

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: Majja 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.

© 1983, 1989, 1994 by Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson.

All rights reserved. No part of this book covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the publisher.

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

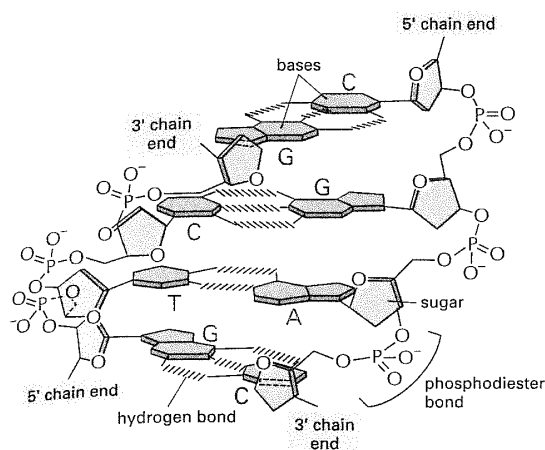
It has been obvious for as long as humans have sown crops or raised animals that each seed or fertilized egg must contain a hidden plan, or design, for the development of the organism. In modern times the science of genetics grew up around the premise of invisible information-containing elements, called **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?

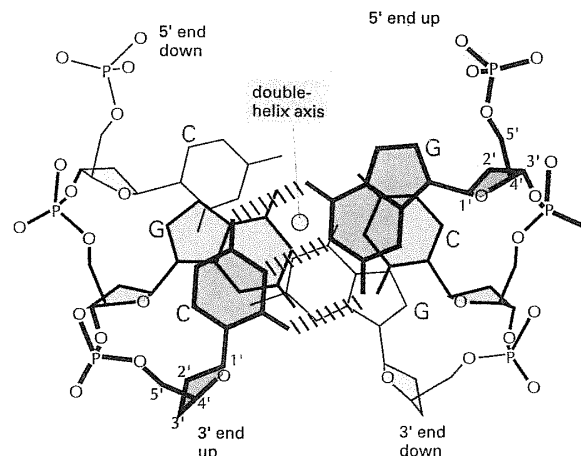
By the end of the nineteenth century biologists had recognized that the carriers of inherited information were the *chromosomes* that become visible in the nucleus as a cell begins to divide. But the evidence that the deoxyribonucleic acid (DNA) in these chromosomes is the substance of which genes are 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



(A)



(B)

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.

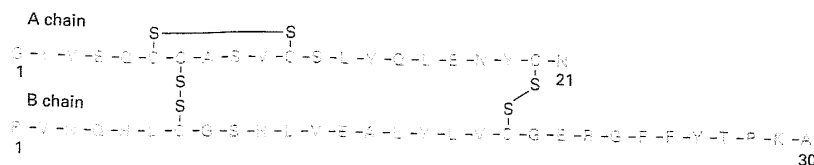
An essential feature of the model was that all of the bases of the DNA molecule are on the *inside* of the double helix, with the sugar phosphates on the outside. This demands that the bases on one strand be extremely close to those on the other, and the fit proposed required specific *base-pairing* 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 Heredity¹¹

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 3-10 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).



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-

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.

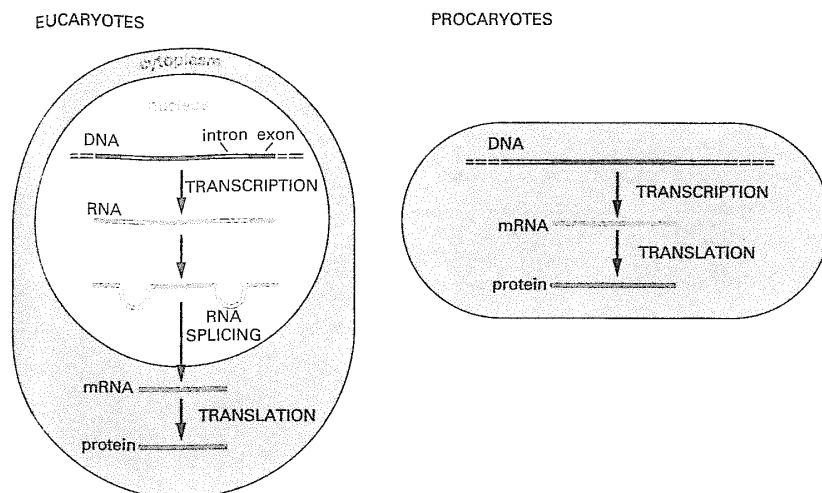


Figure 3-15 The transfer of information from DNA to protein. The transfer proceeds by means of an RNA intermediate called messenger RNA (mRNA). In procaryotic cells the process is simpler than in eucaryotic cells. In eucaryotes the coding regions of the DNA (in the *exons*, shown in color) are separated by noncoding regions (the *introns*). As indicated, these introns must be removed by an enzymatically catalyzed RNA-splicing reaction to form the mRNA.

stranded. Moreover, RNA molecules are relatively short compared to DNA molecules since they are copied from a limited region of the DNA—enough to make one or a few proteins (Figure 3-15). RNA transcripts that direct the synthesis of protein molecules are called **messenger RNA (mRNA)** molecules, while other RNA transcripts serve as *transfer RNAs (tRNAs)* or form the RNA components of ribosomes (rRNA) or smaller ribonucleoprotein particles.

The amount of RNA made from a particular region of DNA is controlled by *gene regulatory proteins* that bind to specific sites on DNA close to the coding sequences of a gene. In any cell at any given time, some genes are used to make RNA in very large quantities while other genes are not transcribed at all. For an active gene thousands of RNA transcripts can be made from the same DNA segment in each cell generation. Because each mRNA molecule can be translated into many thousands of copies of a polypeptide chain, the information contained in a small region of DNA can direct the synthesis of millions of copies of a specific protein. The protein *fibroin*, for example, is the major component of silk. In each silk gland cell a single fibroin gene makes 10^4 copies of mRNA, each of which directs the synthesis of 10^5 molecules of fibroin—producing a total of 10^9 molecules of fibroin in just 4 days.

Eucaryotic RNA Molecules Are Spliced to Remove Intron Sequences¹⁵

In bacterial cells most proteins are encoded by a single uninterrupted stretch of DNA sequence that is copied without alteration to produce an mRNA molecule. In 1977 molecular biologists were astonished by the discovery that most eucaryotic genes have their coding sequences (called *exons*) interrupted by noncoding sequences (called *introns*). To produce a protein, the entire length of the gene, including both its introns and its exons, is first transcribed into a very large RNA molecule—the *primary transcript*. Before this RNA molecule leaves the nucleus, a complex of RNA-processing enzymes removes all of the intron sequences, thereby producing a much shorter RNA molecule. After this RNA-processing step, called **RNA splicing**, has been completed, the RNA molecule moves to the cytoplasm as an mRNA molecule that directs the synthesis of a particular protein (see Figure 3-15).

This seemingly wasteful mode of information transfer in eucaryotes is presumed to have evolved because it makes protein synthesis much more versatile. The primary RNA transcripts of some genes, for example, can be spliced in various ways to produce different mRNAs, depending on the cell type or stage of

1st position (5' end)	2nd position				3rd position (3' end)
	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C U G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C U G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C U G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C U G

Figure 3-16 The genetic code. Sets of three nucleotides (*codons*) in an mRNA molecule are translated into amino acids in the course of protein synthesis according to the rules shown. The codons GUG and GAG, for example, are translated into valine and glutamic acid, respectively. Note that those codons with U or C as the second nucleotide tend to specify the more hydrophobic amino acids (compare with Panel 2-5, pp. 56-57).

development. This allows different proteins to be produced from the same gene. Moreover, because the presence of numerous introns facilitates genetic recombination events between exons, this type of gene arrangement is likely to have been profoundly important in the early evolutionary history of genes, speeding up the process whereby organisms evolve new proteins from parts of preexisting ones instead of evolving totally new amino acid sequences.

Sequences of Nucleotides in mRNA Are “Read” in Sets of Three and Translated into Amino Acids¹⁶

The rules by which the nucleotide sequence of a gene is translated into the amino acid sequence of a protein, the so-called **genetic code**, were deciphered in the early 1960s. The sequence of nucleotides in the mRNA molecule that acts as an intermediate was found to be read in serial order in groups of three. Each triplet of nucleotides, called a **codon**, specifies one amino acid. Since RNA is a linear polymer of four different nucleotides, there are $4^3 = 64$ possible codon triplets (remember that it is the *sequence* of nucleotides in the triplet that is important). However, only 20 different amino acids are commonly found in proteins, so that most amino acids are specified by several codons; that is, the genetic code is *degenerate*. The code (shown in Figure 3-16) has been highly conserved during evolution: with a few minor exceptions, it is the same in organisms as diverse as bacteria, plants, and humans.

In principle, each RNA sequence can be translated in any one of three different *reading frames* depending on where the decoding process begins (Figure 3-17). In almost every case only one of these reading frames will produce a functional protein. Since there are no punctuation signals except at the beginning and end of the RNA message, the reading frame is set at the initiation of the translation process and is maintained thereafter.

tRNA Molecules Match Amino Acids to Groups of Nucleotides¹⁷

The codons in an mRNA molecule do not directly recognize the amino acids they specify in the way that an enzyme recognizes a substrate. The **translation** of mRNA into protein depends on “adaptor” molecules that recognize both an

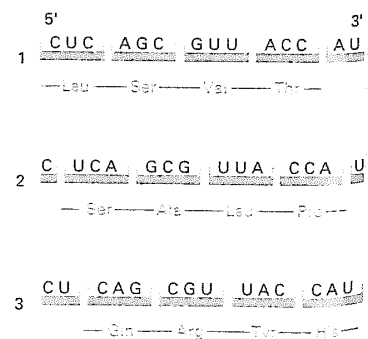


Figure 3-17 The three possible reading frames in protein synthesis. In the process of translating a nucleotide sequence (*blue*) into an amino acid sequence (*green*), the sequence of nucleotides in an mRNA molecule is read from the 5' to the 3' end in sequential sets of three nucleotides. In principle, therefore, the same RNA sequence can specify three completely different amino acid sequences, depending on the “reading frame.”

