CHAPTER **FIVE** **5**

DNA and Chromosomes

Life depends on the ability of cells to store, retrieve, and translate the genetic instructions required to make and maintain a living organism. These instructions are stored within every living cell in its *genes*—the information-bearing elements that determine the characteristics of a species as a whole and of the individuals within it.

At the beginning of the twentieth century, when genetics emerged as a science, scientists became intrigued by the chemical nature of genes. The information in genes is copied and transmitted from a cell to its daughter cells millions of times during the life of a multicellular organism, and passed from generation to generation through the reproductive cells—eggs and sperm. Genes survive this process of replication and transmission essentially unchanged. What kind of molecule could be capable of such accurate and almost unlimited replication, and also be able to direct the development of an organism and the daily life of a cell? What kind of instructions does the genetic information contain? How are these instructions physically organized so that the enormous amount of information required for the development and maintenance of even the simplest organism can be contained within the tiny space of a cell?

The answers to some of these questions began to emerge in the 1940s, when it was discovered from studies in simple fungi that genetic information consists primarily of instructions for making proteins. As described in the previous chapter, proteins perform most of the cell's functions: they serve as building blocks for cell structures; they form the enzymes that catalyze the cell's chemical reactions; they regulate the activity of genes; and they enable cells to move and to communicate with one another. With hindsight, it is hard to imagine what other type of instructions the genetic information could have contained.

The other crucial advance made in the 1940s was the recognition that deoxyribonucleic acid (DNA) is the carrier of the cell's genetic information. But the mechanism whereby the information could be copied for transmission from one generation of cells to the next, and how proteins might be specified by instructions in DNA, remained completely mysterious until 1953, when the structure of DNA was determined by James Watson and Francis Crick. The structure immediately revealed how DNA might be copied, or replicated, and it provided the first clues about how a molecule of DNA might encode the instructions for making proteins. Today, the fact that DNA is the genetic material is so fundamental to our understanding of life that it can be difficult to appreciate what an enormous intellectual gap this discovery filled.

In this chapter, we begin by describing the structure of DNA. We see how, despite its chemical simplicity, the structure and chemical properties of DNA make it ideally suited for carrying genetic information. We then consider how genes and other important segments of DNA are arranged in the single, long DNA molecule that forms each chromosome in the cell. Finally, we discuss how eukaryotic cells fold these long DNA molecules into compact chromosomes inside the nucleus. This packing has to be done in an orderly fashion so that the chromosomes can be apportioned correctly between the two daughter cells at each cell division. At the same time, chromosomal packaging must allow DNA to be accessed by the large number of proteins that replicate and repair it, and that determine the activity of the cell's many genes.

This is the first of five chapters that deal with basic genetic mechanisms—the ways in which the cell maintains and makes use of the genetic information carried in its DNA. In Chapter 6, we discuss the mechanisms by which the cell accurately replicates and repairs its DNA. In Chapter 7, we consider gene expression—how genes are used to produce RNA and protein molecules. In Chapter 8, we describe how a cell controls gene expression to ensure that each of the many thousands of proteins encoded in its DNA is manufactured at the proper time and place. In Chapter 9, we discuss how present-day genes evolved, and, in Chapter 10, we consider some of the ways that DNA can be experimentally manipulated to study fundamental cell processes.

An enormous amount has been learned about these subjects in the past 60 years. Much less obvious, but equally important, is the fact that our knowledge is very incomplete; thus a great deal still remains to be discovered about how DNA provides the instructions to build living things.

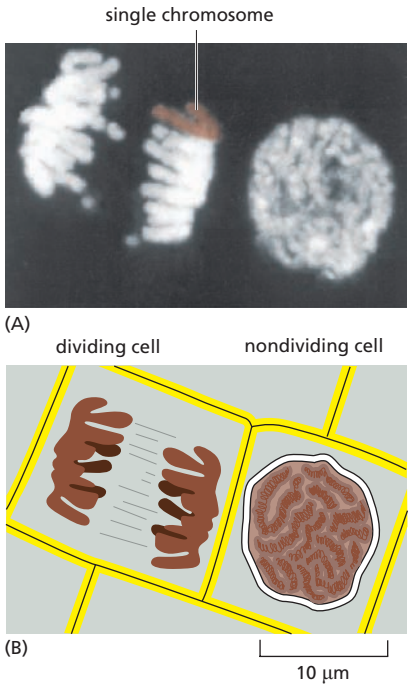


Figure 5–1 Chromosomes become visible as eukaryotic cells prepare to divide. (A) Two adjacent plant cells photographed using a fluorescence microscope. The DNA, which is labeled with a fluorescent dye (DAPI), is packaged into multiple chromosomes; these become visible as distinct structures only when they condense in preparation for cell division, as can be seen in the cell on the *left*. For clarity, a single chromosome has been shaded (*brown*) in the dividing cell. The cell on the *right*, which is not dividing, contains the identical chromosomes, but they cannot be distinguished as individual entities because the DNA is in a much more extended conformation at this phase in the cell's division cycle. (B) Schematic diagram of the outlines of the two cells and their chromosomes. (A, courtesy of Peter Shaw.)

THE STRUCTURE OF DNA

Long before biologists understood the structure of DNA, they had recognized that inherited traits and the genes that determine them were associated with chromosomes. Chromosomes (named from the Greek *chroma*, "color," because of their staining properties) were discovered in the nineteenth century as threadlike structures in the nucleus of eukaryotic cells that become visible as the cells begin to divide (**Figure 5–1**). As biochemical analyses became possible, researchers learned that chromosomes contain both DNA and protein. But which of these components encoded the organism's genetic information was not immediately clear.

We now know that the DNA carries the genetic information of the cell and that the protein components of chromosomes function largely to package and control the enormously long DNA molecules. But biologists in the 1940s had difficulty accepting DNA as the genetic material because of the apparent simplicity of its chemistry (see **How We Know**, pp. 193–195).

DNA, after all, is simply a long polymer composed of only four types of nucleotide subunits, which are chemically very similar to one another.

Then, early in the 1950s, Maurice Wilkins and Rosalind Franklin examined DNA using x-ray diffraction analysis, a technique for determining the three-dimensional atomic structure of a molecule (see Panel 4–6, pp. 168–169). Their results provided one of the crucial pieces of evidence that led, in 1953, to Watson and Crick's model of the double-helical structure of DNA. This structure—in which two strands of DNA are wound around each other to form a helix—immediately suggested how DNA could encode the instructions necessary for life, and how these instructions could be copied and passed along when cells divide. In this section, we examine the structure of DNA and explain in general terms how it is able to store hereditary information.

A DNA Molecule Consists of Two Complementary Chains of Nucleotides

A molecule of **deoxyribonucleic acid** (**DNA**) consists of two long polynucleotide chains. Each *chain*, or *strand*, is composed of four types of nucleotide subunits, and the two strands are held together by hydrogen bonds between the base portions of the nucleotides (**Figure 5–2**).

As we saw in Chapter 2 (Panel 2–7, pp. 78–79), nucleotides are composed of a nitrogen-containing base and a five-carbon sugar, to which a phosphate group is attached. For the nucleotides in DNA, the sugar is deoxyribose (hence the name deoxyribonucleic acid) and the base can be either *adenine* (*A*), *cytosine* (*C*), *guanine* (*G*), or *thymine* (*T*). The

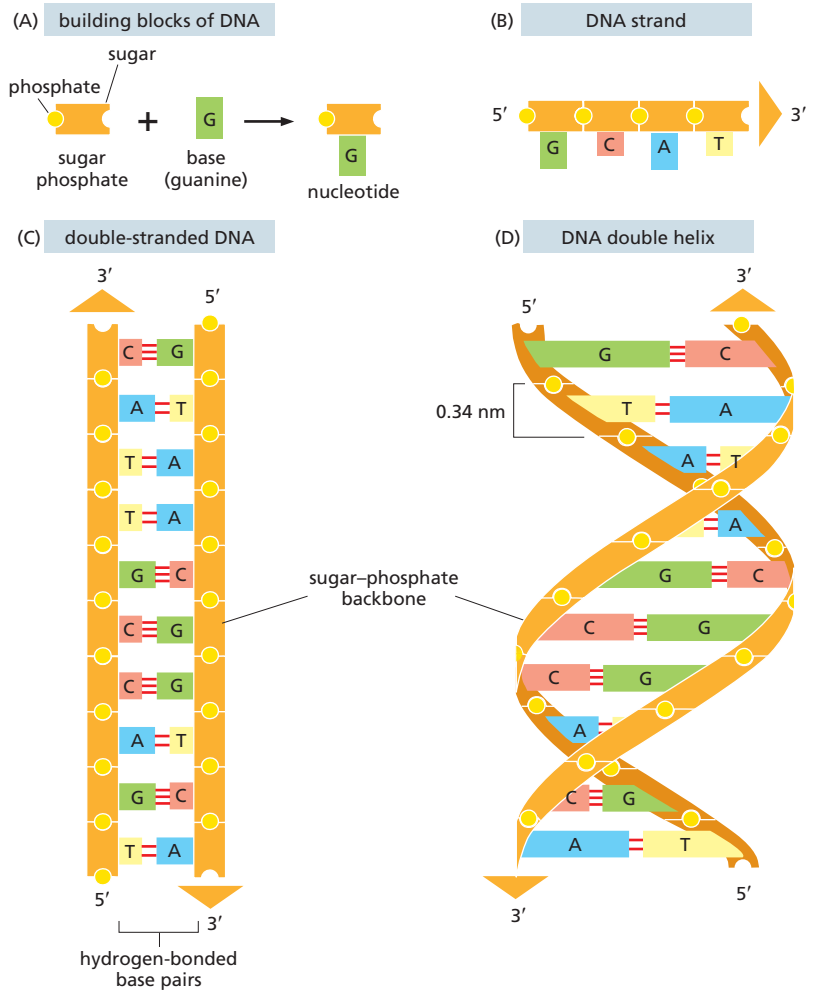


Figure 5–2 DNA is made of four nucleotide building blocks. (A) Each nucleotide is composed of a sugar phosphate covalently linked to a base—guanine (G) in this figure. (B) The nucleotides are covalently linked together into polynucleotide chains, with a sugar–phosphate backbone from which the bases—adenine, cytosine, guanine, and thymine (A, C, G, and T)—extend. (C) A DNA molecule is composed of two polynucleotide chains (DNA strands) held together by hydrogen bonds between the paired bases. The *arrows* on the DNA strands indicate the polarities of the two strands, which run *antiparallel* to each other (with opposite chemical polarities) in the DNA molecule. (D) Although the DNA is shown straightened out in (C), in reality, it is wound into a double helix, as shown here.

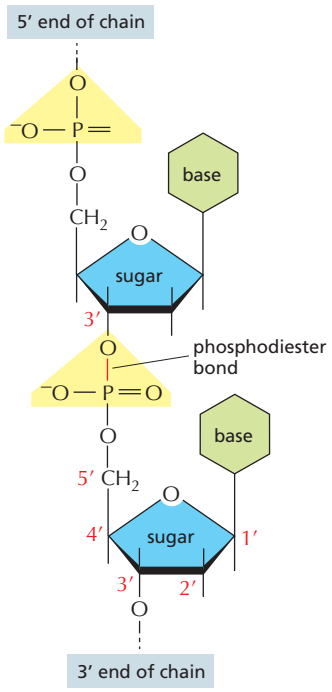


Figure 5–3 The nucleotide subunits within a DNA strand are held together by phosphodiester bonds. These bonds connect one sugar to the next. The chemical differences in the ester linkages—between the 5′ carbon of one sugar and the 3′ carbon of the other—give rise to the polarity of the resulting DNA strand. For simplicity, only two nucleotides are shown here.

nucleotides are covalently linked together in a chain through the sugars and phosphates, which form a backbone of alternating sugar–phosphate–sugar–phosphate (see Figure 5–2B). Because only the base differs in each of the four types of subunits, each polynucleotide chain resembles a necklace: a sugar–phosphate backbone strung with four types of tiny beads (the four bases A, C, G, and T). These same symbols (A, C, G, and T) are also commonly used to denote the four different nucleotides—that is, the bases with their attached sugar phosphates.

The nucleotide subunits within a DNA strand are held together by phosphodiester bonds that link the 5′ end of one sugar with the 3′ end of the next (**Figure 5–3**). Because the ester linkages to the sugar molecules on either side of the bond are different, each DNA strand has a chemical polarity. If we imagine that each nucleotide has a phosphate "knob" and a hydroxyl "hole" (see Figure 5–2A), each strand, formed by interlocking knobs with holes, will have all of its subunits lined up in the same orientation. Moreover, the two ends of the strand can be easily distinguished, as one will have a hole (the 3′ hydroxyl) and the other a knob (the 5′ phosphate). This polarity in a DNA strand is indicated by referring to one end as the *3′ end* and the other as the *5′ end* (see Figure 5–3).

The two polynucleotide chains in the DNA **double helix** are held together by hydrogen-bonding between the bases on the different strands. All the bases are therefore on the inside of the double helix, with the sugar–phosphate backbones on the outside (see Figure 5–2D). The bases do not pair at random, however; A always pairs with T, and G always pairs with C (**Figure 5–4**). In each case, a bulkier two-ring base (a purine, see Panel 2–7, pp. 78–79) is paired with a single-ring base (a pyrimidine). Each purine–pyrimidine pair is called a **base pair**, and this *complementary base-pairing* enables the base pairs to be packed in the energetically most favorable arrangement along the interior of the double helix. In this arrangement, each base pair has the same width, thus holding the sugar–phosphate backbones an equal distance apart along the DNA molecule. For the members of each base pair to fit together within the double helix, the two strands of the helix must run *antiparallel* to each other—that is, be oriented with opposite polarities (see Figure 5–2C and D). The antiparallel sugar–phosphate strands then twist around each other to form a double helix containing 10 base pairs per helical turn (**Figure 5–5**). This twisting also contributes to the energetically favorable conformation of the DNA double helix.

As a consequence of the base-pairing arrangement shown in Figure 5–4, each strand of a DNA double helix contains a sequence of nucleotides that is exactly **complementary** to the nucleotide sequence of its partner strand—an A always matches a T on the opposite strand, and a C always matches a G. This complementarity is of crucial importance when it comes to both copying and maintaining the DNA structure, as we discuss in Chapter 6. An animated version of the DNA double helix can be seen in **Movie 5.1**.

The Structure of DNA Provides a Mechanism for Heredity

The fact that genes encode information that must be copied and transmitted accurately when a cell divides raised two fundamental issues: how

QUESTION 5–1

Which of the following statements are correct? Explain your answers.
A. A DNA strand has a polarity because its two ends contain different bases.
B. G-C base pairs are more stable than A-T base pairs.

