国外大学生物学优秀教材（影印版）

Introduction to Molecular Biology

Peter Paolella

分子生物学导论

清华大学出版社
麦格劳-希尔教育出版集团

INTRODUCTION TO MOLECULAR BIOLOGY

分子生物学导论

Peter Paolella

清华大学出版社

麦格劳-希尔教育出版集团

（京）新登字 158 号

Introduction to Molecular Biology

Peter Paolella

北京市版权局著作权合同登记号：图字 01-2002-0222

书　名：分子生物学导论
作　者：Peter Paolella
出版者：清华大学出版社（北京清华大学学研大厦，邮编 100084）
　　http:// www.tup.tsinghua.edu.cn
印刷者：世界知识印刷厂
发行者：新华书店总店北京发行所
开　本：850×1168 1/16　印张：16
版　次：2002 年 2 月第 1 版　2002 年 2 月第 1 次印刷
书　号：ISBN 7-302-05095-3/Q・19
印　数：0001～3000
定　价：38.00 元

出 版 前 言

为了使生物学教学适应21世纪生命科学发展的需要，同时也为了提高学生阅读专业文献和获取信息的能力，结合当前生物学在高等院校中教学的实际情况，我们精选了一些国外优秀的生物学教材，组织专家进行了评阅和审核，组成国外大学生物学优秀教材系列（影印版）。该系列反映了国外大学生物学教材的最新内容和编写特色，多数教材经过教学实践，被国外很多大学广泛采用，并获得好评，因而不断再版。本书即是其中的一册。

希望这套教材能对高等院校师生和广大科技人员有所帮助，同时对我国的生命科学赶超世界先进水平起到一定的推动作用。

欢迎广大读者将使用本系列教材后的意见反馈给我们，更欢迎国内外专家、教授积极向我社推荐国外的优秀生物学教材，以便我们将国外大学生物学优秀教材系列做得更好。

清华大学出版社

2001年8月

To my wife, Cees, and our children,
Christopher, Andreana, and Eileen
for their love, patience,
and understanding.

A Note from the Editor

It is with great pleasure that I am able to write this note celebrating the publication of *Introduction to Molecular Biology*. Peter would be so proud. We lost our dear friend and colleague, Peter Paolella, in the spring of 1996, only a month after the manuscript for *Introduction to Molecular Biology* was put into production. His love of teaching and passion for writing (he was a novelist at heart) are the enduring qualities found in this text. The devotion and hard work of Peter's friends at Stonehill College carried the book through production and to fruition. It is with my deepest appreciation that I thank Robert Peabody, Craig Almeida, Sandra McAlister, and Maura Geens Tyrrell, all of Stonehill College, for carrying on Peter's dream by taking the manuscript through production—it was a labor of love. I am also greatly indebted to Cecile Paolella, Peter's wife, for putting me in contact with his colleagues at Stonehill College. Without her help, I would never have found such a caring group of individuals who would preserve Peter's heart and soul in this book. Thank you all, and finally, thank you, Peter.

Elizabeth M. Sievers
Editor

PREFACE

This book is aimed at undergraduate biology or science majors, sophomores to seniors, who are not genetics majors. It is intended for a one-semester introductory course for students requiring a knowledge of molecular biology as a basis for the study of other areas of biology. A beginning biology course and a basic organic chemistry course are the only prerequisites.

In the text, molecular genetics is placed within the general framework of metabolism. In this way, the processes and mechanisms of gene function and control are seen as phenomena connected to other cell activities.

The text begins with the structure of genetic material, moves to the function of that material (e.g., expression and regulation of genes), and concludes with the application of the knowledge gained in the study of the first two sections in the relatively new field of genetic engineering, or gene manipulation. In its entirety, then, the text is concerned with the structure, expression, regulation, and manipulation of genetic material, in particular, DNA.

