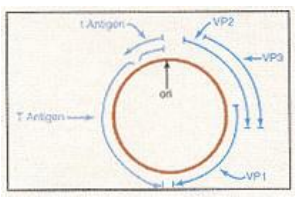
Chapter 17—
Protein Synthesis:
Translation and Posttranslational Modifications

Dohn Glitz



17.1— Overview

Protein biosynthesis is also called **translation** since it involves the biochemical translation of information from the four-letter language and structure of nucleic acids into the 20-letter language and structure of proteins. This process has many requirements: an informational messenger RNA molecule that is exported from the nucleus, several "bilingual" transfer RNA species that read the message, ribosomes that serve as catalytic and organizational centers, a variety of protein factors, and energy. Polypeptides are formed by the sequential addition of amino acids in the specific order determined by the information carried in the nucleotide sequence of the mRNA. The protein is often then matured or processed by a variety of modifications. These may target it to a specific intracellular location or for secretion from the cell, or they may modulate its activity or function. These complex processes are carried out with considerable speed and extreme precision. Levels of translation are regulated, both globally and for specific proteins. Finally, when a protein becomes nonfunctional or is no longer needed, it is degraded and its amino acids are catabolized or recycled into new proteins.

Cells vary in their need and ability to synthesize proteins. At one extreme, terminally differentiated red blood cells have a life span of about 120 days, have no nuclei, do not divide, and do not synthesize proteins because they lack the components of the biosynthetic apparatus. Nondividing cells need to maintain levels of enzymes and other proteins and carry out limited protein synthesis. Growing and dividing cells must synthesize much larger amounts of protein. Finally, some cells synthesize proteins for export as well as for their own use. For example, liver cells synthesize large numbers of enzymes needed for their many metabolic pathways as well as proteins for export, including serum albumin, the major protein of blood plasma or serum. Liver cells are protein factories that are particularly rich in the machinery for synthesis of proteins.

17.2— Components of the Translational Apparatus

Messenger RNA Is the Carrier of Information Present in DNA

Genetic information is stored and transmitted in the nucleotide sequences of DNA. Selective expression of this information requires its transcription into mRNA that carries specific and precise messages from the nuclear "data bank" to the cytoplasmic sites of protein synthesis. In eukaryotes, the messengers, mRNAs, are usually synthesized as significantly larger precursor molecules that are processed prior to export from the nucleus. Eukaryotic mRNA in the cytosol has several identifying characteristics. It is almost always **monocistronic,** that is, encoding a single polypeptide. The 5 end is capped with a specific structure consisting of **7-methylguanosine** linked through a 5 -triphosphate bridge to the 5 end of the messenger sequence (see p. 704). A 5 -nontranslated region, which may be short or up to a few hundred nucleotides in length, separates the cap from the **translational initiation signal,** an **AUG** codon. Usually, but not always, this is the first AUG sequence encountered as the message is read 5 3 . Uninterrupted sequences that specify a unique polypeptide sequence follow the initiation signal until a specific translation termination signal is reached. This is followed by a 3 -untranslated sequence, usually about 100 nucleotides in length, before the mRNA is terminated by a 100- to 200-nucleotide long polyadenylate tail.

Prokaryotic mRNA differs from eukaryotic mRNA in that the 5 terminus is not capped but retains a terminal triphosphate from initiation of its synthesis by RNA polymerase. Also, most messengers are **polycistronic,** that is, encoding several polypeptides, and include more than one initiation AUG sequence. A

ribosome-positioning sequence is located about 10 nucleotides upstream of a valid AUG initiation signal. An untranslated sequence follows the termination signal, but there is no polyadenylate tail.

Ribosomes Are Workbenches for Protein Biosynthesis

Proteins are assembled on particles called ribosomes. These have two dissimilar subunits, each of which contains RNA and many proteins. With one exception, each protein is present in a single copy per ribosome, as is each RNA species. The composition of major ribosome types is shown in Table 17.1, and characteristics of their RNAs are given in Table 16.1.

Ribosome architecture has been conserved in evolution. The similarities between ribosomes and subunits from different sources are more obvious than the differences, and functional roles for each subunit are well defined. Details of ribosome structure and its relationship to function have been learned using many techniques. Overall size and shape can be determined by electron microscopy. The location of many ribosomal proteins, some elements of the RNA, and functional sites on each subunit have been determined by electron microscopy of subunits that are complexed with antibodies against a single ribosomal component. The antibody molecule serves as a physical pointer to the site on the ribosome. Further structural information has been obtained from chemical cross-linking, which identifies near neighbors within the structure, and from neutron diffraction measurements, which quantitate the distances between pairs of proteins. Ribosomes have been crystallized and X-ray structural determination is under way. Sequence comparisons and chemical, immunological, and enzymatic probes give information about RNA conformation. Correlations of structural data with functional measurements in protein synthesis have allowed development of models, such as that in Figure 17.1, that link ribosome morphology to various functions in translation. Each subunit has an RNA core, folded into a specific three-dimensional structure, upon which proteins are positioned through protein–RNA and protein–protein interactions.

Many of these experiments were possible because prokaryotic ribosomes can **self-assemble;** that is, the native structures can be reconstituted from mixtures of purified individual proteins and RNAs. Reconstitution of subunits

TABLE 17.1 Ribosome Classification and Composition

	Subunits		
Ribosome Source	***Monomer Size***	***Small***	***Large***
Eukaryotes			
Cytosol	80S	40S:	60S:
		34 proteins	50 proteins
		18S RNA	28S, 5.8S, 5S RNAs
Mitochondria			
Animals	55S–60S	30S–35S:	40–45S:
		12S RNA	16S RNA
		70–100 proteins	
Higher plants	77S–80S	40S:	60S:
		19S RNA	25S, 5S RNAs
		70–75 proteins	
Chloroplasts	70S	30S:	50S:
		20–24 proteins	34–38 proteins
		16S RNA	23S, 5S, 4.5S RNAs
Prokaryotes			
Escherichia coli	70S	30S:	50S:
		21 proteins	34 proteins
		16S RNA	23S, 5S RNAs

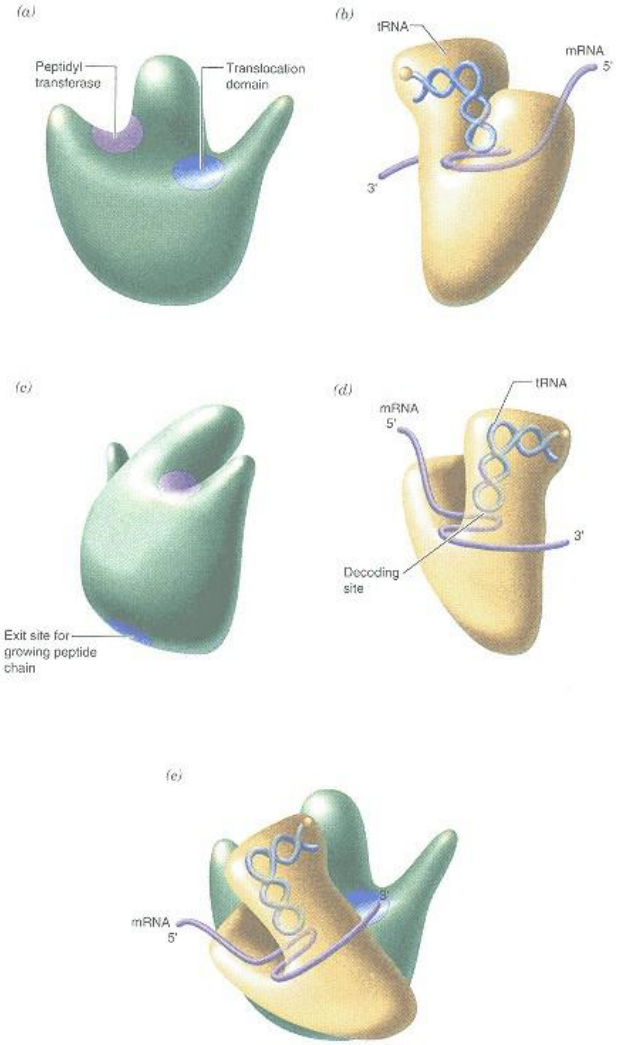


Figure 17.1
Ribosome structure and functional sites.
Top row shows the faces of each subunit that interact in the functional ribosome. In (a) the large subunit is shown; note that sites of peptide bond formation and of binding of the elongation factors are on opposite sides of the bulbous "central protuberance. " The arm-like structure is somewhat flexible or mobile and is seldom visualized in complete ribosomes.
In (b) the small subunit is shown with a "platform" or ledge protruding toward the reader. mRNA and tRNA interact in a "decoding site," deep in the cleft between the platform and subunit body. The orientation of mRNA and tRNA is depicted, although their interaction in the decoding site is obscured by the platform.
In (c) the large subunit has been rotated 90° and the arm projects into the page. The exit site near the base of the subunit is where newly synthesized protein emerges from the subunit. This area of the subunit is in contact with membranes in the "bound" ribosomes of rough endoplasmic reticulum. The site of peptide bond formation, the peptidyltransferase center, is distant from the exit site; the growing peptide passes through a groove or tunnel in the ribosome to reach the exit site.
In (d) the small subunit has been rotated 90° such that the platform projects toward the dish-like face of the large subunit and the cleft is apparent.
In (e) subunits have been brought together to show their relative orientation in the ribosome. Note that tRNA bound by the small subunit is oriented so that the aminoacyl acceptor end is near the peptidyltransferase while the translocational domain (where EF-1α and EF-2 bind) is near the decoding region and the area in which mRNA enters the complex. Drawings are based on electron microscopy of stained and unstained, frozen ribosomes. The latter technique preserves native structure and, perhaps along with X-ray crystallography, should lead to a more detailed and complete model of the ribosome.

from mixtures in which a single component is omitted or modified can show, for example, if a given protein is required for assembly of the subunit or for some specific function. An assembly map for large ribosomal subunits of *Escherichia coli* is shown in Figure 17.2. Total reconstitution of subunits from eukaryotes has not yet been achieved but the general conclusions about how ribosomes function, although determined using bacterial ribosomes, are fully applicable to eukaryotic systems.

Ribosomes are organized in two additional ways. First, several ribosomes often translate a single mRNA molecule simultaneously. Purified mRNA-linked

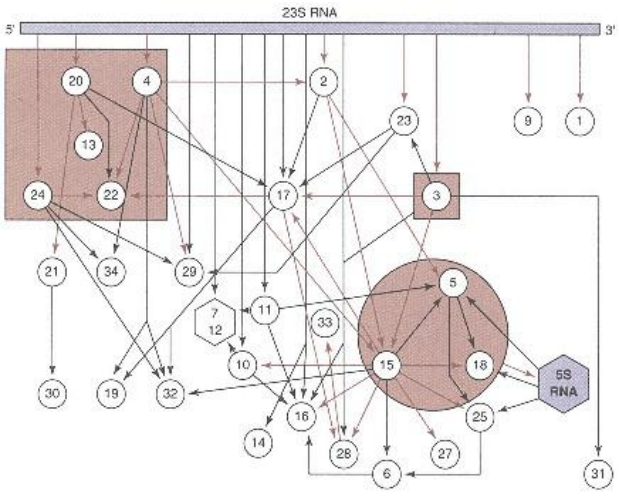


Figure 17.2
Assembly map of the large ribosomal subunit of *E. coli*.
The heavy bar at the top represents the 23S rRNA, and the individual ribosomal proteins are identified by numbers in circles. Arrows that connect components indicate their interaction. Red arrows from RNA to protein indicate that the protein binds directly and strongly to RNA, while black arrows indicate a weaker interaction. Similarly, red arrows between proteins show a strong binding dependence and black arrows show a lesser dependency. For example, protein L4 binds RNA strongly; it then strongly stimulates binding of proteins L2, L22, and L29. Protein L2 in turn stimulates binding of proteins L5 and L15. Proteins L5, L15, and L18 are essential for binding 5S RNA. Proteins within the boxes are required for a conformational transition that occurs during assembly. Diagram shows both orderly progression of the assembly process and interdependence of the components and their specific reactions with other components during the assembly of the subunit.
Adapted from M. Herold and K. Nierbaus, *J. Biol. Chem.* 262–8826, 1987. A similar assembly map for the small subunit was elucidated earlier. (M. Nomura, *Cold Spring Harbor Symp. Quant. Biol.* 52:653, 1987.)

polysomes can be visualized by electron microscopy (Figure 17.3). Second, in eukaryotic cells some ribosomes occur free in the cytosol, but many are bound to membranes of the rough endoplasmic reticulum. In general, **free ribosomes** synthesize proteins that remain within the cell cytosol or become targeted to the nucleus, mitochondria, or some of the other organelles. **Membrane-bound ribosomes** synthesize proteins that will be secreted from the cell or sequestered and function in other cellular membranes or vesicles. In cell homogenates, membrane fragments and the bound ribosomes constitute the **microsome** fraction; detergents that disrupt membranes release these ribosomes.

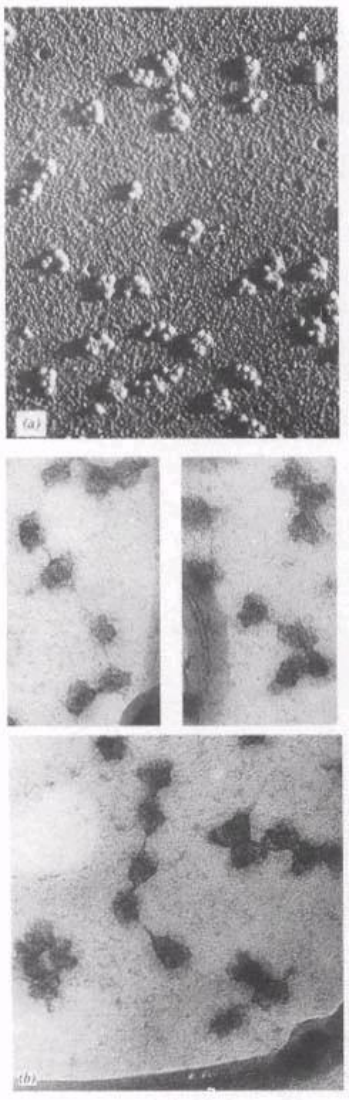


Figure 17.3
Electron micrographs of polysomes.
(a) Reticulocyte polyribosomes shadowed with platinum are seen in clusters of three to six ribosomes, a number consistent with the size of mRNA for a globin chain. (b) Uranyl acetate staining in addition to visualization at a higher magnification shows polysomes in which parts of the mRNA are visible.
Courtesy of Dr. Alex Rich, MIT.

Transfer RNA Acts As a Bilingual Translator Molecule

All tRNA molecules have several common structural characteristics including the 3 -terminal CCA sequence to which amino acids are bound, a highly conserved cloverleaf secondary structure, and an **L**-shaped three-dimensional structure (see p. 682). But each of the many molecular species has a unique nucleotide sequence, giving it individual characteristics that allow great specificity in inter-

actions with mRNA and with the aminoacyl-tRNA synthetase that couples one specific amino acid to it.

The Genetic Code Uses a Four-Letter Alphabet of Nucleotides

Information in the cell is stored in the form of linear sequences of nucleotides in DNA, in a manner that is analogous to the linear sequence of letters of the alphabet in the words you are now reading. The DNA language uses a simple **four-letter alphabet** that comprises the two purines, A and G (adenine and guanine), and the two pyrimidines, C and T (cytosine and thymine). In mRNA the information is encoded in a similar four-letter alphabet, but U (uracil) replaces T. The language of RNA is thus a dialect of the genetic language of DNA. Genetic information is **expressed** predominantly in the form of proteins that derive their properties from their linear sequence of amino acids and to a much lesser extent as RNA species such as tRNA and rRNA. Thus, during protein biosynthesis, the four-letter language of nucleic acids is translated into the 20-letter language of proteins. Implicit in the analogy to language is the directionality of these sequences. By convention, nucleic acid sequences are written in a 5 3 direction, and protein sequences from the amino terminus to the carboxy terminus. These directions in mRNA and protein correspond in both their reading and biosynthetic senses.

Codons in mRNA Are Three-Letter Words

A 1:1 correspondence of nucleotides to amino acids would only permit mRNA to encode four amino acids, while a 2:1 correspondence would encode $4^2 = 16$ amino acids. Neither is sufficient since 20 amino acids occur in most proteins. The actual three-letter **genetic code** has $4^3 = 64$ permutations or words, which is also sufficient to encode start and stop signals, equivalent to punctuation. The three-base words are called **codons** and they are customarily shown in the form of Table 17.2. Only two amino acids are designated by

TABLE 17.2 The Genetic Code[a]

5′ Base	U	C	A	G	3′ Base
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

[a] The genetic code comprises 64 codons, which are permutations of four bases taken in threes. Note the importance of sequence: three bases, each used once per triplet codon, give six permutations: ACG, AGC, GAC, GCA, CAG, and CGA, for threonine, serine, aspartate, alanine, glutamine, and arginine, respectively.

TABLE 17.3 Nonuniversal Codon Usage in Mammalian Mitochondria

Codon	*Usual Code*	*Mitochondrial Code*
UGA	Termination	Tryptophan
AUA	Isoleucine	Methionine
AGA	Arginine	Termination
AGG	Arginine	Termination

single codons: methionine as AUG and tryptophan as UGG. The rest are designated by two, three, four, or six codons. Multiple codons for a single amino acid represent **degeneracy** in the code. The genetic code is nearly **universal.** The same code words are used in all living organisms, prokaryotic and eukaryotic. An exception to universality occurs in mitochondria, in which a few codons have a different meaning than in the cytosol of the same organism (Table 17.3).

Punctuation

Four codons function partly or totally as punctuation, signaling the start and stop of protein synthesis. The **start signal,** AUG, also specifies methionine. An AUG at an appropriate site and within an acceptable sequence in mRNA signifies methionine as the initial, amino-terminal residue. AUG codons elsewhere in the message specify methionine residues within the protein. Three codons, UAG, UAA, and UGA, are **stop signals;** they specify no amino acid and are known as **termination codons** or, less appropriately, as **nonsense codons.**

Codon–Anticodon Interactions Permit Reading of mRNA

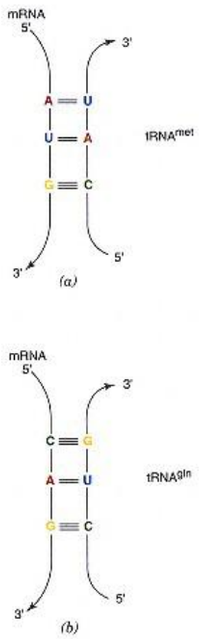


Figure 17.4
Codon–anticodon interactions. Shown are interactions between (a) the AUG (methionine) codon and its CAU anticodon and (b) the CAG (glutamine) codon and a CUG anticodon. Note that these interactions involve antiparallel pairing of mRNA with tRNA.

Translation of the codons of mRNA involves their direct interaction with complementary **anticodon sequences** in tRNA. Each tRNA species carries a unique amino acid, and each has a specific three-base anticodon sequence. Codon–anticodon base pairing is antiparallel, as shown in Figure 17.4. The anticodon is far from the amino acid-acceptor stem in both the tRNA cloverleaf and the **L**-shaped three-dimensional structure of all tRNA molecules. (See Chapter 16, p. 682.) Location of the anticodon and amino acid residue at opposite extremes of the molecule permits the tRNA to conceptually and physically bridge the gap between the nucleotide sequence of the ribosome-bound mRNA and the site of protein assembly on the ribosome.

Since 61 codons designate an amino acid, it might seem necessary to have 61 different tRNA species. This is **not** the case. Variances from standard base pairing are common in codon–anticodon interactions. Many amino acids can be carried by more than one tRNA species, and degenerate codons can be read by more than one tRNA (but always one carrying the correct amino acid). Much of this complexity is explained by the **"wobble" hypothesis,** which permits less stringent base pairing between the third position of a codon and the first position of its anticodon. Thus the first two positions of a codon predominate in tRNA selection and the **degenerate** (third) position is less important. A second modulator of codon–anticodon interactions is the presence of modified nucleotides at or beside the first nucleotide of the anticodon in many tRNA species. A frequent anticodon nucleotide is **inosinic acid** (I), the nucleotide of hypoxanthine, which base pairs with U, C, or A. Wobble base pairing rules are shown in Table 17.4.

TABLE 17.4 Wobble Base Pairing Rules

3′ Codon Base	*5′ Anticodon Bases Possible*
A	U or I
C	G or I
G	C or U
U	A or G or I

If the wobble rules are followed, the 61 nonpunctuation codons could be read by as few as 31 tRNA molecules, but most cells have 50 or more tRNA

species. Some codons are read more efficiently by one anticodon than another. Not all codons are used equally, some being used very rarely. Examination of many mRNA sequences has allowed construction of "codon usage" tables that show that different organisms preferentially use different codons to generate similar polypeptide sequences.

"Breaking" the Genetic Code

The genetic code (Table 17.2) was determined before methods were developed to sequence natural mRNA. These code-breaking experiments provide insight into how proteins are synthesized. Important experiments used simple artificial mRNAs or chemically synthesized trinucleotide codons.

Polynucleotide phosphorylase catalyzes the template-independent and readily reversible reaction:

$$x\text{NDP} \leftrightarrows \text{polynucleotide (pN)}_x + P_i$$

where NDP is any nucleoside 5 -diphosphate or a mixture of two or more. If the nucleoside diphosphate is UDP, a polymer of U designated poly(U) is formed. Under nonphysiological conditions protein synthesis can occur *in vitro* without the initiation components that are normally required. With poly(U) as mRNA, the "protein" polyphenylalanine is made. Similarly, poly(A) encodes polylysine and poly(C) polyproline. An mRNA with a random sequence of only U and C produces polypeptides that contain not only proline and phenylalanine, as predicted, but also serine (from UCU and UCC) and leucine (from CUU and CUC). Because of degeneracy in the code and the complexity of the products, experiments with random sequence mRNAs were difficult to interpret, and so synthetic messengers of defined sequence were transcribed from simple repeating DNA sequences by RNA polymerase. Thus poly(AU), transcribed from a repeating poly(dAT), produces only a repeating copolymer of Ile-Tyr-Ile-Tyr, read from successive triplets AUA UAU AUA UAU and so on. A synthetic poly(CUG) has possible codons CUG for Leu, UGC for Cys, and GCU for Ala, each repeating itself once the **reading frame** has been selected. Since selection of the initiation codon is random in these *in vitro* experiments, three different homopolypeptides are produced: polyleucine, polycysteine, and polyalanine. A perfect poly(CUCG) produces a polypeptide with the sequence (-Leu-Ala-Arg-Ser-) whatever the initiation point. These relationships are summarized in Table 17.5; they show codons to be triplets read in exact sequence, without overlap or omission. Other experiments used chemically synthesized trinucleotide codons as minimal messages. No proteins were made, but the binding of only one amino acid (conjugated to an appropriate tRNA) by the ribosome was stimulated by a given codon. It was thus possible to decipher the meaning of each possible codon and to identify termination codons. All of these conclusions were later verified by the determination of mRNA sequences.

TABLE 17.5 Polypeptide Products of Synthetic mRNAs[a]

mRNA	*Codon Sequence*	*Products*
$-(AU)_n-$	$-\overline{AUA}\,\underline{UAU}\,\overline{AUA}\,\underline{UAU}-$	$-(\text{Ile-Tyr})_{n/3}-$
$-(CUG)_n-$	$-\overline{CUG}\,\underline{CUG}\,\overline{CUG}\,\underline{CUG}-$	$-\text{Leu}_n-$
	$-\overline{UGC}\,\underline{UGC}\,\overline{UGC}\,\underline{UGC}-$	$-\text{Cys}_n-$
	$-\overline{GCU}\,\underline{GCU}\,\overline{GCU}\,\underline{GCU}-$	$-\text{Ala}_n-$
$-(CUCG)_n-$	$\overline{CUC}\,\underline{GCU}\,\overline{GGC}\,\underline{UCG}$	$-(\text{Leu-Ala-Arg-Ser})_{n/3}-$

[a] The horizontal brackets accent the reading frame.

CLINICAL CORRELATION 17.1

Missense Mutation: Hemoglobin

Clinically, the most important missense mutation known is the change from A to U in either the GAA or GAG codon for glutamate to give a GUA or GUG codon for valine in the sixth position of the β chain for hemoglobin. An estimated 1 in 10 African-Americans are carriers of this mutation, which in its homozygous state is the basis for sickle cell disease, the most common of all hemoglobinopathies (see Clin. Corr. 2.3 for the effects of this substitution on the polymerization of deoxygenated hemoglobin). The second most common hemoglobinopathy is hemoglobin C disease, in which a change from G to A in either the GAA or GAG codon for glutamate results in an AAA or AAG codon for lysine in the sixth position of the β chain. Over 600 other hemoglobin missense mutations are now known. Methods for diagnosis of these and other genetic disorders are discussed in Clin. Corr. 16.2. A recent advance in therapy of sickle cell anemia uses hydroxyurea treatment to stimulate synthesis of γchains and thus increase fetal hemoglobin production in affected adults. This decreases the tendency of the HbS in erythrocytes to form linear multimers that result in cell shape distortion—that is, sickling—when the oxygen tension decreases.

Charache, S. et al. Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. *N. Engl. J. Med.* 332:1317–1322, 1995.

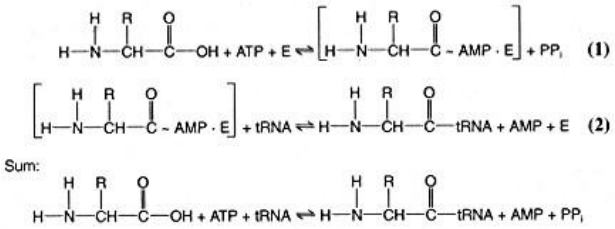
Mutations

An understanding of the genetic code and how it is read provides a basis for understanding the nature of mutations. A mutation is simply a change in a gene. **Point mutations** involve a change in a single base pair in the DNA, and thus a single base in the corresponding mRNA. Sometimes this change occurs in the third position of a degenerate codon and there is no change in the amino acid specified (e.g., UCC to UCA still codes for serine). Such **silent mutations** are only detected by gene sequence determination. They are commonly seen during comparison of genes for similar proteins, for example, hemoglobins from different species. **Missense mutations** arise from a base change that causes incorporation of a different amino acid in the encoded protein (see Clin. Corr. 17.1). Point mutations can also form or destroy a termination codon and thus change the length of a protein. Formation of a termination codon from one that encodes an amino acid (see Clin. Corr. 17.2) is often called a **nonsense mutation;** it results in premature termination and a truncated protein. Mutation of a termination codon to one for an amino acid allows the message to be "read through" until another stop codon is encountered. The result is a larger than normal protein. This phenomenon is the basis of several disorders (see Table 17.6 and Clin. Corr. 17.3).

Insertion or deletion of a single nucleotide within the coding region of a gene results in a **frameshift mutation.** The reading frame is altered at that point and subsequent codons are read in the new context until a termination codon is reached. Table 17.7 illustrates this phenomenon with the mutant hemoglobin Wayne. The significance of reading frame selection is underscored by a phenomenon in some viruses in which a single segment of DNA encodes different polypeptides that are translated using different reading frames. An example is the tumor-causing simian virus SV40 (Figure 17.5), whose small size physically limits the amount of DNA that can be packaged within it.

Aminoacylation of Transfer RNA Activates Amino Acids for Protein Synthesis

In order to be incorporated into proteins, amino acids must first be "activated" by linkage to their appropriate tRNA carriers. This is a two-step process that requires energy and is catalyzed by one of a family of aminoacyl-tRNA synthetases, each of which is specific for a single amino acid and its appropriate tRNA species. The reactions are normally written as follows:



The brackets surrounding the aminoacyl-AMP–enzyme complex indicate that it is a transient, enzyme-bound intermediate. The "squiggle" (~) linkage of amino acid to AMP identifies the aminoacyl-adenylate as a high-energy intermediate, a mixed acid anhydride with carboxyl and phosphoryl components. The aminoacyl ester linkage in tRNA is lower in energy than the aminoacyl-adenylate, but still higher than that of the carboxyl group of the free amino acid. The

CLINICAL CORRELATION 17.2

Disorders of Terminator Codons

In hemoglobin McKees Rocks the UAU or UAC codon normally designating tyrosine in position 145 of the β chain has mutated to the terminator codon UAA or UAG. This results in shortening of the β chain from its normal 146 residues to 144 residues. This change gives the hemoglobin molecule an unusually high oxygen affinity since the normal C-terminal sequence involved in binding 2,3-bisphosphoglycerate is modified. The response to decreased oxygen delivery is secretion of erythropoietin by the kidney and increased red blood cell production that produces a polycythemic phenotype (see Clin. Corr. 22.2).

Another illness that results from a terminator mutation is a variety of β-thalassemia. Thalassemias are a group of disorders characterized at the molecular level by an imbalance in the stoichiometry of α- and β-globin synthesis. In ^{0}-thalassemia no β-globin is synthesized. As a result, α-globin, unable to associate with β-globin to form hemoglobin, accumulates and precipitates in erythroid cells. The precipitation damages cell membranes, causing hemolytic anemia and stimulation of erythropoiesis. One variety of ^{0}-thalassemia, common in Southeast Asia, results from a terminator mutation at codon 17 of the β-globin; the normal codon AAG that designates a lysyl residue at β-17 becomes the stop codon UAG. In contrast to hemoglobin McKees Rocks, in which the terminator mutation occurs late in the β-globin message, the mutation occurs so early in the mRNA that no useful β-globin sequence can be synthesized, and β-globin is absent. This leads to anemia and aggregation of unused α-globin in the red cell precursors. In addition, β-globin mRNA levels are depressed, probably because premature termination of translation leads to instability of the mRNA.

Winslow, R. M., Swenberg, M., Gross, E., et al. Hemoglobin McKees Rocks $(\alpha_2\beta_2^{145\,\mathrm{Tyr}\rightarrow\mathrm{term}})$. A human nonsense mutation leading to a shortened β chain. *J. Clin. Invest.* 57:772, 1976. Chang, J. C., and Kan, Y. W. β-Thalassemia: a nonsense mutation in man. *Proc. Natl. Acad. Sci. USA* 76:2886, 1979.

CLINICAL CORRELATION 17.3

Thalassemia

There are two expressed α-globin genes on each chromosome 16. Many instances of α-thalassemia arise from the deletion of two, three, or all four copies of the α-globin gene. The clinical severity increases with the number of genes deleted. In contrast, the disorders summarized in Table 17.6 are forms of α-thalassemia that arise from abnormally long α-globin molecules, which replace normal α-globin, and are present only in small amounts. These small amounts of α-globin result from a decreased rate of synthesis or more likely from an increased rate of breakdown of the abnormally elongated α-globin. The normal stop codon, UAA, for α-globin mutates to any of four sense codons with resultant placement of four different amino acids at position 142. Normal α-globin is only 141 residues in length, but the four abnormal α-globins are 172 residues in length, presumably because a triplet of nucleotides in the normally untranslated region of the mRNA becomes a terminator codon in the abnormal position 173. Elongated globin chains can also result from frameshift mutations or insertions.

Weatherall, D. J., and Clegg, J. B. The α-chain termination mutants and their relationship to the α-thalassemias. *Philos. Trans. R. Soc. Lond.* 271:411, 1975.

reactions are written to show their reversibility. In reality, **pyrophosphatases** cleave the pyrophosphate released and the equilibrium is strongly shifted toward formation of aminoacyl-tRNA. From the viewpoint of precision in translation, the amino acid, which had only its side chain (R group) to distinguish it, becomes linked to a large, complex, and easily recognized carrrier.

Specificity and Fidelity of Aminoacylation Reactions

Cells contain 20 different **aminoacyl-tRNA synthetases,** each specific for one amino acid, and at most a small family of carrier tRNAs for that amino acid. In translation, codon–anticodon interactions define the amino acid to be incorporated. If an incorrect amino acid is carried by the tRNA, it will be incorporated into the protein. Correct selection of both tRNA and amino acid by the synthetase is necessary to avoid such mistakes. Accuracy of these enzymes is central to the fidelity of protein synthesis.

Aminoacyl-tRNA synthetases share a common mechanism and many are physically associated with one another in the cell. Nevertheless, they are a diverse group of proteins that may contain one, two, or four identical subunits or pairs of dissimilar subunits. Detailed studies indicate that separate structural domains are involved in aminoacyl-adenylate formation, tRNA recognition, and, if it occurs, subunit interactions. In spite of their structural diversity, each enzyme is capable of almost error-free formation of correct aminoacyl-tRNA combinations.

TABLE 17.6 "Read Through" Mutation in Termination Codons Produce Abnormally Long α-Globin Chains

Hemoglobin	α- *Codon 142*	*Amino Acid 142*	α-*Globin Length (Residues)*
A	UAA		141
Constant Spring	CAA	Glutamine	172
Icaria	AAA	Lysine	172
Seal Rock	GAA	Glutamate	172
Koya Dora	UCA	Serine	172

Selection and incorporation of a correct amino acid require great discrimination on the part of some synthetases. While some amino acids be may easily recognized by their bulk (e.g., tryptophan) or lack of bulk (glycine), or by positive or negative charges on the side chains (e.g., lysine and glutamate), others are much more difficult to discriminate. Recognition of valine rather than threonine or isoleucine by the valyl-tRNA synthetase is difficult since the side chains differ by either an added hydroxyl or single methylene group. The amino acid-recognition and -activation sites of each enzyme have great specificity, as is characteristic of many enzymes. Nevertheless, misrecognition does occur. An additional **"proofreading"** or **"editing"** step increases discrimination. This most often occurs through hydrolysis of the aminoacyl-adenylate intermediate, with the release of amino acid and AMP. Valyl-tRNA synthetase efficiently hydrolyzes threonyl-adenylate and it hydrolyzes isoleucyl-adenylate in the presence of bound (but not aminoacylated) $tRNA^{Val}$. In other cases a misacylated tRNA is recognized and deacylated. Valyl- and phenylalanyl-tRNA synthetases deacylate tRNAs that have been mischarged with threonine and tyrosine, respectively. This proofreading is analogous to editing of misincorporated nucleotides by the 3 5 exonuclease activity of DNA polymerases (Chapter 16). Editing is performed by many but not all aminoacyl-tRNA synthetases. The net result is an average level of misacylation of one in 10^4 to 10^5.

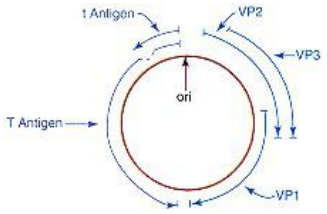


Figure 17.5
Map of genome of simian virus 40 (SV40). DNA of SV40, shown in red, is a double-stranded circle of slightly more than 5000 base pairs that encodes all information needed by the virus for its survival and replication within a host cell. It is an example of extremely efficient use of the information-coding potential of a small genome. Proteins VP1, VP2, and VP3 are structural proteins of the virus; VP2 and VP3 are translated from different initiation points to the same carboxyl terminus. VP1 is translated in a different reading frame so that its amino-terminal section overlaps the VP2 and VP3 genes but its amino acid sequence in the overlapping segment is different from that of VP2 and VP3. Two additional proteins, the large T and small t tumor antigens, which promote transformation of infected cells, have identical amino-terminal sequences. The carboxyl-terminal segment of small t protein is encoded by a segment of mRNA that is spliced out of the large T message, and the carboxyl-terminal sequence of large T is encoded by DNA that follows termination of small t. This occurs through differential processing of a common mRNA precursor. The single site of origin of DNA replication (ori) is outside all coding regions of the genome.

Each synthetase must correctly recognize one to several tRNA species that correctly serve to carry the same amino acid, while rejecting incorrect tRNA species. Given the complexity of tRNA molecules, this should be simpler than selection of a single amino acid. However, recall the conformational similarity and common sequence elements of all tRNAs (p. 682). Different synthetases recognize different elements of tRNA structure. One logical element of tRNA recognition by the synthetase is the anticodon, specific to one amino acid. For example, in the case of $tRNA^{Met}$, changing the anticodon also alters recognition by the synthetase. In other instances, this is at least partly true. Sometimes the anticodon is not a determinant of synthetase-tRNA recognition. Consider, for example, suppressor mutations that "suppress" the expression of classes of chain termination (nonsense) mutations. A point mutation in a glutamine (CAG) codon produces a termination (UAG) codon, which causes the premature termination of the encoded protein. A second **suppressor mutation** in the anticodon of a $tRNA^{Tyr}$, in which the normal GUA anticodon is changed to CUA, allows "read through" of the termination codon. The initial mutation is suppressed as a nearly normal protein is made, with the affected glutamine replaced by tyrosine. Aminoacylation of the mutant $tRNA^{tyr}$ with tyrosine shows that in this case the anticodon does not determine synthetase specificity. In *E. coli* $tRNA^{Ala}$, the primary recognition characteristic is a G_3-U_{70} base pair in the acceptor stem; even if no other changes in the $tRNA^{Ala}$ occur, any variation at this position destroys its acceptor ability with alanine-$tRNA^{Ala}$ synthetase. Incorporation of a G_3-U_{70} base pair in $tRNA^{Cys}$ makes it an alanine acceptor, and even the isolated

TABLE 17.7 A Frameshift Mutation Results in Production of Abnormal Hemoglobin Wayne[a]

Position	*137*	*138*	*139*	*140*	*141*	*142*	*143*	*144*	*145*	*146*	*147*
Normal α-globin amino acid sequence	- Thr	- Ser	- Lys	- Tyr	- Arg						
Normal α-globin codon sequence	- ACP	- UCⓊ	- AAA	- UAC	- CGU	- UAA	- GCU	- GGA	- GCC	- UCG	- GUA
Wayne α-globin codon sequence	- ACP	- UCA	- AAU	- ACC	- GUU	- AAG	- CUG	- GAG	- CCU	- CGG	- UAG
Wayne α-globin amino acid sequence	- Thr	- Ser	- Asn	- Thr	- Val	- Lys	- Leu	- Glu	- Pro	- Arg	

[a] The base deletion causing thre frameshift is encircled. The stop codons are boxed.
P = A, G, U, or C.