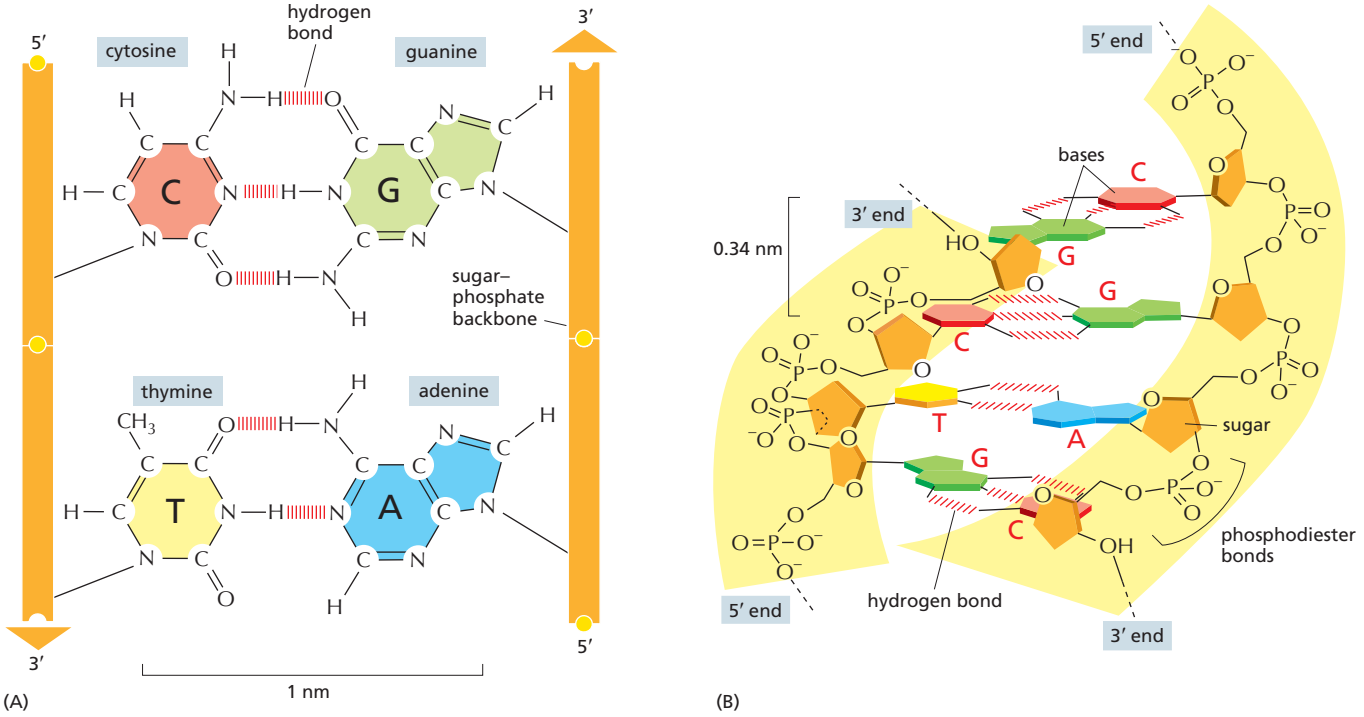


Figure 5–4 The two strands of the DNA double helix are held together by hydrogen bonds between complementary base pairs. (A) Schematic illustration showing how the shapes and chemical structures of the bases allow hydrogen bonds to form efficiently only between A and T and between G and C. The atoms that form the hydrogen bonds between these nucleotides (see Panel 2–3, pp. 70–71) can be brought close together without perturbing the double helix. As shown, two hydrogen bonds form between A and T, whereas three form between G and C. The bases can pair in this way only if the two polynucleotide chains that contain them are antiparallel—that is, oriented in opposite directions. (B) A short section of the double helix viewed from its side. Four base pairs are illustrated; note that they lie perpendicular to the axis of the helix, unlike the schematic shown in (A). As shown in Figure 5–3, the nucleotides are linked together covalently by phosphodiester bonds that connect the 3′-hydroxyl (–OH) group of one sugar and the 5′ phosphate ($–PO_3$) attached to the next (see Panel 2–7, pp. 78–79, to review how the carbon atoms in the sugar ring are numbered). This linkage gives each polynucleotide strand a chemical polarity; that is, its two ends are chemically different. The 3′ end carries an unlinked –OH group attached to the 3′ position on the sugar ring; the 5′ end carries a free phosphate group attached to the 5′ position on the sugar ring.

can the information for specifying an organism be carried in chemical form, and how can the information be accurately copied? The structure of DNA provides the answer to both questions.

Information is encoded in the order, or sequence, of the nucleotides along each DNA strand. Each base—A, C, T, or G—can be considered a letter in a four-letter alphabet that is used to spell out biological messages (**Figure 5–6**). Organisms differ from one another because their respective DNA molecules have different *nucleotide sequences* and, consequently, carry different biological messages. But how is the nucleotide alphabet used to make up messages, and what do they spell out?

Before the structure of DNA was determined, investigators had established that genes contain the instructions for producing proteins. Thus, it was clear that DNA messages must somehow be able to encode proteins. Consideration of the chemical character of proteins makes the problem

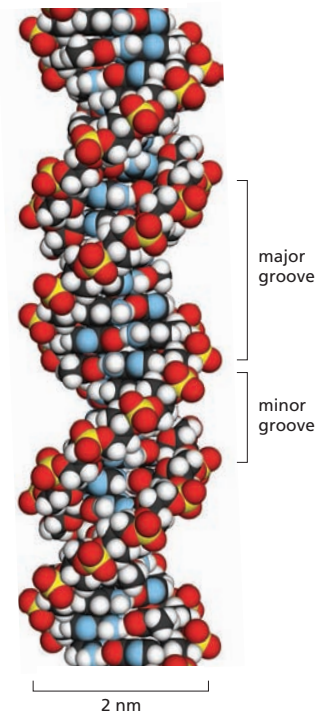


Figure 5–5 A space-filling model shows the conformation of the DNA double helix. The two DNA strands wind around each other to form a right-handed helix (see Figure 4–14) with 10 bases per turn. Shown here are 1.5 turns of the DNA double helix. The coiling of the two strands around each other creates two grooves in the double helix. The wider groove is called the major groove and the smaller one the minor groove. The colors of the atoms are: N, *blue*; O, *red*; P, *yellow*; H, *white*; and C, *black*. (See **Movie 5.1**.)

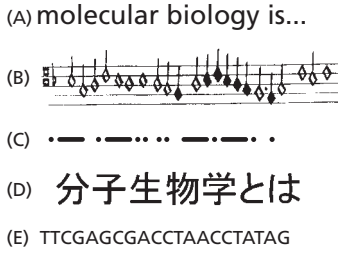


Figure 5–6 Linear messages come in many forms. The languages shown are (A) English, (B) a musical score, (C) Morse code, (D) Japanese, and (E) DNA.

easier to define. As discussed in Chapter 4, the function of a protein is determined by its three-dimensional structure, which in turn is determined by the sequence of the amino acids in its polypeptide chain. The linear sequence of nucleotides in a gene, therefore, must somehow spell out the linear sequence of amino acids in a protein.

The exact correspondence between the 4-letter nucleotide alphabet of DNA and the 20-letter amino acid alphabet of proteins—the **genetic code**—is not at all obvious from the structure of the DNA molecule. It took more than a decade of clever experiments after the discovery of the double helix to work this code out. In Chapter 7, we describe the genetic code in detail when we discuss **gene expression**—the process by which the nucleotide sequence of a gene is *transcribed* into the nucleotide sequence of an RNA molecule—and then, in most cases, *translated* into the amino acid sequence of a protein (**Figure 5–7**).

The amount of information in an organism's DNA is staggering: written out in the four-letter nucleotide alphabet, the nucleotide sequence of a very small protein-coding gene from humans occupies a quarter of a page of text, while the complete human DNA sequence would fill more than 1000 books the size of this one. Herein lies a problem that affects the architecture of all eukaryotic chromosomes: How can all this information be packed neatly into the cell nucleus? In the remainder of this chapter, we discuss the answer to this question.

THE STRUCTURE OF EUKARYOTIC CHROMOSOMES

Large amounts of DNA are required to encode all the information needed to make a single-celled bacterium, and far more DNA is needed to encode the information to make a multicellular organism like you. Each human cell contains about 2 meters (m) of DNA; yet the cell nucleus is only 5–8 μm in diameter. Tucking all this material into such a small space is the equivalent of trying to fold 40 km (24 miles) of extremely fine thread into a tennis ball.

In eukaryotic cells, very long, double-stranded DNA molecules are packaged into **chromosomes**. These chromosomes not only fit handily inside the nucleus, but, after they are duplicated, they can be accurately apportioned between the two daughter cells at each cell division. The complex task of packaging DNA is accomplished by specialized proteins that bind to and fold the DNA, generating a series of coils and loops that provide increasingly higher levels of organization and prevent the DNA from becoming a tangled, unmanageable mess. Amazingly, this DNA is folded in a way that allows it to remain accessible to all of the enzymes and other proteins that replicate and repair it, and that cause the expression of its genes.

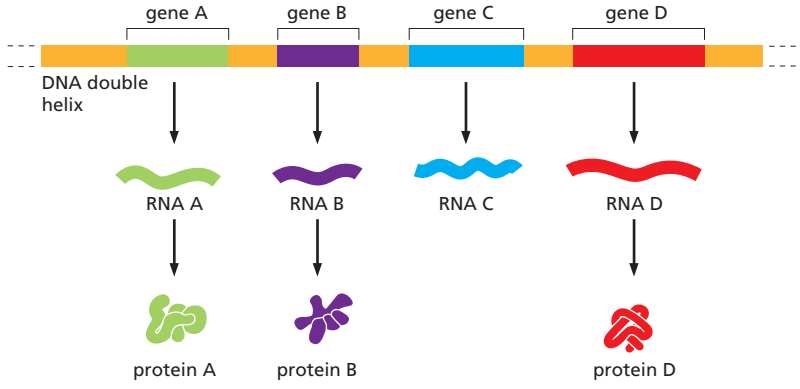


Figure 5–7 Most genes contain information to make proteins. As we discuss in Chapter 7, protein-coding genes each produce a set of RNA molecules, which then direct the production of a specific protein molecule. Note that for a minority of genes, the final product is the RNA molecule itself, as shown here for gene C. In these cases, gene expression is complete once the nucleotide sequence of the DNA has been transcribed into the nucleotide sequence of its RNA.

Bacteria typically carry their genes on a single, circular DNA molecule. This molecule is also associated with proteins that condense the DNA, but these bacterial proteins differ from the ones that package eukaryotic DNA. Although this prokaryotic DNA is called a bacterial "chromosome," it does not have the same structure as eukaryotic chromosomes, and less is known about how it is packaged. Our discussion of chromosome structure in this chapter will therefore focus entirely on eukaryotic chromosomes.

Eukaryotic DNA Is Packaged into Multiple Chromosomes

In eukaryotes, such as ourselves, nuclear DNA is distributed among a set of different chromosomes. The DNA in a human nucleus, for example, is parceled out into 23 or 24 different types of chromosome, depending on an individual's sex (males, with their *Y chromosome*, have an extra type of chromosome that females do not). Each of these chromosomes consists of a single, enormously long, linear DNA molecule associated with proteins that fold and pack the fine thread of DNA into a more compact structure. This complex of DNA and protein is called *chromatin*. In addition to the proteins involved in packaging the DNA, chromosomes also associate with many other proteins involved in DNA replication, DNA repair, and gene expression.

With the exception of the gametes (sperm and eggs) and highly specialized cells that lack DNA entirely (such as mature red blood cells), human cells each contain two copies of every chromosome, one inherited from the mother and one from the father. The maternal and paternal versions of each chromosome are called *homologous chromosomes* (*homologs*). The only nonhomologous chromosome pairs in humans are the sex chromosomes in males, where a Y chromosome is inherited from the father and an *X chromosome* from the mother. (Females inherit one X chromosome from each parent and have no Y chromosome.) Each full set of human chromosomes contains a total of approximately 3.2×10^9 nucleotide pairs of DNA—which together comprise the *human genome*.

In addition to being different sizes, the different human chromosomes can be distinguished from one another by a variety of techniques. Each chromosome can be "painted" a different color using sets of chromosome-specific DNA molecules coupled to different fluorescent dyes (**Figure 5–8A**). An earlier and more traditional way of distinguishing one chromosome from another involves staining the chromosomes with dyes that bind to certain types of DNA sequences. These dyes mainly distinguish between DNA that is rich in A-T nucleotide pairs and DNA that is G-C rich, and they produce a predictable pattern of bands along each type of chromosome. The resulting patterns allow each chromosome to be identified and numbered.

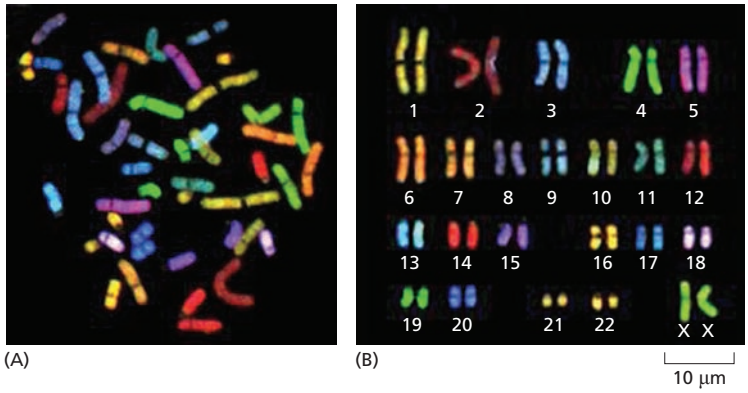


Figure 5–8 Each human chromosome can be "painted" a different color to allow its unambiguous identification. The chromosomes shown here were isolated from a cell undergoing nuclear division (mitosis) and are therefore in a highly compact (condensed) state. Chromosome painting is carried out by exposing the chromosomes to a collection of single-stranded DNA molecules that have been coupled to a combination of fluorescent dyes. For example, single-stranded DNA molecules that match sequences in chromosome 1 are labeled with one specific dye combination, those that match sequences in chromosome 2 with another, and so on. Because the labeled DNA can form base pairs (hybridize) only with its specific chromosome (discussed in Chapter 10), each chromosome is differently colored. For such experiments, the chromosomes are treated so that the individual strands of its double-helical DNA partly separate to enable base-pairing with the labeled, single-stranded DNA.

(A) Micrograph showing the array of chromosomes as they originally spilled from the lysed cell. (B) The same chromosomes artificially lined up in their numerical order. This arrangement of the full chromosome set is called a karyotype. (Adapted from N. McNeil and T. Ried, *Expert Rev. Mol. Med.* 2:1–14, 2000. With permission from Cambridge University Press.)

