Лекция 6.
Биосинтез нуклеиновых кислот

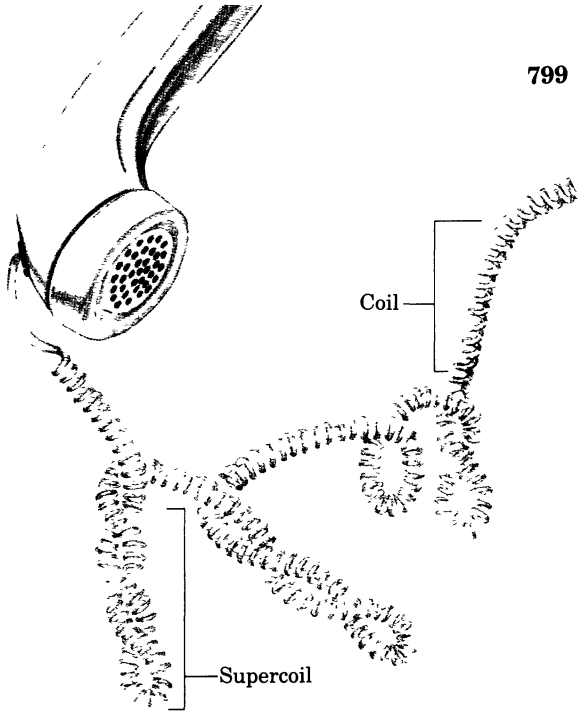


Figure 23–9 Supercoils. A typical phone cord is a coil. A phone cord twisted as shown is a supercoil. The illustration is especially appropriate, because an examination of the twisting of phone cords helped lead Jerome Vinograd and colleagues to the insight that many properties of small, circular DNAs could be explained by supercoiling. They first detected DNA supercoiling in small, circular viral DNAs in 1965.

As can be seen in Figure 23–8, the introns of this particular gene are much longer than the exons; altogether the introns make up 85% of the DNA of this gene. Most eukaryotic genes examined thus far appear to contain introns that vary in number, position, and the fraction of the total length of the gene they occupy. For example, the serum albumin gene contains 6 introns, the gene for the protein conalbumin of the chicken egg contains 17 introns, and a collagen gene has been found to have over 50 introns. Genes for histones provide an example of a family of genes that appear to have no introns. Only a few prokaryotic genes contain introns. In most cases the function of introns is not clear.

DNA Supercoiling

From the examples given above, it is clear that cellular DNA must be very tightly compacted just to fit into the cell. This implies a high degree of structural organization. It is not enough just to fold the DNA into a small space, however. The packaging must permit access to the information in the DNA for processes such as replication and transcription. Before considering how this is accomplished, we must examine an important property of DNA structure that we have not yet considered—DNA supercoiling.

The term "supercoiling" means literally the coiling of a coil. A telephone cord for example, is typically a coiled wire. The twisted path often taken by that wire as it goes from the base of the phone to the receiver generally describes a supercoil (Fig. 23–9). DNA is coiled in the form of a double helix. Let us define an axis about which both strands of the DNA coil. A bending or twisting of that axis upon itself (Fig. 23–10) is referred to as **DNA supercoiling.** As detailed below, DNA supercoiling is generally a manifestation of structural strain. Conversely, if there is no net bending of the DNA axis upon itself, the DNA is said to be in a **relaxed** state.

It is probably apparent that DNA compaction must involve some form of supercoiling. Perhaps less apparent is the fact that replicating or transcribing DNA also must induce some degree of supercoiling.

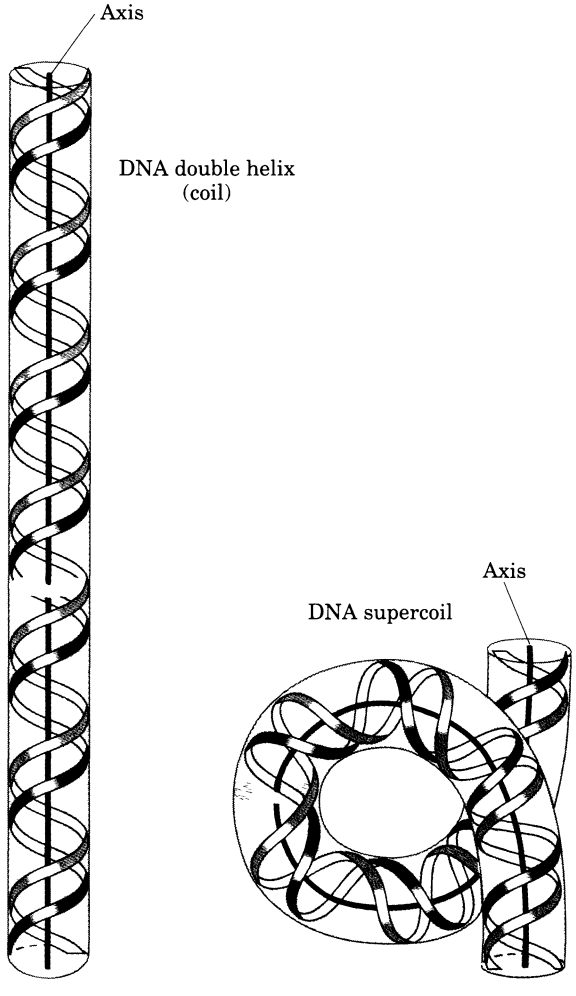


Figure 23–10 Supercoiling of DNA. Supercoiling is the twisting of the DNA axis upon itself.

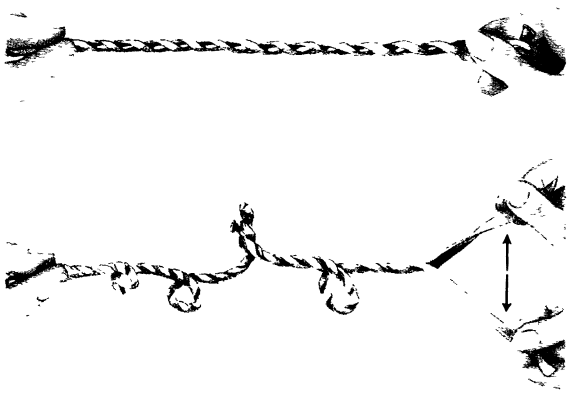


Figure 23–11 Supercoiling induced by separating the strands of a helical structure. Twist two linear strands of rubber band into a right-handed double-helix as shown. Fix the left end by having a friend hold onto it. If the two strands are pulled apart at the right end, the resulting strain will produce supercoiling as shown.

Replication and transcription both require a transient separation of the strands of DNA, and this is not a simple process in a DNA structure in which the two strands are helically interwound. Figure 23–11 illustrates this point.

That supercoiling must occur in cellular DNA would seem almost trivial were it not for one additional fact: many circular DNA molecules remain highly supercoiled even after they are purified from protein and other cellular components. Supercoiling is an important and intrinsic aspect of DNA tertiary structure that is ubiquitous in cellular DNAs and highly regulated by each cell.

A number of quantifiable properties of supercoiling have been established, the study of which 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 properties of an object that do not change under continuous deformations. In the case of DNA, a topological property is one that is not affected by twisting and turning of the DNA axis and can only be changed by breakage and rejoining of the DNA backbone. We now turn to an examination of the fundamental properties of supercoiling and the physical origin of the phenomenon itself.

Most Cellular DNA Is Underwound

To understand supercoiling we must now focus on the properties of small, circular DNAs such as plasmids and the DNAs derived from many small DNA viruses. When these DNAs contain no breaks in either strand, they are called **closed-circular DNAs.** If the DNA making up a closed-circular molecule conforms closely to the B-form structure (see Fig. 12–15), with one turn of the double helix for each 10.5 base pairs, the DNA will be relaxed rather than supercoiled (Fig. 23–12). Supercoiling is not a random process and does not occur unless the DNA is subject to some form of structural strain. When purified, however, closed-circular DNAs are rarely relaxed regardless of their biological origin. Furthermore, the degree of supercoiling tends to be well defined and characteristic of DNAs derived from a given cellular source. These facts suggest that the DNA structure is strained in some way to induce the supercoiling, and that the degree of strain introduced is regulated by the cell.

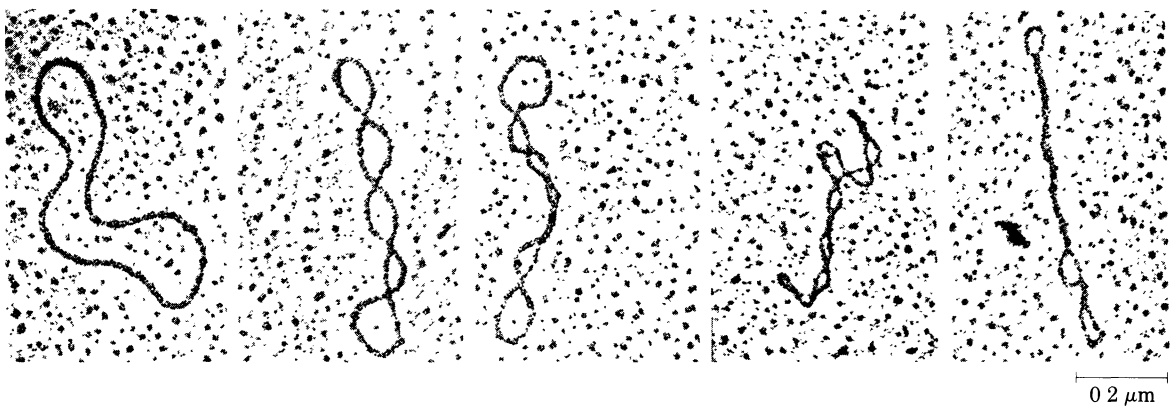


Figure 23–12 Electron micrographs of relaxed and supercoiled plasmid DNAs. The molecule at the left is relaxed, and the degree of supercoiling increases from left to right.

In almost every instance, the strain is a result of an **underwinding** of the DNA in the closed circle. In other words, there are *fewer* helical turns in the DNA than would be expected for the B-form structure. The effect of underwinding is illustrated in Figure 23–13 for an 84 base pair segment of a circular DNA. If the DNA were relaxed, this segment would contain eight double-helical turns, or one for every 10.5 base pairs. If one of these turns is removed, there will be 84/7 or about 12.0 base pairs per turn rather than the 10.5 found in B-DNA. This is a deviation from the most stable DNA form, and the molecule is thermodynamically strained as a result. The strain can be accommodated in one of two ways. First, the two strands can simply separate over the distance corresponding to one turn of B-DNA—10.5 base pairs (Fig. 23–13). Alternatively, the DNA can form a supercoil. When the axis of the DNA is twisted on itself in a certain manner, neighboring base pairs in underwound DNA can stack in positions that more closely approximate those they would assume in B-DNA.

Every cell actively underwinds its DNA with the aid of enzymatic processes to be described below. The resulting strained state of the DNA represents a form of stored energy. In isolated closed-circular DNA, strain introduced by underwinding generally is accommodated by supercoiling rather than strand separation, because twisting the axis of the DNA usually requires less energy than breaking the hydrogen bonds that stabilize paired bases. As we shall see below, however, the underwinding of DNA in vivo makes it easier to separate DNA strands and thereby gain access to the information they contain. Facilitating strand separation is one important reason for maintaining DNA in an underwound state.

The underwound state can be maintained only if the DNA is a closed circle or if it is bound and stabilized by proteins such that the strands are not free to rotate about each other. If there is a break in one of the strands of a 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, however, the number of helical turns present is fixed and cannot be changed without at least transiently breaking one of the DNA strands. The number of helical turns in DNA is quantifiable and leads to a more precise description of supercoiling.

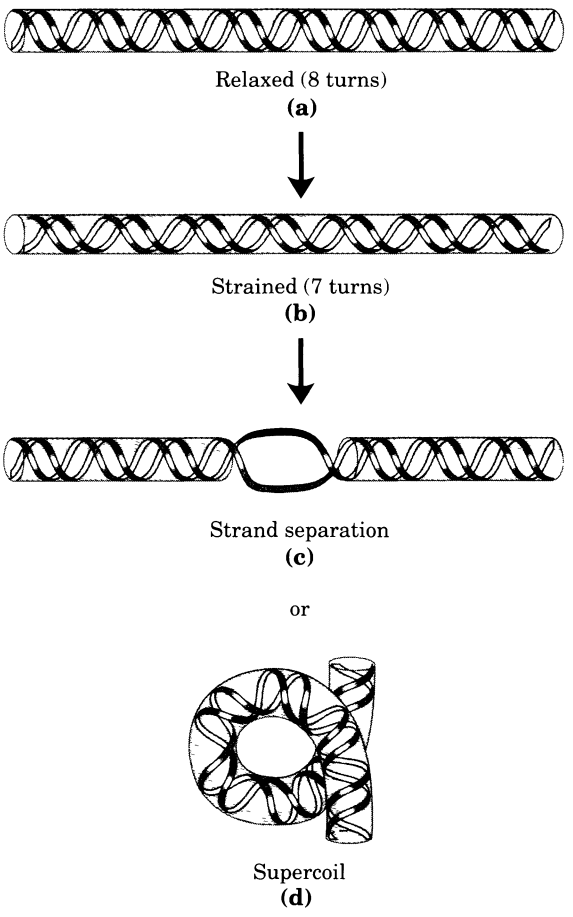
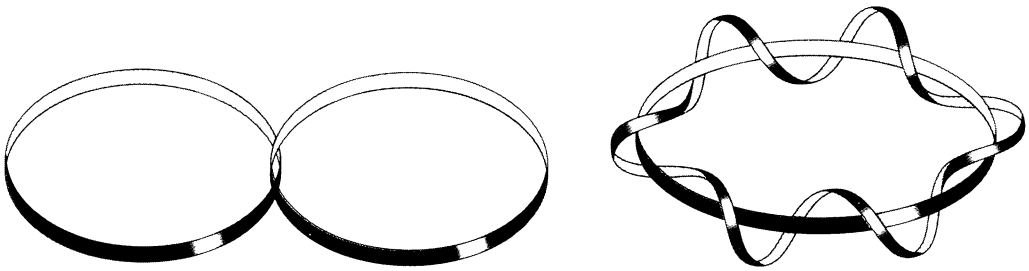


Figure 23–13 The effects of DNA underwinding. **(a)** A segment of DNA, 84 base pairs long, in its relaxed form with eight helical turns. **(b)** Removal of one turn induces structural strain that can be accommodated by **(c)** strand separation over 10.5 base pairs or by **(d)** formation of a supercoil.

DNA Underwinding Is Defined by Topological Linking Number

The branch of mathematics called topology provides a number of ideas that are useful in this discussion. Perhaps foremost among these is the concept of **linking number.** The linking number of a DNA molecule rigorously specifies the number of helical turns in a closed-circular DNA, in the absence of any supercoiling. Linking number is a topological property because it does not vary when double-stranded DNA is twisted or deformed in any way, as long as both DNA strands remain intact.

The concept of linking number (*Lk*) is illustrated in Figure 23–14. We begin by separating the two strands of a double-stranded circular DNA. If these two strands are linked as shown in Figure 23–14a, they are effectively joined by what can be described as a topological bond.



$Lk = 1$
(a)

$Lk = 6$
(b)

Figure 23–14 Linking number, Lk. The molecule in **(a)** has a linking number of 1. The molecule in **(b)** has a linking number of 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 one definition of linking number.

Even if all hydrogen bonds and base-stacking interactions are abolished such that the strands are not in physical contact, this topological bond will still link the two strands. If one of the circular strands is thought of as the boundary of an imaginary surface (much as a soap film might span the space framed by a circular wire), the linking number can be defined rigorously as the number of times the second strand pierces this surface. For the molecule in Figure 23–14a $Lk = 1$; for that in Figure 23–14b $Lk = 6$. 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 (+). Conversely, for strands interwound as a left-handed helix the linking number is negative (−). Given that left-handed Z-DNA occurs only rarely, negative linking numbers are not encountered in studies of DNA for all practical purposes.

We can now extend these ideas to a closed-circular DNA with 210 base pairs (Fig. 23–15). For a closed-circular DNA molecule that is relaxed, the linking number is simply the number of base pairs divided by 10.5; in this case, $Lk = 20$. 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, it is possible in principle to unravel the strands and separate them completely (Fig. 23–15b). Clearly, no topological bond exists in this case, and Lk is undefined.

We can now describe DNA underwinding in terms of changes in the linking number. The linking number in relaxed DNA is used as a reference and called Lk_0. In the molecule shown in Figure 23–15a, $Lk_0 = 20$; if two turns are removed from this molecule, Lk will equal 18. The change can be described by the equation

$$\Delta Lk = Lk - Lk_0 = 18 - 20 = -2$$

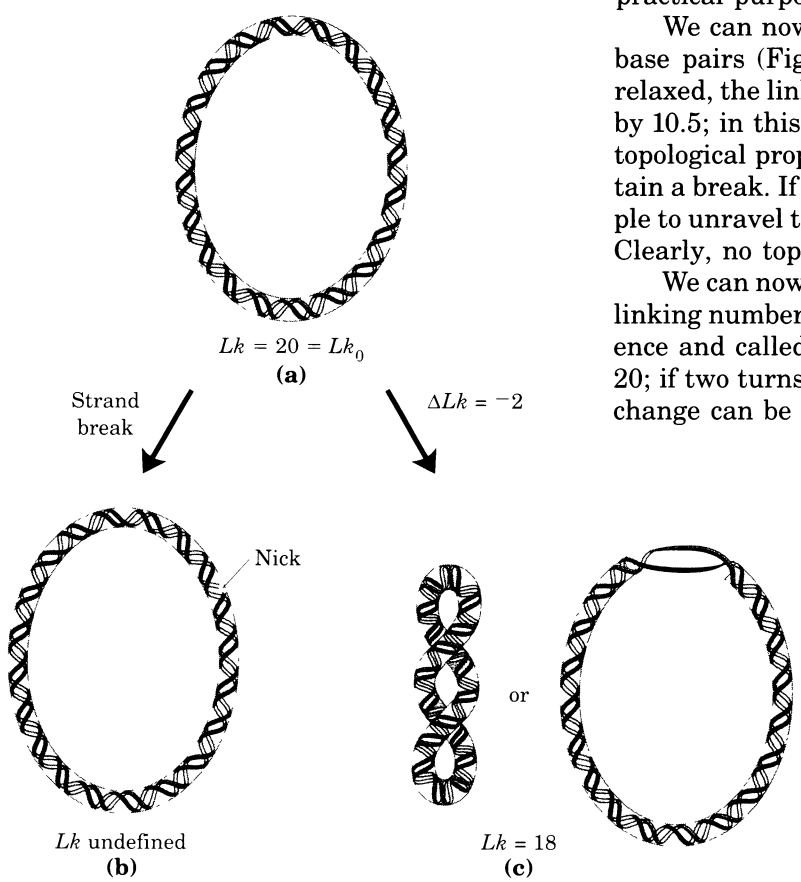


Figure 23–15 Linking number applied to closed-circular DNA molecules. A 210 base pair circular DNA is shown in three forms: **(a)** relaxed, $Lk = 20$; **(b)** relaxed with a nick (break) in one strand, Lk undefined; **(c)** underwound by two turns, $Lk = 18$. The underwound molecule can occur as a supercoiled (left) or strand-separated (right) structure.

It is often convenient to express the change in linking number in terms of a length-independent quantity called the **specific linking difference** (σ), which is a measure of the turns removed relative to those present in relaxed DNA. The term σ is also called the superhelical density and is defined as

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

In the example in Figure 23–15c, $\sigma = -0.10$, which means that 10% of the helical turns present in the DNA (in its B form) have been removed. The degree of underwinding in cellular DNAs generally falls into the range of 5 to 7%; that is, $\sigma = -0.05$ to -0.07. The negative sign of σ denotes that the change in linking number comes about as a result of underwinding the DNA. The supercoiling induced by underwinding is therefore defined as negative supercoiling. Conversely, under some conditions DNA can be overwound, and the resulting supercoiling is defined as positive. 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. 23–16). 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.

The 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 indeed on the number of atoms in the circular DNA molecule. Two forms of a given 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 (W_r) and twist (T_w) (Fig. 23–17). These are more difficult to describe intuitively than linking number, but to a first approximation W_r may be thought of as a measure of the coiling of the helix axis and T_w as determining the local twisting or spatial relationship of neighboring base pairs. When a change in linking number occurs, some of the resulting strain is usually compensated by writhe (supercoiling) and some by changes in twist, giving rise to the equation

$$Lk = T_w + W_r$$

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

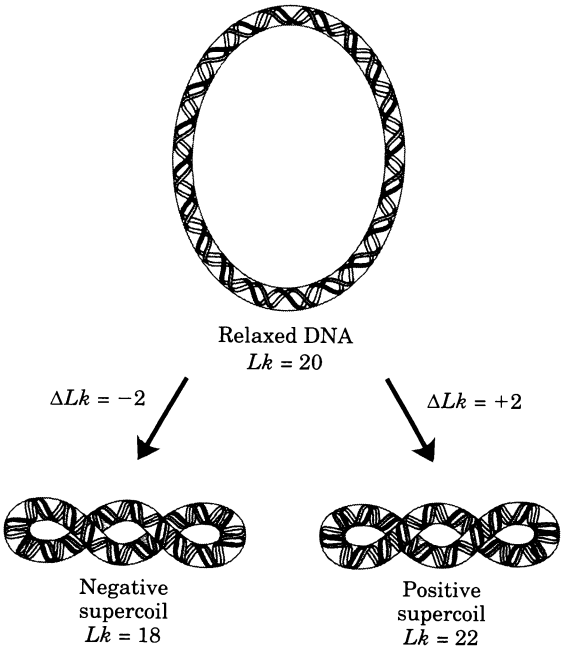


Figure 23–16 For the relaxed DNA molecule of Figure 23–15a, underwinding or overwinding by two helical turns (Lk = 18 or 22) will produce negative or positive supercoiling as shown. Note that the twisting of the DNA axis is opposite in sign in the two cases.

