

Bronze sculpture  
of the Watson–Crick  
model for double-helical DNA.



# 10

## DNA Structure and Analysis

### CHAPTER CONCEPTS

- Except in some viruses, DNA serves as the genetic material in all living organisms on Earth.
- According to the Watson–Crick model, DNA exists in the form of a right-handed double helix.
- The strands of the double helix are antiparallel and are held together by hydrogen bonding between complementary nitrogenous bases.
- The structure of DNA provides the means of storing and expressing genetic information.
- RNA has many similarities to DNA but exists mostly as a single-stranded molecule.
- In some viruses, RNA serves as the genetic material.
- Many techniques have been developed that facilitate the analysis of nucleic acids, most based on detection of the complementarity of nitrogenous bases.

Up to this point in the text, we have described chromosomes as containing genes that control phenotypic traits transmitted through gametes to future offspring. Logically, genes must contain some sort of information that, when passed to a new generation, influences the form and characteristics of each individual. We refer to that information as the **genetic material**. Logic also suggests that this same information in some way directs the

many complex processes that lead to an organism's adult form.

Until 1944, it was not clear what chemical component of the chromosome makes up genes and constitutes the genetic material. Because chromosomes were known to have both a nucleic acid and a protein component, both were candidates. In 1944, however, direct experimental evidence emerged showing that the nucleic acid DNA serves as the informational basis for the process of heredity.

Once the importance of DNA to genetic processes was realized, work was intensified with the hope of discerning not only the structure of this molecule but also the relationship of its structure to its function. Between 1944 and 1953, many scientists sought information that might answer the most significant and intriguing question in the history of biology: How does DNA serve as the genetic basis for living processes? Researchers believed the answer must depend strongly on the chemical structure of the DNA molecule, given the complex but orderly functions ascribed to it.

These efforts were rewarded in 1953, when James Watson and Francis Crick put forth their hypothesis for the double-helical nature of DNA. The assumption that the molecule's functions would be easier to clarify once its general structure was determined proved to be correct. In this chapter, we first review the evidence that DNA is the genetic material and then discuss the elucidation of its structure. We conclude the chapter with a discussion of

various analytical techniques useful during the investigation of the nucleic acids, DNA and RNA.

Please note that some of the topics discussed in this chapter are explored in greater depth later in the text (see Special Topic Chapter 2—Emerging Roles of RNA).

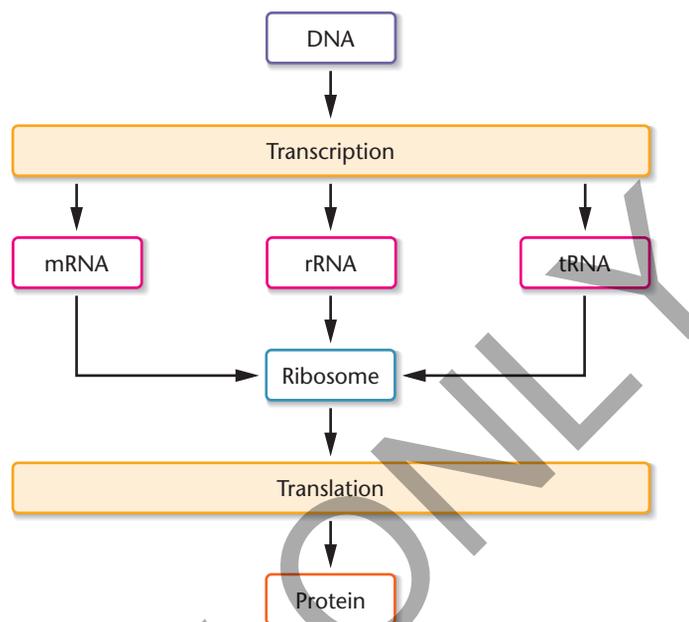
## 10.1 The Genetic Material Must Exhibit Four Characteristics

For a molecule to serve as the genetic material, it must exhibit four crucial characteristics: **replication, storage of information, expression of information, and variation by mutation.** *Replication* of the genetic material is one facet of the cell cycle and as such is a fundamental property of all living organisms. Once the genetic material of cells replicates and is doubled in amount, it must then be partitioned equally—through mitosis—into daughter cells. The genetic material is also replicated during the formation of gametes, but is partitioned so that each cell gets only one-half of the original amount of genetic material—the process of *meiosis* (discussed in Chapter 2). Although the products of mitosis and meiosis are different, these processes are both part of the more general phenomenon of cellular reproduction.

*Storage of information* requires the molecule to act as a repository of genetic information that may or may not be expressed by the cell in which it resides. It is clear that while most cells contain a complete copy of the organism's genome, at any point in time they express only a part of this genetic potential. For example, in bacteria many genes “turn on” in response to specific environmental cues and “turn off” when conditions change. In vertebrates, skin cells may display active melanin genes but never activate their hemoglobin genes; in contrast, digestive cells activate many genes specific to their function but do not activate their melanin genes.

Inherent in the concept of storage is the need for the genetic material to be able to encode the vast variety of gene products found among the countless forms of life on our planet. The chemical language of the genetic material must have the capability of storing such diverse information and transmitting it to progeny cells and organisms.

*Expression* of the stored genetic information is a complex process that is the underlying basis for the concept of **information flow** within the cell (Figure 10-1). The initial event in this flow of information is the **transcription** of DNA, in which three main types of RNA molecules are synthesized: messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). Of these, mRNAs are translated into proteins, by means of a process mediated by the tRNA and rRNA. Each mRNA is the product of a specific gene and leads to the synthesis of a different protein. In



**FIGURE 10-1** Simplified diagram of information flow (the central dogma) from DNA to RNA to produce the proteins within cells.

**translation**, the chemical information in mRNA directs the construction of a chain of amino acids, called a polypeptide, which then folds into a protein. Collectively, these processes serve as the foundation for the **central dogma of molecular genetics**: “DNA makes RNA, which makes proteins.”

The genetic material is also the source of *variability* among organisms, through the process of mutation. If a mutation—a change in the chemical composition of DNA—occurs, the alteration is reflected during transcription and translation, affecting the specific protein. If a mutation is present in a gamete, it may be passed to future generations and, with time, become distributed in the population. Genetic variation, which also includes alterations of chromosome number and rearrangements within and between chromosomes (as discussed in Chapter 8), provides the raw material for the process of evolution.

## 10.2 Until 1944, Observations Favored Protein as the Genetic Material

The idea that genetic material is physically transmitted from parent to offspring has been accepted for as long as the concept of inheritance has existed. Beginning in the late nineteenth century, research into the structure of biomolecules progressed considerably, setting the stage for describing the genetic material in chemical terms. Although proteins and nucleic acid were both considered major candidates for the role of genetic material, until the 1940s many geneticists favored proteins.

This is not surprising, since proteins were known to be both diverse and abundant in cells, and much more was known about protein than about nucleic acid chemistry.

DNA was first studied in 1869 by a Swiss chemist, Friedrich Miescher. He isolated cell nuclei and derived an acidic substance, now known to contain DNA, that he called **nuclein**. As investigations of DNA progressed, however, showing it to be present in chromosomes, the substance seemed to lack the chemical diversity necessary to store extensive genetic information.

This conclusion was based largely on Phoebus A. Levene's observations in 1910 that DNA contained approximately equal amounts of four similar molecules called *nucleotides*. Levene postulated incorrectly that identical groups of these four components were repeated over and over, which was the basis of his **tetranucleotide hypothesis** for DNA structure. Attention was thus directed away from DNA, thereby favoring proteins. However, in the 1940s, Erwin Chargaff showed that Levene's proposal was incorrect when he demonstrated that most organisms do not contain precisely equal proportions of the four nucleotides. We shall see later that the structure of DNA accounts for Chargaff's observations.

### 10.3 Evidence Favoring DNA as the Genetic Material Was First Obtained during the Study of Bacteria and Bacteriophages

The 1944 publication by Oswald Avery, Colin MacLeod, and Maclyn McCarty concerning the chemical nature of a "transforming principle" in bacteria was the initial event leading to the acceptance of DNA as the genetic material. Their work, along with subsequent findings of other research teams, constituted the first direct experimental proof that DNA, and not protein, is the biomolecule responsible for heredity. It marked the beginning of the *era of molecular genetics*, a period of discovery in biology that made biotechnology feasible and has moved us closer to an understanding of the basis of life. The impact of their initial findings on future research and thinking paralleled that of the publication of Darwin's theory of evolution and the subsequent rediscovery of Mendel's postulates of transmission genetics. Together, these events constitute three great revolutions in biology.

#### Transformation: Early Studies

The research that provided the foundation for Avery, MacLeod, and McCarty's work was initiated in 1927 by Frederick Griffith, a medical officer in the British Ministry of Health. He performed experiments with several different

strains of the bacterium *Diplococcus pneumoniae*.<sup>\*</sup> Some were *virulent*, that is, infectious, strains that cause pneumonia in certain vertebrates (notably humans and mice), whereas others were *avirulent*, or noninfectious strains, which do not cause illness.

The difference in virulence depends on the presence of a polysaccharide capsule; virulent strains have this capsule, whereas avirulent strains do not. The nonencapsulated bacteria are readily engulfed and destroyed by phagocytic cells in the host animal's circulatory system. Virulent bacteria, which possess the polysaccharide coat, are not easily engulfed; they multiply and cause pneumonia.

The presence or absence of the capsule causes a visible difference between colonies of virulent and avirulent strains. Encapsulated bacteria form smooth, shiny-surfaced colonies (S) when grown on an agar culture plate; non-encapsulated strains produce rough colonies (R). Thus, virulent and avirulent strains are easily distinguished by standard microbiological culture techniques.

Each strain of *Diplococcus* may be one of dozens of different types called *serotypes* that differ in the precise chemical structure of the polysaccharide constituent of the thick, slimy capsule. Serotypes are identified by immunological techniques and are usually designated by Roman numerals. In the United States, types I and II are the most common in causing pneumonia. Griffith used types IIR and IIIS in his critical experiments that led to new concepts about the genetic material. **Table 10.1** summarizes the characteristics of Griffith's two strains, while **Figure 10-2** on p. 234 depicts his experiment.

Griffith knew from the work of others that only living virulent cells would produce pneumonia in mice. If heat-killed virulent bacteria are injected into mice, no pneumonia results, just as living avirulent bacteria fail to produce the disease. Griffith's critical experiment (Figure 10-2) involved an injection into mice of living IIR (avirulent) cells combined with heat-killed IIIS (virulent) cells. Since neither cell type caused death in mice when injected alone, Griffith expected that the double injection would not kill the mice. But, after five days, all of the mice that received both types of cells were dead. Paradoxically, analysis of their blood revealed a large number of living type IIIS (virulent) bacteria.

As far as could be determined, these IIIS bacteria were identical to the IIIS strain from which the heat-killed cell preparation had been made. The control mice, injected

**TABLE 10.1** Strains of *Diplococcus pneumoniae* Used by Frederick Griffith in His Original Transformation Experiments

Serotype	Colony Morphology	Capsule	Virulence
IIR	Rough	Absent	Avirulent
IIIS	Smooth	Present	Virulent

<sup>\*</sup>This organism is now named *Streptococcus pneumoniae*.

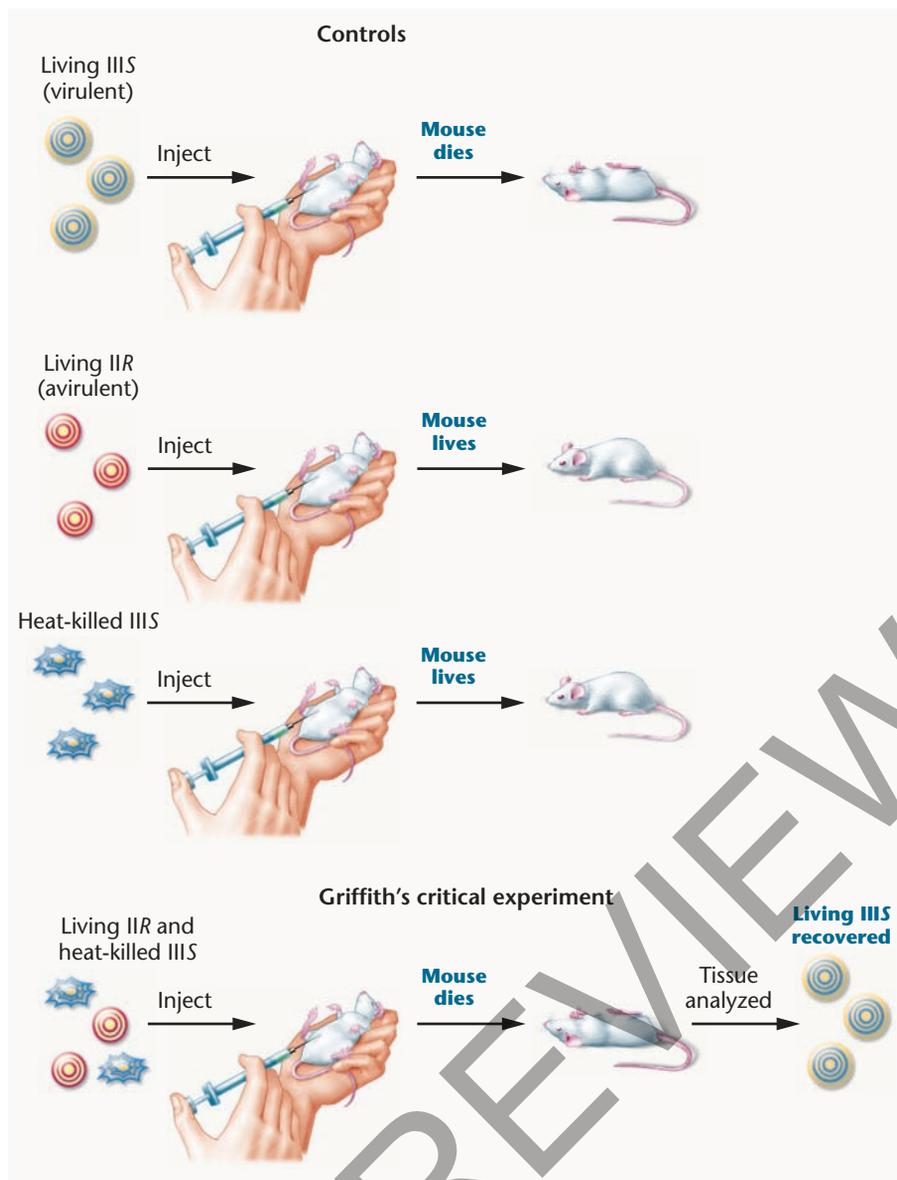


FIGURE 10-2 Griffith's transformation experiment.

only with living avirulent IIR bacteria for this set of experiments, did not develop pneumonia and remained healthy. This ruled out the possibility that the avirulent IIR cells simply changed (or mutated) to virulent III S cells in the absence of the heat-killed III S bacteria. Instead, some type of interaction had taken place between living IIR and heat-killed III S cells.

Griffith concluded that the heat-killed III S bacteria somehow converted live avirulent IIR cells into virulent III S cells. Calling the phenomenon **transformation**, he suggested that the *transforming principle* might be some part of the polysaccharide capsule or a compound required for capsule synthesis, although the capsule alone did not cause pneumonia. To use Griffith's term, the transforming principle from the dead III S cells served as a "pabulum"—that is, a nutrient source—for the IIR cells.