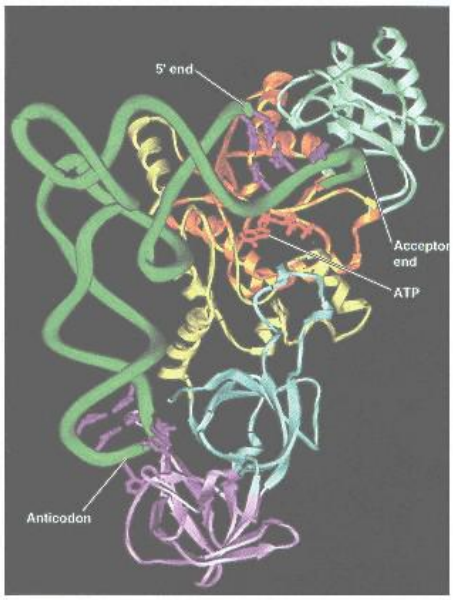


Figure 17.6
Interaction of a tRNA with its cognate aminoacyl-tRNA synthetase.
Figure shows sugar–phosphate backbone of E. coli glutaminyl tRNA in green and the peptide backbone of the glutamine $tRNA^{Gln}$ synthetase in multiple colors. Note the strong interactions of the synthetase with both the partially unwound acceptor stem and the anticodon loop of the tRNA, and placement of ATP, shown in red, within a few angstroms of the 3 end of tRNA. Space-filling models of the enzyme and tRNA would show both molecules to be solid objects with several sites of direct contact. Adapted from J. Perona, M. Rould, and T. Steitz, *Biochemistry* 32:8758, 1993.

acceptor stem of $tRNA^{Ala}$ can be aminoacylated. Other tRNA identification features include additional elements of the acceptor stem and sometimes parts of the variable loop or the D-stem/loop. Usually multiple structural elements contribute to recognition, but many are not absolute determinants. The X-ray structure of the glutaminyl synthetase–tRNA complex shown in Figure 17.6 shows binding at the concave tRNA surface, which is typical and compatible with the biochemical observations.

17.3— Protein Biosynthesis

Translation Is Directional and Colinear with mRNA

In the English language words are read from left to right and not from right to left. Similarly, mRNA sequences are written 5 3 and in the translation process they are read in the same direction. Amino acid sequences are both written and biosynthesized from the amino-terminal residue to the carboxy terminus. This was first demonstrated by following the incorporation of radioactive amino acids into specific sites in hemoglobin as a function of time. Only full length, complete globin chains were isolated and analyzed. Completed chains that incorporated radioactive amino acids during the shortest exposures to the radioactive precursor were near to being finished at the time of the pulse and were found to have radioactive amino acids only in the carboxy-terminal segments. Longer pulses with radioactive amino acids resulted also in labeling of central segments of the protein, and the longest pulse time, still corresponding to less than that needed to synthesize a full-length polypeptide, showed radioactivity approaching the amino-terminal segments. Again, this amino- to carboxy-terminal directionality became obvious as details of translation were clarified.

The existence of stable polysomes and the directional nature of translation imply that each ribosome remains bound to an mRNA molecule and moves along the length of the mRNA until it is fully read. Comparison of mRNA sequences with sequences of the proteins they encode shows a perfect, colinear, gap-free correspondence of the mRNA coding sequence and that of the synthesized polypeptide. In fact, it is common to deduce the sequence of a protein solely from the nucleotide sequence of its mRNA or the DNA of the gene encoding it. However, the deduced sequence may differ from the genuine protein because of posttranslational events and modifications.

Initiation of Protein Synthesis Is a Complex Process

A good novel can be analyzed in terms of its beginning, its development or middle section, and its satisfactory ending. Protein biosynthesis will be described in a similar conceptual and mechanical framework: initiation of the process, elongation during which the great bulk of the protein is formed, and termination of synthesis and release of the finished polypeptide. We will then examine the posttranslational modifications that a protein may undergo.

Initiation requires bringing together a small (40S) ribosomal subunit, the mRNA, and a tRNA complex of the amino-terminal amino acid, all in a proper orientation. This is followed by association of the large (60S) subunit to form a completed initiation complex on an 80S ribosome. The ordered process is shown in Figure 17.7; it also requires a complex group of proteins, known as **initiation factors,** that participate only in initiation. They are not ribosomal proteins, although many of them bind transiently to ribosomes during initiation steps. There are many eukaryotic initiation factors and the specific functions of some remain unclear; prokaryotic protein synthesis provides a useful and less complex model for comparison.

As a first step, **eukaryotic initiation factor 2a (eIF-2a)** binds to GTP and one species of $tRNA^{Met}$, designated $fMet\text{-}tRNA_i^{Met}$ is recognized by prokaryotic IF-2.

The second step in initiation requires 40S ribosomal subunits associated with a very complex protein, **eIF-3.** Mammalian eIF-3 includes eight different polypeptides and has a mass of 600–650 kDa. In electron micrographs eIF-3 is seen bound to the 40S subunit surface that will contact the larger 60S subunit, thus physically blocking association of 40S and 60S subunits. Hence eIF-3 is also called a ribosome **anti-association factor,** as is **eIF-6,** which binds to 60S subunits. A complex that includes eIF-2a · $Met\text{-}tRNA_i^{Met}$ · GTP ternary complex, correctly oriented mRNA, and several protein factors.

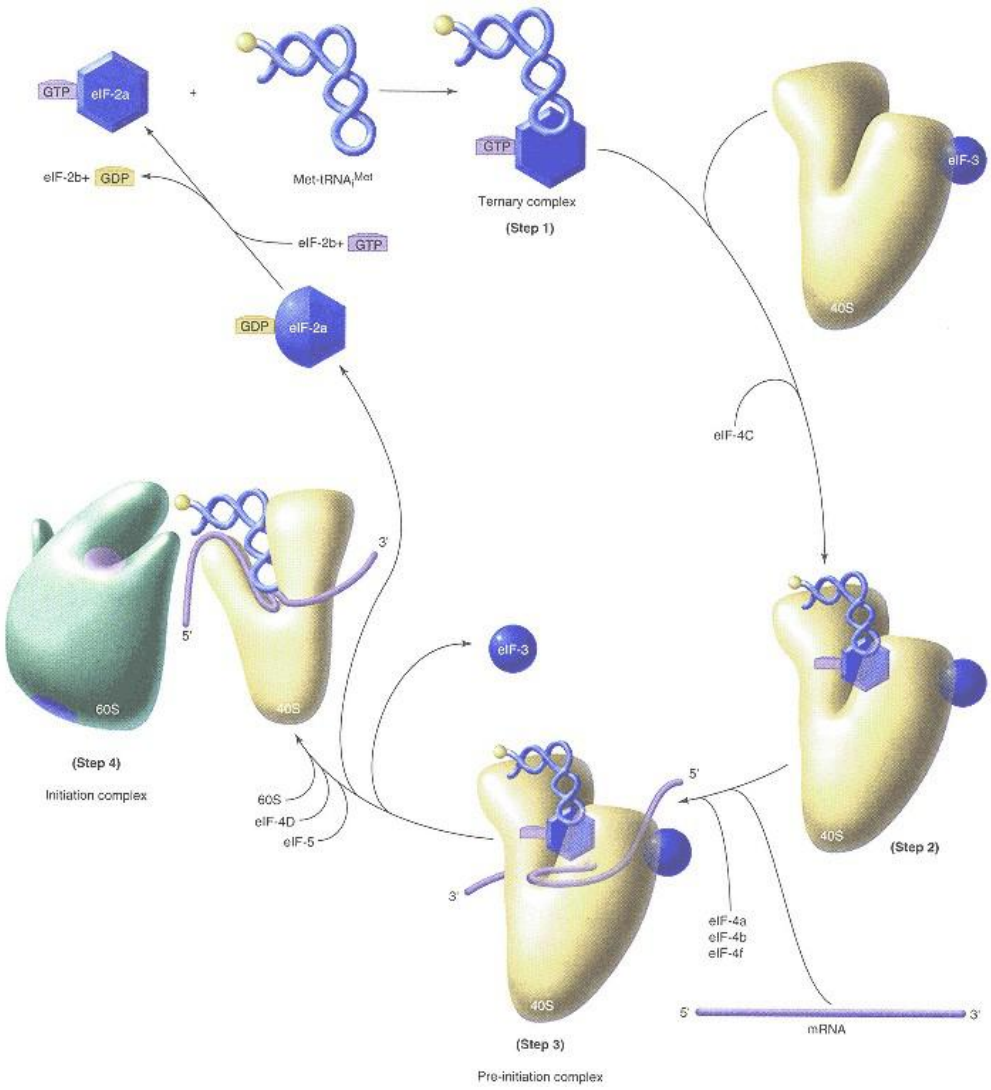


Figure 17.7
Initiation of translation in eukaryotes.
Details are given in the text. Ternary complex (step 1) first combines with small ribosomal subunit to place the initiator tRNA (step 2). Figure shows interaction with a naked mRNA molecule to form a preinitiation complex (step 3); additional small subunits later complex with the same mRNA as polysomes are formed. Formation of the initiation complex (is shown in step 4). The different shape of eIF-2a in complexes with GTP and GDP indicates that conformational change in the protein occurs upon hydrolysis of triphosphate.

Formation of the complete **initiation complex** now proceeds with involvement of a 60S subunit and an additional factor, **eIF-5.** Protein eIF-5 first interacts with the preinitiation complex; GTP is hydrolyzed to GDP and P_i, and eIF-2a · GDP, eIF-3, and other factors are released. The 40S · $Met\text{-}tRNA_i^{Met}$ · mRNA complex interacts with a 60S subunit and initiation factor eIF-4d to generate an 80S ribosome with the mRNA and initiator tRNA correctly positioned on the ribosome. The eIF-2a · GDP that is released interacts with the **guanine nucleotide exchange factor** eIF-2b and GTP to regenerate eIF-2a · GTP for another round of initiation.

Prokaryotes use fewer nonribosomal factors and a slightly different order of interaction. Their 30S subunits complexed with a simpler IF-3 first bind mRNA. Orientation of the mRNA relies in part on base pairing between a pyrimidine-rich sequence of eight nucleotides in 16S rRNA and a purine-rich **"Shine–Dalgarno"** sequence (named for its discoverers) about 10 nucleotides upstream of the initiator AUG codon. Complementarity between rRNA and the message-positioning sequence of an mRNA may include several mismatches but, as a first approximation, the better the complementary pairing the more efficient initiation at that AUG will be. It is interesting that eukaryotes do not utilize an mRNA–rRNA base pairing mechanism, but instead use many protein factors to position mRNA correctly. After the mRNA is bound by a 30S subunit, a ternary complex of IF-2, $fMet\text{-}tRNA_i^{Met}$, and GTP is bound. A third initiation factor, IF-1, also participates in formation of the preinitiation complex. A 50S subunit is now bound; in the process, GTP is hydrolyzed to GDP and P_i, and the initiation factors are released.

Elongation Is the Stepwise Formation of Peptide Bonds

Protein synthesis now occurs by stepwise elongation to form a polypeptide chain. At each step ribosomal **peptidyltransferase** transfers the growing peptide (or in the first step the initiating methionine residue) from its carrier tRNA to the α-amino group of the amino acid residue of the aminoacyl-tRNA specified by the next codon. Efficiency and fidelity are enhanced by nonribosomal protein **elongation factors** that utilize the energy released by GTP hydrolysis to ensure selection of the proper aminoacyl-tRNA species and to move the mRNA and associated tRNAs through the decoding region of the ribosome. Elongation is illustrated in Figure 17.8.

At a given moment, up to three different tRNA molecules may be bound at specific sites that span both ribosomal subunits. The initiating methionyl-tRNA is placed in position so that its methionyl residue may be transferred (or donated) to the free α-amino group of the incoming aminoacyl-tRNA; it thus occupies the donor site, also called the **peptidyl site** or **P site** of the ribosome. The aminoacyl-tRNA specified by the next codon of the message is bound at the acceptor site, also called the **aminoacyl site** or **A site** of the ribosome. Selection of the correct aminoacyl-tRNA is enhanced by **elongation factor 1 (EF-1);** a component of EF-1, **EF-1α,** first forms a ternary complex with aminoacyl-tRNA and GTP. The EF-1α · aminoacyl-tRNA · GTP complex binds to the ribosome and if codon–anticodon interactions are correct, the aminoacyl-tRNA is placed at the A site, GTP is hydrolyzed to GDP and P_i, and the EF-1α · GDP complex dissociates. The initiating methionyl-tRNA and the incoming amino-acyl-tRNA are now juxtaposed on the ribosome. Their anticodons are paired with successive codons of the mRNA in the **decoding region** of the small subunit, and their amino acids are beside one another at the **peptidyltransferase site** of the large subunit. Peptide bond formation now occurs. Peptidyltransferase catalyzes the attack of the α-amino group of the aminoacyl-tRNA onto the carbonyl carbon of the methionyl-tRNA. The result is transfer of the methionine to the amino group of the aminoacyl-tRNA, which then occupies a "hybrid"

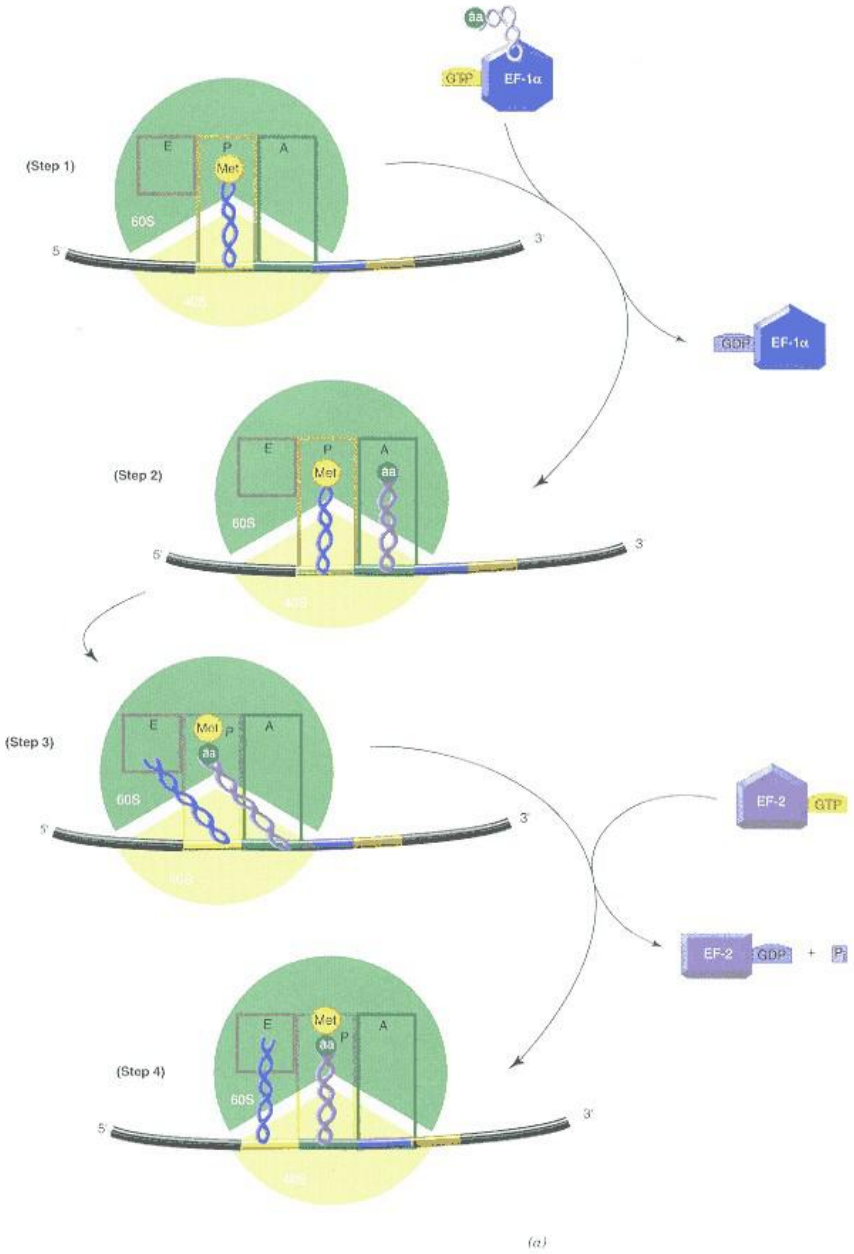
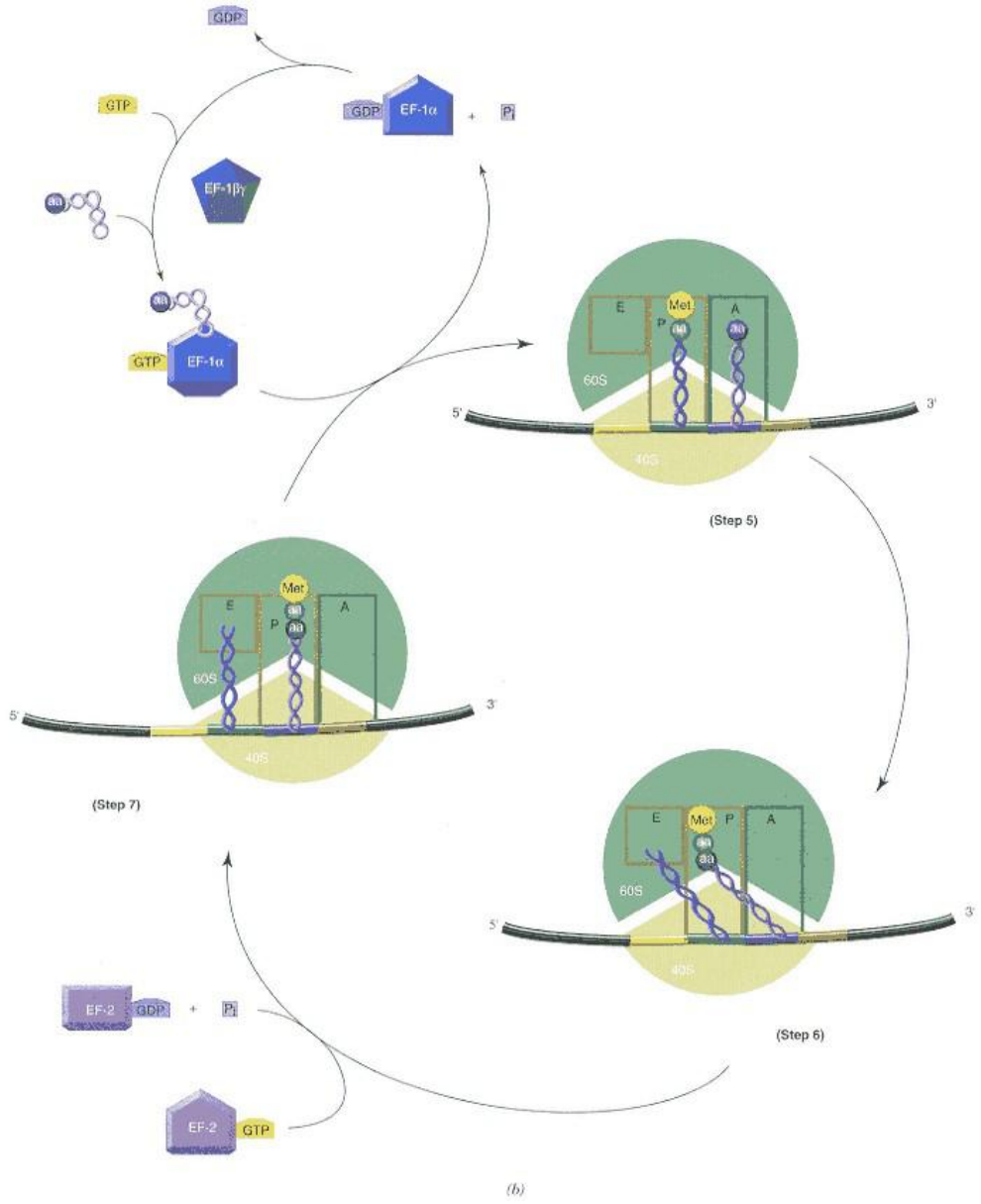


Figure 17.8
Elongation steps in eukaryotic protein synthesis.

(a) First cycle of elongation is shown. Step 1 shows completed initiation complex with methionyl $tRNA_i^{Met}$ in 80S P site. At step 2 an aminoacyl-tRNA has been placed in the ribosomal A site with participation of EF-1α.
Change in shape of EF-1α shows its conformational change upon GTP hydrolysis. At step 3 the first peptide bond has been formed, new peptidyl tRNA occupies a hybrid (A/P) site on the ribosome, and the deacylated acceptor stem of the $tRNA_i^{Met}$ is displaced to the E site of the large subunit. At step 4 mRNA–peptidyl tRNA complex has been fully translocated to the P site while deacylated initiator tRNA is moved to the E site.



(b) Further rounds of elongation are depicted. Binding of aminoacyl-tRNA probably causes concomitant release of deacylated tRNA from the E site, resulting in complex at step 5. Formation of the next peptide bond again results in the new peptidyl RNA occupying a hybrid A/P site on the ribosome (step 6), and translocation moves mRNA and new peptidyl tRNA in register into the P site (step 7). Additional amino acids are added by successive repetitions of the cycle. For further details see text.

position on the ribosome. The anticodon remains in the 40S A site, while the acceptor end and the attached peptide are in the 60S P site. The anticodon of the deacylated tRNA remains in the 40S P site, and its acceptor end is located in the 60S exit or **E site.**

The mRNA and the dipeptidyl-tRNA at the 40S A site must now be repositioned to permit another elongation cycle to begin. This is done by **elongation factor 2 (EF-2),** also called **translocase.** EF-2 moves the messenger and dipeptidyl-tRNA, in codon–anticodon register, from the 40S A site to the P site. In the process, GTP is hydrolyzed to GDP plus P_i, providing energy for the movement, and the A site is fully vacated. As the dipeptidyl-tRNA is moved to the P site, the deacylated donor (methionine) tRNA is also moved to the E site, which only exists on the 60S subunit. The ribosome can now enter a new cycle. The next aminoacyl-tRNA specified by the mRNA is delivered by EF-1α to the A site and the deacylated tRNA in the E site is probably released. Peptide transfer again occurs. Successive cycles of binding of aminoacyl-tRNA, peptide bond formation, and translocation result in the stepwise elongation of the polypeptide toward its eventual carboxyl terminus. Note that whatever the length of the growing chain, peptide bond formation always occurs through attack of the α-amino group of the incoming aminoacyl-tRNA on the peptide carboxyl-tRNA linkage; hence the geometric arrangement of the reacting molecules at the peptidyltransferase site remains constant.

Peptide bond formation does not require any additional energy source such as ATP or GTP. The energy of the methionyl (or peptidyl) ester linkage to tRNA drives the reaction toward peptide bond formation; recall that ATP is used to form each aminoacyl-tRNA and that these reactions are reversible. Isolated 60S subunits can catalyze peptidyltransferase activity, and nonribosomal factors are not involved in the reaction. Yet peptidyltransferase has never been dissociated from the large subunit or identified as a specific ribosomal protein. Reconstitution of *E. coli* peptidyltransferase activity requires only five to six different large subunit proteins and the rRNA. Omission or significant modification of the rRNA or any of these proteins causes the loss of peptidyltransferase activity, while other proteins can be deleted with little or no effect. The discovery of catalytic RNA molecules (Chapter 16) led to speculation that the primordial ribosome was an RNA particle in which peptide bond formation was catalyzed by the RNA. Experiments with very conformationally "stable" large subunit RNA from a thermophilic bacterium suggest that the rRNA may be the catalytic component of peptidyltransferase, while the proteins serve to stabilize RNA folding; however, this hypothesis remains controversial and not fully proved.

As determined with their prokaryotic equivalents, the role of GTP in the action of EF-1α and EF-2 probably relates to conformational changes in these proteins. Crystallographic studies have shown that a large rearrangement of domains with movements of several angstroms occurs upon GTP hydrolysis in **EF-Tu**, the prokaryotic equivalent of EF-1α. Both EF-1α and EF-2 bind ribosomes tightly as GTP complexes, while GDP complexes dissociate from the ribosome more easily. Viewed another way, GTP stabilizes a protein conformation that confers upon EF-1α high affinity toward aminoacyl-tRNA and the ribosome, while GDP stabilizes a conformation with lower affinity for aminoacyl-tRNA and ribosome, thus allowing tRNA delivery and factor dissociation. Restoration of the higher affinity GTP-associated conformation of EF-1α requires participation **EF-1$\beta\gamma$** (Figure 17.9). This protein displaces GDP from EF-1α, forming an EF-1α · EF-1 complex. GTP then displaces EF-1 , forming an EF-1α · GTP complex that can successively bind an aminoacyl-tRNA and then a ribosome. Prokaryotes use a similar mechanism in which EF-Tu binds GTP and aminoacyl-tRNA and EF-Ts displaces GDP and helps recycle the carrier molecule. Prokaryotes also utilize a GTP-dependent translocase, equivalent to EF-2 but called **EF-G** or G factor.

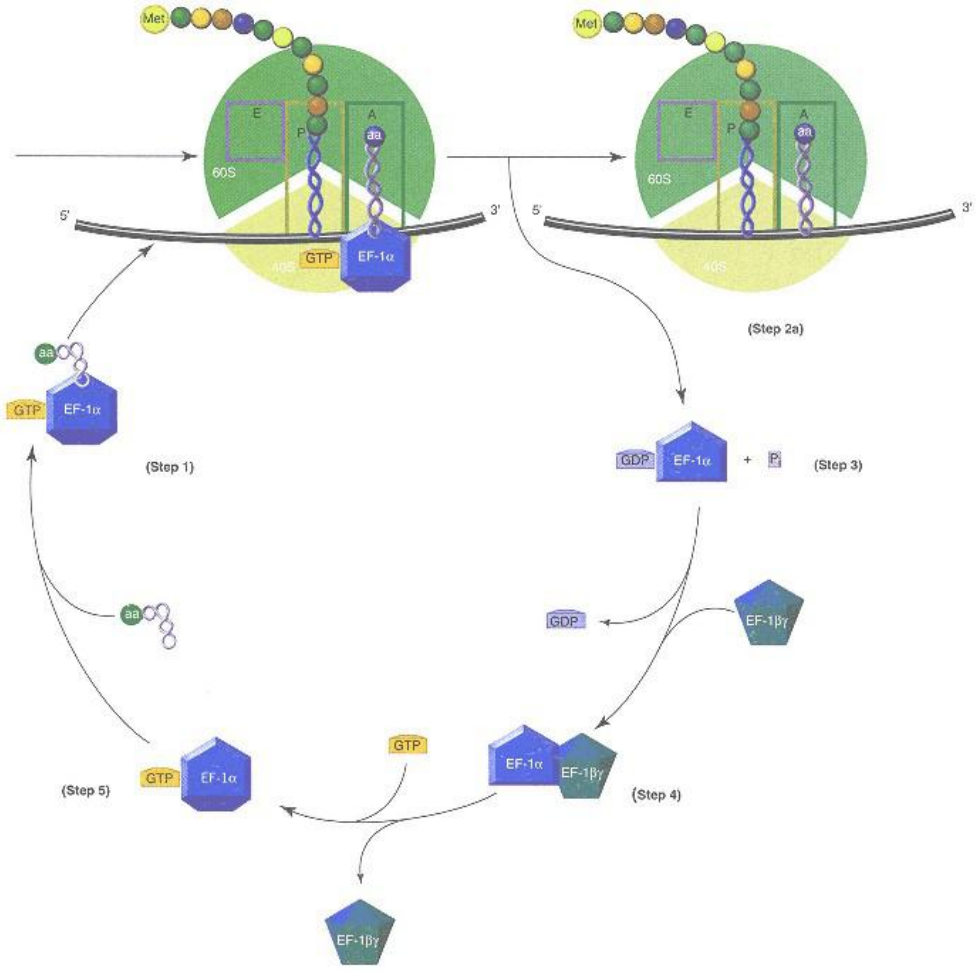


Figure 17.9
EF-1 in elongation cycle.

EF-1α · GTP · aminoacyl-tRNA complex (step 1) binds the ribosome (step 2) and transfers aminoacyl-tRNA to the ribosome (step 2a) with concomitant hydrolysis of GTP and a change in conformation of EF-1α (step 3) that reduces its affinity for tRNA and ribosome. The GDP is then displaced from EF-1α by EF-1 , resulting in the complex at step 4. Binding of GTP then displaces EF-1 (step 5) and allows binding of an aminoacyl-tRNA by EF-1α in its higher affinity conformation (step 1). In prokaryotes a similar cycle exists; EF-Tu functions as the carrier of aminoacyl-tRNA and EF-Ts is guanine nucleotide exchange factor.

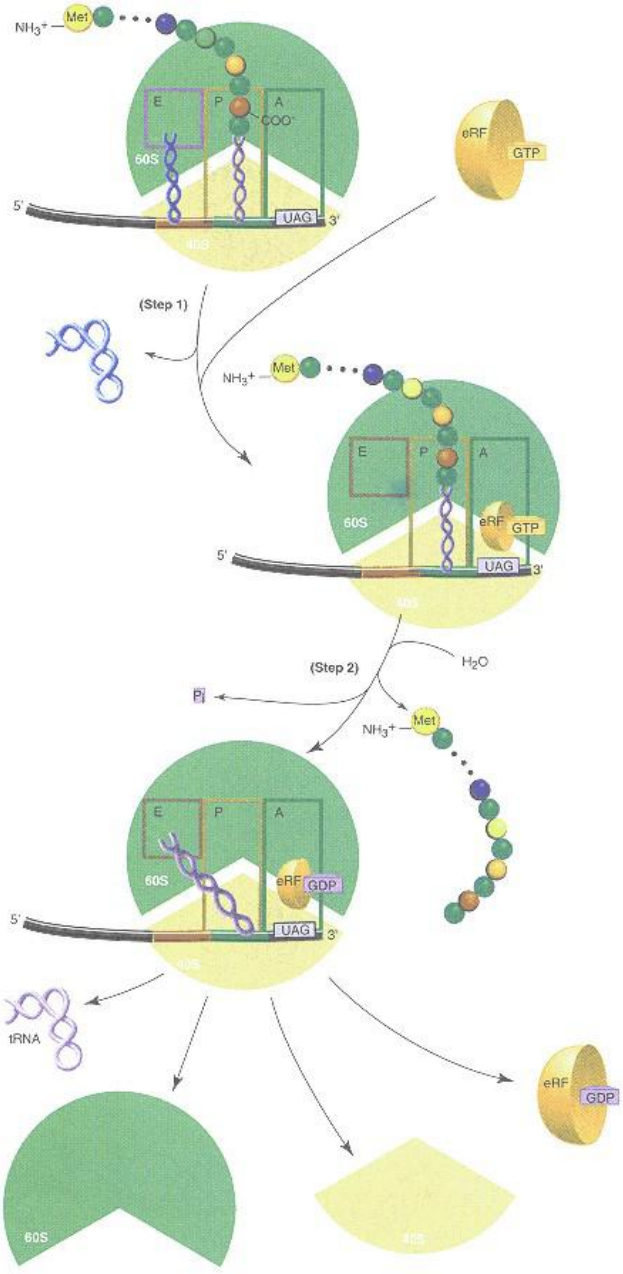


Figure 17.10
Model of termination of protein biosynthesis.
When a termination codon (UAG, UAA, or UGA) in mRNA occupies the ribosomal A site, binding of release factor–GTP complex occurs (step 1), probably with concomitant release of deacylated tRNA from the ribosomal E site. In step 2 peptidyltransferase now functions as a hydrolase; protein is released by hydrolysis of the ester bond linking it to tRNA, and acceptor end of deacylated tRNA is probably displaced. GTP is hydrolyzed to GDP and P_i, presumably altering the conformation of the release factor. Complex is now dissociated and components can enter additional rounds of protein synthesis.

Termination of Polypeptide Synthesis Requires a Stop Codon

A chain-terminating UAG, UAA, or UGA codon in the A site does not promote binding of any tRNA species. Instead, another complex nonribosomal protein, **release factor** (eRF), binds the ribosome as an eRF · GTP complex (Figure 17.10). The peptide–tRNA ester linkage is cleaved through the action of peptidyl transferase, acting here as a hydrolase, and the completed polypeptide is released from its carrier tRNA and the ribosome. Dissociation of eRF from the ribosome requires hydrolysis of the GTP and frees the ribosome to dissociate into subunits and then reenter the protein synthesis cycle at the initiation stage. In prokaryotes three release factors, RF-1, RF-2, and RF-3, carry out the termination function. The factor RF-1 acts in response to UAG or UAA codons, RF-2 acts in response to UGA or UAA codons, and RF-3 is a GTPase that activates RF-1 and RF-2.

Translation Has Significant Energy Cost

There is a considerable use of energy in synthesis of a polypeptide. Amino acid activation converts an ATP to AMP and pyrophosphate, which is normally hydrolyzed to P_i; the net cost is two high-energy phosphates. Two more high-energy bonds are hydrolyzed in the actions of EF-1α and EF-2, for a total of four per peptide bond formed. Posttranslational modifications may add to the energy cost, and of course energy is needed for biosynthesis of the multi-use mRNA, tRNAs, ribosomes, and protein factors, but these costs are distributed among the proteins formed during their lifetime.

Protein Synthesis in Mitochondria Differs Slightly

Many characteristics of mitochondria suggest that they are descendants of aerobic prokaryotes that invaded and set up a symbiotic relationship within a eukaryotic cell. Some of their independence and prokaryotic character are retained. Human mitochondria have a circular DNA genome of 16,569 base pairs that encodes 13 proteins, 22 tRNA species, and two mitochondrion-specific rRNA species. Their independent apparatus for protein synthesis includes RNA polymerase, aminoacyl-tRNA synthetases, tRNAs, and ribosomes. Although the course of protein biosynthesis in mitochondria is like that in the cytosol, some details are different. The synthetic components, tRNAs, aminoacyl-tRNA synthetases, and ribosomes, are unique to the mitochondrion. The number of tRNA species is small and the genetic code is slightly different (see Table 17.3). Mitochondrial ribosomes are smaller and the rRNAs are shorter than those of either the eukaryotic cytosol or of prokaryotes (see Table 17.1). An initiator $\text{fMet-tRNA}_i^{\text{Met}}$. Most mitochondrial proteins are encoded in nuclear DNA and synthesized in the cytosol, but mitochondrial protein synthesis is clearly important (see Clin. Corr. 17.4). Cells must also coordinate protein synthesis within mitochondria with the cytosolic synthesis of proteins destined for import into mitochondria.

Some Antibiotics and Toxins Inhibit Protein Biosynthesis

Protein biosynthesis is central to the continuing life and reproduction of cells. An organism can gain a biological advantage by interfering in the ability of its competitors to synthesize proteins, and many antibiotics and toxins function in this way. Some are selective for prokaryotic rather than eukaryotic protein synthesis and so are extremely useful in clinical practice. Examples of antibiotic action are listed in Table 17.8.

Several mechanisms of interfering in ribosome subunit–tRNA interactions are utilized by different antibiotics. **Streptomycin** binds the small subunit of

CLINICAL CORRELATION 17.4

Mutation in Mitochondrial Ribosomal RNA Results in Antibiotic-Induced Deafness

In some regions of China a significant percentage of irreversible cases of deafness has been linked to use of normally safe and effective amounts of aminoglycoside antibiotics such as streptomycin and gentamicin. The unusual sensitivity to aminoglycosides is transmitted only through women. This maternal transmission suggests a mitochondrial locus, since sperm do not contribute mitochondria to the zygote. Aminoglycosides are normally targeted to bacterial ribosomes, so the mitochondrial ribosome is a logical place to look for a mutation site.

A single A G point mutation at nucleotide 1555 of the gene on mitochondrial DNA for the rDNA of the large subunit has been identified in three families with this susceptibility to aminoglycosides. The mutation site is in a highly conserved region of the rRNA sequence that is known to be involved in aminoglycoside binding; some mutations in the same region confer resistance to the antibiotics, and the RNA region is part of the ribosomal A site. It is hypothesized that the mutation makes the region more "prokaryote-like," increasing its affinity for aminoglycosides and the ability of the antibiotic to interfere in protein synthesis in the mitochondrion. Proteins synthesized in the mitochondrion are needed to form the enzyme complexes of the oxidative phosphorylation system, so affected cells are starved of ATP. Aminoglycosides accumulate in the cochlea, making this a particularly sensitive target and leading to sensorineural deafness.

Fischel-Ghodsian, N., Prezant, T., Bu., X., and Öztas, S. Mitochondrial ribosomal RNA gene mutation in a patient with sporadic aminoglycoside ototoxicity. *Am. J. Otolaryngol.* 14:399, 1993. Prezant, T., Agapian, J., Bohlman. M., et al. Mitochondrial ribosomal RNA mutation associated with both antibiotic-induced and non-syndromic deafness. *Nature Genetics* 4:289, 1993.

prokaryotic ribosomes, interferes with the initiation of protein synthesis, and causes misreading of mRNA. Although streptomycin does not directly bind ribosomal protein S12 of the small subunit, mutations in this protein or in the small subunit rRNA can confer resistance to or even dependence on streptomycin. Protein S12 is involved in tRNA binding, and streptomycin alters the interactions of tRNA with the ribosomal subunit and mRNA, probably by affecting subunit conformation. Other **aminoglycoside antibiotics,** such as the **neomycins** or **gentamicins,** also cause mistranslation; they interact with the small ribosomal subunit, but at sites that differ from that for streptomycin. The aminoglycoside **kasugamycin** binds small subunits and inhibits the initiation of translation. Kasugamycin sensitivity depends on base methylation that normally occurs on two adjacent adenine moieties of small subunit rRNA. **Tetracyclines** bind directly to ribosomes and interfere in aminoacyl-tRNA binding.

