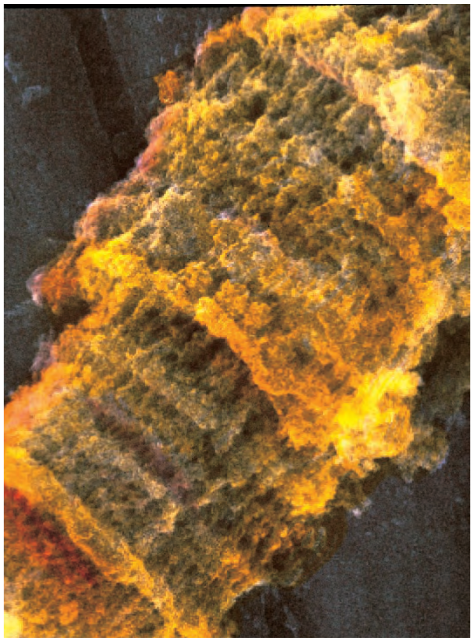
15

THE EUKARYOTIC CHROMOSOME

STUDY OBJECTIVES

1. To examine the arrangement of DNA and proteins comprising the eukaryotic chromosome 440
2. To look at the nature of centromeres and telomeres in eukaryotic chromosomes 453
3. To analyze the nature of the DNA in eukaryotic chromosomes 457

STUDY OUTLINE



Artificially colored scanning electron micrograph of part of a polytene salivary gland chromosome from a fruit fly (*Drosophila*), revealing the underlying banding pattern. (© Professors P. Motta and T. Naguro/SPL/Photo Researchers. inc.)

In chapter 14, we looked at the control of gene expression in prokaryotes and bacteriophages. Compared to eukaryotes, bacteriophages and prokaryotes are relatively simple. Of fundamental importance is that, in these lower forms, the operon model of induction and repression of transcription is a unifying theme for control of gene expression. Despite nuances such as catabolite repression and attenuator control, the operon model provides a relatively clear picture of how genes are turned on and off in phages and prokaryotes. This model does not exist for eukaryotes. In attempting to elucidate models for control of gene expression in eukaryotes, we must take one very important factor into account: the complexity of the structure of the eukaryotic chromosome. In this chapter, we cover the current understanding of how these very large structures are organized.

THE EUKARYOTIC CELL

Eukaryotes and prokaryotes are the two superkingdoms of organisms. The following comparisons, using *E. coli* as a general model for prokaryotes, show how much more complex eukaryotes are:

1. An *E. coli* chromosome contains approximately 4.2×10^6 base pairs of DNA. The haploid human genome contains nearly one thousand times as much DNA.
2. Eukaryotic DNA is in the form of nucleoprotein, a DNA-histone protein complex. Although a few histonelike proteins have been found in *E. coli,* its chromosomal DNA is not complexed with protein to anywhere near the same extent.
3. An *E. coli* cell has very little internal structure. Eukaryotes have a number of internal organelles and an extensive lipid membrane system, including the nuclear envelope itself.
4. An *E. coli* cell is small (0.5 to 5.0 μm in length for bacteria). Eukaryotic cells are generally larger than prokaryotes (10 to 50 μm in length for animal tissue cells; box 15.1).
5. The messenger RNA of *E. coli* is translated while it is being transcribed. Eukaryotic messenger RNA is modified within the nucleus before it is transported out for translation in the cytoplasm.
6. Almost no messenger RNA isolated from eukaryotic cells, including the messenger RNA of animal viruses, has been found to be polycistronic (containing many genes). Most prokaryotic messenger RNAs are polycistronic.
7. Most *E. coli* genes are parts of inducible or repressible operons; there are almost no operons in eukaryotes.
8. *E. coli* exists as a simple, single cell. Although some prokaryotes do aggregate, sporulate, and show a few other limited forms of differentiation, they are primarily one-celled organisms. And, although some eukaryotes are single-celled (e.g., yeast), the essence of eukaryotes is differentiation. In human beings, a zygote gives rise to every other cell type in the body in a relatively predictable manner.

To fully appreciate the complexity of eukaryotes, we begin by looking at the eukaryotic chromosome. In the next chapter, we look at the patterns of development in eukaryotes and some mechanisms of control of gene expression.

THE EUKARYOTIC CHROMOSOME

DNA Arrangement

Evidence that the eukaryotic chromosome is **uninemic**—that is, contains one double helix of DNA—comes from several sources. The best data are provided by radioactive-labeling studies, first done by J. Taylor and his colleagues in 1957. If a eukaryote is allowed to undergo one DNA replication in the presence of tritiated (^{3}H-) thymidine, each of the daughter chromatids would be expected to contain a double helix with one unlabeled DNA template strand and one labeled strand of newly synthesized bases (fig. 15.1). This configuration is expected on the basis of semiconservative replication, with each chromatid containing one double helix. A second round of DNA replication, in the absence of ^{3}H-thymidine, should produce chromosomes in which one chromatid would have unlabeled DNA and one would have labeled DNA. Figure 15.2 shows the chromosomes after this second replication in nonlabeled media. As expected, one chromatid of every pair is labeled and one is not.

In another kind of experiment, R. Kavenoff, L. Klotz, and B. Zimm demonstrated that *Drosophila* nuclei contained pieces of DNA of the size predicted from their DNA content, based on the premise that each chromosome contains one DNA molecule. They isolated the DNA and measured the size of the largest DNA molecules using the *viscoelastic* property of DNA, the rate at which


Ruth Kavenoff (1944–). (Courtesy of Dr. Ruth Kavenoff.)

BOX 15.1

Experimental Methods

How Big Is Big, How Small Is Small?

Generally, eukaryotic cells are large, and prokaryotic cells are small. For example, an average eukaryotic cell is about 50 µm in diameter, whereas an average bacterium is about 5 µm in length. The average virus is about 0.05 µm in diameter. These size differences occur because eukaryotic cells have complex substructures and internal architecture that prokaryotic cells lack. Since we believe that prokaryotic cells depend on diffusion to exchange materials with the environment, they would have to be small. And viruses, intracellular parasites, would of necessity be very small. There are, of course, exceptions.

In 1999, a team of scientists from Germany, Spain, and the United States isolated large sulfur bacteria off the Namibian coast of Africa and named them *Thiomargarita namibiensis,* the sulfur pearl of Namibia. These bacteria can be almost half a millimeter in diameter, the size of the period at the end of this sentence (fig. 1). Based on the sequence of 16S ribosomal DNA, these bacteria were shown to be close relatives of other marine sulfur bacteria. They are almost one hundred times the volume of the bacteria previously believed to be largest, *Epulicium fishelsoni,* known only from the intestine of the brown surgeonfish.

The smallest prokaryotes are the Mycoplasmas, at about 0.2 µm in diameter, rivaling the viruses in size. They are animal pathogens and decomposing organisms. The smallest eukaryote, *Ostreococcus tauri,* a green alga found in the plankton, was discovered in 1994 from a water sample in a French lagoon on the Mediterranean Sea by a group of French scientists. These organisms are less than 1 µm in diameter. Scientists believe that the lower limit on the size of a cell (not counting viruses) is about 200 nm (0.2 µm), based on the size of DNA and ribosomes that a cell must contain.

With *T. namibiensis* as the largest prokaryote, we note that the largest eukaryotic cell with a single nucleus is most likely the ostrich egg. The largest organisms are the blue whale, *Balaenoptera musculus,* weighing in at 118,000 kilograms; giant redwood trees, *Sequoiadendron giganteum,* 100 meters tall and weighing 5.5 million kilograms; a quaking aspen clone, *Populus tremuloides,* weighing 6 million kilograms; and *Armillaria bulbosa,* a fungus. In 1992, three scientists from the University of Toronto and Michigan Technological University, using restriction fragment length polymorphisms (RFLPs) and polymerase chain reaction (PCR) techniques, showed that the huge hyphal mass of this tree-root colonizing fungus growing in a forest in northern Michigan was a single organism. It covered about eight hectares, probably weighed more than 10,000 kilograms, and probably has existed for more than 1,500 years.

Although we don't want to get distracted by the oddities and extremes of nature, size differences are remarkable.

Figure 1 The bacterium *Thiomargarita namibiensis* shown with a fruit fly (*Drosophila virilis,* 3 mm in length) for size comparison. The arrow points to a single bacterial cell, 0.5 mm wide, bright with sulfur inclusions. Above the cell are empty sheaths of dead bacteria. (From H.N. Schulz, et al., "Dense populations of a giant sulfur bacterium in Namibian Shelf Sediments" in *Science,* Vol. 284, pp. 493–95, April 16, 1999. Reprinted by permission of the American Association for the Advancement of Science.)

stretched molecules relax. From other sources, primarily UV absorbance studies, it was estimated that the largest *Drosophila* chromosome had about 43×10^9 daltons of DNA. Results from the viscoelastic measurements indicated the presence of DNA molecules of between 38 and 44×10^9 daltons. Viscoelastic measurements of inversions, which changed the ratio of the arms but not the overall size of the chromosome, yielded similar results. However, a translocation that radically changed the size of the chromosome to 59×10^9 daltons resulted in an

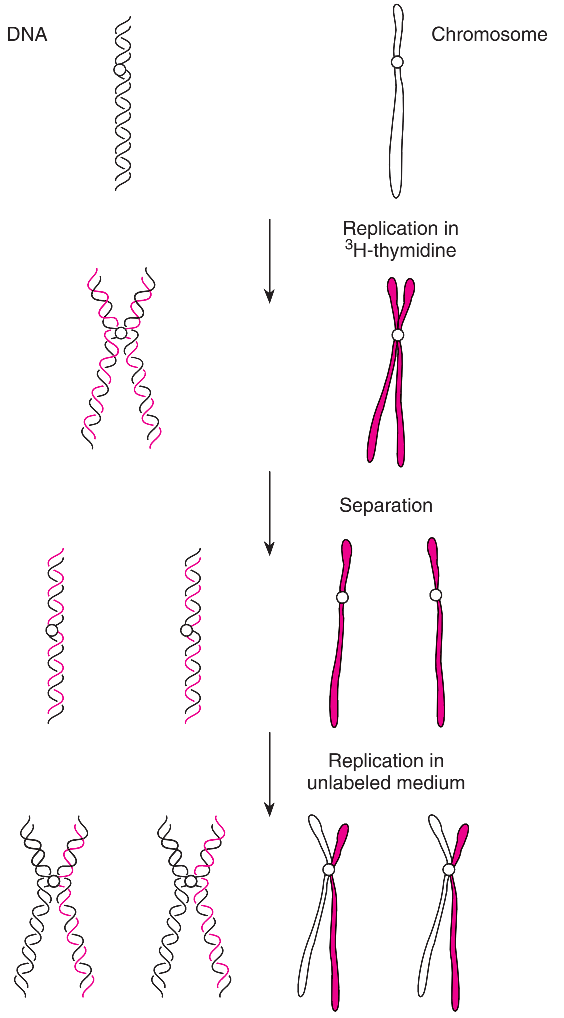


Figure 15.1 Radioactive labeling of a uninemic eukaryotic chromosome following semiconservative replication. Replication occurs first in the presence of ^{3}H-thymidine and then in its absence. *Red* represents labeling. After the second round of replication, one chromatid of each chromosome is labeled, whereas the other is not, confirming that there is only one DNA molecule per chromatid and that the chromosome is thus uninemic.

equivalent change in the viscoelastic estimates to between 52 and 64 × 10^9 daltons.

The conclusion from these studies is that the largest *Drosophila* chromosome, and by extension every eukaryotic chromosome, contains a single DNA molecule running from end to end, encompassing both arms. The viscoelastic values were corroborated by carefully isolating and measuring the lengths of long DNA molecules, an especially difficult task given DNA's propensity to break.

Figure 15.2 Second metaphase in hamster cells in culture after one replication in the presence of ^{3}H-thymidine followed by one in nonradioactive medium, verifying the uninemic nature of the eukaryotic chromosome. Cases in which the label apparently switches from one chromatid to the other are caused by sister chromatid exchanges (at *arrows*). (Source: G. Marin and D. M. Prescott, "The frequency of sister chromatid exchanges following exposure to varying doses of ^{3}H-thymidine or X-ray," *Journal of Cell Biology,* 21, (1964): 159–67, by copyright permission of the Rockefeller University Press.)

The longest molecule that the investigators found was 1.2 cm long, equivalent to between 24 and 32 × 10^9 daltons (fig. 15.3), close to the predicted size. Thus, the evidence is in complete concordance with the simple uninemic model of eukaryotic chromosomal structure (box 15.2).

