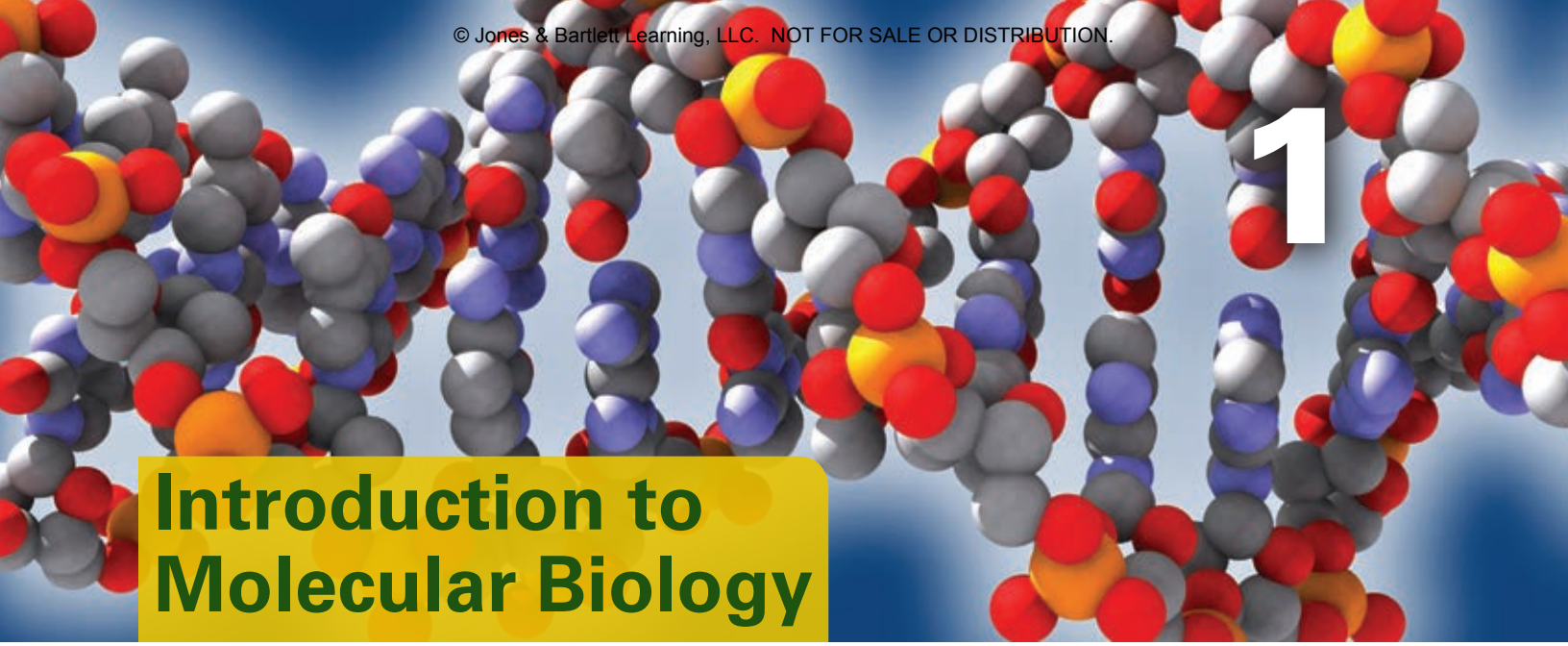


1 Introduction to Molecular Biology

CHAPTER OUTLINE

The first recorded use of the term **molecular biology** appears to have been in a 1938 report prepared for the Rockefeller Foundation by Warren Weaver, the director of the Foundation's Natural Science Division. The report proposed that the Foundation should start to fund research efforts to seek molecular explanations for biological processes. Weaver used the term "molecular biology" to describe a research approach he envisioned would use the physical sciences to address fundamental biological problems. Weaver's proposal was remarkably farsighted, especially when one considers many of his contemporaries believed living cells possessed a vital force that could not be explained by chemical or physical laws that govern the inanimate world.

At the time of Weaver's report two new disciplines—genetics and biochemistry—were altering the way biologists thought about living systems. Geneticists established that the functional and physical unit of heredity is the gene. However, they did not know the chemical nature of genes, the way that hereditary information is stored in genes, how genes are replicated so they can be transmitted to the next generation, or how information stored in genes determines a specific physical trait such as eye color. Biochemists had delivered a major blow to the vital force theory by demonstrating that cell-free extracts can perform many of the same functions as intact cells. However, they knew very little about protein or nucleic acid structure and function.

Neither genetics nor biochemistry had the power to investigate the chemical basis of heredity on its own. In fact, it took an interdisciplinary effort involving specialists in many fields of the life sciences, including genetics, biochemistry, biophysics, microbiology, chemistry, x-ray crystallography, virology, developmental biology, and immunology, to solve the hereditary problem. This interdisciplinary effort resulted in the creation of a new discipline, molecular biology, which has had a remarkable impact on all areas of the life sciences from agronomy to zoology. The scope of this book is limited to an examination of the contributions that molecular biology has made to understanding the flow of information from genes to proteins.

1.1 Intellectual Foundation

Two studies performed in the 1860s provided the intellectual underpinning for molecular biology.

The earliest intellectual roots of molecular biology can be traced back to the work of two investigators in the 1860s. No connection was apparent between the experiments performed by the two investigators for more than 75 years, but when the connection finally was made, the result was the birth of molecular biology and the beginning of a scientific revolution that continues today.

The work of the first investigator, Gregor Mendel, an Austrian monk and botanist, is familiar to all biology students and so is only summarized briefly here. Mendel discovered three basic laws of inheritance by studying the way in which simple physical traits are passed on from one generation of pea plants to the next. For convenience, Mendel's laws of inheritance will be described using two modern biological terms, **gene** for a unit of heredity and **chromosome** for a structure bearing several linked genes.

1. *The law of segregation:* A specific gene may exist in alternate forms called **alleles**. An organism inherits one allele for each trait from each parent. The two alleles, which may be the same or different, segregate (or separate) during germ cell (sperm or egg) formation and combine again as a result of fertilization so that each parent transmits one allele to each offspring.
2. *The law of independent assortment:* Specific physical traits such as plant size and color are inherited independently of one another. Mendel was fortunate to have selected physical traits that were determined by genes that were on different chromosomes.
3. *The law of dominance:* An allele may be **dominant** or **recessive**. A dominant allele produces its characteristic physical trait whether it is paired with an identical allele or a recessive allele. In contrast, a recessive allele produces its characteristic physical trait only when paired with an identical allele. In pea plants, tallness is dominant and shortness recessive. When Mendel allowed pea plants with one allele for tallness and one allele for shortness to self-fertilize, he observed three times as many progeny were tall as short. Today, we know there are exceptions to the law of dominance. Sometimes neither allele is dominant. For instance, a plant that inherits a gene for a red flower and a gene for a white flower may produce an intermediate phenotype, a pink flower.

Unfortunately, scientists failed to recognize the significance of Mendel's work during his lifetime. His paper remained little-known until about 1900, when scientists rediscovered Mendel's laws of inheritance, giving birth to the science of genetics.

The second investigator, the Swiss physician Friedrich Miescher, performed experiments that led to the discovery of deoxyribonucleic acid (DNA), which we, of course, now know is the hereditary material. Miescher did not set out to discover the hereditary material but instead was interested in studying cell nuclei from white blood cells, which he collected from pus discharges on discarded bandages used to cover infected wounds. Miescher used a combination of protease (an enzyme that hydrolyzes proteins) digestion and solvent extraction to disrupt and fractionate the white blood cells. One fraction, which he called nuclein, contained an acidic material with unusually high phosphorus content. Miescher later found that salmon sperm cells, which have remarkably large cell nuclei, are also an excellent source of nuclein. In 1889 Miescher's student, Richard Altmann, separated nuclein into protein and a substance with a very high phosphorous content that he named **nucleic acid**.

1.2 One Gene–One Polypeptide Hypothesis

Each gene is responsible for the synthesis of a single polypeptide.

Mendel's experiments showed that the genetic makeup of an organism, its **genotype**, determines the organism's physical traits, its **phenotype**. However, his experiments did not show how genes are able to determine complex physical traits such as plant color or size. Archibald Garrod, an English physician, was the first to provide an explanation for the relationship between genotype and phenotype. Garrod uncovered this relationship in the early 1920s while studying alkaptonuria, a rare inherited human disorder in which the urine of affected individuals becomes very dark upon standing due to the oxidation of homogentisic acid, a breakdown product of the amino acid tyrosine. Garrod correctly proposed that alkaptonuria results from a recessive gene, which causes a deficiency in an enzyme needed to convert homogentisic acid to colorless products.

Garrod's work was generally ignored until the early 1940s, when two American scientists, George Beadle and Edward Tatum, provided additional experimental proof for a relationship between genes and enzymes. Beadle and Tatum thought that if a gene really does specify an enzyme, it should be possible to create genetic mutants that cannot carry out specific enzymatic reactions in a biochemical pathway. They decided to work with the bread mold, *Neurospora crassa*, because it could be subjected to genetic analysis. In addition, *N. crassa* has very simple nutritional requirements. It can grow in a minimal medium that contains a single carbon source such as sucrose, inorganic salts, and the vitamin biotin. Beadle and Tatum began by irradiating *N. crassa* with x-rays to generate mutants with additional nutritional growth requirements. A mutant that requires a specific nutritional supplement that is not required by the parent cell is called an **auxotroph**.

In one set of experiments Beadle and Tatum focused particular attention on arginine auxotrophs. Genetic analysis of these auxotrophs revealed that mutations in three different genes produce auxotrophs for the amino acid arginine. Each kind of auxotroph was blocked at a specific step in the arginine biosynthetic pathway and accumulated large quantities of the substance formed just before the blocked step. Beadle and Tatum thus had replicated in the bread mold the same type of situation Garrod had observed in alkaptonuria. A defective gene caused a defect in a specific enzyme that resulted in the abnormal accumulation of an intermediate in a metabolic pathway.