Theoretical considerations and experimental data, which are better left to courses for genetics majors, are kept to a minimum. Instead, the idea is to give the student a frame of reference to either continue the study of molecular biology or genetics or to use the information acquired to further an understanding of other courses. When, however, the experimental evidence is crucial to an understanding of how advances in the field occurred, the experiments are described in some detail (e.g., the Meselson-Stahl experiments proving the semiconservative model of DNA replication).

No book is the product of a single person's labors. I have had the assistance of numerous reviewers. Their comments and recommendations have been carefully considered, though not always followed. I am, nonetheless, grateful to them.

Helen H. Benford
Tuskegee University

Douglas J. Burks
Wilmington College

Richard Crawford
Trinity College

Christopher Cullis
Case Western Reserve University

Beth DeStasio
Lawrence University

Carolyn Jones
Vincennes University

Karen Kurvink
Moravian College

Joel Piperberg
Millersville University

Roger Sloboda
Dartmouth College

Steven Woeste
Scholl College

Last, I thank Meg Johnson, who first showed an interest in this project and ultimately got it approved by Wm. C. Brown, and thanks to Robin Steffek, my most capable and enthusiastic editor, who guided me to the completion of the book with as little pain as possible.

Peter Paolella
Johnston, Rhode Island

CONTENTS

CHAPTER
ONE

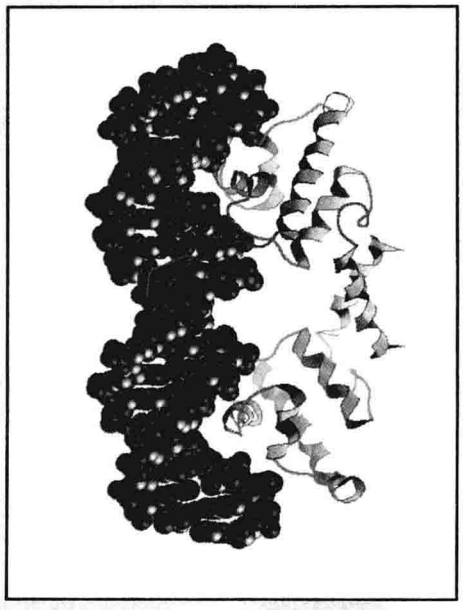
THE NATURE OF DNA

OUTLINE

CHAPTER OBJECTIVES

This chapter will discuss:

- The contributions of various scientists to our understanding of the nature of DNA
- The chemical and physical characteristics of DNA
- The biological functions of genetic material
- How to calculate the number of nucleotides in a DNA molecule
- The nature of the gene



INTRODUCTION

On April 25, 1953, in the British journal *Nature*, a paper, two columns in length, appeared. It was entitled "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid" and was authored by the American James D. Watson and the Englishman Francis H. C. Crick (see Figure 1.1). The structure they proposed has, they say in the first paragraph, "novel features which are of considerable biological interest." And at the end of the paper is the statement, "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

This paper was the culmination of work that stretched back 85 years to Friedrich Miescher, the German scientist who had reported his discovery of a nucleic acid. He called it nuclein because it was isolated from nuclei of pus cells and salmon sperm. Miescher reported his findings in 1871. In 1866, Gregor Mendel had published his work that led to the principles of independent segregation and assortment of genes.

The late 1800s are considered the time of the birth of genetics. And at its birth, the new science was already started in two directions. Mendel's work would lay the foundation of what has been called classical genetics, and Miescher's had begun what is now called molecular genetics. The two scientists apparently worked without knowledge of the other's discoveries.

The classical geneticists have focused on how genes are transferred from one generation to the next **(inheritance),** gene location within **chromosomes,** chromosomal rearrangements, and the concept of dominance. The molecular geneticists, on the other hand, have focused on the structure of genes and on how genes work and are regulated.

