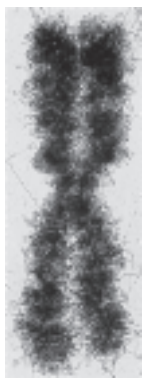


chapter 24

GENES AND CHROMOSOMES

DNA topoisomerases are the magicians of the DNA world. By allowing DNA strands or double helices to pass through each other, they can solve all of the topological problems of DNA in replication, transcription and other cellular transactions.

—James Wang, article in Nature Reviews in Molecular Cell Biology, *2002*

Supercoiling, in fact, does more for DNA than act as an executive enhancer; it keeps the unruly, spreading DNA inside the cramped confines that the cell has provided for it.

—Nicholas Cozzarelli, Harvey Lectures, *1993*

Almost every cell of a multicellular organism contains the same complement of genetic material—its **genome.** Just look at any human individual for a hint of the wealth of information contained in each human cell. **Chromosomes,** the nucleic acid molecules that are the repository of an organism's genetic information, are the largest molecules in a cell and may contain thousands of genes as well as considerable tracts of intergenic DNA. The 16 chromosomes in the relatively small genome of the yeast *Saccharomyces cerevisiae* have molecular masses ranging from 1.5×10^8 to 1×10^9 daltons, corresponding to DNA molecules with 230,000 to 1,532,000 contiguous base pairs (bp). Human chromosomes range up to 279 million bp.

The very size of DNA molecules presents an interesting biological puzzle, given that they are generally much longer than the cells or viral packages that contain them (Fig. 24–1). In this chapter we shift our focus from the secondary structure of DNA, considered in Chapter 8, to the extraordinary degree of organization required for the tertiary packaging of DNA into chromosomes. We first examine the elements within viral and cellular chromosomes, then assess their size and organization. We next consider DNA topology, providing a

FIGURE 24–1 Bacteriophage T2 protein coat surrounded by its single, linear molecule of DNA. The DNA was released by lysing the bacteriophage particle in distilled water and allowing the DNA to spread on the water surface. An undamaged T2 bacteriophage particle consists of a head structure that tapers to a tail by which the bacteriophage attaches itself to the outer surface of a bacterial cell. All the DNA shown in this electron micrograph is normally packaged inside the phage head.

description of the coiling of DNA molecules. Finally, we discuss the protein-DNA interactions that organize chromosomes into compact structures.

24.1 Chromosomal Elements

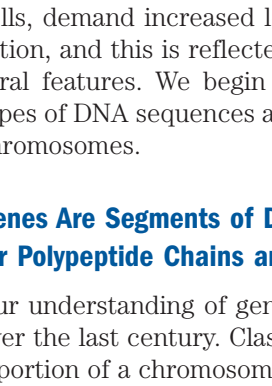
Cellular DNA contains genes and intergenic regions, both of which may serve functions vital to the cell. The more complex genomes, such as those of eukaryotic cells, demand increased levels of chromosomal organization, and this is reflected in the chromosome's structural features. We begin by considering the different types of DNA sequences and structural elements within chromosomes.

Genes Are Segments of DNA That Code for Polypeptide Chains and RNAs

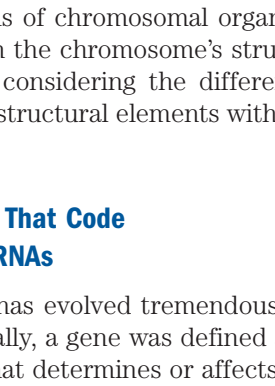
Our understanding of genes has evolved tremendously over the last century. Classically, a gene was defined as a portion of a chromosome that determines or affects a single character or **phenotype** (visible property), such as eye color. George Beadle and Edward Tatum proposed a molecular definition of a gene in 1940. After exposing spores of the fungus *Neurospora crassa* to x rays and other agents known to damage DNA and cause alterations in DNA sequence **(mutations),** they detected mutant fungal strains that lacked one or another specific enzyme, sometimes resulting in the failure of an entire metabolic pathway. Beadle and Tatum concluded that a gene is a segment of genetic material that determines or codes for one enzyme: the **one gene–one enzyme** hypothesis. Later this concept was broadened to **one gene–one polypeptide,** because many genes code for proteins that are not enzymes or for one polypeptide of a multisubunit protein.

The modern biochemical definition of a gene is even more precise. A gene is all the DNA that encodes the primary sequence of some final gene product, which can be either a polypeptide or an RNA with a structural or catalytic function. DNA also contains other segments or sequences that have a purely regulatory function. **Regulatory sequences** provide signals that may denote the beginning or the end of genes, or influence the transcription of genes, or function as initiation points for replication or recombination (Chapter 28). Some genes can be expressed in different ways to generate multiple gene products from one segment of DNA. The special transcriptional and translational mechanisms that allow this are described in Chapters 26 through 28.

We can make direct estimations of the minimum overall size of genes that encode proteins. As described in detail in Chapter 27, each amino acid of a polypeptide chain is coded for by a sequence of three consecutive nucleotides in a single strand of DNA (Fig. 24–2), with these "codons" arranged in a sequence that corresponds to the sequence of amino acids in the polypeptide that the gene encodes. A polypeptide chain of 350 amino acid residues (an average-size chain) corre-



George W. Beadle, 1903–1989



Edward L. Tatum, 1909–1975

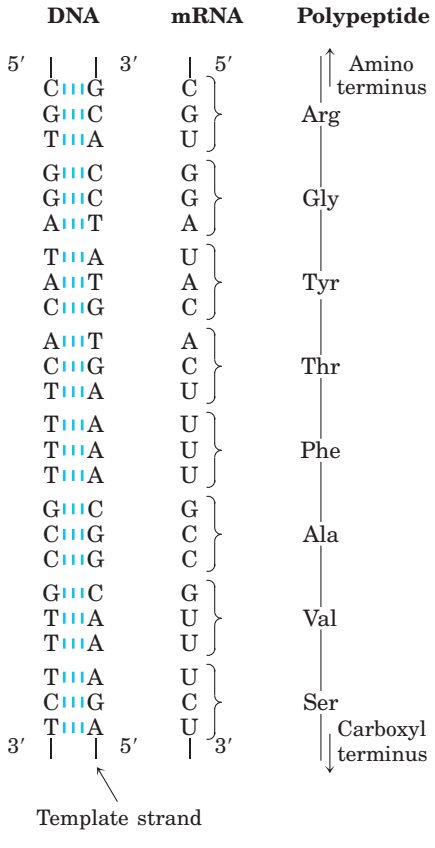


FIGURE 24–2 Colinearity of the coding nucleotide sequences of DNA and mRNA and the amino acid sequence of a polypeptide chain. The triplets of nucleotide units in DNA determine the amino acids in a protein through the intermediary mRNA. One of the DNA strands serves as a template for synthesis of mRNA, which has nucleotide triplets (codons) complementary to those of the DNA. In some bacterial and many eukaryotic genes, coding sequences are interrupted at intervals by regions of noncoding sequences (called introns).

sponds to 1,050 bp. Many genes in eukaryotes and a few in prokaryotes are interrupted by noncoding DNA segments and are therefore considerably longer than this simple calculation would suggest.

How many genes are in a single chromosome? The *Escherichia coli* chromosome, one of the prokaryotic genomes that has been completely sequenced, is a circular DNA molecule (in the sense of an endless loop rather than a perfect circle) with 4,639,221 bp. These base pairs encode about 4,300 genes for proteins and another 115 genes for stable RNA molecules. Among eukaryotes, the approximately 3.2 billion base pairs of the human genome include 30,000 to 35,000 genes on 24 different chromosomes.

DNA Molecules Are Much Longer Than the Cellular Packages That Contain Them

Chromosomal DNAs are often many orders of magnitude longer than the cells or viruses in which they are found (Fig. 24–1; Table 24–1). This is true of every class of organism or parasite.

Viruses Viruses are not free-living organisms; rather, they are infectious parasites that use the resources of a host cell to carry out many of the processes they require to propagate. Many viral particles consist of no more than a genome (usually a single RNA or DNA molecule) surrounded by a protein coat.

Almost all plant viruses and some bacterial and animal viruses have RNA genomes. These genomes tend to be particularly small. For example, the genomes of mammalian retroviruses such as HIV are about 9,000 nucleotides long, and that of the bacteriophage Qβ has 4,220 nucleotides. Both types of viruses have single-stranded RNA genomes.

The genomes of DNA viruses vary greatly in size (Table 24–1). Many viral DNAs are circular for at least part of their life cycle. During viral replication within a host cell, specific types of viral DNA called **replicative forms** may appear; for example, many linear DNAs become circular and all single-stranded DNAs become double-stranded. A typical medium-sized DNA virus is bacteriophage λ (lambda), which infects *E. coli.* In its replicative form inside cells, λ DNA is a circular double helix. This double-stranded DNA contains 48,502 bp and has a contour length of 17.5 μm. Bacteriophage ϕX174 is a much smaller DNA virus; the DNA in the viral particle is a single-stranded circle, and the double-stranded replicative form contains 5,386 bp. Although viral genomes are small, the contour lengths of their DNAs are much greater than the long dimensions of the viral particles that contain them. The DNA of bacteriophage T4, for example, is about 290 times longer than the viral particle itself (Table 24–1).

Bacteria A single *E. coli* cell contains almost 100 times as much DNA as a bacteriophage λ particle. The chromosome of an *E. coli* cell is a single double-stranded circular DNA molecule. Its 4,639,221 bp have a contour length of about 1.7 mm, some 850 times the length of the *E. coli* cell (Fig. 24–3). In addition to the very large, circular DNA chromosome in their nucleoid, many bacteria contain one or more small circular DNA molecules that are free in the cytosol. These extrachromosomal elements are called **plasmids** (Fig. 24–4; see also p. 311). Most plasmids are only a few thousand base pairs long, but some contain more than 10,000 bp. They carry genetic information and undergo replication to yield daughter plasmids, which pass into the daughter cells at cell division. Plasmids have been found in yeast and other fungi as well as in bacteria.

In many cases plasmids confer no obvious advantage on their host, and their sole function appears to be self-propagation. However, some plasmids carry genes that are useful to the host bacterium. For example, some plasmid genes make a host bacterium resistant to antibacterial agents. Plasmids carrying the gene for the enzyme β-lactamase confer resistance to β-lactam antibiotics such as penicillin and amoxicillin (see Box 20–1). These and similar plasmids may pass from an antibiotic-resistant cell to an antibiotic-sensitive cell of the same or another bacterial species, making the recipient cell antibiotic resistant. The extensive use of antibiotics

TABLE 24–1 The Sizes of DNA and Viral Particles for Some Bacterial Viruses (Bacteriophages)

Virus	*Size of viral DNA (bp)*	*Length of viral DNA (nm)*	*Long dimension of viral particle (nm)*
ϕX174	5,386	1,939	25
T7	39,936	14,377	78
λ (lambda)	48,502	17,460	190
T4	168,889	60,800	210

Note: Data on size of DNA are for the replicative form (double-stranded). The contour length is calculated assuming that each base pair occupies a length of 3.4 Å (see Fig. 8–15).

FIGURE 24–3 The length of the *E. coli* chromosome (1.7 mm) depicted in linear form relative to the length of a typical *E. coli* cell (2 μm).

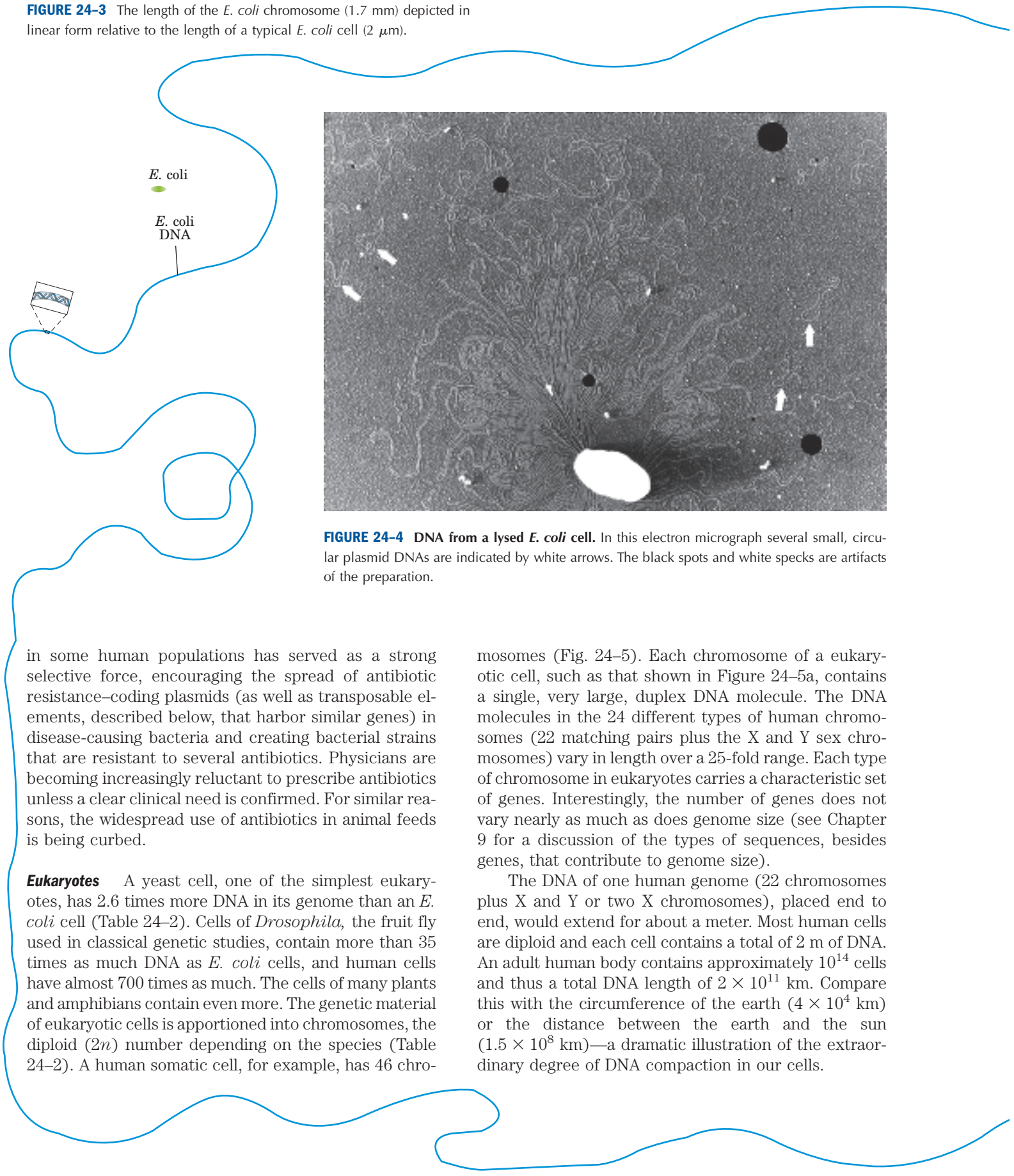
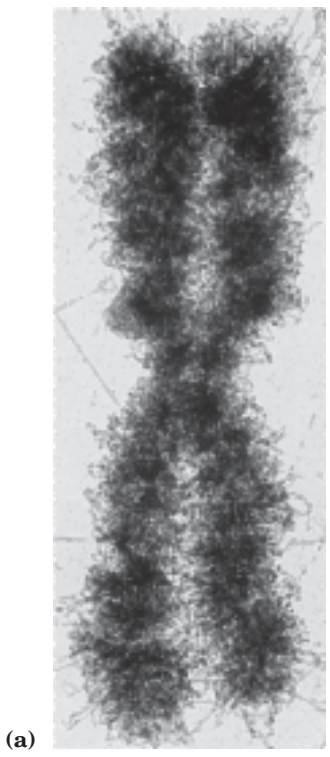


FIGURE 24–4 DNA from a lysed *E. coli* cell. In this electron micrograph several small, circular plasmid DNAs are indicated by white arrows. The black spots and white specks are artifacts of the preparation.

in some human populations has served as a strong selective force, encouraging the spread of antibiotic resistance–coding plasmids (as well as transposable elements, described below, that harbor similar genes) in disease-causing bacteria and creating bacterial strains that are resistant to several antibiotics. Physicians are becoming increasingly reluctant to prescribe antibiotics unless a clear clinical need is confirmed. For similar reasons, the widespread use of antibiotics in animal feeds is being curbed.