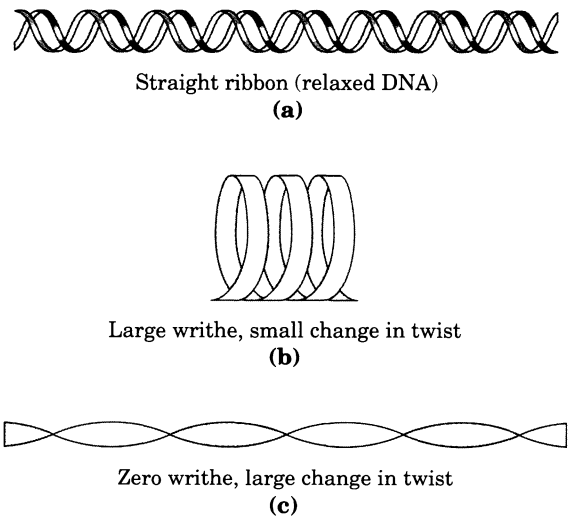


Figure 23–17 A ribbon model for illustrating twist and writhe. The ribbon in **(a)** represents the axis of a relaxed DNA molecule. Strain introduced by twisting the ribbon (underwinding the DNA) can be manifested as a change in writhe **(b)** or a change in twist **(c).** Changes in linking number are usually accompanied by changes in both writhe and twist.

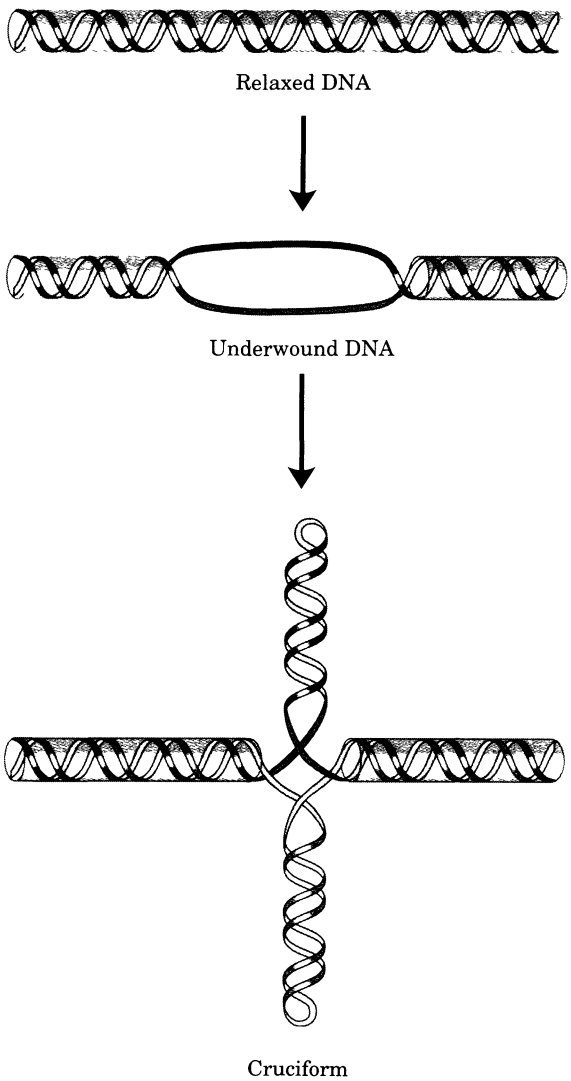


Figure 23–18 DNA underwinding promotes cruciform structures. In relaxed DNA, cruciforms seldom occur because the linear DNA accommodates more paired bases than does the cruciform structure. Underwinding the DNA facilitates the partial strand separation needed to promote cruciform formation at appropriate sequences (palindromes).

The concepts outlined above can be summarized by considering the supercoiling of a typical bacterial plasmid DNA. Plasmids are generally closed-circular DNA molecules. Because DNA is a right-handed helix, a plasmid will have a positive linking number. When the DNA is relaxed, the linking number or Lk_0 is simply the number of base pairs divided by 10.5. A typical plasmid, however, is generally underwound in the cell. Therefore, Lk is less than Lk_0, σ is negative, and the plasmid is negatively supercoiled. Typically for a bacterial plasmid, $\sigma = -0.05$ to -0.07.

Underwinding DNA facilitates a number of structural changes in the molecule. Strand separation occurs more readily in underwound DNA. This is critical to the processes of replication and transcription, and represents a major reason why DNA is maintained in an underwound state. Other structural changes are of less physiological importance but help illustrate the effects of underwinding. A cruciform (see Fig. 12–21) generally contains a few unpaired bases, and DNA underwinding helps to maintain the required strand separation (Fig. 23–18). In addition, underwinding a right-handed DNA helix facilitates the formation of short regions of left-handed Z-DNA, where the DNA sequence is consistent with Z-DNA formation (Chapter 12).

Topoisomerases Catalyze Changes in the Linking Number of DNA

In every cell, DNA supercoiling is a precisely regulated process that influences many aspects of DNA metabolism. Not surprisingly, there are enzymes in every cell whose sole purpose is to underwind and/or relax DNA. The enzymes that increase or decrease the extent of DNA underwinding are called **topoisomerases,** and the property of DNA they affect 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 1 topoisomerases act by transiently breaking one of the two DNA strands, rotating one of the ends about the unbroken strand, and rejoining the broken ends; they change Lk in increments of 1. Type 2 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. 23–19). A population of identical plasmid DNAs with the same linking number will migrate as a discrete band during electrophoresis. Topoisomers with Lk values differing by as little as 1 can be separated by this method. In this way changes in linking number induced by topoisomerases can readily be observed.

There are at least four different topoisomerases in *E. coli,* distinguished by Roman numerals I through IV. The type 1 topoisomerases (topoisomerases I and III) generally relax DNA by removing negative supercoils (they increase Lk). One bacterial type 2 enzyme, called topoisomerase II or, alternatively, DNA gyrase, can introduce negative supercoils (decrease Lk). It uses the energy of ATP and a surprising mechanism to accomplish this (Fig. 23–20). The superhelical density of bacterial DNA is balanced by regulation of the net activity of topoisomerases I and II.

Eukaryotic cells also have type 1 and type 2 topoisomerases; in most eukaryotes there is one known example of each type, called topoisomerase I and II, respectively. The type 2 enzymes in eukaryotic cells cannot underwind DNA (introduce negative supercoils), although both types can relax both positive and negative supercoils. We will consider one probable origin of negative supercoils in eukaryotic cells in our discussion of chromatin.

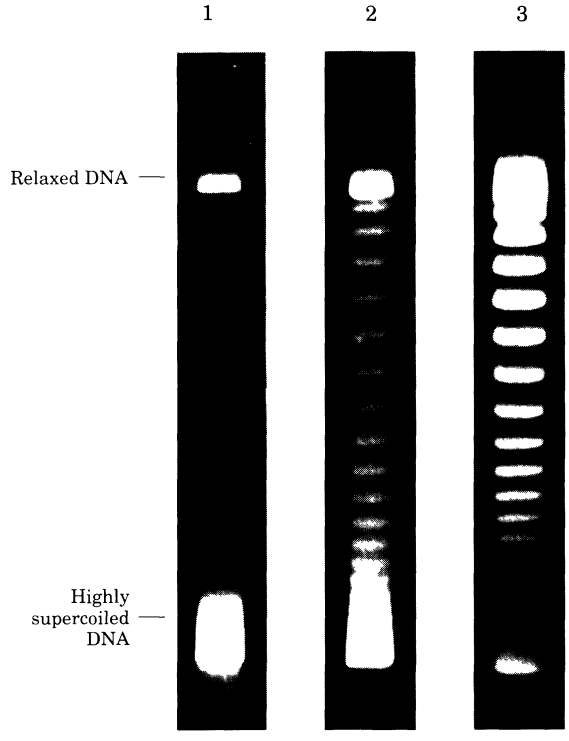


Figure 23–19 Circular DNA molecules that differ in linking number can be separated by gel electrophoresis. All the DNA molecules shown here have the same number of base pairs. Because supercoiled DNA molecules are more compact, they migrate more rapidly in a gel than the corresponding relaxed molecules. Gels such as those shown here separate topoisomers only over a limited range of superhelical density, so that highly supercoiled DNA migrates in a single band (lane 1) even though many different topoisomers may be 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).

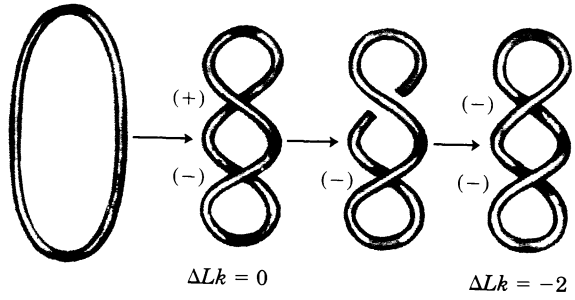


Figure 23–20 *E. coli* topoisomerase II (DNA gyrase) alters the linking number of circular DNA molecules by an unusual mechanism. Two regions of a DNA molecule are overlaid in a specific configuration in the bound complex (a positive (+) node). A compensating (−) node forms spontaneously elsewhere in the DNA molecule. As shown, both strands of one DNA segment are broken, the other segment is passed through the break, and the break is then resealed. The product now contains two minus nodes, and a comparison with Fig. 23–16 shows that the DNA now contains two negative supercoils. The change in structure reflects a change in Lk of −2.

DNA Compaction Requires a Special Form of Supercoiling

Supercoiled DNA molecules are remarkably uniform in many respects; the characteristic form is illustrated in Figure 23–21. Supercoils are right-handed in a negatively supercoiled DNA molecule (Fig. 23–16). Supercoiled DNA also tends to be extended and narrow rather than compacted, and it often exhibits multiple branches. At superhelical densities normally encountered in cells, the length of the supercoil axis, including branches, is about 40% the length of the DNA itself. This type of supercoiling is referred to as plectonemic (from the Greek *plektos,* "twisted," and *nema,* "thread") supercoiling.

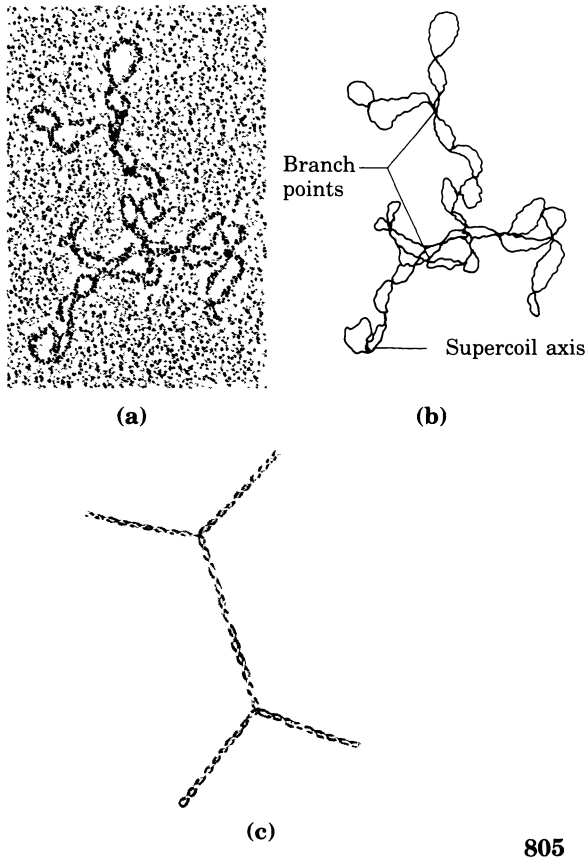


Figure 23–21 **(a)** An electron micrograph of plectonemically supercoiled plasmid DNA with **(b)** an interpretation of the observed structure. The blue lines define the axis of the supercoil. Note the branching of this molecule. **(c)** An idealized representation of this structure.

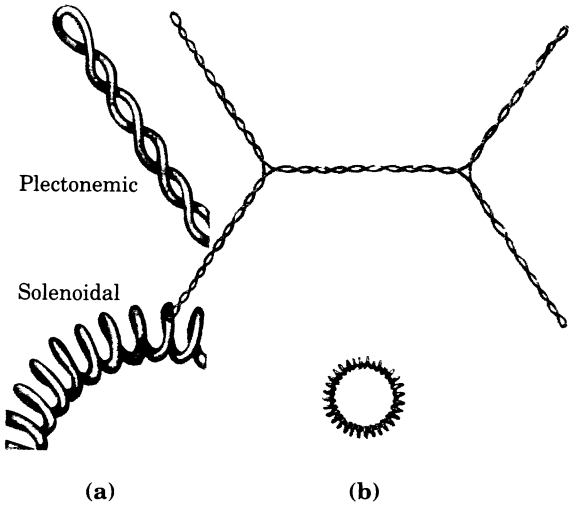


Figure 23–22 **(a)** Plectonemic and solenoidal forms of supercoiling. Solenoidal negative supercoiling takes the form of tight left-handed turns about an imaginary tubelike structure. The two forms are readily interconverted, although the solenoidal form is generally not observed unless certain proteins are bound to the DNA. **(b)** Plectonemic and solenoidal supercoiling of the same DNA molecule, drawn to scale. Note that solenoidal supercoiling provides a much greater degree of compaction.

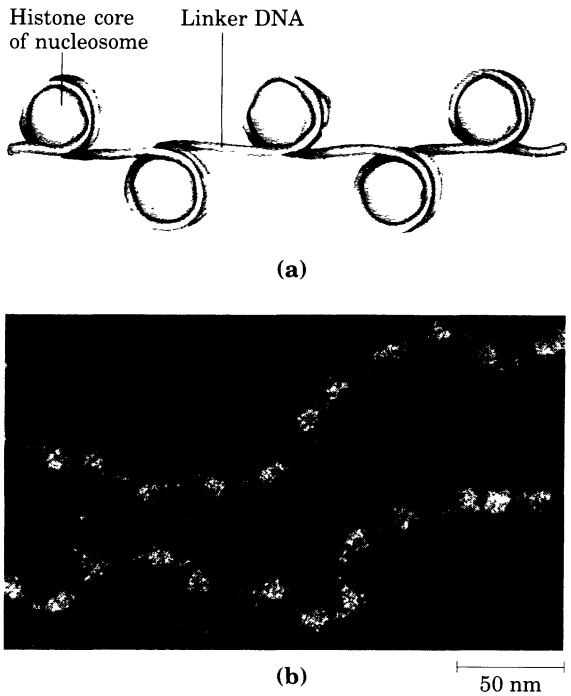


Figure 23–23 Regularly spaced nucleosomes, consisting of histone complexes bound to DNA. **(a)** Schematic illustration; **(b)** electron micrograph.

Although plectonemic coiling is the form observed in underwound DNAs in solution, it does not give the compaction required to package DNA in the cell. A second form of supercoiling, called **solenoidal supercoiling** (Fig. 23–22), can be adopted by an underwound DNA. Instead of the extended right-handed supercoils characteristic of the plectonemic form, solenoidal supercoiling involves tighter, left-handed turns. The structure is similar to that taken up by a garden hose neatly wrapped on a reel. Although their structures are dramatically different, plectonemic and solenoidal supercoiling represent two forms of negative supercoiling that can be taken up by the same 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. 23–22b). Solenoidal supercoiling explains how underwinding contributes to actual DNA compaction.

Chromatin and Nucleoid Structure

The term chromosome today refers to the nucleic acid molecule that is the repository of the genetic information of a virus, a bacterium, a eukaryotic cell, or an organelle. But the word chromosome was originally used in another sense, to refer to the densely colored bodies in eukaryotic nuclei that can be visualized with the light microscope after the cells are stained with a dye. Eukaryotic chromosomes, in the original sense of the word, appear as sharply defined bodies in the nucleus during the period just before and during mitosis, the process of nuclear division in somatic cells (see Fig. 2–14). In nondividing eukaryotic cells, the chromosomal material, called **chromatin,** is amorphous and appears to be randomly dispersed throughout the nucleus. But when the cells prepare to divide, the chromatin condenses and assembles itself into a species-specific number of well-defined chromosomes (see Fig. 23–4).

Chromatin has been isolated and analyzed. It consists of fibers that contain protein and DNA in approximately equal masses, plus 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. 23–23). Also found in chromatin are many nonhistone proteins, some of which regulate the expression of specific genes (Chapter 27). 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 the DNA packaging in bacterial cells.

Histones Are Small, Basic Proteins

Found in the chromatin of all eukaryotic cells, histones have molecular weights of 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). Five major classes of histones are found in all eukaryotic cells, differing in molecular weight and amino acid composition (Table 23–3). The H3 histones are nearly identical in amino acid sequence in all eukaryotes, as are the H4 histones, suggesting strict conservation of their functions. Comparing the 102 amino

The DNA molecules in chromosomes are the largest macromolecules in cells. Many smaller DNAs also occur in cells, in the form of viral DNAs, plasmids, and (in eukaryotes) mitochondrial or chloroplast DNAs. Many DNAs, especially those in bacteria, mitochondria, and chloroplasts, are circular. Viral and chromosomal DNAs have one major feature in common: they are generally much longer than the viral particles or cells in which they are packaged. The total DNA content of a eukaryotic cell is much greater than that of a bacterial cell.

Genes are segments of a chromosome that contain the information for a functional polypeptide or RNA molecule. In addition to these structural genes, chromosomes contain a variety of regulatory sequences involved in replication, transcription, and other processes. In eukaryotic chromosomes, there are two important special-function repetitive DNA sequences: centromeres, which are attachment points for the mitotic spindle, and telomeres, which occur at the ends of the linear chromosomes. Many genes in eukaryotic cells, and occasionally in bacteria, are interrupted by noncoding sequences called introns. The coding segments separated by introns are called exons.

Most cellular DNAs are supercoiled. Supercoiling is a manifestation of structural strain imparted by the underwinding of the DNA molecule. Underwinding is a decrease in the total number of helical turns in the DNA relative to the relaxed or B form. To maintain an underwound state, DNA must be a closed circle or be bound with protein. Supercoils resulting from underwinding are defined as negative supercoils. Underwinding is quantified by a topological parameter called linking number, Lk. The linking number of a relaxed, closed-circular DNA is used as a reference (Lk_0) and is equal to the number of base pairs divided by 10.5. Underwinding is measured in terms of the specific linking difference or σ, which equals $(Lk - Lk_0)/Lk_0$. For cellular DNAs, σ typically equals -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 for processes such as transcription or replication. The plectonemic supercoils in negatively supercoiled DNA in solution are right-handed, and the overall structure is narrow and extended. An alternative form called solenoidal supercoiling provides a much greater degree of compaction, and this form predominates in the cell.

DNAs that differ only in their linking number are called topoisomers. The enzymes that underwind and/or relax DNA are called topoisomerases, and they act by catalyzing changes in linking number. There are two classes, type 1 and type 2, which change Lk in increments of 1 or 2, respectively. In a bacterial cell, the superhelical density of the DNA represents a regulated balance between the activities of topoisomerases that increase and decrease linking number.

In the chromatin of eukaryotic cells, the fundamental unit of organization is the nucleosome, which consists of DNA and a protein particle containing eight histones, two copies each of histones H2A, H2B, H3, and H4. The segment of DNA (about 146 base pairs) wrapped around the protein core is in the form of a left-handed solenoidal supercoil. Nucleosomes are organized into 30 nm fibers, and the fibers themselves 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 large amounts of histone H1 and topoisomerase II. Bacterial chromosomes are also extensively compacted into a structure called a 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.

Further Reading

General

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J.D. (1989) *Molecular Biology of the Cell,* 2nd edn, Garland Publishing, Inc., New York.
An excellent general reference.

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

Singer, M. & Berg, P. (1991) *Genes and Genomes: A Changing Perspective,* University Science Books, Mill Valley, CA.
An up-to-date discussion of genes, chromosome structure, and many other topics.

Genes and Chromosomes

Blackburn, E.H. (1990) Telomeres: structure and synthesis. *J. Biol. Chem.* **265,** 5919–5921.

Jelinek, W.R. & Schmid, C.W. (1982) Repetitive sequences in eukaryotic DNA and their expression. *Annu. Rev. Biochem.* **51,** 813–844.

Murray, A.W. & Szostak, J.W. (1987) Artificial chromosomes. *Sci. Am.* **257** (November), 62–68.

Novick, R.P. (1980) Plasmids. *Sci. Am.* **243** (December), 102–127.

Sharp, P.A. (1985) On the origin of RNA splicing and introns. *Cell* **42,** 397–400.

Ullu, E. & Tschudi, C. (1984) *Alu* sequences are processed 7SL RNA genes. *Nature* **312,** 171–172.

Supercoiling and Topoisomerases

Bauer, W.R., Crick, F.H.C., & White, J.H. (1980) Supercoiled DNA. *Sci. Am.* **243** (July), 118–133.

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.

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.
This provides 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.

Liu, L.F. (1989) DNA topoisomerase poisons as antitumor drugs. *Annu. Rev. Biochem.* **58,** 351–375.
A review of eukaryotic topoisomerases and the use of topoisomerase inhibitors in cancer chemotherapy.

Wang, J.C. (1985) DNA topoisomerases. *Annu. Rev. Biochem.* **54,** 665–697.

Wang, J.C. (1991) DNA topoisomerases: Why so many? *J. Biol. Chem.* **266,** 6659–6662.
A good short summary of topoisomerase functions.

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.

Kornberg, R.D. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science* **184,** 868–871.
The classic paper that introduced the subunit model for chromatin.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., & Klug, A. (1984) Structure of the nucleosome core particle at 7Å resolution. *Nature* **311,** 532–537.

van Holde, K.E. (1989) *Chromatin,* Springer-Verlag, New York.

Problems

1. *How Long Is the Ribonuclease Gene?* What is the minimum number of nucleotide pairs in the gene for pancreatic ribonuclease (124 amino acids long)? Suggest a reason why the number of nucleotide pairs in the gene might be much larger than your answer.

2. *Packaging of DNA in a Virus* The DNA of bacteriophage T2 has a molecular weight of 120×10^6. The head of the T2 phage is about 210 nm long. Assuming the molecular weight of a nucleotide pair is 650, calculate the length of T2 DNA and compare it with the length of the T2 head. Your answer will show the necessity of very compact packaging of DNA in viruses (see Fig. 23–1).

3. *The DNA of Phage M13* Bacteriophage M13 DNA has the following base composition: A, 23%; T, 36%; G, 21%; C, 20%. What does this information tell us about the DNA of this phage?

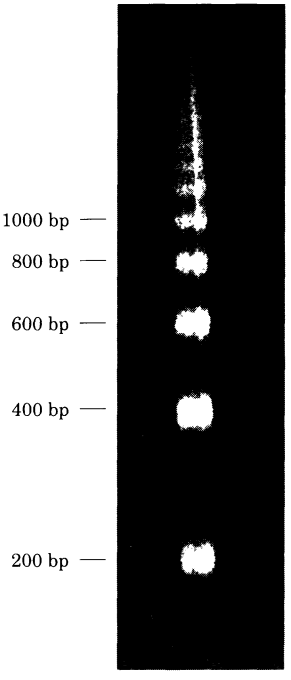
4. *Base Composition of ϕX174 DNA* Bacteriophage ϕX174 DNA occurs in two forms, single-stranded in the isolated virion and double-stranded during viral replication in the host cell. Would you expect them to have the same base composition? Give your reasons.