Summary

Before the synthesis of a particular protein can begin, the corresponding mRNA molecule must be produced by DNA transcription. Then a small ribosomal subunit binds to the mRNA molecule at a start codon (AUG) that is recognized by a unique initiator tRNA molecule. A large ribosomal subunit binds to complete the ribosome and initiate the elongation phase of protein synthesis. During this phase aminoacyl tRNAs, each bearing a specific amino acid, sequentially bind to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. Each amino acid is added to the carboxyl-terminal end of the growing polypeptide by means of a cycle of three sequential steps: aminoacyl-tRNA binding, followed by peptide bond formation, followed by ribosome translocation. The ribosome progresses from codon to codon in the 5'-to-3' direction along the mRNA molecule until one of three stop codons is reached. A release factor then binds to the stop codon, terminating translation and releasing the completed polypeptide from the ribosome.

Eucaryotic and procaryotic ribosomes are highly homologous, despite substantial differences in the number and size of their rRNA and protein components. The predominant role of rRNA in ribosome structure and function is likely to reflect the ancient origin of protein synthesis, which is thought to have evolved in an environment dominated by RNA-mediated catalysis.

DNA Repair¹⁶

The long-term survival of a species may be enhanced by genetic changes, but the survival of the individual demands genetic stability. Maintaining genetic stability requires not only an extremely accurate mechanism for replicating the DNA before a cell divides, but also mechanisms for repairing the many accidental lesions that occur continually in DNA. Most such spontaneous changes in DNA are temporary because they are immediately corrected by processes collectively called *DNA repair*. Only rarely do the cell's DNA maintenance processes fail and allow a permanent change in the DNA. Such a change is called a **mutation**, and it can destroy an organism if the change occurs in a vital position in the DNA sequence.

Before examining the mechanisms of DNA repair, we briefly discuss the maintenance of DNA sequences from one generation to the next.

DNA Sequences Are Maintained with Very High Fidelity¹⁷

The rate at which stable changes occur in DNA sequences (the *mutation rate*) can be estimated only indirectly. One way is to compare the amino acid sequence of the same protein in several species. The fraction of the amino acids that are different can then be compared with the estimated number of years since each pair of species diverged from a common ancestor, as determined from the fossil record. In this way one can calculate the number of years that elapse, on average, before an inherited change in the amino acid sequence of a protein becomes fixed in the species. Because each such change will commonly reflect the alteration of a single nucleotide in the DNA sequence of the gene encoding that protein, this value can be used to estimate the average number of years required to produce a single, stable mutation in the gene.

Such calculations always will substantially underestimate the actual mutation rate because most mutations will spoil the function of the protein and vanish from the population through natural selection. But there is one family of proteins whose sequence does not seem to matter, and so the genes that encode them can accumulate mutations without being selected against. These proteins are the **fibrinopeptides**—20-residue-long fragments that are discarded from the protein *fibrinogen* when it is activated to form *fibrin* during blood clotting. Since the function of fibrinopeptides apparently does not depend on their amino acid se-

quence, they can tolerate almost any amino acid change. Sequence analysis of the fibrinopeptides indicates that an average-sized protein 400 amino acids long would be randomly altered by an amino acid change roughly once every 200,000 years. More recently, DNA sequencing technology has made it possible to compare corresponding nucleotide sequences in regions of the genome that do not code for protein. Comparisons of such sequences in several mammalian species produce estimates of the mutation rate during evolution that are in excellent agreement with those obtained from the fibrinopeptide studies.

The Observed Mutation Rates in Proliferating Cells Are Consistent with Evolutionary Estimates¹⁸

The mutation rate can be estimated more directly by observing the rate at which spontaneous genetic changes arise in a large population of cells followed over a relatively short period of time. This can be done either by estimating the frequency with which new mutants arise in very large animal populations (in a colony of fruit flies or mice, for example) or by screening for changes in specific proteins in cells growing in culture. Although they are only approximate, the numbers obtained in both cases are consistent with an error frequency of 1 base-pair change in roughly 10^9 base pairs for each cell generation. Consequently, a single gene that encodes an average-sized protein (containing about 10^3 coding base pairs) would suffer a mutation once in about 10^6 cell generations. This number is at least roughly consistent with the evolutionary estimate described above, in which one mutation appears in an average gene in the germ line every 200,000 years.

Most Mutations in Proteins Are Deleterious and Are Eliminated by Natural Selection¹⁹

When the number of amino acid differences in a particular protein is plotted for several pairs of species against the time since the species diverged, the result is a reasonably straight line. That is, the longer the period since divergence, the larger the number of differences. For convenience, the slope of this line can be expressed in terms of the "unit evolutionary time" for that protein, which is the average time required for 1 amino acid change to appear in a sequence of 100 amino acid residues. When various proteins are compared, each shows a different but characteristic rate of evolution (Figure 6-32). Since all DNA base pairs are thought to be subject to roughly the same rate of random mutation, these different rates must reflect differences in the probability that an organism with a random mutation over the given protein will survive and propagate. Changes in amino acid sequence are evidently much more harmful for some proteins than for others. From Table 6-2 we can estimate that about 6 of every 7 random amino acid changes are harmful over the long term in hemoglobin, about 29 of every 30 amino acid changes are harmful in cytochrome c, and virtually all amino acid changes are harmful in histone H4. We assume that individuals who carried such harmful mutations have been eliminated from the population by natural selection.

Low Mutation Rates Are Necessary for Life as We Know It¹⁹

Since most mutations are deleterious, no species can afford to allow them to accumulate at a high rate in its germ cells. We discuss later why the observed mutation frequency, low though it is, nevertheless, is thought to limit the number of essential proteins that any organism can encode in its germ line to about 60,000. By an extension of the same arguments, a mutation frequency tenfold higher would limit an organism to about 6000 essential proteins. In this case evolution would probably have stopped at an organism no more complex than a fruit fly.

DNA Repair

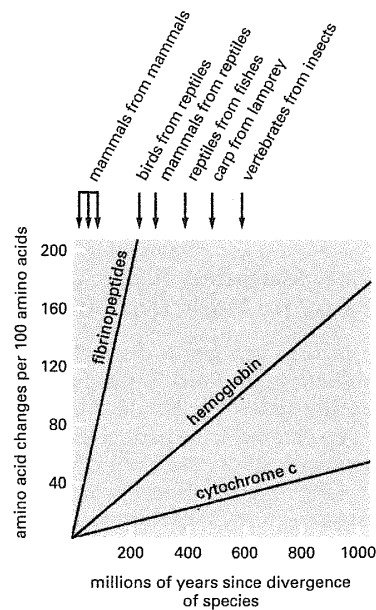


Figure 6-32 Different proteins evolve at very different rates. A comparison of the rates of amino acid change found in hemoglobin, cytochrome c, and the fibrinopeptides. Hemoglobin and cytochrome c have changed much more slowly during evolution than the fibrinopeptides. In determining rates of change per year (as in Table 6-2), it is important to realize that two species that diverged from a common ancestor 100 million years ago are separated by 200 million years of evolutionary time.

4. *DNA cloning*, whereby a single DNA molecule can be copied to generate many billions of identical molecules.
5. *DNA engineering*, by which DNA sequences are altered to make modified versions of genes, which are reinserted back into cells or organisms.

In this chapter we explain how recombinant DNA technology has generated the new experimental approaches that have revolutionized cell biology.

The Fragmentation, Separation, and Sequencing of DNA Molecules¹

Before the 1970s the goal of isolating a single gene from a large chromosome seemed unattainable. Unlike a protein, a gene does not exist as a discrete entity in cells, but rather as a small region of a much larger DNA molecule. Although the DNA molecules in a cell can be randomly broken into small pieces by mechanical force, a fragment containing a single gene in a mammalian genome would still be only one among a hundred thousand or more DNA fragments, indistinguishable in their average size. How could such a gene be purified? Since all DNA molecules consist of an approximately equal mixture of the same four nucleotides, they cannot be readily separated, as proteins can, on the basis of their different charges and binding properties. Moreover, even if a purification scheme could be devised, vast amounts of DNA would be needed to yield enough of any particular gene to be useful for further experiments.

The solution to all of these problems began to emerge with the discovery of *restriction nucleases*. These enzymes, which can be purified from bacteria, cut the DNA double helix at specific sites defined by the local nucleotide sequence, producing double-stranded DNA fragments of strictly defined sizes. Different species of bacteria make restriction nucleases with different sequence specificities, and it is relatively simple to find a restriction nuclease that will create a DNA fragment that includes a particular gene. The size of the DNA fragment can then

Table 7-1 Some Major Steps in the Development of Recombinant DNA Technology

1869	Miescher isolated DNA for the first time.
1944	Avery provided evidence that DNA, rather than protein, carries the genetic information during bacterial transformation.
1953	Watson and Crick proposed the double-helix model for DNA structure based on x-ray results of Franklin and Wilkins.
1957	Kornberg discovered DNA polymerase, the enzyme now used to produce labeled DNA probes.
1961	Marmur and Doty discovered DNA renaturation, establishing the specificity and feasibility of nucleic acid hybridization reactions.
1962	Arber provided the first evidence for the existence of DNA restriction nucleases, leading to their later purification and use in DNA sequence characterization by Nathans and H. Smith.
1966	Nirenberg, Ochoa, and Khorana elucidated the genetic code.
1967	Gellert discovered DNA ligase, the enzyme used to join DNA fragments together.
1972-1973	DNA cloning techniques were developed by the laboratories of Boyer, Cohen, Berg, and their colleagues at Stanford University and the University of California at San Francisco.
1975	Southern developed gel-transfer hybridization for the detection of specific DNA sequences.
1975-1977	Sanger and Barrell and Maxam and Gilbert developed rapid DNA-sequencing methods.
1981-1982	Palmiter and Brinster produced transgenic mice; Spradling and Rubin produced transgenic fruit flies.
1985	Mullis and co-workers invented the polymerase chain reaction (PCR).