For nearly a century, the work of trying to elucidate the structure of Miescher's nuclein went on. Nuclein's role in inheritance and in the metabolism of the cell was not universally accepted. In fact, once the two components of chromosomes, nucleic acid and **protein,** were discovered, it was argued that, of the two, the nucleic acid was much too simple a component to explain the differences between chromosomes and, thus, the differences between individuals whether of the same or of different species. Rather, the protein component, itself composed of a larger variety of subunits, the **amino acids,** was thought to more easily account for the differences between chromosomes and individuals.

Watson and Crick's paper ended one search while simultaneously beginning another. The search for the molecular structure of inheritance had ended; the search for the molecular functions of the nucleic acid had begun. One of these functions, **replication,** by which information is passed on to the next generation was proposed in the paragraph quoted earlier: "It has not escaped our notice . . . " The other functions—how genes work and how they are regulated as an integral part of the cell's metabolism—are still under intense study.

After Miescher's discovery of nuclein (DNA), another 60 years were to elapse before the basic components of DNA were identified and their relationships to one another determined.

At the turn of the century, A. Kossel had demonstrated that a nucleic acid was composed of a nitrogenous base **(adenine, guanine, cytosine, thymine,** or **uracil)** (Figure 1.2) a sugar, and a phosphate group (Figure 1.3).

Then, by the early 1930s, largely as the result of the work of P. A. T. Levene, the arrangement of the bases, sugar, and phosphate was discovered. A single base is linked to the sugar, which in turn is linked to the phosphate. The resulting structure is a **nucleotide** (Figure 1.4), the fundamental unit of nucleic acids.

Levene, along with other workers, also discovered that the sugar of nuclein is **deoxyribose.** And he discovered that there are in fact two nucleic acids: **ribonucleic acid,** or **RNA** (actually discovered by Kossel), and **deoxyribonucleic acid,** or **DNA** (Miescher's nuclein).

The discovery of the components of nucleic acids, in particular DNA, then led to the first models of the structure of DNA. Takahashi, in 1930, proposed the "tetranucleotide" structure for DNA. In this model, the nucleotides of adenine, guanine, cytosine, and thymine repeat in a regular pattern. Thus, the idea that DNA is composed of simple parts arranged in a simple way was born.

The year 1953 proved to be the year of the structure of DNA. In addition to James Watson and Francis Crick, L. Linus Pauling and R. Corey, and Fraser also proposed models. Both these attempts, however, consisted of three intertwining chains. The Pauling and Corey model also suggested that the phosphates were located along the axis of the molecule and the bases were placed on the outside. Although Fraser had positioned the phosphates and bases correctly, Watson and Crick considered the model to be "rather ill-defined."

THE PATH TO THE WATSON AND CRICK MODEL

As we have seen, by the 1930s, much evidence had accumulated regarding the components of DNA, but what was it that was driving scientists in the mid-twentieth century to search for the structure of DNA? The answer is: there was a preponderance of evidence pointing to DNA as the genetic material.

As early as the 1880s, there had been speculation that the chromosomes were involved in inheritance. Wilhelm Roux suggested that it was unlikely that the mechanisms of **mitosis** and **meiosis** had evolved without some good reason. He proposed the chromosomes as the genetic material.

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

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

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

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

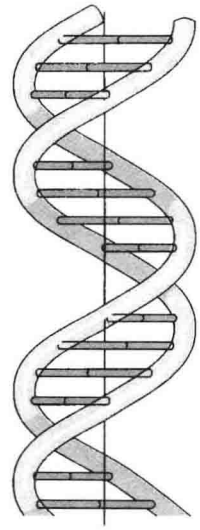
We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate di-ester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

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

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

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

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



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

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

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

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

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on inter-atomic 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.
April 2.

[1] Pauling, L., and Corey, R. B. *Nature*, 171, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, 38, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, 6, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawernian, G., and Chargaff, E., *Biochim. et Biophys. Acta*, 9, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physici.*, 36, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, 10, 182 (1953).

