Gene Expression and the Synthesis of Proteins

n the middle of the 20th century, DNA was identified as the genetic material, and its structure was determined. Using this knowledge, researchers then discovered the mechanisms by which genetic information is inherited and expressed. During the last quarter of the 20th century, our understanding of this critical area of science, known as molecular biology, grew at an increasingly rapid pace. We now have techniques to probe the human genome that will completely revolutionize the way medicine is practiced in the 21st century.

The genome of a cell consists of all its genetic information, encoded in DNA (deoxyribonucleic acid). In eukaryotes, DNA is located mainly in nuclei, but small amounts are also found in mitochondria. Nuclear genes are packaged in chromosomes that contain DNA and protein in tightly coiled structures (Chapter 12).

The molecular mechanism of inheritance involves a process known as replication, in which the strands of parental DNA serve as templates for the synthesis of DNA copies (Fig. 1) (Chapter 13). After DNA replication, cells divide, and these DNA copies are passed to daughter cells. Alterations in genetic material occur by recombination (the exchange of genetic material between chromosomes) and by mutation (the result of chemical changes that alter DNA). DNA repair mechanisms correct much of this damage, but, nevertheless, many gene alterations are passed to daughter cells.

The expression of genes within cells requires two processes: transcription and translation (see Fig. 1) (Chapters 14 and 15). DNA is transcribed to produce ribonucleic acid (RNA). Three major types of RNA are transcribed from DNA and subsequently participate in the process of translation (The synthesis of proteins). Messenger RNA (mRNA) carries the genetic information from the nucleus to the cytoplasm, where translation occurs on ribosomes, structures containing proteins complexed with ribosomal RNA (rRNA). Transfer RNA (tRNA) carries individual amino acids to the ribosomes, where they are joined in peptide linkage to form proteins. During translation, the sequence of nucleic acid bases in mRNA is read in sets of three (each set of three bases constitutes a codon). The sequence of codons in the mRNA dictates the sequence of amino acids in the protein. Proteins function in cell structure, signaling, and catalysis and, therefore, determine the appearance and behavior of cells and the organism as a whole. The regulation of gene expression (Chapter 16) determines which proteins are synthesized and the amount synthesized at any time, thus allowing cells to undergo development and differentiation and to respond to changing environmental conditions.

Research in molecular biology has produced a host of techniques, known collectively as recombinant DNA technology, biotechnology, or genetic engineering, that can be used for the diagnosis and treatment of disease (Chapter 17). These techniques can detect a number of genetic diseases that previously could only be diagnosed after symptoms appeared. Diagnosis of these diseases can now be made with considerable accuracy even before birth, and carriers of these diseases also can be identified.



Fig. 1. Replication, transcription, and translation. Replication: DNA serves as a template for producing DNA copies. Transcription: DNA serves as a template for the synthesis of RNA. Translation: RNA provides the information for the synthesis of proteins.

Many drugs used in medicine to treat bacterial infections are targeted to interfere with their ability to synthesize RNA and proteins. Thus, medical students need to know the basics of bacterial DNA replication, RNA synthesis, and protein synthesis.

Ethical dilemmas have come along with technological advances in moleuclar biology. Consider the case of a patient with a mild case of ornithine transcarbamoylase deficiency, a urea cycle defect that, if untreated, leads to elevated ammonia levels and nervous system dysfunction. The patient was being effectively treated by dietary restriction of protein. However, in 1999, he was treated with a common virus carrying the normal gene for ornithine transcarbamoylase. The patient developed a severe immune response to the virus and died as a result of the treatment. This case history raises the issues of appropriate patient consent, appropriate criteria to be included in this type of study, and the types of diseases for which gene therapy is appropriate. These are issues that you, the student, will be facing as you enter your practice of medicine.



Tumors may be benign or malignant. A tumor is malignant if it invades locally or if cells break

away from the tumor and travel to other parts of the body, where they establish new growths (a process called metastasis), resulting in destruction of the tissues they invade. Many of the drugs used to treat malignant tumors are directed toward inhibition of DNA replication. These chemotherapeutic drugs are more toxic to cancer cells than normal cells, because the cancer cells divide more rapidly. However, such drugs also may inhibit normal rapidly dividing cells, such as the cells of the bone marrow (causing a decrease in white blood cell count) or cells in the hair follicles (resulting in hair loss during chemotherapy). With recent developments in the field of gene therapy, diseases that for centuries have been considered hopeless are now potentially curable. Much of the therapy for these diseases is currently experimental. However, during the 21st century, physicians may be using genetic engineering techniques routinely for both the diagnosis and treatment of their patients.

Replication and cell division are highly regulated processes in the human. Cancer is a group of diseases in which a cell in the body has been transformed and begins to grow and divide out of control (Chapter 18). It results from multiple mutations or changes in DNA structure in the genes that activate cell growth, called proto-oncogenes, and those that ensure that DNA replication and repair are normal, called growth suppressor or tumor suppressor genes. Mutations that activate protooncogenes to oncogenes disturb the regulation of the cell cycle and the rate of cell proliferation. Mutations disrupting tumor suppressor genes lead to an increased incidence of these proto-oncogene–activating mutations. Such mutations may be inherited, causing a predisposition to a type of cancer. They also may arise from DNA replication or copying errors that remain uncorrected, from chemicals or radiation that damages DNA, from translocation of pieces of chromosomes from one chromosome to another during replication, or from incorporation of viral encoded DNA into the genome.

12 Structure of the Nucleic Acids

Nucleotides in DNA and RNA. Nucleotides are the monomeric units of the nucleic acids, DNA (deoxyribonucleic acid) and RNA (ribonucleic acid). Each nucleotide consists of a heterocyclic nitrogenous base, a sugar, and phosphate. DNA contains the purine bases adenine (A) and guanine (G) and the pyrimidine bases cytosine (C) and thymine (T). RNA contains A, G, and C, but it has uracil (U) instead of thymine. In DNA, the sugar is deoxyribose, whereas in RNA it is ribose.