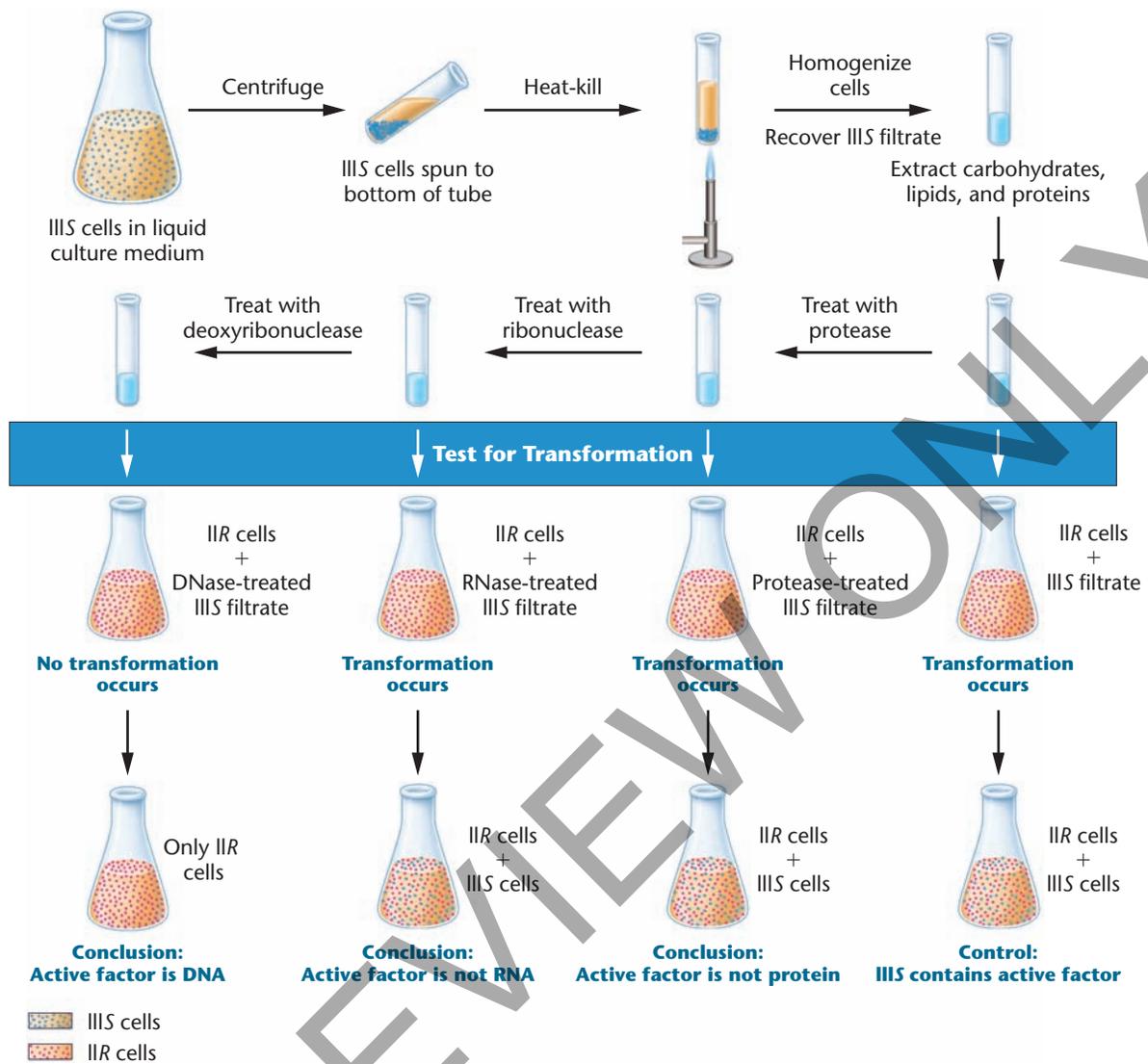
Griffith's work led other physicians and bacteriologists to research the phenomenon of transformation. By 1931, M. Henry Dawson at the Rockefeller Institute had confirmed Griffith's observations and extended his work one step further. Dawson and his coworkers showed that transformation could occur *in vitro* (in a test tube). When heat-killed III S cells were incubated with living IIR cells, living III S cells were recovered. Therefore, injection into mice was not necessary for transformation to occur. By 1933, J. Lionel Alloway had refined the *in vitro* experiments by using crude extracts of III S cells and living IIR cells. The soluble filtrate from the heat-killed III S cells was as effective in inducing transformation as were the intact cells. Alloway and others did not view transformation as a genetic event, but rather as a physiological modification of some sort. Nevertheless, the experimental evidence that a chemical substance was responsible for transformation was quite convincing.

### Transformation: The Avery, MacLeod, and McCarty Experiment

The critical question, of course, was what molecule serves as the trans-

forming principle? In 1944, after 10 years of work, Avery, MacLeod, and McCarty published their results in what is now regarded as a classic paper in the field of molecular genetics. They reported that they had obtained the transforming principle in a purified state and that beyond reasonable doubt it was DNA.

The details of their work, sometimes called the Avery, MacLeod, and McCarty experiment, are outlined in **Figure 10-3**. These researchers began their isolation procedure with large quantities (50–75 liters) of liquid cultures of type III S virulent cells. The cells were centrifuged, collected, and heat killed. Following homogenization and several extractions with the detergent deoxycholate (DOC), the researchers obtained a soluble filtrate that retained the ability to induce transformation of type IIR avirulent cells. Protein was removed from the active filtrate by several chloroform extractions, and polysaccharides were enzymatically digested and



**FIGURE 10-3** Summary of Avery, MacLeod, and McCarty's experiment demonstrating that DNA is the transforming principle.

removed. Finally, precipitation with ethanol yielded a fibrous mass that still retained the ability to induce transformation of type IIR avirulent cells. From the original 75-liter sample, the procedure yielded 10 to 25 mg of this "active factor."

Further testing clearly established that the transforming principle was DNA. The fibrous mass was first analyzed for its nitrogen: phosphorus ratio, which was shown to coincide with the ratio of "sodium desoxyribonucleate," the chemical name then used to describe DNA. To solidify their findings, Avery, MacLeod, and McCarty sought to eliminate, to the greatest extent possible, all probable contaminants from their final product. Thus, it was treated with the proteolytic enzymes trypsin and chymotrypsin and then with an RNA-digesting enzyme, called **ribonuclease (RNase)**. Such treatments destroyed any remaining activity of proteins and RNA. Nevertheless, transforming activity still remained. Chemical testing of the final product gave strong positive reactions for DNA.

The final confirmation came with experiments using crude samples of the DNA-digesting enzyme **deoxyribonuclease (DNase)**, which was isolated from dog and rabbit sera. Digestion with this enzyme destroyed the transforming activity of the filtrate—thus Avery and his coworkers were certain that the active transforming principle in these experiments was DNA.

The great amount of work involved in this research, the confirmation and reconfirmation of the conclusions drawn, and the unambiguous logic of the experimental design are truly impressive. Avery, MacLeod, and McCarty's conclusion in the 1944 publication was, however, very simply stated: "The evidence presented supports the belief that a nucleic acid of the desoxyribose\* type is the fundamental unit of the transforming principle of *Pneumococcus* Type III."

\*Desoxyribose is now spelled deoxyribose.

Avery and his colleagues recognized the genetic and biochemical implications of their work. They observed that “nucleic acids of this type must be regarded not merely as structurally important but as functionally active in determining the biochemical activities and specific characteristics of pneumococcal cells.” This suggested that the transforming principle interacts with the IIR cell and gives rise to a coordinated series of enzymatic reactions culminating in the synthesis of the type IIIS capsular polysaccharide. Avery, MacLeod, and McCarty emphasized that, once transformation occurs, the capsular polysaccharide is produced in successive generations. Transformation is therefore heritable, and the process affects the genetic material.

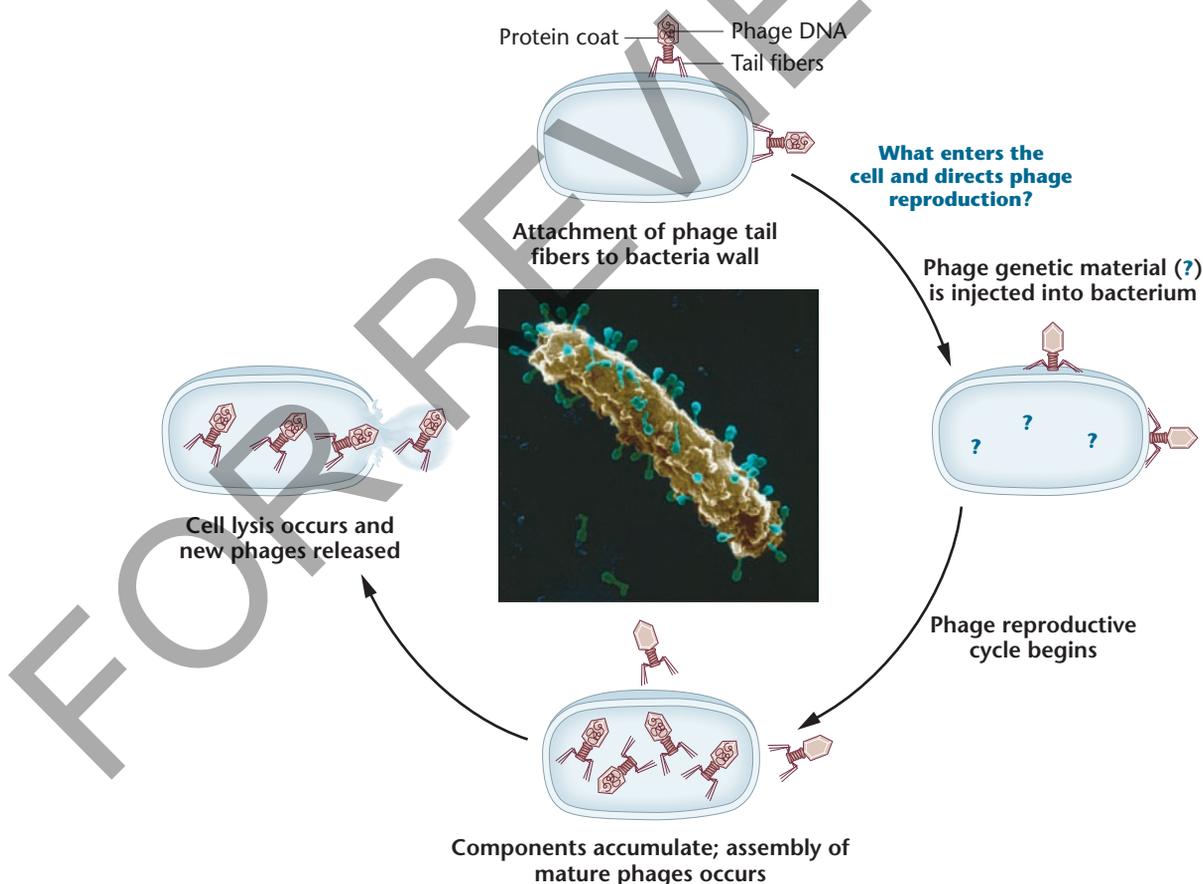
Immediately after publication of the report, several investigators turned to, or intensified, their studies of transformation in order to clarify the role of DNA in genetic mechanisms. In particular, the work of Rollin Hotchkiss was instrumental in confirming that the critical factor in transformation was DNA and not protein. In 1949, in a separate study, Harriet Taylor isolated an **extremely rough** (*ER*) mutant strain from a rough (*R*) strain. This *ER* strain produced colonies that were more irregular than the *R* strain. The DNA from *R* accomplished the transformation

of *ER* to *R*. Thus, the *R* strain, which served as the recipient in the Avery experiments, was shown also to be able to serve as the DNA donor in transformation.

Transformation has now been shown to occur in *Haemophilus influenzae*, *Bacillus subtilis*, *Shigella paradysenteriae*, and *Escherichia coli*, among many other microorganisms. Transformation of numerous genetic traits other than colony morphology has also been demonstrated, including traits involving resistance to antibiotics. These observations further strengthened the belief that transformation by DNA is primarily a genetic event rather than simply a physiological change. We will pursue this idea again in the “Insights and Solutions” section at the end of this chapter.

### The Hershey–Chase Experiment

The second major piece of evidence supporting DNA as the genetic material was provided during the study of the bacterium *Escherichia coli* and one of its infecting viruses, **bacteriophage T2**. Often referred to simply as a **phage**, the virus consists of a protein coat surrounding a core of DNA. Electron micrographs reveal that the phage’s external structure is composed of a hexagonal head plus a tail. **Figure 10–4** shows as much of the life cycle as was known



**FIGURE 10–4** Life cycle of a T-even bacteriophage, as known in 1952. The electron micrograph shows an *E. coli* cell during infection by numerous T2 phages (shown in blue).

in 1952 for a T-even bacteriophage such as T2. Briefly, the phage adsorbs to the bacterial cell, and some genetic component of the phage enters the bacterial cell. Following infection, the viral component “commandeers” the cellular machinery of the host and causes viral reproduction. In a reasonably short time, many new phages are constructed and the bacterial cell is lysed, releasing the progeny viruses. This process is referred to as the **lytic cycle**.

In 1952, Alfred Hershey and Martha Chase published the results of experiments designed to clarify the events leading to phage reproduction. Several of the experiments clearly established the independent functions of phage protein and nucleic acid in the reproduction process associated with the bacterial cell. Hershey and Chase knew from existing data that:

1. T2 phages consist of approximately 50 percent protein and 50 percent DNA.
2. Infection is initiated by adsorption of the phage by its tail fibers to the bacterial cell.
3. The production of new viruses occurs within the bacterial cell.

It appeared that some molecular component of the phage—DNA or protein (or both)—entered the bacterial cell and directed viral reproduction. Which was it?

Hershey and Chase used the radioisotopes  $^{32}\text{P}$  and  $^{35}\text{S}$  to follow the molecular components of phages during infection. Because DNA contains phosphorus (P) but not sulfur,  $^{32}\text{P}$  effectively labels DNA; because proteins contain sulfur (S) but not phosphorus,  $^{35}\text{S}$  labels protein. *This is a key feature of the experiment.* If *E. coli* cells are first grown in the presence of  $^{32}\text{P}$  or  $^{35}\text{S}$  and then infected with T2 viruses, the progeny phages will have *either* a radioactively labeled DNA core *or* a radioactively labeled protein coat, respectively. These labeled phages can be isolated and used to infect unlabeled bacteria (Figure 10–5).

When labeled phages and unlabeled bacteria were mixed, an adsorption complex was formed as the phages attached their tail fibers to the bacterial wall. These complexes were isolated and subjected to a high shear force in a blender. The force stripped off the attached phages so that the phages and bacteria could be analyzed separately. Centrifugation separated the lighter phage particles from the heavier bacterial cells (Figure 10–5). By tracing the radioisotopes, Hershey and Chase were able to demonstrate that most of the  $^{32}\text{P}$ -labeled DNA had been transferred into the bacterial cell following adsorption; on the other hand, almost all of the  $^{35}\text{S}$ -labeled protein remained outside the bacterial cell and was recovered in the phage “ghosts” (empty phage coats) after the blender treatment.

Following this separation, the bacterial cells, which now contained viral DNA, were eventually lysed as new phages were produced. These progeny phages contained  $^{32}\text{P}$ , but not  $^{35}\text{S}$ .

Hershey and Chase interpreted these results as indicating that the protein of the phage coat remains outside the host cell and is not involved in directing the production of new phages. On the other hand, and most important, phage DNA enters the host cell and directs phage reproduction. Hershey and Chase had demonstrated that the genetic material in phage T2 is DNA, not protein.

These experiments, along with those of Avery and his colleagues, provided convincing evidence that DNA was the molecule responsible for heredity. This conclusion has since served as the cornerstone of the field of molecular genetics.

#### NOW SOLVE THIS

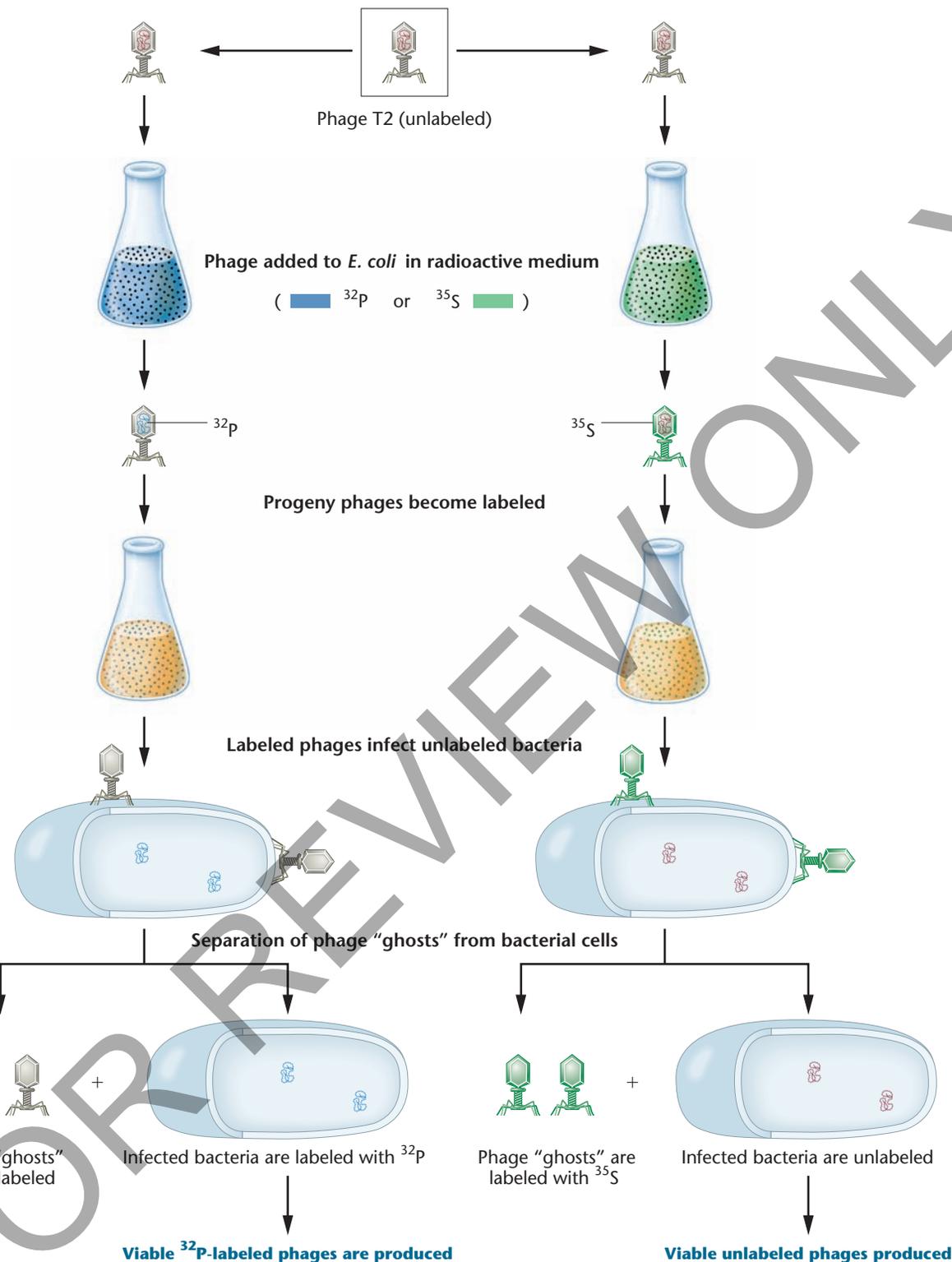
**10–1** Would an experiment similar to that performed by Hershey and Chase work if the basic design were applied to the phenomenon of transformation? Explain why or why not.