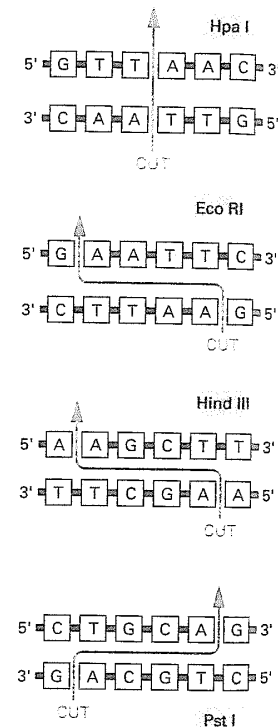


Figure 7-1 The DNA nucleotide sequences recognized by four widely used restriction nucleases. As in the examples shown, such sequences are often six base pairs long and "palindromic" (that is, the nucleotide sequence is the same if the helix is turned by 180 degrees around the center of the short region of helix that is recognized). The enzymes cut the two strands of DNA at or near the recognition sequence. For some enzymes, such as Hpa I, the cleavage leaves blunt ends; for others, such as Eco RI, Hind III, and Pst I, the cleavage is staggered and creates cohesive ends. Restriction nucleases are obtained from various species of bacteria: Hpa I is from *Hemophilus parainfluenzae*, Eco RI is from *Escherichia coli*, Hind III is from *Hemophilus influenzae*, and Pst I is from *Providencia stuartii*.

be used as a basis for partially purifying the gene from a mixture. Most important, the DNA fragment usually serves as the starting material for the production of the highly purified gene in unlimited amounts by *DNA cloning*.

We begin this section by discussing how restriction nucleases are used to produce specific DNA fragments and how these (and other) DNA molecules are separated according to their size. We then explain how, after the purification and amplification of a DNA fragment by DNA cloning, this DNA can be *sequenced* to determine the order of its nucleotides.

Restriction Nucleases Hydrolyze DNA Molecules at Specific Nucleotide Sequences²

Many bacteria make **restriction nucleases**, which protect the bacterial cell from viruses by degrading the viral DNA. Each such nuclease recognizes a specific sequence of four to eight nucleotides in DNA. These sequences, where they occur in the genome of the bacterium itself, are protected from cleavage by methylation at an A or a C residue; where the sequences occur in foreign DNA, they are generally not methylated and so are cleaved by the restriction nucleases. Large numbers of restriction nucleases have been purified from various species of bacteria; more than 100, most of which recognize different nucleotide sequences, are now available commercially.

Some restriction nucleases produce staggered cuts, which leave short single-stranded tails at the two ends of each fragment (Figure 7-1). Ends of this type are known as *cohesive ends*, as each tail can form complementary base pairs with the tail at any other end produced by the same enzyme (Figure 7-2). The cohesive ends generated by restriction enzymes allow any two DNA fragments to be easily joined together, as long as the fragments were generated with the same restriction nuclease (or with another nuclease that produces the same cohesive ends). DNA molecules produced by splicing together two or more DNA fragments in this way are called *recombinant DNA molecules*; they have made possible many new types of cell biological studies.

Restriction Maps Show the Distribution of Short Marker Nucleotide Sequences Along a Chromosome³

A particular restriction nuclease will cut any double-helical DNA molecule extracted from a cell into a series of specific DNA fragments (known as *restriction fragments*). By comparing the sizes of the DNA fragments produced from a particular genetic region after treatment with a combination of different restriction nucleases, a **restriction map** of that region can be constructed showing the location of each cutting (restriction) site in relation to neighboring restriction sites (Figure 7-3). The different short DNA sequences recognized by different restriction nucleases serve as convenient markers, and the restriction map reflects their arrangement in the region. This allows one to compare the same region of DNA in different individuals (by comparing their restriction maps) without having to determine the nucleotide sequences in detail. By comparing the restriction maps illustrated in Figure 7-4, for example, we know that the chromosomal regions that code for hemoglobin chains in humans and various other primates have remained largely unchanged during the 5 to 10 million years since these species first diverged. Restriction maps are also used in DNA cloning and DNA engineering, where they make it possible to locate a gene of interest on a particular restriction fragment and thus facilitate its isolation.

Gel Electrophoresis Separates DNA Molecules of Different Sizes⁴

In the early 1970s it was found that the length and purity of DNA molecules could be accurately determined by the same types of gel electrophoresis methods that

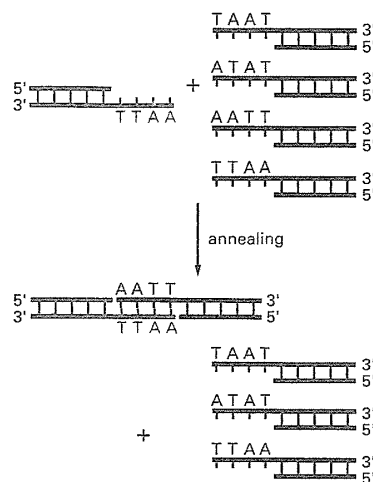
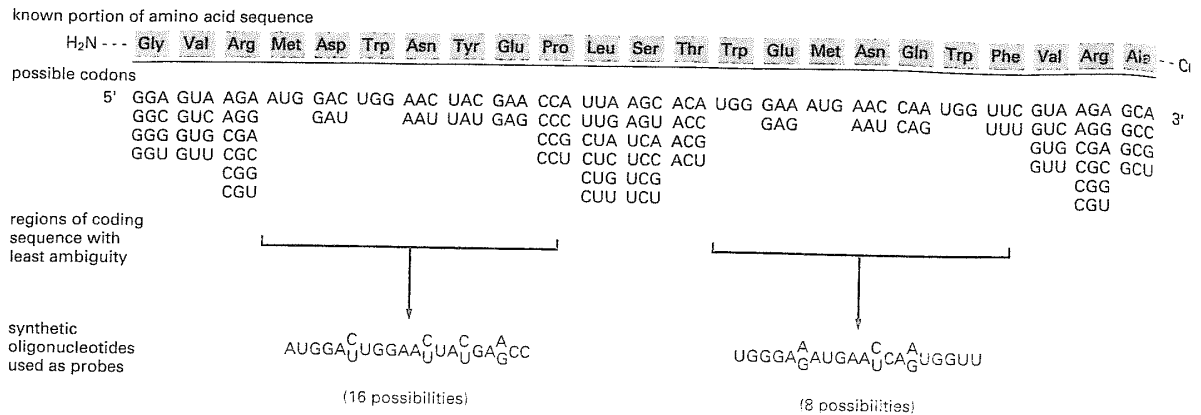


Figure 7-2 Many kinds of restriction nucleases produce DNA fragments with cohesive ends. DNA fragments with the same cohesive ends can readily join by complementary base-pairing between their cohesive ends as illustrated. The two DNA fragments that join in this example were both produced by the *Eco* RI restriction nuclease, whereas the three other fragments were produced by different restriction nucleases that generate different cohesive ends (see Figure 7-1).



The Selection of Overlapping DNA Clones Allows One to “Walk” Along the Chromosome to a Nearby Gene of Interest ²²

Many of the most interesting genes—for example, those that control development—are known only from genetic analysis of mutants in such organisms as the fruit fly *Drosophila* and the nematode *Caenorhabditis elegans*. The protein products of these genes are unknown and may be present in very small quantities in a few cells or produced only at one stage of development. A study of the genetic linkage between different mutations, however, can be used to generate *chromosome maps*, which give the relative locations of the genes (see Figure 7-16). Once one mapped gene has been cloned, the clones in a genomic DNA library that correspond to neighboring genes can be identified using a technique called **chromosome walking**. The methods described in this chapter can then be used to deduce the exact structure and function of the gene of interest and the protein that it encodes.

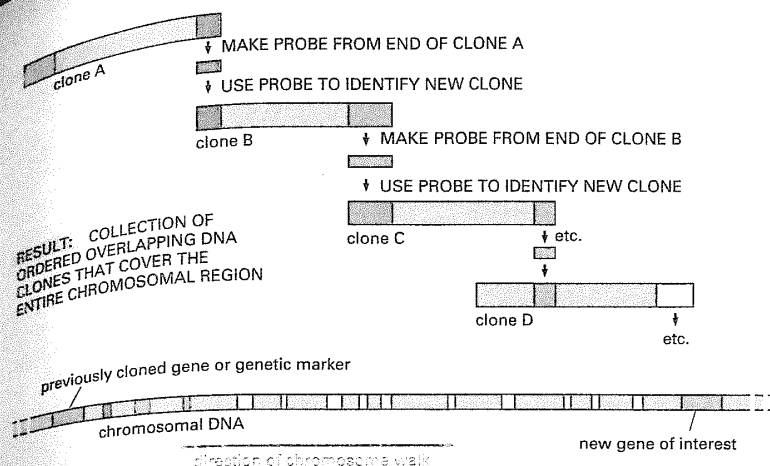
In chromosome walking one starts with a DNA clone corresponding to a gene or an RFLP marker that is known to be as close as possible to the gene of interest. One end of this clone is used to prepare a DNA probe, which is then used in DNA hybridization experiments to find an overlapping clone in a genomic DNA clone library. The DNA from this second DNA clone is purified, and its far end is used to prepare a second DNA probe, which is used to find a clone that is overlapping, and so on. In this way one can walk along a chromosome one clone at a time, in steps of 30,000 base pairs or more in either direction (Figure 7-28).