FIGURE 1.1
Watson and Crick's 1953 article on the structure of DNA.

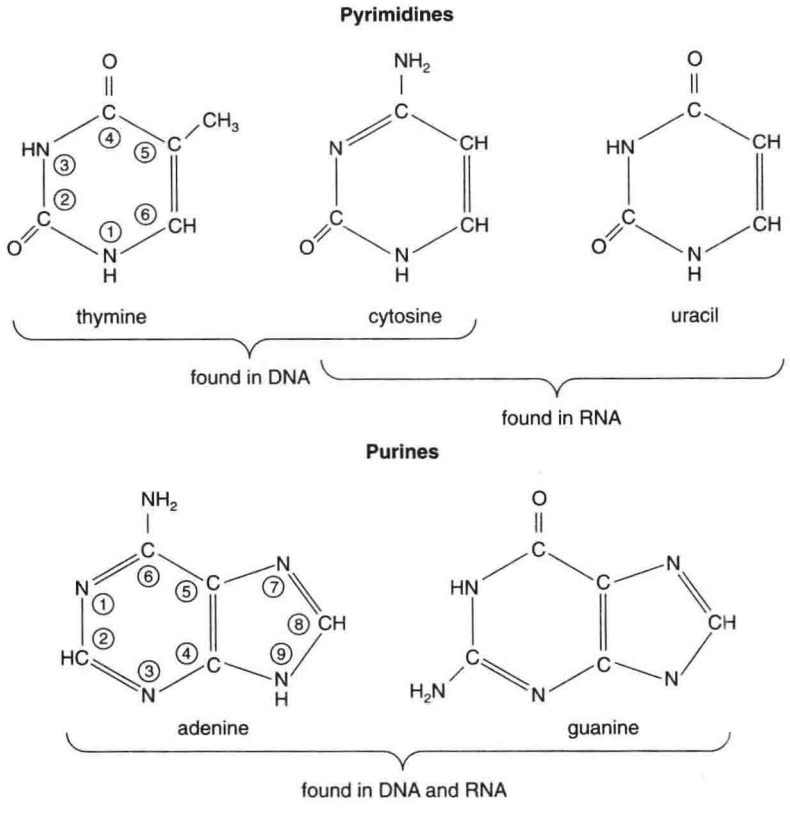


FIGURE 1.2
The nitrogenous bases of nucleic acids. Pyrimidines contain one carbon-nitrogen ring; purines contain two rings.

And, apparently with no knowledge of Mendel's work, he proposed a "linear arrangement of hereditary units along the chromosomal threads."

But not until Frederick Griffith's 1928 discovery of transformation in pneumococci was there experimental evidence pointing to DNA as the genetic material. Griffith was an English physician working in the British Public Health Service. His work laid the foundation for the investigations of other scientists including O. T. Avery and coworkers, E. Chargaff, A. D. Hershey and M. Chase, and Franklin, about whom we will hear later.

Let's now take a brief look at the experimental evidence produced by these workers to see how it led to the Watson and Crick model of the structure of DNA and to the hypothesis of a "possible copying mechanism."

Griffith worked with the pneumococcal bacteria. These organisms come in a variety of "types" that are distinguishable by the nature of the polysaccharide that surrounds the cell. The types are designated by roman numerals. This enveloping polysaccharide is called the capsule. Whether a particular bacterium is encapsulated determines whether it will be able to produce a pneumonia that can kill its host. Cells surrounded by a capsule are virulent (able to produce pneumonia); those without the capsule are not.

The capsular material enables an investigator to quickly determine whether a colony growing on agar media is virulent or avirulent (unable to produce pneumonia). The virulent encapsulated cells produce colonies that have a smooth, mucoid appearance; these are designated *S* colonies. Nonencapsulated, avirulent cells, on the other hand, produce colonies that have a rough, dry appearance; these are designated *R* colonies.