■ **HINT:** *This problem involves an understanding of the protocol of the Hershey–Chase experiment as applied to the investigation of transformation. The key to its solution is to remember that in transformation, exogenous DNA enters the soon-to-be transformed cell and that no cell-to-cell contact is involved in the process.*

### Transfection Experiments

During the eight years following publication of the Hershey–Chase experiment, additional research using bacterial viruses provided even more solid proof that DNA is the genetic material. In 1957, several reports demonstrated that if *E. coli* is treated with the enzyme lysozyme, the outer wall of the cell can be removed without destroying the bacterium. Enzymatically treated cells are naked, so to speak, and contain only the cell membrane as their outer boundary. Such structures are called **protoplasts** (or **sphe-roplasts**). John Spizizen and Dean Fraser independently reported that by using protoplasts, they were able to initiate phage reproduction with disrupted T2 particles. That is, provided protoplasts were used, a virus did not have to be intact for infection to occur. Thus, the outer protein coat structure may be essential to the movement of DNA through the intact cell wall, but it is not essential for infection when protoplasts are used.

Similar, but more refined, experiments were reported in 1960 by George Guthrie and Robert Sinsheimer. DNA was purified from bacteriophage  $\phi\text{X174}$ , a small phage



**FIGURE 10-5** Summary of the Hershey–Chase experiment demonstrating that DNA, and not protein, is responsible for directing the reproduction of phage T2 during the infection of *E. coli*.

that contains a single-stranded circular DNA molecule of some 5386 nucleotides. When added to *E. coli* protoplasts, the purified DNA resulted in the production of complete  $\phi\text{X174}$  bacteriophages. This process of infection by only the viral nucleic acid, called **transfection**, proves conclusively

that  $\phi\text{X174}$  DNA alone contains all the necessary information for production of mature viruses. Thus, the evidence that DNA serves as the genetic material was further strengthened, even though all direct evidence to that point had been obtained from bacterial and viral studies.

## 10.4 Indirect and Direct Evidence Supports the Concept that DNA Is the Genetic Material in Eukaryotes

In 1950, eukaryotic organisms were not amenable to the types of experiments that used bacteria and viruses to demonstrate that DNA is the genetic material. Nevertheless, it was generally assumed that the genetic material would be a universal substance serving the same role in eukaryotes. Initially, support for this assumption relied on several circumstantial observations that, taken together, indicated that DNA does serve as the genetic material in eukaryotes. Subsequently, direct evidence established unequivocally the central role of DNA in genetic processes.

### Indirect Evidence: Distribution of DNA

The genetic material should be found where it functions—in the nucleus as part of chromosomes. Both DNA and protein fit this criterion. However, protein is also abundant in the cytoplasm, whereas DNA is not. Both mitochondria and chloroplasts are known to perform genetic functions, and DNA is also present in these organelles. Thus, DNA is found only where primary genetic functions occur. Protein, on the other hand, is found everywhere in the cell. These observations are consistent with the interpretation favoring DNA over proteins as the genetic material.

Because it had earlier been established that chromosomes within the nucleus contain the genetic material, a correlation was expected to exist between the ploidy ( $n$ ,  $2n$ , etc.) of a cell and the quantity of the substance that functions as the genetic material. Meaningful comparisons can be made between gametes (sperm and eggs) and somatic or body cells. The latter are recognized as being diploid ( $2n$ ) and containing twice the number of chromosomes as gametes, which are haploid ( $n$ ).

**Table 10.2** compares, for a variety of organisms, the amount of DNA found in haploid sperm to the amount found in diploid nucleated precursors of red blood cells. The amount of DNA and the number of sets of chromosomes is closely

**TABLE 10.2** DNA Content of Haploid versus Diploid Cells of Various Species\*

Organism	$n$ (pg)	$2n$ (pg)
Human	3.25	7.30
Chicken	1.26	2.49
Trout	2.67	5.79
Carp	1.65	3.49
Shad	0.91	1.97

\*Sperm ( $n$ ) and nucleated precursors to red blood cells ( $2n$ ) were used to contrast ploidy levels.

correlated. No such consistent correlation can be observed between gametes and diploid cells for proteins. These data thus provide further circumstantial evidence favoring DNA over proteins as the genetic material of eukaryotes.

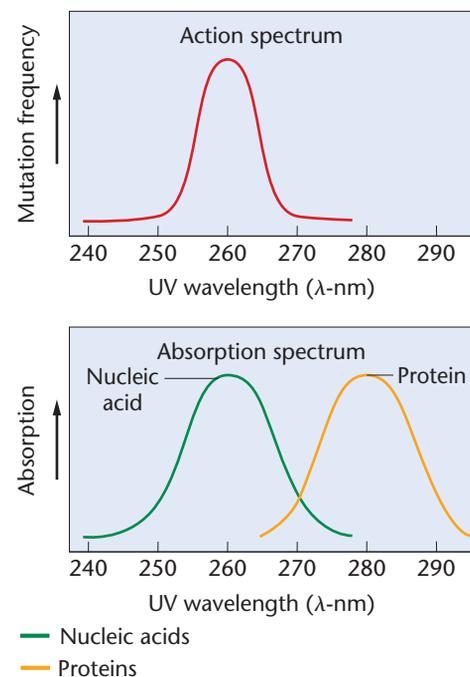
### Indirect Evidence: Mutagenesis

**Ultraviolet (UV) light** is one of a number of agents capable of inducing mutations in the genetic material. Simple organisms such as yeast and other fungi can be irradiated with various wavelengths of ultraviolet light and the effectiveness of each wavelength measured by the number of mutations it induces. When the data are plotted, an **action spectrum** of UV light as a mutagenic agent is obtained. This action spectrum can then be compared with the **absorption spectrum** of any molecule suspected to be the genetic material (**Figure 10-6**). *The molecule serving as the genetic material is expected to absorb at the wavelength(s) found to be mutagenic.*

UV light is most mutagenic at the wavelength  $\lambda$  of 260 nanometers (nm), and both DNA and RNA absorb UV light most strongly at 260 nm. On the other hand, protein absorbs most strongly at 280 nm, yet no significant mutagenic effects are observed at that wavelength. This indirect evidence supports the idea that a nucleic acid, rather than protein, is the genetic material.

### Direct Evidence: Recombinant DNA Studies

Although the circumstantial evidence just described does not constitute direct proof that DNA is the genetic material



**FIGURE 10-6** Comparison of the action spectrum (which determines the most effective mutagenic UV wavelength) and the absorption spectrum (which shows the range of wavelength where nucleic acids and proteins absorb UV light).

in eukaryotes, those observations spurred researchers to forge ahead using this supposition as the underlying hypothesis. Today, there is no doubt of its validity; DNA is the genetic material in all eukaryotes. The strongest evidence is provided by molecular analysis utilizing **recombinant DNA technology**. In this procedure, segments of eukaryotic DNA corresponding to specific genes are isolated and spliced into bacterial DNA. The resulting complex can be inserted into a bacterial cell, and then its genetic expression is monitored. If a eukaryotic gene is introduced, the subsequent production of the corresponding eukaryotic protein product demonstrates directly that the eukaryotic DNA is now present and functional in the bacterial cell. This has been shown to be the case in countless instances. For example, the products of the human genes specifying insulin and interferon are produced by bacteria after the human genes that encode these proteins are inserted. As the bacterium divides, the eukaryotic DNA replicates along with the bacterial DNA and is distributed to the daughter cells, which also express the human genes by creating the corresponding proteins.

The availability of vast amounts of DNA coding for specific genes, derived from recombinant DNA research, has led to other direct evidence that DNA serves as the genetic material. Work in the laboratory of Beatrice Mintz has demonstrated that DNA encoding the human  $\beta$ -globin gene, when microinjected into a fertilized mouse egg, is later found to be present and expressed in adult mouse tissue and transmitted to and expressed in that mouse's progeny. These mice are examples of what are called **transgenic animals**. Other work has introduced rat DNA encoding a growth hormone into fertilized mouse eggs. About one-third of the resultant mice grew to twice their normal size, indicating that foreign DNA was present and functional. Subsequent generations of mice inherited this genetic information and also grew to a large size. This clearly demonstrates that DNA meets the requirement of expression of genetic information in eukaryotes. Later, we will see exactly how DNA is stored, replicated, expressed, and mutated.

## 10.5 RNA Serves as the Genetic Material in Some Viruses

Some viruses contain an RNA core rather than a DNA core. In these viruses, it appears that RNA serves as the genetic material—an exception to the general rule that DNA performs this function. In 1956, it was demonstrated that when purified RNA from **tobacco mosaic virus (TMV)** was spread on tobacco leaves, the characteristic lesions caused by viral infection subsequently appeared. Thus, it was concluded that RNA is the genetic material of this virus.

In 1965 and 1966, Norman Pace and Sol Spiegelman demonstrated that RNA from the phage Q $\beta$  can be isolated and replicated *in vitro*. Replication depends on an enzyme, **RNA replicase**, which is isolated from host *E. coli* cells following normal infection. When the RNA replicated *in vitro* is added to *E. coli* protoplasts, infection and viral multiplication (*transfection*) occur. Thus, RNA synthesized in a test tube serves as the genetic material in these phages by directing the production of all the components necessary for viral reproduction. While many viruses, such as the T2 virus used by Hershey and Chase, use DNA as their hereditary material, another group of RNA-containing viruses bears mention. These are the **retroviruses**, which replicate in an unusual way. Their RNA serves as a template for the synthesis of the complementary DNA molecule. The process, **reverse transcription**, occurs under the direction of an RNA-dependent DNA polymerase enzyme called **reverse transcriptase**. This DNA intermediate can be incorporated into the genome of the host cell, and when the host DNA is transcribed, copies of the original retroviral RNA chromosomes are produced. Retroviruses include the human immunodeficiency virus (HIV), which causes AIDS, as well as several RNA tumor viruses.

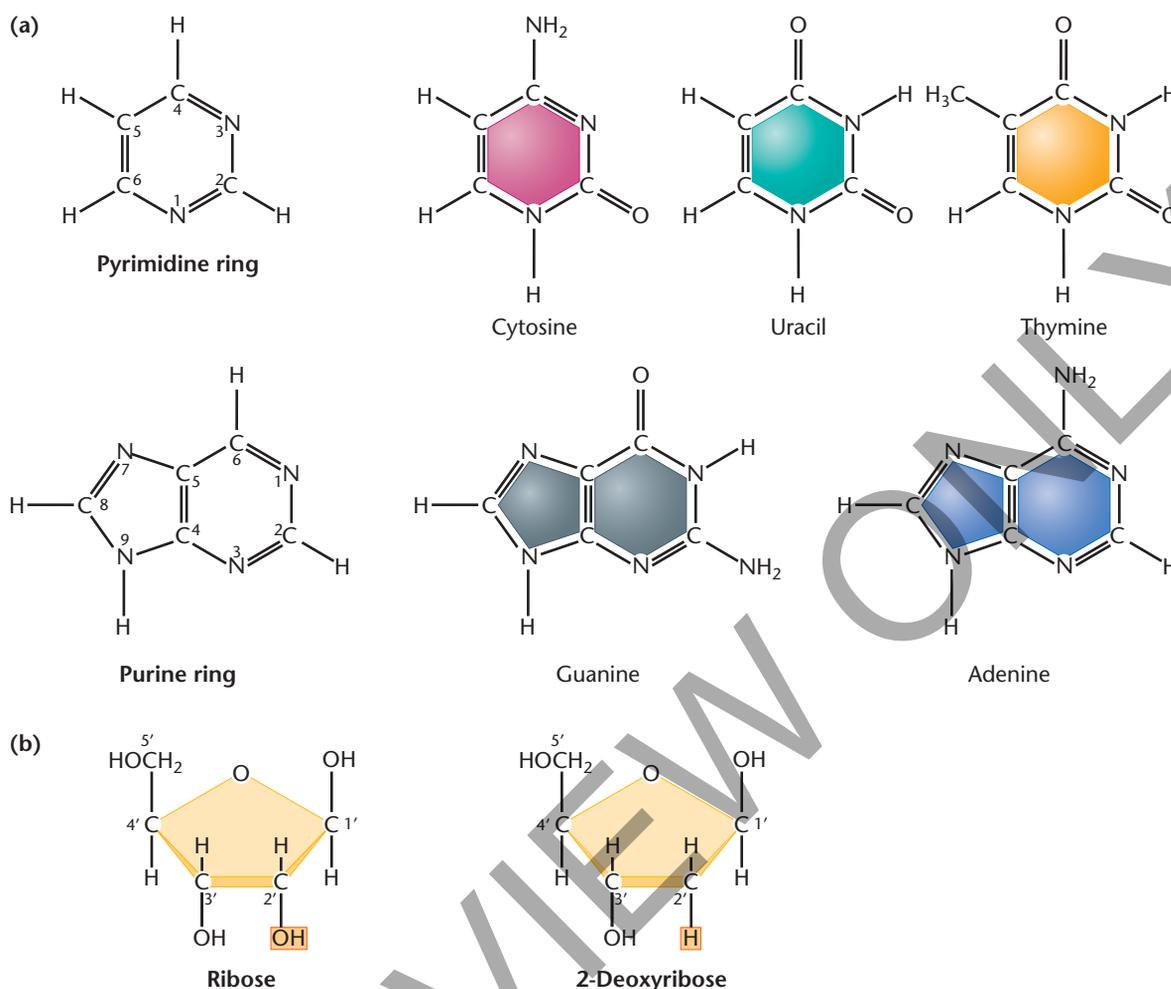
## 10.6 Knowledge of Nucleic Acid Chemistry Is Essential to the Understanding of DNA Structure

Having established the critical importance of DNA and RNA in genetic processes, we will now take a brief look at the chemical structures of these molecules. As we shall see, the structural components of DNA and RNA are very similar. This chemical similarity is important in the coordinated functions played by these molecules during gene expression. Like the other major groups of organic biomolecules (proteins, carbohydrates, and lipids), nucleic acid chemistry is based on a variety of similar building blocks that are polymerized into chains of varying lengths.

### Nucleotides: Building Blocks of Nucleic Acids

DNA is a nucleic acid, and **nucleotides** are the building blocks of all nucleic acid molecules. Sometimes called mononucleotides, these structural units consist of three essential components: a **nitrogenous base**, a **pentose sugar** (a 5-carbon sugar), and a **phosphate group**. There are two kinds of nitrogenous bases: the nine-member double-ring **purines** and the six-member single-ring **pyrimidines**.

Two types of purines and three types of pyrimidines are commonly found in nucleic acids. The two purines are **adenine** and **guanine**, abbreviated A and G. The three



**FIGURE 10-7** (a) Chemical structures of the pyrimidines and purines that serve as the nitrogenous bases in RNA and DNA. The convention for numbering carbon and nitrogen atoms making up the two categories of bases is shown within the structures that appear on the left. (b) Chemical ring structures of ribose and 2-deoxyribose, which serve as the pentose sugars in RNA and DNA, respectively.