As a result of their work with the *N. crassa* mutants, Beadle and Tatum proposed the one gene–one enzyme hypothesis, which states that each gene is responsible for synthesizing a single enzyme. We now know that many enzymes are made of more than one type of polypeptide chain and that a single mutation may affect just one of the polypeptide chains. The original one gene–one enzyme hypothesis hence was modified to become a *one gene–one polypeptide hypothesis*. As we will see later, however, even the one gene–one polypeptide hypothesis is an oversimplification.

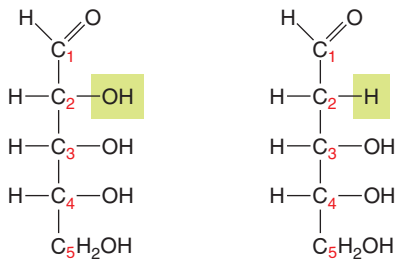


FIGURE 1.1 The two sugars present in nucleic acids. (a) Ribose and (b) deoxyribose each contain five carbon atoms and an aldehyde group and therefore belong to aldopentose family of simple sugars. The only difference between the two sugars is that ribose has a hydroxyl group at carbon-2 and deoxyribose as a hydrogen atom at carbon-2.

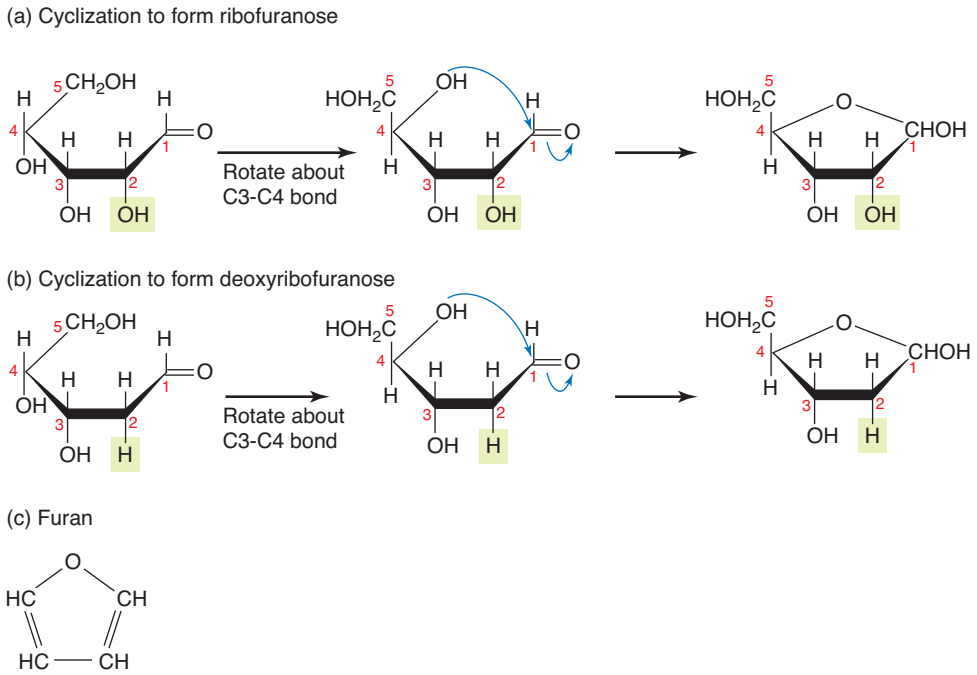


FIGURE 1.2 Conversion of straight chain ribose and deoxyribose to cyclic forms. (a) Conversion of ribose to ribofuranose and (b) conversion of deoxyribose to deoxyribofuranose. (c) Structure of furan, a five-atom ring system.

1.3 Introduction to Nucleic Acids

DNA contains the sugar deoxyribose and RNA contains the sugar ribose.

Investigators slowly came to realize that nucleic acids could be divided into two major groups: DNA and ribonucleic acid (RNA). The principal difference between DNA and RNA is that the former contains deoxyribose and the latter contains ribose (**FIGURE 1.1**). The two five-carbon sugars (pentoses) differ in only one substituent: deoxyribose has a hydrogen atom at carbon-2, whereas ribose has a hydroxyl group at this position. By convention, the aldehyde group in the pentose is C-1, the next carbon is C-2, and so forth.

Each pentose chain can close to form a five-carbon ring in which an oxygen bridge joins C-1 to C-4 (**FIGURE 1.2ab**). Because the sugar rings are derivatives of furan (Figure 1.2c) they are called **furanoses.** Ribofuranose and deoxyribofuranose are often depicted as Haworth structures (named after Walter N. Haworth, the investigator who devised the representations). As illustrated in **FIGURE 1.3**, a Haworth structure represents a cyclic sugar as a flat ring perpendicular to the plane of the page with the ring-oxygen in the back and C-1 to the right. The ring's thick lower edge projects toward the viewer, its upper edge projects back behind the page, and its substituents are visualized as being either above or below the plane of the ring. A line is used to

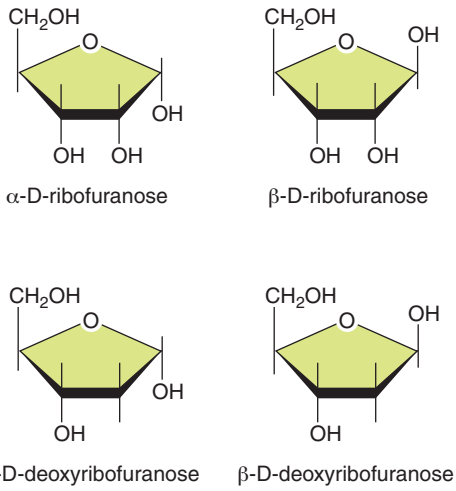


FIGURE 1.3 Haworth structures for ribofuranose and deoxyribofuranose. Haworth structure represents a cyclic sugar as a flat ring perpendicular to the plane of the page with the ring-oxygen in the back and C-1 to the right. A hydrogen atom attached to the sugar ring is represented by a line. Ring formation can lead to two different stereochemical arrangements at C-1: one in which the hydroxyl at C-1 points down and another in which it points up.

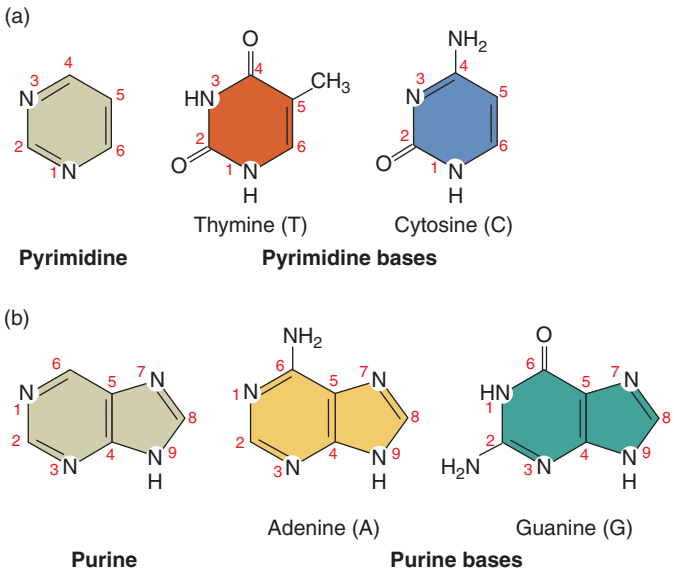


FIGURE 1.4 Pyrimidine and purine bases in DNA. (a) Pyrimidine bases and (b) purine bases.

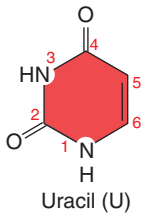


FIGURE 1.5 Uracil (U).

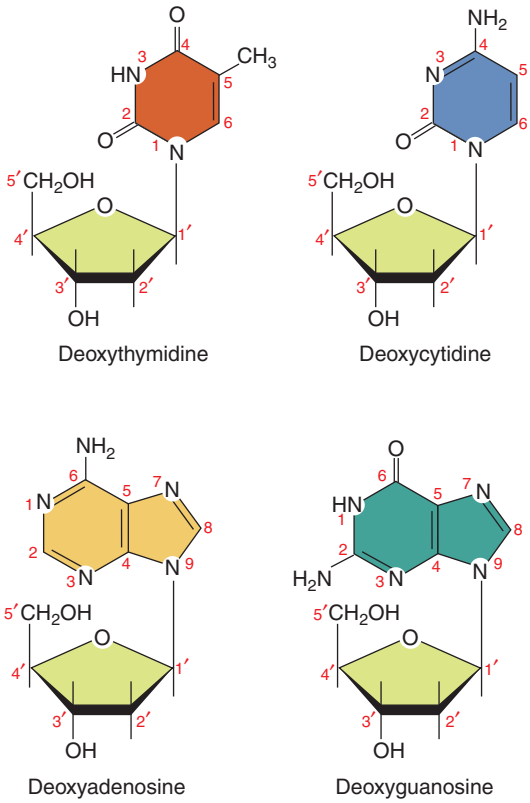


FIGURE 1.6 Deoxyribonucleosides.

represent a hydrogen atom attached to the sugar ring. Ring formation can lead to two different stereochemical arrangements at C-1. One arrangement, the α-anomer, is represented by drawing the hydroxyl group attached to C-1 below the plane of the ring and the other, the β-anomer, by drawing it above the plane of the ring. (Although ribofuranose and deoxyribofuranose actually have puckered rather than planar conformations, Haworth structures are convenient representations for the pentose rings when precise three-dimensional information is not required and, therefore, are used throughout this book.)

A nucleoside is formed by attaching a purine or pyrimidine base to a sugar.

