Chromatin and chromosomes

10

Benjamin Lewin
Founding Publisher/Editor,
Cell Press and Virtual Text

Jocelyn E. Krebs
Professor of Biological Sciences,
University of Alaska, Anchorage, AK

A HUMAN CHROMOSOME. (Photo © Biophoto Associates/Science Source)

CHAPTER OUTLINE

10.1 Introduction

All cellular genetic material exists as a compact mass in a relatively confined volume. In bacteria, the genetic material is seen in the form of a **nucleoid** that forms a discrete clump within the cell. In eukaryotic cells, it is seen as the mass of **chromatin** within the nucleus at interphase. The packaging of chromatin is flexible and changes during the eukaryotic cell cycle. Interphase chromatin becomes even more tightly packaged at the time of division (mitosis or meiosis), when individual **chromosomes** become visible as discrete entities.

A chromosome is a device for segregating genetic material at cell division. The crucial structural feature by which this is accomplished is the **centromere**, often visible as a constriction in the length of the chromosome under the light microscope. At a greater level of detail, the centromere can be seen to include the **kinetochore**, a structure by which it is attached to microtubules. A eukaryotic chromosome usually consists of a very long linear segment of DNA, and another crucial feature is the **telomere**, which stabilizes the ends and is extended by special mechanisms that bypass the difficulties of replicating the ends of linear DNA.

The density of DNA is high. In a bacterial nucleoid it is ~10 mg/mL, in a eukaryotic nucleus it is ~100 mg/mL, and in the head of the phage T4 virus it is >500 mg/mL. Such a concentration in solution is equivalent to a gel of great viscosity and has implications (not fully understood) for the ability of proteins to find their binding sites on DNA. The various activities of DNA, such as replication and transcription, must be accomplished within this confined space. The organization of the material must accommodate transitions between inactive and active states. FIGURE 10.1 shows the range of genome sizes and makes the point that they are divided into chromosomes varying greatly in DNA content.

Organism	Genome (Mb)	Haploid chromosomes	Range of chromosome length (Mb DNA)	Total genes
E. coli	4.6	1	4.6	4,401
S. cerevisiae	12.1	16	(0.2)–1.5	6,702
D. melanogaster	165	4	(1.3)–28	14,399
Rice	389	12	24–45	37,544
Mouse	2,500	20	60–195	26,996
Man	2,900	23	49–245	24,194

FIGURE 10.1 The number of chromosomes in the haploid genome and the chromosome size vary extensively.

The length of the DNA as an extended molecule would vastly exceed the dimensions of the region that contains it. Its condensed state results from its binding to basic proteins. The positive charges of these proteins neutralize the negative charges of the nucleic acid. The structure of the nucleoprotein complex is determined by the interactions of proteins that condense the DNA into a tightly coiled structure. Therefore,in contrast with the customary picture of DNA as an extended double helix, structural deformation of DNA to bend or fold into a more compact form is the rule rather than exception.

Most chromatin has a relatively dispersed appearance; this material is called **euchromatin**, and it contains the majority of active genes. Some regions of chromatin are more densely packed; this material is called **heterochromatin** and is usually not transcriptionally active.

What is the general structure of chromatin, and what is the difference between active and inactive sequences? The high overall packing ratio of the genetic material immediately suggests that DNA cannot be directly packaged into the final structure of chromatin. There must be *hierarchies* of organization. A major question concerns the *specificity* of packaging. Is the DNA folded into a *particular* pattern, or is it different in each individual copy of the genome? How does the pattern of packaging change when a segment of DNA is replicated or transcribed?

The building block of chromatin is the **nucleosome**, and it has the same fundamental structure in all eukaryotes. The nucleosome contains ~200 base pair (bp) of DNA, organized by an octamer of small, basic proteins into a beadlike structure. The protein components are the **histones**. They form an interior core; the DNA lies on the surface of the particle. Nucleosomes are an invariant component of euchromatin and heterochromatin in the interphase nucleus and of mitotic chromosomes. The nucleosome provides the first level of organization. It packages 67 nm of DNA into a body of diameter 11 nm. Its components and structure are well characterized. A linear string of nucleosomes forms a structure referred to as the "10-nm fiber."

The second level of organization is the coiling of the series of nucleosomes into a helical array to constitute the fiber of diameter ~30 nm that is found in interphase chromatin as well as in mitotic chromosomes such as the one shown in FIGURE 10.2. This condenses the nucleosomes by a factor of 6 to 73 per unit length.

FIGURE 10.2 The sister chromatids of a mitotic pair each consist of a fiber (~30 nm in diameter) compactly folded into the chromosome. (Photo © Biophoto Associates/ Science Source)

The final packing ratio is determined by the third level of organization, the packaging of the 30-nm fiber itself. Euchromatin is about 50 times more condensed relative to the 30-nm fiber. Euchromatin is cyclically interchangeable with packing into mitotic chromosomes, which are about 5–10 times more compact. Heterochromatin generally has the same packing density as mitotic chromosomes.

The mass of chromatin contains up to twice as much protein as DNA. Approximately half of the protein mass is accounted for by the nucleosomes. The mass of RNA is <10% of the mass of DNA. Much of the RNA consists of nascent transcripts still associated with the template DNA.

Changes in chromatin structure are accomplished by association with additional proteins or by modifications of existing chromosomal proteins. Both replication and transcription require unwinding of DNA and, thus, must involve an unfolding of the structure that allows the relevant enzymes to manipulate the DNA. This is likely to involve changes in all levels of organization.

The **nonhistones** include all the proteins of chromatin except the histones. Non-histones are more variable between tissues and species and comprise a smaller proportion of the mass than the histones. They also comprise a much larger number of proteins, so that any individual protein is present in amounts much smaller than any histone.

10.2 Chromatin is divided into euchromatin and heterochromatin

Key concepts

- Individual chromosomes can be seen only during mitosis.
- During interphase, the general mass of chromatin is in the form of euchromatin, which is less tightly packed than mitotic chromosomes.
- Regions of heterochromatin remain densely packed throughout interphase.

Each chromosome contains a single, very long duplex of DNA that is folded into a fiber that runs continuously throughout the chromosome. In accounting for interphase chromatin and mitotic chromosome structure, we have to explain the packaging of a single, exceedingly long molecule of DNA into a form in which it can be transcribed and replicated and can become cyclically more or less compressed.

Individual eukaryotic chromosomes are visible as such only during the act of cell division, when each can be seen as a compact unit. Figure 10.2 is an electron micrograph of a sister chromatid pair, captured at metaphase. (The sister chromatids are daughter chromosomes produced by the previous replication event, still joined together at this stage of mitosis.) Each consists of a fiber with a diameter of ~30 nm and a nubbly appearance. The DNA is 5–10 times more condensed in chromosomes than in interphase chromatin.

During most of the life cycle of the eukaryotic cell, however, its genetic material occupies an area of the nucleus in which individual chromosomes cannot be distinguished. The 30-nm fiber from which chromatin is constructed is similar or identical to that of the mitotic chromosomes.

Chromatin can be divided into two types of material, which can be visualized by staining

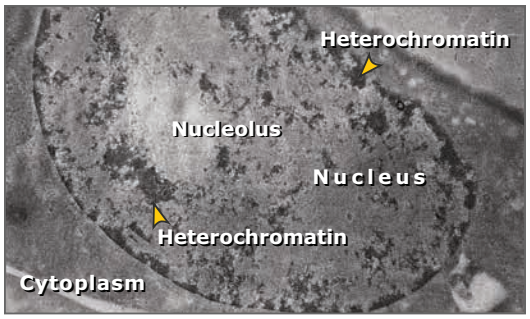

FIGURE 10.3 A thin section through a nucleus stained with a Feulgen-like material shows heterochromatin as compact regions clustered near the nucleolus and nuclear envelope. (Photo courtesy of Edmund Puvion, Centre National de la Recherche Scientifique.)

with DNA-specific dyes, as seen in the nuclear section of **FIGURE 10.3** and as given below:

1. In most regions, the fibers are much less densely packed than in the mitotic chromosome. This material is called euchromatin. It has a relatively dispersed appearance in the nucleus and occupies most of the nuclear region in Figure 10.3.
2. Some regions of chromatin are very densely packed with fibers, displaying a condition comparable with that of the chromosome at mitosis. This material is called heterochromatin. It is typically found at centromeres but also occurs at other locations, such as telomeres and highly repetitive sequences. It passes through the cell cycle with relatively little change in its degree of condensation. It forms a series of discrete clumps in Figure 10.3, but often the various heterochromatic regions aggregate into a densely staining chromocenter. (This description applies to regions that are always heterochromatic, called constitutive heterochromatin; in addition, there is another class of heterochromatin, called facultative heterochromatin, in which regions of euchromatin are converted to a heterochromatic state.)

The same fibers run continuously between euchromatin and heterochromatin, because these states represent different degrees of condensation of the genetic material. In the same way, euchromatic regions exist in different states of condensation during interphase and during mitosis. Therefore, the genetic material is organized in a manner that permits alternative states to be maintained side by side in chromatin and allows cyclic changes to occur in the packaging of euchromatin between interphase and division. We discuss the molecular basis for these states later in this chapter.

The structural condition of the genetic material is correlated with its activity. The common features of constitutive heterochromatin are as follows:

- It is permanently condensed.
- It often consists of multiple repeats of a few sequences of DNA that are not transcribed or are transcribed at very low levels.
- Probably resulting from the condensed state, it replicates later than euchromatin and has a reduced frequency of genetic recombination relative to euchromatic gene-rich areas of the genome.
- The density of genes in this region is very much reduced compared with euchromatin, and genes that are translocated into or near it are often inactivated. The one dramatic exception to this is the ribosomal DNA (rDNA) in the nucleolus, which has the general compacted appearance and behavior of heterochromatin (such as late replication) yet is engaged in very active transcription.

Numerous molecular markers exist for changes in the properties of the DNA and protein components in heterochromatic regions (see *10.27 Heterochromatin depends on interactions with histones*). They include reduced acetylation of histone proteins, increased methylation at particular sites on histone proteins, and methylation of cytosine bases in specific regions of DNA (see Figure 10.72). These molecular changes cause the condensation of the chromatin, which is responsible for its inactivity.

Although active genes are contained within euchromatin, only a subset of the sequences in euchromatin are transcribed at any time. Therefore, although location in euchromatin is *necessary* for expression of many genes, it is not *sufficient* for it.

In addition to the general distributions observed for heterochromatin and euchromatin, studies have addressed whether there is an overall chromosome organization within the nucleus. The answer in many cases is yes; chromosomes appear to occupy distinct three-dimensional spaces known as **chromosome territories**. The chromosomes occupying these territories are not entangled with each other

but do share areas of interaction and some common functional organization. For example, heterochromatic and other silent regions are found primarily at the nuclear periphery, whereas gene-dense regions are internally located. Active genes are often found at the borders of territories, sometimes clustered together in interchromosomal spaces that are enriched in transcriptional machinery, known as "transcription factories." How chromosome territories are established and how they vary by cell cycle and cell type are not yet understood.

Concept and Reasoning Check

1. Why are genes not transcribed in heterochromatin? What do you predict happens to transcription in euchromatin during mitosis?

10.3 Chromosomes have banding patterns

Key concepts

- Certain staining techniques cause the chromosomes to have the appearance of a series of striations called G-bands.
- The bands are lower in G • C content than the interbands.
- Genes are concentrated in the G • C–rich interbands.

Because of the diffuse state of chromatin, we cannot directly determine the specificity of its organization. But we can ask whether the structure of the mitotic chromosome is ordered. Do particular sequences always lie at particular sites, or is the folding of the fiber into the overall structure a more random event?

At the level of the chromosome, each member of the complement has a different and reproducible ultrastructure. When subjected to certain treatments and then stained with the chemical dye Giemsa, chromosomes generate a series of **G-bands**. FIGURE 10.4 presents an example of the human set.

Until the development of this technique, chromosomes could be distinguished only by their overall size and the relative location of the centromere. G-banding allows each chromosome to be identified by its characteristic banding pattern. This pattern allows translocations from one chromosome to another to be identified by comparison with the original diploid set. FIGURE 10.5 shows a diagram of the bands of the human X chromosome. The bands are large structures, each ~10^7 bp of DNA, which could include many hundreds

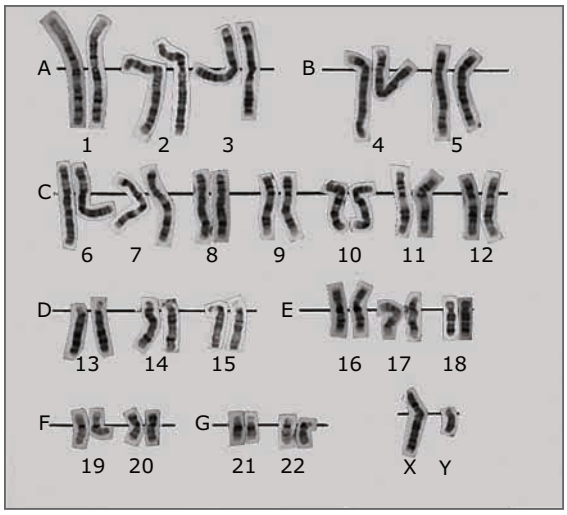

FIGURE 10.4 G-banding generates a characteristic lateral series of bands in each member of the chromosome set. (Photo courtesy of Lisa Shaffer, Signature Genomic Laboratories, Spokane.)

of genes. Figure 10.5 also shows the nomenclature used to identify genetic positions on individual chromosomes. A given location is indicated by its position on the long (q) or short (p) arm and then by the region of arm, the band, and subband(s).

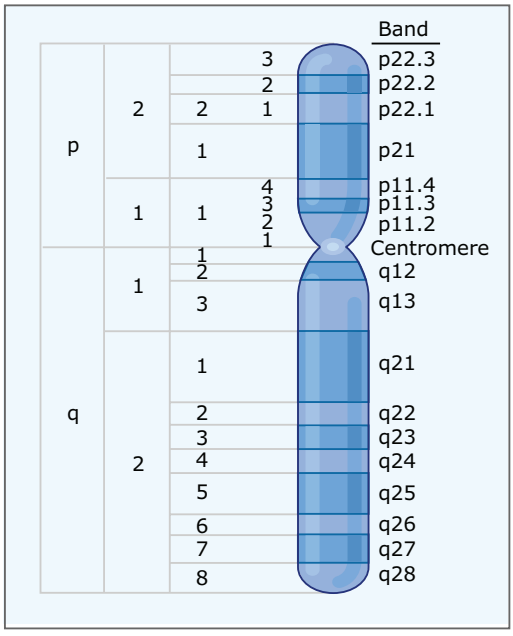

FIGURE 10.5 The human X chromosome can be divided into distinct regions by its banding pattern. The short arm is *p* and the long arm is *q*; each arm is divided into larger regions that are further subdivided. This map shows a low-resolution structure; at higher resolution, some bands are further subdivided into smaller bands and interbands; for example, *p*21 is divided into *p*21.1, *p*21.2, and *p*21.3.

The banding technique is of enormous practical use, but the mechanism of banding remains a mystery. All that is certain is that the dye stains untreated chromosomes more or less uniformly. Thus, the generation of bands depends on a variety of treatments (such as proteolytic digestion) that change the response of the chromosome (presumably by extracting the component that binds the stain from the nonbanded regions). Similar bands can be generated by a variety of treatments.

The only known feature that distinguishes bands from interbands is that the bands have a lower G • C content than the interbands. If there are ~10 bands on a large chromosome with a total content of ~100 Mb, this means the chromosome is divided into regions of ~5 Mb in length that alternate between low G • C (band) and high G • C (interband) content. There is a tendency for genes to be located in the interband regions. All of this argues for some long-range sequence-dependent organization.

The human genome sequence confirms the basic observation. **FIGURE 10.6** shows distinct fluctuations in G • C content when the genome is divided into small tranches. The average of 41% G • C is common to mammalian genomes. There are regions as low as 30% or as high as 65%. When longer tranches are examined, there is less variation. The average length of regions with >43% G • C is 200–250 kb. This makes it clear that the band/interband structure does not represent homogeneous segments that alternate in G • C content, although the bands do contain a higher content of low G • C segments. Genes are concentrated in regions of higher G • C content. We have yet to understand how the G • C content affects chromosome structure.

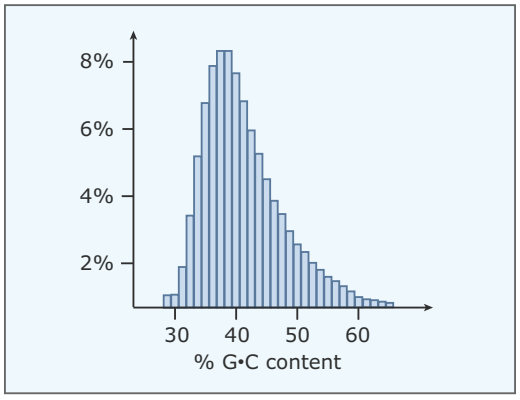

FIGURE 10.6 There are large fluctuations in G • C content over short distances. Each bar shows the percent of 20-kb fragments with the given G • C content.

Concept and Reasoning Check

1. How can G-banding be used to detect chromosomal deletions, inversions, and translocations?

10.4 Eukaryotic DNA has loops and domains attached to a scaffold

Key concepts

- DNA of interphase chromatin is negatively supercoiled into independent domains of ~85 kb.
- Metaphase chromosomes have a protein scaffold to which the loops of supercoiled DNA are attached.

The characteristic banded structure of every chromosome results from the folding of a deoxyribonucleoprotein fiber. The fiber is organized in a series of loops by a proteinaceous matrix. This looped organization can be revealed by the properties of material released by a gentle lysis of the cells.

When nuclei are lysed on top of a sucrose gradient, the eukaryotic genome can be isolated as a single, compact body. As isolated from *Drosophila melanogaster*, it can be visualized as a compactly folded fiber (10 nm in diameter), consisting of DNA bound to proteins. A revealing feature is that the DNA of the isolated chromatin behaves as a closed duplex structure, as judged by its response to ethidium bromide. This small molecule intercalates between base pairs to generate positive superhelical turns in "closed" circular DNA molecules, that is, molecules in which both strands have covalent integrity. (In "open" circular molecules, which contain a nick in one strand, or with linear molecules, the DNA can rotate freely in response to the intercalation, thus relieving the tension.)

Some nicks occur in the DNA during its isolation; they can also be generated by limited treatment with DNase. But this does not abolish the ability of ethidium bromide to introduce positive supercoils. This capacity of the genome to retain its response to ethidium bromide even after accidental or DNase-dependent nicking means it must have many independent chromosomal domains and the supercoiling in each domain is not affected by events in the other domains. Each domain consists of a loop of DNA, the ends of which are secured in some (unknown) way that does not allow rotational events to propagate from one domain to another.

In a natural closed DNA that is negatively supercoiled, the intercalation of ethidium bromide first removes the negative supercoils and then introduces positive supercoils. The amount of ethidium bromide needed to achieve zero supercoiling is a measure of the original density of negative supercoils. In a typical eukaryotic genome, supercoiling measured by the response to ethidium bromide corresponds to about one negative supercoil per 200 bp. These supercoils can be removed by nicking with DNase, although the DNA remains in the form of the 10-nm fiber. This suggests the supercoiling is caused by the arrangement of the fiber in space and represents the existing torsion.

Full relaxation of the supercoils requires one nick per 85 kb, identifying the average length of "closed" DNA. This region could comprise a loop or domain similar in nature to those identified in the bacterial genome. Loops can be seen directly when most proteins are extracted from mitotic chromosomes. The resulting complex consists of the DNA associated with ~8% of the original protein content. As seen in FIGURE 10.7, the histone-depleted chromosomes take the form of a central **scaffold** surrounded by a halo of DNA.

The metaphase scaffold consists of a dense network of fibers. Threads of DNA emanate from the scaffold, apparently as loops with an average length of 10–30 μm (30–90 kb). The DNA can be digested without affecting the integrity of the scaffold, which consists of a set of specific proteins. This suggests a form of organization in which loops of DNA of ~60 kb are anchored in a central proteinaceous scaffold.

The appearance of the scaffold resembles a mitotic pair of sister chromatids. The sister scaffolds usually are tightly connected but sometimes are separate, joined only by a few fibers. Could this be the structure responsible for maintaining the shape of the mitotic chromosomes? Could it be generated by bringing together the protein components that usually secure the bases of loops in interphase chromatin?

Concept and Reasoning Check

1. Even though linear DNA cannot be supercoiled, linear eukaryotic chromosomes have less than one negative supercoil per 200 bp. Explain.

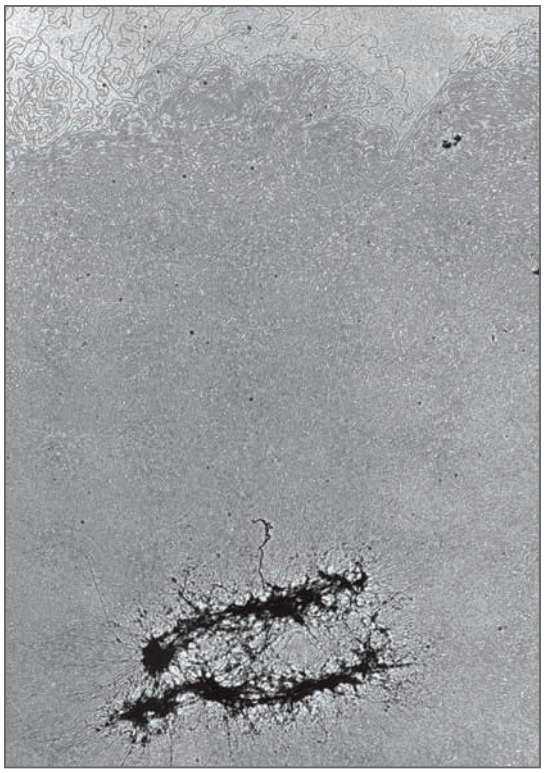

FIGURE 10.7 Histone-depleted chromosomes consist of a protein scaffold to which loops of DNA are anchored. (Photo courtesy of Ulrich K. Laemmli, University of Geneva, Switzerland.)

10.5 Specific sequences attach DNA to an interphase matrix or a metaphase scaffold

Key concepts

- DNA is attached to the nuclear matrix at specific sequences called matrix attachment regions or scaffold attachment regions.
- The matrix attachment regions are A • T–rich but do not have any specific consensus sequence.

Is DNA attached to the scaffold via specific sequences? DNA sites attached to proteinaceous structures in interphase nuclei are called **matrix attachment regions** (**MARs**); they are sometimes also called scaffold attachment regions, or S/MARs, because the same sequences appear to attach to protein substructures in both metaphase and interphase cells. The nature of the structure in interphase cells to which they are connected is not clear. Chromatin appears to be attached to an underlying matrix, and there have been suggestions that this attachment is necessary for transcription or replication. When nuclei are depleted of proteins, the DNA extrudes as loops from a residual proteinaceous

structure. Attempts to relate the proteins found in this preparation to structural elements of intact cells have not been successful.

Are particular DNA regions associated with this matrix? In vivo and in vitro approaches are summarized in **FIGURE 10.8**. Both start by isolating the matrix as a crude nuclear preparation containing chromatin and nuclear proteins. Different treatments can then be used to characterize DNA in the matrix or to identify DNA able to attach to it.

To analyze the existing S/MARs, the chromosomal loops can be decondensed by extracting the proteins. Removal of the DNA loops by treatment with restriction endonucleases leaves only the (presumptive) in vivo S/MAR sequences attached to the matrix.

The complementary approach is to remove all DNA from the matrix by treatment with DNase. Next, isolated fragments of DNA can be tested for their ability to bind to the matrix in vitro. The same sequences should be associated with the matrix in vivo or in vitro. Once a potential S/MAR has been identified, the size of the minimal region needed for association in vitro can be determined by deletions. We can also then identify proteins that bind to the S/MAR sequences.

A surprising feature is the lack of conservation of sequence in S/MAR fragments. They are usually ~70% A•T–rich but otherwise lack any consensus sequences. There are, however, other interesting sequences often seen in the DNA stretch containing the S/MAR *cis*-acting sites that regulate transcription. A recognition site for topoisomerase II is usually present in the S/MAR. It is therefore possible that an S/MAR serves more than one function, not only providing a site for attachment to the matrix but also containing other sites at which topological changes in DNA are effected.

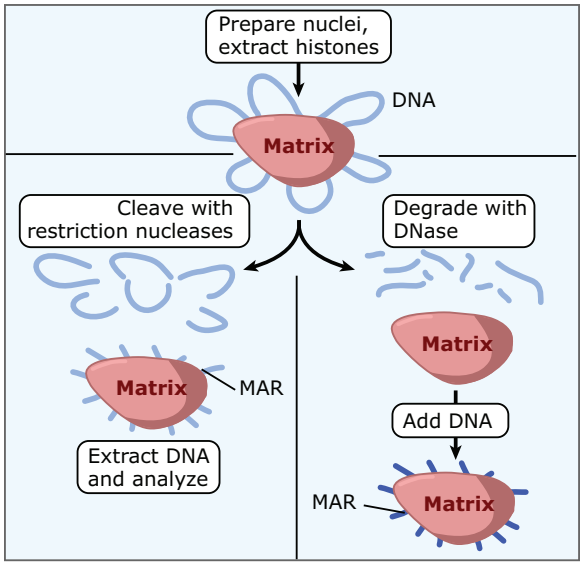

FIGURE 10.8 Matrix-associated regions may be identified by characterizing the DNA retained by the matrix isolated in vivo or by identifying the fragments that can bind to the matrix from which all DNA has been removed.

What is the relationship between the chromosome scaffold of dividing cells and the matrix of interphase cells? Are the same DNA sequences attached to both structures? In several cases, the same DNA fragments found with the nuclear matrix in vivo can be retrieved from the metaphase scaffold, and fragments that contain S/MAR sequences can bind to a metaphase scaffold. It therefore seems likely that DNA contains a single type of attachment site, which in interphase cells is connected to the nuclear matrix and in mitotic cells is connected to the chromosome scaffold. Research has shown that not all S/MARs are attached to the matrix at all times in vivo and that in fact some S/MARs may dynamically associate and disassociate from the matrix depending on the transcription of genes in the area.

The nuclear matrix and chromosome scaffold consist of different proteins, although there are some common components. Topoisomerase II is a prominent component of the chromosome scaffold and is a constituent of the nuclear matrix, suggesting the control of topology is important in both cases. Researchers have also begun to identify proteins, such as SATB1, that can regulate the dynamic association of S/MARs with the matrix to regulate transcription. This suggests that the organization of chromatin on the matrix may play a key regulatory function.

10.6 The eukaryotic chromosome is a segregation device

Key concepts

- A eukaryotic chromosome is held on the mitotic spindle by the attachment of microtubules to the kinetochore that forms in its centromeric region.
- Centromeres often have heterochromatin that is rich in satellite DNA sequences.

During mitosis, sister chromatids move to opposite poles of the cell. Their movement depends on the attachment of the chromosome to microtubules, which are connected at

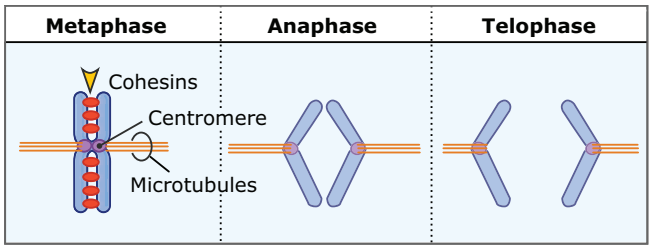

FIGURE 10.9 Chromosomes are pulled to the poles via microtubules that attach at the centromeres. The sister chromatids are held together until anaphase by glue proteins (cohesins). The centromere is shown here in the middle of the chromosome (metacentric) but can be located anywhere along its length, including close to the end (acrocentric) and at the end (telocentric).

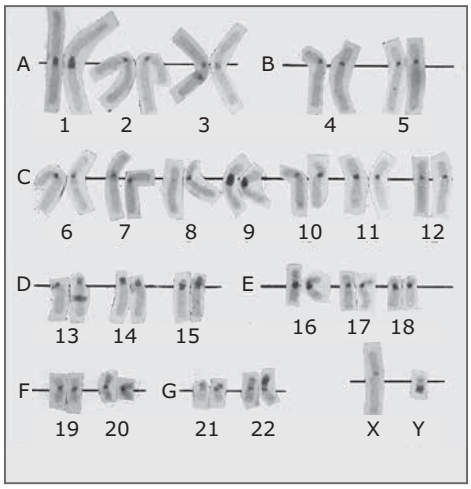

FIGURE 10.10 C-banding generates intense staining at the centromeres of all chromosomes. (Photo courtesy of Lisa Shaffer, Signature Genomic Laboratories, Spokane.)

their other end to the poles. (The microtubules constitute a cellular filamentous system, reorganized at mitosis so they connect the chromosomes to the poles of the cell.) The sites in the two regions where microtubule ends are organized—in the vicinity of the centrioles at the poles and at the chromosomes—are called microtubule organizing centers.

FIGURE 10.9 illustrates the separation of sister chromatids as mitosis proceeds from metaphase to telophase. The region of the chromosome responsible for its segregation during mitosis and meiosis is called the centromere. The centromeric region on each sister chromatid is moved along microtubules toward the opposite pole. Opposing this motive force, proteins called cohesins hold the sister chromatids together. Initially, the sister chromatids separate at their centromeres, and then are released completely from one another during anaphase when the cohesins are degraded, allowing the sister chromatids to be "dragged" along by the centromere. The chromosome provides a device for attaching a large number of genes to the apparatus for division. It contains the site at which the sister chromatids are held together before the separation of the individual chromosomes. This appears as a constricted region connecting all four chromosome arms, as seen in Figure 10.2, which shows the sister chromatids at the metaphase stage of mitosis.

The centromere is essential for segregation, as shown by the behavior of chromosomes that have been broken. A single break generates one piece that retains the centromere and another, an **acentric fragment**, that lacks it. The acentric fragment does not become attached to the mitotic spindle, and as a result, it fails to be included in either of the daughter nuclei. When chromosome movement relies on discrete centromeres, there can be *only* one centromere per chromosome. In some species, however (such as the nematode *Caenorhabditis elegans*), the centromeres are *holocentric*: diffuse and spread along the entire length of the chromosome. Species with holocentric chromosomes still make spindle fiber attachments for mitotic chromosome separation but do not require one and only one regional or point centromere per chromosome. Most of the molecular analysis of centromeres has been done on canonical point (budding yeast) or regional (fly, mammalian, rice) centromeres, and holocentric centromeres are still poorly understood.

The regions flanking the centromere are often rich in satellite DNA sequences and display a considerable amount of heterochromatin. Because the entire chromosome is condensed, centromeric heterochromatin is not immediately evident in mitotic chromosomes. It can be visualized, however, by a technique that generates **C-bands**. In the example of **FIGURE 10.10**, all centromeres show as darkly staining regions. Although it is common, heterochromatin cannot be identified around *every* known centromere, which suggests it is unlikely to be essential for the division mechanism.

The region of the chromosome at which the centromere forms is defined by DNA sequences (although the sequences have been defined in only a very small number of cases). The centromeric DNA binds specific proteins that are responsible for establishing the structure that attaches the chromosome to the microtubules. This structure is called the kinetochore. It is a

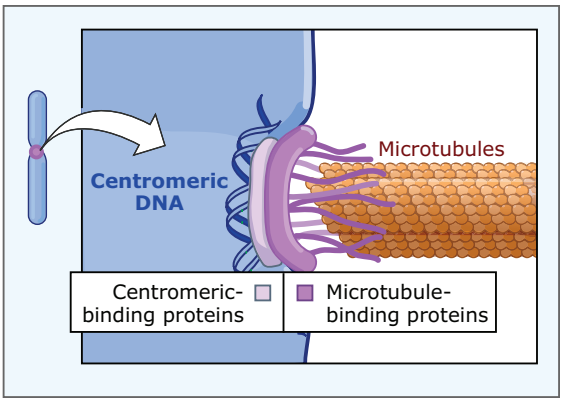

FIGURE 10.11 The centromere is identified by a DNA sequence that binds specific proteins. These proteins do not themselves bind to microtubules but establish the site at which the microtubule binding proteins in turn bind.

darkly staining fibrous object of diameter or length ~400 nm. The kinetochore provides the microtubule organizing centers on a chromosome. **FIGURE 10.11** shows the hierarchy of organization that connects centromeric DNA to the microtubules. Proteins bound to the centromeric DNA bind other proteins that bind to microtubules (see *10.8 The centromere binds a protein complex*).

Concept and Reasoning Check

1. What would happen to a chromosome that, as a result of a translocation, contains two centromeres?

10.7 Point centromeres have short DNA sequences in *Saccharomyces cerevisiae*

Key concepts

- Centromeric DNA region (*CEN*) elements are identified in *S. cerevisiae* by the ability to allow a plasmid to segregate accurately at mitosis.
- *CEN* elements consist of short conserved sequences *CDE-I* and *CDE-III* that flank the T-rich region *CDE-II*.

If a centromeric sequence of DNA is responsible for segregation, any molecule of DNA possessing this sequence should move properly at cell division, whereas any DNA lacking it will fail to segregate. This prediction has been used to isolate centromeric DNA in the budding yeast, *S. cerevisiae*. Budding yeast chromosomes do not display visible kinetochores comparable with those of multicellular eukaryotes but otherwise divide at mitosis and segregate at meiosis by the same mechanisms.

