C. This, and chemical modeling, led Watson and Crick to propose that DNA was double stranded and that A in one strand is always paired with T in the other. Similarly, G is always paired with C. Watson and Crick published their landmark paper in *Nature* (Fig. 3.08) in 1953.



NATURE

No. 4356 April 25, 1953

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3', 5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed gelices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrmidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3-6] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5-6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on inter-atomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. Watson
F. H. C. Crick

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

[1]Pauling, L., and Corey, R. B., *Nature*, 171, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, 39, 84 (1953).
[2]Furberg, S., *Acta Chem. Scand.*, 6, 634 (1952).
[3]Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, 9, 402 (1952).
[4]Wyatt, G. R., *J. Gen. Physiol.*, 36, 201 (1952).
[5]Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6]Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, 10, 192 (1953).

FIGURE 3.08 ***DNA Is a Double Helix***

This one-page paper published in *Nature* described the now-famous double helix. *J. D. Watson & F. H. C. Crick*, Molecular Structure of Nucleic Acids, A Structure for Deoxyribose Nucleic Acid, Nature 171 (1953) 737.

In 2003 the Double Helix celebrated its 50th anniversary. In Great Britain, the Royal Mail issued a set of five commemorative stamps illustrating the double helix together with some of the technological advances that followed, such as comparative genomics and genetic engineering. In addition, the Royal Mint issued a £2 coin depicting the DNA double helix itself (Fig. 3.09).

FIGURE 3.09 ***Double Helix—50th Anniversary Coin***

A £2 coin commemorating the discovery of the double helix was issued in 2003 by Great Britain.

or nitrogen as the atoms that carry the hydrogen, giving three alternative arrangements: O–H–O, N–H–N and O–H–N.

Each base pair consists of one larger purine base paired with a smaller pyrimidine base. So, although the bases themselves differ in size, all of the allowed base pairs are the same width, providing for a uniform width of the helix. The A-T base pair has two hydrogen bonds and the G-C base pair is held together by three, as shown in Figure 3.10. Before the hydrogen bonds form and the bases pair off, the shared hydrogen atom is found attached to one or the other of the two bases (shown by the complete lines in Fig. 3.10). During **base pairing**, this hydrogen also bonds to an atom of the second base (shown by the dashed lines).

Although RNA is normally single-stranded, many RNA molecules fold up, giving double-stranded regions. In addition, a strand of RNA may be found paired with one of DNA under some circumstances. Furthermore, the genome of certain viruses consists of double-stranded RNA (see Ch. 17). In all of these cases, the uracil in RNA will base pair with adenine. Thus the base-pairing properties of the uracil found in RNA are identical to those of the thymine of DNA.

Complementary Strands Reveal the Secret of Heredity

Due to the rules for base-pairing, the sequence of a DNA strand can be deduced if the sequence of its partner is known.

If one of the bases in a base pair of double stranded DNA is known, then the other can be deduced. If one strand has an A, then the other will have a T, and vice versa. Similarly, G is always paired with C. This is termed complementary base pairing. The significance is that if the base sequence of either one of the strands of a DNA molecule is known, the sequence of the other strand can be deduced. Such mutually deducible sequences are known as **complementary sequences**. It is this complementary nature of a DNA double helix that allows genetic information to be inherited. Upon cell division, each daughter cell must receive a copy of the parental genome.

base pairing A pair of two complementary bases (A with T or G with C) held together by hydrogen bonds
complementary sequences Two nucleic acid sequences whose bases pair with each other because A, T, G, C in one sequence correspond to T, A, C, G, respectively, in the other

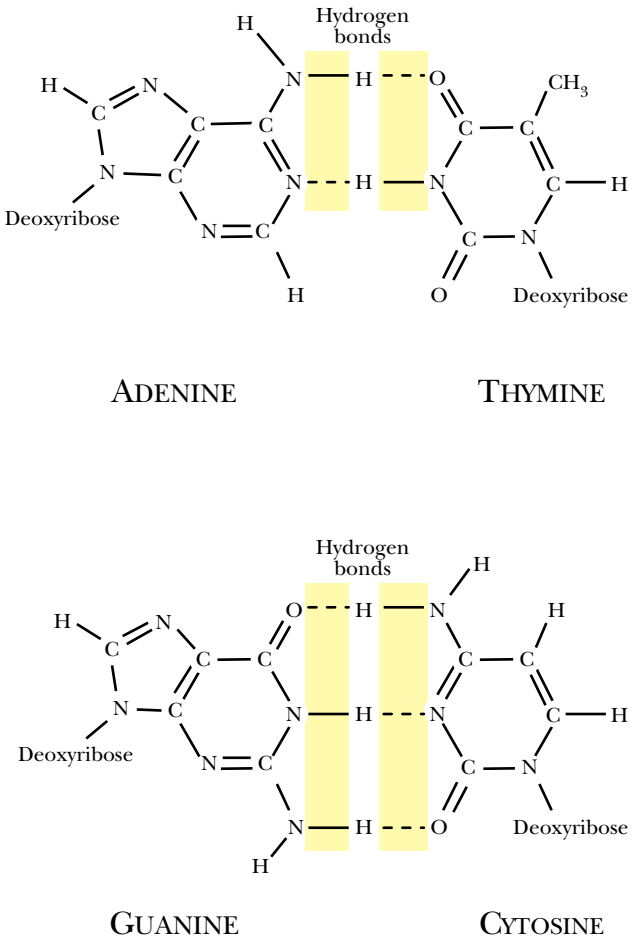


FIGURE 3.10 ***Base Pairing by Hydrogen Bond Formation***

Purines (adenine and guanine) pair with pyrimidines (thymine and cytosine) by hydrogen bonding (colored regions). When the purines and pyrimidines first come together, they form the bonds indicated by the dotted lines.

This requires accurate duplication or replication of the DNA (Fig. 3.11). This is achieved by separating the two strands of DNA and using complementary base pairing to make a new partner for each original strand (see Ch. 5 for details).

Constituents of Chromosomes

Genetic information includes both the genes themselves and regions of DNA involved in controlling gene expression.

Genes are segments of large DNA molecules known as **chromosomes** (Fig. 3.12). Each chromosome is thus an exceedingly long single molecule of DNA. In addition to the DNA, which comprises the genes themselves, the chromosome has some accessory protein molecules, which help maintain its structure. The term **chromatin** refers to this mixture of DNA and protein, especially as observed with the microscope in the nuclei of eukaryotic cells. The genes are arranged in linear order. In front of each gene is a **regulatory region** of DNA involved in switching the gene on or off. Between genes are spacer regions of DNA often referred to as **intergenic regions**. In prokaryotes, groups of genes may be clustered close together with no intergenic regions. Such clusters are

chromatin Complex of DNA plus protein which constitutes eukaryotic chromosomes
chromosome Structure containing the genes of a cell and made of a single molecule of DNA
intergenic region DNA sequence between genes
regulatory region DNA sequence in front of a gene, used for regulation rather than to encode a protein