Polynucleotides such as DNA and RNA are linear sequences of nucleotides linked by **3'- to 5'-phosphodiester bonds** between the sugars (Fig. 12.1). The bases of the nucleotides can interact with other bases or with proteins.

DNA Structure. Genetic information is encoded by the sequence of different nucleotide bases in DNA. **DNA** is **double-stranded**; it contains **two antiparallel polynucleotide strands**. The two strands are joined by hydrogen bonding between their bases to form **base-pairs**. Adenine pairs with thymine, and guanine pairs with cytosine. The two DNA strands run in opposite directions. One strand runs 5' to 3', and the other strand runs 3' to 5'. The two DNA strands wind around each other, forming a double helix.

Transcription of a gene generates a **single-stranded RNA** that is identical in nucleotide sequence to one of the strands of the duplex DNA. The three major types of RNA are **messenger RNA** (mRNA), **ribosomal RNA** (rRNA), and **transfer RNA** (tRNA).

RNA Structures. mRNAs contain the nucleotide sequence that is converted into the amino acid sequence of a protein in the process of translation. Eukaryotic mRNA has a structure known as a **cap** at the 5'-end, a sequence of adenine nucleotides (a **poly(A) tail**) at the 3'-end, and a **coding region** in the center containing **codons** that dictate the sequence of amino acids in a protein or relay a signal. Each codon in the genetic code is a different sequence of three nucleotides.

rRNAs and tRNAs are part of the apparatus for protein synthesis, but do not encode proteins. **rRNA** has **extensive internal base pairing** and complexes with proteins to form **ribonucleoprotein particles** called **ribosomes**. The ribosomes bind mRNA and tRNAs during translation. Each **tRNA** binds and **activates a specific amino acid** for insertion into the polypeptide chain and therefore has a somewhat different nucleotide sequence than other tRNAs. A unique trinucleotide sequence on each tRNA called an **anticodon** binds to a complementary codon on the mRNA, thereby ensuring insertion of the correct amino acid. In spite of their differences, all tRNAs contain a number of unusual nucleotides and assume a similar **cloverleaf** structure.



Fig. 12.1. Structure of a polynucleotide. The 5'-carbon of the top sugar and the 3'-carbon of the bottom sugar are indicated.



THE WAITING ROOM

Ivy Sharer is a 26-year-old intravenous (IV) drug abuser who admitted to sharing unsterile needles with another addict for several years. Five months before presenting to the hospital emergency department with soaking night sweats, she experienced a 3-week course of a flu-like syndrome with fever, malaise, and muscle aches. Four months ago, she noted generalized lymph node enlargement associated with chills, anorexia, and diarrhea, which led to a 22-lb weight loss. Tests were positive for human immunodeficiency virus (HIV). Because her symptoms indicated that she now had the acquired immunodeficiency syndrome (AIDS), a multidrug regimen including zidovudine (ZDV), formerly called azidothymidine (AZT), was initiated.

Colin Tuma had intestinal polyps at age 45, which were removed via a colonoscope. However, he did not return for annual colonoscopic examinations as instructed. At age 56, he reappeared, complaining of tar-colored stools (melena), which are caused by intestinal bleeding. The source of the blood loss was an adenocarcinoma growing from a colonic polyp of the large intestine. At surgery, it was found that the tumor had invaded the gut wall and perforated the visceral peritoneum. Several pericolic lymph nodes contained cancer cells, and several small nodules of metastatic cancer were found in the liver. After resection of the tumor, the oncologist began treatment with 5-fluorouracil (5-FU) combined with other chemotherapeutic agents.

Agneu ("neu") Moania complains to his physician of a fever and cough. His cough produces thick yellow-brown sputum. A stain of his sputum shows many Gram-positive, bullet-shaped diplococci. A sputum culture confirms that he has pneumonia, a respiratory infection caused by *Streptococcus pneumoniae*, which is sensitive to penicillin, erythromycin, tetracycline, and other antibiotics. Because of a history of penicillin allergy, he is started on oral erythromycin therapy.

I. DNA STRUCTURE

A. Location of DNA

DNA and RNA serve as the genetic material for prokaryotic and eukaryotic cells, for viruses, and for plasmids, each of which stores it in a different arrangement or location. In prokaryotes, DNA is not separated from the rest of the cellular contents. In eukaryotes, however, DNA is located in the nucleus, where it is separated from the rest of the cell by the nuclear envelope (see Fig. 10.20). Eukaryotic DNA is bound to proteins, forming a complex called chromatin. During interphase (when cells are not dividing), some of the chromatin is diffuse (euchromatin) and some is dense (heterochromatin), but no distinct structures can be observed. However, before mitosis (when cells divide), the DNA is replicated, resulting in two identical chromosomes called sister chromatids. During metaphase (a period in mitosis), these condense into discrete, visible chromosomes.

Less than 0.1% of the total DNA in a cell is present in mitochondria. The genetic information in a mitochondrion is encoded in less than 20,000 base pairs of DNA; the information in a human haploid nucleus (i.e., an egg or a sperm cell) is encoded in approximately 3×10^9 (3 billion) base pairs. The DNA and protein synthesizing systems in mitochondria more closely resemble the systems in bacteria, which do

An adenoma is a mass of rapidly proliferating cells, called a neoplasm (neo = new; plasm = growth), that is formed from epithelial cells growing into a glandlike structure. The cells lining all the external and internal organs are epithelial cells, and most human tumors are adenocarcinomas. Adenematous polyps are adenomas that grow into the lumen of the colon or rectum. The term malignant applied to a neoplasm refers to invasive unregulated growth. Colin Tuma has an adenocarcinoma, which is a malignant adenoma that has started to grow through the wall of the colon into surrounding tissues. Cells from adenocarcinomas can break away and spread through the blood or lymph to other parts of the body, where they form "colony" tumors. This process is called metastasis.