Plasmids can be engineered in yeast by making circular DNAs that have origins of replication and that are replicated like chromosomal sequences. They are, however, unstable at mitosis and meiosis, disappearing from most cells because they segregate erratically. Fragments of chromosomal DNA containing centromeres have been isolated by their ability to confer mitotic stability on these plasmids.

A centromeric DNA region (*CEN*) fragment is identified as the minimal sequence that can confer stability on such a plasmid. Another way to characterize the function of such sequences is to modify them in vitro and then reintroduce them into the yeast cell, where they replace the corresponding centromere on the chromosome. This allows the sequences required for *CEN* function to be defined directly in the context of the chromosome.

A *CEN* fragment derived from one chromosome can replace the centromere of another chromosome with no apparent consequence. This result suggests that centromeres are interchangeable. They are used simply to attach the chromosome to the spindle and play no role in distinguishing one chromosome from another.

The sequences required for centromeric function fall within a stretch of ~120 bp. The centromeric region is packaged into a nuclease-resistant structure, and it binds a single microtubule. This highly condensed, functional region is why these centromeres are referred to as "point" centromeres. We may therefore look to the *S. cerevisiae* centromeric region to identify proteins that bind centromeric DNA and proteins that connect the chromosome to the spindle.

Three types of sequence element may be distinguished in the *CEN* region, as summarized in **FIGURE 10.12**. They are as follows:

1. Cell cycle–dependent element (*CDE*)-*I* is a sequence of 9 bp that is conserved with minor variations at the left boundary of all centromeres.
2. *CDE-II* is a >90% A • T–rich sequence of 80–90 bp found in all centromeres; its function could depend on its length rather than exact sequence. Its constitution is reminiscent of some short tandemly repeated (satellite) DNAs.
3. *CDE-III* is an 11-bp sequence highly conserved at the right boundary of all centromeres. Sequences on either side of the element are less well conserved and may also be needed for centromeric function.

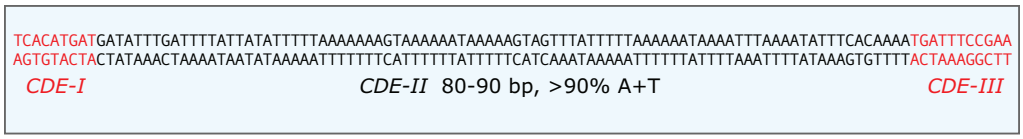

FIGURE 10.12 Three conserved regions can be identified by the sequence homologies between yeast *CEN* elements.

Mutations in *CDE-I* or *CDE-II* reduce but do not inactivate centromere function, but point mutations in the central CCG of *CDE-III* completely inactivate the centromere.

Concept and Reasoning Check

1. Why might the length of the *CDE-II* region, but not its precise sequence, have importance for *CEN* function?

10.8 The centromere binds a protein complex

Key concepts

- A specialized protein complex containing the histone variant Cse4 is formed at *CDE-II*.
- The CBF3 protein complex that binds to *CDE-III* is essential for centromeric function.
- The proteins that connect these two complexes serve as the site of assembly of the kinetochore and provide the connection to microtubules.

Can we identify proteins necessary for the function of *CEN* sequences? There are several genes in which mutations affect chromosome segregation and whose proteins are localized at centromeres. The contributions of these proteins to the centromeric structure are summarized in **FIGURE 10.13**.

A specialized chromatin structure is built by binding the *CDE-II* region to a protein called Cse4, which is a variant of one of the histone proteins that constitute the basic subunits of chromatin (see *10.18 Histone variants produce alternative nucleosomes*). A protein called Scm3 is required for proper association of Cse4 with *CEN*. Inclusion of histone H3 variants related to Cse4, such as CENP-A in higher eukaryotes, is a universal aspect of centromere construction in all species. The basic interaction consists of bending the DNA of the *CDE-II* region around a protein aggregate; the reaction is probably assisted by the occurrence of intrinsic bending in the *CDE-II* sequence.

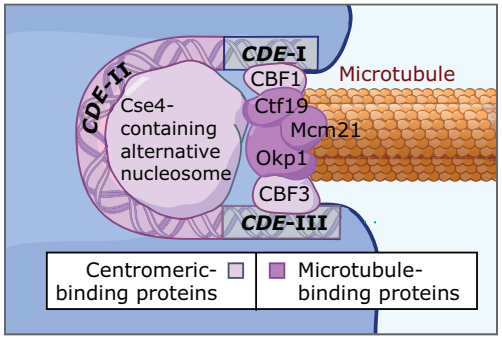

FIGURE 10.13 DNA at *CDE-II* is wound around a specialized nucleosome including the H3 variant Cse4, *CDE-III* is bound to CBF3, and *CDE-I* is bound to CBF1. These proteins are connected by the group of Ctf19, Mcm21, and Okp1.

CDE-I is bound by a homodimer of the Cbf1 protein; this interaction is not essential for centromere function, but in its absence the fidelity of chromosome segregation is reduced ~10 times. A 240-kDa complex of four proteins, called CBF3, binds to *CDE-III*. This interaction is essential for centromeric function.

The proteins bound at *CDE-I, CDE-II,* and *CDE-III* interact with another group of proteins (Ctf19, Mcm21, and Okp1), which in turn link the centromeric complex to the kinetochore proteins (~70 individual kinetochore proteins have been identified in yeast) and to the microtubule.

The overall model suggests that the complex is localized at the centromere by a protein structure that resembles the normal building block of chromatin (the nucleosome). The bending of DNA at this structure allows proteins bound to the flanking elements to become part of a single complex. The DNA binding components of the complex form a scaffold for the assembly of the kinetochore, linking the centromere to the microtubule. The construction of kinetochores follows a similar pattern and uses related components in a wide variety of organisms.

10.9 Regional centromeres contain repetitive DNA

Key concepts

- Centromeres are epigenetically determined.
- Centromeres in higher eukaryotic chromosomes contain large amounts of repetitive DNA and unique histone variants.
- The function of the repetitive DNA is not known.

The region of the chromosome at which the centromere forms was originally thought to be

defined solely by DNA sequences, yet recent studies in plants, animals, and fungi have shown that centromeres are specified epigenetically by chromatin structure. Centromere-specific histone H3 (Cse4 in yeast as described in the previous section, CENP-A in higher eukaryotes, and more generically as CenH3) appears to be a primary determinant in establishing functional centromeres and kinetochore assembly sites. This finding explains the old puzzle of why specific DNA sequences could not be identified as "the centromeric DNA" and why there is so much variation in centromere-associated DNA sequences among closely related species. This has given rise to new questions: What controls the sites of CenH3 assembly? How do chromosomes maintain only one such region per chromosome when not dictated by an underlying sequence?

Centromeres are highly specialized chromatin structures that occupy the same site for many generations. In eukaryotic chromosomes, the centromere-specific histone H3 variant CenH3 replaces the normal H3 histone at sites where centromeres reside and kinetochores attach chromosomes to spindle fibers. This specialized centromeric chromatin is the foundation for binding of other centromere-associated proteins. In addition, other histones at the centromere (including H2A and canonical H3) are subject to posttranslational modifications required for normal binding of centromeric proteins and accurate chromosome segregation, indicating the epigenetic pattern that defines a centromere is complex. This view represents a paradigm shift in how we understand centromere formation, identity, and function. CenH3 is a nucleosomal protein and not a DNA sequence per se; thus, the centromere is now regarded as being primarily epigenetic in its specification. The role of satellite DNA sequences, which are also characteristic of centromeres, remains difficult to ascertain, despite their prevalence and conservation.

The length of DNA required for centromeric function is often quite long. The short, discrete elements of the budding yeast *S. cerevisiae* may be an exception to the general rule. *S. cerevisiae* is the only case so far in which centromeric DNA can be identified by its ability to confer stability on plasmids. However, a related approach has been used with the fission yeast *Schizosaccharomyces pombe*. *S. pombe* has only three chromosomes, and the region containing each centromere has been identified by deleting most sequences of each chromosome to create a stable minichromosome. This approach locates the centromeres within regions of 40–100 kb that consist largely or entirely of repetitive DNA. It is not clear how much of each of these rather long regions is required for chromosome segregation at mitosis and meiosis. Attempts to localize centromeric functions in *Drosophila* chromosomes suggest they are dispersed in an even larger region, consisting of 200–600 kb. The large size of this type of centromere suggests it is likely to contain several separate specialized functions, including sequences required for kinetochore assembly, sister chromatid pairing, and so on.

The size of the centromere in *Arabidopsis* is comparable. Each of the five chromosomes has a centromeric region in which recombination is very largely suppressed. This region occupies >500 kb. Clearly, it includes the centromere, but we have no direct information as to how much of it is required. There are expressed genes within these regions, which casts some doubt on whether the entire region is part of the centromere. At the center of the region is a series of 180-bp repeats; this is the type of structure generally associated with centromeres. It is too early to say how these structures relate to centromeric function.

The primary motif composing the heterochromatin of primate centromeres is the α-satellite DNA, which consists of tandem arrays of a 170-bp repeating unit. There is significant variation between individual repeats, although those at any centromere tend to be better related to one another than to members of the family in other locations.

Current models for regional centromere organization and function invoke alternating chromatin domains, with clusters of CenH3 nucleosomes interspersed among clusters of nucleosomes with H3 and some of the histone variant H2A.Z. Different histones are subject to centromere-specific patterns of modification. The CenH3 nucleosomes form the chromatin foundation for recruitment and assembly of the other proteins that eventually comprise a functional kinetochore. The formation of neocentromeres that contain CenH3 but not α-satellite DNA provide important evidence for the idea of centromeres being epigenetically determined. Key questions remain as to the role of repetitive DNA and alternating chromatin domains in forming the large bipartite kinetochore structure on replicated sister centromeres.

10.10 Telomeres are replicated by a special mechanism

Key concepts

- The telomere is required for the stability of the chromosome end.
- A telomere consists of a simple repeat where a C+A–rich strand has the sequence $C_{>1}(A/T)_{1–4}$.
- The protein TRF2 catalyzes a reaction in which the 3′ repeating unit of the G+T–rich strand forms a loop by displacing its homologue in an upstream region of the telomere.

Another essential feature in all chromosomes is the telomere, which "seals" the chromosome ends. We know the telomere must be a special structure, because chromosome ends generated by breakage are "sticky" (recombinogenic) and tend to react with other chromosomes, whereas natural ends are stable.

We can apply the following two criteria in identifying a telomeric sequence:

1. It must lie at the end of a chromosome.
2. It must confer stability on a linear molecule.

The problem of finding a system that offers an assay for function again has been brought to the molecular level by using yeast. All the plasmids that survive in yeast (by virtue of possessing autonomously replicating sequences and *CEN* elements) are circular DNA molecules. Linear plasmids are unstable (because they are degraded). We can identify telomeres as sequences that confer stability on these plasmids. Fragments from yeast DNA that prove to be located at chromosome ends can be identified by such an assay, and a region from the end of a known natural linear DNA molecule—the extrachromosomal rDNA of *Tetrahymena*—is able to render a yeast plasmid stable in linear form.

Telomeric sequences have been characterized from a wide range of lower and higher eukaryotes. The same type of sequence is found in plants and humans, so the construction of the telomere seems to follow a nearly universal principle (*Drosophila* telomeres are an exception, consisting of terminal arrays of retrotransposons). Each telomere consists of a long series of short, tandemly repeated sequences. There may be 100–1,000 repeats, depending on the organism.

All telomeric sequences can be written in the general form $C_n(A/T)_m$, where $n > 1$ and $m = 1–4$. **FIGURE 10.14** shows one example (this is the telomeric sequence of the ciliate *Tetrahymena*). One unusual property of the telomeric sequence is the extension of the G-T–rich strand, usually

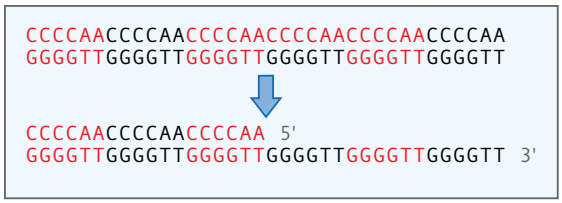

FIGURE 10.14 A typical telomere has a simple repeating structure with a G-T–rich strand that extends beyond the C-A–rich strand. The G-tail is generated by a limited degradation of the C-A–rich strand.

for 14–16 bases as a single strand. The G-tail is probably generated by a specific limited degradation of the C-A–rich strand.

The telomere is replicated by a special mechanism. Telomerase is a ribonucleoprotein enzyme that carries a template RNA with the same sequence as the C-A–rich strand. The template RNA pairs with its complement in the telomere, which forms a primer that is extended by the enzyme's reverse transcriptase activity, as shown in **FIGURE 10.15**. The processivity of

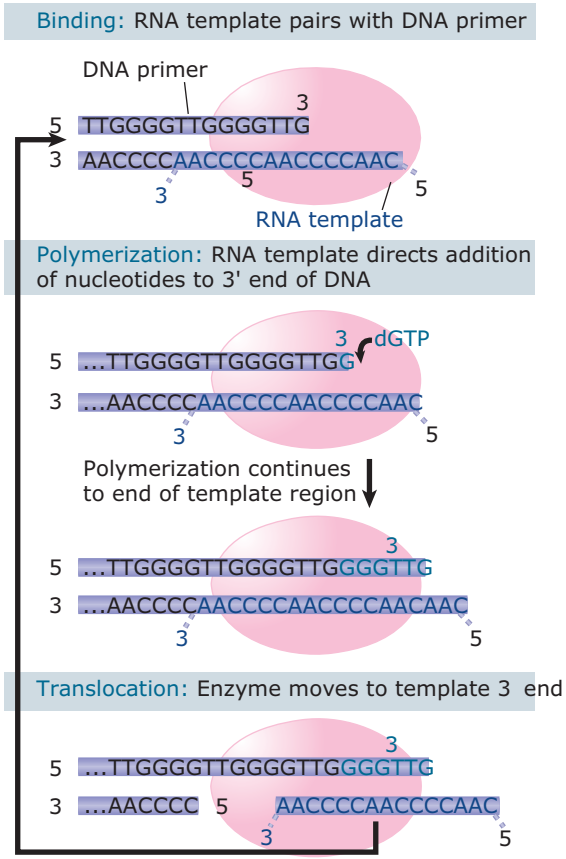

FIGURE 10.15 Telomerase positions itself by base pairing between the RNA template and the protruding single-stranded DNA primer. It adds G and T bases one at a time to the primer as directed by the template. The enzymes repositions and the cycle starts again when one repeating unit has been added.

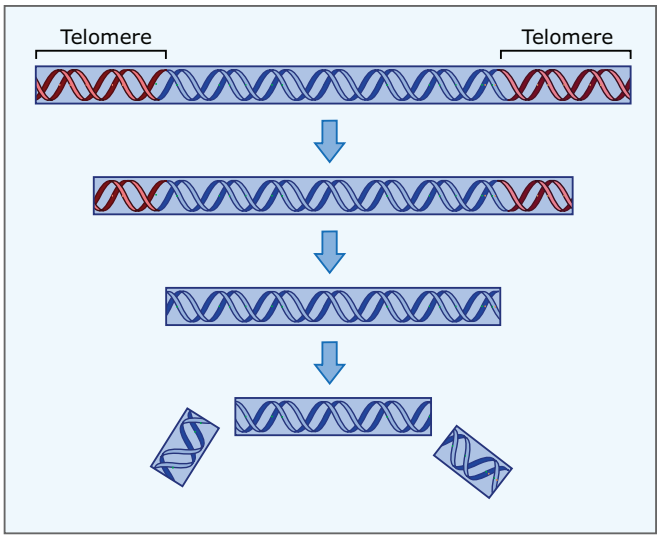

FIGURE 10.16 Mutation in telomerase causes telomeres to shorten in each cell division. Eventual loss of the telomere causes chromosome breaks and rearrangements.

the enzyme and the number of repeats added are controlled by ancillary proteins.

Because DNA replication cannot start at the very end of a linear molecule, when a chromosome is replicated, the number of repeats at the telomere is reduced. This can be demonstrated directly by eliminating telomerase activity. **FIGURE 10.16** shows that if telomerase is mutated in a dividing cell, the telomeres become gradually shorter with each cell division.

The ability of telomerase to add repeats to a telomere by de novo synthesis counteracts the loss of repeats resulting from failure to replicate up to the end of the chromosome. Extension and shortening are in dynamic equilibrium. If telomeres are continually being lengthened (and shortened), their exact sequence may be irrelevant. All that is required is for the end to be recognized as a suitable substrate for addition. Telomerase activity is found in all dividing cells and is generally turned off during cell differentiation in multicellular organisms. Differentiated cells divide slowly (if at all), and the telomere lengths in these cells is a measure of how many divisions they have undergone since differentiation.

In addition to providing the solution to the problem of replicating the ends of linear DNA, telomeres also stabilize the ends of the chromosome. Isolated telomeric fragments do not behave as though they contain single-stranded DNA; instead, they show aberrant electrophoretic mobility and other properties.

FIGURE 10.17 shows that a loop of DNA forms at the telomere. The absence of any free end may be the crucial feature that stabilizes

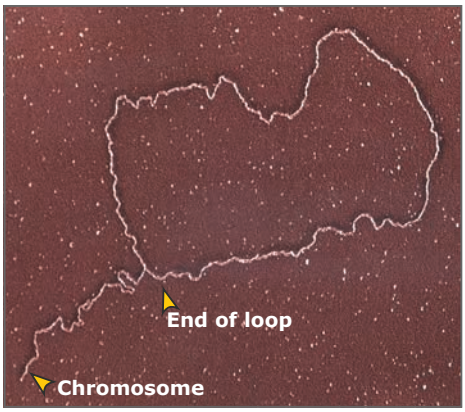

FIGURE 10.17 A loop forms at the end of chromosomal DNA. (Photo courtesy of Jack Griffith, University of North Carolina at Chapel Hill.)

the end of the chromosome. The average length of the loop in animal cells is 5–10 kb.

FIGURE 10.18 shows the loop is formed when the 3′ single-stranded end of the telomere $(TTAGGG)_n$ displaces the same sequence in an upstream region of the telomere. This converts the duplex region into a structure in which a series of TTAGGG repeats are displaced to form a single-stranded region, and the tail of the telomere is paired with the homologous strand.

The reaction is catalyzed by the telomere-binding protein TRF2, which together with other proteins forms a complex (called "shelterin" in mammals) that stabilizes the chromosome ends. Shelterin complexes protect telomeres from DNA damage repair pathways designed to act on free DNA ends. TRF2/shelterin also regulate telomere length by modulating telomerase activity.

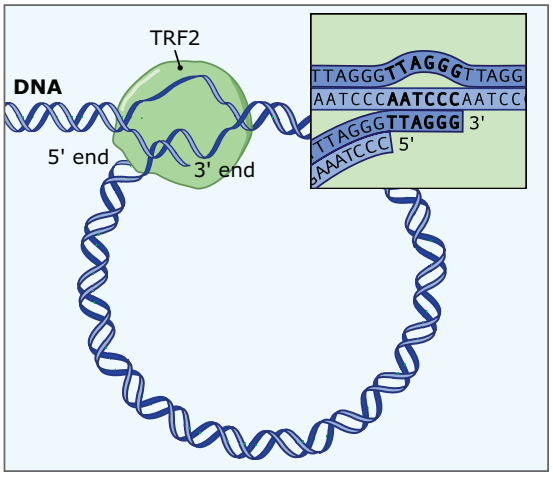

FIGURE 10.18 The 3′ single-stranded end of the telomere $(TTAGGG)_n$ displaces the homologous repeats from duplex DNA to form a t-loop. The reaction is catalyzed by TRF2.

Concept and Reasoning Check

1. What would be the effect on the telomere of a mutation in the telomerase template RNA that changes its sequence?

10.11 Lampbrush chromosomes are extended

Key concept

- Sites of gene expression on lampbrush chromosomes show loops that are extended from the chromosomal axis.

It would be extremely useful to visualize gene expression in its natural state, to see what structural changes are associated with transcription. The compression of DNA in chromatin, coupled with the difficulty of identifying particular genes within it, makes it impossible to visualize the transcription of individual active genes.

Gene expression can be visualized directly in certain unusual situations, in which the chromosomes are found in a highly extended form that allows individual loci (or groups of loci) to be distinguished. Lateral differentiation of structure is evident in many chromosomes when they first appear for meiosis. At this stage, the chromosomes resemble a series of beads on a string. The beads are densely staining granules, properly known as **chromomeres**. Usually there is little gene expression at meiosis, and it is not practical to use this material to identify the activities of individual genes. An exceptional situation, however, that allows the material to be examined is presented by **lampbrush chromosomes**, which have been best characterized in certain amphibians and birds.

Lampbrush chromosomes are formed during an unusually extended meiosis, which can last up to several months. During this period, the chromosomes are maintained in a stretched-out form in which they can be visualized in the light microscope. Later during meiosis, the chromosomes revert to their usual compact size. So the extended state essentially proffers an unfolded version of the normal condition of the chromosome.

The lampbrush chromosomes are meiotic bivalents, each consisting of two pairs of sister chromatids. FIGURE 10.19 shows an example in which the sister chromatid pairs have mostly separated so they are held together only by chiasmata. Each sister chromatid pair forms a series of ellipsoidal chromomeres, ~1–2 µm in diameter, which are connected by a very fine thread. This thread contains the two sister duplexes of DNA and runs continuously along the chromosome, through the chromomeres.

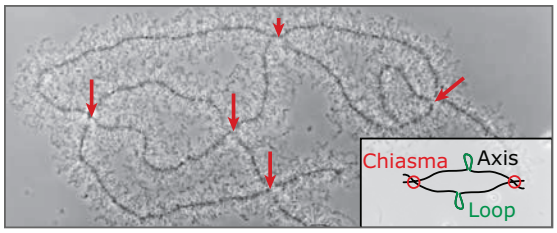

FIGURE 10.19 A lampbrush chromosome is a meiotic bivalent in which the two pairs of sister chromatids are held together at chiasmata (indicated by arrows). (Photo courtesy of Joseph G. Gall, Carnegie Institution.)

The lengths of the individual lampbrush chromosomes in the newt *Notophthalmus viridescens* range from 400 to 800 µm, compared with the range of 15–20 µm seen later in meiosis. Therefore, the lampbrush chromosomes are ~30 times less tightly packed. The total length of the entire lampbrush chromosome set is 5–6 mm, organized into ~5,000 chromomeres.

The lampbrush chromosomes take their name from the lateral loops that extrude from the chromomeres at certain positions. (These resemble a lampbrush, an extinct object.) The loops extend in pairs, one from each sister chromatid. The loops are continuous with the axial thread, which suggests they represent chromosomal material extruded from its more compact organization in the chromomere.

The loops are surrounded by a matrix of ribonucleoproteins. These contain nascent RNA chains. Often, a transcription unit can be defined by the increase in the length of the ribonucleoprotein moving around the loop. An example is shown in FIGURE 10.20.

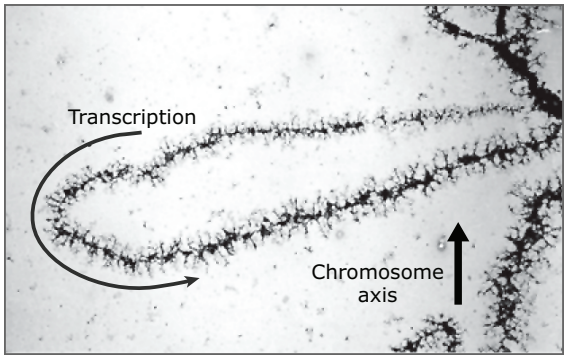

FIGURE 10.20 A lampbrush chromosome loop is surrounded by a matrix of ribonucleoprotein. (Photo courtesy of Oscar Miller.)

Therefore, the loop is an extruded segment of DNA that is being actively transcribed. In some cases, loops corresponding to particular genes have been identified. Then the structure of the transcribed gene, and the nature of the product, can be scrutinized in situ.

Concept and Reasoning Check

1. Why does transcription occur in the loops, but not the unlooped regions, of lampbrush chromosomes?

10.12 Polytene chromosomes form bands

Key concept

- Polytene chromosomes of dipterans have a series of bands that can be used as a cytologic map.

The interphase nuclei of some tissues of the larvae of dipteran ("two-winged") flies contain chromosomes that are greatly enlarged relative to their usual condition. They possess both increased diameter and greater length. **FIGURE 10.21** shows an example of a chromosome set from the salivary gland of *D. melanogaster.* These are called polytene chromosomes.

Each member of the polytene set consists of a visible series of bands (more properly, but rarely, described as chromomeres). The bands range in size from the largest with a breadth of ~0.5 µm to the smallest of ~0.05 µm. (The smallest can be distinguished only under an electron microscope.) The bands contain most of the mass of DNA and stain intensely with appropriate reagents. The regions between them stain more lightly and are called interbands. There are ~5,000 bands in the *D. melanogaster* set.

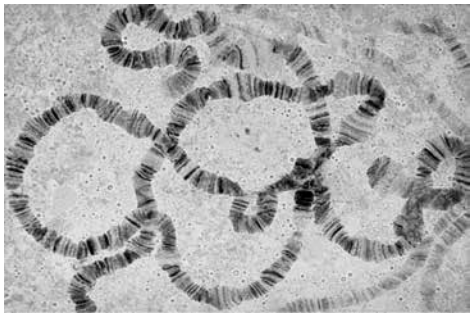

FIGURE 10.21 The polytene chromosomes of *D. melanogaster* form an alternating series of bands and interbands. (Photo courtesy of José Bonner, Indiana University.)

The centromeres of all four chromosomes of *D. melanogaster* aggregate to form a chromocenter that consists largely of heterochromatin (in the male it includes the entire Y chromosome). Allowing for this, ~75% of the haploid DNA set is organized into alternating bands and interbands. The DNA in extended form would stretch for ~40,000 µm. The length of the chromosome set is ~2,000 µm, about 100 times longer than the length of the haploid set at mitosis. This demonstrates vividly the extension of the genetic material relative to the usual states of interphase chromatin or mitotic chromosomes.

What is the structure of these giant chromosomes? Each is produced by the successive replications of a synapsed diploid pair. The replicas do not separate but remain attached to each other in their extended state. At the start of the process, each synapsed pair has a DNA content of 2C (where C represents the DNA content of the individual chromosome). This then doubles up to nine times, at its maximum giving a content of 1,024C. The number of doublings is different in the various tissues of the *D. melanogaster* larva. This process is known as *endoreduplication.*

Each chromosome can be visualized as a large number of parallel fibers running longitudinally, tightly condensed in the bands and less condensed in the interbands. Probably each fiber represents a single (C) haploid chromosome. This gives rise to the name polytene. The degree of polyteny is the number of haploid chromosomes contained in the giant chromosome.

The banding pattern is characteristic for each strain of *Drosophila.* The constant number and linear arrangement of the bands was first noted in the 1930s, when it was realized they form a cytologic map of the chromosomes. Rearrangements—such as deletions, inversions, or duplications—result in alterations of the order of bands.

The linear array of bands can be equated with the linear array of genes. Therefore, genetic rearrangements, as seen in a linkage map, can be correlated with structural rearrangements of the cytologic map. Ultimately, a particular mutation can be located in a particular band. Because the total number of genes in *D. melanogaster* exceeds the number of bands, there are probably multiple genes in most or all bands.

The positions of particular genes on the cytologic map can be determined directly by the technique of **fluorescent in situ**

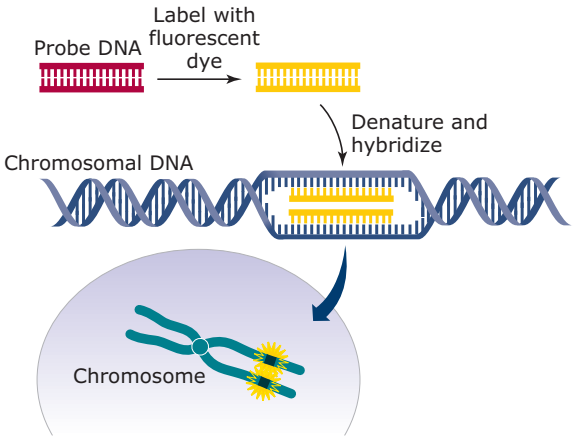

FIGURE 10.22 Fluorescence in situ hybridization (FISH). (Adapted from an illustration by Daryl Leja, National Human Genome Research Institute [www.genome.gov].)

hybridization (**FISH**). The protocol is summarized in **FIGURE 10.22**. A fluorescent probe representing a gene (most often a labeled complementary DNA clone derived from messenger RNA [mRNA]) is hybridized with the denatured DNA of the polytene chromosomes in situ. The position or positions of the corresponding genes are then detected at a particular band or bands by visualizing the fluorescent probe using a microscope that excites the fluorescent label at a suitable wavelength. With this type of technique, it is possible to determine directly the band within which a particular sequence lies. **BOX 10.1** describes variations on FISH that allow fluorescent detection of whole chromosomes and how these methods are applied in the study of cancer. Another powerful in situ method (immunohistochemistry) detects proteins bound to DNA rather than the DNA sequences themselves. This method uses antibodies (which can also be fluorescently labeled) to detect the proteins. An example is shown in **FIGURE 10.23**.

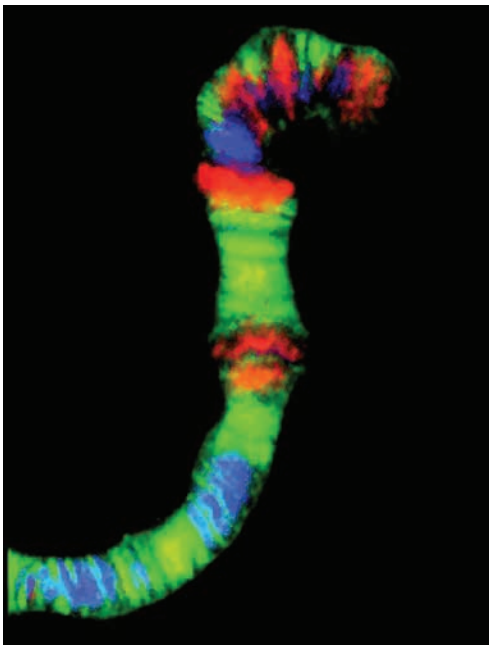

FIGURE 10.23 Detection of proteins bound to DNA in situ. In this example, two proteins involved in transcription are detected using antibodies that are conjugated to fluorescent dyes (red and green), whereas the DNA is counterstained (blue). (Reproduced from *Mol. Cell. Biol.*, 2003, vol. 23, pp. 3305–3319, DOI and reproduced with permission from American Society for Microbiology. Photo courtesy of Jerry L. Workman, Howard Hughes Medical Institute.)

Concept and Reasoning Check

1. If you made a FISH probe that hybridized to *Drosophila* centromeric DNA, where would the FISH signal appear in a polytene chromosome spread?

10.13 Polytene chromosomes expand at sites of gene expression

Key concept

- Bands that are sites of gene expression on polytene chromosomes expand to give "puffs."

One of the intriguing features of the polytene chromosomes is that active sites can be visualized. Some of the bands pass transiently through an expanded state in which they appear like a **puff** on the chromosome, when chromosomal material is extruded from the axis. An example of some very large puffs (called Balbiani rings) is shown in **FIGURE 10.24**.

What is the nature of the puff? It consists of a region in which the chromosome fibers unwind from their usual state of packing in the band. The fibers remain continuous with those in the chromosome axis. Puffs usually emanate from single bands, although when they are very large, as typified by Balbiani rings, the swelling may be so extensive as to obscure the underlying array of bands.

The pattern of puffs is related to gene expression. During larval development, puffs appear and regress in a definite, tissue-specific pattern. A characteristic pattern of puffs is found in each tissue at any given time. Puffs are induced by the hormone ecdysone that

BOX 10.1. MEDICAL APPLICATIONS: CHROMOSOME PAINTING AND SPECTRAL KARYOTYPING

Classic staining techniques such as G-banding (see Figure 10.4) have been used for decades to identify major chromosomal aberrations, such as large deletions, insertions, or translocations. These methods, however, can only detect large changes in chromosome organization, and many smaller aberrations are undetectable. Many cancers (particularly leukemias) are characterized by specific chromosome translocations that can play causative roles in cancer development or progression. Researchers have long been interested in developing methods for sensitive mapping of chromosome aberrations across the genome.

FISH is a classic method for detecting individual genes or sets of repeated sequences in intact chromosomes (see Figure 10.22). An important breakthrough came when researchers realized FISH did not need to be restricted to individual genes or segments of chromosomes but could instead be applied to whole chromosomes at once. This variation of FISH, called chromosome painting, uses a fluorescent dye bound to DNA probes that bind all along a particular chromosome. **FIGURE 10.B1** shows an example of the use of chromosome painting to visualize a translocation between chromosomes 6 and 13 in a patient with chronic lymphocytic leukemia.

