BASIC GENETIC MECHANISMS

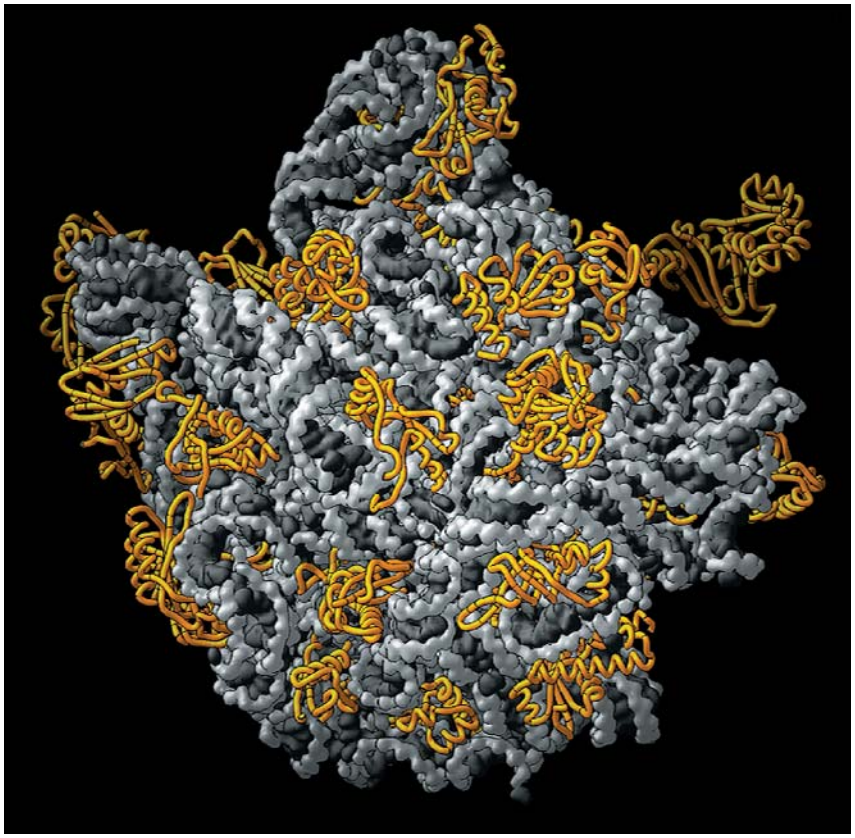
PART II

4 DNA AND CHROMOSOMES

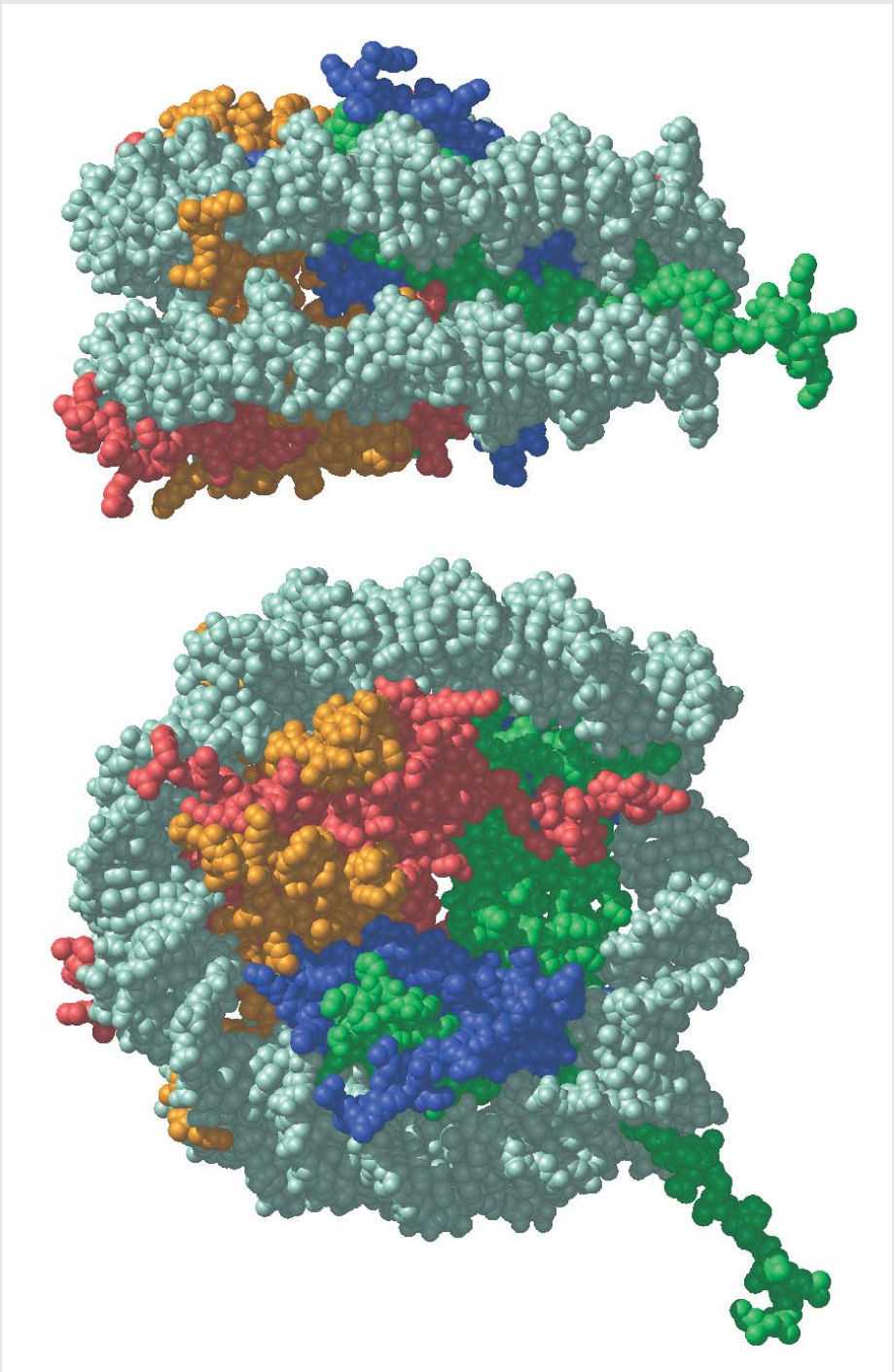
5 DNA REPLICATION, REPAIR, AND RECOMBINATION

6 HOW CELLS READ THE GENOME: FROM DNA TO PROTEIN

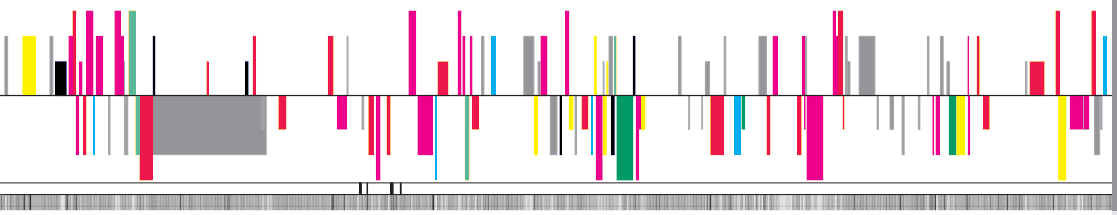
7 CONTROL OF GENE EXPRESSION



The ribosome. The large subunit of the bacterial ribosome is shown here with the ribosomal RNA shown in *gray* and the ribosomal proteins in *gold*. The ribosome is the complex catalytic machine at the heart of protein synthesis. (From N. Ban et al., *Science* 289:905–920, 2000. © AAAS.)



The nucleosome. The basic structural unit of all eucaryotic chromosomes is the nucleosome. The DNA double helix *(gray)* is wrapped around a core particle of histone proteins (colored) to create the nucleosome. Nucleosomes are spaced roughly 200 nucleotide pairs apart along the chromosomal DNA. (Reprinted by permission from K. Luger et al., *Nature* 389:251–260, 1997. © Macmillan Magazines Ltd.)



4

DNA AND CHROMOSOMES

THE STRUCTURE AND FUNCTION OF DNA

CHROMOSOMAL DNA AND ITS PACKAGING IN THE CHROMATIN FIBER

THE GLOBAL STRUCTURE OF CHROMOSOMES

Life depends on the ability of cells to store, retrieve, and translate the genetic instructions required to make and maintain a living organism. This *hereditary* information is passed on from a cell to its daughter cells at cell division, and from one generation of an organism to the next through the organism's reproductive cells. These instructions are stored within every living cell as its **genes**, the information-containing elements that determine the characteristics of a species as a whole and of the individuals within it.

As soon as genetics emerged as a science at the beginning of the twentieth century, scientists became intrigued by the chemical structure of genes. The information in genes is copied and transmitted from cell to daughter cell millions of times during the life of a multicellular organism, and it survives the process essentially unchanged. What form 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 researchers discovered, from studies in simple fungi, that genetic information consists primarily of instructions for making proteins. Proteins are the macromolecules that perform most cellular functions: they serve as building blocks for cellular structures and form the enzymes that catalyze all of the cell's chemical reactions (Chapter 3), they regulate gene expression (Chapter 7), and they enable cells to move (Chapter 16) and to communicate with each other (Chapter 15). The properties and functions of a cell are determined almost entirely by the proteins it is able to make. 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 identification of **deoxyribonucleic acid (DNA)** as the likely carrier of genetic information. But the mechanism whereby the hereditary information is copied for transmission from cell to cell, and how proteins are specified by the instructions in the DNA, remained completely mysterious. Suddenly, in 1953, the mystery was solved when the structure of DNA was determined by James Watson and Francis Crick. As mentioned in Chapter 1, the structure of DNA immediately solved the problem of how the information in this molecule might be copied, or *replicated*. It also provided the first clues as to 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 biological thought that it is difficult to realize what an enormous intellectual gap this discovery filled.

Well before biologists understood the structure of DNA, they had recognized that genes are carried on *chromosomes*, which were discovered in the nineteenth century as threadlike structures in the nucleus of a eucaryotic cell that become visible as the cell begins to divide (Figure 4–1). Later, as biochemical analysis became possible, chromosomes were found to consist of both DNA and protein. We now know that the DNA carries the hereditary information of the cell (Figure 4–2). In contrast, the protein components of chromosomes function largely to package and control the enormously long DNA molecules so that they fit inside cells and can easily be accessed by them.

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 as the raw material of genes. The genes of every cell on Earth are made of DNA, and insights into the relationship between DNA and genes have come from experiments in a wide variety of organisms. We then consider how genes and other important segments of DNA are arranged on the long molecules of DNA that are present in chromosomes. Finally, we discuss how eucaryotic cells fold these long DNA molecules into compact chromosomes. This packing has to be done in an orderly fashion so that the chromosomes can be replicated and apportioned correctly between the two daughter cells at each cell division. It must also allow access of chromosomal DNA to enzymes that repair it when it is damaged and to the specialized proteins that direct the expression of its many genes.

This is the first of four chapters that deal with basic genetic mechanisms—the ways in which the cell maintains, replicates, expresses, and occasionally improves the genetic information carried in its DNA. In the following chapter (Chapter 5) we discuss the mechanisms by which the cell accurately replicates and repairs DNA; we also describe how DNA sequences can be rearranged through the process of genetic recombination. Gene expression—the process through which the information encoded in DNA is interpreted by the cell to guide the synthesis of proteins—is the main topic of Chapter 6. In Chapter 7, we describe how gene expression is controlled by the cell to ensure that each of the many thousands of proteins encrypted in its DNA is manufactured only at the proper time and place in the life of the cell. Following these four chapters on basic genetic mechanisms, we present an account of the experimental techniques used to study these and other processes that are fundamental to all cells (Chapter 8).

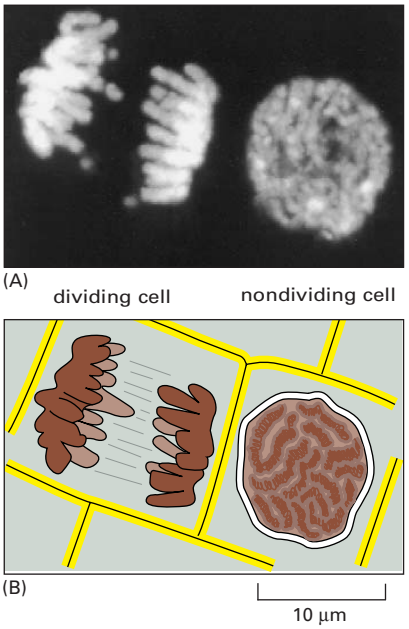


Figure 4–1 Chromosomes in cells. (A) Two adjacent plant cells photographed through a light microscope. The DNA has been stained with a fluorescent dye (DAPI) that binds to it. The DNA is present in chromosomes, which become visible as distinct structures in the light microscope only when they become compact structures in preparation for cell division, as shown on the left. The cell on the right, which is not dividing, contains identical chromosomes, but they cannot be clearly distinguished in the light microscope at this phase in the cell's life cycle, because they are in a more extended conformation. (B) Schematic diagram of the outlines of the two cells along with their chromosomes. (A, courtesy of Peter Shaw.)

THE STRUCTURE AND FUNCTION OF DNA

Biologists in the 1940s had difficulty in accepting DNA as the genetic material because of the apparent simplicity of its chemistry. DNA was known to be a long polymer composed of only four types of subunits, which resemble one another chemically. Early in the 1950s, DNA was first examined by x-ray diffraction analysis, a technique for determining the three-dimensional atomic structure of a molecule (discussed in Chapter 8). The early x-ray diffraction results indicated that DNA was composed of two strands of the polymer wound into a helix. The observation that DNA was double-stranded was of crucial significance and

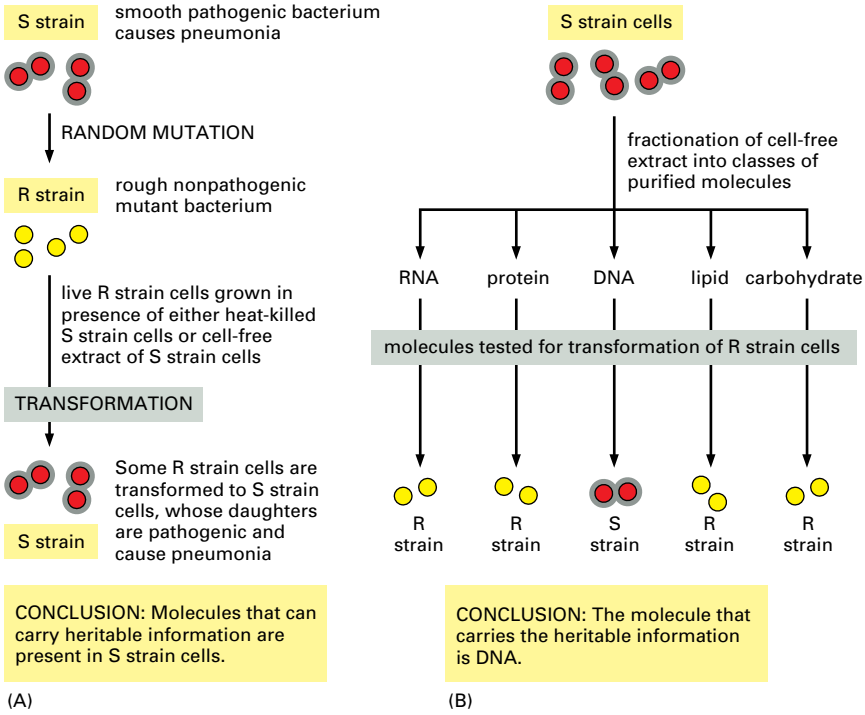


Figure 4–2 Experimental demonstration that DNA is the genetic material. These experiments, carried out in the 1940s, showed that adding purified DNA to a bacterium changed its properties and that this change was faithfully passed on to subsequent generations. Two closely related strains of the bacterium *Streptococcus pneumoniae* differ from each other in both their appearance under the microscope and their pathogenicity. One strain appears smooth (S) and causes death when injected into mice, and the other appears rough (R) and is nonlethal. (A) This experiment shows that a substance present in the S strain can change (or transform) the R strain into the S strain and that this change is inherited by subsequent generations of bacteria. (B) This experiment, in which the R strain has been incubated with various classes of biological molecules obtained from the S strain, identifies the substance as DNA.

provided one of the major clues that led to the Watson–Crick structure of DNA. Only when this model was proposed did DNA's potential for replication and information encoding become apparent. In this section we examine the structure of the DNA molecule and explain in general terms how it is able to store hereditary information.

A DNA Molecule Consists of Two Complementary Chains of Nucleotides

A DNA molecule consists of two long polynucleotide chains composed of four types of nucleotide subunits. Each of these chains is known as a *DNA chain*, or a *DNA strand*. *Hydrogen bonds* between the base portions of the nucleotides hold the two chains together (Figure 4–3). As we saw in Chapter 2 (Panel 2–6, pp. 120-121), nucleotides are composed of a five-carbon sugar to which are attached one or more phosphate groups and a nitrogen-containing base. In the case of the nucleotides in DNA, the sugar is deoxyribose attached to a single phosphate group (hence the name deoxyribonucleic acid), and the base may be either *adenine (A), cytosine (C), guanine (G),* or *thymine (T)*. The nucleotides are covalently linked together in a chain through the sugars and phosphates, which thus form a "backbone" of alternating sugar–phosphate–sugar–phosphate (see Figure 4–3). Because only the base differs in each of the four types of subunits, each polynucleotide chain in DNA is analogous to a necklace (the backbone) strung with four types of 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 and phosphate groups.

The way in which the nucleotide subunits are lined together gives a DNA strand a chemical polarity. If we think of each sugar as a block with a protruding knob (the 5′ phosphate) on one side and a hole (the 3′ hydroxyl) on the other (see Figure 4–3), each completed chain, formed by interlocking knobs with holes, will have all of its subunits lined up in the same orientation. Moreover, the two ends of the chain will be easily distinguishable, as one has a hole (the 3′ hydroxyl) and the other a knob (the 5′ phosphate) at its terminus. This polarity in a DNA chain is indicated by referring to one end as the *3′ end* and the other as the *5′ end*.

The three-dimensional structure of DNA—**the double helix**—arises from the chemical and structural features of its two polynucleotide chains. Because

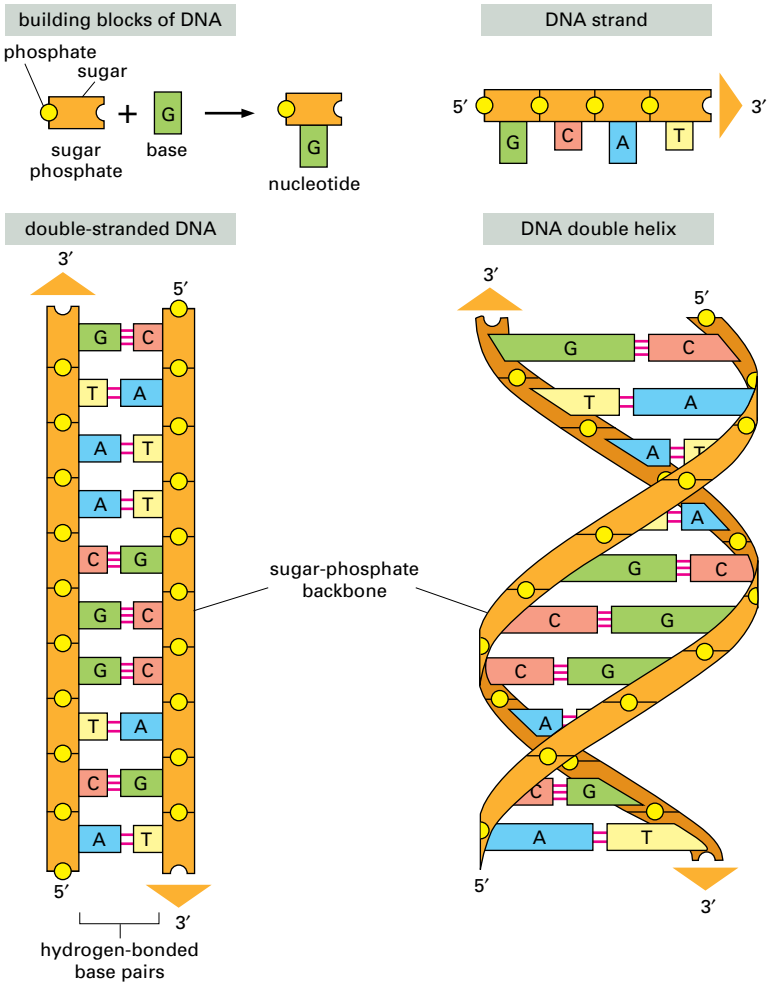


Figure 4–3 DNA and its building blocks. DNA is made of four types of nucleotides, which are linked covalently into a polynucleotide chain (a DNA strand) with a sugar-phosphate backbone from which the bases (A, C, G, and T) extend. A DNA molecule is composed of two DNA strands held together by hydrogen bonds between the paired bases. The *arrowheads* at the ends of the DNA strands indicate the polarities of the two strands, which run antiparallel to each other in the DNA molecule. In the diagram at the bottom left of the figure, the DNA molecule is shown straightened out; in reality, it is twisted into a double helix, as shown on the right. For details, see Figure 4–5.

these two chains are held together by hydrogen bonding between the bases on the different strands, all the bases are on the inside of the double helix, and the sugar-phosphate backbones are on the outside (see Figure 4–3). In each case, a bulkier two-ring base (a purine; see Panel 2–6, pp. 120–121) is paired with a single-ring base (a pyrimidine); A always pairs with T, and G with C (Figure 4–4). This *complementary base-pairing* enables the **base pairs** to be packed in the

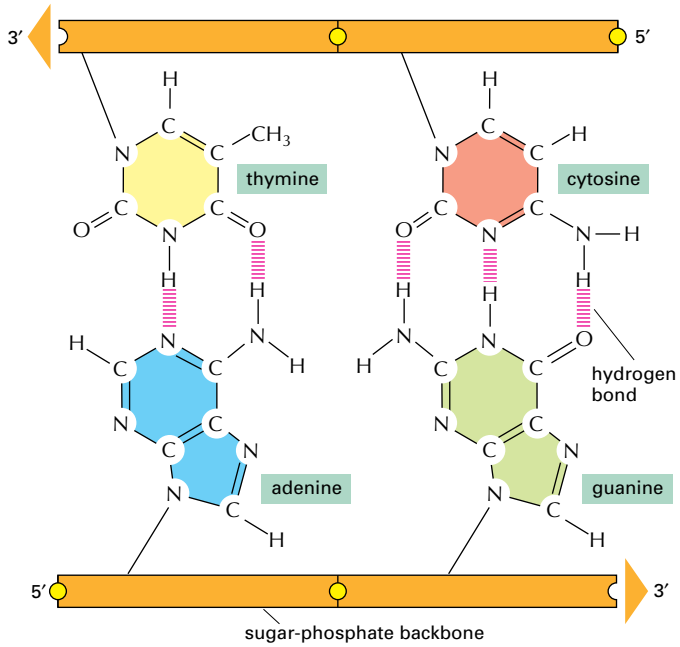


Figure 4–4 Complementary base pairs in the DNA double helix. The shapes and chemical structure of the bases allow hydrogen bonds to form efficiently only between A and T and between G and C, where atoms that are able to form hydrogen bonds (see Panel 2–3, pp. 114–115) can be brought close together without distorting the double helix. As indicated, two hydrogen bonds form between A and T, while three form between G and C. The bases can pair in this way only if the two polynucleotide chains that contain them are antiparallel to each other.

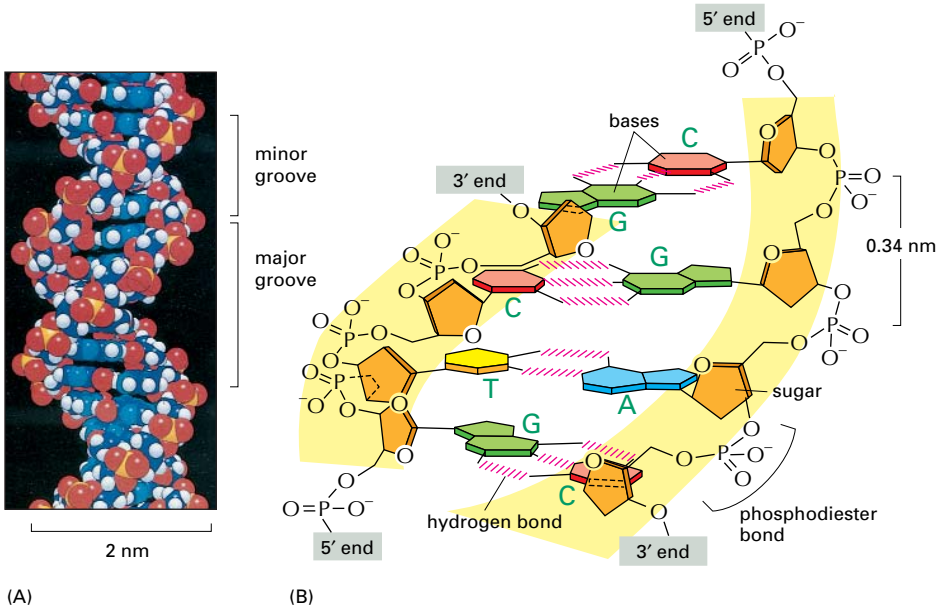


Figure 4–5 The DNA double helix. (A) A space-filling model of 1.5 turns of the DNA double helix. Each turn of DNA is made up of 10.4 nucleotide pairs and the center-to-center distance between adjacent nucleotide pairs is 3.4 nm. The coiling of the two strands around each other creates two grooves in the double helix. As indicated in the figure, the wider groove is called the major groove, and the smaller the minor groove. (B) A short section of the double helix viewed from its side, showing four base pairs. The nucleotides are linked together covalently by phosphodiester bonds through the 3′-hydroxyl (–OH) group of one sugar and the 5′-phosphate (P) of the next. Thus, each polynucleotide strand has 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.

energetically most favorable arrangement in the interior of the double helix. In this arrangement, each base pair is of similar width, thus holding the sugar-phosphate backbones an equal distance apart along the DNA molecule. To maximize the efficiency of base-pair packing, the two sugar-phosphate backbones wind around each other to form a double helix, with one complete turn every ten base pairs (Figure 4–5).

The members of each base pair can fit together within the double helix only if the two strands of the helix are **antiparallel**—that is, only if the polarity of one strand is oriented opposite to that of the other strand (see Figures 4–3 and 4–4). A consequence of these base-pairing requirements is that each strand of a DNA molecule contains a sequence of nucleotides that is exactly **complementary** to the nucleotide sequence of its partner strand.

The Structure of DNA Provides a Mechanism for Heredity

Genes carry biological information that must be copied accurately for transmission to the next generation each time a cell divides to form two daughter cells. Two central biological questions arise from these requirements: how can the information for specifying an organism be carried in chemical form, and how is it accurately copied? The discovery of the structure of the DNA double helix was a landmark in twentieth-century biology because it immediately suggested answers to both questions, thereby resolving at the molecular level the problem of heredity. We discuss briefly the answers to these questions in this section, and we shall examine them in more detail in subsequent chapters.

DNA encodes information through the order, or sequence, of the nucleotides along each strand. Each base—A, C, T, or G—can be considered as a letter in a four-letter alphabet that spells out biological messages in the chemical structure of the DNA. As we saw in Chapter 1, 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 messages, and what do they spell out?

As discussed above, it was known well before the structure of DNA was determined that genes contain the instructions for producing proteins. The DNA messages must therefore somehow encode proteins (Figure 4–6). This relationship immediately makes the problem easier to understand, because of the chemical character of proteins. As discussed in Chapter 3, the properties of a protein, which are responsible for its biological function, are determined by its three-dimensional structure, and its structure is determined in turn by the linear

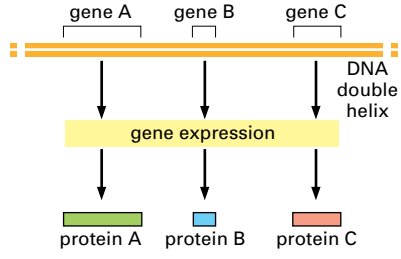


Figure 4–6 The relationship between genetic information carried in DNA and proteins.