Griffith's experiments required injecting mice with the different cell types and, after a day or so, sampling the heart blood of dead mice to determine what organisms were present. The experiments can be summarized as follows:

1. Injection with type$_{III}$ S cells produced dead mice + Type$_{III}$ S cells.
2. Injection with type$_{II}$ R cells produced live mice.
3. Injection with heat-killed type$_{III}$ S cells produced live mice.
4. Injection with heat-killed type$_{III}$ S cells + type$_{II}$ R cells produced dead mice + type$_{III}$ S cells.

Griffith concluded that the presence of the heat-killed S cells must have caused the "transformation" of the living R cells. The living R_{II} cells were transformed into S_{III} cells. **Mutation** did not explain the results, because when cells

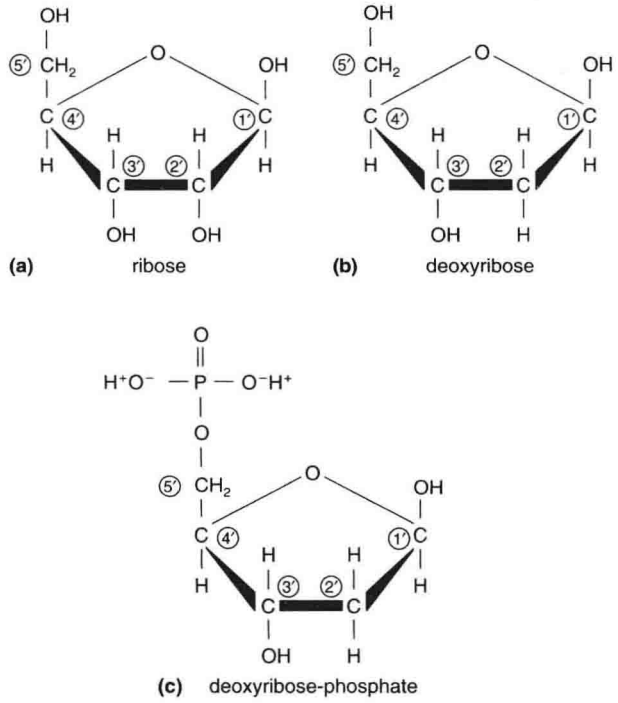


FIGURE 1.3
(a) The sugar ribose has an —OH linked at the 2′ carbon. (b) the sugar deoxyribose has a single hydrogen at the 2′ carbon. (c) Deoxyribose with a phosphate group.

mutated from the R to the S form or from the S to the R form the capsular type was not lost. For example, a mutation would cause R_{II} cells to become S_{II} cells not S_{III} cells.

Griffith's results were published in the British *Journal of Hygiene* in 1928. A few years later, J. L. Alloway reported that crude extracts of S cells from which all cells were removed by filtration was capable of transforming R cells.

Then, in 1944, O. T. Avery, C. M. MacLeod, and M. McCarty published the results of their experiments. The results were, in Watson's phrase, "Avery's bombshell." The conclusion "met with great surprise and disbelief, because at that time hardly anyone was prepared to accept such an informational role for DNA." Avery and his colleagues had demonstrated that Griffith's transforming principle was DNA.

Avery, Macleod, and McCarty used a variety of enzymatic reactions designed to destroy particular kinds of molecules. They reported that "only those preparations shown to contain an enzyme capable of depolymerizing authentic samples of desoxyribonucleic [*sic*] acid were found to inactivate the transforming principle" (Avery et al. 1944). By the use of chemical, enzymatic, and serological reactions, electrophoresis, ultracentrifugation, and UV (ultraviolet) spectroscopy, they could eliminate protein, unbound lipid, RNA, and polysaccharides as the transforming principle.

