MOLECULAR BIOLOGY OF
THE CELL
THIRD EDITION

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



Sarepta Exhibit 1056, page 1

GARLAND STAFF
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.

© 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
A member of the Taylor & Francis Group
19 Union Square West, New York, NY 10003

Printed in the United States of America
15 14 13 12 10 9 8

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.)

Recombinant DNA Technology

7

- The Fragmentation, Separation, and Sequencing of DNA Molecules
- Nucleic Acid Hybridization
- DNA Cloning
- DNA Engineering

Until the early 1970s DNA was the most difficult cellular molecule for the biochemist to analyze. Enormously long and chemically monotonous, the nucleotide sequence of DNA could be approached only by indirect means, such as by protein or RNA sequencing or by genetic analysis. Today the situation has changed entirely. From being the most difficult macromolecule of the cell to analyze, DNA has become the easiest. It is now possible to excise a specific region of DNA, to produce a virtually unlimited number of copies of it, and to determine the sequence of its nucleotides at a rate of hundreds of nucleotides a day. By variations of the same techniques, an isolated gene can be altered (engineered) at will and transferred back into cells in culture. With more difficulty, the redesigned gene can be inserted into the germ line of an animal or plant, so as to become a functional and heritable part of the organism's genome.

These technical breakthroughs have had a dramatic impact on all aspects of cell biology by allowing the study of cells and their macromolecules in previously unimagined ways. They have led to the discovery of whole new classes of genes and proteins and have revealed that many proteins have been much more highly conserved in evolution than had been suspected. They have provided new means to determine the functions of proteins and of individual domains within proteins, revealing a host of unexpected relationships between them. By making available large amounts of any protein, they have shown the way to efficient mass production of protein hormones and vaccines. Finally, by allowing the regulatory regions of genes to be dissected, they have provided biologists with an important tool for unraveling the complex mechanisms by which eucaryotic gene expression is regulated.

Recombinant DNA technology comprises a mixture of techniques, some new and some borrowed from other fields such as microbial genetics (Table 7–1). The most important of these techniques are the following:

1. Cleavage of DNA at specific sites by *restriction nucleases*, which greatly facilitates the isolation and manipulation of individual genes.
2. Rapid *sequencing* of all the nucleotides in a purified DNA fragment, which makes it possible to determine the boundaries of a gene and the amino acid sequence it encodes.
3. *Nucleic acid hybridization*, which makes it possible to find a specific sequence of DNA or RNA with great accuracy and sensitivity on the basis of its ability to bind a complementary nucleic acid sequence.

Sarepta Exhibit 1056, page 3

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 Molecules [1]

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).

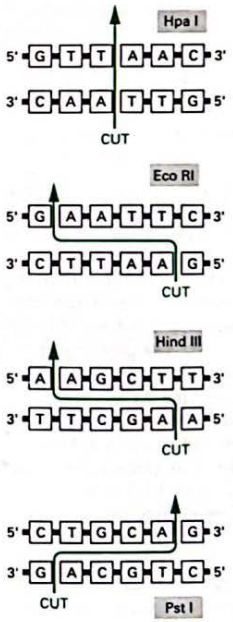


Figure 7–1 The DNA nucleotide sequences recognized by four widely used restriction nucleases. As in the examples shown, such sequences are often six base pairs long and "palindromic" (that is, the nucleotide sequence is the same if the helix is turned by 180 degrees around the center of the short region 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*.

Sarepta Exhibit 1056, page 4

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 Sequences [2]

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.

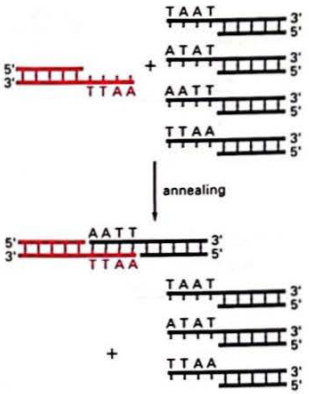


Figure 7–2 Many kinds of restriction nucleases produce DNA fragments with cohesive ends. DNA fragments with the same cohesive ends can readily join by complementary base-pairing between their cohesive ends as illustrated. The two DNA fragments that join in this example were both produced by the Eco RI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

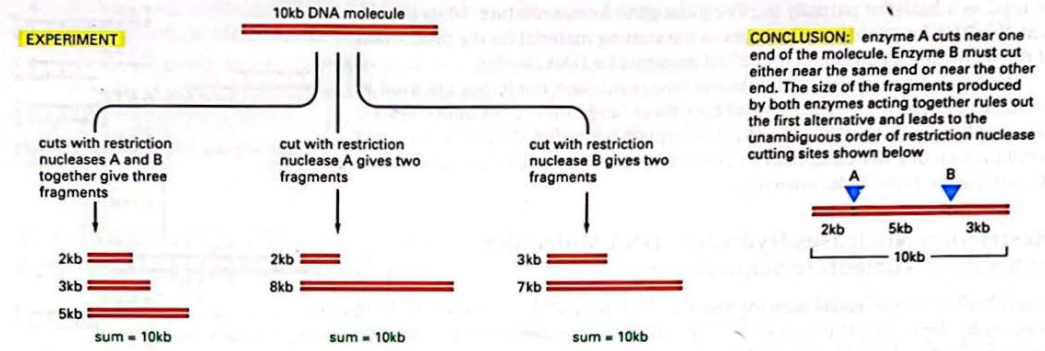
Restriction Maps Show the Distribution of Short Marker Nucleotide Sequences Along a Chromosome [3]

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 Sizes [4]

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

Sarepta Exhibit 1056, page 5



had proved so useful in the analysis of proteins. The procedure is actually simpler than for proteins: because each nucleotide in a nucleic acid molecule already carries a single negative charge, there is no need to add the negatively charged detergent SDS that is required to make protein molecules move uniformly toward the positive electrode. For DNA fragments less than 500 nucleotides long, specially designed polyacrylamide gels allow molecules that differ in length by as little as a single nucleotide to be separated from each other (Figure 7–5A). The pores in polyacrylamide gels, however, are too small to permit very large DNA molecules to pass; to separate these by size, the much more porous gels formed by dilute solutions of agarose (a polysaccharide isolated from seaweed) are used (Figure 7–5B). These DNA separation methods are widely used for both analytical and preparative purposes.

Figure 7–3 Restriction mapping. A simple example illustrating how the positions of cutting sites for different restriction nucleases (known as *restriction sites*) are determined relative to one another on double-helical DNA molecules to create a *restriction map*. (kb = kilobases, an abbreviation designating either 1000 nucleotides or 1000 nucleotide pairs.)

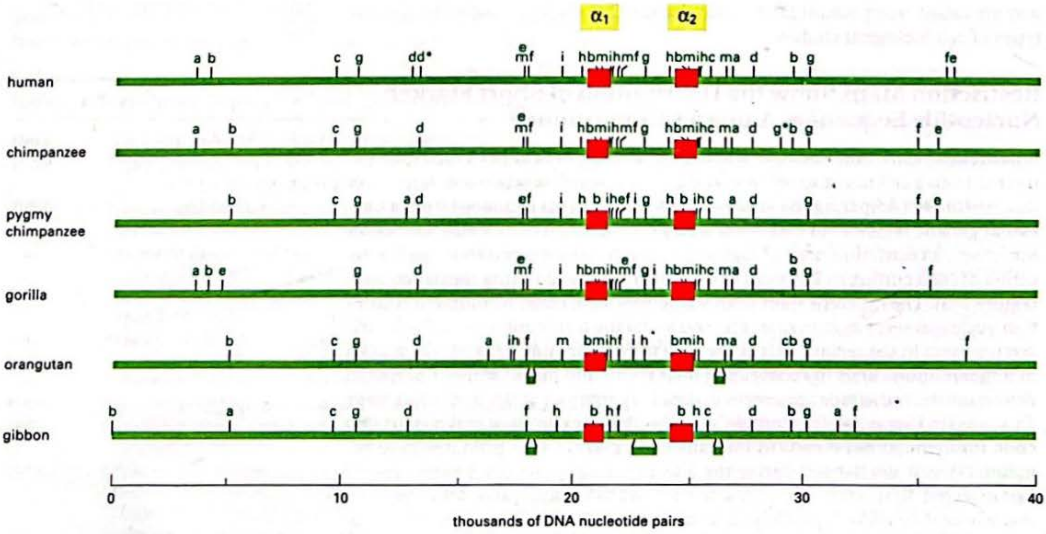


Figure 7–4 Restriction maps of human and various primate DNAs in a cluster of genes coding for hemoglobin. The two *red squares* in each map indicate the positions of the DNA corresponding to the two α-globin genes. Each letter stands for a site cut by a different restriction nuclease. As in Figure 7–3, the location of each cut was determined by comparing the sizes of the DNA fragments generated by treating the DNAs with the various restriction nucleases, individually and in combinations. Note that the chimpanzee, which is most closely related to humans, has the most similar restriction map, whereas the gibbon is more distantly related and has the most diverged map—including three DNA insertions. (Courtesy of Elizabeth Zimmer and Allan Wilson.)

Sarepta Exhibit 1056, page 6

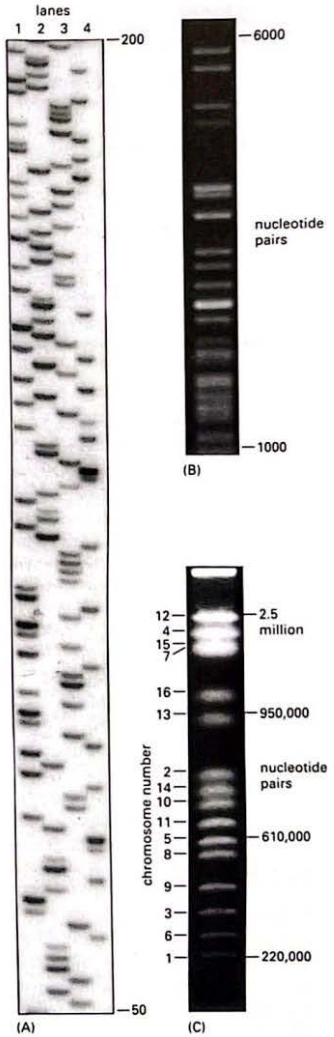


Figure 7–5 Gel electrophoresis techniques for separating DNA molecules by size. In the three examples shown, electrophoresis is from top to bottom, so that the largest—and so slowest-moving—DNA molecules are near the top of the gel. In (A) a polyacrylamide gel with small pores is used to fractionate single-stranded DNA. In the size range 10 to 500 nucleotides, DNA molecules that differ in size by only a single nucleotide can be separated from each other. In the example the four lanes represent sets of DNA molecules synthesized in the course of a DNA-sequencing procedure. The DNA to be sequenced has been artificially replicated from a fixed start site up to a variable stopping point, producing a set of partial replicas of differing lengths. (Figure 7–8 explains how such sets of partial replicas are synthesized.) Lane 1 shows all the partial replicas that terminate in a G; lane 2, all those that terminate in an A; lane 3, all those that terminate in a T; and lane 4, all those that terminate in a C. Since the DNA molecules used in these reactions are radiolabeled, their positions can be determined by autoradiography, as shown. In (B) an agarose gel with medium-sized pores is used to separate double-stranded DNA molecules. This method is most useful in the size range 300 to 10,000 nucleotide pairs. These DNA molecules are restriction fragments produced from the genome of a bacterial virus, and they have been detected by their fluorescence when stained with the dye ethidium bromide. In (C) the technique of pulsed-field agarose gel electrophoresis has been used to separate 16 different yeast (*Saccharomyces cerevisiae*) chromosomes that range in size from 220,000 to 2.5 million nucleotide pairs. The DNA was stained as in (B). DNA molecules as large as 10^7 nucleotide pairs can be separated in this way. (A, courtesy of Leander Lauffer and Peter Walter; B, courtesy of Ken Kreuzer; C, from D. Vollrath and R.W. Davis, *Nucleic Acids Res.* 15:7876, 1987, by permission of Oxford University Press.)

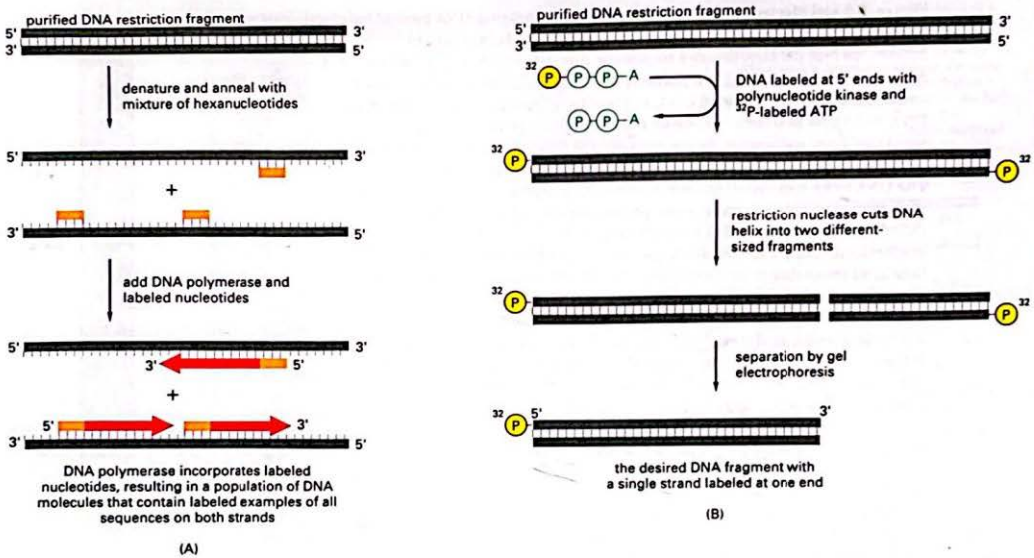
A variation of agarose gel electrophoresis, called *pulsed-field gel electrophoresis*, makes it possible to separate even extremely long DNA molecules. Ordinary gel electrophoresis fails to separate such molecules because the steady electric field stretches them out so that they travel end-first through the gel in snakelike configurations at a rate that is independent of their length. In pulsed-field gel electrophoresis, by contrast, the direction of the electric field is changed periodically, which forces the molecules to reorient before continuing to move snakelike through the gel. This reorientation takes much more time for larger molecules, so that progressively longer molecules move more and more slowly. As a consequence, even entire bacterial or yeast chromosomes separate into discrete bands in pulsed-field gels and so can be sorted and identified on the basis of their size (Figure 7–5C). Although a typical mammalian chromosome of 10^8 base pairs is too large to be sorted even in this way, large subregions of these chromosomes are readily separated and identified if the chromosomal DNA is first cut with a restriction nuclease selected to recognize sequences that occur only extremely rarely (once every 10^6 to 10^7 nucleotide pairs, for example).

The DNA bands on agarose or polyacrylamide gels are invisible unless the DNA is labeled or stained in some way. One sensitive method of staining DNA is to expose it to the dye *ethidium bromide*, which fluoresces under ultraviolet light when it is bound to DNA (see Figure 7–5B and C). An even more sensitive detection method involves incorporating a radioisotope into the DNA molecules before electrophoresis; ^{32}P is often used as it can be incorporated into DNA phosphates and emits an energetic β particle that is easily detected by autoradiography (Figure 7–5A).

Purified DNA Molecules Can Be Specifically Labeled with Radioisotopes or Chemical Markers *in Vitro* [5]

Two procedures are widely used to add distinct labels to isolated DNA molecules. In the first the DNA is copied by an *E. coli* enzyme, *DNA polymerase I*, in the presence of nucleotides that are either radioactive (usually labeled with ^{32}P) or

Sarepta Exhibit 1056, page 7



chemically tagged (Figure 7–6A). In this way "DNA probes" containing many labeled nucleotides can be produced for nucleic acid hybridization reactions (see below). The second procedure uses the bacteriophage enzyme *polynucleotide kinase* to transfer a single ^{32}P-labeled phosphate from ATP to the 5′ end of each DNA chain (Figure 7–6B). Because only one ^{32}P atom is incorporated by the kinase into each DNA strand, the DNA molecules labeled in this way are often not radioactive enough to be used as DNA probes; because they are labeled at only one end, however, they are invaluable for DNA sequencing and DNA footprinting, as we see next.

Figure 7–6 Two procedures are used routinely for making DNA molecules radioactive. (A) A purified DNA polymerase enzyme labels all the nucleotides in a DNA molecule and can thereby produce highly radioactive DNA probes. (B) Polynucleotide kinase labels only the 5′ ends of DNA strands; therefore, when labeling is followed by restriction nuclease cleavage, as shown, DNA molecules containing a single 5′-end-labeled strand can be readily obtained. The method in (A) is also used to produce nonradioactive DNA molecules that carry a specific chemical marker that can be detected with an appropriate antibody (see Figure 7–18).

Isolated DNA Fragments Can Be Rapidly Sequenced [6]

In the late 1970s methods were developed that allow the nucleotide sequence of any purified DNA fragment to be determined simply and quickly. They have made it possible to determine the complete DNA sequences of thousands of genes, including those coding for such well-known proteins as insulin, hemoglobin, interferon, and cytochrome c. The volume of DNA sequence information is already so large (many tens of millions of nucleotides) that computers must be used to store and analyze it. Several continuous stretches of DNA sequence have been determined that each contain more than 10^5 nucleotide pairs; these include the entire genomes of the Epstein-Barr virus (which infects humans and causes infectious mononucleosis) and of a plant chloroplast, as well as an entire chromosome of yeast. Two powerful DNA sequencing methods were originally developed: the principle underlying the *chemical method* is illustrated in Figure 7–7, and the *enzymatic method* is explained in Figure 7–8. The latter method, which is based on *in vitro* DNA synthesis carried out in the presence of chain-terminating nucleoside triphosphates, has now become the standard procedure for sequencing DNA.

DNA sequencing methods are so rapid and reliable that the easiest and most accurate way to determine the amino acid sequence of a protein is to determine the nucleotide sequence of its gene: a cDNA clone is made from the appropriate mRNA (see below), its nucleotide sequence is determined, and the genetic

Sarepta Exhibit 1056, page 8

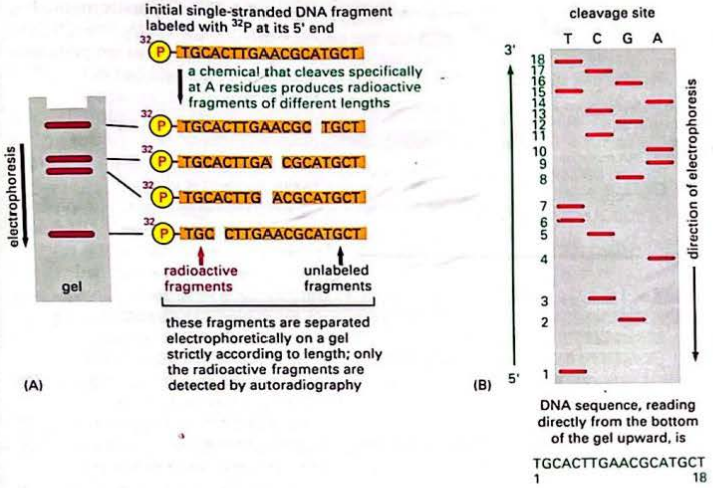


Figure 7–7 The chemical method for sequencing DNA. (A) The procedure starts with a set of identical end-labeled double-stranded DNA molecules produced by the method outlined in Figure 7–6B. In the first step the strands of the double helix are dissociated and exposed to mild treatment with a chemical that destroys one of the four bases (in this case, A residues) in the DNA. Because the treatment is mild, usually only one of the A residues in each molecule is destroyed at random. This generates a family of DNA fragments of different lengths, reflecting the different sites at which A residues occur in the original DNA. These fragments are separated on a gel and detected by autoradiography: only the fragments possessing a 5′-terminal ^{32}P-phosphate group show up on the gel, and their sizes reveal the distances from the labeled end at which A residues occur. (B) To determine the full sequence, similar procedures are carried out simultaneously on four separate samples of the same 5′-end-labeled DNA molecule using chemicals that cleave DNA preferentially at T for the first sample, C for the second, G for the third, and A for the fourth. The resulting fragments are separated in parallel lanes of a gel like that shown in Figure 7–5A, giving a pattern of radioactive DNA bands from which the DNA sequence is read. The nucleotide closest to the 5′ end of the sequence is determined by looking across the gel at level 1 (at the bottom of the gel) and seeing in which lane a band appears (T). The same procedure is repeated for level 2, then level 3, and so on, to obtain the sequence. The method has been idealized here; the actual chemical treatments are less specific than shown.

code is then used as a dictionary to convert this sequence back to an amino acid sequence. Although in principle there are six different reading frames in which a DNA sequence can be translated into protein (three on each strand), the correct one is generally recognizable as the only one lacking frequent *stop codons* (Figure 7–9). A limited amount of amino acid sequence can be determined from the purified protein to confirm the sequence determined from the DNA.

As techniques for DNA sequencing have improved, scientists have begun to envisage the possibility of determining the entire DNA sequence of a human genome of 3×10^9 nucleotides. Because this sequence specifies all of the possible RNA and protein molecules that are used to construct the body, knowledge of it will provide us with a "dictionary of the human being," which will greatly expedite future studies of human cells and tissues. DNA sequencing on such a scale will require the development of highly automated methods that will lower the current cost of DNA sequencing by at least fivefold.

DNA Footprinting Reveals the Sites Where Proteins Bind on a DNA Molecule [7]

A modification of the chemical method for DNA sequencing can be used to determine the nucleotide sequences recognized by DNA-binding proteins. Some of these proteins play a central part in determining which genes are active in a particular cell by binding to regulatory DNA sequences, which are usually located outside the coding regions of a gene. In analyzing how such a protein functions, it is important to identify the specific sequences to which it binds. A method used for this purpose is called **DNA footprinting.** First, a pure DNA fragment that is labeled at one end with ^{32}P is isolated (see Figure 7–6B); this molecule is then cleaved with a nuclease or a chemical that makes random single-stranded cuts in the DNA. After the DNA molecule is denatured to separate its two strands, the resultant subfragments from the labeled strand are separated on a gel and detected by autoradiography. The pattern of bands from DNA cut in the presence of a DNA-binding protein is compared with that from DNA cut in its absence. When the protein is present, it covers the nucleotides at its binding site and protects their phosphodiester bonds from cleavage. As a result, the labeled fragments that terminate in the binding site will be missing, leaving a gap in the gel pattern called a "footprint" (Figure 7–10A). The footprint of a DNA-binding protein that activates the transcription of a eucaryotic gene is shown in Figure 7–10B.

Sarepta Exhibit 1056, page 9

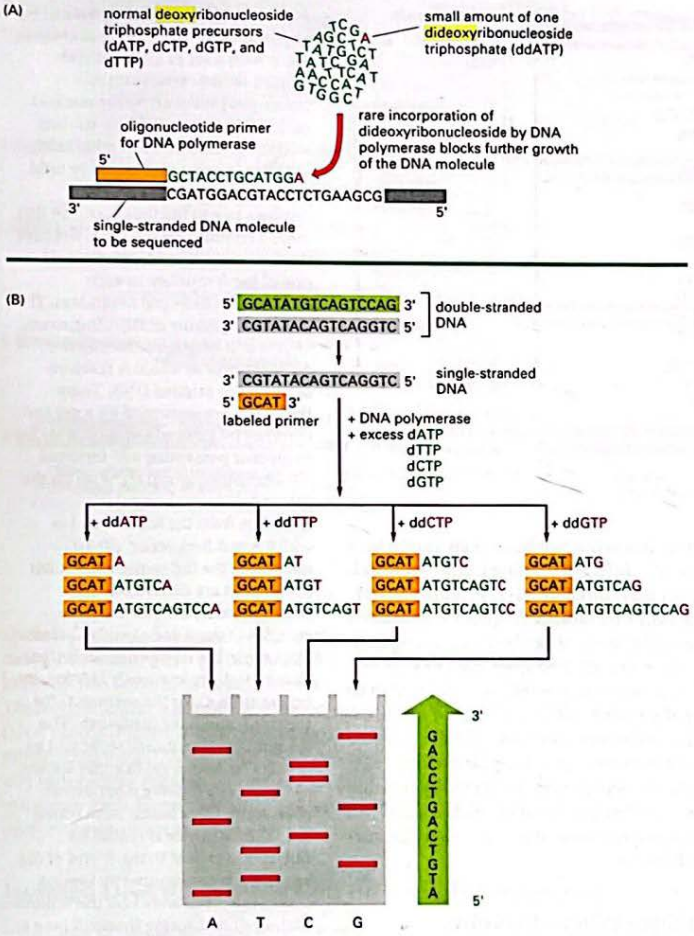
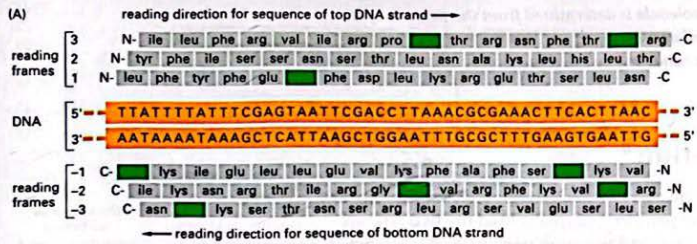


Figure 7–8 The enzymatic method for sequencing DNA. (A) The DNA to be sequenced is used as template for the *in vitro* synthesis, by DNA polymerase, of a set of partial replicas, all beginning at the same place, but terminating at different points along the DNA chain. The key to this method is the use of dideoxyribonucleoside triphosphates in which the deoxyribose 3′-OH group present in normal nucleotides is missing; when such a modified nucleotide is incorporated into a DNA chain, it blocks the addition of the next nucleotide. In the example illustrated, dideoxy ATP (ddATP) competes with an excess of deoxy ATP (dATP), so that each newly synthesized DNA strand made in a test tube by DNA polymerase will stop at a randomly selected A in the sequence. This reaction therefore generates a ladder of DNA fragments similar to that shown previously for the chemical method (Figure 7–7); these fragments are detected by a label (chemical or radioactive) that is either incorporated into the oligonucleotide primer (*orange*), or into one of the deoxyribonucleoside triphosphates (*green*) used to extend the DNA chain. (B) To determine the full sequence, four different chain-terminating nucleoside triphosphates (*red*) are used in separate DNA synthesis reactions on the same primed single-stranded DNA template. When the products of these four reactions are analyzed by electrophoresis in four parallel lanes of a polyacrylamide gel, the DNA sequence can be derived in the manner illustrated for the chemical method in Figure 7–7B.

Summary

Recombinant DNA technology has revolutionized the study of the cell. The development of this technology was neither planned nor anticipated. Instead, steady advances in the ability of researchers to manipulate DNA molecules were made on many different fronts until the combination of techniques became powerful enough to allow researchers to pick out any gene at will and, after an amplification step, to determine the exact molecular structure of both the gene and its products. Crucial elements in this technology are the ability to fragment chromosomes into DNA molecules with specific ends and the means to separate and sequence the resulting DNA fragments. The fragmentation step is carried out by proteins that are normally produced by bacteria in order to destroy invading DNA molecules. These proteins, called restriction nucleases, are purified and used to cut the DNA double helix at a specific short nucleotide sequence. The resulting DNA fragments can be separated from one another according to their size by using electrophoresis to move the mixture of fragments through the pores of a gel. Under certain conditions, even DNA molecules that differ in length by only a single nucleotide can be detected as separate bands. DNA sequencing takes advantage of this powerful separation method: after electrophoresis,

Sarepta Exhibit 1056, page 10



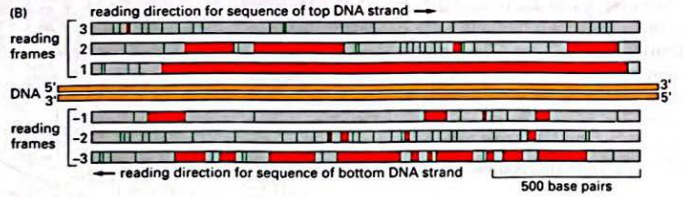


Figure 7–9 Finding the regions in a DNA sequence that encode a protein. (A) Any region of the DNA sequence can, in principle, code for six different amino acid sequences, because any one of three different reading frames can be used to interpret the nucleotide sequence on each strand. Note that a nucleotide sequence is always read in the 5′-to-3′ chain direction and encodes a polypeptide from the amino (N) to the carboxyl (C) terminus. For a random nucleotide sequence, a stop signal for protein synthesis will be encountered, on average, about once every 21 amino acids (once every 63 nucleotides). In this sample sequence of 48 base pairs, each such signal (*stop codon*) is colored *green*, and only reading frame 2 lacks a stop signal. (B) Search of a 1700 base-pair DNA sequence for a possible protein-encoding sequence. The information is displayed as in (A), with each stop signal for protein synthesis denoted by a *green line*. In addition, all of the regions between possible start and stop signals for protein synthesis (see p. 234) are displayed as *red bars*. Only reading frame 1 actually encodes a protein, which is 475 amino acid residues long.

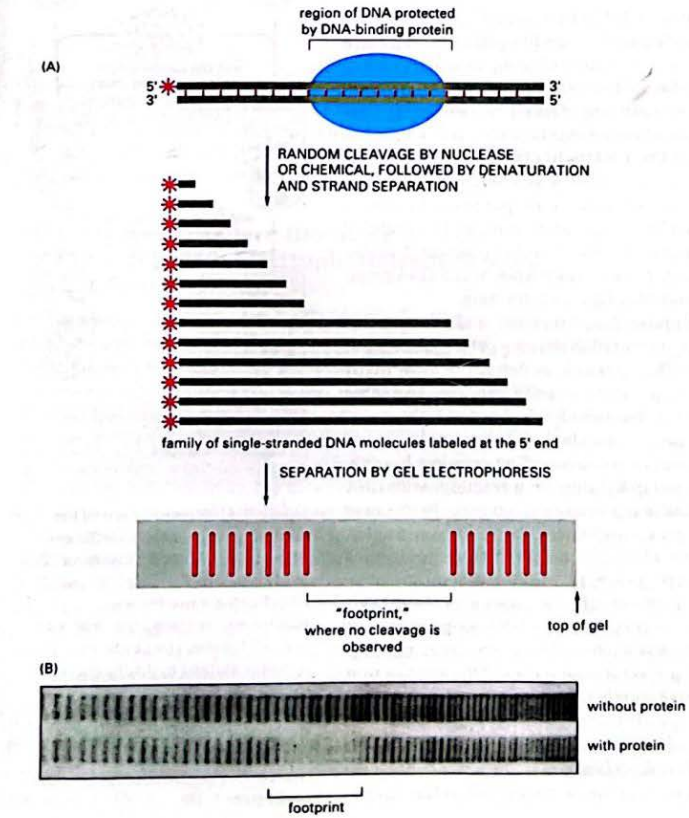


Figure 7–10 The DNA footprinting technique. (A) This technique requires a DNA molecule that has been radioactively labeled at one end (see Figure 7–6B). The protein shown binds tightly to a specific DNA sequence that is seven nucleotides long, thereby protecting these seven nucleotides from the cleaving agent. If the same reaction were carried out without the DNA-binding protein, a complete ladder of bands would be seen on the gel (not shown). (B) An actual footprint used to determine the binding site for a human protein that stimulates the transcription of specific eucaryotic genes. These results locate the binding site about 60 nucleotides upstream from the start site for RNA synthesis. The cleaving agent was a small, iron-containing organic molecule that normally cuts at every phosphodiester bond with nearly equal frequency. (B, courtesy of Michele Sawadogo and Robert Roeder.)

the sequence of nucleotides in a DNA molecule is determined from the pattern of labeled DNA bands observed when a fixed end of one of its DNA strands has been labeled and the other end terminates at a randomly selected nucleotide of a single type (A, G, C, or T).

Nucleic Acid Hybridization [8]

When an aqueous solution of DNA is heated at 100°C or exposed to a very high pH (pH ≥ 13), the complementary base pairs that normally hold the two strands of the double helix together are disrupted and the double helix rapidly dissociates into two single strands. This process, called *DNA denaturation*, was for many years thought to be irreversible. In 1961, however, it was discovered that complementary single strands of DNA will readily re-form double helices (a process called **DNA renaturation** or **hybridization**) if they are kept for a prolonged period at 65°C. Similar hybridization reactions will occur between any two single-stranded nucleic acid chains (DNA/DNA, RNA/RNA, or RNA/DNA), provided that they have a complementary nucleotide sequence. In this section we explain how these specific hybridization reactions can be used to detect and characterize specific nucleotide sequences in both RNA and DNA molecules.

Nucleic Acid Hybridization Reactions Provide a Sensitive Way of Detecting Specific Nucleotide Sequences [9]

The rate of double-helix formation during hybridization reactions is limited by the rate at which two complementary nucleic acid chains happen to collide. Since this collision rate is proportional to the concentration of complementary chains in the solution, hybridization rates can be used to determine the concentration of any desired DNA or RNA sequence in a mixture of other sequences. The assay requires a pure single-stranded DNA fragment that is complementary in sequence to the desired nucleic acid (DNA or RNA). This DNA fragment can be obtained by cloning, or, if the sequence is short, it can be synthesized by chemical means. In either case the DNA fragment must carry a unique marker (either a radioisotope or a chemical label), so that its incorporation into double-stranded molecules can be followed during the course of a hybridization reaction. A single-stranded DNA molecule used as an indicator in this way is known as a *DNA probe;* it can be anywhere from fifteen to thousands of nucleotides long.

Hybridization reactions using DNA probes are so sensitive and selective that complementary sequences present at a concentration as low as one molecule per cell can be detected (Figure 7–11). It is thus possible to determine how many copies of a particular DNA sequence are present in a cell's genome. The same technique can be used to search for related but nonidentical genes; once an interesting gene has been cloned from a mouse or a chicken, for example, part of its sequence can be used as a probe to find the corresponding gene in a human.

Alternatively, DNA probes can be used in hybridization reactions with RNA rather than DNA to find out whether a cell is *expressing* a given gene. In this case a DNA probe that contains part of the gene's sequence is hybridized with RNA purified from the cell in question to see whether the RNA includes molecules matching the probe DNA and, if so, in what quantities. In somewhat more elaborate procedures the DNA probe is treated with specific nucleases after the hybridization is complete to determine the exact regions of the DNA probe that have paired with cellular RNA molecules. One can thereby determine the start and stop sites for RNA transcription, as well as the precise boundaries of the regions that are cut out of the RNA transcripts by *RNA splicing* (the intron sequences) (Figure 7–12).

Large numbers of genes are switched on and off in elaborate patterns as an embryo develops. The hybridization of DNA probes to cellular RNAs allows one to determine whether a particular gene is off or on; moreover, when the expres-

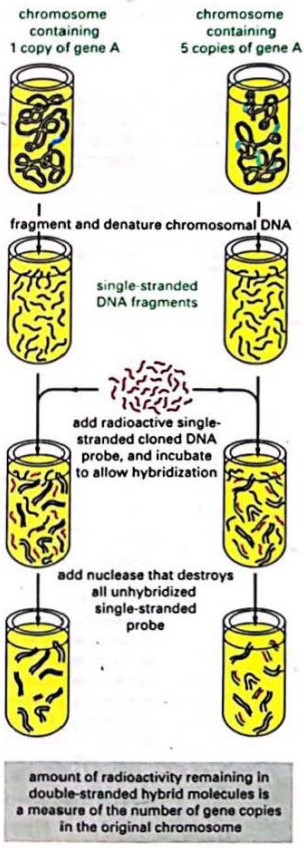


Figure 7–11 Measurement of the number of copies of a specific gene in a sample of DNA by means of DNA hybridization. The single-stranded DNA fragment used in such experiments is commonly referred to as a *DNA probe.* The probe must be specially marked to distinguish it from the vast excess of the chromosomal DNA. In this example the label is a radioactive isotope; alternatively, chemically labeled nucleotides can be used as the label (see Figure 7–18).

Sarepta Exhibit 1056, page 12

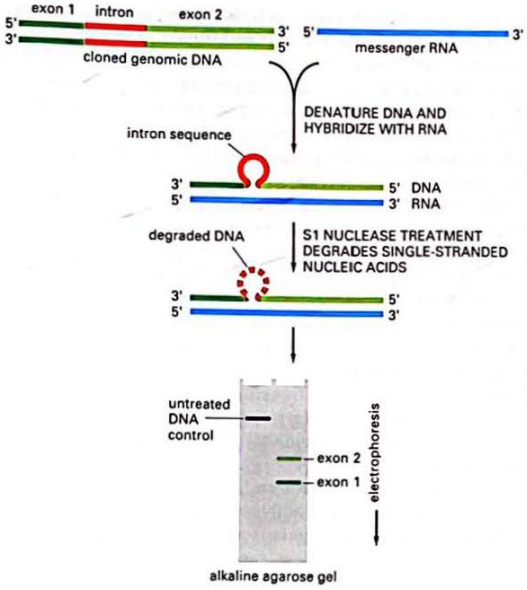


Figure 7–12 The use of nucleic acid hybridization to determine the region of a cloned DNA fragment that is present in an mRNA molecule. The method shown requires a nuclease that cuts the DNA chain only where it is not base-paired to a complementary RNA chain. The positions of the introns (intervening sequences) in eucaryotic genes are mapped by the method shown; the beginning and the end of an RNA molecule can be determined in the same way.

sion of a gene changes, one can determine whether the change is due to controls that act on the transcription of DNA, the splicing of the gene's RNA, or the translation of its mature mRNA molecules into protein. Hybridization methods are in such wide use in cell biology today that it is difficult to imagine what it would be like to study gene structure and expression without them.

Northern and Southern Blotting Facilitate Hybridization with Electrophoretically Separated Nucleic Acid Molecules [10]

DNA probes are often used following gel electrophoresis to detect the nucleic acid molecules with sequences that are complementary to all or part of the probe. The electrophoresis fractionates the many different RNA or DNA molecules in a crude mixture according to their size before the hybridization reaction is carried out; if molecules of only one or a few sizes become labeled with the probe, one can be certain that the hybridization was indeed specific. Moreover, the size information obtained can be invaluable in itself. An example will illustrate this point.

Suppose that one wishes to determine the nature of the defect in a mutant mouse that produces abnormally low amounts of *albumin*, a protein that liver cells normally secrete into the blood in large amounts. First, one collects identical samples of liver tissue from mutant and normal mice (the latter serving as controls) and disrupts the cells in a strong detergent to inactivate cellular nucleases that might otherwise degrade the nucleic acids. Next, one separates the RNA and DNA from all of the other cell components: the proteins present are completely denatured and removed by repeated extractions with phenol—a potent organic solvent that is partly miscible with water; the nucleic acids, which remain in the aqueous phase, are then precipitated with alcohol to separate them from the small molecules of the cell. Then one separates the DNA from the RNA by their different solubilities in alcohols and degrades any contaminating nucleic acid of the unwanted type by treatment with a highly specific enzyme—either an RNase or a DNase.

Sarepta Exhibit 1056, page 13

To analyze the albumin-encoding RNAs with a DNA probe, a technique called **Northern blotting** is used. First, the intact RNA molecules purified from mutant and control liver cells are fractionated according to their size into a series of bands by gel electrophoresis. Then, to make the RNA molecules accessible to DNA probes, a replica of the pattern of RNA bands on the gel is made by transferring ("blotting") the fractionated RNA molecules onto a sheet of nitrocellulose paper (nylon). The RNA molecules that hybridize to the labeled DNA probe (because they are complementary to part of the normal albumin gene sequence) are then located by incubating the paper with a solution containing the probe and detecting the hybridized probe by autoradiography or by chemical means (Figure 7–13). The size of the RNA molecules in each band that binds the probe can be determined by reference to bands of RNA molecules of known size (*RNA standards*) that are electrophoresed side by side with the experimental sample. In this way one might discover that liver cells from the mutant mice make albumin RNA in normal amounts and of normal size; alternatively, albumin RNA of normal size might be detected in greatly reduced amounts. Another possibility is that the mutant albumin RNA molecules might be abnormally short and therefore move unusually quickly through the gel; in this case the gel blot could be retested with a series of shorter DNA probes, each corresponding to small portions of the gene, to reveal which part of the normal RNA is missing.

An analogous gel-transfer hybridization method, called **Southern blotting,** analyzes DNA rather than RNA. Isolated DNA is first cut into readily separable fragments with restriction nucleases. The double-stranded fragments are then

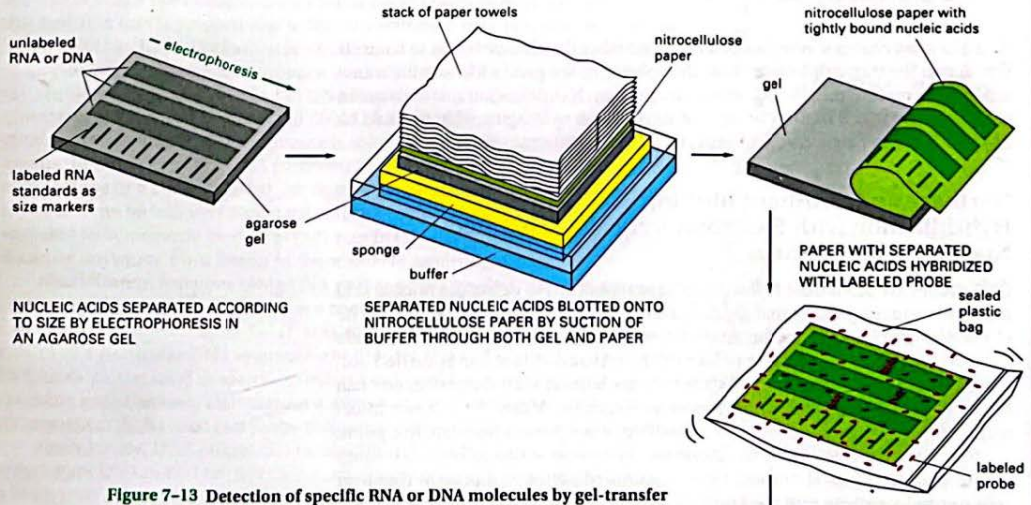


Figure 7–13 Detection of specific RNA or DNA molecules by gel-transfer hybridization. A mixture of either single-stranded RNA molecules (*Northern blotting*) or the double-stranded DNA molecules created by restriction nuclease treatment (*Southern blotting*) is fractionated by electrophoresis. The many different RNA or DNA molecules present are then transferred to nitrocellulose, or nylon, paper by blotting. The paper sheet is exposed to a labeled DNA probe for a prolonged period under conditions favoring hybridization. The sheet is washed thoroughly afterward, so that only those immobilized RNA or DNA molecules that hybridize to the probe become labeled and show up as bands on the paper sheet. For Southern blotting, the strands of the double-stranded DNA molecules on the paper must be separated prior to the hybridization process; this is done by exposing the DNA to alkaline denaturing conditions after the gel has been run (not shown).

Sarepta Exhibit 1056, page 14

separated according to their size by gel electrophoresis, and those complementary to a DNA probe are identified by blotting and hybridization, as just described for RNA (see Figure 7–13). To characterize the structure of the albumin gene in the mutant mice, an albumin-specific DNA probe would be used (after cutting the DNA with different restriction nucleases) to construct a detailed *restriction map* of the genome in the region of the albumin gene. From this map one could determine if the albumin gene has been rearranged in the defective animals—for example, by the deletion or the insertion of a short DNA sequence; most single base changes, however, could not be detected in this way.

RFLP Markers Greatly Facilitate Genetic Approaches to the Mapping and Analysis of Large Genomes [11]

Very large genomes can be mapped either physically or genetically. **Physical maps** are based on direct analysis of the DNA molecules that constitute each chromosome. They include restriction maps and ordered libraries of genomic DNA clones (see Figure 7–29), both of which can be viewed as incomplete representations of the ultimate physical map, which is the complete nucleotide sequence of the genome. **Genetic linkage maps,** by contrast, are based on the frequency of coinheritance of two or more features of the organism that serve as *genetic markers* when individuals interbreed. A genetic marker can be any site in the genome where there is variation in the DNA sequence that is detectable as differences between individuals in a population. If the difference is a rare one, it is called a *mutation;* if it is a common one, it is called a *polymorphism.* Genetic linkage maps are constructed by following the pattern of inheritance of such genetic variants.

Traditionally, linkage mapping has depended on the identification of mutations by their effects on characteristics such as eye color or blood groups. More recently, recombinant DNA methods have made it possible to use as genetic markers short DNA sequences that differ between normal individuals without having a detectable effect on the organism. A widely used genetic marker of this type depends on the way small differences in DNA sequence can alter restriction-enzyme cutting patterns. A single base-pair difference in a particular chromosomal position, for example, may eliminate a restriction-enzyme cutting site, giving rise to a large difference in the lengths of certain restriction fragments produced from the DNA at that position. Similarly, short deletions or insertions will change the size of any DNA restriction fragment in which they occur. Such small differences between individuals are known as **restriction fragment length polymorphisms (RFLPs).** An RFLP is most useful when it is very common in the population, so that there is a high probability that the parents of an individual will carry distinguishable markers. For this reason, short, tandemly repeated sequences whose exact length is highly variable in the population are the basis for the most useful RFLP markers (Figure 7–14).

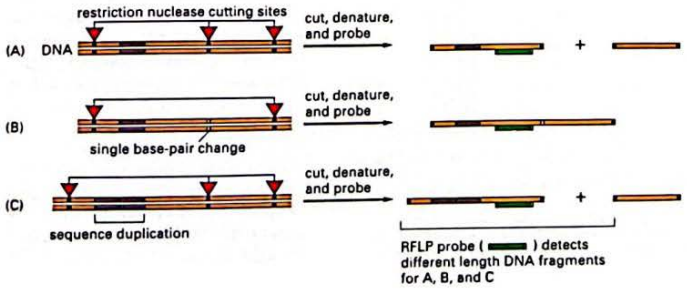


Figure 7–14 Some DNA sequence changes that produce a restriction fragment length polymorphism (RFLP). (A) In many individuals in the population, a restriction nuclease cuts the chromosomal DNA shown at three sites, producing two DNA fragments; one of these DNA fragments is detected with a labeled DNA probe after gel electrophoresis and Southern blotting. (B) In other examples of the same chromosome, a single base-pair change in the short sequence recognized by the restriction nuclease prevents cutting at the central restriction site. The labeled probe now detects a much longer DNA fragment as an RFLP. (C) In yet other examples of the same chromosome, a sequence duplication increases the length of the DNA fragment that is detected by the labeled probe. This type of RFLP is commonly found in regions containing runs of short repeated sequences, such as GTGTGT . . . ; the number of such tandem repeats is found to be highly variable in the population, with each individual typically containing a different number of copies (generally 4 to 40) of the repeat at each particular locus. A repeat of this type is commonly referred to as either a *microsatellite* or a *VNTR (variable number of tandem repeat)* sequence.

Sarepta Exhibit 1056, page 15

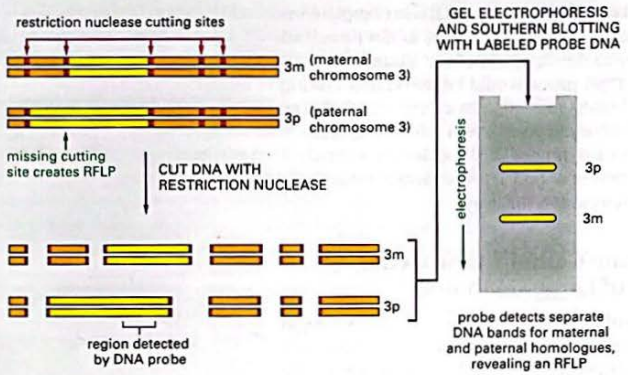


Figure 7–15 Detection of an RFLP by Southern blotting. In this example, the technique is being used to distinguish the maternally and paternally inherited copies of a region of chromosome 3 that are present in a single individual. A variant of this method uses the PCR technique (see Figure 7–32) to amplify the appropriate DNA region selectively before the restriction nuclease treatment. The relevant restriction fragments are then vastly more abundant than other DNA fragments in the preparation and can be seen simply by staining the gel, avoiding the need for blotting and labeled probes.

RFLPs can be used for genetic mapping because the fragment size differences that an individual inherits are readily detected by Southern blotting with a DNA probe complementary to a specific DNA sequence in the region (Figure 7–15). RFLPs provide an immediate way to relate a genetic linkage map to a physical map: on the one hand, they serve as heritable genetic markers, like eye color; on the other hand, the DNA probes that detect them represent DNA sequences whose positions on the physical map are easily found by DNA hybridization.

If two genetic markers are on different chromosomes, the inheritance of a variation in either of them will be *unlinked*—that is, they will have only a 50-50 chance of being inherited together. The same is true for the variants of markers at opposite ends of a single chromosome because of the high probability that they will be separated during the frequent crossing-over that occurs during meiosis in the development of eggs and sperm. The closer together two markers are on the same chromosome, the greater the chance that the form of each marker in the parent will not be separated by crossover events and will therefore be

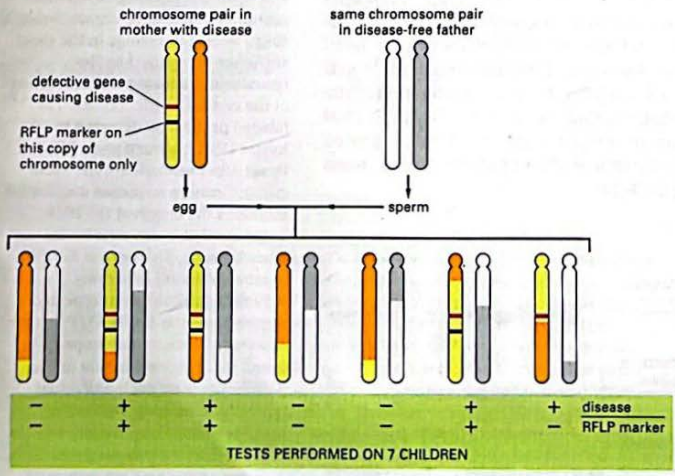


Figure 7–16 Genetic linkage analysis using an RFLP marker. In this procedure one studies the coinheritance of a specific human phenotype (here a genetic disease) with an RFLP marker. If individuals who inherit the disease nearly always inherit the RFLP marker, then the gene causing the disease and the RFLP marker are likely to be close together on the chromosome, as shown here. To prove that an observed linkage is statistically significant, hundreds of individuals may need to be examined. Note that the linkage will not be absolute unless the RFLP marker is located in the gene itself. Thus, occasionally the RFLP marker will be separated from the disease gene by meiotic crossing-over during the formation of the egg or sperm: this has happened in the case of the chromosome pair on the far right.

Sarepta Exhibit 1056, page 16

coinherited by each progeny (*linked*). By screening large family groups for the coinheritance of a gene of interest (such as one associated with a disease) and a large number of individual RFLPs, a few RFLP markers can be identified that are unambiguously coinherited with the gene (this requires the analysis of many individuals, as explained in Figure 7–16). DNA sequences that surround the gene can thereby be located. Eventually the DNA corresponding to the gene itself can be found by techniques to be described later (see Figure 7–30). Many genes that cause human diseases are being isolated in this way, allowing the proteins they encode to be analyzed in detail.

It is clear from Figure 7–16 that the closer an RFLP marker is to the mutant gene of interest, the easier it is to locate the gene unambiguously. To facilitate these and other studies, an intensive effort is underway to prepare a *high-resolution RFLP map* of the human genome, with thousands of RFLP markers spaced an average of 10^6 nucleotide pairs apart. On average, two markers separated by this distance will be coinherited by 99 of every 100 progeny. Such a map will make it relatively easy to use genetic linkage studies to locate a gene that has been identified only by the effect of a mutation in humans, allowing it to be mapped to one or a few large DNA clones in an ordered genomic DNA library. Isolation of the gene might then be accomplished relatively quickly.

Synthetic DNA Molecules Facilitate the Prenatal Diagnosis of Genetic Diseases [12]

At the same time that microbiologists were developing DNA cloning techniques, organic chemists were improving the methods for synthesizing short DNA chains. Today, such synthetic *DNA oligonucleotides* are routinely produced by machines that can automatically synthesize any DNA sequence up to 120 nucleotides long overnight. This ability to produce DNA molecules of a desired sequence makes it possible to redesign genes at will, an important aspect of genetic engineering, as explained later. Such synthetic oligonucleotides can also be used as labeled probes to detect corresponding genomic sequences by DNA hybridization. By varying the temperature at which the hybridization reaction is run, it is possible to vary the *stringency* of the hybridization: above a certain temperature, only perfectly matched sequences will hybridize, and this can make it possible, for example, to detect a mutant gene in the prenatal diagnosis of genetic disease.

More than 3000 human genetic diseases are attributable to single-gene defects. In most of these the mutation is recessive: that is, it shows its effect only when an individual inherits two defective copies of the gene, one from each parent. One goal of modern medicine is to identify those fetuses that carry two copies of the defective gene long before birth so that the mother, if she wishes, can have the pregnancy terminated. In sickle-cell anemia, for example, the exact nucleotide change in the mutant gene is known (the sequence GAG is changed to GTG at a specific point in the DNA strand that codes for the β chain of hemoglobin). For prenatal diagnosis, two DNA oligonucleotides are synthesized—one corresponding to the normal gene sequence in the region of the mutation and the other corresponding to the mutant sequence. If the oligonucleotides are kept short (about 20 nucleotides), they can be hybridized with DNA at a temperature selected so that only the perfectly matched helix will be stable. Such oligonucleotides can thus be used as labeled probes to distinguish between the two forms of the gene by Southern blotting on DNA isolated from fetal cells collected by amniocentesis. A fetus carrying two copies of the mutant β-chain gene can be readily recognized because its DNA will hybridize *only* with the oligonucleotide that is complementary to the mutant DNA sequence.

For many genetic abnormalities the exact nucleotide sequence change is not known. For an increasing number of these, prenatal diagnosis is still possible by using Southern blotting to assay for specific variations in the human genome (the *RFLPs* in Figure 7–16) that are known to be closely linked to the defective gene.

Sarepta Exhibit 1056, page 17

The same techniques can also be used to detect an individual's susceptibility to future disease. Individuals who have inherited abnormal copies of certain genes, for example, have a greatly increased risk of cancer. They need to take protective measures to improve their prospects for a healthy life.

Hybridization at Reduced Stringency Allows Even Distantly Related Genes to Be Identified [13]

New genes arise during evolution by the duplication and divergence of old genes and by the reutilization of portions of old genes in new combinations. For this reason, most genes have a family of close relatives elsewhere in the genome, many of which are likely to have a related function. Laborious methods are usually required to isolate a DNA clone corresponding to the first member of such a *gene family*. Other members of the family, however, can then often be isolated relatively easily by using sequences from the first gene as DNA probes. Because the new genes are unlikely to have *identical* sequences, hybridizations with the DNA probe are usually carried out under conditions of "reduced stringency"—that is, conditions that allow even an imperfect match with the probe sequence to form a stable double helix (Figure 7–17). Although using reduced stringency for hybridization carries the risk of obtaining a false signal from a chance region of short sequence homology in an unrelated DNA sequence, such false positives can be eliminated by further investigation.

This technique is one of the most powerful uses of recombinant DNA technology. It has led, for example, to the isolation of a whole family of DNA-binding proteins that function as master regulators of gene expression during embryonic development in *Drosophila* and also made it possible to isolate members of this same gene family from a variety of other organisms, including mice and humans.

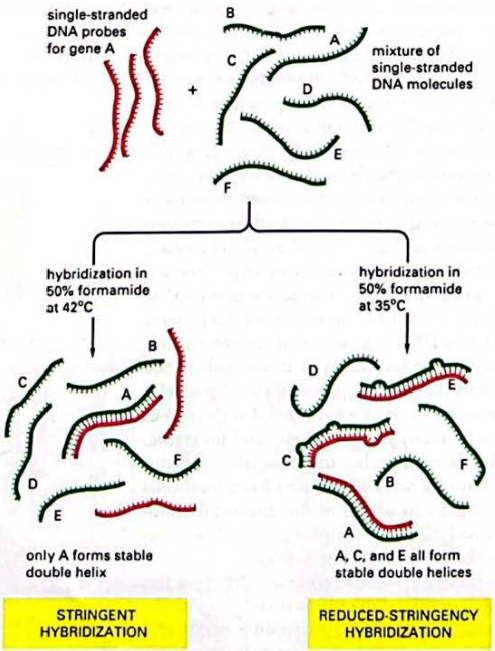


Figure 7–17 Comparison of stringent and reduced-stringency hybridization conditions. In the reaction on the left (stringent conditions), the solution is kept only a few degrees below the temperature at which a perfect DNA helix denatures (its *melting temperature*), so that the imperfect helices that can form under the conditions of reduced stringency are unstable. Only the hybridization conditions on the right can be used to find genes that are nonidentical but related to gene A.