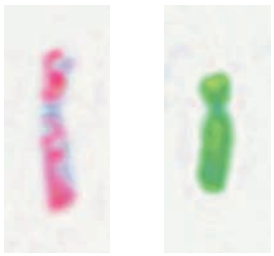


Figure 5–9 Abnormal chromosomes are associated with some inherited genetic disorders. (A) Two normal human chromosomes, chromosome 6 and chromosome 4, have been subjected to chromosome painting as described in Figure 5–8. (B) In an individual with a reciprocal chromosomal translocation, a segment of one chromosome has been swapped with a segment from the other. Such chromosomal translocations are a frequent event in cancer cells. (Courtesy of Zhenya Tang and the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research.)

An ordered display of the full set of 46 human chromosomes is called the human **karyotype** (**Figure 5–8B**). If parts of a chromosome are lost, or switched between chromosomes, these changes can be detected. Cytogeneticists analyze karyotypes to detect chromosomal abnormalities that are associated with some inherited disorders (**Figure 5–9**) and with certain types of cancer (as we see in Chapter 20).

Chromosomes Organize and Carry Genetic Information

The most important function of chromosomes is to carry genes—the functional units of heredity. A **gene** is often defined as a segment of DNA that contains the instructions for making a particular protein or RNA molecule. Most of the RNA molecules encoded by genes are subsequently used to produce a protein. In some cases, however, the RNA molecule is the final product (see Figure 5–7). Like proteins, these RNA molecules have diverse functions in the cell, including structural, catalytic, and gene regulatory roles, as we discuss in later chapters.

Together, the total genetic information carried by a complete set of the chromosomes present in a cell or organism constitutes its **genome**. Complete genome sequences have been determined for thousands of organisms, from *E. coli* to humans. As might be expected, some correlation exists between the complexity of an organism and the number of genes in its genome. For example, the total number of genes is about 500 for the simplest bacterium and about 24,000 for humans. Bacteria and some single-celled eukaryotes, including the budding yeast *S. cerevisiae*, have especially compact genomes: the DNA molecules that make up their chromosomes are little more than strings of closely packed genes (**Figure 5–10**). However, chromosomes from many eukaryotes—including humans—contain, in addition to genes and the specific nucleotide sequences required for normal gene expression, a large excess of interspersed DNA (**Figure 5–11**). This extra DNA is sometimes erroneously called "junk DNA," because its usefulness to the cell has not yet been demonstrated. Although this spare DNA does not code for protein, much of it may serve some other biological function. Comparisons of the genome sequences from many different species reveal that small portions of this extra DNA are highly conserved among related species, suggesting their importance for these organisms.

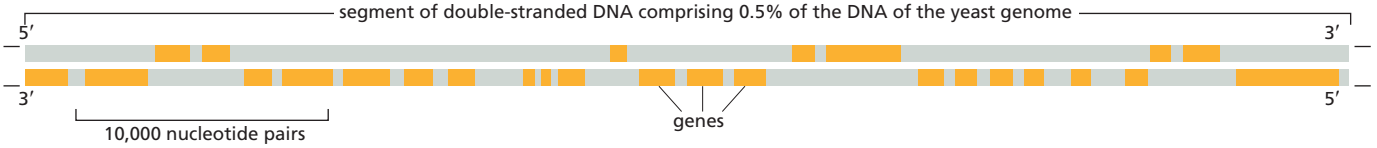


Figure 5–10 In yeast, genes are closely packed along chromosomes. This figure shows a small region of the DNA double helix in one chromosome from the budding yeast *S. cerevisiae*. The *S. cerevisiae* genome contains about 12.5 million nucleotide pairs and 6600 genes—spread across 16 chromosomes. Note that, for each gene, only one of the two DNA strands actually encodes the information to make an RNA molecule. This coding region can fall on either strand, as indicated by the *light red* bars. However, each "gene" is considered to include both the "coding strand" and its complement. The high density of genes is characteristic of *S. cerevisiae*.

Figure 5–11 In many eukaryotes, genes include an excess of interspersed, noncoding DNA. Presented here is the nucleotide sequence of the human β-globin gene. This gene carries the information that specifies the amino acid sequence of one of the two types of subunits found in hemoglobin, a protein that carries oxygen in the blood. Only the sequence of the coding strand is shown here; the noncoding strand of the double helix carries the complementary sequence. Starting from its 5′ end, such a sequence is read from left to right, like any piece of English text. The segments of the DNA sequence that encode the amino acid sequence of β-globin are highlighted in *yellow*. We will see in Chapter 7 how this information is transcribed and translated to produce a full-length β-globin protein.

CCCTGTGGAGCCACACCCTAGGGTTGGCCA
ATCTACTCCCAGGAGCAGGGAGGGCAGGAG
CCAGGGCTGGGCATAAAAGTCAGGGCAGAG
CCATCTATTGCTTACATTTGCTTCTGACAC
AACTGTGTTCACTAGCAACTCAAACAGACA
CCATGGTGCACCTGACTCCTGAGGAGAAGT
CTGCCGTTACTGCCCTGTGGGGCAAGGTGA
ACGTGGATGAAGTTGGTGGTGAGGCCCTGG
GCAGGTTGGTATCAAGGTTACAAGACAGGT
TTAAGGAGACCAATAGAAACTGGGCATGTG
GAGACAGAGAAGACTCTTGGGTTTCTGATA
GGCACTGACTCTCTCTGCCTATTGGTCTAT
TTTCCCACCCTTAGGCTGCTGGTGGTCTAC
CCTTGGACCCAGAGGTTCTTTGAGTCCTTT
GGGGATCTGTCCACTCCTGATGCTGTTATG
GGCAACCCTAAGGTGAAGGCTCATGGCAAG
AAAGTGCTCGGTGCCTTTAGTGATGGCCTG
GCTCACCTGGACAACCTCAAGGGCACCTTT
GCCACACTGAGTGAGCTGCACTGTGACAAG
CTGCACGTGGATCCTGAGAACTTCAGGGTG
AGTCTATGGGACCCTTGATGTTTTCTTTCC
CCTTCTTTTCTATGGTTAAGTTCATGTCAT
AGGAAGGGGAGAAGTAACAGGGTACAGTTT
AGAATGGGAAACAGACGAATGATTGCATCA
GTGTGGAAGTCTCAGGATCGTTTTAGTTTC
TTTTATTTGCTGTTCATAACAATTGTTTTC
TTTTGTTTAATTCTTGCTTTCTTTTTTTTT
CTTCTCCGCAATTTTTACTATTATACTTAA
TGCCTTAACATTGTGTATAACAAAAGGAAA
TATCTCTGAGATACATTAAGTAACTTAAAA
AAAAACTTTACACAGTCTGCCTAGTACATT
ACTATTTGGAATATATGTGTGCTTATTTGC
ATATTCATAATCTCCCTACTTTATTTTCTT
TTATTTTTAATTGATACATAATCATTATAC
ATATTTATGGGTTAAAGTGTAATGTTTTAA
TATGTGTACACATATTGACCAAATCAGGGT
AATTTTGCATTTGTAATTTTAAAAAATGCT
TTCTTCTTTTAATATACTTTTTTGTTTATC
TTATTTCTAATACTTTCCCTAATCTCTTTC
TTTCAGGGCAATAATGATACAATGTATCAT
GCCTCTTTGCACCATTCTAAAGAATAACAG
TGATAATTTCTGGGTTAAGGCAATAGCAAT
ATTTCTGCATATAAATATTTCTGCATATAA
ATTGTAACTGATGTAAGAGGTTTCATATTG
CTAATAGCAGCTACAATCCAGCTACCATTC
TGCTTTTATTTTATGGTTGGGATAAGGCTG
GATTATTCTGAGTCCAAGCTAGGCCCTTTT
GCTAATCATGTTCATACCTCTTATCTTCCT
CCCACAGCTCCTGGGCAACGTGCTGGTCTG
TGTGCTGGCCCATCACTTTGGCAAAGAATT
CACCCCACCAGTGCAGGCTGCCTATCAGAA
AGTGGTGGCTGGTGTGGCTAATGCCCTGGC
CCACAAGTATCACTAAGCTCGCTTTCTTGC
TGTCCAATTTCTATTAAAGGTTCCTTTGTT
CCCTAAGTCCAACTACTAAACTGGGGGATA
TTATGAAGGGCCTTGAGCATCTGGATTCTG
CCTAATAAAAAACATTTATTTTCATTGCAA
TGATGTATTTAAATTATTTCTGAATATTTT
ACTAAAAAGGGAATGTGGGAGGTCAGTGCA
TTTAAAACATAAAGAAATGATGAGCTGTTC
AAACCTTGGGAAAATACACTATATCTTAAA
CTCCATGAAAGAAGGTGAGGCTGCAACCAG
CTAATGCACATTGGCAACAGCCCCTGATGC
CTATGCCTTATTCATCCCTCAGAAAAGGAT
TCTTGTAGAGGCTTGATTTGCAGGTTAAAG
TTTTGCTATGCTGTATTTTACATTACTTAT
TGTTTTAGCTGTCCTCATGAATGTCTTTTC

Specialized DNA Sequences Are Required for DNA Replication and Chromosome Segregation

To form a functional chromosome, a DNA molecule must do more than simply carry genes: it must be able to be replicated, and the replicated copies must be separated and partitioned equally and reliably into the two daughter cells at each cell division. These processes occur through an ordered series of events, known collectively as the **cell cycle**. This cycle of cell growth and division is summarized—very briefly—in **Figure 5–13** and will be discussed in detail in Chapter 18. Only two broad stages of the cell cycle need concern us in this chapter: *interphase*, when chromosomes are duplicated, and *mitosis*, the much more brief stage when the duplicated chromosomes are distributed, or segregated, to the two daughter nuclei.

During interphase, chromosomes are extended as long, thin, tangled threads of DNA in the nucleus and cannot be easily distinguished in the light microscope (see Figure 5–1). We refer to chromosomes in this extended state as *interphase chromosomes*. It is during interphase that DNA replication takes place. As we discuss in Chapter 6, two specialized DNA sequences, found in all eukaryotes, ensure that this process occurs efficiently. One type of nucleotide sequence, called a **replication origin**, is where DNA replication begins; eukaryotic chromosomes contain many replication origins to allow the long DNA molecules to be replicated rapidly (**Figure 5–14**). Another DNA sequence forms the **telomeres** that mark the ends of each chromosome. Telomeres contain repeated nucleotide sequences that are required for the ends of chromosomes to be fully replicated. They also serve as a protective cap that keeps the chromosome tips from being mistaken by the cell as broken DNA in need of repair.

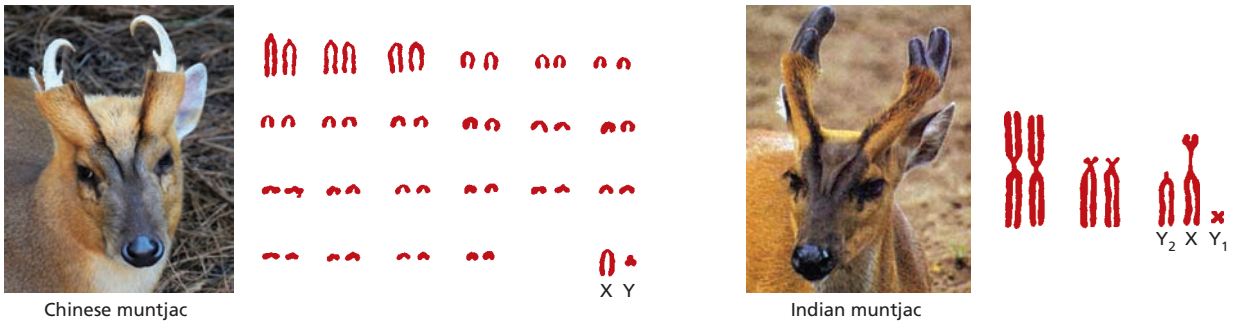


Figure 5–12 Two closely related species can have similar genome sizes but very different chromosome numbers. In the evolution of the Indian muntjac deer, chromosomes that were initially separate, and that remain separate in the Chinese species, fused without having a major effect on the number of genes—or the animal. (Image left, courtesy of Deborah Carreno, Natural Wonders Photography; image right, courtesy of Beatrice Bourgery.)

Eukaryotic chromosomes also contain a third type of specialized DNA sequence, called the **centromere**, that allows duplicated chromosomes to be separated during M phase (see Figure 5–14). During this stage of the cell cycle, the DNA coils up, adopting a more and more compact structure, ultimately forming highly compacted, or condensed, *mitotic chromosomes* (**Figure 5–15**). This is the state in which the duplicated chromosomes can be most easily visualized (see Figure 5–1). Once the chromosomes have condensed, the centromere allows the mitotic spindle to attach to each duplicated chromosome in a way that directs one copy of each chromosome to be segregated to each of the two daughter cells (see Figure 5–13). We describe the central role that centromeres play in cell division in Chapter 18.

Interphase Chromosomes Are Not Randomly Distributed Within the Nucleus

Interphase chromosomes are much longer and finer than mitotic chromosomes. They are nevertheless organized within the nucleus in several ways. First, although interphase chromosomes are constantly undergoing dynamic rearrangements, each tends to occupy a particular region, or territory, of the interphase nucleus (**Figure 5–16**). This loose organization prevents interphase chromosomes from becoming extensively

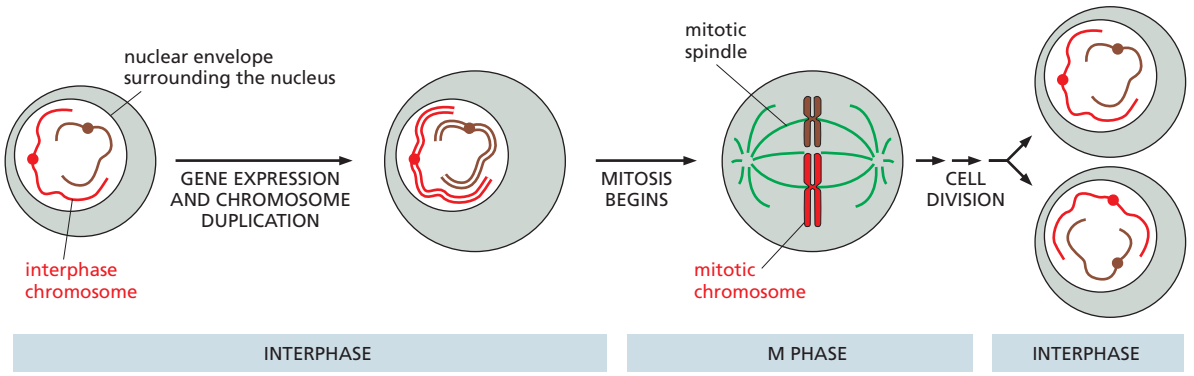


Figure 5–13 The duplication and segregation of chromosomes occurs through an ordered cell cycle in proliferating cells. During interphase, the cell expresses many of its genes, and—during part of this phase—it duplicates its chromosomes. Once chromosome duplication is complete, the cell can enter *M phase*, during which nuclear division, or mitosis, occurs. In mitosis, the duplicated chromosomes condense, gene expression largely ceases, the nuclear envelope breaks down, and the mitotic spindle forms from microtubules and other proteins. The condensed chromosomes are then captured by the mitotic spindle, one complete set is pulled to each end of the cell, and a nuclear envelope forms around each chromosome set. In the final step of M phase, the cell divides to produce two daughter cells. Only two different chromosomes are shown here for simplicity.