DNA is a double-stranded molecule that forms base pairs (bp)

between strands. The bp designation is often used to indicate the size of a DNA molecule. For example, in a stretch of DNA 200 bp long, both strands are included with 200 bases in each strand, for a total of 400 bases.

Purines

 NH_2

not have membrane-enclosed organelles, than those in the eukaryotic nucleus and cytoplasm. It has been suggested that mitochondria were derived from ancient bacterial invaders of primordial eukaryotic cells.

Viruses are small infectious particles consisting of a DNA or RNA genome (but not both), proteins required for pathogenesis or replication, and a protein coat. They lack, however, complete systems for replication, transcription, and translation and, consequently, viruses must invade other cells and commandeer their DNA, RNA, and protein-synthesizing machinery to reproduce. Both eukaryotes and prokaryotes can be infected by viruses. Viruses that infect bacteria are known as bacteriophage (or more simply as phage).

Plasmids are small, circular DNA molecules that can enter bacteria and replicate autonomously, that is, outside the host genome. In contrast to viruses, plasmids are not infectious; they do not convert their host cells into factories devoted to plasmid production. Genetic engineers use plasmids as tools for transfer of foreign genes into bacteria because segments of DNA can readily be incorporated into plasmids.

B. Determination of the Structure of DNA

In 1865, Frederick Meischer first isolated DNA, obtaining it from pus scraped from surgical bandages. Initially, scientists speculated that DNA was a cellular storage form for inorganic phosphate, an important but unexciting function that did not spark widespread interest in determining its structure. In fact, the details of DNA structure were not fully determined until 1953, almost 90 years after it had first been isolated, but only 9 years after it had been identified as the genetic material.

Early in the 20th century, the bases of DNA were identified as the purines adenine (A) and guanine (G), and the pyrimidines cytosine (C) and thymine (T) (Fig. 12.2). The sugar was found to be deoxyribose, a derivative of ribose, lacking a hydroxyl group on carbon 2 (Fig. 12.3).

Nucleotides, composed of a base, a sugar, and phosphate, were found to be the monomeric units of the nucleic acids (Table 12.1). In nucleosides, the nitrogenous base is linked by an *N*-glycosidic bond to the anomeric carbon of the sugar, either ribose or deoxyribose. A nucleotide is a nucleoside with an inorganic phosphate attached to a 5'-hydroxyl group of the sugar in ester linkage (Fig. 12.4). The names and abbreviations of nucleotides specify the base, the sugar, and the number of phosphates attached (MP, monophosphate; DP, diphosphate; TP, triphosphate). In deoxynucleotides, the prefix "d" precedes the abbreviation. For example, GDP is guanosine diphosphate (the base guanine attached to a ribose that has two phosphate groups) and dATP is deoxyadenosine triphosphate (the base adenine attached to a deoxyribose with three phosphate groups).









Fig. 12.2. Purine and pyrimidine bases in DNA.

Table 12.1. Names of Bas	ses and Their Corresponding Nucleosides"
Base	Nucleoside
Adenine (A)	Adenosine
Guanine (G)	Guanosine
Cytosine (C)	Cytidine

Thymine (T)

Hypoxanthine (I)

Uracil (U)

^a If the sugar is deoxyribose rather than ribose, the nucleoside has "deoxy" as a prefix (e.g., deoxyadenosine). Nucleotides are given the name of the nucleoside plus mono, di, or triphosphate (e.g., adenosine triphosphate or deoxyadenosine triphosphate).

Thymidine

Uridine

Inosine^b

^b The base hypoxanthine is not found in DNA but is produced during degradation of the purine bases. It is found in certain tRNA molecules. Its nucleoside, inosine, is produced during synthesis of the purine nucleotides (see Chapter 41).



Fig. 12.3. Deoxyribose and ribose, the sugars of DNA and RNA. The carbon atoms are numbered from 1 to 5. When the sugar is attached to a base, the carbon atoms are numbered from 1' to 5' to distinguish it from the base. In deoxyribose the X = H; in ribose the X = OH.



Fig. 12.4. Nucleoside and nucleotide structures. Shown with ribose as the sugar. The corresponding deoxyribonucleotides are abbreviated dNMP, dNDP, and dNTP. N = any base (A, G, C, U, or T).

In 1944, after Oswald Avery's experiments establishing DNA as the genetic material were published, interest in determining the structure of DNA intensified. Digestion with enzymes of known specificity proved that inorganic phosphate joined the nucleotide monomers, forming a phosphodiester bond between the 3'-carbon of one sugar and the 5'-carbon of the next sugar along the polynucleotide chain (Fig. 12.5). Another key to DNA structure was provided by Erwin Chargaff. He analyzed the base composition of DNA from various sources and concluded that, on a molar basis, the amount of adenine was always equal to the amount of thymine, and the amount of guanine was equal to the amount of cytosine.

During this era, James Watson and Francis Crick joined forces and, using the x-ray diffraction data of Maurice Wilkins and Rosalind Franklin, incorporated the available information into a model for DNA structure. In 1953, they published a brief paper, describing DNA as a double helix consisting of two polynucleotide strands joined by pairing between the bases (adenine with thymine and guanine with cytosine). The model of base-pairing they proposed is the basis of modern molecular biology.

C. Concept of Base-Pairing

As proposed by Watson and Crick, each DNA molecule consists of two polynucleotide chains joined by hydrogen bonds between the bases. In each base pair, a purine on one strand forms hydrogen bonds with a pyrimidine on the other strand. In one type of base pair, adenine on one strand pairs with thymine on the other strand (Fig. 12.6). This base pair is stabilized by two hydrogen bonds. The other base pair, formed between guanine and cytosine, is stabilized by three hydrogen bonds. As a consequence of base-pairing, the two strands of DNA are complementary, that is, adenine on one strand corresponds to thymine on the other strand, and guanine corresponds to cytosine.

The concept of base-pairing proved to be essential for determining the mechanism of DNA replication (in which the copies of DNA are produced that are distributed to daughter cells) and the mechanisms of transcription and translation (in

W th so lar biology.

