9

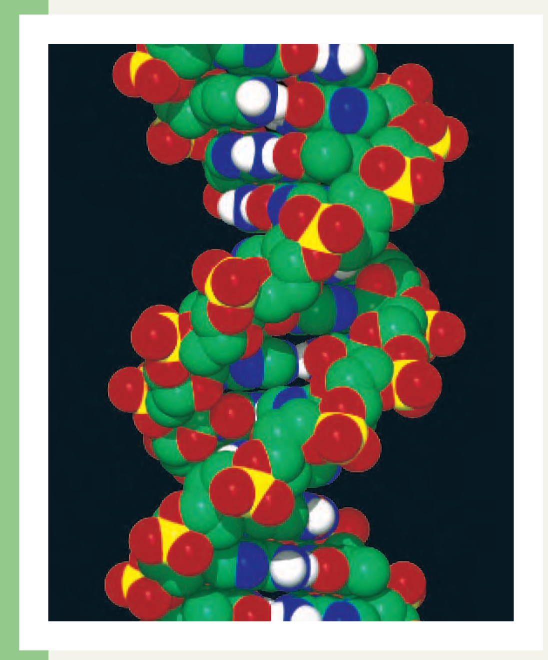
CHEMISTRY OF THE GENE

STUDY OBJECTIVES

1. To understand the properties that a genetic material must have 205
2. To examine the structure of DNA, the genetic material 211
3. To investigate the way in which DNA replicates 220

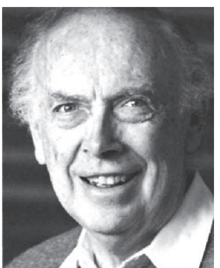
STUDY OUTLINE

In Search of the Genetic Material 205
Required Properties of a Genetic Material 205
Evidence for DNA as the Genetic Material 209
Chemistry of Nucleic Acids 211
Biologically Active Structure 214
Requirements of Genetic Material 218
Alternative Forms of DNA 219
DNA Replication—The Process 220
The Meselson and Stahl Experiment 220
Autoradiographic Demonstration of DNA Replication 222
DNA Replication—The Enzymology 225
Continuous and Discontinuous DNA Replication 225
The Origin of DNA Replication 229
Events at the Y-Junction 232
Supercoiling 234
Termination of Replication 236
DNA Partitioning in *E. coli* 238
Replication Structures 238
Rolling-Circle Model 238
D-Loop Model 238
Eukaryotic DNA Replication 238
Summary 240
Solved Problems 240
Exercises and Problems 241
Critical Thinking Questions 242
Box 9.1 Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid 206
Box 9.2 Prions: The Biological Equivalent of Ice-Nine 213
Box 9.3 Multiple-Stranded DNA 221

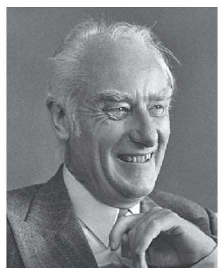


A computer-generated image of deoxyribonucleic acid, DNA. (© Professor K. Seddon & Dr. T. Evans/ Queen's University Belfast/SPL/Photo Researchers.)

In 1953, James Watson and Francis Crick published a two-page paper in the journal *Nature* entitled "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid." It began as follows: "We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest." This paper, which first put forth the correct model of DNA structure, is a milestone in the modern era of molecular genetics, compared by some to the work of Mendel and Darwin (box 9.1). (Watson, Crick, and X-ray crystallographer Maurice Wilkins won Nobel Prizes for this work; Rosalind Franklin, also an X-ray crystallographer, was acknowledged, posthumously, to have played a major role in the discovery of the structure of DNA.) Once the structure of the genetic material had been determined, an understanding of its method of replication and its functioning quickly followed.



James D. Watson (1928–). (Cold Spring Harbor Laboratory Research Library Archives. Margot Bennet, photographer.)



Francis Crick (1916–). (Reproduced by permission of Herb Weitman, Washington University, St. Louis, Missouri.)

Maurice H. F. Wilkins (1916–). (Courtesy of Dr. Maurice H. F. Wilkins and Biophysics Department, King's College, London.)

IN SEARCH OF THE GENETIC MATERIAL

This chapter begins a sequence of nine chapters on the molecular structure of the genetic material, its replication, its expression, and the control of its expression. In this chapter, we look at the evidence that DNA is the genetic material, the chemistry of DNA, and the way in which DNA replicates, including the general enzymatic processes. We look first at prokaryotic, then at eukaryotic, DNA replication. Note that we concentrate on the molecular structure of DNA because, generally, structure reveals function: molecules have shapes that define how they work.

Required Properties of a Genetic Material

We begin with a look at the properties that a genetic material must have and review the evidence that nucleic acids make up the genetic material. To comprise the genes, DNA must carry the information to control the synthesis of the enzymes and proteins within a cell or organism; self-replicate with high fidelity, yet show a low level of mutation; and be located in the chromosomes.

Control of the Proteins

The growth, development, and functioning of a cell are controlled by the proteins within it, primarily its enzymes. Thus, the nature of a cell's phenotype is controlled by the protein synthesis within that cell. The genetic material must therefore determine the need for and effective amounts of the enzymes in a cell. For example, given inorganic salts and glucose, an *E. coli* cell can synthesize, through its enzyme-controlled biochemical pathways, all of the compounds it needs for growth, survival, and reproduction. In contrast, a mammalian red blood cell primarily produces hemoglobin.

At this point we need to review some basic information regarding enzymes. An enzyme is a protein that acts as a catalyst for a specific metabolic process without itself being markedly altered by the reaction. Most reactions that enzymes catalyze could occur anyway, but only under conditions too extreme to take place within living systems. For example, many oxidations occur naturally at high temperatures. Enzymes allow these reactions to occur within the cell by lowering the **free energy of activation ($\Delta G^{\ddagger}$)** of a particular reaction. In other words, an enzyme allows a reaction to take place without needing the boost in energy that heat usually supplies (fig. 9.1).

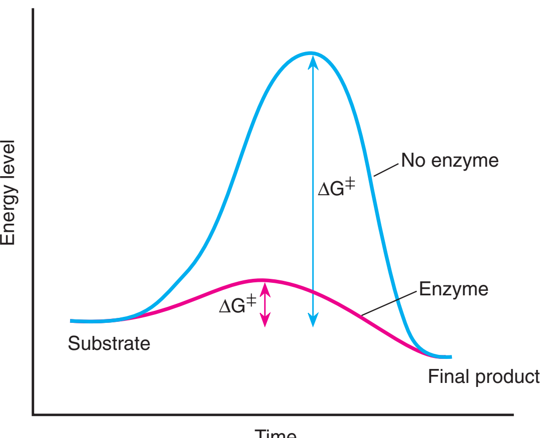


Figure 9.1 An enzyme lowers the free energy of activation ($\Delta G^{\ddagger}$) for a particular reaction.

BOX 9.1

Historical Perspectives

Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey.[1] They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram [fig. 1]). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′, 5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration,' the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3.4 A in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

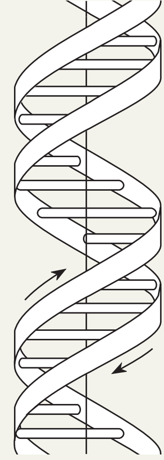


Figure 1 This figure is purely diagrammatic. The two *ribbons* symbolize the two phosphate-sugar chains, and the *horizontal rods* represent the pairs of bases holding the chains together. The *vertical line* marks the fiber axis. (Reprinted with permission from *Nature,* Vol. 171, No. 4356. Watson and Crick, "Molecular Structure of Nucleic Acids," pp. 737–738. Copyright © 1953 Macmillan Magazines Limited.)

Most metabolic processes, such as the biosynthesis or degradation of molecules, occur in pathways, with enzyme facilitating each step in the pathway (see chapter 2). The metabolic pathway for the conversion of threonine into isoleucine (two amino acids) appears in figure 9.2. Each reaction product in the pathway is altered by an enzyme that converts it to the next product. The enzyme threonine dehydratase, for example, converts threonine into α-ketobutyric acid. Enzymes are composed of folded polymers of amino acids. The average protein is three hundred to five hundred amino acids long; only twenty naturally occurring amino acids are used in constructing these proteins. The sequence of amino acids determines the final structure of an enzyme. (We discuss the structure of proteins in more detail in chapter 11.) The genetic material determines the sequence of the amino acids.

The three-dimensional structure of enzymes permits them to perform their function. An enzyme combines with its substrate or substrates (the molecules it works on) at a part of the enzyme called the **active site** (fig. 9.3). The substrates "fit" into the active site, which has a shape that allows only the specific substrates to enter. This view of the way an enzyme interacts with its substrates is called the *lock-and-key model* of enzyme functioning.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations), it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid. It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereo-chemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their coworkers at King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. Watson
F. H. C. Crick

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge. April 2.

Reprinted by permission from *Nature,* 171:737–38. Copyright © 1953 Macmillan Magazines Ltd.

1. Pauling, L., and Corey, R. B., *Nature,* 171, 346 (1953); *Proc. U.S. Nat. Acad. Sci.,* 39, 84 (1953).
2. Furberg, S., *Acta Chem. Scand.,* 6, 634 (1952).
3. Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta,* 9, 402 (1952).
4. Wyatt, G. R., *J. Gen. Physiol.,* 36, 201 (1952).
5. Astbury, W. T., *Symp. Soc. Exp. Biol. 1, Nucleic Acid* 66 (Camb. Univ. Press, 1947).
6. Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys.* Acta, 10, 192 (1953).

When the substrates are in their proper position in the active site of the enzyme, the particular reaction that the enzyme catalyzes takes place. The reaction products then separate from the enzyme and leave it free to repeat the process. Enzymes can work at phenomenal speeds. Some can catalyze as many as a million reactions per minute.

Not all of the cell's proteins function as catalysts. Some are structural proteins, such as keratin, the main component of hair. Other proteins are regulatory—they control the rate at which other enzymes work. Still others are involved in different functions; albumins, for example, help regulate the osmotic pressure of blood.

Replication

The genetic material must be capable of precisely directing its own replication so that every daughter cell receives an exact copy. Some **mutability,** or the ability to change, is also required, because we know that the genetic material has changed, or evolved, over the history of life on earth. In their 1953 paper, Watson and Crick had already worked out the replication process based on the structure of DNA. The fidelity of the replication process is so great that the error rate is only about one in a billion.

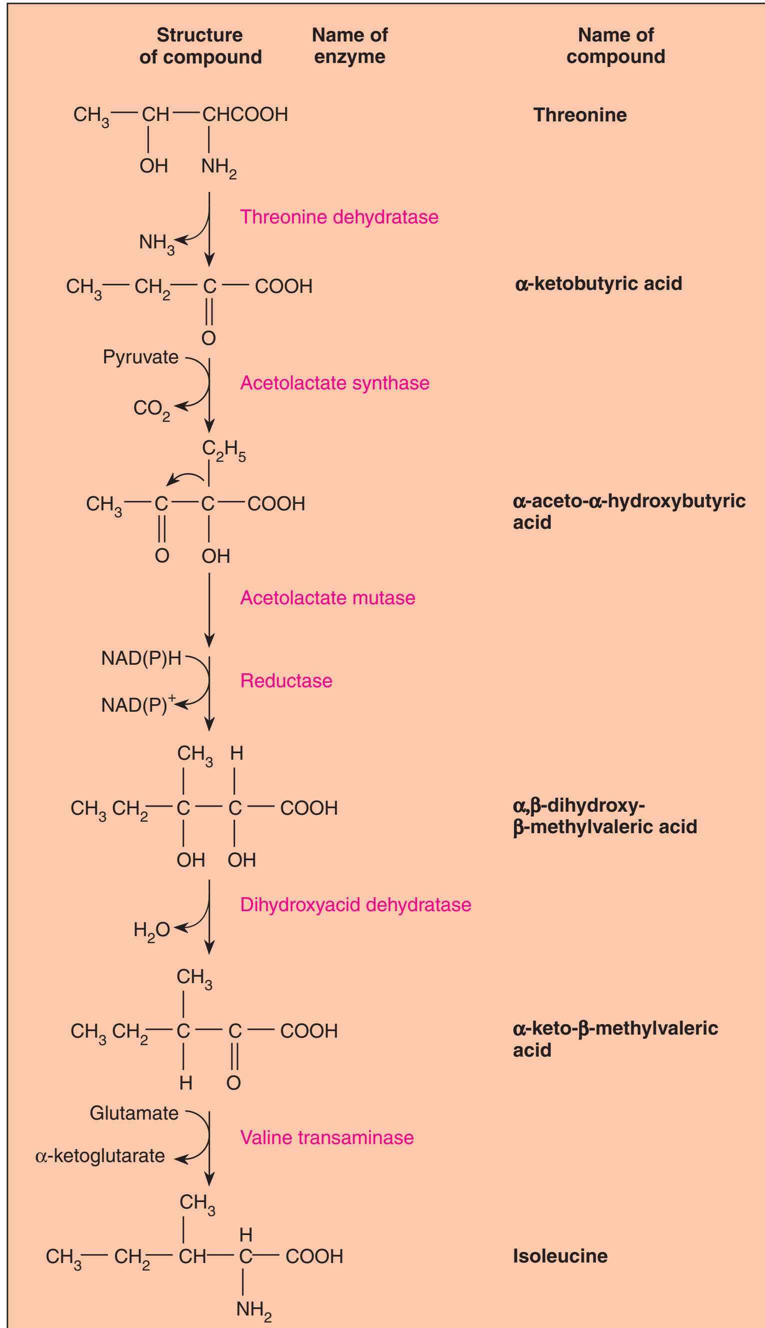


Figure 9.2 Metabolic pathway of conversion of the amino acid threonine into isoleucine.

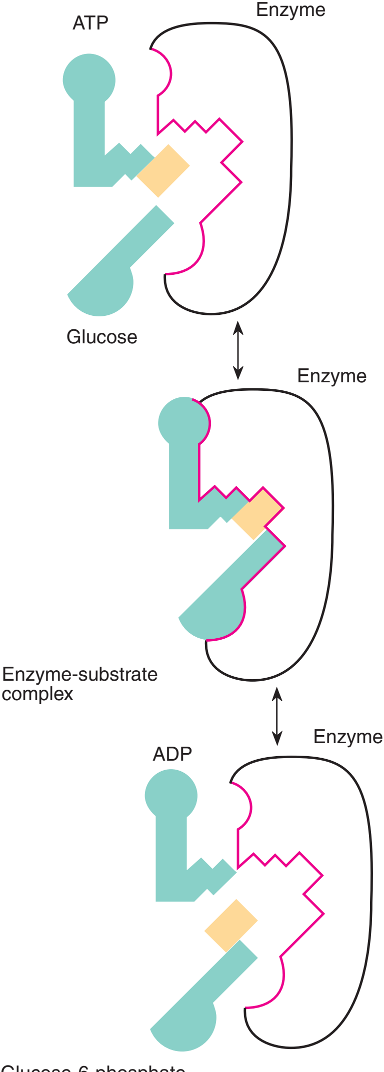


Figure 9.3 The active site of an enzyme recognizes a specific substance. In this case, ATP plus glucose is converted into ADP and glucose-6-phosphate by the enzyme hexokinase. The active site is diagrammed in *red.* The terminal phosphate group of ATP is *tan*.

Location

It has been known since the turn of the century that genes, the discrete functional units of genetic material, are located in chromosomes within the nuclei of eukaryotic cells: the way chromosomes behave during the cellular division stages of mitosis and meiosis mimics the behavior of genes. Thus, the genetic material in eukaryotes must be a part of the chromosomes.

For a long time, proteins were considered the most probable genetic material because they have the neces-

Oswald T. Avery (1877–1955). (Courtesy of the National Academy of Sciences.)

sary molecular complexity. The twenty naturally occurring amino acids can be combined in an almost unlimited variety, creating thousands and thousands of different proteins. The first proof that the genetic material is deoxyribonucleic acid (DNA) came in 1944 from Oswald Avery and his colleagues. The Watson and Crick model in 1953 ended a period when many thought DNA was the genetic material, but its structure was unknown.

Evidence for DNA as the Genetic Material

Transformation

In 1928, F. Griffith reported that heat-killed bacteria of one type could "transform" living bacteria of a different type. Griffith demonstrated this transformation using two strains of the bacterium *Streptococcus pneumoniae.* One strain (S) produced smooth colonies on media in a petri plate because the cells had polysaccharide capsules. It caused a fatal bacteremia (bacterial infection) in mice. Another strain (R), which lacked polysaccharide capsules, produced rough colonies on petri plates (fig. 9.4); it did not have a pathological effect on mice. Bacteria of the rough strain are engulfed by the mice's white blood cells; bacteria of the virulent smooth strain survive because their polysaccharide coating protects them.

Griffith found that neither heat-killed S-type nor live R-type cells, by themselves, caused bacteremia in mice. However, if he injected a mixture of live R-type and heat-killed S-type cells into mice, the mice developed a bacteremia identical to that caused by living S-type cells (fig. 9.5). Thus, something in the heat-killed S cells transformed the R-type bacteria into S-type cells.

In 1944, Oswald Avery and two of his associates, C. MacLeod and M. McCarty, reported the nature of the transforming substance. Avery and his colleagues did their work **in vitro** (literally, in glass), using colony morphology on culture media rather than bacteremia in mice as evidence of transformation. They ruled out proteins, carbohydrates, and lipids by their extraction procedure, by the chemical analysis of the transforming material, and by demonstrating that the only enzymes that destroyed the transforming ability were enzymes that destroyed DNA. This study provided the first experimental evidence that DNA was the genetic material: DNA transformed R-type bacteria into S-type bacteria.

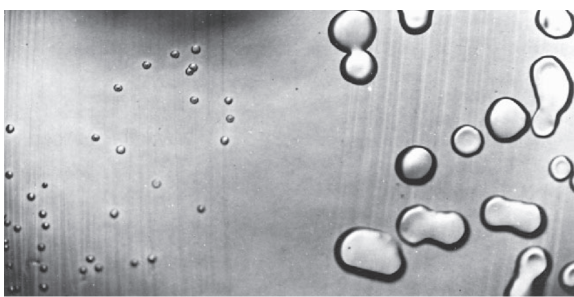
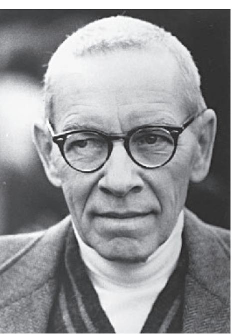


Figure 9.4 Petri plate with smooth and rough colonies of *Streptococcus pneumoniae.* R (rough) strain colonies appear on the *left* and S (smooth) colonies on the *right* on the same agar. Magnification 3.5×. (O. T. Avery, C. M. Macleod, and M. McCarty, "Studies on the chemical nature of the substance inducing transformation of pneumococcal types." Reproduced from the *Journal of Experimental Medicine* 79 (1944):137–58, fig. 1 by copyright permission of the Rockefeller University Press. Reproduced by permission. Photograph made by Mr. Joseph B. Haulenbeek.)

Phage Labeling

Valuable information about the nature of the genetic material has also come from viruses. Of particular value are studies of bacterial viruses—the bacteriophages, or phages. Since phages consist only of nucleic acid surrounded by protein, they lend themselves nicely to the determination of whether the protein or the nucleic acid is the genetic material.

A. D. Hershey and M. Chase published, in 1952, the results of research that supported the notion that DNA is the genetic material and, in the process, helped to



A. D. Hershey (1908–1997). (Courtesy of Dr. A. D. Hershey.)

Martha Chase. (Courtesy of Cold Springs Harbor Laboratory Archives.)