Each deoxyribose group in DNA is attached to one of four different nitrogen heterocyclic molecules, referred to as bases because they can act as proton acceptors. Two bases, **thymine** (T) and **cytosine** (C), are derivatives of **pyrimidine** (FIGURE 1.4a) and the other two, **adenine** (A) and **guanine** (G), are derivatives of **purine** (Figure 1.4b). RNA also contains cytosine, adenine, and guanine, but the pyrimidine **uracil** (U) replaces thymine (FIGURE 1.5). The only difference between uracil and thymine is that the latter contains a methyl group attached to carbon-5. T, C, A, and G combine with deoxyribose to form a class of compounds called deoxyribonucleosides (FIGURE 1.6), and U, C, A, and G combine with ribose to form a related class of compounds called ribonucleosides (FIGURE 1.7). Each base is linked to the pentose ring by a bond that joins a specific nitrogen atom on the base (N-1 in pyrimidines and N-9 in purines) to C-1 on the furanose ring. This bond is termed an **N-glycosylic bond**. Because each nucleoside has two ring systems (the sugar and the base attached to it), a method is required to distinguish between atoms in each ring system. This problem is solved

by adding a prime (′) after the sugar atoms. The first carbon atom in the sugar thus becomes 1′, the second 2′, and so forth.

A nucleotide is formed by attaching a phosphate group to the sugar group in a nucleoside.

Nucleosides that have a phosphate group attached to the sugar group are called nucleotides. Ribonucleoside and deoxyribonucleoside derivatives are called ribonucleotides and deoxyribonucleotides, respectively (**FIGURE 1.8**). The pentose carbon atom to which the phosphate group is attached is given as part of the nucleotide's name. Thus, the phosphate group is attached to C-5′ in uridine-5′-monophosphate (5′-UMP) and thymidine-5′-monophosphate (5′-dTMP) and to C-3′ in uridine-3′-monophosphate (3′-UMP) and thymidine-3′-monophosphate (3′-dTMP). As indicated in **TABLE 1.1**, nucleoside monophosphates have two alternative names. For example, cytidine-5′-monophosphate (5′-CMP) is also known as 5′-cytidylate. Each nucleoside monophosphate is

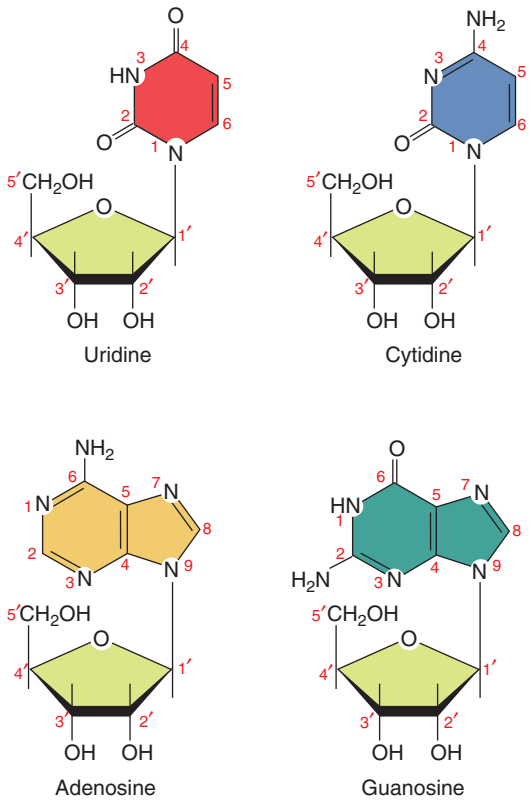


FIGURE 1.7 Ribonucleosides.

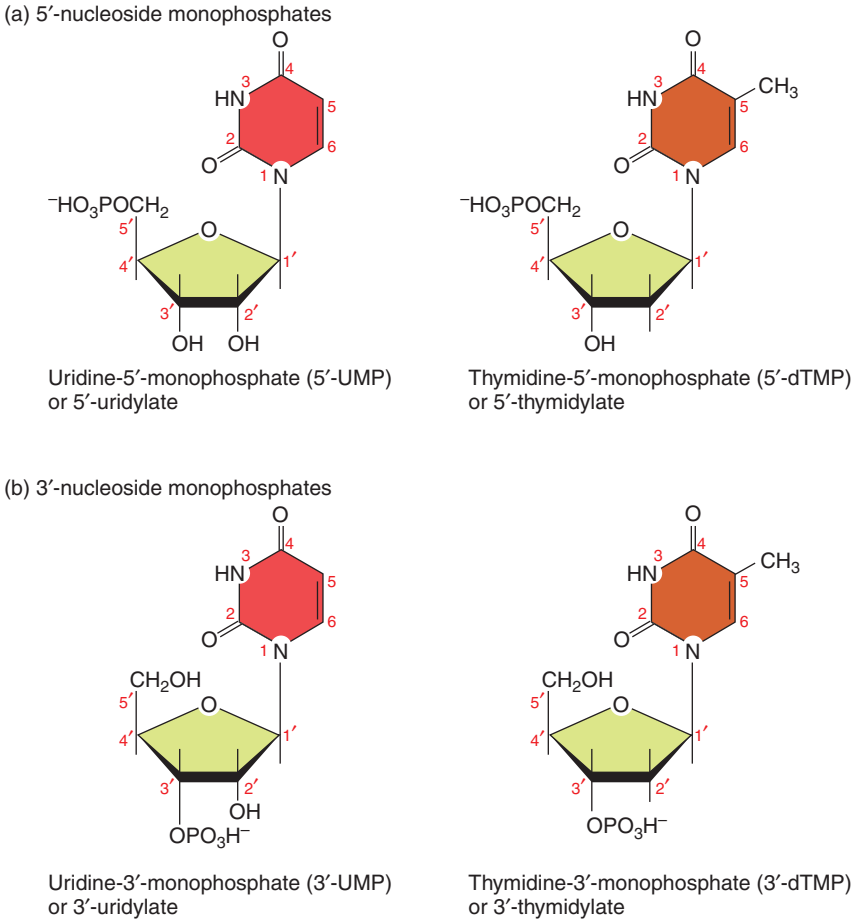


FIGURE 1.8 Nucleotides. Nucleotides are formed by adding a phosphate group to the pentose ring in a nucleoside. (a) Nucleotides formed by adding a phosphate group to the 5′-hydroxyl group in uridine or thymidine. (b) Nucleotides formed by adding a phosphate group to the 3′-hydroxyl group in uridine or thymidine.

TABLE 1.1 Bases, Nucleosides, and Nucleotides

Base	Sugar	Nucleoside	5′-Mononucleotide
Uracil (U)	ribose	uridine	Uridine-5′-monophosphate or 5′-uridylate (5′-UMP)
Cytosine (C)	ribose	cytidine	Cytidine-5′-monophosphate or 5′-cytidylate (5′-CMP)
Adenine (A)	ribose	adenosine	Adenosine-5′-monophosphate or 5′-adenylate (5′-AMP)
Guanine (G)	ribose	guanosine	Guanosine-5′-monophosphate or 5′-guanylate (5′-GMP)
Thymine (T)	deoxyribose	deoxythymidine[1]	Deoxythymidine-5′-monophosphate or 5′-deoxythymidylate (5′-dTMP)[1]
Cytosine (C)	deoxyribose	deoxycytidine	Deoxycytidine-5′-monophosphate or 5′-deoxycytidylate (5′-dCMP)
Adenine (A)	deoxyribose	deoxyadenosine	Deoxyadenosine-5′-monophosphate or 5′-deoxyadenylate (5′-dAMP)
Guanine (G)	deoxyribose	deoxyguanosine	Deoxyguanosine-5′-monophosphate or 5′-deoxyguanylate (5′-dGMP)

[1]Deoxythymidine and deoxythymidine-5′-monophosphate are also called thymidine and thymidine-5′-monophosphate, respectively. When thymine is attached to ribose, the nucleoside is called ribothymidine and the nucleotide is called ribothymidylate. This nomenclature convention follows from the fact that thymine is most frequently attached to deoxyribose.

phosphomonoester because it has a single sugar group attached to a phosphate by an ester bond.

DNA is a linear chain of deoxyribonucleotides.

The deoxyribonucleotide groups in DNA form a linear polymer in which phosphate groups join 5′- and 3′-carbons of neighboring deoxyribonucleosides (**FIGURE 1.9a**). The bond that joins neighboring nucleosides to a common phosphate group is called a **phosphodiester bond**. *Each linear DNA chain has a 5′- and a 3′-terminus*. This directionality will be very important in future discussions of DNA structure and function. The structure of RNA is very similar to that for DNA (**FIGURE 1.10a**).

Just 3 years after the discovery that nucleic acids are linear chains of nucleotides, Frederick Sanger, working in England, showed that each kind of polypeptide molecule is a linear chain of amino acids arranged in a specific order. The amino acid order is responsible for the unique properties of each type of protein, including its ability to catalyze specific reactions, protect cells from foreign organisms and substances, transport materials, or support the cellular infrastructure. The awareness that DNA molecules are made of linear chains of