Sarepta Exhibit 1056, page 18

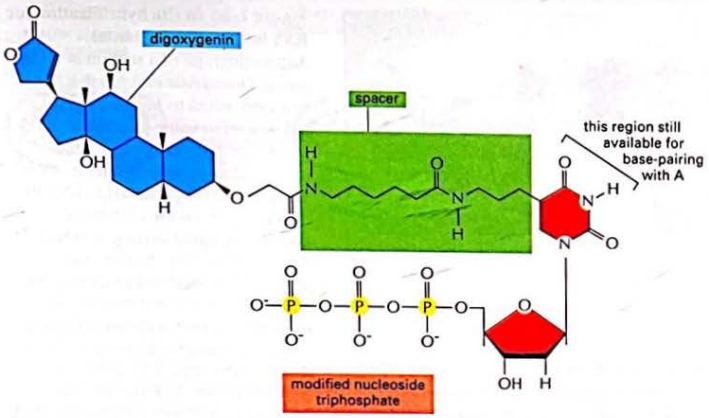


Figure 7–18 A chemical label for DNA probes. The modified nucleotide shown can be incorporated into DNA by DNA polymerase so as to allow the DNA molecule to serve as a probe that can be readily detected. The base on the nucleoside triphosphate shown is an analogue of thymine, in which the methyl group on T has been replaced by a spacer arm linked to the plant steroid digoxygenin. To visualize the probe, the digoxygenin is detected by a specific antibody coupled to a visible marker such as a fluorescent dye. Other chemical labels such as biotin can be attached to nucleotides and used in essentially the same way.

In Situ Hybridization Techniques Locate Specific Nucleic Acid Sequences in Cells or on Chromosomes [14]

Nucleic acids, no less than other macromolecules, occupy precise positions in cells and tissues, and a great deal of potential information is lost when these molecules are extracted by homogenization. For this reason, techniques have been developed in which nucleic acid probes are used in much the same way as labeled antibodies to locate specific nucleic acid sequences *in situ*, a procedure called ***in situ* hybridization.** This can now be done both for DNA in chromosomes and for RNA in cells. Labeled nucleic acid probes can be hybridized to chromosomes that have been exposed briefly to a very high pH to disrupt their DNA base pairs. The chromosomal regions that bind the probe during the hybridization step are then visualized. Originally, this technique was developed using highly radioactive DNA probes, which were detected by autoradiography. The spatial resolution of the technique, however, can be greatly improved by labeling the DNA probes chemically instead of radioactively. For this purpose the probes are synthesized with special nucleotides that contain a modified side chain (Figure 7–18), and the hybridized probes are detected with an antibody (or other ligand) that specifically recognizes this side chain (Figure 7–19).

In situ hybridization methods have also been developed that reveal the distribution of specific RNA molecules in cells in tissues. In this case the tissues are not exposed to a high pH, so the chromosomal DNA remains double-stranded and cannot bind the probe. Instead the tissue is gently fixed so that its RNA is retained in an exposed form that will hybridize when the tissue is incubated with a complementary DNA or RNA probe. In this way the patterns of differential gene expression can be observed in tissues. In the *Drosophila* embryo, for example, such patterns have provided new insights into the mechanisms that create distinctions between cells in different positions during development (Figure 7–20).



Figure 7–19 *In situ* hybridization to locate specific genes on chromosomes. Here, six different DNA probes have been used to mark the location of their respective nucleotide sequences on human chromosome 5 at metaphase. The probes have been chemically labeled and detected with fluorescent antibodies. Both copies of chromosome 5 are shown, aligned side by side. Each probe produces two dots on each chromosome, since a metaphase chromosome has replicated its DNA and therefore contains two identical DNA helices (see Figure 8–4). (Courtesy of David C. Ward.)

Summary

In nucleic acid hybridization reactions, single strands of DNA or RNA randomly collide with one another, rapidly testing out billions of possible alignments. Under stringent hybridization conditions (a combination of solvent and temperature where a perfect double helix is barely stable), two strands will pair to form a "hybrid" helix only if their nucleotide sequences are almost perfectly complementary. The enormous specificity of this hybridization reaction allows any single-stranded sequence of nucleotides to be labeled with a radioisotope or chemical and used as a probe to find

Sarepta Exhibit 1056, page 19

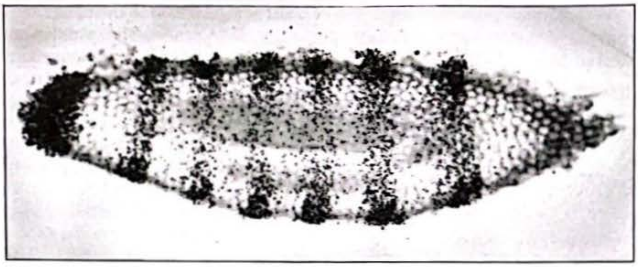


Figure 7–20 ***In situ*** **hybridization for RNA localization in tissues.** Autoradiograph of a section of a very young *Drosophila* embryo that has been subjected to *in situ* hybridization using a radioactive DNA probe complementary to a gene involved in segment development. The probe has hybridized to RNA in the embryo, and the pattern of autoradiographic silver grains reveals that the RNA made by the gene (called *ftz*) is localized in alternating stripes across the embryo that are three or four cells wide. At this stage of development (cellular blastoderm), the embryo contains about 6000 cells. (From E. Hafen, A. Kuriowa, and W.J. Gehring, *Cell* 37:833–841, 1984. © Cell Press.)

a complementary partner strand, even in a cell or cell extract that contains millions of different DNA and RNA sequences. Probes of this type are widely used to detect the nucleic acids corresponding to specific genes, both to facilitate the purification and characterization of the genes after cell lysis and to localize them in cells, tissues, and organisms. Moreover, by carrying out hybridization reactions under conditions of "reduced stringency," a probe prepared from one gene can be used to find its evolutionary relatives—both in the same organism, where the relatives form part of a gene family, and in other organisms, where the evolutionary history of the nucleotide sequence can be traced.

DNA Cloning [15]

In **DNA cloning,** a DNA fragment that contains a gene of interest is inserted into the purified DNA genome of a self-replicating genetic element—generally a virus or a plasmid. A DNA fragment containing a human gene, for example, can be joined in a test tube to the chromosome of a bacterial virus, and the new *recombinant DNA molecule* can then be introduced into a bacterial cell. Starting with only one such recombinant DNA molecule that infects a single cell, the normal replication mechanisms of the virus can produce more than 10^{12} identical virus DNA molecules in less than a day, thereby amplifying the amount of the inserted human DNA fragment by the same factor. A virus or plasmid used in this way is known as a *cloning vector,* and the DNA propagated by insertion into it is said to have been *cloned.*

A DNA Library Can Be Made Using Either Viral or Plasmid Vectors [16]

In order to clone a specific gene, one begins by constructing a *DNA library*—a comprehensive collection of cloned DNA fragments, including (one hopes) at least one fragment that contains the gene of interest. The library can be constructed using either a virus or a plasmid vector and is generally housed in a population of bacterial cells. The principles underlying the methods used for cloning genes are the same for either type of cloning vector, although the details may be different. For simplicity, in this chapter we ignore these differences and illustrate the methods with reference to plasmid vectors.

The **plasmid vectors** used for gene cloning are small circular molecules of double-stranded DNA derived from larger plasmids that occur naturally in bacterial cells. They generally account for only a minor fraction of the total host bacterial cell DNA, but they can easily be separated on the basis of their small size from chromosomal DNA molecules, which are large and precipitate as a pellet upon centrifugation. For use as cloning vectors, the purified plasmid DNA circles are first cut with a restriction nuclease to create linear DNA molecules. The cel-

Sarepta Exhibit 1056, page 20

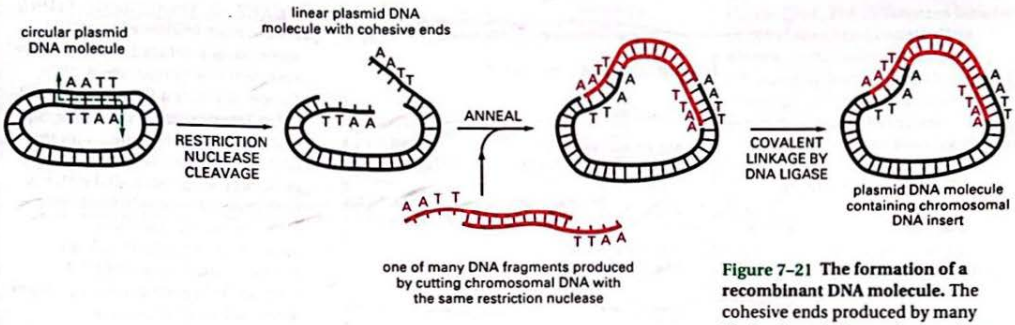


Figure 7–21 The formation of a recombinant DNA molecule. The cohesive ends produced by many kinds of restriction nucleases allow two DNA fragments to join by complementary base-pairing (see Figure 7–2). DNA fragments joined in this way can be covalently linked in a highly efficient reaction catalyzed by the enzyme DNA ligase. In this example a recombinant plasmid DNA molecule containing a chromosomal DNA insert is formed.

lular DNA to be used in constructing the library is cut with the same restriction nuclease, and the resulting restriction fragments (including those containing the gene to be cloned) are then added to the cut plasmids and annealed via their cohesive ends to form recombinant DNA circles. These recombinant molecules containing foreign DNA inserts are then covalently sealed with the enzyme DNA ligase (Figure 7–21).

In the next step in preparing the library, the recombinant DNA circles are introduced into bacterial cells that have been made transiently permeable to DNA; such cells are said to be *transfected* with the plasmids. As these cells grow and divide, doubling in number every 30 minutes, the recombinant plasmids also replicate to produce an enormous number of copies of DNA circles containing the foreign DNA (Figure 7–22). Many bacterial plasmids carry genes for antibiotic resistance, a property that can be exploited to select those cells that have been successfully transfected; if the bacteria are grown in the presence of the antibiotic, only cells containing plasmids will survive. Each original bacterial cell that was initially transfected will, in general, contain a different foreign DNA insert; this insert will be inherited by all of the progeny cells of that bacterium, which together form a small colony in a culture dish.

The mixture of many different surviving bacteria contains the DNA library, composed of a large number of different DNA inserts. The problem is that only a few of the bacteria will harbor the particular recombinant plasmids that contain the desired gene. One needs to be able to identify these rare cells in order to recover the DNA of interest in pure form and in useful quantities. Before discussing how this is achieved, we need to describe a second strategy for generating a DNA library that is commonly used in gene cloning.

Two Types of DNA Libraries Serve Different Purposes [17]

Cleaving the entire genome of a cell with a specific restriction nuclease as just described is sometimes called the "shotgun" approach to gene cloning. It produces a very large number of DNA fragments—on the order of a million for a mammalian genome—which will generate millions of different colonies of transfected bacterial cells. Each of these colonies will be composed of a *clone* derived from a single ancestor cell and therefore will harbor a recombinant plasmid with the same inserted genomic DNA sequence. Such a plasmid is said to contain a **genomic DNA clone,** and the entire collection of plasmids is said to constitute

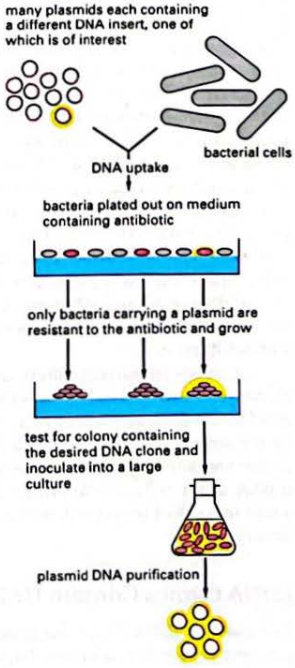


Figure 7–22 Purification and amplification of a specific DNA sequence by DNA cloning in a bacterium. Each bacterial cell carrying a recombinant plasmid develops into a colony of identical cells, visible as a spot on the nutrient agar. By inoculating a single colony of interest into a liquid culture, one can obtain a large number of identical plasmid DNA molecules, each containing the same DNA insert.

Sarepta Exhibit 1056, page 21

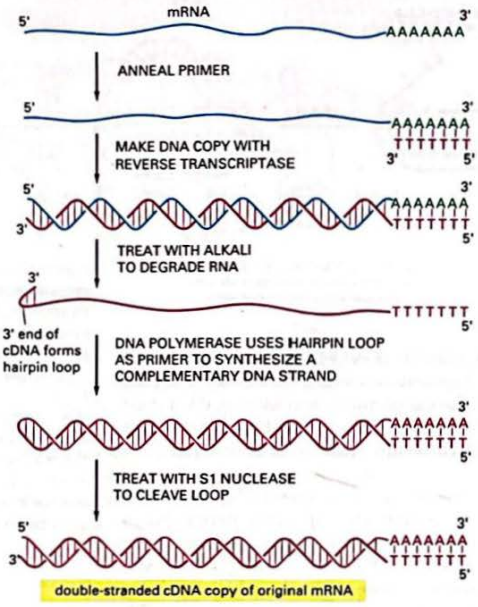


Figure 7–23 The synthesis of cDNA. A DNA copy (cDNA) of an mRNA molecule is produced by the enzyme reverse transcriptase (see p. 282), thereby forming a DNA/RNA hybrid helix. Treating the DNA/RNA hybrid with alkali selectively degrades the RNA strand into nucleotides. The remaining single-stranded cDNA is then copied into double-stranded cDNA by the enzyme DNA polymerase. As indicated, both reverse transcriptase and DNA polymerase require a primer to begin their synthesis. For reverse transcriptase a small oligonucleotide is used; in this example oligo(dT) has been annealed with the long poly-A tract at the 3′ end of most mRNAs. Note that the double-stranded cDNA molecule produced here lacks cohesive ends; such blunt-ended DNA molecules can be cloned by one of several procedures that are analogous to (but less efficient than) that shown in Figure 7–21.

a **genomic DNA library.** But because the genomic DNA is cut into fragments at random, only some fragments will contain genes; many will contain only a portion of a gene, while most of the genomic DNA clones obtained from the DNA of a higher eucaryotic cell will contain only noncoding DNA, which, as we shall discuss in Chapter 8, makes up most of the DNA in such genomes.

An alternative strategy is to begin the cloning process by selecting only those DNA sequences that are transcribed into RNA and thus are presumed to correspond to genes. This is done by extracting the mRNA (or a purified subfraction of the mRNA) from cells and then making a **complementary DNA (cDNA)** copy of each mRNA molecule present; this reaction is catalyzed by the *reverse transcriptase* enzyme of retroviruses, which synthesizes a DNA chain on an RNA template. The single-stranded DNA molecules synthesized by the reverse transcriptase are converted into double-stranded DNA molecules by DNA polymerase, and these molecules are inserted into a plasmid or virus vector and cloned (Figure 7–23). Each clone obtained in this way is called a **cDNA clone,** and the entire collection of clones derived from one mRNA preparation constitutes a **cDNA library.**

There are important differences between genomic DNA clones and cDNA clones, as illustrated in Figure 7–24. Genomic clones represent a random sample of all of the DNA sequences in an organism and, with very rare exceptions, will be the same regardless of the cell type used to prepare them. By contrast, cDNA clones contain only those regions of the genome that have been transcribed into mRNA; as the cells of different tissues produce distinct sets of mRNA molecules, a different cDNA library will be obtained for each type of cell used to prepare the library.

cDNA Clones Contain Uninterrupted Coding Sequences [18]

The use of a cDNA library for gene cloning has several advantages. First, some proteins are produced in very large quantities by specialized cells. In this case,

Sarepta Exhibit 1056, page 22

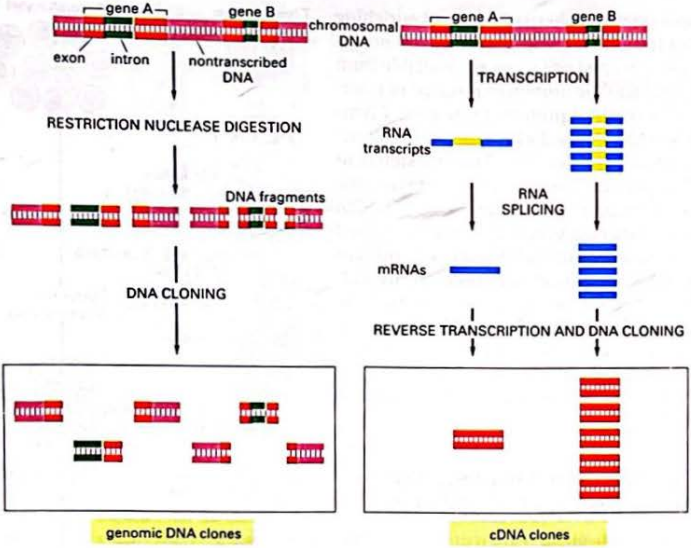


Figure 7–24 The differences between cDNA clones and genomic DNA clones. In this example gene A is infrequently transcribed while gene B is frequently transcribed, and both genes contain introns (*green*). In the genomic DNA clones both the introns and the nontranscribed DNA are included, and most clones will contain only part of the coding sequence of a gene. In the cDNA clones the intron sequences have been removed by RNA splicing during the formation of the mRNA, and a continuous coding sequence is therefore present.

the mRNA encoding the protein is likely to be produced in such large quantities that a cDNA library prepared from the cells will be highly enriched for the cDNA molecules encoding the protein, greatly reducing the problem of identifying the desired clone in the library (see Figure 7–24). Hemoglobin, for example, is made in large amounts by developing erythrocytes (red blood cells); for this reason the globin genes were among the first to be cloned.

By far the most important advantage of cDNA clones is that they contain the uninterrupted coding sequence of a gene. Eucaryotic genes usually consist of short coding sequences of DNA (exons) separated by longer noncoding sequences (introns); the production of mRNA entails the removal of the noncoding sequences from the initial RNA transcript and the splicing together of the coding sequences. Neither bacterial nor yeast cells will make these modifications to the RNA produced from a gene of a higher eucaryotic cell. Thus, if the aim of the cloning is either to deduce the amino acid sequence of the protein from the DNA or to produce the protein in bulk by expressing the cloned gene in a bacterial or yeast cell, it is much preferable to start with cDNA.

Genomic and cDNA libraries are inexhaustible resources that are widely shared among investigators. Today, many such libraries are also available from commercial sources.

cDNA Libraries Can Be Prepared from Selected Populations of mRNA Molecules [19]

When cDNAs are prepared from cells that express the gene of interest at extremely high levels, the majority of cDNA clones may contain the gene sequence, which can therefore be selected with minimal effort. For less abundantly transcribed genes, various methods can be used to enrich for particular mRNAs before making the cDNA library. If an antibody against the protein is available, for example, it can be used to precipitate selectively those polyribosomes (see pp. 237–238) that have the appropriate growing polypeptide chains attached to them. Since these polyribosomes will also have attached to them the mRNA coding for the protein, the precipitate may be enriched in the desired mRNA by as much as 1000-fold.

Sarepta Exhibit 1056, page 23

Subtractive hybridization provides a powerful alternative way of enriching for particular nucleotide sequences prior to cDNA cloning. This selection procedure can be used, for example, if two closely related cell types are available from the same organism, only one of which produces the protein or proteins of interest. It was first used to identify cell-surface receptor proteins present on T lymphocytes but not on B lymphocytes. It can also be used wherever a cell that expresses the protein has a mutant counterpart that does not. The first step is to synthesize cDNA molecules using the mRNA from the cell type that makes the protein of interest. These cDNAs are then hybridized with a large excess of mRNA molecules from the second cell type. Those rare cDNA sequences that fail to find a complementary mRNA partner are likely to represent mRNA sequences present only in the first cell type. Because these cDNAs remain unpaired after the hybridization, they can be purified by a simple biochemical procedure (a hydroxyapatite column) that separates single-stranded from double-stranded nucleic acids (Figure 7–25). Besides providing a powerful way to clone genes whose products are known to be restricted to a specific differentiated cell type, cDNA libraries prepared after subtractive hybridization are useful for defining the differences in gene expression between any two related types of cells.

Either a DNA Probe or a Test for Expressed Protein Can Be Used to Identify the Clones of Interest in a DNA Library [20]

The most difficult part of gene cloning is often the identification of the rare colonies in the library that contain the DNA fragment of interest. This is especially true in the case of a genomic library, where one has to identify one bacterial cell in a million to select a specific mammalian gene. The technique most frequently used is a form of *in situ* hybridization that takes advantage of the exquisite specificity of the base-pairing interactions between two complementary nucleic acid molecules. Culture dishes containing the growing bacterial colonies are blotted with a piece of filter paper, to which some members of each bacterial colony adhere. The adhering colonies, known as *replicas,* are treated with alkali to disrupt the cells and to separate the strands of their DNA molecules; the paper is then incubated with either a radioactive or a chemically labeled DNA probe containing part of the sequence of the gene being sought (Figure 7–26). If necessary, millions of bacterial clones can be screened in this way to find the one clone that hybridizes with the probe.

In order to find the clone of interest, a specific probe must be made. How this is done will depend on the information that is available about the gene to be cloned. In many cases the protein of interest has been identified by biochemical studies and purified in small amounts. Only a few micrograms of pure protein are often enough to determine the sequence of 30 or so amino acid residues. From this amino acid sequence the corresponding nucleotide sequence can be deduced using the genetic code (with some ambiguities corresponding to amino acids that can be represented by several alternative codons). Two sets of DNA oligonucleotides, chosen to match different parts of the predicted nucleotide sequence of the gene, are then synthesized by chemical methods (Figure 7–27). Colonies of cells that hybridize with both sets of DNA probes are strong candidates for containing the desired gene and are saved for further characterization (see below).

Probes can also be obtained in other ways. If an antibody is available that recognizes the protein produced by the gene, it can be labeled and used as a probe to find a clone that is producing the protein, which therefore contains the desired gene. Any other ligand that is known to bind to the protein encoded by the gene can also be used as a probe: if the gene encodes a receptor protein, for example, the ligand that normally binds to the receptor can, in principle, be used as a probe.

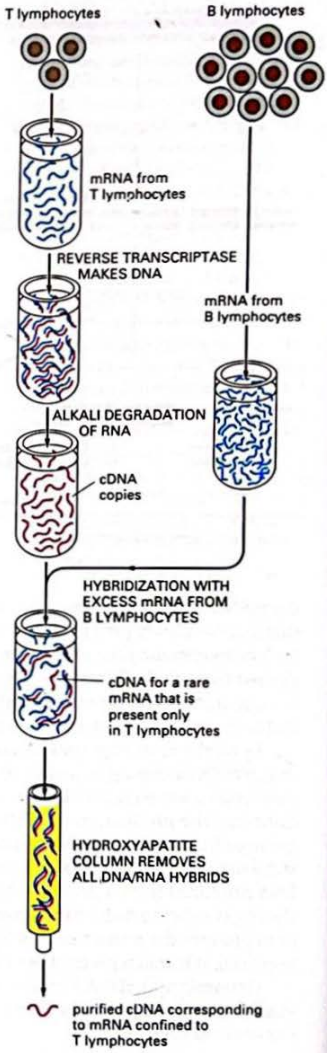


Figure 7–25 Subtractive hybridization. In this example the technique is used to purify rare cDNA clones corresponding to mRNA molecules present in T lymphocytes but not in B lymphocytes. Because the two cell types are very closely related, most of the mRNAs will be common to both cell types; subtractive hybridization is thus a powerful way to enrich for those specialized molecules that distinguish the two cells.

Sarepta Exhibit 1056, page 24

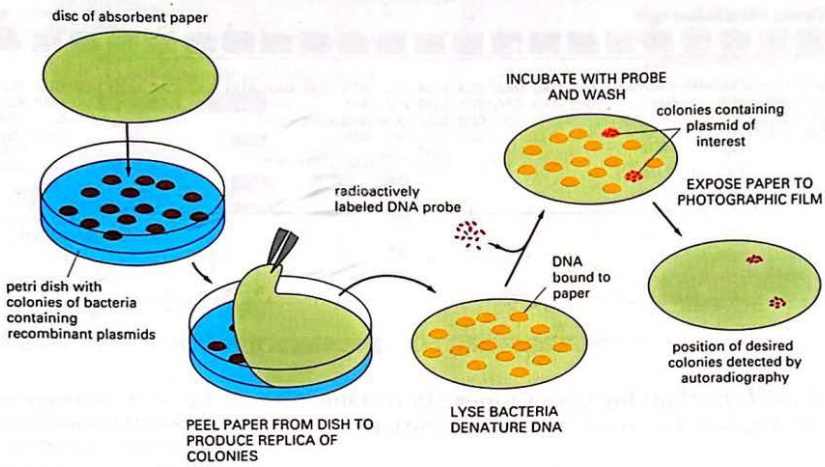


Figure 7–26 An efficient technique commonly used to detect a bacterial colony carrying a particular DNA clone. A replica of the culture is made by pressing a piece of absorbent paper against the surface. This replica is treated with alkali (to disrupt the cells and denature the plasmid DNA) and then hybridized to a highly radioactive DNA probe. Those bacterial colonies that have bound the probe are identified by autoradiography. (See also Figure 7–22.)

Whenever the protein product of a gene is to be detected rather than the gene itself, a special type of cDNA library is required. It is prepared in a special plasmid or virus called an *expression vector*, which directs the transfected bacterium to synthesize large amounts of the protein encoded by the foreign DNA insert contained within the vector's DNA, as we shall discuss later.

In Vitro Translation Facilitates Identification of the Correct DNA Clone [21]

Any method that is used to find a specific clone from a cDNA or genomic DNA library will usually pick out many false positive clones. Further ingenuity is required to discriminate between these and the authentic clones desired. The task is easiest when the desired clone encodes a protein that has already been characterized by other means. In this case each candidate DNA can be tested by one of several methods for its ability to encode the appropriate protein. The cloned DNA can be inserted into an expression vector, for example, so that the protein that it encodes is produced in large amounts in a bacterium. Alternatively, the cloned DNA can be used to obtain a corresponding RNA molecule, either through *in vitro* synthesis with a purified RNA polymerase (see Figure 7–36) or by a technique called *hybrid selection*. In the latter method a mixture of cellular RNAs is added to an excess of single strands of the candidate DNA, and DNA/RNA hybridization is used to purify complementary mRNA molecules from the mixture. In either case the mRNA obtained is allowed to direct protein synthesis in a cell-free system using radioactive amino acids, and the radioactive protein produced is then characterized and compared with the expected protein product of the desired clone. A match in any of these tests allows one to conclude that a cloned DNA fragment encodes the correct protein.

Sarepta Exhibit 1056, page 25

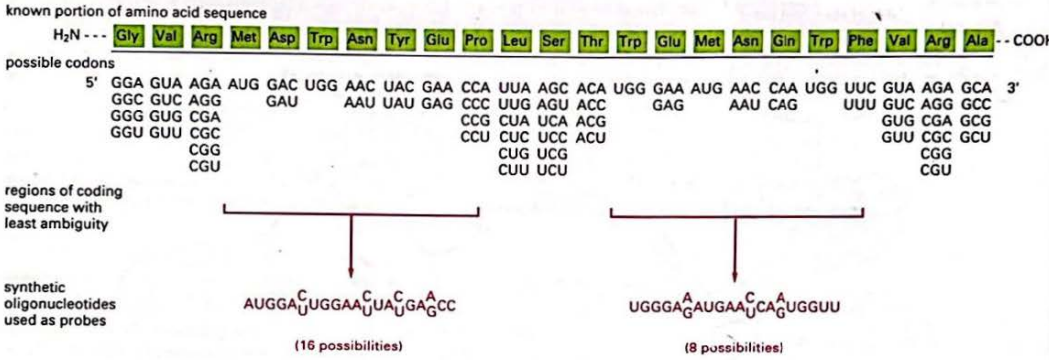


Figure 7–27 Selecting regions of a known amino acid sequence to make synthetic oligonucleotide probes. Although only one nucleotide sequence will actually code for the protein, the degeneracy of the genetic code means that several different nucleotide sequences will give the same amino acid sequence, 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 5′ 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).

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

Sarepta Exhibit 1056, page 26

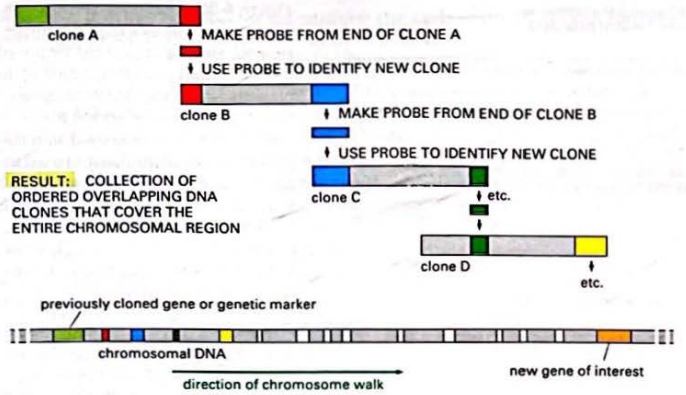


Figure 7–28 The use of overlapping DNA clones to find a new gene by "chromosome walking." To speed up the walk, genomic libraries containing very large cloned DNA molecules are optimal. To probe for the next clone in the walk by DNA hybridization, a short DNA fragment (labeled with a chemical or a radioisotope) 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.

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 Functions [24]

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

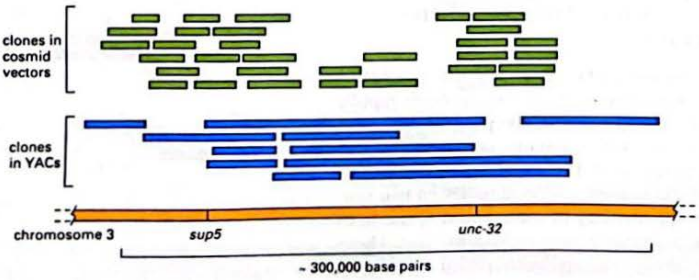


Figure 7–29 Overlapping genomic DNA clones. The collection of clones shown covers a small region of a chromosome of the nematode worm *Caenorhabditis elegans* and represents 0.3% of the total genome. (Adapted from J. Sulston et al., *Nature* 356:37–41, 1992. © 1992 Macmillan Magazines Ltd.)

Sarepta Exhibit 1056, page 27

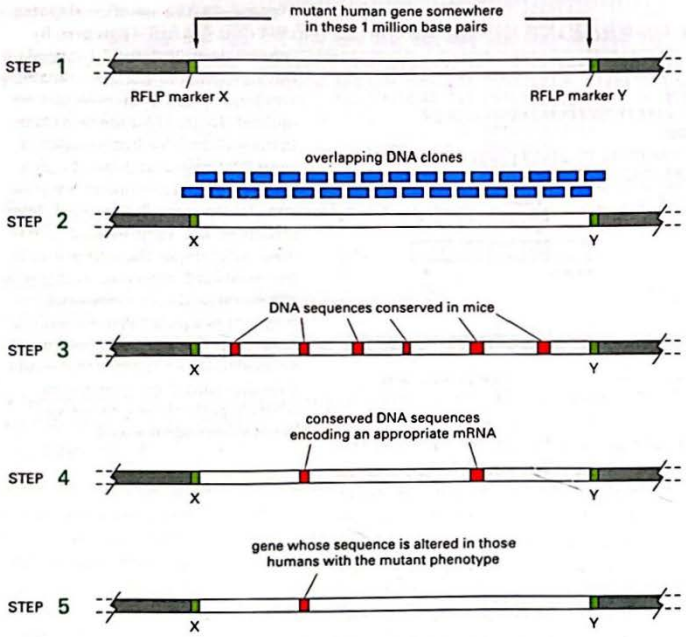


Figure 7–30 Positional cloning. The procedure requires a mutant human gene whose inheritance can be traced in many family groups by virtue of the phenotype that the mutation causes. Step 1, *genetic mapping:* RFLP markers that are coinherited with the phenotype are identified and used to position the 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.

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 Reaction [25]

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

Sarepta Exhibit 1056, page 28

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).

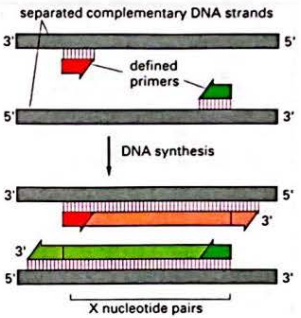
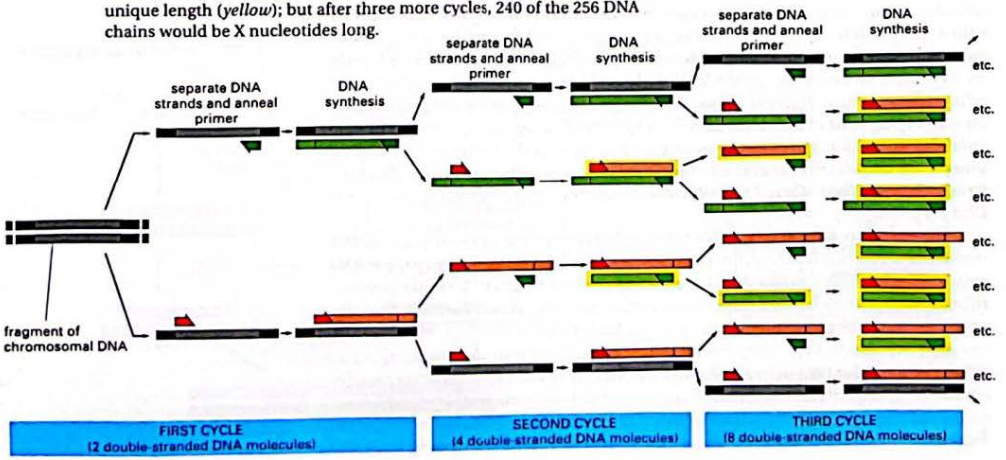


Figure 7–31 The start of the polymerase chain reaction (PCR) for amplifying specific nucleotide sequences *in vitro*. DNA isolated from cells is heated to separate its complementary strands. These strands are then annealed with an excess of two DNA oligonucleotides (each 15 to 20 nucleotides long) that 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.



Sarepta Exhibit 1056, page 29

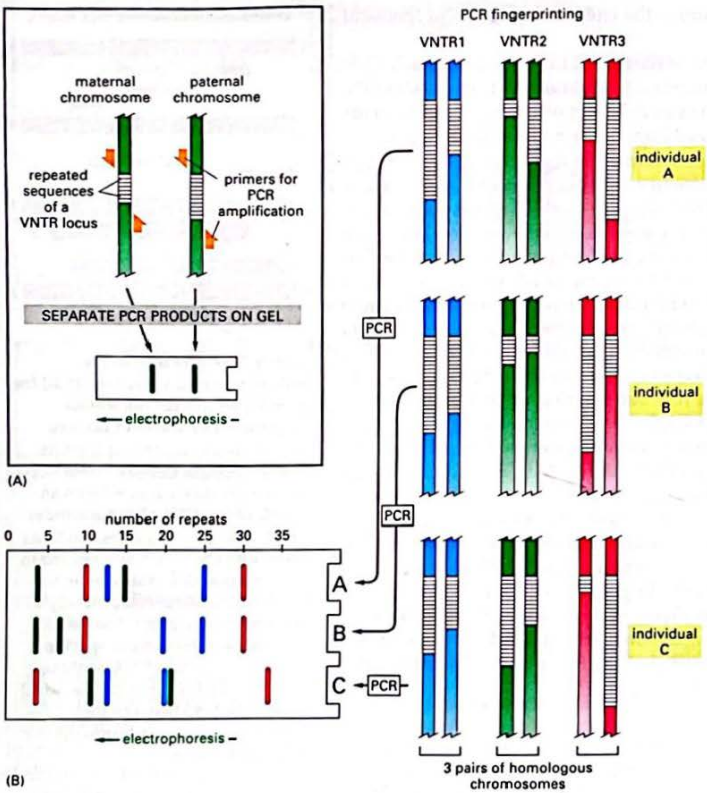


Figure 7–33 The use of PCR in forensic science. (A) A PCR reaction using two primers that bracket a particular microsatellite, or VNTR, sequence (see Figure 7–14C) produces a different pair of DNA bands from each individual. One of these bands contains the repeated VNTR sequence that was inherited from the individual's mother and the other contains the repeated VNTR sequence that was inherited from the individual's father. (B) The large set of DNA bands obtained from a set of different PCR reactions, each of which amplifies the DNA from a different VNTR sequence, can serve as a "fingerprint" to identify each individual nearly uniquely. The starting material for the PCR reaction can be a single hair that was left at the scene of a crime.

Summary

DNA cloning allows a copy of any specific part of a DNA or RNA sequence to be selected from the millions of other sequences in a cell and produced in unlimited amounts in pure form. DNA sequences are amplified after cutting chromosomal DNA with a restriction nuclease and inserting the resulting DNA fragments into the chromosome of a self-replicating genetic element (a plasmid or a virus). When a plasmid vector is used, the resulting "genomic DNA library" is housed in millions of bacterial cells, each carrying a different cloned DNA fragment. The bacterial colony containing a DNA fragment of interest is identified by hybridization using a DNA probe or, following expression of a cloned gene or gene fragment in the bacterial host cell, by using a test that detects the desired protein product. The cells in the identified bacterial colony are then allowed to proliferate, producing large amounts of the desired DNA fragment.

The procedure used to obtain DNA clones that correspond in sequence to mRNA molecules are the same except that the starting material is a DNA copy of the mRNA sequence, called cDNA, rather than fragments of chromosomal DNA. Unlike genomic DNA clones, cDNA clones lack intron sequences, making them the clones of choice for expressing and characterizing the protein product of a gene.

PCR is a new form of DNA cloning that is carried out outside cells using a purified, thermostable DNA polymerase enzyme. This type of DNA amplification requires a prior knowledge of gene sequence, since two synthetic oligonucleotide primers must be synthesized that bracket the DNA sequence to be amplified. PCR cloning, however, has the advantage of being much faster and easier than standard cloning methods.

Sarepta Exhibit 1056, page 30

DNA Engineering [26]

The methods described thus far in this chapter make it possible, in principle, to determine the exact organization and nucleotide sequence of any chromosome, including all of the chromosomes that constitute the human genome. In this section we describe how extensions of these methods have revolutionized all other aspects of cell biology by providing new ways to study the functions of genes, RNA molecules, and proteins.

New DNA Molecules of Any Sequence Can Be Formed by Joining Together DNA Fragments [27]

As we have seen, recombinant DNA molecules are generally constructed by using the enzyme DNA ligase to join together two DNA molecules with matching cohesive ends. For production of a DNA library, one of the two DNA molecules to be joined is the *vector*, derived from a bacterial plasmid or virus, while the other is either a fragment of a chromosome or a cDNA molecule (see Figure 7–21). A recombinant DNA molecule can in turn serve as a vector for cloning additional DNA molecules, and by the stepwise repetition of the cloning procedure, a series of DNA fragments can be joined together end to end. In this way new DNA molecules can be generated that are different from any molecule that occurs naturally (Figure 7–34).

Automated oligonucleotide synthesizers rapidly produce any DNA molecule containing up to about 100 nucleotides. The sequence of such synthetic DNA molecules is entirely determined by the experimenter, and they can be joined together by repeated DNA cloning steps in various combinations to produce long custom-designed DNAs of any sequence. But the DNA sequences of interest to cell biologists are mostly those that code for protein domains that already exist in nature. The polymerase chain reaction (PCR) can be used both to amplify any natural nucleotide sequence and to redesign its two ends, so that any two natu-

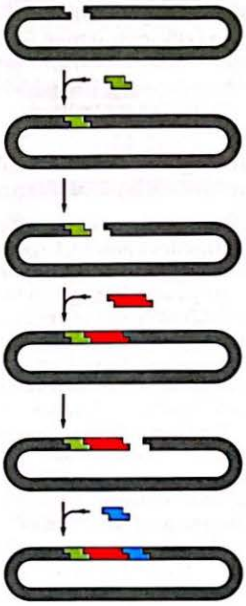


Figure 7–34 Serial DNA cloning can be used to splice together a set of DNA fragments derived from different genes. All of the DNA molecules shown here are double-stranded. After each DNA insertion step, the plasmid is cloned to purify and amplify the new recombinant DNA molecule. The recombinant molecule is then cut once with a restriction nuclease, as indicated, and used as a cloning vector for the next DNA fragment.

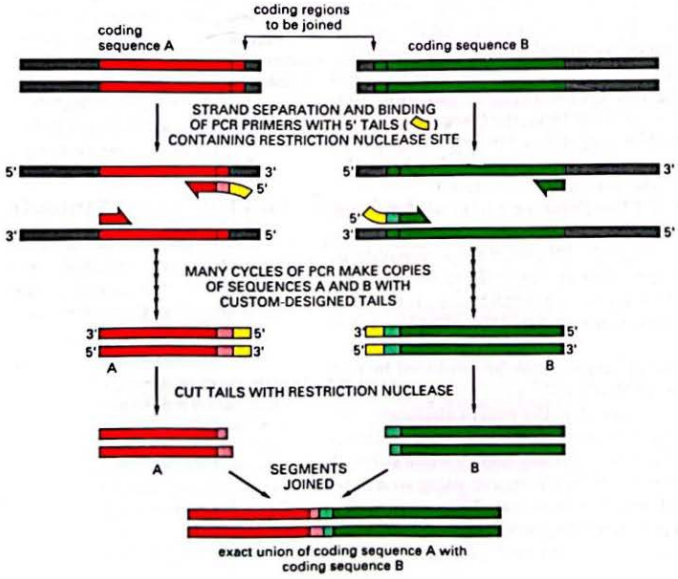


Figure 7–35 Tailoring the ends of DNA fragments facilitates their precise joining. The PCR reaction allows any DNA segment to be amplified with modified ends in preparation for cloning. The synthetic oligonucleotide primers used here contain 5′ tails whose sequence has been chosen to create a particular restriction nuclease cutting site.

Sarepta Exhibit 1056, page 31

rally occurring DNA sequences can be amplified and spliced together very rapidly and efficiently (Figure 7–35). Thus, rather than generating novel DNA molecules from chemically synthesized fragments, it is almost always easier to produce them using PCR and repeated cloning steps to link selected naturally occurring DNA segments.

Homogeneous RNA Molecules Can Be Produced in Large Quantities by DNA Transcription *in Vitro* [28]

Pure RNA molecules are useful for many types of cell biological studies. If the RNA encodes a protein, large amounts of the pure species will facilitate *in vitro* studies of RNA splicing or protein synthesis; if the RNA has a catalytic function, its catalytic mechanism can be tested in cell-free systems. But most species of RNA molecules are present in only tiny quantities in cells, and they are very difficult to purify away from the many thousands of other species present in a typical cell extract. By making available pure DNA templates that can be used to produce large amounts of any RNA molecule, genetic engineering provides the best path to a pure RNA species.

The technique uses the unusually efficient RNA polymerase molecules produced by certain bacterial viruses. To prepare a suitable DNA template, the DNA sequence encoding the RNA is cloned in a way that joins it to a second DNA segment that contains the start signal (promoter) for the viral RNA polymerase enzyme. The pure recombinant DNA molecule is then mixed with a pure preparation of the RNA polymerase plus the four ribonucleoside triphosphates used in RNA synthesis. During a prolonged incubation at 37°C, large amounts of the desired RNA are then generated by *in vitro* transcription (Figure 7–36).

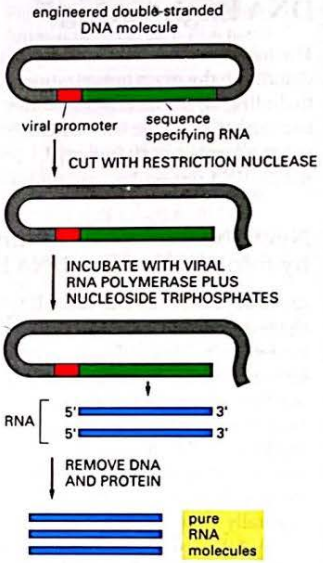


Figure 7–36 Large amounts of any RNA molecule can be prepared *in vitro* using a viral RNA polymerase. The use of this unusually efficient enzyme requires that the DNA template be engineered so that the short DNA sequence that specifies the viral promoter is adjacent to the DNA sequence that encodes the RNA molecule to be synthesized. As indicated, the 3′ end of the RNA is determined by cutting the DNA template at a unique site with a restriction nuclease.

Rare Cellular Proteins Can Be Made in Large Amounts Using Expression Vectors [29]

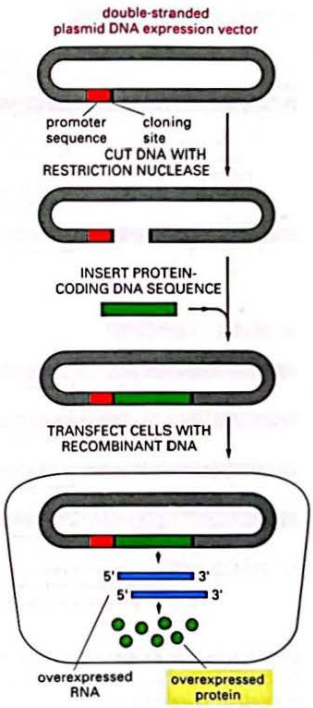
Until recently, the only proteins in a cell that could be studied easily were the relatively abundant ones. Starting with several hundred grams of cells, a major protein—one that constitutes 1% or more of the total cellular protein—can be purified by sequential chromatography steps to yield perhaps 0.1 g (100 mg) of pure protein. This amount is sufficient for conventional amino acid sequencing, for detailed analysis of biological activity (if known), and for the production of antibodies, which can then be used to localize the protein in the cell. Moreover, if suitable crystals can be grown (usually a difficult task), the three-dimensional structure of the protein can be determined by x-ray diffraction techniques. In this way the structure and function of many abundant proteins—including hemoglobin, trypsin, immunoglobulin, and lysozyme—have been analyzed.

The vast majority of the thousands of different proteins in a eucaryotic cell, however, including many with crucially important functions, are present in very small amounts. For most of them it is extremely difficult, if not impossible, to obtain more than a few micrograms of pure material. One of the most important contributions of DNA cloning and genetic engineering to cell biology is that they have made it possible to produce any of the cell's proteins, including the minor ones, in large amounts.

Large amounts of a pure protein could, in principle, be produced by first synthesizing a large amount of the mRNA that encodes the protein and then translating it in a cell-free system that contains all of the many components required for protein synthesis, including ribosomes, transfer RNAs, translation enzymes, and so on. While this approach is occasionally used, a more efficient way is to produce both the mRNA and the protein in a living cell, using an *expression vector* (Figure 7–37). The vector is designed to produce a large amount of a stable mRNA that will be efficiently translated into protein in the transfected bacterial, yeast, insect, or mammalian cell. To prevent the high level of the for-

Sarepta Exhibit 1056, page 32

Figure 7–37 Production of large amounts of a protein by cloning the protein-coding DNA sequence in a plasmid expression vector. The plasmid has been engineered to contain a highly active promoter, which causes unusually large amounts of mRNA encoding the protein to be synthesized in the transfected cells.



eign protein from interfering with the transfected cell's growth, the expression vector is often designed so that the synthesis of the foreign mRNA and protein can be delayed until shortly before the cells are harvested (Figure 7–38).

Because the desired protein made from an expression vector is produced inside a cell, it must be purified away from the host cell proteins by chromatography following cell lysis; but because it is such a plentiful species in the cell lysate (often 1% to 10% of the total cell protein), the purification is usually easy to accomplish in only a few steps. A variety of expression vectors are available, each engineered to function in the type of cell in which the protein is to be made. In this way cells can be induced to make vast quantities of medically useful proteins—such as human insulin and growth hormone, interferon, and viral antigens for vaccines. More generally, these methods make it possible to produce every protein in large enough amounts for the kinds of detailed structural and functional studies that were previously possible for only a rare few.

Reporter Genes Enable Regulatory DNA Sequences to Be Dissected [30]

The transcription of a gene is controlled by regulatory DNA sequences that are not transcribed. These sequences determine which cells will express the gene and under what conditions, as explained in Chapter 9. In higher eucaryotes these regulatory sequences can be located many thousands of base pairs upstream or downstream from the transcribed sequences, and they act in combinations to control gene transcription. The regulatory DNA sequences can be identified for any particular gene by a manipulation that involves replacing part of the gene's coding sequence with a different coding sequence (a so-called *reporter sequence*), selected because the protein it encodes is easily detected by either a simple cytological stain or an enzyme assay. The regulatory sequences for the gene of interest can then be identified by attaching the reporter sequence to various fragments of DNA sequence taken from upstream or downstream of the gene. When such recombinant DNA molecules are used to transfect cells, the level, timing, and cell specificity of reporter protein production will reflect the action of the regulatory sequences that each DNA construct contains (Figure 7–39).

Mutant Organisms Best Reveal the Function of a Gene [31]

Suppose that one has cloned a gene that codes for a protein whose function is unknown. How can one discover what the protein does in the cell? This has become a common problem in cell biology and one that is surprisingly difficult to solve, since neither the three-dimensional structure of the protein nor the com-

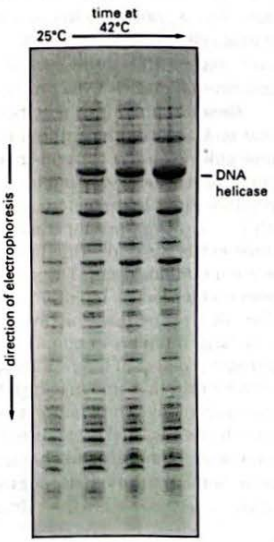


Figure 7–38 Production of large amounts of a protein using a plasmid expression vector. In this example bacterial cells have been transfected with the coding sequence for an enzyme, DNA helicase (*arrow*); transcription from this coding sequence is under the control of a viral promoter that becomes active only at temperatures of 37°C or higher. The total cell protein has been analyzed by SDS-polyacrylamide gel electrophoresis, either from bacteria grown at 25°C (no helicase protein made), or after a shift of the same bacteria to 42°C for up to two hours. (Photograph courtesy of Jack Barry.)

Sarepta Exhibit 1056, page 33

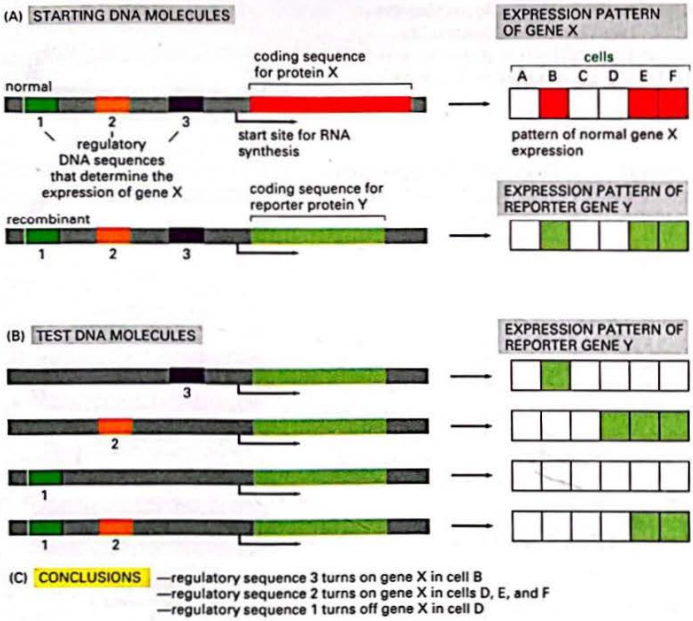


Figure 7–39 Using a reporter protein to locate the regulatory DNA sequences that determine the pattern of a gene's expression. In this example the coding sequence for protein X is replaced by the coding sequence for protein Y, and various fragments of DNA containing candidate regulatory sequences are added in combinations. The recombinant DNA molecules are then tested for expression after their transfection into a variety of different types of mammalian cells. For experiments in eucaryotic cells, two commonly used reporters are the enzymes β-galactosidase (*β-gal*) and chloramphenicol acetyltransferase (*CAT*). Because these are bacterial enzymes, their presence can be monitored by simple and sensitive assays of enzyme activity, without any interference from host cell enzymes.

plete nucleotide sequence of its gene is usually sufficient to deduce the protein's function. Moreover, many proteins—such as those that have a structural role in the cell or normally form part of a large multienzyme complex—will have no obvious activity by themselves.

One approach, discussed in Chapter 4, is to inactivate the protein by injecting into a cell a specific antibody that recognizes the protein and then observing how the cell is affected. Although this provides a useful way to test protein function, the effect is transitory because the injected antibody is eventually diluted out during cell proliferation or destroyed by intracellular proteolytic enzymes. Moreover, many antibodies—even those that bind tightly to some part of the given protein—fail to block the protein's function.

Genetics provides a much more powerful solution to this problem. Mutants that lack a particular protein may quickly reveal the function of the normal molecule. Even more useful are mutants in which the abnormal protein is temperature sensitive, so that it is inactivated by a small increase or decrease in temperature, since in these mutants the abnormality can be switched on and off simply by changing the temperature. Before the advent of gene cloning technology, most genes were identified in this way, according to the processes disrupted when mutations occur. The genetic approach is most easily applicable to organisms that reproduce rapidly—such as bacteria, yeasts, nematode worms, and fruit flies. By treating these organisms with agents that alter their DNA (*mutagens*), very large numbers of mutants can be created quickly and then screened for a particular defect of interest. By screening populations of mutagen-treated bacteria for cells that stop making DNA when they are shifted from 30°C to 42°C, for example, many temperature-sensitive mutants were isolated in the genes that encode the bacterial proteins required for DNA replication. These mutants were later used to identify and characterize the corresponding DNA replication proteins. Similarly, the genetic approach has been used to demonstrate the function of enzymes involved in the principal metabolic pathways of bacteria, as well as

Sarepta Exhibit 1056, page 34

to discover many of the gene products responsible for the orderly development of the *Drosophila* embryo.

Humans do not reproduce rapidly, and they are not, as a rule, intentionally treated with mutagens. Moreover, any human with a serious defect in an essential process, such as DNA replication, would die long before birth. Many mutations that are compatible with life, however—for example, tissue-specific defects in lysosomes or in cell-surface receptors—have arisen spontaneously in the human population. Analyses of the phenotypes of the affected individuals, together with studies of their cultured cells, have provided many unique insights into important cell functions. Although such mutants are rare, they are very efficiently discovered because of a unique human property: the mutant individuals call attention to themselves by seeking special medical care.

Cells Containing Mutated Genes Can Be Made to Order [32]

Although in rapidly reproducing organisms it is often not difficult to obtain mutants that are deficient in a particular process, such as DNA replication or eye development, it can take a long time to trace the defect to a particular altered protein. Recently, recombinant DNA technology has made possible a different type of genetic approach. Instead of starting with a randomly generated mutant and using it to identify a gene and its protein, one can start with a particular gene and proceed to make mutations in it, creating mutant cells or organisms so as to analyze the gene's function. Because the new approach reverses the traditional direction of genetic discovery—proceeding from genes and proteins to mutants, rather than vice versa—it is commonly referred to as *reverse genetics*.

Reverse genetics begins with a cloned gene or a protein with interesting features that has been isolated from a cell. If the starting point is a protein, the gene encoding it is first cloned and its nucleotide sequence is determined. The gene sequence is then altered by biochemical means to create a mutant gene that codes for an altered version of the protein. The mutant gene is transferred into a cell, where it can integrate into a chromosome by genetic recombination to become a permanent part of the cell's genome. If the gene is expressed, the cell and all of its descendants will now synthesize an altered protein.

If the original cell used for the gene transfer is a fertilized egg, whole multicellular organisms can be obtained that contain the mutant gene, and some of these *transgenic organisms* will pass the gene on to their progeny as a permanent part of their germ line. Such *genetic transformations* are now routinely performed with organisms as complex as fruit flies and mammals (see Figure 7–45). Technically, even humans could now be transformed in this way, although such procedures are not undertaken, even for therapeutic purposes, for fear of the unpredictable aberrations that might occur in such individuals.

Genes Can Be Redesigned to Produce Proteins of Any Desired Sequence [33]

To facilitate reverse genetic studies of protein function, both the coding sequence of a gene and its regulatory regions can be altered to change the functional properties of the protein product, the amount of protein made, or the particular cell type in which the protein is produced.