Other antibiotics interfere with elongation. **Puromycin** (Figure 17.11) resembles an aminoacyl-tRNA; it binds at the ribosomal A site and acts as an acceptor in the peptidyltransferase reaction. However, since it does not interact with mRNA it cannot be translocated, and since its aminoacyl derivative is not in an ester linkage to the nucleoside it cannot serve as a peptide donor. Thus puromycin prematurely terminates translation, leading to release of peptidyl-puromycin. **Chloramphenicol** directly inhibits peptidyltransferase by binding the transferase center; no transfer occurs, and peptidyl-tRNA remains associated

TABLE 17.8 Some Inhibitors of Protein Biosynthesis

Inhibitor	*Processes Affected*	*Site of Action*
Streptomycin	Initiation, elongation	Prokaryotes: 30S subunit
Neomycins	Translation	Prokaryotes: multiple sites
Tetracyclines	Aminoacyl-tRNA binding	30S or 40S subunits
Puromycin	Peptide transfer	70S or 80S ribosomes
Erythromycin	Translocation	Prokaryotes: 50S subunit
Fusidic acid	Translocation	Prokaryotes: EF-G
Cycloheximide	Elongation	Eukaryotes: 80S ribosomes
Ricin	Multiple	Eukaryotes: 60S subunit

with the ribosome. The translocation step is also a potential target. **Erythromycin,** a macrolide antibiotic, interferes with translocation on prokaryotic ribosomes. Eukaryotic translocation is inhibited by **diphtheria toxin,** a protein toxin produced by *Corynebacterium diphtheriae*, the toxin binds at the cell membrane and a subunit enters the cytoplasm and catalyzes the ADP-ribosylation and inactivation of EF-2, as represented in the reaction:

$$\underset{\text{(active)}}{\text{EF-2}} + \text{NAD}^+ \rightleftarrows \underset{\text{(inactive)}}{\text{ADP-ribosyl EF-2}} + \text{nicotinamide} + \text{H}^+$$

ADP-ribose is attached to EF-2 at a posttranslationally modified histidine residue known as diphthamide. Posttranslational events are discussed in the next section.

A third group of toxins attack the rRNA. **Ricin** (from castor beans) and related toxins are *N*-glycosidases that cleave a single adenine from the large subunit rRNA backbone. The ribosome is inactivated by this apparently minor damage. A fungal toxin, **α-sarcin,** cleaves large subunit rRNA at a single site and similarly inactivates the ribosome. Some *E. coli* strains make extracellular toxins that affect other bacteria. One of these, **colicin E3,** is a ribonuclease that cleaves 16S RNA near the mRNA-binding sequence and decoding region; it thus inactivates the small subunit and halts protein synthesis in competitors of the colicin-producing cell.

17.4— Protein Maturation: Modification, Secretion, and Targeting

Some proteins emerge from the ribosome ready to function, while others undergo a variety of **posttranslational modifications.** These alterations may result in conversion to a functional form, direction to a specific subcellular compartment, secretion from the cell, or an alteration in activity or stability. Information that determines the posttranslational fate of a protein resides in its structure: that is, the amino acid sequence and conformation of the polypeptide determine whether a protein will be a substrate for a modifying enzyme and/or identify it for direction to a subcellular or extracellular location.

Proteins for Export Follow the Secretory Pathway

Proteins destined for export are synthesized on membrane-bound ribosomes of the rough endoplasmic reticulum (ER) (Figure 17.12). A ribosome has no means of classifying the polypeptide it is about to synthesize, so initiation and elongation begin on free cytosolic ribosomes. Proteins of the secretory pathway have a hydrophobic **signal peptide,** usually at or near their amino terminus. There is no unique signal peptide sequence, but its characteristics include a positively charged N terminus, a core of 8–12 hydrophobic amino acids, and a more polar C-terminal segment that eventually serves as a cleavage site for excision of the signal peptide.

The signal peptide of 15–30 amino acids emerges from the ribosome early during polypeptide synthesis. As it appears it is bound by a cytosolic **signal recognition particle (SRP)** (see Figure 17.13). The SRP is an elongated particle made up of six different proteins plus a small (7S) RNA molecule that serves as a backbone. Binding to SRP halts protein synthesis and the ribosome moves to the ER. SRP recognizes and binds to an **SRP receptor** or **"docking protein,"** localized at the cytosolic surface of the ER membrane, in a reaction that requires GTP hydrolysis and presumably involves conformational changes in the SRP and/or the receptor. The ribosome is transferred to a **"translocon,"** a ribosome receptor on the membrane that serves as a passageway through the membrane. Both SRP and docking protein are freed to direct other ribosomes to the ER,

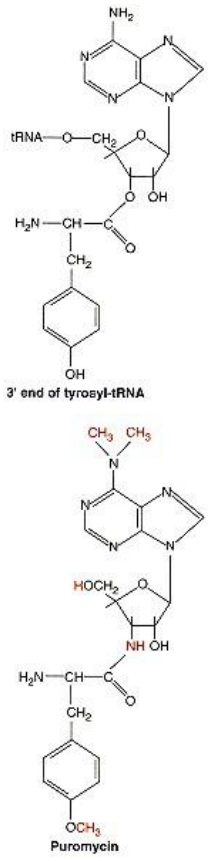


Figure 17.11
Puromycin (right) interferes with protein synthesis by functioning as an analog of aminoacyl-tRNA, here tyrosyl-tRNA (left) in peptidyltransferase reaction.

and the translational block caused by SRP binding is relieved. The hydrophobic signal sequence, probably complexed by a receptor protein, is inserted into the membrane, further anchoring the ribosome to the ER. Translation and extrusion into or through the membrane are now coupled. Translocon proteins form a pore or channel through which the growing polypeptide passes; even very hydrophilic or ionic segments are directed through the hydrophobic membrane into the ER lumen and folding into secondary and tertiary structures begins.

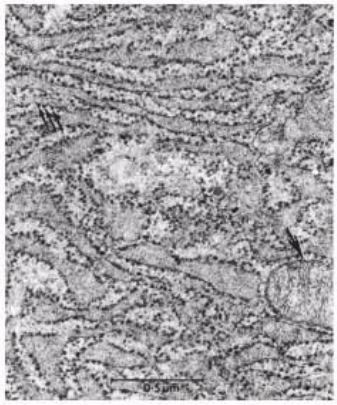


Figure 17.12
Rough endoplasmic reticulum of a plasma cell.
Three parallel arrows indicate three ribosomes among the many attached to the extensive membranes. Single arrow indicates a mitochondrion for comparison.
Courtesy of Dr. U. Jarlfors, University of Miami.

The completed export-destined protein within the ER lumen will probably be anchored to the membrane by the signal peptide. A cleavage site on the protein is hydrolyzed by **signal peptidase,** an integral membrane protein located at the luminal surface of the ER. The protein completes folding into a three-dimensional conformation, disulfide bonds can form, and components of multisubunit proteins may assemble. Other steps may include proteolytic processing and glycosylation that occur within the ER lumen and during transit of the protein through the Golgi apparatus and into secretory vesicles.

Glycosylation of Proteins Occurs in the Endoplasmic Reticulum and Golgi Apparatus

Glycosylation of proteins to form glycoproteins (see p. 60) is important for two reasons. Glycosylation alters the properties of proteins, changing their stability, solubility, and physical bulk. In addition, carbohydrates of glycoproteins act as recognition signals that are central to aspects of protein targeting and for cellular recognition of proteins and other cells. Glycosylation can involve addition of a few carbohydrate residues or the formation of large branched oligosaccharide chains. Sites and types of glycosylation are determined by the presence on a protein of appropriate amino acids and sequences, and by availability of enzymes and substrates to carry out the glycosylation reactions.

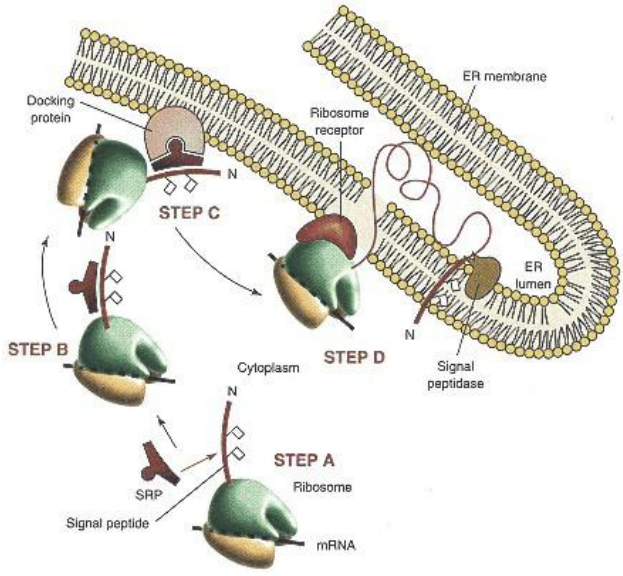


Figure 17.13
Secretory pathway: signal peptide recognition.
At step A a hydrophobic signal peptide emerges from the exit site of a free ribosome in the cytosol. Signal recognition particle (SRP) recognizes and binds the peptide and peptide elongation is temporarily halted (step B). The ribosome moves to the ER membrane where docking protein binds to SRP (step C). In step D the ribosome is transferred to a ribosome receptor or translocon, protein biosynthesis is resumed, and newly synthesized protein is extruded through the membrane into the ER lumen.

TABLE 17.9 Glycosyltransferases in Eukaryotic Cells

Sugar Transferred	*Abbreviation*	*Donors*	*Glycosyltransferase*
Mannose	Man	GDP-Man	Mannosyltransferase
		Dolichol-Man	
Galactose	Gal	UDP-Gal	Galactosyltransferase
Glucose	Glc	UDP-Glc	Glucosyltransferase
		Dolichol-Glc	
Fucose	Fuc	GDP-Fuc	Fucosyltransferase
N-Acetylgalactosamine	GalNAc	UDP-GalNac	*N*-acetylgalactosaminyltransferase
N-Acetylglucosamine	GlcNAc	UDP-GlcNAc	*N*-acetylglucosaminyltransferase
N-Acetylneuraminic acid (or sialic acid)	NANA or NeuNAc SA	CMP-NANA CMP-SA	*N*-Acetylneuraminyltransferase (sialyltransferase)

Glycosylation involves many **glycosyltransferases,** classes of which are summarized in Table 17.9. Up to 100 different enzymes each carry out a similar basic reaction in which a sugar is transferred from an activated donor substrate to an acceptor, usually another sugar residue that is part of an oligosaccharide under construction. The enzymes show three kinds of specificity: for the monosaccharide that is transferred, for structure and sequence of the acceptor molecule, and for the site and configuration of the anomeric linkage formed.

One class of glycoproteins has sugars linked through the amide nitrogen of asparagine residues in the process of ***N*-linked glycosylation.** The antibiotic **tunicamycin,** which prevents *N*-glycosylation, has been valuable in elucidating the biosynthetic pathway. Formation of *N*-linked oligosaccharides begins in the ER lumen and continues after transport of the protein to the Golgi apparatus. A specific sequence, Asn-X-Thr (or Ser) in which X may be any amino acid except proline or aspartic acid, is required for *N*-glycosylation. Not all Asn-X-Thr/Ser sequences are glycosylated because some may be unavailable due to protein conformation.

Biosynthesis of *N*-linked oligosaccharides begins with the synthesis of a lipid-linked intermediate (Figure 17.14). **Dolichol phosphate** (structure on p. 350) at the cytoplasmic surface of the ER membrane serves as glycosyl acceptor of *N*-acetylglucosamine. The GlcNAc-pyrophosphoryldolichol is an acceptor for stepwise glycosylation and formation of a branched $(Man)_5(GlcNAc)_2$-pyro-phosphoryldolichol on the cytosolic side of the membrane. This intermediate is then reoriented to the luminal surface of the ER membrane, and four additional mannose and then three glucose residues are sequentially added to complete the structure. The complete oligosaccharide is then transferred from its dolichol carrier to an asparagine residue of the polypeptide as it emerges into the ER lumen. Thus *N*-glycosylation is **cotranslational,** that is, occurs as the protein is being synthesized, hence it can affect protein folding.

Processing or **modification** of the oligosaccharide by **glycosidases** involves removal of some sugar residues from the newly transferred structure. The glucose residues, which were required for transfer of the oligosaccharide from the dolichol carrier, are sequentially removed, as is one mannose. These alterations mark the glycoprotein for transport to the Golgi apparatus where further trimming by glycosidases may occur. Additional sugars may also be added by a variety of glycosyltransferases. The resulting *N*-linked oligosaccharides are diverse, but two classes are distinguishable. Each has a common core region ($GlcNAc_2Man_3$) linked to asparagine and originating from the dolichol-linked intermediate. The **high-mannose** type includes mannose residues in a variety of linkages and shows less processing from the dolichol-linked intermediate. The **complex** type is more highly processed and diverse, with a larger variety of sugars and linkages. Examples of mature oligosaccharides are shown in Figure 17.15.

The second major class of glycoproteins have sugars that are bound through either serine or threonine hydroxyl groups. Such O-linked glycosylation occurs only after the protein has reached the Golgi apparatus, hence ***O*-glycosylation** is posttranslational and occurs only on fully folded proteins. *O*-linked carbohydrates always involve *N*-acetylgalactosamine attachment to a serine or threonine residue of the protein. There is no defined amino acid sequence in which the

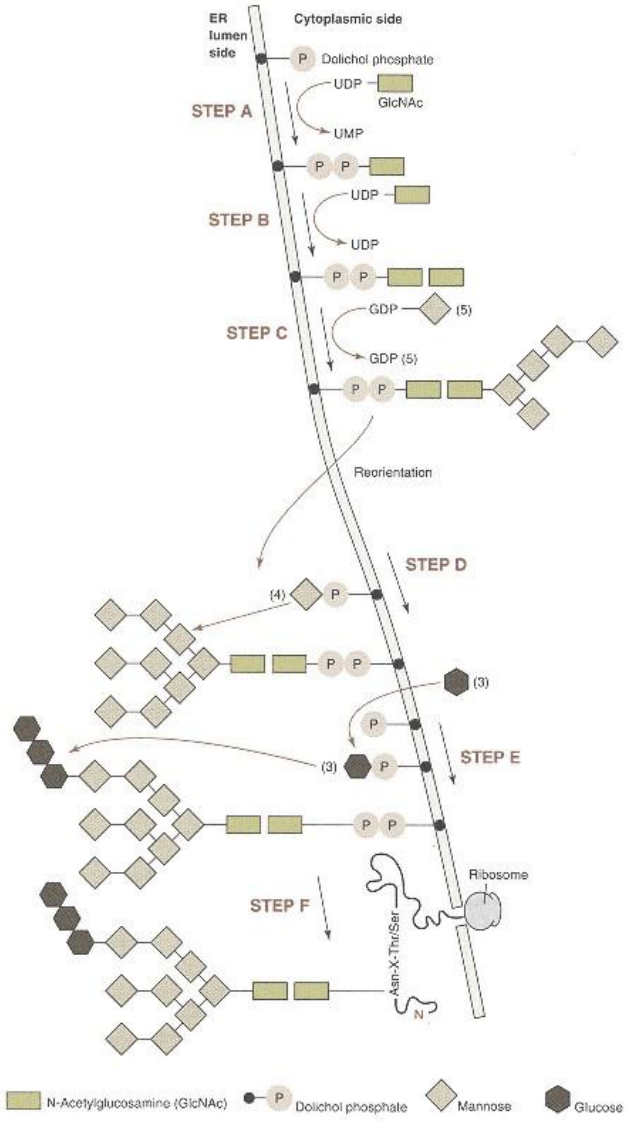


Figure 17.14
Biosynthesis of *N*-linked oligosaccharides at the surface of the endoplasmic reticulum.
Synthesis is initiated on the cytoplasmic face of the ER membrane by transfer of *N*-acetylglucosamine phosphate to a dolichol acceptor (step A) followed by formation of the first glycosidic bond upon transfer of a second residue of *N*-acetylglucosamine (step B). Five residues of mannose are then added sequentially (step C) from a GDP mannose carrier. At this stage lipid-linked oligosaccharide is reoriented to the luminal face of the membrane, and additional mannose (step D) and glucose (step E) residues are transferred from dolichol-linked intermediates. Dolichol sugars are generated from cytosol nucleoside diphosphate sugars. The completed oligosaccharide is finally transferred to a protein in the process of being synthesized at the membrane surface; signal peptide may have already been cleaved at this point.

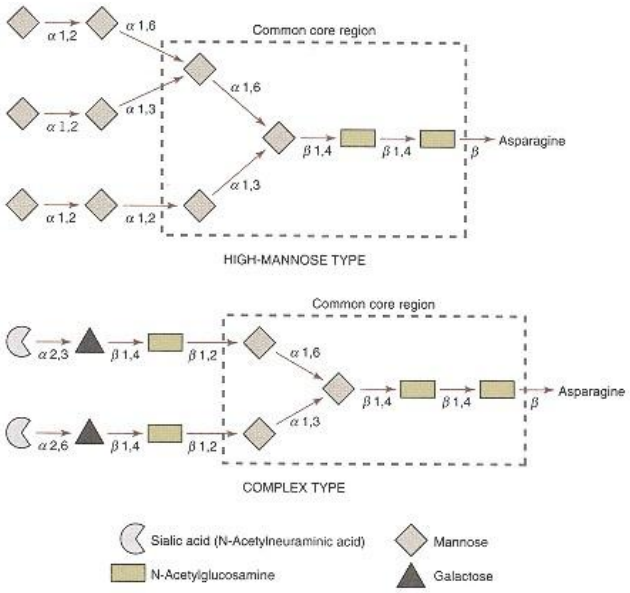


Figure 17.15
Structure of *N*-linked oligosaccharides.
Basic structures of both types of *N*-linked oligosaccharides are shown. In each case structure is derived from that of the initial dolichol-linked oligosaccharide through action of glycosidases and glycosyltransferases. Note the variety of glycosidic linkages involved in these structures.

serine or threonine must occur, but only residues whose side chains are in an appropriate environment on the protein surface serve as acceptors for the GalNAc-transferase.

Sequential addition of sugars to the GalNAc acceptor follows, using the same glycosyltransferases that modified *N*-linked oligosaccharides in the Golgi apparatus. The structures synthesized depend on types and amounts of glycosyltransferases in a given cell. If an acceptor is a substrate for more than one transferase, the amount of each transferase controls the competition between them. Some oligosaccharides may be formed that are not acceptors for any glycosyltransferase present, hence no further growth of the chain occurs. Other structures may be excellent acceptors that continue to grow until completed by one of a number of nonacceptor termination sequences. These processes can lead to many different oligosaccharide structures on otherwise identical proteins, so heterogeneity in glycoproteins is common. Examples are shown in Figure 17.16.

17.5— Organelle Targeting and Biogenesis

Sorting of Proteins Targeted for Lysosomes Occurs in the Secretory Pathway

Protein transport from ER to Golgi apparatus occurs through carrier vesicles that bud from the ER. This transport requires GTP; inhibitors of oxidative phosphorylation cause proteins to accumulate in the ER and vesicles. Sorting of proteins for their ultimate destinations occurs in conjunction with their glycosylation and proteolytic trimming as they pass through the cis, medial, and trans elements of the Golgi apparatus.

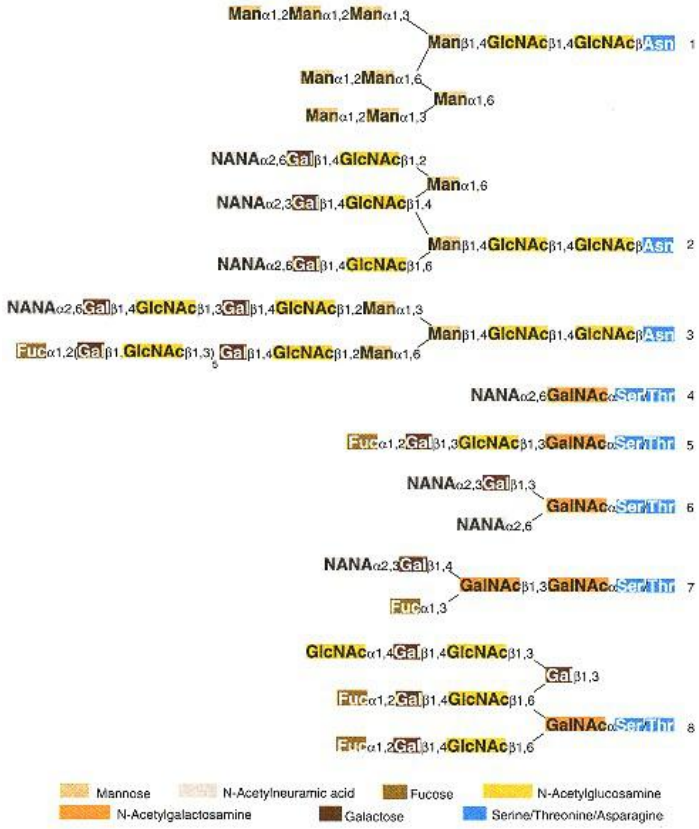


Figure 17.16
Examples of oligosaccharide structure.
Structures 1–3 are typical *N*-linked oligosaccharides of high-mannose (1) and complex types (2, 3); note the common core structure from the protein asparagine residue through the first branch point. Structures 4–8 are common O-linked oligosaccharides that may be quite simple or highly complex. Note that although the core structure (GalNAc-Ser/Thr) is unlike that of N-linked oligosaccharides, the termini can be quite similar (e.g., structures 2 and 6, 3, and 7). Abbreviations: Man = mannose; Gal = galactose; Fuc = fucose; GlcNAc = *N*-acetylglucosamine; GalNAc = *N*-acetylgalactosamine; NANA = *N*-acetylneuraminic acid (sialic acid).
Adapted from J. Paulson, *Trends Biochem. Sci.* 14:272, 1989.

CLINICAL CORRELATION 17.5

I-Cell Disease

I-cell disease (mucolipidosis II) and pseudo-Hurler polydystrophy (mucolipidosis III) are related diseases that arise from defects in lysosomal enzyme targeting because of a deficiency in the enzyme that transfers *N*-acetylglucosamine phosphate to the high mannose-type oligosaccharides of proteins destined for the lysosome. Fibroblasts from affected individuals show dense inclusion bodies (hence I-cells) and are defective in multiple lysosomal enzymes that are found secreted into the medium. Patients have abnormally high levels of lysosomal enzymes in their sera and other body fluids. The disease is characterized by severe psychomotor retardation, many skeletal abnormalities, coarse facial features, and restricted joint movement. Symptoms are usually observable at birth and progress until death, usually by age 8. Pseudo-Hurler polydystrophy is a much milder form of the disease. Onset is usually delayed until the age of 2–4 years, the disease progresses more slowly, and patients survive into adulthood. Prenatal diagnosis of both diseases is possible, but there is as yet no definitive treatment.

For a review of lysosomal enzyme trafficking, see Kornfeld, S. *J. Clin. Invest.* 77:1, 1986. For a comprehensive review of these diseases, see Kornfeld, S., and Sly, W. S., I-cEll disease and pseudo-Hurler polydystrophy. In: C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle (Eds.), *The Molecular and Metabolic Basis of Inherited Disease*, 7th ed. New York: McGraw-Hill, 1995, pp. 2495–2508.

The best understood sorting process is targeting of specific glycoproteins to **lysosomes.** In the cis Golgi some aspect of tertiary structure allows lysosomal proteins to be recognized by a glycosyltransferase that attaches *N*-acetylglucosamine phosphate (GlcNAc-P) to high-mannose type oligosaccharides. A glycosidase then removes the GlcNAc, forming an oligosaccharide that contains **mannose 6-phosphate** (Figure 17.17) that is recognized by a receptor protein responsible for compartmentation and vesicular transport of these proteins to lysosomes. Other oligosaccharide chains on the proteins may be further processed to form complex type structures, but the mannose 6-phosphate determines the lysosomal destination of these proteins. Patients with **I-cell disease** lack the GlcNAc-P glycosyltransferase and cannot correctly mark lysosomal enzymes for their destination. Thus the enzymes are secreted from the cell (see Clin. Corr. 17.5).

Other sorting signals are reasonably well understood. Proteins are retained in the ER lumen in response to a C-terminal KDEL (Lys-Asp-Glu-Leu) sequence, and a different sequence in an exposed C terminus signals retention in the ER membrane. Transmembrane domains have been identified that result in reten-

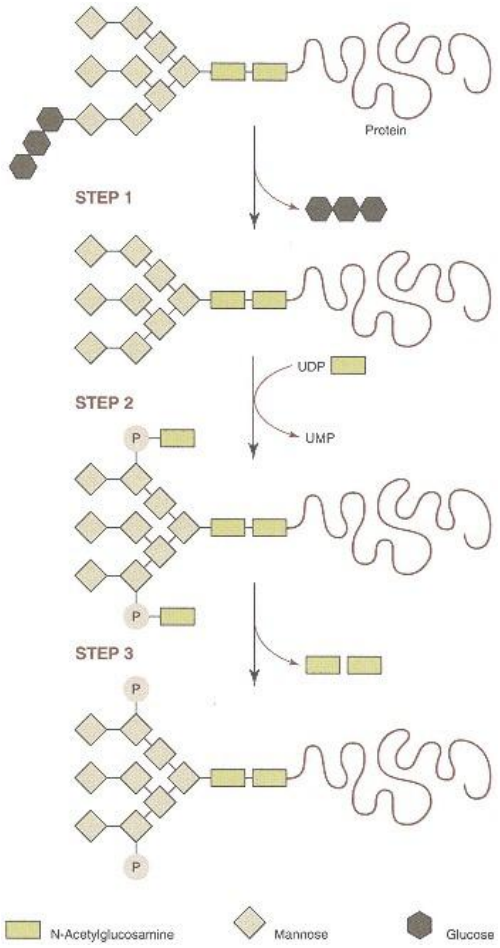


Figure 17.17
Targeting of enzymes to lysosomes.
Completed *N*-linked glycoprotein is released from ER membrane, and during transport to and through the Golgi apparatus the oligosaccharide is modified by glycosidases that remove glucose residues (step 1). Some mannose residues may also be removed. An element of protein structure is then recognized by a glycosyltransferase that transfers one or sometimes two *N*-acetylglucosamine phosphate residues to the oligosaccharide (step 2). A glycosidase removes *N*-acetylglucosamine, leaving one or two mannose 6-phosphate residues on the oligosaccharide (step 3). The protein is then recognized by a mannose 6-phosphate receptor and directed to lysosomes.
Adapted from R. Kornfeld and S. Kornfeld, *Annu. Rev. Biochem.* 54:631, 1985.

tion in the Golgi. Polypeptide-specific glycosylation and sulfation of some glycoprotein hormones in the anterior pituitary mediate their sorting into storage granules. Polysialic acid modification of a neural cell adhesion protein appears to be both specific to the protein and regulated developmentally. Many other sorting signals must still be deciphered to explain fully how the Golgi apparatus directs proteins to its own subcompartments, various storage and secretory granules, and specific elements of the plasma membrane.

The secretory pathway directs proteins to lysosomes, the plasma membrane, or outside the cell. Proteins of the ER and Golgi apparatus are targeted through partial use of the pathway. For example, localization of proteins on either side of or spanning the ER membrane can utilize the signal recogni-

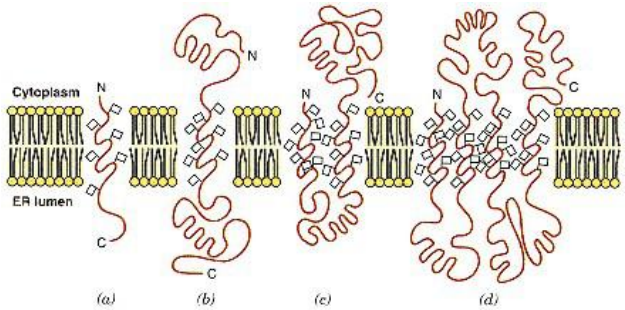


Figure 17.18
Topology of proteins at membranes of endoplasmic reticulum.
Proteins are shown in several orientations with respect to the membrane. In (a) the protein is anchored to the luminal surface of the membrane by an uncleaved signal peptide.
In (*b*) the signal sequence is not near the N terminus; a domain of the protein was synthesized before emergence of signal peptide. Insertion of the internal signal sequence, followed by completion of translation, resulted in a protein with a cytoplasmic N-terminal domain, a membrane-spanning central segment, and a C-terminal domain in the ER lumen.
Diagram (c) shows a protein with the opposite orientation: an N-terminal signal sequence, which might also have been cleaved by signal peptidase, resulted in extrusion of a segment of protein into the ER lumen. A second hydrophobic anchoring sequence remained membrane associated and prevented passage of the rest of the protein through the membrane, thus allowing formation of a C-terminal cytoplasmic domain.
In (d), several internal signal and anchoring sequences allow various segments of the protein to be oriented on each side of the membrane.

tion particle in slightly different ways (Figure 17.18). If the signal sequence is downstream from the amino terminus of the protein, the amino end may not be inserted into the membrane and may remain on the cytoplasmic surface. Internal hydrophobic anchoring sequences within a protein can allow much of the sequence either to remain on the cytoplasmic surface or to be retained, anchored on the luminal surface of the ER membrane. Multiple anchoring sequences in a single polypeptide can cause it to span the membrane several times and thus be largely buried in it. Such hydrophobic sequences are separated by polar loops whose orientation is determined by positively charged flanking residues that predominate on the cytoplasmic side of the membrane.

Import of Proteins by Mitochondria Requires Specific Signals

Mitochondria provide a particularly complex targeting problem since specific proteins are located in the mitochondrial matrix, inner or outer membrane, or intermembrane space. Most of these proteins are synthesized in the cytosol on free ribosomes and imported into the mitochondrion, and most are synthesized as larger preproteins; N-terminal presequences mark the protein not only for the mitochondrion but also for a specific subcompartment. The **mitochondrial matrix targeting signal** is not a specific sequence, but rather a positively charged amphiphilic α-helix. With the aid of a protein chaperone, it is recognized by a **mitochondrial receptor** and the protein is translocated across both membranes and into the mitochondrial matrix in an energy-dependent reaction. Passage occurs at adhesion sites where the inner and outer membranes are close together. Proteases remove the matrix targeting signal but may leave other sequences that further sort the protein within the mitochondrion. For example, a clipped precursor of cytochrome-b_2 is moved back across the inner membrane in response to a hydrophobic signal sequence. Further proteolysis frees the protein in the intermembrane space. In contrast, cytochrome-*c* apoprotein (without heme) binds at the outer membrane and is passed into the intermembrane space. There it acquires its heme and undergoes a conformational change that prevents return to the cytosol. Outer membrane localization can utilize the matrix targeting mechanism to translocate part of the protein, but a large apolar sequence blocks full transfer and leaves a membrane-bound protein with a C-terminal domain on the surface of the mitochondrion.

Targeting to Other Organelles Requires Specific Signals

Nuclei must import many proteins involved in their own structure and for DNA replication, transcription, and ribosome biogenesis. Nuclear pores permit the

passage of small proteins, but larger proteins are targeted by nuclear localization signals that include clusters of basic amino acids. Some nuclear proteins may be retained in the nucleus by forming complexes within the organelle. Peroxisomes contain a limited array of enzymes. One targeting signal is a carboxy-terminal tripeptide, Ser-Lys-Leu (SKL). An N-terminal targeting signal also exists, and others may yet be discovered.

A different targeting problem exists for proteins that reside in more than one subcellular compartment. Sometimes gene duplication and divergence have resulted in different targeting signals on closely related mature polypeptides. **Alternative transcription initiation sites** or pre-mRNA splicing can generate different messages from a single gene. An example of the latter is seen in a calcium–calmodulin-dependent protein kinase; alternatively spliced mRNAs differ with respect to an internal segment that encodes a nuclear localization signal. Without this segment, the protein remains in the cytosol. **Alternative translation initiation sites** lead to two forms of rat liver fumarase, one of which includes a mitochondrial targeting sequence while the other does not and remains in the cytosol. A suboptimal localization signal can lead to inefficient targeting and a dual location, as is seen in the partial secretion of an inhibitor of the plasminogen activator. Finally, some proteins contain more than one targeting signal, which must compete with each other.

17.6— Further Posttranslational Protein Modifications

Several additional maturation events may modify newly synthesized polypeptides to help generate their final, functional structures. Many of these events are very common, while others are specialized to one or a few known instances.

CLINICAL CORRELATION 17.6

Familial Hyperproinsulinemia

Familial hyperproinsulinemia, an autosomal dominant condition, results in approximately equal amounts of insulin and an abnormally processed proinsulin being released into the circulation. Although affected individuals have high levels of proinsulin in their blood, they are apparently normal in terms of glucose metabolism, being neither diabetic nor hypoglycemic. The defect was originally thought to result from a deficiency of one of the proteases that process proinsulin. Three enzymes process proinsulin: endopeptidases that cleave the Arg31–Arg32 and Lys64–Arg65 peptide bonds, and a carboxypeptidase. In several families the defect is the substitution of Arg65 by His or Leu, which prevents cleavage between the C-peptide and the A chain of insulin, resulting in secretion of a partially processed proinsulin. In one family a point mutation (His10 Asp10) causes the hyperproinsulinemia, but how this mutation interferes with processing is not known.

Steiner, D. F., Tager, H. S., Naujo, K., Chan, S. J., and Rubenstein, A. H. Familial syndromes of hyperproinsulinemia with mild diabetes. In: C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle (Eds.), *The Molecular and Metabolic Basis of Inherited Disease*, 7th ed. New York: McGraw-Hill, 1995, pp. 897–904.

Insulin Biosynthesis Involves Partial Proteolysis

Partial proteolysis of proteins is a common maturation step. Sequences can be removed from either end or from within the protein. Proteolysis in the ER and Golgi apparatus helps to mature the protein hormone insulin (Figure 17.19). **Preproinsulin** encoded by mRNA is inserted into the ER lumen. Signal peptidase cleaves the signal peptide to generate **proinsulin,** which folds to form the correct disulfide linkages. Proinsulin is transported to the Golgi apparatus where it is packaged into secretory granules. An internal connecting peptide **(C peptide)** is removed by proteolysis, and mature insulin is secreted. In familial hyperproinsulinemia, processing is incomplete (see Clin. Corr. 17.6).

This pathway for insulin biosynthesis has advantages over synthesis and binding of two separate polypeptides. First, it ensures production of equal amounts of A and B chains without coordination of two translational activities. Second, proinsulin folds into a three-dimensional structure in which the cysteine residues are placed for correct disulfide bond formation. Proinsulin can be reduced and denatured but refolds correctly to form proinsulin. Renaturation of reduced and denatured insulin is less efficient, and incorrect disulfide linkages are also formed. Correct formation of insulin from separately synthesized chains might have required evolution of a helper protein or molecular chaperone.

Proteolysis Leads to Zymogen Activation

Precursor protein cleavage is a common means of enzyme activation. Digestive proteases are classic examples of this phenomenon (see p. 1059). Inactive **zymogen** precursors are packaged in storage granules and activated by proteol-

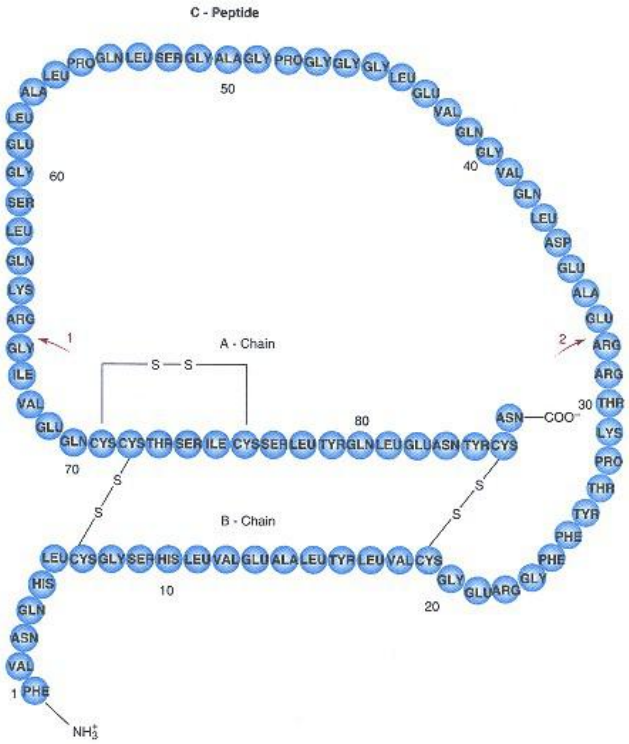


Figure 17.19
Maturation of human proinsulin.
After cleavage at two sites indicated by arrows, the arginine residues 31, 32, and 65 and lysine residue 64 are removed to produce insulin and C-peptide.
Redrawn from G. I. Bell, W. F. Swain, R. Pictet, B. Cordell, H. M. Goodman, and W. J. Putter, *Nature* 282:525, 1979.

ysis upon secretion. Thus trypsinogen is cleaved to give an amino-terminal peptide plus trypsin, and chymotrypsinogen is cleaved to form chymotrypsin and two peptides.

Amino Acids Can Be Modified after Incorporation into Proteins

Only 20 amino acids are encoded genetically and incorporated during translation. **Posttranslational modification** of proteins, however, leads to formation of 100 or more different amino acid derivatives in proteins. Modification may be permanent or highly reversible. The amounts of modified amino acids may be small, but they often play a major functional role in proteins. Examples are listed in Table 17.10.

Protein amino termini are frequently modified. Protein synthesis is initiated using methionine, but in the majority of proteins the amino-terminal residue is not methionine; proteolysis has occurred. The amino terminus is then sometimes modified by, for example, acetylation or myristoylation. Amino-terminal glutamine residues spontaneously cyclize; one possible result is the stabilization of the protein. Amino terminal sequences are occasionally lengthened by the addition of an amino acid (see Section 17.8, Protein Degradation and Turnover).

Posttranslational disulfide bond formation is catalyzed by a **disulfide isomerase.** The cystine-containing protein is conformationally stabilized. Disulfide formation can prevent unfolding of proteins and their passage across membranes, so it also becomes a means of localization. As seen in the case of insulin, disulfide bonds can covalently link separate polypeptides and be necessary

for biological function. Cysteine modification also occurs; multiple sulfatase deficiency arises from reduced ability to carry out a posttranslational modification (see Clin. Corr 17.7).

