Concepts of Genetics

TWELFTH EDITION

Klug • Cummings • Spencer • Palladino • Killian

Brief Contents

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 UV 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 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 in eukaryotes, over a half century of research has provided irrefutable evidence that DNA serves this role. In fact, this simple concept of genetics is at the foundation of modern genetic research and its applications.

For example, **recombinant DNA technology**, which involves splicing together DNA sequences from different organisms (see Chapter 20), has combined the DNA sequence encoding the human hormone insulin with bacterial DNA sequences. When this recombinant DNA molecule is introduced into bacteria, the bacteria replicate the DNA and pass it to daughter cells at each cell division. In addition, the bacteria express the recombinant DNA molecule and thus produce human insulin. This example of biotechnology clearly establishes how a specific DNA sequence confers heritable information responsible for a product of a gene.

Genomics (see Chapter 21), which can provide the full set of DNA sequences of organisms, is the basis of still another example in support of this concept. We have known the full sequence of the human genome since 2001. However, geneticists are still uncovering new clues as to how the 3.2 billion base pairs of human DNA serve as the basis for human life. For example, the sequencing of genomes from individuals with specific heritable disorders and their comparison to genomes of healthy individuals has provided many insights about which DNA sequences (or genes) harbor mutations responsible for these genetic disorders. The underlying premise of such studies is that DNA is the genetic material.

By the mid-1970s the concept that DNA is the genetic material in eukaryotes was accepted, and since then, no information has been forthcoming to dispute that conclusion. In the upcoming chapters, we will see exactly how DNA is stored, replicated, mutated, repaired, and expressed.

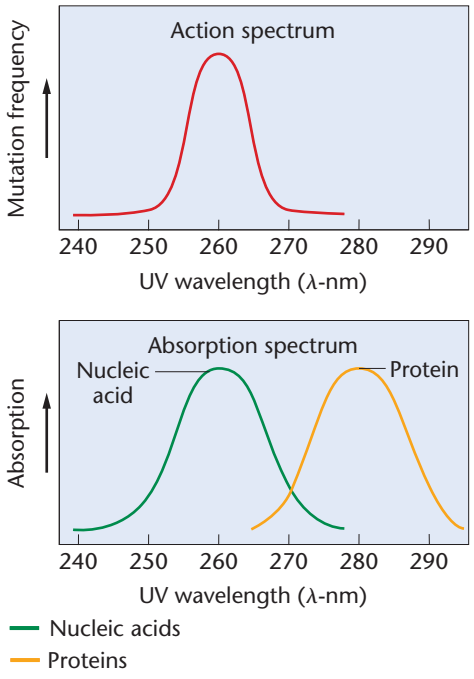


FIGURE 10.6 Comparison of the action spectrum (which determines the most effective mutagenic UV wavelength) and the absorption spectrum (which shows the range of wavelength where nucleic acids and proteins absorb UV light).

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 Qβ 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 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 **guanine** and **adenine,** abbreviated **G** and **A**. The three 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