How does one know when the gene of interest (identified originally by a deleterious mutation) has been reached, given that the walk is generally too long for complete DNA sequencing to be practicable? For experimental organisms such as fruit flies, nematodes, *Arabidopsis*, yeast, and mice, the ultimate proof of the correct gene is to transfer the normal form of the gene (as a cloned DNA molecule) into a chromosome of the mutant organism, producing a transgenic organism (see Figures 7-45 and 7-49). If the original mutation was a recessive one, the correct DNA should reverse the original mutant phenotype. Other, less stringent criteria, however, are often used and are necessary in the case of human genes, as described later (see Figure 7-30).

Ordered Genomic Clone Libraries Are Being Produced for Selected Organisms ²³

The whole task of identifying mutant genes should become vastly easier as knowledge of the sequence of the normal genome becomes more complete and systematic. By using methods related to those described for chromosome walking, it has been possible to order (map) a complete, or nearly complete, set of large

Figure 7-27 Selecting regions of a known amino acid sequence to make synthetic oligonucleotide probes. Although only one nucleotide sequence will actually code for the protein, the degeneracy of the genetic code means that several different nucleotide sequences will give the same amino acid sequence, and it is impossible to tell in advance which is the correct one. Because it is desirable to have as large a fraction of the correct nucleotide sequence as possible in the mixture of oligonucleotides to be used as a probe, those regions with the fewest possibilities are chosen, as illustrated. In this example the mixture of 8 closely related oligonucleotides shown might be synthesized and used to probe a clone library, and the indicated mixture of 16 oligonucleotides would be used to reprobe all positive clones to find those that actually code for the desired protein. After the oligonucleotide mixture is synthesized by chemical means, the end of each oligonucleotide is radioactively labeled (see Figure 7-6B); alternatively, the probe can be marked with a chemical label by incorporating a modified nucleotide during its synthesis (see Figure 7-18).

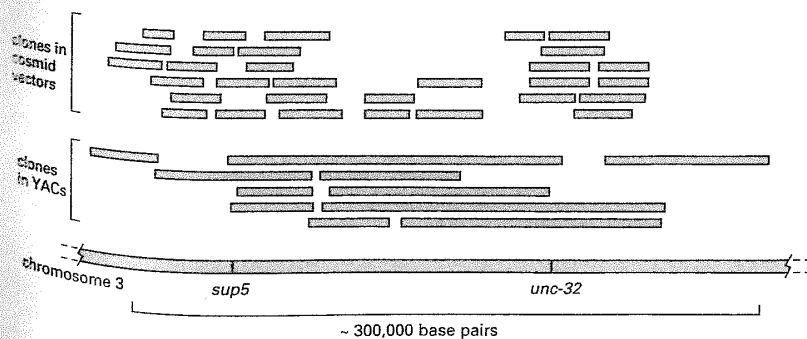


genomic clones along the chromosomes of the *E. coli* bacterium, the yeast *Saccharomyces cerevisiae*, the fruit fly *Drosophila*, the plant *Arabidopsis*, and the nematode *C. elegans*. Such large clones, each about 30,000 base pairs in length, are usually prepared in bacteriophage lambda vectors called *cosmids*, which are specially designed to accept only large DNA inserts. It takes a few thousand such clones to cover the entire genome of an organism such as *C. elegans* or *Drosophila*. To map the entire human genome in this way would require ordering more than 100,000 clones in cosmids, which is very time consuming but technically feasible. DNA fragments that are more than 10 times larger than these clones (300,000 to 1.5 million base pairs) can be cloned in yeast cells as YACs (*yeast artificial chromosomes*) (Figure 7-29); in principle, the human genome could be represented as about 10,000 clones of this type (see Figure 8-5).

In the near future, ordered sets of genomic clones will no doubt be available from centralized DNA libraries for use by all research workers. Eventually, a complete library will be available for each commonly studied organism, with each DNA fragment catalogued according to its chromosome of origin and numbered sequentially with respect to the positions of all other DNA fragments derived from the same chromosome. One will then begin a "chromosome walk" simply by obtaining from the library all the clones covering the region of the genome that contains the mutant gene of interest.

Positional DNA Cloning Reveals Human Genes with Unanticipated Functions²⁴

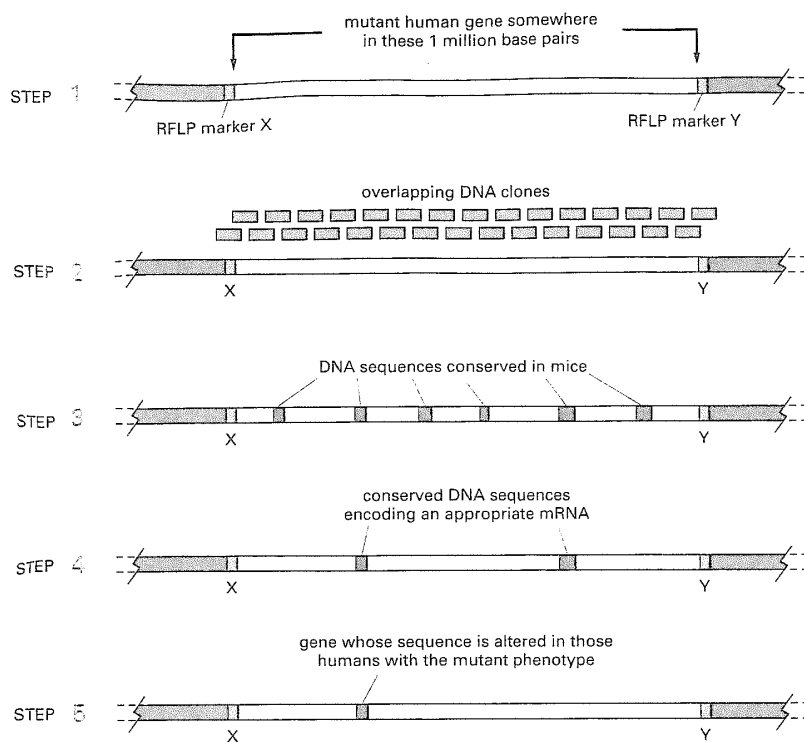
Thousands of human diseases are caused by alterations in single genes. Our understanding of these genetic diseases is being revolutionized by recombinant



DNA Cloning

Figure 7-28 The use of overlapping DNA clones to find a new gene by "chromosome walking." To speed up the walk, genomic libraries containing very large cloned DNA molecules are optimal. To probe for the next clone in the walk by DNA hybridization, a short DNA fragment (labeled with a chemical or a radioisotope) from one end of the previously identified clone is purified: If a "right-handed" end is used, for example, the walk will go in the "rightward" direction, as shown in this example. Use of a small end fragment as a probe also reduces the probability that the probe will contain a repeated DNA sequence that would hybridize with many clones from different parts of the genome and thereby interrupt the walk.

Figure 7-29 Overlapping genomic DNA clones. The collection of clones shown covers a small region of a chromosome of the nematode worm *Caenorhabditis elegans* and represents 0.3% of the total genome. (Adapted from J. Sulston et al., *Nature* 356:37-41, 1992. © 1992 Macmillan Magazines Ltd.)



DNA methods, which allow the altered DNA to be cloned and sequenced, revealing the precise defect in each patient. In this way, for example, Duchenne's muscular dystrophy was shown to be due to an abnormal cytoskeletal protein in muscle cells, and cystic fibrosis to be an abnormal chloride channel in epithelial cells. This knowledge not only improves the accuracy of diagnosis but also makes it possible, in principle at least, to design treatments.

Although the techniques used to find human disease genes have differed depending on the disease, a standard approach has recently been developed that makes it possible to isolate any human gene that is responsible by itself for a specific trait or disease. It is called **positional cloning** because it starts with genetic linkage mapping to locate the gene in the genome (Figure 7-30). While the approach is straightforward, it presently requires 10 to 100 person-years to isolate a gene in this way. It will become much easier once DNA sequencing is highly automated and the full DNA sequence of the human genome is known: genetic linkage mapping will reveal immediately which genes are prime suspects, and the sequences of these genes can then be analyzed directly in individual patients. The functions of thousands of human genes are likely to be identified in this way.

Selected DNA Segments Can Be Cloned in a Test Tube by a Polymerase Chain Reaction²⁵

The availability of purified DNA polymerases and chemically synthesized DNA oligonucleotides has made it possible to clone specific DNA sequences rapidly without the need for a living cell. The technique, called the **polymerase chain reaction (PCR)**, allows the DNA from a selected region of a genome to be amplified a billionfold, provided that at least part of its nucleotide sequence is already known. First, the known part of the sequence is used to design two synthetic DNA oligonucleotides, one complementary to each strand of the DNA double helix and lying on opposite sides of the region to be amplified. These oligonucleotides serve as primers for *in vitro* DNA synthesis, which is catalyzed

Figure 7-30 Positional cloning. The procedure requires a mutant human gene whose inheritance can be traced in many family groups by virtue of the phenotype that the mutation causes. Step 1, *genetic mapping*: RFLP markers that are coinherited with the phenotype are identified and used to position the gene within about 10^6 base pairs (one megabase, or about 1% the length of a typical human chromosome). Step 2, *assembly of an ordered clone library*: genomic DNA clones are obtained that cover the entire region between two RFLP markers that bracket the gene. Step 3, *search for conserved DNA sequences*: the portions of each DNA clone that hybridize with mouse DNA are identified; only those regions of the human chromosome whose nucleotide sequence is important will have been sufficiently conserved during evolution to form such a hybrid DNA helix (one strand human and the other mouse). Step 4, *search for appropriate mRNAs*: the subset of conserved DNA sequences that encode an mRNA in tissues where the mutant phenotype is expressed are the most likely to represent the mutant gene. Step 5, *finding a difference in the DNA sequence of mutated genes*: When the deleterious mutations in a typical human gene are analyzed, about 1 in 10 turns out to be a deletion that is easily detected as a change in the size of a restriction fragment detected by Southern blotting. For this reason one generally begins by screening the DNA of many human patients with the same disease using probes identified in step 4, looking for such a change. If the mutation is not detectable as a deletion, other, more laborious methods that are capable of detecting single base changes must be used to identify the gene of interest.

by a DNA polymerase, and they determine the ends of the final DNA fragment that is obtained (Figure 7-31).