5. *Size of Eukaryotic Genes* An enzyme present in rat liver has a polypeptide chain of 192 amino acid residues. It is coded for by a gene having 1,440 base pairs. Explain the relationship between the number of amino acid residues in this enzyme and the number of nucleotide pairs in its gene.

6. *DNA Supercoiling* A covalently closed circular DNA molecule has an *Lk* of 500 when it is relaxed. Approximately how many base pairs are in this DNA? How will the linking number be altered (increase, decrease, no change, become undefined) if (a) a protein complex is bound to form a nucleosome, (b) one DNA strand is broken, (c) DNA gyrase is added with ATP, or (d) the double helix is denatured (base pairs are separated) by heat?

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

8. *Chromatin* One of the important early pieces of evidence that helped define the structure of the nucleosome is illustrated by the agarose gel shown below, in which the thick bands represent DNA. It was generated by treating chromatin briefly 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 (in base pairs) would migrate. What does this gel tell you about chromatin structure? Why are the DNA bands thick and spread out rather than sharp?



CHAPTER

DNA Metabolism

As the repository of genetic information, DNA occupies a unique and central place among biological macromolecules. The nucleotide sequences of DNA ultimately describe the primary structures of all cellular RNAs and proteins, and through enzymes can indirectly affect the synthesis of all other cellular constituents, determining the size, shape, and function of every living thing.

The structure of DNA is a marvelous device for the stable storage of genetic information. The phrase "stable storage," however, conveys a static and incomplete picture of the biochemical role of DNA in the cell. A proper description of DNA function must also explain how that information is transmitted from one generation of cells to the next. The term "DNA metabolism" can be used to describe the process by which faithful copies of DNA molecules are made (replication), along with the processes that affect the structure of the information within (repair and recombination). Together they are the focus of this chapter.

Perhaps more than any other factor, it is the requirement for an exquisite degree of accuracy that shapes these processes. At the level of joining one nucleotide to the next, the chemistry of DNA replication is simple and elegant, almost deceptively so. But as we will see, the synthesis of all macromolecules that contain information involves complex devices to ensure that the information is transmitted intact. If left uncorrected, errors in DNA synthesis can have dire consequences because they are essentially permanent. The enzymes that synthesize DNA must copy DNA molecules that often contain millions of bases, and they do so with great fidelity and speed. They must also act on a DNA substrate that is highly compacted and bound with other proteins. The enzymes that catalyze the formation of phosphodiester bonds are therefore only part of an elaborate system involving myriad proteins and enzymes.

The importance of maintaining the integrity of the information stored in DNA is underscored when the discussion turns to repair. As detailed in Chapter 12, DNA is susceptible to many types of damaging reactions. Though generally slow, they are nevertheless significant because of the very low biological tolerance for changes in DNA sequence. DNA is the only macromolecule for which repair systems exist, and their number, diversity, and complexity reflect the wide range of insults to which a DNA molecule is subject.

The processes by which genetic information is rearranged, collectively called recombination, seem to belie the principles just established. If the integrity of the genetic information is paramount, why rearrange it? One explanation seems to be the need for maintaining a

level of genetic diversity by providing new combinations of alleles, the alternative forms of a single gene. Even without this explanation, however, recombination is not really so renegade a set of processes. Most recombination events are conservative in the sense that genetic information is neither lost nor gained. Indeed, with a closer look at a recombination event, one often finds a DNA repair or gene regulation process in disguise.

Special emphasis is given in this chapter to the enzymes that catalyze these processes. They are well worth getting acquainted with if for no other reason than their everyday use as reagents in a wide range of modern biochemical technologies. Because many of the seminal discoveries in DNA metabolism have been made with *E. coli,* the well-understood enzymes obtained from this bacterium are generally used here to illustrate the ground rules. A quick look at the relevant genes on the *E. coli* genetic map (Fig. 24–1) provides just a hint of what is to come.

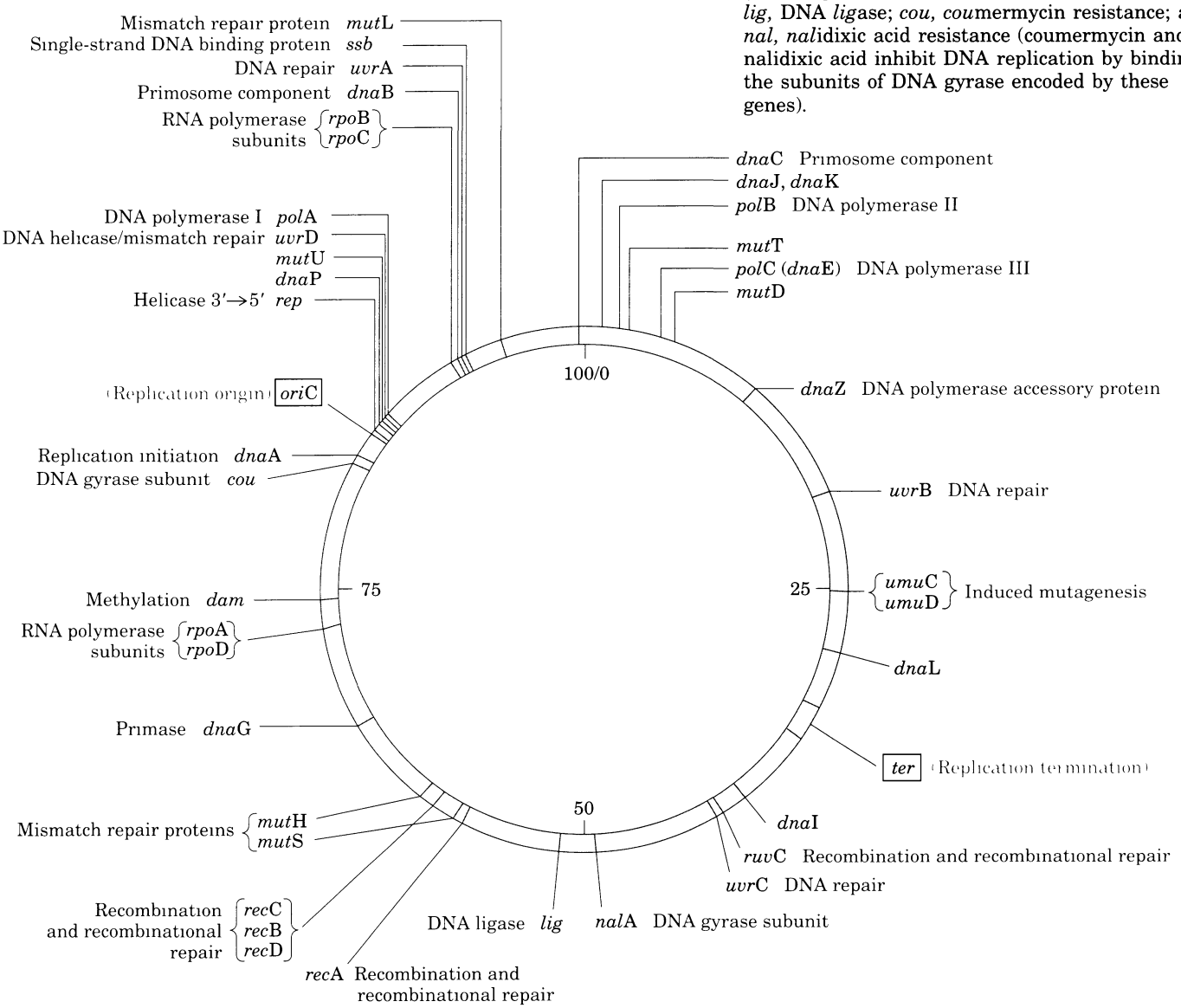


Figure 24–1 A map of the *E. coli* chromosome, showing the relative positions of genes encoding some of the proteins important in DNA metabolism. The number of known genes involved provides a hint of the complexity of these processes. The numbers 0 to 100 denote a genetic measurement called minutes, with each minute corresponding to about 40,000 base pairs. The acronyms consisting of three lowercase letters generally reflect some aspect of the gene's function. These include *mut,* *mut*agenesis; *dna,* DNA replication; *pol,* DNA *pol*ymerase; *rpo,* *R*NA *po*lymerase; *uvr,* *UV*-resistance; *rec,* *rec*ombination; *ter,* *ter*mination of replication; *ori,* *ori*gin of replication; *dam,* *D*NA *a*denine *m*ethylation; *lig,* DNA *lig*ase; *cou,* *cou*mermycin resistance; and *nal,* *nal*idixic acid resistance (coumermycin and nalidixic acid inhibit DNA replication by binding to the subunits of DNA gyrase encoded by these genes).

Before moving on to replication, we must entertain two short digressions. The first concerns the use of acronyms in naming genes and proteins. Bacterial genetics is a powerful tool that has facilitated much of the work described in this chapter. Bacterial genes that affect a given cellular process such as replication often have been identified before the roles of their protein products were understood. By convention, acronyms used to identify bacterial (and sometimes eukaryotic) genes are generally three lowercase, italicized letters that reflect function, such as *dna, uvr,* or *rec* for genes that affect *DNA* replication, *r*esistance to the damaging effects of *UV* radiation, or *rec*ombination, respectively. In the case of multiple genes that affect the same process, the designation A, B, C, etc., is added, usually reflecting the temporal order of gene discovery rather than a reaction sequence. In most cases, the protein product of each gene is ultimately isolated and characterized. Sometimes the product is identified as a previously isolated protein. The *dna*E gene, for example, was found to encode the polymerizing subunit of DNA polymerase III; consequently, the *dna*E gene was renamed *pol*C to reflect that function more clearly. In many cases the protein product has turned out to be novel, with an activity not easily described by a simple enzyme name. In a practice that can be confusing, these proteins often retain the name of their genes; for example, the products of the *dna*A and *rec*A genes are simply called the DnaA and RecA proteins, respectively. Many examples of this practice are found in this chapter. Here we use the convention that names in italics refer to genes or important DNA sequences, and roman type is used when the name refers to a protein.*

The second digression is needed to introduce enzymes that degrade DNA rather than synthesize it, because directed DNA degradation plays a significant role in all of the processes described in this chapter. These enzymes are called **nucleases** or, alternatively, **DNases** if they are specific for DNA. Every cell contains several different nucleases, and these fall into two broad classes: exonucleases and endonucleases. **Exonucleases** degrade DNA from one end of the molecule. Many are specific for degradation in either the 5′→3′ or 3′→5′ direction; that is, they remove nucleotides specifically from the 5′ or 3′ end, respectively, of one strand of a double-stranded nucleic acid (see Fig. 12–7). **Endonucleases** act in the interior of nucleic acids, reducing them to smaller and smaller fragments. A few exonucleases and endonucleases degrade only single-stranded DNA. There are also a few important classes of endonucleases that cleave only at specific nucleotide sequences (e.g., the restriction endonucleases considered in Chapter 28). Many types of nucleases will be encountered in this and subsequent chapters.

DNA Replication

Long before the structure of DNA became known, scientists had wondered first at the ability of organisms to create reasonable copies of themselves, and later at the ability of cells to produce many identical

* For eukaryotic proteins, these naming conventions are somewhat different and vary sufficiently from one organism to the next that no single convention can be presented here.

copies of large and complex macromolecules. Speculation about these problems centered around the concept of a **template.** The molecular template had to be a surface upon which molecules could be lined up in a specific order and joined to create a macromolecule with a unique structure and function.

The process of DNA replication provided the first biological example of the use of a molecular template to guide the synthesis of a macromolecule. The 1940s brought the revelation that DNA was the genetic molecule, but not until James Watson and Francis Crick deduced its structure did it become clear how DNA could act as a template for the replication and transmission of genetic information. *One strand is the complement of the other.* The strict base-pairing rules mean that the use of one strand as a template will result in another strand with a predictable, complementary sequence.

The fundamental properties of the DNA replication process and the mechanisms used by the enzymes that catalyze it have proven to be essentially identical in all organisms. This mechanistic unity will be a major theme as we proceed from general properties of the replication process to *E. coli* replication enzymes and finally to replication in eukaryotes.

DNA Replication Is Governed by a Set of Fundamental Rules

DNA Replication Is Semiconservative If each DNA strand serves as a template for the synthesis of a new strand, two new DNA molecules will result, each with one new strand and one old strand. This is called **semiconservative replication.**

The hypothesis of semiconservative replication was proposed by Watson and Crick soon after publication of their paper on the structure of DNA; the theory was proven in ingeniously designed experiments by Matthew Meselson and Franklin Stahl in 1957 (Fig. 24–2). Meselson and Stahl grew *E. coli* cells for many generations in a medium in which the sole nitrogen source (NH_4Cl) contained ^{15}N, the "heavy" isotope of nitrogen, instead of the normal, more abundant "light" isotope ^{14}N. The DNA isolated from these cells had a density about 1% greater than that of normal [^{14}N]DNA. Although this is only a small difference, a mixture of heavy [^{15}N]DNA and light [^{14}N]DNA can be separated by centrifugation to equilibrium in a cesium chloride density gradient.

The *E. coli* cells grown in the ^{15}N medium were transferred to a fresh medium containing only the ^{14}N isotope, where they were allowed to grow until the cell population had just doubled. The DNA isolated from these first-generation cells formed a single band in the CsCl gradient at a position indicating that the double-helical DNAs of the daughter cells were hybrids containing one new ^{14}N strand and one parental ^{15}N strand (Fig. 24–2).

This result argued against conservative replication, an alternative hypothesis in which one progeny DNA molecule would consist of two newly synthesized DNA strands and the other would contain the two parental strands; this would never yield hybrid DNA molecules in the Meselson–Stahl experiment. The semiconservative replication hypothesis was further supported in the next step of the experiment. Cells were allowed to double in number again in the ^{14}N medium, and the isolated DNA product of this second cycle of replication exhibited *two* bands, one having a density equal to that of light DNA and the other having the density of the hybrid DNA observed after the first cell doubling.

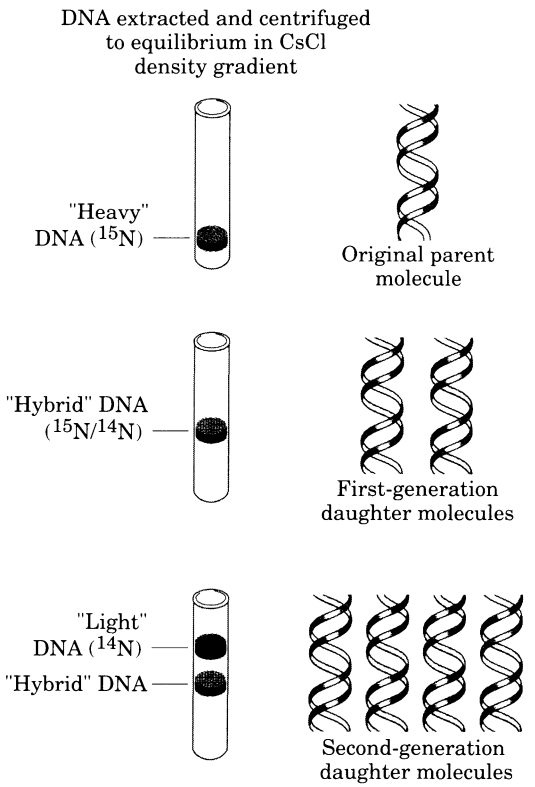
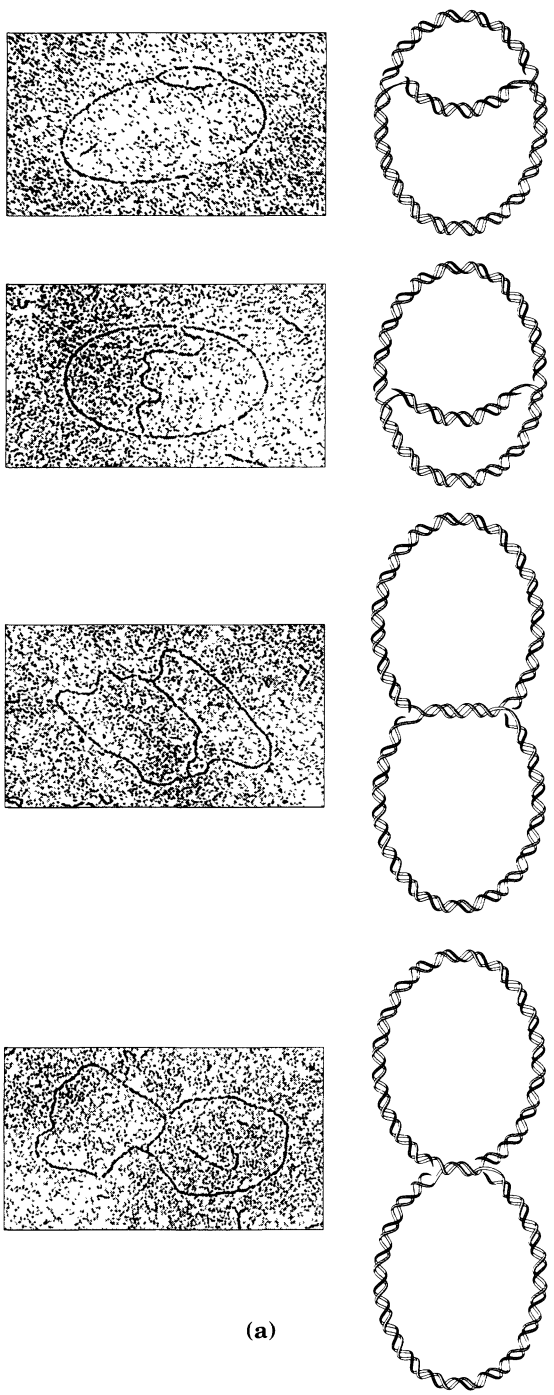
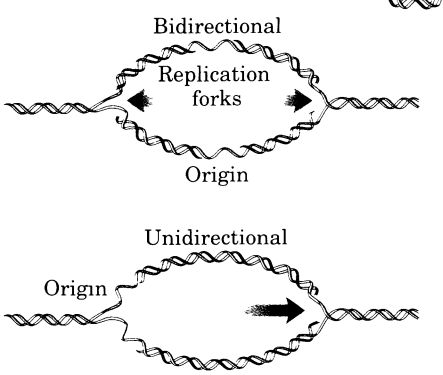


Figure 24–2 The Meselson–Stahl experiment was designed to distinguish between two alternative DNA replication mechanisms. Cells were grown for many generations in a medium containing only heavy nitrogen, ^{15}N, so that all the nitrogen in the DNA was ^{15}N. The cells were then transferred to a medium containing only light nitrogen, ^{14}N, and the density of the DNA was monitored closely for the next two cell generations. Cellular DNA was isolated after the first and second generations and centrifuged to equilibrium in a CsCl density gradient. The [^{15}N]DNA (shown in blue) came to equilibrium at a lower position in the CsCl gradient than [^{14}N]DNA (shown in red). Hybrid DNA equilibrated in an intermediate position. If DNA replication were conservative, each of the two heavy strands of parental DNA would be replicated to yield the original heavy duplex DNA and a DNA duplex containing two new light strands. Continuation of conservative replication would yield in the next generation one heavy DNA and three light DNAs but no hybrid DNAs. The Meselson–Stahl experiment, however, showed that replication is semiconservative, resulting in two daughter duplexes each containing one parental heavy strand and one new light strand. The next generation yielded two hybrid DNAs and two light DNAs.



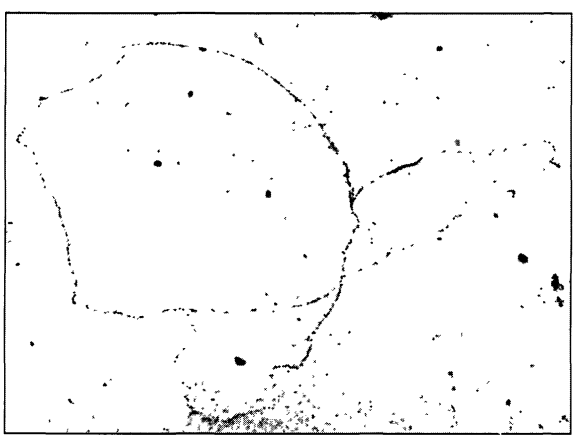
(a)



(b)

Figure 24–3 Replication of a circular chromosome produces a structure resembling the Greek letter theta (θ). **(a)** Labeling with tritium (^3H) shows that both strands are replicated at the same time (new strands shown in red). The electron micrographs illustrate the replication of a circular *E. coli* plasmid as visualized by autoradiography. **(b)** Addition of ^3H for a short period just before the reaction is stopped allows a distinction to be made between unidirectional and bidirectional replication, by determining whether label (in red) is found at one or both replication forks seen in autoradiograms. This technique has revealed bidirectional replication in *E. coli, B. subtilis,* and other bacteria. **(c)** Autoradiogram of a replicating *E. coli* chromosome taken from a culture grown for two generations in [^3H]thymidine.

Replication Begins at an Origin and Usually Proceeds Bidirectionally A host of questions now arises. Are the parental DNA strands completely unwound before each is replicated? Does replication begin at random places or at a unique point? After initiation at any point in the DNA, does replication proceed in one direction or both? An early indication that replication is a highly coordinated process in which the parental strands are unwound and replicated simultaneously was provided by John Cairns using the technique of autoradiography. He made the DNA of *E. coli* cells radioactive by growing them in a medium containing thymidine labeled with tritium (^3H). When the DNA was carefully isolated, spread, and overlaid with a photographic emulsion, and left for several weeks, the radioactive thymidine residues generated "tracks" of silver grains in the emulsion, producing an image of the DNA molecule. These tracks revealed that the intact chromosome of *E. coli* is a single giant circle, 1.7 mm long (see Fig. 23–2). Radioactive DNA isolated from cells during replication showed an extra radioactive loop (Fig. 24–3). The amount of radioactivity in the loop relative to the remainder of the DNA led Cairns to conclude that the loop in the DNA was the result of the formation of two radioactive daughter strands, each complementary to a parent strand. One or both ends of the loop are dynamic points, termed **replication forks,** where parental DNA is being unwound and the separated strands quickly repli-



(c)

cated. This demonstrated that both DNA strands are replicated simultaneously, and a variation of this experiment (Fig. 24–3b) indicated that replication of bacterial chromosomes is bidirectional: both ends of the loop have active replication forks.