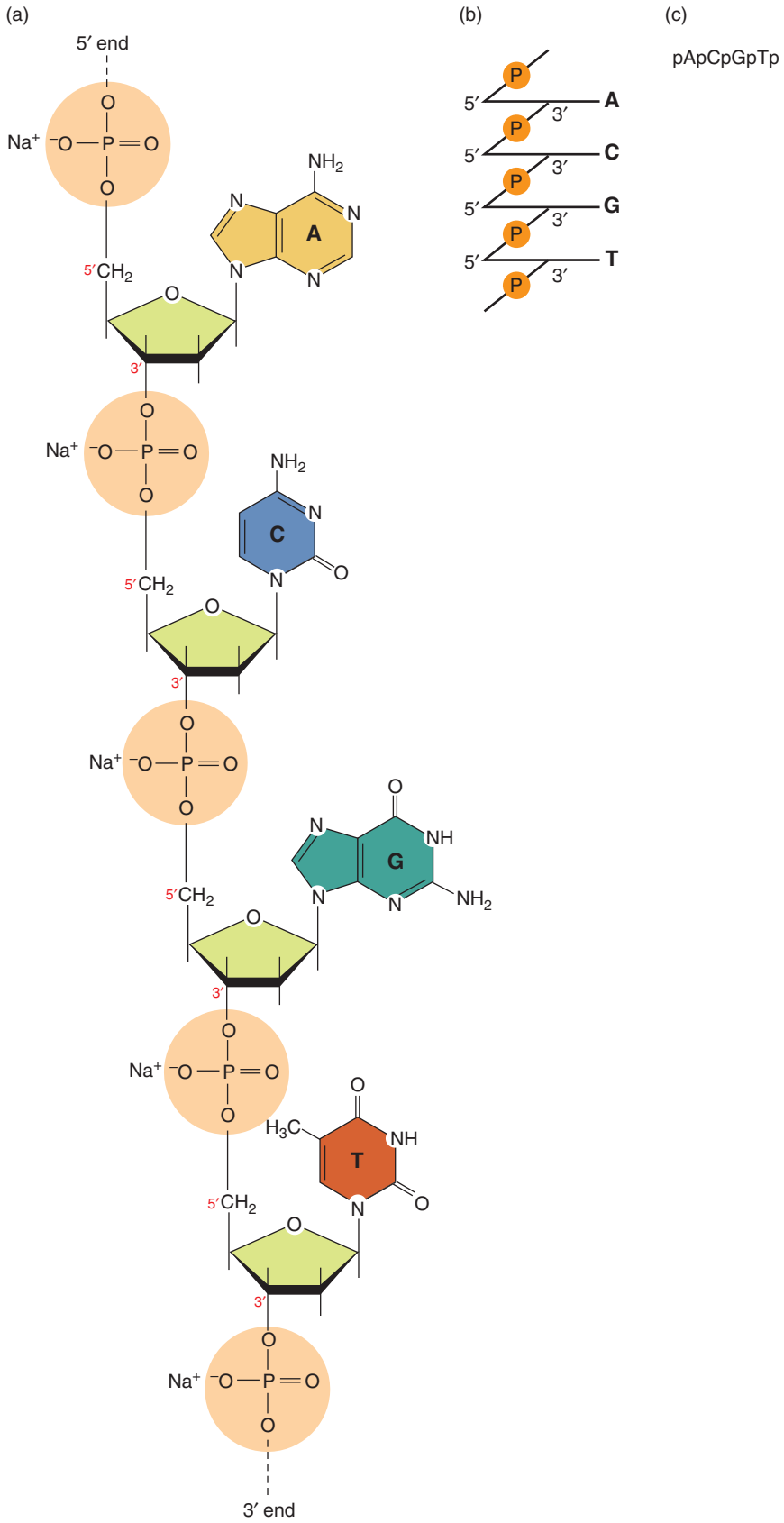


FIGURE 1.9 Segment of a polydeoxyribonucleotide. (a) Extended structure as a sodium salt, (b) stick figure structure, and (c) an abbreviated structure.

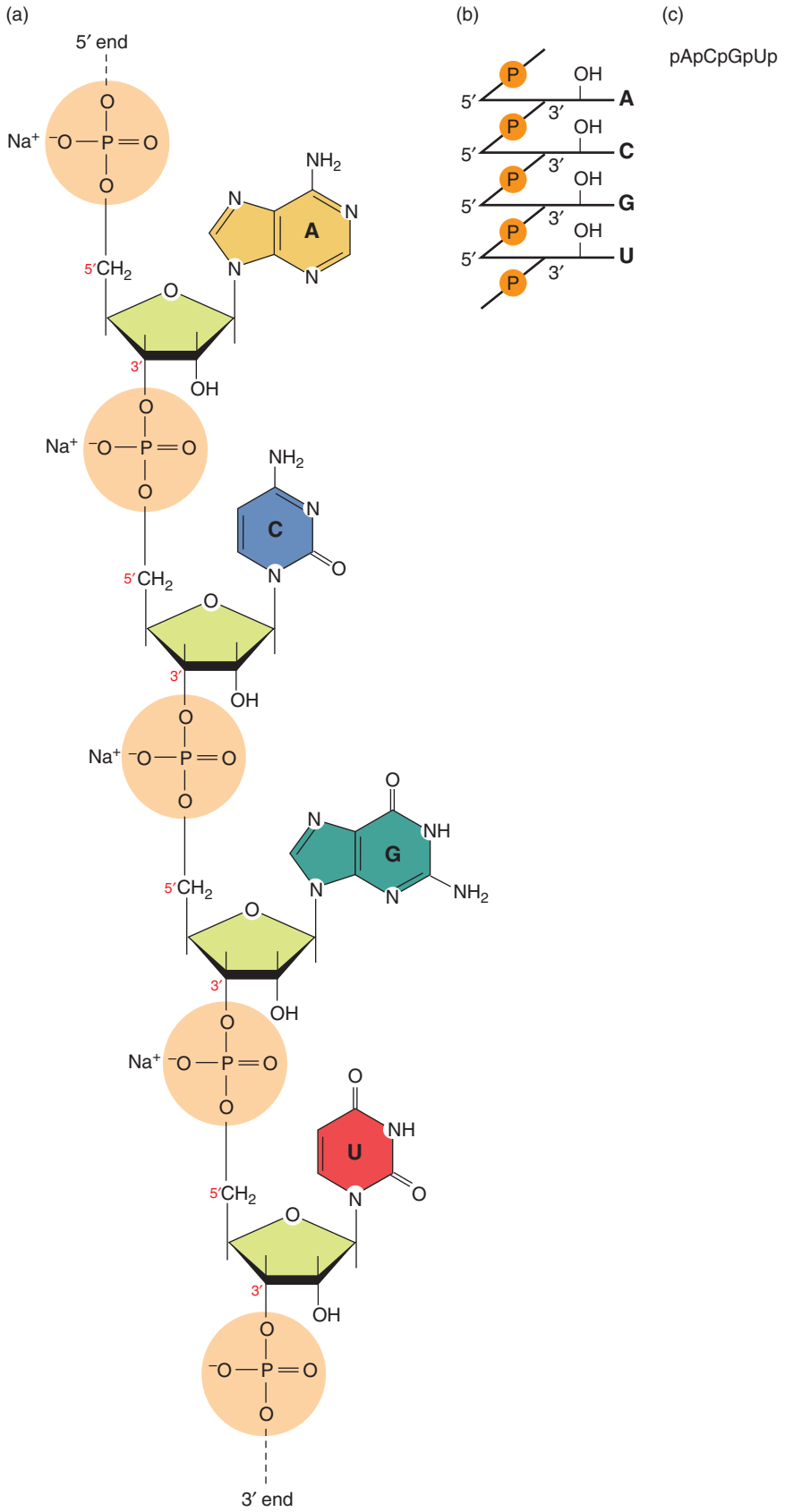


FIGURE 1.10 Segment of a polyribonucleotide. (a) Extended structure as a sodium salt, (b) stick figure structure, and (c) an abbreviated structure.

nucleotides and polypeptides are made of linear chains of amino acids led to the **sequence hypothesis,** which proposes nucleotide sequences specify amino acid sequences.

Drawing extended structures for DNA and RNA chains requires considerable time and space. It is more convenient to draw stick figure structures, which are adequate representations for many purposes (Figure 1.9b and 1.10b). The few simple conventions for drawing stick figure representations for DNA and RNA are as follows: (1) a single line (shown as horizontal in Figures 1.9b and 1.10b) represents the pentose ring; (2) the letter A, G, C, U, or T, at one end of the line, represents the purine or pyrimidine attached to C-1′ of the pentose ring; (3) the letter P, connected by short diagonal lines to adjacent lines, represents the 5′→3′ phosphodiester bond; and (4) the symbol OH represents a hydroxyl group. An even simpler method for indicating nucleotide sequence is to write the letters corresponding to the bases (Figure 1.9c and 1.10c).

1.4 DNA: Hereditary Material

Transformation experiments led to the discovery that DNA is the hereditary material.

The first step on the path leading to the discovery that genes are made of DNA was an observation made in 1928 by Fred Griffith, who was studying *Streptococcus pneumoniae,* a bacterial strain responsible for human pneumonia. This bacterium's virulence was known to depend on a surrounding polysaccharide capsule that protects the bacterium from the body's defense systems. The capsule also causes the bacterium to produce smooth-edged colonies on an agar surface. Bacteria from smooth-edged colonies, called S bacteria, normally kill mice. Griffith isolated a *S. pneumoniae* mutant that produced rough-edged colonies. Cells from rough-edged colonies, called R bacteria, proved to be both non-encapsulated and nonlethal to mice. Additional experiments revealed that although either live R or heat-killed S bacteria are nonlethal to mice, a mixture of the two is lethal (**FIGURE 1.11**). Furthermore, when bacteria were isolated from a mouse that had died from such a mixed infection, the bacteria were live S and R. Live R bacteria, therefore, had somehow either been replaced by or transformed to S bacteria.

Several years later investigators showed the mouse itself is not needed to mediate this transformation, because when a mixture containing live R bacteria and heat-killed S bacteria is incubated in culture medium, living S cells are produced. One possible explanation for this surprising phenomenon is that the R cells restore the viability of the dead S cells. This hypothesis was eliminated by the observation that living S cells appear even when the heat-killed S bacteria in the mixture are replaced by an S cell extract, which had been centrifuged to remove both intact cells and the capsular polysaccharide. Based on these studies it appeared that the S cell extract contains a **transforming principle** of unknown chemical composition.

