NCBI Bookshelf. A service of the National Library of Medicine, National Institutes of Health.

Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002.

# Chromosomal DNA and Its Packaging in the Chromatin Fiber

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

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

# **Eucaryotic DNA Is Packaged into a Set of Chromosomes**

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

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

With the exception of the germ cells, and a few highly specialized cell types that cannot multiply and lack DNA altogether (for example, red blood cells), each human cell contains two copies of each chromosome, one inherited from the mother and one from the father. The maternal and paternal chromosomes of a pair are called **homologous chromosomes (homologs)**. The only nonhomologous chromosome pairs are the sex chromosomes in males, where a *Y chromosome* is inherited from the father and an *X chromosome* from the mother. Thus, each human cell contains a total of 46 chromosomes—22 pairs common to both males and females, plus two so-called sex

chromosomes (X and Y in males, two Xs in females). *DNA hybridization* (described in detail in Chapter 8) can be used to distinguish these human chromosomes by "painting" each one a different color (Figure 4-10). Chromosome painting is typically done at the stage in the cell cycle when chromosomes are especially compacted and easy to visualize (mitosis, see below).



### Figure 4-10

Human chromosomes. These chromosomes, from a male, were isolated from a cell undergoing nuclear division (mitosis) and are therefore highly compacted. Each chromosome has been "painted" a different color to permit its unambiguous identification (more...)

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

## Figure 4-11

The banding patterns of human chromosomes. Chromosomes 1–22 are numbered in approximate order of size. A typical human somatic (non-germ line) cell contains two of each of these chromosomes, plus two sex chromosomes—two X chromosomes in (more...)

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

# Figure 4-12

An aberrant human chromosome. (A) Two pairs of chromosomes, stained with Giemsa (see Figure 4-11), from a patient with ataxia, a disease characterized by progressive deterioration of motor skills. The patient has a normal pair of chromosome 4s *(left-hand* (more...)

# **Chromosomes Contain Long Strings of Genes**

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

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



#### Figure 4-13

The genome of S. cerevisiae (budding yeast). (A) The genome is distributed over 16 chromosomes, and its complete nucleotide sequence was determined by a cooperative effort involving scientists working in many different locations, as indicated *(gray, Canada; (more...)* 

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

The apportionment of the genome over chromosomes also differs from one eucaryotic species to the next. For example, compared with 46 for humans, somatic cells from a species of small deer contain only 6 chromosomes, while those from a species of carp contain over 100. Even closely related species with similar genome sizes can have very different numbers and sizes of chromosomes (Figure 4-14). Thus, there is no simple relationship between chromosome number, species complexity, and total genome size. Rather, the genomes and chromosomes of modern-day species have each been shaped by a unique history of seemingly random genetic events, acted on by selection pressures.

# W 1111 W 1114

## Figure 4-14

Two closely related species of deer with very different chromosome numbers. In the evolution of the Indian muntjac, initially separate chromosomes fused, without having a major effect on the animal. These two species have roughly the same number of genes. (more...)

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

When the DNA sequence of human chromosome 22, one of the smallest human chromosomes (see Figure 4-11), was completed in 1999, it became possible for the first time to see exactly how genes are arranged along an entire vertebrate chromosome (Figure 4-15 and Table 4-1). With the publication of the "first draft" of the entire human genome in 2001, the genetic landscape of all human chromosomes suddenly came into sharp focus. The sheer quantity of information provided by the Human Genome Project is unprecedented in biology (Figure 4-16 and Table 4-1); the human genome is 25 times larger than any other genome sequenced so far, and is 8 times as large as the sum of all previously sequenced genomes. At its peak, the Human Genome Project generated raw nucleotide sequences at a rate of 1000 nucleotides per second around the clock. It will be many decades before this information is fully analyzed, but it will continue to stimulate many new experiments and has already affected the content of all the chapters in this book.



## Figure 4-15

The organization of genes on a human chromosome. (A) Chromosome 22, one of the smallest human chromosomes, contains  $48 \times 10^6$  nucleotide pairs and makes up approximately 1.5% of the entire human genome. Most of the left arm of chromosome 22 consists (more...)

