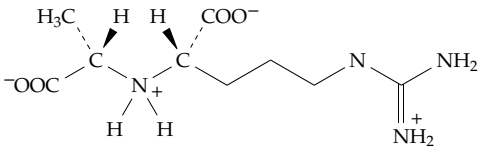
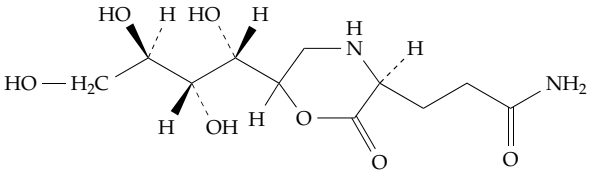
(transferred region) known as T-DNA that encodes enzymes for production of the auxin indoleacetic acid (Fig. 25-12), the cytokinin isopentenyl-AMP, and the compounds known as **opines**. These are reduction products of Schiff bases of amino acids and 2-oxoacids or sugars.



Octopine: derived from pyruvate and arginine
Histopine: derived from pyruvate and histidine[150]
Lysopine: derived from pyruvate and lysine
Nopaline: derived from 2-oxoglutarate and arginine
Leucopine: derived from 2-oxoglutarate and leucine



Agropine: derived from D-mannose and L-glutamine

The auxin and cytokinin, whose production is normally controlled, are now formed in large amounts and cause uncontrolled tumor growth. The opines are used by *Agrobacterium* as a unique source of energy and of metabolites for biosynthesis. The host plant cells, however, cannot catabolize the opines.

Upon entrance into a plant cell the T-region is excised from the Ti plasmid and can become integrated into the DNA of the host plant much as occurs during bacterial conjugation (Fig. 26-3).[151–154a] By deleting the genes for synthesis of auxin, cytokinin, and opines, the Ti plasmid loses its tumorigenic property but is still able to transfer genes into the plant genome. Fusion of an *E. coli* plasmid vehicle into the modified Ti plasmid creates a useful plant shuttle vehicle. Also see below under "physical methods."

Transferring genes with engineered animal viruses. There is special interest in transferring cloned DNA into human cells to correct genetic defects.[155] Transfer of human genes into other animals is also important for a variety of reasons. For example, human proteins of therapeutic value could be produced in animal milk. Animals can be engineered to have defects that mimic those in humans and which can then be studied in animals.[156]

A number of different viruses have been used to transfer foreign genes into eukaryotic cells and also to create stable plasmids for cloning. One of the first was the 5380 bp SV40. The relatively small size of SV40 limits the amount of DNA that can be incorporated. However, many of the functions of this virus can be performed by simultaneous infection with a **helper virus**, often an adenovirus which is itself defective. With the helper virus making proteins essential for SV40 replication, all but 85 base pairs at the origin of replication of SV40 can be deleted and replaced by other DNA. More recently engineered **adenoviruses** and **retroviruses** have been widely used. In both cases genes required for replication of the viruses have been deleted.[155] Adenovirus vehicles can carry 7 kbp or more of passenger DNA. They are efficiently taken up into endosomes and into the nucleus by both replicating and nonreplicating cells. The DNA is not incorporated into the genome of the host cell.[155,157,158] Replication-deficient avian **retroviruses**, or **retrotransposons** (Chapter 28), can carry up to 9 kbp of passenger DNA and are incorporated into the host genome by the process of homologous recombination (Chapter 27).[155,159,160] Both adenovirus and retroviral vehicles suffer from serious problems. The adenovirus proteins can induce fatal inflammatory reactions,[160a] and DNA transfer from the retroviral vehicles is inefficient. A nonpathogenic human parvovirus, the **adeno-associated virus** (AAV), is also being developed as a gene therapy vector.[161,161a,b] Another possibility is to engineer a **lentivirus**, even HIV, to provide efficient integration (Chapter 28).[162,162a] Safety concerns have delayed human testing of this possibility.

A group of insect viruses, the **baculoviruses**, are being used in cultured insect cells for large-scale production of proteins. The cloned genes are placed under the control of the promoter region of the gene that encodes the major viral coat protein.[135,163] The **baculovirus** vehicles can carry over 15 kbp of passenger DNA and may also be useful in human gene therapy.[164,165] Filamentous bacteriophages have been reengineered for the same purpose.[166]

Nonviral methods for DNA incorporation into the human genome may utilize **transposons** (Chapter 27)[166a] or mobile **group II introns** (Chapter 28).[166b,c]

Physical methods of gene transfer. Genes can often be transferred without the use of a cloning vehicle. This is especially important for certain plant cells, such as those of cereal grains, for which transfer of genes via the Ti plasmid has been difficult.[167] If DNA, which may be in a plasmid, is coprecipitated with calcium phosphate, it can often be taken up directly either by animal cells or by plant protoplasts.[168,169] Polycations also facilitate DNA uptake; cationic **liposomes** seem to be especially effective.[170] In the widely used **electroporation** technique a short electrical pulse of a few hundred volts / cm is applied to create transient pores in the plasma membrane through which the DNA can enter a cell.[108,171–175] Chromosomes can be transferred by cell fusion and either

entire chromosomes, isolated DNA, or cloned DNA can be transferred into egg cells by microinjection.[176] Following transfer of DNA or of intact chromosomes, recombination sometimes leads to stable incorporation of some of the transferred genes into the host cell's genome.

A very important technique is the use of high-velocity microprojectiles shot from a particle gun. Spherical tungsten particles of ~0.5–1 μm diameter are coated with recombinant plasmids containing the genes to be transferred. The particles are then shot, using a gunpowder cartridge, into intact recipient cells. The particles penetrate cell membranes, mitochondria, and chloroplasts without serious damage if the number of particles is not too great.[177,178] The technique is very useful for genetic engineering of plants.[179]

5. Genomic Libraries

The human genome contains about 3×10^6 kb of DNA, about the average amount for a eukaryotic organism. If the entire genome is digested to completion with a restriction endonuclease, whose cleavage sites are distributed more or less randomly, the resulting restriction fragments constitute a "library" for that genome. If the average length of a fragment is 17 kb, about 1.8×10^5 unique fragments will be produced. To make a practical library these must be cloned into a suitable vehicle. Derivatives of phage λ or cosmids are most often used. The cloned fragments can then be packaged using the *in vitro* packaging system to form infectious phage particles, which can be propagated as plasmids in *E. coli* cells.

What is the probability that a given fragment among the total produced will be found in one of the recombinant phage in the library? From simple probability theory the number of clones that must be isolated and screened is given by Eq. 26-2.[180] Here N is the number of clones needed, p is the probability of

$$N = \ln(1-p)\,\ln(1-f) \qquad (26\text{-}2)$$

having the desired fragment in the library, and f is the fractional proportion of the genome represented by the fragment sought. For our example of a 17-kb fragment of the human genome this is 17 / 3000. From this equation one can calculate that to have a 99% chance of finding our fragment ($p = 0.99$) we need 8×10^5 clones.[99] What the equation does not show is that there will probably be some long fragments that cannot be cloned in the selected vehicle. These will be missing from the library.

Large numbers of clones obtained can be screened rapidly by colony hybridization using a labeled DNA probe. Thus, if it is desired to isolate a gene for a particular protein and some part of that protein has been sequenced, a synthetic DNA probe can be made. The phage containing the recombinant fragments can be plated, and after plaques form a nitrocellulose filter can be laid on the plate to form a replica. After release of DNA from the phage, denaturation by NaOH, and neutralization the single-stranded DNA fragments can be hybridized with the probe. Another screening method uses a probe carried in a plasmid that promotes homologous recombination between the probe sequence and restriction fragments with a similar sequence.[181] A problem that arises in preparing genomic libraries is that certain sequences, e.g., those involving highly polymorphic regions and inverted repetions, often cannot be propagated in most lambda cloning vehicles.[182]

In addition to genomic libraries, **cDNA libraries** can be prepared from mixed mRNAs. The total RNA of cells is isolated and passed through an affinity column containing oligo(dT) chains. These bind to the 3'-poly(A) tails of the mRNAs, allowing them to be isolated. The mixed mRNAs can then be cloned using a poly(AT)-tailed plasmid vehicle and a reverse transcriptase.[183,184]

6. Probes

The first step in screening the recombinant DNA in a library is use of some probe for detecting the desired DNA fragment. The most direct way is to synthesize a radioactive or fluorescent labeled oligonucleotide[185] complementary to a short known sequence in the protein. The number of codons for a single amino acid varies from 1 to 6 (Table 5-6). It is therefore desirable to prepare a probe complementary to segments of DNA containing a high proportion of codons for Trp and Met (1 codon each) and Asn, Asp, Cys, Glu, Gln, His, Lys, Phe, and Tyr (2 codons each). A popular procedure is to synthesize a mixture of probes containing all of the possible nucleotide sequences coding for the selected sequence of amino acids. The probe may be a mixture of more than 1024 different nucleotide sequences.[186] See also Chapter 5, Section H,4.

Antibodies are another popular type of probe. Antibodies to a specific protein may be utilized in isolating mRNA from ribosomes that are making that protein (Chapter 29). Thus one or more strongly binding antibodies may have already been obtained before the library clones are to be screened. To use this technique for screening recombinant DNA, the cloning must be done in a vehicle that causes expression of the gene, e.g., as proteins fused to *E. coli* galactosidase. One type of expression library is created by insertion of the cDNAs into copies of a bacteriophage gene that permits the expressed proteins to be displayed on surfaces of the phages[186a] (see Fig. 3-16).

7. Studies of Restriction Fragments

By cleavage with the correct restriction enzyme, cloned DNA fragments can be released from the vehicle in which they are carried in the library. What can be done with these fragments? The first obvious use is to sequence a gene that has been located with a probe. In many instances the gene will be longer than the cloned piece. However, the isolated restriction fragment, if labeled and denatured, becomes a highly specific probe for locating other restriction fragments that overlap it. For example, a fragment from an *Eco*RI library may bind to two or more fragments from a *Hin*dIII library. This permits "walking" along the chromosome to locate adjacent fragments. Cosmid vehicles that facilitate ^{32}P-labeling at the ends of the passenger DNA are useful.[187] A related approach called "jumping" depends upon converting very large DNA restriction fragments into circular molecules, digesting with restriction enzymes, and cloning the junction fragments of the circles. These fragments contain segments that may have been separated by as much as 100 kb in the genomic DNA, and enable the investigator to walk or jump again from a new location.[188]

Locating mutations. The study of restriction fragments provides a way of locating many mutations. For point mutations in genes for known proteins, sequencing reveals the exact defect. Some mutations, especially deletions, lead to changes in the lengths of restriction fragments. If the mutation causes loss of a restriction site, a longer piece of DNA will be present than in a digest of normal DNA. Such differences in length of restriction fragments are usually referred to as **restriction fragment length polymorphism** (RFLP, usually used in the plural as **RFLPs**). These polymorphisms are readily detectable by differences in mobility on gel electrophoresis.[96,189] They can often be mapped to particular chromosomes by hybridization *in situ* (Fig. 26-14),[190,191] by study of naturally occurring translocations of chromosome fragments, or other techniques.[192] An example is provided by the human hemoglobin abnormalities known as **thalassemias** (Chapter 32). Here, deletions remove certain restriction sites leading to observation of RFLP. Occasionally a point mutation is linked to RFLP at a nearby site. For example, in the United States most carriers of the sickle cell trait have a 13-kb *Hpa*I fragment that carries the globin S gene. However, noncarriers have their globin gene on a 7.6-kb *Hpa*I fragment.[99] Although the association is fortuitous, the linkage between the hemoglobin S gene and the mutated restriction site is broken only rarely by crossing-over during meiosis. RFLPs have been linked to many other human genetic defects and have also provided the basis for the first linkage maps of human chromosomes.[96,193]

Positional cloning. By using enzymes that cut at relatively rare sites, genomic DNA can be cut into very large restriction fragments. From studies of inheritance within families carrying specific genetic traits it is sometimes possible to find linkages between those traits and polymorphic restriction fragments.[194–196] This has been accomplished for a number of defective human genes including those responsible for sickle cell anemia,[197] cystic fibrosis (Box 26-A), Duchenne muscular dystrophy (Box 19-A), Huntington's disease,[198,199] X-linked chronic granulomatous disease[200] (p. 1072), neurofibromatosis (elephant man's disease),[201] the hereditary cancer retinoblastoma,[202] and others. By 1997 nearly 100 hereditary disease loci had been located by positional cloning.[203] These astonishing successes provided a major impetus for what became the Human Genome Project (Section G).[193]

Serious problems were met in actually locating these disease genes. Crossing-over is infrequent, occurring only at about 50 locations during each meiosis.[192] Therefore, linkage analysis does not tell us with any precision how close the linked gene is to a known DNA probe within a restriction fragment that may be up to 2000 kb in length. Finer restriction mapping or chromosome walking can be used to locate the precise piece of DNA that is defective.[204] This can still be a formidable problem. However, if the defective protein can be identified it can be sequenced. A specific oligonucleotide probe can be made for its gene and can be used to establish the exact chromosome location.

8. Directed Mutation

In addition to the developments of cloning and sequencing of DNA, a third technique is essential to the present revolution in molecular genetics. That is the ability to mutate any gene at any point in a specific way. Because of its precise nature the technique is called **directed mutation** in this book. However, the term **site-directed mutagenesis** is often used. Mutations can be introduced randomly in DNA in many ways including treatment with nitrous acid, bisulfite, formic acid, or hydrazine or by incorporation of nucleotide analogs.[205] Efficient procedures have been devised for isolating the mutants.[206] For many purposes **oligonucleotide-directed mutation** is the preferred technique.[207–210] An oligonucleotide of ~16–20 nucleotide length is synthesized with a sequence complementary to the coding strand containing the desired site of mutation. At that site the codon for the new amino acid is present. Despite this mismatch the oligonucleotide can be successfully hybridized with a single-stranded DNA such as that cloned in an M13 vehicle. Now the Klenow fragment of DNA polymerase I or a viral DNA polymerase (Chapter 27) is used to convert the single-stranded circular DNA into

a double-stranded replicating form. Many of the single-stranded progeny will contain the mutated DNA. They can be screened with a labeled probe made from the oligonucleotide used to induce the mutation. It will hybridize most tightly to the correctly mutated gene.[207] Use of the PCR reaction simplifies the procedure.[210–212] In another screening procedure the template strand is synthesized in an M13 phage vehicle using uracil rather than thymine. The circular heteroduplex obtained after synthesis with T4 DNA polymerase and T4 DNA ligase is taken up by *E. coli* cells, which select against the uracil-containing strand and, therefore, in favor of the mutated strand.[208]

A third approach is to completely synthesize a gene for the protein under study. The sequence does not have to be exactly the natural one but can be made with restriction enzyme cleavage sites that permit easy excision and readdition of particular fragments. Synthetic fragments containing various mutations can then be grafted in at will. Genes of this type have been made for rhodopsin and related proteins (Chapter 23) and for numerous other proteins.

What are the uses of directed mutation? As we have seen in previous chapters, the technique is being used in every area of biochemistry to bring new understanding of protein functions and of the chemical basis of disease. Together with complete synthesis of genes it provides the basis for genetic engineering of specific proteins of plants, animals, and microorganisms. Many protein products can probably be improved. For example, enzymes can be made more stable.[213] Specificities can be changed, but this is difficult.

Targeting and replacing genes. One goal of human genetic therapy is to replace a defective gene in body cells with a good gene. Is this really possible? It is essential to engineer the DNA that is to be transferred, so that it contains all of the components needed for efficient expression in the host following its incorporation as a **transgene**. A transcription-initiation region with suitable promoter, both 5' and 3' untranslated regions, start and stop codons, and polyadenylation site (Chapter 28) must all be present. It is hoped that a correctly constructed promoter region will allow the transgene to be picked up by the machinery of homologous recombination (discussed in Chapter 27) and be incorporated into the host's DNA and expressed in the appropriate tissues.[214,215] Only some cells will take up the new gene and discard the old. However, it may happen enough to benefit a patient. Targeted gene replacement has been very successful in mice.[216]

Knockout mice. If targeted DNA is injected into a fertilized mouse egg, there is a chance that the mouse will have the targeted gene replaced in one chromosome, and that it will be stably transmitted to some of its progeny. If the transgene is totally nonfunctional, the mouse will be a "knockout mouse," suffering from a hereditary defect that can be transmitted through carriers such as its mother. The standard knockout technique is to inactivate the gene of choice in cultured embryonic stem cells and to inject these into mouse embryos. Some progeny will carry the inactivated gene in their germ cells. A refinement of the technique utilizes the **Cre recombinase** or related enzymes discussed in Chapter 27 to selectively remove pieces of DNA from genes in specialized tissues of mice.[217,218] By 1996 several hundred different knockout mice had been created.[215] Nevertheless, interpretations of results of gene knockout are sometimes complex.[219]

F. The Genetics of Eukaryotic Organisms

Whereas DNA synthesis takes place almost continuously in a rapidly growing bacterium, replication of DNA occupies a more limited part of the **cell cycle** of eukaryotes (Fig. 11-15). In a mammalian cell mitosis proper (Fig. 26-11) may require about one hour. It is followed by the "gap" period, G_1, whose length is variable and depends upon the cell type, the nutritional state of the cell, and other factors. About 10 h is typical. During the S phase (~9 h) active DNA replication takes place. This is followed by a second gap (G_2) that occupies 4 h in the 24 h cell cycle shown in Fig. 11-15. The length of the different segments of the cell cycle varies widely among different organisms. Indeed, the concept of a cell cycle can be criticized.[219a] It is only in a rapidly growing culture that all, or most, cells can follow the same cycle. In the adult body most cells are inhibited from division (or are not stimulated to divide) most of the time.

1. Mitosis

The distribution of chromosomes to daughter cells of somatic cells undergoing division is accomplished by mitosis whose successive phases are referred to as **prophase**, **metaphase**, **anaphase**, and **telophase** (Fig. 26-11). As the chromosomes condense during prophase, it is seen that each one actually consists of two separate entities coiled together. These are the identical **chromatids,** which are formed from the two identical double-stranded DNA molecules formed by replication of the DNA of the chromosome during the S phase of the cell cycle. As the folding of the chromosomes occurs (during prophase), the nuclear envelope completely fragments or dissolves in many species.

An important event that *precedes* the main stages of mitosis is the formation of **poles** in the cell. In animal cells, the poles are formed by the **centrioles,**

which move apart and take up positions at opposite sides of the cell. Each of the centrioles is accompanied by a smaller "daughter" centriole lying at right angles to the larger parent. In plant cells, which lack centrioles, a more diffuse pole is formed. As the cell prepares for mitosis, fine microtubules (15 nm diameter) can be seen radiating from the poles. At the end of prophase the microtubules run from one pole to the other to form the **spindle**. Microtubules also become attached to the chromosomes at the **centromeres**.

At metaphase the chromosomes are precisely lined up in the center of the cell to form the **metaphase plate**. Now each centromere divides, permitting the sister chromatids to be completely separated. A protein complex **cohesin**, which holds the sister chromatids together, undergoes proteolysis by a **separase** at this stage.[220,220a,b] During anaphase the separated chromatids, now referred to as **daughter chromosomes**, move to opposite poles as if pulled by contraction of the spindle fibers. Telophase is the final stage in which new nuclear envelopes are formed around each set of daughter chromosomes. In humans and many other species the cell pinches in two. In plants and fungi new plasma membranes and cell walls are constructed through the center of the cell. A partitioning of mitochondria and of Golgi components between cells must also occur.[220c]

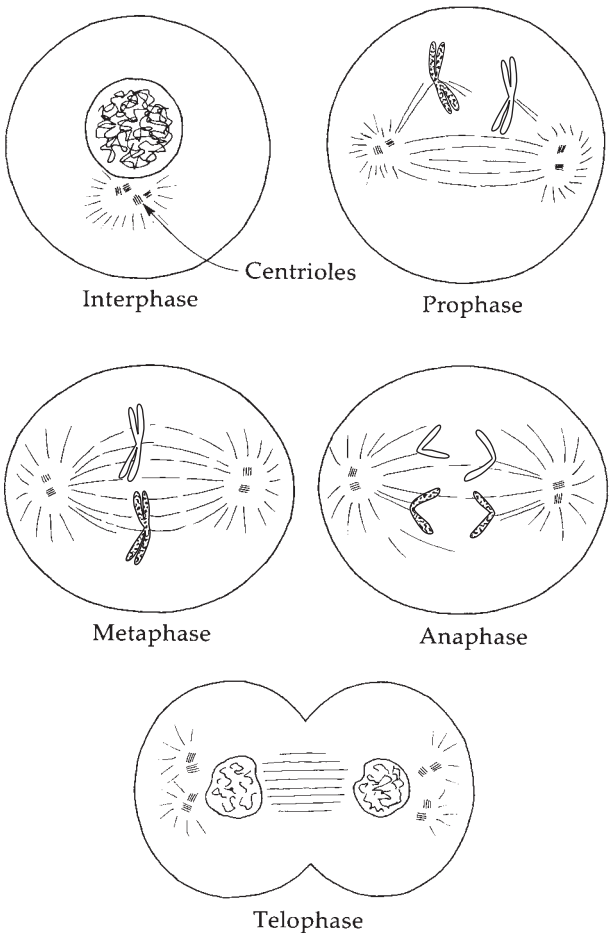


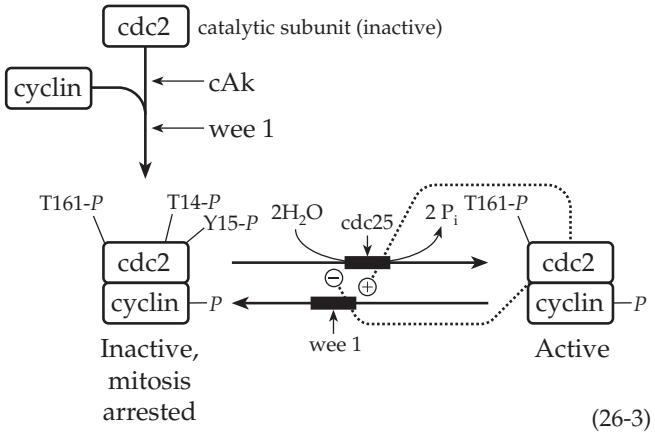
Figure 26-11 Mitosis. Illustrated for a cell with one homologous pair of chromosomes. After Mazia.[221]

The foregoing description overlooks the extreme complexity of mitosis, each stage of which must occur with precision and in the correct sequence.[222–225a] The replication of DNA, which takes place in the S phase of the cell cycle (and is discussed in Chapter 27) must be completed before mitosis begins. This is followed by condensation of the DNA into chromosomes (Chapter 27), breakdown of the nuclear membrane,[226,226a] assembly of the **kinetochores** by which the chromosomes attach to the spindle,[222] assembly of the spindle, attachment of chromosomes to the spindle, segregation of the chromosomes to opposite poles in anaphase, and finally the cleavage of the cell.

Cyclin-dependent kinases. As is shown in Fig. 11-15, the cell cycle is controlled by a series of complexes of the 30- to 45-kDa proteins called **cyclins** with **cyclin-dependent protein kinases** (**CDKs**). These kinases contain ~300-residue catalytic cores that resemble protein kinase A (Fig. 12-32). Like that kinase they transfer phospho groups from ATP to serine and threonine side chains of target proteins.[227–228] The kinases are inactive until a complex with the appropriate cyclin is formed and is activated by phosphorylation.[229] One or more additional kinases are required for this activation. Each stage in the cell cycle is controlled by one or more different cyclin–CDK complexes (Fig. 11-15). One of the best known CDKs is human **CDK2**, which functions in a complex with cyclin A during the S phase of the cycle.[228,230–232] Binding of the cyclin greatly alters the conformation of CDK2, opening the catalytic cleft and exposing threonine 160. Its hydroxyl group can be phosphorylated by the action of the **CDK-activating kinase** (**CAK**) with a 100-fold increase in catalytic activity.[228] Control of the CDKs also depends upon proteins that act as specific inhibitors and upon precise elimination of both cyclins and inhibitors via the ubiquitin system.[228a,b,c]

Checking for completion of replication and for DNA damage. The first checkpoint in the cell cycle is the **start** or **G_1 DNA damage checkpoint** (Fig. 11-15). Replication of DNA does not begin until the cell has had time to repair as much damage to DNA as is possible.[228d] As mentioned in Box 11-D, the cancer suppressor protein p53 is an essential component of the checking process.[224] A second checkpoint, the **G_2 checkpoint**, at the end of the G_2-phase (Fig. 11-15) requires verification that all DNA of all chromosomes has been replicated, checked for damage, and repaired if necessary. Control of these processes is accomplished by a mechanism first identified in

fission yeasts, such as *Schizosaccharomyces pombe*, and which has been conserved in metazoa.[223,233–236d] The CDK known as **cdc2** (fission yeasts), CDC28 in budding yeast, or CDC2 or CDK1 in mammals, functions from G_2 through the DNA damage checkpoint to mitosis (Fig. 11-15) and is sometimes described as the master cell-cycle switch.[236a] The catalytic subunit of cdc2, a serine / threonine protein kinase, is often called $p34^{cdc2}$. It is inactive unless complexed with a suitably phosphorylated cyclin (Eq. 26-3) and phosphorylated on Thr 161 by the action of kinase CAK (Eq. 26-3). However, when cdc2 becomes phosphorylated on Tyr 15 (in *Schizosaccharomyces pombe*) by the action of protein kinase wee 1 (Eq. 26-3) and on Thr 14 by another kinase, cdc2 is inhibited and mitosis is arrested at the G_2 checkpoint. This allows time to verify completion of replication as well as for repair before the replicated DNA strands are separated in mitosis. Hydrolytic removal of the phospho groups from Tyr 15 and Thr 14 of cdc2 by the action of phosphatase cdc25 then allows mitosis to begin.[223,237,237a] This phosphatase is also controlled by a phosphorylation-dephosphorylation cycle. The checkpoint kinase **Chk1** phosphorylates and inhibits cdc25.[238,238a]



(26-3)

When damaged DNA is present, protein p53 accumulates, just as at the G_1 checkpoint, and activates transcription of the seven phosphoserine-binding proteins of the 14-3-3 family.[224] These bind to the phosphorylated cdc25 phosphatase preventing activation of cdc2 by dephosphorylation.[224,238–240] In fission yeast one of the 14-3-3 proteins, **Rad 24**, apparently binds to the phosphorylated cdc25 and induces its export from the nucleus preventing mitosis.[241] Dephosphorylation of cdc25 and return to the nucleus allows mitosis to occur. A somewhat different regulatory mechanism is used by the budding yeast *Saccharomyces cerevisiae*.[242] In animal cells there are three cdc25 isoforms.[240,240a,b] In each case we probably have only a glimpse of a very complex process of checking for DNA damage and repair (see also Chapter 27).

Mitotic spindle formation and the spindle assembly checkpoint. Separation of the two copies of each replicated chromosome depends upon the spindle fibers. Their formation is preceded by replication of centrioles, if present, and formation of the two poles of the cell.[243–244b] Both γ-tubulin (Fig. 7-34) and acidic Ca^{2+}-binding proteins called **centrins** are involved.[244c–e] The microtubules appear to grow outward from the poles with their minus ends at the poles and their plus ends (Fig. 7-33) available for binding to the kinetochores, specialized protein complexes that assemble around the centromeric DNA (Chapter 27).[245–247] Each chromosome has two kinetochores, one for each daughter chromatid. These must be attached to spindle fibers coming from opposite poles. It has usually been assumed that random encounters of microtubule plus ends with kinetochores leads to correct linkage.[245,246] However, neither centrosomes or kinetochores are always essential to spindle assembly, and self-assembly occurs by motor-driven sorting according to polarity of the microtubules.[246,247] In addition to microtubules formation of spindles may require specialized matrix proteins.[247a]

The spindle isn't formed until replication is complete. A small G protein called **Ran**, a relative of Ras (Fig. 11-7), regulates spindle formation. Ran, in turn, depends upon a nucleotide exchange factor called **RCC1** (**regulator of chromatin condensation**).[248–250] RCC1 may signal that the replicated DNA is folding into chromosomes indicating that replication is complete. At the **spindle assembly checkpoint** (Fig. 11-15) the cell verifies that the metaphase spindle has been assembled correctly.[250a] All of the microtubules that pull the sister chromatids toward one pole or the other must be correctly attached to a kinetochore.[220,250b] In addition, there are interdigitated microtubules coming from both poles. Specific motor molecules (Chapter 19) then push the two poles apart. During the assembly and complex movements of the spindle both cytosolic dynein and four different types of kinesin-like motor molecules are required.[251,252] One of these is Kar3 (see Fig. 19-17). After assembly of the spindle is complete, a signal must be sent to the mitotic apparatus to move into anaphase. A clue to the nature of the signal has come from the observation that a single kinetochore lacking a spindle fiber connection causes arrest of mitosis.[253] Apparently unattached kinetochores send a "wait" signal, perhaps via the cytoskeleton.[253–255]

Anaphase. After the spindle has been checked, a sudden loss of cohesion between the sister chromatid pairs allows them to move toward the opposite poles. This process is catalyzed by the **anaphase-promoting complex** (**APC**, or cyclosome) and its activator protein **Cdc20**, a large multiprotein complex.[225] The APC also promotes proteolytic breakdown of cyclins and other

proteins by a ubiquitin- and proteosome-dependent mechanism.[225,256–259c] Its E_3-ubiquitin ligase (Box 10-C) targets the mitotic cyclins and other proteins for destruction.[260] A specific E_2 ubiquitin-conjugating enzyme (Box 10-C) is required for degradation of cyclin B and exit from mitosis.[261] The centrosome also plays an active role in cytokinesis, the final step in cell division.[261a]

The cell cycle encompasses so many different processes that it is clearly impossible to describe it by the single diagram of Fig. 11-15 or by the text written here. The cycle is influenced by a host of growth factors and external stimuli, many of which act on transcription of cyclins and other essential proteins. Transcription factors such as those of the Fos / Jun (AP-1) family in response to the MAP cascade (Fig. 11-13) are among those that control the transcription of cyclins.[262–263c] However, during mitosis most transcription of any genes is repressed.[264]

Among other factors influencing the cell cycle is the size of the cell and the availability of nutrients including purine and pyrimidine nucleotides.[263b,c,264a] Lack of cholesterol decreases the cdc2 kinase activity and causes apoptosis.[265] A cell cycle regulator in *S. pombe* known as **suc1** is essential for cell cycle progression. Although its three-dimensional structure is known, its function (like that of its human homolog CksHs2) is uncertain.[266,267]

2. Meiosis

The mechanism by which chromosomes are distributed during the formation of **gametes** (egg and sperm cells) is known as meiosis (Chapter 1; Fig. 26-12). Formation of gametes involves a halving of the chromosome content of a cell, each gamete receiving only one chromosome of each homologous pair. Genes found in the same chromosome are said to be **linked** because of their tendency to be passed together to the offspring. Genes present in different chromosomes are not linked, and their inheritance follows the pattern of **random segregation** established in Mendel's famous studies.

The simple fact that the genetic material is put up in several different packages (chromosomes) is sufficient to provide for considerable mixing of genetic information between different individuals in sexual reproduction. However, it doesn't provide a means for exchanging genes on the same chromosomes. Mixing of genetic information within chromosomes occurs by genetic recombination occurring during **crossing-over**, an aspect of meiosis with an essential biological role. In the S phase preceding meiosis, DNA is duplicated just as it is prior to mitosis. This provides sufficient genetic material to produce *four haploid cells*. These are formed during meiosis by two consecutive cell divisions (Fig. 26-12). Crosssing-over occurs prior to the first of these divisions, at the four-strand stage. The two homologous chromosomes of a pair come together to form what is called a **bivalent** or **tetrad** made up of four chromatids. For each chromosome, at least one chromatid is seen to come into intimate contact with a chromatid in the other homologous chromosome at points known as **chiasmata** (Fig. 26-13). During metaphase of the first meiotic cell division the homologous chromosomes, each still containing two chromatids, separate. Each chromatid now carries with it some genetic information that was previously found in the other member of the homologous pair and vice versa (Fig. 26-12). Now, without further replication of DNA in the second meiotic cell division, the chromatids separate to form haploid cells.

The process of crossing-over provides a means by which genes that are linked on the same chromosome can be separated, providing offspring with mixtures of genetic traits other than those predicted by simple Mendelian theory. The effects of crossing-over were first studied extensively by T. H. Morgan with the fruit fly *Drosophila melanogaster*. Morgan discovered his first mutant, a white-eyed fly,[267a] in 1910. The first genetic maps were made by assuming a direct relationship between the frequency of crossing-over and the linear distance between genes in a chromosome. Thus, the same approach to genetic mapping that was used later with *E. coli*, i.e., the measurement of recombination frequencies, was applied much earlier to crossing-over in the chromosomes of *Drosophila*. Extensive genetic maps involving many mutations were obtained for the four chromosomes of this organism, and similar techniques have been applied to many other organisms. The unit of distance in these chromosome maps is the **morgan** (named for T. H. Morgan[2,267a]). One centimorgan is the distance that allows recovery of 1% of recombinant progeny.[2] In the human chromosomes, this is ~1000 kb (1 Mb).

Biochemical and genetic studies of meiosis have been conducted in many organisms including fission[268,269] and budding[270–271a] yeasts, *Drosophila*,[272,273] starfish,[274] *Xenopus*,[275] and the mouse.[276] Meiosis can be viewed as a modification of mitosis but with the added initial step of crossing over and recombination. In addition, the S-phase of the cell cycle is absent in the second meiotic division. As was mentioned in Chapter 1, meiosis may occur at different stages of the life cycle of organisms. An important advantage in using fungi for genetic studies is that, like prokaryotes, they are haploid during much of their life cycle. Biochemical defects such as the inability to synthesize a particular nutrient can be recognized readily at this stage. At the same time genetic crosses can be made, and crossing-over frequencies can be measured and used for genetic mapping. The onset of meiosis may

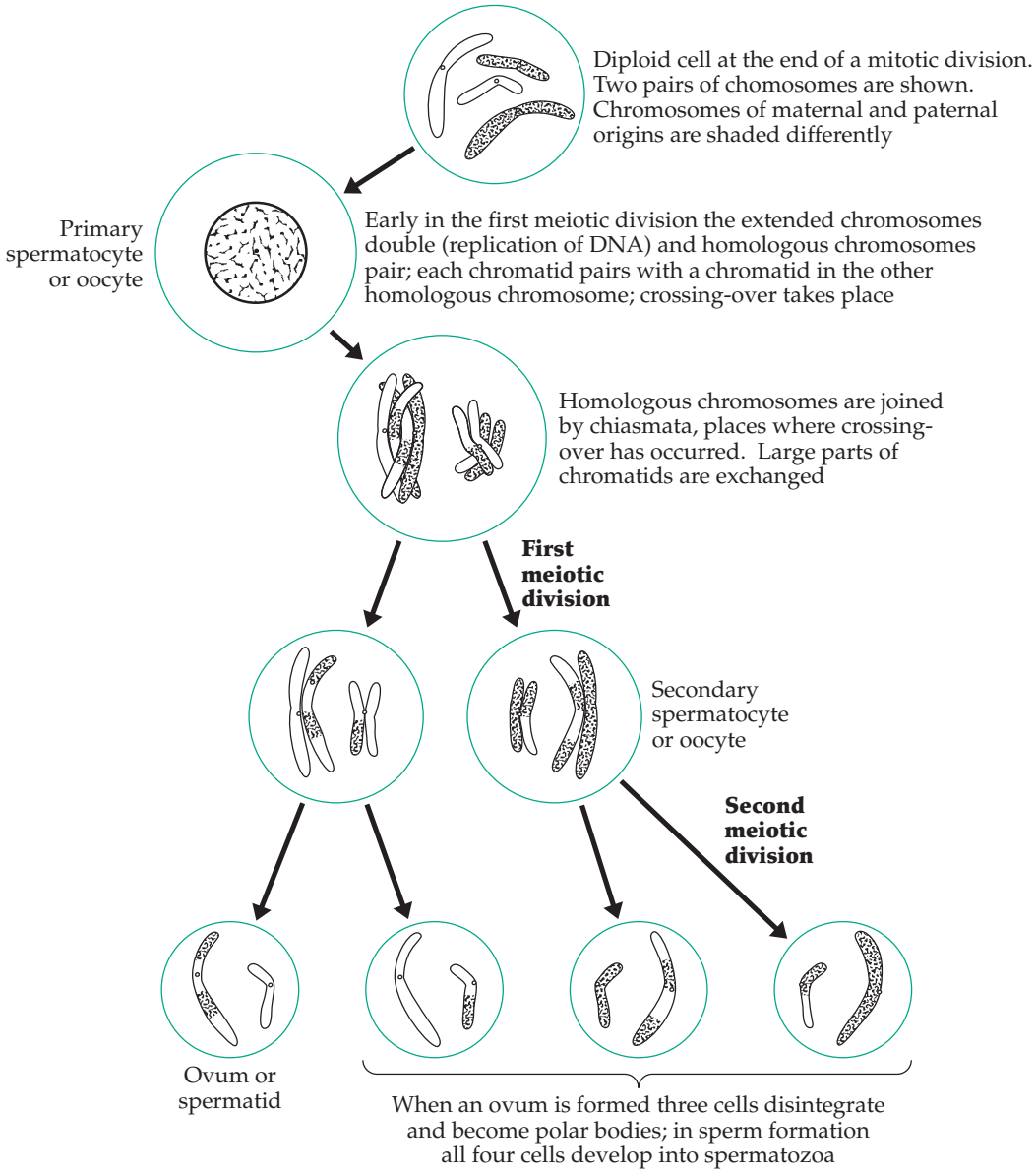


Figure 26-12 Meiosis. Cell division leading to formation of haploid gametes.

vary in time with diploid organisms. Meiosis in males may occur quickly. However, after chromatid pairing and recombination meiotic divisions of oocytes are arrested at the G_2 stage, in some species for many years.[274,275] The arrest is ended by hormonal stimulation, e.g., via progesterone acting through a cyclin B–cdc2 complex.[276] **Meiosis-activating sterols**, intermediates in lanosterol metabolism (Fig. 22-8) also accumulate and are thought to participate in control of meiosis.[276a] In haploid strains of fission yeasts sexual development is induced by starvation, especially of nitrogen. Cells of opposite mating types then fuse to form zygotes, which usually undergo meiosis immediately. Starvation is apparently signaled by the **cyclic AMP–protein kinase A cascade** (Fig. 11-13).[268]

During the prophase of the first meiotic division (meiosis I) two homologous pairs of partially "condensed" chromosomes must find each other and pair with appropriate orientation. A protein in the telomeres of the chromosomes seems to be involved.[269,277] The key structure in meiotic crossing-over is the ribbonlike **synaptonemal complex** formed by the pairs of homologous chromatids.[271,278–279b] This complex, in which a proteinaceous core or **axial element** separates the greatly extended chromatid pairs (Fig. 26-13), is fully formed in the **pachytene stage** of meiosis. Formation of the synaptonemal complex is preceded by development of a few double-stranded breaks in

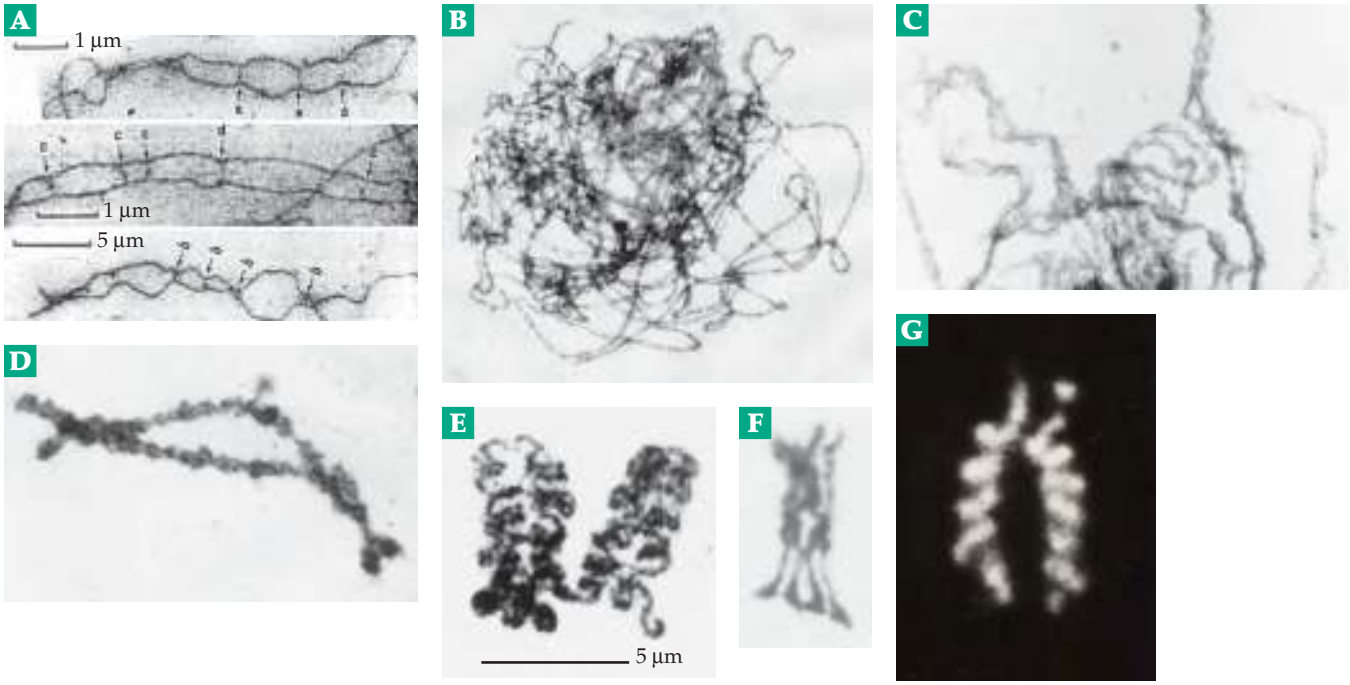


Figure 26-13 Synaptonemal complexes. (A) Aligned pairs of homologous chromatids lying ~0.4 µm apart in *Allium cepa*. Arrows indicate "recombination nodules" which may be involved in initiating formation of crossovers. Portions of meiotic chromosomes of lily are shown at successive stages: (B) Pachytene. (C) Portion of diplotene nucleus. (D) A bivalent at diplotene. (E) Two bivalents at diakinesis. Pairs of sister chromatids are coiled with appropriate handedness. (F) Sister chromatid cores are far apart in preparation for separation. A chiasma is present between the two central strands. (B) through (F) courtesy of Stephen Stack.[279,279d] (G) Pair of sister chromatids coiled with opposite handedness at metaphase. These are immunostained with anti-topoisomerase II antibodies. From Boy de la Tour and Laemmli.[280] Courtesy of U. K. Laemmli.

the DNA.[278,279c] The 3' ends of the DNA chains then invade the homologous chromatids to initiate the exchange via **Holliday junctions** (Fig. 5-28 and Fig. 27-11). The details of the recombination process are considered further in Chapter 27. The first meiotic division is a long process, for example, lasting ten days in mouse spermatocytes.[281] As cells pass to the metaphase stage of meiosis I, the chromosomes become much more compact, but the attachments between the homologous chromatids are still visible as chiasmata (Fig. 26-13). The chromatids then separate, the two homologs appearing as coils of opposite handedness. Within these coils the pairs of sister chromatids continue to be held together at their centromeres through metaphase and anaphase of meiosis I. One or more specific proteins, which must release their hold in meiosis II, are required.[270,277] A leading cause of embryonic and fetal death as well as severe mental and physical problems after birth is incorrect segregation of chromosomes during meiosis. If a gamete contains two copies of any chromosome the embryo will have three. Down syndrome, trisomy of chromosome 21, is the most frequent example. The spindle assembly checkpoint during meiosis I may sometimes be at fault.[281a,b]

3. Polytene Chromosomes

While most cells of higher organisms are normally diploid, the chromosome number may sometimes be doubled or increased even more. A cell with twice the diploid number of chromosomes is **tetraploid,** and with higher multiples of the haploid number it is **polyploid**. Plant breeders have succeeded in producing many tetraploid varieties of flowering plants often with increased size. One tetraploid mammal, the red viscacha rat, is known.[281] While most of our body cells are diploid, we, too, have polyploid cells. For example, some are always found in the liver. The most spectacular example of an increase in the normal DNA content of cells is provided by the giant **polytene chromosomes** of dipteran (fly) larvae. The DNA of cells in the salivary glands and some other parts of these organisms doubles as many as 11–14 times without cell division to give a several thousandfold (i.e.,

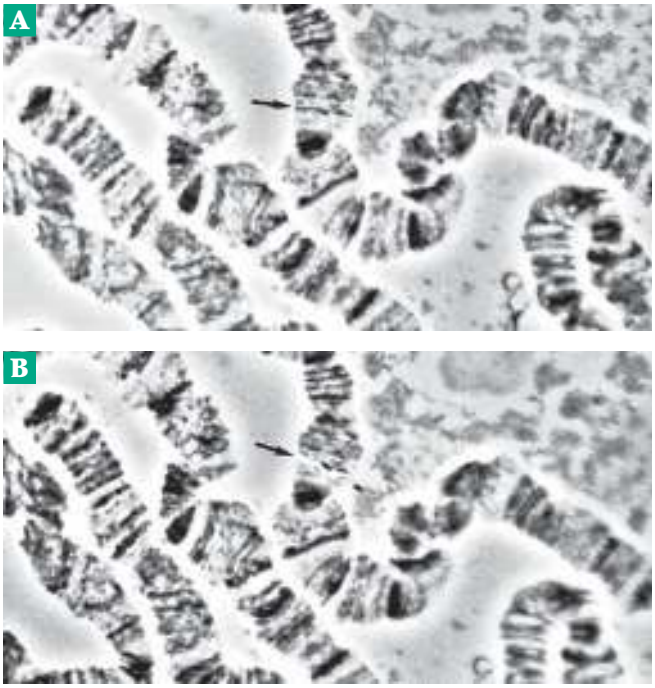


Figure 26-14 Microdissection of a *Drosophila* salivary gland chromosome. (A) Before the cut. (B) After cutting a medium size band. The arrow indicates the site of the cut. From Pirrotta.[282]

2^{14}-fold) increase. The supercoiled, duplicated DNA molecules all line up side by side in a much more extended form than in ordinary chromosomes. They can be seen readily by light microscopy. The total length of the four giant chromosomes of *Drosophila* is ~2 µm, compared to 7.5 cm in diploid cells. The giant chromosomes have a banded structure, ~5000 bands being visible along the length of the chromosome (Fig. 26-14). An average band contains ~36 kb of DNA in each of the strands. Since it has been possible to correlate visible changes in the appearance of these bands with particular mutations in the DNA, study of polytene chromosomes provided a second important method of mapping genes of the fruit fly. The maps produced by the two methods agree well.

Another use of polytene chromosomes is microdissection of DNA for cloning (Fig. 26-14). A piece of DNA containing 100–400 kb can be cut out of any desired spot, cleaved with restriction enzymes, and cloned.[282] Since it has been estimated that *Drosophila* may contain only ~9000–17,000 genes there may be 2–3 genes per band in these chromosomes.[283] The technique has been extended to human and other mammalian chromosomes.[284]

4. Cytoplasmic Inheritance

Not all hereditary traits follow the Mendelian patterns expected for chromosomal genes. Some are inherited directly from the maternal cell because their genes are carried in the cytoplasm rather than the nucleus. There are three known locations for cytoplasmic genes: the mitochondria, the chloroplasts, and certain other membrane-associated sites.[285,286] An example of the last is found in "killer" strains of yeast. Cells with the killer trait release a toxin that kills sensitive cells but are themselves immune. The genes are carried in **double-stranded RNA** rather than DNA, but are otherwise somewhat analogous to the colicin factors of enteric bacteria (Box 8-D). Similar particles (κ factors) are found in *Paramecium*.[287]

Mitochondrial and chloroplast genes are discussed in Chapters 18 and 23, respectively.

G. The Human Genome Project

The discovery of site-specific restriction endonucleases in 1970 and the development of efficient DNA sequencing methods in 1977 sparked a revolution in biology. On Oct 1, 1990, the 15-year project to sequence the complete human genome was officially begun.[288–290] The project is far ahead of schedule. "First drafts" of the genome were published in 2001.[290a–e] A more complete version will include **annotation**, a listing of predicted exons, mRNA transcripts,[290f] functions of recognizable genes, etc.[290e] Such documentation was provided first for chromosome 22[290e] and is now available for other human genes.[290g] It was a surprise to discover, at least initially, that only 30,000–40,000 protein-coding genes could be recognized.[290a,290h] Earlier 50,000–150,000 genes had been predicted (Table 3-1). A nearly complete version of the human genome is anticipated for spring of 2003.[290i]

Included in the genome project are completion of the sequences for *Caenorhabditis elegans*,[290j] *Drosophila melanogaster*, and the mouse (*Mus musculus*).[291,291a–e] However, substantial difficulties remain in filling hard-to-sequence gaps, and some new approaches may need to be developed.[292] More recently the project has been extended to include many additional species. Among them are the rat,[292a] the zebrafish,[292b] the pufferfish *Fugu rubripes*,[292c] the human malaria parasite *Plasmodium falciparum*,[292d,e] the rodent malaria *Plasmodium yoelii, yoelli*,[292f] and the mosquito *Anopheles gambia*, which carries the parasite.[292g,h] Plant genome sequences are also being determined (p. 1511).

1. The Mammalian Genome and Human Health

Human beings and other mammals all have about

3500 Mbp of DNA containing perhaps 40,000 protein-coding genes. However, ~ 97% of the genome is repetitive DNA and other DNA that, at present, has no recognizable function. The genome project will reveal all of these sequences and will doubtless provide us with many surprises.

A primary goal of the genome project is to understand the relationships between gene sequences and human diseases and health. Until recently little was known about the locations or structures of genes in human chromosomes. Human genetic diseases provided the first clues. Although systematic genetic experiments cannot be done with human beings, almost the entire population is under some kind of medical care. Genetic defects are being detected more and more often, and the inheritance within families is traced with increasing frequency. Consequently, a huge body of knowledge of mutations to the human genome is available.[293] By 1995 ~4500 human genetic disorders had been discovered.[294] Many more have been found since then.

The first mapping of human genetic defects came with the recognition that sex-linked traits are encoded on the X-chromosome. Some linkage analysis was also possible from studies of inheritance within families. For example, among individuals who have two X-linked traits, e.g., color blindness plus one other, naturally occurring crossing-over occasionally breaks the linkage in some individuals within the family. **Somatic cell fusion**[191] provided an additional approach. Human lymphocytes can be fused with rodent cells under the influence of inactivated Sendai virus, which causes the cells to adhere and then to fuse. From such human–mouse or human–hamster hybrid cells, strains in which the nuclei have also fused can be selected. Although such cell lines can be propagated for many generations, they tend to lose chromosomes, especially those of human origin. By observing loss of particular human enzymes or other proteins (separable from the hamster enzymes by electrophoresis), it was possible to assign genes to specific chromosomes. This also required identification of the chromosomes lost at each stage in the experiment. New staining techniques made it possible to identify each of the 22 pairs of human autosomes as well as the X and Y chromosomes (Fig. 26-15).[295] Using a variety of techniques, 500 loci in human chromosomes had been mapped by 1983.[296]

Separation of individual human chromosomes with a fluorescence-activated cell sorter (Box 3-B)[299] permitted the preparation of libraries of cloned DNA from individual chromosomes. Fragments, of an average length of 4 kbp, from digestion by *Eco*R I or *Hin*dIII were packaged into the Charon 21A cloning vehicles and "amplified" by culturing infectious phage particles in *E. coli*.[282] However, this approach was inadequate for sequencing the entire genome. By 1987 the use of yeast artificial chromosomes (YACs)[300] and later bacterial artificial chromosomes (BACs)[301,302] and radiation hybrid mapping[104,303] provided the means for sequencing the 34- to 260-Mbp human chromosomes.[288,304] Also essential was the development of high-speed automatic sequencing machines[305–305b] and of computers adequate to assist in compiling the sequences.[302,306,306a] Sets of overlapping and redundant clones define a continuous sequenced segment that is called a **contig**. As the contigs are enlarged the gaps are filled.

In 1987, the first global genetic linkage map of the human genome was published.[288,307] It was based on 403 polymorphic loci, of which 393 were RFLPs studied in 21 three-generation families. By 1996 a linkage map containing 5264 loci, in the form of polymorphism in **short tandem** repeats $(AC/TG)_n$ represented by **microsatellite DNA**, was available.[304] Also mapped by 1998 were 2227 **single nucleotide polymorphisms** (SNPs), where the two alternative bases occur with a frequency >1%.[308] This was increased to 1.4 million by 2001.[308a] As many as 5.3 million common SNPs may occur, each with a frequency of 10–50%. On the average there may be one SNP for every 600 base pairs.[308b] These SNPs account for a large fraction of the diversity in human DNA. By 1995 a physical map with >15,000 **sequence tagged sites** (STSs) with an average spacing of 199 kb had been constructed.[303,309] An STS is defined as a cloned fragment in a YAC (or BAC) library that has been amplified by PCR and tested to establish that it contains a known locus. DNA sequences known as **expressed sequence tags** have also been mapped (see p. 1490). Using STSs, RFLPs, ESTs, and the growing contigs, more than 16,000 human genes had been mapped by 1996.[310] By 1998 more than 1,060,000 ESTs had been reported.[311] In 1999, the nucleotide sequence of the smallest human chromosome, the 33.4-Mbp chromosome 22, was completed. It contains at least 545 genes and 134 pseudogenes.[312] In the entire genome >1000 seed contigs had been assembled, and completion of the first phase of the Human Genome Project was in sight.[313] In 2001 a whole-genome clone-based physical map was published.[313a]

2. Understanding Gene Sequences

The vast array of sequence data coming from the human genome and from genes of other species are deposited, as they are reported in the scientific literature, in the Human Genome Central. It can be found on the Web at the National Center for Biotechnology Information (NCBI) at http://www.ncbi.nlm.nih.gov/genome/guide and European Bioinformatics Institute (EBI) at http://www.ensembl.org/genome/central/. The data have been doubling every 15 months. By

2000 almost 1.3 million human ESTs were included as were sequence data on more than 25,000 other species.[311,314] A problem is that there are many errors in the data with some seriously incorrect sequences.[315] It is hard to manage the mass of new data, but in time most of the errors will probably be corrected. Books[316–319] and computer programs[320,321–321c] are available to help understand genomes. The widely used programs **BLAST** and **FASTA**, available on the Web or in computational packages, routinely compare protein-coding genes with known genes in order to predict function.[322–325] An inexpensive high-powered desktop computer and an internet connection will enable a person to do complex biocomputing.[325a] See also Chapter 5, Section H,7.

Genes must be transcribed, and most transcripts must be spliced, modified, and translated by the ribosomal machinery. Genes cannot be fully understood without considering these processes, which are dealt with in Chapters 28 and 29.

Human variation. Between any two unrelated human beings there are on the average one base pair that is different out of every 500–1000 nucleotides. This amounts to $\sim 4 \times 10^6$ differences in the whole genome.[305,326] In addition to these single nucleotide polymorphisms (SNPs) there are many differences in the >100,000 nearly randomly dispersed short tandem repeats of microsatellite DNA.[189,327] The latter form the basis of DNA fingerprinting (Box 5-D), and together with the SNPs are helpful in tracking genetic diseases. One of the more difficult goals is to identify genes that confer increased susceptibility to cancer and to the complex syndromes of diabetes and mental problems. Searching for correlations with SNPs has proved difficult.[327a,b] A new approach termed **haplotype mapping** may permit correlations of disease susceptibility with larger blocks of conserved DNA sequence known as haplotypes.[327c,d] Some controversy surrounds possible uses of these maps, as also is the case for proposals to conduct extensive genetic testing of populations including newborn infants.[327e,f,g] Another planned project is the sequencing of 1000 individual human genomes.[327h]

Studies of DNA have also shed light on human evolution.[328] Mitochondrial DNA,[329] as well as a variety of nuclear genes,[330] is being studied in attempts to establish approximate dates of evolutionary divergence. If we assume a constant rate of incorporation of nucleotide substitutions during evolution, we can use sequence data as a **molecular clock** to construct phylogenetic trees such as that of Fig. 1-5, which is based on the gene for 16S ribosomal RNA. A problem is that mutation rates of genes are not all the same, and there are not yet enough data to draw firm conclusions. Estimates based upon the maternally transmitted mitochondrial DNA (see Chapter 18) suggest that we are all descended from Africans who lived 100,000 to 200,000 years ago. Limited data from the Y chromosome, which is transmitted through males, agree with the mitochondrial DNA data.[328,328a] However, study of the pyruvate dehydrogenase gene has been interpreted to indicate a more ancient divergence of African and other peoples.[331] Analysis of DNA from a fossil bone shows that the mtDNA of Neandertal people differed from modern mtDNAs in an average of 26 positions. From this figure the mitochondrial DNA molecular clock predicts that the extinct Neandertal line diverged from ours ~550,000 to 690,000 years ago. The most ancient mtDNA found is Australian.[331a,b] Unfortunately, DNA in fossils is very unstable and has not been recovered from older fossils, e.g., from dinosaurs. However, the search for ancient DNA goes on.[331c,d]

We diverged from our closest ape relative, the chimpanzee, about 4–6 million years ago.[196,332] Chimpanzee and bonobo DNA sequences are ~98.8% identical to those of humans.[332a–e] If differences in inserted and deleted segments of genes are included, however, the identity drops to 95%.[332f] One of the longest DNA sequences to be compared among humans and apes is a 7-kb length around the pseudogene in the β globin cluster (Fig. 27-10). In this sequence chimpanzees are closest to humans with gorillas being the next closest.[195] We may well ask in what way we differ from these apes? Some specific differences have been found. Notably, human beings do not hydroxylate the glycolyl groups of sialic acids (Chapter 4) to form *N*-glycolylneuraminic acid residues on glycoproteins.[328,333] Could this really be the most important difference between us and the apes? More genomic analysis may tell.

What does DNA analysis tell us about race? Most investigators conclude that there is only one human race with no detectable boundaries between the group commonly referred to as races.[334–336] As Pääbo put it, "in terms of the variation at most loci, we all seem to be Africans, either living on that continent or in recent exile."[328] The differences in skin color seem to reflect adaptation to the environment in which people live.[336a] Variations in the level of melanocyte-stimulating hormone receptor, one regulator of skin color (Box 25-A), are especially high in Africans.[328]

DNA analysis has also been useful in tracing human migration.[337–338e] For example, a genetic marker in the Y chromosome is carried by 85% of native Americans, suggesting that they are all descended from a man who lived ~20,000 years ago, probably an immigrant from Siberia.[339] Contrary to usual assumptions women, more often than men, seem to have spread their DNA to new locations in the world.[327b] Studies of cattle and of the wild ox reveal information about domestification of these animals about 10,000 years ago in Europe, Asia, and Africa.[340]

Other evolutionary relationships. Studies of chromosome banding, chromosome maps, restriction fragments, and detailed sequences provide many insights into relationships among species. For example, the chromosome banding pattern and also DNA sequences show close similarities between human beings and the mouse.[194,312] The latter is often regarded as the premier organism for the study of mammalian genetics and development.[291a–c,341] Dense genetic maps are available for both the mouse and rat as are moderate-resolution maps for livestock, companion animals, and other mammals.[327,342,343] Comparative gene maps are being constructed for more than 40 mammals[344] and other species of animals, plants, and fungi. Comparisons of these genomes reveal much of interest. For example, the pufferfish *Fugu rupripes* has a genome only 1/9 the size of the human genome. However, both species seem to contain about the same number of genes. Many of them can be directly correlated and some human-disease-causing mutations have been identified first in the Fugu.[344a]

Evolutionary history is being rewritten in molecular terms. Comparison of sequences of individual proteins allows evolutionary relationships of their genes to be traced. Many families of **homologous** genes can be identified.[344b] These include both **orthologs**, genes in different species that have evolved from a common ancestor, and **paralogs**, genes related by gene duplication within a genome. Orthologs have the same function in different organisms, but paralogs have different functions within a single species.[345]

Sophisticated molecular clock studies suggest that the evolution of metazoan organisms began earlier than had been supposed. Ancestral primates appeared at least 65 million years ago.[344] Gene sequence data for many species suggest that a great variety of mammals lived 100 million years ago in the age of dinosaurs,[346] a view also supported by new fossil evidence.[344,347] Metazoans appeared earlier than the "Cambrian explosion" generally thought to have occurred ~550 million years ago.[348] New geochemical data suggest that cyanobacteria diverged from other bacteria as early as 2.1×10^9 years ago.[349] Gram-negative bacteria diverged from gram-positive microbes ~3.2×10^9 years ago,[350] *Salmonella* from *Escherichia* only 0.1×10^9 years ago. However, DNA analysis shows that within this latter time period many genes from other microorganisms have been inserted into the *E. coli* chromosome and into other bacterial chromosomes.[351,352] Some of these transfers have occurred with the help of bacteriophages.[353] A puzzle is the fact that among eukaryotic cells the enzymes catalyzing the genetic **information transfer** via transcription and protein synthesis resemble those of archaea. However, **operational enzymes** that catalyze other basic metabolic processes tend to be more similar to those of eubacteria such as *E. coli*.[353,354] So much gene transfer between organisms has occurred that it is difficult to establish the earliest parts of a phylogenetic tree of the type shown in Fig. 1-5. Another factor that confuses our study of bacteria is that less than 1% of all living microorganisms have been grown in pure cultures.[355,356]

We are still dangerously ignorant of the complexity of the microbial world, which both threatens us with diseases and sustains our environment. We do have complete genome sequences for more than 60 different bacteria with hundreds more expected within a few years.[356,356a] They contain from 0.58 Mbp (*Mycoplasma genitalium*) to 8.7 Mbp for the antibiotic-producing *Streptomyces coelicolor*,[356b] a size shared by the legume symbiont *Bradyzhizobium japonicum*.[356c] The latter is one of many bacteria that have genomes split into two or more parts, often a major chromosome plus one or more plasmids. For example, the 6.7-Mbp genome of *Sinorhizobium melitoti*, an alfalfa symbiont, comprises a 3.65-Mbp chromosome and 1.35- and 1.68-Mbp megaplasmids.[356d] The 5.67-Mbp genome of *Agrobacterium tumefaciens*, much used in genetic engineering (Section E,4), consists of a circular chromosome, a linear chromosome, and two plasmids.[146a–c] The genome sequence revealed a close similarity to those of the above-mentioned legume symbionts.

Sequences of many pathogenic bacteria are known.[356e–o] These include the causative agents of cholera,[356c] typhoid fever,[356f] plague,[356g] brucellosis,[356h] leprosy,[356i] tuberculosis,[356i] and anthrax.[356j] Also included is a virulent strain of *Streptococcus pneumonia* (respiratory infections, ear aches, meningitis),[356k] *Pseudomonas aeruginosa* (a common "opportunistic" pathogen),[356l] *Listeria monocytogenes*, which causes a severe food-borne disease (Box 19-C),[356m] the pathogenic *E. coli* 0157:H7 (see Fig. 1-2),[356n] and a tiny mucosal pathogen *Ureaplasma urealyticum*.[256o] Many surprises were found in the genome sequences. For example, *E. coli* 0157:H7 is related to the laboratory strain *E. coli* K-12, and the two strains share a ~4.1 Mbp common "backbone" sequence. However, the pathogenic strain has hundreds of "islands" of additional DNA spread throughout the genome and amounting to ~1.54 Mbp. Many of these carry genes associated with virulence.[356n]

Specialized features appear in virtually every genome.[356a] For example, the mycoplasma *U. urealyticum* contains genes for enzymes that allow the bacterium to obtain almost all of its ATP from hydrolysis of urea.[356o] Both *Caulobacter crescentus*[356p] and *E. coli* develop specialized structures involved in motility (Figs. 19-3, 32-1). *C. crescentus* and also spore-forming bacteria have alternative developmental plans (Fig. 32-1). At the low end of the genome size range are species of *Buchnera*, which are endocellular symbionts of aphids.[356q] Their genome is only slightly larger than

that of *M. genitalium*. The extremely salt-tolerant *Halobacterium*[356r] and the heat-tolerant *Thermoplasma volcanium*[356s] are among the archaea for which complete genome sequences are known. Species of *Xanthomonas*, whose genomes have been sequenced, are economically important plant pathogens. *Xanthomonas campestris* is also grown commercially to produce xanthan gum (p. 179).[358a] Determination of the genomes of mycobacteria has been challenging.[356i,t] Among these slow-growing organisms is *Mycobacterium tuberculosis*, the causative agent of human tuberculosis. A large fraction of its ~4000 coding genes is devoted to metabolism of lipids and to the synthesis of unusual proteins and lipids of its cell wall.[356t]

The larger genomes code for many proteins of unknown function, but over 80% of the "ORFs" (presumed genes) of *Haemophilus influenzae* have been identified, as have their presumed functions.[357,358] The encoded proteins appear to catalyze 488 metabolic reactions on 343 different metabolites. Together these systems provide a **metabolic genotype**.[357] Results of such analyses are accumulating in **metabolic databases**.[359,359a] What is the minimum number of metabolic reactions needed for support of life? Transposon insertions can inactivate all but from 265 to 350 of the protein-encoding genes of *M. genitalium* without killing the bacteria.[360] From comparisons of a variety of bacterial chromosomes it seems likely that ~256 of these genes are truly essential.[350,361] This conclusion leads to an interesting question. Is it ethical to now try to generate such a minimal bacterium?[361a] Are there hazards, e.g., that its escape might endanger our health or the environment? On the other hand, genetic engineering of bacteria, which is already practiced, can provide useful improvements in bacteria used in foods and in industry.[362,362a]

Modeling with the aid of data available on the World Wide Web is leading to development of new mathematical descriptions of metabolic networks.[359a] An ambitious new project is to model the entire *E. coli* cell. Many experimental data will be required and it has been estimated that ten years will be needed. The effort involves investigators in many laboratories and will be at least ten times as complex as the determination of the human genome.[362b]

Metabolic studies of eukaryotic cells. The yeast, *Saccharomyces cerevisiae*, contains ~6200 genes of which, until recently, only 40% had been assigned a function. Now a variety of methods are being employed to understand this little fungus.[363–365] A useful approach is to systematically inactivate or "knock out" genes. Davis and associates[363] used a PCR-based strategy to delete one gene at a time of 2026 yeast genes. Of these genes 1620 were found not essential for growth in a rich medium. Ross-McDonald and coworkers engineered a transposon, a 274-bp derivative of the *E. coli* Tn3, and allowed it to be inserted into the genes of yeast cells by homologous recombination (Chapter 27, Section D). The transposon carried DNA for a short peptide tag in the form of a specific immunological epitope that could identify the transformed cells. More than 11,000 strains with disruptions in nearly 2000 genes were obtained. These and other deletion mutants are now available for study.[364a,365] A second yeast, *Schizosaccharomyces pombe*, has assumed major scientific importance in studies of the cell cycle and of metabolism.[365a] Its 13.8-Mbp genome is only a little smaller than that of *S. cerevisiae*, but it has ~1400 fewer recognized genes, a total of 4824.[365b] There are smaller eukaryotic genomes. That of the tiny marine chordate *Oikopleura dioica* may be only 51 Mbp. However, there may be a total of ~15,000 genes.[365c]

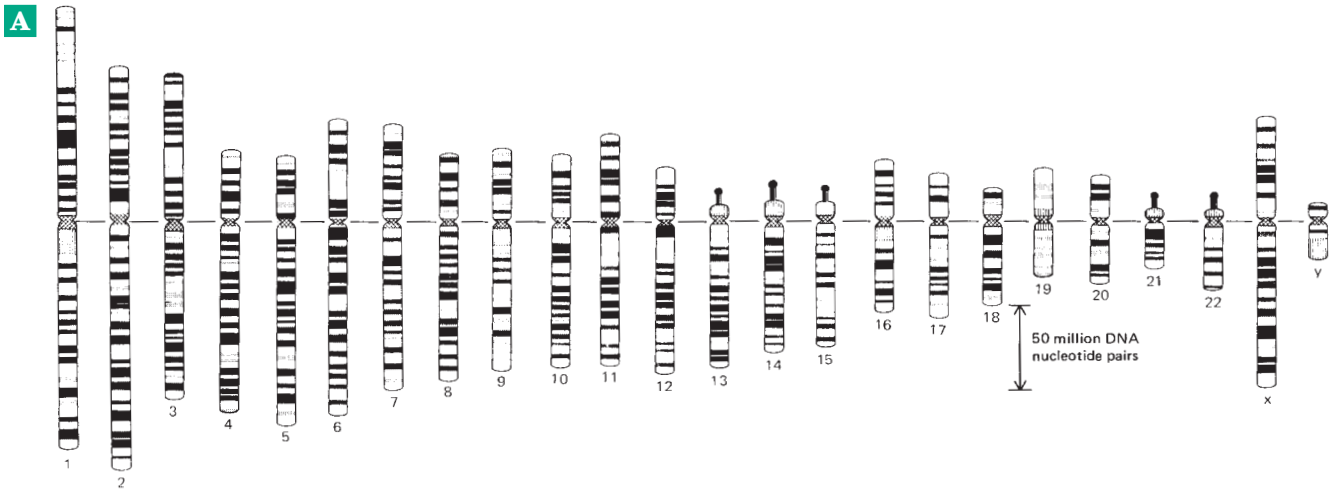
Transposon-induced mutations have also been created in nearly one-fourth of the ~12,000 genes of *Drosophila melanogaster*.[283] Studies of the expression of genes in both *Drosophila* and in the nematode *Caenorhabditis elegans* are directed toward understanding of development and differentiation. Of the nematode's predicted 19,293 genes, only 7% have been studied at the biochemical level. To understand what happens during development we need methods for studying simultaneously the expression of all of these genes. One approach is to look at messenger RNAs that are formed at different times during development. More than 9000 mRNAs have been identified in cells of *C. elegans*, and their patterns of expression have been observed using DNA microarrays.[366,367] A similar technique has been applied to yeast. DNA sequences of fragments of ~6400 yeast genes were amplified by PCR and printed onto a glass plate to form a "DNA chip." From mRNAs formed at different times during growth, fluorescent cDNA copies were made, and their amounts were checked by use of the DNA chip.[368] Another approach is to look directly at the proteins formed. Walhout and coworkers have devised a large-scale automated system for cloning all of the genes of *C. elegans*, expressing the protein products, and testing them in the yeast two-hybrid system (Box 29-F) for protein–protein interactions.[369] Another project is to use large-scale sequence comparisons between proteins of *C. elegans* and of other organisms to identify nematode genes that encode extracellular matrix proteins involved in cell adhesion and to trace their evolution.[370]

The genomes and the metabolism of the two insects *Drosophila melanogaster* and *Anopheles gambia*[370a] can now be compared. Many differences can be seen but almost half of the genes are orthologs. Many of these can be related also to those of pufferfish, mice,[370b] humans, and other species.

Among plant genomes that of *Arabidopsis thaliana* has been studied most. The sequences of its five chromosomes have been determined and analyzed.[371–375c]

About 25,498 genes encoding proteins from 11 different families have been found.[375a] More than 14,000 ESTs were established from cDNAs.[375b] Many of the genes represent new families, some of which may be peculiar to plants. However, many others are homologous to those of *C. elegans* and *H. sapiens*. For example, developmentally important homeotic genes, marked by a **homeo sequence** as in animal genes, are present as are thousands of cell surface receptors. However, only Ser / Thr and histidine kinases are present in *Arabidopsis*. No tyrosine kinases have been identified.[375]

The ~125-Mbp *Arabidopsis* genome is tiny compared to the 3000 Mbp of DNA present in the genomes of maize and of many other plants.[374] However, the rice genome, only ~15% as large as that of maize,[376] has been chosen for complete sequencing.[375d,e] Draft sequences for the genomes of two subspecies of rice (*Oryza sativa L.*), ssp. *indica*[375f] and ssp. *japonica*,[375g] were published in 2002. The much larger genomes of other cereal grains (2500, 4900, and 16,000 for maize, barley, and wheat, respectively) will probably all be sequenced within a few years. The genes of maize are being mapped and studied using transposon-tagging



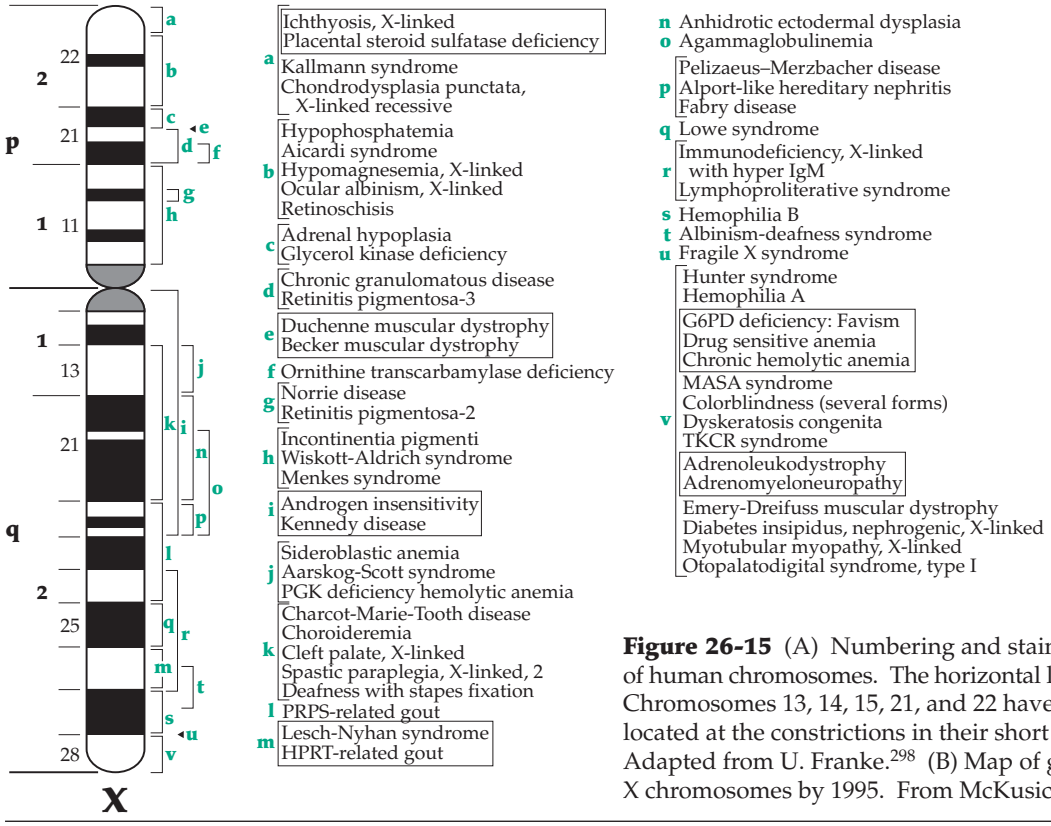


Figure 26-15 (A) Numbering and staining patterns (G-banding) of human chromosomes. The horizontal line marks the centromeres. Chromosomes 13, 14, 15, 21, and 22 have nucleolar organizers located at the constrictions in their short arms. From Alberts *et al.*[297] Adapted from U. Franke.[298] (B) Map of genetic defects identified in X chromosomes by 1995. From McKusick and Amberger.[294]

methods.[377] The smaller, more compact genomes may have just as many genes as the longer ones but have less repetitive DNA.[378] Some ferns have 307,000 Mbp of DNA, nearly 100 times that of a human; bony fishes have ~307,000 Mbp and the ameba >200,000 Mbp. These organisms appear to lose unneeded repetitive DNA faster than those with smaller genomes.[378]

3. Understanding Human Genetic Diseases

Genetic diseases have always been with us, but it was not until 1949 that the first disease, sickle cell anemia (Box 7-B), was understood at the molecular level. A single base substitution in DNA and the resultant single amino acid substitution in hemoglobin causes this disastrous disease. It was soon recognized that defects in single proteins are causes of other inherited diseases. Many of the hundreds of other known genetic disorders[201,203,379] are discussed elsewhere in this book. Among them are muscular dystrophies and cardiomyopathies (Box 19-A),[380] lysosomal deficiency diseases (Chapter 22), problems with ion transporters[381–383] and channels (cystic fibrosis, Box 26-A), defective collagens (Box 8-E),[384] neurological disorders (Chapter 30), and defects in defense systems (X-linked granulomatous disease, Chapter 18, Section G). Many of these were first recognized by their frequent occurrence in boys. Some of these X-linked deficiencies are mapped in Fig. 26-15B.

One insight into molecular disease was the recognition that mutations that cause many diseases, e.g.,

BOX 26-A CYSTIC FIBROSIS

One of the commonest of genetic diseases, cystic fibrosis affects persons all over the world. The incidence is unusually high in persons of northern European descent, about one in 2500 children being born with the defect. The inheritance pattern showed that cystic fibrosis is recessive and is caused by a single-gene defect that is carried by almost 5% of white Americans. In the United States there are ~30,000 persons with the disease. Many die in early childhood and even with careful treatment only 50% live into their late twenties or beyond.[a,b] Through extensive linkage analysis the cystic fibrosis gene was mapped to chromosome 7, and its location was narrowed further to a 1600-kbp region between the oncogene *met* and another marker designated J3.11.[c] Random searching located closer markers, and "chromosome walking and jumping" led to identification and characterization of the gene in 1989.[c,d] The large 250-kbp gene contains 27 exons. The transcribed mRNA is 6129 bp in length, and the gene product is a 1480-residue amino acid protein,[e] which is known as the **cystic fibrosis transmembrane conductance regulator** (**CFTR**).[f]

Children with cystic fibrosis lose excessive amounts of salt in perspiration and become dehydrated readily. A salty taste of the skin and an elevated chloride concentration of sweat are traditional diagnostic symptoms.[a] More serious problems arise from progressive respiratory failure and inadequate pancreatic secretion. Lung infections with *Pseudomonas aeruginosa* are the major cause of death.[g] The CFTR gene is expressed in many tissues, especially those of the mucous membranes. An alternatively spliced isoform may form chloride channels in heart.[h,i] As mentioned in Chapter 8, Section C,5 the CFTR protein is a member of the ATP-dependent ABC transporter family. However, it is atypical because it also contains a regulated chloride channel.[j] In secretory epithelia of intestines, pancreas, lungs, sweat glands, and kidneys Cl^- enters epithelial cells through their basolateral surfaces using an $Na^+ + K^+ + 2\ Cl^-$ cotransporter and exits the cells through their apical surfaces using the CFTR channel. Absorptive epithelia also contain both the cotransporter and the CFTR channel, but Cl^- flows into the cells from the exterior surface, and the distribution of the cotransporter and CFTR between basolateral and apical surfaces is opposite to that in secretory cells.[f]

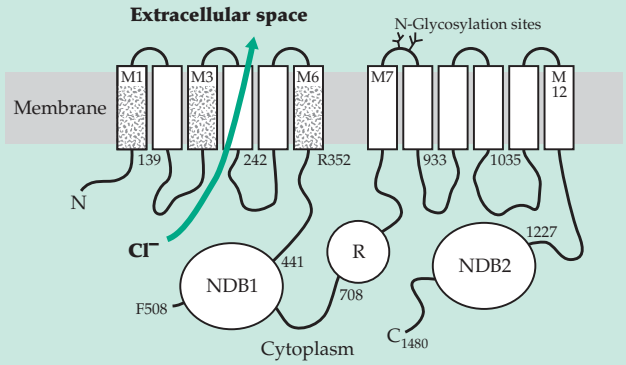
From the amino acid sequence the CFTR protein is predicted to form two 6-helix transmembrane domains, two ~240-residue cytosolic ATP-binding domains, and a cytosolic regulatory (R) domain that contains at least five serine residues that may be phosphorylated by the cAMP-dependent **protein kinase A** (Chapter 11, Section C; Fig. 11-4).[k,l,m] See adjacent scheme. The two ATP-binding domains resemble those of myosin and other ATP-hydrolyzing proteins. The chloride channel, which is probably formed by helices M1, M3, and M6,[f] remains closed unless serine residues of the regulatory domain are phosphorylated. However, opening of the channels also requires binding of ATP in the nucleotide-binding domains. Binding of either vanadate or BeF_x stabilizes the open state of the channels.[k] There are four small 55- to 65-residue cytosolic loops, labeled 1– 4 in the diagram. Transmembrane loop 3 also appears to function in regulation of the channel,[n] which involves a complex regulatory mechanism. Two sites of *N*-linked glycosylation are present in the short extracellular loop between helices M7 and M8.

cystic fibrosis, while affecting a single protein, occur at many places in the gene that encodes the protein. Not all mutations are caused by base substitutions; they are often a result of deletion or insertion of DNA. A whole group of diseases are caused by accumulation of repetitive DNA, often of nucleotide triplets, within genes or in control regions of genes. Another important insight is that understanding a newly discovered and very rare disease may help us to understand other related disorders. For example, after the Duchenne muscular dystrophy gene was located, the encoded protein dystrophin was found to have mutations resulting in other milder dystrophies (Box 19-A).

Mutations are only rarely beneficial, but we know that many mutations alter the properties of proteins very little. We can anticipate that most genes may undergo mutations that cause some loss of good health and vitality without being diagnosed as causing disease.

We have also come to understand that many complex diseases such as diabetes, **polycystic ovary syndrome**,[385] **Crohn's disease** (inflammatory bowel disease),[386] and schizophrenia are in fact multiple diseases. Diabetes is a syndrome that can arise from causes such as defective insulin receptors or defective glucose transporters or from as yet unknown metabolic problems (Chapter 17).[387] Many cancers have a

BOX 26-A CYSTIC FIBROSIS (continued)



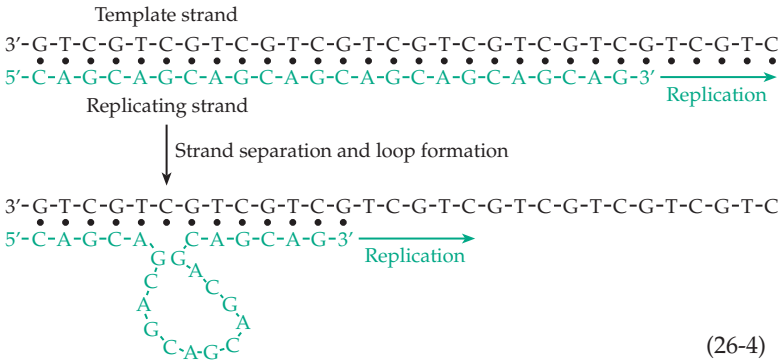
The CFTR protein can undergo endocytosis into clathrin-coated vesicles as part of its regulatory mechanism.[o] Since HCO_3^- is usually exchanged for Cl^- in epithelial ion transport, regulation of HCO_3^- uptake is also a significant aspect of CFTR function.[p]

Mutations that cause cystic fibrosis are found at many locations in the gene. However, ~70% of the mutations are caused by the absence of phenylalanine 508, as a result of a three-nucleotide deletion, in the first nucleotide-binding domain.[e,k,p] This deletion causes misfolding of the CFTR to give an inactive protein.[p] Hundreds of other mutations, some in the regulatory domain and some in the cytosolic loops,[q,r] also cause the disease. Cystic fibrosis is one of the diseases for which targeted gene transfer may become an effective treatment.[a,s] Cystic fibrosis induced in mice by targeted disruption of the CFTR gene has been successfully treated by gene therapy.[t] However, the gene is large and efficiency of transfer into animals is low. Nevertheless, human gene therapy for cystic fibrosis is being pursued cautiously.[s]

The CFTR protein has additional medical significance. Its stimulation by bacterial toxins is responsible for **secretory diarrhea**, which kills 3 million young children annually.[f]

[a] Welsh, M. J., and Smith, A. E. (1995) *Sci. Am.* **273**(Dec), 52–59
[b] Koh, J., Sferra, T. J., and Collins, F. S. (1993) *J. Biol. Chem.* **268**, 15912–15921
[c] Marx, J. L. (1989) *Science* **245**, 923–925
[d] Riordan, J. R., Rommens, J. M., Kerem, B.-s, Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M. L., Iannuzzi, M. C., Collins, F. S., and Tsui, L.-C. (1989) *Science* **245**, 1066–1073
[e] Collins, F. S. (1992) *Science* **256**, 774–779
[f] Akabas, M. H. (2000) *J. Biol. Chem.* **275**, 3729–3732
[g] Ernst, R. K., Yi, E. C., Guo, L., Lim, K. B., Burns, J. L., Hackett, M., and Miller, S. I. (1999) *Science* **286**, 1561–1565
[h] Hart, P., Warth, J. D., Levesque, P. C., Collier, M. L., Geary, Y., Horowitz, B., and Hume, J. R. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6343–6348
[i] Howell, J. M., and Gawthorne, J. M., eds. (1987) *Copper in Animals and Man*, Vol. I and II, CRC Press, Boca, Raton,Florida
[j] Li, C., Ramjeesingh, M., Wang, W., Garami, E., Hewryk, M., Lee, D., Rommens, J. M., Galley, K., and Bear, C. E. (1996) *J. Biol. Chem.* **271**, 28463–28468
[k] Gadsby, D. C., and Nairn, A. C. (1994) *Trends Biochem. Sci.* **19**, 513–518
[l] Ikuma, M., and Welsh, M. J. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8675–8680
[m] Ostedgaard, L. S., Baldursson, O., and Welsh, M. J. (2001) *J. Biol. Chem.* **276**, 7689–7692
[n] Aleksandrov, L., Mengos, A., Chang, X.-b, Aleksandrov, A., and Riordan, J. R. (2001) *J. Biol. Chem.* **276**, 12918–12923
[o] Weixel, K. M., and Bradbury, N. A. (2000) *J. Biol. Chem.* **275**, 3655–3660
[p] Luo, X., Choi, J. Y., Ko, S. B. H., Pushkin, A., Kurtz, I., Ahn, W., Lee, M. G., and Muallem, S. (2001) *J. Biol. Chem.* **276**, 9808–9816
[q] Seibert, F. S., Jia, Y., Mathews, C. J., Hanrahan, J. W., Riordan, J. R., Loo, T. W., and Clarke, D. M. (1997) *Biochemistry* **36**, 11966–11974
[r] Cotten, J. F., Ostedgaard, L. S., Carson, M. R., and Welsh, M. J. (1996) *J. Biol. Chem.* **271**, 21279–21284
[s] Verma, I. M., and Somia, N. (1997) *Nature (London)* **389**, 239–242
[t] Hyde, S. C., Gill, D. R., Higgins, C. F., Trezise, A. E. O., MacVinish, L. J., Cuthbert, A. W., Ratcliff, R., Evans, M. J., and Colledge, W. H. (1993) *Nature (London)* **362**, 250–255

strong hereditary component, and many of these are being mapped to specific DNA locations.[388] Some specific proteins, such as **Ras** (Chapter 11) that are mutated in many cancers were first recognized as avian oncogenes. The tumor suppressors **Rb** (retinoblastoma protein, Box 11-D)[388,389] and **p53** (Fig. 11-15)[390,391] are major sites of mutation in cancer. See also Chapters 11 and 31. Since cancers contain multiple mutations, they are complex diseases. However, many specific susceptibility loci are being located, including some for breast cancer (Box 11-D),[392] prostate cancer,[393] and familial adenomatous polyposis, a hereditary disease leading to colon cancer.[388,394] Cancer has long been known to be associated with chromosome instability including deletion and insertion mutations at simple repeat sequences, frame-shift mutations,[395] DNA breakage, translocation,[396] and losses or gains of whole chromosomes.[397]



(26-4)

From a practical viewpoint the understanding that we are gaining will help us to provide better treatments of genetic diseases. At present almost every human genetic disease can be mimicked in a knockout mouse.[291a,398] A turnabout is that rare human hair-

TABLE 26-4
Some Human Triplet Repeat and Related Diseases[a]

Name	Repeat sequence (5' to 3')	Transmission[b]	Location
Fragile X syndrome (FRAXA)	$(CGG)_n$	XD	by 5' side of gene *FMR*-1 + C_pG island methylation
Type E (FRAXE)	$(CGG)_n$	XR	5' side of gene *FMR*-2
Synpolydactyly	$(GCG)_n$		HOXD13 gene (polyalanine)
Myotonic dystrophy (DM1)	$(CTG)_n$	AD	3' untranslated region of gene for cAMP-dependent protein kinase
DM2	$(CCTG)_n$	AD	Intron in a zinc finger protein gene
Friedreich ataxia	$(GAA)_n$	AR	Intron in frataxin, a mitochondrial protein
Huntington disease (HD)	$(CAG)_n$	AD	Huntingtin (polyglutamine)
Spinocerebellar ataxia, SCA-1	$(CAG)_n$	AD	Ataxin-1 (polyglutamine)
SCA-2	$(CAG)_n$	AD	Ataxin-2 (polyglutamine)
Machado–Joseph disease (SCA-3)	$(CAG)_n$	AD	Ataxin-3
SCA-6	$(CAG)_n$	AD	Calcium channel
Spinobulbar muscular atrophy (Kennedy disease; SBMA)	$(CAG)_n$	XR	Atrophin (androgen receptor)
Dentato-rubro-pallido-luysian atrophy (DRPLA)	$(CAG)_n$	AD	
Progressive myoclonus epilepsy	(G+C)-rich oligonucleotide repeat		Cystatin B

[a] See Mandel, J.-L. (1997) *Nature (London)* **386**, 767–769 and Richards, R. I., and Sutherland, G. R. (1997) *Trends Biochem. Sci.* **22**, 432–436
[b] XD, X dominant; XR, X recessive; AD, autosomal dominant; AR, autosomal recessive

less people have a mutation in a gene homologous to the well-known *hairless* gene of mice.[399] Genes homologous to those of many human genetic defects have also been identified in yeast.[400] Studies of both mutant mice and of mutant yeast cells can also help in understanding diseases and in devising therapies.

Triplet repeat diseases. With the exception of Down disease (extra chromosome 21, affecting 1 in 600 children) the most prevalent cause of mental retardation is the **fragile X syndrome,** which affects ~1 of 2000 newborn males. A fragile site,[401,402] where the X chromosome breaks easily, marks the location of the defect. Identification of nearby RFLPs led to cloning of the gene in 1991. The defect was found to be a repeated trinucleotide sequence 5'-$(CGG)_n$ in the DNA. The value of n varies, in most healthy individuals averaging ~30. However, for some normal individuals n may be 200–300. These persons may transmit the fragile X disease to their offspring in whom n increases from one generation to the next with increasing severity of the disease. There may be 2000 or more CGG triplets.[403] The defect lies at the 5' end in an untranslated part of the gene for the fragile X mental retardation protein (FMRP). A cytoplasmic RNA-binding protein, FMRP, may enter the nucleus and have an as yet uncertain function. In rare cases the fragile X syndrome arises from deletions or missense mutations in the *FMRP* gene.[404] There are actually two similar genes *FMRP*1 and *FMRP*2. Mutations in the latter are associated with a milder form of fragile X disease. FMRP genes with expanded $(CGG)_n$ tracts are not expressed, evidently because the mutation induces methylation of an adjacent "CpG island" (see Chapter 27) as well as of sites within the FMRP gene. Both FMRP proteins are apparently needed for normal brain function. The FMRP1 defect is genetically dominant, and female heterozygotes also suffer from the fragile X syndrome. However, FMRP2, also encoded on the X chromosome, is recessive.[405] FMRP appears to function in neuronal dendrites (Chapter 30) where it binds to polysomal aggregates and participates in regulation of translation of mRNA.[405a–c] Like DNA, mRNA may contain **G quartets** (Fig. 5-8, p. 227).[405d]

Twelve or more additional triplet repeat diseases, many with neurological symptoms, have been identified (Table 26-4).[405–407a] These involve other trinucleotide repeats 5'-$(GCG)_n$, 5'-$(CTG)_n$, 5'-$(GAA)_n$, and 5'-$(CAG)_n$. In **synpolydactyly**, an inherited developmental defect causing malformation of hands and feet, an expansion of a GCG trinucleotide occurs within the gene *HoxD*13. This results in incorporation of a polyalanine tract near the N terminus of the protein.[408] **Myotonic dystrophy** DM1 (Box 19-A) results from expansion of CTG to 6 kbp or more within the untranslated 3' region of a gene for cAMP-dependent protein kinase.[405] The mRNA transcripts accumulate in the nucleus and may bind to a CUG-binding protein that is involved in splicing other mRNAs, thereby poisoning the cell.[409] DM2 is caused by expansion of a CCTG quartet in an intron of a zinc-finger protein.[408a] Expansion of the GAA triplet is associated with the neurological disease **Friedreich ataxia**, which has a prevalence of ~1 in 50,000. The defect lies within an intron found in the gene for the 210-residue mitochondrial protein **frataxin**. The function of the protein is unknown.[409a] However, the defect in Freidrich ataxia leads to a deficit in mitochondrial ATP synthesis.[409b] Studies of a corresponding protein in yeast suggests that frataxin is an iron storage protein.[409c–e] Apparently the DNA defect in the intron interferes with splicing of the mRNA transcript.[410–413] The polyGAA strand in the triplet repeat region of the DNA is able to form various alternative structures including a parallel (GAA)•(TTC) duplex.[410a] Such structures in mRNA may interfere with proper splicing.

A defect in an α-tocopherol-transfer protein causes a similar set of neurological symptoms. Oxidative damage may therefore be a component of this disease.[410b]

Expansion of $(CAG)_n$ sequences causes a series of neurodegenerative diseases (Table 26-4),[413–416c] the commonest of which is **Huntington disease** (HD).[414–414c] In Huntington's disease the $(CAG)_n$ tract is found in the **huntington** gene. The protein is enormous, with over 3140 residues. It is essential for nerve development, but the function remains uncertain. The first exon at the HD gene encodes a polyglutamine tract of 6–35 residues, which when expanded to 36–100 or more causes the disease. A similar situation holds for the **spinocerebellar ataxias** (SCA) and other $(CAG)_n$ diseases (Table 26-4).[414,417–417b] The encoded proteins have a variety of functions. The SCA-1 protein ataxin-1 functions in the nuclear matrix.[418] The X-linked **spinobulbar muscular atrophy** (SBMA) gene encodes an androgen receptor protein.[414,417] The polyglutamine sequences inserted in those proteins seem to be toxic, but the mechanism of toxicity is uncertain. It may result from aggregation of the proteins within cells.[414] It has also been suggested that cleavage of these chains by **caspases** (cysteine proteases; Chapter 12) may produce truncated proteins, which induce apoptosis.[415]

A form of epilepsy (Table 27-4) appears to be a result of repeats of a (G + C)-rich sequence that may be a dodecamer.[405] Dinucleotide repeats and other "minisatellite" DNA sequences are also associated with instability of DNA and may undergo expansion.[419–421] A pentanucleotide repeat $(CCTTT)_n$ is associated with increased expression of the nitric oxide synthase gene NOS2A. Persons with $n = 14$ were found to have enhanced resistance to development of diabetic retinopathy. This seems to be a case of a beneficial "gain of function" mutation.[422]

How do repeat sequences expand from one generation to the next? There is probably more than one mechanism. One is **strand slippage** during DNA replication. If single-strand loops are present at any time one strand could be displaced (slipped) relative to the other. Replication could then either expand the repetitive sequence or cause a deletion (Eq. 26-4).[405,423–425] Expansion could also occur by gene conversion during homologous recombination (Chapter 27, Section D)[401,426] or during DNA repair.[421,425] Repeat sequences may also prevent proper formation of nucleosomes.[402]

Cloned genes and diagnosis. The first areas in which cloned DNA has affected medicine are in the diagnosis of genetic diseases and in the production of medicinally useful proteins in bacteria, yeast, or cultured cells. One of the first applications was for the diagnosis of the sickle-cell trait by use of PCR on DNA isolated from blood.[43] This was followed quickly by methods for recognition of other inherited diseases and by automated procedures. Intense efforts are now being made to develop DNA "chips" (usually small glass plates with an array of DNA fragments bound to the surface) that can recognize a great variety of defects. For example, DNA polymorphisms in the 16.6-kbp human mitochondrial genome can be recognized by a plate containing 135,000 oligonucleotide probes assembled in a regular grid by photolithography and solid-state synthesis (Chapter 3).[427] Cancers may be classified quickly.[428] Many systems are under development for binding and recognizing genomic DNA, cDNA, or mRNA using microchip arrays of up to 400,000 or more oligonucleotides on a 2 cm x 2 cm plate.[429–430b] One chip combines PCR with use of "zip-code primers" that direct the PCR products to specific zip-code addresses on the chip to give a universal array able to detect low abundance of mutations in any gene of interest.[431] Mass spectrometry is also being harnessed to identify oligonucleotides bound at any address on a chip.[432] Commercial chips are expensive, $100–2000 apiece, and good for only one use. The price will fall. To build a machine to make your own chips go to http//cmgm.stanford.edu/pbrown/mguide/.[433,434]

An alternative to DNA chips is to miniaturize DNA sequencing and analysis machinery. Using nanoliter droplets of fluid passing through microchannels built by photolithographic techniques of computer chip construction, these nanolaboratories may be the size of a credit card but able to cleave DNA and conduct PCR reactions, gel electrophoresis, and sequence determinations.[435,436]

Vaccines, hormones, and other medicines. A myriad of products of recombinant DNA technology are already in use in medical diagnostics.[437] In the past problems have arisen because vaccines and hormones can cause allergic reactions and may harbor viruses. A small percentage of diabetics are allergic to animal insulins, but human insulin produced in bacteria is now available. A number of children receiving human growth hormone isolated from cadavers contracted the fatal Creutzfeld–Jakob disease,[438] a neurological disorder caused by a prion (Chapter 29). One way in which AIDS has been spread is through contamination of the blood-clotting factors VIII or IX needed by hemophiliacs. These sources of contamination are being eliminated by the use of bacterially produced proteins.[439]

Recombinant DNA techniques can be used in two ways in the production of vaccines. The first is to find a protein in the virus or other infective agent that is a good inducer of antibody formation, i.e., a good antigen. This protein, or even a fragment of it, can then be produced from its cloned gene or can be made synthetically. Since the cloned protein can be purified highly, it may make a superior vaccine to those made from killed cells, inactivated virus particles, or mixtures of proteins.[440] The first commercial vaccine of this type was against viral hepatitis B, a major cause of liver cancer.[441] Particles consisting of viral envelope proteins can be produced in yeast and be used for vaccination. A DNA encoding these proteins has been transferred, using the Ti plasmid (Section E, 4), into lettuce. Human volunteers produced anti-hepatitis antibodies after eating the lettuce.[442] This suggests that vaccination through ingestion of antigenic proteins in food crops may be feasible. Injection of a small piece of DNA carrying the gene may also lead to antibody formation.[443,444] Using either purified antigens, proteins, or DNA, it may be possible to develop effective vaccines against Rocky Mountain spotted fever,[445] a rickettsial disease, and against malaria, for which there are no satisfactory vaccines.

The vaccinia virus, formerly used to vaccinate against smallpox, has been engineered for use against other diseases.[446] Much of its 187-kb DNA can be excised and replaced with passenger DNA. The virus particles are stable and highly infectious. The vaccinia virus is unusual in carrying genes for both RNA and DNA polymerases and other proteins that permit it to undergo replication and transcription of its own genes in the cytoplasm of the infected cells. One application of a recombinant vaccinia virus is oral vaccination of wild foxes and racoons against rabies.[447]

A major advance based on cloned genes is the production of new medicines previously unavailable or available in only small amounts. Among these are the **interferons**[287] (Chapter 31) and many hormones such as the **interleukins** produced by lymphocytes[448] and the **atrial natriuretic hormones** (Chapter 23). Another candidate is the **α1-protease inhibitor** (Chapter 12). Perhaps better inhibitors than the natural one can be devised and produced in bacteria.[449]

4. Gene Therapy

A few years ago it seemed like fantasy, but there is little doubt that we will soon be able to routinely treat some genetic illnesses by introducing new copies of genes into the body. A current goal is to insert cloned genes into body cells (somatic cells) to correct specific hereditary defects. For example, juvenile diabetics would benefit from introduction of genes for insulin production into cells that could replace their atrophied pancreatic beta cells. At present we don't know how to do this. However, genes have been transferred into human beings lacking adenosine deaminase and showing severe **combined immunodeficiency** (Chapter 31) and those lacking hypoxanthine guanine phosphoribosyltransferase and displaying the **Lesch–Nyhan syndrome** (Chapter 25, Section E,2). Corrections of **Gaucher disease** and other deficiencies of lysosomal enzymes is also an early goal (Chapter 20, Section G,2).

By 1986 more than 5000 children had received bone marrow transplants from close relatives to correct severe combined immunodeficiency caused by a defective adenosine deaminase gene.[450] However, the patients must receive chemotherapy or irradiation to suppress their immune system before the transplantation. Hospitalization may last for 30–60 days. Using genetic therapy some bone marrow cells can be removed from the patient with a needle. The cells can be treated to introduce the corrected genes using a suitable retrovirus. Clones of corrected bone marrow stem cells, which will give rise to lymphocytes in the body, can be selected, cutured and reintroduced into the patient. As discussed on p. 1498, retrovirus vehicles that locate and become integrated at appropriate sites in the genome have been developed.[216,451] Use of homologous recombination (Chapter 27) introduces the cloned gene into its normal chromosomal location.[452] The first children were treated by transfer of the adenosine deaminase gene in 1991. The procedure has been partially successful.[453,454]

By 1999 more than 400 clinical gene therapy trials were planned or in progress.[455] Nevertheless, development of suitable vehicles for gene delivery has been slow.[456,456a] Uncertainty about the safety of adenovirus vectors is one problem.[160a,454,457,457a] Poor efficiency of gene transfer is another. A glycogen storage disease of knockout mice has been cured by transfer of human α-glucosidase (Box 20-D) using an adenovirus vector.[458] Mice have also been used in developing gene therapy for hemophilia,[456a,457a] sickle-cell disease,[458a] and aspartylglycosaminuria.[458b] However, gene targeting in animals other than the mouse has been difficult.[459] Genetic therapy may be most effective when the gene is transferred into stem cells, which can then take up residence within the body (Chapter 32). Young stem cells, which can be obtained from unbilical cord blood at birth, can be used.[460,461] For many diseases therapy will probably be best soon after birth, or even prior to birth during the second trimester of pregnancy.[455]

An alternative approach is to synthesize highly specific hydrolysis-resistant DNA analogs that can form triple helical structures with DNA (Chapter 5). If these can be made specific enough they might bind to a targeted DNA site, such as the sickle cell anemia locus and induce a "back mutation" from the faulty A•T base pair (see Fig. 7-23) to a T•A pair in at least some of the hematopoietic cells that give rise to hemoglobin.[462] Another possibility is to make oligonucleotide analogs that serve as mimics of **antisense RNA** (discussed in Chapter 28). This could impede translation of bad mRNAs, such as that giving rise to polyglutamine chains in the triple repeat diseases.

At this time no effort is being made to alter the DNA in human germ cells. It seems undesirable to experiment with such changes.[160a,462a] However, as methods are developed for genetic therapy of somatic cells, we will rear more and more healthy carriers of serious genetic defects. Eventually we may need to develop therapy for germ cells.

5. Genetic Engineering of Bacteria, Plants, and Animals

Many improvements in bacteria used in industrial fermentations have been made.[463] The number of copies of a useful gene may be increased, and repressed genes may be made more active by deletion mutations in regulatory genes (Chapter 28). New genes are being transferred between bacteria and into plants and animals. For example, *Bacillus thuringensis* produces crystalline protein toxins, which are harmless to mammals but active against many insects. The *Bt* gene for this toxin has already been transferred into many different plants and is having a major effect on agriculture in allowing for decreased use of insecticides.[464,465] Genetic engineering may allow the toxin to be made more specific for particular insect species.

Genetic engineering of plant genes[466] may improve the quality of storage proteins in cereal grains[467] and the flavors of fruits,[468] provide more drought-resistant plants,[469] and offer increased resistance of crop plants to particular herbicides[464] (Chapter 25, Section A,1) and to viruses.[470] Many things can be done to enhance the nutritional qualities of foods.[464] For example, rice has been engineered to produce and store β-carotene in the grain, a development that may benefit 400 million people in the world deficient in vitamin A and often suffering from infections and blindness.[471] (See also p. 1240.) Specialty products such as poly-β-hydroxybutyrate and many others can also be produced in plants.[472] Genetic engineering of animals has produced not only knockout mice but the

possibility for nutritional improvement[156] and for production of useful products such as vaccines in milk.

6. Ethical Problems

With a host of new medicines and agricultural products coming and with the ability to alter genomes at will, we face new ethical problems. For example, who owns the human genome?[473–475] Should patenting of human genes be allowed?[476,477] How private are genetic data?[478] Who has access to human DNA data?[479] Insurance companies?[480,481] Should routine screening with new genetic tests be used (as is now the case for phenylketonuria, diabetes, and others), even if no effective treatment is available?[482,483]

Ethical questions arise in application of experimental forms of genetic therapy to patients dying of inherited disorders. What if the therapy simply provides a longer life of suffering to patients? Some biologists foresee a future in which human beings learn to control their own genes and prevent genetic deterioration resulting from the accumulation of harmful mutations. They find it exciting to think that we can, in an intelligent way, elect to continue our evolution in a desired direction. Others caution that our present knowledge is such that attempts to eliminate all "bad" genes from the population might be disastrous.[484] They point to the hemoglobin S gene (Box 7-B) and the role it once had in preserving life in a malaria-infested environment and urge that at this stage we allow maximum heterogeneity of genetic types. There are other dangers in allowing genetics to control human lives. In the past "eugenic" doctrines have been used to justify racist laws and (in Nazi Germany) genocide.

Problems have resulted from the availability of hormones and other products of genetic engineering. For example, should parents be allowed to request human growth hormone for short children who do not have a pituitary deficiency? Should it be available to athletes and to aging people? Use of stem cells in genetic therapy raises another set of questions. Present guidelines of the National Institutes of Health (NIH), which supports much of the gene therapy research, allows use of stem cells in therapy but doesn't allow their use in reproductive cloning or the combining of human stem cells with animal embryos.[485] Dangers from retroviruses may lurk in attempts to transplant animal organs (e.g., pig hearts) into people.[486–488]

In attempting to improve crop plants, proteins causing allergic reactions in some individuals may be tranferrred.[489] This is a compelling reason that mandatory monitoring and labeling of genetically modified foods is recommended.[490] The danger is emphasized by the highly popularized contamination of foods with "starlink" maize.