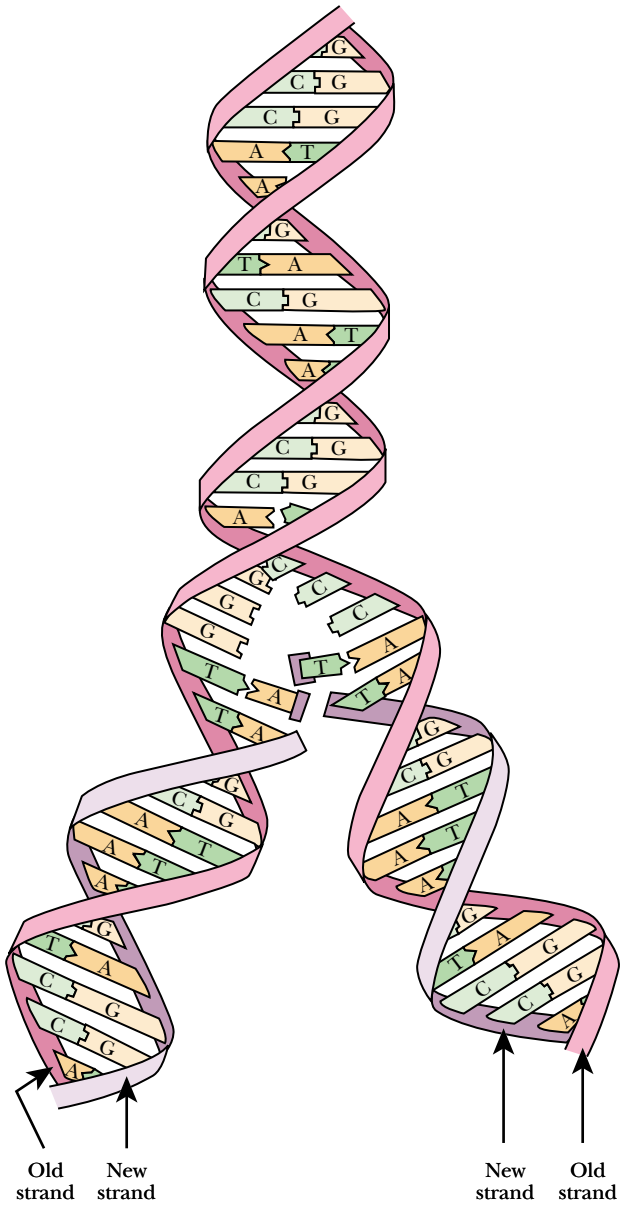


FIGURE 3.11 ***Complementary Strands Allow Duplication***

Because DNA strands are complementary, double-stranded DNA can be split into single strands each carrying sufficient information to recreate the original molecule. Complementary base pairing allows the synthesis of two new strands so restoring double-stranded DNA.

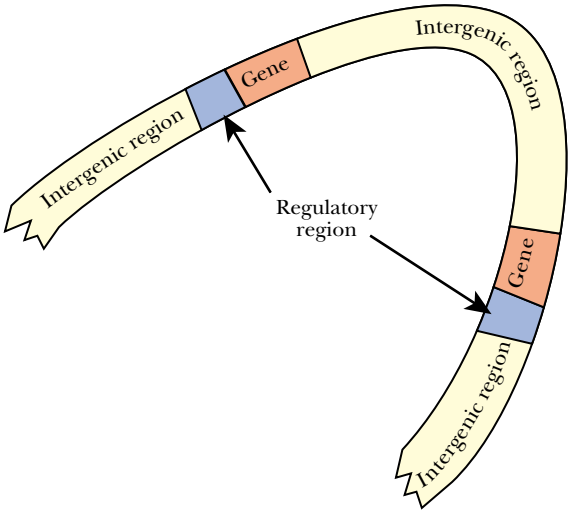


FIGURE 3.12 ***The General Pattern of Information on a Chromosome***

Genes are normally preceded by regions of DNA involved in regulation. Between the genes are regions of DNA that apparently do not carry useful genetic information. These are called intergenic regions and vary greatly in size.

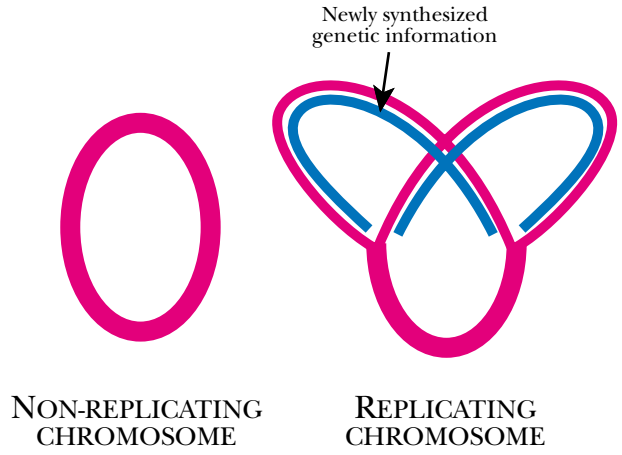


FIGURE 3.13 ***The Circular Bacterial Chromosome and Its Replication***

The bacterial chromosome is circular and not linear. When the double stranded DNA is duplicated, the chromosome is opened forming loops that allow replication of each DNA strand.

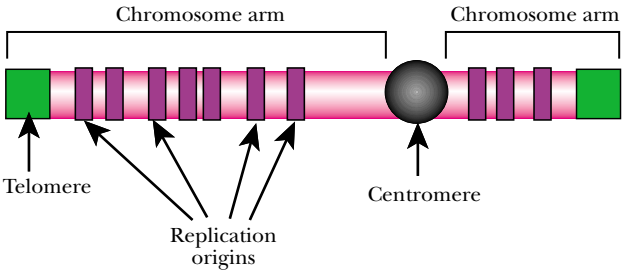


FIGURE 3.14 ***Structural Components of a Eukaryotic Chromosome***

The eukaryotic chromosome is a linear molecule with specific DNA sequences called telomeres at each end. More or less in the center is an organized region called the centromere that is involved in chromosome division. Along the chromosome are multiple regions where replication is initiated.

Details of replication mechanism and structure vary between the linear chromosomes of eukaryotes and the circular chromosomes of most bacteria.

called **operons** and each is under the control of a single regulatory region. Operons are transcribed to give single mRNA molecules, each consisting of several genes.

Chromosomes from bacteria are circular molecules of double-stranded DNA. Since bacteria generally have only around 3,000–4,000 genes, and the intergenic regions are very short, one chromosome is sufficient to accommodate all of their genes. When bacteria divide, the chromosome opens up at the origin of replication and replication proceeds around the circle in both directions (Fig. 3.13).

Chromosomes from higher organisms such as animals and plants are linear molecules of double stranded DNA. They have a **centromere**, usually located more or less in the middle, and structures known as **telomeres** at the two ends (see Fig. 3.14). Both centromeres and telomeres contain special repetitive DNA sequences allowing their recognition by particular proteins. [One exception to this rule is that the yeast, *Saccharomyces*, lacks repetitive sequences at its centromere. However, this is not a general property of fungi, as other fungi do have repetitive centromere sequences.] The centromere is used at cell division when the chromosomes replicate. The newly divided daughter chromosomes are pulled apart by spindle fibers (or microtubules) attached to the centromeres via protein structures known as **kinetochores**. Due to the mechanism of initiating DNA synthesis (see Ch. 5), the far ends of linear DNA molecules are shortened by a few bases each round of replication. In those cells that are permitted to continue growing and dividing, the end sequences are repaired by the enzyme **telomerase**. Telomeres are critically important in cell differentiation, cancer and aging.