The principle of the PCR technique is illustrated in Figure 7-32. Each cycle of the reaction requires a brief heat treatment to separate the two strands of the genomic DNA double helix (step 1). The success of the technique depends on the use of a special DNA polymerase isolated from a thermophilic bacterium that is stable at much higher temperatures than normal, so that it is not denatured by the repeated heat treatments. A subsequent cooling of the DNA in the presence of a large excess of the two primer DNA oligonucleotides allows these oligonucleotides to hybridize to complementary sequences in the genomic DNA (step 2). The annealed mixture is then incubated with DNA polymerase and the four deoxyribonucleoside triphosphates so that the regions of DNA downstream from each of the two primers are selectively synthesized (step 3). When the procedure is repeated, the newly synthesized fragments serve as templates in their turn, and within a few cycles the predominant product is a single species of DNA fragment whose length corresponds to the distance between the two original primers. In practice, 20 to 30 cycles of reaction are required for effective DNA amplification. Each cycle doubles the amount of DNA synthesized in the previous cycle. A single cycle requires only about 5 minutes, and an automated procedure permits "cell-free molecular cloning" of a DNA fragment in a few hours, compared with the several days required for standard cloning procedures.

The PCR method is extremely sensitive; it can detect a single DNA molecule in a sample. Trace amounts of RNA can be analyzed in the same way by first transcribing them into DNA with reverse transcriptase. The PCR cloning technique is rapidly replacing Southern blotting for the diagnosis of genetic diseases and for the detection of low levels of viral infection. It also has great promise in forensic medicine as a means of analyzing minute traces of blood or other tissues—even as little as a single cell—and identifying the person from whom they came by his or her genetic "fingerprint" (Figure 7-33).

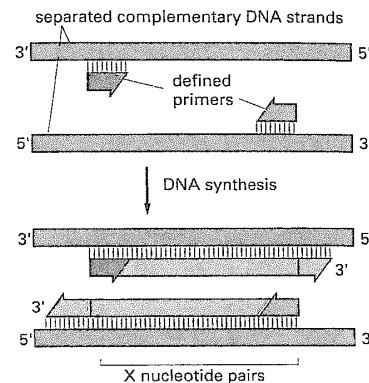
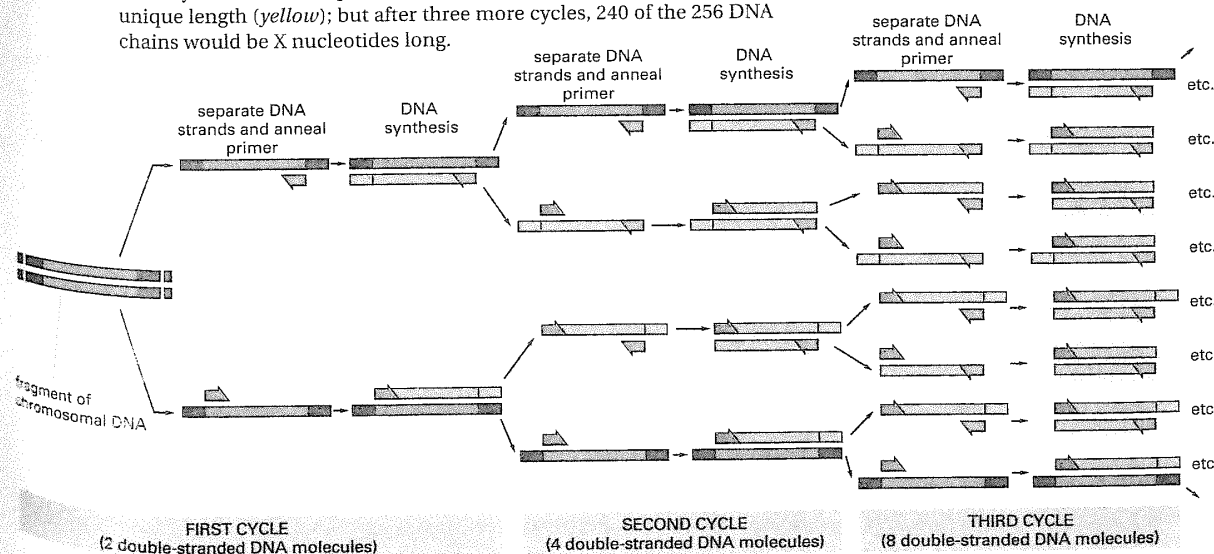


Figure 7-31 The start of the polymerase chain reaction (PCR) for amplifying specific nucleotide sequences *in vitro*. DNA isolated from cells is heated to separate its complementary strands. These strands are then annealed with an excess of two DNA oligonucleotides (each 15 to 20 nucleotides long) that have been chemically synthesized to match sequences separated by X nucleotides (where X is generally between 50 and 2000). The two oligonucleotides serve as specific primers for *in vitro* DNA synthesis catalyzed by DNA polymerase, which copies the DNA between the sequences corresponding to the two oligonucleotides.

Figure 7-32 PCR amplification. PCR produces an amount of DNA that doubles in each cycle of DNA synthesis and includes a uniquely sized DNA species. Three steps constitute each cycle, as described in the text. After many cycles of reaction, the population of DNA molecules becomes dominated by a single DNA fragment, X nucleotides long, provided that the original DNA sample contains the DNA sequence that was anticipated when the two oligonucleotides were designed. In the example illustrated, three cycles of reaction produce 16 DNA chains, 8 of which have this unique length (yellow); but after three more cycles, 240 of the 256 DNA chains would be X nucleotides long.



Nucleic acids were erroneously believed to contain only a simple repeating nucleotide sequence (such as AGCTAGCTAGCT...). We now know, however, that a DNA molecule is an enormously long, unbranched, linear polymer that can contain many millions of nucleotides arranged in an irregular but nonrandom sequence and that the genetic information of a cell is contained in the linear order of the nucleotides in its DNA. The genetic code, written in "words" of three nucleotides (*codons*) that each specify an amino acid, discussed in Chapter 6), neatly solves the problem of storing a large amount of genetic information in a small amount of space: every million "letters" (nucleotides) take up a linear distance of only 3.4×10^5 nm (0.034 cm) and occupy a total volume of about 10^6 nm³ (10^{-15} cm³).

Each DNA molecule is packaged in a separate **chromosome**, and the total genetic information stored in the chromosomes of an organism is said to constitute its **genome**. The genome of the *E. coli* bacterium contains 4.7×10^6 nucleotide pairs of DNA, present in a single double-helical DNA molecule (one chromosome). The human genome, in contrast, contains about 3×10^9 nucleotide pairs, organized as 24 chromosomes (22 different autosomes and 2 different sex chromosomes), and thus consists of 24 different DNA molecules—each containing from 50×10^6 to 250×10^6 nucleotide pairs of DNA. DNA molecules of this size are 1.7 to 8.5 cm long when uncoiled, and even the slightest mechanical force will break them once the chromosomal proteins have been removed.

In diploid organisms such as ourselves, there are two copies of each different chromosome, one inherited from the mother and one from the father (except for the sex chromosomes in males, where a Y chromosome is inherited from the father and an X from the mother). A typical human cell thus contains a total of 46 chromosomes and about 6×10^9 nucleotide pairs of DNA. Other mammals have genomes of similar size. This amount of DNA could in theory be packed into a cube 1.9 μ m on each side. By comparison, 6×10^9 letters in this book would occupy more than a million pages, thus requiring more than 10^{17} times as much space.

In this section we consider the relationship between DNA molecules, genes, and chromosomes, and we discuss how the DNA is folded into a compact and orderly structure—the chromosome—while still allowing access to its genetic information. Throughout the discussion, it is important to bear in mind that the chromosomes in a cell change their structure and activities according to the stage of the cell-division cycle: in mitosis, or *M phase*, they are very highly condensed and transcriptionally inactive; in the other, much longer part of the division cycle, called *interphase*, they are less condensed and are continuously active in directing RNA synthesis.

