MOLECULAR BIOLOGY OF THE CELL

THIRD EDITION

Bruce Alberts • Dennis Bray
Julian Lewis • Martin Raff • Keith Roberts
James D. Watson

Garland Publishing, Inc.
New York & London

Text Editor: Miranda Robertson
Managing Editor: Ruth Adams
Illustrator: Nigel Orme
Molecular Model Drawings: Kate Hesketh-Moore
Director of Electronic Publishing: John M-Roblin
Computer Specialist: Chuck Bartelt
Disk Preparation: Carol Winter
Copy Editor: Shirley M. Cobert
Production Editor: Douglas Goertzen
Production Coordinator: Perry Bessas
Indexer: Maija Hinkle

Bruce Alberts received his Ph.D. from Harvard University and is currently President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. *Dennis Bray* received his Ph.D. from the Massachusetts Institute of Technology and is currently a Medical Research Council Fellow in the Department of Zoology, University of Cambridge. *Julian Lewis* received his D.Phil. from the University of Oxford and is currently a Senior Scientist in the Imperial Cancer Research Fund Developmental Biology Unit, University of Oxford. *Martin Raff* received his M.D. from McGill University and is currently a Professor in the MRC Laboratory for Molecular Cell Biology and the Biology Department, University College London. *Keith Roberts* received his Ph.D. from the University of Cambridge and is currently Head of the Department of Cell Biology, the John Innes Institute, Norwich. *James D. Watson* received his Ph.D. from Indiana University and is currently Director of the Cold Spring Harbor Laboratory. He is the author of *Molecular Biology of the Gene* and, with Francis Crick and Maurice Wilkins, won the Nobel Prize in Medicine and Physiology in 1962.

Library of Congress Cataloging-in-Publication Data
Molecular biology of the cell / Bruce Alberts . . . [et al.].—3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-8153-1619-4 (hard cover).—ISBN 0-8153-1620-8 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology. QH 581.2 M718 1994]
QH581.2.M64 1994
574.87—dc20
DNLM/DLC
for Library of Congress 93-45907
CIP

Published by Garland Publishing, Inc.
717 Fifth Avenue, New York, NY 10022

Printed in the United States of America

15 14 13 12 10 9 8 7 6 5 4

Front cover: The photograph shows a rat nerve cell in culture. It is labeled (*yellow*) with a fluorescent antibody that stains its cell body and dendritic processes. Nerve terminals (*green*) from other neurons (not visible), which have made synapses on the cell, are labeled with a different antibody. (Courtesy of Olaf Mundigl and Pietro de Camilli.)

Dedication page: Gavin Borden, late president of Garland Publishing, weathered in during his mid-1980s climb near Mount McKinley with MBoC author Bruce Alberts and famous mountaineer guide Mugs Stump (1940–1992).

Back cover: The authors, in alphabetical order, crossing Abbey Road in London on their way to lunch. Much of this third edition was written in a house just around the corner. (Photograph by Richard Olivier.)

Summary

The sequence of subunits in a macromolecule contains information that determines the three-dimensional contours of its surface. These contours in turn govern the recognition between one molecule and another, or between different parts of the same molecule, by means of weak, noncovalent bonds. The attractive forces are of four types: ionic bonds, van der Waals attractions, hydrogen bonds, and an interaction between nonpolar groups caused by their hydrophobic expulsion from water. Two molecules will recognize each other by a process in which they meet by random diffusion, stick together for a while, and then dissociate. The strength of this interaction is generally expressed in terms of an equilibrium constant. Since the only way to make recognition infallible is to make the energy of binding infinitely large, living cells constantly make errors; those that are intolerable are corrected by specific repair processes.

Nucleic Acids [8]

Genes Are Made of DNA [9]

It has been obvious for as long as humans have sown crops or raised animals that each seed or fertilized egg must contain a hidden plan, or design, for the development of the organism. In modern times the science of genetics grew up around the premise of invisible information-containing elements, called **genes,** that are distributed to each daughter cell when a cell divides. Therefore, before dividing, a cell has to make a copy of its genes in order to give a complete set to each daughter cell. The genes in the sperm and egg cells carry the hereditary information from one generation to the next.

The inheritance of biological characteristics must involve patterns of atoms that follow the laws of physics and chemistry: in other words, genes must be formed from molecules. At first the nature of these molecules was hard to imagine. What kind of molecule could be stored in a cell and direct the activities of a developing organism and also be capable of accurate and almost unlimited replication?