To determine whether the loops originated at a unique point in the DNA, landmarks were needed in the DNA "string." These were provided by a technique called **denaturation mapping,** developed by Ross Inman and colleagues. Using the 48,502 base pair chromosome from bacteriophage λ, Inman showed that DNA could be selectively denatured at sequences unusually rich in A=T base pairs. This generates a reproducible pattern of single-stranded bubbles (see Fig. 12–30). When isolated DNAs containing replication loops are partially denatured in this way, the progress of the replication forks can be measured and mapped using the denatured regions as points of reference. The technique revealed that the replication loops always initiate at a unique point, called an **origin.** In addition, this work reinforced the earlier observation that replication is usually bidirectional. For circular DNA molecules, the two replication forks meet at a point on the side of the circle opposite to the origin.

DNA Synthesis Proceeds in a 5′→3′ Direction and Is Semidiscontinuous A new strand of DNA is always synthesized in the 5′→3′ direction (the 5′ and 3′ ends of a DNA strand are defined as shown in Figure 12–7). Because the two DNA strands are antiparallel, the strand acting as template is being read from its 3′ end toward its 5′ end.

If synthesis always proceeds in the 5′→3′ direction, how can both strands be synthesized simultaneously? If both were synthesized continuously as the replication fork moved, one would have to undergo 3′→5′ synthesis. This problem was resolved by Reiji Okazaki and colleagues in the 1960s. Okazaki found that one of the new DNA strands is synthesized in short pieces, now called **Okazaki fragments.** This work ultimately led to the conclusion that one strand is synthesized continuously and the other discontinuously (Fig. 24–4). The continuous or **leading strand** is the one in which 5′→3′ synthesis proceeds in the same direction as replication fork movement. The discontinuous or **lagging strand** is the one in which 5′→3′ synthesis proceeds in the direction opposite to the direction of fork movement. Okazaki fragments range in length from a few hundred to a few thousand nucleotides, depending on the cell type.

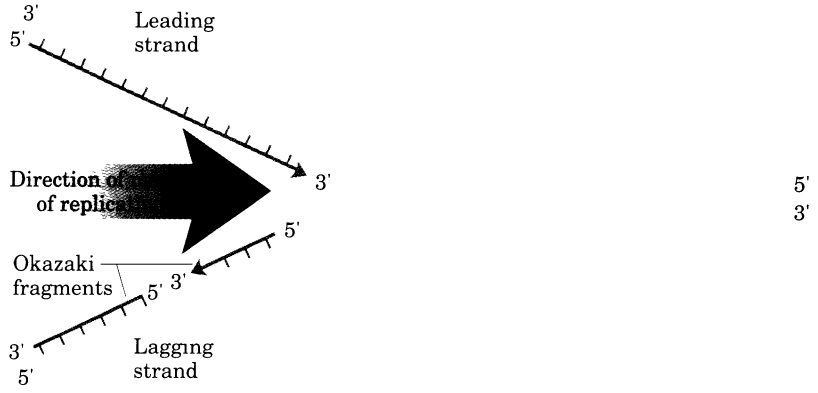
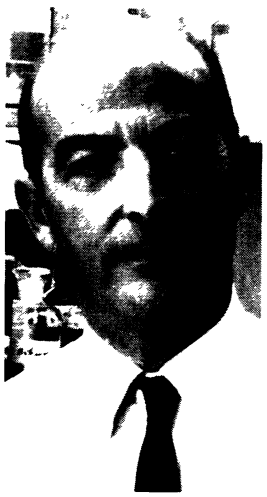


Figure 24–4 A new DNA strand (red) is always synthesized in the 5′→3′ direction. The template is copied in the opposite direction: 3′→5′. The strand that is continuously synthesized (in the direction taken by the replication fork) is the leading strand. The other strand, the lagging strand, is synthesized discontinuously in short pieces (Okazaki fragments) in a direction opposite to the direction of replication fork movement. The Okazaki fragments are then spliced together by DNA ligase. In bacteria the Okazaki fragments are about 1,000 to 2,000 nucleotides long. In eukaryotic cells they are 150 to 200 nucleotides long.



Arthur Kornberg

DNA Is Synthesized by DNA Polymerases

The search for an enzyme that could synthesize DNA was initiated in 1955 by Arthur Kornberg and colleagues. This work led to the purification and characterization of DNA polymerase from *E. coli* cells, a single-polypeptide enzyme now called **DNA polymerase I** (M_r 103,000). Much later, it was found that *E. coli* contains at least two other distinct DNA polymerases, which will be described below.

Detailed studies of DNA polymerase I revealed features of the DNA synthetic process that have proven to be common to all DNA polymerases. The fundamental reaction is a nucleophilic attack by the 3′-hydroxyl group of the nucleotide at the 3′ end of the growing strand on the 5′-α-phosphorus of the incoming deoxynucleoside 5′-triphosphate (Fig. 24–5). Inorganic pyrophosphate is released in the reaction. The general reaction equation is

$$\underset{\text{DNA}}{(\text{dNMP})_n} + \text{dNTP} \longrightarrow \underset{\substack{\text{Lengthened}\\ \text{DNA}}}{(\text{dNMP})_{n+1}} + \text{PP}_i \qquad (24\text{–}1)$$

where dNMP and dNTP are deoxynucleoside 5′-monophosphate and 5′-triphosphate, respectively.

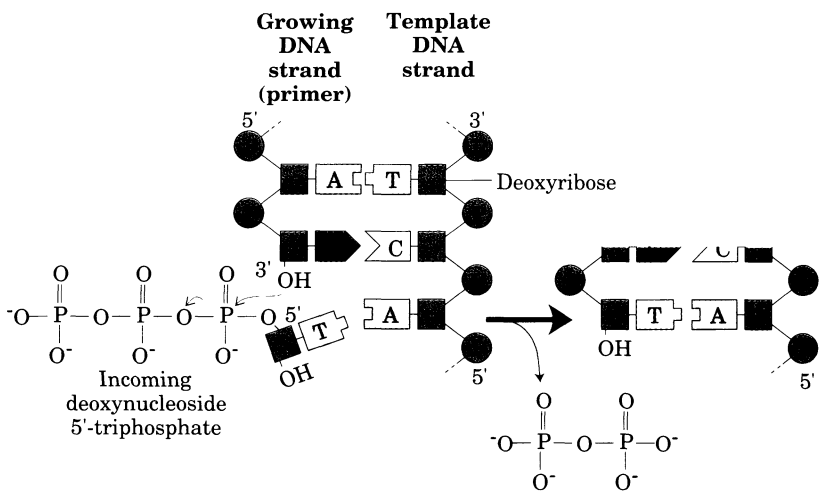


Figure 24–5 Elongation of a DNA chain. A single unpaired strand is required to act as template, and a primer strand is needed to provide a free 3′ end to which new nucleotide units are added. Each incoming nucleotide is selected by virtue of base pairing to the appropriate nucleotide in the template strand.

Early work on DNA polymerase I led to the definition of two central requirements for DNA polymerization. First, all DNA polymerases require a **template** (Fig. 24–5). The polymerization reaction is guided by a template DNA strand according to the base-pairing rules predicted by Watson and Crick: where a guanine is present in the template, a cytosine is added to the new strand, and so on. This was a particularly important discovery, not only because it provided a chemical basis for semiconservative DNA replication, but because it represented the first example of the use of a template to guide a biosynthetic reaction. Second, a **primer** is required. A primer is a segment of new strand (complementary to the template) with a 3′-hydroxyl group to which nucleotides can be added. The 3′ end of the primer is called the **primer terminus.** In other words, part of the new strand must already be in place; the polymerase can only add nucleotides to a preexisting strand. This has proven to be the case for all DNA polymerases,

and this discovery provided an interesting wrinkle in the DNA replication story. No DNA-synthesizing enzyme can initiate synthesis of a new DNA strand. As we will see later in this chapter, enzymes that synthesize RNA do have the capability of initiating synthesis, and as a consequence, primers are often oligonucleotides of RNA.

After a nucleotide is added to a growing DNA strand, the DNA polymerase must either dissociate or move along the template and add another nucleotide. Dissociation and reassociation of the polymerase can limit the overall reaction rate, thus the rate generally increases if a polymerase adds additional nucleotides without dissociating from the template. The number of nucleotides added, on average, before a polymerase dissociates is defined as its **processivity.** DNA polymerases vary greatly in processivity, with some adding just a few nucleotides and others adding many thousands before dissociation occurs.

Polymerization Is a Thermodynamically Favorable Reaction

Throughout this book we have emphasized the importance of noncovalent as well as covalent interactions in biochemical processes. A discussion of the energetics of the polymerization reaction can be deceptive if only the covalent bonds are considered. The rearrangement of covalent bonds is straightforward: one phosphoric anhydride bond (in the dNTP) is hydrolyzed and one phosphodiester bond (in the DNA) is formed. This results in a slightly positive (unfavorable) change in standard free energy ($\Delta G^{\circ\prime} \approx 2$ kJ/mol) for the overall reaction shown in Equation 24–1. Hydrolysis of the pyrophosphate to two molecules of inorganic phosphate by the pyrophosphatases present in all cells yields a $\Delta G^{\circ\prime}$ of -30 kJ/mol, and by coupling these two reactions the cell can provide a strong thermodynamic pull in the direction of polymerization, with a net $\Delta G^{\circ\prime}$ of -28 kJ/mol. This is important to the cell, but in this case it is not the whole story. If this calculation were complete, polymerases would tend to catalyze DNA degradation in the absence of pyrophosphate hydrolysis. Purified DNA polymerases, however, carry out polymerization very efficiently in vitro in the absence of pyrophosphatases. The explanation of this seeming paradox now is clear: noncovalent interactions not considered in the calculation above make an important thermodynamic contribution to the polymerization reaction. Every new nucleotide added to the growing chain is held there not just by the new phosphodiester bond but also by hydrogen bonds to its partner in the template and base-stacking interactions with the adjacent nucleotide in the same chain (p. 330). The additional energy released by these multiple weak interactions helps drive the reaction in the direction of polymerization.

DNA Polymerases Are Very Accurate

Replication *must* proceed with a very high degree of fidelity. In *E. coli*, a mistake is made only once for every 10^9 to 10^{10} nucleotides added. For the *E. coli* chromosome of about 4.7×10^6 base pairs, this means that an error will be made only once per 1,000 to 10,000 replications. During polymerization, discrimination between correct and incorrect nucleotides relies upon the hydrogen bonds that specify the correct pairing between complementary bases. Incorrect bases will not form the correct hydrogen bonds and can be rejected before the phosphodiester bond is formed. The accuracy of the polymerization reaction itself,

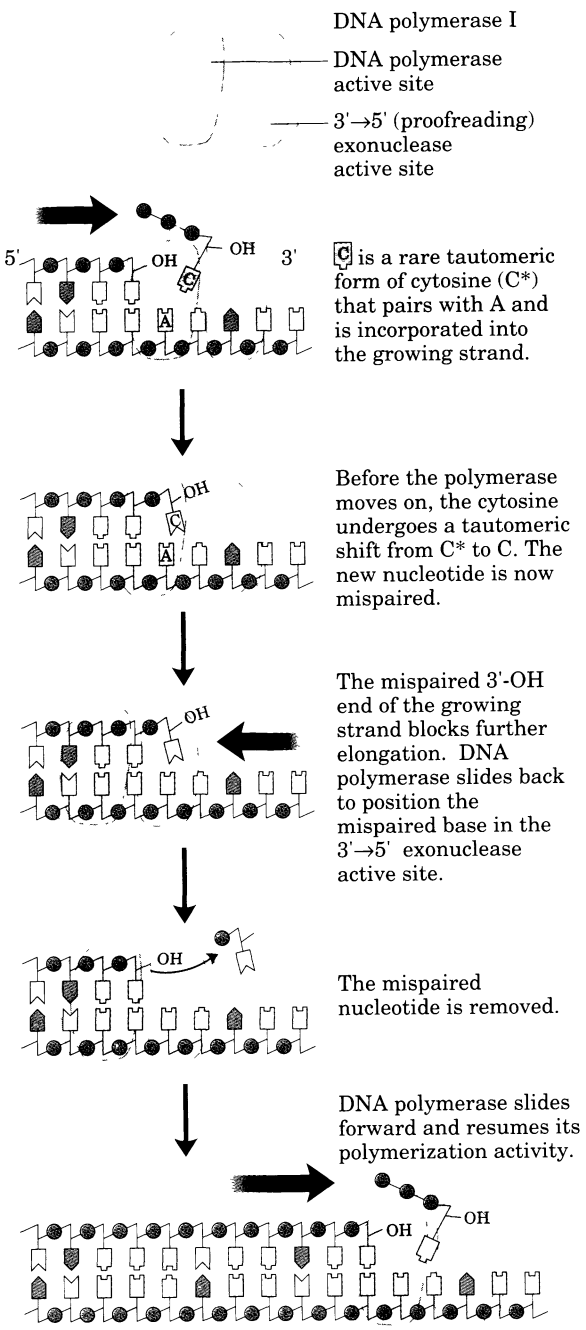


Figure 24–6 An example of error correction by the 3′→5′ exonuclease activity of DNA polymerase I. Structural analysis has located the exonuclease activity ahead of the polymerization activity as the enzyme is oriented in its movement along the DNA. A mismatched base (here, a C–A mismatch) impedes translocation of the enzyme to the next site. Sliding backward, the enzyme corrects the mistake with its 3′→5′ exonuclease activity, then resumes its polymerase activity in the 5′→3′ direction.

however, is insufficient to account for the high degree of fidelity in replication. Careful measurements in vitro have shown that DNA polymerases insert one incorrect nucleotide for every 10^4 to 10^5 correct ones. These mistakes sometimes occur because a base is briefly in an unusual tautomeric form (see Fig. 12–9), allowing it to hydrogen-bond with an incorrect partner. The error rate is reduced further in vivo by additional enzymatic mechanisms.

One mechanism intrinsic to virtually all DNA polymerases is a separate 3′→5′ exonuclease activity that serves to double-check each nucleotide after it is added. This nuclease activity permits the enzyme to remove a nucleotide just added and is highly specific for mismatched base pairs (Fig. 24–6). If the wrong nucleotide has been added, translocation of the polymerase to the position where the next nucleotide is to be added is inhibited. The 3′→5′ exonuclease activity removes the mispaired nucleotide, and the polymerase begins again. This activity, called **proofreading,** is not simply the reverse of the polymerization reaction, because pyrophosphate is not involved. The polymerizing and proofreading activities of a DNA polymerase can be measured separately. Such measurements have shown that proofreading improves the inherent accuracy of the polymerization reaction by 10^2- to 10^3-fold.

The discrimination between correct and incorrect bases during proofreading depends on the same base-pairing interactions that are used during polymerization. This strategy of enhancing fidelity by using complementary noncovalent interactions for discrimination twice in successive steps is common in the synthesis of information-containing molecules. A similar strategy is used to ensure the fidelity of protein synthesis (Chapter 26).

Overall, a DNA polymerase makes about one error for every 10^6 to 10^8 bases added. The measured accuracy of replication in *E. coli* cells, however, is still higher. The remaining degree of accuracy is accounted for by a separate enzyme system that repairs mismatched base pairs remaining after replication. This process, called mismatch repair, is described with other DNA repair processes later in this chapter.

E. coli Has at Least Three DNA Polymerases

More than 90% of the DNA polymerase activity in *E. coli* extracts can be accounted for by DNA polymerase I. Nevertheless, almost immediately after the isolation of this enzyme in 1955, evidence began to accumulate that it is not suited for replication of the large *E. coli* chromosome. First, the rate at which nucleotides are added by this enzyme (600 nucleotides/min) is too slow, by a factor of 20 or more, to account for observed rates of fork movement in the bacterial cell. Second, DNA polymerase I has a relatively low processivity; only about 50 nucleotides are added before the enzyme dissociates. Third, genetic studies have shown that many genes, and therefore many proteins, are involved in replication: DNA polymerase I clearly does not act alone. Finally, and most important, in 1969 John Cairns isolated a bacterial strain in which the gene for DNA polymerase I was altered, inactivating the enzyme. This strain was nevertheless viable!

A search for other DNA polymerases led to the discovery of *E. coli* **DNA polymerase II** and **DNA polymerase III** in the early 1970s. DNA polymerase II appears to have a highly specialized DNA repair function (described later in this chapter). DNA polymerase III is the

primary replication enzyme in *E. coli*. Properties of the three DNA polymerases are compared in Table 24–1. DNA polymerase III is a much more complex enzyme than polymerase I. It is a multimeric enzyme with at least ten different subunits (Table 24–2). Notably, the polymerization and proofreading activities of DNA polymerase III are located in separate subunits. The β subunit of this complex enzyme has been crystallized. Its structure is depicted in Fig. 24–7.

Table 24–1 Comparison of DNA polymerases of *E. coli*

	DNA polymerase		
	I	II	III
Structural gene*	*pol*A	*pol*B (*din*A)	*pol*C (*dna*E)
Subunits	1	≥4	≥10
M_r	103,000	88,000†	~900,000
3′→5′ Exonuclease (proofreading)	Yes	Yes	Yes
5′→3′ Exonuclease	Yes	No	No
Polymerization rate (nucleotides/s)	16–20	~7	250–1000
Processivity (nucleotides added before dissociation)	3–200	≥10,000	≥500,000

* For enzymes with more than one subunit, the gene listed encodes the subunit with polymerization activity Gene names in parentheses represent earlier designations of the same gene (p 815) The acronym *din* stands for *d*amage *in*ducible, *din* genes were originally identified as those that were induced as part of the SOS response to heavy DNA damage, as described later in this chapter

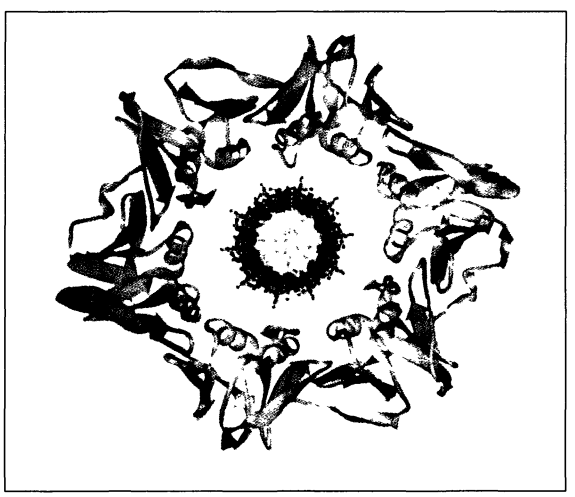
† Polymerization subunit only DNA polymerase II shares several subunits with DNA polymerase III, including the β, γ, and δ subunits (see Table 24–2) and possibly others

Table 24–2 Subunits of DNA polymerase III of *E. coli*

Subunit	M_r	Gene	Function*	
α	132,000	*pol*C (*dna*E)	Polymerization activity	Core subunits
ϵ	27,000	*dna*Q (*mut*D)	3′→5′ Proofreading exonuclease	Core subunits
θ	10,000	*hol*E		
τ	71,000	*dna*X	Stable template binding; core enzyme dimerization	
γ	52,000	*dna*X†	Enhanced processivity	
δ	35,000	*hol*A	Enhanced processivity	
δ'	33,000	*hol*B		
χ	15,000	*hol*C		
ψ	12,000	*hol*D		
β	37,000	*dna*N	Required for optimal processivity	

* Where no function is listed, the molecular role of the subunit is not entirely clear.

† The γ subunit is encoded by a portion of the gene for the τ subunit, such that the amino-terminal 80% of the τ subunit has the same amino acid sequence as the γ subunit The γ subunit is generated by a translational frameshifting mechanism (see Box 26–1) that leads to premature translational termination



(a)

(b)

Figure 24–7 The two β subunits of *E. coli* polymerase III (shown in gray and light blue) form a circular clamp that surrounds DNA, shown in **(a)** from above as a ribbon structure and in **(b)** from the side as a space-filling model. The clamp slides along the DNA, enhancing the processivity of the polymerase by preventing its dissociation. The γ and δ subunits of DNA polymerase III facilitate the binding of the β subunits to DNA.

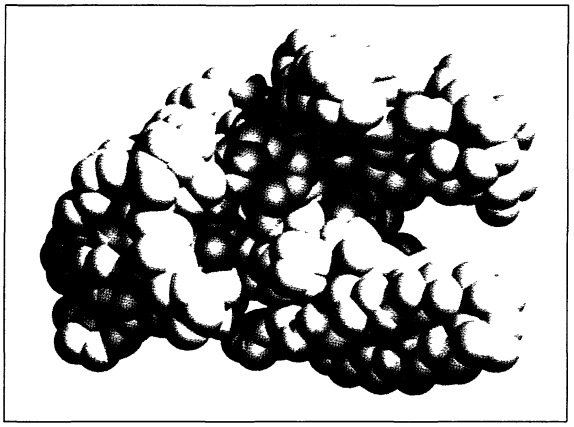


Figure 24–8 The Klenow fragment of *E. coli* DNA polymerase I, produced by proteolytic treatment of the polymerase, includes the polymerization activity of the enzyme. The horizontal groove evident on this face of the protein is the likely binding site for DNA.

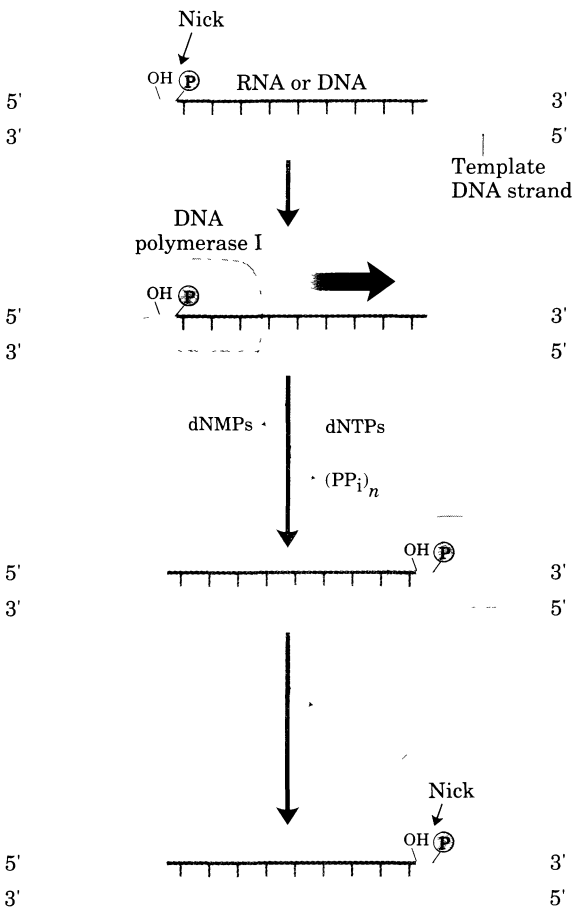


Figure 24–9 The 5′→3′ exonuclease of DNA polymerase I can remove or degrade an RNA or DNA strand paired to the template, as the polymerase activity simultaneously replaces the degraded strand. These activities are important for the role of DNA polymerase I in DNA repair and in removal of RNA primers during replication, as described later in this chapter. The strand of nucleic acid (DNA or RNA) to be removed is shown in green; the replacement strand is shown in red. A nick (a phosphodiester bond broken to leave a free 3′ OH and 5′ phosphate) is found where DNA synthesis starts. After synthesis, a nick remains where DNA polymerase I dissociates. This action of polymerase I has effectively extended the nontemplate DNA strand and moved the nick down the DNA, a process that is sometimes called nick translation.