The major shortcoming of FISH and chromosome painting is they cannot be used to study all chromosomes at the same time, because there are not enough fluorescent dyes with sufficient color differences to mark all 23 chromosomes in a unique color. This problem was solved in 1996 by labeling the painting probes for each chromosome with a different assortment of fluorescent dyes, a technique called spectral karyotyping (SKY). When the fluorescent probes hybridize to a chromosome, each kind of chromosome is labeled with a different assortment of fluorescent dye combinations. Stained chromosomes are then viewed through a series of filters,

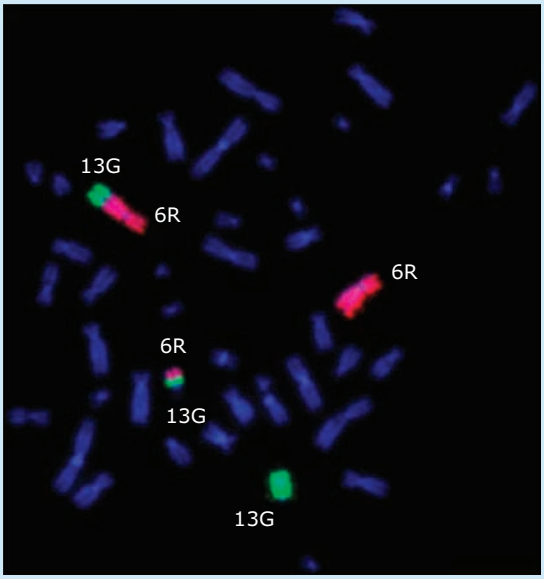

FIGURE 10.B1 Whole chromosome painting detects translocations between chromosome 6 (red) and chromosome 13 (green) in a patient with chronic lymphocytic leukemia. This study revealed loss of key sequences in the region of the translocation that would not be detectable by conventional banding techniques. (Photo courtesy of Claudia Haferlach, Munich Leukemia Laboratory, Germany.)

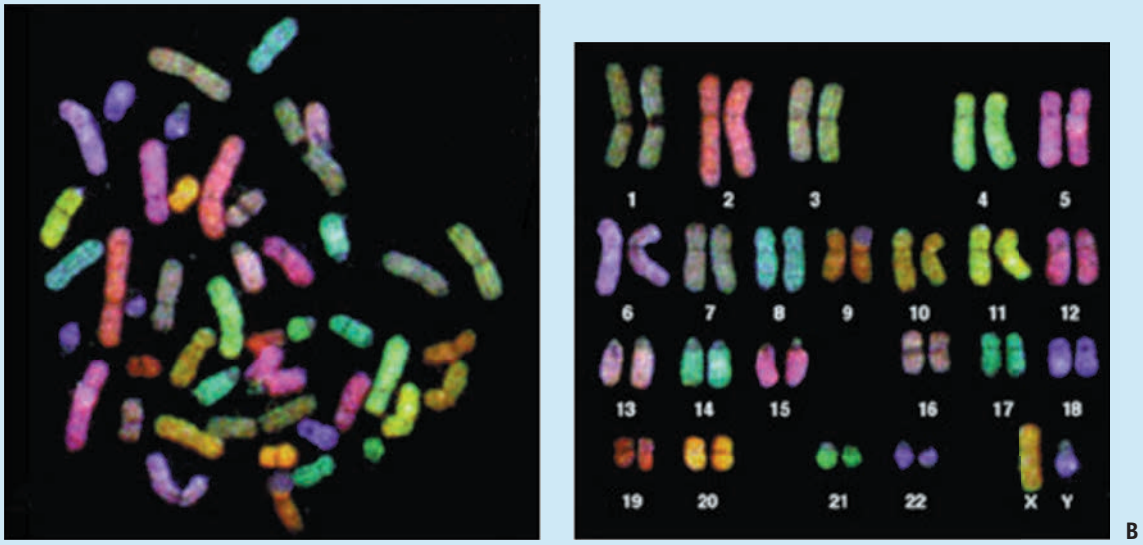

FIGURE 10.B2 SKY, an application of chromosome painting. (A) A chromosome spread hybridized with SKY fluorescent probes. (B) The same labeled chromosomes as in A, sorted by size to show the karyotype. (Photos courtesy of Johannes Wienberg, Ludwig-Maximilians-University, and Thomas Ried, National Institutes of Health.)

BOX 10.1. MEDICAL APPLICATIONS: CHROMOSOME PAINTING AND SPECTRAL KARYOTYPING

or an interferometer determines the full spectrum of light emitted by the stained chromosome. Then, a computer provides a composite picture that shows different chromosome pairs as if they were stained in different colors, as shown in **FIGURE 10.B2**.

SKY has allowed the systematic evaluation of chromosome aberrations in many types of cancers, revealing novel rearrangements, previously unrecognized translocations, and expansions or deletions of key chromosomal loci. Some chromosomal aberrations are highly complex, making SKY a powerful tool for unraveling these complicated rearrangements. An example is shown in **FIGURE 10.B3**, which shows the highly aberrant chromosomes from a pancreatic cancer cell line. As more and more data accumulates, SKY and other results can now be submitted to a public database (http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi) that allows researchers to share results and identify recurrent patterns of aberrations that can lead to new understanding of how specific aberrations can contribute to carcinogenesis and how ongoing chromosomal instability affects the progression of cancer.

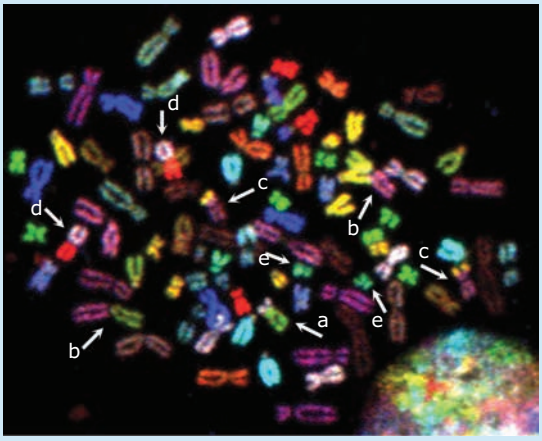

FIGURE 10.B3 SKY analysis of chromosome aberrations in a pancreatic cancer cell line. Arrows indicate numerous rearrangements, which are mostly unbalanced translocations. (Reproduced from C. A. Griffin, et al., Cytogenet. Genome Res. 118 (2007): 148–156. Copyright 2007, S. Karger AG, Basel. Photo courtesy of Constance A. Griffin, Division of Molecular Pathology, Johns Hopkins Hospital.)

controls *Drosophila* development. Some puffs are induced directly by the hormone; others are induced indirectly by the products of earlier puffs. The puffs are sites where RNA is being synthesized. The accepted view of puffing has been that expansion of the band is a consequence of the need to relax its structure to synthesize RNA. Puffing has therefore been viewed as a consequence of transcription. A puff can be generated by a single active gene. The sites of puffing differ from ordinary bands in accumulating additional proteins, which include RNA polymerase II and other proteins associated with transcription.

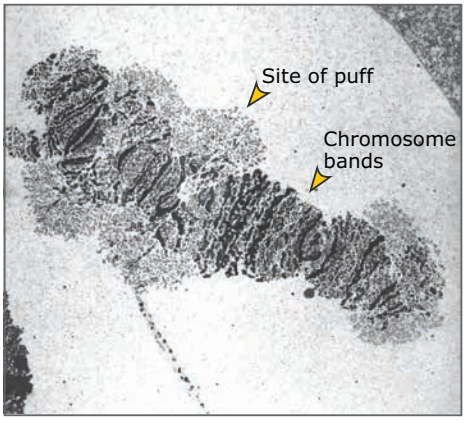

FIGURE 10.24 Chromosome IV of the insect *C. tentans* has three Balbiani rings in the salivary gland. (Reprinted from *Cell,* vol. 4, B. Daneholt, Transcription in polytene chromosomes, pp. 1–9, Copyright (1975) with permission from Elsevier [http://www.sciencedirect.com/science/journal/00928674]. Photo courtesy of Bertil Daneholt, Karolinska Institutet.)

The features displayed by lampbrush and polytene chromosomes suggest a general conclusion. To be transcribed, the genetic material is dispersed from its usual more tightly packed state. The question to keep in mind is whether this dispersion at the gross level of the chromosome mimics the events that occur at the molecular level within the mass of ordinary interphase euchromatin.

Do the bands of a polytene chromosome have a functional significance, that is, does each band correspond to some type of genetic unit? You might believe the answer would be immediately evident from the sequence of the fly genome, because by mapping interbands to the sequence, it should be possible to determine whether a band has any fixed type of identity.

So far, however, no pattern has been found that identifies a functional significance for the bands.

Concept and Reasoning Check

1. How do the bands on polytene chromosomes differ from the G-bands described in *10.3 Chromosomes have banding patterns?*

10.14 The nucleosome is the subunit of all chromatin

Key concepts

- Micrococcal nuclease releases individual nucleosomes from chromatin as ~10-nm particles.
- A nucleosome contains ~200 bp of DNA, two copies of each core histone (H2A, H2B, H3, and H4), and one copy of H1.
- DNA is wrapped around the outside surface of the protein octamer.

Chromatin and chromosomes are constructed from a deoxyribonucleoprotein fiber that has several hierarchies of organization. Its most complex, fully folded state is seen in the banded structure of the mitotic chromosome. Its basic subunit is the same in all eukaryotes: a nucleosome consisting of ~200 bp of DNA and histone proteins. The histones are small, basic proteins (rich in arginine and lysine residues), resulting in a high affinity for DNA. Nonhistone proteins also contribute to the folding of the thread of nucleosomes into the structures of higher-order fibers.

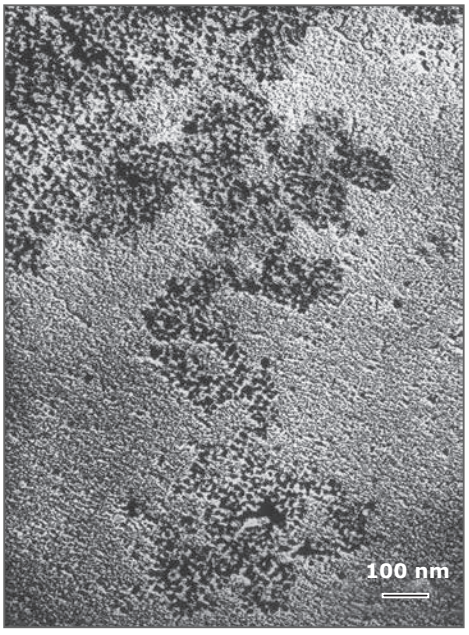

FIGURE 10.25 Chromatin spilling out of lysed nuclei consists of a compactly organized series of particles. (Reprinted from *Cell,* vol. 4, P. Oudet, M. Gross-Bellard, and P. Chambon, Electron microscopic and biochemical evidence ... , pp. 281–300, Copyright (1975) with permission from Elsevier [http://www.sciencedirect.com/science/journal/00928674]. Photo courtesy Pierre Chambon, College of France.)

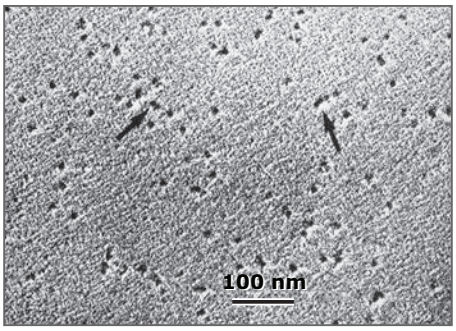

FIGURE 10.26 Individual nucleosomes are released by digestion of chromatin with micrococcal nuclease. (Reprinted from *Cell,* vol. 4, P. Oudet, M. Gross-Bellard, and P. Chambon, Electron microscopic and biochemical evidence ... , pp. 281–300, Copyright (1975) with permission from Elsevier [http://www.sciencedirect.com/science/journal/00928674]. Photo courtesy Pierre Chambon, College of France.)

When interphase nuclei are suspended in a solution of low ionic strength, they swell and rupture to release fibers of chromatin. **FIGURE 10.25** shows a lysed nucleus in which fibers are streaming out. In some regions, the fibers consist of tightly packed material, but in regions that have become stretched, they can be seen to consist of discrete particles. These are the nucleosomes. In especially extended regions, individual nucleosomes are connected by a fine thread, a free duplex of DNA. A continuous duplex thread of DNA runs through the series of particles.

Individual nucleosomes can be obtained by treating chromatin with the endonuclease **micrococcal nuclease**. It cuts the DNA thread at the junction between nucleosomes. First, it releases groups of particles; finally, it releases single nucleosomes. Individual nucleosomes can be seen in **FIGURE 10.26** as compact particles of ~10–11 nm in diameter.

The nucleosome contains ~200 bp of DNA associated with a histone octamer that consists of two copies each of H2A, H2B, H3, and H4. These are known as the **core histones**. Their association is illustrated diagrammatically in **FIGURE 10.27**. This model explains the stoichiometry of the core histones in chromatin: H2A, H2B, H3, and H4 are present in equimolar amounts, with two molecules of each per ~200 bp of DNA.

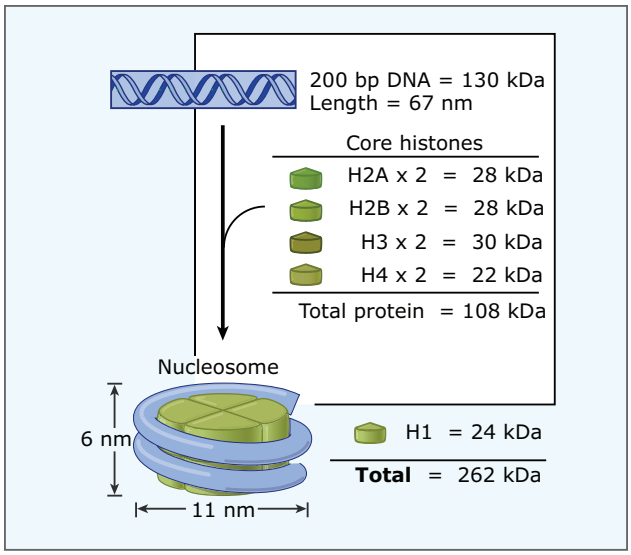

FIGURE 10.27 The nucleosome consists of approximately equal masses of DNA and histones (including H1). The predicted mass of the nucleosome is 262 kDa.

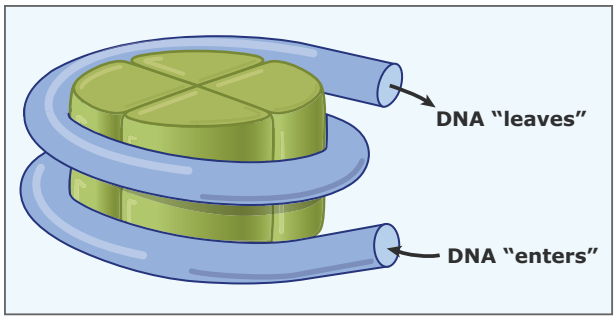

FIGURE 10.28 The nucleosome is roughly cylindrical with DNA organized into 1¾ turns around the surface.

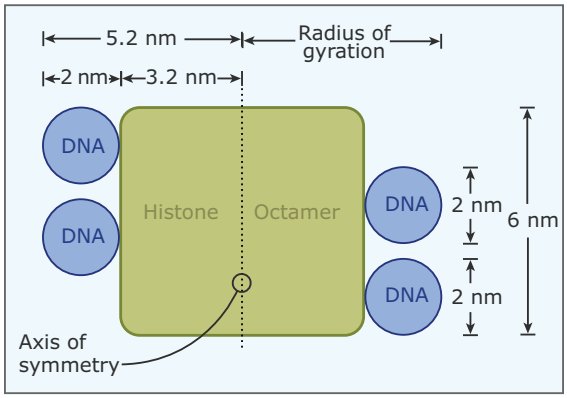

FIGURE 10.29 The two turns of DNA on the nucleosome lie close together.

Histones H3 and H4 are among the most conserved proteins known, and the core histones are responsible for DNA packaging in all eukaryotes. H2A and H2B are also conserved among eukaryotes but show appreciable species-specific variation in sequence, particularly in the histone tails. The core regions of the histones are even conserved in archaea and appear to play a similar role in compaction of archaeal DNA.

The **linker histones**, typified by histone H1, are a set of related proteins that show appreciable variation between tissues and between species. The role of linker histone is different from the core histones. It is present in half the amount of a core histone and can be extracted more readily from chromatin (typically with a moderate salt [0.5 M] solution). Linker histone can be removed without affecting the structure of the nucleosome, which suggests its location is external to the particle.

The shape of the nucleosome corresponds to a flat disk or cylinder, of diameter 11 nm and height 6 nm. The length of the DNA is roughly twice the ~34-nm circumference of the particle. The DNA follows a symmetric path around the octamer. **FIGURE 10.28** shows the DNA path diagrammatically as a helical coil that makes approximately 1¾ turns around the cylindrical octamer. Note the DNA "enters" and "leaves" the nucleosome at points somewhat close together on one side of the nucleosome. Histone H1 may be located in this region (see *10.16 Nucleosomes have a common structure*).

Considering this model in terms of a cross-section through the nucleosome, in **FIGURE 10.29** we see that the two circumferences made by the DNA lie close to one another. The height of the cylinder is 6 nm, of which 4 nm is occupied by the two turns of DNA (each of diameter 2 nm).

The pattern of the two turns has a possible functional consequence. Because one turn around the nucleosome takes ~80 bp of DNA, two points separated by 80 bp in the free double helix may actually be close on the nucleosome surface, as illustrated in **FIGURE 10.30**.

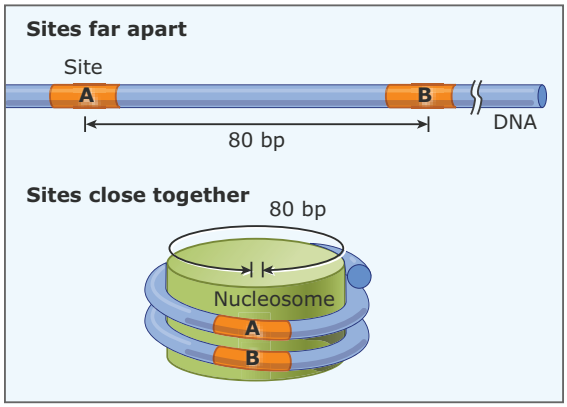

FIGURE 10.30 Sequences on the DNA that lie on different turns around the nucleosome may be close together.

10.15 DNA is coiled in arrays of nucleosomes

Key concepts

- Greater than 95% of the DNA is recovered in nucleosomes or multimers when micrococcal nuclease cleaves DNA of chromatin.
- The length of DNA per nucleosome varies for individual tissues (or species) in a range from 154 to 260 bp.

When chromatin is digested with the enzyme micrococcal nuclease, the DNA is cleaved into integral multiples of a unit length. Fractionation by gel electrophoresis reveals the "ladder" presented in **FIGURE 10.31**. Such ladders typically extend for ~10 steps, and the unit length, determined by the increments between successive steps, is ~200 bp.

FIGURE 10.32 shows the ladder is generated by groups of nucleosomes. When nucleosomes are fractionated on a sucrose gradient, they give a series of discrete peaks that correspond to monomers, dimers, trimers, and so on. When the DNA is extracted from the individual fractions and electrophoresed, each fraction yields a band of DNA whose size corresponds with a step on the micrococcal nuclease ladder. The monomeric nucleosome contains DNA of the unit length, the nucleosome dimer contains DNA of twice the unit length, and so on.

Therefore, each step on the ladder represents the DNA derived from a discrete number of nucleosomes. We take the existence of the 200-bp ladder in any chromatin to indicate that the DNA is organized into nucleosomes. The micrococcal ladder is generated when only ~2% of the DNA in the nucleus is rendered acid-soluble (degraded to small fragments) by the enzyme. Therefore, a small proportion of the DNA is specifically attacked; it must represent especially susceptible regions.

When chromatin is spilled out of nuclei, we often see a series of nucleosomes connected by a thread of free DNA (the beads on a string).

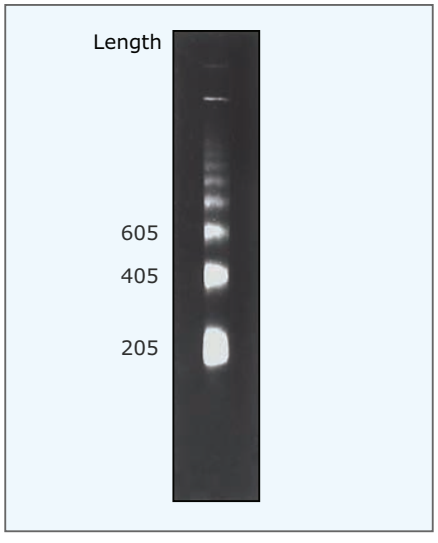

FIGURE 10.31 Micrococcal nuclease digests chromatin in nuclei into a multimeric series of DNA bands that can be separated by gel electrophoresis. (Photo courtesy of Markus Noll, Universität Zürich.)

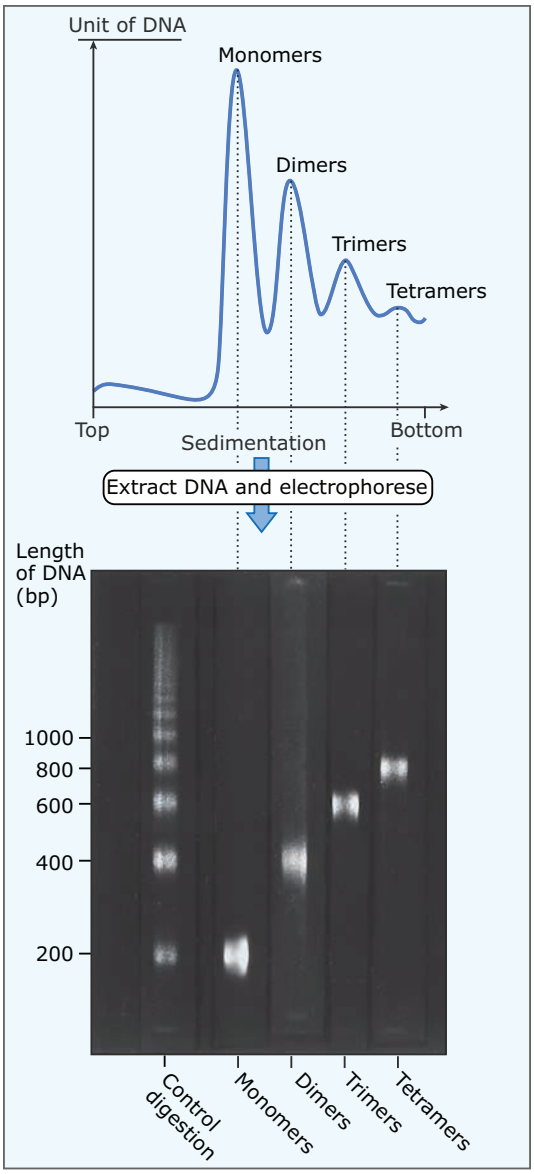

FIGURE 10.32 Each multimer of nucleosomes contains the appropriate number of unit lengths of DNA. (Photo courtesy of John Finch, MRC Laboratory of Molecular Biology.)

The need for tight packaging of DNA in vivo suggests that there is usually little (if any) free DNA.

This view is confirmed by the fact that >95% of the DNA of chromatin can be recovered in the form of the 200-bp ladder. Almost all DNA must therefore be organized in nucleosomes. In their natural state, nucleosomes are likely to be closely packed, with DNA passing directly from one to the next. Free DNA is extremely rare in vivo, although some is probably generated by the loss of some histone octamers during isolation.

The length of DNA present in the nucleosome varies somewhat from the "typical" value of 200 bp. The chromatin of any particular cell type has a characteristic average value (±5 bp). The average most often is between 180 and 200 bp, but there are extremes as low as 154 bp (in a fungus) or as high as 260 bp (in a sea urchin sperm). The average value may be different in individual tissues of the adult organism. Also, there can be differences between different parts of the genome in a single cell type. Variations from the genome average include tandemly repeated sequences, such as clusters of 5S RNA genes.

Concept and Reasoning Check

1. If you performed MNase digestion on a species with extremely heterogeneous linker DNA lengths, how would this affect the appearance of the MNase ladder?

10.16 Nucleosomes have a common structure

Key concepts

- Nucleosomal DNA is divided into the core DNA and linker DNA depending on its susceptibility to micrococcal nuclease.
- There are 1.65 turns of DNA wound around the histone octamer.
- The core DNA is the length of 146 bp that is found on the core particles produced by prolonged digestion with micrococcal nuclease.
- Linker DNA is the region of 8–114 bp that is susceptible to early cleavage by the enzyme.
- Changes in the length of linker DNA account for the variation in total length of nucleosomal DNA.
- Linker histone is associated with linker DNA and may lie at the point where DNA enters or leaves the nucleosome.
- The structure of the DNA is altered so it has an increased number of base pairs per turn in the middle but a decreased number at the ends.

A common structure underlies the varying amount of DNA contained in nucleosomes from different sources. The association of DNA with the histone octamer forms a core particle containing 146 bp of DNA, irrespective of the total length of DNA in the nucleosome. The variation in total length of DNA per nucleosome is superimposed on this basic core structure.

The amount of DNA in the core particle can be defined by the effects of micrococcal nuclease on the nucleosome monomer. The initial reaction of the enzyme is to cut between nucleosomes, but if it is allowed to continue after monomers have been generated, then it proceeds to digest some of the DNA of the individual nucleosome. This occurs by a reaction in which DNA is "trimmed" from the ends of the nucleosome.

The length of the DNA is reduced in discrete steps, as shown in **FIGURE 10.33**. With rat liver nuclei, the nucleosome monomers initially have 205 bp of DNA. Some monomers are found, however, in which the length of DNA has been reduced to ~165 bp. Finally, this is reduced to the length of the DNA of the core particle, 146 bp. (The core is reasonably stable, but continued digestion generates a "limit digest" in which the longest fragments are the 146 bp DNA of the core, whereas the shortest are as small as 20 bp.)

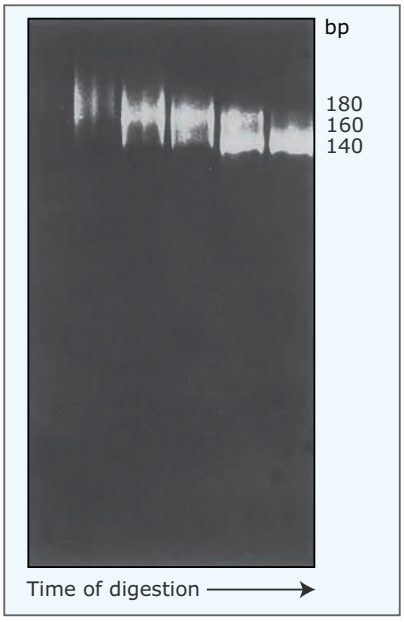

FIGURE 10.33 Micrococcal nuclease reduces the length of nucleosome monomers in discrete steps. (Photo courtesy of Roger Kornberg, Stanford University.)

This analysis suggests nucleosomal DNA can be divided into the following two regions:

1. Core DNA has an invariant length of 146 bp and is relatively resistant to digestion by nucleases.
2. Linker DNA comprises the rest of the repeating unit. Its length varies from as little as 8 bp to as much as 114 bp per nucleosome.

The sharp size of the band of DNA generated by the initial cleavage with micrococcal nuclease suggests that the region immediately available to the enzyme is restricted. It represents only part of each linker. (If the entire linker DNA were susceptible, the band would range from 146 to >200 bp.) Once a cut has been made in the linker DNA, however, the rest of this region becomes susceptible, and it can be removed relatively rapidly by further enzyme action. The connection between nucleosomes is represented in **FIGURE 10.34**.

As suggested by their names, linker histones such as H1 are thought to interact with linker DNA. H1 is lost during the degradation of nucleosome monomers. It can be retained on monomers that still have 165 bp of DNA but is always lost with the final reduction to the 146-bp core particle. This suggests that H1 could be located in the region of the linker DNA immediately adjacent to the core DNA. Although the precise positioning of linker histones remains somewhat controversial, recent models suggest H1 may interact with either the entry or exit DNA in addition to the central turn of DNA on the nucleosome, as shown in **FIGURE 10.35**. In this position, H1 has the potential to influence the angle of DNA entry or exit, which may contribute to the formation of higher-order structures (see *10.19 The path of nucleosomes in the chromatin fiber*).

Whereas MNase cleaves linker DNA, other nucleases can cleave DNA that is exposed on the surface of the nucleosome. The reaction with nucleases that attack single strands has been especially informative. These enzymes (such as DNase I) make single-strand nicks in DNA; they cleave a bond in one strand, but the other strand remains intact at this point. Therefore, no effect is visible in the double-stranded DNA. Upon denaturation, however, short fragments are released instead of full-length single strands. If the DNA has been labeled at its ends, the end fragments can be identified by autoradiography, as summarized in **FIGURE 10.36**.

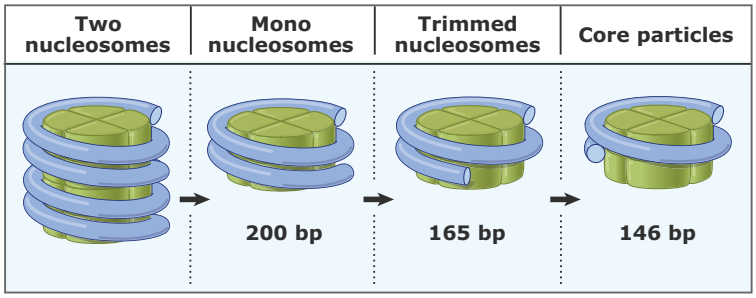

FIGURE 10.34 Micrococcal nuclease initially cleaves between nucleosomes. Mononucleosomes typically have ~200 bp DNA. End-trimming reduces the length of DNA first to ~165 bp and then generates core particles with 146 bp.

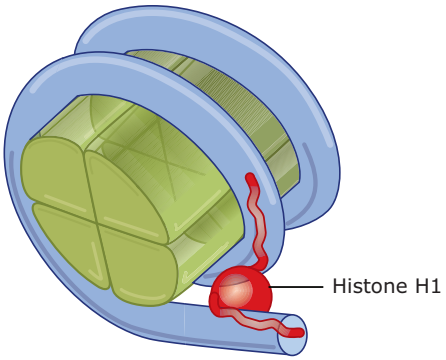

FIGURE 10.35 Model for histone H1 interaction with the nucleosome. H1 interacts with linker DNA near the entry or exit site and may also interact with the central turn of DNA near the dyad axis of the nucleosome.

When free DNA in solution is treated with enzymes like DNase I, it is nicked (relatively) at random. The DNA on nucleosomes also can be nicked by the enzymes, *but only at regular intervals*. When the points of cutting are determined by using radioactively end-labeled DNA and then DNA is denatured and electrophoresed, a ladder of the sort displayed in **FIGURE 10.37** is obtained. The interval between successive steps on the ladder is 10–11 bases. The ladder extends for the full distance of core DNA. The cleavage sites are numbered S1 through S13 (where S1 is ~10 bases from the labeled 5′ end, S2 is ~20 bases from it, and so on). The enzymes DNase I and DNase II, as well as hydroxyl radical cleavage, all generate essentially the same ladder. This shows that the pattern of cutting represents a unique series of targets in DNA, determined by its organization, rather than preferences for particular sites imposed by the individual enzyme or other treatment. The lack of reaction at particular target sites results from the structure of the nucleosome, in which certain positions on DNA are rendered inaccessible.

When DNA is immobilized on a flat surface, sites are cut with a regular separation.

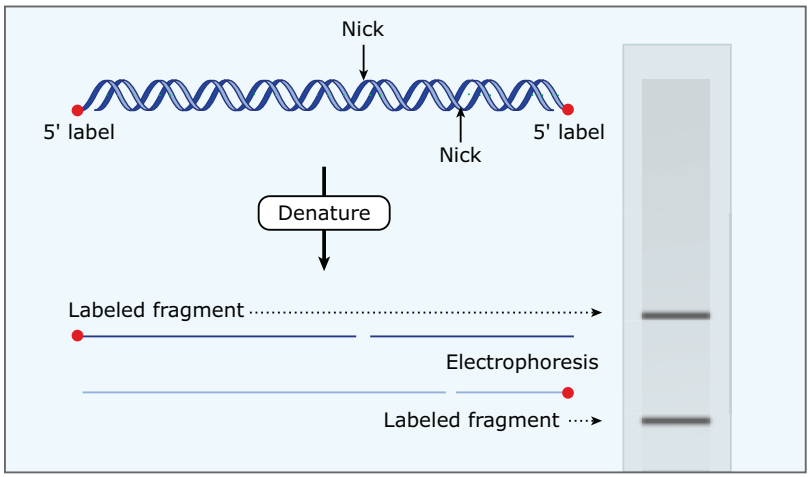

FIGURE 10.36 Nicks in double-stranded DNA are revealed by fragments when the DNA is denatured to give single strands. If the DNA is labeled at (say) 5′ ends, only the 5′ fragments are visible by autoradiography. The size of the fragment identifies the distance of the nick from the labeled end.

FIGURE 10.38 suggests that this reflects the recurrence of the exposed site with the helical periodicity of B-form DNA (the classical form of duplex DNA discovered by Watson and Crick). The cutting periodicity (the spacing between cleavage points) coincides with—indeed, is a reflection of—the structural periodicity (the number of base pairs per turn of the double helix). So the distance between the sites corresponds to the number of base pairs per turn. Measurements of this type yield the average value for double-helical B-type DNA of 10.5 bp per turn.