	CHR
DNA length	48 × pairs
Number of genes	appro
Smallest protein-coding gene	1000
Largest gene	583,0
Mean gene size	19,00
Smallest number of exons per gene	1

# Table 4-1

Vital Statistics of Human Chromosome 22 and the Entire Human Genome.

-
90

# Figure 4-16

Scale of the human genome. If each nucleotide pair is drawn as 1 mm as in (A), then the human genome would extend 3200 km (approximately 2000 miles), far enough to stretch across the center of Africa, the site of our human origins *(red line* in B). At (more...)

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



# Figure 4-17

Representation of the nucleotide sequence content of the human genome. LINES, SINES, retroviral-like elements, and DNA-only transposons are all mobile genetic elements that have multiplied in our genome by replicating themselves and inserting the new (more...)

A second notable feature of the human genome is the large average gene size of 27,000 nucleotide pairs. As discussed above, a typical gene carries in its linear sequence of nucleotides the

information for the linear sequence of the amino acids of a protein. Only about 1300 nucleotide pairs are required to encode a protein of average size (about 430 amino acids in humans). Most of the remaining DNA in a gene consists of long stretches of noncoding DNA that interrupt the relatively short segments of DNA that code for protein. The coding sequences are called *exons*; the intervening (noncoding) sequences are called *introns* (see Figure 4-15 and Table 4-1).

The majority of human genes thus consist of a long string of alternating exons and introns, with most of the gene consisting of introns. In contrast, the majority of genes from organisms with compact genomes lack introns. This accounts for the much smaller size of their genes (about one-twentieth that of human genes), as well as for the much higher fraction of coding DNA in their chromosomes. In addition to introns and exons, each gene is associated with *regulatory DNA sequences*, which are responsible for ensuring that the gene is expressed at the proper level and time, and the proper type of cell. In humans, the regulatory sequences for a typical gene are spread out over tens of thousands of nucleotide pairs. As would be expected, these regulatory sequences are more compressed in organisms with compact genomes. We discuss in Chapter 7 how regulatory DNA sequences work.

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

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

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

very long natural "experiment," comparative DNA sequencing studies highlight the most interesting regions in genomes.

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



## Figure 4-18

Conserved synteny between the human and mouse genomes. Regions from different mouse chromosomes (indicated by the colors of each mouse in B) show conserved synteny (gene order) with the indicated regions of the human genome (A). For example the genes (more...)



### Figure 4-19

A proposed evolutionary history of human chromosome 3 and its relatives in other mammals. (A) At the *lower left* is the order of chromosome 3 segments hypothesized to be present on a chromosome of a mammalian ancestor. Along the *top* are the patterns of (more...)

# Chromosomes Exist in Different States Throughout the Life of a Cell

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



#### Figure 4-20

A simplified view of the eucaryotic cell cycle. During interphase, the cell is actively expressing its genes and is therefore synthesizing proteins. Also, during interphase and before cell division, the DNA is replicated and the chromosomes are duplicated. (more...) During the portions of the cell cycle when the cell is not dividing, the chromosomes are extended and much of their chromatin exists as long, thin tangled threads in the nucleus so that individual chromosomes cannot be easily distinguished (Figure 4-21). We refer to chromosomes in this extended state as *interphase chromosomes*.



## Figure 4-21

A comparison of extended interphase chromatin with the chromatin in a mitotic chromosome. (A) An electron micrograph showing an enormous tangle of chromatin spilling out of a lysed interphase nucleus. (B) A scanning electron micrograph of a mitotic chromosome: (more...)

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

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



## Figure 4-22

The three DNA sequences required to produce a eucaryotic chromosome that can be replicated and then segregated at mitosis. Each chromosome has multiple origins of replication, one centromere, and two telomeres. Shown here is the sequence of events a typical (more...)