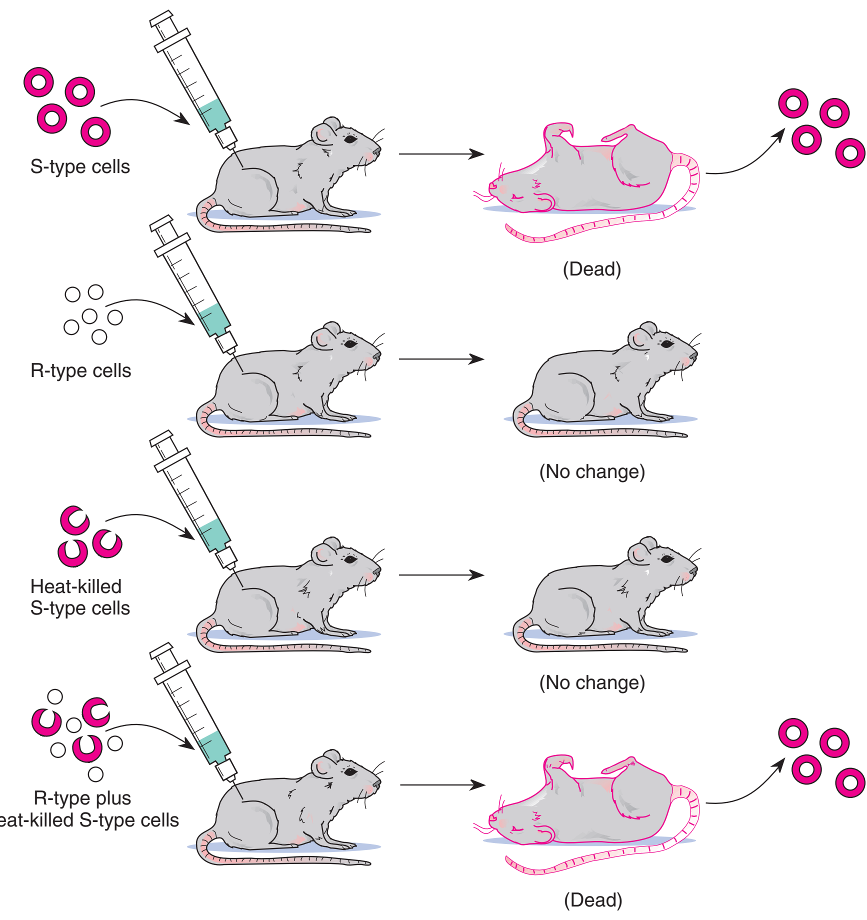


Figure 9.5 Griffith's experiment with *Streptococcus.* S-type cells will kill mice; so will heat-killed S-type cells injected with live R-type cells. S-type cells are recovered from dead mice in both cases.

explain the nature of the viral infection process. Since all nucleic acids contain phosphorus, whereas proteins do not, and since most proteins contain sulfur (in the amino acids cysteine and methionine), whereas nucleic acids do not, Hershey and Chase designed an experiment using radioactive isotopes of sulfur and phosphorus to keep separate track of the viral proteins and nucleic acids during the infection process. They used the T2 bacteriophage and the bacterium *Escherichia coli.* The phages were labeled by having them infect bacteria growing in culture medium containing the radioactive isotopes ^{35}S or ^{32}P. Hershey and Chase then proceeded to identify the material injected into the cell by phages attached to the bacterial wall.

When ^{32}P-labeled phages were mixed with unlabeled *E. coli* cells, Hershey and Chase found that the ^{32}P label entered the bacterial cells and that the next generation of phages that burst from the infected cells carried a significant amount of the ^{32}P label. When ^{35}S-labeled phages were mixed with unlabeled *E. coli,* the researchers found that the ^{35}S label stayed outside the bacteria for the most part. Hershey and Chase thus demonstrated that the outer protein coat of a phage does not enter the bacterium it infects, whereas the phage's inner material, consisting of DNA, does enter the bacterial cell (fig. 9.6). Since the DNA is responsible for the production of the new phages during the infection process, the DNA, not the protein, must be the genetic material.

RNA as Genetic Material

In some viruses, RNA (ribonucleic acid) is the genetic material. The tobacco mosaic virus that infects tobacco plants consists only of RNA and protein. The single, long RNA molecule is packaged within a rodlike structure formed by over two thousand copies of a single protein. No DNA is present in tobacco mosaic virus particles (fig. 9.7*a*). In 1955, H. Fraenkel-Conrat and R. Williams

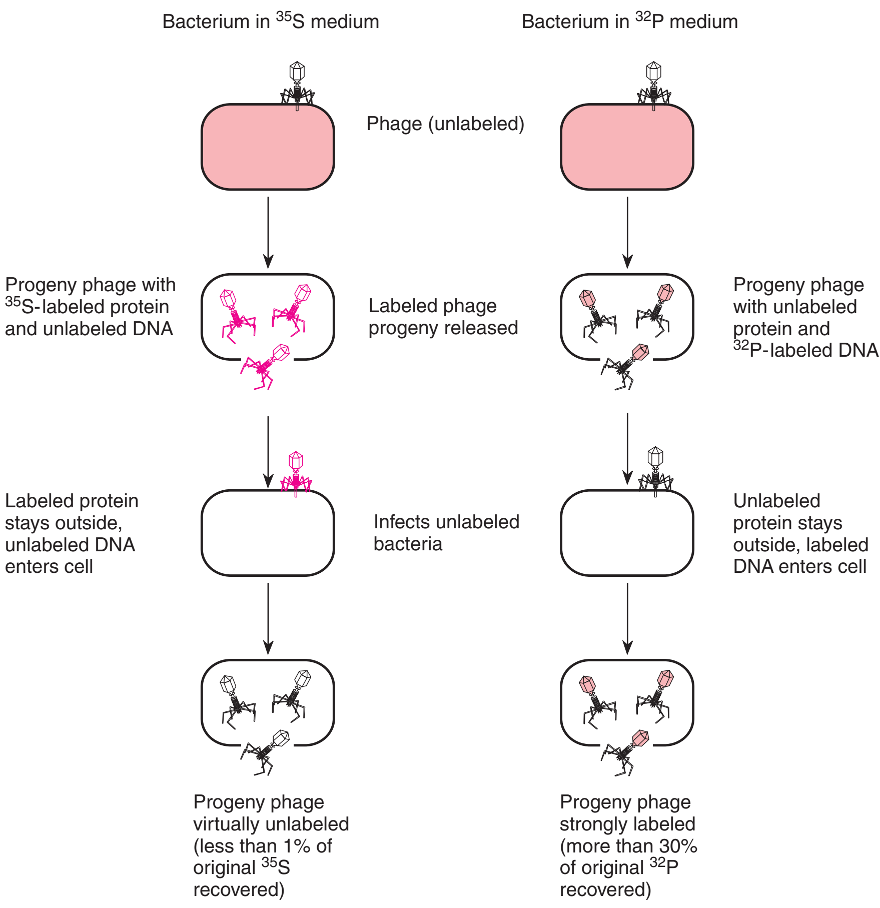


Figure 9.6 The Hershey and Chase experiments using ^{35}S-labeled and ^{32}P-labeled T2 bacteriophages. The nucleic acid label (^{32}P) enters the *E. coli* bacteria during infection; the protein label (^{35}S) does not.

showed that a virus can be separated, in vitro, into its component parts and reconstituted as a viable virus. This finding led Fraenkel-Conrat and B. Singer to reconstitute tobacco mosaic virus with parts from different strains (fig. 9.7*b*). For example, they combined the RNA from the common tobacco mosaic virus with the protein from the masked (M) strain of tobacco mosaic virus. They then made the reciprocal combination of common-type protein and M-type RNA. In both cases, the tobacco mosaic virus produced during the process of infection was the type associated with the RNA, not with the protein. Thus, it was the nucleic acid (RNA in this case) that was the genetic material. Subsequently, scientists rubbed pure tobacco mosaic virus RNA into plant leaves. Normal infection and a new generation of typical, protein-coated tobacco mosaic virus resulted, confirming RNA as the genetic material for this virus.

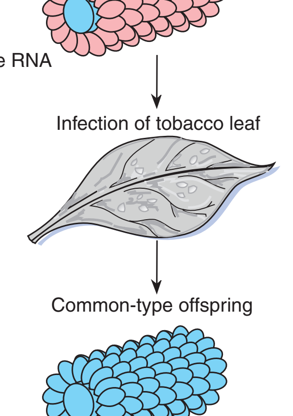
We thus conclude that DNA is the genetic material. In the few viruses that do not have DNA, RNA serves as the genetic material. The only exception to these statements is one type of disease that is transmitted by a protein without accompanying DNA or RNA (box 9.2).

CHEMISTRY OF NUCLEIC ACIDS

Having identified the genetic material as the nucleic acid DNA (or RNA), we proceed to examine the chemical structure of these molecules. Their structure will tell us a good deal about how they function.

Nucleic acids are made by joining **nucleotides** in a repetitive way into long, chainlike polymers. Nucleotides are made of three components: phosphate, sugar, and a nitrogenous base (table 9.1 and fig. 9.8). When incorporated into a nucleic acid, a nucleotide contains one of each of the three components. But, when free in the cell

(a)



(b)

Figure 9.7 (*a*) Electron micrograph of tobacco mosaic virus. Magnification 37,428×. (*b*) Reconstitution experiment of Fraenkel-Conrat and Singer. The nucleic acid (RNA), not the protein component of the virus, controls inheritance.

([*a*] © Biology Media/Photo Researchers, Inc.)

Table 9.1 Components of Nucleic Acids

	Phosphate	Sugar	Base: Purines	Base: Pyrimidines
DNA	Present	Deoxyribose	Guanine Adenine	Cytosine Thymine
RNA	Present	Ribose	Guanine Adenine	Cytosine Uracil

pool, nucleotides usually occur as triphosphates. The energy held in the extra phosphates is used, among other purposes, to synthesize the polymer. A **nucleoside** is a sugar-base compound. Nucleotides are therefore nucleoside phosphates (fig. 9.9). (Note that ATP, adenosine triphosphate, the energy currency of the cell, is a nucleoside triphosphate.)

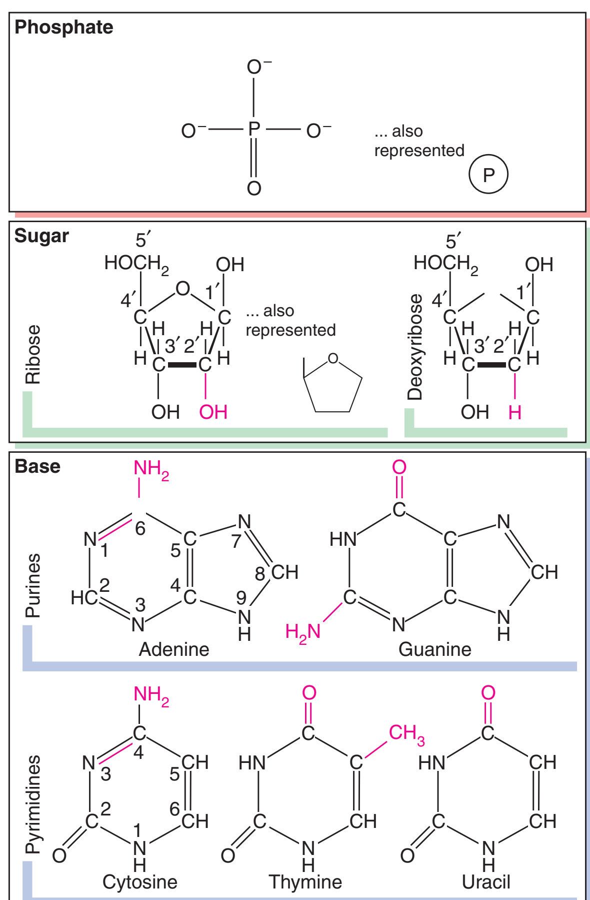


Figure 9.8 Components of nucleic acids: phosphate, sugars, and bases. Primes are used to number the ring positions in the sugars to differentiate them from the ring positions in the bases.

BOX 9.2

Biomedical Applications

Prions: The Biological Equivalent of Ice-Nine

Without exception, the genetic material is either DNA or RNA; it is RNA only in a few viruses. Since virtually all transmissible diseases are of bacterial or viral origin, this means that transmissible diseases are also caused by organisms with DNA or RNA as their genetic material. However, in one interesting situation, a transmissible disease appears to be caused by an agent without genetic material. Four human neurological diseases and six similar animal diseases are caused, we believe, by proteins without DNA or RNA. (Two conditions in yeast are probably caused in a similar way.) The human diseases are kuru, Creutzfeldt-Jakob disease, Gerstmann-Sträussler-Scheinker syndrome, and a recently discovered fatal familial insomnia. The animal diseases are scrapie (sheep and goats), four encephalopathies (bovine, feline, ungulate, and mink), and chronic wasting disease (deer and elk). All of these diseases are extremely slow to develop, all are fatal, and all are believed to be caused either by the ingestion of a protein from an infected individual or from a mutation of the normal gene. None of the diseases as yet has a cure, and the mechanism of action is not completely understood.

The diseases appear to be caused by a protein, similar to one normally produced in the brain of healthy individuals. The term **prion** (taken from *pro*teinaceous *in*fectious particle) has been given to these agents by Stanley Prusiner at the University of California in San Francisco, a 1997 Nobel laureate. He, along with colleagues, isolated the prion protein (PrP) and recently located the gene that codes for the protein on the short arm of chromosome 20. In addition to the infective form, a familial (inherited) form of these diseases can result from a mutation of the gene that codes for the prion protein active in normal individuals (probably at least all mammals). The normal protein is termed PrP^C, and the mutated form is referred to as PrP^{Sc}. Normally, PrP^C is a glycoprotein found on the membrane surface of the cells of the brain and some other tissues.

Although no cures exist for these diseases, kuru, at least, seems to be almost eradicated. It was found only among people in part of New Guinea who practiced cannibalism. Once the people stopped this practice, the spread of the disease also ceased; kuru does not seem to be generated to any major extent by mutation. By controlling feeding practices, it is believed, bovine spongiform encephalopathy will also disappear. In the past, cows were fed protein supplements contaminated by material from infected animals.

In England, a recent epidemic of bovine spongiform encephalopathy (BSE, or mad cow disease) peaked in 1992–1993, affecting over 160,000 cattle. At least fourteen cases of a variant of Creutzfeldt-Jakob disease in people in England and France were attributed to eating affected beef, creating a panic in England. With a change away from using animal matter in cattle feed and a culling of cattle herds, the epidemic has ended. However, new human cases may show up in the future owing to the long incubation period of this prion disease.

The obvious question is, how does a protein that does not appear to contain genetic material cause a transmissible disease when ingested? Prusiner has suggested several mechanisms that would allow an infective protein to induce copies of the normal protein to become infective. One of these mechanisms involves a cascade in which an infective PrP^{Sc} binds with a normal PrP^C, resulting in two infective PrP^{Sc} proteins. From this, one produces two, two produce four, four produce eight, and so on. As Nancy Touchette, writing in *The Journal of NIH Research,* pointed out, this is the way Kurt Vonnegut described the behavior of the mythical ice-nine in his 1963 book, *Cat's Cradle.* In this fictional account, a single seed caused all of the water on earth, by a chain reaction cascade, to form into a novel type of ice. We have not yet resorted to science fiction to answer the mystery of prion function; however, it seems reasonable to guess that an eventual understanding of the mechanism of prion function will provide us with a biological novelty.

The sugars differ only in the presence (ribose in RNA) or absence (deoxyribose in DNA) of an oxygen in the 2′ position. (The carbons of the sugars are numbered 1′ to 5′. The primes are used to avoid confusion with the numbering system of the bases; see fig. 9.8.) DNA and RNA both have four bases (two **purines** and two **pyrimidines**) in their nucleotide chains. Both molecules have the purines **adenine** and **guanine** and the pyrimidine **cytosine.** DNA has the pyrimidine **thymine;** RNA has the pyrimidine **uracil.** Thus, three of the nitrogenous bases are found in both DNA and RNA, whereas thymine is unique to DNA, and uracil is unique to RNA.

A nucleotide is formed in the cell when a base attaches to the 1′ carbon of the sugar and a phosphate attaches to

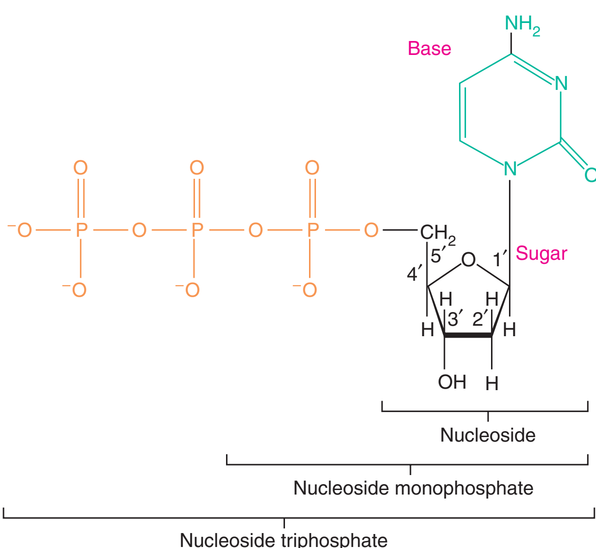
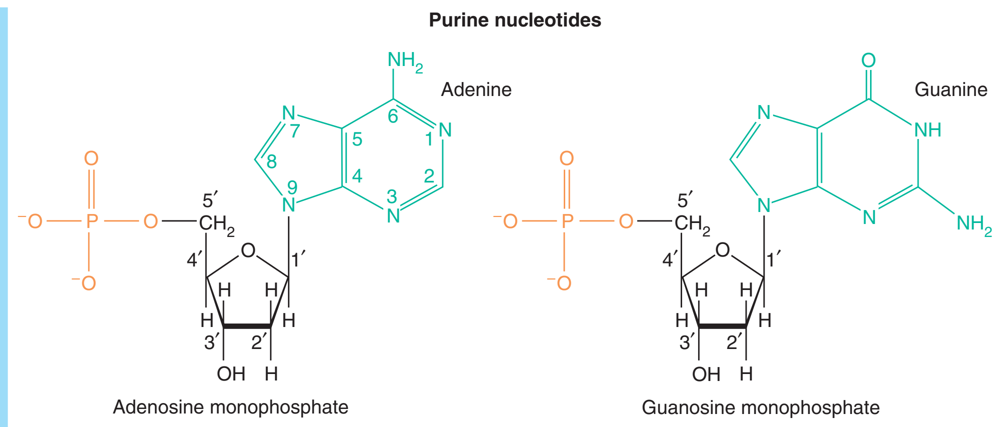


Figure 9.9 The structure of a nucleoside and two nucleotides: a nucleoside monophosphate and a nucleoside triphosphate.

the 5′ carbon of the same sugar (fig. 9.10); the nucleotide takes its name from the base (table 9.2). Nucleotides are linked together **(polymerized)** by the formation of a bond between the phosphate at the 5′ carbon of one nucleotide and the hydroxyl (OH) group at the 3′ carbon of an adjacent molecule. Very long strings of nucleotides can be polymerized by this **phosphodiester bonding** (fig. 9.11).