Nucleoprotein Composition

Nucleosome Structure

Since each eukaryotic chromosome consists of a single, relatively long piece of duplex DNA, the average diploid cell contains many of these long pieces of DNA. For chromosomes to be properly distributed to each daughter cell during mitosis and meiosis, they must be condensed into structures that are more easily managed. Wrapping the DNA around "spools" of protein constitutes the first step in a series of coiling and folding processes that eventually result in the fully compacted chromosome we see at metaphase.

Interphase nuclei can be disrupted by placing them in a hypotonic liquid such as water. When this happens, chromatin material is released. When this material is observed under the electron microscope, small particles

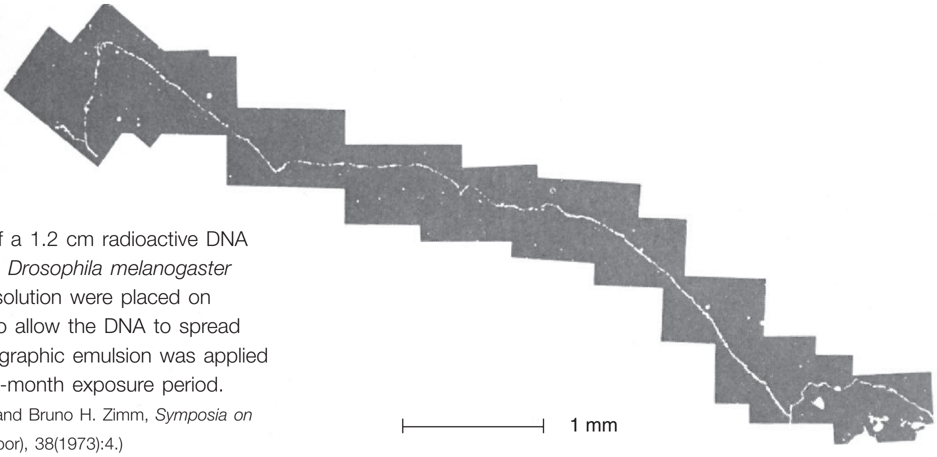


Figure 15.3 Autoradiograph of a 1.2 cm radioactive DNA molecule carefully isolated from *Drosophila melanogaster* chromosomes. Drops of DNA solution were placed on microscope slides, then tilted to allow the DNA to spread slowly down the slide. A photographic emulsion was applied and later developed after a five-month exposure period. (From Ruth Kavenoff, Lynn C. Klotz, and Bruno H. Zimm, *Symposia on Quantitative Biology* (Cold Spring Harbor), 38(1973):4.)

called **nucleosomes** can be seen (fig. 15.4). These are the spools that the DNA is wrapped around. They are made of **histone** proteins and associated DNA (table 15.1). The histones, a group of arginine- and lysine-rich basic proteins, have been well characterized. They are especially well suited to bind to the negatively charged DNA (table 15.2).

When chromatin is treated with micrococcal nuclease, individual nucleosomes can be isolated, indicating that the DNA between nucleosomes is accessible to digestion. The results of these studies indicate that a length of 168 base pairs (bp) of DNA, the core DNA, is intimately associated with the nucleosome, and another 50 to 75 base pairs, depending on species, connects the nucleosomes (linker DNA; fig. 15.5). When the quantities of the various histones were measured, there were two each of histones H2A, H2B, H3, and H4 per nucleosome and only one molecule of histone H1. Reconstitution and degradation studies have indicated that histone H1 is not a necessary component in the formation of nucleosomes. We believe that histone H1 is associated with the linker DNA as it enters and emerges from the nucleosome (fig. 15.6), although its exact position is not known with certainty. Histone H1 may be more off center and internally located than illustrated. The term **chromatosome** has been suggested for the core nucleosome plus the H1 protein, a unit that includes approximately 168 base pairs of DNA. Nucleosomes, then, are a first-order packaging of DNA; they reduce its length and undoubtedly make the coiling and contraction required during mitosis and meiosis more efficient (fig. 15.7).

When DNA is replicated, twice as many nucleosomes are needed since one double helix becomes two. Recent studies indicate that a parental nucleosome is partly disassembled during DNA replication and reassembled on one or the other daughter strand, apparently randomly. The other DNA strand has a new nucleosome constructed

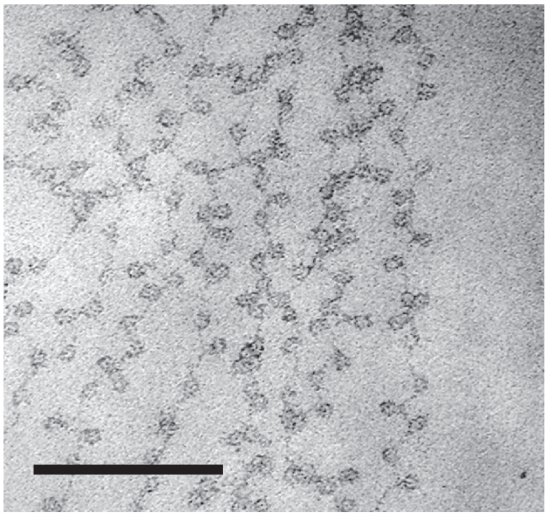


Figure 15.4 Electron micrograph of chromatin fibers. Photo shows nucleosome structures (*spheres*) and connecting strands of DNA called linkers. The bar is 100 nm long. (Source: D. E. Olins and A. L. Olins, "Nucleosomes: The structural quantum in chromosomes," *American Scientist,* 66: 704–11, November 1978. Reproduced by permission.)

Table 15.1 **The Constituency of Calf Thymus Chromatin**

Constituent	Relative Weight*
DNA	100
Histone proteins	114
Nonhistone proteins	33
RNA	1

* Weight relative to 100 units of DNA.

BOX 15.2

Experimental Methods

High-Speed Chromosomal Sorting

To facilitate the creation of recombinant genomic libraries, for mapping purposes, and for other reasons, it is useful to be able to isolate individual human chromosomes. To these ends, several methods have been developed to isolate chromosomes. Here we discuss a high-speed sorting method based on fluorescent staining and flow cytometry.

DNA can be treated with several fluorescent dyes. Chromosomes can then be recognized individually by their relative fluorescent intensities. The dyes Hoechst 33258 and chromomycin A3 are a valuable combination because they respond to different wavelengths of light and they bind DNA differently. Hoechst binds preferentially to DNA rich in adenine and thymine, whereas chromomycin binds preferentially to DNA rich in guanine and cytosine. Thus, since every human chromosome has a unique ratio of bases, the relative intensity of each chromosome is different when fluoresced.

Chromomycin fluoresces in the presence of a laser tuned to 458 nm, and Hoechst fluoresces in the presence of a UV laser. The chromosomes can be identified when their relative fluorescence in the two lasers is plotted, producing a *flow karyotype* (fig. 1). Modern flow cytometry techniques then allow the isolation of these identified chromosomes.

In practice, chromosomes are isolated in large numbers from cells that have been arrested in metaphase by treatment with colcemid, which inhibits spindle formation. These chromosomes are then purified in buffer and treated with the two dyes. The chromosomes are separated at high speed (two hundred chromosomes per second) in a flow cytometry device (fig. 2). As the chromosome-containing buffer passes through the laser beams, identification is made. The liquid is then forced to form minute droplets (215,000 per second) by passing through a vibrator. Specific droplets carrying the identified chromosomes are then charged, either positively or negatively, and passed between deflection plates. Positively charged droplets pass one way, and negatively charged droplets pass the other way, thus allowing the simultaneous isolation of two different chromosomes. At a rate of two hundred chromosomes per second, it

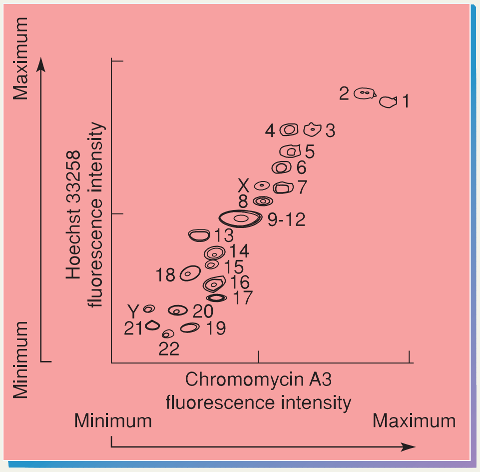


Figure 1 Flow karyotype of human chromosomes at very high resolution, measured under low-speed sorting (fifteen to thirty-five chromosomes per second). The ordinate is Hoechst 33258 fluorescence intensity, and the abscissa is chromomycin A3 fluorescence intensity. All chromosomes are resolved except numbers 9–12. (Reprinted with permission from J. W. Gray, et al., "High-Speed Chromosome Sorting," *Science*, 238:323–329, 1987. Copyright © 1987 American Association for the Advancement of Science.)

of histones from the cellular pool with the help of proteins called **chromatin assembly factors;** at least three of these factors are known.

For example, in fruit flies, a protein complex called the **replication-coupling assembly factor** assembles new nucleosomes. In addition, a protein complex called **condensin** is needed for the condensation of interphase chromosomes to mitotic chromosomes. This complex includes two **SMC proteins** (for *s*tructural *m*aintenance of *c*hromosomes) and two non-SMC proteins. SMC proteins also aid other chromosomal activities, such as mitotic segregation, sister-chromatid adhesion, dosage compensation, and recombination. Thus, a diverse array of proteins is involved in creating nucleosomes and chromatois possible to isolate 0.1 g of DNA in less than an hour; 0.1 g of DNA is adequate for library construction and represents about 5 × 10^5 average chromosomes.

The technique is not perfect. During isolation, debris and clumps of chromosomes are produced that cause contamination problems. Then, some chromosomes are so similar in their fluorescence that they are hard to separate. This is true, for example, for chromosomes 9 to 12. Also, chromosome 21 is hard to separate because its fluorescence tends to fall into the debris area.

Some of these problems, however, can be overcome by using hybrid cell lines of hamsters, for example, containing only one human chromosome. It is much easier to isolate the human chromosome from the hybrid line. Purity values of 90% are not unreasonable, with some in excess of 95%.

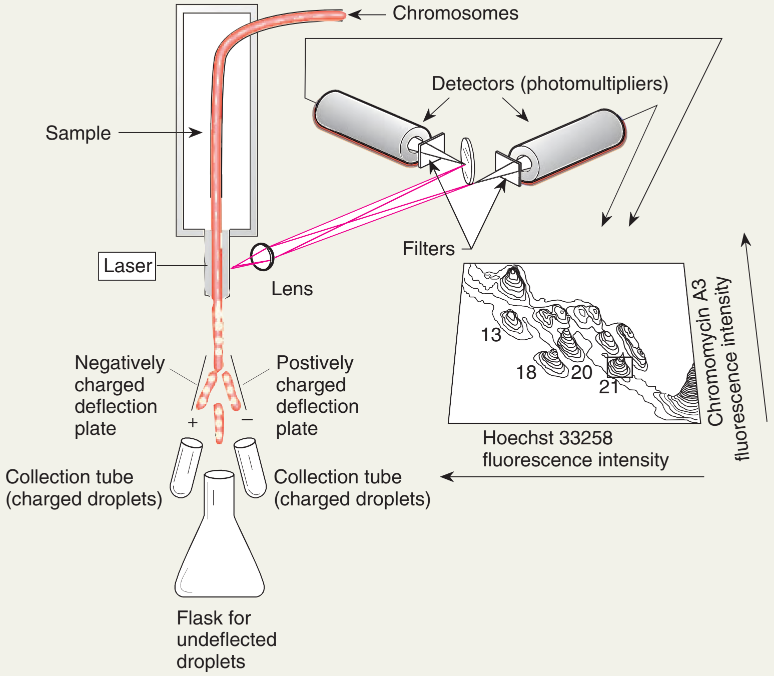


Figure 2 The flow cytometry device used to separate chromosomes at high speed. A buffer with chromosomes enters the device. Lasers cause fluorescence that is analyzed with the aid of the photomultiplier tubes. Droplet formation is induced by vibration, and, based on a flow rate of 50 m/sec, appropriate drops are charged. Charged drops are then separated by charged deflection plates and collected. Uncharged droplets pass through. (Reprinted with permission from J. W. Gray, et al., "High-Speed Chromosome Sorting," *Science,* 238:323–329, 1987. Copyright © 1987 American Association for the Advancement of Science.)

somes, condensing interphase chromosomes, and performing numerous other activities of chromosomes.

Dosage compensation has recently been associated with a change in nucleosome structure. The inactivated X chromosome appears to have a different type of histone present. Histone H2A is replaced by a variant called mH2A. The details of this mechanism are under study.