Should we rush commercial developments of genetic engineering to increase food supply at a time when there is an excess worldwide? Should we engineer increased herbicide-resistance into plants when drinking water supplies are being contaminated by herbicides already in use? On the other hand, less toxic herbicides effective at low levels can be used on plants engineered to resist them.[465] Will we generate superweeds resistant to herbicides by gene transfer? With so many genes being engineered, are there other dangers in their release into the environment?[465] A **terminator** technology that produces plants with sterile seeds has been rejected because it would force poor farmers to purchase seeds year after year from multinational corporations. Yet, it could eventually have beneficial consequences in preventing cross-fertilization with genetically modified plants.[491] Since plants with the *Bt* gene kill insects, will their widespread use decimate populations of butterflies or of various beneficial insects or adversely affect soil organisms in the **rhizosphere**?[492,493]

The darkest prospect may be the possibility that military organizations will develop new biological warfare weapons using recombinant DNA methods.[494] An international treaty bans such weapons. However, governments often respond with the argument "We must do it because we know that 'they' are doing it." Couldn't such an attitude bring on unparalleled disaster? Could new types of viruses spread throughout the world and literally tear the human genome to bits? Does any kind of imagined danger justify preparation to attack populations with new viruses or new toxins? What about the smallpox virus? Should remaining stocks in the United States and Russia be destroyed?[495,496]

References

1. Horowitz, N. H. (1995) *Protein Sci.* **4**, 1017–1019
2. Atherly, A. G., Girton, J. R., and McDonald, J. F. (1999) *The Science of Genetics*, Saunders, Philadelphia, Pennsylvania

2a. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002) *Molecular Biology of the Cell*, 4th ed., Garland Science, New York

2b. Lander, E. S., and Weinberg, R. A. (2000) *Science* **287**, 1777 – 1782

3. Mirsky, A. E. (1968) *Sci. Am.* **218**(Jun), 78–88
4. Portugal, F. H., and Cohen, J. S. (1977) *A Century of DNA: A History of the Discovery of the Structure and Function of the Genetic Substance*, MIT Press, Cambridge, Massachusetts
5. Adams, R. L. P., Knowler, J. T., and Leader, D. P. (1992) *The Biochemistry of the Nucleic Acids*, 11th ed., Chapman & Hall, London
6. Olby, R. (1975) *The Path to the Double Helix*, Univ. of Washington Press, Seattle, Washington
7. Stent, G. S., and Calendar, R. (1978) *Molecular Genetics*, 2nd ed., Freeman, San Francisco, California
8. Chargaff, E. (1971) *Science* **172**, 637–642
9. Olby, R. (1974) *Nature (London)* **248**, 782–785
10. McCarty, M. (1985) *The Transforming Principle*, Norton, New York
11. Cohen, S. S. (1984) *Trends Biochem. Sci.* **9**, 334–336
12. Chambers, D. A., Reid, K. B. M., and Cohen, R. L. (1994) *FASEB J.* **8**, 1219–1226
13. Hershey, A. D., and Chase, M. (1952) *J. Gen. Physiol.* **36**, 39–56
14. Hayes, W. (1968) *The Genetics of Bacteria and Their Viruses; Studies in Basic Genetics and Molecular Biology*, 2nd ed., Wiley, New York
15. Chargaff, E. (1950) *Experientia* **6**, 201–209
16. Brinton, C. C. (1971) *Crit. Rev. Microbiol* **1**, 105–160
17. Sayre, A. (1975) *Rosalind Franklin and DNA*, Norton, New York

17a. Maddox, B. (2002) *Rosalind Franklin: The Dark Lady of DNA*, HarperCollins, New York

17b. Blow, D. (2002) *Nature (London)* **418**, 725 – 726

18. Watson, J. D., and Crick, F. H. C. (1953) *Nature (London)* **171**, 737–738
19. Crick, F. (1974) *Nature (London)* **248**, 766–769
20. Watson, J. D. (1968) *The Double Helix*, Atheneum, New York

20a. Morange, M. (2000) *A History of Molecular Biology*, Harvard Univ. Press, Cambridge,MA (translated by Matthew Cobb)

20b. Echols, H. (2001) *Operators and Promoters: The Story of Molecular Biology and Its Creators*, Univ. of California Press, Berkeley, California (edited by Carol A. Gross)

21. Caspersson, T. (1941) *Naturwissenschaften* **29**, 33–43
22. Brachet, J. (1942) *Arch. Biol.* **53**, 207–257
23. Brachet, J. (1987) *Trends Biochem. Sci.* **12**, 244–246
24. Prescott, D. M. (1964) *Prog. Nucleic Acid Res. Mol. Biol.* **3**, 33–57
25. Sirlin, J. L. (1972) *Biology of RNA*, Academic Press, New York
26. Hoagland, M. B., Keller, E. B., and Zamecnik, P. (1956) *J. Biol. Chem.* **218**, 345–358
27. Belozersky, A. N., and Spirin, A. S. (1960) in *The Nucleic Acids*, Vol. 3 (Chargaff, E., and Davidson, J. N., eds), pp. 147–185, Academic Press, New York
28. Davern, C. I., and Meselson, M. (1960) *J. Mol. Biol.* **2**, 153–160
29. Volkin, E., and Astrachan, L. (1956) *Virology* **2**, 149–161
30. Volkin, E. (1995) *Trends Biochem. Sci.* **20**, 206–209
31. Rouwenhorst, R. J., Pronk, J. T., and van Dijken, J. P. (1989) *Trends Biochem. Sci.* **14**, 416–418
32. Jacob, F., and Monod, J. (1961) *J. Mol. Biol.* **3**, 318–356
33. Cohen, G. N. (1995) *FASEB J.* **9**, 981–982
34. Brenner, S., Jacob, F., and Meselson, M. (1961) *Nature (London)* **190**, 576–581
35. Crick, F. H. C. (1957) *Biochem. Soc. Symp.* **14**, 25–26
36. Hoagland, M. (1996) *Trends Biochem. Sci.* **21**, 77–80
37. Nirenberg, M. W., and Matthaei, J. H. (1961) *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1588–1602
38. Nirenberg, M., and Leder, P. (1964) *Science* **145**, 1399–1407
39. Matthaei, J. H., Voigt, H. P., Heller, G., Neth, R., Schöch, G., Kübler, H., Amelunxen, F., Sander, G., and Parmeggiani, A. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 25–38
40. Khorana, H. G., Büchi, H., Ghosh, H., Gupta, N., Jacob, T. M., Kössel, H., Morgan, R., Narang, S. A., Ohtsuka, E., and Wells, R. D. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 39–49
41. Edgar, R. S., and Epstein, R. H. (1965) *Sci. Am.* **212**(Feb), 70–78
42. Fraenkel-Conrat, H. (1964) *Sci. Am.* **211**(Oct), 47–54
43. Wu, D. Y., Ugozzoli, L., Pal, B. K., and Wallace, R. B. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2757–2760
44. Lewin, B. (1974) *Gene Expression*, Vol. 2, Wiley, New York (p. 258)
45. Knight, R. D., Freeland, S. J., and Landweber, L. F. (1999) *Trends Biochem. Sci.* **24**, 241–247
46. Seid-Akhavan, M., Winter, W. P., Abramson, R. K., and Rucknagel, D. L. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 882–886
47. Benzer, S. (1962) *Sci. Am.* **206**(Jan), 70–84
48. Hayes, W. (1968) *The Genetics of Bacteria and Their Viruses*, 2nd ed., Wiley, New York (pp. 52–54)
49. Weintraub, S. B., and Frankel, F. R. (1972) *J. Mol. Biol.* **70**, 589–615
50. Ennis, H. L., and Kievitt, K. D. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1468–1472
51. Beadle, G. W., and Tatum, E. L. (1941) *Proc. Natl. Acad. Sci. U.S.A.* **27**, 499–
52. Yanofsky, C. (1967) *Sci. Am.* **216**(May), 80–94
53. Yanofsky, C. (1976) in *Reflections on Biochemistry* (Kornberg, A., ed), pp. 263–269, Pergamon, New York
54. Sarabhai, A. S., Stretton, A. O. W., Brenner, S., and Bolle, A. (1964) *Nature (London)* **210**, 13–17
55. Crick, F. H. C., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961) *Nature (London)* **192**, 1227–1232

55a. Atkins, J. F., and Gesteland, R. F. (2001) *Nature (London)* **414**, 693

56. Watson, J. O. (1987) *Molecular Biology of the Gene*, 4th ed., Benjamin, New York
57. Eiserling, F. A., and Dickson, R. C. (1972) *Ann. Rev. Biochem.* **41**, 467–502
58. Wood, W. B., and Edgar, R. S. (1967) *Sci. Am.* **217**(Jul), 60–74
59. Sober, H. A., ed. (1968) *CRC Handbook of Biochemistry*, 1st ed., Chem. Rubber Publ. Co., Cleveland, Ohio (p. I25)
60. Garen, A. (1968) *Science* **160**, 149–159
61. Littauer, U. Z., and Inouye, H. (1973) *Ann. Rev. Biochem.* **42**, 439–470
62. Hartman, P. E., and Roth, J. R. (1973) *Adv. Genet.* **17**, 1–105
63. Lewin, B. (1974) *Gene Expression*, Vol. 1, Wiley, New York (p. 213)
64. Yanofsky, C., and Crawford, I. P. (1972) in *The Enzymes*, 3rd ed., Vol. 7 (Boyer, P. D., ed), Academic Press, New York (pp. 1–31)
65. Sherratt, D. (1975) *Nature (London)* **254**, 559–560
66. Lederberg, J., and Tatum, E. L. (1946) *Nature (London)* **158**, 558
67. Kornberg, A. (1974) *DNA Synthesis*, Freeman, San Francisco, California (p. 242–248)
68. Harrington, L. C., and Rogerson, A. C. (1990) *J. Bacteriol.* **172**, 7263–7264
69. Eisenbrandt, R., Kalkum, M., Lai, E.-M., Lurz, R., Kado, C. I., and Lanka, E. (1999) *J. Biol. Chem.* **274**, 22548–22555
70. Hendrix, R. W., Roberts, J. W., Stahl, F. W., and Weisberg, R., eds. (1983) *Lambda II*, Cold spring Harbor Lab, Cold Spring Harbor, New York
71. Hershey, A. D., ed. (1971) *The Bacteriophage Lambda*, Cold Spring Harbor Lab, Cold Spring Harbor, New York
72. Echols, H. (1971) *Ann. Rev. Biochem.* **40**, 827–854
73. O'Brien, S. J., ed. (1993) *Genetic Maps*, 6th ed., Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York ((Books 1–6))
74. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) *Science* **277**, 1453–1462
75. Bachmann, B. J., Low, K. B., and Taylor, A. L. (1976) *Bacteriol. Rev.* **40**, 116–167
76. Bachmann, B. J. (1983) *Bacteriol. Rev.* **47**, 180–230
77. Smith, C. L., Econome, J. G., Schutt, A., Klco, S., and Cantor, C. R. (1987) *Science* **236**, 1448–1453
78. Kohara, Y., Akiyama, K., and Isono, K. (1987) *Cell* **50**, 495–508
79. Low, K. B. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), Am. Soc. for Microbiology, Washington, D. C.
80. Willetts, N., and Skurray, R. (1987) in *Escherichia coli and Salmonella typhimurium*, Vol. 2 (Neidhardt, F. C., ed), pp. 1110–1133, Am. Soc. for Microbiology, Washington, D. C.

80a. Bist, P., Sistla, S., Krishnamurthy, V., Acharya, A., Chandrakala, B., and Rao, D. N. (2001) *J. Mol. Biol.* **310**, 93 – 109

81. Davies, G. P., Martin, I., Sturrock, S. S., Cronshaw, A., Murray, N. E., and Dryden, D. T. F. (1999) *J. Mol. Biol.* **290**, 565–579
82. Janscak, P., MacWilliams, M. P., Sandmeier, U., Nagaraja, V., and Bickle, T. A. (1999) *EMBO J.* **18**, 2638–2647
83. Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I., and Aggarwal, A. K. (1995) *Science* **269**, 656–663
84. Bozic, D., Grazulis, S., Siksnys, V., and Huber, R. (1996) *J. Mol. Biol.* **255**, 176–186
85. Luria, S. E. (1970) *Sci. Am.* **222**(Jan), 88–102
86. McClarin, J. A., Frederick, C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J., and Rosenberg, J. M. (1986) *Science* **234**, 1526–1541
87. Hager, P. W., Reich, N. O., Day, J. P., Coche, T. G., Boyer, H. W., Rosenberg, J. M., and Greene, P. J. (1990) *J. Biol. Chem.* **265**, 21520–21526
88. Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B.-C., Greene, P., Boyer, H. W., and Rosenberg, J. M. (1984) *Nature (London)* **309**, 327–331
89. Frederick, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., and Rich, A. (1988) *J. Biol. Chem.* **263**, 17872–17879

References

90. Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I., and Aggarwal, A. K. (1994) *Nature (London)* **368**, 660–664
91. Kostrewa, D., and Winkler, F. K. (1995) *Biochemistry* **34**, 683–696
91a. Horton, N. C., Otey, C., Lusetti, S., Sam, M. D., Kohn, J., Martin, A. M., Ananthnarayan, V., and Perona, J. J. (2002) *Biochemistry* **41**, 10754 – 10763
92. Stanford, N. P., Halford, S. E., and Baldwin, G. S. (1999) *J. Mol. Biol.* **288**, 105–116
93. Sam, M. D., and Perona, J. J. (1999) *Biochemistry* **38**, 6576–6586
94. Kim, Y., Grable, J. C., Love, R., Greene, P. J., and Rosenberg, J. M. (1990) *Science* **249**, 1307–1309
95. Kim, S. C., Podhajska, A. J., and Szybalski, W. (1988) *Science* **240**, 504–506
96. Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980) *Am. J. Hum. Genet.* **32**, 314–331
97. Cohen, D., Chumakov, I., and Weissenbach, J. (1993) *Nature (London)* **366**, 698–701
98. Westmoreland, B. C., Szybalski, W., and Ris, H. (1969) *Science* **163**, 1343–1348
99. Glover, D. M., ed. (1987) *DNA Cloning: A Practical Approach*, Vol. III, IRL Press, Oxford
100. Allemand, J.-F., Bensimon, D., Jullien, L., Bensimon, A., and Croquette, V. (1997) *Biophys. J.* **73**, 2064–2070
101. Michalet, X., Ekong, R., Fougerousse, F., Rousseaux, S., Schurra, C., Hornigold, N., van Slegtenhorst, M., Wolfe, J., Povey, S., Beckmann, J. S., and Bensimon, A. (1997) *Science* **277**, 1518–1523
102. Samad, A. H., Cai, W. W., Hu, X., Irvin, B., Jing, J., Reed, J., Meng, X., Huang, J., Huff, E., Porter, B., Shenkar, A., Anantharaman, T., Mishra, B., Clarke, V., Dimalanta, E., Edington, J., Hiort, C., Rabbah, R., Skiada, J., and Schwartz, D. C. (1995) *Nature (London)* **378**, 516–517
103. Lin, J., Qi, R., Aston, C., Jing, J., Anantharaman, T. S., Mishra, B., White, O., Daly, M. J., Minton, K. W., Venter, J. C., and Schwartz, D. C. (1999) *Science* **285**, 1558–1561
104. Cox, D. R., Burmeister, M., Price, E. R., Kim, S., and Myers, R. M. (1990) *Science* **250**, 245–250
105. Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., and Venter, J. C. (1991) *Science* **252**, 1651–1656
106. Mann, M. (1996) *Trends Biochem. Sci.* **21**, 494–495
107. Watson, J. D., Gilman, M., Witkowski, J., and Zoller, M., eds. (1992) *Recombinant DNA*, 2nd ed., Scientific American Publ. Co., New York
108. Sambrook, J., and Russell, D. (2000) *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Lab. Press, Plainview, New York
109. Berger, S. L., and Kimmel, A. R., eds. (1987) *Guide to Molecular Cloning Techniques*, Academic Press, San Diego
110. Perbal, B. (1988) *A Practical Guide to Molecular Cloning*, Wiley, New York
111. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry, Part II*, Freeman, San Francisco, California (pp. 409–433)
112. Greene, J. J., and Rao, V. B., eds. (1998) *Recombinant DNA Principles and Methodologies*, Dekker, New York
113. Miller, J. H. (1992) *A Short Course in Bacterial Genetics*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
114. Wu, R., ed. (1987) *Methods in Enzymology, "Recombinant DNA,"* Vols. 153–155, Academic Press, San Diego
115. Pouwels, P. H., Enger-Valk, B. E., and Brammar, W. J., eds. (1985) *Cloning Vectors: A Laboratory Manual*, Elsevier, Amsterdam
116. Adolph, K. W., ed. (1993) *Methods in Molecular Genetics: Gene and Chromosome Analysis*, Vols. 1 and 2, Academic Press, San Diego, California
117. Birren, B., Green, E. D., Klapholz, S., Myers, R. M., and Roskams, J., eds. (1997) *Analyzing DNA: A Laboratory Manual*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
118. Birren, B., Green, E. D., Klapholz, S., Myers, R. M., Riethman, H., and Roskams, J., eds. (1999) *Cloning Systems: A Laboratory Manual*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
119. Lathe, R., Kieny, M. P., Skory, S., and Lecocq, J. P. (1984) *DNA* **3**, 173–182
120. Nobbs, T. J., and Halford, S. E. (1995) *J. Mol. Biol.* **252**, 399–411
121. Winnacker, E. L. (1987) *From Genes to Clones*, VCH, New York
122. Midgley, C. A., and Murray, N. E. (1985) *EMBO J.* **4**, 2695–2703
123. Bollum, F. J. (1981) *Trends Biochem. Sci.* **6**, 41–43
123a. Shanks, O. C., Bissonette, L., and Ream, W. (2000) *Science* **289**, 413
124. Bolivar, F., Rodriguez, R. L., Coreene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H., and Falkow, S. (1977) *Gene* **2**, 95–113
125. Peden, K. W. C. (1983) *Gene* **22**, 277–280
126. Vieira, J., and Messing, J. (1982) *Gene* **19**, 259–268
127. Messing, J. (1981) *Nucleic Acids Res.* **9**, 309–321
128. Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchison, C. A., III, Slocombe, P. M., and Smith, M. (1978) *J. Mol. Biol.* **125**, 225–246
129. Blattner, F. R., Williams, B. G., Blechi, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L.-A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L., and Smithies, O. (1977) *Science* **196**, 161–169
130. Short, J. M. (1988) *Nucleic Acids Res.* **15**
131. Collins, J., and Hohn, B. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4242–4246
132. Wahl, G. M., Lewis, K. A., Ruiz, J. C., Rothenberg, B., Zhao, J., and Evans, G. A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2160–2164
133. Fernandez, J. M., and Hoeffler, J. P., eds. (1998) *Gene Expression Systems*, Academic Press, San Diego, California
134. Moss, B., Elroy-Stein, O., Mizukami, T., Alexander, W. A., and Fuerst, T. R. (1990) *Nature (London)* **348**, 91–92
135. O'Reilly, D. R., Miller, L. K., and Luckow, V. A. (1992) *Baculovirus Expression Vectors*, Freeman, New York
136. Kriegler, M. (1990) *Gene Transfer and Expression*, Stockton Press, New York
137. Burke, D. T., Carle, G. F., and Olson, M. V. (1987) *Science* **236**, 806–812
138. Murray, A. W., and Szostak, J. W. (1987) *Sci. Am.* **257**(Nov), 62–68
139. Forget, B. G. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7909–7911
140. Anderson, C. (1993) *Science* **259**, 1684–1687
141. Sternberg, N. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 103–107
142. Pierie, J. C., and Sternberg, N. L. (1992) *Methods Enzymol.* **216**, 549–
143. Heller, R., Brown, K. E., Burgtorf, C., and Brown, W. R. A. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7125–7130
144. Chilton, M.-D. (1983) *Sci. Am.* **248**(Jun), 51–59
145. Weiler, E. W., and Schröder, J. (1987) *Trends Biochem. Sci.* **12**, 271–275
146. Zambryski, P. C. (1992) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **43**, 465–490
146a. Wood, D. W., and 50 other authors. (2001) *Science* **294**, 2317–2323
146b. Goodner, B., and 30 other authors. (2001) *Science* **294**, 2323–2328
146c. Moriguchi, K., Maeda, Y., Satou, M., Hardayani, N. S. N., Kataoka, M., Tanaka, N., and Yoshida, K. (2001) *J. Mol. Biol.* **307**, 771–784
147. Slightom, J. L., Durand-Tardif, M., Jouanin, L., and Tepfer, D. (1986) *J. Biol. Chem.* **261**, 108–121
148. Hess, K. M., Dudley, M. W., Lynn, D. G., Joerger, R. D., and Binns, A. N. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7854–7858
149. Stachel, S. E., Messens, E., Van Montagu, M., and Zambryski, P. (1985) *Nature (London)* **318**, 624–629
150. Bates, H. A., Kaushal, A., Deng, P.-N., and Sciaky, D. (1984) *Biochemistry* **23**, 3287–3290
151. Tinland, B., Schoumacher, F., Gloeckler, V., Bravo-Angel, A. M., and Hohn, B. (1995) *EMBO J.* **14**, 3585–3595
152. Scheiffele, P., Pansegrau, W., and Lanka, E. (1995) *J. Biol. Chem.* **270**, 1269–1276
153. Bundock, P., den Dulk-Ras, A., Beijersbergen, A., and Hooykaas, P. J. J. (1995) *EMBO J.* **14**, 3206–3214
154. Fullner, K. J., Lara, J. C., and Nester, E. W. (1996) *Science* **273**, 1107–1109
154a. Ward, D. V., Draper, O., Zupan, J. R., and Zambryski, P. C. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11493–11500
155. Crystal, R. G. (1995) *Science* **270**, 404–410
156. Westphal, H. (1989) *FASEB J.* **3**, 117–120
157. Cotton, M., Wagner, E., Zatloukal, K., Phillips, S., Curiel, D. T., and Birnstiel, M. L. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6094–6098
158. Kojima, H., Ohishi, N., and Yagi, K. (1998) *Biochem. Biophys. Res. Commun.* **246**, 868–872
159. Yu, S.-F., von Rüden, T., Kantoff, P. W., Garber, C., Seiberg, M., Rüther, U., Anderson, W. F., Wagner, E. F., and Gilboa, E. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3194–3198
160. Chakraborty, A. K., Zink, M. A., Boman, B. M., and Hodgson, C. P. (1993) *FASEB J.* **7**, 971–977
160a. Ferber, D. (2001) *Science* **294**, 1638–1642
161. Daly, T. M., Vogler, C., Levy, B., Haskins, M. E., and Sands, M. S. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2296–2300
161a. Yan, Z., Zhang, Y., Duan, D., and Engelhardt, J. F. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6716–6721
161b. Xie, Q., Bu, W., Bhatia, S., Hare, J., Somasundaram, T., Azzi, A., and Chapman, M. S. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10405–10410
162. Amado, R. G., and Chen, I. S. Y. (1999) *Science* **285**, 674–676
162a. Lois, C., Hong, E. J., Pease, S., Brown, E. J., and Baltimore, D. (2002) *Science* **295**, 868–872
163. Jasny, B. R. (1987) *Science* **238**, 1653
164. Hofmann, C., Sandig, V., Jennings, G., Rudolph, M., Schlag, P., and Strauss, M. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10099–10103
165. Yin, H. L., and Stull, J. T. (1999) *J. Biol. Chem.* **274**, 32529–32530
166. Poul, M.-A., and Marks, J. D. (1999) *J. Mol. Biol.* **288**, 203–211
166a. Kaminski, J. M., Huber, M. R., Summers, J. B., and Ward, M. B. (2002) *FASEB J.* **16**, 1242–1247
166b. Saldanha, R., Chen, B., Wank, H., Matsuura, M., Edwards, J., and Lambowitz, A. M. (1999) *Biochemistry* **38**, 9069–9083
166c. Guo, H., Karberg, M., Long, M., Jones, J. P., III, Sullenger, B., and Lambowitz, A. M. (2000) *Science* **289**, 452–457
167. Cocking, E. C., and Davey, M. R. (1987) *Science* **236**, 1259–1262

References

168. Smith, H. O., Danner, D. B., and Deich, R. A. (1981) *Ann. Rev. Biochem.* **50**, 41–68
169. Wahl, G. M., de Saint Vincent, B. R., and DeRose, M. L. (1984) *Nature (London)* **307**, 516–520
170. Felgner, P. L., and Ringold, G. M. (1989) *Nature (London)* **337**, 387–388
171. Miller, J. F., Dower, W. J., and Tompkins, L. S. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 856–860
172. Solioz, M., and Bienz, D. (1990) *Trends Biochem. Sci.* **15**, 175–177
173. Hengen, P. N. (1995) *Trends Biochem. Sci.* **20**, 248–249
174. Collombet, J.-M., Wheeler, V. C., Vogel, F., and Coutelle, C. (1997) *J. Biol. Chem.* **272**, 5342–5347
175. Chang, D. C., Hunt, J. R., Zheng, Q., and Gao, P. Q. (1992) in *Guide to Electroporation and Electrofusion* (Chang, D. C., ed), pp. 303–326, Academic Press, San Diego, California
176. Klobutcher, L. A., and Ruddle, F. H. (1981) *Ann. Rev. Biochem.* **50**, 533–554
177. Klein, T. M., Wolf, E. D., Wu, R., and Sanford, J. C. (1987) *Nature (London)* **327**, 70–73
178. Butow, R. A., and Fox, T. D. (1990) *Trends Biochem. Sci.* **15**, 465–468
179. Kohli, A., Leech, M., Vain, P., Laurie, D. A., and Christou, P. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7203–7208
180. Hohn, B., and Murray, K. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3259–3263
181. Lutz, C. T., Hollifield, W. C., Seed, B., Davie, J. M., and Huang, H. V. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4379–4383
182. Wyman, A. R., Wolfe, L. B., and Botstein, D. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2880–2884
183. Okayama, H., and Berg, P. (1982) *Mol. Cell. Biochem.* **2**, 161–170
184. Gubler, U., and Hoffman, B. J. (1983) *Gene* **25**, 263–269
185. Mason, W. T., ed. (1999) *Fluorescent and Luminescent Probes*, 2nd ed., Academic Press, San Diego, California
186. Keller, G. H., and Manak, M. M. (1989) *DNA Probes*, Stockton Press, New York
186a. Santi, E., Capone, S. Mennuni, C., Lahm, A., Tramontano, A, Luzzago, A., and Nicosia, A. (2000) *J. Mol. Biol.* **296**, 497 – 508
187. Little, P. F. R., and Cross, S. H. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3159–3163
188. Richards, J. E., Gilliam, T. C., Cole, J. L., Drumm, M. L., Wasmuth, J. J., Gusella, J. F., and Collins, F. S. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6437–6441
189. Housman, D. (1995) *N. Engl. J. Med.* **332**, 318–320
190. Viegas-Pequignot, E., Dutrilleax, B., Magdelenat, H., and Coppey-Moisam, M. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 582–586
191. Abbott, C., and Povey, S. (1995) *Somatic Cell Hybrids*, IRL Press, Oxford
192. Williamson, B. (1981) *Nature (London)* **293**, 10–11
193. White, R., and Lalouel, J.-M. (1988) *Sci. Am.* **258**(Feb), 40–48
194. Gusella, J. F. (1986) *Ann. Rev. Biochem.* **55**, 831–854
195. Davies, K. E., ed. (1986) *Human Genetic Diseases*, IRL Press, Washington, D.C.
196. Martin, J. B. (1987) *Science* **238**, 765–772
197. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N. (1985) *Science* **230**, 1350–1354
198. Gusella, J. F., Tanzi, R. E., Bader, P. I., Phelan, M. C., Stevenson, R., Hayden, M. R., Hofman, K. J., Faryniarz, A. G., and Gibbons, K. (1985) *Nature (London)* **318**, 75–78
199. Gilliam, T. C., Bucan, M., MacDonald, M. E., Zimmer, M., Haines, J. L., Cheng, S. V., Pohl, T. M., Meyers, R. H., Whaley, W. L., Allitto, B. A., Faryniarz, A., Wasmuth, J. J., Frischauf, A.-M., Conneally, P. M., Lehrach, H., and Gusella, J. F. (1987) *Science* **238**, 950–952
200. Royer-Pokora, B., Kunkel, L. M., Monaco, A. P., Goff, S. C., Newburger, P. E., Baehner, R. L., Cole, F. S., Curnutte, J. T., and Orkin, S. H. (1986) *Nature (London)* **322**, 32–38
201. Reich, N. O., and Mashboon, N. (1993) *J. Biol. Chem.* **268**, 9191–9193
202. Lee, W.-H., Shew, J.-Y., Hong, F. D., Sery, T. W., Donoso, L. A., Young, L.-J., Bookstein, R., and Lee, E. Y.-H. P. (1987) *Nature (London)* **329**, 642–645
203. Thomson, G., and Esposito, M. S. (1999) *Trends Biochem. Sci.* **24**, M17–M20
204. Little, P. (1986) *Nature (London)* **321**, 558–559
205. Zaccolo, M., Williams, D. M., Brown, D. M., and Gherardi, E. (1996) *J. Mol. Biol.* **255**, 589–603
206. Myers, R. M., Lerman, L. S., and Maniatis, T. (1985) *Science* **229**, 242–247
207. Botstein, D., and Shortle, D. (1985) *Science* **229**, 1193–1201
208. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367–382
209. Zoller, M. J., and Smith, M. (1983) *Methods Enzymol.* **100**, 468–500
210. Higuchi, R., Krummel, B., and Saiki, R. (1988) *Nucleic Acids Res.* **16**, 7351–7367
211. Jones, D. H., and Howard, B. H. (1990) *BioTechniques* **8**, 178–183
212. Jones, D. H., Sakamoto, K., Vorce, R. L., and Howard, B. H. (1990) *Nature (London)* **344**, 793–794
213. Wells, J. A., and Powers, D. B. (1986) *J. Biol. Chem.* **261**, 6564–6570
214. Shuldiner, A. R. (1996) *N. Engl. J. Med.* **334**, 653–655
215. Jasin, M., Moynahan, M. E., and Richardson, C. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8804–8808
216. Capecchi, M. R. (1994) *Sci. Am.* **270**(Mar), 52–59
217. Travis, J. (1992) *Science* **256**, 1392–1394
218. Barinaga, M. (1994) *Science* **265**, 26–28
219. Weissmann, C., and Aguzzi, A. (1999) *Science* **286**, 914–915
219a. Smith, J. A., and Martin, L. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1263–1267
220. Nasmyth, K. (2002) *Science* **297**, 559–565
220a. Hauf, S., Waizenegger, I. C., and Peters, J.-M. (2001) *Science* **293**, 1320–1323
220b. Campbell, J. L., and Cohen-Fix, O. (2002) *Trends Biochem. Sci.* **27**, 492–495
220c. Seemann, J., Pypaert, M., Taguchi, T., Malsam, J., and Warren, G. (2002) *Science* **295**, 848–851
221. Mazia, D. (1961) *Sci. Am.* **205**(Sep), 101–120
222. Earnshaw, W. C., and Mackay, A. M. (1994) *FASEB J.* **8**, 947–956
223. Russell, P. (1998) *Trends Biochem. Sci.* **23**, 399–402
224. Piwnica-Worms, H. (1999) *Nature (London)* **401**, 535–537
225. Nasmyth, K. (1999) *Trends Biochem. Sci.* **24**, 98–104
225a. Cortez, D., and Elledge, S. J. (2000) *Nature (London)* **406**, 354–356
226. Warren, G. (1985) *Trends Biochem. Sci.* **10**, 439–443
226a. Lippincott-Schwartz, J. (2002) *Nature (London)* **416**, 31–32
227. Morgan, D. O. (1995) *Nature (London)* **374**, 131–133
227a. Kong, M., Barnes, E. A., Ollendorff, V., and Donoghue, D. J. (2000) *EMBO J.* **19**, 1378–1388
227b. Healy, J. M. S., Menges, M., Doonan, J. H., and Murray, J. A. H. (2001) *J. Biol. Chem.* **276**, 7041–7047
228. Pavletich, N. P. (1999) *J. Mol. Biol.* **287**, 821–828
228a. Nash, P., Tang, X., Orlicky, S., Chen, Q., Gertler, F. B., Mendenhall, M. D., Sicheri, F., Pawson, T., and Tyers, M. (2001) *Nature (London)* **414**, 514–521
228b. Bartek, J., and Lukas, J. (2001) *Science* **294**, 66–67
228c. Schwab, M., and Tyers, M. (2001) *Nature (London)* **413**, 268–269
228d. Zhou, B.-B. S., and Elledge, S. J. (2000) *Nature (London)* **408**, 433–439
229. Fotedar, A., Cannella, D., Fitzgerald, P., Rousselle, T., Gupta, S., Dorée, M., and Fotedar, R. (1996) *J. Biol. Chem.* **271**, 31627–31637
230. De Bondt, H. L., Rosenblatt, J., Jancarik, J., Jones, H. D., Morgan, D. O., and Kim, S.-H. (1993) *Nature (London)* **363**, 595–602
231. Jeffrey, P. D., Russo, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massagué, J., and Pavletich, N. P. (1995) *Nature (London)* **376**, 313–320
232. Shiffman, D., Brooks, E. E., Brooks, A. R., Chan, S. C., and Milner, P. G. (1996) *J. Biol. Chem.* **271**, 12199–12204
233. Nurse, P. (1990) *Nature (London)* **344**, 503–507
234. Pines, J. (1993) *Trends Biochem. Sci.* **18**, 195–197
235. Solomon, M. J. (1994) *Trends Biochem. Sci.* **19**, 496–500
236. Lew, J., and Wang, J. H. (1995) *Trends Biochem. Sci.* **20**, 33–37
236a. Crawford, D. F., and Piwnica-Worms, H. (2001) *J. Biol. Chem.* **276**, 37166–37177
236b. Karsenti, E., and Vernos, I. (2001) *Science* **294**, 543–547
236c. Kastan, M. B. (2001) *Nature (London)* **410**, 766–767
236d. Davenport, R. J. (2001) *Science* **292**, 2415–2417
237. Morris, M. C., and Divita, G. (1999) *J. Mol. Biol.* **286**, 475–487
237a. Jacobs, H. W., Keidel, E., and Lehner, C. F. (2001) *EMBO J.* **20**, 2376–2386
238. Lopez-Girona, A., Furnari, B., Mondesert, O., and Russell, P. (1999) *Nature (London)* **397**, 172–175
238a. Yang, Q., Manicone, A., Coursen, J. D., Linke, S. P., Nagashima, M., Forgues, M., and Wang, X. W. (2000) *J. Biol. Chem.* **275**, 36892–36898
239. Reynolds, R. A., Yem, A. W., Wolfe, C. L., Deibel, M. R., Jr., Chidester, C. G., and Watenpaugh, K. D. (1999) *J. Mol. Biol.* **293**, 559–568
240. Hagting, A., Karlsson, C., Clute, P., Jackman, M., and Pines, J. (1998) *EMBO J.* **17**, 4127–4138
240a. Qian, Y.-W., Erikson, E., and Maller, J. L. (1998) *Science* **282**, 1701–1704
240b. Körner, K., Jérôme, V., Schmidt, T., and Müller, R. (2001) *J. Biol. Chem.* **276**, 9662–9669
241. Pines, J. (1999) *Nature (London)* **397**, 104 – 105
242. Sanchez, Y., Bachant, J., Wang, H., Hu, F., Liu, D., Tetzlaff, M., and Elledge, S. J. (1999) *Science* **286**, 1166–171
243. Kellogg, D. R., Moritz, M., and Alberts, B. M. (1994) *Ann. Rev. Biochem.* **63**, 639–674
244. Geier, B. M., Wiech, H., and Schiebel, E. (1996) *J. Biol. Chem.* **271**, 28366–28374
244a. Murray, A. W. (2001) *Science* **291**, 1499–1502
244b. Hinchcliffe, E. H., Miller, F. J., Cham, M., Khodjakov, A., and Sluder, G. (2001) *Science* **291**, 1547–1550
244c. Middendorp, S., Paoletti, A., Schiebel, E., and Bornens, M. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9141–9146
244d. do Carmo Avides, M., and Glover, D. M. (1999) *Science* **283**, 1733–1735

References

244e. Flory, M. R., Moser, M. J., Monnat, R. J., Jr., and Davis, T. N. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5919–5923
245. Nicklas, R. B. (1997) *Science* **275**, 632–637
246. Heald, R., Tournebize, R., Blank, T., Sandaltzopoulos, R., Becker, P., Hyman, A., and Karsenti, E. (1996) *Nature (London)* **382**, 420–425
246a. Fraschini, R., Beretta, A., Sironi, L., Musacchio, A., Lucchini, G., and Piatti, S. (2001) *EMBO J.* **20**, 6648–6659
247. Hyams, J. (1996) *Nature (London)* **382**, 397–398
247a. Bloom, K. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4757–4759
248. Dasso, M. (1993) *Trends Biochem. Sci.* **18**, 96–101
249. Ohba, T., Nakamura, M., Nishitani, H., and Nishimoto, T. (1999) *Science* **284**, 1356–1358
250. Wilde, A., and Zheng, Y. (1999) *Science* **284**, 1359–1362
250a. Gruneberg, U., Campbell, K., Simpson, C., Grindlay, J., and Schiebel, E. (2000) *EMBO J.* **19**, 6475–6488
250b. Martin-Lluesma, S., Stucke, V. M., and Nigg, E. A. (2002) *Science* **297**, 2267–2270
251. Barton, N. R., and Goldstein, L. S. B. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1735–1742
252. Lockhart, A., and Cross, R. A. (1996) *Biochemistry* **35**, 2365–2373
253. Pennisi, E. (1998) *Science* **279**, 477–478
254. Elledge, S. J. (1998) *Science* **279**, 999–1000
255. Hwang, L. H., Lau, L. F., Smith, D. L., Mistrot, C. A., Hardwick, K. G., Hwang, E. S., Amon, A., and Murray, A. W. (1998) *Science* **279**, 1041–1044
256. Yamano, H., Gannon, J., and Hunt, T. (1996) *EMBO J.* **15**, 5268–5279
257. Bachant, J. B., and Elledge, S. J. (1999) *Nature (London)* **398**, 757–758
258. Funabiki, H., Kumada, K., and Yanagida, M. (1996) *EMBO J.* **15**, 6617–6628
259. Page, A. M., and Hieter, P. (1999) *Ann. Rev. Biochem.* **68**, 583–609
259a. Prinz, S., and Amon, A. (1999) *Nature (London)* **402**, 133–135
259b. Nakaseko, Y., and Yanagida, M. (2001) *Nature (London)* **412**, 291–292
259c. Morgan, D. O., and Roberts, J. M. (2002) *Nature (London)* **418**, 495–496
260. Grossberger, R., Gieffers, C., Zachariae, W., Podtelejnikov, A. V., Schleiffer, A., Nasmyth, K., Mann, M., and Peters, J.-M. (1999) *J. Biol. Chem.* **274**, 14500–14507
261. Jiang, F., and Basavappa, R. (1999) *Biochemistry* **38**, 6471–6478
261a. Piel, M., Nordberg, J., Euteneuer, U., and Bornens, M. (2001) *Science* **291**, 1550–1553
262. Wisdom, R., Johnson, R. S., and Moore, C. (1999) *EMBO J.* **18**, 188–197
263. Takenaka, K., Moriguchi, T., and Nishida, E. (1998) *Science* **280**, 599–602
263a. Wilkinson, M. G., and Millar, J. B. A. (2000) *FASEB J.* **14**, 2147–2157
263b. Graves, L. M. (2000) *Nature (London)* **403**, 328–332
263c. Whitmarsh, A. J., and Davis, R. J. (2000) *Nature (London)* **403**, 255–256
264. Gottesfeld, J. M., and Forbes, D. J. (1997) *Trends Biochem. Sci.* **22**, 197–202
264a. Goldstein, L. S. B. (2001) *Science* **291**, 2102–2103
265. Martínez-Botas, J., Suárez, Y., Ferruelo, A. J., Gómez-Coronado, D., and Lasunción, M. A. (1999) *FASEB J.* **13**, 1359–1370
266. Bourne, Y., Arvai, A. S., Bernstein, S. L., Watson, M. H., Reed, S. I., Endicott, J. E., Noble, M. E., Johnson, L. N., and Tainer, J. A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10232–10236
267. Parge, H. E., Arvai, A. S., Murtari, D. J., Reed, S. I., and Tainer, J. A. (1993) *Science* **262**, 387–395
267a. Rubin, G. M., and Lewis, E. B. (2000) *Science* **287**, 2216–2218
268. Yamamoto, M. (1996) *Trends Biochem. Sci.* **21**, 18–22
269. de Lange, T. (1998) *Nature (London)* **392**, 753–754
270. Conrad, M. N., Dominguez, A. M., and Dresser, M. E. (1997) *Science* **276**, 1252–1255
271. McKim, K. S., Green-Marroquin, B. L., Sekelsky, J. J., Chin, G., Steinberg, C., Khodosh, R., and Hawley, R. S. (1998) *Science* **279**, 876–878
271a. Lindgren, A., Bungard, D., Pierce, M., Xie, J., Vershon, A., and Winter, E. (2000) *EMBO J.* **19**, 6489–6497
272. Orr-Weaver, T. L. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10443–10449
273. McKim, K. S., and Hawley, R. S. (1995) *Science* **270**, 1595–1601
274. Picard, A., Galas, S., Peaucellier, G., and Dorée, M. (1996) *EMBO J.* **15**, 3590–3598
275. Lenormand, J.-L., Dellinger, R. W., Knudsen, K. E., Subramani, S., and Donoghue, D. J. (1999) *EMBO J.* **18**, 1869–1877
276. Sette, C., Barchi, M., Bianchini, A., Conti, M., Rossi, P., and Geremia, R. (1999) *J. Biol. Chem.* **274**, 33571–33579
276a. Xu, R., Wilson, W. K., and Matsuda, S. P. T. (2002) *J. Am. Chem. Soc.* **124**, 918–919
277. Watanabe, Y., and Nurse, P. (1999) *Nature (London)* **400**, 461–464
278. Roeder, G. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10450–10456
278a. Albini, S. M., and Jones, G. H. (1987) *Chromosoma* **95**, 324–328
279. Kleckner, N. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8167–8174
279a. Nabeshima, K., Kakihara, Y., Hiraoka, Y., and Nojima, H. (2001) *EMBO J.* **20**, 3871–3881
279b. Yuan, L., Liu, J.-G., Hoja, M.-R., Wilbertz, J., Nordqvist, K., and Höög, C. (2002) *Science* **296**, 1115–1118
279c. Gerton, J. L., DeRisi, J., Shroff, R., Lichten, M., Brown, P. O., and Petes, T. D. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11383–11390
279d. Stack, S. M. (1991) *Genome* **34**, 900–908
280. Boy de la Tour, E., and Laemmli, U. K. (1988) *Cell* **55**, 937–944
281. Gallardo, M. H., Bickham, J. W., Honeycutt, R. L., Ojeda, R. A., and Köhler, N. (1999) *Nature (London)* **401**, 341
281a. Shonn, M. A., McCarroll, R., and Murray, A. W. (2000) *Science* **289**, 300–303
281b. Sluder, G., and McCollum, D. (2000) *Science* **289**, 254–255
282. Pirrotta, V. (1984) *Trends Biochem. Sci.* **9**, 220–221
283. Burtis, K. C., and Hawley, R. S. (1999) *Nature (London)* **401**, 125–127
284. Kao, F.-t, and Yu, J.-w. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1844–1848
285. Sager, R. (1965) *Sci. Am.* **212**(Jan), 71–79
286. Goodenough, U. W., and Levine, R. P. (1970) *Sci. Am.* **223**(Nov), 22–29
287. Nagata, S., Taira, H., Hall, A., Johnsrud, L., Streuli, M., Ecsodi, J., Boll, W., Cantell, K., and Weissmann, C. (1980) *Nature (London)* **284**, 316–320
288. Olson, M. V. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4338–4344
289. Collins, F. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10821–10823
290. Guyer, M. S., and Collins, F. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10841–10848
290a. International Human Genome Sequencing Consortium. (2001) *Nature (London)* **409**, 860–921
290b. Pennisi, E. (2001) *Science* **291**, 1177–1180
290c. Venter, J. C., and many others. (2001) *Science* **291**, 1304–1351
290d. Olivier, M., and 54 other authors. (2001) *Science* **291**, 1298–1302
290e. Shoemaker, D. D., and 34 other authors. (2001) *Nature (London)* **409**, 922–927
290f. Caron, H., and 12 other authors. (2001) *Science* **291**, 1289–1292
290g. Deloukas, P., and 126 other authors. (2001) *Nature (London)* **414**, 865–871
290h. Claverie, J.-M. (2001) *Science* **291**, 1255–1257
290i. Pennisi, E. (2002) *Science* **296**, 1600–1601
290j. Kim, S. K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J. M., Eizinger, A., Wylie, B. N., and Davidson, G. S. (2001) *Science* **293**, 2087–2092
291. Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., and Walters, L. (1998) *Science* **282**, 682–689
291a. Malakoff, D. (2000) *Science* **288**, 248–253
291b. The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. (2001) *Nature (London)* **409**, 685–690
291c. The International Mouse Mutagenesis Consortium. (2001) *Science* **291**, 1251–1255
291d. Gregory, S. G., and 85 other authors (2002) *Nature (London)* **418**, 743–750
291e. Mural, R. J., and 176 other authors (2002) *Science* **296**, 1661–1671
292. Roach, J. C., Siegel, A. F., van den Engh, G., Trask, B., and Hood, L. (1999) *Nature (London)* **401**, 843–845
292a. Marshall, E. (2001) *Science* **291**, 1872
292b. Vogel, G. (2000) *Science* **290**, 1671
292c. Aparicio, S., and 40 other authors (2002) *Science* **297**, 1301–1310
292d. Gardner, M. J., and 44 other authors (2002) *Nature (London)* **419**, 498–511
292e. Doolittle, R. F. (2002) *Nature (London)* **419**, 493–494
292f. Carlton, J. M., and 43 other authors (2002) *Nature (London)* **419**, 512–519
292g. Holt, R. A., and 121 other authors (2002) *Science* **298**, 129–149
292h. Kaufman, T. C., Severson, D. W., and Robinson, G. E. (2002) *Science* **298**, 97–115
293. White, R., and Caskey, C. T. (1988) *Science* **240**, 1483–1488
294. McKusick, V. A., and Amberger, J. S. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 119–125, McGraw-Hill, New York
295. Sawyer, J. R., and Hozier, J. C. (1986) *Science* **232**, 1632–1635
296. Tunnacliffe, A., Benham, F., and Goodfellow, P. (1984) *Trends Biochem. Sci.* **9**, 5–7
297. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. (1994) *Molecular Biology of the Cell*, 3rd ed., Garland, New York
298. Franke, U. (1981) *Cytogenet. Cell Genet.* **31**, 24–32
299. Gray, J. W., and Langlois, R. G. (1986) *Ann. Rev. Biophys. Biophys. Chem.* **15**, 195–235
300. Chumakov, I., and plus 35 other authors (1992) *Nature (London)* **359**, 380–387
301. Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8794–8797
302. Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith, H. O., and Hunkapiller, M. (1998) *Science* **280**, 1540–1542
303. Marx, J. (1995) *Science* **270**, 1919–1920

References

304. Dib, C., Fauré, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J., and Weissenbach, J. (1996) *Nature (London)* **380**, 152–154, iii–v + supplement
305. Rogers, J. (1999) *Science* **286**, 429–432
305a. Liu, S., Ren, H., Gao, Q., Roach, D. J., Loder, R. T., Jr., Armstrong, T. M., Mao Q, Blaga, I., Barker, D. L., and Jovanovich, S. B. (2000) *Proc. Natl. Acad. Sci., U.S.A.* **97**, 5369 – 5374
305b. Paegel, B. M., Emrich, C. A., Wedemayer, G. J., Scherer, J. R., and Mathies, R. A. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 574 – 579
306. Marshall, E. (1999) *Science* **284**, 1906–1909
306a. Green, P. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4143 – 4144
307. Donis-Keller, H., and 32 other authors (1987) *Cell* **51**, 319–337
308. Wang, D. G., and 26 other authors (1998) *Science* **280**, 1077–1082
308a. The International SNP Map Working Group. (2001) *Nature (London)* **409**, 928–933
308b. Patil, N., and 21 other authors (2001) *Science* **294**, 1719–1723
309. Hudson, T. J., and 50 other authors (1995) *Science* **270**, 1945–1954
310. Schuler, G. D., and 103 other authors (1996) *Science* **274**, 540–546
311. Wang, S. M., and Rowley, J. D. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11909–11914
312. Dunham, I., Shimizu, N., Roe, B. A., Chossoe, S., and 117 other authors (1999) *Nature (London)* **402**, 489–495
313. Butler, D. (1999) *Nature (London)* **402**, 447–448
313a. The International Human Genome Mapping Consortium. (2001) *Nature (London)* **409**, 934–941
314. Pandey, A., and Lewitter, F. (1999) *Trends Biochem. Sci.* **24**, 276–280
315. Pennisi, E. (1999) *Science* **286**, 447–450
316. Baxevanis, A. D., and Ouellette, B. F. F., eds. (1998) *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Wiley, New York
317. Bishop, M. J., ed. (1999) *Genetics Databases*, Academic Press, San Diego, California
318. Durbin, R., Eddy, S., Krogh, A., and Mitchinson, G., eds. (1998) *Biological Sequence Analysis*, Cambridge Univ. Press., Cambridge
319. Bishop, M. J., ed. (1998) *Human Genome Computing*, 2nd ed., Academic Press, San Diego, California
320. Calvet, J. P. (1998) *Science* **282**, 1057–1058
321. Kereil, D. P., and Etzold, T. (1999) *Trends Biochem. Sci.* **24**, 155–157
321a. Tamames, J., and Tramontano, A. (2000) *Trends Biochem. Sci.* **25**, 402–403
321b. Jones, D. T., and Swindells, M. B. (2002) *Trends Biochem. Sci.* **27**, 161–164
321c. Zafar, N., Mazumder, R., and Seto, D. (2001) *Trends Biochem. Sci.* **26**, 514–516
322. Boguski, M. S. (1995) *N. Engl. J. Med.* **333**, 645–647
323. Borodovsky, M., Koonin, E. V., and Rudd, K. E. (1994) *Trends Biochem. Sci.* **19**, 309–313
324. Brown, N. P., Whittaker, A. J., Newell, W. R., Rawlings, C. J., and Beck, S. (1995) *J. Mol. Biol.* **249**, 342–359
325. Pennisi, E. (1998) *Science* **280**, 1692–1693
325a. Matthiessen, M. W. (2002) *Trends Biochem. Sci.* **27**, 586 – 588
326. Little, P. (1999) *Nature (London)* **402**, 467–468
327. O'Brien, S. J., Menotti-Raymond, M., Murphy, W. J., Nash, W. G., Wienberg, J., Stanyon, R., Copeland, N. G., Jenkins, N. A., Womack, J. E., and Graves, J. A. M. (1999) *Science* **286**, 458–481
327a. Gura, T. (2001) *Science* **293**, 593–595
327b. Buetow, K. H., Edmonson, M., MacDonald, R., Clifford, R., Yip, P., Kelley, J., Little, D. P., Strausberg, R., Koester, H., Cantor, C. R., and Braun, A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 581–584
327c. Couzin, J. (2002) *Science* **296**, 1391,1393
327d. Adam, D. (2001) *Nature (London)* **412**, 105
327e. Yan, H., Kinzler, K. W., and Vogelstein, B. (2000) *Science* **289**, 1890–1892
327f. Wright, A. F., and Van Heyningen, V. (2001) *Nature (London)* **414**, 705–706
327g. Marshall, E. (2001) *Science* **294**, 2272–2274
327h. Trager, R. S. (2002) *Science* **298**, 947
328. Pääbo, S. (1999) *Trends Biochem. Sci.* **24**, M13–M16
328a. Templeton, A. R. (2002) *Nature (London)* **416**, 45–51
329. Olsen, G. J., and Woese, C. R. (1993) *FASEB J.* **7**, 113–123
330. Shenk, M. A., and Steele, R. E. (1993) *Trends Biochem. Sci.* **18**, 459–463
331. Harris, E. E., and Hey, J. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3320–3324
331a. Adcock, G. J., Dennis, E. S., Easteal, S., Huttley, G. A., Jermiin, L. S., Peacock, W. J., and Thorne, A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 537–542
331b. Relethford, J. H. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 390–391
331c. Jones, M. (2001) *The Molecule Hunt: Archaeology and the Search for Ancient DNA*, Allen Lane,
331d. Fish, S. A., Shepherd, T. J., McGenity, T. J., and Grant, W. D. (2002) *Nature (London)* **417**, 432–436
332. Kahn, P., and Gibbons, A. (1997) *Science* **277**, 176–178
332a. Gibbons, A. (2001) *Science* **292**, 627–629
332b. Fujiyama, A., Watanabe, H., Toyoda, A., Taylor, T. D., Itoh, T., Tsai, S.-F., and 11 other authors. (2002) *Science* **295**, 131–134
332c. Enard, W., Khaitovich, P., Klose, J., Zöllner, S., and 9 other authors. (2002) *Science* **296**, 340–343
332d. Cyranoski, D. (2002) *Nature (London)* **418**, 910–912
332e. Stone, A. C., Griffiths, R. C., Zegura, S. L., and Hammer, M. F. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 43–48
332f. Britten, R. J. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13633–13635
333. Gibbons, A. (1998) *Science* **281**, 1432–1434
334. Kennedy, E. P. (1992) *Ann. Rev. Biochem.* **61**, 1–28
335. Barbujani, G., Magagni, A., Minch, E., and Cavalli-Sforza, L. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4516–4519
336. Molnar, S. (1998) *Human Variation, Races, Types and Ethnic Groups*, 4th ed., Prentice Hall, Upper Saddle River, New Jersey (pp. 230–247)
336a. Jablonski, N. G., and Chaplin, G. (2002) *Sci. Am.* **287**(Oct), 74–81
337. Rannala, B., and Mountain, J. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9197–9201
338. Sugimoto, C., Kitamura, T., Guo, J., Al-Ahdal, M. N., Shchelkunov, S. N., Otova, B., Ondrejka, P., Chollet, J.-Y., El-Safi, S., Ettayebi, M., Grésenguet, G., Kocagöz, T., Chaiyarasamee, S., Thant, K. Z., Thein, S., Moe, K., Kobayashi, N., Taguchi, F., and Yogo, Y. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9191–9196
338a. Pääbo, S. (2001) *Science* **291**, 1219–1220
338b. Pennisi, E. (2001) *Science* **291**, 1733–1734
338c. Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J., and Feldman, M. W. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7360–7365
338d. Gibbons, A. (2000) *Science* **290**, 1080–1081
338e. Cann, R. L. (2001) *Science* **291**, 1742–1747
339. De Mendoza, D. H., and Braginski, R. (1999) *Science* **283**, 1439–1440
340. Bradley, D. G., MacHugh, D. E., Cunningham, P., and Loftus, R. T. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5131–5135
341. Makalowski, W., and Boguski, M. S. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9407–9412
342. Nadeau, J. H., Grant, P. L., Mankala, S., Reiner, A. H., Richardson, J. E., and Eppig, J. T. (1995) *Nature (London)* **373**, 363–365
343. Collins, F. S., and Jegalian, K. G. (1999) *Sci. Am.* **281**(Dec), 86–91
344. O'Brien, S. J., and Stanyon, R. (1999) *Nature (London)* **402**, 365–366
344a. Hedges, S. B., and Kumar, S. (2002) *Science* **297**, 1283 – 1285
344b. Knight, J. (2002) *Nature (London)* **417**, 374–376
345. Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) *Science* **278**, 631–637
346. Gibbons, A. (1998) *Science* **280**, 675–676
347. Rowe, T. (1999) *Nature (London)* **398**, 283–284
348. Bromham, L., Rambaut, A., Fortey, R., Cooper, A., and Penny, D. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12386–12389
349. Knoll, A. H. (1999) *Science* **285**, 1025–1026
350. Maniloff, J. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10004–10006
351. Lawrence, J. G., and Ochman, H. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9413–9417
352. Doolittle, W. F. (1999) *Trends Biochem. Sci.* **24**, M5–M8
353. Lake, J. A., Jain, R., and Rivera, M. C. (1999) *Science* **283**, 2027–2028
354. Pennisi, E. (1998) *Science* **280**, 672–674
355. Stein, J. L., and Simon, M. I. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6228–6230
356. Davies, J. (1999) *Trends Biochem. Sci.* **24**, M2–M5
356a. Doolittle, R. F. (2002) *Nature (London)* **416**, 697–700
356b. Bentley, S. D. and 42 other authors (2002) *Nature (London)* **417**, 141–147
356c. Barnett, M. J., and 25 other authors (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9883–9888
356d. Galibert, F., and 55 other authors (2001) *Science* **293**, 668–672
356e. Heidelberg, J. F., and 31 other authors (2000) *Nature (London)* **406**, 477–483
356f. Parkhill, J., and 40 other authors (2001) *Nature (London)* **413**, 848–852
356g. Parkhill, J., and 34 other authors (2001) *Nature (London)* **413**, 523–527
356h. Moreno, E., and Moriyón, I. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1–3
356i. Cole, S. T., and 43 other authors (2001) *Nature (London)* **409**, 1007–1011
356j. Enserink, M. (2002) *Science* **295**, 1442–1443
356k. Tettelin, H., and 38 other authors (2001) *Science* **293**, 498–506
356l. Stover, C. K., and 30 other authors (2000) *Nature (London)* **406**, 959–964
356m. Glaser, P., and 54 other authors (2001) *Science* **294**, 849–853
356n. Perna, N. T., and 27 other authors (2001) *Nature (London)* **409**, 529–533
356o. Glass, J. I., Lefkowitz, E. J., Glass, J. S., Heiner, C. R., Chen, E. Y., and Cassell, G. H. (2000) *Nature (London)* **407**, 757–761
356p. Nierman, W. C., and 36 other authors (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4136–4141
356q. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., and Ishikawa, H. (2000) *Nature (London)* **407**, 81–86
356r. Ng, W. V., Kennedy, S. P., Mahairas, G. G., and 40 other authors. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12176 – 12181

References

356s. Kawashima, T., Amano, N., Koike, H., Makino, S.-i, Higuchi, S., and 10 other authors. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14257– 14262

356t. Cole, S. T., and 41 other authors. (1998) *Nature (London)* **393**, 537 – 544

357. Edwards, J. S., and Palsson, B. O. (1999) *J. Biol. Chem.* **274**, 17410 – 17416

358. Strauss, E. J., and Falkow, S. (1997) *Science* **276**, 707 – 711

358a. da Silva, A. C. R., and 64 other authors. (2002) *Nature (London)* **417**, 459 – 463

359. Karp, P. D. (1998) *Trends Biochem. Sci.* **23**, 114 – 116

359a. Covert, M. W., Schilling, C. H., Famili, I., Edwards, J. S., Goryanin, I. I., Selkov, E., and Palsson, B. O. (2001) *Trends Biochem. Sci* **26**, 179 – 186

360. Hutchison, C. A., III, Peterson, S. N., Gill, S. R., Cline, R. T., White, O., Fraser, C. M., Smith, H. O., and Venter, J. C. (1999) *Science* **286**, 2165 – 2169

361. Mushegian, A. R., and Koonin, E. V. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10268 – 10273

361a. Szostak, J. W., Bartel, D. P., and Luisi, P. L., (2001) *Nature (London)* **409**, 387 – 390

362. Cho, M. K., Magnus, D., Caplan, A. L., McGee, D., and Ethics of Genomics Group. (1999) *Science* **286**, 2087 – 2090

362a. Stephanopoulos, G., and Kelleher, J. (2001) *Science* **292**, 2024 – 2025

362b. Holden, C. (2002) *Science* **297**, 1459 – 1460

363. Winzeler, E. A., and 51 other authors (1999) *Science* **285**, 901 – 906

364. Hieter, P. (1999) *Nature (London)* **402**, 362 – 363

364a. Ross-Macdonald, P., and 17 other authors (1999) *Nature (London)* **402**, 413 – 418

365. Glaever, G., and 72 other authors. (2002) *Nature (London)* **418**, 387 – 391

365a. Eisen, J. A. (2002) *Nature (London)* **415**, 845 – 848

365b. Wood, V., and 132 other authors (2002) *Nature (London)* **415**, 871– 880

365c. Seo, H.-C., Kube, M., Edvardsen, R. B., Jensen, M. F., Beck, A., Spriet, E., Gorsky, G., Thompson, E. M., Leharch, H., Reinhardt, R., and Chourrout, D. (2001) *Science* **294**, 2506

366. Fields, S., Kohara, Y., and Lockhart, D. J. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8825 – 8826

367. Tabara, H., Motohashi, T., and Kohara, Y. (1996) *Nucleic Acids Res.* **24**, 2119 – 2124

368. DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997) *Science* **278**, 680 – 686

369. Walhout, A. J. M., Sordella, R., Lu, X., Hartley, J. L., Temple, G. F., Brasch, M. A., Thierry-Mieg, N., and Vidal, M. (2000) *Science* **287**, 116 – 122

370. Hutter, H., Vogel, B. E., Plenefisch, J. D., Norris, C. R., Proenca, R. B., Spieth, J., Guo, C., Mastwal, S., Zhu, X., Scheel, J., and Hedgecock, E. M. (2000) *Science* **287**, 989 – 994

370a. Zdobnov, E. M., and 35 other authors. (2002) *Science* **298**, 149 – 159

370b. Miki, R., and 29 other authors. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2199 – 2204

371. Lin, X., and 36 other authors (1999) *Nature (London)* **402**, 761 – 768

372. European Union Arabidopsis Genome Sequencing Consortium, and Cold Spring Harbor, W. U. in St. L. and P. E. B. A. S. C. (1999) *Nature (London)* **402**, 769 – 777

373. Mathé, C., Peresetsky, A., Déhais, P., Van Montagu, M., and Rouzé, P. (1999) *J. Mol. Biol.* **285**, 1977 – 1991

374. Gura, T. (2000) *Science* **287**, 412 – 414

375. Meyerowitz, E. M. (1999) *Nature (London)* **402**, 731 – 732

375a. The Arabidopsis Genome Initiative. (2000) *Nature (London)* **408**, 796 – 815

375b. Seki, M., and 19 other authors (2002) *Science* **296**, 141– 145

375c. Pennisi, E. (2000) *Science* **290**, 32 – 35

375d. Normile, D., and Pennisi, E. (2002) *Science* **296**, 32 – 39

375e. Bennetzen, J. (2002) *Science* **296**, 60 – 63

375f. Yu, J., and 99 other authors (2002) *Science* **296**, 79 – 92

375g. Goff, S. A., and 54 other authors (2002) *Science* **296**, 92 – 100

376. Somerville, C., and Somerville, S. (1999) *Science* **285**, 380 – 383

377. Pennisi, E. (1998) *Science* **282**, 652 – 654

378. Capy, P. (2000) *Science* **287**, 985 – 986

379. Scriver, C. R., Kaufman, S., Eisensmith, R. C., and Woo, S. L. C. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1015 – 1075, McGraw-Hill, New York

380. Barinaga, M. (1998) *Science* **281**, 32 – 34

381. Schultheis, P. J., Lorenz, J. N., Meneton, P., Nieman, M. L., Riddle, T. M., Flagella, M., Duffy, J. J., Doetschman, T., Miller, M. L., and Shull, G. E. (1998) *J. Biol. Chem.* **273**, 29150 – 29155

382. Chairoungdua, A., Segawa, H., Kim, J. Y., Miyamoto, K.-i, Haga, H., Fukui, Y., Mizoguchi, K., Ito, H., Takeda, E., Endou, H., and Kanai, Y. (1999) *J. Biol. Chem.* **274**, 28845 – 28848

383. Chambers, E. J., Bloomberg, G. B., Ring, S. M., and Tanner, M. J. A. (1999) *J. Mol. Biol.* **285**, 1289 – 1307

384. Annunen, S., Paassilta, P., Lohiniva, J., Perälä, M., Pihlajamaa, T., Karppinen, J., Tervonen, O., Kröger, H., Lähde, S., Vanharanta, H., Ryhänen, L., Göring, H. H. H., Ott, J., Prockop, D. J., and Ala-Kokko, L. (1999) *Science* **285**, 409 – 412

385. Ward, C. J., Turley, H., Ong, A. C. M., Comley, M., Biddolph, S., Chetty, R., Ratcliffe, P. J., Gatter, K., and Harris, P. C. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1524 – 1528

386. Hugot, J.-P., 16 other authors, and Groupe d'Etude Thérapeutique des Affections Inflammatoires Digestives. (1996) *Nature (London)* **379**, 821 – 823

387. Ghosh, S., and 39 other authors (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2198 – 2203

388. Kinzler, K. W., and Vogelstein, B. (1993) *Nature (London)* **363**, 495

389. Brehm, A., and Kouzarides, T. (1999) *Trends Biochem. Sci.* **24**, 142 – 145

390. Cho, Y., Gorina, S., Jeffrey, P. D., and Pavletich, N. P. (1994) *Science* **265**, 346 – 355

391. Oren, M. (1999) *J. Biol. Chem.* **274**, 36031 – 36034

392. Venkitaraman, A. R. (1999) *Science* **286**, 1100 – 1102

393. Smith, J. R., and 22 other authors (1996) *Science* **274**, 1371 – 1374

394. Novelli, M. R., Williamson, J. A., Tomlinson, I. P. M., Elia, G., Hodgson, S. V., Talbot, I. C., Bodmer, W. F., and Wright, N. A. (1996) *Science* **272**, 1187 – 1190

395. Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J. C., and Perucho, M. (1997) *Science* **275**, 967 – 969

396. Rabbitts, T. H. (1994) *Nature (London)* **372**, 143 – 149

397. Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) *Nature (London)* **396**, 643 – 649

398. Kappel, C. A., Bieberich, C. J., and Jay, G. (1994) *FASEB J.* **8**, 583 – 592

399. Ahmad, W., and 17 other authors (1998) *Science* **279**, 720 – 724

400. Bassett, D. E., Jr., Boguski, M. S., and Hieter, P. (1996) *Nature (London)* **379**, 589 – 590

401. Richards, R. I., and Sutherland, G. R. (1997) *Trends Biochem. Sci.* **22**, 432 – 436

402. Wang, Y.-H., Gellibolian, R., Shimizu, M., Wells, R. D., and Griffith, J. (1996) *J. Mol. Biol.* **263**, 511 – 516

403. Weisman-Shomer, P., Naot, Y., and Fry, M. (2000) *J. Biol. Chem.* **275**, 2231 – 2238

404. Fridell, R. A., Benson, R. E., Hua, J., Bogerd, H. P., and Cullen, B. R. (1996) *EMBO J.* **15**, 5408 – 5414

405. Mandel, J.-L. (1997) *Nature (London)* **386**, 767 – 769

405a. Moine, H., and Mandel, J.-L. (2001) *Science* **294**, 2487– 2488

405b. Greenough, W. T., Klintsova, A. Y., Irwin, S. A., Galvez, R., Bates, K. E., and Weiler, I. J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7101– 7106

405c. Sohn, E. (2001) *Science* **294**, 1809

405d. Deng, J., Xiong, Y., and Sundaralingam, M. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13665 – 13670

406. Wells, R. D. (1996) *J. Biol. Chem.* **271**, 2875 – 2878

407. Timchenko, L. T., and Caskey, C. T. (1996) *FASEB J.* **10**, 1589 – 1597

407a. Sinden, R. R. (2001) *Nature (London)* **411**, 757– 758

408. Muragaki, Y., Mundlos, S., Upton, J., and Olsen, B. R. (1996) *Science* **272**, 548 – 551

408a. Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., and Ranum, L. P. W. (2001) *Science* **293**, 864 – 867

409. Singer, R. H. (1998) *Science* **280**, 696 – 697

409a. Dhe-Paganon, S., Shigeta, R., Chi, Y.-I., Ristow, M., and Shoelson, S. E. (2000) *J. Biol. Chem.* **275**, 30753 – 30756

409b. Lodi, R., Cooper, J. M., Bradley, J. L., Manners, D., Styles, P., Taylor, D. J., and Schapira, A. H. V. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11492 – 11495

409c. Chen, O. S., and Kaplan, J. (2000) *J. Biol. Chem.* **275**, 7626 – 7632

409d. Foury, F., and Talibi, D. (2001) *J. Biol. Chem.* **276**, 7762 – 7768

409e. Gakh, O., Adamec, J., Gacy, A. M., Twesten, R. D., Owen, W. G., and Isaya, G. (2002) *Biochemistry* **41**, 6798 – 6804

410. Campuzano, V., and 26 other authors (1996) *Science* **271**, 1423 – 1427

410a. LeProust, E. M., Pearso, C. E., Sinden, R. R., and Gao, X. (2000) *J. Mol. Biol.* **302**, 1063 – 1080

410b. Rosenberg, R. N. (1995) *N. Engl. J. Med.* **333**, 1351– 1353

411. Mariappan, S. V. S., Catasti, P., Silds, L. A., III, Bradbury, E. M., and Gupta, G. (1999) *J. Mol. Biol.* **285**, 2035 – 2052

412. Dürr, A., Cossee, M., Agid, Y., Campuzano, V., Mignard, C., Penet, C., Mandel, J.-L., Brice, A., and Koenig, M. (1996) *N. Engl. J. Med.* **335**, 1169 – 1175

413. Rosenberg, R. N. (1996) *N. Engl. J. Med.* **335**, 1222 – 1224

414. Perutz, M. F. (1999) *Trends Biochem. Sci.* **24**, 58 – 63

414a. Holbert, S., Denghien, I., Kiechle, T., Rosenblatt, A., Wellington, C., Hayden, M. R., Margolis, R. L., Ross, C. A., Dausset, J., Ferrante, R. J., and Néri, C. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1811– 1816

414b. Peters, M. F., and Ross, C. A. (2001) *J. Biol. Chem.* **276**, 3188 – 3194

414c. Bates, G. P. (2001) *Nature (London)* **413**, 691– 694

415. Wellington, C. L., and 20 other authors (1998) *J. Biol. Chem.* **273**, 9158 – 9167

References

416. Karlin, S., and Burge, C. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1560–1565
416a. Kaytor, M. D., and Warren, S. T. (1999) *J. Biol. Chem.* **274**, 37507–37510
416b. Perutz, M. F., and Windle, A. H. (2001) *Nature (London)* **412**, 143–144
416c. Chai, Y., Wu, L., Griffin, J. D., and Paulson, H. L. (2001) *J. Biol. Chem.* **276**, 44889–44897
417. Diamond, M. L., Robinson, M. R., and Yamamoto, K. R. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 657–661
417a. Fernandez-Funez, P., and 14 other authors (2000) *Nature (London)* **408**, 101–106
417b. Mushegian, A. R., Vishnivetskiy, S. A., and Gurevich, V. V. (2000) *Biochemistry* **39**, 6809–6813
418. Skinner, P. J., Koshy, B. T., Cummings, C. J., Klement, I. A., Helin, K., Servadio, A., Zoghbi, H. Y., and Orr, H. T. (1997) *Nature (London)* **389**, 971–974
419. Collick, A., Norris, M. L., Allen, M. J., Bois, P., Barton, S. C., Surani, M. A., and Jeffreys, A. J. (1994) *EMBO J.* **13**, 5745–5753
420. Sutherland, G. R., and Richards, R. I. (1994) *N. Engl. J. Med.* **331**, 191–193
421. McMurray, C. T. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1823–1825
422. Warpeha, K. M., Xu, W., Liu, L., Charles, I. G., Patterson, C. C., Ah-Fat, F., Harding, S., Hart, P. M., Chakravarthy, U., and Hughes, A. E. (1999) *FASEB J.* **13**, 1825–1832
423. Petruska, J., Hartenstine, M. J., and Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 5204–5210
424. Pearson, C. E., Eichler, E. E., Lorenzetti, D., Kramer, S. F., Zoghbi, H. Y., Nelson, D. L., and Sinden, R. R. (1998) *Biochemistry* **37**, 2701–2708
425. Iyer, R. R., Pluciennik, A., Rosche, W. A., Sinden, R. R., and Wells, R. D. (2000) *J. Biol. Chem.* **275**, 2174–2184
426. Jakupciak, J. P., and Wells, R. D. (1999) *J. Biol. Chem.* **274**, 23468–23479
427. Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., and Fodor, S. P. A. (1996) *Science* **274**, 610–614
428. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. (1999) *Science* **286**, 531–537
429. Service, R. F. (1998) *Science* **282**, 396–399
430. Gerhold, D., Rushmore, T., and Caskey, C. T. (1999) *Trends Biochem. Sci.* **24**, 168–173
430a. Pennisi, E. (2002) *Science* **297**, 1985, 1987
430b. Schena, M. (2002) *Microarray Analysis*, Wiley, New York
431. Gerry, N. P., Witowski, N. E., Day, J., Hammer, R. P., Barany, G., and Barany, F. (1999) *J. Mol. Biol.* **292**, 251–262
432. Tang, K., Fu, D.-J., Julien, D., Braun, A., Cantor, C. R., and Köster, H. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10016–10020
433. Marshall, E. (1999) *Science* **286**, 444–447
434. Dalton, R. (2000) *Nature (London)* **403**, 234
435. Burns, M. A., Johnson, B. N., Brahmasandra, S. N., Handique, K., Webster, J. R., Krishnan, M., Sammarco, T. S., Man, P. M., Jones, D., Heldsinger, D., Mastrangelo, C. H., and Burke, D. T. (1998) *Science* **282**, 484–487
436. Service, R. F. (1998) *Science* **282**, 399–401
437. Landegren, U., Kaiser, R., Caskey, C. T., and Hood, L. (1988) *Science* **242**, 229–237
438. Kolata, G. (1986) *Science* **234**, 22–24
439. Schnieke, A. E., Kind, A. J., Ritchie, W. A., Mycock, K., Scott, A. R., Ritchie, M., Wilmut, I., Colman, A., and Campbell, K. H. S. (1997) *Science* **278**, 2130–2133
440. Arnon, R. (1986) *Trends Biochem. Sci.* **11**, 521–524
441. McAleer, W. J., Buynak, E. B., Maigetter, R. Z., Wampler, D. E., Miller, W. J., and Hilleman, M. R. (1984) *Nature (London)* **307**, 178–180
442. Kapusta, J., Modelska, A., Figlerowicz, M., Pniewski, T., Letellier, M., Lisowa, O., Yusibov, V., Koprowski, H., Plucienniczak, A., and Legocki, A. B. (1999) *FASEB J.* **13**, 1796–1799
443. Weiner, D. B., and Kennedy, R. C. (1999) *Sci. Am.* **281**(Jul), 50–57
444. Taubes, G. (1997) *Science* **278**, 1711–1714
445. McDonald, G. A., Anacker, R. L., and Garjian, K. (1987) *Science* **235**, 83–85
446. Brown, F., Schild, G. C., and Ada, G. L. (1986) *Nature (London)* **319**, 549–550
447. Rupprecht, C. E., Wiktor, T. J., Johnston, D. H., Hamir, A. N., Dietzschold, B., Wunner, W. H., Glickman, L. T., and Koprowski, H. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7947–7950
448. Waldmann, T. A. (1986) *Science* **232**, 727–732
449. Travis, J., Owen, M., George, P., Carrell, R., Rosenberg, S., Hallewell, R. A., and Barr, P. J. (1985) *J. Biol. Chem.* **260**, 4384–4389
450. Parkman, R. (1986) *Science* **232**, 1373–1378
451. Sandig, V., Youil, R., Bett, A. J., Franlin, L. L., Oshima, M., Maione, D., Wang, F., Metzker, M. L., Savino, R., and Caskey, C. T. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1002–1007
452. Verma, I. M., and Somia, N. (1997) *Nature (London)* **389**, 239–242
453. Anderson, W. F. (1995) *Sci. Am.* **273**(Sep), 124–128
454. Marshall, E. (1995) *Science* **269**, 1050–1055
455. Zanjani, E. D., and Anderson, W. F. (1999) *Science* **285**, 2084–2088
456. Finkel, T., and Epstein, S. E. (1995) *FASEB J.* **9**, 843–851
456a. Miller, D. G., and Stamatoyannopoulos, G. (2001) *N. Engl. J. Med.* **344**, 1782–1783
457. Miller, H. I. (2000) *Science* **287**, 591–592
457a. Gura, T. (2001) *Science* **291**, 1692–1697
458. Amalfitano, A., McVie-Wylie, A. J., Hu, H., Dawson, T. L., Raben, N., Plotz, P., and Chen, Y. T. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8861–8866
458a. Pawliuk, R., Westerman, K. A., Fabry, M. E., Payen, E., Tighe, R., Bouhassira, E. E., Acharya, S. A., Ellis, J., London, I. M., Eaves, C. J., Humphries, R. K., Beuzard, Y., Nagel, R. L., and Leboulch, P. (2001) *Science* **294**, 2368–2371
458b. Dunder, U., Kaartinen, V., Valtonen, P., Väänänen, E., Kosma, V.-M., Heisterkamp, N., Groffen, J., and Mononen, I. (2000) *FASEB J.* **14**, 361–367
459. Capecchi, M. R. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 956–957
460. Marshall, E. (1996) *Science* **271**, 586–588
461. Kohn, D. B., and Parkman, R. (1997) *FASEB J.* **11**, 635–639
462. Broitman, S., Amosova, O., Dolinnaya, N. G., and Fresco, J. R. (1999) *J. Biol. Chem.* **274**, 21763–21768
462a. Knight, J. (2001) *Nature (London)* **413**, 12–15
463. Alper, J. (1999) *Science* **283**, 1625–1626
464. Kishore, G. M., and Shewmaker, C. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5968–5972
465. Trewavas, A. (1999) *Nature (London)* **402**, 231–232
466. Goodman, R. M., Hauptli, H., Crossway, A., and Knauf, V. C. (1987) *Science* **236**, 48–54
467. Jones, M. G. K. (1985) *Nature (London)* **317**, 579–580
468. Roberts, L. (1988) *Science* **241**, 1290
469. Frommer, W. B., Ludewig, U., and Rentsch, D. (1999) *Science* **285**, 1222–1223
470. Kling, J. (1996) *Science* **274**, 180–181
471. Gura, T. (1999) *Science* **285**, 994–995
472. Simon Moffat, A. (1992) *Science* **256**, 770–771
473. Thomas, S. M., Davies, A. R. W., Birtwistle, N. J., Crowther, S. M., and Burke, J. F. (1996) *Nature (London)* **380**, 387–388
474. Marshall, E. (1996) *Science* **273**, 1788–1789
475. Enserink, M. (1998) *Science* **281**, 890–891
476. Dickson, D. (1993) *Nature (London)* **366**, 391
477. Poste, G. (1995) *Nature (London)* **378**, 534–536
478. Fuller, B. P., Ellis Kahn, M. J., Barr, P. A., Biesecker, L., Crowley, E., Garber, J., Mansoura, M. K., Murphy, P., Murray, J., Phillips, J., Rothenberg, K., Rothstein, M., Stopfer, J., Swergold, G., Weber, B., Collins, F. S., and Hudson, K. L. (1999) *Science* **285**, 1359–1361
479. Butler, D. (1998) *Nature (London)* **391**, 727
480. Hudson, K. L., Rothenberg, K. H., Andrews, L. B., Kahn, M. J. E., and Collins, F. S. (1995) *Science* **270**, 391–393
481. Lapham, E. V., Kozma, C., and Weiss, J. O. (1996) *Science* **274**, 621–624
482. Fost, N. (1992) *FASEB J.* **6**, 2813–2817
483. Holtzman, N. A. (1999) *Science* **286**, 409
484. Wills, C. (1970) *Sci. Am.* **222**(Mar), 98–107
485. Vogel, G. (1999) *Science* **286**, 2050–2051
486. Stoye, J. P. (1997) *Nature (London)* **386**, 126–127
487. Weiss, R. A. (1999) *Science* **285**, 1221–1222
488. Paradis, K., Langford, G., Long, Z., Heneine, W., Sandstrom, P., Switzer, W. M., Chapman, L. E., Lockey, C., Onions, D., XEN 111 Study Group, and Otto, E. (1999) *Science* **285**, 1236–1241
489. Enserink, M. (1999) *Science* **286**, 1662–1668
490. Haslberger, A. G. (2000) *Science* **287**, 431–432
491. Jackson, A., and Inglehearn, C. (1999) *Nature (London)* **402**, 457
492. Crawley, M. J. (1999) *Nature (London)* **400**, 501–502
493. Saxena, D., Flores, S., and Stotzky, G. (1999) *Nature (London)* **402**, 480
494. Smith, R. J. (1984) *Science* **224**, 1215–1216
495. Many, B. W. J., Almond, J. W., Berns, K. I., Chanock, R. M., Lvov, D. K., Pettersson, R. F., Schatzmayr, H. G., and Fenner, F. (1993) *Science* **262**, 1223–1224
496. Joklik, W. K., Moss, B., Fields, B. N., Bishop, D. H. L., and Sandakhchiev, L. S. (1993) *Science* **262**, 1225–1226
497. Jesson, L. K., and Barrett, S. C. H. (2002) *Nature (London)* **417**, 707
498. Human Genome Organization Ethics Committee. (2000) *Science* **290**, 49
499. Williamson, R., and Duncan, R. (2002) *Nature (London)* **418**, 585–586
500. Adam, D. (2002) *Nature (London)* **417**, 370
501. Aldhous, P. (2002) *Nature (London)* **418**, 355–356
502. Rothenberg, K. H., and Terry, S. F. (2002) *Science* **297**, 196–197
503. Nowlan, W. (2002) *Science* **297**, 195–196
504. McDowell, N. (2002) *Nature (London)* **416**, 571
505. Stokstad, E. (2002) *Science* **297**, 1797, 1799
506. Kaiser, J. (2001) *Science* **292**, 34–36
507. Enserink, M. (2002) *Science* **297**, 30–31
508. Clarke, T. (2002) *Nature (London)* **419**, 429–430
509. Rieger, M. A., Lamond, M., Preston, C., Powles, S. B., and Roush, R. T. (2002) *Science* **296**, 2386–2388
510. Spurgeon, D. (2001) *Nature (London)* **409**, 749

Study Questions

1. The two chromatids in Fig. 26-13G are said to be coiled with opposite handedness. Can you draw this conclusion from Fig. 26 alone? What are the biological implications for mitosis?[280] Does the DNA have a differing chirality at the molecular level? Compare this observation with the existence of snail shells or flowers[497] with both right and left handedness within the same species.

2. Will the sequencing of the human genome ever be complete?

3. Should patenting of human genes be allowed? Under what circumstances should patents be allowed on genetically engineered genes?

4. Should the human genome be regarded as a **common heritage** such that there is a guarantee that the medical and other benefits arising from genetic research are available to all persons on earth? See Human Genome Organization Ethics Committee.[498]

5. DNA testing (Box 5-D) is widely used by police throughout the world. It has been estimated that if ten loci in the DNA are tested the chance of a random match between two people is one in a billion. In the United Kingdom it is planned to hold DNA profiles on record for one of every 15 people. Is this wise? Or should DNA profiles be recorded for all people?[499] If DNA profiles are on record how can we be sure that they are not used dishonestly? Should police have access to DNA data bases? See Adam.[500]

6. Should "race" be used as a variable in biomedical studies? See Aldhous.[501]

7. Should insurance companies be allowed to have access to genetic information about insured people? Companies usually obtain other medical information. See Rothenberg and Terry,[502] Adam,[500] and Nowlan.[503]

8. Should genetically engineered fish be allowed in "farms" that are set up in ocean waters? See McDowell[504] and Stokstad.[505]

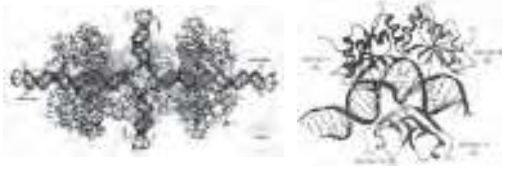
9. Do transgenic trees pose a threat to natural ecosystems? See Kaiser.[506]

10. Should we attempt to replace wild populations of mosquitoes with genetically engineered mosquitoes that can not transmit malaria? See Enserick[507] and Clarke.[508]

11. Could dispersal of pollen from genetically modified plants lead to undesirable "genetic pollution" of the environment? See Rieger *et al*.[509]

12. To what extent is genetic modification of plants and animals equivalent to changes made by conventional breeding? See Spurgeon.[510]

13. Choose one of the ethical questions that can be raised about application of our new knowledge about the genome (e.g., see pp. 1518, 1519, Chapter 32, and study questions on this page). Study literature available to you and prepare a recommendation to the public, Congress, or to local regulatory agencies. Follow a scientific approach. Try to find true facts that can be verified. Consider all viewpoints. State some of the uncertainties in your recommendation. Present your proposal to a class or to a friend for criticism. Then publish your view in a newspaper if you wish to.

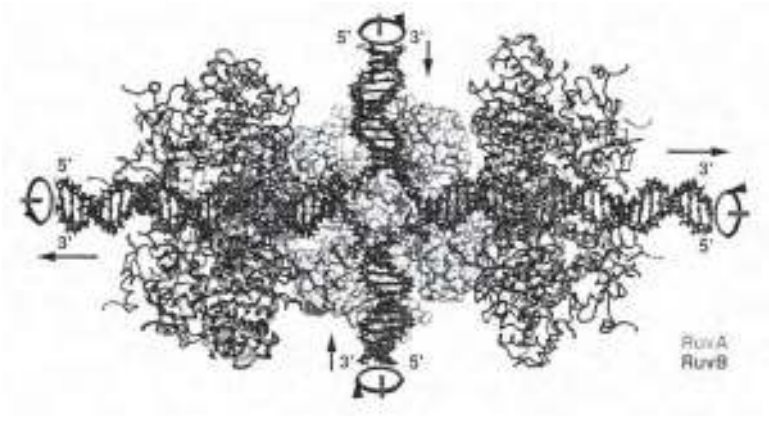


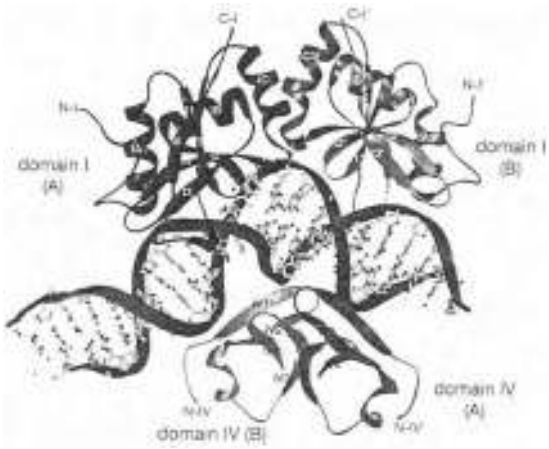
Many proteins interact with DNA. These include polymerases that replicate DNA, helicases that unwind double helices, topoisomerases that cut and reseal DNA strands to avoid entangling, and enzymes that repair damage to DNA. At left is a motor complex consisting of the tetrameric binding protein RuvA (light shading) and the hexameric helicase RuvB (two copies in darker shading). See also Fig. 27-26B. From Putnam *et al.* (2001) *J. Mol. Biol.*, **311**, 297–310. Right bacterial protein (from *Thermus aquaticus*) MutS recognizes mispaired bases in DNA and initiates their removal. The DNA is bound in a bent conformation. In this duplex one strand contains an unpaired thymine (at top of bend), which would cause a mutation if not removed. From Obmolova *et al.* (2000) *Nature*, **407**, 703–7.

Contents

Organization, Replication, Transposition, and Repair of DNA

27





Both the replication and transcription of DNA are complex processes. Although the basic chemistry is relatively simple many enzymes and other proteins are required. In part this reflects organizational and topological problems[1] associated with the huge amount of DNA present as a single molecule within a chromosome.

A. The Topology and Environment of DNA

Although we can isolate DNA in the form of simple double helical fragments, the topology of natural DNA is always more complex. Covalently closed circular DNA such as that in plasmids, mitochondria, and bacterial chromosomes is supercoiled (Chapter 5) and bound to proteins. The DNA of chromosomes and bacteriophage particles is folded further into more compact forms. For example, the chromosome of *E. coli* contains DNA about 1.5 mm in length folded within a cell that is only 2 µm long. The diploid length of DNA in a 20 µm cell of a human is about 1.5 meters. At the time of cell division human DNA must all be replicated and packaged into chromosomes, 23 pairs in each cell. The density of the compacted DNA varies. A bacterial nucleoid may contain 10–30 mg / ml of DNA.[2] In chromatin of a eukaryotic nucleus there may be 200 mg / ml of DNA and in nucleosomes 330–400 mg / ml.[2a] The tightly compacted head of the T4 bacteriophage (Box 7-C) contains 520 mg / ml.[3]

1. DNA in Viruses

In the simplest filamentous DNA viruses such as M13 the DNA is coated by a helical protein sheath (Fig. 7-7), as it is extruded from a cell. The sheath is peeled off as the virus enters another cell. However, in the large tailed phage (Box 7-C), which contains ~160 kb of polynucleotide chains, the DNA is closely packed within the heads. In a model that seems to accommodate most experimental results, the DNA rod bends sharply into a series of folds, which are laid down around the long axis of the head in spirally arranged shells (Fig. 27-1).[4,4a] The end of the DNA that enters the phage head first appears to be located in the center with successive shells of DNA around it.[4] In the large bacteriophage G the DNA appears to be folded to form 12 icosahedrally arranged pear-shaped rings in the corners of the capsid.[5] The 2.0 nm diameter double helical segments of DNA lie roughly parallel and are separated by only 0.5–1.0 nm of solvent,[6] which contains cations such as the polycation of spermidine. Another possibility is that the DNA may be wound as on a spool of thread.[7] The DNA chains have an external diameter of ~2 nm with 0.6 nm additional for a hydration layer. In a phage head the adjacent parallel chains are 2.6–2.7 nm apart. Thus, the packing is very tight. Even so, capsids tend to be only about half filled with DNA. An exception is provided by the tobacco mosaic virus (Fig. 7-8) in which the RNA genome is held precisely by protein subunits, which dissociate to release the RNA during infection.[3]

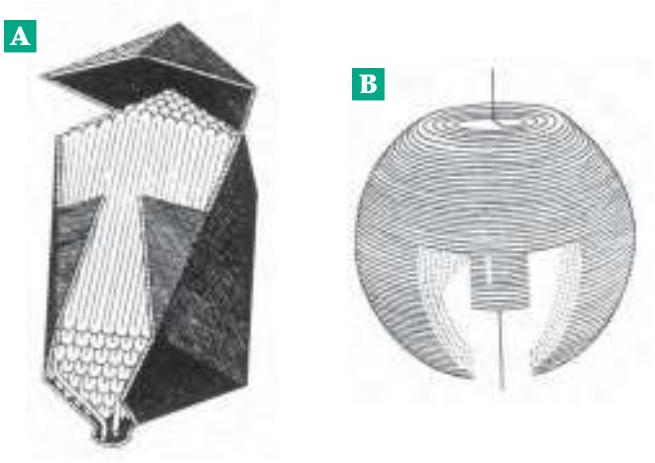


Figure 27-1 Possible ways of packing DNA into the heads of bacteriophage particles. (A) Spiral-fold. (B) Concentric shell model From Black *et al.*[4]

2. Bacterial Chromosomes and Plasmids

Most DNA in living organisms, whether bacteria or eukaryotes, is underwound. That is, the superhelix density (Chapter 5) is about −0.05 or one supercoil per 200 base pairs. In eukaryotes this negative supercoiling can be accounted for by the winding of DNA around the histones within the nucleosomes (Figs. 5-21, 27-4). The situation in bacteria is not as clear. There are many bacterial DNA-binding proteins,[8,9] but one of them known as HU is particularly abundant.[8–12] In *E. coli* it exists as mixed dimers and tetramers of 9.5 (α) and 9.2 (β) kDa subunits. Each HU tetramer can bind ~60 bp of DNA. There are about 60,000 HU monomers per cell, enough to coat ~20% of the genome. Possible modes of interaction with DNA have been proposed on the basis of the X-ray structure of HU.[8,13–14a] Binding to HU causes the DNA to be more tightly wound, introducing additional negative supercoils. The resulting unreleased torsional stress may be important in the functioning of the DNA.[15–17a] Binding is strongest to four-way junctions,[17a] and to DNA with nicks and gaps or to structures induced by supercoiling.[17b] Other basic histonelike proteins may also bind to the DNA. However, there are no structures that resemble eukaryotic nucleosomes.

If bacterial cells are lysed under certain conditions, e.g., in 1 M NaCl or in the presence of a "physiological" 5 mM spermidine, the entire bacterial chromosome can be isolated.[10] The DNA in these isolated chromosomes retains some torsional tension that, however, can be relaxed by nicking with nucleases or by γ-irradiation. However, a single nick relaxes the DNA very little. The explanation appears to be that the DNA is held by proteins of the nucleoid matrix in a series of loops (Fig. 27-2). A single nick relaxes just one loop. On this basis there are 43 ± 10 loops per genome with ~100 kb of DNA per loop.[18] A 136-residue protein **H-NS** is involved in condensation of bacterial DNA.[18a] It may act as a scaffolding protein, but it also functions in controlling transcription.[18b]

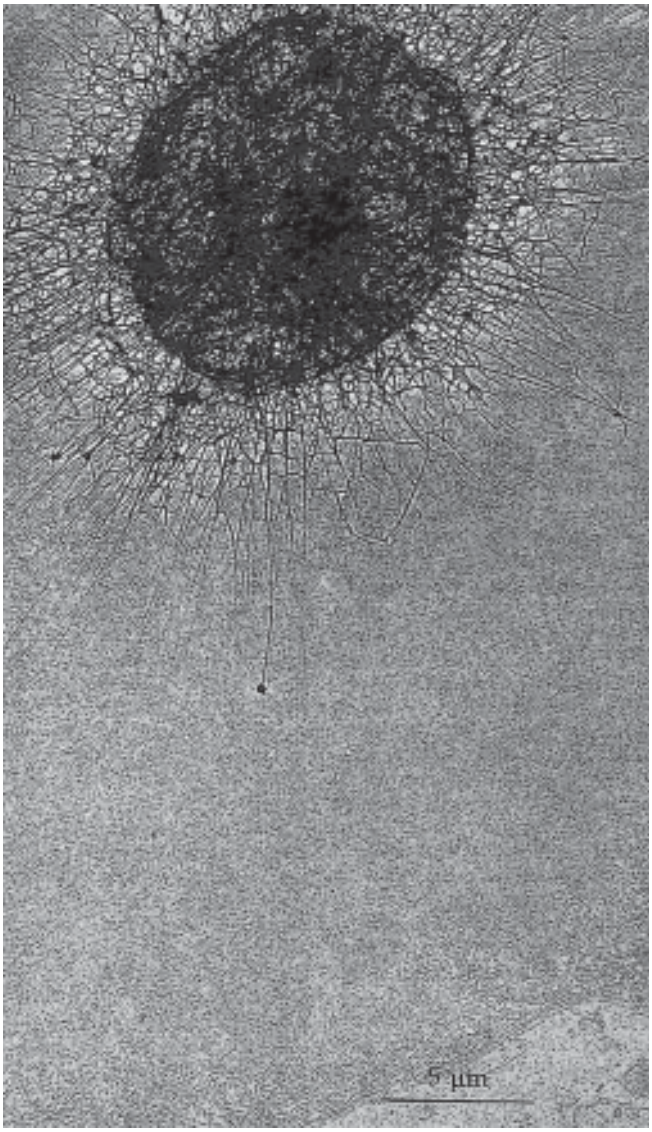


Figure 27-2 Electron micrograph of a bacterial nucleoid. The DNA is usually contained within the 'cage', but has been spread, using Kleinschmidt's procedure, to yield a surrounding skirt. The cage contains a protein network which includes elements of the cytoskeleton, enclosing the residual nuclear substructures. The denser fibrils radiating from this cage disappear to nuclease digestion and are probably aggregates of DNA fibers which merge with individual DNA strands at the extremities of the skirt. These strands are highly supercoiled, indicative of intact DNA. From Jackson and Patel, provided by Dr. S. J. McCready.[19]

3. Protamines, Histones, and Nucleosomes

Within bacterial cells the negatively charged phosphate groups of the DNA are neutralized to a large extent by the positively charged polyamines,[19a] by cations such as K^+ and Mg^{2+}, and by basic proteins such as HU. Within the mature heads of sperm cells of fish, the tightly packaged DNA is neutralized by the **protamines**, small ~5-kDa proteins rich in arginine.[4a,20–21a] Similar basic proteins are found in mammalian sperm.[22,23] However, within most eukaryotic cells, the charges on DNA are balanced principally by a group of basic proteins, first isolated and named **histone** by Kossel in 1884.[24]

There are five classes of histones, which range in molecular mass from ~11 to 21.5 kDa:[25,26]

H1 (including $H1^0$, H5; lysine-rich "linker" histone)
H2A, H2B (moderately lysine-rich) } "core" histones
H3, H4 (arginine-rich)

All of the core histones share a conserved 65-residue **histone fold.**[27,28] The arginine-rich histones have a strongly conserved amino acid sequence, histone H4 from pea seedlings differing from that of the bovine thymus by only two amino acids. On the other hand, the lysine-rich H1 is almost species-specific in its sequence. Differentiated tissues contain at least seven variant forms of histone H1 including proteins designated $H1^0$, H1t, and H5.[29–31]

The N-terminal 25–40 amino acids of the core histones are positively charged and highly conserved.[32] The 135-residue histone H3 of calf thymus carries a net charge of +18 within the first 53 residues. This is probably the portion that binds to DNA. On the other hand, the carboxyl terminal end is hydrophobic and only slightly basic.[33] Histones undergo substantial amounts of **micromodification** including phosphorylation, acetylation, and methylation.[33a,b] Mono-, di-, and tri-methyllysine residues may be present.[33c,d] The core histones all undergo acetylation on specific lysyl side chains. Nuclear histone acetyltransferases[34–38b] transfer acetyl groups from acetylCoA and hydrolytic deacetylases may remove them.[39–40b] The amount of acetylation varies during different stages of the cell cycle, suggesting a regulatory role.[41] Acetylation sites in H3 and H4 are highly conserved in all eukaryotes.[42]

A small fraction of histone H2A undergoes phosphorylation and dephosphorylation continuously,[43] but H1 and H3 are phosphorylated and dephosphorylated at specific stages of the cell cycle. Phosphorylation of H1 has been thought essential for "condensation" of chromatin,[44,45] the folding into the tightly packed chromosome structures. However, more recent experiments point to the N terminus of histone H2B as the required site of phosphorylation for chromosome condensation.[45a] In addition, histone H1 may interact with membrane lipids.[46] Histone H2A exhibits the greatest heterogeneity and appears to function in regulation of transcription, of gene silencing, and of repair of double-strand breaks in DNA.[46a] Each of the histones appears to be regulated separately. In animal chromatin ~10% of histone H2A and small fractions of H2B and of tissue-specific histones are covalently linked to ubiquitin (Box 10-C). However, this monoubiquitination may not be related to proteolytic degradation.[47] Archea contain histones that dimerize and bind DNA to form nucleosomes.[47a,b]

Nucleosomes. An early idea of the function of histones was that they serve as gene repressors. To some extent this view is still valid. However, the large quantity of histone and uniform distribution over the DNA suggested some other role. This was clarified when electron micrographs showed that chromatin fibers form **nucleosomes**,[48–50] regular repeating structures resembling beads on a string. The same structure is seen in the "minichromosomes" formed from virus SV40 (Fig. 27-3).[51–53] Two molecules each of histones H2A, H2B, H3, and H4 form the core of the nucleosome around which ~146 bp of dsDNA is coiled into approximately two negative, left-handed toroidal superhelical turns (Figs. 5-21, 27-4).

Digestion of chromatin by nucleases causes rapid cleavage into ~200-bp fragments and slower cleavage to 146 ± 20-bp fragments. This suggested that ~200-bp segments of DNA are folded around a histone octomer, contracting the 68 nm extended length of relaxed B-DNA into a 10-nm nucleosome. A short linker region of variable length, up to 80 bp, lies between the nucleosomes.[54] The fifth histone, H1 (or H5 in some species), may bind to this linker DNA. A nucleosome with bound H1 is sometimes called a **chromatosome**.[55]

The superhelix density of ~0.05 observed for DNA extracted from eukaryotic cells is just equal to one negative superhelix turn per nucleosome. For example, the number of nucleosomes seen in the minichromosome of Fig. 27-3 matches the numbers of supercoils in the SV40 DNA (Fig. 5-20). If there are two negative supercoils per nucleosome, as shown in Fig. 27-4, the DNA in the nucleosome must be wound more tightly than in relaxed DNA (10.0 bp per turn instead of the 10.6 of relaxed DNA).[56] NMR data suggest that within the nucleosome the regular base pairing in the DNA may be partially disrupted and that some parts of the histone have a high degree of mobility.[57]

Although nucleosomes are distributed rather evenly along the DNA of a cell, there are some DNA sequences that favor nucleosome formation. The resulting **positioned nucleosomes** are often found in the vicinity of gene promoters, enhancers, and other

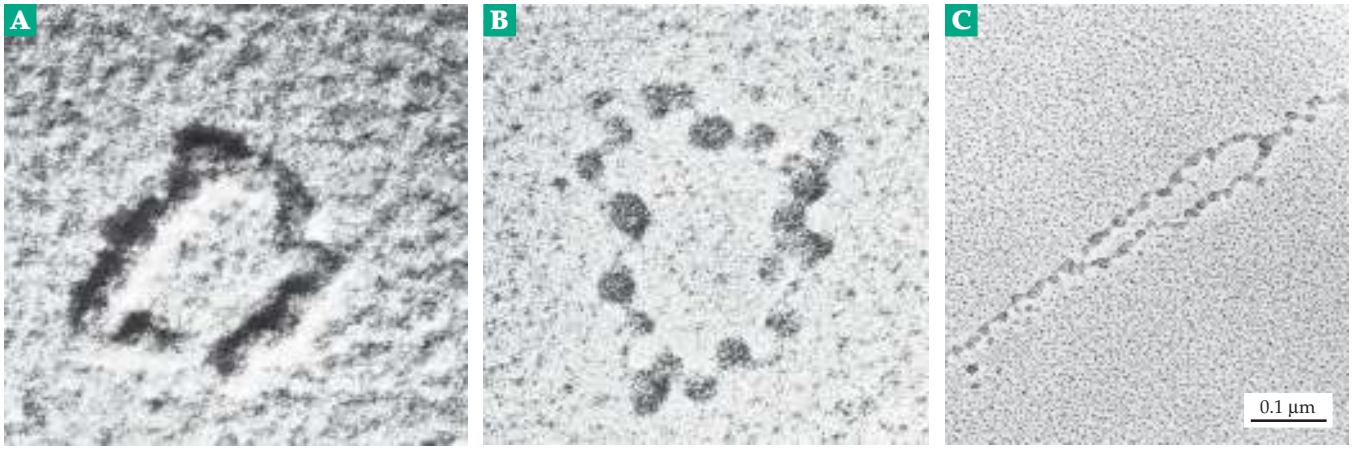


Figure 27-3 (A) Electron micrograph of "minichromosome" formed from virus SV40 growing in monkey cells in culture.[51] In this native form the nucleoprotein fiber is ~11 nm in diameter and ~210 nm in length. (B) Beaded form of minichromosome observed when the ionic strength was lowered. The 20 beads have diameters of ~11 nm and are joined by bridges roughly 2 nm in diameter and 13 nm long. Deproteinization and relaxation of the DNA revealed that the overall length of the DNA present is seven times the length of the native minichromosome. (C) Electron micrograph of chromatid of a blastoderm-stage embryo of *Drosophila melanogaster* in the process of replication. Nucleosomal particles are visible immediately adjacent to the replication forks. Courtesy of Steven L. McKnight and Oscar L. Miller, Jr.

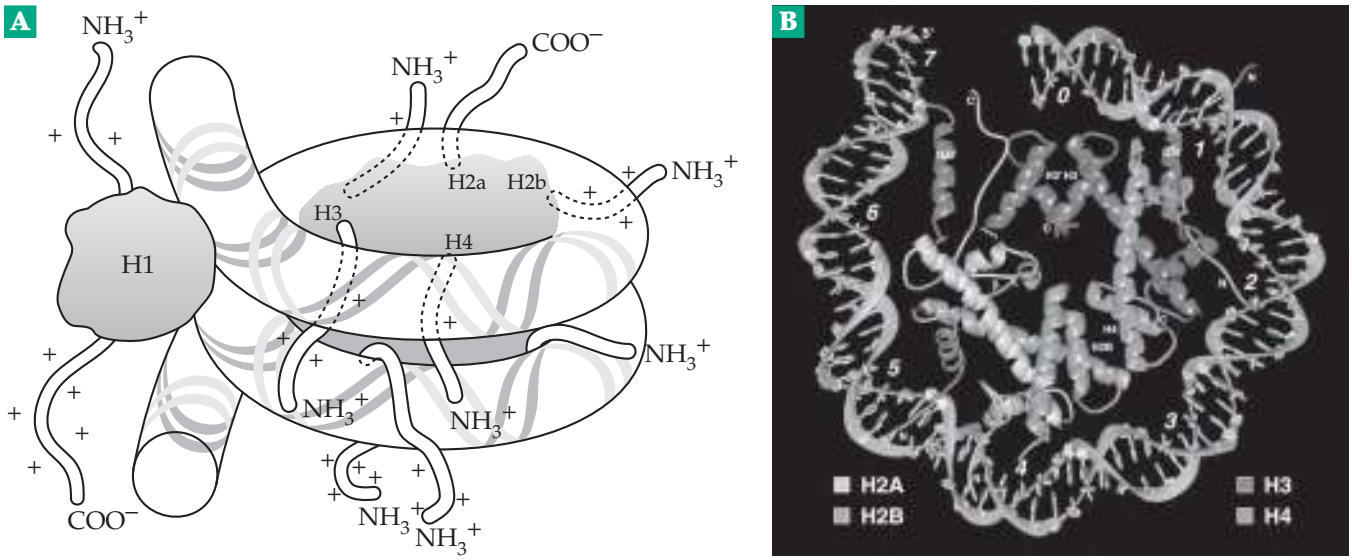


Figure 27-4 (A) A nucleosome is formed when dsDNA, shown schematically as a tube, wraps roughly twice around a histone octomer. This complex, or nucleosomal core particle, includes at its center two copies of histone H3, two of histone H4, and two H2A–H2B dimeric pairs, one of which is not visible. Ends of each histone molecule are thought to protrude like tails from the core, ready to interact with other molecules. In many organisms, histone H1, portrayed at left in one possible position, helps to anchor DNA to the core and promotes further compaction of the DNA into a 30-nanometer fiber. See Grunstein.[49] (B) Nucleosome core particle: 73-bp half. The view is down the superhelix axis with the pseudo dyad axis aligned vertically. The central base pair through which the dyad passes is above the superhelix axis location labeled 0. Each additional numerical label 1–7 represents one further DNA double helix turn. The complete histone proteins (except for the tail regions) that are primarily associated with the 73-bp superhelix half are shown. The two copies of each histone pair are distinguished as unprimed and primed, where the histone fold of the unprimed copy is primarily associated with the 73-bp DNA half and the primed copy with the 72-bp half. Four-helix bundles are labeled as H3', H3, and H2B, H4; histone-fold extensions of H3 and H2B are labeled as αN′, αN, and αC, respectively; the interface between the H2A docking domain and the H4 C terminus as β; and N- and C-terminal tail regions as N or C. From Luger *et al.*[50]

control sequences.[58–61a] It seems likely that positioning of nucleosomes is related to control of transcription or of other activities. The same basic structure for chromatin has been found in animals, fungi, and green plants.