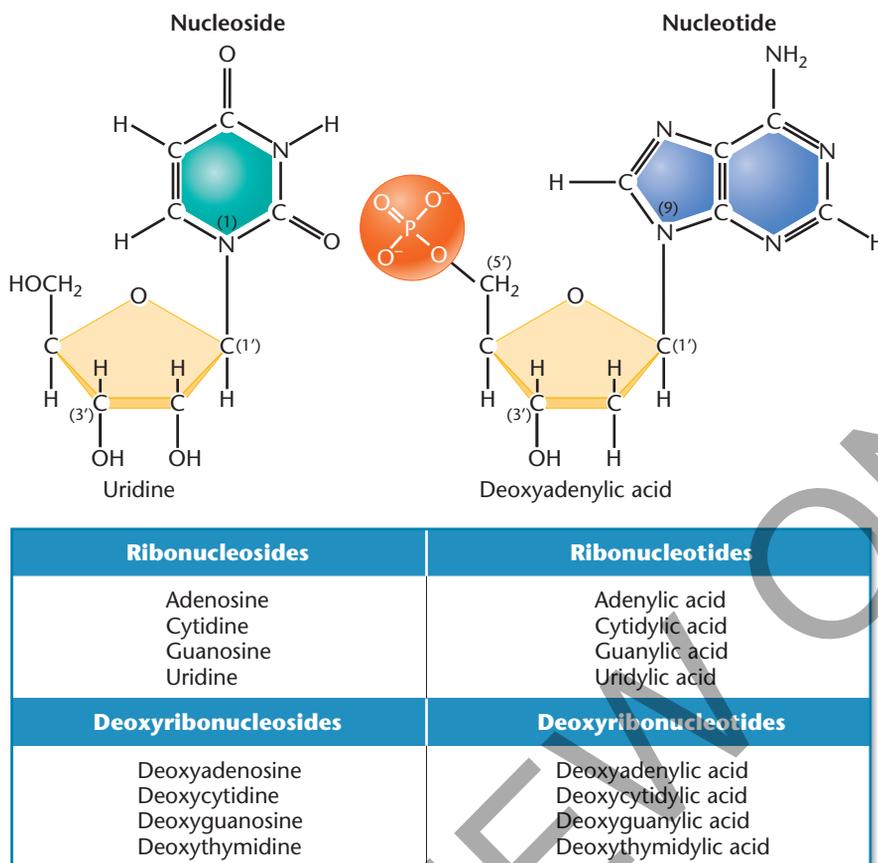
pyrimidines are **cytosine**, **thymine**, and **uracil**, abbreviated **C**, **T**, and **U**, respectively. The chemical structures of A, G, C, T, and U are shown in **Figure 10-7(a)**. Both DNA and RNA contain A, C, and G, but only DNA contains the base T and only RNA contains the base U. Each nitrogen or carbon atom of the ring structures of purines and pyrimidines is designated by an unprimed number. Note that corresponding atoms in the two rings are numbered differently in most cases.

The pentose sugars found in nucleic acids give them their names. Ribonucleic acids (RNA) contain **ribose**, while deoxyribonucleic acids (DNA) contain **deoxyribose**. **Figure 10-7(b)** shows the ring structures for these two pentose sugars. Each carbon atom is distinguished by a number with a prime sign (e.g., C-1', C-2'). Compared with ribose, deoxyribose has a hydrogen atom rather than a hydroxyl group at the C-2' position. The absence of a hydroxyl group at the C-2' position thus distinguishes DNA from RNA. In

the absence of the C-2' hydroxyl group, the sugar is more specifically named **2-deoxyribose**.

If a molecule is composed of a purine or pyrimidine base and a ribose or deoxyribose sugar, the chemical unit is called a **nucleoside**. If a phosphate group is added to the nucleoside, the molecule is now called a **nucleotide**. Nucleosides and nucleotides are named according to the specific nitrogenous base (A, T, G, C, or U) that is part of the molecule. The structures of a nucleoside and a nucleotide and the nomenclature used in naming nucleosides and nucleotides are given in **Figure 10-8**.

The bonding between components of a nucleotide is highly specific. The C-1' atom of the sugar is involved in the chemical linkage to the nitrogenous base. If the base is a purine, the N-9 atom is covalently bonded to the sugar; if the base is a pyrimidine, the N-1 atom bonds to the sugar. In deoxyribonucleotides, the phosphate group may be bonded to the C-2', C-3', or C-5' atom of the sugar. The



**FIGURE 10-8** Structures and names of the nucleosides and nucleotides of RNA and DNA.

C-5' phosphate configuration is shown in Figure 10-8. It is by far the prevalent form in biological systems and the one found in DNA and RNA.

### Nucleoside Diphosphates and Triphosphates

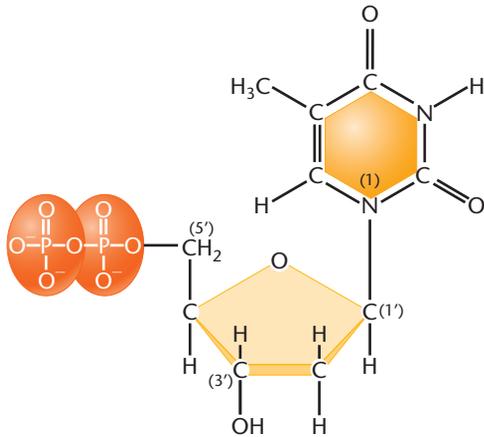
Nucleotides are also described by the term **nucleoside monophosphate (NMP)**. The addition of one or two phosphate groups results in **nucleoside diphosphates (NDPs)** and **triphosphates (NTPs)**, respectively, as shown in Figure 10-9. The triphosphate form is significant because it serves as the precursor molecule during nucleic acid synthesis within the cell (see Chapter 11). In addition, **adenosine triphosphate (ATP)** and **guanosine triphosphate (GTP)** are important in cell bioenergetics because of the large amount of energy involved in adding or removing the terminal phosphate group. The hydrolysis of ATP or GTP to ADP or GDP and inorganic phosphate ( $P_i$ ) is accompanied by the release of a large amount of energy in the cell. When these chemical conversions are coupled to other reactions, the energy produced is used to drive the reactions. As a result, both ATP and GTP are involved in many cellular activities, including numerous genetic events.

### Polynucleotides

The linkage between two mononucleotides consists of a phosphate group linked to two sugars. It is called a **phosphodiester bond** because phosphoric acid has been joined to two alcohols (the hydroxyl groups on the two sugars) by an ester linkage on both sides. Figure 10-10(a) shows the phosphodiester bond in DNA. The same bond is found in RNA. Each structure has a **C-5' end** and a **C-3' end**. Two joined nucleotides form a **dinucleotide**; three nucleotides, a **trinucleotide**; and so forth. Short chains consisting of up to approximately 30 nucleotides linked together are called **oligonucleotides**; longer chains are called **polynucleotides**.

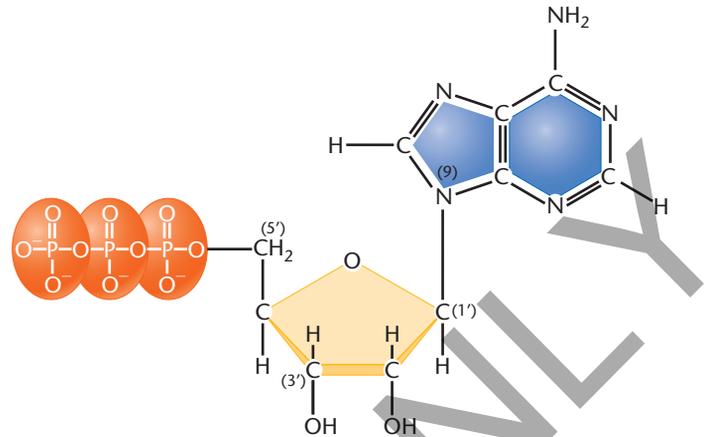
Because drawing polynucleotide structures, as shown in Figure 10-10(a), is time consuming and complex, a schematic shorthand method has been devised [Figure 10-10(b)]. The nearly vertical lines represent the pentose sugar; the nitrogenous base is attached at the top, in the C-1' position. A diagonal line with the P in the middle of it is attached to the C-3' position of one sugar and the C-5' position of the neighboring sugar; it represents the phosphodiester bond. Several modifications of this shorthand method are in use, and they can be understood in terms of these guidelines.

## Deoxynucleoside diphosphate (dNDP)



Deoxythymidine diphosphate (dTDP)

## Nucleoside triphosphate (NTP)



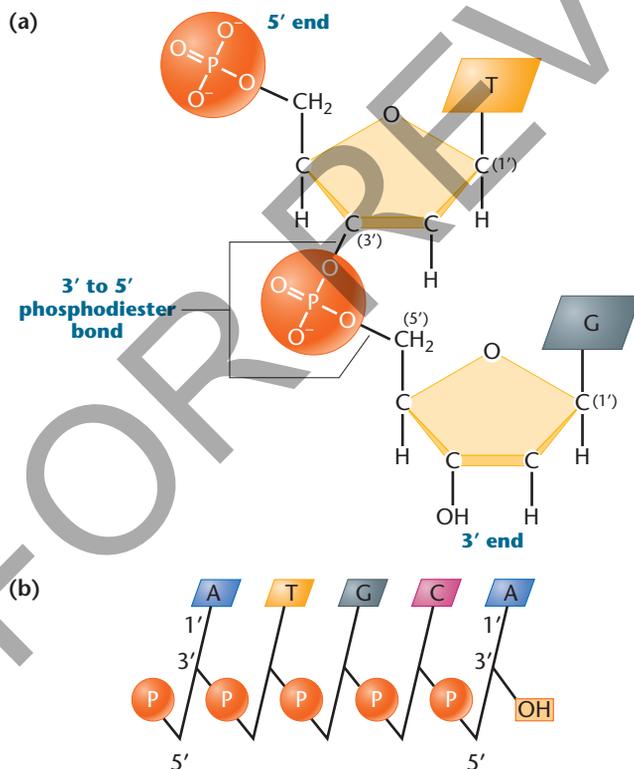
Adenosine triphosphate (ATP)

**FIGURE 10-9** Structures of nucleoside diphosphates and triphosphates. Deoxythymidine diphosphate and adenosine triphosphate are diagrammed here.

Although Levene's tetranucleotide hypothesis (described earlier in this chapter) was generally accepted before 1940, research in subsequent decades revealed it to be incorrect. It was shown that DNA does not necessarily contain equimolar quantities of the four bases. In addition, the molecular weight of DNA molecules was determined to be in the range

of  $10^6$  to  $10^9$  daltons, far in excess of that of a tetranucleotide. The current view of DNA is that it consists of exceedingly long polynucleotide chains.

Long polynucleotide chains account for the large molecular weight of DNA and explain its most important property—storage of vast quantities of genetic information. If each nucleotide position in this long chain can be occupied by any one of four nucleotides, extraordinary variation is possible. For example, a polynucleotide only 1000 nucleotides in length can be arranged  $4^{1000}$  different ways, each one different from all other possible sequences. This potential variation in molecular structure is essential if DNA is to store the vast amounts of chemical information necessary to direct cellular activities.



**FIGURE 10-10** (a) Linkage of two nucleotides by the formation of a C-3'-C-5' (3'-5') phosphodiester bond, producing a dinucleotide. (b) Shorthand notation for a polynucleotide chain.

## 10.7 The Structure of DNA Holds the Key to Understanding Its Function

The previous sections in this chapter have established that DNA is the genetic material in all organisms (with certain viruses being the exception) and have provided details as to the basic chemical components making up nucleic acids. What remained to be deciphered was the precise structure of DNA. That is, how are polynucleotide chains organized into DNA, which serves as the genetic material? Is DNA composed of a single chain or more than one? If the latter is the case, how do the chains relate chemically to one another? Do the chains branch? And more important, how does the structure of this molecule relate to the various genetic functions served by DNA (i.e., storage, expression, replication, and mutation)?

From 1940 to 1953, many scientists were interested in solving the structure of DNA. Among others, Erwin Chargaff, Maurice Wilkins, Rosalind Franklin, Linus Pauling, Francis Crick, and James Watson sought information that might answer what many consider to be the most significant and intriguing question in the history of biology: *How does DNA serve as the genetic basis for life?* The answer was believed to depend strongly on the chemical structure and organization of the DNA molecule, given the complex but orderly functions ascribed to it.

In 1953, James Watson and Francis Crick proposed that the structure of DNA is in the form of a double helix. Their model was described in a short paper published in the journal *Nature*. (The article is reprinted in its entirety on page 248.) In a sense, this publication was the finish of a highly competitive scientific race. Watson's book *The Double Helix* recounts the human side of the scientific drama that eventually led to the elucidation of DNA structure.

The data available to Watson and Crick, crucial to the development of their proposal, came primarily from two sources: (1) base composition analysis of hydrolyzed samples of DNA and (2) X-ray diffraction studies of DNA. Watson and Crick's analytical success can be attributed to their focus on building a model that conformed to the existing data. If the correct solution to the structure of DNA is viewed as a puzzle, Watson and Crick, working at the Cavendish Laboratory in Cambridge, England, were the first to fit the pieces together successfully.

Before learning the details of this far-reaching discovery, you may find some of the background on James Watson and Francis Crick to be of interest. Watson began his undergraduate studies at the University of Chicago at age 15, and was originally interested in ornithology. He then pursued his Ph.D. at Indiana University, where he studied viruses. He was only 24 years old in 1953, when he and Crick proposed the double-helix theory. Crick, now considered one of the great theoretical biologists of our time, had studied undergraduate physics at University College, London, and went on to perform military research during World War II. At the time of his collaboration with Watson, he was 35 years old and was performing X-ray diffraction studies of polypeptides and proteins as a graduate student. Immediately after they made their major discovery, Crick is reputed to have walked into the Eagle Pub in Cambridge, where the two frequently lunched, and announced for all to hear, "We have discovered the secret of life." It turns out that more than 50 years later, many scientists would quite agree!

### Base-Composition Studies

Between 1949 and 1953, Erwin Chargaff and his colleagues used chromatographic methods to separate the four bases in DNA samples from various organisms. Quantitative methods were then used to determine the amounts of the four bases from each source. **Table 10.3(a)** lists some of Chargaff's original data. Parts (b) and (c) of the table show

**TABLE 10.3** DNA Base-Composition Data

(a) Chargaff's Data*					(c) G + C Content in Several Organisms			
Organism/Source	Molar Proportions <sup>a</sup>				Organism	%G + C		
	1	2	3	4			7	8
	A	T	G	C				
Ox thymus	26	25	21	16	Phage T2	36.0		
Ox spleen	25	24	20	15	<i>Drosophila</i>	45.0		
Yeast	24	25	14	13	Maize	49.1		
Avian tubercle bacilli	12	11	28	26	<i>Euglena</i>	53.5		
Human sperm	29	31	18	18	<i>Neurospora</i>	53.7		

(b) Base Compositions of DNAs from Various Sources								
Organism	Base Composition				Base Ratio		Combined Base Ratios	
	1	2	3	4	5	6	7	8
	A	T	G	C	A/T	G/C	(A + G)/(C + T)	(A + T)/(C + G)
Human	30.9	29.4	19.9	19.8	1.05	1.00	1.04	1.52
Sea urchin	32.8	32.1	17.7	17.3	1.02	1.02	1.02	1.58
<i>E. coli</i>	24.7	23.6	26.0	25.7	1.04	1.01	1.03	0.93
<i>Sarcina lutea</i>	13.4	12.4	37.1	37.1	1.08	1.00	1.04	0.35
T7 bacteriophage	26.0	26.0	24.0	24.0	1.00	1.00	1.00	1.08

\*Source: Data from Chargaff, 1950.

<sup>a</sup>Moles of nitrogenous constituent per mole of P. (Often, the recovery was less than 100 percent.)

more recently derived base-composition information that reinforces Chargaff's findings. As we shall see, Chargaff's data were critical to the success of Watson and Crick as they devised the double-helical model of DNA. On the basis of these data, the following conclusions may be drawn:

1. As shown in **Table 10.3(b)**, the amount of adenine residues is proportional to the amount of thymine residues in DNA (columns 1, 2, and 5). Also, the amount of guanine residues is proportional to the amount of cytosine residues (columns 3, 4, and 6).
2. Based on this proportionality, the sum of the purines (A + G) equals the sum of the pyrimidines (C + T) as shown in column 7.
3. The percentage of (G + C) does not necessarily equal the percentage of (A + T). As you can see, this ratio varies greatly among organisms, as shown in column 8 and in part (c) of Table 10.3.

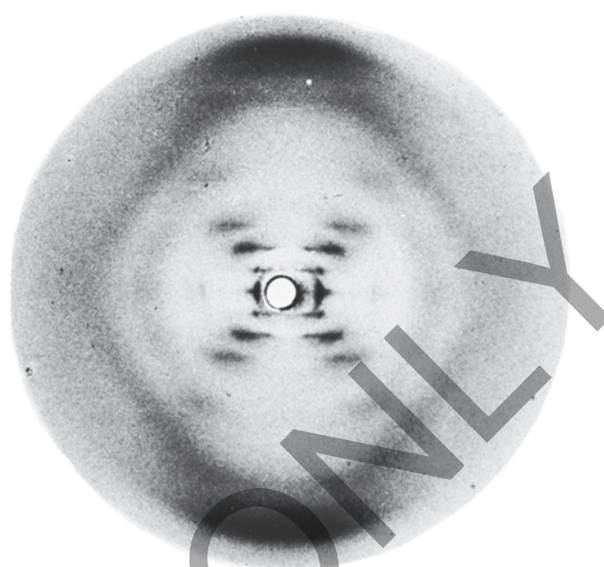
These conclusions indicate a definite pattern of base composition in DNA molecules. The data provided the initial clue to the DNA puzzle. In addition, they directly refute Levene's tetranucleotide hypothesis, which stated that all four bases are present in equal amounts.