Figure 4–7 The nucleotide sequence of the human β-globin gene. This gene carries the information for the amino acid sequence of one of the two types of subunits of the hemoglobin molecule, which carries oxygen in the blood. A different gene, the α-globin gene, carries the information for the other type of hemoglobin subunit (a hemoglobin molecule has four subunits, two of each type). Only one of the two strands of the DNA double helix containing the β-globin gene is shown; the other strand has the exact complementary sequence. By convention, a nucleotide sequence is written from its 5′ end to its 3′ end, and it should be read from left to right in successive lines down the page as though it were normal English text. The DNA sequences highlighted in *yellow* show the three regions of the gene that specify the amino sequence for the β-globin protein. We see in Chapter 6 how the cell connects these three sequences together to synthesize 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

sequence of the amino acids of which it is composed. The linear sequence of nucleotides in a gene must therefore somehow spell out the linear sequence of amino acids in a protein. The exact correspondence between the four-letter nucleotide alphabet of DNA and the twenty-letter amino acid alphabet of proteins—the genetic code—is not obvious from the DNA structure, and it took over a decade after the discovery of the double helix before it was worked out. In Chapter 6 we describe this code in detail in the course of elaborating the process, known as *gene expression*, through which a cell translates the nucleotide sequence of a gene into the amino acid sequence of a protein.

The complete set of information in an organism's DNA is called its **genome**, and it carries the information for all the proteins the organism will ever synthesize. (The term genome is also used to describe the DNA that carries this information.) The amount of information contained in genomes is staggering: for example, a typical human cell contains 2 meters of DNA. Written out in the four-letter nucleotide alphabet, the nucleotide sequence of a very small human gene occupies a quarter of a page of text (Figure 4–7), while the complete sequence of nucleotides in the human genome would fill more than a thousand books the size of this one. In addition to other critical information, it carries the instructions for about 30,000 distinct proteins.

At each cell division, the cell must copy its genome to pass it to both daughter cells. The discovery of the structure of DNA also revealed the principle that makes this copying possible: because each strand of DNA contains a sequence of nucleotides that is exactly complementary to the nucleotide sequence of its partner strand, each strand can act as a **template**, or mold, for the synthesis of a new complementary strand. In other words, if we designate the two DNA strands as S and S′, strand S can serve as a template for making a new strand S′, while strand S′ can serve as a template for making a new strand S (Figure 4–8).

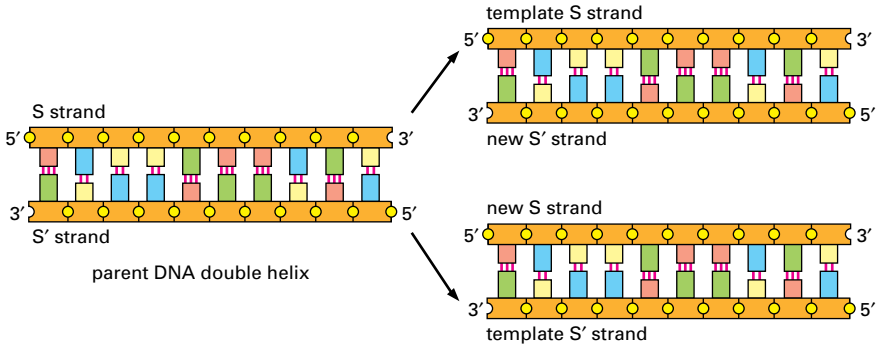


Figure 4–8 DNA as a template for its own duplication. As the nucleotide A successfully pairs only with T, and G with C, each strand of DNA can specify the sequence of nucleotides in its complementary strand. In this way, double-helical DNA can be copied precisely.

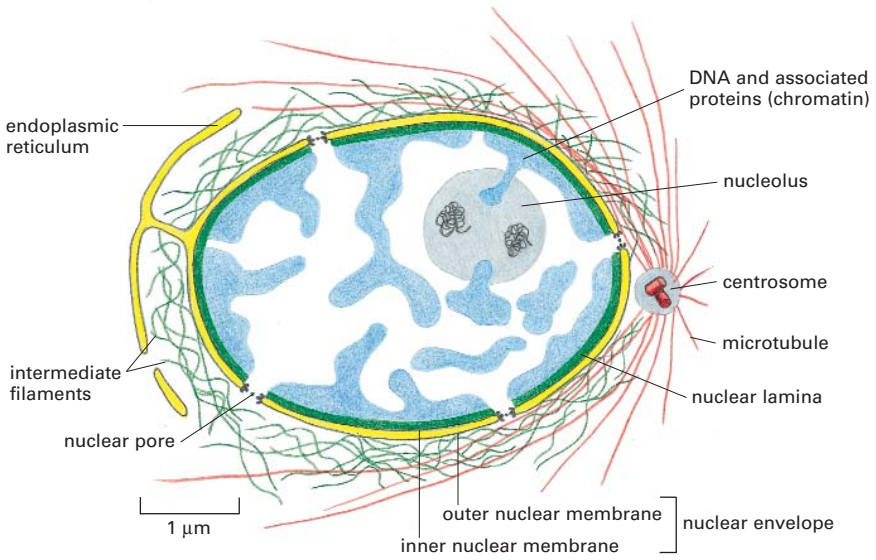


Figure 4–9 A cross-sectional view of a typical cell nucleus. The nuclear envelope consists of two membranes, the outer one being continuous with the endoplasmic reticulum membrane (see also Figure 12–9). The space inside the endoplasmic reticulum (the ER lumen) is colored *yellow;* it is continuous with the space between the two nuclear membranes. The lipid bilayers of the inner and outer nuclear membranes are connected at each nuclear pore. Two networks of intermediate filaments *(green)* provide mechanical support for the nuclear envelope; the intermediate filaments inside the nucleus form a special supporting structure called the nuclear lamina.

Thus, the genetic information in DNA can be accurately copied by the beautifully simple process in which strand S separates from strand S′, and each separated strand then serves as a template for the production of a new complementary partner strand that is identical to its former partner.

The ability of each strand of a DNA molecule to act as a template for producing a complementary strand enables a cell to copy, or *replicate*, its genes before passing them on to its descendants. In the next chapter we describe the elegant machinery the cell uses to perform this enormous task.

In Eucaryotes, DNA is Enclosed in a Cell Nucleus

Nearly all the DNA in a eucaryotic cell is sequestered in a nucleus, which occupies about 10% of the total cell volume. This compartment is delimited by a *nuclear envelope* formed by two concentric lipid bilayer membranes that are punctured at intervals by large nuclear pores, which transport molecules between the nucleus and the cytosol. The nuclear envelope is directly connected to the extensive membranes of the endoplasmic reticulum. It is mechanically supported by two networks of intermediate filaments: one, called the *nuclear lamina*, forms a thin sheetlike meshwork inside the nucleus, just beneath the inner nuclear membrane; the other surrounds the outer nuclear membrane and is less regularly organized (Figure 4–9).

The nuclear envelope allows the many proteins that act on DNA to be concentrated where they are needed in the cell, and, as we see in subsequent chapters, it also keeps nuclear and cytosolic enzymes separate, a feature that is crucial for the proper functioning of eucaryotic cells. Compartmentalization, of which the nucleus is an example, is an important principle of biology; it serves to establish an environment in which biochemical reactions are facilitated by the high concentration of both substrates and the enzymes that act on them.

Summary

Genetic information is carried in the linear sequence of nucleotides in DNA. Each molecule of DNA is a double helix formed from two complementary strands of nucleotides held together by hydrogen bonds between G-C and A-T base pairs. Duplication of the genetic information occurs by the use of one DNA strand as a template for formation of a complementary strand. The genetic information stored in an organism's DNA contains the instructions for all the proteins the organism will ever synthesize. In eucaryotes, DNA is contained in the cell nucleus.

CHROMOSOMAL DNA AND ITS PACKAGING IN THE CHROMATIN FIBER

The most important function of DNA is to carry genes, the information that specifies all the proteins that make up an organism—including information about when, in what types of cells, and in what quantity each protein is to be made. The genomes of eucaryotes are divided up into chromosomes, and in this section we see how genes are typically arranged on each chromosome. In addition, we describe the specialized DNA sequences that allow a chromosome to be accurately duplicated and passed on from one generation to the next.

We also confront the serious challenge of DNA packaging. Each human cell contains approximately 2 meters of DNA if stretched end-to-end; yet the nucleus of a human cell, which contains the DNA, is only about 6 μm in diameter. This is geometrically equivalent to packing 40 km (24 miles) of extremely fine thread into a tennis ball! 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, preventing the DNA from becoming an unmanageable tangle. Amazingly, although the DNA is very tightly folded, it is compacted in a way that allows it to easily become available to the many enzymes in the cell that replicate it, repair it, and use its genes to produce proteins.

Eucaryotic DNA Is Packaged into a Set of Chromosomes

In eucaryotes, the DNA in the nucleus is divided between a set of different **chromosomes**. For example, the human genome—approximately 3.2×10^9 nucleotides—is distributed over 24 different chromosomes. Each chromosome consists of a single, enormously long linear DNA molecule associated with proteins that fold and pack the fine DNA thread into a more compact structure. The complex of DNA and protein is called *chromatin* (from the Greek *chroma*, "colored," because of its staining properties). In addition to the proteins involved in packaging the DNA, chromosomes are also associated with many proteins required for the processes of gene expression, DNA replication, and DNA repair.

Bacteria carry their genes on a single DNA molecule, which is usually circular (see Figure 1–30). This DNA is associated with proteins that package and condense the DNA, but they are different from the proteins that perform these functions in eucaryotes. Although often called the bacterial "chromosome," it does not have the same structure as eucaryotic chromosomes, and less is known about how the bacterial DNA is packaged. Even less is known about how DNA is compacted in archaea. Therefore, our discussion of chromosome structure will focus almost entirely on eucaryotic chromosomes.

With the exception of the germ cells, and a few highly specialized cell types that cannot multiply and lack DNA altogether (for example, red blood cells), each human cell contains two copies of each chromosome, one inherited from the mother and one from the father. The maternal and paternal chromosomes of a pair are called **homologous chromosomes (homologs)**. The only nonhomologous chromosome pairs are the sex chromosomes in males, where a *Y chromosome* is inherited from the father and an *X chromosome* from the mother. Thus, each human cell contains a total of 46 chromosomes—22 pairs common to both males and females, plus two so-called sex chromosomes (X and Y in males, two Xs in females). *DNA hybridization* (described in detail in Chapter 8) can be used to distinguish these human chromosomes by "painting" each one a different color (Figure 4–10). Chromosome painting is typically done at the stage in the cell cycle when chromosomes are especially compacted and easy to visualize (mitosis, see below).

Another more traditional way to distinguish one chromosome from another is to stain them with dyes that produce a striking and reliable pattern of bands along each mitotic chromosome (Figure 4–11). The structural bases for these banding patterns are not well understood, and we return to this issue at the end of the chapter. Nevertheless, the pattern of bands on each type of chromosome is unique, allowing each chromosome to be identified and numbered.

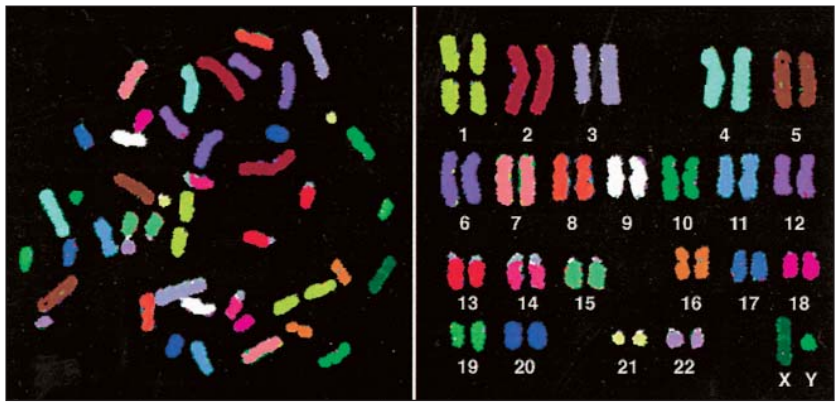


Figure 4–10 Human chromosomes. These chromosomes, from a male, were isolated from a cell undergoing nuclear division (mitosis) and are therefore highly compacted. Each chromosome has been "painted" a different color to permit its unambiguous identification under the light microscope. Chromosome painting is performed by exposing the chromosomes to a collection of human DNA molecules that have been coupled to a combination of fluorescent dyes. For example, DNA molecules derived from chromosome 1 are labeled with one specific dye combination, those from chromosome 2 with another, and so on. Because the labeled DNA can form base pairs, or hybridize, only to the chromosome from which it was derived (discussed in Chapter 8), each chromosome is differently labeled. For such experiments, the chromosomes are subjected to treatments that separate the double-helical DNA into individual strands, designed to permit base-pairing with the single-stranded labeled DNA while keeping the chromosome structure relatively intact. (A) The chromosomes visualized 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. (From E. Schröck et al., *Science* 273:494–497, 1996. © AAAS.)

The display of the 46 human chromosomes at mitosis is called the human **karyotype**. If parts of chromosomes are lost, or switched between chromosomes, these changes can be detected by changes in the banding patterns or by changes in the pattern of chromosome painting (Figure 4–12). Cytogeneticists use these alterations to detect chromosome abnormalities that are associated with inherited defects or with certain types of cancer that arise through the rearrangement of chromosomes in somatic cells.

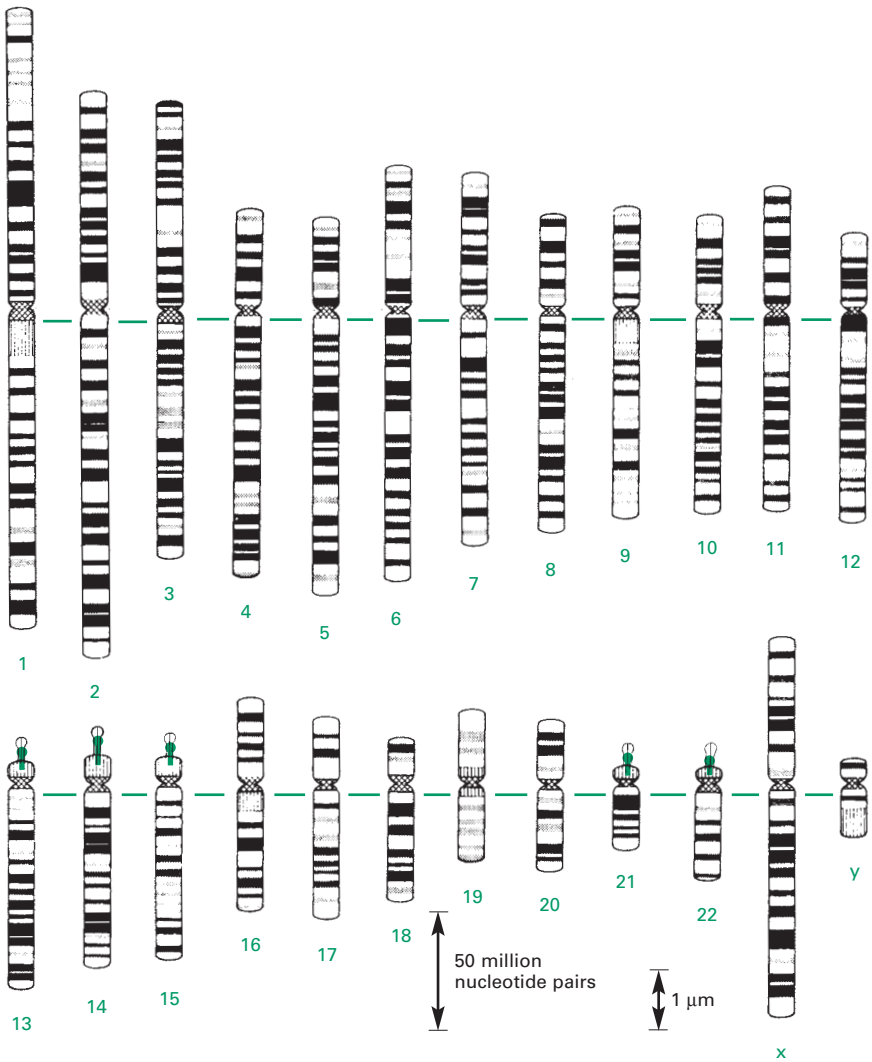


Figure 4–11 The banding patterns of human chromosomes. Chromosomes 1–22 are numbered in approximate order of size. A typical human somatic (non-germ line) cell contains two of each of these chromosomes, plus two sex chromosomes—two X chromosomes in a female, one X and one Y chromosome in a male. The chromosomes used to make these maps were stained at an early stage in mitosis, when the chromosomes are incompletely compacted. The *horizontal green line* represents the position of the centromere (see Figure 4–22), which appears as a constriction on mitotic chromosomes; the knobs on chromosomes 13, 14, 15, 21, and 22 indicate the positions of genes that code for the large ribosomal RNAs (discussed in Chapter 6). These patterns are obtained by staining chromosomes with Giemsa stain, and they can be observed under the light microscope. (Adapted from U. Franke, *Cytogenet. Cell Genet.* 31:24–32, 1981.)

Chromosomes Contain Long Strings of Genes

The most important function of chromosomes is to carry genes—the functional units of heredity. A gene is usually defined as a segment of DNA that contains the instructions for making a particular protein (or a set of closely related proteins). Although this definition holds for the majority of genes, several percent of genes produce an RNA molecule, instead of a protein, as their final product. Like proteins, these RNA molecules perform a diverse set of structural and catalytic functions in the cell, and we discuss them in detail in subsequent chapters.

As might be expected, a correlation exists between the complexity of an organism and the number of genes in its genome (see Table 1–1). For example, total gene numbers range from less than 500 for simple bacteria to about 30,000 for humans. Bacteria and some single-celled eucaryotes have especially compact genomes; the complete nucleotide sequence of their genomes reveals that the DNA molecules that make up their chromosomes are little more than strings of closely packed genes (Figure 4–13; see also Figure 1–30). However, chromosomes from many eucaryotes (including humans) contain, in addition to genes, a large excess of interspersed DNA that does not seem to carry critical information. Sometimes called junk DNA to signify that its usefulness to the cell has not been demonstrated, the particular nucleotide sequence of this DNA may not be important; but the DNA itself, by acting as spacer material, may be crucial for the long-term evolution of the species and for the proper expression of genes. These issues are taken up in detail in Chapter 7.

In general, the more complex the organism, the larger its genome, but because of differences in the amount of excess DNA, the relationship is not systematic (see Figure 1–38). For example, the human genome is 200 times larger than that of the yeast *S. cerevisiae*, but 30 times smaller than that of some plants and amphibians and 200 times smaller than a species of amoeba. Moreover, because of differences in the amount of excess DNA, the genomes of similar organisms (bony fish, for example) can vary several hundredfold in their DNA content, even though they contain roughly the same number of genes. Whatever the excess DNA may do, it seems clear that it is not a great handicap for a higher eucaryotic cell to carry a large amount of it.

The apportionment of the genome over chromosomes also differs from one eucaryotic species to the next. For example, compared with 46 for humans,

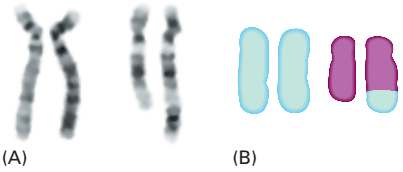


Figure 4–12 An aberrant human chromosome. (A) Two pairs of chromosomes, stained with Giemsa (see Figure 4–11), from a patient with ataxia, a disease characterized by progressive deterioration of motor skills. The patient has a normal pair of chromosome 4s *(left-hand pair),* but one normal chromosome 12 and one aberrant chromosome 12, as seen by its greater length *(right-hand pair).* The additional material contained on the aberrant chromosome 12 was deduced, from its pattern of bands, as a piece of chromosome 4 that had become attached to chromosome 12 through an abnormal recombination event, called a chromosomal translocation. (B) The same two chromosome pairs, "painted" *blue* for chromosome 4 DNA and *purple* for chromosome 12 DNA. The two techniques give rise to the same conclusion regarding the nature of the aberrant chromosome 12, but chromosome painting provides better resolution, and the clear identification of even short pieces of chromosomes that have become translocated. However, Giemsa staining is easier to perform. (From E. Schröck et al., *Science* 273:494–497, 1996. © AAAS.)

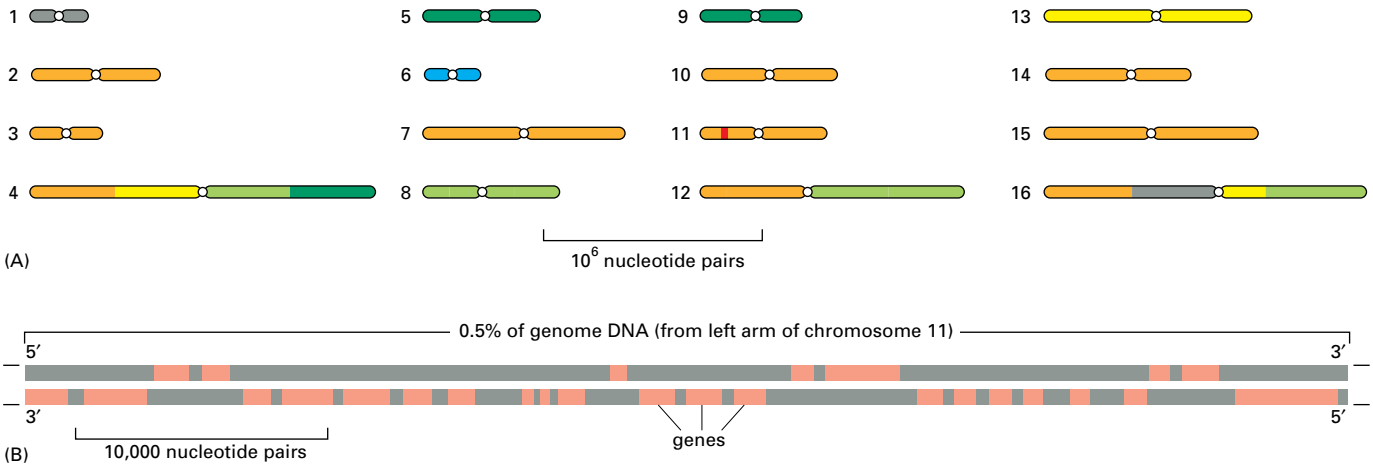


Figure 4–13 The genome of *S. cerevisiae* (budding yeast). (A) The genome is distributed over 16 chromosomes, and its complete nucleotide sequence was determined by a cooperative effort involving scientists working in many different locations, as indicated *(gray,* Canada; *orange,* European Union; *yellow,* United Kingdom; *blue,* Japan; *light green,* St Louis, Missouri; *dark green,* Stanford, California). The constriction present on each chromosome represents the position of its centromere (see Figure 4–22). (B) A small region of chromosome 11, highlighted in *red* in part A, is magnified to show the high density of genes characteristic of this species. As indicated by *orange,* some genes are transcribed from the lower strand (see Figure 1–5), while others are transcribed from the upper strand. There are about 6000 genes in the complete genome, which is 12,147,813 nucleotide pairs long.

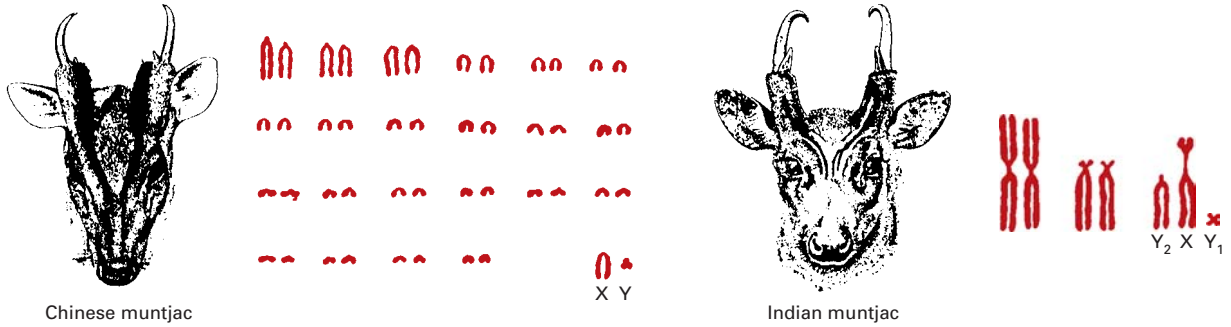


Figure 4–14 Two closely related species of deer with very different chromosome numbers. In the evolution of the Indian muntjac, initially separate chromosomes fused, without having a major effect on the animal. These two species have roughly the same number of genes. (Adapted from M.W. Strickberger, Evolution, 3rd edition, 2000, Sudbury, MA: Jones & Bartlett Publishers.)

somatic cells from a species of small deer contain only 6 chromosomes, while those from a species of carp contain over 100. Even closely related species with similar genome sizes can have very different numbers and sizes of chromosomes (Figure 4–14). Thus, there is no simple relationship between chromosome number, species complexity, and total genome size. Rather, the genomes and chromosomes of modern-day species have each been shaped by a unique history of seemingly random genetic events, acted on by selection pressures.

The Nucleotide Sequence of the Human Genome Shows How Genes Are Arranged in Humans

When the DNA sequence of human chromosome 22, one of the smallest human chromosomes (see Figure 4–11), was completed in 1999, it became possible for the first time to see exactly how genes are arranged along an entire vertebrate chromosome (Figure 4–15 and Table 4–1). With the publication of the "first draft" of the entire human genome in 2001, the genetic landscape of all human chromosomes suddenly came into sharp focus. The sheer quantity of information

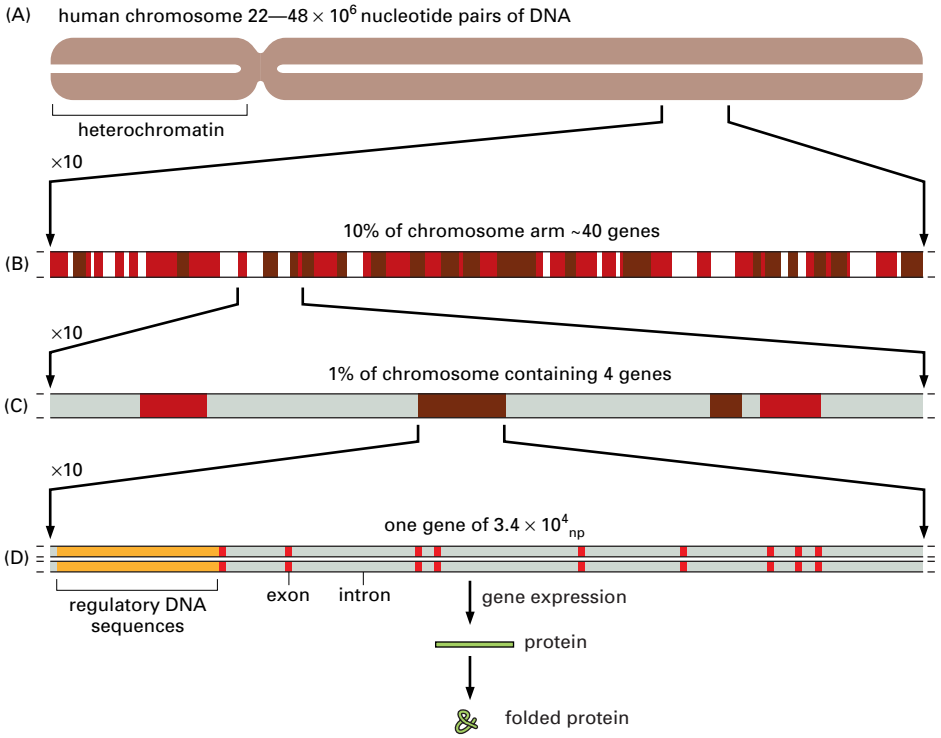


Figure 4–15 The organization of genes on a human chromosome. (A) Chromosome 22, one of the smallest human chromosomes, contains 48×10^6 nucleotide pairs and makes up approximately 1.5% of the entire human genome. Most of the left arm of chromosome 22 consists of short repeated sequences of DNA that are packaged in a particularly compact form of chromatin (heterochromatin), which is discussed later in this chapter. (B) A tenfold expansion of a portion of chromosome 22, with about 40 genes indicated. Those in *dark brown* are known genes and those in *light brown* are predicted genes. (C) An expanded portion of (B) shows the entire length of several genes. (D) The intron–exon arrangement of a typical gene is shown after a further tenfold expansion. Each exon *(red)* codes for a portion of the protein, while the DNA sequence of the introns *(gray)* is relatively unimportant. The entire human genome (3.2×10^9 nucleotide pairs) is distributed over 22 autosomes and 2 sex chromosomes (see Figures 4–10 and 4–11). The term *human genome sequence* refers to the complete nucleotide sequence of DNA in these 24 chromosomes. Being diploid, a human somatic cell therefore contains roughly twice this amount of DNA. Humans differ from one another by an average of one nucleotide in every thousand, and a wide variety of humans contributed DNA for the genome sequencing project. The published human genome sequence is therefore a composite of many individual sequences. (Adapted from International Human Genome Sequencing Consortium, *Nature* 409:860–921, 2001.)

provided by the Human Genome Project is unprecedented in biology (Figure 4–16 and Table 4–1); the human genome is 25 times larger than any other genome sequenced so far, and is 8 times as large as the sum of all previously sequenced genomes. At its peak, the Human Genome Project generated raw nucleotide sequences at a rate of 1000 nucleotides per second around the clock. It will be many decades before this information is fully analyzed, but it will continue to stimulate many new experiments and has already affected the content of all the chapters in this book.