Special techniques are required to alter a gene in such subtle ways: it is often desirable, for example, to change the protein the gene encodes by one or a few amino acids. The first step in redesigning a gene in this way is the chemical synthesis of a short DNA molecule containing the altered portion of the gene's nucleotide sequence. This synthetic DNA oligonucleotide is hybridized with single-stranded plasmid DNA that contains the DNA sequence to be altered, using conditions that allow imperfectly matched DNA strands to pair (Figure 7–40). The synthetic oligonucleotide will now serve as a primer for DNA synthesis

Sarepta Exhibit 1056, page 35

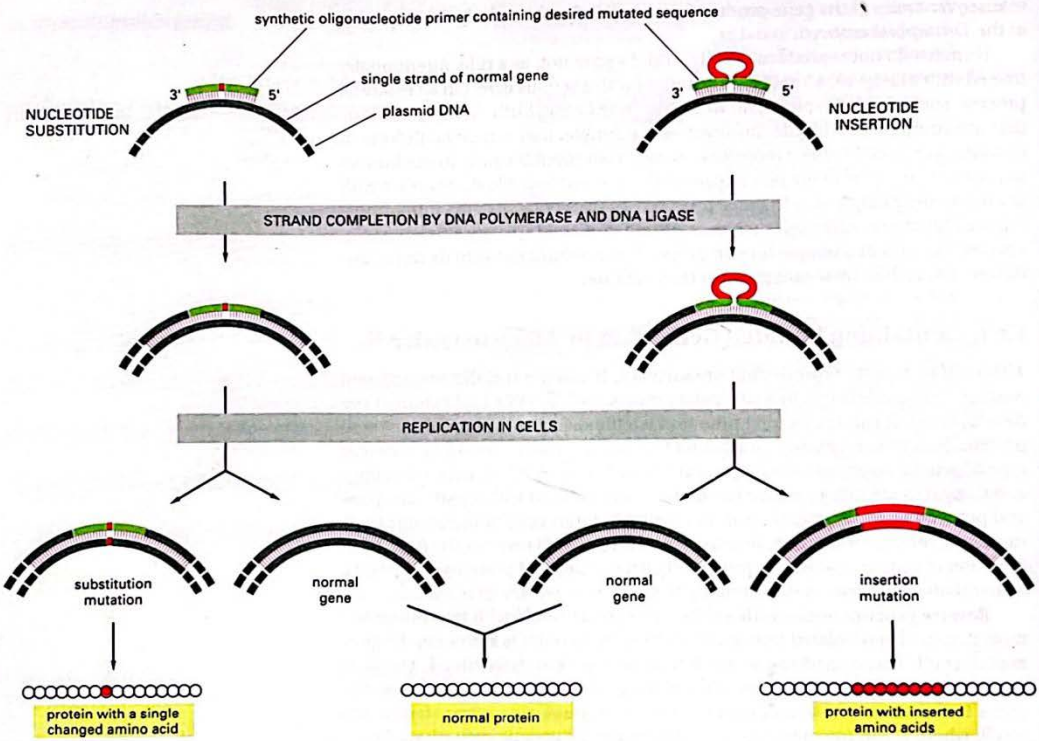


Figure 7–40 The use of synthetic oligonucleotides to modify the protein-coding regions of genes. Only two of the many types of changes that can be engineered in this way are shown. With an appropriate oligonucleotide, for example, more than one amino acid substitution can be made at a time, or one or more amino acids can be deleted. As indicated, because only one of the two DNA strands in the original recombinant plasmid is altered by this procedure, only half of the transfected cells will end up with a plasmid that contains the desired mutant gene. Note that most of the plasmid DNA sequence is not illustrated here.

by DNA polymerase, thereby generating a DNA double helix that incorporates the altered sequence into one of its two strands. After transfection, plasmids that carry the fully modified gene sequence are obtained, and the appropriate DNA is inserted into an expression vector so that the redesigned protein can be produced in the appropriate type of cells for detailed studies of its function. By changing selected amino acids in a protein in this way, one can analyze which parts of the polypeptide chain are important in such fundamental processes as protein folding, protein-ligand interactions, and enzymatic catalysis.

Fusion Proteins Are Often Useful for Analyzing Protein Function [34]

A protein will often contain short amino acid sequences that determine its location in a cell, or its stability after the protein has been synthesized. These special regions of the protein can be identified by fusing them to an easily detected reporter protein that lacks such regions and then following the behavior of the reporter protein in a cell (Figure 7–41). Such **fusion proteins** are produced by the recombinant DNA techniques discussed previously. Most nuclear proteins, for example, contain one or more specific short sequences of amino acids that serve as signals for their import into the nucleus after their synthesis in the cytosol. By artificially attaching different segments of such a nuclear protein to a cytoplasmic protein using gene-fusion techniques, the "signal peptides" responsible for nuclear import have been identified.

Not all of the important signals carried by proteins, however, can be transferred to a large reporter protein in the form of a short amino acid sequence. The

Sarepta Exhibit 1056, page 36

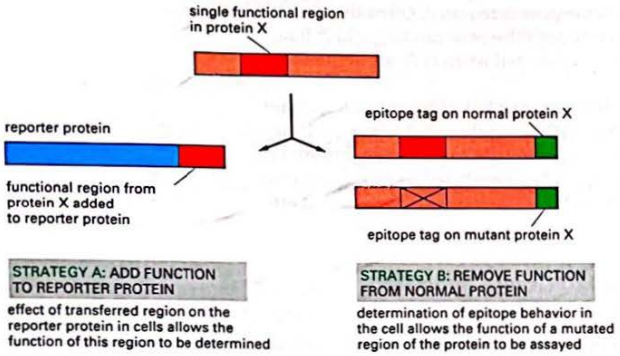


Figure 7–41 Two strategies for the analysis of protein function that exploit genetic engineering. In strategy A, a new function is imparted to a reporter protein by the addition of a small segment of normal protein X. In strategy B, minor alterations are made in the normal protein X and its behavior is followed by detection of a short epitope tag and compared to that of the normal protein. Both strategies have been used to analyze the structure and function of the nuclear localization signals found on nuclear proteins (see Chapter 12).

information that moves newly synthesized lysosomal proteins to lysosomes from the Golgi apparatus, for example, resides in a "signal patch," whose formation requires that the lysosomal protein fold correctly so as to bring distant portions of the polypeptide chain into close proximity. In these cases a different strategy, called **epitope tagging,** can often be used. Again, a fusion protein is produced, but it contains the entire protein being analyzed plus a short peptide of 8 to 12 amino acids (the "epitope") that can be recognized by an antibody—hence its name. The fusion protein can therefore be specifically followed in cells with the anti-epitope antibody, even in the presence of a large excess of the normal protein, so that one can determine the effect of altering any of the amino acids in the fusion protein on its intracellular location (strategy B in Figure 7–41).

Normal Genes Can Be Easily Replaced by Mutant Ones in Bacteria and Some Lower Eucaryotes [35]

Unlike higher eucaryotes (which are multicellular and diploid), bacteria, yeasts, and the cellular slime mold *Dictyostelium* generally exist as haploid single cells. In these organisms an artificially introduced DNA molecule carrying a mutant gene will, with a relatively high frequency, replace the single copy of the normal gene by *homologous recombination* (see p. 263), so that it is easy to produce cells in which the mutant gene has replaced the normal gene (Figure 7–42A). In this way cells can be made to order that produce an altered form of any specific protein or RNA molecule instead of the normal form of the molecule. If the mutant gene is completely inactive and the gene product normally performs an essential function, the cell will die; but in this case a less severely mutated version of the gene can be used to replace the normal gene, so that the mutant cell survives

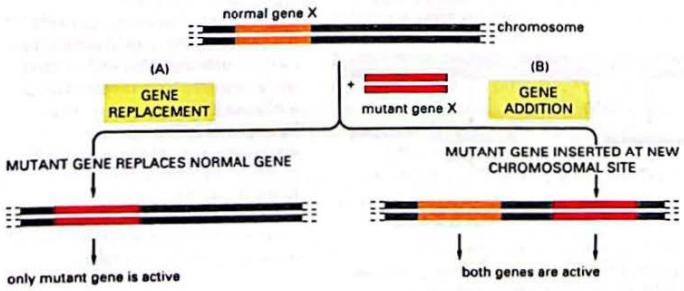


Figure 7–42 Gene replacement and gene addition. A gene whose nucleotide sequence has been altered can be inserted back into the chromosomes of an organism. In bacteria and some haploid eucaryotes such as yeast, the altered gene frequently replaces the normal gene, a process called *gene replacement* (A); in these cases, only the mutant gene remains in the cell. In higher eucaryotes, *gene addition* (B) generally occurs instead of gene replacement; the transformed cell or organism now contains the mutated gene in addition to the normal gene.

Sarepta Exhibit 1056, page 37

but is abnormal in the process for which the gene is required. Often the mutant of choice is one that produces a temperature-sensitive gene product, which functions normally at one temperature but is inactivated when cells are shifted to a higher or lower temperature.

The ability to perform direct *gene replacements* in lower eucaryotes, combined with the power of standard genetic analyses in these haploid organisms, in large part explains why studies in these types of cells have been so important for working out the details of those processes that are shared by all eucaryotes. Gene replacements occur more rarely in higher eucaryotes, for reasons that are not known.

Engineered Genes Can Be Used to Create Specific Dominant Mutations in Diploid Organisms [36]

Higher eucaryotes, such as mammals or fruit flies, are diploid and therefore have two copies of each chromosome. Moreover, transfection with an altered gene generally leads to *gene addition* rather than gene replacement: the altered gene inserts at a random location in the genome, so that the cell (or the organism) ends up with the mutated gene in addition to its normal gene copies (Figure 7–42B).

Because gene addition is much more easily accomplished than gene replacement in higher eucaryotic cells, it would be enormously useful to be able to create specific **dominant negative mutations** in which a mutant gene eliminates the activity of its normal counterparts in the cell. One ingenious and promising approach exploits the specificity of hybridization reactions between two complementary nucleic acid chains. Normally, only one of the two DNA strands in a given portion of double helix is transcribed into RNA, and it is always the same strand for a given gene. If a cloned gene is engineered so that the opposite DNA strand is transcribed instead, it will produce **antisense RNA molecules** that have a sequence complementary to the normal RNA transcripts. Antisense RNA, when synthesized in large enough amounts, will often hybridize with the "sense" RNA made by the normal genes and thereby inhibit the synthesis of the corresponding protein (Figure 7–43). A related method is to synthesize short antisense nucleic acid molecules by chemical or enzymatic means and then inject (or otherwise deliver) them into cells, again blocking (though only temporarily) production of the corresponding protein.

For unknown reasons the antisense RNA approach frequently fails to inactivate the desired gene. An alternative way of producing a dominant negative mutation takes advantage of the fact that most proteins function as part of a

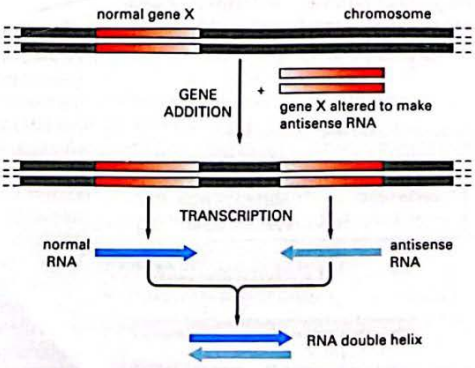


Figure 7–43 The antisense RNA strategy for generating dominant negative mutations. Mutant genes that have been engineered to produce antisense RNA, which is complementary in sequence to the RNA made by the normal gene X, can cause double-stranded RNA to form inside cells. If a large excess of the antisense RNA is produced, it can hybridize with—and thereby inactivate—most of the normal RNA produced by gene X. Although in the future it may become possible to inactivate any gene in this way, at present the technique seems to work for some genes but not others.

Sarepta Exhibit 1056, page 38

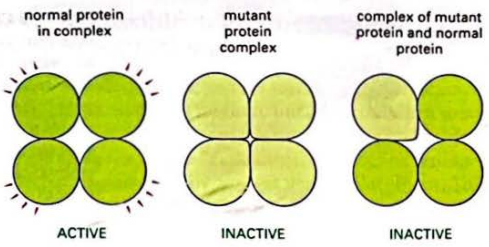


Figure 7–44 A dominant negative effect of a protein. Here a gene is engineered to produce a mutant protein that prevents the normal copies of the same protein from performing their function. In this simple example the normal protein must form a multi-subunit complex to be active, and the mutant protein blocks function by forming a mixed complex that is inactive. In this way a single copy of a mutant gene located anywhere in the genome can inactivate the normal products produced by other gene copies.

larger protein complex. Such complexes can be inactivated by the inclusion of just one nonfunctional component. Therefore, by designing a gene that produces large quantities of a mutant protein that is inactive but still able to assemble into the complex, it is often possible to produce a cell in which all the complexes are inactivated despite the presence of both normal and mutant copies of the protein (Figure 7–44).

If a protein is required for the survival of the cell (or the organism), a dominant negative mutant will die, making it impossible to test the function of the protein. To avoid this problem, one can couple the mutant gene to control sequences that have been engineered to produce the gene product only on command—for example, in response to an increase in temperature or to the presence of a specific signaling molecule. Cells or organisms containing such an *inducible* dominant mutant gene can be deprived of a specific protein at a particular time, and the effect can then be followed. In the future, techniques for producing dominant negative mutations to inactivate specific genes are likely to be widely used to determine the functions of proteins in higher organisms.

Engineered Genes Can Be Permanently Inserted into the Germ Line of Mice or Fruit Flies to Produce Transgenic Animals [37]

The ultimate test of the function of an altered gene is to reinsert it into an organism and see what effect it has. Ideally one would like to be able to replace the normal gene with the altered one so that the function of the mutant protein can be analyzed in the absence of the normal protein. As discussed above, this can be readily accomplished in some haploid organisms, but in higher eucaryotic cells an integrative event leading to a gene replacement occurs only very rarely. Foreign DNA can, however, rather easily be randomly integrated into the genome. In mammals, for example, linear DNA fragments introduced into cells are rapidly ligated end to end by intracellular enzymes to form long tandem arrays, which usually become integrated into a chromosome at an apparently random site. Fertilized mammalian eggs behave like other mammalian cells in this respect. A mouse egg injected with 200 copies of a linear DNA molecule will often develop into a mouse containing, in many of its cells, a tandem array of copies of the injected gene integrated at a single random site in one of its chromosomes (Figure 7–45). If the modified chromosome is present in the germ line cells (eggs or sperm), the mouse will pass these foreign genes on to its progeny. Animals that have been permanently altered in this way are called **transgenic organisms,** and the foreign genes are called **transgenes.** Because the normal gene generally remains present, only dominant effects of the alteration will show up. Nevertheless, such transgenic animals have already provided important insights into how mammalian genes are regulated and how certain altered genes (called oncogenes) cause cancer.

It is also possible to produce transgenic fruit flies, in which single copies of a gene are inserted at random into the *Drosophila* genome. The trick in this case is first to insert the DNA fragment between the two terminal sequences of a par-

Sarepta Exhibit 1056, page 39

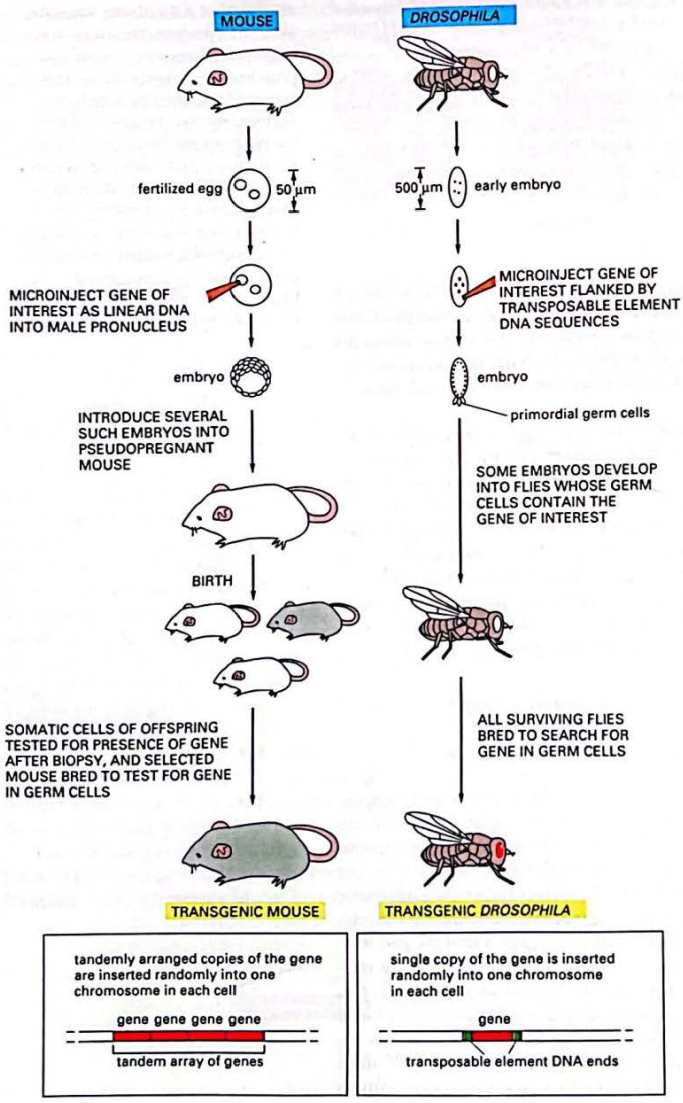


Figure 7–45 Comparison of the standard procedures used to make transgenic mice and transgenic *Drosophila*. In these examples the gene injected into the mouse egg causes a change in coat color, whereas the gene injected into the fly embryo causes a change in eye color. In both organisms some of the transgenic animals are found to have DNA insertions at more than one chromosomal site.

ticular *Drosophila* transposable element, called the *P element*. The terminal sequences enable the P element to integrate into *Drosophila* chromosomes if the P element transposase enzyme is also present (see p. 285). To make transgenic fruit flies, therefore, the appropriately modified DNA fragment is injected into a very young fruit fly embryo along with a separate plasmid containing the gene encoding the transposase. When this is done, the injected gene often enters the germ line in a single copy as the result of a transposition event (Figures 7–45 and 7–46).

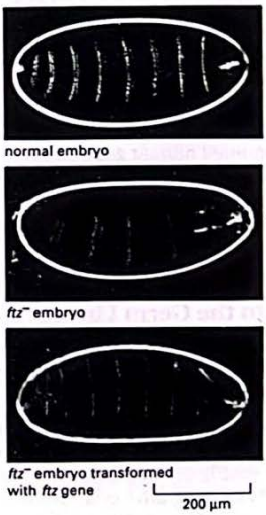


Figure 7–46 Genetic rescue. Comparison of a normal *Drosophila* larva and two mutant larvae that contain defective *ftz* genes. One of the defective (*ftz*⁻) larvae has been rescued by the injection of a DNA clone containing the normal *ftz* gene sequence into the egg of one of its ancestors. This added DNA sequence has become permanently integrated into one of the fly's chromosomes and is therefore faithfully inherited and expressed. The *ftz* gene is required for normal development, and the addition of this gene to the mutant genome restores the larval segments that are missing in the *ftz*⁻ organism. For the technique used to make transgenic animals, see Figure 7–45. (Courtesy of Walter Gehring.)

Sarepta Exhibit 1056, page 40

Gene Targeting Makes It Possible to Produce Transgenic Mice That Are Missing Specific Genes [38]

If a DNA molecule carrying a mutated mouse gene is transferred into a mouse cell, it usually inserts into the chromosomes at random, but about once in a thousand times, it will replace one of the two copies of the normal gene by homologous recombination. By exploiting these rare "gene targeting" events, any specific gene can be inactivated in a mouse cell by a direct gene replacement. This technique can be extended to produce a mouse with a missing gene by means of the following two-step pathway.

In the first step, a DNA fragment containing a desired mutant gene (or part of a gene) is inserted into a vector and then introduced into a special line of embryo-derived mouse stem cells (called *embryonic stem cells*, or *ES cells*) that grow in cell culture and are capable of producing cells of any tissue. After a period of cell proliferation, the rare colonies of cells in which a homologous recombination event is likely to have caused a gene replacement to occur are isolated by a double drug selection (Figure 7–47). The correct colonies among these are identified by PCR or by Southern blotting: they will contain recombinant DNA sequences in which the inserted fragment has replaced all or part of one copy of the normal gene. In the second step, individual cells from the identified colony are taken up into a fine micropipette and injected into an early mouse embryo. The transfected embryo-derived stem cells collaborate with the cells of the host embryo to produce a normal-looking mouse (see Figure 21–32); large parts of this chimeric animal, including in favorable cases cells of the germ line, will often derive from the artificially altered stem cells. Such mice are bred to produce both a male and a female animal, each heterozygous for the gene replacement (that is, they have one normal and one mutant copy of the gene). When these two mice are mated, one-fourth of their progeny will be homozygous for the altered gene. Studies of these homozygotes allow the function of the altered gene to be examined in the absence of the corresponding normal gene.

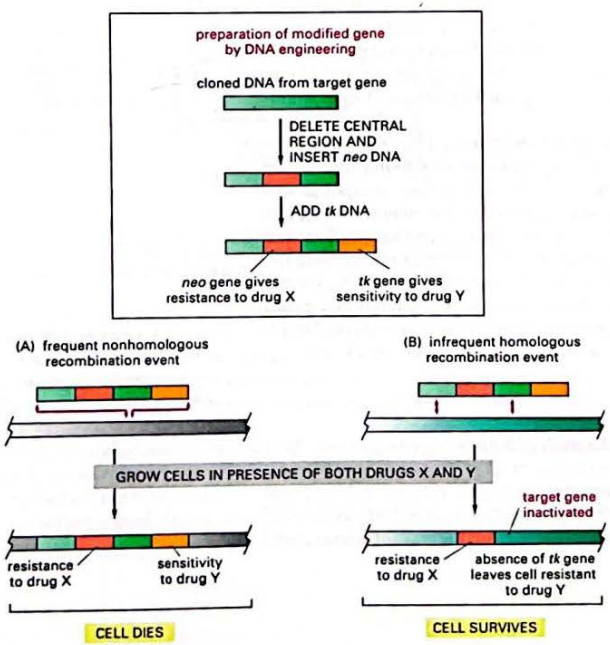


Figure 7–47 Selective gene knockout by homologous recombination in the mouse. Cloned DNA from the mouse gene to be mutated is modified by genetic engineering so that it contains a bacterial gene within it, called *neo*, whose integration into a mouse chromosome makes the mouse cells *resistant* to a drug that otherwise kills them (drug X). A viral gene, called *tk*, is also added, attached to one end of the mouse DNA; the integration of *tk* into a mouse chromosome makes the cells *sensitive* to a different drug (drug Y). Most insertions occur into random sites in the mouse chromosome, and these nearly always include both ends of the engineered DNA fragment, as shown (A). By selecting for those rare mouse cells that grow in the presence of *both* drugs, colonies of cells are obtained in which homologous recombination has incorporated the center of the engineered DNA fragment *without* the ends; most of these cells will turn out to carry the targeted gene replacement shown in (B).

Sarepta Exhibit 1056, page 41

The ability to prepare transgenic mice lacking a known normal gene has been a major advance, and the technique is now being widely used to dissect the functions of specific mammalian genes (Figure 7–48).

Transgenic Plants Are Important for Both Cell Biology and Agriculture [39]

When a plant is damaged, it can often repair itself by a process in which mature differentiated cells "dedifferentiate," proliferate, and then redifferentiate into other cell types. In some circumstances the dedifferentiated cells can even form an apical meristem, which can then give rise to an entire new plant, including gametes. This remarkable plasticity of plant cells can be exploited to generate transgenic plants from cells growing in culture.

When a piece of plant tissue is cultured in a sterile medium containing nutrients and appropriate growth regulators, many of the cells are stimulated to proliferate indefinitely in a disorganized manner, producing a mass of relatively undifferentiated cells called a *callus*. If the nutrients and growth regulators are carefully manipulated, one can induce the formation of a shoot and then root apical meristems within the callus, and, in many species, a whole new plant can be regenerated.

Callus cultures can also be mechanically dissociated into single cells, which will grow and divide as a suspension culture. In a number of plants—including tobacco, petunia, carrot, potato, and *Arabidopsis*—a single cell from such a suspension culture can be grown into a small clump (a clone) from which a whole plant can be regenerated. Just as mutant mice can be derived by genetic manipulation of embryonic stem cells in culture, so transgenic plants can be created from single plant cells transfected with DNA in culture (Figure 7–49).

The ability to produce transgenic plants has greatly accelerated progress in many areas of plant cell biology. It has played an important part, for example, in isolating receptors for growth regulators and in analyzing the mechanisms of morphogenesis and of gene expression in plants. It has also opened up many new possibilities in agriculture that could benefit both the farmer and the consumer. It has made it possible, for example, to modify the lipid, starch, and protein storage reserved in seeds, to impart pest and virus resistance to plants, and to create modified plants that tolerate extreme habitats such as salt marshes or waterlogged soil.

Many of the major advances in understanding animal development have come from studies on the fruit fly *Drosophila* and the nematode worm *Caenorhabditis elegans*, which are amenable to extensive genetic analysis as well as to experimental manipulation. Progress in plant developmental biology has been relatively slow by comparison. Many of the organisms that have proved most amenable to genetic analysis, such as maize and tomato, have long life cycles and very large genomes, which have made both classical and molecular genetic analysis time-consuming. Increasing attention is consequently being paid to a fast-growing small weed, the common wall cress (*Arabidopsis thaliana*), which has several major advantages as a "model plant" (see Figure 21–93).

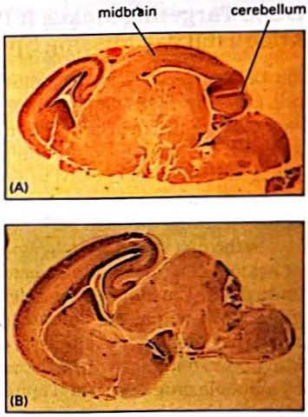
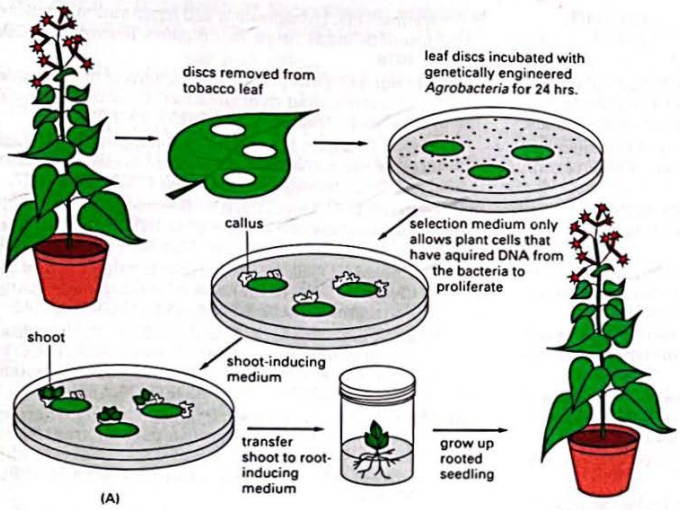


Figure 7–48 A transgenic mouse in which both copies of the *Wnt-1* growth factor gene have been eliminated by homologous recombination. (A) Section of the brain from a normal embryo. (B) Section of the brain from a mutant embryo, which lacks a cerebellum and most of the midbrain, and dies *in utero*. (Photographs courtesy of Mario Capecchi.)

Summary

Genetic engineering has revolutionized the study of both cells and organisms. DNA engineering techniques can be used to create any mutant gene and insert it into a cell's chromosomes so that it becomes a permanent part of the genome. If the cell used for this gene transfer is a fertilized egg (for an animal) or a totipotent plant cell in culture, transgenic organisms can be produced that express the mutant gene and will

Sarepta Exhibit 1056, page 42



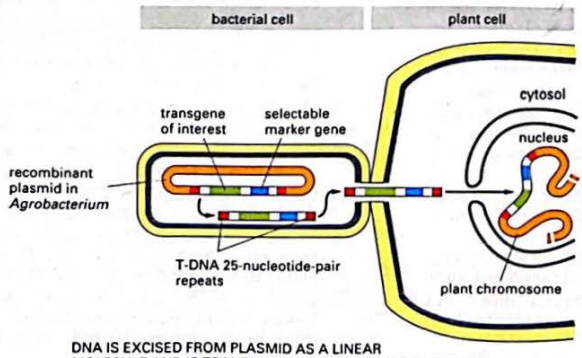


Figure 7–49 A procedure used to make a transgenic plant. (A) Outline of the process. A disc is cut out of a leaf and incubated in culture with *Agrobacteria* that carry a recombinant plasmid with both a selectable marker and a desired transgene. The wounded cells at the edge of the disc release substances that attract the *Agrobacteria* and cause them to inject DNA into these cells. Only those plant cells that take up the appropriate DNA and express the selectable marker gene survive to proliferate and form a callus. The manipulation of growth factors supplied to the callus induces it to form shoots that subsequently root and grow into adult plants carrying the transgene. (B) The preparation of the recombinant plasmid and its transfer to plant cells. An *Agrobacterium* plasmid that normally carries the T-DNA sequence is modified by substituting a selectable marker (such as the kanamycin-resistance gene) and a desired transgene between the 25-nucleotide-pair T-DNA repeats. When the *Agrobacterium* recognizes a plant cell, it efficiently passes a DNA strand that carries these sequences into the plant cell, using the special machinery that normally transfers the plasmid's T-DNA sequence.

pass it on to their progeny. Especially important for cell biology is the power that such technology gives the experimenter to alter cells and organisms in highly specific ways—allowing one to discern the effect on the cell or the organism of a designed change in a single protein that has been intentionally mutated by "reverse genetic" techniques.

The consequences of genetic engineering are far-reaching in other ways as well. Plants can be altered in ways that increase their value as food crops. Bacteria, yeasts, or mammalian cells can be engineered to synthesize any desired protein in large quantities, making it possible to analyze the structure and function of the protein in detail or to use the protein as a drug or a vaccine for medical purposes.

Sarepta Exhibit 1056, page 43

References

Cited

1. Sambrook, J.; Fritsch, E.F.; Maniatis, T. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1989.

   Watson, J.D.; Tooze, J. The DNA Story: A Documentary History of Gene Cloning. New York: W.H. Freeman, 1981.

   Watson, J.D.; Gilman, M.; Witkowski, J.; Zoller, M. Recombinant DNA, 2nd ed. New York: W.H. Freeman, 1992.

2. Kessler, C.; Manta, V. Specificity of restriction endonucleases and DNA modification methyltransferases—a review (Edition 3). *Gene* 92:1–248, 1990.

3. Danna, K.J. Determination of fragment order through partial digests and multiple enzyme digests. *Methods Enzymol.* 65:449–467, 1980.

   Nathans, D.; Smith, H.O. Restriction endonucleases in the analysis and restructuring of DNA molecules. *Annu. Rev. Biochem.* 44:273–293, 1975.

4. Andrews, A.T. Electrophoresis, 2nd ed. Oxford, UK: Clarendon Press, 1986.

   Evans, G.A. Physical mapping of the human genome by pulsed field gel analysis. *Curr. Opin. Genet. Dev.* 1:75–81, 1991.

   Southern, E.M. Gel electrophoresis of restriction fragments. *Methods Enzymol.* 68:152–176, 1979.

5. Lehman, I.R. DNA polymerase I of *Escherichia coli.* In The Enzymes, Vol. 14A (P.D. Boyer, ed.), pp. 16–38. New York: Academic Press, 1981.

   Richardson, C.C. Polynucleotide kinase from *Escherichia coli* infected with bacteriophage T4. *Nucleic Acids Res.* 2:815, 1971.

   Rigby, P.W.; Dieckmann, M.; Rhodes, C.; Berg, P. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* 113:237–251, 1977.

6. Griffin, H.G.; Griffin, A.M. DNA sequencing. Recent innovations and future trends. *Appl. Biochem. Biotech.* 38:147–159, 1993.

   Maxam, A.M.; Gilbert, W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* 74:560–564, 1977.

   Maxam, A.M.; Gilbert, W. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499–560, 1980.

   Prober, J.M.; Trainor, G.; Dam, R.; et al. A system for rapid DNA sequencing with fluorescent chain terminating dideoxynucleotides. *Science* 238:336–341, 1987.

   Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463–5467, 1977.

7. Cartwright I.L.; Kelly, S.E. Probing the nature of chromosomal DNA-protein contacts by *in vivo* footprinting. *Biotechniques* 11:188–203, 1991.

   Galas, D.J.; Schmitz, A. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* 5:3157–3170, 1978.

   Tullius, T.D. Physical studies of protein-DNA complexes by footprinting. *Annu. Rev. Biophys. Biophys. Chem.* 18:213–237, 1989.

8. Schildkraut, C.L.; Marmur, J.; Doty, P. The formation of hybrid DNA molecules and their use in studies of DNA homologies. *J. Mol. Biol.* 3:595–617, 1961.

   Wetmur, J.G. Hybridization and renaturation kinetics of nucleic acids. *Annu. Rev. Biophys. Bioeng.* 5:337–361, 1976.

   Wetmur, J.G. DNA probes: applications of the principles of nucleic acid hybridization. *Critical Reviews in Biochem. Mol. Biol.* 26:227–259, 1991.

9. Berk, A.J.; Sharp, P.A. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. *Cell* 12:721–732, 1977.

   Gerhard, D.S.; Kawasaki, E.S.; Bancroft, F.C.; Szabo, P. Localization of a unique gene by direct hybridization. *Proc. Natl. Acad. Sci. USA* 78:3755–3759, 1981.

   Pardue, M.L.; Gall, J.G. Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc. Natl. Acad. Sci. USA* 64:600–604, 1969.

   Wallace, R.B.; Shaffer, J.; Murphy, R.F.; et al. Hybridization of synthetic oligodeoxyribonucleotides to ϕX174 DNA: the effect of a single base pair mismatch. *Nucleic Acids Res.* 6:3543–3557, 1979.

10. Alwine J.C.; Kemp, D.J.; Stark, G.R. Method for detection of specific RNAs in agarose gels by transfer to diabenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. USA* 74:5350–5354, 1977.

    Southern, E.M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–517, 1975.

    Thomas, P.S. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* 77:5201–5205, 1980.

11. Chang, C; Meyerowitz, E.M. Plant genome studies: restriction fragment length polymorphism and chromosome mapping information. *Curr. Opin. Genet. Dev.* 1:112–118, 1991.

    Kidd, K.K. Progress towards completing the human linkage map. *Curr. Opin. Genet. Dev.* 1:99–104, 1991.

    Pourzand, C.; Cerutti, P. Genotypic mutation analysis by RFLP/PCR. *Mutat. Res.* 288:113–121, 1993.

12. Gilbert, F.; Marinduque, B. DNA prenatal diagnosis. *Curr. Opin. Obstet. Gynecol.* 2:226–235, 1990.

    Lathe, R. Synthetic oligonucleotide probes deduced from amino acid sequence data. Theoretical and practical considerations. *J. Mol. Biol.* 183:1–12, 1985.

13. McGinnis, W.; Garber, R.L.; Wirz, J.; et al. A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell* 37:403–408, 1984.

14. Stephenson, E.C.; Pokrywka, N.J. Localization of bicoid message during *Drosophila* oogenesis. *Curr. Top. Dev. Biol.* 26:23–34, 1992.

    Trask, B.J. Gene mapping by *in situ* hybridization. *Curr. Opin. Genet. Dev.* 1:82–87, 1991.

15. Ausubel, F.M.; Brent, R.; Kingston, R.E.; et al., eds. Current Protocols in Molecular Biology. New York: Wiley, 1993.

    Drlica, K. Understanding DNA and Gene Cloning. New York: Wiley, 1984.

16. Cohen, S.N. The manipulation of genes. *Sci. Am.* 233(1):24–33, 1975.

    Foster, T.J. Plasmid-determined resistance to antimicrobial drugs and toxic metal ions in bacteria. *Microbiol. Rev.* 47:361–409, 1983.

    Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166:557–580, 1983.

    Novick, R.P. Plasmids. *Sci. Am.* 243(6):102–127, 1980.

Sarepta Exhibit 1056, page 44

17. Lennon, G.G.; Lehrach, H. Hybridization analyses of arrayed cDNA libraries. *Trends Genet.* 7:314–317, 1991.

Maniatis, T., et al. The isolation of structural genes from libraries of eucaryotic DNA. *Cell* 15:687–701, 1978.

18. Okayama, H.; Berg, P. High-efficiency cloning of full-length cDNA. *Mol. Cell Biol.* 2:161–170, 1982.

Southern, E.M. Genome mapping: cDNA approaches. *Curr. Opin. Genet. Dev.* 2:412–416, 1992.

19. Calvet, J.P. Molecular approaches for analyzing differential gene expression: differential cDNA library construction and screening. *Pediat. Nephrol.* 5:751–757, 1991.

Hedrick, S.M.; Cohen, D.I.; Nielsen, E.A.; Davis, M.M. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* 308:149–153, 1984.

20. Yang, J.; Ye, J.; Wallace, D.C. Computer selection of oligonucleotide probes from amino acid sequences for use in gene library screening. *Nucleic Acids Res.* 12:837–843, 1984.

Young, R.A.; Davis, R.W. Efficient isolation of genes using antibody probes. *Proc. Natl. Acad. Sci. USA* 80:1194–1198, 1983.

21. Hope, I.A.; Struhl, K. GCN4 protein, synthesized *in vitro*, binds HIS3 regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast. *Cell* 43:177–188, 1985.

Ricciardi, R.P.; Miller, J.S.; Roberts, B.E. Purification and mapping of specific mRNAs by hybridization-selection and cell-free translation. *Proc. Natl. Acad. Sci. USA* 76:4927–4931, 1979.

22. Stubbs, L. Long-range walking techniques in positional cloning strategies. *Mammalian Genome* 3:127–142, 1992.

23. Burke, D.T. The role of yeast artificial chromosomes in generating genome maps. *Curr. Opin. Genet. Dev.* 1:69–74, 1991.

Carrano, A.V.; de Jong, P.J.; Branscomb, E.; et al. Constructing chromosome- and region-specific cosmid maps of the human genome. *Genome* 31:1059–1065, 1989.

Coulson, A.; Kozono, Y.; Lutterbach, B.; et al. YACs and the *C. elegans* genome. *Bioessays* 13:413–417, 1991.

Hauge, B.M.; Hanley, S.; Giraudat, J.; et al. Mapping the *Arabidopsis* genome. *Symp. Soc. Exper. Biol.* 45:45–56, 1991.

24. Harris, A.; Argent, B.E. The cystic fibrosis gene and its product CFTR. *Semin. Cell Biol.* 4:37–44, 1993.

Hyser, C.L. Unraveling the mysteries of Duchenne and Becker muscular dystrophy. *Mol. Chem. Neuropathol.* 10:15–20, 1989.

Wicking, C.; Williamson, B. From linked marker to gene. *Trends Genet.* 7:288–293, 1991.

25. Allen, R.W.; Wallhermfechtel, M.; Miller, W.V. The application of restriction fragment length polymorphism mapping to parentage testing. *Transfusion* 30:552–564, 1990.

Bottema, C.D.; Sommer, S.S. PCR amplification of specific alleles: rapid detection of known mutations and polymorphisms. *Mutat. Res.* 288:93–102, 1993.

Nelson, D.L. Applications of polymerase chain reaction methods in genome mapping. *Curr. Opin. Genet. Dev.* 1:62–68, 1991.

Pourzand, C.; Cerutti, P. Genotypic mutation analysis by RFLP/PCR. *Mutat. Res.* 288:113–121, 1993.

Saiki, R.K.; Gelfand, D.H.; Stoffel, S.; et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487–491, 1988.

26. Olson, M.V. The human genome project. *Proc. Natl. Acad. Sci. USA* 90:4338–4344, 1993.

Szostak, J.W. *In vitro* genetics. *Trends Biochem. Sci.* 17:89–93, 1992.

27. Itakura, K.; Rossi, J.J.; Wallace, R.B. Synthesis and use of synthetic oligonucleotides. *Annu. Rev. Biochem.* 53:323–356, 1984.

28. Melton, D.A.; Krieg, P.A.; Rebagliati, M.R.; et al. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* 12:7035–7056, 1984.

29. Abelson, J.; Butz, E., eds. Recombinant DNA. *Science* 209:1317–1438, 1980.

Balbas, P.; Bolivar, F. Design and construction of expression plasmid vectors in *E. coli*. *Methods Enzymol.* 185:14–37, 1990.

Gilbert, W.; Villa-Komaroff, L. Useful proteins from recombinant bacteria. *Sci. Am.* 242(4):74–94, 1980.

Miller, L.K. Baculoviruses: high-level expression in insect cells. *Curr. Opin. Genet. Dev.* 3:97–101, 1993.

Rose, A.B.; Broach, J.R. Propagation and expression of cloned genes in yeast: 2-micron circle-based vectors. *Methods Enzymol.* 185:234–279, 1990.

30. Alam, J.; Cook, J.L. Reporter genes: application to the study of mammalian gene transcription. *Anal. Biochem.* 188:245–254, 1990.

Bellen, H.J.; Wilson, C.; Gehring, W.J. Dissecting the complexity of the nervous system by enhancer detection. *Bioessays* 12:199–204, 1990.

31. Jockusch, B.M.; Zurek, B.; Zahn, R.; Westmeyer, A.; Fuchtbauer, A. Antibodies against vertebrate microfilament proteins in the analysis of cellular motility and adhesion. *J. Cell Sci.* S14:41–47, 1991.

Lederberg, J.; Lederberg, E.M. Replica plating and indirect selection of bacterial mutants. *J. Bacteriol.* 63:399–406, 1952.

Nusslein-Volhard, C.; Weischaus, E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287:795–801, 1980.

32. Kaiser, K. From gene to phenotype in *Drosophila* and other organisms. *Bioessays* 12:297–301, 1990.

Landel, C.P.; Chen, S.; Evans, G.A. Reverse genetics using transgenic mice. *Annu. Rev. Physiol.* 52:841–851, 1990.

Provost, G.S.; Kretz, P.L.; Hamner, R.T.; et al. Transgenic systems for *in vivo* mutation analysis. *Mutat. Res.* 288:133–149, 1993.

33. Baase, W.A.; Eriksson, A.E.; Zhang, X.J.; et al. Dissection of protein structure and folding by directed mutagenesis. *Faraday Discuss.* 93:173–181, 1992.

Sharon, J.; Kao C.Y.; Sompuram, S.R. Oligonucleotide-directed mutagenesis of antibody combining sites. *Int. Rev. Immunol.* 10:113–127, 1993.

34. Garoff, H. Using recombinant DNA techniques to study protein targeting in the eucaryotic cell. *Annu. Rev. Cell Biol.* 1:403–445, 1985.

Sarepta Exhibit 1056, page 45

Kelly, J.H.; Darlington, G.J. Hybrid genes: molecular approaches to tissue-specific gene regulation. *Annu. Rev. Genet.* 19:273–296, 1985.

35. Leclerc, D.; Brakier-Gingras, L. Study of the function of *Escherichia coli* ribosomal RNA through site-directed mutagenesis. *Biochem. Cell Biol.* 68:169–179, 1990.

Scherer, S.; Davis, R.W. Replacement of chromosome segments with altered DNA segments constructed *in vitro*. *Proc. Natl. Acad. Sci. USA* 76:4951–4955, 1979.

Struhl, K. The new yeast genetics. *Nature* 305:391–397, 1983.

36. Herskowitz, I. Functional inactivation of genes by dominant negative mutations. *Nature* 329:219–222, 1987.

Weintraub, H.; Izant, J.G.; Harland, R.M. Anti-sense RNA as a molecular tool for genetic analysis. *Trends Genet.* 1:22–25, 1985.

37. Boyd, A.L.; Samid, D. Review: molecular biology of transgenic animals. *J. Anim. Sci.* 71(S3):1–9, 1993.

Palmiter, R.D.; Brinster, R.L. Germ line transformation of mice. *Annu. Rev. Genet.* 20:465–499, 1986.

Rubin, G.M.; Sprading, A.C. Genetic transformation of *Drosophila* wth transposable element vectors. *Science* 218:348–353, 1982.

Smith, P.A.; Corces, V.G. *Drosophila* transposable elements: mechanisms of mutagenesis and interactions with the host genome. *Adv. Genet.* 29:229–300, 1991.

38. Babinet, C.; Morello, D.; Renard, J.P. Transgenic mice. *Genome* 31:938–949, 1989.

Gossen, J.; Vijg, J. Transgenic mice as model systems for studying gene mutations *in vivo*. *Trends Genet.* 9:27–31, 1993.

Gridley, T. Insertional versus targeted mutagenesis in mice. *New Biol.* 3:1025–1034, 1991.

39. Davey, M.R.; Rech, E.L.; Mulligan, B.J. Direct DNA transfer to plant cells. *Plant Mol. Biol.* 133:273–285, 1989.

Walden, R.; Schell, J. Techniques in plant molecular biology—progress and problems. *Eur. J. Biochem.* 192:563–576, 1990.

Sarepta Exhibit 1056, page 46

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

GARLAND STAFF
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.

© 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 3 2

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.)

The Cell Nucleus

8

- Chromosomal DNA and Its Packaging
- The Global Structure of Chromosomes
- Chromosome Replication
- RNA Synthesis and RNA Processing
- The Organization and Evolution of the Nuclear Genome

The DNA in a eucaryotic cell is sequestered in the nucleus, which occupies about 10% of the total cell volume. The nucleus is delimited by a *nuclear envelope* formed by two concentric membranes. These membranes are punctured at intervals by nuclear pores, which actively transport selected molecules to and from the cytosol. The envelope is directly connected to the extensive membranes of the endoplasmic reticulum, and it is supported by two networks of intermediate filaments: one called the *nuclear lamina* forms a thin shell just inside the nucleus underlying the inner nuclear membrane, while the other, less regularly organized, surrounds the outer nuclear membrane (Figure 8–1).

Like modern procaryotes, the ancestors of eucaryotic cells almost certainly lacked a nucleus, and one can only speculate on why a separate nuclear compartment evolved, segregating the DNA from the activities in the cytoplasm. Two special features of eucaryotic cells suggest possible reasons. One is the eucaryotic cytoskeleton, which is composed mainly of microtubules and actin filaments and mediates cell movements. Bacteria, whose DNA is in direct contact with the cytoplasm, lack a cytoskeleton and move by means of external structures such as flagella. One function of the nuclear envelope may therefore be to protect the long, fragile DNA molecules from the mechanical forces generated by the cytoplasmic filaments in eucaryotes.

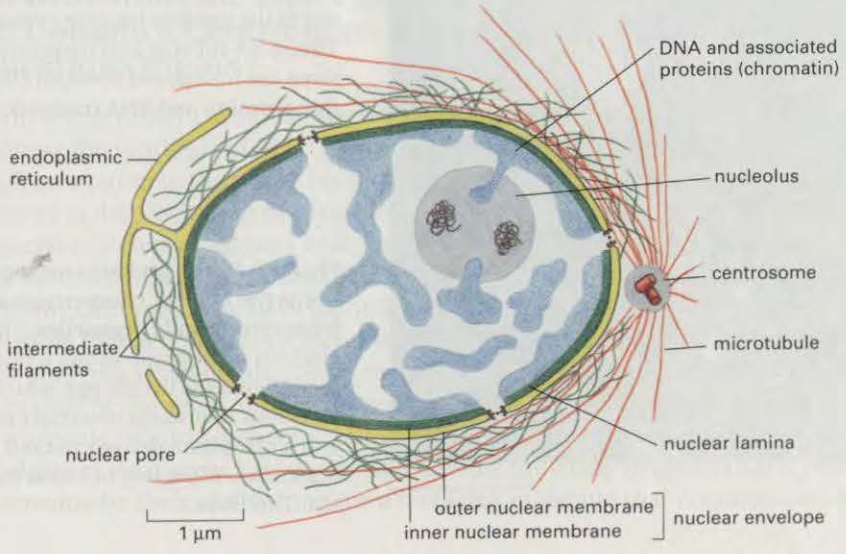


Figure 8–1 Cross-section of a typical cell nucleus. The nuclear envelope consists of two membranes, the outer one being continuous with the endoplasmic reticulum membrane (see also Figure 12–9). The lipid bilayers of the inner and outer nuclear membranes are connected at each nuclear pore. Two networks of intermediate filaments (*green*) provide mechanical support for the nuclear envelope; the intermediate filaments inside the nucleus form a sheetlike *nuclear lamina*. The space inside the endoplasmic reticulum (the ER lumen) is colored *yellow*.

Sarepta Exhibit 1056, page 49

A second special feature of eucaryotic cells is the extensive processing that RNA molecules undergo before they are translated into protein. In procaryotic cells RNA synthesis (*transcription*) and protein synthesis (*translation*) occur concurrently: ribosomes translate the 5′ end of an RNA molecule into protein while the 3′ end of the RNA molecule is still being synthesized. Consequently, there is relatively little opportunity to alter the RNA transcripts before they are translated into protein. In eucaryotes, by contrast, transcription (in the nucleus) is separated both temporally and spatially from translation (in the cytoplasm). The RNA transcripts in the nucleus are immediately packaged into ribonucleo-protein complexes and subjected to *RNA splicing*, in which certain portions of the nucleotide sequence are removed. Only when splicing is complete are the packaging proteins removed and the RNA molecules transported out of the nucleus to the cytosol, where ribosomes begin translating the RNA into protein (Figure 8–2). As we discuss later, RNA splicing is an important intermediate step in the transfer of genetic information in eucaryotes. It provides a number of advantages for the cell, including the potential for a single gene to make several different proteins. This may help explain why eucaryotic cells have a nucleus, where splicing can occur without interference from ribosomes (Figure 8–3).

In this chapter we describe how proteins package eucaryotic DNA into chromosomes, how chromosomes are folded and organized in the nucleus, and how they are replicated during each cell-division cycle. We then discuss RNA synthesis and RNA splicing, and, finally, we describe how genetic information is organized in the eucaryotic genome and how this organization may have arisen during evolution. The nuclear envelope is discussed in detail in Chapter 12 in connection with the selective transport of macromolecules into and out of the nucleus, and a hypothetical scheme of how the nuclear compartment might have evolved is presented there (see Figure 12–5).

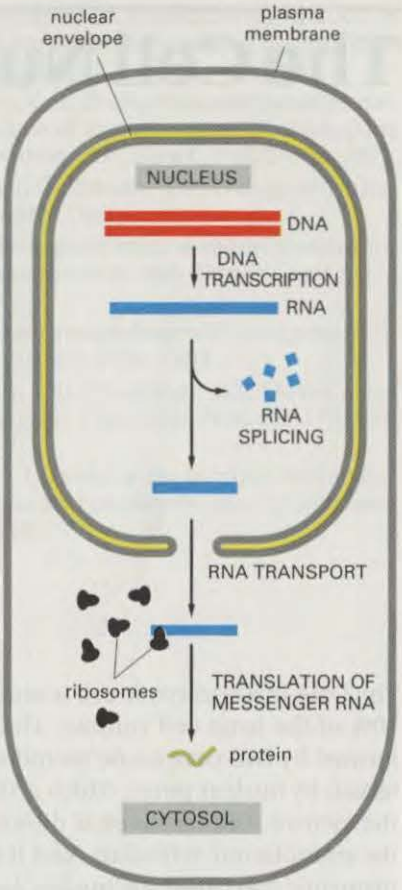


Figure 8–2 Protein synthesis (DNA → RNA → protein) in eucaryotes. Eucaryotic cells have evolved numerous membrane-bounded compartments that segregate their various chemical reactions so as to make them more efficient, and the nucleus is one such compartment. The nuclear envelope keeps functional ribosomes out of the nucleus, preventing RNA transcripts from being translated into protein until they have been extensively processed (spliced) and transported out of the nucleus into the cytosol. Thus RNA splicing and transport steps are interposed between DNA transcription and RNA translation.

Chromosomal DNA and Its Packaging [1]

For the first 40 years of this century, biologists tended to dismiss the possibility that DNA could carry the genetic information in chromosomes, partly because

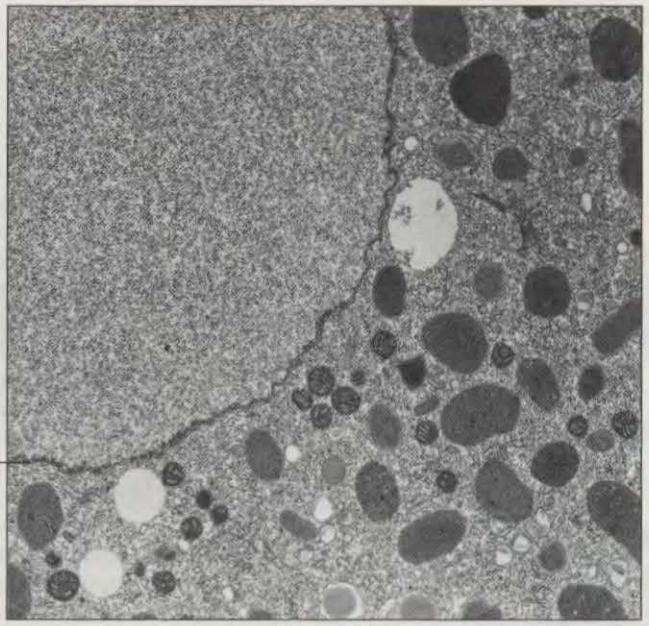


Figure 8–3 The nuclear envelope keeps the nuclear compartment free from cytoplasmic organelles. This electron micrograph shows a thin section of a sea urchin egg, which has a nucleus that stains unusually evenly and a cytoplasm densely packed with organelles. (Courtesy of David Begg and Tim Hunt.)

Sarepta Exhibit 1056, page 50

nucleic acids were erroneously believed to contain only a simple repeating tetranucleotide 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 nm^3 (10^{-15} cm^3).

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 μm 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

Sarepta Exhibit 1056, page 51

them to replicate when free of the host chromosome. A second sequence element, called a **centromere,** attaches any DNA molecule that contains it to the mitotic spindle during cell division. Each yeast chromosome contains a single centromere; when this sequence is inserted into a plasmid, it guarantees that each daughter cell will receive one of the two copies of the newly replicated plasmid DNA molecule when the yeast cell divides.

The third required sequence element is a **telomere,** which is needed at each end of a linear chromosome. If a circular plasmid that contains a replication origin and a centromere is broken at a single site to create two free ends in the double helix, it will still replicate and attach to the mitotic spindle, but it will eventually be lost from the progeny cells. This is because replication on the lagging strand of a replication fork requires the presence of some DNA ahead of the sequence to be copied to serve as the template for an RNA primer (see Figure 6–44). Since there can never be such a template for the last few nucleotides of a linear DNA molecule, special mechanisms are required to prevent each such DNA strand from becoming shorter with each replication cycle. Bacteria and many viruses solve this "end-replication problem" by having a circular DNA molecule as their chromosome. Eucaryotic cells have instead evolved a specialized telomeric DNA sequence at each chromosome end. This simple repeating sequence is periodically extended by an enzyme, *telomerase,* thus compensating for the loss of a few nucleotides of telomeric DNA in each cycle and permitting a linear chromosome to be completely replicated.

Figure 8–4 summarizes the functions of the three DNA sequence elements that are required for a linear chromosome to propagate itself from generation to generation in a yeast cell. These sequence elements are relatively short (typically less than 1000 base pairs each) and therefore utilize only a tiny fraction of the information-carrying capacity of a chromosome. The same three types of sequence elements are thought to operate in human chromosomes, but to date only the human telomere sequences have been well defined. Although the yeast versions of these sequences do not function in higher eucaryotic cells, recombinant DNA methods allow the yeast sequence elements to be added to human DNA molecules, which can then replicate in yeast cells as **artificial chromosomes.** In this way yeast cells can be used to prepare human genomic DNA libraries (see p. 315) in which each DNA clone (propagated as an artificial chromosome) contains as many as a million nucleotide pairs of human DNA sequence (Figure 8–5).

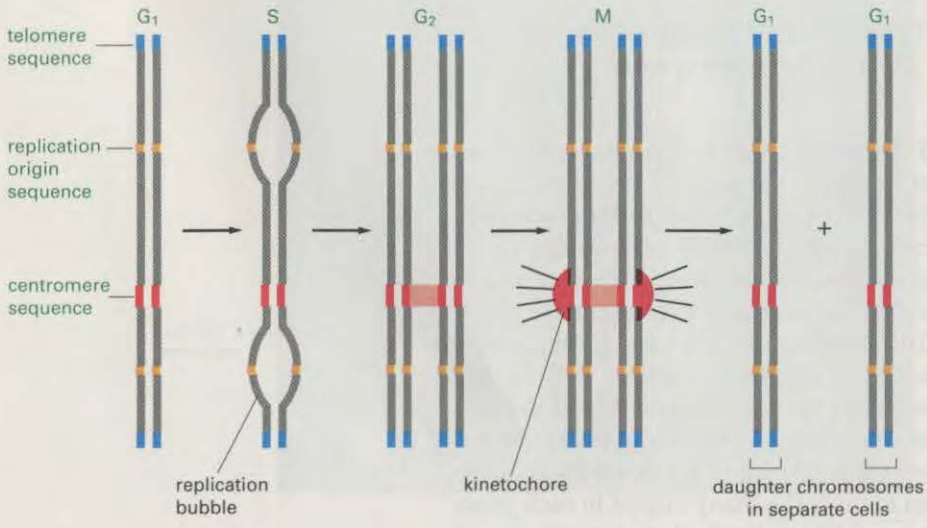


Figure 8–4 The functions of the three DNA sequence elements needed to produce a stable linear eucaryotic chromosome. Each chromosome has many origins of replication, one centromere, and two telomeres. The centromere serves to hold the two copies of the duplicated chromosome together and to attach them, via a protein complex called a kinetochore, to the mitotic spindle in such a way that one copy is distributed to each daughter cell at mitosis. The phases of the cell-division cycle corresponding to the events of chromosome replication and segregation are shown above the diagrams.