Biologically Active Structure

Although the identities of the nucleotides that polymerized to form a strand of DNA or RNA were known, the actual structures of these nucleic acids when they function as the genetic material remained unknown until 1953. The general feeling was that the biologically active structure of DNA was more complex than a single string of nucleotides linked together by phosphodiester bonds, and that several interacting strands were involved. In 1953, Linus Pauling, a Nobel laureate who had discovered the



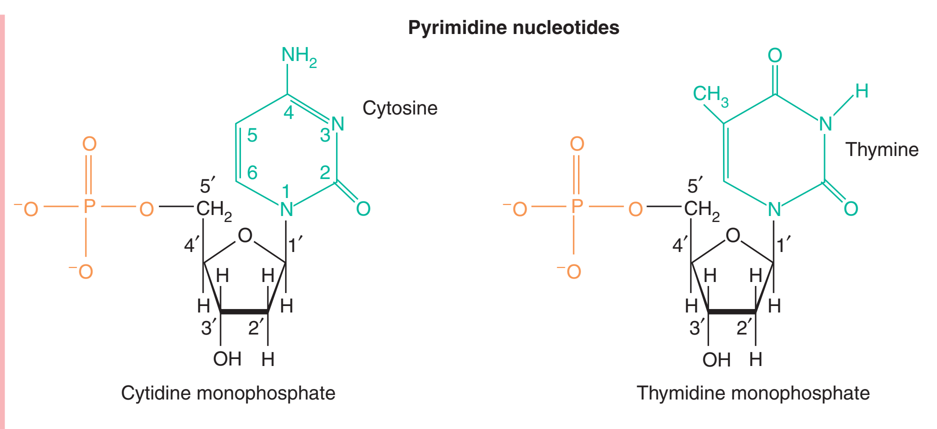


Figure 9.10 Structure of the four deoxyribose nucleotides.

Table 9.2 **Nucleotide Nomenclature**

Base	Nucleotide (nucleoside monophosphate)	Abbreviation					
		Monophosphate		Diphosphate		Triphosphate	
		Ribose	Deoxyribose	Ribose	Deoxyribose	Ribose	Deoxyribose
Guanine	Guanosine monophosphate	GMP		GDP		GTP	
	Deoxyguanosine monophosphate		dGMP		dGDP		dGTP
Adenine	Adenosine monophosphate	AMP		ADP		ATP	
	Deoxyadenosine monophosphate		dAMP		dADP		dATP
Cytosine	Cytidine monophosphate	CMP		CDP		CTP	
	Deoxycytidine monophosphate		dCMP		dCDP		dCTP
Thymine	Deoxythymidine monophosphate		dTMP		dTDP		dTTP
Uracil	Uridine monophosphate	UMP		UDP		UTP	

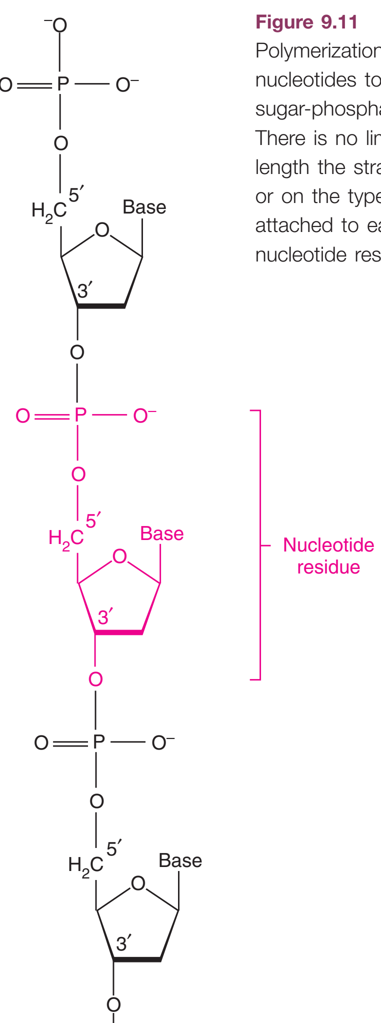


Figure 9.11 Polymerization of adjacent nucleotides to form a sugar-phosphate strand. There is no limit to the length the strand can be or on the type of base attached to each nucleotide residue.

helical structure of proteins, was investigating a three-stranded structure for the genetic material, whereas Watson and Crick had decided that a two-stranded structure was more consistent with available evidence. Three lines of evidence directed Watson and Crick: the chemical nature of the components of DNA, X-ray crystallography, and Chargaff's ratios.

DNA X-Ray Crystallography

All the time Watson and Crick were studying DNA structure, Maurice Wilkins, Rosalind Franklin, and their colleagues were using **X-ray crystallography** to analyze the structure of DNA. The molecules in a crystal are arranged in an orderly way, so that when a beam of X rays is aimed at the crystal, the beam scatteres in an orderly fashion. The scatter pattern can be recorded on photographic film or computer-controlled devices. The nature of this pattern depends on the structure of the crystal. The cross in the center of the photograph in figure 9.12 indicates that the molecule is a helix; the dark areas at the top and bottom come from the bases, stacked perpendicularly to the main axis of

Rosalind E. Franklin (1920–1958). (Courtesy of Cold Spring Harbor Laboratory.)

the molecule. This image of the DNA molecule stimulated Watson and Crick's understanding of its structure.

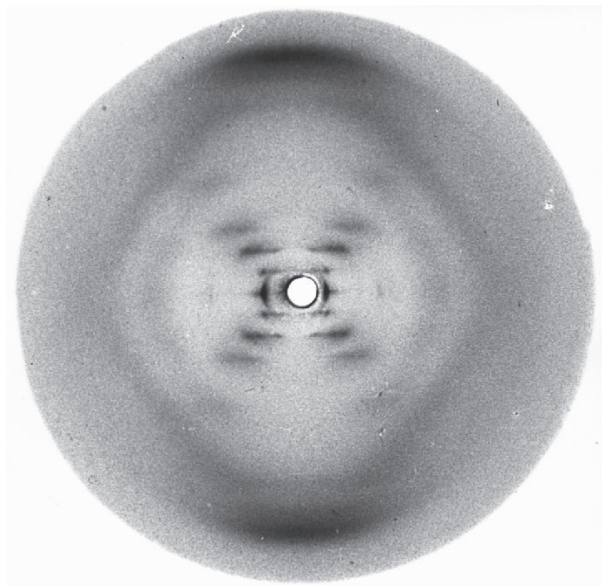


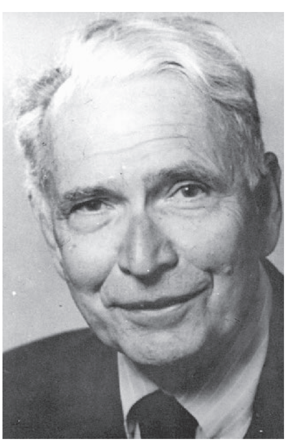
Figure 9.12 Scatter pattern of a beam of X rays passed through crystalline DNA. (Source: Reprinted by permission from R. E. Franklin and R. Gosling, "Molecular configuration in sodium thymonucleate," *Nature* 171:740–41. Copyright 1953 by Macmillan Journals Limited.)

Chargaff's Ratios

Until Erwin Chargaff's work, scientists had labored under the erroneous **tetranucleotide hypothesis.** This hypothesis proposed that DNA was made up of equal quantities of the four bases; therefore, a subunit of this DNA consisted of one copy of each base. Chargaff carefully analyzed the base composition of DNA in various species (table 9.3). He found that although the relative amount of a given nucleotide differs among species, the amount of adenine equaled that of thymine and the amount of guanine equaled that of cytosine. That is, in the DNA of all the organisms studied, a 1:1 correspondence exists between the purine and pyrimidine bases. This is known as **Chargaff's rule.** Chargaff's observations disproved the tetranucleotide hypothesis; the four bases of DNA did not occur in a 1:1:1:1 ratio. His results gave insight to Watson and Crick in the development of their model.

The Watson-Crick Model

With the information available, Watson and Crick began constructing molecular models. They found that a possible structure for DNA was one in which two helices coiled around one another (a **double helix**), with the sugar-phosphate backbones on the outside and the bases on the inside. This structure would fit the dimensions X-ray crystallography had established for DNA if the bases from the two strands were opposite each other and formed "rungs" in a helical "ladder" (fig. 9.13). The diameter of the helix could only be kept constant at about 20 Å (10 angstrom units = 1 nanometer) if one purine and one pyrimidine base made up each rung. Two purines per rung would be too big, and two pyrimidines would be too small.



Erwin Chargaff (1905–). (Courtesy of Dr. Erwin Chargaff.)

After further experimentation with models, Watson and Crick found that the hydrogen bonding necessary to form the rungs of their helical ladder could occur readily between certain base pairs, the pairs that Chargaff found in equal frequencies. (Hydrogen bonds are very weak bonds in which two electronegative atoms, such as O and N, share a hydrogen atom between them. They have 3 to 5% of the strength of a covalent bond.) Thermodynamically stable

Table 9.3 Percentage Base Composition of Some DNAs

Species	Adenine	Thymine	Guanine	Cytosine
Human being (liver)	30.3	30.3	19.5	19.9
Mycobacterium tuberculosis	15.1	14.6	34.9	35.4
Sea urchin	32.8	32.1	17.7	18.4

Source: From E. Chargaff and J. Davidson, *The Nucleic Acids,* Academic Press, 1955.

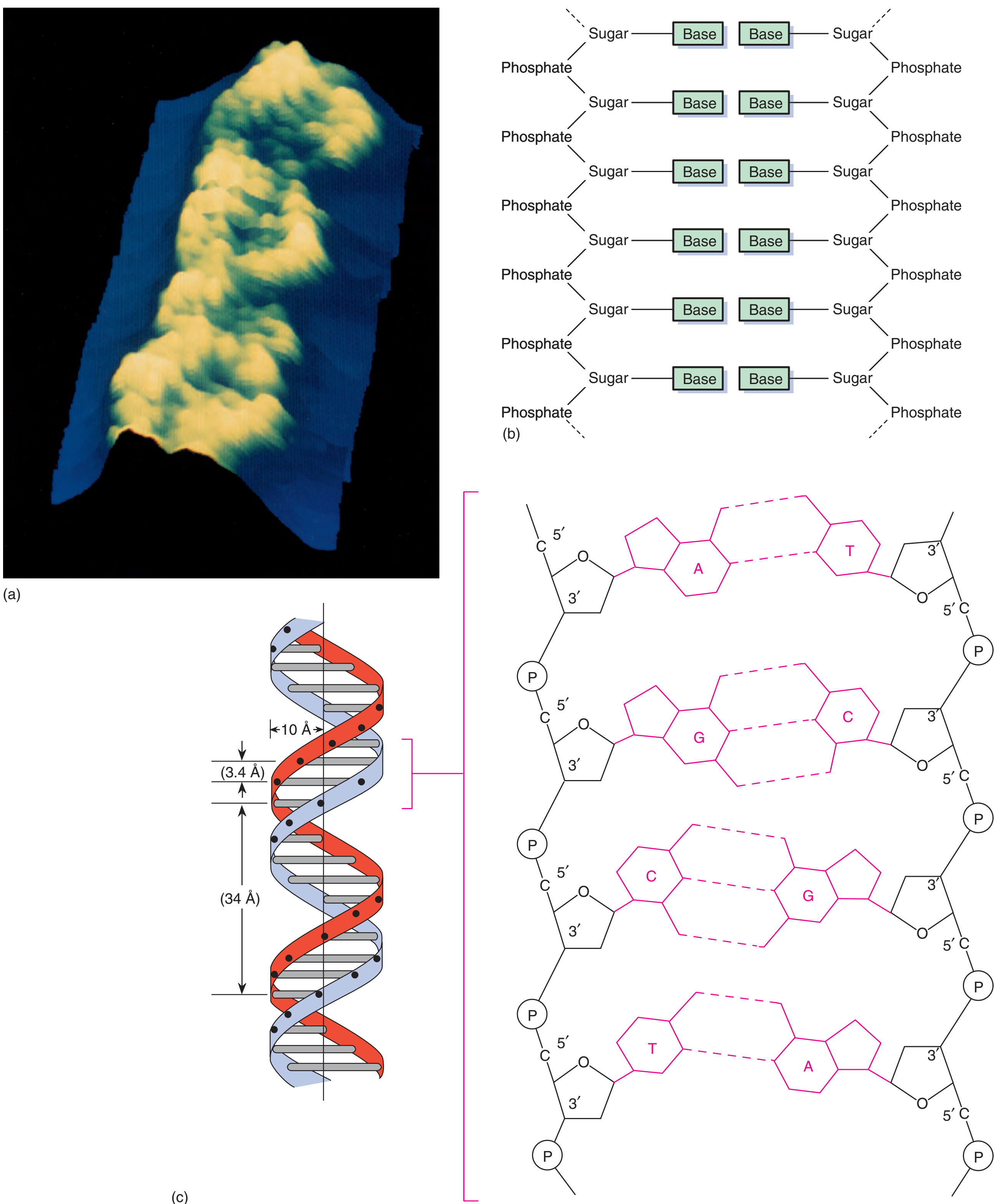


Figure 9.13 Double helical structure of DNA. (*a*) DNA magnified twenty-five million times by scanning tunneling microscopy. (*b*) Component parts. (*c*) Line drawing. ([*a*] © John D. Baldeschweiler.)

© The McGraw–Hill Companies, 2001

hydrogen bonding occurs between thymine and adenine and between cytosine and guanine (fig. 9.14). The relationship is one of **complementarity.** There are two hydrogen bonds between adenine and thymine and three between cytosine and guanine.

Another point about DNA structure relates to the **polarity** that exists in each strand. That is, one end of a DNA strand has a 5′ phosphate and the other end has a 3′ hydroxyl group. Watson and Crick found that hydrogen bonding would occur if the polarity of the two strands ran in opposite directions; that is, if the two strands were **antiparallel** (fig. 9.15).

DNA Denaturation

Denaturation studies indicated that the hydrogen bonding in DNA occurs in the way Watson and Crick suggested. Hydrogen bonds, although individually very weak, give structural stability to a molecule in large enough numbers. However, the hydrogen bonds can be broken and the DNA strands separated when the DNA molecule is heated in water. At a certain point, the thermal agitation overcomes the hydrogen bonding, and the molecule becomes **denatured** (or "melts"). It is logical that the more hydrogen bonds DNA contains the higher the temperature needed to denature it. It thus follows that since a G-C (guanine-cytosine) base pair has three hydrogen bonds to every two in an A-T (adenine-thymine) base pair, the higher the G-C content in a given molecule of DNA, the higher the temperature required to denature it. This relationship exists (fig. 9.16).

Requirements of Genetic Material

Let us now return briefly to the requirements we have said a genetic material needs to meet: (1) control of protein synthesis, (2) self-replication, and (3) location on the chromosomes in the nucleus (in organisms with nuclei). Does DNA (or when DNA is absent, RNA) meet these requirements?

Control of Enzymes

In the next several chapters, we examine the details of protein synthesis. We will see that DNA does possess the complexity required to direct protein synthesis. Although complementarity restricts the base opposite a given base in a double helix, there are no restrictions on the sequence of bases on a given strand. Later, we will show that each sequence of three bases in DNA specifies a particular amino acid during protein synthesis. The ge-

Hydrogen bond
To deoxyribose
To deoxyribose
Adenine
Thymine
To deoxyribose
To deoxyribose
Guanine
Cytosine

Figure 9.14 Hydrogen bonding between the nitrogenous bases in DNA.

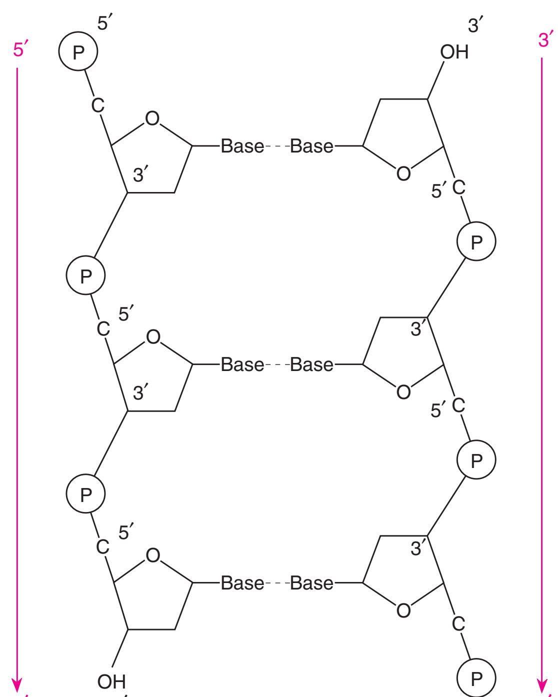


Figure 9.15 Polarity of the DNA strands. Polarity is established by the 3′ and 5′ carbons of a given sugar. For example, moving down the left strand, the polarity is 5′ → 3′ (read as *five-prime to three-prime*). Moving down the right strand, the polarity is 3′ → 5′.

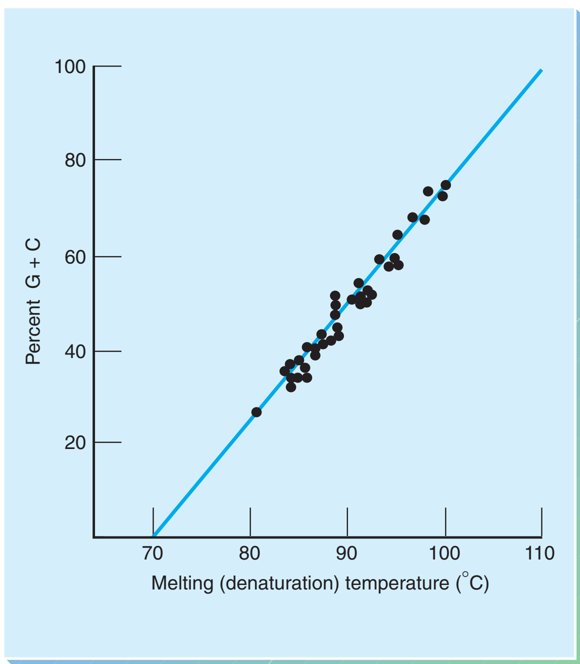


Figure 9.16 Relationship of the number of hydrogen bonds (G-C content) and the thermal stability of DNA from different sources. (From J. Marmur and P. Doty, Jr., "Relationship of the Number of Hydrogen Bonds and the Thermal Bonds and the Thermal Stability of DNA from Different Sources," *Journal of Molecular Biology,* 5:109–112. Copyright © 1962 Academic Press LTD.)

netic code gives the relationship of DNA bases to the amino acids in proteins.

Replication

Watson and Crick hinted in their 1953 paper how DNA might replicate. Their observation stemmed from the property of complementarity. Since the base sequence on one strand is complementary to the base sequence on the opposite strand, each strand could act as a template for a new double helix if the molecule simply "unzipped," allowing each strand to specify the sequence of bases on a new strand by complementarity (fig. 9.17). Mutability would occur due to mispairings, other errors in replication, or damage to the DNA.

Location

DNA must reside in the nucleus of eukaryotes, where the genes occur on chromosomes, or in the chromosomes of prokaryotes and viruses. In both prokaryotes and eukaryotes, the majority of the cell's DNA is in the chromosomes. And all viruses contain either DNA or RNA. Thus, DNA fulfills all the requirements of a genetic material. RNA can fulfill the same requirements in RNA viruses and viroids.

Alternative Forms of DNA

The form of DNA we have described so far is called **B DNA.** It is a right-handed helix: it turns in a clockwise manner when viewed down its axis. The bases are stacked almost exactly perpendicular to the main axis, with about ten base pairs per turn (34 Å; see fig. 9.13*c*). However, DNA can exist in other forms. If the water content increases to about 75%, the A form of DNA **(A DNA)** occurs. In this form, the bases tilt in regard to the axis, and there are more base pairs per turn. However, this and other known forms of DNA are relatively minor variations on the right-handed B form.