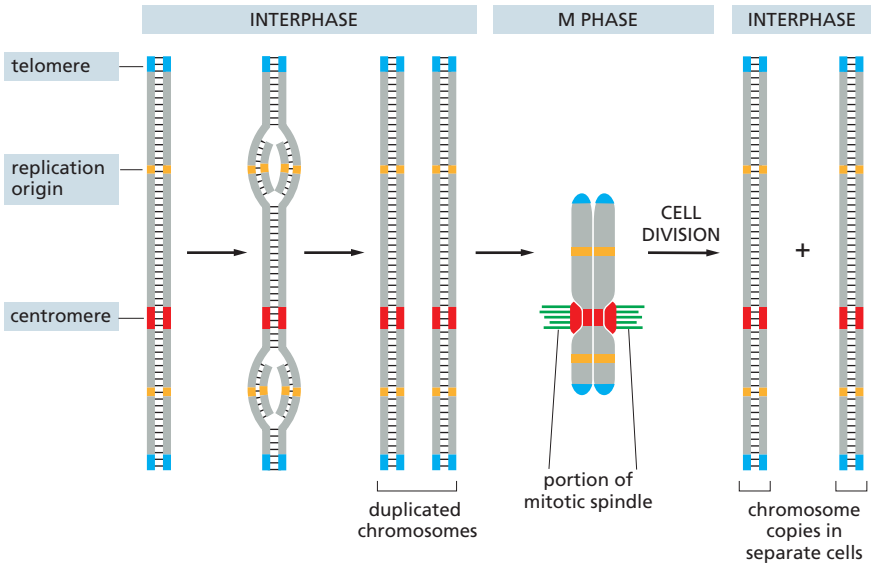


Figure 5–14 Three DNA sequence elements are needed to produce a eukaryotic chromosome that can be duplicated and then segregated at mitosis. Each chromosome has multiple origins of replication, one centromere, and two telomeres. The sequence of events that a typical chromosome follows during the cell cycle is shown schematically. The DNA replicates in interphase, beginning at the origins of replication and proceeding bidirectionally from each origin along the chromosome. In M phase, the centromere attaches the compact, duplicated chromosomes to the mitotic spindle so that one copy will be distributed to each daughter cell when the cell divides. Prior to cell division, the centromere also helps to hold the duplicated chromosomes together until they are ready to be pulled apart. Telomeres contain DNA sequences that allow for the complete replication of chromosome ends.

entangled, like spaghetti in a bowl. In addition, some chromosomal regions are physically attached to particular sites on the *nuclear envelope*—the pair of concentric membranes that surround the nucleus—or to the underlying *nuclear lamina*, the protein meshwork that supports the envelope (discussed in Chapter 17). These attachments also help interphase chromosomes remain within their distinct territories.

The most obvious example of chromosomal organization in the interphase nucleus is the **nucleolus**—a structure large enough to be seen in the light microscope (**Figure 5–17A**). During interphase, the parts of different chromosomes that carry genes encoding ribosomal RNAs come together to form the nucleolus. In human cells, several hundred copies of these genes are distributed in 10 clusters, located near the tips of five different chromosome pairs (**Figure 5–17B**). In the nucleolus, ribosomal RNAs are synthesized and combine with proteins to form ribosomes, the cell's protein-synthesizing machines. As we discuss in Chapter 7, ribosomal RNAs play both structural and catalytic roles in the ribosome.

The DNA in Chromosomes Is Always Highly Condensed

As we have seen, all eukaryotic cells, whether in interphase or mitosis, package their DNA tightly into chromosomes. Human chromosome 22, for example, contains about 48 million nucleotide pairs; stretched out end-to-end, its DNA would extend about 1.5 cm. Yet, during mitosis, chromosome 22 measures only about 2 μm in length—that is, nearly 10,000 times more compact than the DNA would be if it were extended to its full length. This remarkable feat of compression is performed by proteins that coil and fold the DNA into higher and higher levels of organization.

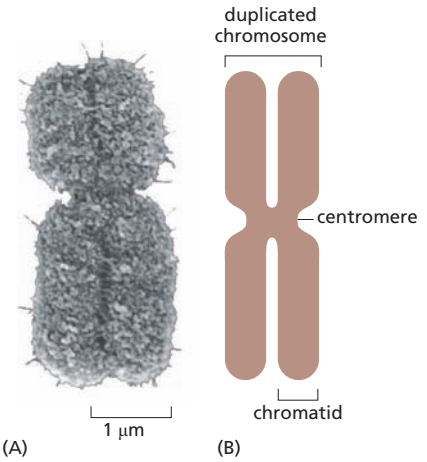


Figure 5–15 A typical duplicated mitotic chromosome is highly compact. Because DNA is replicated during interphase, each mitotic chromosome contains two identical duplicated DNA molecules (see Figure 5–14). Each of these very long DNA molecules, with its associated proteins, is called a *chromatid*; as soon as the two sister chromatids separate, they are considered individual chromosomes. (A) A scanning electron micrograph of a mitotic chromosome. The two chromatids are tightly joined together. The constricted region reveals the position of the centromere. (B) A cartoon representation of a mitotic chromosome. (A, courtesy of Terry D. Allen.)

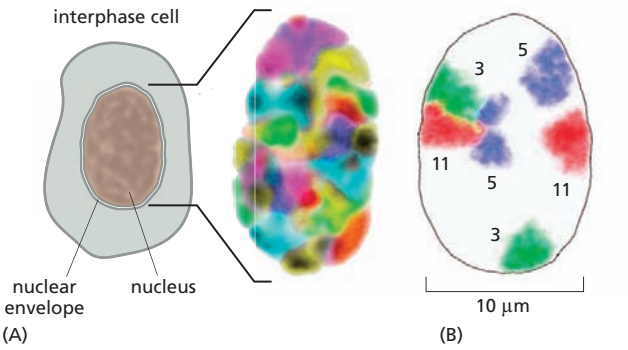


Figure 5–16 Interphase chromosomes occupy their own distinct territories within the nucleus. DNA probes coupled with different fluorescent markers are used to paint individual interphase chromosomes in a human cell. (A) Viewed in a fluorescence microscope, the nucleus is seen to be filled with a patchwork of discrete colors. (B) To highlight their distinct locations, three sets of chromosomes are singled out: chromosomes 3, 5, and 11. Note that pairs of homologous chromosomes, such as the two copies of chromosome 3, are not generally located in the same position. (Adapted from M.R. Hübner and D.L. Spector, *Annu. Rev. Biophys.* 39:471–489, 2010.)

Although the DNA of interphase chromosomes is packed tightly into the nucleus, it is about 20 times less condensed than that of mitotic chromosomes (**Figure 5–18**).

In the next sections, we introduce the specialized proteins that make this compression possible. Bear in mind, though, that chromosome structure is dynamic. Not only do chromosomes condense and decondense during the cell cycle, but chromosome packaging must be flexible enough to allow rapid, on-demand access to different regions of the interphase chromosome, unpacking enough to allow protein complexes access to specific, localized nucleotide sequences for DNA replication, DNA repair, or gene expression.

Nucleosomes Are the Basic Units of Eukaryotic Chromosome Structure

The proteins that bind to DNA to form eukaryotic chromosomes are traditionally divided into two general classes: the **histones** and the *nonhistone chromosomal proteins*. Histones are present in enormous quantities (more than 60 million molecules of several different types in each human cell), and their total mass in chromosomes is about equal to that of the DNA itself. Nonhistone chromosomal proteins are also present in large numbers; they include hundreds of different chromatin-associated proteins. In contrast, only a handful of different histone proteins are present in eukaryotic cells. The complex of both classes of protein with nuclear DNA is called **chromatin**.

Histones are responsible for the first and most fundamental level of chromatin packing: the formation of the **nucleosome**. Nucleosomes convert the DNA molecules in an interphase nucleus into a *chromatin fiber* that

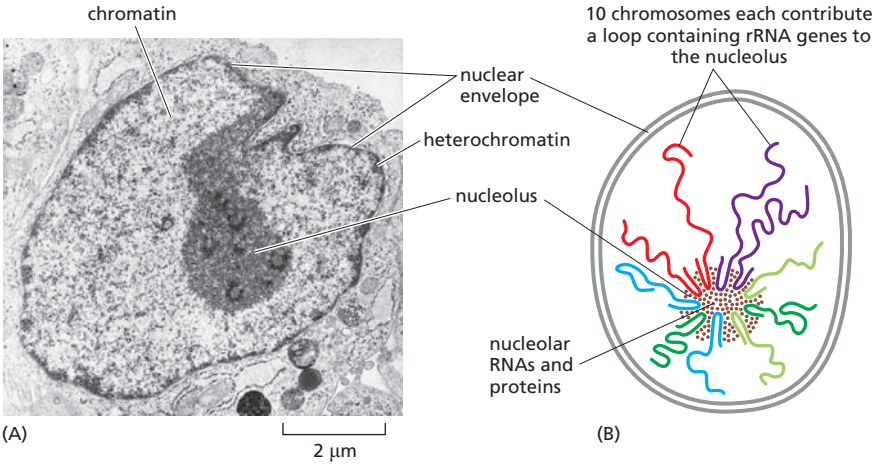


Figure 5–17 The nucleolus is the most prominent structure in the interphase nucleus. (A) Electron micrograph of a thin section through the nucleus of a human fibroblast. The nucleus is surrounded by the nuclear envelope. Inside the nucleus, the chromatin appears as a diffuse speckled mass; regions that are especially dense are called heterochromatin (dark staining). Heterochromatin contains few genes and is located mainly around the periphery of the nucleus, immediately under the nuclear envelope. The large, dark region within the nucleus is the nucleolus, which contains the genes for ribosomal RNAs. (B) Schematic illustration showing how ribosomal RNA genes, which are clustered near the tips of five different human chromosomes (13, 14, 15, 21, and 22), come together to form the nucleolus, which is a biochemical subcompartment produced by the aggregation of a set of macromolecules—DNA, RNAs, and proteins (see Figure 4–54). (A, courtesy of E.G. Jordan and J. McGovern.)

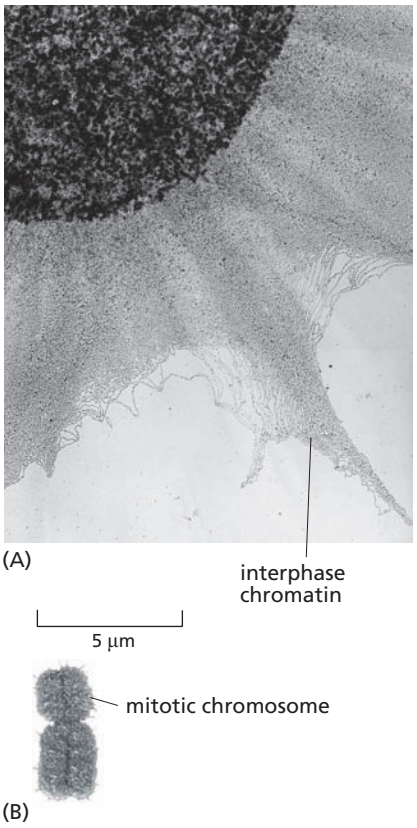


Figure 5–18 DNA in interphase chromosomes is less compact than in mitotic chromosomes. (A) An electron micrograph showing an enormous tangle of chromatin (DNA with its associated proteins) spilling out of a lysed interphase nucleus. (B) For comparison, a compact, human mitotic chromosome is shown at the same scale. (A, courtesy of Victoria Foe; B, courtesy of Terry D. Allen.)

is approximately one-third the length of the initial DNA. These chromatin fibers, when examined with an electron microscope, contain clusters of closely packed nucleosomes (**Figure 5–19A**). If this chromatin is then subjected to treatments that cause it to unfold partially, it can then be seen in the electron microscope as a series of "beads on a string" (**Figure 5–19B**). The string is DNA, and each bead is a *nucleosome core particle*, which consists of DNA wound around a core of histone proteins.

To determine the structure of the nucleosome core particle, investigators treated chromatin in its unfolded, "beads-on-a-string" form with enzymes called nucleases, which cut the DNA by breaking the phosphodiester bonds between nucleotides. When this nuclease digestion is carried out for a short time, only the exposed DNA between the core particles—the *linker DNA*—will be cleaved, allowing the core particles to be isolated. An individual nucleosome core particle consists of a complex of eight histone proteins—two molecules each of histones H2A, H2B, H3, and H4—along with a segment of double-stranded DNA, 147 nucleotide pairs long, that winds around this *histone octamer* (**Figure 5–20**). The high-resolution structure of the nucleosome core particle was solved in 1997, revealing in atomic detail the disc-shaped histone octamer around which the DNA is tightly wrapped, making 1.7 turns in a left-handed coil (**Figure 5–21**). The linker DNA between each nucleosome core particle can vary in length from a few nucleotide pairs up to about 80. Technically speaking, a "nucleosome" consists of a nucleosome core particle plus one of its adjacent DNA linkers, as shown in Figure 5–20; however, the term is often used to refer to the nucleosome core particle itself.

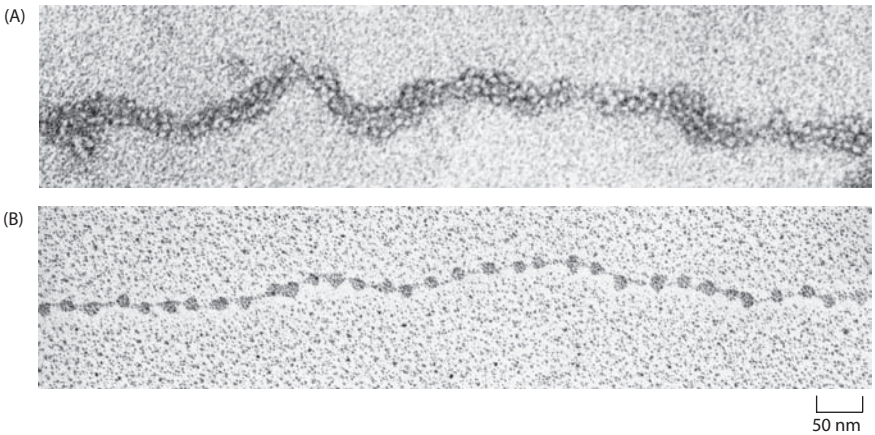


Figure 5–19 Nucleosomes can be seen in the electron microscope. (A) Chromatin isolated directly from an interphase nucleus can appear in the electron microscope as a chromatin fiber, composed of packed nucleosomes. (B) Another electron micrograph shows a length of a chromatin fiber that has been experimentally unpacked, or decondensed, after isolation to show the "beads-on-a-string" appearance of the nucleosomes. (A, courtesy of Barbara Hamkalo; B, courtesy of Victoria Foe.)