Although there are many aspects to analyzing the human genome, here we simply make a few generalizations regarding the arrangement of genes in human chromosomes. The first striking feature of the human genome is how little of it (only a few percent) codes for proteins or structural and catalytic RNAs (Figure 4–17). Much of the remaining chromosomal DNA is made up of short, mobile pieces of DNA that have gradually inserted themselves in the chromosome over evolutionary time. We discuss these *transposable elements* in detail in later chapters.

A second notable feature of the human genome is the large average gene size of 27,000 nucleotide pairs. As discussed above, a typical gene carries in its linear sequence of nucleotides the information for the linear sequence of the amino acids of a protein. Only about 1300 nucleotide pairs are required to encode a protein of average size (about 430 amino acids in humans). Most of the remaining DNA in a gene consists of long stretches of noncoding DNA that interrupt the relatively short segments of DNA that code for protein. The coding sequences are called *exons*; the intervening (noncoding) sequences are called *introns* (see Figure 4–15 and Table 4–1).

The majority of human genes thus consist of a long string of alternating exons and introns, with most of the gene consisting of introns. In contrast, the majority of genes from organisms with compact genomes lack introns. This accounts for the much smaller size of their genes (about one-twentieth that of human genes), as well as for the much higher fraction of coding DNA in their chromosomes. In addition to introns and exons, each gene is associated with *regulatory DNA sequences*, which are responsible for ensuring that the gene is

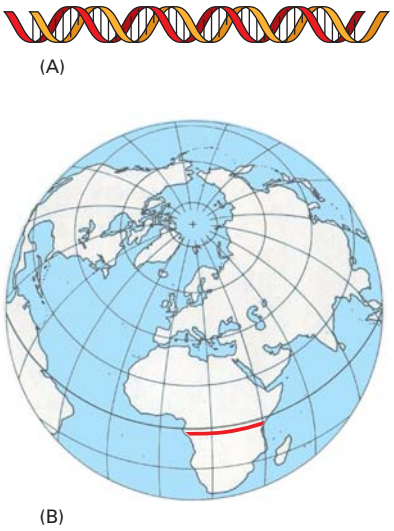


Figure 4–16 Scale of the human genome. If each nucleotide pair is drawn as 1 mm as in (A), then the human genome would extend 3200 km (approximately 2000 miles), far enough to stretch across the center of Africa, the site of our human origins *(red line* in B). At this scale, there would be, on average, a protein-coding gene every 300 m. An average gene would extend for 30 m, but the coding sequences in this gene would add up to only just over a meter.

TABLE 4–1 Vital Statistics of Human Chromosome 22 and the Entire Human Genome

	CHROMOSOME 22	HUMAN GENOME
DNA length	48×10^6 nucleotide pairs*	3.2×10^9
Number of genes	approximately 700	approximately 30,000
Smallest protein-coding gene	1000 nucleotide pairs	not analyzed
Largest gene	583,000 nucleotide pairs	2.4×10^6 nucleotide pairs
Mean gene size	19,000 nucleotide pairs	27,000 nucleotide pairs
Smallest number of exons per gene	1	1
Largest number of exons per gene	54	178
Mean number of exons per gene	5.4	8.8
Smallest exon size	8 nucleotide pairs	not analyzed
Largest exon size	7600 nucleotide pairs	17,106 nucleotide pairs
Mean exon size	266 nucleotide pairs	145 nucleotide pairs
Number of pseudogenes**	more than 134	not analyzed
Percentage of DNA sequence in exons (protein coding sequences)	3%	1.5%
Percentage of DNA in high-copy repetitive elements	42%	approximately 50%
Percentage of total human genome	1.5%	100%

* The nucleotide sequence of 33.8×10^6 nucleotides is known; the rest of the chromosome consists primarily of very short repeated sequences that do not code for proteins or RNA.
** A pseudogene is a nucleotide sequence of DNA closely resembling that of a functional gene, but containing numerous deletion mutations that prevent its proper expression. Most pseudogenes arise from the duplication of a functional gene followed by the accumulation of damaging mutations in one copy.

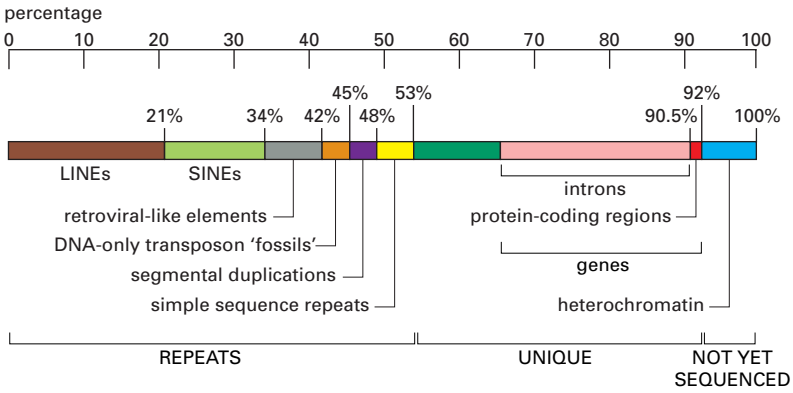


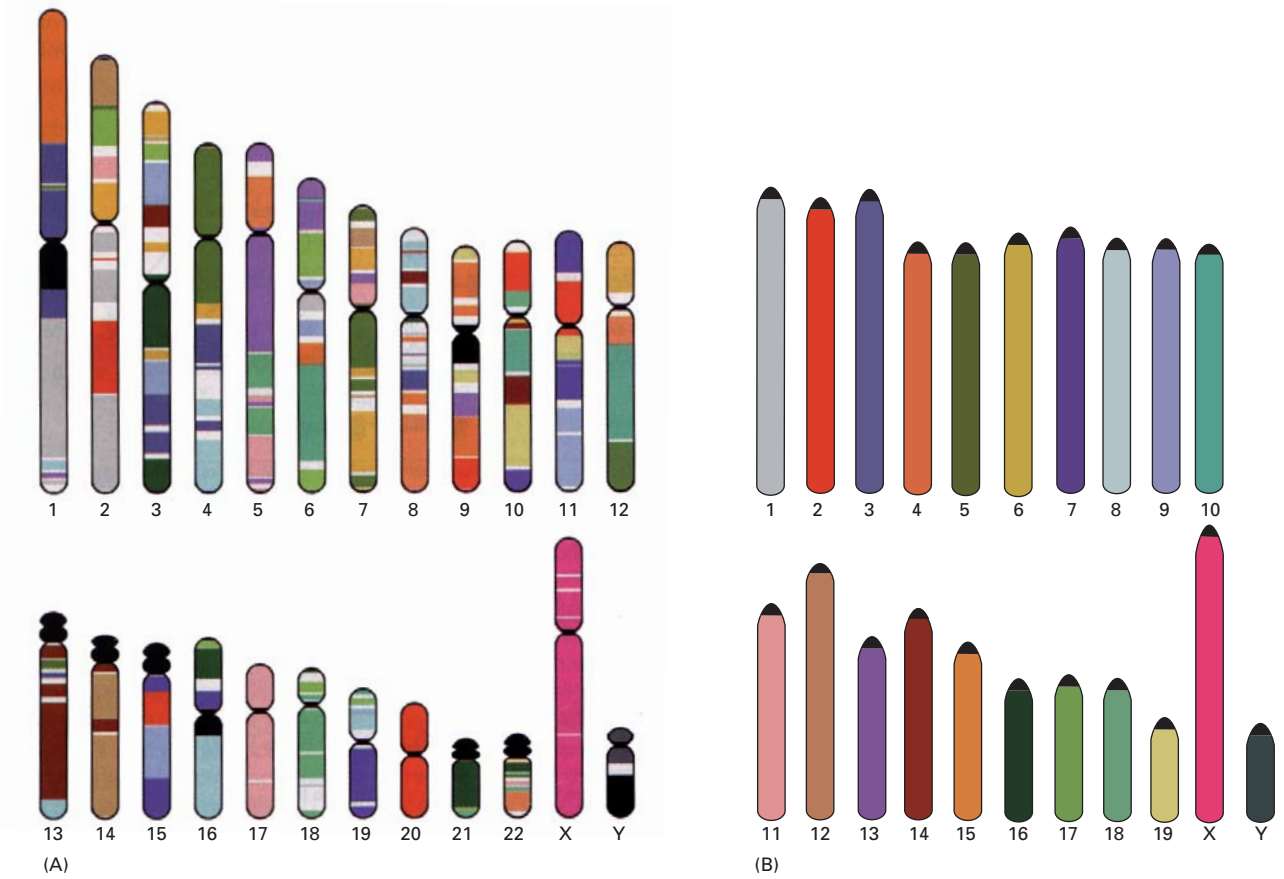
Figure 4–17 Representation of the nucleotide sequence content of the human genome. LINES, SINES, retroviral-like elements, and DNA-only transposons are all mobile genetic elements that have multiplied in our genome by replicating themselves and inserting the new copies in different positions. Mobile genetic elements are discussed in Chapter 5. Simple sequence repeats are short nucleotide sequences (less than 14 nucleotide pairs) that are repeated again and again for long stretches. Segmental duplications are large blocks of the genome (1000–200,000 nucleotide pairs) that are present at two or more locations in the genome. Over half of the unique sequence consists of genes and the remainder is probably regulatory DNA. Most of the DNA present in heterochromatin, a specialized type of chromatin (discussed later in this chapter) that contains relatively few genes, has not yet been sequenced. (Adapted from Unveiling the Human Genome, Supplement to the Wellcome Trust Newsletter. London: Wellcome Trust, February 2001.)

expressed at the proper level and time, and the proper type of cell. In humans, the regulatory sequences for a typical gene are spread out over tens of thousands of nucleotide pairs. As would be expected, these regulatory sequences are more compressed in organisms with compact genomes. We discuss in Chapter 7 how regulatory DNA sequences work.

Finally, the nucleotide sequence of the human genome has revealed that the critical information seems to be in an alarming state of disarray. As one commentator described our genome, "In some ways it may resemble your garage/bedroom/refrigerator/life: highly individualistic, but unkempt; little evidence of organization; much accumulated clutter (referred to by the uninitiated as 'junk'); virtually nothing ever discarded; and the few patently valuable items indiscriminately, apparently carelessly, scattered throughout."

Comparisons Between the DNAs of Related Organisms Distinguish Conserved and Nonconserved Regions of DNA Sequence

A major obstacle in interpreting the nucleotide sequences of human chromosomes is the fact that much of the sequence is probably unimportant. Moreover, the coding regions of the genome (the exons) are typically found in short segments (average size about 145 nucleotide pairs) floating in a sea of DNA whose exact nucleotide sequence is of little consequence. This arrangement makes it very difficult to identify all the exons in a stretch of DNA sequence; even harder is the determination of where a gene begins and ends and how many exons it spans. Accurate gene identification requires approaches that extract information from the inherently low signal-to-noise ratio of the human genome, and we describe some of them in Chapter 8. Here we discuss the most general approach, one that has the potential to identify not only coding sequences but also additional DNA sequences that are important. It is based on the observation that sequences that have a function are conserved during evolution, whereas those without a function are free to mutate randomly. The strategy is therefore to compare the human sequence with that of the corresponding regions of a related genome, such as that of the mouse. Humans and mice are thought to have diverged from a common mammalian ancestor about 100×10^6 years ago, which is long enough for the majority of nucleotides in their genomes to have been changed by random mutational events. Consequently, the only regions that will have remained closely similar in the two genomes are those in which mutations would have impaired function and put the animals carrying them at a disadvantage, resulting in their elimination from the population by natural selection. Such closely similar regions are known as *conserved regions*. In general, conserved regions represent functionally important exons and regulatory sequences. In contrast, *nonconserved regions* represent DNA whose sequence is generally not critical for function. By revealing in this way the results of a very long natural "experiment," comparative DNA sequencing studies highlight the most interesting regions in genomes.



Comparative studies of this kind have revealed not only that mice and humans share most of the same genes, but also that large blocks of the mouse and human genomes contain these genes in the same order, a feature called *conserved synteny* (Figure 4–18). Conserved synteny can also be revealed by chromosome painting, and this technique has been used to reconstruct the evolutionary history of our own chromosomes by comparing them with those from other mammals (Figure 4–19).

Figure 4–18 Conserved synteny between the human and mouse genomes. Regions from different mouse chromosomes (indicated by the colors of each mouse in B) show conserved synteny (gene order) with the indicated regions of the human genome (A). For example the genes present in the upper portion of human chromosome 1 *(orange)* are present in the same order in a portion of mouse chromosome 4. Regions of human chromosomes that are composed primarily of short, repeated sequences are shown in *black*. Mouse centromeres (indicated in *black* in B) are located at the ends of chromosomes; no known genes lie beyond the centromere on any mouse chromosome. For the most part, human centromeres, indicated by constrictions, occupy more internal positions on chromosomes (see Figure 4–11). (Adapted from International Human Genome Sequencing Consortium, *Nature* 409:860–921, 2001.)

Chromosomes Exist in Different States Throughout the Life of a Cell

We have seen how genes are arranged in chromosomes, but to form a functional chromosome, a DNA molecule must be able to do more than simply carry genes: it must be able to replicate, and the replicated copies must be separated and reliably partitioned into daughter cells at each cell division. This process occurs through an ordered series of stages, collectively known as the **cell cycle**. The cell cycle is briefly summarized in Figure 4–20, and discussed in detail in Chapter 17. Only two of the stages of the cycle concern us in this chapter. During *interphase* chromosomes are replicated, and during *mitosis* they become highly condensed and then are separated and distributed to the two daughter nuclei. The highly condensed chromosomes in a dividing cell are known as *mitotic chromosomes*. This is the form in which chromosomes are most easily visualized; in fact, all the images of chromosomes shown so far in the chapter are of chromosomes in mitosis. This condensed state is important in allowing the duplicated chromosomes to be separated by the mitotic spindle during cell division, as discussed in Chapter 18.

During the portions of the cell cycle when the cell is not dividing, the chromosomes are extended and much of their chromatin exists as long, thin tangled

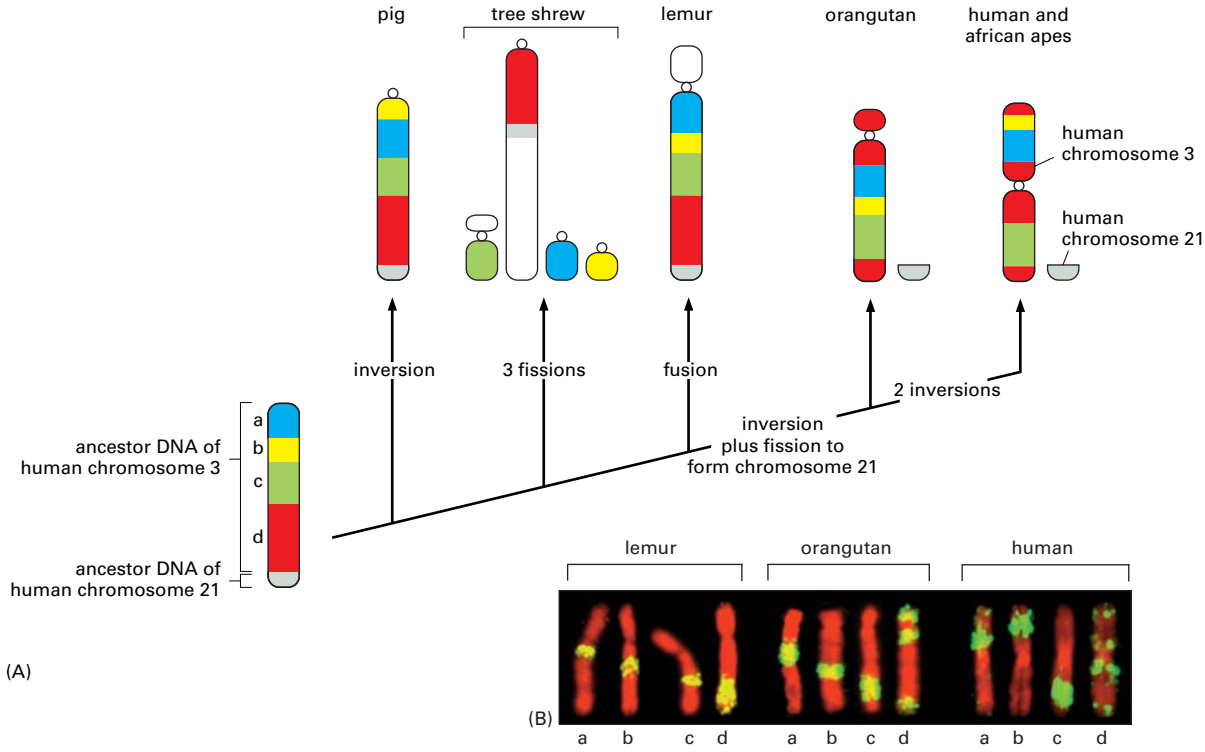


Figure 4–19 A proposed evolutionary history of human chromosome 3 and its relatives in other mammals. (A) At the *lower left* is the order of chromosome 3 segments hypothesized to be present on a chromosome of a mammalian ancestor. Along the *top* are the patterns of chromosome sequences found in the chromosomes of modern mammals. The minimum changes necessary to account for the appearance of the modern chromosomes from the hypothetical ancestor are marked along each branch. In mammals, these types of changes in chromosome organization are thought to occur once every 5–10 × 10^6 years. The *small circles* depicted in the modern chromosomes represent the positions of centromeres. (B) Some of the chromosome painting experiments that led to the diagram in (A). Each image shows the chromosome most closely related to human chromosome 3, painted *green* by hybridization with different segments of DNA, lettered a, b, c, and d along the *bottom* of the figure. These letters correspond to the colored segments of the diagram in (A). (From S. Müller et al., *Proc. Natl. Acad. Sci. USA* 97:206–211, 2000. © National Academy of Sciences.)

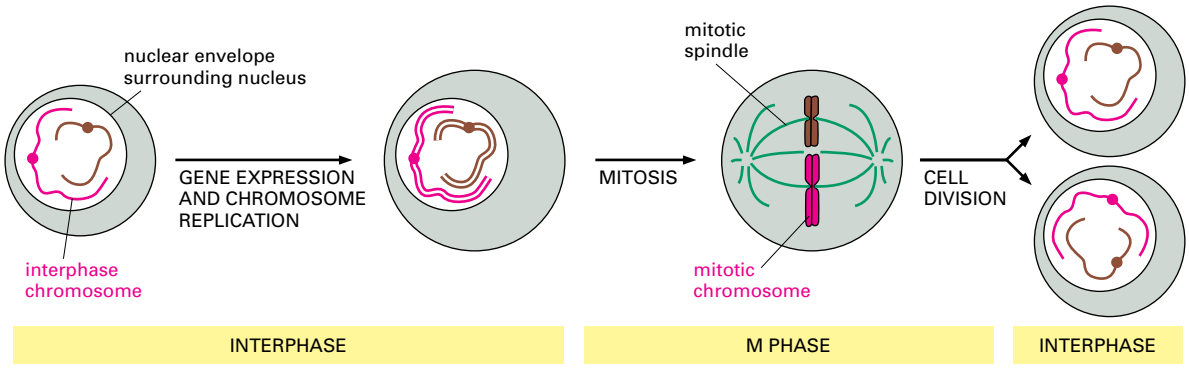


Figure 4–20 A simplified view of the eucaryotic cell cycle. During interphase, the cell is actively expressing its genes and is therefore synthesizing proteins. Also, during interphase and before cell division, the DNA is replicated and the chromosomes are duplicated. Once DNA replication is complete, the cell can enter *M phase,* when mitosis occurs and the nucleus is divided into two daughter nuclei. During this stage, the chromosomes condense, the nuclear envelope breaks down, and the mitotic spindle forms from microtubules and other proteins. The condensed mitotic chromosomes are captured by the mitotic spindle, and one complete set of chromosomes is then pulled to each end of the cell. A nuclear envelope re-forms around each chromosome set, and in the final step of M phase, the cell divides to produce two daughter cells. Most of the time in the cell cycle is spent in interphase; M phase is brief in comparison, occupying only about an hour in many mammalian cells.

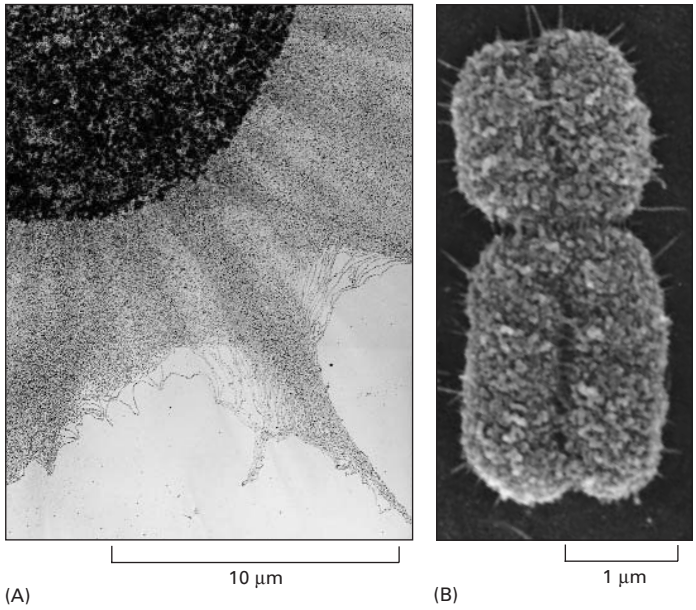


Figure 4–21 A comparison of extended interphase chromatin with the chromatin in a mitotic chromosome. (A) An electron micrograph showing an enormous tangle of chromatin spilling out of a lysed interphase nucleus. (B) A scanning electron micrograph of a mitotic chromosome: a condensed duplicated chromosome in which the two new chromosomes are still linked together (see Figure 4–22). The constricted region indicates the position of the centromere. Note the difference in scales. (A, courtesy of Victoria Foe; B, courtesy of Terry D. Allen.)

threads in the nucleus so that individual chromosomes cannot be easily distinguished (Figure 4–21). We refer to chromosomes in this extended state as *interphase chromosomes*.

Each DNA Molecule That Forms a Linear Chromosome Must Contain a Centromere, Two Telomeres, and Replication Origins

A chromosome operates as a distinct structural unit: for a copy to be passed on to each daughter cell at division, each chromosome must be able to replicate, and the newly replicated copies must subsequently be separated and partitioned correctly into the two daughter cells. These basic functions are controlled by three types of specialized nucleotide sequence in the DNA, each of which binds specific proteins that guide the machinery that replicates and segregates chromosomes (Figure 4–22).

Experiments in yeasts, whose chromosomes are relatively small and easy to manipulate, have identified the minimal DNA sequence elements responsible for each of these functions. One type of nucleotide sequence acts as a DNA **replication origin**, the location at which duplication of the DNA begins. Eucaryotic chromosomes contain many origins of replication to ensure that the entire chromosome can be replicated rapidly, as discussed in detail in Chapter 5.

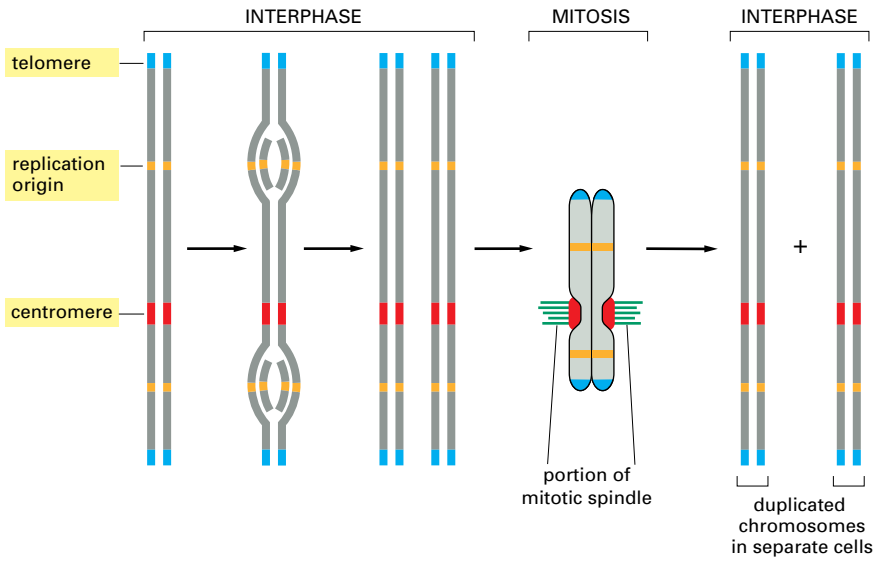


Figure 4–22 The three DNA sequences required to produce a eucaryotic chromosome that can be replicated and then segregated at mitosis. Each chromosome has multiple origins of replication, one centromere, and two telomeres. Shown here is the sequence of events a typical chromosome follows during the cell cycle. The DNA replicates in interphase beginning at the origins of replication and proceeding bidirectionally from the origins across the chromosome. In M phase, the centromere attaches the duplicated chromosomes to the mitotic spindle so that one copy is distributed to each daughter cell during mitosis. The centromere also helps to hold the duplicated chromosomes together until they are ready to be moved apart. The telomeres form special caps at each chromosome end.

After replication, the two daughter chromosomes remain attached to one another and, as the cell cycle proceeds, are condensed further to produce mitotic chromosomes. The presence of a second specialized DNA sequence, called a **centromere**, allows one copy of each duplicated and condensed chromosome to be pulled into each daughter cell when a cell divides. A protein complex called a *kinetochore* forms at the centromere and attaches the duplicated chromosomes to the mitotic spindle, allowing them to be pulled apart (discussed in Chapter 18).

The third specialized DNA sequence forms **telomeres**, the ends of a chromosome. Telomeres contain repeated nucleotide sequences that enable the ends of chromosomes to be efficiently replicated. Telomeres also perform another function: the repeated telomere DNA sequences, together with the regions adjoining them, form structures that protect the end of the chromosome from being recognized by the cell as a broken DNA molecule in need of repair. We discuss this type of repair and the other features of telomeres in Chapter 5.

In yeast cells, the three types of sequences required to propagate a chromosome are relatively short (typically less than 1000 base pairs each) and therefore use only a tiny fraction of the information-carrying capacity of a chromosome. Although telomere sequences are fairly simple and short in all eucaryotes, the DNA sequences that specify centromeres and replication origins in more complex organisms are much longer than their yeast counterparts. For example, experiments suggest that human centromeres may contain up to 100,000 nucleotide pairs. It has been proposed that human centromeres may not even require a stretch of DNA with a defined nucleotide sequence; instead, they may simply create a large, regularly repeating protein–nucleic acid structure. We return to this issue at the end of the chapter when we discuss in more general terms the proteins that, along with DNA, make up chromosomes.