Methylation of lysine -amino groups occurs in histone proteins and may modulate their interactions with DNA. A fraction of the H2A histone is also modified through isopeptide linkage of a small protein, ubiquitin, from its C-terminal glycine to a lysine -amino group on the histone. A role in DNA interactions is postulated. Biotin is also linked to proteins through amide linkages to lysine.

Serine and threonine hydroxyl groups are major sites of glycosylation and of reversible phosphorylation by protein kinases and protein phosphatases. A classic example of phosphorylation of a serine residue is glycogen phosphorylase, which is modified by phosphorylase kinase (see p. 322). Tyrosine kinase activity is a property of many growth factor receptors; growth factor binding stimulates cell division and the proliferation of specific cell types. Oncogenes, responsible in part for the proliferation of tumor cells, often have tyrosine kinase activity and show strong homology with normal growth factor receptors Dozens of other examples exist; together the protein kinases and protein phosphatases control the activity of many proteins that are central to normal and abnormal cellular development.

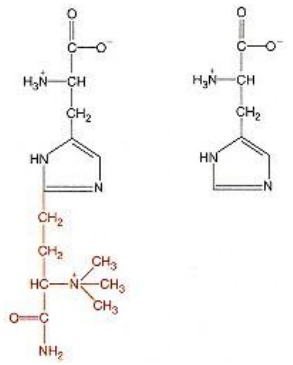


Figure 17.20
Diphthamide (left) is a posttranslational modification of a specific residue of histidine (right) in EF-2.

ADP-ribosylation of EF-2 at a modified histidine residue represents a doubling of posttranslational modifications. First, a specific EP-2 histidine residue is modified to generate the diphthamide derivative (Figure 17.20) of the functional protein. This modification is probably not absolutely required since yeast mutants that cannot make diphthamide survive. ADP-ribosylation of the diphtham-

TABLE 17.10 Modified Amino Acids in Proteins[a]

Amino Acid	*Modifications Found*
Amino terminus	Formylation, acetylation, aminoacylation, myristoylation, glycosylation
Carboxyl terminus	Methylation, glycosyl-phosphatidylinositol anchor formation, ADP-ribosylation
Arginine	*N*-Methylation, ADP-ribosylation
Asparagine	*N*-Glycosylation, *N*-methylation, deamidation
Aspartic acid	Methylation, phosphorylation, hydroxylation
Cysteine	Cystine formation, selenocysteine formation, palmitoylation, linkage to heme, *S*-glycosylation, prenylation
Glutamic acid	Methylation, γ-carboxylation, ADP-ribosylation
Glutamine	Deamidation, cross-linking, pyroglutamate formation
Histidine	Methylation, phosphorylation, diphthamide formation, ADP-ribosylation
Lysine	*N*-acetylation, *N*-methylation, oxidation, hydroxylation, cross-linking, ubiquitination, allysine formation
Methionine	Sulfoxide formation
Phenylalanine	β-Hydroxylation and glycosylation
Proline	Hydroxylation, glycosylation
Serine	Phosphorylation, glycosylation, acetylation
Threonine	Phosphorylation, glycosylation, methylation
Tryptophan	β-Hydroxylation, dione formation
Tyrosine	Phosphorylation, iodination, adenylation, sulfonylation, hydroxylation

Source: Adapted from R. G. Krishna and F. Wold, Post-translational modification of proteins. In: A. Meister (Ed.), *Advances in Enzymology*, Vol. 67. New York: Wiley-Interscience, 1993, pp. 265–298.

[a] The listing is not comprehensive and some of the modifications are very rare. Note that no derivatives of alanine, glycine, isoleucine, and valine have been identified in proteins.

CLINICAL CORRELATION 17.7

Absence of Posttranslational Modification: Multiple Sulfatase Deficiency

A variety of biological molecules are sulfated; examples include glycosaminoglycans, steroids, and glycolipids. Ineffective sulfation of the glycosaminoglycans chondroitin sulfate and keratan sulfate of cartilage results in major skeletal deformities. Degradation of sulfated molecules depends on the activity of a group of related sulfatases, most of which are located in lysosomes. Multiple sulfatase deficiency is a rare lysosomal storage disorder that combines features of metachromatic leukodystrophy and mucopolysaccharidosis. Affected individuals develop slowly and from their second year of life lose the abilities to stand, sit, or speak; physical deformities and neurological deficiencies develop and death before age 10 is usual. Biochemically, multiple sulfatase deficiency is characterized by severe lack of all the sulfatases. In contrast, deficiencies in individual sulfatases are also known, and several distinct diseases are linked to single enzyme defects.

The molecular defect in multiple sulfatase deficiency arises from a deficiency in a posttranslational modification that is common to all sulfatase enzymes and is necessary for their enzymatic activity. In each case a cysteine residue of the enzyme is normally converted to 2-amino-3-oxopropionic acid; the $-CH_2SH$ side chain of cysteine becomes a –CHO (aldehyde) group, which may itself react with amino or hydroxyl groups of the enzyme, a cofactor, and so on. Fibroblasts from individuals with multiple sulfatase deficiency catalyze this modification with significantly lowered efficiency, and the unmodified sulfatases are catalytically inactive.

Schmidt, B., Selmer, T., Ingendoh, A, and von Figura, K. A novel amino acid modification in sulfatases that is deficient in multiple sulfatase deficiency. *Cell* 82:271–278, 1995.

ide by diphtheria toxin then inhibits EF-2 activity. Other instances of physiological ADP-ribosylation not mediated by bacterial toxins are reversible.

Formation of γ-carboxyglutamate from glutamic acid residues occurs in several blood-clotting proteins including prothrombin and factors VII, IX, and X. The γ-carboxyglutamate residues chelate calcium ion, which is required for normal blood clotting (see p. 963). In each case the modification requires vitamin K and can be blocked by coumarin derivatives, which antagonize vitamin K. As a result, the rate of coagulation is greatly decreased.

Collagen Biosynthesis Requires Many Posttranslational Modifications

Collagen, the most abundant protein (or family of related proteins) in the human body, is a fibrous protein that provides the structural framework for tissues and organs. It undergoes a wide variety of posttranslational modifications that directly affect its structure and function, and defects in its modification result in serious diseases. Collagen is an excellent example of the importance of posttranslational modification.

Different species of collagen, designated types I, II, III, IV, and so on (see Table 2.11) are encoded on several chromosomes and expressed in different tissues. Their amino acid sequences differ, but their overall structural similarity suggests a common evolutionary origin. Each collagen polypeptide, designated an α chain, has a repeating sequence Gly-X-Y that is about 1000 residues long. Every third residue is glycine, about one-third of the X positions are occupied by proline and a similar number of Y positions are 4-hydroxyproline, a posttranslationally modified form of proline. Proline and hydroxyproline residues impart considerable rigidity to the structure, which exists as a polyproline type II helix (Figure 17.21; see also p. 52). A collagen molecule includes three α chains intertwined in a collagen triple helix in which the glycine residues occupy the center of the structure.

Procollagen Formation in the Endoplasmic Reticulum and Golgi Apparatus

Collagen α chain synthesis starts in the cytosol, where the amino-terminal signal sequences bind signal recognition particles. Precursor forms, designated, for example, prepro α1(I), are extruded into the ER lumen and the signal peptides are cleaved. Hydroxylation of proline and lysine residues occurs **cotranslationally,** before assembly of a triple helix. Prolyl 4-hydroxylase requires an -X-Pro-

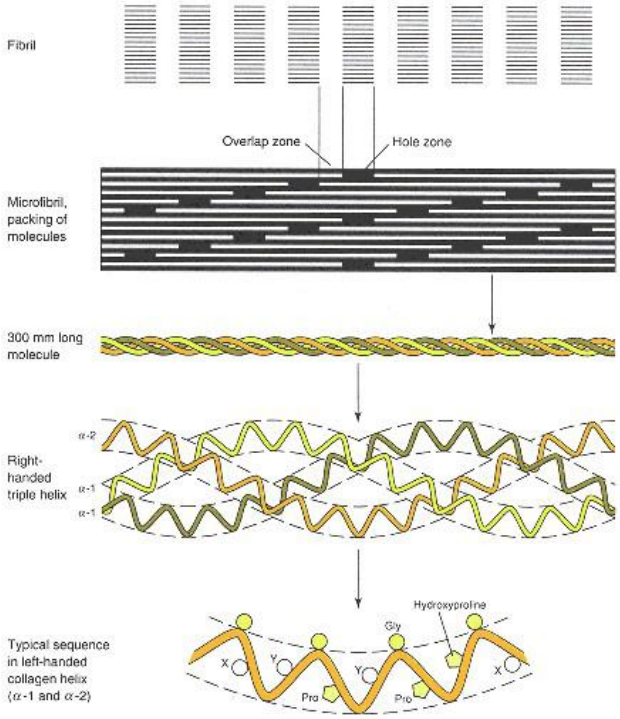


Figure 17.21
Collagen structure, illustrating the regularity of the primary sequence, the left-handed helix, the right-handed triple helix, the 300-nm molecule, and the organization of molecules in a typical fibril, within which collagen molecules are cross-linked.

Gly- sequence (hence 4-hydroxyproline is found only at Y positions in the -Gly-X-Y- sequence). Also present in the ER is a prolyl 3-hydroxylase, which modifies a smaller number of proline residues, and a lysyl hydroxylase, which modifies some of the Y-position lysine residues. These hydroxylases require Fe^{2+} and ascorbic acid, the extent of modification depending on the specific α-chain type. Proline hydroxylation stabilizes collagen and lysine hydroxylation provides sites for interchain cross-linking and for glycosylation by specific glycosyl transferases of the ER. Asparagine residues are also glycosylated at this point, eventually leading to high mannose-type oligosaccharides.

Triple helix assembly occurs after the polypeptide chains have been completed. Carboxy-terminal globular proprotein domains fold and disulfide bonds are formed. Interaction of these domains initiates winding of the triple helix from the carboxyl end toward the amino terminus. The completed triple helix, with globular proprotein domains at each end, moves to the Golgi apparatus where oligosaccharides are processed and matured. Sometimes tyrosine residues are modified by sulfation and some serines are phosphorylated. The completed procollagen is then released from the cell via secretory vesicles.

Collagen Maturation

Conversion of procollagen to collagen occurs extracellularly. The amino-terminal and carboxyl-terminal propeptides are cleaved by separate proteases that may also be type specific. Concurrently, the triple helices assemble into fibrils

TABLE 17.11 Selected Disorders in Collagen Biosynthesis and Structure

Disorder	*Collagen Defect*	*Clinical Manifestations*
Osteogenesis imperfecta 1	Decreased synthesis of type I	Long bone fractures prior to puberty
Osteogenesis imperfecta 2	Point mutations and exon rearrangements in triple helical regions	Perinatal lethality; malformed and soft, fragile bones
Ehlers–Danlos IV	Poor secretion, premature degradation of type III	Translucent skin, easy bruising, arterial and colon rupture
Ehlers–Danlos VI	Decreased hydroxylysine in types I and III	Hyperextensive skin, joint hypermobility
Ehlers–Danlos VII	Type I procollagen accumulation: N-terminal propeptide not cleaved	Joint hypermobility and dislocation
Cutis laxa (occipital horn syndrome)	Decreased hydroxylysine due to poor Cu distribution	Lax, soft skin; occipital horn formation in adolescents

and the collagen is stabilized by extensive cross-linking (see Figure 2.39). Lysyl oxidase converts some lysine or hydroxylysine to the reactive aldehydes, allysine, or hydroxyallysine. These residues condense with each other or with lysine or hydroxylysine residues in adjacent chains to form Schiff's base and aldol cross-links. Further and less well-characterized reactions can involve other residues including histidines and can link three α chains. Defects at many of these steps are known. Some of the best characterized are listed in Table 17.11 and described in Clin. Corr. 17.8.

17.7— Regulation of Translation

Translation requires considerable energy, and the formation of functioning proteins has significant consequences for the cell. It is logical that the process is carefully controlled, both globally and for specific proteins. The most efficient and common mechanism of regulation is at the initiation stage.

The best understood means of overall regulation of translation involves the reversible phosphorylation of eIF-2a. Under conditions that include nutrient starvation, heat shock, and viral infection, eIF-2a is phosphorylated by a specific kinase. Phosphorylated eIF-2a · GDP binds tightly to eIF-2b, the guanine nucleotide exchange factor, which is present in limiting amounts. Since eIF-2b is unavailable for nucleotide exchange, no eIF-2a · GTP is available for initiation. Phosphorylation can be catalyzed by a **heme-regulated inhibitor kinase,** which, in the absence of heme, is activated by autophosphorylation. This kinase is present in many cells but is best studied in reticulocytes that synthesize hemoglobin. Deficiencies in energy supply or any heme precursor activate the kinase. A related **double-stranded RNA-dependent kinase** is autophosphorylated and activated in response to binding of ds-RNA that results from many viral infections. Production of this kinase is also induced by interferon. Initiation factor eIF-4e (a component of the cap binding protein eIF-4f) is activated by phosphorylation in response to, for example, growth factors and is inactivated by a protein phosphatase following, for example, viral infection. These effects may be greatest in the translation of mRNAs with long, highly structured leader sequences that need to be unwound to allow identification of a translational start site.

Regulation of translation of specific genes also occurs. A clear example is the regulation by iron of synthesis of the iron-binding protein, ferritin. In

CLINICAL CORRELATION 17.8

Defects in Collagen Synthesis

Ehlers–Danlos Syndrome, Type IV

Ehlers–Danlos syndrome is a group of at least ten disorders that are clinically, genetically, and biochemically distinguishable, but that share manifestations of structural weaknesses in connective tissue. The usual problems are fragility and hyperextensibility of skin and hypermobility of the joints. The weaknesses result from defects in collagen structure. For example, type IV Ehlers–Danlos syndrome is caused by defects in type III collagen, which is particularly important in skin, arteries, and hollow organs. Characteristics include thin, translucent skin through which veins are easily seen, marked bruising, and sometimes an appearance of aging in the hands and skin. Clinical problems arise from arterial rupture, intestinal perforation, and rupture of the uterus during pregnancy or labor. Surgical repair is difficult because of tissue fragility. The basic defects in type IV Ehlers–Danlos appear to be due to changes in the primary structure of type III chains. These arise from point mutations that result in replacement of glycine residues and thus disruption of the collagen triple helix, and from exon-skipping, which shortens the polypeptide and can result in inefficient secretion and decreased thermal stability of the collagen, and in abnormal formation of type III collagen fibrils. In some cases type III collagen is accumulated in the rough ER, overmodified, and degraded very slowly.

Superti-Furga, A., Gugler, E., Gitzelmann, R., and Steinmann, B. Ehlers–Danlos syndrome type IV: a multi-exon deletion in one of the two COL 3A1 alleles affecting structure, stability, and processing of type III procollagen. *J. Biol. Chem.* 263:6226, 1988.

Osteogenesis Imperfecta

Osteogenesis imperfecta is a group of at least four clinically, genetically, and biochemically distinguishable disorders, all characterized by multiple fractures with resultant bone deformities. Several variants result from mutations producing modified α(I) chains. In the clearest example a deletion mutation causes absence of 84 amino acids in the α1(I) chain. The shortened α1(I) chains are synthesized, because the mutation leaves the reading frame in register. The short α1(I) chains associate with normal α1(I) and α2(I) chains, thereby preventing normal collagen triple helix formation, with resultant degradation of all the chains, a phenomenon aptly named "protein suicide." Three-fourths of all the collagen molecules formed have at least one short (defective) α1(I) chain, an amplification of the effect of a heterozygous gene defect. Other forms of osteogenesis imperfecta result from point mutations that substitute another amino acid for one of the glycines. Since glycine has to fit into the interior of the collagen triple helix, these substitutions destabilize that helix.

Barsh, G. S., Roush, C. L., Bonadio, J., Byers, P. H., and Gelinas, R. E. Intron mediated recombination causes an α(I) collagen deletion in a lethal form of osteogenesis imperfecta. *Proc. Natl. Acad. Sci. USA* 82:2870, 1985.

Scurvy and Hydroxyproline Synthesis

Scurvy results from dietary deficiency of ascorbic acid. Most animals can synthesize ascorbic acid from glucose but humans have lost this enzymatic mechanism. Among other problems, ascorbic acid deficiency causes decreased hydroxyproline synthesis because prolyl hydroxylase requires ascorbic acid. The hydroxyproline provides additional hydrogen-bonding atoms that stabilize the collagen triple helix. Collagen containing insufficient hydroxyproline loses temperature stability, becoming less stable than normal collagen at body temperature. The resultant clinical manifestations are distinctive and understandable: suppression of the orderly growth process of bone in children, poor wound healing, and increased capillary fragility with resultant hemorrhage, particularly in the skin. Severe ascorbic acid deficiency leads secondarily to a decreased rate of procollagen synthesis.

Crandon, J. H., Lund, C. C., and Dill, D. B. Experimental human scurvy. *N. Engl. J. Med.* 223:353, 1940.

Deficiency of Lysyl Hydroxylase

In type VI Ehlers–Danlos syndrome lysyl hydroxylase is deficient. As a result type I and III collagens in skin are synthesized with decreased hydroxylysine content, and subsequent cross-linking of collagen fibrils is less stable. Some cross-linking between lysine and allysine occurs but these are not as stable and do not mature as readily as do hydroxylysine-containing cross-links. In addition, carbohydrates add to the hydroxylysine residues but the function of this carbohydrate is unknown. The clinical features include marked hyperextensibility of the skin and joints, poor wound healing, and musculoskeletal deformities. Some patients with this form of Ehlers–Danlos syndrome have a mutant form of lysyl hydroxylase with a higher Michaelis constant for ascorbic acid than the normal enzyme. Accordingly, they respond to high doses of ascorbic acid.

Pinnell, S. R., Krane, S. M., Kenzora, J. E., and Glimcher, M. J. A heritable disorder of connective tissue: hydroxylysine-deficient collagen disease. *N. Engl. J. Med.* 286:1013, 1972.

Ehlers–Danlos Syndrome, Type VII

In Ehlers–Danlos syndrome, type VII, skin bruises easily and is hyperextensible, but the major manifestations are dislocations of major joints, such as hips and knees. Laxity of ligaments is caused by incomplete removal of the amino-terminal propeptide of the procollagen chains. One variant of the disease results from deficiency of procollagen *N*-protease. A similar deficiency occurs in the autosomal recessive disease called dermatosparaxis of cattle, sheep, and cats, in which skin fragility is so extreme as to be lethal. In other variants the proα1(I) and proα2(I) chains lack amino acids at the cleavage site because of skipping of one exon in the genes. This prevents normal cleavage by procollagen *N*-protease.

Cole, W. G., Chan, W., Chambers, G. W., Walker, I. D., and Bateman, J. F. Deletion of 24 amino acids from the proα(I) chain of type I procollagen in a patient with the Ehlers–Danlos syndrome type VII. *J. Biol. Chem.* 261:5496, 1986.

Occipital Horn Syndrome

In type IX Ehlers–Danlos syndrome and in Menke's (kinky-hair) syndrome there is thought to be a deficiency in lysyl oxidase activity. In type IX Ehlers–Danlos syndrome there are consequent cross-linking defects manifested in lax, soft skin and in the appearance during adolescence of bony occipital horns. Copper-deficient animals have deficient cross-linking of elastin and collagen, apparently because of the requirement for cuprous ion by lysyl oxidase.

(continued)

In Menke's (kinky-hair) syndrome there is a defect in intracellular copper transport that results in low activity of lysyl oxidase, and in occipital horn syndrome there is also a defect in intracellular copper distribution. A woman taking high doses of the copper-chelating drug, *d*-penicillamine, gave birth to an infant with an acquired Ehlers–Danlos-like syndrome, which subsequently cleared. Side effects of *d*-penicillamine therapy include poor wound healing and hyperextensible skin.

Peltonen, L., Kuivaniemi, H., Palotie, A., et al. Alterations of copper and collagen metabolism in the Menkes syndrome and a new subtype of Ehlers–Danlos syndrome. *Biochemistry* 22:6156, 1983. For a detailed overview of collagen disorders see: Byers, P. H. Disorders of collagen biosynthesis and structure. In: C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle (Eds.) *The Metabolic and Molecular Basis of Inherited Disease*, Vol. III, 7th ed. New York: McGraw-Hill, 1995, pp. 4029–4077.

the absence of iron, a repressor protein binds to the **iron-responsive element** (IRE), a stem-loop structure in the 5 leader sequence of ferritin mRNA. This mRNA is sequestered for future use. -Aminolevulinic acid synthase, an enzyme of heme biosynthesis, is also regulated by a 5 -IRE in its mRNA. In contrast, more ferritin receptor mRNA is needed if iron is limited; it has IREs in its 3 -nontranslated region. Binding of the repressor protein stabilizes the mRNA and prolongs its useful lifetime. Many growth-regulated mRNAs, including those for ribosomal proteins, have a polypyrimidine tract in their leader sequence. A polypyrimidine-binding protein helps regulate their translation.

17.8— Protein Degradation and Turnover

Proteins have finite lifetimes. They are subject to environmental damage such as oxidation, proteolysis, conformational denaturation, or other irreversible modifications. Equally important, cells need to change their protein complements in order to respond to different needs and situations. Specific proteins have very different lifetimes. Cells of the eye lens are not replaced and their proteins are not recycled. Hemoglobin in red blood cells lasts the life of these cells, about 120 days. Other proteins have lifetimes measured in days, hours, or even minutes. Some blood-clotting proteins survive for only a few days, so hemophiliacs are only protected for a short period by transfusions or injections of required factors. Diabetics require insulin injections regularly since the hormone is metabolized. Metabolic enzymes vary quantitatively depending on need; for example, urea cycle enzyme levels change in response to diet. Most amino acids produced by protein degradation are recycled to synthesize new proteins but some degradation products will be excreted. In either case, proteolysis first reduces the proteins in question to peptides and eventually amino acids. Several proteolytic systems accomplish this end.

Intracellular Digestion of Some Proteins Occurs in Lysosomes

Digestive proteases such as pepsin, trypsin, chymotrypsin, and elastase hydrolyze dietary protein and have no part in intracellular protein turnover within an organism (see Chapter 25). Intracellular digestion of proteins from the extracellular environment occurs within **lysosomes.** Material that is impermeable to the plasma membrane is imported by endocytosis. In **pinocytosis** large particles, molecular aggregates, or other molecules present in the extracellular fluid are ingested by engulfment. Macrophages ingest bacteria and dead cells by this mechanism. **Receptor-mediated endocytosis** uses cell surface receptors to bind specific molecules. Endocytosis occurs at pits in the cell surface that are coated internally with the multisubunit protein clathrin. Uptake is by invagination of the plasma membrane and the receptors to form intracellular coated vesicles. One fate of such vesicles is fusion with a lysosome and degradation of the contents. Some intracellular protein turnover may also occur within

lysosomes, and under some conditions significant amounts of cellular material can be mobilized via lysosomes. For example, serum starvation of fibroblasts in culture or starvation of rats leads to the lysosomal degradation of a subpopulation of cellular proteins. Recognition of a specific peptide sequence is involved, indicating that the lifetime of a protein is ultimately encoded in its sequence. This concept will be more apparent in the next section on ubiquitin-dependent proteolysis.

Although lysosomal degradation of cellular proteins occurs, it is not the main route of protein turnover. Calcium-dependent proteases, also called calpains, are present in most cells. Activators and inhibitors of these enzymes are also present, and calpains are logical candidates for enzymes involved in protein turnover. However, their role in these processes is not quantitatively established. Golgi and ER proteases degrade peptide fragments that arise during maturation of proteins in the secretory pathway. They could also be involved in turnover of ER proteins. **Apoptosis,** programmed cell death, requires several proteases. It is likely that other uncharacterized mechanisms exist in both the cytosol and in the mitochondrion.

Ubiquitin Is a Marker in ATP-Dependent Proteolysis

One well-described proteolytic pathway requires ATP hydrolysis and the participation of **ubiquitin,** a highly conserved protein containing 76 amino acids. One function of ubiquitin is to mark proteins for degradation. Ubiquitin has other roles; as an example, linkage of ubiquitin to histones H2A and H2B is unrelated to turnover since the proteins are stable, but modification may affect chromatin structure or transcription.

The ubiquitin-dependent proteolytic cycle is shown in Figure 17.22. Ubiquitin is activated by enzyme E1 to form a thioester; ATP is required and a transient AMP–ubiquitin complex is involved. The ubiquitin is then passed to enzyme E2, and finally via one of a group of E3 enzymes it is coupled to a targeted protein. Linkage of ubiquitin is through isopeptide bonds between -amino groups of lysine residues of the protein and the carboxyl-terminal glycine residues of ubiquitin. Several ubiquitin molecules may be attached to the protein and to each other. ATP-dependent proteases then degrade the tagged protein and free the ubiquitin for further degradation cycles.

Ubiquitin-dependent proteolysis plays a major role in the regulation of cellular events. **Cyclins** are involved in control of progress through the cell cycle. The ubiquitin-dependent destruction of a cyclin allows cells to pass from the M phase into G1. Other proteins known to be degraded by ubiquitin-dependent proteolysis include transcription factors, the p53 tumor suppressor and other oncoproteins, a protein kinase, and immune system and other cell surface receptors.

Damaged or mutant proteins are rapidly degraded via the ubiquitin pathway. In **cystic fibrosis** a mutation that results in deletion of one amino acid greatly alters the stability of a protein (see Clin. Corr. 17.9), but it is not always clear how native proteins are identified for degradation. Selectivity occurs at the level of the E3 enzyme, but most specific recognition signals are obscure. One determinant is simply the identity of the amino-terminal amino acid. Otherwise identical β-galactosidase proteins with different amino-terminal residues are degraded at widely differing rates. Amino termini may be modified to alter the lifetime of the protein, and some residues serve as aminoacyl acceptors for a destabilizing residue from an aminoacyl-tRNA. Internal sequences and conformation are also likely to be important; destabilizing PEST sequences (rich in Pro, Glu, Ser, and Thr) have been identified in several short-lived proteins.

The ATP-dependent degradation of ubiquitin-marked proteins occurs in a 26S organelle called the **proteasome.** Proteasomes are dumbbell-shaped complexes of about 25 polypeptides; a proteolytically active 20S cylindrical

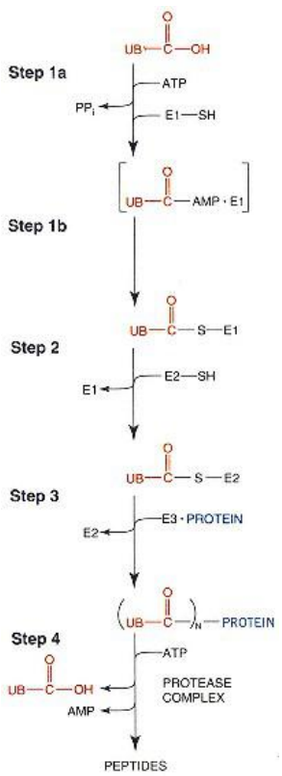


Figure 17.22
ATP and ubiquitin-dependent protein degradation.
Ubiquitin is first activated in a two-step reaction involving formation of a transient mixed anhydride of AMP and the carboxy terminus of ubiquitin (step 1a), followed by generation of a thioester with enzyme E1 (step 1b). Enzyme E2 can now form a thioester with ubiquitin (step 2) and serve as a donor in E3-catalyzed transfer of ubiquitin to a targeted protein (step 3). Several ubiquitin molecules are usually attached to different lysine residues of a targeted protein at this stage. Ubiquitinylated protein is now degraded by ATP-dependent proteolysis (step 4); ubiquitin is not degraded and can reenter the process at step 1.

Figure 17.23
Model of proteasome.
A 20S central segment is made up of four stacked heptameric rings of two types. The core is hollow and includes 12–15 different polypeptides; several proteases with different specificities are localized within the rings. V-shaped segments at each end cap the cylinder and are responsible for ATP-dependent substrate recognition, unfolding, and translocation into the proteolytic core. Upper cap structure is also in contact with the central segment but it is shown displaced from it in order to illustrate the hollow core of the cylinder. Adapted from D. Rubin and D. Finley, *Curr. Biol.* 5:854, 1995; and J.-M. Peters, *Trends Biochem. Sci.* 19:377, 1994.

core is capped at each end by **V**-shaped complexes that bestow ATP dependence (Figure 17.23). It is speculated that the cap structure is involved in recognizing and unfolding polypeptides and transporting them to the proteolytic core. The complex *E. coli* proteases Lon and Clp and similar enzymes in other microorganisms (and in mitochondria) also require ATP hydrolysis for their action, but ubiquitin is absent in prokaryotes and the means of identification of proteins for degradation is still obscure. It is likely that protein degradation will turn out to be as complex and important a problem as protein biosynthesis.

CLINICAL CORRELATION 17.9

Deletion of a Codon, Incorrect Posttranslational Modification, and Premature Protein Degradation: Cystic Fibrosis

Cystic fibrosis (CF) is the most common autosomal recessive disease in Caucasians, with a frequency of almost 1 per 2000. The CF gene is 230 kb in length and includes 27 exons encoding a protein of 1480 amino acids. The protein known as the cystic fibrosis transmembrane conductance regulator or CFTR is a member of a family of ATP-dependent transport proteins and it includes two membrane-spanning domains, two nucleotide-binding domains that interact with ATP, and one regulatory domain that includes several phosphorylation sites. CFTR functions as a cyclic AMP-regulated chloride channel. CF epithelia are characterized by defective electrolyte transport. The organs most strongly affected include the lungs, pancreas, and liver, and the most life-threatening effects involve thick mucous secretions that lead to chronic obstructive lung disease and persistent infections of lungs.

In about 70% of affected individuals the problem is traced to a three-nucleotide deletion that results in deletion of a single amino acid, phenylalanine 508, normally located in ATP-binding domain 1 on the cytoplasmic side of the plasma membrane. As with several other CF mutations, the Phe 508 deletion protein is not properly glycosylated or transported to the cell surface. Instead, it is only partially glycosylated, and it is degraded within the endoplasmic reticulum. It is postulated that the mutant protein does not fold properly and is marked for degradation rather than movement to the plasma membrane.

Ward, C., Omura, S., and Kopito, R. Degradation of CFTR by the ubiquitin–proteasome pathway. Cell 83:121, 1995.

Bibliography

Ribosomes and Transfer RNA

Filonenko, V. V., and Deutscher, M. P. Evidence for similar structural organization of the multienzyme aminoacyl-tRNA complex *in vivo* and *in vitro*. *J. Biol. Chem.* 269:17375, 1994.

Frank, J., Zhu, J., Penczek, P., et al. A model of protein synthesis based on a new cryo-electron microscopy reconstruction of the *E. coli* ribosome. *Nature* 376:441, 1995.

Freist, W. Mechanisms of aminoacyl-tRNA synthetases: a critical consideration of recent results. *Biochemistry* 28:6787, 1989.

Gesteland, R. F., and Atkins, J. F. (Eds.). *The RNA World*. Cold Spring Harbor, NY: Cold Spring Harbor Press, 1993.

Herold, M., and Nierhaus, K. H. Incorporation of six additional proteins to complete the assembly map of the 50S subunit from *Escherichia coli* ribosomes. *J. Biol. Chem.* 262:8826, 1987.

Hou, Y.-M., Francklyn, C., and Schimmel, P. Molecular dissection of a transfer RNA and the basis for its identity. *Trends Biochem. Sci.* 14:233, 1989.

Jakubowski, H., and Goldman, E. Editing errors in selection of amino acids for protein synthesis. *Microbiol. Rev.* 56:412, 1992.

Nierhaus, K. H., Franceschi, F., Subramanian, A. R., Erdmann, V. A., and Wittmann-Liebold, B. (Eds.). *The Translational Apparatus: Structure, Function, Regulation, Evolution*. New York: Plenum Press, 1993.

Nomura, M. The role of RNA and protein in ribosome function: a review of early reconstitution studies and prospects for future studies. *Cold Spring Harbor Symp. Quant. Biol.* 52:653, 1987.

Perona, J., Rould, M., and Steitz, T. Structural basis for transfer RNA aminoacylation by *Escherichia coli* glutaminyl-tRNA synthetase. *Biochemstry* 32:8758, 1993.

Stark, H., Mueller, F., Orleva, E. V., et al. The 70S *Escherichia coli* ribosome at 23Å resolution: fitting the ribosomal RNA. *Structure* 3:815, 1995.

Verschoor, A., and Frank, J. Three-dimensional structure of the mammalian cytoplasmic ribosome. *J. Mol. Biol.* 214:737, 1990.

Protein Biosynthesis and its Regulation

Barrell, B., Anderson, S., Bankier, A. T., et al. Different pattern of codon recognition by mammalian mitochondrial tRNAs. *Proc. Natl. Acad. Sci. USA* 77:3164, 1980.

Berchtold, H., Reshetnikova, L., Reiser, C. O., et al. Crystal structure of active elongation factor T_u reveals major domain rearrangements. *Nature* 365:126, 1993.

Gnirke, A., Geigenmuller, U., Rheinberger, H., and Nierhaus, K. The allosteric three-site model for the ribosomal elongation cycle. *J. Biol. Chem.* 264:7291, 1989.

Merrick, W. C. Mechanism and regulation of eukaryotic protein synthesis. *Microbiol. Rev.* 56:291, 1992.

Merrick, W. C. Eukaryotic protein synthesis: an *in vitro* analysis. *Biochimie* 76:822, 1994.

Rhoads, R. E. Regulation of eukaryotic protein biosynthesis by initiation factors. *J. Biol. Chem.* 268:3017, 1993.

Samuel, C. E. The eIF2a protein kinases as regulators of protein synthesis in eukaryotes from yeasts to humans. *J. Biol. Chem.* 268:7603, 1993.

Zhouravleva, G., Frolova, L., LeGoff, X., et al. Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO J* 14:4065, 1995.

Ziff, E. B. Transcription and RNA processing by the DNA tumour viruses. Nature 287:491, 1980.

Protein Targeting and Posttranslational Modification

Dahms, N. M., Lobel, P., and Kornfeld, S. Mannose 6-phosphate receptors and lysosomal enzyme targeting. *J. Biol. Chem.* 264:12115, 1989.

Danpure, C. J. How can products of a single gene be localized to more than one subcellular compartment? *Trends Cell Biol.* 5:230, 1995.

Kornfeld, R., and Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 54:631, 1985.

Krishna, R. G., and Wold, F. Post-translational modification of proteins. *Adv. Enzymol.* 67:265, 1993.

Kuhn, K. The classical collagens. In: R. Mayne and R. E. Burgeson (Eds.), *Structure and Function of Collagen Types*. Orlando, FL: Academic Press, 1987, pp. 1–42.

Nothwehr, S. F., and Stevens, T. H. Sorting of membrane proteins in the yeast secretory pathway. *J. Biol. Chem.* 269:10185, 1994.

Paulson, J. C. Glycoproteins: what are the sugar chains for? *Trends Biochem. Sci.* 14:272, 1989.

Paulson, J. C., and Colley, K. J. Glycosyltransferases: structure, localization, and control of cell type-specific glycosylation. *J. Biol. Chem.* 264:17615, 1989.

Pfanner, N., Craig, E., and Meijer, M. The protein import machinery of the mitochondrial inner membrane. *Trends Biochem. Sci.* 19:368, 1994.

Pfeffer, S. R., and Rothman, J. E. Biosynthetic protein transport and sorting by the endoplasmic reticulum and Golgi. *Annu. Rev. Biochem.* 56:829, 1987.

Rachubinski, R. A., and Subramani, S. How proteins penetrate peroxisomes. *Cell* 83:525, 1995.

Rogenkamp, R. Targeting signals for protein import into peroxisomes. *Cell Biochem. Function* 10:193, 1992.

Rudd, P. M., et al. The effects of variable glycosylation on the functional activities of ribonuclease, plasminogen and tissue plasminogen activator. *Biochim. Biophys. Acta* 1248:1, 1995.

von Heijne, G. Signals for protein targeting into and across membranes. In: A. H. Maddy and J. R. Harris (Eds.), *Subcellular Biochemistry, Vol. 22: Membrane Biochemistry*. New York: Plenum Press, 1994, pp. 1–19.

Walter, P., and Johnson, A. E. Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* 10:87, 1994.

Wolin, S. From the elephant to *E. coli*: SRP-dependent protein targeting. *Cell* 77:787, 1994.

Protein Turnover and Proteasomes

Ciechanover, A. The ubiquitin–proteasome proteolytic pathway. *Cell* 79:13, 1994.

Dice, J. F. Molecular determinants of protein half-lives in eukaryotic cells. *FASEB J.* 1:349, 1987.

Gonda, D. K., Bachmair, A., Wünning, I., et al. Universality and structure of the *N*-end rule. *J. Biol. Chem.* 264:16700, 1989.

Jentsch, S. The ubiquitin-conjugation system. *Annu. Rev. Genet.* 26:179, 1992.

Kessel, M., Maurizi, M. R., Kim, B., et al. Homology in structural organization between *E. coli* C1pAP protease and the eukaryotic 26S proteasome. *J. Mol. Biol.* 250:587, 1995.

Löwe, J., Stock, D., Jap, B., et al. Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4A resolution. *Science* 268:533, 1995.

Rechsteiner, M. Natural substrates of the ubiquitin proteolytic pathway. *Trends Biochem. Sci.* 66:615, 1991.

Rock, K., Gramm, C., Rothstein, L., et al. Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class II molecules. *Cell* 78:761, 1994.

Rogers, S., Wells, R., and Rechsteiner, M. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* 234:364, 1986.

Wlodawer, A. Proteasome: a complex protease with a new fold and a distinct mechanism. *Structure* 3:417, 1995.