Watson and Crick's one-page paper, published in 1953, contained little more than 900 words. However it triggered a major revolution in the biologic sciences and produced the conceptual foundation for the discipline of moleculogy.

After Ivy Sharer was diagnosed with AIDS, she was treated with a mixture of drugs including zidovudine (ZDV), formerly called AZT. This drug is an analog of the thymine nucleotide found in DNA (the modified group is shown in the dashed box). ZDV is phosphorylated in the body by the kinases that normally phosphorylate nucleosides and nucleotides. As the viral DNA chain is being synthesized in a human cell, ZDV is then added to the growing 3'-end by viral reverse transcriptase. However, ZDV lacks a 3' OH group and, therefore, no additional nucleotides can be attached through a $5' \rightarrow 3'$ bond. Thus, chain elongation of the DNA is terminated. Reverse transcriptase has a higher affinity for ZDV than does normal human cellular DNA polymerases, enabling the drug to target viral replication more specifically than cellular replication.





Fig. 12.5. A segment of a polynucleotide chain of DNA. The dashes at the 5'- and 3'-ends indicate that the molecule contains more nucleotides than are shown.



Fig. 12.6. Base pairs of DNA. Note that the pyrimidine bases are "flipped over" from the positions in which they are usually shown (see Fig. 12.5). The bases must be in this orientation to form base pairs. The dotted lines indicate hydrogen bonds between the bases. Although the hydrogen bonds hold the bases and thus the two DNA strands together, they are weaker than covalent bonds and allow the DNA strands to separate during replication and transcription.

"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." J.D. Watson and F.H.C. Crick, *Nature,* April 25, 1953.



Fig. 12.7. DNA strands serve as templates. During replication, the strands of the helix separate in a localized region. Each parental strand serves as a template for the synthesis of a new DNA strand.



Multi-drug regimens used to treat cancers (e.g., lymphomas) sometimes include the drug doxorubicin

(adriamycin). It is a natural product with a complex multi-ring structure that intercalates or slips in between the stacked base pairs of DNA and inhibits replication and transcription. which mRNA is produced from genes and used to direct the process of protein synthesis). Obviously, as Watson and Crick suggested, base-pairing allows one strand of DNA to serve as a template for the synthesis of the other strand (Fig. 12.7). Basepairing also allows a strand of DNA to serve as a template for the synthesis of a complementary strand of RNA.

D. DNA Strands Are Antiparallel

As concluded by Watson and Crick, the two complementary strands of DNA run in opposite directions. On one strand, the 5'-carbon of the sugar is above the 3'-carbon (Fig. 12.8). This strand is said to run in a 5' to 3' direction. On the other strand, the 3'-carbon is above the 5'-carbon. This strand is said to run in a 3' to 5' direction. Thus, the strands are antiparallel (that is, they run in opposite directions.) This concept of directionality of nucleic acid strands is essential for understanding the mechanisms of replication and transcription.

E. The Double Helix

Because each base pair contains a purine bonded to a pyrimidine, the strands are equidistant from each other throughout. If two strands that are equidistant from each other are twisted at the top and the bottom, they form a double helix (Fig. 12.9). In the double helix of DNA, the base pairs that join the two strands are stacked like a spiral staircase along the central axis of the molecule. The electrons of the adjacent base pairs interact, generating stacking forces that, in addition to the hydrogen bonding of the base pairs, help to stabilize the helix.

The phosphate groups of the sugar-phosphate backbones are on the outside of the helix (see Fig. 12.9). Each phosphate has two oxygen atoms forming the phosphodiester bonds that link adjacent sugars. However, the third -OH group on the phosphate is free and dissociates a hydrogen ion at physiologic pH. Therefore, each DNA helix has negative charges coating its surface that facilitate the binding of specific proteins.

The helix contains grooves of alternating size, known as the major and minor grooves (see Fig. 12.9). The bases in these grooves are exposed and therefore can interact with proteins or other molecules.



Fig. 12.8. Antiparallel strands of DNA. For the strand on the left, the 5'-carbon of each sugar is above the 3'-carbon, so it runs 5' to 3'. For the strand on the right, the 3'-carbon of each sugar is above the 5'-carbon, so it runs 3' to 5'.

Watson and Crick described the B form of DNA, a right-handed helix, containing 3.4 Å between base pairs and 10.4 base pairs per turn. Although this form predominates in vivo, other forms also occur (Fig. 12.10). The A form, which predominates in DNA-RNA hybrids, is similar to the B form, but is more compact (2.3 Å between base pairs and 11 base pairs per turn). In the Z form, the bases of the two DNA strands are positioned toward the periphery of a left-handed helix . There are 3.8 Å between base pairs and 12 base pairs per turn in Z DNA. This form of the helix was designated "Z" because, in each strand, a line connecting the phosphates "zigs" and "zags."

F. Characteristics of DNA

Both alkali and heat cause the two strands of the DNA helix to separate (denature). Many techniques employed to study DNA or to produce recombinant DNA molecules make use of this property. Although alkali causes the two strands of DNA to



If you look up through the bottom of a helix along the central axis and the helix spirals away from you in a clockwise direction (toward the arrowhead in the drawing), it is a right-handed helix. If it spirals away from you in a counterclockwise direction, it is a lefthanded helix.



Fig. 12.10. Z, B, and A forms of DNA. The solid black lines connect one phosphate group to the next. Modified from Saenger W. Principles of nucleic acid structure. New York: Springer Verlag, 1984:257–286.



Fig. 12.9. Two DNA strands twist to form a double helix. The distance between the two phosphodiester backbones, shown with a ribbon, is about 11 Å. The hydrogen-bonded base pairs, shown in blue, create stacking forces with adjacent base pairs. Each phosphate group contains one negatively charged oxygen atom that provides the phosphodiester backbone with a negative charge. Because of the twisting of the helix, grooves are formed along the surface, the larger one being the major groove, and the smaller one the minor groove.