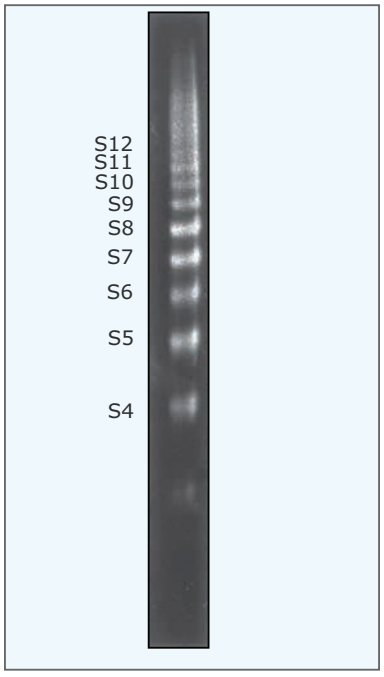

FIGURE 10.37 Sites for nicking lie at regular intervals along core DNA, as seen in a DNase I digest of nuclei. (Photo courtesy of Leonard Lutter, Henry Ford Hospital.)

A similar analysis of DNA on the surface of the nucleosome reveals striking variation in the structural periodicity at different points. At the ends of the DNA, the average distance between pairs of DNase I digestion sites is about 10.0 bases each, significantly less than the usual 10.5 bp per turn. In the center of the particle, the separation between cleavage sites averages 10.7 bases. This variation in cutting periodicity along the core DNA means there is variation in the structural periodicity of core DNA. The DNA has more base pairs per turn than its solution value in the middle but has fewer base pairs per turn at the ends. The average periodicity over the entire nucleosome is only 10.17 bp per turn, which is significantly less than the 10.5 bp per turn of DNA in solution.

The crystal structure of the core particle suggests that DNA is organized as a flat superhelix, with 1.65 turns wound around the histone octamer (see *10.17 Organization of the histone octamer*). A high-resolution structure of the nucleosome core shows in detail how the structure of DNA is distorted. The central 129 bp are in the form of B-DNA, but with a substantial curvature that is needed to form the superhelix. The major groove is smoothly bent, but the minor groove contains some abrupt

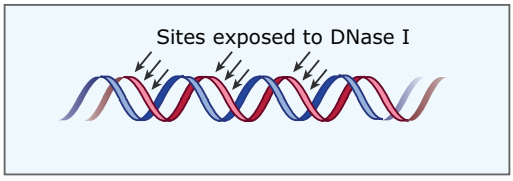

FIGURE 10.38 The most exposed positions on DNA recur with a periodicity that reflects the structure of the double helix. (For clarity, sites are shown for only one strand.)

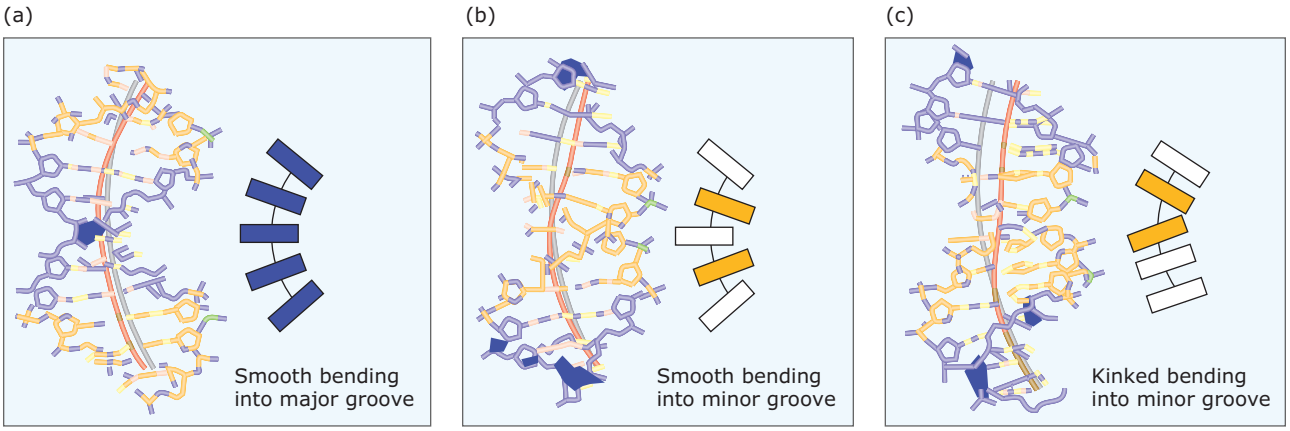

FIGURE 10.39 DNA bending in nucleosomal DNA. Structures (left) and schematic representations (right) show uniformity of curvature along the major groove (red) and both smooth and kinked bending into the minor groove (yellow). Also indicated are the DNA axes for the experimental (gold) and ideal (white) superhelices. (Adapted from T. J. Richmond and C. A. Davey, *Nature* 423 (2003): 145–150.)

kinks, as shown in **FIGURE 10.39**. These conformational changes may explain why the central part of nucleosomal DNA is not usually a target for binding by regulatory proteins, which typically bind to the terminal parts of the core DNA or to the linker sequences.

Concept and Reasoning Check

1. Explain the different digestion patterns generated by MNase and DNase on nucleosomal DNA.

10.17 Organization of the histone octamer

Key concepts

- The histone octamer has a structural core of an $H3_2 \cdot H4_2$ tetramer associated with two H2A · H2B dimers.
- Each histone is extensively interdigitated with its partner.
- All core histones have the structural motif of the histone fold.
- The histone N-terminal tails extend out of the nucleosome.

So far, we have considered the construction of the nucleosome from the perspective of how the DNA is organized on the surface. From the perspective of protein, we need to know how the histones interact with each other and with DNA.

The core histones form two types of complexes in the absence of DNA. H3 and H4 form a tetramer ($H3_2 \cdot H4_2$), whereas H2A and H2B primarily form a dimer (H2A · H2B). Complete histone octamers can be obtained either by extraction from chromatin or by assembling histones in vitro under conditions of high salt and high protein concentrations. The octamer easily dissociates to generate a hexamer of histones that has lost an H2A · H2B dimer. Then, the other H2A · H2B dimer is lost separately, leaving the $H3_2 \cdot H4_2$ tetramer. This argues for a form of organization in which the nucleosome has a central "kernel" consisting of the $H3_2 \cdot H4_2$ tetramer. This tetramer can organize DNA in vitro into particles that display some of the properties of the core particle.

Structural studies show the overall shape of the isolated histone octamer is similar to that of the core particle. This suggests that the histone–histone interactions establish the general structure. This can be seen in the space-filling model of the 3.1-Å resolution crystal structure of the histone octamer, shown in **FIGURE 10.40**. Tracing the paths of the individual polypeptide backbones in the crystal structure indicates that the histones are not organized as individual globular proteins but that each is interdigitated with its partner: H3 with H4 and H2A with H2B. Figure 10.40 distinguishes the $H3_2 \cdot H4_2$ tetramer (white) from the H2A · H2B dimers (blue) but does not show individual histones. The $H3_2 \cdot H4_2$ tetramer accounts for the diameter of the octamer. It forms the shape of a horseshoe. The H2A · H2B pairs fit in as two dimers, but only one can be seen in this view. In the side view, the responsibilities of the $H3_2 \cdot H4_2$ tetramer and of the separate H2A · H2B dimers can be distinguished. The protein forms a sort of spool, with a superhelical path corresponding to the binding site for DNA. The model displays twofold symmetry about an axis that would run perpendicular through the side view.

A more detailed view of the positions of the histones (based on a crystal structure at 2.8 Å) is summarized in **FIGURE 10.41**. The upper

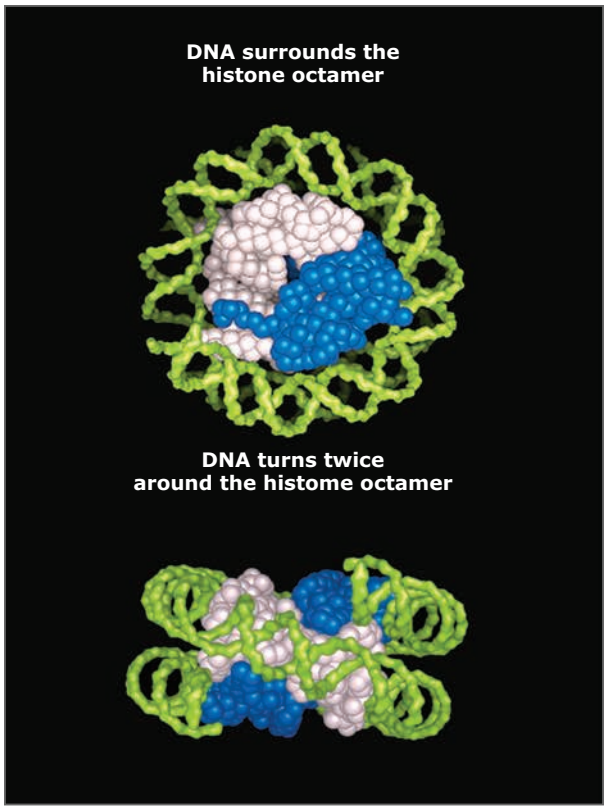

FIGURE 10.40 The crystal structure of the histone core octamer is represented in a space-filling model with the $H3_2 \cdot H4_2$ tetramer shown in white and the H2A • H2B dimers shown in blue. Only one of the H2A • H2B dimers is visible in the top view, because the other is hidden underneath. The path of the DNA is shown in green. (Photos courtesy of E. N. Moudrianakis, John Hopkins University.)

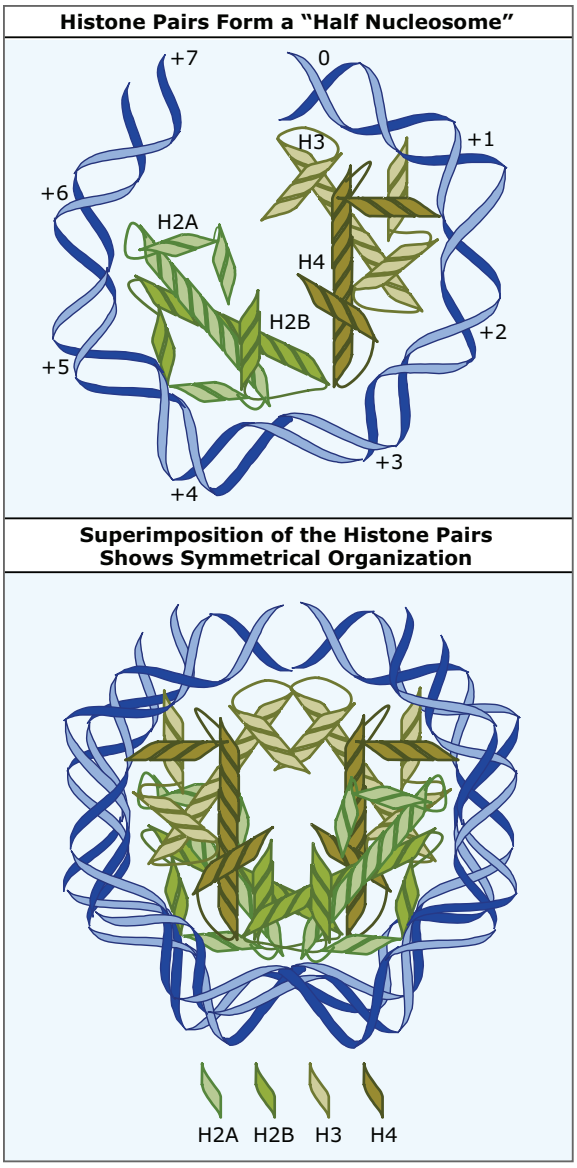

FIGURE 10.41 Histone positions in a top view show H3 • H4 and H2A • H2B pairs in a half nucleosome; the symmetric organization can be seen in the superimposition of both halves.

view shows the position of one histone of each type relative to one turn around the nucleosome (numbered from 0 to +7). All four core histones show a similar type of structure in which three α-helices are connected by two loops: This is called the **histone fold**. These regions interact to form crescent-shaped heterodimers; each heterodimer binds 2.5 turns of the DNA double helix. Binding is mostly to the phosphodiester backbones (consistent with the need to package any DNA irrespective of sequence). The $H3_2 \cdot H4_2$ tetramer is formed by interactions between the two H3 subunits, as can be seen in the lower part of Figure 10.41.

In addition to the globular histone core domains, all histones also contain flexible N-terminal tails (H2A and H2B have substantial C-terminal tails as well), which contain sites for modification that are important in chromatin function. The tails, which account for about one-fourth of the protein mass, are too flexible to be visualized by x-ray crystallography. Therefore, their positions in the nucleosome are not well defined, and they are generally depicted schematically as seen in **FIGURE 10.42**. To the contrary, points at which the tails exit the nucleosome core are known, and the tails of both H3 and H2B can be seen to pass between the turns of the DNA superhelix and extend out of the nucleosome, as seen in **FIGURE 10.43.** The tails of H4 and H2A extend from both faces of the nucleosome. All histone tails are positively charged and can interact with DNA on the nucleosome or in the linker. The tail of H4 is able to contact an H2A • H2B dimer in an adjacent nucleosome, and the tails are required for formation of higher-order structure (see *10.19 The path of nucleosomes in the chromatin fiber*).

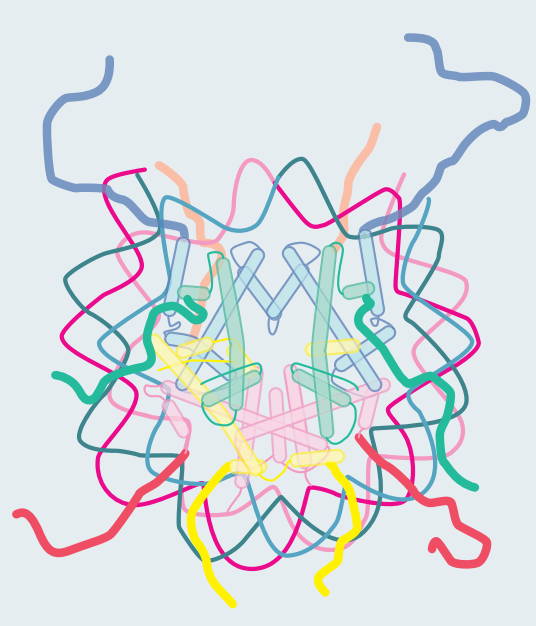

FIGURE 10.42 The histone-fold domains of the histones are located in the core of the nucleosome. The N- and C-terminal tails, which carry many sites for modification, are flexible and their positions cannot be determined by crystallography.

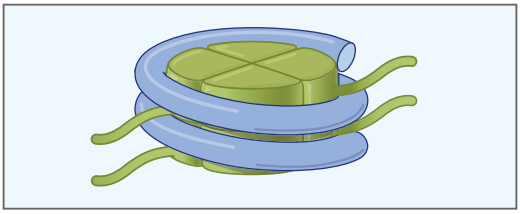

FIGURE 10.43 The N-terminal histone tails are disordered and exit from the nucleosome between turns of the DNA.

Concept and Reasoning Check

1. Histones contain many basic amino acids (and therefore carry a positive charge). Why?

10.18 Histone variants produce alternative nucleosomes

Key concepts

- All core histones except H4 are members of families of related variants.
- Histone variants can be closely related or highly divergent from canonical histones.
- Different variants serve different functions in the cell.

Although all nucleosomes share a related core structure, some nucleosomes exhibit subtle or dramatic differences resulting from the incorporation of histone variants. Histone variants comprise a large group of histones that are related to the histones we have already discussed but have differences in sequence from the "canonical" histones. These sequence differences can be small (as few as four amino acid differences) or extensive (such as alternative tail sequences).

Variants have been identified for all core histones except histone H4. The best-characterized histone variants are summarized in **FIGURE 10.44**. Most variants have significant differences between them, particularly in the N- and C-terminal tails. At one extreme, macroH2A is nearly three times larger than conventional H2A and contains a large C-terminal tail that is not related to any other histone. At the other end of the spectrum, the H3.3 variant differs from canonical H3 (also known as H3.1) at only four amino acid positions, three in the histone core, and one in the N-terminal tail.

Histone variants have been implicated in a number of different functions, and their incorporation changes the nature of the chromatin containing the variant. We have already discussed one type of histone variant, the centromeric H3 (or CenH3) histone, known as Cse4 in yeast and CENP-A in higher eukaryotes. CenH3 histones are incorporated into specialized nucleosomes present at centromeres in all eukaryotes (discussed in *10.8 The centromere binds a protein complex*). In yeast, these centromeric nucleosomes have been shown to consist of Cse4, H4, and a nonhistone protein Scm3, which replaces H2A/H2B dimers. In *Drosophila*, the centromeric chromatin may consist of "hemisomes" containing one copy each of CenH3, H4, H2A, and H2B.

The other major H3 variant is histone H3.3. H3.3 is expressed throughout the cell cycle, in contrast to most histones, which are expressed exclusively during S phase, when new chromatin assembly is required during DNA replication. As a result, H3.3 is available for assembly at any time in the cell cycle and is incorporated at sites of active transcription, where nucleosomes become disrupted. Because of this, H3.3 is often referred to as a "replacement" histone, in contrast to the "replicative" histone H3.1 (discussed further in *10.20 Replication of chromatin requires assembly of nucleosomes*).

The H2A variants are the largest and most diverse family of core histone variants and have

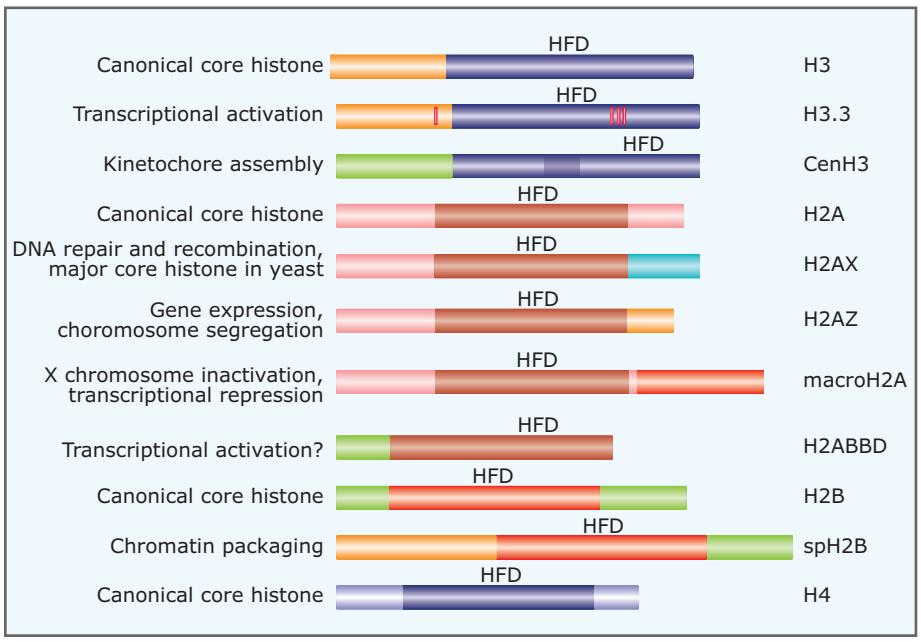

FIGURE 10.44 The major core histones contain a conserved histone-fold domain. In the histone H3.3 variant, the residues that differ from the major histone H3 (also known as H3.1) are highlighted in yellow. The centromeric histone CenH3 has a unique N-terminus, which does not resemble other core histones. Most H2A variants contain alternative C-termini, except H2ABbd, which contains a distinct N-terminus. The sperm-specific SpH2B has a long N-terminus. Proposed functions of the variants are listed. (Adapted from K. Sarma and D. Reinberg, *Nat. Rev. Mol. Cell Biol.* 6 (2005): 139–149.)

been implicated in a variety of distinct functions. One of the best studied is the variant H2AX. H2AX is normally present in only 10%–15% of the nucleosomes in multicellular eukaryotes, although this subtype is the major H2A present in yeast (this is also true for H3.3). H2AX has a C-terminal tail that is distinct from the canonical H2A, characterized by a SQEL/Y motif at the end. This motif is the target of phosphorylation by ATM/ATR kinases, activated by DNA damage, and this histone variant is involved in DNA repair, particularly repair of double-strand breaks. H2AX phosphorylated at the SQEL/Y motif is referred to as "γ-H2AX" and is required to stabilize binding of various repair factors at DNA breaks and to maintain checkpoint arrest.

Other H2A variants have different roles. The H2AZ variant, which has ~60% sequence identity with canonical H2A, has been shown to be important in several processes, such as gene activation, heterochromatin–euchromatin boundary formation, and cell cycle progression. The vertebrate-specific macroH2A is named for its extremely long C-terminal tail, which contains a leucine-zipper dimerization motif that may mediate chromatin compaction by facilitating internucleosome interactions. Mammalian macroH2A is enriched in the inactive X chromosome in females, which is assembled into a silent, heterochromatic state (see *10.28 X chromosomes undergo global changes*). In contrast, the mammalian H2ABbd variant is *excluded* from the inactive X and forms a less stable nucleosome than canonical H2A. This histone may be designed to be more easily displaced in transcriptionally active regions of euchromatin.

Still other variants are expressed in limited tissues, such as SpH2B, present in sperm and required for chromatin compaction. The presence and distribution of histone variants show that individual chromatin regions, entire chromosomes, or even specific tissues can have unique "flavors" of chromatin specialized for different function. In addition, the histone variants, like the canonical histones, are subject to numerous covalent modifications (see *10.25 Histones are covalently modified*), adding levels of complexity to the roles chromatin plays in nuclear processes.

Concept and Reasoning Check

1. Why would a less stable nucleosome (such as one containing H2ABbd) more likely be present in euchromatin than in heterochromatin?

10.19 The path of nucleosomes in the chromatin fiber

Key concepts

- Ten-nanometer chromatin fibers are unfolded from 30-nm fibers and consist of a string of nucleosomes.
- Thirty-nanometer fibers have six nucleosomes per turn, organized into a double solenoid.
- Histone H1 promotes the formation of the 30-nm fiber.

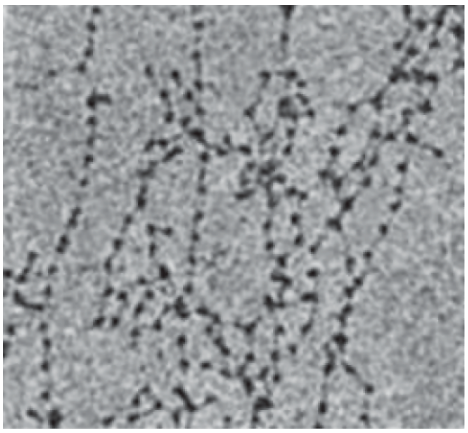

FIGURE 10.45 The 10-nm fiber in the partially unwound state can be seen to consist of a string of nucleosomes. (Photo courtesy of Barbara A. Hamkalo, University of California, Irvine.)

When chromatin is examined in the electron microscope, two types of fibers are seen: 10-nm fiber and 30-nm fiber. They are described by the approximate diameter of the thread (that of the 30-nm fiber actually varies from ~25 to 30 nm).

The **10-nm fiber** is essentially a continuous string of nucleosomes. Sometimes, indeed, it runs continuously into a more stretched-out region in which nucleosomes are seen as a string of beads, as indicated in the example of **FIGURE 10.45**. The 10-nm fiber structure is obtained under conditions of low ionic strength and does not require the presence of linker histone. This means it is a function strictly of the nucleosomes themselves. It may be visualized essentially as a continuous series of nucleosomes, as in **FIGURE 10.46**.

When chromatin is visualized in conditions of greater ionic strength, the **30-nm fiber** is obtained. An example is given in **FIGURE 10.47**. The fiber can be seen to have an underlying coiled structure. It has less than six nucleosomes for every turn, which corresponds to a packing ratio of 40 (i.e., each micron along the axis of the fiber contains 40 µm of DNA). The formation of this fiber from the 10-nm fiber requires the histone tails, which are involved in internucleosomal contacts, and is facilitated

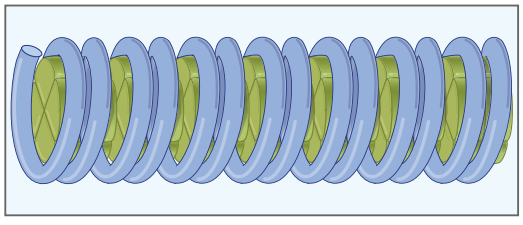

FIGURE 10.46 The 10-nm fiber is a continuous string of nucleosomes.

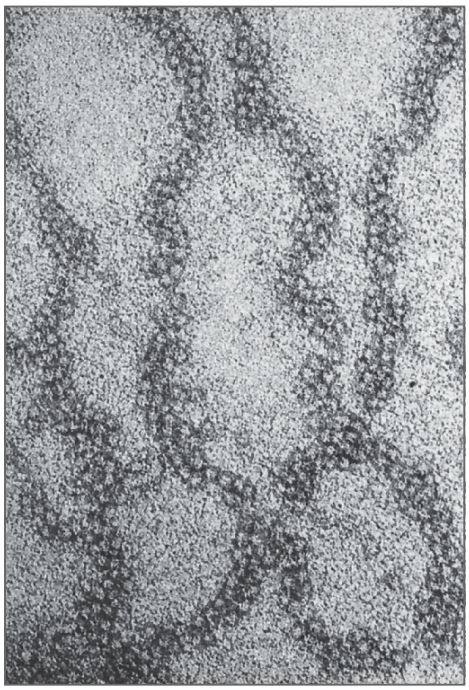

FIGURE 10.47 The 30-nm fiber has a coiled structure. (Photo courtesy of Barbara A. Hamkalo, University of California, Irvine.)

by high ionic strength and the presence of a linker histone such as H1. This fiber is thought to be the basic constituent of both interphase chromatin and mitotic chromosomes, although it is challenging to visualize this directly in vivo.

The most likely arrangement for packing nucleosomes into the fiber is a solenoid, in which the nucleosomes turn in a helical array, coiled around a central cavity. The two main forms of a solenoid are a single-start, which forms from a single linear array, and a two-start, which in effect consists of a double row of nucleosomes, with linker DNA crisscrossing the cavity inside the solenoid. **FIGURE 10.48** shows a two-start model suggested by cross-linking and x-ray crystallography data identifying a double stack of nucleosomes in the 30-nm fiber.

Although the presence of linker histone may be necessary for the formation of the 30-nm fiber in vivo, information about its location is conflicting. Its relative ease of extraction from chromatin seems to argue that it is present on the outside of the superhelical fiber axis. But diffraction data, and the fact that it is harder to find in 30-nm fibers than in 10-nm fibers that retain it, would argue for an interior location.

Concept and Reasoning Check

1. Why might histone tails be important for formation of the 30-nm fiber?

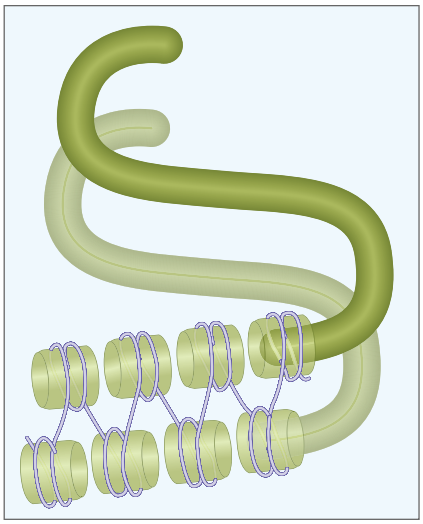

FIGURE 10.48 The 30-nm fiber is a helical ribbon consisting of two parallel rows of nucleosomes coiled into a solenoid.

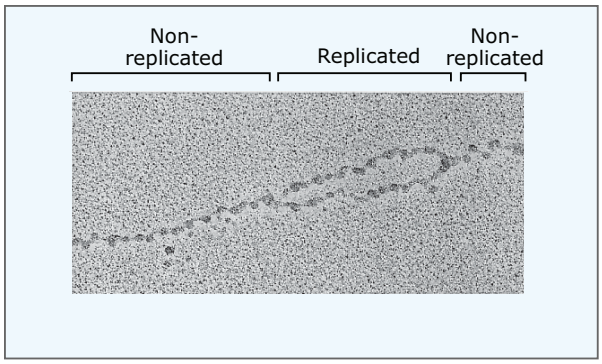

FIGURE 10.49 Replicated DNA is immediately incorporated into nucleosomes. (Photo courtesy of Steven L. McKnight, UT Southwestern Medical Center.)

10.20 Replication of chromatin requires assembly of nucleosomes

Key concepts

- Histone octamers are not conserved during replication, but H2A • H2B dimers and $H3_2$ • $H4_2$ tetramers are conserved.
- There are different pathways for the assembly of nucleosomes during replication and independently of replication.
- Accessory proteins are required to assist the assembly of nucleosomes.
- CAF-1 and ASF1 are histone assembly proteins that are linked to the replication machinery.
- The HIRA assembly protein and the histone H3.3 variant are used for replication-independent assembly.

Replication separates the strands of DNA and, therefore, must inevitably disrupt the structure of the nucleosome. The structure of the replication fork is distinctive. It is more resistant to micrococcal nuclease and is digested into bands that differ in size from nucleosomal DNA. The region that shows this altered structure is confined to the immediate vicinity of the replication fork. This suggests that a large protein complex is engaged in replicating the DNA, but the nucleosomes reform more or less immediately behind as it moves along. This point is illustrated by the electron micrograph of **FIGURE 10.49**, which shows a recently replicated stretch of DNA, already packaged into nucleosomes on both daughter duplex segments.

Both biochemical analysis and visualization of the replication fork therefore suggest that the disruption of nucleosome structure is limited to a short region immediately around the fork. Progress of the fork disrupts nucleosomes, but they form very rapidly on the daughter duplexes as the fork moves forward. In fact, the assembly of nucleosomes is directly linked to the replisome that is replicating DNA.

How do histones associate with DNA to generate nucleosomes? Do the histones *preform* a protein octamer around which the DNA is subsequently wrapped? Or does the histone octamer assemble on DNA from free histones? **FIGURE 10.50** shows that two pathways can be used in vitro to assemble nucleosomes, depending on the conditions used. In one pathway, a preformed octamer binds to DNA. In the other pathway, a tetramer of $H3_2$ • $H4_2$ binds first, and then two H2A • H2B dimers are added. The latter pathway is used in replication in vivo.

Accessory proteins assist histones to associate with DNA, acting as "molecular chaperones" that bind to the histones to release either individual histones or complexes ($H3_2$ • $H4_2$ or H2A • H2B) to the DNA in a controlled manner. This could be necessary because the histones, as basic proteins, have a general high affinity for DNA. Such interactions allow histones to form nucleosomes without becoming trapped in other kinetic intermediates (i.e., other complexes resulting from indiscrete binding of histones to DNA).

A number of histone chaperones have been identified. Chromatin assembly factor

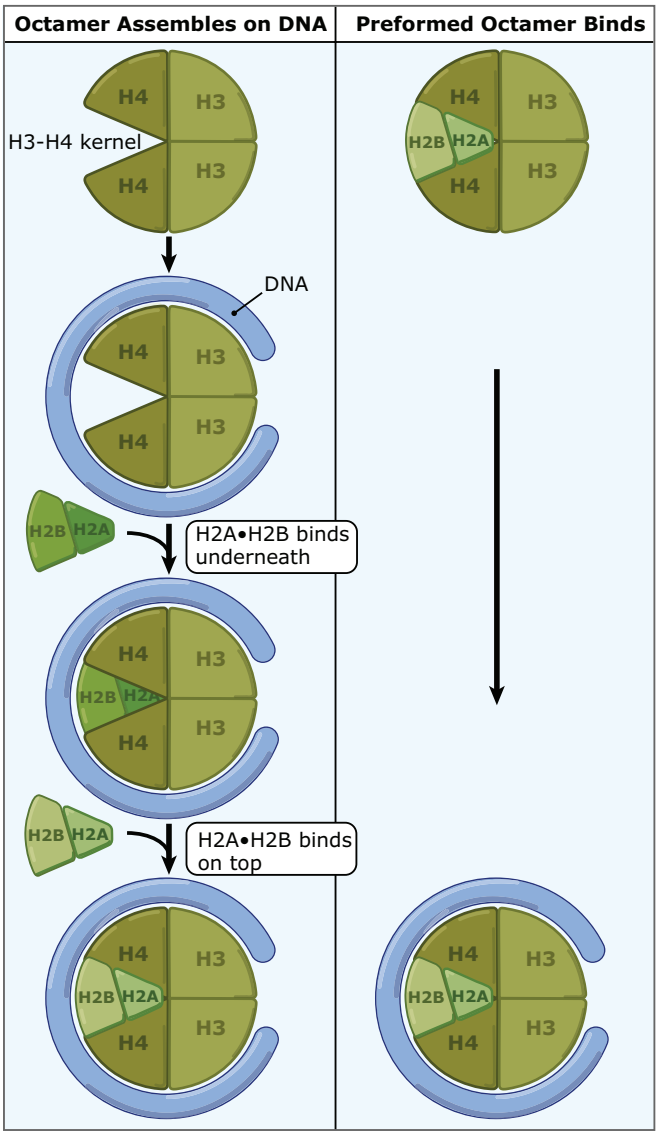

FIGURE 10.50 In vitro, DNA can either interact directly with an intact (cross-linked) histone octamer or can assemble with the $H3_2 \cdot H4_2$ tetramer, after which two H2A • H2B dimers are added.