Questions

J. Baggott and C. N. Angstadt

1. Degeneracy of the genetic code denotes the existence of:

A. multiple codons for a single amino acid.

B. codons consisting of only two bases.

C. base triplets that do not code for any amino acid.

D. different protein synthesis systems in which a given triplet codes for different amino acids.

E. codons that include one or more of the "unusual" bases.

2. Deletion of a single base from a coding sequence of mRNA may result in a polypeptide product with any of the following EXCEPT:

A. a sequence of amino acids that differs from the sequence found in the normal polypeptide.

B. more amino acids.

C. fewer amino acids.

D. a single amino acid replaced by another amino acid.

3. During initiation of protein synthesis.

A. methionyl-tRNA appears at the A site of the 80S initiation complex.

B. eIF-3 and the 40S ribosomal subunit participate in forming a preinitiation complex.

C. eIF-2 is phosphorylated by GTP.

D. the same methionyl-tRNA is used as is used during elongation.

E. a complex consisting of mRNA, the 60S ribosomal subunit, and certain initiation factors is formed.

4. Requirements for eukaryotic protein synthesis include all of the following EXCEPT:

A. mRNA.

B. ribosomes.

C. GTP.

D. 20 different amino acids in the form of aminoacyl-tRNAs.

E. E. $fMet\text{-}tRNA_i^{Met}$.

5. During the elongation stage of eukaryotic protein synthesis:

A. the incoming aminoacyl-tRNA binds to the P site.

B. a new peptide bond is synthesized by peptidyl transferase site of the large ribosomal subunit in a GTP-requiring reaction.

C. the peptide, still bound to a tRNA molecule, is transloated to a different site on the ribosome.

D. streptomycin can cause premature release of the incomplete peptide.

E. peptide bond formation occurs by the attack of the carboxyl group of the incoming aminoacyl-tRNA on the amino group of the growing peptide chain.

6. Diphtheria toxin:

A. acts catalytically.

B. releases incomplete polypeptide chains from the ribosome.

C. inhibits translocase.

D. prevents release factor from recognizing termination signals.

E. attacks the RNA of the large subunit.

7. How many high-energy bonds are expended in the formation of one peptide bond?

A. 1

B. 2

C. 3

D. 4

E. 5

8. Formation of mature insulin includes all of the following EXCEPT:

A. removal of a signal peptide.

B. folding into a three-dimensional structure.

C. disulfide bond formation.

D. removal of a peptide from an internal region.

E. γ-carboxylation of glutamate residues.

9. 4-Hydroxylation of specific prolyl residues during collagen synthesis requires all of the following EXCEPT:

A. Fe^{2+}.

B. a specific amino acid sequence.

C. ascorbic acid.

D. succinate.

E. individual α-chains, not yet assembled into a triple helix.

10. In the formation of an aminoacyl-tRNA:

A. ADP and P_i are products of the reaction.

B. aminoacyl adenylate appears in solution as a free intermediate.

C. the aminoacyl-tRNA synthetase is believed to recognize and hydrolyze incorrect aminoacyl-tRNAs it may have produced.

D. there is a separate aminoacyl-tRNA synthetase for every amino acid appearing in the final, functional protein.

E. there is a separate aminoacyl-tRNA synthetase for every tRNA species.

11. During collagen synthesis, events that occur extracellularly include all of the following EXCEPT:

A. modification of prolyl residues.

B. amino-terminal peptide cleavage.

C. carboxyl-terminal peptide cleavage.

D. modification of lysyl residues.

E. covalent cross-linking.

12. In the functions of ubiquitin all of the following are true EXCEPT:

A. ATP is required for activation of ubiquitin.

B. ubiquitin-dependent degradation of proteins occurs in the lysosomes.

C. linkage of a protein to ubiquitin does not always mark it for degradation.

D. the identity of the N-terminal amino acid is one determinant of selection for degradation.

E. ATP is required by the protease that degrades the tagged protein.

Match each of the following numbered markers with the appropriate lettered target site.

A. export from the cell

B. lysosomes

C. mitochondria

D. nucleus

E. peroxisomes

13. Clusters of lysine and arginine amino acid residues.

14. Mannose 6-phosphate.

15. Positively charged amphiphilic α-helix.

16. Ser-Lys-Leu (SKL).

Answers

1. A This is the definition of degeneracy (p. 719). B and E are not known to occur, although sometimes tRNA reads only the first two bases of a triplet (wobble), and sometimes unusual bases occur in anticodons (p. 719). C denotes the stop (nonsense) codons (p. 719). D is a deviation from universality of the code, as found in mitochondria (p. 719).

2. D Deletion of a single base causes a frameshift mutation (p. 721). The frameshift would destroy the original stop codon; another one would be generated before or after the original location. In contrast, replacement of one base by another would cause replacement of one amino acid (missense mutation), unless a stop codon is thereby generated (p. 721).

3. B A: Methionyl-$tRNA_e^{Met}$ is used internally. E: mRNA associates first with the 40S subunit (p. 725).

4. E f-Met-$tRNA_i^{Met}$ is involved in initiation of protein synthesis in prokaryote (p. 725).

5. C A: The incoming aminoacyl-tRNA binds to the A site. B: Peptide bond formation requires no energy source other than the aminoacyl-tRNA (pp. 727 and 730). D: Streptomycin inhibits formation of the prokaryotic 70S initiation complex (analogous to the eukaryotic 80S complex) and causes misreading of the genetic code when the initiation complex is already formed (p. 734). E: The electron pair of the amino group carries out a nucleophilic attack on the carbonyl carbon.

6. A This toxin catalyzes the formation of an ADP ribosyl derivative of translocase, which irreversibly inactivates the translocase (p. 735).

7. D One ATP is converted to AMP during activation of an amino acid (p. 721), and two GTP are converted to GDP during elongation (pp. 727 and 730). The ATP AMP counts as two high-energy bonds expended.

8. E See p. 743. γ-Carboxylation is of special importance in several blood clotting proteins (p. 746).

9. D See pp. 746–747.

10. C. C. Bonds between a tRNA and an incorrect smaller amino acid may form but are rapidly hydrolyzed (p. 723). A and B: ATP and the amino acid react to form an enzyme-bound aminoacyl adenylate; PP_i is released into the medium (p. 721). D: Some amino acids, such as hydroxyproline and hydroxylysine, arise by co- or posttranslational modification (p. 747). E: An aminoacyl-tRNA synthetase may recognize any of several tRNAs specific for a given amino acid (p. 722).

11. A See p. 747. Some modification of lysyl residues also occurs intracellularly (p. 747).

12. E A–D: True (see p. 751). C: Linkage to histones does not result in their degradation.

13. D (see p. 743).

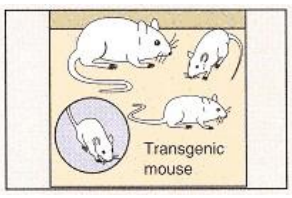
14. B (see p. 740).

15. C (see p. 742).

16. E (see p. 743). This tripeptide must occur at the carboxyl terminal.

Chapter 18—
Recombinant DNA and Biotechnology

Gerald Soslau



18.1— Overview

By 1970, the stage was set for modern molecular biology based on studies of numerous scientists in the previous 30 years, during which ignorance of what biochemical entity orchestrated the replication of life forms with such fidelity gave way to a state where sequencing and manipulating the expression of genes would be feasible. The relentless march toward a full understanding of gene regulation under normal and pathological conditions has moved with increasing rapidity since the 1970s. Deoxyribonucleic acid, composed of only four different nucleotides covalently linked by a sugar–phosphate backbone, is deceptively complex. Complexity is conferred on the DNA molecule by the nonrandom sequence of its bases, multiple conformations that exist in equilibrium in the biological environment, and specific proteins that recognize and associate with selected regions. By the 1970s biochemical knowledge of the cellular processes and their macromolecular components had established several facts required for the surge forward. It was clear that gene expression was highly regulated. Enzymes involved in DNA replication and RNA transcription had been purified and their function in the synthetic process defined. The genetic code had been broken. Genetic maps of prokaryotic chromosomes had been established based on gene linkage studies with thousands of different mutants. Finally, RNA species could be purified, enzymatically hydrolyzed into discrete pieces, and laboriously sequenced. It was evident that further progress in the understanding of gene regulation would require techniques to selectively cut DNA into homogeneous pieces. Even small, highly purified viral DNA genomes were too complex to decipher. The thought of tackling the human genome with more than 3×10^9 base pairs was all the more onerous.

Identification, purification, and characterization of restriction endonucleases that faithfully hydrolyze DNA molecules at specific sequences permitted the development of recombinant DNA methodologies. Development of DNA sequencing opened the previously tightly locked molecular biology gates to the secrets held within the organization of diverse biological genomes. Genes could finally be sequenced, but perhaps more importantly so could the flanking regions that regulate their expression. Sequencing regulatory regions of numerous genes defined consensus sequences such as those found in promoters, enhancers, and many binding sites for regulatory proteins (see Chapter 19). Each gene contains an upstream promoter where a DNA-dependent RNA polymerase binds prior to initiation of transcription. While some DNA regulatory sites lie just upstream of the transcription initiation site, other regulatory regions are hundreds to thousands of bases removed and still others are downstream.

This chapter presents many of the sophisticated techniques, developed in the past 25 years, that allow for the dissection of complex genomes into defined fragments with the complete analysis of the nucleotide sequence and function of these DNA regions. The modification and manipulation of genes, that is, genetic engineering, facilitates the introduction and expression of genes in both prokaryotic and eukaryotic cells. Many methodological approaches in genetic engineering have been greatly simplified by employment of a method that rapidly amplifies selected regions of DNA—the polymerase chain reaction (PCR). Proteins for experimental and clinical uses are readily produced by these procedures and it is anticipated that in the not too distant future these methods will allow for the rapid increase of treatment modalities of genetic diseases with gene replacement therapy. Current and potential uses of recombinant DNA technologies are also described. The significance to our society of advancements in the understanding of genetic macromolecules and their manipulation cannot be overstated.

18.2— Polymerase Chain Reaction

The rapid production of large quantities of a specific DNA sequence took a leap forward with the development of the **polymerase chain reaction (PCR).** The PCR requires two nucleotide oligomers that hybridize to the complementary DNA strands in a region of interest. The oligomers serve as primers for a DNA polymerase that extends each strand. Repeated cycling of the PCR yields large amounts of each DNA molecule of interest in a matter of hours as opposed to days and weeks associated with cloning techniques.

The PCR *amplification of a specific DNA sequence* can be accomplished with a purified DNA sample or a small region within a complex mixture of DNA. The principles of the reaction are shown in Figure 18.1. The nucleotide sequence of the DNA to be amplified must be known or it must be cloned in a vector (see p. 778) where the sequence of the flanking DNA has been established. The product of PCR is a double-stranded DNA molecule and the reaction is completed in each cycle when all of the template molecules have been copied. In order to initiate a new round of replication the sample is heated to melt the double-stranded DNA and, in the presence of excess oligonucleotide primers, cooled to permit hybridization of the single-stranded template with free oligomers. A new cycle of DNA replication will initiate in the presence of DNA polymerase and all four dNTPs. Heating to about 95°C as required for melting DNA inactivates most DNA polymerases, but a heat stable polymerase,

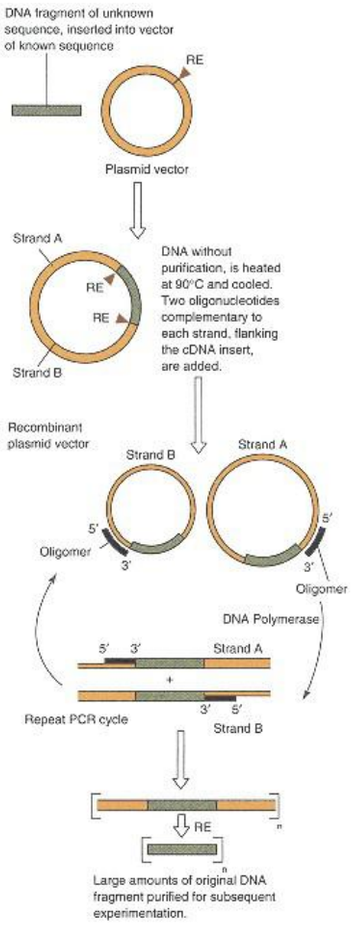


Figure 18.1
Polymerase chain reaction (PCR).
A DNA fragment of unknown sequence is inserted into a vector of known sequence by normal recombinant methodologies. The recombinant DNA of interest does not need to be purified from contaminating DNA species. The DNA is heated to 90°C to dissociate the double strands and cooled in the presence of excess amounts of two different complementary oligomers that hybridize to the known vector DNA sequences flanking the foreign DNA insert. Only recombinant single-stranded DNA species can serve as templates for DNA replication, yielding double-stranded DNA fragments of foreign DNA bounded by the oligomer DNA sequences. The heating–replication cycle is repeated many times to rapidly produce greatly amplified amounts of the original foreign DNA. The DNA fragment of interest can be purified from the polymerase chain reaction mixture by cleaving it with the original restriction endonuclease (RE), electrophoresing the DNA mixture through an agarose gel, and eluting the band of interest from the gel.

CLINICAL CORRELATION 18.1

Polymerase Chain Reaction and Screening for Human Immunodeficiency Virus

Use of the polymerase chain reaction (PCR) to amplify minute quantities of DNA has revolutionized the ability to detect and analyze DNA species. With PCR it is possible to synthesize sufficient DNA for analysis. Conventional methods for detection and identification of the human immunodeficiency virus (HIV), such as Southern blot–DNA hybridization and antigen analysis, are labor intensive, expensive, and have low sensitivity. An infected individual, with no sign of AIDS (acquired immunodeficiency syndrome), may test false negative for HIV by these procedures. Early detection of HIV infections in these individuals is crucial to initiate treatment and/or monitor the progression of their disease. In addition, a sensitive method is required to be certain that contributed blood from donors does not contain HIV. PCR amplification of potential HIV DNA sequences within DNA isolated from an individual's white blood cells permits the identification of viral infections prior to the presence of antibodies, the so-called seronegative state. Current methodologies are too costly to apply this testing to large-scale screening of donor blood samples. PCR can also be used to increase the sensitivity to detect and characterize DNA sequences of any other human infectious pathogen.

Kwok, S., and Sninsky, J. J. Application of PCR to the detection of human infectious diseases. In: H. A. Erlich (Ed.), *PCR Technology*. New York: Stockton Press, 1989, p. 235.

termed Taq DNA polymerase isolated from *Thermus aquaticus*, is now employed, obviating the need for fresh polymerase after each cycle. This has permitted the automation of PCR with each DNA molecule capable of being amplified one million-fold.

When the DNA to be amplified is present in very low concentrations relative to the total DNA in the sample, it is possible to amplify the DNA region of interest along with other spurious sequences. In this situation the specificity of the amplification reaction can be enhanced by **nested PCR.** After conducting the first PCR with one set of primers for 10–20 cycles, a small aliquot is removed for a second PCR. However, the second PCR is conducted with a new set of primers that are complementary to the template DNA just downstream of the first set of primers, or "nested" between the original set of primers. This process amplifies the DNA region of interest twice with a greatly enhanced specificity.

PCR has many applications including gene diagnosis, forensic investigations where only a drop of dried blood or a single hair is available, and evolutionary studies with preserved biological material. Use of PCR for screening for human immunodeficiency virus is presented in Clin. Corr. 18.1.

18.3— Restriction Endonuclease and Restriction Maps

Restriction Endonucleases Permit Selective Hydrolysis of DNA to Generate Restriction Maps

Nature possesses a diverse set of tools, the **restriction endonucleases,** capable of selectively dissecting DNA molecules of all sizes and origin into smaller fragments. These enzymes confer some protection on bacteria against invading viruses, that is, bacteriophage. The bacterial DNA sequences normally recognized by the restriction endonuclease may be protected from cleavage in the host cell by methylation of bases within the enzyme recognized palindrome while the unmethylated viral DNA is recognized as foreign and is hydrolyzed. Numerous Type II restriction endonucleases, with differing sequence specificities, have been identified and purified; many are now commercially available (see p. 609 for discussion of restriction endonuclease activities).

Restriction endonuclease permits construction of a new type of genetic map, the **restriction map,** in which the site of enzyme cleavage within the DNA is identified. Purified DNA species that contain restriction endonuclease sequences are subjected to restriction endonuclease cleavage. By regulating the time of exposure of the purified DNA molecules to restriction endonuclease cleavage, a population of DNA fragments that are partially to fully hydrolyzed can be generated. Separation of these enzyme-generated fragments by agarose gel electrophoresis allows for the construction of restriction maps; an example of this procedure with circular DNA is presented in Figure 18.2. Analysis of a DNA completely hydrolyzed by a restriction endonuclease establishes how many sites the restriction endonuclease recognizes within the molecule and what size fragments are generated. The size distribution of composite fragments generated by the partial enzymatic cleavage of the DNA molecules demonstrates linkage of all potential fragments. The sequential use of different restriction endonucleases has permitted a detailed restriction map of numerous circular DNA species including bacterial plasmids, viruses, and mitochondrial DNA. The method is also equally amenable to linear DNA fragments that have been purified to homogeneity.

Restriction Maps Permit the Routine Preparation of Defined Segments of DNA

Restriction maps may yield little information as to the genes or regulatory elements within the various DNA fragments. They have been used to demonstrate sequence diversity of organelle DNA, such as mitochondrial DNA, within species (see Clin. Corr. 18.2). Restriction maps can also be used to detect deletion mutations where a defined DNA fragment from the parental strain

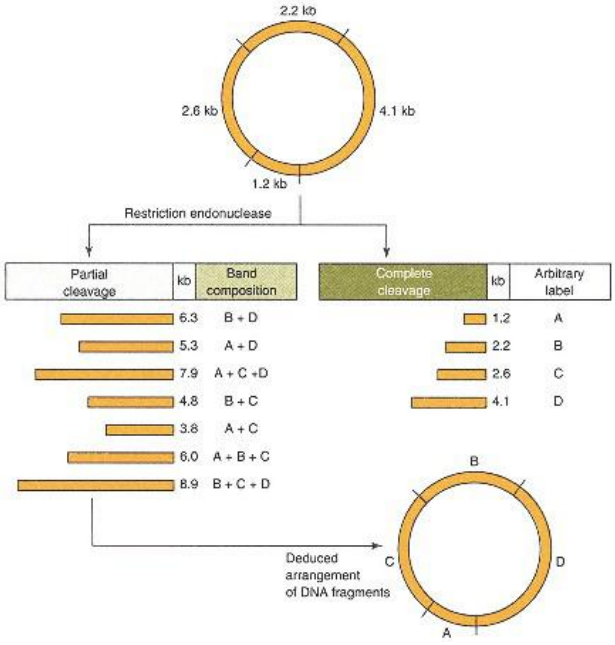


Figure 18.2
Restriction endonuclease mapping of DNA.
Purified DNA is subjected to restriction endonuclease digestion for varying times, which generates partially to fully cleaved DNA fragments. The DNA fragments are separated by agarose gel electrophoresis and stained with ethidium bromide. The DNA bands are visualized with a UV light source and photographed. The size of the DNA fragments is determined by the relative migration through the gel as compared to co-electrophoresed DNA standards. The relative arrangement of each fragment within the DNA molecule can be deduced from the size of the incompletely hydrolyzed fragments.

CLINICAL CORRELATION 18.2

Restriction Mapping and Evolution

In the past, evolutionary studies of species have depended solely on anatomical changes observed in fossil records and on carbon dating. More recently, these studies are being supported by molecular analysis of the sequence and size of selected genes or whole DNA molecules. Evolutionary alterations of a selected DNA molecule from different species can be rapidly assessed by restriction endonuclease mapping. Generation of restriction endonuclease maps requires a pure preparation of DNA. Mammalian mitochondria contain a covalently closed circular DNA molecule of approximately 16,000 base pairs that can rapidly be purified from cells. The mitochondrial DNA (mtDNA) can be employed directly for the study of evolutionary changes in DNA without the need of cloning a specific gene.

Mitochondrial DNA has been purified from the Guinea baboon, rhesus macaque, guenon, and human and cleaved with 11 different restriction endonucleases. Restriction maps were constructed for each species. The maps were all aligned relative to the direction and nucleotide site where DNA replication is initiated. A comparison of shared and altered restriction endonuclease sites allowed for calculation of the degree of divergence in nucleotide sequence between species. It was found that the rate of base substitution (calculated from the degree of divergence versus the time of divergence) has been about tenfold greater than changes in the nuclear genome. This high rate of mutation of the readily purified mtDNA molecule makes it an excellent model to study evolutionary relationships between species.

Brown, W. M., George, M. Jr., and Wilson, A. C. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 76:1967, 1979.

migrates as a smaller fragment in the mutated strain. Most importantly, the enzymatic microscissors used to generate restriction maps cut DNA into defined homogeneous fragments that can be readily purified. These maps are crucial for cloning and for sequencing genes and their flanking DNA regions.

18.4— DNA Sequencing

To determine the complexities of regulation of gene expression and to seek the basis for genetic diseases, techniques were necessary to determine the exact sequence of bases in DNA. In the late 1970s two different sequencing techniques were developed, one by A. Maxam and W. Gilbert, the chemical cleavage approach, and the other by F. Sanger, the enzymatic approach. Both procedures may employ the labeling of a terminal nucleotide, followed by the separation and detection of generated oligonucleotides.

Chemical Cleavage Method: Maxam–Gilbert Procedure

Requirements for this procedure include (1) labeling of the terminal nucleotide, (2) selective hydrolysis of the phosphodiester bond for each nucleotide separately to produce fragments with 1, 2, 3, or more bases, (3) quantitative separation of the hydrolyzed fragments, and (4) a qualitative determination of the label added in Step 1. The following describes one approach of the **Maxam–Gilbert procedure.** The overall approach is presented in Figure 18.3.

One end of each strand of DNA can be selectively radiolabeled with ^{32}P. This is accomplished when a purified double helix DNA fragment contains restriction endonuclease sites on either side of the region to be sequenced. Hydrolysis of the DNA with two different restriction endonucleases then results in different staggered ends, each with a different base in the first position of the single-stranded region. Labeling of the 3 end of each strand is accomplished with addition of the next nucleotide as directed by the corresponding base sequence on the complementary DNA strand. A fragment of *E. coli* DNA polymerase I, termed the **Klenow fragment,** will catalyze this reaction. The Klenow fragment, produced by partial proteolysis of the polymerase holoenzyme, lacks 5 3 exonuclease activity but retains the 3 5 exonuclease and polymerase activity. Each strand can therefore be selectively labeled in separate experiments. The complementary unlabeled strand will not be detectable when analyzing the sequence of the labeled strand.

The hydrolysis of the labeled DNA into different lengths is accomplished by first selectively destroying one or two bases of the four nucleotides. The procedure used exposes the phosphodiester bond connecting adjoining bases and permits selective cleavage of the DNA at the altered base. In separate chemical treatments, samples of labeled DNA are treated to alter purines and pyrimidines without disrupting the sugar–phosphate backbone; a method is not currently available to specifically alter adenine or thymine. Conditions for base modification are selected such that only one or a few bases are destroyed randomly within any one molecule. The four separate DNA samples are then reacted with piperidine, which chemically breaks the sugar–phosphate back-bone at sites where a base has been destroyed, generating fragments of different sizes. Since labeling is specific at the end while the chemical alteration of the base is random and not total, some of the fragments will be end labeled. For example, wherever a cytosine residue had been randomly destroyed in the appropriate reaction tube a break will be introduced into the DNA fragment. The series of chemically generated, end-labeled DNA fragments from each of the four tubes are electrophoresed through a polyacrylamide gel. Bases destroyed near the end-labeled nucleotide will generate fragments that migrate faster through the gel, as low molecular weight species, while fragments derived

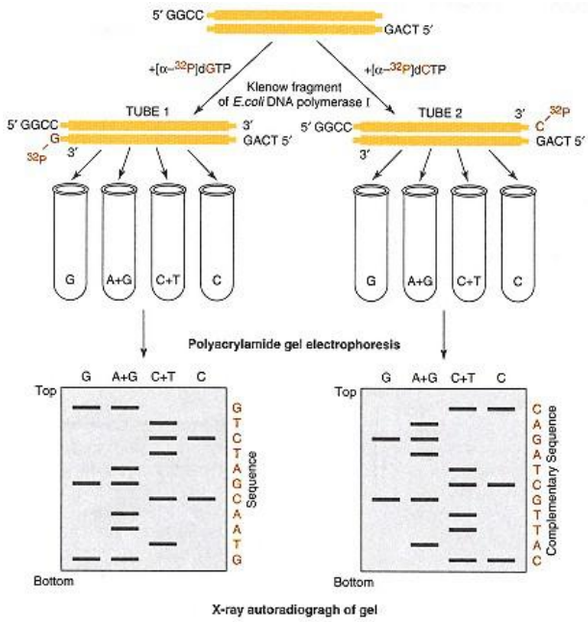


Figure 18.3
Maxam–Gilbert chemical method to sequence DNA.
A double-stranded DNA fragment to be sequenced is obtained by restriction endonuclease cleavage and purified. Both strands are sequenced by selectively labeling the ends of each DNA strand. One strand of DNA is end-labeled with [^{32}P]dGTP in reaction tube 1 while the other is end-labeled with [^{32}P]dCTP in reaction tube 2. The end-labeled DNA is then subdivided into four fractions where the different bases are chemically destroyed at random positions within the single-stranded DNA molecule. The less selective chemical destruction of adenine simultaneously destroys G and the destruction of thymine destroys the C bases. The single-stranded DNA is cleaved at the sites of the destroyed bases. This generates end-labeled fragments of all possible lengths corresponding to the distance from the end to the sites of base destruction. Labeled DNA fragments are separated according to size by electrophoresis. The DNA sequence can then be determined from the electrophoretic patterns detected on autoradiograms.

from bases destroyed more distant from the end will migrate through the gel more slowly as higher molecular weight molecules. The gel is then exposed to X-ray film, which detects the ^{32}P, and the radioactively labeled bands within the gel can be visualized. The sequence can be read manually or by automated methods directly from the X-ray autoradiograph beginning at the bottom (smaller fragments) and proceeding toward the top of the film (larger fragments). Sequencing the complementary strand checks the correctness of the sequence.

Interrupted Enzymatic Cleavage Method: Sanger Procedure

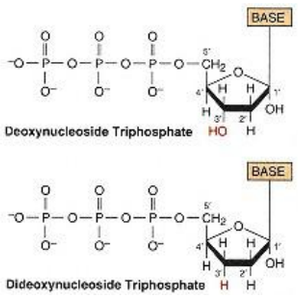


Figure 18.4
Structure of deoxynucleoside triphosphate and dideoxynucleoside triphosphate.
The 3 -OH group is lacking on the ribose component of the dideoxynucleoside triphosphate (ddNTP). This molecule can be incorporated into a growing DNA molecule through a phosphodiester bond with its 5 -phosphates. Once incorporated, the ddNTP blocks further synthesis of the DNA molecule since it lacks the 3 -OH acceptor group for an incoming nucleotide.

The **Sanger procedure** of DNA sequencing is based on the random termination of a DNA chain during enzymatic synthesis. The technique is possible because the dideoxynucleotide analog of each of the four normal nucleotides (Figure 18.4) can be incorporated into a growing DNA chain by DNA polymerase. The ribose of the **dideoxynucleotide triphosphate (ddNTP)** has the OH group at both the 2 and 3 positions replaced with a proton, whereas dNTP has only a single OH group replaced by a proton at the 2 position. Thus the ddNTP incorporated into the growing chain is unable to form a phosphodiester bond with another dNTP because the 3 position of the ribose does not contain an OH group. The growing DNA molecule can be terminated at random points, from the first nucleotide incorporated to the last, by including in the reaction system both the normal nucleotide and the ddNTP (e.g., dATP and ddATP) at concentrations such that the two nucleotides compete for incorporation.

Identification of DNA fragments requires labeling of the 5 end of the DNA molecules or the incorporation of labeled nucleotides during synthesis. The technique, outlined in Figure 18.5, is best conducted with pure single-stranded DNA; however, denatured double-stranded DNA can be used. Today, the DNA

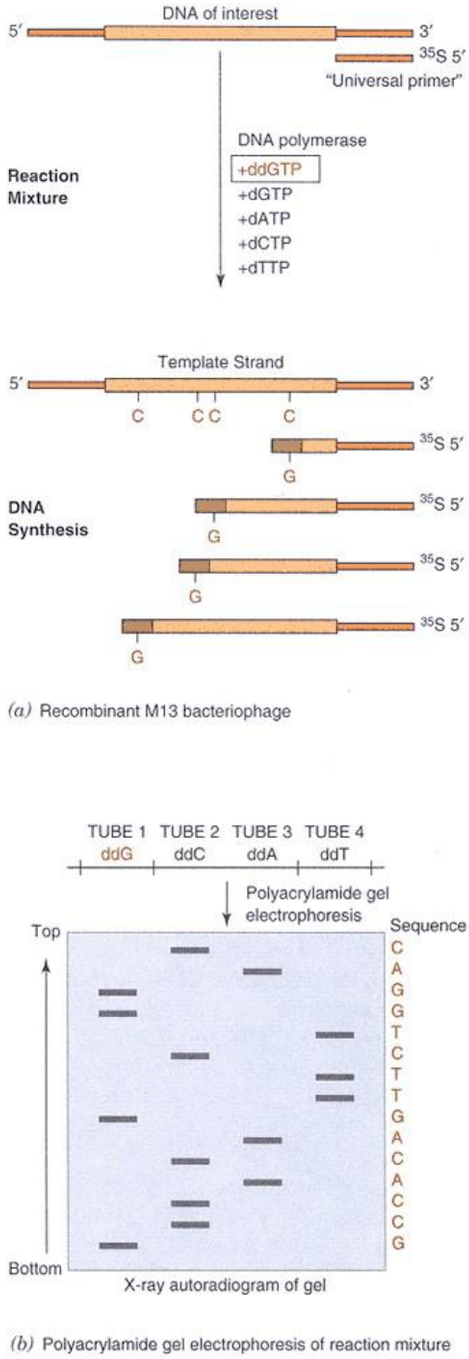


Figure 18.5
Sanger dideoxynucleoside triphosphate method to sequence DNA.
The DNA region of interest is inserted into bacteriophage DNA molecule. Replicating bacteriophage produces a single-stranded recombinant DNA molecule that is readily purified. The known sequence of the bacteriophage DNA downstream of the DNA insert serves as a hybridization site for an end-labeled oligomer with a complementary sequence, a universal primer. Extension of this primer is catalyzed with a DNA polymerase in the presence of all four deoxynucleoside triphosphates plus one dideoxynucleoside triphosphate, for example, ddGTP. Synthesis stops whenever a dideoxynucleoside triphosphate is incorporated into the growing molecule. Note that the dideoxynucleotide competes for incorporation with the deoxynucleotide. This generates end-labeled DNA fragments of all possible lengths that are separated by electrophoresis. The DNA sequence can then be determined from the electrophoretic patterns.

to be sequenced is frequently isolated from a recombinant single-stranded bacteriophage (see p. 778) where a region flanking the DNA of interest contains a sequence that is complementary to a universal primer. The primer can be labeled with either ^{32}P or ^{35}S nucleotide. Primer extension is accomplished with one of several different available DNA polymerases; one with great versatility is a genetically engineered form of the bacteriophage T7 DNA polymerase. The reaction mixture, composed of the target DNA, labeled primer, and all four deoxynucleoside triphosphates, is divided into four tubes, each containing a different dideoxynucleoside triphosphate. The ddNTPs are randomly incorporated during the enzymatic synthesis of DNA and cause termination of the chain.

Since the ddNTP is present in the reaction tube at a low level, relative to the corresponding dNTP, termination of DNA synthesis occurs randomly at all possible complementary sites to the DNA template. This yields DNA molecules of varying sizes, labeled at the 5 end, that can be separated by polyacrylamide gel electrophoresis. The labeled species are detected by X-ray autoradiography and the sequence is read.

Initially, this method required a single-stranded DNA template, production of a specific complementary oligonucleotide primer, and the need for a relatively pure preparation of the Klenow fragment of *E. coli* DNA polymerase I. These difficulties have been overcome and modifications have simplified the approach. The Sanger method can rapidly sequence as many as 400 bases while the Maxam–Gilbert method is limited to about 250 bases.

The PCR and Sanger methods can be combined for **direct sequencing** of small DNA regions of interest. The double-stranded PCR product is employed directly as template. Conditions are set such that one strand of melted DNA (template) anneals with the primer in preference to reannealing of template with its complementary strand, which would reform the original double-stranded DNA. Sequencing then follows the standard dideoxy chain termination reaction (typically with Sequenase in lieu of the Klenow polymerase) with synthesis of random-length chains occurring as extensions of the PCR primer. This method has been employed successfully for the diagnosis of genetic disorders (see Clin. Corr. 18.3).

18.5— Recombinant DNA and Cloning

DNA from Different Sources Can Be Ligated to Form a New DNA Species: Recombinant DNA

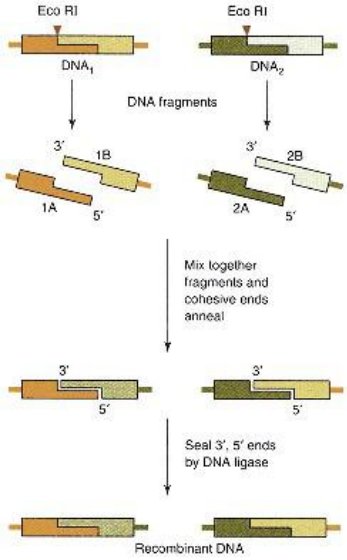


Figure 18.6
Formation of recombinant DNA from restriction endonuclease -generated fragments containing cohesive ends.
Many restriction endonucleases hydrolyze DNA in a staggered fashion, yielding fragments with single-stranded regions at their 5 and 3 ends. DNA fragments generated from different molecules with the same restriction endonuclease have complementary single-stranded ends that can be annealed and covalently linked together with a DNA ligase. All different combinations are possible in a mixture. When two DNA fragments of different origin combine it results in a recombinant DNA molecule.

The ability to selectively hydrolyze a population of DNA molecules with a battery of restriction endonucleases led to the development of a technique for joining two different DNA molecules termed **recombinant DNA.** This technique combined with the various techniques for replication, separation, and identification permits the production of large quantities of purified DNA fragments. The combined techniques, referred to as **recombinant DNA technologies,** allow the removal of a piece of DNA out of a larger complex molecule, such as the genome of a virus or human, and amplification of the DNA fragment. Recombinant DNAs have been prepared with DNA fragments from bacteria combined with fragments from humans, viruses with viruses, and so on. The ability to join two different pieces of DNA together at specific sites within the molecules is achieved with two enzymes, a restriction endonuclease and a **DNA ligase.** There are a number of different restriction endonucleases, varying in their nucleotide sequence specificity, that can be used (Section 18.3). Some hydrolyze the two strands of DNA in a staggered fashion, producing "*sticky* or *cohesive*" *ends* (Figure 18.6), while others cut both strands symmetrically, producing a *blunt end*. A specific restriction enzyme cuts DNA at exactly the same nucleotide sequence site regardless of the source of the DNA (bacteria, plant, mammal, etc.). A DNA molecule may have one, several, hundreds, thousands, or no recognition sites for a particular restriction endonuclease. The staggered cut results in a fragmented DNA molecule with ends that are single stranded. When different DNA fragments generated by the same restriction endonuclease are mixed, their single-stranded ends can hybridize, that is, anneal together. In the presence of DNA ligase the two fragments are connected covalently, producing a recombinant DNA molecule.

The DNA fragments produced from restriction endonuclease that form blunt ends can also be ligated but with much lower efficiency. The efficiency can be increased by enzymatically adding a poly(dA) tail to one species of DNA and a poly(dT) tail to the ends of the second species of DNA. The DNA fragments

CLINICAL CORRELATION 18.3

Direct Sequencing of DNA for Diagnosis of Genetic Disorders

The X-linked recessive hemorrhagic disorder hemophilia B is caused by a coagulation factor IX deficiency. The factor IX gene has been cloned and sequenced and contains 8 exons spanning 34 kb that encode a glyco-protein secreted by the liver. Over 300 mutations of the gene have been discovered of which about 85% are single base substitutions and the rest are complete or partial gene deletions. Several methods have been employed to identify carriers of a defective gene copy and for prenatal diagnoses. Unfortunately, these methods were costly, time consuming, and all too often inaccurate. Direct sequencing of PCR amplified genomic DNA has been employed to circumvent these diagnostic shortcomings. Between 0.1 and 1 g of genomic DNA can readily be isolated from patient blood samples and each factor IX exon can be PCR amplified with appropriate primers. The amplified DNA can then be used for direct sequencing to determine if a mutation in the gene exists that would be diagnostic of one of the forms of hemophilia B. For example, a patient with a moderate hemophilia B (London 6) had an A G transition at position 10442 that led to a substitution of Asp 64 by Gly.

Green, P. M., Bentley, D. R., Mibashan, R. S., Nilsson, I. M., and Gianelli, F. Molecular pathology of hemophilia b. *EMBO J.* 8:1067, 1989.

with complementary tails can be annealed and ligated in the same manner as fragments with restriction enzyme-generated cohesive ends.

Recombinant DNA Vectors Can Be Produced in Significant Quantities by Cloning

Synthesis of a recombinant DNA opens the way for production of significant quantities of interesting DNA fragments. By incorporating a recombinant DNA into a cellular system that allows replication of recombinant DNA, amplification of DNA of interest can be achieved. A carrier DNA, termed a **cloning vector,** is employed. Bacterial plasmids are ideally suited as recombinant DNA vectors. Many bacteria contain a single circular chromosome of approximately 4 million base pairs and minicircular DNA molecules called **plasmids.** Plasmids are usually composed of only a few thousand base pairs and are rarely associated with the large chromosomal molecule. Genes within the plasmid have various functions; one of the most useful is the ability to confer antibiotic resistance to the bacterium, an attribute useful in selecting specific colonies of the bacteria. Plasmids replicate independently of replication of the main bacterial chromosome. One type of plasmid, the **relaxed-control** plasmids, may be present in tens to hundreds of copies per bacterium, and replication is dependent solely on host enzymes that have long half-lives. Therefore replication of **"relaxed" plasmids** can occur in the presence of a protein synthesis inhibitor. Bacteria can accumulate several thousand plasmid copies per cell under these conditions. Other plasmid types are subjected to **stringent control** and their replication is dependent on the continued synthesis of plasmid-encoded proteins. These plasmids replicate at about the same rate as the large bacterial chromosome, and only a low number of copies occur per cell. The former plasmid type is routinely used for recombinant DNA studies.