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

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

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

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

# **DNA Molecules Are Highly Condensed in Chromosomes**

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

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

# Nucleosomes Are the Basic Unit of Eucaryotic Chromosome Structure

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

Histones are responsible for the first and most basic level of chromosome organization, the nucleosome, which was discovered in 1974. When interphase nuclei are broken open very gently and their contents examined under the electron microscope, most of the chromatin is in the form of a fiber with a diameter of about 30 nm (Figure 4-23A). If this chromatin is subjected to treatments that cause it to unfold partially, it can be seen under the electron microscope as a series of "beads on a string" (Figure 4-23B). The string is DNA, and each bead is a "nucleosome core particle" that consists of DNA wound around a protein core formed from histones. The beads on a string represent the first level of chromosomal DNA packing.



# Figure 4-23

Nucleosomes as seen in the electron microscope. (A) Chromatin isolated directly from an interphase nucleus appears in the electron microscope as a thread 30 nm thick. (B) This electron micrograph shows a length of chromatin that has been experimentally (more...)

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



## Figure 4-24

Structural organization of the nucleosome. A nucleosome contains a protein core made of eight histone molecules. As indicated, the nucleosome core particle is released from chromatin by digestion of the linker DNA with a nuclease, an enzyme that breaks (more...)

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

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

The high-resolution structure of a nucleosome core particle, solved in 1997, revealed a disc-shaped histone core around which the DNA was tightly wrapped 1.65 turns in a left-handed coil (Figure 4-25). All four of the histones that make up the core of the nucleosome are relatively small proteins (102–135 amino acids), and they share a structural motif, known as the *histone fold*, formed from three  $\alpha$  helices connected by two loops (Figure 4-26). In assembling a nucleosome, the histone folds first bind to each other to form H3–H4 and H2A-H2B dimers, and the H3–H4 dimers combine to form tetramers. An H3–H4 tetramer then further combines with two H2A-H2B dimers to form the compact octamer core, around which the DNA is wound (Figure 4-27).



# Figure 4-25

The structure of a nucleosome core particle, as determined by x-ray diffraction analyses of crystals. Each histone is colored according to the scheme of Figure 4-24, with the DNA double

helix in *light gray.* (Reprinted by permission from K. Luger et al., (more...)



## Figure 4-26

The overall structural organization of the core histones. (A) Each of the core histones contains an N-terminal tail, which is subject to several forms of covalent modification, and a histone fold region, as indicated. (B) The structure of the histone (more...)



## Figure 4-27

The assembly of a histone octamer. The histone H3–H4 dimer and the H2A-H2B dimer are formed from the handshake interaction. An H3-H4 tetramer forms the scaffold of the octamer onto which two H2A-H2B dimers are added, to complete the assembly. (more...)

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

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

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

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

the high rates of DNA transcription and DNA replication that occur during these early stages of development.

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

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



## Figure 4-28

The bending of DNA in a nucleosome. The DNA helix makes 1.65 tight turns around the histone octamer. This diagram is drawn approximately to scale, illustrating how the minor groove is compressed on the inside of the turn. Owing to certain structural features (more...)

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

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

# Nucleosomes Are Usually Packed Together into a Compact Chromatin Fiber

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

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



#### Figure 4-29

Variations on the Zigzag model for the 30-nm chromatin fiber. (A and B) Electron microscopic evidence for the *top* and *bottom-left* model structures depicted in (C). (C) Zigzag variations. An interconversion between these three variations is proposed to (more...)



## Figure 4-30

Irregularities in the 30-nm fiber. This schematic view of the 30nm fiber illustrates its interruption by sequence-specific DNAbinding proteins. How these proteins bind tightly to DNA is explained in Chapter 7. The interruptions in the 30-nm fiber may (more...)

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



## Figure 4-31

A speculative model for how histone H1 could change the path of DNA as it exits from the nucleosome. Histone H1 *(green)* consists of a globular core and two extended tails. Part of the effect of H1 on the compaction of nucleosome organization may result (more...)

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



A speculative model for histone tails in the formation of the 30-nm fiber. (A) The approximate exit points of the eight histone tails, four from each histone subunit, that extend from each nucleosome. In the high-resolution structure of the nucleosome (more...)