(CAF)-1 and antisilencing function 1 (ASF1) are two chaperones that function at the replication fork. CAF-1 is a conserved three-subunit complex that is directly recruited to the replication fork by proliferating cell nuclear antigen, the processivity factor for DNA polymerase. ASF1 interacts with the replicative helicase that unwinds the replication fork. Furthermore, CAF-1 and ASF1 interact with each other, and ASF1 stimulates histone deposition by CAF-1. These interactions provide the link between replication and nucleosome assembly, ensuring that nucleosomes are assembled as soon as DNA has been replicated.

CAF-1 acts stoichiometrically and functions by binding to newly synthesized H3 and H4. This suggests that new nucleosomes form by assembling first the $H3_2 \cdot H4_2$ tetramer and then adding the H2A • H2B dimers. When chromatin is replicated, however, a stretch of DNA *already associated with nucleosomes* is replicated, giving rise to two daughter duplexes, and the preexisting nucleosomes are displaced during this process.

The working model for the pattern of disassembly and reassembly is illustrated in **FIGURE 10.51**. The replication fork displaces histone octamers, which then dissociate into $H3_2 \cdot H4_2$ tetramers and H2A • H2B dimers. These "old" tetramers and dimers enter a pool that also includes "new" tetramers and dimers, assembled from newly synthesized histones. Nucleosomes assemble ~600 bp behind the replication fork. Assembly is initiated when $H3_2 \cdot H4_2$ tetramers bind to each of the daughter duplexes, assisted by CAF-1 or ASF1. Some "old" tetramers may be transferred directly to the newly synthesized regions with the assistance of chaperones; ASF1 appears to play an important role in this transfer of parental nucleosomes from ahead of the replication fork to the newly synthesized region behind the fork, although ASF1 can bind and assemble newly synthesized histones as well. Two H2A • H2B dimers then bind to each $H3_2 \cdot H4_2$ tetramer to complete the histone octamer. The assembly of tetramers and dimers is random with respect to "old" and "new" subunits, resulting in a mixture of old and new histones in the replicated octamers. It appears that nucleosomes are disrupted and reassembled in a similar way during transcription, using a different set of chaperones (see *10.23 Histone octamers are displaced and reassembled during transcription*).

During S phase (the period of DNA replication) in a eukaryotic cell, the duplication of chromatin requires synthesis of sufficient histone proteins to package an entire genome—basically, the same quantity of histones must be synthesized that are already contained in nucleosomes. The synthesis of histone mRNAs is controlled as part of the cell cycle and increases enormously in S phase. The pathway for assembling chromatin from this equal mix of old and new histones during S phase is called the replication-coupled (RC) pathway.

Another pathway, called the replication-independent (RI) pathway, exists for assembling nucleosomes during other phases of the cell cycle, when DNA is not being synthesized. This may become necessary as the result of damage to DNA or because nucleosomes are displaced during transcription. The assembly process must necessarily have some

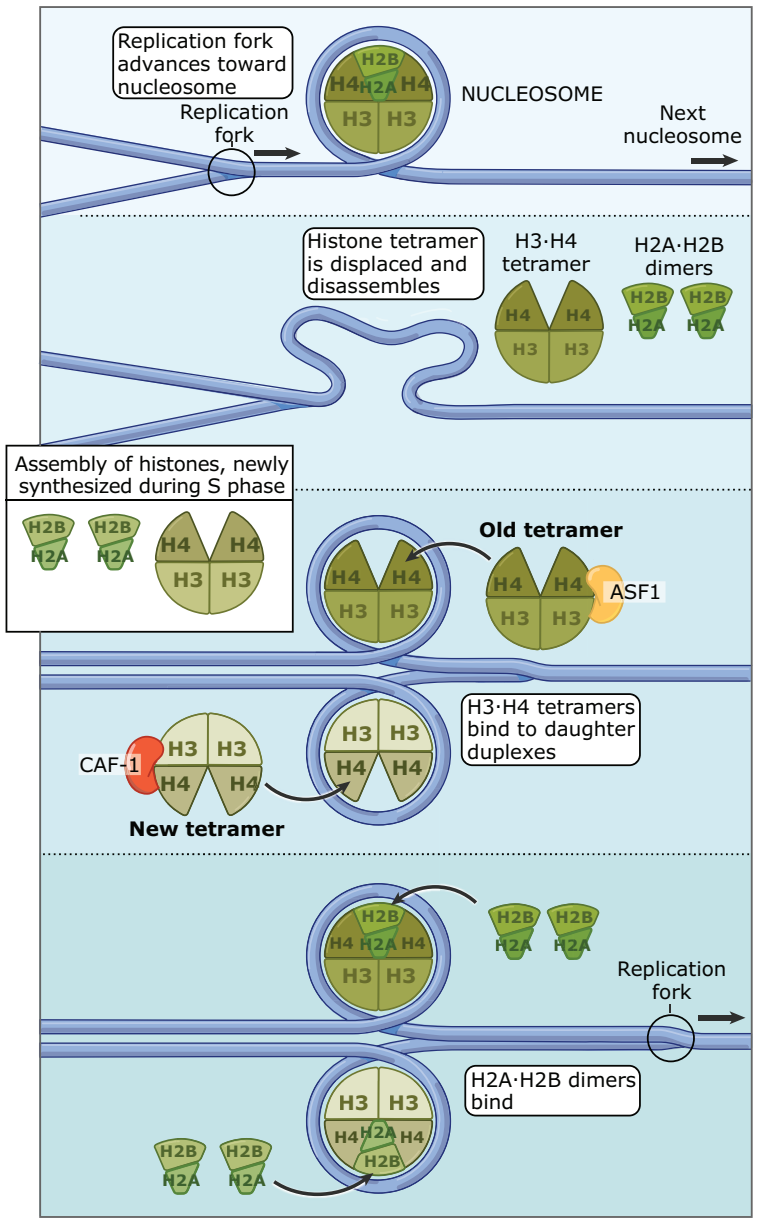

FIGURE 10.51 Replication fork passage displaces histone octamers from DNA. They disassemble into H3 • H4 tetramers and H2A • H2B dimers. Newly synthesized histones are assembled into H3 • H4 tetramers and H2A • H2B dimers. The old and new tetramers and dimers are assembled with the aid of CAF-1 and ASF1 at random into new nucleosomes immediately behind the replication fork. H2A-H2B chaperones have not been identified.

differences from the RC pathway, because it cannot be linked to the replication apparatus. One of the most interesting features of the RI pathway is that it uses different variants of some of the histones from those used during replication.

The histone H3.3 variant differs from the highly conserved H3 histone at four amino acid positions (see Figure 10.44). H3.3 slowly replaces H3 in differentiating cells that do not have replication cycles. This happens as the result of assembly of new histone octamers to replace those that have been displaced from DNA for whatever reason. The mechanism that is used to ensure the use of H3.3 in the RI pathway is different in two cases that have been investigated.

In the protozoan *Tetrahymena,* histone usage is determined exclusively by availability. Histone H3 is synthesized only during the cell cycle; the variant replacement histone is synthesized only in nonreplicating cells. In *Drosophila,* however, an active pathway ensures the usage of H3.3 by the RI pathway. New nucleosomes containing H3.3 assemble at sites of transcription, presumably replacing nucleosomes that were displaced by RNA polymerase. The assembly process discriminates between H3 and H3.3 on the basis of their sequences, specifically excluding H3 from being utilized. By contrast, RC assembly uses both types of H3 (although H3.3 is available at much lower levels than H3 and therefore enters only a small proportion of nucleosomes).

CAF-1 is not involved in RI assembly. (In organisms such as yeast and *Arabidopsis* the gene encoding CAF-1 is not essential, implying that alternative assembly processes may be used in RC assembly.) Instead, a protein called HIRA is used in RI assembly. Depletion of HIRA from in vitro systems for nucleosome assembly inhibits the formation of nucleosomes on nonreplicated DNA but not on replicating DNA, indicating that the pathways do indeed use different assembly mechanisms. Like CAF-1 and ASF1, HIRA functions as a chaperone to assist the incorporation of histones into nucleosomes. This pathway appears to be generally responsible for RI assembly; for example, HIRA is required for the decondensation of the sperm nucleus, when protamines are replaced by histones, leading to the generation of chromatin that is competent to be replicated after fertilization.

RI assembly is also used for assembly of centromeric nucleosomes that incorporate the CenH3 variant (discussed in *10.18 Histone variants produce alternative nucleosomes*). Centromeric DNA replicates early during the replication phase of the cell cycle (in contrast with the surrounding heterochromatic sequences that replicate later). The incorporation of H3 at the centromeres is inhibited, and instead the CenH3 variant is preferentially used (CENP-A in higher eukaryotic cells, Cid in *Drosophila*, Cse4 in yeast). This occurs by the RI assembly pathway, apparently because the RC pathway is inhibited for a brief period of time while centromeric DNA replicates.

Concept and Reasoning Check

1. Why is an RI pathway of chromatin assembly important?

10.21 Do nucleosomes lie at specific positions?

Key concepts

- Nucleosomes may form at specific positions as the result either of the local structure of DNA or of proteins that interact with specific sequences.
- A common cause of nucleosome positioning is when proteins binding to DNA establish a boundary.
- Positioning may affect which regions of DNA are in the linker and which face of DNA is exposed on the nucleosome surface.
- DNA sequence determinants (exclusion or preferential binding) may be responsible for many of the observed nucleosome positions in vivo.

Does a particular DNA sequence always lie in a certain position in vivo with regard to the topography of the nucleosome? Or, are nucleosomes arranged randomly on DNA so that a particular sequence may occur at any location, for example, in the core region in one copy of the genome and in the linker region in another?

To investigate these questions, it is necessary to use a defined sequence of DNA; more precisely, we need to determine the position relative to the nucleosome of a defined point in the DNA. **FIGURE 10.52** illustrates the principle of a procedure used to achieve this.

Suppose the DNA sequence is organized into nucleosomes in only one particular configuration, so that each site on the DNA always is located at a particular position on the nucleosome. This type of organization is called **nucleosome**

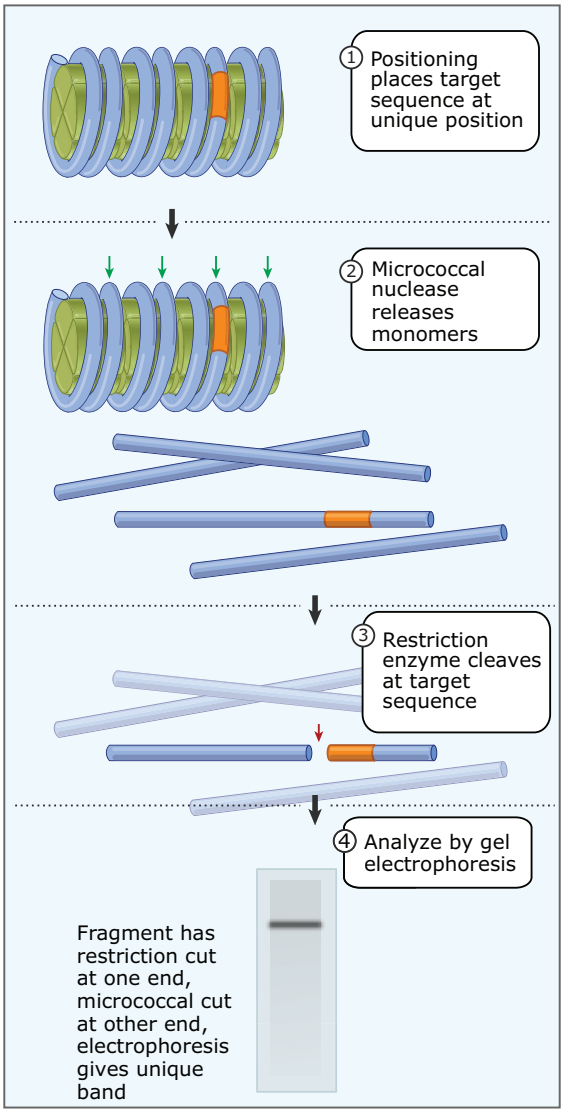

FIGURE 10.52 Nucleosome positioning places restriction sites at unique positions relative to the linker sites cleaved by micrococcal nuclease.

positioning (or sometimes nucleosome phasing). In a series of positioned nucleosomes, the linker regions of DNA comprise unique sites.

Consider the consequences for just a single nucleosome. Cleavage with micrococcal nuclease generates a monomeric fragment that constitutes a *specific sequence*. If the DNA is isolated and cleaved with a restriction enzyme that has only one target site in this fragment, it should be cut at a unique point. This produces two fragments, each of unique size.

The products of the micrococcal/restriction double digest are separated by gel electrophoresis. A probe representing the sequence on one side of the restriction site is used to identify the corresponding fragment in the double digest. This technique is called **indirect end labeling**.

Reversing the argument, the identification of a single sharp band demonstrates that the position of the restriction site is uniquely defined with respect to the end of the nucleosomal DNA (as defined by the micrococcal nuclease cut). So the nucleosome has a unique sequence of DNA.

What happens if the nucleosomes do *not* lie at a single position? Now the linkers consist of *different* DNA sequences in each copy of the genome. Therefore, the restriction site lies at a different position each time; in fact, it lies at all possible locations relative to the ends of the monomeric nucleosomal DNA. FIGURE 10.53 shows the double cleavage then generates a broad smear, ranging from the smallest detectable fragment (~20 bases) to the length of the monomeric DNA.

Nucleosome positioning might be accomplished in either of the following two ways:

1. *Intrinsic mechanisms:* Nucleosomes are deposited specifically at particular DNA sequences. This modifies our view of the nucleosome as a complex able to form equally well at any DNA sequence.
2. *Extrinsic mechanisms:* The first nucleosome in a region is preferentially assembled at a particular site due to action of other protein(s). A preferential starting point for nucleosome positioning can result from the exclusion of a nucleosome from a particular region due to binding of another factor or from the specific deposition of a nucleosome at a given site. The excluded region or the positioned nucleosome provides a boundary that restricts the positions available to the adjacent nucleosome. Then, a series of nucleosomes may be assembled sequentially, with a defined repeat length.

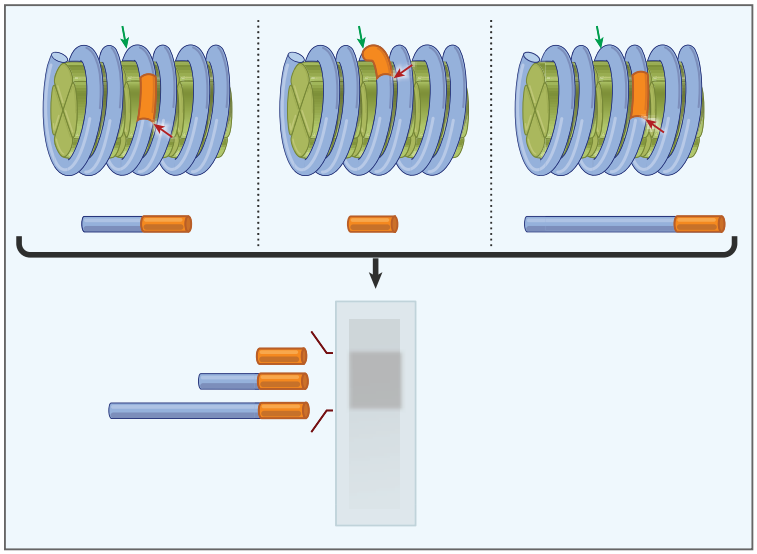

FIGURE 10.53 In the absence of nucleosome positioning, a restriction site lies at all possible locations in different copies of the genome. Fragments of all possible sizes are produced when a restriction enzyme cuts at a target site (red) and micrococcal nuclease cuts at the junctions between nucleosomes (green).

It is now clear that the deposition of histone octamers on DNA is not random with regard to sequence. The pattern is intrinsic in cases in which it is determined by structural features in DNA. It is extrinsic in other cases, resulting from the interactions of other proteins with the DNA and/or histones.

Certain structural features of DNA affect placement of histone octamers. DNA has intrinsic tendencies to bend in one direction rather than another; thus, A · T–rich regions orient so the minor groove faces in toward the octamer, whereas G · C–rich regions are arranged so the minor groove points out. Long runs of dA · dT (>8 bp) avoid positioning in the central superhelical turn of the core. It is not yet possible to sum all relevant structural effects and, thus, entirely to predict the location of a particular DNA sequence with regard to the nucleosome. Sequences that cause DNA to take up more extreme structures have effects such as the exclusion of nucleosomes and thus cause boundary effects or nucleosome-free regions. Specific sequences, such as a portion of the 5S rDNA in some species, and certain simple sequence satellites can robustly position nucleosomes. Recent genome-wide studies have begun to reveal patterns of intrinsic positioning, including the prevalence of nucleosome-excluding sequences in critical promoter regions.

Positioning of nucleosomes near boundaries is common. If there is some variability in the construction of nucleosomes—for example,

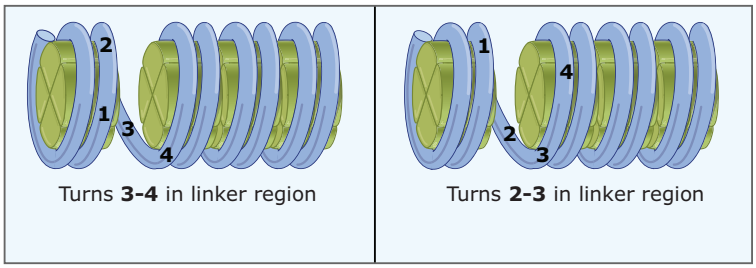

FIGURE 10.54 Translational positioning describes the linear position of DNA relative to the histone octamer. Displacement of the DNA by 10 bp changes the sequences that are in the more exposed linker regions but does not alter which face of DNA is protected by the histone surface and which is exposed to the exterior.

if the length of the linker can vary by, say, 10 bp—the specificity of location would decline proceeding away from the first, defined nucleosome at the boundary. In this case, we might expect the positioning to be maintained rigorously only relatively near the boundary.

The location of DNA on nucleosomes can be described in two ways: translational positioning and rotational positioning. **FIGURE 10.54** shows that **translational positioning** describes the position of DNA with regard to the boundaries of the nucleosome. In particular, it determines which sequences are found in the linker regions. Shifting the DNA by 10 bp brings the next turn into a linker region. Therefore, translational positioning determines which regions are more accessible (at least as judged by sensitivity to micrococcal nuclease).

Because DNA lies on the outside of the histone octamer, one face of any particular sequence is obscured by the histones, but the other face is accessible. Depending on its positioning with regard to the nucleosome, a site in DNA that must be recognized by a regulator protein could be inaccessible or available. The exact position of the histone octamer with respect to DNA sequence may, therefore, be important. **FIGURE 10.55** shows the effect of **rotational positioning** of the double helix with regard to the octamer surface. If the DNA is moved by a partial number of turns (imagine the DNA as rotating relative to the protein surface), there is a change in the exposure of sequence to the outside.

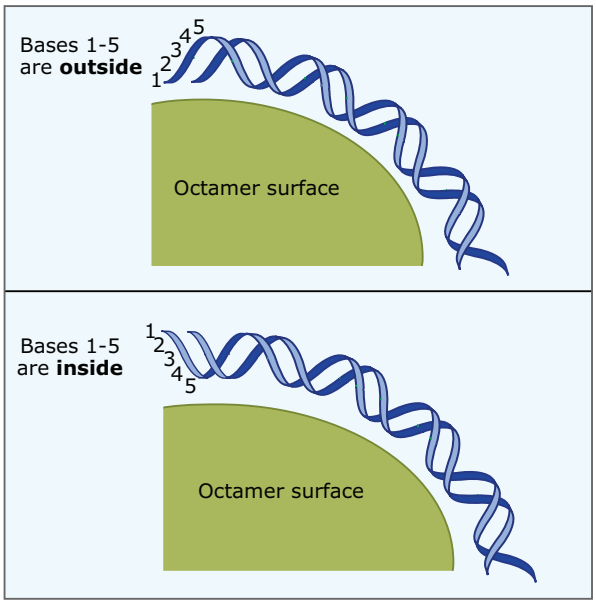

FIGURE 10.55 Rotational positioning describes the exposure of DNA on the surface of the nucleosome. Any movement that differs from the helical repeat (~10.2 bp per turn) displaces DNA with reference to the histone surface. Nucleotides on the inside are more protected against nucleases than nucleotides on the outside.

Both translational and rotational positioning can be important in controlling access to DNA. The best-characterized cases of positioning involve the specific placement of nucleosomes at promoters. Translational positioning and/or the exclusion of nucleosomes from a particular sequence may be necessary to allow a transcription complex to form. Some regulatory factors can bind to DNA only if a nucleosome is excluded to make the DNA freely accessible, and this creates a boundary for translational positioning. In other cases, regulatory factors can bind to DNA on the surface of the nucleosome, but rotational positioning is important to ensure the face of DNA with the appropriate contact points is exposed.

Gene promoters (and some other structures) often have short regions that exclude nucleosomes. These regions typically form a boundary next to which nucleosome positions are restricted. A survey of an extensive region in the *S. cerevisiae* genome (mapping 2,278 nucleosomes over 482 kb of DNA) showed that in fact 60% of the nucleosomes have specific positions as the result of boundary effects, most often from promoters. Nucleosome positioning is a complex output of intrinsic and extrinsic positioning mechanisms; thus, it has been difficult to predict nucleosome positioning based on sequence alone, although there have been some successes.

Concept and Reasoning Check

1. How can a boundary effect be created by either extrinsic or intrinsic mechanisms?

10.22 Domains of nuclease sensitivity define regions that contain active genes

Key concepts

- DNase hypersensitive sites are found at the promoters of expressed genes.
- Hypersensitive sites are generated by the binding of factors that exclude histone octamers.
- A domain containing a transcribed gene is defined by increased sensitivity to degradation by DNase I.

Numerous changes occur to chromatin in active or potentially active regions. These include distinctive structural changes at specific sites associated with initiation of transcription or with certain structural features in DNA. These changes were first detected by the effects of digestion with very low concentrations of the enzyme DNase I.

When chromatin is digested with DNase I, the first effect is the introduction of breaks in the duplex at specific **hypersensitive sites**. Because susceptibility to DNase I reflects the availability of DNA in chromatin, we take these sites to represent chromatin regions in which the DNA is particularly exposed because it is not organized in the usual nucleosomal structure. A typical hypersensitive site is 100 times more sensitive to enzyme attack than bulk chromatin. These sites are also hypersensitive to other nucleases and to chemical agents.

The locations of hypersensitive sites can be determined by the technique of indirect end-labeling that we introduced earlier in the context of nucleosome positioning. This application of the technique is recapitulated in FIGURE 10.56. In this case, cleavage at the hypersensitive site by DNase I is used to generate one end of the fragment, and its distance is measured from the other end that is generated by cleavage with a restriction enzyme.

Hypersensitive sites are created by the local structure of chromatin, which may be tissue specific. Most hypersensitive sites are related to gene expression. Every active gene has a site, or sometimes more than one site, in the region of the promoter. Most hypersensitive sites are found only in chromatin of cells in which the associated gene is being expressed; they do not occur when the gene is inactive. Hypersensitive site(s) typically appear in the 5′ region before transcription begins and the DNA sequences contained within the hypersensitive sites are required for gene expression.

A particularly well-characterized nuclease-sensitive region lies on the simian virus (SV)40 "minichromosome," the DNA genome of the

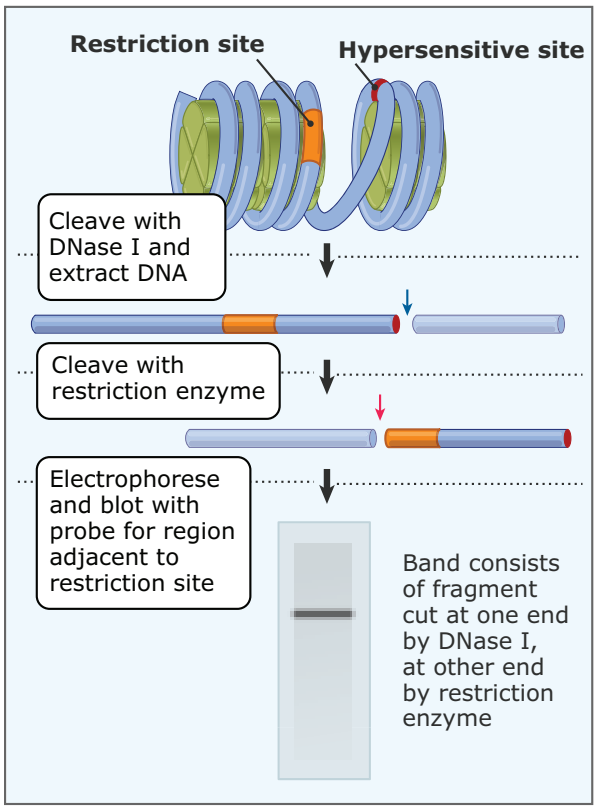

FIGURE 10.56 Indirect end-labeling identifies the distance of a DNase hypersensitive site from a restriction cleavage site. The existence of a particular cutting site for DNase I generates a discrete fragment, whose size indicates the distance of the DNase I hypersensitive site from the restriction site.

SV40 virus. A short segment near the origin of replication, just upstream of the promoter for a particular transcription unit, is cleaved preferentially by DNase I, micrococcal nuclease, and other nucleases (including restriction enzymes).

The state of the SV40 minichromosome can be visualized by electron microscopy. In up to 20% of the samples, a "gap" is visible in the nucleosomal organization, as evident in FIGURE 10.57. The gap is a region of ~120 nm

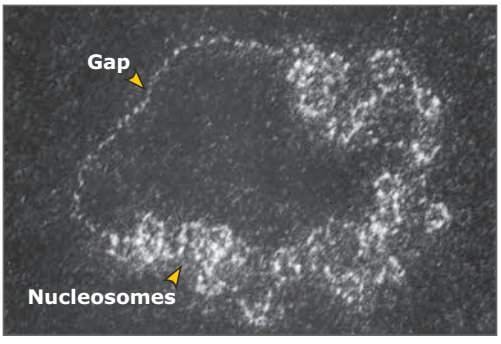

FIGURE 10.57 The SV40 minichromosome has a nucleosome gap. (Photo courtesy of Moshe Yaniv, Pasteur Institute.)

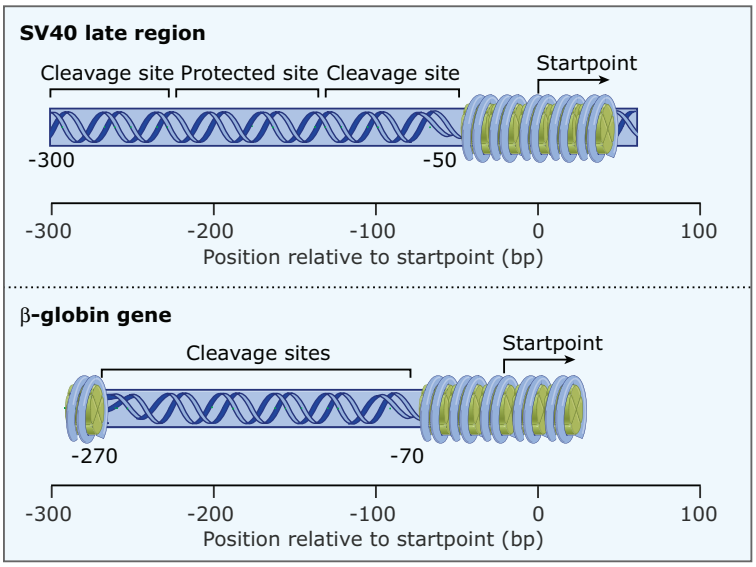

FIGURE 10.58 The SV40 gap includes hypersensitive sites, sensitive regions, and a protected region of DNA. The hypersensitive site of a chicken β-globin gene comprises a region that is susceptible to several nucleases.

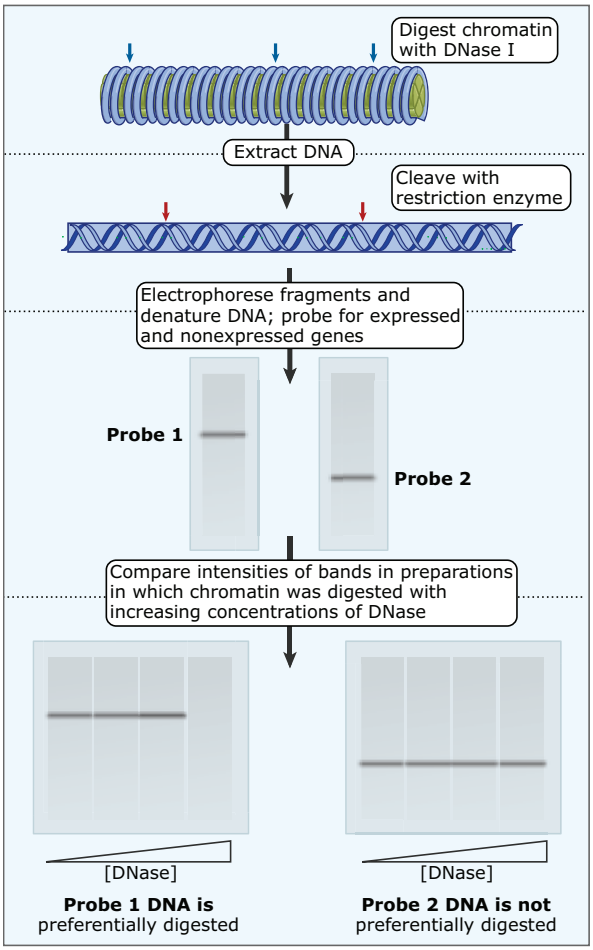

FIGURE 10.59 Sensitivity to DNase I can be measured by determining the rate of disappearance of the material hybridizing with a particular probe.

in length (about 350 bp), surrounded on either side by nucleosomes. The visible gap corresponds with the nuclease-sensitive region. This shows directly that increased sensitivity to nucleases is associated with the exclusion of nucleosomes.

A hypersensitive site is not necessarily uniformly sensitive to nucleases. **FIGURE 10.58** shows the maps of two hypersensitive sites. Within the SV40 gap of ~300 bp, there are two DNase I hypersensitive sites and a "protected" region. The protected region presumably reflects the association of (nonhistone) protein(s) with the DNA. The gap is associated with the DNA sequence elements that are necessary for promoter function.

A hypersensitive site at the β-globin promoter is preferentially digested by several enzymes, including DNase I, DNase II, and micrococcal nuclease. A region extending from about –70 to –270 bp is preferentially accessible to nucleases when the gene is poised for transcription.

What is the structure of the hypersensitive site? Its preferential accessibility to nucleases indicates that it is not protected by histone octamers, but this does not necessarily imply it is free of protein. A region of free DNA might be vulnerable to damage, and in any case, how would it be able to exclude nucleosomes? We assume the hypersensitive site results from the binding of specific regulatory proteins that exclude nucleosomes. Indeed, the binding of such proteins is probably the basis for the existence of the protected region within the hypersensitive site.

The proteins that generate hypersensitive sites are likely to be regulatory factors of various types, because hypersensitive sites are found to be associated with promoters and other elements that regulate transcription, origins of replication, centromeres, and sites with other structural significance. In some cases, they are associated with more extensive organization of chromatin structure. A hypersensitive site may provide a boundary for a series of positioned nucleosomes. Hypersensitive sites associated with transcription may be generated by transcription factors when they bind to the promoter as part of the process that makes it accessible to RNA polymerase.

In addition to detecting hypersensitive sites, DNase I digestion can also be used to assess the relative accessibility of a genomic region. DNase I sensitivity defines a chromosomal **domain**, a region of altered structure including at least one active transcription unit and sometimes extending farther. (Note that use of the term *domain* does not imply any

necessary connection with the structural domains identified by the loops of chromatin or chromosomes.)

When chromatin is digested with DNase I, it is eventually degraded into very small fragments of DNA, however, different regions of the genome are degraded at different rates. Specifically, active genes are preferentially degraded. The fate of individual genes can be followed by quantitating the amount of DNA that survives to react with a specific probe. The protocol is outlined in **FIGURE 10.59**. The principle is that the loss of a particular band indicates the corresponding region of DNA has been degraded by the enzyme.

FIGURE 10.60 shows what happens to β-globin genes and an ovalbumin gene in chromatin extracted from chicken red blood cells (in which globin genes are expressed and the ovalbumin gene is inactive). The restriction fragments representing the β-globin genes are rapidly lost, whereas those representing the ovalbumin gene show little degradation. (The ovalbumin gene in fact is digested at the same rate as the bulk of DNA.) Therefore, the bulk of chromatin is relatively resistant to DNase I and contains nonexpressed genes (as well as other sequences). *A gene becomes relatively susceptible to degradation specifically in the tissue(s) in which it is expressed.*

Is preferential susceptibility a characteristic only of rather actively expressed genes, such as globin, or of all active genes? Experiments using probes representing the entire cellular mRNA population suggest that all active genes, whether coding for abundant or rare mRNAs, are preferentially susceptible to DNase I. (There are, however, variations in the degree of susceptibility.) Because rarely expressed genes are likely to have very few RNA polymerase molecules actually engaged in transcription at any moment, this implies that sensitivity to DNase I does not result from the act of transcription but is a feature of *genes that are able to be transcribed.*

What is the extent of the preferentially sensitive region? This can be determined by using a series of probes representing the flanking regions as well as the transcription unit itself. The sensitive region always extends over the entire transcribed region; an additional region of several kilobases on either side may show an intermediate level of sensitivity (probably as the result of spreading effects).