DNA polymerase I is far from irrelevant, however. This enzyme serves a host of "clean-up" functions during replication, recombination, and repair, as discussed later in the chapter. These special functions are enhanced by an additional enzymatic activity of DNA polymerase I, a 5′→3′ exonuclease activity. This activity is distinct from the 3′→5′ proofreading exonuclease and is located in a distinct structural domain that can be separated from the enzyme by mild protease treatment. When the 5′→3′ exonuclease domain is removed, the remaining fragment (M_r 68,000) retains the polymerization and proofreading activities, and is called the **large** or **Klenow fragment.** The structure of the Klenow fragment has been determined, and it is this fragment of DNA polymerase I that is depicted in Figure 24–8. The 5′→3′ exonuclease activity of intact DNA polymerase I permits it to extend DNA strands even if the template is already paired to an existing strand of nucleic acid (Fig. 24–9). Using this activity, DNA polymerase I can degrade or displace a segment of DNA (or RNA) paired to the template and replace it with newly synthesized DNA. Most other DNA polymerases, including DNA polymerase III, lack a 5′→3′ exonuclease activity.

DNA Replication Requires Many Enzymes and Protein Factors

We now know that replication in *E. coli* requires not just a single DNA polymerase but 20 or more different enzymes and proteins, each performing a specific task. Although not yet obtained as a physical entity, the entire complex has been called the **DNA replicase system** or the **replisome.** The enzymatic complexity of replication reflects the requirements imposed on the process by the structure of DNA. We will introduce some of the major classes of replication enzymes by considering the problems that they overcome.

To gain access to the DNA strands that are to act as templates the two parent strands must be separated. This is generally accomplished by enzymes called **helicases,** which move along the DNA and separate the strands using chemical energy from ATP. Strand separation creates topological stress in the helical DNA structure, which is relieved by the action of **topoisomerases** (Chapter 23). The separated strands are stabilized by **DNA-binding proteins.** Primers must be present or synthesized before DNA polymerases can synthesize DNA. The primers are generally short segments of RNA laid down by enzymes called **primases.** Ultimately, the RNA primers must be removed and replaced by DNA. In *E. coli,* this is one of the many functions of DNA polymerase I. After removal of the RNA segments and filling in of the gap with DNA, there remain points in the DNA backbone where a phosphodiester bond is broken. These breaks, called nicks, must be sealed by enzymes called **DNA ligases.** All of these processes must be coordinated and regulated. The interplay of these and other enzymes has been best characterized in the *E. coli* system.

Replication of the *E. coli* Chromosome Proceeds in Stages

The synthesis of a DNA molecule can be divided into three stages: initiation, elongation, and termination. These are distinguished by differences in the reactions taking place and in the enzymes required. In the next two chapters we will see that the synthesis of the other major biological polymers, RNAs and proteins, can be similarly broken down into the same three stages, each with unique characteristics. The events described below reflect information derived from in vitro experiments using purified *E. coli* proteins.

Initiation The *E. coli* replication origin, called *ori*C, consists of 245 base pairs, many of which are highly conserved among bacteria. The general arrangement of the conserved sequences is illustrated in Figure 24–10. The key sequences for this discussion are two series of short repeats; three repeats of a 13 base pair sequence and four repeats of a 9 base pair sequence.

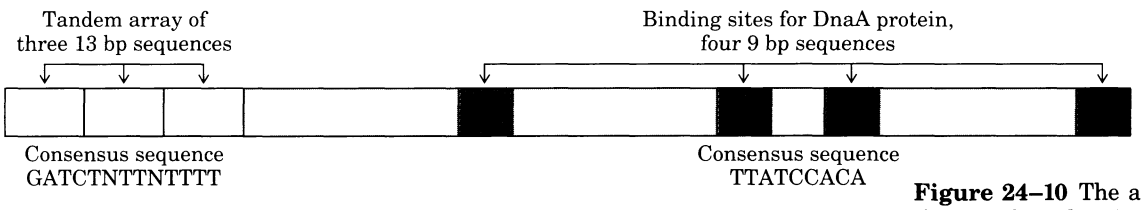


Figure 24–10 The arrangement of sequences in the *E. coli* replication origin, called *ori*C. The repeated sequences are shaded in color. The term "consensus sequence" is used to describe a repeated sequence that varies somewhat from one copy to the next; it depicts the most common nucleotide residues found at each position in the sequence (N represents any of the four nucleotides). Individual copies of the repeated sequence may differ from the consensus at one or several positions. The arrows indicate the orientations of the nucleotide sequences.

At least eight different enzymes or proteins (summarized in Table 24–3) participate in the initiation phase of replication. They open the DNA helix at the origin and establish a prepriming complex that sets the stage for subsequent reactions. The key component in the initiation

Table 24–3 Proteins required to initiate replication at the *E. coli* origin

Protein	M_r	Number of subunits	Function
DnaA protein	50,000	1	Opens duplex at specific sites in origin
DnaB protein (helicase)	300,000	6*	Unwinds DNA
DnaC protein	29,000	1	Required for DnaB binding at origin
HU	19,000	2	Histonelike protein; stimulates initiation
Primase (DnaG protein)	60,000	1	Synthesizes RNA primers
SSB	75,600	4*	Binds single-stranded DNA
RNA polymerase	454,000	6	Facilitates DnaA activity
DNA topoisomerase II (gyrase)	400,000	4	Relieves torsional strain generated by DNA unwinding

* Subunits in these cases are identical

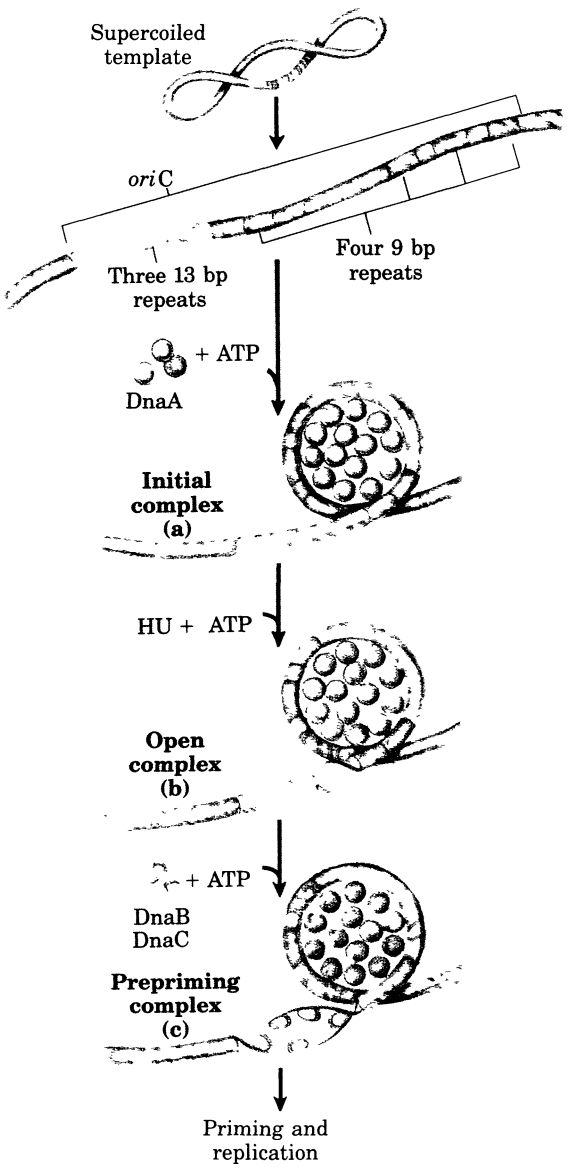


Figure 24–11 A model for initiation of replication at the *E. coli* origin, *ori*C. **(a)** About 20 DnaA protein molecules, each with a bound ATP, bind at the four 9 base pair repeats. The DNA is wrapped around this complex. **(b)** The three 13 base pair repeats are then denatured sequentially to give the open complex. **(c)** The DnaB protein binds to the open complex, with the aid of DnaC protein, and the DnaB helicase activity further unwinds the DNA in preparation for priming and DNA synthesis.

process is the DnaA protein (Fig. 24–11). A complex of about 20 DnaA protein molecules binds to the four 9 base pair repeats in the origin. In a reaction that requires ATP and is facilitated by the bacterial histone-like protein HU, the DnaA protein recognizes and successively denatures the DNA in the region of the three 13 base pair repeats, which are rich in A=T pairs. The DnaB protein then binds to this region in a reaction that requires the DnaC protein. The DnaB protein is a helicase that unwinds the DNA bidirectionally, creating two potential replication forks. If the *E. coli* single-strand DNA-binding protein (SSB) and DNA gyrase (DNA topoisomerase II) are added to this reaction in vitro, thousands of base pairs are rapidly unwound by the DnaB helicase, proceeding out from the origin. Multiple molecules of SSB bind cooperatively to single-stranded DNA, stabilizing the separated DNA strands and preventing renaturation. Gyrase relieves the topological stress created by the DnaB helicase reaction. When additional replication proteins are added as described below, the DNA unwinding mediated by DnaB protein is coupled to replication.

DNA replication must be precisely regulated so that it occurs once and only once in each cell cycle. Initiation is the only phase of replication that is regulated, but the mechanism is not yet well understood. Biochemical studies have provided a few insights. The DnaA protein hydrolyzes its tightly bound ATP slowly (about 1 hour) to form an inactive DnaA–ADP complex. Reactivating this complex (replacing ADP with ATP) is facilitated by an interaction between DnaA protein and acidic phospholipids in the bacterial plasma membrane. Initiation at inappropriate times is prevented by the presence of the inactive DnaA-ADP complex, by the binding of a protein called IciA (*i*nhibitor of *c*hromosomal *i*nitiation) to the 13 base pair repeats, and perhaps by other factors. Deciphering the complex interactions in this regulatory network remains an active area of research.

Elongation The elongation phase of replication consists of two seemingly similar operations that are mechanistically quite distinct: leading strand synthesis and lagging strand synthesis. Several enzymes at the replication fork are important to the synthesis of both strands. DNA helicases unwind the parental DNA. DNA topoisomerases relieve the topological stress induced by the helicases, and SSB stabilizes the separated strands. In other respects, synthesis of DNA in the two strands is sharply different. We will begin with leading strand synthesis, the more straightforward of the two.

Leading strand synthesis begins with the synthesis by primase of a short (10 to 60 nucleotide) RNA primer at the replication origin. Deoxyribonucleotides are then added to this primer by DNA polymerase III. Once begun, leading strand synthesis proceeds continuously, keeping pace with the replication fork (Fig. 24–12).

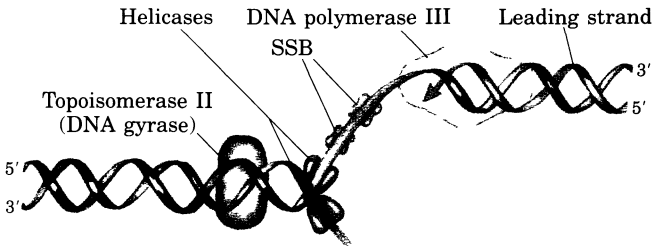


Figure 24–12 Synthesis of the leading strand. DNA polymerase III keeps pace with the replication fork. Helicases separate the two DNA strands at the fork, molecules of SSB bind to and stabilize the separated strands, and DNA topoisomerase II acts to relieve torsional stress generated by the helicases.

Lagging strand synthesis, which must be accomplished in short fragments (Okazaki fragments) synthesized in the direction opposite to fork movement, is a more intricate problem. It is solved by a protein machine that incorporates several specialized proteins in addition to polymerase III. Each fragment must have its own RNA primer, synthesized by primase, and positioning of the primers must be controlled and coordinated with fork movement. The regulatory apparatus for lagging strand synthesis is a traveling protein machine called a **primosome,** which consists of seven different proteins including the DnaB protein, DnaC protein, and primase mentioned above (Table 24–4). The primosome moves along the lagging strand template in the 5′→3′ direction, keeping pace with the replication fork. As it moves, the primosome at intervals compels primase to synthesize a short (10 to 60) residue RNA primer to which DNA is then added by DNA poly-

Table 24–4 ***E. coli*** **proteins at the replication fork**

Protein	M_r	Number of subunits	Function
SSB	75,600	4	Binding to single-stranded DNA
Protein i (DnaT protein)	66,000	3	Primosome constituent
Protein n	28,000	2	Primosome assembly and function
Protein n′	76,000	1	Primosome constituent
Protein n″	17,000	1	Primosome constituent
DnaC protein	29,000	1	Primosome constituent
DnaB protein (helicase)	300,000	6	DNA unwinding; primosome constituent
Primase (DnaG protein)	60,000	1	RNA primer synthesis; primosome constituent
DNA polymerase III	900,000	2 × 10	Processive chain elongation
DNA polymerase I	103,000	1	Filling of gaps, excision of primers
DNA ligase	74,000	1	Ligation
DNA topoisomerase II (gyrase)	400,000	4	Supercoiling
Rep (helicase)	65,000	1	Unwinding
DNA helicase II	75,000	1	Unwinding
DNA topoisomerase I	100,000	4	Relaxing negative supercoils

* Modified from Kornberg, A (1982) *Supplement to DNA Replication,* Table S11–2, W H Freeman and Company, New York.

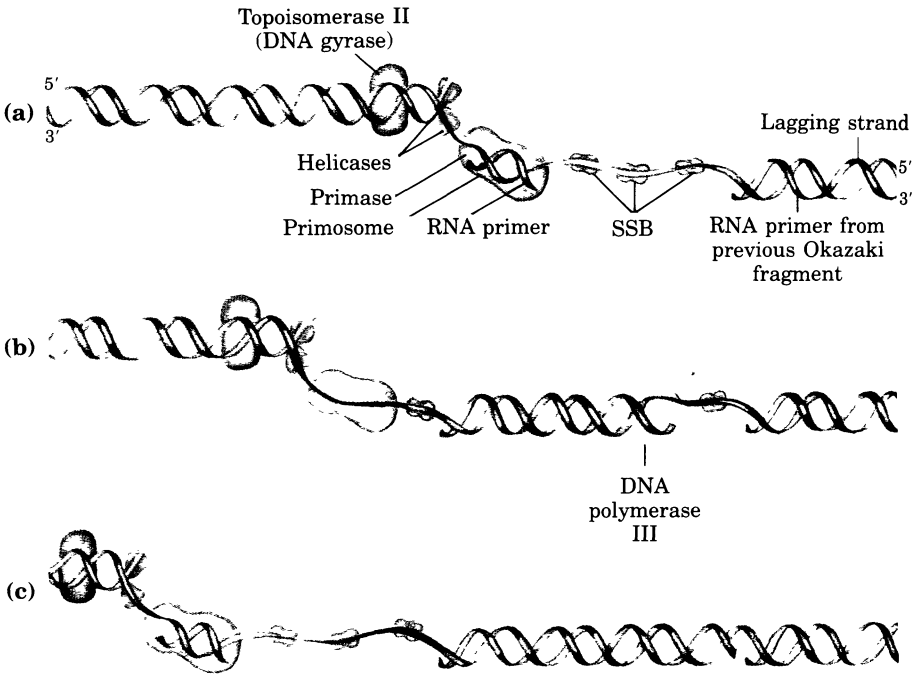


Figure 24–13 Synthesis of Okazaki fragments. The multiprotein primosome complex travels in the same direction as the replication fork. **(a)** At intervals, primase synthesizes an RNA primer for a new Okazaki fragment. Note that this synthesis formally proceeds in the direction opposite to fork movement. **(b)** Each primer is extended by DNA polymerase III. **(c)** DNA synthesis continues until the primer of the previously added Okazaki fragment is encountered. (Helicases, DNA topoisomerase II, and SSB have the functions outlined in Fig. 24–12.)

merase III (Fig. 24–13). Note that the direction of the synthetic reactions of primase and polymerase III is opposite to the direction of primosome movement. When the new Okazaki fragment is complete, the RNA primer is removed by DNA polymerase I (using its 5′→3′ exonuclease activity) and is replaced with DNA by the same enzyme. The remaining nick is sealed by DNA ligase (Fig. 24–14). The proteins acting at the replication fork are summarized in Table 24–4.

DNA ligase catalyzes the formation of a phosphodiester bond between a 3′ hydroxyl at the end of one DNA strand and a 5′ phosphate at the end of another strand. In *E. coli* the phosphate must be activated using NAD^+ (ATP is used in some organisms) to supply the required chemical energy. The reaction pathway, as established by I. Robert Lehman and colleagues, is shown in Figure 24–15. The use by the ligase of *E. coli* of the nucleotide NAD^+—a cofactor that normally functions in hydride transfer reactions (see Fig. 13–16)—as the source of the AMP activating group is unusual. DNA ligase is another enzyme of DNA metabolism that has become an important reagent in recombinant DNA experiments (Chapter 28).

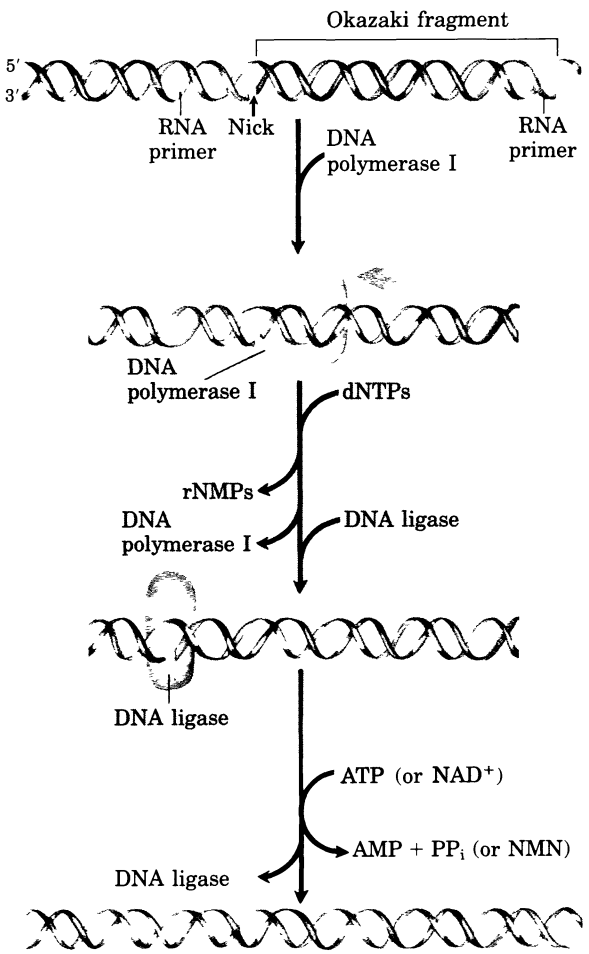


Figure 24–14 Removal of RNA primers in the lagging strand. The RNA primer is removed by the 5′→3′ exonuclease activity of DNA polymerase I, and it is replaced with DNA by the same enzyme. The remaining nick is sealed by DNA ligase. The role of NAD^+ is shown in Fig. 24–15.

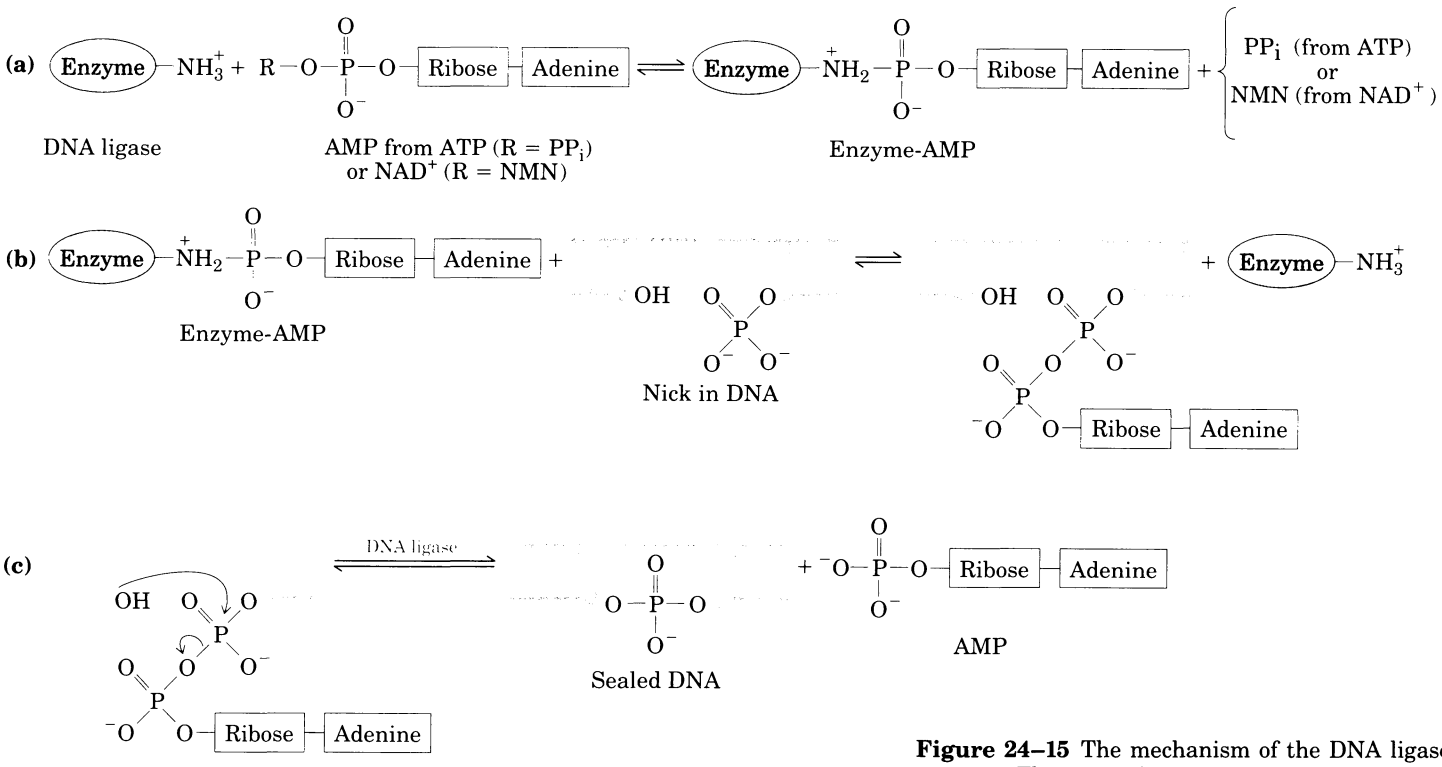


Figure 24–15 The mechanism of the DNA ligase reaction. There are three steps, and in each step one phosphodiester bond is formed at the expense of another. Steps **(a)** and **(b)** lead to activation of the 5′ phosphate in the nick. An AMP group is transferred first to a Lys residue on the enzyme and then to the 5′ phosphate in the nick. **(c)** The 3′-OH group then attacks this phosphate and displaces AMP, leading to the formation of a phosphodiester bond to seal the nick. The AMP is derived from NAD^+ in the case of *E. coli* DNA ligase. The DNA ligases isolated from a number of other prokaryotic and eukaryotic sources use ATP rather than NAD^+, and release pyrophosphate rather than nicotinamide mononucleotide (NMN) in step **(a).**

In *E. coli,* synthesis of the leading and lagging strands may actually be coupled as shown in Figure 24–16. This can be accomplished by looping the lagging strand template so that synthesis can be carried out concurrently on both strands by a single dimeric polymerase III acting in concert with the primosome and all of the other proteins at the replication fork (Table 24–4).