The linker histones. A nucleosome is pictured in Fig. 27-4A as if held in a compact configuration by the binding of histone H1 at a position that marks a pseudo twofold axis that lies in the plane of the nucleosome. However, this is only one of several possible locations for linker histones of the H1, H5 family.[62–67] The structure of the linker histones is somewhat different from that of core histones. They have an 80-residue globular domain with long N-terminal *and* C-terminal chains, both of which are rich in basic residues and evidently available for binding to DNA.[68] Perhaps they bind both to the DNA entering the nucleosome and to that leaving the nucleosome, reducing electrostatic repulsion of those two parts of the DNA superhelix.[69]

Another possible location for histone H1 or H5 is above the histone surface as shown in Fig. 27-4 and *inside* the DNA loop.[65] A third suggested location for the globular linker core is *between* the two turns of the DNA strand.[66] While one function of the linker histones may be to stabilize mononucleosomes, they may also play a role in compaction of the DNA into the 30-nm fibers universally seen in nuclei of cells.[62,64–64b]

Histone H1 can also be regarded as a general repressor, holding chromatin tightly folded and preventing transcription. The possible roles of acetylation, phosphorylation, methylation, ubiquitination, and other modifications of histones in controlling transcription, replication, and DNA repair are receiving increasing attention.[70–73] **Active chromatin**, where transcription is occurring, has an altered nucleosome structure and increased susceptibility to nuclease action. It appears to be less tightly packed than inactive chromatin and to contain regions called **hypersensitive sites** that are accessible to nucleases or chemical modification reagents.[74–76] There seems to be a direct link between increased acetylation of histones and enhanced initiation of transcription by RNA polymerase II.[35,38,77–82] Conversely, deacetylation is associated with repression of transcription. Both histone acetylase and deacetylase activities have been found in transcriptional regulators.[80–82]

The observation that H1 becomes phosphorylated during the initiation step of mitosis suggests another control mechanism for its repressor functions.[83] Several multiprotein complexes that "remodel" chromatin have been identified.[73,84] These complexes contain

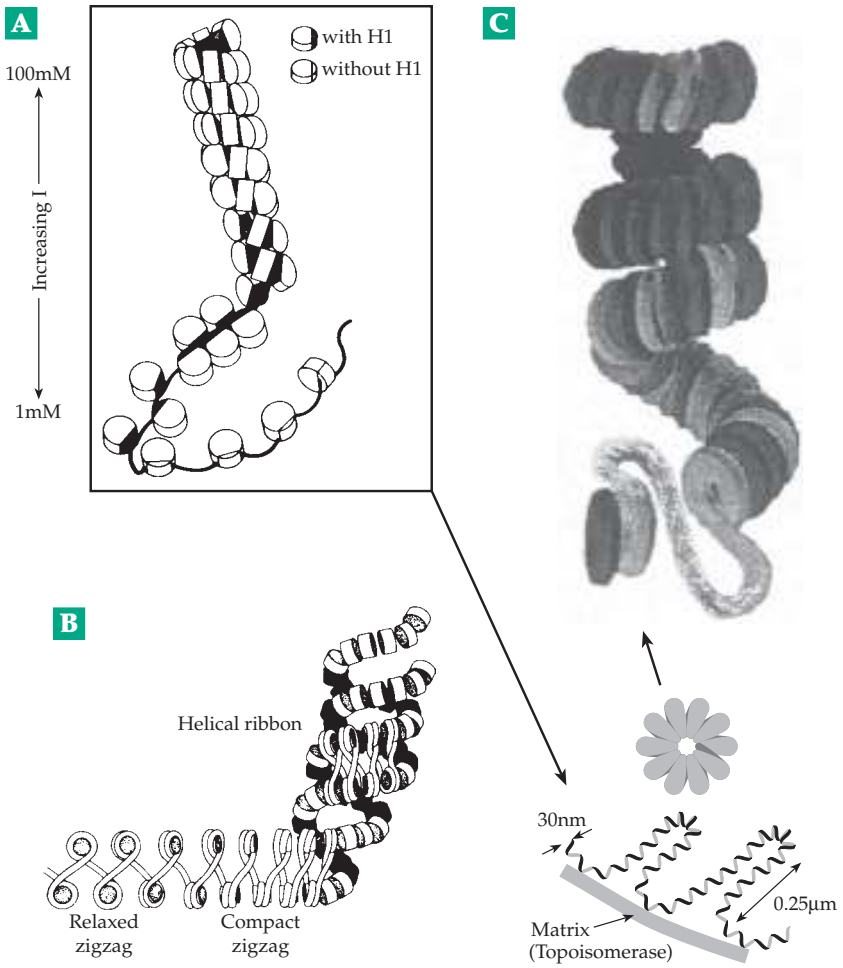


Figure 27-5 (A, B) Two possible models of the 30-nm chromatin fiber.[55] (A) Thoma *et al*.[85] (B) Woodcock *et al*.[64,87] The fully compacted structure is seen at the top of each figure. The bottom parts of the figures illustrate proposed intermediate steps in the ionic strength-induced compaction. (C) Possible organization of the DNA within a metaphase chromosome. Six nucleosomes form each turn of a solenoid in the 30-nm filament as in (A). The 30-nm filament forms ~30 kb-loop domains of DNA and some of these attach at the base to the nuclear matrix that contains topoisomerase II. About ten of the loops form a helical radial array of 250-nm diameter around the core of the chromosome. Further winding of this helix into a tight coil ~700 nm in diameter, as at the top in (C), forms a metaphase chromatid. From Manuelidis[91].

ATP-dependent **helicase** activities that open up DNA for transcription (see Chapter 28, Section B).

Folding of nucleosome chains; chromosomes. Electron microscopy shows that chromatin is packed into the nucleus largely as fibers of ~30- to 36-nm diameter.[55,84a] Thoma *et al.*[85] proposed that the fibers could be formed by winding the string of nucleosomes into a simple one-start helix with six nucleosomes per turn and a pitch of ~11 nm (Fig. 27-5A). The H1 molecules would be close together in the center. An alternative model, one of several suggested,[55,86] is shown in Fig. 27-5B. Its zigzag pattern of adjacent nucleosomes would generate a two-start helix.[64,87,88] Another alternative model envisions larger solenoids with interdigitated nucleosomes.[89]

A somewhat similar structure appears to be present in the specialized eukaryotic **lampbrush chromosomes** (Fig. 27-6), which are observed during the meiotic prophase of oocytes. They have been studied intensively in amphibians such as *Xenopus*. A lampbrush chromosome is actually a homologous pair of chromosomes, each one in turn consisting of two closely associated chromatids. The chromosomes are highly expanded, and about 5% of the DNA is extended in the form of ~4000 perfectly paired loops visible with an electron microscope. Each loop consists of ~50 μm or ~150 kb of extended DNA. No evidence of any breaks in the DNA is seen, a fact that supports the belief that a single DNA molecule extends from one end of the chromosome to the other through all of the loops.

Like the puffs of polytene chromosomes (Chapter 28), which may have a similar structure, lampbrush chromosomes appear to be actively engaged in transcription. Approximately 3% of the DNA may be functional in producing mRNA that is accumulated within the oocyte and is used as a template for protein synthesis during early embryonic development.[90]

A different arrangement is present in metaphase chromosomes, which appear as two dense parallel sister chromatids of ~700-nm diameter.[91] The DNA must be highly folded. In the model shown in Fig. 27-5C the 30-nm fiber is folded into ~30-kb loops, each one formed from ~25 turns of the 30-nm helix. The loops then form a helical array 250 nm in diameter with ~ ten loops (300 kb) per turn. This helix is further wound into a tight helix of ~700-nm diameter. A single turn of this helix may contain as much as 9 Mb of DNA.[91] About four hundred coils (an average of 18 coils per human haploid chromosome) could accommodate the entire genome. A group of five proteins, some of which are designated **SMC** (structural maintenance of chromosomes) proteins, form a complex called **cohesin**. SMC proteins, large multidomain proteins found in all eukaryotes, are also present in bacteria.[91a,b]

Interphase chromatin must be much less tightly

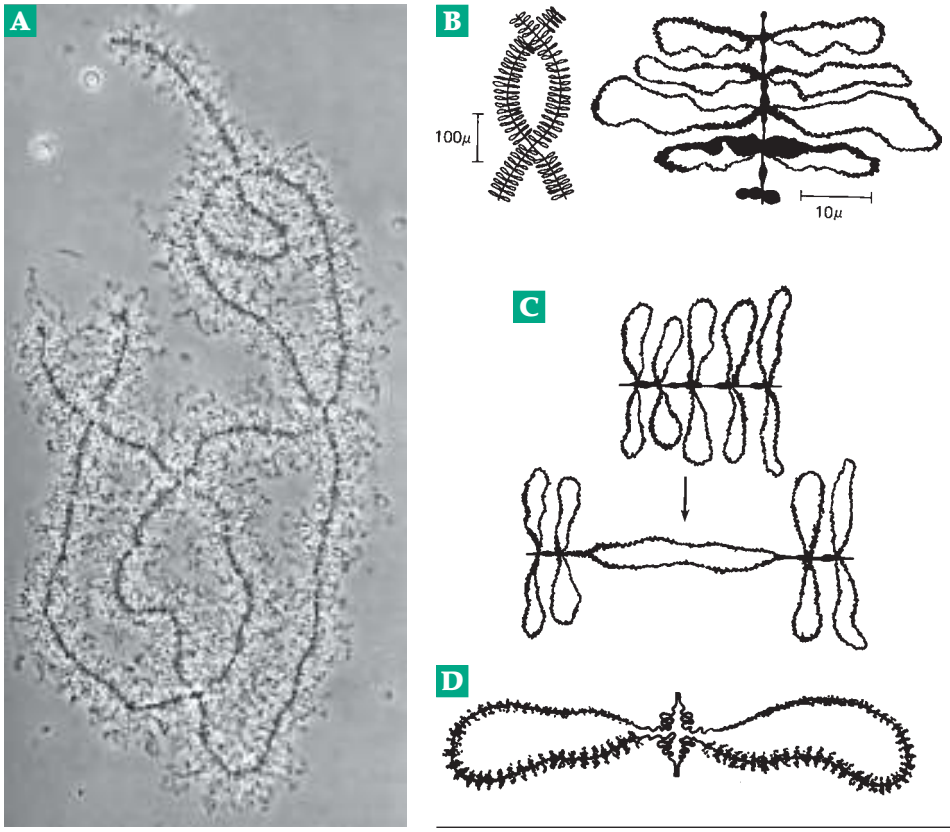


Figure 27-6 (A) Photomicrograph of a lampbrush chromosome from the nucleus of an oocyte of the newt *Triturus*. From L. M. Mays, *Genetics, A Molecular Approach*, Macmillan, New York, 1981, p. 227. (B–D) Diagrammatic views of lampbrush chromosomes. (B) The two homologous chromosomes (left) are held together by two chiasmata. A portion of the central chromosome axis (right) shows that two loops with identical morphology emerge at a given point, evidence that each chromosome has already split into two chromatids. (C) Accidental stretching of a chromosome reveals the continuity of the loop axis with the central axis. (D) A single loop pair, showing the single DNA molecules on which RNA chains (indicated by fuzzy shading) are being transcribed. From J. Gall, *Brookhaven Symp. Biol.* **8**, 17 (1955).

packed and contains regions in which large loops, e.g., of 20–120-kb size, are uncoiled enough to allow transcription factors and other proteins to locate their target sequences.[91c] Many models of interphase chromatin have been proposed.[92,93]

4. The Cell Nucleus

It has been clear for many years that the nucleus has a well-organized structure. However, techniques such as fluorescent labeling have only recently been used to provide important details. These studies show that individual interphase chromosomes occupy discrete territories within the nucleus.[94–96] Parts of the chromosomes may be unfolded and active in transcription, while others are more tightly coiled. Some parts, known as **heterochromatin**,[97–98a] are very tightly coiled and metabolically almost inert. These regions include the highly repetitive DNA of telomeres and centromeres as well as other regions and complete inactivated female X chromosomes. Replication, transcription, and RNA splicing complexes are found at distinct locations within cells.[99–102a] They are apparently fixed, perhaps attached to the inner nuclear membrane, while the DNA passes through the complexes.

The nuclear matrix. The lipid bilayers, the histones and other soluble proteins, and the DNA can all be removed from nuclei by extraction and enzymatic digestion. An insoluble residue, the **nuclear matrix,** is left.[103–107] Largely protein in nature, this matrix is spread throughout the nucleus. Remnants of the membranes remain in the form of proteins that were in or along the bilayer. The nucleolus is clearly defined. The DNA appears to be bound to the nuclear matrix proteins. A specific 320-kb piece of a *Drosophila* chromosome has been mapped and used to locate nontranscribed scaffold (or matrix) attachment regions of DNA bound to matrix proteins. These were found at intervals of 26–112 kb, the intervening loops containing up to five or more genes.[105,106,108,109] A 120-kDa protein together with topoisomerase II (Section C,2)[102,104] may be components of a **nuclear scaffold** that constrains the loops of DNA. The scaffold may also provide locations for the complexes of proteins involved in replication and other processes.[103] The **matrix attachment regions** (MARs) may also act as **insulators** that shield promoters for transcription within certain loops from control elements such as **enhancers** that may be present in adjacent loops.[91b,110] At least one nuclear matrix component becomes phosphorylated and moves to the nuclear poles during mitosis.[111]

Other nonhistone nuclear proteins. Polyacrylamide gel electrophoresis revealed more than 450 components in HeLa cell nuclei. Most are present in small amounts of <10,000 molecules per cell and are not detectable in cytoplasm.[112] Among the more acidic proteins are many enzymes including RNA polymerases. There are also gene repressors, hormone-binding proteins, protein kinases, and topoisomerases.[113] Among the six most abundant nonhistone nuclear proteins in the rat are the cytoskeletal proteins myosin, actin, tubulin, and tropomyosin.[114]

A group of small (<30 kDa) proteins, the **high mobility group (HMG)** proteins,[112,115] can be extracted from chromatin with 0.35 M NaCl. Two pairs, HMG-1 + HMG-2 R (renamed **HMGB**)[115a,b] and HMG-14 + HMG-17 (renamed **HMGN**),[115b,c] are present in nuclei of all mammals and birds. HMG-14

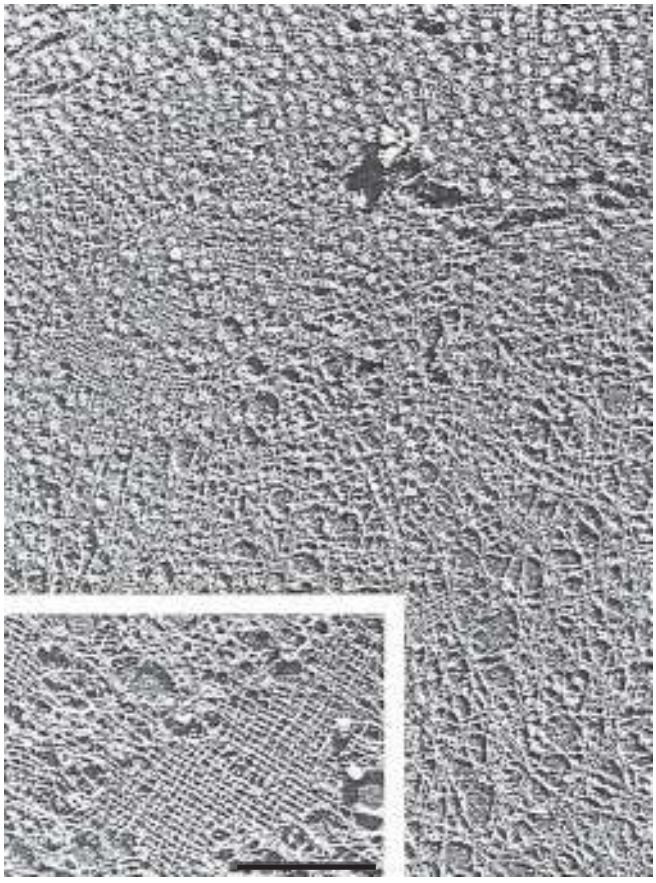


Figure 27-7 Native nuclear lamina of *Xenopus* oocytes. Freeze-dried metal-shadowed nuclear envelope extracted with Triton X-100, revealing the nuclear lamina meshwork partially covered with arrays of nuclear pore complexes. Inset, relatively well-preserved area of the meshwork of nearly orthogonal filaments from which pore complexes have been mechanically removed. Bar, 1 µm. From Aebi *et al.*[121]

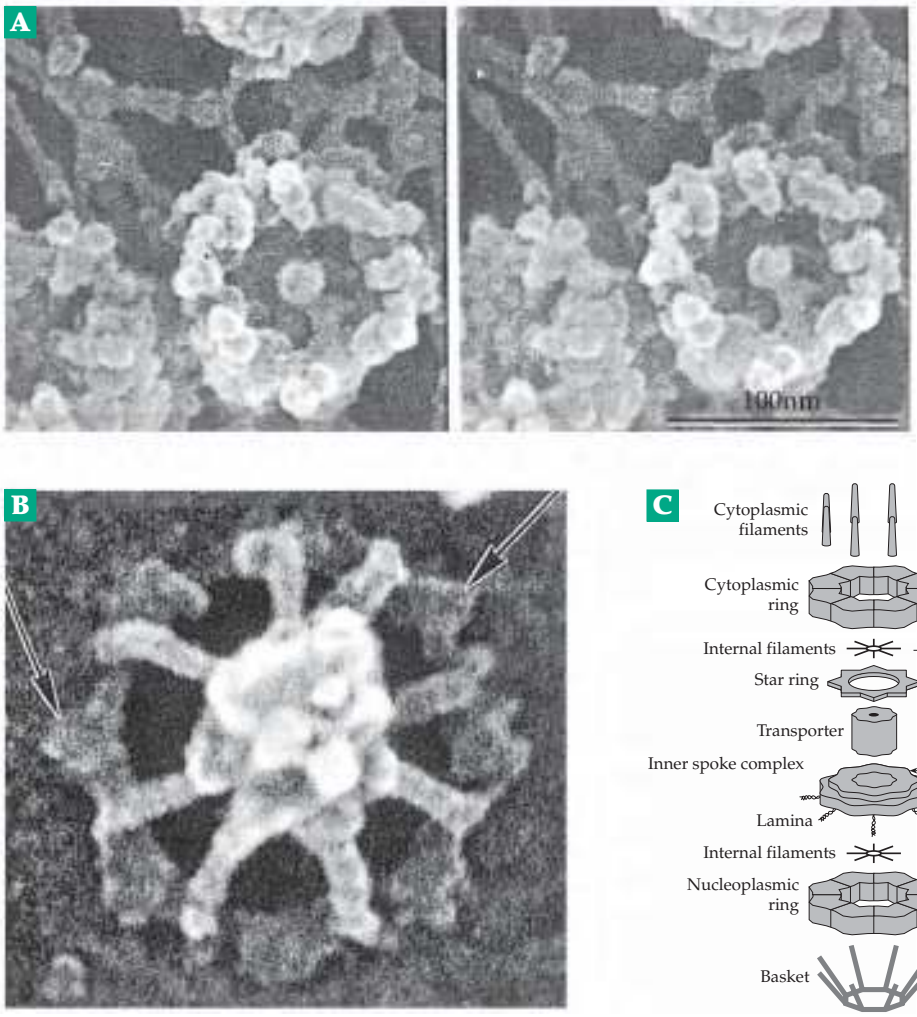


Figure 27-8 Electron micrographs of nuclear pores from a *Xenopus* oocyte. (A) Cytoplasmic face of a detergent-extracted membrane. Stereo view showing structure of a single pore. Lamin fibers are seen in the background. (B) Nucleoplasmic face of a trypsin-treated sample. The trypsin-resistant basket is clearly visible. Arrows point to the triangular pieces of the nucleoplasmic ring. (C) Diagram illustrating various structural features of the pores and the way in which they may fit together. From Goldberg and Allen.[120] Courtesy of Martin W. Goldberg.

and HMG-17 appear to be concentrated where active transcription is occurring. HMG-17 undergoes a variety of modifications including acetylation, methylation, phosphorylation, ubiquitination, glycosylation, and ADP-ribosylation, suggesting that it may assist in regulation of transcription. Other members of the HMG family bind more specifically to certain DNA sequences or structures.[115d,e]

Nuclear membrane, pores, and lamina. The nuclear membrane consists of two bilayer membranes separated by a 40- to 60-nm perinuclear space.[116,117] Both of the membranes and the intervening space are penetrated by large proteinaceous **nuclear pores**.[118–120b] The two membranes are fused together around the pores. On the inside of the innermost nuclear membrane is a relatively insoluble meshwork of intermediate filaments, the **nuclear lamina** (Fig. 27-7). It acts as a scaffold for the pores and may also interact with chromatin.[113,121–123a] Three proteins, the **lamins** A, B, and C, are the major components of the nuclear lamina. The A and C lamins, which differ only at their C termini, are homologous to cytoplasmic intermediate filament proteins and may exist as a network of coiled-coil polymers.[124] Lamins are phosphorylated, and the lamina is disassembled during the prophase of mitosis, apparently under the influence of the cdc2–cyclin B complex (Eq. 26-3),[125] and reappears in telophase at the end of mitosis.

Nuclear pores consist of octameric rings of protein subunits with a complex structure and an outside diameter of 120 nm and an inside diameter of ~80 nm (Fig. 27-8). A pore may consist of as many as 80 to 100 different proteins and have a mass of ~125 MDa.[119,126,126a,b] The pores are sometimes seen to be blocked by ~35 nm granules, perhaps pre-ribosomes. Transport through nuclear pores occurs in both directions. Numerous proteins enter the nucleus where many bind to RNA and are then exported as ribosomal subunits. Transfer RNAs and messenger RNAs must also be exported. A family of nuclear transporters known as **importins** and **exportins** mediate the movement of protein and RNA through the pores.[119,126b,127–130] They depend upon the G protein **Ran**, which also functions in spindle formation (p. 1503), and the hydrolysis of GTP.[131–132b] Large

conformational changes may be linked to the transport cycle[133,134] and also have been observed in response to changes in the Ca^{2+} level.[135,136]

The nucleolus. This organelle is a dynamic structure, which breaks down and reforms in each mitotic cycle. It is organized around clusters of genes for the 28S, 18S, and 5.8S ribosomal RNA subunits. Both chromatin and ribonucleoproteins that transport the rRNA out to the cytoplasm are present.[95,137] The major nucleolar protein **nucleolin** is characterized by its glycine-rich C terminus, which contains seven repeats of GGRGG and also contains N^{γ},N^{γ}-dimethylarginine.[138,139] Nucleolin may control transcription of the DNA that carries the rRNA genes. It appears to be required for ribosome synthesis (Chapter 29) and for attachment to the nuclear matrix. A variety of other proteins in smaller amounts are also needed in nucleoli. In addition to the true nucleoli other regions described as **speckles** and **coiled bodies** may develop around RNA splicing centers.[95,139a]

B. Organization of DNA

A vast amount of new information on DNA sequences and structures is available for yeast, *Arabidopsis*,[140] *Drosophila*, the nematode *Caenorhabditis elegans*,[141] human beings, and other species (see Chapter 26). Nevertheless, it may be worthwhile to consider older discoveries, some dating back 20 years or more.

When DNA is cut into ~10-kb fragments by shearing and the fragments are denatured by heat, the renaturation of the resulting single-stranded fragments upon cooling takes place in two or more steps. Some material reforms double helices rapidly, whereas other material is slow to renature (Fig. 5-46). At least four kinetically distinct fractions have been recognized: (1) About 70% of mammalian DNA appears to exist largely as single copies, i.e., with unique sequences. (2) About 20% of the total DNA is moderately repetitious, containing sequences that may be present 10^3–10^5 times. (3) About 5% reassociates very rapidly and is identified as highly repetitive **satellite DNA** of which there may be ~10^6 copies. (4) A smaller fraction contains long palindromic sequences. As a consequence, the single-stranded pieces can fold back almost instantaneously to form hairpin structures.[142,143] These differences also show up in density gradient centrifugation in CsCl, whether on large pieces of DNA[144] in the presence of ligands such as Ag^+ or on smaller restriction fragments which have often been studied by polyacrylamide gel electrophoresis.

If there are ~35,000 human genes, protein or RNA coding sequences must occupy only 2–3% of the genome.[145] The pufferfish *Fuga rubripes* probably has almost as many genes as we but only ~13% as much DNA. In further contrast the newt *Triturus cristatus* has six times as much DNA as a human.[146] The compact genome of the green plant *Arabidopsis thaliana* occupies only 120 Mb. In contrast are the 415-, 2500-, and 5300-Mb genomes of rice, maize, and barley, respectively.[147] What is the function of all of the apparently noncoding DNA present in some organisms? It is often viewed as "junk," whose only function may be to facilitate evolutionary changes in the genome. However, there is doubtless important undiscovered information in these regions.[145] It has been very hard to determine sequences, in part because of the large amount of highly repetitive DNA. Some regions have been "unclonable" in prokaryotic systems because of the presence of transposon-like sequences or "kinkable" elements (TG•CA steps), palindromes, etc.[145a,b]

1. Repetitive DNA

Rapidly renaturing DNA fragments often have a different base composition than the bulk of the DNA and, consequently, often separate as small satellite bands upon centrifugation in a CsCl gradient. Satellite DNA is usually associated with regions of the chromosome that do not unravel in telophase as does the bulk of the DNA. Satellite DNA usually consists of short highly repetitive sequences,[148] which occur in large clusters of up to 100 Mb of DNA, often near centromeres or telomeres or on a Y chromosome.[146] The DNA of a satellite band from the kangaroo rat contains the sequence 5'-GGACACAGCG-3' repeated so often that it accounts for 11% of the entire DNA of the cell. Longer repeating sequences ~170 bp are also often present as are **microsatellites** of 2- to 5-base-pair repeats. At least 30,000 microsatellite loci are present in the human genome.[149]

Centromeres. The attachment of spindle fibers to chromosomes depends upon the segments of DNA known as centromeres to organize the attached kinetochores. In the yeast *Saccharomyces cerevisiae* a 120-bp region containing three short conserved sequences is present in the centrosomes of all ten chromosomes.[150,151] This may fold into a distinctive looped structure.[152] Human centrosomes are large and complex,[153] but the DNA is highly repetitive, giving rise to **α-satellite DNA**.[154–157b] Sequences such as $(TGGAA)_n$ are repeated many times. Such sequences can form self-complementary looped structures containing some unpaired guanines that intercalate and stack between sheared G•A pairs.[158,158a] Complex regional centromeres involving kilobases or megabases of DNA have also been identified in fission yeasts, *Drosophila*, and green plants.[156,157] The great variability indicates that centromeric sequences undergo rapid

evolution. This may be related to the fact that of the four cells formed by female meiosis only one becomes an egg.[158b] A series of unique **centromeric proteins** (CENP-A to CENP-G) bind to the DNA sequences of centromeres[159–162a] and direct the formation of the kinetochores. Even for the simpler centromere of budding yeasts, kinetochores have a complex structure.[150] The CENP proteins were first identified as autoantigens in sera of patients with the autoimmune disorder **scleroderma** (Chapter 31).[160,162]

Telomeres. The DNA sequences at the chromosome ends have a TG-rich strand, such as the $(TTGGGG)_{50–70}$ (*Tetrahymena*)[163] and the $(TTAGGG)_n$ of both human and trypanosome chromosomes.[164–166] The complementary DNA strand is CA-rich. The *S. cerevisiae* telomers have ~350 base pairs containing the sequences ($TG_{1–3}$ / $C_{1–3}A$) as well as one or more copies of a 6.7-kb nonrepetitive sequence and other elements.[166,167] In many species the repetitive telomeric sequences have 3' poly(G) tails at the ends of the DNA molecules. These tails are able to form G quartet structures (Fig. 5-26 and Chapter 5, Section C,4). A variety of telomere-binding proteins have been isolated.[167a–d] Some of these bind to G quartet structures[168–169a] and some, such as **RAP1** of yeast,[170–171a] to double-stranded telomere repeat regions. Special problems associated with replication of telomeres are discussed in Section C,8.

Short interspersed sequences (SINES). Much of the reiterated DNA is present in repeated segments 100–500 bp in length that lie between 1- to 2-kb segments of unique DNA.[172–174] The best known example is the human ***Alu*** family,[175–177b] so-called because it contains a site for cleavage by restriction enzyme *Alu*I. Sequences of the *Alu* family also exists in other primates and in rodents. The ~300-bp *Alu* sequence is reiterated over 500,000 times in the human genome with various sequence alterations, but an 80–90% homology. This sequence (Fig. 27-9) consists of two similar ~130-bp segments called the "left monomer" and "right monomer." The right monomer contains a 31-bp insertion and the left end carries a poly(dA) sequence. In addition there is a short 7- to 20-bp sequence, which is variable between different *Alu* sequences but is directly repeated at each end of a given *Alu* sequence.

The *Alu* sequence has strong homology with the 7S RNA (Table 5-4) that is part of the signal recognition particle involved in transport of newly synthesized peptide chains across the membranes of the ER (Chapter 10). *Alu* sequences are transcribed into hnRNA, the precursors to mRNA. Some *Alu* sequences are present in intervening sequences (introns) within genes and others are in noncoding sequences between genes. Sharp[176] suggested that specific proteins in the nucleus may bind to the *Alu* sequences, preventing hnRNA from leaving the nucleus before it has been processed to remove introns and other sequences absent in mRNA. However, the presence of the poly(dA) regions and the direct terminal repeats suggests that the whole *Alu* sequence is **pseudogene** derived from a **retroposon**, a type of transposon that originated from an RNA molecule. *Alu* and other SINES contain an RNA polymerase III promoter and are transcribed. Active retroposons form reverse transcripts (cDNAs) that can be integrated at various points in the genome (see Section D,4 and Chapter 28). One theory is that retroposons have no biological function but have invaded the genome at random locations.

Many SINES and other families of repetitive sequences have been characterized by the presence of a restriction enzyme cleavage site in each copy of the sequence. *Eco*RI cuts the previously mentioned α-satellite DNA.[177] A 319-bp reiterated sequence in the human genome surrounds a *Hin*f site,[178] and *Sau*3A cleaves a 849-bp sequence with ~1000 copies per haploid genome.[179] In the hermit crab 30% of the genome consists of repeated sequences, one 156-bp unit occuring ~7 million times. Many identical 14-bp GC-rich inverted repeats are present.[180] *Neurospora* and yeast both contain many copies of the GC-rich palindromic sequence 5'-CCCTGCAGTACTGCAGGG-3', which contains the two underlined *Pst*I sites.[181] Some repetitive sequences such as the **CAT** family and the **homeodomain sequence** (Chapter 28, Section C,6) are found in the control regions preceding the 5' ends of groups of genes that are regulated coordinately.[182] Some repetitive sequences seem to be unstable in the genome and may be excised and lost or may increase in number during aging.[183] Repeated DNA

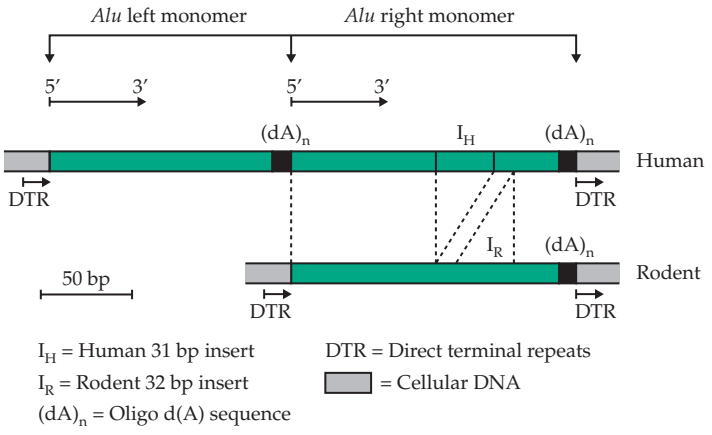


Figure 27-9 Structure of human and rodent *Alu* sequences in DNA. From Ullu.[175]

sequences have apparently originated not only with 7S RNA but also from mRNA or tRNA molecules.[172] A characteristic SINE in rodent DNA is known as the ID (identifier) sequence because it was once thought to be a marker for genes transcribed in neural tissues. Rat DNA contains ~130,000 copies of the ID sequence.[172a] Other **clustered repeats** are frequently found between genes that are present in large numbers. These include the genes for ribosomal RNA, tRNA, small nuclear RNAs, and histones. Triplet repeats are considered in Chapter 26, Section G,3.

Long interspersed repeat sequences (LINES). These moderately repetitive sequences may be several kb in length. Just one type seems to be abundant in each mammalian species.[122] The human L1 sequence is over 6 kb in length. Like the *Alu* sequence it has a poly(dA) sequence at the 3'-end, but it does not contain short terminal direct repeats. These interspersed repeat sequences usually contain genes which may be functional, but many of the copies contain pseudogenes or genes that are randomly truncated, having lost a segment from the 5' end. These sequences, too, seem to have been dispersed throughout the genome by retroposons.[122,184,185] Human chromosome 22, sequenced in 1999, contains within its 33.4 Mb of euchromatin at least 545 genes, 134 pseudogenes, and 8043 L1 sequences (9.7% of the DNA), as well as 20,188 *Alu* sequences (16.8% of the DNA) and many other interspersed repeats.[186] Repetitive DNA sequences are less common in prokaryotes, but they do exist. For example, many dispersed and clustered repeating units are present in *Halobacterium* DNA.[187]

2. Genes for Ribosomal RNA and Small RNA Molecules

Most genes are present as one copy each per haploid genome. However, there are many copies of the genes for ribosomal RNA and tRNA. In *Xenopus* DNA there are ~450 repeats of the 28 S and 18 S rRNA genes on one chromosome and ~24,000 copies of the 5 S RNA genes at the ends of the long arms of most of the chromosomes.[188] Nontranscribed spacer regions lie between the repeats of the 28S and 18S gene pairs, as can be seen with the electron microscope in Fig. 28-17. The gene for 5 S RNA has a high GC content. A denaturation map of the DNA shows easily melted regions separated by shorter 120-bp sequences, apparently of high GC content and presumably coding of the 5 S RNA. The easily denatured AT-rich spacers are ~630 bases long. Using restriction enzymes much of this DNA can be cut into segments that contain repeats within repeats. One 15-unit polynucleotide contains the sequence $A_4CUCA_3CU_3G$ repeated about 30 times.[189]

About 200 copies of rRNA genes per haploid genome are located at the constrictions in the short arms of human chromosomes 13, 14, 15, 21, and 22 (Chapter 26, banner). As in *Xenopus*, clusters of tandemly repeated 5S RNA genes are found at the ends of the long arms of most chromosomes. A similar organization of rRNA genes is found in the rat.[190] The ten different chromosome ends carrying rRNA genes in the human diploid nucleus come together to form the nucleolus, the site of synthesis of ribosomal subunits.[191] At first there are ten small nucleoli, but these fuse to form the single highly structured but membraneless nucleolus. The rRNA genes of the macronucleus of *Tetrahymena* have been "amplified" and are found on linear 21-kb palindromic molecules, about 10^4 copies being present per cell. Initiation of transcription begins near the center and proceeds outward in both directions.[192] These short chromosomes contain typical telomeric ends.

Genes for the tRNAs are spread throughout the genome of bacteria, mitochondria (Fig. 18-3), chloroplasts,[193] and eukaryotic nuclei. They sometimes occur in clusters but more often are far apart. In *Drosophila* there are probably at least 600 tRNA genes, many occurring in pairs of opposite polarity, i.e., as inverted repeat sequences. The genes for small nuclear RNAs U1–U6 (Chapter 28) are organized in a variety of ways. The ~100 human U1 genes occur on a single chromosome, perhaps organized in a tandem array and interspersed with as many as 10,000 defective pseudogenes.[194]

3. Other Gene Clusters and Pseudogenes

Closely related structural genes often occur in clusters.[195] Among these are clusters of **immunoglobulin** genes (Fig. 31-17) and clustered genes for the α and β **globins** (Fig. 27-10), which encode the protein sequences for the hemoglobins. The human α globin gene cluster occupies about 30 kb on chromosome 16 and the β globin genes 60 kb on chromosome 11.[196–197d] The α1 and α2 genes (Fig. 27-10) encode identical peptides, while the related ζ gene encodes the corresponding subunit of embryonic hemoglobin. The β cluster includes, in addition to the adult gene, a pair of fetal globin genes (γ^G and γ^A) differing by only one amino acid (Gly vs Ala) at position 136, the embryonic ε chain, and the minor adult δ chain.

Besides the functional genes the globin cluster contains **pseudogenes**, which are given the prefix ψ in Fig. 27-10. These are nonfunctional genes, which appear to encode peptides homologous to the known globins. However, they contain mutations that prevent expression. For example, deletion of a single nucleotide near the beginning of the pseudogene will scramble the genetic message by changing the reading

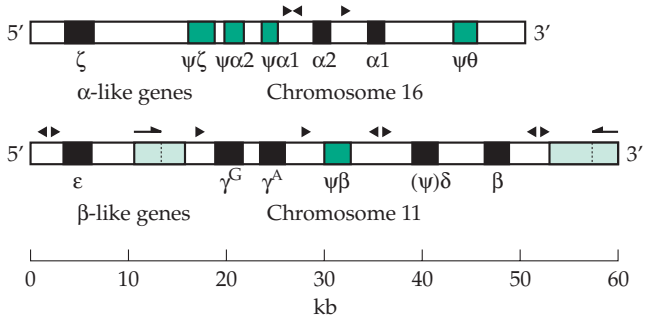


Figure 27-10 Organization of the globin genes on human chromosomes 11 and 16. The composition of the various embryonic, fetal, and adult hemoglobins is also indicated. Closed boxes indicate active genes and open boxes pseudogenes. The triangles (▶) indicate *Alu* repetitive sequences and their orientation. The shaded boxes indicate Kpn repeat sequences and the half-arrows their respective orientation. The Kpn sequence between the ε and γ^G genes in fact consists of two tandemly linked Kpn repeats. From Karlsson and Nienhuis[196] and Proudfoot.[197]

frame. Globin pseudogene ζψ has a nonsense mutation at codon 6, ψα1 contains a whole set of mutations that prevent transcription and translation, while ψα2 is mutated almost to the point of being unrecognizable as a globin relative.[198]

Do pseudogenes have a physiological function? The presence of pseudogenes in close proximity to functional genes suggests that there may be an as yet unrecognized controlling function.[198] On the other hand, they may be relics of evolution. If a gene family arises by duplication and mutation of duplicated copies, we may expect to find gene copies that have not been selected as useful but are still present in the genome.

There are regions of the genome that seem to be largely single copy DNA with few reiterated sequences or gene clusters.[144] While some gene families exist as clusters others are dispersed throughout the genome.

4. Introns, Exons, and Overlapping Genes

A major difference between prokaryotes and eukaryotes is the presence of intervening sequences (introns) between the coding sequences (exons) in eukaryotes. Introns are especially numerous in higher organisms. For example, the gene for the myosin heavy chain present in rat embryonic skeletal muscle[199] encodes a 1939-residue peptide but occupies a length of 22 kb of DNA. The gene is split into 41 exons, whose transcribed RNA must be cut and spliced at 40 places to form the mRNA. By comparison the corresponding gene from the nematode *Caenorhabdatis elegans* is far less fragmented. It is not clear why some genes have so many introns and others so few. The very compact chromosomes of viruses and most bacteria do not contain introns. However, they are sometimes present in mitochondrial and chloroplast genes.

A surprise, which was first recognized in viral RNA and DNA, is that genes sometimes overlap. For example, two proteins, one long and one shorter, are synthesized starting at the same point in the RNA genome of phage Qβ.[200] In the DNA of phage ϕX174 the third nucleotide of the stop signals for some genes are also the first nucleotides for the start signal for translation of the next gene. Pairs of genes have been found in which one of the genes of the pair is found completely inside the other gene but is translated in a different reading frame.

A four-base overlap between dihydrofolate reductase and thymidylate synthase has been found[201] in the DNA of phage T4. A transposable DNA insertion sequence (see Section D,5) in *E. coli* encodes two genes, one of which is contained within the other and which is transcribed from the opposite strand of DNA.[202] The double-stranded RNA of a reovirus produces two peptides from the same sequence using two different AUG initiation codons in different reading frames.[203]

Overlapping obviously limits severely the mutational alterations that are allowable. Perhaps it is for this reason that eukaryotic genes seem to be dispersed in widely separated locations and overlap is rare. However, the human type IV collagen genes for the α1 and α2 chains are encoded on opposite strands with their 5' flanking regions overlapping.[204] Some introns contain genes.[205,206] Chlorarachniophyte algae contain a multimembraned chloroplast thought to be a vestigial remnant of an endosymbiont. Its 380-kb genome consists of three short chromosomes that encode overlapping genes and contain the shortest introns (18–20 base pairs) known.[207]

5. DNA of Organelles

Mitochondrial DNA (mtDNA), discussed in Chapter 18, varies in size from <6 kb to 367 kb.[208] The small circular mtDNA of animals is extremely efficiently packed with genes for tRNAs, rRNA, and a small number of protein subunits (Fig. 18-3). However, the 78-kb yeast mtDNA contains many long AT-rich spacers as well as long introns, some of which contain genes for splicing enzymes. Otherwise, it is similar to animal mitochondrial DNA. Mitochondrial genomes of higher plants are much larger; that of maize is a 570-kb circle, which contains both direct and inverted repeat sequences. Recombination between these sequences is apparently the origin of smaller incom-

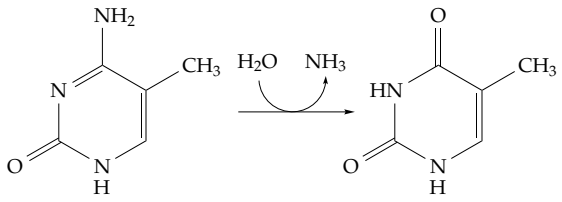
plete circular genomes that are also present. Plant mtDNA appears to contain the genes present in animal mitochondria plus additional genes. For example, one encodes a 5S RNA. The most unusual mtDNA is found in hemoflagellates such as *Trypanosoma* and *Leishmania*. The single large mitochondrion or **kinetoplast** located near the base of the flagellum contains a network of catenated circular DNA molecules (Fig. 5-16).[209,210] The genetic makeup is similar to that of other mtDNA, but the rRNA genes are unusually short. Chloroplast DNA, which varies in size from 120 to 160 kb, is discussed in Chapter 23.

6. Methylation of DNA

As mentioned in Chapter 5, a significant fraction of the pyrimidine and purine bases in DNA is methylated. One function of such methylation in bacteria, discussed in Chapter 26, is to protect against the action of restriction endonucleases.[211] For example, the gene for the well-known *Eco*RI endonuclease is carried in *E. coli* cells by an R factor. This plasmid also carries (just 29 base pairs away) the gene for a 326-residue **N^6-adenine methylase**.[212–213a] This enzyme uses *S*-adenosylmethionine (AdoMet) to methylate the two adenines (marked by asterisks) in the six-base-pair recognition sequence 5'-G A A* T T C / 3'-C T T A* AG converting them to N^6-methyladenines (m^6A). Other DNA methyltransferases place methyl groups on N-4 of cytosine or on C-5 of cytosine.[214–215a] The latter utilizes a mechanism illustrated in Eq. 12-4.[216,217] Such enzymes are components of both type II and type I restriction–methylation systems.[218] However, most of the m^6A in the *E. coli* chromosome arises from action of a different methylase, one that recognizes the palindromic sequence 5'-GATC and methylates adenines in both chains.[219] This **DNA adenine methylase**, a product of the *dam* gene, plays an important role in mismatch repair, transposition, regulation of transcription, and initiation of DNA replication.[220] The same methylase regulates at least 20 genes induced during infection by *Salmonella typhimurium*. Some of these genes are essential to virulence.[221] A similar methyltransferase appears to control differentiation of the stalked *Caulobacteria* cells.[220]

CG doublets. The only modified base commonly found in eukaryotes is 5-methylcytosine,[222,223] which upon deamination becomes thymine (Eq. 27-1). Most methylation occurs when C is followed by G. Usually 60–90% of all 5'-CG sequences (CpG sequences) in eukaryotic DNA are methylated. However, the fraction of methylated cytosine varies from almost zero for *Drosophilia, Caenorhabditis*, and *Saccharomyces* to as much as 30% in higher plants.[224]

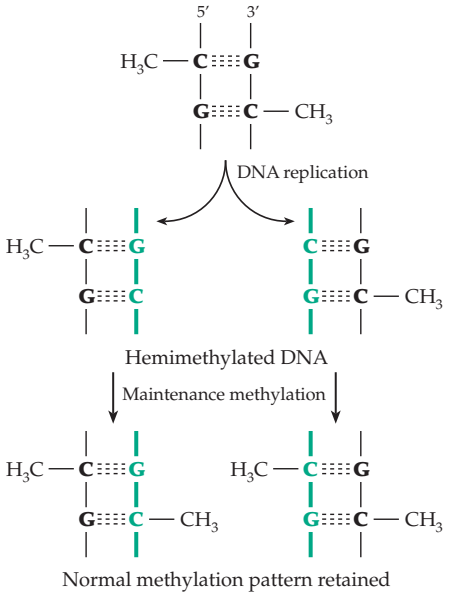
When CG pairs are methylated on both cytosines,



(27-1)

an interesting result arises upon replication of the DNA. The methyl groups don't prevent replication, but only one of the DNA strands in each daughter duplex is methylated (Eq. 27-2). However, additional methylation of DNA occurs within three hours of DNA replication by a **maintenance methyltransferase** that recognizes methylated CpG sequences in the old DNA strand and methylates the cytosine in the 3′-GpC of the newly synthesized strand.[220,225,226] Mammalian cells contain at least three different DNA (cytosine-5) methyltransferases.[227–229b] Enzymatic **demethylation**, which converts m^5C residues back to cytosine, may also occur.[230–233] It has been difficult to demonstrate[233a] but enzymatic demethylation of 5-methylcytosine has been reported.[233b,c] The observed loss of methylation during embryonic development (Chapter 32) may be a result of loss of the maintenance methylase. Most CpG doublets are found in large "islands" of several hundred bases to about 2 kb in length, which are unusually rich in G + C.[234,235] These islands lie near the 5' ends of ~60% of all human genes[236] and near origins of replication.[237] They are also found in regions of the DNA that are compacted into heterochromatin. The genes in methylated regions tend to be "silent," i.e., they are not actively transcribed.[236] Demethylation may permit transcription.

The degree of methylation of the CG doublets is variable both in position within a chromosome and



(27-2)

with stage of development. Cytosine methylation is essential to embryonic development,[238] and mice lacking the maintenance DNA methytransferase are developmentally retarded and die at mid-gestation.[239] Not all CG sequences are methylated and various patterns of DNA methylation are generated at different stages of development by rounds of methylation and by the action of demethylases.[232] The maintenance methylase (Eq. 27-2) ensures that a stable methylation pattern persists until altered by new rounds of methylase or demethylase action. A parallel is found in the prokaryotic *Caulobacter* in which three chromosomal sites are fully methylated in swarmer cells, become hemimethylated in stalked cells, and are fully methylated again just prior to cell division.[220] See Fig. 32-1.

DNA methylation affects transcription, either directly by preventing the binding of transcription factors or indirectly via a series of binding proteins specific for methylated CG doublets. In early stages of embryonic development there is very little methylation, but some genes are quickly silenced as methylation takes place. Heterochromatic regions including inactivated X chromosomes are heavily methylated. However, additional alteration in chromatin is required for complete silencing of genes.[236,240] Recent studies indicate that an abundant mammalian protein binds to the methylated DNA along with a **histone deacetylase**.[239,241–244] The latter acts on acetylated histones to free lysine side chains, which may interact in an inhibitory manner with the DNA.

In female mammalian cells most of the genes on one of the two X-chromosomes are completely inactivated. DNA methylation plays a major role in this process.[244,245] A perfect correlation has been observed between 5'-methylation of cytosines in CpG islands and inactivation of X-chromosome genes.[246] Methylation may also play a role in recombination and repair.[247] Methylation of DNA decreases with increasing age.[248] It increases as a result of oncogenic transformation of cells.[249] Some other modifications of DNA largely limited to bacteriophages are discussed on p. 234.[247,250]

Imprinting. With the exception of X-linked genes each person has two copies of each gene, one of maternal and one of paternal origin. Both copies of most of these genes are expressed. However, a few of the genes receive from one parent or the other an **imprint**, a mark that distinguishes the parental origin.[251–253] Such imprints are maintained in cells through embryonic development but are erased in embryonic gonads to allow for a new imprint in the germ cells. Imprinting depends upon DNA methylation, and all imprinted genes show the presence of differentially methylated regions.[253] See also Chapter 32, Section A,1.

C. Replication

Following the discovery of the double helix and the enthusiasm that it engendered many people thought that the synthesis of DNA was simple. The nucleotide precursors would align themselves along separated DNA template strands and perhaps spontaneously snap onto the growing chains. In fact, replication is a complex process that requires the cooperative action of many different gene products and perhaps an association with membrane sites. The matter is made more complex by the fact that some of the enzymes involved in replication are also required in the processes of genetic recombination, in repair of damaged DNA molecules, and in defensive systems of cells.

1. Early Studies

That the DNA content doubles prior to cell division was established by microspectrophotometry. It was clear that both daughter cells must receive one or more identical molecules of DNA. However, it was not known whether the original double-stranded DNA molecule was copied in such a way that an entirely new double-stranded DNA was formed or whether, as we now know to be the case, the two chains of the original molecule separated. The latter is called **semiconservative** replication, each of the separated strands having a new complementary strand synthesized along it to form the two identical double-stranded molecules.

The first definitive evidence for semiconservative replication was reported by Meselson and Stahl[254] in 1958. Cells of *E. coli* were grown on a medium containing isotopically pure $^{15}NH_4^+$ ions as the sole source of nitrogen. After a few generations of growth in this medium the DNA contained exclusively ^{15}N. Then the cells were transferred abruptly to a medium containing $^{14}NH_4^+$ and were allowed to grow and to double and quadruple in number. At various stages DNA was isolated and subjected to ultracentrifugation in a CsCl gradient. Small but easily detectable differences in density led to separation of dsDNA molecules containing only ^{15}N from those containing partly ^{15}N and from those containing only ^{14}N. At the beginning of the experiment only DNA containing entirely ^{15}N was present. However, after one generation of growth in the ^{14}N-containing medium, the density of *all* the DNA was such as to indicate a content of one-half ^{14}N and one-half ^{14}N. After a second generation of growth half of the DNA still contained both nitrogen isotopes in equal quantity, whereas half contained only ^{14}N, exactly the result expected for semiconservative replication. A similar experiment using 5-bromodeoxyuridine, a thymidine analog, is shown in Fig. 27-11.

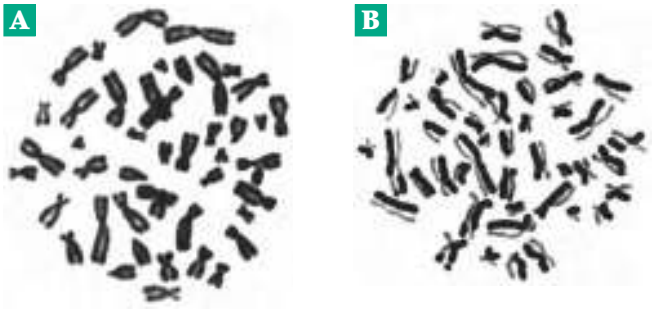
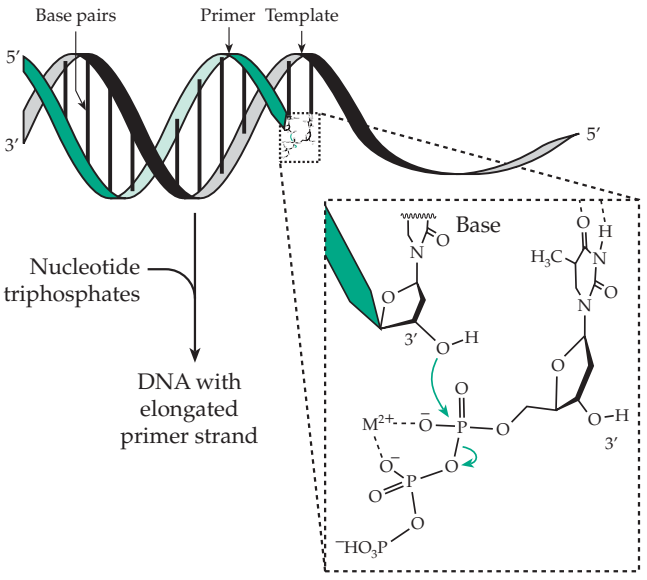


Figure 27-11 (A) Human chromosomes after one replication in the presence of 5-bromodeoxyuridine (BrdU). Both chromatids of each chromosome contain BrdU in one strand of the DNA duplex and normal thymidine in the other. (B) After two replications in the presence of BrdU one chromatid of each chromosome contains BrdU in both strands of the duplex and stains strongly with a special differential staining procedure. The other chromatid contains only normal thymidine in one strand of the duplex and is not stained. Courtesy of Carolina Biologicals.

Autoradiography. Later a technique of direct autoradiography of DNA using ^{3}H-labeled thymidine[255] was applied by Cairns.[256,257] Cells of *E. coli* were grown on a medium containing the radioactive thymidine for various times but typically for 1 h (~2 generations). The cells were then ruptured, the DNA was spread on a thin membrane filter, and autoradiograms were prepared. When the DNA contained [^{3}H]thymidine the exposed trace in the autoradiogram could be followed around the entire 1.1–1.4 mm circumference of the spread DNA molecule. Molecules partially labeled and in the process of replication could also be identified. After two hours of growth in the presence of ^{3}H-labeled thymidine about half of the bacterial DNA was fully labeled, but half contained regions that were labeled only half as heavily. They presumably contained ^{3}H label in a single strand and, therefore, represented unreplicated regions. All of the molecules had undergone one round of replication with tritiated thymidine to yield lightly labeled molecules; parts of the molecules had not completed the second round. The more heavily labeled regions were interpreted as fully replicated. The shapes of the "replication figures" suggested that DNA is synthesized in a continuous manner starting from one point and continuing around the circular molecule at a constant rate. Although subsequent experiments (considered in Section 4) show that replication is usually **bidirectional**, the experiments of Cairns were important because they introduced a technique for direct visualization of replication *in vivo*.

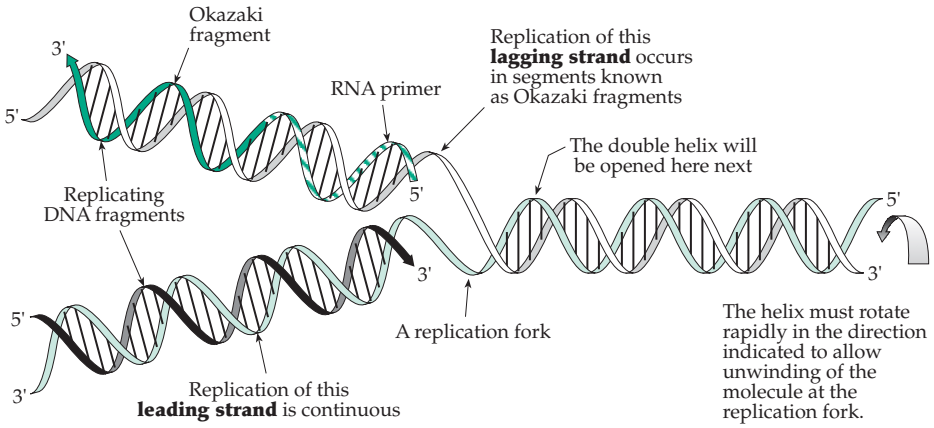


(27-3)

The chemistry of DNA polymerization. What are precursors of DNA? Early experiments showed that the nucleoside [^{3}H]thymidine was efficiently incorporated into DNA, but for energetic reasons it seemed unlikely that thymidine was an immediate precursor. Evidence favoring the nucleoside triphosphates was provided in 1958 when Arthur Kornberg identified a DNA polymerase from *E. coli*. Kornberg's enzyme, now known as **DNA polymerase I**, was isolated in the amount of 600 mg from 90 kg of bacterial cells[258,259] (over 400 molecules of enzyme per cell). The 928-residue enzyme displayed many of the properties expected of a DNA-synthesizing enzyme. It requires a **template strand** of DNA as well as a shorter **primer strand**. As indicated in Eq. 27-3, the enzyme recognizes the 3' end of the primer strand and binds the proper nucleoside triphosphate to pair with the next base in the template strand. Then it catalyzes the displacement of pyrophosphate, at the same time linking the new nucleotide unit onto the 3' end of the primer strand. Continuing in this way, the enzyme is able to turn a single-stranded template DNA into a double-stranded DNA in which the newly synthesized strand contains, at each point, the base complementary to the one in the template strand.[259a]

Although the action of the DNA polymerase I, according to Eq. 27-3, provided a straightforward way to form a complementary strand of DNA, it did not explain how double-stranded DNA could be copied. One problem is that the two strands must be separated and unwound. If unwinding and replication occured at a single **replication fork** in the DNA, as indicated by Cairns' experiment, the entire molecule would have to spin at a speed of 300 revolutions per second to permit replication of the *E. coli* chromosome in 20 min. It also required that some kind of a **swivel**, or at least a

nick in one chain, be present in the chromosome as indicated in the following drawing.



Another problem was posed by the fact that the two chains in DNA have opposite orientations. Thus, at the replication fork one of the new chains might be expected to grow by addition of a new nucleotide at the 3' end, while the other chain would grow at the 5' end. *If so, there should be two DNA polymerases*, one specific for polymerization at each of the two kinds of chain end. Nevertheless, despite intensive search the only DNA polymerases found added new units only at the 3' end.

Discontinuous replication and RNA primers. In 1968, Okazaki reported that during the time that replication of DNA is taking place bacterial cells contain short fragments of DNA. These are now called **Okazaki fragments** or **replication fragments**.[260] A second development was the discovery of the enzyme **DNA ligase**,[261,262] which is able to join two pieces of DNA to form a continuous chain. These two discoveries provided an explanation for the lack of a second kind of DNA polymerase. One strand, the **leading strand**, of the replicating DNA could be synthesized continuously in the 5' to 3' direction while the other strand, the **lagging strand**, would have to be synthesized in segments (the Okazaki fragments), which could then be joined by the DNA ligase. In 1971 Brutlag *et al*. reported that initiation of synthesis of the DNA of phage M13 in *E. coli* required formation of a short segment of RNA as a primer.[263] It was subsequently shown that *with few exceptions priming by RNA synthesis is always required for replication*.

Newer studies have revealed great complexity in the mechanisms of replication. As for synthesis of any polymer there are distinct steps of *initiation*, *elongation*, and *termination*. Topological problems associated with the unwinding and rewinding of the double helices and with disconnection of catenated circles and untying of knots are solved with the aid of special enzymes, the **helicases** and **topoisomerases**. Replication requires both DNA and RNA polymerases, a ligase, and ancillary proteins, some of whose functions aren't yet clear. Many of these proteins associate to form large multiprotein complexes, which are given names such as **primosome** (for priming) and **replisome**. Many bacteriophages and plasmids also replicate within cells of *E. coli* utilizing bacterial proteins as well as proteins encoded in the viral or plasmid genome.

2. DNA Polymerases

Immediately after its identification DNA polymerase I was generally assumed to be the principal chain elongation enzyme. However, an *amber* mutant of *E. coli* deficient in DNA polymerase I (gene *pol*A; Fig. 26-4A) synthesized DNA normally. This finding stimulated an intensive search for new polymerases. Two were found: DNA polymerases II (gene *pol B*)[264] and III. Both are present in amounts less than 25% of that of DNA polymerase I.[265,266] Both have properties similar to those of polymerase I, but there are important differences. By now DNA polymerases have been isolated from many organisms, many genes have been cloned and many sequences, both of bacterial and eukaryotic polymerases are known. Comparisons of both sequences and three-dimensional structures,[266a,b] a few of which are shown in Fig. 27-12, suggest that the polymerases belong to at least six families (Table 27-1). These include the RNA-dependent DNA polymerases known as **reverse transcriptases** as well as some **RNA polymerases**.[267–268b]

Some of the polymerases exist as single polypeptide chains, while others function only as large complexes. In every case a two-metal ion catalytic mechanism with in-line nucleotidyl transfer,[269] illustrated in Fig. 27-13, appears to be used by the enzymes.[267,270] Two-metal ion catalysis is also observed for phosphatases and ribozymes (Chapter 12).

Exonuclease activities, proofreading, and editing. DNA polymerase I not only catalyzes the growth of DNA chains at the 3' end of a primer strand but also, at about a 10-fold slower rate, the hydrolytic removal of nucleotides from the 3' end (**3'- 5' exonuclease activity**). The same enzyme also catalyzes hydrolytic removal of nucleotides from the 5' end of DNA chains. This latter **5'- 3' exonuclease activity**, the DNA polymerase activity, and the 3'-5' exonuclease activity all arise from separate active sites in the protein. DNA polymerases II and III do not catalyze

TABLE 27-1
Families of DNA Polymerases[a,]

Class	Name	Function	Molecular mass (kDa)
A	*E. coli* polymerase I (Pol I)	DNA excision repair	103
	Klenow fragment		68
	Bacillus subtilis Pol I[b]	DNA excision repair	
	Thermus aquaticus DNA polymerase (Taq)[c,d]	DNA excision repair	
	T7 DNA polymerase[e]	Virus replication	80
	T7 RNA polymerase[f,g]		99
	Eukaryotic Pol γ (gamma)	Mitochondrial replication	
	Eukaryotic Pol θ (theta)	DNA repair	
B	Eukaryotic Pol α (alpha)	DNA replication	180 (core)
	Eukaryotic Pol δ (delta)	DNA replication	
	Eukaryotic Pol ε (epsilon)	DNA replication	
	Eukaryotic Pol ζ (zeta)	Bypass synthesis	
	Bacteriophage T4 DNA Pol + accessory proteins[h,i]	DNA replication	43
	E. coli Pol II[j]		90
C	Bacterial DNA Pol III[p] + accessory proteins	DNA replication	~900 (holoenzyme)
D	Euryarchaeotic Pol II		
X	Eukaryotic DNA Pol β[n,o]	DNA repair	39 x 2
	Eukaryotic Pol λ (lambda)	Base excision repair	
	Eukaryotic Pol μ (mu)	Non-homologous end-joining	
	Eukaryotic Pol σ (sigma)	Sister chromatid cohesion	
Y	*E. coli* UmuC protein	SOS response	
	Eukaryotic Pol η (eta, XP-V, RAD 30)	Bypass synthesis	
	Eukaryotic Pol ι (iota)	Bypass synthesis	
	Eukarytoic Pol κ (kappa)	Bypass synthesis	
Reverse transcriptase family			
	HIV reverse transcriptase[k,l]		
	Telomerase[m]		
	RNA-dependent RNA polymerases		

[a] Based on reviews by Burgers, P. M. J. *et al.* (2001) *J. Biol. Chem.* **276**, 43487–43490 and Steitz, T. A. (1999) *J. Biol. Chem.* **274**, 17395–17398
[b] Kiefer, J. R., Mao, C., Braman, J. C., and Beese, L. S. (1998) *Nature (London)* **391**, 304–307
[c] Eom, S. H., Wang, J., and Steitz, T. A. (1996) *Nature (London)* **382**, 278–281
[d] Li, Y., Korolev, S., and Waksman, G. (1998) *EMBO J.* **17**, 7514–7525
[e] Doublié, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature (London)* **391**, 251–258
[f] Sousa, R., Chung, Y. J., Rose, J. P., and Wang, B.-C. (1993) *Nature (London)* **364**, 593–599
[g] Sastry, S., and Ross, B. M. (1999) *Biochemistry* **38**, 4972–4981
[h] Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H., and Steitz, T. A. (1996) *Biochemistry* **35**, 8110–8119
[i] Jing, D. H., Dong, F., Latham, G. J., and von Hippel, P. H. (1999) *J. Biol. Chem.* **274**, 27287–27298
[j] Cai, H., Yu, H., McEntee, K., Kunkel, T. A., and Goodman, M. F. (1995) *J. Biol. Chem.* **270**, 15327–15335
[k] Morris, M. C., Berducou, C., Mery, J., Heitz, F., and Divita, G. (1999) *Biochemistry* **38**, 15097–15103
[l] Rodgers, D. W., Gamblin, S. J., Harris, B. A., Ray, S., Culp, J. S., Hellmig, B., Woolf, D. J., Debouck, C., and Harrison, S. C. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1222–1226
[m] Lundblad, V. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8415–8416
[n] Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H., and Kraut, J. (1994) *Science* **264**, 1930–1935
[o] Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J., and Pelletier, H. (1997) *Biochemistry* **36**, 11205–11215
[p] Kornberg, A., and Baker, T. A. (1992) *DNA Replication*, 2nd ed., Freeman, New York

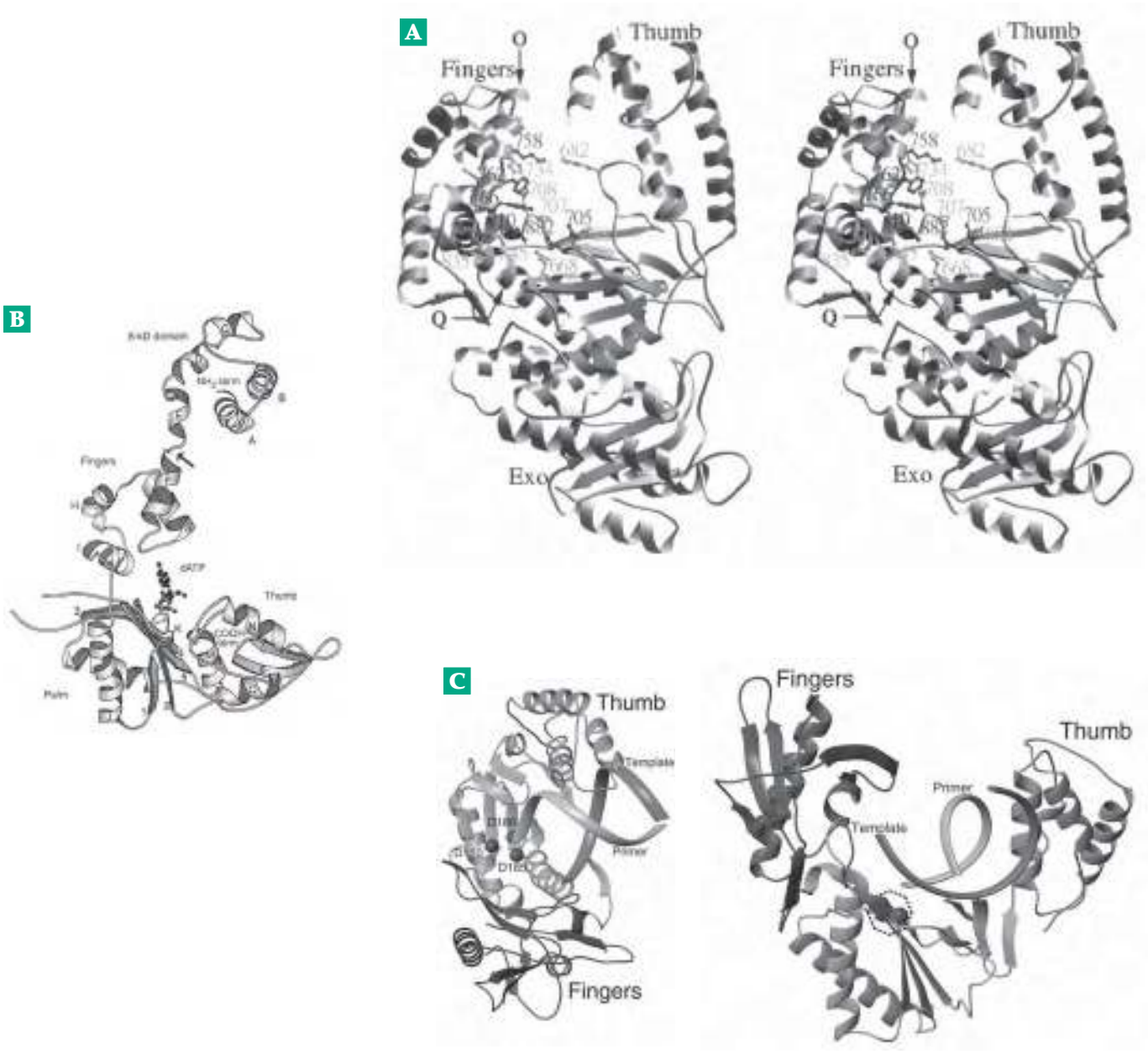


Figure 27-12 (A) Stereoscopic MolScript ribbon drawing showing three-dimensional structure of the 605-residue Klenow fragment of *E. coli* DNA polymerase I and illustrating the characteristic "right hand" structure of DNA polymerases. Fingers and thumb domains are marked. The polymerase active site is between these in the palm region and involves essential catalytic residues Asp 705, Asp 882, and Lys 758. The 3'-5' exonuclease domain is in the foreground. From Astatke *et al*.[271] Courtesy of Catherine M. Joyce. (B) Ribbon drawing of rat DNA polymerase β with dATP in the active site. From Sawaya *et al*.[273] Courtesy of Michael R. Sawaya. (C) Ribbon drawings of two views of the HIV reverse transcriptase with a DNA fragment in the active site. The 3' end of the primer is adjacent to the catalytic triad Asp 110, Asp 185, and Asp 186 (dark spheres). From Singh and Modak.[272] Courtesy of Kamlendra Singh. (D) Ribbon drawing of PCNA, the eukaryotic processivity factor or "sliding clamp" (see pp. 1549, 1562). Courtesy of John Kuriyan.[272a]

hydrolysis from the 5' end.

Treatment with proteolytic enzymes cuts a 323-residue piece containing the 5'-3' exonuclease from DNA polymerase I leaving a larger C-terminal piece known as the **Klenow fragment**. This fragment retains the polymerase activity as well as the 3'-5' exonuclease activity and is widely used in genetic engineering. Its three-dimensional structure[271,275] is shown in Fig. 27-12A. In the Klenow fragment the large C-terminal domain contains the polymerase. The N-terminal domain contains the 3', 5'-exonuclease activity, which is thought to fulfill a **proofreading** and **editing** function.[275a] The polymerase acts at the 3' end of the growing DNA chain. Before moving on to the next position, the enzyme verifies that the correct base pair has been formed in the preceding polymerization event. If it has not, the exonuclease action removes the incorrect nucleotide and allows the polymerase to add the correct one. Thus, each base pair is checked twice, first before polymerization and then after polymerization.

A puzzle was the fact that structural studies indicate that the editing center is over 3 nm away from the catalytic center in Pol I.[267] The proposed solution to the puzzle is illustrated in Fig. 27-14. When the catalytic center "identifies" a nucleotide triphosphate as able to form a proper Watson–Crick nucleotide pair, it catalyzes the formation of the new nucleotide linkage. Then it releases the 3'-chain end, which sometimes "melts" and over a 10- to 100-ms time interval is able to reach over into the exonuclease site where the newly added nucleotide may be hydrolytically removed. However, if the newly formed nucleotide is properly paired, it will be less likely to melt, and the new nucleotide will be more likely to be retained.[259a,265,267] DNA polymerase I and other related polymerases utilize **processive mechanisms**, moving from one site to the next without diffusing away from the DNA. The schematic picture shown in Fig. 27-14 also indicates how the 5'-3' exonuclease activity can come into play, when the polymerase reaches the end of a single-stranded gap.

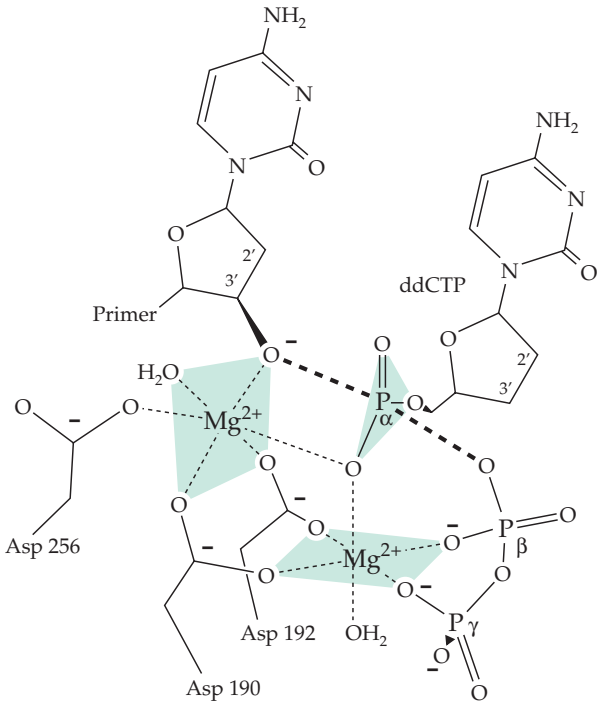


Figure 27-13 Proposed mechanism and transition state structure for the synthetic nucleotidyltransfer activity of DNA polymerase β (and other DNA polymerases). The chain-terminating inhibitor dideoxy CTP is reacting with the 3'-OH group of a growing polynucleotide primer chain. This –OH group (as $-O^-$) makes an in-line nucleophilic attack on P_α of the dideoxy-CTP. Notice the two metal ions, which interact with the phospho groups and which are held by three aspartate side chains. Two of the latter, Asp 190 and Asp 256, are present in similar positions in all of the polymerases. The active centers for the hydrolytic 3'-5' and 5'-3' exonuclease activities of some of the polymerases also appear to involve two-metal catalysis and in-line displacement. See Sawaya *et al.*[274]

Other Class A polymerases. The *Thermus aquaticus* (***Taq***) polymerase is best known for its widespread use in the polymerase chain reaction (PCR; Fig. 5-47). Like *E. coli* I the enzyme is a large multidomain protein. The structure of the catalytic domains of the two enzymes are nearly identical, but the *Taq* polymerase has poor 3'-5' editing activity.[276] The enzyme has been carefully engineered to improve its characteristics for use in the PCR reaction.[277]

Some bacteriophage encode their own DNA polymerases. However, they usually rely on the host cell to provide accessory proteins. The sequence of the DNA polymerase from phage T7 is closely homologous to that of the Klenow fragment and the 3D structures are similar. The 80-kDa T7 polymerase requires the 12-kDa thioredoxin from the host cell as an additional subunit. It has been genetically engineered to improve its usefulness in DNA sequencing.[278]

About 45% of the sequence of the **RNA polymerase** encoded by phage T7, which transcribes RNA from the phage DNA, is also similar to that of the Klenow fragment. Sequences of these DNA polymerases are distantly related to those of reverse transcriptases.[279,280] The 136-kDa polymerase γ functions in mitochondria but is encoded in a nuclear gene. It is the only DNA polymerase that is inhibited by antiviral nucleotide analogs such as AZT (Box 28-C).[280a,b]

Polymerases of Class B. Although *E. coli* polymerase II is a member of this family, relatively little is known about its function. It may participate in DNA repair in the "SOS" response (Section E).[265]

The catalytic subunits of the major *eukaryotic* DNA

polymerases as well as of archaeal DNA polymerases are members of the B family.[267,281–282a] Eukaryotic cells contain at least 13 DNA polymerases which are designated by Greek letters (Table 27-1). Polymerases α, δ, and ε are essential for nuclear DNA synthesis.[283–286b] They function together with accessory proteins in primase and replisome complexes considered in Section 8. Others participate in DNA repair (pp. 1583, 1584).

Phage T4 encodes a DNA polymerase much used in the laboratory because of its ability to polymerize using a long single-stranded template. It also depends upon accessory factors provided by the bacterial host (Section 5).[287,287a]

Other DNA polymerases. **Reverse transcriptases** synthesize DNA using an RNA template strand. They are best known for their function in retroviruses (Chapter 28). The HIV reverse transcriptase is a heterodimer of 51- and 66-kDa subunits. The larger subunit contains a **ribonuclease H** domain.[288–289a] The enzyme is a prime target for drugs such as AZT and others.[290,291] A different reverse transcriptase is found in all eukaryotic cells in **telomerase**, an enzyme essential for replication of chromosome ends. Reverse transcriptases have also been found in rare L1 sequences that are functioning retrotransposons (Section D).[292]

The ~335-residue catalytic subunit of **eukaryotic polymerase β** , which has a DNA repair function, is the simplest known DNA polymerase. The active enzyme (Fig. 27-12B) is as small as 38 kDa. It lacks proofreading and is less accurate than the eukaryotic replicative polymerases listed in Table 27-1.[293–293c] Structural analysis revealed a folding pattern (Class X) related to that of a **nucleotidyltransferase superfamily** that includes enzymes such as terminal deoxynucleotidyltransferase (Chapter 12) and the glutamine synthase adenylytransferase (Fig. 24-7).[294] However, its active site (Fig. 27-13) is similar to that of other DNA polymerases.

DNA polymerase III. This Class C enzyme is the major bacterial polymerase for DNA replication. In its complete holoenzyme form it can synthesize new DNA strands at rates as high as ~1 kilobase s^{-1} without dissociation from a template.[294a] A genetic approach has provided important information about DNA replication.[265,266] A series of temperature-sensitive mutants of *E. coli* unable to carry out DNA synthesis were obtained. From these mutations, genes *dnaA*, *F*, *G*, *I*, *J*, *K*, *L*, *P*, *I*, *X*, and *Z* were identified and located at various points on the chromosome map. Genes *C* and *D* map close together at 89 min, and it now appears that they are one gene. Gene *F* encodes a ribonucleotide reductase (Eq. 16-22). The functions of genes *A*–*E*, *G*, *X*, and *Z* are discussed in the following sections. None of these genes code for DNA polymerase I, but gene *dnaE* was identified as that of DNA polymerase III, which is now known to be the major polymerase in bacterial DNA replication.[265,295] To obtain rapid error-

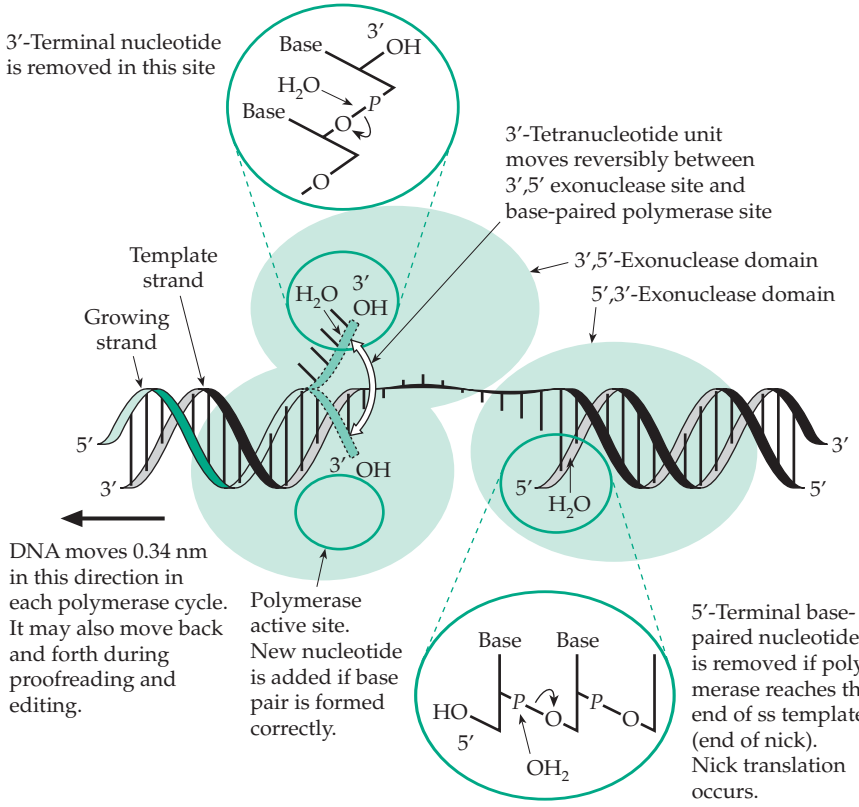


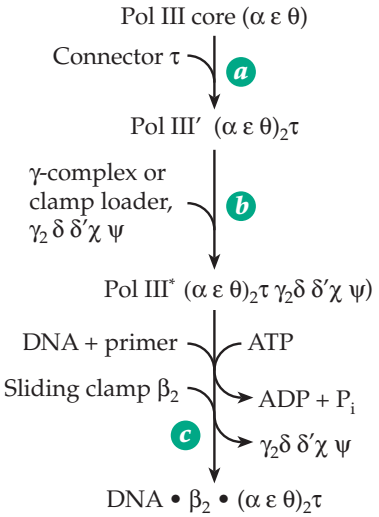
Figure 27-14 Schematic representation of DNA polymerase action on a nicked strand of DNA in which the nick has been enlarged. At the catalytic center new nucleotide units are added at the 3' end of a growing strand. At the 3'-5' exonuclease site the 3' terminal nucleotide may be removed hydrolytically. This will happen to the greatest extent if the nucleotide is poorly paired in the duplex. At the 5'-3' exonuclease site nucleotides are hydrolytically removed from the 5' end of a strand in the chain.[265,267]

free synthesis it must be combined with a number of other subunits to form **polymerase III holoenzyme** (or replisome). Ten different subunits, some as two or more copies, form the holoenzyme.[294a,296–299] Subunits are listed in Table 27-2. The polymerase catalytic center is in the 132-kDa α subunit. The 27.5-kDa ε subunit contains the 3'-5' exonuclease editing activity.[298,300,300a] Mutation in its gene (*dnaQ*) leads to a high spontaneous mutation rate in bacteria. Together with the θ subunit, α and ε form the polymerase III core. This complex has polymerase activity and improved proofreading ability but is still unable to act rapidly, accurately, and processively. Full catalytic activity requires at least the additional β, γ, δ, and τ subunits (Table 27-2). The presence of the τ subunit causes the core complex to dimerize to form Pol III′ (Eq. 27-4, step *a*). Pol III′ can add only about ten nucleotides to a growing DNA chain before it dissociates.[301] The presence of the β_2 dimer, known as the **processivity factor** or **sliding clamp**,[301a,b] is essential for highly processive polymerization. The β protein forms a ring around the duplex DNA and interacts with the polymerase clamping it tightly to the DNA. Putting the β_2 clamp onto the DNA is an ATP-dependent process that first involves binding of the **γ-complex** or **clamp loader** (Eq. 27-4, step *b*), ATP-dependent opening of the β_2 ring, and insertion of the DNA duplex (step *c*).[301a,c] This complex may form a replisome structure that acts simultaneously on the leading and lagging strands (see Fig. 27-19).[302,303] The *Bacillus subtilis* replisome appears to contain two different catalytic (α) subunits, perhaps one for each strand.[303a]

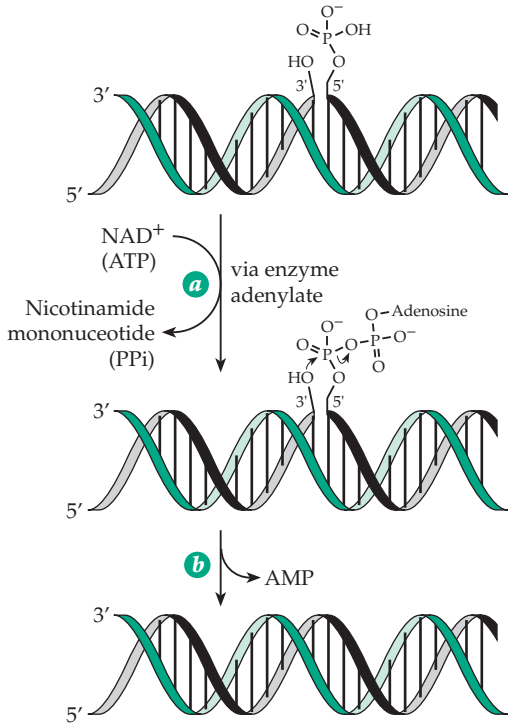
3. Other Replication Proteins

DNA ligases. These enzymes, which are essential to replication, have a specific function of repairing "nicked" DNA.[261,304–307a] Such DNA, as indicated in Eq. 27-5, has a break in one strand and contains a 3'-hydroxyl group and a 5'-phosphate group, which must be rejoined. The ligase from *E. coli* activates the phosphate group in an unusual way by transfer of an adenylyl group from NAD^+, with displacement of nicotinamide mononucleotide (Eq. 27-5, step *a*). The reaction is completed by displacement of AMP as indicated in Eq. 27-5, step *b*. Cells infected by bacteriophage T4 synthesize a virally encoded ligase, which utilizes ATP rather than NAD^+ as the activating reagent. The ~190-kDa mammalian DNA ligase I has been found deficient in some patients with the Bloom syndrome, a condition associated with poor DNA repair and a high incidence of cancer (see also p. 1550).[304]

Single-strand binding proteins. Genetic analysis of replication of the DNA of phage T4 within cells of *E. coli* revealed that at least five genes of the virus are required. One of these, gene 43 specifies the T4 DNA polymerase, while gene 32 codes for a single-strand binding protein, also known as the DNA unwinding, melting, or helix-destabilizing protein. It has a greater affinity for ssDNA than for dsDNA and binds to a length of ssDNA causing unwinding of the double helix and exposure of the purine and pyrimidine bases of the template strand.[308] The protein is required for replication, genetic recombination, and repair of DNA. Similar proteins are encoded in the genomes of many viruses.[309] The 87-residue single-strand binding protein encoded by gene 5 of phage M13 forms a dimer, which completely coats newly synthesized ssDNA preventing the DNA polymerase system of the host



(27-4)



(27-5)

bacteria from converting it into dsDNA. The polynucleotide chain binds into a groove in the protein with one tyrosine intercalated between the DNA bases.[310,311]

The *E. coli* single-strand binding protein, another helix-destabilizing protein that is usually called simply **SSB**, is a tetramer of 18.5-kDa subunits.[265,312,313] It is essential to DNA replication. About 35 nucleotides may bind to each tetramer.[314] The situation is not as clear in eukaryotes where DNA is largely coiled around histones in the nucleosomes. Several single-strand binding proteins have been identified,[315] but the need for SSB proteins in eukaryotic nuclear replication is uncertain.[316] A human mitochondrial SSB resembles that of *E. coli*.[317]

Helicases and topoisomerases. Cells of *E. coli* contain at least 12 DNA-dependent ATPases that cause unwinding of DNA at the expense of hydrolysis of ATP.[318] The activity of these **helicases** is essential to replication, repair, and recombination of DNA in all organisms.[319–320b] The primary replicative helicase of *E. coli*, which unwinds DNA ahead of the replication fork, is encoded by gene *dnaB*. The active form is a hexamer, which exists in at least two conformational states.[321,321b] A segment of ~20 nucleotide units of DNA binds to one hexamer.[322] Helicases have ATPase activity, and the dnaB hexamer contains six ATP-binding sites. However, only three of them may be occupied.[323,324] Many helicases have a hexameric ringlike structure; that of *Bacillus subtilis* is seen clearly in electron micrographs.[325] Three-dimensional structures are known for some.[321a,325a] Although these enzymes may bind to duplex DNA, they also bind to and move along single-stranded DNA in either the 5'→3' or 3'→5' direction. The directionality of a helicase can be determined by annealing two small pieces of ssDNA to the 5' and 3' ends of a longer strand of ssDNA. A 5'→3' helicase will translocate along the long ssDNA and displace the oligonucleotide annealed at the 3' end of the strand, while a 3'→5' helicase will displace the oligonucleotide annealed at the 5' end.[265]

The dnaB protein is a 5'→3' helicase. However, the first helicase identified, the product of the *E. coli rep* gene, is a 3'→5' helicase. It is one of the host proteins needed for the propagation of phages such as ϕX174 and M13. It catalyzes the unwinding of the double-stranded replication forms of these viruses. It binds to a stretch of ~20 nucleotides in a single-stranded region of nicked DNA. The hydrolysis of ATP moves the enzyme along the bound strand in a 3' to 5' direction opening up the DNA at a replication fork. Another *E. coli* 3'→5' helicase is protein *priA* (also called n′), a component of some primasome structures involved in replication of viruses.[318]

The bacteriophage T4 gene 41 protein, a 5'→3' helicase, functions together with the gene 61 primase in replication of that virus.[326–326b] The phage T7 gene 4 protein and virus SV40 large T antigen are also hexameric ringlike helicases. The *E. coli* protein **RecQ** is required for various aspects of recombination and is the prototype of a large group of helicases present in both prokaryotes and eukaryotes.[326c] The bacterial ruvB protein is a hexameric helicase that propels branch migration in Holliday junctions (Fig. 5-28) during genetic recombination,[327–327b] while helicase rho is required for termination of RNA synthesis.[328] Numerous eukaryotic helicases have been identified and purified.[329–331] Helicase DNA2 is needed for DNA replication in nuclear extracts from yeast.[332] Human 3'→5' DNA helicases, members of the RecQ family, are defective in some patients with **Bloom syndrome** and **Werner Syndrome**[332a] (Box 27-A). Bloom syndrome causes growth retardation, immunodeficiency, sensitivity to sunlight, and a predisposition to skin cancers and leukemias.[331a,b] The yeast (*S. cerevisiae*) genome contains genes for 134 different proteins that are probably helicases.[332b] RNA helicases are also known.[333]

A characteristic of all helicases is their ATPase activity, which apparently provides energy for "melting" the DNA. The mechanisms are not clear, but rapid separation of the stacked and hydrogen-bonded base pairs may be impossible without some assistance from an ATP-dependent process. In the case of the rep protein hydrolysis of two molecules of ATP seems to be required to melt one base pair.[323] It isn't clear whether one strand of DNA passes through the center of the oligomeric ring, as shown in Fig. 27-15, or whether both strands pass through. Helicases vary in their amino acid sequences, but they all possess several characteristic **signature sequences** including the Walker A and B motifs found in other ATPases, in

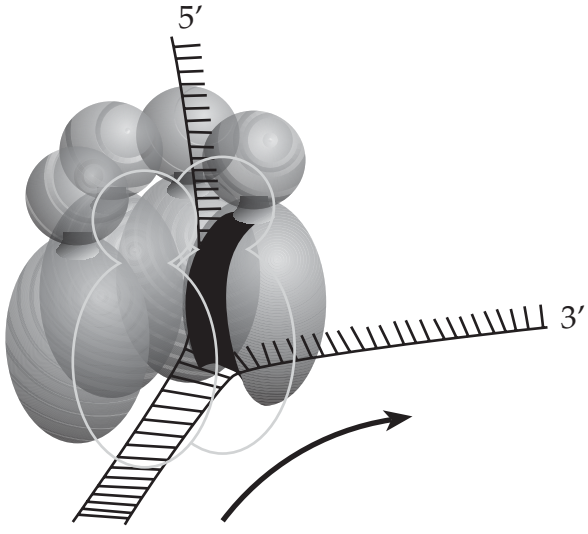
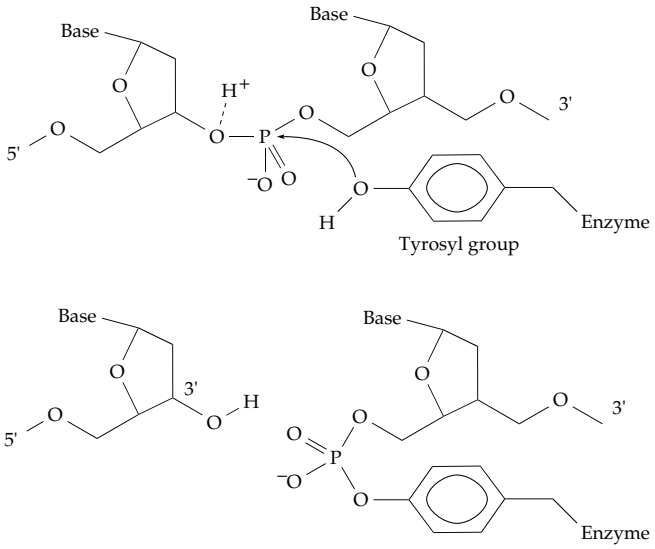


Figure 27-15 Model of DnaB hexamer proposed by Jezewska *et al*.[322] The arrow indicates the direction of movement of the DNA relative to the position of the helicase.

sequences related to the *E. coli* **recA** proteins (Section D) and in synthases (Chapter 18).[328,334]

While helicases can cause the unwinding of linear DNA duplexes, they cannot alter the linking number of covalently closed circular double-stranded DNA. The latter is the function of **topoisomerases**, which have been found in organisms from bacteria to higher plants and animals and may also be encoded by viruses.[265,335–337] There are two basic types of DNA topoisomerase. Those of type I change the linking number in steps of 1. One way that this might be accomplished is for an enzyme to nick one strand in the DNA allowing one of the cut ends to swivel around the unbroken strand, then to reseal the chain. However, it was found that topoisomerases can also cause catenation or decatenation of circular duplex DNA as long as at least one of the reacting DNA molecules is nicked. This observation suggested that a topoisomerase binds to a single-stranded region at a nick and cuts the chain but does not release the ends. This permits either a single strand or a duplex to pass through the broken strand, which is then resealed.

Topoisomerases of type I usually act most rapidly on negatively supercoiled DNA. They relax it by decreasing the number of negatively supercoiled turns one at a time. Negative supercoiling presumably facilitates binding of the enzyme to a single-stranded region by unwinding of the duplex. No ATP or other obvious source of energy is needed by type I topoisomerases. The chain cleavage involves a simple nucleophilic displacement by an –OH group of a tyrosine side chain (Tyr 319 of *E. coli* topoisomerase I),[338] which attacks a phosphorus atom in the DNA chain (Eq. 27-6). The result is covalent attachment of the enzyme to the 5'-end of the cut strand in type IA topoisomerases and at the 3'-end in type IB topoisomerases.[336,337,339] After the passage of the other



(27-6)

TABLE 27-2
Some Proteins of DNA Replication in *E. coli*[a]

Name	Gene	Mass of monomer (kDa)	Map location (min) Fig. 26-4
Polymerase I (also has 3'-5' and 5'-3' exonuclease, and RNase activities)	*polA*	103	87
Polymerase III			
Core			
α subunit (polymerase)	*dnaE*	130	4
ε subunit (3'-5' exonuclease)	*dnaQ*	27.5	5
θ subunit	*hol E*	10	
Sliding clamp, β_2	*dnaN*	37 × 2	83
Connector τ (ATPase)	*dnaX*	71	
Clamp loader (γ complex)			
γ subunit	*dnaX*	47.5 × 2	11
σ subunit		35	
σ′ subunit		33	
χ subunit		15	
ψ subunit		12	
DNA binding proteins			
Single-strand, SSB	*ssb*	18.5 × 4	92
Double-strand, HU			
α subunit		9.5 × 2	
β subunit		9.5 × 2	
Helicases (ATP-dependent)			
Primary replicative	*dnaB*	52 × 6	92
Dna C protein	*dnaC*		99
PriA (n′), primosome	*priA*	76	88
Rep	*rep*	76.4	85
Initiation and priming proteins			
Dna A protein	*dnaA*	52	83
Primase (an RNA polymerase)	*dnaG*	60	67
PriB (n) primosome	*priB*		96
PriC (n″) primosome	*priC*		
DnaT (primosome assembly)	*dnaT*		99
Ribonuclease HI	*rnhA*		
DNA ligase	*lig*	75	52
Topoisomerases			
Type I	*topA*		28
Type II, DNA gyrase ($\alpha_2\beta_2$)			
Subunit α	*gyrA*		97
Subunit β	*gyrB*		90

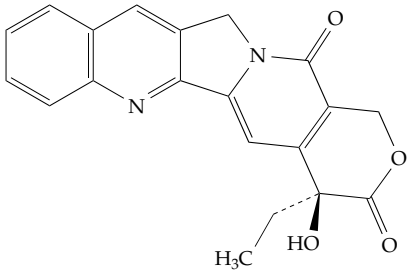
[a] In large part from Kornberg, A., and Baker, T. A. (1992) *DNA Replication*, 2nd ed., Freeman, New York

strand through the gap in the cut strand of a type IB topoisomerase (Fig. 27-16) the free 3'-OH oxygen atom (or 5'-OH) attacks the phosphorus atom in the phosphotyrosine diester to reform the chain and release the enzyme. X-ray diffraction studies show that both the Type IA *E. coli* topoisomerase I and the human type IB topoisomerase[339–341a] are large proteins with holes of appropriate diameter for a DNA double helix. As illustrated in Fig. 27-16B, the protein may open to allow a double helix to enter and occupy a suitable position for cleavage of one chain by the active-site Tyr 319. Topoisomerases are metalloenzymes, usually functioning best with Mg^{2+}. *E. coli* topoisomerase I also contains 3–4 tightly bound Zn^{2+} ions.

Topoisomerases of type II change the linking number by 2 in either the positive or negative direction and hydrolyze ATP in the process. The best known of these is the *E. coli* **DNA gyrase**, an $\alpha_2\beta_2$ dimer of 97-kDa (α) and 90-kDa (β) subunits.[343,344] The enzyme catalyzes the ATP-dependent introduction of negative supercoils into DNA. It also relaxes negatively supercoiled DNA slowly in the absence of ATP. Type II topoisomerases are found in all organisms.[335,345] They are encoded by some bacteriophage such as T4[346] and by plasmids.[347] However, most differ from bacterial DNA gyrase in not coupling DNA supercoiling to ATP hydrolysis. They require ATP but like topoisomerase I cause a relaxation of the supercoiling. Strands of one segment of DNA (called the "gate" or **G-segment**) are cleaved by the enzyme with staggered cuts four base pairs apart. Another segment of DNA (the "transport" or **T-segment**) is then passed through the gate and is thought to be released from a second gate in the complex (Fig. 27-16C).[342,348] The enzyme subunits bind through phosphotyrosine linkages as in Eq. 27-5 to the 5' phospho groups of the two cleaved chains, while the subunits bind ATP and may like tiny muscles twist the DNA.[349] Topoisomerases II are large dimeric proteins. The subunits of yeast topoisomerase II (Fig. 27-16C) are 1200-residue multidomain proteins.[342,350] Mechanisms of DNA cleavage by types I and II topoisomerases appear to be related.[351] However, the ATP-dependent conformational changes involved in a two-gate mechanism are unique to topoisomerases II.[348,348a] The bacterium *Sulfolobus* contains a type I topoisomerase that is called **reverse gyrase** because it utilizes ATP to introduce *positive* supercoils into DNA.[352] This is in contrast to gyrase, a type II topoisomerase that introduces negative supercoils.

Type II topoisomerases are essential and function in replication, DNA repair, transcription, and chromosome segregation at mitosis.[345,349] Yeast with a *top2* mutation dies during mitosis with hopelessly entangled daughter chromosomes.[353] A fluorescent antibody to eukaryotic topoisomerase II binds to chromosomes, probably at the bases of the radial loops present during mitosis,[353] and also to centrosomes[353a] Higher organisms contain more than one topoisomerase II.[354] Their specific functions are uncertain, but one appears to be to unknot entangled chromosomal DNA. In the crowded conditions of a cell nucleus topoisomerase can also cause inadvertent *formation* of knots.[355]

The functional role of topoisomerases of type I is less clear. Staining with fluorescent antibodies to the enzyme has revealed its presence in the transcriptionally active "puffs" of polytene chromosomes (p. 1635)[356] and in centromeres of mitotic cells.[357] A current hypothesis is that in *E. coli* class I topoisomerases act to relax negatively supercoiled strands of DNA behind transcription complexes, while gyrase acts to generate superhelical twists, which favor opening of the duplex ahead of transcription complexes.[354,358] Transcription of a supercoiled rRNA gene *in vitro* is diminished by the selective topoisomerase I inhibitor **camptothecin**, one of a group of antitumor drugs directed against topoisomerase of both types I and II.[349,354,359]



Camptothecin, a topoisomerase I inhibitor

In the autoimmune diseases scleroderma and systematic lupus erythematosus antigens to nuclear proteins or nucleic acids are present in the blood. Many patients with severe scleroderma have an antibody against topoisomerase I.[360]

Primases, initiator proteins, and ribonucleases. The priming segment needed for initiation of DNA replication is either a short segment of RNA or an oligonucleotide containing a mixture of ribonucleotide and deoxyribonucleotide units. The enzyme forming the primer is an RNA polymerase called **primase**. In *E. coli* it is encoded by gene *dna*G.[361–362a] Under some circumstances other RNA polymerases can act as the primase. Bacteriophages and plasmids may also encode primases (Table 27-1). For example, gene 61 of phage T4 (Fig. 26-2) encodes a primase, which together with the T4 helicase forms the priming particle or **primosome**.[363] The phage T7 gene 4 encodes a 63-kDa mutlifunctional protein that is both a primase and a helicase.[364–366b] The primase active site is on the outside of the hexameric ring. Additional proteins repre-

senting products of genes *dna*A and C are also required for initiation of replication in *E. coli*. Several molecules of the dnaA **initiator protein** bind to a specific DNA **origin** sequence and participate in assembling a primosome that also contains the hexameric dnaB helicase and, transiently, protein dnaC.[367–370] Replication of some single-stranded phages, such as ϕX174, in *E. coli* also require the *E. coli* priA, priB, and priC proteins (Table 27-2).[265,371,372] For successful completion of replication chaperonins the products of genes *dna*J and *grp* E are needed[373] as is a ribonuclease that digests the primer segments after they have been used as replication initiators.[374–376]

4. Replication of Bacterial DNA

The basic mechanisms of replication implied in Eq. 27-3 seems to be universal, but several questions had to be asked. "Is replication initiated at a fixed point or points in a chromosome?" and "Does replication occur in one direction only or do two forks form at the point

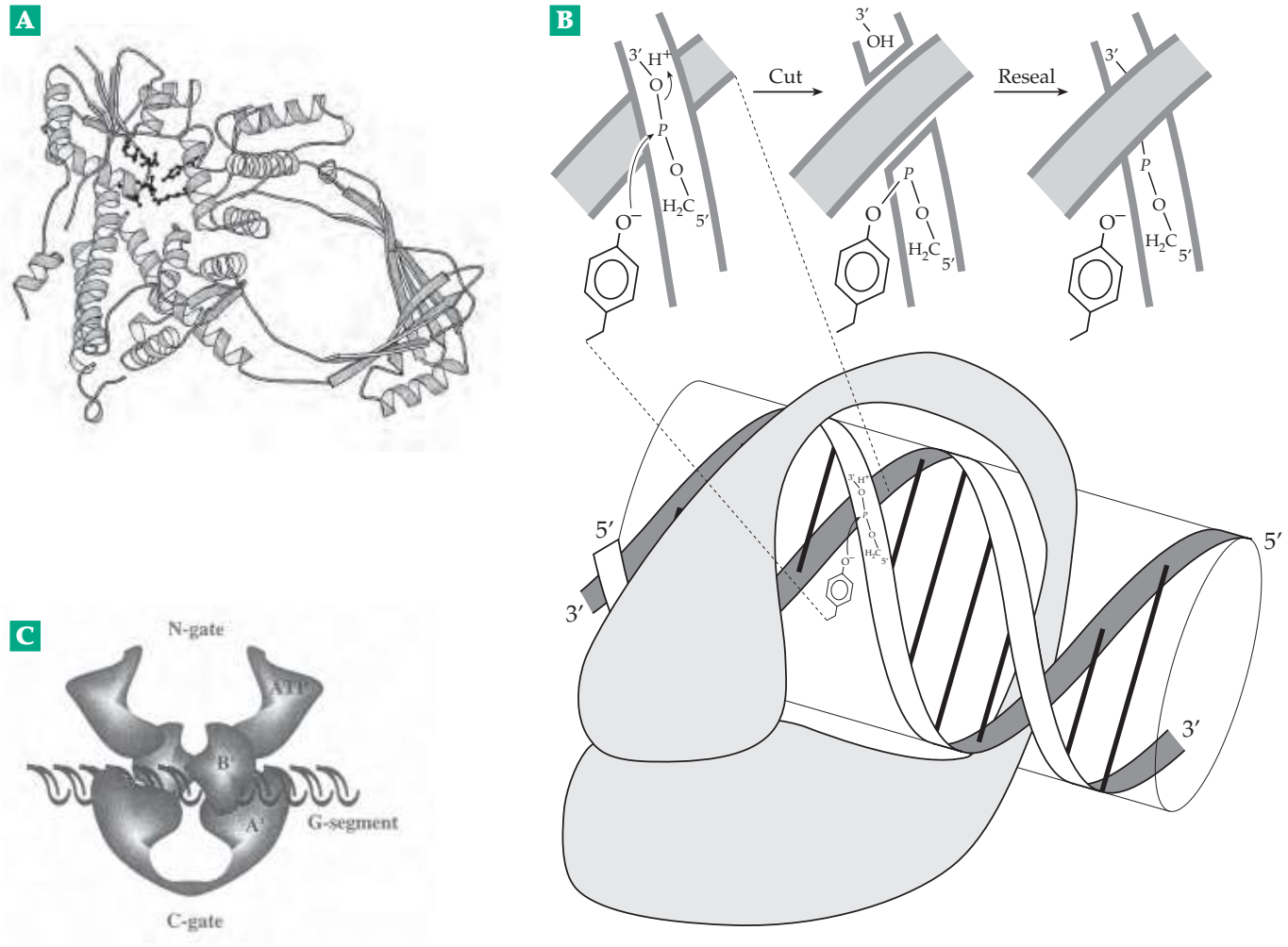


Figure 27-16 (A) Ribbon drawing of a large 67-kDa fragment of the 97-kDa (864-residue) *E. coli* topoisomerase I showing the position of the active-site tyrosine 319 and an adjacent arginine. From Lima *et al.*[341] Courtesy of Alfonso Mondragon. (B) Schematic diagram indicating a way in which topoisomerases of type 1 may pass one strand of DNA through another. The protein is shown binding to a single strand of a DNA duplex. This binding is facilitated by negative supercoiling. The enzyme then cuts the same strand by means of a nucleophilic displacement on a phosphorus atom using a tyrosinate side chain. The other cut end is held noncovalently by the enzyme, while the second strand passes through the gap. Then the gap is resealed by a reversal of the cleavage reaction. From Lima *et al.*[341] (C) Schematic model of a type II topoisomerase bound to a G-segment of DNA. This double helix is cut and another double helical strand, the T-strand, enters the N-gate. The gate then closes, and the central gate opens to allow the T-strand to pass through and exit through the C-gate. The shapes of the three domains are based on crystallographic data for the yeast enzyme. The ATPase, B', and A' domains consist of residues 1 to 409, 410 to 660, and 660 to 1200, approximately. From Olland and Wang.[342] Courtesy of James C. Wang. See also Champoux.[340c]

of origin and travel in opposite directions?" To answer these questions both genetic methods and electron microscopy have been employed.