The critical concept implicit in the description of the domain is that a region of high sensitivity to DNase I extends over a considerable distance. Often we think of regulation as residing in events that occur at a discrete site in DNA—for example, in the ability to initiate transcription at the promoter. Even if this is true, such regulation must determine, or must be accompanied by, a more wide-ranging change in structure.

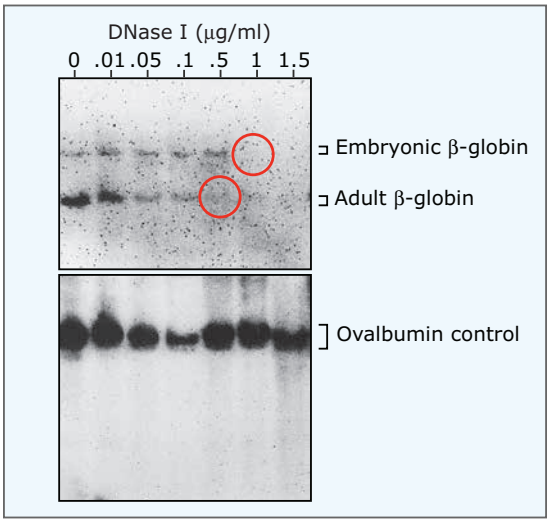

FIGURE 10.60 In adult erythroid cells, the adult β-globin gene is highly sensitive to DNase I digestion, the embryonic β-globin gene is partially sensitive (probably due to spreading effects), but ovalbumin is not sensitive. (Photos courtesy of Harold Weintraub and Mark Groudine, Fred Hutchinson Cancer Research Center.)

Concept and Reasoning Check

1. Describe the difference between a DNase I–hypersensitive site and a DNase I–sensitive domain.

10.23 Histone octamers are displaced and reassembled during transcription

Key concepts

- Most transcribed genes retain a nucleosomal structure.
- Some heavily transcribed genes appear to be exceptional cases devoid of nucleosomes.
- RNA polymerase displaces histone octamers during transcription in a model system, but octamers reassociate with DNA as soon as the polymerase has passed.
- Nucleosomes are reorganized when transcription passes through a gene.
- Ancillary factors are required both for RNA polymerase to displace octamers during transcription and for the histones to reassemble into nucleosomes after transcription.

Transcription involves the unwinding of DNA and may require the fiber to unfold in restricted regions of chromatin. A simplistic view suggests that some "elbow room" must be needed for the process.

Heavily transcribed chromatin adopts structures too extended to still be contained in nucleosomes. For example, in the intensively transcribed genes coding for ribosomal RNA (rRNA), shown in **FIGURE 10.61**, the extreme packing of RNA polymerases makes it hard to see the DNA. We cannot directly measure the lengths of the rRNA transcripts because the RNA is compacted by proteins, but we know (from the sequence of the rRNA) how long the transcript must be. The length of the transcribed DNA segment, measured by the length of the axis of the "Christmas tree," is ~85% of the length of the rRNA. This means the DNA is almost completely extended.

On the other hand, transcription complexes of SV40 minichromosomes can be extracted from infected cells. They contain the usual complement of histones and display a classic beads-on-a-string structure of the 10-nm fiber. Chains of RNA can be seen to extend from the minichromosome, as in the example of **FIGURE 10.62**. This argues that transcription can proceed while the SV40 DNA is organized into nucleosomes. Of course, the SV40 minichromosome is transcribed less intensively than the rRNA genes.

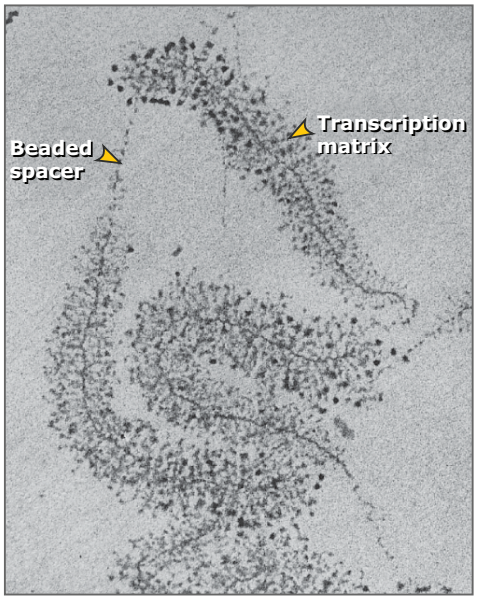

FIGURE 10.61 The extended axis of an rDNA transcription unit alternates with the only slightly less extended nontranscribed spacer. (Photo courtesy of Yean Chooi and Charles Laird.)

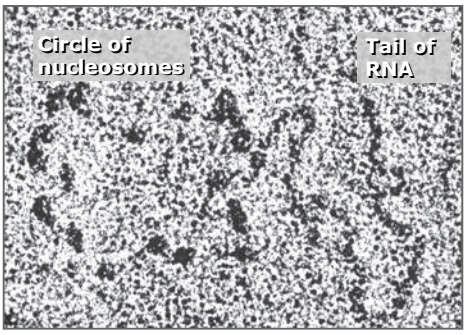

FIGURE 10.62 An SV40 minichromosome can be transcribed. (Reprinted from *J. Mol. Biol.*, vol. 131, P. Gariglio, et al., The template of the isolated native simian virus 40... , pp. 75–105, Copyright (1979) with permission from Elsevier [http://www.sciencedirect.com/science/journal/00222836]. Photo courtesy of Pierre Chambon, College of France.)

Experiments to test whether an RNA polymerase can transcribe directly through a nucleosome suggest that the histone octamer is displaced by the act of transcription. **FIGURE 10.63** shows what happens when the phage T7 RNA polymerase transcribes a short piece of DNA containing a single octamer core in vitro. The core remains associated with the DNA but is found in a different location. The core is most likely to rebind to the same DNA molecule from which it was displaced.

FIGURE 10.64 shows a model for polymerase progression based on these studies. DNA is displaced as the polymerase enters the nucleosome, but the polymerase reaches a point at which the DNA loops back and reattaches, forming a closed region. As polymerase advances further, unwinding the DNA, it creates positive supercoils in this loop; the effect could be dramatic, because the closed loop is only ~80 bp, so each base pair through which the polymerase advances makes a significant addition to the supercoiling. In fact, the polymerase progresses easily for the first 30 bp into the nucleosome. It then proceeds more slowly, as though encountering increasing difficulty in progressing. Pauses arise every 10 bp, suggesting that the structure of the loop imposes a constraint related to rotation around each turn of DNA. When the polymerase reaches the midpoint of the nucleosome (the next bases to be added are essentially at the axis of dyad symmetry), pausing ceases, and the polymerase advances rapidly. This suggests that the midpoint of the nucleosome marks the point at which the octamer is displaced

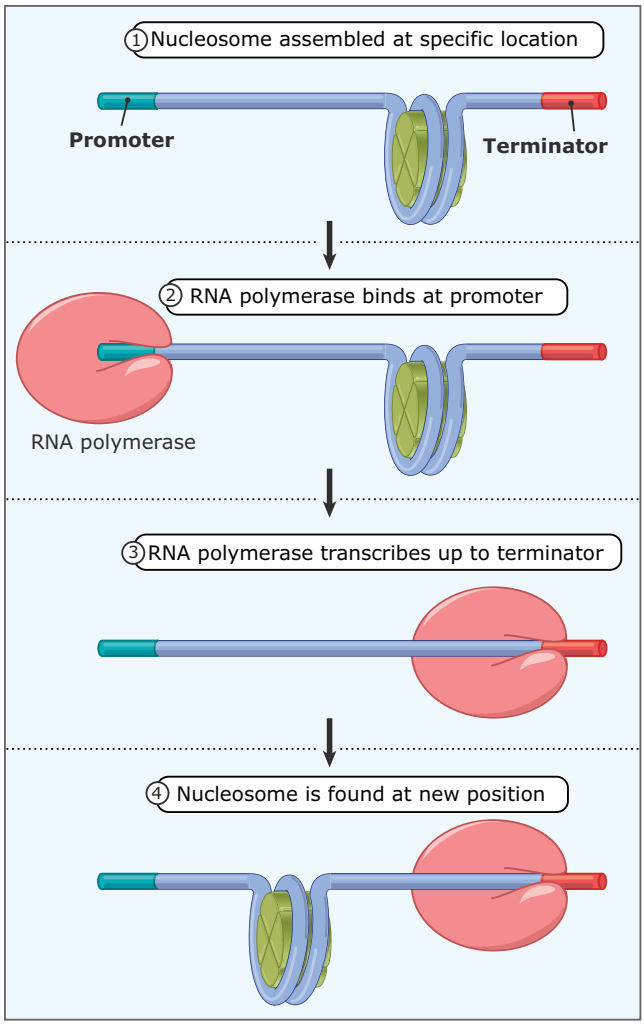

FIGURE 10.63 A protocol to test the effect of transcription on nucleosomes shows that the histone octamer is displaced from DNA and rebinds at a new position.

(possibly, because positive supercoiling has reached some critical level that expels the octamer from DNA). This releases tension ahead of the polymerase and allows it to proceed. The octamer then binds to the DNA behind the polymerase and no longer presents an obstacle to progress. Possibly, the octamer changes position without ever completely losing contact with the DNA.

These studies show that a small RNA polymerase can displace a single nucleosome, which reforms behind it, during transcription. Of course, the situation is more complex in a eukaryotic nucleus. RNA polymerase is very much larger, and the impediment to progress is a string of connected nucleosomes, which can also be folded into higher order structures. Overcoming this obstacle requires additional factors that act on chromatin.

The organization of nucleosomes can be dramatically altered by transcription. **FIGURE 10.65** shows what happens to the yeast *URA3* gene when it is transcribed under control of an inducible promoter. Positioning is examined by using micrococcal nuclease to examine cleavage sites relative to a restriction site at the 5′ end of the gene. Initially, the gene displays a pattern of nucleosomes that are organized from the promoter for a significant distance across the gene; positioning is lost in the 3′ regions. When the gene is expressed, a general smear replaces the positioned pattern of nucleosomes. Therefore, nucleosomes are present at the same density but are no longer organized in phase. This suggests that transcription destroys the nucleosomal positioning. When repression is reestablished, positioning appears within 10 minutes (although it is not complete).

The unifying model supposes that RNA polymerase displaces histone octamers (either as a whole or as dimers and tetramers) as it progresses. If the DNA behind the polymerase is available, the nucleosome is reassembled

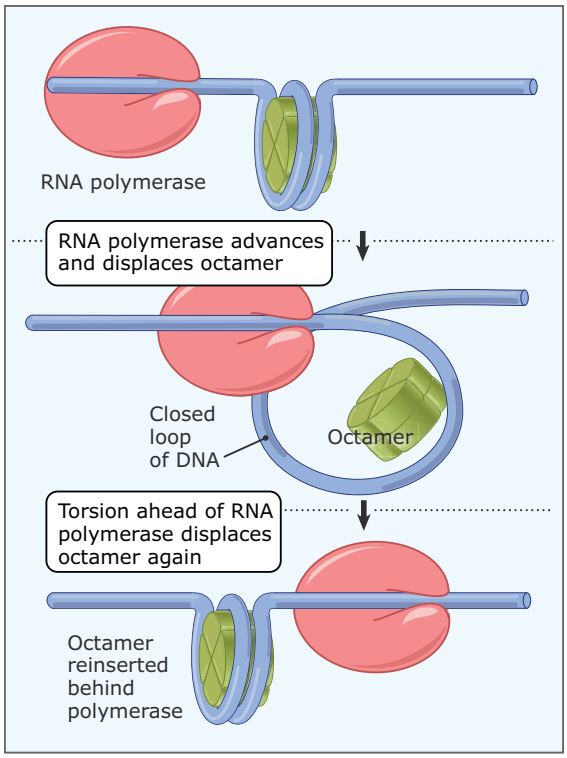

FIGURE 10.64 RNA polymerase displaces DNA from the histone octamer as it advances. The DNA loops back and attaches (to polymerase or to the octamer) to form a closed loop. As the polymerase proceeds, it generates positive supercoiling ahead. This displaces the octamer, which keeps contact with DNA and/or polymerase, and is inserted behind the RNA polymerase.

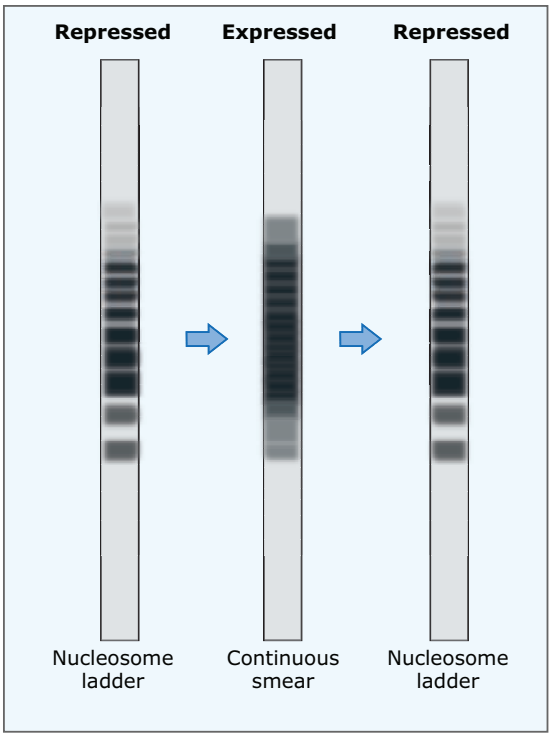

FIGURE 10.65 The *URA3* gene has positioned nucleosomes before transcription. When transcription is induced, nucleosome positions are randomized. When transcription is repressed, the nucleosomes resume their particular positions.

there. If the DNA is not available, for example, because another polymerase continues immediately behind the first, then the octamer may be permanently displaced, and the DNA may persist in an extended form.

The displacement and reassembly of nucleosomes does not occur solely as a result of the passage of RNA polymerase but is facilitated by factors that help regulate this process. The first of these to be characterized is a heterodimeric factor called FACT that behaves like a transcription elongation factor. FACT (*fa*cilitates *c*hromatin *t*ranscription) is not part of RNA polymerase but associates with it specifically during the elongation phase of transcription. FACT consists of two subunits that are well conserved in all eukaryotes. It is associated with the chromatin of active genes.

When FACT is added to isolated nucleosomes, it causes them to lose H2A · H2B dimers. During transcription in vitro, it converts nucleosomes to "hexasomes" that have lost H2A · H2B dimers. This suggests that FACT is part of a mechanism for displacing octamers during transcription. FACT may also be involved in the reassembly of nucleosomes after transcription, because it assists formation of nucleosomes from core histones, acting like a histone chaperone. This suggests the model shown in **FIGURE 10.66**, in which FACT (or a similar factor) detaches H2A · H2B from a nucleosome in front of RNA polymerase and then helps to add it to a nucleosome that is reassembling behind the enzyme. Other factors must be required to complete the process.

Several other factors have been identified that play key roles in either nucleosome displacement or reassembly during transcription. These include the Spt6 protein, a factor involved in "resetting" chromatin structure after transcription. Spt6, like FACT, co-localizes with actively transcribed regions and can act as a histone chaperone to promote nucleosome assembly.

Concept and Reasoning Check

1. Why might it be important to reassemble nucleosomes behind an elongating RNA polymerase?

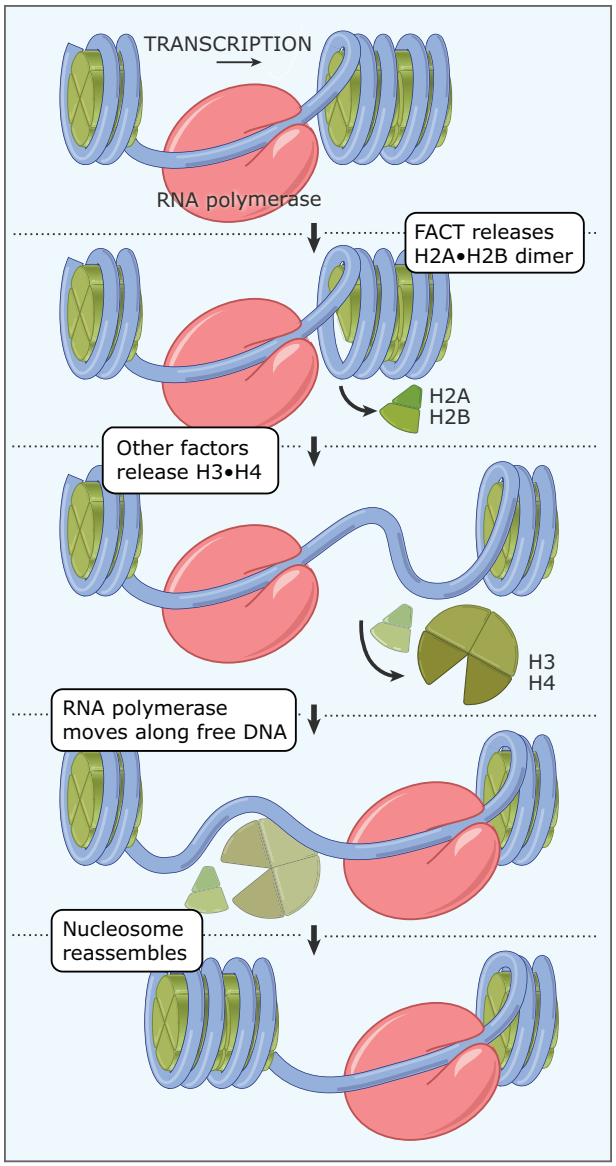

FIGURE 10.66 Histone octamers are disassembled ahead of transcription to remove nucleosomes. They re-form after transcription. Release of H2A • H2B dimers probably initiates the disassembly process.

10.24 Chromatin remodeling is an active process

Key concepts

- Chromatin structure is changed by remodeling complexes that use energy provided by hydrolysis of ATP.
- All remodeling complexes contain a related ATPase subunit and are grouped into subfamilies containing more closely related ATPases.
- Most remodeling complexes do not have specificity for any particular target site but must be recruited by a component of the transcription apparatus.
- Remodeling complexes can alter, slide, or displace nucleosomes or exchange histones within nucleosomes.

Transcriptional activators face a challenge when trying to bind to their recognition sites in chromatin. In a sense, histones function as generalized repressors of transcription. Activation of a gene requires changes in the state of chromatin: The essential issue is how the transcription apparatus gains access to the promoter DNA.

Whether a gene is expressed depends on the structure of chromatin both locally (at the promoter) and in the surrounding domain. Chromatin structure correspondingly can be regulated by individual activation events or by changes that affect a wide chromosomal region. The most localized events concern an individual target gene, where changes in nucleosomal structure and organization occur in the immediate vicinity of the promoter. More general changes may affect regions as large as a whole chromosome.

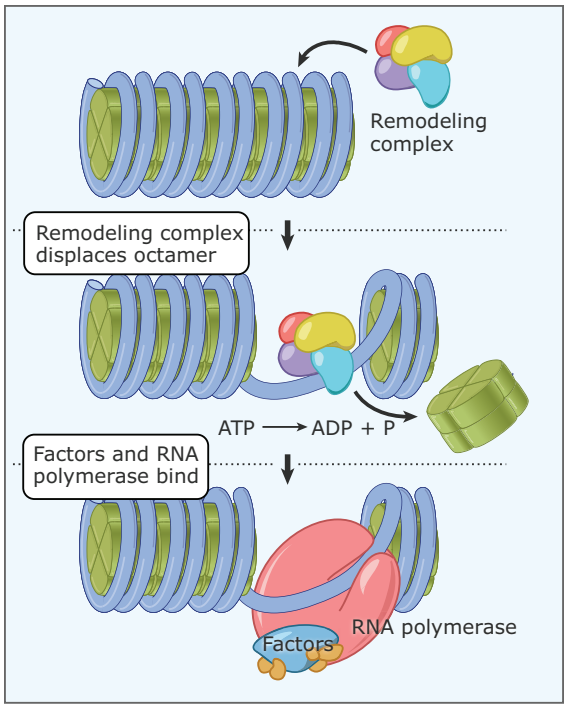

FIGURE 10.67 The dynamic model for transcription of chromatin relies on factors that can use energy provided by hydrolysis of ATP to displace nucleosomes from specific DNA sequences.

Changes that affect large regions control the potential of a gene to be expressed. The term **silencing** is used to refer to repression of gene activity in a local chromosomal region. Silenced regions are typically assembled into heterochromatin, which results from additional proteins binding to chromatin and either directly or indirectly preventing transcription factors and RNA polymerase from activating promoters in the region.

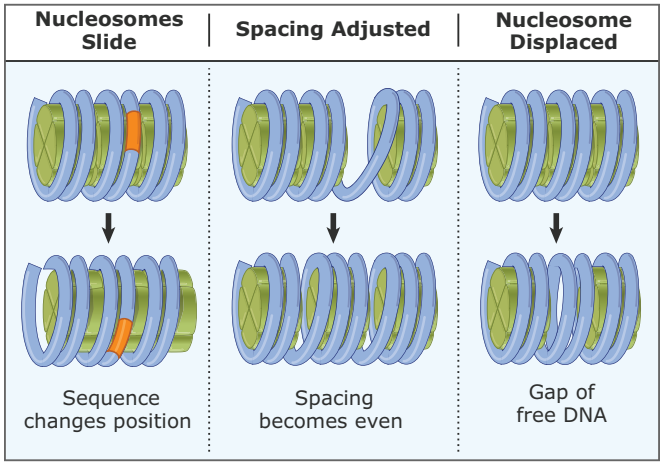

FIGURE 10.68 Remodeling complexes can cause nucleosomes to slide along DNA, can displace nucleosomes from DNA, or can reorganize the spacing between nucleosomes.

Changes at an individual promoter control whether transcription is initiated for a particular gene. These changes may be either activating or repressing. There is an intimate and continuing connection between initiation of transcription and chromatin structure. Some activators of gene transcription directly modify histones; in particular, acetylation of histones is associated with gene activation (see *10.25 Histones are covalently modified*). Conversely, some repressors of transcription function by deacetylating histones. Therefore, a reversible change in histone structure in the vicinity of the promoter is involved in the control of gene expression. This is one mechanism by which a gene is maintained in an active or inactive state.

The general process of inducing changes in chromatin structure is called **chromatin remodeling**. This consists of mechanisms for moving or displacing histones that depend on the input of energy. Many protein–protein and protein–DNA contacts need to be disrupted to release histones from chromatin. There is no free ride: The energy must be provided to disrupt these contacts. **FIGURE 10.67** illustrates the principle of dynamic remodeling by a factor that hydrolyzes ATP. When the histone octamer is released from DNA, other proteins (in this case transcription factors and RNA polymerase) can bind.

FIGURE 10.68 summarizes several alternative outcomes of chromatin remodeling:

- Histone octamers may *slide* along DNA, changing the relationship between the nucleic acid and protein. This can alter both the rotational and translational position of a particular sequence on the nucleosomal surface.
- The *spacing* between histone octamers may be changed, again with the result that the positions of individual sequences are altered relative to the histone octamer.
- The most extensive change is that an octamer(s) may be *displaced* entirely from DNA to generate a nucleosome-free gap. Alternatively, one or both H2A–H2B dimers can be displaced, leaving an $H3_2$–$H4_2$ tetramer on the DNA.

A major role of chromatin remodeling is to change the organization of nucleosomes at the promoter of a gene that is to be transcribed. This is required to allow the transcription apparatus to gain access to the promoter. Remodeling, however, is also required to enable other manipulations of chromatin, including replication or repair of damaged DNA.

Remodeling often takes the form of sliding or displacing one or more histone octamers. This can result in the creation of a site that is hypersensitive to cleavage with DNase I (see *10.22 Domains of nuclease sensitivity define regions that contain active genes*). Sometimes there are less dramatic changes, such as a change in rotational positioning of a single nucleosome; this may be detected by loss or alteration of the DNase I 10-base ladder. Therefore, changes in chromatin structure may extend from altering the positions of nucleosomes to removing them altogether.

Chromatin remodeling is undertaken by **ATP-dependent chromatin remodeling complexes** that use ATP hydrolysis to provide the energy for remodeling. The heart of the remodeling complex is its ATPase subunit. The ATPase subunits of all remodeling complexes are related members of a large superfamily of proteins, which is divided into subfamilies of more closely related members. Remodeling complexes are classified according to the subfamily of ATPase that they contain as their catalytic subunit. There are many subfamilies, but the four largest subfamilies (SWI/SNF, ISWI, CHD, and INO80/SWR1) are shown in **FIGURE 10.69**. The chromatin-remodeling superfamily is large and diverse, and most species have multiple complexes in different subfamilies. For example, at least eight different ISWI complexes have been identified thus far in mammals. Remodeling complexes range from small heterodimeric complexes (the ATPase subunit plus a single partner) to massive complexes of 10 or more subunits. Each type of complex may undertake a different range of remodeling activities.

SWI/SNF was the first remodeling complex to be identified. Its name reflects the fact that many of its subunits are encoded by genes originally identified by *swi* or *snf* mutations in *S. cerevisiae*. These mutations also show genetic interactions with mutations in genes that code for components of chromatin, in particular *SIN1*, which codes for a nonhistone protein, and *SIN2*, which codes for histone H3. The *SWI* and *SNF* genes are required for expression of a variety of individual loci (~120 or 2% of *S. cerevisiae* genes are affected). Expression of these loci may require the SWI/SNF complex to remodel chromatin at their promoters.

SWI/SNF acts catalytically in vitro, and there are only ~150 complexes per yeast cell. All genes encoding the SWI/SNF subunits are nonessential, which implies that yeast must also have other ways of remodeling chromatin. The RSC (remodel the structure of chromatin) complex is more abundant and is essential for viability. It acts at ~700 target loci.

Type of Complex	SWI/SNF	ISW1	IN080	CHD
Yeast	SWI/SNF RSC	ISW1 ISW2	IN080 SWRI	CHD1
Fly	dSWI/SNF (Brahma)	NURF CHRAC ACF		CHD1 CHD3,4
Human	hSWI/SNF	RSF hACF/WCFR hCHRAC WICH NURF NORC	IN080	NURD CHD1-4
Frog		CHRAC ACF WICH		NURD

FIGURE 10.69 Remodeling complexes can be classified by their ATPase subunits.

Different subfamilies of remodeling complexes have distinct modes of remodeling, reflecting differences in their ATPase subunits as well as effects of other proteins in individual remodeling complexes. SWI/SNF complexes can remodel chromatin in vitro without overall loss of histones or can displace histone octamers. Both types of reaction may pass through the same intermediate in which the structure of the target nucleosome is altered, leading either to reformation of a (remodeled) nucleosome on the original DNA or to displacement of the histone octamer to a different DNA molecule. The SWI/SNF complex alters nucleosomal sensitivity to DNase I at the target site and induces changes in protein–DNA contacts that persist after it has been released from the nucleosomes. The Swi2 (also known as Snf2) subunit is the ATPase that provides the energy for remodeling by SWI/SNF. In contrast, the ISWI family primarily affects nucleosome positioning via sliding, without displacement of octamers.

There are many contacts between DNA and a histone octamer—14 are identified in the crystal structure. All of these contacts must be broken for an octamer to be released or for it to move to a new position. How is this achieved? The ATPase subunits are distantly related to helicases (enzymes that unwind double-stranded nucleic acids), but remodeling complexes do not have any unwinding activity. Present thinking is that remodeling complexes in the SWI/SNF and ISWI classes use the hydrolysis of ATP to *twist* DNA on the nucleosomal surface. This twisting creates a mechanical force that allows a small region of DNA to be released from the surface and then repositioned. This mechanism creates transient loops of DNA on the surface of the octamer; these loops are themselves accessible to interact with

other factors, or they can propagate along the nucleosome, ultimately resulting in nucleosome sliding. In the case of SWI/SNF complexes, this activity can also result in nucleosome disassembly, first by displacement of the H2A/H2B dimers and then of the H3/H4 tetramer.

How are remodeling complexes targeted to specific sites on chromatin? They do not themselves contain subunits that bind specific DNA sequences. This suggests the model shown in **FIGURE 10.70** in which they are recruited by activators or (sometimes) by repressors. This is accomplished by a "hit and run" mechanism, in which the activator or repressor may be released after the remodeling complex has bound.

Different remodeling complexes have different roles in the cell. SWI/SNF complexes are generally involved in transcriptional activation, whereas some ISWI complexes act as repressors, using their remodeling activity to slide nucleosomes *onto* promoter regions to prevent transcription. Members of the CHD (chromodomain helicase DNA binding) family have also been implicated in repression, particularly the Mi-2/NuRD complexes, which contain both chromatin remodeling and histone deacetylase activities (see *10.25 Histones are covalently modified*). Remodelers in the SWR1/INO80 class have a unique activity: In addition to their normal remodeling capabilities, some members of this class also have *histone exchange* capability, in which individual histones (usually H2A/H2B dimers) can be replaced in a nucleosome, typically with a histone variant (described in *10.18 Histone variants produce alternative nucleosomes*).

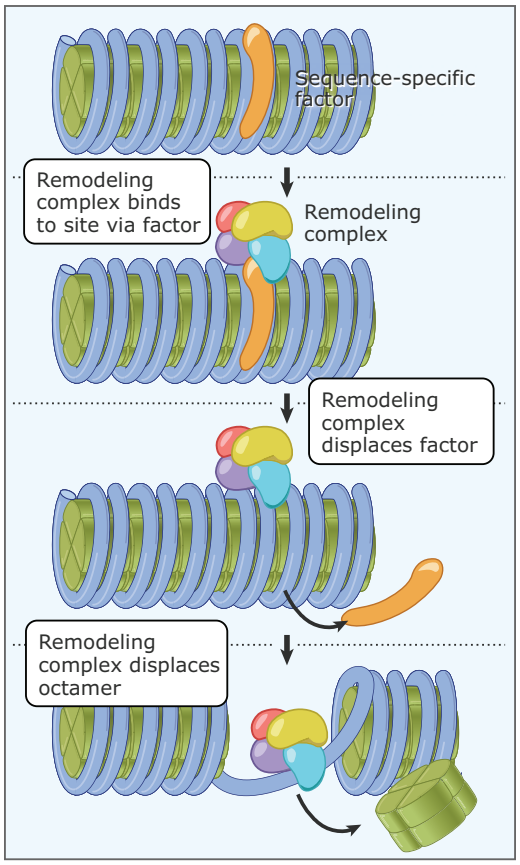

FIGURE 10.70 A remodeling complex binds to chromatin via an activator (or repressor).

Concept and Reasoning Check

1. How can remodeling complexes act as either activators or repressors?

10.25 Histones are covalently modified

Key concepts

- Histones are modified by acetylation, methylation, phosphorylation, and other modifications.
- Combinations of specific histone modifications define the function of local regions of chromatin; this is known as the histone code.
- Histone acetylation occurs transiently at replication.
- Histone acetylation is associated with activation of gene expression.
- Deacetylated chromatin may have a more condensed structure.
- Transcription activators are associated with histone acetyltransferase (HAT) activities in large complexes.
- HATs vary in their target specificity.
- Deacetylation is associated with repression of gene activity.

In addition to the ATP-dependent chromatin remodeling discussed in the section above, chromatin structure is also modulated by making covalent changes on the nucleosome, in particular by modifying the N- and C-terminal tails of the histones. The histone tails consist of 15–30 amino acids at the N-termini of all four core histones and the C-termini of H2A and H2B. The N-terminal tails of H2B and H3 pass between the turns of DNA (see Figure 10.43). They can be modified at several sites, by methylation, acetylation, phosphorylation, or a number of other modifications. Lysines in the histone tails are the most common targets of modification. Acetylation, methylation, ubiquitylation, and sumoylation all occur on the free epsilon (ε) amino group of lysine. Acetylation neutralizes the positive charge that resides on the NH_3 of the ε-amino group. In contrast, lysine methylation retains the positive charge, and lysine can be mono-, di-, or trimethylated. Although different histone modifications

have known roles in replication, chromatin assembly, transcription, splicing, and DNA repair, functions of a number of specific modifications are yet to be characterized. Histone modification can result in structural changes in chromatin (primarily due to changes in charge) as well as the creation of binding sites for the attachment of nonhistone proteins that then change the properties of chromatin.

The range of nucleosomes targeted for modification can vary. Modification can be a local event, for example, restricted to nucleosomes at the promoter. Or it can be a general event, extending, for example, to an entire chromosome. **FIGURE 10.71** shows the general correlation in which acetylation is associated with active chromatin, whereas methylation at specific sites is associated with inactive chromatin. This is not a simple rule, however, and the particular sites that are modified, as well as combinations of specific modifications, are very important. For example, there are several cases in which histones methylated at a certain position are found in active chromatin; this is actually always the case in *S. cerevisiae*, which have no methylation associated with inactive chromatin.

The specificity of the modifications is controlled by the fact that many of the enzymes responsible for these modifications have individual target sites in specific histones.