Eukaryotes A yeast cell, one of the simplest eukaryotes, has 2.6 times more DNA in its genome than an *E. coli* cell (Table 24–2). Cells of *Drosophila,* the fruit fly used in classical genetic studies, contain more than 35 times as much DNA as *E. coli* cells, and human cells have almost 700 times as much. The cells of many plants and amphibians contain even more. The genetic material of eukaryotic cells is apportioned into chromosomes, the diploid ($2n$) number depending on the species (Table 24–2). A human somatic cell, for example, has 46 chromosomes (Fig. 24–5). Each chromosome of a eukaryotic cell, such as that shown in Figure 24–5a, contains a single, very large, duplex DNA molecule. The DNA molecules in the 24 different types of human chromosomes (22 matching pairs plus the X and Y sex chromosomes) vary in length over a 25-fold range. Each type of chromosome in eukaryotes carries a characteristic set of genes. Interestingly, the number of genes does not vary nearly as much as does genome size (see Chapter 9 for a discussion of the types of sequences, besides genes, that contribute to genome size).

The DNA of one human genome (22 chromosomes plus X and Y or two X chromosomes), placed end to end, would extend for about a meter. Most human cells are diploid and each cell contains a total of 2 m of DNA. An adult human body contains approximately 10^{14} cells and thus a total DNA length of 2×10^{11} km. Compare this with the circumference of the earth (4×10^4 km) or the distance between the earth and the sun (1.5×10^8 km)—a dramatic illustration of the extraordinary degree of DNA compaction in our cells.



(a)

(b)

FIGURE 24–5 **Eukaryotic chromosomes.** **(a)** A pair of linked and condensed sister chromatids from a human chromosome. Eukaryotic chromosomes are in this state after replication and at metaphase during mitosis. **(b)** A complete set of chromosomes from a leukocyte from one of the authors. There are 46 chromosomes in every normal human somatic cell.

Eukaryotic cells also have organelles, mitochondria (Fig. 24–6) and chloroplasts, that contain DNA. Mitochondrial DNA (mtDNA) molecules are much smaller than the nuclear chromosomes. In animal cells, mtDNA contains fewer than 20,000 bp (16,569 bp in human mtDNA) and is a circular duplex. Each mitochondrion typically has two to ten copies of this mtDNA molecule, and the number can rise to hundreds in certain cells when an embryo is undergoing cell differentiation. In a few organisms (trypanosomes, for example) each mitochondrion contains thousands of copies of mtDNA, organized into a complex and interlinked matrix known as a kinetoplast. Plant cell mtDNA ranges in size from 200,000 to 2,500,000 bp. Chloroplast DNA (cpDNA) also exists as circular duplexes and ranges in size from 120,000 to 160,000 bp. The evolutionary origin of mitochondrial and chloroplast DNAs has been the subject of much speculation. A widely accepted view is that they are vestiges of the chromosomes of ancient bacteria that gained access to the cytoplasm of host cells and became the precursors of these organelles (see Fig. 1–36).

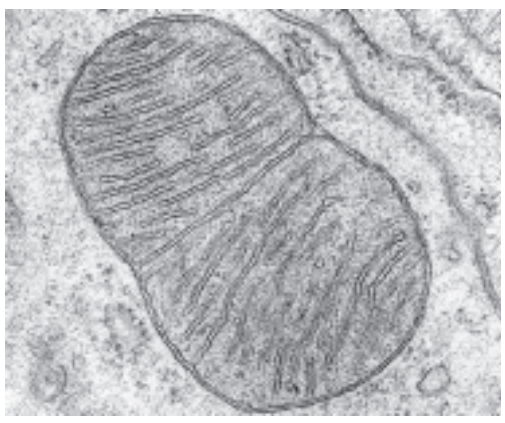


FIGURE 24–6 **A dividing mitochondrion.** Some mitochondrial proteins and RNAs are encoded by one of the copies of the mitochondrial DNA (none of which are visible here). The DNA (mtDNA) is replicated each time the mitochondrion divides, before cell division.

TABLE 24–2 DNA, Gene, and Chromosome Content in Some Genomes

	Total DNA (bp)	*Number of chromosomes*[*]	*Approximate number of genes*
Bacterium (*Escherichia coli*)	4,639,221	1	4,405
Yeast (*Saccharomyces cerevisiae*)	12,068,000	16[†]	6,200
Nematode (*Caenorhabditis elegans*)	97,000,000	12[‡]	19,000
Plant (*Arabidopsis thaliana*)	125,000,000	10	25,500
Fruit fly (*Drosophila melanogaster*)	180,000,000	18	13,600
Plant (*Oryza sativa*; rice)	480,000,000	24	57,000
Mouse (*Mus musculus*)	2,500,000,000	40	30,000–35,000
Human (*Homo sapiens*)	3,200,000,000	46	30,000–35,000

Note: This information is constantly being refined. For the most current information, consult the websites for the individual genome projects.
[*]The diploid chromosome number is given for all eukaryotes except yeast.
[†]Haploid chromosome number. Wild yeast strains generally have eight (octoploid) or more sets of these chromosomes.
[‡]Number for females, with two X chromosomes. Males have an X but no Y, thus 11 chromosomes in all.

Mitochondrial DNA codes for the mitochondrial tRNAs and rRNAs and for a few mitochondrial proteins. More than 95% of mitochondrial proteins are encoded by nuclear DNA. Mitochondria and chloroplasts divide when the cell divides. Their DNA is replicated before and during division, and the daughter DNA molecules pass into the daughter organelles.

Eukaryotic Genes and Chromosomes Are Very Complex

Many bacterial species have only one chromosome per cell and, in nearly all cases, each chromosome contains only one copy of each gene. A very few genes, such as those for rRNAs, are repeated several times. Genes and regulatory sequences account for almost all the DNA in prokaryotes. Moreover, almost every gene is precisely colinear with the amino acid sequence (or RNA sequence) for which it codes (Fig. 24–2).

The organization of genes in eukaryotic DNA is structurally and functionally much more complex. The study of eukaryotic chromosome structure, and more recently the sequencing of entire eukaryotic genomes, has yielded many surprises. Many, if not most, eukaryotic genes have a distinctive and puzzling structural feature: their nucleotide sequences contain one or more intervening segments of DNA that do not code for the amino acid sequence of the polypeptide product. These nontranslated inserts interrupt the otherwise colinear relationship between the nucleotide sequence of the gene and the amino acid sequence of the polypeptide it encodes. Such nontranslated DNA segments in genes are called **intervening sequences** or **introns,** and the coding segments are called **exons.** Few prokaryotic genes contain introns.

In higher eukaryotes, the typical gene has much more intron sequence than sequences devoted to exons. For example, in the gene coding for the single polypeptide chain of the avian egg protein ovalbumin (Fig. 24–7), the introns are much longer than the exons; altogether, seven introns make up 85% of the gene's DNA. In the gene for the β subunit of hemoglobin, a single intron contains more than half of the gene's DNA. The gene for the muscle protein titin is the intron champion, with 178 introns. Genes for histones appear to have no introns. In most cases the function of introns is not clear. In total, only about 1.5% of human DNA is "coding" or exon DNA, carrying information for protein or RNA products. However, when the much larger introns are included in the count, as much as 30% of the human genome consists of genes.

The relative paucity of genes in the human genome leaves a lot of DNA unaccounted for. Figure 24–8 provides a summary of sequence types. Much of the nongene DNA is in the form of repeated sequences of several kinds. Perhaps most surprising, about half the human genome is made up of moderately repeated sequences that are derived from transposable elements—segments of DNA, ranging from a few hundred to several thousand base pairs long, that can move from one location to another in the genome. Transposable elements (transposons) are a kind of molecular parasite, efficiently making a home within the host genome. Many have genes encoding proteins that catalyze the transposition process, described in more detail in Chapters 25 and 26. Some transposons in the human genome are active, moving at a low frequency, but most are inactive relics, evolutionarily altered by mutations. Although these elements generally do not encode proteins or RNAs that are used in human cells, they have played a

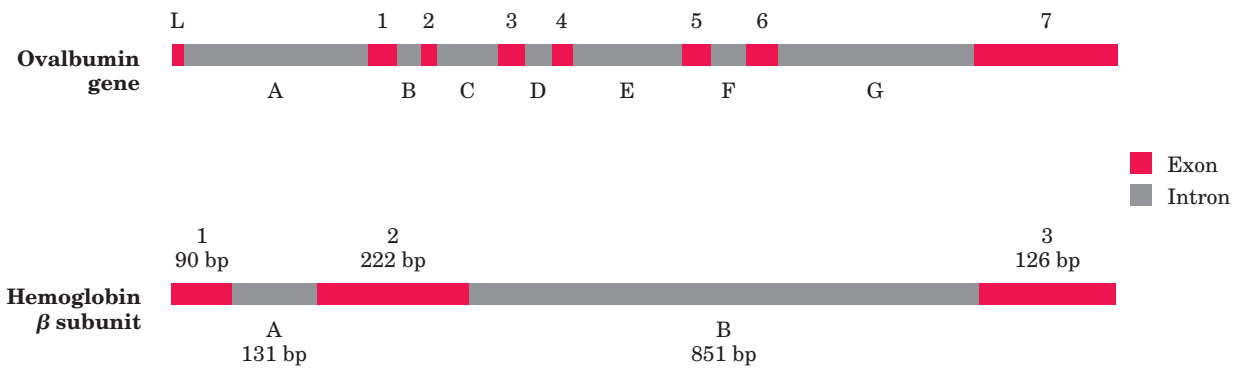


FIGURE 24–7 Introns in two eukaryotic genes. The gene for ovalbumin has seven introns (A to G), splitting the coding sequences into eight exons (L, and 1 to 7). The gene for the β subunit of hemoglobin has two introns and three exons, including one intron that alone contains more than half the base pairs of the gene.

major role in human evolution: movement of transposons can lead to the redistribution of other genomic sequences.

Another 3% or so of the human genome consists of **highly repetitive** sequences, also referred to as **simple-sequence DNA** or **simple sequence repeats (SSR).** These short sequences, generally less than 10 bp long, are sometimes repeated millions of times per cell. The simple-sequence DNA has also been called **satellite DNA,** so named because its unusual base composition often causes it to migrate as "satellite" bands (separated from the rest of the DNA) when fragmented cellular DNA samples are centrifuged in a cesium chloride density gradient. Studies suggest that simple-sequence DNA does not encode proteins or RNAs. Unlike the transposable elements, the highly repetitive DNA can have identifiable functional importance in human cellular metabolism, because much of it is associated with two defining features of eukaryotic chromosomes: centromeres and telomeres.

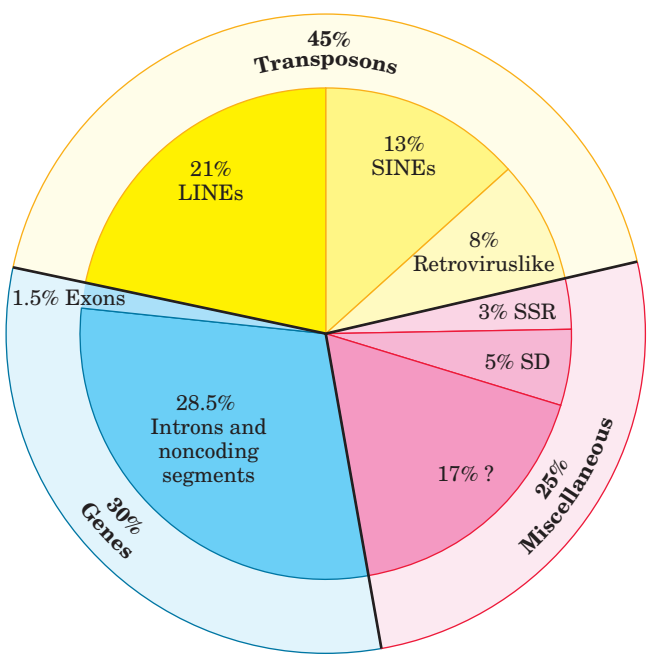


FIGURE 24–8 Types of sequences in the human genome. This pie chart divides the genome into transposons (transposable elements), genes, and miscellaneous sequences. There are four main classes of transposons. Long interspersed elements (LINEs), 6 to 8 kbp long (1 kbp = 1,000 bp), typically include a few genes encoding proteins that catalyze transposition. The genome has about 850,000 LINEs. Short interspersed elements (SINEs) are about 100 to 300 bp long. Of the 1.5 million in the human genome more than 1 million are Alu elements, so called because they generally include one copy of the recognition sequence for *Alu*I, a restriction endonuclease (see Fig. 9–3). The genome also contains 450,000 copies of retroviruslike transposons, 1.5 to 11 kbp long. Although these are "trapped" in the genome and cannot move from one cell to another, they are evolutionarily related to the retroviruses (Chapter 26), which include HIV. A final class of transposons (making up <1% and not shown here) consists of a variety of transposon remnants that differ greatly in length.