Heating and cooling cycles are used to separate and reanneal DNA strands in the polymerase chain reaction (PCR), a technique for obtaining large quantities of DNA from very small samples for research or for clinical and forensic testing (see Chapter 17).



Fig. 12.11. Effect of alkali on DNA and RNA. DNA strands stay intact, but they separate. RNA strands are degraded to nucleotides.



Fig. 12.12. Hybridization of DNA and complementary RNA.



If histones contain large amounts of arginine and lysine, will their net charge be positive or negative? separate, it does not break the phosphodiester bonds (Fig. 12.11). In contrast, the phosphodiester bonds of RNA are cleaved by alkali. Therefore, alkali is used to remove RNA from DNA and to separate DNA strands before, or after, electrophoresis on polyacrylamide or agarose gels.

Heat alone converts double-stranded DNA to single-stranded DNA. The separation of strands is called melting, and the temperature at which 50% of the DNA is separated is called the T_m . If the temperature is slowly decreased, complementary single strands can realign and base-pair, re-forming a double helix essentially identical to the original DNA. This process is known as renaturation, reannealing, or hybridization. The process by which a single-stranded DNA anneals with complementary strands of RNA is also called hybridization (Fig. 12.12). Hybridization is used extensively in research and clinical testing.

II. STRUCTURE OF CHROMOSOMES

A. Size of DNA Molecules

A prokaryotic cell generally contains a single chromosome composed of doublestranded DNA that forms a circle. These circular DNA molecules are extremely large. The entire chromosome of the bacterium *Escherichia coli*, composed of a single, circular double-stranded DNA molecule, contains over 4×10^6 base pairs. Its molecular weight is over $2,500 \times 10^6$ g/mol (compared to the molecular weight for a glucose molecule of 180 g/mol). If this molecule were linear, its length would measure almost 2 mm.

DNA from eukaryotic cells is approximately 1,000 times larger than that from bacterial cells. In eukaryotes, each chromosome contains one continuous, linear DNA helix. The DNA of the longest human chromosome is over 7 cm in length. In fact, if the DNA from all 46 chromosomes in a diploid human cell were placed end to end, our total DNA would span a distance of about 2 m (over 6 feet). Our total DNA contains about 6×10^9 base pairs.

B. Packing of DNA

DNA molecules require special packaging to enable them to reside within cells because the molecules are so large. In *E. coli*, the circular DNA is supercoiled and attached to an RNA-protein core. Packaging of eukaryotic DNA is much more complex because it is larger and must be contained within the nucleus of the eukaryotic cell. Eukaryotic DNA binds to an equal weight of histones, which are small basic proteins containing large amounts of arginine and lysine. The complex of DNA and proteins is called chromatin. The organization of eukaryotic DNA into chromatin is essential for controlling transcription, as well as for packaging. When chromatin is extracted from cells, it has the appearance of beads on a string (Fig. 12.13). The beads with DNA protruding from each end are known as nucleosomes, and the beads themselves are known as nucleosome cores (Fig. 12.14). Two molecules of each of four histone classes (histones H2A, H2B, H3, and H4) form the center of the core around which approximately 140 base pairs of double-stranded DNA are wound. The DNA wrapped around the nucleosome core is continuous and joins one

DNA consists of a double helix, with the two strands of DNA wrapping around each other to form a helical structure. To compact, the DNA molecule coils about itself to form a structure called a supercoil. A telephone cord, which connects the handpiece to the phone, displays supercoiling when the coiled cord wraps about itself. When the strands of a DNA molecule separate and unwind over a small local region (which occurs during DNA replication), supercoils are introduced into the remaining portion of the molecule, thereby increasing stress on this portion. Enzymes known as topoisomerases relieve this stress so that unwinding of the DNA strands can occur. nucleosome core to the next. The DNA joining the cores is complexed with the fifth type of histone, H1. Further compaction of chromatin occurs as the strings of nucleosomes wind into helical, tubular coils called solenoid structures.

Although complexes of DNA and histones form the nucleosomal substructures of chromatin, other types of proteins are also associated with DNA in the nucleus. These proteins were given the unimaginative name of "non-histone chromosomal proteins." The cells of different tissues contain different amounts and types of these proteins, which include enzymes that act on DNA and factors that regulate transcription.

C. The Human Genome

The genome, or total genetic content, of a human haploid cell (a sperm or an egg) is distributed in 23 chromosomes. Haploid cells contain one copy of each chromosome. The haploid egg and haploid sperm cells combine to form the diploid zygote, which continues to divide to form our other cells (mitosis), which are diploid. Diploid cells thus contain 22 pairs of autosomal chromosomes, with each pair composed of two homologous chromosomes containing a similar series of genes (Fig. 12.15). In addition to the autosomal chromosomes, each diploid cell has two sex chromosomes, designated X and Y. A female has two X chromosomes, and a male has one X and one Y chromosome. The total number of chromosomes per diploid cell is 46.

Genes are arranged linearly along each chromosome. A gene, in genetic terms, is the fundamental unit of heredity. In structural terms, a gene encompasses the DNA sequence encoding the structural components of the gene product (whether it be a polypeptide chain or RNA molecule) along with the DNA sequences adjacent to the 5' end of the gene which regulates its expression. A genetic locus is a specific position or location on a chromosome. Each gene on a chromosome in a diploid cell is matched by an alternate version of the gene at the same genetic locus on the homologous chromosome (Fig. 12.16). These alternate versions of a gene are called alleles. We thus have two alleles of each

At physiologic pH, arginine and lysine carry positive charges on their side chains; therefore, histones have a net positive charge. The arginine and lysine residues are clustered in regions of the histone molecules. These positively charged regions of the histones interact with the negatively charged DNA phosphate groups.



Fig. 12.13. Chromatin showing "beads on a string" structure.



Fig. 12.14. A polynucleosome, indicating the histone cores and linker DNA.