The first practical recombinant DNA molecule that could be cloned involved as a vector the *E. coli* **plasmid pSC101,** which contains a single EcoRI restriction endonuclease site and a gene that encodes for a protein that confers antibiotic resistance to the bacteria. This plasmid contains an origin of replication and associated DNA regulatory sequences that are referred to as a **replicon.** This vector, however, suffers from a number of limiting factors. The single restriction endonuclease site limits the DNA fragments that can be cloned and the one antibiotic-resistance selectable marker reduces the convenience in selection; in addition, it replicates poorly.

Plasmid vectors with broad versatilities have been constructed using recombinant DNA technology. The desirable features of a plasmid vector include a relatively low molecular weight (3–5 kb) to accommodate larger fragments; several different restriction endonuclease sites useful in cloning a variety of restriction enzyme-generated fragments; multiple selectable markers to aid in selecting bacteria with recombinant DNA molecules; and a high rate of replication. The first plasmid constructed (Figure 18.7) to satisfy these requirements was **pBR322** and this plasmid has been used for the subsequent generation of newer vectors in use today. Most currently employed vectors contain an inserted sequence of DNA termed **polylinker, restriction site bank,** or **polycloning site,** which contains numerous restriction endonuclease sites unique to the plasmid.

DNA Can Be Inserted into Vector DNA in a Specific Direction: Directional Cloning

Directional cloning reduces the number of variable "recombinants" and enhances the probability of selection of the desired recombinant. Insertion of foreign DNA, with a defined polarity, into a plasmid vector in the absence of the plasmid resealing itself can be accomplished by employing two restriction

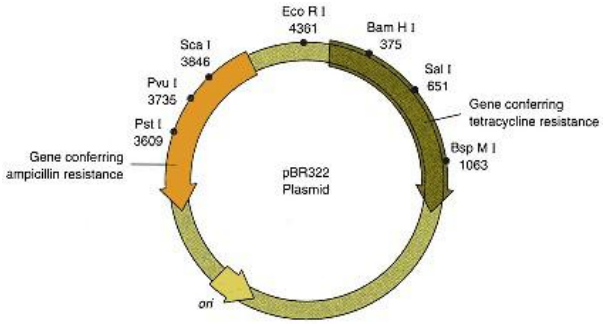


Figure 18.7
The pBR322 plasmid constructed in the laboratory to contain features that facilitate cloning foreign DNA fragments.
By convention, the numbering of the nucleotides begins with the first T in the unique EcoRI recognition sequence (GAATTC) and the positions on the map refer to the 5 base of the various restriction endonuclease-recognition sequences. Only a few of the unique restriction sites within the antibiotic resistance genes and none of the numerous sites where an enzyme cuts more than once within the plasmid are shown.

endonucleases to cleave the plasmids (Figure 18.8); vectors with polylinkers are ideally suited for this purpose. The use of two enzymes yields DNA fragments and linearized plasmids with different "sticky" ends. Under these conditions the plasmid is unable to reanneal with itself. In addition, the foreign DNA can be inserted into the vector in only one orientation. This is extremely important when one clones a potentially functional gene downstream from the promoter-regulatory elements in expression vectors (see p. 778).

Bacteria Can Be Transformed with Recombinant DNA

The process of artificially introducing DNA into bacteria is referred to as **transformation.** It is accomplished by briefly exposing the cells to divalent cations that make them transiently permeable to small DNA molecules. Recombinant plasmid molecules, containing foreign DNA, can be introduced into bacteria where it would replicate normally.

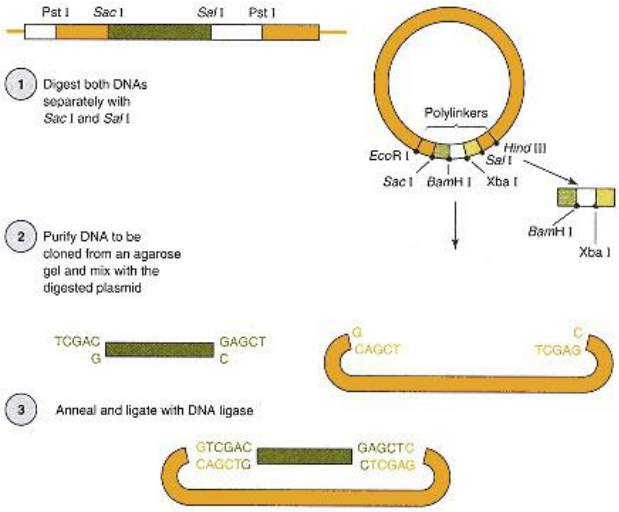


Figure 18.8
Directional cloning of foreign DNA into vectors with a specified orientation.
Insertion of a foreign DNA fragment into a vector with a specified orientation requires two different annealing sequences at each end of the fragment and the corresponding complementary sequence at the two ends generated in the vector. A polylinker with numerous unique restriction endonuclease sites within the vector facilitates directional cloning. Knowledge of the restriction map for the DNA of interest allows for selection of appropriate restriction endonucleases to generate specific DNA fragments that can be cloned in a vector.

It Is Necessary to Be Able to Select Transformed Bacteria

Once the plasmid has been introduced into the bacterium, both can replicate. Methods are available to select those bacteria that carry the recombinant DNA molecules. In the recombinant process some bacteria may not be transformed or may be transformed with a vector not carrying foreign DNA; in preparing the vector some may reanneal without inclusion of the DNA of interest. In some experimental conditions one can generate DNA fragments that can be readily purified for recombinant studies. Such fragments can be generated from small, highly purified DNA species, for example, some DNA viruses. More typically, however, a single restriction endonuclease will generate hundreds to hundreds of thousands of DNA fragments, depending on the size and complexity of DNA being studied. Individual fragments cannot be isolated from these samples to be individually incorporated into the plasmid. Methods have therefore been developed to select those bacteria containing the desired DNA.

Restriction endonucleases do not necessarily hydrolyze DNA into fragments containing intact genes. If the fragment contains an entire gene it may not contain the required flanking regulatory sequences, such as the promoter region. If the foreign gene is of mammalian origin, its regulatory sequences would not be recognized by the bacterial synthetic machinery. The primary gene transcript (pre-mRNA) can also contain introns that cannot be processed by the bacteria.

Recombinant DNA Molecules in a Gene Library

When a complex mixture of thousands of different genes, arranged on different chromosomes, as in the human genome, is subjected to hydrolysis with a single restriction endonuclease, thousands of DNA fragments are generated. These DNA fragments are annealed with a plasmid vector that has been cleaved to a linear molecule with the same restriction endonuclease. By adjusting the ratio of plasmid to foreign DNA the probability of joining at least one copy of each DNA fragment within a cyclized recombinant-plasmid DNA approaches one. Usually, only one out of the multiple DNA fragments is inserted into each plasmid vector. Bacteria are transformed with the recombinant molecules such that only one plasmid is taken up by a single bacterium. Each recombinant molecule can now be replicated within the bacterium and the bacterium will give rise to progeny, each carrying multiple copies of the recombinant DNA. The total population of bacteria now contain fragments of DNA that may represent the entire human genome. This is termed a **gene library.** As in any library containing thousands of volumes, a selection system must be available to retrieve the book or gene of interest.

Plasmids are commonly employed to clone DNA fragments generated from molecules of limited size and complexity, such as viruses, and to subclone large DNA fragments previously cloned in other vectors. Genomic DNA fragments are usually cloned from other vectors capable of carrying larger foreign DNA fragments than plasmids (see p. 780).

PCR May Circumvent the Need to Clone DNA

Cloning and amplification of a DNA fragment carried within a vector may be employed for subcloning, mutagenesis, and sequencing. The PCR has, in many instances, replaced the need to amplify recombinant DNA in a replicating biological system, greatly reducing the time and preparative steps required. It is not necessary to know the sequence of the DNA insert (up to 6 kb) to amplify it by the PCR, since the sequence of the vector DNA flanking the insert is known.

In some instances the PCR completely circumvents the need to clone the DNA of interest. For instance, a gene that has previously been cloned and sequenced can readily be analyzed in patient DNA for the detection of mutations

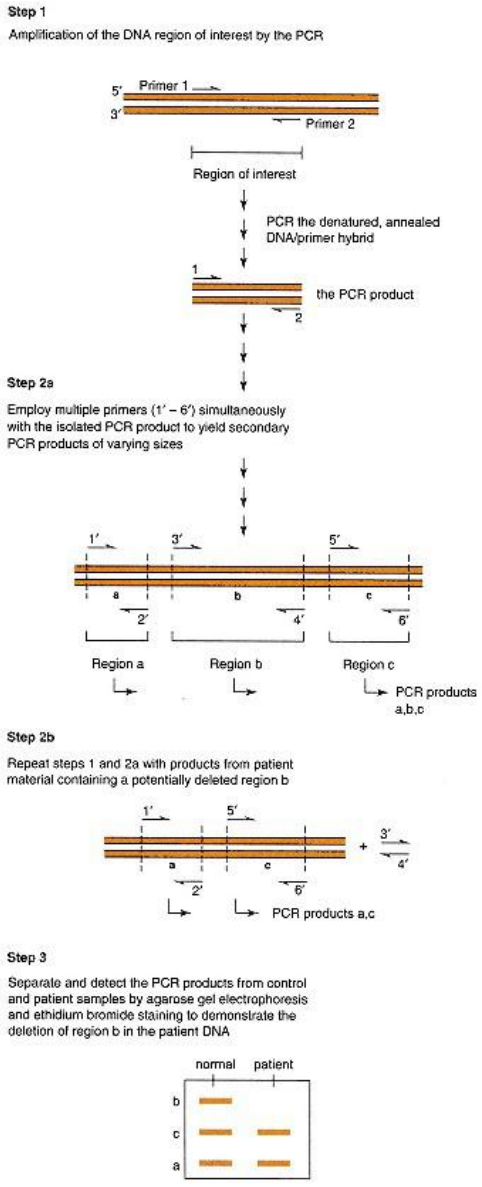


Figure 18.9
A multiplex PCR strategy to analyze a DNA region of interest for mutated alterations.
A region of DNA within a complex DNA molecule, derived from any source, can be amplified by the PCR with specific primers that are complementary to sequences flanking the DNA region of interest (Step 1). After multiple PCR cycles the amplified DNA (PCR product) can then be used as a template simultaneously for multiple pairs of primers (Step 2a) that are complementary throughout the DNA (here they cover three segments—a, b, and c). This procedure requires prior knowledge of the sequence of the normal DNA/gene. Step 2a is repeated for DNA derived from a patient with potential mutation(s) in the DNA region of interest (Step 2b). The amplified DNA products from the multiplex PCR step (Steps 2a and 2b) are then analyzed by agarose gel electrophoresis to ascertain if the patient sample contains a mutation (Step 3).

within this gene by a **multiplex PCR strategy.** DNA is isolated from patient blood cells and multiple pairs of oligonucleotide primers are synthesized to amplify the entire gene or selected regions within the gene (Figure 18.9). Analysis of the amplified DNA fragments by agarose gel electrophoresis would

CLINICAL CORRELATION 18.4

Multiplex PCR Analysis of HGPRTase Gene Defects in Lesch–Nyhan Syndrome

Lesch–Nyhan syndrome, as described in Clin. Corr. 12.2, results from a deficiency in hypoxanthine–guanine phosphoribosyl-transferase (HGPRTase) activity. Several variant forms of HGPRTase defects have been detected. Multiplex PCR amplification of the *HGPRT*gene locus has been employed to analyze this gene in cells derived from Lesch–Nyhan patients and results account for the variability of the HGPRTase. The gene, comprised of 9 exons, can be multiplex amplified using 16 different primers in a single PCR. The products can be separated by agarose gel electrophoresis. Analysis of the *HGPRT*gene locus by multiplex amplification of DNA derived from cells of several patients detected great variations in deletions of different exons to total absence of the exons.

Rossiter, B. J. F., et al. In: M. J. McPherson, P. Quirke, and G. R. Taylor (Eds.), *PCR. A Practical Approach*, Vol. 1. Oxford, England: Oxford University Press, 1994, p. 67.

allow one to detect any potential deletion mutation as compared to the normal gene products. Direct sequencing of multiple PCR products can be employed to detect point mutations in the patient gene. Multiplex PCR has been used to detect various defects in the HGPRTase gene in Lesch–Nyhan patients (see Clin. Corr. 18.4).

18.6— Selection of Specific Cloned DNA in Libraries

Loss of Antibiotic Resistance Is Used to Select Transformed Bacteria

When a single transformed bacterium carrying a recombinant molecule multiplies, its progeny are all genetically the same. If the transformed bacterium carries a recombinant DNA, all progeny will carry copies of the same recombinant plasmid. The foreign DNA has been amplified and is derived from a single cloned DNA fragment. The problem is how to identify the one colony containing the desired plasmid in a field of thousands to millions of different bacterial colonies. The plasmid construct pBR322 and its descendants carry two genes that confer antibiotic resistance. Within these antibiotic-resistant genes are DNA sequences sensitive to restriction endonuclease. When a fragment of foreign DNA is inserted into a restriction site within the gene for antibiotic resistance, the gene becomes nonfunctional. Bacteria carrying this recombinant plasmid are sensitive to the antibiotic (Figure 18.10). The second antibiotic resistance gene within the plasmid, however, remains intact and the bacteria will be resistant to this antibiotic. This technique of **insertional inactivation** of plasmid gene products affords a method to select bacteria that carry recombinant plasmids.

pBR322 contains genes that confer resistance to ampicillin (amp^r) and tetracycline (tet^r). A gene library with cellular DNA fragments inserted within the tet^r gene can be selected and screened in two stages (Figure 18.10). First, the bacteria are grown in an ampicillin-containing growth medium. Bacteria that are not transformed by a plasmid (they lack a normal or recombinant plasmid) during the construction of the gene library will not grow in the presence of the antibiotic, thus eliminating this population of bacteria. This, however, does not indicate which of the remaining viable bacteria carry a recombinant plasmid vector versus a plasmid with no DNA insert. The second step is to identify bacteria carrying recombinant vectors with nonfunctional tet^r genes, which are therefore sensitive to tetracycline.

Bacteria insensitive to ampicillin are plated and grown on agar plates containing ampicillin (Figure 18.10). Replica plates can be made by touching the colonies on the original agar plate with a filter and then touching additional sterile plates with the filter. All the plates will contain portions of each original colony at identifiable positions on the plates. The replica plate can contain tetracycline, which will not support the growth of bacteria harboring recombinant plasmids with their tet^r gene disrupted. Comparison of replica plates with and without tetracycline will indicate which colonies on the original ampicillin plate contain recombinant plasmids. Thus individual colonies containing the recombinant DNA can be selected, cultured, and analyzed.

Either **DNA** or **RNA probes** (see pp. 583 and 773) can be utilized to identify the DNA of interest. Ampicillin-resistant bacterial colonies on agar can be replica plated onto a nitrocellulose filter and adhering cells from each colony can be lysed with NaOH (Figure 18.10). DNA within the lysed bacteria is also denatured by the NaOH and becomes firmly bound to the filter. A labeled DNA or RNA probe that is complementary to the DNA of interest can be hybridized to the nitrocellulose-bound DNA. The filter is exposed to X-ray autoradiography. Any colony carrying the cloned DNA of interest will appear as a developed signal on the X-ray film. These spots would then correspond to the colony on the

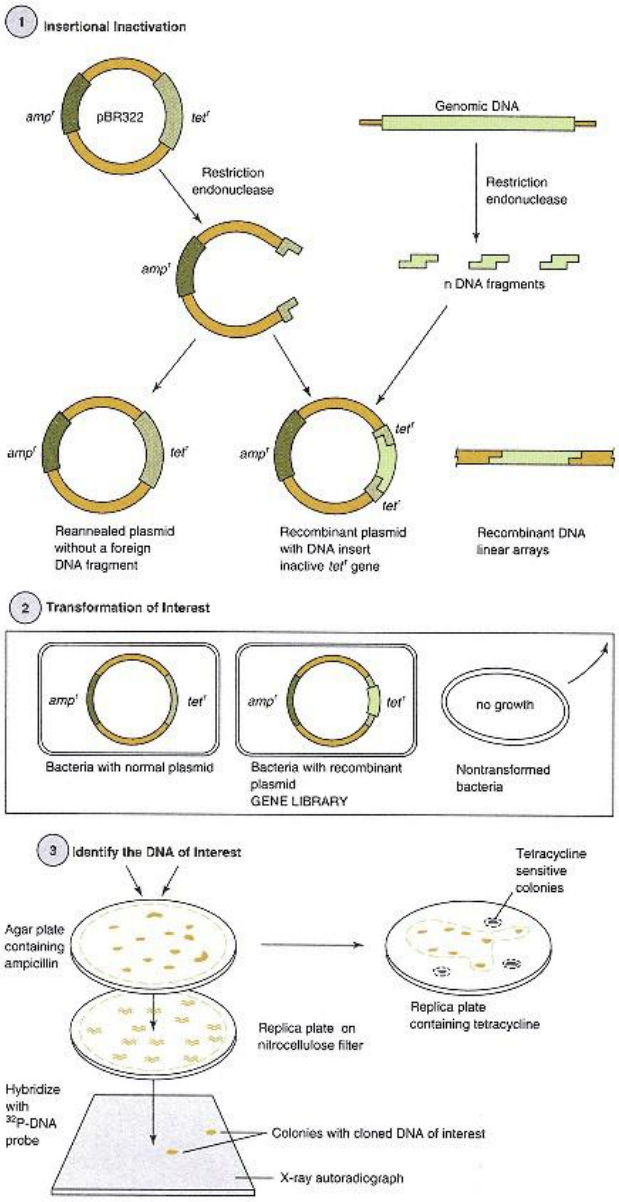


Figure 18.10
Insertional inactivation of recombinant plasmids and detection of transformed bacteria carrying a cloned DNA of interest.
When the insertion of a foreign DNA fragment into a vector disrupts a functional gene sequence, the resulting recombinant DNA does not express the gene. The gene that codes for antibiotic resistance to tetracycline (*tet*^{r}) is destroyed by DNA insertion while the ampicillin resistance gene (*amp*^{r}) remains functional. Destruction of one antibiotic resistance gene and retention of a second antibiotic resistance gene allow for the detection of bacterial colonies carrying the foreign DNA of interest within the replicating recombinant vector.

original agar plate that can then be grown in a large-scale culture for further manipulation.

Cloned and amplified DNA fragments usually do not contain a complete gene and are not expressed. The DNA inserts, however, can readily be purified for sequencing or used as probes to detect genes within a mixture of genomic DNA, transcription levels of mRNA, and pathological conditions via clinical diagnostic tests.

α-Complementation for Selecting Bacteria Carrying Recombinant Plasmids

Other selection techniques can identify bacteria carrying recombinant DNA molecules. Vectors have been constructed (the pUC series) such that selected bacteria transformed with these vectors carrying foreign DNA inserts can be identified visually (Figure 18.11). The pUC plasmids contain the regulatory sequences and a portion of the 5 -end coding sequence (N-terminal 146 amino acids) for the β-galactosidase gene (*lacZ* gene) of the *lac* operon (Chapter 19, p. 802). The translated N-terminal 146 amino acid fragment of β-galactosidase is an inactive polypeptide. Mutant *E. coli* that code for the missing inactive carboxy-terminal portion of β-galactosidase can be transformed with the pUC plasmids. The translation of the host cell and plasmid portions of the β-galactosidase in response to an inducer, isopropylthio-β-D-galactoside, complement each other, yielding an active enzyme. The process is referred to as **α-complementation.** When these transformed bacteria are grown in the presence of a chromogenic substrate (5-bromo-4-chloro-3-indolyl-β-D-galactoside[X-gall) for β-galactosidase they form blue colonies. If, however, a foreign DNA fragment is inserted into the base sequence for the N-terminal portion of β-galactosidase,

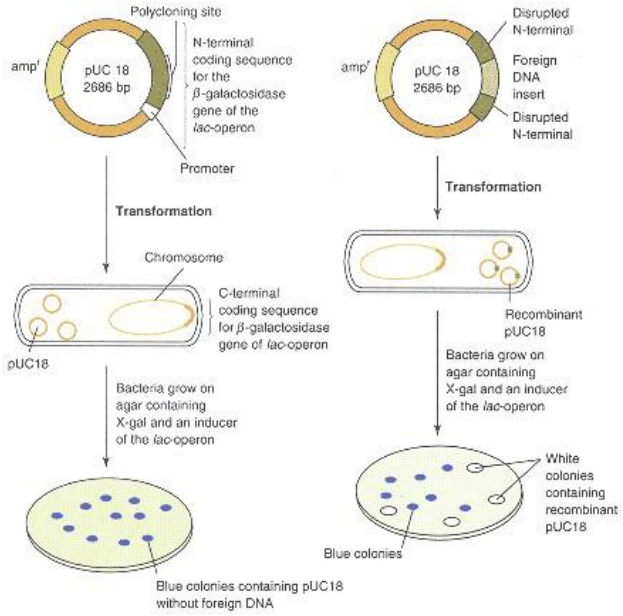


Figure 18.11
α-Complementation for detection of transformed bacteria.
A vector has been constructed (pUC 18) that expresses the N-terminal coding sequence for the enzyme β-galactosidase of the lac operon. Bacterial mutants coding for the C-terminal portion of β-galactosidase are transformed with pUC 18. These transformed bacteria, grown in the presence of a special substrate for the intact enzyme (X-gal), result in blue colonies because they contain the enzyme to react with substrate. The functional N-terminal and C-terminal coding sequences for the gene complement each other to yield a functional enzyme. If, however, a foreign DNA fragment insert disrupts the pUC 18 N-terminal coding sequence for β-galactosidase, bacteria transformed with this recombinant molecule will not produce a functional enzyme. Bacterial colonies carrying these recombinant vectors can then be visually detected as white colonies.

the active enzyme cannot be formed. Bacteria transformed with these recombinant plasmids and grown on X-gal will yield white colonies and can be selected visually from nontransformed blue colonies.

18.7— Techniques for Detection and Identification of Nucleic Acids

Nucleic Acids Can Serve As Probes for Specific DNA or RNA Sequences

Selection of bacteria harboring recombinant DNAs of interest, analysis of mRNAs expressed in a cell, or identification of the presence of DNA sequences within a genome require sensitive and specific detection methods. DNA and RNA **probes** meet these requirements. These probes contain nucleotide sequences complementary to the target nucleic acid and will thus hybridize with the nucleic acid of interest. The degree of complementarity of a probe with the DNA under investigation determines the tightness of binding of the probe. The probe does not need to contain the entire complementary sequence of the DNA. The probe, RNA or DNA, can be labeled, usually with ^{32}P. Nonradioactive labels are also employed that depend on enzyme substrates coupled to nucleotides, which when incorporated into the nucleic acid can be detected by an enzyme-catalyzed reaction.

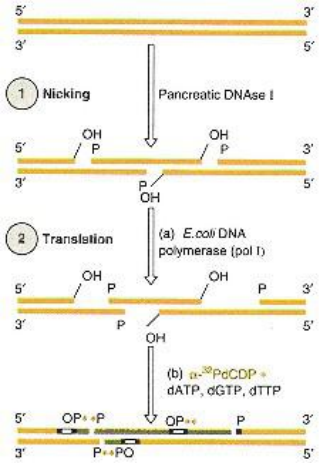


Figure 18.12
Nick translation to label DNA probes. Purified DNA molecules can be radioactively labeled and used to detect, by hybridization, the presence of complementary RNA or DNA in experimental samples. (1) Nicking step: introduces random single-stranded breaks in the DNA. (2) Translation step: (a) *E. coli* DNA polymerase (pol I) has 5 3 exonucleolytic activity that hydrolyzes nucleotides from the 5 end of the nick; (b) pol I simultaneously fills in the single-stranded gap with radioactively labeled nucleotides using the 3 end as a primer.

Labeled probes can be produced by **nick translation** of double-stranded DNA. Nick translation (Figure 18.12) involves the random enzymatic hydrolysis of a phosphodiester bond in the backbone of one strand of DNA by DNase I; the enzymatic breaks in the DNA backbone are referred to as nicks. A second enzyme, *E. coli* DNA polymerase I, with its 5 3 exonucleolytic activity and its DNA polymerase activity, creates single-strand gaps by hydrolyzing nucleotides from the 5 side of the nick and then filling in the gaps with its polymerase activity. The polymerase reaction is usually carried out in the presence of one α-^{32}P-labeled deoxynucleotide triphosphate and three unlabeled deoxynucleotide triphosphates. The DNA employed in this method is usually purified and is derived from cloned DNA, viral DNA, or cDNA.

Another method to label DNA probes, **random primer labeling of DNA,** has distinct advantages over the nick translation method. The random primer method typically requires only 25 ng of DNA as opposed to 1–2 g of DNA for nick translation and results in labeled probes with a specific activity ($>10^9$ cmp g^{-1}) approximately ten times higher. This method generally produces longer labeled DNA probes. The double-stranded probe is melted and hybridized with a mixture of random hexanucleotides containing all possible sequences (ACTCGG, ACTCGA, ACTCGC, etc.). The hybridized hexanucleotides serve as primers for DNA synthesis with a DNA polymerase, such as the Klenow enzyme, in the presence of one or more radioactively labeled deoxynucleoside triphosphates.

Labeled RNA probes have advantages over DNA probes. For one, relatively large amounts of RNA can be transcribed from a template, which may be available in very limited quantities. A double-stranded DNA (dsDNA) probe must be denatured prior to hybridization with the target DNA and rehybridization with itself competes for hybridization with the DNA of interest. No similar competition occurs with the single-stranded RNA probes that hybridize with complementary DNA or RNA molecules. Synthesis of an RNA probe requires DNA as a template. To be transcribed the template DNA must be covalently linked to an upstream promoter that can be recognized by a DNA-dependent RNA polymerase. Vectors have been constructed that are well suited for this technique.

A labeled DNA or RNA probe can be hybridized to nitrocellulose-bound nucleic acids and identified by the detection of the labeled probe. The nucleic

acids of interest can be transferred to nitrocellulose from bacterial colonies grown on agar or from agarose gels where the nucleic acid species have been electrophoretically separated by size.

Southern Blot Technique Is Useful for Identifying DNA Fragments

A technique to transfer DNA species, separated by agarose gel electrophoresis, to a filter for analysis was developed in the 1970s, and it is an indispensable tool. The method, developed by E. M. Southern, is referred to as the **Southern blot technique** (Figure 18.13). A DNA mixture of discrete restriction endonuclease-generated fragments from any source and complexity can be separated according to size by electrophoresis through an agarose gel. The DNA is dena-

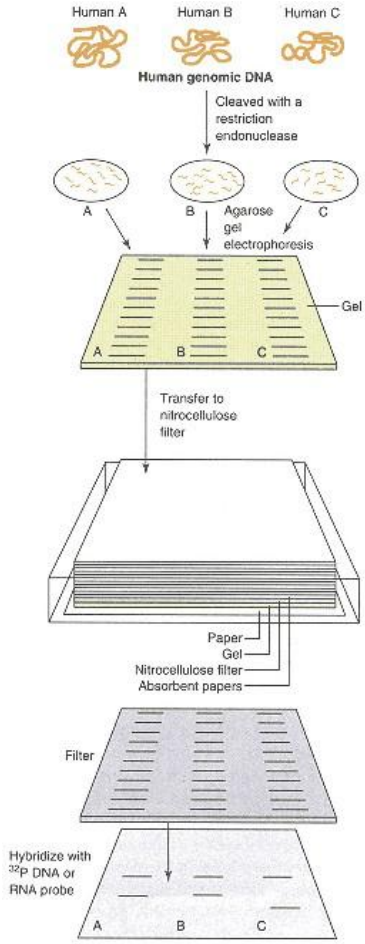


Figure 18.13
Southern blot to transfer DNA from agarose gels to nitrocellulose.
Transfer of DNA to nitrocellulose, as single-stranded molecules, allows for the detection of specific DNA sequences within a complex mixture of DNA. Hybridization with nick translated labeled probes can demonstrate if a DNA sequence of interest is present in the same or different regions of the genome.

tured by soaking the gel in alkali. The gel is then placed on absorbent paper and a nitrocellulose filter placed directly on top of the gel. Several layers of absorbent paper are placed on top of the nitrocellulose filter. The absorbent paper under the gel is kept wet with a concentrated salt solution that by capillary action is pulled up through the gel, the nitrocellulose, and into the absorbent paper layers above. The DNA is eluted from the gel by the upward movement of the high salt solution onto the nitrocellulose filter directly above, where it becomes bound. The position of the DNA bound to the nitrocellulose filter is exactly that which was present in the agarose gel. In its single-stranded membrane-bound form, the DNA can be analyzed with labeled probes.

The Southern blot technique is invaluable in analytical procedures for detection of the presence and determination of the number of copies of particular sequences in complex genomic DNA, confirming DNA cloning results, and demonstrating the polymorphic DNA arrangements of the human genome that correspond to pathological states. An example of the use of Southern blots is shown in Figure 18.13. Here whole human genomic DNA, isolated from three individuals, was digested with a restriction endonuclease, generating thousands of fragments. These fragments were distributed throughout the agarose gel according to size in an electric field. The DNA was transferred (blotted) to a nitrocellulose filter and hybridized with a ^{32}P-labeled DNA or RNA probe that represents a portion of a gene of interest. The probe detected two bands in all three individuals, indicating that the gene of interest is cleaved at one site within its sequence. Individuals A and B presented a normal pattern while patient C had one normal band and one lower molecular weight band. This is an example of altered DNA within different individuals of a single species, **restriction fragment length polymorphism (RFLP),** and implies a deletion in a segment of the gene that may be associated with a pathological state. The gene from this patient can be cloned, sequenced, and fully analyzed to characterize the altered nature of the DNA (see Clin. Corr. 18.5).

Other techniques that employ the principles of Southern blot are the transfer of RNA (Northern blots) and of proteins (Western blots) to nitrocellulose filters or nylon membranes.

Single-Strand Conformation Polymorphism

Southern blot analysis and detection of base changes in DNA from different individuals by RFLP analysis is dependent on alteration of a restriction endonuclease site. Often a base substitution, deletion, or insertion does not occur within a restriction endonuclease site. However, these modifications can readily be detected by **single-strand conformation polymorphism (SSCP).** This technique takes advantage of the fact that single-stranded DNA, smaller than 400 bases long, subjected to electrophoresis through a polyacrylamide gel migrates with a mobility partially dependent on its conformation. A single base alteration usually modifies the DNA conformation sufficiently to be detected as a mobility shift upon electrophoresis through a nondenaturing polyacrylamide gel. The analysis of a small region of genomic DNA or cDNA for SSCP can be accomplished by PCR amplification of the region of interest. Sense and antisense oligonucleotide primers are synthesized that flank the region of interest and this DNA is amplified by PCR in the presence of radiolabeled nucleotide(s). The resulting purified radiolabeled double-stranded PCR product is then heat denatured in 80% formamide and immediately loaded onto a nondenaturing polyacrylamide gel. The mobilities of control products are compared to samples derived from experimental/patient samples. Detection of mutations in patient samples can identify genetic lesions. This method depends on prior knowledge of the sequence of the gene/gene fragment of interest, while analysis by RFLP requires only restriction map analysis of DNA.

CLINICAL CORRELATION 18.5

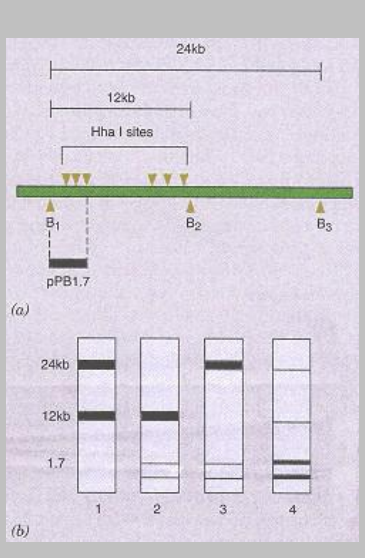
Restriction Fragment Length Polymorphisms Determine the Clonal Origin of Tumors

It is generally assumed that most tumors are monoclonal in origin; that is, a rare event alters a single somatic cell genome in such a fashion that the cells grow abnormally into a tumor mass with all-daughter cells carrying the identically altered genome. Proof that a tumor is of monoclonal origin versus polyclonal in origin can help to distinguish hyperplasia (increased production and growth of normal cells) from neoplasia (growth of new or tumor cells). The detection of restriction fragment length polymorphisms (RFLPs) of Southern blotted DNA samples allows one to define the clonal origin of human tumors. If tumor cells were collectively derived from different parental cells they should contain a mixture of DNA markers characteristic of each cell of origin. However, an identical DNA marker in all tumor cells would indicate a monoclonal origin. The analysis is limited to females where one can take advantage of the fact that each cell carries only one active X chromosome of either paternal or maternal origin with the second X chromosome being inactivated. Activation occurs randomly during embryogenesis and is faithfully maintained in all-daughter cells with one-half the cells carrying an activated maternal X chromosome and the other one-half an activated paternal X chromosome.

Analysis of the clonal nature of a human tumor depends on the fact that activation of an X chromosome involves changes in the methylation of selected cytosine (C) residues within the DNA molecule. Several restriction endonucleases, such as Hha I, which cleaves DNA at GCGC sites, will not cleave DNA at their recognition sequences if a C is methylated within this site. Therefore the methylated state (activated versus inactivated) of the X chromosome can be probed with restriction endonucleases. Furthermore, the paternal X chromosome can be distinguished from the maternal X chromosome in a significant number of individuals based on differences in the electrophoretic migration of restriction endonuclease generated fragments derived from selected regions of the chromosome. These DNA fragments are identified on a Southern blot by hybridization with a DNA probe that is complementary to this region of the X chromosome. An X-linked gene that is amenable to these studies is the hypoxanthine phosphoribosyltransferase (*HPRTase*) gene. The *HPRTase* gene consistently has two BamHI restriction endonuclease sites (B_1 and B_3 in figure), but in many individuals a third site (B_2) is also present (see figure).

The presence of site B_2 in only one parental X chromosome HPRT allows for the detection of restriction enzyme-generated polymorphisms. Therefore a female cell may carry one X chromosome with the *HPRT* gene possessing two BamHI sites (results in a single detectable DNA fragment of 24 kb) or three BamHI sites (results in a single detectable DNA fragment of 12 kb). This figure depicts the expected results for the analysis of tumor cell DNA to determine its monoclonal or polyclonal origin. As expected, three human tumors examined by this method were shown to be of monoclonal origin.

Vogelstein, B., Fearon, E. R., Hamilton, S. R., and Feinberg, A. B. Use of restriction fragment length polymorphism to determine the clonal origin of tumors. *Science 227*:642, 1985.



Analysis of genomic DNA to determine the clonal origin of tumors. (*a*) The X chromosome-linked *HPRTase* gene contains two invariant BamHI restriction endonuclease sites (B_1 and B_3) while in some individuals a third site, B_2, is also present. The *HPRTase* gene also contains several HhaI sites; however, all of these sites, except H1, are usually methylated in the active X chromosome. Therefore only the H1 site would be available for cleavage by HhaI in the active X chromosome. A cloned, labeled probe, pPB1.7, is employed to determine which form of the HPRTase gene is present in a tumor and if it is present on an active X chromosome. (*b*) Restriction endonuclease patterns predicted for monoclonal versus polyclonal tumors are as follows: (1) Cleaved with BamHI alone; 24-kb fragment derived from *HPRTase* gene containing only B_1 and B_3 sites and 12-kb fragment derived from *HPRTase* gene containing extra B_2 site. Pattern is characteristic for heterozygous individual. (2) Cleaved with BamHI plus HhaI, monoclonal tumor with the 12-kb fragment derived from an active X chromosome (methylated). (3) Cleaved with BamHI plus HhaI; monoclonal tumor with the 24-kb fragment derived from an active X chromosome (methylated). (4) Cleaved with BamHI plus HhaI; polyclonal tumor. All tumors studied displayed patterns as in Lane 2 or Lane 3.

18.8— Complementary DNA and Complementary DNA Libraries

The insertion of specific functional eukaryotic genes into vectors that can be expressed in a prokaryotic cell could produce large amounts of "genetically engineered" proteins with significant medical, agricultural, and experimental potential. Hormones and enzymes are currently produced by these methods, including insulin, erythropoietin, thrombopoietin, interleukins, interferons, and tissue plasminogen activator. Unfortunately, it is impossible, except in rare instances, to clone functional genes from genomic DNA. One reason for this is that most genes within the mammalian genome yield transcripts that contain introns that must be spliced out of the primary mRNA transcript. Prokaryotic systems cannot splice out the introns to yield functional mRNA transcripts. This problem can be circumvented by synthesizing **complementary DNA (cDNA)** from functional eukaryotic mRNA.

mRNA Is Used As a Template for DNA Synthesis Using Reverse Transcriptase

Messenger RNA can be reverse transcribed to cDNA and the cDNA inserted into a vector for amplification, identification, and expression. Mammalian cells normally contain 10,000–30,000 different species of mRNA molecules at any time during the cell cycle. In some cases, however, a specific mRNA species may approach 90% of the total mRNA, such as mRNA for globin in reticulocytes. Many mRNAs are normally present at only a few (1–14) copies per cell. A **cDNA library** can be constructed from the total cellular mRNA but if only a few copies per cell of mRNA of interest are present, the cDNA may be very difficult to identify. Methods that enrich the population of mRNAs or their corresponding cDNAs permit reduction of the number of different cDNA species within a cDNA library and greatly enhance the probability of identifying the clone of interest.

Desired mRNA in a Sample Can Be Enriched by Separation Techniques

Messenger RNA can be separated by size by gel electrophoresis or centrifugation. Utilization of mRNA in a specific molecular size range will enrich several-fold an mRNA of interest. Knowledge of the molecular weight of the protein encoded by the gene of interest gives a clue to the approximate size of the mRNA transcript or its cDNA; variability in the predicted size, however, will arise from differences in the length of the untranslated regions of the mRNAs.

Enrichment of a specific mRNA molecule can also be accomplished by immunological procedures but requires the availability of antibodies against the protein encoded by the gene of interest. Antibodies added to an *in vitro* protein synthesis mixture will react with the growing polypeptide chain associated with the polysome and precipitate it. The mRNA can be purified from the immunoprecipitated polysomal fraction.