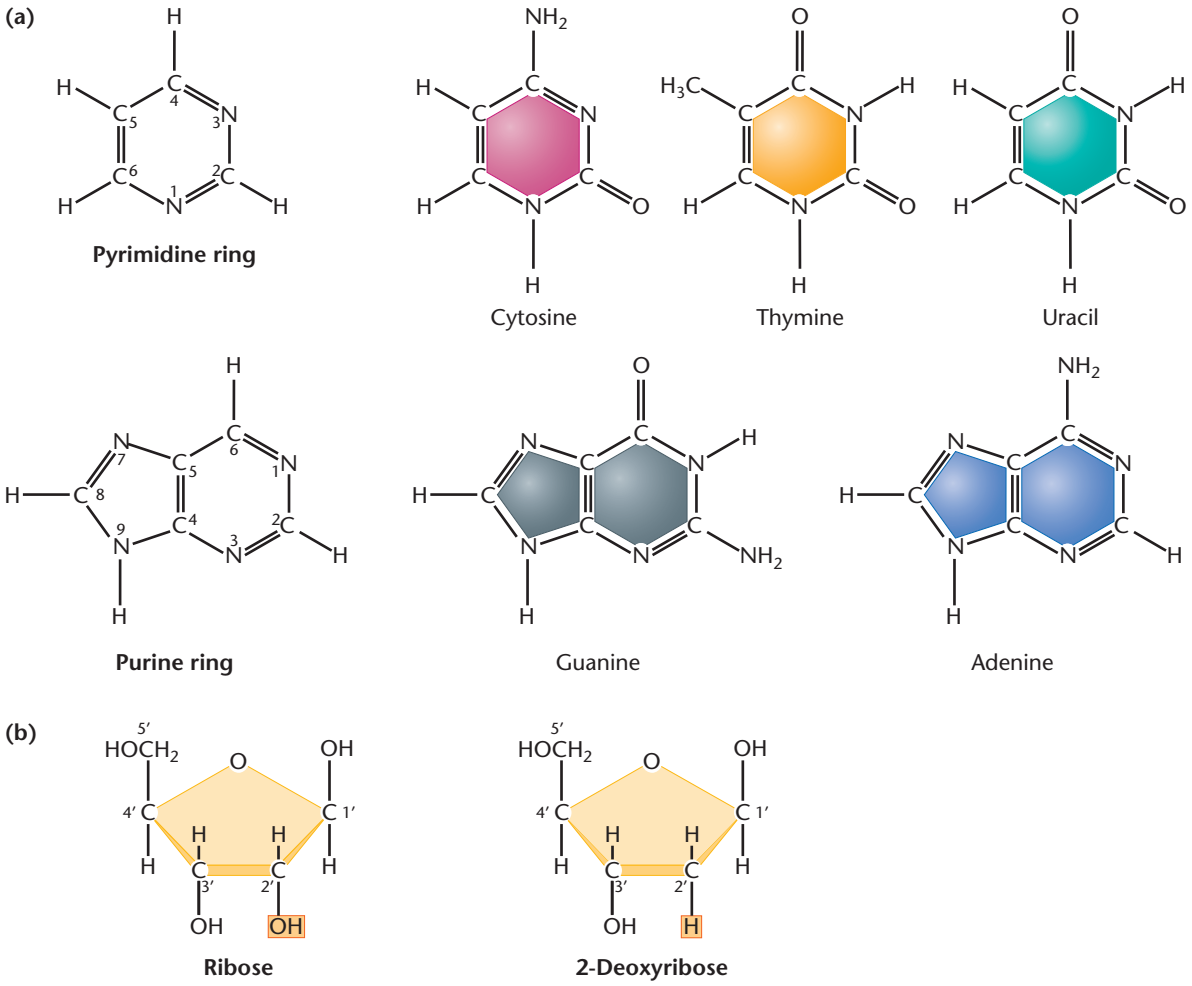


FIGURE 10.7 (a) Chemical structures of the pyrimidines and purines that serve as the nitrogenous bases in RNA and DNA. The convention for numbering carbon and nitrogen atoms making up the two categories of bases is shown within the structures that appear on the left. (b) Chemical ring structures of ribose and 2-deoxyribose, which serve as the pentose sugars in RNA and DNA, respectively.

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.

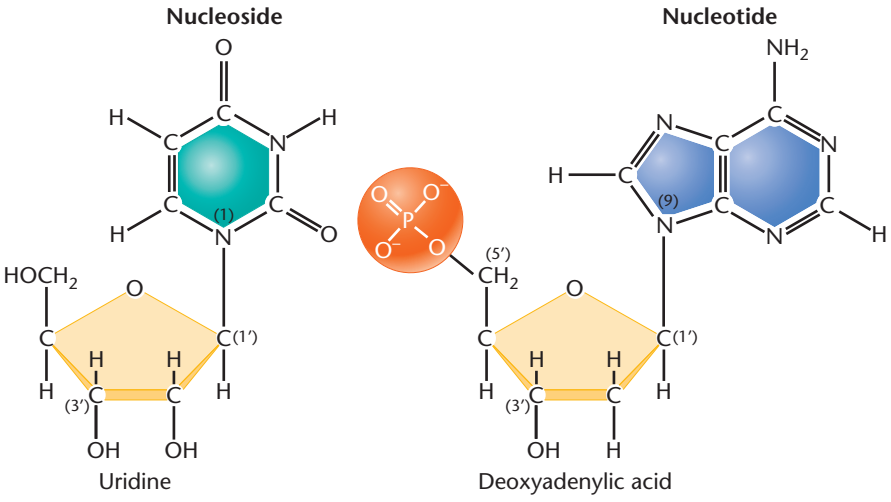
The pentose sugars found in nucleic acids give them their names. Ribonucleic acids (RNA) contain **ribose,** while deoxyribonucleic acids (DNA) contain **deoxyribose**. **Figure 10.7(b)** shows the ring structures for these two pentose sugars. Each carbon atom is distinguished by a number with a prime 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 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