Nucleosomes apparently play a major role in controlling gene expression; DNA with nucleosomes has a much lower transcription rate than DNA without nucleosomes. It makes sense that the positions of nucleosomes can provide or prevent access to promoters. There are regions of the DNA, known as **nuclease-hypersensitive sites,** that appear to be nucleosome free. These sites, usually mutiples of a nucleosomal region of about two hundred base pairs, are particularly sensitive to digestion by different nucleases. When these regions are isolated, they usually have sequences that control functions in replication, transcription, or other activities of DNA. For example, numerous promoter regions in *Drosophila,* mouse, and human DNA are in nuclease-hypersensitive sites. Hence, some specific DNA sequences are kept free of nucleosomes, and these sequences appear to be recognized by various enzymes such as RNA polymerase. In many other cases, however, nucleosomes do appear to cover promoters and repress transcription. For transcription to occur in these cases, some form of **chromatin remodeling** must take place.

Two general classes of proteins are involved in chromatin remodeling. First are proteins that acetylate the N-terminal tails of the histones, a process that may cause the nucleosomes to bind the DNA less tightly and thus make it available for attachment of transcription factors. These enzymes are called **histone acetyl transferases (HATs).** Deacetylating enzymes have the reverse effect: They act to repress transcription. Second, a class of ATP-dependent proteins such as the SWI/SNF complex in yeast also affect chromatin remodeling. (Some workers called the proteins SWI because they were involved in mating type *swi*tching, and others called them SNF for *s*ucrose *n*on*f*ermenting.) The SWI/SNF complex is a group of eleven proteins involved in transcription activation in many genes, presumably allowing transcription

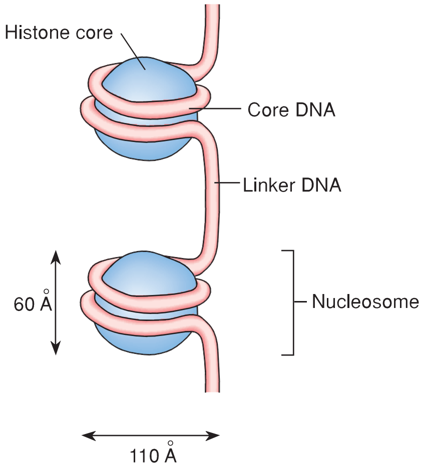


Figure 15.5 The eukaryotic chromosome is associated with histone proteins to form nucleosomes. The protein core is wrapped with 1.7 loops of DNA and connected with a length of DNA called a linker.

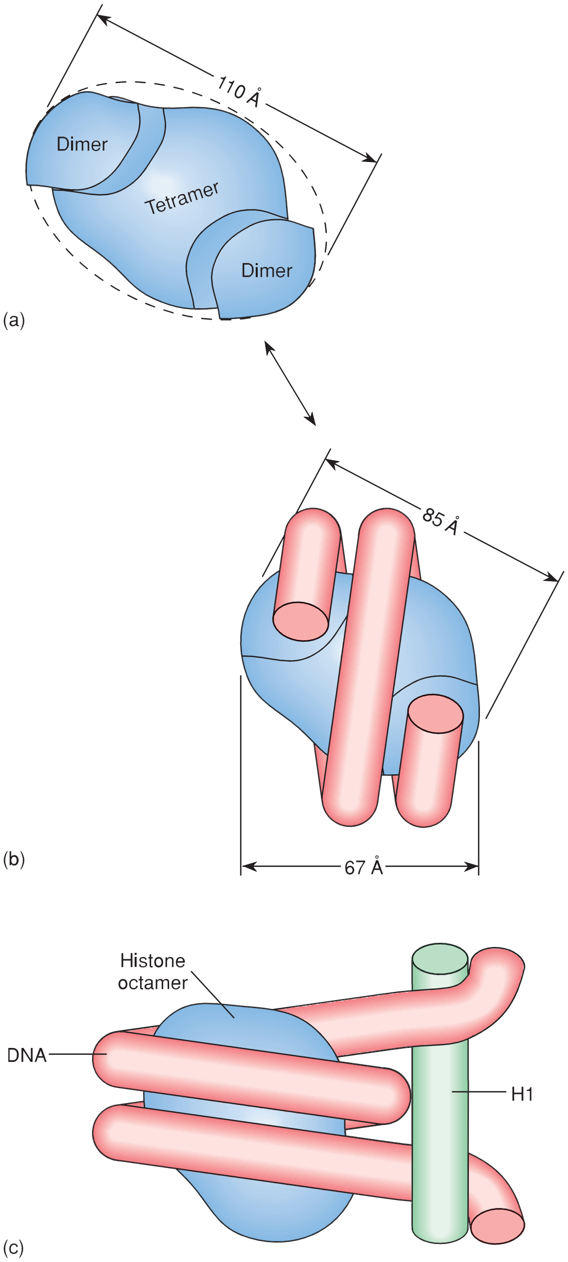


Figure 15.6 Nucleosome structure. (*a*) Schematic comparison of the eight histones comprising the nucleosome in salt solution. A dimer consists of one H2A and one H2B histone molecule; a tetramer consists of two H3 and two H4 histones. (*b*) DNA fits in surface grooves on the more compacted structure found in physiological conditions. (*c*) The diagram shows the presumed position of the H1 histone, encompassing 166 base pairs of DNA.

Table 15.2 Composition of Histones

Fraction	Class	Number of Amino Acids	Percentage of Basic Amino Acids
H1	Very lysine rich	213	30
H2A	Lysine, arginine rich	129	23
H2B	Moderately lysine rich	125	24
H3	Arginine rich	135	24
H4	Arginine, glycine rich	102	27

Figure 15.7 Nucleosome core particle at 2.8 Å resolution. Shown are 146 base pairs of DNA (*brown* and *turquoise*) and the eight histone protein chains (*purple:* H3; *green:* H4; *yellow:* H2A; and *red:* H2B). Note the protein tails (N-terminal ends) of the histone polypeptides extending out of the nucleosome. On the *left* is the view down the DNA helix, and on the *right* is the perpendicular view. (From Karolin Luger, et al., "Crystal structure of the nucleosome core particle at 2.8A resolution" in *Nature,* 389:251–260, September 18, 1997, fig. 1a p. 252. Reprinted by permission of Macmillan, Ltd.)

factors to access promoters by remodeling chromatin. These proteins are able to reposition a nucleosome on DNA by sliding the nucleosome down the DNA.

We thus conclude that although nucleosomes serve as a general, first-order packing mechanism in eukaryotic DNA, they can be positioned precisely and can attenuate transcription. It is interesting to note that once transcription begins, RNA polymerase apparently moves along nucleosomed DNA by translocation of the histones by 75 to 80 base pairs without disrupting the nucleosome itself. This seems to be accomplished by the RNA polymerase moving the DNA and then re-forming the nucleosome in its wake (fig. 15.8).

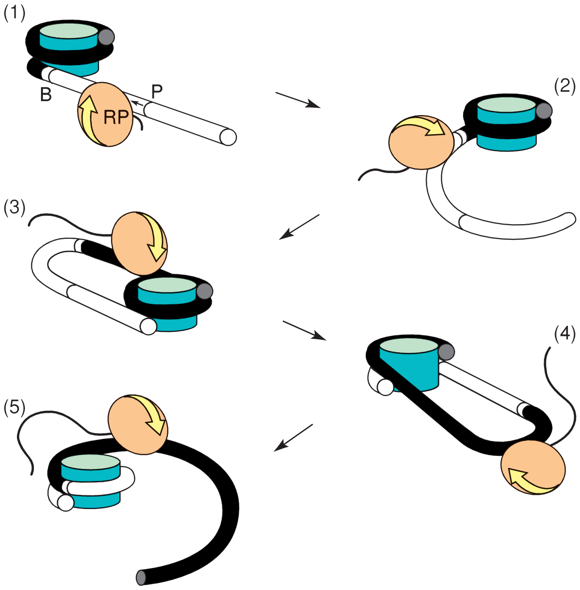


Figure 15.8 RNA polymerase steps around a nucleosome without disrupting it. (*1*) The RNA polymerase begins at a promoter (*P*) and heads for the nucleosomed DNA (filled in), whose border is noted with a line and the letter *B.* As the polymerase encounters the nucleosome, it begins to unwrap the DNA from the histones (*2*). The displaced DNA then reencounters the histones (*3*), about seventy-five to eighty base pairs upstream from the original point of nucleosome formation. The polymerase continues on its way (*4* and *5*), and the nucleosome re-forms in its displaced position without disrupting the histones or ever fully losing contact with the core DNA. (From Vasily M. Studitsky, et al., "A histone octamer can step around a transcribing polymerase without leaving the template," *Cell,* 76: 371–82, January 28, 1994. Copyright © 1994 by Cell Press. Reprinted by permission.)

Higher-Order Structure of Chromatin

Since the nucleosome has a width of only 110 Å, and metaphase chromosomes appear to be constructed of a fiber having a diameter of about 2,400 Å (fig. 15.9), several additional levels of chromatin compaction lead to the metaphase chromosome. Various experiments, which change the ionic strength the chromatin is subjected to, indicate that the 110 Å DNA spontaneously forms a 300 Å, solenoidlike fiber with increased ionic strength. It seems that this fiber results from the coiling of the nucleosomal DNA (fig. 15.10). This 300 Å fiber is not, however, the final form of the DNA. We can account for the contraction of the 300 Å fiber to the 2,400 Å fiber found in metaphase chromosomes by the formation of a second solenoidlike structure from the winding of the 300 Å fiber (fig. 15.11).

If the histones are removed from a chromosome, the DNA billows out, leaving a proteinaceous structure termed a **scaffold** (fig. 15.12). This scaffold structure is formed from **nonhistone proteins;** two of them predominate, namely SC1 and SC2. SC1 has been identified as topoisomerase II. It would not be unreasonable to expect several hundred different proteins, many in minute quantities, to be associated with the chromosome and involved in replication, repair, and transcription.

Figure 15.9 Chinese hamster chromosome. Note the fibers making up the chromosome; they are approximately 2,400 Å in diameter. Magnification 11,800×. (Source: Courtesy of Dr. Hans Ris.)

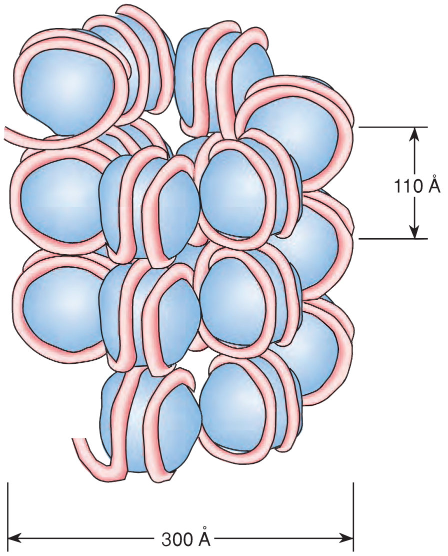


Figure 15.10 Solenoid model for the formation of the 300 Å chromatin fiber. Nucleosomal DNA wraps in a helical fashion, forming a hollow core. Although histone H1 is not shown, it is known to be on the inside of the solenoid.

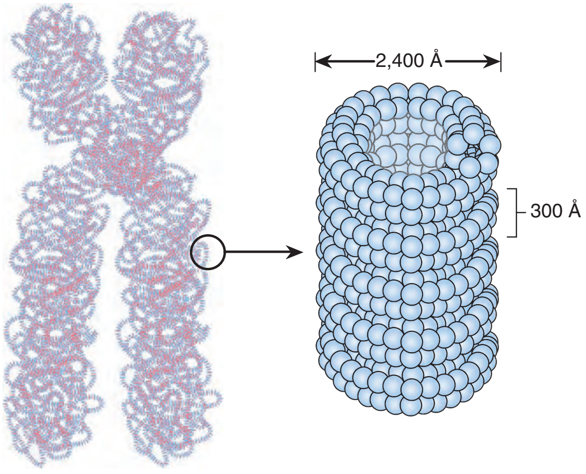


Figure 15.11 The 2,400 Å fiber of the eukaryotic chromosome is a hollow, solenoidlike structure. It is formed by the coiling of the 300 Å fiber, which itself is a solenoid.