Fig. 12.16. Homologous chromosomes and their protein products. A set of homologous chromosomes is shown diagrammatically. (Of course, during interphase when they are producing their protein products, they cannot be visualized as discrete entities.) Four genes are shown as examples on each homologue. The genes of the homologues are alleles (e.g., AA, Bb, CC, dD). They may be identical (e.g., AA, CC), or they may differ (e.g., Bb, dD) in DNA sequence. Thus the corresponding protein products may be identical or they may differ in amino acid sequence.



Fig. 12.15. Human chromosomes from a male dipolid cell. Each diploid cell contains 22 pairs of autosomes (the numbered chromosomes 1-22) plus one X and one Y. Each female diploid cell contains two X chromosomes. Each haploid cell contains chromosomes 1 through 22 plus either an X or a Y. From Gelehrter TD, Collins FS, Ginsburg D. Principles of Human Genetics, 2nd Ed. Baltimore: Williams & Wilkins, 1998.

gene, one from our mother and one from our father. If the alleles are identical in base sequence, we are homozygous for this gene. If the alleles differ, we are heterozygous for this gene and may produce two versions of the encoded protein that differ somewhat in primary structure.

The genomes of prokaryotic and eukaryotic cells differ in size. The genome of the bacterium E. coli contains approximately 3,000 genes. All of this bacterial DNA has a function; it either codes for proteins, rRNA, and tRNA, or it serves to regulate the synthesis of these gene products. In contrast, the genome of the human haploid cell contains between 30,000 and 50,000 genes, 10 to 15 times the number in E. coli. The function of most of this extra DNA has not been determined (an issue considered in more detail in Chapter 15).

III. STRUCTURE OF RNA

A. General Features of RNA

RNA is similar to DNA. Like DNA, it is composed of nucleotides joined by 3'- to 5'-phosphodiester bonds, the purine bases adenine and guanine, and the pyrimidine base cytosine. However, its other pyrimidine base is uracil rather than thymine. Uracil and thymine are identical bases except that thymine has a methyl group at position 5 of the ring (Fig. 12.17). In RNA, the sugar is ribose, which contains a hydroxyl group on the 2'-carbon (see Fig 12.3. The prime refers to the position on the ribose ring). (The presence of this hydroxyl group allows RNA to be cleaved to its constituent nucleotides in alkaline solutions.)

RNA chains are usually single-stranded and lack the continuous helical structure of double-stranded DNA. However, RNA still has considerable secondary and tertiary structure because base pairs can form in regions where the strand loops back on itself. As in DNA, pairing between the bases is complementary and antiparallel. But in RNA, adenine pairs with uracil rather than thymine (Fig. 12.18). Basepairing in RNA can be extensive, and the irregular looped structures generated are



Will Sichel has sickle cell anemia (see Chapters 6 and 7). He has two alleles for the β -globin gene that both generate the mutated form of hemoglobin, HbS. His younger sister Amanda, a carrier for sickle cell trait, has one normal allele (that produces HbA) and one that pro-

duces HbS. A carrier would theoretically be expected to produce HbA:HbS in a 50:50 ratio. However, what is generally seen in electrophoresis is 60:40 ratio of HbA:HbS. Dramatic deviations from this ratio imply the occurrence of an additional hemoglobin mutation (e.g., thalassemia).



Fig. 12.17. Comparison of the structures of uracil and thymine. They differ in structure only by a methyl group, outlined in blue.



Fig. 12.18. A uracil-adenine base pair in RNA.

important for the binding of molecules, such as enzymes, that interact with specific regions of the RNA.

The three major types of RNA (mRNA, rRNA, and tRNA) participate directly in the process of protein synthesis. Other less abundant RNAs are involved in replication or in the processing of RNA, that is, in the conversion of RNA precursors to their mature forms.

Some RNA molecules are capable of catalyzing reactions. Thus, RNA, as well as protein, can have enzymatic activity. Certain rRNA precursors can remove internal segments of themselves, splicing the remaining fragments together. Because this RNA is changed by the reaction that it catalyzes, it is not truly an enzyme and therefore has been termed a "ribozyme." Other RNAs act as true catalysts, serving as ribonucleases that cleave other RNA molecules or as a peptidyl transferase, the enzyme in protein synthesis that catalyzes the formation of peptide bonds.

B. Structure of mRNA

Each **mRNA** molecule contains a nucleotide sequence that is converted into the amino acid sequence of a polypeptide chain in the process of translation. In eukaryotes, messenger RNA (mRNA) is transcribed from protein-coding genes as a long primary transcript that is processed in the nucleus to form mRNA. The various processing intermediates, which are mRNA precursors, are called pre-mRNA or hnRNA (heterogenous nuclear RNA). mRNA travels through nuclear pores to the cytoplasm, where it binds to ribosomes and tRNAs and directs the sequential insertion of the appropriate amino acids into a polypeptide chain.

Eukaryotic mRNA consists of a leader sequence at the 5' end, a coding region, and a trailer sequence at the 3' end (Fig 12.19). The leader sequence begins with a guanosine cap structure at its 5' end. The coding region begins with a trinucleotide start codon that signals the beginning of translation, followed by the trinucleotide codons for amino acids, and ends at a termination signal. The trailer terminates at its 5' end with a poly(A) tail that may be up to 200 nucleotides long. Most of the leader sequence, all of the coding region, and most of the trailer are formed by transcription of the complementary nucleotide sequence in DNA. However, the terminal guanosine in the cap structure and the poly(A) tail do not have complementary sequences; they are added posttranscriptionally.

C. Structure of rRNA

Ribosomes are subcellular ribonucleoprotein complexes on which protein synthesis occurs. Different types of ribosomes are found in prokaryotes and in the cytoplasm and mitochondria of eukaryotic cells (Fig. 12.20). Prokaryotic ribosomes contain three types of rRNA molecules with sedimentation coefficients of 16, 23, and 5S. The 30S ribosomal subunit contains the 16S rRNA complexed with proteins, and

Colin Tuma is being treated with 5fluorouracil (5-FU), a pyrimidine base similar to uracil and thymine. 5-FU inhibits the synthesis of the thymine nucleotides required for DNA replication. Thymine is normally produced by a reaction catalyzed by thymidylate synthase, an enzyme that converts deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP). 5-FU is converted in the body to F-dUMP, which binds tightly to thymidylate synthase in a transition state complex and inhibits the reaction (recall that thymine is 5-methyl uracil). Thus, thymine nucleotides cannot be generated for DNA synthesis, and the rate of cell proliferation decreases.