Sarepta Exhibit 1056, page 52

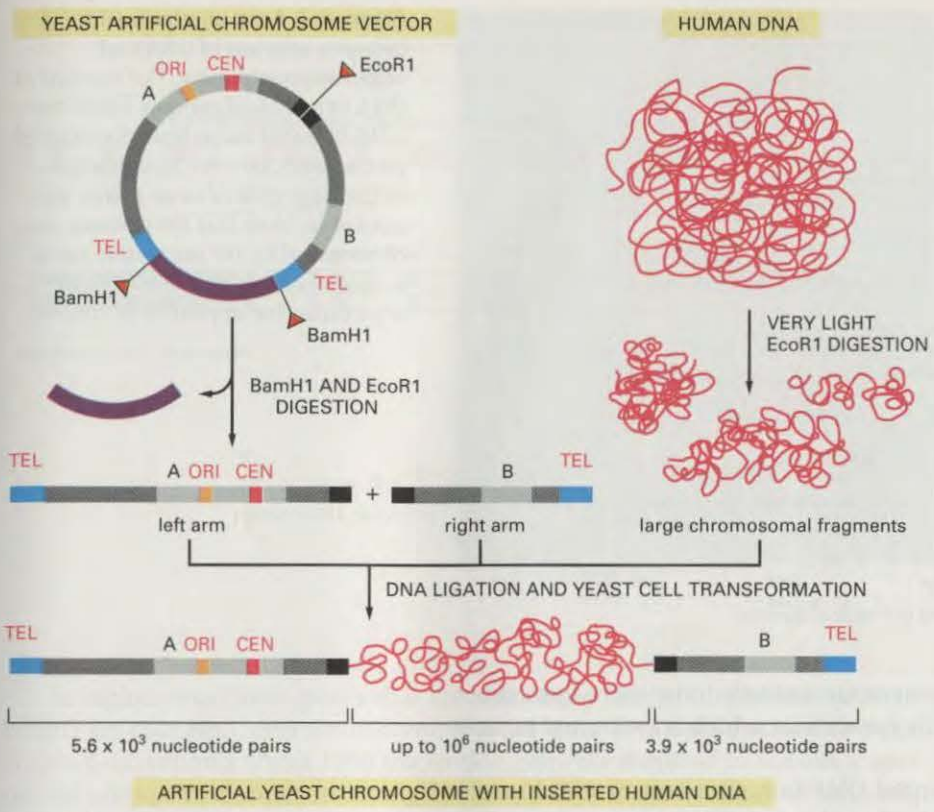


Figure 8–5 The making of a yeast artificial chromosome (YAC). A YAC vector allows the cloning of very large DNA molecules. TEL, CEN, and ORI are the telomere, centromere, and replication origin sequence elements, respectively, for the yeast *Saccharomyces cerevisiae*. BamH1 and EcoR1 are sites where 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 RNAs [3]

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.

Sarepta Exhibit 1056, page 53

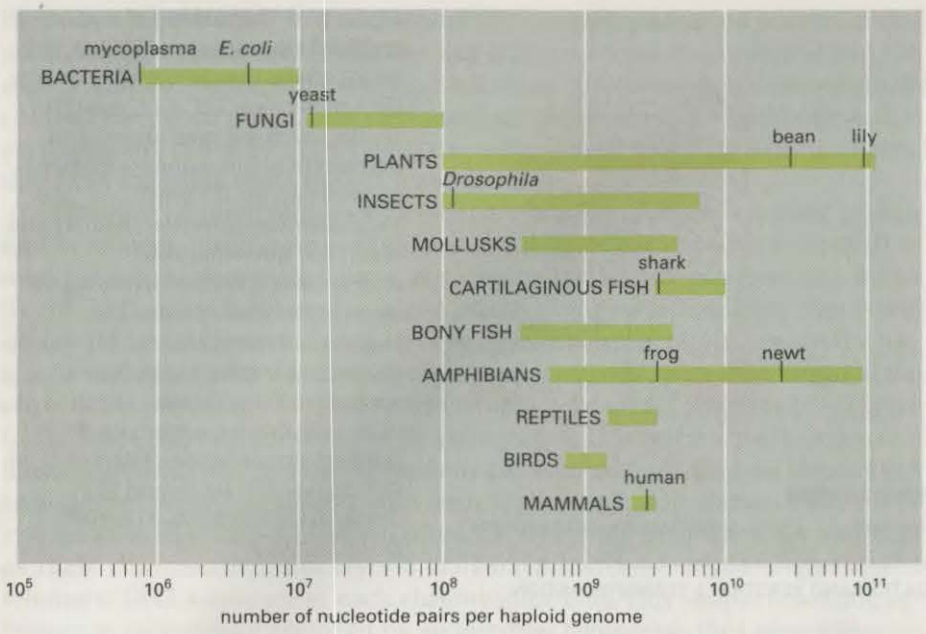


Figure 8–6 Lack of relationship between amount of DNA and organism complexity. The amount of DNA in a haploid genome varies over a 100,000-fold range from the smallest procaryotic cell—the mycoplasma—to the large cells of some plants and amphibia. Note that the genome size of humans (3×10^9 nucleotide pairs) is much smaller than that of many organisms that appear to be simpler.

Thus, from a genetic point of view, humans are unlikely to be more than about 10 times more complex than the fruit fly *Drosophila*, which is estimated to have about 5000 essential genes.

Whatever the remaining nonessential DNA in higher eucaryotic chromosomes may do (we discuss this later), the data shown in Figure 8–6 make it clear that it is not a great handicap for a higher eucaryotic cell to carry a large amount of extra DNA. Indeed, even the essential coding regions are often interrupted by long stretches of noncoding DNA.

Each Gene Produces an RNA Molecule [4]

The primary function of the genome is to specify RNA molecules. Selected portions of the DNA nucleotide sequence are copied into a corresponding RNA nucleotide sequence, which either encodes a protein (if it is an mRNA) or forms a "structural" RNA, such as a transfer RNA (tRNA) or ribosomal RNA (rRNA) molecule. Each region of the DNA helix that produces a functional RNA molecule constitutes a **gene.**

Table 8–1 The Size of Some Human Genes in Thousands of Nucleotides

	Gene Size	mRNA Size	Number of Introns
β-Globin	1.5	0.6	2
Insulin	1.7	0.4	2
Protein kinase C	11	1.4	7
Albumin	25	2.1	14
Catalase	34	1.6	12
LDL receptor	45	5.5	17
Factor VIII	186	9	25
Thyroglobulin	300	8.7	36
Dystrophin*	more than 2000	17	more than 50

*An altered form of this gene causes Duchenne muscular dystrophy.
The size specified here for a gene includes both its transcribed portion and nearby regulatory DNA sequences. (Compiled from data supplied by Victor McKusick.)

Sarepta Exhibit 1056, page 54

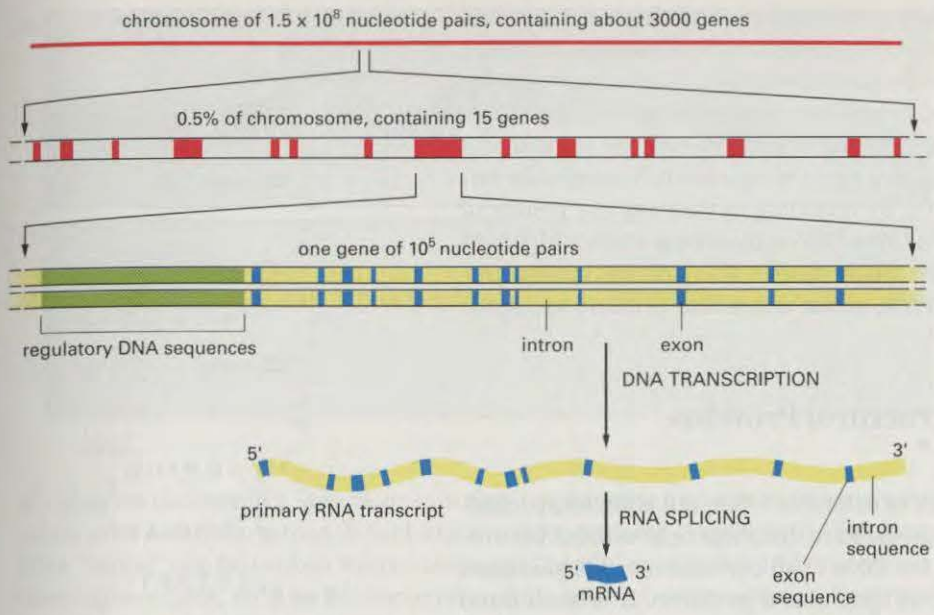


Figure 8–7 The organization of genes on a typical vertebrate chromosome. Proteins that bind to the DNA in regulatory regions determine whether a gene is transcribed; although often located on the 5′ side of a gene, as shown here, regulatory regions can also be located in introns, in exons, or on the 3′ side of a gene. Intron sequences are removed from primary RNA transcripts to produce messenger RNA (mRNA) molecules. The figure given here for the number of genes per chromosome is a minimal estimate.

In higher eucaryotes genes that are more than 100,000 nucleotide pairs in length are common, and some contain more than 2 million nucleotide pairs (Table 8–1); yet only about 1000 nucleotide pairs are required to encode a protein of average size (one containing 300 to 400 amino acids). Most of the extra length consists of long stretches of noncoding DNA that interrupt the relatively short segments of coding DNA. The coding sequences are called **exons**; the intervening (noncoding) sequences are called **introns.** The RNA molecule (called a *primary RNA transcript*) synthesized from such a gene is altered to remove the intron sequences during its conversion to an mRNA molecule (see Figure 8–2) in the process of *RNA splicing*, as we discuss later.

Large genes consist of a long string of alternating exons and introns, with most of the gene consisting of introns. In addition, each gene is associated with *regulatory DNA sequences*, which are responsible for ensuring that the gene is transcribed at the proper time and in the appropriate cell type. We discuss in Chapter 9 how these regulatory sequences work. Many of them are located "upstream" (on the 5′ side) of the site where RNA transcription begins, but they can also be located "downstream" (on the 3′ side) of the site where RNA transcription ends, or even in introns or exons. A typical vertebrate chromosome is illustrated schematically in Figure 8–7, along with one of its many genes.

Comparisons Between the DNAs of Related Organisms Distinguish Conserved and Nonconserved Regions of DNA Sequence [5]

Technical improvements in DNA sequencing are expected to allow the routine sequencing of stretches of chromosomal DNA that are millions of nucleotide pairs long, so that in the foreseeable future the sequence of all 3×10^9 nucleotides of the human genome will be determined. As more than 90% of this sequence is probably unimportant, it will be crucial to have some way of identifying the small proportion of sequence that is important. One approach to this problem is based on the observation that important sequences are conserved during evolution, while unimportant ones are free to mutate randomly. The strategy, therefore, is to compare the human sequence with that of the corresponding regions of a related genome, such as that of the mouse. Humans and mice are thought to have diverged from a common mammalian ancestor about 80×10^6 years ago, which is long enough for roughly two out of every three nucleotides to have been changed by random mutational events. Consequently, the only regions that will

Sarepta Exhibit 1056, page 55

have remained closely similar in the two genomes are those where mutations would impair function and put animals carrying them at a disadvantage, resulting in their elimination from the population by natural selection. Such closely similar regions are known as *conserved regions*. In general, *nonconserved* regions represent noncoding DNA—both between genes and in introns—whose sequence is not critical for function, whereas conserved regions represent functionally important exons and regulatory sequences. By revealing in this way the results of a very long natural "experiment," comparative DNA sequencing studies highlight the most interesting regions in genomes. Such studies also provide strong support for the conclusion that only about 10% of the vertebrate genome sequence is vitally important to the organism.

Histones Are the Principal Structural Proteins of Eucaryotic Chromosomes [6]

If chromosomes were composed simply of extended DNA, it is difficult to imagine how they could be replicated and segregated to daughter cells without becoming severely tangled or broken. In fact, the DNA of all chromosomes is packaged into a compact structure with the aid of specialized proteins. It is traditional to divide the DNA-binding proteins in eucaryotes into two general classes: the **histones** and the *nonhistone chromosomal proteins*. The complex of both classes of proteins with the nuclear DNA of eucaryotic cells is known as **chromatin.** Histones are unique to eucaryotes. They are present in such enormous quantities (about 60 million molecules of each type per cell, compared to 10,000 molecules per cell for a typical sequence-specific DNA-binding protein) that their total mass in chromatin is about equal to that of the DNA.

Histones are relatively small proteins with a very high proportion of positively charged amino acids (lysine and arginine); the positive charge helps the histones bind tightly to DNA (which is highly negatively charged), regardless of its nucleotide sequence. Histones probably only rarely dissociate from the DNA, and so they are likely to have an influence on any reaction that occurs on chromosomes.

The five types of histones fall into two main groups—the *nucleosomal histones* and the *H1 histones*. The **nucleosomal histones** are small proteins (102–135 amino acids) responsible for coiling the DNA into *nucleosomes*, as discussed later. These four histones are designated **H2A, H2B, H3,** and **H4.** H3 and H4 are among the most highly conserved of all known proteins (Figure 8–8). This evolutionary conservation suggests that their functions involve nearly all of their amino acids, so that a change in any position is deleterious to the cell. This suggestion has been tested in yeasts, where it is possible to mutate a given histone gene *in vitro* and introduce it into the yeast genome in place of the normal gene. As predicted, many mutations are found to be lethal; some that are not lethal cause changes in the normal pattern of gene expression (discussed in Chapter 9). The H1 histones are larger (containing about 220 amino acids) and have been less conserved during evolution than the nucleosomal histones.

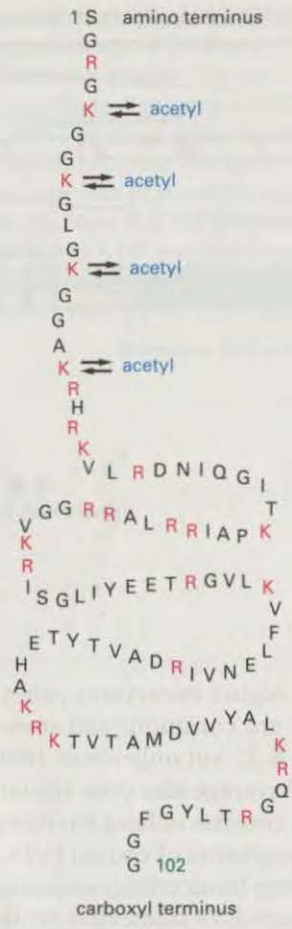


Figure 8–8 The amino acid sequence of histone H4. The amino acids are designated by their single-letter abbreviations, with the positively charged amino acids colored for emphasis. As in the three other nucleosomal histones, an elongated amino-terminal "tail" is reversibly modified in the cell by the acetylation of selected lysines, which removes the lysine's positive charge. The sequence from a cow is shown; the sequence is the same in peas except that one valine is changed to an isoleucine and one lysine is changed to an arginine.

Histones Associate with DNA to Form Nucleosomes, the Unit Particles of Chromatin [7]

If it were stretched out, the DNA double helix in each human chromosome would span the cell nucleus thousands of times. Histones play a crucial part in packing this very long DNA molecule in an orderly way into a nucleus only a few micrometers in diameter. Their role in DNA folding is also important for a second reason. As we shall see, not all the DNA is folded in exactly the same way, and the manner in which a region of the genome is packaged into chromatin in a particular cell seems to influence the activity of the genes the region contains.

A major advance in our understanding of chromatin structure came in 1974 with the discovery of the fundamental packing unit known as the **nucleosome,**

Sarepta Exhibit 1056, page 56

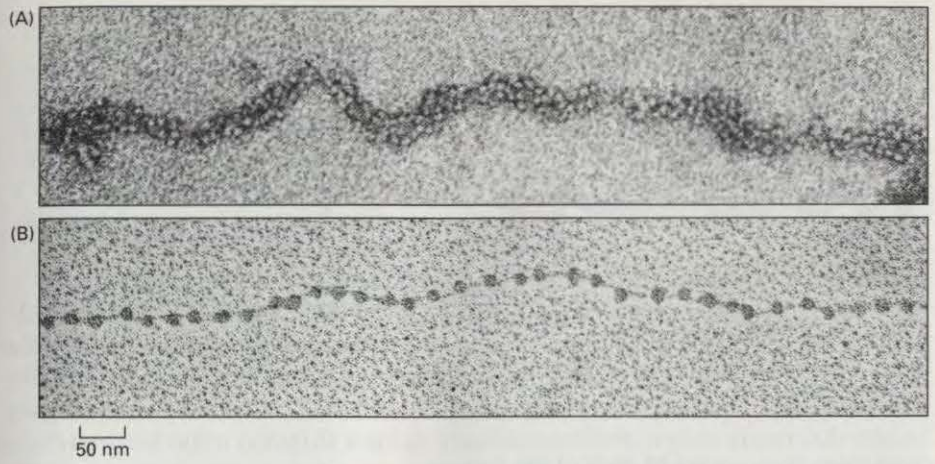


Figure 8–9 Nucleosomes as seen in the electron microscope. These electron micrographs show chromatin strands before and after treatments that unpack, or "decondense," the native structure to produce the "beads-on-a-string" form. The native structure, known as the 30-nm fiber (discussed later), is shown in (A). The decondensed, "beads-on-a-string" form of chromatin is shown at the same magnification in (B). For a schematic drawing of the relation between these two chromatin forms, see Figure 8–30. These electron micrographs were taken by modifications of the procedure outlined in Figure 8–45. (A, courtesy of Barbara Hamkalo; B, courtesy of Victoria Foe.)

which gives chromatin a "beads-on-a-string" appearance in electron micrographs taken after treatments that unfold higher-order packing (Figure 8–9). The long DNA "string" can be broken into nucleosome "beads" by digestion with enzymes that degrade DNA, such as the bacterial enzyme micrococcal nuclease. (Enzymes that degrade both DNA and RNA are called *nucleases;* enzymes that degrade only DNA are *deoxyribonucleases,* or *DNases.*) After digestion for a short period with micrococcal nuclease, only the DNA between the nucleosome beads is degraded. The rest is protected from digestion and remains as double-helical DNA fragments 146 nucleotide pairs long bound to a specific complex of eight nucleosomal histones (the *histone octamer*). The nucleosome beads obtained in this way have been crystallized and analyzed by x-ray diffraction. Each is a disc-shaped

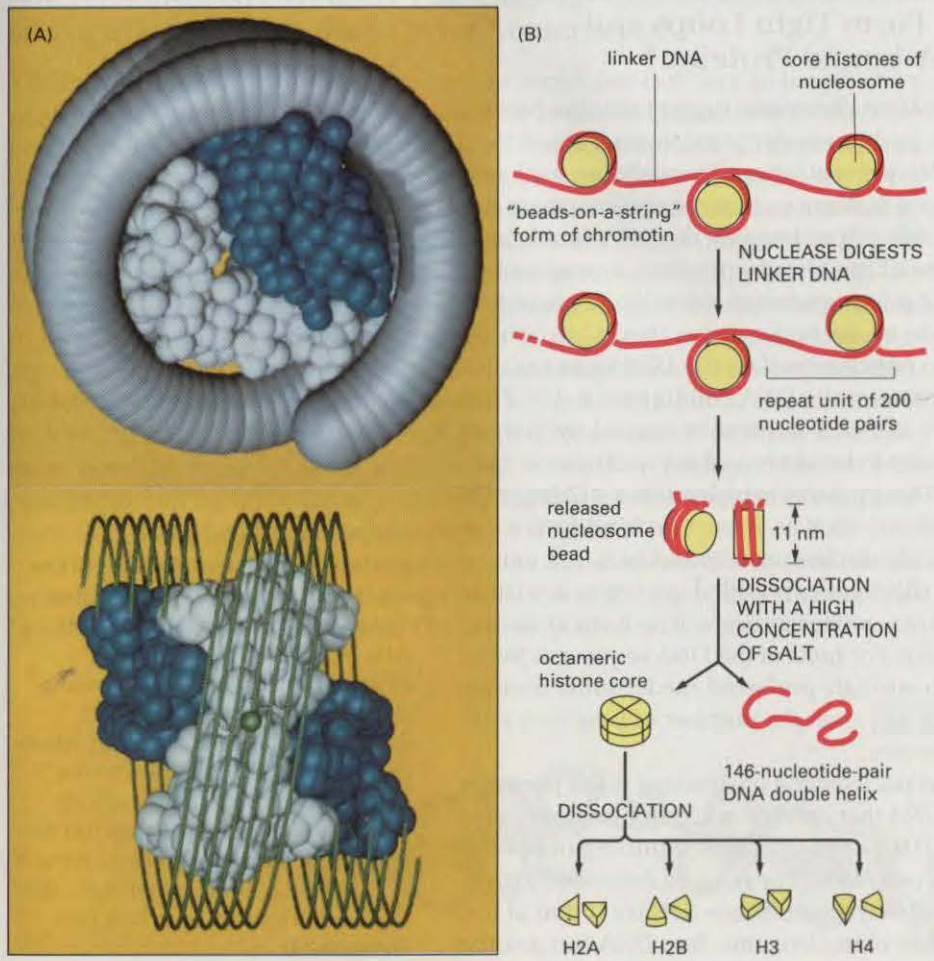


Figure 8–10 The nature of the nucleosome. (A) depicts two views of the three-dimensional structure of the histone octamer; the general path of the DNA wrapped around it is indicated by a coiled tube (*top*) and a series of parallel lines (*bottom*). Two H2A-H2B dimers (*blue*) flank an H3-H4 tetramer. The histone octamer is thus composed of two each of histones H2A, H2B, H3, and H4, with a total mass of about 100,000 daltons. (B) The nucleosome consists of two full turns of DNA (83 nucleotide pairs per turn) wound around an octameric histone core, plus the adjacent "linker DNA." The part of the nucleosome referred to here as the "nucleosome bead" is released from chromatin by digestion of the DNA with micrococcal nuclease. In each nucleosome bead 146 nucleotide pairs of DNA double helix (about 1.8 turns) remain wound around the octameric histone core. (A, courtesy of Evangelos Moudrianakis.)

Sarepta Exhibit 1056, page 57

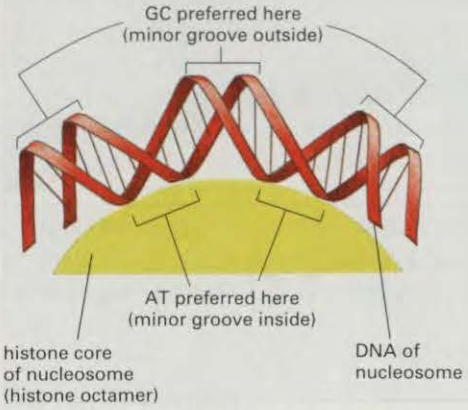


Figure 8–11 The bending of DNA in a nucleosome. The DNA helix makes two tight turns around the histone octamer. This diagram is drawn approximately to scale to illustrate how the minor groove is compressed on the inside of the turn. Due to certain structural features of the DNA molecule, A-T base pairs are preferentially accommodated in a narrow minor groove.

particle with a diameter of about 11 nm containing two copies of each of the four nucleosomal histones—H2A, H2B, H3, and H4. This **histone octamer** forms a protein core around which the double-stranded DNA helix is wound twice (Figure 8–10).

In undigested chromatin the DNA extends as a continuous double-helical thread from nucleosome to nucleosome. Each nucleosome is separated from the next by a region of *linker DNA*, which can vary in length from 0 to 80 nucleotide pairs. On average, nucleosomes repeat at intervals of about 200 nucleotide pairs (see Figure 8–10). Thus, a typical eucaryotic gene of 10,000 nucleotide pairs will be associated with 50 nucleosomes, and each human cell with 6×10^9 DNA nucleotide pairs contains 3×10^7 nucleosomes.

The Positioning of Nucleosomes on DNA Is Determined by the Propensity of the DNA to Form Tight Loops and by the Presence of Other DNA-bound Proteins [8]

Experiments performed *in vitro* with isolated chromatin suggest that the histone octamers generally remain fixed in one position under physiological conditions, inasmuch as their tight binding to DNA prevents them from sliding back and forth along the helix. There are two main influences that determine where nucleosomes form in the DNA. One is the difficulty of bending the DNA double helix into two tight turns around the outside of the histone octamer, a process that requires substantial compression of the minor groove of the helix. Because A-T-rich sequences in the minor groove are easier to compress than G-C-rich sequences, each histone octamer tends to position itself on the DNA so as to maximize A-T-rich minor grooves on the inside of the DNA coil (Figure 8–11). Thus a segment of DNA that contains short A-T-rich sequences spaced by integral numbers of DNA turns will be much easier to bend around the nucleosome than a segment of DNA lacking this feature. This probably explains some striking cases of very precise positioning of nucleosomes, such as those that bind to the tiny 5S rRNA genes, each of which has a single nucleosome bound to it in a unique location. If the DNA containing the 5S rRNA genes is added *in vitro* to a mixture of the four purified nucleosomal histones, nucleosomes will re-form at the exact position where they are located *in vivo*. For most of the DNA sequences found in chromosomes, however, there is no strongly preferred nucleosome binding site; instead, a nucleosome can occupy any one of a number of positions relative to the DNA sequence.

The second important influence on nucleosome positioning is the presence of other tightly bound proteins on the DNA that prevent nucleosomes from forming. For this reason some regions of DNA appear to lack a nucleosome even though they are hundreds of nucleotide pairs long. They can be detected by treating cell nuclei with trace amounts of a deoxyribonuclease (DNase I) that at low concentrations will digest long stretches of nucleosome-free DNA but not the

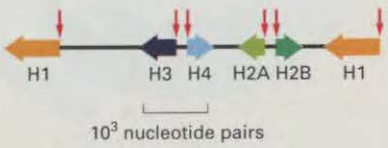


Figure 8–12 The location of nuclease-hypersensitive sites in the regulatory regions of active genes. The genes shown encode histones (H1, H2A, H2B, H3, and H4) in *Drosophila*. The *horizontal arrows* denoting each gene point in the direction of DNA transcription, which always proceeds from the 5′ to the 3′ end of the transcript. Although the nuclease-hypersensitive sites (*vertical red arrows*) are often present on the 5′ side of a gene, as illustrated here, they can also be located elsewhere (see Figure 9–34).

Sarepta Exhibit 1056, page 58

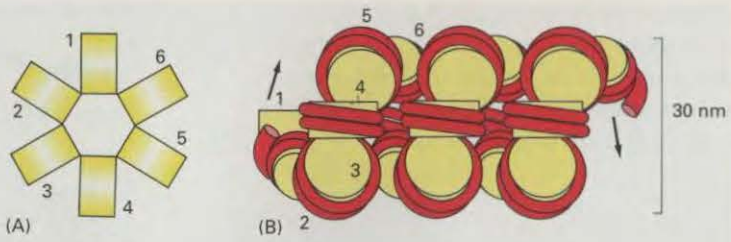


Figure 8–13 The 30-nm chromatin fiber. A model to explain how the "beads-on-a-string" form of nucleosomes is packed to form the 30-nm fiber seen in electron micrographs (see Figure 8–9A) in top (A) and side view (B). This type of packing requires one molecule of histone H1 per nucleosome (not shown). Although the position where the H1 attaches to the nucleosome has been defined (see Figure 8–15), the location of the H1 molecules in this fiber is unknown. (See also Figure 8–14.)

short stretches of linker DNA between nucleosomes. Such **nuclease-hypersensitive sites** often lie in the regulatory regions of genes (Figure 8–12). The first evidence for this idea came from studies of the monkey virus SV40, whose circular DNA chromosome binds to histones produced by its host cells. The SV40 chromosome often contains a single nucleosome-free region about 300 nucleotide pairs long very near the sequences at which viral DNA synthesis and RNA synthesis begin. Although several sequence-specific DNA-binding proteins are bound to this region, they do not protect long stretches of DNA against nuclease attack as do the nucleosomes, which is why the site is DNase-I sensitive.

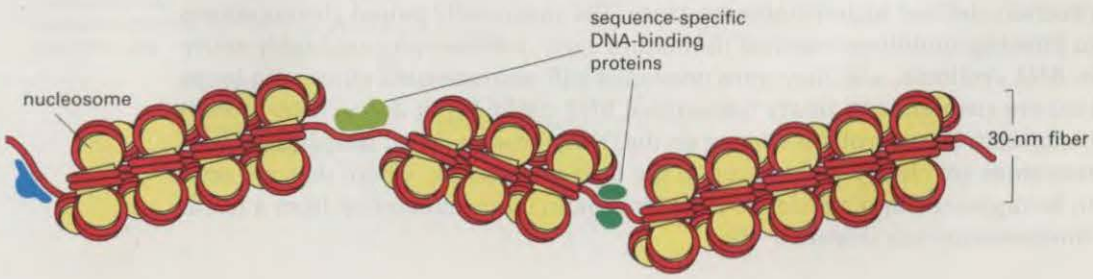
The default state of the DNA in eucaryotic cells is to be fully covered with nucleosomes, and most nucleosome-free regions are specifically created by gene regulatory proteins as part of the process of activating DNA transcription (discussed in Chapter 9). Wherever nucleosomes are specifically positioned by the DNA sequence itself, there may have been evolutionary pressure to keep the adjacent linker DNA free of a nucleosome so as to facilitate its recognition by sequence-specific DNA-binding proteins.

Nucleosomes Are Usually Packed Together by Histone H1 to Form Regular Higher-Order Structures [9]

The linker DNA that connects adjacent nucleosomes can vary in length since nucleosomes position themselves according to the local flexibility of the DNA helix and the distribution of other proteins bound to specific DNA sequences. Although long strings of nucleosomes form on most chromosomal DNA, in the living cell chromatin probably rarely adopts the extended "beads-on-a-string" form. Instead, the nucleosomes are packed upon one another to generate regular arrays in which the DNA is even more highly condensed. Thus, when nuclei are very gently lysed onto an electron microscope grid, most of the chromatin is seen to be in the form of a fiber with a diameter of about 30 nm, which is considerably wider than chromatin in the "beads-on-a-string" form. One of several models proposed to explain how nucleosomes are packed in the 30-nm chromatin fiber is illustrated in Figure 8–13. Such models represent an idealized structure, since both the range of linker lengths that result from preferred nucleosome positioning and the presence of occasional nucleosome-free sequences will punctuate the 30-nm fiber with irregular features (Figure 8–14).

The histone H1 molecules, of which there are about six closely related subtypes in a mammalian cell, are thought to be responsible for pulling nucleosomes together to form the 30-nm fiber. The H1 molecule has an evolutionarily con-

Figure 8–14 Nucleosome-free regions in 30-nm fibers. A schematic section of chromatin illustrating the interruption of its regular nucleosomal structure by short regions where the chromosomal DNA is unusually vulnerable to digestion by DNase I. At each of these nuclease-hypersensitive sites, a nucleosome appears to have been excluded from the DNA by one or more sequence-specific DNA-binding proteins. How these proteins bind DNA tightly is discussed in Chapter 9.



Sarepta Exhibit 1056, page 59

served globular central region linked to less conserved extended amino-terminal and carboxyl-terminal "arms." Each H1 molecule binds through its globular portion to a unique site on a nucleosome, and its arms extend to contact other sites on the histone cores of adjacent nucleosomes, so that the nucleosomes are pulled together into a regular repeating array (Figure 8–15).

Summary

A gene is defined as a nucleotide sequence in a DNA molecule that acts as a functional unit for the production of an RNA molecule. A chromosome is formed from a single, enormously long DNA molecule that contains a series of many genes. A chromosomal DNA molecule also contains three other types of functionally important nucleotide sequences: replication origins and telomeres allow the DNA molecule to be replicated, while a centromere is needed to attach the DNA molecule to the mitotic spindle, ensuring its accurate segregation to daughter cells. The human haploid genome contains 3×10^9 DNA nucleotide pairs, divided among 22 different autosomes and 2 sex chromosomes. Only a small percentage of this DNA is thought to code for proteins.

The DNA in eucaryotes is tightly bound to an equal mass of histones, which form a repeating array of DNA-protein particles called nucleosomes. The nucleosome is made up of an octameric core of histone proteins around which the DNA is wrapped twice. The ease with which a segment of DNA can undergo the severe bending required for this wrapping varies with its nucleotide sequence. Despite irregularities such as this, nucleosomes are usually packed together, with the aid of histone H1 molecules, into regular arrays to form a 30-nm fiber.

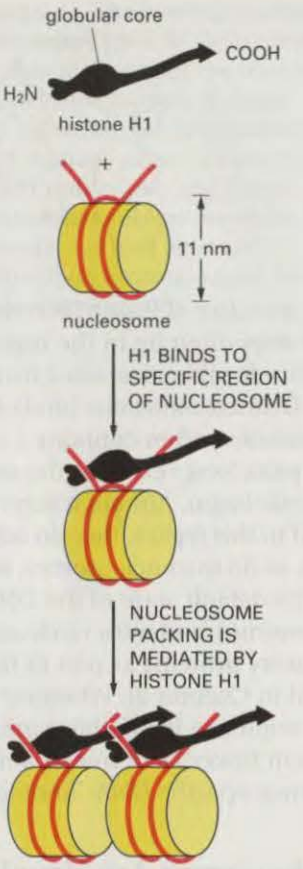


Figure 8–15 The way histone H1 is thought to help pack adjacent nucleosomes together. The globular core of H1 binds to each nucleosome near the site where the DNA helix enters and leaves the histone octamer. When H1 is present on the nucleosomes, 166 nucleotide pairs of DNA are protected from micrococcal nuclease digestion, compared with 146 nucleotide pairs for nucleosomes lacking H1 (see Figure 8–10).

The Global Structure of Chromosomes

Having discussed the DNA and protein molecules from which chromosomes are made, we now turn to the organization of the chromosome on a more global scale. As a 30-nm fiber, the typical human chromosome would be 0.1 cm in length and could span the nucleus more than 100 times. Clearly, there must be a still higher level of folding. The DNA is not only packaged with histones into regularly repeating nucleosomes that are packed into 30-nm fibers; it is also elaborately folded and organized by other proteins into a series of subdomains of distinct character. This higher-order packaging is one of the most fascinating but also most poorly understood aspects of chromatin. Although its molecular basis is still a mystery, this packaging almost certainly plays a crucial part in the regulation of gene transcription. In this section we discuss what is known about higher-order chromatin structure and examine evidence that shows it is functionally important.

Lampbrush Chromosomes Contain Loops of Decondensed Chromatin [10]

Although packaged into chromatin, most chromosomes in interphase cells (cells not in mitosis) are too fine and too tangled to be visualized clearly. In a few exceptional cases, however, interphase chromosomes can be seen to have a precisely defined higher-order structure. The meiotically paired chromosomes in growing amphibian oocytes (immature eggs), for example, are highly active in RNA synthesis, and they form unusually stiff and extended chromatin loops that are covered with newly transcribed RNA packed into dense RNA-protein complexes. Because of this coating on the DNA, these so-called **lampbrush chromosomes** are clearly visible even in the light microscope, where they are seen to be organized into a series of large chromatin loops emanating from a linear chromosomal axis (Figure 8–16).

Sarepta Exhibit 1056, page 60

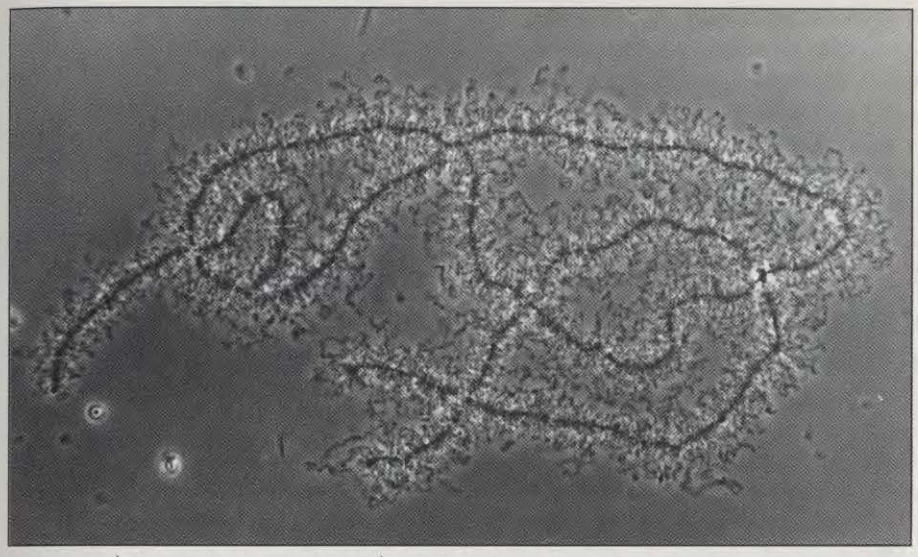


Figure 8–16 Light micrograph of lampbrush chromosomes in an amphibian oocyte. Early in oocyte differentiation each chromosome replicates to begin meiosis, and the homologous replicated chromosomes pair to form this highly extended structure containing a total of four replicated DNA molecules, or chromatids. The lampbrush chromosome stage persists for months or years as the oocyte builds up a supply of mRNA and other materials required for its ultimate development into a new individual. (Courtesy of Joseph G. Gall.)

The organization of the lampbrush chromosome is illustrated schematically in Figure 8–17. Nucleic acid hybridization experiments show that a given loop always contains the same DNA sequence and that it remains extended in the same manner as the oocyte grows. Other experiments demonstrate that most of the looped DNA is being actively transcribed into RNA. The majority of the chro-

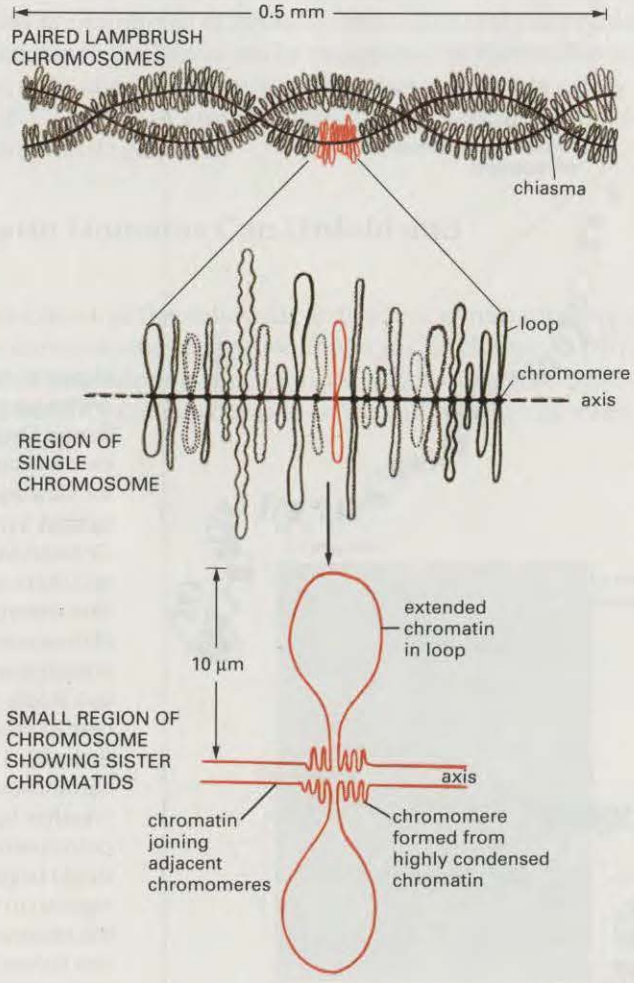


Figure 8–17 Lampbrush chromosome structure. The set of lampbrush chromosomes in many amphibians contains a total of about 10,000 chromatin loops, although most of the DNA in each chromosome remains highly condensed in the *chromomeres*. Each loop corresponds to a particular DNA sequence. Four copies of each loop are present in each cell, since each of the two chromosomes shown at the top consists of two closely apposed sister chromatids. This four-stranded structure is characteristic of this stage of development of the oocyte (the diplotene stage of meiosis, see Figure 20–9).

Sarepta Exhibit 1056, page 61

matin, however, is not in loops but remains highly condensed in the *chromomeres*, which are generally not transcribed. A general chromosome model based on these studies has been proposed in which loops of 30-nm fibers extend at an angle from the main axis of the chromosome (Figure 8–18).

Lampbrush chromosomes illustrate the recurrent themes of this section—when chromatin is actively transcribed, it has an extended structure; when the chromatin is condensed, it is inactive; and the structural units of regulation are large, precisely defined domains.

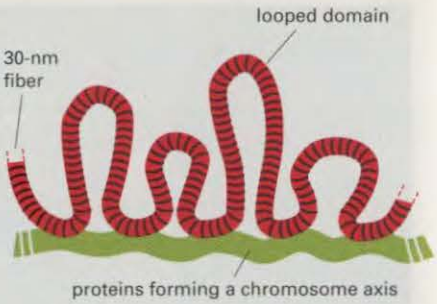


Figure 8–18 A model of chromosome structure. A section of a chromosome is shown folded into a series of looped domains, each containing perhaps 20,000 to 100,000 nucleotide pairs of double-helical DNA condensed in a 30-nm chromatin fiber.

Orderly Domains of Interphase Chromatin Also Can Be Seen in Insect Polytene Chromosomes [11]

Because of a specialization (which is different from that of the lampbrush chromosome), chromatin structure is also unusually visible in certain insect cells.

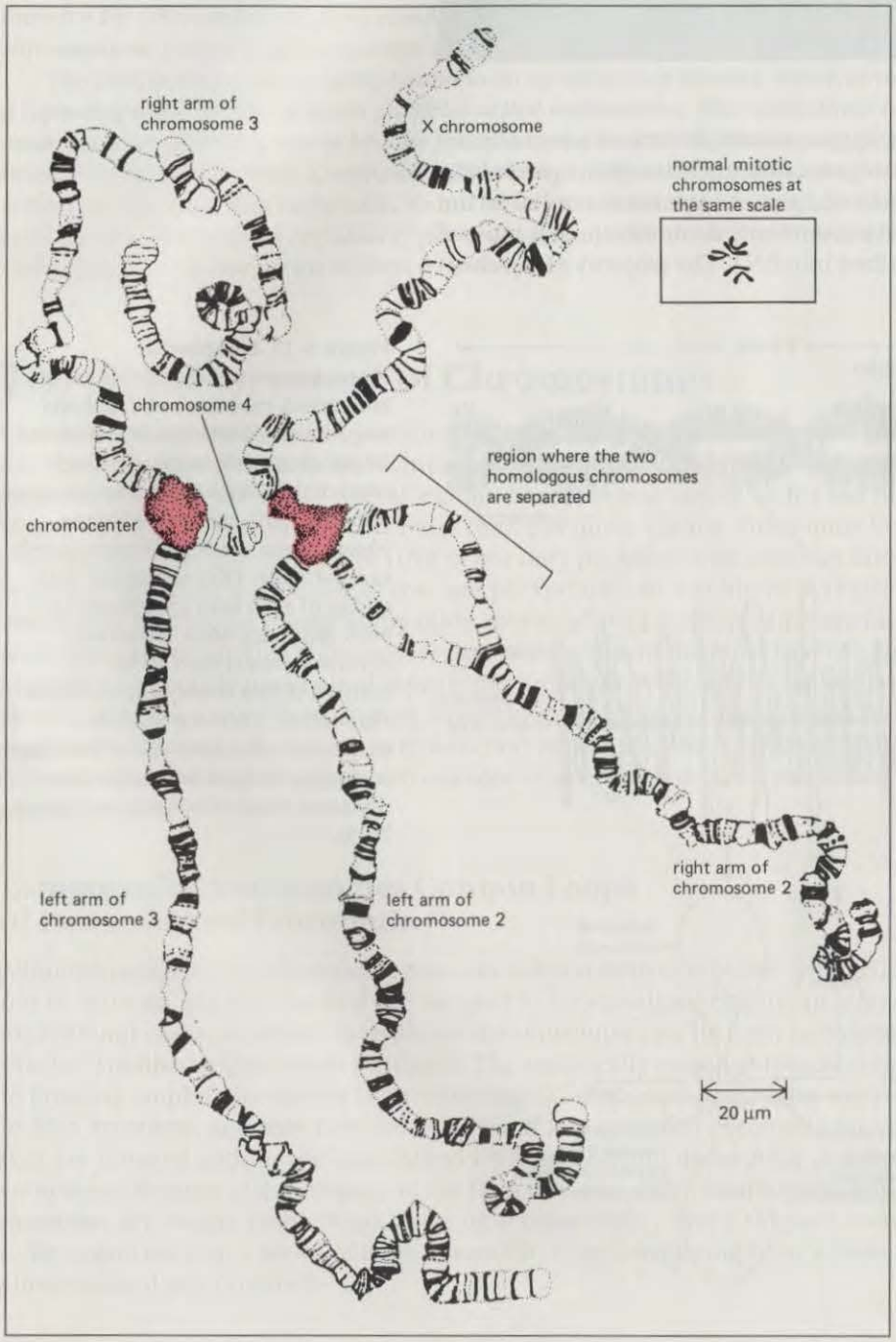


Figure 8–19 A detailed sketch of the entire set of polytene chromosomes in one *Drosophila* salivary cell. These chromosomes have been spread out for viewing by squashing them against a microscope slide. *Drosophila* has four chromosomes, and there are four different chromosome pairs present. But each chromosome is tightly paired with its homologue (so that each pair appears as a single structure), which is not the case in most nuclei (except in meiosis). The four polytene chromosomes are normally linked together by regions near their centromeres that aggregate to create a single large "chromocenter" (*colored region*); in this preparation, however, the chromocenter has been split into two halves by the squashing procedure used. (Modified from T.S. Painter, *J. Hered.* 25:465–476, 1934.)

Sarepta Exhibit 1056, page 62

Many of the cells of the larvae of flies grow to an enormous size through multiple cycles of DNA synthesis without cell division. The resulting giant cells contain as much as several thousand times the normal DNA complement. Cells with more than the normal DNA complement are said to be *polyploid* when, as is usually the case, they contain increased numbers of standard chromosomes. In several types of secretory cells of fly larvae, however, all the homologous chromosome copies remain side by side, creating a single giant *polytene chromosome*. The fact that some large insect cells can undergo a direct polytene-to-polyploid conversion demonstrates that these two chromosomal states are closely related and that the basic structure of a polytene chromosome must be similar to that of a normal chromosome.

Polytene chromosomes are easy to see in the light microscope because they are so large and because the precisely aligned side-to-side adherence of individual chromatin strands greatly elongates the chromosome axis and prevents tangling. Like lampbrush chromosomes, these chromosomes are active in RNA synthesis. Polyteny has been most studied in the salivary gland cells of *Drosophila* larvae, in which the DNA in each of the four *Drosophila* chromosomes has been replicated through 10 cycles without separation of the daughter chromosomes, so that 1024 ($= 2^{10}$) identical strands of chromatin are lined up side by side (Figure 8–19).

When polytene chromosomes are viewed in the light microscope, distinct alternating dark *bands* and light *interbands* are visible (Figure 8–20). Each band and interband represents a set of 1024 identical DNA sequences arranged in register. About 85% of the DNA in polytene chromosomes is in bands, and 15% is in interbands. The chromatin in each band stains darkly because it is much more condensed than the chromatin in the interbands (Figure 8–21). Depending on their size, individual bands are estimated to contain 3000 to 300,000 nucleotide pairs per chromatin strand. Since the bands can be recognized by their different thicknesses and spacings, each one has been given a number to generate a polytene chromosome "map." There are approximately 5000 bands and 5000 interbands in the total *Drosophila* genome.

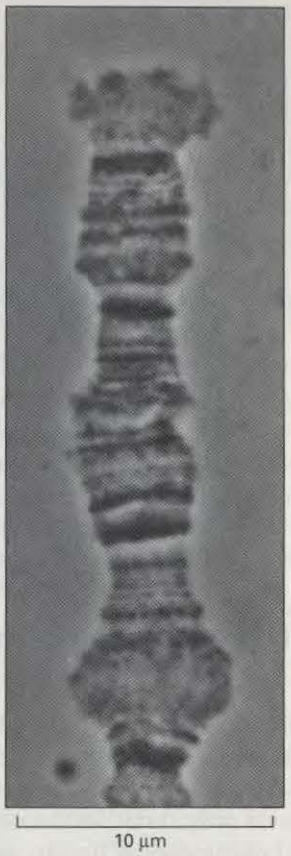


Figure 8–20 Light micrograph of a portion of a polytene chromosome from *Drosophila* salivary glands. The distinct patterns recognizable in different chromosome bands are readily seen. The bands are regions of increased chromatin concentration. They occur in interphase chromosomes and are a special property of the giant polytene chromosomes. (Courtesy of Joseph G. Gall.)

Individual Chromatin Domains Can Unfold and Refold as a Unit [12]

Long before anything was known at the molecular level about chromatin structure, studies of polytene chromosomes suggested that a major change in DNA packing accompanies gene transcription, since individual chromosome bands often expand when the genes they contain become active and recondense when these genes become quiescent.

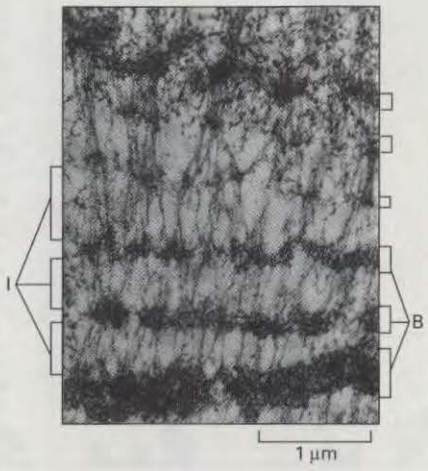


Figure 8–21 Electron micrograph of a small section of a *Drosophila* polytene chromosome seen in thin section. Bands (B) of very different thickness can be readily distinguished, separated by interbands (I), which contain less condensed chromatin. (Courtesy of Viekko Sorsa.)

Sarepta Exhibit 1056, page 63

The regions on a polytene chromosome being transcribed at any instant can be identified by labeling the cells briefly with the radioactive RNA precursor ^{3}H-uridine and locating the growing RNA transcripts by autoradiography (Figure 8–22). This analysis (see below) reveals that the most active chromosomal regions are decondensed, forming distinctive **chromosome puffs.**

One of the main factors controlling the activity of genes in polytene chromosomes of *Drosophila* is the insect steroid hormone *ecdysone,* the levels of which rise and fall periodically during larval development, inducing the transcription of various genes coding for proteins that the larva requires for each molt and for pupation. As the organism progresses from one developmental stage to another, new puffs arise and old puffs recede, as transcription units are activated and deactivated and different mRNAs and proteins are made (Figure 8–23). From inspection of each puff when it is relatively small and the banding pattern of the chromosome is still discernible, it seems that most puffs arise from the uncoiling of a single chromosome band (Figure 8–24). Electron microscopy of thin sections of such puffs shows that the DNA in the chromatin is much less condensed than it would be in the 30-nm chromatin fiber. These observations suggest that the chromatin in a band can decondense as a unit during transcription.

Both Bands and Interbands in Polytene Chromosomes Are Likely to Contain Genes [13]

The fixed pattern of bands and interbands in a *Drosophila* polytene chromosome suggested to early cytogeneticists that each band might correspond to a single gene. Mutational analyses that allowed geneticists to estimate that *Drosophila* contains only about 5000 essential genes, a number roughly equal to the number of chromosome bands, supported this view. Moreover, when an intensive effort was made to isolate as many mutants as possible in a small chromosomal region, about 50 essential genes were genetically identified in a region that contains about 50 visible bands. Although it is not possible with these techniques to determine whether a particular gene lies in a band or an interband region, these observations suggested that an average band might contain the DNA coding sequences for only one essential protein.



Figure 8–22 Synthesis of RNA along a giant polytene chromosome. In this autoradiograph of a chromosome (from the salivary gland of the insect *Chironomus tentans*) labeled with ^{3}H-uridine, sites of RNA synthesis are covered with dark silver grains in proportion to their activity. (From C. Pelling, *Chromosoma* 15:71–122, 1964.)

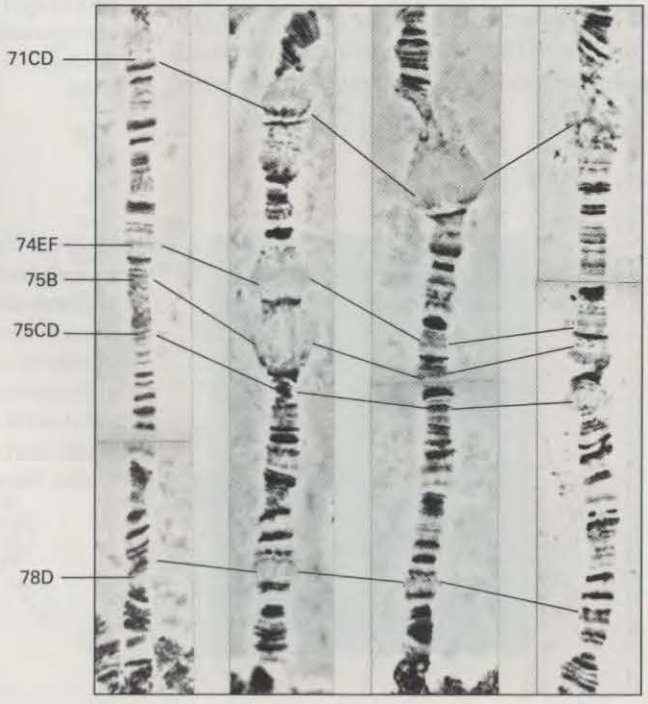


Figure 8–23 Chromosome puffs. A temporal series of photographs illustrating how puffs arise and recede in the polytene chromosomes of *Drosophila melanogaster* during larval development. A region of the left arm of chromosome 3 is shown. It exhibits five very large puffs in salivary gland cells, each active for only a short developmental period. The series of changes shown occur over a period of 22 hours, appearing in a reproducible pattern as the organism develops. (Courtesy of Michael Ashburner.)

Sarepta Exhibit 1056, page 64

Genes that are not essential for survival in the laboratory (where predators are absent and both food and mates are provided), however, would have been missed in these genetic analyses, and more recent data have made the simple "one-band, one-gene" hypothesis seem unlikely. A continuous 315,000-nucleotide-pair region of the *Drosophila* genome has been cloned, for example, and fragments have been used as DNA probes to catalogue the mRNAs produced from the region: three times as many separate mRNAs as bands were identified, indicating that each band probably contains several genes. In addition, it has now been shown directly that mRNAs are produced from both interbands and bands in polytene chromosomes.

Although still controversial, it has been proposed that the DNA in polytene chromosome bands is arranged in loops in a manner analogous to that of the lampbrush chromosomes. According to this model, puff formation would correspond to the decondensation of one or more looped domains. Whether or not this model is correct, the above findings show that at least some interphase chromosomes are arranged in a complex, precisely ordered sequence of structural domains, each typically containing a small number of genes whose transcription is regulated in a coordinated way. It may be that all interphase chromosomes are packaged into ordered structures according to these same principles.

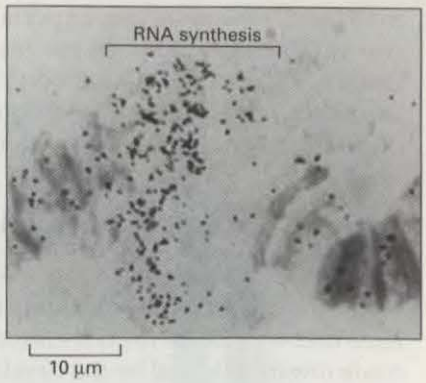


Figure 8–24 Autoradiogram of a single puff in a polytene chromosome. The portion of the chromosome indicated is undergoing RNA synthesis and has therefore become labeled with ^{3}H-uridine. (Courtesy of Jose Bonner.)

Transcriptionally Active Chromatin Is Less Condensed [14]

The studies of puffs on insect polytene chromosomes suggest that the chromatin structure of transcribed genes is selectively decondensed. Experiments of a completely different type support this suggestion. When nuclei isolated from vertebrate cells are exposed to the enzyme DNase I, about 10% of the genome is preferentially degraded. Although a small subset of these degraded sequences corresponds to the hypersensitive sites described earlier, most of them represent genes that are being transcribed: different DNA sequences, for example, are digested by the nuclease in different cells of the same organism, depending on the genes that are being transcribed in the different cell types. Remarkably, even genes that are transcribed only a few times in every cell generation are sensitive

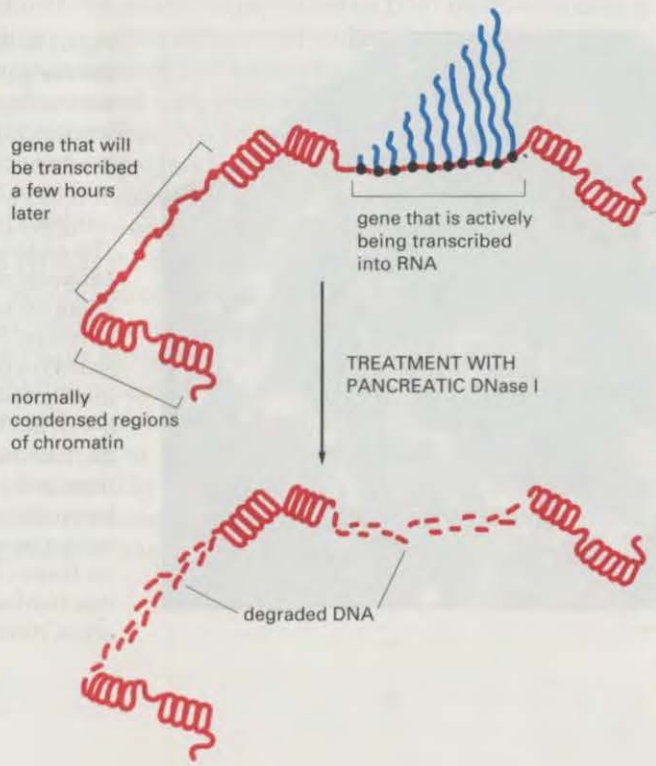


Figure 8–25 Digestion of chromatin with DNase I. The enzyme first cuts at nuclease-hypersensitive sites (not shown, but see Figure 8–12) and then selectively degrades the entire DNA sequence both of genes that are actively being transcribed and of genes that have been activated but are not yet being transcribed.