Higher organisms have much more DNA than bacteria. This is partly because they possess more genes–higher eukaryotes may have up to 50,000 genes. However, the major reason is that eukaryotes have much longer intergenic regions and other **non-coding DNA**. In fact, as shall be discussed later (Ch. 4), in higher eukaryotes, the genes are only a small proportion of the total DNA. Consequently, higher organisms need

centromere Region of eukaryotic chromosome, usually more or less central, where the microtubules attach during mitosis and meiosis
kinetochore Protein structure that attaches to the DNA of the centromere during cell division and also binds the microtubules
non-coding DNA DNA sequences that do not code for proteins or functional RNA molecules
operon A cluster of prokaryotic genes that are transcribed together to give a single mRNA (i.e. polycistronic mRNA)
telomerase Enzyme that adds DNA to the end, or telomere, of a chromosome
telomere Specific sequence of DNA found at the end of linear eukaryotic chromosomes

FIGURE 3.15 ***A Set of Human Chromosomes***

A human karyotype is a complete set of chromosomes containing 22 pairs plus one "X" and one "Y" chromosome (lower right) if the individual is male (as shown here). Females possess two "X" chromosomes. Courtesy of Alfred Pasieka, Science Photo Library.

several chromosomes to accommodate all their DNA. Since eukaryotes are usually diploid, their chromosomes come in pairs. For example, humans have two duplicate sets of 23 different chromosomes, making a total of 46 chromosomes. The chromosomes are only visible under the light microscope during cell division and it is then that a complete set of chromosomes can be visualized (Fig. 3.15). The complete set of chromosomes found in the cells of a particular individual is known as the **karyotype**.

Humans have vast amounts of DNA making up 46 linear chromosomes. However, most of this is non-coding DNA as discussed further in Chapters 4 and 24.

Chromosomes and specific regions of chromosomes may be identified by their staining patterns after using specific stains that emphasize regions lacking genes. This **chromosome banding technique** has been used to identify major chromosome abnormalities (Fig. 3.16).

As shown in Chapter 2, chromosomes of higher organisms are found in a separate membranous compartment of the cell, the nucleus. The nucleus is divided off from the rest of the cell by the nuclear envelope, consisting of two concentric membranes. The genes control the rest of the cell by dispatching genetic information in the form of special messenger molecules, the **messenger RNA**, through pores in the nuclear envelope.

The Central Dogma Outlines the Flow of Genetic Information

Genetic information flows from DNA to RNA to protein during cell growth. In addition, all living cells must replicate their DNA when they divide. The **central dogma** of molecular biology is a scheme showing the flow of genetic information during both the growth and division of a living cell (Fig. 3.17). During cell division each daughter cell receives a copy of the genome of the parent cell. As the genome is present in the form of DNA, cell division involves the duplication of this DNA. **Replication** is the process by which two identical copies of DNA are made from an original molecule of DNA. Replication occurs prior to cell division. An important point is that information does not flow from protein to RNA or DNA. However, information flow from RNA "backwards" to DNA is possible in certain special circumstances due to the operation of reverse transcriptase. In addition, replication of RNA occurs in viruses with an RNA genome (neither complication is shown in Fig. 3.17).

Under normal circumstances, genetic information flows from DNA to RNA to protein. As a result, proteins are often referred to as "gene products". Some RNA molecules are also "gene products" as they act without being translated into protein.

The genetic information stored as DNA is not used directly to make protein. During cell growth and metabolism, temporary, working copies of the genes known as

central dogma Basic plan of genetic information flow in living cells which relates genes (DNA), message (RNA) and proteins
chromosome banding technique Visualization of chromosome bands by using specific stains that emphasize regions lacking genes
karyotype The complete set of chromosomes found in the cells of a particular individual
messenger RNA (mRNA) The molecule that carries genetic information from the genes to the rest of the cell
replication Duplication of DNA prior to cell division

FIGURE 3.16 ***Banding Patterns of Human Chromosomes***

Representation of the banding patterns seen in metaphase chromosomes during meiosis. The bands are originally visualized by dyes. The relative distances between these bands are the same for an individual chromosome, so this is a useful way of identifying a particular chromosome. Courtesy of Dept. of Clinical Cytogenetics, Addenbrookes Hospital, Cambridge, UK, Science Photo Library.

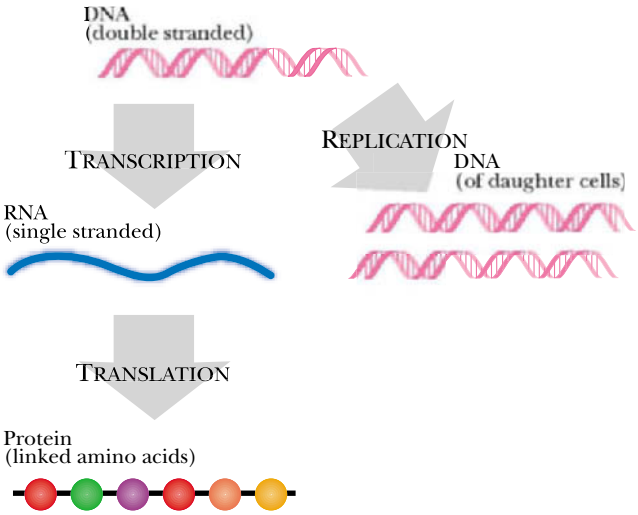


FIGURE 3.17 ***The Central Dogma (Simple Version)***

The information flow in cells begins with DNA, which may either be replicated, giving a duplicate molecule of DNA, or be transcribed to give RNA. The RNA is read (translated) as a protein is built.

messenger RNA (mRNA) are used. These are RNA copies of genetic information stored by the DNA and are made by a process called **transcription**. The messenger RNA molecules carry information from the genome to the cytoplasm, where the information is used by the **ribosomes** to synthesize **proteins**. In eukaryotes, mRNA is not made directly. Instead, transcription yields precursor RNA molecules (pre-mRNA) that must be processed, to produce the actual mRNA as detailed in Ch. 12.

The DNA that carries the primary copy of the genes is present as gigantic molecules, each carrying hundreds or thousands of genes. In contrast, any individual messenger RNA molecule carries only one or a few genes' worth of information. Thus, in practice, multiple short segments of DNA are transcribed simultaneously to give many different messenger RNA molecules. In eukaryotes, each mRNA normally carries only a single gene, whereas in prokaryotes, anywhere from one to a dozen genes may be transcribed as a block to give an mRNA molecule carrying several genes, usually with related functions (Fig. 3.18).

Translation is the synthesis of proteins using genetic information carried by messenger RNA. Proteins consist of one or more polymer chains known as **polypeptides**. These are made from subunits called **amino acids**. Translation thus involves transfer of information from nucleic acids to an entirely different type of macromolecule. This decoding process is carried out by ribosomes. These submicroscopic machines read the messenger RNA and use the information to make a polypeptide chain. Proteins, which make up about two-thirds of the organic matter in a typical cell, are directly responsible for most of the processes of metabolism. Proteins perform most of the enzyme reactions and transport functions of the cell. They also provide many structural components and some act as regulatory molecules, as described below.