It might be assumed that this work led to universal acceptance of DNA as the genetic material. But there was still some criticism that Avery's preparations might have been

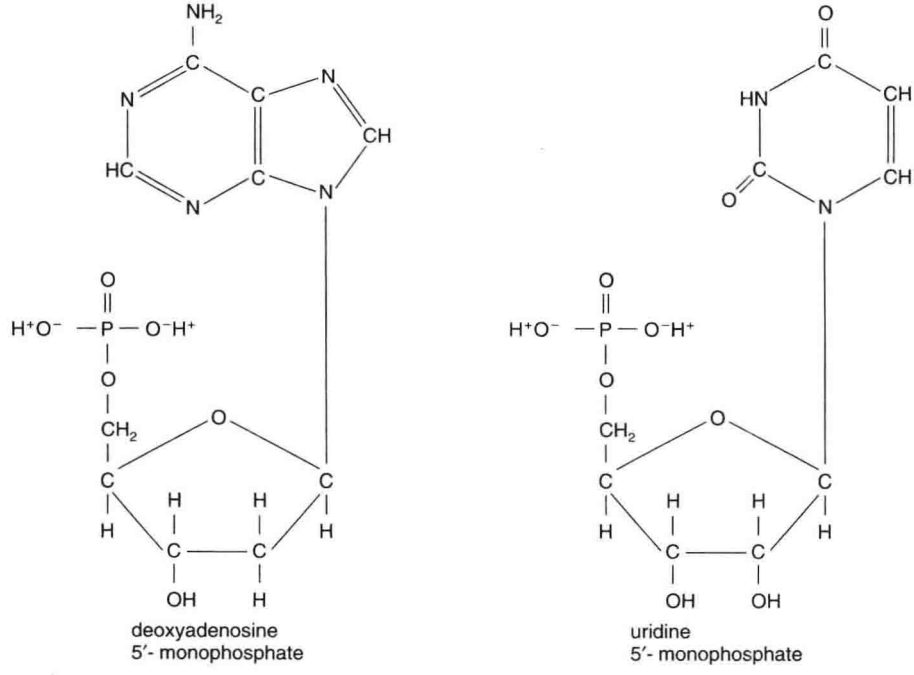


FIGURE 1.4
Nucleotides.

contaminated by undetectable amounts of protein. In fact, Avery, MacLeod, and McCarty concluded their paper with the statement: "*If* the results of the present study on the chemical nature of the transforming principle are confirmed, then nucleic acids must be regarded as possessing biological specificity the chemical basis of which is as yet undetermined" [emphasis added]. Their work and data were, nonetheless, a stimulant for the biochemist Erwin Chargaff.

A technique called paper chromatography is used to determine the identity of unknown materials. Substances dissolved in various solvents move up or down paper strips; the distance they move relative to the starting point and to one another controls their identities.

Chargaff gathered DNA from several sources, subjected the samples to acid hydrolysis at high temperatures and high pressures to release the adenines, guanines, cytosines, and thymines. He identified the compounds by means of paper chromatography and determined the amounts of each in the 2–40 μg range using absorption spectra techniques.

In 1950, he published his results. Chargaff noted that DNA extracted from different species was "of a composition constant for different organs of the same species and characteristic of the species." Further, he said, "The results serve to disprove the tetranucleotide hypothesis." That idea required that the amounts of adenine, guanine, cytosine, and thymine be equal.

Instead, he wrote, "It is noteworthy that in all the desoxypentose [*sic*] nucleic acids examined thus far the molar ratios of total **purines** to total **pyrimidines,** and also of adenine to thymine and of guanine to cytosine were not far from 1." This observation has come to be known as **Chargaff's rule.**

This was the first published report of so important a structural feature of DNA since Levene's work on nucleotides in the 1930s. Still, the idea that the genetic material might be protein and not DNA would not die.

The research of Alfred Hershey and Martha Chase finally put that notion to rest. Their experiments culminated in a paper published in 1952 in *The Journal of General Physiology* entitled "Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage." This may have been the paper that spurred Watson and Crick to redouble their efforts to propose a model of DNA in 1953.