Sarepta Exhibit 1056, page 65

to DNase I, indicating that a special state of the chromatin, rather than the process of DNA transcription itself, makes these regions unusually accessible to digestion (Figure 8–25). Chromatin in this accessible state is often termed **active chromatin,** and its structure appears to be altered in a way that makes the packing of the nucleosomes less condensed.

Active Chromatin Is Biochemically Distinct [15]

What distinguishes active from inactive chromatin? The answer to this question will probably require purification and characterization of the chromosomal proteins that are unique to each chromatin conformation. Some progress has been made toward this goal by the development of biochemical methods designed to isolate active chromatin by taking advantage of its relatively decondensed state. The analysis of the chromosomal proteins in the active chromatin fraction has suggested the following: (1) Histone H1 seems to be less tightly bound to at least some active chromatin, and particular subtypes of this histone may be specific for active chromatin. (2) Although the four nucleosomal histones are present in normal amounts in active chromatin, they appear to be unusually highly acetylated when compared with the same histones in inactive chromatin. (3) The nucleosomal histone H2B in active chromatin appears to be less phosphorylated than it is in inactive chromatin. (4) Active chromatin is highly enriched in a minor variant form of histone H2A that is found in many species, including *Drosophila* and humans. (5) The nucleosomes in active chromatin selectively bind two closely related small chromosomal proteins, called HMG 14 and HMG 17. Consistent with their presence only in active chromatin, these "high-mobility-group" (HMG) proteins occur in roughly the quantities required to bind to 1 in every 10 nucleosomes in total chromatin; moreover, their amino acid sequences have been highly conserved during evolution, implying that they have important functions.

Any or all of these changes might play an important part in uncoiling the chromatin of active genes, helping to make the DNA available as a template for RNA synthesis, but more direct experiments are needed to test this idea. In Chapter 9 we discuss mechanisms by which the transition between inactive and active chromatin may be controlled.

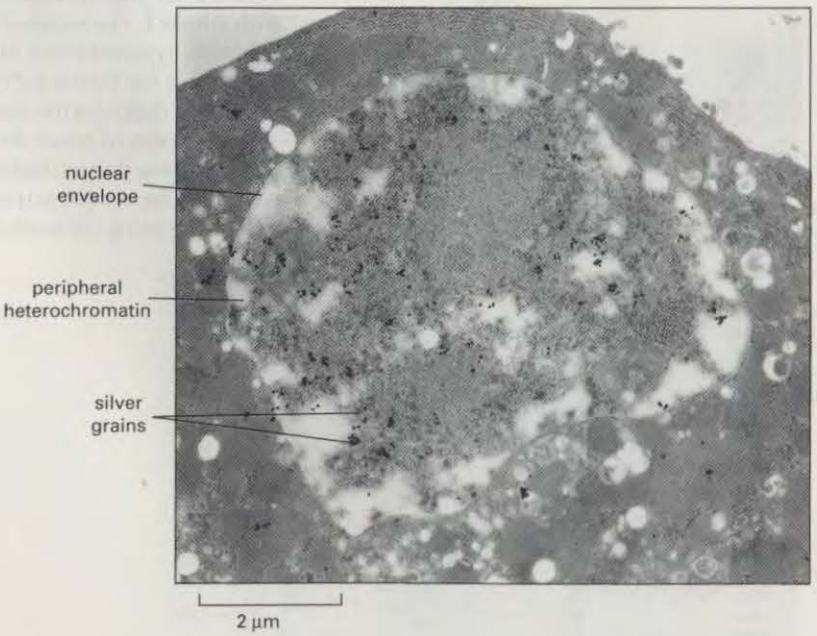


Figure 8–26 Heterochromatin is transcriptionally inactive. An autoradiograph is shown of a thin section of a nucleus from a cell that has been pulse-labeled with ^{3}H-uridine in order to label sites of RNA synthesis (seen as black silver grains). The *white areas* are regions of heterochromatin, which tend to pack along the inside of the nuclear envelope. Normally, heterochromatin stains unusually darkly in electron micrographs (for example, see Figure 8–71), but because of the particular method used for sample preparation here, areas containing heterochromatin have been bleached. As can be seen, most of the RNA synthesis occurs in the euchromatin that borders these heterochromatic areas. (Courtesy of Stan Fakan.)

Sarepta Exhibit 1056, page 66

Heterochromatin Is Highly Condensed and Transcriptionally Inactive [16]

Light microscopic studies in the 1930s distinguished two types of chromatin in interphase nuclei of higher eucaryotic cells: a highly condensed form called **heterochromatin** and all the rest, which is less condensed, called **euchromatin.** Heterochromatin was originally identified because it remains unusually compact during interphase; it was later found to be transcriptionally inactive (Figure 8–26). In a typical cell in interphase, approximately 10% of the genome is packed into heterochromatin. It is now clear, as described above, that the euchromatin exists in at least two forms: about 10% is in the form of active chromatin, which is the least condensed, while the rest is inactive euchromatin, which is more condensed than active chromatin but less condensed than heterochromatin. Heterochromatin is therefore thought to be a special variety of transcriptionally inactive chromatin with distinctive functions. In mammals, for example, and in many other higher eucaryotes, the DNA surrounding each centromere is composed of relatively simple repeating nucleotide sequences. These so-called *satellite DNAs* constitute a major portion of the heterochromatin in these organisms.

Mitotic Chromosomes Are Formed from Chromatin in Its Most Condensed State [17]

With the exception of a few specialized cases, such as the lampbrush and polytene chromosomes discussed above, most *interphase* chromosomes are too extended, thin, and entangled to be readily detectable. In contrast, chromosomes from nearly all cells are visible during mitosis, when they coil up to form much more condensed structures. This coiling, which reduces a 5-cm length of DNA to about 5 μm, makes it possible for the mitotic spindle to segregate the chromosomes to separate daughter cells without breaking them. The condensation is accompanied by the phosphorylation of all of the histone H1 in the cell at five of its serine residues. Because histone H1 helps to pack nucleosomes together (see Figure 8–15), its phosphorylation during mitosis is suspected to play a role in chromosome condensation.

Figure 8–27 depicts a typical **mitotic chromosome** at the metaphase stage of mitosis. The two daughter DNA molecules produced by DNA replication during the S phase of the cell-division cycle are separately folded to produce two sister chromosomes (called *sister chromatids*) held together at their *centromeres.* These chromosomes are normally covered with a variety of molecules, including large amounts of ribonucleoproteins. Once this covering is stripped away, each chromatid can be seen in electron micrographs to be organized into loops of chromatin emanating from a central axis (Figures 8–28 and 8–29). Several types of experiments demonstrate that the order of visible features along a mitotic chromosome at least roughly reflects the order of the genes along the DNA molecule. To illustrate how such an organized folding of the long DNA helix might be achieved, we have drawn each chromatid as a closely packed series of looped domains wound in a tight helix in Figure 8–30, which also presents a schematic view of all the different folding processes that would contribute to this structure.

The condensed form of chromatin found in mitotic chromosomes is thought to resemble heterochromatin in its high degree of packing. Not surprisingly, mitotic chromosomes are found to be transcriptionally inactive: all RNA synthesis ceases as the chromosomes condense. Presumably, the condensation prevents RNA polymerase from gaining access to the DNA, although other control factors may also be involved.

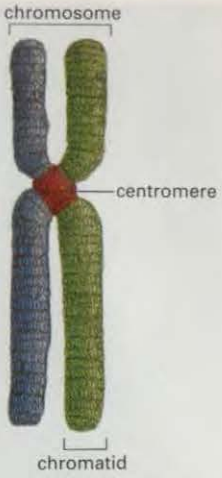


Figure 8–27 Drawing of a typical metaphase chromosome. Each sister chromatid contains one of two identical daughter DNA molecules generated earlier in the cell cycle by DNA replication.



Figure 8–28 Scanning electron micrograph of a region near one end of a typical mitotic chromosome. Each knoblike projection is believed to represent the tip of a separate looped domain. Note that the two identical paired chromatids diagrammed in Figure 8–27 can be clearly distinguished. (From M.P. Marsden and U.K. Laemmli, *Cell* 17:849–858, 1979. © Cell Press.)

Sarepta Exhibit 1056, page 67

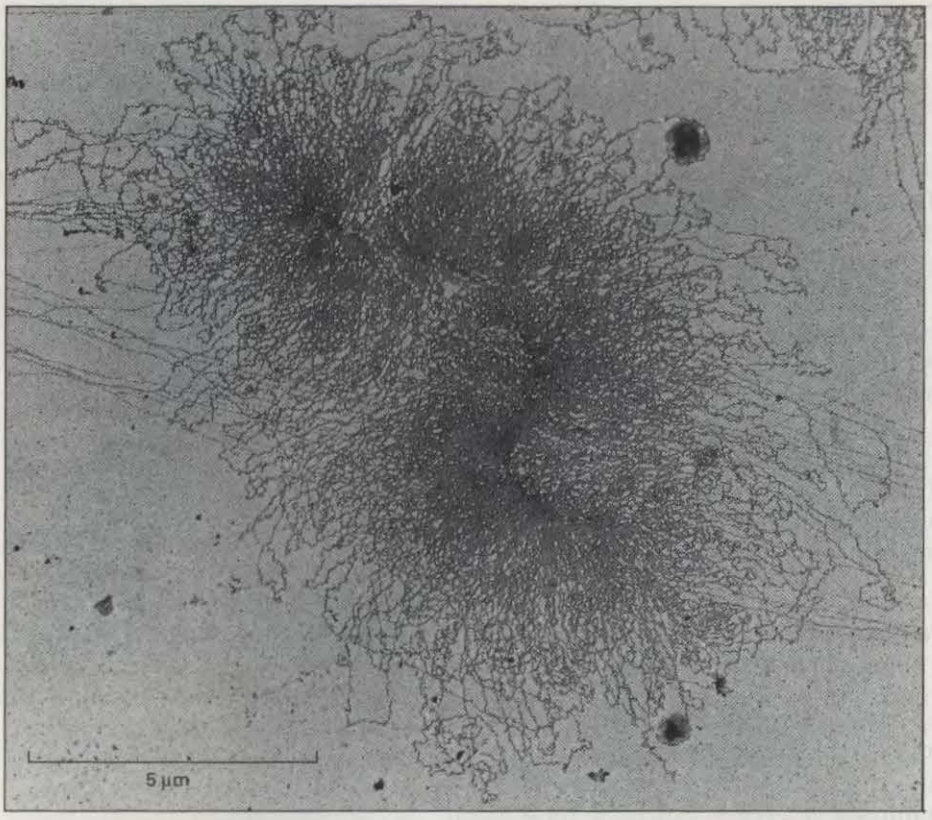


Figure 8–29 Electron micrograph of a single chromatid of a mitotic chromosome. The chromosome (from an insect) was treated to reveal loops of chromatin fibers that emanate from a central axis of the chromatid. Such micrographs support the idea that the chromatin in all chromosomes is folded into a series of looped domains (see Figure 8–18). (Courtesy of Victoria Foe.)

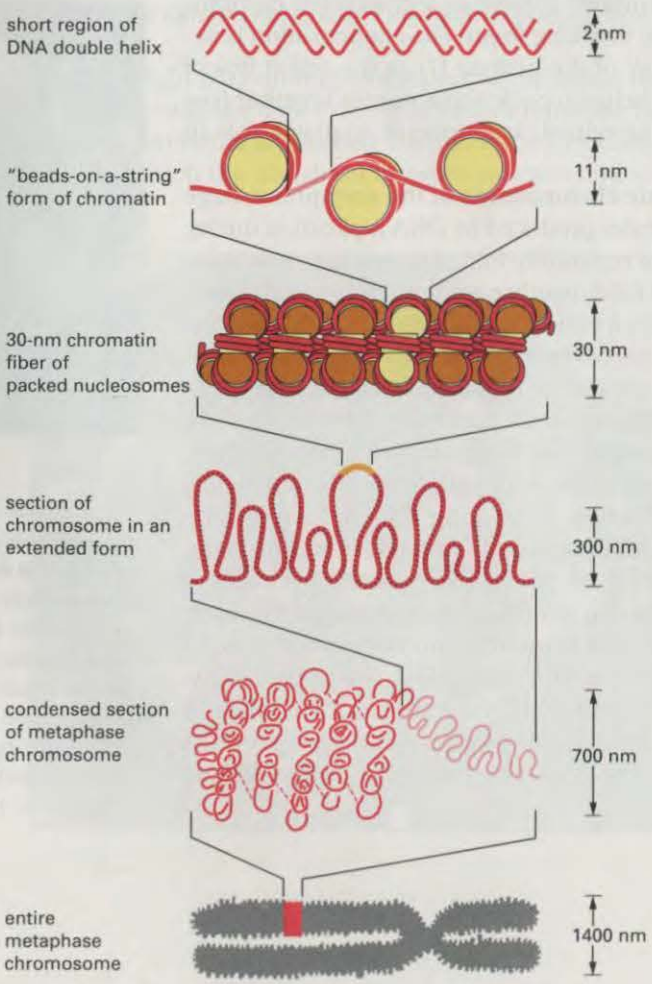


Figure 8–30 Model of chromatin packing. This schematic drawing shows some of the many orders of chromatin packing postulated to give rise to the highly condensed mitotic chromosome.

Sarepta Exhibit 1056, page 68

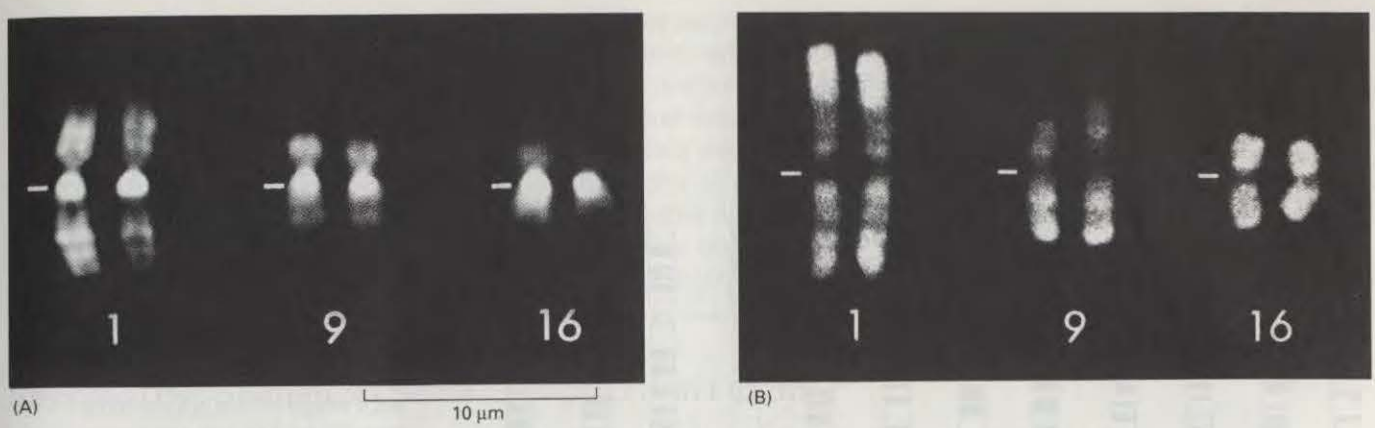


Figure 8–31 Fluorescence micrographs showing the banding patterns in three pairs of human mitotic chromosomes. In (A) the chromosomes were stained with the A-T-base-pair-specific dye Hoechst 33258, which stains the G bands, and in (B) with the G-C-base-pair-specific dye olivomycin, which stains the R bands. The bars indicate the position of the centromere. Note that these banding patterns are the reverse of each other: the bands that are bright in (A) are dark in (B), and vice versa. The G bands also stain with Giemsa stain—hence their name—while the R bands were so named because they formed the "reverse" of the G-band pattern. (From K.F. Jorgenson, J.H. van de Sande, and C.C. Lin, *Chromosoma* 68:287–302, 1978.)

Each Mitotic Chromosome Contains a Characteristic Pattern of Very Large Domains [18]

The display of the 46 human chromosomes at mitosis is called the human **karyotype.** Staining methods developed over the last 25 years permit unambiguous identification of each individual chromosome. Some of these methods involve staining mitotic chromosomes with dyes that fluoresce only when they bind to certain types of DNA sequences. Although these dyes have very low specificity and appear mainly to distinguish DNA that is rich in A-T nucleotide pairs (**G bands**) from DNA that is rich in G-C nucleotide pairs (**R bands**), they produce a striking and reproducible banding pattern along each mitotic chromosome (Figure 8–31). In this way each chromosome can be identified and numbered, as illustrated in Figure 8–32. Both the G bands and the R bands are known to contain genes. These bands are unrelated to those described earlier for the insect polytene chromosomes, which are thought to correspond to regions of condensed chromatin; in the human mitotic chromosomes, all of the chromatin is condensed and the bands are formed by the selective binding of specific dyes. Whereas a band in a polytene chromosome contains only a few genes, a typical band in a human karyotype contains several hundred.

By examining human chromosomes very early in mitosis, when they are less condensed than at metaphase, it has been possible to estimate that the total haploid genome contains about 2000 distinguishable A-T-rich bands. These progressively coalesce as condensation proceeds during mitosis to produce fewer and thicker bands.

Chromosome bands are detected in mitotic chromosomes from species as diverse as humans and flies. Moreover, the exact pattern of bands in a chromosome has remained unchanged over long periods of evolutionary time. Every human chromosome, for example, has a clearly recognizable counterpart with a nearly identical banding pattern in the chromosomes of the chimpanzee, gorilla, and orangutan (although there has been a single chromosome fusion that gives humans 46 chromosomes instead of the apes' 48). This conservation suggests that the long-range spatial organization of chromosomes may be important for chromosomal function. But why such bands exist at all is a mystery. Even the thinnest of the bands diagrammed in Figure 8–32 probably contain more than a million nucleotide pairs, which is nearly the size of a bacterial genome, and the average nucleotide-pair composition would be expected to be random over such long stretches of DNA sequence.

Summary

Chromosomes are generally decondensed during interphase, so that their structure is difficult to discern. Notable exceptions are the specialized lampbrush chromosomes of vertebrate oocytes and the polytene chromosomes of insect giant secretory cells. Studies of these two types of interphase chromosomes suggest that each long DNA

Sarepta Exhibit 1056, page 69



Figure 8–32 A standard map of the banding pattern of each chromosome in the human karyotype. This map was determined at the prometaphase stage of mitosis. Chromosomes 1 through 22 are labeled in the approximate order of their size; a diploid cell contains two of each of these autosomes plus two sex chromosomes—two X chromosomes (female) or an X and a Y chromosome (male). The 850 bands shown here are G bands, which stain with reagents that appear to be specific for A-T-rich DNA sequences. The *red knobs* on chromosomes 13, 14, 15, 21, and 22 indicate the positions of the genes that encode the large ribosomal RNAs; the *red lines* mark the centromere on each chromosome. (Adapted from U. Franke, *Cytogenet. Cell Genet.* 31:24–32, 1981.)

molecule in a chromosome is divided into a large number of discrete domains that are folded differently. In both lampbrush and polytene chromosomes the regions that are actively synthesizing RNA are least condensed. Likewise, as judged by nuclease sensitivity, about 10% of the DNA in interphase vertebrate cells is in a relatively uncondensed conformation that correlates with DNA transcription in these regions. Such "active chromatin" is biochemically distinct from the more condensed inactive regions of chromatin.

All chromosomes adopt a highly condensed conformation during mitosis. When they are specially stained, these mitotic chromosomes have a banded structure that allows each individual chromosome to be recognized unambiguously; these bands contain millions of DNA nucleotide pairs, and they reflect a coarse heterogeneity of chromosome structure that is not understood.

Chromosome Replication

Before a cell can divide, it must produce a new copy of each of its chromosomes, and it does this during a specific part of interphase called the DNA-synthesis phase, or **S phase,** of the cell-division cycle; the part of interphase preceding S phase is called Gap 1, or $\mathbf{G_1}$, and the part following S phase is called Gap 2, or $\mathbf{G_2}$ (see Figure 17–3). In a typical higher eucaryotic cell the S phase lasts for about 8 hours. By its end each chromosome has been replicated to produce two complete copies, which remain joined together at their centromeres until the M phase

Sarepta Exhibit 1056, page 70

that soon follows (see Figure 8–27). Chromosome duplication requires both the replication of the long DNA molecule in each chromosome and the assembly of a new set of chromosomal proteins onto the DNA to form chromatin. In Chapter 6 we discussed the enzymology of DNA replication and described the structure of the *DNA replication fork*, where DNA synthesis occurs by a semiconservative process (see Figures 6–48 and 6–49). In Chapter 17 we consider how entry into the S phase is regulated as part of a more general discussion of how the cell-division cycle is controlled. In this section we describe the timing and pattern of eucaryotic chromosome replication and its relation to chromosome structure.

Specific DNA Sequences Serve as Replication Origins [19]

We saw in Chapter 6 that replication origins have been precisely defined in bacteria as specific DNA sequences that allow the DNA replication machinery to assemble on the DNA double helix and move in opposite directions to produce *replication forks*. By analogy, one would expect the replication origins in eucaryotic chromosomes to be specific DNA sequences too. As mentioned earlier, the search for replication origins in the chromosomes of eucaryotic cells has been most successful in the yeast *Saccharomyces cerevisiae*. The powerful selection methods that have been devised to find them make use of mutant yeast cells that are defective for an essential gene: they will survive in a selective medium only if they are provided with a plasmid carrying a functional copy of the missing gene. If a bacterial plasmid that carries the essential yeast gene is introduced into the mutant yeast cells, it will not be able to replicate free of the host chromosome because the bacterial replication origin in the plasmid DNA cannot function in a yeast cell. If random pieces of yeast DNA are inserted into the plasmid, however, a small proportion of the plasmid DNA molecules will contain a yeast replication origin and therefore will be able to replicate in the yeast cells. The yeast cells that carry such plasmids will be able to proliferate since they will have been provided with the essential gene (Figure 8–33). A DNA sequence identified by its presence in a plasmid isolated from these cells is called an *autonomously replicating sequence (ARS)*. A number of ARSs recently have been shown to be authentic chromosomal origins of replication, thereby justifying the strategy used to obtain them.

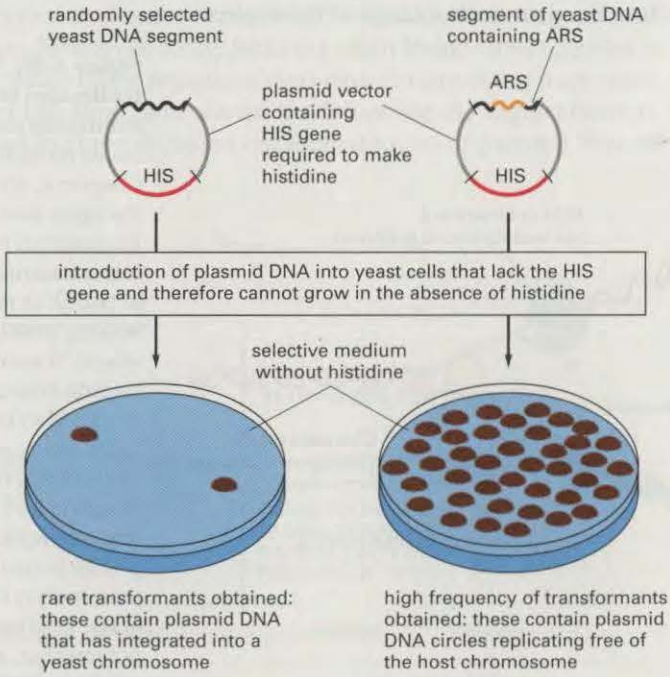


Figure 8–33 The strategy used to identify replication origins in yeast cells. A DNA sequence identified in this way was initially called an autonomously replicating sequence (ARS) since it enables a plasmid that contains it to replicate freely in the host cell without having to be incorporated into its chromosomes.

Sarepta Exhibit 1056, page 71

An 11-nucleotide "core consensus sequence" has been shown to be essential for yeast replication-origin function, and all known origins contain multiple near matches of this sequence, spaced throughout a region of about 100 nucleotides. These DNA sequences are recognized by a large protein complex that is thought to function in the initiation process. Although some replication-origin sequences in higher eucaryotic cells have been tentatively identified, the proteins that bind to them are not known.

A Mammalian Cell-free System Replicates the Chromosome of a Monkey Virus [20]

In bacteria a multienzyme complex that includes DNA polymerase, DNA primase, and DNA helicase moves the replication fork (see Figures 6–48 and 6–49); this complex assembles at a replication origin in a reaction involving an initiator protein that binds specifically to the origin DNA (see Figure 6–52).

Many attempts have been made to reconstitute mammalian DNA replication in a test tube. To do so would make it possible to identify all the enzymes required and to define their actions at a replication fork. The most successful *in vitro* system developed thus far replicates the small circular chromosome of the monkey virus SV40. This virus parasitizes its mammalian host cell for all but one of the proteins that the virus needs to replicate; the one exception is the *SV40 T-antigen*, a large multifunctional protein that allows the virus to bypass normal controls and thereby replicate faster than the host cell. Two hexamers of the T-antigen bind specifically to the SV40 replication origin, recognizing the outside of the DNA double helix. Then, in a process that is not well understood, the bound T-antigen undergoes a conformational change that pulls apart the DNA strands of the origin. This remarkable protein has a third function: utilizing the energy of ATP hydrolysis, it acts as a DNA helicase, moving along the DNA and opening the DNA double helix as it moves, forming a *replication bubble* (Figure 8–34). The cellular components of the replication machinery can now assemble in the bubble to form two **replication forks**, which then move away from the origin in opposite directions.

Although the replication forks are similar to those of procaryotes, the SV40 experiments have revealed that at least two distinct types of DNA polymerase are needed in eucaryotes: *DNA polymerase α (alpha)* on the lagging strand and *DNA polymerase δ (delta)* on the leading strand (Figure 8–35). In procaryotes, by contrast, a single type of DNA polymerase functions on both strands of the replication fork (see Figure 6–48).

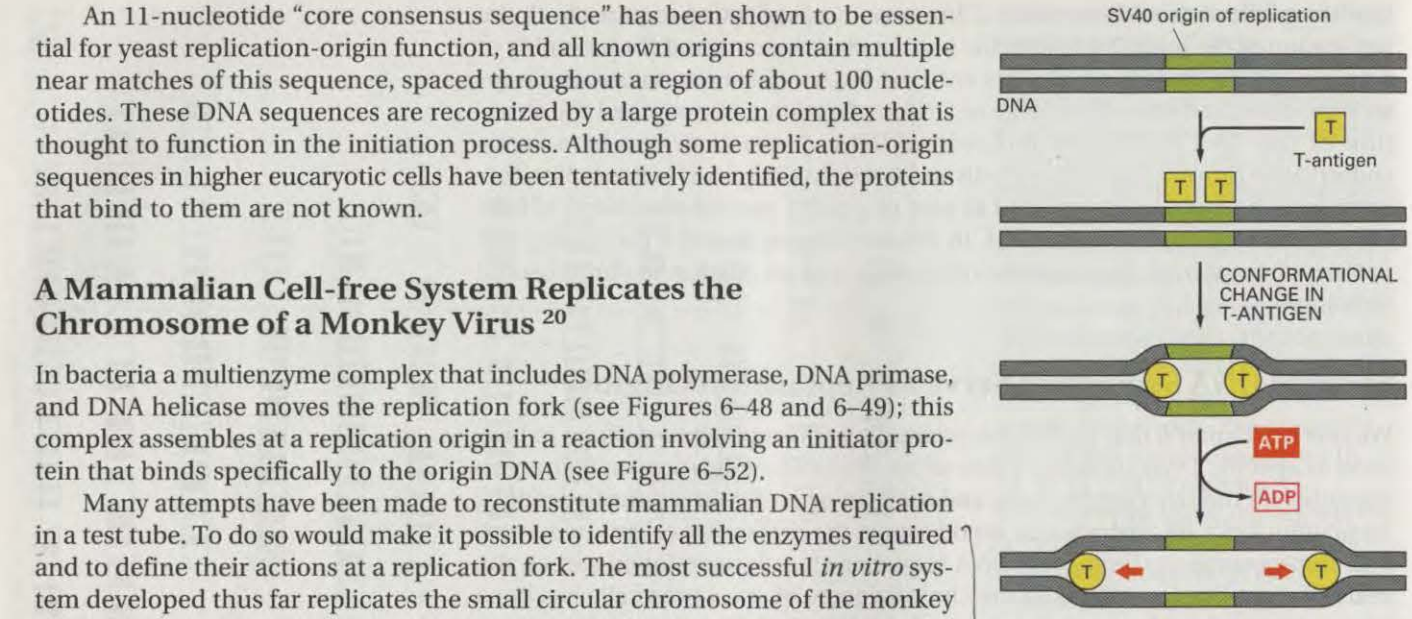


Figure 8–34 Initiation of DNA replication on the SV40 viral genome. The initiation protein (T-antigen) is encoded by the virus, and it specifically recognizes the viral origin of replication. After opening the DNA double helix at the origin, the T-antigen acts as a DNA helicase, traveling away from the origin and opening the double helix as it moves. Subsequent to the steps depicted, host cell components of the replication machinery assemble in the "replication bubble" to form two replication forks (see Figure 8–35).

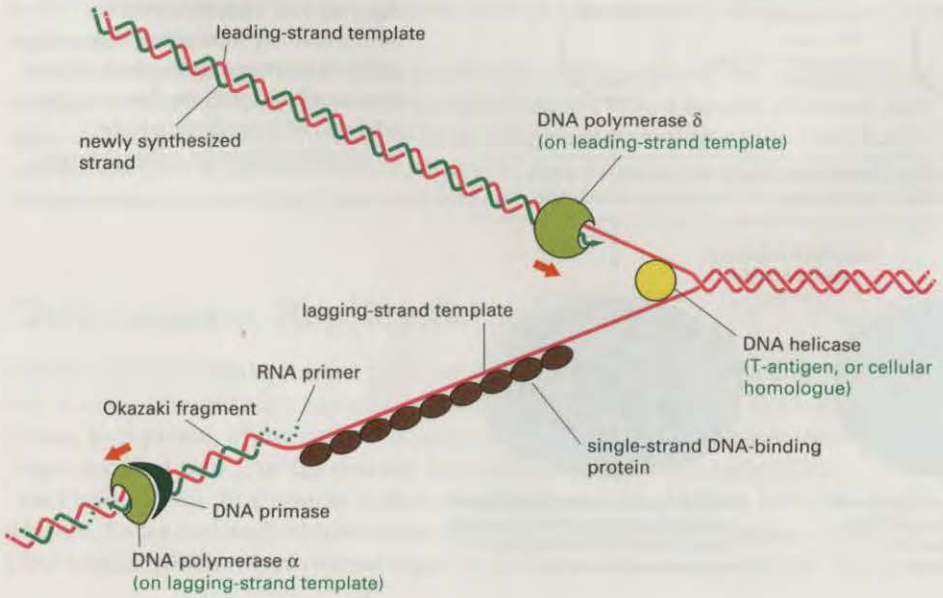


Figure 8–35 A mammalian replication fork. The fork is drawn to emphasize its similarity to the bacterial replication fork described in Chapter 6. Although both forks utilize the same basic components, the mammalian fork differs in two important respects. First, it makes use of two DNA polymerases, one for the leading strand and one for the lagging strand. It seems likely that the leading-strand polymerase is designed to keep a tight hold on the DNA, whereas that on the lagging strand must be able to release the template and then rebind each time that a new Okazaki fragment (see p. 256) is synthesized. Second, the mammalian DNA primase is a subunit of the lagging-strand DNA polymerase, while that of bacteria is associated with the DNA helicase.

Sarepta Exhibit 1056, page 72

The protein that normally initiates DNA replication of mammalian chromosomes has not yet been discovered. It is therefore not known whether it resembles the T-antigen in functioning as a DNA helicase at the fork or whether, as in procaryotes, a separate initiator protein and DNA helicase are involved.

Replication Origins Are Activated in Clusters on Higher Eucaryotic Chromosomes [21]

In Chapter 6 we describe how in bacteria two replication forks begin at each origin and proceed in opposite directions, moving away from the origin at a rate of about 500 nucleotides per second until all of the DNA in the single circular bacterial chromosome is replicated. The bacterial genome is so small that these two replication forks can duplicate it in less than 40 minutes. Because of the much larger size of most eucaryotic chromosomes, it seemed unlikely that their replication would begin from a single origin.

A method for determining the general pattern of eucaryotic chromosome replication was developed in the early 1960s. Human cells growing in culture are labeled for a short time with ^{3}H-thymidine so that the DNA synthesized during this period becomes highly radioactive. The cells are then gently lysed, and the DNA is streaked onto the surface of a glass slide, which is coated with a photographic emulsion, so that the pattern of labeled DNA can be determined by autoradiography. Since the time allotted for radioactive labeling is chosen to allow each replication fork to move several micrometers along the DNA, the replicated DNA can be detected in the light microscope as lines of silver grains, even though the DNA molecule itself is too thin to be visible. In this way both the rate and the direction of replication-fork movement can be determined (Figure 8–36). From the rate at which tracks of replicated DNA increase in length with increasing labeling time, the replication forks are estimated to travel at a speed of about 50 nucleotides per second. This is one-tenth the rate at which bacterial replication forks move, possibly reflecting the increased difficulty of replicating DNA that is packaged tightly in chromatin.

As discussed previously, an average human chromosome is composed of a single DNA molecule containing about 150 million nucleotide pairs. To replicate such a DNA molecule from end to end with a single replication fork moving at a rate of 50 nucleotides per second would require $0.02 \times 150 \times 10^6 = 3.0 \times 10^6$ seconds (about 800 hours). As expected, therefore, the autoradiographic experiments just described reveal that many forks are moving simultaneously on each eucaryotic chromosome. Moreover, many forks are often found close together in the same DNA region, while other regions of the same chromosome have none. Further experiments of this type have shown the following: (1) Replication origins tend to be activated in clusters (called *replication units*) of perhaps 20 to 80

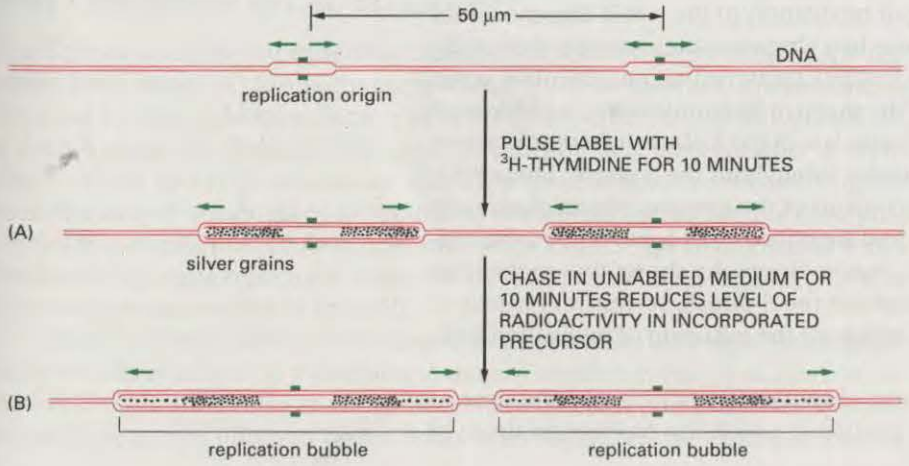


Figure 8–36 The experiments that demonstrated the pattern in which replication forks move during the S phase. The new DNA made in human cells in culture was briefly labeled with a pulse of highly radioactive thymidine (^{3}H-thymidine). In the experiment illustrated in (A), the cells were lysed and the DNA was stretched out on a glass slide that was subsequently covered with a photographic emulsion. After several months the emulsion was developed, revealing a line of silver grains over the radioactive DNA. The experiment in (B) was the same except that a further incubation in unlabeled medium allowed additional DNA, with a lower level of radioactivity, to be replicated. The pairs of dark tracks in (B) were found to have silver grains tapering off in opposite directions, demonstrating bidirectional fork movement from a central replication origin (see Figure 6–51). The *red* DNA in this figure is shown only to help with the interpretation of the autoradiograph; the unlabeled DNA is invisible in such experiments. A replication fork is thought to stop only when it encounters a replication fork moving in the opposite direction or when it reaches the end of the chromosome; in this way all the DNA is eventually replicated.

Sarepta Exhibit 1056, page 73

origins. (2) New replication units seem to be activated throughout the S phase until all of the DNA is replicated. (3) Within a replication unit, individual origins are spaced at intervals of 30,000 to 300,000 nucleotide pairs from one another. (4) As in bacteria, replication forks are formed in pairs and create a replication bubble as they move in opposite directions away from a common point of origin, stopping only when they collide head-on with a replication fork moving in the opposite direction (or reach a chromosome end). In this way many replication forks can operate independently on each chromosome and yet form two complete daughter DNA helices (see Figure 8–36).

Different Regions on the Same Chromosome Replicate at Distinct Times [22]

The replication of DNA in the region between one replication origin and the next should normally require only about an hour to complete, given the rate at which a replication fork moves and the largest distances measured between the replication origins in a replication unit. Yet the S phase usually lasts for about eight hours in a mammalian cell. This implies that the replication origins are not all activated simultaneously and that the DNA in each replication unit (which contains a cluster of perhaps 20 to 80 replication origins) is replicating for only a small part of the total S-phase interval.

Are different replication units activated at random, or is there a specified order in which different regions of the genome are replicated? One way to answer this question is to use the thymidine analogue bromodeoxyuridine (BrdU) to label the newly synthesized DNA in synchronized cell populations for different short periods throughout the S phase. Later, in M phase, those regions of the mitotic chromosomes that have incorporated BrdU into their DNA can be recognized by their altered staining properties or by means of anti-BrdU antibodies. The results show that different regions of each chromosome are replicated in a reproducible order during S phase (Figure 8–37). Moreover, as one would expect from the clusters of replication forks seen in DNA autoradiographs, the timing of replication is coordinated over large regions of the chromosome.

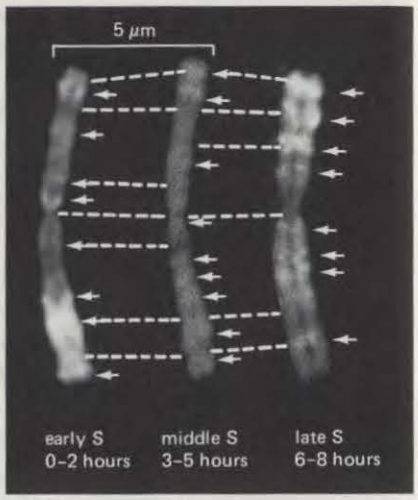


Figure 8–37 Different regions of a chromosome are replicated at different times in S phase. These light micrographs show stained mitotic chromosomes in which the replicating DNA has been differentially labeled during different defined intervals of the preceding S phase. In these experiments cells were first grown in the presence of BrdU (a thymidine analogue) and in the absence of thymidine to label the DNA uniformly. The cells were then briefly pulsed with thymidine in the absence of BrdU during early, middle, or late S phase. Because the DNA made during the thymidine pulse is a double helix with thymidine on one strand and BrdU on the other, it stains more darkly than the remaining DNA (which has BrdU on both strands) and shows up as a bright band (*arrows*) on these negatives. Dashed lines connect corresponding positions on the three copies of the chromosome shown. (Courtesy of Elton Stubblefield.)

Highly Condensed Chromatin Replicates Late, While Genes in Active Chromatin Replicate Early [23]

It seems that the order in which replication origins are activated, at least in part, depends on the chromatin structure in which the origins reside. We have seen, for example, that heterochromatin remains in a highly condensed conformation (similar to that at mitosis) during interphase, while active chromatin has an especially decondensed conformation, which is apparently required to allow RNA synthesis. Heterochromatin is replicated very late in the S phase, suggesting that the timing of replication is related to the packing of the DNA in chromatin. This suggestion is supported by the timing of replication of the two X chromosomes in a female mammalian cell. While these two chromosomes contain essentially the same DNA sequences, one is active for DNA transcription and the other is not (discussed in Chapter 9). Nearly all of the inactive X chromosome is condensed into heterochromatin and its DNA replicates late in the S phase, whereas its active homologue is less condensed and replicates throughout the S phase. These findings support the hypothesis that those regions of the genome whose chromatin is least condensed during interphase, and therefore most accessible to the replication machinery, are replicated first. Autoradiography shows that replication forks move at comparable rates throughout the S phase, so that the extent of chromosome condensation appears to influence the initiation of replication forks but not their speed once formed.

The suggested relationship between chromatin structure and the time of DNA replication is also supported by studies in which the replication times of

Sarepta Exhibit 1056, page 74



Figure 8–38 Large numbers of replication origins operate in rapidly dividing embryonic nuclei. This electron micrograph of spread chromatin from an early *Drosophila* embryo shows that replication bubbles (*arrows*) are extremely closely spaced. Only about 10 minutes elapse between some of the successive nuclear divisions in this embryo. Since the replication origins used here are so closely spaced (separated by a few thousand nucleotide pairs), it should require only about a minute to replicate all the DNA between them. (Courtesy of Victoria Foe.)

specific genes are measured. The results show that so-called "housekeeping" genes, which are those active in all cells, replicate very early in S phase in all cells tested. Genes that are active in only a few cell types, in contrast, generally replicate early in the cells in which the genes are active and later in other types of cells.

The Late-replicating Replication Units Coincide with the A-T-rich Bands on Metaphase Chromosomes [24]

Many of the replication units seem to correspond to distinct chromosome bands made visible by the various fixation and staining procedures used for karyotyping. As discussed previously, as many as 2000 dark-staining *A-T-rich bands (G bands)* can be detected early in mitosis in the haploid set of mammalian chromosomes, and these are separated by an equal number of light-staining *G-C-rich bands (R bands)*. It is intriguing that the A-T-rich DNA and the G-C-rich DNA differ in the time of their replication during S phase. Experiments like the one shown in Figure 8–37 suggest that most G-C-rich bands replicate during the first half of S phase, while most A-T-rich bands replicate during the second half of S phase. It has therefore been suggested that housekeeping genes are located mostly in G-C-rich bands, while many cell-type-specific genes—the vast majority of which will be inactive in most cells—are located in A-T-rich bands. As stated earlier, it is a complete mystery why the mammalian genome should be segregated into such large alternating blocks of chromatin—many nearly equal in size to an entire bacterial genome. It is also not known how the many replication origins present in each replication unit are activated all at once. Perhaps the chromatin in a late-replicating unit remains condensed even after the end of M phase and decondenses only in mid S phase, making all the replication origins in the unit simultaneously accessible. If this is the case, the all-or-none coordinated replication of the DNA in a single replication unit could reflect the stepwise decondensation of large chromatin domains.

The Controlled Timing of DNA Replication May Contribute to Cell Memory [25]

The S phase is completed extremely rapidly in the cleaving eggs of many species, where large stores of chromatin components (such as histones) are present, as required for the rapid manufacture of new nuclei. As illustrated in Figure 8–38, a short S phase also requires the use of an exceptionally large number of replication origins spaced at intervals of only a few thousand nucleotide pairs (rather than the tens or hundreds of thousands of nucleotide pairs found between the replication origins later in development). Since any foreign DNA injected into a fertilized frog egg is replicated, a specific DNA sequence may not be required to form a replication origin in this cell.

Thus, DNA replication can be viewed as a potentially very rapid process that in most cells is subject to a complex system of regulation that restrains the initiation of replication forks in a way that causes different portions of the genome to replicate at very different times. It has been suggested, for example, that the

Sarepta Exhibit 1056, page 75

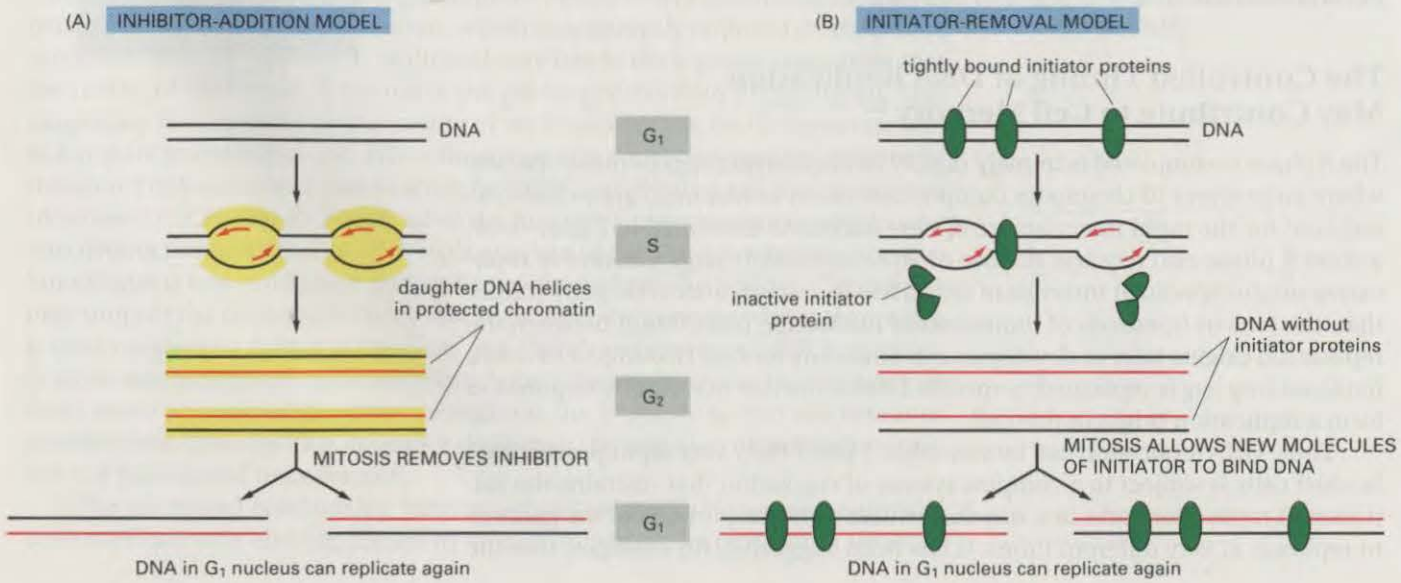
chromatin that replicates early is assembled in part from a special store of chromosomal proteins produced during the G_1 phase and that once a chromosomal region becomes decondensed into active chromatin, its early replication causes the region to pick up proteins that help it to maintain a decondensed state from one cell generation to the next. In this view, the timing of DNA replication contributes to a cellular memory process that facilitates the continued transcription of expressed genes.

Chromatin-bound Factors Ensure That Each Region of the DNA Is Replicated Only Once [26]

In a normal S phase the whole genome must be replicated exactly once and no more. As we have just seen, DNA replication in most eucaryotic cells is an asynchronous process that takes a relatively long time to complete. Because the replication origins are used at different times in different chromosomal regions, in the middle of S phase some parts of a chromosome will not yet have begun replication, while other parts will have replicated completely. An enormous "bookkeeping" problem consequently arises during the middle and late stages of the S phase. Those replication origins already used have been duplicated and, at least with respect to their DNA sequences, are presumably identical to other replication origins not yet used. But each replication origin must be used only once in each S phase. How is this accomplished?

Cell-fusion experiments have provided an important clue. When an S-phase cell is fused with a G_1-phase cell (which has not yet begun S phase), DNA synthesis is induced in the G_1-phase nucleus, suggesting that the transition from G_1 to S phase is mediated by a diffusible activator of DNA synthesis. In contrast, when the S-phase cell is fused with a G_2-phase cell (that is, a cell that has just completed S phase), the G_2 nucleus is not stimulated to synthesize DNA. Since DNA synthesis continues undisturbed in the S-phase nucleus, this implies that the G_2 nucleus is prevented from entering further rounds of replication because every part of its DNA is somehow blocked. Some nondiffusible inhibitor of replication, for example, may have become tightly bound to its DNA. Such an inhibitor, if applied locally during S phase in the wake of each replication fork, would neatly solve the bookkeeping problem: by modifying the chromatin of freshly replicated DNA, it would ensure that once replicated DNA is not replicated again in the same S period (Figure 8–39A). An equally plausible alternative is that tightly bound initiator proteins, or "licensing factors," required for replication are de-

Figure 8–39 Two possible mechanisms to explain the "re-replication block." This block, which protects replicated DNA from further replication in the same cell cycle, is crucial to replication bookkeeping, but its molecular nature is not known. Normally, the block is removed at mitosis, but in a few specialized types of cells (the salivary gland cells of *Drosophila* larvae, for example), it is removed without mitosis, leading to the formation of giant *polytene* chromosomes. (A) A model based on the addition of an inhibitor to all newly replicated chromatin; (B) a model based on tightly bound initiator proteins, or "licensing factors," that act only once and that can be added to the DNA only during mitosis.



Sarepta Exhibit 1056, page 76

stroyed or inactivated by the passage of a replication fork (Figure 8–39B). Whatever its nature, the DNA re-replication block must be removed at or near the time of mitosis, since after cell division the DNA in the G_1 nuclei that emerges in the daughter cells is no longer protected.

The fragments of bacterial DNA that replicate when injected into a fertilized frog egg can be shown to be affected by the re-replication block. Therefore, the mechanism responsible for the block cannot require a highly specific replication origin. The block does not affect the SV40 virus, presumably because the T-antigen made by the virus supplies both initiator and DNA helicase functions for replication, substituting for analogous host cell components and thereby evading the controls to which they are subject.

New Histones Are Assembled into Chromatin as DNA Replicates [27]

A large amount of new histone, approximately equal in mass to the newly synthesized DNA, is required to make new chromatin in each cell cycle. For this reason most organisms possess multiple copies of the gene for each histone. Vertebrate cells, for example, have about 20 repeated sets, each set containing all five histone genes.

Unlike most proteins, which are made continuously throughout interphase, the histones are synthesized mainly in the S phase, when the level of histone mRNA increases about 50-fold as a result of both increased transcription and decreased mRNA degradation. By a mechanism that depends on special properties of their 3′ ends (discussed in Chapter 9), the major histone mRNAs become highly unstable and are degraded within minutes when DNA synthesis stops at the end of S phase (or when inhibitors are added to stop DNA synthesis prematurely). In contrast, the histone proteins themselves are remarkably stable and may survive for the entire life of a cell. The tight linkage between DNA synthesis and histone synthesis may be due, at least in part, to a feedback mechanism that monitors the level of free histone to ensure that the amount of histone made is appropriate for the amount of new DNA synthesized.

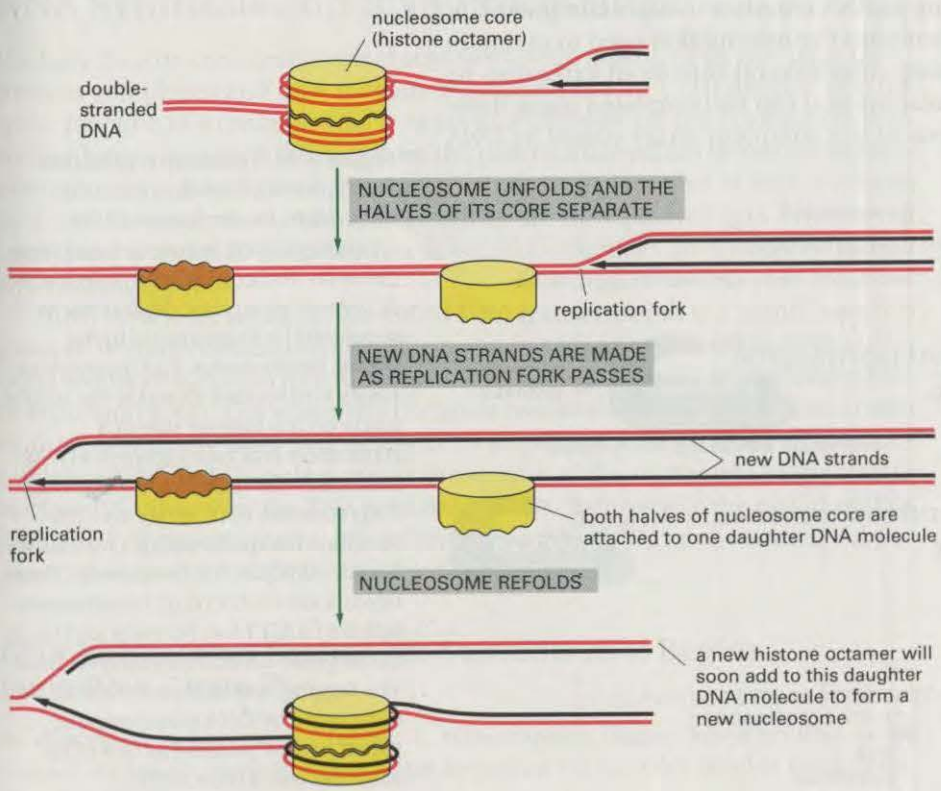


Figure 8–40 Speculative model showing how a nucleosome might open up to permit DNA replication. After the replication fork passes, the nucleosome reassembles. In this way the histones of the nucleosome core remain permanently bound to the DNA. Although in this diagram the old nucleosome has been inherited intact by the DNA helix made on the leading strand, there is evidence that an intact nucleosome can be inherited by either daughter DNA molecule. Moreover, this is only one of many possible models.

Sarepta Exhibit 1056, page 77

Once they are assembled in nucleosomes, histone molecules rarely, if ever, leave the DNA to which they are bound. As a replication fork advances, therefore, it must somehow pass through the parental nucleosomes, which seem to have been designed to allow both transcription and replication to proceed past them. According to one hypothesis, each nucleosome transiently unfolds into two half-nucleosomes during DNA replication, thereby allowing the DNA polymerase to copy the uncoiled nucleosomal DNA (Figure 8–40).

The newly synthesized DNA behind a replication fork inherits some old histones, but since the amount of DNA has doubled, it also needs to bind an equal amount of new histones to complete its packaging into chromatin (see Figure 8–40). There is some evidence to suggest that a special nucleosome assembler may travel with the replication fork, packaging the newly synthesized DNA as soon as it emerges from the replication machinery. This newly formed chromatin, however, requires as much as an hour to become fully mature; until then, for example, the new histones are more susceptible than usual to histone-modifying enzymes. The chemical differences between mature and immature chromatin are unknown, although it is thought that maturation involves both covalent modifications of histones and changes in the binding of other proteins to the chromatin.

Telomeres Consist of Short G-rich Repeats That Are Added to Chromosome Ends by Telomerase [28]

As discussed earlier, it is not possible for DNA polymerase to replicate the end of a linear DNA molecule completely in the ordinary way, and this has led to the evolution of special DNA sequences, called **telomeres,** at the ends of eucaryotic chromosomes. These sequences, which are similar in organisms as diverse as protozoa, fungi, plants, and mammals, consist of many tandem repeats of a short sequence that contains a block of neighboring G nucleotides. In humans this sequence is GGGTTA.

The problem of replicating the ends of chromosomes is solved in an ingenious way by an enzyme called **telomerase.** This enzyme recognizes the G-rich strand of an existing telomere repeat sequence and elongates it in the 5′-to-3′ direction. In the absence of a complementary DNA strand, the telomerase synthesizes a new copy of the repeat using an RNA template that is a component of the enzyme itself. The enzyme thus contains the information used to maintain the characteristic telomere sequences. After several rounds of extension by telomerase, replication of the chromosome end can be completed using these extensions as a template for synthesis of the complementary strand by DNA

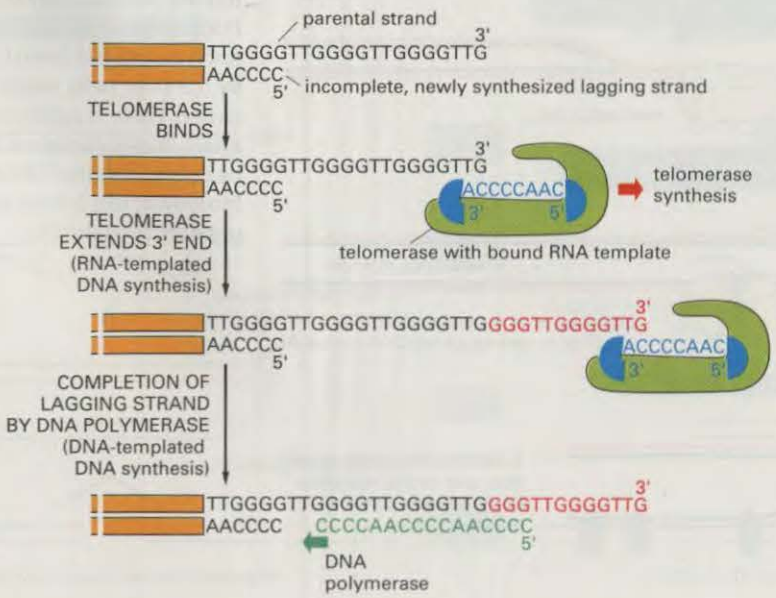


Figure 8–41 Telomere replication. The figure outlines the reactions involved in the formation of the repeating G-rich sequences that form the ends of chromosomes (telomeres) of diverse eucaryotic organisms, as suggested by experiments in the ciliate *Tetrahymena.* The incomplete, newly synthesized strand is the strand made on the lagging side of a replication fork (see Figure 6–41). As indicated, the telomerase is a protein-RNA complex that carries an RNA template for synthesizing a repeating, G-rich telomere DNA sequence. These repeats are GGGTTG in *Tetrahymena* but are GGGTTA in humans and $G_{1\text{-}3}A$ in the yeast *Saccharomyces cerevisiae.* The lagging strand is presumed to be completed by DNA polymerase α, which carries primase as one of its subunits (see Figure 8–35).