Ribosomes Read the Genetic Code

Proteins are made by a subcellular machine, the ribosome, that uses the genetic code to read information encoded by nucleic acids.

This introductory section will summarize protein synthesis as it occurs in bacteria. It should be noted that the details of protein synthesis differ between bacteria and higher organisms (see Ch. 8). The bacterial ribosome, as described in more detail below, consists of two subunits, small (30S) and large (50S). **S-values** tell how fast a particle

amino acid Monomer from which the polypeptide chains of proteins are built
polypeptide chain A polymer that consists of amino acids
protein Polymer made from amino acids; may consist of several polypeptide chains
ribosome The cell's machinery for making proteins
S-value The sedimentation coefficient is the velocity of sedimentation divided by the centrifugal field. It is dependent on mass and is measured in Svedberg units
transcription Conversion of information from DNA into its RNA equivalent
translation Making a protein using the information provided by messenger RNA

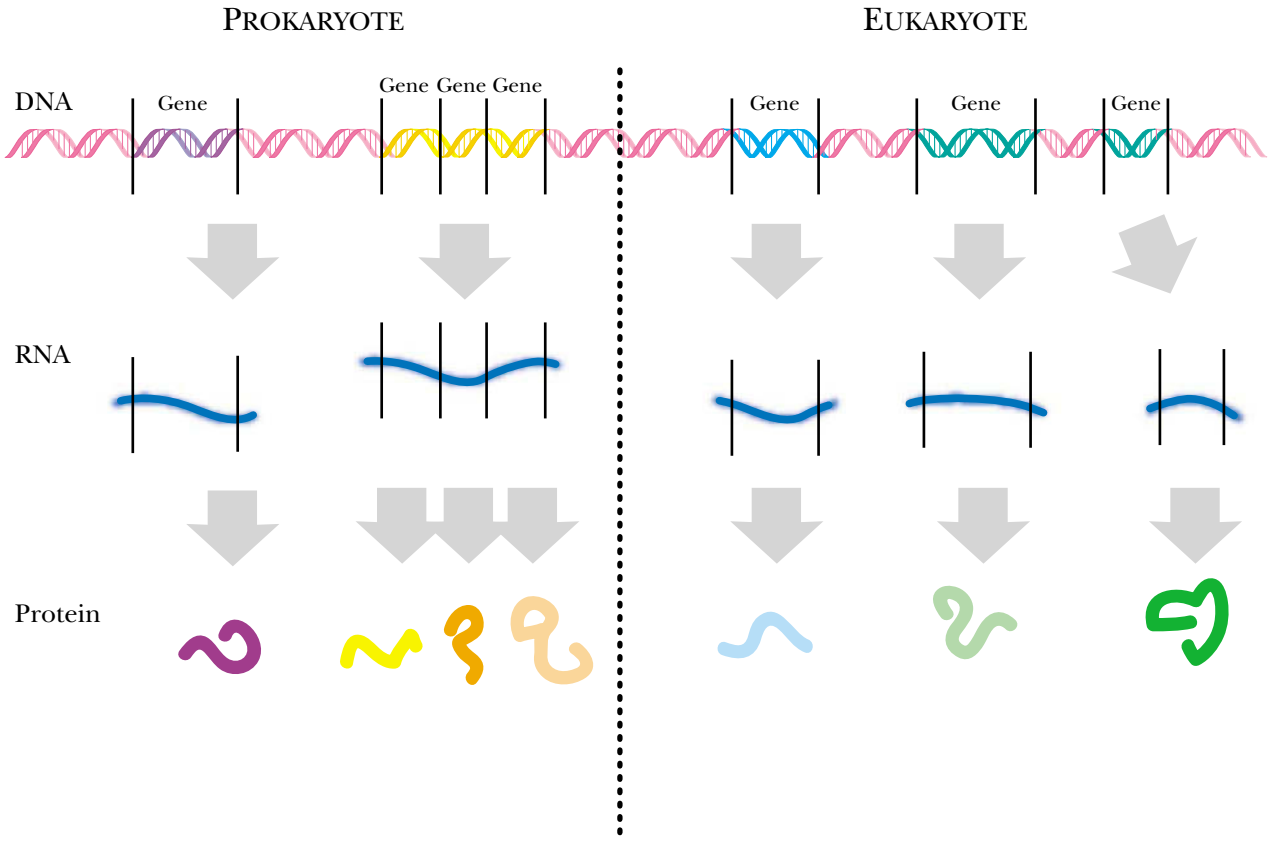


FIGURE 3.18 ***Differing Patterns of Transcription***

In eukaryotes, each gene is transcribed to give a separate mRNA that encodes only a single protein. In prokaryotes, an mRNA molecule may carry information from a single gene or from several genes that are next to each other on the chromosome.

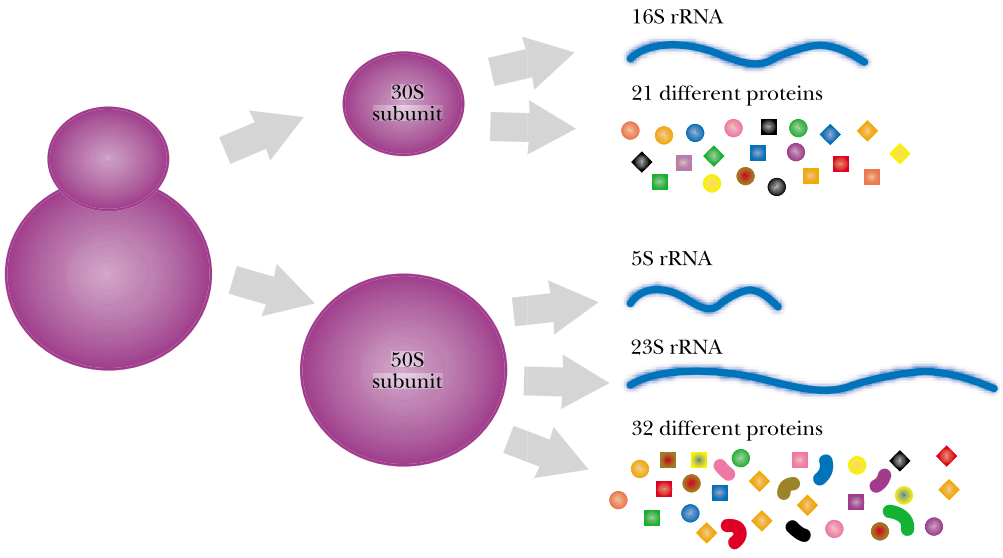


FIGURE 3.19 ***Structural Components of a Ribosome***

The bacterial ribosome can be broken down into two smaller subunits and finally into RNA molecules and proteins.