A number of bacterial viruses, called bacteriophage, or simply phage, were known to be constructed of two macromolecules, protein and DNA. Like all viruses, bacteriophage are intracellular parasites; they can replicate only within a living cell. The viruses first adsorb to the cell surface and then penetrate the cell. Once inside, however, the viruses are noninfective. That is, shortly after infection of the cell, if the cells are ruptured, no viral particles capable of infecting other cells can be found. This noninfective, intracellular form was seen "as the connecting link between parental and progeny phage, and the elucidation of its structure and function became the central problem in the study of viral growth."

Hershey and Chase hit on the idea of using bacteriophage T2 grown in such a way that both the protein and DNA would be made radioactive but with different radioisotopes. They were able to do this because the viral proteins contain sulfur in the amino acids methionine and **cysteine,** and the viral DNA contains phosphorus. What is also critical is that the proteins of T2 contain no phosphorus and its DNA contains no sulfur.

The radioactive T2 phage was allowed to infect its host cell, the bacterium ***Escherichia coli*** (abbreviated ***E. coli***), and at a predetermined time the cells were put into a Waring blender and subjected to violent shearing forces in order to release the attached viruses.

The investigators then checked for the following:

1. The ability of the infected cells to yield viral progeny: it was unaffected, so the intracellular form was present and functioning.
2. The removal of sulfur: the shearing forces stripped away 70%–80% of the attached radioactive sulfur from infected cells.
3. The removal of phosphorus: only 20%–35% of the radioactive phosphorus was stripped from the bacteria. (This result can be explained by the fact that not all bacteriophage succeed in penetrating the host cell. So some radioactive phosphorus is expected to be left on the cell surface and subsequently released by the shearing forces.)

Hershey and Chase interpreted these results to mean that the viral DNA leaves the protein coat of the virus and enters the cell. To bolster this interpretation, they allowed viral particles to adsorb to isolated bacterial membranes or to heat-killed bacteria and showed that adsorption to both caused release of DNA. The nucleic acid was identified by means of its susceptibility to the enzyme DNase.

From these experiments, Hershey and Chase (1952) concluded "that one of the first steps in the growth of T2 is the release from its protein coat of the nucleic acid of the viral particle, after which the bulk of the sulfur-containing protein has no further function."

In other words, DNA, not protein, is the genetic material of T2. The argument was at last settled in a decisive manner in favor of DNA. From this time on, no one again seriously questioned the genetic role of DNA.

The last piece of evidence Watson and Crick would need to convince themselves that they were on the right track was the work of Rosalind Franklin. She used a technique first demonstrated in 1913 by W. H. Brag and

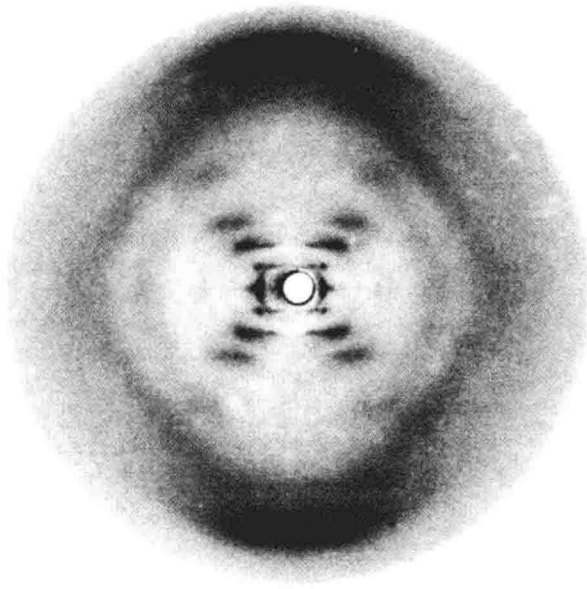


FIGURE 1.5
X-ray diffraction photograph of DNA, by Franklin.
Courtesy M. H. F. Wilkins, Kings College, London.