Directions of replication. One technique for establishing the direction of replication in *E. coli* was to insert a λ prophage at the *att* site (Fig. 26-4, 17 min) and DNA from phage Mu-1 at a variety of other sites around the chromosome.[377] Phage Mu-1 was especially useful because it can be integrated at many different locations within the well-mapped genome. Integration within a gene inactivates that gene (an addition mutation) and allows the localization of the Mu prophage. Bacteria were prepared containing both λ and Mu-1 prophage, the latter at various sites. The bacteria were also auxotrophic for certain amino acids. Because of this, replication could be stopped by amino acid starvation. The bacteria usually completed any replication cycle in progress and then stopped. When the missing amino acids were added, replication began, again starting from the replication origin. Bromouracil, which enters DNA in place of thymine, was added at the same time. Consequently, the newly synthesized DNA strands were denser than the parent strands (see Fig. 27-11). After various times of replication the newly formed strands were separated by centrifugation in a CsCl gradient and were tested for hybridization with both λ and Mu-1 DNA. Since the cells did not all begin replication at the same time after addition of amino acids, a variety of lengths of newly replicated DNA were present. Nevertheless, from the observed ratios of Mu-1 DNA to λ DNA for the various strains it was possible to map the progress of replication beginning at an **origin** *ori*C near gene *ilv* at 74 min (Fig. 26-4). Replication was found to progress bidirectionally around the chromosome and to terminate between genes *trp* and *his* at ~25 min.

The use of autoradiographic methods confirmed bidirectional replication. Strains of amino acid auxotrophs with small nucleoside triphosphate pools were used. The addition of amino acids after starvation led to initiation of replication with only a 6-min lag. The cells were labeled with [^{3}H]thymidine, and after the replication forks had moved a short distance from the origin of replication the cells were given a pulse of "super-hot" [^{3}H]thymidine. Using autoradiography it was possible to observe the clearly bidirectional replication forks[378] (Fig. 27-17). Replication in other bacteria is also bidirectional.

Origins of replication. Replication of the *E. coli* chromosome begins and proceeds bidirectionally from its defined origin *ori*C. Replication of linear phage T7 is also bidirectional and begins at a point 17% of the way from one end.[265] In mammalian mitochondrial DNA the origin of replication for the H-strand is in the D-loop but that for the L-strand is 2 / 3 of the way around the circular chromosome within a cluster of tRNA genes (Fig. 18-3).[381,382] The single-stranded circular DNA genomes of Ff, ϕX174, and G4 phage also have distinct origins for initiation of replication to give RF circles.[265]

Most origins have quasi-palindromic nucleotide sequences, perhaps so that DNA can be looped out from the main duplex as is shown in Fig. 27-18A and B. The lengths of *ori* sequences vary, as does the complexity of their possible folding patterns. Plasmids have been constructed which not only contain the *E. coli* origin, but are dependent upon that origin for their own replication.[382a] Study of those plasmids indicate

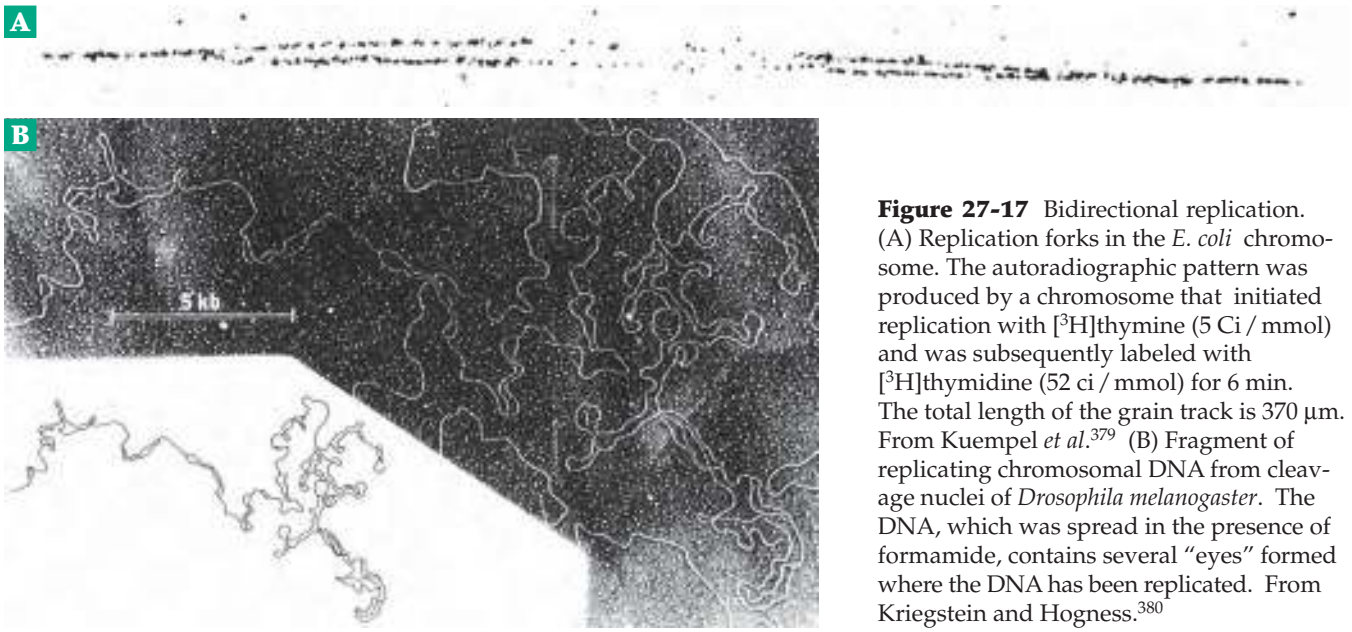


Figure 27-17 Bidirectional replication. (A) Replication forks in the *E. coli* chromosome. The autoradiographic pattern was produced by a chromosome that initiated replication with [^{3}H]thymine (5 Ci / mmol) and was subsequently labeled with [^{3}H]thymidine (52 ci / mmol) for 6 min. The total length of the grain track is 370 μm. From Kuempel *et al.*[379] (B) Fragment of replicating chromosomal DNA from cleavage nuclei of *Drosophila melanogaster*. The DNA, which was spread in the presence of formamide, contains several "eyes" formed where the DNA has been replicated. From Kriegstein and Hogness.[380]

that a 245-bp *ori*C sequence is essential.[265,383,383a] This sequence, which is shown in Fig. 27-18, contains several repeated oligonucleotides including 11 GATC sequences, which are sites of adenine N^6-methylation (see Section B, 6), and four "9-mers" (commonly known as dnaA "boxes") with the consensus sequence

$$TTAT\begin{matrix}A\\C\end{matrix}CA\begin{matrix}A\\C\end{matrix}C$$

These appear in both 5' to 3' and 3' to 5' orientations, allowing them to form two base paired "stems." In addition, there are three direct repeats of a 13-residue consensus sequence 5'-GATCTNTTNTTTT (shaded in green in Fig. 27-18), which form an AT rich duplex. Other bacterial replication origins often follow a pattern similar to that in Fig. 27-18.[384] However, the origin for *Mycoplasma genitalium* has been hard to detect.[385] Replication origins of archaea have characteristics similar to those of bacteria and of organelles.[385a]

Priming and initiation of DNA synthesis. The first step in initiating a new round of replication and a new cell cycle in *E. coli* appears to be the binding of the initiator protein, dnaA, to the 9-mers in the negatively supercoiled origin.[295,383,386–388] The resulting complex is visible in an electron microscope and may consist of a core of up to 20–40 molecules of dnaA protein with the DNA wrapped around them (Fig. 27-19). ATP is also required and is bound to the protein where it hydrolyzes slowly. Similar initiator proteins are found

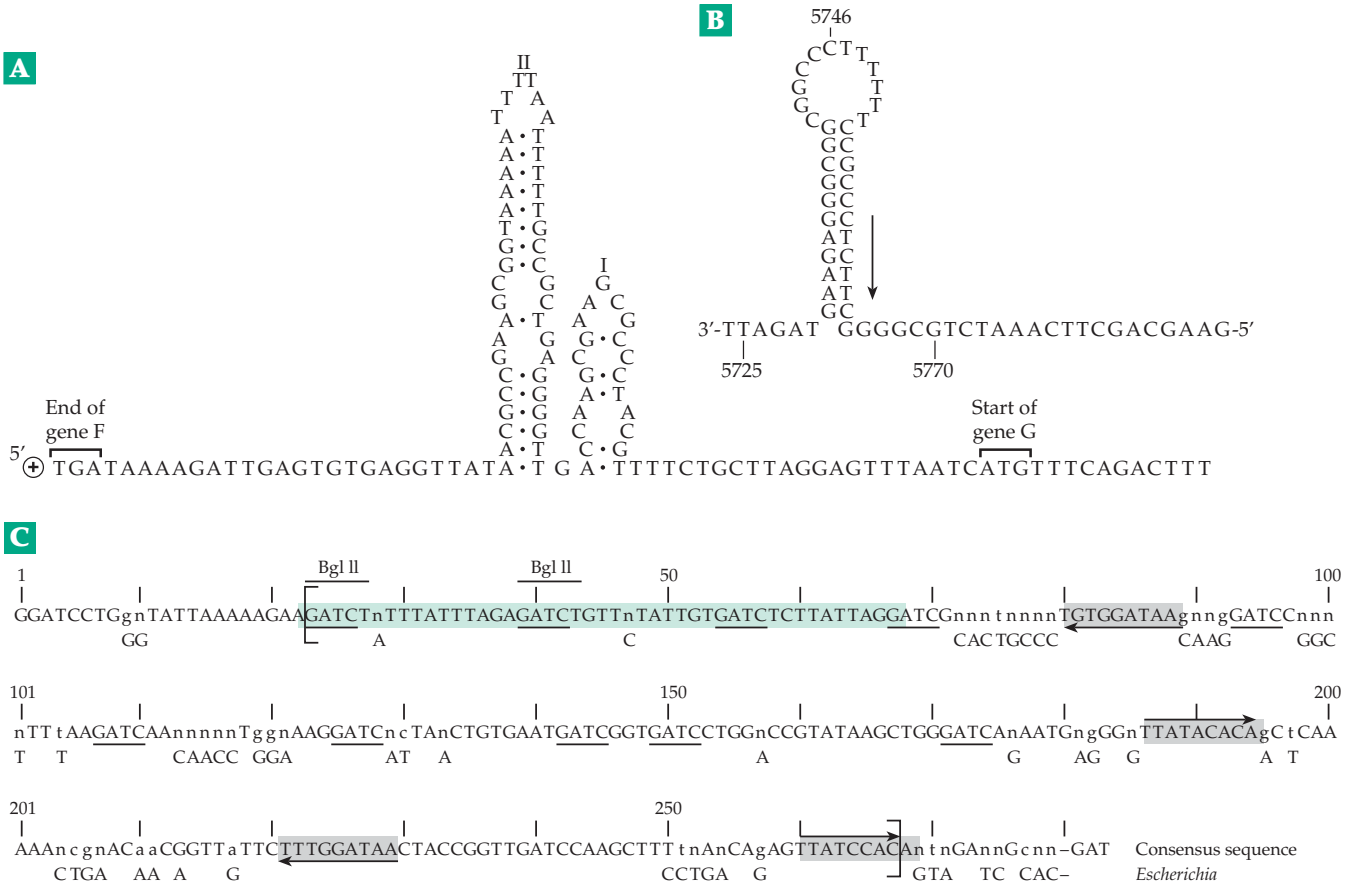


Figure 27-18 Nucleotide sequences and proposed secondary structures for some origins of replication. (A) Bacteriophage ϕX174. The ends of the neighboring genes F and G are shown. From Kornberg and Baker.[265] (B) Human mitochondrial DNA; origin of replication of the L strand. The sequence shown is for the H strand. The arrow indicates the direction of replication. From Hixson *et al*.[381] (C) Bacterial *ori* region. The consensus sequence of origins of replication that will function in *E. coli*. Derived from sequences from *E. coli, Salmonella typhimurium, Enterobacter aerogenes, Klebsiella pneumoniae,* and *Erwinia carotovora.* In the consensus sequence a large capital letter means that the same nucleotide is present in all six origins; a lowercase letter means that the same nucleotide is present in three of the five origins; n means that any of the four nucleotides may be present; – represents a deletion, G being present here in the *Enterobacter* origin. Underlined are *dam* methylation sites. Green shaded sequences, three 13-mers in a region that undergoes easy opening of the duplex for insertion of replication proteins; gray boxes with arrows, four 9-mers that are specific sites for binding of replication protein A. Brackets enclose a 245-bp minimum origin. From Kornberg and Baker.[265] Data from Zyskind *et al*.[390]

in many bacteria and are encoded by some viruses and plasmids. In archaea, as well as in humans, the initiator proteins seem to combine functions of the *E. coli* dnaA protein with those of SSB.[391,391a]

Following the binding of the dnaA protein the hexameric helicase dnaB (Fig. 27-15) is loaded onto the adjacent DNA in the region of the 13-residue repeated sequences, which are labeled 1, 2, and 3 in Fig. 27-19. As shown in this figure introduction of the helicase is assisted by protein dnaC, which forms a complex with the helicase. Additional binding of protein HU and a temperature of >30°C are essential for tight binding of the proteins. Although the dnaC protein is needed for formation of the prepriming complex, it dissociates after the dnaB protein becomes firmly bound. One dnaB hexamer binds to each single strand of the DNA duplex with opposite orientations (Fig. 27-19).[383a] With the dnaB helicase in place on each strand, this ATP-driven enzyme processes along the DNA, unwinding the duplex in both di-rections. (Perhaps it may be more accurate to say that the *DNA moves through the helicase*.) The resulting "bubble" is held open by binding of SSB tetramers. The primase (dnaG protein) then binds adjacent to the helicase and synthesizes the RNA primer along each strand of DNA. As the helicase processes to the right in the fork shown in Fig. 27-19, the complex with the dnaA protein dissociates, permitting primer synthesis into the origin region. Alternatively, RNA polymerase (Chapter 28) can prime replication by initiating transcription on both strands of the DNA. Suitable promoters are present and oriented in opposite directions on the two strands.[392]

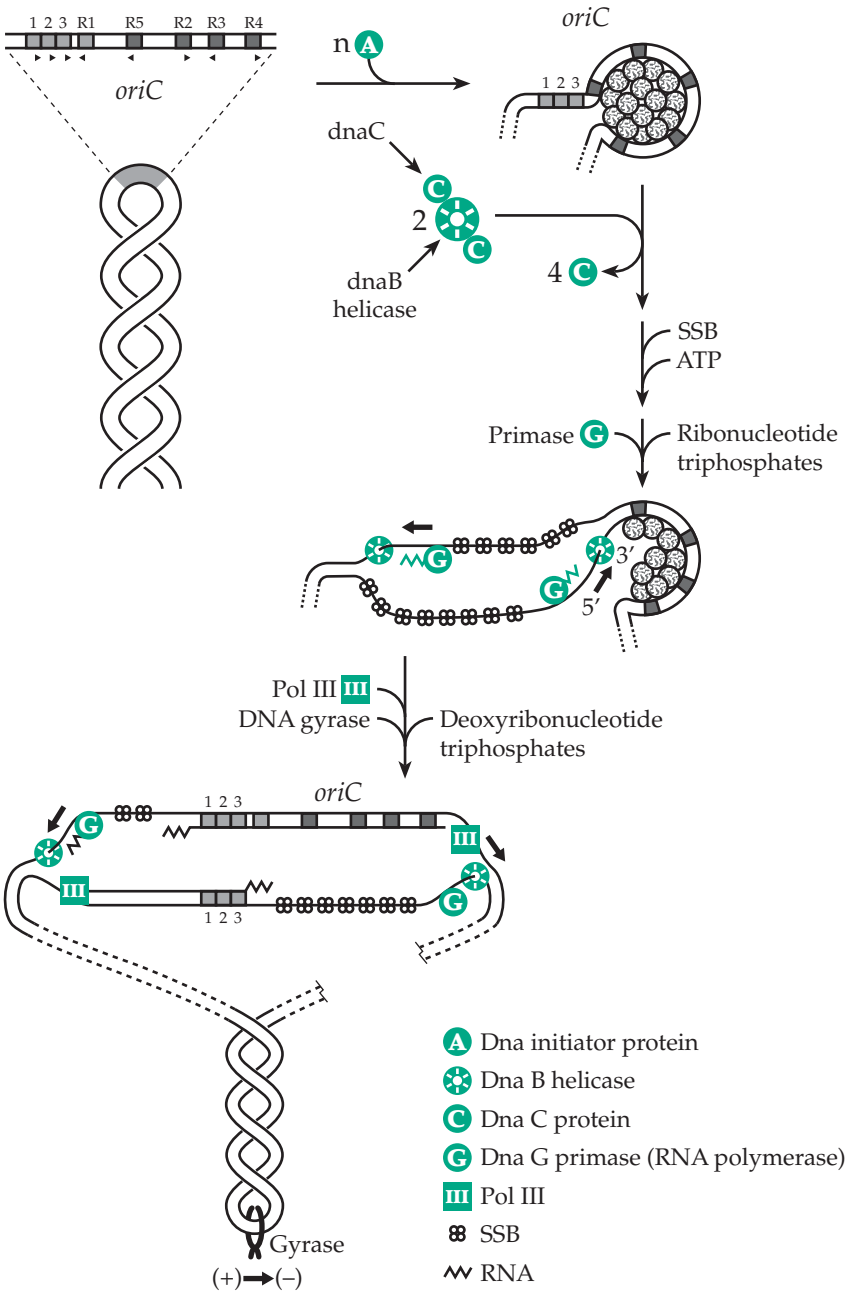


Figure 27-19 Hypothetical scheme for initiation of bidirectional replication in *E. coli*. The closed boxes R1 through R5 represent the 9-residue recognition sequences for the *E. coli* dnaA protein. The open boxes 1, 2, and 3 represent the three 13-residue repeats, possible sites for binding of the dnaB-dnaC protein complex. From McMacken *et al.*[383] Redrawn in simplified form.

It has long been postulated that the bacterial chromosome is attached to the plasma membrane. At least one such attachment site may be at or near the origin of replication.[393] Furthermore, the exchange of ADP for ATP in the dnaA protein is catalyzed specifically by cardiolipin and phosphatidylglycerol containing the unsaturated oleic acid.[386,393] Inositol polyphosphates may also play a role.[394]

Elongation of DNA chains.

DNA polymerase III in its holoenzyme form is the major polymerase for DNA replication. It elongates the primer chains rapidly and processively leaving only very small gaps at the ends of single-stranded regions. The rate of elongation, which is ~3 nucleotides / s for 8 kb

oriC plasmids, may be determined by the rate of action of the dnaB helicase. A completely unwound *ori*C plasmid, bound to SSB, undergoes primer elongation ~ 10 times faster.[383] However, the rate for the intact *E. coli* replisome is nearly 1000 nucleotide s^{-1} with a rate of misincorporation of only one in 10^9 nucleotides.[394a]

Small *ori*C plasmids need to be primed at only one location, but the large bacterial chromosome must undergo priming at many sites on the lagging strand to permit DNA polymerase III to act on that strand with formation of the Okazaki fragments. The DNA polymerase complex may be a dimer that works on both strands at once. The lagging strand may be looped out to allow it to lie parallel with the leading strand (Fig. 27-20). The appearance of the electron micrograph in Fig. 27-20B supports this suggestion. However, the manner in which the lagging strand can be shifted to bring the next primed initiation site to the replication complex is not clear. DNA polymerase III holoenzyme itself may be organized as an asymmetric oligomer[395] that operates on both strands in a complex such as is shown in Fig. 27-20. The primase, either the dnaG protein[265,383] or the primosome used by phage ϕX174 (Section 5),[396] may synthesize the RNA primers on the lagging strand.

Termination of replication. As each Okazaki fragment is completed along the template for the lagging strand, the RNA primer piece is digested out, replaced by DNA, and the nick sealed by action of DNA ligase. Ribonuclease H, which is found in both bacteria and eukaryotic cells, specifically degrades the RNA–DNA hybrid regions.[396a] In bacteria another mechanism for primer removal is available. The 5'- 3' exonuclease activity of DNA polymerase I will cut out the RNA segment, while the 5'- 3' polymerase activity of the same enzyme will fill the gap.

Replication of *oriC* plasmids may occur by simply allowing replication of the leading strands at both replication forks to continue all the way around the circle.[397] However, in *E. coli* bidirectional replication continues only until the two replication forks converge. This can occur anywhere between two **terminators**, T1 and T2, located at 28.1 min and 35.6 min. The terminators slow replication in the counterclockwise and clockwise directions, respectively. A gene (*tus*) near T1 encodes a **terminator utilization substance**, a DNA-binding protein that associates with T1 and T2 and causes termination.[398–401] Another problem may be the separating of catenated DNA circles by action of a topoisomerase. Finally, it is essential to **partition** the original chromosome and its replica, one to each daughter cell. This requires at least three other gene products including one large 170-kDa protein.[402]

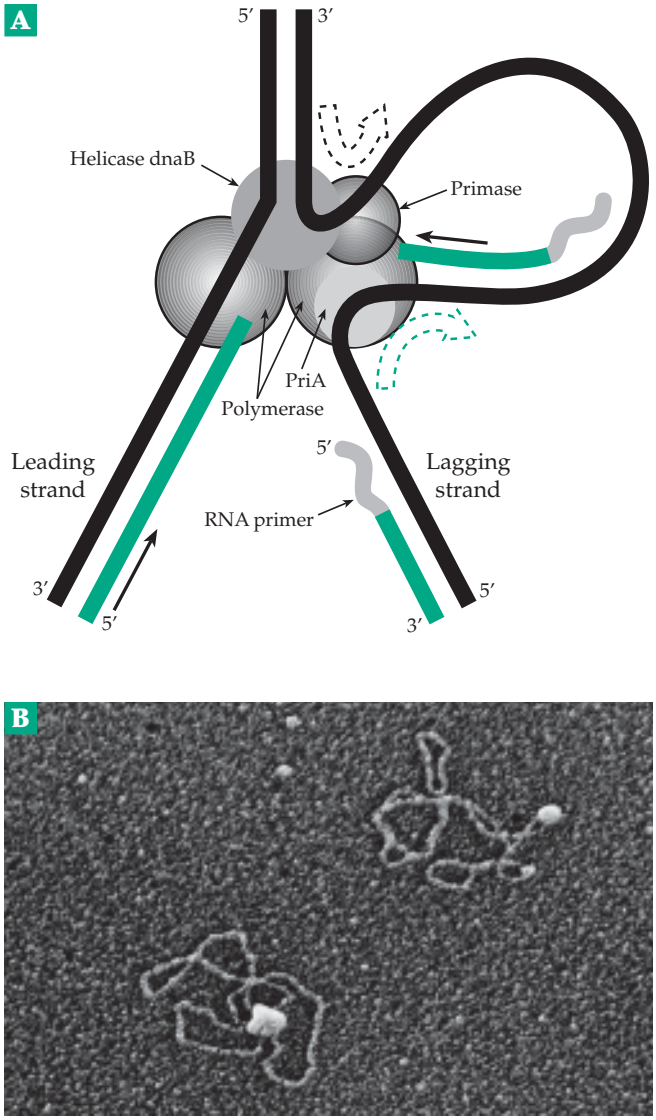


Figure 27-20 (A) Hypothetical replisome for concurrent replication of leading and lagging strands by a dimeric polymerase associated with helicase dnaB and a primosome. Open arrows indicate directions of movement of DNA, which is forming a loop as the polymerase fills a gap to complete an Okazaki fragment. The primase will then form a new primer and a new loop. From Kornberg and Baker.[265] (B) Electron micrograph of the primosome bound to covalently closed ϕX174 duplex replicative form. These enzymatically synthesized duplexes invariably contain a single primosome with one or two associated small DNA loops. From A. Kornberg in Hubscher and Spadari,[266] pp. 9,10.

5. The Replication of Viral DNA

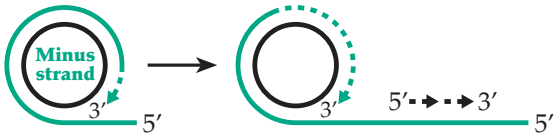
The replication of viral DNA usually depends upon the genes of both the host organism and the virus. For example, *ts* mutations in the *E. coli* genes *dnaB*, *D*, *E*, *F*, and *G* lead to a loss in ability to support

growth of phage λ as well as loss of ability to reproduce under conditions where the *ts* gene products are inactivated. However, the phage can replicate in *E. coli* with mutated genes *dnaA* and *C* because phage λ encodes its own initiator proteins by genes *O* and *P* (marked on the gene map in Fig. 26-4). In addition to these two proteins, seven *E. coli* proteins are required to initiate replication at the lambda origin *ori* λ and to complete replication.[403–407] The *E. coli* dnaB helicase and the dnaC protein are needed, as in Fig. 27-19. The chaperonins dnaJ, dnaK, and GrpE are also necessary for replication of phage λ and other viruses.[408,409] As we have seen (Section 2) many viruses contain genes specifying their own DNA polymerases and primases, which function in cooperation with host proteins. For many dsDNA viruses the origins of replication, priming reactions, and chain elongation processes closely resemble those of *E. coli*.[363,410]

In contrast, the first step in replication of the filamentous Ff viruses (f1, fd, and M13) or the small icosahedral ϕX174 or G4 is conversion of the single-stranded closed circular DNA molecules of the infecting virus particles into circular double-stranded **replicative forms** (RF).[265,411,412] This occurs as the DNA enters the bacterial cell and is accomplished entirely by the enzymes of the host cell. Phage G4 DNA, whose replication has the simplest known requirements, contains a tight hairpin region at its origin. The rest of the DNA must be coated by SSB for replication to occur. The hairpin resists melting and serves as a binding site for *E. coli* primase. This is the only known case in which no other priming proteins are needed.[265] Primase synthesizes up to a 29-ribonucleotide primer after which DNA polymerase III holoenzyme copies the rest of the chain.

Replication of the closely related ϕX174 is more complex. It requires assembly of a **primosome** made up of at least seven host proteins[265,371,413] : dnaB, dnaC, primase (dnaG), and proteins priB, A, C (n, n′, n″), and dnaT (i), Table 27-2. The 76-kDa helicase priA (n′) may locate the **primosome assembly sequences**, which are ~70 nucleotides in length,[414] and displace SSB from them. These sequences can adopt secondary structures with a pair of hairpin loops.[265] Kornberg and associates suggested that the same kind of primosome formed at these sites in the ϕX DNA may participate in replication of the lagging strand of the chromosomal DNA. If so, helicase priA presumably functions in the replisome on one strand and dnaB helicase on the other as depicted in Fig. 27-20.

In the second and third stages of replication ϕX174 RF molecules are themselves replicated and are then used for synthesis of new viral (+) strands. At both stages a virally encoded **gene A protein**, which has endonuclease activity, nicks the duplex. Cutting the (+) strand it leaves a free 3'-OH on DNA residue 4305, while the 5'-phospho group of residue 4306 becomes covalently attached to a tyrosyl residue in the A protein.[411,415] The free 3'-OH serves as the primer for a **rolling-circle synthesis** (Eq. 27-7).[412,416,416a] As a new viral strand is synthesized along the complementary (-) strand as a template, the original viral DNA (+) strand is displaced (Eq. 27-7) as a single-stranded tail.



(27-7)

A strand complementary to the single-stranded tail is then formed in segments. A complete turn of the circle produces a viral strand twice the normal length. Cleavage by the endonuclease activity of the gene A protein and closure of the circle completes the replication. The displaced (+) strand can be cut off and either incorporated into a progeny phage or converted into another RF circle. The A protein, attached to the 5'-terminus of the (+) strand, is involved in either case. It can participate in repeated sequences of initiation and termination of viral (+) strand synthesis.[265,411] Once double-stranded circles are formed, they undergo several replications to give additional RF circles, which serve as templates for the synthesis of many single strands of viral (+) DNA, which are incorporated into the mature viruses. This synthesis of additional RF circles requires transcription of some viral RF genes.

In the final stage of replication the single-stranded (+) chains formed by the rolling-circle mechanism are packaged into phage particles. The gene 5 single-strand binding protein of M13 coats the DNA chains as they are formed, evidently preventing their conversion to RF circles. In the case of ϕX174 the new single-stranded DNA circles are packaged as they are synthesized (Fig. 7-28) to form complete icosahedral virus particles. See Kornberg and Baker[265] for details about these and many other virus replication systems.

Replication of the larger tailed viruses, which have many genes, is complex and varied. The lytic phage λ resembles the smaller viruses in using the host replication enzymes.[265,417,418] In the final stages a rolling-circle mechanism is utilized to form **concatemers** consisting of linear DNA duplexes with numerous successive copies of the viral DNA. The ssDNA that is formed in the rolling circle is converted to dsDNA as it is formed. Finally a **terminase** cleaves the DNA at specific *cos* sites, using staggered cuts, to form cohesive ends.[417,419–420a] However, there are uncertainties.[421] The linear dsDNA enters an empty preformed procapsid, apparently pumped in an ATP-dependent fashion, perhaps by a rotating portal ring.[421a]

6. Packaging of Viral Genomes

The construction of intact virus particles from the genomic DNA and protein subunits is often a complex process. It is simplest for the small filamentous ssDNA viruses (Fig. 7-7). The subunits are synthesized as soluble proteins, which enter the cell membrane, then lose their leader sequences. As the viral DNA coated by the viral gene 5 ssDNA-binding protein enters the membrane, the binding protein is replaced by the coat subunits.[422–424]

The process is somewhat more complex for the icosahedral viruses. In the ϕX, G4, α3 family the icosahedral procapsid is constructed with the aid of both internal and external **scaffolding proteins**[425,426] as is illustrated in Fig. 7-28. In phage that replicate via concatameric dsDNA the terminase that cleaves the DNA also interacts in a precise way with the packaging apparatus of the prohead.[427–429] For the tailed phage the ring-shaped oligomeric head–tail connector (Fig. 7-29), together with an ATPase, may function as a rotatory pump to feed the DNA into the prohead.[430–430c] This has been demonstrated for phage ϕ29.[430a–c] In some cases the terminase produces new phage DNA of unit size, but in other cases, e.g., with phage T4, the DNA may be cut more randomly when the head is full or when another piece is needed to fill it.[431] After the DNA is packaged the virus capsids usually expand and become stronger.[432,432a]

7. Plasmids

Most bacteria contain plasmids which are self-replicating but stably maintained at well-defined numbers of copies per cell.[265,433] They are usually not essential to the cell but may carry traits such as antibiotic resistance or toxin formation that benefit the bacterial host. A plasmid always carries in its DNA an origin (*ori*) of replication and a gene, usually designated *rep*, for an initiator protein. It usually encodes other proteins as well but may depend largely on host proteins for replication. Plasmids may use the *oriC* copied from the bacterial host's DNA, the origin from phage λ, yeast autonomously replicating sequences (ARSs, Section 10), or other origins. Replication of the small 6.6-kb plasmid ColE1, which is present at ~20 copies per cell, depends entirely on the host-cell replication machinery.[265] However, the control of copy number depends upon synthesis of **antisense RNA** and its reaction with the plasmid DNA (see also Chapter 28, p. 1615).[434,435] Similar copy number control is used by the larger ~100-kb resistance factor **R plasmids**.[436] Some plasmids use replication systems very similar to those of viruses such as ϕX174, often using rolling-circle replication.[437] However, the plasmids lack the proteins for virus coat formation and maturation.[438]

The F factor plasmids, discussed in Chapter 26, are large 100-kb circular DNA molecules containing ~60 genes, about 20 of which encode proteins involving transfer of DNA into another bacterial cell (Fig. 26-3).[265,439,440] F plasmids display strict copy number control with only 1–2 copies per host chromosome. The controls lie in a region known as the partition locus, which resembles regions of the host chromosomes that are involved in partition of the bacterial genome. They have repetitive sequences suggesting a similarity to centromeres of eukaryotic chromosomes.[441]

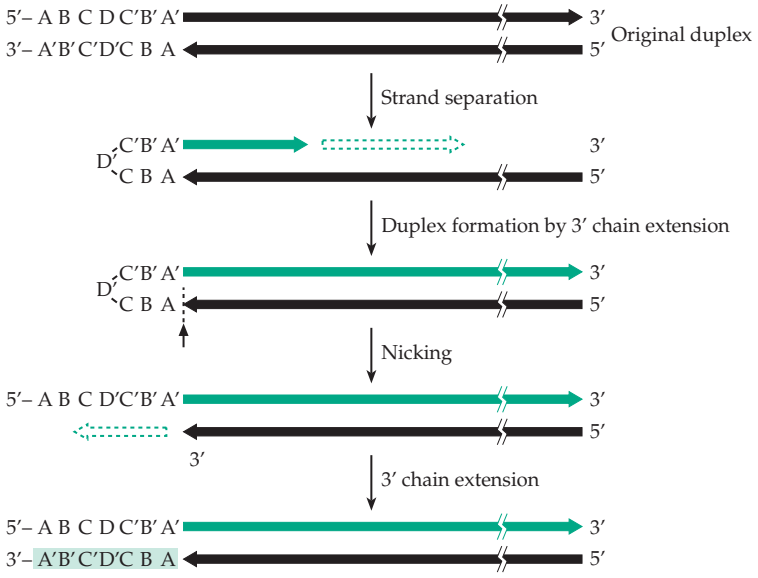
8. Chromosome Ends

The T-odd bacteriophages T1, T3, T5, and T7 are medium-sized phage with linear duplex DNA genomes. Replication of linear DNA in these and in many other genomes presents a problem. Even if the RNA primer segment is made at the very 3' end of the template strand, there will be a gap in the final replicated strand when the primer is digested out. Since there is no known enzyme that will add to the 3' end of a chain, this gap will remain unfilled. The problem is solved by **terminal redundancy**, the presence of a common 260-nucleotide

sequence at both ends. Several daughter DNA molecules with gaps at the 5' ends can be joined by their cohesive ends to form a long **concatamer**. DNA polymerase I fills any gaps, and the chains can then be ligated and cleaved at different points to generate complete 5' ends.[442]

Another mechanism, which is utilized by some single-stranded parvoviruses,[416,443] obviates the need for an RNA primer by use of a palindromic sequence to form a hairpin loop (Eq. 27-8).

Yet another solution to this problem is used by some viruses. **Phage ϕ29** of *Bacillus subtilis* primes the replication of its 19,285 bp dsDNA at both ends by a **terminal protein,** which is linked covalently through its Ser 232 –OH group to dAMP.[443a] The 3'–OH of the deoxyadenosyl group primes the DNA replication. In a similar fashion replication of the eukaryotic **adenoviruses**, whose genome is a 35- to 36-kb linear DNA duplex, starts at the ends and is primed by one residue of dCMP covalently attached through a 5'-phosphodiester linkage to a serine side chain in a 80-kDa **preterminal protein**. It substitutes for the RNA oligonucleotides that prime most DNA synthesis.[444–447] The dCMP pairs with guanine at the 3' terminus of the template strand and provides the initiating 3'-OH group. During the replication the



(27-8)

preterminal protein is cleaved to the 55-kDa **terminal protein,** which remains covalently attached, one molecule at the 5' end of each strand. The genome can be replicated *in vitro* by five proteins: the virally encoded preterminal protein, DNA polymerase, DNA-binding protein, and two cellular transcription factors that bind in the adenovirus origin region.

The chromosome end problem is solved in another way in eukaryotes. As discussed in Section B,1, **telomeres** (chromosome ends) contain repeated sequences of variable length. One DNA strand is always G-rich. For example, in human cells the sequence 5'-$(TTAGGG)_n$–3', where n may be ~20, occurs at the 3' ends of the G-rich strands. The other strand, whose 5' end is at the telomere end, is C-rich and has the complementary sequence 3'-$(AACTCCC)_n$–5'. The 3' end of the G-rich strand is always longer by 12–16 nucleotides than the end of the C-rich strand. This 3' extension may fold back to form non-Watson–Crick structures that apparently involve G-quartets (Chapter 5).[447a,b] The shorter 5' end is thought to result from the need for a short RNA primer during replication. As shown in Fig. 27-21, when replicated in the normal fashion the full G-rich leading strand will be formed, but the C-rich lagging strand will be 8–12 nucleotides short, when the RNA primer is digested away. A result of this is that human somatic cells gradually lose telomeric repeats. However, in tumor cells, germline cells, and unicellular organisms the enzyme **telomerase** prevents this telomere loss.[448–449a]

Telomerase is a reverse transcriptase that copies the DNA sequence of the telomeric repeats from a small **guide RNA** that is part of the enzyme. The first telomerase studied was the relatively abundant enzyme from *Tetrahymena*. It contains a 159-nucleotide RNA with the sequence 5'-CAACCCCAA-3' at positions 43–51. This sequence is complementary to the 5'-TTGGGG-3' repeat sequence of the *Tetrahymena* telomeres.[450,451] A 127-kDa human protein contains a similar guide RNA with the sequence 5'-CUAACCCUACC-3', which is complementary to the human telomere repeat sequence as is illustrated in Fig. 27-21.[452–454] Telomerases[455,456] evidently allow the cell to elongate the telomere 5'-ends using the

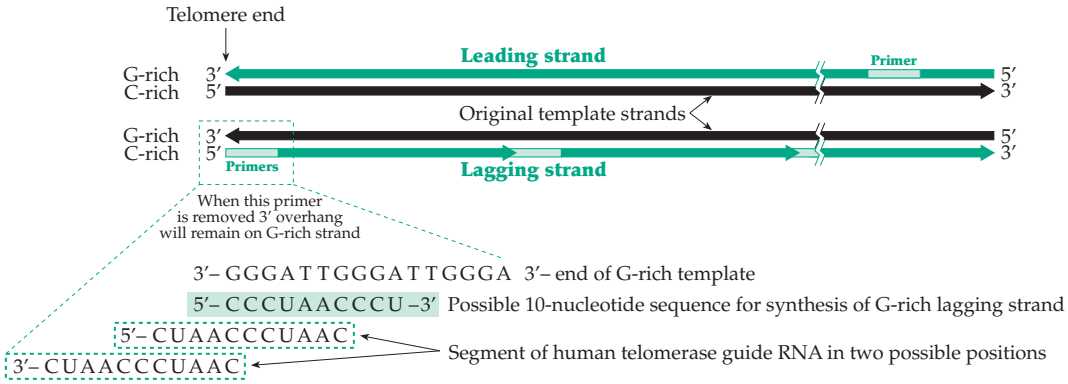


Figure 27-21 Aspects of telomere synthesis. The end of the chromosome and the 5' end of the C-rich strand is at the left. This is the template for replication of the leading G-rich strand (green). The primer lies far back in the chromosome. The C-rich strand is replicated in segments from several RNA oligonucleotide primers, one of which lies at the 5'-terminus. This first primer is removed by RNase activity leaving a 12–16 nucleotide 3'-overhang. The telomerase guide RNA can hybridize with the 3'-end of the G-rich strands providing a template that allows additional growth of the G-rich strand and extension of the C-rich strand also in the next replication.

guide sequence and the reverse transcriptase activity of the telomerase. Any number of additional repeats may be added to the 5' ends. The shortened 3' ends can also be lengthened in the next round of replication.

The control of telomerase must be important. The enzyme is active in early embryonic cells and some stem cells. However, most normal cells have little or no telomerase activity and lose telomere length throughout their lifespan with eventual growth arrest.[448,454,457] On the other hand, excessive telomerase activity may induce cancer.[451,458,459] Certain mutations in the telomerase guide RNA can cause greatly increased telomerase activity.[459] The control of telomerase is still poorly understood but involves specific telomere-binding proteins.[448,454,460–460b]

9. Mitochondrial and Chloroplast DNA

Replication of the ~16-kb mammalian mtDNA begins with RNA priming within a small **displacement loop** or D-loop. One daughter strand, the heavier or H strand, starts to grow on the primer. As it does, the parental H strand is displaced and the D-loop is enlarged. The H-strand grows until ~70% of the parental H strand has been displaced and the L-strand origin (Fig. 18-3) is uncovered. Then a new light L strand is laid down to form the second daughter duplex. The rate of formation of the new L strand is only 10 nucleotides / s, an hour being required to complete the process. The DNA formed is initially relaxed, another 40 min being needed to introduce the 100 superhelical turns present in the finished chromosome.[461]

The kinetoplast DNA of trypanosomes (Fig. 5-16) consists of thousands of catenated circular DNA molecules. Among these the smaller minicircles always contain the sequence GGGGTTGGTGTA at their origins of replication. The minicircles are individually removed from the mass prior to replication. The two progeny circles are then both recatenated into the mass.[462]

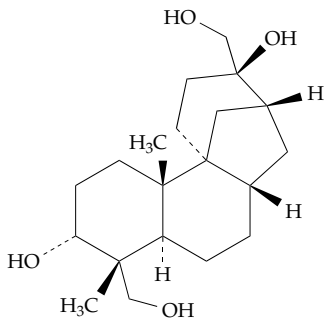
Chloroplasts contain large 120- to 169-kb circular genomes encoding about 100 proteins (Chapter 23). A characteristic feature of most chloroplast DNA is the presence of long inverted repeat sequences (10,058 bp in the liverwort, 25,339 bp in tobacco).[463,464] These are separated by 19,813 and 81,095 bp single copy regions in the liverwort and by similar sized regions in tobacco. Plastid DNA exists as a mixture of monomeric molecules with smaller amounts of dimers, trimers, and tetramers.[464]

Ethidium bromide inhibits the replication of chloroplast DNA and causes partial degradation of existing DNA in chloroplasts without interfering with replication of DNA in the nucleus. The effect is similar to that of the same drug on mitochondrial DNA. However, cells of *Chlamydomonas* treated with ethidium bromide are able later to regenerate their chloroplast DNA. This result has been interpreted to mean that there may be one or a few "master copies" of chloroplast DNA in specially protected locations. The result should also be considered in relationship to the following observation. Although nuclear and organelle DNA molecules replicate at different times in the cell cycle, constant proportions of the organelle and nuclear DNA tend to be maintained. Thus, there must be some kind of control mechanism leading to a coupling of DNA replication in nuclei, mitochondria, and chloroplasts.[465]

10. Replication of Eukaryotic Nuclear and Viral DNA

Replication in eukaryotes is similar in many ways to that in bacteria.[284,466–467a] However, the ~10^6 kb of DNA in a typical eukaryotic genome is divided into many **replicons**, segments of DNA 30–150 kb in length, each having its own origin of replication. In the relatively small ~14-Mb yeast genome there are ~400 replicons,[468,468a] but in mammalian DNA and also in plant DNA[464] there are probably thousands. DNA synthesis is initiated at different times during the S-phase of the cell cycle at the various origins in an ordered pattern.[469,469a,b] An important unanswered question is how the cell is able to replicate all segments of all of the chromosomes just once before entering mitosis.

Many of the proteins of eukaryotic replication are closely related in sequences and functions to those of bacteria. There are initiator proteins analogous to *E. coli* dnaA (Fig. 27-19). The DNA polymerases have been discussed in Section C, 2 (see Table 27-1). Eukaryotic polymerases α and δ and possibly ε are essential for replication.[470,471] Polymerase α, which is inhibited specifically by the fungal metabolite **aphidocolin**,[472] is a complex of a ~170-kDa DNA polymerase core, an RNA-synthesizing primase consisting



Aphidocolin

of 58- and 49-kDa subunits, and a 70-kDa subunit of uncertain function.[473–474a] The complex makes an RNA–DNA primer consisting of ~10 nucleotides of RNA and ~30 of DNA.[475] This pol α/primer is replaced early in replication by the highly processive polymerase δ and, perhaps, under some conditions by polymerase ε. The ringlike processivity factor or "clamp" that is provided in *E. coli* by protein dnaB is called the **proliferating cell nuclease antigen** (**PCNA**) in eukaryotes.[476] It is loaded onto the DNA by a **clamp loader**, the 5-subunit replication factor C (RFC).[476a] As in *E. coli* (Fig. 27-19) an SSB type protein known as **replication protein A** (**RPA**) is also essential.[467,477] PCNA is not only essential to eukaryotic replication but is also required for recombination and repair.[476] The polα/primer primes leading strand synthesis initially but then switches to replication origins on the lagging strand where, together with other proteins, it primes the formation of the Okazaki fragments. Polα/primer is also a logical participant in the control of the initiation of the S-phase of the cell cycle.[478]

Eukaryotic viruses. Investigation of viruses provided the first insights into eukaryotic DNA replication. Most of the factors needed for replication of the DNA of adenoviruses, simian virus 40 (SV40), and polyomavirus[447] within animal cells are supplied by the host. Replication of the 5-kb SV40 DNA, whose DNA forms typical nucleosomes (Fig. 27-3), appears to be an excellent model for eukaryotic replication in general.[479–482] The single SV40 origin of replication is a 64-bp sequence containing the 5-bp sequence GAGG C four times as pairs of inverted repeats. These are recognized by the 95-kDa virally encoded initiation protein which also has helicase activity and is known as the **T antigen**.[483,483a] A nearby 17-bp sequence containing only AT pairs is presumably the region of entry of the host cell's polymerase α/primer. Single-stranded DNA regions are coated with the replication protein A. After the primer is formed, the RFC complex loads the sliding clamp PCNA, and polymerase α is replaced by polymerase δ on both leading and lagging DNA strands permitting highly processive bidirectional chain elongation. Topoisomerase activity is required to decatenate the replicated chromosomes. Since SV40 DNA forms typical nucleosome (Fig. 27-3), its replication is thought to mimic chromosomal replication quite closely. The more complex herpes simplex virus HSV-1 has a 153-kb genome, a linear DNA duplex. It has ~75 genes and encodes its own DNA polymerase, origin-binding protein, SSB, and other proteins needed for replication within eukaryotic cells.[484,485]

Artificial chromosomes. Another approach to understanding eukaryotic replication, similar to the use of *ori*C plasmids in *E. coli*, is to study **autonomously replicating sequences** (**ARSs**)[469,486] and plasmids[468] and **artificial chromosomes**[487] made from them. ARS sequences were first found in the budding yeast *S. cerevisiae*. Plasmids containing an ARS, whose core consensus sequence is 5'-(A/T)TT TAT(A/G)TTT(A/T), replicate autonomously during S-phase. Such plasmids have been genetically engineered, providing them with telomeres and some kind of functional centromere, to form artificial chromosomes. **Yeast artificial chromosomes** (**YACs**) have become extremely important as cloning vehicles (Chapter 26), and they also serve as important tools for studying eukaryotic replication and its control. They can be cultured in yeast cells or can be transferred into animal cells, etc.

As mentioned in Section B, 1, human centromeres are rich in the repetitive α-satellite DNA. By joining α-satellite DNA-containing fragments of the X-chromosome to cloned telomeric DNA, human **minichromosomes** have been created.[488] These have been developed into **human artificial chromosomes**,[489] which may be practical vehicles for gene transfer in human therapy.

Replication of nuclear DNA. The budding yeast *Saccharomyces cerevisiae* has permitted the most detailed picture of DNA replication in a eukaryote. The complete genome sequences are known and the ARSs have been physically mapped.[468a] For example, in chromosome VI there are nine origins that differ in frequency of initiation and which replicate sequentially during the S-phase of the cell cycle.[490] The initiation (**replicator**) regions surround the 11-bp consensus sequence of the ARSs, each occupying at most ~150 bp. However, in metazoa and even in the fission yeast *Schizosaccharomyces pombe* the ARSs range from 500 to 1500 bp in length. These origin regions frequently overlap the **promoter** sequences, which control initiation of transcription (Chapter 28).[491] This association with transcription origins has also been observed in metazoan cells, where replication origins are often clustered.[492] However, there is no sequence homology between the ARSs of *S. cerevisiae* and replication origins in other species, even those of *S. pombe*.[490]

The study of replication in yeast ARSs and artificial chromosomes has revealed that initiation of replication requires not only an initiator protein but a complex of six proteins that form an **origin recognition complex** (**ORC**).[493–495a] This complex, which is essential to initiation of replication, may be joined by additional proteins in a **prereplication complex**. At least some of the ORC proteins have their homologous counterparts in metazoa, suggesting a highly conserved initiation machinery.[495a–c]

One or more of the proteins that bind to the ORC may constitute a **license** to replicate. The licensing

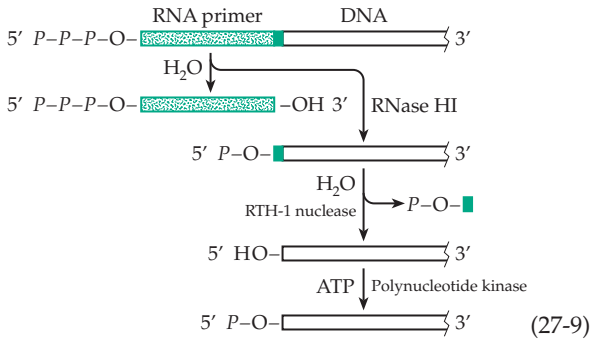
concept states that when replication occurs the license is destroyed and the origin involved cannot initiate replication again without a new license. The **replication licensing factor** (**LRF**) is postulated to be unable to pass through the nuclear membrane.[494,496] It can only reach the replication origin after the S-phase has concluded and mitosis has taken place. At this time the membrane has been disrupted. A second signal, the **S-phase promoting factor** (**SPF**), cannot act without an intact nucleus and a license in place.[469,495,497–500] This system ensures that DNA is replicated only once per cell cycle. Among the proteins involved in the licensing is a group of **minichromosome maintenance** (**MCM**) **proteins**, so-named because of their importance to replication of ARSs and artificial chromosomes.[495] Six of these proteins (MCM2–MCM7) can form a hexameric complex with one subunit of each type as well as other complexes, e.g., $(MCM4,6,7)_2$.[500a–c] The latter acts as an ATP-dependent helicase. A somewhat simpler MCM complex is found in archaea.[500d] Some of the cell cycle proteins (Chapter 26, Section F,1), including Cdc6 and the protein kinases Cdc7 and Cdc28 as well as other proteins, are also required for regulation of replication origins. Proteins homologous to those of yeast have been identified in humans and other eukaryotes. Licensing of replication involves association of the MCM helicases with each ORC during the G1 phase of the cell cycle. Binding of the initiation factor Cdc6/18 and of a recently discovered loading factor **Cdt-1** apparently completes the licensing. Once licensing has occurred both cdc6/18 and Cdt-1 can dissociate from the DNA.[496] Removal of Cdt is facilitated by its binding to another protein **Geminen**, found first in the frog *X. laevis*.[500e–g] The ORC complexes may remain at the origins. It has been estimated that a yeast cell contains ~400–600 molecules of the very stable ORC, about one ORC per replication origin. However, a large excess of MCM proteins may be present. Their concentrations may regulate the number of ORC molecules that associate with DNA. Replication of DNA is not the only aspect of cell growth. For example, as DNA is replicated histones must be synthesized and assembled. This synthesis occurs during the S-phase and is tightly coupled to replication.[500h]

Initiation of replication in metazoans is still confusing. Almost any piece of DNA will be replicated if introduced into a *Xenopus* egg, where initiation appears to occur just once at a random position.[501,501a,b] However, in differentiated tissue the origins of replication seem to be fewer in number and more specifically located. A possible explanation is that high concentrations of ORC and MCM proteins in the embryo may lead to many relatively nonspecific origins and a replicon size of ~7 kb. The lower concentrations of these auxiliary factors in somatic cells may lead to fewer but more specific origins with a replicon size of ~170 kb.[495] Three distinct mammalian origins have been studied in detail. One is in the β globin locus (Fig. 27-10).[500] A second is near the dihydrofolate reductase gene.[502] A third, which is activated early in S-phase, is at the 3' end of the lamin B2 gene.[503] The latter has been localized to a 500-bp region. These findings suggest that the replicon concept, as developed for yeast, may be generally applicable.

Replication of the intact genome of *Drosophila* has been studied in rapidly dividing nuclei by electron microscopy.[504] The replication rate in these nuclei is ~300,000 bases / s, but it has been estimated that replication forks in animal chromosomes move no faster than ~50 bases / s. Thus, we would anticipate at least 6000 forks, or one fork per 10 kK bases. Indeed, this number of forks has been observed.[505] They occur in pairs with many short regions containing single-stranded DNA as if one strand at the fork is replicated more rapidly than the other as in mitochondrial DNA. The arrangement of the ssDNA regions at the two forks in a pair suggests bidirectional replication. However, replication forks are rarely seen in higher eukaryotes, but extensive regions of single-stranded DNA are often visible. Benbow and associates suggested that in higher eukaryotes the strands of duplex DNA may be separated throughout a whole looped domain of DNA. Replication could then occur with initiation at many points along each strand.[505]

Replication reactions are similar in bacteria and eukaryotes, but some details differ. In eukaryotes at least two DNA polymerases, α and δ, are required. In budding yeast polymerase ε is also essential.[506] Both polymerases δ and ε may replicate separate strands at the replication fork.[506a] Processing of Okazaki fragments also differs from that in bacteria, where either RNase H or the 5' to 3' exonuclease activity of DNA pol I removes the RNA primer (Fig. 27-14). This exonuclease activity is lacking in eukaryotic polymerases. Replication primers are removed in a two-step process by **RNase HI**, which makes an endonucleolytic attack that removes all but one nucleotide residue of the primer in a single piece, leaving a 5'-phospho group on the remaining ribonucleotides. That residue is removed by a 5' to 3' exonuclease designated RTH-1 nuclease (Eq. 27-9).[467,507,507a,b] This is a homolog of the yeast RAD27 protein. A polynucleotide kinase may then phosphorylate the 5' end of the DNA fragment.[508]

Another difference between bacterial and eukaryotic replication is the presence of nucleosomes in eukaryotes. Some evidence suggests that nucleosomes may open and close to allow replication forks to pass through.[509] Studies of SV40 minichromosomes indicate that passage of the replication machinery does destabilize nucleosomes, which must be partially reconstructed about 260 nucleotides past the elongation point.[510] Another factor is the variable extent and location of modifications to histones, in particular to



(27-9)

the H3 and H4 histone tails (Section A,3). A code has been proposed according to which certain modifications would favor transcription or mitosis, while lack of modification would silence the genes.[72]

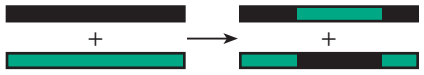
Much of the control of replication is at the initiation stage. Growth factors and other mitogenic stimuli acting at the plasma membrane can stimulate expression of such nuclear proteins as those encoded by the proto-oncogenes *c-myc*, *c-myb*, and *c-fos*. These may initiate a regulatory cascade (Fig. 11-13) and trigger mitosis.[511] As indicated in Chapter 26, the 34-kDa protein kinase encoded by fission yeast gene *cdc*2 (budding yeast CDC28) is essential for progression of the cell cycle through the G1 phase into mitosis (Eq. 26-3). A single oscillation in this kinase activity induced by a B-type cyclin can promote both replication and mitosis. However, in *S.cerevisiae* there are 14 different cyclinlike proteins, and their individual functions are not clear.[512] The signal that is sent to the ORCs is likewise unclear.[513] However, theoretical models involving Eq. 26-3 and many additional components have been proposed.[514] Multiple phosphorylations may occur, some on the RPA initiator protein.[515] Many proteins required for replication, including DNA polymerase and primase, are associated with the nuclear matrix.[516] The nuclear membrane may also be important in controlling replication.

D. Integration, Excision, and Recombination of DNA

The exchange of genetic information between chromosomes, plasmids, and viruses occurs in many ways, which are described collectively as recombination.[517–520] Mutants of *E. coli* deficient in recombinational ability often have defects in genes designated *recA, B, C,* etc. (for recombination), or *ruvA, B, C* . . . (for resistance to ultraviolet light). Some of these mutants are unusually sensitive to ultraviolet light because of their inability to repair damage to DNA. Several of the recombination enzymes are used for repair of ultraviolet damage and of double-strand breaks in DNA arising from other causes.[521,522] In eukaryotes recombination occurs during meiosis. Many viruses, including phage, also carry genes for their own general recombination systems. In addition, the DNA of some viruses, such as the temperate phage λ, undergoes recombination with the host DNA. This can happen during the processes of integration of the viral DNA into the host genome or excision of the viral DNA during the lysogenic cycle of replication (Chapter 28, Section B). Recombination occurs around specific sites in the chromosomes of both the virus and its host and is called **site-specific recombination**. Genetic recombination is essential to the development of genomic diversity, to the survival of a species, and to evolution.

1. Recombination Mechanisms

How can the homologous regions of two different DNA duplexes be brought together? As illustrated schematically in Eq. 27-10, the strand exchange must occur at exactly the same point in each duplex. An early attempt to explain this postulated a "copy choice" mechanism of replication. It was assumed that replication occurred along one DNA strand up to some random point at which the polymerase jumped and



(27-10)

began to copy from the second of a pair of homologous chromosomes. The newly formed DNA molecule would be complementary to different parts of both parental DNA duplexes. To test the idea, Meselson and Weigle infected *E. coli* with two strains of phage λ containing ^{13}C- and ^{15}N-labeled DNA, respectively.[523] Recombinant DNA was found to contain some ^{13}C and some ^{15}N, as judged by density gradient centrifugation. It was clear that DNA from both parents was incorporated into DNA of recombinant progeny, a finding that ruled out the copy choice hypothesis.

If recombination occurs instead by enzymatic cutting of two homologous duplex DNA molecules followed by rejoining, how is it possible to avoid inactivation of genes by addition or deletion of genetic material? Recombination cannot depend upon the random action of a nonspecific enzyme with random rejoining. Yet, general recombination can occur at any point and with a roughly constant frequency throughout the DNA chain. The explanation of these facts lies in the occurrence of base pairing between at least some short homologous regions of strands of the two different DNA duplexes.

The Holliday recombination intermediate. In 1964, Holliday suggested a recombination process that would give rise to characteristic H-shaped intermediates.[524] Recombination could be initiated at special points on the duplexes, recognizable by a recombination enzyme (Fig. 27-22). A short amount of unraveling would be followed by strand exchange with the two broken strands being rejoined by a ligase as indicated in Fig. 27-22. The crossover points would then migrate up or down the chains as the two helices turned about their own axes. Long regions of heteroduplex DNA could be generated in this way, and the process could be terminated at a random distance from the starting point, accounting for the observed uniformity of genetic recombination events. Chain cleavage and rejoining of two of the strands would terminate the process. If these were the same strands broken in the initiation event (cleavage at points *aa*′ in Fig. 27-22), genes lying outside the heteroduplex region would not be recombined, but cleavage of the other chains (at points *bb*′) would lead to their recombination. Intermediates of the type predicted by the Holliday model were soon observed by electron microscopy (Fig. 27-23).[525] Three-dimensional structures have been determined by X-ray crystallography[525a] and have been studied by atomic force microscopy.[525b]

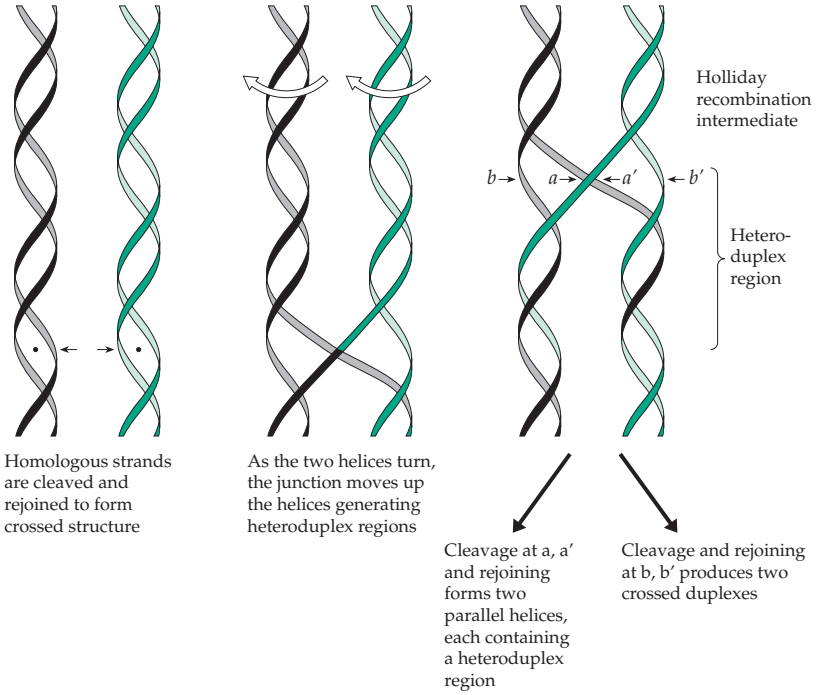
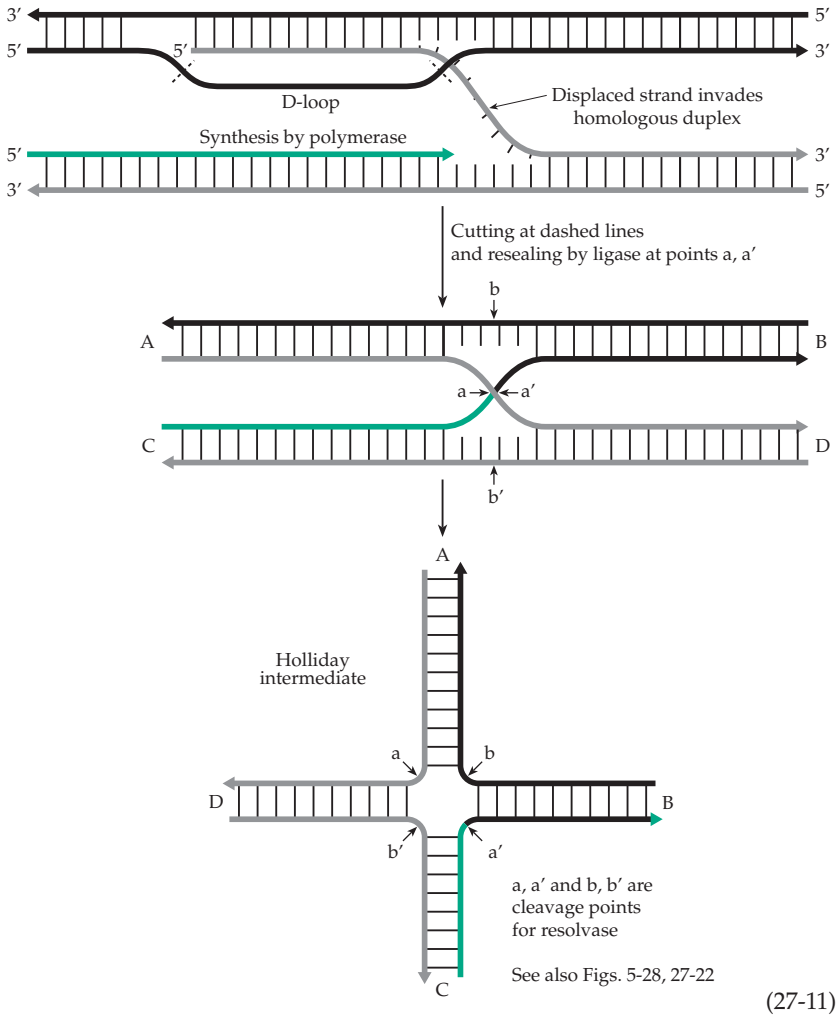


Figure 27-22 A recombination mechanism involving single-stranded exchanges. After Holliday.[524]

The cross-stranded structure shown in Fig. 27-22 can be formed with all base pairs in both duplexes intact.[526,527] All that is required is formation of a nick in each of the two polynucleotide chains and a rejoining of the backbones across the close gap between the duplexes. This model also accounts for the cutting of the two crossed strands at exactly equivalent points to terminate the process. Various mechanisms of recombination exist, and most make use of the key Holliday recombination intermediate (Fig. 27-22, Fig. 5-28, Eq. 27-11).[521,528–530a] Such four-way junctions can arise in several ways.[530a–d] For example, a 3' or 5' single-stranded tail in a piece of dsDNA can "invade"

Figure 27-23 A chi form of DNA from the colicin E1 plasmid. These forms are thought to be derived from recombination intermediates of the Holliday type, which appear as "figure eight"-shaped molecules twice the length of the colicin genome. This figure eight form was cut at a specific site that occurs only once in the genome (twice in the figure eight) by restriction enzyme *Eco*R1 to give the chi form. The pairs of long and short arms are believed to represent homologous duplexes. The single strands in the crossover have pulled apart revealing the strand connections clearly. Such a structure would be expected from the Holliday intermediate (upper right corner of Fig. 27-22), e.g., if one of the two vertical duplexes were rotated end over end. From Potter and Dressler.[525]



(27-11)

another dsDNA that has a homologous sequence as indicated in Eq. 27-11. In this drawing a 5' tail has been displaced during repair of a gap in one strand. The resulting D loop may be trimmed out and a new connection made to give the Holliday intermediate. The cleavage points a, a′, b, b′ marked in Eq. 27-11 correspond to those in Fig. 27-22. Holliday junctions may also be formed in stalled replication forks and must be removed to allow replication and transcription to continue.[530e] The existence of the Holliday intermediate has been supported not only by electron micrographs such as that of Fig. 27-23 but also by the identification of endonucleases that carry out the necessary cleavages of synthetic Holliday intermediates that have been made artificially (see Chapter 5.) Endonucleases with a high specificity for Holliday junctions have been found in bacteria, among proteins encoded by viruses, and in a wide variety of eukaryotic cells. Additional proteins including helicases, DNA-binding proteins, and specialized **strand exchange proteins** are also required to catalyze the individual steps in the recombination process.

The main **RecBCD pathway** of recombination in *E. coli* depends upon a dsDNA nuclease and an unwinding complex consisting of proteins RecB, C, and D.[531–533] The complex is a powerful *exonuclease*, which can digest the ends of a DNA duplex. It degrades the 3' ends most rapidly, leaving 5'-tails that can invade other homologous duplexes as in Eq. 27-11. The RecBCD complex is also an ATP-dependent helicase, which unwinds the DNA, preparing ssDNA for reaction with the strand-exchange protein **RecA**. The RecBCD complex also functions to completely degrade foreign dsDNA such as that from invading bacteriophages.[532] Why doesn't it also degrade the genomic DNA of the *E. coli* cell in which it functions? The answer lies in an eight-base DNA sequence, a recom-bination "hot spot" called **chi (χ)**: 5'–GCTGGTG G–3'. This χ sequence occurs 761 times in the leading strands for DNA replication in the *E. coli* genome.[533a–534a] When the RecBCD complex reaches a χ sequence, when approaching it from the 3' end, the enzyme stops its exonuclease action by inactivating the nucleolytic activity of the D subunit and promotes recombination about five- to ten-fold as fast as at other sites.[533,535]

RecA and other strand-exchange proteins. The 352-residue product of the *E. coli RecA* gene is a multifunctional **recombinase,** which is required both for recombination and also for DNA repair.[536–540a] In its repair function the RecA protein acts as a DNA-dependent protease that cleaves a number of repressors in response to damage to DNA. It has a quite different role in recombination where it (1) brings a piece of single-stranded DNA (an end or a gap) together with a duplex; (2) locates homologous sequences; and (3) forms a synaptic complex in which strand exchange can occur. Electron microscopic observations[536] show that the RecA protein binds to either single-stranded or duplex DNA in a cooperative manner to form long rodlike spiral filaments (Fig. 27-24). Measurement of the lengths of RecA protein-covered duplexes shows that the DNA is underwound and stretched by about 50%. It contains ~18 nucleotides per turn.[536,539,539a] Similar filaments are formed with single-stranded DNA, 3–4 nucleotides being bound per RecA protein monomer. Formation of this ssDNA complex, which may be regarded as an initiation

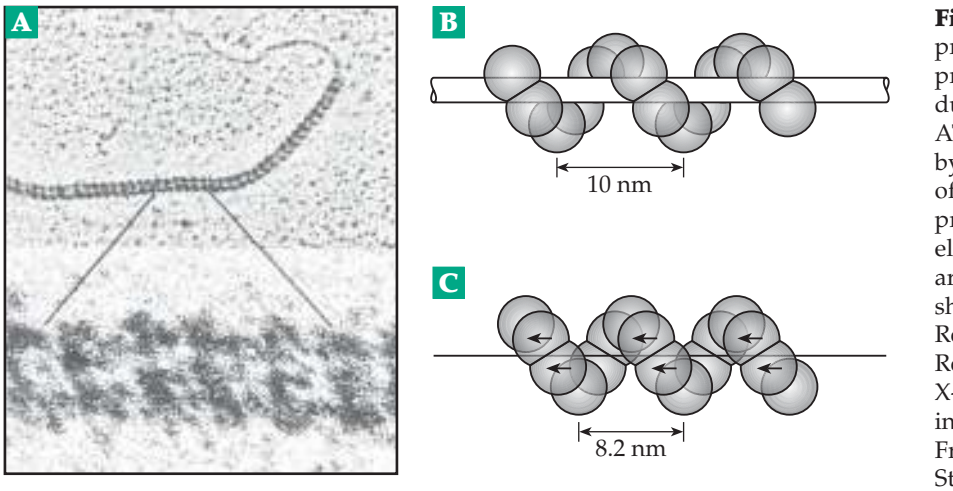
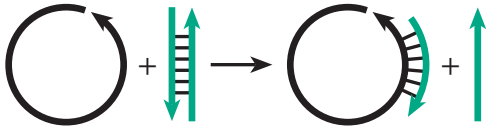


Figure 27-24 Structures of RecA protein spiral filaments. (A) RecA protein filament formed on circular duplex DNA in the presence of ATP(γ-S), shadowed with Pt and seen by electron microscopy. (B) Diagram of RecA bound to duplex DNA in the presence of ATP(γ-S), as determined by electron microscopy. RecA monomers are shown as spheres, but their exact shape is unknown. (C) Diagram of RecA spiral filament in crystals of RecA protein free of DNA, based on X-ray crystallographic data. Arrows indicated alignment of monomers. From Howard-Flanders, West, and Stasiak.[536]

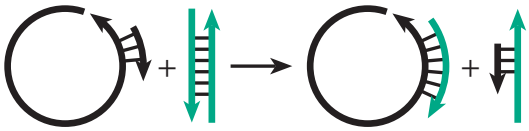
complex for recombination, requires MgATP and is faciliated by prior coating of the DNA with SSB protein. The RecA protein subunits are added in the 5'- 3' direction of the DNA, and SSB is displaced in the process.[541]

The initiation complex binds to duplex DNA rapidly and more slowly promotes strand exchange. In related reactions a single-stranded SSB-coated circular DNA will bind to RecA protein, then exchange strands with a linear duplex (Eq. 27-12). The strand exchange requires ATP and advances in the 5' to 3'



(27-12)

direction along the original single strand at the rate of a few bases / s. Strand exchange can also occur between two duplexes if there is a suitable gap in one strand, e.g., as is illustrated in Eq. 27-13.



(27-13)

A possible mechanism of strand exchange is illustrated in Figs. 27-25 and 27-26. The RecA protein has binding sites that can accommodate nucleotides from two DNA molecules, one single-stranded and the other a duplex. It may also accommodate two DNA duplexes.[542] As shown in Fig. 27-25, the RecA protein could test the hydrogen-bonding between many base pairs at once in a search of homologous regions. The two DNA chains would have to either slide past each other or repeatedly dissociate and reassociate[543] until a homologous region was found. Then strand exchange could occur. As is shown in Fig. 27-26, the single strand may be wound into the major groove of the duplex to form an interwound triplex. The matching of hydrogen-bonding atoms may be an attractive way of searching for homology, but the actual search seems to substitute speed for precision. Base substitutions are quite permissive. The need for precise hydrogen-bonding has not been demonstrated, and the exact recognition mechanisms in homologous recombination remain uncertain.[544–544b] Whole chromosomes must be aligned and checked rapidly in the homology search.[545]

Many proteins similar to the RecA protein and with similar functions have been found.[546,547] These include the products of gene *uvsX* of phage T4,[548,549] the β protein of phage lambda,[550] the yeast RAD51[551,551a] and human RAD51 proteins,[552,553] a meiosis-specific human homolog of the RecA protein,[554] and corresponding proteins from plastids of higher plants.[555] Both the UvsX protein of phage T4 and human RAD51 protein yield strands of coated DNA similar to those in Fig. 27-24.

Processing the Holliday junction. Completion of the recombination process requires "resolution" of the Holliday intermediate by endonuclease action followed by ligation and perhaps by gap repair. The major recombination pathway in *E. coli* employs a binding protein, a nuclease, and a helicase encoded by genes *RuvA*, *B*, and *C*.[528] **RuvA** is a DNA binding protein specific for symmetric Holliday junctions.[529] **RuvB** is a closely associated ATP-dependent helicase.[556–558] On the basis of genetic and X-ray crystallographic evidence it is now evident that some of the functions previously attributed to RecA are carried out by the RuvABC complex. As indicated in Fig. 27-26B, RuvA binds to the Holliday junctions, holding it in the symmetric square configuration in which branch

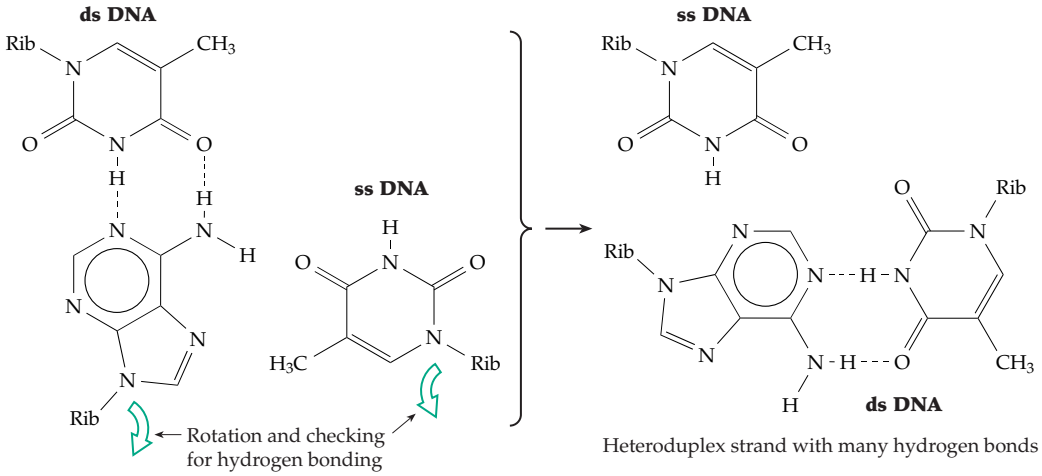


Figure 27-25 A possible mechanism for homologous pairing of an ssDNA with a duplex DNA and strand exchange. The ssDNA (right) binds together with a hydrogen-bonded duplex (left). The RecA protein rotates the bases into the heteroduplex configuration, where hydrogen bonds may be formed in many of the base pairs. After Howard-Flanders, West, and Stassiak.[536]

migration is possible. Two molecules of the oligomeric RuvB helicase apparently rotate the DNA, causing branch migration and movement of the DNA through the RuvAB complex.[559,559a,b] Under some circumstances a different helicase, encoded by *E. coli* gene **RecG,** moves Holliday junctions in the opposite direction.[560] In addition to RuvA a variety of other Holliday junction-binding proteins are known.[561] These include **p53** and the nuclear HMG proteins.[560,561]

RuvC is an endonuclease that is highly specific for Holliday junctions. It is a **resolvase** that cuts at either points a,a′ or b,b′ of Eq. 27-11 to form either "patched" or "spliced" recombinant DNA (Fig. 27-26C). Similar resolvases process bacteriophage DNA[562–564] and have also been found in yeasts and in mammals.[565,566] All are dimeric metal ion-dependent proteins.[567]

2. Nonreciprocal Recombination and Unequal Crossing-Over

The phenomenon of **gene conversion** or **nonreciprocal recombination**[568,569] was first recognized in genetic studies of fungi for which the four haploid miotic products can be examined individually (tetrad analysis; p. 20). Instead of the normal Mendelian ratio of 2:2 for the gene distribution in the progeny at a heterozygous locus a ratio of 3:1 is sometimes observed. One of the recombinant chromosomes appears to have been altered to a parental type. A reasonable mechanism by which this can occur arises from the fact that heteroduplex regions, which are present in recombination intermediates, contain defects in base pairing. One strand of the heteroduplex will have a base that does not properly pair with the base in the other strand, or will have an extra base that loops out from the heteroduplex. Since cells contain repair mechanisms that search for defects and carry out a repair process, there is a likelihood that one strand in the heteroduplex region will be altered to restore perfect base pairing, thus causing the observed gene conversion. "Flanking" genetic markers outside of the heteroduplex region are unaffected by gene conversion, and during meiosis crossing-over between these markers occurs in about 50% of gene conversion events as would be predicted from the model of Fig. 27-21. Data from yeast show that nonreciprocal recombination during meiosis may also result from double-strand breaks and gap formation followed by repair synthesis using both strands of the homologous chromosome as templates.[569]

Recombination is not limited to meiosis but can occur between homologous chromosomes during mitosis, during the G_1 period preceding mitosis, or even during the G_2 period.[570,571] Certain mutations in yeast abolish meiotic recombination but have much less effect on mitotic recombination.[572] Thus, the two processes are not identical. It has been suggested that mitotic recombination is utilized to maintain sequence homogeneity between repeated eukaryotic genes.[572,573]

Since DNA contains many repeated sequences, crossing-over sometimes occurs between locations that are not the same in the two duplexes. Such **unequal crossing-over** has the effect of lengthening one duplex and shortening the other. This may be very important in evolution. It may also, surprisingly, function to preserve homogeneity of chromosomes within a species.[574,575] For example, tandem arrays of ribosomal RNA genes (Section B,3) in yeast have 140 identical copies of their 9-kb repeat unit.[575] Unequal crossing-over between either sister chromatids or homologous chromosomes, when repeated often enough, can lead statistically to a highly homogeneous population.

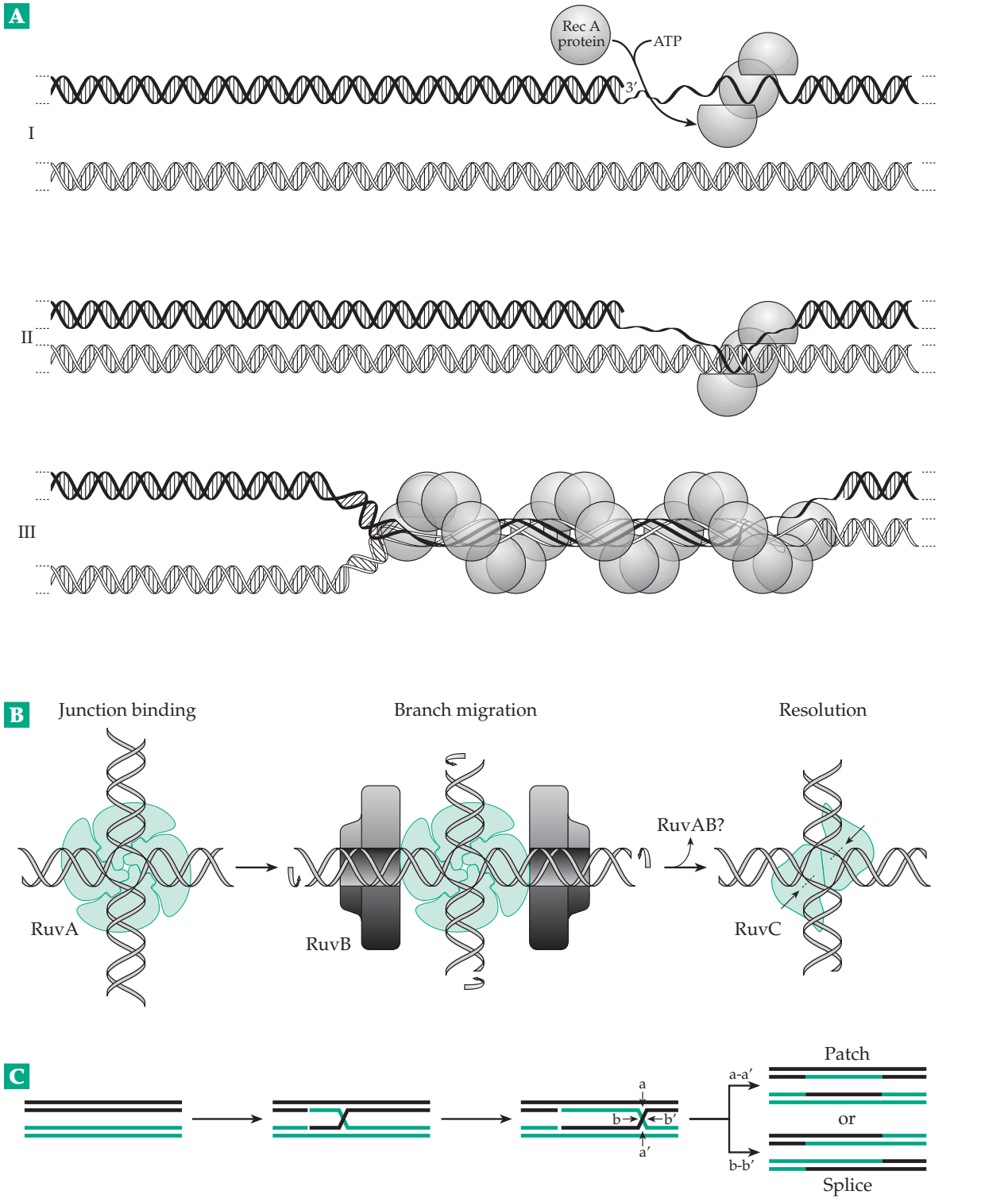


Figure 27-26 (A) Model for genetic recombination proposed by Howard-Flanders *et al.*[536] (I) RecA protein binds cooperatively to the single strand in a gapped duplex to form an initiation complex in preparation for pairing. (II) The initiation complex binds to the intact duplex, making transient contacts until a homologous site is reached. For clarity, the initiation complex is drawn with only a few protein monomers, but in reality it is likely to extend over hundreds or thousands of nucleotides. (III) When homologous contacts are made and the strands become paired locally, the initiation complex acts as nucleus for further cooperative binding, which extends the RecA spiral filament around all three or perhaps all four interacting strands. (B) Arrangement of the proteins and DNA during three stages of recombination catalyzed by the RuvABC system. The two RuvB hexameric rings are shown in cross section with the DNA passing through their centers. After Rafferty *et al.*[529] See also chapter banner, p. 1527. (C) Scheme of DNA rearrangement during homologous recombination in *E. coli*.

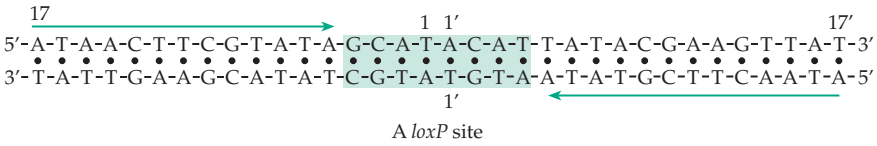
3. Site-Specific Recombination and the Integration and Excision of DNA

Recombination at specific sites in DNA is responsible for integration of DNA from viruses into the genome and for the cutting out of viral DNA and other pieces of DNA from the genome. The temperate bacteriophage λ and the F factors and R factors of bacteria can all be integrated into the genomic DNA of the host in this way. Genes encoded by the phage or plasmid are required. In the case of phage λ the viral genes *int* and *xis* are required for integration and excision, respectively.[576] These are not the same as the enzymes of the *rec* loci of the bacterium or the general recombination genes *exo* and *bet* of the phage. In addition, both integration and excision require an *E. coli* protein called integration host factor (**IHF**), a DNA bending protein resembling the DNA-binding HU.[18b,265,577,577a]

Integration of λ DNA (Fig. 27-27) occurs at the ~25-bp site *att B* in *E. coli* (Fig. 26-4) and the ~240-bp site *att P* in the λ chromosome (Fig. 28-11). These two sites contain identical 15-bp core sequences within which the recombination occurs. In a manner similar to that of the recA protein a homologous region is located by the complex of the Int (**integrase**) protein and IHF. Several molecules of Int protein bind and, together with the IHF protein, hold the phage DNA in a nucleosomelike structure (an **intasome**) in which the recombination occurs.[406,578–580b] Strand cleavage and rejoining occur within the short core sequence (Fig. 27-27). Parts of both *att P* and *att B* are recombined to give sites *att L* (left) and *att R* (right) in the DNA of the integrated prophage. In the integration complex the two core *att* sequences are aligned, and single-strand cuts are made at one of the points *a* or *b* that are indicated by the small arrows located on opposite strands and seven bp apart in the core sequence shown in Fig. 27-27. Rejoining of strands from the opposite duplex yields a Holliday intermediate. That this really occurs is shown by the fact that the Int protein cleaves synthetic Holliday intermediates derived from the *att* core and reseals the strands to give the expected products.[581] Cleavage of the Holliday intermediate at points a (Fig. 27-27) will lead to excision of the viral circle, but cutting at points b followed by resealing with opposite strands, as is observed, will yield integrated prophage. Although Int and IHF proteins are sufficient to promote integration, the **excisionase** encoded by phage λ gene *xis* is needed together with the Int protein for excision of the λ prophage.[576,582]

The integrase (tyrosine recombinase) family.

The lambda integrase is the first recognized member of a family of a hundred or more closely related enzymes that are involved not only in integration and excision of phage DNA but also in converting multimeric forms of bacterial and plasmid chromosomes into monomers. One of the best known integrases is the 38.5-kDa **Cre recombinase**, which functions to keep the lysogenic phage P1 in a monomeric form by recombination between pairs of 34-bp core sequences designated *loxP*.



A *loxP* site

Since the reaction doesn't require accessory protein factors and can be performed *in vitro* with a variety of DNA substrates, the *Cre–loxP* system is much used in genetic engineering.[583,584] A pair of related integrase subunits known as **XerC** and **XerD** perform a similar function for the *E. coli* chromosome as well as for multicopy plasmids.[585,586] The XerC / XerD system is

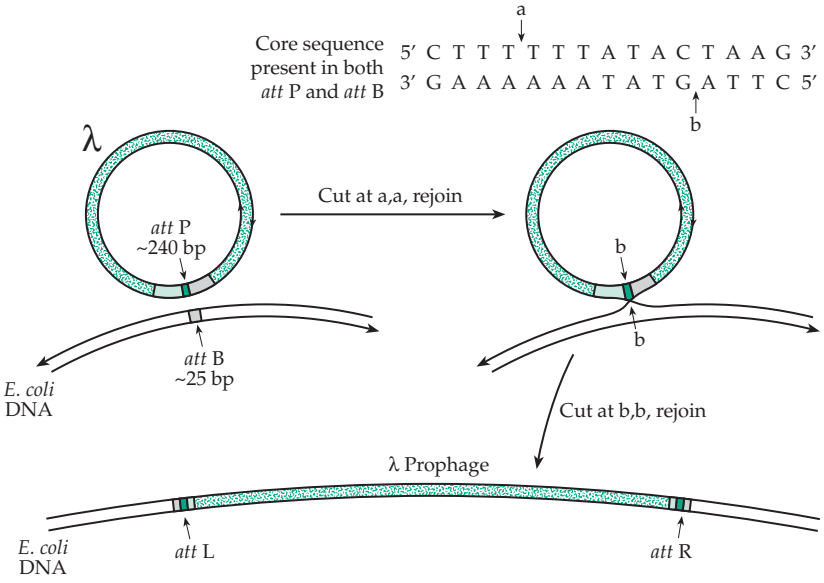


Figure 27-27 Integration of the phage λ genome into the *E. coli* chromosome at site *att*B. The same recognition sequences are present at *attB* and *attP*. These are cut at points *a* with rejoining to give the structure at center right. This is cleaved at points *b* with rejoining to give the integrated prophage.

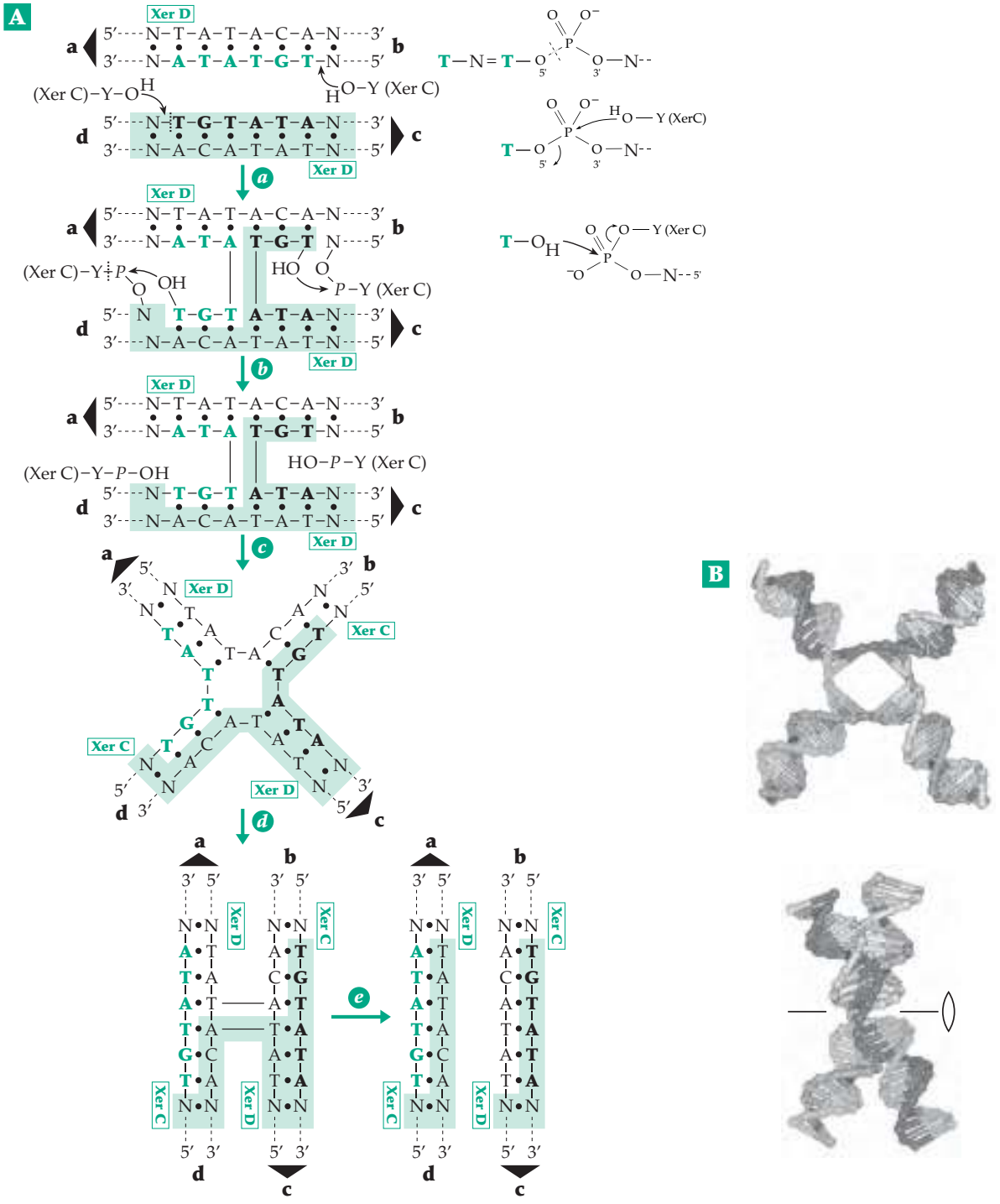


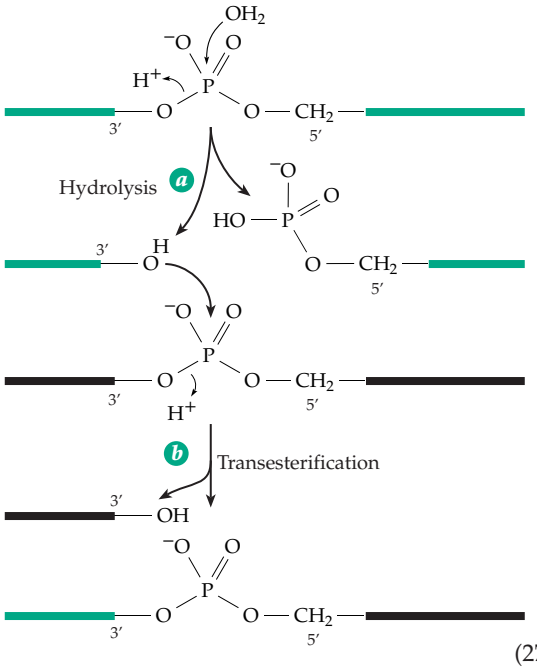
Figure 27-28 (A) Action of the integrase XerC / XerD on a pair of *E. coli dif* sites, containing the central six bp sequences TATACA/ATATGT, which are shown in an antiparallel orientation. In step *a* the active site tyrosine hydroxyl groups (Y-OH) of a pair of XerC subunits carry out transesterification reactions on the 5'-terminal thymidylate residues of the central hexanucleotide sequences to yield 3'-phosphodiester linkages to the XerC tyrosines. In step *b* the cut 5'-ends, containing free thymidylate 5'–OH groups, fold back to form new base pairs, and the strands are resealed in a second transesterification. This generates a Holliday junction, which *isomerizes* (steps *c* + *d*) to an isomeric species that is acted on by a second pair of transesterification steps (*e*) that are catalyzed by protein XerD, again with folding back of the central trinucleotides. After Arciszewska *et al.*[586] (B) The two isoforms of an antiparallel stacked X Holliday junction are shown. These can be reached from the symmetric square form shown in Fig. 5-28 and schematically in (A) by folding into X-conformations in which all base pairs are stacked either in pairs a,b and c,d or a,c and b,d. From Eichman *et al.*[525a]

atypical because it utilizes a pair of integrase subunits rather than just one. However, the basic chemistry (Fig. 27-28) is the same for the entire family.[587,587a] As with the λ integrase (Fig. 27-27) the XerC / XerD complex acts on a pair of identical core sequences that are aligned in an antiparallel fashion. Active sites in all of the integrases contain the conserved amino acid sequence Arg-His-Arg-Tyr. All of these residues are essential for catalysis.[585] Staggered cuts are made sequentially in the core sequences, e.g., at points a and b in Fig. 27-27 and adjacent to the 5'-terminal thymidylate residues in Fig. 27-28A. A transesterification reaction forms 3'-phosphotyrosyl linkages from the cut DNA to the integrase protein. The freed 5'-OH groups on the other cut end fold back and recombine with the bound 3' ends of the second duplex to generate a Holliday junction. In the mechanism proposed in this figure, three nucleotide units are involved in the folding back. The hydrogen bonds of their initial base pairs are broken, and new bonds are formed. This "base swapping" process is the equivalent of a short branch migration and verifies homology of the two recombination sites.

The Holliday junction can isomerize readily between two forms. In one a base-stacked double helix runs between ends a and b in Fig. 27-28B and another runs between ends c and d. The core hexanucleotide sequences lie between the marked XerC and XerD cleavage sites. In the other isomer (lower drawing) one helix has ends a and d and the other b and c. Following the isomerization (steps *c* and *d* in Fig. 27-28A) the XerD active sites act in two transesterification reactions (not shown but analogous to those in steps *a* and *b*) with base swapping of the trinucleotides at the cut ends. This generates the two separate recombinant duplexes. These might be two circular chromosomes or plasmids formed by recombination from a double length chromosome or plasmid. The previously mentioned λ amd Cre recombinases appear to act by closely similar mechanisms.

Tyrosine recombinases of the lambda family also function in eukaryotes. Best known is the **FLP (Flip) recombinase,** which is encoded by the 2-μm plasmid of *Saccharomyces cerevisiae* and is thought to function in amplifying the number of plasmid copies.[265] The 6.3-kbp plasmid contains a unique DNA sequence that lies between two 599-bp repeats in inverted orientation. Embedded in each repeat is an FLP recombination target (**FRT**) sequence, which is recognized by the plasmid recombinase. Each FRT segment includes inverted repeats 13 bp in length with an 8-bp spacer between them. As with other integrase systems the 8-bp spacer or **strand exchange region** is **asymmetric** and establishes the orientation of the recombination sites.[587–590] The role of the recombinase is to invert one of the 599-bp repeats with respect to the other (see Eq. 27-15). This switches replication of the plasmid to a rolling circle pattern.[265]

The resolvase/invertase family and invertible DNA sequences. A second large family of recombinases act by cleaving a target DNA sequence hydrolytically leaving a free 3'-OH end (Eq. 27-14, step *a*). This free end then attacks a phosphodiester linkage in a second strand of DNA, cleaving that strand with an in-line nucleophilic displacement (step *b*). Active sites usually contain a characteristic cluster of aspartate and

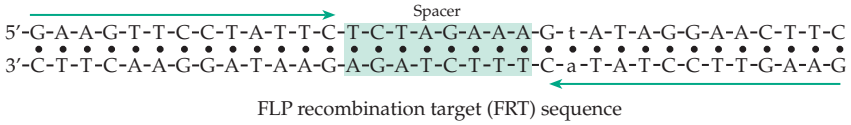


(27-14)

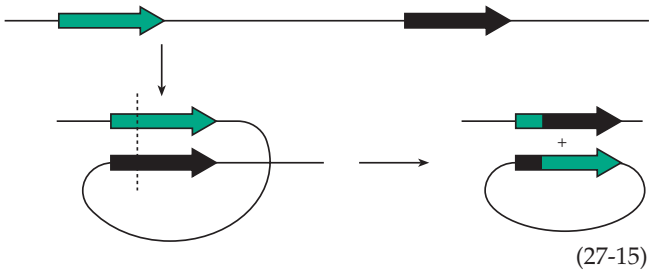
glutamate (DDE) side chains, which probably act together with a metal ion, perhaps as in Fig. 27-13.[591] Enzymes in this resolvase/invertase family act either to resolve cointegrates in transposon action (next section) or to invert DNA sequences.

If recombination occurs within a piece of DNA at two homologous sites such as the *attL* and *attR* sites at the boundaries of the λ prophage, the intervening DNA will be excised as a circular particle (Eq. 27-15). In this instance the two homologous regions must be repeated in the same direction, as is indicated by the arrow structures in Eq. 27-15. If the homologous sequences are oriented in opposite directions, i.e., they are inverted repeats, excision will not occur but the piece of DNA between the repeats will be inverted (Eq. 27-16).

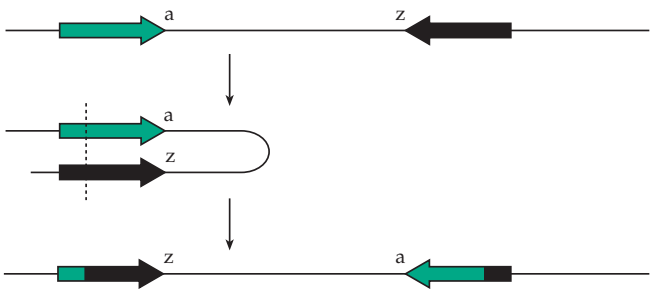
A number of such invertible



FLP recombination target (FRT) sequence



(27-15)



(27-16)

DNA recombination systems are known and are sometimes used to control specific genes.[592] For example, many strains of *Salmonella* have two types of flagella, which are composed of flagellin subunits encoded by genes H1 and H2 respectively.[592–594] On rare occasions an individual bacterium switches from one flagellar "phase" to the other. This occurs by recombination, as in Eq. 27-16, between two 26-recombination sites *hixL* and *hixR*, each of which contains a 14-bp inverted repeat. The 993-bp invertible segment encodes a recombinase gene called *hin* and a promoter, i.e., an mRNA initiation site, which has a specific orientation. In one orientation mRNA is transcribed from a short operon that includes the right inverted repeat IRR, the *H2* flagellin gene and gene *rh*1, which encodes a repressor for flagellin gene *H1*. Consequently, only gene *H2* is expressed. In the other orientation the RNA transcription is in the opposite direction so that neither *H2* nor *rh*1 is expressed. However, *H1*, which is located elsewhere, is expressed freely.

Two other proteins are required for efficient inversion by the Hin recombinase. A dimer of a 98-residue helix–turn–helix DNA binding protein called **Fis** (factor for inversion stimulation), a relative of protein HU,[18b] must bind to an enhancer, a 65-bp DNA segment. Binding of Fis to the enhancer helps to hold the supercoiled DNA and the recombinase in a correct orientation for reaction.[576,595,596] Protein HU is also needed. The same Fis protein binds to an enhancer for a 3-kbp invertible DNA sequence, which controls alternative host preferences for bacteriophage Mu (Fig. 27-29).[576,596] The chemistry of the inversion reaction is related to that of the replicative transposons discussed in Section 4.

Microorganisms sometimes control the synthesis of surface proteins using segments of invertible DNA. The pathogenic bacterium *Campylobacter fetus* utilizes DNA rearrangements to allow one of a large family of surface layer (S-layer) proteins to be formed.[597] The yeast FLP recombinase, mentioned in the preceding section, also inverts the sequence flanked by the 599-bp repeats.[589]

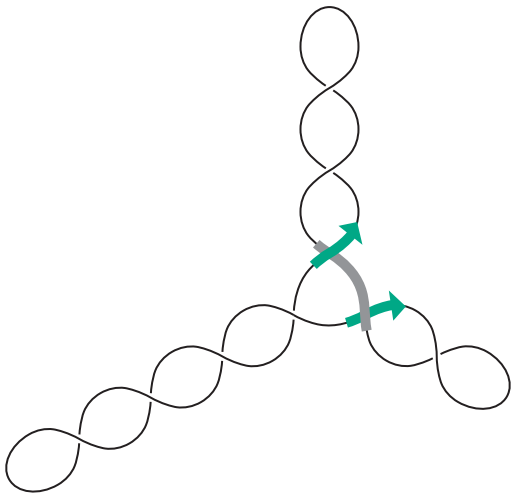
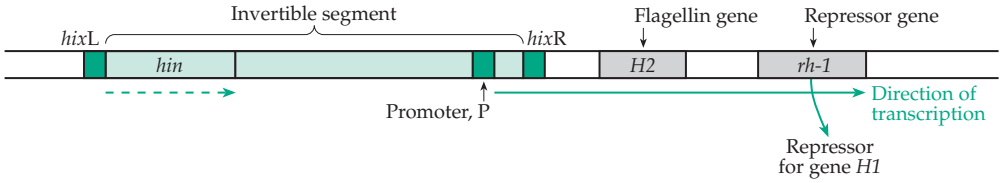


Figure 27-29 Formation of a synaptic complex of a supercoiled circular DNA containing the sites *gix* (green), which pass over and under the enhancer (gray). The recombinase Gin and the enhancer-binding Fis form a synaptic complex with DNA in this form as seen directly by electron microscopy. From Sadowski.[576]

4. Transposons and Insertion Sequences

The first evidence that some genes can move from one location to another within the genome came from studies of *Zea mays* by Barbara McClintock in the late 1940s.[598–602] She concluded that the variegated kernels found in some colored maize were a result of **controlling elements**, which could move from place to place turning on or inhibiting expression of various genes including some of those determining anthocyanin pigment formation. Two of these systems have been



studied especially intensively: the "**activator (Ac)–dissociation (Ds)**" system, discovered by McClintock and the "enhancer (En)–inhibitor (I)" system, discovered by Peterson[599,600] and independently by McClintock.[603] Each contains two genetic elements (segments of DNA) of which Ac and En are autonomous, i.e., they can move by themselves. Ds and I, however, cannot move unless the other element of the pair is also present. Both Ds and En have now been cloned and sequenced.

The general importance of transposable genetic elements was not appreciated by most molecular biologists until about 20 years after McClintock's discoveries. Then several moveable **insertion sequences** (Is elements) were found in enteric bacteria. Like the controlling elements of maize these small (0.8–1.4 kb) sequences can move and insert themselves at many points in the genome, often inactivating genes which they enter.[602–608] The *E. coli* genome contains eight copies of IS1, and five of IS2, as well as several others. Most species of *Shigella* contain more than 40 copies of IS1.[604] Mobile elements similar to those present in maize also exist in archaea.[604a]

Following the discovery of the IS elements it was found that transposable elements named **transposons** could transfer resistance to antibiotics between bacteria. All of these transposable elements have inverted repeat sequences at the ends. For example, IS1 contains the following sequence at both ends but with opposite orientation as if in a

$$\text{CAACTTA}\begin{matrix}\text{C}\\\text{T}\end{matrix}\text{TGAT}$$

palindrome.[605] Some complex transposons have an IS sequence at each end. For example, Tn10 contains IS10, a relative of IS1, at both ends.[608] These provide the characteristic inverted repeat termini. Figure 27-30 shows a schematic drawing of a bacterial drug resistance plasmid containing IS1, IS2, and IS10 as well as transposons Tn3, Tn10, and Tn55. The two IS1 elements surround the large resistance determinant, which can be transferred as a block.[602] To be autonomously mobile a transposon must contain a **transposase** that enables it to be transferred. It usually carries other genes as well and may also contain one or more signals for transcriptional regulation such as promoters.