1st base	2nd (middle) base U	C	A	G	3rd base
U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA stop UAG stop	UGU Cys UGC Cys UGA stop UGG Trp	U C A G
C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg	U C A G
A	AUU Ile AUC Ile AUA Ile AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg	U C A G
G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gly GGC Gly GGA Gly GGG Gly	U C A G

FIGURE 3.20 ***The Genetic Code***

A codon consisting of three base pairs determines each amino acid to be added to a growing polypeptide chain. The codon table shows the 64 different codons, in RNA language, alongside the amino acids they encode. Three of the codons act as stop signals. AUG (methionine) and GUG (valine) act as start codons.

sediments in an ultracentrifuge. They give a rough indication of size but are not linearly related to molecular weight. A complete ribosome with a 30S and a 50S subunit has an S-value of 70S (not 80S).

By weight, the ribosome itself consists of about two-thirds **ribosomal RNA** (**rRNA**) and one-third protein. In bacteria, the large subunit has two rRNA molecules, 5S rRNA and 23S rRNA, and the small subunit has just the one 16S rRNA. In addition to the rRNA, there are 52 different proteins, 31 in the large subunit and the other 21 in the small subunit (Fig. 3.19). The rRNA molecules are NOT themselves translated into protein; instead, they form part of the machinery of the ribosome that translates the mRNA.

The Genetic Code Dictates the Amino Acid Sequence of Proteins

The bases of DNA or RNA are grouped in threes for decoding.

There are 20 amino acids in proteins but only four different bases in the messenger RNA. So nature cannot simply use one base of a nucleic acid to code for a single amino acid when making a protein. During translation, the bases of mRNA are read off in groups of three, which are known as **codons**. Each codon represents a particular amino acid. Since there are four different bases, there are 64 possible groups of three bases; that is, 64 different codons in the **genetic code**. However, there are only 20 different amino acids making up proteins, so some amino acids are encoded by more than one codon. In addition, three of the codons are used for punctuation to stop the growing chain of amino acids (Fig. 3.20). In addition, the codon, AUG, encoding methionine, acts as a start codon. Thus newly made polypeptide chains start with the amino acid methionine. [Much less often, GUG encoding valine, may also act as the start codon. However, even if the start codon is GUG the first amino acid of the newly made protein is methionine (not valine).]

To read the codons a set of adapter molecules is needed. These molecules, known as **transfer RNA** (**tRNA**), recognize the codon on the mRNA at one end and carry the corresponding amino acid attached to their other end (Fig. 3.21). These adapters represent a third class of RNA and were named transfer RNA since they transport amino acids to the ribosome in addition to recognizing the codons of mRNA. Since there are numerous codons, there many different tRNAs. [Actually, there are fewer different tRNA molecules than codons as some tRNA molecules can read multiple codons—see Ch. 8 for details.] At one end, the tRNA has an **anticodon** consisting of three bases that are complementary to the three bases of the codon on the messenger RNA. The

anticodon Group of three complementary bases on tRNA that recognize and bind to a codon on the mRNA
codon Group of three RNA or DNA bases that encodes a single amino acid
genetic code The code for converting the base sequence in nucleic acids, read in groups of three, into the sequence of a polypeptide chain
ribosomal RNA (rRNA) Class of RNA molecule that makes up part of the structure of a ribosome
transfer RNA (tRNA) RNA molecules that carry amino acids to a ribosome

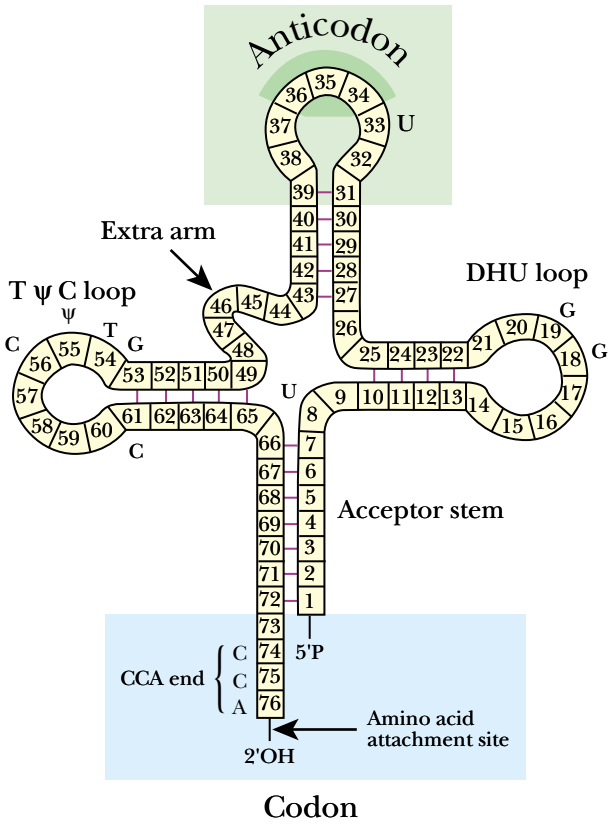


FIGURE 3.21 ***Transfer RNA Contains the Anticodon***

Each transfer RNA molecule has an anticodon that is complementary to the codon carried on the messenger RNA. The codon and anticodon bind together by base pairing. At the far end of the tRNA is the acceptor stem ending in the bases CCA (cytosine, cytosine, adenine). Here is attached the amino acid that corresponds to the codon on the mRNA.

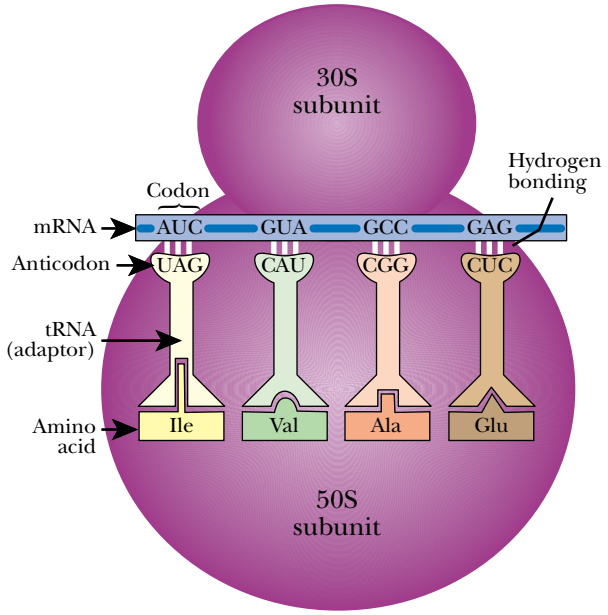


FIGURE 3.22 ***Stylized Relationship of Charged tRNA to mRNA and the Ribosome***

Note: This figure does not show the correct physical arrangement—instead it illustrates the coding relationships between the tRNA and mRNA. The mRNA binds to the 30S subunit of the ribosome. The anticodons of the tRNAs carrying amino acids bind to the corresponding codons on the mRNA. In real life only two tRNAs are present on the ribosome at any given time and the codons on the mRNA are contiguous, with no gaps between.

codon and anticodon recognize each other by base pairing and are held together by hydrogen bonds. At its other end, each tRNA carries the amino acid corresponding to the codon it recognizes.

The small (30S) subunit binds the messenger RNA and the large (50S) subunit is responsible for making the new polypeptide chain. Figure 3.22 shows the relationship between the mRNA and the tRNAs in a stylized way. In practice, only two tRNA molecules are base-paired to the messenger RNA at any given time. After binding to the mRNA, the ribosome moves along it, adding a new amino acid to the growing polypeptide chain each time it reads a codon from the message (Fig. 3.23). A more detailed account of protein synthesis is given in Chapter 8.