About 30% of the genome consists of sequences included in genes for proteins, but only a small fraction of this DNA is in exons (coding sequences). Miscellaneous sequences include simple-sequence repeats (SSR) and large segmental duplications (SD), the latter being segments that appear more than once in different locations. Among the unlisted sequence elements (denoted by a question mark) are genes encoding RNAs (which can be harder to identify than genes for proteins) and remnants of transposons that have been evolutionarily altered so that they are now hard to identify.

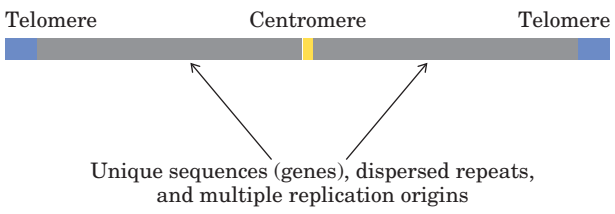


FIGURE 24–9 **Important structural elements of a yeast chromosome.**

The **centromere** (Fig. 24–9) is a sequence of DNA that functions during cell division as an attachment point for proteins that link the chromosome to the mitotic spindle. This attachment is essential for the equal and orderly distribution of chromosome sets to daughter cells. The centromeres of *Saccharomyces cerevisiae* have been isolated and studied. The sequences essential to centromere function are about 130 bp long and are very rich in A═T pairs. The centromeric sequences of higher eukaryotes are much longer and, unlike those of yeast, generally contain simple-sequence DNA, which consists of thousands of tandem copies of one or a few short sequences of 5 to 10 bp, in the same orientation. The precise role of simple-sequence DNA in centromere function is not yet understood.

Telomeres (Greek *telos*, "end") are sequences at the ends of eukaryotic chromosomes that help stabilize the chromosome. The best-characterized telomeres are those of the simpler eukaryotes. Yeast telomeres end with about 100 bp of imprecisely repeated sequences of the form

$$(5')(T_xG_y)_n$$
$$(3')(A_xC_y)_n$$

where x and y are generally between 1 and 4. The number of telomere repeats, n, is in the range of 20 to 100 for most single-celled eukaryotes and generally more than 1,500 in mammals. The ends of a linear DNA molecule cannot be routinely replicated by the cellular replication machinery (which may be one reason why bacterial DNA molecules are circular). Repeated telomeric sequences are added to eukaryotic chromosome ends primarily by the enzyme telomerase (see Fig. 26–35).

Artificial chromosomes (Chapter 9) have been constructed as a means of better understanding the functional significance of many structural features of eukaryotic chromosomes. A reasonably stable artificial linear chromosome requires only three components: a centromere, telomeres at each end, and sequences that allow the initiation of DNA replication. Yeast artificial chromosomes (YACs; see Fig. 9–8) have been developed as a research tool in biotechnology. Similarly, human artificial chromosomes (HACs) are being developed for the treatment of genetic diseases by somatic gene therapy.

SUMMARY 24.1 Chromosomal Elements

- Genes are segments of a chromosome that contain the information for a functional polypeptide or RNA molecule. In addition to genes, chromosomes contain a variety of regulatory sequences involved in replication, transcription, and other processes.
- Genomic DNA and RNA molecules are generally orders of magnitude longer than the viral particles or cells that contain them.
- Many genes in eukaryotic cells, and a few in bacteria, are interrupted by noncoding sequences called introns. The coding segments separated by introns are called exons.
- Less than one-third of human genomic DNA consists of genes. Much of the remainder consists of repeated sequences of various types. Nucleic acid parasites known as transposons account for about half of the human genome.
- Eukaryotic chromosomes have two important special-function repetitive DNA sequences: centromeres, which are attachment points for the mitotic spindle, and telomeres, located at the ends of chromosomes.

24.2 DNA Supercoiling

Cellular DNA, as we have seen, is extremely compacted, implying a high degree of structural organization. The folding mechanism must not only pack the DNA but also permit access to the information in the DNA. Before considering how this is accomplished in processes such as replication and transcription, we need to examine an important property of DNA structure known as **supercoiling.**

Supercoiling means the coiling of a coil. A telephone cord, for example, is typically a coiled wire. The path taken by the wire between the base of the phone and the receiver often includes one or more supercoils (Fig. 24–10). DNA is coiled in the form of a double helix, with both strands of the DNA coiling around an axis. The further coiling of that axis upon itself (Fig. 24–11) produces DNA supercoiling. As detailed below, DNA supercoiling is generally a manifestation of structural strain. When there is no net bending of the DNA axis upon itself, the DNA is said to be in a **relaxed** state.

We might have predicted that DNA compaction involved some form of supercoiling. Perhaps less predictable is that replication and transcription of DNA also affect and are affected by supercoiling. Both processes

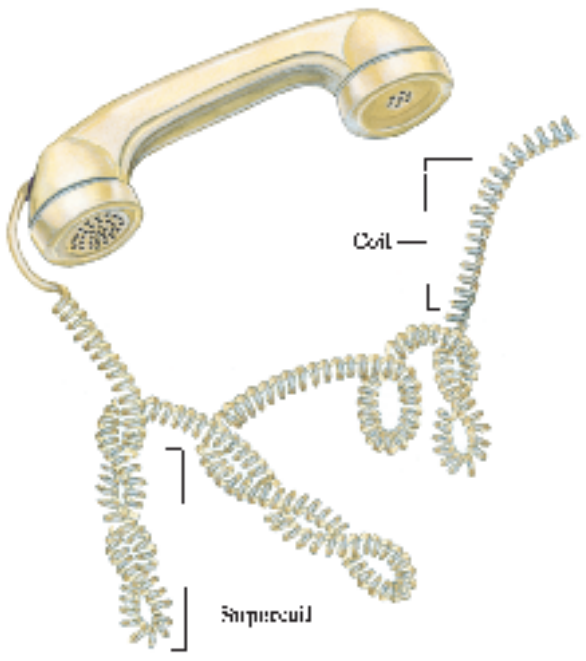


FIGURE 24–10 **Supercoils.** A typical phone cord is coiled like a DNA helix, and the coiled cord can itself coil in a supercoil. The illustration is especially appropriate because an examination of phone cords helped lead Jerome Vinograd and his colleagues to the insight that many properties of small circular DNAs can be explained by supercoiling. They first detected DNA supercoiling, in small circular viral DNAs, in 1965.

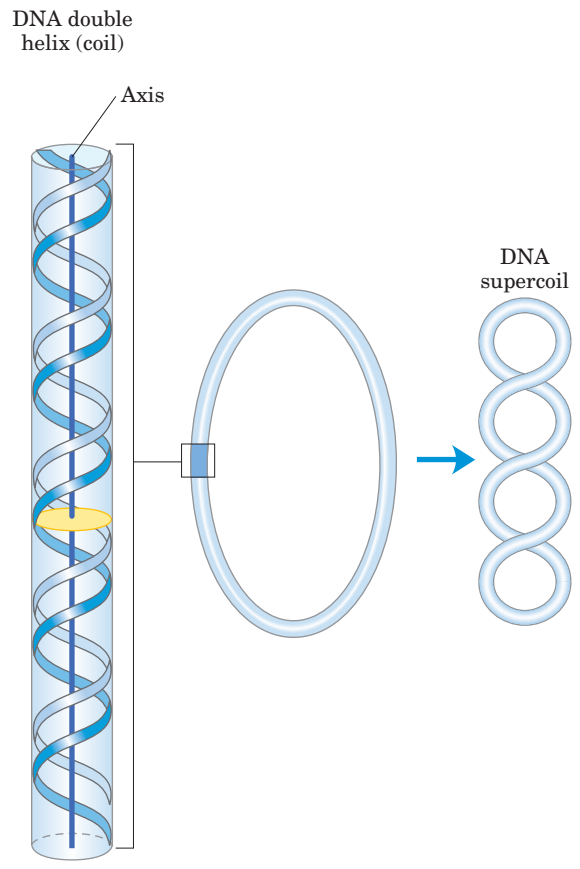


FIGURE 24–11 **Supercoiling of DNA.** When the axis of the DNA double helix is coiled on itself, it forms a new helix (superhelix). The DNA superhelix is usually called a supercoil.

require a separation of DNA strands—a process complicated by the helical interwinding of the strands (as demonstrated in Fig. 24–12).

That DNA would bend on itself and become supercoiled in tightly packaged cellular DNA would seem logical, then, and perhaps even trivial, were it not for one additional fact: many circular DNA molecules remain highly supercoiled even after they are extracted and purified, freed from protein and other cellular components. This indicates that supercoiling is an intrinsic property of DNA tertiary structure. It occurs in all cellular DNAs and is highly regulated by each cell.

A number of measurable properties of supercoiling have been established, and the study of supercoiling has provided many insights into DNA structure and function. This work has drawn heavily on concepts derived from a branch of mathematics called **topology,** the study of the properties of an object that do not change under continuous deformations. For DNA, continuous deformations include conformational changes due to thermal motion or an interaction with proteins or other molecules; discontinuous deformations involve DNA strand breakage. For circular DNA molecules, a topological property is one that is unaffected by deformations

FIGURE 24–12 **Supercoiling induced by separating the strands of a helical structure.** Twist two linear strands of rubber band into a right-handed double helix as shown. Fix one end by having a friend hold onto it, then pull apart the two strands at the other end. The resulting strain will produce supercoiling.

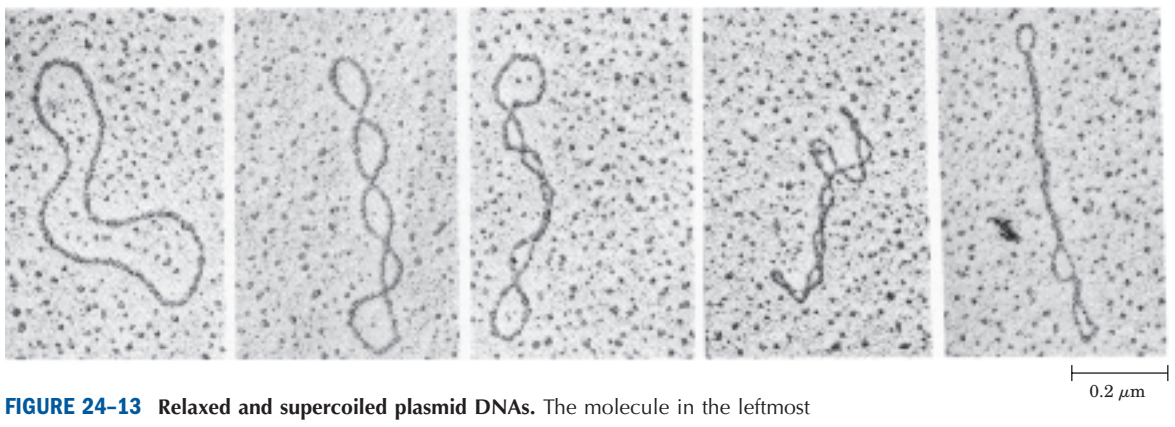


FIGURE 24–13 Relaxed and supercoiled plasmid DNAs. The molecule in the leftmost electron micrograph is relaxed; the degree of supercoiling increases from left to right.

of the DNA strands as long as no breaks are introduced. Topological properties are changed only by breakage and rejoining of the backbone of one or both DNA strands.

We now examine the fundamental properties and physical basis of supercoiling.

Most Cellular DNA Is Underwound

To understand supercoiling we must first focus on the properties of small circular DNAs such as plasmids and small viral DNAs. When these DNAs have no breaks in either strand, they are referred to as **closed-circular DNAs.** If the DNA of a closed-circular molecule conforms closely to the B-form structure (the Watson-Crick structure; see Fig. 8–15), with one turn of the double helix per 10.5 bp, the DNA is relaxed rather than supercoiled (Fig. 24–13). Supercoiling results when DNA is subject to some form of structural strain. Purified closed-circular DNA is rarely relaxed, regardless of its biological origin. Furthermore, DNAs derived from a given cellular source have a characteristic degree of supercoiling. DNA structure is therefore strained in a manner that is regulated by the cell to induce the supercoiling.

In almost every instance, the strain is a result of **underwinding** of the DNA double helix in the closed circle. In other words, the DNA has *fewer* helical turns than would be expected for the B-form structure. The effects of underwinding are summarized in Figure 24–14. An 84 bp segment of a circular DNA in the relaxed state would contain eight double-helical turns, or one for every 10.5 bp. If one of these turns were removed, there would be (84 bp)/7 = 12.0 bp per turn, rather than the 10.5 found in B-DNA (Fig. 24–14b). This is a deviation from the most stable DNA form, and the molecule is thermodynamically strained as a result. Generally, much of this strain would be accommodated by coiling the axis of the DNA on itself to form a supercoil (Fig. 24–14c; some of the strain in this 84 bp segment would simply become dispersed in the untwisted structure of the larger DNA molecule). In principle, the strain could also be accommodated by separating the two DNA strands over a distance of about 10 bp (Fig. 24–14d). In isolated closed-circular DNA, strain introduced by underwinding is generally accommodated by supercoiling rather than strand separation, because coiling the axis of the DNA usually requires less energy than breaking the hydrogen bonds that stabilize paired bases. Note, however, that the underwinding of DNA in vivo makes

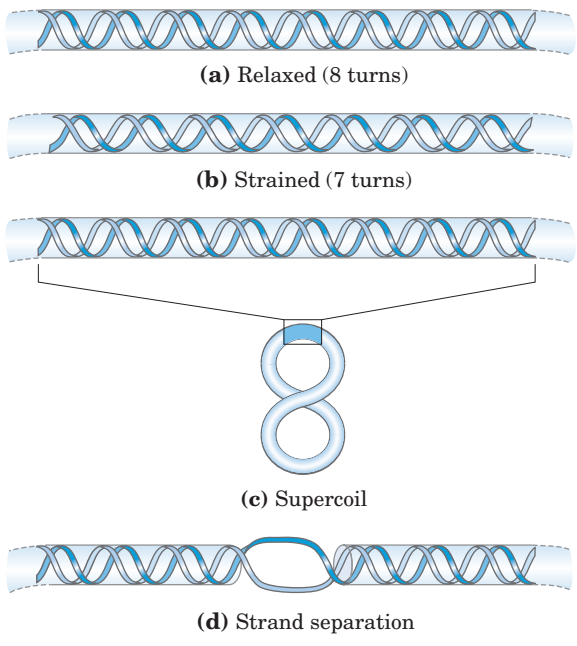


FIGURE 24–14 Effects of DNA underwinding. (a) A segment of DNA within a closed-circular molecule, 84 bp long, in its relaxed form with eight helical turns. **(b)** Removal of one turn induces structural strain. **(c)** The strain is generally accommodated by formation of a supercoil. **(d)** DNA underwinding also makes the separation of strands somewhat easier. In principle, each turn of underwinding should facilitate strand separation over about 10 bp, as shown. However, the hydrogen-bonded base pairs would generally preclude strand separation over such a short distance, and the effect becomes important only for longer DNAs and higher levels of DNA underwinding.

it easier to separate DNA strands, giving access to the information they contain.