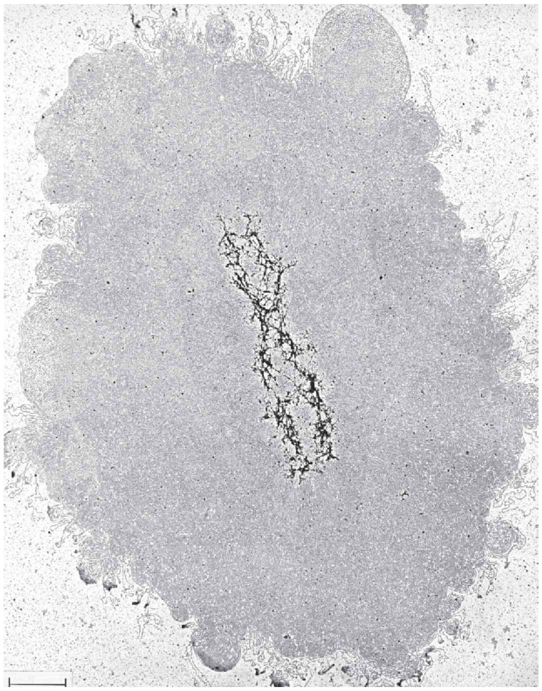


Figure 15.12 Scaffold protein. When the histones are removed from a eukaryotic chromosome, a fibrous scaffold remains. The DNA loops out from this scaffold. The bar is 2 μm long. (J. Paulson and U. Laemmli, "The structure of histone-depleted metaphase chromosomes," *Cell,* 12:817–28, 1977. Micrograph courtesy of James R. Paulson.)

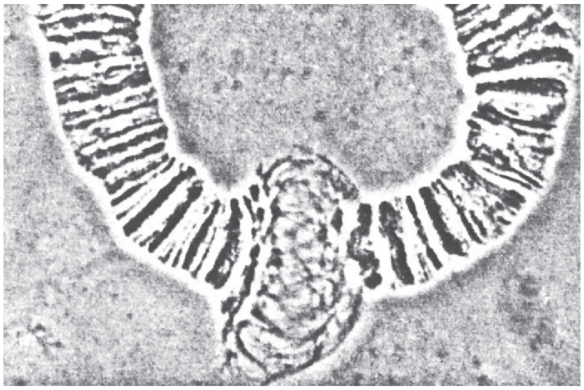


Figure 15.13 A chromosome puff on the left arm of chromosome 3 of the midge *Chironomus pallidivittatus.* (Jan-Erik Edström, et al., *Developmental Biology* 91:131–37, 1982, Figure 1B, Academic Press.)

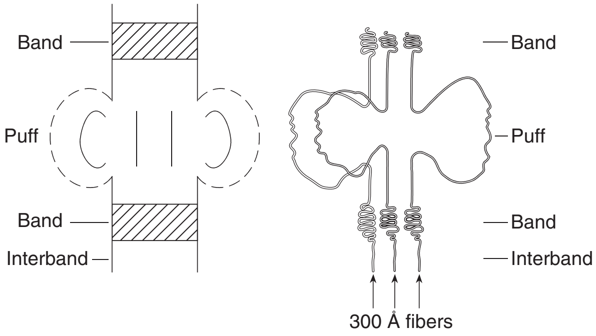


Figure 15.14 Polytene chromosome with bands and a puff. Three of the approximately one thousand synapsed chromatids are shown diagrammatically on the *right.*

Polyteny, Puffs, and Balbiani Rings

Drosophila's salivary glands, as well as some other tissues of *Drosophila* and other diptera, contain giant banded chromosomes (see fig. 6.12) that result from the replication of the chromosomes and the synapsis of homologues without cell division (endomitosis). These chromosomes consist of more than one thousand copies of the same chromatid and appear as alternating dark bands and lighter interband regions. The dark bands are referred to as *chromomeres.* Also seen are diffuse areas called **chromosome puffs** (fig. 15.13). Chromosome puffs are also referred to as **Balbiani rings.** These rings were originally defined as puffs in the midge, *Chironomus,* whose polytene chromosomes were discovered by E. G. Balbiani in 1881. Currently, the term applies to all puffs, or at least the larger puffs, in all species with polytene chromosomes.

The structure of the polytene chromosome can be explained by the diagram in figure 15.14. Dark bands (chromomeres) are due to tight coiling of the 300 Å fiber; light interband regions are due to looser coiling. The figure also shows how chromosome puffs would come about as fibers unfold in regions of active transcription.

Staining with reagents specific for RNA, such as toluidine blue, or autoradiography with tritiated (^{3}H) uridine, have been used to demonstrate that active transcription is going on in the puffs but not in neighboring regions of the polytene chromosomes. The messenger RNA isolated from cells with puffs has also been shown to hybridize only to the puffed regions of the chromosomes. Thus, these regions of the DNA are complementary to the messenger RNA (fig. 15.15) and represent areas of active transcription. Modern recombinant DNA techniques have also shown that many puffs probably represent the transcription of only one gene, although there are exceptions.

Puffs generally fall into four categories. *Stage-specific puffs* appear during a certain stage of development, such as molting. *Tissue-specific puffs* are active in one tissue but not another. (In dipteran larvae, tissues other than the salivary glands, such as the midgut and Malpighian tubules, have polytene chromosomes.) *Constitutive puffs* are active almost all the time in a specific tissue. And *environmentally induced puffs* appear after some environmental change, such as heat shock (fig. 15.16). In *Drosophila,* about 80% of the puffs are stage specific; in *Chironomus,* only about 20% are. For example, at the time of molt in insects, the hormone ecdysone is secreted by the prothoracic gland. At the same time, many puff patterns change (fig. 15.17). Similar changes in puff patterns can be induced by the injection of ecdysone. Hence, molting, a stage-specific developmental sequence, is related to a sequential transcription sequence in the chromosomes.

Lampbrush Chromosomes

Lampbrush chromosomes, which occur in amphibian oocytes, are so named because their looped-out configuration has the appearance of a brush for cleaning lamps, now a relatively uncommon household item (fig. 15.18). The loops of the lampbrush chromosomes are covered by an RNA matrix and are the sites of active transcription. Presumably, the loops are unwindings of the single chromosome, similar to the unwindings in the polytene chromosome shown in figure 15.14. Thus, under certain circumstances, such as in polytene chromosomal puffs

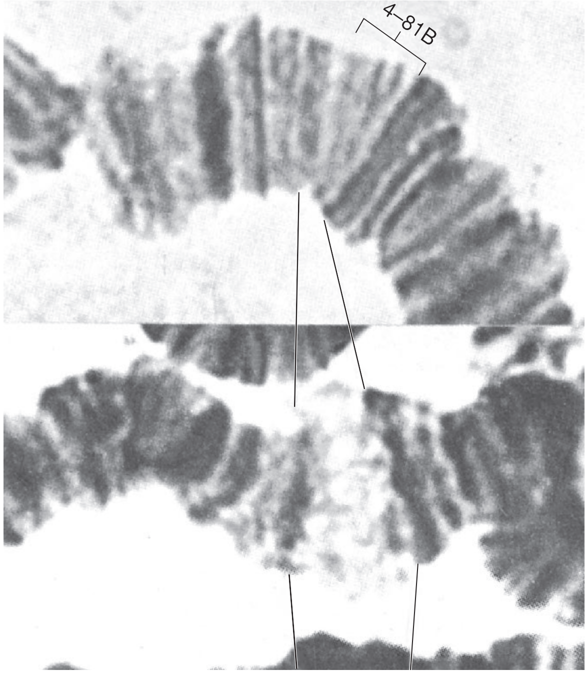


Figure 15.16 Puff 4-81B of the salivary gland in *Drosophila hydei* is induced by heat shock (37° C for one-half hour). At the *top,* normal activity. At the *bottom,* temperature shock in vitro, resulting in the puff. (Source: H. D. Berendes, et al., "Experimental puffs in salivary gland chromosomes of *Drosophila hydei,*" *Chromosoma* [Berl.] 16:35–46, Fig. 4a–b, 1965. © Springer-Verlag.)

Figure 15.15 Hybridization at a *Chironomus tentans* salivary gland chromosome puff. The chromosomal DNA is hybridized with labeled RNA (*black dots*) transcribed from the locus. The activity of the locus is forming the puff. (Reprinted by permission from B. Lambert, "Repeated DNA sequences in a Balbiani ring," *Journal of Molecular Biology,* 72:65–75, 1972. Copyright by Academic Press, Inc. (London) Ltd.)

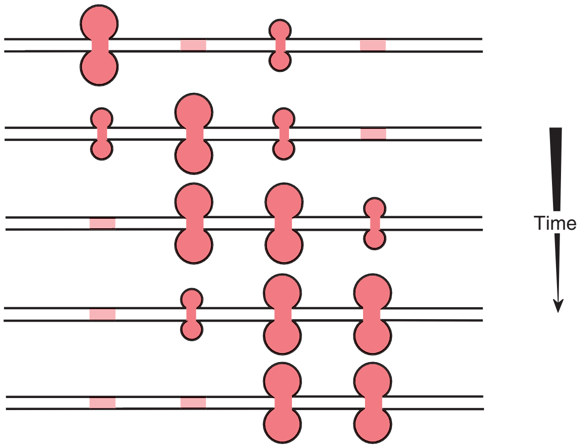


Figure 15.17 Puff patterns on a segment of a *Chironomus tentans* salivary gland chromosome during molt. As time proceeds, puffs appear and disappear and change in size.

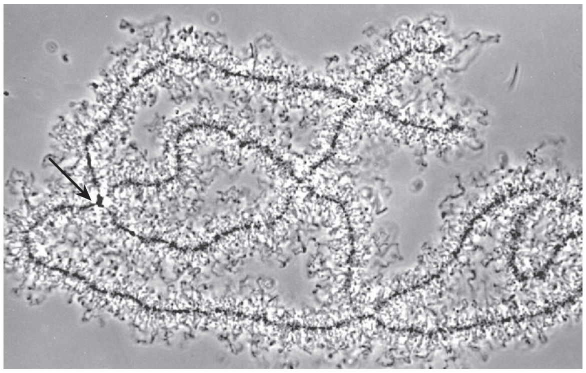


Figure 15.18 Lampbrush chromosome of the newt, *Notophthalmus viridescens.* Centromere is at the the *left* (*arrow*); the two long homologues are held together by three chiasmata. Magnification 238×. (Source: Joseph G. Gall, figure 2 in D. M. Prescott, ed., *Methods in Cell Physiology,* vol. 2 [New York: Academic Press, 1966], 39. Reproduced by permission.)

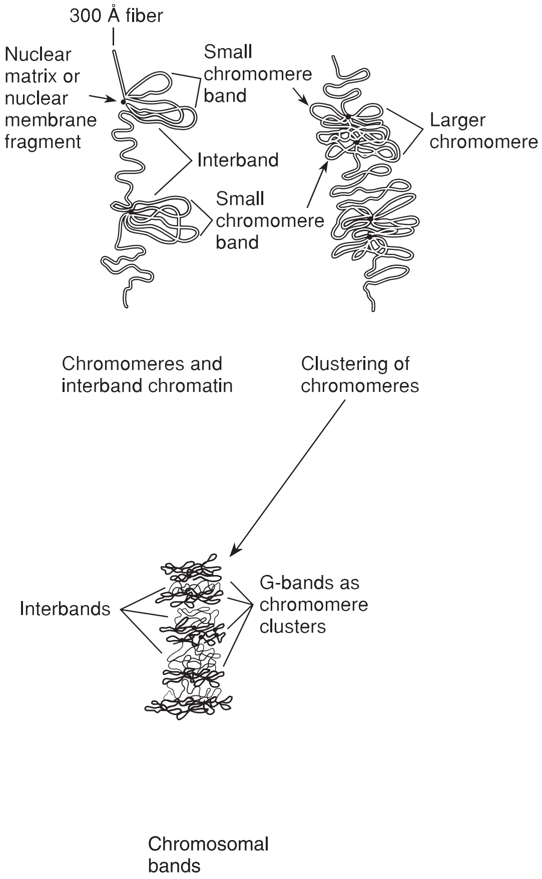


Figure 15.19 Model of eukaryotic (mammalian) chromosomal banding. G-bands are chromomere clusters, which result from the contraction of smaller chromomeres. These, in turn, result from looping of the 300 Å fiber. (Reproduced with permission, from the *Annual Review of Genetics,* Volume 12, © 1978 by Annual Reviews, Inc.)

and in lampbrush chromosomes, active transcription can be seen in the light microscope. Since only certain bands puff at any one moment in polytene chromosomes, and since the loops of lampbrush chromosomes are of various sizes (with some regions not looped at all), we have evidence of specific transcription. However, we have no indication, so far, of the nature of the control of that transcription.

Chromosomal Banding

Several chromosomal staining techniques reveal consistent banding patterns. By means of these patterns, all of the human chromosomes can be differentiated (see fig. 5.1). Of possibly greater importance is the fact that these staining techniques have provided some insight into the structure of the chromosome. The techniques for staining the C, G, and R chromosomal bands will serve as an illustration.