Complementary DNA Synthesis

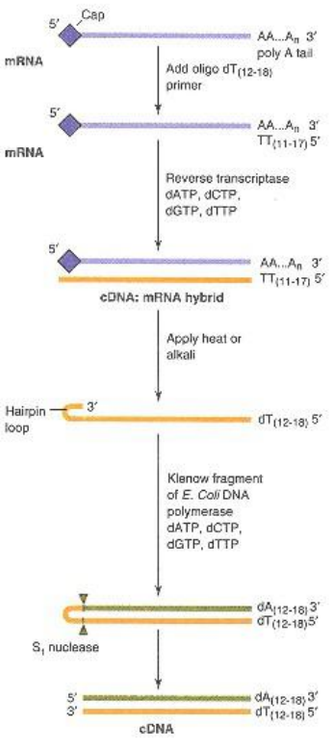


Figure 18.14
Synthesis of cDNA from mRNA.
The 3′ poly(A) tail of mRNA is hybridized with an oligomer of dT [oligo(dT)12–18] that serves as a primer for reverse transcriptase, which catalyzes the synthesis of the complementary DNA (cDNA) strand in the presence of all four deoxynucleotide triphosphates (dNTPs). The resulting cDNA–mRNA hybrid is separated into single-stranded cDNA by melting with heat or hydrolyzing the mRNA with alkali. The 3′ end of the cDNA molecule forms a hairpin loop that serves as a primer for the synthesis of the second DNA strand catalyzed by the Klenow fragment of *E. coli* DNA polymerase. The single-stranded unpaired DNA loop is hydrolyzed by S_1 nuclease to yield a double-stranded DNA molecule.

An isolated mRNA mixture is used as a template to synthesize a complementary strand of DNA using RNA-dependent DNA polymerase, reverse transcriptase (Figure 18.14). A primer is required for the reaction; advantage is taken of the poly(A) tail at the 3′ terminus of eukaryotic mRNA. An oligo(dT) with 12–18 bases is employed as the primer that will hybridize with the poly(A) sequence. After cDNA synthesis, the hybrid is denatured or the mRNA hydrolyzed in alkali in order to obtain the single-stranded cDNA. The 3′ termini of single-stranded cDNAs form a hairpin loop that serves as a primer for the synthesis of the second strand of the cDNA. Either the Klenow fragment or a reverse transcriptase can be used for this step. The resulting double-stranded cDNA contains a single-

stranded loop that is selectively recognized and digested by S1 nuclease. The ends of the cDNA must be modified prior to cloning in a vector. One method involves incubating blunt-ended cDNA molecules with linker molecules and a bacteriophage T4 DNA ligase that catalyzes the ligation of blunt-ended molecules (Figure 18.15). The synthetic linker molecules contain restriction endonuclease sites that can now be hydrolyzed with the appropriate enzyme for insertion of the cDNA into a compatibly cut vector.

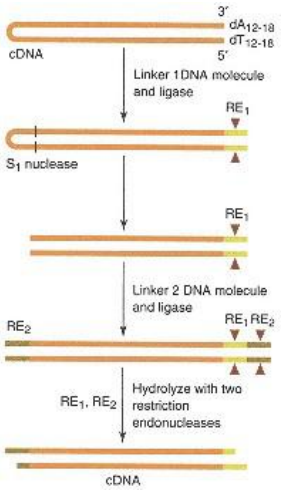


Figure 18.15
Modification of cDNA for cloning. The procedure begins with double-stranded DNA containing a hairpin loop. A linker DNA containing a restriction endonuclease site (RE_1) is added to the free end of the cDNA by blunt-end ligation. The single-stranded hairpin loop is next hydrolyzed with S_1 nuclease. A second linker with a different restriction endonuclease site within (RE_2) is blunt-end ligated to the newly created free cDNA. The second linker will probably bind to both ends but will not interfere with the first restriction endonuclease site. The modified DNA is hydrolyzed with the two restriction endonucleases and can be inserted into a plasmid or bacteriophage DNA by directional cloning.

Bacteriophage DNA (see p. 779) is the most convenient and efficient vector to create cDNA libraries because they can readily be amplified and stored indefinitely. Two bacteriophage vectors, **λgt10** and **λgt11,** and their newer constructs have been employed to produce cDNA libraries. The cDNA libraries in gt10 can be screened only with labeled nucleic acid probes, whereas those in gt11, an expression vector, can also be screened with antibody for the production of the protein or antigen of interest.

Total Cellular RNA May Be Used As a Template for DNA Synthesis Using RT-PCR

Alternative methods to construct cDNA libraries employ a **reverse transcriptase–PCR (RT-PCR) technique** and obviate the need to purify mRNA. One such strategy is depicted in Figure 18.16 and begins with the reverse transcriptase production of a DNA–mRNA hybrid. The method then adds a dG homopolymer tail to the 3 end catalyzed by terminal transferase and the subsequent hydrolysis of the mRNA. PCR primers are synthesized to hybridize with the dG, dA tails and terminate with two different restriction endonuclease sequences. The resulting PCR-amplified cDNA can then be hydrolyzed with the two different restriction endonucleases for directional cloning (see p. 765, Section 18.5) into an appropriate vector.

18.9— Bacteriophage, Cosmid, and Yeast Cloning Vectors

Detection of noncoding sequences in most eukaryotic genes and distant regulatory regions flanking the genes necessitated new cloning strategies to package larger DNA fragments than could be cloned in plasmids. Plasmids can accommodate foreign DNA inserts with a maximum length in the range of 5–10 kb (kilobases). Portions of recombinant DNA fragments larger than this are randomly deleted during replication of the plasmid within the bacterium. Thus alternate vectors have been developed.

Bacteriophage As Cloning Vectors

Bacteriophage λ (λ phage)—a virus that infects and replicates in bacteria—is an ideal vector for DNA inserts of approximately 15-kb lengths. The phage selectively infects bacteria and can replicate by either a lytic or nonlytic (lysogenic) pathway. The phage contains a self-complementary 12-base single-stranded tail (cohesive termini) at both ends of its 50-kb double-stranded DNA molecule. Upon infection of the bacteria the cohesive termini (cos sites) of a single phage DNA molecule self-anneal and the ends are covalently linked with the host cell DNA ligase. The circular DNA molecule serves as a template for transcription and replication. The phage, with restriction endonuclease-generated fragments representing a cell's whole genomic DNA inserted into it, is used to infect bacteria. Recombinant bacteriophages, released from the lysed cells, are collected and constitute a genomic library in phage. The phage library can be screened more rapidly than a plasmid library due to the increased size of the DNA inserts.

Numerous phage vectors have been constructed for different cloning strategies. For the sake of simplicity only a generic phage vector will be

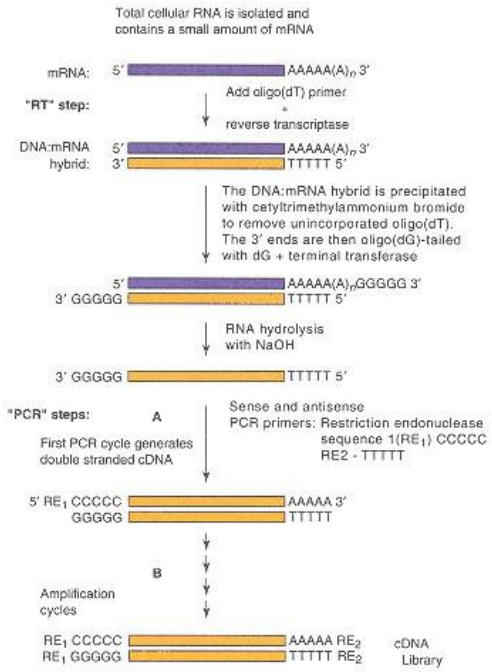


Figure 18.16
Generation of cDNA by reverse transcriptase–PCR (RT-PCR). Total cellular RNA or mRNA can be used to generate cDNA by RT-PCR. The mRNA with an oligo rA tail is reverse transcribed with an oligo dT primer. An oligo dG tail is added to the 3 ends of the RNA and DNA strands and the RNA strand is subsequently hydrolyzed with NaOH. Sense and antisense primers, modified with restriction site sequences, are then employed to amplify the cDNA by the PCR. The products can be hydrolyzed with the specific restriction endonucleases (RE_1 and RE_2) for cloning and subsequent studies.

described here. Cloning large fragments of DNA in phage takes advantage of the fact that a 15–25 kb segment of the phage DNA can be replaced without impairing its replication in *E. coli* (Figure 18.17). Packaging of phage DNA into the virus particle is constrained by its total length, which must be approximately 50 kb. The linear phage DNA can be digested with specific restriction endonucleases that generate small terminal fragments with their **cos sites** (arms), which are separated from the larger intervening fragments. Cellular genomic DNA is partially digested with the appropriate restriction enzymes to permit annealing and ligation with the phage arms. Genomic DNA is not enzymatically hydrolyzed completely in order to randomly generate fragments that can be properly packaged into phage particles. The DNA fragments that are smaller or larger than 15–25 kb can hybridize with the cos arms but are excluded from being packaged into infectious bacteriophage particles. All of the information required for phage infection and replication in bacteria is carried within the cos arms. The recombinant phage DNA is mixed with phage proteins *in vitro*, which assemble into infectious virions. The infectious recombinant phage particles are then propagated in an appropriate *E. coli* strain to yield a **λ phage library.** Many different *E. coli* strains have been genetically altered to sustain replication of specific recombinant virions.

Screening Bacteriophage Libraries

The bacteriophage library can be screened by plating the virus on a continuous layer of bacteria (a bacterial lawn) grown on agar plates (Figure 18.18). The individual phage will infect, replicate, and lyse one cell. The progeny virions will then infect and subsequently lyse bacteria immediately adjacent to the site of the first infected cell, creating a clear region or plaque in the opaque bacterial

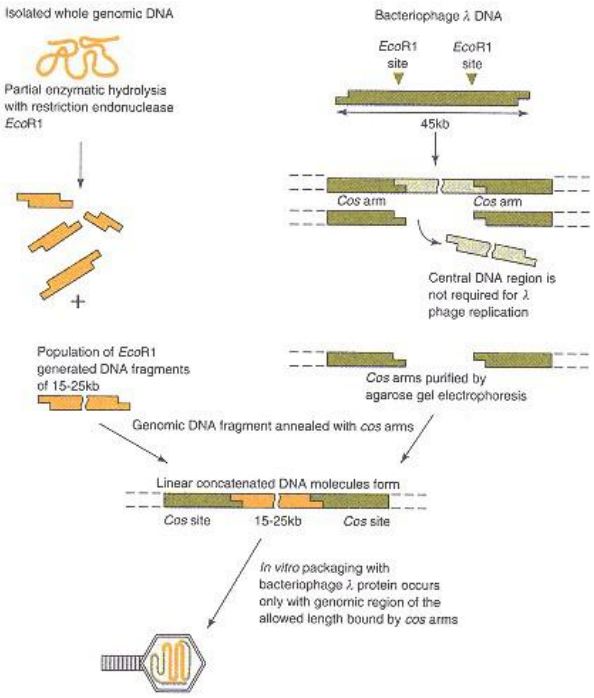


Figure 18.17
Cloning genomic DNA in bacteriophage λ.
Whole genomic DNA is incompletely digested with a restriction endonuclease (e.g., EcoRI). This results in DNA of random size fragments with single-stranded sticky ends. DNA fragments, cos arms, are generated with the same restriction endonuclease from bacteriophage DNA. The purified cos arm fragments carry sequence signals required for packaging DNA into a bacteriophage virion. The genomic fragments are mixed with the cos arms, annealed, and ligated, forming linear concatenated DNA arrays. The *in vitro* packaging with bacteriophage proteins occurs only with genomic DNA fragments of allowed lengths (15–25 kb) bounded by cos arms.

field. Phage, within each plaque, can be picked up on a nitrocellulose filter (as for replica plating) and the DNA fixed to the filter with NaOH. The location of cloned DNA fragments of interest is determined by hybridizing the filter-bound DNA with a labeled DNA or RNA probe followed by autoradiography. Bacteriophages in the plaque corresponding to the labeled filter-bound hybrid are picked up and amplified in bacteria for further analysis. Complementary DNA libraries in bacteriophage are also constructed that contain the phage cos arms. If the cDNA is recombined with phage DNA that permits expression of the gene, such as gt11, then plaques can be screened immunologically with antibodies specific for the antigen of interest.

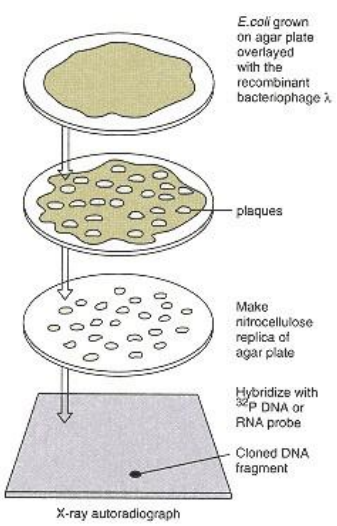


Figure 18.18
Screening genomic libraries in bacteriophage λ.
Competent *E. coli* are grown to confluence on an agar plate and then overlayed with the recombinant bacteriophage. Plaques develop where bacteria are infected and subsequently lysed by the phage . Replicas of the plate can be made by touching the plate with a nitrocellulose filter. The DNA is denatured and fixed to the nitrocellulose with NaOH. The fixed DNA is hybridized with a ^{32}P-labeled probe and exposed to X-ray film. The autoradiograph identifies the plaque(s) with recombinant DNA of interest.

Cloning DNA Fragments into Cosmid and Yeast Artificial Chromosome Vectors

Even though phage are the most commonly used vectors to construct genomic DNA libraries, the lengths of many genes exceed the maximum size of the

DNA that can be inserted between the phage arms. A **cosmid vector** can accommodate foreign DNA inserts of approximately 45 kb. **Yeast artificial chromosomes (YACs)** have been developed to clone DNA fragments of 200–500 kb lengths. While cosmid and yeast artificial chromosome vectors are difficult to work with, their libraries permit the cloning of large genes with their flanking regulatory sequences, as well as families of genes or contiguous genes.

Cosmid vectors are a cross between plasmid and bacteriophage vectors. Cosmids contain an antibiotic-resistance gene for selection of recombinant DNA molecules, an origin of replication for propagation in bacteria, and a cos site for packaging of recombinant molecules in bacteriophage particles. The bacteriophage with recombinant cosmid DNA can infect *E. coli* and inject its DNA into the cell. Cosmid vectors contain only approximately 5 kb of the 50-kb bacteriophage DNA and therefore cannot direct replication and assembly of new infectious phage particles. Instead, the recombinant cosmid DNA circularizes and replicates as a large plasmid. Bacterial colonies with recombinants of interest can be selected and amplified by methods similar to those described for plasmids.

Standard cloning procedures and some novel methods are employed to construct YACs. Very large foreign DNA fragments are joined to yeast DNA sequences, one that functions as a telomere (distal extremity of chromosome arm) and another that functions as a centromere and as an origin of replication. The recombinant YAC DNA is introduced into the yeast by transformation. The YAC constructs are designed so that yeast transformed with recombinant chromosomes grow as visually distinguishable colonies. This facilitates selection and analysis of cloned DNA fragments.

18.10— Techniques to Further Analyze Long Stretches of DNA

Subcloning Permits Definition of Large Segments of DNA

Complete analysis of functional elements in a cloned DNA fragment requires sequencing of the entire molecule. Current techniques can sequence 200–400 bases in a DNA fragment, yet cloned DNA inserts are frequently much larger. Restriction maps of the initial DNA clone are essential for cleaving the DNA into smaller pieces to be recloned, or **subcloned** for further analysis. The sequences of each of the small subcloned DNA fragments can be determined. Overlapping regions of the subcloned DNA properly align and confirm the entire sequence of the original DNA clone.

Sequencing can often be accomplished without subcloning. Antisense primers can be synthesized that are complementary to the initially sequenced 3 ends of the cloned DNA. This process is repeated until the full length of the cloned DNA has been sequenced. This method obviates the need to prepare subclones but it requires synthesis/purchase of numerous primers. On the other hand, the subcloned DNA is always inserted back into the same region of the plasmid. Therefore one set of primers complementary to the plasmid DNA sequences flanking the inserted DNA can be used for all of the sequencing reactions with subcloned DNA.

Chromosome Walking Is a Technique to Define Gene Arrangement in Long Stretches of DNA

Knowledge of how genes and their regulatory elements are arranged in a chromosome should lead to an understanding of how sets of genes may be coordinately regulated. Currently, it is difficult to clone DNA fragments large enough to identify contiguous genes. The combination of several techniques allows for the analysis of very long stretches of DNA (50–100 kb). The method, **chromosome walking,** is possible because phage or cosmid libraries contain

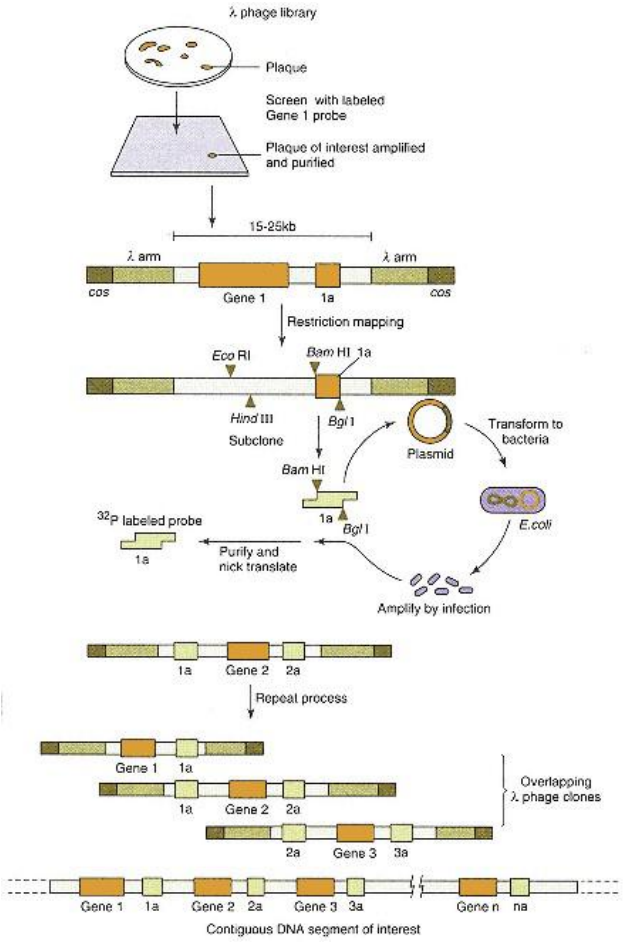


Figure 18.19
Chromosome walking to analyze contiguous DNA segments in a genome. Initially, a DNA fragment is labeled by nick translation to screen a library for recombinant phage carrying a gene of interest. The amplified DNA is mapped with a battery of restriction endonucleases to select a new region (1a) within the original cloned DNA that can be recloned (subcloned). The subcloned DNA (1a) is used to identify other DNA fragments within the original library that would overlap the initially amplified DNA region. The process can be repeated many times to identify contiguous DNA regions upstream and downstream of the initial DNA (gene 1) of interest.

partially cleaved genomic DNA cut at specific restriction endonuclease sites. The cloned fragments will contain overlapping sequences with other cloned fragments. Overlapping regions are identified by restriction mapping, subcloning, screening phage or cosmid libraries, and sequencing procedures.

The overall procedure of chromosome walking is shown in Figure 18.19. Initially the phage library is screened for a sequence of interest with a DNA or RNA probe. The cloned DNA is restriction mapped and a small segment is subcloned in a plasmid, amplified, purified, and labeled by nick translation. This labeled probe is then used to rescreen the phage library for complementary sequences, which are then cloned. The newly identified overlapping cloned DNA is then treated in the same fashion as the initial DNA clone to search for other overlapping sequences. Caution must be taken that the subcloned DNA does not contain a sequence common to the large numbers of repeating DNA

sequences in higher eukaryotic genomes. If a subcloned DNA probe contains a repeat sequence it hybridizes to numerous bacteriophage plaques and prevents the identification of a specific overlapping clone.

18.11— Expression Vectors and Fusion Proteins

Recombinant DNA methodology described to this point has dealt primarily with screening, amplification, and purification of cloned DNA species. An important goal of recombinant DNA studies, as stated earlier, is to have a foreign gene expressed in bacteria with the product in a biologically active form. Sequencing the DNA of many bacterial genes and their flanking regions has identified the spatial arrangement of regulatory sequences required for expression of genes. A promoter and other regulatory elements upstream of the gene are required to transcribe a gene (Chapter 19, Section 19.3). mRNA transcript of a recombinant eukaryotic gene, however, is not translated in a bacterial system because it lacks the bacterial recognition sequence, the Shine–Dalgarno sequence, required to properly orient it with a functional bacterial ribosome. Vectors that facilitate the functional transcription of DNA inserts, termed **expression vectors,** have been constructed such that a foreign gene can be inserted into the vector downstream of a regulated promoter but within a bacterial gene, commonly the *lacZ* gene. The mRNA transcript of the recombinant DNA contains the *lacZ* Shine–Dalgarno sequence, codons for a portion of the 3 end of the *lacZ* gene protein, followed by the codons of the complete foreign gene of interest. The protein product is a **fusion protein** that contains a few N-terminal amino acids of the *lacZ* gene protein and the complete amino acid sequence of the foreign gene product.

Foreign Genes Can Be Expressed in Bacteria Allowing Synthesis of Their Encoded Proteins

Many plasmid and bacteriophage vectors have been constructed to permit expression of eukaryotic genes in bacterial cells. Rapidly replicating bacteria can serve as a biological factory to produce large amounts of specific proteins, which have research, clinical, and commercial value. As an example, human protein hormones are produced by recombinant technologies, which serve as replacement or supplemental hormones in patients with aberrant or missing hormone production. Figure 18.20 depicts a generalized plasmid vector for the expression of a mammalian gene. Recall that the inserted foreign gene must be in the form of cDNA from its corresponding mRNA since the bacterial system cannot remove the introns in the pre-mRNA transcript. The DNA must be inserted in register with the codons of the 3 -terminal codons of the bacterial protein when creating a fusion protein. That is, insertion must occur after a triplet codon of the bacterial protein and at the beginning of a triplet codon of the eukaryotic gene protein to ensure proper translation. Finally, the foreign gene must be inserted in the proper orientation relative to the promoter to yield a functional transcript. This can be achieved by directional cloning.

Eukaryotic proteins synthesized within bacteria are often unstable and are degraded by intracellular proteases. Fusion protein products, however, are usually stable. The fusion protein amino acids encoded by the prokaryotic genome may be cleaved from the purified protein of interest by enzymatic or chemical procedures. An alternative cloning strategy to circumvent the intracellular instability of some proteins is to produce a foreign protein that is secreted. This requires cloning the foreign gene in a vector such that the fusion protein synthesized contains a signal peptide that can be recognized by the bacterial signal peptidase that properly processes the protein for secretion.

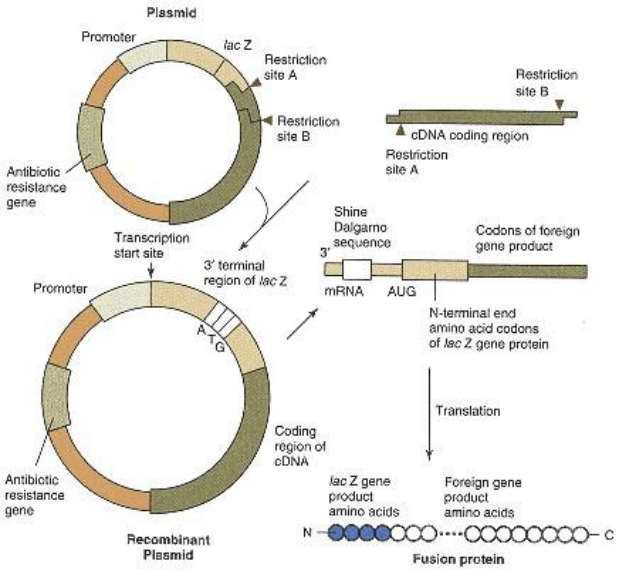


Figure 18.20
Construction of a bacterial expression vector.
A cDNA coding region of a protein of interest is inserted downstream of bacterial regulatory sequences (promoter, P) for the *lacZ* gene, the coding sequence for the mRNA Shine–Dalgarno sequence, the AUG codon, and a few codons for the N-terminal amino acids of the *lacZ* gene protein. The mRNA produced from this expression vector will therefore direct synthesis of a foreign protein in the bacterium with a few of its N-terminal amino acids of bacterial protein origin (a fusion protein).

18.12— Expression Vectors in Eukaryotic Cells

Mammalian genetic diseases result from missing or defective intracellular proteins. To utilize recombinant techniques to treat these diseases, vectors have to be constructed that can be incorporated into mammalian cells. In addition, these vectors have to be selective for the tissue or cells containing the aberrant protein. Numerous vectors permit the expression of foreign DNA genes in mammalian cells grown in tissue culture. These vectors have been used extensively for elucidation of the posttranslational processing and synthesis of proteins in cultured eukaryotic cells. Unfortunately, the goal to selectively express genes in specific tissues or at specific developmental stages within an animal has met with very limited success.

Several types of expression vectors have been developed that allow the replication, transcription, and translation of foreign genes in eukaryotic cells grown *in vitro,* including both RNA and DNA viral vectors that contain a foreign DNA insert. These viral vectors are able to infect and then replicate in a host cell. Experimentally constructed vectors that contain essential DNA elements, usually derived from a viral genome, permit expression of foreign gene inserts. **Shuttle vectors** contain both bacterial and eukaryotic replication signals, thus permitting replication of the vector in both bacteria and mammalian cells. A shuttle vector allows a gene to be cloned and purified in large quantities from a bacterial system and then the same recombinant vector can be expressed in a mammalian cell. Some expression vectors become integrated into the host cell genome while others remain as extrachromosomal entities (episomes) with stable expression of their recombinant gene in the daughter cells. Other expression vectors remain as episomal DNA, permitting only transient expression of their foreign gene prior to cell death.

Foreign DNA, such as viral expression vectors, may be introduced into the cultured eukaryotic cells by **transfection,** a process that is analogous to transformation of DNA into bacterial cells. The most commonly employed

transfection methods involve the formation of a complex of DNA with calcium phosphate or diethylaminoethyl (DEAE)-dextran, which is then taken up by the cell by endocytosis. The DNA is subsequently transferred from the cytoplasm to the nucleus, where it is replicated and expressed. The details of the mechanism of transfection are not known. Both methods are employed to establish transiently expressed vectors while the calcium phosphate procedure is also used for permanently expressed foreign genes. Typically, 10–20% of the cells in culture can be transfected by these procedures.

DNA Elements Required for Expression of Vectors in Mammalian Cells

Expression of recombinant genes in mammalian cells requires the presence of DNA-controlling elements within the vector that are not necessary in the bacterial system. To be expressed in a eukaryotic cell the cloned gene is inserted in the vector in the proper orientation relative to control elements, including a promoter, polyadenylation signals, and an enhancer sequence. Expression may be improved by the inclusion of an intron. Some or all of these DNA elements may be present in the recombinant gene if whole genomic DNA is used for cloning. A particular cloned fragment generated by restriction endonuclease cleavage, however, may not contain the required controlling elements. A cDNA would not possess these required DNA elements. It is therefore necessary that the expression vector to be used in mammalian cells be constructed such that it contains all of the required controlling elements.

An expression vector can be constructed by insertion of required DNA-controlling elements into the vector by recombinant technologies. Enhancer and promoter elements, engineered into an expression vector, should be recognized by a broad spectrum of cells in culture for the greatest applicability of the vector. Controlling elements derived from viruses with a broad host range are used for this purpose and are usually derived from the **papovavirus, simian virus 40 (SV40), Rous sarcoma virus,** or the **human cytomegalovirus.**

The vector must replicate so as to increase the number of copies within each cell or to maintain copies in daughter cells. The vector therefore is constructed to contain DNA sequences that promote its replication in the eukaryotic cell. This DNA region is usually derived from a virus and is referred to as the origin of replication (Ori). Specific protein factors, encoded by genes engineered into the vector or previously introduced into the host genome, recognize and interact with the ori sequences to initiate DNA replication.

Transfected Eukaryotic Cells Can Be Selected by Utilizing Mutant Cells That Require Specific Nutrients

It is important to have a means of selectively growing the transfected cells since they often represent only 10–20% of the cell population. As was the case for the bacterial plasmid, a gene can be incorporated into the vector that encodes an enzyme that confers resistance to a drug or confers selective growth capability to the cells carrying the vector. Constructing vectors that express both a selectable marker and a foreign gene is difficult. **Cotransfection** circumvents this problem. Two different vectors are efficiently taken up by those cells capable of being transfected. In most cases greater than 90% of transfected cells carry both vectors, one with the selectable marker and the second carrying the gene of interest.

Two of the more commonly employed selectable markers are the thymidine kinase (*tk*) and the dihydrofolate reductase gene. The *tk* gene product, thymidine kinase, is expressed in most mammalian cells and participates in the salvage pathway for thymidine. Several mutant cell lines have been isolated that lack a functional thymidine kinase gene (tk^-) and in growth medium containing hypoxanthine, aminopterin, and thymidine these cells will not survive. Only

those tk^- mutant cells cotransfected with a vector carrying a *tk* gene, usually of herpes simplex virus origin, will grow in the medium. In most instances, these cells have been cotransfected with the gene of interest.

The dihydrofolate reductase gene (*dhfr*) is required to maintain cellular concentrations of tetrahydrofolate for nucleotide biosynthesis (see Chapter 13). Cells lacking this enzyme will only survive in media containing thymidine, glycine, and purines. Mutant cells ($dhfr^-$), which are transfected with the *dhfr* gene, can therefore be selectively grown in a medium lacking these supplements. Expressing foreign genes in mutant cells, cotransfected with selectable markers, is limited to cell types that can be isolated with the required gene defect. Normal cells, however, transfected with a vector carrying the *dhfr* gene, are also resistant to methotrexate, an inhibitor of dihydrofolate reductase, and these cells can be selected for by growth in methotrexate.

Another approach for selecting nonmutated cells involves the use of a bacterial gene coding for aminoglycoside 3 -phosphotransferase (APH) for co-transfection. Cells expressing APH are resistant to aminoglycoside antibiotics such as neomycin and kanamycin, which inhibits protein synthesis in both prokaryotes and eukaryotes. Vectors carrying an *APH* gene can therefore be used as a selectable marker in both bacterial and mammalian cells.

Foreign Genes Can Be Expressed in Eukaryotic Cells by Utilizing Virus Transformed Cells

Figure 18.21 depicts the transient expression of a transfected gene in COS cells, a commonly used system to express foreign eukaryotic genes. The COS cells are permanently cultured simian cells, transformed with an origin-defective SV40 genome. The defective viral genome has integrated into the host cell genome and constantly expresses viral proteins. Infectious viruses, which are normally lytic to infected cells, are not produced because the viral origin of replication is defective. The SV40 proteins expressed by the transformed COS cell will recognize and interact with a normal SV40 ori carried in a vector transfected into these cells. These SV40 proteins will therefore promote the repeated replication of the vector. A transfected vector containing both an SV40 ori and a gene of interest may reach a copy number in excess of 10^5 molecules/cell. Transfected COS cells die after 3–4 days, possibly due to a toxic overload of the episomal vector DNA.

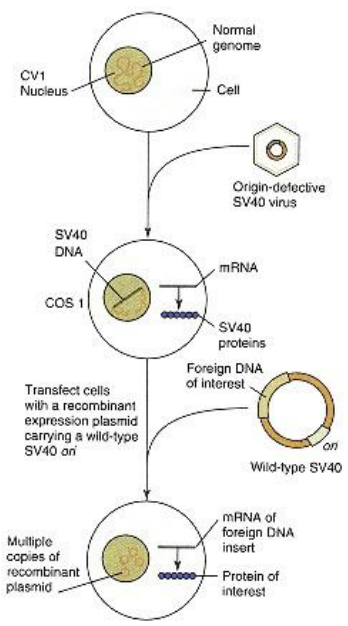


Figure 18.21
Expression of foreign genes in the eukaryotic COS cell.
CV1, an established tissue culture cell line of simian origin, can be infected and supports the lytic replication of the simian DNA virus, SV40. Cells are infected with an origin (ori)-defective mutant of SV40 whose DNA permanently integrates into the host CV1 cell genome. The defective viral DNA continuously codes for proteins that can associate with a normal SV40 ori to regulate replication. Due to its defective ori, the integrated viral DNA will not produce viruses. The SV40 proteins synthesized in the permanently altered CV1 cell line, COS-1, can, however, induce the replication of recombinant plasmids carrying a wild-type SV40 ori to a high copy number (as high as 105 molecules per cell). The foreign protein synthesized in the transfected cells may be detected immunologically or enzymatically.

18.13— Site-Directed Mutagenesis

By mutating selected regions or single nucleotides within cloned DNA, it is possible to define the role of DNA sequences in gene regulation and amino acid sequences in protein function. **Site-directed mutagenesis** is the controlled alteration of selected regions of a DNA molecule. It may involve the insertion or deletion of selected DNA sequences or the replacement of a specific nucleotide with a different base. A variety of chemical methods mutate DNA *in vitro* and *in vivo* usually at random sites within the molecule.

Role of Flanking Regions in DNA Can Be Evaluated by Deletion and Insertion Mutations

Site-directed mutagenesis can be carried out in various regions of a DNA sequence including the gene itself or the flanking regions. Figure 18.22 depicts a simple deletion mutation strategy where the sequence of interest is selectively cleaved with restriction endonuclease, the specific sequence removed, and the altered recombinant vector recircularized with DNA ligase. The role of the

deleted sequence can be determined by comparing the level of expression (translation) of the gene product, measured immunologically or enzymatically, to the unaltered recombinant expression vector. A similar technique is used to insert new sequences at the site of cleavage. Deletion of a DNA sequence within the flanking region of a cloned gene can help to define its regulatory role in gene expression. The presence or absence of a regulatory sequence may not be sufficient to evaluate its role in controlling expression. The spatial arrangement of regulatory elements to one another, to the gene, and to its promoter may be important in the regulation of gene expression (see Chapter 19).

Analysis of potential regulatory sequences is conveniently conducted by inserting the sequence of interest upstream of a reporter gene in an expression vector. A **reporter gene,** usually of prokaryotic origin, encodes for a gene product that can readily be distinguished from proteins normally present in the nontransfected cell and for which there is a convenient and rapid assay. A commonly used reporter gene is the chloramphenicol acetyltransferase (*CAT*) gene of bacteria. The gene product catalyzes the acetylation and inactivation of chloramphenicol, a protein synthesis inhibitor of prokaryotic cells. The ability of a regulatory element to enhance or suppress expression of the *CAT* gene can be determined by assaying the level of acetylation of chloramphenicol in extracts prepared from transfected cells. The regulatory element can be mutated prior to insertion into the vector carrying the reporter gene to determine its spatial and sequence requirements as a regulator of gene expression.

A difficulty encountered in analysis of regulatory elements is the lack of restriction endonuclease sites at useful positions within the cloned DNA. **Deletion mutations** can be made, in the absence of appropriately positioned restriction endonuclease sites, by linearizing cloned DNA with a restriction endonuclease downstream of the potential regulatory sequence of interest. The DNA can then be systematically truncated with an exonuclease, which hydrolyzes nucleotides from the free end of both strands of the linearized DNA. Increasing times of digestion generates smaller DNA fragments. Figure 18.23 demonstrates how larger deletion mutations (yielding smaller fragments) can be tested for functional activity. The enzymatic hydrolysis of the double strand of DNA occurs at both ends of the linearized recombinant vector, destroying the original restriction endonuclease site (RE_2). A unique restriction endonuclease site is reestablished to recircularize the truncated DNA molecule for further manipulations to evaluate the function of the deleted sequence. This is accomplished by ligating the blunt ends with a linker DNA, a synthetic oligonucleotide containing one or more restriction endonuclease sites. The ligated linkers are cut with the appropriate enzyme permitting recircularization and ligation of the DNA.

Site-Directed Mutagenesis of a Single Nucleotide

The previously discussed procedures can elucidate the functional role of small to large DNA sequences. Frequently, however, one wants to evaluate the role of a single nucleotide at selected sites within the DNA molecule. A single base change permits evaluation of the role of specific amino acids in a protein (see Clin. Corr. 18.6). This method also allows one to create or destroy a restriction endonuclease site at specific locations within a DNA sequence. The site-directed mutagenesis of a specific nucleotide is a multistep process that begins with cloning the normal type gene in a bacteriophage (Figure 18.24). The M13 series of recombinant bacteriophage vectors are commonly employed for these studies. M13 is a filamentous bacteriophage that specifically infects male *E. coli* that express sex pili encoded for by a plasmid (F factor). M13 bacteriophage contains DNA in a single-stranded or replicative form, which is replicated to double-stranded DNA within an infected cell. The double-stranded form of the

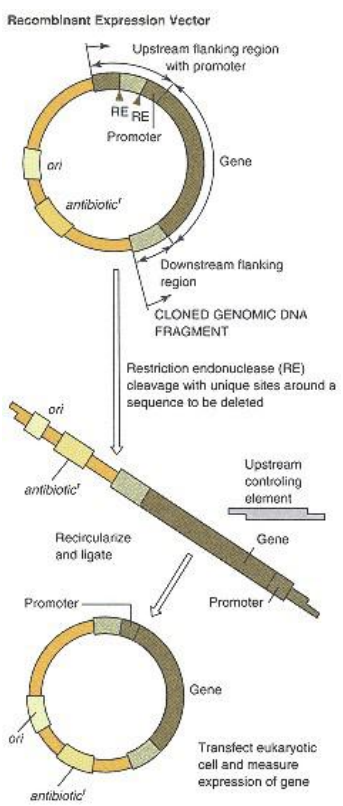


Figure 18.22
Use of expression vectors to study DNA regulatory sequences.
The gene of interest along with upstream and/or downstream DNA flanking regions is inserted and cloned in an expression vector and the baseline expression of the gene in an appropriate cell is determined. Defined regions of potential regulatory sequences can be removed by restriction endonuclease cleavage and the truncated recombinant DNA vector can be recircularized, ligated, and transfected into an appropriate host cell. The level of gene expression in the absence of the potential regulator is determined and compared to controls to ascertain the regulatory role of the deleted flanking DNA sequence.