By the end of the nineteenth century biologists had recognized that the carriers of inherited information were the *chromosomes* that become visible in the nucleus as a cell begins to divide. But the evidence that the deoxyribonucleic acid (DNA) in these chromosomes is the substance of which genes are made came only much later, from studies on bacteria. In 1944 it was shown that adding purified DNA from one strain of bacteria to a second, slightly different bacterial strain conferred heritable properties characteristic of the first strain upon the second. Because it had been commonly believed that only proteins have enough conformational complexity to carry genetic information, this discovery came as a surprise, and it was not generally accepted until the early 1950s. Today the idea that DNA carries genetic information in its long chain of nucleotides is so fundamental to biological thought that it is sometimes difficult to realize the enormous intellectual gap that it filled.

DNA Molecules Consist of Two Long Chains Held Together by Complementary Base Pairs [10]

The difficulty that geneticists had in accepting DNA as the substance of genes is understandable, considering the simplicity of its chemistry. A DNA chain is a long, unbranched polymer composed of only four types of subunits. These are the deoxyribonucleotides containing the bases adenine (A), cytosine (C), guanine (G), and thymine (T). The nucleotides are linked together by covalent phosphodiester bonds that join the 5′ carbon of one deoxyribose group to the 3′ carbon of the next (see Panel 2–6, pp. 58–59). The four kinds of bases are attached to this

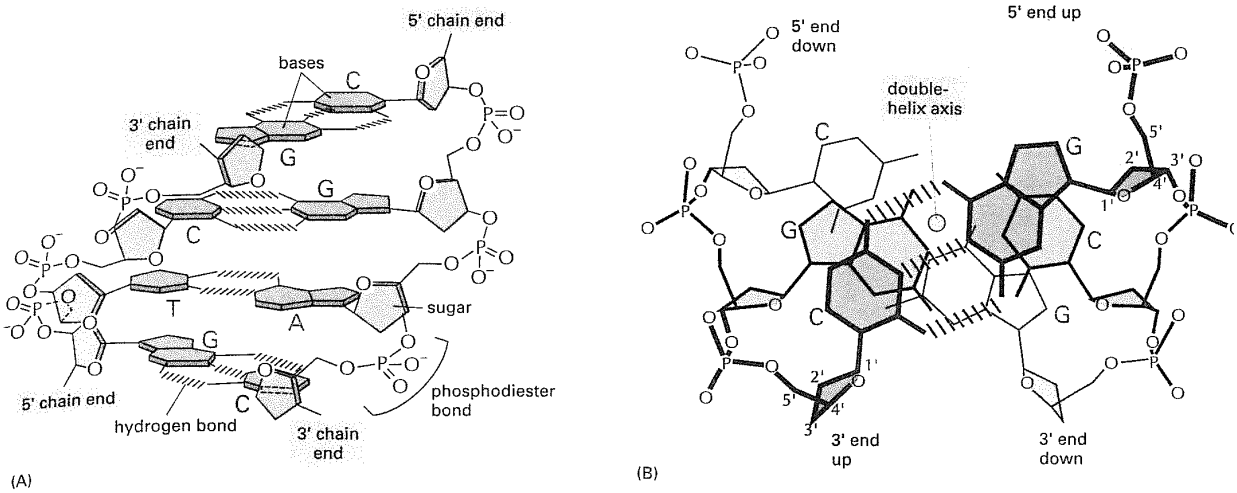

Figure 3–10 **The DNA double helix.** (A) A short section of the helix viewed from its side. Four complementary base pairs are shown. The bases are shown in *green,* while the deoxyribose sugars are *blue.* (B) The helix viewed from an end. Note that the two DNA strands run in opposite directions and that each base pair is held together by either two or three hydrogen bonds (see also Panel 3–2, pp. 100–101).

repetitive sugar-phosphate chain almost like four kinds of beads strung on a necklace.

How can a long chain of nucleotides encode the instructions for an organism or even a cell? And how can these messages be copied from one generation of cells to the next? The answers lie in the structure of the DNA molecule.

Early in the 1950s x-ray diffraction analyses of specimens of DNA pulled into fibers suggested that the DNA molecule is a helical polymer composed of two strands. The helical structure of DNA was not surprising since, as we have seen, a helix will often form if each of the neighboring subunits in a polymer is regularly oriented. But the finding that DNA is two-stranded was of crucial significance. It provided the clue that led, in 1953, to the construction of a model that fitted the observed x-ray diffraction pattern and thereby solved the puzzle of DNA structure and function.

An essential feature of the model was that all of the bases of the DNA molecule are on the *inside* of the double helix, with the sugar phosphates on the outside. This demands that the bases on one strand be extremely close to those on the other, and the fit proposed required specific *base-pairing* between a large purine base (A or G, each of which has a double ring) on one chain and a smaller pyrimidine base (T or C, each of which has a single ring) on the other chain (Figure 3–10).