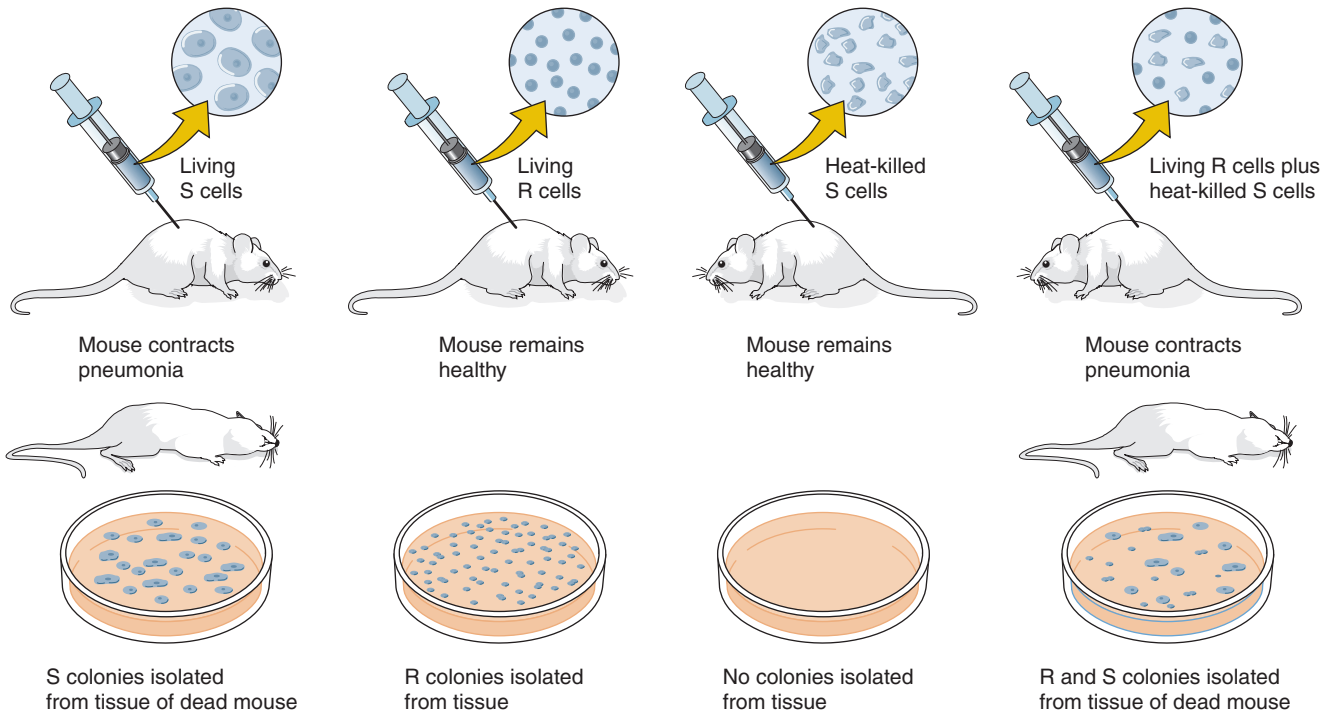


FIGURE 1.11 Griffith's experiment demonstrating bacterial transformation. A mouse dies from pneumonia if injected with the virulent S (smooth) strain of *Streptococcus pneumoniae*. However, the mouse remains healthy if injected with either the nonvirulent R (rough) strain or the heat-killed S strain. R cells in the presence of heat-killed S cells are transformed into the virulent S strain, killing the mouse.

The next development occurred in 1944 when Oswald Avery, Colin MacLeod, and Maclyn McCarty determined the chemical nature of the transforming principle. They did so by purifying the transforming principle from S cells and adding it to live R bacterial cultures (**FIGURE 1.12**). After allowing the mixture to incubate for a period of time, they placed samples on an agar surface and incubated them until colonies appeared. Some of the colonies (about 1 in 10^4) that grew were S type. To show this was a permanent genetic change, Avery and coworkers dispersed many of the newly formed S colonies and placed them on a second agar surface. The resulting colonies were again S type. If an R colony arising from the original mixture was dispersed, only R bacteria grew in subsequent generations. R colonies, hence, retained the R character, whereas the transformed S colonies bred true as S cells. Chemical analysis showed the transforming principle was a deoxyribose-containing nucleic acid, and physical measurements showed it was a highly viscous substance having the properties of DNA. Because the transforming principle contained other kinds of molecules, it was necessary to provide evidence that the transformation was actually caused by DNA and not an impurity. This evidence was provided by the following three procedures:

1. Polysaccharides isolated from S cells did not transform R cells.
2. Incubation with trypsin or chymotrypsin, enzymes that catalyze protein hydrolysis, or with ribonuclease (RNase), an enzyme that catalyzes RNA hydrolysis, did not affect

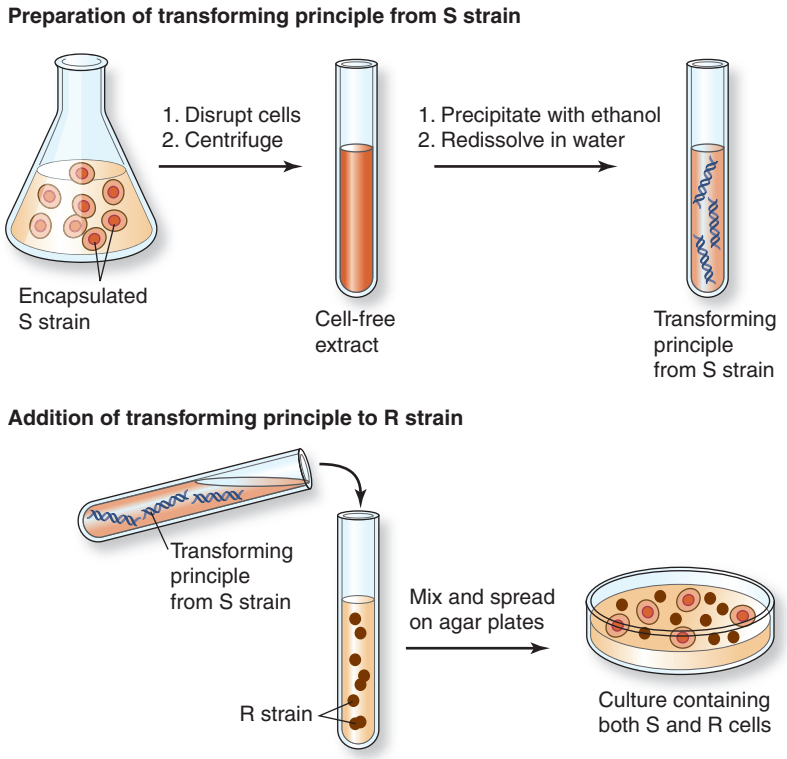


FIGURE 1.12 Avery, MacLeod, and McCarty transformation experiment.

transforming activity. The transforming principle thus is neither protein nor RNA.

3. Incubation with deoxyribonuclease (DNase), an enzyme that catalyzes DNA hydrolysis, inactivated the transforming principle.

In drawing conclusions from their experiments, Avery, MacLeod, and McCarty did not explicitly state that DNA is the hereditary material. One reason for their restraint may have been the crude analytical tools then available seemed to indicate that adenine, guanine, thymine, and cytosine were present in DNA in equimolar concentrations. Based on this and other evidence, many contemporary investigators mistakenly thought DNA consisted of a repeating tetranucleotide sequence and so could not possibly exhibit the variation needed by the hereditary material. Instead, the prevailing theory was that genetic information was stored in proteins because protein composition and structure were known to vary among organisms.

Investigators who supported the genes-as-protein theory posed two alternative explanations for the transformation results: (1) the transforming principle might not be DNA but rather one of the proteins invariably contaminating the DNA sample and (2) DNA somehow affected capsule formation directly by acting in the metabolic pathway for biosynthesis of the polysaccharide and permanently altering this pathway. The first point should have already been discounted by the original work because the experiments showed insensitivity to

proteolytic enzymes and sensitivity to DNase. Because the DNase was not a pure enzyme, however, the possibility could not be eliminated conclusively.

The transformation experiment was repeated 5 years later by Rollin Hotchkiss using a DNA sample with a protein content that was only 0.02%, and it was found that this extensive purification did not reduce the transforming activity. This result supported the view of Avery, MacLeod, and McCarty but still did not prove it. The second alternative, however, was clearly eliminated—also by Hotchkiss—with an experiment in which he transformed a penicillin-sensitive bacterial strain to penicillin resistance. Because penicillin resistance is totally distinct from the rough-smooth character of the bacterial capsule, this experiment showed the transforming ability of DNA was not limited to capsule synthesis. Interestingly enough, most biologists still remained unconvinced that DNA was the genetic material. It was not until Erwin Chargaff showed in 1950 that a wide variety of chemical structures in DNA were possible—thus allowing biological specificity—that this idea was accepted.

Chemical experiments also supported the hypothesis that DNA is the hereditary material.

The hypothesis of a tetranucleotide structure for DNA arose for two reasons. First, in the chemical analysis of DNA, the technique used to separate the bases before identification did not resolve them very well, so the quantitative analysis was poor. Second, the DNA analyzed was usually isolated from animals, plants, and yeast in which the four bases are present in nearly equimolar concentration or from bacterial species such as *Escherichia coli* that also happen to have nearly equimolar base concentrations. Using the DNA from a wide variety of organisms, Chargaff applied new separation and analytical techniques and showed that the molar content of bases (generally called the base composition) could vary widely. The base composition of DNA from a particular organism is usually expressed as a fraction of all bases that are G • C pairs. This fraction, called the G + C content, can be expressed as follows:

$$\text{G + C content} = ([\text{G}] + [\text{C}])/[\text{all bases}]$$

where the square brackets ([ ]) denote molar concentrations. Base compositions of DNAs from many different organisms have been determined. Generally speaking, the value of the G + C content is near 0.50 for the higher organisms, ranging from 0.49 to 0.51 for primates. For lower organisms the value of the G + C content varies widely from one genus to another. For example, for bacteria the extremes are 0.27 for the genus *Clostridium* and 0.76 for the genus *Sarcina; E. coli* DNA has the value 0.50. Thus, it was demonstrated that DNA could have variable composition, a primary requirement for the hereditary material.

Chargaff's studies also revealed one other remarkable fact about DNA base composition. In each of the DNA samples Chargaff studied, he found that [A] = [T] and [G] = [C]. Chargaff's findings can

be summarized by two rules, known as **Chargaff's rules**, which state: (1) double stranded DNA has equimolar adenine and thymine concentrations as well as equimolar guanine and cytosine concentrations; and (2) DNA composition varies from one genus to another. Although the significance of Chargaff's rules, was not immediately apparent, they would later help to confirm the structure of the DNA molecule.