In 1979, Alexander Rich and his colleagues at MIT discovered a left-handed helix that they called **Z DNA** because its backbone formed a zigzag structure (fig. 9.18). Z DNA was found by X-ray crystallographic analysis of very small DNA molecules composed of repeating G-C sequences on one strand with the complementary C-G sequences on the other (alternating purines and pyrimidines). Z DNA looks like B DNA with each base rotated 180 degrees, resulting in a zigzag, left-handed structure (fig. 9.19). (The original configuration of the bases is referred to as the *anti* configuration; the rotated configuration is called the *syn* configuration.)

Originally, it was thought that Z DNA would not prove of interest to biologists because it required very high salt concentrations to become stable. However, it was found that Z DNA can be stabilized in physiologically normal conditions if methyl groups are added to the cytosines. Z DNA may be involved in regulating gene expression in eukaryotes. We return to this topic in chapter 16 (box 9.3).

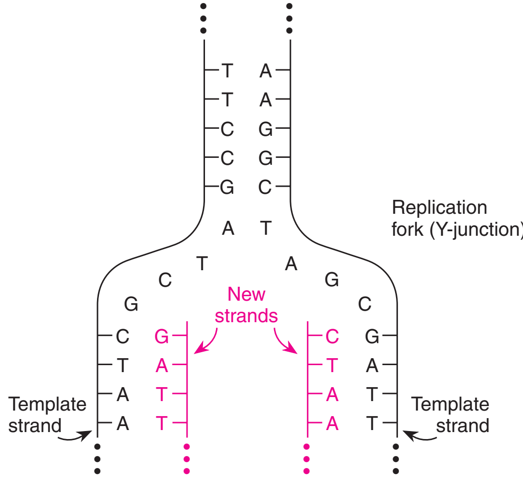


Figure 9.17 Complementarity provides a possible mechanism for accurate DNA replication. The parent duplex opens, and each strand becomes a template for a new duplex.

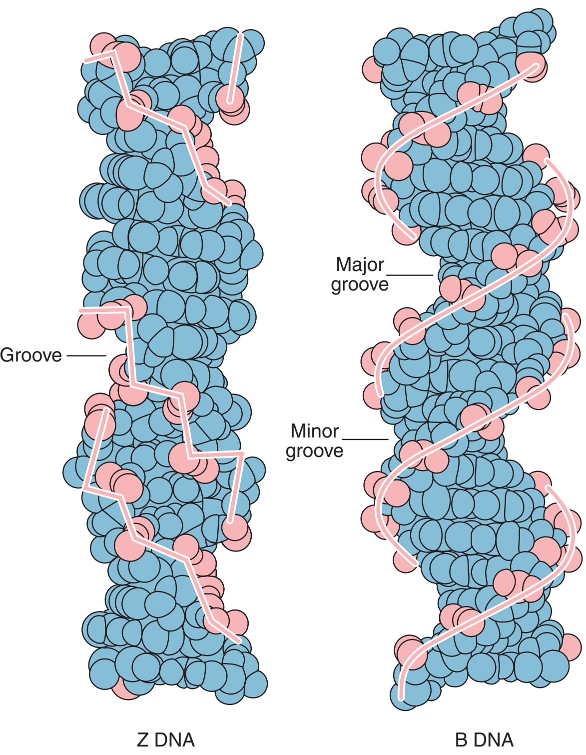


Figure 9.18 Z (*left*) and B (*right*) DNA. The *lines* connect phosphate groups. (Reproduced with permission from the *Annual Review of Biochemistry,* Volume 53, © 1984 by Annual Reviews, Inc.)

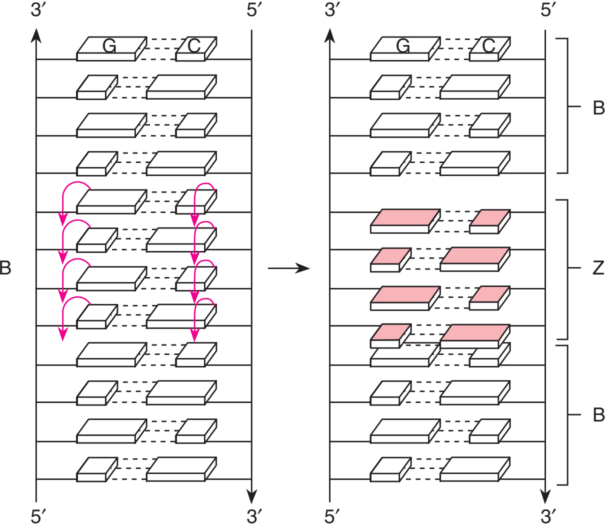


Figure 9.19 B DNA converts to Z DNA by the rotation of bases as indicated by the curved arrows. (Reproduced with permission from *Annual Review of Biochemistry,* Volume 53, © 1984 by Annual Reviews, Inc.)

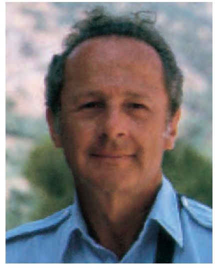
DNA REPLICATION— THE PROCESS

In their 1953 paper, Watson and Crick hinted that the replication of the double helix could take place as the DNA unwinds, so that each strand would form a new double helix by acting as a **template** for a newly synthesized strand (see fig. 9.17). For example, when a double helix is unwound at an adenine-thymine (A-T) base pair, one unwound strand would carry A and the other would carry T. During replication, the A in the template DNA would pair with a T in a newly replicated DNA strand, giving rise to another A-T base pair. Similarly, the T in the other template strand would pair with an A in the other newly replicated strand, giving rise to another A-T base pair. Thus, one A-T base pair in one double helix would result in two A-T base pairs in two double helices. This process would repeat at every base pair in the double helix of the DNA molecule.

This mechanism is called **semiconservative** replication because, although the entire double helix is not conserved in replication, each strand is. Every daughter DNA molecule has an intact template strand and a newly replicated strand. This is not the only way that replication could occur. The alternative methods are **conservative** and **dispersive.** In conservative replication, in which the whole original double helix acts as a template for a new one, one daughter molecule would consist of the original parental DNA, and the other daughter would be totally new DNA. In dispersive replication, some parts of the original double helix are conserved, and some parts are not. Daughter molecules would consist of part template and part newly synthesized DNA. In reality, the dispersive category is the all-inclusive "other" category, including any possibility other than conservative and semiconservative replication.

The Meselson and Stahl Experiment

In 1958, M. Meselson and F. Stahl reported the results of an experiment designed to determine the mode of DNA



Matthew Meselson (1930–). (Courtesy of Dr. Matthew Meselson. Photograph by Bud Gruce.)

Franklin W. Stahl (1929–). (Courtesy of Dr. Franklin W. Stahl.)

BOX 9.3

Biomedical Applications

Multiple-Stranded DNA

Under natural conditions, single-stranded RNA and double-stranded DNA are the rule. However, under laboratory conditions, it is possible to induce a third strand of DNA to interdigitate itself into the major groove of the double helix of normal DNA in a sequence-specific fashion. That is, the third strand of DNA will not just interdigitate anywhere, but will form a stable triplex at a specific sequence (fig. 1). The rules of binding are a little less precise than normal; not all sequences are recognized, and recognition can depend on surrounding sequences. However, a thymine in the third strand will usually recognize an adenine in an adenine-thymine base pair (T•A-T), and a cytosine in the third strand will recognize a guanine in a guanine-cytosine base pair (C+G-C).

Triple-stranded nucleotide chains were first created in 1957 by three scientists at the National Institutes of Health—Alexander Rich, David Davies, and Gary Felsenfeld—while they were creating artificial nucleic acids. At the time, triple-stranded DNA seemed like a laboratory curiosity. Now it seems of interest because it may have valuable uses both experimentally and clinically. (Rich apparently had the same experience in his codiscovery of Z DNA, which at first seemed like an oddity but now is the focus of some attention—see chapter 16.) Now, researchers are able to form triplexes in naturally occurring DNA. Two applications of this technology are actively being pursued.

Both applications arise because a single strand of DNA is capable of recognizing a relatively long sequence of the double-stranded DNA in a chromosome. Thus, it is possible to selectively locate a particular genic sequence. Once the third strand locates a particular sequence on a chromosome, two things can happen. First, triplex DNA formation can prevent a particular gene from expressing itself. By the same technique, triplex DNA can also be an abortifacient, a safe method for preventing the implantation of a fetus by preventing the expression of genes under the control of the hormone progesterone.

The second use of triplex DNA is to cut DNA at a specific place by adding a cleaving compound to both ends of the third strand of DNA. Once the third strand has interdigitated, it can then break the original double helix. For example, S. Strobel and P. Dervan at the California Institute of Technology have used a chemical complex containing iron attached to both 3′ and 5′ ends of the third strand of DNA. The addition of a third chemical then initiates the cleavage reaction. The cleavage of the original duplex can be of value in modern recombinant DNA technology (see chapter 13). Whether triplex DNA will ever be of value is not certain at this time. However, it seems to have good potential for therapeutic use and to help in studying and mapping the human genome.

More recently, four-stranded DNA molecules have been found, in which double helices of certain sequences interdigitate to form four-stranded structures. These may be of importance in the formation of crossover sites or in the structures at the ends of eukaryotic chromosomes (see chapter 15).

continued

replication. Some historians and philosophers of science consider this the most elegant scientific experiment ever designed. Meselson and Stahl grew *E. coli* in a medium containing a heavy isotope of nitrogen, ^{15}N. (The normal form of nitrogen is ^{14}N.) After growing for several generations on the ^{15}N medium, the DNA of *E. coli* was denser. The researchers determined the density of the strands using a technique known as **density-gradient centrifugation.** In this technique, a cesium chloride (CsCl) solution is spun in an ultracentrifuge at high speed for several hours. Eventually an equilibrium arises between centrifugal force and diffusion, so that a density gradient is established in the tube with an increasing concentration of CsCl from top to bottom. If DNA (or any other substance) is added, it concentrates and forms a band in the tube at the point where its density is the same as that of the CsCl. If several types of DNA with different densities are added, they form several bands. The bands are detectable under ultraviolet light at a wavelength of 260 nm (nanometers), which nucleic acids absorb strongly.

Meselson and Stahl transferred the bacteria with heavy (^{15}N) DNA to a medium containing only ^{14}N. The new DNA, replicated in the ^{14}N medium, was intermediate in density between light (^{14}N) and heavy (^{15}N) DNA, because the replication was semiconservative (fig. 9.20). If replication had been conservative, two bands would have appeared at the first generation of replication—an original ^{15}N DNA and a new ^{14}N double helix. And, throughout the experiment, if the method of replication had been conservative, the original DNA would have continued to show up as a ^{15}N band. (This, of course, did not happen.) If the method of replication had been dispersive, various multiple-banded patterns would have appeared, depending on the degree of dispersiveness. The results figure 9.20 shows are completely consistent with semiconservative replication and only with semiconservative replication.

BOX 9.3 CONTINUED

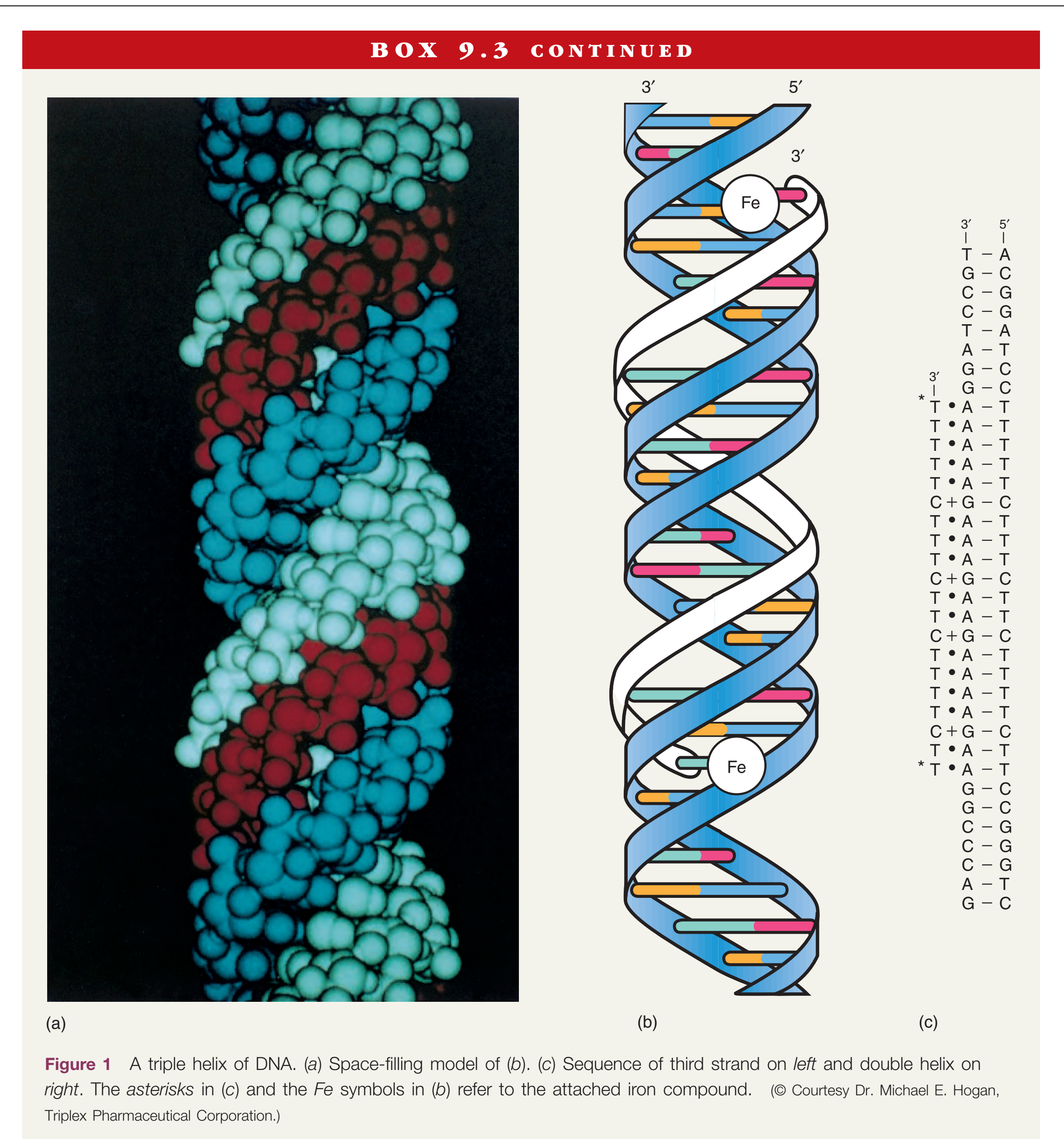


Figure 1 A triple helix of DNA. (*a*) Space-filling model of (*b*). (*c*) Sequence of third strand on *left* and double helix on *right*. The *asterisks* in (*c*) and the *Fe* symbols in (*b*) refer to the attached iron compound. (© Courtesy Dr. Michael E. Hogan, Triplex Pharmaceutical Corporation.)

Autoradiographic Demonstration of DNA Replication

In 1963, J. Cairns used **autoradiography** to verify the semiconservative method of replication photographically. This technique makes use of the fact that radioactive atoms expose photographic film. The visible silver grains on the film can then be counted to provide an estimate of the quantity of radioactive material present. Cairns grew *E. coli* bacteria in a medium containing radioactive thymine, a component of one of the DNA nucleotides. The radioactivity was in tritium (^{3}H). Cairns then carefully extracted the DNA from the bacteria and placed it on photographic emulsion for a period of time. He developed the emulsion to produce autoradiographs

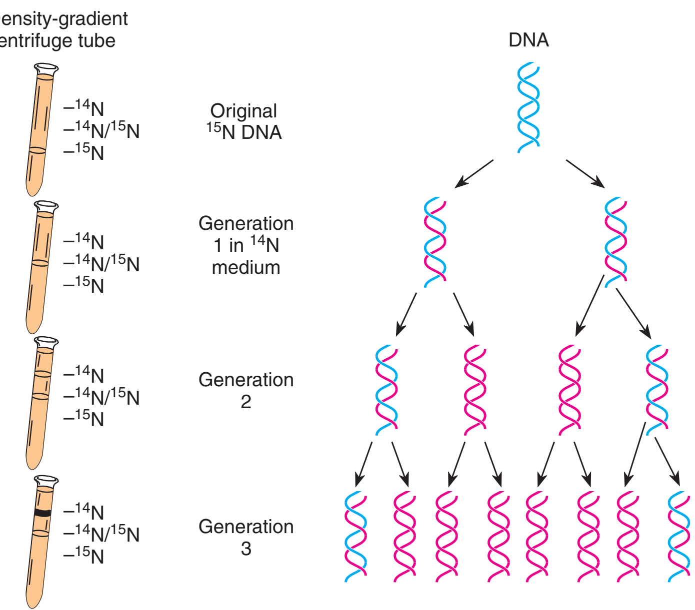
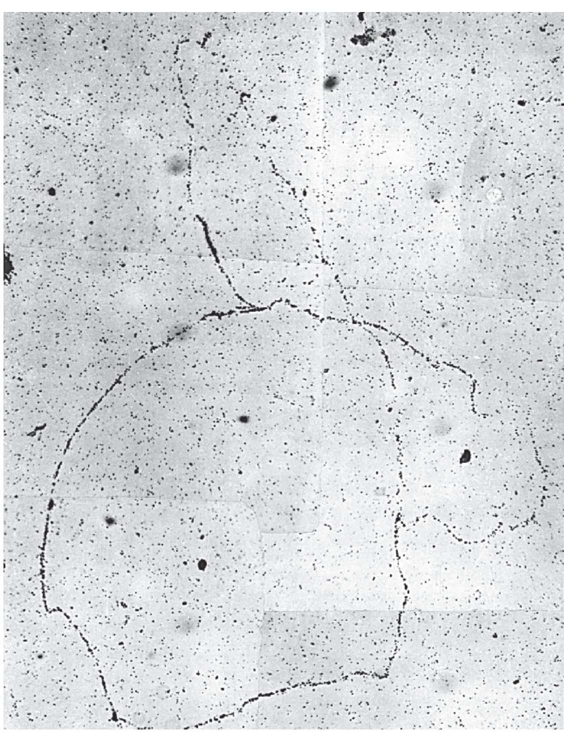
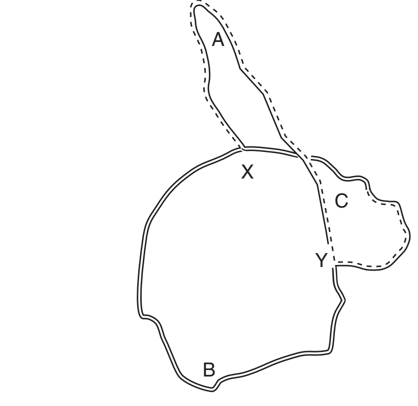


Figure 9.20 The Meselson and Stahl experiment to determine the mode of DNA replication. The bands in the centrifuge tube are visible under ultraviolet light. The pattern of bands (*left*) comes about from semiconservative DNA replication (*right*) of ^{15}N DNA (*blue*) replicating in a ^{14}N medium (*red*).



(a)



(b)

Figure 9.21 (*a*) Autoradiograph of *E. coli* DNA during replication. (*b*) Diagram has labels on the three segments, *A, B,* and *C,* created by the existence of two forks, *X* and *Y,* in the DNA. Forks are created when the circle opens for replication. Length of the chromosome is about 1,300 μm. ([*a*] From J. Cairns, "The chromosome of *E. coli*", *Cold Spring Harbor Symposia on Quantitative Biology,* 28. Copyright © 1963 by Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Reprinted by permission.)

that he then examined under the electron microscope (fig. 9.21). Each grain of silver represents a radioactive decay. Interpretation of this autoradiograph reveals several points. The first, known at the time, is that the *E. coli* DNA is a circle. The second point is that the DNA is replicated while maintaining the integrity of the circle. That is, the circle does not appear to break during the process of DNA replication; an intermediate **theta structure** forms (topologically similar in shape to the Greek letter theta, θ). Third, replication of the DNA seems to be occurring at one or two moving **Y-junctions** in the circle, which further supports the semiconservative mode of replication. The DNA is unwound at a given point, and replication proceeds at a Y-junction, in a semiconservative manner, in one or both directions (see fig. 9.17).