G-bands are obtained with **Giemsa stain,** a complex of stains specific for the phosphate groups of DNA. Treatment of fixed chromatin with trypsin or hot salts brings out the G-bands. Giemsa stain enhances banding that is already visible in mitotic chromosomes. The banding pattern is caused by the arrangement of chromomeres. Under careful observation, the major G-bands prove to consist of many smaller chromomeres. This banding appearance has led D. Comings to suggest the mechanism of chromosomal folding shown in figure 15.19.

C-bands are Giemsa-stained bands after the chromosomes are treated with NaOH. The *C* is for "centromere," because these bands represent constitutive heterochromatin surrounding the centromeres (fig. 15.20). The DNA is also usually satellite rich. **Satellite DNA** differs in **buoyant density** from the major portion of cellular DNA. When eukaryotic DNA is isolated and centrifuged in CsCl, forming a density gradient, the majority of the DNA forms one band in the gradient at a single buoyant density. The buoyancy is determined by the G-C content of the DNA. However, smaller secondary bands are also usually present, indicating regions of DNA having sequences different from the majority of the cell's DNA (fig. 15.21). DNA isolated this way is referred to as *satellite DNA* because of the secondary, or satellite, bands formed in the density gradient. As we will see, this DNA is found primarily around centromeres and consists of numerous repetitions of a short sequence.

R-bands are visible with a technique that stains the regions between G-bands. The chromosomes are fixed, stained with Giemsa, and then viewed with a phase contrast microscope. Since the dark-light pattern is the opposite of the G-band pattern, these bands are called *reverse bands.*

From the information gleaned from these staining techniques, D. Comings distinguished between three basic chromatin types: euchromatin, constitutive heterochromatin, and intercalary heterochromatin (table 15.3). Presumably, the only chromatin involved in transcription is **euchromatin. Constitutive heterochromatin** surrounds the centromere and is rich in

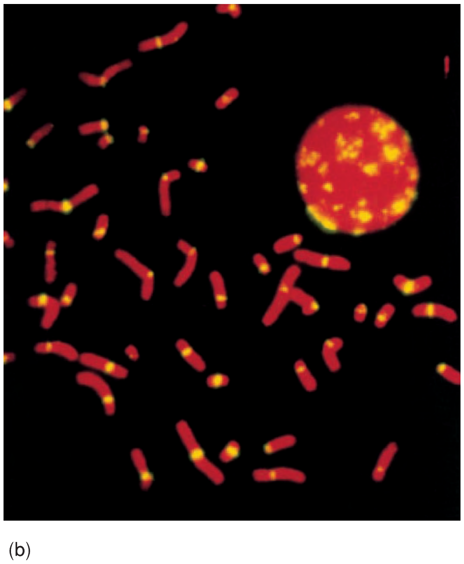
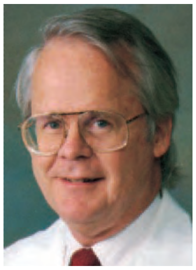

(b)

Figure 15.20 (*a*) C banding of chromosomes from a cell in the bone marrow of the house mouse, *Mus musculus.* The *arrow* indicates that the Y chromatids have already separated into two chromosomes. (*b*) Yellow fluorescence indicates a satellite DNA probe in human chromosomes (centromeres). ([*a*] B. Vig, "Sequence of centromere separation: Role of centromeric heterochromatin," *Genetics,* 102:795–806, 1982. [*b*] Photograph Courtesy of Oncor, Inc. Gaithersburg, Maryland.)


David E. Comings (1935–).
(Courtesy Dr. David E. Comings.)

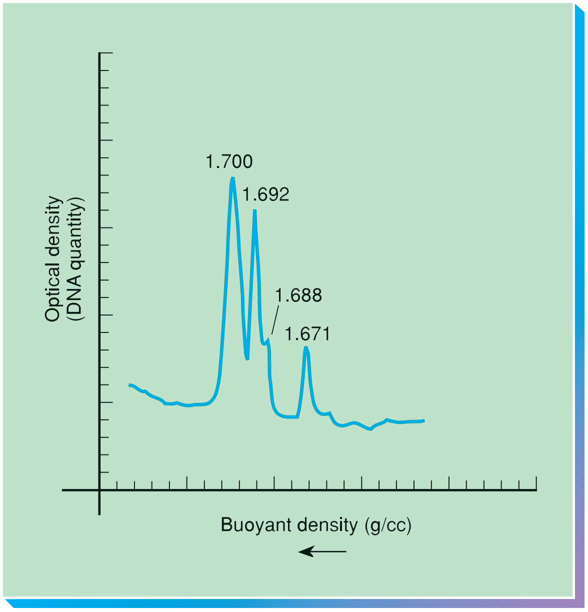


Figure 15.21 Satellite DNA in *Drosophila virilis.* The quantity of DNA is graphed against the buoyant density (g/cc), resulting in four peaks. The large peak (at *left*) is the major DNA component of the cell; the other three bands are satellite DNA. The left-most of the satellite peaks (1.692) is DNA with a repeating sequence of ACAAACT; the middle satellite peak (1.688) is a sequence of ATAAACT; and the right-most satellite peak (1.671) has a sequence of ACAAATT. (From Joseph G. Gall, et al., *Cold Spring Harbor Laboratory Symposia on Quantitative Biology,* 38:417–21. Copyright © 1974 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. Reprinted by permission.)

Table 15.3 **The Three Major Types of Chromatin in Eukaryotic Chromosomes**

	Euchromatin	Centromeric Constitutive Heterochromatin	Intercalary Heterochromatin
Relation to bands	In R-bands	In C-bands	In G-bands
Location	Chromosome arms	Usually centromeric	Chromosome arms
Condition during interphase	Usually dispersed	Condensed	Condensed
Genetic activity	Usually active	Inactive	Probably inactive
Relation to chromomeres	Interchromomeric	Centromeric chromomere	Intercalary chromomeres

satellite DNA. **Intercalary heterochromatin** is found throughout the chromosome. Thus, it becomes apparent that the eukaryotic chromosome is a relatively complex structure.

Centromeres and Telomeres

Centromeres

Two regions of the eukaryotic chromosome have specific functions—the centromere and the telomeres. The centromere is involved in chromosomal movement during mitosis and meiosis, whereas the telomeres terminate the chromosomes. As we pointed out in chapter 3, the terms *centromere* and *kinetochore,* while occasionally used interchangeably, are distinct. The kinetochore is the interface between the visible constriction in the chromosome (the centromere) and the microtubules of the spindle. The kinetochore of higher organisms (e.g., mammals) contains proteins and some RNA. Microscopically, it is a trilaminar structure, attached to chromatin at the inner layer and to microtubules at the outer layer (fig. 15.22).

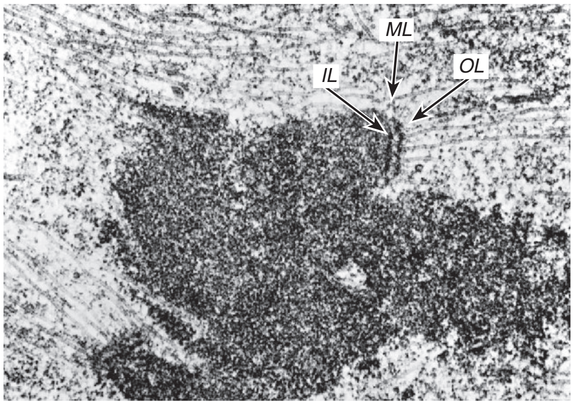


Figure 15.22 The kinetochore of a metaphase chromosome of the rat kangaroo. *IL, ML,* and *OL* refer to *inner, middle,* and *outer layers,* respectively, of the kinetochore. Note the microtubules attached to the kinetochore and the large mass of dark-staining chromatin making up most of the figure. Magnification 30,800×. (From B. R. Brinkley and J. Cartwright, Jr., *J. Cell Biology,* 50:416–31, 1971.)

Most of our knowledge of the genetics of centromeres has come from work in yeast *(Saccharomyces cerevisiae).* Cells did not maintain most artificially created yeast plasmids because they were lost during mitosis. However, plasmids were isolated that did replicate normally during cell division. Presumably, they contained centromeres, allowing them to replicate and move in synchrony with the host's chromosomes. Further genetic engineering made it possible to isolate smaller and smaller regions that could serve as centromeres. After sequencing the centromeres of fifteen of the sixteen yeast chromosomes, it was possible to conclude that the centromere from yeast is about 250 base pairs long with three consensus regions (fig. 15.23); we are defining a centromere as a sequence of DNA called the *CEN* locus or CEN region. Recent data indicate that this region may contain a single, modified nucleosome associated with region II. The 250 base-pair length of the CEN regions of yeast chromosomes is about 200 Å, the same as the diameter of a microtubule, indicating that only one microtubule attaches to each centromere during mitosis or meiosis in a yeast cell. This region is called a **point centromere** (fig. 15.24).

Higher eukaryotes have larger centromeric regions that attach more microtubules. These regions are referred to as **regional centromeres** (see figs. 15.22 and also 3.12). Regional centromeres range from nineteen to one hundred kilobases (kb; 19,000–100,000 bases) with unique and satellite (repeated sequence) DNA that is heterochromatic and may include expressed genes. We know much less about regional centromeres than we do about point centromeres.

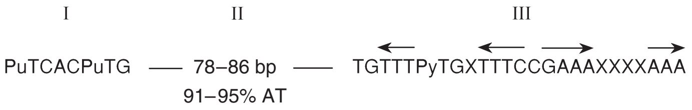


Figure 15.23 Consensus sequence for the three regions (I–III) of fifteen yeast centromeres. *Pu* represents any purine, *Py* represents any pyrimidine, and *X* represents any base. The *arrows* appear over inverted repeat sequences. (Source: Data from L. Clarke and J. Carbon, "The structure and function of yeast centromeres," *Annual Review of Genetics,* 19:29–56, 1985.)

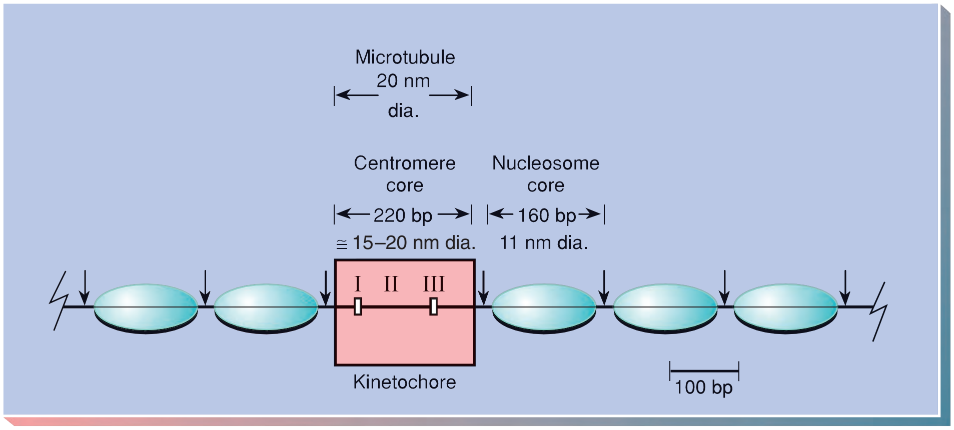


Figure 15.24 Schematic view of a yeast centromeric region. The *arrows* are the nuclease-hypersensitive sites. A microtubule is about the same width as the centromeric region. (With permission, from the *Annual Review of Genetics,* Volume 19 © 1985 by Annual Reviews www.AnnualReviews.org)

Telomeres

Since eukaryotic chromosomes are linear, each has two ends, referred to as **telomeres,** that not only mark the termination of the linear chromosome but also have several specific functions (fig. 15.25). Telomeres must prevent the chromosomal ends from acting in a "sticky" fashion, the way that broken chromosomal ends act (see chapter 8). In other words, chromosomal ends must not elicit a DNA repair response (see chapter 12). Telomeres must also prevent the ends of chromosomes from being degraded by exonucleases and must allow chromosomal ends to be properly replicated.

Most telomeres isolated so far are repetitions of sequences of five to eight bases. In human beings, the telomeric sequence is TTAGGG, repeated 300 to 5,000 times at the end of each chromosome. The human telomere was discovered by R. Moyzis and his colleagues when they probed the highly repetitive segment of human DNA. (Highly repetitive DNA, as its name implies, consists of numerous copies of a single sequence and usually comprises the satellite components of the cell's DNA; see next section.) When a probe for this sequence was applied to human chromosomes, the sequence was found at the tip of each chromosome in roughly the same quantity (fig. 15.26). This is a highly conserved sequence, found in all vertebrates studied as well as in unicellular trypanosomes. Similar sequences are found in various other eukaryotes (table 15.4); the first sequence was isolated by E. Blackburn and J. Gall in 1978.