DNA Molecules Are Highly Condensed in Chromosomes

All eucaryotic organisms have elaborate ways of packaging DNA into chromosomes. Recall from earlier in this chapter that human chromosome 22 contains about 48 million nucleotide pairs. Stretched out end to end, its DNA would extend about 1.5 cm. Yet, when it exists as a mitotic chromosome, chromosome 22 measures only about 2 μm in length (see Figures 4–10 and 4–11), giving an end-to-end compaction ratio of nearly 10,000-fold. This remarkable feat of compression is performed by proteins that successively coil and fold the DNA into higher and higher levels of organization. Although less condensed than mitotic chromosomes, the DNA of interphase chromosomes is still tightly packed, with an overall compaction ratio of approximately 1000-fold. In the next sections we discuss the specialized proteins that make the compression possible.

In reading these sections it is important to keep in mind that chromosome structure is dynamic. Not only do chromosomes globally condense in accord with the cell cycle, but different regions of the interphase chromosomes condense and decondense as the cells gain access to specific DNA sequences for gene expression, DNA repair, and replication. The packaging of chromosomes must therefore be accomplished in a way that allows rapid localized, on-demand access to the DNA.

Nucleosomes Are the Basic Unit of Eucaryotic Chromosome Structure

The proteins that bind to the DNA to form eucaryotic chromosomes are traditionally divided into two general classes: the **histones** and the *nonhistone chromosomal proteins*. The complex of both classes of protein with the nuclear DNA of eucaryotic cells is known as **chromatin**. Histones are present in such enormous quantities in the cell (about 60 million molecules of each type per human cell) that their total mass in chromatin is about equal to that of the DNA.

Histones are responsible for the first and most basic level of chromosome organization, the **nucleosome**, which was discovered in 1974. When interphase

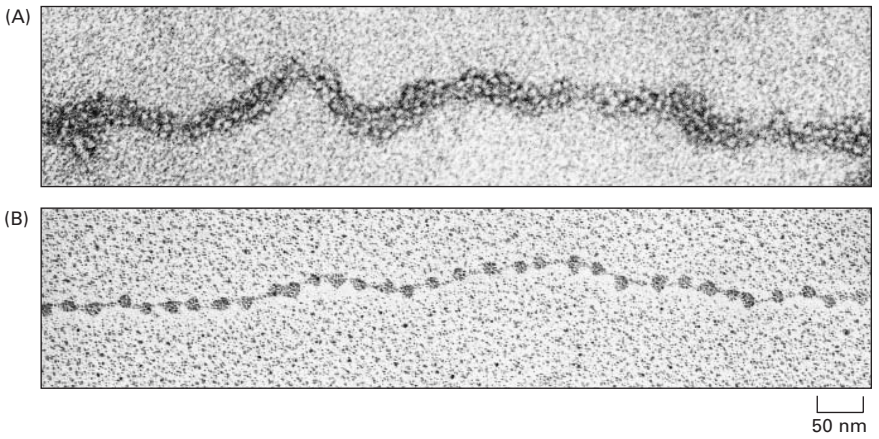


Figure 4–23 Nucleosomes as seen in the electron microscope.
(A) Chromatin isolated directly from an interphase nucleus appears in the electron microscope as a thread 30 nm thick. (B) This electron micrograph shows a length of chromatin that has been experimentally unpacked, or decondensed, after isolation to show the nucleosomes. (A, courtesy of Barbara Hamkalo; B, courtesy of Victoria Foe.)

nuclei are broken open very gently and their contents examined under the electron microscope, most of the chromatin is in the form of a fiber with a diameter of about 30 nm (Figure 4–23A). If this chromatin is subjected to treatments that cause it to unfold partially, it can be seen under the electron microscope as a series of "beads on a string" (Figure 4–23B). The string is DNA, and each bead is a "nucleosome core particle" that consists of DNA wound around a protein core formed from histones. The beads on a string represent the first level of chromosomal DNA packing.

The structural organization of nucleosomes was determined after first isolating them from unfolded chromatin by digestion with particular enzymes (called nucleases) that break down DNA by cutting between the nucleosomes. After digestion for a short period, the exposed DNA between the nucleosome core particles, the *linker DNA*, is degraded. Each individual nucleosome core particle consists of a complex of eight histone proteins—two molecules each of histones H2A, H2B, H3, and H4—and double-stranded DNA that is 146 nucleotide pairs long. The *histone octamer* forms a protein core around which the double-stranded DNA is wound (Figure 4–24).

Each nucleosome core particle is separated from the next by a region of linker DNA, which can vary in length from a few nucleotide pairs up to about 80. (The term *nucleosome* technically refers to a nucleosome core particle plus one of its adjacent DNA linkers, but it is often used synonymously with nucleosome core particle.) On average, therefore, nucleosomes repeat at intervals of about 200 nucleotide pairs. For example, a diploid human cell with 6.4×10^9 nucleotide pairs contains approximately 30 million nucleosomes. The formation of nucleosomes converts a DNA molecule into a chromatin thread about one-third of its initial length, and this provides the first level of DNA packing.

The Structure of the Nucleosome Core Particle Reveals How DNA Is Packaged

The high-resolution structure of a nucleosome core particle, solved in 1997, revealed a disc-shaped histone core around which the DNA was tightly wrapped 1.65 turns in a left-handed coil (Figure 4–25). All four of the histones that make up the core of the nucleosome are relatively small proteins (102–135 amino

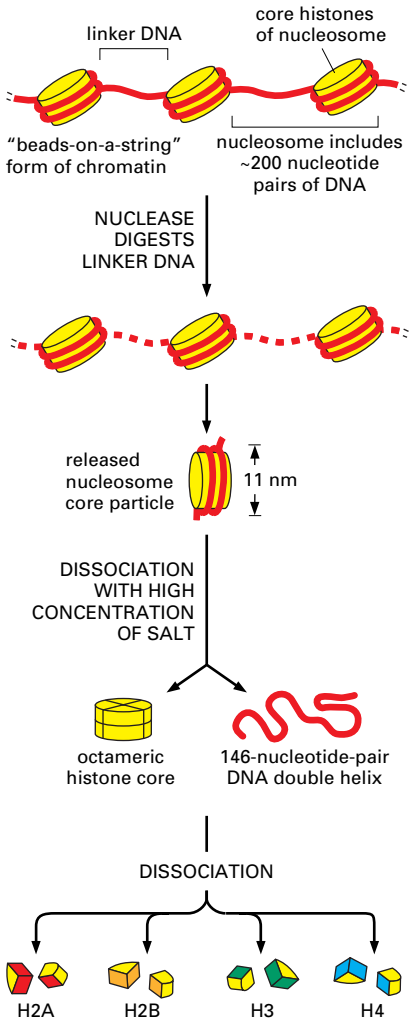


Figure 4–24 Structural organization of the nucleosome.
A nucleosome contains a protein core made of eight histone molecules. As indicated, the nucleosome core particle is released from chromatin by digestion of the linker DNA with a nuclease, an enzyme that breaks down DNA. (The nuclease can degrade the exposed linker DNA but cannot attack the DNA wound tightly around the nucleosome core.) After dissociation of the isolated nucleosome into its protein core and DNA, the length of the DNA that was wound around the core can be determined. This length of 146 nucleotide pairs is sufficient to wrap 1.65 times around the histone core.

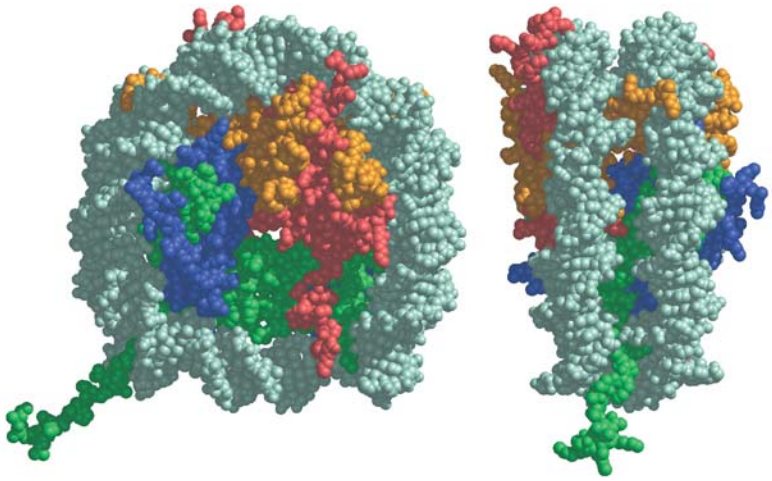


Figure 4–25 The structure of a nucleosome core particle, as determined by x-ray diffraction analyses of crystals. Each histone is colored according to the scheme of Figure 4–26, with the DNA double helix in *light gray.* (Reprinted by permission from K. Luger et al., *Nature* 389:251–260, 1997. © Macmillan Magazines Ltd.)

acids), and they share a structural motif, known as the *histone fold*, formed from three α helices connected by two loops (Figure 4–26). In assembling a nucleosome, the histone folds first bind to each other to form H3–H4 and H2A–H2B dimers, and the H3–H4 dimers combine to form tetramers. An H3–H4 tetramer then further combines with two H2A–H2B dimers to form the compact octamer core, around which the DNA is wound (Figure 4–27).

The interface between DNA and histone is extensive: 142 hydrogen bonds are formed between DNA and the histone core in each nucleosome. Nearly half of these bonds form between the amino acid backbone of the histones and the phosphodiester backbone of the DNA. Numerous hydrophobic interactions and salt linkages also hold DNA and protein together in the nucleosome. For example, all the core histones are rich in lysine and arginine (two amino acids with basic side chains), and their positive charges can effectively neutralize the negatively charged DNA backbone. These numerous interactions explain in part why DNA of virtually any sequence can be bound on a histone octamer core. The path of the DNA around the histone core is not smooth; rather, several kinks are seen in the DNA, as expected from the nonuniform surface of the core.

In addition to its histone fold, each of the core histones has a long N-terminal amino acid "tail", which extends out from the DNA-histone core (see Figure 4–27). These histone tails are subject to several different types of covalent modifications, which control many aspects of chromatin structure. We discuss these issues later in the chapter.

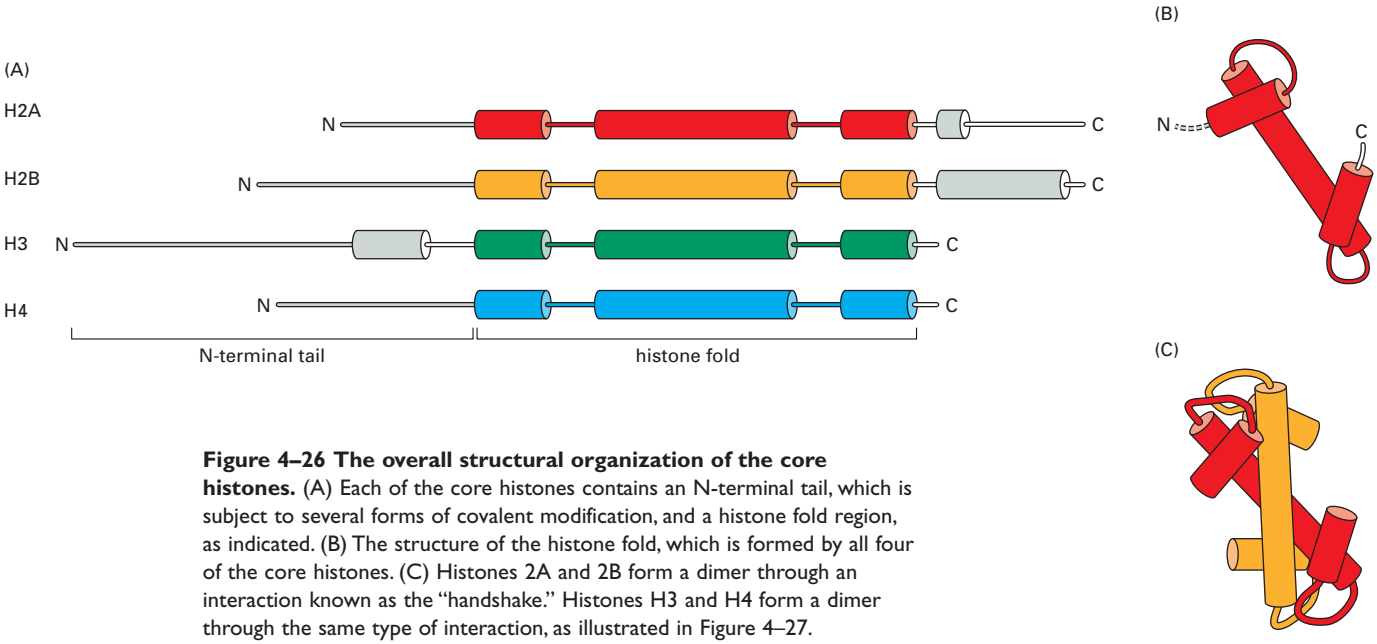


Figure 4–26 The overall structural organization of the core histones. (A) Each of the core histones contains an N-terminal tail, which is subject to several forms of covalent modification, and a histone fold region, as indicated. (B) The structure of the histone fold, which is formed by all four of the core histones. (C) Histones 2A and 2B form a dimer through an interaction known as the "handshake." Histones H3 and H4 form a dimer through the same type of interaction, as illustrated in Figure 4–27.

As might be expected from their fundamental role in DNA packaging, the histones are among the most highly conserved eucaryotic proteins. For example, the amino acid sequence of histone H4 from a pea and a cow differ at only at 2 of the 102 positions. This strong evolutionary conservation suggests that the functions of histones involve nearly all of their amino acids, so that a change in any position is deleterious to the cell. This suggestion has been tested directly in yeast cells, in which it is possible to mutate a given histone gene *in vitro* and introduce it into the yeast genome in place of the normal gene. As might be expected, most changes in histone sequences are lethal; the few that are not lethal cause changes in the normal pattern of gene expression, as well as other abnormalities.

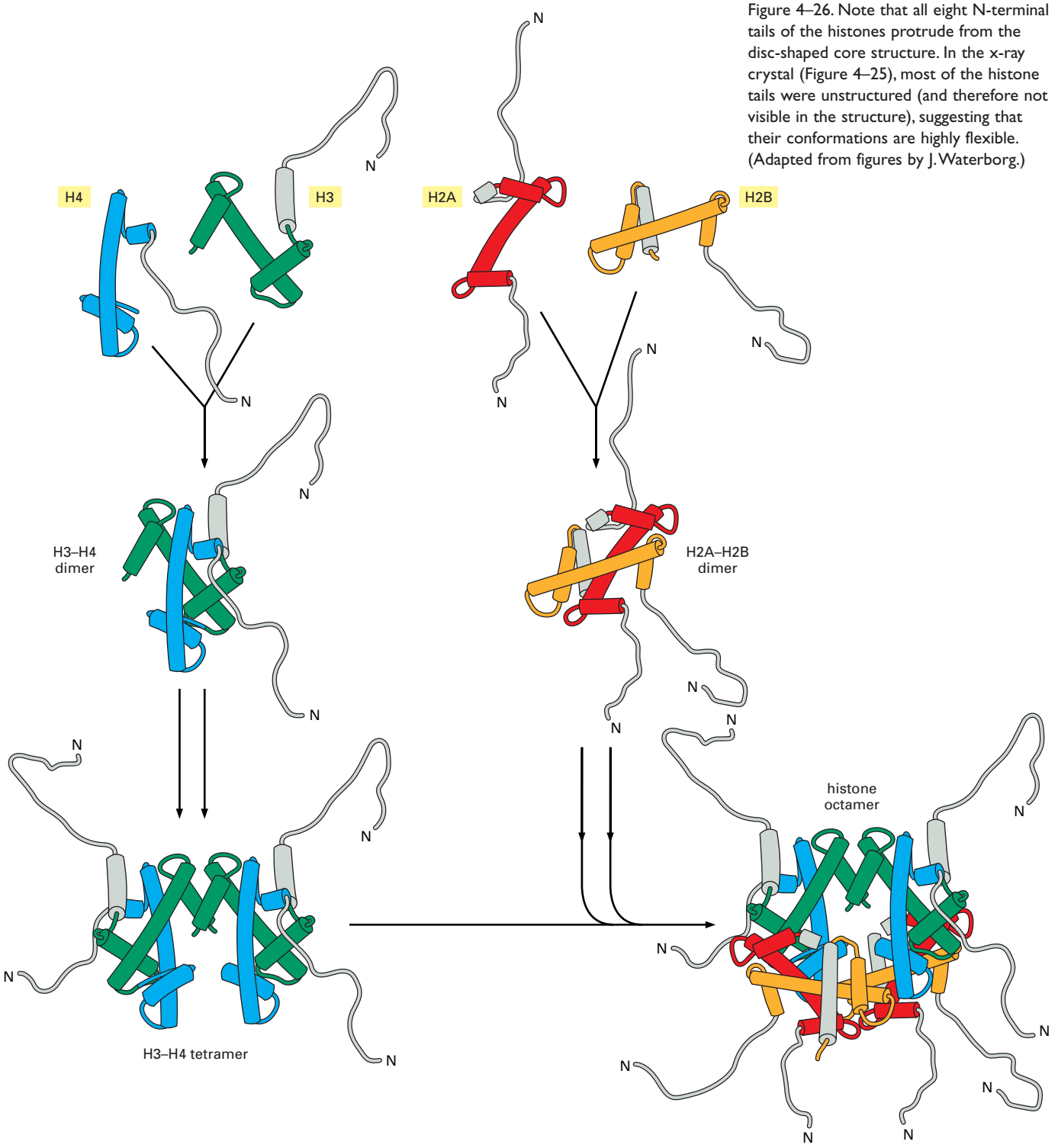


Figure 4–27 The assembly of a histone octamer. The histone H3–H4 dimer and the H2A–H2B dimer are formed from the handshake interaction. An H3–H4 tetramer forms the scaffold of the octamer onto which two H2A–H2B dimers are added, to complete the assembly. The histones are colored as in Figure 4–26. Note that all eight N-terminal tails of the histones protrude from the disc-shaped core structure. In the x-ray crystal (Figure 4–25), most of the histone tails were unstructured (and therefore not visible in the structure), suggesting that their conformations are highly flexible. (Adapted from figures by J. Waterborg.)

Despite the high conservation of the core histones, many eucaryotic organisms also produce specialized variant core histones that differ in amino acid sequence from the main ones. For example, the sea urchin has five histone H2A variants, each of which is expressed at a different time during development. It is thought that nucleosomes that have incorporated these variant histones differ in stability from regular nucleosomes, and they may be particularly well suited for the high rates of DNA transcription and DNA replication that occur during these early stages of development.

The Positioning of Nucleosomes on DNA Is Determined by Both DNA Flexibility and Other DNA-bound Proteins

Although nearly every DNA sequence can, in principle, be folded into a nucleosome, the spacing of nucleosomes in the cell can be irregular. Two main influences determine where nucleosomes form in the DNA. One is the difficulty of bending the DNA double helix into two tight turns around the outside of the histone octamer, a process that requires substantial compression of the minor groove of the DNA helix. Because A-T-rich sequences in the minor groove are easier to compress than G-C-rich sequences, each histone octamer tends to position itself on the DNA so as to maximize A-T-rich minor grooves on the inside of the DNA coil (Figure 4–28). Thus, a segment of DNA that contains short A-T-rich sequences spaced by an integral number of DNA turns is easier to bend around the nucleosome than a segment of DNA lacking this feature. In addition, because the DNA in a nucleosome is kinked in several places, the ability of a given nucleotide sequence to accommodate this deformation can also influence the position of DNA on the nucleosome.

These features of DNA probably explain some striking, but unusual, cases of very precise positioning of nucleosomes along a stretch of DNA. For most of the DNA sequences found in chromosomes, however, there is no strongly preferred nucleosome-binding site; a nucleosome can occupy any one of a number of positions relative to the DNA sequence.

The second, and probably most important, influence on nucleosome positioning is the presence of other tightly bound proteins on the DNA. Some bound proteins favor the formation of a nucleosome adjacent to them. Others create obstacles that force the nucleosomes to assemble at positions between them. Finally, some proteins can bind tightly to DNA even when their DNA-binding site is part of a nucleosome. The exact positions of nucleosomes along a stretch of DNA therefore depend on factors that include the DNA sequence and the presence and nature of other proteins bound to the DNA. Moreover, as we see below, the arrangement of nucleosomes on DNA is highly dynamic, changing rapidly according to the needs of the cell.

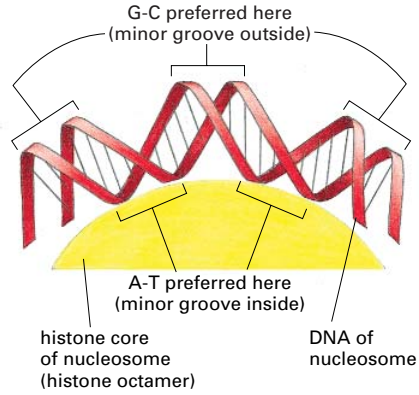


Figure 4–28 The bending of DNA in a nucleosome. The DNA helix makes 1.65 tight turns around the histone octamer. This diagram is drawn approximately to scale, illustrating how the minor groove is compressed on the inside of the turn. Owing to certain structural features of the DNA molecule, A-T base pairs are preferentially accommodated in such a narrow minor groove.

Nucleosomes Are Usually Packed Together into a Compact Chromatin Fiber

Although long strings of nucleosomes form on most chromosomal DNA, chromatin in a living cell probably rarely adopts the extended "beads on a string" form. Instead, the nucleosomes are packed on top of one another, generating regular arrays in which the DNA is even more highly condensed. Thus, when nuclei are very gently lysed onto an electron microscope grid, most of the chromatin is seen to be in the form of a fiber with a diameter of about 30 nm, which is considerably wider than chromatin in the "beads on a string" form (see Figure 4–23).

Several models have been proposed to explain how nucleosomes are packed in the 30-nm chromatin fiber; the one most consistent with the available data is a series of structural variations known collectively as the Zigzag model (Figure 4–29). In reality, the 30-nm structure found in chromosomes is probably a fluid mosaic of the different zigzag variations. We saw earlier that the linker DNA that connects adjacent nucleosomes can vary in length; these differences in linker length probably introduce further local perturbations into the zigzag structure. Finally, the presence of other DNA-binding proteins and DNA sequence that

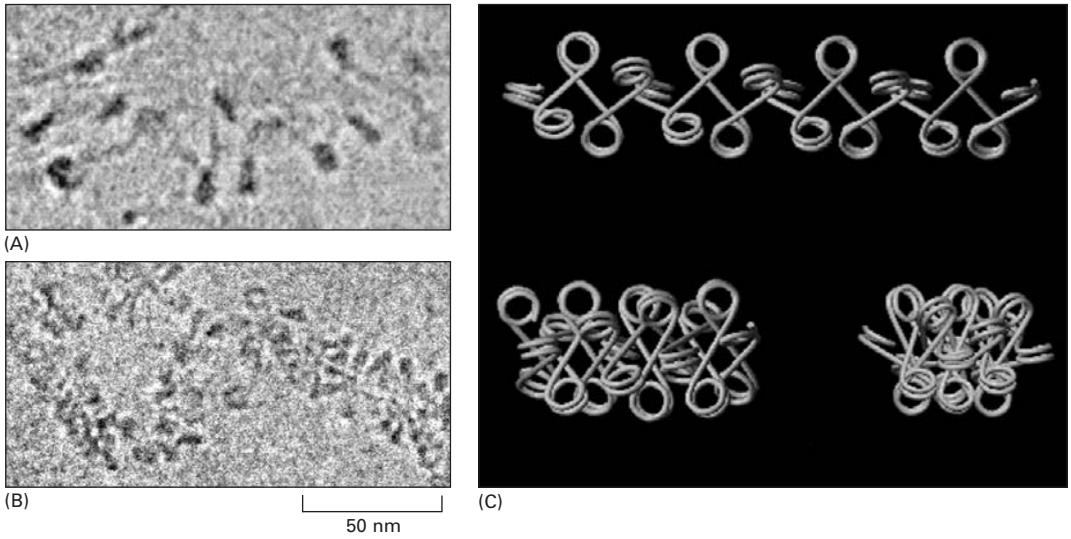


Figure 4–29 Variations on the Zigzag model for the 30-nm chromatin fiber. (A and B) Electron microscopic evidence for the *top* and *bottom-left* model structures depicted in (C). (C) Zigzag variations. An interconversion between these three variations is proposed to occur by an accordion-like expansion and contraction of the fiber length. Differences in the length of the linker between adjacent nucleosome beads can be accommodated by snaking or coiling of the linker DNA, or by small local changes in the width of the fiber. Formation of the 30-nm fiber requires both histone H1 and the core histone tails; for simplicity, neither is shown here, but see Figures 4–30 and 4–32. (From J. Bednar et al., *Proc. Natl. Acad. Sci. USA* 95:14173–14178, 1998. © National Academy of Sciences.)

are difficult to fold into nucleosomes punctuate the 30-nm fiber with irregular features (Figure 4–30).

Several mechanisms probably act together to form the 30-nm fiber from a linear string of nucleosomes. First, an additional histone, called histone H1, is involved in this process. H1 is larger than the core histones and is considerably less well conserved. In fact, the cells of most eucaryotic organisms make several histone H1 proteins of related but quite distinct amino acid sequences. A single histone H1 molecule binds to each nucleosome, contacting both DNA and protein, and changing the path of the DNA as it exits from the nucleosome. Although it is not understood in detail how H1 pulls nucleosomes together into the 30-nm fiber, a change in the exit path in DNA seems crucial for compacting nucleosomal DNA so that it interlocks to form the 30-nm fiber (Figure 4–31).

A second mechanism for forming the 30-nm fiber probably involves the tails of the core histones, which, as we saw above, extend from the nucleosome. It is thought that these tails may help attach one nucleosome to another—thereby allowing a string of them, with the aid of histone H1, to condense into the 30-nm fiber (Figure 4–32).

ATP-driven Chromatin Remodeling Machines Change Nucleosome Structure

For many years biologists thought that, once formed in a particular position on DNA, a nucleosome remained fixed in place because of the tight association

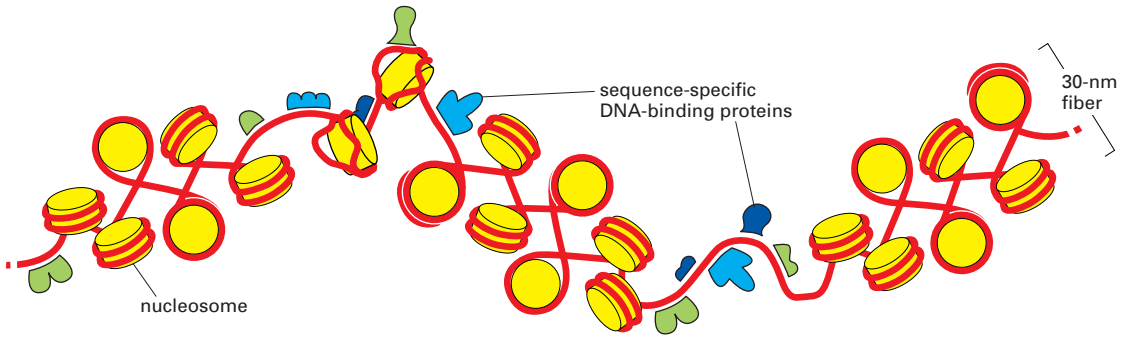


Figure 4–30 Irregularities in the 30-nm fiber. This schematic view of the 30-nm fiber illustrates its interruption by sequence-specific DNA-binding proteins. How these proteins bind tightly to DNA is explained in Chapter 7. The interruptions in the 30-nm fiber may be due to regions of DNA that lack nucleosomes altogether or, more probably, to regions that contain altered or remodeled nucleosomes. Regions of chromatin that are nucleosome free or contain remodeled nucleosome can often be detected experimentally by the unusually high susceptibility of their DNA to digestion by nucleases—as compared with the DNA in nucleosomes.