Figure 9.22 diagrams the way in which the two Y-junctions move along the circle to the final step, forming two new circles. The steps by themselves do not support either a unidirectional or a bidirectional mode of replication. That is, a theta structure will develop if either one or both Y-junctions is active in replication. But with autoradiography, it is possible to determine whether new growth is occurring in only one or in both directions.

In some cases, radioactivity was not applied to the cell until DNA replication had already begun. In these cases, the radioactive label appeared after the theta structure

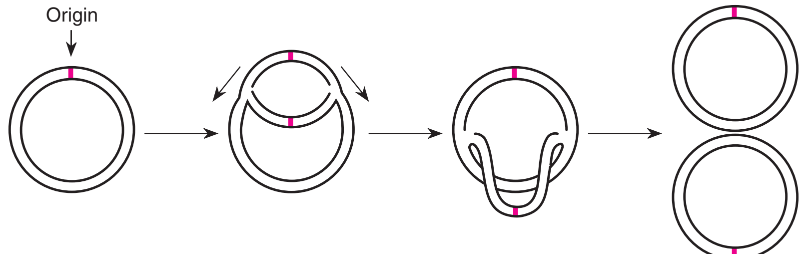


Figure 9.22 Observable stages in the DNA replication of a circular chromosome, assuming bidirectional DNA synthesis. The intermediate figures are called theta structures.

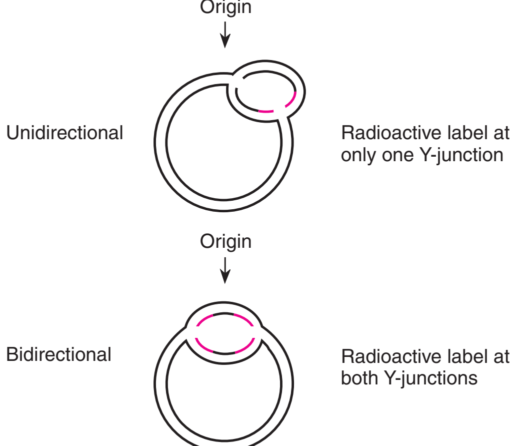


Figure 9.23 Radioactive labels distinguish unidirectional from bidirectional DNA replication. In these hypothetical experiments, DNA replication was allowed to begin, and then a radioactive label was added. After a short period of time, the process was stopped and the autoradiographs prepared. In bidirectional replication (the actual case), the label appears at both Y-junctions.

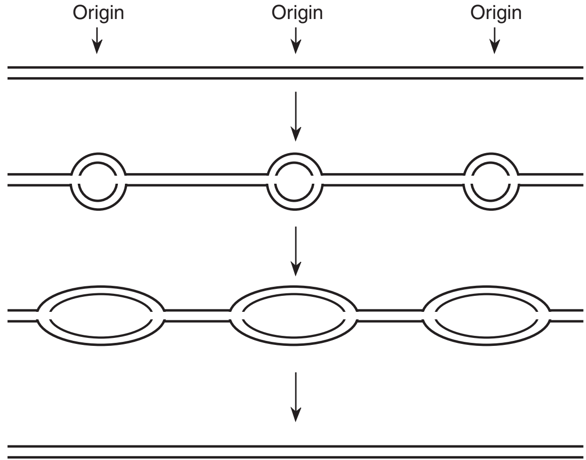


Figure 9.24 Replication bubbles. (*a*) Formation of bubbles (eyes) in eukaryotic DNA because of multiple DNA synthesis sites of origin. (*b*) Electron micrograph (and explanatory line drawing) of replicating *Drosophila* DNA showing these bubbles. ([*b*] H. Kreigstein and D. Hogness, "Mechanism of DNA replication in *Drosophila* chromosomes: Structure of replication forks and evidence for bidirectionality," *Proceeding of the National Academy of Sciences USA,* 71 (1974):135–39. Reproduced by permission.)

had already begun forming. Figure 9.23 illustrates hypothetical outcomes for either unidirectional or bidirectional replication. By counting silver grains in autoradiographs, Cairns found growth to be bidirectional. Both autoradiographic and genetic analysis have subsequently verified this finding.

In eukaryotes, the DNA molecules (chromosomes) are larger than in prokaryotes and are not circular; there are also usually multiple sites for the initiation of replication. Thus, each eukaryotic chromosome is composed of many replicating units, or **replicons**—stretches of DNA with a single origin of replication. In comparison, the *E. coli* chromosome is composed of only one replicon. In eukaryotes, these replicating units form "bubbles" (or "eyes") in the DNA during replication (fig. 9.24).

DNA REPLICATION—THE ENZYMOLOGY

Let us turn now to the details of the processes that take place during DNA replication. Like virtually all metabolic processes, DNA replication is under the control of enzymes. The evidence for the details we describe comes from physical, chemical, and biochemical studies of enzymes and nucleic acids and from the analysis of mutations that influence the replication processes. More recent techniques of recombinant DNA technology and nucleotide sequencing have allowed us to determine the nucleotide sequences of many of these key regions in DNA and RNA. We will look first at *E. coli.*

There are three major enzymes that will polymerize nucleotides into a growing strand of DNA in *E. coli.* These enzymes are **DNA polymerase** I, II, and III. DNA polymerase I, discovered by Arthur Kornberg, who subsequently won the Nobel Prize for his work, is primarily utilized in filling in small DNA segments during replication and repair processes. DNA polymerase II can serve as an alternative repair polymerase; it can also replicate DNA if the template is damaged. DNA polymerase III is the primary polymerase during normal DNA replication.

Arthur Kornberg (1918–). (Courtesy of Dr. Arthur Kornberg. Photograph by Karsh.)

In the simplest model of DNA replication, new nucleotides would be simultaneously added, according to the rules of complementarity, on both strands of newly synthesized DNA at the replication fork as the DNA opens up. But a problem exists, created by DNA's antiparallel nature; the two strands of a DNA double helix run in opposite directions. Going in one direction on the duplex, for example, one strand is a 5′ → 3′ strand, whereas the other is a 3′ → 5′ strand. These directions refer to the numbering of carbon atoms across the sugar. In figure 9.25, going from the bottom of the figure to the top, the left-hand strand is a 3′ → 5′ strand, and the right-hand strand is a 5′ → 3′ strand. Since DNA replication involves the formation of two new antiparallel strands with the old single strands as templates, one new strand would have to be replicated in the 5′ → 3′ direction and the other in the 3′ → 5′ direction.

However, all the known polymerase enzymes add nucleotides in only the 5′ → 3′ direction. That is, the polymerase catalyzes a bond between the first 5′-PO_4 group of a new nucleotide and the 3′-OH carbon of the last nucleotide in the newly synthesized strand (fig. 9.25). The polymerases cannot create the same bond with the 5′ phosphate of a nucleotide already in the DNA and the 3′ end of a new nucleotide. Thus, the simple model needs some revision.

Continuous and Discontinuous DNA Replication

Autoradiographic evidence leads us to believe that replication occurs simultaneously on both strands. **Continuous replication** is, of course, possible on the 3′ → 5′ template strand, which begins with the necessary 3′-OH **primer.** (Primer is double-stranded DNA—or, as we shall see, a DNA-RNA hybrid—continuing as single-stranded DNA template. The strand being synthesized has a 3′-OH available; fig. 9.26.) A **discontinuous** form of replication takes place on the complementary strand, where it occurs in short segments, moving backward, away from the Y-junction (fig. 9.27). These short segments, called **Okazaki fragments** after R. Okazaki, who first saw them, average about 1,500 nucleotides in prokaryotes and 150 in eukaryotes. The strand synthesized continuously is referred to as the **leading strand,** and the strand synthesized discontinuously is referred to as the **lagging strand.**

Once initiated, continuous DNA replication can proceed indefinitely. DNA polymerase III on the leading-strand template has what is called high **processivity:** once it attaches, it doesn't release until the entire strand is replicated. Discontinuous replication, however, requires the repetition of four steps: primer synthesis, elongation, primer removal with gap filling, and ligation.

Primer Synthesis and Elongation

To synthesize Okazaki fragments, a primer must be created *de novo* (Latin: from the beginning). None of the DNA polymerases can create that primer. Instead, **primase,** an RNA polymerase coded for by the *dnaG* gene, creates the primer, ten to twelve nucleotides, at the site

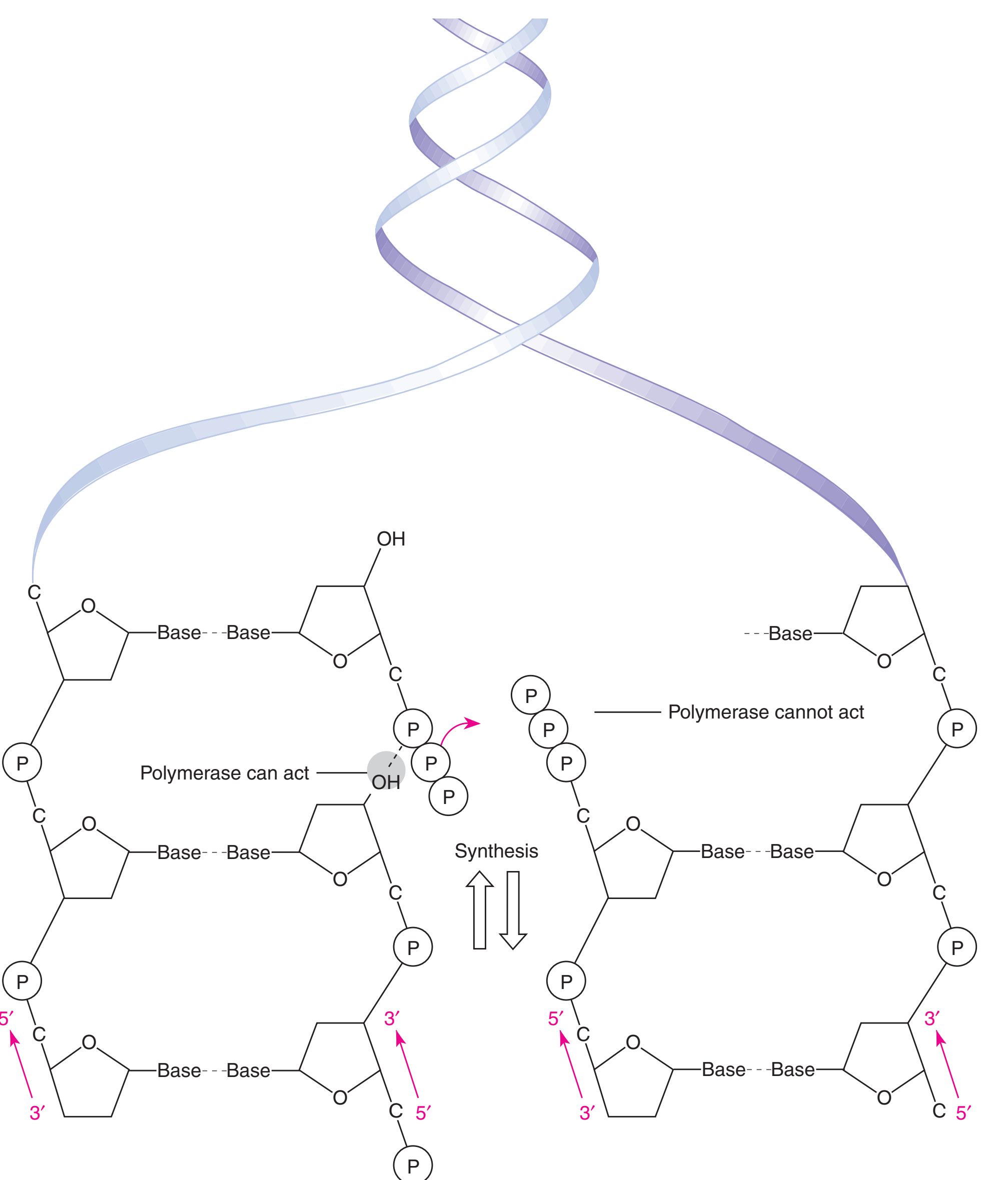


Figure 9.25 New nucleotides can be added to DNA only during replication in the 5′ → 3′ direction.

of Okazaki fragment initiation (fig. 9.28). The result is a short RNA primer that provides the free 3′-OH group that DNA polymerase III needs in order to synthesize the Okazaki fragment. DNA polymerase III continues until it reaches the primer RNA of the previously synthesized Okazaki fragment. At that point, it stops and releases from the DNA.

All three prokaryotic polymerases not only can add new nucleotides to a growing strand in the 5′ → 3′ direction, but also can remove nucleotides in the opposite 3′ → 5′ direction. This property is referred to as *3′ → 5′ exonuclease activity.* Enzymes that degrade nucleic acids are nucleases. They are classified as **exonucleases** if they remove nucleotides from the end of a nucleotide strand or as **endonucleases** if they can break the sugar-phosphate backbone in the middle of a nucleotide strand. At first glance, exonuclease activity seems like an extremely curious property for a polymerase to have—

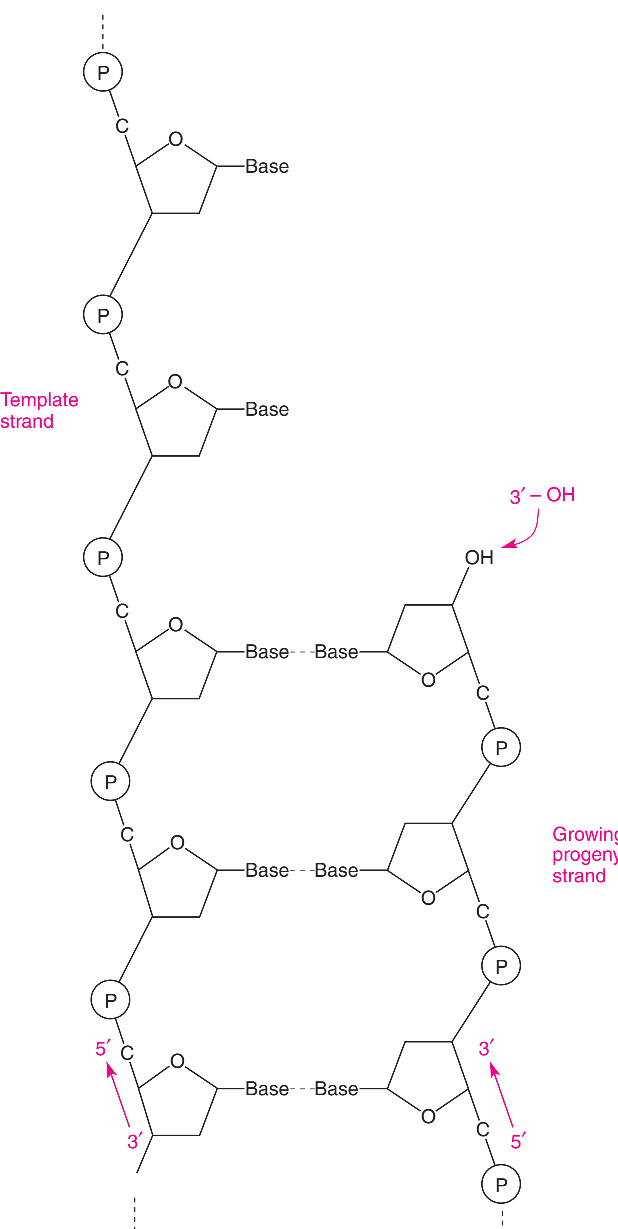


Figure 9.26 Primer configuration for DNA replication. A 3′-OH group must be available on the nascent progeny strand opposite a continuing single-stranded template.

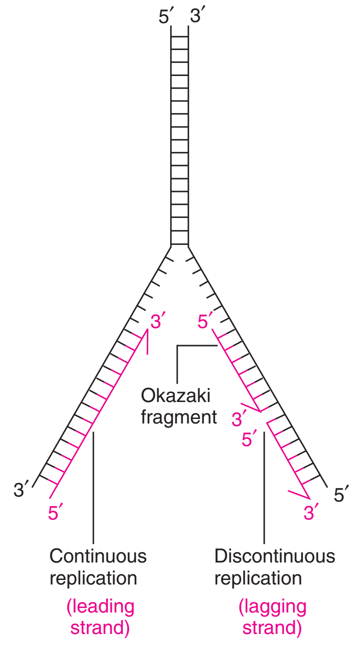


Figure 9.27 Discontinuous model of DNA replication. Lagging-strand replication requires Okazaki fragments to form going backward, away from the Y-junction.

curious unless we think about its ability to check complementarity. If the complementarity is improper, meaning that the wrong nucleotide has been inserted, the polymerase can remove the incorrect nucleotide, put in the proper one, and continue on its way. This is known as the **proofreading** function of DNA polymerase. In addition, exonuclease activity can remove the RNA primers of Okazaki fragments.

Primer Removal with Gap Filling

DNA polymerase I is a polymerase when it adds nucleotides, one at a time, and an exonuclease when it removes nucleotides one at a time. To complete the Okazaki fragment, DNA polymerase I acts in both capacities. (DNA polymerase I mutants cannot properly connect Okazaki fragments.) DNA polymerase I completes the Okazaki fragment by removing the previous RNA primer and replacing it with DNA nucleotides (fig. 9.29). When DNA polymerase I has completed its nuclease and polymerase activity, the two previous Okazaki fragments are almost complete. All that remains is for a single phosphodiester bond to form.

Ligation

DNA polymerase I cannot make the final bond to join two Okazaki fragments. The configuration needing completion is shown in figure 9.30. An enzyme, **DNA ligase,**

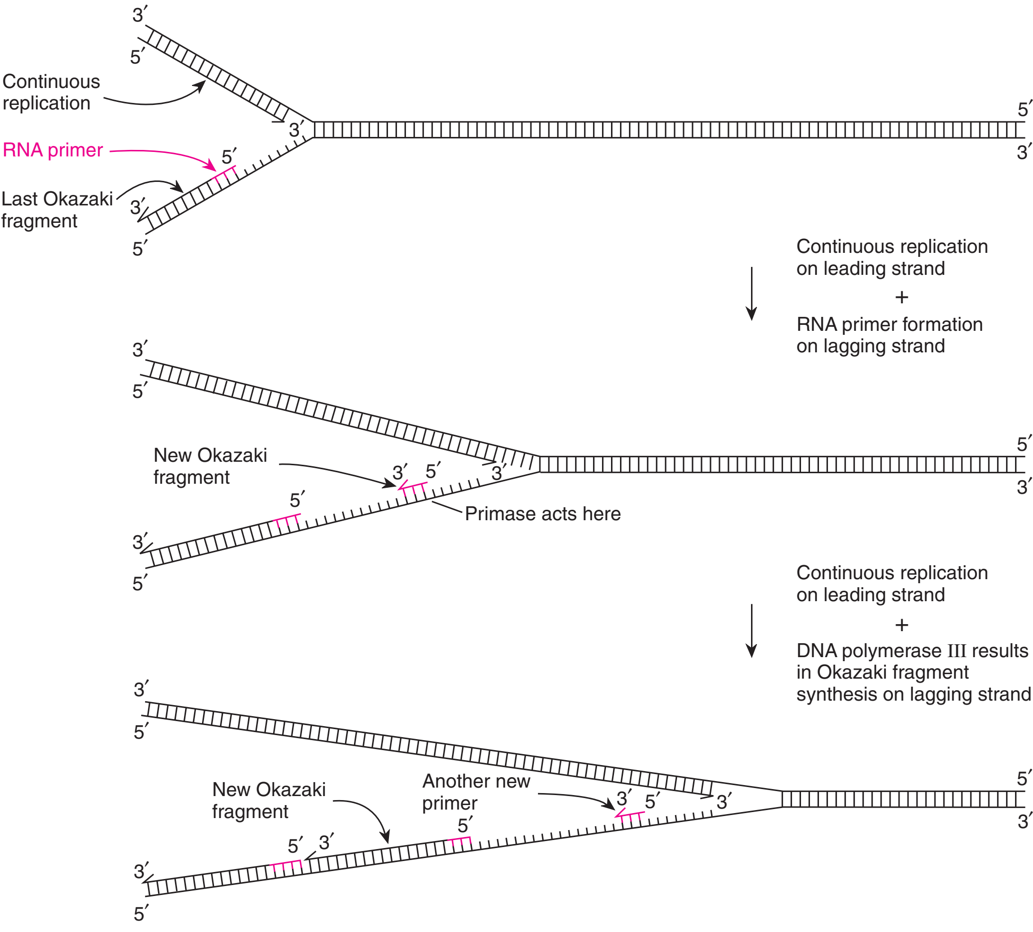


Figure 9.28 Primer formation and elongation create an Okazaki fragment during discontinuous DNA replication.

completes the task by making the final phosphodiester bond in an energy-requiring reaction.