When a linear DNA molecule is replicated, the $3' \rightarrow 5'$ strand can be replicated to the end (see chapter 9). The $5' \rightarrow 3'$ strand, however, is replicated with RNA primers that are then degraded, leaving a short gap on the progeny strand (fig. 15.27). It is always the G-rich strand of telomeric DNA that ends up single-stranded, forming a $3'$ overhang of twelve to sixteen nucleotides. Thus, the normal replication process of a linear DNA molecule leaves an incomplete terminus. Hence, scientists suspected that there would be a unique mechanism for the replication of telomeres.

Telomeric sequences appear to be added de novo without, DNA template assistance by an enzyme called **telomerase,** discovered by E. Blackburn and her colleagues. This was seen when telomeres from another species were engineered into yeast cells. After a cell cycle, the yeast telomeric sequence had been added on at

Figure 15.25 Polytene chromosome from the salivary gland of a *Drosophila* larva showing bands, interbands, puffs, and telomeres. (© David M. Phillips/Visuals Unlimited.)

Elizabeth H. Blackburn (1948–). (Courtesy of Dr. Elizabeth H. Blackburn.)

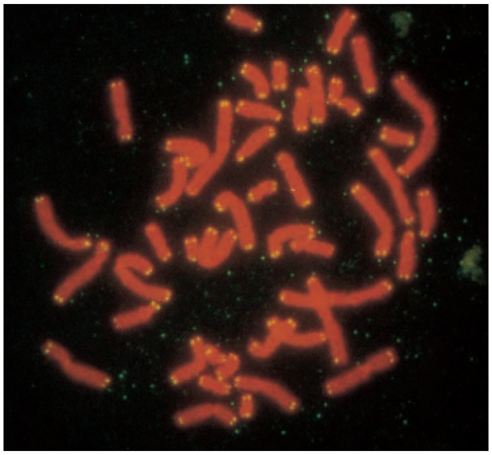


Figure 15.26 The human genome probed for the telomeric sequence, TTAGGG, using fluorescent staining techniques. The *yellow dots* at the tips of the chromosomes are the probes. (From Robert K. Moyzis, et al., *Proceedings of the National Academy of Science, USA,* 85:6622–26, 1988. Figure 4, left.)

Table 15.4 **Telomeric Sequences in Eukaryotes; The G-Rich Strand of the Double Helix Is Shown**

Organism	Telomeric Repeat
Human beings, other mammals, birds, reptiles	TTAGGG
Trypanosomes	TTAGGG
Holotrichous ciliates (*Tetrahymena*)	GGGGTT
Hypotrichous ciliates (*Stylonychia*)	GGGGTTTT
Yeast	GT, GGT, and GGGT
Plants	TTTAGGG

the ends of the foreign chromosome, the result, presumably, of the telomerase enzyme.

When Blackburn and her colleagues isolated telomerase, they discovered that a segment of RNA, about 160 base pairs, is an integral part of the enzyme. That RNA has a region that is complementary to the G-rich repeat of the telomeric DNA sequence of the species. After careful experimentation, including modifying the gene for the telomerase RNA, Blackburn and her colleagues concluded that telomerase uses its RNA as a template for adding telomeric repeats to the ends of chromosomes. Telomerase is thus a reverse transcriptase, using RNA nucleotides as a template to polymerize DNA nucleotides.

Blackburn and her colleagues proposed that the first step in telomere extension is hybridization of the 3′ end of the telomere with the RNA component of telomerase (fig. 15.28*a*). Then, with the telomerase RNA as a template, the 3′ end of the telomere is extended (fig. 15.28*b*). Finally, a translocation step takes place that displaces the telomere in respect to the RNA, returning to the configuration at the beginning of the process (fig. 15.28*c*). The single-stranded C-rich strand is then synthesized with DNA polymerase and DNA ligase.

Once telomeres have been added to the ends of eukaryotic chromosomes, different organisms use any of three different methods known to protect the ends of

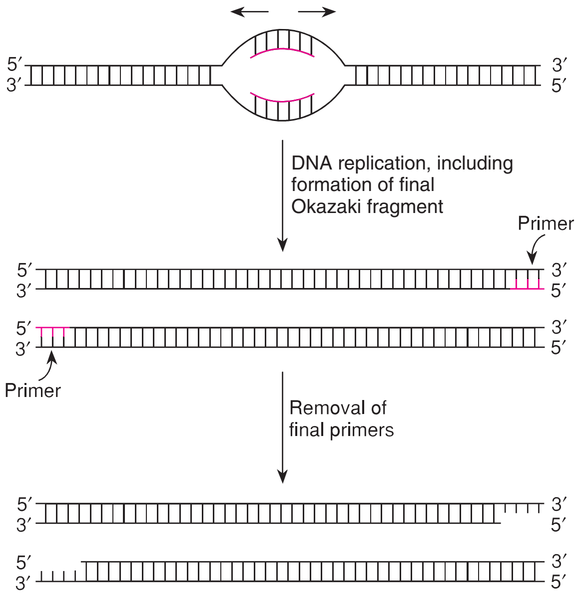


Figure 15.27 Removal of final primers after the replication of linear DNA creates single-stranded ends.

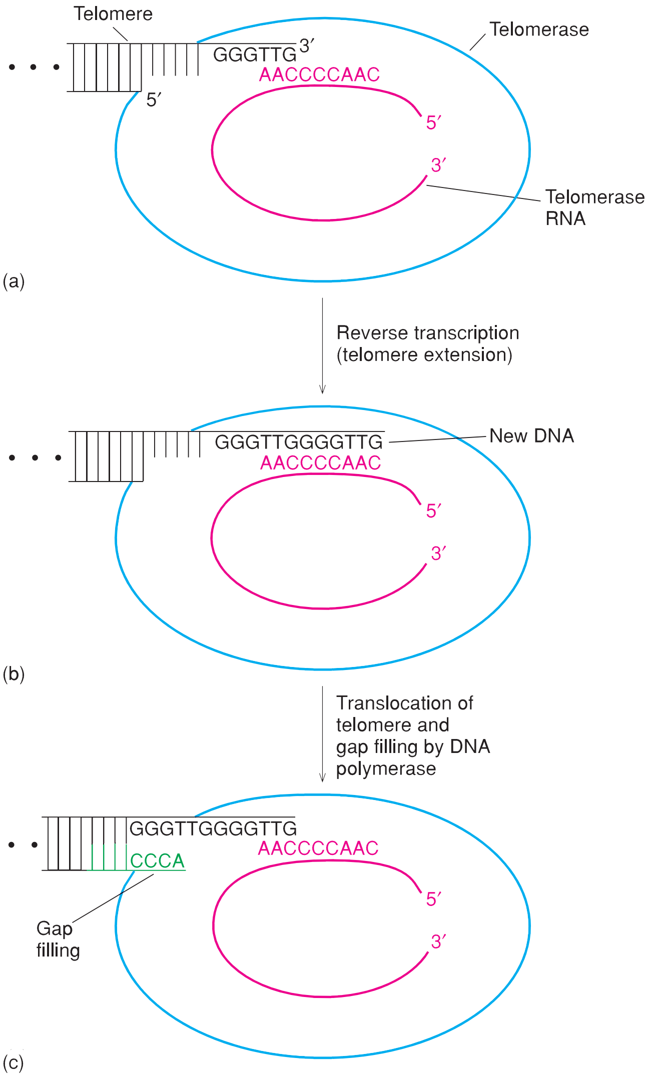


Figure 15.28 Telomerase extends telomeres using telomerase RNA (*red*) as a template. Gap filling (*green*) by DNA polymerase and ligase complete the double helix. (Source: Data from Shippen-Lentz and Blackburn, *Science,* 247:550, 1990.)

the chromosomes. First, the guanine-rich DNA can form complex structures. Biochemists have discovered that four guanines can form a planar **G-tetraplex,** with the four bases hydrogen bonded to each other (fig. 15.29). Several structures have been hypothesized to explain the novel ends of these chromosomes (fig. 15.30). Second, proteins have been discovered that bind to the 3′ ends of telomeres. In the ciliate *Oxytricha nova,* a protein called the telomere end-binding protein (TEBP) attaches to the 3′ ends of telomeres and protects them (fig. 15.31). Finally, a novel structure called the ***t-loop*** has been discovered at the end of mammalian telomeres. This loop forms at the ends of chromosomes under the direction of a protein called TRF2 (*t*elomere *r*epeat-binding *f*actor), which causes the 3′ end of the chromosome to loop around and interdigitate into the double helix, forming the loop (fig. 15.32).

How do cells keep track of the number of their telomeric repeats? Proteins have been isolated that bind to telomeres (Rap1 in *Saccharomyces cerevisiae,* TRF1 in human beings). By mutating these proteins or the telomeric sequences, scientists have changed the equilibrium number of telomeric repeats. This led to the current model that the cell counts the number of these proteins bound to the telomeres, not the number of telomeres directly, to know whether telomeres should be added. This is a very active area of research.

In yeast, protozoa, and other single-celled organisms, telomerase is active, keeping the ends of the chromosomes at the appropriate lengths. These cells can divide potentially forever. However, in most cells of higher organisms, telomerase is not active, and the ends of the chromosomes get shorter with each cell division. At a certain telomeric length, the cells no longer divide. However, if telomerase becomes active, and the ends of the chromosomes lengthen, a signal is conveyed to keep cells dividing, which can lead to cancerous growth. In fact, human telomerase was isolated from an immortal cell line (HeLa) derived from cervical cancer cells. Thus,

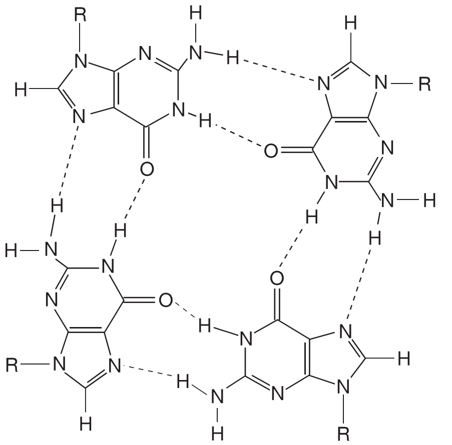


Figure 15.29 A G-tetraplex can form from four guanines in a plane, hydrogen bonded with each other. (Source: Data from Yong Wang and Dinshaw J. Patel, "Solution structure of the human telomeric repeat d[AG$_3$ (T$_2$AG$_3$)$_3$] G-tetraplex," *Structure,* 1:263–82, December 15, 1993.)

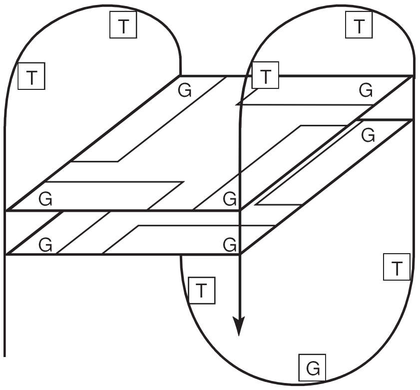


Figure 15.30 Based on G-tetraplexes (fig. 15.29), the illustrated structure can form at the very tip of a telomere. The sequence d(GGTTGGTGTGGTTGG) is shown forming a four-stranded structure. (Reproduced, with permission, from the *Annual Review of Biophysics and Biomolecular Structure,* Volume 23, © 1994 by Annual Reviews, Inc.)

attention is now turning to the possible clinical application of this knowledge: If telomerase can be deactivated in tumor cells, the cells may stop dividing or die, thereby eliminating the cancer. Further, studying normal telomere shortening, which appears to act as a biological clock, may help us understand the aging process and senescence.

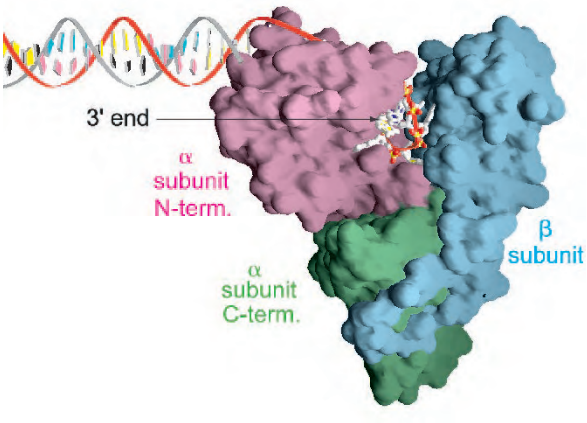


Figure 15.31 Telomere of the ciliate *Oxytricha nova* shown bound by the dimeric protein called telomere end-binding protein (TEBP). The α and β subunits of the protein form a deep cleft in which the 3′ end of the telomere lies. The folding of the protein into its final form around the DNA may only occur after the DNA has bound, explaining how the DNA could be recognized and placed into such a deep cleft. (Reprinted courtesy of Dr. Martin Horvath.)