FIGURE 3.23 ***Ribosome Elongating a Polypeptide Chain***

A new amino acid is added to the polypeptide chain each time a new tRNA arrives at the ribosome, bringing its attached amino acid. The anticodon of the tRNA binds to the mRNA. The large subunit cross links the incoming amino acid to the growing chain, such that the incoming tRNA ends up carrying the growing polypeptide chain. The 30S subunit of the ribosome then moves one step along the mRNA. This results in ejection of the left-most tRNA and readies the mRNA to accept the next incoming tRNA. The polypeptide chain continues to grow until a "stop codon" is reached.

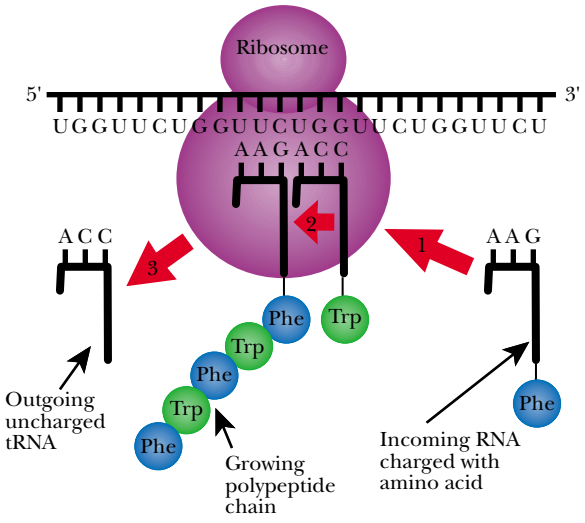


TABLE 3.02 Major Classes of Non-Translated RNA

Name	Function
Ribosomal RNA	comprises major portion of ribosome and is involved in synthesis of polypeptide chains
Transfer RNA	carries amino acids to ribosome and recognizes codons on mRNA
Small nuclear RNA	involved in the processing of messenger RNA molecules in the nucleus of eukaryotic cells (also called snRNA, or "snurps")
Guide RNA	involved in processing of RNA or DNA in some organisms
Regulatory RNA	functions in the regulation of gene expression by binding to proteins or DNA or to other RNA molecules
Antisense RNA	functions in regulating gene expression by base pairing to mRNA
Recognition RNA	part of a few enzymes (e.g., telomerase); enables them to recognize certain short DNA sequences
Ribozymes	enzymatically active RNA molecules

Various Classes of RNA Have Different Functions

RNA is not so simple after all. Several classes of RNA exist that carry out a variety of roles in addition to carrying information for protein synthesis. See especially Chapter 11 for novel insights into the role of RNA in regulation.

Originally, genes were regarded as units of heredity and alleles were defined as alternative versions of a gene. However, these concepts have been broadened as knowledge of genome structure has increased. Molecular insights led first to the view of genes as segments of DNA encoding proteins—the one gene—one enzyme model of Beadle and Tatum. In this case, messenger RNA acts as an intermediary between the DNA, which is used for storage of genetic information, and the protein, which functions in running the cell. The concept of a gene was then further extended to include segments of DNA that encode RNA molecules that are not translated into protein but function as RNA. The most common examples are the ribosomal RNA and transfer RNA involved in protein synthesis. The term "gene products" therefore refers to such non-translated RNA molecules as well as proteins. For convenience, the major classes of non-translated RNA are summarized in Table 3.02.

In addition to the chemical differences discussed above (ribose instead of deoxyribose and uracil instead of thymine), RNA differs from DNA in several respects. RNA is usually single stranded, although most RNA molecules do fold up, thus producing double stranded regions. RNA molecules are usually much shorter than DNA and only

antisense RNA RNA complementary in sequence to messenger RNA and which, therefore, base pairs with it and prevents translation
ribozyme RNA molecule that acts as an enzyme

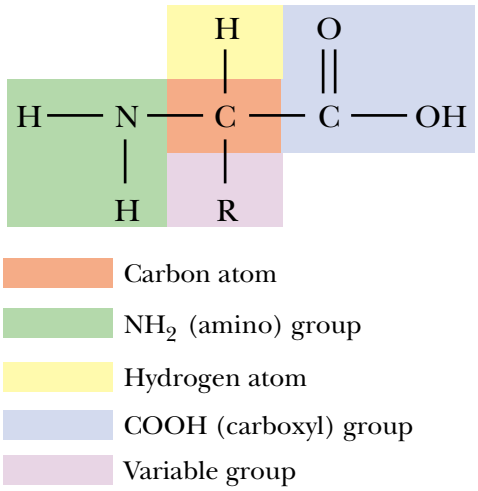


FIGURE 3.24 ***General Features of Amino Acids***

Almost all amino acids found in proteins have the features shown in common. In glycine, the simplest amino acid, the R group is a single hydrogen atom. In proline, the R group consists of a ring structure that bonds to the nitrogen atom shown. This therefore only has a single attached hydrogen and becomes an imino group (—NH—).

carry the information for one or a few genes. Moreover, RNA is usually much shorter-lived than DNA, which is used for long-term storage of the genome. Some classes of RNA molecules, especially tRNA, contain unusual, chemically modified bases that are never found in DNA (see Ch. 8).

The above differences in function between RNA and DNA apply to living cells. However, certain viruses carry their genomes as either single or double-stranded RNA. In such cases, multiple genes will obviously be present on these RNA genomes. Furthermore, double-stranded viral RNA can form a double helix, similar though not identical in structure to that of DNA. The properties of viruses and the novel aspects of their genomes are discussed more fully in Chapter 17.

Proteins, Made of Amino Acids, Carry Out Many Cell Functions

Typically, about 60% of the organic matter in a cell is protein. Most of the cell's activities and many of its structures depend on its proteins.

Proteins are multifunctional biological polymers that consist of one or more polypeptide chains. The information carried by messenger RNA is translated to give a polypeptide chain. The linear sequence of nucleotides in the RNA (read in groups of three—i.e. codons) corresponds to the linear sequence of the amino acids that make up the polypeptide chain. That is, the mRNA and the polypeptide chain are co-linear.

Some proteins act as **enzymes** to catalyze biochemical reactions including the generation of energy and the synthesis of nucleotides and their assembly into nucleic acids. Other proteins are structural, or transport nutrients or take part in cell movement (mechanical proteins). Finally, there are proteins involved in information processing. Molecules whose primary role is to carry information (nucleic acids like DNA and messenger RNA) are basically linear molecules with a regular repeating structure. Molecules that form cellular structures or have active roles carrying out reactions are normally folded into three-dimensional (3-D) structures. These include both proteins and most non-translated RNA molecules, including tRNA and rRNA.