### X-Ray Diffraction Analysis

When fibers of a DNA molecule are subjected to X-ray bombardment, the X rays scatter (diffract) in a pattern that depends on the molecule's atomic structure. The pattern of diffraction can be captured as spots on photographic film and analyzed for clues to the overall shape of and regularities within the molecule. This process, **X-ray diffraction analysis**, was applied successfully to the study of protein structure by Linus Pauling and other chemists. The technique had been attempted on DNA as early as 1938 by William Astbury. By 1947, he had detected a periodicity of 3.4-angstrom unit (3.4-Å) repetitions\* within the structure of the molecule, which suggested to him that the bases were stacked like coins on top of one another.

Between 1950 and 1953, Rosalind Franklin, working in the laboratory of Maurice Wilkins, obtained improved X-ray data from more purified samples of DNA (**Figure 10-11**). Her work confirmed the 3.4-Å periodicity seen by Astbury and suggested that the structure of DNA was some sort of helix. However, she did not propose a definitive model. Pauling had analyzed the work of Astbury and others and incorrectly proposed that DNA was a triple helix.

\*Today, measurement in nanometers (nm) is favored (1 nm = 10 Å).



**FIGURE 10-11** X-ray diffraction photograph by Rosalind Franklin using the B form of purified DNA fibers. The strong arcs on the periphery represent closely spaced aspects of the molecule, allowing scientists to estimate the periodicity of the nitrogenous bases, which are 3.4 Å apart. The inner cross pattern of spots reveals the grosser aspects of the molecule, indicating its helical nature.

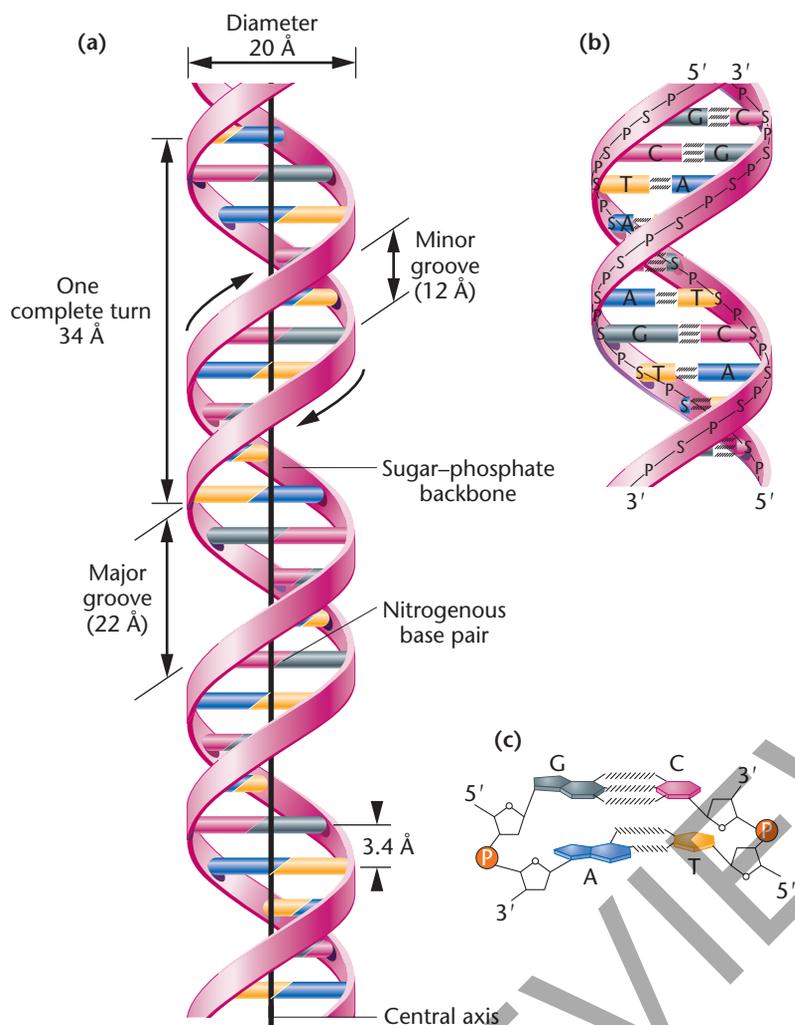
### The Watson-Crick Model

Watson and Crick published their analysis of DNA structure in 1953. By building models based on the above-mentioned parameters, they arrived at the double-helical form of DNA shown in **Figure 10-12(a)** on p. 246.

This model has the following major features:

1. Two long polynucleotide chains are coiled around a central axis, forming a right-handed double helix.
2. The two chains are **antiparallel**; that is, their C-5'-to-C-3' orientations run in opposite directions.
3. The bases of both chains are flat structures lying perpendicular to the axis; they are "stacked" on one another, 3.4 Å (0.34 nm) apart, on the inside of the double helix.
4. The nitrogenous bases of opposite chains are *paired* as the result of the formation of hydrogen bonds; in DNA, only A = T and G ≡ C pairs occur.
5. Each complete turn of the helix is 34 Å (3.4 nm) long; thus, each turn of the helix is the length of a series of 10 base pairs.
6. A larger **major groove** alternating with a smaller **minor groove** winds along the length of the molecule.
7. The double helix has a diameter of 20 Å (2.0 nm).

The nature of the base pairing (point 4 above) is the model's most significant feature in terms of explaining



**FIGURE 10-12** (a) The DNA double helix as proposed by Watson and Crick. The ribbon-like strands represent the sugar-phosphate backbones, and the horizontal rungs depict the nitrogenous base pairs, of which there are 10 per complete turn. The major and minor grooves are apparent. A solid vertical line shows the central axis. (b) A detailed view depicting the bases, sugars, phosphates, and hydrogen bonds of the helix. (c) A demonstration of the antiparallel arrangement of the chains and the horizontal stacking of the bases.

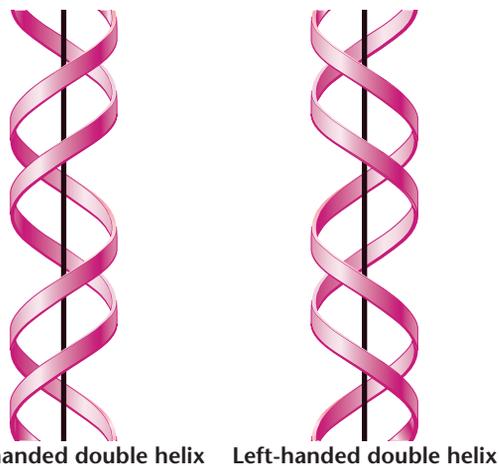
its genetic functions. Before we discuss it, however, several other important features warrant emphasis. First, the antiparallel arrangement of the two chains is a key part of the double-helix model. While one chain runs in the 5'-to-3' orientation (what seems right side up to us), the other chain goes in the 3'-to-5' orientation (and thus appears upside down). This is indicated in **Figure 10-12(b)** and **(c)**. Given the bond angles in the structures of the various nucleotide components, the double helix could not be constructed easily if both chains ran parallel to one another.

Second, the right-handed nature of the helix modeled by Watson and Crick is best appreciated by comparison with its left-handed counterpart, which is a mirror image,

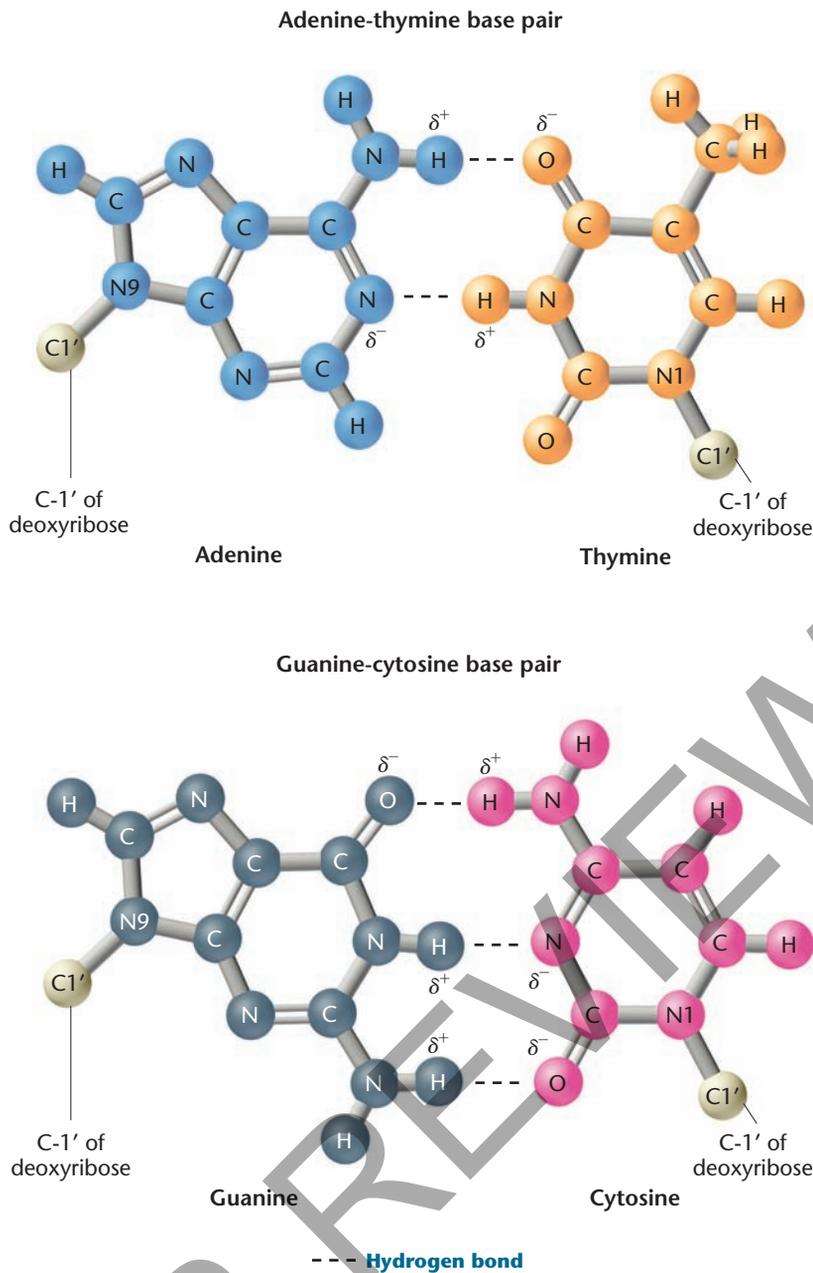
as shown in **Figure 10-13**. The conformation in space of the right-handed helix is most consistent with the data that were available to Watson and Crick, although an alternative form of DNA (Z-DNA) does exist as a left-handed helix, as we discuss below.

The key to the model proposed by Watson and Crick is the specificity of base pairing. Chargaff's data suggested that A was equal in amount to T and that G was equal in amount to C. Watson and Crick realized that pairing A with T and C with G would account for these proportions, and that such pairing could occur as a result of hydrogen bonds between base pairs [Figure 10-12(b)], which would also provide the chemical stability necessary to hold the two chains together. Arrangement of the components in this way produces the major and minor grooves along the molecule's length. Furthermore, a purine (A or G) opposite a pyrimidine (T or C) on each "rung of the spiral staircase" in the proposed helix accounts for the 20-Å (2-nm) diameter suggested by X-ray diffraction studies.

The specific A=T and G=C base pairing is described as **complementarity** and results from the chemical affinity that produces the hydrogen bonds in each pair of bases. As we will see, complementarity is very important in the processes of DNA replication and gene expression.



**FIGURE 10-13** The right- and left-handed helical forms of DNA. Note that they are mirror images of one another.



**FIGURE 10-14** Ball-and-stick models of A=T and G=C base pairs. The dashes (---) represent the hydrogen bonds that form between bases.

Two questions are particularly worthy of discussion. First, why aren't other base pairs possible? Watson and Crick discounted the pairing of A with G or of C with T because these would represent purine-purine and pyrimidine-pyrimidine pairings, respectively. Such pairings would lead to aberrant diameters of, in one case, more than and, in the other case, less than 20 Å because of the respective sizes of the purine and pyrimidine rings. In addition, the three-dimensional configurations that would be formed by such pairings would not produce an alignment that allows sufficient hydrogen-bond formation. It is for this reason that A=C and G=T pairings

were also discounted, even though those pairs would each consist of one purine and one pyrimidine.

The second question concerns the properties of hydrogen bonds. Just what is the nature of such a bond, and is it strong enough to stabilize the helix? A **hydrogen bond** is a very weak electrostatic attraction between a covalently bonded hydrogen atom and an atom with an unshared electron pair. The hydrogen atom assumes a partial positive charge, while the unshared electron pair—characteristic of covalently bonded oxygen and nitrogen atoms—assumes a partial negative charge. These opposite charges are responsible for the weak chemical attraction that is the basis of the hydrogen bond. As oriented in the double helix, adenine forms two hydrogen bonds with thymine, and guanine forms three hydrogen bonds with cytosine (**Figure 10-14**). Although two or three hydrogen bonds taken alone are energetically very weak, thousands of bonds in tandem (as found in long polynucleotide chains) provide great stability to the helix.

Another stabilizing factor is the arrangement of sugars and bases along the axis. In the Watson-Crick model, the hydrophobic (“water-fearing”) nitrogenous bases are stacked almost horizontally on the interior of the axis and are thus shielded from the watery environment that surrounds the molecule within the cell. The hydrophilic (“water-loving”) sugar-phosphate backbones are on the outside of the axis, where both components may interact with water. These molecular arrangements provide significant chemical stabilization to the helix.

A more recent and accurate analysis of the form of DNA that served as the basis for the Watson-Crick model has revealed a minor structural difference between the substance and the model. A precise measurement of the number of base pairs per turn has demonstrated a value of 10.4, rather than the 10.0 predicted by Watson and Crick. In the classic model, each base pair is rotated 36° around the helical axis relative to the adjacent base pair, but the new finding requires a rotation of 34.6°. This results in slightly more than 10 base pairs per 360° turn.

The Watson-Crick model had an instant effect on the emerging discipline of molecular biology. Even in

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

We wish to suggest a structure for the salt of deoxyribose nucleic acid (DNA). This structure has novel features which are of considerable biological interest. A structure for nucleic acid has already been proposed by Pauling and Corey.<sup>1</sup> 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. We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining  $\alpha$ -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<sup>2</sup> 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 Å 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 Å the distance of a phosphorus atom from the fibre axis is 10 Å. As the phosphates are on the outside, cations have easy access to them.

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-coordinates. 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<sup>3,4</sup> 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<sup>5,6</sup> 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 stereochemical 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 co-workers 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, England.

<sup>1</sup>Pauling, L., and Corey, R. B., *Nature*, 171, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, 39, 84 (1953).

<sup>2</sup>Furberg, S., *Acta Chem. Scand.*, 6, 634 (1952).

<sup>3</sup>Chargaff, E. For references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, 9, 402 (1952).

<sup>4</sup>Wyatt, G. R., *J. Gen. Physiol.*, 36, 201 (1952).

<sup>5</sup>Astbury, W. T., *Symp. Soc. Exp. Biol. 1, Nucleic Acid*, 66 (Camb. Univ. Press, 1947).

<sup>6</sup>Wilkins, M.H.F., and Randall, J. T., *Biochim. et Biophys. Acta*, 10, 192 (1953).

Source: Reprinted by permission from Macmillan Publishers Ltd: *Nature*, "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid" by F.H.C. Crick and J.D. Watson, *Nature*, Vol. 171, No. 4356, pp. 737-38. Copyright 1953. www.nature.com

their initial 1953 article in *Nature*, the authors observed, “It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.” Two months later, Watson and Crick pursued this idea in a second article in *Nature*, suggesting a specific mechanism of replication of DNA—the *semiconservative mode of replication* (described in Chapter 11). The second article also alluded to two new concepts: (1) the storage of genetic information in the sequence of the bases and (2) the mutations or genetic changes that would result from an alteration of bases. These ideas have received vast amounts of experimental support since 1953 and are now universally accepted.