The C-Value Paradox

Why do eukaryotes have so much DNA, and why is there huge variation in the DNA content between species of comparable complexity? These questions define the **C-value paradox,** in which *C* refers to the quantity of DNA in a cell. For an example of the paradox, although human beings have 3.3 billion base pairs in the haploid genome, an amoeba has more than 200 billion base pairs. And although an average bony fish has over 300 billion base pairs of DNA in its haploid genome, the Japanese puffer fish has less than half a billion base pairs. If the basic bony fish pattern can be created with less than half a billion base pairs, why does the average bony fish have over 600 times that much DNA? What is this excess DNA doing? To explain the C-value paradox, researchers examined the repetitiveness of DNA, and more recently, probed and sequenced DNA to understand its properties.

DNA-DNA Hybridization

R. Britten and his colleagues, using the technique of **DNA-DNA hybridization,** first systematically analyzed the repetitiveness of the DNA within eukaryotes. When DNA is heated, it denatures or unwinds into single strands; when it cools, it renatures. The rate of renaturation depends on the DNA sequences. If the sample contains DNA with repeated sequences, it will hybridize

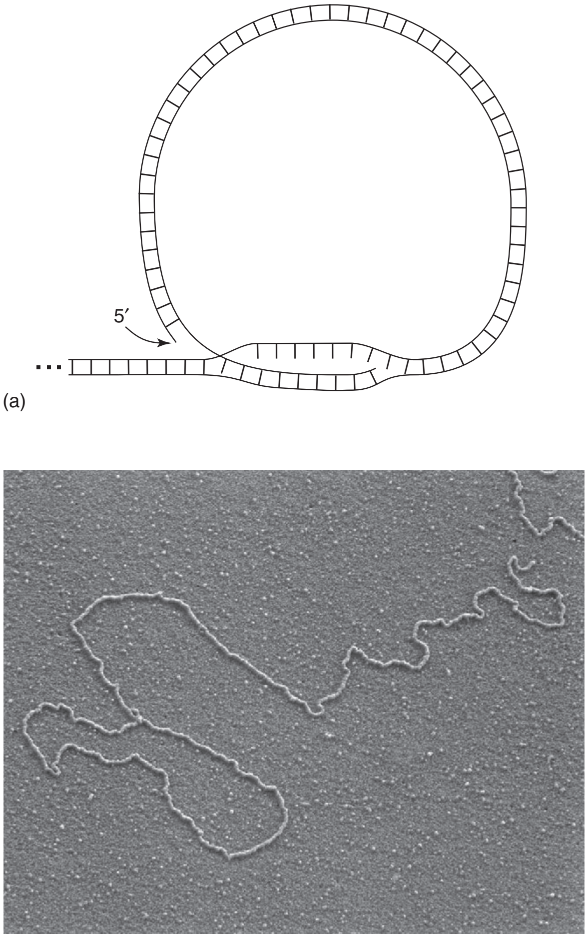
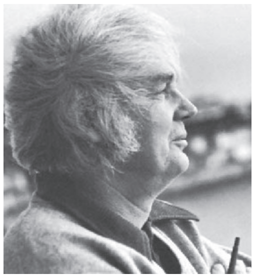


Figure 15.32 The t-loop at the end of the mammalian telomere. (*a*) A diagram of how the t-loop is formed by the interdigitation of the 3′ end of the telomere into the double helix. (*b*) Electron micrograph of a t-loop from a mouse liver cell. The loop is about 10,000 bases around. ([b] From Jack D. Griffith, et al., "Mammalian telomeres end in a large duplex loop" in *Cell,* 97:503–14, May 14, 1999. Copyright © Cell Press.)

faster than DNA that does not have repeated sequences. From these studies, Britten and his colleagues found that eukaryotic chromosomes contain regions of unique, moderately repetitive, and highly repetitive DNA. **Unique DNA** is, as its name implies, DNA with unrepeated sequences. **Repetitive DNA** is DNA whose sequences are repeated in the genome.

Satellite DNA, found around centromeres (see fig. 15.20), is highly repetitive DNA with a unique repeat length of about two hundred base pairs. Given the quantity of satellite DNA per cell, there must be more than one million repetitions of this two-hundred-nucleotide sequence in higher eukaryotes. At the other end of the spectrum is unique DNA, which makes up most of the transcribed genes of an organism. The rest of the DNA is repetitive DNA in a few to several hundred thousand copies. This repetitive DNA comprises at least three categories. One is "junk" DNA, DNA that is not useful to the organism, made up of untranscribed and parasitic sequences (selfish DNA). Another category is transcribed genes in many copies that have diverged from each other, such as antibody, collagen, and globin genes. We use the term **gene family** to refer to genes that have arisen by duplication, with or without divergence, from an ancestral gene. And finally, transcribed genes in many copies that are virtually identical, such as ribosomal RNA and histone genes, make up a third category of repetitive DNA.



Roy J. Britten (1919–). (Courtesy of Dr. Roy J. Britten.)

Junk DNA

We saw in chapter 13 that transposons in prokaryotes are generally viewed as selfish or parasitic: They serve no purpose to the cell. The transposons replicate on their own, increasing in number. Eukaryotic transposons are mostly **retrotransposons,** transposable elements that move by way of an RNA intermediate. That is, the retrotransposon is transcribed into RNA and then, by reverse transcription, converted to a cDNA that is then inserted into the genome. These elements can make up 50% of the eukaryotic genome, existing in hundreds of thousands of copies. They generally fall into two categories: LINES and SINES. **Long interspersed elements (LINES),** are up to seven thousand base pairs each and contain genes for reverse transcription, RNA binding, and endonuclease activity. They thus have the ability to jump by way of an RNA intermediate. Human DNA is believed to be composed of about 15% LINES.

Short interspersed elements (SINES) are generally derivatives of transfer RNA genes and do not have the ability to retrotranspose on their own. That is, in the past, their transcripts were modified, converted to cDNA by reverse transcription, and then reinserted into the host's genome. They rely on the reverse transcriptase provided by the genes of LINES or retroviruses. One group of SINEs not derived from transfer RNA is derived from the RNA of the signal recognition particle (see chapter 11); members of this group occur in human beings in about five-hundred thousand copies of a three-hundred-base-pair sequence. Because these sequences are cleaved by the restriction endonuclease *Alu*I, they are called the **Alu family.** The human genome is also permeated by remnants of at least a dozen distinct families of ancient retroviruses scattered throughout our chromosomes.

At this point, we can see some potential explanations for the C-value paradox. Much eukaryotic DNA is junk, apparently doing no harm. In some cases, 97% of the host genome is composed of junk DNA. Recent work seems to indicate that gross differences in DNA content between higher organisms may be due to the differing abilities of different species to rid themselves of this parasitic DNA. If it builds up without being removed, the DNA content of the species can soar. Thus, the wide differences in DNA content among higher eukaryotes mentioned at the beginning of this section have little to do with the complexity of the organism, but rather with the ability of the organism to remove junk DNA as it forms.

Expressed Genes in Many Copies

Several types of genes create a product that is needed in such large quantity that one copy of the gene could not fulfill the cell's needs. We are familiar with the nucleolus, the site of the ribosomal RNA genes (see fig. 10.20). Human beings have about two hundred copies of the major ribosomal RNA gene and about two thousand copies of the 5S ribosomal RNA gene. Fruit flies have about two hundred and one hundred copies, respectively, of the two genes.

In some cases, the normal number of multiple copies of a gene is still not enough. The cell must then resort to **gene amplification,** a process whereby the cell increases the number of copies of the gene. For example, during oogenesis, ribosomal RNA genes (rDNA) are often amplified. In *Xenopus,* rDNA is amplified about one thousand times, which allows an oocyte to accumulate about 10^{12} ribosomes. The amplified DNA is in the form of small, circular, extrachromosomal molecules of DNA. Several models have been proposed as to how cells actually amplify their DNA. One model relies on unequal crossing over (as in *Bar* eye in *Drosophila*), whereas another model is based on unscheduled extra DNA replication in a region, followed by recombinational events that generate linear and circular forms of the excess DNA. It is not presently clear which model is correct.

In addition to ribosomal RNA genes, other genes are repeated, ensuring adequate gene products. The number and location of repeated genes are usually discovered by hybridization studies using probes, similar to the way that telomeric DNA was shown to be at the tips of the chromosomes (see fig. 15.26). Repeated genes include the genes for transfer RNAs and histones. The average transfer RNA is repeated about a dozen times in *Drosophila.* Human beings have over thirteen hundred copies of transfer RNA genes in the haploid genome. In many species, the five histone genes form a repeated cluster, although each gene is transcribed independently (fig. 15.33), while prokaryotic operons are transcribed as a unit. The arrangement of histone genes may be more complex in higher forms. There are indications that in mammals, histone genes may lie in small groups or even as individual genes.

Several types of genes occur in similar but not identical forms—that is, an original gene was duplicated but, unlike histone or ribosomal RNA genes, the copies diverged in function. These gene families include globin genes, immunoglobulin genes (see chapter 16), chorion protein (insect eggshell) genes, and *Drosophila* heat shock genes.

The Globin Gene Family

Globins are oxygen-transporting and storage molecules found in animals, some plants, and microorganisms. In higher vertebrates, there are two types of globins: myoglobin, which stores oxygen in muscles, and hemoglobin, found in red blood cells. Myoglobins function as single molecules, whereas hemoglobins occur as tetramers, two each of two protein chains. Evolution in the globin gene family can be traced by comparative studies of globins in different species as well as molecular studies of globins within a species (see chapter 21). Studying hemoglobins has provided a great deal of information on gene expression and evolution. We turn our attention to the globin gene family in human beings.

During human development, four major hemoglobins appear: embryonic hemoglobin, Hb F, Hb A, and Hb A_2

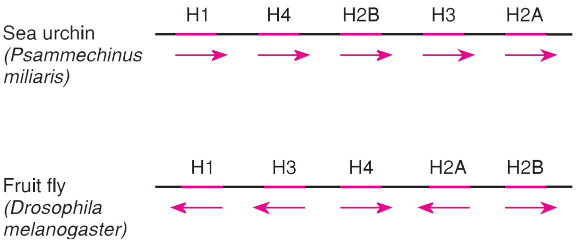


Figure 15.33 The arrangement of histone genes (*red*) within the five-gene cluster in sea urchins and fruit flies. *Arrows* indicate the direction of transcription. Spacer DNA (*black*) separates the genes.

(table 15.5). Structurally, the ζ (Greek, zeta) subunit (a component of embryonic hemoglobin) is α-like, whereas the rest are β-like (fig. 15.34; see also fig. 10.29). Fetal hemoglobin has a higher affinity for oxygen than does adult hemoglobin, thus allowing fetuses to draw oxygen from their mother's blood. From a comparative study of the DNA sequences, the evolution of the various hemoglobin genes has been inferred (fig. 15.35).

The α genes are located in a cluster on chromosome 16; the β genes are located in a cluster on chromosome 11 (fig. 15.36). These two clusters provide a clear case history of gene duplication, presumably by unequal crossing over, followed by divergence. Having a second or third copy of a gene allows one of the duplicates to diverge (and perhaps to become nonfunctional in the process), whereas the original still performs the required function.

Many diseases of genetic interest involve the hemoglobins. In fact, hemoglobinopathies, including sickle-cell anemia and the thalassemias, are the most common genetic disorders in the world population. The best-known mutation of a hemoglobin gene itself is the one that causes sickle-cell anemia, a mutation of the sixth amino acid of the β chain. In the homozygous state, the disease is usually fatal. However, heterozygotes show an

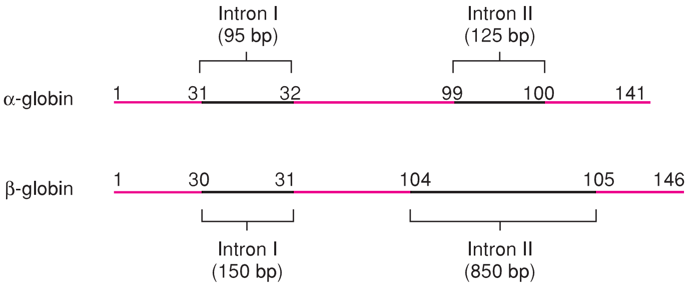


Figure 15.34 The structure of adult human α- and β-globin genes. The *numbers* refer to amino acids (or translated codons).