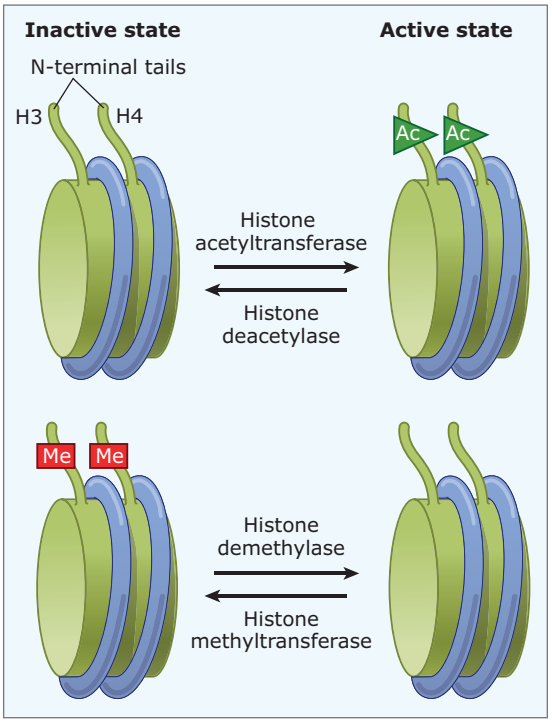

FIGURE 10.71 Acetylation of H3 and H4 is associated with active chromatin, whereas methylation is associated with inactive chromatin.

Histone	Site	Modification	Function
H3	Lys-4	Methylation	Activation
	Lys-9	Methylation	Chromatin condensation; required for DNA methylation
		Acetylation	Activation
	Ser-10	Phosphorylation	Activation
	Lys-14	Acetylation	Prevents methylation at Lys-9, activation
	Lys-79	Methylation	Telomeric silencing, DNA repair
H4	Arg-3	Methylation	
	Lys-5	Acetylation	Assembly
	Lys-12	Acetylation	Assembly
	Lys-16	Acetylation	Fly X activation

FIGURE 10.72 Most modified sites in histones have a single, specific type of modification, but some sites can have more than one type of modification. Individual functions can be associated with some of the modifications.

FIGURE 10.72 summarizes the known effects of just a handful of the many known modifications. Some modified sites are subject to only a single type of modification, whereas other sites can be modified in different ways under different conditions. In some cases, modification of one site may activate or inhibit modification of another site within the same histone or even another histone. The idea that combinations of signals may be used to define chromatin types has sometimes been called the *histone code*.

Of all the many histone modifications, acetylation may be the best studied. All the core histones can be acetylated on lysine residues in the tails (and occasionally within the globular core). Acetylation occurs on free histones when they are first synthesized during S phase and in the context of nucleosomes when genes are activated.

When chromosomes are replicated during S phase, new histones are transiently acetylated in the cytoplasms on specific lysines, such as lysines 5 and 12 (K5 and K12) of histone H4 (see Figure 10.72). **FIGURE 10.73** shows that this acetylation occurs before the histones are incorporated into nucleosomes. We know that histones H3 and H4 are acetylated at the stage

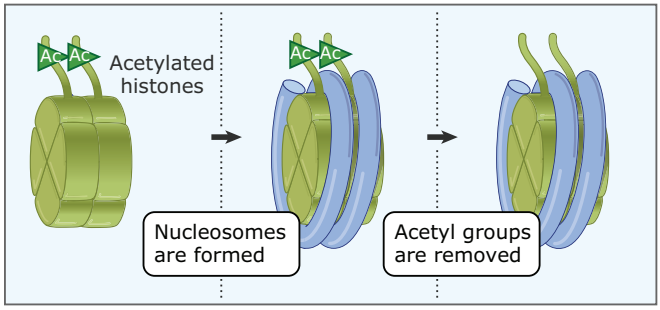

FIGURE 10.73 Acetylation at replication occurs on histones before they are incorporated into nucleosomes.

when they are associated with one another in the $H3_2 \cdot H4_2$ tetramer. The tetramer is then incorporated into nucleosomes. Quite soon after, the acetyl groups are removed.

The importance of the acetylation is indicated by the fact that preventing acetylation of both histones H3 and H4 during replication causes loss of viability in yeast. The two histones are redundant as substrates, because yeast can manage perfectly well as long as they can acetylate either one of these histones during S phase. There are two possible roles for the acetylation: It could be needed for the histones to be recognized by factors that incorporate them into nucleosomes, or it could be required for the assembly and/or structure of the new nucleosome.

The factors that are known to be involved in chromatin assembly do not distinguish between acetylated and nonacetylated histones, suggesting the modification is more likely to be required for subsequent interactions. It has been thought for a long time that acetylation might be needed to help control protein–protein interactions that occur as histones are incorporated into nucleosomes. Some evidence for such a role is that the yeast SAS (something about silencing) histone acetylase complex binds to chromatin assembly complexes at the replication fork, where it acetylates K16 of histone H4. This may be part of the system that establishes the histone acetylation patterns after replication.

Outside of S phase, acetylation of histones in chromatin is generally correlated with the state of gene expression. The correlation was first noticed because histone acetylation is increased in a domain containing active genes, and acetylated chromatin is more sensitive to nucleases. **FIGURE 10.74** shows that this involves the acetylation of histone tails in nucleosomes. We now know this occurs largely because of acetylation of the nucleosomes in the vicinity of the promoter when a gene is activated.

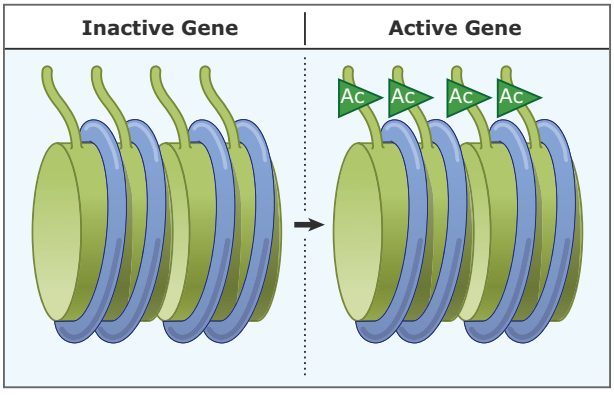

FIGURE 10.74 Acetylation associated with gene activation occurs by directly modifying histones in nucleosomes.

In addition to events at individual promoters, global changes in acetylation occur on sex chromosomes. This is part of the mechanism by which the activities of genes on the X chromosome are altered to compensate for the presence of two X chromosomes in one sex but only one X chromosome (in addition to the Y chromosome) in the other sex (see *10.28 X chromosomes undergo global changes*). The inactive X chromosome in female mammals has underacetylated H4. The superactive X chromosome in *Drosophila* males has increased acetylation of H4. This suggests the presence of acetyl groups may be a prerequisite for a less condensed, active structure. In male *Drosophila,* the X chromosome is acetylated specifically at K16 of histone H4. The responsible enzyme is called MOF, which is recruited to the chromosome as part of a large protein complex. This "dosage compensation" complex is responsible for introducing general changes in the X chromosome that enable it to be more highly expressed. The increased acetylation is only one of its activities.

Acetylation is reversible. Each direction of the reaction is catalyzed by a specific type of enzyme. Enzymes that can acetylate histones are called **histone acetyltransferases** (**HATs**); these can also be more generally referred to as lysine (K) acetyltransferases, because some of these enzymes have nonhistone substrates as well. The acetyl groups are removed by **histone deacetylases** (**HDACs**). There are two groups of HAT enzymes: group A enzymes act on histones in chromatin and are involved with the control of transcription, whereas group B enzymes act on newly synthesized histones in the cytosol and are involved with nucleosome assembly.

Like the chromatin-remodeling enzymes discussed previously, group A HATs are typically members of large complexes that must be targeted to their sites of action in chromatin. **FIGURE 10.75** shows a simplified model for their behavior. Typically, a site-specific binding protein recognizes its DNA target and directly interacts with the HAT complex. This determines the target for the HAT. HAT complexes sometimes also contain other subunits (and other enzymatic activities) that affect chromatin structure or act directly on transcription.

Just as activation of transcription is associated with acetyltransferases, so is inactivation associated with deacetylases. This is true both for individual genes and for heterochromatin. Repression at individual promoters may be accomplished

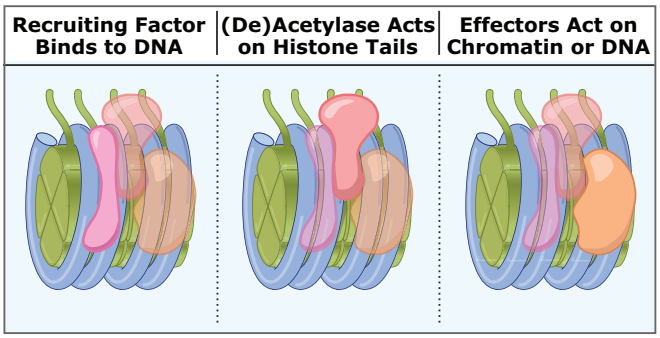

FIGURE 10.75 Complexes that modify chromatin structure or activity are recruited to their sites of action. HAT or HDAC enzymes acetylate or deacetylate histones, and effector subunits may have other actions on chromatin or DNA.

by complexes that have deacetylase activities acting on localized regions in the vicinity of the promoter; these HDACs are typically recruited to their sites of action by sequence-specific repressors, just as HATs are recruited by activators. Absence of histone acetylation in heterochromatin is true of both constitutive heterochromatin (typically involving regions of centromeres or telomeres) and facultative heterochromatin (regions that are inactivated in one cell although they may be active in another). Typically, the N-terminal tails of histones H3 and H4 are not acetylated in heterochromatic regions.

There can also be direct interactions between remodeling complexes and histone-modifying complexes (or their modifications). Binding by the SWI/SNF remodeling complex may lead in turn to binding by an acetylase complex. Acetylation of histones may then in fact stabilize the association with the SWI/SNF complex, leading to a mutual reinforcement of the changes in the components at the promoter.

We can connect some of the events at the promoter into the series summarized in **FIGURE 10.76**. The initiating event is usually binding of a sequence-specific component (which is able to find its target DNA sequence in the context of chromatin). In this example, this activator recruits a remodeling complex. Changes (rotational, sliding, displacement) occur in nucleosome structure. An acetylating complex binds, and the acetylation of target histones provides a covalent mark that the locus has been activated. This is just a simplified example. In many cases, histone modification precedes ATP-dependent remodeling, and histone modifications during transcription also include dynamic ubiquitylation and methylation; only acetylation is presented in this case.

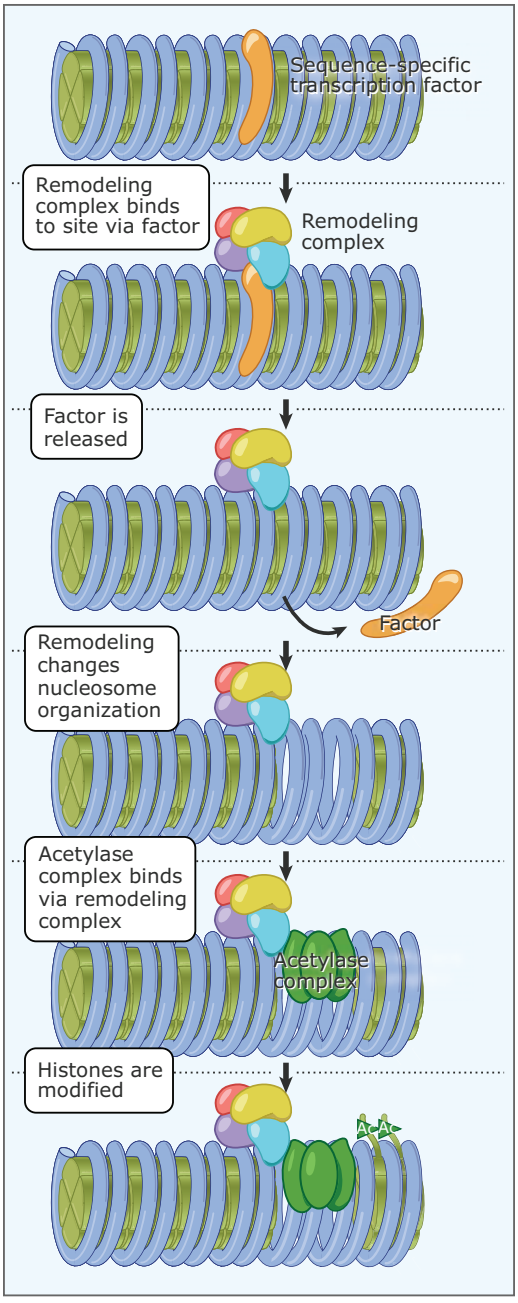

FIGURE 10.76 Promoter activation involves binding of a sequence-specific activator, recruitment and action of a remodeling complex, and recruitment and action of an acetylating complex.

Concept and Reasoning Check

1. What would be the effect on transcription of adding an HDAC inhibitor to a cell?

10.26 Heterochromatin propagates from a nucleation event

Key concepts

- Heterochromatin is nucleated at a specific sequence, and the inactive structure propagates along the chromatin fiber.
- Genes within regions of heterochromatin are inactivated.
- Because the length of the inactive region varies from cell to cell, inactivation of genes in this vicinity causes position effect variegation.
- Similar spreading effects occur at telomeres and at the silent cassettes in yeast mating type.

An interphase nucleus contains both euchromatin and heterochromatin. The condensation state of heterochromatin is close to that of mitotic chromosomes. Heterochromatin remains condensed in interphase, is transcriptionally repressed, replicates late in S phase, and may be localized to the nuclear periphery. Centromeric heterochromatin typically consists of satellite DNAs, however, the formation of heterochromatin is not rigorously defined by sequence. When a gene is transferred, either by a chromosomal translocation or by transfection and integration, into a position adjacent to heterochromatin, it may become inactive as the result of its new location, implying it has become heterochromatic.

Such inactivation is the result of an **epigenetic** effect. Epigenetic inheritance describes the ability of different states, which may have different phenotypic consequences, to be inherited without any change in the sequence of DNA. This means two individuals with the same DNA sequence at the locus that controls the effect may show different phenotypes. The basic cause of this phenomenon is the existence of a self-perpetuating structure in one of the individuals that does not depend on DNA sequence.

Epigenetic states can differ between individual cells in an animal. In the most extreme case, cells of different tissue types exhibit completely different phenotypes due to stable changes in activation and silencing of their identical genomes, mediated by developmentally determined epigenetic changes. Alternatively, cells in the same tissue can exist in distinct epigenetic states, resulting in the phenomenon of **position effect variegation** (**PEV**). This has been well characterized in *Drosophila*. FIGURE 10.77 shows an example of position effect variegation in the fly eye, in which some regions lack color and others are red, because the *white* gene is inactivated by adjacent heterochromatin in some cells, whereas it remains active in other cells.

FIGURE 10.77 Position effect variegation in eye color results when the *white* gene is integrated near heterochromatin. Cells in which *white* is inactive give patches of white eye, whereas cells in which *white* is active give red patches. The severity of the effect is determined by the closeness of the integrated gene to heterochromatin. (Photo courtesy of Steven Henikoff, Fred Hutchinson Cancer Research Center.)

The explanation for this effect is shown in FIGURE 10.78. Inactivation spreads from heterochromatin into the adjacent region for a variable distance. In some cells it goes far enough to inactivate a nearby gene, but in others it does not. This happens at a certain point in embryonic development, and after that point, the state of the gene is inherited by all the progeny cells. Cells descended from an ancestor in which the gene was inactivated form patches corresponding to the phenotype of loss-of-function (in the case of *white*, absence of color).

The closer a gene lies to heterochromatin, the higher the probability that it will be inactivated. This is because the formation of heterochromatin is a two-stage process: A *nucleation* event occurs at a specific sequence or region (triggered by binding of a protein that recognizes the DNA sequence or other features in the region), and then the inactive structure *propagates* along the chromatin fiber. The distance for which the inactive structure extends is not precisely determined and may be stochastic, being influenced by parameters such as the quantities of limiting protein components. One factor that may affect the spreading process is the activation of promoters

in the region; an active promoter may inhibit spreading. Genes closer to heterochromatin are more likely to be inactivated and will therefore be inactive in a greater proportion of cells.

The effect of **telomeric silencing** in yeast is analogous to position effect variegation in *Drosophila;* genes translocated to a telomeric location show the same sort of variable loss of activity. This results from a spreading effect that propagates from the telomeres, described in *10.27 Heterochromatin depends on interactions with histones.*

A second form of silencing occurs in yeast. Yeast mating type is determined by the activity of a single active locus (*MAT*), but the genome contains two other copies of the mating type sequences (*HML* and *HMR*), which are maintained in an inactive form. The silent loci *HML* and *HMR* nucleate heterochromatin via binding of several proteins, which then lead to propagation of heterochromatin similar to that at telomeres. Heterochromatin in yeast exhibits features typical of heterochromatin in other species, such as transcriptional inactivity and self-perpetuating protein structures superimposed on nucleosomes (which are generally deacetylated). The only notable difference between yeast heterochromatin and that of most other species is that histone methylation in budding yeast is not associated with silencing, whereas specific sites of histone methylation are a key feature of heterochromatin formation in most multicellular eukaryotes, as well as in the fission yeast *S. pombe.*

Concept and Reasoning Check

1. A strongly transcribed gene can sometimes prevent heterochromatin from spreading. Why might this be?

10.27 Heterochromatin depends on interactions with histones

Key concepts

- HP1 is the key protein in forming mammalian heterochromatin and acts by binding to methylated histone H3.
- Histone methylation and DNA methylation are linked in heterochromatin.
- Rap1 initiates formation of heterochromatin in yeast by binding to specific target sequences in DNA.
- The targets of Rap1 include telomeric repeats and silencers at *HML* and *HMR*.

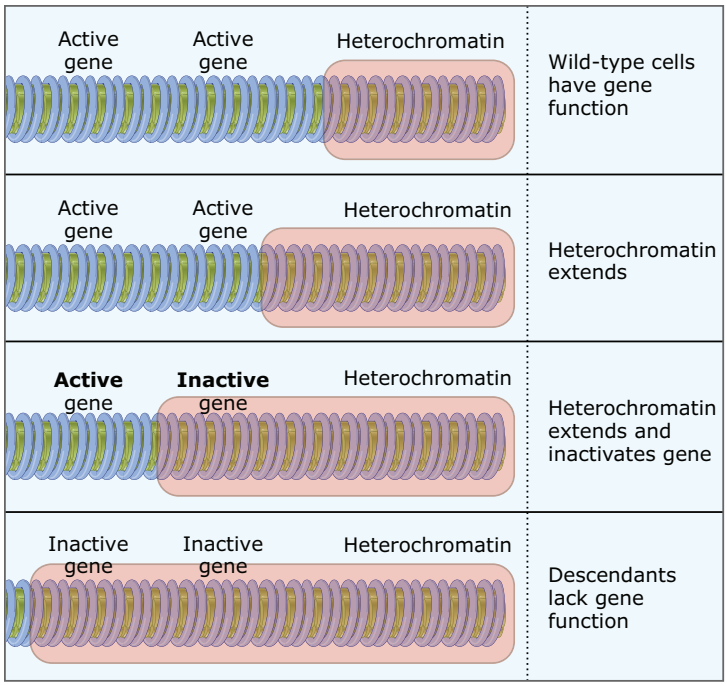

FIGURE 10.78 Extension of heterochromatin inactivates genes. The probability that a gene will be inactivated depends on its distance from the heterochromatic region.

- Rap1 recruits Sir4, which interacts with the N-terminal tails of H3 and H4.
- Sir2 deacetylates the N-terminal tails of H3 and H4 and promotes spreading of Sir3 and Sir4.
- RNA interference (RNAi) pathways promote heterochromatin formation at centromeres.

Inactivation of chromatin occurs by the addition of proteins to the nucleosomal fiber. The inactivation may be due to a variety of effects, including condensation of chromatin to make it inaccessible to the apparatus needed for gene expression, addition of proteins that directly block access to regulatory sites, or the presence of proteins that directly inhibit transcription.

Two systems have been characterized at the molecular level. They involve heterochromatin protein 1 (HP1) in mammals and the silent information regulator (SIR) complex in yeast. Although there are no detailed similarities between the proteins involved in each system, the general mechanism of reaction is similar: the points of contact in chromatin are the N-terminal tails of the histones.

HP1 was originally identified as a protein that is localized to heterochromatin by staining polytene chromosomes with an antibody directed against the protein. The original protein identified as HP1 is now called HP1α, because two related proteins, HP1β and HP1γ, have since been found.

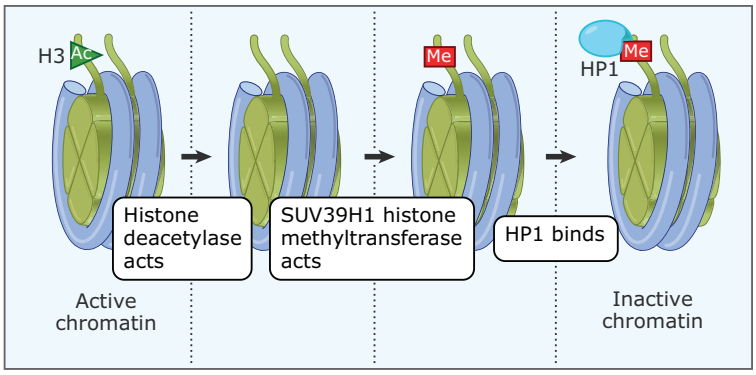

FIGURE 10.79 SUV39H1 is a histone methyltransferase that acts on ^{9}Lys of histone H3. HP1 binds to the methylated histone.

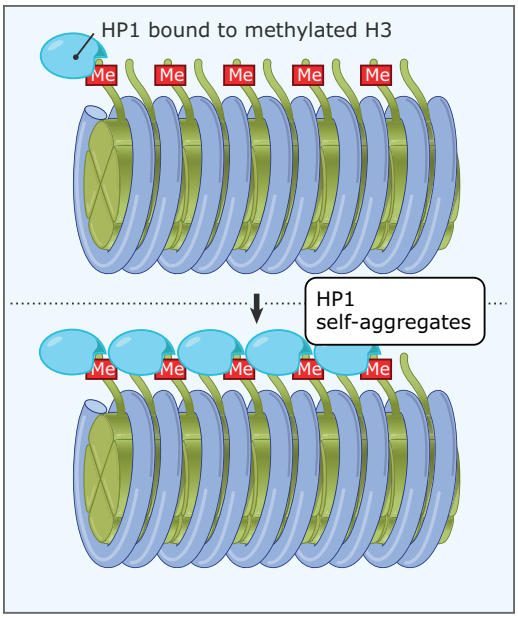

FIGURE 10.80 Binding of HP1 to methylated histone H3 forms a trigger for silencing because further molecules of HP1 aggregate on the nucleosome chain.

HP1 contains a region called the *chromodomain* near the N-terminus and another domain that is related to it, called the *chromoshadow* domain, at the C-terminus (see Figure 10.80). The importance of the chromodomain is indicated by the fact that it is the location of many mutations in HP1. The chromodomain is a common protein motif of 60 amino acids. It is found in proteins involved with either activating or repressing chromatin, suggesting that it represents a motif that participates in protein–protein interactions with targets in chromatin. Chromodomain(s) targets proteins to heterochromatin by recognizing methylated lysines in histone tails.

Mutation of a deacetylase that acts on the H3 N-terminus prevents the methylation at K9. H3 that is trimethylated at K9 binds the protein HP1 via the chromodomain. This suggests the model for initiating formation of heterochromatin shown in **FIGURE 10.79**. First, the deacetylase acts to remove the acetylation in the H3 tail. Then a specific methylase (SUV39H1) acts on K9 of histone H3 to create the methylated signal to which HP1 will bind. This is a trigger for forming inactive chromatin. **FIGURE 10.80** shows that the inactive region may then be extended by the ability of further HP1 molecules to interact with one another.

Modification of DNA is also linked to silencing and heterochromatin formation in higher eukaryotes (and in the fission yeast *S. pombe* but not in *S. cerevisiae*). Methylation of cytosine at CpG doublets is associated with gene inactivity and represents a major epigenetic mark. Methylation of DNA and methylation of histones is connected in a mutually reinforcing circuit. When HP1 binds to methylated K4 of histone H3, it can recruit DNA methyltransferases to modify CpG doublets. In turn, DNA methylation can result in histone methylation. Some histone methyltransferase complexes (as well as some HDAC complexes) contain binding domains that recognize the methylated CpG doublet, so the DNA methylation reinforces the circuit by providing a target for the HDACs and methyltransferases to bind.

Although the yeast *S. cerevisiae* does not exhibit DNA methylation or heterochromatin-specific histone methylation, in other respects yeast heterochromatin is very similar to that of other eukaryotes. Heterochromatin formation at telomeres and silent mating type loci in yeast relies on an overlapping set of genes, known as *silent information regulators* (*SIR* genes). Mutations in *SIR2*, *SIR3*, or *SIR4* cause the two silent loci (*HML* and *HMR*) to become activated and also relieve the inactivation of genes that have been integrated near telomeric heterochromatin. The products of these loci therefore function to maintain the inactive state of both types of heterochromatin.

FIGURE 10.81 proposes a model for actions of these proteins. Only one of them, Rap1, is a sequence-specific DNA binding protein. Rap1 binds to the $C_{1-3}A$ repeats at the telomeres and binds to the *cis*-acting silencer elements needed for repression of *HML* and *HMR*. The proteins Sir3 and Sir4 interact with Rap1 and with one another (they may function as a heteromultimer). Sir3/Sir4 interact with the N-terminal tails of the histones H3 and H4, with a preference for unacetylated tails. Another SIR protein, Sir2, is a

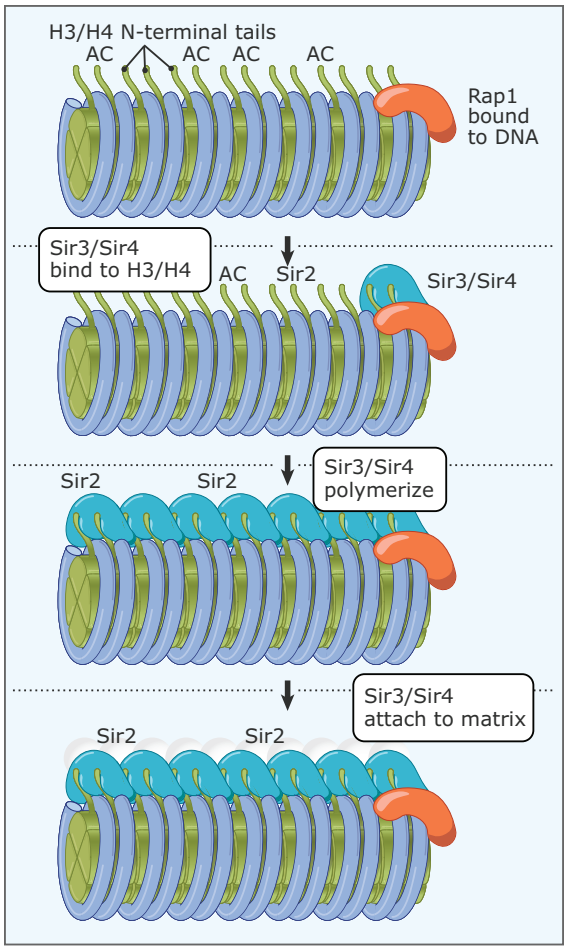

FIGURE 10.81 Formation of heterochromatin is initiated when Rap1 binds to DNA. Sir2 deacetylates histones to facilitate Sir3/Sir4 binding. Sir3/Sir4 bind to RAP1 and also to histones H3/H4. The complex polymerizes along chromatin and may connect telomeres to the nuclear matrix.

deacetylase, and its activity is necessary to maintain binding of the Sir3/4 complex to chromatin.

Rap1 has the crucial role of identifying the DNA sequences at which heterochromatin forms. It recruits Sir4, which in turn recruits both its binding partner Sir3 and the HDAC Sir2. Sir3/Sir4 then interact directly with histones H3 and H4. Once Sir3/Sir4 have bound to histones H3/H4, the complex (including Sir2) may polymerize further and spread along the chromatin fiber. This inactivates the region, either because coating with Sir3/Sir4 itself has an inhibitory effect or because Sir2-dependent deacetylation represses transcription. The C-terminus of Sir3 has a similarity to nuclear lamin proteins (constituents of the nuclear matrix) and may be responsible for tethering heterochromatin to the nuclear periphery.

A similar series of events forms the silenced regions at *HMR* and *HML*. Three sequence-specific factors are involved in triggering formation of the complex: Rap1, Abf1 (a transcription factor), and ORC (the origin replication complex). In this case, Sir1 binds to a sequence-specific factor and recruits Sir2, Sir3, and Sir4 to form the repressive structure.

How does a silencing complex repress chromatin activity? It could condense chromatin so regulator proteins cannot find their targets. The simplest case would be to suppose the presence of a silencing complex is mutually incompatible with the presence of transcription factors and RNA polymerase. The cause could be that silencing complexes block remodeling (and, thus, indirectly prevent factors from binding) or that they directly obscure the binding sites on DNA for the transcription factors. The situation may not be this simple, however, because transcription factors and RNA polymerase can be found at promoters in silenced chromatin. This could mean the silencing complex prevents the factors from working rather than from binding. In fact, there may be competition between gene activators and the repressing effects of chromatin, so that activation of a promoter inhibits spread of the silencing complex.

Centromeric heterochromatin is particularly interesting, because it is not necessarily nucleated by simple sequences (as is the case for telomeres and the mating type loci in yeast) but instead depends on more complex mechanisms, some of which are RNAi dependent. The specialized chromatin structure that forms at the centromere (see *10.8 The centromere binds a protein complex*) may be associated with the formation of heterochromatin in the region. In human cells, the centromere-specific protein CENP-B is required to initiate modifications of histone H3 (deacetylation of K9 and K14, followed by methylation of K9) that trigger an association with HP1 that leads to the formation of heterochromatin in the region. Moreover, heterochromatin and RNAi are required to establish the human CenH3 variant, CENP-A, at centromeres. Heterochromatin is often present near CENP-A chromatin and the RNAi-directed heterochromatin flanking the central kinetochore domain is required for kinetochore assembly. Several factors, such as the Suv39 methyltransferase, HP1, and components of the RNAi pathway are required to form the CENP-A chromatin.

Concept and Reasoning Check

1. What is the role of sequence-specific binding proteins in heterochromatin formation?

10.28 X chromosomes undergo global changes

Key concepts

- One of the two X chromosomes is inactivated at random in each cell during embryogenesis of eutherian mammals.
- In exceptional cases where there are more than two X chromosomes, all but one are inactivated.
- The *Xic* (X-inactivation center) is a *cis*-acting region on the X chromosome that is necessary and sufficient to ensure that only one X chromosome remains active.
- *Xic* includes the *Xist* gene, which codes for an RNA that is found only on inactive X chromosomes.
- The mechanism responsible for preventing *Xist* RNA from accumulating on the active chromosome is unknown.

For species with chromosomal sex determination, the sex of the individual presents an interesting problem for gene regulation, because of the variation in the number of X chromosomes. If X-linked genes were expressed equally well in each sex, females would have twice as much of each product as males. The importance of avoiding this situation is shown by the existence of **dosage compensation**, which equalizes the level of expression of X-linked genes in the two sexes. Mechanisms used in different species are summarized in FIGURE 10.82 as given below:

- In mammals, one of the two female X chromosomes is inactivated. The result is that females have only one active X chromosome, which is the same situation found in males. The active X chromosome of females and the single X chromosome of males are expressed at the same level.
- In *Drosophila*, the expression of the single male X chromosome is doubled relative to the expression of each female X chromosome.
- In *C. elegans*, the expression of each female X chromosome is halved relative to the expression of the single male X chromosome.

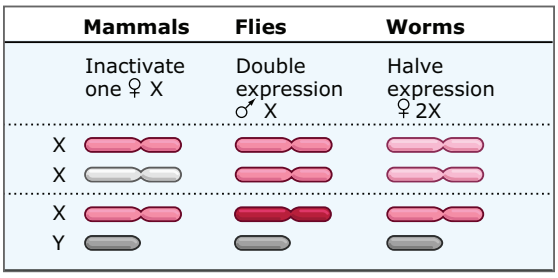

FIGURE 10.82 Different means of dosage compensation are used to equalize X chromosome expression in male and female.

The common feature in all these mechanisms of dosage compensation is that *the entire chromosome is the target for regulation.* A global change occurs that quantitatively affects all the promoters on the chromosome. We know most about the inactivation of the X chromosome in mammalian females, where the entire chromosome becomes heterochromatic.

The twin properties of heterochromatin are its condensed state and associated inactivity, as described earlier in this chapter. Heterochromatin can be divided into the following two types:

1. *Constitutive* heterochromatin contains specific sequences that have no coding function. These include satellite DNAs, which are often found at the centromeres. These regions are invariably heterochromatic because of their intrinsic nature.
2. *Facultative* heterochromatin takes the form of chromosome regions or entire chromosomes that are inactive in one cell lineage, although they can be expressed in other lineages. The example *par excellence* is the mammalian X chromosome. The inactive X chromosome is perpetuated in a heterochromatic state, whereas the active X chromosome is part of the euchromatin. Thus, identical DNA sequences are involved in both states. Once the inactive state has been established, it is inherited by descendant cells. This is an example of epigenetic inheritance, because it does not depend on the DNA sequence.