Termination Eventually, the two replication forks meet at the other side of the circular *E. coli* chromosome. Very little is known about this stage of the reaction, though the action of a type 2 topoisomerase called DNA topoisomerase IV appears to be necessary for final separation of the two completed circular DNA molecules. Nor is much understood about the process of partitioning the two DNA molecules into daughter cells at division.

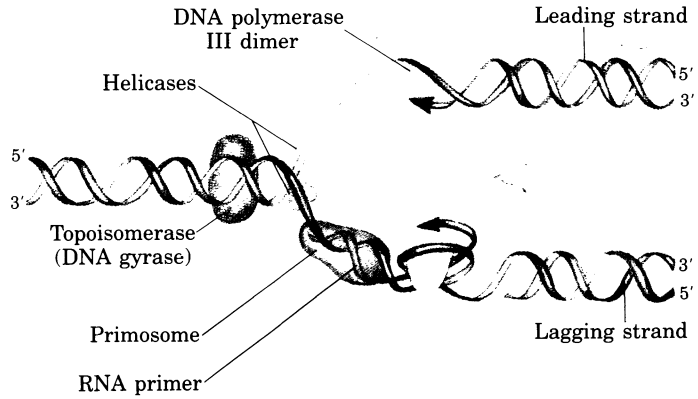


Figure 24–16 Coupling the synthesis of leading and lagging strands with a dimeric DNA polymerase III. The template for the lagging strand is looped tightly so that the direction of synthesis has the same orientation for both strands. As polymerization proceeds, the loop grows until the previous Okazaki fragment is encountered. Here, the polymerase synthesizing the lagging strand must dissociate and reinitiate at a new primer and with a new tight loop. This must be coordinated to keep pace with that part of the polymerase synthesizing the leading strand.

Replication in Eukaryotic Cells Is More Complex

The DNA molecules in eukaryotic cells are considerably larger than those in bacteria and are organized into complex nucleoprotein structures (chromatin) (Chapter 23). The essential features of DNA replication are the same in eukaryotes and prokaryotes. However, some interesting variations on the general principles discussed above promise new insights into the regulation of replication and its link with the cell cycle.

Origins of replication, called *a*utonomously *r*eplicating *s*equences (ARS), have been identified and studied in yeast. ARS elements span regions of about 300 base pairs and contain several conserved sequences that are essential for ARS function. There are about 400 ARS elements in yeast, with most chromosomes having several. Proteins that specifically bind the ARS region have been identified in yeast, although their functions are not yet understood.

The rate of replication fork movement in eukaryotes (~50 nucleotides/s) is only one-tenth that observed in *E. coli*. At this rate, replication of an average human chromosome proceeding from a single origin would take more than 500 hours. Instead, replication of human chromosomes proceeds bidirectionally from multiple origins spaced 30,000 to 300,000 base pairs apart. With the exception of the ARS elements of yeast, the structure of the origins of replication in eukaryotes is not known. Because eukaryotic chromosomes are almost uniformly much larger than bacterial chromosomes, the presence of multiple origins on a eukaryotic chromosome is probably a general rule.

As in bacteria, there are several types of DNA polymerases in eukaryotic cells. Some have been linked to special functions such as the replication of the DNA in mitochondria. The replication of nuclear chromosomes involves an enzyme called **DNA polymerase α,** in association with another polymerase called **DNA polymerase δ.** DNA polymerase α is typically a four-subunit enzyme with similar structure and properties in all eukaryotic cells. One of the subunits has a primase activity. The largest subunit (M_r ~ 180,000) contains the polymerization activity. DNA polymerase δ has two subunits. This enzyme exhibits a very interesting association with and stimulation by a protein called *p*roliferating *c*ell *n*uclear *a*ntigen (PCNA; M_r 29,000) found in large amounts in the nuclei of proliferating cells. The PCNA from yeast will function with DNA polymerase δ from calf thymus, and the calf thymus PCNA with yeast DNA polymerase δ, suggesting a conservation of the structure and function of these key components of the cell division apparatus in all eukaryotic cells. PCNA appears to have a function analogous to the β subunit of *E. coli* DNA polymerase III (see Fig. 24–7), forming a circular clamp that greatly enhances the processivity of DNA polymerase δ.

DNA polymerase δ, which has a 3′→5′ proofreading exonuclease activity, appears to carry out leading strand synthesis. DNA polymerase α has a relatively low processivity, and with its associated primase it may carry out lagging strand synthesis as part of a eukaryotic replisome. Another polymerase called **DNA polymerase ϵ,** may replace DNA polymerase δ in some situations, such as in DNA repair.

Two other protein complexes, called RFA and RFC (RF stands for replication factor), have been implicated in eukaryotic DNA replication. Both have been found in organisms ranging from yeast to mammals. RFA is a eukaryotic single-stranded DNA-binding protein, with a function equivalent to the *E. coli* SSB protein. RFC appears to facilitate the assembly of active replication complexes.

Summary

The integrity of the structure and nucleotide sequence of DNA is of utmost importance to the cell. This is reflected in the complexity and redundancy of the enzyme systems that participate in DNA replication, repair, and recombination.

Replication of DNA occurs with very high fidelity and within a designated time period in the cell cycle. Replication is semiconservative, with each strand acting as a template for a new daughter strand. The reaction starts at a sequence in the DNA called the origin, and usually proceeds bidirectionally from that point. DNA is synthesized in the 5′→3′ direction by DNA polymerases. At the replication fork, the leading strand is synthesized continuously and in the same direction as replication fork movement. The lagging strand is synthesized discontinuously. The fidelity of DNA replication is maintained by (1) base selection by the polymerase, (2) a 3′→5′ proofreading exonuclease activity that is part of most DNA polymerases, and (3) a specific repair system that repairs any mismatches left behind after replication.

Most cells have several DNA polymerases. In *E. coli,* DNA polymerase III is the primary replication enzyme. DNA polymerase I is responsible for special functions during replication, recombination, and repair. DNA polymerase II has a specialized replication activity that allows it to replicate past DNA lesions in error-prone DNA repair. Replication of the *E. coli* chromosome involves many enzymes and protein factors organized into complexes. Initiation of replication requires binding of DnaA protein to the origin, strand separation, and the entry of the DnaB and DnaC proteins to set up two replication forks. The action of DnaA is associated with the *E. coli* membrane and is regulated by the action of acidic phospholipids. Initiation is the only phase of replication that is regulated. The process of elongation has different requirements for each strand. DNA strands are separated by helicases, and the resulting topological strain is relieved by topoisomerases. Single-strand DNA binding proteins stabilize the separated strands. In synthesis of the lagging strand, the primosome protein complex moves with the fork and regulates the synthesis of RNA primers by primase. Synthesis of the leading and lagging strands by DNA polymerase III may be coupled. RNA primers are removed and replaced with DNA by DNA polymerase I, and nicks are sealed by DNA ligase.

A similar pattern of replication occurs in eukaryotic cells, but eukaryotic chromosomes have multiple replication origins. Several eukaryotic DNA polymerases have been identified.

Every cell also has multiple and sometimes redundant systems for DNA repair. Mismatch repair in *E. coli* is directed by transient undermethylation of (5′)GATC sequences on the newly synthesized strand after replication. Other systems recognize and repair damage caused by environmental agents such as radiation and alkylating agents, and damage caused by spontaneous reactions of nucleotides. Some repair systems recognize and excise only damaged or incorrect bases (e.g., uracil), leaving an AP (apurinic or apyrimidinic) site in the DNA. This is repaired by excising and replacing the segment of DNA containing the AP site. Other excision repair systems recognize and remove pyrimidine dimers and other modified nucleotides. Some types of DNA damage can also be repaired by direct reversal of the reaction causing the damage: pyrimidine dimers are directly converted to monomeric pyrimidines by photolyase, and the methyl group in O^6-methylguanine is removed by a specific methyltransferase. Error-prone repair is a specialized and mutagenic replication process observed when DNA damage is so heavy that the need for some replication outweighs the need to avoid errors.

DNA sequences are rearranged in recombination reactions. Homologous genetic recombination occurs between any two DNAs that share sequence homology. This reaction takes place in meiosis (in eukaryotes) and is one of the processes that creates genetic diversity. Homologous recombination also is needed for repair of some types of DNA damage. A Holliday intermediate in which a crossover has occurred between the strands of two homologous DNAs is formed during the process. In *E. coli,* the RecA protein promotes formation of Holliday intermediates and branch migration to extend heteroduplex DNA.

Site-specific recombination occurs only at specific target sequences and can also involve a Holliday intermediate. The recombinases cleave the DNA at specific points and ligate the strands to new partners. This type of recombination is found in virtually all cells, and its many functions include DNA integration and regulation of gene expression. In vertebrates, a programmed recombination reaction related to site-specific recombination is used to join immunoglobulin gene segments to form immunoglobulin genes during B-lymphocyte differentiation. Some small segments of DNA, called transposons, are capable of moving from one point in a chromosome to another point in the same or another chromosome. These elements are found in virtually all cells.

Further Reading

General

Kornberg, A. & Baker, T.A. (1991) *DNA Replication,* 2nd edn, W.H. Freeman and Company, New York.

An excellent primary source.

Kucherlapati, R. & Smith, G.R. (eds) (1988) *Genetic Recombination,* American Society of Microbiology, Washington, DC.

Excellent reviews on a wide assortment of recombination topics.

Richardson, C.C. & Lehman, I.R. (eds) (1990) *Molecular Mechanisms in DNA Replication and Recombination,* Alan R. Liss, Inc., New York.

A collection of a papers from a major symposium on the topic.

Replication

Bramhill, D. & Kornberg, A. (1988) A model for initiation at origins of DNA replication. *Cell* **54,** 915–918.

Burgers, P.M.J. (1989) Eukaryotic DNA polymerases α and δ: conserved properties and interactions, from yeast to mammalian cells. *Prog. Nucleic Acid Res. Mol. Biol.* **37,** 235–280.

Campbell, J. (1988) Eukaryotic DNA replication: yeast bares its ARSs. *Trends Biochem. Sci.* **13,** 212–217.

Echols, H. & Goodman, M.F. (1991) Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem.* **60,** 477–511.

Marians, K.J. (1992) Prokaryotic DNA replication. *Annu. Rev. Biochem.* **61,** 673–719.

McHenry, C.S. (1991) DNA polymerase III holoenzyme. *J. Biol. Chem.* **266,** 19127–19130.

Radman, M. & Wagner, R. (1988) The high fidelity of DNA duplication. *Sci. Am.* **259** (August), 40–46.

Wang, T.S.-F. (1991) Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* **60,** 513–552.

Repair

Friedberg, E.C. (1985) *DNA Repair,* W.H. Freeman and Company, New York.

Modrich, P. (1991) Mechanisms and biological effects of mismatch repair. *Annu. Rev. Genet.* **25,**

Sancar, A. & Sancar, G.B. (1988) DNA repair enzymes. *Annu. Rev. Biochem.* **57,** 29–67.

Recombination

Berg, D.E. & Howe, M.M. (eds) (1989) *Mobile DNA,* American Society for Microbiology, Washington, DC.

Reviews covering many topics related to transposition.

Cox, M.M. & Lehman, I.R. (1987) Enzymes of genetic recombination. *Annu. Rev. Biochem.* **56,** 229–262.

Craig, N.L. (1988) The mechanism of conservative site-specific recombination. *Annu. Rev. Genet.* **22,** 77–105.

Landy, A. (1989) Dynamic, structural, and regulatory aspects of λ site-specific recombination. *Annu. Rev. Biochem.* **58,** 913–949.

A thorough description of the protein–DNA complexes involved in this reaction.

Mizuuchi, K. (1992) Transpositional recombination: insights from Mu and other elements. *Annu. Rev. Biochem.* **61,** 1011–1051.

Radding, C.M. (1991) Helical interactions in homologous pairing and strand exchange driven by RecA protein. *J. Biol. Chem.* **266,** 5355–5358.

A good, short summary.

Roca, A.I. & Cox, M.M. (1990) The RecA protein: structure and function. *Crit. Rev. Biochem. Mol. Biol.* **25,** 415–456.

Taylor, A.F. (1992) Movement and resolution of Holliday junctions by enzymes from *E. coli. Cell* **69,** 1063–1065.

Problems

1. *Conclusions from the Meselson–Stahl Experiment* The Meselson–Stahl experiment proved that DNA undergoes semiconservative replication in *E. coli*. In the "dispersive" model of DNA replication, the parent DNA strands are cleaved into pieces of random size and are then joined with pieces of the newly replicated DNA to yield daughter duplexes in which, in the Meselson–Stahl experiment, both strands would contain random segments of both heavy and light DNA. Explain how the results of the Meselson–Stahl experiment ruled out such a model.

2. *Number of Turns in the* E. coli *Chromosome* How many turns must be unwound during replication of the *E. coli* chromosome? The chromosome contains about 4.7×10^6 base pairs.

3. *Replication Time in* E. coli From the data in this chapter, how long would it take to replicate the *E. coli* chromosome at 37 °C, if two replication forks start from the origin? Under some conditions *E. coli* cells can divide every 20 min. Can you suggest how this is possible?

4. *Base Composition of DNAs Made from Single-Stranded Templates* Determine the base composition you might expect in the total DNA synthesized by DNA polymerase on templates provided by an equimolar mixture of the two complementary strands of circular bacteriophage ϕX174 DNA. The base composition of one strand is A, 24.7%; G, 24.1%; C, 18.5%; and T, 32.7%. What assumption is necessary to answer this problem?

5. *Okazaki Fragments* In the replication of the *E. coli* chromosome, about how many Okazaki fragments would be formed? What factors guarantee that the numerous Okazaki fragments are assembled in the correct order in the new DNA?

6. *Leading and Lagging Strands* List and compare the precursors and enzymes needed to make the leading versus lagging strands during DNA replication in *E. coli*.

7. *Fidelity of Replication of DNA* What factors participate in ensuring the fidelity of replication during the synthesis of the leading strand of a new DNA? Would you expect the lagging strand to be made with the same fidelity as the leading strand? Give reasons for your answers.

8. *DNA Repair Mechanisms* Vertebrate and plant cells often methylate cytosine in DNA to form 5-methylcytosine (see Fig. 12–5a). In these same cells, there is a specialized repair system that recognizes G–T mismatches and repairs them to G≡C base pairs. Rationalize this repair system in terms of the presence of 5-methylcytosine in the DNA.

9. *Holliday Intermediates* How are the Holliday intermediates formed in homologous genetic recombination and in site-specific recombination different?

10. *DNA Recombination* A circular DNA molecule is converted to two smaller circles by an enzyme or enzymes in a crude cellular extract. What types of recombination could account for this reaction, and what else must you know to determine which type it is?

CHAPTER

RNA Metabolism

The expression of the genetic information contained in a segment of DNA always involves the generation of a molecule of RNA. At first glance, a strand of RNA may seem quite similar to a strand of DNA, differing only in the hydroxyl group at the 2′ position and the substitution of uracil for thymine. As we will see, however, these small differences confer on RNA the potential for much greater structural diversity than DNA, a diversity that allows RNA to assume a variety of cellular functions. RNA molecules not only carry and express genetic information, they can also act as catalysts.

RNA is the only macromolecule known to have both informational and catalytic functions, leading to much speculation that it may have been the essential chemical intermediate in the development of life on this planet. The discovery of catalytic RNAs has changed the very definition of the word "enzyme." Many RNAs are also complexed with proteins, forming complicated biochemical machines with a wide variety of functions.

With the exception of the RNA genomes of certain viruses, all RNA molecules are derived from information permanently stored in DNA. In a process called **transcription,** an enzyme system converts the genetic information of a segment of DNA into an RNA strand with a base sequence complementary to one of the DNA strands. Three major kinds of RNA are produced. **Messenger RNA (mRNA)** carries the sequences that encode the amino acid sequence of one or more polypeptides specified by a gene or set of genes in the chromosomes. **Transfer RNA (tRNA)** is an adapter that reads the information encoded in the mRNA and transfers the appropriate amino acid to the growing polypeptide chain during protein synthesis. **Ribosomal RNA (rRNA)** molecules associate with proteins to form the intricate protein synthetic machine, the ribosome. In addition, there are many specialized RNAs with regulatory or catalytic functions.

Replication and transcription differ in one important respect. During replication the entire chromosome is copied to yield daughter DNAs identical to the parent DNA, whereas transcription is selective: only particular genes or groups of genes are transcribed at any one time. The transcription of DNA can therefore be regulated so that only genetic information needed by the cell at a particular moment is transcribed. Specific regulatory sequences indicate the beginning and end of the segments of DNA to be transcribed, as well as which DNA strand is to be used as template. Regulation also involves a variety of proteins that will be described in more detail in Chapter 27.

In this chapter we begin by describing the synthesis of RNA on a DNA template, a process similar in many respects to DNA synthesis. We then turn to postsynthetic processing and turnover of RNA molecules. Many of the specialized functions of RNA will be encountered in this discussion of the posttranscriptional reactions. Indeed, the substrates for RNA enzymes are generally other RNA molecules. We conclude the chapter with an examination of systems in which RNA rather than DNA serves as a template for the transfer of genetic information. Here, the information pathways are expanded and come full circle, and template-directed nucleic acid synthesis is revealed as a process with standard rules that apply regardless of whether the template or product is RNA or DNA. This biological interconversion of DNA and RNA as information carriers leads finally to a discussion of the origin of biological information.

DNA-Dependent Synthesis of RNA

We can most usefully begin our discussion of RNA synthesis by comparing it with DNA replication as described in Chapter 24. Transcription is very similar to replication in terms of chemical mechanism, polarity (direction of synthesis), and use of a template. The two processes differ, however, in that transcription does not require a primer, it generally involves only short segments of a DNA molecule, and within those segments only one of the two DNA strands serves as a template. We begin our discussion by introducing the enzymes responsible for transcription.

RNA Is Synthesized by RNA Polymerases

The discovery of DNA polymerase and its dependence on a DNA template encouraged a search for an enzyme that synthesizes an RNA strand complementary to a DNA template. Such an enzyme, capable of forming an RNA polymer from ribonucleoside 5′-triphosphates, was isolated from bacterial extracts in 1959 by four independent research groups. This enzyme, **DNA-directed RNA polymerase,** requires, in addition to a DNA template, all four ribonucleoside 5′-triphosphates (ATP, GTP, UTP, and CTP) as precursors of the nucleotide units of RNA, as well as Mg^{2+}. The purified enzyme also contains Zn^{2+}. The fundamental chemistry of RNA synthesis has much in common with DNA synthesis. RNA polymerase elongates an RNA strand by adding ribonucleotide units to the 3′-hydroxyl end of the RNA chain and thus builds RNA chains in the 5′→3′ direction. The 3′-hydroxyl group acts as nucleophile, attacking at the α-phosphate of the incoming ribonucleoside triphosphate (as illustrated for DNA synthesis in Fig. 24–5) and releasing pyrophosphate. The overall reaction is

$$\underset{\text{RNA}}{(\text{NMP})_n} + \text{NTP} \longrightarrow \underset{\substack{\text{Lengthened}\\ \text{RNA}}}{(\text{NMP})_{n+1}} + \text{PP}_\text{i} \qquad (25\text{–}1)$$

RNA polymerase requires DNA for activity and is most active with a double-stranded DNA as template. Only one of the two DNA strands is used as a template, copied in the 3′→5′ direction (antiparallel to the new RNA strand) just as in DNA replication. Each nucleotide in the newly formed RNA is selected by Watson–Crick base-pairing interac-

tions; uridylate (U) residues are inserted in the RNA opposite to adenylate residues in the DNA template, adenylate residues are inserted opposite to thymidylate residues. Guanylate and cytidylate residues in DNA specify cytidylate and guanylate, respectively, in the new RNA strand.

Unlike DNA polymerase, RNA polymerase does not require a primer to initiate synthesis. Initiation of RNA synthesis, however, occurs only at specific sequences called promoters (described below). RNA synthesis usually starts with a GTP or ATP residue, whose 5′-triphosphate group is not cleaved to release PP_i but remains intact throughout transcription. During transcription the new RNA strand base-pairs temporarily with the DNA template to form a short length of hybrid RNA–DNA double helix, which is essential to the correct readout of the DNA strand (Fig. 25–1). The RNA in this hybrid duplex "peels off" shortly after its formation.