The chemistry of transposition is more complex than that of simple site-specific recombination. Transposition can occur at many sites in a genome, and no homology with the transposon termini is required. Transposition is accompanied by *duplication of a short sequence of the recipient DNA* exactly at the ends of the transposon. Usually 5, 9, or 11 base pairs are duplicated as in the hypothetical example of Eq. 27-17. This happens because staggered cuts 5, 9, or 11 bp apart in the recipient DNA are made during recombination. These are indicated by the small arrows in Eq. 27-17. Transposons causing 5-bp duplications are Tn3, Tn7,[608a] γδ, phage Mu of *E. coli*, and Ty1 of yeast[609]; IS1, Tn10,[609a] and Tn5 of *E. coli* cause 9-bp duplications.[610–611b] When a transposon of one major group moves to a new location, the original copy remains. In this case transposition involves a combination of replication and site-specific (for the transposon) recombination. As a consequence, a circular DNA molecule containing the transposon will often react with a second circular DNA to form a large circle, a **cointegrate**, which contains two copies of the transposon. However, another group of transposons utilize a "cut-and-paste" mechanism that doesn't require extensive DNA duplication.

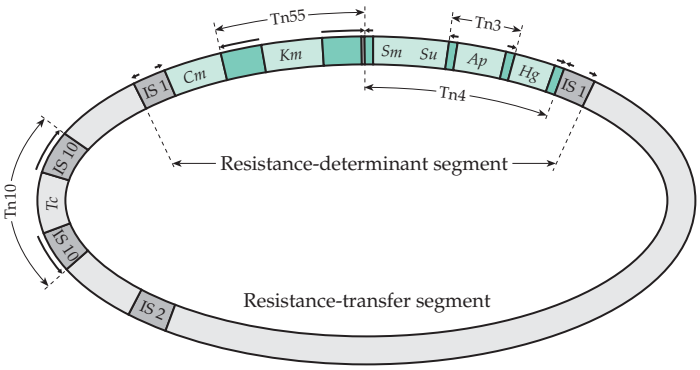
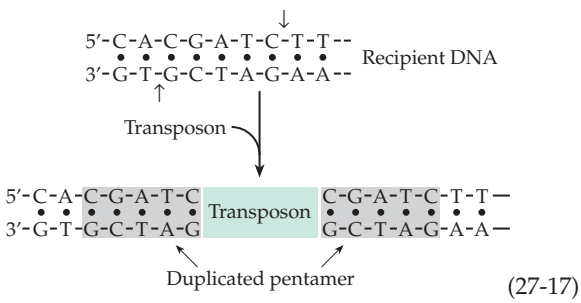


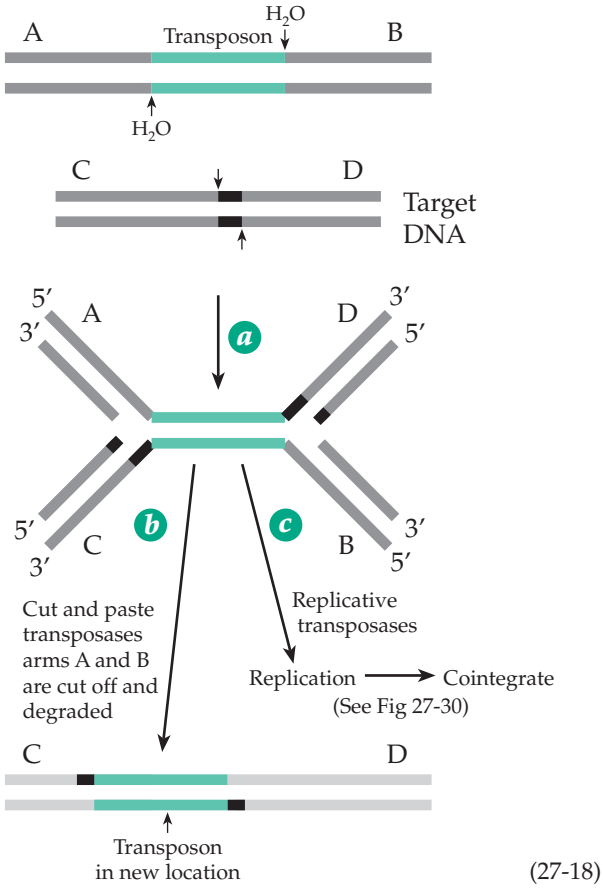
Figure 27-30 Transposons in an antibiotic-resistance plasmid. The plasmid appears to have been formed by the joining of a resistance-determinant segment and a resistance-transfer segment; there are insertion elements (*IS1*) at the junctions, where the two segments sometimes dissociate reversibly. Genes encoding resistance to the antibiotics chloramphenicol (*Cm*), kanamycin (*Km*), streptomycin (*Sm*), sulfonamide (*Su*), and ampicillin (*Ap*) and to mercury (*Hg*) are clustered on the resistance-determinant segment, which consists of multiple transposable elements; inverted-repeat termini are designated by arrows pointing outward from the element. A transposon encoding resistance to tetracycline (*Tc*) is on the resistance-transfer segment. Transposon Tn3 lies within Tn4. Each transposon can be transferred independently. From Cohen and Shapiro.[602]



(27-17)

Cut-and-paste (nonreplicative) transposons. The transposases of *E. coli* Tn5, Tn7, and Tn10 act by hydrolytically cutting both strands of duplex DNA at the transposon ends leaving the phospho groups attached to the 5' cut ends, as is depicted in detail in Eq. 27-14, step *a*. The two 3' ends then carry out transesterification reactions, as in Eq. 27-14, step *b*. These two steps are used to nick both strands of the DNA carrying the transposon and to join them to a target DNA sequence to give a branched intermediate (Eq. 27-18, step *a*). Nonreplicative transposons apparently cut off two arms, e.g., A and B, and heal up the small gaps by repair synthesis, leaving the transposon in a new location between C and D. The gap repair accounts for the duplication of the end sequences of the cut target DNA.[610–612a]

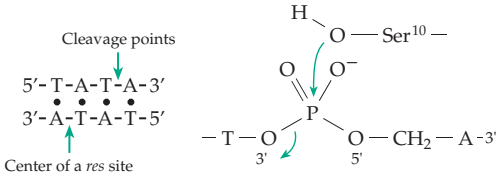
Replicative transposons. In 1979 Shapiro proposed the mechanism illustrated in Fig. 27-31 for replicative transposons. The two inversely repeated segments (green) at the ends of the transposon are aligned with the recipient DNA whose ends are labeled C and D. In fact, the recombining DNA molecules must be supercoiled.[1,613] Staggered cuts are made in the recipient DNA at points *a* and *b*, which are 5, 9, or 11 bp apart, depending upon the specific recombinase. Nicks are also made in the transposon ends. The 3' ends from the transposon are resealed with the 5' ends from the recipient DNA (step *a*) to give a structure that in effect has two replication forks. Replication (step *b*) yields the cointegrate, which contains two copies of the transposon as indicated. In a third step (step *c*) recombination between the two integrated transposons yields a copy of the original transposon-containing donor and the recipient DNA, which now also contains a copy of the transposon.



(27-18)

When a transposon reacts with another part of the same DNA circle there are two possibilities. The piece of DNA lying between the transposon and the recipient site may be excised as a circle containing a copy of the transposon. Alternatively, there will be inversion of that sequence as well as replication of the transposon (compare with Eqs. 27-15, 27-16).

The closely related transposons Tn3 and γδ are understood best. They contain not only a transposase gene but also a **resolvase** gene.[613–616] The transposase carries out the recombination reactions that yield the cointegrates, while the resolvase catalyzes site-specific recombination between the two transposons in the cointegrate to complete the transposition. Several subunits of the resolvases bind the two **resolution (res) sites** of the supercoiled DNA in a parallel orientation with the DNA supercoiled as in Fig. 27-29. However, no enhancer-binding protein is needed, and the two *res* sites must be supercoiled. Purified γδ resolvase uses hydroxyl groups of the serine-10 side chains as the nucleophiles to cleave the DNA by displacement at specific *res* sites to give transient enzyme-bound phosphodiester linkages.[615]



The synaptic complex contains 240 bp of DNA and at least two resolvase dimers. All four DNA chains are cut to give eight ends. Four of these are bound to the serine side chains in phosphodiester linkage. In the second step the freed 3'-OH groups react with the bound ends of the other duplex via a transesterification reaction to form the recombinant chains.

The resolvases act on supercoiled cointegrated DNA molecules that contain two directly repeated res sites to produce two singly linked circles (which are still supercoiled) each containing one res site as shown in Fig. 27-32. The two *res* (resolution) sites within the transposons are aligned, the open circle of DNA shown at the upper left being folded as shown in the lower part of the drawing. The DNA substrate is not knotted. However, after recombination it is catentated and will require action of a topoisomerase to separate

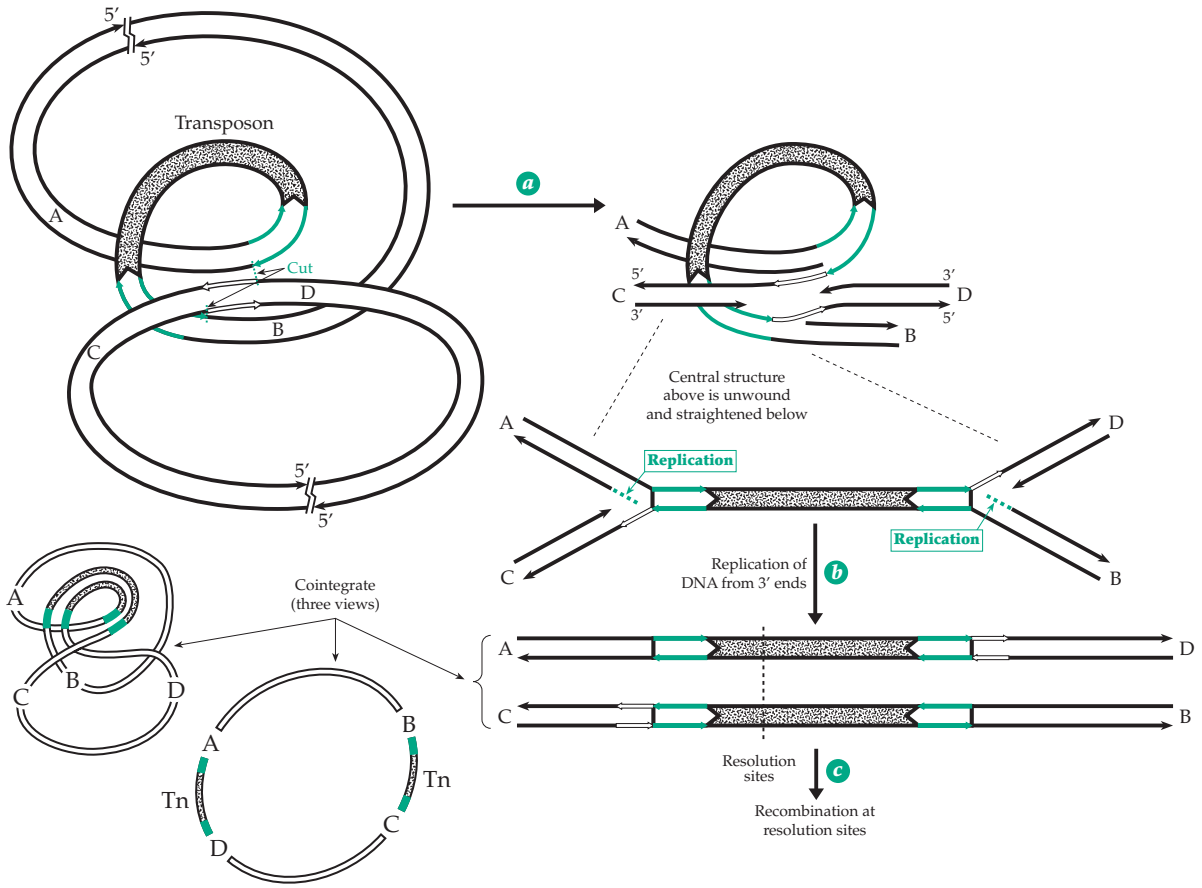


Figure 27-31 Scheme for integration of a transposon (stippled duplex) present in a piece of DNA with ends A and B into another piece of DNA with ends C and D and containing a suitable recognition sequence (open bars). Inversely repeated sequences in the transposon are shown as solid bars with a direction arrowhead. Arrowheads point toward 5′ strand ends. Cleavage and rejoining at points a,a and b,b yield an intermediate with two replication forks. Replication through the transposon yields one unchanged DNA segment with ends A and B and a transposon inserted into the other DNA segment. If A is continuous with B, a cointegrate structure is formed. See Cohen and Shapiro.[602]

the two products. Occasionally additional recombination events occur, perhaps processively along interwound double helices. This produces various knotted products.[617] An electron micrograph of one of these is shown in Fig. 5-17.

The temperate bacteriophage Mu. The most efficient transposon known is the 37-kb genome of the **mutator bacteriophage Mu**.[265,618–621] Once it becomes integrated into a bacterial chromosome, it replicates by repeated rounds of transposition within the host bacterium. During the lytic cycle some of the replicating DNA is excised as extrachromosomal circles of various sizes, which are packaged into virus particles by a headfull mechanism. The circles contain copies of some host DNA, but this is left behind when a virus particle infects a new cell.[265] The phage Mu DNA is integrated into host DNA by a cut-and-paste mechanism in a transpososome that contains a tetramer of the virally encoded transposase (MuA protein) bound to a supercoiled DNA. The transpososome resembles that in Fig. 27-29 and contains cleavage sites at the ends of the transposon and also an enhancer sequence.[620,621] Several steps involving conformational alterations occur in the transpososome. One is an ATP-dependent action of a second Mu-encoded protein MuB.[622] During the lytic cycle replicative transposition predominates.

Some other transposons. Transposons have a variety of biological functions. For example, haploid cells of the yeast *S. cerevisiae* exist as one of two mating types *a* or α. The mating type is established by transposition of one of two "cassettes" of genes from two different "silent" locations to a location from which they can be expressed.[623,624] See Chapter 28.

One of the best known eukaryotic transposons is the P element of *Drosophila*, which transposes only within the germ line cells of developing embryos, somatic cells being unaffected.[265,625,626] It belongs to

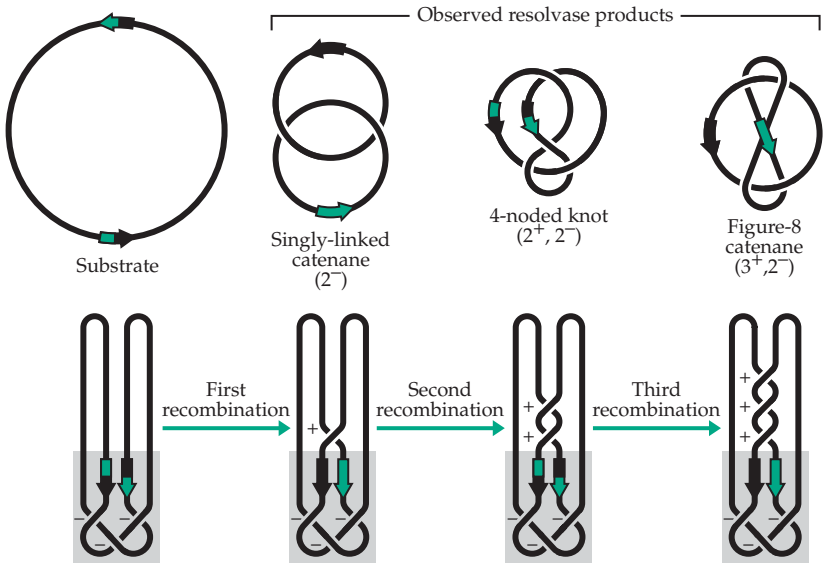


Figure 27-32 Scheme for resolution of an unknotted cointegrate molecule by a resolvase that cuts the transposons at resolution (*res*) sites and recombines them. The resolvase may act only once or repeatedly as shown. In the upper row, the duplex DNA substrate and products are represented in standard topological form as they might appear after nicking. In the lower row the DNAs are depicted as folded forms bound to the resolvase with the two directly repeated *res* sites (thick arrows) dividing the substrate into two domains (thick and thin regions). The substrate at synapsis has three (–) supercoils that entail crossing of the two domains. Successive rounds of recombination, each introducing a single (+) interdomainal node (see Fig. 5-17), are drawn in the lower row. Bound resolvase maintains the three synaptic supercoils. After dissociation from the resolvase at any stage, the product supercoil nodes either cancel with ones of opposite sign or are removed by subsequent nicking. The node composition is indicated in parentheses. From Wasserman *et al.*[617]

the same family as Barbara McClintock's Ac element of maize and the *Tc*1 family of nematodes.[627] P elements use a nonreplicative cut-and-paste method of transposition. The 87-kDa transposase protein requires GTP and Mg^{2+} for activity.[628] It was only in the past few decades that P elements have been found in *D. melanogaster*. They may have entered this fruit fly from another species, possibly transferred by a mite.[627,629]

A second *Drosophila* transposon called **mariner**[630] typifies the *mariner* / *Tc*1 transposon superfamily, which also contains members from nematodes,[631] other invertebrates, fishes,[632] amphibia,[633] and possibly human beings.[634] These transposons encode a transposase containing a D, D, D or D, D, E motif[630] but no other proteins. They contain short ~30-bp terminal inverted repeats and become inserted into host TA sequences.[631] Movement of some repetitive sequences of the LINE[635] and SINE[636] families within the human genome may be assisted by *mariner* transposons.[637]

The maize transposon *ac* is widely used as a means of inactivating genes and placing a "tag" that can be used to map the gene and to permit it to be cloned and sequenced.[598,638] Although initially of use only in maize the method has been extended to other plants,[639–641] and genetically engineered transposons have allowed it to be utilized in animals.[642]

A different kind of transposition controls self-sterility in maize. The cause of self-sterility in one strain has been traced to the presence of two linear episomes called S-1 (614 kb) and S-2 (514 kb) within mitochondria.[643] These have inverted terminal 208 bp repeats. On rare occasions they recombine with the circular mtDNA converting it to a linear form with the episomes covalently linked to one end. The change is accompanied by reversion to fertility.

As described in Section B,1, mammalian DNA contains many **retrotransposons** (retroposons) that lie within short direct repeats characteristic of transposons. However, they contain a poly(A) tail at the 3' end, an indication of their relationship to RNA transcripts, and are discussed in Chapter 28.

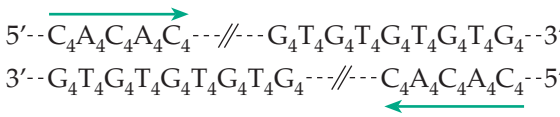
5. Other Causes of Genetic Recombination

Because their integration sites in host DNA do not depend upon homology with the transposon ends, transposition is sometimes called "**illegitimate recombination.**" Although no homology is required, there are prefered sites for integration. For example, Tn10 is transposed most readily into certain "hot spots" in the *Salmonella* DNA among which is the sequence 5'-GC^{m5}CAGGC.[608] Illegitimate recombination can also be induced by other processes that involve DNA chain cleavage, e.g., by topoisomerases.[644] Whenever a DNA chain breaks, it must be repaired, a process that often also involves recombination. Recombination is often observed to occur between direct repeat sequences, which are a major cause of instability in the genome.[645]

Under some circumstances selected segments of the genome are **amplified** by repeated replication of a gene or genes.[646,647] Amplication of specific genes occurs in viruses,[648,649] in bacteria where it may provide

for adaptation to conditions of stress,[650] and in eukaryotes.[651] In oocytes of amphibia, such as *Xenopus*, excess DNA accumulates around the nucleoli and later breaks up to form 1000 or more separate nucleoli. As many as 3000 copies of the rDNA (which forms a distinct satellite band upon centrifugation) may be present. Much of this DNA exists as extrachromosomal rings containing 1–20 rDNA units. Using these genes as many as 10^{12} ribosomes per oocyte are synthesized.

Tetrahymena contains only one set of rRNA genes per haploid genome in its diploid **micronucleus**. Following sexual conjugation the chromosomes of the micronucleus undergo multiple replications to form polytene chromosomes containing thousands of copies. However, about 5% of the resulting DNA is excised as linear pieces with characteristic inverted repeat sequences and 3' single-stranded tails:



These tails are joined by a protein to form the circles, which are segregated in the **macronucleus**. The other 95% of the amplified DNA is degraded. At the next meiosis the macronucleus is discarded entirely, and a new one is formed at the next diploid stage.

Toxic drugs often cause cells to amplify genes that help resist the drug.[647] This can be a major problem in the chemotherapy of cancer. For example, a culture of human leukemia cells grown in the presence of increasing concentrations of methotrexate increased its level of dihydrofolate reductase 240-fold.[652] The cause is an increase in the number of copies of a chromosomal region containing the gene.[653] Cancer cells tend to amplify oncogenes such as c-*myc*.[654]

There are several mechanisms for gene amplification. The formation and breakup of polytene chromosomes in *Tetrahymena* is one. The circular copies of rDNA in *Xenopus* and many other species are generated by a rolling circle mechanism similar to that in Eq. 27-7. If each circle excised from the original chromosomes contains an origin of replication, many copies can be formed. Another possible mechanism is replication of a local region of DNA several times followed by excision of pieces of the DNA. A mechanism that may give rise to homogeneously staining regions is unequal crossing-over (Section 2) repeated several times within a gene cluster. Transposition can cause excision of DNA in a circular form, which can be amplified by a rolling-circle mechanism.

E. Damage and Repair of DNA

A characteristic of living things is their high degree of mutability. Harmful mutations take a toll of human life at an early age, and the very high incidence of cancer in older persons is largely a result of the accumulation of somatic mutations. Mutations are a major factor in aging and are continuously introducing new genetic defects into the population. Mutations can be described as **base substitutions**, **deletions**, or **additions**. Base substitution mutations are classified as **transitions**, in which a pyrimidine in one strand is replaced by a different pyrimidine. In the complementary strand a purine is replaced by the other purine, e.g.,

$$C\cdot G \rightarrow T*G \xrightarrow{\text{Replication}} T\cdot A + C\cdot G \qquad (27\text{-}19)$$

$$A\cdot T \rightarrow \underset{\text{Mispair}}{G*T} \longrightarrow \underset{\text{Mutation}}{G\cdot C} + A\cdot T \qquad (27\text{-}20)$$

In a *transversion* a purine in one chain is replaced by a pyrimidine, while the pyrimidine in the complementary chain is replaced by a purine:

$$C\cdot G \rightarrow A*G \longrightarrow A\cdot T + C\cdot G \qquad (27\text{-}21)$$

Here the central asterisks designate mispairs and the green shade marks mismatched bases and also the resulting mutant base pair formed in the next replication cycle.

1. Causes of Mutations

DNA can be damaged in many ways.[655] Spontaneous hydrolysis of the glycosidic bonds between nucleic acid bases and the deoxyribose to which it is connected cause the loss of ~10^5 purines and pyrimidine rings per day from the DNA in a mammalian cell.[656] About 100 residues per day of cytosine are deaminated by such agents as nitrite or bisulfite (Chapter 5, Section H,3) to form uracil. Like thymine, uracil will pair with adenine causing a C • G → T • G transition mutation as in Eq. 27-19. A few adenines per day are also deaminated to form hypoxanthine. Oxidized bases are formed by attack of HO• radicals and other species of reduced oxygen.[657–659] Alkylating agents from the external environment as well as *S*-adenosylmethionine carry out slow, nonspecific alkylation of purine and pyrimidine bases. Polycyclic aromatic hydrocarbons and other carcinogens are converted to metabolites that alkylate DNA, and alkylated bases often mispair during replication.[660,661] Ultraviolet light induces formation of photohydrates (Eq. 23-25),[662] pyrimidine dimers (Eq. 23-26), and other photochemical products.[663] X-rays and gamma rays cleave nucleic acid bases and break chromosomes.[663a]

Natural radioactivity has a measurable effect.[663b] Mistakes caused by mispairing[664] or misalignment[665] are made during replication and recombination and even during repair of DNA. Errors in replication are especially numerous in highly repetitive DNA sequences. Some errors probably arise as a result of tautomerization and others from incorporation of uracil in place of thymine. Thus, to keep its DNA in repair a cell must continuously deal with missing bases, wrong bases, altered bases or sugars, pyrimidine dimers and other crosslinkages, deletions, and insertions.

2. Fidelity of Replication

During DNA replication in *E. coli* only one mistake is made on the average during polymerization of 10^9–10^{10} nucleotides.[666–668] The rate varies among different sites in the genome.[668a] In eukaryotes the error rate may be only 1 in 10^{-10} base pair or less per generation.[669,669a] Early workers often attributed the specificity in base pairing and the resultant high precision of replication entirely to the strength of the two or three hydrogen bonds formed together with the stabilization provided by the adjacent helix. However, the Gibbs energy of formation of the base pairs is small (Chapter 5), and the additional energy of binding to the end of an existing helix is insufficient to account for the specificity of pairing.[670,671] Thus, according to Eq. 6-30 a difference in ΔG of binding between the correct nucleotide and an incorrect one of 11 kJ / mol would give an error rate of 1 in 100.

The role of polymerases. The polymerase enzymes play a major role in ensuring correct pairing during replication, transcription, and protein synthesis. RNA and DNA polymerases are large molecules. The binding site on the enzyme can completely surround the double helix. Water will be excluded as the enzyme folds around the base pairs. This may have the effect of greatly increasing differences in the Gibbs energies of binding and thereby enhancing selectivity.[671] It has traditionally been assumed that formation of proper hydrogen bonds is essential for the base selection. A hypothetical active site is portrayed in Fig. 27-33. A guanine ring of the template strand of DNA is portrayed at the point where the complementary DNA strand is growing from the 3' end. The proper nucleoside triphosphate must be fitted in to form the correct GC base pair before the displacement reaction takes place to link the new nucleotide unit to the growing chain. Let us suppose that the enzyme possesses binding sites for the deoxyribose unit of the template nucleotide and for the sugar unit of the incoming nucleoside triphosphate and that the two binding sites are held at a fixed distance one from the other. As indicated in Fig. 27-33, some group H-Y might also be present at each binding site to hydrogen bond to the nitrogen or oxygen indicated by the heavy green arrows. All four of the bases could form hydrogen bonds of this type in the same position. Hydrophobic interactions could provide additional stabilization. With such an arrangement the correct nucleoside triphosphate could be selected no matter which one of the four bases occupied the binding site on the left side of the figure. (The outlines of the thymine and adenine rings have been drawn in with dotted and dashed lines, respectively.) If a purine is present on the left side, as is shown in the drawing, there is room on the right side only for a pyrimidine ring. Thus, A and G are excluded, and the choice is only between C and U (or T). However, U will be excluded because the dipoles needed to form the hydrogen bonds point in the wrong direction. These dipolar groups are hydrated in solution. They are unlikely to give up their associated water molecules unless hydrogen bonds can be formed within the base pair. Not only would a molecule of U (or T) be unable to form the stabilizing hydrogen bonds within the vacant site but also the electrostatic repulsion of the like ends of the dipoles would tend to prevent association. This would lower the affinity of the polymerase for mispaired bases. Verification that the proper base pair has been formed could be accomplished by using its tautomeric properties to sense an electronic

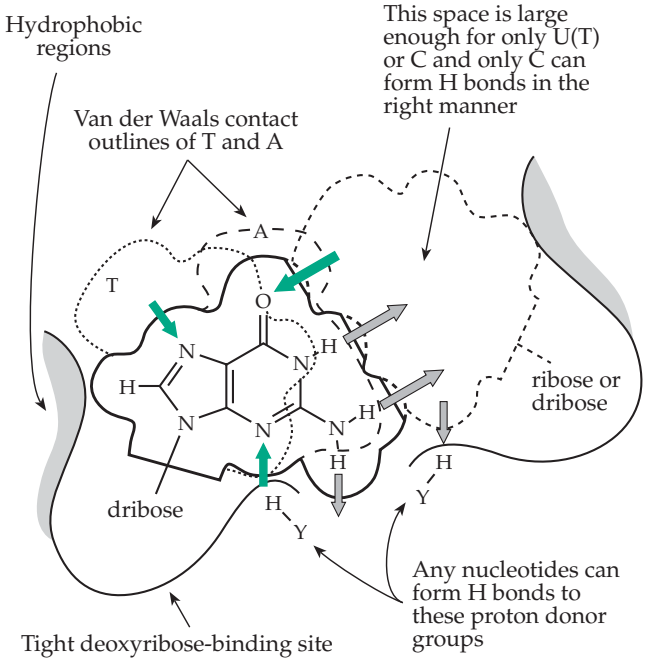
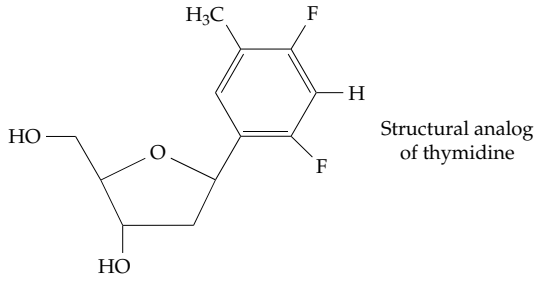


Figure 27-33 Selecting the right nucleotide for the next unit in a growing RNA or DNA chain. A deoxyguanosine unit of the template chain is shown bound to a hypothetical site of a DNA polymerase.

displacement through the hydrogen-bonded network.[672,673] Considerable experimental evidence suggests, however, that hydrogen bonding may not be important to the selectivity of DNA polymerases.[666,674–676] For example, the triphosphate of the following analog of thymidine is efficiently incorporated by *E. coli* DNA polymerase I into DNA opposite the thymine base-pair partner adenine.[675]



Structural analog of thymidine

Similar results have been obtained for a variety of other analogs with poor hydrogen bonding characteristics. Apparently *shape* is most important.[666,674,676,676a] Binding strengths in transition state structures must also be considered.[676b]

Editing (proofreading). As is discussed in Section C,2, base-pairing is checked after a new monomer is added at the 3' end of a growing polynucleotide chain, as well as before polymerization occurs. If the wrong base pair has been formed, the newly created linkage is hydrolyzed, and the incorrect nucleotide is released. It has been estimated[677] that in *E. coli* the error rate for DNA polymerase III holoenzyme is $\sim 1 \times 10^{-7}$ per base pair of which proofreading by the ε subunit may provide $\sim 10^{-2}$. Additional mismatch repair reduces the error rate 200- to 300-fold to give the overall error rate of $<10^{-9}$. Fidelity of replication, which seems to be higher on the lagging strand than on the leading strand,[678] may also be improved by operation of other proteins, such as the sliding clamp, that are part of the replication apparatus.[679] Replicative proteins also appear to be designed to minimize **frameshift mutations** that could arise by slippage of the template versus the replicated strand in long runs of dA•dT pairs.[680]

3. Repair of Damaged DNA

A final check of the fidelity of replication is made after a new strand has been formed. Mismatched base pairs are identified, and the incorrect nucleotides are cut out and replaced by correct ones.[655,670,681–683] Some of the thymine dimers created by the action of light are also repaired photochemically by photolyases (see Chapter 23). **Photoreactivation** was the first DNA repair process recognized.[684] However, most thymine dimers in human DNA must be excised and replaced. Loops, gaps, and double-stranded breaks are also sensed, and appropriate corrections are made. If necessary the sequence of a badly damaged segment of DNA may be copied from the DNA in a sister chromatid before mitosis is completed or from the homologous chromosome.[685]

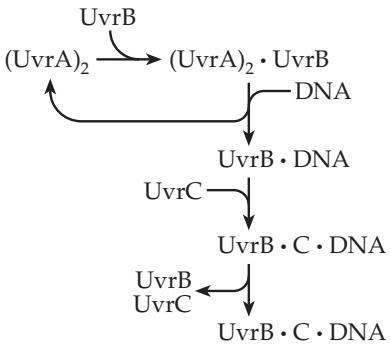
Much of our understanding of DNA repair comes from investigations of mutant strains of *E. coli*, which display an elevated incidence of mutations. On this basis, the "mutator" genes *dam, mutD, mutH*, *mutL*, *mutS*, *mutU* (also called *uvrD*), and *mutY* were implicated in DNA repair.[677,686] An additional series of genes, *uvrA*, *B*, *C*, and *D*, were identified by their participation in resistance to ultraviolet radiation damage and were also recognized as being involved in DNA repair. Many corresponding genes were found later in the yeast, *Drosophila*, and human genomes. The study of human DNA repair has also been facilitated by the recognition of a group of human inherited defects (Box 27-A). There are 130 known human DNA repair genes.[686a]

Methyl-directed and other mismatch repair. In *E. coli* the methylase encoded by gene *dam* (Fig. 29-4), within ~7 min of replication, methylates adenine at N-6 in the sequence GATC, which occurs at many locations.[670,686,687] This methylation provides a label for the template strand in a newly replicated duplex. The GATC sites in the newly synthesized strand will not yet be methylated, and repair enzymes recognize it as the strand on which to carry out excision repair. Another system, which may function during recombination, acts on fully *dam* methylated DNA.[655] Since eukaryotes do not use methylation of GATC sites to distinguish the old template and newly replicated DNA strands, other mechanisms must operate.[669]

Ten proteins are required for methyl-directed mismatch correction. Of the required *mut* genes, *mut*D (also called *dnaQ*) was found to be the structural gene for the proofreading subunit ε of DNA polymerase III (Table 27-2).[688] *MutH*, *L*, and *S* as well as ATP are also needed, as is a helicase, an exonuclease, DNA pol III, and a ligase. Homodimers of proteins MutS and MutL preferentially bind to DNA with mismatched base pairs such as Pur-Pur (e.g., A•G) or Pyr•Pyr (less well recognized). Some of these mismatched pairs may contain modified bases. After **recognition** by the MutS•MutL complex (see banner, p. 1527)[688a–c] the endonuclease MutH is activated and cuts the DNA chain at the nearest unmethylated GATC sequence.[689] An exonuclease then degrades the chain past the mismatched base leaving a gap that must be filled by the action of DNA polymerase and DNA ligase. This **long-patch mismatch repair** (**MMR**) is utilized also by eukaryotes,[688a,b,690–692a] using proteins homologous to *E. coli* proteins MutS and MutL. However, methyla-

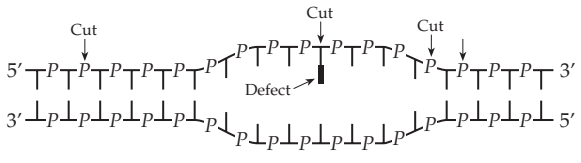
tion is not involved. In both yeast and human beings there are at least six MutS homologs, some of which have functions other than repair. For example, in yeast MutS- and MutL-related proteins are essential for normal levels of meiotic crossing-over.[692] Defects in human mismatch proteins may cause hereditary nonpolyposis colorectal cancer (Box 27-A).[692b]

Excision repair. The *E. coli* mismatch repair is a type of excision repair. However, a different **nucleotide excision repair** system (**NER**) is utilized by all organisms from bacteria to human to remove a variety of defects. These include thymine dimers, photohydrates, oxidized bases, adducts of cisplatin (Box 5-B), mutagens derived from polycyclic aromatic compounds,[683] and poorly recognized C•C mismatched pairs.[692] In *E. coli* this excision repair process depends upon proteins encoded by genes ***UvrA, B, C,*** and ***D*** and also DNA polymerase I and DNA ligase.[693–695a] A dimer of protein UvrA forms a complex with helicase UvrB (Eq. 27-22).[696,696a]



(27-22)

The helicase, driven by ATP hydrolysis, may move along the DNA chain with its associated UvrA protein in search of defects.[696] When one is located UvrB binds tightly.[697] UvrA then dissociates, and the nuclease UvrC binds and cleaves the DNA chain in two places as in the following scheme. One is at the fourth or fifth phosphodiester linkage in the 3' direction from the defect. The other is at the eighth phosphodiester on the 5' side.[698] The resulting gap is filled by a DNA polymerase and DNA ligase. Another helicase, encoded by *UvrD*, is also required.[699]



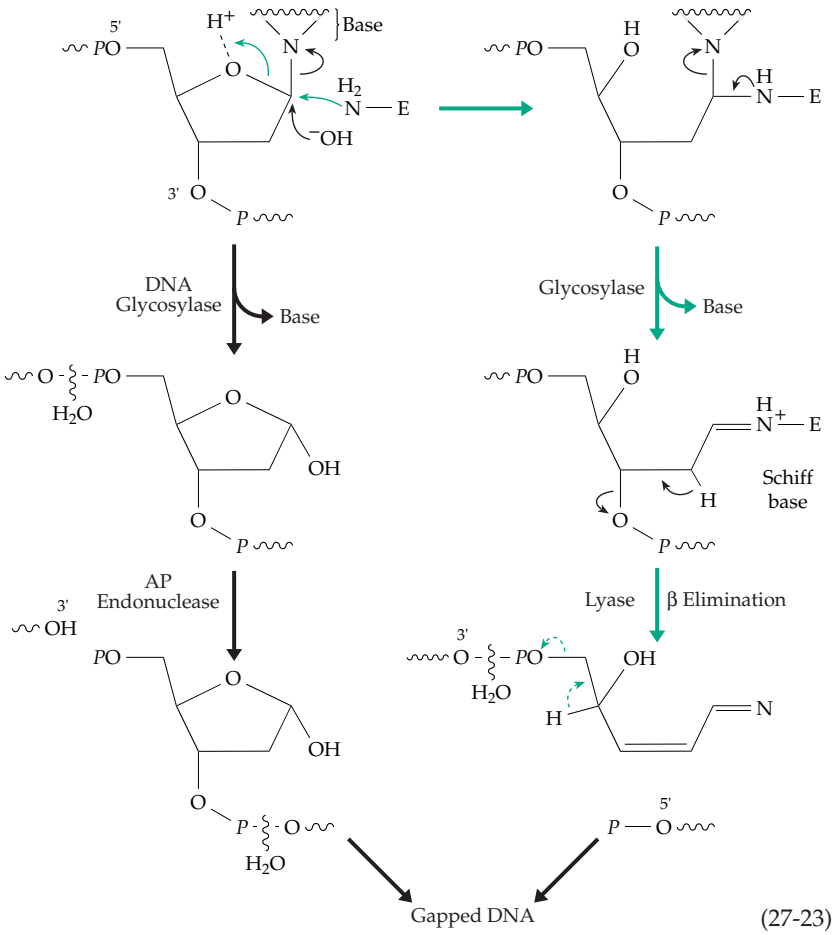
UvrBC preincision complex

Both yeast and human cells have similar but more complex systems of nucleotide excision repair.[700–703] The dual incision steps require at least six components. Several of these (XPA, C, F, G) have been found defective in various forms of the inherited disease **xeroderma pigmentosum** (Box 27-A), in which there is a high incidence of UV-induced skin cancer.[700,703a] Replication protein A (Section C,10), a three-subunit ssDNA binding protein, is also essential.[704] In eukaryotic excision repair the DNA is cleaved at the same position as in the bacterial preincision complex (preceding scheme), on the 3' side of the defect but further out on the 5' side, the excised polynucleotide being ~24–32 nucleotides in length. Repair of the resulting gap in the DNA duplex is accomplished by synthesis mediated by DNA polymerase δ or ε and action of a DNA ligase.

One of the lesions removed by nucleotide excision is the thymine photodimer.[705] In the fission yeast *S. pombe* an **alternative excision repair** system, specialized for removal of thymine dimers and 6–4 photoproducts (Chapter 23), produces two-nucleotide gaps with 3'-OH and 5'-phospho-group ends.[702,705a] Alternative NER pathways are also employed by bacteria.[705b]

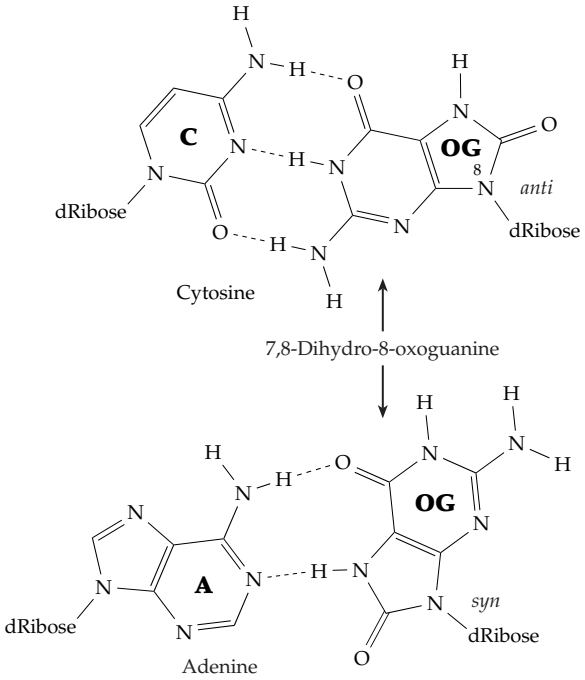
Base excision repair (BER). The *N*-glycosyl linkages of the purine and pyrimidine bases to the deoxyribose residues of the sugar–phosphate backbone of DNA are subject to spontaneous hydrolysis, one important source of damage to DNA. Similar hydrolytic reactions are catalyzed by **DNA glycosylases**, which remove many mismatched or damaged bases.[706–708a] At least seven enzymes of this type are present in cells of *E. coli*. One of them, a **uracil-DNA glycosylase**, hydrolyzes the glycosyl linkage wherever uracil has been incorporated accidentally in place of thymine or has been produced by deamination of cytosine.[706,709–711d] Acting via a nucleophilic displacement mechanism (Eq. 27-23, black arrows), it removes the uracil, leaving the DNA backbone intact but with an **apyrimidinic site** in one chain. The resulting apyrimidinic (apurinic sites) are recognized and cleaved by **apurinic/apyrimidinic DNA endonucleases** (AP nucleases).[708,712–714] These enzymes cut the DNA backbone, some leaving a 3'-phosphate group and a 5'-OH terminus and others a 3'-OH and 5'-phosphate. The resulting gap can be filled, for example by the action of DNA polymerase I and a ligase. To prevent the eliminated uracil from being converted to dUTP and reincorporated into DNA a **deoxyUTPase**, essential to both *E. coli* and yeast, hydrolyzes dUTP to dUMP and inorganic phosphate, decreasing the concentration of dUTP.[715] Other enzymes, known as **DNA glycosylase/AP lyases,**[706,716–718] use an amino group of an enzyme side chain as a nucleophile in a ring-opening reaction that is followed by β elimination of the nucleotide base with formation of a Schiff base intermediate (green arrow, Eq. 27-23). An example is the bacteriophage T4

enzyme **T4 endonuclease V**.[706,719] Others include the *E. coli* **endonuclease III**, encoded by the **Nth** gene, its yeast homolog **Nth-Spo**,[717–718] and human DNA pol ι (iota).[718a] After the Schiff base is formed, the DNA backbone is cut on the 3' side of the abasic site by a second β elimination reaction. A δ elimination (green dashed arrows in Eq. 27-23) may also occur, cleaving the DNA chain on the 5' side of the lesion. Alternatively, a hydrolytic cleavage is possible as is also indicated in Eq. 27-23. Either type of DNA glycosylase forms a single-nucleotide gap or a gap missing just a few nucleotides. This gap can also be filled by polymerase and ligase action.

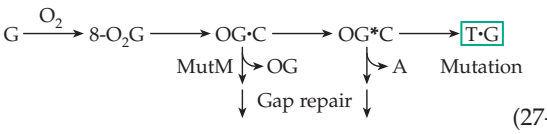


(27-23)

Both NER and BER forms of excision repair remove a great variety of defects, many of which are a result of oxidative damage.[657,720] Most prominent among these is **7,8-dihydro-8-oxoguanine** (8-OG), which is able to base pair with either cytosine (with normal Watson–Crick hydrogen bonding) or with adenine, which will yield a purine–purine mismatch and a C•G → A•T transversion mutation (Eq. 27-24), a frequent mutation in human cancers.[721,722]



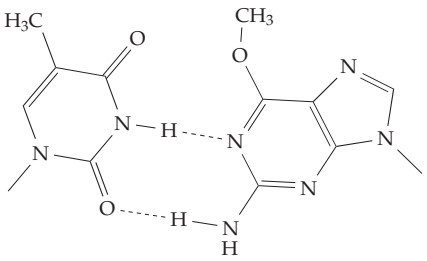
In *E. coli* three enzymes protect against 8-OG DNA mismatches. A glycosyltransferase encoded by the *mutM* (or *fpg*) gene removes 8-OG from DNA.[723,724] Because some A*OG base pairs may remain, a second glycosyltransferase (**MutY**) removes adenine from them.[723,725–725b] Interestingly, MutY contains an Fe_4S_4 cluster, essential to its function.[725,726] Both yeast and human cells have a corresponding enzyme.[727,728] The third *E. coli* enzyme **(MutT)** is a nucleoside triphosphate pyrophosphohydrolase, similar to the previously mentioned dUTPase. It preferentially hydrolyzes free 8-oxo-dGTP, preventing its incorporation into DNA.[729]



(27-24)

Other DNA glycosylases remove thymine rings that have been converted to saturated or fragmented forms by oxidizing agents or ionizing radiation.[720,730,731] Among these are 5,6-hydrated thymine and urea, which are still attached in ribosyl linkage. Purines such as hypoxanthine and 3-methyladenine can all be removed by glycosylases.

Reactions of alkylated bases. The O^6 alkylation of guanine rings in DNA is highly mutagenic, presumably because mispairing with thymine will cause C•G → T•A mutations.



Thymine–O^6-methylguanine pair

Both bacteria and higher eukaryotes have an enzyme-like **O^6-methylguanine–DNA methyltransferase,** whose synthesis may be induced by culture of cells in the presence of alkylating agents. This 354-residue Zn^{2+}-containing **Ada protein** acts as both acceptor and catalyst for transfer of the methyl group off from the O^6-methylguanine onto the sulfur atom of a cysteine side chain in the protein.[732–734b] The presence of alkylating agents also induces synthesis of DNA glycosylases that remove 3-methyladenine and other alkylated bases.[735,736] 5-Methylcytosine, present in CpG islands in human DNA, will occasionally be hydrolyzed to thymine, giving T∗G mismatches. The mismatched thymines are removed in cells of *E. coli* by a specialized endonuclease (**vsr** gene),[737] which hydrolyzes the phosphodiester linkage preceding the mismatch. In eukaryotic cells a **thymine–DNA glycosylase**[738,739] accomplishes the removal.

Repair of double-strand breaks. Following replication both in bacteria and in eukaryotes, recombinational exchanges may occur between sister chromatid duplexes or between homologous pairs of chromosomes.[740] A newly replicated chain may have gaps because of defects in the template strand from which it was copied. Copying from a sister duplex may permit assembly of a correct chromosome sequence and survival for the cell.[551a] Recombinational repair is also a major mechanism for preventing loss of genetic information from **double-strand breaks**. Such breaks are induced by ionizing radiation or chemical damage. They are often created at replication forks, stalling DNA synthesis with potentially disastrous consequences to the cell. Repair of these breaks, which is required for completion of replication, depends upon a large complex of replication and recombination proteins.[741–741b] In *E. coli* this includes protein RecF.[741,742] Repair of double-strand breaks by homologous recombination may be the most frequently employed type of DNA repair in bacteria, yeasts such as *S. cerevisiae*[530d] and also in mammals and other organisms.[742a–d]

A second repair pathway, called **nonhomologous end-joining** (**NHEJ**), is also utilized and may be relatively more important in higher organisms.[743–746b] The NHEJ mechanism repairs double-strand breaks caused by ionizing radiation and is also employed in specialized genomic rearrangements of developing B and T cells of the immune system. This V(D)J recombination (Chapter 31) is used to create the huge array of antibodies and antigen receptors required for immunological recognition and protection.[745,746a,747,748] NHEJ doesn't depend upon templates but simply rejoins duplex ends. Errors can be made if the ends have been damaged or if the wrong ends are joined. Several special proteins are required. One is a DNA-activated protein kinase (**DNA-PK**)[749] and a heterodimeric **end-binding protein** with subunits **KU70** and **KU80**.[745,750,750a] The serine / threonine protein kinase ATM (Box 27-A) is also activated and phosphorylates protein p53 in response to γ-irradiation of cells.[750] This is thought to be part of a signaling pathway for control of the cell cycle following DNA damage. Mitosis may be delayed while repair is completed. Alternatively, the damaged cell may be killed by apoptosis.

Subjects of current interest are the mechanisms by which completion of DNA synthesis is signaled at the various checkpoints in the cell cycle.[750b,c] These include points in addition to those marked in Fig. 11-15. At every point signals must accumulate to indicate that replication must be delayed to allow repair or that the cell must be allowed to die by apoptosis. Failure to accomplish repair may lead to cancer.[750d] Even one double-strand break will prevent the completion of mitosis.[750b] Among other problems faced by a dividing cell are slow replication and stalling at replication forks.[750c] A range of chemical alterations in chromatin and other proteins are associated with DNA damage and repair.[750b,e–h] Among these are conjugation of PCNA with ubiquitin and SUMO (Fig. 27-12D).[750i]

The SOS response and translesion repair. If cells of bacteria or eukaryotes are heavily damaged by UV, X-ray irradiation, or mutagenic chemicals, an emergency or **SOS response** is initiated.[711a,741,751–753a] In *E. coli* the two proteins specified by genes *lexA* and *recA* initiate the response. The *lexA* protein is a repressor that prevents transcription of a group of SOS genes (see Chapter 28). It is thought that some product from damaged DNA activates the RecA protease activity. The activated RecA protein then cleaves the lexA protein allowing transcription of the SOS genes. The SOS response is transient but complex. It includes increased recombinational activity, alterations in replication initiation, inhibition of nucleases, and induction of an **error-prone DNA synthesis**. The cell now

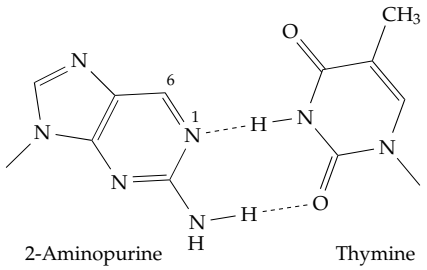
replicates DNA more rapidly than normal but with an increased frequency of errors.[754,754a] For example, it will bypass thymine dimers and other defects, and it will continue the DNA synthesis even though incorrect bases have been put into the chain opposite the thymine dimer. Later the errors may be corrected by recombinational repair or by photoreactivation. In *E. coli* this translesional repair depends upon DNA polymerase III, the RecA protein, and two proteins encoded by genes *umuC* and *umuD*. These genes were recognized as providing resistance to mutations induced by UV radiation.[753–754b] Study of similar genes in yeast led to the discovery of two new and unusual DNA polymerases essential for translesional repair. DNA polymerase ζ (zeta), encoded by gene *REV3* and *REV7* subunits, bypasses abasic sites by inserting a dCMP residue into the growing DNA chain. Because C may be the wrong base the process is "error-prone." Polymerase η (eta) bypasses thymine photodimers by placing two consecutive dAMP residues in the growing strand. This is an error-free process.[755,756] Genes for human polymerases ζ and η have also been identified,[757–759b] and additional polymerases have been found in yeast and other organisms.[760–760e] The human XP-V gene for polymerase η is defective in xeroderma pigmentosa variant type (Box 27-A). Polymerase ι may act sequentially with pol η to bypass highly distorting lesions.[760b] The very imprecise pol θ may also be used to bypass hard-to-remove lesions.[760c] Polymerase λ is thought to play a role in DNA repair during meiosis,[760d] while pol κ is needed in some way to provide cohesion between sister chromatids.[760e] Some of these DNA polymerases are very inaccurate.[760a] It has been suggested that human pol ι may participate in hypermutation of immunoglobin variable genes.[760f] Polymerase ϕ of yeast may be required for synthesis of ribosomal RNA.[760g]

Poly (ADP-ribose). A eukaryotic peculiarity, which is not well understood, is the synthesis of poly(ADP-ribose) chains attached to many sites in nuclear proteins (see also Eq. 15-16). Increased synthesis is observed following damage to DNA.[761–763a] The poly (ADP-ribose) polymerase binds to DNA near strand breaks or nicks and, using NAD^+ as a substrate, synthesizes the highly branched polymer attached to a small number of nuclear target proteins.

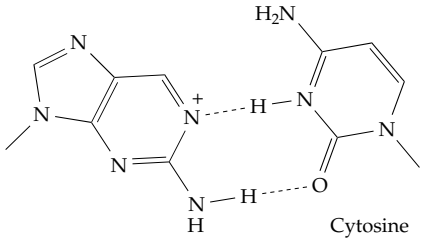
F. Mutagens in the Environment

More than 500 new chemicals are introduced into the environment industrially each year. Some widely used drugs, e.g., **hycanthone** (Fig. 5-22), are mutagenic. Powerful mutagens are present naturally in some foods.[764–766] Others have been added through ignorance. Although many of these have now been removed, the problem cannot be ignored.

One way in which chemical compounds can induce base substitution mutation is through their incorporation into the structure of DNA itself. Thus, 5-bromodeoxyuridine (or bromouracil) can replace thymidine in DNA, where it serves as an efficient mutagenic agent.[767] 2-Aminopurine, an analog of adenine, pairs with thymine, just as does adenine when incorporated into DNA.



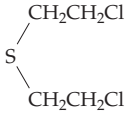
However, if it protonates on N-1, it can pair with cytosine, causing mutation.[768,769]



Another mutagenic base is N^6-hydroxylaminopurine.[770]

Many alkylating agents are powerful mutagens. Alkylation can occur at many places in DNA, but the N-7 position of guanine is especially susceptible (Eq. 5-18). The resultant positive charge on the imidazole ring portion of the alkylated guanine causes hydrolysis of the *N*-glycosyl linkage and depurination. However, this may be a lethal rather than a mutagenic event. The previously mentioned methylation on O-6 of guanine is probably more important in inducing mutations.[771]

Among the most biologically reactive alkylating agents are the nitrogen and sulfur "mustards" such as bis-(2-chloroethyl)sulfide. These toxic bifunctional compounds cause lethal crosslinking of DNA chains



bis-(2-chloroethyl)sulfide

(see Eq. 5-20). The monofunctional half-mustards are mutagenic but are less acutely toxic. Another group of alkylating agents, the **nitrosamines** (Chapter 5), are

BOX 27-A DEFICIENCIES IN HUMAN DNA REPAIR

Problems with human repair systems account for several serious diseases. One is the rare skin condition **xeroderma pigmentosum** (**XP**), an autosomal recessive hereditary defect. Homozygous individuals are extremely sensitive to ultraviolet radiation and have a high incidence of multiple carcinomas.[a–e] Defects in at least seven human genes, listed in the accompanying table, can cause XP. Most of the genes have been cloned and found to be homologs of genes for nucleotide excision repair (NER). In rodents the *ERCC* genes have the same functions as do genes for resistance to ultraviolet light in yeast (*RAD* genes). This information has elucidated the functions of the XP genes and the basis for this group of human diseases.[a,b,f]

The XP-A protein (see table) is a zinc-finger-containing DNA-binding protein, which interacts with human replication protein A (Section C,10). It may also interact with proteins corresponding to the rodent ERCC1 and ERCC4 proteins to form a complex that recognizes DNA defects.[g–i] Proteins XPC and XPE may be specialized DNA binding proteins for cyclobutane dimers and other photoproducts.[j,k] The XPG protein is an endonuclease that probably cuts on the 3' side of the DNA defect.[a,l] In human cell lines with a variety of XP mutations the DNA problems have been corrected by transfer of the corresponding normal genes into the cells.[m,n]

While defects in protein XPD often cause typical XP symptoms, some defects in the same protein lead to **trichothiodystrophy** (TTD, brittle hair disease). The hair is sulfur deficient, and scaly skin (ichthyosis, Box 8-F), mental retardation, and other symptoms are observed.[o] Like their yeast counterparts (proteins RAD3 and RAD25), XPB and XPD are both DNA helicases.[o] They also constitute distinct subunits of the human transcription factor TFIIH[p], which is discussed in Chapter 28. It seems likely that XPD is involved in **transcription-coupled repair** (**TCR**) of DNA.[o,q–s] This is a subpathway of the nucleotide excision repair (NER) pathway, which allows for rapid repair of the transcribed strand of DNA. This is important in tissues such as skin, where the global NER process may be too slow to keep up with the need for rapid protein synthesis. Transcription-coupled repair also appears to depend upon proteins CSA and CSB, defects which may result in the rare **cockayne syndrome**.[b,o,t,u] Patients are not only photosensitive but have severe mental and physical retardation including skeletal defects and a wizened appearance.

In a variant form of XP, designated XP-V, nucleotide excision repair is normal, but DNA replication is very slow. Postreplicational translesional repair (bypass repair) is also slow, and patients are cancer-prone.[v,w] The recently discovered DNA polymerase η may be defective.[x]

Children with **ataxia telangiectasia** (**AT**) have progressive neurological problems, a weak immune system, premature aging, and a high incidence of cancer.[y] Their skin fibroblasts are deficient in the ability to repair X-ray damage, which causes many double-strand breaks. Apparently in this disease cells do not wait until DNA repair has been carried out after exposure to ionizing radiation but attempt to replicate the damaged DNA.[y] The defective protein **ATM** (ataxia telangiectasia, mutated) has been identified as a large 370-kDa Ser / Thr protein kinase with a carboxyl terminal domain similar to phosphatidylinositol 3-kinase.[z,aa] It appears to play a crucial role in the cell cycle DNA damage checkpoints (Fig. 11-15) by participating in the detection of double-strand breaks and in delay of replication, while they are repaired by homologous recombination.[y,bb,cc] Although AT is an autosomal recessive disease, women heterozygous for the ATM gene have an increased susceptibility to breast cancer. This observation led to the discovery that the proteins encoded by the well known breast cancer genes *Brca1* and *Brca2* form a complex with ATM. Phosphorylation of BRCA1 by ATM may initiate a signaling pathway through the p53, c-Abl, and Chk2 proteins that cause cell cycle arrest (Fig. 11-15). BRCA1 and BRCA2 also form a complex with protein RAD51, a RecA homolog necessary for homologous recombination. BRCA1 also may be essential to transcription-coupled repair.[dd,de]

There are two major forms of hereditary susceptibility to colon cancer.[ee] Familial adenomatous polyposis is caused by defects in the *APC* gene (see Chapter 32). The more common **hereditary nonpolyposis colorectal cancer** (**HNPCC**), which includes many endometrial, stomach, and urinary tract tumors, results from defects in DNA mismatch repair.[ff–jj] The proteins hMSH2 and hMSL1 are homologs of the *E.coli* MutS and MutL (main text).

Cells of patients with **Bloom syndrome** (**BS**) have many chromosome breaks and a high frequency of sister chromatid exchanges, perhaps in an effort to correct these breaks. The body is small but well-proportioned.[kk] A somewhat similar disease, the **Werner syndrome** (**WS**), is associated with premature aging.[ll] The Bloom's protein **BLM** and the WS gene product **WRN** are both helicases related to *E.coli* RecQ. Protein BLM colocalizes with replication protein A as discrete foci in the meiotic synaptonemal complex.[mm] Protein WRN also seems to be associated with DNA replication. Defects

BOX 27-A DEFICIENCIES IN HUMAN DNA REPAIR (continued)

appear to increase homologous and illegitimate recombination.[nn] Both proteins may also function in transcription.[oo]

Many other diseases leading to a high incidence of cancer are known. Among them is the Nijmegan breakage syndrome, in which chromosomes are hypersensitive to breakage by ionizing radiation. The gene has been identified by positional cloning, and its protein is apparently involved in repair of double strand breaks.[pp] **Fanconi anemia**, **Gardner syndrome**, and hereditary **retinoblastoma** (Box 11-D) may also involve defects in DNA repair.

Some Human Hereditary Defects of DNA Repair

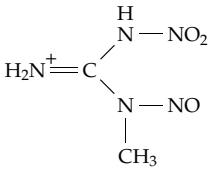
Human disease	Human gene involved		Yeast gene	Function
Xeroderma pigmentosum (XP)				
XP-A	*XPA*		*RAD*14	DNA-binding, damage recognition
XP-B	*XPB*	*ERCC*3	*RAD*25 (*SSL*2)	DNA helicase
XP-C	*XPC*		*RAD*4	DNA-binding, thymine dimers
XP-D	*XPD*	*ERCC*2	*RAD*3	DNA helicase
XP-E	*XPE*		*RAD16*	
XP-F	*XPF*		*RAD*10	DNA nuclease complex
XP-G	*XPG*		*RAD*2	DNA nuclease
XP-V	*XPV*		*RAD*30	DNA polymerase η
Cockayne syndrome (CS)				
CSA	*CSA*			Interaction with helicase CSB and with TFIIH
CSB	*CSB*			DNA helicase
Trichothiodystrophy (TID, brittle hair disease)				
	XPDE			
Ataxia telangiectasia (AT)				
	ATM			Cell cycle delay for repair of ds breaks
Human nonpolyposis colorectal cancer (HNPCC)				
	*hMSH*2		*MSH*2	Mismatch repair
	*hMSL*1		*MSL*1	
	GTBP			

[a] Hoeijmakers, J. H. J., and Bootsma, D. (1994) *Nature (London)* **371**, 654–655
[b] Tanaka, K., and Wood, R. D. (1994) *Trends Biochem. Sci.* **19**, 83–86
[c] de Vries, A., van Oostrom, C. T. M., Hofhuis, F. M. A., Dortant, P. M., Berg, R. J. W., de Gruijl, F. R., Wester, P. W., van Kreijl, C. F., Capel, P. J. A., van Steeg, H., and Verbeek, S. J. (1995) *Nature (London)* **377**, 169–173
[d] Nakane, H., and 19 other authors. (1995) *Nature (London)* **377**, 165–168
[e] Lehmann, A. R. (1995) *Trends Biochem. Sci.* **20**, 402–405
[f] Bootsma, D., and Hoeijmakers, J. H. J. (1993) *Nature (London)* **363**, 114–115
[g] Bankmann, M., Prakash, L., and Prakash, S. (1992) *Nature (London)* **355**, 555–558
[h] Li, L., Elledge, S. J., Peterson, C. A., Bales, E. S., and Legerski, R. J. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5012–5016
[i] Lee, S.-H., Kim, D.-K., and Drissi, R. (1995) *J. Biol. Chem.* **270**, 21800–21805
[j] Emmert, S., Kobayashi, N., Khan, S. G., and Kraemer, K. H. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2151–2156
[k] Hwang, B. J., and Chu, G. (1993) *Biochemistry* **32**, 1657–1666
[l] O'Donovan, A., Scherly, D., Clarkson, S. G., and Wood, R. D. (1994) *J. Biol. Chem.* **269**, 15965–15968
[m] Evans, M. K., Robbins, J. H., Ganges, M. B., Tarone, R. E., Nairn, R. S., and Bohr, V. A. (1993) *J. Biol. Chem.* **268**, 4839–4847

BOX 27-A (continued)

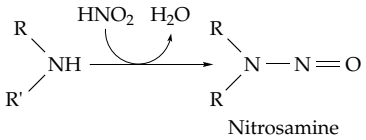
[n] Tebbs, R. S., Zhao, Y., Tucker, J. D., Scheerer, J. B., Siciliano, M. J., Hwang, M., Liu, N., Legerski, R. J., and Thompson, L. H. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6354–6358
[o] van Gool, A. J., van der Horst, G. T. J., Citterio, E., and Hoeijmakers, J. H. J. (1997) *EMBO J.* **16**, 4155–4162
[p] Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J. H. J., Chambon, P., and Egly, J.-M. (1993) *Science* **260**, 58–63
[q] Guzder, S. N., Sung, P., Prakash, S., and Prakash, L. (1995) *J. Biol. Chem.* **270**, 17660–17663
[r] Winkler, G. S., Araújo, S. J., Fiedler, U., Vermeulen, W., Coin, F., Egly, J.-M., Hoeijmakers, J. H. J., Wood, R. D., Timmers, H. T. M., and Weeda, G. (2000) *J. Biol. Chem.* **275**, 4258–4266
[s] Reardon, J. T., Bessho, T., Kung, H. C., Bolton, P. H., and Sancar, A. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9463–9468
[t] Mu, D., and Sancar, A. (1997) *J. Biol. Chem.* **272**, 7570–7573
[u] Cooper, P. K., Nouspikel, T., Clarkson, S. G., and Leadon, S. A. (1997) *Science* **275**, 990–993
[v] Masutani, C., Araki, M., Yamada, A., Kusumoto, R., Nogimori, T., Maekawa, T., Iwai, S., and Hanaoka, F. (1999) *EMBO J.* **18**, 3491–3501
[w] Johnson, R. E., Kondratick, C. M., Prakash, S., and Prakash, L. (1999) *Science* **285**, 263–265
[x] Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999) *Nature (London)* **399**, 700–704
[y] Beamish, H., Williams, R., Chen, P., and Lavin, M. F. (1996) *J. Biol. Chem.* **271**, 20486–20493
[z] Lavin, M. F., Khanna, K. K., Beamish, H., Spring, K., and Watters, D. (1995) *Trends Biochem. Sci.* **20**, 382–383
[aa] Chong, M. J., Murray, M. R., Gosink, E. C., Russell, H. R. C., Srinivasan, A., Kapsetaki, M., Korsmeyer, S. J., and McKinnon, P. J. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 889–894
[bb] Morrison, C., Sonoda, E., Takao, N., Shinohara, A., Yamamoto, K.-i, and Takeda, S. (2000) *EMBO J.* **19**, 463–471
[cc] Cortez, D., Wang, Y., Qin, J., and Elledge, S. J. (1999) *Science* **286**, 1162–1166
[dd] Gowen, L. C., Avrutskaya, A. V., Latour, A. M., Koller, B. H., and Leadon, S. A. (1998) *Science* **281**, 1009–1012
[de] Davies, A. A., Masson, J.-Y., McIlwraith, M. J., Stasiak, A. Z., Stasiak, A., Venkitaraman, A. R., and West, S. C. (2001) *Molecular Cell* **7**, 273–282
[ee] Peltomäki, P., Aaltonen, L. A., Sistonen, P., Pylkkänen, L., Mecklin, J.-P., Järvinen, H., Green, J. S., Jass, J. R., Weber, J. L., Leach, F. S., Petersen, G. M., Hamilton, S. R., de la Chapelle, A., and Vogelstein, B. (1993) *Science* **260**, 810–819
[ff] Papadopoulos, N., and 19 other authors. (1994) *Science* **18**, 1625–1629
[gg] Prolla, T. A., Pang, Q., Alani, E., Kolodner, R. D., and Liskay, R. M. (1994) *Science* **265**, 1091–1092
[hh] Kolodner, R. D. (1995) *Trends Biochem. Sci.* **20**, 397–401
[ii] Palombo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D'Arrigo, A., Truong, O., Hsuan, J. J., and Jiricny, J. (1995) *Science* **268**, 1912–1914
[jj] Mellon, I., Rajpal, D. K., Koi, M., Boland, C. R., and Champe, G. N. (1996) *Science* **272**, 557–560
[kk] German, J., Roe, A. M., Leppert, M. F., and Ellis, N. A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6669–6673
[ll] Lebel, M., Spillare, E. A., Harris, C. C., and Leder, P. (1999) *J. Biol. Chem.* **274**, 37795–37799
[mm] Walpita, D., Plug, A. W., Neff, N. F., German, J., and Ashley, T. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5622–5627
[nn] Yamagata, K., Kato, J.-I., Shimamoto, A., Goto, M., Furuichi, Y., and Ikeda, H. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8733–8738
[oo] Lee, S.-K., Johnson, R. E., Yu, S.-L., Prakash, L., and Prakash, S. (1999) *Science* **286**, 2339–2342
[pp] Dong, Z., Zhong, Q., and Chen, P.-L. (1999) *J. Biol. Chem.* **274**, 19513–19516

highly mutagenic.[772,773] Much used in the laboratory as a mutagen is ***N*-methyl-*N'*-nitro-*N*-nitroso-guanidine**:



N-Ethyl-*N*-nitrosourea is one of the most potent carcinogens known.[773,774] These compounds, as well as diazomethane and other related substances, probably act via a common intermediate: $CH_3N_2^+$ or $C_2H_5N_2^+$.[775]

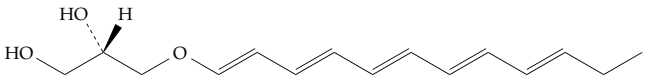
Any secondary amine will react with nitrous acid to form a nitrosamine (Eq. 27-25). Tertiary amines can also react with loss of one alkyl group. This can occur in the stomach, and the nitrosamines may be absorbed into the system. All plants contain some nitrate and some, such as spinach and beets, have large amounts. Bacon and other cured R' R meats contain both nitrites and nitrates, and many drugs and natural food



Nitrosamine (27-25)

constituents are secondary amines. Cigarette smoke also contains a nitrosamine.[776] There is a possibility that these substances may induce human cancer.[772,777]

About 3% of residents of North America excrete feces that contain mutagenic unsaturated glyceryl ethers, which have been named **fecapentaenes**.[778,779]

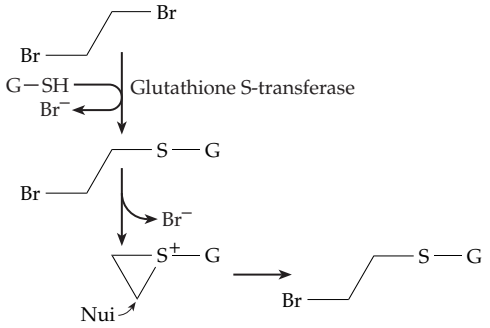


A fecapentaene

They could be among the causative agents of colorectal cancer. Protonation of a double bond will produce a reactive carbocation, which may be an active alkylating agent.[779]

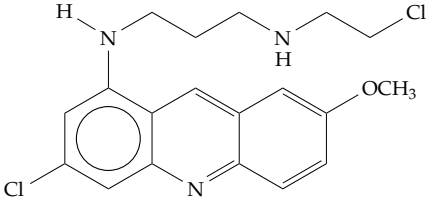
Many halogenated compounds are carcinogenic. Among these is **1,2-dibromoethane,** which has been

produced in the United States in quantities as great as 10^8 kg / year. It has been widely used as a fumigant for foods, as an industrial solvent, and as an additive to gasoline. It is a **procarcinogen**, whose carcinogenic properties are expressed only after **metabolic activation**. One pathway of activation is reaction with glutathione to form a thioether, which can cyclize to a sulfonium compound.[780] The latter alkylates nucleophilic groups in DNA as in Eq. 27-26. Other bifunctional electrophiles such as acrolein, malondialdehyde, vinyl chloride, and urethane are procarcinogens.[781–784] Styrene, another procarcinogen, is converted to the carcinogenic styrene oxide by action of a cytochrome P450.[785] Metabolites of glucose such as **pyruvaldehyde**, **methylglyoxal**, and other reactive aldehydes can also attack DNA.[786,787]



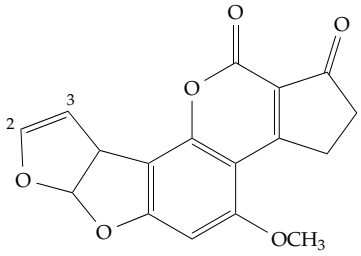
(27-26)

Less common than base pair exchanges are frame-shift mutations. A characteristic of such mutations is that they do not revert (back mutate) as readily as do base substitution mutations, and reversion is not induced by chemicals known to cause base substitutions. However, reversion of frame-shift mutations is induced by acridines and other flat molecules that are known to act as intercalating agents in DNA helices and which promote frame-shift mutations. They are especially effective in causing mutations of regions in which long repeated sequences of a single base such as $(A)_n$ or $(G)_n$ occur.[770,788] For example, deletion of two base pairs from a "hot spot" (site of frequent mutation) in the *Salmonella* histidine[789] operon with the following sequence is induced by 2-nitrosofluorene and causes reversion of a (–1) histidine-requiring mutant: 5'-CGCGCGCG. Whereas simple intercalating agents are often not very mutagenic, compounds that are both an intercalating agent and an alkylating agent are especially potent. An example is the following compound, which contains a half-mustard side chain:



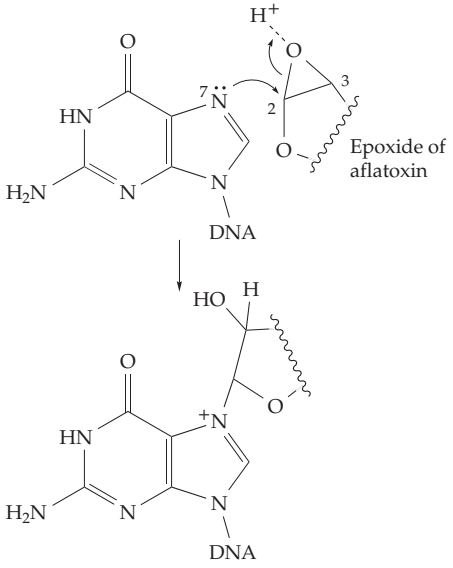
When the CH_2Cl group of the side chain is replaced by CH_2OH, the compound is 100 times less mutagenic.

The carcinogenic **aflatoxins**, which are produced by *Aspergillus flavus*, may be present in infected peanuts and other foodstuffs.[790] Like many other compounds that are carcinogenic or mutagenic, the aflatoxins are not unusually reactive chemically.



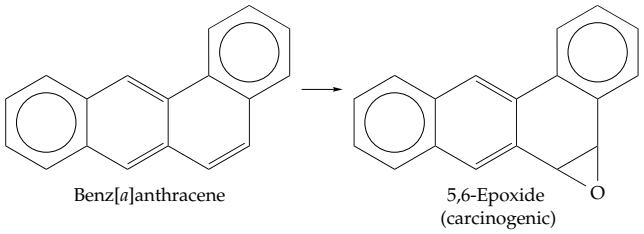
Aflatoxin B_1

However, they are activated in the animal body by hydroxylation and formation of 2,3-epoxides. The latter may react with N-7 of guanyl residues in DNA (Eq. 27-27).[791,792]

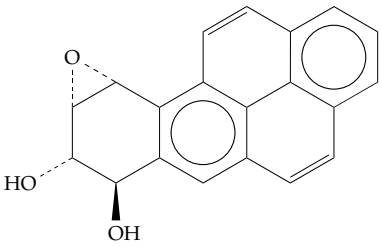


(27-27)

Benz(*a*)anthracene is also converted in the body to a carcinogenic epoxide (Eq. 27-28).[793] Benzo(*a*)pyrene was isolated from coal tar in 1929 and in 1930 provided the first demonstration of the carcinogenicity of a pure chemical compound.[794] It can also be activated by conversion in the ER to the 7,8-dihydrodiol 9,10-



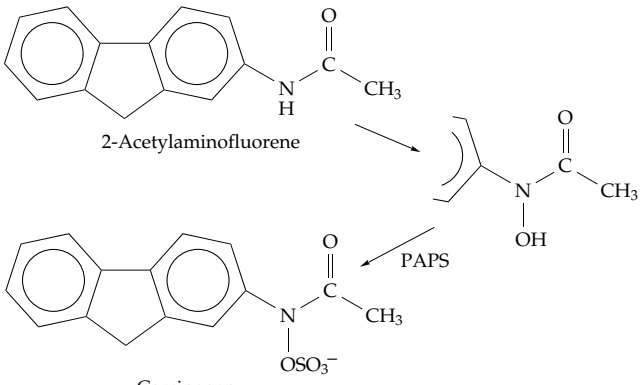
(27-28)



Benzo(*a*)pyrene diol epoxide

epoxide.[795–797] Hydroxylation activates 2-acetylaminofluorene to a carcinogenic *N*-sulfate (Eq. 27-29),[788,798] which reacts with guanine rings.[788,799]

Although it seems incongruous, many of the most potent antitumor drugs are powerful mutagens.



(27-29)

Among these are intercalating agents such as daunomycin (Figs. 5-22, 5-23), neocarzinostatin, and bleomycin (Box 5-B). These are alkylating reagents[800] or attack DNA in other ways. The fact that such compounds are in use for chemotherapy emphasizes the need for new approaches to cancer treatment.

How can compounds be recognized as mutagenic? This is an important question that is difficult to answer and which has led to many controversies. For example, how dangerous is formaldehyde in the environment?[801] Butadiene?[802] Dioxin?[790,803] Is fluoride a carcinogen?[804,805]

A quick test for mutagenic activity makes use of **tester strains** of bacteria developed by Ames and associates. These are *Salmonella* mutants that are unable to synthesize their own histidine, but which can grow when a mutagenic agent produces a back mutation.[806,807] One of the strains can be mutated by agents causing base exchanges, while the other three, which contain different types of frame-shift mutations, are affected differently by various mutagens. About 10^9 bacteria are spread on a Petri plate, and a small amount of the mutagenic chemical is introduced in the center of the plate. Where back mutation has occurred, a colony of the bacteria appears. The strains all carry a mutation in the main DNA excision-repair system so that most mutations are not repaired, and the test is very sensitive. Addition of a liver homogenate plus an NADPH-generating system to the sample tested allows activation of many aromatic chemicals by hydroxylation.[808] Feeding of mutagens to *Drosophila* eye color mutants permits testing for back mutation in a eukaryote.[809]

The bacterial tests have been widely used and have been of great value. For example, they revealed that certain chemicals that were being used as flame retardants in children's clothing are mutagens,[810] and that mutagens can be generated during cooking of meat and other foods.[811] However, there are good reasons for using other methods of identifying mutagens together with the bacterial tests.[812] Government regulatory agencies in the United States have relied largely on tests with rodents. Compounds are tested at very high doses with these short-lived animals. Extrapolation to human exposures at very low levels has been criticized.[764,765,813,814] However, rodent tests have identified many true carcinogens.[815] A broader range of tests are now in use.[816] Direct monitoring of the accumulation of defects in animal and human cells is now possible. For example, in the **^{32}P-postlabeling technique** DNA is enzymatically digested to the nucleotide level, and the adducts with mutagens are labeled with ^{32}P, separated, and their quantities measured.[817–819] Laser-excited fluorescence from such adducts, GC / mass spectrometry, and immunological methods can also be used to identify DNA adducts.[819] Careful vigilance is needed to keep our mutation rate at as low a level as possible.

References

1. Wasserman, S. A., and Cozzarelli, N. R. (1986) *Science* **232**, 951–960
2. Reich, Z., Wachtel, E. J., and Minsky, A. (1994) *Science* **264**, 1460–1463

2a. Daban, J.-R. (2000) *Biochemistry* **39**, 3862–3866

3. Moody, M. F. (1999) *J. Mol. Biol.* **293**, 401–433
4. Black, L. W., Newcomb, W. W., Boring, J. W., and Brown, J. C. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7960–7964

4a. Hud, N. V., and Downing, K. H. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14925–14930

5. Sun, M., and Serwer, P. (1997) *Biophys. J.* **72**, 958–963
6. Marx, K. A., and Reynolds, T. C. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6484–6488
7. Widom, J., and Baldwin, R. L. (1983) *J. Mol. Biol.* **171**, 419–437
8. Kamashev, D., Balandina, A., and Rouviere-Yaniv, J. (1999) *EMBO J.* **18**, 5434–5444
9. Kobryn, K., Lavoie, B. D., and Chaconas, G. (1999) *J. Mol. Biol.* **289**, 777–784
10. Schmid, M. B. (1988) *Trends Biochem. Sci.* **13**, 131–135
11. Claret, L., and Rouviere-Yaniv, J. (1996) *J. Mol. Biol.* **263**, 126–139
12. Esser, D., Rudolph, R., Jaenicke, R., and Böhm, G. (1999) *J. Mol. Biol.* **291**, 1135–1146
13. Tanaka, I., Appelt, K., Dijk, J., White, S. W., and Wilson, K. S. (1984) *Nature (London)* **310**, 376–381
14. White, S. W., Wilson, K. S., Appelt, K., and Tanaka, I. (1999) *Acta Crystallogr. D* **55**, 801–809

14a. Grove, A., and Saavedra, T. C. (2002) *Biochemistry* **41**, 7597–7603

15. Ueguchi, C., Suzuki, T., Yoshida, T., Tanaka, K.-i, and Mizuno, T. (1996) *J. Mol. Biol.* **263**, 149–162
16. Starich, M. R., Sandman, K., Reeve, J. N., and Summers, M. F. (1996) *J. Mol. Biol.* **255**, 187–203
17. Krueger, J. K., McCrary, B. S., Wang, A. H.-J., Shriver, J. W., Trewhella, J., and Edmondson, S. P. (1999) *Biochemistry* **38**, 10247–10255

17a. Kamashev, D., and Rouviere-Yaniv, J. (2000) *EMBO J.* **19**, 6527–6535

17b. Grove, A., and Lim, L. (2001) *J. Mol. Biol.* **311**, 491–502

18. Pettijohn, D. E. (1988) *J. Biol. Chem.* **263**, 12793–12796

18a. Renzoni, D., Esposito, D., Pfuhl, M., Hinton, J. C. D., Higgins, C. F., Driscoll, P. C., and Ladbury, J. E. (2001) *J. Mol. Biol.* **306**, 1127–1137

18b. Schröder, O., and Wagner, R. (2000) *J. Mol. Biol.* **298**, 737–748

19. Jackson, D. A., and Patel, S. B. (1982) *Trends Biochem. Sci.* **7**, 272–274

19a. Korolev, N., Lyubartsev, A. P., Nordenskiöld, L., and Laaksonen, A. (2001) *J. Mol. Biol.* **308**, 907–917

20. Hunt, J. G., Kasinsky, H. E., Elsey, R. M., Wright, C. L., Rice, P., Bell, J. E., Sharp, D. J., Kiss, A. J., Hunt, D. F., Arnott, D. P., Russ, M. M., Shabanowitz, J., and Ausió, J. (1996) *J. Biol. Chem.* **271**, 23547–23557
21. Harborne, J. B. (1993) *Introduction to Ecological Biochemistry*, 4th ed., Academic Press, San Diego, California

21a. Prieto, C., Saperas, N., Arnan, C., Hills, M. H., Wang, X., Chiva, M., Aligué, R., Subirana, J. A., and Ausió, J. (2002) *Biochemistry* **41**, 7802–7810

22. Bellvé, A. R., McKay, D. J., Renaux, B. S., and Dixon, G. H. (1988) *Biochemistry* **27**, 2890–2897
23. de Yebra, L., Ballescà, J. L., Vanrell, J. A., Bassas, L., and Oliva, R. (1993) *J. Biol. Chem.* **268**, 10553–10557
24. Doenecke, D., and Karlson, P. (1984) *Trends Biochem. Sci.* **9**, 404–405
25. Kornberg, R. D., and Thomas, J. O. (1974) *Science* **184**, 865–868
26. Isenberg, I. (1979) *Ann. Rev. Biochem.* **48**, 159–191
27. Arents, G., and Moudrianakis, E. N. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11170–11174
28. Zemzoumi, K., Frontini, M., Bellorini, M., and Mantovani, R. (1999) *J. Mol. Biol.* **286**, 327–337
29. Wellman, S. E., Song, Y., and Mamoon, N. M. (1999) *Biochemistry* **38**, 13112–13118
30. Lindner, H., Sarg, B., Hoertnagl, B., and Helliger, W. (1998) *J. Biol. Chem.* **273**, 13324–13330
31. Zlatanova, J., and Doenecke, D. (1994) *FASEB J.* **8**, 1260–1268
32. Hansen, J. C., Tse, C., and Wolffe, A. P. (1998) *Biochemistry* **37**, 17637–17641
33. DeLange, R. J., Hooper, J. A., and Smith, E. L. (1973) *J. Biol. Chem.* **248**, 3261–3274

33a. Wu, J., and Grunstein, M. (2000) *Trends Biochem. Sci.* **25**, 619–623

33b. Jenuwein, T., and Allis, C. D. (2001) *Science* **293**, 1074–1080

33c. Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtier, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000) *Nature (London)* **406**, 593–599

33d. Jacobs, S. A., and Khorasanizadeh, S. (2002) *Science* **295**, 2080–2083

34. Kelner, D. N., and McCarty, K. S., Sr. (1984) *J. Biol. Chem.* **259**, 3413–3418
35. Sternglanz, R. (1996) *Trends Biochem. Sci.* **21**, 357–358
36. Strenglanz, R., and Schindelin, H. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8807–8808
37. Champagne, N., Bertos, N. R., Pelletier, N., Wang, A. H., Vezmar, M., Yang, Y., Heng, H. H., and Yang, X.-J. (1999) *J. Biol. Chem.* **274**, 28528–28536
38. Wade, P. A., Pruss, D., and Wolffe, A. P. (1997) *Trends Biochem. Sci.* **22**, 128–132

38a. Marmorstein, R. (2001) *J. Mol. Biol.* **311**, 433–444

38b. Kruhlak, M. J., Hendzel, M. J., Fischle, W., Bertos, N. R., Hameed, S., Yang, X.-J., Verdin, E., and Bazett-Jones, D. P. (2001) *J. Biol. Chem.* **276**, 38307–38319

39. Hay, C. W., and Candido, E. P. M. (1983) *J. Biol. Chem.* **258**, 3726–3734
40. Kölle, D., Brosch, G., Lechner, T., Pipal, A., Helliger, W., Taplick, J., and Loidl, P. (1999) *Biochemistry* **38**, 6769–6773

40a. Huang, X., and Kadonaga, J. T. (2001) *J. Biol. Chem.* **276**, 12497–12500

40b. Furumai, R., Komatsu, Y., Nishino, N., Khochbin, S., Yoshida, M., and Horinouchi, S. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 87–92

41. Osley, M. A. (1991) *Ann. Rev. Biochem.* **60**, 827–861
42. Gorovsky, M. A. (1986) in *Molecular Biology of Ciliated Protozoas* (Gall, J. G., ed), Academic Press, Orlando, Florida
43. Prentice, D. A., Loechel, S. C., and Kitos, P. A. (1982) *Biochemistry* **21**, 2412–2420
44. Swank, R. A., Th'ng, J. P. H., Guo, X.-W., Valdez, J., Bradbury, E. M., and Gurley, L. R. (1997) *Biochemistry* **36**, 13761–13768
45. Allera, C., Lazzarini, G., and Patrone, E. (1995) *Biochemistry* **34**, 301–311

45a. de la Barre, A.-E., Angelov, D., Molla, A., and Dimitrov, S. (2001) *EMBO J.* **20**, 6383–6393

46. Koiv, A., Palvimo, J., and Kinnunen, P. K. J. (1995) *Biochemistry* **34**, 8018–8027

46a. Ausió, J., and Abbott, D. W. (2002) *Biochemistry* **41**, 5945–5949

47. Chen, H. Y., Sun, J.-M., Zhang, Y., Davie, J. R., and Meistrich, M. L. (1998) *J. Biol. Chem.* **273**, 13165–13169

47a. Soares, D. J., Sandman, K., and Reeve, J. N. (2000) *J. Mol. Biol.* **297**, 39–47

47b. Bell, S. D., Botting, C. H., Wardleworth, B. N., Jackson, S. P., and White, M. F. (2002) *Science* **296**, 148–151

48. Kornberg, R. D., and Klug, A. (1981) *Sci. Am.* **244**(Feb), 52–64
49. Grunstein, M. (1992) *Sci. Am.* **267**(Oct), 68–74B
50. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) *Nature (London)* **389**, 251–260
51. Griffith, J. D. (1975) *Science* **187**, 1202–1203
52. Krude, T., and Knippers, R. (1994) *J. Biol. Chem.* **269**, 21021–21029
53. Givens, R. M., Saavedra, R. A., and Huberman, J. A. (1996) *J. Mol. Biol.* **257**, 53–65
54. Sen, D., Mitra, S., and Crothers, D. M. (1986) *Biochemistry* **25**, 3441–3447
55. Felsenfeld, G., and McGhee, J. D. (1986) *Cell* **44**, 375–377
56. Wasserman, S. A., White, J. H., and Cozzarelli, N. R. (1988) *Nature (London)* **334**, 448–450
57. McMurray, C. T., van Holde, K. E., Jones, R. L., and Wilson, W. D. (1985) *Biochemistry* **24**, 7037–7044
58. Fitzgerald, D. J., and Anderson, J. N. (1999) *J. Mol. Biol.* **293**, 477–491
59. Lowary, P. T., and Widom, J. (1998) *J. Mol. Biol.* **276**, 19–42

59a. Thåström, A., Lowary, P. T., Widlund, H. R., Cao, H., Kubista, M., and Widom, J. (1999) *J. Mol. Biol.* **288**, 213–229

60. Flaus, A., and Richmond, T. J. (1998) *J. Mol. Biol.* **275**, 427–441
61. Widlund, H. R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P. E., Kahn, J. D., Crothers, D. M., and Kubista, M. (1997) *J. Mol. Biol.* **267**, 807–817

61a. Negri, R., Buttinelli, M., Panetta, G., De Arcangelis, V., Di Mauro, E., and Travers, A. (2001) *J. Mol. Biol.* **307**, 987–999

62. Leuba, S. H., Zlatanova, J., and van Holde, K. (1993) *J. Mol. Biol.* **229**, 917–929
63. Crane-Robinson, C. (1997) *Trends Biochem. Sci.* **22**, 75–77
64. Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., and Woodcock, C. L. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14173–14178

64a. Vila, R., Ponte, I., Jiménez, M. A., Rico, M., and Suau, P. (2000) *Protein Sci.* **9**, 627–636

64b. Bharath, M. M. S., Ramesh, S., Chandra, N. R., and Rao, M. R. S. (2002) *Biochemistry* **41**, 7617–7627

65. Pruss, D., Bartholomew, B., Persinger, J., Hayes, J., Arents, G., Moudrianakis, E. N., and Wolffe, A. P. (1996) *Science* **274**, 614–617
66. Zhou, Y.-B., Gerchman, S. E., Ramakrishnan, V., Travers, A., and Muyldermans, S. (1998) *Nature (London)* **395**, 402–405
67. Hayes, J. J. (1996) *Biochemistry* **35**, 11931–11937
68. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., and Sweet, R. M. (1993) *Nature (London)* **362**, 219–223
69. DeLucia, F., Alilat, M., Sivolob, A., and Prunell, A. (1999) *J. Mol. Biol.* **285**, 1101–1119
70. Pennisi, E. (1997) *Science* **275**, 155–157
71. Polach, K. J., and Widom, J. (1995) *J. Mol. Biol.* **254**, 130–149
72. Strahl, B. D., and Allis, C. D. (2000) *Nature (London)* **403**, 41–45
73. Cairns, B. R. (1998) *Trends Biochem. Sci.* **23**, 20–24
74. Kellenberger, E. (1987) *Trends Biochem. Sci.* **12**, 105–107

References

75. Gross, D. S., and Garrard, W. T. (1988) *Ann. Rev. Biochem.* **57**, 159–197
76. Felsenfeld, G., Boyes, J., Chung, J., Clark, D., and Studitsky, V. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9384–9388
77. Wu, C. (1997) *J. Biol. Chem.* **272**, 28171–28174
78. Grunstein, M. (1997) *Nature (London)* **389**, 349–352
79. Edmondson, D. G., and Roth, S. Y. (1996) *FASEB J.* **10**, 1173–1182
80. Nightingale, K. P., Wellinger, R. E., Sogo, J. M., and Becker, P. B. (1998) *EMBO J.* **17**, 2865–2876
81. Zhang, W., Bone, J. R., Edmondson, D. G., Turner, B. M., and Roth, S. Y. (1998) *EMBO J.* **17**, 3155–3167
82. Utley, R. T., Ikeda, K., Grant, P. A., Côté, J., Steger, D. J., Eberharter, A., John, S., and Workman, J. L. (1998) *Nature (London)* **394**, 498–502
83. Mueller, R. D., Yasuda, H., and Bradbury, E. M. (1985) *J. Biol. Chem.* **260**, 5081–5086
84. Muchardt, C., and Yaniv, M. (1999) *J. Mol. Biol.* **293**, 187–198
84a. Horn, P. J., and Peterson, C. L. (2002) *Science* **297**, 1824 – 1827
85. Thoma, F., Koller, T. H., and Klug, A. (1979) *J. Cell Biol.* **83**, 403–427
86. Walker, P. R., and Sikorska, M. (1987) *J. Biol. Chem.* **262**, 12218–12222
87. Woodcock, C. L. F., Frado, L.-L. Y., and Rattner, J. B. (1984) *J. Cell Biol.* **99**, 42–52
88. Rydberg, B., Holley, W. R., Mian, I. S., and Chatterjee, A. (1998) *J. Mol. Biol.* **284**, 71–84
89. Daban, J.-R., and Bermúdez, A. (1998) *Biochemistry* **37**, 4299–4304
90. Sirlin, J. L. (1972) *Biology of RNA*, Academic Press, New York (pp. 162–164)
91. Manuelidis, L. (1990) *Science* **250**, 1533–1540
91a. Uhlmann, F. (2002) *Nature (London)* **417**, 135–136
91b. Antes, T. J., Namciu, S. J., Fournier, R. E. K., and Levy-Wilson, B. (2001) *Biochemistry* **40**, 6731–6742
91c. Löwe, J., Cordell, S. C., and van den Ent, F. (2001) *J. Mol. Biol.* **306**, 25–35
92. Münkel, C., Eils, R., Dietzel, S., Zink, D., Mehring, C., Wedemann, G., Cremer, T., and Langowski, J. (1999) *J. Mol. Biol.* **285**, 1053–1065
93. van Holde, K., and Zlatanova, J. (1995) *J. Biol. Chem.* **270**, 8373–8376
94. Nagele, R., Freeman, T., McMorrow, L., and Lee, H.-y. (1995) *Science* **270**, 1831–1835
95. Lamond, A. I., and Earnshaw, W. C. (1998) *Science* **280**, 547–553
96. Wagner, R. P., Maguire, M. M., and Stallings, R. L. (1993) *Chromosomes: A Synthesis*, Wiley-Liss, New York
97. Cook, K. R., and Karpen, G. H. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5219–5221
98. Motta, M. C., Landsberger, N., Merli, C., and Badaracco, G. (1998) *J. Biol. Chem.* **273**, 18028–18039
98a. Ringrose, L., and Paro, R. (2001) *Nature (London)* **412**, 493–494
99. Wei, X., Samarabandu, J., Devdhar, R. S., Siegel, A. J., Acharya, R., and Berezney, R. (1998) *Science* **281**, 1502–1505
100. Baskin, Y. (1995) *Science* **268**, 1564–1565
101. Lemon, K. P., and Grossman, A. D. (1998) *Science* **282**, 1516–1519
102. Hirano, T. (1995) *Trends Biochem. Sci.* **20**, 357–361
102a. Misteli, T. (2001) *Science* **291**, 843 – 847
103. Tsutsui, K., Tsutsui, K., and Muller, M. T. (1988) *J. Biol. Chem.* **263**, 7235–7241
104. Nelson, W. G., Pienta, K. J., Barrack, E. R., and Coffey, D. S. (1986) *Ann. Rev. Biophys. Biophys. Chem.* **15**, 457–475
105. Pemov, A., Bavykin, S., and Hamlin, J. L. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14757–14762
106. Benham, C., Kohwi-Shigematsu, T., and Bode, J. (1997) *J. Mol. Biol.* **274**, 181–196
107. Pederson, T. (1998) *J. Mol. Biol.* **277**, 147–159
108. Mirkovitch, J., Spierer, P., and Laemmli, U. K. (1986) *J. Mol. Biol.* **190**, 255–258
109. Kay, V., and Bode, J. (1994) *Biochemistry* **33**, 367–374
110. Udvardy, A. (1999) *EMBO J.* **18**, 1–8
111. Harborth, J., Wang, J., Gueth-Hallonet, C., Weber, K., and Osborn, M. (1999) *EMBO J.* **18**, 1689–1700
112. Peterson, J. L., and McConkey, E. H. (1976) *J. Biol. Chem.* **251**, 548–554
113. Hnilica, L. S., ed. (1983) *Chromosomal Nonhistone Proteins*, Vol. III, CRC Press, Boca Raton, Florida
114. Quick, D. P., Orchard, P. J., and Duerre, J. A. (1981) *Biochemistry* **20**, 4724–4729
115. Landsman, D., McBride, O. W., Soares, N., Crippa, M. P., Srikantha, T., and Bustin, M. (1989) *J. Biol. Chem.* **264**, 3421–3427
115a. Müller, S., Scaffidi, P., Degryse, B., Bonaldi, T., Ronfani, L., Agresti, A., Beltrame, M., and Bianchi, M. E. (2001) *EMBO J.* **20**, 4337–4340
115b. Thomas, J. O., and Travers, A. A. (2001) *Trends Biochem. Sci.* **26**, 167–174
115c. Bustin, M. (2001) *Trends Biochem. Sci.* **26**, 431–437
115d. Banks, G. C., Li, Y., and Reeves, R. (2000) *Biochemistry* **39**, 8333–8346
115e. Webb, M., Payet, D., Lee, K.-B., Travers, A. A., and Thomas, J. O. (2001) *J. Mol. Biol.* **309**, 79–88
116. Dingwall, C., and Laskey, R. (1992) *Science* **258**, 942–947
117. Ashery-Padan, R., Weiss, A. M., Feinstein, N., and Gruenbaum, Y. (1997) *J. Biol. Chem.* **272**, 2493–2499
118. Davis, L. I. (1995) *Ann. Rev. Biochem.* **64**, 865–896
119. Görlich, D. (1998) *EMBO J.* **17**, 2721–2727
120. Goldberg, M. W., and Allen, T. D. (1996) *J. Mol. Biol.* **257**, 848–865
120a. Raska, I., Aebi, U., and Earnshaw, W. C. (2000) *EMBO J.* **19**, 3843 – 3848
120b. Shahin, V., Danker, T., Enss, K., Ossig, R., and Oberleithner, H. (2001) *FASEB J.* **15**, 1895 – 1901
121. Aebi, U., Cohn, J., Buhle, L., and Gerace, L. (1986) *Nature (London)* **323**, 560–564
122. Worman, H. J., Lazaridis, I., and Georgatos, S. D. (1988) *J. Biol. Chem.* **263**, 12135–12141
123. Gerace, L. (1986) *Trends Biochem. Sci.* **11**, 443–446
123a. Georgatos, S. D. (2001) *EMBO J.* **20**, 2989 – 2994
124. McKeon, F. D., Kirschner, M. W., and Caput, D. (1986) *Nature (London)* **319**, 463–468
125. Schneider, U., Mini, T., Jenö, P., Fisher, P. A., and Stuurman, N. (1999) *Biochemistry* **38**, 4620–4632
126. Favreau, C., Worman, H. J., Wozniak, R. W., Frappier, T., and Courvalin, J.-C. (1996) *Biochemistry* **35**, 8035–8044
126a. Rout, M. P., and Aitchison, J. D. (2001) *J. Biol. Chem.* **276**, 16593–16596
126b. Allen, N. P. C., Huang, L., Burlingame, A., and Rexach, M. (2001) *J. Biol. Chem.* **276**, 29268–29274
127. Weis, K. (1998) *Trends Biochem. Sci.* **23**, 185–189
128. Cingolani, G., Petosa, C., Weis, K., and Müller, C. W. (1999) *Nature (London)* **399**, 221–228
129. Chook, Y. M., and Blobel, G. (1999) *Nature (London)* **399**, 230–237
130. Panté, N., and Aebi, U. (1996) *Science* **273**, 1729–1732
131. Moore, M. S. (1998) *J. Biol. Chem.* **273**, 22857–22860
132. Vetter, I. R., Nowak, C., Nishimoto, T., Kuhlmann, J., and Wittinghofer, A. (1999) *Nature (London)* **398**, 39–46
132a. Geyer, M., Assheuer, R., Klebe, C., Kuhlmann, J., Becker, J., Wittinghofer, A., and Kalbitzer, H. R. (1999) *Biochemistry* **38**, 11250–11260
132b. Goldberg, M. W., Rutherford, S. A., Hughes, M., Cotter, L. A., Bagley, S., Kiseleva, E., Allen, T. D., and Clarke, P. R. (2000) *J. Mol. Biol.* **300**, 519–529
133. Akey, C. W. (1995) *J. Mol. Biol.* **248**, 273–293
134. Kiseleva, E., Goldberg, M. W., Daneholt, B., and Allen, T. D. (1996) *J. Mol. Biol.* **260**, 304–311
135. Perez-Terzic, C., Pyle, J., Jaconi, M., Stehno-Bittel, L., and Clapham, D. E. (1996) *Science* **273**, 1875–1877
136. Stoffler, D., Goldie, K. N., Feja, B., and Aebi, U. (1999) *J. Mol. Biol.* **287**, 741–752
137. Cook, P. R. (1999) *Science* **284**, 1790–1795
138. Chang, J.-H., Dumbar, T. S., and Olson, M. O. J. (1988) *J. Biol. Chem.* **263**, 12824–12827
139. Serin, G., Joseph, G., Ghisolfi, L., Bauzan, M., Erard, M., Amalric, F., and Bouvet, P. (1997) *J. Biol. Chem.* **272**, 13109–13116
139a. Wang, I.-F., Reddy, N. M., and Shen, C.-K. J. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13583–13588
140. Somerville, C., and Somerville, S. (1999) *Science* **285**, 380–383
141. C. elegans Sequencing Consortium. (1998) *Science* **282**, 2012–2018
142. Long, E. O., and Dawid, I. B. (1980) *Ann. Rev. Biochem.* **49**, 727–764
143. Jelinek, W. R., and Schmid, C. W. (1982) *Ann. Rev. Biochem.* **51**, 813–844
144. Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M., and Rodier, F. (1985) *Science* **228**, 953–956
145. Nowak, R. (1994) *Science* **263**, 608–610
145a. Razin, S. V., Ioudinkova, E. S., Trifonov, E. N., and Scherrer, K. (2001) *J. Mol. Biol.* **307**, 481–486
145b. Mashkova, T. D., Oparina, N. Y., Lacroix, M.-H., Fedorova, L. I., Tumeneva, I. G., Zinovieva, O. L., and Kisselev, L. L. (2001) *J. Mol. Biol.* **305**, 33–48
146. Charlesworth, B., Sniegowski, P., and Stephan, W. (1994) *Nature (London)* **371**, 215–220
147. Barakat, A., Matassi, G., and Bernardi, G. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10044–10049
148. Rosenberg, H., Singer, M., and Rosenberg, M. (1978) *Science* **200**, 394–402
149. Inoue, S., Kitajima, K., and Inoue, Y. (1996) *J. Biol. Chem.* **271**, 24341–24344
150. Pluta, A. F., Mackay, A. M., Ainsztein, A. M., Goldberg, I. G., and Earnshaw, W. C. (1995) *Science* **270**, 1591–1594
151. Xiao, Z.-X., and Fitzgerald-Hayes, M. (1995) *J. Mol. Biol.* **248**, 255–263
152. Tal, M., Shimron, F., and Yagil, G. (1994) *J. Mol. Biol.* **243**, 179–189
153. Ferrer, N., Azorín, F., Villasante, A., Gutiérrez, C., and Abad, J. P. (1995) *J. Mol. Biol.* **245**, 8–21
154. Catasti, P., Gupta, G., Garcia, A. E., Ratliff, R., Hong, L., Yau, P., Moyzis, R. K., and Bradbury, E. M. (1994) *Biochemistry* **33**, 3819–3830
155. Chou, S.-H., Zhu, L., and Reid, B. R. (1996) *J. Mol. Biol.* **259**, 445–457
156. Copenhaver, G. P., Nickel, K., Kuromori, T., Benito, M.-I., Kaul, S., Lin, X., Beven, M., Murphy, G., Harris, B., Parnell, L. D., McCombie, W. R., Martienssen, R. A., Marra, M., and Preuss, D. (1999) *Science* **286**, 2468–2474