A question of evolutionary interest is why RNA is used to prime DNA synthesis. Why not use DNA directly and avoid the exonuclease and resynthesis activity seen in figure 9.29? Probably, making use of RNA primers lowers the error rate of DNA replication. That is, priming is an inherently error-prone process since nucleotides are initially added without a stable primer configuration. To prevent long-term errors in the DNA, an RNA primer is put in that can later be recognized and removed. Resynthesis by polymerase I is in a much more stable primer configuration (a long primer) and thus makes very few errors.

Another question of evolutionary interest is why DNA synthesis cannot take place in the $3' \rightarrow 5'$ direction. Probably, the answer has to do with proofreading and the exonuclease removal of mismatched nucleotides. When an incorrect nucleotide is found and removed, the next nucleotide brought in, in the $5' \rightarrow 3'$ direction, has a triphosphate end available to provide the energy for its own incorporation (see fig. 9.25). Consider what would happen if the polymerase were capable of adding nucleotides in the opposite direction. The energy for the phosphodiester bond would be coming from the triphosphate already attached in the growing $3' \rightarrow 5'$ strand (see fig. 9.25). Then, if an error in complementarity were detected and the polymerase removed the most recently added nucleotide from the $3' \rightarrow 5'$ strand, the last nucleotide in the double helix would no longer have a triphosphate available to provide energy for the diester bond with the next nu-

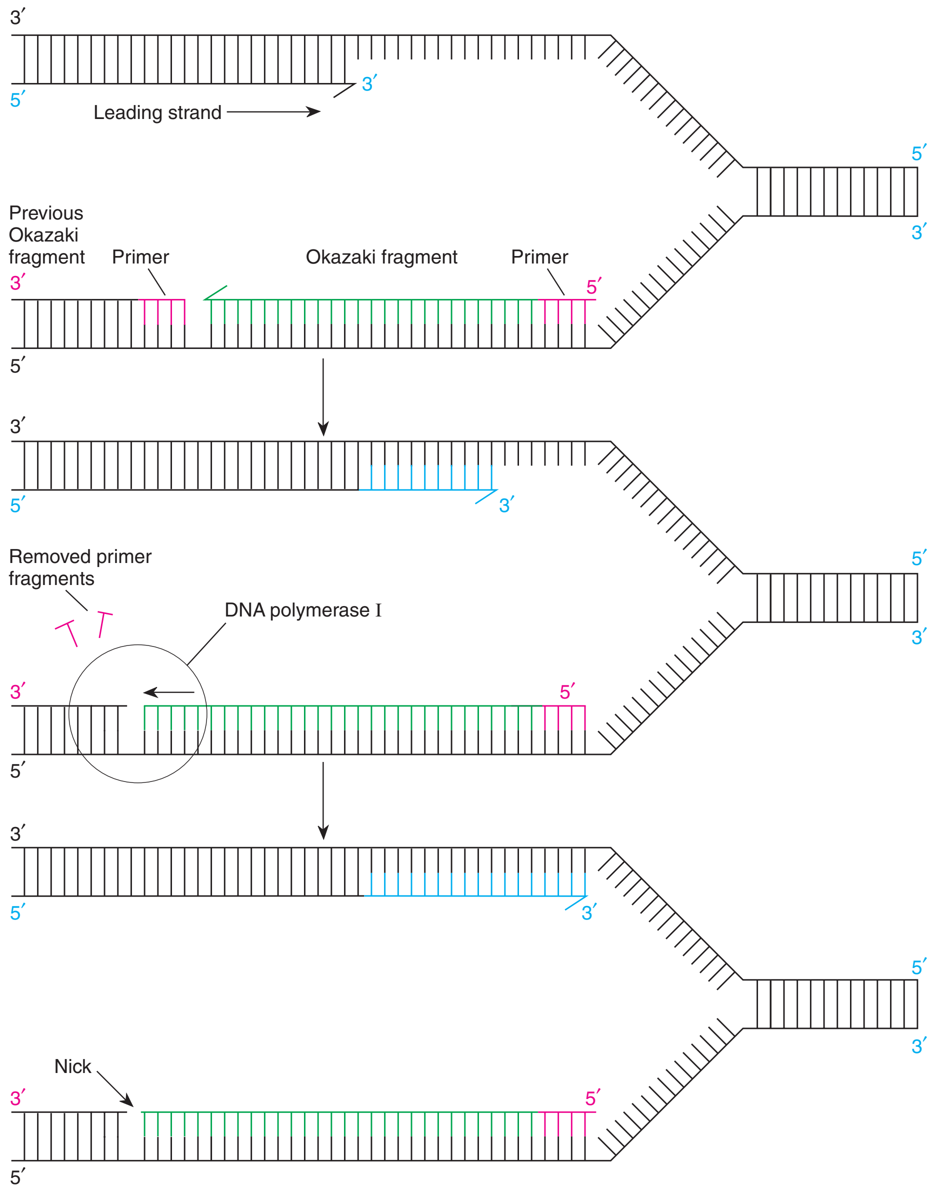


Figure 9.29 The completion of an Okazaki fragment requires that DNA polymerase I replace the RNA primer base by base with DNA nucleotides. A final nick in the DNA backbone remains (*arrow*).

cleotide. Continued polymerization would thus require additional enzymatic steps to provide the energy needed for the process to continue. This could stop or slow the process down considerably. As it is, the process incorporates about four hundred nucleotides per second with an error rate of about one incorrect pairing per 10^9 bases. (Other repair systems further improve this error rate—see chapter 12.)

The Origin of DNA Replication

Each replicon (e.g., the *E. coli* chromosome, or a segment of a eukaryotic chromosome with an origin of replication) must have a region where DNA replication initiates. In *E. coli,* this region is referred to as the genetic locus *oriC;* it occurs at map location 84 minutes (see fig. 7.27). For DNA replication to begin, several steps must occur.

First, the appropriate initiation proteins must recognize the specific origin site. Then the site must be opened and stabilized. And, finally, a replication fork must be initiated in both directions, involving continuous and discontinuous DNA replication. Although most of the proteins involved are known, there are still a few gaps in our knowledge.

OriC, the origin of replication in *E. coli,* is about 245 base pairs long and is recognized by **initiator proteins.** These proteins, the product of the *dnaA* locus, open up the double helix. (Other DNA-binding proteins are also involved here.) The initiator proteins then take part in the attachment of DNA **helicase,** the product of the *dnaB* gene, which unwinds DNA at the Y-junction. Helicase is then responsible for recruiting (binding) the rest of the proteins that form the replication initiation complex. First is primase, which creates RNA primers. Together, the helicase and primase comprise a **primosome,** attached to the lagging-strand template. As the primosomes move along, they create RNA primers that

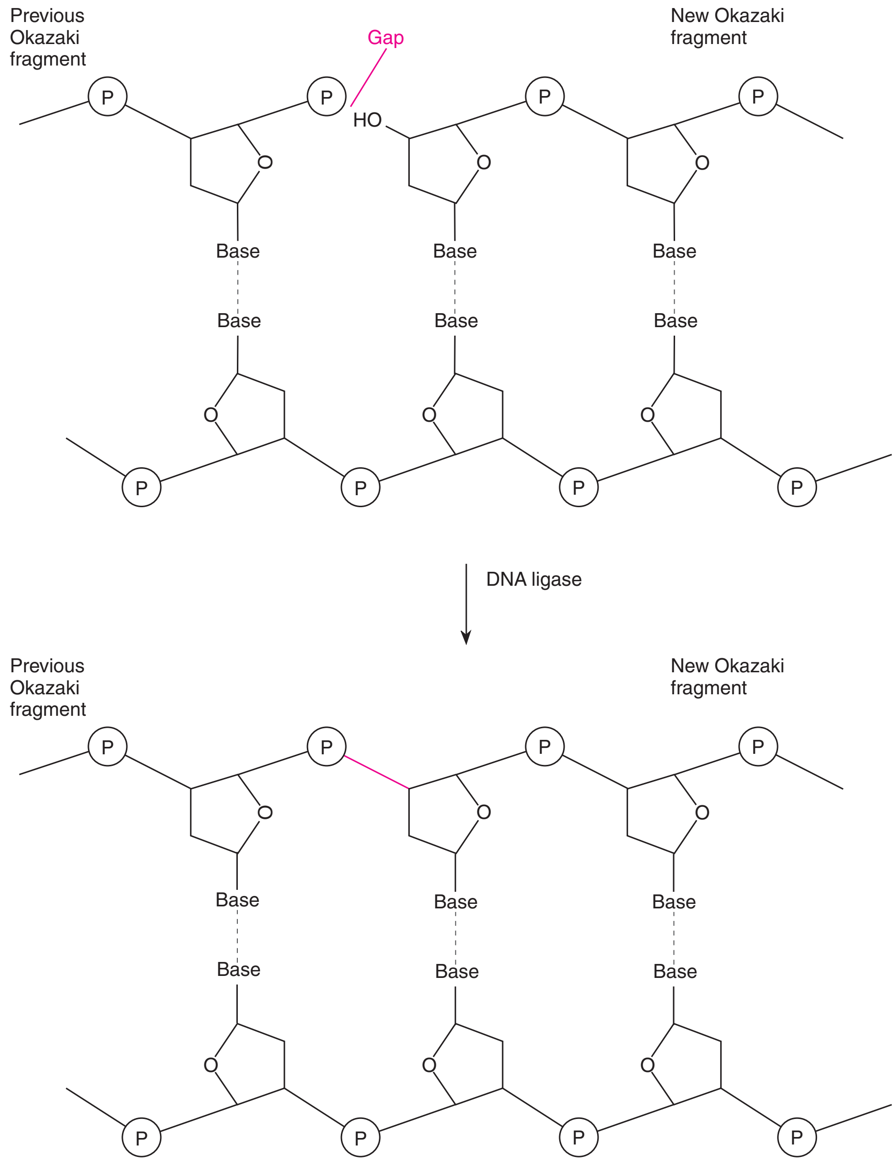


Figure 9.30 After DNA polymerase I removes the RNA primer to complete an Okazaki fragment, a final gap remains. DNA ligase closes it.

DNA polymerase III uses to initiate leading-strand synthesis. As primers are being laid down on the lagging-strand template, Okazaki fragment synthesis begins, and Y-junction activity then proceeds as outlined earlier (see figs. 9.28, 9.29, and 9.30).

DNA polymerase III **holoenzyme** is a very large protein composed of ten subunits (table 9.4). Three of the subunits, α, ε, and θ, form the polymerization core, with both 5′ → 3′ polymerase activity and 3′ → 5′ exonuclease activity. One subunit, the β subunit, is a "processivity clamp." As a dimer (two identical copies attached head to tail), the protein forms a "doughnut" around the DNA so it can move freely on the DNA. When it is attached to the core enzyme, the polymerase is held tightly to the DNA and shows high processivity (fig. 9.31): the leading strand is usually synthesized entirely without the enzyme leaving the template (fig. 9.32). The remaining subunits are involved in processivity control and replisome formation. They allow the polymerase to move off and on the DNA of the lagging-strand template as Okazaki fragments are completed (a process known as **polymerase cycling**).

Eukaryotes have evolved at least nine DNA polymerases, named DNA polymerase α, β, γ, δ, ε, ζ, η, θ, and ι. DNA polymerase δ seems to be the major replicating enzyme in eukaryotes, forming replisomes as in *E. coli.* In eukaryotes, the polymerase α-primase complex adds the Okazaki fragment primers, first adding an RNA primer and then a short length of DNA nucleotides. Polymerase ε may be involved in repair or in normal DNA replication, as is polymerase δ. DNA polymerase γ appears to replicate mitochondrial DNA. The remaining polymerases are probably involved in DNA repair, with polymerase β being the major repair polymerase, as polymerase I is in

Table 9.4 Summary of the Enzymes Involved in DNA Replication in *E. coli*

Enzyme or Protein	Genetic Locus	Function
DNA polymerase I	*polA*	Gap filling and primer removal
DNA polymerase II	*polB*	Replicating damaged templates
DNA polymerase III		
α subunit	*dnaE*	Polymerization core; 5′ → 3′ polymerase
ε subunit	*dnaQ*	Polymerization core; 3′ → 5′ exonuclease
θ subunit	*holE*	Polymerization core
β subunit	*dnaN*	Processivity clamp (as a dimer)
τ subunit	*dnaX*	Preinitiation complex; dimerization of core
γ subunit	*dnaX*	Preinitiation complex; loads clamp
δ subunit	*holA*	Processivity core
δ′ subunit	*holB*	Processivity core
χ subunit	*holC*	Processivity core
ψ subunit	*holD*	Processivity core
Helicase	*dnaB*	Primosome; unwinds DNA
Primase	*dnaG*	Primosome; creates Okazaki fragment primers
Initiator protein	*dnaA*	Binds at origin of replication
DNA ligase	*lig*	Closes Okazaki fragments
Ssb protein	*ssb*	Binds single-stranded DNA
DNA topoisomerase I	*topA*	Relaxes supercoiled DNA
DNA topoisomerase type II		
(DNA Gyrase)		
α subunit	*gyrA*	Relaxes supercoiled DNA; ATPase
β subunit	*gyrB*	Relaxes supercoiled DNA
Topoisomerase IV	*parE*	Unconcatenates DNA circles
Termination protein	*tus*	Binds at termination sites

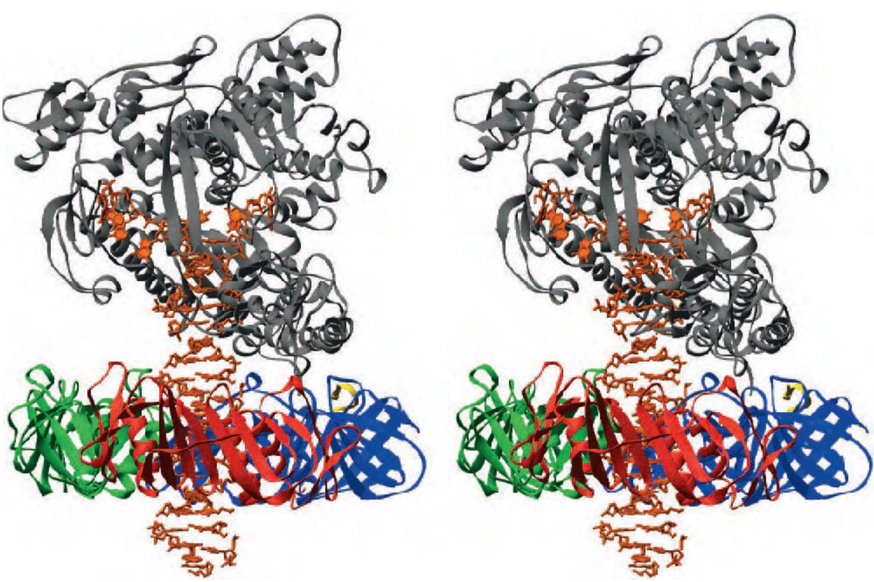


Figure 9.31 Stereo view of sliding clamp, DNA polymerase, and DNA from bacteriophage RB69. The clamp (*red, blue, green*) surrounds the DNA (*brown*) like a doughnut. The clamp is attached to the proximal segment of the DNA polymerase (*gray*). (From Yousif Shamoo and Thomas A. Steitz, "Building a replisome from interacting pieces" in *Cell,* 99:155–166, October 15. Reprinted by permission of *Cell*.)

E. coli. Several of the polymerases most likely both replicate and repair DNA.

Eukaryotes also have a clamp-loader complex, called replication factor C, and a six-unit clamp called the proliferating cell nuclear antigen. The RNA primers are removed during Okazaki fragment completion (maturation) by mechanisms similar to those in prokaryotes. In eukaryotes, RNAase enzymes remove the RNA primers in Okazaki fragments; a repair polymerase fills gaps; and a DNA ligase forms the final seal. Helicases, topoisomerases, and single-strand binding proteins play roles similar to those they play in prokaryotes. The completion of the replication of linear eukaryotic chromosomes involves the formation of specialized structures at the tips of the chromosomes, which we discuss in chapter 15. Thus, all of the enzymatic processes are generally the same in prokaryotes and eukaryotes. DNA replication developed in prokaryotes and was refined as prokaryotes evolved into eukaryotes.

T. Steitz and his colleagues have done much X-ray crystallography work that has given us an excellent look at the structure of a polymerase. (Most work has actually been done on a fragment of DNA polymerase I called the **Klenow fragment.**) The enzyme is shaped like a cupped right hand with enzymatic activity taking place in two places, separated by a distance of about two to three nucleotides (fig. 9.33). It is proposed that when the polymerization site senses a mismatch, the DNA is moved so that the 3′ end enters the exonuclease site, where the incorrect nucleotide residue is then cleaved. Polymerization then continues. There may be a general mode of polymerase action among diverse polymerases.

The replication of the *E. coli* chromosome may be controlled by the methylation state of several sequences within *oriC.* As we discuss in chapter 13, certain enzymes add methyl groups to specific DNA bases, and the presence or absence of these methyl groups can serve as signals to other enzymes.