Proteins are made from a linear chain of monomers, known as amino acids (Fig. 3.24), and are folded into a variety of complex 3-D shapes. A chain of amino acids is called a **polypeptide chain** (Fig. 3.25). There are 20 different amino acids used in making proteins. All have a central carbon atom, the **alpha carbon**, surrounded by a hydrogen atom, an amino group (NH_2), a carboxyl group (COOH), and a variable side chain, the **R-group** (Fig. 3.24). Amino acids are joined together by **peptide bonds** (Fig. 3.25). The first amino acid in the chain retains its free amino group and this end is often called the

alpha carbon The central carbon atom of an amino acid, to which the amino, carboxyl and R groups are attached
enzyme A protein or RNA molecule that catalyses a chemical reaction
peptide bond Type of chemical linkage holding amino acids together in a protein molecule
polypeptide chain Polymeric chain of amino acids
R-group Chemical group forming side chain of amino acid

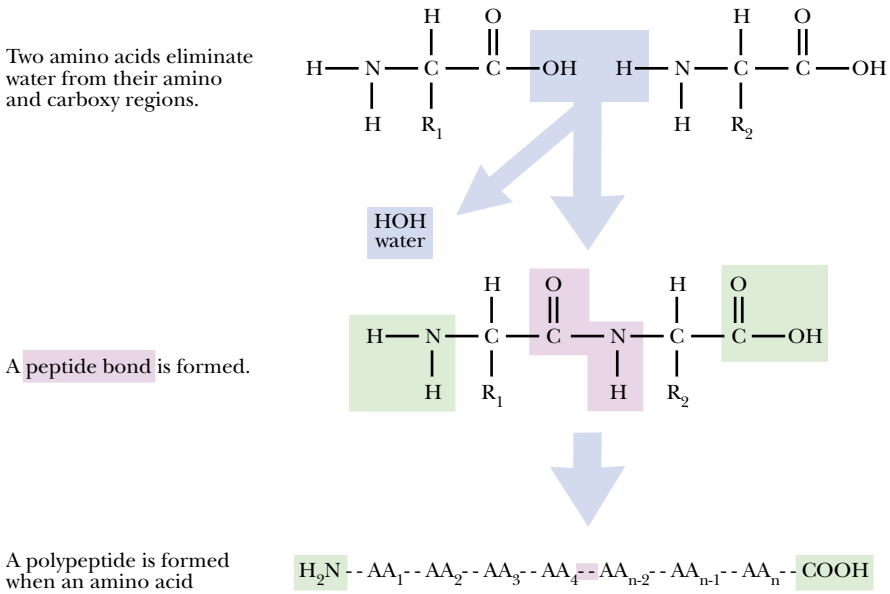


FIGURE 3.25 ***Formation of a Polypeptide Chain***

A polypeptide chain is formed as amino and carboxyl groups on two neighboring amino acids combine and eliminate water. The linkage formed is known as a peptide bond. No matter how many amino acids are added, the growing chain always has an N- or amino terminus, and a C- or carboxy terminus.

amino- or **N-terminus** of the polypeptide chain. The last amino acid to be added is left with a free carboxyl group and this end is often called the **carboxy-** or **C-terminus**.

Some proteins consist of a single polypeptide chain; others contain more than one. To function properly, many proteins need extra components, called **cofactors** or **prosthetic groups**, which are not made of amino acids. Many proteins use single metal atoms as cofactors; others need more complex organic molecules.

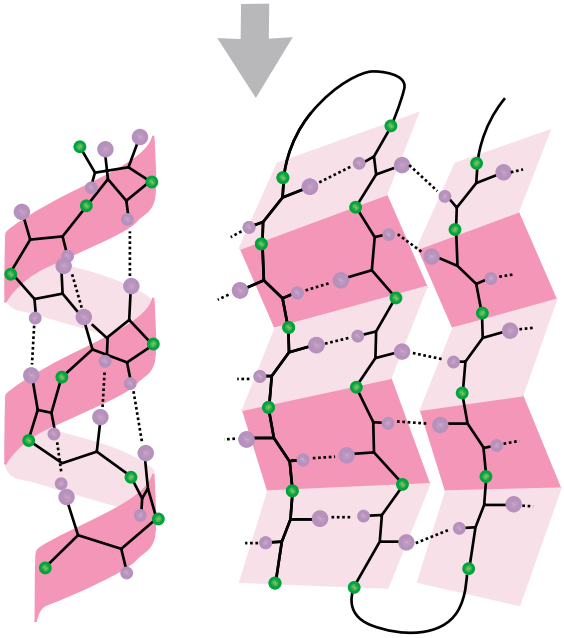
The Structure of Proteins Has Four Levels of Organization

For a protein to be functional, the polypeptide chains must be folded into their correct 3-D structures. The structures of biological polymers, both protein and nucleic acid, are often divided into levels of organization (Fig. 3.26). The first level, or **primary structure**, is the linear order of the monomers—i.e., the sequence of the amino acids for a protein, or of the nucleotides in the case of DNA or RNA. **Secondary structure** is the folding or coiling of the original polymer chains by means of hydrogen bonding. Although DNA is not a protein, hydrogen bonding between base pairs forms the famous double helix. In proteins, hydrogen bonding between peptide groups results in several possible helical or wrinkled sheet-like structures (see Ch. 7 for details).

The next level is the **tertiary structure**. The polypeptide chain, with its preformed regions of secondary structure, is then folded to give the final 3-D structure. This level of folding depends on the side chains of the individual amino acids. In certain cases, proteins known as chaperonins help other proteins to fold correctly (see Ch. 7). As there are 20 different amino acids, a great variety of final 3-D conformations is possible. Nonetheless, many proteins are roughly spherical. Lastly, **quaternary structure** is the assembly of several individual polypeptide chains to give the final structure. Not

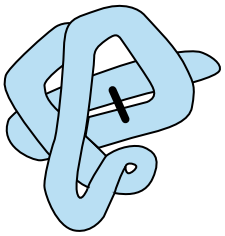
amino- or N-terminus The end of a polypeptide chain that is made first and that has a free amino group
carboxy- or C-terminus The end of a polypeptide chain that is made last and has a free carboxy-group
cofactor Extra chemical group non-covalently attached to a protein that is not part of the polypeptide chain
primary structure The linear order in which the subunits of a polymer are arranged
prosthetic group Extra chemical group covalently attached to a protein that is not part of the polypeptide chain
quaternary structure Aggregation of more than one polymer chain in final structure
secondary structure Initial folding up of a polymer due to hydrogen bonding
tertiary structure Final 3-D folding of a polymer chain

(a) Primary structure

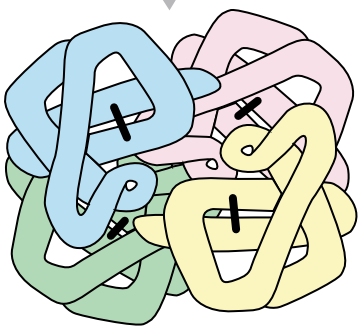


α helix β sheet

(b) Secondary structures



(c) Tertiary structure



(d) Quaternary structure

FIGURE 3.26 ***Four Levels of Protein Structure***

The final protein structure is best understood by following the folding process from simple to complex. The primary structure is the specific order of the amino acids (a). The secondary structure is due to regular folding of the polypeptide chain due to hydrogen bonding (b). The tertiary structure results from further folding of the polypeptide due to interactions between the amino acid side chains (c). Finally, the quaternary structure is the assembly of multiple polypeptide chains (d).