Upon publication of Chargaff's results, the tetranucleotide hypothesis quietly died and the DNA–gene idea began to catch on. Shortly afterward, workers in several laboratories found for a wide variety of organisms that somatic cells have twice the DNA content of germ cells, a characteristic to be expected of the genetic material, given the tenets of classical chromosome genetics. Although it could apply just as well to any component of chromosomes, once this result was revealed objections to the work of Avery, MacLeod, and McCarty were no longer heard, and the hereditary nature of DNA rapidly became the accepted idea.

1.5 Watson-Crick Model

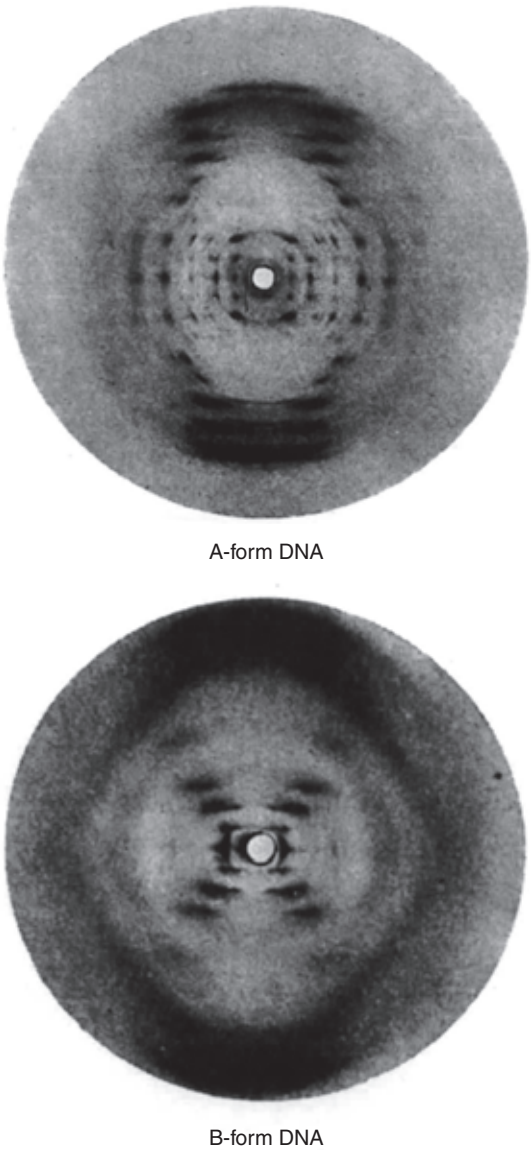
Rosalind Franklin and Maurice Wilkins obtained x-ray diffraction patterns of extended DNA fibers.

At the same time chemists were attempting to learn something about the composition of DNA, crystallographers were trying to obtain a three-dimensional image of the molecule. Rosalind Franklin and Maurice Wilkins obtained some excellent x-ray diffraction patterns of extended DNA fibers in the early 1950s (**FIGURE 1.13**). One might predict that all x-ray diffraction patterns would look alike, but this was not the case. DNA structure and therefore x-ray diffraction patterns depend on several variables. One of the most important of these is the relative humidity of the chamber in which DNA fibers are placed.

Two types of DNA structure are of particular interest. **B-DNA** is stable at a relative humidity of about 92%, whereas **A-DNA** appears as the relative humidity falls to about 75%. Crystallographers did not know whether A-DNA or B-DNA was present in the living cell. Partly for this reason, Wilkins turned his attention toward taking x-ray diffraction pictures of DNA in sperm cells. Franklin focused her attention on x-ray diffraction patterns of A-DNA because they appeared to provide more detail. She believed that careful analyses of the detailed patterns would eventually lead to the solution of DNA's structure.

James Watson and Francis Crick proposed that DNA is a double-stranded helix.

The American biologist, James D. Watson, and the English crystallographer, Francis H. C. Crick, working together in England, took a different approach to determining DNA's structure. They tried to obtain as much information as they could from the x-ray diffraction patterns and then to build a model consistent with this information.



A-form DNA

B-form DNA

FIGURE 1.13 X-ray diffraction patterns of the A and B forms of the sodium salt of DNA. (Reproduced from Franklin, R. E., and Gosling, R. G. 1953. *Acta Crystallographica* 6:673–677. Photos courtesy of International Union of Crystallography.)

The term "model" has a special meaning to scientists. A **model** is a hypothesis or tentative explanation of the way a system works, usually including the components, interactions, and sequences of events. A successful model suggests additional experiments and allows investigators to make predictions that can be tested in the laboratory. If predictions do not agree with experimental results, the model must be considered incorrect in its current form and modified. A model cannot be proved to be correct merely by showing it makes a correct prediction. If it makes many correct predictions, however, it is probably nearly, if not completely, correct.

Watson and Crick focused their attention on Franklin's x-ray diffraction patterns of B-DNA. This pattern indicated that B-DNA has a helical structure, a diameter of approximately 2.0 nm, and a repeat distance of 0.34 nm. Their model would have to account for these structural features. Watson and Crick still had to work out the number of DNA chains in a DNA molecule, the location of the bases, and the position of the phosphate and deoxyribose groups. The density of DNA seemed to be consistent with one, two, or three chains per molecule. Watson and Crick tried to build a two-chain model with hydrogen bonds (weak electrostatic attractions; see Chapter 2) holding the bases together. They were unsuccessful until their colleague, Jerry Donohue, suggested they use the *keto tautomeric* forms of T and G in their models. At first, Watson tried to link two purines together and two pyrimidines together in what he called a "like-with-like" model. A dramatic turning point occurred in 1953 when Watson realized adenine forms hydrogen bonds with thymine and guanine forms hydrogen bonds with cytosine. Watson describes this turning point in his book, *The Double Helix* (pages 194–195):

> When I got to our still empty office the following morning, I quickly cleared away the paper from my desk top so that I would have a large, flat surface on which to form pairs of bases held together by hydrogen bonds. Though I initially went back to my like-with-like prejudices, I saw all too well that they led nowhere. When Jerry [Donohue] came in I looked up, saw that it was not Francis [Crick], and began shifting the bases in and out of various other pairing possibilities. Suddenly I became aware that an adenine-thymine pair held together by two hydrogen bonds was identical in shape to a guanine-cytosine base pair held together by at least two hydrogen bonds. All the hydrogen bonds seemed to form naturally; no fudging was required to make the two types of base pairs identical in shape.

With the realization that adenine-thymine and cytosine-guanine base pairs have the same width, Watson and Crick were quickly able to construct a double helix model of DNA that fit Franklin's x-ray diffraction data (**FIGURE 1.14**).

The key features of the Watson-Crick Model for B-DNA are as follows:

1. Two polydeoxyribonucleotide strands twist about each other to form a double helix.
2. Phosphate and deoxyribose groups form a backbone on the outside of the helix.

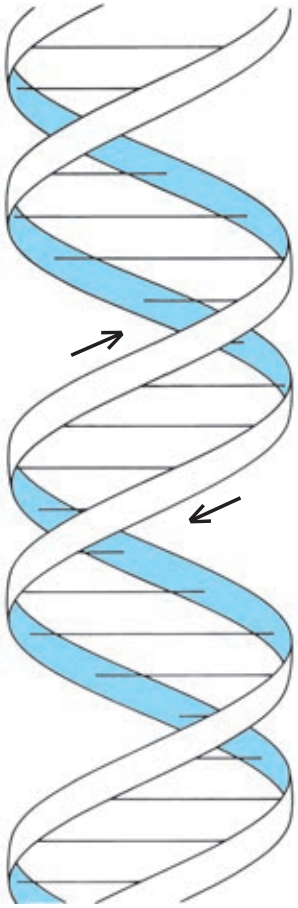


FIGURE 1.14 DNA structure based on the 1953 paper by Watson and Crick in *Nature* showing their double-helix model for DNA for the first time.

3. Purine and pyrimidine base pairs stack inside the helix and form planes perpendicular to the helix axis and the deoxyribose groups.
4. The helix diameter is 2.0 nm (or 20 Å).
5. Adjacent base pairs are separated by an average distance of 0.34 nm (or 3.4 Å) along the helix axis. The structure repeats itself after about 10 base pairs, or about once every 3.4 nm (or 34 Å) along the helix axis.
6. Adenine always pairs with thymine and guanine with cytosine. The original model showed two hydrogen bonds stabilizing each kind of base pair. Although this was an accurate description for A-T base pairs, later work showed that G-C base pairs are stabilized by three hydrogen bonds (**FIGURE 1.15**). (These base pairing relationships explained Chargaff's observation that the molar ratios of adenine to thymine and guanine to cytosine are one.)
7. The two strands are antiparallel, which means the strands run in opposite directions. That is, one strand runs 3′→5′ in one direction, whereas the other strand runs 5′→3′ in the same direction. Because the strands are antiparallel, a convention is needed for stating the sequence of bases of a single chain. The convention is to write a sequence with the 5′ terminus at the left; for example, ATC denotes the trinucleotide 5′-ATC-3′. This is also often written as pApTpC, again using the conventions that the left side of each base is the 5′-terminus of the nucleotide and that a phosphodiester group is represented by a "p" between two capital letters.
8. A **major groove** and a **minor groove** wind about the cylindrical outer helical face. The two grooves are of about equal depth, but the major groove is much wider than the minor groove.