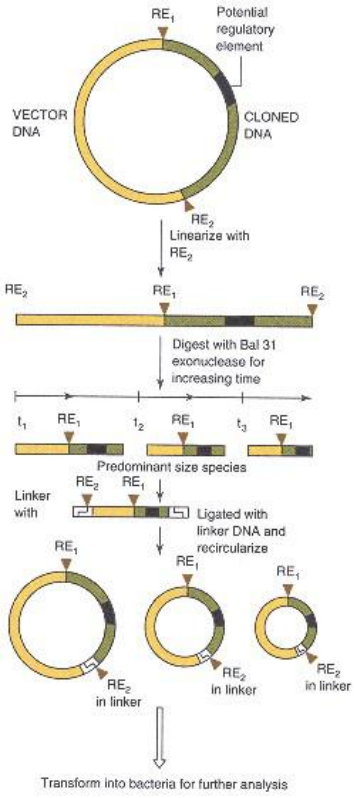


Figure 18.23
Enzymatic modification of potential DNA regulatory sequences.
A purified recombinant DNA molecule with a suspected gene regulatory element within flanking DNA regions is cleaved with a restriction endonuclease (RE_2). The linearized recombinant DNA is digested for varying time periods with the exonuclease, Bal31, reducing the size of the DNA flanking the potential regulatory element. The resulting recombinant DNA molecules of varying reduced sizes have small DNA oligomers (linkers) containing a restriction endonuclease sequence for RE_2 ligated to their ends. The linker-modified DNA is hydrolyzed with RE_2, creating complementary single-stranded sticky ends that permit recircularization of recombinant vectors. The potential regulatory element, bounded by various reduced-sized flanking DNA sequences, can be amplified, purified, sequenced, and inserted upstream of a competent gene in an expression vector. Modification of expression of the gene in an appropriate transfected cell can then be monitored to evaluate the role of the potential regulatory element placed at varying distances from the gene.

DNA is isolated from infected cells and used for cloning the gene to be mutated. The plaques of interest can be visually identified by α-complementation (see p. 772).

The M13 carrying the cloned gene of interest is used to infect susceptible *E. coli*. The progeny bacteriophages are released into the growth medium and contain single-stranded DNA. An oligonucleotide (18–30 nucleotides long) is synthesized that is complementary to a region of interest except for the nucleotide to be mutated. This oligomer, with one mismatched base, will hybridize to the single-stranded gene cloned in the M13 DNA and serves as a primer. Primer extension is accomplished with the bacteriophage T4 DNA polymerase and the resulting double-stranded DNA can be transformed into susceptible *E. coli*, where the mutated DNA strand serves as a template to replicate new (+) strands now carrying the mutated nucleotide.

The bacteriophage plaques, containing the mutated DNA, are screened by hybridizing with a labeled probe of the original oligonucleotide. By adjusting the wash temperature of the hybridized probe only the perfectly matched hybrid will remain complexed while the wild-type DNA–oligomer with mismatched nucleotide will dissociate. The M13 carrying the mutated gene is then replicated in bacteria, the DNA purified, and the mutated region of the gene sequenced

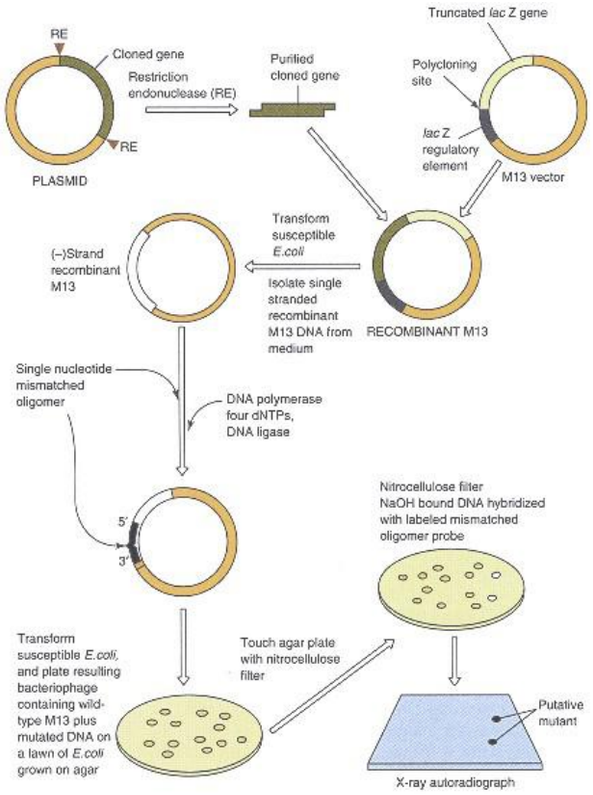


Figure 18.24
Site-directed mutagenesis of a single nucleotide and detection of the mutated DNA. The figure is a simplified overview of the method. This process involves the insertion of an amplified pure DNA fragment into a modified bacteriophage vector, M13. Susceptible *E. coli*, transformed with the recombinant M13 DNA, synthesize the (+) strand DNA packaged within the viron bacteriophage proteins. The bacteriophages are isolated from the growth medium and the single-stranded recombinant M13 DNA is purified. The recombinant M13 DNA serves as a template for DNA replication in the presence of DNA polymerase, deoxynucleoside triphosphates (dNTPs), DNA ligase, and a special primer The DNA primer (mismatched oligomer) is synthesized to be exactly complementary to a region of the DNA (gene) of interest except for the one base intended to be altered (mutated). The newly synthesized M13 DNA therefore contains a specifically mutated base, which when reintroduced into susceptible *E. coli* will be faithfully replicated. The transformed *E. coli* are grown on agar plates with replicas of the resulting colonies picked up on a nitrocellulose ilter. DNA associated with each colony is denatured and fixed to the filter with NaOH and the filter-bound DNA is hybridized with a ^{32}P-labeled mismatched DNA oligomer probe. The putative mutants are then identified by exposing the filter to X-ray film.

CLINICAL CORRELATION 18.6

Site-Directed Mutagenesis of HSV I gD

The structural and functional roles of a carbohydrate moiety covalently linked to a protein can be studied by site-directed mutagenesis. The gene that codes for a glycoprotein whose asparagine residue(s) is normally glycosylated (N-linked) must first be cloned. The herpes simplex virus type I (HSV I) glycoprotein D (gD) may contain as many as three N-linked carbohydrate groups. The envelope bound HSV I gD appears to play a central role in virus absorption and penetration. Carbohydrate groups may play a role in these processes.

The cloned *HSV I gD* gene has been modified by site-directed mutagenesis to alter codons for the asparagine residue at the three potential glycosylation sites. These mutated genes, cloned within an expression vector, were transfected into eukaryotic cells (COS-1), where the gD protein was transiently expressed. The mutated HSV I gD, lacking one or all of its normal carbohydrate groups, can be analyzed with a variety of available monoclonal anti-gD antibodies to determine if immunological epitopes (specific sites on a protein recognized by an antibody) have been altered. Altered epitopes would indicate that the missing carbohydrate moiety is directly associated with the normal recognition site or played a role in the protein's native conformation. An altered protein conformation can impact on immunogenicity (e.g., for vaccines) and protein processing (movement of the protein from the endoplasmic reticulum, where it is synthesized, to the membrane, where it is normally bound). Mutations at two of the glycosylation sites altered the native conformation of the protein such that it was less reactive with selected monoclonal antibodies. Alteration at a third site had no apparent effect on protein structure, and loss of the carbohydrate chain at all three sites did not prevent normal processing of the protein.

Sodora, D. L., Cohen, G. H., and Eisenberg, R. J. Influence of asparagine-linked oligosaccharides on antigenicity, processing, and cell surface expression of herpes simplex virus type I glycoprotein D. *J. Virol.* 63:5184, 1989.

to confirm the identity of the mutation. Many modifications have been developed to improve the efficiency of site-directed mutagenesis of a single nucleotide including a method to selectively replicate the mutated strand. M13 bacteriophage, replicated in a mutant *E. coli*, incorporates some uracil residues into its DNA in place of thymine due to a metabolic defect in the synthesis of dTTP from dUTP and the lack of an enzyme that normally removes uracil residues from DNA. The purified single-stranded M13 uracil-containing DNA is hybridized with a complementary oligomer containing a mismatched base at the nucleotide to be mutated. The oligomer serves as the primer for DNA replication *in vitro* with the template (+) strand containing uracils and the new (–) strand containing thymines. When this double-stranded M13 DNA is transformed into a wild-type *E. coli*, the uracil-containing strand is destroyed and the mutated (–) strand serves as the template for the progeny bacteriophages, most of which will carry the mutation of interest.

The polymerase chain reaction can also be employed for site-directed mutagenesis. Strategies have readily been developed to incorporate a mismatched base into one of the oligonucleotides that primes the PCR. Some of these procedures employ M13 bacteriophage and follow the principles described in Figure 18.24. A variation of these PCR methods, **inverse PCR mutagenesis,** has been applied to small recombinant plasmids (4–5 kb) (Figure 18.25). The method is very rapid with 50–100% of the generated colonies containing the mutant sequence. The two primers are synthesized so that they anneal back-to-back with one primer carrying the mismatched base.

18.14— Applications of Recombinant DNA Technologies

The practical uses of recombinant DNA methods in biological systems are limited only by one's imagination. Recombinant DNA methods are applicable to numerous biological disciplines including agriculture, studies of evolution, forensic biology, and clinical medicine. Genetic engineering can introduce new or altered proteins into crops (e.g., corn), so that they contain amino acids essential to humans but often lacking in plant proteins. Toxins that are lethal to specific insects but harmless to humans can be introduced into crops to protect plants without the use of environmentally destructive pesticides. The DNA isolated from cells in the amniotic fluid of a pregnant woman can be analyzed for the presence or absence of genetic defects in the fetus. Minuscule quantities of DNA can be isolated from biological samples that have been preserved in ancient tar pits or frozen tundra and can be amplified and sequenced for evolutionary studies at the molecular level. The DNA from a single hair, a drop of blood, or sperm from a rape victim can be isolated, amplified, and mapped to aid in identifying felons. Current technologies in conjunction with future invented methods should permit the selective introduction of genes into cells with defective or absent genes. Developing methodologies are also likely to become available to introduce nucleic acid sequences into cells to selectively turn off the expression of detrimental genes.

Antisense Nucleic Acids Hold Promise As Research Tools and in therapy

Recently, a new tool, **antisense nucleic acids,** has been introduced to study the intracellular expression and function of specific proteins. Natural and synthetic antisense nucleic acids that are complementary to mRNAs will hybridize within the cell, inactivate the mRNA, and block translation. The introduction of antisense nucleic acids into cells has opened new avenues to explore how proteins, whose expression has been selectively repressed in a cell, function within that cell. This method also holds great promise in control of diseased processes

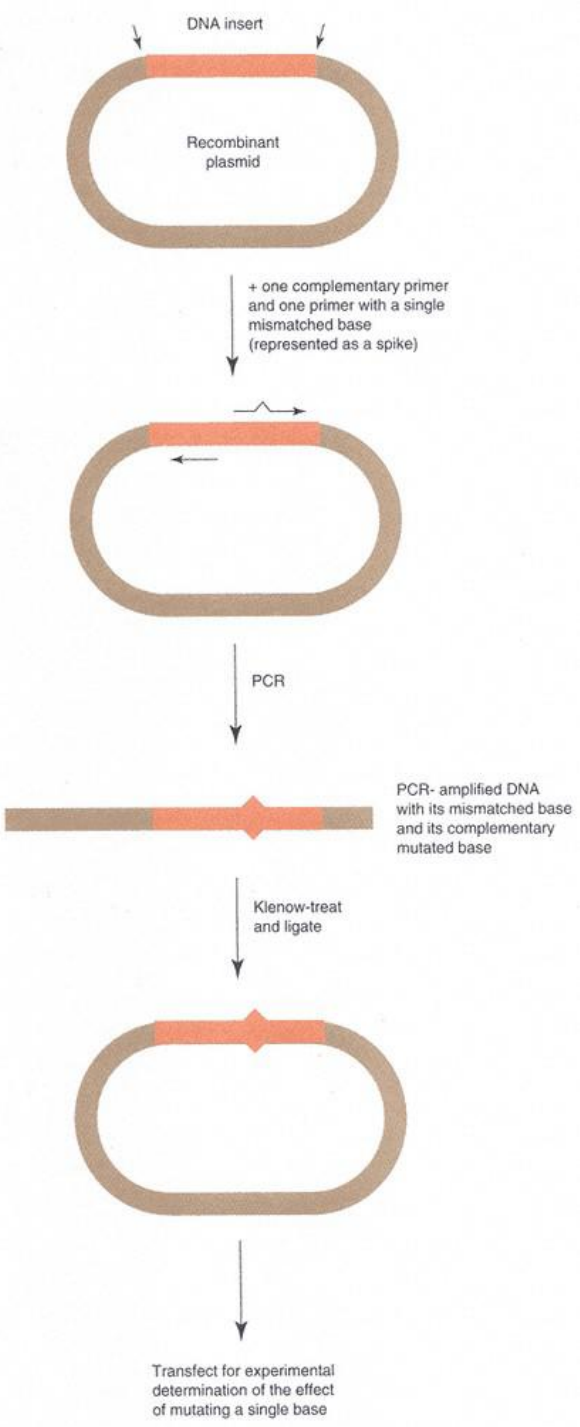


Figure 18.25
Inverse PCR mutagenesis.
A single base can be mutated in recombinant DNA plasmids by inverse PCR. Two primers are synthesized with their antiparallel 5 ends complementary to adjacent bases on the two strands of DNA. One of the two primers carries a specific mismatched base that is faithfully copied during the PCR amplification steps, yielding ultimately a recombinant plasmid with a single mutated base.

such as viral infections. Antisense technology, along with site-directed mutagenesis, are part of a new approach termed **reverse genetics.** Reverse genetics (from gene to phenotype) selectively modifies a gene to evaluate its function, as opposed to classical genetics, which depends on the isolation and analysis of cells carrying random mutations that can be identified. A second use of the term reverse genetics refers to the mapping and ultimate cloning of a human gene associated with a disease where no prior knowledge of the molecular agents causing the disease exists. The use of the term "reverse genetics" in this latter case is likely to be modified.

Antisense RNA can be introduced into a cell by common cloning techniques. Figure 18.26 demonstrates one method. A gene of interest is cloned in an expression vector in the wrong orientation. That is, the sense or coding strand that is normally inserted into the expression vector downstream of a promoter is intentionally inserted in the opposite direction. This now places the complementary or antisense strand of the DNA under the control of the promoter with expression or transcription yielding antisense RNA. Transfection of cells with the antisense expression vector introduces antisense RNA that is capable of hybridizing with normal cellular mRNA. The mRNA–antisense RNA complex is not translated due to a number of reasons, such as its inability to bind to ribosomes, blockage of normal processing, and rapid enzymatic degradation.

DNA oligonucleotides have also been synthesized that are complementary to the known sequences of mRNAs of selected genes. Introduction of specific DNA oligomers to cells in culture have inhibited viral infections including infections by the human immunodeficiency virus (HIV). It is conceivable that one day bone marrow cells will be removed from AIDS patients and antisense HIV nucleic acids will be introduced into their cells in culture. These "protected" cells can then be reintroduced into the AIDS patient's bone marrow (autologous bone marrow transplantation) and replace those cells normally destroyed by

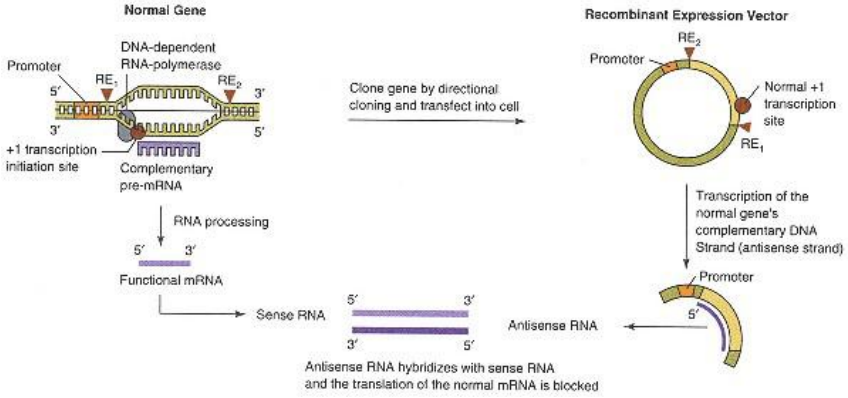


Figure 18.26
Production of antisense RNA.
A gene, or a portion of it, is inserted into a vector by directional cloning downstream of a promoter and in the reversed orientation to that normally found in the cell of origin. Transfection of this recombinant DNA into the parental cell carrying the normal gene results in the transcription of RNA (antisense RNA) from the cloned reversed-polarity DNA along with a normal cellular mRNA (sense RNA) transcript. The two anti-parallel complementary RNAs hybridize within the cell, resulting in blocked expression (translation)of the normal mRNA transcript.

the virus. Experimental progress is also being made with antisense nucleic acids that can regulate the expression of oncogenes, genes involved in the cancer-forming process. Harnessing antisense technologies holds great promise for treatment of human diseases.

Normal Genes Can Be Introduced into Cells with a Defective Gene in Gene Therapy

It is sometimes desirable for the transfected recombinant DNA to replicate to high copy numbers independent of the cell cycle. In other situations it is preferable for only one or few copies to integrate into the host genome with its replication regulated by the cell cycle. Individuals who possess a defective gene resulting in a debilitating or fatal condition could theoretically be treated by supplying their cells with a normal gene. **Gene therapy** is in its infancy; however, the successful transfer of a normal gene to humans has been accomplished employing retroviral vectors (see Clin. Corr 18.7). The success of gene transfer depends, in part, on integration of the gene into the host genome. This is directed by the retroviral integration system. Integration, however, is normally a random event that could result in deleterious sequelae. Exciting studies are in progress that indicate that the viral integration machinery can be selectively tethered to specific target sequences within the host DNA by protein–protein interactions to obviate these potential problems.

CLINICAL CORRELATION 18.7

Normal Genes Can Be Introduced into Cells with Defective Genes in Gene Therapy

More than 4000 different genetic diseases are known, many of which are debilitating or fatal. Most are currently incurable. With the advent of new technologies in molecular biology, the clinical application of gene transfer and gene therapy is becoming a reality. Adenosine deaminase (ADA) deficiency and Gaucher's disease are but two of many genetic diseases that may readily be cured by gene therapy.

ADA is important in purine salvage, catalyzing the conversion of adenosine to inosine or deoxyadenosine to deoxyinosine. It is a protein of 363 amino acids with highest activity in thymus and other lymphoid tissues. A defect in the *ADA* gene is inherited as an autosomal recessive disorder. Over 30 mutations are associated with the disease. ADA deficiency causes a severe combined immunodeficiency disease (SCID), by an unknown mechanism. These immune-compromised children usually die in the first few years of life due to overwhelming infections. The first authorized gene therapy in humans began on September 14, 1990 with the treatment of a four-year-old girl with ADA deficiency. The patient's peripheral blood T cells were expanded in tissue culture with appropriate growth factors. The *ADA* gene was introduced within these cells by retroviral mediated gene transfer. A modified retrovirus was constructed to contain the human *ADA* gene such that it would be expressed in human cells without virus replication. (These viruses that cannot replicate are first propagated in a cell line that contains a helper virus to produce "infectious" viruses. The "infectious" viruses with foreign genetic information can now infect and transfer information to cells without helper virus functions and, therefore, cannot replicate.) Transfer of the *ADA* gene to the patient's T cells was mediated by retroviral infection. Modified T cells carrying a normal *ADA* gene were then reintroduced to the patient by autologous transfusion. Levels of ADA as low as 10% of normal are sufficient to normalize the patient.

Gaucher's disease is an autosomal recessive lysosomal storage disorder caused by a deficiency of lysosomal glucocerebrosidase (GC). Clinical problems include hepatosplenomegaly, pancytopenia, and bone deterioration. The enzyme is a lysosomal membrane glycoprotein that contains 497 amino acids. Over 36 mutations, mostly missense, that decrease catalytic activity are associated with the disease. The disorder can be treated with enzyme replacement; however, this is very expensive and the patient must be subjected to intravenous therapy throughout life. Viral constructs, similar to the ADA protocol, have been made that carry the *GC* gene and have been successfully transduced into a Gaucher patient's hematopoietic cells in culture with very high efficiencies. These studies indicate that Gaucher patients may be normalized by gene therapy in the near future. The genetically altered cells would become endogenous factories capable of continuously synthesizing *GC*, thus obviating the need for intravenous delivery of the missing enzyme.

Blaese, R.M. Progress toward gene therapy. *Clin. Immunol. Immunopathol* 61:574, 1991; Mitani, K., Wakamiya, M., and Caskey, C. T. Long-term expression of retroviral-transduced adenosine deaminase in human primitive hematopoietic progenitors. *Hum. Gene Ther*. 4:9, 1993; and Xu, L., Stahl, S. K., Dave, H. P., Schiffman, R., Correll, P. H., Kessler, S., and Karlsson, S. Correction of the enzyme deficiency in hematopoietic cells of Gaucher patients using a clinically acceptable retroviral supernatant transduction protocol. *Exp. Hematol.* 22:223, 1994.

Transgenic Animals

Recombinant DNA methods allow production of large amounts of foreign gene products in bacteria and cultured cells. These methods also facilitate evaluation of the role of a specific gene product in cell structure or function. In order to investigate the role of a selected gene product in the growth and development of a whole animal, the gene must be introduced into the fertilized egg. Foreign genes can be inserted into the genome of a fertilized egg. Animals that develop from a fertilized egg with a foreign gene insert carry that gene in every cell and are referred to as **transgenic animals.**

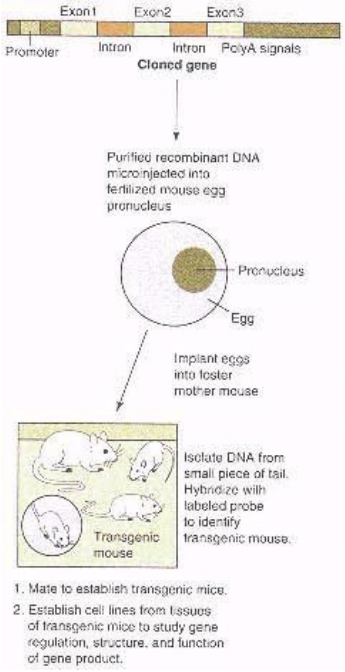


Figure 18.27
Production of transgenic animals.
Cloned, amplified, and purified functional genes are microinjected into several fertilized mouse egg pronuclei *in vitro*. The eggs are implanted into a foster mother. DNA is isolated from a small piece of each offspring pup's tail and hybridized with a labeled probe to identify animals carrying the foreign gene (transgenic mouse). The transgenic mice can be mated to establish a new strain of mice. Cell lines can also be established from tissues of transgenic mice to study gene regulation and the structure/function of the foreign gene product.

The most commonly employed method to create transgenic animals is outlined in Figure 18.27. The gene of interest is usually a cloned recombinant DNA molecule that includes its own promoter or is cloned in a construct with a different promoter that can be selectively regulated. Multiple copies of the foreign gene are microinjected into the pronucleus of the fertilized egg. The foreign DNA inserts randomly within the chromosomal DNA. If the insert disrupts a critical cellular gene the embryo will die. Usually, nonlethal mutagenic events result from the insertion of the foreign DNA into the chromosome.

Transgenic animals are currently being used to study several different aspects of the foreign gene, including the analysis of DNA regulatory elements, expression of proteins during differentiation, tissue specificity, and the potential role of oncogene products on growth, differentiation, and induction of tumorigenesis. Eventually, it is expected that these and related technologies will allow for methods to replace defective genes in the developing embryo (see Clin. Corr. 18.8).

Recombinant DNA in Agriculture Will Have Significant Commercial Impact

Perhaps the greatest gain to all humanity would be the practical use of recombinant technologies to improve our agricultural crops. Genes must be identified and isolated that code for properties that include higher crop yield, rapid plant growth, resistance to adverse conditions such as arid conditions or cold periods, and plant size. New genes, not common to plants, may be engineered into plants that confer resistance to insects, fungi, or bacteria. Finally, genes encoding existing structural proteins can be modified to contain essential amino acids not normally present in the plant, without modifying the protein function. The potential to produce plants with new genetic properties depends on the ability to introduce genes into plant cells that can differentiate into whole plants.

New genetic information carried in **crown gall plasmids** can be introduced into plants infected with soil bacteria known as agrobacteria. Agrobacteria naturally contain a crown gall or Ti (tumor-inducing) plasmid whose genes integrate into an infected cell's chromosome. The plasmid genes direct the host plant cell to produce new amino acid species that are required for bacterial growth. A crown gall, or tumor mass of undifferentiated plant cells, develops at the site of bacterial infection. New genes can be engineered into the Ti plasmid, and the recombinant plasmid introduced into plant cells upon infection with the agrobacteria. Transformed plant cells can then be grown in culture and under proper conditions can be induced to redifferentiate into whole plants. Every cell would contain the new genetic information and would represent a transgenic plant.

Some limitations in producing plants with improved genetic properties must be overcome before significant advances in our world food supply can be realized. Clearly, proper genes must yet be identified and isolated for desired characteristics. Also, important crops such as corn and wheat cannot be transformed by Ti plasmids; therefore other vectors must be identified. However, significant success has been achieved in recent years in designing crop plants

CLINICAL CORRELATION 18.8

Transgenic Animal Models

Transgenic animal model systems hold promise for future methodologies to correct genetic diseases early in fetal development. These animals are used to study the regulation of expression and function of specific gene products in a whole animal and have the potential for creating new breeds of commercially valuable animals. Transgenic mice have been developed from fertilized mouse eggs with rat growth hormone (*GH*) genes microinjected into their male pronuclei (see p. 835). The rat *GH* gene DNA, fused to the mouse metallothionein-I (MT-I) promoter region, was purified from the plasmid in which it had been cloned. Approximately 600 copies of the promoter–gene complex were introduced into each egg, which was then inserted into the reproductive tract of a foster mother mouse. The resulting transgenic mouse was shown to carry the rat *GH* gene within its genome by hybridizing a labeled DNA probe to mouse DNA that had been purified from a slice of the tail, restriction endonuclease digested, electrophoresed, and Southern blotted. The diet of the animals was supplemented with $ZnSO_4$ at 33 days postparturition. The $ZnSO_4$ presumably can activate the mouse MT-I promoter to initiate transcription of the rat *GH* gene. The continuous overexpression of rat GH in some transgenic animals produced mice nearly twice the size of littermates that did not carry the rat *GH* gene. A transgenic mouse transmitted the rat *GH* gene to one-half of its offspring, indicating that the gene stably integrated into the germ cell genome and that new breeds of animals can be created.

Palmiter, R. D., Brinster, R. L., Hammer, R. E., et al. Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. *Nature* 300:611, 1982.

with resistance to insects and viruses. Of equal importance is the very recent genetic engineering feat of inserting a foreign gene into pea plants that now produce a protein that inhibits the feeding of weevil larvae on the pea seeds. Peas and other legume seeds will be able to be stored without the need of protective chemical fumigants (currently Brazilian farmers lose 20–40% of their stored beans to pests).

18.15— Concluding Remarks

The old cliché, so close and yet so far away, seems appropriate for our current juncture in molecular biology. The eukaryotic yeast genome, which consists of approximately 14 million base pairs of DNA distributed among 16 chromosomes, will be entirely sequenced by the mid-1990s. Equally impressive is the fact that the entire human genome will likely be sequenced in the next decade or so (see Clin. Corr. 15.11). Two human chromosomes, 16 and 19, have been fully mapped and it is anticipated they will be the first chromosomes to be fully sequenced. More than 100,000 cDNA clones are available for sequencing, which ultimately will provide landmarks of the huge human genetic map. More than 100 clinical trials in gene therapy have been initiated since the apparent success with ADA. Genetic diseases now identified and to be identified should eventually be curable by gene replacement therapy when the technical roadblocks are surmounted. If one looks at the enormous advances made in molecular biology in just the past two decades it is reasonable to believe the "when" will not be that far off.

Bibliography

Askari, F. K., and McDonnell, W. M. Molecular medicine: Antisense oligonucleotide therapy. *N. Engl. J. Med.* 334:316, 1996.

Brown, W. M., George, M. Jr., and Wilson, A. C. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 76:1967, 1979.

Bushman, F. Targeting retroviral integration. *Science* 267:1443, 1995.

Davis, L. G., Kuehl, W. M., and Battey, J. F. *Basic Methods in Molecular Biology*, 2nd ed. Norwalk, CT: Appleton & Lange, 1994.

Erlich, H. A. (Ed.). *PCR Technology. Principles and Applications for DNA Amplification*. New York: Stockton Press, 1989.

Feinberg, A., and Vogelstein, B. Addendum: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem.* 137:266, 1984.

Jaenisch, R. Transgenic animals. *Science* 240:1468, 1988.

Kreeger, K. Y. Influential consortium's cDNA clones praised as genome research time-saver. *The Scientist* 9:1, 1995.

Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* 82:488, 1985.

Marshall, E. Gene therapy's growing pains. *Science* 269:1050, 1995.

Maxam, A. M., and Gilbert, W. A new method of sequencing DNA. *Proc. Natl. Acad. Sci. USA* 74:560, 1977.

McPherson, M. J., Quirke, P., and Taylor, G. R. (Eds.). *PCR. A Practical Approach*, Vol. 1. Oxford, England: Oxford University Press, 1994.

Mulligan, R. C. The basic science of gene therapy. *Science* 260:926, 1993.

Palmiter, R. D., Brinster, R. L., Hammer, R. E., et al. Dramatic growth of mice that develop from eggs microinjected with metallothionein–growth hormone fusion genes. *Nature* 300:611, 1982.

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

Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning. A Laboratory Manual*, 2nd ed. New York: Cold Spring Harbor Laboratory Press, 1989.

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

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

Vogelstein, B., Fearon, E. R., Hamilton, S. R., and Feinberg, A. P. Use of restriction fragment length polymorphism to determine the clonal origin of human tumors. *Science* 227:642, 1985.

Watson, J. D., Tooze, J., and Kurtz, D. T. *Recombinant DNA: A Short Course* San Francisco: Scientific American Books/Freeman, 1983.

Weintraub, H. M. Antisense RNA and DNA. *Sci. Am.* 262:40, 1990.

Williams, N. Closing in on the complete yeast genome sequence. *Science* 268:1560, 1995.

Zhang, Y., and Yunis, J. J. Improved blood RNA extraction micro-technique for RT-PCR. *Biotechniques* 18:788, 1995.

Questions

J. Baggott and C. N. Angstadt

1. Development of recombinant DNA methodologies is based on discovery of:

   A. the polymerase chain reaction (PCR).

   B. restriction endonucleases.

   C. plasmids.

   D. complementary DNA (cDNA).

   E. yeast artificial chromosomes (YACs).

2. The essential property of the DNA polymerase employed in the polymerase chain reaction (PCR) is that it:

   A. does not require a primer.

   B. is unusually active.

   C. is thermostable.

   D. replicates double-stranded DNA.

   E. can replicate both eukaryotic and prokaryotic DNA.

3. Construction of a restriction map of DNA requires all of the following EXCEPT:

   A. partial hydrolysis of DNA.

   B. complete hydrolysis of DNA.

   C. electrophoretic separation of fragments on a gel.

   D. staining of an electrophoretic gel to locate DNA.

   E. cyclic heating and cooling of the reaction mixture.

4. In the Maxam–Gilbert method of DNA sequencing:

   A. cleavage of the DNA backbone occurs randomly at only some of the sites where the base had been destroyed.

   B. all nucleotides produced during cleavage of the DNA backbone are detected by radioautography.

   C. electrophoretic separation of DNA fragments is due to differences in both size and charge.

   D. the sequence of bands in the four lanes of the autoradiogram contains the base sequence information.

   E. dideoxynucleoside triphosphates are used.

5. The Sanger and Maxam–Gilbert methods of DNA sequencing differ in that:

   A. the Maxam–Gilbert method involves labeling the 5 end, while the Sanger method requires labeling the 3 end of the DNA.

   B. only the Maxam–Gilbert method involves electrophoresing a mixture of fragments of different sizes.

   C. the Sanger method employs DNA cleavage, while the Maxam–Gilbert method employs interrupted DNA synthesis.

   D. only the Maxam–Gilbert method uses radioautography to detect fragments in which one of the termini is radioactively labeled.

   E. in the Maxam–Gilbert method, a complete DNA chain is cleaved, while in the Sanger method, synthesis of the chain is interrupted at different points.

6. Preparation of recombinant DNA requires:

   A. restriction endonucleases that cut in a staggered fashion.

   B. restriction endonucleases that cleave to yield blunt-ended fragments.

   C. poly (dT).

   D. DNA ligase.

   E. cDNA.

7. In the selection of colonies of bacteria that carry cloned DNA in plasmids, such as pBR322, that contain two antibiotic resistance genes:

   A. one antibiotic resistance gene is nonfunctional in the desired bacterial colonies.

   B. untransformed bacteria are antibiotic resistant.

   C. both antibiotic resistance genes are functional in the desired bacterial colonies.

   D. radiolabeled DNA or RNA probes play a role.

   E. none of the above.

8. A technique for defining gene arrangement in very long stretches of DNA (50–100 kb) is:

   A. RFLP.

   B. chromosome walking.

   C. nick translation.

   D. Southern blotting.

   E. SSCP.

9. Which of the following pairs of vectors and DNA insert sizes is correct?

   A. plasmids 5–10 kb

   B. cosmids 15 kb

   C YACs 2000–5000 kb

   D. bacteriophage 45 kb

   E. none of the above

10. Expression of a eukaryotic gene in prokaryotes involves which of the following:

A. a SD sequence in mRNA.

B. absence of introns.

C. regulatory elements upstream of the gene.

D. a fusion protein.

E. all of the above.

Refer to the following for Questions 11–15.

A. antisense nucleic acid

B. polymerase chain reaction

C. site-directed mutagenesis

D. shuttle vector

E. transfecton

11. Contains both bacterial and eukaryotic replication signals.

12. Complementary to mRNA and will hybridize to it, thus blocking translation.

13. Can rapidly produce large quantities of a specific DNA.

14. Oligomer with one mismatched base is used as a primer.

15. A process that introduces foreign DNA into a eukaryotic genome.

Answers

1. B The ability to cleave DNA predictably at specific sites is essential to recombinant DNA technology (p. 760).

2. C PCR requires cycling between low temperatures, where hybridization of template DNA and oligomer primers occurs, and high temperatures, where DNA melts (p. 759). The Taq DNA polymerase, isolated from a thermophilic organism discovered in a hot spring on federal land, is stable at high temperatures and makes the cycling possible with no addition of fresh polymerase after each cycle. The lucrative commercialization of this publicly owned natural resource, with no royalties accruing to the public (i.e., taxpayers') coffers, has evoked criticism from some observers.

3. E Cyclic heating and cooling are part of the PCR process, not of restriction mapping (p. 761). A and B: Restriction mapping involves all degrees of hydrolysis. Partial hydrolysis gives fragments of varying sizes, and complete hydrolysis gives the smallest possible fragments. C and D: Fragments are electrophoretically separated by size on agarose gel, which is stained to reveal the DNA.

4 D The relative positions of G are given by the bands in the lane corresponding to the destruction of G; of A by the bands in the AG lane that are not duplicated in the G lane; of C by the bands in the C lane; of T by the bands in the CT lane that are not duplicated in the C lane. A: Cleavage occurs at all such sites. Limited destruction of the bases is random (p. 762). B: Only the nucleotides that contain the labeled 5 terminal are detected. Other nucleotides are produced but are not detected by the method and do not contribute information to the analysis (p. 762). C: Although charge is, of course, required to produce movement of a particle in a field, the separation of these fragments is not due to charge differences, but to size differences, with the smallest fragments migrating farthest (pp. 762–763). E: This is part of the Sanger method (p. 763).

5. D They both use radioautography to detect fragments in which one of the termini is radioactively labeled. A: The Sanger method involves a labeled 5 end. With the Maxam–Gilbert method either end could be labeled. Here we show labeling of the 3 end. B: Both methods do this. C: This statement reverses the methodologies. E: See pp. 762–765.

6. D DNA ligase covalently connects fragments held together by interaction of cohesive ends (p. 765). A: This is the most desirable type of restriction endonuclease to use, but it is not essential. B: Restriction nucleases that make blunt cuts can also be used if necessary. C: This is used in conjunction with poly (dA) if restriction endonucleases that make blunt cuts are employed, but it is not essential to all of recombinant DNA preparation.

7. A The foreign DNA is inserted into one antibiotic resistance gene, thus destroying it (p. 770). B: Resistance is due to the plasmids. C: See the comment for A above. D: Radiolabeling detects the DNA of interest, not the colonies that contain cloned DNA (p. 770).

8. B A: Restriction fragment length polymorphism (RFLP) is a characteristic of DNA, not a technique (p. 775). C: Nick translation is used to label DNA during chromosome walking (p. 773). D: Southern blotting is a method for analyzing DNA (p. 774). E: Single-strand conformation polymorphism (SSCP) is a method for detecting base changes in DNA that do not alter restriction endonuclease sites.

9. A B: Cosmids will accept a 45-kb insert (p. 781). C: YACs will accept a 200–500 kb insert (p. 781). D: Bacteriophage will accept a 15-kb insert (p. 779).

10. E A: The SD sequence is necessary for the bacterial ribosome to recognize the mRNA. B: Bacteria do not have the intracellular machinery to remove introns from mRNA. C: Appropriate regulatory elements are necessary to allow the DNA to be transcribed. D: A fusion protein may be a product of the reaction (p. 783).

11. D (see p. 784).

12. A (see p. 790).

13. B (see p. 759).

14. C (see p. 788).

15. E (see p. 784).