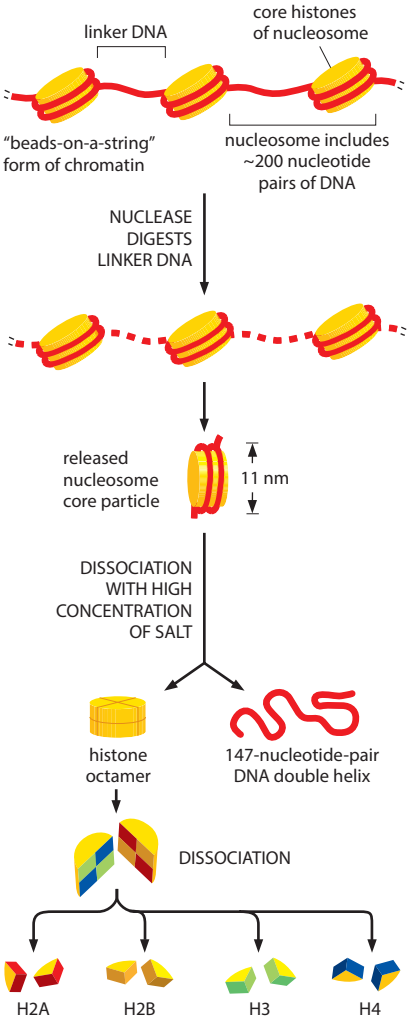


Figure 5–20 Nucleosomes contain DNA wrapped around a protein core of eight histone molecules. In a test tube, the nucleosome core particle can be released from chromatin by digestion of the linker DNA with a nuclease, which cleaves the exposed linker DNA but not the DNA wound tightly around the nucleosome core. When the DNA around each isolated nucleosome core particle is released, its length is found to be 147 nucleotide pairs; this DNA wraps around the histone octamer that forms the nucleosome core nearly twice.

All four of the histones that make up the octamer are relatively small proteins with a high proportion of positively charged amino acids (lysine and arginine). The positive charges help the histones bind tightly to the negatively charged sugar–phosphate backbone of DNA. These numerous electrostatic interactions explain in part why DNA of virtually any sequence can bind to a histone octamer. Each of the histones in the octamer also has a long, unstructured N-terminal amino acid "tail" that extends out from the nucleosome core particle (see the H3 tail in Figure 5–21). These histone tails are subject to several types of reversible, covalent chemical modifications that control many aspects of chromatin structure.

The histones that form the nucleosome core are among the most highly conserved of all known eukaryotic proteins: there are only two differences between the amino acid sequences of histone H4 from peas and cows, for example. This extreme evolutionary conservation reflects the vital role of histones in controlling eukaryotic chromosome structure.

Chromosome Packing Occurs on Multiple Levels

Although long strings of nucleosomes form on most chromosomal DNA, chromatin in the living cell rarely adopts the extended beads-on-a-string form seen in Figure 5–19B. Instead, the nucleosomes are further packed on top of one another to generate a more compact structure, such as the chromatin fiber shown in Figure 5–19A and **Movie 5.2**. This additional packing of nucleosomes into a chromatin fiber depends on a fifth

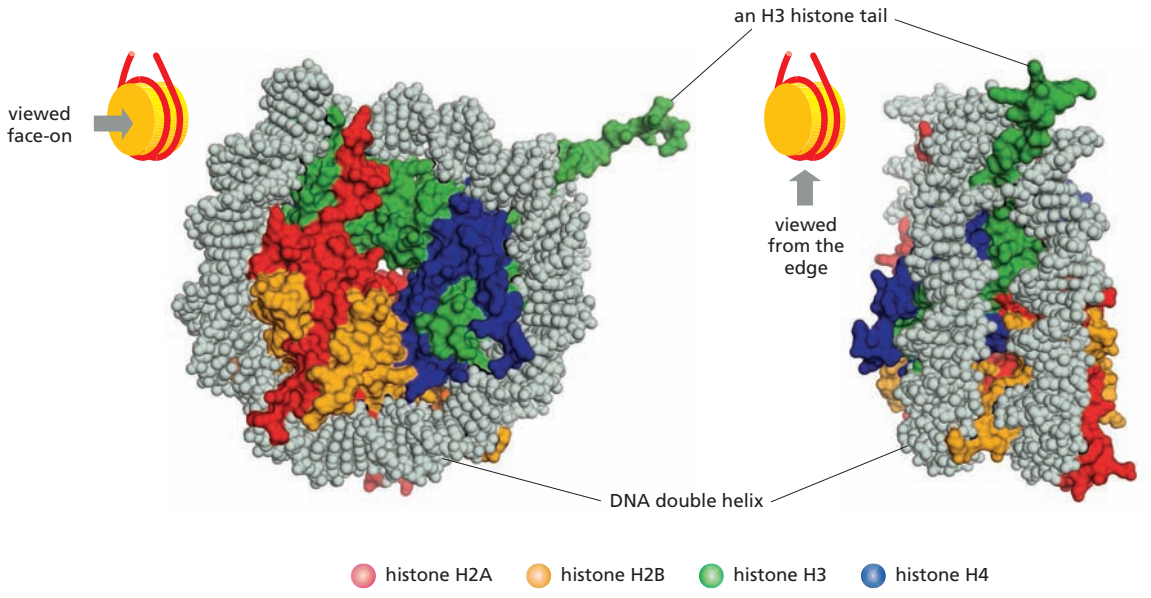


Figure 5–21 The structure of the nucleosome core particle, as determined by x-ray diffraction analysis, reveals how DNA is tightly wrapped around a disc-shaped histone octamer. Two views of a nucleosome core particle are shown here. The two strands of the DNA double helix are shown in *gray*. A portion of an H3 histone tail (*green*) can be seen extending from the nucleosome core particle, but the tails of the other histones have been truncated. (From K. Luger et al., *Nature* 389:251–260, 1997.)

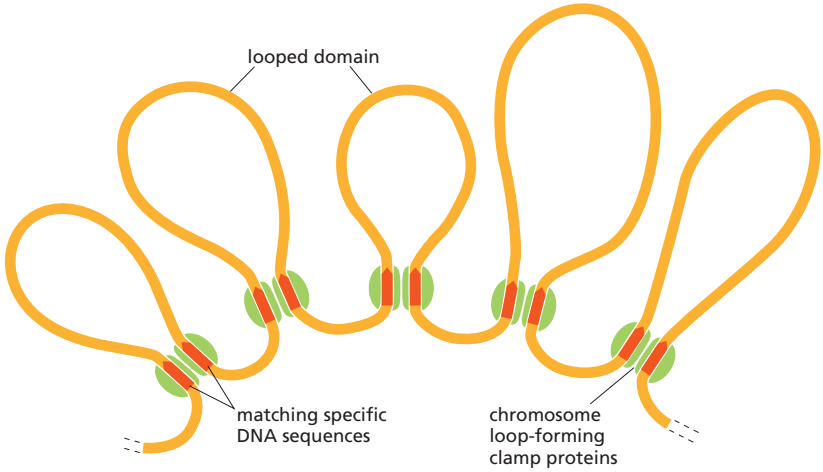


Figure 5–22 The chromatin in human chromosomes is folded into looped domains. These loops are established by special nonhistone chromosomal proteins that bind to specific DNA sequences, creating a clamp at the base of each loop.

histone called histone H1, which is thought to pull adjacent nucleosomes together into a regular repeating array. This "linker" histone changes the path the DNA takes as it exits the nucleosome core, allowing it to form a more condensed chromatin fiber.

We saw earlier that, during mitosis, chromatin becomes so highly condensed that individual chromosomes can be seen in the light microscope. How is a chromatin fiber folded to produce mitotic chromosomes? Although the answer is not yet known in detail, it is known that specialized nonhistone chromosomal proteins fold the chromatin into a series of loops (**Figure 5–22**). These loops are further condensed to produce the interphase chromosome. Finally, this compact string of loops is thought to undergo at least one more level of packing to form the mitotic chromosome (**Figure 5–23**).

QUESTION 5–2

Assuming that the histone octamer (shown in Figure 5–20) forms a cylinder 9 nm in diameter and 5 nm in height and that the human genome forms 32 million nucleosomes, what volume of the nucleus (6 μm in diameter) is occupied by histone octamers? (Volume of a cylinder is $\pi r^2 h$; volume of a sphere is $4/3\ \pi r^3$.) What fraction of the total volume of the nucleus do the histone octamers occupy? How does this compare with the volume of the nucleus occupied by human DNA?

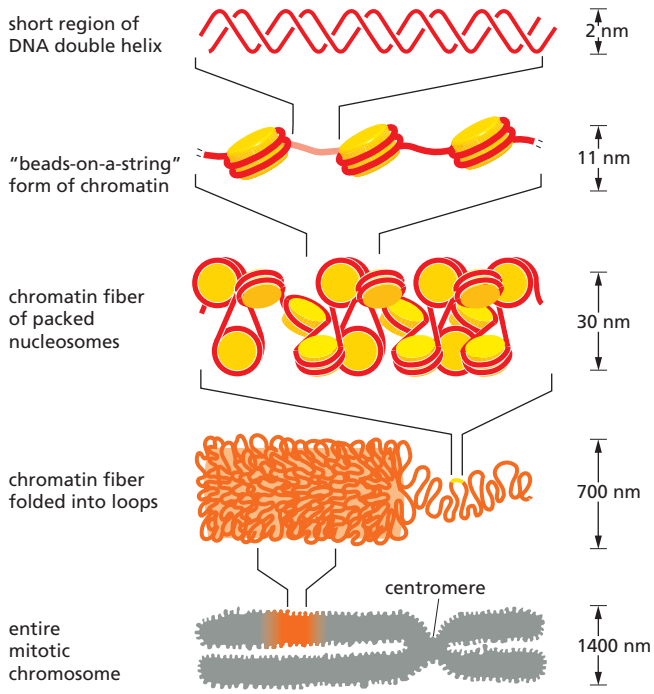


Figure 5–23 DNA packing occurs on several levels in chromosomes. This schematic drawing shows some of the levels thought to give rise to the highly condensed mitotic chromosome. Both histone H1 and a set of specialized nonhistone chromosomal proteins are known to help drive these condensations, including the chromosome loop-forming clamp proteins and the abundant non-histone protein condensin (see Figure 18–18). However, the actual structures are still uncertain.

THE REGULATION OF CHROMOSOME STRUCTURE

So far, we have discussed how DNA is packed tightly into chromatin. We now turn to the question of how this packaging can be adjusted to allow rapid access to the underlying DNA. The DNA in cells carries enormous amounts of coded information, and cells must be able to retrieve this information as needed.

In this section, we discuss how a cell can alter its chromatin structure to expose localized regions of DNA and allow access to specific proteins and protein complexes, particularly those involved in gene expression and in DNA replication and repair. We then discuss how chromatin structure is established and maintained—and how a cell can pass on some forms of this structure to its descendants, helping different cell types to sustain their identity. Although many of the details remain to be deciphered, the regulation and inheritance of chromatin structure play crucial roles in the development of eukaryotic organisms.

Changes in Nucleosome Structure Allow Access to DNA

Eukaryotic cells have several ways to adjust rapidly the local structure of their chromatin. One way takes advantage of a set of ATP-dependent **chromatin-remodeling complexes**. These protein machines use the energy of ATP hydrolysis to change the position of the DNA wrapped around nucleosomes (**Figure 5–24**). By interacting with both the histone octamer and the DNA wrapped around it, chromatin-remodeling complexes can locally alter the arrangement of the nucleosomes, rendering the DNA more accessible (or less accessible) to other proteins in the cell. During mitosis, many of these complexes are inactivated, which may help mitotic chromosomes maintain their tightly packed structure.

Another way of altering chromatin structure relies on the reversible chemical modification of histones, catalyzed by a large number of different **histone-modifying enzymes**. The tails of all four of the core histones are particularly subject to these covalent modifications, which include the addition (and removal) of acetyl, phosphate, or methyl groups

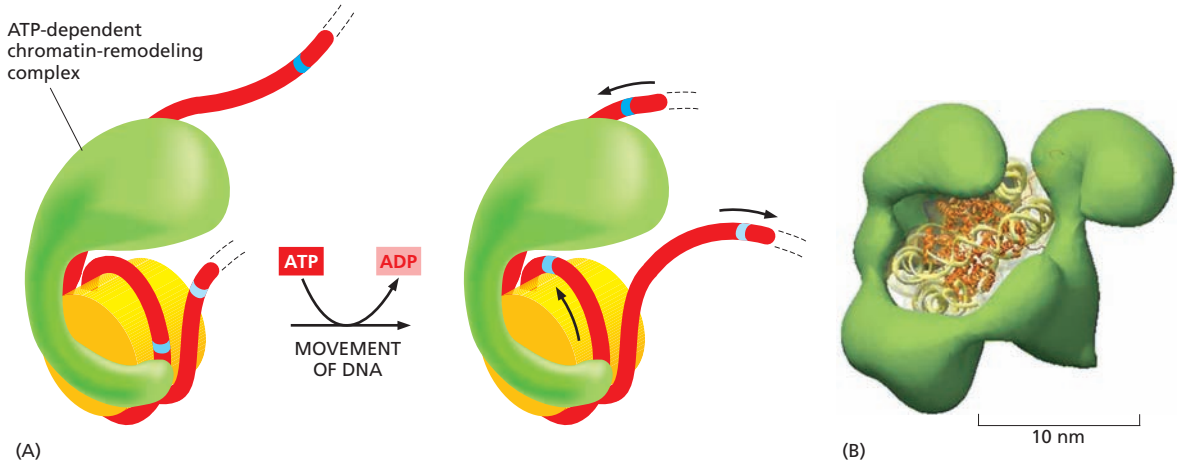


Figure 5–24 Chromatin-remodeling complexes locally reposition the DNA wrapped around nucleosomes. (A) The complexes use energy derived from ATP hydrolysis to loosen the nucleosomal DNA and push it along the histone octamer. In this way, the enzyme can expose or hide a sequence of DNA, controlling its availability to other DNA-binding proteins. The *blue* stripes have been added to show how the DNA shifts its position. Many cycles of ATP hydrolysis are required to produce such a shift. (B) The structure of a chromatin-remodeling complex, showing how the enzyme cradles a nucleosome core particle, including a histone octamer (*orange*) and the DNA wrapped around it (*light green*). This large complex, purified from yeast, contains 15 subunits, including one that hydrolyzes ATP and four that recognize specific covalently modified histones. (B, adapted from A.E. Leschziner et al., *Proc. Natl. Acad. Sci. USA* 104:4913–4918, 2007.)