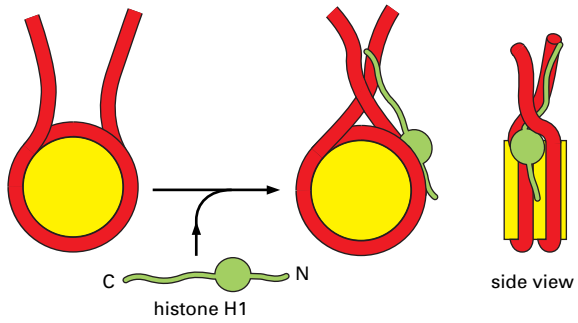


Figure 4–31 A speculative model for how histone H1 could change the path of DNA as it exits from the nucleosome. Histone H1 *(green)* consists of a globular core and two extended tails. Part of the effect of H1 on the compaction of nucleosome organization may result from charge neutralization: like the core histones, H1 is positively charged (especially its C-terminal tail), and this helps to compact the negatively charged DNA. Unlike the core histones, H1 does not seem to be essential for cell viability; in one ciliated protozoan the nucleus expands nearly twofold in the absence of H1, but the cells otherwise appear normal.

between the core histones and DNA. But it has recently been discovered that eucaryotic cells contain *chromatin remodeling complexes*, protein machines that use the energy of ATP hydrolysis to change the structure of nucleosomes temporarily so that DNA becomes less tightly bound to the histone core. The remodeled state may result from movement of the H2A–H2B dimers in the nucleosome core; the H3–H4 tetramer is particularly stable and would be difficult to rearrange (see Figure 4–27).

The remodeling of nucleosome structure has two important consequences. First, it permits ready access to nucleosomal DNA by other proteins in the cell, particularly those involved in gene expression, DNA replication, and repair. Even after the remodeling complex has dissociated, the nucleosome can remain in a "remodeled state" that contains DNA and the full complement of histones—but one in which the DNA–histone contacts have been loosened; only gradually does this remodeled state revert to that of a standard nucleosome. Second, remodeling complexes can catalyze changes in the positions of nucleosomes along DNA (Figure 4–33); some can even transfer a histone core from one DNA molecule to another.

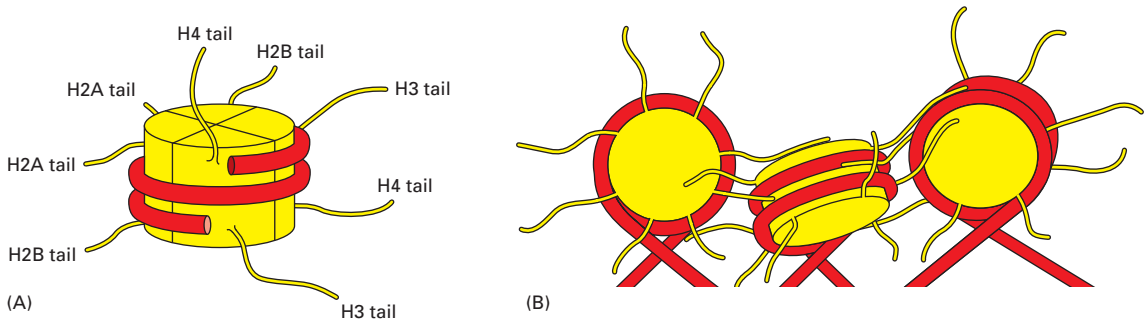
Cells have several different chromatin remodeling complexes that differ subtly in their properties. Most are large protein complexes that can contain more than ten subunits. It is likely that they are used whenever a eucaryotic cell needs direct access to nucleosome DNA for gene expression, DNA replication, or DNA repair. Different remodeling complexes may have features specialized for each of these roles. It is thought that the primary role of some remodeling complexes is to allow access to nucleosomal DNA, whereas that of others is to re-form nucleosomes when access to DNA is no longer required (Figure 4–34).

Chromatin remodeling complexes are carefully controlled by the cell. We shall see in Chapter 7 that, when genes are turned on and off, these complexes can be brought to specific regions of DNA where they act locally to influence chromatin structure. During mitosis, at least some of the chromatin-remodeling complexes are inactivated by phosphorylation. This may help the tightly packaged mitotic chromosomes maintain their structure.

Covalent Modification of the Histone Tails Can Profoundly Affect Chromatin

The N-terminal tails of each of the four core histones are highly conserved in their sequence, and perform crucial functions in regulating chromatin structure.

Figure 4–32 A speculative model for histone tails in the formation of the 30-nm fiber. (A) The approximate exit points of the eight histone tails, four from each histone subunit, that extend from each nucleosome. In the high-resolution structure of the nucleosome (see Figure 4–25), the tails are largely unstructured, suggesting that they are highly flexible. (B) A speculative model showing how the histone tails may help to pack nucleosomes together into the 30-nm fiber. This model is based on (1) experimental evidence that histone tails aid in the formation of the 30-nm fiber, (2) the x-ray crystal structure of the nucleosome, which showed that the tails of one nucleosome contact the histone core of an adjacent nucleosome in the crystal lattice, and (3) evidence that the histone tails interact with DNA.



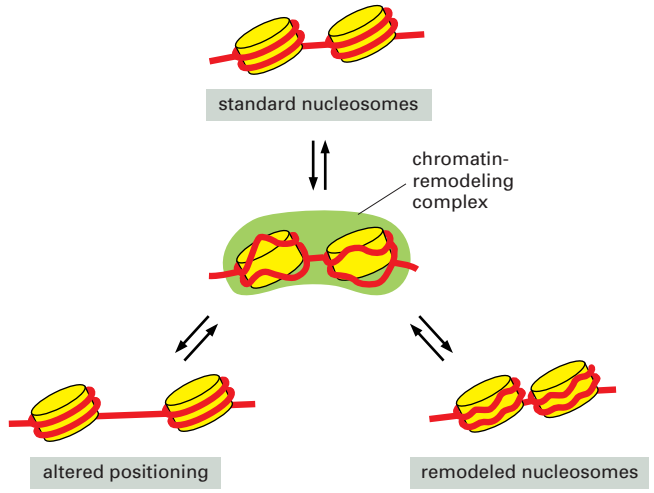


Figure 4–33 Model for the mechanism of some chromatin remodeling complexes. In the absence of remodeling complexes, the interconversion between the three nucleosomal states shown is very slow because of a high activation energy barrier. Using ATP hydrolysis, chromatin-remodeling complexes *(green)* create an activated intermediate (shown in the center of the figure) in which the histone–DNA contacts have been partly disrupted. This activated state can then decay to any one of the three nucleosomal configurations shown. In this way, the remodeling complexes greatly increase the rate of interconversion between different nucleosomal states. The remodeled state, in which the histone–DNA contacts have been loosened, has a higher free energy level than that of standard nucleosomes and will slowly revert to the standard nucleosome conformation, even in the absence of a remodeling complex. Cells have many different chromatin remodeling complexes, and they differ in their detailed biochemical properties; for example, not all can change the position of a nucleosome, but all use the energy of ATP hydrolysis to alter nucleosome structure. (Adapted from R.E. Kingston and G.J. Narlikar, *Genes Dev.* 13:2339–2352, 1999.)

Each tail is subject to several types of covalent modifications, including acetylation of lysines, methylation of lysines, and phosphorylation of serines (Figure 4–35A). Histones are synthesized in the cytosol and then assembled into nucleosomes. Some of the modifications of histone tails occur just after their synthesis, but before their assembly. The modifications that concern us, however, take place once the nucleosome has been assembled. These nucleosome modifications are added and removed by enzymes that reside in the nucleus; for example, acetyl groups are added to the histone tails by histone acetyl transferases (HATs) and taken off by histone deacetylases (HDACs).

The various modifications of the histone tails have several important consequences. Although modifications of the tails have little direct effect on the stability of an individual nucleosome, they seem to affect the stability of the

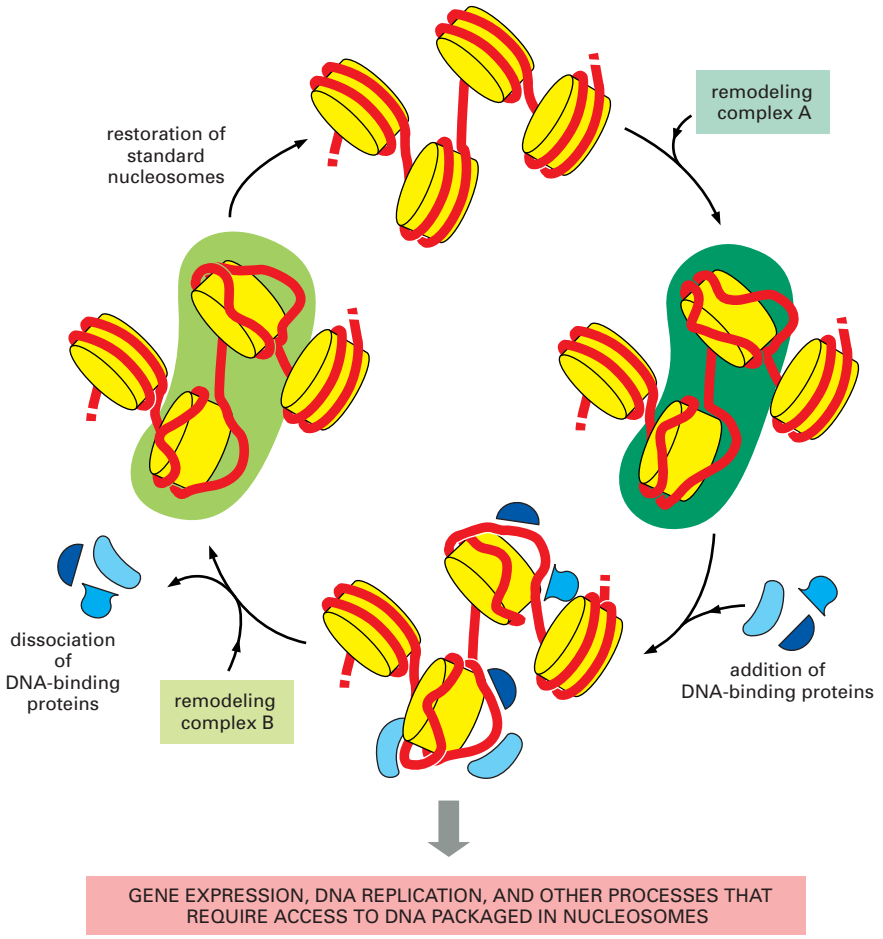
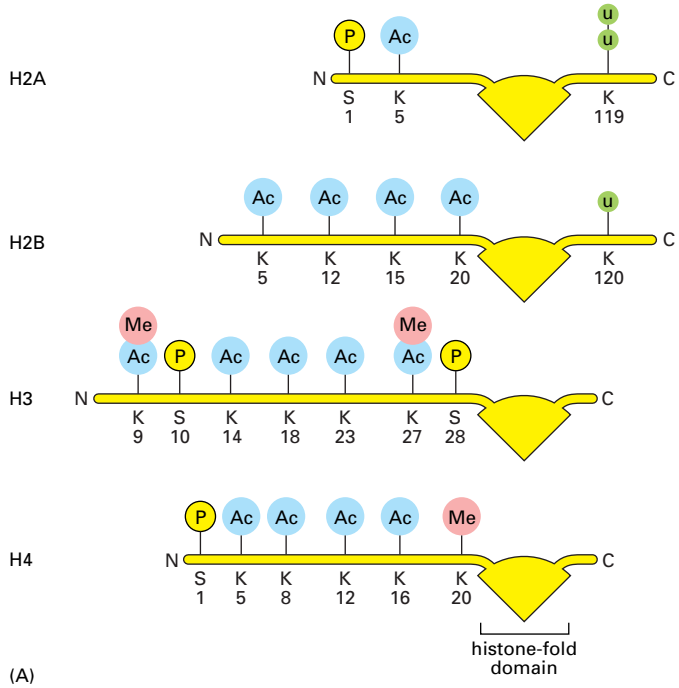
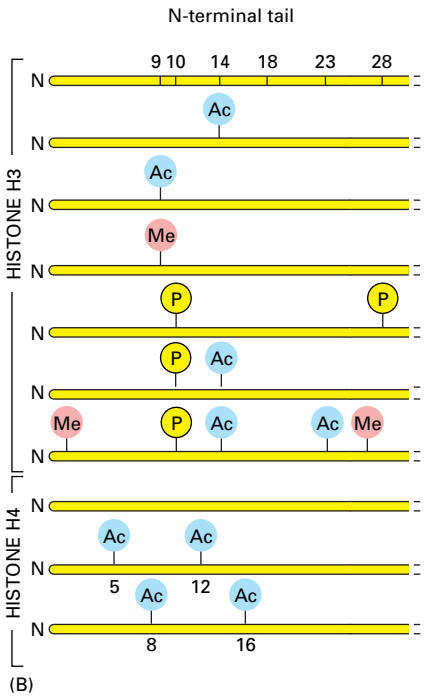


Figure 4–34 A cyclic mechanism for nucleosome disruption and re-formation. According to this model, different chromatin remodeling complexes disrupt and re-form nucleosomes, although, in principle, the same complex might catalyze both reactions. The DNA-binding proteins could function in gene expression, DNA replication, or DNA repair, and in some cases their binding could lead to the dissociation of the histone core to form nucleosome-free regions of DNA like those illustrated in Figure 4–30. (Adapted from A. Travers, *Cell* 96:311–314, 1999.)

30-nm chromatin fiber and of the higher-order structures discussed below. For example, histone acetylation tends to destabilize chromatin structure, perhaps in part because adding an acetyl group removes the positive charge from the lysine, thereby making it more difficult for histones to neutralize the charges on DNA as chromatin is compacted. However, the most profound effect of modified histone tails is their ability to attract specific proteins to a stretch of chromatin that has been appropriately modified. Depending on the precise tail modifications, these additional proteins can either cause further compaction of the chromatin or can facilitate access to the DNA. If combinations of modifications are taken into account, the number of possible distinct markings for each histone tail is very large. Thus, it has been proposed that, through covalent modification of the histone tails, a given stretch of chromatin can convey a particular meaning to





modification state	"meaning"
unmodified	gene silencing?
acetylated	gene expression
acetylated	histone deposition
methylated	gene silencing/ heterochromatin
phosphorylated	mitosis/meiosis
phosphorylated/ acetylated	gene expression
higher-order combinations	?
unmodified	gene silencing?
acetylated	histone deposition
acetylated	gene expression

Figure 4–35 Covalent modification of core histone tails. (A) Known modifications of the four histone core proteins are indicated: Me = methyl group, Ac = acetyl group, P = phosphate, u = ubiquitin. Note that some positions (e.g., lysine 9 of H3) can be modified in more than one way. Most of these modifications add a relatively small molecule onto the histone tails; the exception is ubiquitin, a 76 amino acid protein also used in other cellular processes (see Figure 6–87). The function of ubiquitin in chromatin is not well understood: histone H2B can be modified by a single ubiquitin molecule; H2A can be modified by the addition of several ubiquitins. (B) A histone code hypothesis. Histone tails can be marked by different combinations of modifications. According to this hypothesis, each marking conveys a specific meaning to the stretch of chromatin on which it occurs. Only a few of the meanings of the modifications are known. In Chapter 7, we discuss the way a doubly-acetylated H4 tail is "read" by a protein required for gene expression. In another well-studied case, an H3 tail methylated at lysine 9 is recognized by a set of proteins that create an especially compact form of chromatin, which silences gene expression.

The acetylation of lysine 14 of histone H3 and lysines 8 and 16 of histone H4—usually associated with gene expression—is performed by the type A histone acetylases (HATs) in the nucleus. In contrast, the acetylation of lysines 5 and 12 of histone H4 and a lysine of histone H3 takes place in the cytosol, after the histones have been synthesized but before they have been incorporated into nucleosomes; these modifications are catalyzed by type B HATs. These modified histones are deposited onto DNA after DNA replication (see Figure 5–41), and their acetyl groups are taken off shortly afterwards by histone deacetylases (HDACs). Thus, the acetylation at these positions signals newly replicated chromatin.

Modification of a particular position in a histone tail can take on different meanings depending on other features of the local chromatin structure. For example, the phosphorylation of position 10 of histone H3 is associated not only with the condensation of chromosomes that takes place in mitosis and meiosis but also with the expression of certain genes. Some histone tail modifications are interdependent. For example methylation of H3 position 9 blocks the phosphorylation of H3 position 10, and vice versa.

the cell (Figure 4–35B). For example, one type of marking could signal that the stretch of chromatin has been newly replicated, and another could signal that gene expression should not take place. According to this idea, each different marking would attract those proteins that would then execute the appropriate functions. Because the histone tails are extended, and are therefore probably accessible even when chromatin is condensed, they provide an especially apt format for such messages.

As with chromatin remodeling complexes, the enzymes that modify (and remove modifications from) histone tails are usually multisubunit proteins, and they are tightly regulated. They are brought to a particular region of chromatin by other cues, particularly by sequence-specific DNA-binding proteins. We can thus imagine how cycles of histone tail modification and demodification can allow chromatin structure to be dynamic—locally compacting and decompacting it, and, in addition, attracting other proteins specific for each modification state. It is likely that histone-modifying enzymes and chromatin remodeling complexes work in concert to condense and recondense stretches of chromatin; for example, evidence suggests that a particular modification of the histone tail attracts a particular type of remodeling complex. Moreover, some chromatin remodeling complexes contain histone modification enzymes as subunits, directly connecting the two processes.

Summary

A gene is a nucleotide sequence in a DNA molecule that acts as a functional unit for the production of a protein, a structural RNA, or a catalytic RNA molecule. In eucaryotes, protein-coding genes are usually composed of a string of alternating introns and exons. A chromosome is formed from a single, enormously long DNA molecule that contains a linear array of many genes. The human genome contains 3.2×10^9 DNA nucleotide pairs, divided between 22 different autosomes and 2 sex chromosomes. Only a small percentage of this DNA codes for proteins or structural and catalytic RNAs. A chromosomal DNA molecule also contains three other types of functionally important nucleotide sequences: replication origins and telomeres allow the DNA molecule to be completely replicated, while a centromere attaches the daughter DNA molecules to the mitotic spindle, ensuring their accurate segregation to daughter cells during the M phase of the cell cycle.

The DNA in eucaryotes is tightly bound to an equal mass of histones, which form a repeating array of DNA–protein particles called nucleosomes. The nucleosome is composed of an octameric core of histone proteins around which the DNA double helix is wrapped. Despite irregularities in the positioning of nucleosomes along DNA, nucleosomes are usually packed together (with the aid of histone H1 molecules) into quasi-regular arrays to form a 30-nm fiber. Despite the high degree of compaction in chromatin, its structure must be highly dynamic to allow the cell access to the DNA. Two general strategies for reversibly changing local chromatin structures are important for this purpose: ATP-driven chromatin remodeling complexes, and an enzymatically catalyzed covalent modification of the N-terminal tails of the four core histones.

THE GLOBAL STRUCTURE OF CHROMOSOMES

Having discussed the DNA and protein molecules from which the 30-nm chromatin fiber is made, we now turn to the organization of the chromosome on a more global scale. As a 30-nm fiber, the typical human chromosome would still be 0.1 cm in length and able to span the nucleus more than 100 times. Clearly, there must be a still higher level of folding, even in interphase chromosomes. This higher-order packaging is one of the most fascinating—but also one of the most poorly understood—aspects of chromosome structure. Although its molecular basis is still largely a mystery, it almost certainly involves the folding of the 30-nm fiber into a series of loops and coils, as we see below. Our discussion of this higher-order packing continues an important theme in chromosome

architecture: interphase chromatin structure is fluid, exposing at any given moment the DNA sequences directly needed by the cell.

We first describe several rare cases in which the overall structure and organization of interphase chromosomes can be easily visualized, and we explain that certain features of these exceptional cases may be representative of the structures of all interphase chromosomes. Next we describe the different forms of chromatin that make up a typical interphase chromosome. Finally we discuss the additional compaction that interphase chromosomes undergo during the process of mitosis.

Lampbrush Chromosomes Contain Loops of Decondensed Chromatin

Most chromosomes in interphase cells are too fine and too tangled to be visualized clearly. In a few exceptional cases, however, *interphase chromosomes* can be seen to have a precisely defined higher-order structure, and it is thought that certain characteristics of these higher-order structures are representative of all interphase chromosomes. The meiotically paired chromosomes in growing amphibian oocytes (immature eggs), for example, are highly active in gene expression, and they form unusually stiff and extended chromatin loops. These so-called **lampbrush chromosomes** (the largest chromosomes known) are clearly visible even in the light microscope, where they are seen to be organized into a series of large chromatin loops emanating from a linear chromosomal axis (Figure 4–36).

The organization of a lampbrush chromosome is illustrated in Figure 4–37. A given loop always contains the same DNA sequence, and it remains extended in the same manner as the oocyte grows. Other experiments demonstrate that

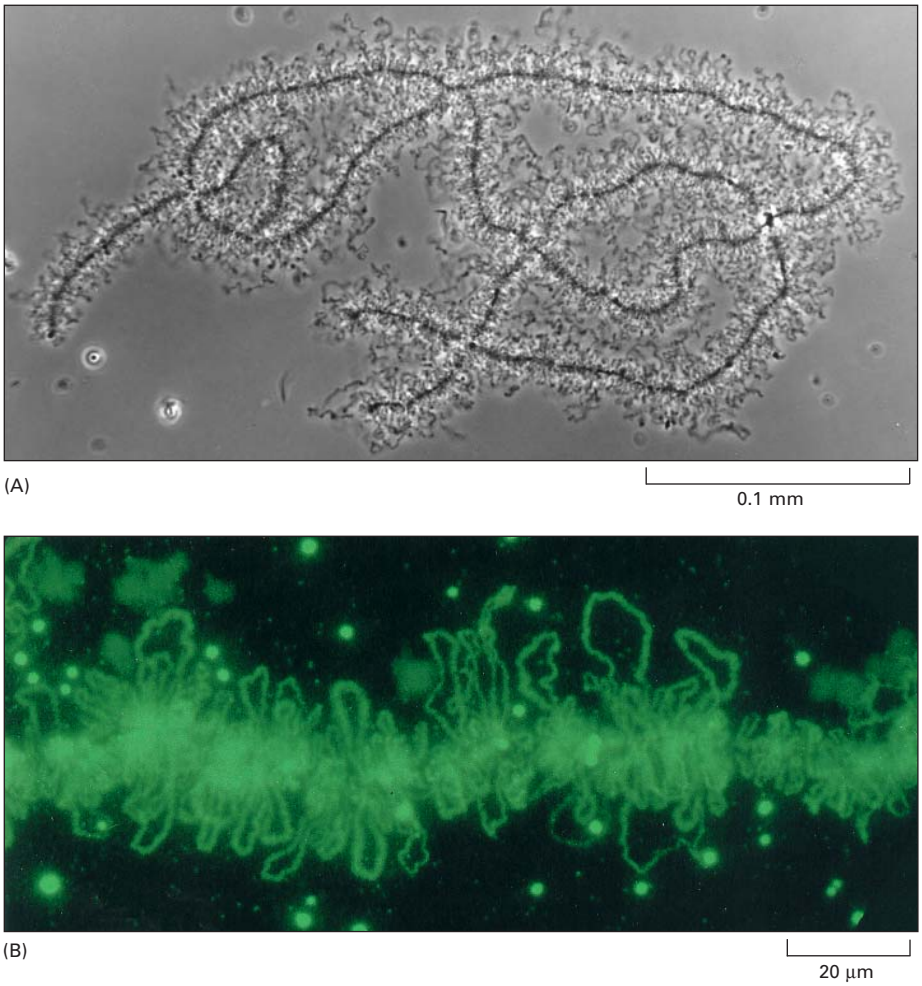


Figure 4–36 Lampbrush chromosomes. (A) A light micrograph of lampbrush chromosomes in an amphibian oocyte. Early in oocyte differentiation, each chromosome replicates to begin meiosis, and the homologous replicated chromosomes pair to form this highly extended structure containing a total of four replicated DNA molecules, or chromatids. The lampbrush chromosome stage persists for months or years, while the oocyte builds up a supply of materials required for its ultimate development into a new individual. (B) Fluorescence light micrograph showing a portion of an amphibian lampbrush chromosome. The regions of the chromosome that are being actively expressed are stained *green* by using antibodies against proteins that process RNA during one of the steps of gene expression (discussed in Chapter 6). The round granules are thought to correspond to large complexes of the RNA-splicing machinery that will also be discussed in Chapter 6. (A, courtesy of Joseph G. Gall; B, courtesy of Joseph G. Gall and Christine Murphy.)

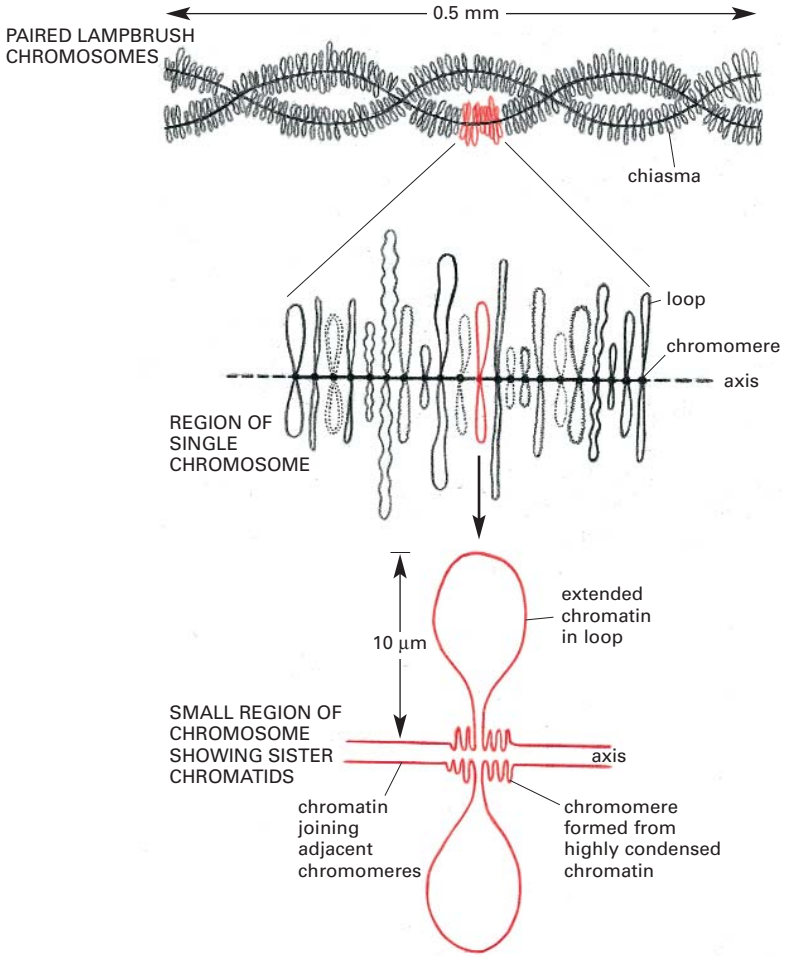


Figure 4–37 A model for the structure of a lampbrush chromosome. The set of lampbrush chromosomes in many amphibians contains a total of about 10,000 chromatin loops, although most of the DNA in each chromosome remains highly condensed in the chromomeres. Each loop corresponds to a particular DNA sequence. Four copies of each loop are present in each cell, since each of the two chromosomes shown at the top consists of two closely apposed, newly replicated chromosomes. This four-stranded structure is characteristic of this stage of development of the oocyte, the diplotene stage of meiosis; see Figure 20–12.

most of the genes present in the DNA loops are being actively expressed (see Figure 4–36B). Most of the DNA, however, is not in loops but remains highly condensed in the *chromomeres* on the axis, which are generally not expressed. Lampbrush chromosomes illustrate a recurrent theme of this chapter—when the DNA in a region of chromatin is in use (in this case, for gene expression), that part of the chromatin has an extended structure; otherwise, the chromatin is condensed. In lampbrush chromosomes, the structural units of this regulation are large, precisely defined loops.

Relatively few species undergo the specialization that produces lampbrush chromosomes. However, when injected into amphibian oocytes, the DNA from organisms that normally do not produce lampbrush chromosomes (e.g., DNA from a fish) is packaged into lampbrush chromosomes. On the basis of this type of experiment, it has been proposed that the interphase chromosomes of all eucaryotes are arranged in loops that are normally too small and fragile to be easily observed. It may be possible in the future to coax the DNA from a mammal such as a mouse to form lampbrush chromosomes by introducing it into amphibian oocytes. This could allow a detailed correlation of loop structure, gene arrangement, and DNA sequence, and we could begin to learn how the packaging into loops reflects the sequence content of our DNA.