References

157. Jiang, J., Nasuda, S., Dong, F., Scherrer, C. W., Woo, S.-S., Wing, R. A., Gill, B. S., and Ward, D. C. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14210–14213
157a. Schueler, M. G., Higgins, A. W., Rudd, M. K., Gustashaw, K., and Willard, H. F. (2001) *Science* **294**, 109–115
157b. Pennisi, E. (2001) *Science* **294**, 30–31
158. Zhu, L., Chou, S.-H., and Reid, B. R. (1995) *J. Mol. Biol.* **254**, 623–637
158a. Gao, Y.-G., Robinson, H., Sanishvili, R., Joachimiak, A., and Wang, A. H.-J. (1999) *Biochemistry* **38**, 16452–16460
158b. Henikoff, S., and Malik, H. S. (2002) *Nature (London)* **417**, 227
159. Pluta, A. F., Cooke, C. A., and Earnshaw, W. C. (1990) *Trends Biochem. Sci.* **15**, 181–185
160. Sugimoto, K., Hagishita, Y., and Himeno, M. (1994) *J. Biol. Chem.* **269**, 24271–24276
161. Pluta, A. F., and Earnshaw, W. C. (1996) *J. Biol. Chem.* **271**, 18767–18774
162. Fukagawa, T., Pendon, C., Morris, J., and Brown, W. (1999) *EMBO J.* **18**, 4196–4209
162a. Takahashi, K., Chen, E. S., and Yanagida, M. (2000) *Science* **288**, 2215–2219
163. Blackburn, E. H., and Szostak, J. W. (1984) *Ann. Rev. Biochem.* **53**, 163–194
164. Moyzis, R. K. (1991) *Sci. Am.* **265**(Aug), 48–55
165. Blackburn, E. H. (1991) *Trends Biochem. Sci.* **16**, 378–381
166. Zakian, V. A. (1995) *Science* **270**, 1601–1607
167. Vega-Palas, M. A., Venditti, S., and Di Mauro, E. (1998) *J. Biol. Chem.* **273**, 9388–9392
167a. de Lange, T. (2001) *Science* **292**, 1075–1076
167b. Stansel, R. M., de Lange, T., and Griffith, J. D. (2001) *EMBO J.* **20**, 5532–5540
167c. Tomaska, L., Makhov, A. M., Nosek, J., Kucejova, B., and Griffith, J. D. (2001) *J. Mol. Biol.* **305**, 61–69
167d. Mitton-Fry, R. M., Anderson, E. M., Hughes, T. R., Lundblad, V., and Wuttke, D. S. (2002) *Science* **296**, 145–147
168. Schierer, T., and Henderson, E. (1994) *Biochemistry* **33**, 2240–2246
169. Frantz, J. D., and Gilbert, W. (1995) *J. Biol. Chem.* **270**, 20692–20697
169a. Baumann, P., and Cech, T. R. (2001) *Science* **292**, 1171–1175
170. Shore, D. (1998) *Science* **281**, 1818–1819
171. König, P., and Rhodes, D. (1997) *Trends Biochem. Sci.* **22**, 43–47
171a. Rossetti, L., Cacchione, S., De Menna, A., Chapman, L., Rhodes, D., and Savino, M. (2001) *J. Mol. Biol.* **306**, 903–913
172. Ohshima, K., and Okada, N. (1994) *J. Mol. Biol.* **243**, 25–37
172a. Kim, J., Martignetti, J. A., Shen, M. R., Brosius, J., and Deininger, P. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3607–3611
173. Ünsal, K., and Morgan, G. T. (1995) *J. Mol. Biol.* **248**, 812–823
174. Izsvák, Z., Ivics, Z., Garcia-Estefania, D., Fahrenkrug, S. C., and Hackett, P. B. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1077–1081
175. Ullu, E. (1982) *Trends Biochem. Sci.* **7**, 216–219
176. Sharp, P. A. (1983) *Nature (London)* **301**, 471–472
177. Adams, D. S., Eickbash, T. H., Herrea, R. S., and Lizandi, P. M. (1986) *J. Mol. Biol.* **187**, 465–478
177a. Carroll, M. L., Roy-Engel, A. M., Nguyen, S. V., Salem, A.-H., Vogel, E., Vincent, B., Myers, J., Ahmad, Z., Nuguyen, L., Sammarco, M., Watkins, W. S., Henke, J., Makalowski, W., Jorde, L. B., Dininger, P. L., and Batzer, M. A. (2001) *J. Mol. Biol.* **311**, 17–40
177b. Martinez, J., Dugaiczyk, L. J., Zielinski, R., and Dugaiczyk, A. (2001) *J. Mol. Biol.* **308**, 587–596
178. Shimizu, Y., Yoshida, K., Ren, C.-S., Fujinaga, K., Rejagopalan, S., and Chinnadurai, G. (1983) *Nature (London)* **302**, 587–590
179. Kiyama, R., Matsui, H., and Oishi, M. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4665–4669
180. Fowler, R. F., and Skinner, D. M. (1985) *J. Biol. Chem.* **260**, 1296–1303
181. Nargang, F. E., Bell, J. B., Stohl, L. L., and Lambowitz, A. M. (1983) *J. Biol. Chem.* **258**, 4257–4260
182. Wildeman, A. G., Rasquinha, I., and Nazar, R. N. (1986) *J. Biol. Chem.* **261**, 13401–13403
183. Reis, R. J. S., Lumpkin, C. K., Jr., McGill, J. R., Riabowol, K. T., and Goldstein, S. (1983) *Nature (London)* **301**, 394–398
184. Casavant, N. C., and Hardies, S. C. (1994) *J. Mol. Biol.* **241**, 390–397
185. Dhellin, O., Maestre, J., and Heidmann, T. (1997) *EMBO J.* **16**, 6590–6602
186. Dunham, I., Shimizu, N., Roe, B. A., Chossoe, S., and 117 other authors. (1999) *Nature (London)* **402**, 489–495
187. Sapienza, C., and Doolittle, W. F. (1982) *Nature (London)* **295**, 384–389
188. Wellauer, P. K., Reeder, R. H., Carroll, D., Brown, D. D., Deutch, A., Higashinakagawa, T., and Dawid, I. B. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2823–2827
189. Carroll, D., and Brown, D. D. (1976) *Cell* **7**, 467–475
190. Mroczka, D. L., Cassidy, B., Busch, H., and Rothblum, L. E. (1984) *J. Mol. Biol.* **174**, 141–162
191. Sommerville, J. (1985) *Nature (London)* **318**, 410–411
192. Budarf, M. L., and Blackburn, E. H. (1986) *J. Biol. Chem.* **261**, 363–369
193. Karabin, G. D., and Hallick, R. B. (1983) *J. Biol. Chem.* **258**, 5512–5518
194. Mattaj, I. W. (1984) *Trends Biochem. Sci.* **9**, 435–437
195. Dandekar, T., Snel, B., Huynen, M., and Bork, P. (1998) *Trends Biochem. Sci.* **23**, 324–328
196. Karlsson, S., and Nienhuis, A. W. (1985) *Ann. Rev. Biochem.* **54**, 1071–1108
197. Proudfoot, N. (1986) *Nature (London)* **321**, 730–731
197a. Wang, Z., and Liebhaber, S. A. (1999) *EMBO J.* **18**, 2218–2228
197b. Li, J., Noguchi, C. T., Miller, W., Hardison, R., and Schechter, A. N. (1998) *J. Biol. Chem.* **273**, 10202–10209
197c. Filipe, A., Li, Q., Deveaux, S., Godin, I., Roméo, P.-H., Stamatoyannopoulos, G., and Mignotte, V. (1999) *EMBO J.* **18**, 687–697
197d. Ristaldi, M. S., Drabek, D., Gribnau, J., Poddie, D., Yannoutsous, N., Cao, A., Grosveld, F., and Imam, A. M. A. (2001) *EMBO J.* **20**, 5242–5249
198. Vanin, E. F., Goldberg, G. I., Tucker, P. W., and Smithies, O. (1980) *Nature (London)* **286**, 222–226
199. Strehler, E. E., Strehler-Page, M.-A., Prriard, J.-C., Periasamy, M., and Nadal-Ginard, B. (1986) *J. Mol. Biol.* **190**, 291–317
200. Kolata, G. B. (1977) *Science* **196**, 1187–1188
201. Purohie, S., and Mathews, C. K. (1984) *J. Biol. Chem.* **259**, 6261–6266
202. Rak, B., Lusky, M., and Hable, M. (1982) *Nature (London)* **297**, 124–128
203. Ernst, H., and Shatkin, A. J. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 48–52
204. Soininen, R., Huotari, M., Hostikka, S. L., Prockop, D. J., and Tryggvason, K. (1988) *J. Biol. Chem.* **263**, 17217–17220
205. Henikoff, S., Keene, M. A., Fechtel, K., and Fristrom, J. W. (1986) *Cell* **44**, 33–42
206. Borst, P., and Grivell, L. A. (1981) *Nature (London)* **289**, 439–440
207. Gilson, P. R., and McFadden, G. I. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7737–7742
208. Gray, M. W., Burger, G., and Lang, B. F. (1999) *Science* **283**, 1476–1481
209. Chen, K. K., and Donelson, J. E. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2445–2449
210. de la Cruz, V. F., Lake, J. A., Simpson, A. M., and Simpson, L. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1401–1405
211. Jeltsch, A., Friedrich, T., and Roth, M. (1998) *J. Mol. Biol.* **275**, 747–758
212. Newman, A. K., Rubin, R. A., Kim, S.-H., and Modrich, P. (1981) *J. Biol. Chem.* **256**, 2131–2139
213. Surby, M. A., and Reich, N. O. (1996) *Biochemistry* **35**, 2201–2208
213a. Gowher, H., and Jeltsch, A. (2000) *J. Mol. Biol.* **303**, 93–110
214. Schumann, J., Walter, J., Willert, J., Wild, C., Koch, D., and Trautner, T. A. (1996) *J. Mol. Biol.* **257**, 949–959
215. Ahmad, I., and Rao, D. N. (1996) *J. Mol. Biol.* **259**, 229–240
215a. Pues, H., Bleimling, N., Holz, B., Wölcke, J., and Weinhold, E. (1999) *Biochemistry* **38**, 1426–1434
216. Erlanson, D. A., Chen, L., and Verdine, G. L. (1993) *J. Am. Chem. Soc.* **115**, 12583–12584
217. Lau, E. Y., and Bruice, T. C. (1999) *J. Mol. Biol.* **293**, 9–18
218. Powell, L. M., Connolly, B. A., and Dryden, D. T. F. (1998) *J. Mol. Biol.* **283**, 947–961
219. Herman, G. E., and Modrich, P. (1982) *J. Biol. Chem.* **257**, 2605–2612
220. Zweiger, G., Marczynski, G., and Shapiro, L. (1994) *J. Mol. Biol.* **235**, 472–485
221. Heithoff, D. M., Sinsheimer, R. L., Low, D. A., and Mahan, M. J. (1999) *Science* **284**, 967–970
222. Burdon, R. H., and Adams, R. L. P. (1980) *Trends Biochem. Sci.* **5**, 294–297
223. Doerfler, W. (1983) *Ann. Rev. Biochem.* **52**, 93–124
224. Antequera, F., Tamame, M., Villaneuva, J. R., and Santos, T. (1984) *J. Biol. Chem.* **259**, 8033–8036
225. Selker, E. U., Jensen, B. C., and Richardson, G. A. (1987) *Science* **238**, 48–53
226. Holliday, R. (1989) *Sci. Am.* **260**(Jun), 60–73
227. Hsu, D.-W., Lin, M.-J., Lee, T.-L., Wen, S.-C., Chen, X., and Shen, C.-K., J. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9751–9756
228. Yoder, J. A., Soman, N. S., Verdine, G. L., and Bestor, T. H. (1997) *J. Mol. Biol.* **270**, 385–395
229. Pradhan, S., Bacolla, A., Wells, R. D., and Roberts, R. J. (1999) *J. Biol. Chem.* **274**, 33002–33010
229a. Gowher, H., and Jeltsch, A. (2001) *J. Mol. Biol.* **309**, 1201–1208
229b. Saito, Y., Kanai, Y., Sakamoto, M., Saito, H., Ishii, H., and Hirohashi, S. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10060–10065
230. Gjerset, R. A., and Martin, D. W., Jr. (1982) *J. Biol. Chem.* **257**, 8581–8583
231. Ramchandani, S., Bhattacharya, S. K., Cervoni, N., and Szyf, M. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6107–6112
232. Cervoni, N., Bhattacharya, S., and Szyf, M. (1999) *J. Biol. Chem.* **274**, 8363–8366
233. Matsuo, K., Silke, J., Georgiev, O., Marti, P., Giovannini, N., and Rungger, D. (1998) *EMBO J.* **17**, 1446–1453
233a. Smith, S. S. (2000) *J. Mol. Biol.* **302**, 1–7
233b. Bhattacharya, S. K., Ramchandani, S., Cervoni, N., and Szyf, M. (1999) *Nature (London)* **397**, 579–583
233c. Cedar, H., and Verdine, G. L. (1999) *Nature (London)* **397**, 568–569
234. Bird, A. P. (1986) *Nature (London)* **321**, 209–213

References

235. Shiraishi, M., Lerman, L. S., and Sekiya, T. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4229–4233
236. Pieper, R. O., Patel, S., Ting, S. A., Futscher, B. W., and Costello, J. F. (1996) *J. Biol. Chem.* **271**, 13916–13924
237. Rein, T., Kobayashi, T., Malott, M., Leffak, M., and DePamphilis, M. L. (1999) *J. Biol. Chem.* **274**, 25792–25800
238. Mayer-Jung, C., Moras, D., and Timsit, Y. (1998) *EMBO J.* **17**, 2709–2718
239. Wakefield, R. I. D., Smith, B. O., Nan, X., Free, A., Soteriou, A., Uhrin, D., Bird, A. P., and Barlow, P. N. (1999) *J. Mol. Biol.* **291**, 1055–1065
240. Davey, C., Pennings, S., and Allan, J. (1997) *J. Mol. Biol.* **267**, 276–288
241. Bester, T. H. (1998) *Nature (London)* **393**, 311–312
242. Razin, A. (1998) *EMBO J.* **17**, 4905–4908
243. Naruse, Y., Aoki, T., Kojima, T., and Mori, N. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13691–13696
244. O'Neill, L. P., Keohane, A. M., Lavender, J. S., McCabe, V., Heard, E., Avner, P., Brockdorff, N., and Turner, B. M. (1999) *EMBO J.* **18**, 2897–2907
245. Herzing, L. B. K., Romer, J. T., Horn, J. M., and Ashworth, A. (1997) *Nature (London)* **386**, 272–275
246. Jegalian, K., and Page, D. C. (1998) *Nature (London)* **394**, 776–780
247. Gommers-Ampt, J. H., and Borst, P. (1995) *FASEB J.* **9**, 1034–1042
248. Wilson, V. L., Smith, R. A., Ma, S., and Cutler, R. G. (1987) *J. Biol. Chem.* **262**, 9948–9951
249. Bakin, A. V., and Curran, T. (1999) *Science* **283**, 387–390
250. Song, H. K., Sohn, S. H., and Suh, S. W. (1999) *EMBO J.* **18**, 1104–1113
251. Sapienza, C. (1990) *Sci. Am.* **263**(Oct), 52–60
252. Barlow, D. P. (1993) *Science* **260**, 309–310
253. Birger, Y., Shemer, R., Perk, J., and Razin, A. (1999) *Nature (London)* **397**, 84–88
254. Meselson, M., and Stahl, F. W. (1958) *Proc. Natl. Acad. Sci. U.S.A.* **44**, 671–682
255. Taylor, J. H. (1997) *Trends Biochem. Sci.* **22**, 447–450
256. Cairns, J. (1963) *J. Mol. Biol.* **6**, 208–213
257. Zimm, B. H. (1999) *Trends Biochem. Sci.* **24**, 121–123
258. Kornberg, A. (1969) *Science* **163**, 1410–1418
259. Lehman, I. R., Bessman, M. J., Simms, E. S., and Kornberg, A. (1958) *J. Biol. Chem.* **233**, 163–170
259a. Patel, P. H., Suzuki, M., Adman, E., Shinkai, A., and Loeb, L. A. (2001) *J. Mol. Biol.* **308**, 823–837
260. Sugimoto, K., Okazaki, T., and Okazaki, R. (1968) *Proc. Natl. Acad. Sci. U.S.A.* **60**, 1356–1362
261. Modrich, P., Anraku, Y., and Lehman, I. R. (1973) *J. Biol. Chem.* **248**, 7495–7501
262. Lehman, I. R. (1974) *Science* **186**, 790–797
263. Brutlag, D., Schekman, R., and Kornberg, A. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 2826–2829
264. Cai, H., Yu, H., McEntee, K., Kunkel, T. A., and Goodman, M. F. (1995) *J. Biol. Chem.* **270**, 15327–15335
265. Kornberg, A., and Baker, T. A. (1992) *DNA Replication*, 2nd ed., Freeman, New York
266. Hubscher, U., and Spadari, S., eds. (1984) *Proteins Involved in DNA Replication*, Plenum, New York
266a. Friedberg, E. C., Feaver, W. J., and Gerlach, V. L. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5681–5683
266b. Burgers, P. M. J., Koonin, E. V., Bruford, E., Blanco, L., Burtis, K. C., Christman, M. F., Copeland, W. C., Friedberg, E. C., and 13 other authors. (2001) *J. Biol. Chem.* **276**, 43487–43490
267. Steitz, T. A. (1999) *J. Biol. Chem.* **274**, 17395–17398
268. Sousa, R. (1996) *Trends Biochem. Sci.* **21**, 186–190
268a. Yang, G., Franklin, M., Li, J., Lin, T.-C., and Konigsberg, W. (2002) *Biochemistry* **41**, 10256–10261
268b. Hübscher, U., Maga, G., and Spadari, S. (2002) *Ann. Rev. Biochem.* **71**, 133–163
269. Brody, R. S., and Frey, P. A. (1981) *Biochemistry* **20**, 1245–1252
270. Steitz, T. A. (1998) *Nature (London)* **391**, 231–232
271. Astatke, M., Ng, K., Grindley, N. D. F., and Joyce, C. M. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3402–3407
272. Singh, K., and Modak, M. J. (1998) *Trends Biochem. Sci.* **23**, 277–281
272a. Krishna, T. S. R., Kong, X-P, Gary, S., Burgers, P. M., and Kuriyan, J. (1994) *Cell* **79**, 1233–1243
273. Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H., and Kraut, J. (1994) *Science* **264**, 1930–1935
274. Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J., and Pelletier, H. (1997) *Biochemistry* **36**, 11205–11215
275. Beese, L. S., Derbyshire, V., and Steitz, T. A. (1993) *Science* **260**, 352–355
275a. Spratt, T. E. (2001) *Biochemistry* **40**, 2647–2652
276. Eom, S. H., Wang, J., and Steitz, T. A. (1996) *Nature (London)* **382**, 278–281
277. Li, Y., Mitaxov, V., and Waksman, G. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9491–9496
278. Tabor, S., and Richardson, C. C. (1990) *J. Biol. Chem.* **265**, 8322–8328
279. Sousa, R., Chung, Y. J., Rose, J. P., and Wang, B.-C. (1993) *Nature (London)* **364**, 593–599
280. Osumi-Davis, P. A., Sreerama, N., Volkin, D. B., Middaugh, C. R., Woody, R. W., and Woody, A.-Y. M. (1994) *J. Mol. Biol.* **237**, 5–19
280a. Longley, M. J., Ropp, P. A., Lim, S. E., and Copeland, W. C. (1998) *Biochemistry* **37**, 10529–10539
280b. Fan, L., and Kaguni, L. S. (2001) *Biochemistry* **40**, 4780–4791
281. Kelman, Z., Pietrokovski, S., and Hurwitz, J. (1999) *J. Biol. Chem.* **274**, 28751–28761
282. Hopfner, K.-P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R., and Angerer, B. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3600–3605
282a. Rodriguez, A. C., Park, H.-W., Mao, C., and Beese, L. S. (2000) *J. Mol. Biol.* **299**, 447–462
283. Maga, G., and Hübscher, U. (1996) *Biochemistry* **35**, 5764–5777
284. Bambara, R. A., Murante, R. S., and Henriksen, L. A. (1997) *J. Biol. Chem.* **272**, 4647–4650
285. Mozzherin, D. J., and Fisher, P. A. (1996) *Biochemistry* **35**, 3572–3577
286. Dua, R., Levy, D. L., and Campbell, J. L. (1998) *J. Biol. Chem.* **273**, 30046–30055
286a. MacNeill, S. A., Baldacci, G., Burgers, P. M., and Hübscher, U. (2001) *Trends Biochem. Sci.* **26**, 16–17
286b. Xie, B., Mazloum, N., Liu, L., Rahmeh, A., Li, H., and Lee, M. Y. W. T. (2002) *Biochemistry* **41**, 13133–13142
287. Alley, S. C., Trakselis, M. A., Mayer, M. U., Ishmael, F. T., Jones, A. D., and Benkovic, S. J. (2001) *J. Biol. Chem.* **276**, 39340–39349
287a. Pietroni, P., Young, M. C., Latham, G. J., and von Hippel, P. H. (2001) *J. Mol. Biol.* **309**, 869–891
288. Davies, J. F., II, Hostomska, Z., Hostomsky, Z., Jordan, S. R., and Matthews, D. A. (1991) *Science* **252**, 88–95
289. Wang, J., Smerdon, S. J., Jäger, J., Kohlstaedt, L. A., Rice, P. A., Friedman, J. M., and Steitz, T. A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7242–7246
289a. Sarafianos, S. G., Das, K., Tantillo, C., Clark, A. D., Jr., Ding, J., Whitcomb, J. M., Boyer, P. L., Hughes, S. H., and Arnold, E. (2001) *EMBO J.* **20**, 1449–1461
290. Tantillo, C., Ding, J., Jacobo-Molina, A., Nanni, R. G., Boyer, P. L., Hughes, S. H., Pauwels, R., Andries, K., Janssen, P. A. J., and Arnold, E. (1994) *J. Mol. Biol.* **243**, 369–387
291. Carroll, S. S., Olsen, D. B., Bennett, C. D., Gotlib, L., Graham, D. J., Condra, J. H., Stern, A. M., Shafer, J. A., and Kuo, L. C. (1993) *J. Biol. Chem.* **268**, 276–281
292. Mathias, S. L., Scott, A. F., Kazazian, H. H., Jr., Boeke, J. D., and Gabriel, A. (1991) *Science* **254**, 1808–1810
293. Pelletier, H., Sawaya, M. R., Wolfle, W., Wilson, S. H., and Kraut, J. (1996) *Biochemistry* **35**, 12742–12761
293a. Shah, A. M., Conn, D. A., Li, S.-X., Capaldi, A., Jäger, J., and Sweasy, J. B. (2001) *Biochemistry* **40**, 11372–11381
293b. Liu, J., and Tsai, M.-D. (2001) *Biochemistry* **40**, 9014–9022
293c. Eckert, K. A., Mowery, A., and Hile, S. E. (2002) *Biochemistry* **41**, 10490–10498
294. Holm, L., and Sander, C. (1995) *Trends Biochem. Sci.* **20**, 345–347
294a. Song, M.-S., Pham, P. T., Olson, M., Carter, J. R., Franden, M. A., Schaaper, R. M., and McHenry, C. S. (2001) *J. Biol. Chem.* **276**, 35165–35175
295. Kornberg, A. (1988) *J. Biol. Chem.* **263**, 1–4
296. Kelman, Z., and O'Donnell, M. (1995) *Ann. Rev. Biochem.* **64**, 171–200
297. Kim, D. R., and McHenry, C. S. (1996) *J. Biol. Chem.* **271**, 20699–20704
298. Perrino, F. W., Harvey, S., and McNeill, S. M. (1999) *Biochemistry* **38**, 16001–16009
299. Turner, J., Hingorani, M. M., Kelman, Z., and O'Donnell, M. (1999) *EMBO J.* **18**, 771–783
300. Brenowitz, S., Kwack, S., Goodman, M. F., O'Donnell, M., and Echols, H. (1991) *J. Biol. Chem.* **266**, 7888–7892
300a. DeRose, E. F., Li, D., Darden, T., Harvey, S., Perrino, F. W., Schaaper, R. M., and London, R. E. (2002) *Biochemistry* **41**, 94–110
301. LaDuca, R. J., Crute, J. J., McHenry, C. S., and Bambara, R. A. (1986) *J. Biol. Chem.* **261**, 7550–7557
301a. Leu, F. P., and O'Donnell, M. (2001) *J. Biol. Chem.* **276**, 47185–47194
301b. Matsumiya, S., Ishino, Y., and Morikawa, K. (2001) *Protein Sci.* **10**, 17–23
301c. Bertram, J. G., Bloom, L. B., Hingorani, M. M., Beechem, J. M., O'Donnell, M., and Goodman, M. F. (2000) *J. Biol. Chem.* **275**, 28413–28420
302. Hingorani, M. M., Bloom, L. B., Goodman, M. F., and O'Donnell, M. (1999) *EMBO J.* **18**, 5131–5144
303. Kuriyan, J., and O'Donnell, M. (1993) *J. Mol. Biol.* **234**, 915–925
303a. Dervyn, E., Suski, C., Daniel, R., Bruand, C., Chapuis, J., Errington, J., Jannière, L., and Ehrlich, S. D. (2001) *Science* **294**, 1716–1718
304. Lindahl, T., and Barnes, D. E. (1992) *Ann. Rev. Biochem.* **61**, 251–281

References

305. Husain, I., Tomkinson, A. E., Burkhart, W. A., Moyer, M. B., Ramos, W., Mackey, Z. B., Besterman, J. M., and Chen, J. (1995) *J. Biol. Chem.* **270**, 9683–9690
306. Arabshahi, A., and Frey, P. A. (1999) *J. Biol. Chem.* **274**, 8586–8588
307. Doherty, A. J., and Wigley, D. B. (1999) *J. Mol. Biol.* **285**, 63–71
307a. Doherty, A. J., and Dafforn, T. R. (2000) *J. Mol. Biol.* **296**, 43–56
308. Shamoo, Y., Friedman, A. M., Parsons, M. R., Konigsberg, W. H., and Steitz, T. A. (1995) *Nature (London)* **376**, 362–366
309. Folmer, R. H. A., Nilges, M., Konings, R. N. H., and Hilbers, C. W. (1995) *EMBO J.* **14**, 4132–4142
310. Guan, Y., Zhang, H., and Wang, A. H.-J. (1995) *Protein Sci.* **4**, 187–197
311. Olah, G. A., Gray, D. M., Gray, C. W., Kergil, D. L., Sosnick, T. R., Mark, B. L., Vaughan, M. R., and Trewhella, J. (1995) *J. Mol. Biol.* **249**, 576–594
312. Lohman, T. M., and Ferrari, M. E. (1994) *Ann. Rev. Biochem.* **63**, 527–570
313. Raghunathan, S., Ricard, C. S., Lohman, T. M., and Waksman, G. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6652–6657
314. Lohman, T. M., Bujalowski, W., and Overman, L. B. (1988) *Trends Biochem. Sci.* **13**, 250–255
315. Cobianchi, F., SenGupta, D. N., Zmudzka, B. Z., and Wilson, S. H. (1986) *J. Biol. Chem.* **261**, 3536–3543
316. Richter, A., Sapp, M., and Knippers, R. (1986) *Trends Biochem. Sci.* **11**, 283
317. Kelly, T. J., Simancek, P., and Brush, G. S. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14634–14639
318. Subramanya, H. S., Bird, L. E., Brannigan, J. A., and Wigley, D. B. (1996) *Nature (London)* **384**, 379–383
319. Lohman, T. M., and Bjornson, K. P. (1996) *Ann. Rev. Biochem.* **65**, 169–214
320. Marians, K. J. (1997) *Structure* **5**, 1129–1134
320a. Soultanas, P., and Wigley, D. B. (2001) *Trends Biochem. Sci.* **26**, 47–54
320b. Soultanas, P., Dillingham, M. S., Wiley, P., Webb, M. R., and Wigley, D. B. (2000) *EMBO J.* **19**, 3799–3810
321. Yu, X., Jezewska, M. J., Bujalowski, W., and Egelman, E. H. (1996) *J. Mol. Biol.* **259**, 7–14
321a. Donate, L.-E., Llorca, O., Bárcena, M., Brown, S. E., Dixon, N. E., and Carazo, J.-M. (2000) *J. Mol. Biol.* **303**, 383–393
321b. Kaplan, D. L. (2000) *J. Mol. Biol.* **301**, 285–299
322. Jezewska, M. J., Rajendran, S., Bujalowska, D., and Bujalowski, W. (1998) *J. Biol. Chem.* **273**, 10515–10529
323. Hsieh, J., Moore, K. J. M., and Lohman, T. M. (1999) *J. Mol. Biol.* **288**, 255–274
324. Biswas, E. E., and Biswas, S. B. (1999) *Biochemistry* **38**, 10929–10939
325. Bárcena, M., San Martín, C., Weise, F., Ayora, S., Alonso, J. C., and Carazo, J. M. (1998) *J. Mol. Biol.* **283**, 809–819
325a. Niedenzu, T., Röleke, D., Bains, G., Scherzinger, E., and Saenger, W. (2001) *J. Mol. Biol.* **306**, 479–487
326. Richardson, R. W., and Nossal, N. G. (1989) *J. Biol. Chem.* **264**, 4725–4731
326a. Jones, C. E., Mueser, T. C., and Nossal, N. G. (2000) *J. Biol. Chem.* **275**, 27145–27154
326b. Delagoutte, E., and von Hippel, P. H. (2001) *Biochemistry* **40**, 4459–4477
326c. Harmon, F. G., and Kowalczykowski, S. C. (2001) *J. Biol. Chem.* **276**, 232–243
327. Stasiak, A., Tsaneva, I. R., West, S. C., Yu, X., and Engelman, E. H. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7618–7622
327a. Parsons, C. A., Stasiak, A., Bennet, R. J., and West, S. C. (1995) *Nature (London)* **374**, 375–378
327b. Yu, X., West, S. C., and Egelman, E. H. (1997) *J. Mol. Biol.* **266**, 217–222
328. West, S. C. (1996) *Nature (London)* **384**, 316–317
329. Seo, Y.-S., and Hurwitz, J. (1993) *J. Biol. Chem.* **268**, 10282–10295
330. Poll, E. H. A., Harrison, J., Umthun, A., Dobbs, D. L., and Benbow, R. M. (1994) *Biochemistry* **33**, 3841–3847
331. Karow, J. K., Chakraverty, R. K., and Hickson, I. D. (1997) *J. Biol. Chem.* **272**, 30611–30614
331a. Wang, W., Seki, M., Narita, Y., Sonoda, E., Takeda, S., Yamada, K., Masuko, T., Katada, T., and Enomoto, T. (2000) *EMBO J.* **19**, 3428–3435
331b. Wu, L., and Hickson, I. D. (2001) *Science* **292**, 229–230
332. Braguglia, D., Heun, P., Pasero, P., Duncker, B. P., and Gasser, S. M. (1998) *J. Mol. Biol.* **281**, 631–649
332a. Orren, D. K., Theodore, S., and Machwe, A. (2002) *Biochemistry* **41**, 13483–13488
332b. Egelman, E. H. (2001) *Nature (London)* **409**, 573,575
333. de la Cruz, J., Kressler, D., Tollervey, D., and Linder, P. (1998) *EMBO J.* **17**, 1128–1140
334. Gorbalenya, A. E., and Koonin, E. V. (1993) *Current Opinion in Structural Biology* **3**, 419–429
335. Wang, J. C. (1996) *Ann. Rev. Biochem.* **65**, 635–692
336. Roca, J. (1995) *Trends Biochem. Sci.* **20**, 156–160
337. Cheng, C., and Shuman, S. (1999) *Biochemistry* **38**, 16599–16612
338. Chen, S.-J., and Wang, J. C. (1998) *J. Biol. Chem.* **273**, 6050–6056
339. Stewart, L., Redinbo, M. R., Qiu, X., Hol, W. G. J., and Champoux, J. J. (1998) *Science* **279**, 1534–1541
340. Redinbo, M. R., Stewart, L., Champoux, J. J., and Hol, W. G. J. (1999) *J. Mol. Biol.* **292**, 685–696
340a. Redinbo, M. R., Champoux, J. J., and Hol, W. G. J. (2000) *Biochemistry* **39**, 6832–6840
340b. Dekker, N. H., Rybenkov, V. V., Duguet, M., Crisona, N. J., Cozzarelli, N. R., Bensimon, D., and Croquette, V. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12126–12131
340c. Champoux, J. J. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11998–12000
341. Lima, C. D., Wang, J. C., and Mondragón, A. (1994) *Nature (London)* **367**, 138–146
341a. Changela, A., DiGate, R. J., and Mondragón, A. (2001) *Nature (London)* **411**, 1077–1081
342. Olland, S., and Wang, J. C. (1999) *J. Biol. Chem.* **274**, 21688–21694
343. Smith, C. V., and Maxwell, A. (1998) *Biochemistry* **37**, 9658–9667
344. Williams, N. L., and Maxwell, A. (1999) *Biochemistry* **38**, 13502–13511
345. Nichols, M. D., DeAngelis, K., Keck, J. L., and Berger, J. M. (1999) *EMBO J.* **18**, 6177–6188
346. Huang, W. M., Wei, L. S., and Casjens, S. (1985) *J. Biol. Chem.* **260**, 8973–8977
347. Li, Z., Hiasa, H., Kumar, U., and DiGate, R. J. (1997) *J. Biol. Chem.* **272**, 19582–19587
348. Baird, C. L., Harkins, T. T., Morris, S. K., and Lindsley, J. E. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13685–13690
348a. Williams, N. L., Howells, A. J., and Maxwell, A. (2001) *J. Mol. Biol.* **306**, 969–984
349. Morris, S. K., and Lindsley, J. E. (1999) *J. Biol. Chem.* **274**, 30690–30696
350. Berger, J. M., Gamblin, S. J., Harrison, S. C., and Wang, J. C. (1996) *Nature (London)* **379**, 225–232
351. Berger, J. M., Fass, D., Wang, J. C., and Harrison, S. C. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7876–7881
352. Kikuchi, A., and Asai, K. (1984) *Nature (London)* **309**, 677–681
353. North, G. (1985) *Nature (London)* **316**, 394–395
353a. Barthelmes, H. U., Grue, P., Feineis, S., Straub, T., and Boege, F. (2000) *J. Biol. Chem.* **275**, 38823–38830
354. Froelich-Ammon, S. J., and Osheroff, N. (1995) *J. Biol. Chem.* **270**, 21429–21432
355. Sogo, J. M., Stasiak, A., Martínez-Robles, M. L., Krimer, D. B., Hernández, P., and Schvartzman, J. B. (1999) *J. Mol. Biol.* **286**, 637–643
356. Fleischmann, G., Pflugfelder, G., Steiner, E. K., Javaherian, K., Howard, G. C., Wang, J. C., and Elgin, S. C. R. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6958–6962
357. Brooks, M., and Dumas, L. B. (1989) *J. Biol. Chem.* **264**, 3602–3610
358. Massé, E., and Drolet, M. (1999) *J. Mol. Biol.* **294**, 321–332
359. Bailly, C., Carrasco, C., Hamy, F., Vezin, H., Prudhomme, M., Saleem, A., and Rubin, E. (1999) *Biochemistry* **38**, 8605–8611
360. Shero, J. H., Bordwell, B., Rothfield, N. F., and Earnshaw, W. C. (1986) *Science* **231**, 737–740
361. Hiasa, H., and Marians, K. J. (1999) *J. Biol. Chem.* **274**, 27244–27248
362. Sun, W., and Godson, G. N. (1998) *J. Mol. Biol.* **276**, 689–703
362a. Keck, J. L., Roche, D. D., Lynch, A. S., and Berger, J. M. (2000) *Science* **287**, 2482–2486
363. Jing, D. H., Dong, F., Latham, G. J., and von Hippel, P. H. (1999) *J. Biol. Chem.* **274**, 27287–27298
364. Frick, D. N., and Richardson, C. C. (1999) *J. Biol. Chem.* **274**, 35889–35898
365. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature (London)* **391**, 251–258
366. Hacker, K. J., and Johnson, K. A. (1997) *Biochemistry* **36**, 14080–14087
366a. Picha, K. M., Ahnert, P., and Patel, S. S. (2000) *Biochemistry* **39**, 6401–6409
366b. VanLoock, M. S., Chen, Y.-J., Yu, X., Patel, S. S., and Egelman, E. H. (2001) *J. Mol. Biol.* **311**, 951–956
367. Skarstad, K., Thöny, B., Hwang, D. S., and Kornberg, A. (1993) *J. Biol. Chem.* **268**, 5365–5370
368. Tougu, K., Peng, H., and Marians, K. J. (1994) *J. Biol. Chem.* **269**, 4675–4682
369. Tougu, K., and Marians, K. J. (1996) *J. Biol. Chem.* **271**, 21398–21405
370. Komori, H., Matsunaga, F., Higuchi, Y., Ishiai, M., Wada, C., and Miki, K. (1999) *EMBO J.* **18**, 4597–4607
371. Ng, J. Y., and Marians, K. J. (1996) *J. Biol. Chem.* **271**, 15642–15648
372. Allen, G. C., Jr., and Kornberg, A. (1993) *J. Biol. Chem.* **268**, 19204–19209
373. Hupp, T. R., and Kaguni, J. M. (1993) *J. Biol. Chem.* **268**, 13143–13150
374. Katayanagi, K., Ishikawa, M., Okumura, M., Ariyoshi, M., Kanaya, S., Kawano, Y., Suzuki, M., Tanaka, I., and Morikawa, K. (1993) *J. Biol. Chem.* **268**, 22092–22099
375. Keck, J. L., Goedken, E. R., and Marqusee, S. (1998) *J. Biol. Chem.* **273**, 34128–34133
376. Ohtani, N., Haruki, M., Morikawa, M., Crouch, R. J., Itaya, M., and Kanaya, S. (1999) *Biochemistry* **38**, 605–618
377. Bird, R. E., Louarn, J., Martuscelli, J., and Caro, L. (1972) *J. Mol. Biol.* **70**, 549–566
378. Prescott, D. M., and Kuempel, P. L. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2842–2845

References

379. Kuempel, P. L. et al. (1972) in *DNA Synthesis in Vitro* (Wells, R., and Inman, R., eds), pp. 463–472, University Park Press, Baltimore, Maryland
380. Kriegstein, H. J., and Hogness, D. S. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 135–139
381. Hixson, J. E., Wong, T. W., and Clayton, D. A. (1986) *J. Biol. Chem.* **261**, 2384–2390
382. Lee, D. Y., and Clayton, D. A. (1998) *J. Biol. Chem.* **273**, 30614–30621
382a. Krüger, R., Konieczny, I., and Filutowicz, M. (2001) *J. Mol. Biol.* **306**, 945–955
383. McMacken, R., Silver, L., and Georgopoulos, C. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), pp. 565–612, Am. Soc. Microbiol., Washington,DC
383a. Smelkova, N., and Marians, K. J. (2001) *J. Biol. Chem.* **276**, 39186–39191
384. Krause, M., Rückert, B., Lurz, R., and Messer, W. (1997) *J. Mol. Biol.* **274**, 365–380
385. Lobry, J. R. (1996) *Science* **272**, 745–746
385a. Kelman, Z. (2000) *Trends Biochem. Sci.* **25**, 521–523
386. Yung, B. Y.-M., and Kornberg, A. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7202–7205
387. Sutton, M. D., Carr, K. M., Vicente, M., and Kaguni, J. M. (1998) *J. Biol. Chem.* **273**, 34255–34262
388. Sutton, M. D., and Kaguni, J. M. (1997) *J. Mol. Biol.* **274**, 546–561
389. Radzicka, A., and Wolfenden, R. (1988) *Biochemistry* **27**, 1664–1670
390. Zyskind, J. W., Cleary, J. M., Brusilow, W. S. A., Harding, N. E., and Smith, D. W. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1164–1168
391. Chédin, F., Seitz, E. M., and Kowalczykowski, S. C. (1998) *Trends Biochem. Sci.* **23**, 273–277
391a. Bochkareva, E., Belegu, V., Korolev, S., and Bochkarev, A. (2001) *EMBO J.* **20**, 612–618
392. Lother, H., and Messer, W. (1981) *Nature (London)* **294**, 376–378
393. Xia, W., and Dowhan, W. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 783–787
394. York, J. D., Saffitz, J. E., and Majerus, P. W. (1994) *J. Biol. Chem.* **269**, 19992–19999
394a. Li, X., and Marians, K. J. (2000) *J. Biol. Chem.* **275**, 34757–34765
395. Maki, H., Maki, S., and Kornberg, A. (1988) *J. Biol. Chem.* **263**, 6570–6578
396. Lasken, R. S., and Kornberg, A. (1988) *J. Biol. Chem.* **263**, 5512–5518
396a. Chapados, B. R., Chai, Q., Hosfield, D. J., Qiu, J., Shen, B., and Tainer, J. A. (2001) *J. Mol. Biol.* **307**, 541–556
397. O'Donnell, M. E., and Kornberg, A. (1985) *J. Biol. Chem.* **260**, 12884–12889
398. Lee, E. H., and Kornberg, A. (1992) *J. Biol. Chem.* **267**, 8778–8784
399. Kamada, K., Horiuchi, T., Ohsumi, K., Shimamoto, N., and Morikawa, K. (1996) *Nature (London)* **383**, 598–603
400. Duggin, I. G., Andersen, P. A., Smith, M. T., Wilce, J. A., King, G. F., and Wake, R. G. (1999) *J. Mol. Biol.* **286**, 1325–1335
401. Manna, A. C., Pai, K. S., Bussiere, D. E., White, S. W., and Bastia, D. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3253–3258
402. Yamazoe, M., Onogi, T., Sunako, Y., Niki, H., Yamanaka, K., Ichimura, T., and Hiraga, S. (1999) *EMBO J.* **18**, 5873–5884
403. Dodson, M., McMacken, R., and Echols, H. (1989) *J. Biol. Chem.* **264**, 10719–10725
404. Learn, B. A., Um, S.-J., Huang, L., and McMacken, R. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1154–1159
405. Wickner, S., and McKenney, K. (1987) *J. Biol. Chem.* **262**, 13163–13167
406. Echols, H. (1990) *J. Biol. Chem.* **265**, 14697–14700
407. Murialdo, H. (1991) *Ann. Rev. Biochem.* **60**, 125–153
408. Wickner, S. H. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2690–2694
409. Wickner, S., Skowyra, D., Hoskins, J., and McKenney, K. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10345–10349
410. Nossal, N. G. (1992) *FASEB J.* **6**, 871–878
411. Brown, D. R., Roth, M. J., Reinberg, D., and Hurwitz, J. (1984) *J. Biol. Chem.* **259**, 10545–10555
412. Higashitani, A., Greenstein, D., Hirokawa, H., Asano, S., and Horiuchi, K. (1994) *J. Mol. Biol.* **237**, 388–400
413. Masai, H., Nomura, N., Kubota, Y., and Arai, K.-i. (1990) *J. Biol. Chem.* **265**, 15124–15133
414. Abarzua, P., Soeller, W., and Marians, K. J. (1984) *J. Biol. Chem.* **259**, 14286–14292
415. Roth, M. J., Brown, D. R., and Hurwitz, J. (1984) *J. Biol. Chem.* **259**, 10556–10568
416. Mok, M., and Marians, K. J. (1987) *J. Biol. Chem.* **262**, 2304–2309
416a. Campos-Olivas, R., Louis, J. M., CLérot, D., Gronenborn, B., and Gronenborn, A. M. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10310–10315
417. Cue, D., and Feiss, M. (1998) *J. Mol. Biol.* **280**, 11–29
418. Yang, Q., Berton, N., Manning, M. C., and Catalano, C. E. (1999) *Biochemistry* **38**, 14238–14247
419. Higgins, R. R., and Becker, A. (1995) *J. Mol. Biol.* **252**, 31–46
420. Yeo, A., and Feiss, M. (1995) *J. Mol. Biol.* **245**, 126–140
420a. Cue, D., and Feiss, M. (2001) *J. Mol. Biol.* **311**, 233–240
421. Franklin, J. L., Haseltine, D., Davenport, L., and Mosig, G. (1998) *J. Mol. Biol.* **277**, 541–557
421a. Rao, V. B., and Mitchell, M. S. (2001) *J. Mol. Biol.* **314**, 401–411
422. Marvin, D. A. (1998) *Current Opinion in Structural Biology* **8**, 150–158
423. Papavoine, C. H. M., Christiaans, B. E. C., Folmer, R. H. A., Konings, R. N. H., and Hilbers, C. W. (1998) *J. Mol. Biol.* **282**, 401–419
424. Rakonjac, J., Feng, J.-n, and Model, P. (1999) *J. Mol. Biol.* **289**, 1253–1265
425. Burch, A. D., Ta, J., and Fane, B. A. (1999) *J. Mol. Biol.* **286**, 95–104
426. Dokland, T., Bernal, R. A., Burch, A., Pletnev, S., Fane, B. A., and Rossmann, M. G. (1999) *J. Mol. Biol.* **288**, 595–608
427. Chai, S., Lurz, R., and Alonso, J. C. (1995) *J. Mol. Biol.* **252**, 386–398
428. Valle, M., Kremer, L., Martínez-A, C., Roncal, F., Valpuesta, J. M., Albar, J. P., and Carrascosa, J. L. (1999) *J. Mol. Biol.* **288**, 899–909
429. Caetano de Sousa, P., Jr., Tuma, R., Prevelige, P. E., Jr., Silva, J. L., and Foguel, D. (1999) *J. Mol. Biol.* **287**, 527–538
430. Grimes, S., and Anderson, D. (1997) *J. Mol. Biol.* **266**, 901–914
430a. Rodríguez-Casado, A., Moore, S. D., Prevelige, P. E., Jr., and Thomas, G. J., Jr. (2001) *Biochemistry* **40**, 13583–13591
430b. Simpson, A. A., Tao, Y., Leiman, P. G., Badasso, M. O., He, Y., Jardine, P. J., Olson, N. H., Morais, M. C., Grimes, S., Anderson, D. L., Baker, T. S., and Rossmann, M. G. (2000) *Nature (London)* **408**, 745–750
430c. Smith, D. E., Tans, S. J., Smith, S. B., Grimes, S., Anderson, D. L., and Bustamante, C. (2001) *Nature (London)* **413**, 748–752
431. Leffers, G., and Rao, V. B. (1996) *J. Mol. Biol.* **258**, 839–850
432. Jardine, P. J., and Coombs, D. H. (1998) *J. Mol. Biol.* **284**, 661–672
432a. Conway, J. F., Wikoff, W. R., Cheng, N., Duda, R. L., Hendrix, R. W., Johnson, J. E., and Steven, A. C. (2001) *Science* **292**, 744–748
433. Burian, J., Stuchlík, S., and Kay, W. W. (1999) *J. Mol. Biol.* **294**, 49–65
434. Merlin, S., and Polisky, B. (1995) *J. Mol. Biol.* **248**, 211–219
435. Paulsson, J., and Ehrenberg, M. (1998) *J. Mol. Biol.* **279**, 73–88
436. Nordström, K., and Wagner, E. G. H. (1994) *Trends Biochem. Sci.* **19**, 294–300
437. Kramer, M. G., Khan, S. A., and Espinosa, M. (1997) *EMBO J.* **16**, 5784–5795
438. Novick, R. P. (1998) *Trends Biochem. Sci.* **23**, 434–438
439. Willetts, N., and Skurray, R. (1987) in *Escherichia coli and Salmonella typhimurium*, Vol. 2 (Neidhardt, F. C., ed), pp. 1110–1133, Am. Soc. for Microbiology, Washington, D. C.
440. Lanka, E., and Wilkins, B. M. (1995) *Ann. Rev. Biochem.* **64**, 141–169
441. Yates, P., Lane, D., and Biek, D. P. (1999) *J. Mol. Biol.* **290**, 627–638
442. White, J. H., and Richardson, C. C. (1987) *J. Biol. Chem.* **262**, 8851–8860
443. Willwand, K., Mumtsidu, E., Kuntz-Simon, G., and Rommelaere, J. (1998) *J. Biol. Chem.* **273**, 1165–1174
443a. Dufour, E., Méndez, J., Lázaro, J. M., de Vega, M., Blanco, L., and Salas, M. (2000) *J. Mol. Biol.* **304**, 289–300
444. King, A. J., and van der Vliet, P. C. (1994) *EMBO J.* **13**, 5786–5792
445. Pombo, A., Ferreira, J., Bridge, E., and Carmo-Fonseca, M. (1994) *EMBO J.* **13**, 5075–5085
446. Monaghan, A., and Hay, R. T. (1996) *J. Biol. Chem.* **271**, 24242–24248
447. Voyles, B. A. (1993) *The Biology of Viruses*, Mosby, St. Louis, Missouri
447a. Parkinson, G. N., Lee, M. P. H., and Neidle, S. (2002) *Nature (London)* **417**, 876–880
447b. Patel, D. J. (2002) *Nature (London)* **417**, 807–808
448. Greider, C. W. (1996) *Ann. Rev. Biochem.* **65**, 337–365
449. Kowald, A. (1997) *J. Mol. Biol.* **273**, 814–825
449a. Aigner, S., Lingner, J., Goodrich, K. J., Grosshans, C. A., Shevchenko, A., Mann, M., and Cech, T. R. (2000) *EMBO J.* **19**, 6230–6239
450. Shippen-Lentz, D., and Blackburn, E. H. (1990) *Science* **247**, 546–552
451. Greider, C. W., and Blackburn, E. H. (1996) *Sci. Am.* **274**(Feb), 92–97
452. Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A. A., Chiu, C.-P., Adams, R. R., Chang, E., Allsopp, R. C., Yu, J., Le, S., West, M. D., Harley, C. B., Andrews, W. H., Greider, C. W., and Villeponteau, B. (1995) *Science* **269**, 1236–1241
453. Bachand, F., and Autexier, C. (1999) *J. Biol. Chem.* **274**, 38027–38031
454. Shay, J. W., and Wright, W. E. (1999) *Science* **286**, 2284–2285
455. Singer, M. S., and Gottschling, D. E. (1994) *Science* **266**, 404–409
456. Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T. R. (1997) *Science* **276**, 561–567
457. Lee, H.-W., Blasco, M. A., Gottlieb, G. J., Horner, J. W., II, Greider, C. W., and DePinho, R. A. (1998) *Nature (London)* **392**, 569–574
458. Lustig, A. J. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3339–3341
459. McEachern, M. J., and Blackburn, E. H. (1995) *Nature (London)* **376**, 403–409
460. Smith, S., Giriat, I., Schmitt, A., and de Lange, T. (1998) *Science* **282**, 1484–1487
460a. Blackburn, E. H. (2000) *Nature (London)* **408**, 53–56

References

460b. Lundblad, V. (2000) *Science* **288**, 2141–2142
461. King, T. C., and Low, R. L. (1987) *J. Biol. Chem.* **262**, 6204–6213
462. Ntambi, J. M., Shapiro, T. A., Ryan, K. A., and Englund, P. T. (1986) *J. Biol. Chem.* **261**, 11890–11895
463. Palmer, J. D. (1983) *Nature (London)* **301**, 92–93
464. Dey, P. M., and Harborne, J. B., eds. (1997) *Plant Biochemistry*, Academic Press, San Diego, California
465. Klein, A., and Bonhoeffer, F. (1972) *Ann. Rev. Biochem.* **41**, 301–332
466. DePamphilis, M. L., ed. (1996) *DNA Replication in Eukaryotic Cells*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
466a. DePamphilis, M. L., ed. (1999) *Concepts of Eukaryotic DNA Replication*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
467. Waga, S., and Stillman, B. (1998) *Ann. Rev. Biochem.* **67**, 721–751
467a. Bell, S. P., and Dutta, A. (2002) *Ann. Rev. Biochem.* **71**, 333–374
468. Campbell, J. L. (1988) *Trends Biochem. Sci.* **13**, 212–217
468a. Wyrick, J. J., Aparicio, J. G., Chen, T., Barnett, J. D., Jennings, E. G., Young, R. A., Bell, S. P., and Aparicio, O. M. (2001) *Science* **294**, 2357–2360
469. Diller, J. D., and Raghuraman, M. K. (1994) *Trends Biochem. Sci.* **19**, 320–325
469a. Raghuraman, M. K., Winzeler, E. A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D. J., Davis, R. W., Brewer, B. J., and Fangman, W. L. (2001) *Science* **294**, 115–121
469b. Stillman, B. (2001) *Science* **294**, 2301–2304
470. Weiser, T., Gassmann, M., Thömmes, P., Ferrari, E., and Hafkemeyer, P. (1991) *J. Biol. Chem.* **266**, 10420–10428
471. Sugino, A. (1995) *Trends Biochem. Sci.* **20**, 319–323
472. Spadari, S., Sala, F., and Pedrali-Noy, G. (1982) *Trends Biochem. Sci.* **7**, 29–32
473. Copeland, W. C., and Tan, X. (1995) *J. Biol. Chem.* **270**, 3905–3913
474. Arezi, B., Kirk, B. W., Copeland, W. C., and Kuchta, R. D. (1999) *Biochemistry* **38**, 12899–12907
474a. Arezi, B., and Kuchta, R. D. (2000) *Trends Biochem. Sci.* **25**, 572–576
475. Copeland, W. C., and Wang, T. S.-F. (1993) *J. Biol. Chem.* **268**, 11028–11040
476. Kelman, Z., and Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238
476a. Shiomi, Y., Usukura, J., Masamura, Y., Takeyasu, K., Nakayama, Y., Obuse, C., Yoshikawa, H., and Tsurimoto, T. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14127–14132
477. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M., and Frappier, L. (1997) *Nature (London)* **385**, 176–181
478. Foiani, M., Lucchini, G., and Plevani, P. (1997) *Trends Biochem. Sci.* **22**, 424–427
479. Kelly, T. J. (1988) *J. Biol. Chem.* **263**, 17889–17892
480. Hurwitz, J., Dean, F. B., Kwong, A. D., and Lee, S.-H. (1990) *J. Biol. Chem.* **265**, 18043–18046
481. Waga, S., and Stillman, B. (1994) *Nature (London)* **369**, 207–212
482. Challberg, M. D., and Kelly, T. J. (1989) *Ann. Rev. Biochem.* **58**, 671–717
483. SenGupta, D. J., and Borowiec, J. A. (1992) *Science* **256**, 1656–1661
483a. Weisshart, K., Förster, H., Kremmer, E., Schlott, B., Grosse, F., and Nasheuer, H.-P. (2000) *J. Biol. Chem.* **275**, 17328–17337
484. Lehman, I. R., and Boehmer, P. E. (1999) *J. Biol. Chem.* **274**, 28059–28062
485. Murata, L. B., and Dodson, M. S. (1999) *J. Biol. Chem.* **274**, 37079–37086
486. Toyn, J. H., Toone, W. M., Morgan, B. A., and Johnston, L. H. (1995) *Trends Biochem. Sci.* **20**, 70–73
487. Murray, A. W., and Szostak, J. W. (1987) *Sci. Am.* **257**(Nov), 62–68
488. Farr, C. J., Bayne, R. A. L., Kipling, D., Mills, W., Critcher, R., and Cooke, H. J. (1995) *EMBO J.* **14**, 5444–5454
489. Warburton, P. E., and Kipling, D. (1997) *Nature (London)* **386**, 553–555
490. Shirahige, K., Hori, Y., Shiraishi, K., Yamashita, M., Takahashi, K., Obuse, C., Tsurimoto, T., and Yoshikawa, H. (1998) *Nature (London)* **395**, 618–621
491. Gómez, M., and Antequera, F. (1999) *EMBO J.* **18**, 5683–5690
492. Kim, S.-M., and Huberman, J. A. (1999) *J. Mol. Biol.* **288**, 867–882
493. Gavin, K. A., Hidaka, M., and Stillman, B. (1995) *Science* **270**, 1667–1671
494. Marx, J. (1995) *Science* **270**, 1585–1587
495. Tye, B. K. (1999) *Ann. Rev. Biochem.* **68**, 649–686
495a. Dhar, S. K., Delmolino, L., and Dutta, A. (2001) *J. Biol. Chem.* **276**, 29067–29071
495b. Blow, J. J. (2001) *EMBO J.* **20**, 3293–3297
495c. Krude, T. (2000) *J. Biol. Chem.* **275**, 13699–13707
496. Prokhorova, T. A., and Blow, J. J. (2000) *J. Biol. Chem.* **275**, 2491–2498
497. Coverley, D., and Laskey, R. A. (1994) *Ann. Rev. Biochem.* **63**, 745–776
498. Chong, J. P. J., Thömmes, P., and Blow, J. J. (1996) *Trends Biochem. Sci.* **21**, 102–106
499. Stillman, B. (1994) *J. Biol. Chem.* **269**, 7047–7050
500. Aladjem, M. I., Rodewald, L. W., Kolman, J. L., and Wahl, G. M. (1998) *Science* **281**, 1005–1009
500a. Tye, B. K., and Sawyer, S. (2000) *J. Biol. Chem.* **275**, 34833–34836
500b. Sato, M., Gotow, T., You, Z., Komamura-Kohno, Y., Uchiyama, Y., Yabuta, N., Nojima, H., and Ishimi, Y. (2000) *J. Mol. Biol.* **300**, 421–431
500c. Lee, J.-K., and Hurwitz, J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 54–59
500d. Chong, J. P. J., Hayashi, M. K., Simon, M. N., Xu, R.-M., and Stillman, B. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1530–1535
500e. Wohlschlegel, J. A., Dwyer, B. T., Dhar, S. K., Cvetic, C., Walter, J. C., and Dutta, A. (2000) *Science* **290**, 2309–2312
500f. Lygerou, Z., and Nurse, P. (2000) *Science* **290**, 2271–2273
500g. Blow, J. J., and Tada, S. (2000) *Nature (London)* **404**, 560–561
500h. Keller, C., and Krude, T. (2000) *J. Biol. Chem.* **275**, 35512–35521
501. DePamphilis, M. L. (1993) *J. Biol. Chem.* **268**, 1–4
501a. Lucas, I., Chevrier-Miller, M., Sogo, J. M., and Hyrien, O. (2000) *J. Mol. Biol.* **296**, 769–786
501b. Gilbert, D. M. (2001) *Science* **294**, 96–100
502. Wu, J.-R., and Gilbert, D. M. (1996) *Science* **271**, 1270–1272
503. Dimitrova, D. S., Giacca, M., Demarchi, F., Biamonti, G., Riva, S., and Falaschi, A. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1498–1503
504. Kasamatson, H., and Vinograd, J. (1973) *Nature (London)* **241**, 103–105
505. Gaudette, M. F., and Benbow, R. M. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5953–5957
506. Turchi, J. J., Huang, L., Murante, R. S., Kim, Y., and Bambara, R. A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9803–9807
506a. Karthikeyan, R., Vonarx, E. J., Straffon, A. F. L., Simon, M., Faye, G., and Kunz, B. A. (2000) *J. Mol. Biol.* **299**, 405–419
507. Huang, L., Rumbaugh, J. A., Murante, R. S., Lin, R. J. R., Rust, L., and Bambara, R. A. (1996) *Biochemistry* **35**, 9266–9277
507a. Gomes, X. V., and Burgers, P. M. J. (2000) *EMBO J.* **19**, 3811–3821
507b. Bae, S.-H., and Seo, Y.-S. (2000) *J. Biol. Chem.* **275**, 38022–38031
508. Pohjanpelto, P., and Hölttä, E. (1996) *EMBO J.* **15**, 1193–1200
509. Bonne-Andrea, C., Wong, M. L., and Alberts, B. M. (1990) *Nature (London)* **343**, 719–726
510. Gasser, R., Koller, T., and Sogo, J. M. (1996) *J. Mol. Biol.* **258**, 224–239
511. Murray, A. W. (1987) *Nature (London)* **327**, 14–15
512. Fisher, D. L., and Nurse, P. (1996) *EMBO J.* **15**, 850–860
513. Santocanale, C., Neecke, H., Longhese, M. P., Lucchini, G., and Plevani, P. (1995) *J. Mol. Biol.* **254**, 595–607
514. Novak, B., and Tyson, J. J. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9147–9152
515. Brush, G. S., Morrow, D. M., Hieter, P., and Kelly, T. J. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 15075–15080
516. Collins, J. M., and Chu, A. K. (1987) *Biochemistry* **26**, 5600–5607
517. Stahl, F. W. (1987) *Sci. Am.* **256**(Feb), 91–101
518. Kucherlapati, R., and Smith, G. R., eds. (1988) *Genetic Recombination*, Am. Soc. for Microbiology, Washington, DC
519. Kowalczykowski, S. C. (1991) *Ann. Rev. Biophys. Biophys. Chem.* **20**, 539–575
520. Edelmann, W., and Kucherlapati, R. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6225–6227
521. Shinohara, A., and Ogawa, T. (1995) *Trends Biochem. Sci.* **20**, 387–391
522. Haber, J. E. (1999) *Nature (London)* **398**, 665–667
523. Meselson, M., and Weigle, J. J. (1961) *Proc. Natl. Acad. Sci. U.S.A.* **47**, 857–868
524. Holliday, R. (1964) *Genet. Res. Camb.* **5**, 282–304
525. Potter, H., and Dressler, D. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3000–3004
525a. Eichman, B. F., Vargason, J. M., Mooers, B. H. M., and Ho, P. S. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3971–3976
525b. Sha, R., Liu, F., and Seeman, N. C. (2002) *Biochemistry* **41**, 5950–5955
526. Sigal, N., and Alberts, B. (1972) *J. Mol. Biol.* **71**, 789–793
527. Meselson, M. (1972) *J. Mol. Biol.* **71**, 795–798
528. Shinagawa, H., and Iwasaki, H. (1996) *Trends Biochem. Sci.* **21**, 107–111
529. Rafferty, J. B., Sedelnikova, S. E., Hargreaves, D., Artymiuk, P. J., Baker, P. J., Sharples, G. J., Mahdi, A. A., Lloyd, R. G., and Rice, D. W. (1996) *Science* **274**, 415–421
530. Ishimori, K., Sommer, S., Bailone, A., Takahashi, M., Cox, M. M., and Devoret, R. (1996) *J. Mol. Biol.* **264**, 696–712
530a. Mcllwraith, M. J., and West, S. C. (2001) *J. Mol. Biol.* **305**, 23–31
530b. Postow, L., Ullsperger, C., Keller, R. W., Bustamante, C., Vologodskii, A. V., and Cozzarelli, N. R. (2001) *J. Biol. Chem.* **276**, 2790–2796
530c. Neschastnova, A. A., Markina, V. K., Popenko, V. I., Danilova, O. A., Sidorov, R. A., Belitsky, G. A., and Yakubovskaya, M. G. (2002) *Biochemistry* **41**, 7795–7801
530d. Mcllwraith, M. J., Van Dyck, E., Masson, J.-Y., Stasiak, A. Z., Stasiak, A., and West, S. C. (2000) *J. Mol. Biol.* **304**, 151–164
530e. Bolt, E. L., Sharples, G. J., and Lloyd, R. G. (2000) *J. Mol. Biol.* **304**, 165–176

References

531. Farah, J. A., and Smith, G. R. (1997) *J. Mol. Biol.* **272**, 699–715
532. Wang, J., Chen, R., and Julin, D. A. (2000) *J. Biol. Chem.* **275**, 507–513
533. Anderson, D. G., and Kowalczykowski, S. C. (1998) *J. Mol. Biol.* **282**, 275–285
533a. Arnold, D. A., Handa, N., Kobayashi, I., and Kowalczykowski, S. C. (2000) *J. Mol. Biol.* **300**, 469–479
534. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) *Science* **277**, 1453–1462
534a. Flores, M.-J., Bierne, H., Ehrlich, S. D., and Michel, B. (2001) *EMBO J.* **20**, 619–629
535. Köppen, A., Krobitsch, S., Thoms, B., and Wackernagel, W. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6249–6253
536. Howard-Flanders, P., West, S. C., and Stasiak, A. (1984) *Nature (London)* **309**, 215–220
537. Story, R. M., Weber, I. T., and Steitz, T. A. (1992) *Nature (London)* **355**, 318–325
538. Cui, X., Wise, R. P., and Schnable, P. S. (1996) *Science* **272**, 1334–1336
539. Arenson, T. A., Tsodikov, O. V., and Cox, M. M. (1999) *J. Mol. Biol.* **288**, 391–401
539a. Egelman, E. H. (2001) *J. Mol. Biol.* **309**, 539–542
540. Hegner, M., Smith, S. B., and Bustamante, C. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10109–10114
540a. Lusetti, S. L., and Cox, M. M. (2002) *Ann. Rev. Biochem.* **71**, 71 – 100
541. Register, J. C., III, and Griffith, J. (1985) *J. Biol. Chem.* **260**, 12308–12312
542. Zaitsev, E. N., and Kowalczykowski, S. C. (1999) *J. Mol. Biol.* **287**, 21–31
543. Adzuma, K. (1998) *J. Biol. Chem.* **273**, 31565–31573
544. Rao, B. J., Chiu, S., Bazemore, L. R., Reddy, G., and Radding, C. M. (1995) *Trends Biochem. Sci.* **20**, 109–113
544a. Voloshin, O. N., Wang, L., and Camerini-Otero, R. D. (2000) *J. Mol. Biol.* **303**, 709–720
544b. Rice, K. P., Eggler, A. L., Sung, P., and Cox, M. M. (2001) *J. Biol. Chem.* **276**, 38570–38581
545. Engels, W. R., Preston, C. R., and Johnson-Schlitz, D. M. (1994) *Science* **263**, 1623–1625
546. Kowalczykowski, S. C., and Eggleston, A. K. (1994) *Ann. Rev. Biochem.* **63**, 991–1043
547. Beernink, H. T. H., and Morrical, S. W. (1999) *Trends Biochem. Sci.* **24**, 385–389
548. Birkenkamp, K., and Kemper, B. (1996) *J. Mol. Biol.* **259**, 622–631
549. Ando, R. A., and Morrical, S. W. (1999) *Biochemistry* **38**, 16589–16598
550. Passy, S. I., Yu, X., Li, Z., Radding, C. M., and Egelman, E. H. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4279–4284
551. Sung, P. (1994) *Science* **265**, 1241–1243
551a. Masson, J.-Y., and West, S. C. (2001) *Trends Biochem. Sci.* **26**, 131 – 136
552. Benson, F. E., Stasiak, A., and West, S. C. (1994) *EMBO J.* **13**, 5764–5771
553. Baumann, P., and West, S. C. (1998) *Trends Biochem. Sci.* **23**, 247–251
554. Li, Z., Golub, E. I., Gupta, R., and Radding, C. M. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11221–11226
555. Cerutti, H., Osman, M., Grandoni, P., and Jagendorf, A. T. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8068–8072
556. Rafferty, J. B., Ingleston, S. M., Hargreaves, D., Artymiuk, P. J., Sharples, G. J., Lloyd, R. G., and Rice, D. W. (1998) *J. Mol. Biol.* **278**, 105–116
557. George, H., Mézard, C., Stasiak, A., and West, S. C. (1999) *J. Mol. Biol.* **293**, 505–519
558. Qiu, X.-B., Lin, Y.-L., Thome, K. C., Pian, P., Schlegel, B. P., Weremowicz, S., Parvin, J. D., and Dutta, A. (1998) *J. Biol. Chem.* **273**, 27786–27793
559. Rafferty, J. B., Sedelnikova, S. E., Hargreaves, D., Artymiuk, P. J., Baker, P. J., Sharples, G. J., Hahdi, A. A., Lloyd, R. G., and Rice, D. W. (1996) *Science* **274**, 415–421
559a. Putnam, C. D., Clancy, S. B., Tsuruta, H., Gonzalez, S., Wetmur, J. G., and Tainer, J. A. (2001) *J. Mol. Biol.* **311**, 297–310
559b. Yamada, K., Kunishima, N., Mayanagi, K., Ohnishi, T., Nishino, T., Iwasaki, H., Shinagawa, H., and Morikawa, K. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1442–1447
560. Lee, S., Cavallo, L., and Griffith, J. (1997) *J. Biol. Chem.* **272**, 7532–7539
561. Zlatanova, J., and van Holde, K. (1998) *FASEB J.* **12**, 421–431
562. Golz, S., and Kemper, B. (1999) *J. Mol. Biol.* **285**, 1131–1144
563. Raaijmakers, H., Vix, O., Töro, I., Golz, S., Kemper, B., and Suck, D. (1999) *EMBO J.* **18**, 1447–1458
564. Bolt, E. L., Sharples, G. J., and Lloyd, R. G. (1999) *J. Mol. Biol.* **286**, 403–415
565. Whitby, M. C., and Dixon, J. (1998) *J. Biol. Chem.* **273**, 35063–35073
566. Komori, K., Sakae, S., Shinagawa, H., Morikawa, K., and Ishino, Y. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8873–8878
567. Kvaratskhelia, M., George, S. J., Cooper, A., and White, M. F. (1999) *Biochemistry* **38**, 16613–16619
568. Fincham, J. R. S., and Oliver, P. (1989) *Nature (London)* **338**, 14–15
569. Nicolas, A., Treco, D., Schultes, N. P., and Szostak, J. W. (1989) *Nature (London)* **338**, 35–39
570. Fabre, F. (1978) *Nature (London)* **272**, 795–798
571. Groden, J., Nakamura, Y., and German, J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4315–4319
572. Jackson, J. A., and Fink, G. R. (1981) *Nature (London)* **292**, 306–311
573. Klein, H. L. (1984) *Nature (London)* **310**, 748–753
574. Robertson, M. (1981) *Nature (London)* **293**, 333–334
575. Szostak, J. W., and Wu, R. (1980) *Nature (London)* **284**, 426–430
576. Sadowski, P. D. (1993) *FASEB J.* **7**, 760–767
577. Wang, S., Cosstick, R., Gardner, J. F., and Gumport, R. I. (1995) *Biochemistry* **34**, 13082–13090
577a. Read, E. K., Gumport, R. I., and Gardner, J. F. (2000) *J. Biol. Chem.* **275**, 33759–33764
578. Nash, H. A. (1990) *Trends Biochem. Sci.* **15**, 222–227
579. Tirumalai, R. S., Kwon, H. J., Cardente, E. H., Ellenberger, T., and Landy, A. (1998) *J. Mol. Biol.* **279**, 513–527
580. Crisona, N. J., Weinberg, R. L., Peter, B. J., Sumners, D. W., and Cozzarelli, N. R. (1999) *J. Mol. Biol.* **289**, 747–775
580a. Bankhead, T., and Segall, A. M. (2000) *J. Biol. Chem.* **275**, 36949–36956
580b. Lorbach, E., Christ, N., Schwikardi, M., and Dröge, P. (2000) *J. Mol. Biol.* **296**, 1175–1181
581. Hsu, P. L., and Landy, A. (1984) *Nature (London)* **311**, 721–726
582. Dorgai, L., Sloan, S., and Weisberg, R. A. (1998) *J. Mol. Biol.* **277**, 1059–1070
583. Guo, F., Gopaul, D. N., and Van Duyne, G. D. (1997) *Nature (London)* **389**, 40–46
584. Gopaul, D. N., Guo, F., and Van Duyne, G. D. (1998) *EMBO J.* **17**, 4175–4187
585. Subramanya, H. S., Arciszewska, L. K., Baker, R. A., Bird, L. E., Sherratt, D. J., and Wigley, D. B. (1997) *EMBO J.* **16**, 5178–5187
586. Arciszewska, L. K., Grainge, I., and Sherratt, D. J. (1997) *EMBO J.* **16**, 3731–3743
587. Jayaram, M. (1994) *Trends Biochem. Sci.* **19**, 78–82
587a. Ferreira, H., Sherratt, D., and Arciszewska, L. (2001) *J. Mol. Biol.* **312**, 45–57
588. Zhu, X.-D., Pan, G., Luetke, K., and Sadowski, P. D. (1995) *J. Biol. Chem.* **270**, 11646–11653
589. Huffman, K. E., and Levene, S. D. (1999) *J. Mol. Biol.* **286**, 1–13
590. Lee, J., Jayaram, M., and Grainge, I. (1999) *EMBO J.* **18**, 784–791
591. Allingham, J. S., Pribil, P. A., and Haniford, D. B. (1999) *J. Mol. Biol.* **289**, 1195–1206
592. Watson, M. (1984) *Trends Biochem. Sci.* **9**, 82–83
593. Lim, H. M., Hughes, K. T., and Simon, M. I. (1992) *J. Biol. Chem.* **267**, 11183–11190
594. Feng, J.-A., Johnson, R. C., and Dickerson, R. E. (1994) *Science* **263**, 348–355
595. Kostrewa, D., Granzin, J., Koch, C., Choe, H.-W., Raghunathan, S., Wolf, W., Labahn, J., Kahmann, R., and Saenger, W. (1991) *Nature (London)* **349**, 178–180
596. Safo, M. K., Yang, W.-Z., Corselli, L., Cramton, S. E., Yuan, H. S., and Johnson, R. C. (1997) *EMBO J.* **16**, 6860–6873
597. Dworkin, J., and Blaser, M. J. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 985–990
598. Fedoroff, N., ed. (1992) *The Dynamic Genome Barbara McClintock's Ideas in the Century of Genetics*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
599. Fedoroff, N. V. (1984) *Sci. Am.* **250**(Jun), 85–98
600. Burr, B., and Burr, F. A. (1983) *Trends Biochem. Sci.* **8**, 429–431
601. Kingsman, A. J., Chater, K. F., and Kingsman, S. M., eds. (1988) *Transposition*, Cambridge Univ. Press, London and New York
602. Cohen, S. N., and Shapiro, J. A. (1980) *Sci. Am.* **243**(Feb), 40–49
603. Watson, M. (1985) *Trends Biochem. Sci.* **10**, 178
604. Ohtsubo, H., Nyman, K., Doroszkiewicz, W., and Ohtsubo, E. (1981) *Nature (London)* **292**, 640–643
604a. Redder, P., She, Q., and Garrett, R. A. (2001) *J. Mol. Biol.* **306**, 1–6
605. Gamas, P., Galas, D., and Chandler, M. (1985) *Nature (London)* **317**, 458–460
606. Kröger, M., and Hobom, G. (1982) *Nature (London)* **297**, 159–162
607. Phadnis, S. H., and Berg, D. E. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9118–9122
608. Lee, S. Y., Butler, D., and Kleckner, N. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7876–7880
608a. Rao, J. E., and Craig, N. L. (2001) *J. Mol. Biol.* **307**, 1161–1170
609. Farabaugh, P. J., and Fink, G. R. (1980) *Nature (London)* **286**, 352–356
609a. Pribil, P. A., and Haniford, D. B. (2000) *J. Mol. Biol.* **303**, 145–159
610. Sakai, J., Chalmers, R. M., and Kleckner, N. (1995) *EMBO J.* **14**, 4374–4383
611. Bhasin, A., Goryshin, I. Y., and Reznikoff, W. S. (1999) *J. Biol. Chem.* **274**, 37021–37029
611a. Davies, D. R., Goryshin, I. Y., Reznikoff, W. S., and Rayment, I. (2000) *Science* **289**, 77–85
611b. Bhasin, A., Goryshin, I. Y., Steiniger-White, M., York, D., and Reznikoff, W. S. (2000) *J. Mol. Biol.* **302**, 49–63
612. Gary, P. A., Biery, M. C., Bainton, R. J., and Craig, N. L. (1996) *J. Mol. Biol.* **257**, 301–316
612a. Williams, T. L., and Baker, T. A. (2000) *Science* **289**, 73 – 74
613. Benjamin, K. R., Abola, A. P., Kanaar, R., and Cozzarelli, N. R. (1996) *J. Mol. Biol.* **256**, 50–65
614. Arnold, P. H., Blake, D. G., Grindley, N. D. F., Boocock, M. R., and Stark, W. M. (1999) *EMBO J.* **18**, 1407–1414

References

615. Boocock, M. R., Zhu, X., and Grindley, N. D. F. (1995) *EMBO J.* **14**, 5129–5140
616. Watson, M. A., Boocock, M. R., and Stark, W. M. (1996) *J. Mol. Biol.* **257**, 317–329
617. Wasserman, S. A., Dungan, J. M., and Cozzarelli, N. R. (1985) *Science* **229**, 171–174
618. Savilahti, H., Rice, P. A., and Mizuuchi, K. (1995) *EMBO J.* **14**, 4893–4903
619. Krementsova, E., Giffin, M. J., Pincus, D., and Baker, T. A. (1998) *J. Biol. Chem.* **273**, 31358–31365
620. Naigamwalla, D. Z., Coros, C. J., Wu, Z., and Chaconas, G. (1998) *J. Mol. Biol.* **282**, 265–274
621. Jiang, H., Yang, J.-Y., and Harshey, R. M. (1999) *EMBO J.* **18**, 3845–3855
622. Stellwagen, A. E., and Craig, N. L. (1998) *Trends Biochem. Sci.* **23**, 486–490
623. Dranginis, A. M. (1986) *Trends Biochem. Sci.* **11**, 328–331
624. Sugawara, N., Ivanov, E. L., Fishman-Lobell, J., Ray, B. L., Wu, X., and Haber, J. E. (1995) *Nature (London)* **373**, 84–86
625. Flavell, A. (1986) *Nature (London)* **320**, 397
626. Sherratt, D. J., ed. (1995) *Mobile Genetic Elements*, Oxford Univ. Press, New York
627. Gloor, G. B., Nassif, N. A., Johnson-Schlitz, D. M., Preston, C. R., and Engels, W. R. (1991) *Science* **253**, 1110–1117
628. Beall, E. L., and Rio, D. C. (1998) *EMBO J.* **17**, 2122–2136
629. Marx, J. (1991) *Science* **253**, 1092–1093
630. Lohe, A. R., De Aguiar, D., and Hartl, D. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1293–1297
631. van Pouderoyen, G., Ketting, R. F., Perrakis, A., Plasterk, R. H. A., and Sixma, T. K. (1997) *EMBO J.* **16**, 6044–6054
632. Goodier, J. L., and Davidson, W. S. (1994) *J. Mol. Biol.* **241**, 26–34
633. Lam, W. L., Seo, P., Robison, K., Virk, S., and Gilbert, W. (1996) *J. Mol. Biol.* **257**, 359–366
634. Oosumi, T., and Belknap, W. R. (1995) *Nature (London)* **378**, 672
635. Holzman, D. (1991) *Science* **254**, 1728–1729
636. Lykke-Andersen, J., Garrett, R. A., and Kjems, J. (1997) *EMBO J.* **16**, 3272–3281
637. Morgan, G. T. (1995) *J. Mol. Biol.* **254**, 1–5
638. Jones, J. D. G., Carland, F. M., Maliga, P., and Dooner, H. K. (1989) *Science* **244**, 204–207
639. Spaink, H. P. (1999) *Nature (London)* **402**, 135–136
640. Chopra, S., Brendel, V., Zhang, J., Axtell, J. D., and Peterson, T. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15330–15335
641. Koes, R., Souer, E., van Houwelingen, A., Mur, L., Spelt, C., Quattrocchio, F., Wing, J., Oppedijk, B., Ahmed, S., Maes, T., Gerats, T., Hoogeveen, P., Meesters, M., Kloos, D., and Mol, J. N. M. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8149–8153
642. Morgan, B. A., Conlon, F. L., Manzanares, M., Millar, J. B. A., Kanuga, N., Sharpe, J., Krumlauf, R., Smith, J. C., and Sedgwick, S. G. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2801–2806
643. Schardl, C. L., Lonsdale, D. M., Pring, D. R., and Rose, K. R. (1984) *Nature (London)* **310**, 292–296
644. Ikeda, H. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 922–926
645. Bi, X., and Liu, L. F. (1996) *J. Mol. Biol.* **256**, 849–858
646. Stark, G. R., and Wahl, G. M. (1984) *Ann. Rev. Biochem.* **53**, 447–491
647. Schimke, R. T. (1988) *J. Biol. Chem.* **263**, 5989–5992
648. Syu, L.-J., and Fluck, M. M. (1997) *J. Mol. Biol.* **271**, 76–99
649. Wu, C. H. H., and Black, L. W. (1995) *J. Mol. Biol.* **247**, 604–617
650. Andersson, D. I., Slechta, E. S., and Roth, J. R. (1998) *Science* **282**, 1133–1135
651. Huang, T., and Campbell, J. L. (1995) *J. Biol. Chem.* **270**, 9607–9614
652. Sheriff, S., and Herriott, J. R. (1981) *J. Mol. Biol.* **145**, 441–451
653. Subramanian, V., Liu, T.-N., Yeh, W.-K., Narro, M., and Gibson, D. T. (1983) *J. Biol. Chem.* **256**, 2723–2730
654. Alitalo, K. (1985) *Trends Biochem. Sci.* **10**, 194–197
655. Friedberg, E. C., Walker, G. C., and Siede, W. (1995) *DNA Repair and Mutagenesis*, ASM Press, Materials Park, Ohio
656. Kuhnlein, U. (1985) *J. Biol. Chem.* **260**, 14918–14924
657. Beckman, K. B., and Ames, B. N. (1997) *J. Biol. Chem.* **272**, 19633–19636
658. Purmal, A. A., Lampman, G. W., Bond, J. P., Hatahet, Z., and Wallace, S. S. (1998) *J. Biol. Chem.* **273**, 10026–10035
659. Jourdan, M., Garcia, J., Defrancq, E., Kotera, M., and Lhomme, J. (1999) *Biochemistry* **38**, 3985–3995
660. Huff, A. C., and Topal, M. D. (1987) *J. Biol. Chem.* **262**, 12843–12850
661. Rosenblatt, J., and Mitchison, T. J. (1998) *Nature (London)* **393**, 739–740
662. Boorstein, R. J., Hilbert, T. P., Cunningham, R. P., and Teebor, G. W. (1990) *Biochemistry* **29**, 10455–10460
663. Doetsch, P. W., Zastawny, T. H., Martin, A. M., and Dizdaroglu, M. (1995) *Biochemistry* **34**, 737–742
663a. Zhou, H., Suzuki, M., Randers-Pehrson, G., Vannais, D., Chen, G., Trosko, J. E., Waldren, C. A., and Hei, T. K. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14410–14415
663b. Forster, L., Forster, P., Lutz-Bonengel, S., Willkomm, H., and Brinkmann, B. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13950–13954
664. Lai, M.-D., and Beattie, K. L. (1988) *Biochemistry* **27**, 1722–1728
665. Kunkel, T. A., and Soni, A. (1988) *J. Biol. Chem.* **263**, 14784–14789
666. Echols, H., and Goodman, M. F. (1991) *Ann. Rev. Biochem.* **60**, 477–511
667. Kunkel, T. A. (1992) *J. Biol. Chem.* **267**, 18251–18254
668. Timsit, Y. (1999) *J. Mol. Biol.* **293**, 835–853
668a. Viswanathan, M., Lacirignola, J. J., Hurley, R. L., and Lovett, S. T. (2000) *J. Mol. Biol.* **302**, 553–564
669. Radman, M., and Wagner, R. (1988) *Sci. Am.* **259**(Aug), 40–46
669a. Patel, P. H., Kawate, H., Adman, E., Ashbach, M., and Loeb, L. A. (2001) *J. Biol. Chem.* **276**, 5044–5051
670. Fersht, A. R. (1980) *Trends Biochem. Sci.* **5**, 262–265
671. Petruska, J., Goodman, M. F., Boosalis, M. S., Sowers, L. C., Cheong, C., and Tinoco, I., Jr. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6252–6256
672. Metzler, D. E. (1977) *Biochemistry; The Chemical Reactions of Living Cells*, 1st ed., Academic Press, New York (pp. 912–915)
673. Metzler, D. E. (1979) *Adv. Enzymol.* **50**, 1–40
674. Goodman, M. F. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10493–10495
675. Moran, S., Ren, R. X.-F., and Kool, E. T. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10506–10511
676. Matray, T. J., and Kool, E. T. (1998) *J. Am. Chem. Soc.* **120**, 6191–6192
676a. Kool, E. T. (2002) *Ann. Rev. Biochem.* **71**, 191–219
676b. Showalter, A. K., and Tsai, M.-D. (2002) *Biochemistry* **41**, 10571–10576
677. Mo, J.-Y., and Schaaper, R. M. (1996) *J. Biol. Chem.* **271**, 18947–18953
678. Fijalkowska, I. J., Jonczyk, P., Tkaczyk, M. M., and Bialoskorska, M. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10020–10025
679. Kroutil, L. C., Frey, M. W., Kaboord, B. F., Kunkel, T. A., and Benkovic, S. J. (1998) *J. Mol. Biol.* **278**, 135–146
680. Seki, M., Akiyama, M., Sugaya, Y., Ohtsubo, E., and Maki, H. (1999) *J. Biol. Chem.* **274**, 33313–33319
681. Modrich, P. (1987) *Ann. Rev. Biochem.* **56**, 435–466
682. Bohr, V. A., and Wassermann, K. (1988) *Trends Biochem. Sci.* **13**, 429–433
683. Thoma, F. (1999) *EMBO J.* **18**, 6585–6598
684. Friedberg, E. C. (1995) *Trends Biochem. Sci.* **20**, 381
685. Arbel, A., Zenvirth, D., and Simchen, G. (1999) *EMBO J.* **18**, 2648–2658
686. Modrich, P. (1989) *J. Biol. Chem.* **264**, 6597–6600
686a. Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. (2001) *Science* **291**, 1284–1289
687. Modrich, P., and Lahue, R. (1996) *Ann. Rev. Biochem.* **65**, 101–133
688. Schaaper, R. M. (1993) *J. Biol. Chem.* **268**, 23762–23765
688a. Obmolova, G., Ban, C., Hsieh, P., and Yang, W. (2000) *Nature (London)* **407**, 703–711
688b. Lamers, M. H., Perrakis, A., Enzlin, J. H., Winterwerp, H. H. K., de Wind, N., and Sixma, T. K. (2000) *Nature (London)* **407**, 711–717
688c. Bowers, J., Tran, P. T., Joshi, A., Liskay, R. M., and Alani, E. (2001) *J. Mol. Biol.* **306**, 957–968
689. Ban, C., and Yang, W. (1998) *EMBO J.* **17**, 1526–1534
690. Modich, P. (1997) *J. Biol. Chem.* **272**, 24727–24730
691. Wang, H., Lawrence, C. W., Li, G.-M., and Hays, J. B. (1999) *J. Biol. Chem.* **274**, 16894–16900
692. Nakagawa, T., Datta, A., and Kolodner, R. D. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14186–14188
692a. Coïc, E., Gluck, L., and Fabre, F. (2000) *EMBO J.* **19**, 3408–3417
692b. Rich, T., Allen, R. L., and Wyllie, A. H. (2000) *Nature (London)* **407**, 777–783
693. Ahn, B., and Grossman, L. (1996) *J. Biol. Chem.* **271**, 21462–21470
694. Sancar, A. (1996) *Ann. Rev. Biochem.* **65**, 43–81
695. Zou, Y., and Van Houten, B. (1999) *EMBO J.* **18**, 4889–4901
695a. Verhoeven, E. E. A., Wyman, C., Moolenaar, G. F., Hoeijmakers, J. H. J., and Goosen, N. (2001) *EMBO J.* **20**, 601–611
696. Theis, K., Chen, P. J., Skorvaga, M., Van Houten, B., and Kisker, C. (1999) *EMBO J.* **18**, 6899–6907
696a. Villani, G., and Le Gac, N. T. (2000) *J. Biol. Chem.* **275**, 33185–33188
697. Shi, Q., Thresher, R., Sancar, A., and Griffith, J. (1992) *J. Mol. Biol.* **226**, 425–432
698. Verhoeven, E. E. A., van Kesteren, M., Moolenaar, G. F., Visse, R., and Goosen, N. (2000) *J. Mol. Biol.* **275**, 5120–5123
699. Ali, J. A., Maluf, N. K., and Lohman, T. M. (1999) *J. Mol. Biol.* **293**, 815–834
700. Wood, R. D. (1997) *J. Biol. Chem.* **272**, 23465–23468
701. Sancar, A. (1995) *J. Biol. Chem.* **270**, 15915–15918
702. Doetsch, P. W. (1995) *Trends Biochem. Sci.* **20**, 384–386
703. Guzder, S. N., Bailly, V., Sung, P., Prakash, L., and Prakash, S. (1995) *J. Biol. Chem.* **270**, 8385–8388
703a. Cappelli, E., Degan, P., Thompson, L. H., and Frosina, G. (2000) *Biochemistry* **39**, 10408–10412

References

704. Lao, Y., Gomes, X. V., Ren, Y., Taylor, J.-S., and Wold, M. S. (2000) *Biochemistry* **39**, 850–859
705. Zamble, D. B., Mu, D., Reardon, J. T., Sancar, A., and Lippard, S. J. (1996) *Biochemistry* **35**, 10004–10013
705a. Kunz, C., and Fleck, O. (2001) *J. Mol. Biol.* **313**, 241–253
705b. Van Houten, B., Eisen, J. A., and Hanawalt, P. C. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2581–2583
706. Seeberg, E., Eide, L., and Bjorås, M. (1995) *Trends Biochem. Sci.* **20**, 391–397
707. McCullough, A. K., Dodson, M. L., and Lloyd, R. S. (1999) *Ann. Rev. Biochem.* **68**, 255–285
708. Wilson, D. M., III, and Thompson, L. H. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12754–12757
708a. Parker, A., Gu, Y., Mahoney, W., Lee, S.-H., Singh, K. K., and Lu, A.-L. (2001) *J. Biol. Chem.* **276**, 5547–5555
709. Savva, R., McAuley-Hecht, K., Brown, T., and Pearl, L. (1995) *Nature (London)* **373**, 487–493
710. Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. (1996) *Nature (London)* **384**, 87–92
711. Drohat, A. C., Xiao, G., Tordova, M., Jagadeesh, J., Pankiewicz, K. W., Watanabe, K. A., Gilliland, G. L., and Stivers, J. T. (1999) *Biochemistry* **38**, 11876–11886
711a. Otterlei, M., Kavli, B., Standal, R., Skjelbred, C., Bharati, S., and Krokan, H. E. (2000) *EMBO J.* **19**, 5542–5551
711b. Garvish, J. F., and Lloyd, R. S. (2000) *J. Mol. Biol.* **295**, 479–488
711c. Werner, R. M., and Stivers, J. T. (2000) *Biochemistry* **39**, 14054–14064
711d. Dinner, A. R., Blackburn, G. M., and Karplus, M. (2001) *Nature (London)* **413**, 752–755
712. Mol, C. D., Kuo, C.-F., Thayer, M. M., Cunningham, R. P., and Tainer, J. A. (1995) *Nature (London)* **374**, 381–386
713. Otterlei, M., Warbrick, E., Nagelhus, T. A., Haug, T., Slupphaug, G., Akbari, M., Aas, P. A., Steinsbekk, K., Bakke, O., and Krokan, H. E. (1999) *EMBO J.* **18**, 3834–3844
714. Carey, D. C., and Strauss, P. R. (1999) *Biochemistry* **38**, 16553–16560
715. Harris, J. M., McIntosh, E. M., and Muscat, G. E. O. (1999) *J. Mol. Biol.* **288**, 275–287
716. Williams, S. D., and David, S. S. (1999) *Biochemistry* **38**, 15417–15424
717. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1995) *EMBO J.* **14**, 4108–4120
717a. Dizdaroglu, M., Bauche, C., Rodriguez, H., and Laval, J. (2000) *Biochemistry* **39**, 5586–5592
718. Karahalil, B., Roldán-Arjona, T., and Dizdaroglu, M. (1998) *Biochemistry* **37**, 590–595
718a. Bebenek, K., Tissier, A., Frank, E. G., McDonald, J. P., Prasad, R., Wilson, S. H., Woodgate, R., and Kunkel, T. A. (2001) *Science* **291**, 2156–2159
719. Fuxreiter, M., Warshel, A., and Osman, R. (1999) *Biochemistry* **38**, 9577–9589
720. Croteau, D. L., and Bohr, V. A. (1997) *J. Biol. Chem.* **272**, 25409–25412
721. Lipscomb, L. A., Peek, M. E., Morningstar, M. L., Verghis, S. M., Miller, E. M., Rich, A., Essigmann, J. M., and Williams, L. D. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 719–723
722. Bruner, S. D., Norman, D. P. G., and Verdine, G. L. (2000) *Nature (London)* **403**, 859–866
723. Hatahet, Z., Zhou, M., Reha-Krantz, L. J., Morrical, S. W., and Wallace, S. S. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8556–8561
724. Bhagwat, M., and Gerlt, J. A. (1996) *Biochemistry* **35**, 659–665
725. Porello, S. L., Cannon, M. J., and David, S. S. (1998) *Biochemistry* **37**, 6465–6475
725a. Miyako, K., Takamatsu, C., Umeda, S., Tajiri, T., Furuichi, M., Nakabeppu, Y., Sekiguchi, M., Hamasaki, N., Takeshige, K., and Kang, D. (2000) *J. Biol. Chem.* **275**, 12326–12330
725b. Zharkov, D. O., Gilboa, R., Yagil, I., Kycia, J. H., Gerchman, S. E., Shoham, G., and Grollman, A. P. (2000) *Biochemistry* **39**, 14768–14778
726. Manuel, R. C., Czerwinski, E. W., and Lloyd, R. S. (1996) *J. Biol. Chem.* **271**, 16218–16226
727. Guibourt, N., Castaing, B., Auffret Van Der Kemp, P., and Boiteux, S. (2000) *Biochemistry* **39**, 1716–1724
728. Kang, D., Nishida, J.-i, Iyama, A., Nakabeppu, Y., Furuichi, M., Fujiwara, T., Sekiguchi, M., and Takeshige, K. (1995) *J. Biol. Chem.* **270**, 14659–14665
729. Bessman, M. J., Frick, D. N., and O'Handley, S. F. (1996) *J. Biol. Chem.* **271**, 25059–25062
730. Masaoka, A., Terato, H., Kobayashi, M., Honsho, A., Ohyama, Y., and Ide, H. (1999) *J. Biol. Chem.* **274**, 25136–25143
731. Kung, H. C., and Bolton, P. H. (1997) *J. Biol. Chem.* **272**, 9227–9236
732. Demple, B., and Karran, P. (1983) *Trends Biochem. Sci.* **8**, 137–139
733. Spratt, T. E., Wu, J. D., Levy, D. E., Kanugula, S., and Pegg, A. E. (1999) *Biochemistry* **38**, 6801–6806
734. Myers, L. C., Verdine, G. L., and Wagner, G. (1993) *Biochemistry* **32**, 14089–14094
734a. Hashimoto, H., Inoue, T., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., and Kai, Y. (1999) *J. Mol. Biol.* **292**, 707–716
734b. Lin, Y., Dötsch, V., Wintner, T., Peariso, K., Myers, L. C., Penner-Hahn, J. E., Verdine, G. L., and Wagner, G. (2001) *Biochemistry* **40**, 4261–4271
735. Asaeda, A., Ide, H., Asagoshi, K., Matsuyama, S., Tano, K., Murakami, A., Takamori, Y., and Kubo, K. (2000) *Biochemistry* **39**, 1959–1965
736. Hollis, T., Ichikawa, Y., and Ellenberger, T. (2000) *EMBO J.* **19**, 758–766
737. Lutsenko, E., and Bhagwat, A. S. (1999) *J. Biol. Chem.* **274**, 31034–31038
738. Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J. J., Wiebauer, K., and Jiricny, J. (1996) *J. Biol. Chem.* **271**, 12767–12774
739. Um, S., Harbers, M., Benecke, A., Pierrat, B., Losson, R., and Chambon, P. (1998) *J. Biol. Chem.* **273**, 20728–20736
740. Haber, J. E. (1999) *Trends Biochem. Sci.* **24**, 271–275
741. Cox, M. M., Goodman, M. F., Kreuzer, K. N., Sherratt, D. J., Sandler, S. J., and Marians, K. J. (2000) *Nature (London)* **404**, 37–41
741a. Connelly, J. C., and Leach, D. R. F. (2002) *Trends Biochem. Sci.*, **27**, 410–418
741b. Hopfner, K.-P., Craig, L., Moncalian, G., Zinkel, R. A., Usui, T., Owen, B. A. L., Karcher, A., Henderson, B., Bodmer, J.-L., McMurray, C. T., Carney, J. P., Petrini, J. H. J., and Tainer, J. A. (2002) *Nature (London)* **418**, 562–566
742. Webb, B. L., Cox, M. M., and Inman, R. B. (1999) *J. Biol. Chem.* **274**, 15367–15374
742a. Johnson, R. D., and Jasin, M. (2000) *EMBO J.* **19**, 3398–3407
742b. Arnaudeau, C., Lundin, C., and Helleday, T. (2001) *J. Mol. Biol.* **307**, 1235–1245
742c. Kondo, T., Wakayama, T., Naiki, T., Matsumoto, K., and Sugimoto, K. (2001) *Science* **294**, 867–870
742d. Hoege, C., Pfander, B., Moldovan, G.-L, Pyrowolakis, G., and Jentsch, S. (2002) *Nature (London)* **419**, 135–141
743. Yamaguchi-Iwai, Y., Sonoda, E., Sasaki, M. S., Morrison, C., Haraguchi, T., Hiraoka, Y., Yamashita, Y. M., Yagi, T., Takata, M., Price, C., Kakazu, N., and Takeda, S. (1999) *EMBO J.* **18**, 6619–6629
743a. Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M., Devine, S. K., Day J. P., Wilkinson, A., d'Adda di Fagagna, F., Devine, K. M., Bowater, R. P., Jeggo, P. A., Jackson, S. P., and Doherty, A. J. (2002) *Science* **297**, 1686–1689
744. Hays, S. L., Firmenich, A. A., and Berg, P. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6925–6929
745. Critchlow, S. E., and Jackson, S. P. (1998) *Trends Biochem. Sci.* **23**, 394–398
746. Wilson, T. E., and Lieber, M. R. (1999) *J. Biol. Chem.* **274**, 23599–23609
746a. Fugmann, S. D. (2002) *Nature (London)* **416**, 691–694
746b. Ooi, S. L., Shoemaker, D. D., and Boeke, J. D. (2001) *Science* **294**, 2552–2556
747. Jackson, S. P., and Jeggo, P. A. (1995) *Trends Biochem. Sci.* **20**, 412–415
748. Chu, G. (1997) *J. Biol. Chem.* **272**, 24097–24100
749. Hammarsten, O., DeFazio, L. G., and Chu, G. (2000) *J. Biol. Chem.* **275**, 1541–1550
750. Brown, K. D., Lataxes, T. A., Shangary, S., Mannino, J. L., Giardina, J. F., Chen, J., and Baskaran, R. (2000) *J. Biol. Chem.* **275**, 6651–6656
750a. Walker, J. R., Corpina, R. A., and Goldberg, J. (2001) *Nature (London)* **412**, 607–614
750b. Rouse, J., and Jackson, S. P. (2002) *Science* **297**, 547–551
750c. Kolodner, R. D., Putnam, C. D., and Myung, K. (2002) *Science* **297**, 552–557
750d. Carr, A. M. (2002) *Science* **297**, 557–558
750e. Cha, R. S., and Kleckner, N. (2002) *Science* **297**, 602–606
750f. Sogo, J. M. Lopes, M., and Foiani, M. (2002) *Science* **297**, 599–602
750g. Celeste, A., and 20 other authors (2002) *Science* **296**, 922–927
750h. Bird, A. W., Yu, D. Y., Pray-Grant, M. G., Qiu, Q., Harmon, K. E., Megee, P. C., Grant, P. A., Smith, M. M., and Christman, M. F. (2002) *Nature (London)* **419**, 411–415
750i. Pickart, C. M. (2002) *Nature (London)* **419**, 120–121
751. Flanders, P. H. (1981) *Sci. Am.* **245**(Nov), 72–80
752. Kenyon, C. J. (1983) *Trends Biochem. Sci.* **8**, 84–87
753. Walker, G. C. (1995) *Trends Biochem. Sci.* **20**, 416–420
753a. Yasuda, T., Morimatsu, K., Kato, R., Usukura, J., Takahashi, M., and Ohmori, H. (2001) *EMBO J.* **20**, 1192–1202
754. McDonald, J. P., Peat, T. S., Levine, A. S., and Woodgate, R. (1999) *J. Mol. Biol.* **285**, 2199–2209
754a. Goodman, M. F. (2002) *Ann. Rev. Biochem.* **71**, 17–50
754b. Livneh, Z. (2001) *J. Biol. Chem.* **276**, 25639–25642
755. Johnson, R. E., Washington, M. T., Prakash, S., and Prakash, L. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12224–12226
756. Baynton, K., and Fuchs, R. P. P. (2000) *Trends Biochem. Sci.* **25**, 74–79
757. Murakumo, Y., Roth, T., Ishii, H., Rasio, D., Numata, S.-i, Croce, C. M., and Fishel, R. (2000) *J. Biol. Chem.* **275**, 4391–4397
758. Wood, R. D. (1999) *Nature (London)* **399**, 639–640
759. Daube, S. S., Tomer, G., and Livneh, Z. (2000) *Biochemistry* **39**, 348–355
759a. Matsuda, T., Bebenek, K., Masutani, C., Rogozin, I. B., Hanaoka, F., and Kunkel, T. A. (2001) *J. Mol. Biol.* **312**, 335–346

References

759b. Kusumoto, R., Masutani, C., Iwai, S., and Hanaoka, F. (2002) *Biochemistry* **41**, 6090–6099
760. Hübscher, U., Nasheuer, H.-P., and Syväoja, J. E. (2000) *Trends Biochem. Sci.* **25**, 143–147
760a. Friedberg, E. C., Wagner, R., and Radman, M. (2002) *Science* **296**, 1627–1630
760b. Haracska, L., Johnson, R. E., Unk, I., Phillips, B. B., Hurwitz, J., Prakash, L., and Prokash, S. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14256–14261
760c. Johnson, R. E., Prakash, S., and Prakash, L. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3838–3843
760d. García-Díaz, M., Domínguez, O., López-Fernández, L. A., de Lera, L. T., Saníger, M. L., Ruiz, J. F., Párraga, M., García-Ortiz, M. J., Kirchhoff, T., del Mazo, J., Bernad, A., and Blanco, L. (2000) *J. Mol. Biol.* **301**, 851–867
760e. Wang, Z., Castano, I. B., De Las Penas, A., Adams, C., and Christman, M. F. (2000) *Science* **289**, 774–779
760f. Frank, E. G., Tissier, A., McDonald, J. P., Rapic-Otrin, V., Zeng, X., Gearhart, P. J., and Woodgate, R. (2001) *EMBO J.* **20**, 2914–2922
760g. Shimizu, K., Kawasaki, Y., Hiraga, S.-I., Tawaramoto, M., Nakashima, N., and Sugino, A. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9133–9138
761. de Murcia, G., and de Murcia, J. M. (1994) *Trends Biochem. Sci.* **19**, 172–176
761a. Jacobson, M. K., and Jacobson, E. L. (1999) *Trends Biochem. Sci.* **24**, 415–417
762. Lindahl, T., Satoh, M. S., Poirier, G. G., and Klungland, A. (1995) *Trends Biochem. Sci.* **20**, 405–411
763. Aoufouchi, S., Yélamos, J., and Milstein, C. (1999) *J. Mol. Biol.* **290**, 943–949
763a. Smith, S. (2001) *Trends Biochem. Sci.* **26**, 174–179
764. Ames, B. N., Magaw, R., and Gold, L. S. (1987) *Science* **236**, 271–280
765. Gold, L. S., Slone, T. H., Stern, B. R., Manley, N. B., and Ames, B. N. (1992) *Science* **258**, 261–265
766. Miller, J. A., and Miller, E. C. (1976) *Fed. Proc.* **35**, 1316–1321
767. Yu, H., Eritja, R., Bloom, L. B., and Goodman, M. F. (1993) *J. Biol. Chem.* **268**, 15935–15943
768. Goodman, M. F., and Ratliff, R. L. (1983) *J. Biol. Chem.* **258**, 12842–12846
769. Sowers, L. C., Fazakerley, G. V., Eritja, R., Kaplan, B. E., and Goodman, M. F. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5434–5438
770. Ripley, L. S., and Clark, A. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6954–6958
771. Ginell, S. L., Kuzmich, S., Jones, R. A., and Berman, H. M. (1990) *Biochemistry* **29**, 10461–10465
772. Wolff, I. A., and Wasserman, A. E. (1972) *Science* **177**, 15–18
773. Russell, W. L., Kelly, E. M., Hunsicker, P. R., Bangham, J. W., Maddux, S. C., and Phipps, E. L. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5818–5819
774. Shaver-Walker, P. M., Urlando, C., Tao, K. S., Zhang, X. B., and Heddle, J. A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11470–11474
775. Wurdeman, R. L., Church, K. M., and Gold, B. (1989) *J. Am. Chem. Soc.* **111**, 6408–6412
776. Klein-Szanto, A. J. P., Iizasa, T., Momiki, S., Garcia-Palazzo, I., Caamano, J., Metcalf, R., Welsh, J., and Harris, C. C. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6693–6697
777. Craddock, V. M. (1983) *Nature (London)* **306**, 638
778. Hirai, N., Kingston, D. G. I., Van Tassell, R. L., and Wilkins, T. D. (1982) *J. Am. Chem. Soc.* **104**, 6149–6150
779. Gupta, I., Suzuki, K., Bruce, W. R., Krepinsky, J. J., and Yates, P. (1984) *Science* **225**, 521–522
780. Ozawa, N., and Guengerich, F. P. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5266–5270
781. Langouët, S., Mican, A. N., Müller, M., Fink, S. P., Marnett, L. J., Muhle, S. A., and Guengerich, F. P. (1998) *Biochemistry* **37**, 5184–5193
782. Leonard, G. A., McAuley-Hecht, K. E., Gibson, N. J., Brown, T., Watson, W. P., and Hunter, W. N. (1994) *Biochemistry* **33**, 4755–4761
783. Weisenseel, J. P., Moe, J. G., Reddy, G. R., Marnett, L. J., and Stone, M. P. (1995) *Biochemistry* **34**, 50–64
784. Palejwala, V. A., Simha, D., and Humayun, M. Z. (1991) *Biochemistry* **30**, 8736–8743
785. Zegar, I. S., Setayesh, F. R., DeCorte, B. L., Harris, C. M., Harris, T. M., and Stone, M. P. (1996) *Biochemistry* **35**, 4334–4348
786. Bucala, R., Model, P., and Cerami, A. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 105–109
787. Shires, T. K., Tresnak, J., Kaminsky, M., Herzog, S. L., and Truc-pham, B. (1990) *FASEB J.* **4**, 3340–3346
788. Napolitano, R. L., Lambert, I. B., and Fuchs, R. P. P. (1994) *Biochemistry* **33**, 1311–1315
789. Isono, K., and Yourno, J. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1612–1617
790. Stoloff, L. (1987) *Science* **237**, 1283
791. Baertschi, S. W., Raney, K. D., Stone, M. P., and Harris, T. M. (1988) *J. Am. Chem. Soc.* **110**, 7929–7931
792. Iyer, R. S., Coles, B. F., Raney, K. D., Thier, R., Guengerich, F. P., and Harris, T. M. (1994) *J. Am. Chem. Soc.* **116**, 1603–1609
793. Heidelberger, C. (1975) *Ann. Rev. Biochem.* **44**, 79–121
794. Phillips, D. H. (1983) *Nature (London)* **303**, 468–472
795. Flowers, L., Bleczinski, W. F., Burczynski, M. E., Harvey, R. G., and Penning, T. M. (1996) *Biochemistry* **35**, 13664–13672
796. Schurter, E. J., Sayer, J. M., Oh-hara, T., Yeh, H. J. C., Yagi, H., Luxon, B. A., Jerina, D. M., and Gorenstein, D. G. (1995) *Biochemistry* **34**, 9009–9020
797. Mao, B., Gu, Z., Gorin, A., Chen, J., Hingerty, B. E., Amin, S., Broyde, S., Geacintov, N. E., and Patel, D. J. (1999) *Biochemistry* **38**, 10831–10842
798. Tang, M.-s, Bohr, V. A., Zhang, X.-s, Pierce, J., and Hanawalt, P. C. (1989) *J. Biol. Chem.* **264**, 14455–14462
799. Cho, B. P., and Zhou, L. (1999) *Biochemistry* **38**, 7572–7583
800. Eis, P. S., Smith, J. A., Rydzewski, J. M., Case, D. A., Boger, D. L., and Chazin, W. J. (1997) *J. Mol. Biol.* **272**, 237–252
801. Perera, F., and Petito, C. (1982) *Science* **216**, 1285–1291
802. Rall, D. P. (1992) *Science* **257**, 1330
803. Hanson, D. J. (1991) *Chem. Eng. News* **Aug 12**, 7–14
804. Marshall, E. (1990) *Science* **247**, 276–277
805. Hileman, B. (1990) *Chem. Eng. News* **May 7**, 4
806. Ames, B. N. (1979) *Science* **204**, 587–593 (see also discussion in Science 224, 659–670, 757–760 (1984))
807. Devoret, R. (1979) *Sci. Am.* **241**(Aug), 40–49
808. Muench, K. F., Misra, R. P., and Humayun, M. Z. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6–10
809. Green, M. M., Todo, T., Ryo, H., and Fujikawa, K. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6667–6671
810. Gold, M. O., Blum, A., and Ames, B. N. (1978) *Science* **200**, 785–787
811. Commoner, B., Vithayathil, A. J., Dolara, P., Nair, S., Madyastha, P., and Cuca, G. C. (1978) *Science* **201**, 913–916
812. Tennant, R. W., Margolin, B. H., Shelby, M. D., Zeiger, E., Haseman, J. K., Spalding, J., Caspary, W., Resnick, M., Stasiewicz, S., Anderson, B., and Minor, R. (1987) *Science* **236**, 933–941
813. Epstein, S. S., Ames, B. N., and Gold, L. S. (1988) *Science* **240**, 1043–1047
814. Ashby, J., and Morrod, R. S. (1991) *Nature (London)* **352**, 185–186
815. Weinstein, I. B. (1991) *Science* **251**, 387–388
816. Stone, R. (1995) *Science* **268**, 356–357
817. Randerath, K., Reddy, M. V., and Gupta, R. C. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6126–6129
818. Kadlubar, F. F. (1992) *Nature (London)* **360**, 189
819. Schut, H. A. J., and Shiverick, K. T. (1992) *FASEB J.* **6**, 2942–2951

Study Questions

1. Demethylation of 5-methylcytosine in DNA during early embryonic development has been proposed (see Chapter 32). Can you suggest one or more mechanisms by which such demethylation could occur?