Every cell actively underwinds its DNA with the aid of enzymatic processes (described below), and the resulting strained state represents a form of stored energy. Cells maintain DNA in an underwound state to facilitate its compaction by coiling. The underwinding of DNA is also important to enzymes of DNA metabolism that must bring about strand separation as part of their function.

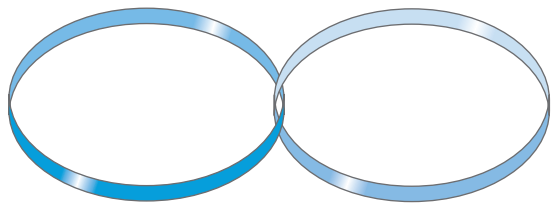
The underwound state can be maintained only if the DNA is a closed circle or if it is bound and stabilized by proteins so that the strands are not free to rotate about each other. If there is a break in one strand of an isolated, protein-free circular DNA, free rotation at that point will cause the underwound DNA to revert spontaneously to the relaxed state. In a closed-circular DNA molecule, however, the number of helical turns cannot be changed without at least transiently breaking one of the DNA strands. The number of helical turns in a DNA molecule therefore provides a precise description of supercoiling.

DNA Underwinding Is Defined by Topological Linking Number

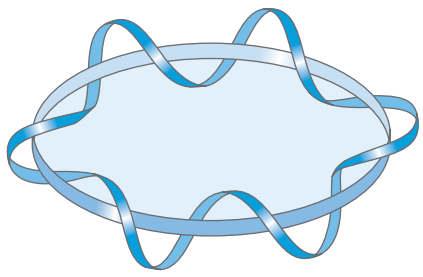
The field of topology provides a number of ideas that are useful to this discussion, particularly the concept of **linking number.** Linking number is a topological property of double-stranded DNA, because it does not vary when the DNA is bent or deformed, as long as both DNA strands remain intact. Linking number (Lk) is illustrated in Figure 24–15.

Let's begin by visualizing the separation of the two strands of a double-stranded circular DNA. If the two strands are linked as shown in Figure 24–15a, they are effectively joined by what can be described as a topological bond. Even if all hydrogen bonds and base-stacking interactions were abolished such that the strands were not in physical contact, this topological bond would still link the two strands. Visualize one of the circular strands as the boundary of a surface (such as a soap film spanning the space framed by a circular wire before you blow a soap bubble). The linking number can be defined as the number of times the second strand pierces this surface. For the molecule in Figure 24–15a, $Lk = 1$; for that in Figure 24–15b, $Lk = 6$. The linking number for a closed-circular DNA is always an integer. By convention, if the links between two DNA strands are arranged so that the strands are interwound in a right-handed helix, the linking number is defined as positive (+); for strands interwound in a left-handed helix, the linking number is negative (−). Negative linking numbers are, for all practical purposes, not encountered in DNA.

We can now extend these ideas to a closed-circular DNA with 2,100 bp (Fig. 24–16a). When the molecule is relaxed, the linking number is simply the number of base pairs divided by the number of base pairs per turn, which is close to 10.5; so in this case, $Lk = 200$. For a circular DNA molecule to have a topological property such as linking number, neither strand may contain a break. If there is a break in either strand, the strands can, in principle, be unraveled and separated completely. In this case, no topological bond exists and Lk is undefined (Fig. 24–16b).



(a) $Lk = 1$



(b) $Lk = 6$

FIGURE 24–15 Linking number, *Lk*. Here, as usual, each blue ribbon represents one strand of a double-stranded DNA molecule. For the molecule in **(a),** $Lk = 1$. For the molecule in **(b),** $Lk = 6$. One of the strands in **(b)** is kept untwisted for illustrative purposes, to define the border of an imaginary surface (shaded blue). The number of times the twisting strand penetrates this surface provides a rigorous definition of linking number.

We can now describe DNA underwinding in terms of changes in the linking number. The linking number in relaxed DNA, Lk_0, is used as a reference. For the molecule shown in Figure 24–16a, $Lk_0 = 200$; if two turns are removed from this molecule, $Lk = 198$. The change can be described by the equation

$$\Delta Lk = Lk - Lk_0 = 198 - 200 = -2$$

It is often convenient to express the change in linking number in terms of a quantity that is independent of the length of the DNA molecule. This quantity, called the **specific linking difference (σ),** or **superhelical density,** is a measure of the number of turns removed relative to the number present in relaxed DNA:

$$\sigma = \frac{\Delta Lk}{Lk_0}$$

In the example in Figure 24–16c, $\sigma = -0.01$, which means that 1% (2 of 200) of the helical turns present

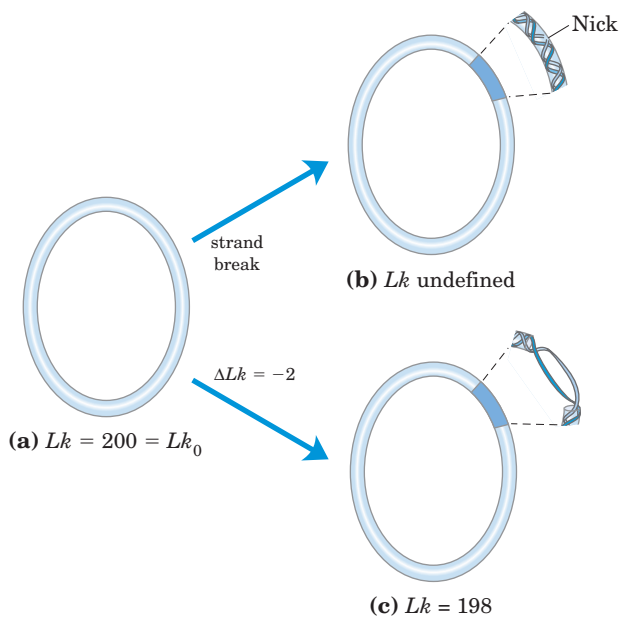


FIGURE 24–16 Linking number applied to closed-circular DNA molecules. A 2,100 bp circular DNA is shown in three forms: **(a)** relaxed, $Lk = 200$; **(b)** relaxed with a nick (break) in one strand, Lk undefined; and **(c)** underwound by two turns, $Lk = 198$. The underwound molecule generally exists as a supercoiled molecule, but underwinding also facilitates the separation of DNA strands.

in the DNA (in its B form) have been removed. The degree of underwinding in cellular DNAs generally falls in the range of 5% to 7%; that is, $\sigma = -0.05$ to -0.07. The negative sign indicates that the change in linking number is due to underwinding of the DNA. The supercoiling induced by underwinding is therefore defined as negative supercoiling. Conversely, under some conditions DNA can be overwound, resulting in positive supercoiling. Note that the twisting path taken by the axis of the DNA helix when the DNA is underwound (negative supercoiling) is the mirror image of that taken when the DNA is overwound (positive supercoiling) (Fig. 24–17). Supercoiling is not a random process; the path of the supercoiling is largely prescribed by the torsional strain imparted to the DNA by decreasing or increasing the linking number relative to B-DNA.

Linking number can be changed by ±1 by breaking one DNA strand, rotating one of the ends 360° about the unbroken strand, and rejoining the broken ends. This change has no effect on the number of base pairs or the number of atoms in the circular DNA molecule. Two forms of a circular DNA that differ only in a topological property such as linking number are referred to as **topoisomers.**

Linking number can be broken down into two structural components called writhe (Wr) and twist (Tw) (Fig. 24–18). These are more difficult to describe than linking number, but writhe may be thought of as a measure of the coiling of the helix axis and twist as determining the local twisting or spatial relationship of neighboring base pairs. When the linking number changes, some of the resulting strain is usually compensated for by writhe (supercoiling) and some by changes in twist, giving rise to the equation

$$Lk = Tw + Wr$$

Tw and Wr need not be integers. Twist and writhe are geometric rather than topological properties, because they may be changed by deformation of a closed-circular DNA molecule.

In addition to causing supercoiling and making strand separation somewhat easier, the underwinding of

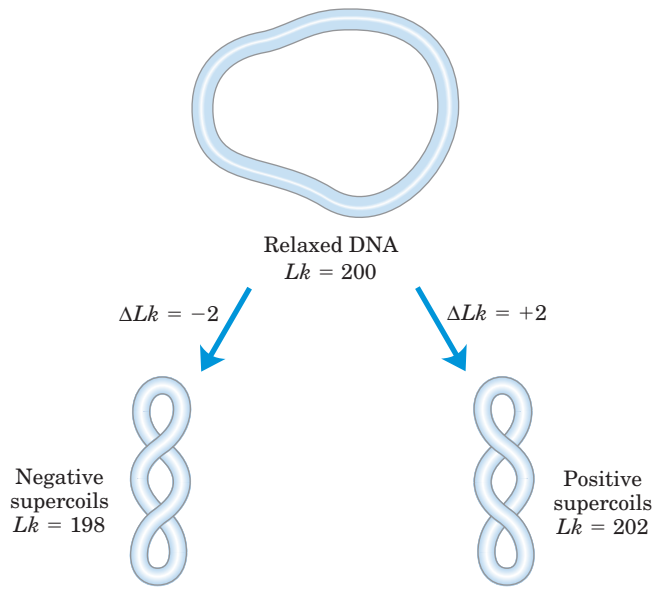


FIGURE 24–17 Negative and positive supercoils. For the relaxed DNA molecule of Figure 24–16a, underwinding or overwinding by two helical turns ($Lk = 198$ or 202) will produce negative or positive supercoiling, respectively. Note that the DNA axis twists in opposite directions in the two cases.

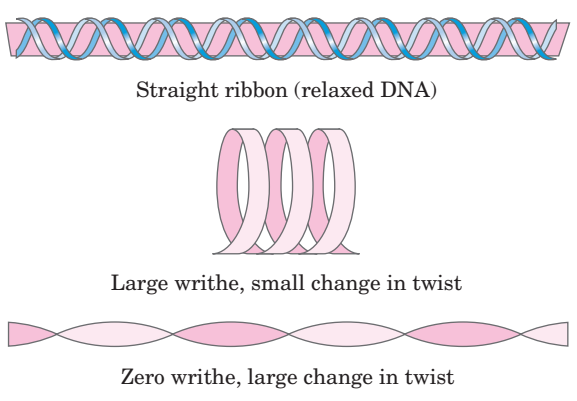


FIGURE 24–18 Ribbon model for illustrating twist and writhe. The pink ribbon represents the axis of a relaxed DNA molecule. Strain introduced by twisting the ribbon (underwinding the DNA) can be manifested as writhe or twist. Changes in linking number are usually accompanied by changes in both writhe and twist.

DNA facilitates a number of structural changes in the molecule. These are of less physiological importance but help illustrate the effects of underwinding. Recall that a cruciform (see Fig. 8–21) generally contains a few unpaired bases; DNA underwinding helps to maintain the required strand separation (Fig. 24–19). Underwinding of a right-handed DNA helix also facilitates the formation of short stretches of left-handed Z-DNA in regions where the base sequence is consistent with the Z form (Chapter 8).

Topoisomerases Catalyze Changes in the Linking Number of DNA

DNA supercoiling is a precisely regulated process that influences many aspects of DNA metabolism. Every cell has enzymes with the sole function of underwinding and/or relaxing DNA. The enzymes that increase or decrease the extent of DNA underwinding are **topoisomerases;** the property of DNA that they change is the linking number. These enzymes play an especially important role in processes such as replication and DNA packaging. There are two classes of topoisomerases. **Type I topoisomerases** act by transiently breaking one of the two DNA strands, passing the unbroken strand through the break, and rejoining the broken ends; they change *Lk* in increments of 1. **Type II topoisomerases** break both DNA strands and change *Lk* in increments of 2.

The effects of these enzymes can be demonstrated using agarose gel electrophoresis (Fig. 24–20). A population of identical plasmid DNAs with the same linking number migrates as a discrete band during electrophoresis. Topoisomers with *Lk* values differing by as little as 1 can be separated by this method, so changes in linking number induced by topoisomerases are readily detected.

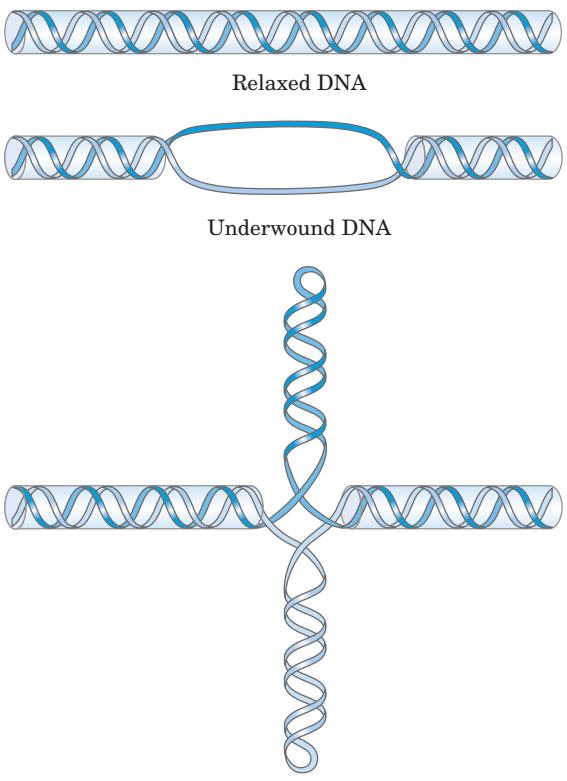


FIGURE 24–19 Promotion of cruciform structures by DNA underwinding. In principle, cruciforms can form at palindromic sequences (see Fig. 8–21), but they seldom occur in relaxed DNA because the linear DNA accommodates more paired bases than does the cruciform structure. Underwinding of the DNA facilitates the partial strand separation needed to promote cruciform formation at appropriate sequences.