Drosophila Polytene Chromosomes Are Arranged in Alternating Bands and Interbands

Certain insect cells also have specialized interphase chromosomes that are readily visible, although this type of specialization differs from that of lampbrush chromosomes. For example, many of the cells of certain fly larvae grow to an enormous size through multiple cycles of DNA synthesis without cell division. The resulting giant cells contain as much as several thousand times the normal DNA complement. Cells with more than the normal DNA complement are said

to be *polyploid* when they contain increased numbers of standard chromosomes. In several types of secretory cells of fly larvae, however, all the homologous chromosome copies are held side by side, like drinking straws in a box, creating a single **polytene chromosome**. The fact that, in some large insect cells, polytene chromosomes can disperse to form a conventional polyploid cell demonstrates that these two chromosomal states are closely related, and that the basic structure of a polytene chromosome must be similar to that of a normal chromosome.

Polytene chromosomes are often easy to see in the light microscope because they are so large and because the precisely aligned side-by-side adherence of individual chromatin strands greatly elongates the chromosome axis and prevents tangling. Polyteny has been most studied in the salivary gland cells of *Drosophila* larvae, in which the DNA in each of the four *Drosophila* chromosomes has been replicated through 10 cycles without separation of the daughter chromosomes, so that 1024 (2^{10}) identical strands of chromatin are lined up side by side (Figure 4–38).

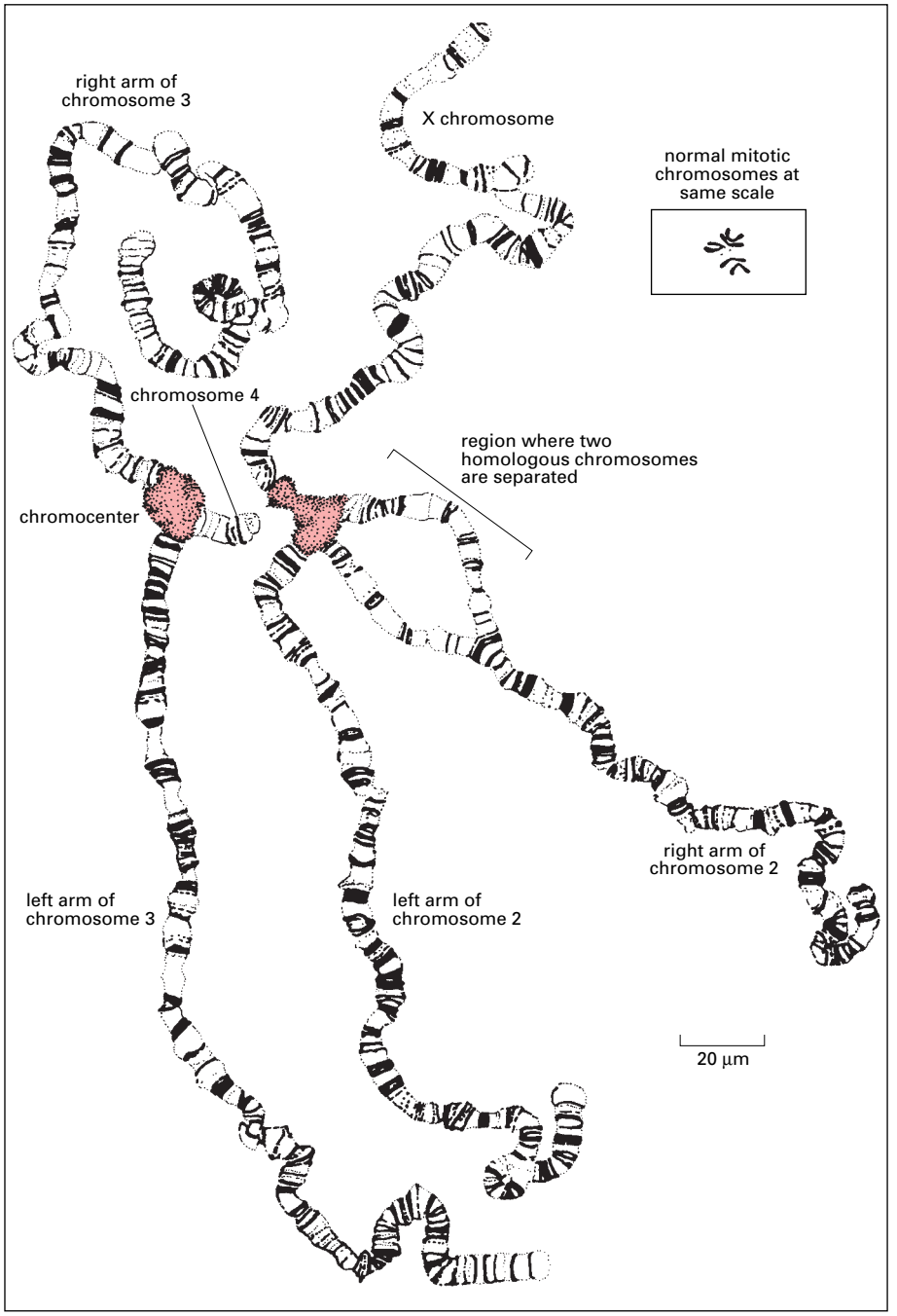


Figure 4–38 The entire set of polytene chromosomes in one *Drosophila* salivary cell. These chromosomes have been spread out for viewing by squashing them against a microscope slide. *Drosophila* has four chromosomes, and there are four different chromosome pairs present. But each chromosome is tightly paired with its homolog (so that each pair appears as a single structure), which is not true in most nuclei (except in meiosis). The four polytene chromosomes are normally linked together by regions near their centromeres that aggregate to create a single large chromocenter *(pink region)*. In this preparation, however, the chromocenter has been split into two halves by the squashing procedure used. (Modified from T.S. Painter, *J. Hered.* 25:465–476, 1934.)

When polytene chromosomes are viewed in the light microscope, distinct alternating dark *bands* and light *interbands* are visible (Figure 4–39). Each band and interband represents a set of 1024 identical DNA sequences arranged in register. About 95% of the DNA in polytene chromosomes is in bands, and 5% is in interbands. The chromatin in each band appears dark, either because it is much more condensed than the chromatin in the interbands, or because it contains a higher proportion of proteins, or both (Figure 4–40). Depending on their size, individual bands are estimated to contain 3000–300,000 nucleotide pairs in a chromatin strand. The bands of *Drosophila* polytene chromosomes can be recognized by their different thicknesses and spacings, and each one has been given a number to generate a chromosome "map." There are approximately 5000 bands and 5000 interbands in the complete set of *Drosophila* polytene chromosomes.

Both Bands and Interbands in Polytene Chromosomes Contain Genes

The reproducible pattern of bands and interbands seen in *Drosophila* polytene chromosomes means that these interphase chromosomes are highly organized. Since the 1930s, scientists have debated the nature of this organization, and we still do not have a clear answer. Because the number of bands in *Drosophila* chromosomes was once thought to be roughly equal to the number of genes in the genome, it was initially thought that each band might correspond to a single gene; however, we now know this simple idea is incorrect. There are nearly three times more genes in *Drosophila* than chromosome bands, and genes are found in both band and interband regions. Moreover, some bands contain multiple genes, and some bands seem to lack genes altogether.

It seems likely that the band–interband pattern reflects different levels of gene expression and chromatin structure along the chromosome, with genes in the less compact interbands being expressed more highly than those in the more compact bands. In any case, the remarkable appearance of fly polytene chromosomes is thought to reflect the heterogeneous nature of the chromatin compaction found along all interphase chromosomes. In the next section we see how the appearance of a band can change dramatically when the gene or genes within it become highly expressed.

Individual Polytene Chromosome Bands Can Unfold and Refold as a Unit

A major factor controlling gene expression in the polytene chromosomes of *Drosophila* is the insect steroid hormone *ecdysone*, the levels of which rise and fall periodically during larval development. When ecdysone concentrations rise,

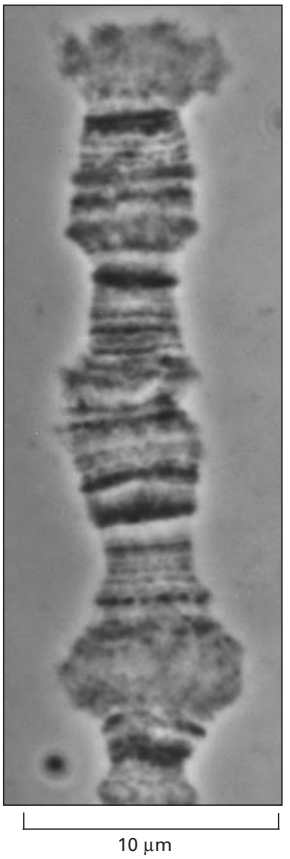


Figure 4–39 A light micrograph of a portion of a polytene chromosome from *Drosophila* salivary glands. The distinct pattern produced by bands and interbands is readily seen. The bands are regions of increased chromatin concentration that occur in interphase chromosomes. Although they are detectable only in polytene chromosomes, it is thought that they reflect a structure common to the chromosomes of most eucaryotes. (Courtesy of Joseph G. Gall.)

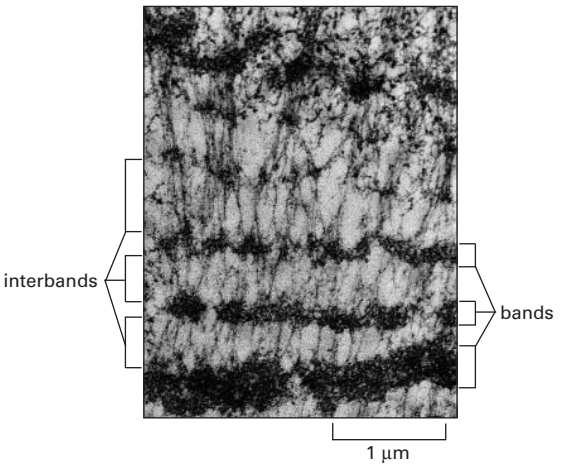


Figure 4–40 An electron micrograph of a small section of a *Drosophila* polytene chromosome seen in thin section. Bands of very different thickness can be readily distinguished, separated by interbands, which contain less condensed chromatin. (Courtesy of Veikko Sorsa.)

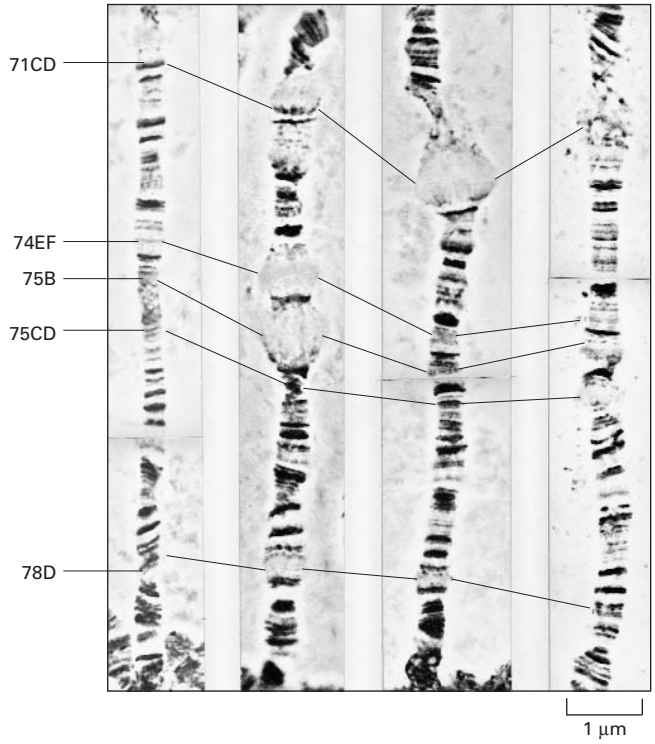
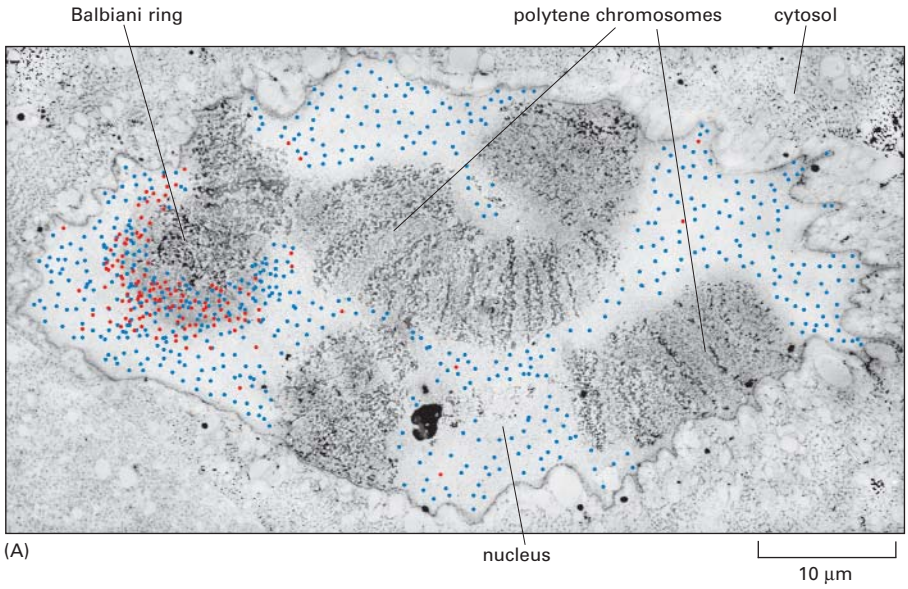


Figure 4–41 Chromosome puffs. This series of time-lapse photographs shows how puffs arise and recede in the polytene chromosomes of *Drosophila* during larval development. A region of the *left* arm of chromosome 3 is shown. It exhibits five very large puffs in salivary gland cells, each active for only a short developmental period. The series of changes shown occur over a period of 22 hours, appearing in a reproducible pattern as the organism develops. The designations of the indicated bands are given at the left of the photographs. (Courtesy of Michael Ashburner.)

they induce the expression of genes coding for proteins that the larva requires for each molt and for pupation. As the organism progresses from one developmental stage to another, distinctive *chromosome puffs* arise and old puffs recede as new genes become expressed and old ones are turned off (Figure 4–41). From inspection of each puff when it is relatively small and the banding pattern is still discernible, it seems that most puffs arise from the decondensation of a single chromosome band.

The individual chromatin fibers that make up a puff can be visualized with an electron microscope. For technical reasons, this is easier in the polytene chromosomes from a different insect, *Chironomus tentans*, a midge. Electron micrographs of certain puffs, called Balbiani rings, of *Chironomus* salivary gland polytene chromosomes show the chromatin arranged in loops (Figures 4–42 and 4–43), much like those observed in the amphibian lampbrush chromosomes discussed earlier. Additional experiments suggest that each loop contains a



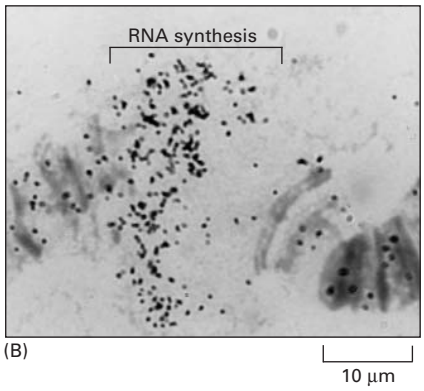


Figure 4–42 RNA synthesis in chromosome puffs. (A) Polytene chromosomes from the salivary glands of the insect *C. tentans*. As outlined in Chapter 1 and described in detail in Chapter 6, the first step in gene expression is the synthesis of an RNA molecule using the DNA as a template. In this electron micrograph, newly synthesized RNA from a Balbiani ring gene is indicated in *red.* Cells were exposed to a brief pulse of BrUTP (an analog of UTP), which was incorporated into RNA. Cells were then fixed and the newly synthesized RNAs were identified by using antibodies against BrU. Balbiani ring RNAs could be distinguished from other RNAs by their characteristic shape (see Figure 4–43). The *blue dots* in the figure represent positions of Balbiani ring RNAs that were synthesized before the addition of BrUTP. The experiment shows that Balbiani ring RNAs are synthesized in puffs and then diffuse through the nucleoplasm. (B) An autoradiogram of a single puff in a polytene chromosome. The portion of the chromosome indicated is undergoing RNA synthesis and has therefore become labeled with ^{3}H-uridine. (A, courtesy of B. Daneholt, from O.P Singh et al., *Exp. Cell Res.* 251:135–146, 1999. © Academic Press; B, courtesy of José Bonner.)

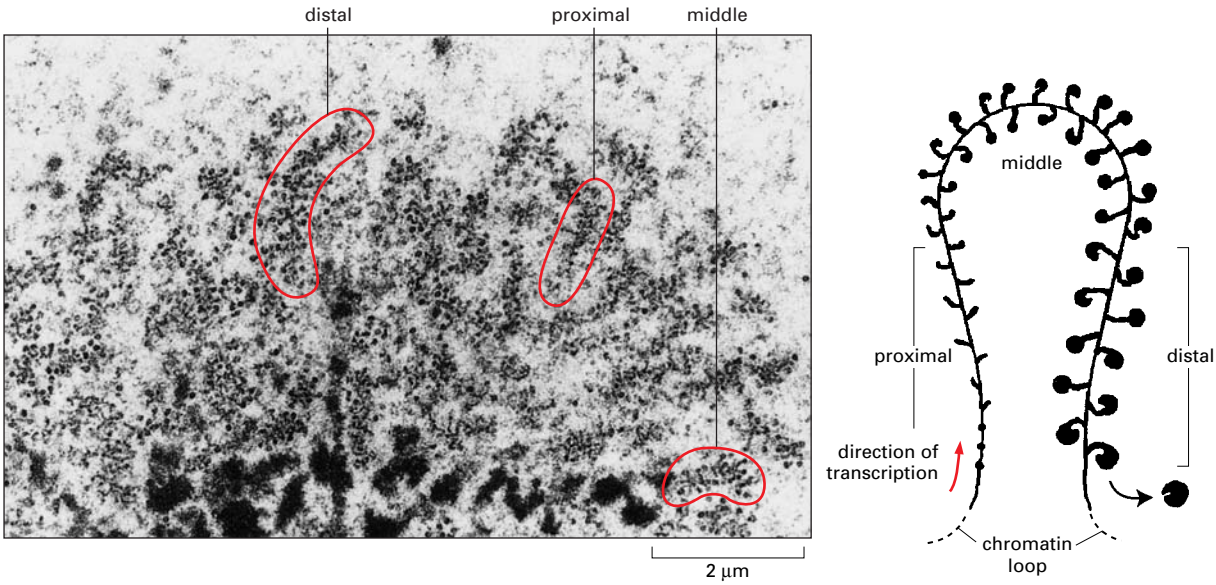


Figure 4–43 Polytene chromosomes from *C. tentans*. The electron micrograph shows a thin section of the chromatin in a Balbiani ring, a chromosome puff very active in gene expression. The Balbiani ring gene codes for secretory proteins the larvae uses to spin a protective tube. The chromatin is arranged in loops, but because the sample has been sectioned, only portions of the loops are visible. As they are synthesized on the chromatin, the RNA molecules are bound up by protein molecules, making them visible as knobs on stalks in the electron microscope. From the size of the RNA–protein complex, the extent of RNA synthesis (transcription) can be inferred; a whole chromatin loop (shown on the right) can then be reconstructed from a set of electron micrograph sections such as that shown here. (Courtesy of B. Daneholt, from U. Skoglund et al., *Cell* 34:847–855, 1983. © Elsevier.)

single gene. When not expressed, the loop of DNA assumes a thickened structure, possibly a folded 30-nm fiber, but when gene expression is occurring, the loop becomes more extended. Both types of loops contain the four core histones and histone H1.

It seems likely that the default loop structure is a folded 30-nm fiber and that the histone modifying enzymes, chromatin remodeling complexes, and other proteins required for gene expression all help to convert it to a more extended form whenever a gene is expressed. In electron micrographs, the chromatin located on either side of the decondensed loop appears considerably more compact, which is consistent with the idea that a loop constitutes an independent functional domain of chromatin structure.

Although controversial, it has been proposed that all of the DNA in polytene chromosomes is arranged in loops that condense and decondense according to when the genes within them are expressed. It may be that all interphase chromosomes from all eucaryotes are also packaged into an orderly series of looped domains, each containing a small number of genes whose expression is regulated in a coordinated way (Figure 4–44). We shall return to this issue in Chapter 7 when we discuss the ways in which gene expression is regulated by the cell.

Heterochromatin Is Highly Organized and Usually Resistant to Gene Expression

Having described some features of interphase chromosomes inferred from a few rare cases, we now turn to characteristics of interphase chromosomes that can be observed in a wide variety of organisms. Light-microscope studies in the 1930s distinguished between two types of chromatin in the interphase nuclei of many higher eucaryotic cells: a highly condensed form, called **heterochromatin**, and all the rest, which is less condensed, called **euchromatin**. Euchromatin is composed of the types of chromosomal structures—30-nm fibers and looped domains—that we have discussed so far. Heterochromatin, in contrast, includes additional proteins and probably represents more compact levels of organization that are just beginning to be understood. In a typical mammalian cell, approximately 10% of the genome is packaged into heterochromatin. Although present in many locations along chromosomes, it is concentrated in specific regions, including the centromeres and telomeres.

Most DNA that is folded into heterochromatin does not contain genes. However, genes that do become packaged into heterochromatin are usually resistant to being expressed, because heterochromatin is unusually compact. This does not mean that heterochromatin is useless or deleterious to the cell; as we see

below, regions of heterochromatin are responsible for the proper functioning of telomeres and centromeres (which lack genes), and its formation may even help protect the genome from being overtaken by "parasitic" mobile elements of DNA. Moreover, a few genes require location in heterochromatin regions if they are to be expressed. In fact, the term *heterochromatin* (which was first defined cytologically) is likely to encompass several distinct types of chromatin structures whose common feature is an especially high degree of organization. Thus, heterochromatin should not be thought of as encapsulating "dead" DNA, but rather as creating different types of compact chromatin with distinct features and roles.

Heterochromatin's resistance to gene expression makes it amenable to study even in organisms in which it cannot be directly observed. When a gene that is normally expressed in euchromatin is experimentally relocated into a region of heterochromatin, it ceases to be expressed, and the gene is said to be *silenced*. These differences in gene expression are examples of **position effects**, in which the activity of a gene depends on its position along a chromosome. First recognized in *Drosophila*, position effects have now been observed in many organisms and they are thought to reflect an influence of the different states of chromatin structure along chromosomes on gene expression. Thus, chromosomes can be considered as mosaics of distinct forms of chromatin, each of which has a special effect on the ability of the DNA it contains to be addressed by the cell.

Many position effects exhibit an additional feature called *position effect variegation*, which is responsible for the mottled appearance of the fly eye and the sectoring of the yeast colony in the examples shown in Figure 4–45. These patterns can result from patches of cells in which a silenced gene has become reactivated; once reactivated, the gene is inherited stably in this form in daughter cells. Alternatively, a gene can start out in euchromatin early in development, and then be selected more or less randomly for packaging into heterochromatin, causing its inactivation in a cell and all of its daughters.

The study of position effect variegation has revealed two important characteristics of heterochromatin. First, heterochromatin is dynamic; it can "spread" into a region and later "retract" from it at low but observable frequencies. Second, the state of chromatin—whether heterochromatin or euchromatin—tends to be inherited from a cell to its progeny. These two features are responsible for position effect variegation, as explained in Figure 4–46. In the next section, we discuss several models to account for the self-sustaining nature of heterochromatin, once it has been formed.

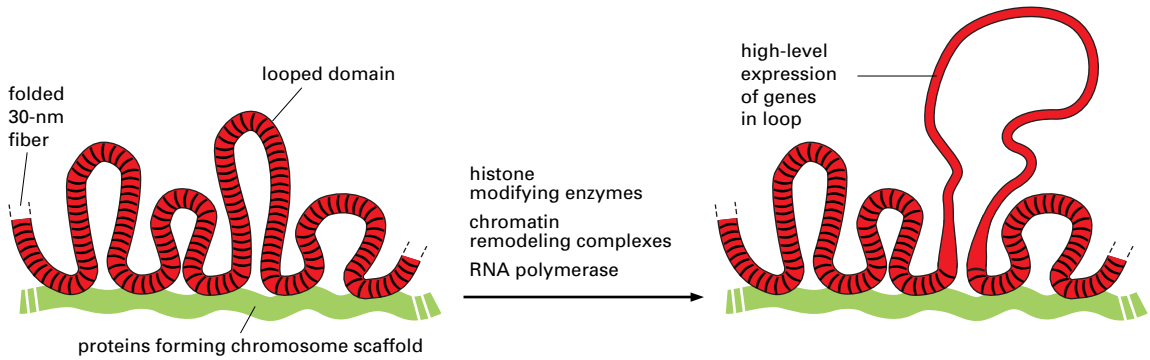


Figure 4–44 A model for the structure of an interphase chromosome. A section of an interphase chromosome is shown folded into a series of looped domains, each containing 20,000–100,000 nucleotide pairs of double-helical DNA condensed into a 30-nm fiber. Individual loops can decondense, perhaps in part through an accordionlike expansion of the 30-nm fiber (see Figure 4–29), when the cell requires direct access to the DNA packaged in these loops. This decondensation is brought about by enzymes that directly modify chromatin structure—as well as by proteins, such as RNA polymerase (discussed in Chapter 6), that act directly on the underlying DNA. It is not understood how the folded 30-nm fiber is anchored to the chromosome axis, but evidence suggests that the base of chromosomal loops is rich in DNA topoisomerases, which are enzymes that allow DNA to swivel when anchored (see pp. 251–253).

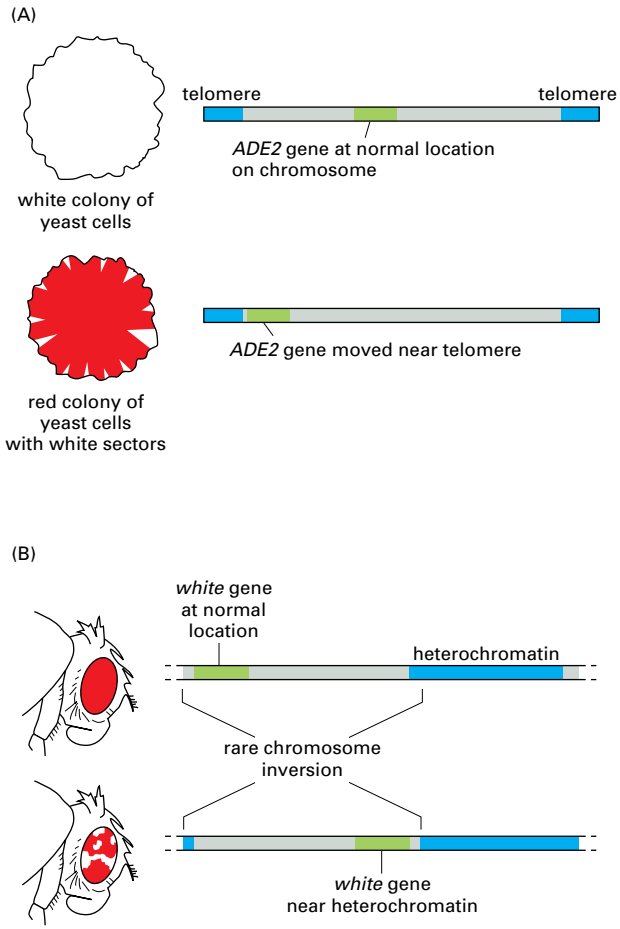


Figure 4–45 Position effects on gene expression in two different eucaryotic organisms. (A) The yeast *ADE2* gene at its normal chromosomal location is expressed in all cells. When moved near the end of a yeast chromosome, which is inferred to be folded into a form of heterochromatin, the gene is no longer expressed in most cells of the population. The *ADE2* gene codes for one of the enzymes of adenine biosynthesis, and the absence of the *ADE2* gene product leads to the accumulation of a red pigment. Therefore, a colony of cells that expresses *ADE2* is *white,* and one composed of cells where the *ADE2* gene is not expressed is *red.* The *white* sectors that fan out from the middle of the *red* colony grown on an agar surface represent the descendants of cells in which the *ADE2* gene has spontaneously become active. These white sectors are thought to result from a heritable change to a less tightly packed state of chromatin near the *ADE2* gene in these cells. Although yeast chromosomes are too small to be seen under the light microscope, the chromatin structure at the ends of yeast chromosomes is thought to have many of the same structural features as the heterochromatin in the chromosomes of larger organisms.