5–Fluorouracil, an analogue of uracil or thymine





Fig. 12.19. The regions of eukaryotic mRNA. The wavy line indicates the polynucleotide chain of the mRNA and the As constituting the poly(A) tail. The 5'-cap consists of a guanosine residue linked at its 5' hydroxyl group to three phosphates, which are linked to the 5'-hydroxyl group of the next nucleotide in the RNA chain. The start and stop codons represent where protein synthesis is initiated and terminated from this mRNA.





Fig. 12.20. Comparison of prokaryotic and eukaryotic ribosomes. The cytoplasmic ribosomes of eukaryotes are shown. Mitochondrial ribosomes are similar to prokaryotic ribosomes, but they are smaller (55S rather than 70S).



Erythromycin, the antibiotic used to treat **Neu Moania**, inhibits protein synthesis on prokaryotic ribo-

somes, but not on eukaryotic ribosomes. It binds to the 50S ribosomal subunit, which is absent in eukaryotes. Therefore, it will selectively inhibit bacterial growth. However, because mitochondrial ribosomes are similar to those of bacteria, mitochondrial protein synthesis can also be inhibited. This fact is important in understanding some of the side effects of antibiotics that work by inhibiting bacterial protein synthesis. A sedimentation coefficient is a measure of the rate of sedimentation of a macromolecule in a high-speed centrifuge (an ultracentrifuge). It is expressed in Svedberg units (S). Although larger macromolecules generally have higher sedimentation coefficients than do smaller macromolecules, sedimentation coefficients are not additive. Because frictional forces acting on the surface of a macromolecule slow its migration through the solvent, the rate of sedimentation depends not only on the density of the macromolecule, but also on its shape.

the 50S ribosomal subunit contains the 23S and 5S rRNAs complexed with proteins. The 30S and 50S ribosomal subunits join to form the 70S ribosome, which participates in protein synthesis.

Cytoplasmic ribosomes in eukaryotes contain four types of rRNA molecules of 18, 28, 5, and 5.8S. The 40S ribosomal subunit contains the 18S rRNA complexed with proteins, and the 60S ribosomal subunit contains the 28, 5, and 5.8S rRNAs complexed with proteins. In the cytoplasm, the 40S and 60S ribosomal subunits combine to form the 80S ribosomes that participate in protein synthesis.

Mitochondrial ribosomes, with a sedimentation coefficient of 55S, are smaller than cytoplasmic ribosomes. Their properties are similar to those of the 70S ribosomes of bacteria.

rRNAs contain many loops and exhibit extensive base-pairing in the regions between the loops (Fig. 12.21). The sequences of the rRNAs of the smaller ribosomal subunits exhibit secondary structures that are common to many different genera.

D. Structure of tRNA

During protein synthesis, tRNA molecules carry amino acids to ribosomes and ensure that they are incorporated into the appropriate positions in the growing polypeptide chain (Fig. 12.22). This is done through base-pairing of three bases of the tRNA (the anticodon) with the three base codons within the coding region of the mRNA. Therefore, cells contain at least 20 different tRNA molecules that differ somewhat in nucleotide sequence, one for each of the amino acids found in proteins. Many amino acids have more than one tRNA.

tRNA molecules contain not only the usual nucleotides, but also derivatives of these nucleotides that are produced by posttranscriptional modifications. In eukaryotic cells, 10 to 20% of the nucleotides of tRNA are modified. Most tRNA molecules contain ribothymidine (T), in which a methyl group is added to uridine to form ribothymidine. They also contain dihydrouridine (D), in which one of the double bonds of the base is reduced; and pseudouridine (Ψ), in which uracil is attached to ribose by a carbon–carbon bond rather than a nitrogen–carbon bond (see Chapter 14). The base at the 5'-end of the anticodon of tRNA is frequently modified.

tRNA molecules are rather small compared with both mRNA and the large rRNA molecules. On average, tRNA molecules contain approximately 80 nucleotides and have a sedimentation coefficient of 4S. Because of their small size and high content of modified nucleotides, tRNAs were the first nucleic acids to be sequenced. Since 1965 when Robert Holley deduced the structure of the first tRNA, the nucleotide sequences of many different tRNAs have been determined. Although their primary sequences differ, all tRNA molecules can form a structure resembling a cloverleaf (discussed in more detail in Chapter 14).

E. Other Types of RNA

In addition to the three major types of RNA described above, other RNAs are present in cells. These RNAs include the oligonucleotides that serve as primers for DNA replication and the RNAs in the small nuclear ribonucleoproteins (snRNPs or snurps) that are involved in the splicing and modification reactions that occur during the maturation of RNA precursors (see Chapter 14).

CLINICAL COMMENTS

Ivy Sharer. Ivy Sharer's clinical course was typical for the development of full-blown AIDS, in this case caused by the use of needles contaminated with HIV. The progressive immunologic deterioration that accompanies this disease ultimately results in life-threatening opportunistic infections with fungi (e.g., *Candida*, cryptococcus), other viruses (e.g., cytomegalovirus, herpes simplex), and bacteria (e.g., *Mycobacterium, Pneumocystis carinii, Salmonella*). The immunologic incompetence also frequently results in the development of certain neoplasms (e.g., Kaposi's sarcoma, non-Hodgkin's lymphoma) as well as meningitis, neuropathies, and neuropsychiatric disorders causing cognitive dysfunction. Although recent advances in drug therapy can slow the course of the disease, no cure is yet available.