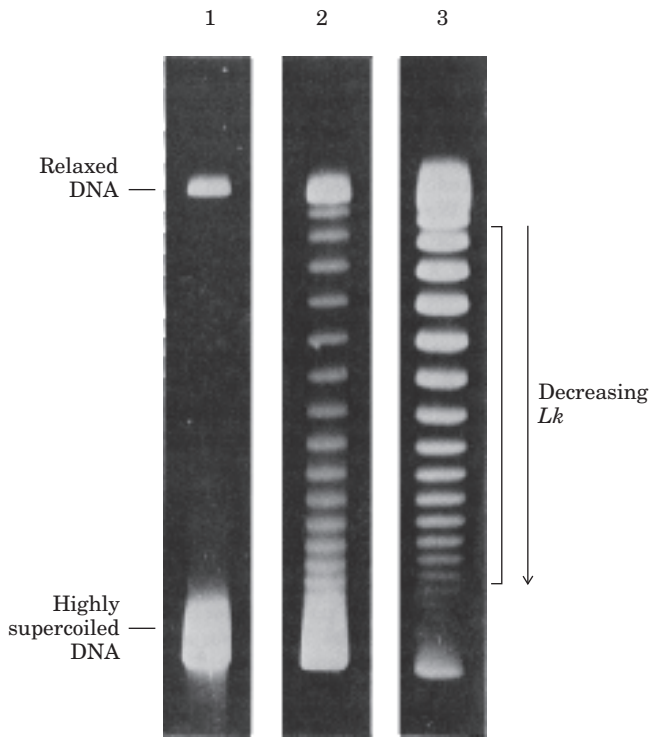
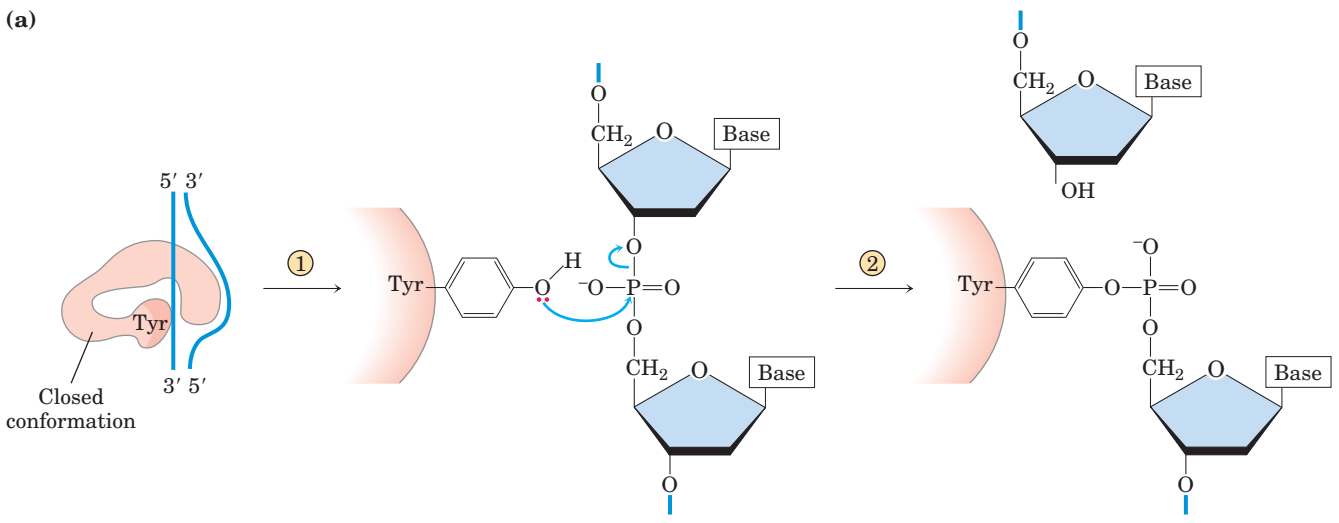


FIGURE 24–20 Visualization of topoisomers. In this experiment, all DNA molecules have the same number of base pairs but exhibit some range in the degree of supercoiling. Because supercoiled DNA molecules are more compact than relaxed molecules, they migrate more rapidly during gel electrophoresis. The gels shown here separate topoisomers (moving from top to bottom) over a limited range of superhelical density. In lane 1, highly supercoiled DNA migrates in a single band, even though different topoisomers are probably present. Lanes 2 and 3 illustrate the effect of treating the supercoiled DNA with a type I topoisomerase; the DNA in lane 3 was treated for a longer time than that in lane 2. As the superhelical density of the DNA is reduced to the point where it corresponds to the range in which the gel can resolve individual topoisomers, distinct bands appear. Individual bands in the region indicated by the bracket next to lane 3 each contain DNA circles with the same linking number; the linking number changes by 1 from one band to the next.



After DNA binds (step ①), an active-site Tyr attacks a phosphodiester bond on one DNA strand in step ②, cleaving it, creating a covalent 5′-Ⓟ–Tyr protein-DNA linkage, and liberating the 3′-hydroxyl group of the adjacent nucleotide.

In step ③ the enzyme switches to its open conformation, and the unbroken DNA strand passes through the break in the first strand.

With the enzyme in the closed conformation, the liberated 3′-hydroxyl group attacks the 5′-Ⓟ–Tyr protein-DNA linkage in step ④ to religate the cleaved DNA strand.

MECHANISM FIGURE 24–21 Bacterial type I topoisomerases alter linking number. A proposed reaction sequence for the bacterial topoisomerase I is illustrated. The enzyme has closed and open conformations. **(a)** A DNA molecule binds to the closed conformation and one DNA strand is cleaved. **(b)** The enzyme changes to its open conformation, and the other DNA strand moves through the break in the first strand. **(c)** In the closed conformation, the DNA strand is religated.

E. coli has at least four different individual topoisomerases (I through IV). Those of type I (topoisomerases I and III) generally relax DNA by removing negative supercoils (increasing *Lk*). The way in which bacterial type I topoisomerases change linking number is illustrated in Figure 24–21. A bacterial type II enzyme, called either topoisomerase II or DNA gyrase, can introduce negative supercoils (decrease *Lk*). It uses the energy of ATP to accomplish this. To alter DNA linking number, type II topoisomerases cleave both strands of a DNA molecule and pass another duplex through the break. The degree of supercoiling of bacterial DNA is maintained by regulation of the net activity of topoisomerases I and II.

Eukaryotic cells also have type I and type II topoisomerases. The type I enzymes are topoisomerases I and III; the type II enzymes are topoisomerases IIα and IIβ. The eukaryotic type II topoisomerases cannot underwind DNA (introduce negative supercoils), but they can relax both positive and negative supercoils. We consider one probable origin of negative supercoils in eukaryotic cells in our discussion of chromatin in Section 24.3. The process catalyzed by eukaryotic type II topoisomerases is illustrated in Figure 24–22.

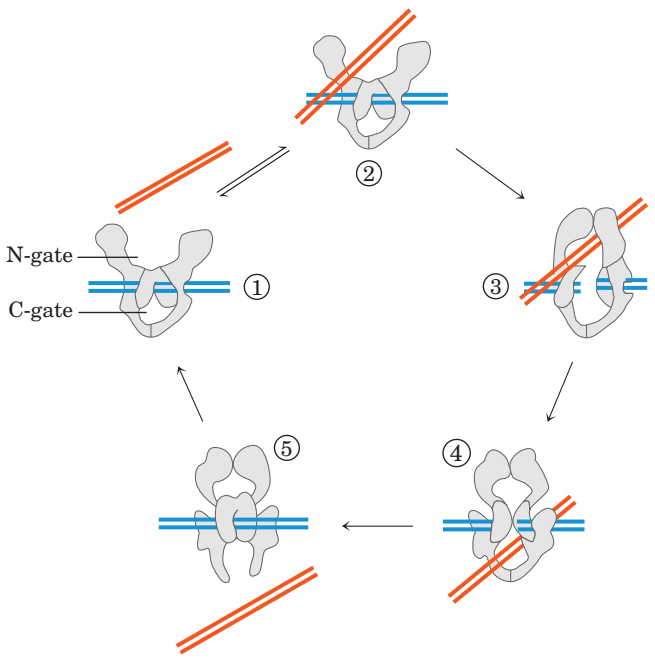


FIGURE 24–22 Proposed mechanism for the alteration of linking number by eukaryotic type IIA topoisomerases. ① The multisubunit enzyme binds one DNA molecule (blue). Gated cavities above and below the bound DNA are called the N-gate and the C-gate. ② A second segment of the same DNA molecule (red) is bound at the N-gate and ③ trapped. Both strands of the first DNA are now cleaved (the chemistry is similar to that in Fig. 24–20b), and ④ the second DNA segment is passed through the break. ⑤ The broken DNA is religated, and the second DNA segment is released through the C-gate. Two ATPs are bound and hydrolyzed during this cycle; it is likely that one is hydrolyzed in the step leading to the complex in step ④. Additional details of the ATP hydrolysis component of the reaction remain to be worked out.

DNA Compaction Requires a Special Form of Supercoiling

Supercoiled DNA molecules are uniform in a number of respects. The supercoils are right-handed in a negatively supercoiled DNA molecule (Fig. 24–17), and they tend to be extended and narrow rather than compacted, often with multiple branches (Fig. 24–23). At the superhelical densities normally encountered in cells, the length of the supercoil axis, including branches, is about 40% of the length of the DNA. This type of supercoiling is referred to as **plectonemic** (from the Greek *plektos,* "twisted," and *nema,* "thread"). This term can be applied to any structure with strands intertwined in some simple and regular way, and it is a good description of the general structure of supercoiled DNA in solution.

Plectonemic supercoiling, the form observed in isolated DNAs in the laboratory, does not produce sufficient compaction to package DNA in the cell. A second form of supercoiling, **solenoidal** (Fig. 24–24), can be adopted by an underwound DNA. Instead of the

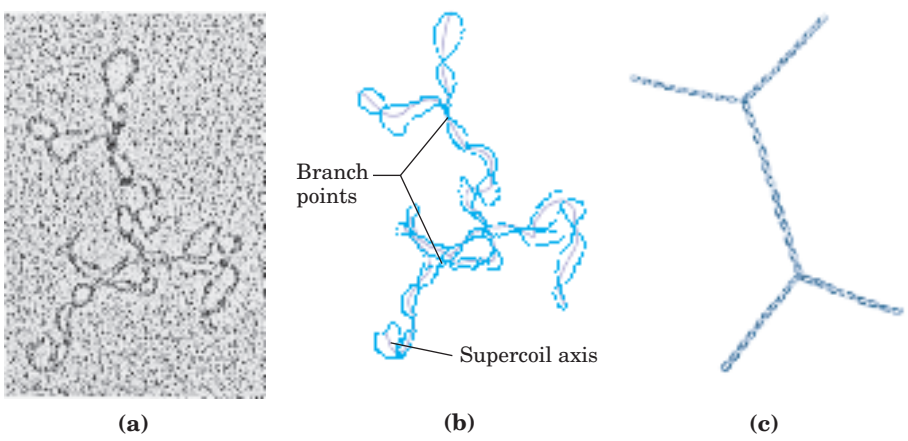


FIGURE 24–23 Plectonemic supercoiling. **(a)** Electron micrograph of plectonemically supercoiled plasmid DNA and **(b)** an interpretation of the observed structure. The purple lines show the axis of the supercoil; note the branching of the supercoil. **(c)** An idealized representation of this structure.

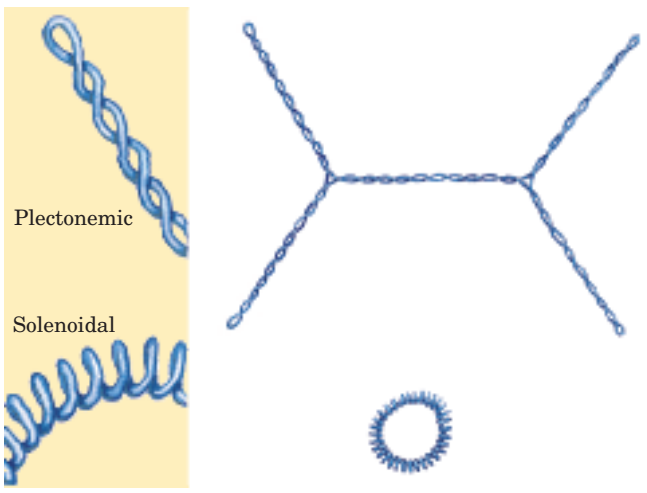


FIGURE 24–24 Plectonemic and solenoidal supercoiling. (a) Plectonemic supercoiling takes the form of extended right-handed coils. Solenoidal negative supercoiling takes the form of tight left-handed turns about an imaginary tubelike structure. The two forms are readily interconverted, although the solenoidal form is generally not observed unless certain proteins are bound to the DNA. **(b)** Plectonemic (top) and solenoidal supercoiling of the same DNA molecule, drawn to scale. Solenoidal supercoiling provides a much greater degree of compaction.

extended right-handed supercoils characteristic of the plectonemic form, solenoidal supercoiling involves tight left-handed turns, similar to the shape taken up by a garden hose neatly wrapped on a reel. Although their structures are dramatically different, plectonemic and solenoidal supercoiling are two forms of negative supercoiling that can be taken up by the *same* segment of underwound DNA. The two forms are readily interconvertible. Although the plectonemic form is more stable in solution, the solenoidal form can be stabilized by protein binding and is the form found in chromatin. It provides a much greater degree of compaction (Fig. 24–24b). Solenoidal supercoiling is the mechanism by which underwinding contributes to DNA compaction.

SUMMARY 24.2 DNA Supercoiling

- Most cellular DNAs are supercoiled. Underwinding decreases the total number of helical turns in the DNA relative to the relaxed, B form. To maintain an underwound state, DNA must be either a closed circle or bound to protein. Underwinding is quantified by a topological parameter called linking number, *Lk*.
- Underwinding is measured in terms of specific linking difference, σ (also called superhelical density), which is $(Lk - Lk_0)/Lk_0$. For cellular DNAs, σ is typically -0.05 to -0.07, which means that approximately 5% to 7% of the helical turns in the DNA have been removed. DNA underwinding facilitates strand separation by enzymes of DNA metabolism.
- DNAs that differ only in linking number are called topoisomers. Enzymes that underwind and/or relax DNA, the topoisomerases, catalyze changes in linking number. The two classes of topoisomerases, type I and type II, change *Lk* in increments of 1 or 2, respectively, per catalytic event.

24.3 The Structure of Chromosomes

The term "chromosome" is used to refer to a nucleic acid molecule that is the repository of genetic information in a virus, a bacterium, a eukaryotic cell, or an organelle. It also refers to the densely colored bodies seen in the nuclei of dye-stained eukaryotic cells, as visualized using a light microscope.