Sarepta Exhibit 1056, page 78

polymerase (Figure 8–41). Because the processes that shorten and restore the telomere sequence are only approximately balanced, each chromosome end contains a variable number of the tandem repeats, which generally extend for hundreds of nucleotide pairs.

Summary

In vitro studies using the monkey virus SV40 as a model system suggest that, in eucaryotes as in procaryotes, DNA replication begins with the loading of a DNA helicase onto the DNA by an initiator protein bound to a replication origin. A replication bubble forms at such an origin as two replication forks move away from each other. During S phase in higher eucaryotes, neighboring replication origins appear to be activated in clusters known as replication units, with the origins spaced an average of about 100,000 nucleotide pairs apart. Since the replication fork moves at about 50 nucleotides per second, only about an hour should be required to complete the DNA synthesis in a replication unit. Throughout a typical 8-hour S phase different replication units are activated in a sequence determined in part by the structure of the chromatin, the most condensed regions of chromatin being replicated last. The correspondence between replication units and the bands containing millions of nucleotide pairs seen on mitotic eucaryotic chromosomes suggests that replication units may correspond to structurally distinct domains in interphase chromatin.

After the replication fork passes, chromatin structure is re-formed by the addition of new histones and other chromosomal proteins to the old histones inherited on the daughter DNA molecules. A DNA re-replication block of unknown nature acts locally to prevent a second round of replication from occurring until a chromosome has passed through mitosis; this block is needed to ensure that each region of the DNA is replicated only once in each S phase.

The problem of replicating the ends of chromosomes is solved by a specialized end structure (the telomere) and an enzyme (telomerase) that extends this structure using an RNA template that is part of the telomerase.

RNA Synthesis and RNA Processing

We have thus far considered how chromosomes are organized as very large DNA-protein complexes and how they are duplicated before a cell divides. But the main function of a chromosome is to act as a template for the synthesis of RNA molecules, since only in this way does the genetic information stored in chromosomes become directly useful to the cell. There is a great deal of RNA synthesis in a cell: the total rate at which nucleotides are incorporated into RNA during interphase is about 20 times the rate at which nucleotides are incorporated into DNA during S phase.

RNA synthesis, which is also called **DNA transcription,** is a highly selective process. In most mammalian cells, for example, only about 1% of the DNA nucleotide sequence is copied into functional RNA sequences (mature messenger RNA or structural RNA). The selectivity occurs at two levels, which we discuss in turn in this section: (1) only part of the DNA sequence is transcribed to produce nuclear RNAs, and (2) only a minor proportion of the nucleotide sequences in nuclear RNAs survives the RNA processing steps that precede the export of RNA molecules to the cytoplasm. We begin by describing *RNA polymerases,* the enzymes that catalyze all DNA transcription.

RNA Polymerase Exchanges Subunits as It Begins Each RNA Chain [29]

As described in outline in Chapter 6, transcription begins when an RNA polymerase molecule binds to a **promoter** sequence on the DNA double helix. Next,

Sarepta Exhibit 1056, page 79

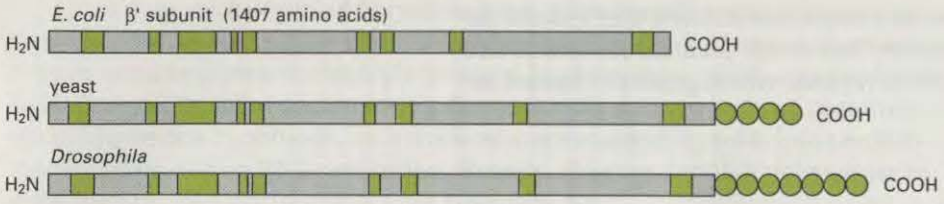


Figure 8–42 The common origin of bacterial and eucaryotic RNA polymerases. Amino acid sequence similarities between the β′ subunit of *E. coli* RNA polymerase and the largest subunit of eucaryotic RNA polymerase II are among the comparisons that reveal a common evolutionary origin for the bacterial and eucaryotic enzymes. The β′ subunit is thought to bind to DNA. The regions of the sequence shown as *green bars* are more than 70% identical between yeast and *Drosophila* and more than 40% identical between *Drosophila* and *E. coli*. A uniquely eucaryotic sequence (of seven amino acid residues) is repeated 26 times at the carboxyl terminus of the yeast subunit and more than 40 times in the *Drosophila* subunit (indicated here by *green circles*); these repeats are phosphorylated as part of the process that starts an RNA chain in eucaryotes (see Figure 9–30). (After A.L. Greenleaf et al., in RNA Polymerase and the Regulation of Transcription [W.S. Reznikoff et al., eds.], pp. 459–464. New York: Elsevier, 1987.)

in a step that is not well understood, the two strands of the DNA are separated locally to form an *open complex*. At this stage the template strand is exposed, and synthesis of the complementary RNA chain can begin. The polymerase then moves along the template strand, extending its growing RNA chain in the 5′-to-3′ direction by the stepwise addition of ribonucleoside triphosphates until it reaches a *stop (termination) signal*, at which point the newly synthesized RNA chain and the polymerase are released from the DNA. Each RNA molecule thus represents a single-strand copy of the nucleotide sequence of one DNA strand in a relatively short region of the genome (see Figure 6–2). This transcribed segment of DNA is called a *transcription unit*.

RNA polymerases are generally formed from multiple polypeptide chains and have masses of 500,000 daltons or more. The enzymes in bacteria and eucaryotes are evolutionarily related (Figure 8–42). Since the bacterial enzyme has been far easier to study, its properties provide a basis for understanding its eucaryotic relatives. The *E. coli* enzyme contains four different subunits, α, β, β′, and σ, there being two copies of α and one each of the others. The complete amino acid sequence of each subunit has been determined from the nucleotide sequence of its gene.

The sigma (σ) subunit of the *E. coli* polymerase has a specific role in the initiation of transcription: it enables the enzyme to find promoter sequences to which it binds. After about eight nucleotides of an RNA molecule have been synthesized (step 4 in Figure 8–43), the σ subunit dissociates and a number of *elongation factors*—important for chain elongation and termination—become associated with the enzyme instead. The elongation factors include several proteins that function in ways that are incompletely understood. The initiation of transcription is an important control point where the cell can regulate the expression of a gene; for this reason it is discussed in more detail in Chapter 9.

Three Kinds of RNA Polymerase Make RNA in Eucaryotes [30]

Although the mechanism of DNA transcription is similar in eucaryotes and procaryotes such as *E. coli*, the machinery is considerably more complex in eu-

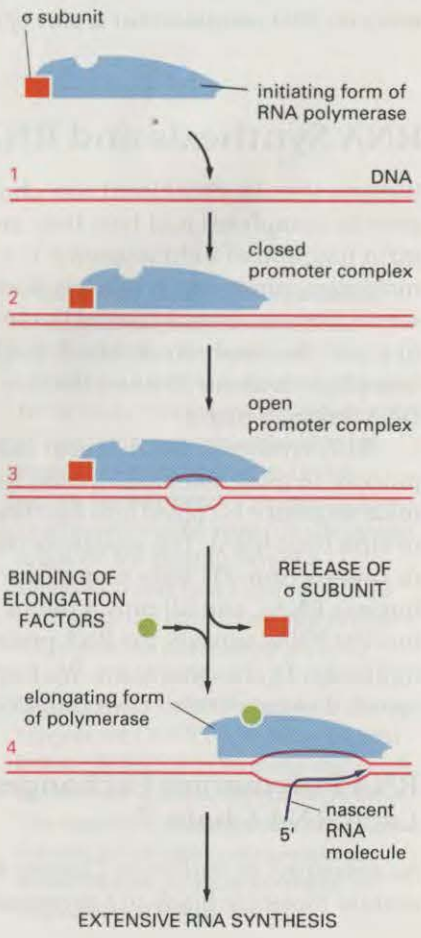


Figure 8–43 Schematic diagram of the steps in the initiation of RNA synthesis (DNA transcription) catalyzed by RNA polymerase. The steps indicated have been revealed by studies of the *E. coli* enzyme. A DNA molecule containing a promoter sequence for the *E. coli* polymerase is shown (see Figure 6–2). The enzyme first forms a *closed complex* in which the two DNA strands remain fully base-paired. In the next step the enzyme catalyzes the opening of a little more than one turn of the DNA helix to form an *open complex*, in which the template DNA strand is exposed for the initiation of an RNA chain. The polymerase containing the bound σ subunit, however, behaves as though it is tethered to the promoter site: it seems unable to proceed with the elongation of the RNA chain and on its own frequently synthesizes and releases short RNA chains. As indicated, the conversion to an actively elongating polymerase requires the release of initiation factors (the σ subunit in the case of the *E. coli* enzyme) and generally involves the binding of other proteins that serve as elongation factors.

Sarepta Exhibit 1056, page 80

caryotes. In eucaryotes as diverse as yeasts and humans, for example, there are three types of RNA polymerases, each responsible for transcribing different sets of genes. These enzymes—denoted as *RNA polymerases I, II, and III*—are structurally similar to one another and have some common subunits, although other subunits are unique. Each is more complex than *E. coli* RNA polymerase and is thought to contain 10 or more polypeptide chains. Another important distinction between the bacterial and eucaryotic enzymes is that, whereas the purified bacterial enzyme can bind to promoters and initiate transcription on its own, the eucaryotic enzymes require the presence of additional initiation proteins that must bind to the promoter before the enzyme can bind. For this reason it was not until 1979 that systems with all the needed components became available so that eucaryotic initiation mechanisms could be analyzed *in vitro*. Because these initiation proteins and their interactions with the polymerases are intimately involved with the control of transcription initiation, we shall defer discussion of them until Chapter 9.

The three eucaryotic RNA polymerases were initially distinguished by their chemical differences during purification and by their sensitivity to *α-amanitin*, a poison isolated from the deadly toadstool *Amanita phalloides*. RNA polymerase I is unaffected by α-amanitin; RNA polymerase II is very sensitive to this poison; and RNA polymerase III is moderately sensitive to it. The sensitivity of RNA synthesis to α-amanitin is still used to determine which polymerase transcribes a gene. Such studies indicate that **RNA polymerase II** transcribes the genes whose RNAs will be translated into proteins. The other two polymerases synthesize RNAs that have structural or catalytic roles, chiefly as part of the protein synthetic machinery: **polymerase I** makes the large ribosomal RNAs, and **polymerase III** makes a variety of very small, stable RNAs—including the small 5S ribosomal RNA and the transfer RNAs. However, most of the small RNAs that form snRNPs, which we discuss later when we consider RNA processing, are made by polymerase II.

Mammalian cells typically contain 20,000 to 40,000 molecules of each of the RNA polymerases, and studies with cultured cells indicate that the concentrations of these enzymes are regulated individually according to the rate of cell growth.

RNA Polymerase II Transcribes Some DNA Sequences Much More Often Than Others [31]

Because RNA polymerase II makes all of the mRNA precursors and thus determines which proteins a cell will make, we shall focus most of our discussion on its activities and on the fate of its products. Although experiments with purified polymerases *in vitro* are essential for establishing the mechanism of transcription, much can also be learned about how the process occurs in a cell by using the electron microscope to examine genes in action, with their bound RNA polymerases caught in the act of transcription.

Ordinary thin-section electron micrographs of interphase nuclei show granular clumps of chromatin (see Figure 8–71) but reveal very little about how genes are transcribed. A much more detailed picture emerges if the nucleus is ruptured and its contents spilled out onto an electron microscope grid (Figures 8–44 and 8–45). At the farthest point from the center of the lysed nucleus, the chromatin is diluted sufficiently to make individual chromatin strands visible in the expanded, beads-on-a-string form shown previously in Figure 8–9B.

RNA polymerase molecules actively engaged in transcription appear as globular particles with a single RNA molecule trailing behind. Particles representing active RNA polymerase II molecules are usually seen as single units, without nearby neighbors. This observation indicates that most genes are transcribed into mRNA precursors only infrequently, so that one polymerase finishes transcription before another one begins. Occasionally, however, many polymerase molecules (and their associated RNA transcripts) are seen clustered together. These clusters occur on the relatively few genes that are transcribed at high frequency



Figure 8–44 A typical cell nucleus visualized by electron microscopy using the procedure shown in Figure 8–45. An enormous tangle of chromatin can be seen spilling out of the lysed nucleus; only the chromatin at the outermost edge of this tangle will be sufficiently dilute for meaningful examination at higher power. (Courtesy of Victoria Foe.)

Sarepta Exhibit 1056, page 81

(Figure 8–46). The length of the attached RNA molecules in such a cluster increases in the direction of transcription, producing a characteristic pattern. This pattern defines the RNA polymerase II start site and direction of transcription for a specific transcription unit (Figure 8–47).

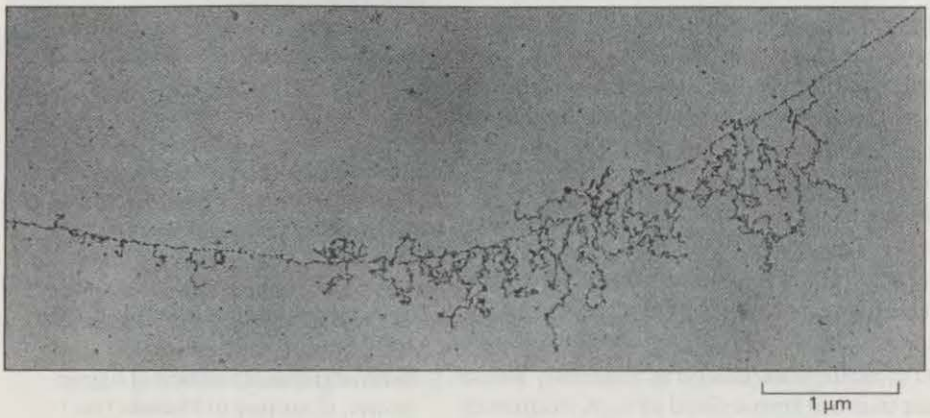
Biochemical studies have confirmed and extended the results obtained by electron microscopy, leading to three major conclusions:

1. Eucaryotic RNA polymerase molecules, like those in procaryotes, begin at specific sites on the chromosome.
2. The average length of the complete RNA molecule produced by RNA polymerase II from a single transcription unit is about 7000 nucleotides, and RNA molecules 10,000 to 20,000 nucleotides long are common. These lengths, which are much longer than the 1200 nucleotides of RNA needed to code for an average protein of 400 amino acid residues, reflect the complex structure of eucaryotic genes and, in particular, the presence of long intron sequences, which, as we discuss later, are later removed from the RNA.
3. Although chain *elongation rates* of about 30 nucleotides per second are observed for all RNAs, different RNA polymerase II start sites have different *initiation frequencies,* so that some genes are transcribed at much higher rates than others. As indicated in Table 8–2, the majority of the genes that are transcribed give rise to very few mRNA molecules.

The Precursors of Messenger RNA Are Covalently Modified at Both Ends [32]

In eucaryotes mature mRNA is produced in several steps. The RNA molecules freshly synthesized by RNA polymerase II in the nucleus are known as *primary transcripts;* the collection of such transcripts was originally called **heterogeneous nuclear RNA (hnRNA)** because of the large variation in RNA size, contrasting with the more uniform and smaller size of the RNA sequences actually needed to encode proteins. We shall see shortly that much of this variation is due to the presence of long intron sequences in the primary transcripts. As they are being synthesized, these transcripts are covalently modified at both their 5′ end and their 3′ end in ways that clearly distinguish them from transcripts made by other RNA polymerases. These modifications will be used later in the cytoplasm as signals that these transcripts are to be translated into protein.

The 5′ end of the RNA molecule (which is the end synthesized first during transcription) is first *capped* by the addition of a methylated G nucleotide. Capping occurs almost immediately, after about 30 nucleotides of RNA have been synthesized, and it involves condensation of the triphosphate group of a molecule of GTP with a diphosphate left at the 5′ end of the initial transcript (Figure 8–48). This **5′ cap** will later play an important part in the initiation of protein synthesis; it also seems to protect the growing RNA transcript from degradation.



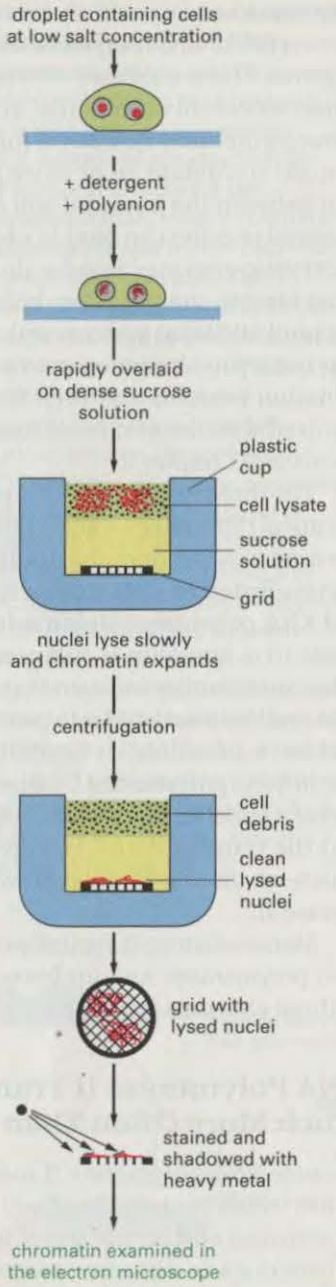


Figure 8–45 A method for examining chromatin in the electron microscope. The nuclei are first lysed, and then the chromatin is freed from cellular debris and spread out on a grid.

Figure 8–46 Electron micrograph of a region of chromatin containing a gene being transcribed at unusually high frequency. Many RNA polymerase II molecules with their growing RNA transcripts are visible. The direction of transcription is from left to right (see Figure 8–47). (From V.E. Foe, L.E. Wilkinson, and C.D. Laird, *Cell* 9:131–146, 1976. © 1976 Cell Press.)

Sarepta Exhibit 1056, page 82

Table 8–2 The Population of mRNA Molecules in a Typical Mammalian Cell

	Copies per Cell of Each mRNA Sequence		Number of Different mRNA Sequences in Each Class		Total Number of mRNA Molecules in Each Class
Abundant class	12,000	×	4	=	48,000
Intermediate class	300	×	500	=	150,000
Scarce class	15	×	11,000	=	165,000

This division of mRNAs into just three discrete classes is somewhat arbitrary, and in many cells a more continuous spread in abundances is seen. However, a total of 10,000 to 20,000 different mRNA species is normally observed in each cell, most species being present at a low level (5 to 15 molecules per cell). Most of the total cytoplasmic RNA is rRNA, and only 3% to 5% is mRNA, a ratio consistent with the presence of about 10 ribosomes per mRNA molecule. This particular cell type contains a total of about 360,000 mRNA molecules in its cytoplasm.

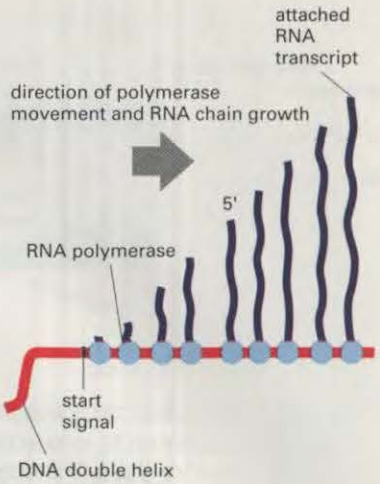


Figure 8–47 An idealized transcription unit. The drawing illustrates how the electron microscope appearance (see Figure 8–46) demonstrates the direction of transcription, as well as the start site of the unit.

The 3′ end of most polymerase II transcripts is defined not by the termination of transcription but by a second modification in which the growing transcript is cleaved at a specific site and a **poly-A tail** is added by a separate polymerase to the cut 3′ end. The signal for the cleavage is the appearance in the RNA chain of the sequence AAUAAA located 10 to 30 nucleotides upstream from the site of cleavage, plus a less well-defined downstream sequence. Immediately after cleavage, a *poly-A polymerase* enzyme adds 100 to 200 residues of adenylic acid (as *poly A*) to the 3′ end of the RNA chain to complete the **primary RNA transcript.** Meanwhile, the polymerase fruitlessly continues transcribing for hundreds or thousands of nucleotides until termination occurs at one of several later sites; the extra piece of functionless RNA transcript thus generated presumably lacks a 5′ cap and is rapidly degraded (Figure 8–49).

The poly-A tail appears to have several functions: (1) as described later, it aids in the export of mature mRNA from the nucleus; (2) it is thought to affect the stability of at least some mRNAs in the cytoplasm; and (3) it seems to serve as a recognition signal for the ribosome that is required for efficient translation of mRNA. The latter feature—in combination with the 5′ cap—would enable a ribosome to determine whether the mRNA was intact before expending energy and precursors to begin its translation.

Even though polymerase II transcripts comprise more than half of the RNA being synthesized by DNA transcription, we shall see below that most of the RNA in these transcripts is unstable and therefore short-lived. Consequently, the hnRNA in the cell nucleus and the cytoplasmic mRNA derived from it constitute only a minor fraction of the total RNA in a cell (Table 8–3). Despite their relative scarcity, these RNA molecules can be readily purified because of their poly-A tails. When the total cellular RNA is passed through a column containing poly dT linked to a solid support, the complementary base-pairing between T and A residues selectively binds the molecules with poly-A tails to the column; the bound molecules can then be released for further analysis. This procedure is widely used to separate the hnRNA and mRNA molecules from the ribosomal and transfer RNA molecules that predominate in cells.

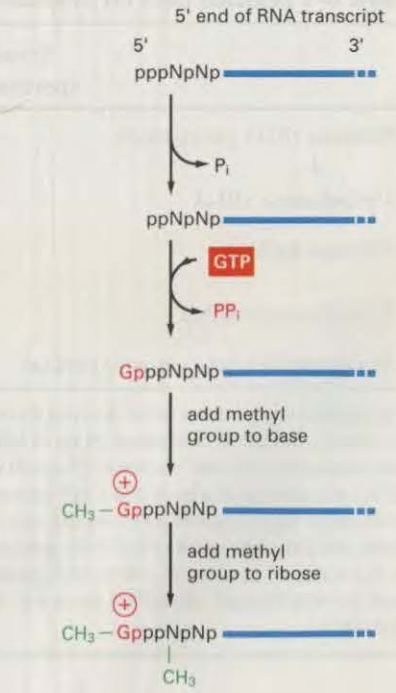


Figure 8–48 The reactions that cap the 5′ end of each RNA molecule synthesized by RNA polymerase II. The final cap contains a novel 5′-to-5′ linkage between the positively charged 7-methyl G residue and the 5′ end of the RNA transcript (see Figure 6–26). At least some of the enzymes required for this process are thought to be bound to polymerase II, since polymerase I and III transcripts are not capped and the indicated reaction occurs almost immediately following initiation of each RNA chain. The letter N is used here to represent any one of the four ribonucleotides, although the nucleotide that starts an RNA chain is usually a purine (an A or a G). (After A.J. Shatkin, *Bioessays* 7:275–277, 1987. © ICSU Press.)

Sarepta Exhibit 1056, page 83

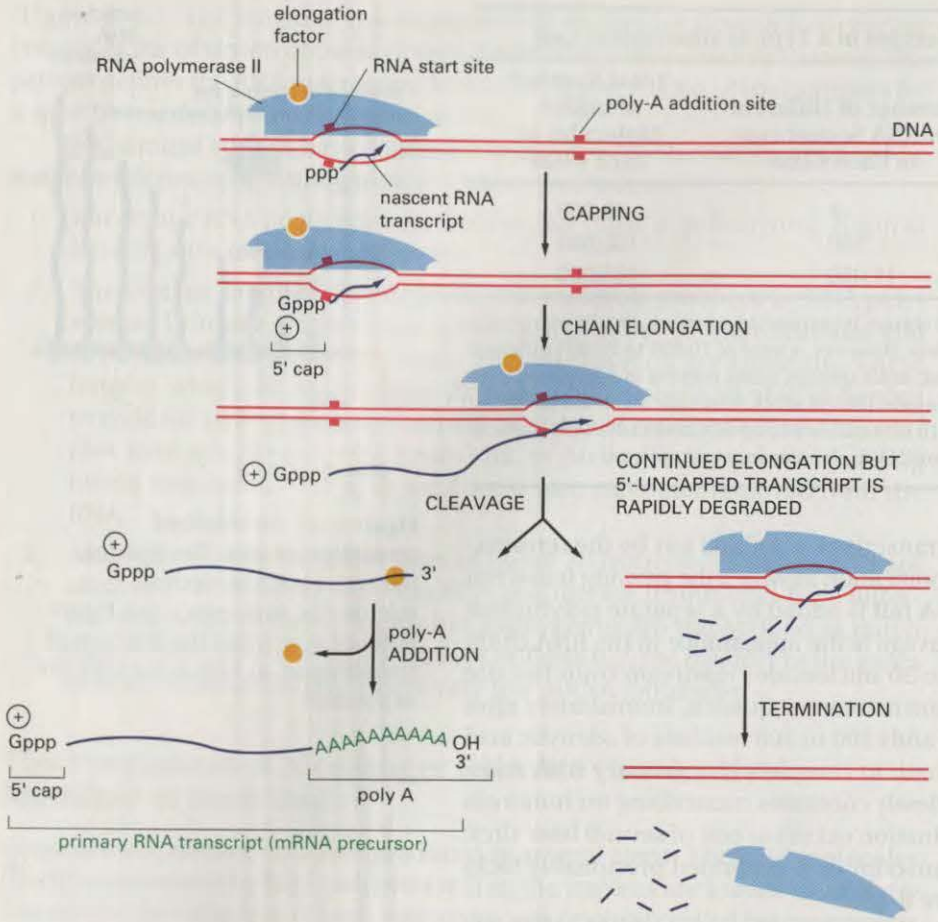


Figure 8–49 Synthesis of a primary RNA transcript (an mRNA precursor) by RNA polymerase II. This diagram starts with a polymerase that has just begun synthesizing an RNA chain (step 4 of Figure 8–43). Recognition of a poly-A addition signal in the growing RNA transcript causes the chain to be cleaved and then polyadenylated as shown. In yeasts the polymerase terminates its RNA synthesis almost immediately thereafter, but in higher eucaryotes it often continues transcription for thousands of nucleotides. It seems likely that the polymerase changes its properties once one RNA chain cleavage has occurred. Thus it cannot trigger poly-A addition to its downstream RNA transcript, and it seems to have a greater probability of responding to the sequences that cause chain termination and polymerase release. The simplest hypothesis, represented here, is that an elongation factor (or factors) is released from the polymerase after cleavage of the transcript.

Table 8–3 Selected Data on Amounts of RNA in a Typical Mammalian Cell

	Steady-State Amount (percent of total cell RNA)	Percent of Total RNA Synthesis
Nuclear rRNA precursors	4	39
↓		
Cytoplasmic rRNA	71	—
Nuclear hnRNA	7	58
↓		
Cytoplasmic mRNA	3	—
Small stable RNAs (mostly tRNAs)	15	3

The figures shown here were derived from the analysis of a mouse fibroblast cell line (L cells) in culture. Each cell contained 26 pg of RNA (5×10^{10} nucleotides of RNA), of which about 14% was located in the cell nucleus. (The cell nucleus thus contains about twice as much DNA as RNA.) An average of about 200×10^6 nucleotides is polymerized into RNA every minute during interphase. This is about 20 times the average rate at which DNA is synthesized during S phase. Note that although most of the RNA synthesized is hnRNA, most of this RNA is rapidly degraded in the nucleus. As a result, the mRNA produced from the hnRNA is only a minor fraction of the total RNA in the cell. (Modified from B.P. Brandhorst and E.H. McConkey, *J. Mol. Biol.* 85:451–563, 1974.)

Sarepta Exhibit 1056, page 84

Only RNA polymerase II transcripts have 5′ caps and 3′ poly-A tails. This seems to be because the capping and cleavage plus poly-A addition reactions are mediated by enzymes that bind selectively to polymerase II. Thus, if a gene that is normally transcribed by polymerase II is separated from its promoter by recombinant DNA methods and fused to a promoter recognized by polymerase I or by polymerase III, the RNA transcripts produced from it by these polymerases are neither capped nor polyadenylated. The requirement for a specific capping and polyadenylation of mRNA precursors may explain why these RNAs are synthesized by a separate type of RNA polymerase molecule in eucaryotes.

RNA Processing Removes Long Nucleotide Sequences from the Middle of RNA Molecules [33]

The discovery of interrupted genes in 1977 was entirely unexpected. Previous studies in bacteria had shown that their genes are composed of a continuous string of the nucleotides needed to encode the amino acids of a protein, and there seemed to be no obvious reason why a gene should be organized in any other way. The first indication that eucaryotic genes are not continuous like bacterial genes came when new methods allowing an accurate comparison of mRNA and DNA sequences were applied to mRNAs produced by a human *adenovirus* (a large DNA virus). The region of the viral DNA producing these RNAs turned out to contain sequences that are not present in the mature RNAs. The possibility that this situation was unique to viruses was quickly eliminated by the finding of similar interruptions in the ovalbumin and β-globin genes of vertebrates. As discussed earlier, the sequences present in the DNA but omitted from the mRNA are called *intron* sequences, while those present in the mRNA are called *exon* sequences (Figures 8–50 and 8–51).

Before the discovery of introns, the significance of hnRNA and its relationship to mRNA had seemed very mysterious. It had long been known that most of the RNA synthesized by RNA polymerase II is rapidly degraded in the nucleus. The hnRNA molecules of cultured cells can be radiolabeled by brief exposure to ^{3}H-uridine and followed over a long period. This sort of experiment showed that the average length of the hnRNA molecules in the labeled population decreases rapidly, starting from about 7000 nucleotides, to reach the size of cytoplasmic mRNA molecules (an average of about 1500 nucleotides) after only about 30 minutes; at about the same time, radioactively labeled RNA molecules begin to leave the nucleus as mRNA molecules. Only about 5% of the mass of the labeled hnRNA ever reaches the cytoplasm, however; the remainder is degraded into small fragments in the nucleus over a period of about an hour. This seemed strangely wasteful. And the puzzle was deepened by the finding that even though the hnRNA molecules became progressively shorter, they retained their 5′ caps and their 3′ poly-A tails.

With the discovery of introns, the explanation became clear: the primary RNA transcript is a faithful copy of the gene, containing both exon and intron sequences, and the latter sequences are cut out of the middle of the RNA transcript to produce an mRNA molecule that codes directly for a protein (see Figure 3–15). Because the coding RNA sequences on either side of an intron sequence are joined to each other after the intron sequence has been cut out, this reaction is known as **RNA splicing.** RNA splicing occurs in the cell nucleus, out of reach of the ribosomes, and RNA is exported to the cytoplasm only when processing is complete.

Because most mammalian genes contain much more intron than exon sequence (see Table 8–1, p. 340), RNA splicing can account for the conversion of the very long nuclear hnRNA molecules to the much shorter cytoplasmic mRNA molecules.

Before discussing the distribution of introns in eucaryotic genes and some of their consequences for cell function, it is necessary to explain how intron sequences are recognized and removed by the splicing machinery.

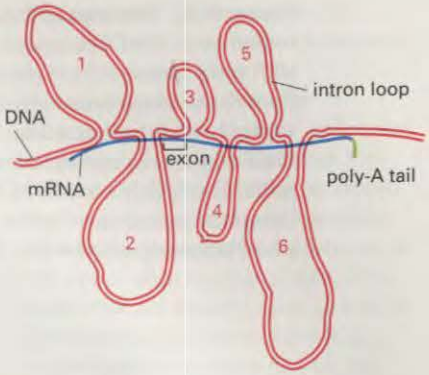
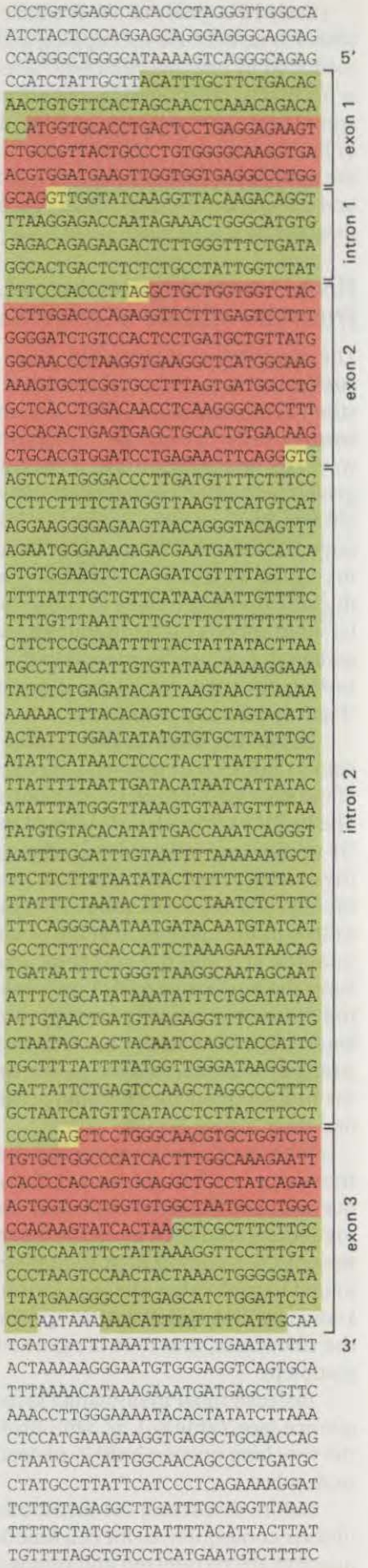


Figure 8–50 Early evidence for the existence of introns in eucaryotic genes. The evidence was provided by the "R-loop technique," in which a base-paired complex between mRNA and DNA molecules is visualized in the electron microscope. An unusually abundant mRNA molecule, such as β-globin mRNA or ovalbumin mRNA, is readily purified from the specialized cells that produce it. When this single-stranded mRNA preparation is annealed in a suitable solvent to a cloned double-stranded DNA molecule containing the gene that encodes the mRNA, the RNA can displace a DNA strand wherever the two sequences match and form regions of RNA-DNA helix. Regions of DNA where no match to the mRNA sequence is possible are clearly visible as large loops of double-stranded DNA. Each of these loops (numbered 1 to 6) represents an intron in the gene sequence.

Sarepta Exhibit 1056, page 85

Figure 8–51 The transcribed portion of the human β-globin gene. The sequence of the DNA strand corresponding to the mRNA sequence is given, with the primary RNA transcript surrounded by a *green line* and the nucleotides in the three exons *shaded yellow*. Note that exon 1 includes a *5′-leader sequence* and that exon 3 includes a *3′-untranslated sequence;* although these sequences are included in the mRNA, they do not code for amino acids. The highly conserved GT and AG nucleotides at the ends of each intron are boxed (see Figure 8–53), along with the cleavage and polyadenylation signal near the 3′ end of the gene (AATAAA, see Figure 8–49).



hnRNA Transcripts Are Immediately Coated with Proteins and snRNPs [34]

Newly made RNA in eucaryotic cells, unlike that in bacteria, appears to become immediately condensed into a string of closely spaced protein-containing particles. Each particle consists of about 500 nucleotides of RNA wrapped around a protein complex that serves to condense and package each growing RNA transcript in a manner reminiscent of the DNA-protein complexes of nucleosomes. The resulting **hnRNP particles** (*heterogeneous nuclear ribonucleoprotein particles*) can be purified after nuclei have been treated with ribonucleases at levels just sufficient to destroy the linker RNA between them. Each particle has a diameter of about 20 nm, which is twice that of a nucleosome, and the protein core is more complex and less well characterized, being composed of a set of at least eight different proteins. Except for histones, the proteins in this core are the most abundant proteins in the cell nucleus. Several of them contain a conserved domain of about 80 amino acids, which is often repeated and shared by many other RNA-binding proteins.

The hnRNP particles are generally distorted by the standard spreading techniques used to view gene transcription in the electron microscope (see Figure 8–45). These micrographs, however, reveal especially stable particles of a less common type, whose position on the RNA strongly implicates them in RNA splicing. These particles form very quickly at specific RNA sequences—at or near the junctions between intron and exon sequences—and, as the RNA transcript elongates, they coalesce in pairs to form a larger assembly that is thought to be the *spliceosome* that catalyzes RNA splicing (Figure 8–52).

Biochemical analysis has revealed that the cell nucleus contains many complexes of proteins with small RNAs (generally RNAs of 250 nucleotides or less), which have arbitrarily been designated U1, U2, . . . , U12 RNAs. These complexes, called **small nuclear ribonucleoproteins** (snRNPs—pronounced "snurps"), resemble ribosomes in that each contains a set of proteins complexed to a stable RNA molecule. They are much smaller than ribosomes, however—only 250,000 daltons compared with 4.5 million daltons for a ribosome. Some proteins are present in several types of snRNPs, whereas others are unique to one type. This was first demonstrated using serum from patients with the autoimmune disease *systemic lupus erythematosus*, who make antibodies directed against one or more of their own snRNP proteins: a single antibody was found that binds the U1, U2, U5, and U4/U6 snRNPs, for example, and we now know that they all contain common proteins.

Individual snRNPs are believed to recognize specific nucleic acid sequences through RNA-RNA base-pairing. Some mediate RNA splicing, one is known to be involved in the cleavage reaction that generates the 3′ ends of newly formed histone RNAs, while the function of others is unknown. The evidence for the role of snRNPs in RNA splicing comes from experiments on RNA splicing *in vitro*, as well as from analyses of yeast cells that are mutant in one of the snRNP components.

Sarepta Exhibit 1056, page 86

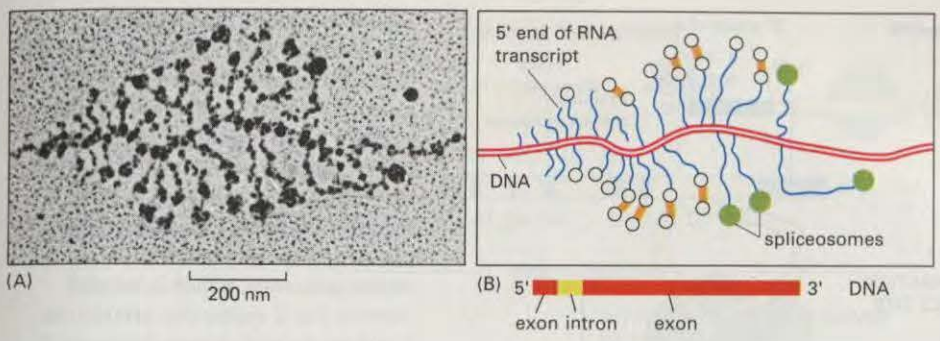


Figure 8–52 Spliceosomes. (A) Electron micrograph of a chromatin spread showing large ribonucleoprotein particles assembling at the 5′ and 3′ splice site regions to form a *spliceosome*. The RNA transcripts are being produced from a gene encoding a *Drosophila* chorion protein, and the positions of the splice sites on the primary RNA transcript are known. As indicated in the drawing (B), most of the RNA transcripts have either one or two large RNP particles near their 5′ ends; a schematic representation of the gene is shown below the drawing. When there are two particles on a transcript [*open circles* in (B)], they average 25 nm in diameter and occur at or very near the positions of the 5′ and 3′ splice sites for the single small intron sequence (228 nucleotides long) near the 5′ end of the transcripts. The more mature, longer transcripts frequently display a single larger particle [*green circles* in (B)] in the region of the intron, which probably results from the stable association of the two smaller particles and represents the assembled spliceosome. Since in some cases (including the example shown in this figure) splicing occurs while the 3′ end of the RNA chain is still being transcribed, the poly A at the 3′ end of hnRNA molecules cannot be required for splicing. The hnRNP proteins have been removed from these transcripts by the spreading conditions used. (Adapted from Y.N. Osheim, O.L. Miller, and A.L. Beyer, *Cell* 43:143–151, 1985. © Cell Press.)

Intron Sequences Are Removed as Lariat-shaped RNA Molecules [35]

Introns range in size from about 80 nucleotides to 10,000 nucleotides or more. Unlike the sequence of an exon, the exact nucleotide sequence of an intron seems to be unimportant. Thus introns have accumulated mutations rapidly during evolution, and it is often possible to alter most of an intron's nucleotide sequence without greatly affecting gene function. This has led to the suggestion that intron sequences have no function at all and are largely genetic "junk," a proposition we shall examine at the end of the chapter. The only highly conserved sequences in introns are those required for intron removal, which are found at or near the ends of an intron and are very similar in all known intron sequences; they generally cannot be altered without affecting the splicing process that normally removes the intron sequence from the primary RNA transcript. These conserved boundary sequences at the **5′ splice site (donor site)** and the **3′ splice site (acceptor site)** of introns from higher eucaryotes are shown in Figure 8–53. The RNA breaking and rejoining reactions must be carried out precisely because an error of even one nucleotide would shift the reading frame in the resulting mRNA molecule and make nonsense of its message.

The pathway by which the intron sequences are removed from primary RNA transcripts has been elucidated by *in vitro* studies in which a pure RNA species containing a single intron is prepared by incubating an appropriately designed DNA fragment with an RNA polymerase (see Figure 7–36). When these RNA molecules are added to a cell extract, they become spliced in a two-step enzymatic reaction that requires prolonged incubation with ATP, the U1, U2, U5, and U4/U6 snRNPs, and a number of additional proteins; these components assemble into a large multicomponent ribonucleoprotein complex, or **spliceosome.** Characterization of the RNA species that appear as intermediates during the reaction, as well as the snRNPs required to produce them, led to the discovery that the intron is excised in the form of a *lariat*, according to the splicing pathway shown in Figures 8–54 and 8–55.

Individual roles have been defined for several of the snRNPs. The U1 snRNP, for example, binds to the 5′ splice site, guided by a nucleotide sequence in the U1 RNA that forms base pairs complementary to the nine-nucleotide splice site consensus sequence (see Figure 8–53). Since RNA is capable of acting like an enzyme, either the RNA or the protein components of the spliceosome could be responsible for catalyzing the breakage and formation of covalent bonds required for RNA splicing.

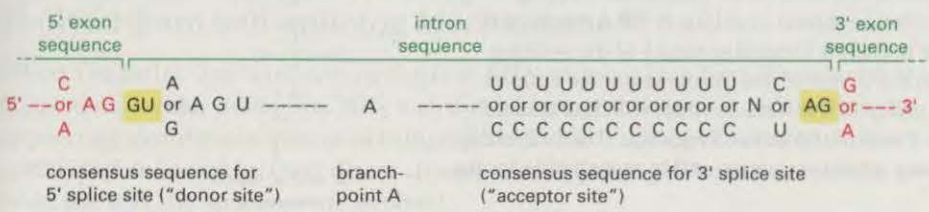


Figure 8–53 Consensus sequences for RNA splicing in higher eucaryotes. The sequence given is that for the RNA chain; the nearly invariant GU and AG dinucleotides at either end of the intron sequence are highlighted in *yellow* (see also Figure 8–51).

Sarepta Exhibit 1056, page 87

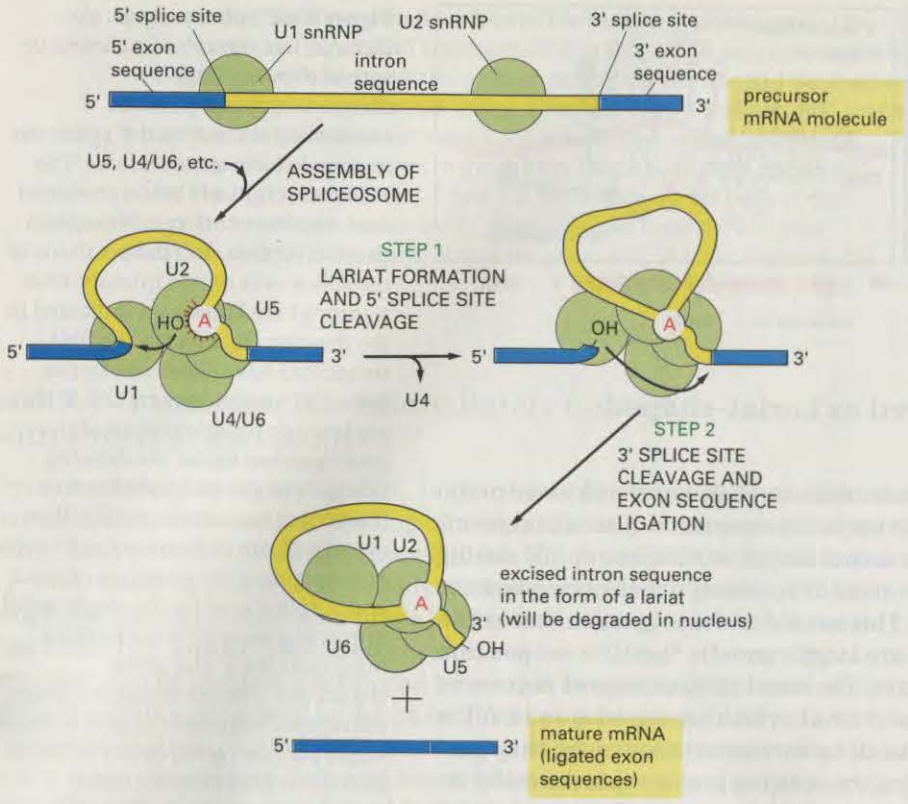


Figure 8–54 The RNA splicing mechanism. RNA splicing is catalyzed by a spliceosome formed from the assembly of U1, U2, U5, and U4/U6 snRNPs (shown as *green circles*) plus other components (not shown). After assembly of the spliceosome, the reaction occurs in two steps: in step 1 the branch-point A nucleotide in the intron sequence, which is located close to the 3′ splice site, attacks the 5′ splice site and cleaves it; the cut 5′ end of the intron sequence thereby becomes covalently linked to this A nucleotide, forming the branched nucleotide shown in Figure 8–55. In step 2 the 3′-OH end of the first exon sequence, which was created in the first step, adds to the beginning of the second exon sequence, cleaving the RNA molecule at the 3′ splice site; the two exon sequences are thereby joined to each other and the intron sequence is released as a lariat. The spliceosome complex sediments at 60S, indicating that it is nearly as large as a ribosome. These splicing reactions occur in the nucleus and generate mRNA molecules from primary RNA transcripts (mRNA precursor molecules).

Multiple Intron Sequences Are Usually Removed from Each RNA Transcript [36]

Because the spliceosome seems mainly to work by recognizing the consensus sequences that mark the two boundaries of an intron sequence (and for all intron sequences these consensus sequences are alike), the 5′ splice site (donor site) at the end of any one intron sequence can in principle be spliced to the 3′ splice site (acceptor site) of any other intron sequence. Indeed, when an RNA molecule is created artificially, with donor and acceptor splice sites from different intron sequences inserted into it, the intervening RNA is often recognized by the spliceosome and removed.

In view of this result, it is surprising that vertebrate genes can contain as many as 50 introns (see Table 8–1, p. 340). If *any* two 5′ and 3′ splice sites were mispaired for splicing, some functional mRNA sequences would be lost, with disastrous consequences. Somehow such mistakes are avoided: the RNA processing machinery normally guarantees that each 5′ splice site pairs only with the 3′ splice site that is closest to it in the downstream (5′-to-3′) direction of the linear RNA sequence (Figure 8–56). How this sequential pairing of splice sites is accomplished is not known, although the assembly of the spliceosome while the RNA transcript is still growing (see Figure 8–52) is presumed to play a major part in ensuring an orderly pairing of the appropriate splice sites. There is also evidence

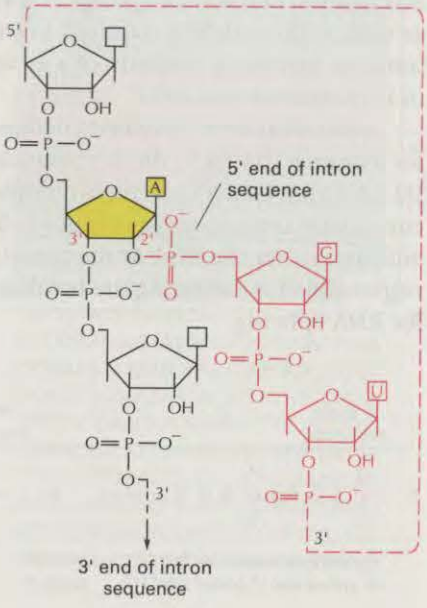


Figure 8–55 Structure of the branched RNA chain that forms during nuclear RNA splicing. The nucleotide shown in *yellow* is the A nucleotide highlighted in Figure 8–54. The branch is formed in step 1 of the splicing reaction illustrated there, when the 5′ end of the intron sequence couples covalently to the 2′-OH ribose group of the A nucleotide, which is located about 30 nucleotides from the 3′ end of the intron sequence. The branched chain remains in the final excised intron sequence and is responsible for its lariat form (see Figure 8–54).

Sarepta Exhibit 1056, page 88

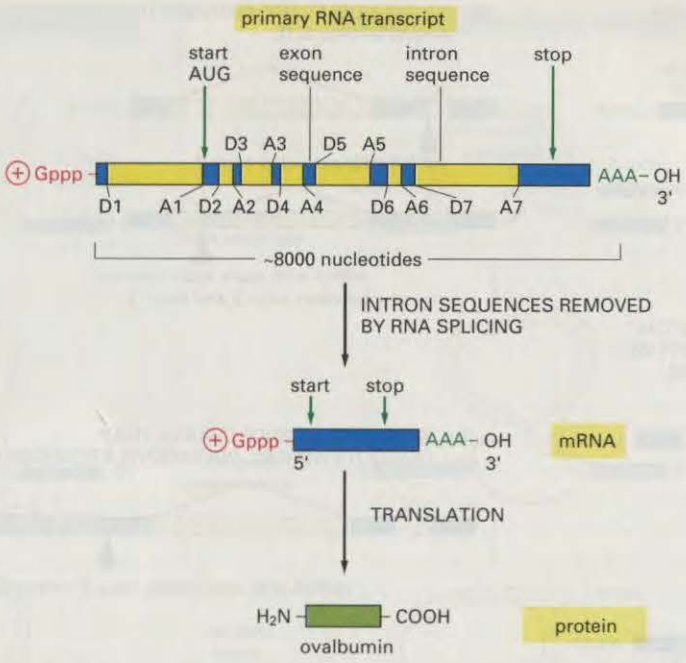


Figure 8–56 Splicing the primary RNA transcript from the chicken ovalbumin gene. The drawing shows the organized removal of seven introns required to obtain a functional mRNA molecule. The 5′ splice sites (donor sites) are denoted by D, and 3′ splice sites (acceptor sites) are denoted by A.

that the exact three-dimensional conformations adopted by the intron and exon sequences in the RNA transcript are important. We shall see in Chapter 9, however, that this simple 5′-to-3′ splicing can be altered by specialized control mechanisms that allow a single gene to produce several different mRNAs and hence several different proteins.

Studies of Thalassemia Reveal How RNA Splicing Can Allow New Proteins to Evolve [37]

Recombinant DNA techniques have made humans with inherited diseases an increasingly important source of material for genetic studies of cellular mechanisms. In a group of human genetic diseases called the **thalassemia syndromes,** for example, patients have an abnormally low level of hemoglobin—the oxygen-carrying protein in red blood cells. The change in the DNA sequence has been determined for more than 50 such mutants, and a large proportion of them have been found to have alterations in the pattern of splicing of globin RNA transcripts. Thus single nucleotide changes have been detected that inactivate a splice site. Surprisingly, analysis of the mRNAs produced in these mutant individuals reveals that the loss of a splice site does not prevent splicing but instead causes its normal partner site to seek out and become joined to a new "cryptic" site nearby; often a number of alternative splices are made in these mutants, causing the mutant gene to produce a set of altered proteins rather than just one (Figure 8–57). Other single nucleotide changes create new splice sites by changing a sequence in an intron or an exon into a consensus splice site. These results demonstrate that RNA splicing is a flexible process in higher eucaryotic cells, and they suggest that changes in the splicing pattern caused by random mutations could be an important pathway in the evolution of genes and organisms.

Spliceosome-catalyzed RNA Splicing Probably Evolved from Self-splicing Mechanisms [38]

When the lariat intermediate in nuclear RNA splicing was first discovered, it puzzled molecular biologists. Why was this bizarre pathway used rather than the apparently simpler alternative of bringing the 5′ and 3′ splice sites together in an initial step, followed by their direct cleavage and rejoining? The answer seems to lie in the way the spliceosome evolved.

Sarepta Exhibit 1056, page 89

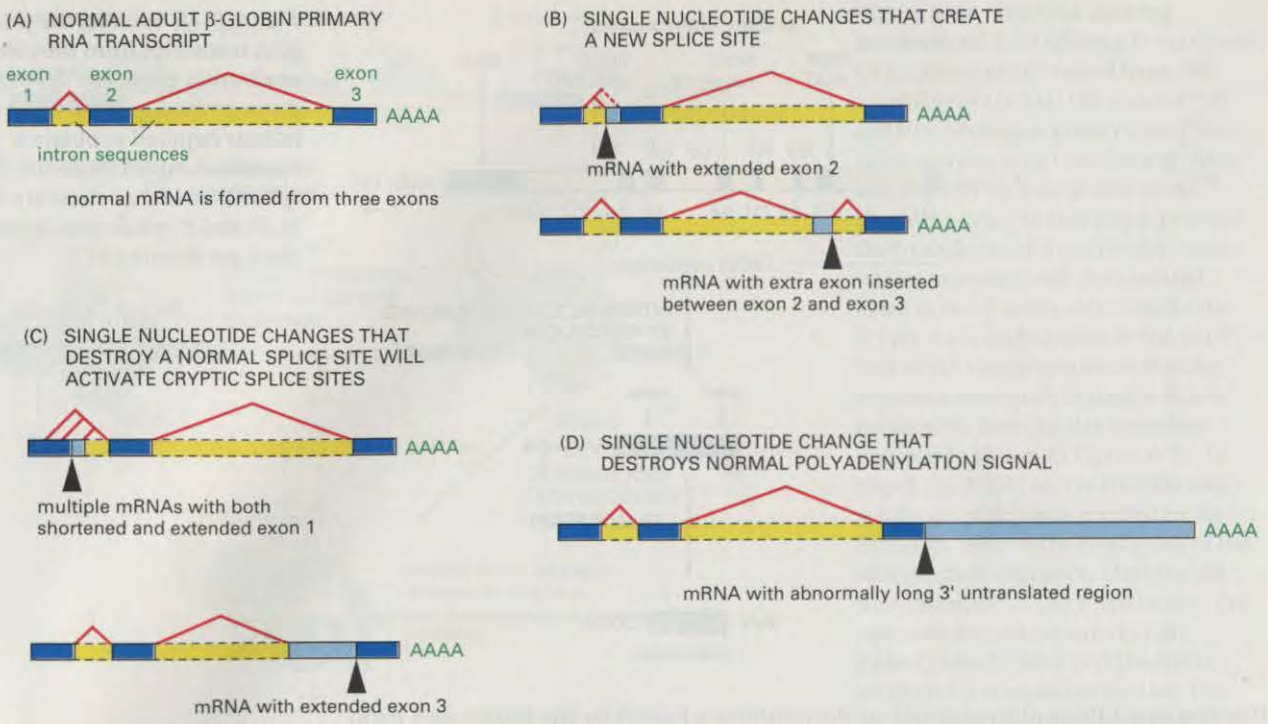


Figure 8–57 Abnormal processing of the β-globin primary RNA transcript in humans with β thalassemia. The site of each mutation is denoted by a *black arrowhead*. The *dark blue boxes* represent the three normal exon sequences illustrated previously in Figure 8–51, and the *red lines* join the 5′ and 3′ splice sites utilized in splicing the primary RNA transcript produced by the gene. The *light blue boxes* depict new nucleotide sequences included in the final mRNA molecule as a result of a mutation. Note that when a mutation leaves a normal splice site without a partner, one or more abnormal "cryptic" splice sites nearby are used as the partner site, as in (C). (After S.H. Orkin, in The Molecular Basis of Blood Diseases [G. Stamatoyannopoulos et al., eds.], pp. 106–126. Philadelphia: Saunders, 1987.)