The Watson-Crick model indicates that when the nucleotide sequence of one strand is known, the sequence of the complementary

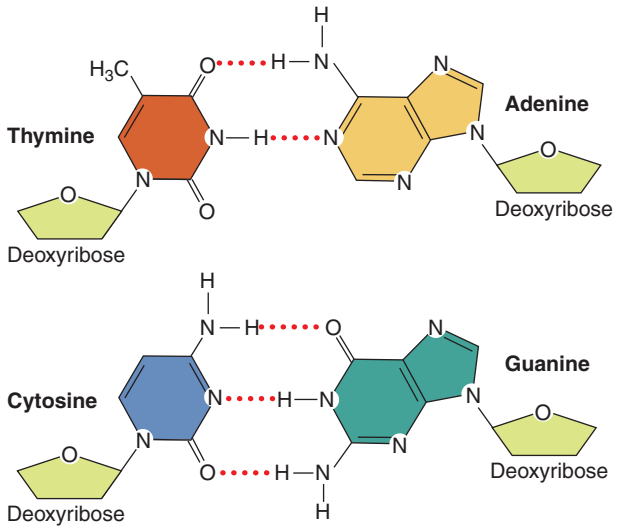


FIGURE 1.15 **Base pairs in DNA.**

strand can be predicted, providing the theoretical framework needed to understand the fidelity of gene replication. Each strand serves as a mold or **template** for the synthesis of the complementary strand (**FIGURE 1.16**). Watson and Crick ended their short paper announcing the double helix model with the following sentence that must be one of the greatest understatements in the scientific literature: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." The double helix showed how establishing a chemical structure can be used to understand biological function and to make predictions that guide new research.

Structure–function relationships remain a central theme of molecular biology. Knowledge of structure helps us to understand function

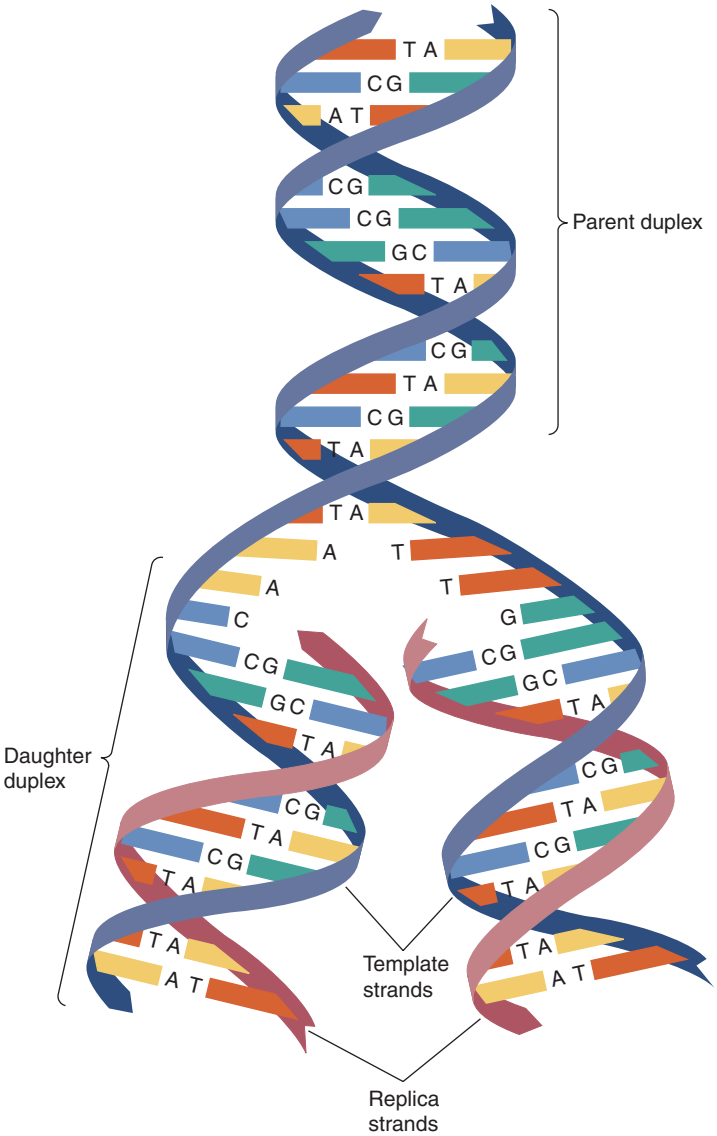


FIGURE 1.16 Replication of DNA. Replication of a DNA duplex as originally envisioned by Watson and Crick. As the parental strands separate, each parental strand serves as a template for the formation of a new daughter strand by means of A-T and G-C base pairing.

and leads us to new insights. Biological structures are sometimes quite complex and difficult to study. It is, however, worth the effort to study the structures because the reward for doing so is so great. The Watson-Crick model also serves as a powerful example of the fundamental principles that help to define molecular biology as a discipline: (1) the same physical and chemical laws apply to living systems and inanimate objects, (2) the same biological principles tend to apply to all organisms, and (3) biological structure and function are intimately related. The Watson-Crick model provided an excellent starting point for the challenging job of elucidating the chemical basis of heredity.

1.6 Central Dogma

The central dogma provides the theoretical framework for molecular biology.

Much of the research in molecular biology in the late 1950s was directed toward discovering the mechanisms by which nucleotide sequences specify amino acid sequences. In a talk given at a 1957 symposium, Crick suggested a model for information flow that provides a theoretical framework for molecular biology. The following excerpt from Crick's presentation ("On Protein Synthesis." *The Symposia of the Society for Experimental Biology* 12, (1958):138–163) states the theory known as the **central dogma** in a clear and concise fashion:

> In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the precise determination of sequence, either of bases in nucleic acid or of amino acid residues in protein.

Crick thus was proposing that genetic information flows from DNA to DNA (DNA replication), from DNA to RNA (transcription), and from RNA to polypeptide (translation) (**FIGURE 1.17**).

By the mid-1960s molecular biologists had obtained considerable experimental support for the central dogma. In particular, they had discovered enzymes that catalyze replication and transcription and elucidated the pathway for translating nucleotide sequences to amino acid sequences. With some variations in detail, all organisms use the same basic mechanism to translate information from nucleotide sequences to amino acid sequences. Three major components of the translation machinery—**messenger RNA** (mRNA), **ribosomes**, and **transfer RNA** (tRNA)—play essential roles in this information transfer (**FIGURE 1.18**). Each gene can serve as a template for the synthesis of specific mRNA molecules. mRNA molecules program ribosomes (protein synthetic factories) to form specific polypeptides. tRNA molecules carry activated amino acids to programmed ribosomes, where under the direction of the mRNA the amino acids join to form a polypeptide chain.

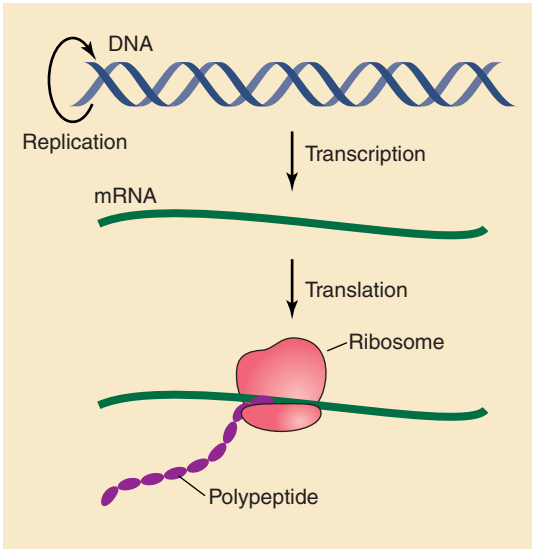


FIGURE 1.17 The "central dogma." The central dogma as originally proposed by Francis Crick postulated information flow from DNA to RNA to protein. The ribosome is an essential part of the translation machinery. Later studies demonstrated that information can also flow from RNA to RNA and from RNA to DNA (reverse transcription).

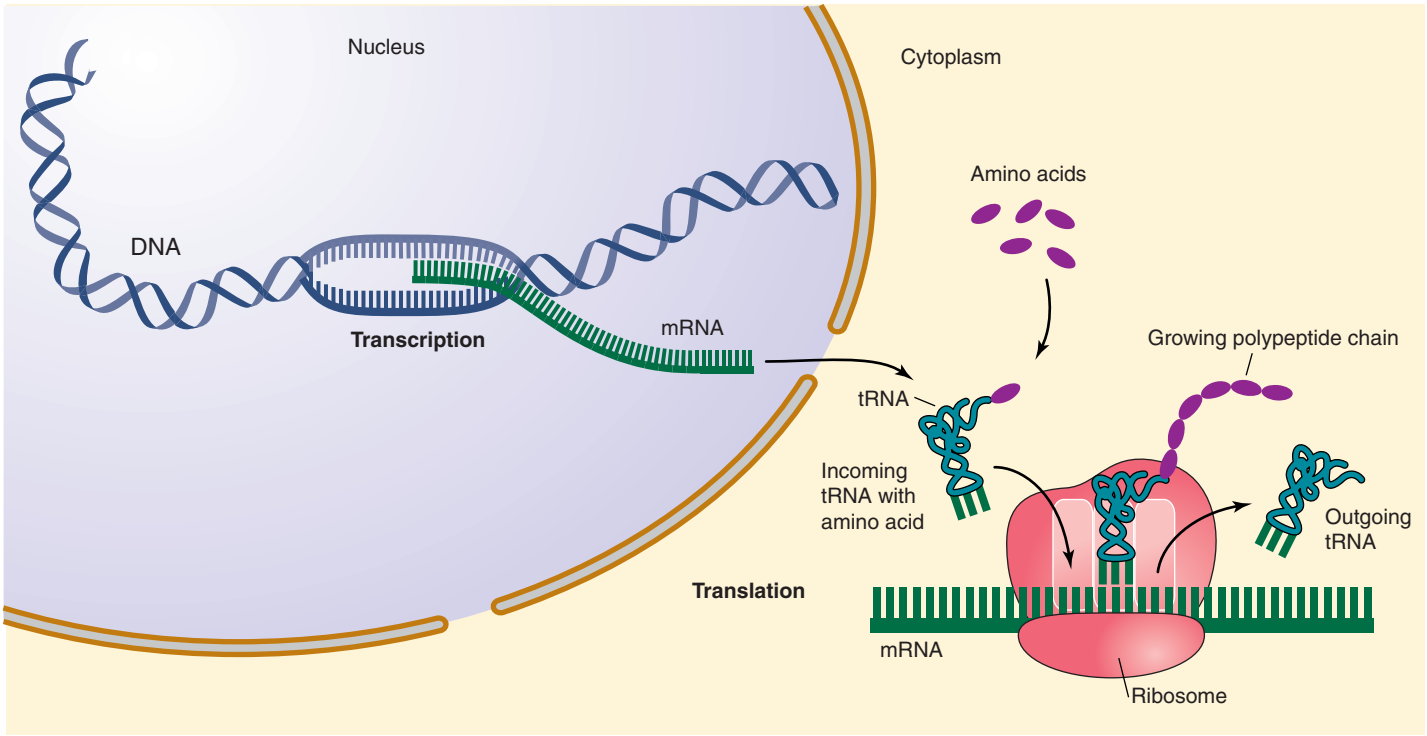


FIGURE 1.18 Simplified schematic diagram of the protein synthetic machinery in an animal, plant, or yeast cell. Transfer RNA (tRNA) carries the next amino acid to be attached to the growing polypeptide chain to the ribosome, which recognizes a match between a three nucleotide sequence in the mRNA (the codon) and a complementary sequence in the tRNA (the anticodon) and transfers the growing polypeptide chain to the incoming amino acid while it is still attached to the tRNA. (Adapted from Secko, D. 2007. *The Science Creative Quarterly 2*. http://www.scq.ubc.ca/a-monks-flourishing-garden-the-basics-of-molecular-biology-explained/, Figure 4.)