Chromatin Consists of DNA and Proteins

The eukaryotic cell cycle (see Fig. 12–41) produces remarkable changes in the structure of chromosomes (Fig. 24–25). In nondividing eukaryotic cells (in G0) and those in interphase (G1, S, and G2), the chromosomal material, **chromatin,** is amorphous and appears to be randomly dispersed in certain parts of the nucleus. In the S phase of interphase the DNA in this amorphous state replicates, each chromosome producing two sister chromosomes (called sister chromatids) that remain associated with each other after replication is complete. The chromosomes become much more condensed during prophase of mitosis, taking the form of a species-specific number of well-defined pairs of sister chromatids (Fig. 24–5).

Chromatin consists of fibers containing protein and DNA in approximately equal masses, along with a small amount of RNA. The DNA in the chromatin is very tightly associated with proteins called **histones,** which package and order the DNA into structural units called **nucleosomes** (Fig. 24–26). Also found in chromatin are many nonhistone proteins, some of which help maintain chromosome structure, others that regulate the expression of specific genes (Chapter 28). Beginning with nucleosomes, eukaryotic chromosomal DNA is packaged into a succession of higher-order structures that ultimately yield the compact chromosome seen with the light microscope. We now turn to a description of this structure in eukaryotes and compare it with the packaging of DNA in bacterial cells.

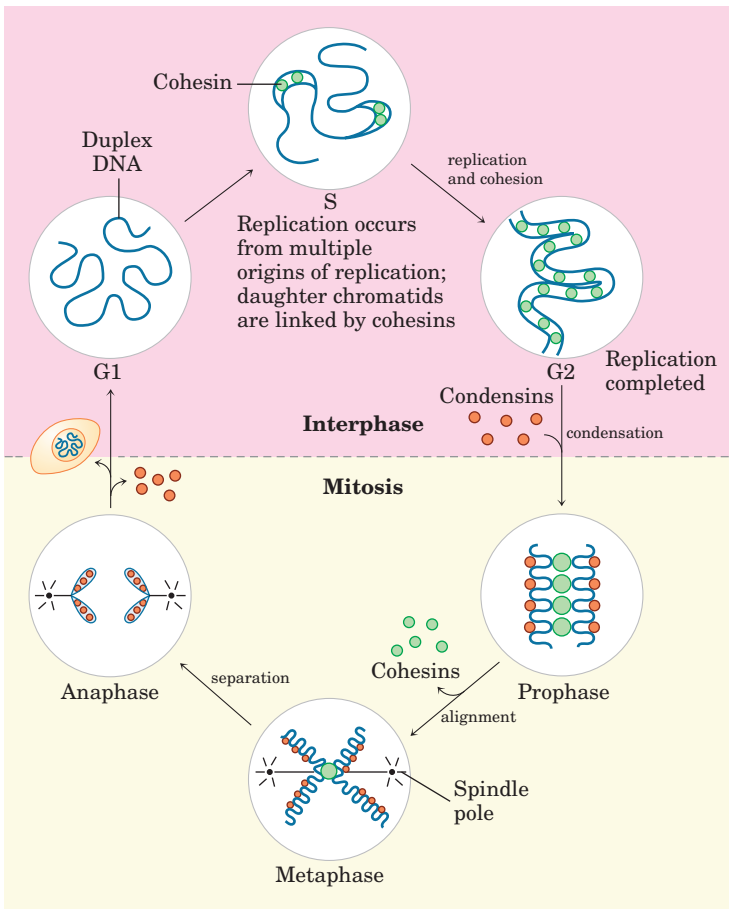


FIGURE 24–25 Changes in chromosome structure during the eukaryotic cell cycle. Cellular DNA is uncondensed throughout interphase. The interphase period can be subdivided (see Fig. 12–41) into the G1 (gap) phase; the S (synthesis) phase, when the DNA is replicated; and the G2 phase, in which the replicated chromosomes cohere to one another. The DNA undergoes condensation in the prophase of mitosis. Cohesins (green) and condensins (red) are proteins involved in cohesion and condensation (discussed later in the chapter). The architecture of the cohesin-condensin-DNA complex is not yet established, and the interactions shown here are figurative, simply suggesting their role in condensation of the chromosome. During metaphase, the condensed chromosomes line up along a plane halfway between the spindle poles. One chromosome of each pair is linked to each spindle pole via microtubules that extend between the spindle and the centromere. The sister chromatids separate at anaphase, each drawn toward the spindle pole to which it is connected. After cell division is complete, the chromosomes decondense and the cycle begins anew.

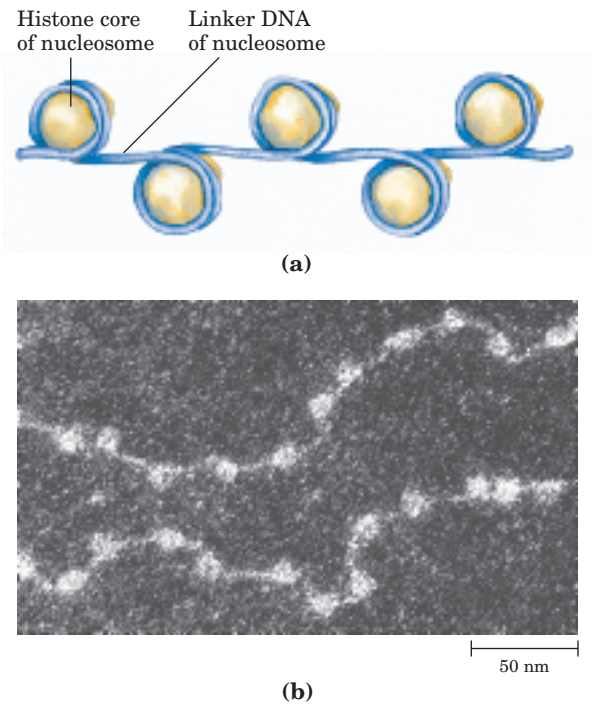


FIGURE 24–26 Nucleosomes. Regularly spaced nucleosomes consist of histone complexes bound to DNA. **(a)** Schematic illustration and **(b)** electron micrograph.

Histones Are Small, Basic Proteins

Found in the chromatin of all eukaryotic cells, histones have molecular weights between 11,000 and 21,000 and are very rich in the basic amino acids arginine and lysine (together these make up about one-fourth of the amino acid residues). All eukaryotic cells have five major classes of histones, differing in molecular weight and amino acid composition (Table 24–3). The H3 histones are nearly identical in amino acid sequence in all eukaryotes, as are the H4 histones, suggesting strict conservation of their functions. For example, only 2 of 102 amino acid residues differ between the H4 histone molecules of peas and cows, and only 8 differ between the H4 histones of humans and yeast. Histones H1, H2A, and H2B show less sequence similarity among eukaryotic species.

Each type of histone has variant forms, because certain amino acid side chains are enzymatically modified by methylation, ADP-ribosylation, phosphorylation, glycosylation, or acetylation. Such modifications affect the net electric charge, shape, and other properties of histones, as well as the structural and functional properties of the chromatin, and they play a role in the regulation of transcription (Chapter 28).

Nucleosomes Are the Fundamental Organizational Units of Chromatin

The eukaryotic chromosome depicted in Figure 24–5 represents the compaction of a DNA molecule about 10^5 μm long into a cell nucleus that is typically 5 to 10 μm in diameter. This compaction involves several levels of highly organized folding. Subjection of chromosomes to treatments that partially unfold them reveals a structure in which the DNA is bound tightly to beads of protein, often regularly spaced (Fig. 24–26). The beads in this "beads-on-a-string" arrangement are complexes of histones and DNA. The bead plus the connecting DNA that leads to the next bead form the nucleosome, the fundamental unit of organization upon which the higher-order packing of chromatin is built. The bead of each nucleosome contains eight histone molecules: two copies each of H2A, H2B, H3, and H4. The spacing of the nucleosome beads provides a repeating unit typically of about 200 bp, of which 146 bp are bound tightly around the eight-part histone core and the remainder serve as linker DNA between nucleosome beads. Histone H1 binds to the linker DNA. Brief treatment of chromatin with enzymes that digest DNA causes preferential degradation of the linker DNA, releasing histone particles containing 146 bp of bound DNA that have been protected from digestion. Researchers have crystallized nucleosome cores obtained in this way, and x-ray diffraction analysis reveals a particle made up of the eight histone molecules with the DNA wrapped around it in the form of a left-handed solenoidal supercoil (Fig. 24–27).

A close inspection of this structure reveals why eukaryotic DNA is underwound even though eukaryotic cells lack enzymes that underwind DNA. Recall that the solenoidal wrapping of DNA in nucleosomes is but one form of supercoiling that can be taken up by underwound (negatively supercoiled) DNA. The tight wrapping of DNA around the histone core requires the removal of about one helical turn in the DNA. When the protein core of a nucleosome binds in vitro to a relaxed, closed-circular DNA, the binding introduces a negative supercoil. Because this binding process does not break the DNA or change the linking number, the formation of a negative solenoidal supercoil must be accompanied by a compensatory positive supercoil in the unbound region of the DNA (Fig. 24–28). As mentioned earlier, eukaryotic topoisomerases can relax positive supercoils. Relaxing the unbound positive supercoil leaves the negative supercoil fixed (through its binding to the nucleosome histone core) and results in an overall decrease in linking number. Indeed, topoisomerases have proved necessary for assembling chromatin from purified histones and closed-circular DNA in vitro.

Another factor that affects the binding of DNA to histones in nucleosome cores is the sequence of the

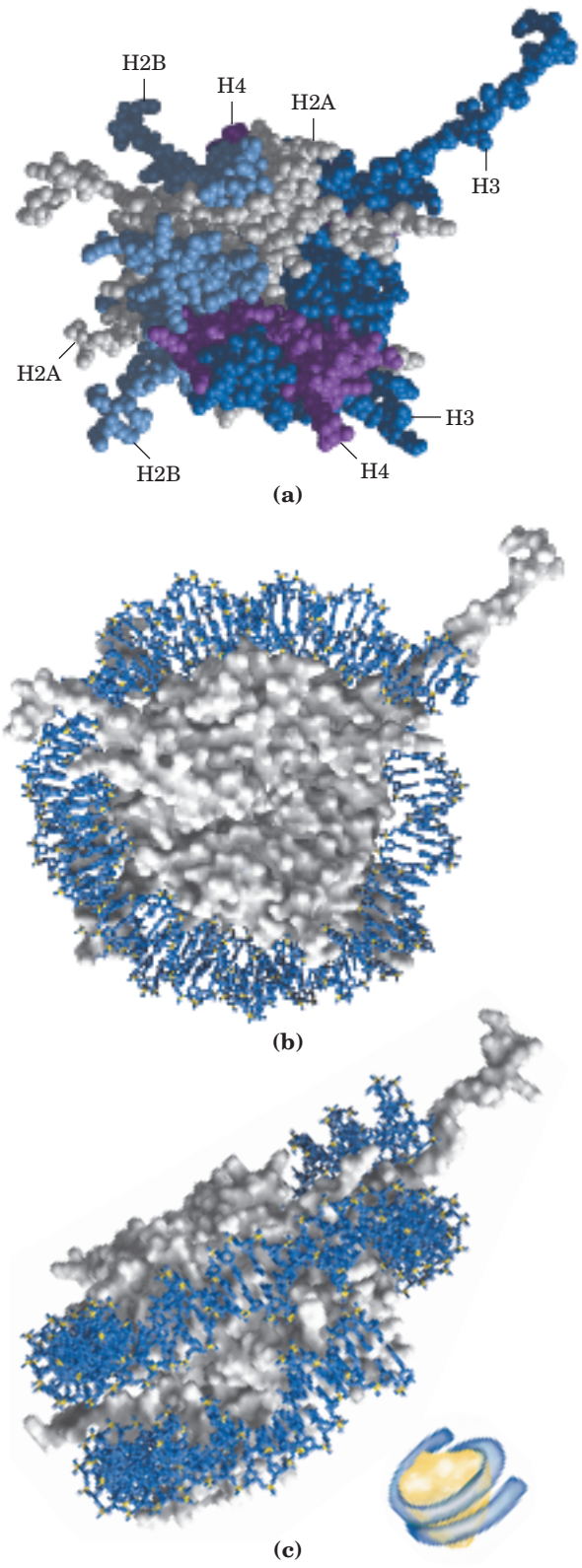


FIGURE 24–27 **DNA wrapped around a nucleosome core.** **(a)** Space-filling representation of the nucleosome protein core, with different colors for the different histones (PDB ID 1AOI). **(b)** Top and **(c)** side views of the crystal structure of a nucleosome with 146 bp of bound DNA. The protein is depicted as a gray surface contour, with the bound DNA in blue. The DNA binds in a left-handed solenoidal supercoil that circumnavigates the histone complex 1.8 times. A schematic drawing is included in **(c)** for comparison with other figures depicting nucleosomes.

TABLE 24–3 Types and Properties of Histones

Histone	Molecular weight	Number of amino acid residues	Content of basic amino acids (% of total)	
			Lys	Arg
H1*	21,130	223	29.5	11.3
H2A*	13,960	129	10.9	19.3
H2B*	13,774	125	16.0	16.4
H3	15,273	135	19.6	13.3
H4	11,236	102	10.8	13.7

*The sizes of these histones vary somewhat from species to species. The numbers given here are for bovine histones.

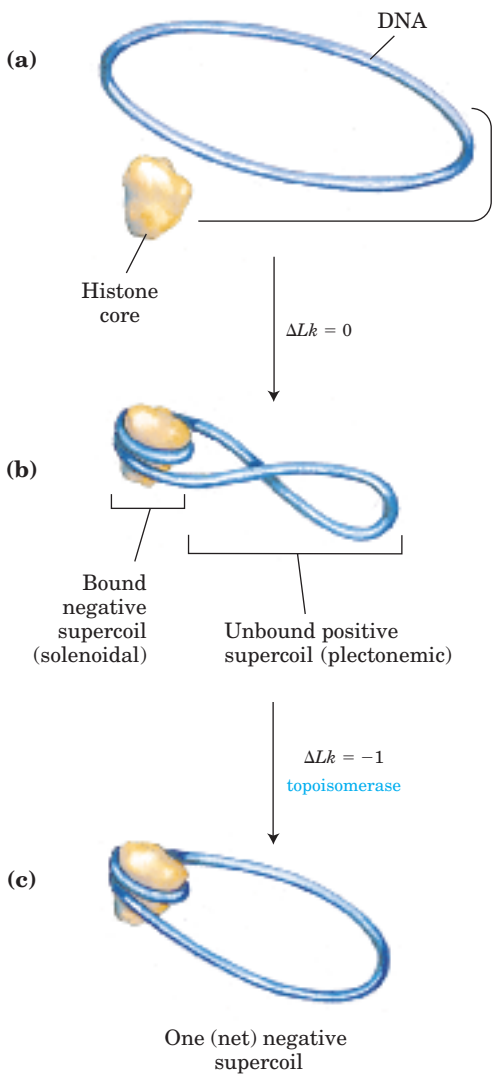


FIGURE 24–28 Chromatin assembly. (a) Relaxed, closed-circular DNA. **(b)** Binding of a histone core to form a nucleosome induces one negative supercoil. In the absence of any strand breaks, a positive supercoil must form elsewhere in the DNA ($\Delta Lk = 0$). **(c)** Relaxation of this positive supercoil by a topoisomerase leaves one net negative supercoil ($\Delta Lk = -1$).

bound DNA. Histone cores do not bind randomly to DNA; rather, they tend to position themselves at certain locations. This positioning is not fully understood but in some cases appears to depend on a local abundance of A═T base pairs in the DNA helix where it is in contact with the histones (Fig. 24–29). The tight wrapping of the DNA around the nucleosome's histone core requires compression of the minor groove of the helix at these points, and a cluster of two or three A═T base pairs makes this compression more likely.