Our basic view of the situation of the female mammalian X chromosomes was formed by the **single X hypothesis** in 1961. Female mice that are heterozygous for X-linked coat color mutations have a variegated phenotype in which some areas of the coat are wild-type but others are mutant. FIGURE 10.83 shows that this can be explained *if one of the two X chromosomes is inactivated at random in each cell of a small precursor population.* Cells in which the X chromosome carrying the wild-type gene is inactivated give rise to progeny that express only the mutant allele on the active chromosome. Cells derived from a precursor where the other chromosome was inactivated have an active wild-type gene. In the case of coat

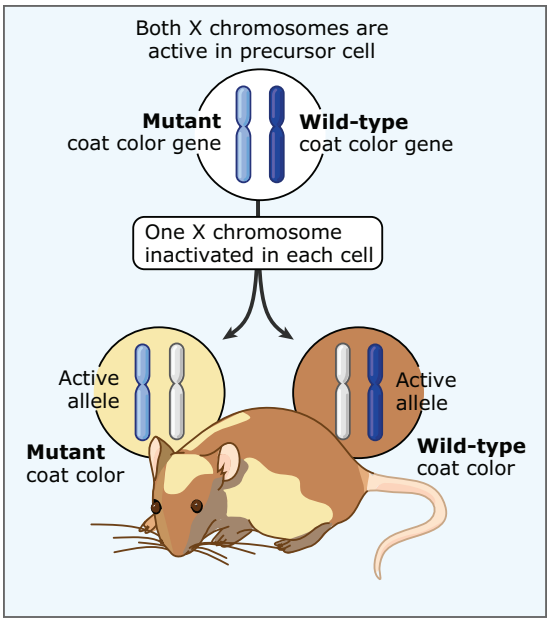

FIGURE 10.83 X-linked variegation is caused by the random inactivation of one X chromosome in each precursor cell. Cells in which the plus allele is on the active chromosome have wild phenotype, but cells in which the minus allele is on the active chromosome have mutant phenotype.

color, cells descended from a particular precursor stay together and thus form a patch of the same color, creating the pattern of visible variegation. In other cases, individual cells in a population express one or the other of X-linked alleles; for example, in heterozygotes for the X-linked locus *G6PD*, any particular red blood cell expresses only one of the two allelic forms. Interestingly, random inactivation of one X chromosome occurs in eutherian mammals, whereas in marsupials the choice is directed: It is always the X chromosome inherited from the father that is inactivated.

Inactivation of the X chromosome in females is governed by the **n–1 rule**; however many X chromosomes are present, all but one will be inactivated. In normal females, there are of course two X chromosomes, but in rare cases where nondisjunction has generated a 3X or greater genotype, only one X chromosome remains active. This suggests a general model in which a specific event is limited to one X chromosome and protects it from an inactivation mechanism that applies to all others.

A single locus on the X chromosome is sufficient for inactivation. When a translocation occurs between the X chromosome and an autosome, this locus is present on only one of the reciprocal products, and only that product can be inactivated. By comparing different translocations, it is possible to map this locus, which is called the *Xic* (X-inactivation center). A cloned region of 450 kb contains all the properties of the *Xic.* When this sequence is inserted as a transgene onto an autosome, the autosome becomes subject to inactivation (at least in a cell culture system). Pairing of *Xic* loci on the two X chromosomes has been implicated in the mechanism for the random choice of X inactivation. Moreover, differences in the sister chromatid cohesion correlate with the outcome of the choice of the X chromosome to be inactivated, indicating that alternate states present before the inactivation process may direct the choice of which X chromosome will become inactivated.

Xic is a *cis*-acting locus that contains the information necessary to count X chromosomes and inactivate all copies but one. Inactivation spreads from *Xic* along the entire X chromosome. When *Xic* is present on an X chromosome-autosome translocation, inactivation spreads into the autosomal regions (although the effect is not always complete).

Xic is a complex genetic locus that expresses several long noncoding RNAs. The most important of these is a gene called *Xist* (X inactive specific transcript) that is stably expressed only on the *inactive* X chromosome. The behavior of this gene is effectively the opposite from all other loci on the chromosome, which are turned off. Deletion of *Xist* prevents an X chromosome from being inactivated, however, it does not interfere with the counting mechanism (because other X chromosomes can be inactivated). So we can distinguish two features of *Xic:* an unidentified element(s) required for counting and the *Xist* gene required for inactivation.

The n–1 rule suggests that stabilization of *Xist* RNA is the "default" and that some blocking mechanism prevents stabilization at one X chromosome (which will be the active X). This means that although *Xic* is necessary and sufficient for a chromosome to be *inactivated,* the products of other loci are necessary for the establishment of an *active* X chromosome.

FIGURE 10.84 illustrates the role of *Xist* RNA in X inactivation. *Xist* codes for an RNA that lacks open reading frames. The *Xist* RNA "coats" the X chromosome from which it is synthesized, suggesting it has a structural role. Before X inactivation, it is synthesized by both female X chromosomes. After inactivation, the RNA is found only on the inactive X chromosome. The transcription rate remains the same before and after inactivation, so the transition depends on

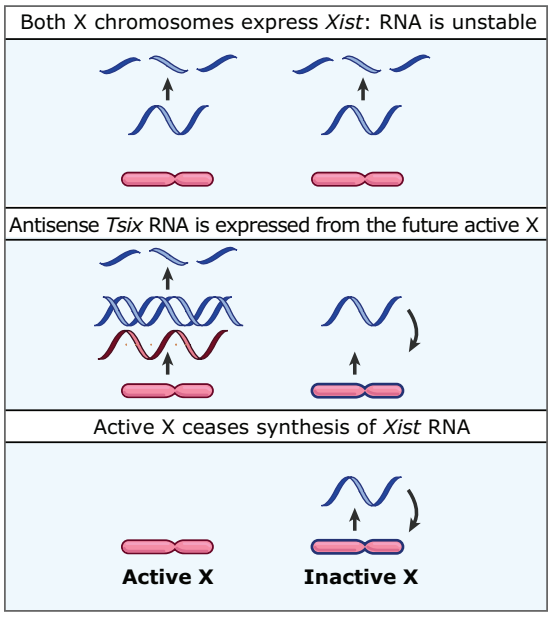

FIGURE 10.84 X inactivation involves stabilization of *Xist* RNA, which coats the inactive chromosome. *Tsix* prevents *Xist* expression on the future active X.

posttranscriptional events. An antisense RNA generated from *Xic, Tsix,* plays a role in regulating the stability of *Xist. Tsix* is active on the future active X but downregulated on the future inactive X; this regulation allows persistence of *Xist* on the inactive X and thus leads to silencing.

Before X inactivation, *Xist* RNA decays with a half-life of ~2 hours. X inactivation is mediated by stabilizing the *Xist* RNA on the inactive X chromosome. The *Xist* RNA shows a punctate distribution along the X chromosome, suggesting that association with proteins to form particulate structures may be the means of stabilization. We do not know yet what other factors may be involved in this reaction, although some other noncoding RNAs and protein factors have been implicated. We also do not know how the *Xist* RNA is limited to spreading in *cis* along the chromosome. Accumulation of *Xist* on the future inactive X results in exclusion of transcription machinery (such as RNA polymerase II) and leads to a series of chromosome-wide histone modifications, including H4 deacetylation and specific methylation of both H3 and H4. Late in the process, an inactive X-specific histone variant, macroH2A, is incorporated into the chromatin (see Figure 10.44), and promoter DNA is methylated (see *10.27 Heterochromatin depends on interactions with histones*). At this point, the heterochromatic state of the inactive X is stable, and *Xist* is not required to maintain the silent state.

Global changes also occur in other types of dosage compensation. In *Drosophila,* a large ribonucleoprotein complex, MSL, is found only in males, where it localizes on the X chromosome. This complex contains two noncoding RNAs, which appear to be needed for localization to the male X (perhaps analogous to the localization of *Xist* to the inactive mammalian X) and a HAT that acetylates histone H4 on K16 throughout the male X. The net result of the action of this complex is the twofold increase in transcription of all genes on the male X. In the next section, we discuss a third mechanism for dosage compensation, a global *reduction* in X-linked gene expression in XX (hermaphrodite) nematodes.

Concept and Reasoning Check

1. The *Drosophila* dosage compensation complex contains an acetyltransferase. Explain how this promotes dosage compensation in *Drosophila*.

10.29 Chromosome condensation is caused by condensins

Key concepts

- SMC proteins are ATPases that include the condensins and the cohesins.
- A heterodimer of SMC proteins associates with other subunits.
- Cohesins are responsible for holding sister chromatids together.
- Condensins cause chromatin to be more tightly coiled by introducing positive supercoils into DNA.
- Condensins are responsible for condensing chromosomes at mitosis.
- Chromosome-specific condensins are responsible for condensing inactive X chromosomes in *C. elegans*.

The structures of entire chromosomes are influenced by interactions with proteins of the **SMC** (structural maintenance of chromosome) family. They are ATPases that fall into two functional groups. **Condensins** are involved with the control of overall structure and are responsible for the condensation into compact chromosomes at mitosis. **Cohesins** are concerned with connections between sister chromatids that must be released at mitosis. Both consist of dimers formed by SMC proteins. Condensins form complexes that have a core of the heterodimer SMC2-SMC4 associated with other (non-SMC) proteins. Cohesins have a similar

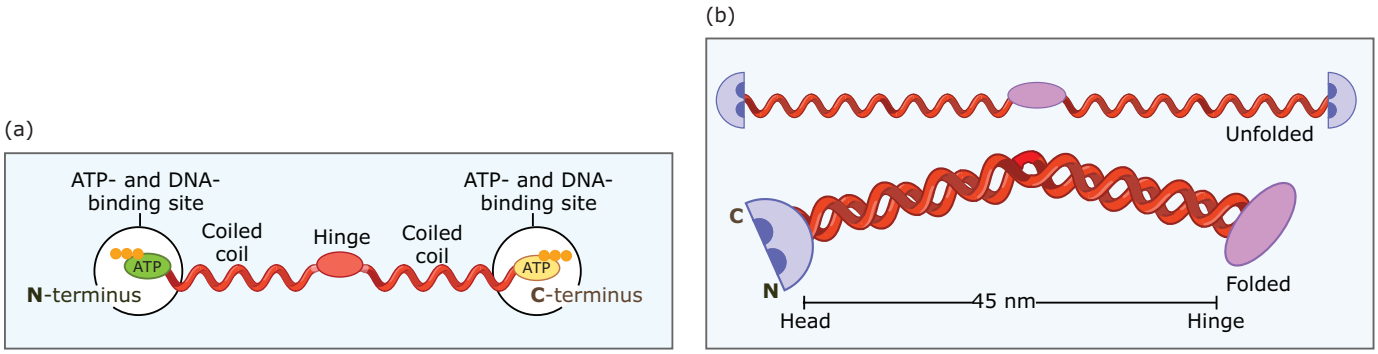

FIGURE 10.85 (A) An SMC protein has an ATP-binding motif and DNA binding site at each end, connected by coiled coils that are linked by a hinge region. (B) SMC monomers fold at the hinge region and interact along the length of the coiled coils. The N- and C-termini interact to form a head domain. (Reprinted, with permission, from the *Annual Reviews of Cell and Developmental Biology,* Volume 24 © 2008 by Annual Reviews www.annualreviews.org. Additional permission courtesy of Itay Onn, Howard Hughes Medical Institute.)

organization based on the heterodimeric core of SMC1-SMC3 and also interact with non-SMC partners.

FIGURE 10.85 shows that an SMC protein has a coiled-coil structure in its center, interrupted by a globular flexible hinge region. Both the amino and carboxyl termini have ATP and DNA binding motifs. SMC monomers fold at the hinge region, forming an antiparallel interaction between the two halves of each coiled coil. This allows the amino and carboxyl termini to interact to form a "head" domain.

Folded SMC proteins form dimers via several different interactions. The most stable association occurs between hydrophobic domains in the hinge regions. **FIGURE 10.86** shows that these hinge-to-hinge interactions result in V-shaped structures. Electron microscopy shows that in solution, cohesins tend to form Vs with the arms separated by large angle, whereas condensins form more linear structures, with only a small gap between the arms. In addition, the heads of the two monomers can interact, closing the V, and the coiled coils of the individual monomers may also interact with each other. Various non-SMC proteins interact with SMC dimers and can influence the final structure of the dimer.

The function of cohesins is to hold sister chromatids together, but it is not yet clear how this is achieved. There are several different models for cohesin function. **FIGURE 10.87** shows one model in which a cohesin could take the form of extended dimers, interacting hinge-to-hinge, that cross-link two DNA molecules. Head-to-head interactions would create tetrameric structures, adding to the stability of cohesion. An alternative "ring" model is shown in **FIGURE 10.88**. In this model, dimers interact at both their head and hinge regions to form a circular structure. Instead of binding directly to DNA, a structure of this type could hold DNA molecules together by encircling them.

Although cohesins act to hold separate sister chromatids together, condensins are responsible for chromatin condensation. **FIGURE 10.89** shows that a condensin could take the form of a V-shaped dimer, interacting via the hinge domains, that pulls together distant

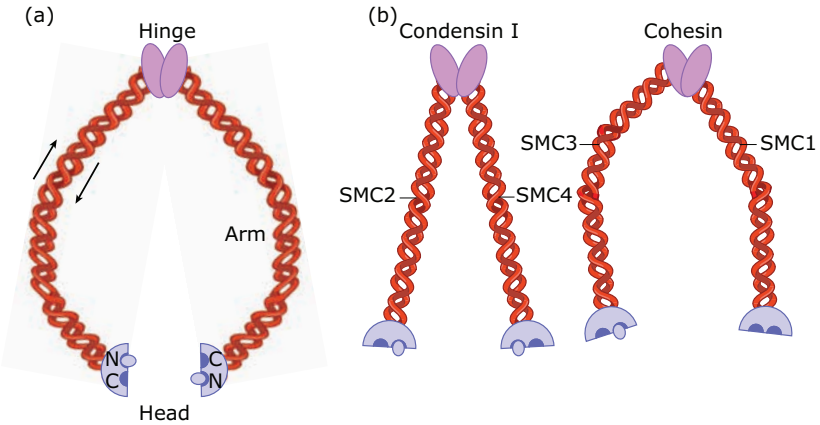

FIGURE 10.86 The basic architecture of condensin and cohesin complexes. Condensin and cohesin consist of V-shaped dimers of two SMC proteins interacting through their hinge domains. The two monomers in a condensin dimer tend to exhibit a very small separation between the two arms of the V, whereas cohesins have a much larger angle of separation between the arms. (Adapted from T. Hirano, *Nat. Rev. Mol. Cell Biol.* 7 (2006): 311–322.)

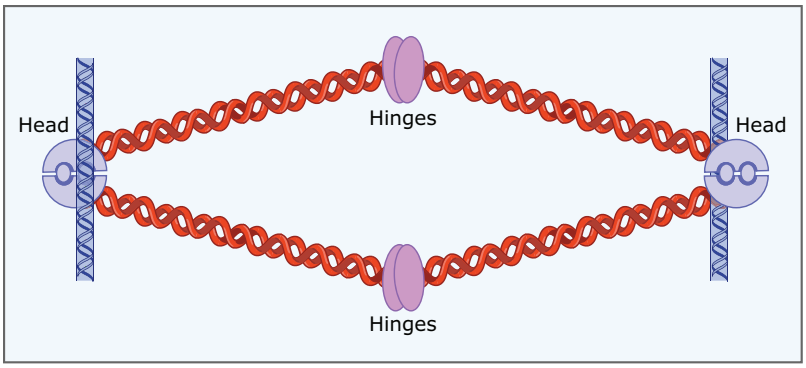

FIGURE 10.87 One model for DNA linking by cohesins. Cohesins may form an extended structure in which each monomer binds DNA and connects via the hinge region, allowing two different DNA molecules to be linked. Head domain interactions can result in binding by two cohesin dimers. (Reproduced from *Annual Reviews* by Itay Onn, et al., Copyright 2008 by Annual Reviews, Inc. Reproduced with permission of Annual Reviews, Inc. in the format of Textbook via Copyright Clearance Center.)

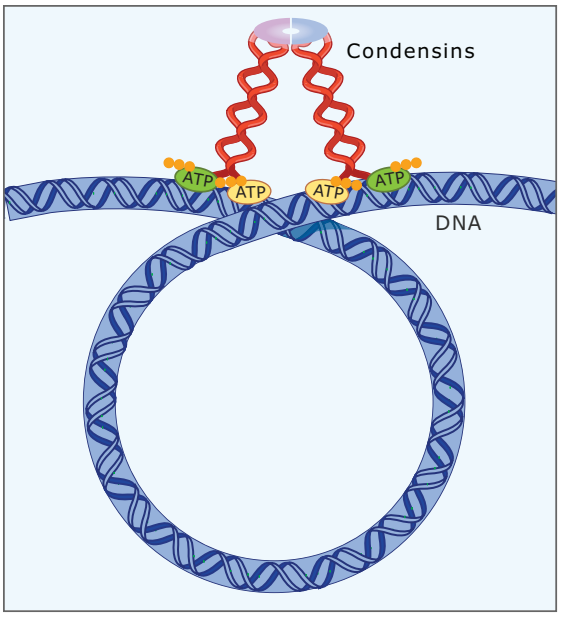

FIGURE 10.89 Condensins may form a compact structure by bending at the hinge, causing DNA to become compacted.

sites on the same DNA molecule, causing it to condense (**FIGURE 10.90**). It is thought that dynamic head-to-head interactions could act to promote the ordered assembly of condensed loops, but the details of condensin action are still far from clear.

Consistent with the looping model described above, the condensin complex has an ability to introduce positive supercoils into DNA in an action that uses hydrolysis of ATP and depends on the presence of topoisomerase I. This ability is controlled by the phosphorylation of non-SMC subunits, which occurs at mitosis. It is now known how this connects with other modifications of chromatin, such as the phosphorylation of histones, which is also linked to chromosome condensation.

We discussed in the previous section the dramatic chromosomal changes that occur during X inactivation in female mammals and in X chromosome upregulation in male flies. In the nematode *C. elegans,* a third approach is used: twofold *reduction* of X-chromosome transcription in XX hermaphrodites relative to XO males. A dosage compensation complex is maternally provided to both XX and XO embryos, but it then associates with both X chromosomes only in XX animals, while remaining diffusely distributed in XO nuclei. The dosage compensation complex contains a core of SMC proteins and is similar to the condensin complexes that are associated with mitotic chromosomes in *C. elegans* and other species. This suggests it has a structural role in causing

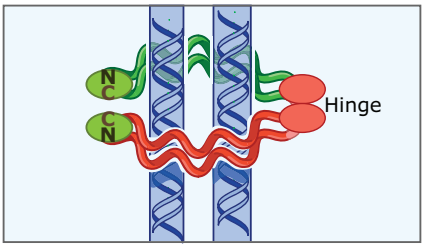

FIGURE 10.88 Cohesins may dimerize by intramolecular connections, then forming multimers that are connected at the heads and at the hinge. Such a structure could hold two molecules of DNA together by surrounding them.

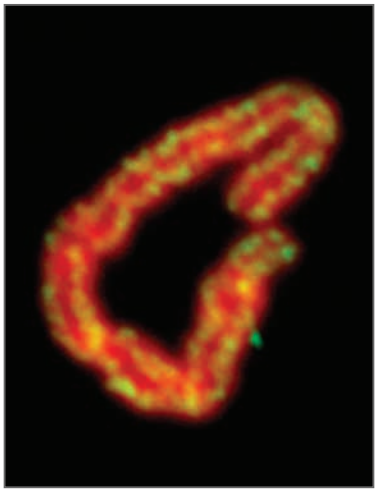

FIGURE 10.90 Condensins are located along the entire length of a mitotic chromosome. DNA is red; condensins are yellow. (Photo courtesy of Ana Losada and Tatsuya Hirano, Cold Spring Harbor Laboratory.)

the chromosome to take up a more condensed, inactive state. Multiple sites on the X chromosome may be needed for the complex to be fully distributed along it, and short DNA sequence motifs have been identified that appear to be key for localization of dosage compensation complex. The complex binds to these sites and then spreads along the chromosome to cover it more thoroughly.

Dosage compensation in mammals and *Drosophila* both entail chromosome-wide changes in histone acetylation and involve noncoding RNAs that play central roles in targeting X chromosomes for global change. In *C. elegans*, chromosome condensation by condensin homologs is used to accomplish dosage compensation. It remains to be seen whether there are also global changes in the histone acetylation or other modifications in XX *C. elegans* that reflect the twofold reduction in transcription of the X chromosomes.

Concept and Reasoning Check

1. SMC homologs have not been implicated in mammalian or *Drosophila* dosage compensation. If SMCs did have a role in dosage compensation, would you expect it to be more likely in mammals or *Drosophila*? Why?

10.30 What's next?

Since the discovery of the nucleosome revolutionized the study of chromatin, there have been two concurrent lines of research:

1. The analysis of structure, both to describe the nucleosome itself and to describe how the nucleosome is organized into higher-order structures
2. The analysis of function in terms of the nucleosome, to relate the events that occur in activating and during transcription to the structure of chromatin

The most immediately pressing questions along these lines are to define the mechanisms that control higher-order folding of the 30-nm chromatin filament and to resolve how the transcription-activating and -inactivating apparatuses act via epigenetic modification of nucleosomes to achieve their effects.

In the 1970s, histones were seen as general repressors of transcription whose effects had to be neutralized for genes to be expressed. There was, however, little insight into how this might be accomplished beyond the belief that they somehow had to be removed from chromatin to enable activators to get at DNA. Consider how much more we know now, and we can see that we are on the verge of achieving a definition of chromatin function in terms of structure. There is terrific progress in defining the roles of the acetylating, deacetylating, and other enzymes that modify histones, and we are beginning to see how they change chromatin structure in a localized way to enable activation of a promoter. We are close to defining promoter activation in terms of these structural interactions. And viewing histones as repressors, we are now beginning to understand how their interactions with other proteins create heterochromatic structures that propagate epigenetically and may inactivate local or even quite wide regions of chromatin. We can expect to see all these events described in terms of increased resolution of the structures of individual components, that is, in terms of the molecular changes in individual histones and other proteins of chromatin.

Understanding of hierarchical organization at higher levels, that is, of the chromosome itself, is still difficult. In spite of the enormous progress made in genome analysis, we do not understand the significance of the structural features of bands and interbands, of the concentrations of G-C–rich regions, and so on. We cannot yet relate these features to interactions in terms of DNA and proteins; even when the structure of the 30-nm fiber is resolved, there will still be much to learn about the higher orders of structure. Other distinctive features of the chromosome, most crucially the centromere and telomere, are yielding to description in terms of molecular components, although we have yet to understand the role of DNA sequences in the higher eukaryotic centromere. As in other areas of biology, the ability to relate increased structural resolution to functional changes is providing powerful insights at deeper levels of understanding.

10.31 Summary

The genetic material of all organisms and viruses takes the form of tightly packaged nucleoprotein. In eukaryotes, transcriptionally active sequences reside within the euchromatin that constitutes the majority of interphase chromatin. The regions of heterochromatin are packaged ~5–10 times more compactly and are transcriptionally inert. All chromatin becomes densely packaged during cell division, when the individual chromosomes can be distinguished. The existence of a reproducible ultrastructure

in chromosomes is indicated by the production of G-bands by treatment with Giemsa stain. The bands are very large regions, ~10^7 bp that can be used to map chromosomal translocations or other large changes in structure.

In eukaryotes, interphase chromatin and metaphase chromosomes both appear to be organized into large loops. Each loop may be an independently supercoiled domain. The bases of the loops are connected to a metaphase scaffold or to the nuclear matrix by specific DNA sites.

The centromeric region contains the kinetochore, which is responsible for attaching a chromosome to the mitotic spindle. The centromere often is surrounded by heterochromatin. Centromeric sequences have been identified only in yeast *S. cerevisiae,* where they consist of short conserved elements—*CDE-I* and *CDE-III,* which bind Cbf1 and the CBF3 complex, respectively—and a long A · T-rich region called *CDE-II* that binds the histone variant Cse4 to form a specialized structure in chromatin. Another group of proteins that binds to this assembly provides the connection to microtubules.

Telomeres make the ends of chromosomes stable. Almost all known telomeres consist of multiple repeats in which one strand has the general sequence $C_n(A/T)_m$, where $n > 1$ and $m = 1 - 4$. The other strand, $G_n(T/A)_m$, has a single protruding end that provides a template for addition of individual bases in defined order. The enzyme telomerase is a ribonucleoprotein, whose RNA component provides the template for synthesizing the G-rich strand. This overcomes the problem of the inability to replicate at the very end of a duplex. The telomere stabilizes the chromosome end because the overhanging single strand $G_n(T/A)_m$ displaces its homologue in earlier repeating units in the telomere to form a loop, so there are no free ends.

Lampbrush chromosomes of amphibians and polytene chromosomes of insects have unusually extended structures with packing ratios < 100. Polytene chromosomes of *D. melanogaster* are divided into ~5,000 bands, varying in size by an order of magnitude, with an average of ~25 kb. Transcriptionally active regions can be visualized in even more unfolded ("puffed") structures, in which material is extruded from the axis of the chromosome. This may resemble the changes that occur on a smaller scale when a sequence in euchromatin is transcribed.

All eukaryotic chromatin consists of nucleosomes. A nucleosome contains a characteristic length of DNA, usually ~200 bp, wrapped around an octamer containing two copies each of histones H2A, H2B, H3, and H4. A single H1 protein can be associated with each nucleosome. Virtually all genomic DNA is organized into nucleosomes. Treatment with micrococcal nuclease shows the DNA packaged into each nucleosome can be divided operationally into two regions. The linker region is digested rapidly by the nuclease; the core region of 146 bp is resistant to digestion. Histones H3 and H4 are the most highly conserved and an $H3_2 \cdot H4_2$ tetramer accounts for the diameter of the particle. The H2A and H2B histones are organized as two H2A · H2B dimers. Octamers are assembled by the successive addition of two H2A · H2B dimers to the $H3_2 \cdot H4_2$ kernel.

Nucleosomes are organized into a fiber of 30-nm diameter, which has six nucleosomes per turn and a packing ratio of 40. Removal of H1 allows this fiber to unfold into a 10-nm fiber that consists of a linear string of nucleosomes. The 30-nm fiber probably consists of the 10-nm fiber wound into a two-start solenoid. The 30-nm fiber is the basic constituent of both euchromatin and heterochromatin; nonhistone proteins are responsible for further organization of the fiber into chromatin or chromosome ultrastructure.

There are two pathways for nucleosome assembly. In the RC pathway, the proliferating cell nuclear antigen processivity subunit of the replisome recruits CAF-1, which is a nucleosome assembly factor. CAF-1 assists the deposition of $H3_2 \cdot H4_2$ tetramers onto the daughter duplexes resulting from replication. The tetramers may be produced either by disruption of existing nucleosomes by the replication fork or as the result of assembly from newly synthesized histones. Similar sources provide the H2A · H2B dimers that then assemble with the $H3_2 \cdot H4_2$ tetramer to complete the nucleosome. Because the $H3_2 \cdot H4_2$ tetramer and the H2A · H2B dimers assemble at random, the new nucleosomes may include both preexisting and newly synthesized histones. HIRA assembles nucleosomes outside of S phase, and ASF1 acts both during and outside replication to assemble chromatin.

RNA polymerase displaces histone octamers during transcription. Nucleosomes reform on DNA after the polymerase has passed, unless transcription is very intensive (such as in rDNA) when they may be displaced completely. The RI pathway for nucleosome assembly is responsible for replacing histone octamers that have been displaced by transcription. It uses the histone variant H3.3 instead of H3. A similar pathway, with another alternative to H3,

is used for assembling nucleosomes at centromeric DNA sequences after replication.

Two types of changes in sensitivity to nucleases are associated with gene activity. Chromatin capable of being transcribed has a generally increased sensitivity to DNase I, reflecting a change in structure over an extensive region that can be defined as a domain containing active or potentially active genes. Hypersensitive sites in DNA occur at discrete locations and are identified by greatly increased sensitivity to DNase I. A hypersensitive site consists of a sequence of ~200 bp from which nucleosomes are excluded by the presence of other proteins. A hypersensitive site forms a boundary that may cause adjacent nucleosomes to be restricted in position. Nucleosome positioning may be important in controlling access of regulatory proteins to DNA.

Genes whose control regions are organized in nucleosomes usually are not expressed. In the absence of specific regulatory proteins, promoters and other regulatory regions are organized by histone octamers into a state in which they cannot be activated. This may explain the need for nucleosomes to be precisely positioned in the vicinity of a promoter, so that essential regulatory sites are appropriately exposed. Some transcription factors have the capacity to recognize DNA on the nucleosomal surface, and a particular positioning of DNA may be required for initiation of transcription.

Active chromatin and inactive chromatin are not in equilibrium. Sudden, disruptive events are needed to convert one to the other. Chromatin remodeling complexes have the ability to slide or displace histone octamers by a mechanism that involves hydrolysis of ATP. Remodeling complexes range from small to extremely large and are classified according to the type of the ATPase subunit. The common types are SWI/SNF, ISWI, CHD, and INO80. A typical form of this chromatin remodeling is to displace one or more histone octamers from specific sequences of DNA, creating a boundary that results in the precise or preferential positioning of adjacent nucleosomes. Chromatin remodeling may also involve changes in the positions of nucleosomes, sometimes involving sliding of histone octamers along DNA.

Extensive covalent modifications occur on histone tails, all of which are reversible. Acetylation of histones occurs at both replication and transcription and could be necessary to form a less compact chromatin structure. Some coactivators that connect transcription factors to the basal apparatus have HAT activity. Conversely, repressors may be associated with deacetylases. The modifying enzymes are usually specific for particular amino acids in particular histones. The most common sites for modification are located in the N-terminal tails of histones H3 and H4, which extrude from nucleosomes between the turns of DNA. The activating (or repressing) complexes are usually large and often contain several activities that undertake different modifications of chromatin.

The formation of heterochromatin occurs by proteins that bind to specific chromosomal regions (such as telomeres) and that interact with histones. The formation of an inactive structure may propagate along the chromatin thread from an initiation center. Similar events occur in silencing of the inactive yeast mating type loci.

Formation of heterochromatin may be initiated at certain sites and then propagated for a distance that is not precisely determined. When a heterochromatic state has been established, it is inherited through subsequent cell divisions. This gives rise to a pattern of epigenetic inheritance, in which two identical sequences of DNA may be associated with different protein structures and therefore have different abilities to be expressed. This explains the occurrence of position effect variegation in *Drosophila.*

Modification of histone tails is a trigger for chromatin reorganization. Acetylation is generally associated with gene activation. HATs are found in activating complexes, and HDACs are found in inactivating complexes. Histone methylation at specific sites is associated with gene inactivation, as is DNA methylation. Some histone modifications may be exclusive or synergistic with others.

Inactive chromatin at yeast telomeres and silent mating type loci appear to have a common structure and involve the interaction of certain proteins with the N-terminal tails of histones H3 and H4. Formation of the inactive complex may be initiated by binding of one protein to a specific sequence of DNA; the other components may then polymerize in a cooperative manner along the chromosome.

Inactivation of one X chromosome in female (eutherian) mammals occurs at random. The *Xic* locus is necessary and sufficient to count the number of X chromosomes. The n–1 rule ensures that all but one X chromosome are inactivated. *Xic* contains the gene *Xist,* which codes for an RNA that is expressed only on the inactive X chromosome. Stabilization of *Xist* RNA is the mechanism by which the inactive X chromosome is distinguished.

SMC proteins can control the global structure of chromosomes. Cohesins hold sister chromatids together, whereas condensins are responsible for chromosome condensation during mitosis (and meiosis). Specialized homologs of condensins are used to reduce X chromosome gene expression for dosage compensation in *C. elegans*.

http://biology.jbpub.com/lewin/cells/3e

To explore these topics in more detail, visit this book's Interactive Student Study Guide.

References

Ahmad, K., and Henikoff, S. 2002. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* v. 9 p. 1191–1200.

Black, B. E., and Bassett, E. A. 2008. The histone variant CENP-A and centromere specification. *Curr. Opin. Cell Biol.* v. 20 p. 91–100.

Hirano, T. 2006. At the heart of the chromosome: SMC proteins in action. *Nat. Rev. Mol. Cell Biol.* v. 7 p. 311–322.

Li, B., Carey, M., and Workman, J. L. 2007. The role of chromatin during transcription. *Cell* v. 128 p. 707–719.

Luger, K., et al. 1997. Crystal structure of the nucleosome core particle at 28 Å resolution. *Nature* v. 389 p. 251–260.

Payer, B., and Lee, J. T. 2008. X chromosome dosage compensation: How mammals keep the balance. *Annu. Rev. Genet.* v. 42 p. 733–772.

Vignali, M., Hassan, A. H., Neely, K. E., and Workman, J. L. 2000. ATP-dependent chromatin-remodeling complexes. *Mol. Cell Biol.* v. 20 p. 1899–1910.

Workman, J. L. 2006. Nucleosome displacement during transcription. *Genes Dev.* v. 20 p. 2507–2512.

Yuan, G. C., et al. 2005. Genome scale identification of nucleosome positions in *S. cerevisiae*. *Science* v. 309 p. 626–630.

Zhang, Y., and Reinberg, D. 2001. Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails. *Genes Dev.* v. 15 p. 2343–2360.