2. Describe the structures and functions of histones and of nucleosomes in eukaryotic cells. Are there comparable proteins and structures in bacteria? Do you think that our knowledge of histones and nucleosomes is nearly complete?

3. List the major molecular components required for replication of DNA in *E. coli.* Describe briefly the functions of each protein or other component.

4. Compare replication in bacteria and in eukaryotes.

5. Compare systhesis of the leading and lagging strands in the elongation phase of DNA replication. Explain why DNA polymerases may have difficulty in replicating the 3'-end of the *lagging* strand of **linear** DNA. How has this problem been solved in many bacterial and viral systems? In eukaryotic cells?

6. The circular chromosome of an *E. coli* cell contains 4.6×10^6 base pairs. If a replication fork moves at a rate of ~1000 nucleotides per second, how much time will be required for replication of the DNA? Cells of *E. coli* can divide every 20 minutes under favorable conditions. How can you explain this rapid rate of growth?

7. DNA ligase, whose reaction is reversible, is able to relax supercoiled circular DNA in the presence of AMP but not in its absence. Outline the chemical mechanism of the ligase reaction. Why is it dependent on AMP? What other DNA ligase mechanism is known?

8. DNA polymerases involved in replication require a primer. Why? What is the nature of the primer?

9. Why is it essential for a cell to have several different mechanisms of DNA repair? Describe some of these mechanisms.

10. Why do cells use error-prone DNA polymerases under some circumstances?

11. Do you see a relationship of some types of DNA repair to the chemical events during meiosis?

12. Is O^6-methyltransferase an enzyme?

13. Why are high rates of mutation observed in regions of DNA that contain 5-methylcytosine?

14. Why is uracil-DNA glycosylase important in DNA repair? Is it important for DNA replication?

15. Why do cells exposed to visible light following irradiation by ultraviolet light have a greater survival rate than cells kept in the dark after UV irradiation?

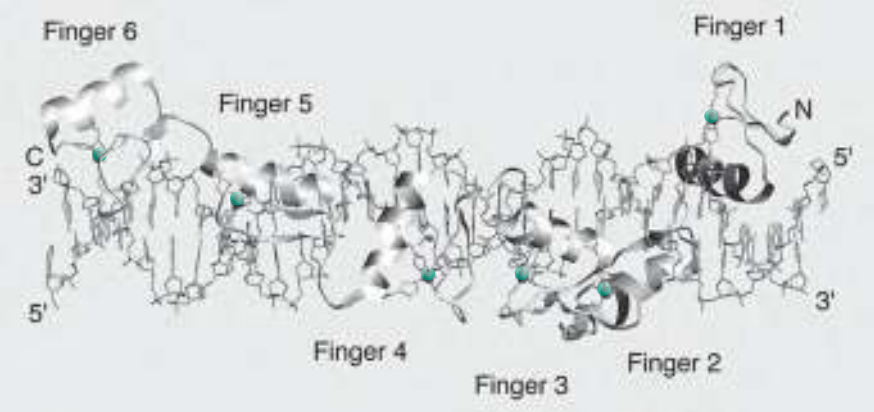
16. Can exposure of *E. coli* to nitrous acid (HNO_2) lead to mutation of a $tRNA^{GLY}$ to an amber suppressor? The Gly codons are GGX (where X = any nucleotide) and the amber codon is UAG.

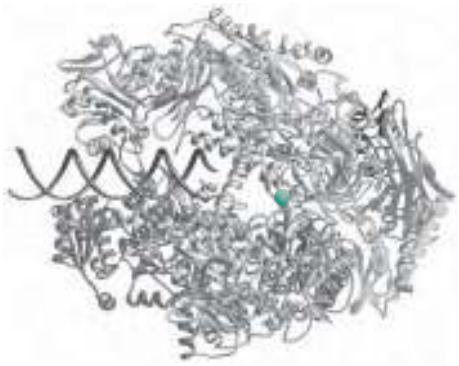
Left. The N-terminal 190-residue fragment of transcription factor TFIIIA of *Xenopus laevis* bound to a 31 bp DNA segment of the promoter region for 5S ribosomal RNA. Six zinc finger motifs (zinc atoms are green) bind in several ways into the major groove of DNA and across the minor groove. From Nolte *et al*, *Proc. Natl. Acad. Sci. USA* **95**, 2938–2943, 1998. Courtesy of Raymond S. Brown. Right. Ribbon drawing of the three-dimensional structure of a 10-subunit form of yeast RNA polymerase II, which transcribes genes to form messenger RNA. A 20 base pair segment of B-DNA has been modeled but the transcription bubble is not shown. The active site Mg^{2+} is green. From Cramer *et al*, *Science* **288**, 640–649, 2000. Courtesy of Roger D. Kornberg.

Contents

The Transcription of Genes

28





The copying of genetic information from DNA into messenger RNA is the initial step in the chain of reactions leading to synthesis of the multitude of proteins and specialized RNA molecules needed by cells. The requirement for these macromolecules varies with conditions, and in eukaryotic cells, with the stage of differentiation. Therefore, it is not surprising that transcription is highly controlled.

Cells make four principal kinds of RNA: ribosomal (rRNA), transfer (tRNA), messenger (mRNA), and a variety of small RNAs. The last, which range in length from a few up to several hundred nucleotide units,[1–1b] are designated variously as sRNAs, ncRNAs, miRNAs, siRNAs, snRNAs, and snoRNAs. The abbreviation *s*, *nc*, *mi*, *si*, *sn*, and *sno* stand for small, noncoding, micro, silencing, small nuclear, and small nucleolar, respectively. All of these RNAs are synthesized as larger transcripts, which undergo cleavage and other modifications within the cell. Therefore, a second major topic in this chapter is the processing of RNA precursors. We will also consider the fact that cells may be hosts for RNA viruses, may occasionally harbor RNA plasmids, and must sometimes transcribe viral DNA.

The absence of a nuclear membrane is a characteristic of bacteria that has a profound effect on transcription. Bacterial transcripts are processed rapidly, and their 5' ends often enter ribosomes and are directing protein synthesis, while the 3' ends of the genes are still being transcribed. In contrast, most eukaryotic RNA transcripts must be processed and transported out of the nucleus before they can function. As consequence, many aspects of the control of transcription differ between prokaryotes and eukaryotes.

A. Transcription and Processing of RNA in Bacteria

Even after the existence of mRNA had been recognized, it was not obvious how formation of single-stranded (ss) RNA would be accomplished, using a double-stranded (ds) template. The fact that purified RNA polymerases can synthesize RNA from the four ribonucleoside triphosphates using ssDNA as the template suggested that transcription, like DNA replication, involves base pairing. In line with this conclusion was the fact that the ssDNA obtained from bacteriophage ϕX174 was converted by RNA polymerases into a dsRNA–DNA hybrid. However, when dsDNA served as the template, free ssRNA was formed. Thus, it appeared likely that at the site of the polymerase action the dsDNA was momentarily pulled apart into single strands and that one of these was copied by the polymerase. More recent experiments have confirmed this view.

1. The *lac* Operon

Much of the terminology used to describe the control of transcription originated with Jacob and Monod. Based on studies of the induction of enzymes in bacteria they proposed the **operon model**.[1c–3] An operon is a regulated cluster of genes, one of which is shown diagrammatically in Fig. 28-1. This is the *lac* operon of *E. coli*. Found at position 8 min on the genetic map of Fig. 26-4, it is probably the most intensively studied group of *E. coli* genes. Three structural genes encode the amino acid sequences of β-galactosidase (*lacZ*),[3a]

permease (*lacY*),[4] and a transacetylase (acetyltransferase, *lacA*), which transfers acetyl groups from acetyl-CoA to β galactosides. To account for the apparently synchronous control of these three genes, Jacob and Monod proposed that they function as a **transcriptional unit** the operon, which encodes a single molecule of mRNA. They proposed that each operon is controlled by a segment of the DNA molecule located at the beginning of the operon, i.e., at the 5' end of the coding chain or 3' end of the template chain. The first part of this **control region** they called the **promoter** (*P*). The promoter is the site of the initial binding of the RNA polymerase to the DNA, the binding constants for the association being very high. The rates of association and of initiation may be influenced strongly by various other proteins. One of these, the **catabolite gene activator protein** (**CAP**; also called cAMP receptor protein, CRP), is important to the *lac* operon. It also binds in the promoter region (Fig. 28-1) and stimulates transcription.

Repression and induction. Immediately adjacent to the promoter is the **operator** (*O*), which is a binding site for a **repressor** (R). When the operator is free, transcription is initiated and proceeds through the operator region and on to the genes coding for the three proteins. On the other hand, if the repressor is bound to the operator, transcription is blocked. When the operon model was first proposed, the chemical nature of the repressor was unknown. However, many repressors have been identified as oligomeric proteins able to undergo allosteric alteration. The *lac* repressor is made up of four identical 360-residue subunits. Each subunit has a helix–turn–helix binding domain that is specific for the DNA sequence of the operator and an allosteric binding site for an effector.[5,6] The drawing in Fig. 28-1 is simplified to show only two of the four subunits (see also Chapter 5, Section F,1).

The *lac* operon is ordinarily subject to repression and is activated by the presence of an **inducer**, now known to be **allolactose**, D-Gal*p*-β1→6-D-Glc. However, in experimental work artificial inducers such as **isopropyl-β-D-thiogalactoside** (**IPTG**) are most often used. Jacob and Monod postulated that the free repressor protein binds to the operator. In the presence of the inducer a conformational change takes place, destroying the affinity of the repressor protein for the operator site. Thus, in the presence of inducer the operator is not blocked, and the transcription takes place. Such an operon is said to be **negatively controlled** and **inducible**.

Important to the control of the operon is the **regulatory gene,** which codes for the synthesis of the repressor protein. In the case of the *lac* operon, the regulatory gene (the *I* gene) is located immediately preceding the operon itself (Fig. 28-1). However, for some operons the regulatory gene is located a considerable distance away. For example, the *gal* operon of *E. coli*, which codes for enzymes of galactose metabolism, is found at map position 17 min, while the regulatory gene is at 61 min.[7]

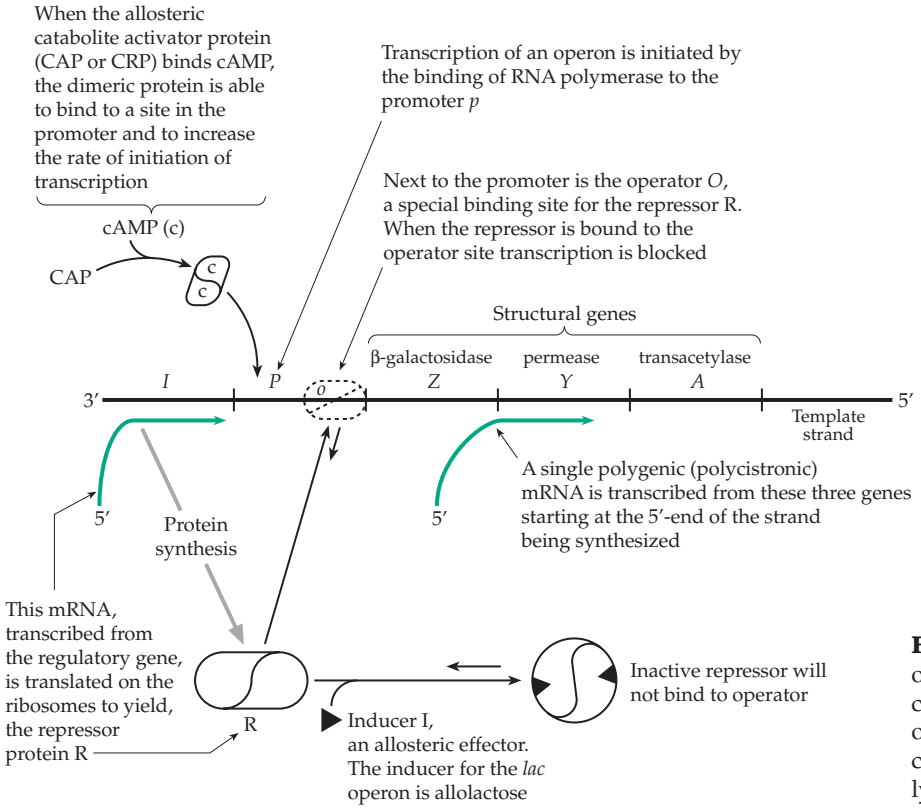


Figure 28-1 Schematic representation of the *lac* operon of *E. coli* and of its control. Here only the template strand of the DNA is shown. However, the coding (nontranscribed) strand is usually the one labeled, as in Fig. 28-2.

Regulatory genes are normally transcribed at a slow but steady rate, presumably because RNA polymerase initiates RNA chains infrequently at the promoter sites of regulatory genes. Thus, each cell of *E. coli* normally contains only about ten molecules of the *lac* repressor protein. A mutation in a regulatory gene may lead to a defective repressor, which no longer binds at the operator. Then, transcription of the operon is uncontrolled, and mRNA is produced in greater amounts. In such a mutant strain (designated I^- in contrast to the normal I^+ strain) production of the enzyme representing the gene product becomes constitutive, just as is the formation of the enzymes of the central pathways of metabolism. The latter enzymes also appear to be produced regularly in large amounts without control by a repressor, establishing that *transcription* rather than translation regulates the expression of these genes. The operon can also become unregulated, i.e., **constitutively expressed**, if a mutation occurs in the operator site and the repressor can no longer bind because of the altered DNA sequence.

Nucleotide sequence of the lac control region.

The sequence of the *E. coli* DNA representing the promoter-operator region of the *lac* operon is shown in Fig. 28-2. It includes the end of the *I* gene, at the left, and the beginning of the *Z* gene, at the right.[2] The series of codons representing the peptide sequence Glu-Ser-Gly-Gln-Stop at the left-hand end corresponds to the known C-terminal sequence of the repressor, while the three codons at the right are those of formyl-Met-Thr-Met, the known N-terminal sequence of the *Z* gene product β-galactosidase. Detailed genetic mapping of the region in 1973 made it possible to assign operator and promoter regions with confidence as indicated. The mRNA transcript begins in the operator region as shown in the figure. The initiation codon for the *Z* gene is 39 bases from the end of the transcript. In this figure the original numbering of the nucleotides is printed in black. Now it is customary to number from the transcription initiation position (+1). Green numbers are used in this way in Fig. 28-2. Positions to the left and toward the 5' end of the nontranscribed coding strand precede the initiation position. They are referred to as **upstream** and are numbered with negative integers. Positions to the right are **downstream** and numbered with positive integers.

The operator region was located by digesting the DNA with deoxyribonuclease in the presence of the repressor protein.[8] The bound repressor protected a region of 27 base pairs as indicated in the figure. The operator is centered on a region of local twofold rotational symmetry (Chapter 5; Fig. 5-34). The symmetry is not perfect, the sequence being **quasipalindromic**. The following precisely symmetric synthetic sequence, which contains an 11-bp inverted repeat of the left half of the *lac* operator sequence, binds *lac* repressor 8-fold more tightly than does the natural *E. coli* operator.[9]

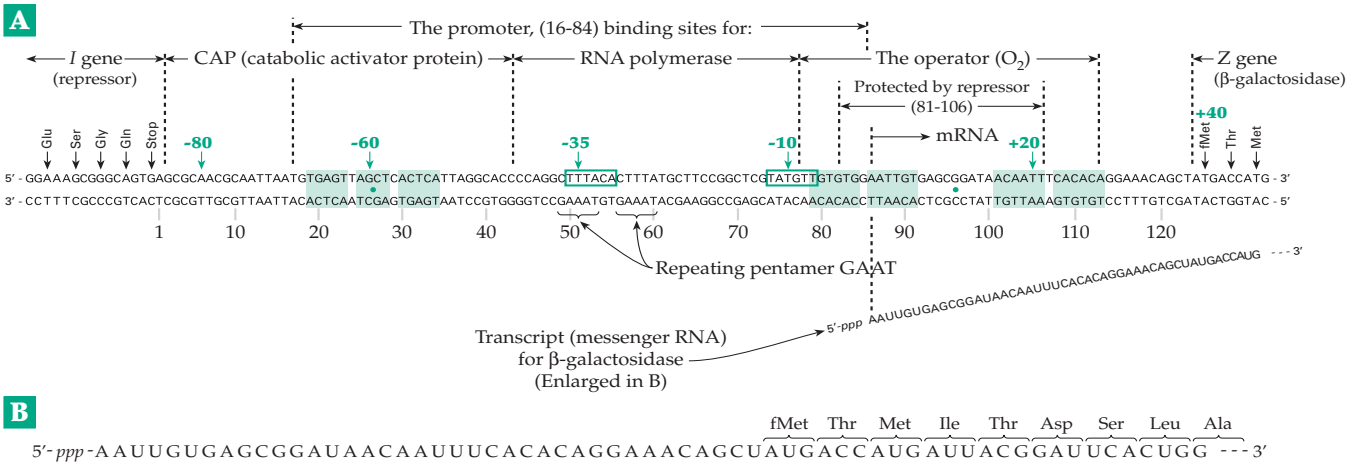
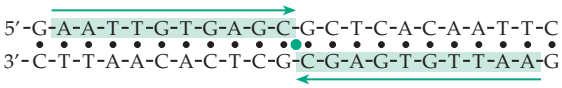


Figure 28-2 (A) Nucleotide sequence of the *lac* promoter-operator region of the *E. coli* chromosome.[7] The proposed locations of the *I* gene, the promoter (which contains CAP and RNA polymerase binding sites), the operator, and the beginning of the Z gene (β-galactosidase) are shown. Note the two regions of local 2-fold rotational symmetry, which are marked by bars and central dots and the repeating pentamer. Positions upstream (–) or downstream (+) from the +1 start position for transcription are marked in green. The –10 (Pribnow) and –35 promoter elements are boxed on the coding strand (top). Labels are usually applied to the coding strand. The mRNA is copied from the complementary template strand (bottom). (B) The sequence of an mRNA molecule initiated in the *lac* promoter-operator region of a mutant strain of *E. coli* with an altered promoter.[10] The peptide initiation amino acid is identified by the symbol fMet, and the successive amino acids from the known N-terminal sequence of β-galactosidase have been matched with the codons.



The dyad symmetry of the operator sequence is probably important in providing tight binding to two subunits of the symmetric tetrameric protein.[11–13] It is also possible that repressor molecules move along DNA chains in a one-dimensional diffusion process, and that the symmetry of the operator site facilitates recognition by a protein moving from either direction.[14,15]

The repressor structure. The *lac* repressor protein is a member of the **helix–turn–helix** family of DNA binding proteins (Fig. 28-3; see also Fig. 5-35). The first helix and the turn of this motif fit into the major groove of the DNA, the side chains from the helix interacting with specificity-determining groups in the major groove. The protein consists of three domains; the N-terminal DNA binding "head" (residues ~1– ~50), a core domain (residues 62–340), and a leucine heptad repeat domain (residues 340–360) that forms the dimer–dimer interface. The dimeric form of the repressor binds to the palindromic operator sequence. The tetramer can bind to two operator sequences.[16–20] The inducer IPTG binds to the core domain near the ONPF site shown in Fig. 28-3. The sequence, structure, and binding site of the core domain resemble those of sugar transport proteins such as the galactose-binding chemoreceptor protein (Fig. 4-18A). Binding of the inducer causes the conformational change that appears to disrupt the interactions of the "hinge helices," seen in the center of Fig. 28-3A, with the DNA. This causes the repressor to dissociate from the operator (Fig. 28-1) and allows RNA polymerase to bind and to initiate transcription.

The *E. coli lac* repressor is one of the most investigated of all proteins. For example, 4000 single-amino-acid mutants have been prepared and studied.[6,17,21] Suppressor mutations were used to determine the function of various portions of the protein.[22,23] Many of the mutant proteins were created using *amber* mutations that were induced in the gene at many positions. The mutated genes were transferred into plasmids for cloning. Each plasmid was used to infect five different strains of bacteria, each carrying a suppressor mutation that would introduce a different amino acid when the (termination) codon UAG was encountered (see Chapter 29, Section C,3). From these infected bacteria large quantities of the mutant forms of the *lac* repressor were isolated. It was found that many mutations near the N-terminal end interfered with binding of the repressor to DNA, whereas mutations near the center interfere with binding to the inducer.

In addition to the main *lac* operator O_1, which is marked in Fig. 28-2, there are two weaker auxiliary operator sequences designated O_2 and O_3 located 401

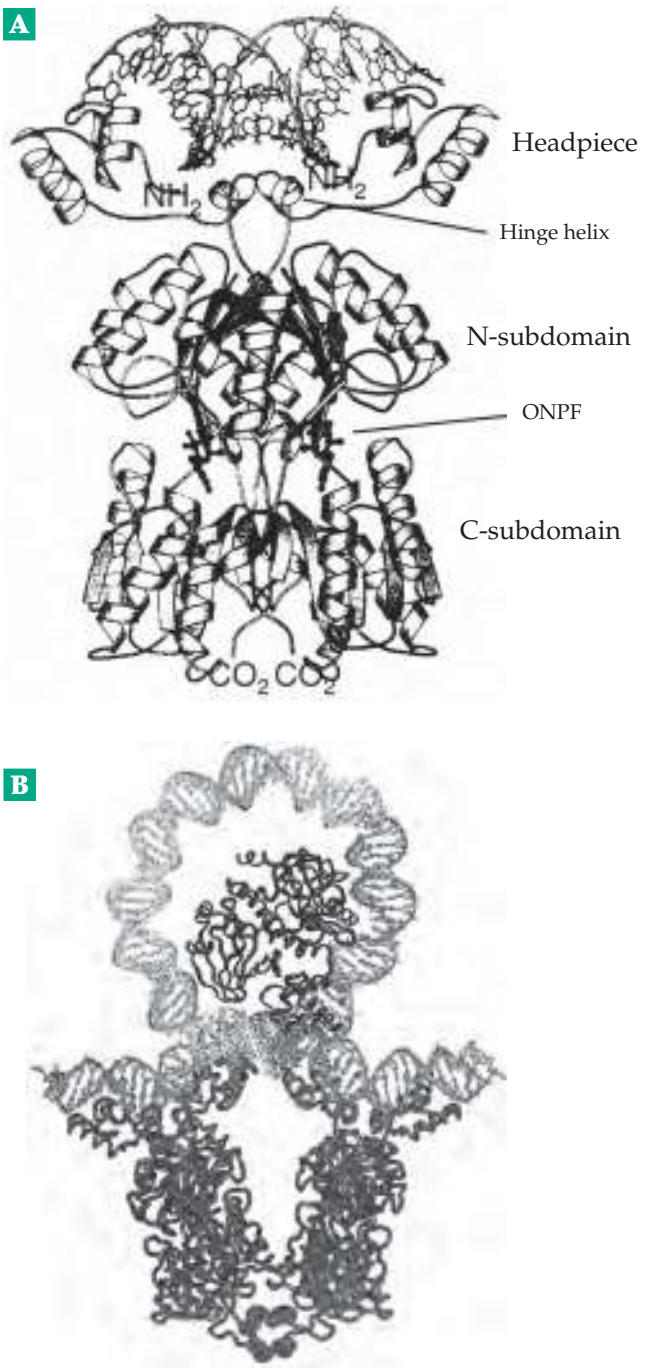


Figure 28-3 (A) Ribbon view of the dimeric *lac* repressor bound to a natural operator and to the anti-inducer *o*-nitrophenylfucoside (ONPF). The headpiece (residues 2–46) and the hinge helix (residues 50–58) form the DNA-binding domains. The core (residues 62–330), which is divided into N- and C-terminal subdomains, forms the binding site for ONPF. The C-terminal residues 334–360, which form a tetramerization domain, are absent from this MolScript drawing. Notice that the hinge helices bind to and widen the minor groove at the center of the operator. From Lewis *et al.*[5a] (B) Model of a 93-bp DNA loop corresponding to residues –82 to +11 of the *lac* operon (Fig. 28-2) bound to the tetrameric *lac* repressor. The active sites of the repressor are bound to the major operator O_1 and to the secondary operator O_3. From Lewis *et al.*[5]

bp downstream of O_1 in the lacZ coding region, and 92 bp upstream of O_1, respectively.[24] The DNA can bind to both O_1 and either O_2 or O_3 with a loop between them as in Fig. 28-3B. Binding of the 10–20 copies of the lac repressor present in a cell of *E. coli* to the operator sequence is so tight that expression of the genes controlled is reduced 1000-fold. However, when placed in front of other operators or in different positions relative to the Pribnow sequence (Fig. 28-2), its effectiveness varied greatly. The extent of repression appears to be affected both by thermodynamic factors[12,13] and by relative rates of repressor binding and of RNA polymerase movement (see also Eq. 28-1).[25]

2. Initiation of Transcription

The rate of RNA synthesis varies from one operon to another. Sequences of promoters, operators, and other control sequences as well as the state of repressors and activator proteins all affect these rates.[10,26,27] However, in every instance the first steps in transcription involve the binding of RNA polymerase to DNA.

Bacterial RNA polymerase. Most RNA polymerases (**RNAPs**) are large multisubunit proteins. However, bacterial viruses sometimes induce their own RNA polymerases, and these may be monomeric. For example, the 99-kDa (883-residue) phage T7-encoded polymerase is a single peptide chain with a structure and two-metal-ion active site resembling those of *E. coli* DNA polymerase I.[28–30a] It is able to carry out all of the steps of the transcription cycle of the virus. In contrast, the most studied bacterial RNAP, that from *E. coli*, consists of five kinds of subunits[31–33d] with the composition $\alpha_2\beta\beta'\sigma\omega$. A similar composition has been found for RNAPs of other bacteria.[33e] Functions of the five subunits can be correlated directly with components of archaeal and eukaryotic RNAPs.[33f] However, the latter contain additional subunits. The two α subunits in the *E. coli* enzyme have identical sequences, but their locations and interactions are different.[33a–c]

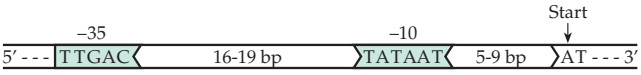
Gene symbol	Subunit	Molecular mass (kDa)	Number of amino acid residues
α	*rpo*A	36.5	329
β	*rpo*B	150.6	1342
β′	*rpo*C	155.2	1407
σ	*rpo*D	70.2	
ω	*rpo*Z	6	

The three-dimensional structure of the *E. coli* RNAP bound to DNA in an initiation complex shows that the enzyme forms a groove into which the DNA can fit. It can then close to form a tunnel in which the template and nontemplate strands are separated (Fig. 28-4). The polymerase is present in large amounts, ~3000 molecules per cell in E. coli.[26] The β and β′ subunits of the *E. coli* enzyme each contain an essential zinc ion. The Zn^{2+} in the β′ subunit is present in a zinc finger motif near the N terminus.[34] It is thought to interact with the template strand of the DNA. The Zn^{2+} in the β strand is more loosely bound.[33,35] The active site is largely in the β subunit. In the assembly of the RNAP complex a dimer of the small α subunit binds to β, after which β′ is added.[36] The sequences of the β and β′ subunits have several highly conserved regions with homologous sequences from bacteria and from functionally equivalent regions in eukaryotic RNAPs. Three aspartate residues in the sequence NADFDGD may chelate two Mg^{2+} ions as in the active site of DNA polymerases (Fig. 27-13).[37] The basic chemistry of all of the polymerases may be similar, but the modular structure of the bacterial RNAP differs markedly from that of the DNA Pol I family.[38]

Of the RNA polymerase subunits σ (sigma) plays a unique role in initiation of transcription. It is required for the recognition of promoter sites.[39] However, it is not needed for elongation of an RNA chain and dissociates from the $\alpha_2\beta\beta'$ core complex soon after transcription is initiated. In a given bacterial species there is one predominant σ factor, but there are often smaller amounts of other σ factors with homologous sequences.[26,40,40a] In *E. coli* σ^{70} (where the superscript number is the molecular mass in kDa) is predominant, but other specialized σ subunits recognize different groups of promoters. For example, $\sigma^N(\sigma^{54})$ binds to promoters that allow transcription of genes involved in assimilation of nitrogen[41,41a] as well as in aromatic catabolism.[42] Synthesis of protein σ^S of *E. coli* is induced by stress such as carbon starvation.[43] At high temperatures (e.g., 40–49°C) σ^{32} is synthesized and permits transcription of genes for "heat shock proteins."[44] Actively growing cells of *Bacillus subtilis* contain at least five different sigma factors. An additional four control gene expression during spore formation.[45]

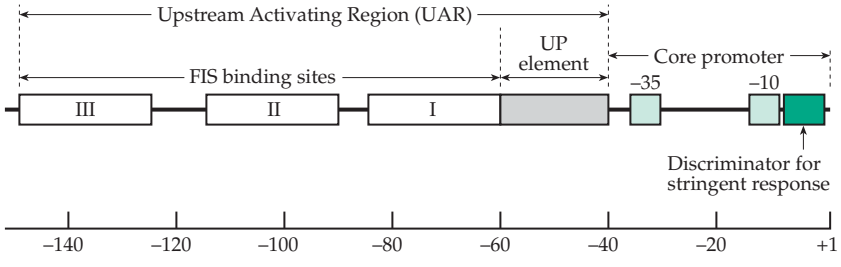
Promoter sequences. In 1975, Pribnow pointed out[46] that a series of six known promoters had a conserved 7-base sequence beginning six nucleotides upstream from the initiation site for transcription. Although this sequence varies somewhat from one promoter to another, it has been found in hundreds of *E. coli* promoters. This is called the **–10 region**, the **Pribnow sequence,** or Pribnow box (the last in recognition of the fact that people like to draw boxes around these special sequences). A typical 6-base consensus Pribnow sequence is 5'-TATAAT as written for the coding strand, whose sequence corresponds to that of the mRNA. Only three of these bases are highly

conserved: 5'-TA_ _ _ T. For example, in the *lac* promoter (Fig. 28-2) the sequence is TATGTT. The nucleotides of the –10 consensus sequence are present with the following frequencies (as percentages): T(80)A(95)T(45)A(60)A(50)T(96).[47] The position of the –10 sequence is not exactly the same in all promoters but usually starts 5–9 bp upstream of the start position for transcription. About 16–19 bp upstream from the Pribnow sequence is another conserved sequence, the **–35 region**. A consensus sequence is TTGACA, the TTG sequence to the left being the most highly conserved. Both –10 and –35 regions are needed for efficient promoter activity. Lewin suggested the following "optimal" promoter sequence.[47] The location of these two regions in transcription initiation complex is shown in Fig. 28-4B.



Naturally occurring promoters usually do not have the exact –10 and –35 consensus sequences, but artificially constructed promoters containing them are highly effective *in vivo*.[26] The fact that most promoters depart from these "ideal" sequences is not surprising because cells need varying amounts of different proteins. Promoter strengths vary over a range of at least 10^4. Much of this variation comes from variations in the specific –10 and –35 sequences, which appear to be specifically recognized by RNA polymerase. This variation includes an upward extension of the –10 region[48] for some promoters. Activator-binding sites are also often present in an **upstream activating region** (**UAR**; see following diagram). This may extend from the –35 sequence through the remainder of the promoter region and as far upstream as –200 to –500 bp. For example, the CAP-binding site in the *lac* operon DNA is centered at –60 (Fig. 28-2). In spite of the variation it was possible to locate 2584 operators (of which only 392 were previously known) and to predict the location of 2405 promoters, when the complete *E. coli* genome sequence became known.[49]

The binding of the various RNAP subunits, repressors, and activators has been studied using mutant promoter sequences,[31,37] antibiotic-binding sites,[50,51] chemical crosslinking reagents,[34,37] and a cysteine-tethered Fe-dependent DNA-cutting reagent.[32,52] Other "footprinting" techniques, e.g., observing cleavage of the DNA by hydroxyl radicals generated by reduction of H_2O_2 by Fe(II) (Fig. 5-50), have also been employed. It was shown that RNA polymerase binds to both the –10 and –35 sequences and also to sequences further upstream. The σ^{70} subunit associates with the DNA, principally the transcribed strand, along a region from about the –25 to the +12 position relative to the transcription start site.[32–32b] The α subunits bind to an UP element from ~–40 to –60 via their C-terminal domains (CTDs). See Fig. 28-4B.[53–54d]



Control of stable RNA synthesis. Whereas most mRNA has a relatively short lifetime, the stable ribosomal RNAs and transfer RNAs have much longer lives. Furthermore, in *E. coli* their transcription is coordinately controlled by seven *rrnB* P1 promoters.[55] The genes for stable RNA have promoters with the usual –10 and –35 sequences, but they contain a complex upstream activating sequence that includes the UP element and three binding sites for protein FIS (**the factor for inversion stimulation**).[56] This name reflects a second function, that of promoting inversion of a DNA segment in the Hin recombinase system (Chapter 27, Section D,3). FIS is a dsDNA-bending protein.[55,57] A dimer of 11.2-kDa subunits, it is an abundant protein. Like HU, IHF, H–NS, and Dps, it coats a significant fraction of the DNA in the *E. coli* chromosome.[58] It binds to the FIS sites using a helix–turn–helix motif. In addition, there is a GC-rich region at positions –7 to –1 with the consensus sequence 5'-GCGCC_C. It has been suggested that this **discriminator** is involved in the **stringent response**, the diminished rate of stable RNA synthesis observed during amino acid starvation.[59,59a]

The very complex stringent response, which involves ribosomes in the synthesis of guanosine 5'-triphosphate-3'-diphosphate (pppGpp) under some conditions,[60] is dealt with further in Chapter 29 (Section C,8). The regulator pppGpp, whose concentration may rise from ~50 μM to ~1 mM within minutes after deprivation of amino acids, may react directly with RNA polymerase at an allosteric site to inhibit transcription.[61–62a] Another possibility is that pppGpp acts on another protein that binds to the discriminator sequence. However, the fact that a phage T7 gene, which is also under stringent control, lacks the discriminator sequence argues against this.[63] pppGpp, whose concentration is usually high under conditions of slow growth, may be a major growth-controlling effector, which acts by inhibiting the replication of the rRNA needed for

ribosome formation and protein synthesis when nutrients are scarce. However, some operons are activated by the same effector.[62]

The initiation reaction. A promoter not only locates the site of initiation but also determines the direction of transcription and, therefore, the strand of the DNA duplex that is to serve as the template. The requirement for two specific recognition sequences ensures this directionality. The RNA polymerase may bind randomly to DNA, then move rapidly along the double helix until it locates a strong binding site[64–66] where it binds to the recognition sequences of the promoter through specific interactions in the major groove of the DNA helix (see Fig. 5-3).

A satisfactory mathematical model for initiation of transcription supposes that the polymerase and DNA bind reversibly to form a complex with formation constant K_f. This initial specific polymerase–promoter complex is referred to as a **closed complex** because it is thought that the bases in the DNA chain are all still paired. It is postulated that in a rate-determining step the closed complex is converted into an **open complex,** which is ready to initiate mRNA synthesis (Eq. 28-1).[26,67] In the open complex the hydrogen bonds holding together the base pairs have been broken, and the bases of the template chain are available for pairing with incoming ribonucleotide triphosphates.

$$\text{E(polymerase)} + \text{P(promoter)} \underset{}{\overset{K_f}{\rightleftharpoons}} \text{EP}_c \text{ (closed)} \xrightarrow{k} \text{EP}_o \text{ (open)} \longrightarrow \text{Transcription} \qquad (28\text{-}1)$$

It is clear from Eq. 28-1 that the efficiency of initiation depends upon both the affinity K_f and the rate constant k for opening of the double helix. Notice that the Pribnow sequence is AT-rich; therefore, opening of the helix at this point would be easier than in a GC-rich region. Thus, the Pribnow sequence may represent a point of entry of RNA polymerase to form the open complex.[67] Other upstream A•T tracts are often present frequently at about the –43 position in the UP element. They also seem to strengthen promoter activity.[68] The open complex is thought to undergo some kind of isomerization to form an **initial transcribing**

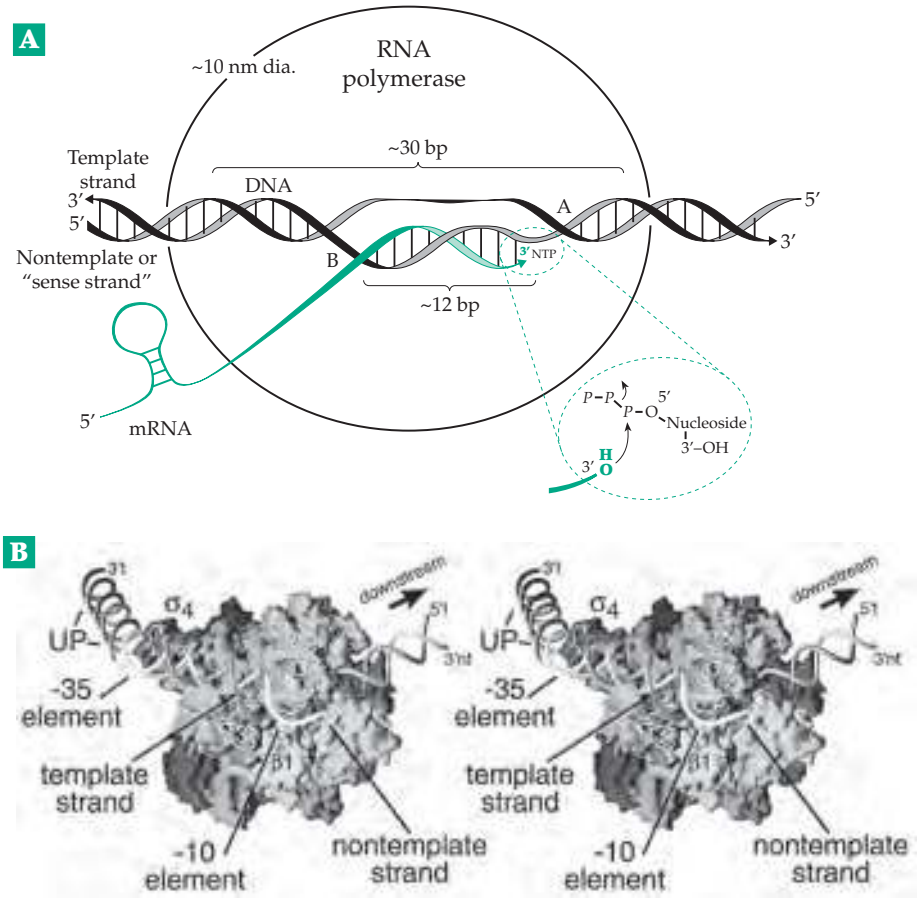


Figure 28-4 (A) Hypothetical structure of a "transcription bubble" formed by an RNA polymerase. Shown is a double-stranded length of DNA with the unwound bubble in the center. This contains a short DNA–RNA hybrid helix formed by the growing mRNA. The DNA double helix is undergoing separation at point A as is the hybrid helix at point B. NTP is the ribonucleotide triphosphate substrate. See Yager and von Hippel.[71] (B) Stereoscopic view of the structure of RNA polymerase from *Thermus aquaticus* in a complex with a promoter DNA. Included are the αI, αII, ω, β, β′, and σ subunits. However, the α C-terminal domains have been omitted. The template (t) strand passes through a tunnel, which is formed by the β and β′ subunits and two of the structural domains of the σ subunit. The nontemplate (nt) strand follows a different path. The position of the –10, –35, and UP elements of the DNA are marked. From Murakami *et al.*[33d] Courtesy of Seth A. Darst.

complex in which the first chemical steps in RNA formation occur.[69,70] These initial steps may involve rearrangements of subunit interactions and untwisting of DNA by torsional movements between subunits.[57]

Initiation of an RNA chain begins by reaction within the transcription bubble of either ATP or GTP with a second ribonucleotide triphosphate (Eq. 28-2) to form a dinucleotide still bearing a triphosphate at the 5' end. Further addition of nucleotide units at the 3' end by the same type

$$\text{ATP(GTP)} + \text{NTP} \rightarrow \text{PP}_i + \textit{ppp}\text{Pu}\textit{p}\text{N} \qquad (28\text{-}2)$$

of reaction leads to rapid transcription at a rate of ~50 nucleotides s^{-1} at 25°C. This is about one-thirtieth the rate of replication. The action of the RNA polymerase is apparently processive, a single molecule of the enzyme synthesizing the entire mRNA transcript.

3. Elongation of RNA Transcripts

After the newly initiated transcript has grown to 8–9 nucleotides, the sigma factor is lost from the RNA polymerase complex, the complex becomes very stable toward increased salt concentrations, and transcription proceeds processively in a **stably elongating mode**. As it does, the strands of the DNA duplex are pulled apart ahead of the polymerase and close up again behind the polymerase, the polymerase itself moving in a "transcription bubble" (Fig. 28-4). This is thought to lie within the ~10 nm diameter RNA polymerase complex and to encompass ~30 bp of DNA.[71] At the leading edge (point A in Fig. 28-4) a "separator" opens the DNA, which then closes behind the bubble. Behind the polymerase active site (next to substrate NTP in Fig. 28-4) the transcribed RNA forms a short hybrid helix with the DNA, presumably with an A type structure (Chapter 5). A severe topological problem is avoided if the transcribed RNA is separated at point B (Fig. 28-4) as the polymerase moves along the double helix. The driving force for the polymerization lies largely in the hydrolysis of the inorganic PP_i formed in the polymerization (Eq. 17-57). Stabilization of the RNA transcript by formation of loops and other secondary structure may also be a factor.

Although this picture seems clear and simple, many uncertainties remain. Transcription does not proceed evenly but by pauses and spurts. This has suggested the possibility of an "inchworm" type of movement of RNA polymerase.[72–75] However, the observations may also be explained by variations in the sequence. There are both pausing or stalling sites[76] and terminator sequences. The concentrations of the needed ribonucleotide triphosphate precursors will also affect the kinetics. In addition, defects in the DNA will be met. Transcriptionally linked repair (Box 27-A)[76a] may have to be called into play before transcription can continue.[77,78] Mispairing of bases can also occur in the growing RNA chain necessitating a pause during which the mismatch is recognized. Specialized proteins **GreA** and **GreB** participate in an editing step during which the RNAP backtracks for a few nucleotides, while a piece is hydrolyzed off from the 3' end before transcription can continue.[69,79,80]

If the transcription complex moves straight along a DNA double helix, separation of the strands will create positive supercoils (overwound DNA) in front of it and negative supercoils behind the bubble.[81] Experimental data support this prediction.[82] In *E. coli* the transcription of a plasmid generates positively supercoiled plasmid DNA when DNA gyrase (Chapter 27) is inhibited selectively.[83] A similar result was observed in yeast.[84] These and other data suggest that DNA gyrase may act to remove these positive supercoils, and that topoisomerase I may function in removing the negative supercoils generated behind the transcription bubble.

The electron micrograph in Fig. 28-5 shows RNA polymerase complexes apparently moving along a DNA strand with ribosomes assembled on the RNA and

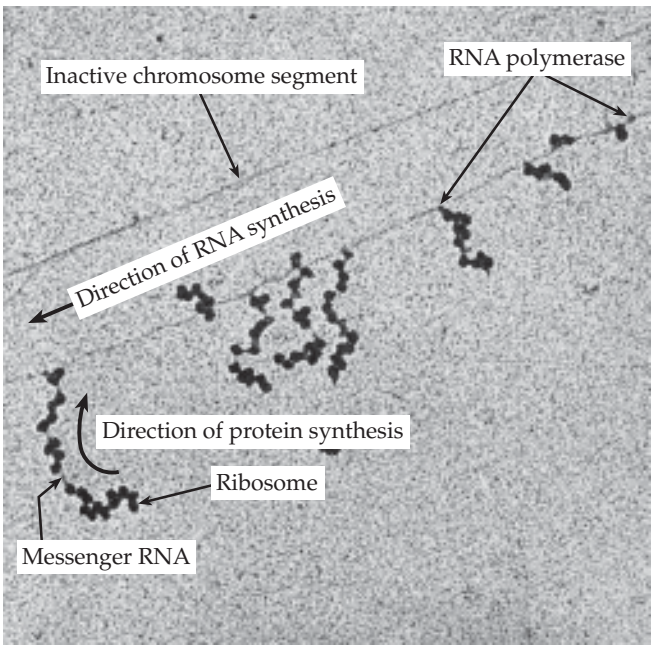


Figure 28-5 Electron micrograph showing transcription from an unidentified operon in *E. coli*. Note the DNA duplexes (horizontal) and the mRNA chains with ribosomes attached. The mRNA chains are shorter at the right side where transcription begins and larger to the left where transcription has proceeded for a longer time. From O. L. Miller, Jr.[85]

presumably synthesizing proteins. Actually, the DNA may be moving through the polymerase complexes.

4. Control of Various Bacterial Operons

With more than 2400 promoters an *E. coli* cell can be expected to utilize a great variety of control mechanisms. The best known of these involve the basic biosynthetic pathways and energy-yielding reactions. Here is a small sample.

Many repressors. Bacteria tend to keep most operons relatively inactive by use of repressors, most of which are proteins. Repressor proteins come in a variety of sizes and three-dimensional structures. Most are oligomers, and all have a DNA-binding motif, often at the N terminus but sometimes at the C terminus or elsewhere. The most studied and perhaps most abundant family[86] have the helix–turn–helix (HTH) motif. Within this family three-dimensional structures have been established for the *lac* repressors, the *trp* repressor (Fig. 5-35), the 66-residue *cro* repressor,[87–89] and a 92-residue DNA-binding "head piece" of the 236-residue lambda (cI) repressor.[90–91] The latter two proteins, both of which occur as symmetric dimers, are involved in preserving the lysogenic state of the λ prophage (Section B,1). The related 71-residue *cro* repressor from phage 434[92] and the head piece from the phage 434 repressor have similar structures.[93,94] All of these proteins have the N-terminal HTH structural motif, which binds to DNA and recognizes the appropriate sequence by interactions in the major groove of the DNA. Experiments in which the recognition helix of the 434 repressor was replaced with the corresponding helix from a phage P22 repressor supported the concept of an N-terminal HTH DNA-binding domain.[95,96] The hybrid repressor bound to the P22 operator rather than to that of phage 434. A similar HTH domain is present near the C terminus in the catabolite-activator protein (Fig. 28-6). The fit of the helix–turn–helix recognition corner varies from one protein to another. The small 53-residue **Arc** repressor of bacteriophage P22 of *Salmonella* has a very different interaction with DNA: the β sheet of an arc dimer fits into the major groove of the operator DNA making specific contacts.[97,98] A similar interaction characterizes the dimeric methionine repressor.[99]

Basic to the functions both of repressors and of activator proteins are allosterically induced conformational changes caused by the binding of inducers or corepressors. The changes begin at the binding sites of these small effector molecules but are transmitted to the DNA-binding heads. In several cases the conformational changes have been observed by X-ray crystallography and are seen to involve a movement of the recognition helices. This is also true for the *lacI* repressor (Fig. 28-3) and for the *TrpR* repressor (Fig. 5-35B) Notice that the tryptophan binds to the aporepressor immediately adjacent to the DNA-binding site, where it may control not only the shape but perhaps also the charge distribution within the recognition motif.

Another important factor in determining the strength of a repressor–operator interaction is the twist of the DNA or any other distortion of its regular helical structure. For example, in the center of the *cro* 434-operator complex the DNA is wound, while at the ends it is underwound.[92] Conformational changes in either the repressor or in the DNA or in both may be needed to provide optimal binding. As is seen in Fig. 28-3 the *lac* repressor causes a distinct bend in the DNA.

A theoretical possibility would be for the DNA in an operator site to be extruded as a cruciform structure like that in Fig. 5-34A. Such structures bind well to certain oligomeric proteins (e.g., see Fig. 27-26B), and they do appear to form in some promoters.[100] However, crystallographic structure determinations have ruled out cruciform structures for many repressors. A change from a linear helical duplex to cruciform structure would require a substantial unwinding of the helix and would mean that negatively superhelical DNA molecules would bind repressor much more tightly than does DNA without superhelical turns. Negative supercoiling does facilitate the *lac* operator–repressor interaction. However, there is only a 50–90° unwinding of the DNA.[101] Some AT-rich palindromes are readily converted to cruciform structures when negative supercoiling is increased. Placement of a promoter sequence within such a structure represses transcription unless the supercoiling is relaxed.[102]

Binding of repressors and activator proteins to DNA control sequences is being studied in many ways. Among them are ^{1}H and ^{19}F NMR measurements on *lac* and *cro* repressors with specific tyrosine side chains replaced by deuterated tyrosine or 3-fluorotyrosine[103] or with 5-F-uracil replacing specific thymines in an operator sequence.[104] Footprinting shows that λ and *cro* repressors bind to only one side of the double helix, as is depicted in Fig. 5-50. Addition of dimethyl sulfate to growing *E. coli* causes methylation at many sites in the DNA and has been used to obtain footprints *in vivo*.[105] Not only are certain guanosines in a promoter protected from methylation when active transcription is occurring but also guanosines on the opposite side of the double helix become unusually reactive. This suggested that the DNA helix is bent in the transcriptional initiation complex and that the resulting distortion makes the bases on the outside of the bend more reactive.[105] There is also evidence that the repressor binds to the opposite face of the DNA that binds RNA polymerase in the –35 region of the promoter.[106]

Repressors may have similar recognition domains but may vary greatly both in size and in the functioning of their other domains, which may react both with small allosteric effectors and with other proteins. The repressor **BirA** of the *E. coli* biotin synthesis operon is an enzyme. The 321-residue protein activates biotin to form biotinyl 5'-adenylate and transfers the biotinyl group to proteins such as acetyl-CoA carboxylase[107–109a] and also represses transcription.

Inducible operons. The operon model as presented in Fig. 28-1 describes the negatively controlled inducible *lac* operon. There are many other examples of this type of control. Among them are control of the utilization of *N*-acetylglucosamine and xylose.[110] Ten genes for catabolism of inositol are encoded in two negatively controlled operons in *Bacillus subtilis*.[111] Similar controls are used by bacteria to protect against the antibiotic tetracycline. The **TetR** repressor controls a membrane associated protein TetA, which acts as an energy-dependent pump to remove tetracycline from the bacterial cell. Synthesis of TetA is normally repressed, but tetracycline, if present, acts as an inducer to cause TetR to dissociate from its operator site, allowing transcription of the transporter gene.[112–113a] The gene order, gene content, and regulatory mechanisms in an operon are often poorly conserved among related species of bacteria. However, functional and regulatory relationships may be maintained.[113b]

Feedback repression. A simple modification of the operon model accounts for **feedback repression** by end products of biosynthetic sequences (Fig. 11-1). The product, e.g., an amino acid, vitamin, purine, or pyrimidine base, acts as a corepressor binding to the aporepressor and causing an allosteric modification that inhibits transcription. It is not the aporepressor but the effector–repressor complex that binds to the operon and blocks transcription. An example is the tryptophan–repressor complex, whose 3-dimensional structure is shown in Fig. 5-35. If free tryptophan accumulates within a bacterial cell, it binds to the allosteric site in the *TrpR* repressor, inducing a conformational change that permits the tryptophan–repressor complex to bind to at least three 21-bp operator sites. Binding of the repressor–corepressor complex at the *aroH* operator locus represses the genes for the first three steps in the aromatic biosynthetic pathway (Fig. 25-1, step *a*). Binding at the *trpEDBCA* operator represses the genes for conversion of chorismate to tryptophan (Fig. 25-2), and binding at the *trpR* locus lowers the rate of synthesis of the trp repressor, providing a counteracting effect that may be important in controlling the growth rate. The *trp* operon is also well known for another type of control called attenuation (see Fig. 28-9 and associated text).

Feedback repression controls methionine synthesis in *E. coli*. The corepressor *S*-adenosylmethionine binds to an aporepressor that recognizes the following palindromic sequence in the DNA. This sequence occurs,

with minor variation, in front of four different operons encoding enzymes of methionine biosynthesis.[99,114] Together they form the methionine **regulon**. Other *E. coli* operons that are negatively regulated by operator–repressor interactions include those involved in biosynthesis of phenylalanine, tyrosine, arginine, threonine, and isoleucine. The *tyr* repressor modulates gene expression in at least eight operons, largely by repression.[115] The *ile* repressor acts on both the *ile* and *thr* operons.[116] The purine repressor, 341-residue PurR, belongs to the LacI family of repressors. Its C-terminal domain also has some sequence similarity to that of periplasmic sugar-binding proteins.[117] PurR represses several steps in the biosynthesis of IMP and in its conversion to GMP and AMP (Fig. 25-15).[118,119] In *E. coli* twelve biosynthetic genes, organized as nine transcriptional units, provide for the synthesis of arginine in eight enzymatic steps (Fig. 24-10).[120,121] As with the *trp* operons, synthesis of the repressor ArgR is autoregulated. The hexameric repressor has an N-terminal winged HTH DNA-binding domain.[121–123]

As is pointed out in Chapter 16, the acquisition of iron and control of its concentration is of crucial importance to bacteria. In *E. coli* the Fe^{2+}-binding protein **Fur** (ferric uptake regulator) represses promoters controlling siderophore biosynthesis as well as other responses. It is a **global regulator** that controls ~40 transcriptional units.[124] Similar proteins repress synthesis of the diphtheria toxin by *Corynebacterium diphtheriae*,[125,126] uptake of iron in these bacteria and in *Mycobacterium tuberculosis*,[127] and uptake of molybdate.[127a]

Positive control by activator proteins. Cyclic AMP in bacterial cells mediates the phenomenon of **catabolite repression**. This is the inhibition of the transcription of genes for enzymes needed in catabolism of lactose and other energy-yielding substrates, when the more efficient energy source glucose is present. Glucose causes a decrease in the concentration of cAMP by a complex mechanism[59,128] that may also cause a decrease in the concentration of inducer. When the glucose concentration decreases, the concentration of cAMP rises and stimulates the initiation of transcription in many operons. This is accomplished through the mediation of the 209-residue catabolite gene activator protein, CAP (also known as cyclic AMP receptor protein, CRP). This protein is a "global"

regulator of gene expression that activates transcription at over 100 promoters in *E. coli*.[129] The CAP–cAMP complex binds to the *lac* promoter adjacent to the RNA polymerase site at the palindromic sequence in the DNA as is indicated in Fig. 28-2. The CAP molecule is a 45-kDa dimer of identical subunits, which resembles the repressors in having the HTH reading head that binds to the DNA.[130–132] However, the HTH motif is at the C terminus. The cAMP binds to two sites in each monomer (Fig. 28-6). Tightly bound cAMP molecules in an *anti* conformation bind in the center of the large N-terminal domains in the major allosteric sites. At higher cAMP concentration the second sites in the C-terminal DNA-binding domains are occupied.[131]

The CAP binds to DNA with the consensus sequence 5′-AAATGTGATCT/5′-AGATCACATTT, which may be located at variable distances from the promoter.[133] How does binding of the CAP–cAMP complex increase the rate of initiation of mRNA transcription? The anwer evidently lies in direct interaction between CAP and the N-terminal domain of the RNAP α subunit.[54d,129] Binding of CAP induces a 90° bend in the DNA, which may facilitate the protein–protein interaction and may lead to looping.[130,134]

The galactose (*gal*) operon of *E. coli* is negatively controlled and inducible by D-galactose or D-fucose, which bind to the *gal* repressor. There are two overlapping promoter sites, one of which is stimulated by adjacent binding of a CAP molecule.[135,136] A surprise came from the discovery that the operator was *upstream* from the promoter, that is, it comes before both the promoter and the structural genes to be transcribed. Later, a second operator sequence was found 90 bp away from the first and within the first structural gene. This suggested that the dimeric gal repressor binds the two operators to form a loop that blocks transcription.

Another example of positive regulation by CAP is provided by the seven proteins required for uptake of maltose and its catabolism by *E. coli*. These are encoded in two operons that are controlled as a single regulon. An apo-activator protein becomes an activator when it binds maltose.[137,138]

An *E. coli* protein known as **FRN** (for fumarate nitrate reduction) is a global transcription regulator homologous to CAP. It is active only under anaerobic conditions in which it controls more than 100 genes.[139–141] FRN contains an $[Fe_4S_4]^{2+}$ cluster, which is required for dimerization and binding to DNA. Exposure to O_2 converts the cluster into an $[Fe_2S_2]^{2+}$ cluster with loss of activator activity. The photosynthetic *Rhodospirillum rubrum* is able to adapt to growth on carbon monoxide as a carbon source. A CAP-like transcriptional activator **CooA** contains heme. It acts as a sensor for CO, which activates transcription, as does cyclic AMP with CAP.[142,142a] Sequence homologies suggest that several other bacterial activator proteins also have the HTH DNA-binding motif near their C termini.

There are other types of transcriptional activators in bacteria. One is transcription factor 1 (TF1) encoded by a *Bacillus subtilis* phage. It is a member of the protein HU family (Chapter 27). However, unlike the nonspecific HU it binds to some sites specifically and activates transcription.[143] The *E. coli* Ada protein is the acceptor protein in removal of methyl groups from DNA (Chapter 27). The same protein is an inducer of transcription of DNA repair enzymes in the large ada regulon. Methylation of Cys 69 of the Ada protein itself converts it into a gene activator.[144]

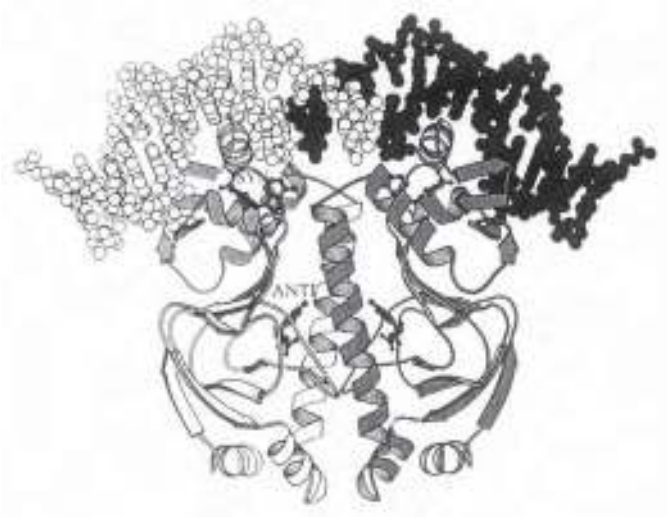


Figure 28-6 MolScript ribbon drawing of the CAP dimer bound to DNA with two molecules of the coactivator cAMP bound per monomer. A *syn*-cAMP molecule is bound to the HTH domain and a loop from the N-terminal domain, while the second *anti*-cAMP is bound more tightly in the center of the larger N-terminal domain. The DNA sequence for each half site is 5'-ATGTCACATTAATTGCGTTGCGC-3'. From Passner and Steitz.[131] Courtesy of Thomas A. Steitz.

Control by looping. The arabinose utilization operon of *E. coli*, *araBAD*, encodes proteins needed for uptake of arabinose and conversion to D-xylulose 5-*P*. The repressor AraC in the absence of arabinose binds at operator 1 (O_1) to prevent further synthesis of repressor (autorepression) and also at the *araI* region to repress transcription of operon *araBAD*. The operator 2 (O_2) site, which is 211 bp upstream from *ara*I, is also needed for this repression.[145–147b] A loop is apparently formed by repressor binding (Fig. 28-7). Binding of arabinose to the repressor converts it into an activator, which stimulates initiation of transcription at the *BAD* promoter. Further stimulation is provided by the CAP-cAMP complex, which binds next at *araI*.

Looping is a recognized control mechanism for a number of other operons as well.[148]

Bacterial enhancers. Positive regulatory DNA sequences that are distant from the genes controlled are often called enhancers.[149,150] Their function is usually independent of position over a range of hundreds or more base pairs either upstream or downstream of the transcription initiation site. Quite common in eukaryotes (Section C,4), enhancers are less often found in bacteria. However, the binding sites for the **nitrogen regulatory protein C** (**NtrC** or Nr_I) of *E. coli* has the characteristics of an enhancer.[151,152] It functions with the rather complex glutamine synthetase (*glnALG*) operon in a major control point for nitrogen metabolism. The enzymology is illustrated in Fig. 24-7. When the supply of nitrogen from NH_3 is low the NtrC protein, a product of gene *glnG*, binds to the enhancer, which is located over 100 bp upstream and is thought to contact the σ^N (σ^{54}) subunit of the RNAP by formation of a loop.[153,153a] The process has been visualized by scanning force microscopy.[152] Another interesting aspect of this control system is the activation of Ntrc by phosphorylation of a specific aspartate side chain (Asp 54). The NtrC-*P* form is the active enhancer-binding protein. NtrC is a member of the family of two-component sensor-response regulator pairs, which frequently control bacterial metabolism and behavior (Chapter 11, Section C, 2; Fig. 19-5). The sensor protein is **NtrB** (NR_{II}), which is an autophosphorylating histidine kinase similar to the CheY protein of bacterial chemotaxis (Fig. 19-5).

NtrC-*P* dimerizes and binds to the enhancer sequence, where it appears to catalyze an ATP-dependent isomerization of the closed to open forms of the transcription initiation complex (Eq. 28-1).[153,154] The isomerization may depend upon looping.[152] Other operons that utilize the σ^N subunit of RNAP often also have upstream or downstream enhancers.[155,156]

Other two-component control systems. More than 60 different sensor–response regulator pairs have been discovered in bacteria. Many are associated with nutrition. For example, more than 30 genes of the phosphate (*pho*) regular are controlled by the sensor kinase **phoR**, which detects low phosphate ion concentrations and phosphorylates the response regulator **phoB**.[157–159] Protein **ArB** senses changes in environmental O_2 levels, and response regulator **ArcA** regulates ~30 operons in response.[159a] One effect is to activate ~30 genes needed for the conjugative transfer of DNA (Fig. 26-3).[160] Transcription of rhizobial *nif* and *fix* genes (Fig. 24-4) is controlled by the O_2 sensor **FixL** and its response regulator **FixJ**.[161] FixL is a heme protein whose kinase is active only when the heme is deoxygenated. The *E. coli* proteins discussed on p. 1075 mediate transcriptional responses to accumulation of superoxide anions or hydrogen peroxide.[161a]

In the absence of O_2 the *E. coli* FNR protein induces proteins of the anaerobic respiration pathways. Nitrate also has its own two-component system that senses nitrate availability and activates transcription of enzymes catalyzing nitrate respiration.[162] An expanded two-component system induces sporulation in *Bacillus subtilis* in response to poor growth conditions.[163,164] The crystal structure of one of two response regulators (**Spo0F**) has a structure closely related to that of CheY and the nitrate response regulator NarL.

Antisense RNA. Another mechanism of control of either transcription or of plasmid replication involves small molecules of RNA that are transcribed from the opposite strand than the template strand used for mRNA synthesis.[1,1a,165–166b] These **antisense** RNA molecules have at least some part of their sequence complementary to that of the mRNA and to the corresponding sequence in DNA. A well-studied example is control of the copy number of the colicin E1 and other plasmids of *E. coli*.[167–169] Two transcripts, RNAI and RNAII, are initiated upstream from the origin of replication (Fig. 28-8). RNA II is a 555-nucleotide primer of replication. It is synthesized as a longer transcript that is cut by RNase H at *ori*. This

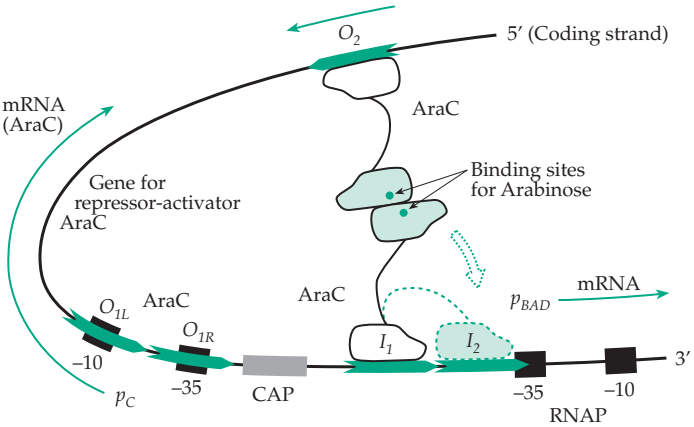


Figure 28-7 Scheme showing regulatory region of the *araCBAD* operon of *E. coli*. In the absence of arabinose the protein AraC acts as a repressor, which binds to upstream region I of the promoter and also to operator O_2, forming a loop of ~210 bp of DNA. Binding of arabinose to the N-terminal domains of the dimeric repressor allows the dimer to dissociate and the N-terminal domain of one subunit to bind to region I_2 of the promoter for genes *B,A,D* (dashed outline) activating their transcription. The AraC protein also binds to promoter P_c, repressing its own synthesis. After Zhang *et al*.[146]

cleavage is inhibited if the 108 nucleotide RNA I forms a duplex with the 5' end of the RNA II, which has a complementary base sequence. The process is more complex than this because both RNA II and RNA I have complex secondary structures and are brought together with the help of the small protein product of gene *gro*, which permits them to recognize each other prior to duplex formation.

The major outer membrane porins of *E. coli* (Fig. 8-20) are encoded by genes *ompC* and *ompF*. A small 174-nucleotide RNA called **mRNA-interfering complementary RNA** (mic RNA), whose gene is upstream of the *ompC* promoter, is transcribed in the opposite direction from that of the porin gene.[167] It is homologous with the 5' end of *ompF* mRNA, and its function is evidently to inhibit translation of the *ompF* mRNA. Since *ompC* and the *micF*RNA gene are apparently regulated coordinately, synthesis of the *ompF* product is inhibited if the *omp*C product is being synthesized, as happens when the bacteria are growing in a medium of high osmotic strength.[166b,170,171]

Many other examples of regulation by antisense RNA are being discovered. Small noncoding RNA molecules (**ncRNAs**) often serve as templates or **guides**[1,171a] in processes ranging from synthesis of telomere ends (Fig. 27-21) to editing,[172] modifying,[173] and splicing mRNA[174] (Section D). Both in bacteria[174a] and in eukaryotes (p. 1640) dsRNAs are often formed and subsequently cleaved to single-stranded antisense RNAs that act as guides to initiate hydrolytic destruction of defective, toxic, or unwanted mRNAs.[171a] The recognition of these natural regulatory mechanisms has led to keen interest in artifical regulation by antisense RNA. Synthetic antisense RNA injected into cells will inhibit expression of selected genes. There is the possibility of effective therapy against viruses, cancer, and inflammation if suitable antisense RNA could be generated within eukaryotic cells or introduced as drugs.[175–176b] Such drugs, which are typically ~15 nucleotides in length, are most satisfactory if they are stable, enter cells, and interact specifically with complementary sequences of DNA or RNA.[177] Stability can be improved by use of linkages other than the natural phosphodiester.[178–181] Phosphorothioates, in which sulfur replaces one nonbridging oxygen atom on phosphorus, is a favorite. Synthetic antisense oligonucleotide mimics may cause adverse reactions with proteins within cells.[178,182] Nevertheless, future successes seem likely.[182a] See also RNA interference (p. 1640).

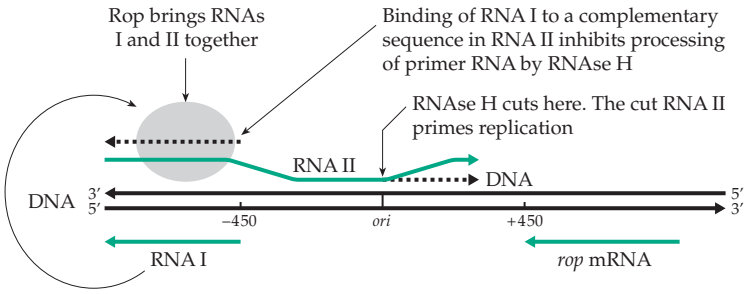


Figure 28-8 Simplified scheme for control of replication of the ColE1 type plasmid by antisense RNA. The primer for DNA synthesis is RNA II whose initial transcript extends past the replication *ori*. It is cut by RNase H at *ori* and then primes replication of the upper strand as shown in the figure. The antisense RNA is RNA I. It binds to protein Rop whose gene location is also indicated in the figure. Rop assists RNA I and RNA II in undergoing a complementary interaction. However, both RNAs apparently maintain a folded tertiary structure, and only some segments interact. The interaction with the Rop protein evidently in some way prevents initiation of replication until the Rop concentration falls because of replication of the host cell.[167,168]

5. Termination of Transcription in Bacteria

Encoded in DNA are not only the initiation signals for transcription but also termination signals or **terminators**.[183] Some of these are sensed by the bacterial RNA polymerase itself while the "reading" of others requires auxiliary proteins. Terminators can be either constitutive or regulatable. The simplest terminators result from GC-rich regions of dyad symmetry in the DNA. The RNA transcript is able to form a stable hairpin loop, possibly within the transcription bubble. If such a loop is followed closely by a series of uracils, the RNA and the polymerase will dissociate from the DNA template terminating transcription. The low stability of AU base pairs may facilitate dissociation, but RNA polymerase may also recognize the terminator loop. Sometimes a terminator will have a series of adenines preceding the loop. This is often a bidirectional terminator; the transcript from the other strand of DNA will have a loop followed by a series of U's. There are many more complex termination mechanisms.

Attenuation. A major mechanism of feedback repression, known as attenuation, depends not upon a repressor protein but upon control of premature termination. It was first worked out in detail by Yanofsky *et al*. for the *trp* operon of *E. coli* and related bacteria.[184–186] Accumulation of tryptophan in the cell represses the *trp* biosynthetic operon by the action of accumulating tryptophanyl-tRNA^{Trp}, which specifically induces termination in the *trp* operon. Other specific "charged" aminoacyl-tRNA molecules induce termination at other amino acid synthesis operons. The first structural gene in the *trp* operon,

*trp*E, is preceded at the 5' end by a 162-bp **leader sequence,** which is transcribed into mRNA. Within this RNA are two adjacent hairpin loops. The second of these loops (labeled **3:4** in Fig. 28-9) has a GC-rich stem and is followed by eight consecutive U's. It is a typical efficient terminator. An RNA polymerase, having just passed this sequence, will interact with the looped RNA formed behind it and will dissociate from the DNA to terminate transcription. However, if the terminator loop is prevented from forming, transcription will continue, and the structural genes of the operon will be expressed.

The *trp* operon contains a short gene for a **leader peptide** preceding the terminator. Its RNA transcript is shaded, and its initiation codon AUG and termination codon AAU are boxed in Fig. 28-9. In bacteria translation of mRNA begins while transcription is still in progress (Fig. 28-5). The 5' end of the mRNA may enter a ribosome before the RNA polymerase reaches the terminator loop, after which the leader peptide will be synthesized. The ninth and tenth codons of the leader peptide gene are of special importance. They lie at the beginning of the **1:2** or "protector" loop (Fig. 28-9) and code for tryptophan. If the level of tryptophan is high, tryptophanyl-tRNA^{Trp} will be formed and the mRNA will move rapidly through the ribosome, these tryptophan codons will pass through, and tryptophan will be incorporated into the peptide. The 1:2 loop will be opened, but the terminator loop will remain intact. The result will be termination of transcription. However, if the tryptophan concentration is low, there will be a shortage of charged tRNA^{Trp}, and peptide synthesis will be "stalled" with these tryptophan codons in the active sites of the ribosome. This will allow time for the attentuator region to assume the alternative secondary structure shown in Fig. 28-9B. Here the **1** limb of the protector loop is stalled in the ribosome allowing the **2** limb to form the altenative **2:3** loop. The terminator has been destroyed, and transcription continues through the rest of the operon. Thus a low level of tryptophan (and of tryptophanyl-tRNA) favors transcription of the *trp* synthetic operon.

Cells of *Bacillus subtilis* also synthesize a *trp* operon transcript that can form either an antiterminator or a terminator loop (Fig. 28-9C).[187,187a–c] Tryptophan, when present in a high enough concentration, binds to a *trp* RNA-binding attenuation protein (*TRAP*). This is an 11-subunit protein, which has 11 tryptophan-binding pockets and also 11 binding sites for GAG or UAG RNA triplets. When tryptophan accumulates within the cell it binds to TRAP, which then wraps ~53 residues of RNA transcript containing 11 GAG or UAG triplets around its perimeter (Fig. 28-9D). This prevents formation of the antiterminator loop but allows the terminator loop to form. At low tryptophan concentrations the antiterminator loop is formed, preventing formation of the terminator loop.[188]

Attenuation is also an important mechanism of control of transcription of biosynthetic operons for histidine, phenylalanine, leucine, isoleucine, and threonine.[189] Like the *trp* attenuator region, attenuators for these operons contain codons for the amino acid whose synthesis is being regulated: seven Phe codons in the *phe* attenuator, seven His codons in the *his* attenuator, four Leu codons in the *leu* attenuator. The *thr* operon, which is sensitive to both threonine and isoleucine, has eight Thr and four Ile codons, while the *ilv* attenuator has four Leu, five Ile, and six Val codons permitting feedback repressor by three kinds of charged tRNA.[190] The pyrimidine synthesis operon *pyr* has three attenuator sequences, one right after the promoter and two others, one just before each of the two genes in the operon. This permits partially independent control of the two genes.[191]

Rho and other termination factors. Termination proteins can also react with specific regions of DNA or of an RNA transcript to terminate transcription.[183] The best known termination factor is the rho protein; a hexamer of 45-kDa subunits. It interacts with transcripts at specific termination sequences, which are often C-rich, and in a process accompanied by hydrolysis of ATP causes release of both RNA and the polymerase from the DNA.[192,193] Additional *E. coli* proteins, products of genes ***nus A*** and *nus G*, cooperate with the rho factor at some termination sequences.[194–196c] The rho hexamer is a helicase that moves along the RNA transcript in the 5' → 3' direction driven by ATP hydrolysis. If it locates an appropriate termination signal, it may utilize its helicase activity to uncoil the DNA–RNA hybrid segment within the transcription bubble (Fig. 28-4).[197–198b]

Cells also contain **antitermination proteins,** which prevent termination of transcription of rRNA or tRNA genes at the many loops of secondary structure that are possible with these transcripts.[59,199–200b] These antitermination factors are also important in regulating transcription during the lytic phase of growth of phage λ (see Section B,1). Also important are rates of hydrolytic degradation of mRNA molecules.[201]

6. Effects of Antibiotics

The antibiotic **rifamycin** (Box 28-A) appears to interfere with initiation by competing for the binding of the initial purine nucleoside 5'-triphosphate. The same bacterial RNAP that synthesizes mRNA also transcribes both rRNA and the tRNAs. Thus, the synthesis of all forms of RNA is inhibited by rifamycin. When a population of bacteria is subjected to this antibiotic, a few individuals survive. These rifamycin-resistant mutants are no longer sensitive to the antibiotic. Among them are some mutants that produce an

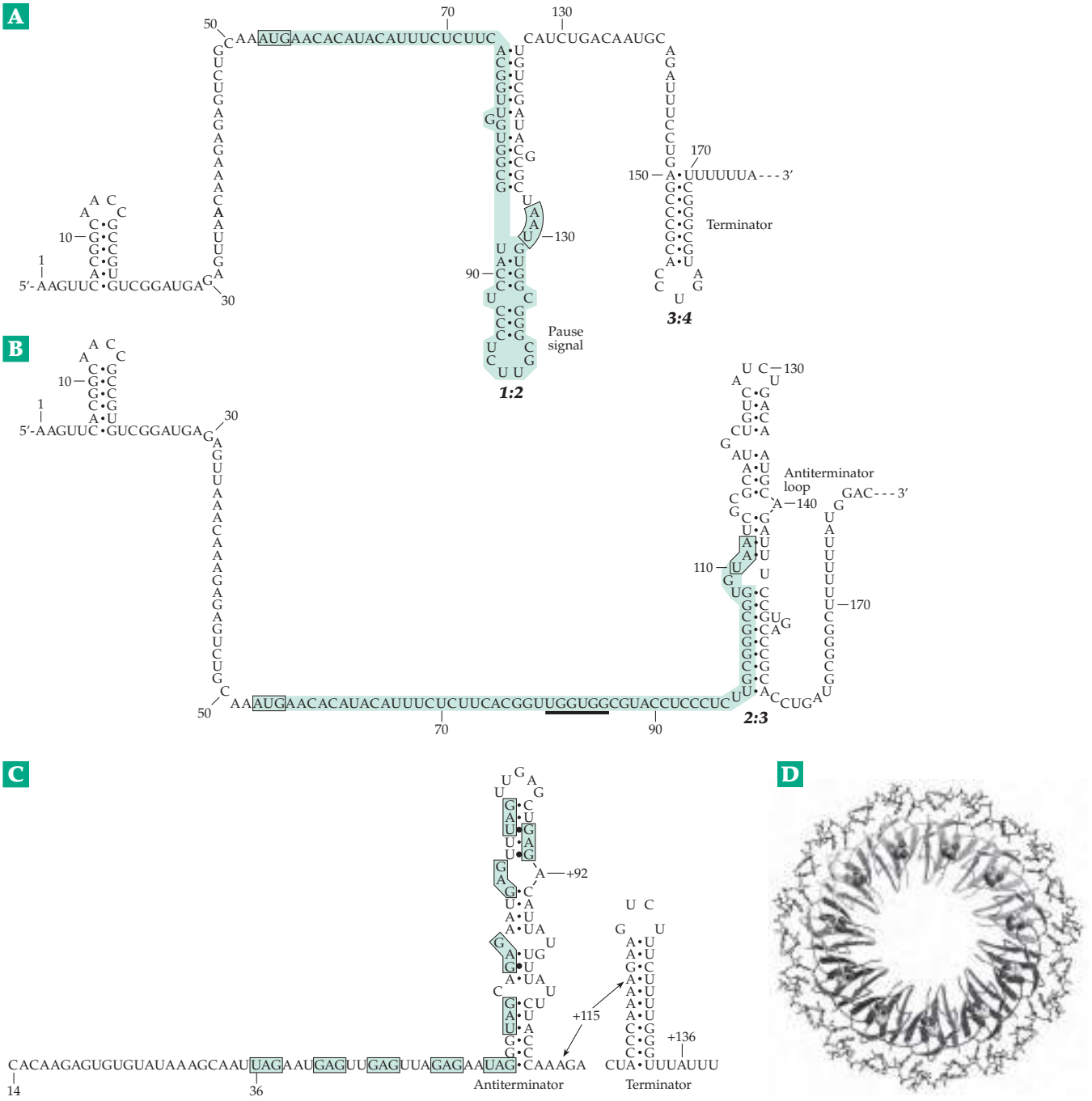


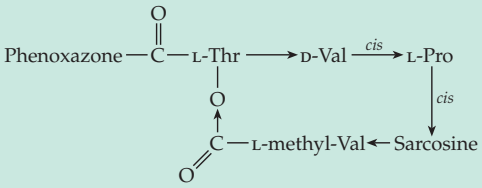
Figure 28-9 (A,B) Alternative leader RNA structures that mediate control of attenuation in the *trp* operon of *Serratia marcescens*. The leader peptide initiation and termination codons are *boxed*. Tryptophan codons are indicated by *underlining* in *B*. Formation of the 5' hairpin structure is predicted by computer analysis but is not implicated in control by attenuation. A structure **1:2** is thought to serve as a transcription pause signal. Structure **3:4** is predicted to form when there is an adequate supply of charged $tRNA^{Trp}$ and is thought to function as the transcription termination signal recognized by RNA polymerase. The **2:3** structure, or **antiterminator loop,** is predicted to form when charged $tRNA^{Trp}$ is unavailable. Its formation presumably precludes formation of structure **3:4**, thereby allowing RNA polymerase to continue transcription into the structural genes of the operon. From Kuroda and Yanofsky.[186] (C) Antiterminator and terminator loops, one of which may form from the leader sequence of the *Bacillus subtilis trp* operon mRNA. Numbering refers to the start point of transcription. The triplet repeats involved in attenuation are shaded. From Baumann *et al.*[187] (D) Structure of the 11-subunit tryptophan RNA-binding attenuation protein (TRAP) as a ribbon diagram with 11 molecules of L-tryptophan shown as van der Waals spheres. The apparently circular RNA structure reflects the fact that the gap between the beginning and end of the bound RNA segment is averaged (randomized) over eleven binding sites in the crystal structure. The 53-residue RNA containing 11 triplet repeats of GAG or UAG is bound around the perimeter and is shown as a ball-and stick model. From Antson *et al.*[188]

BOX 28-A THE ANTIBIOTICS RIFAMYCIN, RIFAMPICIN, AND ACTINOMYCIN D

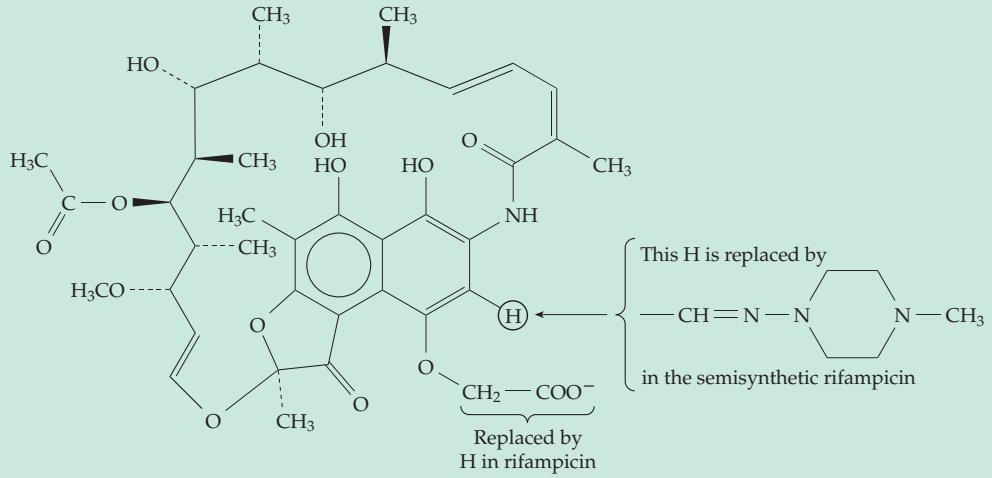
Rifamycin, produced by *Streptomyces mediterranei*, is of medical value because it affects **acid-fast** as well as gram-positive bacteria. The semisynthetic rifampicin has been especially useful in the treatment of tuberculosis. The ether linkage at the bottom of the ring at the right in the structural formula is cleaved, and the resulting hydroquinone is oxidized to a quinone within the bacteria.[a] At a concentration of 2×10^{-8} M rifampicin inhibits bacterial RNA polymerase 50%. It does not prevent the binding of polymerase to DNA but inhibits initiation of transcription. Mutants of *E. coli* resistant to rifampicin produce RNA polymerase whose β subunit has been altered, sometimes with a change in electrophoretic mobility. The related antibiotic **streptolydigin** also binds to the subunit of RNA polymerase and blocks elongation, resistant mutants mapping close to *rif* mutants.

The actinomycins, which are also produced by *Streptomyces*, not only kill bacteria but also are among the most potent antitumor agents known.[b] However, because of their extreme toxicity they are of little medicinal value. Actinomycin D, which is a specific inhibitor of RNA polymerase, contains a planar phenoxazone chromophore bearing two carboxyl groups, each one linked to an identical cyclic peptide made up of L-threonine, D-valine, L-proline, sarcosine (*N*-methylglycine), and L-methylvaline.

An ester linkage joins the methylvaline residue of the peptide to the side-chain hydroxyl of threonine. Two *cis* peptide linkages are present. Ignoring the obvious asymmetry of the phenoxazine ring, actinomycin possesses approximate twofold symmetry.



The antibiotic binds tightly to double-stranded DNA in regions containing guanine. A 2:1 deoxyguanosine–actinomycin complex has been crystallized, and the structure has been determined by X-ray diffraction.[c–e] The phenoxazine ring lies at the center of the complex, one peptide loop extending above it. The twofold symmetry is present in the dideoxyguanosine complex as well as in actinomycin itself. The phenoxazine ring lies between the two flat guanosine rings in van der Waals contact. The two amino groups of the guanine rings form strong hydrogen bonds with the carbonyl groups of the threonine residues. There are also weaker nonlinear hydrogen bonds from the N-3 atoms of the guanines to the NH groups of the same threonines.



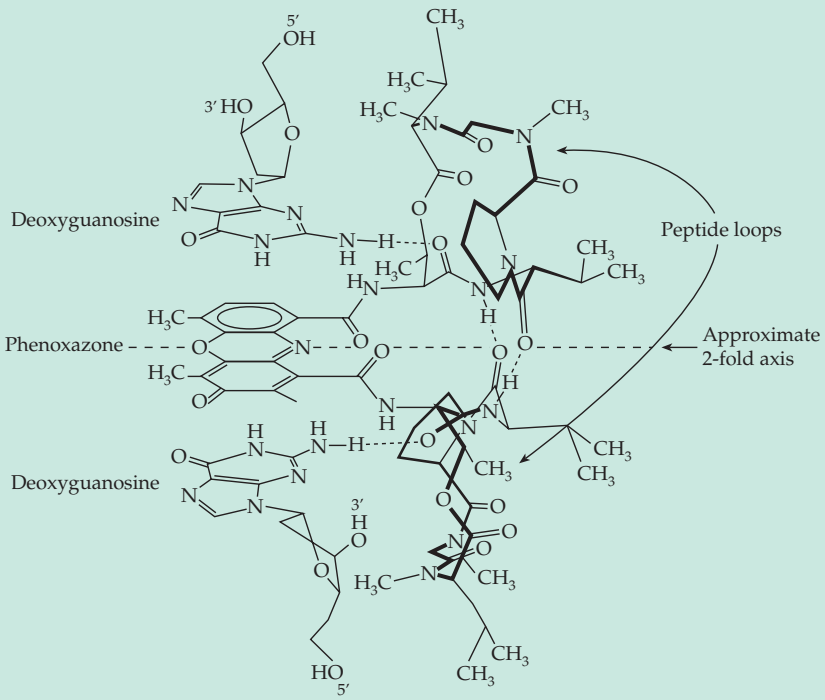
RNAP with an altered β subunit. Since the mutant polymerases do not bind rifamycin, it was concluded that rifamycin binds to the β subunit and that the rifamycin-resistance gene *rpo*B or *rif* (which maps at 89 min) is the gene for the β subunits of RNA polymerase.

Streptolydigin inhibits both initiation and elongation. **Actinomycin D**, by binding to DNA, inhibits DNA polymerases as well as RNA polymerases, the latter at a concentration of only 10^{-6} M (Box 28-A). The eukaryotic RNA polymerases are not inhibited by rifamycin, but RNA polymerases II and III are completely inhibited by the mushroom poison **α-amanitin** (see Box 28-B). Inhibitors of DNA gyrase (Chapter 27) also interfere with transcription as do chain terminators such as cordycepin (3'-deoxyadenosine) and related nucleosides.

BOX 28-A (continued)

A symmetric pair of hydrogen bonds join the two carbonyl and NH groups of the D-valine residues in the peptide loops.

Model building studies show that a similar complex can be formed with double-stranded DNA.[f] While the amino groups of the guanine rings (see drawing) are hydrogen-bonded to the actinomycin, the other hydrogen atoms of the same amino groups as well as the N-1 hydrogen atoms and the carbonyl groups of the guanine ring are available for hydrogen bonding to form a GC base pair. Thus, the structure above can be modified readily into part of the double-stranded DNA molecule in which the phenoxazone ring of actinomycin is intercalated between two CG pairs. To do this the normal DNA structure has to be unwound by 18° at the point of insertion of the extra ring. Binding also occurs at other sites.[f] Sobell suggested that actinomycin binds to a premelted region of the DNA helix within the transcription bubble and immobilizes it. This interferes with the elongation of growing RNA chains.[g]



[a] Goldberg, I. H., and Friedman, P. A. (1971) *Ann. Rev. Biochem.* **40**, 775–810
[b] Perlman, D. (1970) in *Medicinal Chemistry*, 3rd ed. (Burger, A., ed), pp. 309–316, Wiley (Interscience), New York (Part 1)
[c] Sobell, H. M. (1973) *Prog. Nucleic Acid Res. Mol. Biol.* **13**, 153–190
[d] Sobell, H. M., Jain, S. C., Sakore, T. D., and Ponticello, G. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 263–270
[e] Sobell, H. M. (1974) *Sci. Am.* **231**(Aug), 82–91
[f] Robinson, H., Gao, Y.-G., Yang, X.-l, Sanishvili, R., Joachimiak, A., and Wang, A. H.-J. (2001) *Biochemistry* **40**, 5587–5592
[g] Sobell, H. M. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5328–5331

7. Processing of Bacterial RNA

Newly formed rRNA and tRNA molecules are usually not functional but must undergo chain cleavage, methylation, and other alterations before they are "mature." However, most bacterial mRNA does not require processing. Bacteria produce a series of mRNA molecules of variable length, some corresponding to polycistronic (polygenic) and some to monocistronic operons. Most of the mRNA molecules produced are unstable with an average lifetime of about two min; however, some, such as those produced in bacteria about to undergo sporulation, survive much longer. Bacterial mRNA sometimes does undergo processing before it reaches the ribosomes. For example, following infection of *E. coli* cells by phage T7, ribonuclease RNase III cleaves the large 7000 nucleotide "early RNA" transcript from the virus DNA into five defined fragments.[202] Each fragment presumably carries the message for a single viral gene. Genes containing introns have been identified in several archaebacteria[203–205] and in certain phage.[206] The corresponding transcripts must be spliced, as are most eukaryotic transcripts.

Ribosomal RNA. Quantitatively the most im portant RNA, making up 90% of that present in cells, is ribosomal RNA. Synthesis of rRNA must be rapid, for an *E. coli* cell produces 5–10 new ribosomes per second, or 2 x 10^4 molecules of RNA per generation. Bacterial ribosomes contain three pieces of RNA. These are designated, according to their sedimentation constants, as 5S, 16S, and 23S and contain about 120, 1700, and 3300 nucleotides, respectively. All three pieces appear in cells as parts of larger **pre-rRNA** precursor molecules with extra nucleotide sequences at both the 3' and 5' ends.[207,208]

There are seven rRNA regions in the *E. coli* chromosome.[208a] Each region consists of a single transcriptional unit containing a gene each for 16S, 23S, and 5S

rRNA with interspersed tRNA genes as follows: 16S, tRNA, 23S, tRNA, 5S, tRNA. A single transcript (which can, in certain mutant strains, appear as a 30S molecule) is cut by the endonuclease RNAse III into the smaller pre-rRNA molecules.[209,210] Other nucleases trim these to their final sizes, and methylases act to modify 24 residues in *E. coli* rRNAs.[208,211–213] Most RNAs of all organisms contain **pseudouridine** (ψ), which is formed by isomerization of specific uridines present in the RNAs (Eq. 28-3). In *E. coli* there is one pseudouridine in the 16S ribosome RNA and nine in the 23S RNA as well as one or more in most tRNAs.[214–217] The isomerization depends upon a carboxylate group of the enzyme, which evidently adds to the 6-position of the uracil ring to form a pivot around which the ring can rotate after it is eliminated from its attachment to the RNA and before it is reattached with a C–N linkage (Eq. 28-3).[217] Both bacteria and eukaryotes contain several pseudouridine synthases, which act to isomerize specific uridine residues in the precursor RNAs.[218,218a,b] In eukaryotes special **guide RNAs** direct the pseudouridine synthases to specific locations in their substrates.[174,219] The same thing is true for 2'-*O*-methylases that modify selected ribose rings in precursor RNAs.[219,220]

Transfer RNA. The genes for tRNA molecules in both bacteria and mammalian cells are grouped in clusters, which are transcribed as large precursors sometimes containing more than one kind of tRNA or containing tRNA fused to rRNA or mRNA sequences. At least three different nucleases are needed for cutting and trimming to form the mature tRNA. These enzymes may not always act in the same sequence. Thus, for some but not all tRNAs cleavage near the 3' end is needed before cleavage can take place at the 5' end.

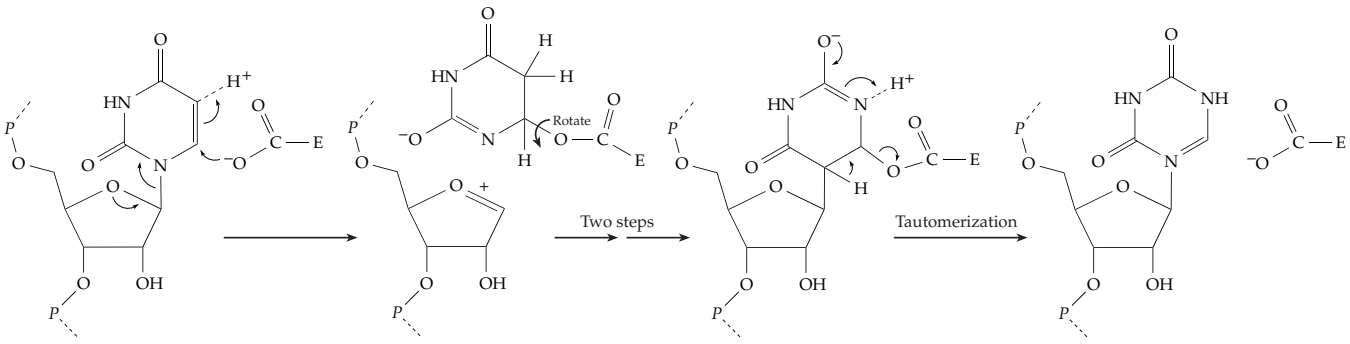
The best known processing nuclease is **RNase P,** which cleaves bacterial tRNA precursors to create the 5' ends of the mature tRNAs. All of the 64 tRNA precursors present in *E. coli* are cleaved by this unusual enzyme,[221–222c] which contains an essential piece of RNA (Chapter 12, Section D,6). Cleavage of polycistronic tRNA precursors by RNase P or of the previously mentioned rRNA and tRNA pecursor by RNAses P and III releases mature tRNAs carrying extra nucleotides at the 3' ends are removed by exoribonucleases, a surprising number of which are present in cells. In *E. coli* they include RNases II, D, BN, and T,[223,224] as well as polynucleotide phosphorylase.[225]

The structure of the precursor to the minor *E. coli* tyrosine $tRNA_1$ is shown in Fig. 28-10. This is encoded by the *amber* suppressor gene *Sup*F (see Chapter 29). Transcription of its gene is initiated by GTP 43 bp upstream of the 5' end of the mature tRNA and usually terminates at a ρ-dependent signal 225 bp beyond the CCA terminus of the tRNA. An endonuclease cuts the transcript a few nucleotides beyond the CCA end. It is then trimmed to an ~130-nucleotide piece still containing 2–3 extra nucleotides at the 3' end. This intermediate is cut by RNase P at the 5' end after which final trimming is done at the 3' end (Fig. 28-10).[226,227]

An important chemical achievement was the synthesis by H. G. Khorana and associates of the double-stranded DNA segment coding for this *E. coli* tyrosine tRNA[228] and its precursor.[229] This was one of the first synthetic genes (Chapter 5). The synthesis was extended to include the gene termination region, which lies beyond the CCA end of the tRNA. Two noteworthy features appeared. There is a local center of dyad symmetry (indicated by vertical bars and a central dot in Fig. 28-10), which may serve as a termination signal. The operator is located in the 29-nucleotide sequence *preceding* the tyrosine tRNA gene.[230–232]

The 3'-terminal group of three nucleotides, CCA, is invariant among all tRNA molecules and is labile, undergoing active removal and resynthesis. The rate of this turnover is sufficient to involve about 20% of the tRNA molecules of a cell per generation, but it is very much slower than the rate of participation of the tRNA molecules in protein synthesis. The physiological significance of end turnover is unknown.[233] While this CCA sequence is encoded in bacterial tRNA genes, it is added in a separate reaction in eukaryotes.[234]

In addition to the cutting and trimming of precursors by nucleases, extensive modification of purine and pyrimidine bases is required to generate mature tRNAs.[235] Some of these modification reactions are



Uridine in an RNA chain — Pseudouridine in RNA (28-3)

discussed in Chapter 5 (see Fig. 5-33). In Fig. 28-10 the modifications in the mature tyrosine suppressor tRNA are shown, and in Fig. 5-30 those in *E. coli* phenylalanine tRNA are indicated. Modification usually begins with tRNA precursors. For example, the precursor in Fig. 28-10 is methylated to form ribothymidine at position 63.[236] Pseudouridine is then introduced at positions 40 and 64 by isomerization of the uridines present in the initial transcript (Eq. 28-3).[237] The Tψ pair at position 63 and 64 of Fig. 28-10 is almost universally found in tRNA,[238] but the positions are usually designated as 54 and 55 as in Fig. 5-30. Position 8 in most tRNAs is occupied by 4-thiouridine. The sulfur atom is transferred from cysteine as S^0 using a PLP-dependent mechanism similar to that in Eq. 14-34 and involving an ***S*-sulfonylcysteine (persulfide) intermediate**.[239,240] Some modifications are completed on the mature tRNA.[241] Some tRNAs require RNA splicing for maturation (see Section D,5).[205]

The hypermodified nucleoside **queuosine** is found in the first (wobble) position of anticodons of most eukaryotic tRNAs for Asn, Asp, His, and Tyr and also in most bacteria.[242–244] Bacteria apparently make aminomethyl-7-deazaguanine from guanine (Eq. 28-4, step *a*) and transfer this compound into the appropriate position in tRNA (step *b*) by a tRNA-guanine transglycosylase.[245–246a] The incorporated nucleoside is then converted to queuine by incorporation of the 5-carbons of the ribosyl group in *S*-adenosylmethionine (Eq. 28-4, step *c*) to form an intermediate epoxide.[247] This is converted to queuine in the vitamin B_{12}-dependent step *d*. Eukaryotes are unable to form queuine and must obtain it as a nutrient factor or from intestinal flora.[244] It is exchanged into tRNA by action of the transglycosylase.[245] Queuine might be considered a

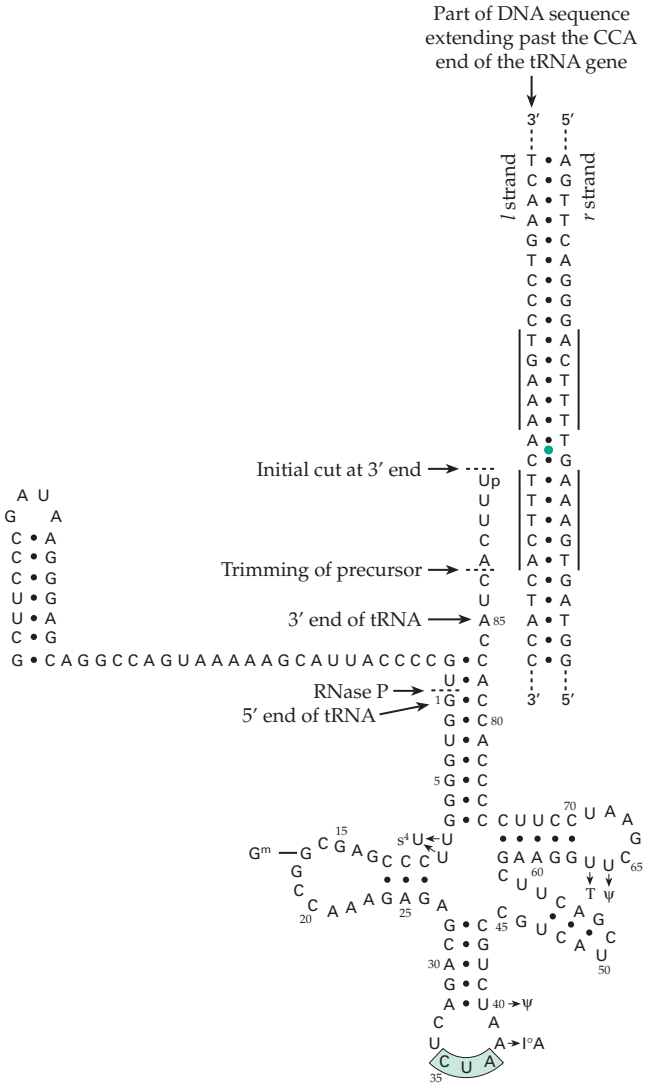
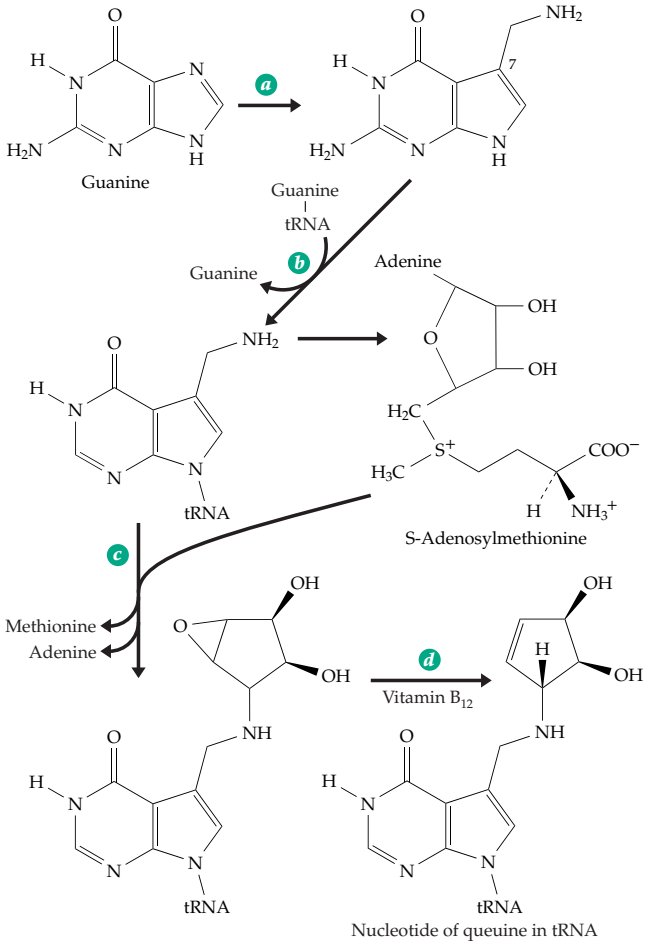


Figure 28-10 Sequence of an *E. coli* tyrosine tRNA precursor drawn in a hypothetical secondary structure. Nucleotides found modified in the mature tRNA are indicated with their modifications (S^4, 4-thiouridine; G^m, 2′-*O*-methylguanosine; I^0, N^6-isopentenyladenosine; ψ, pseudouridine; T, ribothymidine; see also Fig. 5-33).[241] A partial sequence of the tRNA gene past the CCA end is also shown. Note the region of local 2-fold rotational symmetry (indicated by the bars and the dot). The anticodon 3′-CUA (shaded) of this suppressor tRNA pairs with termination codon 5′-UAG.



(28-4)

vitamin. However, germ-free mice reared on a queuine-deficient diet seemed normal. Therefore, the essentiality of queuine in the human diet is in doubt.[242,243]

B. Viral RNA in Prokaryotes

Bacteria not only transcribe their own genes but sometimes transcribe, or assist in transcribing, genes of invading DNA viruses or of integrated proviruses. In addition, they assist in replication of RNA viruses, another process that requires RNA synthesis. Viruses sometimes make use of host RNA polymerases but often synthesize their own catalytic subunits. Bacteriophage T4 uses the *E. coli* RNA polymerase and σ factors but modifies their action through the binding of several phage-encoded proteins.[248] In contrast, phage T7 encodes its own relatively simple RNAP whose initiation complex (Section A,2)[29] and elongation complexes have been studied.[249–249b]

1. The Lysogenic State of Phage λ

The study of bacteriophage lambda has provided many insights into biological process.[250] As we have seen (Chapter 27), the DNA of phage λ can become incorporated into the genomic DNA of *E. coli*. The resulting prophage contains many genes (Fig. 28-11), but they remain largely unexpressed until the SOS signal (Chapter 27) is generated. Certain prophage genes are then expressed with the result that the λ DNA is excised as a replicating virus.[251] How can the λ genes remain unexpressed in the prophage but be expressed rapidly at the time of excision? Part of the answer has been found[252–254] in the *cI* and *Cro* repressors. The short L1 operon (Fig. 28-11) of the λ prophage is transcribed continuously by the *E. coli* RNA polymerase. This operon contains genes *cI* and *rex*, which are transcribed from the *l* strands of the prophage DNA as indicated in Fig. 28-11. The protein C_I (or CI) specified by gene *cI* is the **lambda repressor,** which binds to two operator sites in the prophage DNA. One operator (o_L) is to the left and the other (o_R) to the right of the *cI* gene. From a study of fragments of DNA protected by the repressor, it was concluded that each operator has three subsites, which are filled from left to right at o_L and from right to left at o_R successively by up to six repressor monomers. Each presumed subsite has a similar 17-bp quasipalindromic sequence to which a dimeric repressor can bind. The binding is cooperative, probably because the repressor molecules contact each other, apparently binding the DNA into a loop.

The right operator o_R controls not only the R1 operon but also the L1 operon, which encodes the λ repressor (*cI* gene). The first of the three subsites in the operator is adjacent to the L1 promoter P_{RM}, and binding of λ repressor activates that promoter at the same time that binding to the adjacent subsites blocks transcription of the R1 operon. Thus, the λ repressor positively controls its own synthesis. At the same time, blocking of promoters P_R and P_L prevents synthesis of virally encoded enzymes that catalyze excision of the λ DNA and replication and transcription of the rest of the genes.

The SOS signal causes rapid hydrolytic cleavage of the λ repressor and transcription of the other λ operons. The matter is more complex than this. Gene products *cIII* and *cII*, from the **early left** and **early right operons**, respectively, stimulate the transcription of *cI* and are needed for establishing the lysogenic state initially.[256] Once established these genes do not function since they are never transcribed. There are only a few molecules of the λ repressor present in a cell, but this is ordinarily sufficient to maintain the prophage state. On the other hand, irradiation of the bacterium with ultraviolet light activates the SOS response and results in rapid hydrolytic cleavage of the λ repressor and transcription of other phage operons.