Other proteins are required for the positioning of some nucleosome cores on DNA. In several organisms, certain proteins bind to a specific DNA sequence and then facilitate the formation of a nucleosome core nearby. Precise positioning of nucleosome cores can play a role in the expression of some eukaryotic genes (Chapter 28).

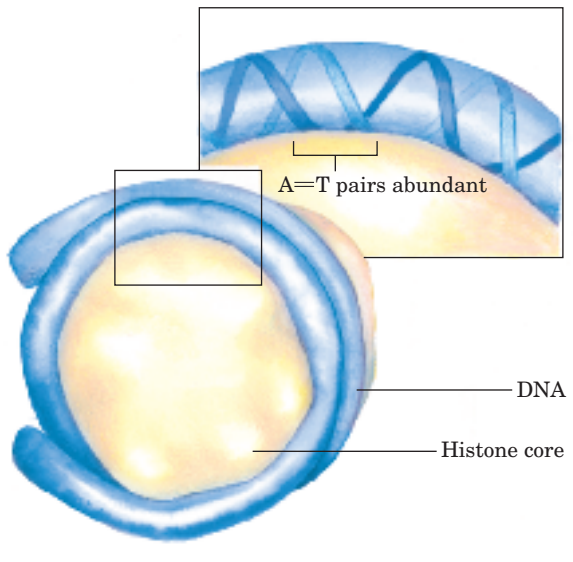
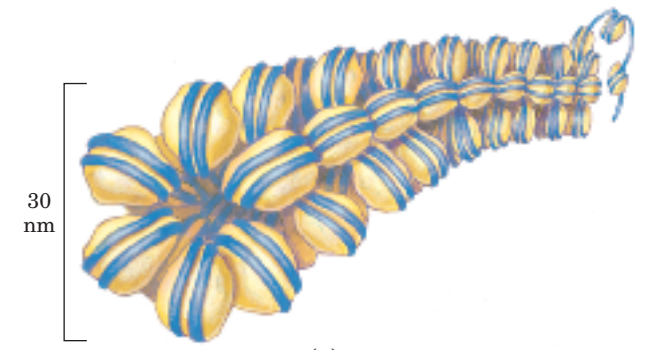
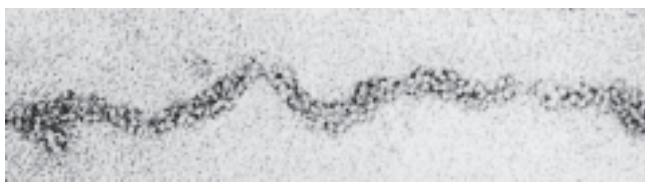


FIGURE 24–29 Positioning of a nucleosome to make optimal use of A═T base pairs where the histone core is in contact with the minor groove of the DNA helix.



(a)



(b)

FIGURE 24–30 The 30 nm fiber, a higher-order organization of nucleosomes. (a) Schematic illustration of the probable structure of the fiber, showing nucleosome packing. **(b)** Electron micrograph.

Nucleosomes Are Packed into Successively Higher Order Structures

Wrapping of DNA around a nucleosome core compacts the DNA length about sevenfold. The overall compaction in a chromosome, however, is greater than 10,000-fold—ample evidence for even higher orders of structural organization. In chromosomes isolated by very gentle methods, nucleosome cores appear to be organized into a structure called the **30 nm fiber** (Fig. 24–30). This packing requires one molecule of histone H1 per nucleosome core. Organization into 30 nm fibers does not extend over the entire chromosome but is punctuated by regions bound by sequence-specific (nonhistone) DNA-binding proteins. The 30 nm structure also appears to depend on the transcriptional activity of the particular region of DNA. Regions in which genes are being transcribed are apparently in a less-ordered state that contains little, if any, histone H1.

The 30 nm fiber, a second level of chromatin organization, provides an approximately 100-fold compaction of the DNA. The higher levels of folding are not yet understood, but it appears that certain regions of DNA associate with a nuclear scaffold (Fig. 24–31). The scaffold-associated regions are separated by loops of DNA with perhaps 20 to 100 kbp. The DNA in a loop may contain a set of related genes. For example, in *Drosophila* complete sets of histone-coding genes seem to cluster together in loops that are bounded by scaffold attachment sites (Fig. 24–32). The scaffold itself appears to contain several proteins, notably large

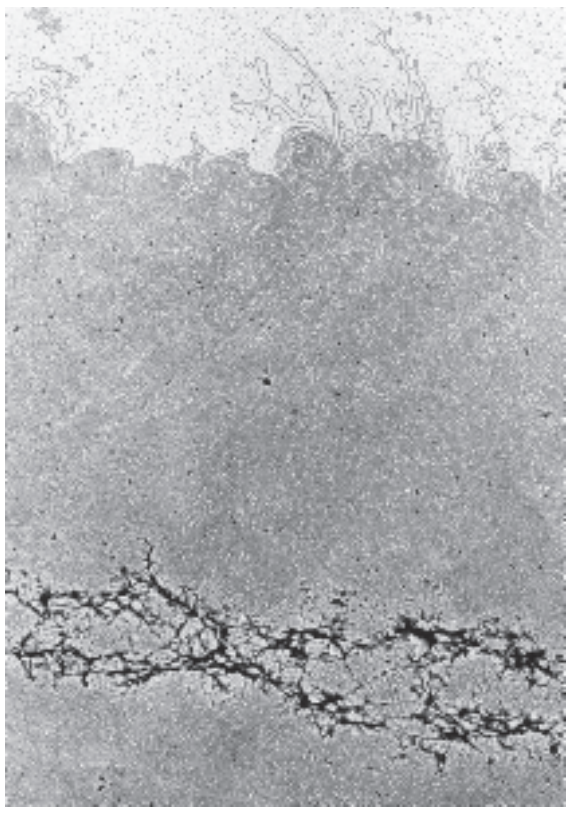


FIGURE 24–31 A partially unraveled human chromosome, revealing numerous loops of DNA attached to a scaffoldlike structure.

amounts of histone H1 (located in the interior of the fiber) and topoisomerase II. The presence of topoisomerase II further emphasizes the relationship between DNA underwinding and chromatin structure. Topoisomerase II is so important to the maintenance of chromatin structure that inhibitors of this enzyme can kill

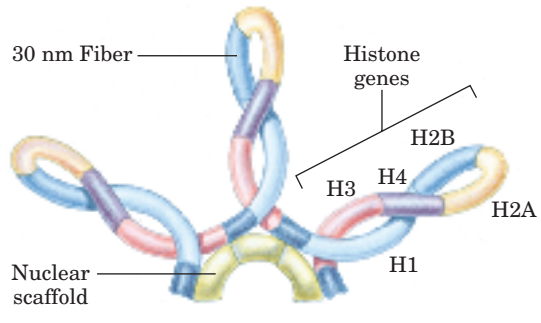


FIGURE 24–32 Loops of chromosomal DNA attached to a nuclear scaffold. The DNA in the loops is packaged as 30 nm fibers, so the loops are the next level of organization. Loops often contain groups of genes with related functions. Complete sets of histone-coding genes, as shown in this schematic illustration, appear to be clustered in loops of this kind. Unlike most genes, histone genes occur in multiple copies in many eukaryotic genomes.

rapidly dividing cells. Several drugs used in cancer chemotherapy are topoisomerase II inhibitors that allow the enzyme to promote strand breakage but not the resealing of the breaks.

Evidence exists for additional layers of organization in eukaryotic chromosomes, each dramatically enhancing the degree of compaction. One model for achieving this compaction is illustrated in Figure 24–33. Higher-order chromatin structure probably varies from chromosome to chromosome, from one region to the next in a single chromosome, and from moment to moment in the life of a cell. No single model can adequately describe these structures. Nevertheless, the principle is clear: DNA compaction in eukaryotic chromosomes is likely to involve coils upon coils upon coils . . . Three-Dimensional Packaging of Nuclear Chromosomes

Condensed Chromosome Structures Are Maintained by SMC Proteins

A third major class of chromatin proteins, in addition to the histones and topoisomerases, is the **SMC proteins** (*s*tructural *m*aintenance of *c*hromosomes). The primary structure of SMC proteins consists of five distinct domains (Fig. 24–34a). The amino- and carboxyl-terminal globular domains, N and C, each of which has part of an ATP hydrolytic site, are connected by two regions of α-helical coiled-coil motifs (see Fig. 4–11) that are joined by a hinge domain. The proteins are generally dimeric, forming a V-shaped complex that is thought to be tied together through their hinge domains (Fig. 24–34b). One N and one C domain come together to form a complete ATP hydrolytic site at each end of the V.

Proteins in the SMC family are found in all types of organisms, from bacteria to humans. Eukaryotes have two major types, cohesins and condensins (Fig. 24–25). The **cohesins** play a substantial role in linking together sister chromatids immediately after replication and keeping them together as the chromosomes condense to metaphase. This linkage is essential if chromosomes are to segregate properly at cell division. The detailed mechanism by which cohesins link sister chromosomes, and the role of ATP hydrolysis, are not yet understood. The **condensins** are essential to the condensation of chromosomes as cells enter mitosis. In the laboratory, condensins bind to DNA in a manner that creates positive supercoils; that is, condensin binding causes the DNA to become overwound, in contrast to the underwinding induced by the binding of nucleosomes. It is not yet clear how this helps to compact the chromatin, although one possibility is presented in Figure 24–35.

Bacterial DNA Is Also Highly Organized

We now turn briefly to the structure of bacterial chromosomes. Bacterial DNA is compacted in a structure called the **nucleoid,** which can occupy a significant

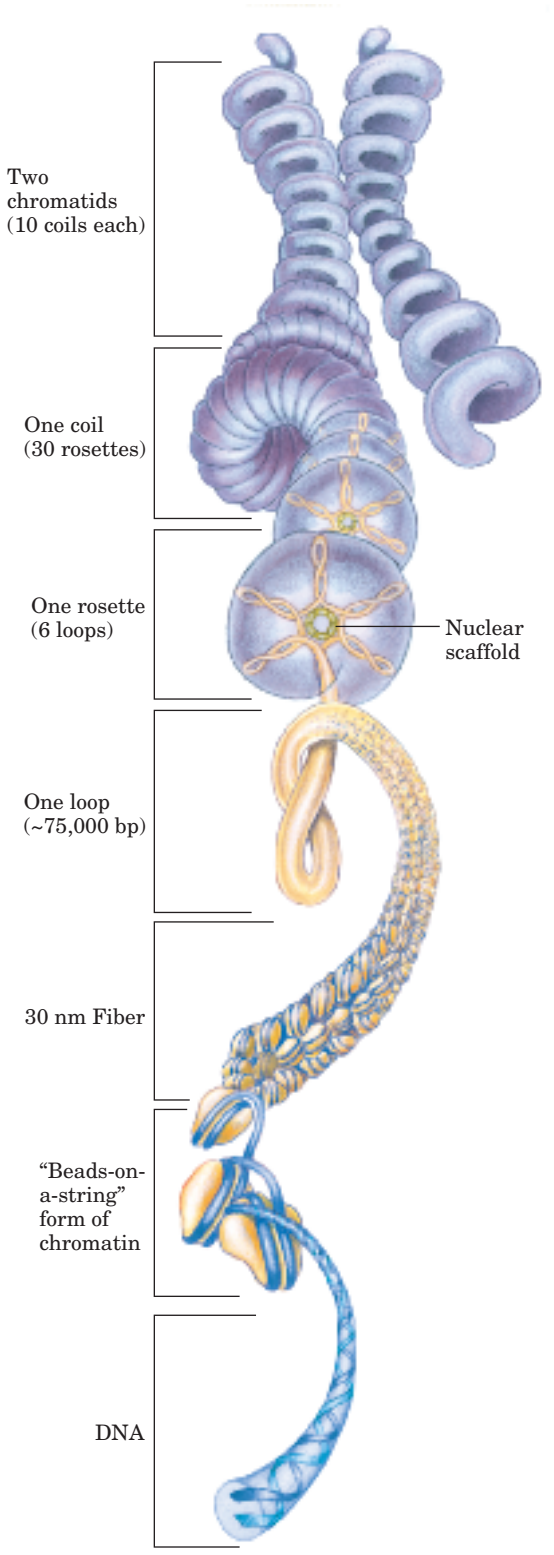


FIGURE 24–33 Compaction of DNA in a eukaryotic chromosome. Model for levels of organization that could provide DNA compaction in the chromosomes of eukaryotes. The levels take the form of coils upon coils. In cells, the higher-order structures (above the 30 nm fibers) are unlikely to be as uniform as depicted here.