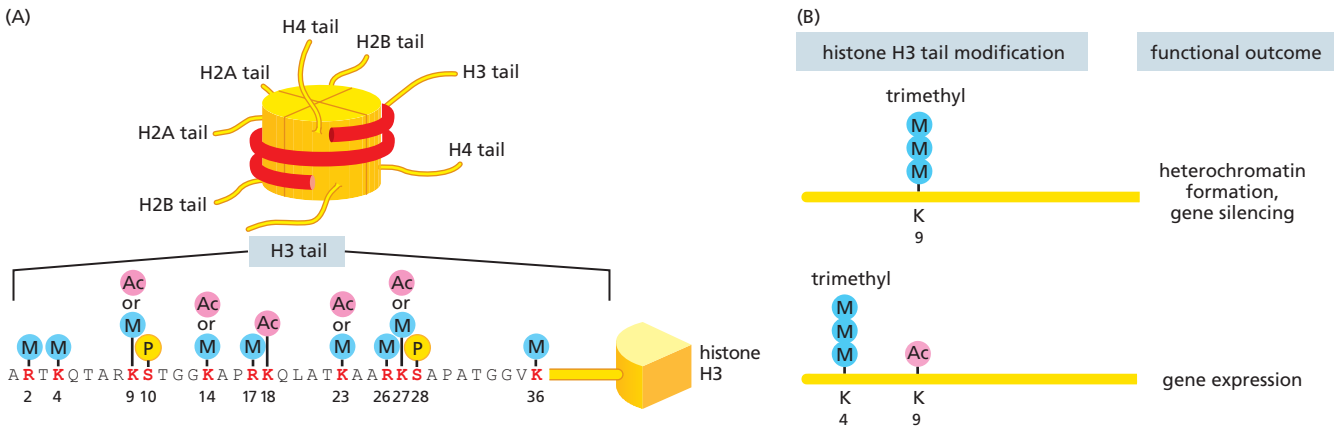


Figure 5–25 The pattern of modification of histone tails can determine how a stretch of chromatin is handled by the cell. (A) Schematic drawing showing the positions of the histone tails that extend from each nucleosome core particle. Each histone can be modified by the covalent attachment of a number of different chemical groups, mainly to the tails. The tail of histone H3, for example, can receive acetyl groups (Ac), methyl groups (M), or phosphate groups (P). The numbers denote the positions of the modified amino acids in the histone tail, with each amino acid designated by its one-letter code. Note that some amino acids, such as the lysine (K) at positions 9, 14, 23, and 27, can be modified by acetylation or methylation (but not by both at once). Lysines, in addition, can be modified with either one, two, or three methyl groups; trimethylation, for example, is shown in (B). Note that histone H3 contains 135 amino acids, most of which are in its globular portion (represented by the wedge); most modifications occur on the N-terminal tail, for which 36 amino acids are shown. (B) Different combinations of histone tail modifications can confer a specific meaning on the stretch of chromatin on which they occur, as indicated. Only a few of these functional outcomes are known.

(**Figure 5–25A**). These and other modifications can have important consequences for the packing of the chromatin fiber. Acetylation of lysines, for instance, can reduce the affinity of the tails for adjacent nucleosomes, thereby loosening chromatin structure and allowing access to particular nuclear proteins.

Most importantly, however, these modifications generally serve as docking sites on the histone tails for a variety of regulatory proteins. Different patterns of modifications attract specific sets of non-histone chromosomal proteins to a particular stretch of chromatin. Some of these proteins promote chromatin condensation, whereas others promote chromatin expansion and thus facilitate access to the DNA. Specific combinations of tail modifications, and the proteins that bind to them, have different functional outcomes for the cell: one pattern, for example, might mark a particular stretch of chromatin as newly replicated; another might indicate that the genes in that stretch of chromatin are being actively expressed; still others are associated with genes that are silenced (**Figure 5–25B**).

Both ATP-dependent chromatin-remodeling complexes and histone-modifying enzymes are tightly regulated. These enzymes are often brought to particular chromatin regions by interactions with proteins that bind to a specific nucleotide sequence in the DNA—or in an RNA transcribed from this DNA (a topic we return to in Chapter 8). Histone-modifying enzymes work in concert with the chromatin-remodeling complexes to condense and relax stretches of chromatin, allowing local chromatin structure to change rapidly according to the needs of the cell.

Interphase Chromosomes Contain both Highly Condensed and More Extended Forms of Chromatin

The localized alteration of chromatin packing by remodeling complexes and histone modification has important effects on the large-scale structure of interphase chromosomes. Interphase chromatin is not uniformly packed. Instead, regions of the chromosome containing genes that are being actively expressed are generally more extended, whereas those that contain silent genes are more condensed. Thus, the detailed structure of an interphase chromosome can differ from one cell type to the next, helping to determine which genes are switched on and which are shut down. Most cell types express only about half of the genes they contain, and many of these are active only at very low levels.

The most highly condensed form of interphase chromatin is called **heterochromatin** (from the Greek *heteros*, "different," chromatin). This highly compact form of chromatin was first observed in the light microscope in the 1930s as discrete, strongly staining regions within the total

QUESTION 5–3

Histone proteins are among the most highly conserved proteins in eukaryotes. Histone H4 proteins from a pea and a cow, for example, differ in only 2 of 102 amino acids. Comparison of the gene sequences shows many more differences, but only two change the amino acid sequence. These observations indicate that mutations that change amino acids must have been selected against during evolution. Why do you suppose that amino-acid-altering mutations in histone genes are deleterious?

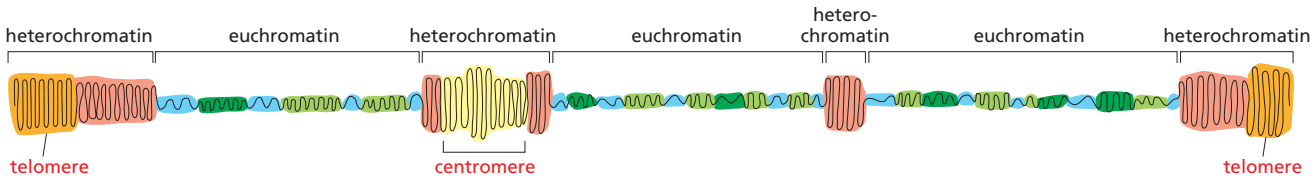


Figure 5–26 The structure of chromatin varies along a single interphase chromosome. As schematically indicated by the path of the DNA molecule (represented by the central *black* line) and the different arbitrarily assigned colors, heterochromatin and euchromatin each represent a set of different chromatin structures with different degrees of condensation. Overall, heterochromatin is more condensed than euchromatin.

chromatin mass. Heterochromatin typically makes up about 10% of an interphase chromosome, and in mammalian chromosomes, it is concentrated around the centromere region and in the telomeric DNA at the chromosome ends (see Figure 5–14).

The rest of the interphase chromatin is called **euchromatin** (from the Greek *eu*, "true" or "normal," chromatin). Although we use the term euchromatin to refer to chromatin that exists in a less condensed state than heterochromatin, it is now clear that both euchromatin and heterochromatin are composed of mixtures of different chromatin structures (**Figure 5–26**).

Each type of chromatin structure is established and maintained by different sets of histone tail modifications, which attract distinct sets of nonhistone chromosomal proteins. The modifications that direct the formation of the most common type of heterochromatin, for example, include the methylation of lysine 9 in the tail of histone H3 (see Figure 5–25B). Once heterochromatin has been established, it can spread to neighboring regions of DNA, because its histone tail modifications attract a set of heterochromatin-specific proteins, including histone-modifying enzymes, which then add the same histone tail modifications on adjacent nucleosomes. These modifications in turn recruit more of the heterochromatin-specific proteins, causing a wave of condensed chromatin to propagate along the chromosome. This extended region of heterochromatin will continue to spread until it encounters a barrier DNA sequence that stops the propagation (**Figure 5–27**). As an example, some barrier sequences contain binding sites for histone-modifying enzymes that add

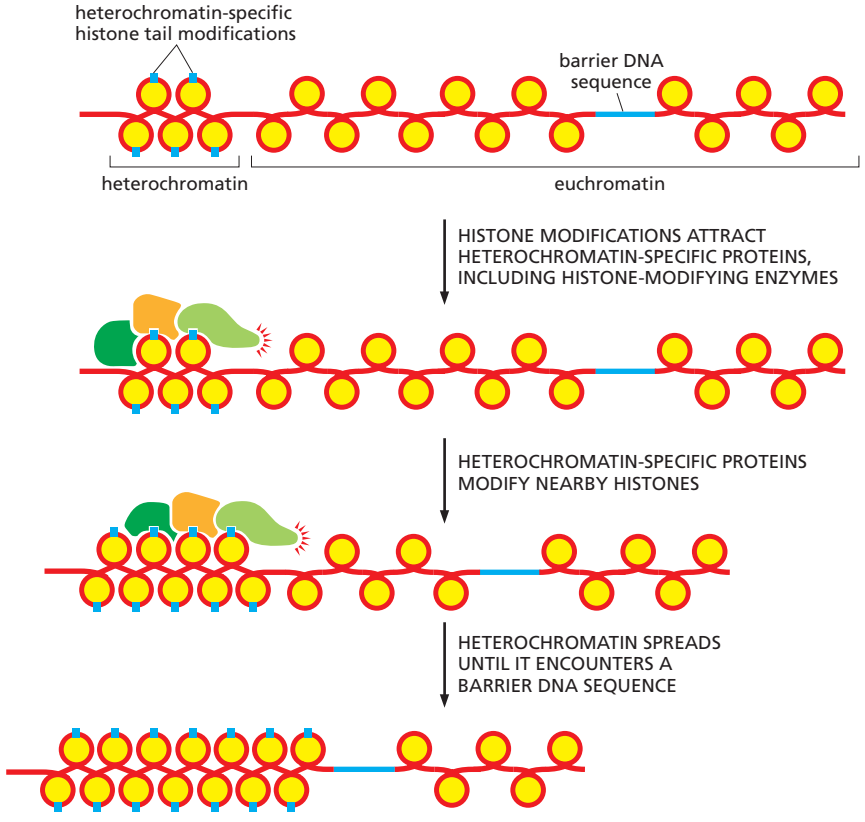


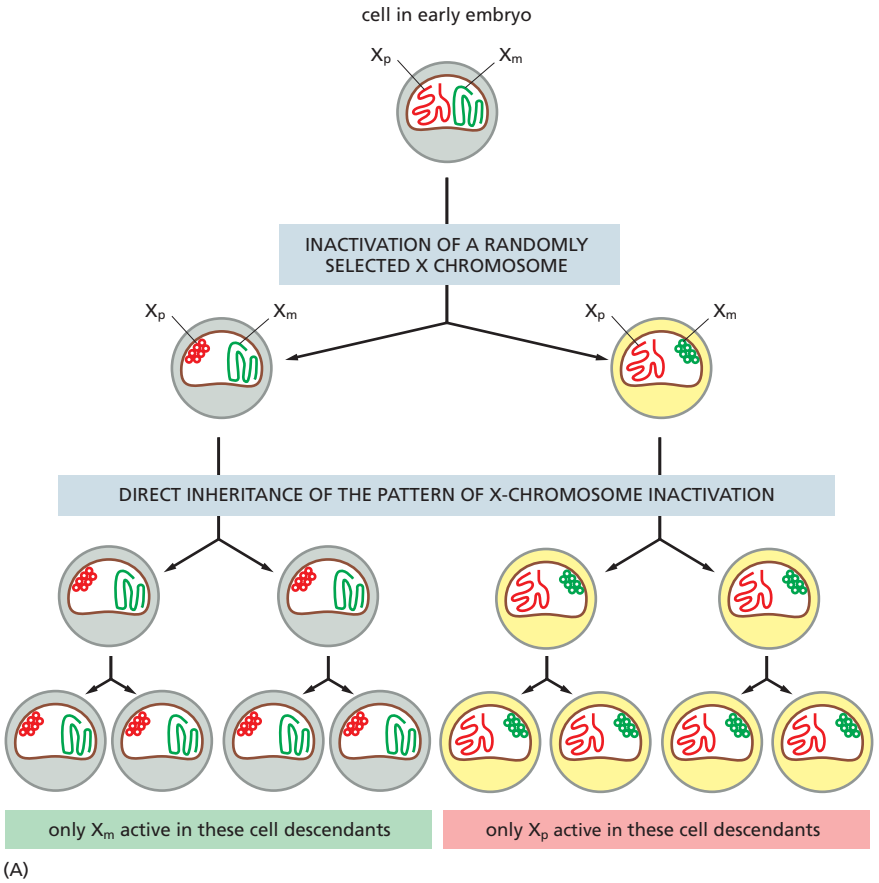
Figure 5–27 Heterochromatin-specific histone modifications allow heterochromatin to form and to spread. These modifications attract heterochromatin-specific proteins that reproduce the same histone modifications on neighboring nucleosomes. In this manner, heterochromatin can spread until it encounters a barrier DNA sequence that blocks further propagation into regions of euchromatin.

an acetyl group to lysine 9 of the histone H3 tail; this modification blocks the methylation of that lysine, preventing any further spread of heterochromatin (see Figure 5–25B).

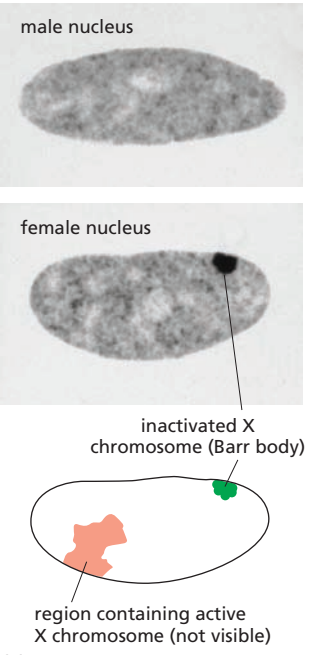
Much of the DNA that is folded into heterochromatin does not contain genes. Because heterochromatin is so compact, genes that accidentally become packaged into heterochromatin usually fail to be expressed. Such inappropriate packaging of genes in heterochromatin can cause disease: in humans, the gene that encodes β-globin—a protein that forms part of the oxygen-carrying hemoglobin molecule—is situated near a region of heterochromatin. In an individual with an inherited deletion of its barrier DNA, that heterochromatin spreads and deactivates the β-globin gene, causing severe anemia.

Perhaps the most striking example of the use of heterochromatin to keep genes shut down, or *silenced*, is found in the interphase X chromosomes of female mammals. In mammals, female cells contain two X chromosomes, whereas male cells contain one X and one Y. A double dose of X-chromosome products could be lethal, and female mammals have evolved a mechanism for permanently inactivating one of the two X chromosomes in each cell. At random, one or other of the two X chromosomes in each nucleus becomes highly condensed into heterochromatin early in embryonic development. Thereafter, the condensed and inactive state of that X chromosome is inherited in all of the many descendants of those cells (**Figure 5–28**). This process of X-inactivation is responsible for the patchwork coloration of calico cats (**Figure 5–29**).