Each DNA Molecule That Forms a Linear Chromosome Must Contain a Centromere, Two Telomeres, and Replication Origins ²

To form a functional chromosome, a DNA molecule must be able to do more than direct the synthesis of RNA: it must be able to propagate itself reliably from one cell generation to the next. This requires three types of specialized nucleotide sequences in the DNA, each of which serves to attach specific proteins that guide the machinery that replicates and segregates chromosomes. Experiments in yeasts, whose chromosomes are relatively small and easy to manipulate by recombinant DNA methods, have identified the minimal DNA sequence elements responsible for each of these functions. Two of the three elements were identified by studying small circular DNA molecules that can be propagated as *plasmids* in cells of the yeast *Saccharomyces cerevisiae*. In order to replicate, such a DNA molecule requires a specific nucleotide sequence to act as a **DNA replication origin**; as we discuss below, one can identify the many origins in each yeast chromosome by their ability to allow a test DNA molecule that contains one of

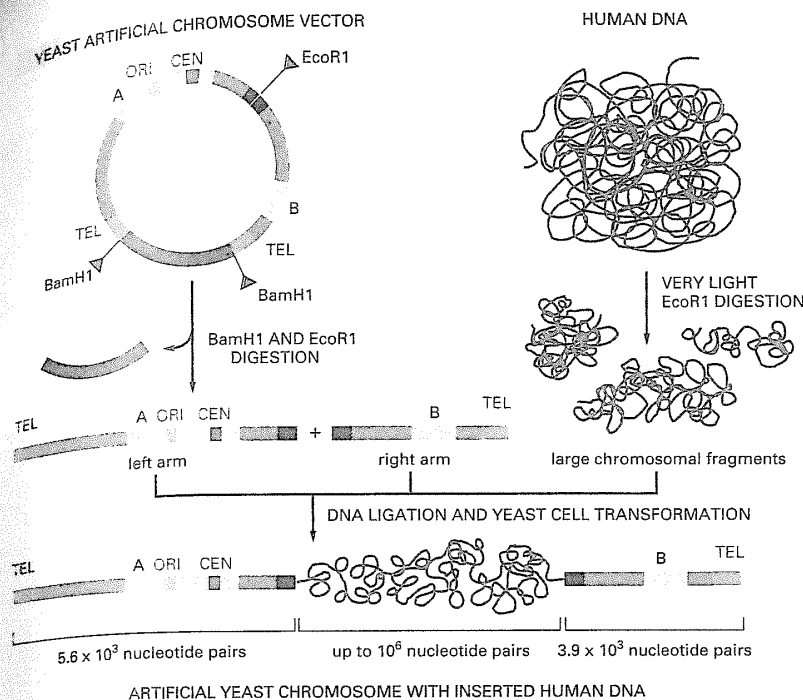


Figure 8-5 The making of a yeast artificial chromosome (YAC). A YAC vector allows the cloning of very large DNA molecules. TEL, CEN, and ORI are the telomere, centromere, and replication origin sequence elements, respectively, for the yeast *Saccharomyces cerevisiae*. BamHI and EcoRI are sites where the corresponding restriction nucleases cut the DNA double helix. The sequences denoted as A and B encode enzymes that serve as selectable markers to allow the easy isolation of yeast cells that have taken up the artificial chromosome. (Adapted from D.T. Burke, G.F. Carle, and M.V. Olson, *Science* 236:806-812, 1987. © 1987 AAAS.)

Most Chromosomal DNA Does Not Code for Proteins or RNAs³

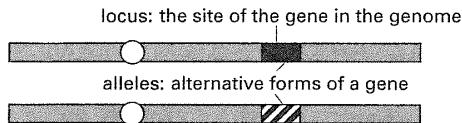
The genomes of higher organisms seem to contain a large excess of DNA. Long before it was possible to examine the nucleotide sequences of chromosomal DNA directly, it was evident that the amount of DNA in the haploid genome of an organism has no systematic relationship to the complexity of the organism. Human cells, for example, contain about 700 times more DNA than the bacterium *E. coli*, but some amphibian and plant cells contain 30 times more DNA than human cells (Figure 8-6). Moreover, the genomes of different species of amphibians can vary 100-fold in their DNA content.

Population geneticists have tried to estimate how much of the DNA in higher organisms codes for essential proteins or RNA molecules on the basis of the following indirect argument. Each gene is inevitably subject to a small risk of accidental mutation, in which nucleotides in the DNA are altered at random. The greater the number of genes, the greater the probability that a mutation will occur in at least one of them. Since most mutations will impair the function of the gene in which they occur, the mutation rate sets an upper limit to the number of essential genes that an organism can depend on for its survival: if there are too many, disaster becomes almost a certainty, as with a complex machine dependent on too many components that are liable to fail. Using this argument and the observed mutation rate, it has been estimated that no more than a small percentage of the mammalian genome can be involved in regulating or encoding essential proteins or RNA molecules. We shall see later that other evidence supports this conclusion.

The most important implication of this estimate is that although the mammalian genome contains enough DNA, in principle, to code for nearly 3 million average-sized proteins (3×10^9 nucleotides), the limited fidelity with which DNA sequences can be maintained means that no mammal (or any other organism) is likely to be constructed from more than perhaps 60,000 essential proteins.

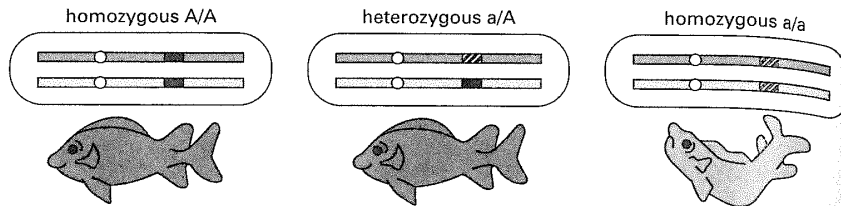
GENES AND PHENOTYPES

Gene: a functional unit of inheritance, usually corresponding to the segment of DNA coding for a single protein.
Genome: an organism's set of genes.



GENOTYPE: the specific set of alleles forming the genome of an individual

PHENOTYPE: the visible character of the individual



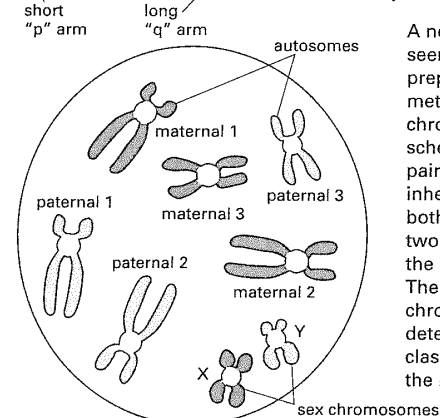
allele A is dominant (relative to a); allele a is recessive (relative to A)

In the example above, the phenotype of the heterozygote is the same as that of one of the homozygotes; in cases where it is different from both, the two alleles are said to be co-dominant.

CHROMOSOMES

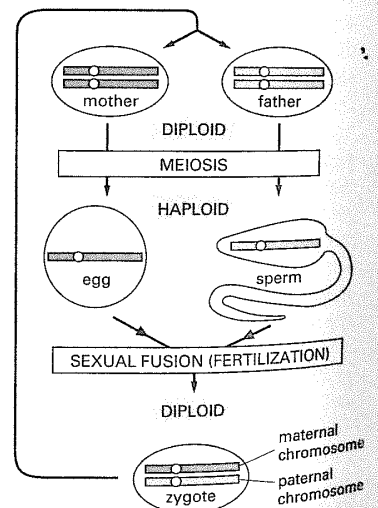
a chromosome at the beginning of the cell cycle, in G₁ phase; the single long bar represents one long double helix of DNA

a chromosome at the end of the cell cycle, in metaphase; it is duplicated and condensed, consisting of two identical sister chromatids (each containing one DNA double helix) joined at the centromere.



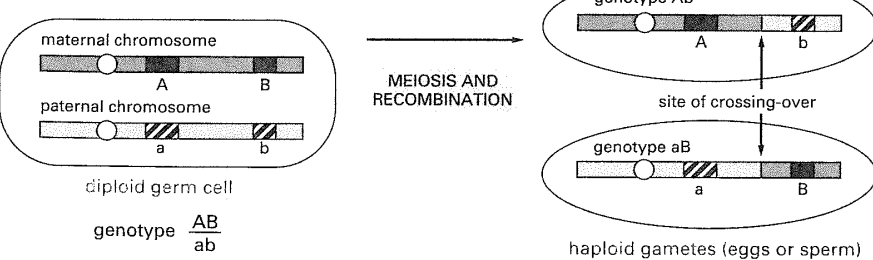
A normal diploid chromosome set, as seen in a metaphase spread, prepared by bursting open a cell at metaphase and staining the scattered chromosomes. In the example shown schematically here, there are three pairs of autosomes (chromosomes inherited symmetrically from both parents, regardless of sex) and two sex chromosomes—an X from the mother and a Y from the father. The numbers and types of sex chromosomes and their role in sex determination are variable from one class of organisms to another, as is the number of pairs of autosomes.

THE HAPLOID-DIPLOID CYCLE OF SEXUAL REPRODUCTION



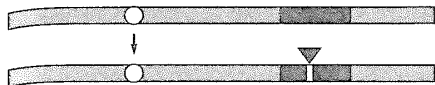
For simplicity, the cycle is shown for only one chromosome/chromosome pair.

MEIOSIS AND GENETIC RECOMBINATION

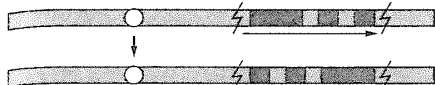


The bigger the distance between two loci, the bigger the chance that they will be recombined by crossing-over occurring at a site between them. If they get recombined in x% of gametes, they are said to be separated by a genetic map distance of x map units (or x centimorgans).

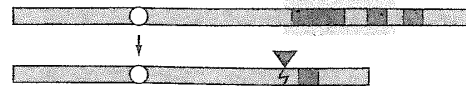
TYPES OF MUTATIONS



point mutation: maps to a single site in the genome, corresponding to a single nucleotide pair or a very small part of a single gene



inversion: inverts a segment of a chromosome



deletion: deletes a segment of a chromosome



translocation: breaks off a segment from one chromosome and attaches it to another

lethal mutation: causes the developing organism to die prematurely.

conditional mutation: produces its phenotypic effect only under certain conditions, called the *restrictive* conditions; under other conditions—the *permissive* conditions—the effect is not seen. For a *temperature-sensitive* mutation, the restrictive condition typically is high temperature, while the permissive condition is low temperature.

loss-of-function mutation: reduces or abolishes the activity of the gene. These are the commonest class of mutations. Loss-of-function mutations are usually *recessive*—the organism can usually function normally so long as it retains at least one normal copy of the affected gene.

null mutation: completely abolishes the activity of the gene.

gain-of-function mutation: increases the activity of the gene or makes it active in inappropriate circumstances; these mutations are usually *dominant*.

dominant negative mutation: dominant-acting mutation that blocks gene activity, causing a loss-of-function phenotype even in the presence of a normal copy of the gene. The phenomenon occurs when the mutant gene product interferes with the function of the normal gene product.

suppressor mutation: suppresses the phenotypic effect of another mutation, so that the double mutant appears normal.