W. L. Brag, who showed that x-ray diffraction patterns can be used to reveal the three-dimensional structure of molecules. The technique is called x-ray crystallography. Crystalline forms of substances are subjected to x-ray beams, which are then deflected from the crystal onto a film surface. The film is, thus, exposed in a pattern characteristic for a particular structural arrangement of atoms.

In the winter of 1952-1953, Franklin produced the now-famous x-ray "picture" of DNA (Figure 1.5). Watson (1976) wrote that the picture was

> the key x-ray photograph involved in the elucidation of the DNA structure. It experimentally confirmed the then current guesses that DNA was helical. The helical form is indicated by the crossways pattern of x-ray reflections in the center of the photograph. The heavy black regions at the top and bottom tell that the 3.4 Å thick purine and pyrimidine bases are regularly stacked next to each other, perpendicular to the helical axis.

THE COMPONENTS OF DNA

As we have already seen, there are three basic components of DNA: a phosphate, a sugar, and a nitrogenous base. The phosphate group gives DNA its acidic properties and a negative charge. In vivo, unless these charges are neutralized, it would not be possible to pack the huge DNA molecule into the cell's nucleus. Neutralization is brought about by reaction of basic proteins with the acidic DNA in both **eukaryotes** (organisms with cells that have a nucleus surrounded by a membrane), such as human beings, and **prokaryotes,** the bacteria. In eukaryotes, histones are the basic proteins involved in packaging DNA (see Chapter 10), and in prokaryotes, polyamines are the basic proteins.

In fact, the smallest of the human chromosomes, with a DNA molecule of approximately 1.4 cm, in its most condensed state during mitosis, is about 1/7000 its extended length. So the packing ratio, the length of the DNA molecule divided by the length of the unit that contains it (the chromosome in this case), is on the order of 7000.

The sugar of DNA is the five-carbon 2′-deoxyribose (Figure 1.3). The "prime" is used to distinguish the sugar carbons from the nitrogenous base atoms (numbered in Figure 1.2).

The commonly occurring nitrogenous bases in DNA (Figure 1.2), usually referred to simply as "bases," are adenine, guanine, cytosine, and thymine. There are other bases sometimes called "rare" bases, but we will not be discussing those.

From these components—phosphate, sugar, base—cells construct the precursors of DNA, the **deoxyribonucleotides,** most often referred to simply as nucleotides. Any one of the nitrogenous bases, plus the deoxyribose yields a **nucleoside** (see deoxyadenosine, Figure 1.6) and if to this a phosphate is added, a nucleotide is formed (see deoxyadenosine 5′-monophosphate, Figure 1.4). (Later, we will discuss nucleotides and nucleosides formed with ribose, also shown in Figures 1.4 and 1.6.)

Deoxyribonucleic acid (DNA) is a **polymer,** or chain, of these nucleotides (Figure 1.7). The sugar is attached via its 1′ carbon to a nitrogen atom of the base, and the phosphate group is attached to the 5′ carbon of the same sugar. In fact, all nucleotides are synthesized as the 5′-monophosphate nucleotides. So, for example, the nucleotide containing adenine as the base is really 2′-deoxyriboadenine-5′-monophosphate, and the nucleotide containing cytosine is 2′-deoxyribocytidine-5′-monophosphate.

In vivo, synthesis of the nucleotides takes place via two pathways, namely, the purine and pyrimidine pathways (Table 1.1). All the nucleotides are synthesized with the phosphate group attached to the 5′ carbon of the sugar. This fact is especially important in the structure and replication of DNA, as we shall see.

For convenience and by consensus, the nucleotides are designated by the capitalized first letter of the base they contain. Thus **A, G, C,** and **T** represent the nucleotides containing adenine, guanine, cytosine, and thymine, respectively. We shall use this notation system throughout the text.