Watson and Crick’s synthesis of ideas was highly significant with regard to subsequent studies of genetics and biology. The nature of the gene and its role in genetic mechanisms could now be viewed and studied in biochemical terms. Recognition of their work, along with that of Wilkins, led to all three receiving the Nobel Prize in Physiology and Medicine in 1962. Unfortunately, Rosalind Franklin had died in 1958 at the age of 37, making her contributions ineligible for consideration, since the award is not given posthumously. The Nobel Prize was to be one of many such awards bestowed for work in the field of molecular genetics.

#### EVOLVING CONCEPT OF THE GENE

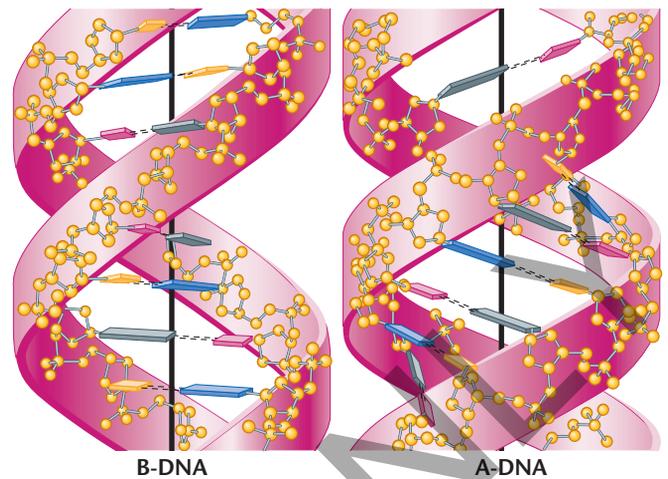
Based on the model of DNA put forward by Watson and Crick in 1953, the gene was viewed for the first time in molecular terms as a sequence of nucleotides in a DNA helix that encodes genetic information. ■

## 10.8 Alternative Forms of DNA Exist

Under different conditions of isolation, different conformations of DNA are seen. At the time when Watson and Crick performed their analysis, two forms—**A-DNA** and **B-DNA**—were known. Watson and Crick’s analysis was based on Rosalind Franklin’s X-ray studies of the B form, which is seen under aqueous, low-salt conditions and is believed to be the biologically significant conformation.

While DNA studies around 1950 relied on the use of X-ray diffraction, more recent investigations have been performed using **single-crystal X-ray analysis**. The earlier studies achieved resolution of about 5 Å, but single crystals diffract X rays at about 1-Å intervals, near atomic resolution. As a result, every atom is “visible,” and much greater structural detail is available.

With this modern technique, A-DNA, which is prevalent under high-salt or dehydration conditions, has now been



**FIGURE 10-15** An artist’s depiction illustrating the orientation of the base pairs of B-DNA and A-DNA. Note that in B-DNA the base pairs are perpendicular to the helix, while they are tilted and pulled away from the helix in A-DNA. In comparison to B-DNA, A-DNA also displays a slightly greater diameter and contains one more base pair per turn.

scrutinized. In comparison to B-DNA, A-DNA is slightly more compact, with 9 base pairs in each complete turn of the helix, which is 23 Å (2.3 nm) in diameter (**Figure 10-15**). While it is also a right-handed helix, the orientation of the bases is somewhat different—they are tilted and displaced laterally in relation to the axis of the helix. As a result, the appearance of the major and minor grooves is modified. It seems doubtful that A-DNA occurs *in vivo* (under physiological conditions).

Still other forms of DNA right-handed helices have been discovered when investigated under various laboratory conditions. These have been designated C-, D-, E-, and most recently P-DNA. **C-DNA** is found under even greater dehydration conditions than those observed during the isolation of A- and B-DNA. Two other forms, **D-DNA** and **E-DNA**, occur in helices lacking guanine in their base composition. And most recently, it has been observed that if DNA is artificially stretched, still another conformation is assumed, called **P-DNA** (named for Linus Pauling).

Finally, still another form of DNA, called **Z-DNA**, was discovered in 1979, when a small synthetic DNA oligonucleotide containing only G≡C base pairs was studied. Z-DNA takes on the rather remarkable configuration of a *left-handed double helix*. Speculation has occasionally run high over the possibility that regions of Z-DNA exist in the chromosomes of living organisms. The unique helical arrangement has the potential to provide an important recognition site for interaction with DNA-binding molecules. However, the extent to which Z-DNA, or any of the forms mentioned above, occur *in vivo* is still not clear.

## NOW SOLVE THIS

**10-2** In sea urchin DNA, which is double stranded, 17.5 percent of the bases were shown to be cytosine (C). What percentages of the other three bases are expected to be present in this DNA?

■ **HINT:** This problem asks you to extrapolate from one measurement involving a unique DNA molecule to three other values characterizing the molecule. The key to its solution is to understand the base-pairing rules in the Watson–Crick model of DNA.

## 10.9 The Structure of RNA Is Chemically Similar to DNA, but Single Stranded

The structure of RNA molecules resembles DNA, with several important exceptions. Although RNA also has nucleotides linked into polynucleotide chains, the sugar ribose replaces deoxyribose, and the nitrogenous base uracil replaces thymine. Another important difference is that most RNA is single stranded, although there are two important exceptions. First, RNA molecules sometimes fold back on themselves to form double-stranded regions of complementary base pairs. Second, some animal viruses that have RNA as their genetic material contain double-stranded helices.

As established earlier (see Figure 10-1), three major classes of cellular RNA molecules function during the expression of genetic information: **ribosomal RNA (rRNA)**, **messenger RNA (mRNA)**, and **transfer RNA (tRNA)**. These molecules all originate as complementary copies of one of the two strands of DNA segments during the process of transcription. That is, their nucleotide sequence is complementary to the deoxyribonucleotide sequence of DNA that served as the template for their synthesis. Because uracil replaces thymine in RNA, uracil is complementary to adenine during transcription and during RNA base pairing.

**Table 10.4** characterizes these major forms of RNA as found in prokaryotic and eukaryotic cells. Different RNAs are distinguished according to their sedimentation behavior in a centrifugal field and by their size (the number of nucleotides each contains). Sedimentation behavior depends on a molecule's density, mass, and shape, and its measure is called the **Svedberg coefficient (S)**. While higher *S* values almost always designate molecules of greater molecular weight, the correlation is not direct; that is, a twofold increase in molecular weight does not lead to a twofold increase in *S*. This is because,

**TABLE 10.4** Principal Classes of RNA

Class	% of Total RNA*	Svedberg Coefficient	Eukaryotic (E) or Prokaryotic (P)	Number of Nucleotides
Ribosomal (rRNA)	80	5	P and E	120
		5.8	E	160
		16	P	1542
		18	E	1874
		23	P	2904
Transfer (tRNA)	15	4	P and E	75-90
		28	E	4718
Messenger (mRNA)	5	varies	P and E	100-10,000

\*In *E. coli*.

in addition to a molecule's mass, the size and shape of the molecule also affect its rate of sedimentation (*S*). As you can see in Table 10.4, RNA molecules come in a wide range of sizes.

Ribosomal RNA usually constitutes about 80 percent of all RNA in an *E. coli* cell. Ribosomal RNAs are important structural components of **ribosomes**, which function as nonspecific workbenches where proteins are synthesized during translation. The various forms of rRNA found in prokaryotes and eukaryotes differ distinctly in size.

Messenger RNA molecules carry genetic information from the DNA of the gene to the ribosome. The mRNA molecules vary considerably in size, reflecting the range in the sizes of the proteins encoded by the mRNA as well as the different sizes of the genes serving as the templates for transcription of mRNA. While Table 10.4 shows that about 5 percent of RNA is mRNA in *E. coli*, this percentage varies from cell to cell and even at different times in the life of the same cell.

Transfer RNA, accounting for up to 15 percent of the RNA in a typical cell, is the smallest class (in terms of average molecule size) of these RNA molecules and carries amino acids to the ribosome during translation. Because more than one tRNA molecule interacts simultaneously with the ribosome, the molecule's smaller size facilitates these interactions.

We will discuss more detailed functions of these three classes of RNA later in the text (see Chapters 13 and 14). While these RNAs represent the major forms of the molecule involved in genetic expression, other unique RNAs exist that perform various roles, especially in eukaryotes. For example, **telomerase RNA** and **RNA primers** are involved in DNA replication at the ends of chromosomes

(Chapter 11), **small nuclear RNA (snRNA)** participates in processing mRNAs (Chapter 13), and **antisense RNA**, **microRNA (miRNA)**, and **short interfering RNA (siRNA)** are involved in gene regulation (Chapter 17). We will also discuss still other recent findings involving RNA later in the text (see Special Topic Chapter 2—Emerging Roles of RNA).

### NOW SOLVE THIS

**10-3** German measles results from an infection of the rubella virus, which can cause a multitude of health problems in newborns. What conclusions can you reach from a viral nucleic acid analysis that reveals an  $A + G/U + C$  ratio of 1.13?

■ **HINT:** This problem asks you to analyze information about the chemical composition of a nucleic acid serving as the genetic material of a virus. The key to its solution is to apply your knowledge of nucleic acid chemistry, in particular your understanding of base-pairing.

For more practice, see Problems 32 and 33.

## 10.10 Many Analytical Techniques Have Been Useful during the Investigation of DNA and RNA

Since 1953, the role of DNA as the genetic material and the role of RNA in transcription and translation have been clarified through detailed analysis of nucleic acids. In this section, we consider several fundamental methods that have been particularly important for accomplishing this analysis. Many of them make use of the unique nature of the hydrogen bond that is so integral to the structure of nucleic acids. Later in the text (see Chapter 20), we will consider still other techniques that have been developed for manipulating DNA in various ways, including cloning.

### Absorption of Ultraviolet Light

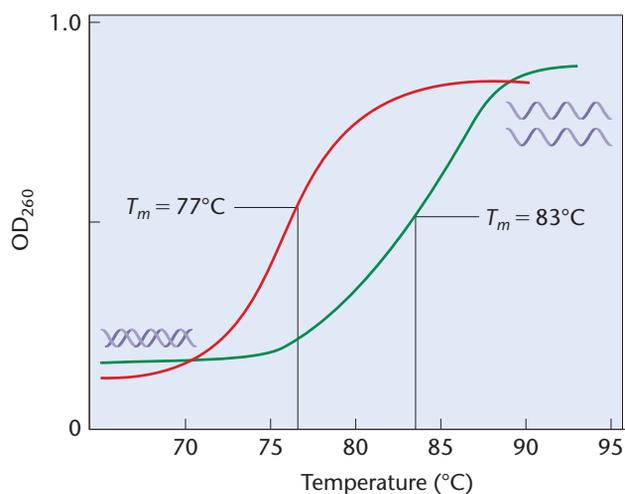
Nucleic acids absorb ultraviolet (UV) light most strongly at wavelengths of 254 to 260 nm (Figure 10-6), due to the interaction between UV light and the ring systems of the purines and pyrimidines. In aqueous solution, peak absorption by DNA and RNA occurs at 260 nm. Thus, UV light can be used in the localization, isolation, and characterization of molecules that contain nitrogenous bases (i.e., nucleosides, nucleotides, and polynucleotides).

Ultraviolet analysis is used in conjunction with many standard procedures that separate and identify molecules. As we shall see in the next section, the use of UV absorption is critical to the isolation of nucleic acids following their separation.

### Denaturation and Renaturation of Nucleic Acids

Heat or other stresses can cause a complex molecule like DNA to **denature**, or lose its function due to the unfolding of its three-dimensional structure. In the denaturation of double-stranded DNA, the hydrogen bonds of the duplex (double-stranded) structure break, the helix unwinds, and the strands separate. However, no covalent bonds break. The viscosity of the DNA decreases, and both the UV absorption and the buoyant density increase. In laboratory studies, denaturation may be caused intentionally either by heating or chemical treatment. (Denaturation as a result of heating is sometimes referred to as *melting*.) In such studies, the increase in UV absorption of heated DNA in solution, called the **hyperchromic shift**, is the easiest change to measure. This effect is illustrated in **Figure 10-16**.

Because  $G \equiv C$  base pairs have one more hydrogen bond than do  $A = T$  pairs (see Figure 10-14), they are more stable under heat treatment. Thus, DNA with a greater proportion of  $G \equiv C$  pairs than  $A = T$  pairs requires higher temperatures to denature completely. When absorption at 260 nm is monitored and plotted against temperature during heating, a **melting profile** of the DNA is obtained. The



**FIGURE 10-16** A melting profile shows the increase in UV absorption versus temperature (the hyperchromic effect) for two DNA molecules with different  $G \equiv C$  contents. The molecule with a melting point ( $T_m$ ) of 83°C has a greater  $G \equiv C$  content than the molecule with a  $T_m$  of 77°C.

midpoint of this profile, or curve, is called the **melting temperature** ( $T_m$ ) and represents the point at which 50 percent of the strands are unwound, or denatured (Figure 10–16). When the curve plateaus at its maximum optical density, denaturation is complete, and only single strands exist. Analysis of melting profiles provides a characterization of DNA and an alternative method of estimating its base composition.

One might ask whether the denaturation process can be reversed; that is, can single strands of nucleic acids reform a double helix, provided that each strand's complement is present? Not only is the answer yes, but such reassociation provides the basis for several important analytical techniques that have contributed much valuable information during genetic experimentation.

If DNA that has been denatured thermally is cooled slowly, random collisions between complementary strands will result in their reassociation. At the proper temperature, hydrogen bonds will re-form, securing pairs of strands into duplex structures. With time during cooling, more and more duplexes will form. Depending on the conditions, a complete match (that is, perfect complementarity) is not essential for duplex formation, provided there are stretches of base pairing on the two reassociating strands.

## Molecular Hybridization

The denaturation and renaturation of nucleic acids are the basis for one of the most powerful and useful techniques in molecular genetics—**molecular hybridization**. This technique derives its name from the fact that single strands need not originate from the same nucleic acid source in order to combine to form duplex structures. For example, if DNA strands are isolated from two distinct organisms and a reasonable degree of base complementarity exists between them, under the proper temperature conditions, double-stranded molecular hybrids will form during renaturation. Such hybridization may also occur between single strands of DNA and RNA. A case in point is when RNA and the DNA from which it has been transcribed are mixed together (Figure 10–17). The RNA will find and attach to its single-stranded DNA complement. In this example, DNA molecules are heated, causing strand separation, and then the strands are slowly cooled in the presence of single-stranded RNA. If the RNA has been transcribed from the DNA used in the experiment, and is therefore complementary to it, molecular hybridization will occur, creating a DNA:RNA duplex. Several methods are available for monitoring the amount of double-stranded molecules produced following strand separation. In early studies,

radioisotopes were utilized to “tag” one of the strands and monitor its presence in the hybrid duplexes that formed.

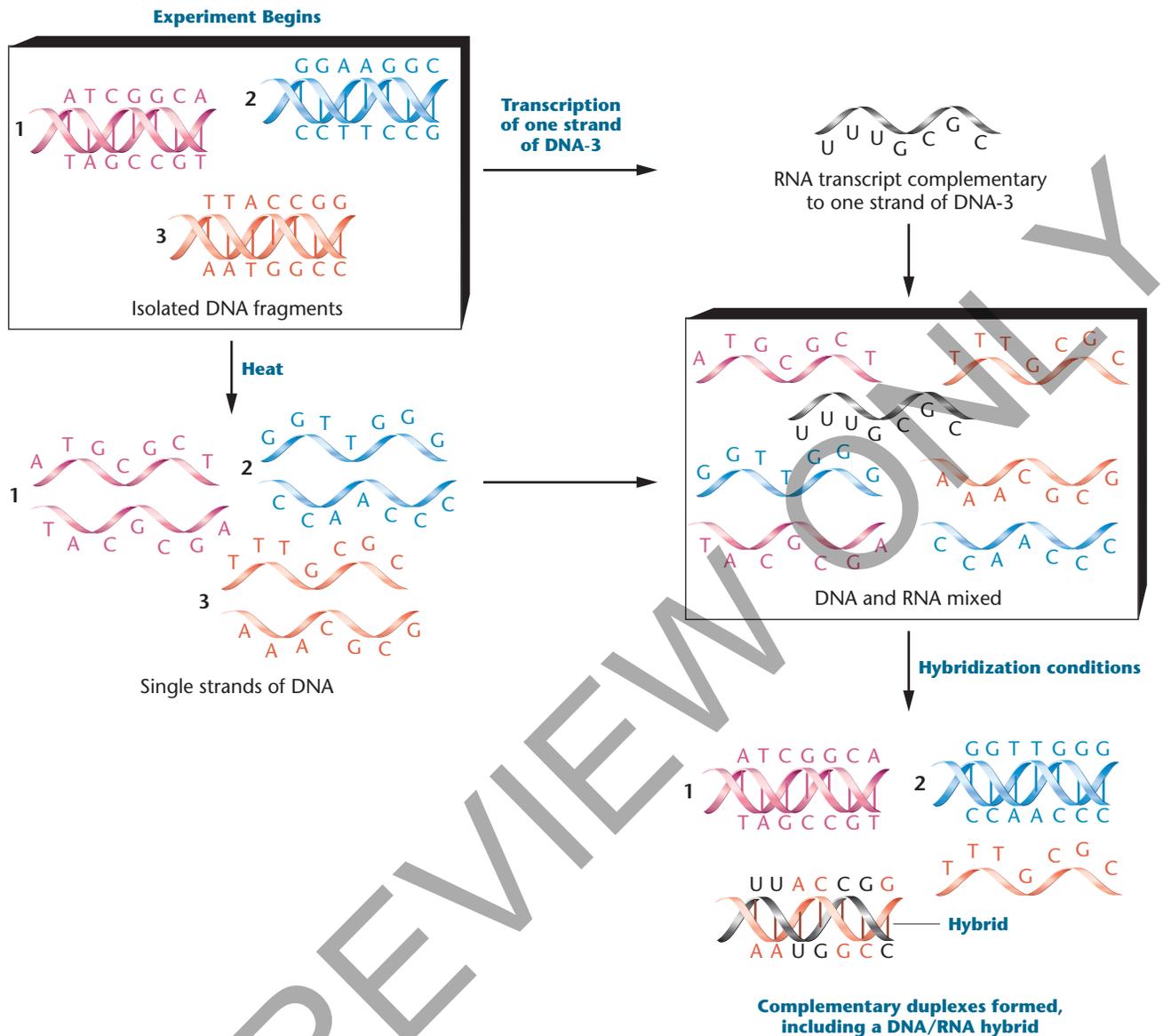
In the 1960s, molecular hybridization techniques contributed to our increased understanding of transcriptional events occurring at the gene level. Refinements of this process have occurred continually, helping to advance the study of both molecular evolution and the organization of DNA in chromosomes. Hybridization can occur in solution or when DNA is bound either to a gel or to a specialized binding filter. Such filters are used in a variety of **DNA blotting procedures**, whereby hybridization serves as a way to “probe” for complementary nucleic acid sequences. Blotting is used routinely in modern genomic analysis. In addition, hybridization will occur even when DNA is present in tissue affixed to a slide, as in the FISH procedure (discussed in the next section), or when affixed to a glass chip, the basis of **DNA microarray analysis** (discussed in Chapter 21). Microarray analysis allows mass screening for a specific DNA sequence from among thousands of cloned genes in a single assay.