To enable RNA polymerase to synthesize an RNA strand complementary to one of the DNA strands, the DNA duplex must unwind over a short distance, forming a transcription "bubble." During transcription, the *E. coli* RNA polymerase generally keeps about 17 base pairs unwound, unwinding the DNA ahead and rewinding it behind. Because the DNA is a helix, this process requires considerable rotation of the nucleic acid molecules (Fig. 25–1a). Rotation is restricted in most DNAs by DNA-binding proteins and other structural barriers, and a moving RNA polymerase generates waves of positive supercoils ahead of and negative supercoils behind the point at which transcription is occurring (Fig. 25–1b). This transcription-driven supercoiling of DNA has been observed both in vitro and, in bacteria, in vivo. In the cell, the

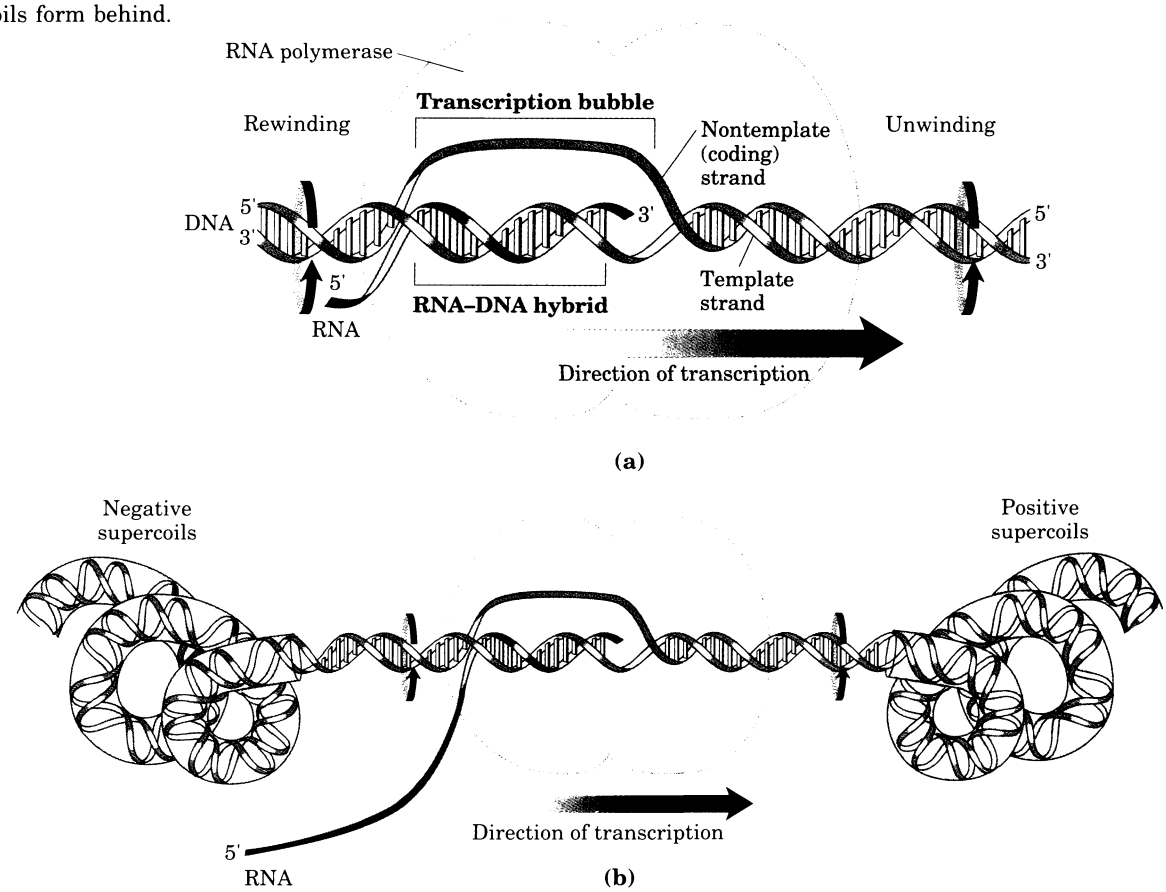


Figure 25–1 Transcription by RNA polymerase in *E. coli*. To synthesize an RNA strand complementary to one of two DNA strands, the DNA is transiently unwound. Strand designations are summarized in Table 25–1. **(a)** About 17 base pairs are unwound at any given time. A short RNA–DNA hybrid (about 12 base pairs) is present in the unwound region. The transcription bubble moves from left to right as shown, keeping pace with RNA synthesis. The DNA is unwound ahead and rewound behind as RNA is transcribed. Arrows show the direction in which the DNA and the RNA–DNA hybrid must rotate to permit this process. As the DNA is rewound, the RNA–DNA hybrid is displaced and the RNA strand is extruded. **(b)** Supercoiling of DNA brought about by transcription. Positive supercoils form ahead of the transcription bubble and negative supercoils form behind.

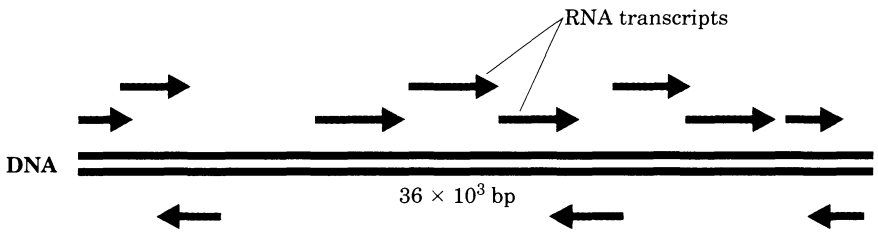


Figure 25–2 The genetic information of the adenovirus is encoded by a double-stranded DNA molecule (36,000 base pairs), both strands of which encode proteins. The information for most proteins is encoded by the top strand (transcribed left to right), but some is encoded by the bottom strand and is transcribed in the opposite direction. Synthesis of mRNAs in adenovirus is actually much more complex than shown here. Many of the mRNAs shown for the upper strand are initially synthesized as one long transcript derived from more than two-thirds of the length of the DNA. The transcript is extensively processed to produce the mRNAs for most of the individual gene products. Adenovirus causes some types of upper respiratory tract infections in some vertebrates.

topological problems caused by transcription are relieved through the action of topoisomerases. Once begun, transcription in *E. coli* proceeds at a rate of about 50 nucleotides per second.

The sequences of two complementary DNA strands are different, and the two strands serve different functions in transcription. A variety of designations are used to distinguish the two strands (Table 25–1). The strand that serves as template for RNA synthesis is called the **template strand** or minus (−) strand. In any chromosome, different genes may use different strands as template (Fig. 25–2). The DNA strand complementary to the template is called the **nontemplate strand** or plus (+) strand. It is identical in base sequence with the RNA transcribed from the gene, with U in place of T (Fig. 25–3). The nontemplate strand is also sometimes called the **coding strand,** even though it has no direct function in either transcription or protein synthesis. The regulatory sequences needed for transcription (described later in this chapter) are by convention given as sequences in the nontemplate (or coding or +) strand.

Table 25–1 Alternative designations for DNA strands in transcription

Template strand	Nontemplate strand
Minus (−) strand	Plus (+) strand
	Coding strand

(5′) CGCTATAGCGTTT(3′)	DNA nontemplate (+) strand
(3′) GCGATATCGCAAA(5′)	DNA template (−) strand
(5′) CGCUAUAGCGUUU(3′)	RNA transcript

Figure 25–3 The two complementary strands of DNA are defined by their function in transcription. The RNA transcript is synthesized on the complementary template (−) strand, and it is identical in sequence (with U in place of T) to the nontemplate (+) or coding strand.

E. coli has a single DNA-directed RNA polymerase that synthesizes all types of RNA. It is a large (M_r 390,000) and complex enzyme, containing five core subunits and a sixth subunit, called σ or σ^{70} (M_r 70,000), that binds transiently to the core and directs the enzyme to specific initiation sites on the DNA (described below). These six subunits constitute the RNA polymerase holoenzyme (Fig. 25–4). RNA polymerases, whether from *E. coli* or other organisms, lack a proofreading 3′→5′ exonuclease activity such as that found in many DNA polymerases. As a result, during transcription about one error is made for every 10^4 to 10^5 ribonucleotides incorporated into RNA. Given that many copies of an RNA are generally produced from a single gene and that all of the RNAs are eventually degraded and replaced, a rare mistake in an RNA molecule is of less consequence to the cell than a mistake in the permanent information stored in DNA.

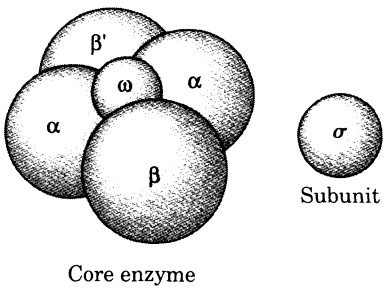


Figure 25–4 The subunit structure of *E. coli* RNA polymerase. The α (of which there are two), β, β', ω, and σ subunits have molecular weights of 36,500, 151,000, 155,000, 11,000, and 70,000, respectively. The σ subunit is also called σ^{70}. The catalytic site for RNA synthesis is believed to be in the β subunit.

RNA Synthesis Is Initiated at Promoters

Initiation of RNA synthesis at random points in a DNA molecule would be an extraordinarily wasteful process. Instead, the RNA polymerase binds to specific sequences in the DNA called **promoters,** which direct the transcription of adjacent segments of DNA (genes). The sequences adjacent to genes where RNA polymerases must bind can be quite variable, and much research has focused on identifying the sequences that are critical to promoter function. Analysis and comparison of sequences in many different bacterial promoters have revealed similarities in two short sequences located about 10 and 35 base pairs away from the point where RNA synthesis is initiated (Fig. 25–5). By convention the base pair that begins an RNA molecule is given the number +1, so these sequences are commonly called the −10 and −35 regions. The sequences are not identical for all bacterial promoters, but certain nucleotides are found much more often than others at each position. The most common nucleotides form what is called a **consensus sequence** (recall the consensus sequences of *ori*C in the *E. coli* chromosome; see Fig. 24–10). For most promoters in *E. coli* and related bacteria, the consensus sequence for the −10 region (also called the Pribnow box) is (5′)TATAAT(3′), and the consensus sequence at the −35 region is (5′)TTGACA(3′).

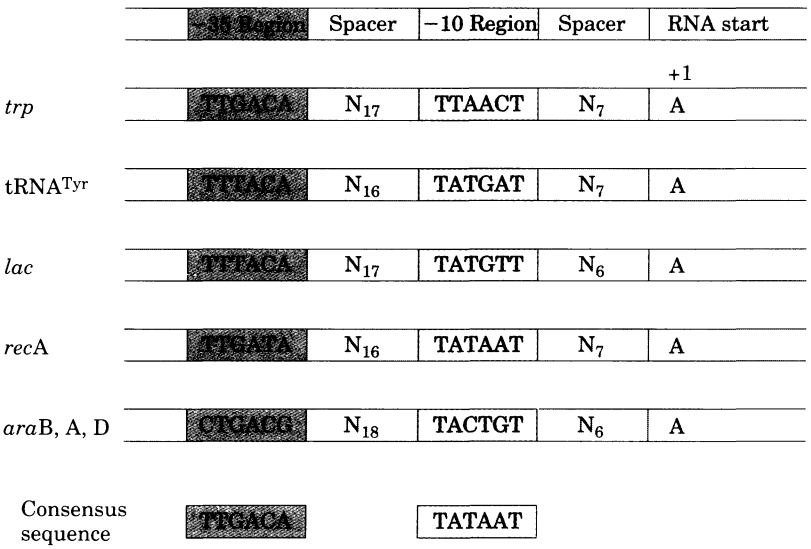
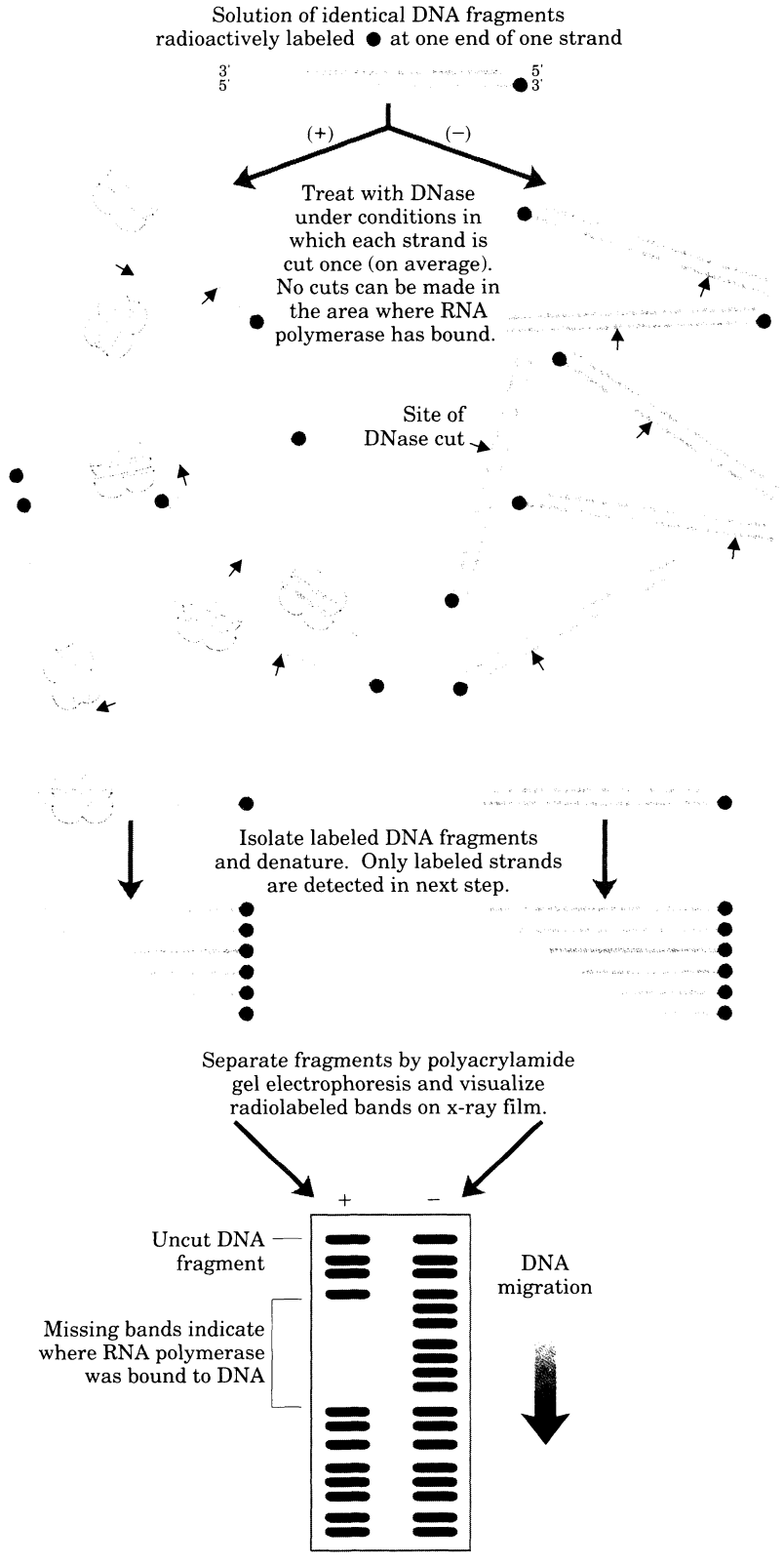


Figure 25–5 The sequences of five *E. coli* promoters. These include promoters for genes involved in *tryp*tophan, *lac*tose, and *ara*binose metabolism. The sequences vary from one promoter to the next, but comparisons of many promoters reveal similarities in the −10 and −35 regions. The consensus sequences of the −10 and −35 regions are shown at the bottom. The −10 region is often called the Pribnow box, after David Pribnow, the investigator who first recognized it in 1975. All sequences shown are those of the coding (nontemplate) strand and read 5′→3′, left to right, as is the convention in representations of this kind. The spacer regions contain variable numbers of nucleotides (N). Only the first nucleotide coding the RNA transcript (at position +1) is shown.

Many independent lines of evidence attest to the functional importance of these sequences. Mutations that affect the function of a given promoter usually involve one of the base pairs in the −35 or −10 region. Natural variations in the consensus sequence also affect the efficiency of RNA polymerase binding and transcription initiation. Differences of a few base pairs can decrease the rate of initiation by several orders of magnitude, providing one means by which *E. coli* can modulate the expression of different genes. In addition, specific binding of RNA polymerase to these sequences has been directly demonstrated in vitro (Box 25–1).

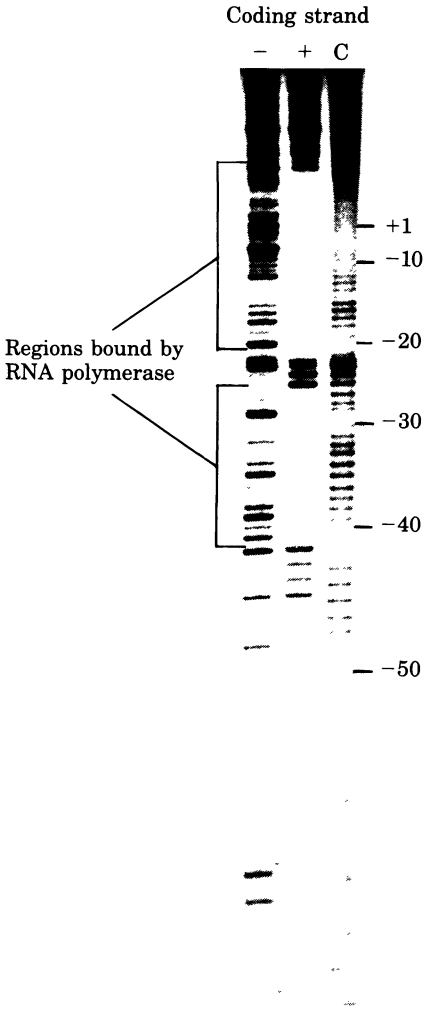
BOX 25–1 RNA Polymerase Leaves Its Footprint on a Promoter



Footprinting, a technique derived from principles used in DNA sequencing (see Fig. 12–35), is used to identify the specific DNA sequences that are bound by a particular protein. A DNA fragment thought to contain sequences recognized by the DNA binding protein is isolated and radiolabeled at one end of one strand (Fig. 1). Chemical or enzymatic cleavage introduces random breaks in the DNA fragment (averaging about one per molecule). Separation of the labeled cleavage products (broken fragments of various lengths) by high-resolution electrophoresis reveals a "ladder" of radioactive bands. In a separate tube the cleavage procedure is repeated on the original DNA fragment to which the protein is bound. The protein prevents cleavage of the DNA in the region to which it is bound. The second set of cleavage prod-

Figure 1 Footprint analysis of the binding site for RNA polymerase on a DNA fragment. Separate experiments are carried out in the presence (+) and absence (−) of RNA polymerase.

Continued on next page



ucts is subjected to electrophoresis side by side with the products of the original reaction. A hole or "footprint" is revealed in the "ladder" of radioactive bands derived from the protein-containing sample. The hole results from the protection of the DNA by protein binding, and it defines the sequences recognized by the protein. The precise location of this binding site can be determined by directly sequencing (see Fig. 12–35) the original DNA fragment and including the sequencing lanes (not shown here) on the same gel with the footprint. Footprinting results for the binding of RNA polymerase to a DNA fragment containing a promoter are shown in Figure 2. The polymerase covers 50 to 60 base pairs; protection by the bound enzyme is concentrated in the −10 and −35 regions.

Figure 2 Footprinting results of RNA polymerase binding to the *lac* promoter (see Fig. 25–5). In this experiment the 5′ end of the coding strand was radioactively labeled. The C lane is a control in which the labeled DNA fragment is cleaved with a chemical reagent that produces a more uniform banding pattern.

RNA polymerase binds to the promoter in at least two distinguishable steps (Fig. 25–6). The holoenzyme first binds the DNA and migrates to the −35 region, forming what is called the "closed complex." The DNA is then unwound for about 17 base pairs beginning at the −10 region, exposing the template strand at the initiation site. The RNA polymerase binds more tightly to this unwound region, forming an "open complex" (the name reflects the state of the DNA). RNA synthesis then begins. The binding of RNA polymerase to promoters is facilitated by the supercoiling (underwinding) of the DNA, which may be one of the reasons why cellular DNA is maintained in an underwound or supercoiled state.

The σ subunit is required only to ensure the specific recognition of the promoter by the RNA polymerase. Once a few phosphodiester bonds are formed the σ subunit dissociates, leaving the core polymerase to complete synthesis of the RNA molecule.

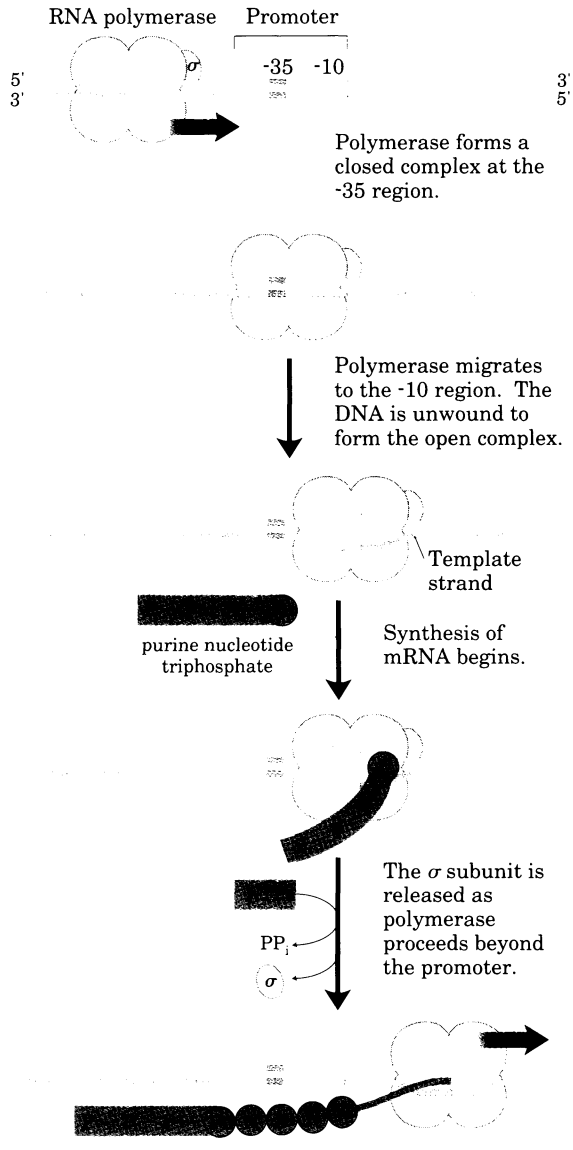


Figure 25–6 Steps in the initiation of transcription by *E. coli* RNA polymerase. RNA polymerase binding to a promoter requires two steps: formation of the closed and open complexes. Messenger RNA synthesis is almost always initiated with a purine (Pu) nucleotide. N is any nucleoside.