X-inactivation is an extreme example of a process that takes place in all eukaryotic cells—one that operates on a much finer scale to help control gene expression. When a cell divides, it can pass along its histone modifications, chromatin structure, and gene expression patterns to the two daughter cells. Such "cell memory" transmits information about which



(A)



(B)

Figure 5–28 One of the two X chromosomes is inactivated in the cells of mammalian females by heterochromatin formation. (A) Each female cell contains two X chromosomes, one from the mother (X_m) and one from the father (X_p). At an early stage in embryonic development, one of these two chromosomes becomes condensed into heterochromatin in each cell, apparently at random. At each cell division, the same X chromosome becomes condensed (and inactivated) in all the descendants of that original cell. Thus, all mammalian females end up as mixtures (mosaics) of cells bearing either inactivated maternal or inactivated paternal X chromosomes. In most of their tissues and organs, about half the cells will be of one type, and the rest will be of the other. (B) In the nucleus of a female cell, the inactivated X chromosome can be seen as a small, discrete mass of chromatin called a Barr body, named after the physician who first observed it. In these micrographs of the nuclei of human fibroblasts, the inactivated X chromosome in the female nucleus (*bottom* micrograph) has been visualized by use of an antibody that recognizes proteins associated with the Barr body. The male nucleus (*top*) contains only a single X chromosome, which is not inactivated and thus not recognized by this antibody. Below the micrographs, a cartoon shows the locations of both the active and the inactive X chromosomes in the female nucleus. (B, adapted from B. Hong et al. *Proc. Natl Acad. Sci. USA* 98:8703–8708, 2001.)

Figure 5–29 The coat color of a calico cat is dictated in large part by patterns of X-inactivation. In cats, one of the genes specifying coat color is located on the X chromosome. In female calicos, one X chromosome carries the form of the gene that specifies black fur, the other carries the form of the gene that specifies orange fur. Skin cells in which the X chromosome carrying the gene for black fur is inactivated will produce orange fur; those in which the X chromosome carrying the gene for orange fur is inactivated will produce black fur. The size of each patch will depend on the number of skin cells that have descended from an embryonic cell in which one or the other of the X chromosomes was randomly inactivated during development (see Figure 5–28). (bluecaterpillar/Depositphotos.)

QUESTION 5–4

Mutations in a particular gene on the X chromosome result in color blindness in men. By contrast, most women carrying the mutation have proper color vision but see colored objects with reduced resolution, as though functional cone cells (the photoreceptor cells responsible for color vision) are spaced farther apart than normal in the retina. Can you give a plausible explanation for this observation? If a woman is color-blind, what could you say about her father? About her mother? Explain your answers.

genes are active and which are not—a process critical for the establishment and maintenance of different cell types during the development of a complex multicellular organism. We discuss some of the mechanisms involved in cell memory in Chapter 8, when we consider how cells control gene expression.

ESSENTIAL CONCEPTS

- Life depends on the stable storage, maintenance, and inheritance of genetic information.
- Genetic information is carried by very long DNA molecules and is encoded in the linear sequence of four nucleotides: A, T, G, and C.
- Each molecule of DNA is a double helix composed of a pair of antiparallel, complementary DNA strands, which are held together by hydrogen bonds between G-C and A-T base pairs.
- The genetic material of a eukaryotic cell—its genome—is contained in a set of chromosomes, each formed from a single, enormously long DNA molecule that contains many genes.
- When a gene is expressed, part of its nucleotide sequence is transcribed into RNA molecules, most of which are translated to produce a protein.
- The DNA that forms each eukaryotic chromosome contains, in addition to genes, many replication origins, one centromere, and two telomeres. These special DNA sequences ensure that, before cell division, each chromosome can be duplicated efficiently, and that the resulting daughter chromosomes can be parceled out equally to the two daughter cells.
- In eukaryotic chromosomes, the DNA is tightly folded by binding to a set of histone and nonhistone chromosomal proteins. This complex of DNA and protein is called chromatin.
- Histones pack the DNA into a repeating array of DNA–protein particles called nucleosomes, which further fold up into even more compact chromatin structures.
- A cell can regulate its chromatin structure—temporarily decondensing or condensing particular regions of its chromosomes—using chromatin-remodeling complexes and enzymes that covalently modify histone tails in various ways.
- The loosening of chromatin to a more decondensed state allows proteins involved in gene expression, DNA replication, and DNA repair to gain access to the necessary DNA sequences.
- Some forms of chromatin have a pattern of histone tail modification that causes the DNA to become so highly condensed that its genes cannot be expressed to produce RNA; a high degree of condensation occurs on all chromosomes during mitosis and in the heterochromatin of interphase chromosomes.

KEY TERMS

base pair
cell cycle
centromere
chromatin
chromatin-remodeling complex
chromosome
complementary
deoxyribonucleic acid (DNA)
double helix
euchromatin
gene
gene expression
genetic code
genome
heterochromatin
histone
histone-modifying enzyme
karyotype
nucleolus
nucleosome
replication origin
telomere

GENES ARE MADE OF DNA

By the 1920s, scientists generally agreed that genes reside on chromosomes. And studies in the late nineteenth century had demonstrated that chromosomes are composed of both DNA and proteins. But because DNA is so chemically simple, biologists naturally assumed that genes had to be made of proteins, which are much more chemically diverse than DNA molecules. Even when the experimental evidence suggested otherwise, this assumption proved hard to shake.

Messages from the dead

The case for DNA began to emerge in the late 1920s, when a British medical officer named Fred Griffith made an astonishing discovery. He was studying *Streptococcus pneumoniae* (pneumococcus), a bacterium that causes pneumonia. As antibiotics had not yet been discovered, infection with this organism was usually fatal. When grown in the laboratory, pneumococci come in two

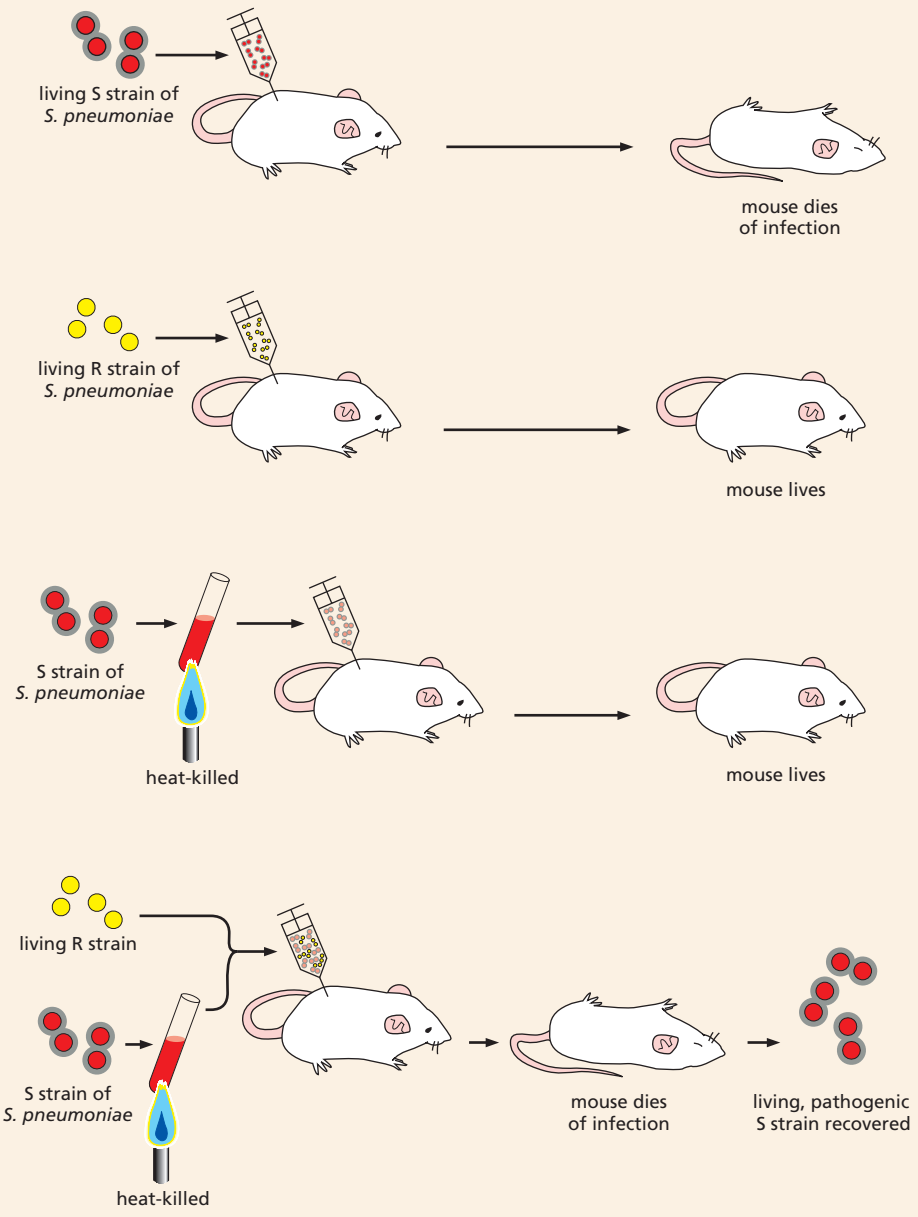


Figure 5–30 Griffith showed that heat-killed infectious bacteria can transform harmless live bacteria into pathogens. The bacterium *Streptococcus pneumoniae* comes in two forms that differ in their microscopic appearance and in their ability to cause disease. Cells of the pathogenic strain, which are lethal when injected into mice, are encased in a slimy, glistening polysaccharide capsule. When grown on a plate of nutrients in the laboratory, this disease-causing bacterium forms colonies that look dome-shaped and smooth; hence it is designated the S form. The harmless strain of the pneumococcus, on the other hand, lacks this protective coat; it forms colonies that appear flat and rough—hence, it is referred to as the R form. As illustrated in this diagram, Griffith found that a substance present in the pathogenic S strain could permanently change, or transform, the nonlethal R strain into the deadly S strain.

forms: a pathogenic form that causes a lethal infection when injected into animals, and a harmless form that is easily conquered by the animal's immune system and does not produce an infection.

In the course of his investigations, Griffith injected various preparations of these bacteria into mice. He showed that pathogenic pneumococci that had been killed by heating were no longer able to cause infection. The surprise came when Griffith injected both heat-killed pathogenic bacteria and live harmless bacteria into the same mouse. This combination proved unexpectedly lethal: not only did the animals die of pneumonia, but Griffith found that their blood was teeming with live bacteria of the pathogenic form (**Figure 5–30**). The heat-killed pneumococci had somehow converted the harmless bacteria into the lethal form. What's more, Griffith found that the change was permanent: he could grow these "transformed" bacteria in culture, and they remained pathogenic. But what was this mysterious material that turned harmless bacteria into killers? And how was this change passed on to progeny bacteria?

Transformation

Griffith's remarkable finding set the stage for the experiments that would provide the first strong evidence that genes are made of DNA. The American bacteriologist Oswald Avery, following up on Griffith's work, discovered that the harmless pneumococcus could be transformed into a pathogenic strain in a test tube by exposing it to an extract prepared from the pathogenic strain. It would take another 15 years, however, for Avery and his colleagues Colin MacLeod and Maclyn McCarty to successfully purify the "transforming principle" from this soluble extract and to demonstrate that the active ingredient was DNA. Because the transforming principle caused a heritable change in the bacteria that received it, DNA must be the very stuff of which genes are made.

The 15-year delay was in part a reflection of the academic climate—and the widespread supposition that the genetic material was likely to be made of protein. Because of the potential ramifications of their work, the researchers wanted to be absolutely certain that the transforming principle was DNA before they announced their findings. As Avery noted in a letter to his brother, also a bacteriologist, "It's lots of fun to blow bubbles, but it's wiser to prick them yourself before someone else tries to." So the researchers subjected the transforming material to a battery of chemical tests (**Figure 5–31**). They found that it exhibited all the chemical properties characteristic of DNA; furthermore, they showed that enzymes that destroy proteins and RNA did not affect the ability of the extract to transform bacteria, while enzymes that destroy DNA inactivated it. And like Griffith before them, the investigators found that their purified preparation changed the bacteria permanently: DNA from the pathogenic species was taken up by the harmless species, and this change was faithfully passed on to subsequent generations of bacteria.

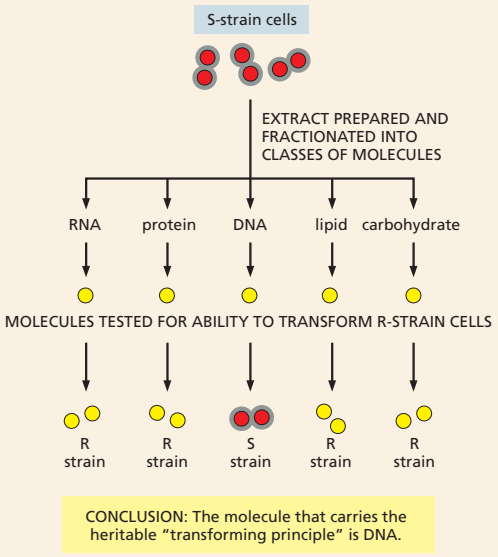


Figure 5–31 Avery, MacLeod, and McCarty demonstrated that DNA is the genetic material. The researchers prepared an extract from the disease-causing S strain of pneumococci and showed that the "transforming principle" that would permanently change the harmless R-strain pneumococci into the pathogenic S strain is DNA. This was the first evidence that DNA could serve as the genetic material.

This landmark study offered rigorous proof that purified DNA can act as genetic material. But the resulting paper, published in 1944, drew strangely little attention. Despite the meticulous care with which these experiments were performed, geneticists were not immediately convinced that DNA is the hereditary material. Many argued that the transformation might have been caused by some trace protein contaminant in the preparations. Or that the extract might contain a mutagen that alters the genetic material of the harmless bacteria—converting them to the pathogenic form—rather than containing the genetic material itself.

Virus cocktails

The debate was not settled definitively until 1952, when Alfred Hershey and Martha Chase fired up their laboratory blender and demonstrated, once and for all, that genes are made of DNA. The researchers were studying T2—a virus that infects and eventually destroys the bacterium *E. coli*. These bacteria-killing viruses behave like tiny molecular syringes: they inject their genetic material into the bacterial host cell, while the empty virus heads remain attached outside (**Figure 5–32A**). Once inside the bacterial cell, the viral genes direct the formation of new virus particles. In less than an hour, the infected cells explode, spewing thousands of new viruses into the medium. These then infect neighboring bacteria, and the process begins again.