## Fluorescent *in situ* Hybridization (FISH)

A refinement in the molecular hybridization technique has led to the use of DNA present in cytological preparations as the “target” for hybrid formation. When this approach is combined with the use of fluorescent probes to monitor hybridization, the technique is called **fluorescent *in situ* hybridization**, or is known simply by the acronym **FISH**. In this procedure, mitotic or interphase cells are fixed to slides and subjected to hybridization conditions. Single-stranded DNA or RNA is added, and hybridization is monitored. The added nucleic acid serves as a “probe,” since it will hybridize only with the specific chromosomal areas for which it is sufficiently complementary. Before the use of fluorescent probes was refined, radioactive probes were used in these *in situ* procedures to allow detection on the slide by autoradiography.

Fluorescent probes are prepared in a unique way. When DNA is used, it is first coupled to the small organic molecule biotin (creating biotinylated DNA). Another molecule (avidin or streptavidin) that has a high binding affinity for biotin is linked with a fluorescent molecule such as fluorescein, and this complex is reacted with the cytological preparation after the *in situ* hybridization has occurred. This procedure represents an extremely sensitive method for localizing the hybridized DNA.

**Figure 10–18** on p. 253 illustrates the results of using FISH to identify the DNA specific to the centromeres of human chromosomes. The resolution of FISH is great



**FIGURE 10-17** Diagrammatic representation of the process of molecular hybridization between DNA fragments and RNA that has been transcribed on one strand of one of the DNA fragments (DNA-3).

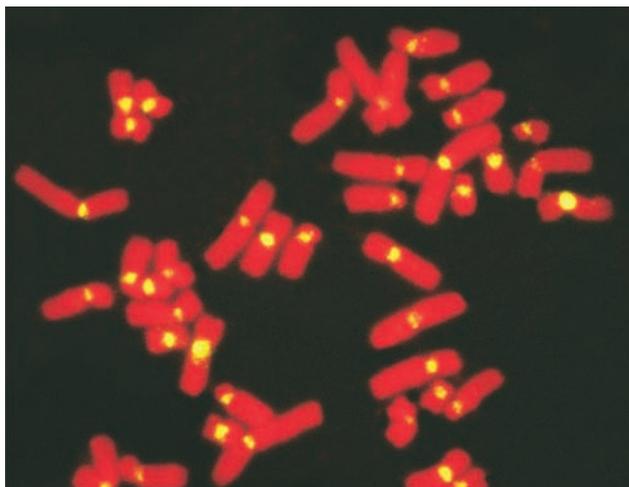
enough to detect just a single gene within an entire set of chromosomes. The use of this technique to identify chromosomal locations of specific genetic information has been a valuable addition to the repertoire of experimental geneticists.

### Reassociation Kinetics and Repetitive DNA

In still another extension of molecular hybridization procedures, the *rate of reassociation* of complementary single DNA strands may be analyzed. This technique, **reassociation kinetics**, was first refined and studied by Roy Britten and David Kohne.

The DNA used in such studies is first fragmented into small pieces by shearing forces introduced during its isolation. The resultant DNA fragments have an average size of several hundred base pairs. The fragments are then dissociated into single strands by heating (denatured), and when the temperature is lowered, reassociation is monitored. During reassociation, pieces of single-stranded DNA randomly collide. If they are complementary, a stable double strand is formed; if not, they separate and are free to encounter other DNA fragments. The process continues until all possible matches are made.

A great deal of information can be obtained from studies that compare the reassociation of DNA of different organisms.



**FIGURE 10-18** Fluorescent *in situ* hybridization (FISH) of human metaphase chromosomes. The probe, specific to centromeric DNA, produces a yellow fluorescence signal indicating hybridization. The red fluorescence is produced by propidium iodide counterstaining of chromosomal DNA.

As genome size increases and there is more DNA, reassociation time is extended. Reassociation occurs more slowly in larger genomes because with random collisions it takes more time for all correct matches to be made. When reassociation kinetics in eukaryotic organisms with much larger genome sizes was first studied, a surprising observation was made. Rather than requiring an extended reassociation time, *some* eukaryotic DNA reassociated even more rapidly than those derived from bacteria. The remaining DNA, as expected because of its complexity, took longer to reassociate.

Based on this observation, Britten and Kohne hypothesized that the rapidly reassociating fraction might represent **repetitive DNA sequences**. This interpretation would explain why these segments reassociate so rapidly—multiple copies of the same sequence are much more likely to make matches, thus reassociating more quickly than single copies. On the other hand, the remaining DNA segments consist of **unique DNA sequences**, present only once in the genome.

It is now known that repetitive DNA is prevalent in eukaryotic genomes and is key to our understanding of how genetic information is organized in chromosomes. Careful study has shown that various levels of repetition exist. In some cases, short DNA sequences are repeated over a million times. In other cases, longer sequences are repeated only a few times, or intermediate levels of sequence redundancy are present. We will return to this topic later in the text (see Chapter 12), where we will discuss the organization of DNA in genes and chromosomes. For now, we will simply point out that the discovery of repetitive DNA was one of the first clues that much of the DNA in eukaryotes

is not contained in genes that encode proteins. We will develop and elaborate on this concept as we proceed with our coverage of the molecular basis of heredity.

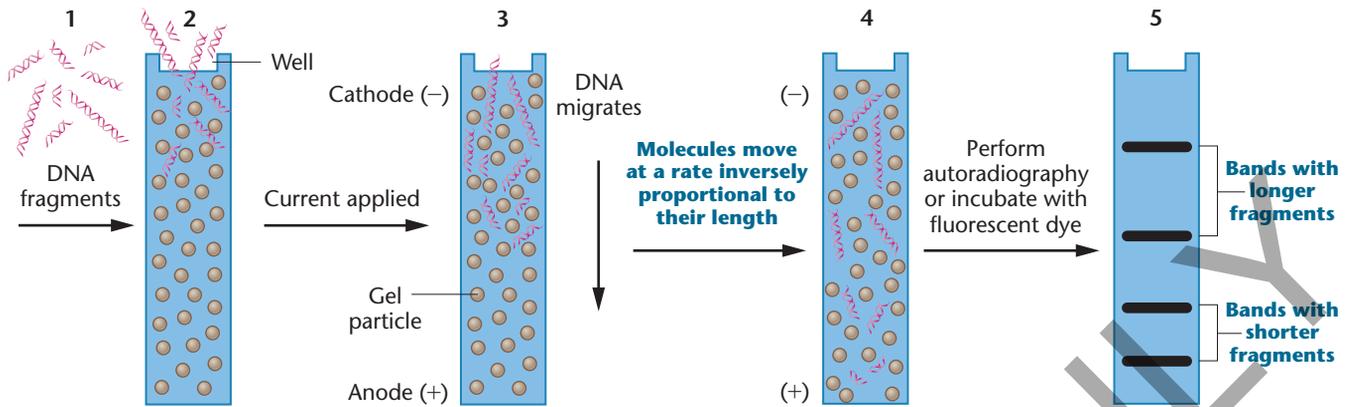
## Electrophoresis of Nucleic Acids

We conclude the chapter by considering **electrophoresis**, a technique that has made essential contributions to the analysis of nucleic acids. Electrophoresis, also useful in protein studies, can be applied to the separation of different-sized fragments of DNA and RNA chains and is invaluable in current research investigations in molecular genetics.

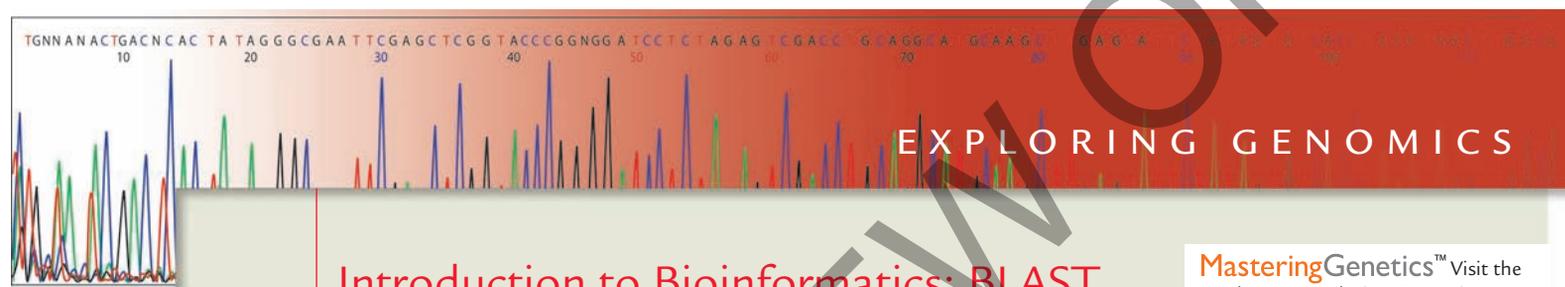
In general, electrophoresis separates, or *resolves*, molecules in a mixture by causing them to migrate under the influence of an electric field. A sample to be analyzed is placed on a porous substance (a piece of filter paper or a semisolid gel) that in turn is placed in a solution that conducts electricity. If two molecules have approximately the same shape and mass, the one with the greatest net charge will migrate more rapidly toward the electrode of opposite polarity.

As electrophoretic technology developed from its initial application (which was protein separation), researchers discovered that using gels of varying pore sizes significantly improved the method's resolution. This advance is particularly useful for mixtures of molecules with a similar charge-mass ratio but different sizes. For example, two polynucleotide chains of different lengths (say, 10 vs. 20 nucleotides) are both negatively charged because of the phosphate groups of the nucleotides. Thus, both chains move to the positively charged pole (the anode), and because they have the same charge-mass ratio, the electric field moves them at similar speeds. Consequently, the separation between the two chains is minimal. However, using a porous medium such as an **agarose gel**, which can be prepared with various pore sizes, enables us to separate the two molecules. *Smaller molecules migrate at a faster rate through the gel than larger molecules* (Figure 10-19). The key to separation is the porous gel matrix, which restricts migration of larger molecules more than it restricts smaller molecules. The resolving power of this method is so great that polynucleotides that vary by even one nucleotide in length may be separated. Once electrophoresis is complete, bands representing the variously sized molecules are currently identified by use of a fluorescent dye that binds to nucleic acids.

Electrophoretic separation of nucleic acids is at the heart of a variety of commonly used research techniques discussed later in the text (Chapters 20 and 21). Of particular note, as you will see in those discussions, are the various “blotting” techniques (e.g., Southern blots and Northern blots), as well as DNA sequencing methods.



**FIGURE 10-19** Electrophoretic separation of a mixture of DNA fragments that vary in length.



## EXPLORING GENOMICS

### Introduction to Bioinformatics: BLAST

MasteringGenetics™ Visit the Study Area: Exploring Genomics

In this chapter, we focused on the structural details of DNA, the genetic material for living organisms. In Chapter 20, you will learn how scientists can clone and sequence DNA—a routine technique in molecular biology and genetics laboratories. The explosion of DNA and protein sequence data that has occurred in the last 15 years has launched the field of *bioinformatics*, an interdisciplinary science that applies mathematics and computing technology to develop hardware and software for storing, sharing, comparing, and analyzing nucleic acid and protein sequence data.

A large number of sequence databases that make use of bioinformatics have been developed. An example is **GenBank** (<http://www.ncbi.nlm.nih.gov/genbank/>), which is the National Institutes of Health sequence database. This global resource, with access to databases in Europe and Japan, currently contains more than 152 billion base pairs of sequence data!

In the Exploring Genomics exercises for Chapter 5, you were introduced to the National Center for Biotechnology Information (NCBI) Genes and Disease site. Now we will use an

NCBI application called **BLAST, Basic Local Alignment Search Tool**. BLAST is an invaluable program for searching through GenBank and other databases to find DNA- and protein-sequence similarities between cloned substances. It has many additional functions that we will explore in other exercises.

#### ■ Exercise I – Introduction to BLAST

1. Access BLAST from the NCBI Web site at <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.

2. Click on “nucleotide blast.” This feature allows you to search DNA databases to look for a similarity between a sequence you enter and other sequences in the database. Do a nucleotide search with the following sequence:

```
CCAGAGTCCAGCTGCTGCTCATA
CTACTGATACTGCTGGG
```

3. Imagine that this sequence is a short part of a gene you cloned in your laboratory. You want to know if this gene or others with similar sequences have been discovered. Enter this sequence into the “Enter Query Sequence” text box at the top of the page. Near the

bottom of the page, under the “Program Selection” category, choose “blastn”; then click on the “BLAST” button at the bottom of the page to run the search. It may take several minutes for results to be available because BLAST is using powerful algorithms to scroll through billions of bases of sequence data! A new page will appear with the results of your search.

4. On the search results page, below the Graphic Summary you will see a category called Descriptions and a table showing significant matches to the sequence you searched with (called the query sequence). BLAST determines significant matches based on statistical measures that consider the length of the query sequence, the number of matches with sequences in the database, and other factors. Significant *alignments*, regions of significant similarity in the query and subject sequences, typically have E values less than 1.0.

5. The top part of the table lists matches to transcripts (mRNA sequences), and the lower part lists matches to genomic DNA sequences, in order of highest to lowest number of matches.

(continued)

## Exploring Genomics—continued

6. Alignments are indicated by horizontal lines. BLAST adjusts for gaps in the sequences, that is, for areas that may not align precisely because of missing bases in otherwise similar sequences. Scroll below the table to see the aligned sequences from this search, and then answer the following questions:

- What were the top three matches to your query sequence?
- For each alignment, BLAST also indicates the percent *identity* and the number of gaps in the match between the query and subject sequences. What was the percent identity for the top three matches? What percentage of each aligned sequence showed gaps indicating sequence differences?
- Click on the links for the first matched sequence (far-right column). These will take you to a

wealth of information, including the size of the sequence; the species it was derived from; a PubMed-linked chronology of research publications pertaining to this sequence; the complete sequence; and if the sequence encodes a polypeptide, the predicted amino acid sequence coded by the gene. Skim through the information presented for this gene. What is the gene's function?

- A BLAST search can also be done by entering the *accession number* for a sequence, which is a unique identifying number assigned to a sequence before it can be put into a database. For example, search with the accession number NM\_007305. What did you find?
- Run a BLAST search using the sequences or accession numbers listed below. In each case, after entering the

accession number or sequence in the “Enter Query Sequence” box, go to the “Choose Search Set” box and click on the “Others” button for database. Then go to the “Program Selection” box and click “megablast” before running your search. These features will allow you to align the query sequence with similar genes from a number of other species. When each search is completed, explore the information BLAST provides so that you can identify and learn about the gene encoded by the sequence.