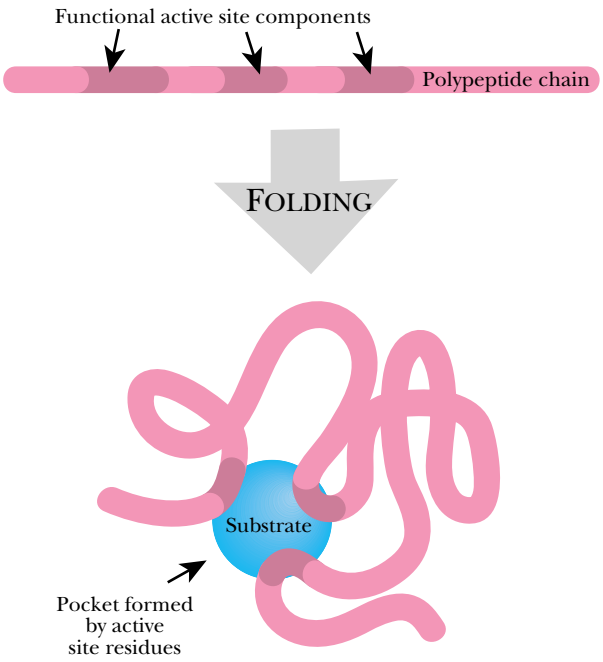


FIGURE 3.27 ***Polypeptide Forms an Active Site after Folding***

Folding of the protein brings together several regions of the polypeptide chain that are needed to perform its biological role. The active site forms a pocket for binding the substrate. Some of the amino acid residues at the active site are also involved in chemical reactions with the substrate.

all proteins have more than one polypeptide chain; some just have one, so they have no quaternary structure.

Proteins Vary in Their Biological Roles

Functionally, proteins may be divided into four main categories: **structural proteins**, enzymes, **regulatory proteins** and **transport proteins**.

1. Structural proteins make up many sub-cellular structures. The flagella with which bacteria swim around, the microtubules used to control traffic flow inside cells of higher organisms, the fibers involved in contractions of a muscle cell, and the outer coats of viruses are a few examples of structures constructed using proteins.
2. Enzymes are proteins that facilitate chemical reactions. An enzyme first binds another molecule, known as its **substrate**, and then performs some chemical operations with it. Some enzymes bind only a single substrate molecule; others may bind two or more, and react them together to make the final product. In any case, the enzyme needs an **active site**, a pocket or cleft in the protein, where the substrate binds and the reaction occurs. The active site of the protein is produced by the folding up of its polypeptide chain correctly so that amino acid residues that were spread out at great distances in the linear chain now come together and will cooperate in binding the substrate to facilitate the enzyme reaction (Fig. 3.27).
3. Although regulatory proteins are not enzymes, they do bind other molecules and so they also need active sites to accommodate these. Regulatory proteins vary enormously. Many of them can bind both small signal molecules and DNA. The presence or absence of the signal molecule determines whether or not the gene is switched on (Fig. 3.28).

active site Special site or pocket on a protein where the substrate binds and the enzyme reaction occurs
regulatory protein A protein that regulates the expression of a gene or the activity of another protein
structural protein A protein that forms part of a cellular structure
substrate The molecule altered by the action of an enzyme
transport protein A protein that carries other molecules across membranes or around the body

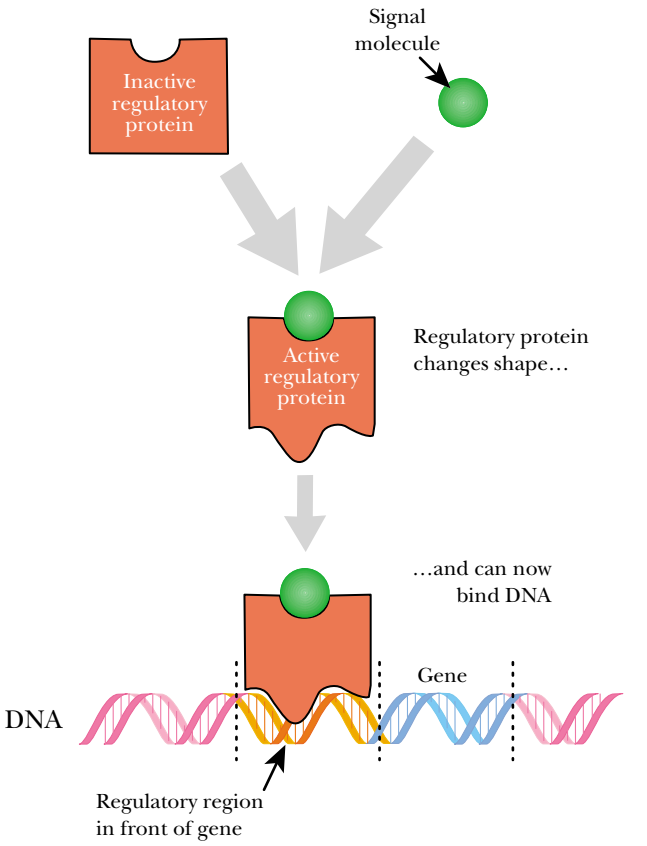


FIGURE 3.28 ***Regulatory Protein***

Regulatory proteins usually exist in two conformations. Receiving a signals promotes a change in shape. The regulatory protein may then bind to DNA and alter the expression of a gene.

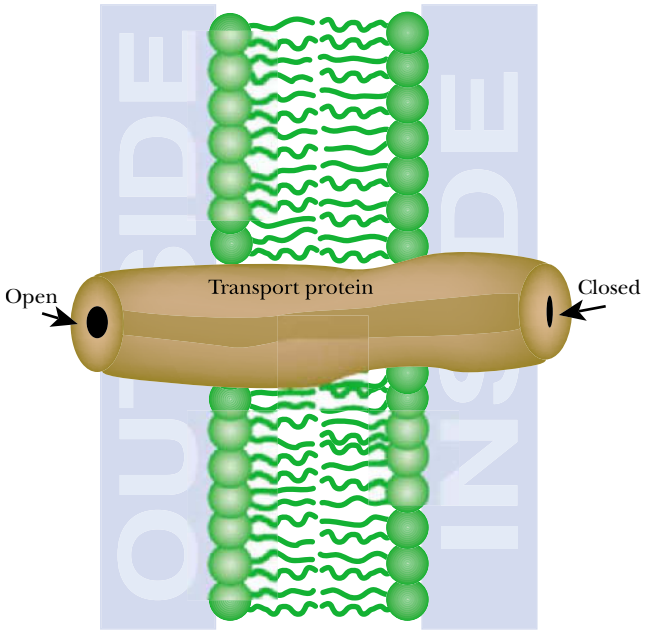


FIGURE 3.29 ***Transport Proteins***

Transport proteins are often found in cell membranes where they are responsible for the entry of nutrients or the export of waste products.

4. Transport proteins are found mostly in biological membranes, as shown in Figure 3.29, where they carry material from one side to the other. Nutrients, such as sugars, must be transported into cells of all organisms, whereas waste products are deported. Multi-cellular organisms also have transport proteins to carry materials around the body. An example is hemoglobin, which carries oxygen in blood.