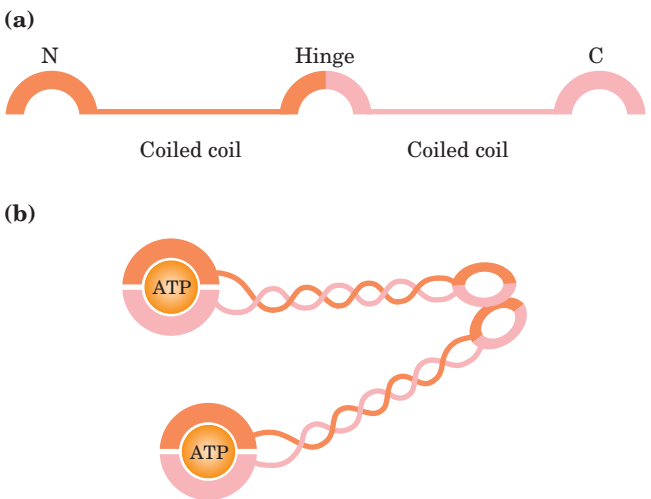


FIGURE 24–34 Structure of SMC proteins. (a) The five domains of the SMC primary structure. N and C denoted the amino-terminal and carboxyl-terminal domains, respectively. **(b)** Each polypeptide is folded so that the two coiled-coil domains wrap around each other and the N and C domains come together to form a complete ATP-binding site. Two of these domains are linked at the hinge region to form the dimeric V-shaped molecule. **(c)** Electron micrograph of SMC proteins from *Bacillus subtilis*.

fraction of the cell volume (Fig. 24–36). The DNA appears to be attached at one or more points to the inner surface of the plasma membrane. Much less is known about the structure of the nucleoid than of eukaryotic chromatin. In *E. coli,* a scaffoldlike structure appears to organize the circular chromosome into a series of looped domains, as described above for chromatin. Bacterial DNA does not seem to have any structure comparable to the local organization provided by nucleosomes in eukaryotes. Histonelike proteins are abundant in *E. coli*—the best-characterized example is a two-subunit protein called HU (M_r 19,000)—but these proteins bind and dissociate within minutes, and no regular, stable DNA-histone structure has been found. The bacterial chromosome is a relatively dynamic molecule, possibly reflecting a requirement for more ready access to its genetic information. The bacterial cell division cycle can be as short as 15 min, whereas a typical eukaryotic cell may not divide for hours or even months. In addition, a much greater fraction of prokaryotic DNA is used to encode RNA and/or protein products. Higher rates of cellular metabolism in bacteria mean that a much higher proportion of the DNA is being transcribed or replicated at a given time than in most eukaryotic cells.

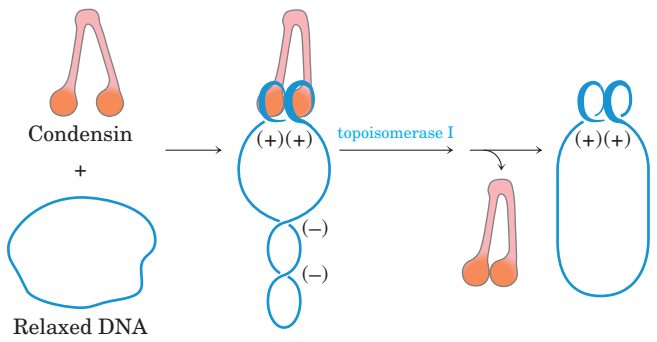


FIGURE 24–35 Model for the effect of condensins on DNA supercoiling. Binding of condensins to a closed-circular DNA in the presence of topoisomerase I leads to the production of positive supercoils (+). Wrapping of the DNA about the condensin introduces positive supercoils because it wraps in the opposite sense to a solenoidal supercoil (see Fig. 24–24). The compensating negative supercoils (−) that appear elsewhere in the DNA are then relaxed by topoisomerase I. In the chromosome, it is the wrapping of the DNA about condensin that may contribute to DNA condensation.

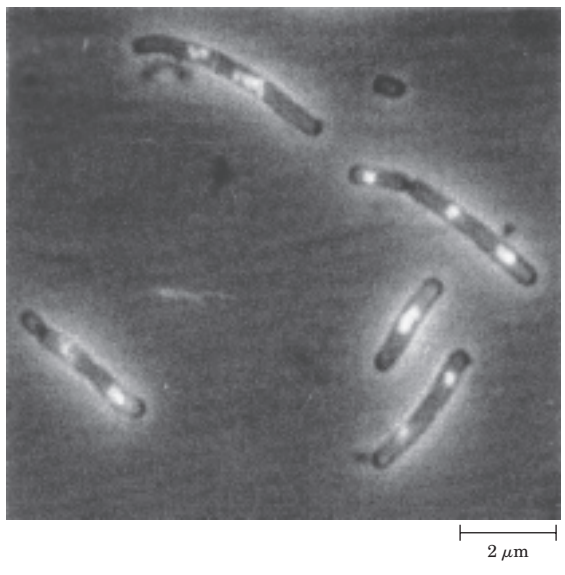


FIGURE 24–36 *E. coli* cells showing nucleoids. The DNA is stained with a dye that fluoresces when exposed to UV light. The light area defines the nucleoid. Note that some cells have replicated their DNA but have not yet undergone cell division and hence have multiple nucleoids.

With this overview of the complexity of DNA structure, we are now ready to turn, in the next chapter, to a discussion of DNA metabolism.

SUMMARY 24.3 The Structure of Chromosomes

- The fundamental unit of organization in the chromatin of eukaryotic cells is the nucleosome, which consists of histones and a 200 bp segment of DNA. A core protein particle containing eight histones (two copies each of histones H2A, H2B, H3, and H4) is encircled by a segment of DNA (about 146 bp) in the form of a left-handed solenoidal supercoil.
- Nucleosomes are organized into 30 nm fibers, and the fibers are extensively folded to provide the 10,000-fold compaction required to fit a typical eukaryotic chromosome into a cell nucleus. The higher-order folding involves attachment to a nuclear scaffold that contains histone H1, topoisomerase II, and SMC proteins.
- Bacterial chromosomes are also extensively compacted into the nucleoid, but the chromosome appears to be much more dynamic and irregular in structure than eukaryotic chromatin, reflecting the shorter cell cycle and very active metabolism of a bacterial cell.

Key Terms

Terms in bold are defined in the glossary.

gene 921
genome 923
chromosome 923
phenotype 924
mutation 924
regulatory sequence 924
plasmid 925
intron 928
exon 928
simple-sequence DNA 929
satellite DNA 929
centromere 930
telomere 930
supercoil 930
relaxed DNA 930
topology 931
underwinding 932
linking number 933
specific linking difference (σ) 933
superhelical density 933
topoisomers 934
topoisomerases 935
plectonemic 937
solenoidal 937
chromatin 938
histones 938
nucleosome 938
30 nm fiber 942
SMC proteins 943
cohesins 943
condensins 943
nucleoid 943

Further Reading

General

Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453–1474.
New secrets of this common laboratory organism are revealed.

Cozzarelli, N.R. & Wang, J.C. (eds) (1990) *DNA Topology and Its Biological Effects,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Kornberg, A. & Baker, T.A. (1991) *DNA Replication*, 2nd edn, W. H. Freeman & Company, New York.
A good place to start for further information on the structure and function of DNA.

Lodish, H., Berk, A., Matsudaira, P., Kaiser, C.A., Krieger, M., Scott, M.P., Zipursky, S.L., & Darnell, J. (2003) *Molecular Cell Biology,* 5th edn, W. H. Freeman & Company, New York.
Another excellent general reference.

Genes and Chromosomes

Bromham, L. (2002) The human zoo: endogenous retroviruses in the human genome. *Trends Ecol. Evolut.* **17,** 91–97.
A thorough description of one of the transposon classes that makes up a large part of the human genome.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. (1996) Life with 6000 genes. *Science* **274,** 546, 563–567.
Report of the first complete sequence of a eukaryotic genome, the yeast *Saccharomyces cerevisiae.*

Greider, C.W. & Blackburn, E.H. (1996) Telomeres, telomerase and cancer. *Sci. Am.* **274** (February), 92–97.

Huxley, C. (1997) Mammalian artificial chromosomes and chromosome transgenics. *Trends Genet.* **13,** 345–347.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921.
One of the first reports on the draft sequence of the human genome, with lots of analysis and many associated articles.

Long, M., de Souza, S.J., & Gilbert, W. (1995) Evolution of the intron-exon structure of eukaryotic genes. *Curr. Opin. Genet. Dev.* **5,** 774–778.

McEachern, M.J., Krauskopf, A., & Blackburn, E.H. (2000) Telomeres and their control. *Annu. Rev. Genet.* **34,** 331–358.

Schmid, C.W. (1996) Alu: structure, origin, evolution, significance and function of one-tenth of human DNA. *Prog. Nucleic Acid Res. Mol. Biol.* **53,** 283–319.

Tyler-Smith, C. & Floridia, G. (2000) Many paths to the top of the mountain: diverse evolutionary solutions to centromere structure. *Cell* **102,** 5–8.

Details of the diversity of centromere structures from different organisms, as currently understood.

Zakian, V.A. (1996) Structure, function, and replication of *Saccharomyces cerevisiae* telomeres. *Annu. Rev. Genet.* **30,** 141–172.

Supercoiling and Topoisomerases

Berger, J.M. (1998) Type II DNA topoisomerases. *Curr. Opin. Struct. Biol.* **8,** 26–32.

Boles, T.C., White, J.H., & Cozzarelli, N.R. (1990) Structure of plectonemically supercoiled DNA. *J. Mol. Biol.* **213,** 931–951.

A study that defines several fundamental features of supercoiled DNA.

Champoux, J.J. (2001) DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* **70,** 369–413.

An excellent summary of the topoisomerase classes.

Cozzarelli, N.R., Boles, T.C., & White, J.H. (1990) Primer on the topology and geometry of DNA supercoiling. In *DNA Topology and Its Biological Effects* (Cozzarelli, N.R. & Wang, J.C., eds), pp. 139–184, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

A more advanced and thorough discussion.

Lebowitz, J. (1990) Through the looking glass: the discovery of supercoiled DNA. *Trends Biochem. Sci.* **15,** 202–207.

A short and interesting historical note.

Wang, J.C. (2002) Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.* **3,** 430–440.

Chromatin and Nucleosomes

Filipski, J., Leblanc, J., Youdale, T., Sikorska, M., & Walker, P.R. (1990) Periodicity of DNA folding in higher order chromatin structures. *EMBO J.* **9,** 1319–1327.

Hirano, T. (2002) The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion and repair. *Genes Dev.* **16,** 399–414.

Description of the rapid advances in understanding of this interesting class of proteins.

Kornberg, R.D. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science* **184,** 868–871.

A classic paper that introduced the subunit model for chromatin.

Nasmyth, K. (2002) Segregating sister genomes: the molecular biology of chromosome separation. *Science* **297,** 559–565.

Wyman, C. & Kanaar, R. (2002) Chromosome organization: reaching out to embrace new models. *Curr. Biol.* **12,** R446–R448.

A good, short summary of chromosome structure and the roles of SMC proteins within it.

Zlatanova, J. & van Holde, K. (1996) The linker histones and chromatin structure: new twists. *Prog. Nucleic Acid Res. Mol. Biol.* **52,** 217–259.

Problems

1. Packaging of DNA in a Virus Bacteriophage T2 has a DNA of molecular weight 120×10^6 contained in a head about 210 nm long. Calculate the length of the DNA (assume the molecular weight of a nucleotide pair is 650) and compare it with the length of the T2 head.

2. The DNA of Phage M13 The base composition of phage M13 DNA is A, 23%; T, 36%; G, 21%; C, 20%. What does this tell you about the DNA of phage M13?

3. The *Mycoplasma* Genome The complete genome of the simplest bacterium known, *Mycoplasma genitalium*, is a circular DNA molecule with 580,070 bp. Calculate the molecular weight and contour length (when relaxed) of this molecule. What is Lk_0 for the *Mycoplasma* chromosome? If $\sigma = -0.06$, what is Lk?

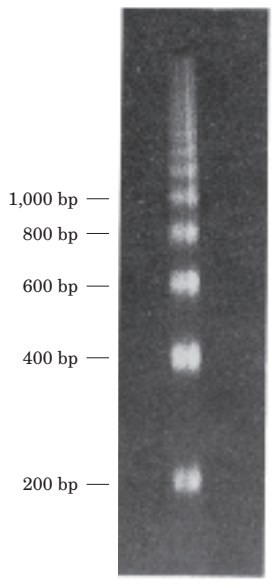
4. Size of Eukaryotic Genes An enzyme isolated from rat liver has 192 amino acid residues and is coded for by a gene with 1,440 bp. Explain the relationship between the number of amino acid residues in the enzyme and the number of nucleotide pairs in its gene.

5. Linking Number A closed-circular DNA molecule in its relaxed form has an Lk of 500. Approximately how many base pairs are in this DNA? How is the linking number altered (increases, decreases, doesn't change, becomes undefined) when (a) a protein complex is bound to form a nucleosome, (b) one DNA strand is broken, (c) DNA gyrase and ATP are added to the DNA solution, or (d) the double helix is denatured by heat?

6. Superhelical Density Bacteriophage λ infects *E. coli* by integrating its DNA into the bacterial chromosome. The success of this recombination depends on the topology of the *E. coli* DNA. When the superhelical density (σ) of the *E. coli* DNA is greater than -0.045, the probability of integration is $<20\%$; when σ is less than -0.06, the probability is $>70\%$. Plasmid DNA isolated from an *E. coli* culture is found to have a length of 13,800 bp and an Lk of 1,222. Calculate σ for this DNA and predict the likelihood that bacteriophage λ will be able to infect this culture.

7. Altering Linking Number (a) What is the Lk of a 5,000 bp circular duplex DNA molecule with a nick in one strand? (b) What is the Lk of the molecule in (a) when the nick is sealed (relaxed)? (c) How would the Lk of the molecule in (b) be affected by the action of a single molecule of *E. coli* topoisomerase I? (d) What is the Lk of the molecule in (b) after eight enzymatic turnovers by a single molecule of DNA gyrase in the presence of ATP? (e) What is the Lk of the molecule in (d) after four enzymatic turnovers by a single molecule of bacterial type I topoisomerase? (f) What is the Lk of the molecule in (d) after binding of one nucleosome?

8. Chromatin Early evidence that helped researchers define nucleosome structure is illustrated by the agarose gel below, in which the thick bands represent DNA. It was generated by briefly treating chromatin with an enzyme that degrades DNA, then removing all protein and subjecting the purified DNA to electrophoresis. Numbers at the side of the gel denote the position to which a linear DNA of the indicated size would migrate. What does this gel tell you about chromatin structure? Why are the DNA bands thick and spread out rather than sharply defined?



9. DNA Structure Explain how the underwinding of a B-DNA helix might facilitate or stabilize the formation of Z-DNA.

10. Maintaining DNA Structure (a) Describe two structural features required for a DNA molecule to maintain a negatively supercoiled state. (b) List three structural changes that become more favorable when a DNA molecule is negatively supercoiled. (c) What enzyme, with the aid of ATP, can generate negative superhelicity in DNA? (d) Describe the physical mechanism by which this enzyme acts.

11. Yeast Artificial Chromosomes (YACs) YACs are used to clone large pieces of DNA in yeast cells. What three types of DNA sequences are required to ensure proper replication and propagation of a YAC in a yeast cell?