(B) Position effects can also be observed for the *white* gene in the fruit fly *Drosophila*. The *white* gene controls eye pigment production and is named after the mutation that first identified it. Wild-type flies with a normal *white* gene *(white*$^{-}$*)* have normal pigment production, which gives them red eyes, but if the *white* gene is mutated and inactivated, the mutant flies *(white*$^{-}$*)* make no pigment and have *white* eyes. In flies in which a normal *white*$^{+}$ gene has been moved near a region of heterochromatin, the eyes are mottled, with both *red* and *white* patches. The *white* patches represent cells in which the *white*$^{+}$ gene has been silenced by the effects of the heterochromatin. In contrast, the red patches represent cells that express the *white*$^{+}$ gene because the heterochromatin did not spread across this gene at the time, early in development, when the heterochromatin first formed. As in the yeast, the presence of large patches of *red* and *white* cells indicates that the state of transcriptional activity of the gene is inherited, once determined by its chromatin packaging in the early embryo.

The Ends of Chromosomes Have a Special Form of Heterochromatin

Unlike the nucleosome and the 30-nm fiber, heterochromatin is not well understood structurally. It almost certainly involves an additional level of folding of 30-nm fiber and requires many proteins in addition to the histones. Although its chromosomes are too small to be seen under the light microscope, the molecular nature of heterochromatin is probably best understood in the simple yeast *S. cerevisiae*. Many experiments with yeast cells have shown that the chromatin extending inward roughly 5000 nucleotide pairs from each chromosome end is resistant to gene expression, and probably has a structure that corresponds to at least one type of heterochromatin in the chromosomes of more complex organisms. Extensive genetic analysis has led to the identification of many of the yeast proteins required for this type of gene silencing.

Mutations in any one of a set of yeast *S*ilent *i*nformation *r*egulator (Sir) proteins prevent the silencing of genes located near telomeres, thereby allowing these genes to be expressed. Analysis of these proteins has led to the discovery of a telomere-bound Sir protein complex that recognizes underacetylated N-terminal tails of selected histones (Figure 4–47A). One of the proteins in this complex is a highly conserved histone deacetylase known as Sir2, which has homologs in diverse organisms, including humans, and presumably has a major role in creating a pattern of histone underacetylation unique to heterochromatin. As discussed earlier in this chapter, deacetylation of the histone tails is thought to allow nucleosomes to pack together into tighter arrays and may also render them less susceptible to some chromatin remodeling complexes. In addition, heterochromatin-specific patterns of histone tail modification are likely to attract additional proteins involved in forming and maintaining heterochromatin (see Figure 4–35).

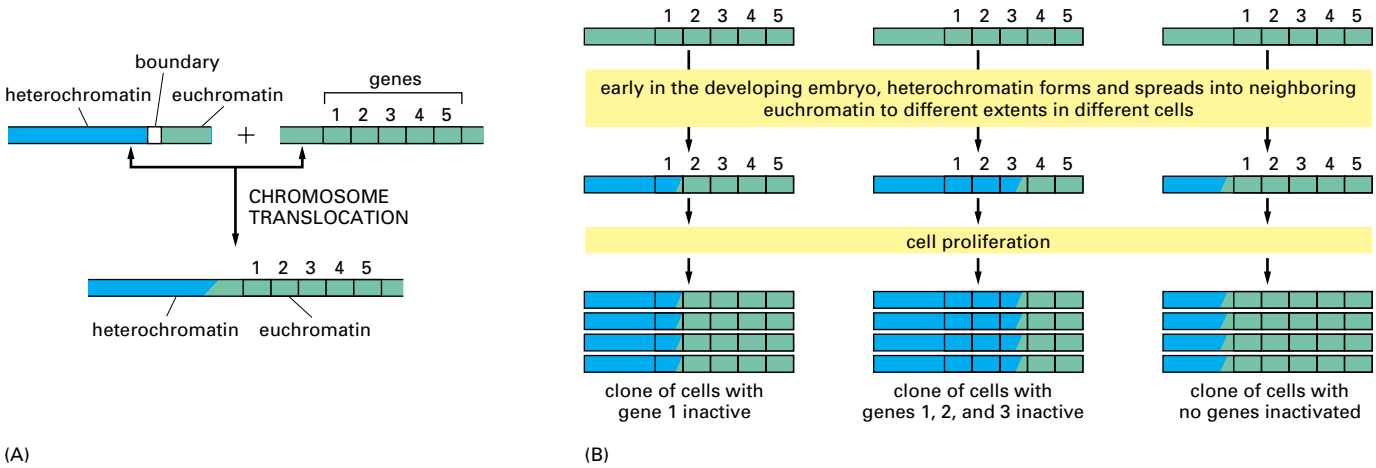


Figure 4–46 The cause of position effect variegation in *Drosophila*. (A) Heterochromatin *(blue)* is normally prevented from spreading into adjacent regions of euchromatin *(green)* by special boundary DNA sequences, which we discuss in Chapter 7. In flies that inherit certain chromosomal rearrangements, however, this barrier is no longer present. (B) During the early development of such flies, heterochromatin can spread into neighboring chromosomal DNA, proceeding for different distances in different cells. This spreading soon stops, but the established pattern of heterochromatin is inherited, so that large clones of progeny cells are produced that have the same neighboring genes condensed into heterochromatin and thereby inactivated (hence the "variegated" appearance of some of these flies; see Figure 4–45B). Although "spreading" is used to describe the formation of new heterochromatin near previously existing heterochromatin, the term may not be wholly accurate. There is evidence that during expansion, heterochromatin can "skip over" some regions of chromatin, sparing the genes that lie within them from repressive effects. One possibility is that heterochromatin can expand across the base of some DNA loops, thus bypassing the chromatin contained in the loop.

But how is the Sir2 protein delivered to the ends of chromosomes in the first place? Another series of experiments has suggested the model shown in Figure 4–47B. A DNA-binding protein that recognizes specific DNA sequences in yeast telomeres also binds to one of the Sir proteins, causing the entire Sir protein complex to assemble on the telomeric DNA. The Sir complex then spreads along the chromosome from this site, modifying the N-terminal tails of adjacent histones

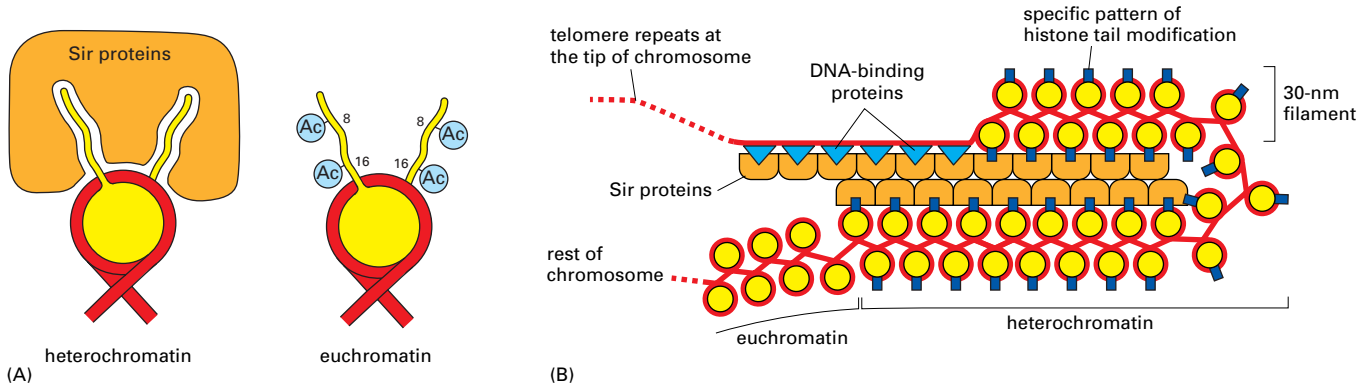


Figure 4–47 Speculative model for the heterochromatin at the ends of yeast chromosomes. (A) Heterochromatin is generally underacetylated, and underacetylated tails of histone H4 are proposed to interact with a complex of Sir proteins, thus stabilizing the association of these proteins with nucleosomes. Although shown as fully unacetylated, the exact pattern of histone H4 tail modification required to bind to the Sir complex is not known with certainty. In some organisms, the methylation of lysine 9 of histone H3 is also a critical signal for heterochromatin formation. In euchromatin, histone tails are typically highly acetylated. Those of H4 are shown as partially acetylated but, in reality, the acetylation state varies across euchromatin. (B) Specialized DNA-binding proteins *(blue triangles)* recognize DNA sequences near the ends of chromosomes and attract the Sir proteins, one of which (Sir2) is an NAD^+-dependent histone deacetylase. This then leads to the cooperative spreading of the Sir protein complex down the chromosome. As this complex spreads, the deacetylation catalyzed by Sir2 helps create new binding sites on nucleosomes for more Sir protein complexes. A "fold back" structure of the type shown may also form.

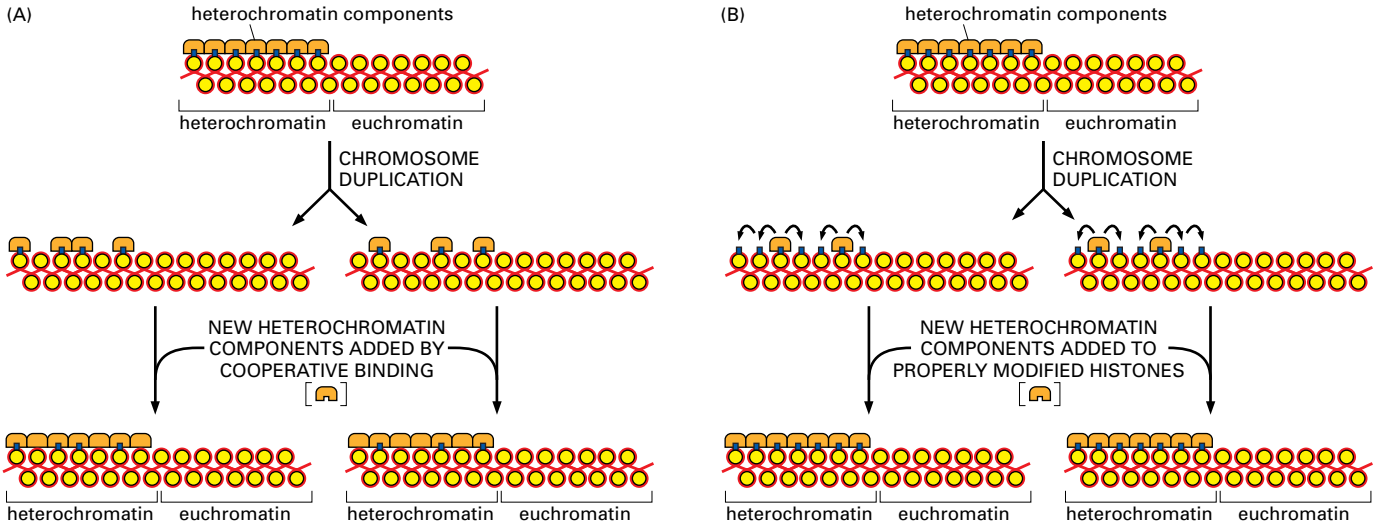


Figure 4–48 Two speculative models for how the tight packaging of DNA in heterochromatin can be inherited during chromosome replication. In both cases, half of the specialized heterochromatin components have been distributed to each daughter chromosome after DNA duplication. (A) In this model, new heterochromatin components bind cooperatively to the inherited components, thereby beginning the process of new heterochromatin formation. The process is completed with the assembly of additional proteins and the eventual covalent modification of the histones (not shown). (B) In this model, the inherited heterochromatin components change the pattern of histone modification on the newly formed daughter nucleosomes nearby, creating new binding sites for free heterochromatin components, which assemble and complete the structure. Both models can account for the spreading effects of heterochromatin, and indeed, both processes may occur simultaneously in cells.

to create the nucleosome-binding sites that the complex prefers. This "spreading effect" is thought to be driven by the cooperative binding of adjacent Sir protein complexes, as well as by the folding back of the chromosome on itself to promote Sir binding in nearby regions (see Figure 4–47B). In addition, the formation of heterochromatin probably requires the action of chromatin remodeling complexes to readjust the positions of nucleosomes as they are packed together.

Unlike most deacetylases, Sir2 requires NAD^+ as a cofactor (see Figure 2–60). The NAD^+ levels in the cell fluctuate with the nutritional health of the cell, increasing as cells become nutritionally deprived. This feature might cause the telomeric heterochromatin to spread in response to starvation (perhaps to silence the expression of genes that are not absolutely required for survival) and then to retract when conditions improve.

The properties of the yeast heterochromatin just described may resemble features of heterochromatin in more complex organisms. Certainly, the spreading of yeast heterochromatin from telomeres is similar in principle to the movement of heterochromatin that causes position effect variegation in animals (see Figure 4–46). Moreover, these properties can be used to explain the heritability of heterochromatin, as outlined in Figure 4–48. Whatever the precise mechanism of heterochromatin formation, it has become clear that covalent modifications of the nucleosome core histones have a critical role in this process. Of special importance in many organisms are the *histone methyl transferases,* enzymes that methylate specific lysines on histones including lysine 9 of histone H3 (see Figure 4–35). This modification is "read" by heterochromatin components (including HP1 in *Drosophila)* that specifically bind this modified form of histone H3 to induce the assembly of heterochromatin. It is likely that a spectrum of different histone modifications is used by the cell to distinguish heterochromatin from euchromatin (see Figure 4–35).

Having the ends of chromosomes packaged into heterochromatin provides several advantages to the cell: it helps to protect the ends of chromosomes from being recognized as broken chromosomes by the cellular repair machinery, it may help to regulate telomere length, and it may assist in the accurate pairing and segregation of chromosomes during mitosis. In Chapter 5 we see that telomeres have additional structural features that distinguish them from other parts of chromosomes.

Centromeres Are Also Packaged into Heterochromatin

Heterochromatin is also observed around centromeres, the DNA sequences that direct the movement of each chromosome into daughter cells every time a cell divides (see Figure 4–22). In many complex organisms, including humans, each

centromere seems to be embedded in a very large stretch of heterochromatin that persists throughout interphase, even though the centromere-directed movement of DNA occurs only during mitosis. The structure and biochemical properties of this so-called *centric heterochromatin* are not well understood, but, like other forms of heterochromatin, it silences the expression of genes that are experimentally placed into it. It contains, in addition to histones (which are typically underacetylated and methylated in heterochromatin), several additional structural proteins that compact the nucleosomes into particularly dense arrangements.

As with telomeres, our best understanding of the chromatin structure of a centromere comes from studies of the much simpler centromeres of the yeast *S. cerevisiae*. Earlier in this chapter we saw that a simple DNA sequence of approximately 125 nucleotide pairs was sufficient to serve as a centromere in this organism. Despite its small size, more than a dozen different proteins assemble on this DNA sequence; the proteins include a histone H3 variant that, along with the other core histones, is believed to form a centromere-specific nucleosome (Figure 4–49A). We do not yet understand what properties this variant type of nucleosome provides to the cell, but similar specialized nucleosomes seem to be present in all eucaryotic centromeres (Figure 4–49B). The additional proteins at the yeast centromere attach it to the spindle microtubules and provide signals that ensure that this attachment is complete before the later stages of mitosis are allowed to proceed (discussed in Chapters 17 and 18).

The centromeres in more complex organisms are considerably larger than those in budding yeasts. For example, fly and human centromeres extend over hundreds of thousands of nucleotide pairs and do not seem to contain a centromere-specific DNA sequence. Rather, most consist largely of short, repeated DNA sequences, known as *alpha satellite DNA* in humans (Figure 4–50). But the same repeat sequences are also found at other (noncentromeric) positions on chromosomes, and how they specify a centromere is poorly understood. Somehow the formation of the inner plate of a kinetochore is "seeded," followed by the cooperative assembly of the entire group of special proteins that form the kinetochore (Figure 4–50B). It seems that centromeres in complex organisms are defined more by an assembly of proteins than by a specific DNA sequence.

There are some striking similarities between the formation and maintenance of centromeres and the formation and maintenance of other regions of heterochromatin. The entire centromere forms as an all-or-none entity, suggesting a highly cooperative addition of proteins after a seeding event. Moreover, once formed, the structure seems to be directly inherited on the DNA as part of each round of chromosome replication. Thus, for example, some regions of our chromosomes contain nonfunctional alpha satellite DNA sequences that seem to be identical to those at the centromere; these sequences are presumed to have arisen from a chromosome-joining event that initially created one chromosome with two centromeres (an unstable, dicentric

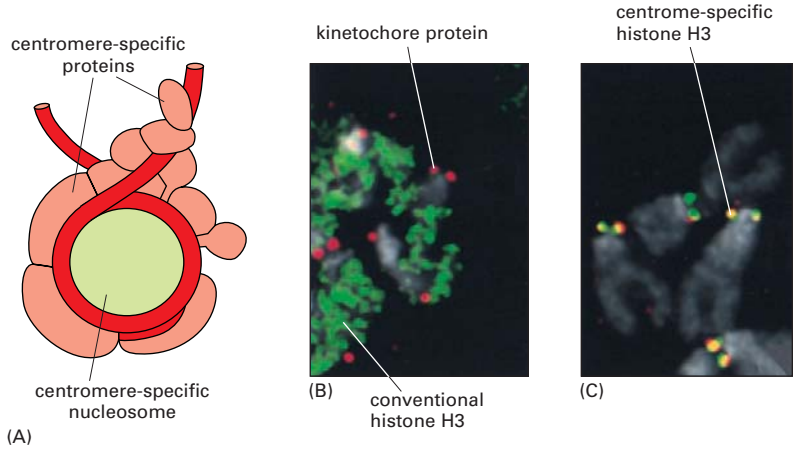


Figure 4–49 The specialized nucleosome formed on centromeres. (A) A model for the proteins that assemble on a yeast centromere. The specialized nucleosome contains an H3 variant (called CENP-A in most organisms), along with core histones H2A, H2B, and H4. The folding of DNA into this nucleosome facilitates the assembly of the other centromere-binding proteins, which form the kinetochore that attaches the centromere to the mitotic spindle (B) The localization of conventional histone H3 on *Drosophila* mitotic chromosomes. The conventional H3 has been fused to a fluorescent protein and appears *green*. A component of the kinetochore has been stained *red* with antibodies against a specific kinetochore protein. (C) The same experiment, but with the centromere-specific histone H3 (instead of the conventional H3) labeled *green*. When the *red* and *green* stains are coincident, the staining appears *yellow*. (A, adapted from P.B. Meluh et al., *Cell* 94:607–613, 1998; B and C, from S. Henikoff et al., *Proc. Natl. Acad. Sci. USA* 97:716–721, 2000. © National Academy of Sciences.)

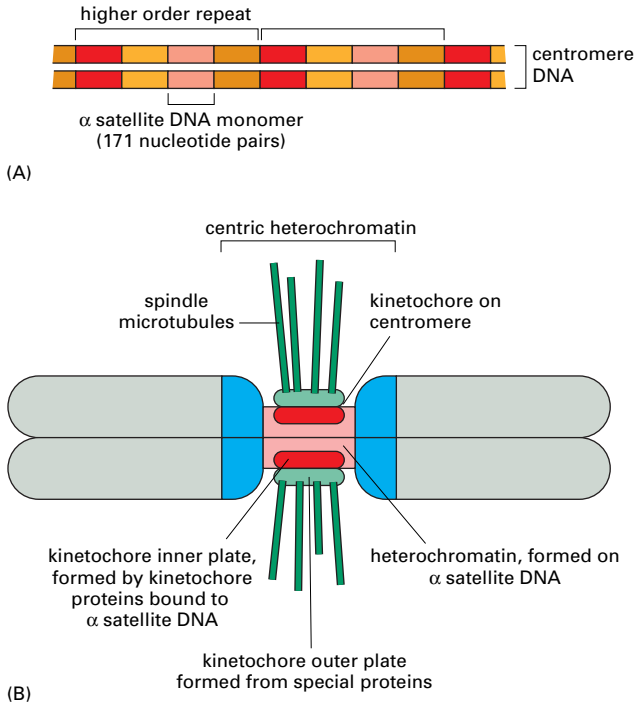


Figure 4–50 The structure of a human centromere. (A) The organization of the alpha satellite DNA sequences, which are repeated many thousands of times at a centromere. (B) An entire chromosome. The alpha satellite DNA sequences *(red)* are AT-rich and consist of a large number of repeats that vary slightly from one another in their DNA sequence. *Blue* represents the position of flanking centric heterochromatin, which contains DNA sequences composed of different types of repeats. As indicated, the kinetochore consists of an inner and an outer plate, formed by a set of kinetochore proteins. The spindle microtubules attach to the kinetochore in M phase of the cell cycle (see Figure 4–22). (B, adapted from T.D. Murphy and G.H. Karpen, *Cell* 93:317–320, 1998.)

chromosome; Figure 4–51A). Moreover, in some unusual cases, new human centromeres (called neocentromeres) have been observed to form spontaneously on fragmented chromosomes. Some of these new positions were originally euchromatic and lack alpha satellite DNA altogether (Figure 4–51B).

To explain these observations it has been proposed that *de novo* centromere formation requires an initial marking (or seeding) event, perhaps the formation of a specialized DNA–protein structure, which, in humans, happens more readily on arrays of alpha satellite DNA than on other DNA sequences. This mark would be duplicated when the chromosome divides, and the same centromere would then function in the next cell division. Very rarely, the mark would be lost after chromosome replication, in which case it would be very difficult to establish again (Figure 4–51C). Although the self-renewing nature of centromeres is not understood in detail, the type of models described for heterochromatin inheritance in Figure 4–48 could also be critical here.

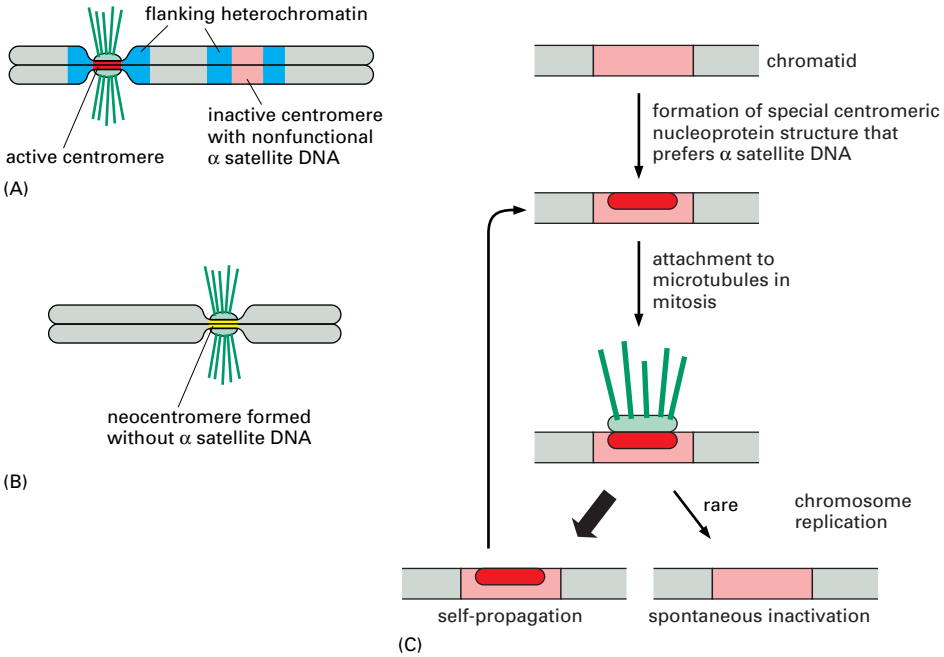


Figure 4–51 The plasticity of human centromere formation. (A) Owing to an ancient chromosome breakage and rejoining event, some human chromosomes contain two blocks of alpha satellite DNA *(red),* each of which presumably functioned as a centromere in its original chromosome. Usually, these dicentric chromosomes are not stably propagated because they are attached improperly to the spindle and are broken apart during mitosis. In those chromosomes that survive, one of the centromeres has spontaneously inactivated, even though it contains all the necessary DNA sequences. This allows the chromosome to be stably propagated. (B) In a small fraction (1/2000) of human births, extra chromosomes are observed in cells of the offspring. Some of these extra chromosomes, which have formed from a breakage event, lack alpha satellite DNA altogether, yet new centromeres have arisen from what was originally euchromatic DNA. (C) A model to explain the plasticity and inheritance of centromeres.

The plasticity of centromeres may provide an important evolutionary advantage. We have seen that chromosomes evolve in part by breakage and rejoining events (see Figure 4–19). Many of these events produce chromosomes with two centromeres, or chromosome fragments with no centromeres at all. Although rare, the inactivation of centromeres and their ability to be activated *de novo* may occasionally allow newly formed chromosomes to be maintained stably and thereby facilitate the process of chromosome evolution.

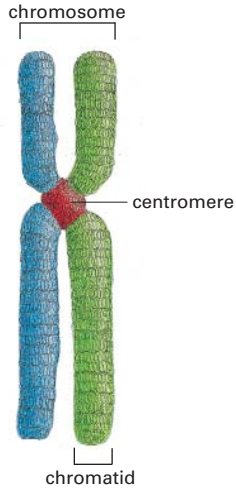


Figure 4–52 A typical mitotic chromosome at metaphase. Each sister chromatid contains one of two identical daughter DNA molecules generated earlier in the cell cycle by DNA replication.

Heterochromatin May Provide a Defense Mechanism Against Mobile DNA Elements

DNA packaged in heterochromatin often consists of large tandem arrays of short, repeated sequences that do not code for protein, as we saw above for the heterochromatin of mammalian centromeres. In contrast, euchromatic DNA is rich in genes and other single-copy DNA sequences. Although this correlation is not absolute (some arrays of repeated sequences exist in euchromatin and some genes are present in heterochromatin), this trend suggests that some types of repeated DNA may be a signal for heterochromatin formation. This idea is supported by experiments in which several hundred tandem copies of genes have been artificially introduced into the germ lines of flies and mice. In both organisms these gene arrays are often silenced, and in some cases, they can be observed under the microscope to have formed regions of heterochromatin. In contrast, when single copies of the same genes are introduced into the same position in the chromosome, they are actively expressed.

This feature, called *repeat-induced gene silencing*, may be a mechanism that cells have for protecting their genomes from being overtaken by mobile genetic elements. These elements, which are discussed in Chapter 5, can multiply and insert themselves throughout the genome. According to this idea, once a cluster of such mobile elements has formed, the DNA that contains them would be packaged into heterochromatin to prevent their further proliferation. The same mechanism could be responsible for forming the large regions of heterochromatin that contain large numbers of tandem repeats of a simple sequence, as occurs around centromeres.

Mitotic Chromosomes Are Formed from Chromatin in Its Most Condensed State

Having discussed the dynamic structure of interphase chromosomes, we now turn to the final level of DNA packaging, that observed for mitotic chromosomes. With the exception of a few specialized cases, such as the lampbrush and polytene chromosomes discussed above, most interphase chromosomes are too extended and entangled for their structures to be clearly seen. In contrast, the chromosomes from nearly all eucaryotic cells are readily visible during mitosis, when they coil up to form highly condensed structures. It is remarkable that this further condensation, which reduces the length of a typical interphase chromosome only about tenfold, produces such a dramatic change in the appearance of chromosomes.