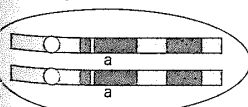
An *intragenic* suppressor mutation lies within the gene affected by the first mutation; an *extragenic* suppressor mutation lies in a second gene—often one whose product interacts directly with the product of the first.

TWO GENES OR ONE?

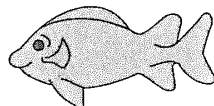
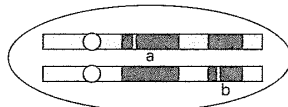
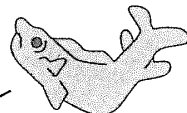
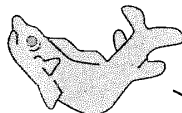
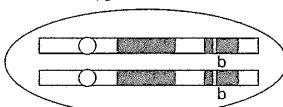
Given two mutations that produce the same phenotype, how can we tell whether they are mutations in the same gene? If the mutations are recessive (as they most often are), the answer can be found by a complementation test.

COMPLEMENTATION: MUTATIONS IN TWO DIFFERENT GENES

homozygous mutant mother



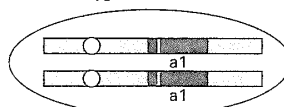
homozygous mutant father



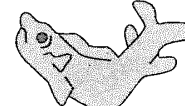
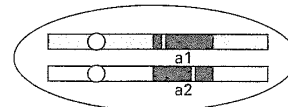
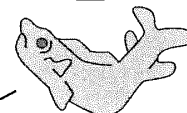
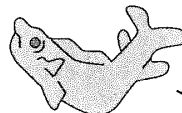
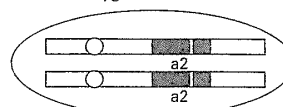
hybrid offspring shows normal phenotype:
one normal copy of each gene is present

NONCOMPLEMENTATION: TWO INDEPENDENT MUTATIONS IN THE SAME GENE

homozygous mutant mother



homozygous mutant father



hybrid offspring shows mutant phenotype:
no normal copies of the mutated gene are present

carrier protein

Membrane transport protein that binds to a solute and transports it across the membrane by undergoing a series of conformational changes.

cartilage

Form of connective tissue composed of cells (chondrocytes) embedded in a matrix rich in type II collagen and chondroitin sulfate.

catabolism

General term for the enzyme-catalyzed reactions in a cell by which complex molecules are degraded to simpler ones with release of energy. Intermediates in these reactions are sometimes called catabolites.

catalyst

Substance that accelerates a chemical reaction without itself undergoing a change. Enzymes are protein catalysts.

***cdc* gene (cell-division-cycle gene)**

Gene that controls a specific step or set of steps in the cell cycle. Originally identified in yeasts.

Cdk protein—see cyclin-dependent protein kinase**cDNA—see complementary DNA****cell-adhesion molecule (CAM)**

Protein on the surface of an animal cell that mediates cell-cell binding.

cell body

Main part of a nerve cell that contains the nucleus. The other parts are axons and dendrites.

cell cortex

Specialized layer of cytoplasm on the inner face of the plasma membrane. In animal cells it is an actin-rich layer responsible for cell-surface movements.

cell cycle

Reproductive cycle of the cell: the orderly sequence of events by which the cell duplicates its contents and divides into two.

cell division

Separation of a cell into two daughter cells. In eucaryotic cells it entails division of the nucleus (mitosis) closely followed by division of the cytoplasm (cytokinesis).

cell fusion

Process in which the plasma membranes of two cells break down at the point of contact between them, allowing the two cytoplasm to mingle.

cell junction

Specialized region of connection between two cells or between a cell and the extracellular matrix.

cell line

Population of cells of plant or animal origin capable of dividing indefinitely in culture.

cell locomotion (cell migration)

Active movement of a cell from one location to another, particularly the migration of a cell over a surface.

cell-mediated immunity

Immune responses mediated by T lymphocytes.

cell plate

Flattened membrane-bounded structure that forms from fusing vesicles in the cytoplasm of a dividing plant cell and is the precursor of the new cell wall.

cell wall

Mechanically strong extracellular matrix deposited by a cell outside its plasma membrane. It is prominent in most plants, bacteria, algae, and fungi. Not present in most animal cells.

cellulose

Structural polysaccharide consisting of long chains of covalently linked glucose units. It provides tensile strength in plant cell walls.

central nervous system (CNS)

Main information-processing organ of the nervous system. In vertebrates it consists of the brain and spinal cord.

centriole

Short cylindrical array of microtubules, closely similar in structure to a basal body. A pair of centrioles is usually found at the center of a centrosome in animal cells.

centromere

Constricted region of a mitotic chromosome that holds sister chromatids together; also the site on the DNA where the kinetochore forms and then captures microtubules from the mitotic spindle.

centrosome (cell center)

Centrally located organelle of animal cells that is the primary microtubule organizing center and acts as the spindle pole during mitosis. In most animal cells it contains a pair of centrioles.

chaperone (molecular chaperone)

Protein that helps other proteins avoid misfolding pathways that produce inactive or aggregated states.

checkpoint

Point in the eucaryotic cell-division cycle where progress through the cycle can be halted until conditions are suitable for the cell to proceed to the next stage.

chelate

Combine reversibly, usually with high affinity, with a metal ion such as iron, calcium, or magnesium.

chemiosmotic coupling

Mechanism in which a gradient of hydrogen ions (a pH gradient) across a membrane is used to drive an energy-requiring process, such as ATP production or the rotation of bacterial flagella.

chemotaxis

Motile response of a cell or an organism that carries it toward or away from a diffusible chemical.

Chlamydomonas

Unicellular green alga with two flagella.

chlorophyll

Light-absorbing pigment that plays a central part in photosynthesis.

chloroplast

Specialized organelle in green algae and plants that contains chlorophyll and performs photosynthesis. It is a specialized form of plastid.

cholesterol

Lipid molecule with a characteristic four-ringed steroid structure that is an important component of the plasma membranes of animal cells. (See Figure 10-8.)

chromaffin cell

Cell that stores adrenaline in secretory vesicles and secretes it in times of stress when stimulated by the nervous system.

chromatid

One copy of a chromosome formed by DNA replication that is still joined at the centromere to the other copy.

chromatin

Complex of DNA, histones, and nonhistone proteins found in the nucleus of a eucaryotic cell. The material of which chromosomes are made.

chromatography

Biochemical technique in which a mixture of substances is separated by charge, size, or some other property by allowing it to partition between a moving phase and a stationary phase.

chromosome

Structure composed of a very long DNA molecule and associated proteins that carries part (or all) of the hereditary information of an organism. Especially evident in plant and animal cells undergoing mitosis or meiosis, where each chromosome becomes condensed into a compact, readily visible thread.

cilium (plural cilia)

Hairlike extension of a cell containing a core bundle of microtubules and capable of performing repeated beating movements. Cilia are found in large numbers on the surface of many eucaryotic cells, and they are responsible for the swimming of many single-celled organisms.

cisterna (plural cisternae)

Flattened membrane-bounded compartment, as found in the endoplasmic reticulum or Golgi apparatus.

citric acid cycle (TCA, or tricarboxylic acid cycle; Krebs cycle)

Central metabolic pathway found in all aerobic organisms. Oxidizes acetyl groups derived from food molecules to CO₂ and H₂O. Occurs in mitochondria in eucaryotic cells.

clathrin

Protein that assembles into a polyhedral cage on the cytoplasmic side of a membrane so as to form a clathrin-coated pit, which buds off to form a clathrin-coated vesicle.

cleavage

(1) Physical splitting of a cell into two. (2) Specialized type of cell division seen in many early embryos whereby a large cell becomes subdivided into many smaller cells without growth.

clone

Population of cells or organisms formed by repeated (asexual) division from a common cell or organism. Also used as a verb: "to clone a gene" means to produce many copies of a gene by repeated cycles of replication.

cloning vector

Genetic element, usually a bacteriophage or plasmid, that is used to carry a fragment of DNA into a recipient cell for the purpose of gene cloning.

coated pit

Invagination of the plasma membrane associated with a bristlelike layer of protein on its cytoplasmic surface. Pinches off to form a coated vesicle in the process of endocytosis.

coated vesicle

Small membrane-bounded organelle formed by the pinching off of a coated region of membrane. Some coats are made of clathrin, whereas others are made from other proteins.

codon

Sequence of three nucleotides in a DNA or messenger RNA molecule that represents the instruction for incorporation of a specific amino acid into a growing polypeptide chain.

coenzyme

Small molecule tightly associated with an enzyme that participates in the reaction that the enzyme catalyzes, often by forming a transient covalent bond to the substrate. Examples include biotin, NAD⁺, and coenzyme A.

coenzyme A

Small molecule used in the enzymatic transfer of acyl groups in the cell. (See also **acetyl CoA** and Figure 2-20.)

cofactor

Inorganic ion or coenzyme that is required for an enzyme's activity.

coiled-coil

Especially stable rodlike protein structure formed by two α helices coiled around each other.

collagen

Fibrous protein rich in glycine and proline that is a major component of the extracellular matrix and connective tissues. Exists in many forms: type I, the most common, is found in skin, tendon, and bone; type II is found in cartilage; type IV is present in basal laminae.

combinatorial

Describes any process that is governed by a specific combination of factors (rather than by any single factor), with different combinations having different effects.

complement

System of serum proteins activated by antibody-antigen complexes or by microorganisms. Helps eliminate pathogenic microorganisms by directly causing their lysis or by promoting their phagocytosis.

complementary DNA (cDNA)

DNA molecule made as a copy of mRNA and therefore lacking the introns that are present in genomic DNA. Used to determine the amino acid sequence of a protein by DNA sequencing or to make the protein in large quantities by cloning followed by expression.

complementary nucleotide sequence

Two nucleic acid sequences are said to be complementary if they can form a perfect base-paired double helix with each other.

complex

Assembly of molecules that are held together by noncovalent bonds. Protein complexes perform most cell functions.

conformation

Spatial location of the atoms of a molecule—for example, the precise shape of a protein or other macromolecule in three dimensions.

connective tissue

Any supporting tissue that lies between other tissues and consists of cells embedded in a relatively large amount of extracellular matrix. Includes bone, cartilage, and loose connective tissue.

connexon

Water-filled pore in the plasma membrane formed by a ring of six protein subunits. Part of a gap junction: connexons from two adjoining cells join to form a continuous channel between the two cells.

free-energy change (ΔG)

Change in the free energy during a reaction: the free energy of the product molecules minus the free energy of the starting molecules. A large negative value of ΔG indicates that the reaction has a strong tendency to occur. (See Panel 14-1, pp. 668-669.)