Ribonucleosides	Ribonucleotides
Adenosine Cytidine Guanosine Uridine	Adenylic acid Cytidylic acid Guanylic acid Uridylic acid
Deoxyribonucleosides	**Deoxyribonucleotides**
Deoxyadenosine Deoxycytidine Deoxyguanosine Deoxythymidine	Deoxyadenylic acid Deoxycytidylic acid Deoxyguanylic acid Deoxythymidylic acid

FIGURE 10.8 Structures and names of the nucleosides and nucleotides of RNA and DNA.

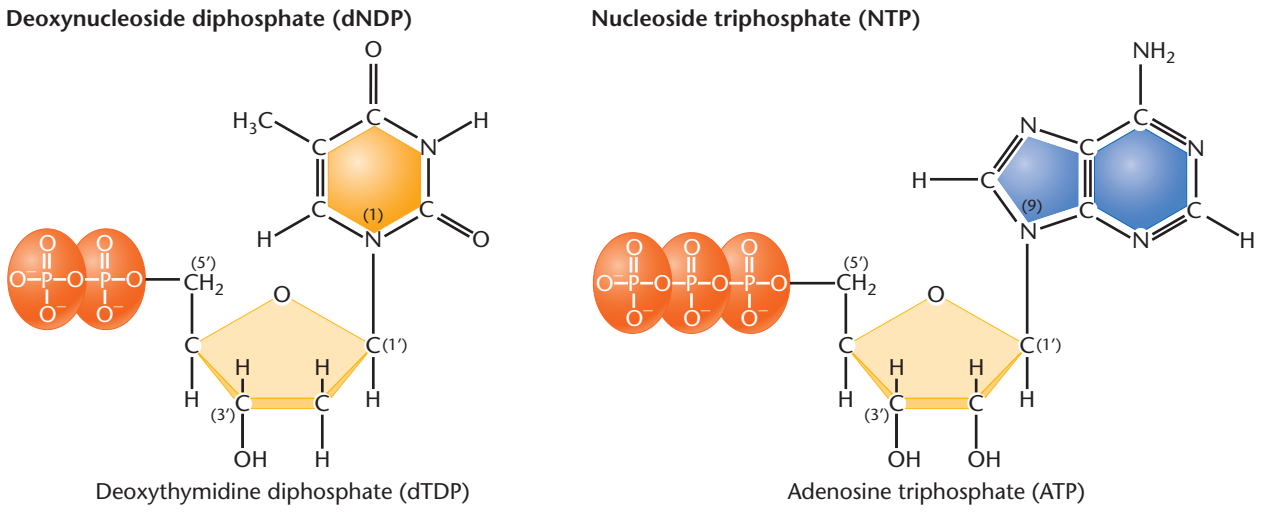


FIGURE 10.9 Structures of nucleoside diphosphates and triphosphates. Deoxythymidine diphosphate and adenosine triphosphate are diagrammed here.

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.

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.