The beauty of T2 is that these viruses contain only two kinds of molecules: DNA and protein. So the genetic material had to be one or the other. But which? The experiment was fairly straightforward. Because the viral genes enter the bacterial cell, while the rest of the virus particle remains outside, the researchers decided to radioactively label the protein in one batch of virus and the DNA in another. Then, all they had to do was follow the radioactivity to see whether viral DNA or viral protein wound up inside the bacteria. To do this, Hershey and Chase incubated their radiolabeled viruses with *E. coli*; after allowing a few minutes for infection to take place, they poured the mix into a Waring blender and hit "puree." The blender's spinning blades sheared the empty virus heads from the surfaces of the bacterial cells. The researchers then centrifuged the sample to separate the heavier, infected bacteria, which formed a pellet at the bottom of the centrifuge tube, from the empty viral coats, which remained in suspension (**Figure 5–32B**).

As you have probably guessed, Hershey and Chase found that the radioactive DNA entered the bacterial cells, while the radioactive proteins remained outside with the empty virus heads. They found that the radioactive DNA was also incorporated into the next generation of virus particles.

This experiment demonstrated conclusively that viral DNA enters bacterial host cells, whereas viral protein does not. Thus, the genetic material in this virus had to be made of DNA. Together with the studies done by Avery, MacLeod, and McCarty, this evidence clinched the case for DNA as the agent of heredity.

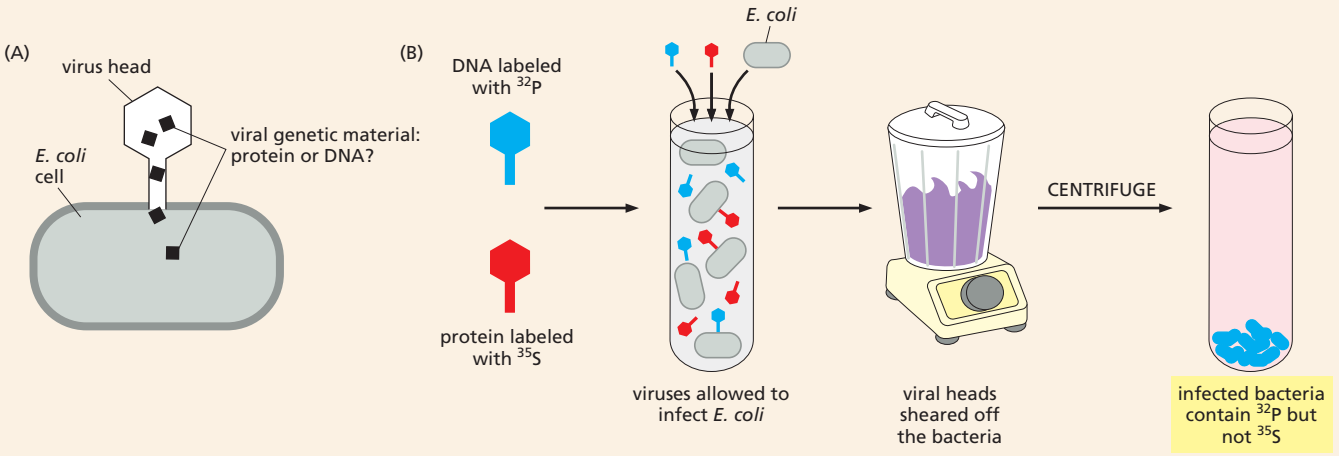


Figure 5–32 Hershey and Chase showed definitively that genes are made of DNA. (A) The researchers worked with T2 viruses, which are made entirely of protein and DNA. Each virus acts as a molecular syringe, injecting its genetic material into a bacterium; the empty viral capsule remains attached to the outside of the cell. (B) To determine whether the genetic material of the virus is made of protein or DNA, the researchers labeled the DNA in one batch of viruses with radioactive phosphorous (^{32}P) and the proteins in a second batch of viruses with radioactive sulfur (^{35}S). Because DNA lacks sulfur and the proteins lack phosphorus, these radioactive isotopes allowed the researchers to distinguish these two types of molecules. The radioactively labeled viruses were allowed to infect *E. coli*, and the mixture was then disrupted by brief pulsing in a Waring blender and centrifuged to separate the infected bacteria from the empty viral heads. When the researchers measured the radioactivity, they found that much of the ^{32}P-labeled DNA had entered the bacterial cells, while the vast majority of the ^{35}S-labeled proteins remained in solution with the spent viral particles. Furthermore, the radioactively labeled DNA also made its way into subsequent generations of virus particles, confirming that DNA is the heritable, genetic material.

QUESTIONS

QUESTION 5–5

A. The nucleotide sequence of one DNA strand of a DNA double helix is 5′-GGATTTTTGTCCACAATCA-3′.

What is the sequence of the complementary strand?

B. In the DNA of certain bacterial cells, 13% of the nucleotides contain adenine. What are the percentages of the other nucleotides?

C. How many possible nucleotide sequences are there for a stretch of single-stranded DNA that is *N* nucleotides long?

D. Suppose you had a method of cutting DNA at specific sequences of nucleotides. How many nucleotides long (on average) would such a sequence have to be in order to make just one cut in a bacterial genome of 3×10^6 nucleotide pairs? How would the answer differ for the genome of an animal cell that contains 3×10^9 nucleotide pairs?

QUESTION 5–6

An A-T base pair is stabilized by only two hydrogen bonds. Hydrogen-bonding schemes of very similar strengths can also be drawn between other base combinations that normally do not occur in DNA molecules, such as the A-C and the A-G pairs shown in Figure Q5–6. What would happen if these pairs formed during DNA replication and the inappropriate bases were incorporated? Discuss why this does not often happen. (Hint: see Figure 5–4.)

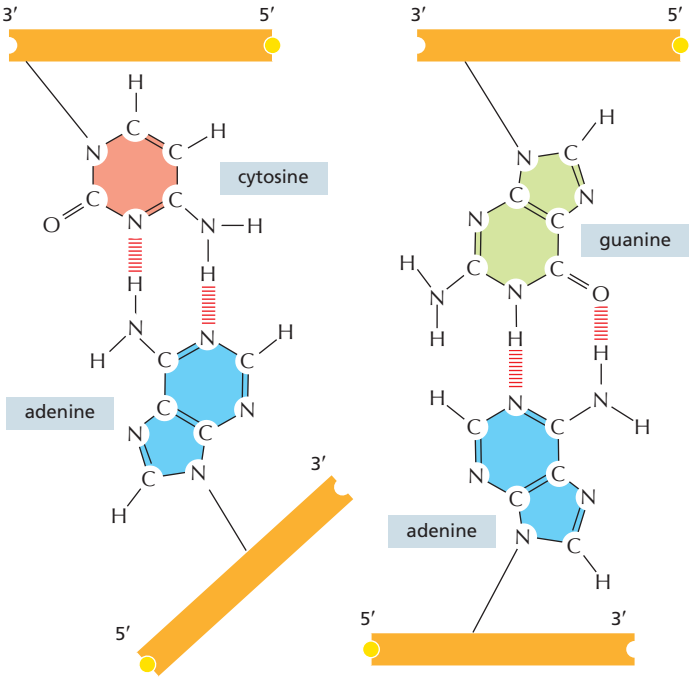


Figure Q5–6

QUESTION 5–7

A. A macromolecule isolated from an extraterrestrial source superficially resembles DNA, but closer analysis reveals that the bases have quite different structures (Figure Q5–7). Bases V, W, X, and Y have replaced bases A, T, G, and C. Look at these structures closely. Could these DNA-like molecules have been derived from a living organism that uses principles of genetic inheritance similar to those used by organisms on Earth?

B. Simply judged by their potential for hydrogen-bonding, could any of these extraterrestrial bases replace terrestrial A, T, G, or C in terrestrial DNA? Explain your answer.

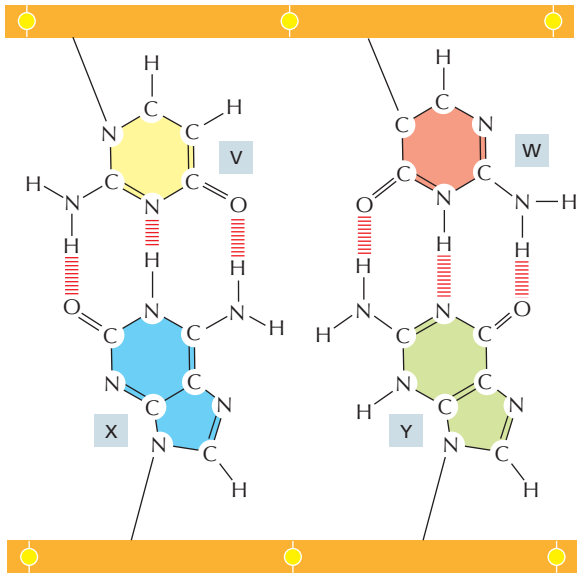


Figure Q5–7

QUESTION 5–8

The two strands of a DNA double helix can be separated by heating. If you raised the temperature of a solution containing the following three DNA molecules, in what order do you suppose they would "melt"? Explain your answer.

A. 5′-GCGGGCCAGCCCGAGTGGGTAGCCCAGG-3′
   3′-CGCCCGGTCGGGCTCACCCATCGGGTCC-5′

B. 5′-ATTATAAAATATTTAGATACTATATTTACAA-3′
   3′-TAATATTTTATAAATCTATGATATAAATGTT-5′

C. 5′-AGAGCTAGATCGAT-3′
   3′-TCTCGATCTAGCTA-5′

QUESTION 5–9

The total length of DNA in one copy of the human genome is about 1 m, and the diameter of the double helix is about 2 nm. Nucleotides in a DNA double helix are stacked (see Figure 5–4B) at an interval of 0.34 nm. If the DNA were enlarged so that its diameter equaled that of an electrical extension cord (5 mm), how long would the extension cord be from one end to the other (assuming that it is completely stretched out)? How close would the bases be to each other? How long would a gene of 1000 nucleotide pairs be?

QUESTION 5–10

A compact disc (CD) stores about 4.8×10^9 bits of information in a 96 cm^2 area. This information is stored as a binary code—that is, every bit is either a 0 or a 1.

A. How many bits would it take to specify each nucleotide pair in a DNA sequence?

B. How many CDs would it take to store the information contained in the human genome?

QUESTION 5–11

Which of the following statements are correct? Explain your answers.

A. Each eukaryotic chromosome must contain the following DNA sequence elements: multiple origins of replication, two telomeres, and one centromere.

B. Nucleosome core particles are 30 nm in diameter.

QUESTION 5–12

Define the following terms and their relationships to one another:

A. Interphase chromosome

B. Mitotic chromosome

C. Chromatin

D. Heterochromatin

E. Histones

F. Nucleosome

QUESTION 5–13

Carefully consider the result shown in Figure Q5–13. Each of the two colonies shown on the *left* is a clump of approximately 100,000 yeast cells that has grown up from a single cell, which is now somewhere in the middle of the colony. The two yeast colonies are genetically different, as shown by the chromosomal maps on the right. The yeast *Ade2* gene encodes one of the enzymes required for adenine biosynthesis, and the absence of the *Ade2* gene product leads to the accumulation of a red pigment. At its normal chromosome location, *Ade2* is expressed in all cells. When it is positioned near the telomere, which is highly condensed, *Ade2* is no longer expressed. How do you think the white sectors arise? What can you conclude about the propagation of the transcriptional state of the *Ade2* gene from mother to daughter cells?

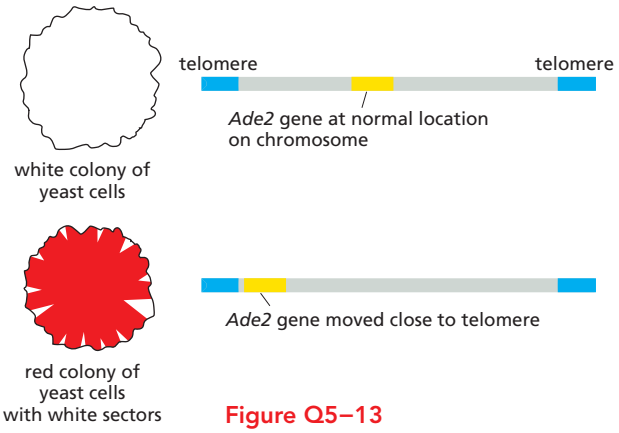


Figure Q5–13

QUESTION 5–14

The two electron micrographs in Figure Q5–14 show nuclei of two different cell types. Can you tell from these pictures which of the two cells is transcribing more of its genes? Explain how you arrived at your answer. (Micrographs courtesy of Don W. Fawcett.)

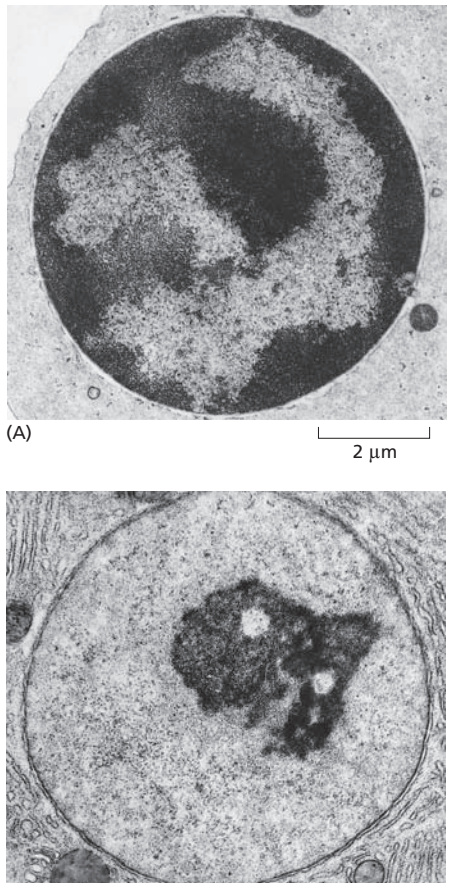


Figure Q5–14

QUESTION 5–15

DNA forms a right-handed helix. Pick out the right-handed helix from those shown in Figure Q5–15.

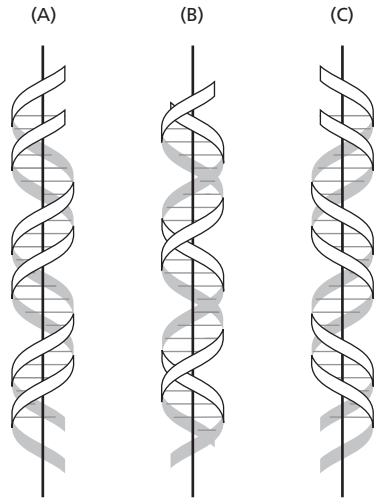


Figure Q5–15

QUESTION 5–16

A single nucleosome core particle is 11 nm in diameter and contains 147 base pairs (bp) of DNA (the DNA double helix measures 0.34 nm/bp). What packing ratio (ratio of DNA length to nucleosome diameter) has been achieved by wrapping DNA around the histone octamer? Assuming that there are an additional 54 bp of extended DNA in the linker between nucleosomes, how condensed is "beads-on-a-string" DNA relative to fully extended DNA? What fraction of the 10,000-fold condensation that occurs at mitosis does this first level of packing represent?