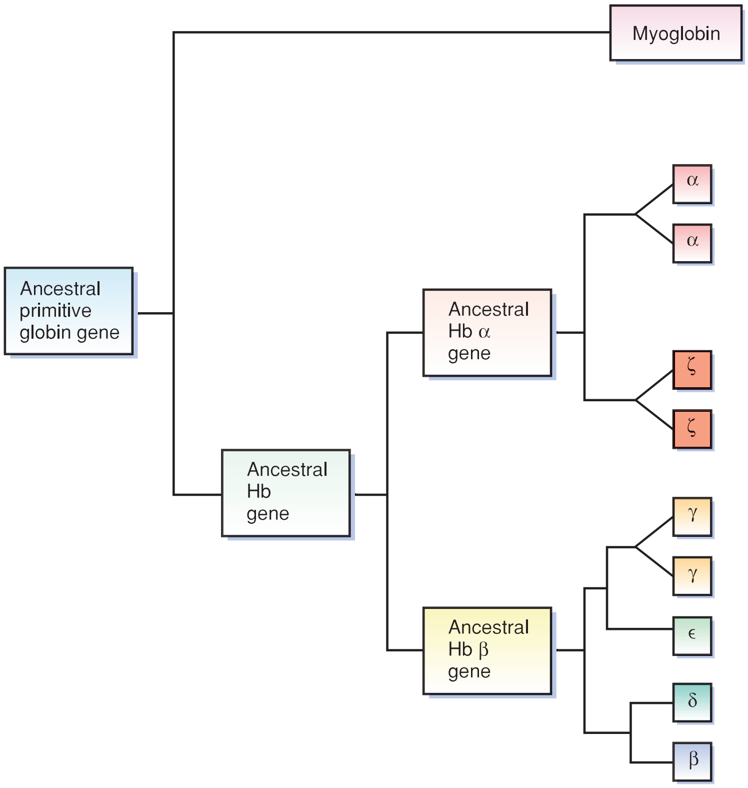


Figure 15.35 The presumed evolution of the various human globin genes from an ancestral primitive gene. The diagram represents a branching tree that begins on the *left* and progresses to the *right.* Each branch point is an evolutionary step in which the genes presumably were duplicated and then either diverged or simply endured as duplicates, as in present-day genes (on the *right*).

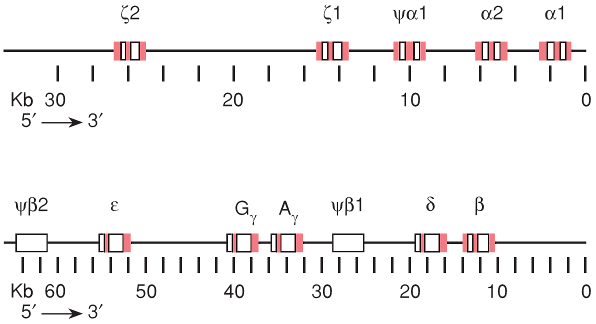


Figure 15.36 The α- and β-globin gene clusters in human beings. The ψβ1 and 2 and the ψα1 refer to nontranscribing genes (pseudogenes). Mutation has rendered the pseudogenes inactive. Within each gene box, *solid color* refers to exons and *open regions* refer to introns. (Reproduced, with permission, from the *Annual Review of Genetics,* Volume 14, © 1980 by Annual Reviews, Inc.)

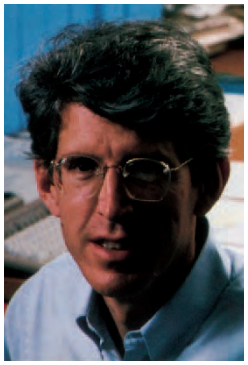
increased resistance to malaria. One of the ramifications is that the sickle-cell allele is maintained at relatively high frequencies in malarial regions (see chapter 21).

The *thalassemias* are a group of diseases that affect the regulation of the α and β hemoglobin genes. (*Thalassemia* comes from the Greek for "sea blood," because the disease is best known in individuals living around the Mediterranean Sea.) In α and β thalassemias, the α or β subunit, respectively, is present in very low quantities or entirely absent. Many of the genetic defects are deletions, possibly due to unequal crossing over within the globin gene complexes. T. Maniatis showed that β thalassemia is caused by a mutation in the β-globin gene that disrupts RNA splicing. The body compensates by forming γ_4 or β_4 hemoglobin in α thalassemias, or $\alpha_2\gamma_2$ or $\alpha_2\delta_2$ in β thalassemias. These are relatively unsuitable or inefficient responses; the diseases range from very mild to very severe and frequently fatal. More information is needed regarding the control of hemoglobin production in the thalassemias.

Table 15.5 Types of Human Hemoglobin

Type	Generally When Present	Composition
Embryonic	Up until eight weeks of gestation and beyond	$\zeta_2\epsilon_2$
Fetal (Hb F)	Eight weeks to birth	$\alpha_2\gamma_2$
Adult (Hb A)	Just before birth and beyond	$\alpha_2\beta_2$
Adult (Hb A_2)	In immature cells	$\alpha_2\delta_2$

Note: Subscripts refer to the numbers of subunits present.



Tom Maniatis (1943–). (Courtesy of Dr. Tom Maniatis.)

S U M M A R Y

STUDY OBJECTIVE 1: To examine the arrangement of DNA and proteins compromising the eukarayotic chromosome 440–452

To study developmental control in eukaryotes, we must understand the eukaryotic chromosome, which is uninemic: It consists of one DNA double helix per chromosome. Nucleoprotein is composed of DNA, histones, and nonhistone proteins. The nucleosome, a uniform packaging of the DNA, is made of histones. The majority of the nonhistone proteins create the scaffold structure of the chromosome and are not involved in gene regulation. Presumably, very small quantities of the nonhistone proteins take part in the regulation of transcription.

Core DNA, wrapped around nucleosomes, is separated by linker DNA between nucleosomes. There are regions of DNA, vulnerable to nucleases, that do not contain nucleosomes; these are referred to as nuclease-hypersensitive sites. Nucleosomes generally inhibit transcription. The 110 Å nucleosomed DNA forms a 300 Å fiber by coiling into a solenoidlike structure. Coiling of this fiber presumably forms the thick, 2,400 Å fiber seen in metaphase chromosomes.

STUDY OBJECTIVE 2: To look at the nature of centromeres and telomeres in eukaroyotic chromosomes 453–457

The centromere and telomeres are specific functional regions of a chromosome. Centromeres isolated from yeast chromosomes have three consensus areas. Telomeres are tandem repeats of a short (five-base-pair to eight-base-pair) segment. Telomeric sequences are added to the ends of chromosomes by the enzyme telomerase, which uses RNA as a template for adding DNA nucleotides. The number of telomeric repeats varies, declining as a cell ages. Telomeric repeat number may control the ability of a cell to replicate and may be implicated in cancerous growth.

Substructuring in the eukaryotic chromosome is demonstrated by G-, C-, and R-banding techniques. C-bands (constitutive heterochromatin) appear to be around the centromeres. These bands consist primarily of satellite DNA, which seems to have a structural role in the chromosome. G-bands (Giemsa bands) presumably represent intercalary heterochromatin and, also presumably, do not have an active transcriptional role. R-bands (reverse bands) appear between the G-bands and represent intercalary euchromatin, the site of transcribed, structural genes.

STUDY OBJECTIVE 3: To analyze the nature of the DNA in eukaryotic chromosomes 457–461

Eukaryotes have very large genomes with huge differences in DNA content between organisms similar in complexity, leading to the C-value paradox. Eukaryotic chromosomes contain both unique and repetitive DNA. Highly repetitive DNA is structural (centromeres, telomeres). Junk DNA is mainly short and long interspersed elements. These SINEs and LINEs are often present in hundreds of thousands of copies and can account for 50% of an organism's DNA. They are retrotransposons, transposons that jump by way of an RNA intermediate. Some functional genes also occur in many copies, such as ribosomal RNA genes, histone genes, and globin genes.

SOLVED PROBLEMS

PROBLEM 1: Why is higher-order chromosomal structure expected in eukaryotes but not prokaryotes?

Answer: The simplest explanation is the difference in amount of the genetic material in prokaryotes and eukaryotes. Since the average human chromosome has several centimeters of DNA, that DNA must be contracted to a size in which it can be moved during mitosis and meiosis without tangling and breaking. Nucleosomes provide the first order of coiling, and then several levels of coiling of the nucleosomed DNA bring it down to a manageable size for nuclear divisional processes.

PROBLEM 2: Why might we expect to see chromosomal puffs that are tissue- and stage-specific, constitutive, and environmentally induced?

Answer: The various patterns of chromosomal puffing are expected because puffing indicates transcription, the activity of specific genes. Thus, since various tissues are different because they have different proteins, each tissue is expected to have a unique suite of active genes and thus a unique suite of puffs. Similarly, different stages in an insect's development would require different genes to be active, and different puffs should therefore appear at different stages of development. Some genes are active all the time because they specify proteins, such as ribosomal protein genes, that are needed all the time. Finally, environmental insults such as heat shock are known to induce a group of genes that are needed to react to the specific insult, resulting in a suite of puffs that respond consistently to an environmental insult.

EXERCISES AND PROBLEMS*

THE EUKARYOTIC CELL

1. Summarize the major differences between eukaryotes and prokaryotes, including the structures of their DNAs.

THE EUKARYOTIC CHROMOSOME

2. Summarize the evidence that the eukaryotic chromosome is uninemic.
3. What results would you get in the experiment shown in figure 15.1 if the eukaryotic chromosome were not uninemic, but instead had some other number of complete DNA molecules (e.g., binemic)?
4. What are the major protein components of the eukaryotic chromosome? What are their functions?
5. What evidence is used to determine the length of DNA associated with a nucleosome? What is a nuclease-hypersensitive site? What functions are associated with these sites?

---

*Answers to selected exercises and problems are on page A-18.

6. What is the protein composition of a nucleosome? What function does histone H1 have?
7. What are the relationships among the 110 Å, 300 Å, and 2,400 Å fibers of the eukaryotic chromosome?
8. Draw a mitotic chromosome during metaphase. Diagram the various kinds of bands that can be brought out by various staining techniques. What information is known about the DNA content of these bands?
9. Give a 300 Å fiber model of the chromosome to account for G-bands.
10. Give a 300 Å fiber model of the chromosome to account for polytene chromosomal puffs.
11. What are the differences among polytene chromosomes, lampbrush chromosomes, puffs, and Balbiani rings? Draw an example of each.
12. Under what circumstances does a chromosomal puff occur? What does it signify?
13. What is satellite DNA? What does it signify?
14. What is a centromere? a kinetochore? What do we know about the sequences within a yeast centromere?
15. What is a telomere? What are its functions? What is its structure?
16. Describe three ways in which cells protect their telomeres.
17. What functions exist in unique, repetitive, and highly repetitive DNAs?
18. How would you use recombinant DNA techniques to locate the number and position of Alu members in the human chromosomes?
19. How could you use modern recombinant DNA technology to determine the direction of transcription of the histone genes in figure 15.33?
20. How many functional globin genes are there in mammals?
21. How could you determine, using modern recombinant DNA techniques, that the α- and β-globin pseudogenes exist?
22. Kavenoff and colleagues determined the size of DNA in *Drosophila* chromosomes in two ways: (1) Spectrophotometric measurements were made on the largest intact chromosome. These measurements were then used to calculate the amount of DNA in each chromosome. (2) Nuclei were gently lysed and chromosomes isolated. The lengths of the longest DNA molecules were measured, and those lengths were used to determine the amount of DNA in each molecule. What results for each method would you expect if
    - **a.** the chromosomes contain one DNA molecule?
    - **b.** the chromosomes contain more than one DNA molecule?
23. What can be said about the base composition of the satellite DNA with a density of 1.671 in figure 15.21?
24. When chromatin is partially digested with an endonuclease, the proteins removed, and the DNA separated in a sizing gel, DNA fragments in multiples of two hundred base pairs are found. Provide an explanation for this observation.
25. If chromatin is digested with an endonuclease to produce two hundred base-pair fragments, and these fragments are then used for transcription experiments, very little RNA is made. Provide an explanation for this observation.
26. Can nucleosomes contain the DNA for one gene? Explain.
27. If radioactive probes are made from highly repetitive DNA, these probes hybridize *in situ* mainly to centromeric and telomeric regions. What does this result suggest about the organization of chromosomes?
28. Would you expect archaeal species to have nucleosomes?
29. What is the C-value paradox, and how is it explained?
30. What are the origins of SINEs and LINEs?

CRITICAL THINKING QUESTIONS

1. How could comparative DNA studies aid us in understanding the roles of the different kinds of DNA present in the eukaryotic chromosome?
2. How could mutations involving telomeres lead to cancer?

Suggested Readings for chapter 15 are on page B-14.