As explained in Chapter 1, it is thought that early cells may have used RNA molecules rather than proteins as their major catalysts and stored their genetic information in RNA rather than DNA sequences. RNA-catalyzed splicing reactions presumably played important roles in these early cells, and some self-splicing RNA introns remain today—for example, in the nuclear rRNA genes of the ciliate *Tetrahymena*, in bacteriophage T4, and in some mitochondrial and chloroplast genes. A self-splicing intron sequence can be identified in a test tube by incubating a pure RNA molecule that contains the intron sequence and observing the splicing reaction; it can also be identified from the RNA sequence, inasmuch as large parts of the intron sequence need to be conserved in order to fold to create a catalytic surface in the RNA molecule. Two major classes of self-splicing intron sequences can be readily distinguished in this way. *Group I intron sequences* begin the splicing reaction by binding a G nucleotide to the intron sequence; the G is thereby activated to form the attacking group that will break the first of the phosphodiester bonds cleaved during splicing (the bond at the 5′ splice site). In *group II intron sequences* a specially reactive A residue in the intron sequence is the attacking group, and a lariat intermediate is generated. Otherwise the reaction pathways for the two types of sequences are the same. Both are presumed to represent vestiges of very ancient mechanisms (Figure 8–58).

In the evolution of nuclear RNA splicing, the reaction pathway used by the group II self-splicing intron sequences seems to have been retained, with the catalytic role of the intron sequences being replaced by separate spliceosome components. Thus the small RNAs U1 and U2, for example, may well be remnants of catalytic RNA sequences that were originally present in intron sequences. Shifting the catalysis from intron sequence to spliceosome presumably lifted most of the constraints on the evolution of introns, allowing many new intron sequences to evolve.

The Transport of mRNAs to the Cytoplasm Is Delayed Until Splicing Is Complete [39]

Finished mRNA molecules are thought to be recognized by receptor proteins in the nuclear pore complex and to be transported actively to the cytoplasm (dis-

Sarepta Exhibit 1056, page 90

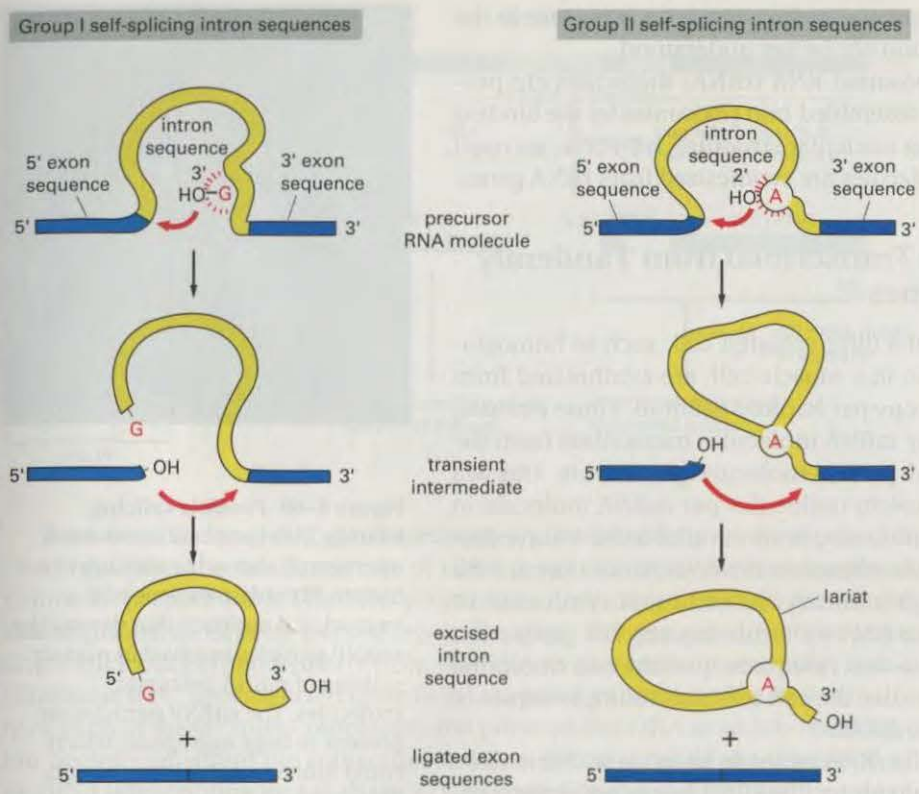


Figure 8–58 The two known classes of self-splicing intron sequences. The group I intron sequences bind a free G nucleotide to a specific site to initiate splicing (see Figure 3–21), while the group II intron sequences use a specially reactive A nucleotide in the intron sequence itself for the same purpose. The two mechanisms have been drawn in a way that emphasizes their similarities. Both are normally aided by proteins that speed up the reaction, but the catalysis is nevertheless mediated by the RNA in the intron sequence. The mechanism used by group II intron sequences forms a lariat and resembles the pathway catalyzed by the spliceosome (compare to Figure 8–54). (After T.R. Cech, *Cell* 44:207–210, 1986. © Cell Press.)

cussed in Chapter 12). The major proteins of the hnRNP particles and various processing molecules bound to the RNA, by contrast, are largely confined to the nucleus, although some of them pass into the cytoplasm with the transported mRNA before being rapidly stripped from the RNA and returned to the nucleus (Figure 8–59).

Studies of mutant yeasts suggest that for RNAs that have splice sites transport out of the nucleus can occur only after the splicing reaction has been completed. When a mutation creates a defect in the splicing machinery so that splicing cannot occur, unspliced mRNA precursors remain in the nucleus, while those mRNAs that do not require splicing (which includes most of the mRNAs in this single-cell eucaryote) are transported normally to the cytosol. This observation suggests that RNAs may be retained in the nucleus by their bound spliceosome components, which seem to form numerous large aggregates throughout the nucleus of higher eucaryotes. These aggregates could serve as "splicing islands" (Figure 8–60); although it is not known how they form or function, they may be

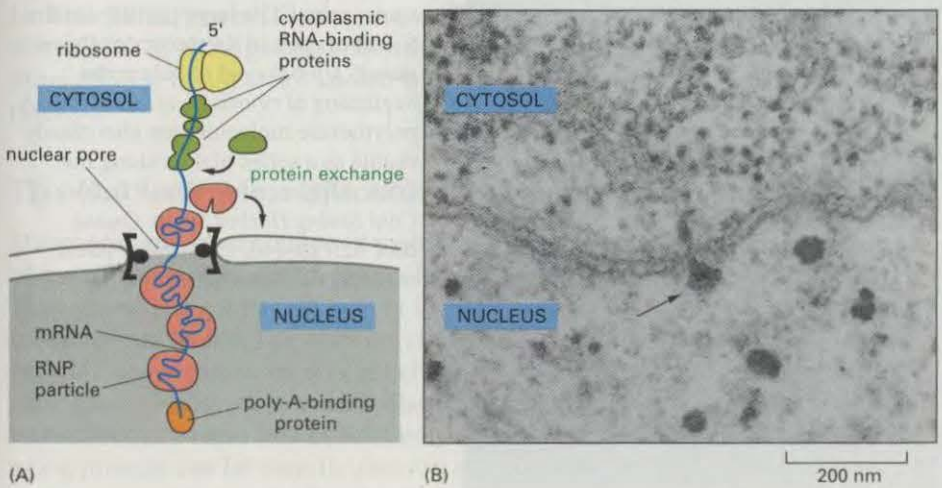


Figure 8–59 The transport of mRNA molecules through nuclear pores. (A) Schematic illustration of the change in the proteins bound to the RNA molecule as it moves out of the nucleus. (B) Electron micrograph of a large mRNA molecule produced in an insect salivary gland cell; this molecule (*arrow*) has apparently been caught in the process of moving to the cytosol. (B, from B.J. Stevens and H. Swift, *J. Cell Biol.* 31:55–77, 1966, by copyright permission of the Rockefeller University Press.)

Sarepta Exhibit 1056, page 91

analogous to the *nucleolus*, a much larger and more prominent structure in the nucleus, whose organization and function are better understood.

The nucleolus is the site where ribosomal RNA (rRNA) molecules are processed from a larger precursor RNA and assembled into ribosomes by the binding of ribosomal proteins. Before discussing nucleolar structure, however, we need to consider how the precursor rRNA molecules are synthesized from rRNA genes.

Ribosomal RNAs (rRNAs) Are Transcribed from Tandemly Arranged Sets of Identical Genes [40]

Many of the most abundant proteins of a differentiated cell, such as hemoglobin in the red blood cell and myoglobin in a muscle cell, are synthesized from genes that are present in only a single copy per haploid genome. These proteins are abundant because each of the many mRNA molecules transcribed from the gene can be translated into as many as 10 protein molecules per minute. This will normally produce more than 10,000 protein molecules per mRNA molecule in each cell generation. Such an amplification step is not available for the synthesis of the intrinsic RNA components of the ribosome, however, since they are the final gene products. Yet a growing higher eucaryotic cell must synthesize 10 million copies of each type of ribosomal RNA molecule in each cell generation in order to construct its 10 million ribosomes. Adequate quantities of ribosomal RNAs, in fact, can be produced only because the cell contains multiple copies of the **rRNA genes** that code for ribosomal RNAs.

Even *E. coli* needs seven copies of its rRNA genes to keep up with the cell's need for ribosomes. Human cells contain about 200 rRNA gene copies per haploid genome, spread out in small clusters on five different chromosomes, while cells of the frog *Xenopus* contain about 600 rRNA gene copies per haploid genome in a single cluster on one chromosome. In eucaryotes the multiple copies of the highly conserved rRNA genes on a given chromosome are located in a tandemly arranged series in which each gene (8000 to 13,000 nucleotide pairs long, depending on the organism) is separated from the next by a nontranscribed region known as *spacer DNA*, which can vary greatly in length and sequence. We shall see later that such multiple copies of tandemly arranged genes tend to co-evolve.



Figure 8–60 Possible splicing islands. This immunofluorescence micrograph shows the staining of a human fibroblast nucleus with a monoclonal antibody that detects the snRNP particles involved in nuclear splicing of mRNA precursor molecules. The snRNP particles are present in large aggregates, which could function as "splicing islands." The antibody detects specific proteins that are present in several of the snRNPs that function in the spliceosome. (Courtesy of N. Ringertz.)

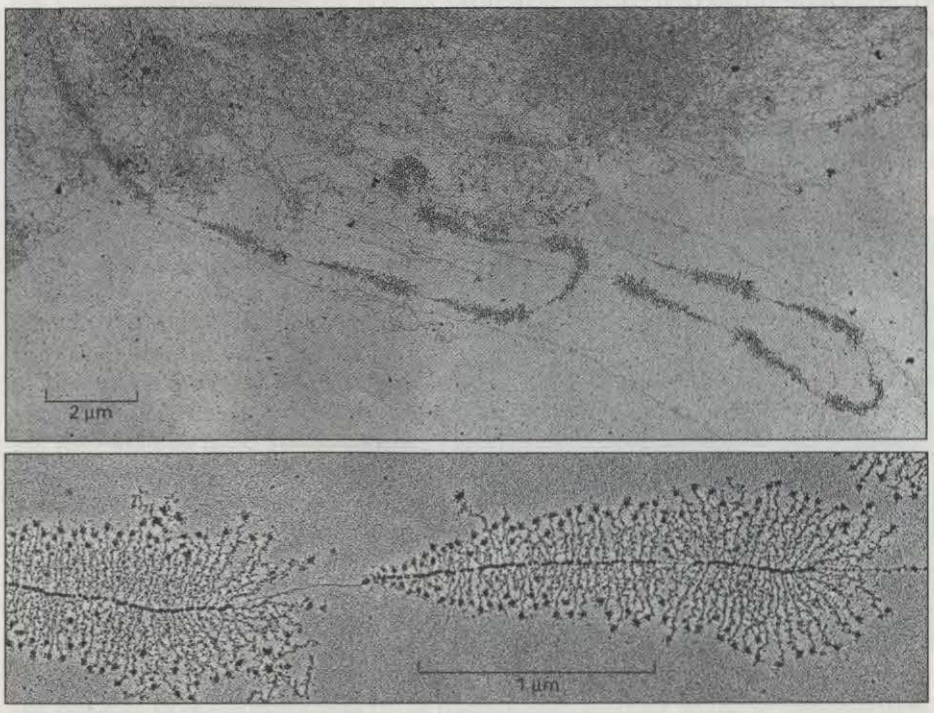


Figure 8–61 Transcription from tandemly arranged rRNA genes, as visualized in the electron microscope. The pattern of alternating transcribed gene and nontranscribed spacer is readily seen in the lower-magnification view in the *upper panel*. The large particles at the 5′ end of each rRNA transcript (*lower panel*) are believed to reflect the beginning of ribosome assembly; RNA polymerase molecules are also clearly visible as a series of dots along the DNA. (Upper panel, from V.E. Foe, *Cold Spring Harbor Symp. Quant. Biol.* 42:723–740, 1978; lower panel, courtesy of Ulrich Scheer.)

Sarepta Exhibit 1056, page 92

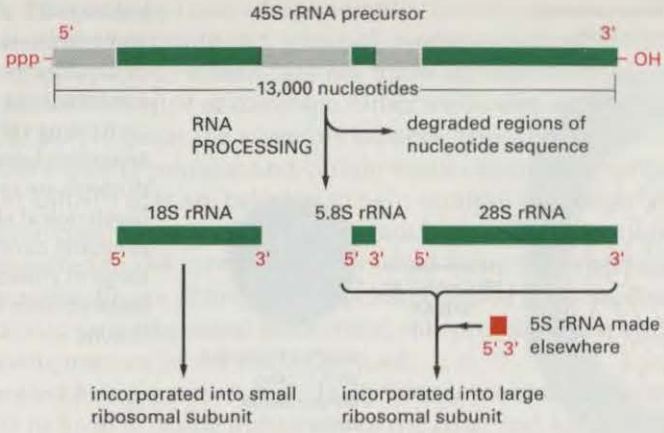


Figure 8–62 The processing of a 45S rRNA precursor molecule into three separate ribosomal RNAs. Nearly half of the nucleotide sequences in the primary RNA transcript are degraded in the nucleus.

Because of their repeating arrangement, and because they are transcribed at a very high rate, the tandem arrays of rRNA genes can easily be seen in spread chromatin preparations. The RNA polymerase molecules and their associated transcripts are so densely packed (typically about 100 per gene) that the transcripts stick out perpendicularly from the DNA to give each transcription unit a "Christmas tree" appearance (Figure 8–61). As noted earlier (see Figure 8–47), the tip of each of these "trees" represents the point on the DNA at which transcription begins and where the transcripts are thus shortest, while the other end of the rRNA transcription unit is sharply demarcated by the sudden disappearance of RNA polymerase molecules and their transcripts.

The rRNA genes are transcribed by *RNA polymerase I*, and each gene produces the same primary RNA transcript. In humans this RNA transcript, known as *45S rRNA*, is about 13,000 nucleotides long. Before it leaves the nucleus in assembled ribosomal particles, the 45S rRNA is cleaved to give one copy each of the 28S rRNA (about 5000 nucleotides), the 18S rRNA (about 2000 nucleotides), and the 5.8S rRNA (about 160 nucleotides) of the final ribosome. The derivation of these three rRNAs from the same primary transcript ensures that they will be made in equal quantities. The remaining part of each primary transcript (about 6000 nucleotides) is degraded in the nucleus (Figure 8–62). Some of these extra RNA sequences are thought to play a transient part in ribosome assembly, which begins immediately as specific proteins bind to the growing 45S rRNA transcripts in the nucleus.

Another set of tandemly arranged genes with nontranscribed spacers codes for the 5S rRNA of the large ribosomal subunit (the only rRNA that is transcribed separately). The 5S rRNA genes are only about 120 nucleotide pairs in length, and like a number of other genes encoding small stable RNAs (most notably the transfer RNA [tRNA] genes), they are transcribed by *RNA polymerase III*. Humans have about 2000 5S rRNA genes tandemly arranged in a single cluster far from all the other rRNA genes. It is not known why this one type of rRNA is transcribed separately.

The Nucleolus Is a Ribosome-producing Machine [40]

The continuous transcription of multiple gene copies ensures an adequate supply of the rRNAs, which are immediately packaged with ribosomal proteins to form ribosomes. The packaging occurs in the nucleus, in a large, distinct structure called the **nucleolus.** The nucleolus contains large loops of DNA emanating from several chromosomes, each of which contains a cluster of rRNA genes. Each such gene cluster is known as a **nucleolar organizer** region. Here the rRNA genes are transcribed at a rapid rate by RNA polymerase I. The beginning of the rRNA packaging process can be seen in electron micrographs of these genes: the 5′ tail of each transcript is encased by a protein-rich granule (see Figure 8–61). These gran-

Sarepta Exhibit 1056, page 93

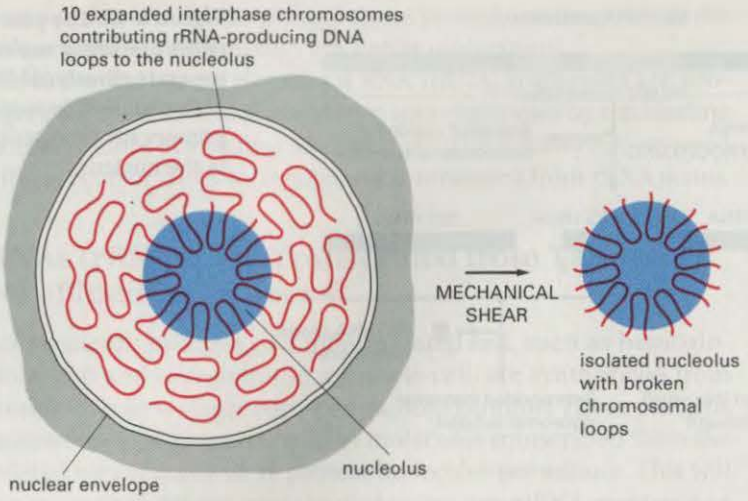


Figure 8–63 The nucleolus. This highly schematic view of a nucleolus in a human cell shows the contributions of loops of chromatin containing rRNA genes from 10 separate chromosomes. Purified nucleoli are very useful for biochemical studies of nucleolar function; to obtain such nucleoli, the loops of chromatin are mechanically sheared from their chromosomes, as shown.

ules, which do not appear on other types of RNA transcripts, presumably reflect the first of the protein-RNA interactions that take place in the nucleolus.

The biosynthetic functions of the nucleolus can be traced by briefly labeling newly made RNA with ^{3}H-uridine. After varying intervals of further incubation without ^{3}H-uridine, a cell fractionation procedure can be used to break the rRNA genes free of their chromosomes, thereby allowing the radioactive nucleoli to be isolated in relatively pure form (Figure 8–63). Such experiments show that the

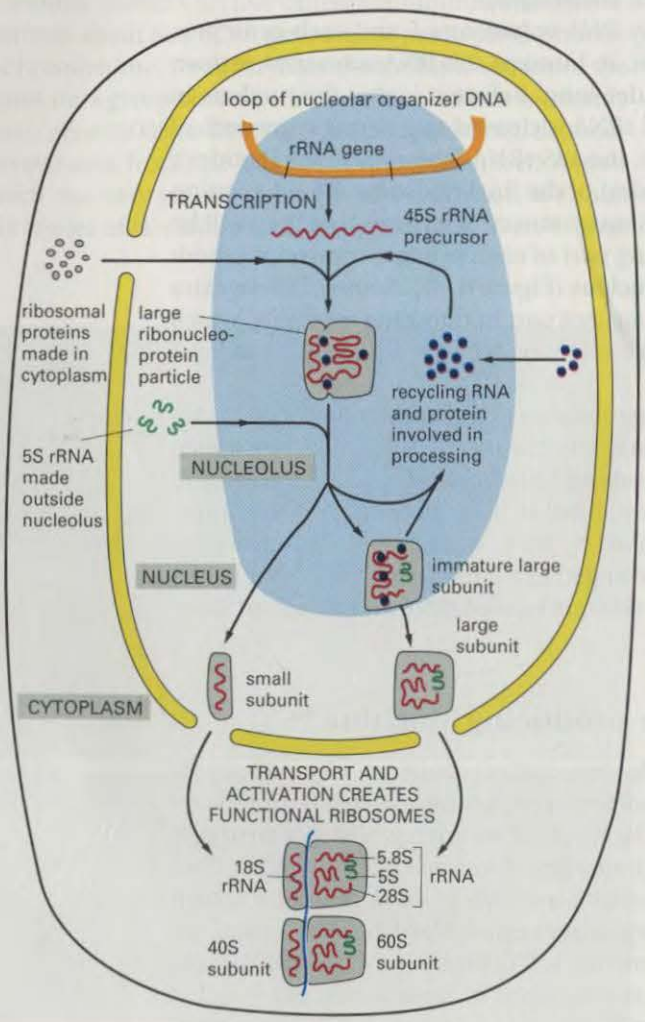


Figure 8–64 The function of the nucleolus in ribosome synthesis. The 45S rRNA transcript is packaged in a large ribonucleoprotein particle containing many ribosomal proteins imported from the cytoplasm. While this particle remains in the nucleolus, selected pieces are discarded as it is processed into immature large and small ribosomal subunits. These two subunits are thought to attain their final functional form only as each is individually transported through the nuclear pores into the cytoplasm.

Sarepta Exhibit 1056, page 94

intact 45S transcript is first packaged into a large complex containing many different proteins imported from the cytoplasm, where all proteins are synthesized. Most of the 80 different polypeptide chains that will make up the ribosome, as well as the 5S rRNAs, are incorporated at this stage. Other molecules are needed to process the 45S rRNA and to guide the assembly process. Thus the nucleolus also contains other RNA-binding proteins and certain small ribonucleoprotein particles (including U3 snRNP) that are believed to help catalyze the construction of ribosomes. These components remain in the nucleolus when the ribosomal subunits are exported to the cytoplasm in finished form. An especially notable component is *nucleolin*, an abundant, well-characterized RNA-binding protein that seems to coat only ribosomal transcripts; this protein stains with silver in the characteristic manner of the nucleolus itself.

As the 45S rRNA molecule is processed, it gradually loses some of its RNA and protein and then splits to form separate precursors of the large and small ribosomal subunits (Figure 8–64). Within 30 minutes of radioactive pulse labeling, the first mature small ribosomal subunits, containing their 18S rRNA, emerge from the nucleolus and appear in the cytoplasm. Assembly of the mature large ribosomal subunit, with its 28S, 5.8S, and 5S rRNAs, takes about an hour to complete. The nucleolus therefore contains many more incomplete large ribosomal subunits than small ones.

The last steps in ribosome maturation occur only as these subunits are transferred to the cytoplasm. This delay prevents functional ribosomes from gaining access to the incompletely processed hnRNA molecules in the nucleus.

The Nucleolus Is a Highly Organized Subcompartment of the Nucleus [41]

As seen in the light microscope, the large spheroidal nucleolus is the most obvious structure in the nucleus of a nonmitotic cell. Consequently, it was so closely scrutinized by early cytologists that an 1898 review could list some 700 references. By the 1940s cytologists had demonstrated that the nucleolus contains high concentrations of RNA and proteins, but its major function in ribosomal RNA synthesis and ribosome assembly was not discovered until the 1960s.

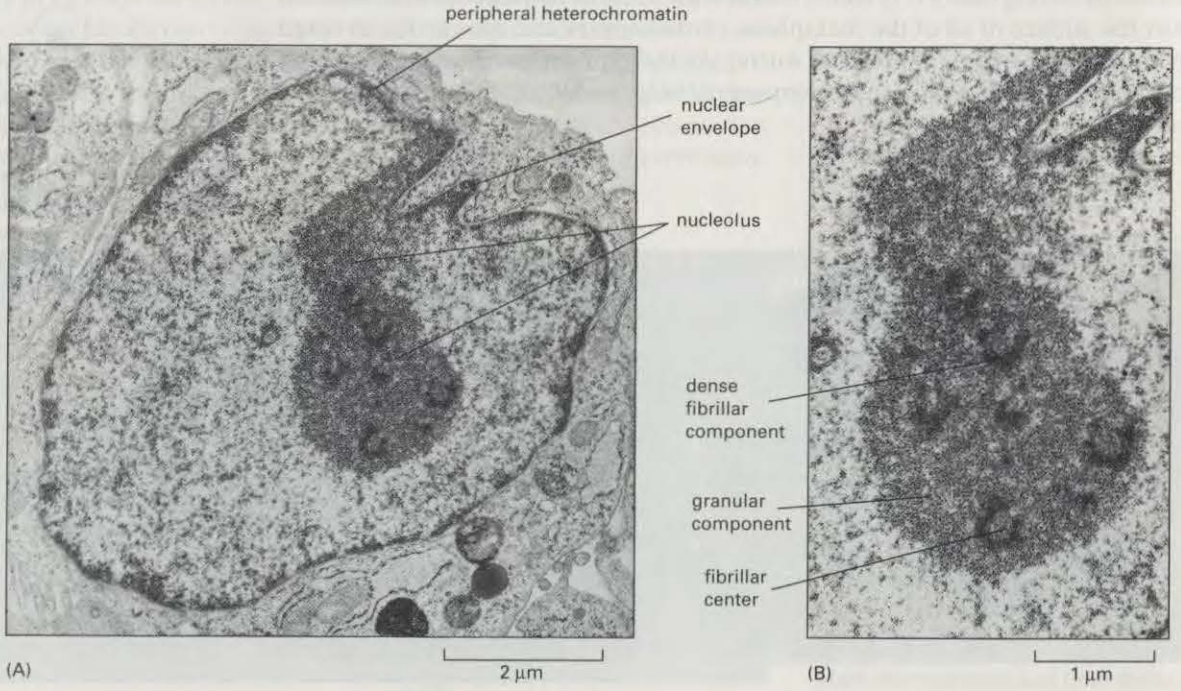


Figure 8–65 Electron micrograph of a thin section of a nucleolus in a human fibroblast, showing its three distinct zones. (A) View of entire nucleus. (B) High-power view of the nucleolus. (Courtesy of E.G. Jordan and J. McGovern.)

Sarepta Exhibit 1056, page 95

Some of the details of nucleolar organization can be seen in the electron microscope. Unlike the cytoplasmic organelles, the nucleolus is not bounded by a membrane; instead, it seems to be constructed by the specific binding of unfinished ribosome precursors to one another to form a large network. In a typical electron micrograph three partially segregated regions can be distinguished (Figure 8–65): (1) a pale-staining *fibrillar center*, which contains DNA that is not being actively transcribed; (2) a *dense fibrillar component*, which contains RNA molecules in the process of being synthesized; and (3) *a granular component*, which contains maturing ribosomal precursor particles.

The size of the nucleolus reflects its activity. Its size therefore varies greatly in different cells and can change in a single cell. It is very small in some dormant plant cells, for example, but can occupy up to 25% of the total nuclear volume in cells that are making unusually large amounts of protein. The differences in size are due largely to differences in the amount of the granular component, which is probably controlled at the level of ribosomal gene transcription: electron microscopy of spread chromatin shows that both the fraction of activated ribosomal genes and the rate at which each gene is transcribed can vary according to circumstances.

The Nucleolus Is Reassembled on Specific Chromosomes After Each Mitosis [42]

The appearance of the nucleolus changes dramatically during the cell-division cycle. As the cell approaches mitosis, the nucleolus first decreases in size and then disappears as the chromosomes condense and all RNA synthesis stops, so that generally there is no nucleolus in a metaphase cell. When ribosomal RNA synthesis restarts at the end of mitosis (in telophase), tiny nucleoli reappear at the chromosomal locations of the ribosomal RNA genes (Figure 8–66).

In humans the ribosomal RNA genes are located near the tips of each of 5 different chromosomes, as shown previously in Figure 8–32 (that is, on 10 of the 46 chromosomes in a diploid cell). Correspondingly, 10 small nucleoli form after mitosis in a human cell, although they are rarely seen as separate entities because they quickly grow and fuse to form the single large nucleolus typical of many interphase cells (Figure 8–67).

What happens to the RNA and protein components of the disassembled nucleolus during mitosis? It seems that at least some of them become distributed over the surface of all of the metaphase chromosomes and are carried as cargo to each of the two daughter cell nuclei. As the chromosomes decondense at telophase, these "old" nucleolar components help reestablish the newly emerging nucleoli.

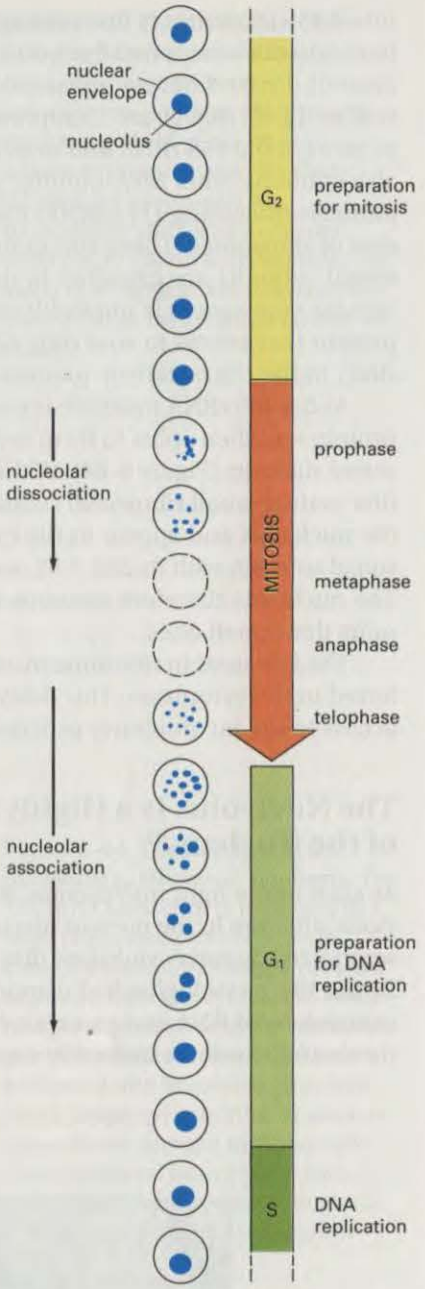


Figure 8–66 Changes in the appearance of the nucleolus in a human cell during the cell cycle. Only the cell nucleus is represented in this diagram. In most eucaryotic cells the nuclear membrane breaks down during mitosis, as indicated by the dashed circles.

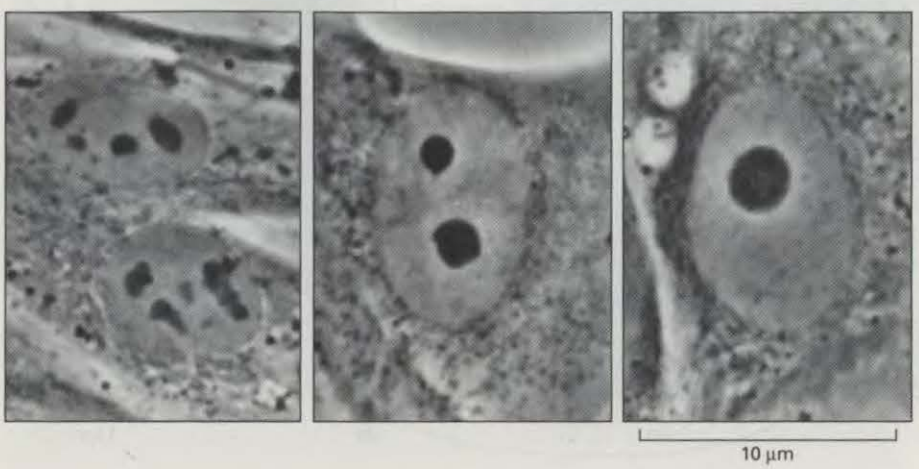


Figure 8–67 Nuclear fusion. These light micrographs of human fibroblasts grown in culture show various stages of nucleolar fusion. (Courtesy of E.G. Jordan and J. McGovern.)

Sarepta Exhibit 1056, page 96

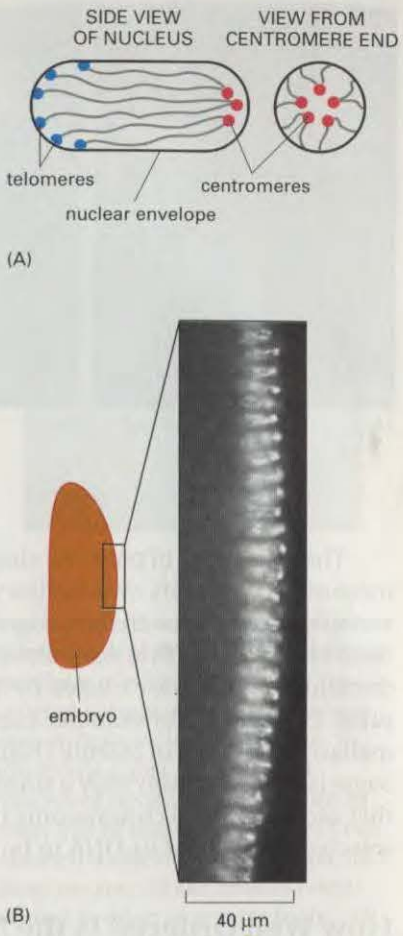


Figure 8–68 The polarized orientation of chromosomes in interphase cells of the early *Drosophila* embryo. (A) Diagrams of the Rabl orientation, with all centromeres facing one nuclear pole and all telomeres pointing toward the opposite pole. In the embryo each nucleus is elongated as shown. (B) Low-magnification light micrograph of a *Drosophila* embryo at the cellular blastoderm stage. The chromosomes in each interphase nucleus have been stained with a fluorescent dye. Note that the most brightly staining region (the chromocenter), which is known to contain the centromeric regions of each of the four chromosomes (see Figure 8–19), is oriented toward the outer surface of the embryo and thus faces the apical plasma membrane of every cell. (Courtesy of John Sedat.)

Individual Chromosomes Occupy Discrete Territories in the Nucleus During Interphase [43]

As we have just seen, specific genes from separate interphase chromosomes are brought together at a single site in the nucleus when the nucleolus forms. Are other parts of chromosomes also nonrandomly ordered in the nucleus? First raised by biologists in the late nineteenth century, this fundamental question still has not been answered satisfactorily.

A certain degree of chromosomal order results from the configuration that the chromosomes always have at the end of mitosis. Just before a cell divides, the condensed chromosomes are pulled to each spindle pole by microtubules attached to the centromeres; thus, as the chromosomes move, the centromeres lead the way and the distal arms (terminating in telomeres) lag behind. The chromosomes in many nuclei tend to retain this so-called *Rabl orientation* throughout interphase, with their centromeres facing one pole of the nucleus and their telomeres pointing toward the opposite pole (Figure 8–68A). In some cases the nucleus is specifically oriented in the cell: in the early *Drosophila* embryo, for example, all the centromeres face apically (Figure 8–68B). Such fixed nuclear orientations might have important effects on cell polarity, but it is difficult to design experiments to test this possibility.

In most cells the various chromosomes are indistinguishable from one another during interphase. Consequently, it is difficult to assess their arrangement in more detail than just described. The giant interphase chromosomes of the polytene cells of *Drosophila* larvae, however, are an exception. Here the individual chromosome bands can be resolved clearly enough to determine the precise positions of specific genes in intact nuclei by microscopic optical-sectioning and reconstruction techniques. The results of such analyses suggest that the interphase chromosome set is not highly ordered: although the Rabl orientation tends to be maintained, two apparently identical cells often have different chromosomes as nearest neighbors.



Figure 8–69 A stereo pair that displays the three-dimensional arrangement of the polytene chromosomes in a single nucleus of a *Drosophila* larval gland cell. The large ball is the nucleolus, and the course of each chromosome arm is represented by a line running along the chromosome axis. The telomeres tend to be on the surface of the nuclear envelope opposite the surface that is nearest the nucleolus, where all the centromeres are located. The chromosomes in such nuclei are never entangled, but their detailed foldings and neighbors are different in otherwise identical nuclei. (For viewing with crossed eyes; courtesy of Mark Hochstrasser and John Sedat.)

Sarepta Exhibit 1056, page 97

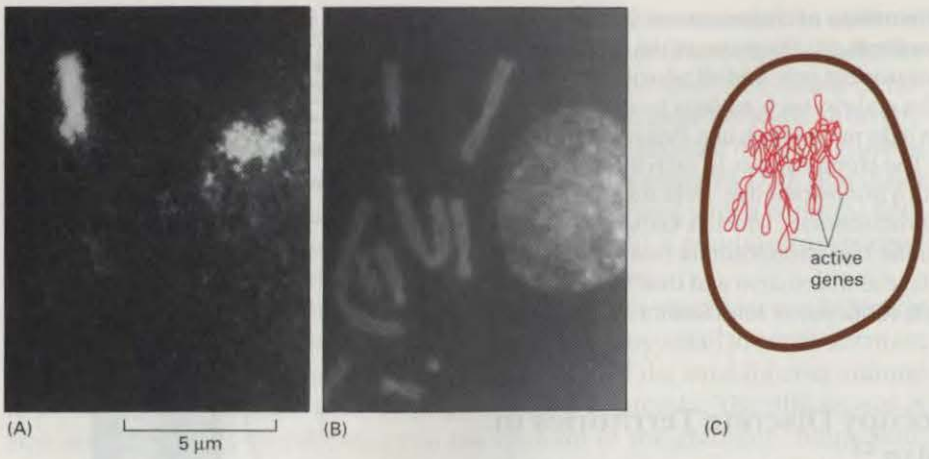


Figure 8–70 Selective labeling of a single chromosome in a cultured mammalian cell nucleus during interphase. In (A) and (B), an interphase nucleus freed from its cytoplasm is shown at the right, with scattered mitotic chromosomes released from a second cell on the left. (A) The results of *in situ* hybridization (using a fluorescent probe) to outline the single human chromosome in a human-hamster hybrid cell line. The same preparation is shown with all of the DNA labeled with a second fluorescent dye in (B). (C) Schematic drawing of the human chromosome detected in the interphase nucleus in (A), shown at a somewhat larger scale. (A and B, courtesy of Joyce A. Kobori and David R. Cox.)

These analyses of polytene chromosomes have also indicated that each chromosome occupies its own territory in the interphase nucleus—that is, the individual chromosomes are not extensively intertwined (Figure 8–69). Other experiments have shown that nonpolytene chromosomes also tend to occupy discrete domains in interphase nuclei. *In situ* hybridization experiments with an appropriate DNA probe, for example, can outline a single chromosome in hybrid mammalian cells grown in culture (Figure 8–70). Most of the DNA of such a chromosome is seen to occupy only a small portion of the interphase nucleus, suggesting that each individual chromosome remains compact and organized while allowing selected portions of its DNA to be active in RNA synthesis.

How Well Ordered Is the Nucleus? [44]

The interior of the nucleus is not a random jumble of its many RNA, DNA, and protein components. We have seen that the nucleolus is organized as an efficient ribosome-construction machine, and clusters of spliceosome components are organized, possibly as discrete RNA-splicing islands (see Figure 8–60). Order is also seen in the electron microscope when one focuses on the regions around nuclear pores: the chromatin that lines the inner nuclear membrane (which is unusually condensed chromatin and therefore clearly visible in electron micrographs) is excluded from a considerable region beneath and around each nuclear pore, clearing a path between the cytoplasm and the nucleoplasm (Figure 8–71). In some special cases, moreover, the nuclear pores are found to be highly organized in the nuclear envelope (Figure 8–72), presumably reflecting a corresponding organization of the nuclear lamina to which the pores are attached.

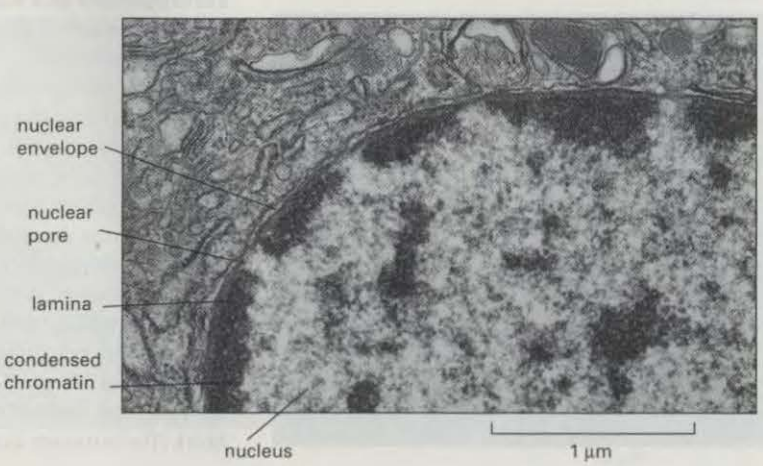


Figure 8–71 Electron micrograph of a mammalian cell nucleus. Note that the condensed chromatin underlying the nuclear envelope is excluded from regions around the nuclear pores. (Courtesy of Larry Gerace.)

Sarepta Exhibit 1056, page 98

Is there an intranuclear framework, analogous to the cytoskeleton, on which nuclear components are organized? Many cell biologists believe there is. The *nuclear matrix*, or *scaffold*, has been defined as the insoluble material left in the nucleus after a series of biochemical extraction steps. Some of the proteins that constitute it can be shown to bind specific DNA sequences called *SARs* or *MARs* (for scaffold- or matrix-associated regions). Such DNA sequences have been postulated to form the base of chromosomal loops (see Figure 8–18). By means of such chromosomal attachment sites, the matrix might help organize chromosomes, localize genes, and regulate DNA transcription and replication within the nucleus. Because the structural components of the matrix have not yet been identified, however, it remains uncertain whether the matrix isolated by cell biologists represents a structure that is present in intact cells.



Figure 8–72 Freeze-fracture electron micrograph of the elongated nuclear envelope of a fern spore. Note the ordered arrangement of the nuclear pore complexes in parallel rows. In other cells either concentrated clusters of nuclear pores or unusual areas free of nuclear pores have been detected in the nuclear envelope, and these are specifically oriented with respect to other structures in the cell. (Courtesy of Don H. Northcote; from K. Roberts and D.H. Northcote, *Microsc. Acta* 71:102–120, 1971.)

Summary

RNA polymerase, the enzyme that catalyzes DNA transcription, is a complex molecule containing many polypeptide chains. In eucaryotic cells there are three RNA polymerases, designated polymerases I, II, and III; they are evolutionarily related to one another and to bacterial RNA polymerase, and they have some subunits in common. After initiating transcription, each enzyme is thought to release one or more subunits and to bind other subunits that are required for RNA chain elongation and termination.

Most of the cell's mRNA is produced by a complex process beginning with the synthesis of heterogeneous nuclear RNA (hnRNA). The primary hnRNA transcript is made by RNA polymerase II. It is then capped by the addition of a special nucleotide to its 5′ end and is cleaved and then polyadenylated at its 3′ end. The modified RNA molecules are usually then subjected to one or more RNA splicing events, in which intron sequences are removed from the middle of the RNA molecule by a reaction catalyzed by a large ribonucleoprotein complex known as a spliceosome. In this process most of the mass of the primary RNA transcript is removed and degraded in the nucleus. As a result, although the rate of production of hnRNA typically accounts for about half of a cell's RNA synthesis, the mRNA produced represents only about 3% of the steady-state quantity of RNA in a cell.

Unlike genes that code for proteins, which are transcribed by polymerase II, the genes that code for most structural RNAs are transcribed by polymerase I and III. These genes are usually repeated many times in the genome and are often clustered in tandem arrays. RNA polymerase III makes a variety of small stable RNAs, including the tRNAs and the small 5S rRNA of the ribosome. RNA polymerase I makes the large rRNA precursor molecule (45S rRNA) containing the major rRNAs. Except for the ribosomes in mitochondria and chloroplasts, all the cell's ribosomes are assembled in the nucleolus—a distinct intranuclear organelle that is formed around the tandemly arranged rRNA genes, which are brought together from several chromosomes.

The Organization and Evolution of the Nuclear Genome

Much of evolutionary history is recorded in the genomes of present-day organisms and can be deciphered from a careful analysis of their DNA sequences. Tens of millions of DNA nucleotides have been sequenced thus far, and we can now see in outline how the genes coding for certain proteins have evolved over hundreds of millions of years. Studies of the occasional changes that occur in present-day chromosomes provide additional clues to the mechanisms that have brought about evolutionary change in the past. In this section we consider some of the general principles that have emerged from such molecular genetic studies, with emphasis on the organization and evolution of the nuclear genome in higher eucaryotes.

Sarepta Exhibit 1056, page 99

Genomes Are Fine-tuned by Point Mutation and Radically Remodeled or Enlarged by Genetic Recombination [45]

DNA nucleotide sequences must be accurately replicated and conserved. In Chapter 6 we discussed the elaborate DNA-replication and DNA-repair mechanisms that enable DNA sequences to be inherited with extraordinary fidelity: only about one nucleotide pair in a thousand is randomly changed every 200,000 years. Even so, in a population of 10,000 individuals, every possible nucleotide substitution will have been "tried out" on about 50 occasions in the course of a million years, which is a short span of time in relation to the evolution of species. Much of the variation created in this way will be disadvantageous to the organism and will be selected against in the population. When a rare variant sequence is advantageous, however, it will be rapidly propagated by natural selection. Consequently, it can be expected that in any given species the functions of most genes will have been optimized by random point mutation and selection.

While point mutation is an efficient mechanism for fine-tuning the genome, evolutionary progress in the long term must depend on more radical types of genetic change. Genetic recombination causes major rearrangements of the genome with surprising frequency: the genome can expand or contract by duplication or deletion, and its parts can be transposed from one region to another to create new combinations. Component parts of genes—their individual exons and regulatory elements—can be shuffled as separate modules to create proteins that have entirely new roles. In addition, duplicated copies of genes tend to diverge by further mutation and become specialized and individually optimized for subtly different functions. By these means the genome as a whole can evolve to become increasingly complex and sophisticated. In a mammal, for example, multiple variant forms of almost every gene exist—different actin genes for the different types of contractile cells, different opsin genes for the perception of lights of different colors, different collagen genes for the different types of connective tissues, and so on. The expression of each gene is regulated according to its own precise and specific rules. Moreover, DNA sequencing studies reveal that many genes share related modular segments but are otherwise very different: common sequence motifs are frequently found in otherwise unrelated proteins.

Genetic recombination, whereby one chromosome exchanges genetic material with another, is fundamental to the creation of such families of genes and gene segments. In Chapter 6 we discussed the molecular mechanisms of both general recombination and site-specific recombination. Here we consider some of their effects on the genome.

Tandemly Repeated DNA Sequences Tend to Remain the Same [46]

Gene duplications are usually attributed to rare accidents catalyzed by some of the enzymes that mediate normal recombination processes. Higher eucaryotes, however, contain an efficient enzymatic system that joins the two ends of a broken DNA molecule together, so that duplications (as well as inversions, deletions, and translocations of DNA segments) can also arise as a consequence of the erratic rejoining of fragments of chromosomes that have somehow become broken in more than one place. When duplicated DNA sequences are joined head to tail, they are said to be *tandemly repeated*. Once a single tandem repeat appears, it can be extended readily into a long series of tandem repeats by unequal crossover events between two sister chromosomes, inasmuch as the large amount of matching sequence provides an ideal substrate for general recombination (Figure 8–73). DNA duplication followed by sequential unequal crossing-over underlies *DNA amplification*, a process that often contributes to the formation of cancer by increasing the number of copies of genes (proto-oncogenes) that promote cancer (see Figure 24–27).

Sarepta Exhibit 1056, page 100

Tandemly repeated genes both increase and decrease in number due to unequal crossing-over (see Figure 8–73). They therefore would be expected to be maintained by natural selection in large numbers only if the extra copies were beneficial to the organism. We have already discussed the hundreds of tandemly repeated genes that code for the vertebrate large ribosomal RNA precursor; these are needed to keep up with a growing cell's demand for new ribosomes. Similarly, vertebrates have clusters of tandemly repeated genes that encode other structural RNAs, including 5S rRNA and the U1 and U2 snRNAs, as well as clusters of repeated histone genes, which produce the large amounts of histones required during each S phase.

One might expect that in the course of evolution the sequences of the genes in a tandem array—and of the nontranscribed *spacer* DNA between them—would drift apart. With many copies of the same gene there should be little selection against random mutations that alter one or a few of the copies, and most nucleotide changes in the long nontranscribed spacer regions would have no functional consequence. In fact, however, the sequences of the tandemly repeated genes and their spacer DNAs are generally almost identical. Two mechanisms are thought to account for this. First, recurring unequal crossing-over events will cause the continued expansion and contraction of tandem arrays, and computer simulations show that this will tend to keep the sequences the same (Figure 8–74A). Second, related DNA sequences can become homogenized through *gene conversion*—the process whereby a portion of the DNA sequence is changed by copying a closely similar sequence present at a different site in the genome, as described in Chapter 6 (Figure 8–74B). Although gene conversion does not require that the genes be tandemly repeated, in higher eucaryotes it seems to occur mainly between genes that are close to each other.

The movement of one gene copy in a tandem array to a new chromosomal location will protect it from both of the above homogenizing influences. Thus, in higher eucaryotes accidental gene translocation is an important step in the evolution of new genes: it allows the translocated DNA sequence to begin to evolve independently, so that it can acquire new functions that might benefit the organism.

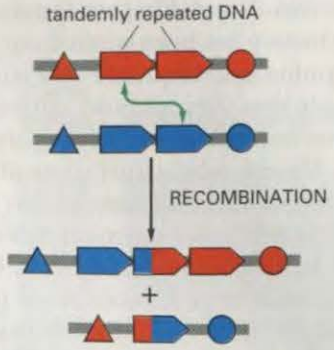


Figure 8–73 A family of tandemly repeated genes frequently loses and gains gene copies due to unequal crossing-over between sister chromosomes containing the genes. This type of event is frequent because the long regions of homologous DNA sequence are good substrates for the general genetic recombination process (discussed in Chapter 6).

The Evolution of the Globin Gene Family Shows How Random DNA Duplications Contribute to the Evolution of Organisms [47]

In addition to generating a number of sets of tandemly repeated genes, DNA duplications have played a more important general role in the evolution of new

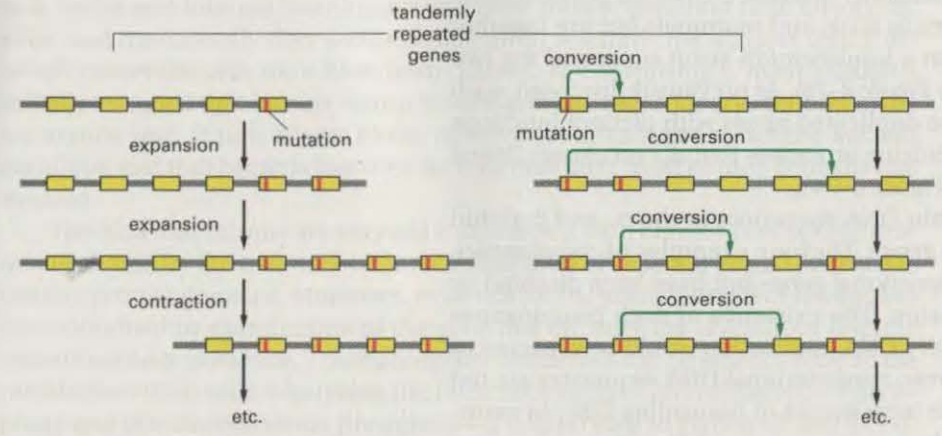


Figure 8–74 Two types of events that help to keep all DNA sequences in a tandem array very similar to one another. (A) The continual expansion and contraction of the number of gene copies in a tandem array caused by unequal crossing-over (see Figure 8–73) tends to homogenize all of the gene sequences in the array. (B) In gene conversion one gene copy acts as a template that passes all or part of its DNA sequence to another gene copy. In higher eucaryotes this process seems to be confined largely to genes that are next to each other on the chromosome; in lower eucaryotes such as fungi, where gene conversion is most readily studied, it is not confined to neighboring genes.

Sarepta Exhibit 1056, page 101

proteins. The globin gene family provides a good example because its evolutionary history has been worked out particularly well. The unmistakable homologies in amino acid sequence and structure among the present-day globin genes indicate that they all must derive from a common ancestral gene, even though some now occupy widely separated locations in the mammalian genome.

We can reconstruct some of the past events that produced the various types of oxygen-carrying hemoglobin molecules by considering the different forms of the protein in organisms at different levels on the phylogenetic scale. A molecule like hemoglobin was necessary to allow multicellular animals to grow to a large size, since large animals could no longer rely on the simple diffusion of oxygen through the body surface to oxygenate their tissues adequately. Consequently, hemoglobinlike molecules are found in all vertebrates and in many invertebrates. The most primitive oxygen-carrying molecule in animals is a globin polypeptide chain of about 150 amino acids, which is found in many marine worms, insects, and primitive fish. The hemoglobin molecule in higher vertebrates, however, is composed of two kinds of globin chains. It appears that about 500 million years ago, during the evolution of higher fish, a series of gene mutations and duplications occurred. These events established two slightly different globin genes, coding for the α- and β-globin chains in the genome of each individual. In modern higher vertebrates each hemoglobin molecule is a complex of two α chains and two β chains (Figure 8–75). The four oxygen binding sites in the $\alpha_2\beta_2$ molecule interact, allowing a cooperative allosteric change in the molecule as it binds and releases oxygen, which enables hemoglobin to take up and to release oxygen more efficiently than the single-chain version.

Still later, during the evolution of mammals, the β-chain gene apparently underwent mutation and duplication to give rise to a second β-like chain that is synthesized specifically in the fetus. The resulting hemoglobin molecule has a higher affinity for oxygen than adult hemoglobin and thus helps in the transfer of oxygen from the mother to the fetus. The gene for the new β-like chain subsequently mutated and duplicated again to produce two new genes, ε and γ, the ε chain being produced earlier in development (to form $\alpha_2\varepsilon_2$) than the fetal γ chain, which forms $\alpha_2\gamma_2$ (see Figure 9–52). A duplication of the adult β-chain gene occurred still later, during primate evolution, to give rise to a δ-globin gene and thus to a minor form of hemoglobin ($\alpha_2\delta_2$) found only in adult primates (Figure 8–76). Each of these duplicated genes has been modified by point mutations that affect the properties of the final hemoglobin molecule, as well as by changes in regulatory regions that determine the timing and level of expression of the gene.

The end result of the gene duplication processes that have given rise to the diversity of globin chains is seen clearly in the genes that arose from the original β gene, which are arranged as a series of homologous DNA sequences located within 50,000 nucleotide pairs of one another. A similar cluster of α-globin genes is located on a separate human chromosome. Because the α- and β-globin gene clusters are on separate chromosomes in birds and mammals but are together in the frog *Xenopus*, it is believed that a translocation event separated the two genes about 300 million years ago (see Figure 8–76). As previously discussed, such translocations probably help stabilize duplicated genes with distinct functions by protecting them from the homogenizing processes that act on closely linked genes of similar DNA sequence (see Figure 8–74).

There are several duplicated globin DNA sequences in the α- and β-globin gene clusters that are not functional genes. They are examples of *pseudogenes*. These have a close homology to the functional genes but have been disabled by mutations that prevent their expression. The existence of such pseudogenes should not be surprising since not every DNA duplication would be expected to lead to a new functional gene. Moreover, nonfunctional DNA sequences are not rapidly discarded, as indicated by the large excess of noncoding DNA in mammalian genomes, discussed previously.

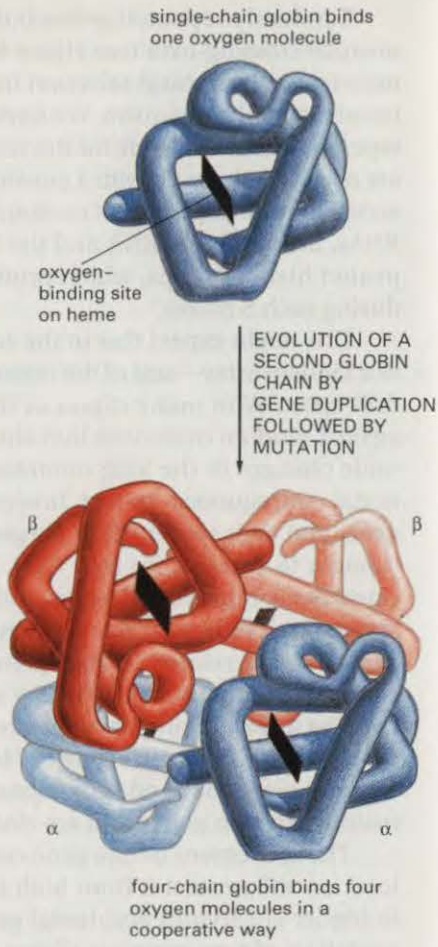


Figure 8–75 A comparison of the structure of one-chain and four-chain globins. The four-chain globin shown is hemoglobin, which is a complex of two α- and two β-globin chains. The one-chain globin in some primitive vertebrates forms a dimer that dissociates when it binds oxygen, representing an intermediate in the evolution of the four-chain globin.