Events at the Y-Junction

We now have the image of DNA replication proceeding as a primosome moves along the lagging-strand template, opening up the DNA (helicase activity), and creating RNA primers (primase activity) for Okazaki fragments. One DNA polymerase III moves along the leading-strand template, generating the leading strand by continuous DNA replication, whereas a second DNA polymerase III moves backward, away from the Y-junction, creating Okazaki fragments. **Single-strand binding proteins** (ssb proteins) keep single-stranded DNA stabilized (open) during

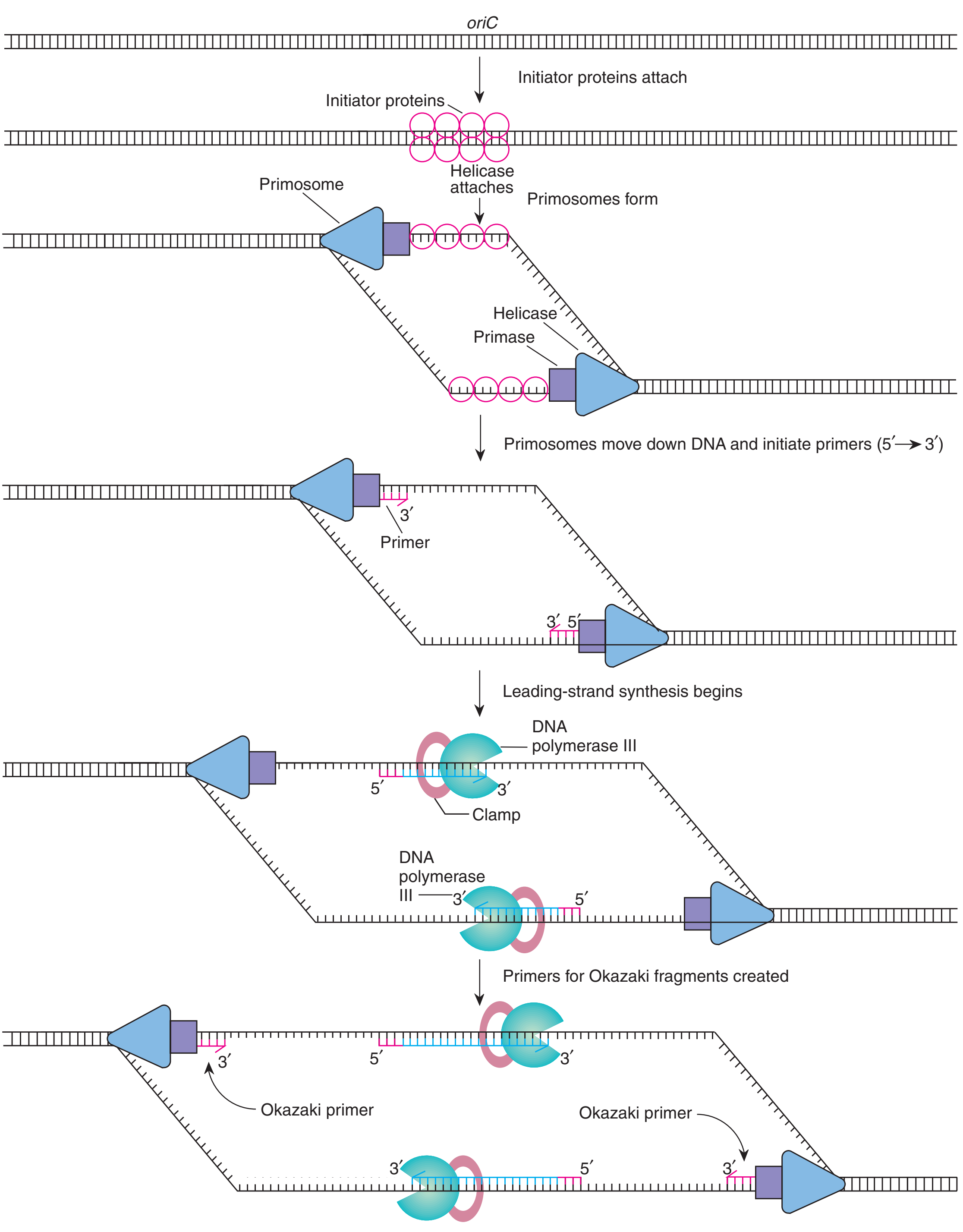


Figure 9.32 Events at the origin of DNA replication in *E. coli.* The DNA opens up at *oriC* to create two moving Y-junctions. Initiator proteins attach and then bind helicase. The helicase then binds primase, forming a primosome. After the primer forms and two copies of DNA polymerase III are bound, the polymerization process begins.

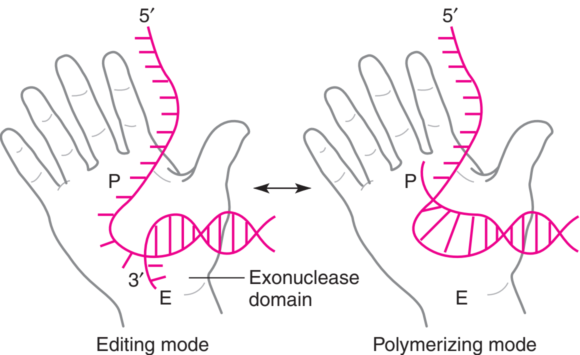


Figure 9.33 DNA polymerase (P) and exonuclease (E) activities of the Klenow fragment of DNA polymerase I in *E. coli.* On the *right,* 5′ → 3′ polymerization is occurring. On the *left,* the 3′ end of the nascent strand has been backed up into the exonuclease site, presumably when a mismatch was detected. (With permission from the *Annual Review of Biochemistry,* Volume 63. ©1994 by Annual Reviews. www.AnnualReviews.org.)

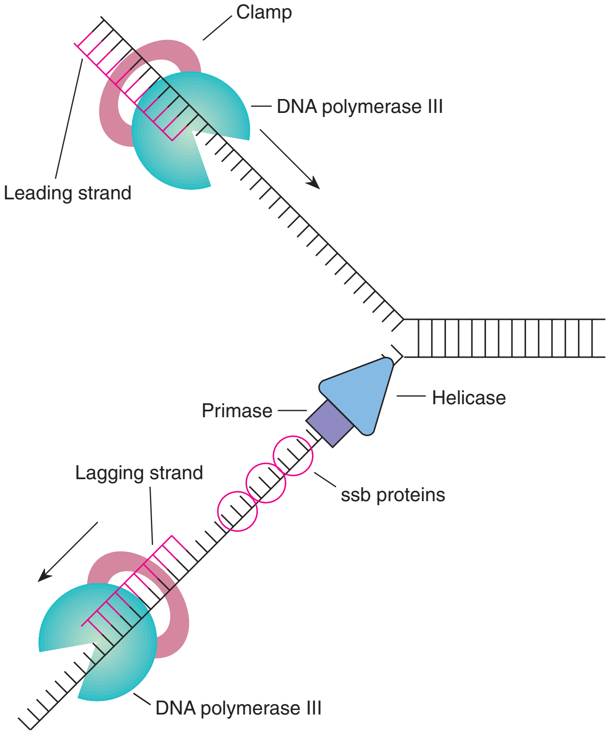


Figure 9.34 Schematic drawing of DNA replication at a Y-junction. Two copies of DNA polymerase III, ssb proteins, and a primosome (helicase + primase) are present.

this process, and DNA polymerase I and ligase connect Okazaki fragments (fig. 9.34).

This simple picture is slightly complicated by the fact that the lagging- and leading-strand synthesis is coordinated. B. Alberts suggested an explanation: the **replisome** model, in which both copies of DNA polymerase III are attached to each other and work in concert with the primosome at the Y-junction (fig. 9.35). According to this model, a single replisome, consisting of two copies of DNA polymerase III, a helicase, and a primase, moves along the DNA. The leading-strand template is immediately fed to a polymerase, whereas the lagging-strand template is not acted on by the polymerase until an RNA primer has been placed on the strand, meaning that a long (fifteen-hundred base) single strand has been opened up (fig. 9.35*a*).

As the replisome moves along, another single-stranded length of the lagging-strand template forms. At about the time that the Okazaki fragment is completed, a new RNA primer has been created (fig. 9.35*b*). The Okazaki fragment is released (fig. 9.35*c*), and a new Okazaki fragment is begun (polymerase cycling), starting with the latest primer (fig. 9.35*d*). This takes the replisome back to the same configuration as in figure 9.35*a,* but one Okazaki fragment farther along.

Figure 9.36 gives us a closer look at the details of the Y-junction at the moment of polymerase cycling. Primase, which is not highly processive, must be in touch with an ssb protein to stay attached to the DNA when forming a primer. At the appropriate moment, after the primer is formed, the clamp loader contacts the ssb, dislodging the primase. The clamp loader also loads a sliding clamp, which then recruits (attaches to) the polymerase that is creating the lagging strand. The polymerase then continues, creating the Okazaki fragment. The primase can later attach at a new point on the lagging-strand template to create the next primer.

Supercoiling

The simplicity and elegance of the DNA molecule masks an inevitable problem: coiling. Since the DNA molecule is made from two strands that wrap about each other, certain operations, such as DNA replication and its termination, face topological difficulties. Up to this point, we have seen the circular *E. coli* chromosome in its “relaxed” state (e.g., figs. 9.21 and 9.22). However, certain enzymes in the cell cause DNA to become overcoiled (positively **supercoiled**) or undercoiled (negatively supercoiled). Positive supercoiling comes about in two ways: either the DNA takes too many turns in a given length, or the molecule wraps around itself (fig. 9.37).

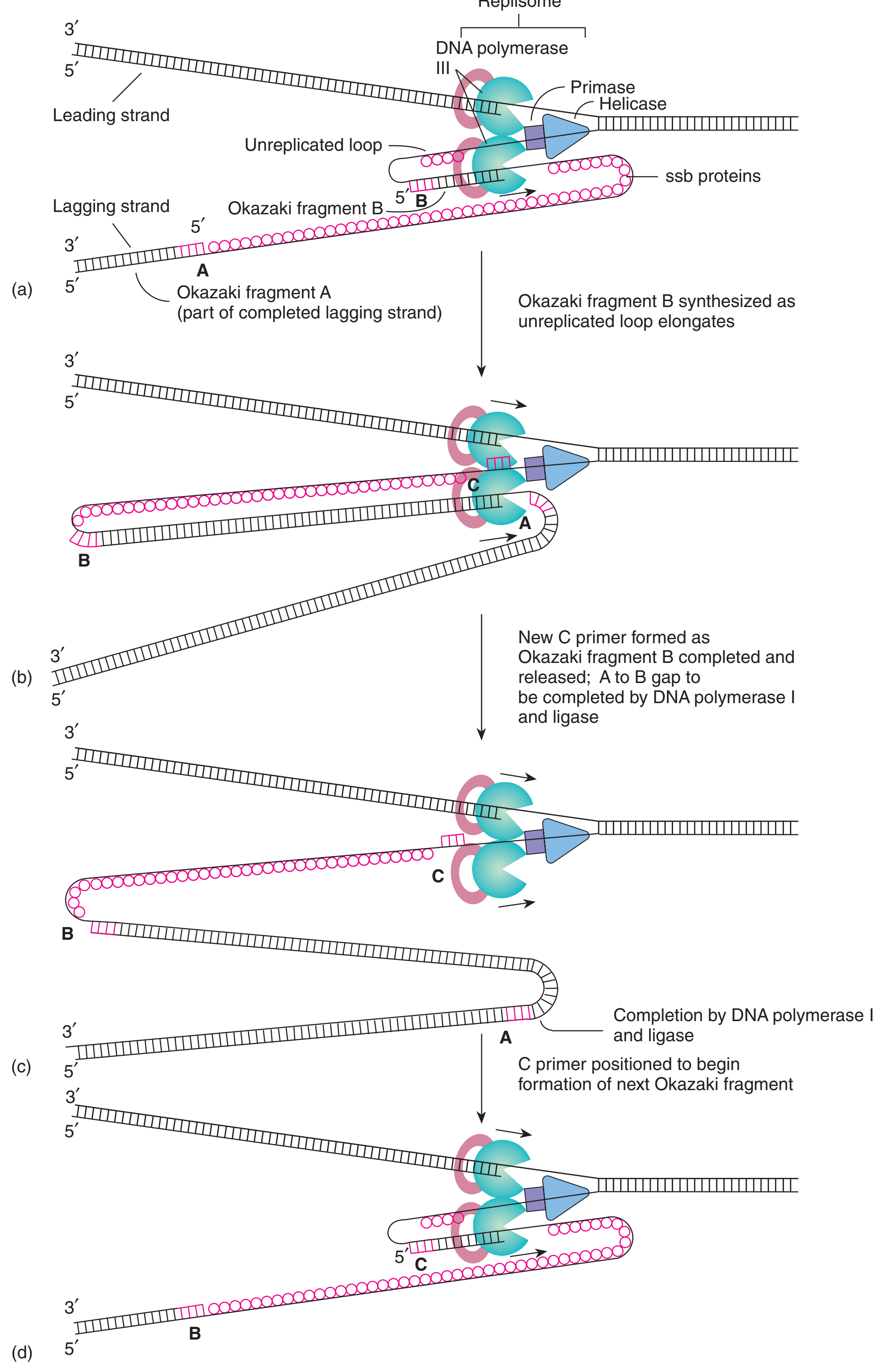


Figure 9.35 The replisome, which consists of two DNA polymerase III holoenzymes and a primosome (helicase + primase), coordinates replication at the Y-junction. Parts *b–d* show "polymerase cycling," in which the polymerase on the lagging-strand template releases a completed Okazaki fragment and then begins the next one.

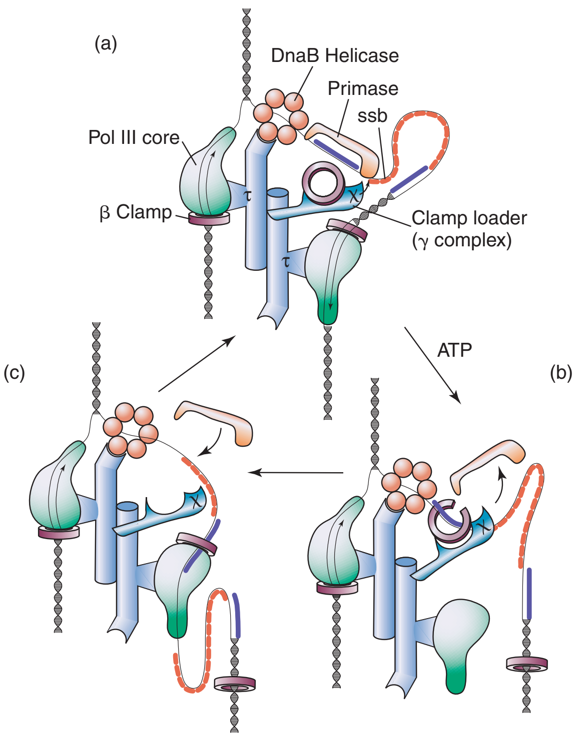


Figure 9.36 A close-up view of the Y-junction during polymerase cycling. The two polymerases (pol III core) are held together by τ subunits. Also pictured are the sliding clamp (β Clamp), clamp loader (γ complex), primase, helicase, and ssb proteins. In (*a*), the primase has just finished creating a primer. The χ subunit of the clamp loader contacts the ssb protein that is touching the primase; the primase is then dislodged (*b*). The clamp is loaded at the new primer and the polymerase on the lagging strand is cycled to the clamp to begin the next Okazaki segment (*c*). (Reprinted from *Cell,* Vol. 96, Yuzhakov et al., "Trading Places on DNA-a Three Point Switch Underlies Primer Handoff from Primase to the Replicative DNA Polymers," pp. 153–163, Copyright © 1999, with permission from Elsevier Science.)

Positive supercoiling occurs when the circular duplex winds about itself in the same direction as the helix twists (right-handed), whereas negative supercoiling comes about when the duplex winds about itself in the opposite direction as the helix twists (left-handed). The former increases the number of turns of one helix around the other (the **linkage number,** L), whereas the latter decreases it. The three forms of DNA in figure 9.37 all have the same sequence, yet they differ in linkage number. Accordingly, they are referred to as topological isomers **(topoisomers).** The enzymes that create or alleviate these states are called **topoisomerases.**

Topoisomerases affect supercoiling by either of two methods. Type I topoisomerases break one strand of a double helix and, while binding the broken ends, pass the other strand through the break. The break is then sealed (fig. 9.38). Type II topoisomerases (e.g., **DNA gyrase** in *E. coli*) do the same sort of thing, only instead of breaking one strand of a double helix, they break both and pass another double helix through the temporary gap. Four topoisomerases are active in *E. coli,* with somewhat confusing nomenclature: topoisomerases I and III are type I; topoisomerases II and IV are type II.

As DNA replication proceeds, positive supercoiling builds up ahead of the Y-junction. This is eliminated by topoisomerases that either create negative supercoiling ahead of the Y-junction in preparation for replication or alleviate positive supercoiling after it has been created.

Termination of Replication

The termination of the replication of a circular chromosome presents no major topological problems. At the end of the theta-structure replication (see fig. 9.22), both Y-junctions have proceeded around the molecule. The region of termination on the *E. coli* chromosome, the terminus region, is 180 degrees from *oriC* on the circular chromosome, between minutes 28 and 36. There are six terminator sites (*Ter*); three arrest the Y-junction from the left, and three arrest the one from the right when bound by a termination protein, the protein product of the *tus* gene. (*Tus* stands for terminus utilization substance; each *Ter* site is about twenty base pairs.) One interesting aspect of the termination of *E. coli* DNA replication is that the cells are viable even if the whole terminator region is deleted. There are fewer viable cells and some growth problems, but in general, *E. coli* can successfully terminate DNA replication even without formal termination sites. A topoisomerase, topoisomerase IV, then releases the two circles, and DNA polymerase I and ligase close them up (fig. 9.39).

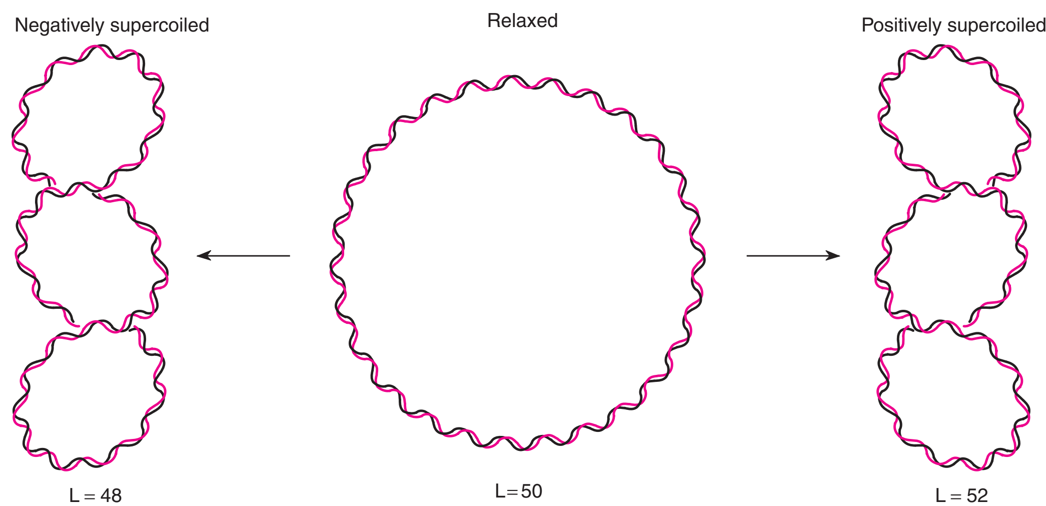


Figure 9.37 Positive and negative supercoils. Enzymes called topoisomerases can take relaxed DNA (*center*) and add negative (*left*) or positive (*right*) supercoils. *L* is the linkage number.