Figure 4–52 depicts a typical **mitotic chromosome** at the metaphase stage of mitosis. The two daughter DNA molecules produced by DNA replication during interphase of the cell-division cycle are separately folded to produce two sister chromosomes, or *sister chromatids*, held together at their centromeres (see also Figure 4–21). These chromosomes are normally covered with a variety of molecules, including large amounts of RNA–protein complexes. Once this covering has been stripped away, each chromatid can be seen in electron micrographs to be organized into loops of chromatin emanating from a central scaffolding (Figures 4–53 and 4–54). Several types of experiment demonstrate that the order of visible features along a mitotic chromosome at least roughly reflects the order of the genes along the DNA molecule. Mitotic chromosome condensation can thus be thought of as the final level in the hierarchy of chromosome packaging (Figure 4–55).

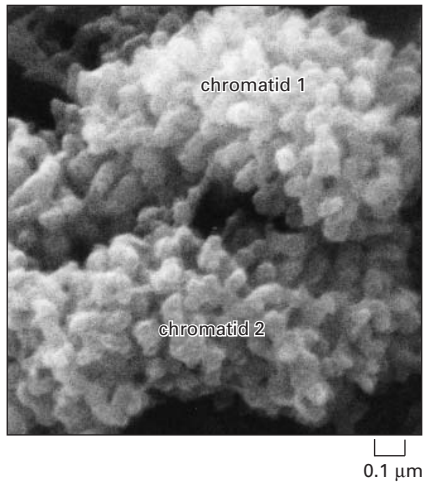


Figure 4–53 A scanning electron micrograph of a region near one end of a typical mitotic chromosome. Each knoblike projection is believed to represent the tip of a separate looped domain. Note that the two identical paired chromatids drawn in Figure 4–52 can be clearly distinguished. (From M.P. Marsden and U.K. Laemmli, *Cell* 17:849–858, 1979. © Elsevier.)

The compaction of chromosomes during mitosis is a highly organized and dynamic process that serves at least two important purposes. First, when condensation is complete (in metaphase), sister chromatids have been disentangled from each other and lie side by side. Thus, the sister chromatids can easily separate when the mitotic apparatus begins pulling them apart. Second, the compaction of chromosomes protects the relatively fragile DNA molecules from being broken as they are pulled to separate daughter cells.

The condensation of interphase chromosomes into mitotic chromosomes occurs in M phase, and it is intimately connected with the progression of the cell cycle, as discussed detail in Chapters 17 and 18. It requires a class of proteins called *condensins* which using the energy of ATP hydrolysis, drive the coiling of each interphase chromosome that produces a mitotic chromosome. Condensins are large protein complexes that contain SMC proteins: long, dimeric protein molecules hinged in the center, with globular domains at each end that bind DNA and hydrolyze ATP (Figure 4–56). When added to purified DNA, condensins use the energy of ATP hydrolysis to make large right-handed loops in the DNA. Although it is not yet known how they act on chromatin, the coiling model shown in Figure 4–56C is based on the fact that condensins are a major structural component of mitotic chromosomes, with one molecule of condensin being present for every 10,000 nucleotides of mitotic DNA.

Each Mitotic Chromosome Contains a Characteristic Pattern of Very Large Domains

As mentioned earlier, the display of the 46 human chromosomes at mitosis is called the human karyotype. When stained with dyes such as Giemsa, mitotic

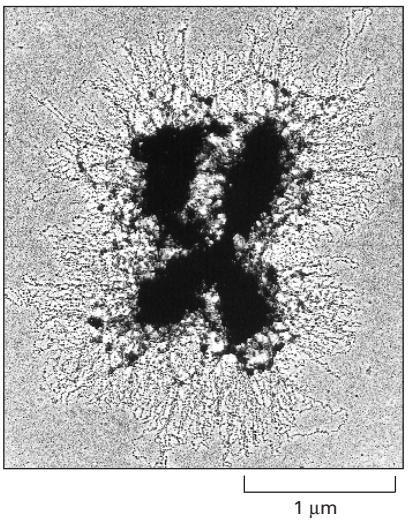
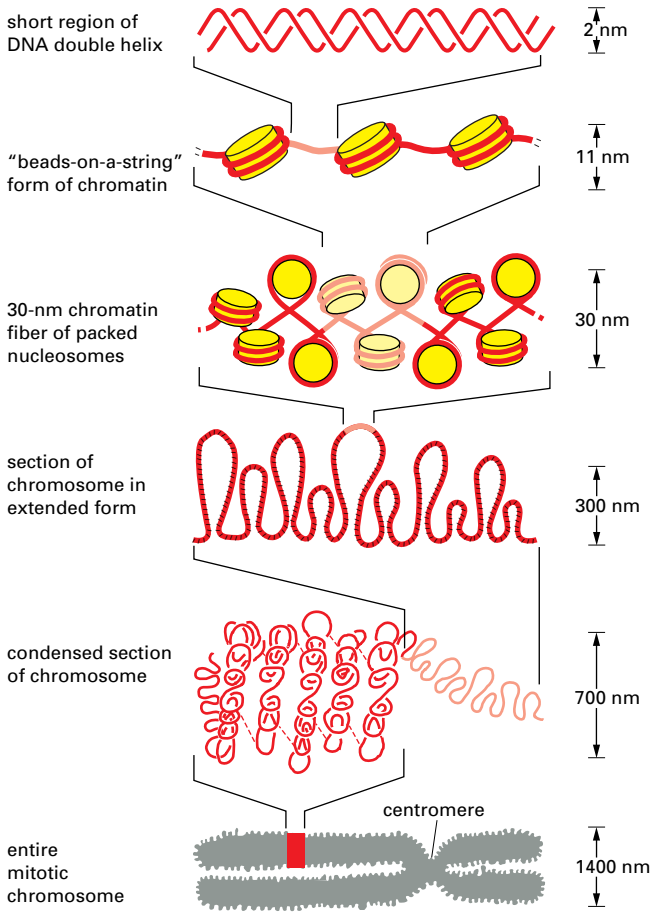


Figure 4–54 An electron micrograph of a mitotic chromosome. This chromosome (from an insect) was treated to reveal loops of chromatin fibers that emanate from a central scaffold of the chromatid. Such micrographs support the idea that the chromatin in all chromosomes is folded into a series of looped domains (see Figure 4–55). (Courtesy of Uli Laemmli.)



NET RESULT: EACH DNA MOLECULE HAS BEEN PACKAGED INTO A MITOTIC CHROMOSOME THAT IS 10,000-FOLD SHORTER THAN ITS EXTENDED LENGTH

Figure 4–55 Chromatin packing. This model shows some of the many levels of chromatin packing postulated to give rise to the highly condensed mitotic chromosome.

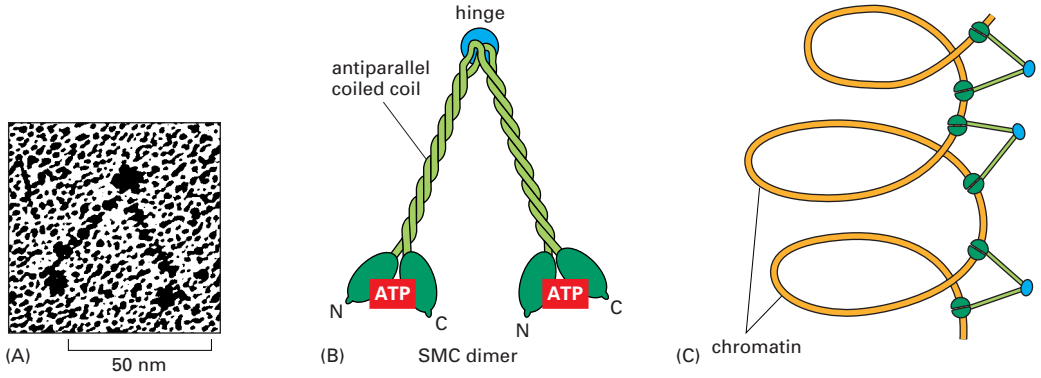


Figure 4–56 The SMC proteins in condensins. (A) Electron micrographs of a purified SMC dimer. (B) The structure of an SMC dimer. The long central region of this protein is an antiparallel coiled coil (see Figure 3–11) with a flexible hinge in its middle, as demonstrated by the electron micrograph in (A). (C) A model for the way in which the SMC proteins in condensins might compact chromatin. In reality, SMC proteins are components of a much larger condensin complex. It has been proposed that, in the cell, condensins coil long strings of looped chromatin domains (see Figure 4–55). In this way the condensins would form a structural framework that maintains the DNA in a highly organized state during M phase of the cell cycle. (A, courtesy of H.P. Erickson; B and C, adapted from T. Hirano, *Genes Dev.* 13:11–19, 1999.)

chromosomes show a striking and reproducible banding pattern along each chromosome, as shown in Figure 4–11. These bands are unrelated to those described earlier for the insect polytene chromosomes, which correspond to relatively small regions of interphase chromatin. In a human mitotic chromosome, all the chromatin is condensed and the bands represent a selective binding of the dyes.

By examining human chromosomes very early in mitosis, when they are less condensed than at metaphase, it has been possible to estimate that the total haploid genome contains about 2000 distinguishable bands. These coalesce progressively as condensation proceeds during mitosis, producing fewer and thicker bands. As we saw earlier, cytogeneticists routinely use the pattern of these chromosome bands to discover in patients genetic alterations such as chromosome inversions, translocations, and other types of chromosomal rearrangements (see Figure 4–12).

Mitotic chromosome bands are detected in chromosomes from species as diverse as humans and flies. Moreover, the pattern of bands in a chromosome has remained unchanged over long periods of evolutionary time. Each human chromosome, for example, has a clearly recognizable counterpart with a nearly identical banding pattern in the chromosomes of the chimpanzee, gorilla, and orangutan—although there are also clear differences, such as chromosome fusion, that give the human 46 chromosomes instead of the ape's 48 (Figure 4–57). This conservation suggests that chromosomes are organized into large domains that may be important for chromosomal function.

Even the thinnest of the bands in Figure 4–11 probably contains more than a million nucleotide pairs, which is nearly the size of a bacterial genome. These bands seem to reflect a rough division of chromosomes into regions of different GC content. The nucleotide sequence of the human genome has revealed large non-random blocks of sequence (some greater than 10^7 nucleotide pairs) that are significantly higher or lower in GC content than the genome-wide average of 41%. The blocks correlate roughly with the staining pattern of metaphase chromosomes. For example, bands that are darkly stained by Giemsa (the so-called G-bands) are correlated with DNA that is low in GC content, whereas lightly stained bands (the R-bands) correspond to DNA of higher than average GC content.

In general, GC-rich regions of the genome have a higher density of genes, especially of "house-keeping" genes, the genes that are expressed in virtually all cell types. On the basis of these observations, it has been proposed that the

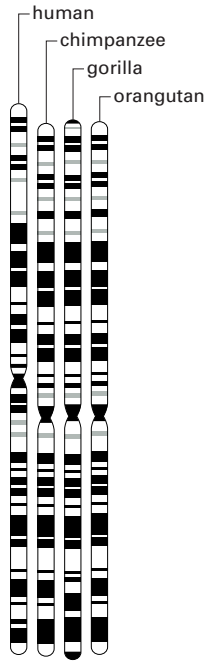


Figure 4–57 Comparison of the Giemsa pattern of the largest human chromosome (chromosome 1) with that of chimpanzee, gorilla, and orangutan. Comparisons among the staining patterns of all the chromosomes indicate that human chromosomes are more closely related to those of chimpanzee than to those of gorilla and that they are more distantly related to those of orangutan. (Adapted from M.W. Strickberger, Evolution, 3rd edn. Sudbury, MA: Jones & Bartlett Publishers, 2000.)

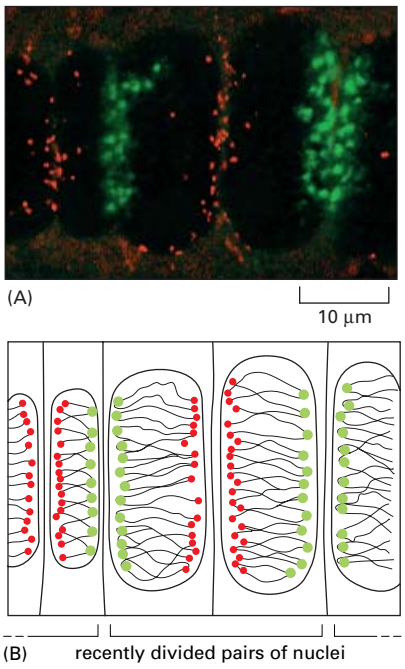


Figure 4–58 The polarized orientation of chromosomes found in certain interphase nuclei. (A) Fluorescent light micrograph of interphase nuclei from the rapidly growing root tip of a plant. Centromeres are stained *green* and teleomeres *red* by *in situ* hybridization of centromere- and telomere-specific DNA sequences coupled to the different fluorescent dyes. (B) Interpretation of (A) showing chromosomes in the Rab1 orientation with all the centromeres facing one side of a nucleus and all the telomeres pointing toward the opposite side. (A, from R. Abranches et al., *J. Cell Biol.* 143:5–12, 1998. © The Rockefeller University Press.)

banding pattern may be related to gene expression. Perhaps the differentiation of chromosomes into G- and R-bands reflects subtle differences, determined by GC content, in the way in which chromatin loops are packaged in these areas. If this idea is correct, the rough division of chromosomes can be seen as a form of compartmentalization, in which the particular cellular components involved in gene expression are more concentrated in the R-bands where their activities are required. In any case, it should be obvious from this discussion that we are only beginning to glimpse the principles of large-scale chromosome organization.

Individual Chromosomes Occupy Discrete Territories in an Interphase Nucleus

We saw earlier in this chapter that chromosomes from eucaryotes are contained in the cell nucleus. However, the nucleus is not simply a bag of chromosomes; rather, the chromosomes—as well as the other components inside the nucleus which we shall encounter in subsequent chapters—are highly organized. The way in which chromosomes are organized in the nucleus during interphase, when they are active and difficult to see, has intrigued biologists since the nineteenth century. Although our understanding today is far from complete, we do know some interesting features of these chromosome arrangements.

A certain degree of chromosomal order results from the configuration that the chromosomes always have at the end of mitosis. Just before a cell divides, the condensed chromosomes are pulled to each spindle pole by microtubules attached to the centromeres; thus, as the chromosomes move, the centromeres lead the way and the distal arms (terminating in the telomeres) lag behind. The chromosomes in some nuclei tend to retain this so-called *Rabl orientation* throughout interphase, with their centromeres facing one pole of the nucleus and their telomeres pointing toward the opposite pole (Figures 4–58 and 4–59).

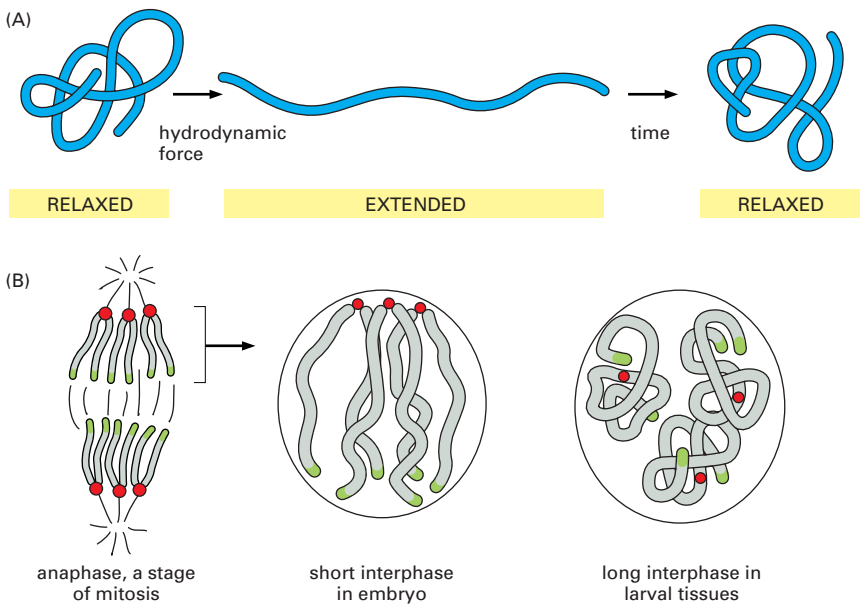


Figure 4–59 A polymer analogy for interphase chromosome organization. (A) The behavior of a polymer in solution. Entropy drives a long polymer into a compact conformation in the absence of an externally applied force. If the polymer is subjected to shear or hydrodynamic force, it becomes extended. But once the force is removed, the polymer chain returns to a more favorable, compact conformation. (B) The behavior of interphase chromosomes may reflect the same simple principles. In *Drosophila* embryos, for example, mitotic divisions occur at intervals of about 10 minutes; during the short intervening interphases, the chromosomes have little time to relax from the Rabl orientation induced by their movement during mitosis. However, in later stages of development, when interphase is much longer, the chromosomes have time to fold up. This folding may be strongly affected by specific associations between different regions of the same chromosome. (Adapted from A.F. Dernburg et al., *Cell* 85:745–759, 1996.)

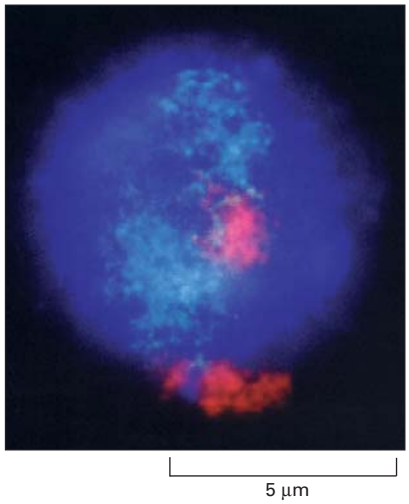


Figure 4–60 Selective "painting" of two interphase chromosomes in a human peripheral lymphocyte. The fluorescent light micrograph shows that the two copies of human chromosome 18 *(red)* and chromosome 19 *(turquoise)* occupy discrete territories of the nucleus. (From J.A. Croft et al., *J. Cell Biol.* 145:1119–1131, 1999. © The Rockefeller University Press.)

The chromosomes in most interphase cells are not found in the Rabl orientation; instead, the centromeres seem to be dispersed in the nucleus. Most cells have a longer interphase than the specialized cells illustrated above, and this presumably gives their chromosomes time to assume a different conformation (see Figure 4–59). Nevertheless, each interphase chromosome does tend to occupy a discrete and relatively small territory in the nucleus: that is, the different chromosomes are not extensively intertwined (Figure 4–60).

One device for organizing chromosomes in the nucleus may be the attachment of certain portions of each chromosome to the nuclear envelope (Figure 4–61). For example, in many cells, telomeres seem bound in this way. But the exact position of a chromosome in a nucleus is not fixed. In the same tissue, for example, two apparently identical cells can have different chromosomes as nearest neighbors.

Some cell biologists believe that there is an intranuclear framework, analogous to the cytoskeleton, on which chromosomes and other components of the nucleus are organized. The *nuclear matrix*, or *scaffold*, has been defined as the insoluble material left in the nucleus after a series of biochemical extraction steps. Some of the proteins that constitute it can be shown to bind specific DNA sequences called *SARs* or *MARs* (scaffold-associated or matrix-associated regions). These DNA sequences have been postulated to form the base of chromosomal loops (see Figure 4–44), or to attach chromosomes to the nuclear envelope and other structures in the nucleus. By means of such chromosomal attachment sites, the matrix might help to organize chromosomes, localize genes, and regulate gene expression and DNA replication. It still remains uncertain, however, whether the matrix isolated by cell biologists represents a structure that is present in intact cells.

Summary

Chromosomes are generally decondensed during interphase, so that their structure is difficult to visualize directly. Notable exceptions are the specialized lampbrush chromosomes of vertebrate oocytes and the polytene chromosomes in the giant secretory cells of insects. Studies of these two types of interphase chromosomes suggest that each long DNA molecule in a chromosome is divided into a large number of discrete domains organized as loops of chromatin, each loop probably consisting of a folded 30-nm chromatin fiber. When genes contained in a loop are expressed, the loop decondenses and allows the cell's machinery easy access to the DNA.

Euchromatin makes up most of interphase chromosomes and probably corresponds to looped domains of 30-nm fibers. However, euchromatin is interrupted by stretches of heterochromatin, in which 30-nm fibers are subjected to additional levels of packing that usually render it resistant to gene expression. Heterochromatin is commonly found around centromeres and near telomeres, but it is also present at other positions on chromosomes. Although considerably less condensed than mitotic chromosomes, interphase chromosomes occupy discrete territories in the cell nucleus; that is, they are not extensively intertwined.

All chromosomes adopt a highly condensed conformation during mitosis. When they are specially stained, these mitotic chromosomes have a banding structure that allows each individual chromosome to be recognized unambiguously. These bands contain millions of DNA nucleotide pairs, and they reflect a poorly-understood coarse heterogeneity of chromosome structure.

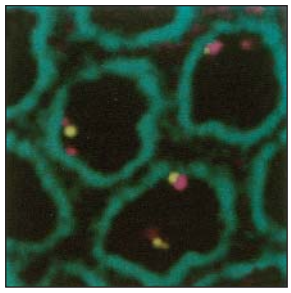


Figure 4–61 Specific regions of interphase chromosomes in close proximity to the nuclear envelope. This high-resolution microscopic view of nuclei from a *Drosophila* embryo shows the localization of two different regions of chromosome 2 *(yellow* and *magenta)* close to the nuclear envelope (stained green with antilamina antibodies). Other regions of the same chromosome are more distant from the envelope. (From W.F. Marshall et al., *Mol. Biol. Cell* 7:825–842, 1996.)

References

General

Hartwell L, Hood L, Goldberg ML et al. (2000) Genetics: from Genes to Genomes. Boston: McGraw Hill.

Lewin B (2000) Genes VII. Oxford: Oxford University Press.

Lodish H, Berk A, Zipursky SL et al. (2000) Molecular Cell Biology, 4th edn. New York: WH Freeman.

Wolfe A (1999) Chromatin: Structure and Function, 3rd edn. New York: Academic Press.

The Structure and Function of DNA

Avery OT, MacLeod CM & McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.* 79, 137.

Meselson M & Stahl FW (1958) The replication of DNA in *E. coli*. *Proc. Natl Acad. Sci. USA* 44, 671–682.

Watson JD & Crick FHC (1953) Molecular structure of nucleic acids. A structure for deoxyribose nucleic acids. *Nature* 171, 737–738.

Chromosomal DNA and Its Packaging in the Chromatin Fiber

Aalfs JD & Kingston RE (2000) What does 'chromatin remodeling' mean? *Trends Biochem. Sci.* 25, 548–555.

Cairns BR (1998) Chromatin remodeling machines: similar motors, ulterior motives. *Trends Biochem. Sci.* 23, 20–25.

Carter NP (1994) Cytogenetic analysis by chromosome painting. *Cytometry* 18, 2–10.

Cheung P, Allis CD & Sassone-Corsi P (2000) Signaling to chromatin through histone modifications. *Cell* 103, 263–271.

Clark MS (1999) Comparative genomics: the key to understanding the Human Genome Project. *Bioessays* 21, 121–130.

DePamphilis ML (1999) Replication origins in metazoan chromosomes: fact or fiction? *Bioessays* 21, 5–16.

Dunham I, Shimizu N, Roe BA et al. (1999) The DNA sequence of human chromosome 22. *Nature* 402, 489–495.

Felsenfeld G (1985) DNA. *Sci. Am.* 253(4), 58–67.

Grunstein M (1992) Histones as regulators of genes. *Sci. Am.* 267(4), 68–74B.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Jenuwein T & Allis CD (2001) Translating the histone code. *Science* 293:1074–1080.

Kingston RE & Narlikar GJ (1999) ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev.* 13, 2339–2352.

Kornberg RD & Lorch Y (1999) Chromatin-modifying and -remodeling complexes. *Curr. Opin. Genet. Dev.* 9, 148–151.

Kornberg RD & Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98, 285–294.

Luger K & Richmond TJ (1998) The histone tails of the nucleosome. *Curr. Opin. Genet. Dev.* 8, 140–146.

Luger K, Mader AW, Richmond RK et al. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.

McEachern MJ, Krauskopf A & Blackburn EH (2000) Telomeres and their control. *Annu. Rev. Genet.* 34, 331–358.

Ng HH & Bird A (2000) Histone deacetylases: silencers for hire. *Trends Biochem. Sci.* 25, 121–126.

O'Brien S, Menotti-Raymond M, Murphy W et al. (1999) The promise of comparative genomics in mammals. *Science* 286, 458–480.

Pidoux AL & Allshire RC (2000) Centromeres: getting a grip of chromosomes. *Curr. Opin. Cell Biol.* 12, 308–319.

Rhodes D (1997) Chromatin structure. The nucleosome core all wrapped up. *Nature* 389, 231–233.

Rice JC & Allis CD (2001) Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr. Opin. Cell Biol.* 13, 263–273.

Ried T, Schrock E, Ning Y & Wienberg J (1998) Chromosome painting: a useful art. *Hum. Mol. Genet.* 7, 1619–1626.

Rubin GM (2001) Comparing species. *Nature* 409, 820–821.

Stewart A (1990) The functional organization of chromosomes and the nucleus—a special issue. *Trends Genet.* 6, 377–379.

Strahl BD & Allis CD (2000) The language of covalent histone modifications. *Nature* 403, 41–45.

Travers AA (1987) DNA bending and nucleosome positioning. *Trends Biochem. Sci.* 12, 108–112.

Wu J & Grunstein M (2000) 25 years after the nucleosome model: chromatin modifications. *Trends Biochem. Sci.* 25, 619–623.

The Global Structure of Chromosomes

Agard DA & Sedat JW (1983) Three-dimensional architecture of a polytene nucleus. *Nature* 302, 676–681.

Ashburner M, Chihara C, Meltzer P & Richards G (1974) Temporal control of puffing activity in polytene chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38, 655–662.

Bickmore WA & Sumner AT (1989) Mammalian chromosome banding—an expression genome organization. *Trends Genet.* 5, 144–148.

Birchler JA, Bhadra MP & Bhadra U (2000) Making noise about silence: repression of repeated genes in animals. *Curr. Opin. Genet. Dev.* 10, 211–216.

Callan HG (1982) Lampbrush chromosomes. *Proc. Roy. Soc. Lond. Ser. B. (Biol.)* 214, 417–448.

Croft JA, Bridger JM, Boyle S et al. (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* 145, 1119–1131.

Griffith JD, Comeau L, Rosenfield S et al. (1999) Mammalian telomeres end in a large duplex loop. *Cell* 97, 503–514.

Grunstein M (1997) Molecular model for telomeric heterochromatin in yeast. *Curr. Opin. Cell Biol.* 9, 383–387.

Hart CM & Laemmli UK (1998) Facilitation of chromatin dynamics by SARs. *Curr. Opin. Genet. Dev.* 8, 519–525.

Henikoff S (1990) Position-effect variegation after 60 years. *Trends Genet.* 6, 422–426.

Henikoff S (1998) Conspiracy of silence among repeated transgenes. *Bioessays* 20, 532–535.

Hirano T (1999) SMC-mediated chromosome mechanics: a conserved scheme from bacteria to vertebrates. *Genes Dev.* 13, 11–19.

Hirano T (2000) Chromosome cohesion, condensation, and separation. *Annu. Rev. Biochem.* 69, 115–144.

Lamond AI & Earnshaw WC (1998) Structure and function in the nucleus. *Science* 280, 547–553.

Lyko F & Paro R (1999) Chromosomal elements conferring epigenetic inheritance. *Bioessays* 21, 824–832.

Marsden M & Laemmli UK (1979) Metaphase chromosome structure: evidence for a radial loop model. *Cell* 17, 849–858.

Pluta AF, Mackay AM, Ainsztein AM et al. (1995) The centromere: hub of chromosomal activities. *Science* 270, 1591–1594.

Saitoh N, Goldberg I & Earnshaw WC (1995) The SMC proteins and the coming of age of the chromosome scaffold hypothesis. *Bioessays* 17, 759–766.

Spector DL (1993) Macromolecular domains within the cell nucleus. *Annu. Rev. Cell Biol.* 9, 265–315.

Thummel CS (1990) Puffs and gene regulation—molecular insights into the Drosophila ecdysone regulatory hierarchy. *Bioessays* 12, 561–568.

Weiler KS & Wakimoto BT (1995) Heterochromatin and gene expression in Drosophila. *Annu. Rev. Genet.* 29, 577–605.

Zhimulev IF (1998) Morphology and structure of polytene chromosomes. *Adv. Genet.* 37, 1–566.