Sarepta Exhibit 1056, page 102

A great deal of our evolutionary history will be discernible in our chromosomes once the DNA sequences of many gene families have been compared in different animals (see also Figure 7–4).

Genes Encoding New Proteins Can Be Created by the Recombination of Exons [45]

The role of DNA duplication in evolution is not confined to the generation of gene families. It can also be important in generating new single genes. The proteins encoded by genes generated in this way can be recognized by the presence of repeating, similar protein domains, which are covalently linked to one another in series. The immunoglobulins (Figure 8–77) and albumins, for example, as well as most fibrous proteins (such as the collagens) are encoded by genes that have evolved by repeated duplications of a primordial DNA sequence.

In genes that have evolved in this way, as well as in many other genes, each separate exon often encodes an individual protein folding unit, or domain. It is believed that the organization of DNA coding sequences as a series of such exons separated by long introns has greatly facilitated the evolution of new proteins. The duplications necessary to form a single gene coding for a protein with repeating domains, for example, can occur by breaking and rejoining the DNA anywhere in the long introns on either side of an exon encoding a useful protein domain; without introns there would be only a few sites in the original gene at which a recombinational exchange between sister DNA molecules could duplicate the domain. By enabling the duplication to occur at many potential recombination sites rather than at just a few, introns increase the probability of a favorable duplication event.

For the same reason, the presence of introns greatly increases the probability that a chance recombination event will generate a functional hybrid gene by joining two initially separated DNA sequences that code for different protein domains in such a way that both domains are preserved in the new protein that the hybrid gene encodes (see Figure 8–81, for example). The presumed results of such recombinations are seen in many present-day proteins (see Figure 3–43). Thus the large separation between the exons encoding individual domains in higher eucaryotes is thought to accelerate the process by which random genetic-recombination events generate useful new proteins. This could help to explain the successful evolution of these very complex organisms.

Most Proteins Probably Originated from Highly Split Genes [48]

The discovery in 1977 of genes split up by introns was unexpected. Previously all genes analyzed in detail were bacterial genes, which lack introns. Bacteria also lack nuclei and internal membranes and have smaller genomes than eucaryotic cells, and traditionally they were considered to resemble the simpler cells from which eucaryotic cells must have been derived. Not surprisingly, most biologists initially assumed that introns were a bizarre and late evolutionary addition to the eucaryotic line. It now seems likely, however, that split genes are the ancient condition and that bacteria lost their introns only after most of their proteins had evolved.

The idea that introns are very old is consistent with current concepts of protein evolution by the trial-and-error recombination of separate exons that encode distinct protein domains. Moreover, evidence for the ancient origin of introns has been obtained by examination of the gene that encodes the ubiquitous enzyme *triosephosphate isomerase*. Triosephosphate isomerase has an essential role in the metabolism of all cells, catalyzing the interconversion of glyceraldehyde-3 phosphate and dihydroxyacetone phosphate—a central step in glycolysis and gluconeogenesis (see Figure 2–21). By comparing the amino acid sequence of this en-

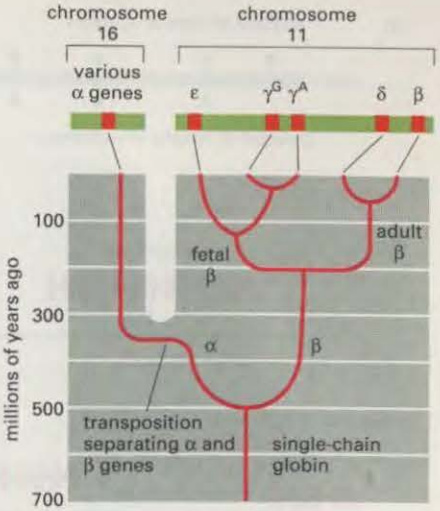


Figure 8–76 An evolutionary scheme for the globin chains that carry oxygen in the blood of animals. The scheme emphasizes the β-like globin gene family. A relatively recent gene duplication of the γ-chain gene produced γ^G and γ^A, which are fetal β-like chains of identical function. The location of the globin genes in the human genome is shown at the top of the figure.

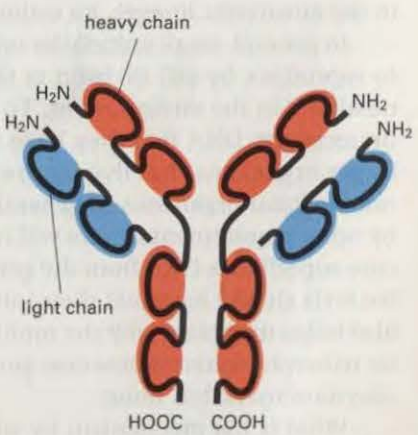


Figure 8–77 Schematic view of an antibody (immunoglobulin) molecule. This molecule is a complex of two identical heavy chains and two identical light chains. Each heavy chain contains four similar, covalently linked domains, while each light chain contains two such domains. Each domain is encoded by a separate exon, and all of the exons are thought to have evolved by the serial duplication of a single ancestral exon.

Sarepta Exhibit 1056, page 103

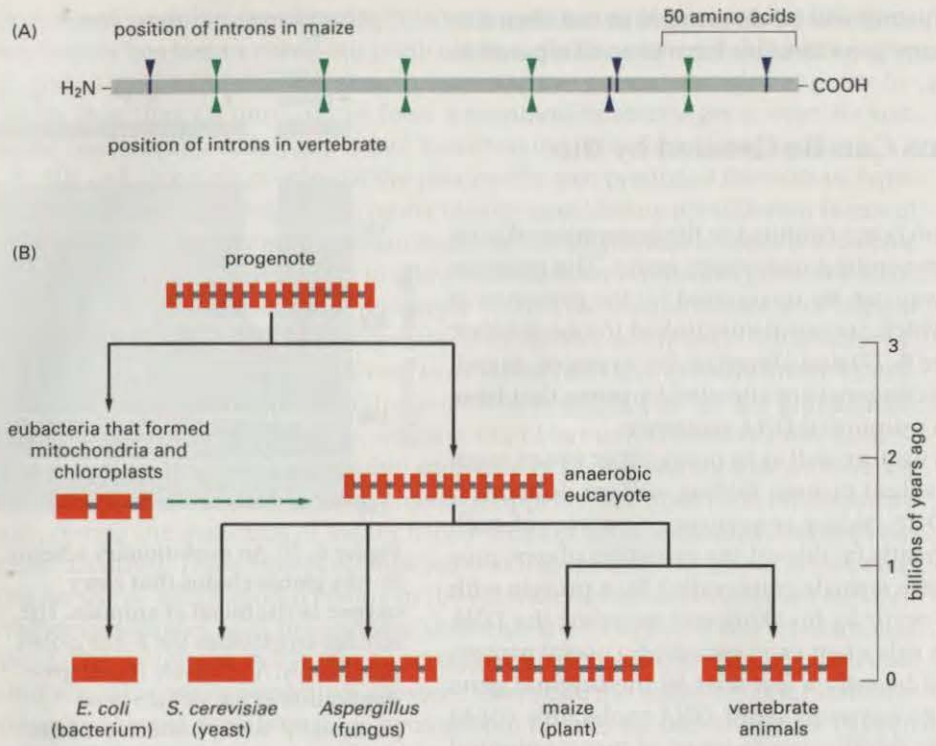


Figure 8–78 The ancient origin of split genes. (A) A comparison of the exon structure of the *triosephosphate isomerase* gene in plants and animals. The intron positions that are identical in maize (corn) and vertebrates are marked with *green arrowheads*, while the intron positions that differ are marked with *blue arrowheads*. Since plants and animals are thought to have diverged from a common ancestor about a billion years ago, the introns that they share must be of very ancient origin. (B) An outline of how a particular gene may have evolved. The exon sequences are shown in *red* and the intron sequences in *gray*. The gene illustrated here codes for a hypothetical protein that is required in all cells. Like *triosephosphate isomerase*, this protein is assumed to have evolved to its final three-dimensional structure before the eubacterial, archaebacterial, and eucaryotic lineages split off from a common ancestor cell—designated here as a "progenote." The dotted *green* line marks the approximate time of the endosymbiotic events that gave rise to mitochondria and chloroplasts (see pp. 20–21). (A, after W. Gilbert, M. Marchionni, and G. McKnight, *Cell* 46:151–154, 1987. © Cell Press.)

zyme in various organisms, it is possible to deduce that the enzyme evolved before the divergence of procaryotes and eucaryotes from a common ancestor; the human and bacterial amino acid sequences are 46% identical. The gene encoding the enzyme contains six introns in vertebrates (chickens and humans), and five of these are in precisely the same positions in maize. This implies that these five introns were present in the gene before plants and animals diverged in the eucaryotic lineage, an estimated 10^9 years ago (Figure 8–78).

In general, small unicellular organisms are under a strong selection pressure to reproduce by cell division at the maximum rate permitted by the levels of nutrients in the environment. To do this, they must minimize the amount of unnecessary DNA that they have to synthesize in each cell-division cycle. For larger organisms that live by predation, where size is an advantage, and for multicellular organisms in general, where rates of cell division are constrained by other requirements, there will not be such strong selection pressure to eliminate superfluous DNA from the genome. This argument may help to explain why bacteria should have lost their introns while eucaryotes have retained them. It also helps to explain why the multicellular fungus *Aspergillus* has five introns in its triosephosphate isomerase gene, whereas its unicellular relative, the yeast *Saccharomyces*, has none.

What is the mechanism by which introns are lost? Precise loss of introns would occur only rarely by piecemeal random deletions of short segments of DNA, yet precise and selective loss of introns seems not uncommon in eucaryotic cells (and perhaps was also frequent in the ancestors of bacteria). Whereas most vertebrates contain only a single insulin gene with two intron sequences, rats, for example, contain a second, neighboring insulin gene with only one intron. The second gene apparently arose by gene duplication relatively recently and subsequently lost one of its introns. Because intron loss requires the exact rejoining of DNA coding sequences, the most likely source of the information needed for such an event is an mRNA transcript of the original gene, from which the intron sequences will have been precisely removed. We know that messenger RNAs may be copied back into DNA through the activity of *reverse transcriptases* (see p. 282), and it is thought that recombination enzymes on occasion allow these DNA copies to become paired with the original sequence, which is

Sarepta Exhibit 1056, page 104

then "corrected" to an intronless form by a gene-conversion type of event. This pathway of intron loss has been demonstrated in the laboratory using the powerful genetic tools available in the yeast *S. cerevisiae.*

Reverse transcriptases are not needed for the central genetic pathways, but they are produced in cells by specific transposable elements (see Table 8–4) as well as by all retroviruses. The generation of DNA copies of segments of the genome by reverse transcription has contributed in several ways to the evolution of the genomes of higher organisms, as we discuss later.

A Major Fraction of the DNA of Higher Eucaryotes Consists of Repeated, Noncoding Nucleotide Sequences [49]

Eucaryotic genomes contain not only introns but also large numbers of copies of other seemingly nonessential DNA sequences that do not code for protein. The presence of such repeated DNA sequences in higher eucaryotes was first revealed by a hybridization technique that measures the number of gene copies. In this procedure the genome is broken mechanically into short fragments of DNA double helix about 1000 nucleotide pairs long, and the fragments are then denatured to produce DNA single strands. The speed with which the single-stranded fragments in the mixture reanneal under conditions in which the double-helical conformation is stable depends on how long it takes each fragment to find a complementary fragment to pair with, which in turn depends on the concentration of suitable fragments in the mixture. For the most part, the reaction is very slow. The haploid genome of a mammalian cell, for example, is represented by about 6 million different 1000-nucleotide-long DNA fragments, and any fragment whose sequence is present in only one copy must randomly collide with 6 million noncomplementary strands for every complementary partner strand that it happens to find.

When the DNA from a human cell is analyzed in this way under conditions that require near perfect matching (high stringency conditions, see Figure 7–17), about 70% of the DNA strands reanneal as slowly as one would expect for a large collection of unique (nonrepeated) DNA sequences, requiring days for complete annealing. But most of the remaining 30% of the DNA strands anneal much more quickly. These strands contain sequences that are repeated many times in the genome, and they thus collide with a complementary partner relatively rapidly. Most of these highly repeated DNA sequences do not encode proteins, and they are of two types: about one-third are the tandemly repeated *satellite DNAs,* to be discussed next; the rest are *interspersed repeated DNAs.* As we shall see, most of the latter DNAs derive from a few transposable DNA sequences that have multiplied to especially high copy numbers in the human genome.

Satellite DNA Sequences Have No Known Function [50]

The most rapidly annealing DNA strands in an experiment of the type just described usually consist of very long tandem series of repetitions of a short nucleotide sequence (Figure 8–79). The repeat unit in a sequence of this type may be composed of only one or two nucleotides, but most repeats are longer, and in mammals they are typically composed of variants of a short sequence organized into a repeat of a few hundred nucleotides. These tandem repeats of a simple sequence are called **satellite DNAs** because the first DNAs of this type to be discovered had an unusual ratio of nucleotides that made it possible to separate them by density-gradient centrifugation from the bulk of the cell's DNA as a minor component (or "satellite"). Satellite DNA sequences generally are not transcribed and are located most often in the heterochromatin associated with the centromeric regions of chromosomes. In some mammals a single type of satellite DNA sequence constitutes 10% or more of the DNA and may even occupy a whole chromosome arm, so that the cell contains millions of copies of the basic repeated sequence.

Sarepta Exhibit 1056, page 105

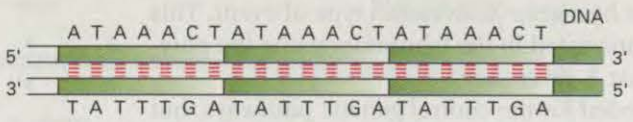


Figure 8–79 Satellite DNA. A simple satellite DNA sequence from *Drosophila* is shown. It consists of many serially arranged repetitions of a sequence seven nucleotide pairs long, and it occurs millions of times in the *Drosophila* haploid genome.

Satellite DNA sequences seem to have changed unusually rapidly and even to have shifted their positions on chromosomes in the course of evolution. When two homologous mitotic chromosomes of any human are compared, for example, some of the satellite DNA sequences usually are found arranged in a strikingly different manner on the two chromosomes. Moreover, in contrast to the high degree of conservation of DNA sequences elsewhere in the genome, generally there are marked differences in the satellite DNA sequences of two closely related species. No function has yet been found for satellite DNA sequences: tests designed to demonstrate a role in chromosome pairing or nuclear organization have failed thus far to reveal any evidence for such a role. It has therefore been suggested that they are an extreme form of "selfish DNA" sequences, whose properties ensure their own retention in the genome but which do nothing to help the survival of the cells containing them. Other sequences that are commonly viewed as selfish are the *transposable elements*, which we discuss next.

The Evolution of Genomes Has Been Accelerated by Transposable Elements [51]

Genomes generally contain many varieties of **transposable elements.** These segments of DNA were first discovered in maize, where several have been sequenced and characterized. Eucaryotic transposable elements have been studied most extensively in *Drosophila*, where more than 30 varieties are known, varying in length between 2000 and 10,000 nucleotide pairs; most are present in 5 to 10 copies per diploid cell.

At least three broad classes of transposable elements can be distinguished by the peculiarities of their sequence organization (Table 8–4). Some elements move from place to place within chromosomes directly as DNA, while many others move via an RNA intermediate, as described in Chapter 6. In either case they

Table 8–4 Three Major Families of Transposable Elements

Structure	Genes in Complete Element	Mode of Movement	Examples
 short inverted repeats at each end	encodes transposase	moves as DNA, either excising or following a replicative pathway	P element (*Drosophila*) Ac-Ds (maize) Tn3 and IS1 (*E. coli*) Tam3 (snapdragon)
 directly repeated long terminal repeats (LTRs) at ends	encodes reverse transcriptase and resembles retrovirus	moves via an RNA intermediate produced by promoter in LTR	Copia (*Drosophila*) Ty (yeast) THE-1 (human) Bs1 (maize)
 Poly A at 3′ end of RNA transcript; 5′ end is often truncated	encodes reverse transcriptase	moves via an RNA intermediate that is presumably produced from a neighboring promotor	F element (*Drosophila*) L1 (human) Cin4 (maize)

These elements range in length from 2000 to about 12,000 nucleotide pairs; each family contains many members, only a few of which are listed here.

Sarepta Exhibit 1056, page 106

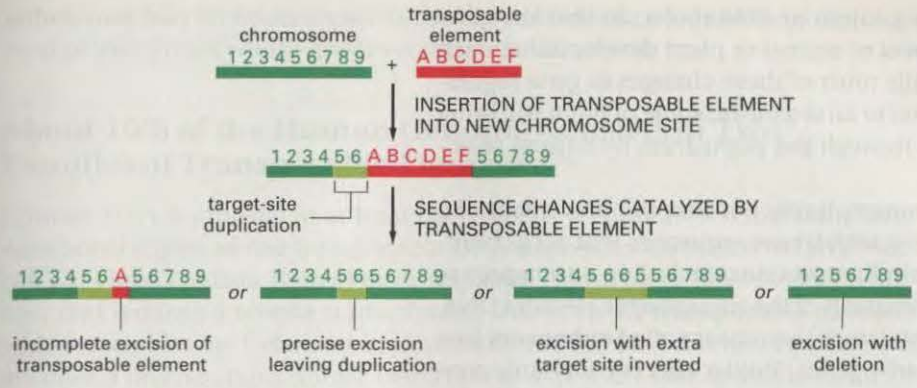


Figure 8–80 Some changes in chromosomal DNA sequences caused by transposable elements. The insertion of a transposable element always produces a short target-site duplication of the chromosomal sequence, which is generally 3 to 12 nucleotide pairs in length depending on the identity of the element. The site-specific recombination enzymes associated with the element can also cause its subsequent excision, which often fails to restore the original chromosomal DNA sequence, as in the four examples shown.

can multiply and spread from one site in a genome to a multitude of other sites, sometimes behaving as disruptive parasites.

Transposable elements seem to make up at least 10% of higher eucaryotic genomes. Although most of these elements move only very rarely, there are so many elements that their movement has a major effect on the variability of a species. More than half of the spontaneous mutations examined in *Drosophila*, for example, are due to the insertion of a transposable element in or near the mutant gene.

Mutations can occur either when an element inserts into a gene or when it exits to move elsewhere. All known transposable elements cause a short "target-site duplication" because of their mechanism of insertion (see Figure 6–70); when they exit, they generally leave behind part of this duplication—often with other local sequence changes as well (Figure 8–80). Thus, as transposable elements move in and out of chromosomes, they cause a variety of short additions and deletions of nucleotide sequences.

Transposable elements have also contributed to genome diversity in another way. When two transposable elements that are recognized by the same site-specific recombination enzyme (*transposase*) integrate into neighboring chromosomal sites, the DNA between them can become subject to transposition by the transposase. Because this provides a particularly effective pathway for the duplication and movement of exons (*exon shuffling*), these elements can help to create new genes (Figure 8–81).

Transposable Elements Often Affect Gene Regulation [52]

A DNA sequence rearrangement caused by a transposable element is often observed to alter the timing, level, or spatial pattern of expression of a nearby gene

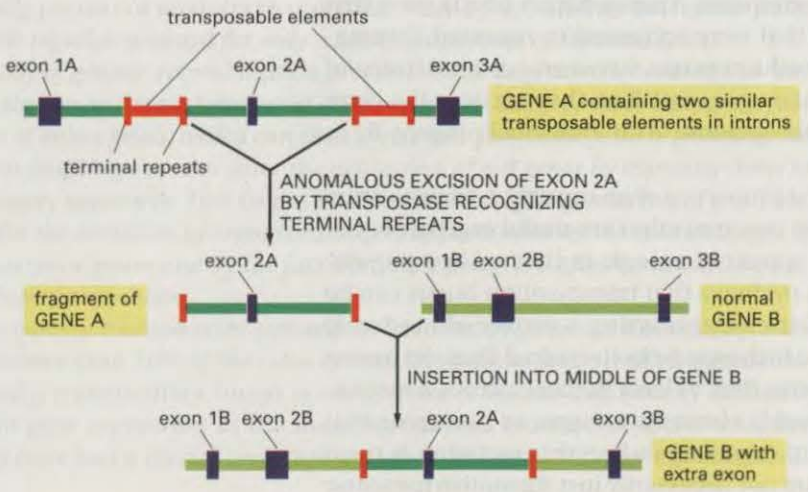


Figure 8–81 An example of the exon shuffling that can be caused by transposable elements. When two elements of the same type (*red* DNA) happen to insert near each other in a chromosome, the transposition mechanism may occasionally use the ends of two different elements (instead of the two ends of the same element) and thereby move the chromosomal DNA between them to a new chromosomal site. Since introns are very large relative to exons, the illustrated insertion of a new exon into a preexisting intron is a frequent outcome.

Sarepta Exhibit 1056, page 107

without affecting the sequence of the protein or RNA molecule that the gene encodes. This can change a subtle aspect of animal or plant development, such as the shape of an eye or a flower. While most of these changes in gene regulation would be expected to be detrimental to an organism, some of them will bring benefits and therefore tend to spread through the population by natural selection.

Effects on gene regulation are common, partly because the movement of a transposable element will generally bring with it new sequences that act as binding sites for sequence-specific DNA-binding proteins, including a transposase and the proteins that regulate the transcription of the transposable element DNA. These sequences can thereby act as regulatory sequences called **enhancers** (see p. 422) to affect the transcription of nearby genes. Similar effects commonly contribute to the evolution of cancer cells, where oncogenes can be created by the transposition of such regulatory sequences into the neighborhood of a proto-oncogene, as we discuss in Chapter 24.

The organization of higher eucaryotic genomes, with long noncoding DNA sequences interspersed with comparatively short coding sequences, provides an accommodating "playground" for the integration and excision of mobile DNA sequences. Because gene transcription can be regulated from distances that are tens of thousands of nucleotide pairs away from a promoter, many of the resulting changes in the genome would be expected to affect gene expression; by contrast, relatively few would be expected to disrupt the short exons that contain the coding sequences.

Might the vast excess of noncoding DNA in higher eucaryotes have been favored by selection during evolution because of the regulatory flexibility that it has provided to organisms with a large variety of transposable elements? What is known about the regulatory systems that control higher eucaryotic genes is consistent with this possibility. Enhancers, like exons, seem to function as separate modules, and the activity of a gene depends on a summation of the influences received at its promoter from a set of enhancers (see Figure 9–44). Transposable elements, by moving such enhancer modules around in a genome, may allow gene regulation to be optimized for the long-term survival of the organism.

Transposition Bursts Cause Cataclysmic Changes in Genomes and Increase Biological Diversity [53]

Another unique feature that distinguishes transposable elements as mutagens is their tendency to undergo long quiescent periods, during which they remain fixed in their chromosomal positions, followed by a period of intense movement. Their transposition, and therefore their mutagenic action, is activated from time to time in a few individuals in a population of organisms. Such cataclysmic changes in genomes, called *transposition bursts*, can involve near simultaneous transpositions of several types of transposable elements. Transposition bursts were first observed in developing maize plants that were subjected to repeated chromosome breakage. They also are observed in crosses between certain strains of flies—a phenomenon known as *hybrid dysgenesis*. When they occur in the germ line, they induce multiple changes in the genome of an individual progeny fly or plant.

By simultaneously changing several properties of an organism, transposition bursts increase the probability that two new traits that are useful together but of no selective value by themselves will appear in a single individual in a population. In several types of plants there is evidence that transposition bursts can be activated by a severe environmental stress, generating a variety of randomly modified progeny organisms, some of which may be better suited than the parent to survive in the new conditions. It seems that, at least in these plants, a mechanism has evolved to activate transposable elements to serve as mutagens that produce an enhanced range of variant organisms when this variation is most needed. Thus transposable elements are not necessarily just disruptive parasites;

Sarepta Exhibit 1056, page 108

rather, they may on occasion act as useful symbionts that aid the long-term survival of the species whose genomes they inhabit.

About 10% of the Human Genome Consists of Two Families of Transposable Elements [54]

Primate DNA is unusual in at least one respect: it contains a remarkably large number of copies of two transposable DNA sequences that seem to have overrun our chromosomes. Both of these sequences move by an RNA-mediated process that requires a reverse transcriptase. One is the **L1 transposable element,** which resembles the F element in *Drosophila* and the Cin4 element in maize and encodes a reverse transcriptase (see Table 8–4, p. 392). Transposable elements have generally evolved with feedback control systems that severely limit their numbers in each cell (thereby saving the cell from potential disaster); the L1 element in humans, however, constitutes about 4% of the mass of the genome.

Even more abundant is the ***Alu* sequence,** which is very short (about 300 nucleotide pairs) and moves like a transposable element, creating target-site duplications when it inserts. It was derived, however, from an internally deleted host-cell RNA gene (7SL), which encodes the RNA component of the signal-recognition particle (SRP) that functions in protein synthesis (see Figure 12–39); it is therefore not clear whether the *Alu* sequence should be considered a transposable element or an unusually mobile pseudogene. It is present in about 500,000 copies in the haploid genome and constitutes about 5% of human DNA; thus it is present on average about once every 5000 nucleotide pairs. The *Alu* DNA is transcribed from the 7SL RNA promoter, a polymerase-III promoter that is internal to the transcript, so that it carries the information necessary for its own transcription wherever it moves. It needs a reverse transcriptase encoded elsewhere, however, to transpose.

Comparisons of the sequence and locations of the L1- and *Alu*-like sequences in different mammals suggest that these sequences have multiplied to high copy numbers relatively recently (Figure 8–82). It is hard to imagine that these highly abundant sequences scattered throughout our genome have not had major effects on the expression of many nearby genes. How many of our uniquely human qualities, for example, do we owe to these parasitic elements?

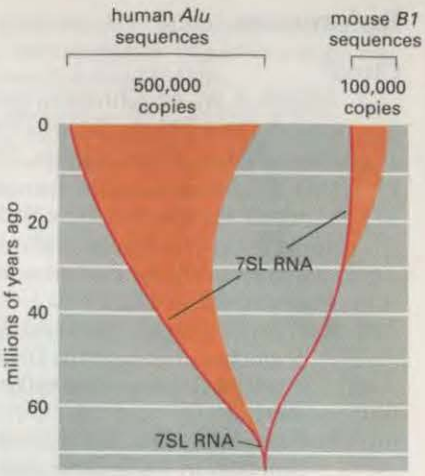


Figure 8–82 The proposed pattern of evolution of the abundant *Alu* sequence found in the human genome. A related transposable element, *B1*, is found in the mouse genome. Both of these transposable DNA sequences are thought to have evolved from the essential 7SL RNA gene. Based on the species distribution and sequence homology of these highly repeated elements, however, the major expansion in copy numbers seems to have occurred independently in mice and humans. (Adapted from P.L. Deininger and G.R. Daniels, *Trends Gen.* 2:76–80, 1986.)

Summary

The functional DNA sequences in the genomes of higher eucaryotes appear to be constructed from small genetic modules of at least two kinds. Modules of coding sequence are combined in many ways to produce proteins, whereas modules of regulatory sequences are scattered throughout long stretches of noncoding sequences and regulate the expression of genes. Both the coding sequences and the regulatory sequences are typically present in modules that are less than a few hundred nucleotide pairs long, which together account for only a small proportion of the total DNA.

A variety of genetic-recombination processes occur in genomes, causing the random duplication and translocation of DNA sequences. Some of these changes create duplicates of entire genes, which can then evolve new functions. Others produce new proteins by shuffling exons or alter the expression of old genes by exposing them to new regulatory sequences. This type of DNA sequence shuffling, which is of great importance for the evolution of organisms, is greatly facilitated by the split structure of higher eucaryotic genes and by the fact that these genes are often controlled by distant regulatory sequences.

Many types of transposable elements are present in genomes. Collectively, they constitute more than 10% of the mass of both* Drosophila *and vertebrate genomes. Occasionally, transposition bursts occur in germ cells and cause many heritable changes in gene expression in the same individual. Transposable elements are thought to have had a special evolutionary role in the generation of organismal diversity.

Sarepta Exhibit 1056, page 109

References

Cited

1. Adolph, K.W., ed. Chromosomes and Chromatin, Vols. 1–3. Boca Raton, FL: CRC Press, 1988.

   Felsenfeld, G. DNA. *Sci. Am.* 253(4):58–67, 1985.

   Hsu, T.C. Human and Mammalian Cytogenetics: A Historical Perspective. New York: Springer-Verlag, 1979.

   Stewart, A. The functional organization of chromosomes and the nucleus—a special issue. *Trends Genet.* 6:377–379, 1990.

2. Burke, D.T.; Carle, G.F.; Olson, M.V. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236:806–812, 1987.

   DePamphilis, M.L. Origins of DNA replication that function in eucaryotic cells. *Curr. Opin. Cell Biol.* 5:434–441, 1993.

   Murray, A.W. Chromosome structure and behavior. *Trends Biochem. Sci.* 10:112–115, 1985.

   Price, C.M. Centromeres and telomeres. *Curr. Opin. Cell Biol.* 4:379–384, 1992.

3. Gall, J.G. Chromosome structure and the C-value paradox. *J. Cell Biol.* 91:3s–14s, 1981.

   Ohta, T.; Kimura, M. Functional organization of genetic material as a product of molecular evolution. *Nature* 233:118–119, 1971.

4. Ruby, S.W.; Abelson, J. Pre-mRNA splicing in yeast. *Trends Genet.* 7:79–85, 1991.

5. Wilson, A.C.; Ochman, H.; Prager, E.M. Molecular time scale for evolution. *Trends Genet.* 3:241–247, 1987.

6. Grunstein, M. Histones as regulators of genes. *Sci. Am.* 267(4):68–74B, 1992.

   Isenberg, I. Histones. *Annu. Rev. Biochem.* 48:159–191, 1979.

   Smith, M.M. Histone structure and function. *Curr. Opin. Cell Biol.* 3:429–437, 1991.

   Wells, D.E. Compilation analysis of histones and histone genes. *Nucleic Acids Res.* 14:r119–r149, 1986.

7. Chromatin. *Cold Spring Harbor Symp. Quant. Biol.*, Vol. 42, 1978.

   Kornberg, R.D.; Klug, A. The nucleosome. *Sci. Am.* 244(2):52–64, 1981.

   McGhee, J.D.; Felsenfeld, G. Nucleosome structure. *Annu. Rev. Biochem.* 49:1115–1156, 1980.

   Richmond, T.J.; Finch, J.T.; Rushton, B.; Rhodes, D.; Klug, A. Structure of the nucleosome core particle at 7Å resolution. *Nature* 311:532–537, 1984.

8. Gross, D.S.; Garrard, W.T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* 57:159–198, 1988.

   Simpson, R.T. Nucleosome position *in vivo* and *in vitro*. *Bioessays* 4:172–176, 1986.

   Svaren, J.; Chalkley, R. The structure and assembly of active chromatin. *Trends Genet.* 6:52–56, 1990.

   Travers, A.A. DNA bending and nucleosome positioning. *Trends Biochem. Sci.* 12:108–112, 1987.

   Zhang, L.; Gralla. J.D. *In situ* nucleoprotein structure involving origin-proximal SV40 DNA control elements. *Nucleic Acids Res.* 18:1797–1803, 1990.

9. Hansen, J.C.; Ausio, J. Chromatin dynamics and the modulation of genetic activity. *Trends Biochem. Sci.* 17:187–191, 1992.

   Pederson, D.S.; Thoma, F.; Simpson, R. Core particle, fiber, and transcriptionally active chromatin structure. *Annu. Rev. Cell Biol.* 2:117–147, 1986.

   Swedlow, J.R.; Agard, D.A.; Sedat, J.W. Chromosome structure inside the nucleus. *Curr. Opin. Cell Biol.* 5:412–416, 1993.

   Zlatanova, J.; Yaneva, J. Histone H1–DNA interactions and their relation to chromatin structure and function. *DNA Cell Biol.* 10:239–248, 1991.

10. Bellini, M.; Lacroix, J.-C.; Gall, J.G. A putative zinc-binding protein on lampbrush chromosome loops. *EMBO J.* 12:107–114, 1993.

    Bostock, C.J.; Sumner, A.T. The Eucaryotic Chromosome, pp. 347–374. Amsterdam: North-Holland, 1978.

    Callan, H.G. Lampbrush chromosomes. *Proc. R. Soc. Lond. (Biol.)* 214:417–448, 1982.

    Reznik, N.A.; Yampol, G.P.; Kiseleva, E.V.; Khristolyubova, N.B.; Gruzdev, A.D. Functional and structural units in the chromomere. *Genetica* 83:293–299, 1991.

11. Agard, D.A.; Sedat, J.W. Three-dimensional architecture of a polytene nucleus. *Nature* 302:676–681, 1983.

    Beermann, W. Chromosomes and genes. In Developmental Studies on Giant Chromosomes (W. Beermann, ed.), pp. 1–33. New York: Springer-Verlag, 1972.

    Zhimulev, I.F.; Belyaeva, E.S. Chromomeric organization of polytene chromosomes. *Genetica* 85:65–72, 1991.

12. Ashburner, M.; Chihara, C.; Meltzer, P.; Richards, G. Temporal control of puffing activity in polytene chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38:655–662, 1974.

    Lamb, M.M.; Daneholt, B. Characterization of active transcription units in Balbiani rings of *Chironomous tentans*. *Cell* 17:835–848, 1979.

    Thummel, C.S. Puffs and gene regulation—molecular insights into the *Drosophila* ecdysone regulatory hierarchy. *Bioessays* 12:561–568, 1990.

13. Bossy, B.; Hall, L.M.; Spierer, P. Genetic activity along 315 kb of the *Drosophila* chromosome. *EMBO J.* 3:2537–2541, 1984.

    Friedman, T.B.; Owens, K.N.; Burnett, J.B.; Saura, A.O.; Wallrath, L.L. The faint band/interband region 28C2 to 28C4–5(–) of the *Drosophila melanogaster* salivary gland polytene chromosomes is rich in transcripts. *Mol. Gen. Genet.* 226:81–87, 1991.

    Judd, B.H.; Young, M.W. An examination of the one cistron: one chromomere concept. *Cold Spring Harbor Symp. Quant. Biol.* 38:573–579, 1974.

14. Croston, G.E.; Kadonaga, J.T. Role of chromatin structure in the regulation of transcription by RNA polymerase II. *Curr. Opin. Cell Biol.* 5:417–423, 1993.

    Nacheva, G.A.; Guschin, D.Y.; Preobrazhenskaya, O.V.; et al. Change in the pattern of histone binding to DNA upon transcriptional activation. *Cell* 58:27–36, 1989.

    Svaren, J.; Horz, W. Histones, nucleosomes and transcription. *Curr. Opin. Genet. Dev.* 3:219–225, 1993.

15. Allis, C.D., et al. hv1 is an evolutionarily conserved H2A variant that is preferentially associated with active genes. *J. Biol. Chem.* 261:1941–1948, 1986.

    Bradbury, E.M. Reversible histone modifications and the chromosome cell cycle. *Bioessays* 14:9–16, 1992.

Sarepta Exhibit 1056, page 110

Tremethick, D.J.; Drew, H.R. High mobility group proteins 14 and 17 can space nucleosomes *in vitro*. *J. Biol. Chem.* 268:11389–11393, 1993.

Turner, B.M. Histone acetylation and control of gene expression. *J. Cell Sci.* 99:13–20, 1991.

16. Brown, S.W. Heterochromatin. *Science* 151:417–425, 1966.

Pimpinelli, S.; Bonaccorsi, S.; Gatti, M.; Sandler, L. The peculiar genetic organization of *Drosophila* heterochromatin. *Trends Genet.* 2:17–20, 1986.

17. Georgiev, G.P.; Nedospasov, S.A.; Bakayev, V.V. Supranucleosomal levels of chromatin organization. In The Cell Nucleus (H. Busch, ed.), Vol. 6, pp. 3–34. New York: Academic Press, 1978.

Kitsberg, D.; Selig, S.; Cedar, H. Chromosome structure and eukaryotic gene organization. *Curr. Opin. Genet. Dev.* 1:534–537, 1991.

Marsden, M.; Laemmli, U.K. Metaphase chromosome structure: evidence for a radial loop model. *Cell* 17:849–858, 1979.

18. Bickmore, W.A.; Sumner, A.T. Mammalian chromosome banding—an expression of genome organization. *Trends Genet.* 5:144–148, 1989.

Holmquist, G. DNA sequences in G-bands and R-bands. In Chromosomes and Chromatin Structure (K.W. Adolph, ed.), Vol. 2, pp. 75–122. Boca Raton, FL: CRC Press, 1988.

Lewin, B. Gene Expression, Vol. 2: Eucaryotic Chromosomes, 2nd ed., pp. 428–440. New York: Wiley, 1980.

19. DePamphilis, M.L. Origins of DNA replication that function in eukaryotic cells. *Curr. Opin. Cell Biol.* 5:434–441, 1993.

Marahrens, Y.; Stillman, B. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* 255:817–823, 1992.

Struhl, K.; Stinchcomb, D.T.; Sherer, S.; Davis. R.W. High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. *Proc. Natl. Acad. Sci. USA* 76:1035–1039, 1979.

20. Dodson, M.; Dean, F.B.; Bullock, P.; Echols, H.; Hurwitz, J. Unwinding of duplex DNA from the SV40 origin of replication by T antigen. *Science* 238:964–967, 1987.

Stillman, B.; Bell, S.P.; Dutta, A.; Marahrens, Y. DNA replication and the cell cycle. *Ciba Found. Symp.* 170:147–156; discussion 156–160, 1992.

Tsurimoto, T.; Stillman, B. Replication factors required for SV40 DNA replication *in vitro*. I. DNA structure-specific recognition of a primer-template junction by eukaryotic DNA polymerases and their accessory proteins. *J. Biol. Chem.* 266:1950–1960, 1991.

21. Bonifer, C.; Hecht, A.; Saueressig, H.; Winter, D.M.; Sippel, A.E. Dynamic chromatin: the regulatory domain organization of eukaryotic gene loci. *J. Cell. Biochem.* 47:99–108, 1991.

Huberman, J.A.; Riggs, A.D. On the mechanism of DNA replication in mammalian chromosomes. *J. Mol. Biol.* 32:327–341, 1968.

22. Craig, J.M.; Bickmore, W.A. Chromosome bands—flavours to savour. *Bioessays* 15:349–354, 1993.

Stubblefield, E. Analysis of the replication pattern of Chinese hamster chromosomes using 5-bromodeoxyuridine suppression of 33258 Hoechst fluorescence. *Chromosoma* 53:209–221, 1975.

23. Lima-de-Faria, A.; Jaworska, H. Late DNA synthesis in heterochromatin. *Nature* 217:138–142, 1968.

24. Bickmore, W.A.; Sumner, A.T. Mammalian chromosome banding—an expression of genome organization. *Trends Genet.* 5:144–148, 1989.

25. Hofmann, A.; Montag, M.; Steinbeisser, H.; Trendelenburg, M.F. Plasmid and bacteriophage lambda-DNA show differential replication characteristics following injection into fertilized eggs of *Xenopus laevis*: dependence on period and site of injection. *Cell Differ. Dev.* 30:77–85, 1990.

Mechali, M.; Kearsey, S. Lack of specific sequence requirement for DNA replication in *Xenopus* eggs compared with high sequence specificity in yeast. *Cell* 38:55–64, 1984.

Riggs, A.D. DNA methylation and late replication probably aid cell memory, and type I DNA reeling could aid chromosome folding and enhancer function. *Phil. Trans. R. Soc. Lond. (Biol.)* 326:285–297, 1990.

26. Blow, J.J. Preventing re-replication of DNA in a single cell cycle: evidence for a replication licensing factor. *J. Cell Biol.* 122:993–1002, 1993.

Coverley, D.; Downes, C.S.; Romanowski, P.; Laskey, R.A. Reversible effects of nuclear membrane permeabilization on DNA replication: evidence for a positive licensing factor. *J. Cell Biol.* 122:985–992, 1993.

Rao, P.N., Johnson, R.T. Mammalian cell fusion: studies on the regulation of DNA synthesis and mitosis. *Nature* 225:159–164, 1970.

Wanka, F. Control of eukaryotic DNA replication at the chromosomal level. *Bioessays* 13:613–618, 1991.

27. Almouzni, G.; Clark, D.J.; Mechali, M.; Wolffe, A.P. Chromatin assembly on replicating DNA *in vitro*. *Nucleic Acids Res.* 18:5767–5774, 1990.

Randall, S.K.; Kelly, T.J. The fate of parental nucleosomes during SV40 DNA replication. *J. Biol. Chem.* 267:14259–14265, 1992.

Russev, G.; Hancock, R. Assembly of new histones into nucleosomes and their distribution in replicating chromatin. *Proc. Natl. Acad. Sci. USA* 79:3143–3147, 1982.

28. Blackburn, E.H. Telomeres. *Trends Biochem. Sci.* 16:378–381, 1991.

Blackburn, E.H.; Szostak, J.W. The molecular structure of centromeres and telomeres. *Annu. Rev. Biochem.* 53:163–194, 1984.

Vogt, P. Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code." *Hum. Genet.* 84:301–336, 1990.

29. Chamberlin, M. Bacterial DNA-dependent RNA polymerases. In The Enzymes, 3rd ed. (P. Boyer, ed.), Vol. 15B, pp. 61–108. New York: Academic Press, 1982.

Gill, S.C.; Yager, T.D.; von Hippel, P.H. Thermodynamic analysis of the transcription cycle in *E. coli*. *Biophys. Chem.* 37:239–250, 1990.

Kerppola, T.K.; Kane, C.M. RNA polymerase: regulation of transcript elongation and termination. *FASEB J.* 5:2833–2842, 1991.

30. Chambon, P. Eucaryotic nuclear RNA polymerases. *Annu. Rev. Biochem.* 44:613–638, 1975.

Geiduschek, E.P.; Tocchini-Valentini, G.P. Transcription by RNA polymerase III. *Annu. Rev. Biochem.* 57:873–914, 1988.

Sentenac, A. Eucaryotic RNA polymerases. *CRC Crit. Rev. Biochem.* 18:31–91, 1985.

Young, R.A. RNA polymerase II. *Annu. Rev. Biochem.* 60:689–715, 1991.

31. Foe, V.E.; Wilkinson, L.E.; Laird, C.D. Comparative organization of active transcription units in *Oncopeltus fasciatus*. *Cell* 9:131–146, 1976.

Lewin, B. Gene Expression, Vol. 2: Eucaryotic Chromosomes, 2nd ed., pp. 708–719. New York: Wiley, 1980.

Miller, O.L. The nucleolus, chromosomes, and visualization of genetic activity. *J. Cell Biol.* 91:15s–27s, 1981.

32. Nevins, J.R. The pathway of eukaryotic mRNA formation. *Annu. Rev. Biochem.* 52:441–466, 1983.

Proudfoot, N.J. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem. Sci.* 14:105–110, 1989.

Takgaki, V.; Ryner, L.C.; Manley, J.L. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell* 52:731–742, 1988.

Wickens, M. How the messenger got its tail: addition of poly(A) in the nucleus. *Trends Biochem. Sci.* 15:277–281, 1990.

33. Crick, F. Split genes and RNA splicing. *Science* 204:264–271, 1979.

Dreyfuss, G.; Matunis, M.J.; Piñol-Roma, S.; Burd, C.G. hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* 62:289–321, 1993.

Hickey, D.A.; Benkel, B.F.; Abukashawa, S.M. A general model for the evolution of nuclear pre-mRNA introns. *J. Theor. Biol.* 137:41–53, 1989.

34. Dreyfuss, G.; Swanson, M.S.; Piñol-Roma, S. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem. Sci.* 13:86–91, 1988.

Guthrie, C. Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science* 253:157–163, 1991.

Guthrie, C.; Patterson, B. Spliceosomal snRNAs. *Annu. Rev. Genet.* 22:387–419, 1988.

Samarina, O.P.; Krichevskaya, A.A.; Georgiev, G.P. Nuclear ribonucleoprotein particles containing messenger ribonucleic acid. *Nature* 210:1319–1322, 1966.

Steitz, J.A. "Snurps." *Sci. Am.* 258(6):56–63, 1988.

35. Balvay, L.; Libri, D.; Fiszman, M.Y. Pre-mRNA secondary structure and the regulation of splicing. *Bioessays* 15:165–169, 1993.

Padgett, R.A.; Grabowski, P.J.; Konarska, M.M.; Seiler, S.; Sharp, P.A. Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* 55:1119–1150, 1986.

36. Aebi, M.; Weissman, C. Precision and orderliness in splicing. *Trends Genet.* 3:102–107, 1987.

Goguel, V.; Liao, X.; Rymond, B.C.; Rosbash, M. U1 snRNP can influence 3′-splice site selection as well as 5′-splice site selection. *Genes Dev.* 5:1430–1438, 1991.

37. Orkin, S.H.; Kazazian, H.H. The mutation and polymorphism of the human β-globin gene and its surrounding DNA. *Annu. Rev. Genet.* 18:131–171, 1984.

Vidaud, M.; Gattoni, R.; Stevenin, J.; et al. A 5′ splice-region G→C mutation in exon 1 of the human beta-globin gene inhibits pre-mRNA splicing: a mechanism for β^+-thalassemia. *Proc. Natl. Acad. Sci. USA* 86:1041–1045, 1989.

38. Cech, T.R. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* 44:207–210, 1986.

Saldanha, R.; Mohr, G.; Belfort, M.; Lambowitz, A.M. Group I and group II introns. *FASEB J.* 7:15–24, 1993.

39. Brown, J.D.; Plumpton, M.; Beggs, J.D. The genetics of nuclear pre-mRNA splicing: a complex story. *Antonie Van Leeuwenhoek* 62:35–46, 1992.

Green, M.R. Pre-mRNA processing and mRNA nuclear export. *Curr. Opin. Cell Biol.* 1:519–525, 1989.

Jimenez-Garcia, L.F.; Spector. D.L. *In vivo* evidence that transcription and splicing are coordinated by a recruiting mechanism. *Cell* 73:47–59, 1993.

Spector, D.L. Nuclear organization of pre-mRNA processing. *Curr. Opin. Cell Biol.* 5:442–447, 1993.

40. Long, E.O.; Dawid, I.B. Repeated genes in eucaryotes. *Annu. Rev. Biochem.* 49:727–764, 1980.

Miller, O.L. The nucleolus, chromosomes, and visualization of genetic activity. *J. Cell Biol.* 91:15s–27s, 1981.

Williams, S.M.; Robbins, L.G. Molecular genetic analysis of *Drosophila* rDNA arrays. *Trends Genet.* 8:335–340, 1992.

41. Fischer, D.; Weisenberger, D.; Scheer, U. Assigning functions to nucleolar structures. *Chromosoma* 101:133–140, 1991.

Hadjiolov, A.A. The Nucleolus and Ribosome Biogenesis. New York: Springer-Verlag, 1985.

Jordan, E.G.; Cullis, C.A., eds. The Nucleolus. Cambridge, UK: Cambridge University Press, 1982.

Larson, D.E.; Zahradka, P.; Sells, B.H. Control points in eucaryotic ribosome biogenesis. *Biochem. Cell Biol.* 69:5–22, 1991.

42. Anastassova-Kristeva, M. The nucleolar cycle in man. *J. Cell Sci.* 25:103–110, 1977.

McClintock, B. The relation of a particular chromosomal element to the development of the nucleoli in Zea Mays. *Z. Zellforsch. Mikrosk. Anat.* 21:294–323, 1934.

43. Cremer, T., et al. Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Hum. Genet.* 60:46–56, 1980.

Haaf, T.; Schmid, M. Chromosome topology in mammalian interphase nuclei. *Exp. Cell Res.* 192:325–332, 1991.

Hochstrasser, M.; Sedat, J.W. Three-dimensional organization of *Drosophila melanogaster* interphase nuclei. II. Chromosome spatial organization and gene regulation. *J. Cell Biol.* 104:1471–1483, 1987.

Manuelidis, L. A view of interphase chromosomes. *Science* 250:1533–1540, 1990.

44. Dessev, G.N. Nuclear envelope structure. *Curr. Opin. Cell Biol.* 4:430–435, 1992.

Gasser, S.M.; Laemmli, U.K. A glimpse at chromosome order. *Trends Genet.* 3:16–22, 1987.

Gerace, L.; Burke, B. Functional organization of the nuclear envelope. *Annu. Rev. Cell Biol.* 4:335–374, 1988.

Sarepta Exhibit 1056, page 112

Newport, J.W.; Forbes, D.J. The nucleus: structure, function, and dynamics. *Annu. Rev. Biochem.* 56:535–566, 1987.

45. Doolittle, R.F. Proteins. *Sci. Am.* 253(4):88–99, 1985.

Holland, S.K.; Blake, C.C. Proteins, exons, and molecular evolution. *Biosystems* 20:181–206, 1987.

Maeda, N.; Smithies, O. The evolution of multigene families: human haptoglobin genes. *Annu. Rev. Genet.* 20:81–108, 1986.

46. Kourilsky, P. Molecular mechanisms for gene conversion in higher cells. *Trends Genet.* 2:60–63, 1986.

Roth, D.B.; Porter, T.N.; Wilson, J.H. Mechanisms of nonhomologous recombination in mammalian cells. *Mol. Cell. Biol.* 5:2599–2607, 1985.

Smith, G.P. Evolution of repeated DNA sequences by unequal crossover. *Science* 191:528–535, 1976.

Stark, G.R.; Wahl, G.M. Gene amplification. *Annu. Rev. Biochem.* 53:447–491, 1984.

47. Dickerson, R.E.; Geis, I. Hemoglobin: Structure, Function, Evolution, and Pathology. Menlo Park, CA: Benjamin-Cummings, 1983.

Efstratiadis, A., et al. The structure and evolution of the human β-globin gene family. *Cell* 21:653–668, 1980.

48. Anderson, C.L.; Carew, E.A.; Powell, J.R. Evolution of the Adh locus in the *Drosophila willistoni* group: the loss of an intron, and shift in codon usage. *Mol. Biol. Evol.* 10:605–618, 1993.

Doolittle, W.F. RNA-mediated gene conversion? *Trends Genet.* 1:64–65, 1985.

Nyberg, A.M.; Cronhjort, M.B. Intron evolution: a statistical comparison of two models. *J. Theor. Biol.* 157:175–190, 1992.

49. Britten, R.J.; Kohne, D.E. Repeated sequences in DNA. *Science* 161:529–540, 1968.

Jelinek, W.R.; Schmid, C.W. Repetitive sequences in eukaryotic DNA and their expression. *Annu. Rev. Biochem.* 51:813–844, 1982.

50. Craig-Holmes, A.P.; Shaw, M.W. Polymorphism of human constitutive heterochromatin. *Science* 174:702–704, 1971.

Hsu, T.C. Human and Mammalian Cytogenetics: A Historical Perspective. New York: Springer-Verlag, 1979.

John, B.; Miklos, G.L.G. Functional aspects of satellite DNA and heterochromatin. *Int. Rev. Cytol.* 58:1–114, 1979.

Orgel, L.E.; Crick, F.H.C. Selfish DNA: the ultimate parasite. *Nature* 284:604–607, 1980.

51. Berg, D.E.; Howe, M.M., eds. Mobile DNA. Washington, DC: American Society for Microbiology, 1989.

Döring, H.-P.; Starlinger, P. Molecular genetics of transposable elements in plants. *Annu. Rev. Genet.* 20:175–200, 1986.

Finnegan, D.J. Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5:103–107, 1989.

McClintock, B. Controlling elements and the gene. *Cold Spring Harbor Symp. Quant. Biol.* 21:197–216, 1956.

52. Coen, E.S.; Carpenter, R. Transposable elements in *Antirrhinum majus*: generators of genetic diversity. *Trends Genet.* 2:292–296, 1986.

Georgiev, G.P. Mobile genetic elements in animal cells and their biological significance. *Eur. J. Biochem.* 145:203–220, 1984.

O'Kane, C.J.; Gehring, W.J. Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 84:9123–9127, 1987.

53. Gerasimova, T.I.; Mizrokhi, L.J.; Georgiev, G.P. Transposition bursts in genetically unstable *Drosophila melanogaster*. *Nature* 309:714–716, 1984.

McClintock, B. The significance of responses of the genome to challenge. *Science* 226:792–801, 1984.

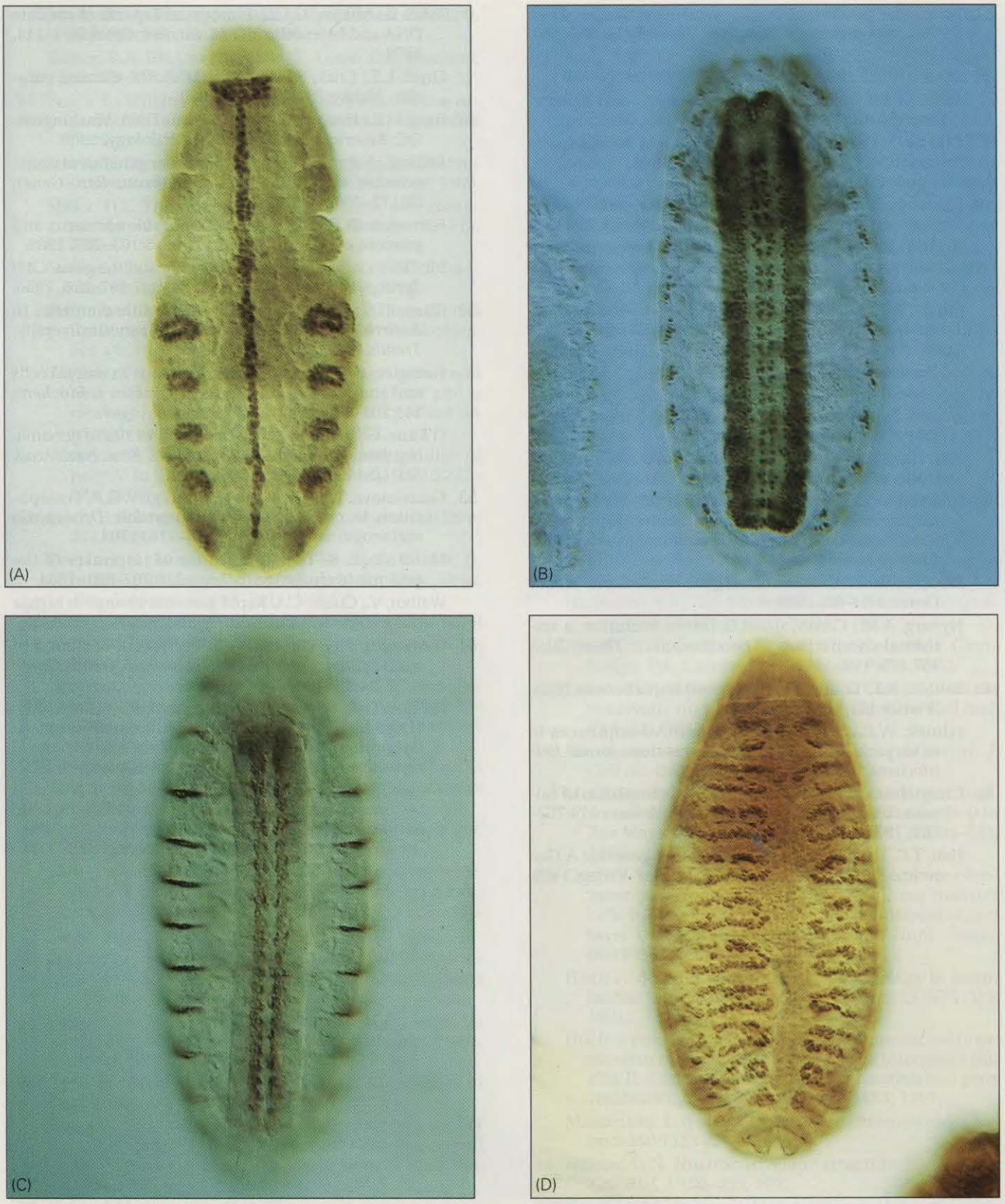
Walbot, V.; Cullis, C.A. Rapid genomic change in higher plants. *Annu. Rev. Plant Physiol.* 36:367–396, 1985.

54. Deininger, P.L.; Daniels, G.R. The recent evolution of mammalian repetitive DNA elements. *Trends Genet.* 2:76–80, 1986.

Ruffner, D.E.; Sprung, C.N.; Minghetti, P.P.; Gibbs, P.E.; Dugaiczyk, A. Invasion of the human albumin-α-fetoprotein gene family by Alu, Kpn, and two novel repetitive DNA elements. *Mol. Biol. Evol.* 4:1–9, 1987.

Weiner, A.M.; Deininger, P.L.; Estratiadis, A. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* 55:631–661, 1986.

Sarepta Exhibit 1056, page 113



Position-dependent patterns of gene expression in four different transgenic *Drosophila* embryos. Each embryo has a single copy of a bacterial β-galactosidase gene inserted into its genome. Depending on the chromosomal insertion site, nearby *Drosophila* enhancers cause the gene to be expressed in a pattern that corresponds either to (A) trachea and mesectoderm, (B) nervous system, (C) glial cells of the central nervous system plus a segmentally repeating pattern, or (D) muscle. (Courtesy of Yuh-Nung Jan.)