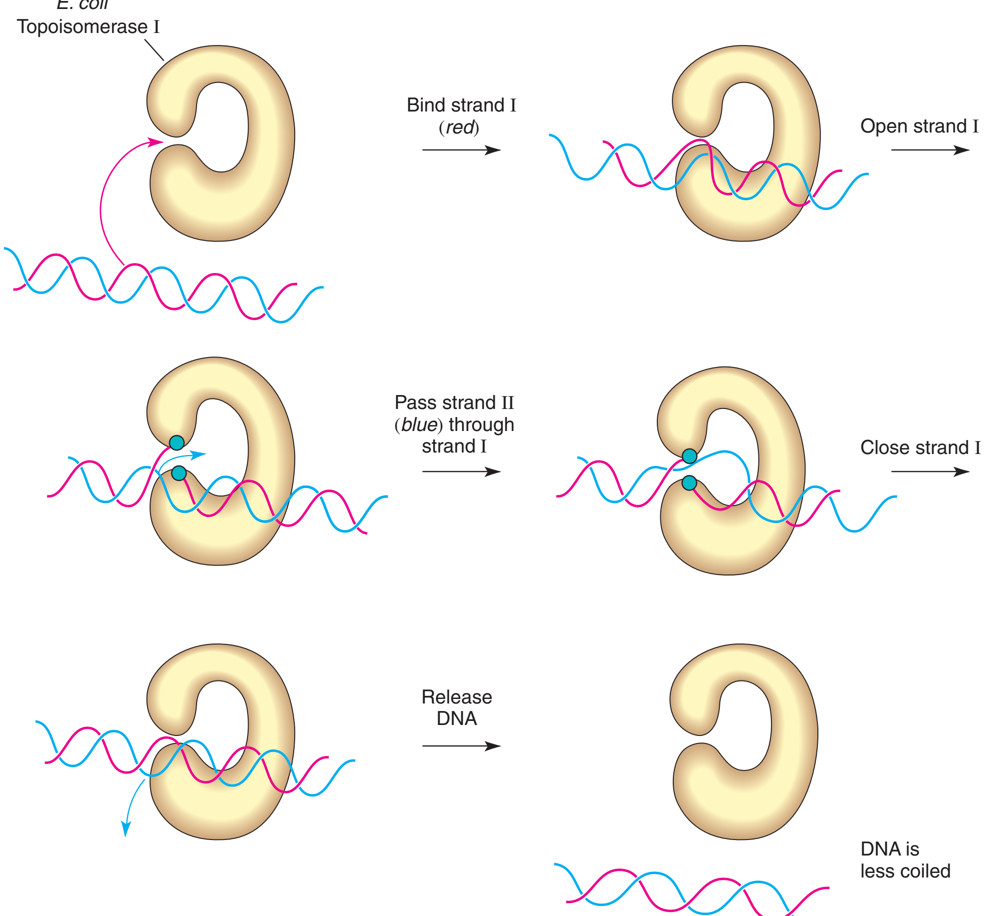


Figure 9.38 Topoisomerase I can reduce DNA coiling by breaking one strand of the double helix and passing the other strand through it.

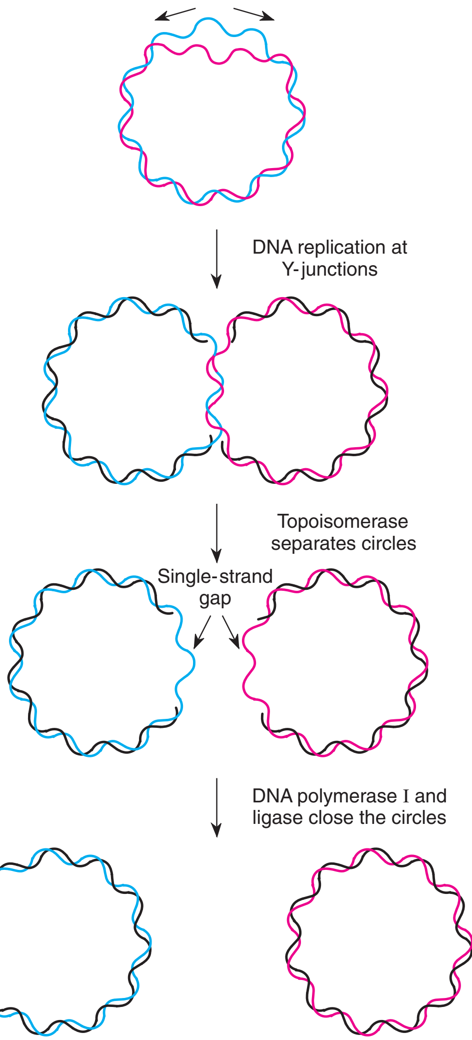


Figure 9.39 The replication of circular DNA terminates when topoisomerase separates the circles and DNA polymerase I and ligase close the gaps in each circle.

DNA Partitioning in E. coli

In chapter 3, we discussed processes that partition eukaryotic chromosomes between daughter cells during mitosis and meiosis. Until very recently, geneticists believed that the partitioning of the *E. coli* chromosome was a passive process, unlike that in eukaryotes. Now, however, we know that more complexity is involved in *E. coli* DNA partitioning. When DNA replication begins, the newly replicated origins of replication are segregated to opposite ends of the bacterial cell, acting as centromeres do. A ring of proteins, the products of the *FtsZ* gene, form a ring at the middle of the cell and begin to create the septum that will divide the cell into two. The full complexity involved in *E. coli* chromosomal partitioning should be uncovered in the near future.

REPLICATION STRUCTURES

The *E. coli* model of DNA replication that we have presented here is by way of the intermediate theta-structure (see fig. 9.22). Two other modes of replication occur in circular chromosomes: rolling-circle and D-loop.

Rolling-Circle Model

In the **rolling-circle** mode of replication, a nick (a break in one of the phosphodiester bonds) is made in one of the strands of the circular DNA, resulting in replication of a circle and a tail (fig. 9.40). This form of replication occurs in the F plasmid or *E. coli* Hfr chromosome during conjugation (see chapter 7). The F^+ or Hfr cell retains the circular daughter while passing the linear tail into the F^- cell, where replication of the tail takes place. Several phages also use this method, filling their heads (protein coats) with linear DNA replicated from a circular parent molecule.

D-Loop Model

Chloroplasts and mitochondria (in eukaryotic cells) have their own circular DNA molecules (see chapter 17) that appear to replicate by a slightly different mechanism. The origin of replication is at a different point on each of the two parental template strands. Replication begins on one strand, displacing the other while forming a displacement loop or **D-loop** structure (fig. 9.41). Replication continues until the process passes the origin of replication on the other strand. Replication then initiates on the second strand, in the opposite direction. Normal Y-junction replication, as described earlier, also occurs in mitochondrial DNA under some growth conditions.

EUKARYOTIC DNA REPLICATION

As we saw earlier, linear eukaryotic chromosomes usually have multiple origins of replication, resulting in figures referred to as "bubbles" or "eyes" (see fig. 9.24). Multiple origins allow eukaryotes to replicate their larger quantities of DNA in a relatively short time, even though eukaryotic DNA replication is considerably slowed by the presence of histone proteins associated

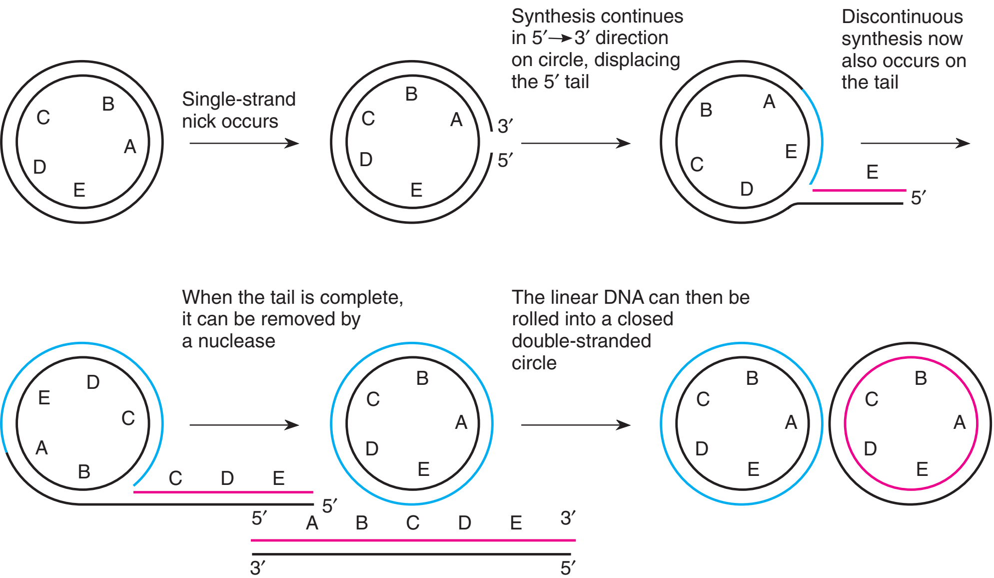


Figure 9.40 Rolling-circle model of DNA replication. The letters *A–E* provide landmarks on the chromosomes.

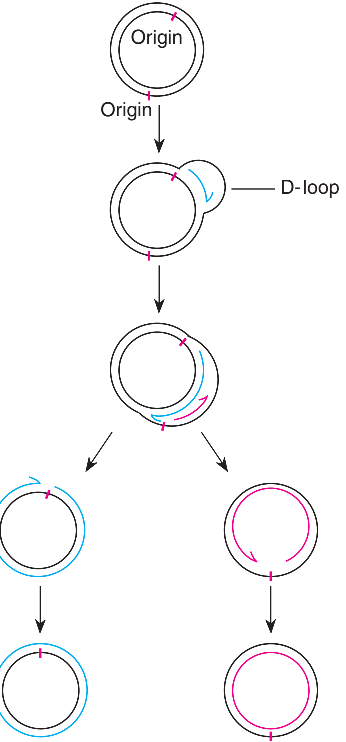


Figure 9.41 D-loops form during mitochondrial and chloroplast DNA replication because the origins of replication are at different places on the two strands of the double helix.

with the DNA to form chromatin (see chapter 15). For example, the *E. coli* replication fork moves through about twenty-five thousand base pairs per minute, whereas the eukaryotic Y-junction moves through only about two thousand base pairs per minute. The number of replicons in eukaryotes varies from about five hundred in yeast to as many as sixty thousand in a diploid mammalian cell.

In budding yeast, a lower eukaryote that is often used as a model organism, DNA replication initiates at sites called **autonomously replicating sequences (ARS).** Each consists of a specific 11-base-pair sequence plus two or three additional short DNA sequences encompassing 100–200 base pairs. Six proteins form a complex that binds to this sequence, referred to as the **origin recognition complex (ORC).** These proteins seem to be bound all the time, and thus additional proteins are needed to initiate DNA replication. Some of these additional proteins are cyclin-dependent kinases, proteins involved in the control of the cell cycle (chapter 3). This makes sense because in eukaryotes, DNA replication can take place only once during the cell cycle, during the S phase. Thus, the initiation of DNA replication must be tightly controlled to avoid multiple replication of some or all replicons.

SUMMARY

STUDY OBJECTIVE 1: To understand the properties that a genetic material must have 205–211.

A genetic material must be able to control the phenotype of a cell or organism (i.e., to direct protein synthesis), it must be able to replicate, and it must be located in the chromosomes. Avery and his colleagues demonstrated that DNA was the genetic material when they showed that the transforming agent was DNA. Griffith had originally demonstrated transformation of *Streptococcus* bacteria in mice. Hershey and Chase demonstrated that the DNA of bacteriophage T2 entered the bacterial cell. Fraenkel-Conrat demonstrated that in viruses without DNA (RNA viruses), such as tobacco mosaic virus, the RNA acted as the genetic material. Thus, by 1953, the evidence strongly suggested nucleic acids (DNA or, in its absence, RNA) as the genetic material.

STUDY OBJECTIVE 2: To examine the structure of DNA, the genetic material 211–224.

Chargaff showed a 1:1 relationship of adenine (A) to thymine (T) and cytosine (C) to guanine (G) in DNA. Wilkins, Franklin, and their colleagues showed, by X-ray crystallography, that DNA was a helix of specific dimensions. Following these lines of evidence, Watson and Crick in 1953 suggested the double-helical model of the structure of DNA. In their model, DNA is made up of two strands, running in opposite directions, with sugar-phosphate backbones and bases facing inward. Bases from the two strands form hydrogen bonds with each other with the restriction that only A and T or G and C can pair. This explains the quantitative relationships that Chargaff found among the bases. Melting temperatures of DNA also support this structural hypothesis because DNAs with higher G-C contents have higher melting, or denaturation, temperatures; G-C base pairs have three hydrogen bonds versus only two in an A-T base pair. The Watson-Crick DNA model represents the B form. DNA can exist in other forms, including the Z form, a left-handed double helix that may be important in controlling eukaryotic gene expression.

STUDY OBJECTIVE 3: To investigate the way in which DNA replicates 220–239.

DNA replicates by unwinding of the double helix, with each strand subsequently acting as a template for a new strand. This works because of complementarity—only A-T, T-A, G-C, or C-G base pairs form stable hydrogen bonds within the structural constraints of the model. This model of replication is *semiconservative.* Meselson and Stahl confirmed it in an experiment with heavy nitrogen. Autoradiographs of replicating DNA showed that replication proceeds bidirectionally from a point of origin. Prokaryotic chromosomes are circular, with a single initiation point of replication. Eukaryotic DNA is linear, with multiple initiation points of replication.

DNA polymerase enzymes add nucleotides only in the $5' \rightarrow 3'$ direction. Replication proceeds in small segments, working backward from the Y-junction on the $5' \rightarrow 3'$ template strand. Presumably, the $5' \rightarrow 3'$ restriction has to do with the proofreading DNA polymerases do to correct errors in complementarity. Polymerase III is the active replicating enzyme, and polymerase I is involved in DNA repair. Many other enzymes help create the Okazaki fragments, unwind DNA, and release the DNA from supercoiling. Prokaryotic and eukaryotic systems follow similar steps.

SOLVED PROBLEMS

PROBLEM 1: What evidence led to the idea that DNA was the genetic material?

Answer: Avery and his colleagues (MacLeod and McCarty) performed experiments showing that DNA was the transforming agent, and they are thus generally given credit for formalizing the notion that DNA, not protein, is the genetic material. Chargaff, Hershey and Chase, Fraenkel-Conrat, and several others also helped shape the general view. At the time that Watson and Crick published their model, the scientific community knew that DNA was the genetic material but didn't know its structure.

PROBLEM 2: How does DNA fulfill the requirements of a genetic material?

Answer: DNA is located in chromosomes, has a structure that is easily and accurately replicated, and has the sequence complexity to code for the fifty thousand or more genes that a eukaryotic organism has.

PROBLEM 3: What enzymes are involved in DNA replication in *E. coli?*

Answer: A replisome, consisting of a primosome (a primase and a helicase) and two polymerase III holoenzymes, forms at a Y-junction on DNA. One polymerase acts processively, synthesizing the leading strand, while the other forms Okazaki fragments initiated by primers created by the primase. DNA polymerase I completes the Okazaki fragments, eliminating the RNA primer of the previous Okazaki fragment and replacing it with DNA. Finally, DNA ligase connects the fragments. Also involved in the process are single-strand binding proteins and topoisomerases that relieve the DNA's supercoiling. Initiation involves initiation proteins at *oriC,* and termination requires termination proteins bound to the termination sites and a topoisomerase.

PROBLEM 4: What can be concluded about the nucleic acids in the following table?

Nucleic Acid Molecule	%A	%T	%G	%C	%U
a.	28	28	22	22	0
b.	31	0	31	17	21
c.	15	15	35	35	0

Answer: We must first look to see if U or T is present, for this will indicate whether the molecule is RNA or DNA, respectively. Molecule b is RNA; a and c are DNA. Now we look at base composition. In double-stranded molecules, A pairs evenly with T (or U) and G pairs with C. This relationship holds for molecules a and c, so they are double-stranded; molecule b is single-stranded. Finally, the melting temperature increases with the amount of G-C, so the melting temperature of c is greater than that of a.

EXERCISES AND PROBLEMS*

CHEMISTRY OF NUCLEIC ACIDS

1. If the tetranucleotide hypothesis were correct regarding the simplicity of DNA structure, under what circumstances could DNA be the genetic material?
2. Nucleic acids, proteins, carbohydrates, and fatty acids could have been mentioned as potential genetic material. What other molecular moieties (units) in the cell could possibly have functioned as the genetic material?
3. In what component parts do DNA and RNA differ?
4. Draw the structure of a short segment of DNA (three base pairs) at the molecular level. Indicate the polarity of the strands.
5. Roughly sketch the shape of B and Z DNA, remembering that B DNA is a right-handed helix and Z DNA is a left-handed helix.
6. Deduce whether each of the nucleic acid molecules in the following table is DNA or RNA and single-stranded or double-stranded.

Nucleic Acid Molecule	%A	%G	%T	%C	%U
a.	33	17	33	17	0
b.	33	33	17	17	0
c.	26	24	0	24	26
d.	21	40	21	18	0
e.	15	40	0	30	15
f.	30	20	15	20	15

7. A double-stranded DNA molecule is 28% guanosine (G).
   - **a.** What is the complete base composition of this molecule?
   - **b.** Answer the same question, but assume the molecule is double-stranded RNA.
8. The following are melting temperatures for five DNA molecules: 73° C, 69° C, 84° C, 78° C, 82° C. Arrange these DNAs in increasing order of percentage of G-C pairs.
9. We normally think that single-stranded nucleic acids should not melt, but many, in fact, do have a T_m. How can you explain this apparent mystery?
10. In a single-stranded DNA molecule, the amount of G is twice the amount of A, the amount of T is three times the amount of C, and the ratio of pyrimidines to purines is 1.5:1. What is the base composition of the DNA?
11. A double-stranded DNA measures 6.5 m in length. Approximately how many base pairs does it contain?

DNA REPLICATION—THE PROCESS

12. Diagram the results that Meselson and Stahl would have obtained (a) if DNA replication were conservative and (b) if it were dispersive.
13. What type of photo would J. Cairns have obtained if DNA replication were conservative? Dispersive?

*Answers to selected exercises and problems are on page A-10.

DNA REPLICATION—THE ENZYMOLOGY

14. Following is a section of a single strand of DNA. Supply a strand, by the rules of complementarity, that would turn this into a double helix. What RNA bases would primase use if this segment initiated an Okazaki fragment? In which direction would replication proceed?

5′-ATTCTTGGCATTCGC-3′

15. What is a primosome in *E. coli*? a replisome? What enzymes make up each? What is the relationship between these structures?

16. What are the differences between continuous and discontinuous DNA replication? Why do both exist?

17. Describe the synthesis of an Okazaki fragment.

18. Describe the enzymology of the origin, continuation, and termination of DNA replication in *E. coli.*

19. Can you think of any other mechanisms besides topoisomerase activity that could release supercoiling in replicating DNA?

20. Draw a diagram showing how topoisomerase II (gyrase) might work.

21. Retroviruses are single-stranded RNA viruses that insert their genomes into the host DNA during their life cycle. But only double-stranded DNA can be inserted into double-stranded DNA.

a. Propose a mechanism that retroviruses could use to insert their genomes.

b. What novel enzymes might such viruses require?

22. Propose a mechanism by which a single strand of DNA can make multiple copies of itself.

23. Progeria is a human disorder that causes affected individuals to age prematurely; a nine-year-old often resembles a sixty- to seventy-year-old individual in appearance and physiology. Suppose you extract DNA from a progeric patient and find mostly small DNA fragments rather than the expected long DNA molecules. What enzyme(s) might be defective in patients with progeria?

REPLICATION STRUCTURES

24. Under what circumstances would you expect to see a DNA theta structure? D-loop? rolling-circle? bubbles? What function does each structure serve?

EUKARYOTIC DNA REPLICATION

25. In developing sea urchins, just after fertilization, the cells divide every thirty to forty minutes. In the adult, the cells divide once every ten to fifteen hours. The amount of DNA per cell is the same in each case, but the DNA obviously replicates much faster in developing cells. Propose an explanation to account for the difference in replication time.

CRITICAL THINKING QUESTIONS

1. Mutants are used to study various aspects of the phenotype and genotype. How can we study genes that are critically important in the functioning of an organism? For example, how do we study mutations in the gene for DNA polymerase III in *E. coli,* when changes in this gene are usually lethal? Remember, to study the genes in bacteria, we need the bacteria to grow and form colonies in order to be scored for their phenotypes.

2. DNA and RNA differ in two major ways: DNA has deoxyribose sugar, whereas RNA has ribose, and DNA has thymine, whereas RNA has uracil. Why might those differences exist other than accidents of evolution?

Suggested Readings for chapter 9 are on page B-5.