The 12 years of extraordinary scientific progress that followed the discovery of DNA's structure were capped by a series of brilliant investigations by Marshal W. Nirenberg and others that culminated in deciphering the genetic code. By 1965 investigators could predict a polypeptide chain's amino acid sequence from a DNA or mRNA molecule's nucleotide sequence. Remarkably, the genetic code was found to be nearly universal. Each particular sequence of three adjacent nucleotides or **codon** specifies the same amino acid in bacteria, plants, and animals. Could one ask for more convincing support for the uniformity of life processes?

1.7 Introduction to Recombinant DNA Technology

Recombinant DNA technology allows us to study complex biological systems.

The second major wave in molecular biology started in the late 1970s with the development of **recombinant DNA technology** (also known as **genetic engineering**). Thanks to recombinant DNA

technology genes can be manipulated in the laboratory just like any other organic molecule. They can be synthesized or modified as desired and their nucleotide sequence determined. Sequence information can save molecular biologists months and perhaps even years of work. For example, when a new gene is discovered that causes a specific disease in humans, molecular biologists may be able to obtain valuable clues to the new gene's function by comparing its nucleotide sequence with sequences of all other known genes. If similarities are found with a gene of known function from some other organism, then it is likely the new gene will have a similar function. Alternatively, a segment of the nucleotide sequence of the new gene may predict an amino acid sequence that is known to have specific binding or catalytic functions.

Development of recombinant DNA technology has allowed the incredible progress in solving problems that once seemed intractable. Recombinant DNA technology has helped investigators to study cell division, cell differentiation, transformation of normal cells to cancer cells, programmed cell death (apoptosis), antibody production, hormone action, and a variety of other fundamental biological processes.

One of the most exciting applications of recombinant DNA technology is in medicine. Until relatively recently many diseases could only be studied in humans because no animal models existed. Very little progress was made in studying these diseases because of obvious ethical constraints associated with studying human diseases. Once investigators learned how to modify and transfer genes, they were able to create animal models for human diseases, thus facilitating study of the diseases. Recombinant DNA technology also has led to the production of new drugs to treat diabetes, anemia, cardiovascular disease, and cancer as well as to the development of diagnostic tools to detect a wide variety of diseases. The list of practical medical applications grows longer with each passing day.

Although recombinant DNA technology promises to change our lives for the better, it also forces us to consider important social, political, ethical, and legal issues. For example, how do we protect the interests of an individual when DNA analysis reveals the individual has alleles likely to cause a serious physical or mental disease in the future, especially if there is no cure or treatment? What impact will the knowledge have on affected individuals and their families? Should insurance companies or potential employers have access to the genetic information? If not, how do we limit access to information and how do we enforce the limitation? Rapid progress in recombinant DNA technology also raises troubling ethical issues. Germ line therapy allows new genes to be introduced into fertilized eggs and thereby alter the genetic characteristics of future generations. This technique has been used to introduce desired traits into plants and animals.

An argument can be made for using germ line therapy to correct human genetic diseases, such as Tay-Sachs disease, cystic fibrosis, or Huntington disease, so affected individuals and their families can be spared the devastating consequences of these diseases. However, we must be very careful about application of germ line therapy to

humans because the technique has the power to do great harm. Who will decide which genetic characteristics are desirable and which are undesirable? Should the technique be used to change physical appearance, intelligence, or personality traits? Should anyone be permitted to make such decisions?

A great deal of molecular biology information is available on the Internet.

Recombinant DNA technology has generated so much information it would be nearly impossible to share all of it in a timely fashion with the entire molecular biology community by conventional means such as publishing in professional journals or writing books. Fortunately, there is an alternate method for sharing large quantities of rapidly accumulating information that is both quick and efficient. A worldwide network of communication networks, the Internet, allows us to gain almost instant access to the information. The Internet also provides many helpful tutorials and instructive animations. This text will include references to helpful Internet sites. Because the addresses for these Web sites tend to change over time, however, this text will refer the reader to a primary site maintained by the publisher.

Questions and Problems

1. Briefly define, describe, or illustrate each of the following terms.
   a. Allele
   b. Phenotype
   c. Genotype
   d. Dominant gene
   e. Recessive gene
   f. Purine
   g. Pyrimidine
   h. Nucleoside
   i. N-glycosylic bond
   j. Nucleotide
   k. Phosphomonoester
   l. Phosphodiester bond
   m. Transformation
   o. Transforming principle
   p. Chargaff's rules
   q. B-DNA
   r. Model
   s. Major groove
   t. Minor groove
   u. Replication
   v. Transcription
   w. Translation
   x. Codon
2. Draw the four bases commonly found in DNA. Which of these bases is a purine and which is a pyrimidine?
3. Draw the base that is commonly present in RNA but not in DNA. Is this base a purine or a pyrimidine?
4. What are the chemical similarities and differences between ribose and deoxyribose?
5. Each purine and pyrimidine base in nucleic acids has distinct chemical features.
   a. Which purine has a single amine group and no oxygen atoms?
   b. Which pyrimidine has a single amine group and one oxygen atom?
   c. What is the major difference between the structures of uracil and thymine?
   d. Which pyrimidine has two oxygen atoms attached to the ring structure?
6. Although Avery, MacLeod, and McCarty did not explicitly state that DNA is the hereditary material, their experiment certainly seemed to show this was the case.
   a. Describe their experiment.
   b. Why were their experiments showing that the transforming principle is resistant to RNase and proteases but sensitive to DNase important?
   c. Why were contemporary investigators slow to accept the idea that DNA is the hereditary material?

d. Why did Chargaff's studies of DNA base composition help to convince investigators that DNA is the hereditary material?
e. How does the Watson-Crick model explain Chargaff's rules?

7. DNA and RNA are linear nucleotide chains.
   a. Draw the extended structure for a segment of a polydeoxyribonucleotide with the abbreviated structure pdTpdApdC. Indicate the 5′- and 3′-ends.
   b. Draw the stick figure structure for a segment of a polydeoxyribonucleotide with the abbreviated structure pdTpdApdC. Indicate the 5′- and 3′-ends.
   c. Draw the extended structure for a segment of a polyribonucleotide with the abbreviated structure pUpApC. Indicate the 5′- and 3′-ends.
   d. Draw the stick figure structure for a segment of a polyribonucleotide with the abbreviated structure pUpApC. Indicate the 5′- and 3′-ends.
   e. Identify two important chemical differences between a polydeoxyribonucleotide and a polyribonucleotide strand.
8. What are the major structural features of the Watson-Crick model for DNA?
9. How does the Watson-Crick model help us to understand DNA function?
10. What functions do ribosomal RNA (rRNA), messenger RNA (mRNA), and transfer RNA (tRNA) play in the cell?

Suggested Reading

General

Choudri, S. 2003. The path from nuclein to human genome: a brief history of DNA with a note on human genome sequencing and its impact on future research in biology. *Bull Sci Technol Soc* 23:360–367.

Crick, F. 1974. The double helix: a personal view. *Nature* 248:766–769.

Crick, F. 1988. *What Mad Pursuit: A Personal View of Scientific Discovery*. New York: Basic Books.

Judson, H. F. 1996. *The Eighth Day of Creation: Makers of the Revolution in Biology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Maddox, B. 2002. *Rosalind Franklin: The Dark Lady of DNA*. New York: Harper Collins.

McCarty, M. 1985. *The Transforming Principle: Discovering That Genes Are Made of DNA*. New York: W. W. Norton.

Olby, R. C. 1994. *The Path to the Double Helix: The Discovery of DNA*. New York: Dover.

Pukkila, P. J. 2001. Molecular biology: the central dogma. *Encyclopedia of Life Sciences*. pp. 1–5. London, UK: Nature.

Sayre, A. 1978. *Rosalind Franklin and DNA*. New York: W. W. Norton.

Stent, G. 1972. Prematurity and uniqueness in scientific discovery. *Sci Am* 227:84–93.

Summers, W. C. 2002. History of molecular biology. *Encyclopedia of Life Sciences*. pp. 1–8. London, UK: Nature.

Watson, J. D. 1968. *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*. New York: Antheneum Books.

Wilkins, M. 2003. *The Third Man of the Double Helix: The Autobiography of Maurice Wilkins*. Oxford, UK: Oxford University Press.

Classic Papers

Avery, O. T., Macleod, C. M., and McCarty, M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *J Exp Med* 79:137–156.

Beadle, G. W., and Tatum, E., 1941. Genetic control of biochemical reactions in Neurospora. *Proc Natl Acad Sci USA* 27:499–506.

Watson, J. D., and Crick F. H. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171:737–738.