Colin Tuma. Colin Tuma's original benign adenomatous polyp was located in the ascending colon, where 10% of large bowel cancers eventually arise. Because Mr. Tuma's father died of a cancer of the colon, his physician had warned him that his risk for developing colon cancer was three times higher than for the general population. Unfortunately, Mr. Tuma neglected to have his annual colonoscopic examinations as prescribed, and he developed an adenocarcinoma that metastasized.

The most malignant characteristic of neoplasms is their ability to metastasize, that is, form a new neoplasm at a noncontiguous site. The initial site of metastases for a tumor is usually at the first capillary bed encountered by the malignant cells once they are released. Thus, cells from tumors of the gastrointestinal tract often pass through the portal vein to the liver, which is Colin Tuma's site of metastasis. Because his adenocarcinoma has metastasized, there is little hope of eradicating it, and his therapy with 5-FU is palliative (directed toward reducing the severity of the disease and alleviation of the symptoms without actually curing the disease.)

Agneu Moania. Neu Moania's infection was treated with erythromycin, a macrolide antibiotic. Because this agent can inhibit mitochondrial protein synthesis in eukaryotic cells, it has the potential to alter host cell function, leading to such side effects as epigastric distress, diarrhea, and, infrequently, cholestatic jaundice.

BIOCHEMICAL COMMENTS

Retroviruses. RNA also serves as the genome for certain types of viruses, including retroviruses (e.g., the human immunodeficiency virus [HIV] that causes AIDS). Viruses must invade host cells to reproduce. They are not capable of reproducing independently. Some viruses that are pathogenic to



Fig. 12.21. Secondary structure of the portion of 16S-type ribosomal RNA that is common to many species. Darkened areas are base-paired. Circles are unpaired loops. Reproduced with permission, from Annu Rev Biochem. 1984;53:137. © 1984, by Annual Reviews, Inc.



Fig 12.22. The typical cloverleaf structure of tRNA.



Colin Tuma completed his first course of intravenous 5-fluorouracil (5-FU) in the hospital. He tolerated the therapy with only mild anorexia and diarrhea and with only a mild leukopenia (a decreased white blood cell count; leuko = white). Thirty days after the completion of the initial course, these symptoms abated and he started his second course of chemotherapy with 5-FU as an out-

Because 5-FU inhibits synthesis of thymine, DNA synthesis is affected in all cells in the human body that are rapidly dividing, such as the cells in the bone marrow that produce leukocytes and the mucosal cells lining the intestines. Inhibition of DNA synthesis in rapidly dividing cells contributes to the side effects of 5-FU and many other chemotherapeutic drugs.



Fig. 12.23. The life cycle of a retrovirus. The virus contains two identical RNA strands, only one of which is shown for clarity. After penetrating the plasma membrane, the single-stranded viral RNA genome is reverse-transcribed to a double-stranded DNA form. The viral DNA migrates to the nucleus and integrates into the chromosomal DNA, where it is transcribed to form a viral RNA transcript. The viral transcript can form the viral RNA genome for progeny viruses, or can be translated to generate viral structural proteins.

humans contain DNA as their genetic material. Others contain RNA as their genetic material.

Some viruses that contain an RNA genome are known as retroviruses. HIV, the human immunodeficiency virus, is the retrovirus that causes AIDS (Fig 12.23). It invades cells of the immune system and prevents the affected individual from mounting an adequate immune response to combat infections.

According to the "central dogma" proposed by Francis Crick, information flows from DNA to RNA to proteins. For the most part, this concept holds true. However, retroviruses provide one violation of this rule. When retroviruses invade cells, their RNA genome is transcribed to produce a DNA copy. The enzyme that catalyzes this process is encoded in the viral RNA and is known as reverse transcriptase. This DNA copy integrates into the genome of the infected cell, and enzymes of the host cell are used to produce many copies of the viral RNA, as well as viral proteins, which can be packaged into new viral particles.

Suggested Readings

Watson JD. The Double Helix. New York: Atheneum, 1968.

- Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature 1953;171:737–738.
- American Cancer Society Web site (http://www.cancer.org): Colorectal cancer. This web site provides information on the cause, treatment and prevention of a number of different cancers, and links to additional resources.
- The Human Genome Project Web site (http://www.nhgri.nih.gov/About_NHGRI/Der/Elsi/): This site discusses the ethical issues involved in gene therapy protocols.

REVIEW QUESTIONS-CHAPTER 12

- 1. For the following DNA sequence, determine the sequence and direction of the complementary strand. 5' ATCGATCGATCGATCG 3'
 - (A) 5' ATCTATCGATCGATCG 3'
 - (B) 3' ATCGATCGATCGATCG 5'
 - $(C) \ 5' CGAUCGAUCAUCGAU 3'$
 - (D) 5' CGATCGATCGATCGAT 3'
 - (E) 3' CGATCGATCGATCGAT 5'
- 2. If the DNA strand shown below serves as a template for the synthesis of RNA, which of the following choices gives the sequence and direction of the RNA?

5' - GCTATGCATCGTGATCGAATTGCGT - 3'

- (A) 5' ACGCAATTCGATCACGATGCATAGC 3'
- (B) 5' UGCGUUAAGCUAGUGCUACGUAUCG 3'
- (C) 5' ACGCAAUUCGAUCACGAUGCAUAGC 3'
- (D) 5' CGAUACGUAGCACUAGCUUAACGCA 3'
- (E) 5' GCTATGCATCGTGATCGAATTGCGT 3'
- 3. In DNA, the bond between the deoxyribose sugar and the phosphate is which of the following?
 - (A) A polar bond
 - (B) An ionic bond
 - (C) A hydrogen bond
 - (D) A covalent bond
 - (E) A van der Waals bond
- 4. How many double-stranded DNA molecules 8 base pairs long are theoretically possible?
 - (A) 12
 - (B) 32
 - (C) 64
 - (D) 256
 - (E) 65,536
- 5. The backbone of a DNA strand is composed of which of the following?
 - (A) Sugars and bases
 - (B) Phosphates and sugars
 - (C) Bases and phosphates
 - (D) Nucleotides and sugars
 - (E) Phosphates and nucleosides