- NM\_001006650. What is the top sequence that aligns with the query sequence of this accession number and shows 100 percent sequence identity?
- DQ991619. What gene is encoded by this sequence?
- NC\_007596. What living animal has a sequence similar to this one?

## CASE STUDY | Zigs and zags of the smallpox virus

Smallpox, a once highly lethal contagious disease, has been eradicated worldwide. However, research continues with stored samples of variola, the smallpox virus, because it is a potential weapon in bioterrorism. Human cells protect themselves from the variola virus (and other viruses) by activating genes that encode protective proteins. It has recently been discovered that in response to variola, human cells create small transitory stretches of Z-DNA at sites that regulate these genes. The smallpox virus can bypass this cellular defense mechanism by specifically targeting the segments of Z-DNA and inhibiting the synthesis of the protective proteins. This discovery raises some interesting questions:

- What is unique about Z-DNA that might make it a specific target during viral infection?
- How might the virus target host-cell Z-DNA formation to block the synthesis of antiviral proteins?
- To study the interaction between viral proteins and Z-DNA, how could Z-DNA-forming DNA be synthesized in the lab?
- How could this research lead to the development of drugs to combat infection by variola and related viruses?

### Summary Points

- Although both proteins and nucleic acids were initially considered as possible candidates for genetic material, proteins were initially favored.
- By 1952, transformation studies and experiments using bacteria infected with bacteriophages strongly suggested that DNA is the genetic material in bacteria and most viruses.
- Although initially only indirect observations supported the hypothesis that DNA controls inheritance in eukaryotes,

subsequent studies involving recombinant DNA techniques and transgenic mice provided direct experimental evidence that the eukaryotic genetic material is DNA.

- RNA serves as the genetic material in some bacteriophages as well as some plant and animal viruses.
- As proposed by Watson and Crick, DNA exists in the form of a right-handed double helix composed of two long antiparallel polynucleotide chains held together

**MasteringGenetics™** For activities, animations, and review quizzes, go to the Study Area.

- by hydrogen bonds formed between complementary, nitrogenous base pairs.
- The second category of nucleic acids important in genetic function is RNA, which is similar to DNA with the exceptions that it is usually single stranded, the sugar ribose replaces the deoxyribose, and the pyrimidine uracil replaces thymine.
  - Various methods of analysis of nucleic acids, particularly molecular hybridization and electrophoresis, have led to studies essential to our understanding of genetic mechanisms.

- Among the techniques used to study double-stranded DNA, reassociation kinetics analysis enabled geneticists to postulate the existence of repetitive DNA in eukaryotes, where certain nucleotide sequences are present many times in the genome.

## INSIGHTS AND SOLUTIONS

*The current chapter, in contrast to preceding chapters, does not emphasize genetic problem solving. Instead, it recounts some of the initial experimental analyses that launched the era of molecular genetics. Accordingly, our “Insights and Solutions” section shifts its emphasis to experimental rationale and analytical thinking, an approach that will continue to be used in later chapters whenever appropriate.*

- Based strictly on your scrutiny of the transformation data of Avery, MacLeod, and McCarty, what objection might be made to the conclusion that DNA is the genetic material? What other conclusion might be considered?
  - What observations, including later ones, argue against this objection?

**Solution:** (a) Based solely on their results, it may be concluded that DNA is essential for transformation. However, DNA might have been a substance that caused capsular formation by *directly* converting nonencapsulated cells to cells with a capsule. That is, DNA may simply have played a catalytic role in capsular synthesis, leading to cells displaying smooth type III colonies.

(b) First, transformed cells pass the trait onto their progeny cells, thus supporting the conclusion that DNA is responsible for heredity, not for the direct production of polysaccharide coats. Second, subsequent transformation studies over a period of five years showed that other traits, such as antibiotic resistance, could be transformed. Therefore, the transforming factor has a broad general effect, not one specific to polysaccharide synthesis. This observation is more in keeping with the conclusion that DNA is the genetic material.

- If RNA were the universal genetic material, how would it have affected the Avery experiment and the Hershey–Chase experiment?

**Solution:** In the Avery experiment, digestion of the soluble filtrate with RNase, rather than DNase, would have eliminated transformation. Had this occurred, Avery and his colleagues would have concluded that

RNA was the transforming factor. Hershey and Chase would have obtained identical results, since  $^{32}\text{P}$  would also label RNA but not protein. Had they been using a bacteriophage with RNA as its nucleic acid, and had they known this, they would have concluded that RNA was responsible for directing the reproduction of their bacteriophage.

- A quest to isolate an important disease-causing organism was successful, and molecular biologists were hard at work analyzing the results. The organism contained as its genetic material a remarkable nucleic acid with a base composition of A = 21 percent, C = 29 percent, G = 29 percent, U = 21 percent. When heated, it showed a major hyperchromic shift, and when the reassociation kinetics were studied, the nucleic acid of this organism reannealed more slowly than that of phage T4 and *E. coli*. T4 contains  $10^5$  nucleotide pairs.

Analyze this information carefully, and draw *all* possible conclusions about the genetic material of this organism, based strictly on the preceding observations. As a test of your model, make one prediction that if upheld would strengthen your hypothesis about the nature of this molecule.

**Solution:** First of all, because of the presence of uracil (U), the molecule appears to be RNA. In contrast to normal RNA, this one has base ratios of  $A/U = G/C = 1$ , suggesting that the molecule may be a double helix. The hyperchromic shift and reassociation kinetics support this hypothesis. In the kinetic study, since none of the nucleic acid segments reannealed more rapidly than bacterial or viral nucleic acid, there is no repetitive sequence RNA. Furthermore, the total length of unique-sequence DNA is greater than that of either phage T4 ( $10^5$  nucleotide pairs) or *E. coli*. A prediction might be made concerning the sugars. Our model suggests that ribose rather than deoxyribose should be present. If so, this observation would support the hypothesis that RNA is the genetic material in this organism.

## Problems and Discussion Questions

MasteringGenetics™ Visit for instructor-assigned tutorials and problems.

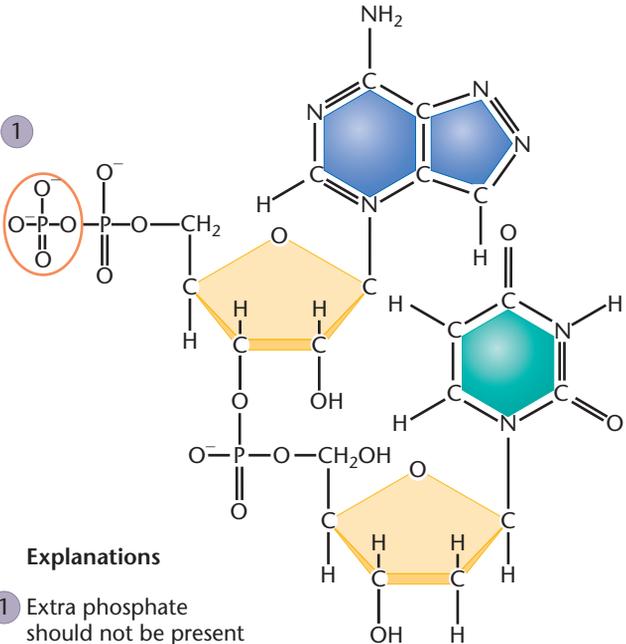
### HOW DO WE KNOW?

- In this chapter, we first focused on the information that showed DNA to be the genetic material and then discussed the structure of DNA as proposed by Watson and Crick. We concluded the chapter by describing various techniques developed to study DNA. Along the way, we found many opportunities to consider the methods and reasoning by which much of this information was acquired. From the explanations given in the chapter, what answers would you propose to the following fundamental questions:
  - How were scientists able to determine that DNA, and not some other molecule, serves as the genetic material in bacteria and bacteriophages?
  - How do we know that DNA also serves as the genetic material in eukaryotes such as humans?
  - How was it determined that the structure of DNA is a double helix with the two strands held together by hydrogen bonds formed between complementary nitrogenous bases?
  - How do we know that G pairs with C and that A pairs with T as complementary base pairs are formed?
  - How do we know that repetitive DNA sequences exist in eukaryotes?

### CONCEPT QUESTION

- Review the Chapter Concepts list on p. 231. Most center around DNA and RNA and their role of serving as the genetic material. Write a short essay that contrasts these molecules, including a comparison of advantages conferred by their structure that each of them has over the other in serving in this role.
- Discuss the reasons proteins were generally favored over DNA as the genetic material before 1940. What was the role of the tetranucleotide hypothesis in this controversy?
- Contrast the various contributions made to an understanding of transformation by Griffith, by Avery and his colleagues, and by Taylor.
- When Avery and his colleagues had obtained what was concluded to be the transforming factor from the IIIS virulent cells, they treated the fraction with proteases, RNase, and DNase, followed in each case by the assay for retention or loss of transforming ability. What were the purpose and results of these experiments? What conclusions were drawn?
- Why were  $^{32}\text{P}$  and  $^{35}\text{S}$  chosen for use in the Hershey–Chase experiment? Discuss the rationale and conclusions of this experiment.
- Does the design of the Hershey–Chase experiment distinguish between DNA and RNA as the molecule serving as the genetic material? Why or why not?
- What observations are consistent with the conclusion that DNA serves as the genetic material in eukaryotes? List and discuss them.
- What are the exceptions to the general rule that DNA is the genetic material in all organisms? What evidence supports these exceptions?
- Draw the chemical structure of the three components of a nucleotide, and then link the three together. What atoms are removed from the structures when the linkages are formed?
- How are the carbon and nitrogen atoms of the sugars, purines, and pyrimidines numbered?
- Adenine may also be named 6-amino purine. How would you name the other four nitrogenous bases, using this alternative system? (O is indicated by “oxy-,” and  $\text{CH}_3$  by “methyl.”)
- Draw the chemical structure of a dinucleotide composed of A and G. Opposite this structure, draw the dinucleotide composed of T and C in an antiparallel (or upside-down) fashion. Form the possible hydrogen bonds.
- Describe the various characteristics of the Watson–Crick double-helix model for DNA.
- What evidence did Watson and Crick have at their disposal in 1953? What was their approach in arriving at the structure of DNA?
- What might Watson and Crick have concluded had Chargaff's data from a single source indicated the following?
 

	A	T	G	C
%	29	19	21	31

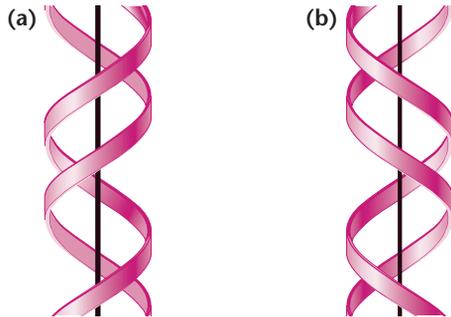
Why would this conclusion be contradictory to Wilkins's and Franklin's data?
- How do covalent bonds differ from hydrogen bonds? Define base complementarity.
- List three main differences between DNA and RNA.
- What are the three major types of RNA molecules? How is each related to the concept of information flow?
- What component of the nucleotide is responsible for the absorption of ultraviolet light? How is this technique important in the analysis of nucleic acids?
- What is the physical state of DNA following denaturation?
- What is the hyperchromic effect? How is it measured? What does  $T_m$  imply?
- Why is  $T_m$  related to base composition?
- What is the chemical basis of molecular hybridization?
- What did the Watson–Crick model suggest about the replication of DNA?
- A genetics student was asked to draw the chemical structure of an adenine- and thymine-containing dinucleotide derived from DNA. His answer is shown here:
 

Explanations

  - Extra phosphate should not be present

The student made more than six major errors. One of them is circled, numbered 1, and explained. Find five others. Circle them, number them 2 through 6, and briefly explain each in the manner of the example given.

27. Considering the information in this chapter on B- and Z-DNA and right- and left-handed helices, carefully analyze structures (a) and (b) below and draw conclusions about their helical nature. Which is right handed and which is left handed?



28. One of the most common spontaneous lesions that occurs in DNA under physiological conditions is the hydrolysis of the

amino group of cytosine, converting the cytosine to uracil. What would be the effect on DNA structure of a uracil group replacing cytosine?

29. In some organisms, cytosine is methylated at carbon 5 of the pyrimidine ring after it is incorporated into DNA. If a 5-methyl cytosine molecule is then hydrolyzed, as described in Problem 28, what base will be generated?
30. Because of its rapid turnaround time, fluorescent *in situ* hybridization (FISH) is commonly used in hospitals and laboratories as an aneuploid screen of cells retrieved from amniocentesis and chorionic villus sampling (CVS). Chromosomes 13, 18, 21, X, and Y (see Chapter 8) are typically screened for aneuploidy in this way. Explain how FISH might be accomplished using amniotic or CVS samples and why the above chromosomes have been chosen for screening.
31. Assume that you are interested in separating short (200–400 nucleotides) DNA molecules from a pool of longer molecules in the 10,000–20,000 nucleotide range. You have two recipes for making your agarose gels: one recipe uses 1.5 percent agarose and would be considered a “hard gel,” while the other uses 0.5 percent agarose and would be considered a loose gel. Which recipe would you consider using and why?

## Extra-Spicy Problems

MasteringGenetics™ Visit for instructor-assigned tutorials and problems.

32. A primitive eukaryote was discovered that displayed a unique nucleic acid as its genetic material. Analysis provided the following information:

- (a) The general X-ray diffraction pattern is similar to that of DNA, but with somewhat different dimensions and more irregularity.
- (b) A major hyperchromic shift is evident upon heating and monitoring UV absorption at 260 nm.
- (c) Base-composition analysis reveals four bases in the following proportions:

Adenine	=	8%
Guanine	=	37%
Xanthine	=	37%
Hypoxanthine	=	18%

- (d) About 75 percent of the sugars are deoxyribose, while 25 percent are ribose.

Postulate a model for the structure of this molecule that is consistent with the foregoing observations.

33. *Newsdate: March 1, 2030.* A unique creature has been discovered during exploration of outer space. Recently, its genetic material has been isolated and analyzed. This material is similar in some ways to DNA in its chemical makeup. It contains in abundance the 4-carbon sugar erythrose and a molar equivalent of phosphate groups. In addition, it contains six nitrogenous bases: adenine (A), guanine (G), thymine (T), cytosine (C), hypoxanthine (H), and xanthine (X). These bases exist in the following relative proportions:

$$A = T = H \text{ and } C = G = X$$

X-ray diffraction studies have established a regularity in the molecule and a constant diameter of about 30 Å. Together, these data have suggested a model for the structure of this molecule.

- (a) Propose a general model of this molecule. Describe it briefly.
- (b) What base-pairing properties must exist for H and for X in the model?
- (c) Given the constant diameter of 30 Å, do you think that *either* (i) both H and X are purines or both pyrimidines, *or* (ii) one is a purine and one is a pyrimidine?
34. You are provided with DNA samples from two newly discovered bacterial viruses. Based on the various analytical techniques discussed in this chapter, construct a research protocol that would be useful in characterizing and contrasting the DNA of both viruses. For each technique that you include in the protocol, indicate the type of information you hope to obtain.
35. During gel electrophoresis, DNA molecules can easily be separated according to size because all DNA molecules have the same charge-to-mass ratio and the same shape (long rod). Would you expect RNA molecules to behave in the same manner as DNA during gel electrophoresis? Why or why not?
36. Electrophoresis is an extremely useful procedure when applied to analysis of nucleic acids as it can resolve molecules of different sizes with relative ease and accuracy. Large molecules migrate more slowly than small molecules in agarose gels. However, the fact that nucleic acids of the same length may exist in a variety of conformations can often complicate the interpretation of electrophoretic separations. For instance, when a single species of a bacterial plasmid is isolated from cells, the individual plasmids may exist in three forms (depending on the genotype of their host and conditions of isolation): superhelical/supercoiled (form I), nicked/open circle

(form II), and linear (form III). Form I is compact and very tightly coiled, with both DNA strands continuous. Form II exists as a loose circle because one of the two DNA strands has been broken, thus releasing the supercoil. All three have the same mass, but each will migrate at a different rate through a gel. Based on your understanding of gel composition and DNA migration, predict the relative rates of migration of the various DNA structures mentioned above.

37. Following is a table (modified from Kropinski, 1973) that presents the  $T_m$  and chemical composition (%G  $\equiv$  C) of DNA from certain bacteriophages. From these data develop a graph that presents %G  $\equiv$  C (ordinate) and  $T_m$  (abscissa). What is the relationship between  $T_m$  and %G  $\equiv$  C for these samples? What might be the molecular basis of this relationship? Visit for instructor-assigned tutorials and problems.

Phage	$T_m$	%G $\equiv$ C
$\alpha$	86.5	44.0
$\kappa$	91.5	53.8
$\lambda$	89	49.2
$\phi 80$	90.5	53.0
$\chi$	92.1	57.4
Mu-1	88	51.4
T1	89	48.0
T3	90	49.6
T7	89.5	48.0

FOR REVIEW ONLY