Both evidence from earlier biochemical experiments and conclusions derived from model building suggested that **complementary base pairs** (also called *Watson-Crick base pairs*) form between A and T and between G and C. Biochemical analyses of DNA preparations from different species had shown that, although the nucleotide composition of DNA varies a great deal (for example, from 13% A residues to 36% A residues in the DNA of different types of bacteria), there is a general rule that quantitatively [G] = [C] and [A] = [T]. Model building revealed that the numbers of effective hydrogen bonds that could be formed between G and C or between A and T were greater than for any other combinations (see Panel 3–2, pp. 100–101). The double-helical model for DNA thus neatly explained the quantitative biochemistry.

The Structure of DNA Provides an Explanation for Heredity [11]

A gene carries biological information in a form that must be precisely copied and transmitted from each cell to all of its progeny. The implications of the discov-

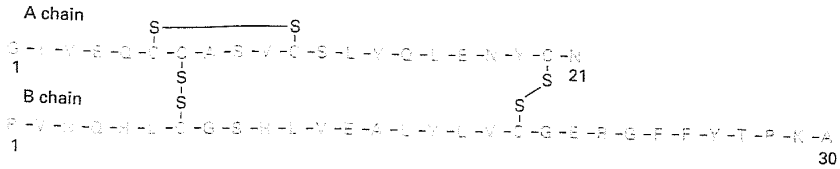

Figure 3–14 **The amino acid sequence of bovine insulin.** Insulin is a very small protein that consists of two polypeptide chains, one 21 and the other 30 amino acid residues long. Each chain has a unique, genetically determined sequence of amino acids. The one-letter symbols used to specify amino acids are those listed in Panel 2–5, pages 56–57; the S—S bonds shown in *red* are disulfide bonds between cysteine residues. The protein is made initially as a single long polypeptide chain (encoded by a single gene) that is subsequently cleaved to give the two chains.

The Nucleotide Sequence of a Gene Determines the Amino Acid Sequence of a Protein [13]

DNA is relatively inert chemically. The information it contains is expressed indirectly via other molecules: DNA directs the synthesis of specific RNA and protein molecules, which in turn determine the cell's chemical and physical properties.

At about the time that biophysicists were analyzing the three-dimensional structure of DNA by x-ray diffraction, biochemists were intensively studying the chemical structure of proteins. It was already known that proteins are chains of amino acids joined together by sequential peptide linkages; but it was only in the early 1950s, when the small protein *insulin* was sequenced (Figure 3–14), that it was discovered that each type of protein consists of a unique sequence of amino acids. Just as solving the structure of DNA was seminal in understanding the molecular basis of genetics and heredity, so sequencing insulin provided a key to understanding the structure and function of proteins. If insulin had a definite, genetically determined sequence, then presumably so did every other protein. It seemed reasonable to suppose, moreover, that the properties of a protein would depend on the precise order in which its constituent amino acids are arranged.

Both DNA and protein are composed of a linear sequence of subunits; eventually, the analysis of the proteins made by mutant genes demonstrated that the two sequences are *co-linear*—that is, the nucleotides in DNA are arranged in an order corresponding to the order of the amino acids in the protein they specify. It became evident that the DNA sequence contains a coded specification of the protein sequence. The central question in molecular biology then became how a cell translates a nucleotide sequence in DNA into an amino acid sequence in a protein.

Portions of DNA Sequence Are Copied into RNA Molecules That Guide Protein Synthesis [14]

The synthesis of proteins involves copying specific regions of DNA (the *genes*) into polynucleotides of a chemically and functionally different type known as **ribonucleic acid**, or **RNA.** RNA, like DNA, is composed of a linear sequence of nucleotides, but it has two small chemical differences: (1) the sugar-phosphate backbone of RNA contains ribose instead of a deoxyribose sugar and (2) the base thymine (T) is replaced by uracil (U), a very closely related base that likewise pairs with A (see Panel 3–2, pp. 100–101).

RNA retains all of the information of the DNA sequence from which it was copied, as well as the base-pairing properties of DNA. Molecules of RNA are synthesized by a process known as **DNA transcription,** which is similar to DNA replication in that one of the two strands of DNA acts as a template on which the base-pairing abilities of incoming nucleotides are tested. When a good match is achieved with the DNA template, a ribonucleotide is incorporated as a covalently bonded unit. In this way the growing RNA chain is elongated one nucleotide at a time.

DNA transcription differs from DNA replication in a number of ways. The RNA product, for example, does not remain as a strand annealed to DNA. Just behind the region where the ribonucleotides are being added, the original DNA helix re-forms and releases the RNA chain. Thus RNA molecules are single-

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.