Some *E. coli* promoters differ greatly from the standard promoters described above, and recognition of these promoters by RNA polymerase is mediated by different σ factors. An example occurs in a set of genes called the heat-shock genes, which are induced (their gene products are made at higher levels) when the cell is under the stress that accompanies an insult such as a sudden temperature jump. RNA polymerase binds to these promoters when its normal σ subunit (designated σ^{70} because it has a molecular weight of 70,000) is replaced with a different σ subunit that is specific for the heat-shock promoters (see Fig. 27–3). This distinct σ subunit has a molecular weight of 32,000 and is therefore called σ^{32}. The use of different σ factors allows the cell to coordinately express sets of genes involved in major changes in cell physiology.

Initiation of Transcription Is Regulated

Under certain conditions and at different developmental stages, the cellular requirements for any given gene product may very greatly. To provide proteins to the cell in the proportions needed, the transcription of each gene is carefully regulated. The variation in affinity of RNA polymerase for promoters due to differences in promoter sequences, as discussed above, is only one level of control. A variety of proteins bind to sequences in and around the promoter and either activate transcription by facilitating RNA polymerase binding or repress transcription by blocking the activity of polymerase. In *E. coli,* an example of a protein that activates transcription is the **catabolite gene activator protein (CAP),** which increases the transcription of genes coding for enzymes that metabolize sugars other than glucose when cells are grown in the absence of glucose. **Repressors,** typified by the Lac repressor, are proteins that block the synthesis of RNA at specific genes. In the case of the Lac repressor, RNA synthesis is blocked at the genes for enzymes involved in lactose metabolism when lactose is unavailable. Because transcription is the first step in a complicated and energy-intensive pathway leading to protein synthesis, much of the regulation of protein levels in both bacterial and eukaryotic cells is directed at transcription initiation. In Chapter 27 we will describe many mechanisms by which this is accomplished.

Eukaryotic Cells Have Three Kinds of RNA Polymerases

The transcriptional machinery in the nucleus of a eukaryotic cell is much more complex than that in bacteria. Eukaryotes have three different RNA polymerases, designated I, II, and III. Each has a specific function and binds to a different promoter sequence. RNA polymerase I (Pol I) is responsible for the synthesis of only one type of RNA, a

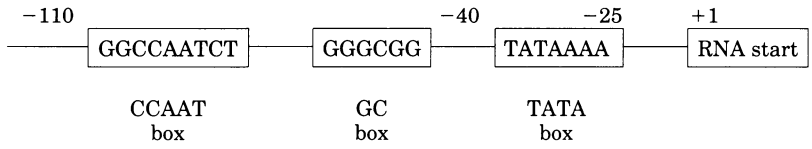


Figure 25–7 The consensus sequences of some common elements in promoters used by eukaryotic RNA polymerase II, derived from a comparison of 100 promoters of this type. A transcription factor (TFIID) binds at the A=T-rich sequence called a TATA box, facilitating the binding of the polymerase. This sequence is commonly found about 25 base pairs before the RNA start site. Two other elements are also sometimes present, found somewhere between −110 and −40: the CCAAT box and GC box are binding sites for other transcription factors that affect polymerase function. Other sequences, some quite distant in the DNA, can affect transcription (Chapter 27). Eukaryotic promoters are more variable than their bacterial counterparts, and some RNA polymerase II promoters lack all of the sequences shown. As in Fig. 25–5, the sequences are those in the coding (nontemplate) strand.

preribosomal RNA transcript that contains the precursor for the 18S, 5.8S, and 28S rRNAs (see Fig. 26–12). Its promoter varies greatly in sequence from one species to another. RNA polymerase II (Pol II) has the central function of synthesizing mRNAs, as well as some special-function RNAs. This enzyme must recognize thousands of promoters, many of which share some key sequence similarities in most eukaryotes (Fig. 25–7). These sequences are generally binding sites for proteins called **transcription factors,** which modulate the binding of RNA polymerase to the promoter. RNA polymerase III (Pol III) makes tRNAs, the 5S rRNA, and some other small specialized RNAs. The promoter recognized by RNA polymerase III is well characterized. Interestingly, some of the sequences required for the regulated initiation of transcription by RNA polymerase III are located within the gene itself, whereas others are found in more conventional locations before the RNA start site (Chapter 27).

Specific Sequences Signal Termination of RNA Synthesis

RNA synthesis proceeds until the RNA polymerase encounters a sequence that triggers its dissociation. This process is not well understood in eukaryotes, and our focus again shifts to bacteria. In *E. coli* there are at least two classes of such termination signals or terminators. One class relies on a protein factor called ρ (rho), and the other is ρ-independent.

The ρ-independent class has two distinguishing features (Fig. 25–8). The first is a region that is transcribed into self-complementary sequences, permitting the formation of a hairpin structure (see Fig. 12–21) centered 15 to 20 nucleotides before the end of the RNA. The second feature is a run of adenylates in the template strand that are transcribed into uridylates at the end of the RNA. It is thought that formation of the hairpin disrupts part of the RNA–DNA hybrid in the transcription complex. The remaining hybrid duplex (oligoribo-U–oligodeoxy-A) contains a particularly unstable combination of bases, and the entire complex simply dissociates.

The ρ-dependent terminators lack the sequence of repeated adenylates in the template but do usually have a short sequence that is transcribed to form a hairpin. RNA polymerase pauses at these sequences, and dissociates if ρ protein is present. The ρ protein has an ATP-dependent RNA–DNA helicase activity and probably disrupts the RNA–DNA hybrid formed during transcription. ATP is hydrolyzed by ρ protein during the termination process, but the detailed mechanism by which the protein acts is not known.

DNA-Directed RNA Polymerase Can Be Selectively Inhibited

The elongation of RNA chains by RNA polymerase in both bacteria and eukaryotes is specifically inhibited by the antibiotic **actinomycin D** (Fig. 25–9). The planar portion of this molecule intercalates (inserts

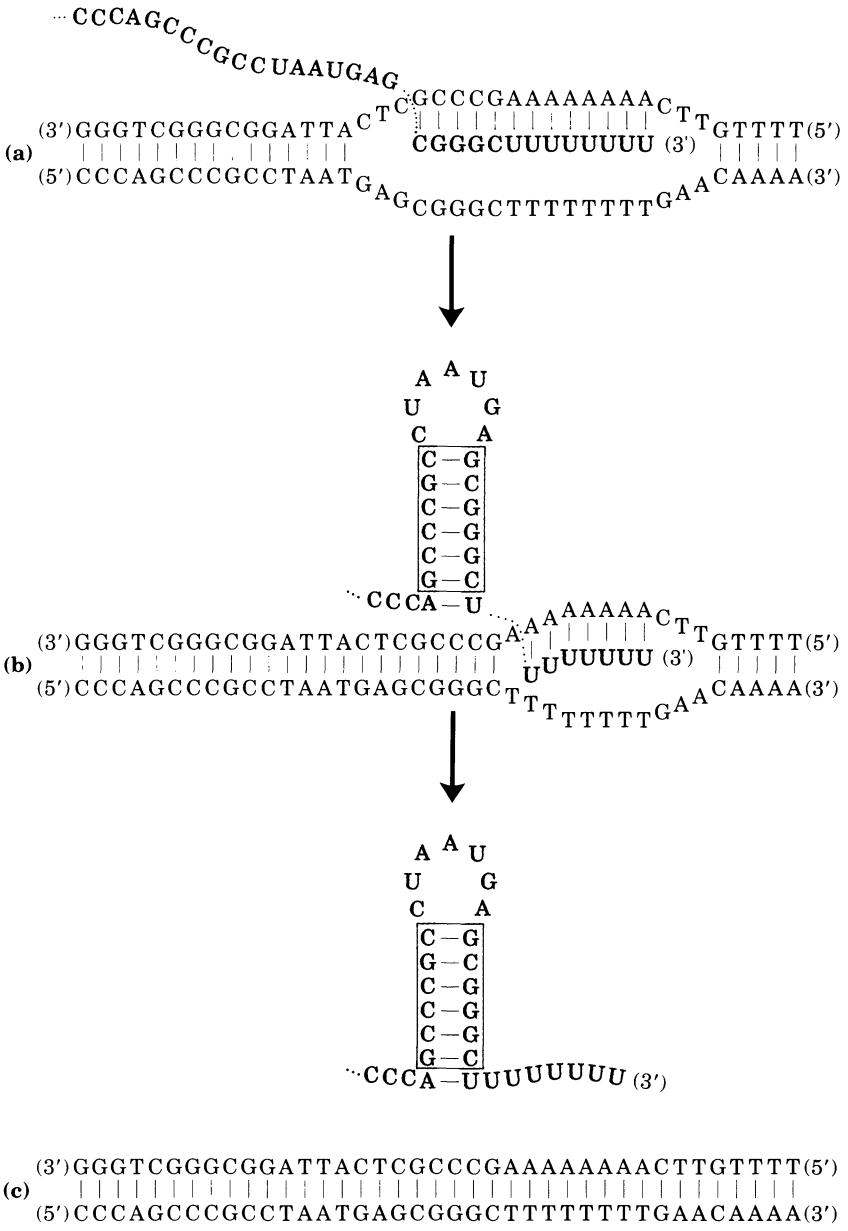
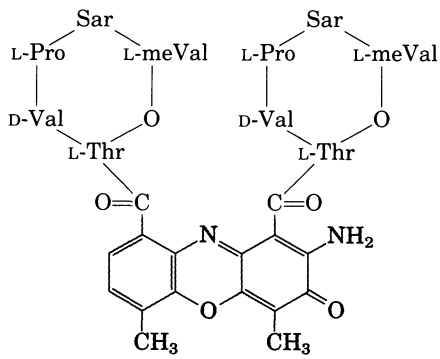
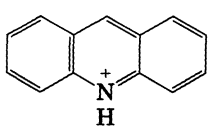


Figure 25–8 A model for ρ-independent termination of transcription in *E. coli*. **(a)** The poly(U) region is synthesized by RNA polymerase. **(b)** Intramolecular pairing of complementary sequences in the RNA forms a hairpin, destroying part of the RNA–DNA hybrid. The remaining A=U hybrid region is relatively unstable, and **(c)** the RNA dissociates completely.



Actinomycin D



Acridine

Figure 25–9 Structure of actinomycin D and acridine, inhibitors of DNA transcription. The shaded portion of actinomycin D is planar and intercalates between two successive G≡C base pairs in duplex DNA. The two cyclic peptide structures of the actinomycin D molecule bind to the minor groove of the double helix. Sarcosine (Sar) is *N*-methylglycine; meVal represents methylvaline. The linkages between sarcosine, L-proline, and D-valine are peptide bonds. Acridine also acts by intercalation in the DNA.

itself) into the double-helical DNA between successive G≡C base pairs, deforming the DNA. This local alteration prevents the movement of the polymerase along the template. In effect, actinomycin D jams the zipper. Because actinomycin D inhibits RNA elongation in intact cells, as well as in cell extracts, it has become very useful for identifying cell processes that depend upon RNA synthesis. **Acridine** inhibits RNA synthesis in a similar fashion (Fig. 25–9).

Rifampicin is an antibiotic inhibitor of RNA synthesis that binds specifically to the β subunit of bacterial RNA polymerases (see Fig. 25–4), preventing the initiation of transcription. A specific inhibitor of RNA synthesis in animal cells is **α-amanitin,** a toxic component of the poisonous mushroom *Amanita phalloides*. It blocks mRNA synthesis by RNA polymerase II and, at higher concentrations, by RNA polymerase III. It does not affect RNA synthesis in bacteria. This mushroom has developed a very effective defense mechanism: a substance that inhibits mRNA formation in organisms that might try to eat it but is evidently harmless to the mushroom's own transcription mechanism.

proteins. Many proteins are clearly derived, at least in part, from exon shuffling during evolution. Walter Gilbert and colleagues have suggested that all present-day proteins may have been assembled from as few as 1,000 to 7,000 primordial exons encoding small polypeptides each 30 to 50 amino acids long.

The origin of life still offers a major intellectual challenge. Even though we cannot go back billions of years and observe the events firsthand, many clues to the puzzle lie buried in the fundamental chemistry of living cells.

Walter Gilbert

Summary

Transcription is catalyzed by DNA-directed RNA polymerase, a complex enzyme that synthesizes RNA complementary to a segment of one strand (the template strand) of duplex DNA, starting from ribonucleoside 5′-triphosphates. To initiate transcription, RNA polymerase binds to a DNA site called a promoter. Bacterial RNA polymerase requires a special subunit for recognizing the promoter. As the first committed step in transcription, binding of RNA polymerase to promoters is subject to many forms of regulation. Eukaryotic cells have three different types of RNA polymerases. Transcription stops at specific sequences called terminators. Many copies of an RNA chain can be transcribed simultaneously from a single gene.

Ribosomal RNAs and transfer RNAs are made from longer precursor RNAs that are trimmed by nucleases, and some bases are modified enzymatically to yield the mature RNAs. In eukaryotes, messenger RNAs are also formed from longer precursors. Primary RNA transcripts often contain noncoding regions called introns, which are removed by splicing. Group I introns are found in rRNAs and their excision requires a guanosine cofactor. Some group I and some group II introns are capable of self-splicing; no protein enzymes are required. Nuclear mRNA precursors have a third class of introns that are spliced with the aid of RNA–protein complexes called snRNPs. The fourth class of introns, found in some tRNAs, are the only ones known to be spliced by protein enzymes. Messenger RNAs are also modified by addition of a 7-methylguanosine residue at the 5′ end, and cleavage and polyadenylation at the 3′ end to form a long poly(A) tail.

The self-splicing introns and the RNA component of RNase P (the enzyme that cleaves the 5′ end of tRNA precursors) form a new class of biological catalysts called ribozymes. These have the properties of true enzymes and are effective catalysts. They promote two types of reaction, hydrolytic cleavage and transesterification, using RNA as substrate. Combinations of these reactions are promoted by the excised group I rRNA intron from *Tetrahymena,* resulting in a type of RNA polymerization reaction. The study of these reactions and of introns themselves has provided insights into likely pathways for biochemical evolution.

Polynucleotide phosphorylase can reversibly form RNA-like polymers from ribonucleoside 5′-diphosphates, adding or removing ribonucleotides at the 3′-hydroxyl end of the polymer. It acts in vivo to degrade RNA.

RNA-directed DNA polymerases, also called reverse transcriptases, are produced in animal cells infected by RNA viruses called retroviruses. These enzymes transcribe the viral RNA into DNA. This process can be used experimentally to form complementary DNA. Many eukaryotic transposons are related to retroviruses, and their mechanism of transposition includes an RNA intermediate. The enzyme that synthesizes telomeres, called telomerase, is a specialized reverse transcriptase that contains an internal RNA template.

RNA-directed RNA polymerases, or replicases, are found in bacterial cells infected with certain RNA viruses. They are template-specific for the viral RNA.

The existence of catalytic RNAs and pathways for the interconversion of RNA and DNA has led to speculation that the earliest living things were made up entirely or largely of RNA molecules that served both for information storage and for catalysis of replication.

General

Darnell, J.E., Jr. (1985) RNA. *Sci. Am.* **253** (October), 68–78.

Evolution of Catalytic Function. (1987) *Cold Spring Harb. Symp. Quant. Biol.* **52.**

An excellent source for articles on catalytic RNA, evolution, and many other topics discussed in this chapter.

Jacob, F. & Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3,** 318–356.

A classic article that introduced many important ideas.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A., & Weiner, A.M. (1987) *Molecular Biology of the Gene,* 4th edn, The Benjamin/Cummings Publishing Company, Menlo Park, CA.

DNA-Directed RNA Synthesis

Conaway, J.W. & Conaway, R.C. (1991) Initiation of eukaryotic messenger RNA synthesis. *J. Biol. Chem.* **266,** 17721–17724.

A good minireview.

McClure, W.R. (1985) Mechanism and control of transcription initiation in prokaryotes. *Annu. Rev. Biochem.* **54,** 171–204.

Platt, T. (1986) Transcription termination and the regulation of gene expression. *Annu. Rev. Biochem.* **55,** 339–372.

Sawadogo, M. & Sentenac, A. (1990) RNA polymerase B (II) and general transcription factors. *Annu. Rev. Biochem.* **59,** 711–754.

A good review of eukaryotic RNA polymerase II.

RNA Processing

Breitbart, R.E., Andreadis, A., & Nadal-Ginard, B. (1987) Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu. Rev. Biochem.* **56,** 467–495.

Cech, T.R. (1986) RNA as an enzyme. *Sci. Am.* **255** (November), 64–75.

Cech, T.R. (1987) The chemistry of self-splicing RNA and RNA enzymes. *Science* **236,** 1532–1539.

Deutscher, M.P. (1990) Ribonucleases, tRNA nucleotidyltransferase, and the 3′ processing of tRNA. *Prog. Nucleic Acid Res. Mol. Biol.* **39,** 209–240.

A good overview of tRNA processing reactions.

Green, M.R. (1986) Pre-mRNA splicing. *Annu. Rev. Genet.* **20,** 671–708.

McCorkle, G.M. & Altman, S. (1987) RNAs as catalysts: a new class of enzyme. *J. Chem. Educ.* **64,** 221–226.

Pace, N.R. & Smith, D. (1990) Ribonuclease P: function and variation. *J. Biol. Chem.* **265,** 3587–3590.

Ross, J. (1989) The turnover of messenger RNA. *Sci. Am.* **260** (April), 48–55.

Sharp, P.A. (1987) Splicing of messenger RNA precursors. *Science* **235,** 766–771.

Wahle, E. & Keller, W. (1992) The biochemistry of 3′-end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem.* **61,** 419–440.

Wickens, M. (1990) How the messenger got its tail: addition of poly(A) in the nucleus. *Trends Biochem. Sci.* **15,** 277–281.

RNA-Directed RNA or DNA Synthesis

Belfort, M. (1991) Self-splicing introns in prokaryotes: migrant fossils? *Cell* **64,** 9–11.

A discussion of the evolutionary significance of prokaryotic introns.

Bishop, J.M. (1991) Molecular themes in oncogenesis. *Cell* **64,** 235–248.

A good overview of oncogenes; it introduces a series of more detailed reviews included in the same issue of Cell.

Blackburn, E.H. (1991) Telomeres. *Trends Biochem. Sci.* **16,** 378–381.

Blackburn, E.H. (1992) Telomerases. *Annu. Rev. Biochem.* **61,** 113–129.

Boeke, J.D. (1990) Reverse transcriptase, the end of the chromosome, and the end of life. *Cell* **61,** 193–195.

The possible role of telomerase in regulating the life span of an organism.

Dorit, R.L., Schoenbach, L., & Gilbert, W. (1990) How big is the universe of exons? *Science* **250,** 1377–1382.

Interesting speculation on the origin and function of exons.

Gallo, R.C. & Montagnier, L. (1988) AIDS in 1988. *Sci. Am.* **259** (October), 40–48.

The introductory article to an entire Scientific American *issue devoted to AIDS.*

Kingsman, A.J. & Kingsman, S.M. (1988) Ty: a retroelement moving forward. *Cell* **53,** 333–335.

Describes a well-studied yeast transposon related to retroviruses.

Pace, N.R. (1991) Origin of life—facing up to the physical setting. *Cell* **65,** 531–533.

A discussion of the conditions believed to have existed when life began evolving.

Perlman, P.S. & Butow, R.A. (1989) Mobile introns and intron-encoded proteins. *Science* **246,** 1106–1109.

This article describes a special class of introns capable of colonizing homologous genes that lack introns.

Temin, H.M. (1976) The DNA provirus hypothesis: the establishment and implications of RNA-directed DNA synthesis. *Science* **192,** 1075–1080.

A discussion of the original proposal for reverse transcription in retroviruses.

Varmus, H. (1987) Reverse transcription. *Sci. Am.* **257** (September), 56–64.

Varmus, H.E. (1989) Reverse transcription in bacteria. *Cell* **56,** 721–724.

Problems

1. *RNA Polymerase* How long would it take for the *E. coli* RNA polymerase to synthesize the primary transcript for *E. coli* rRNAs (6500 bases)?

2. *Error Correction by RNA Polymerases* DNA polymerases are capable of editing and error correction, but RNA polymerases do not appear to have this capacity. Given that a single base error in either replication or transcription can lead to an error in protein synthesis, can you give a possible biological explanation for this striking difference?

3. *The Rate of Transcription* From what you know of the rate at which *E. coli* RNA polymerase synthesizes RNA, predict how far the transcription "bubble" formed by RNA polymerase will move along the DNA in 10 s.

4. *RNA Posttranscriptional Processing* Predict the likely effects of a mutation in the sequence (5′)AAUAAA in a eukaryotic mRNA transcript.

5. *Coding vs. Template Strands* The RNA genome of phage Qβ is the nontemplate or (+) strand, and when introduced into the cell it functions as an mRNA. Suppose the RNA replicase of phage Qβ synthesized primarily (−) strand RNA and uniquely incorporated it into the virus particles, rather than (+) strands. What would be the fate of the (−) strands when they entered a new cell? What enzyme would such a (−) strand virus need to include in the virus particle to successfully invade a host cell?

6. *The Chemistry of Nucleic Acid Biosynthesis* Describe three properties common to the reactions catalyzed by DNA polymerase, RNA polymerase, reverse transcriptase, and RNA replicase.

7. *RNA Splicing* What is the minimum number of transesterification reactions needed to splice an intron from an mRNA transcript? Why?

8. *Telomerase* Assuming that the RNA component of telomerase is fixed within the protein structure, in what respect might the active site of this enzyme differ from the active site of reverse transcriptases, RNA polymerases, and DNA polymerases? (Hint: The latter three enzymes add one nucleotide at a time.)

9. *RNA Genomes* The RNA viruses have relatively small genomes. For example, the single-stranded RNAs of retroviruses have about 10,000 nucleotides and the Qβ RNA is only 4,220 nucleotides long. Given the properties of reverse transcriptase and RNA replicase described in this chapter, can you suggest a reason for the small size of these viral genomes?