# ATP-driven Chromatin Remodeling Machines Change Nucleosome Structure

For many years biologists thought that, once formed in a particular position on DNA, a nucleosome remained fixed in place because of the tight association between the core histones and DNA. But it has recently been discovered that eucaryotic cells contain *chromatin remodeling complexes*, protein machines that use the energy of ATP hydrolysis to change the structure of nucleosomes temporarily so that DNA becomes less tightly bound to the histone core. The remodeled state may result from movement of the H2A-H2B dimers in the nucleosome core; the H3–H4 tetramer is particularly stable and would be difficult to rearrange (see Figure 4-27).

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



# Figure 4-33

Model for the mechanism of some chromatin remodeling complexes. In the absence of remodeling complexes, the interconversion between the three nucleosomal states shown is very slow because of a high activation energy barrier. Using ATP hydrolysis, chromatin-remodeling (more...)

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



## Figure 4-34

A cyclic mechanism for nucleosome disruption and reformation. According to this model, different chromatin remodeling complexes disrupt and re-form nucleosomes, although, in principle, the same complex might catalyze both reactions. The DNA-binding proteins (more...) Chromatin remodeling complexes are carefully controlled by the cell. We shall see in Chapter 7 that, when genes are turned on and off, these complexes can be brought to specific regions of DNA where they act locally to influence chromatin structure. During mitosis, at least some of the chromatin-remodeling complexes are inactivated by phosphorylation. This may help the tightly packaged mitotic chromosomes maintain their structure.

# **Covalent Modification of the Histone Tails Can Profoundly Affect Chromatin**

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

-		
1 arrivered		The second
C	and the second second	and desired.
	-	-
	******	Interesting
		And in case of the local division of the loc
	phone in the local state	-
	Para land	are equal to
9 89 99	141110	
	secultud	-producting" -
1 . 1 .	-	Same Squade
	arrest.	-

### Figure 4-35

Covalent modification of core histone tails. (A) Known modifications of the four histone core proteins are indicated: Me = methyl group, Ac = acetyl group, P = phosphate, u = ubiquitin. Note that some positions (e.g., lysine 9 of H3) can be modified in (more...)

The various modifications of the histone tails have several important consequences. Although modifications of the tails have little direct effect on the stability of an individual nucleosome, they seem to affect the stability of the 30-nm chromatin fiber and of the higher-order structures discussed below. For example, histone acetylation tends to destabilize chromatin structure, perhaps in part because adding an acetyl group removes the positive charge from the lysine, thereby making it more difficult for histones to neutralize the charges on DNA as chromatin is compacted. However, the most profound effect of modified histone tails is their ability to attract specific proteins to a stretch of chromatin that has been appropriately modified. Depending on the precise tail modifications, these additional proteins can either cause further compaction of the chromatin or can facilitate access to the DNA. If combinations of modifications are taken into account, the number of possible distinct markings for each histone tail is very large. Thus, it has been proposed that, through covalent modification of the histone tails, a given stretch of chromatin can convey a particular meaning to the cell (Figure 4-35B). For example, one type of marking could signal that the stretch of chromatin has been newly replicated, and another could signal that gene expression should not take place. According to this idea, each different marking would attract those proteins that would then execute the appropriate functions. Because the histone tails are extended, and are therefore probably accessible even when chromatin is condensed, they provide an especially apt format for such messages.

As with chromatin remodeling complexes, the enzymes that modify (and remove modifications from) histone tails are usually multisubunit proteins, and they are tightly regulated. They are brought to a particular region of chromatin by other cues, particularly by sequence-specific DNA-

binding proteins. We can thus imagine how cycles of histone tail modification and demodification can allow chromatin structure to be dynamic—locally compacting and decompacting it, and, in addition, attracting other proteins specific for each modification state. It is likely that histone-modifying enzymes and chromatin remodeling complexes work in concert to condense and recondense stretches of chromatin; for example, evidence suggests that a particular modification of the histone tail attracts a particular type of remodeling complex. Moreover, some chromatin remodeling complexes as subunits, directly connecting the two processes.

## **Summary**

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

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

By agreement with the publisher, this book is accessible by the search feature, but cannot be browsed.

Copyright © 2002, Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter; Copyright © 1983, 1989, 1994, Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson .

Bookshelf ID: NBK26834