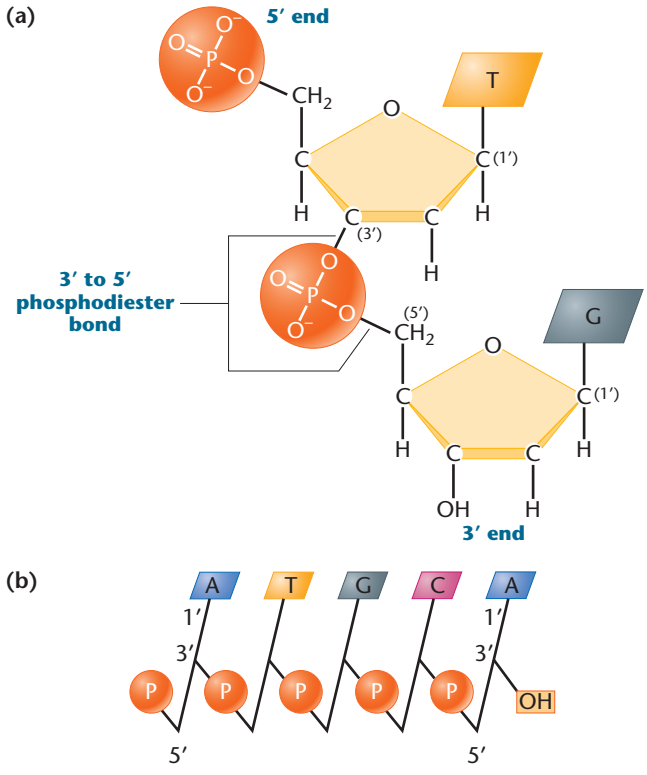


FIGURE 10.10 (a) Linkage of two nucleotides by the formation of a C-3′-to-C-5′ (3′-to-5′) phosphodiester bond, producing a dinucleotide. (b) Shorthand notation for a polynucleotide chain.

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*. 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.

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 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

TABLE 10.3 DNA Base-Composition Data

(a) Chargaff's Data*

	Molar Proportions[a]			
	1	2	3	4
Organism/Source	A	T	G	C
Ox thymus	26	25	21	16
Ox spleen	25	24	20	15
Yeast	24	25	14	13
Avian tubercle bacilli	12	11	28	26
Human sperm	29	31	18	18

(b) Base Compositions of DNAs from Various Sources

	Base Composition				Base Ratio		Combined Base Ratios	
	1	2	3	4	5	6	7	8
Organism	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
E. coli	24.7	23.6	26.0	25.7	1.04	1.01	1.03	0.93
Sarcina lutea	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

(c) G + C Content in Several Organisms

Organism	%G + C
Phage T2	36.0
Drosophila	45.0
Maize	49.1
Euglena	53.5
Neurospora	53.7

*Source: Data from Chargaff, 1950.

[a]Moles of nitrogenous constituent per mole of P. (Often, the recovery was less than 100 percent.)

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.

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. 265.

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.

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

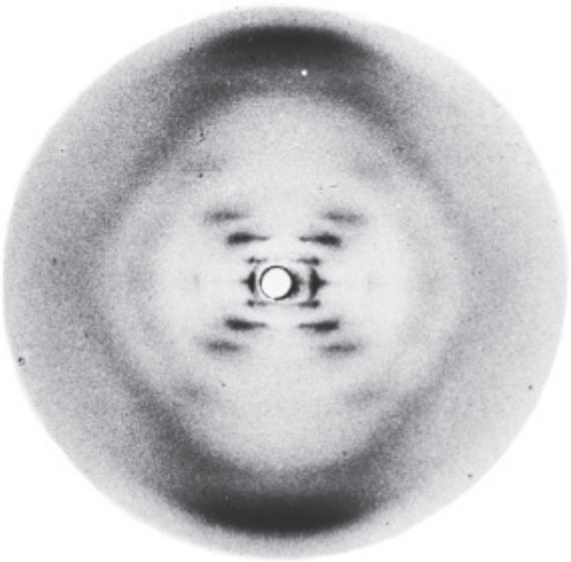


FIGURE 10.11 X-ray diffraction photograph by Rosalind Franklin using the B form of purified DNA fibers. The strong arcs on the periphery represent closely spaced aspects of the molecule, allowing scientists to estimate the periodicity of the nitrogenous bases, which are 3.4 Å apart. The inner cross pattern of spots reveals the grosser aspects of the molecule, indicating its helical nature.

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 C-G 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 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