G₀ phase

(G-"zero" phase) State of withdrawal from the eucaryotic cell-division cycle by entry into a quiescent G₁ phase; often seen in differentiated cells.

G₁ phase

Gap 1 phase of the eucaryotic cell-division cycle, between the end of cytokinesis and the start of DNA synthesis.

G₂ phase

Gap 2 phase of the eucaryotic cell-division cycle, between the end of DNA synthesis and the beginning of mitosis.

G protein

One of a large family of heterotrimeric GTP-binding proteins that are important intermediaries in cell-signaling pathways. Usually activated by the binding of a hormone or other signaling ligand to a seven-pass transmembrane receptor protein.

GAG (glycosaminoglycan)

Long, linear, highly charged polysaccharide composed of a repeating pair of sugars, one of which is always an amino sugar. Mainly found covalently linked to a protein core in extracellular matrix proteoglycans. Examples include chondroitin sulfate, hyaluronic acid, and heparin.

gamete

Specialized haploid cell, either a sperm or an egg, serving for sexual reproduction.

ganglion (plural ganglia)

Cluster of nerve cells and associated glial cells located outside the central nervous system.

ganglioside

Any glycolipid having one or more sialic acid residues in its structure. Found in the plasma membrane of eucaryotic cells and especially abundant in nerve cells.

gap junction

Communicating cell-cell junction that allows ions and small molecules to pass from the cytoplasm of one cell to the cytoplasm of the next.

gastrula

Animal embryo at an early stage of development where cells are invaginating to form the rudiment of a gut cavity. (From Greek *gaster*, belly.)

gene

Region of DNA that controls a discrete hereditary characteristic, usually corresponding to a single protein or RNA. This definition includes the entire functional unit, encompassing coding DNA sequences, noncoding regulatory DNA sequences, and introns.

gene regulatory protein

General name for any protein that binds to a specific DNA sequence to alter the expression of a gene.

general transcription factor

Any of the proteins whose assembly around the TATA box is required for the initiation of transcription of most eucaryotic genes.

genetic code

Set of rules specifying the correspondence between nucleotide triplets (codons) in DNA or RNA and amino acids in proteins.

genome

The totality of genetic information belonging to a cell or an organism; the DNA that carries this information.

genomic DNA

DNA constituting the genome of a cell or an organism. Often used in contrast with cDNA (DNA prepared by reverse transcription from messenger RNA).

genotype

Genetic constitution of an individual cell or organism.

germ cells

Precursor cells that give rise to gametes.

germ line

The lineage of germ cells (which contribute to the formation of a new generation of organisms), as distinct from somatic cells (which form the body and leave no descendants).

giga-

Prefix denoting 10⁹. (From Greek *gigas*, giant.)

glial cells

Supporting cells of the nervous system, including oligodendrocytes and astrocytes in the vertebrate central nervous system and Schwann cells in the peripheral nervous system.

globular protein

Any protein with an approximately rounded shape; contrasts with highly elongated, fibrous proteins such as collagen.

glucose

Six-carbon sugar that plays a major role in the metabolism of living cells. Stored in polymeric form as glycogen in animal cells and as starch in plant cells. (See Panel 2-3, pp. 52-53.)

glutaraldehyde

Small reactive molecule with two aldehyde groups that is often used as a cross-linking fixative.

glycerol

Small organic molecule that is the parent compound of many small molecules in the cell, including phospholipids.

glycocalyx (cell coat)

Carbohydrate-rich layer that forms the outer coat of a eucaryotic cell. Composed of the oligosaccharides linked to intrinsic plasma membrane glycoproteins and glycolipids, as well as glycoproteins and proteoglycans that have been secreted and reabsorbed onto the cell surface.

glycogen

Polysaccharide composed exclusively of glucose units used to store energy in animal cells. Large granules of glycogen are especially abundant in liver and muscle cells.

glycolipid

Membrane lipid molecule with a short carbohydrate chain attached to a hydrophobic tail.

glycolysis

Ubiquitous metabolic pathway in the cytosol in which sugars are incompletely degraded with production of ATP. (Literally, "sugar splitting.")

noncovalent bond

Chemical bond in which, in contrast to a covalent bond, no electrons are shared. Noncovalent bonds are relatively weak, but they can sum together to produce strong, highly specific interactions between molecules.

nonpolar molecule (apolar molecule)

Molecule lacking any asymmetric accumulation of positive or negative charge; such molecules are generally insoluble in water.

Northern blotting

Technique in which RNA fragments separated by electrophoresis are immobilized on a paper sheet; a specific molecule is then detected with a labeled nucleic acid probe.

nuclear envelope

Double membrane surrounding the nucleus. Consists of outer and inner membranes perforated by nuclear pores.

nuclear lamina

Fibrous layer on the inner surface of the inner nuclear membrane made up of a network of intermediate filaments formed from nuclear lamins.

nuclear magnetic resonance—see NMR**nuclear pore**

Channel through the nuclear envelope that allows selected molecules to move between the nucleus and cytoplasm.

nucleation

Critical stage in the process of the assembly of a polymer at which a small cluster of monomers aggregates in the correct arrangement to initiate rapid polymerization; more generally, the rate-limiting step in an assembly process.

nucleic acid

RNA or DNA; consists of a chain of nucleotides joined together by phosphodiester bonds.

nucleolar organizer

Region of a chromosome containing a cluster of ribosomal RNA genes that gives rise to a nucleolus.

nucleolus

Structure in the nucleus where ribosomal RNA is transcribed and ribosomal subunits are assembled.

nucleoside

Compound composed of a purine or pyrimidine base linked to either a ribose or a deoxyribose sugar. (See Panel 2–6, pp. 58–59.)

nucleosome

Structural, beadlike unit of a eucaryotic chromosome composed of a short length of DNA wrapped around a core of histone proteins; the fundamental subunit of chromatin.

nucleotide

Nucleoside with one or more phosphate groups joined in ester linkages to the sugar moiety. DNA and RNA are polymers of nucleotides.

nucleus

Prominent membrane-bounded organelle in a eucaryotic cell, containing DNA organized into chromosomes.

nurse cell

Cell that is connected by cytoplasmic bridges to an oocyte and thereby provides macromolecules to the growing oocyte.

Okazaki fragments

Short lengths of DNA produced on the lagging strand during DNA replication, discovered by R. Okazaki. They are rapidly joined by DNA ligase to form a continuous DNA strand.

oligodendrocyte

Type of glial cell in the vertebrate central nervous system that forms a myelin sheath around axons.

oligomer

Short polymer, usually consisting (in a cell) of amino acids (oligopeptides), sugars (oligosaccharides), or nucleotides (oligonucleotides). (From Greek *oligos*, few, little.)

oncogene

One of a large number of genes that can help make a cell cancerous. Typically, a mutant form of a normal gene (proto-oncogene) involved in the control of cell growth or division.

oocyte

Developing egg; usually a large and immobile cell.

oogenesis

Formation and maturation of oocytes in the ovary.

operator

Short region of DNA in a bacterial chromosome that controls the transcription of an adjacent gene.

operon

In a bacterial chromosome, a group of contiguous genes that are transcribed into a single mRNA molecule.

Organizer—see Spemann's Organizer**osmium tetroxide**

Inorganic compound (OsO_4) used as a fixative for electron microscopy.

osmosis

Net movement of water molecules across a semipermeable membrane driven by a difference in concentration of solute on either side. The membrane must be permeable to water but not to the solute molecules.

osmotic pressure

Pressure that must be exerted on the high solute concentration side of a semipermeable membrane to prevent the flow of water across the membrane due to osmosis.

oxidation (verb oxidize)

Loss of electron density from an atom, as occurs during the addition of oxygen to a molecule or when a hydrogen is removed. Opposite of reduction. (See Figure 2–14.)

oxidative phosphorylation

Process in bacteria and mitochondria in which ATP formation is driven by the transfer of electrons from food molecules to molecular oxygen. Involves the intermediate generation of a pH gradient across a membrane and chemiosmotic coupling.

palindromic sequence

Nucleotide sequence that is identical to its complementary strand when each is read in the same chemical direction
 5' AAAAGATCGAAA 3'
 —for example, GATC. 3' TTTCTACGTTAAA 5'

patch-clamp recording

Electrophysiological technique in which a tiny electrode tip is sealed onto a patch of cell membrane, thereby