Of special significance to the lytic cycle is the Cro repressor gene *cro*, found at the beginning of operon R1. Although it binds to the same operator sequences as does the λ repressor, the Cro repressor has opposite effects.[254] It represses transcription of operon L2 and hence synthesis of λ repressor, but it positively activates P_R and P_L. The earliest proteins synthesized during lytic development are the Cro protein and the product of the first gene *N* in the left operon L2. The N protein, an antiterminator that permits transcription to continue on past points t_L and t_R, is an unstable, short-lived molecule of $t_{1/2}$ ~2 min.[257] Leftward transcription proceeds through genes *exo* and β, which are involved with recombination, and *xis*, which is required for excision. When the DNA is integrated into the *E. coli* chromosome, it is cut at points *aa'* (Fig. 28-11) and is inserted just to the right of the *gal* operon (Fig. 26-4). Prophage transcription can now continue past point a' and into the genes of the bacterium. Translation of the mRNA formed from this early left operon generates the enzymes needed to free the prophage and to permit reformation of the circular replicative form of the phage DNA. The excision is also made near point a', and it is easy to see how the nearby *gal* genes can sometimes be included in the excised λ genome.

The product of gene *N* also permits rightward transcription through genes *O*, *P*, and *Q* and at a slower rate on along the rest of the chromosome to point a. Genes *O* and *P* code for proteins that permit the host replication system to initiate formation of new λ DNA molecules. Replication begins at the point *ori* and occurs in both directions. Gene *Q* codes for a

protein that activates transcription of the **late genes** beginning at promoter P_R.

As indicated in Fig. 28-11, the chromosome can be divided into four major operons, the short one that produces repressor, and the early left, early right, and late operons. The early operons code largely for replication and recombination enzymes and control proteins. The late operon is concerned with production of proteins needed for assembly of the virus particles and must be transcribed at an even higher rate; hence the need for the product of gene *Q*. Within the late operon, genes *A* to *F* are involved in packaging of λ DNA and in formation of heads, while genes *S* to *J* are concerned with the production and assembly of tails. Genes *S* and *R* produce proteins that lead to destruction of the host membrane and to lysis of the cell. During the late stages of lytic growth the early genes are largely shut off by the Cro repressor. We can see that even in a virus the control of transcription can be a complex process.

2. Replication of RNA Bacteriophages

The small icosahedral RNA-containing bacteriophage are of interest because of the small number of

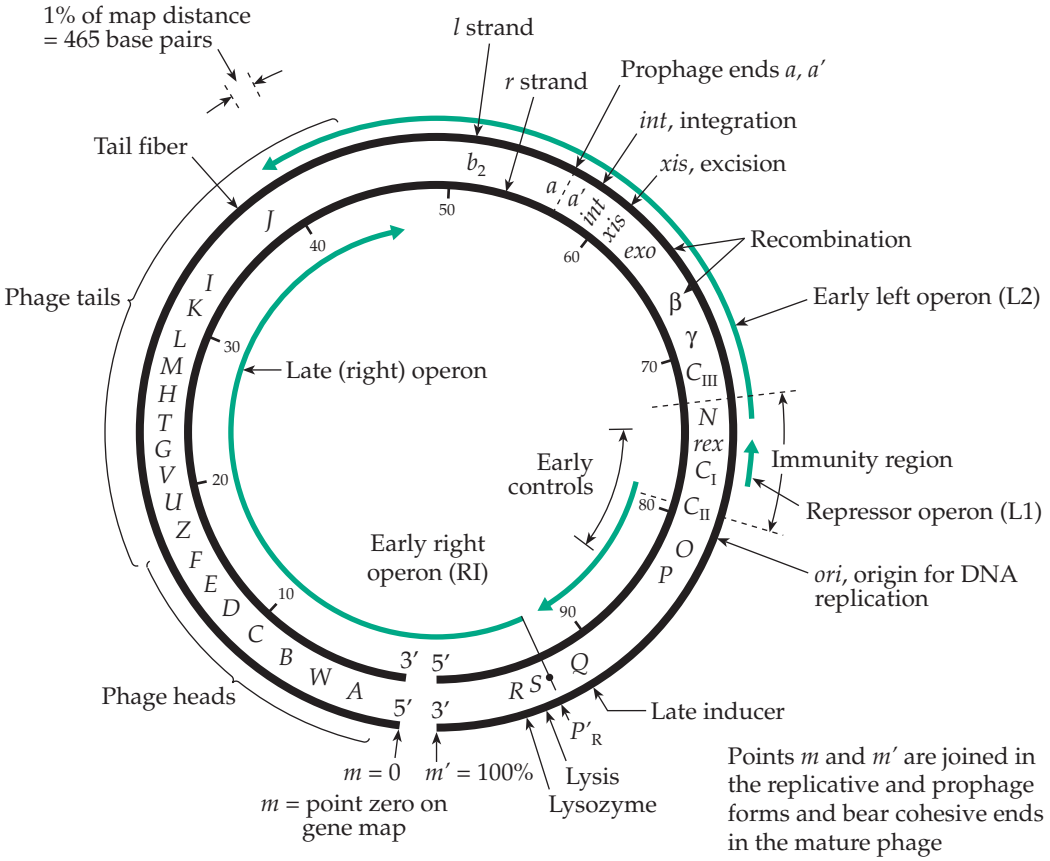


Figure 28-11 Genetic and physical map of the λ phage genome. After Szybalski. See Honigman *et al.*[255] for a more detailed diagram of the immunity region. The gene for the lambda repressor is labeled C_I.

genes and the possibility of obtaining a detailed understanding of their replication. The four genes in the 3569-nucleotide MS2 RNA (Fig. 29-17) code for the A protein (maturation protein), the coat protein, a replicase (RNA polymerase) subunit, and a protein needed for lysis of the host cell. The last gene overlaps both the replicase gene and that of the coat protein.[258,259] The somewhat larger phage Qβ contains a 4.2-kb RNA genome. One subunit of the Qβ replicase is encoded by the virus, but three bacterial proteins are needed to form the complete replicase.[260] They are ribosomal protein S1 and elongation factors EF-Tu and EF-Ts, proteins that normally function within *E. coli* in translation of mRNA (Chapter 29). Their ability to associate with RNA has been exploited by the phage for a quite different purpose.

Replication of a single-stranded virus must take place in two steps. From the (+) strand present in the virus a complementary (–) strand is first formed. Initiation of this step requires another bacterial host factor Hfq[260,261] and GTP. The (–) strands formed do not associate with the (+) strands. They are apparently released from the replicase in a single-stranded form and presumably fold into highly structured molecules with many hairpin loops (as for the (+) strand of MS2 RNA shown in Fig. 29-17). The (–) strands are then copied (Hfq is not needed for this) to make a large number of new (+) strands for incorporation into the finished virus particles. The Qβ replicase is able to synthesize *in vitro* complete complementary strands to either (+) or (–) viral RNA molecules. However, the system is specific for the viral RNA and will not copy any arbitrary nucleotide sequence, certain sequences at the 3' end being essential for initiation of replication. During replication in the test tube mistakes are made including premature termination and mispairing of bases.[262,263] Thus mutation takes place, and it is possible to select RNA molecules much smaller than the original viral RNA that will be replicated readily by the Qβ replicase system. One such fragment contains only 114 nucleotides in a known sequence.[264]

C. Transcription in Eukaryotic Cells and in Archaea

There are three primary domains of life, represented by the bacteria, archaea, and eukaryotes. Some of the clearest evidence for the independent evolution of these three groups of organisms is found in the transcriptional apparatus. While the basic chemistry is the same, the details of initiation and control of transcription in bacteria and in eukaryotes are very different.[264a] The archaea share characteristics of both bacteria and eukaryotes. Archaeal RNA polymerases have a complexity similar to that of eukaryotes and also share a similar mechanism of initiation of transcription.[265–266b] Several of the protein transcription factors of archaea also resemble those of eukaryotes.[267,268] However, in a comparison of DNA sequences from the complete genomes of four archaeal species, it was found that of 280 predicted transcription factors or transcription-associated proteins 168 were homologous to bacterial proteins and only 51 to eukaryotic proteins.[268] This tends to confirm the ancient divergence of the three primary domains of life.

In bacteria transcription and translation are closely linked. Polyribosomes may assemble on single DNA strands as shown in Fig. 28-5. It has often been assumed that RNA synthesis occurs on loops of DNA that extend out into the cytosol. However, recent studies indicate that most transcription occurs in the dense nucleoid and that assembly of ribosomes takes place in the cytosol.[268a] In a similar way eukaryotic transcription occurs in the nucleus and protein synthesis in the cytosol. Nevertheless, some active ribosomes are present in the nucleus.[268b]

1. Eukaryotic Nuclei and Transcription

In cells with true membrane-enclosed nuclei the messenger RNA molecules are relatively long lived. They must move out from the nucleus to the sites of protein synthesis in the cytoplasm. In addition to the need for eukaryotic mRNA to travel further and to last longer than that of bacteria, a number of other differences are evident. Eukaryotic mRNAs are usually transcribed from single genes. Polygenic operons are uncommon in most animals but are numerous in *C. elegans*.[268c] Eukaryotic cells appear to rely less on negative control through specialized protein repressors than do bacteria but use a greater variety of positive control mechanisms. However, most genes are repressed by being held in a **silent state**.

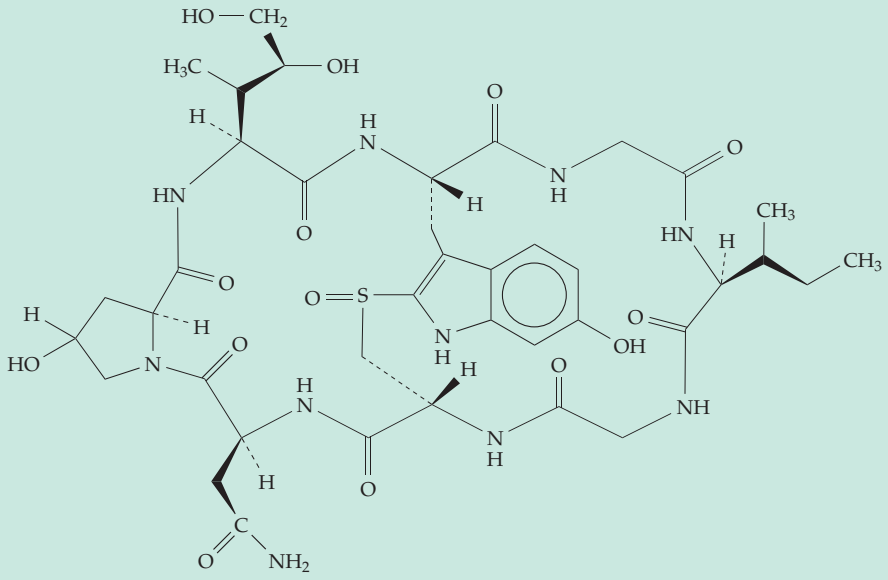
Another characteristic of eukaryotes is the extensive processing of transcripts. Most primary transcripts that give rise to mRNA appear first in the nucleus as **heterogeneous nuclear RNA** (**hnRNA**). Like mRNA it has a base composition resembling that of DNA. The molecular size varies from 1.5 to 30 kb or more. It turns over rapidly, most of it having a half-life of ~10 min. However, some may last as long as 20 h. Only about 5% of the hnRNA ever leaves the nucleus as mRNA, most being degraded without export to the cytoplasm.[269] The processing consists of **capping** at the 5' ends, removal of introns (splicing), **cleavage** by nucleases, **polyadenylation** at the 3' ends, **methylation**, formation of pseudouridines, other covalent base modifications, and sometimes **editing**. Because of the complexity of eukaryotic transcription there are many points at which control can be exerted during initiation of transcription, termination of transcription, splicing and methylation, transport of mRNA out of

the nucleus, and the degradation of mRNA. However, the first points of control are found in the nucleosomes and in the structure of the chromatin.

It has been recognized for many years that genes in heterochromatin (Chapter 27) are usually not transcribed. In most chromatin the great majority of the genes are silent most of the time. Genes are repressed by the folding of the DNA into nucleosomes and by the further folding into higher order folds or coils (Fig. 27-5).[270–272] Various activating transcription factors as well as RNA polymerases must bind to the DNA, displacing it from the histones around which it is wrapped in the nucleosomes. The processes by which inactive nucleosomal DNA becomes active in response to external signals are beginning to be understood. Chemical alterations in the histones in the nuclear matrix and in other nuclear proteins and also in the DNA itself may be involved.[273–275] As pointed out in Chapter 27, the CpG "islands" that lie upstream of many genes are heavily methylated in the silent heterochromatin. Repressor proteins may bind to methyl-CpG groups.[276,276a]

In recent years attention has been focused on the N-terminal "tails" of histones H3 and H4 (Fig. 27-4) in which lysine side chains undergo reversible acetylation and which may also be phosphorylated and

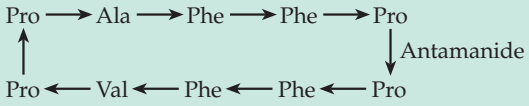
BOX 28-B POWERFUL POISONS FROM MUSHROOMS



Several deadly species of the genus *Amanita* produce colorless toxic octapeptides, the **amanitins**.[a,b] Two residues of glycine, one of L-isoleucine, one of the unusual L-dihydroxyisoleucine, one of L-asparagine, and one of L-hydroxyproline are present in α-amanitin. In the center a modified tryptophan residue has been combined oxidatively with an SH group of a cysteine residue. If the dihydroxyisoleucine residue of α-amanitin is replaced with unhydroxylated leucine, the resulting compound, known as amanullin, is nontoxic. The LD_{50} for mice is 0.3 mg kg^{-1} and 50 g of fresh *Amanita phalloides* may be sufficient to kill a person. Amanitins act slowly, and it is impossible to kill mice in less than 15 h, no matter how high the dose.

α-Amanitin completely blocks transcription by eukaryotic RNA polymerases II and III. Polymerase II is the major nuclear RNA polymerase, and its inhibition prevents almost all protein synthesis by the cell. Note that the amanitin molecule is semisymmetric overall, much as is a actinomycin (Box 28-A), with an aromatic group protruding from behind in the center.

The same mushrooms contain several fast-acting toxic heptapeptides, the **phalloidins**, whose structures are similar to those of the amanitins. However, they contain a reduced sulfur atom (—S—) in the cross-bridge. They are specifically toxic to the liver.[c] The same mushrooms also contain an antidote to the phalloidins, **antamanide**. This cyclic decapeptide, like the toxins, is made up entirely of L-amino acids, and it apparently competes for the



binding site of the phalloidins. Unfortunately, it is of little value in treating cases of mushroom poisoning. Antamanide is a specific sodium-binding ionophore.

[a] Wieland, T., and Wieland, O. (1972) *Microb. Toxins* **8**, 249–280
[b] Wieland, T., and Faulstich, H. (1983) *Handbook of Natural Toxins*, Marcel Dekker, New York
[c] Wieland, T., Nassal, M., Kramer, W., Fricker, G., Bickel, U., and Kurz, G. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5232–5236

methylated.[33b,271,275a–c] The resulting modifications in shape and electrical charge can affect the ways in which the histones interact with the DNA, with each other, and with other proteins in transcription complexes.[271a,275a]The histone tails may interact with adjacent molecules either to pack the chromatin more tightly or to loosen it and allow transcription to take place. Active chromatin has long been known to be highly acetylated, while silent chromatin has a low degree of acetylation, but a high degree of methylation of both histones and DNA (pp. 1541–1542). Methylation of histone H3 on lysine 9 (H3-Lys^9 or H3-K9) is especially significant.[275d,e] See Chapter 32, Section A,1 and C,1 for further discussion.

Several transcriptional activators form complexes with acetylating enzymes, the **histone acetyltransferases** (**HATs**), while transcriptional repressor proteins often associate with **histone deacetylases**.[277–279c] The deacetylases are often found in very large complexes. For example, the mammalian complex **Sin3** contains two histone deacetylases plus at least five other subunits, some of which evidently bind to histones.[277–278a] Sodium butyrate in millimolar concentrations is a powerful inhibitor of these deacetylases.[280] Special **silencer** sequences in the DNA are sometimes present and provide sites for binding of transcriptional repressors. Among these are **silent information repressors** (**Sir** proteins). They regulate regions of DNA that can be converted to a heterochromatin-like state. They have been located in silenced mating type loci in yeast (see p. 1880), in telomeres, and in DNA containing ribosomal RNA genes.[280a,b] Sir proteins have an unusual histone deacetylase activity. The acetyl groups removed from histones are transferred by reaction with NAD^+ to ADP ribose (see Eq. 15-16).[280c,d] Regions of silenced DNA are often set apart by **insulator** or bounding regions.[280e–g]

An important mechanism of silencing some genes is the use of an antisense RNA strand, as is illustrated for a bacterial system in Fig. 28-8. This **RNA interference** is also used in animals and plants, often as a way of blocking replication of viruses.[1,1a,280h–j] The small 20–25 nucleotide siRNAs that function in this way are abundant in *C. elegans* and in *Drosophila*.

Histone acetyltransferases also form large complexes that acetylate not only histones but also other nuclear proteins.[281–283] The **SAGA HAT** complex of *S. cerevisiae* has a molecular mass of ~2.0 MDa and contains at least 14 subunits.[283a] Large acetylating complexes have also been identified in *Tetrahymena*, *Drosophila*, *Arabidopsis*, and mammalian species.[283] Some subunits of these complexes have been identified as previously known transcription factors. Such multiprotein complexes are sometimes described as **cis-regulatory elements** (CREs).[283b,c]

Changes in the structural properties of chromatin observed during silencing of genes or during their activation are often described as **chromatin remodeling**. Large multisubunit complexes are involved. Their action is characteristically dependent upon ATP hydrolysis.[284–284b] Complexes **SWI/SNF** and **RSC**, first found in yeast but also present in human cells, appear to participate in the disruption of nucleosomes needed for initiation of transcription.[284–287c] Among other distinctly different remodeling complexes are **ISWI** of *Drosophila*, its human homolog, and the human Williams syndrome transcription factor **WCRF**.[284] The ATP-dependent component in these complexes has a conserved sequence that is shared with DNA helicases. RNA helicases are also required for all processes that form, modify, or utilize RNA.[288,288a–c] However, the chromatin remodeling complexes appear not to unwind DNA but to open the nucleosomal DNA for initiation of transcription.[272,284] They may act in a processive fashion and be coupled to transcription. **Peptidyl-prolyl isomerases**, such as the cyclophilins (Box 9-F) may also be essential components of chromatin-remodeling complexes.[289]

Just as they participate in driving the cell cycle, ubiquitin and proteasomes also function in the control of transcription. In many cases specific transcription factors are targeted for destruction after they are used to activate or repress a gene.[289a–c] However, a 19S regulatory complex, which consists of a base and a lid of the proteasome (Box 7-A), may participate directly in control of transcription rather than in mediating proteolysis.[289a–e] In addition to ubiquitin a 97-residue relative designated SUMO-1 is linked to proteins by enzymes resembling E_1 and E_2 of the ubiquitin system (Box 10-C). Conjugation with **SUOMO-1** regulates some transcription factors and has other functions,[289f,g] e.g., participation in control of nuclear pores. The 81-residue protein **NEDD8**, which is 80% homologous with ubiquitin, controls some transcriptional processes in heart and skeletal muscle.[289h,i] Furthermore, ubiquitin-like sequences (UBX domain) are present in the C-terminal ends of a variety of specific proteins.[289j]

2. RNA Polymerases

Eukaryotic nuclei contain at least three RNA polymerases[269,290–292] which have the following functions:

Polymerase I	Formation of large pre-rRNAs
Polymerase II	Transcription of most genes to give precursors to mRNA and most small nuclear RNAs (snRNAs and small nucleolar RNAs (snoRNAs)
Polymerase III	Formation of 5S rRNA, tRNAs, and small RNA U6

Polymerase I is localized in the nucleolus.[293,294] Mitochondria contain a fourth RNA polymerase[295,295a] and chloroplasts a fifth.[296]

Like the bacterial polymerase, eukaryotic RNA polymerases are large 500–600 kDa aggregates of 9–14 subunits each. Yeast and human RNA Pol IIs each contain twelve subunits (Fig. 28-12).[290] There are two large nonidentical subunits, which in mammalian cells have masses of 214 kDa and 140 kDa and are homologous to the β′ and β core subunits of the *E. coli* polymerase, respectively.[291,297–299c] The active site contains one or two catalytic Mg^{2+} ions.[290] The largest subunit has an unusual singly glycosylated C-terminal domain (known as the **CTD**). It contains the repeating sequence $(YSPTSPS)_n$. The number of repeats varies: $n = 18$ in plasmodia, 27 in yeast, 45 in *Drosophila*, and 52 in mammals.[300] The numerous serine side chains in this tail domain undergo phosphorylation and dephosphorylation to varying extents during each catalytic cycle.[301] This may be a way of easing the transcriptional complex through nucleosomes,[302] but its most important function appears to be the linking of transcription to pre-mRNA processing.[303,304]

RNA polymerases I and III have properties similar to those of Pol II.[304a,b] Polymerases II and III are very sensitive to inhibition by the lethal mushroom poison **amanitin** (Box 28-B). However, both RNA polymerase I and RNA polymerases of mitochondria resemble the bacterial enzyme in being resistant. Genes transcribed by polymerase I, II, and III are often referred to as genes of classes I, II, and III, respectively. While mRNA is transcribed from class II genes, rRNAs, tRNAs, and some small RNAs, which must undergo processing but are not polyadenylated, are transcribed from genes of classes I and III. Each type of nuclear RNA functions in its own sites as independent "factories."[305,306]

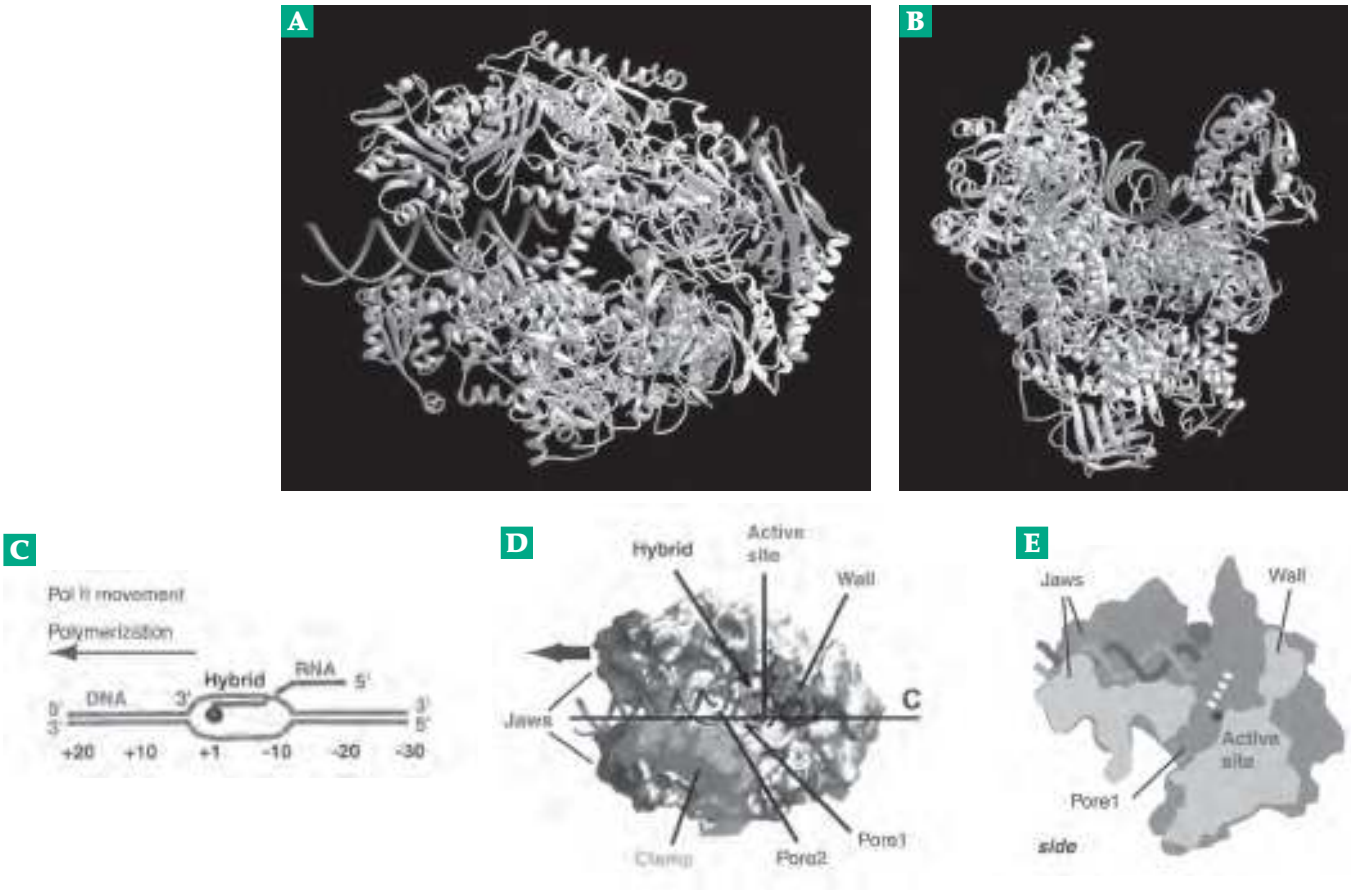
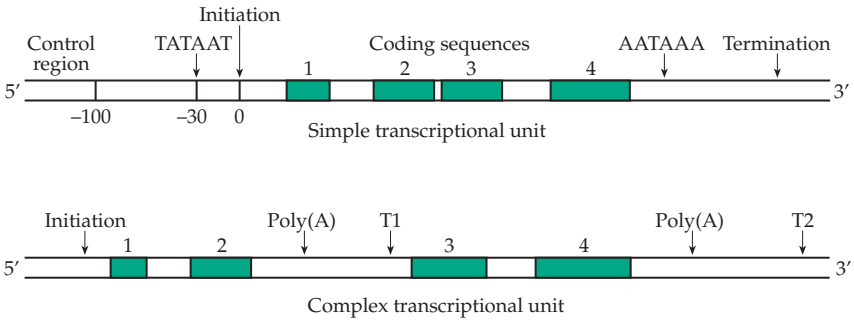


Figure 28-12 Three-dimensional structure of yeast RNA polymerase II. (A) View from the "top," with backbones of ten subunits (of 12) in the 514-kDa protein shown as ribbon drawings. A 20-base-pair segment of B-DNA has been modeled in a location indicated by electron crystallography. (B) Side view looking toward the end of the DNA. Eight zinc atoms as well as an active-site magnesium (green) are visible. (C) Schematic drawing showing the transcription bubble as proposed for a transcribing polymerase. (D) Surface representation of the polymerase viewed as in (A). (E) Side view of a section cut along the line marked C in D. The dashed white line represents the axis of the DNA-RNA hybrid segment. The hybrid axis must lie at an anlge with respect to the axis of the incoming DNA. Pore 1 may be a route for exit of RNA during "backtracking." The nucleotide triphosphate substrates may also enter via pore 1. From Cramer *et al.*[290] Courtesy of Roger D. Kornberg.

3. Transcriptional Units and Initiation of RNA Synthesis

Typical simple **transcriptional units** for class II eukaryotic genes contain the following elements: (a) site of initiation, (b) **TATA sequence** (Goldberg–Hogness sequence) at position ~–30 bp, (c) **upstream regulatory elements**, (d) **enhancers**, (e) a series of **coding sequences** or exons separated by introns, an **AATAAA sequence** that in the RNA transcript may establish the 3' polyadenylation site, (f) a termination region.[307]



The preceding paragraph describes a **simple transcriptional unit**. There are also **complex transcriptional units**.[308] For example, the terminator sequence T1 may be followed by additional exons and a second polyadenylation signal and second termination sequence T2. Termination may sometimes occur at T1 and sometimes at T2 resulting, after splicing, in two mRNAs, one containing only exons 1 and 2 and the other all four exons. In some cases two or more different modes of splicing may occur with one or more exons omitted from the final processed mRNA. Thus, a single transcriptional unit can give rise to two or more different proteins which share some common sequences.[309–312] Multiple initiation sites sometimes exist as well.

Initiation of RNA synthesis is a complex process, which is summarized diagrammatically in Eq. 28-5. Some details are given in Sections 4–7. Elongation of the RNA being formed is also complex, often requiring splicing and other processing steps before synthesis can be terminated. These steps, which are discussed in Section 8 and in Section D, also depend upon large complexes of proteins, which are directly coupled to the RNA polymerase.[312a–c]

4. Promoters, Transcription Factors, Enhancers, and Activators

Eukaryotic promoter sequences are less well defined than are those of bacteria, and the initiation points for transcription are more variable. Most promoters contain the TATA sequence, which is usually required for binding of polymerase II. In addition, there are upstream regulatory elements (Table 28-1). Many promoters contain the sequence 5'- **CCAAT** at about –75, and others have 5'- **GGGCGG** or similar sequences at close to –100.[273, 273a,b] A sequence found upstream from many yeast genes is 5'- **TGACTC** or the longer semipalindromic 5'-**ATGACTCAT**.[313,314] Some sequences are unique to small sets of genes such as the "heat-shock" genes[315] considered in Section 6. Most upstream sequences are not polymerase-binding sites but attachment sites for additional protein transcription factors. Computer programs that help locate pro- moter sequences have been devised.[316,317]

The SV40 early promoter and transcription factor Sp1. The study of transcription in eukaryotes has been difficult because purified polymerases do not initiate transcription at most promoter sites. As a consequence, much of the early work was done with viruses such as adenovirus and SV40. Their genes are transcribed by RNA polymerase II and have unusually effective promoters. A protein known as **Sp1**, isolated from human cultured cells, protects an SV40 promoter from digestion by DNase. Sp1, which is now known as an **acessory factor** or **coactivator** for transcription,[318,318a] protects a region that extends from about –45 to –104 bp and contains the hexanucleotide GGGCGG sequence repeated six times. The Sp1 protein was isolated as a mixture of related 95- and 105-kDa peptides[319] and was found to contain two DNA-binding zinc-finger domains (Figs. 5-37, 5-38).[320] The Sp1 protein is synthesized in most cells and binds not only to SV40 promoters but also to many promoters of host cells[321] (an example is provided by the mouse mitochondrial aspartate aminotransferase gene whose sequence is shown in Fig. 5-4). In genes that lack a TATA sequence the binding of Sp1 or the related Sp2, 3, and 4 to GGGCGG or similar GC-rich sequences is essential to initiation of transcription.[322,323] A possible role is to assist in nucleosome remodeling.[322] Sp1 also binds, together with other transcription factors, to certain enhancer sequences.[324,324a,b] Its effects are modulated by posttranscriptional phosphorylation and glycosylation.[321]

The TATA binding protein and general transcription initiation factors. A slow basal level of transcription can be observed when all but a small part of the control region at the 5' end of a gene is deleted.[325] This minimum promoter, which includes the TATA sequence, is the binding site of both the RNA

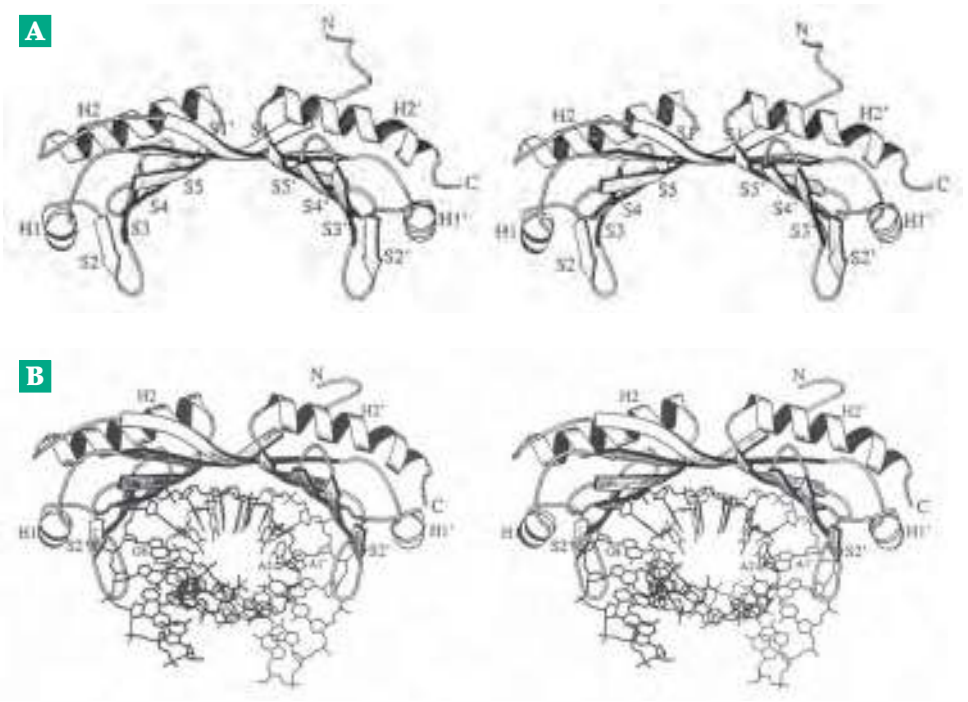
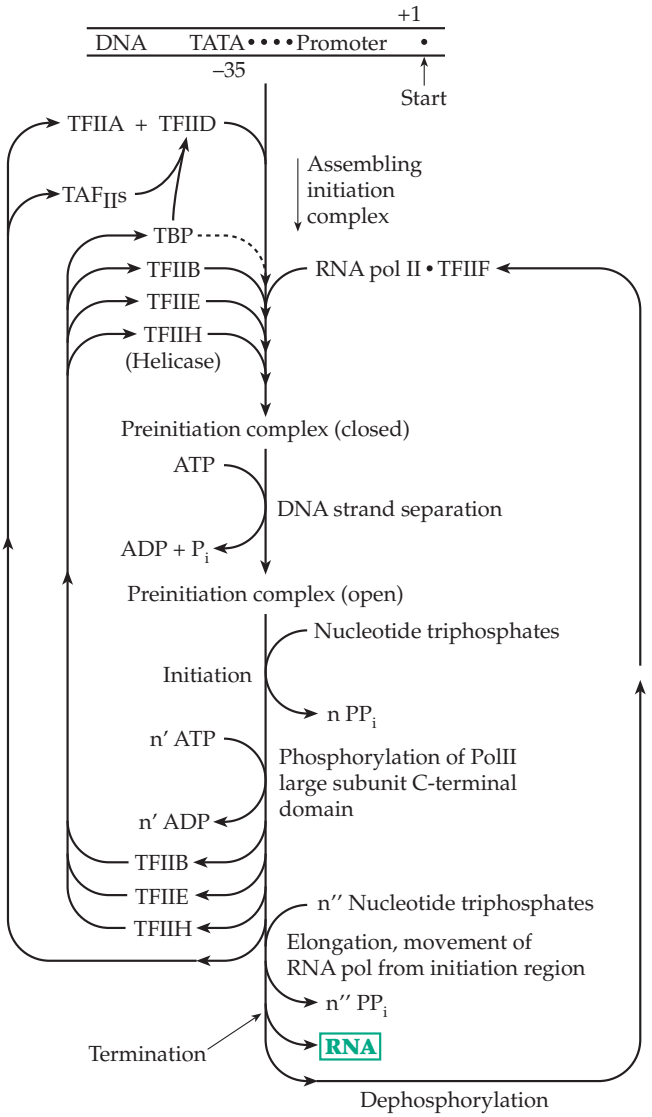


Figure 28-13 (A) Stereoscopic ribbon drawing of the phylogenetically conserved 180-residue C-terminal portion of the TATA-binding protein (TBP) from *Arabidopsis thaliana*. The sequence consists of two direct repeats, giving the protein an approximate twofold symmetry. From Nikolov *et al.*[337] (B) Structure of the corresponding C-terminal core (residues 155–335) of the human TATA-binding protein (TBP) bound to the TATA sequence of a promoter in adenovirus DNA. From Nikolov *et al.*[327] (C) Structure of human transcription factor IIB bound to a TBP from *Arabidopsis thaliana*, which, in turn, is bound to an adenovirus TATA sequence. Hypothetical B DNA extensions have been modeled at both ends of the DNA segment. The +1 at the left end is the transcription start site and the –43 upstream end is to the right. From Nikolov *et al.*[338] Courtesy of Stephen K. Burley.

polymerase and of transcription initiation factors that are designated TFII-A, -B, -D, -E, -F, and H. Because they affect many genes these are called **general transcription factors**.[326–329] TFIID is a large complex of a DNA-binding subunit known as the **TATA-binding protein (TBP)** together with 8–12 additional tightly bound subunits known as **TBP-associated factors** (**TAF_{II}s**). Many of the TAF_{II}s have histone-fold structures. Some possess histone acetyltransferase and other enzymatic activities.[329a] They may stabilize initiation complexes on specific gene promoters.[330–332a] TBP binds specifically to the TATA sequence (Table 28-1),[333] which is found in most promoters for RNA polymerases I, II, and III. Its three-dimensional structure resembles a saddle, which sits astride the TATA sequence (Fig. 28-13). The DNA is bent, untwisted by ~117°, and the minor groove broadened to allow a good fit.[327,328,334] TFIIB is thought to bind first to the DNA–TBP complex, after which the RNA polymerase II complex binds and becomes positioned on a promoter site. Other factors, including the ATP-dependent bidirectional helicase TFIIH,[335,335a] also add (Eq. 28-5). ATP may be needed for more than one step in initiation.[336] TFIID contains TAF_{II} subunits. They may bind along with TBP[336a] as indicated in Eq. 28-5, or they may bind at a later point in an assembly pathway. It is often stated in current literature that the growing initiation complex "recruits" the next subunit. It is important to realize that this simply means that the next subunit that strikes the complex by diffusion sticks, perhaps with cooperativity in binding. The word recruit doesn't mean that the complex advertises a vacancy for the next subunit. The very large complexes (**transcriptosomes**) that are formed vary in composition and in assembly pathways. Proteins related to TBP bind to different promoters.[310] It is also



(28-5)

important to recognize that some promoters bind transcription factors tightly while others do so only weakly. Some are constitutive, always functioning, but most are inducible, acting only upon appropriate stimulation. The many operators present in a genome and the many species of eukaryotes present on earth ensure a vast variety of detailed pathways and control mechanisms.

Transcriptional activators. Many proteins serve as activators of transcription, causing larger increases in rate over those observed with TBP alone. Some of these are listed in Table 28-2.[338a] The table also lists two proteins (Sp1 and NF1), and the DNA sequence CCAAT, which control constitutive or continuously active genes. A large group of transcription factors are active in development. Receptors may be resident in cytoplasm, cell membrane or nucleus, as indicated in Table 28-2. Some cytoplasmic factors are *latent*, becoming active only following stimulation by an external signal. All of the factors in Table 28-2 are positive-acting.[338a] However, some negative-acting factors are known. One, designated **Ssn6-Tup1** in *S. cerevisiae* is a global repressor, affecting many genes.[338b] For example, it opposes the activator GAL4. Proteins related to Ssn6-Tup1 are found in flies, worms, and mammals.

Genetic studies indicate that gene activator proteins often bind to TFIIB, TFIID, and TFIIH.[272] The coactivator **TafII130** (which binds to TFIID), and the bound transcriptional activator Sp1 apparently interact with the protein **huntingtin** in regulation of transcription in the brain.[338c] A defect in huntingtin leads to the fatal neurodegenerative Huntington disease (pp. 1516, Chapter 30). Other activator proteins bind to upstream activator sequences, as in prokaryotes. Among the most studied of these is GAL4, an 881-residue yeast protein that binds to a specific 17-bp palindromic upstream site near the TATA sequence. It activates transcription of genes needed for galactose metabolism.[339–343] The GAL4 protein contains a binuclear metal cluster composed of two Zn^{2+} ions and six cysteine side chains, two of which bridge the pair of metal ions.[341,343] GAL4 is able to activate genes of *Drosophila* and of human cells. The specific GAL4-binding sequence has been introduced into 5' control regions of genes in various positions.[344] It was found that neither an exact distance nor alignment between the GAL4 binding site and the TATA sequence is required, but activation is best when this distance is not too large. The explanation for the lack of a requirement for alignment seems to lie in the flexibility of the C-terminal segment of GAL4, which carries a large negative charge and may bind to the repeated C-terminal sequence of RNA polymerase II to activate it.[345] Many other promoter-specific activators are known.

Mediators and coactivators. Transcriptional activators that act in a crude cell-free system often do not function with purified DNA, RNA polymerase, and the basal transcription factors as indicated in Eq. 28-5. Studies with yeast, *Drosophila*, and human cells revealed that additional large multisubunit complexes known as mediators are needed.[272,346–348] A yeast mediator complex consists of 20 subunits.[349–350b] Many activator proteins bind to the DNA sequences known as enchancers, discussed in the next section. Mediator complexes may also interact with enhancer-bound activators. Individual proteins, such as the TAF subunits, that bind to and cooperate with activator proteins are often called coactivators.[351]

Enhancers. Complex DNA sequences called enhancers help to regulate transcription of many eukaryotic genes. The first of these was discovered in an upstream control region of the virus SV40 DNA

and consists of two repeats of a 72-bp sequence.[47,352,353] The presence of an enhancer sequence may cause as much as 100- to 1000-fold increase in the rate of transcription as compared with the same transcriptional unit from which the enhancer has been deleted. A surprising fact is that enhancers as far as 1–2 kbp upstream or even far downstream of the promoter and in either of the two possible orientations are effective. This finding suggested that enhancers induce long-range conformational alterations in DNA. Alternatively, they might contain points of entry for RNA polymerase or for an initiation factor that could move along the DNA to the promoter region. However, the synthetic DNA molecule shown in Fig. 28-14 contains two copies of an enhancer in opposite orientations in one strand but none in the other strand.[354] The

TABLE 28-1
Nucleotide Sequences of Some Commonly Found Regulatory Elements in Promoters and Upstream Activator Sites

Sequence[a] *transcription*	*Description. Positions are relative to the start site*
RNA polymerase binding sites:	
5'-**TAT**AA**T**[b]	Bacterial –10 or Pribnow promoter sequence
5'-**TT**GACA	Bacterial –35 region promoter sequence
5'-(C/T)TTA(A/T)Ann	Archaeal –30 region, TBP binding site[c]
5'-TATA(A/T)A(A/T) or 5'-TATA@A@n[d]	Eukaryotic –30 region; yeast –60 to –120 region
Upstream eukaryotic promoter sites:	
5'-GGGCGG	–100 region, Sp1
5'-CCAAT	–75 region, CTF
Small GC clusters ~5 bp apart	Binding site for TFIIIA
5'-GCGCC-C	~ –5; "discriminator" sequence: inhibition of gene expression by ppGpp
Enhancer elements and transcription factors:	
5'-ATGA(C/G)TCAT	AP-1, cJun, GCN4 (yeast)
5'-CCCCAGGC	AP-2
5'-CAC(G/T)	Myc / Max heterodimer
5'-ATGACGTCAT	CRE (cAMP responsive element)
5'-GGTCAnnnTGACC	Estrogen-responsive element
5'-GGGTGAnnnGGGTGA	Vitamin D-responsive element; direct repeats
5'-CC$(A/T)_6$GC	SRE (serum response element)
5'-ATGCAAAT	Homeotic genes; Oct-2
5'-GGTCAnnnTGACC	ERE (erythroid responsive element)

[a] The sequences are all for the sense strand of the DNA, n = any base. See Keller, W., König, P., and Richmond, T. J. (1995) *J. Mol. Biol.* **254**, 657–667; Adams, R. L. P., Knowler, J. T., and Leader, D. P. (1992) *The Biochemistry of the Nucleic Acids*, 11th ed., Chapman & Hall, London; Lewin, B. (2000) *Genes VII*, Oxford Univ. Press, New York.

[b] Consensus sequence. The bases in boldface are the most highly conserved.

[c] From DeDecker, B. S., O'Brien, R., Fleming, P. J., Geiger, J. H., Jackson, S. P., and Sigler, P. B. (1996) *J. Mol. Biol.* **264**, 1072–1084.

[d] The symbol @ refers to either A/T or T/A. See Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J., and Dickerson, R. E. (1996) *J. Mol. Biol.* **261**, 239–254.

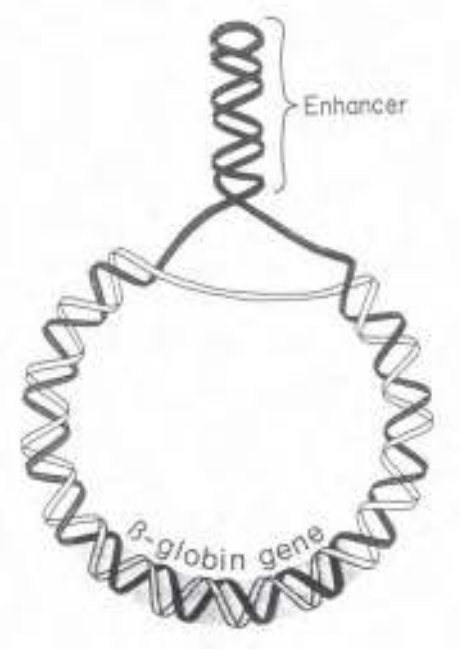
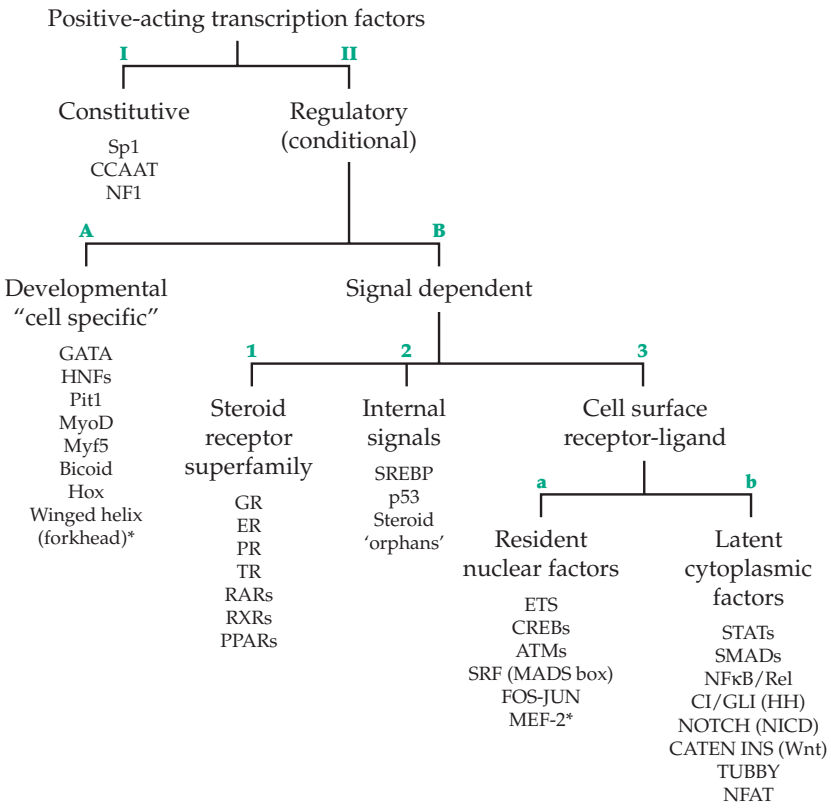


Figure 28-14 A "tailed circle" consisting of an enhancer linked to, but topologically separated from, a gene. One of the DNA strands of this plasmid bears two copies of an SV40 enhancer sequence, one copy inverted with respect to the other. This extra region protrudes from the circle and self-pairs to form a functional enhancer. The main body of the circle contains the β-globin gene, transcription of which is increased by the enhancer. Twisting of the enhancer has no effect on the winding of the strands on the main body of the circle; nevertheless, the enhancer efficiently increases β-globin transcription. From Ptashne.[355]

enhancer is thus topologically separated from the globin gene present in the DNA. Nevertheless, the enhancer functioned efficiently. This suggested that enhancers are protein-binding sequences and that the bound proteins may, perhaps, close a loop to affect the transcription initiation complex.[149,355,356] An enhancer may control a whole loop of DNA in a chromosome affecting many genes.[357] Enhancer DNA sequences are often complex or *modular*, consisting of several shorter elements. For example, the SV40 enhancer contains 8- to 10-bp sequences that are repeated with minor variations and are also present in other enhancers. These are sites for the binding of activator proteins, some of which carry messages from signaling cascades such as the MAP kinase pathway of Fig. 11-13. These pathways are central to the control of cell growth. Their activation is recognized as a mitogenic response. Some enhancers (*nonmodular*) are more compact and directly tied to the RNA pol complex. An example is the human interferon-β (IFN-β) enhancer.[357a,b] It responds cooperatively to three separate activator proteins: NF-κB, interferon regulatory factors, and ATF-2/c-Jun together with an architectural HMG protein. These form an **enhanceosome**, which interacts directly with the transcriptosome complex. The whole complex is sometimes referred to as a cis-regulatory module (CRM). The DNA domains affected by enhancers may be separated by insulator or boundary regions (p. 1626).

TABLE 28-2
Functional Classification of Positive-Acting Eukaryotic Transcription Factors[a]



[a] Major groups are labeled in large type and specific examples are listed below (green). Some bacterial proteins are included. Proteins designated by asterisks can be trapped in the cytoplasm by phosphorylation of serine site chains. From Brivanlou and Darnell.[338a]

5. The Variety of DNA-Binding Proteins

A large number of proteins, often present in very small amounts, bind to DNA and may affect transcription. Some, such as the histones, are relatively nonselective. The more specific transcription factors are often capable of binding tightly, but may do so only upon

allosteric activation. Architectures of transcription factors vary. Some place an α helix in the major groove of B-DNA while for others β strands or peptide loops may interact with the DNA in the wide groove. Many transcription factors bind only to a bent or distorted DNA helix, often with a broadened groove. Intercalation of groups from other domains of a transcription factor within the minor groove may help to bring about necessary distortion.[358] Transcription factors sometimes, perhaps often, have other roles, e.g., as enzymes. For example, a pterin dehydratase acts as a transcriptional cofactor in liver cells.[359]

While architectures vary greatly, the DNA-binding domain or domains are usually rich in positively charged side chains, which may interact directly by hydrogen bonding with the charged phosphate groups of the DNA backbone or indirectly via bound water molecules (see Fig. 5-36).

Leucine zipper transcription factors. A large family of DNA enhancer-binding proteins, which are involved in regulating cell growth in specialized cells of differentiated tissues and also in yeast cells, have related structure domains with a characteristic coiled coil that holds two subunits together.[360,361] An example is the Max structure shown in Fig. 2-21. One of the first proteins of this type to be recognized was the mammalian phorbol ester-induced **cell activator protein AP-1**, which is a heterodimer of the protein **cJun**, encoded by cellular protooncogene *c-jun*, and **cFos**, encoded by protooncogene *c-fos*.[362] The heterodimer binds to the palindromic sequence 5'-TGACTCA. Each of the monomeric proteins c-jun and c-fos, as well as other members of the leucine zipper family, has an N-terminal DNA-binding domain rich in positively charged basic amino acid side chains, an **activation domain** that can interact with other proteins in the initiation complex, and the leucine-rich dimerization domain.[363] The parallel coiled-coil structure (Fig. 2-21) allows for formation of either homodimers or heterodimers. However, cFos alone does not bind to DNA significantly and the cJun/cFos heterodimer binds much more tightly than does cJun alone.[364] The yeast transcriptional activator protein GCN4 binds to the same 5'-TGACTCA sequence as does the mammalian AP-1 and also has a leucine zipper structure.[360,364,365]

Several mammalian leucine zipper proteins bind to the CCAAT sequence (Table 28-1) and are, therefore, as a family designated C/EBP.[361,366,367] A 30-residue segment of C/EBP contains four leucine residues at 7-residue intervals. When plotted as a helical wheel (Fig. 2-20) the four leucines are aligned on one side.[368] Similar sequences are present in the proteins cMyc, cJun, and cFos and in GCN4. These observations suggested that if the peptide sequence forms an α helix, the leucine side chain from two identical subunits or closely related proteins might interdigitate in a knobs-in-holes fashion to form the leucine zipper (Fig. 2-21).[368] The structures of these leucine zipper proteins have now been thoroughly investigated and verified by X-ray and NMR methods.[360,369–371] Mutational alterations in the zipper regions of these proteins decreases both activation[372] and dimerization.[373] The CCAAT sequence is found in many enhancers[366,374,375] and is present in ~25% of all eukaryotic promoters that function in differentiated tissues. A trimeric protein known as NF-Y binds specifically to this sequence.[366] Its subunits contain glutamine-rich domains and histone-fold domains that suggest formation of nucleosome-like structures.[366,376]

Another transcriptional activator of the leucine zipper class is **Myc**, a product of the *c-myc* oncogene (Chapter 11) and a key regulator of both cell growth and programmed cell death (apoptosis).[377–379a] It binds to the sequence 5'-CACGTG, and its binding is greatly enhanced by formation of the heterodimer **Myc/Max** (Fig. 2-21). Max is not an activator and may dimerize with certain other proteins to become an inhibitor of transcription. However, it is ubiquitously present and ready to join with Myc to activate an appropriate series of genes. A related enhancer sequence 5'-TGACGTCA is a **cyclic AMP response element** (CRE) that functions within hormone-responsive tissues that use cAMP as a second messenger.[364,380] Cyclic AMP activates protein kinase A (Fig. 11-4), whose catalytic subunit diffuses into the nucleus and phosphorylates the cyclic AMP-response element-binding protein **CREB**, a coactivator that binds to the CRE and which also contacts the general transcriptional factor complex.[365,381–382b] Transcriptional responses to cAMP are quite complex. For example, activation of the phosphoenolpyruvate carboxykinase (PEPCK) gene is maximal only when CREB, C/EBP, and AP-1 are all bound at adjacent DNA sites.[382c] CREB (also known as ATF) is a family of proteins that control the activities of hundreds of genes. Participating in this control is a coactivator, the CREB-binding protein (**CBP**),[382d] which is, in turn, subject to control by methylation.[383]

Control of growth. A large variety of transcription factors control cell growth. Some of these are indicated in Figs. 11-14 and 11-15 and Table 28-2. Since growth in cell numbers requires completion of the cell cycle, the specialized transcription factors involved are necessary. As is indicated in Fig. 11-5, factor E2F is of central importance. In fact, there are at least six mammalian E2F proteins, five of which have both a conserved DNA-binding domain and conserved activation domains.[383a,b] E2F1, 2, and 3 can all induce the S phase of the cycle. E2F6, in contrast, appears to bind to EF2-binding sites in DNA and also to Myc-binding sites to silence these genes and to help keep the cell in the G_0 state.[383c,d] As shown in Fig. 11-5, the

retinoblastoma protein Rb also binds to EF2 and represses transcription. However, it allows transcription when phosphorylated.[383e]

Among the many other proteins that influence growth are cFos and cJun, which may be activated by the MAP kinase pathway (Fig. 11-13).[383f,g] Binding of cytokines (pp. 1571, 1845) activates signaling pathways from cell membrane receptors to two other families of transcription factors. These are the **STATS** (p. 1845)[383h–j] and **SMADS**.[383h,k,l] Upon activation STATS and SMADS move from the cytosol into the nucleus where they find their binding sites on DNA.

Response elements. DNA binding sites for activator proteins are often described as "response elements."[47] Thus, the site for the cAMP-responsive protein CREB is the response element CRE. The binding site for AP-1 is **TRE**, named after the phorbol ester TPA. HSE is the heat shock response element, GRE the glucocorticoid response element (or the glucose response element), and **SRE** the sterol regulatory element.[384,384a] Response elements tend to be present in many enhancers and cooperate with other enhancer-binding proteins to activate groups of genes. An especially large group of genes respond to the **serum-response element** (also SRE), which is found within the *c-fos* promoter region.[384b,c] It is the DNA binding site for the ubiquitous serum protein SRF (serum response factor),[385–387] which is involved in growth control, cell cycle progression, and wound repair.

Zinc-containing transcription factors. The zinc finger domain (Fig. 5-37), which is also designated Krüppel-like finger in reference to a *Drosophila* protein,[388,388a] is a repeated motif present in many transcription factors (see Fig. 5-38).[389–391] In each finger a Zn^{2+} ion is coordinated by two –SH groups and two imidazole groups to form the Cys_2His_2Zn domain. One of the first proteins in which zinc fingers were recognized is TFIIIA, one of the factors that controls transcription of 5S RNA genes. A large 30-kDa N-terminal domain contains nine ~30-residue repeats of the sequence $XF/YXCX_{2-4}CX_3FX_{4-5}LX_{2-3}HX_{3-4}HX_{2-6}$ where X can be any amino acid. The fact that TFIIIA binds 7–11 Zn^{2+} ions per polypeptide chain suggested that the repeated sequences might be Zn^{2+}-binding domains in which each Zn^{2+} is held by two cysteine and two histidine side chains. Each of the nine Zn^{2+}-binding domains might constitute a metal-binding "finger" able to interact with about five bases in the DNA.[391–394] In agreement with this idea was the observation of an ~5-bp repeat of guanine clusters in the DNA.[395] The three-dimensional structures of a large fragment of the TFIIIA N-terminal domain (see title page banner for this chapter)[396] and of numerous other zinc finger proteins are known.

Yeast proteins often contain a pair of zinc fingers, but in the nematode *Caenorhabditis elegans* and in *Drosophila* there are more proteins with three or more zinc fingers.[391] The previously discussed Sp1 has three.[390] In *C. elegans* there are more than 100 genes that encode proteins with the Cys_2His_2 zinc-binding motif.[391] In addition, there are many proteins with four-cysteine zinc-binding motifs. These include Cys_4 Zn proteins of the **GATA family** of transcription factors, which are found in fungi, plants, and animals.[397,397a] GATA-1 is a specific transcription factor for regulation of erythroid genes. It binds to the consensus sequence 5'(T/A)GATA(A/G) found in globin genes.[398] In fungi members of the GATA family regulate nitrogen metabolism, biosynthesis of siderophores, and uptake of iron.[397,399] Another family (LIM) has $Cys_2HisCysZn$ domains.[391] A widely distributed motif in transcriptional repressors is a $Cys_3HisCys_4Zn_2$ or RING finger domain.[400]

If zinc-containing domains lose their Zn^{2+} they do not bind tightly to DNA. Regulation of the flow of zinc ions from storage sites in metallothioneins (Box 6-E) into transcription factors as well as into more than 300 enzymes poses interesting mechanistic questions.[401]

Winged helix transcription factors. Liver-specific expression of certain genes in rats depends upon **hepatocyte nuclear factor-3** (HNF-3). Related proteins are encoded by the **forkhead family** of genes in *Drosophila*. These proteins have characteristic C-terminal DNA-binding domains, each consisting of three helices, one of which fits into the major groove of DNA. Also present is a twisted three-strand β structure and two flexible loops or "wings."[402,403] The structure of the DNA-binding domain is similar to that of histone H5 and also resembles the HTH domains of prokaryotic repressors (Figs. 5-35, 28-3) and of CAP (Fig. 28-6).[404] There are many members of the HNF-3/Forkhead family of proteins and of the related **Ets-domain** transcription factors.[405–406a]

The NF-κB/Rel proteins. Nuclear factor NF-κB plays a crucial role in cellular immune responses and in inflammatory disease.[407–408c] This transcription factor was first recognized for its function in regulating transcription of the κ light chains of immunoglobulins. It is a member of the larger NF-κB/Rel family, which act in concert with a group of DNA-binding inhibitors of the IκB family.[409] The structure of an NF-κB dimer bound to its DNA target, whose consensus sequence is 5'-GGGRNYYYCC, is shown in Fig. 5-40. Its architecture[410] is quite unlike that of other transcription factors discussed in this book.

HMG proteins as transcription factors. The abundant high mobility group (HMG) nuclear proteins (Chapter 27) bind to DNA, some of them to four-

way junctions. The latter may be present in cruciform structures thought to play a role in regulation of transcription.[411] They are often regarded as modulators of chromatin structure. The ~80-residue HMG domain contains three helices and binds into a flattened, underwound, and bent DNA minor groove.[412,413] HMG proteins also act as transcription factors, which may interact directly with TBP, p53, steroid hormone receptors, and enhancers.[414] Cooperative binding with other DNA-binding proteins is characteristic of the effects of HMG proteins.[414a] Such interactions may be affected by acetylation. Members of the enhancer-binding HMG-14/-17 family undergo acetylation at seven specific sites.[415]

6. The Variety of Transcriptional Responses

Every protein has specialized functions, and specific regulatory mechanisms often control transcription of its genes. A cell must respond to a large number of stimuli, and responses often include activation or repression of transcription.[415a] In some cases an internal signal, such as a change in concentration of a nutrient or a key metabolite, provides the stimulus. In other cases an external stimulus such as heat or light is the inducer. A few examples follow. Others are mentioned throughout the book.

Many types of hormonal response, including those of insulin[415b,416] (Chapter 11), are transcriptionally mediated. This is also true for plants (Chapter 30). Defensive responses of both animals and plants (Chapter 31) are mediated in part by transcriptional responses.

Nuclear hormone receptors. Among the best known transcription factors are a large family of hormone receptors, which not only bind specific hormones but also contain in a central domain a pair of Cys_4Zn fingers that interact with response elements (discussed in Chapter 22).[416a–c] A subfamily of these binding proteins includes receptors for glucocorticoids, progesterone,[417] androgens, and mineralocorticoids. Another member binds the insect hormone 20-hydroxyecdysone (Fig. 22-12) and regulates the puffing seen in giant salivary gland chromosomes of *Drosophila*.[418,418a] A larger subfamily binds estrogens, vitamin D_3, thyroid hormone,[419] and retinoic acid.[420] The vitamin D receptor appears to serve also as a **bile acid sensor**.[420a] The same subfamily also includes many "orphan receptors."[421–424] The latter have been discovered by DNA sequence comparisons and have led through "endocrinology in reverse" to discovery of new hormonal signaling pathways.[425] The binding of a nuclear receptor to its response element in DNA is well illustrated by the estrogen receptor,[426–427b] which binds to a palindromic **estrogen response element** with the consensus sequence 5'-GGTCAnnnTCACC. It is regulated both by hormone-binding and by phosphorylation. The latter is catalyzed by a cyclinA-CDK2 complex in response to cell cycle alterations.[428] The glucocorticoid receptor protein forms a complex with a second protein, which has been identified as the 90-kDa chaperonin hsp90.[429] A **sterol regulatory element-binding protein** functions in a more general way to activate over 20 different genes that encode enzymes needed for synthesis of cholesterol and unsaturated fatty acids by animal cells.[429a,b] The steroid hormones often bind to their receptors in the cytosol, and the resulting complex is translocated into the nucleus (p. 1264).

Nutrient control. In addition to the sterols and fat-soluble vitamins, other dietary constituents are also recognized by transcriptional activators or repressors.[430] These include glucose,[431,432] amino acids,[433] phosphate ions,[434,435] and various metal ions. Best known among the latter is iron. The mammalian **iron response element** (IRE) is a hairpin loop RNA structure, which like the bacterial attenuator system (Fig. 28-9) functions posttranscriptionally.[436–438a] Iron regulatory proteins (IRP1 and IRP2), which contain Fe_4S_4 clusters, bind to the IRE sequences and inhibit translation. IRP1 is identical to cytosolic aconitase. A high intracellular iron concentration promotes assembly of the Fe_4S_4 cluster and binding to the IRE (see also Chapter 16). In the green algae *Chlamydomonas* a **copper response element** (GTAC) in DNA induces expression of genes important to copper uptake.[439] Many nutritional response systems have been recognized first in bacteria. For example, *E. coli* controls uptake of molybdate[127a,438,440] and of phosphate[437,441] as well as of sugars and ammonia (discussed in Section A,4).

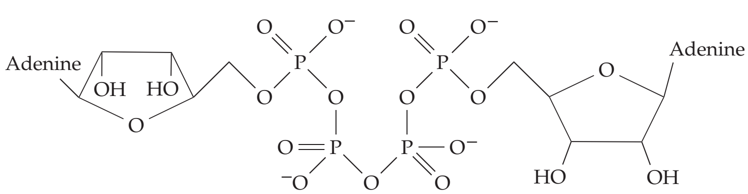
Stress proteins. Ritossa reported in 1962 that when *Drosophila* were suddenly warmed from 25°C to 36–37°C, a series of new puffs could be observed within the polytene chromosomes of the salivary glands.[442] These same puffs, which were also seen after other stresses, appeared within a few minutes and were associated with formation of new mRNAs. This **heat-shock** phenomenon was subsequently found to be universal. It is observed in all species of animals as well as in plants and bacteria.[443–448]

The principal **heat-shock proteins** (hsp), encoded by the new mRNAs, belong to five conserved classes: hsp100, hsp90, hsp70, hsp60, and small heat shock proteins. The function of some proteins as chaperonins has been discussed in Chapter 10. The *E. coli* chaperonin GroEL, a member of the hsp60 group, forms cylindrical aggregates with chambers in which proteins may fold (Box 7-A).[449] In a similar way a 16.5-kDa small heat-shock protein of *Methanococcus*

jannaschii forms 24-subunit hollow spheres with octahedral geometry.[450] The structure of the 68- to 70-kDa hsp 70 (dnaK) in *E. coli* has been conserved with high homology throughout evolution. Using DNA chip technology 77 heat-shock genes have been identified recently in *E. coli*.[451]

Induction of heat-shock proteins depends upon a heat-shock promoter element (**HSE**) that binds an activating transcription factor **HSF**.[452–455] An increase in temperature not only induces synthesis of heat-shock proteins but represses synthesis of most other proteins. Thus, in *E. coli* or *Salmonella* a shift from 30°C to 42°C causes the appearance of 13 heat-shock proteins. At 50°C synthesis of almost all other proteins stops. In *E. coli* transcription of heat-shock genes is controlled by alternative factors, σ^{32} and σ^{E}.[456,456a]

In *S. typhimurium* a series of unusual nucleotides such as P^1, P^4-diadenosine 5'-tetraphosphate (ApppA or Ap_4A) accumulate. The related compounds App-pGpp (with a 3' pyrophosphate), ApppG, AppG, and ApppA also accumulate and appear not only in bacteria but in eukaryotes as well.[457] They are formed as a side product in aminoacylation of transfer RNAs (Chapter 29).[458,459] Lee *et al*.[460] proposed that Ap_4A and related nucleotides are formed in response to oxidative stress and serve as **alarmones** that signal the need to reduce transcription of most genes and to increase transcription of genes for protective proteins. However, there is little correlation between the Ap_4A concentration and the heat shock response.[461] See also p. 1715.



P^1,P^4-Diadenosine 5'-tetraphosphate (A*pppp*A or A*P*$_4$A)

Another type of **stress response protein** is related to glutathione *S*-transferase.[462] Cells of *E. coli* also accumulate a 15.8-kDa **universal stress protein** in response to most types of stress. It is an autophosphorylating phosphoprotein, a member of a little-studied group of phosphoproteins normally present in very small amounts.[463] Cells also respond to various other types of stress such as deprivation of glucose, hypoxia,[464] ultraviolet irradiation, presence of hydrogen peroxide,[465,465a] or change in osmotic pressure.[466] Salt-tolerant plants synthesize new proteins in response to increased salinity of water.[467]

Responses to low oxygen tension in tissues (**hypoxia**) are important to all aerobic organisms.[464,467a–d] In mammals transcription of hypoxia-responsive genes is regulated by **hypoxia inducible factors** HIF-1 and HIF-2. A subunit of HIF-1 undergoes 2-oxoglutarate-dependent hydroxylation on proline and asparagine residues. This may be a step in induction of ubiquitination and destruction of this component of the transcription factor complex. The human **von Hippel-Lindau** (VHL) tumor suppressor is a ubiquitin E3 ligase, which is also present in this transcription factor complex.[467b,d–h] One subunit of HIF-1 also interacts with the tumor suppressor p53.[467i] Together with the VHL protein and the elongation factor elongin (p. 1637)[467j] HIF participates in controlling both production of red blood cells and growth of new blood vessels (angiogenesis, Chapter 32, Section D).[467k]

Light-induced transcription. Light has a strong effect on transcription, especially in plants and photosynthetic bacteria. The photosystem II subunits D1, D2, CP47, and CP43 (see Fig. 23-34) are encoded in the chloroplast genome. D1 and D2 are unstable in light, and their rate of synthesis is increased as a result of elevated levels of transcription that are induced by a blue light response.[468] The light-induced conversion of phytochrome to its far-red absorbing form **Pfr** (Eq. 23-42) causes increased transcription of a variety of plant genes.[469] See also Chapter 23.

Homeotic genes and homeodomain proteins. Geneticists discovered in *D. melanogaster* and other species genes that establish the placement of antennae and legs on particular segments and in general to specify the body plan.[470,471] These homeotic genes encode a series of proteins containing a 60-residue **homeodomain**, a DNA-binding domain of the helix–turn–helix class.[470,472] Some homeodomain-containing proteins bind to DNA containing the octameric sequence shown in Table 28-1 and are known as **octamer-binding transcription factors** (Oct). One of these (Oct-2) is specifically needed for activation of immunoglobulin genes,[473,474] while Oct-1 binds to promoters of various other genes including that of histone H2B, U1, and U2 snRNAs.[475,475a] Another transcription factor, **Pit-1**, which activates genes for growth hormone and for prolactin in the pituitary, binds to the same octamer.[476,476a] Homeotic genes are considered further in Chapter 32 and immunoglobulin genes in Chapter 31.

7. Transcription by RNA Polymerases I and III

Promoters for RNA Pol I, like those of Pol II, lie upstream of the initiation site for transcription. At least two transcription factors have been identified[47,477–478a] and vary among species. The human factors bind to a G•C-rich DNA sequence in the –45 to

+20 region and to a related upstream control element **UCE** at −180 to −107.

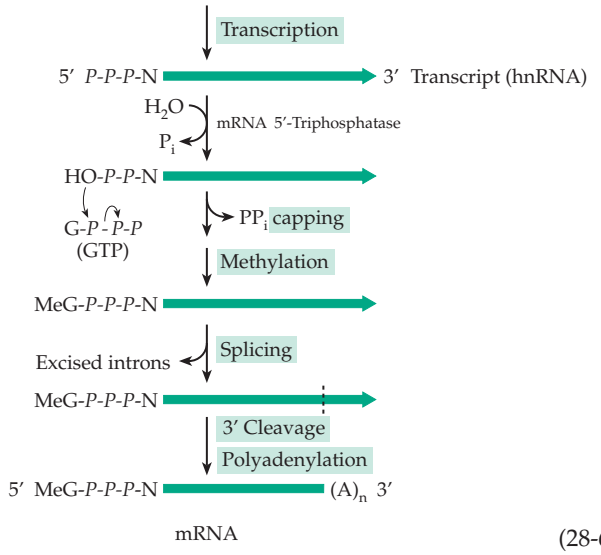
In vertebrate cells transcription by RNA pol III is controlled by three types of promoter: class 1 (5S RNA), class 2 (tRNAs), and class 3 (U6 snRNA).[47,479,480] Initiation of transcription of class 1 genes requires the 40-kDa **transcription factor TF IIIA**.[479] This was the first eukaryotic transcription factor to be purified and characterized,[481,482] and one of the first of the zinc finger proteins to be recognized. It was a surprise to find that TFIIIA does not bind to the promoter region but to a DNA sequence in the center of the 5S RNA gene between positions +55 and +80.[47] The presence of the TFIIIA binding site in the middle of the gene it controls suggests that TFIIIA interacts with other proteins that bind in the promoter region to form a loop. TFIIIA is involved in initiation of transcription of the ~24,000 oocyte type 5S RNA genes of *Xenopus*, but the ~400 somatic type 5S RNA genes are not activated in the same way. TFIIIA is also unusual in that it binds not only to DNA but also to a specific CCUGG sequence in the transcribed 5S rRNA. This RNA is stored as a 7S ribonucleoprotein particle until it is needed after the oocyte is fertilized and begins rapid protein synthesis.[483]

After TFIIIA binds, proteins TFIIIC and then TFIIIB also bind. Although promoters of classes 2 and 3 do not require TFIIIA, all three classes depend upon TFIIIB and TFIIIC.[484] The TATA-binding protein TBF is one of three components present in TFIIIB, which may be regarded as the true initiation factor.[47,485] Both TFIIIA and TFIIIC can be described as **assembly factors**.[47] A silkworm RNA pol III has been reported to require a transcription factor consisting of RNA.[486]

8. Elongation, Modification, and Termination of Transcription

As in prokaryotic transcription (Fig. 28-4) elongation by RNA polymerase II occurs within a transcription bubble of ~20–30 nucleotides in length.[487] Most transcriptionally active DNA is still in the form of nucleosomes, which must be unwound as the transcription bubble moves. Details are still uncertain.[269,488] All of the major steps in processing of the pre-mRNA transcripts, which include capping, splicing, 3'-end cleavage, and polyadenylation (Eq. 28-6), are coupled to transcription. This is apparently accomplished, in part, by physical connections of the necessary proteins to the CTD domain of RNA polymerase II.[304,312a,b] While pre-mRNA usually undergoes all of the steps of Eq. 28-6, rRNA and tRNAs are not capped or polyadenylated and often are not spliced.

Elongation of the RNA chain depends upon five **general elongation factors**, designated **P-TEFb**, **SII** (TFIIS), **TFIIF**, **Elongin** (SIII), and **ELL**.[489–492] Many of



(28-6)

these subunits of the transcriptosome function in the suppression of pausing in the uneven movement of the template DNA through the complex. Some may have to deal with torsional strain induced in the transcribed DNA.[493] Elongin is a heterotrimer of subunits A, B, and C.[494] Most of the other elongation factors are heterodimers,[489] but SII is a monomer that contains two conserved Zn^{2+}-binding **zinc ribbon** motifs.[492] It promotes cleavage of the growing transcript within the transcription bubble at stalled sites, allowing transcription to be restarted from a fresh 3' end. Defects in elongation factors P-TEFb, ELL, and **CSB** (Cockayne syndrome complementation group) have been correlated with human diseases including cancer.[495]

Despite the complexity of the processes represented by Eq. 28-6, a yeast cell is able to transcribe genes at rates of about one in every 6–8 s.[496] This can be compared with a rate of about once in 2–3 s for RNA polymerase of *E. coli*.

The least well defined step in eukaryotic transcription is termination, which follows the various steps of processing, discussed in Section D. The final 3' end of processed transcripts of RNA pol II action in mammals is marked by the sequence AAUAAA, which is found about 10–30 nucleotides upstream from the end.[497] This is usually followed by a polyU or GU-rich sequence.[498] In yeast the termination and polyadenylation signals are less clear.[499,500] The initial transcript almost always continues beyond the AAUAAA signal, sometimes for hundreds of nucleotides. However, the excess RNA is rapidly degraded by a large complex of proteins.[499–502] The precise 3' end cleavage is energy-dependent, requiring creatine phosphate rather than ATP or GTP.[500] Transcription termination by RNA polymerases I and III is more like that of bacteria. Terminator sequences are present in the DNA, and terminator proteins interact with them.[480,503–504b]

9. Conformational Properties of RNA

Newly formed RNA transcripts fold quickly into structures of complex shapes,[505–508b] the folding being influenced by interactions with proteins and with other RNA molecules. RNA chains are flexible, with many sterically allowed conformations.[509] As with proteins folding probably begins with a nucleation event, perhaps involving monovalent or divalent metal ions[510–512] and continues rapidly.[513] Folding is affected by hydrogen bonding,[514] base stacking, and binding of ions, and by formation of pseudoknots (Fig. 5-29).[515]

As is apparent from the structures of tRNAs (Figs. 5-30, 5-31, 28-20), the *Tetrahymena* self-splicing ribozyme (Fig. 12-26), and ribosomal RNA structures (Fig. 29-2), a large fraction of a folded RNA exists as hairpin or **stem–loop** structures. These are A-type structures with largely Watson–Crick base pairs. However, mismatched pairs, triples, and quadruples of bases are also formed. Recently discovered RNA structural elements include **base platforms**, formed by pairing of adjacent bases,[516] interdigitation of unpaired bases (also seen in DNA; Fig. 5-27), and wobble pairs (Chapter 5).[517] Hydrogen bonding between riboses of consecutive nucleotides in two strands may help to form a **ribose zipper**.[437,507] Guanine-rich tetraplexes (pp. 208, 227), cytosine-rich i-motif structures (p. 228), and water-mediated U•C base pairs also arise in RNAs.[517a–c]

The terminal loops, which usually contain the consensus sequence **GNRA**, may constitute up to one-third of the entire molecule.[518] These loops interact with many binding proteins, such as those in the snRNA-protein particles.[519] GNRA loops may also dock into the shallow groove of RNA helices.[518] Adenosines that are not paired in double helices, e.g., those in GNRA loops, are able to interact in a variety of ways with other parts of an RNA molecule or with other molecules. They are involved in helix packing interactions in virtually every RNA studied.[517d] Although examples are still rare, specific mRNA molecules may provide binding pockets for small regulatory molecules, e.g., amino acids and thiamin.[517e]

D. Processing of Eukaryotic RNAs

All RNA found in eukaryotes undergoes major alterations prior to functioning. The cutting out of rRNA and tRNA molecules from larger precursors resembles that in bacteria, but subsequent processing is much more complex, as is that of mRNA.

1. Ribosomal RNA

Eukaryotic ribosomes contain four pieces of RNA (Tables 5-4 and 29-1), which are usually designated by their sedimentation coefficients. The 18S, 5.8S, and 28S RNAs are encoded as single transcriptional units with spacers separating the sequences that encode the mature RNAs. A typical animal cell contains several hundred copies of this transcriptional unit, all located in the DNA in the nucleolus (Fig. 28-15), and each having its own set of promoter sequences, enhancers, and transcription factors.[47,520–522] The promoter sequences vary substantially among different species.[523] The primary transcripts from these units are the sole product of RNA polymerase I.

Electron micrographs of portions of unwound cores of nucleoli have revealed fibrils of RNA coated with protein growing from the DNA strands of the pre-rRNA genes (Fig. 28-16), ~80–100 RNA chains of different length being transcribed concurrently from a single gene. The overall gene length in the electron microscope is 2.3 μm, only a little less than the calculated length for a fully extended DNA molecule in

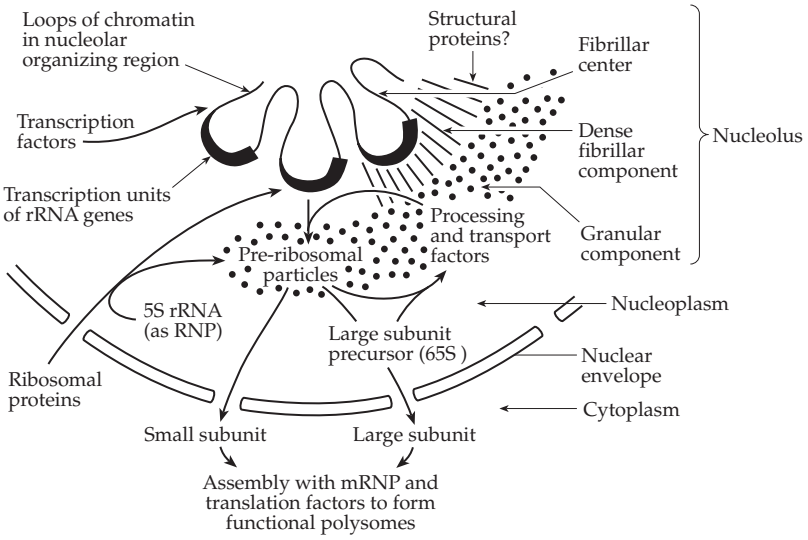


Figure 28-15 Steps in ribosome formation in the nucleolus. From Sommerville.[524] Abbreviations are NOR, nucleolar organizing regions; RNP, ribonucleoproteins.

the B form. However, judging by the lengths of the transcripts formed, the pre-rRNA chains are folded extensively.

The primary eukaryotic rRNA transcripts extend several hundred nucleotides past the 3' termini of the mature 28S RNA molecules. As the primary transcripts, which are formed in the core region of the nucleolus, move away into the outer cortex of the nucleolus, cleavage occurs in a number of steps (Eq. 28-7).[525,526] Electron microscopy provided the first direct confirmation of the relationship of one pre-rRNA molecule to another (Fig. 28-17).[527,528] The 18S portion of the 45S RNA seen in Fig. 28-17 lies nearest to the 5'-end just as does the 16S rRNA in the large transcript of the prokaryotic rRNA genes (Section A,7).

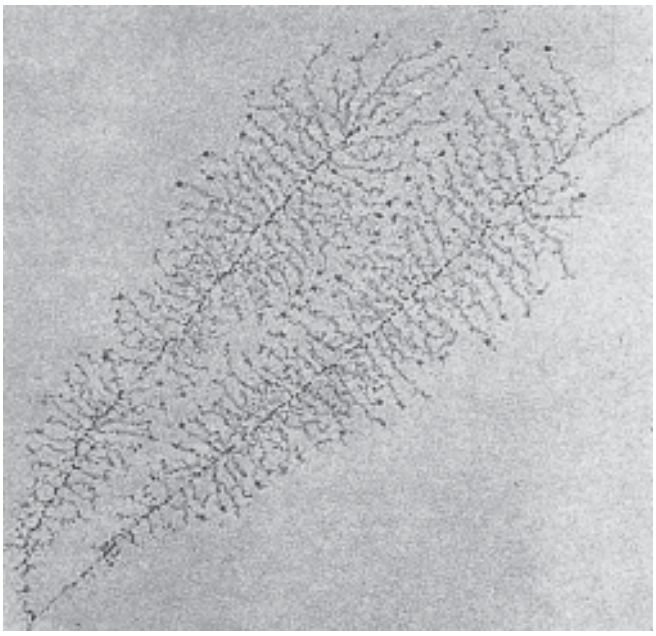
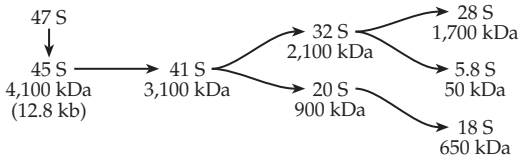


Figure 28-16 Ribosomal RNA genes from an embryo of *Drosophila melanogaster* in the process of transcription. The densely packed ribonucleoprotein strands are shorter where transcription begins and contain increasing lengths of rRNA where transcription has proceeded for a longer time. Also note the characteristic granular knobs at the tips of the strands. From Miller.[85]



(28-7)

As is indicated in Fig. 28-15, transcription is thought to occur from the loops of DNA that form the nucleolar organizing region. The 100-kDa **nucleolin**, the major protein of the nucleolus, binds to the nontranscribed spacer sequences in the DNA.[529–530] It also binds to the newly formed transcripts, as do various proteins that enter the nucleus from the cytoplasm.[524,531] More than 270 proteins, many of which participate in synthesis of ribosomes, have been detected in the nucleolus.[531a] Some of these proteins, acting together with the snoRNAs discussed in the next section, catalyze hydrolytic cleavage of the pre-rRNA molecules. For completion of pre-ribosomal particles additional protein molecules enter the nucleolus and associate with the pre-rRNA particles, then diffuse out of the nucleus.

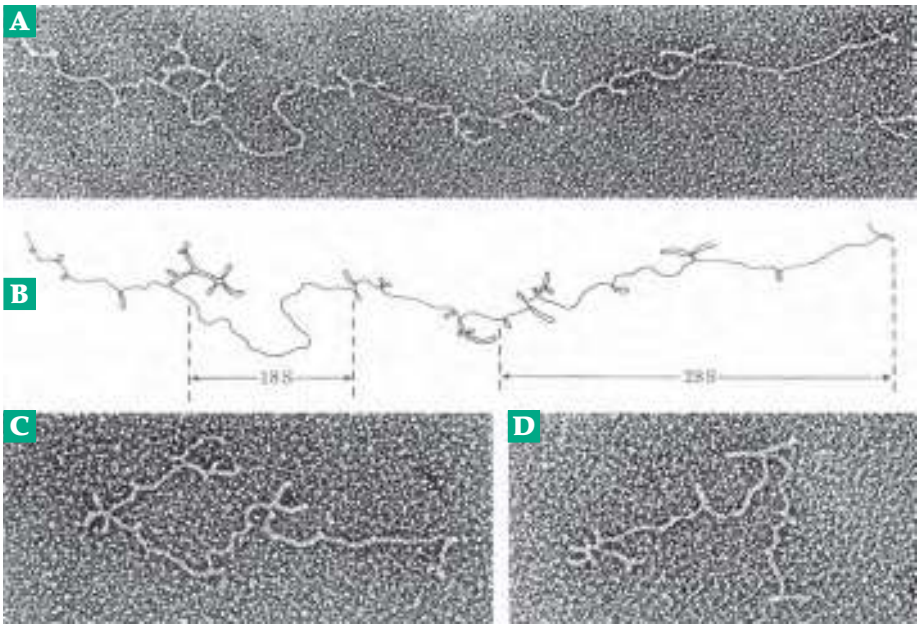


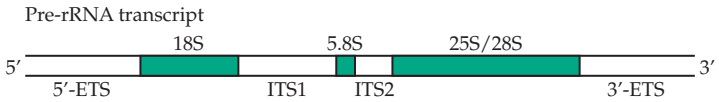
Figure 28-17 (A) Electron micrograph of the 45S precursor of rRNA from HeLa cells after spreading from 80% formamide and 4 M urea. The molecule is shown in reverse contrast. (B) Tracing of molecule in (A) showing several regions of secondary structure as hairpin loops. The 28S and 18S rRNA regions are indicated. (C) 32S rRNA. (D) 28S rRNA. Notice that the same secondary structure can be seen in the 28S RNA as in its 32S and 45S precursors. From Wellauer and Dawid.[527]

2. Small Nuclear, and Nucleolar, and Cytoplasmic RNAs

The nuclei of all eukaryotic cells contain a group of small nuclear RNAs (**snRNAs**), some of which (**snoRNAs**) are localized to the nucleolus and function there.[532–535e] At least ten of these are always present (Table 28-3), and yeast appears to contain more than 100. Some are present in small amounts, and it may be that a large number of snRNAs will also be found in other eukaryotes. The uridine-rich or U series of snRNA are especially abundant (10^4–10^6 molecules per nucleus). Several of them (U1,2,4–6,11,12,16,18) function in RNA splicing. Species U1–U10 contain from ~60 to ~215 nucleotides but some snRNAs from yeast have over 300 nucleotides and one over 1000.[532] The most abundant snRNAs, the metabolically stable U1, U2, U4, U5, and U6, are exported from the nucleus. In the cytoplasm each snRNA associates with a complex of several proteins to form ribonucleoprotein particles (**snRNP particles**). The proteins, known as **Sm proteins**, range in mass from 11 to 70 kDa[536] and are designated B, B′, D1, D2, D3, E, F, and G.[535,537,537a] The proteins may associate to a complex B_2D_2EFG, which then binds the snRNAs. Whereas many proteins are found in most of the snRNP particles, some are associated with specific RNAs. Thus, mature U1 snRNP may have a partial stoichiometry $U1A_2B_2C_2D_4E$.[536,538] The snRNA particle, known as U4 / U6snRNP, contains two snRNAs. After returning to the nucleus the four snRNP complexes U1, U2, U5, and U4 / U6, together with the pre-mRNA molecules, associate via an ordered assembly pathway that gives rise to the large **spliceosomes** in which the removal of introns takes place.[539] Some of the snRNP complexes, e.g. the spliceosomal U6 snRNP and snRNPs involved in mRNA degradation, contain Sm-like (Lsm) proteins.[537a]

More recently small 20–25 nucleotide antisense micro RNAs (**miRNAs**) have become recognized as important in control of the breakdown of mRNA, in formation of heterochromatin, and in control of specific stages in development (Chapter 32), and in cellular defense mechanisms (Chapter 31, Section G).[1a,537b] In animals the 21–23 nt miRNAs are formed by enzymatic cleavage of double-stranded dsRNA by a ribonuclease III-type enzyme called **Dicer**.[171a,537c] In the resulting fragments, which have 5'-phosphate termini, one strand is antisense to a "target" sequence in mRNA. **Small temporal RNAs** (stRNAs) that guide development are cut by Dicer from RNA stem-loop structures. In other cases, e.g., in formation of the siRNAs that silence individual genes, an RNA-dependent RNA polymerase acts on an RNA transcript to form a long dsRNA that can be cleaved by Dicer. This often happens with foreign RNAs, e.g., from viruses or trans-genes. Unwinding of the small fragments formed by Dicer, in a process that may require ATP, provides the single-stranded antisense siRNA molecules. Such antisense molecules bind to their target RNAs together with a group of proteins to form **RNA-induced silencing complexes** (RISCs). These complexes contain an RNAse (different from Dicer) that cuts the dsRNA of an siRNA-target complex, initiating destruction of the mRNA and silencing of its gene.[1a,171a,539a] This **RNA-interference** is widely utilized by plants and animals. It has also provided the basis for development of many practical tools for understanding gene sequences, for genetic engineering, and for design of new drugs.

Processing of ribosomal RNA. Transcripts of rRNA genes vary in size from ~35 to 47S (6–15 kb) and often contain spacer regions at both ends as well as between the 18S, 5.8S, and 28S sequences. For example, human 47S transcripts have 414 extra nucleotides at the 5' ends.[540] One group of snoRNAs participates in the hydrolytic cleavage of pre-rRNA. RNAs U3, U14, snR10, snR30, as well as MRP RNA (Table 28-3) are always required.[541–545] Also needed is U22, an intron-encoded RNA.[531] The reactions represented by Eq. 28-7 are best known for *Saccharomyces cerevisiae*.[525,541–541c] Four spacers, the **5'-external transcribed** spacer (5'-ETS), the first and second **internal transcribed spacers** (ITS1 and ITS2), and the 3'-external transcribed sequence (3'-ETS), must be removed.[525,541,546] Removal of the 5'-ETS depends upon the snoRNA U3, which contains two highly conserved sequences able to form base-paired structures with the 5' end of the 5.8S rRNA region of the pre-rRNA gene.[542–544,546a] Both U8 and U22 are also needed for cleavage of the pre-rRNA.[546b] Although the exact functions of snoRNAs and their associated proteins in the cleavage of pre-rRNA are still uncertain, these RNAs probably act as guide molecules for the cleavage reactions. They may have ribozyme activity and may perhaps be chaperones.



Cleavage at the 5' end of the 5.8S region requires RNase MRP, a relative of the RNase P that cleaves at the 5' ends of tRNAs (Fig. 28-10).[525,547] MRP (**mitochondrial processing protein**) also cleaves primers for mitochondrial DNA replication. The importance of the enzyme is emphasized by the existence of a hereditary defect in the MRP RNA (Table 28-3) that causes abnormalities in bone, cartilage, hair, and the immune system.[547a] Most bacterial rRNA genes have a tRNA gene in the position corresponding to that of 5.8S RNA

TABLE 28-3
Some Eukaryotic Small Nuclear and Nucleolar RNA Molecules

Designation	Number of nucleotides	
	Vertebrate	Yeast
U1[a]	164	568
U2[a]	188	1175
U3[b]	206 – 228	333
U4[a]	142 – 146	160
$U4_{atac}$[c]	131	
U5[a]	116 – 118	183 or 196
U6[a,d]	116	214
$U6_{atac}$[c]	125	
U7[e]	57 – 58	
U8[b]	136 – 140	
U9	130	
U10	60	
U11[c]	131 – 135	
U12[c]	150	
U13[b,f]	105	
U14[b,f]	87– 96	125 – 128
U18[b,f]	67– 70	102
X[b]	150	
U20[b,f]	80	
U21[b,f]	93	
U22(Y)[b]	125	
U24[b,f,g]	77	
U32 – U40[b,f,h]		
SnR10[b]		245
SnR30[b]		605
SnR38[f]		93
SnR39[f]		85
SnR40[f]		96
SnR41[f]		
MRP RNA (RNA 7– 2)	260 – 280	339

[a] Major spliceosomal RNAs.
[b] Fibrillarin-associated SnoRNAs that function in pre-ribosomal RNA processing. See Morrissey, J. P., and Tollervey, D. (1995) *Trends Biochem. Sci.* **20**, 78–82.
[c] Function in AT–AC spliceosomes. See Tarn, W.-Y., and Steitz, J. A. (1997) *Trends Biochem. Sci.* **22**, 132–137 and Fournier, M. J., and Maxwell, E. S. (1993) *Trends Biochem. Sci.* **18**, 131–135.
[d] γ-Monomethyl cap.
[e] Required for 3'-end formation in histone mRNAs.
[f] SnoRNAs with long complementarities to rRNA. C and D sequences are present. See Bachellerie, J.-P., Michot, B., Nicoloso, M., Balakin, A., Ni, J., and Fournier, M. J. (1995) *Trends Biochem. Sci.* **20**, 261–264 and Nicoloso, M., Qu, L.-H., Michot, B., and Bachellerie, J.-P. (1996) *J. Mol. Biol.* **260**, 178–195.
[g] Polyadenylated, noncoding.
[h] Participate in 2'-*O*-ribose methylation.

in eukaryotes. This provides an RNase P-dependent cleavage mechanism, which is alternative to action of other nucleases.[525]

The most abundant protein in the fibrillar regions of the nucleus, where the early stages of pre-rRNA processing occur, is **fibrillarin**.[541,548,549] Many of the snoRNAs are closely associated with this protein. Fibrillarin is also well known as an autoantigen, which can induce formation of destructive antibodies that cause **scleroderma** (Chapter 31, Section F).

Modification guide RNAs. A second group of snoRNAs function in methylation, pseudouridine formation, and other RNA modifications (Section 6).[174,525,541,548] These snoRNAs have long sequences complementary to highly conserved regions of pre-rRNA, enabling them to form helical regions that may guide the docking with modification enzymes. Many of them also contain characteristic conserved sequences: **C**, 5'-UGAUGA; **D**, 5'-CUGA; **H**, 5'-AnAnnA; and 5'-ACA. Sequences C and D are present in snoRNAs that act as methylation guides, while the H and ACA sequences characterize guide RNAs for pseudouridine formation.[174]

Transcription and processing of snRNAs and snoRNAs. In higher organisms each of the snRNAs has several genes,[538] e.g., there are 50 –60 U1 genes in the human haploid genome. However, in yeast there are often single copies.[541] All of the snRNA genes, except for that of U6,[550] are transcribed by RNA polymerase II. The transcripts, which are capped at the 5'-end but are not polyadenylated, pass into the cytoplasm, where they undergo further processing and become associated with proteins. The 3' ends are trimmed, the 7-methylguanosine of the cap is methylated further, and methylation may occur on other bases as well.[551] SnoRNAs are not capped. It was a great surprise to discover that many of the snoRNA genes lie within introns that occur in abundantly expressed genes for functionally unrelated proteins.[173,541,548]

Patients with the autoimmune disease systemic **lupus erythematosus** make autoantibodies directed against the Sm proteins of snRNP particles.[552,553] Antibodies from different patients vary in their specificities, making these antibodies a useful tool in the isolation and study of snRNAs and their protein complexes.[552,554]

3. Processing of 5S RNA and tRNAs

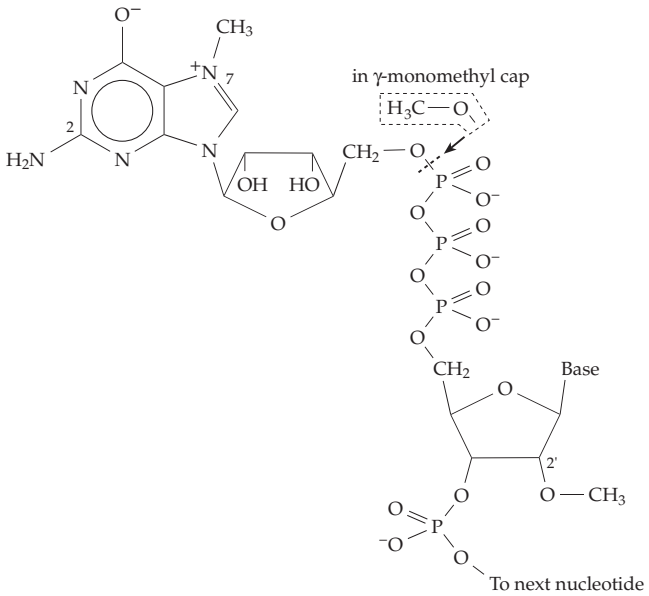
The genes for 5S ribosomal RNA and all of the tRNAs are transcribed by RNA polymerase III. In the yeast genome the 5S RNA genes are located in the spacers between the transcriptional units containing the other rRNAs. However, in animals the 5S RNA

genes are separate from the other rRNA genes and are not located in the nucleolus. In *Drosophila* ~500 copies of the 5S RNA gene are located in the right arm of chromosome 2. In *Xenopus* ~400 genes are active in somatic cells and another 24,000 only in oocytes. These are arranged as large transcriptional units, each containing several thousand copies of the 120-bp 5S RNA gene separated by 720-bp spacers. Cleavage and trimming are required to form the final product, but there are usually no introns to be removed.

Most eukaryotic tRNAs are formed from monomeric precursors, each gene acting as a transcriptional unit. Processing is similar to that in bacteria (Section A, 7). Eukaryotic RNase P usually cleaves the 5' end,[555–558c] and another enzyme cuts at the 3' end.[542,556] The 3' CCA sequence of the mature tRNA is usually not present in the primary transcript but is added.[559,559a,b] As in bacteria (p. 1620) extensive modification of bases also occurs in the tRNA precursors of eukaryotes.[235,560–562] Many tRNA genes contain introns, which must be removed by splicing (Section 5).

4. Messenger RNA, Caps, and Polyadenylate Tails

The first processing event (Eq. 28-6) for most of the pre-mRNA and snRNA transcripts made by RNA polymerase II is addition to the 5' end of a "cap," a terminal structure containing 7-methylguanosine from which a proton has dissociated to form a dipolar ion.[563–565] The cap structure may be abbreviated 5′-m^{7}G(5')pppNm —. The 5' terminal ribose is often methylated on O2', as shown below. More complex caps are methylated at additional sites, e.g., the guanine may be dimethylated on the 2-NH_2 group.[551] Most snRNAs, including the U1–U5 and U7–U13 snRNAs, have such 2,2,7-trimethylguanosine



caps.[551,566] Many viral transcripts, including those of the much-studied vaccinia virus, have similar caps.[564] However, U6 and some other snRNAs, which are transcribed by RNA polymerase III, have γ-monomethyl (me-ppp) caps[567] or undergo a series of additions and deletions of uridylate residues at the 3' ends.[566,568]

Cap synthesis occurs as follows. The 5' end of an RNA transcript initially contains a triphosphate group arising from the fact that a nucleotide triphosphate serves as the primer in initiating transcription. The terminal phospho group is removed by a triphosphatase leaving a diphosphate, which is then guanylated by GTP (Eq. 28-6).[569] The capped transcripts are exported from the nucleus, after which additional methylation may follow.[570]

The cap structure affects several processes.[565,571] A family of cap-binding proteins recognize the structure and may facilitate splicing as well as export from the nucleus.[571a] The cap is very important for ribosome binding and initiation of translation (Chapter 29). The trimethylated caps of snRNAs, on the other hand, may be signals for retention in the nucleus where they function.[563] Following capping is the often elaborate process of splicing to remove introns (Section 5).

A poly(A) "tail" consisting of ~250 residues of adenylic acid is added next by poly(A) polymerase, a component of an enzyme complex that also cleaves the RNA chains.[545,571b] Most eukaryotic mRNA is polyadenylated with the exception of that encoding histones. The function of the poly(A) is unclear. It is needed for transport of mRNA out of the nucleus, but it does confer a greatly increased stability to the mRNA in the cytoplasm where the adenylate units are gradually removed.[307,308] In contrast, in chloroplasts and plant mitochondria polyadenylation is required for rapid degradation of mRNA.[571c,d] Polyadenylation may also increase the efficiency of translation.[572] Polyadenylation occurs rapidly within ~1 min after transcription is completed.

Mature mRNA molecules vary in lifetime.[573,574] Some last for hours or days. Among the latter are mRNAs of maternal origin that accumulate in oocytes and are utilized during the early stages of embryonic development.[575,575a,b] Other mRNAs, e.g., transcripts of the *c-fos* and *c-myc* proto-oncogene products, have half-lives of 30 min.[573,576] Some mRNA molecules are degraded while attached to ribosomes in response to recognition of the synthesized peptide (Chapter 29). Longer lived mRNA molecules may be protected by RNase inhibitors.[577] Hydrolytic removal of caps often initiates degradation, and Sm-like protein complexes participate.[578–579a]

5. Splicing

An essential modification of the precursor forms

of large rRNA, tRNA, and mRNA molecules is the splicing out of intervening sequences. This occurs prior to polyadenylation of mRNA and is usually slow, the half-life of introns varying from a few seconds to 10–20 min.[551] Splicing occurs by at least four distinctly different pathways.[47,580]

Self-splicing RNA. The precursor to the 26S rRNA of *Tetrahymena* contains a 413-nucleotide intron, which was shown by Cech and coworkers to be self-splicing, i.e., not to require a protein catalyst for maturation.[581,582] This pre-rRNA is a ribozyme with true catalytic properties (Chapter 12). It folds into a complex three-dimensional structure which provides a binding site for free guanosine whose 3'–OH attacks the phosphorus at the 5' end of the intron as shown in Fig. 28-18A, step *a*. The reaction is a simple displacement on phosphorus, a transesterification similar to that in the first step of pancreatic ribonuclease action (Eq. 12-25). The resulting free 3'–OH then attacks the phosphorus atom at the other end of the intron (step *b*) to accomplish the splicing and to release the intron as a linear polynucleotide. The excised intron undergoes a third transesterification reaction, of uncertain significance (step *c*), to form a circular polynucleotide and a short displaced 15-residue oligonucleotide. The *Tetrahymena* pre-rRNA intron is a member of a group of similar **Group I introns**, many of which are found in fungal mitochondrial pre-mRNA and pre-rRNA.[583] All are excised by a similar mechanism. Many are self-splicing, but others require a protein catalyst.[584–585a] A similar splicing sequence is involved in removal of a 1017-nucleotide intron from the thymidylate kinase gene of phage T4 and other introns in T-even phage. The later are among the relatively rare introns in prokaryotic systems.[583]

All Group I introns have several small conserved sequences, which suggest a common folded tertiary structure as is indicated in Fig. 28-19A. The conserved sequences are labeled **A**, **B**, **9L**, **2**, **9R**, and **9R′**. **A** is paired with **B**, **9R** with **9R′,** and **9L** with **2**. The sites of chain cleavage at the 3' and 5' ends of the intron are indicated by the heavy arrows. They are evidently selected by formation of the double-stranded regions.[586–588]

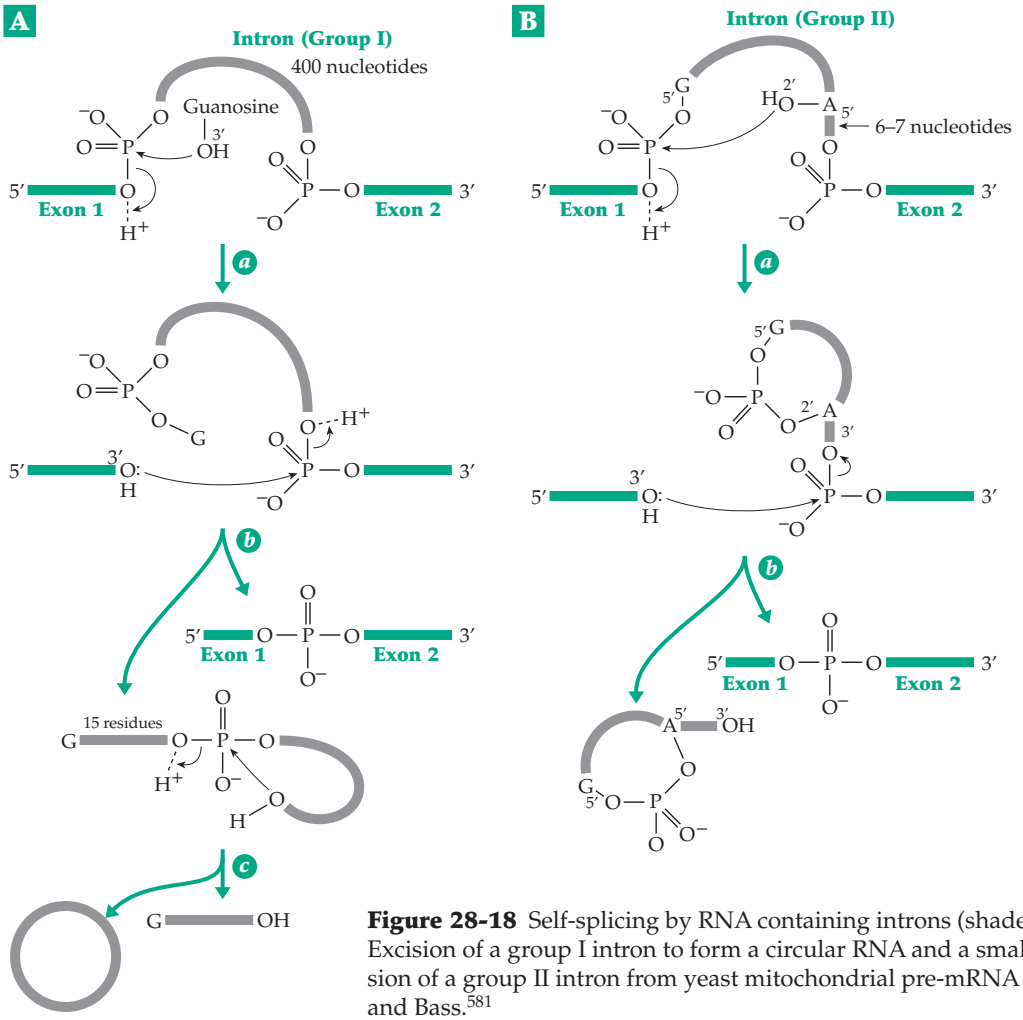


Figure 28-18 Self-splicing by RNA containing introns (shaded) of groups I or II. (A) Excision of a group I intron to form a circular RNA and a small oligonucleotide. (B) Excision of a group II intron from yeast mitochondrial pre-mRNA as a circular RNA. See Cech and Bass.[581]

Another type of intron (Group II) also undergoes self-splicing.[589–590d] The best known example is the last intron in the yeast mitochondrial pre-mRNA. The splicing pathway shown in Fig. 28-18B is similar chemically to that of the group I introns. However, the initial attack is not by free guanosine, but by the 2' OH of an internal adenosine, the intermediate product having a **lariat structure** with a loop at the end. Otherwise, processing is similar to that of group I introns. The same pathway with lariat formation is followed by the more widely used removal of introns from pre-mRNA in spliceosomes (see Fig. 28-22).

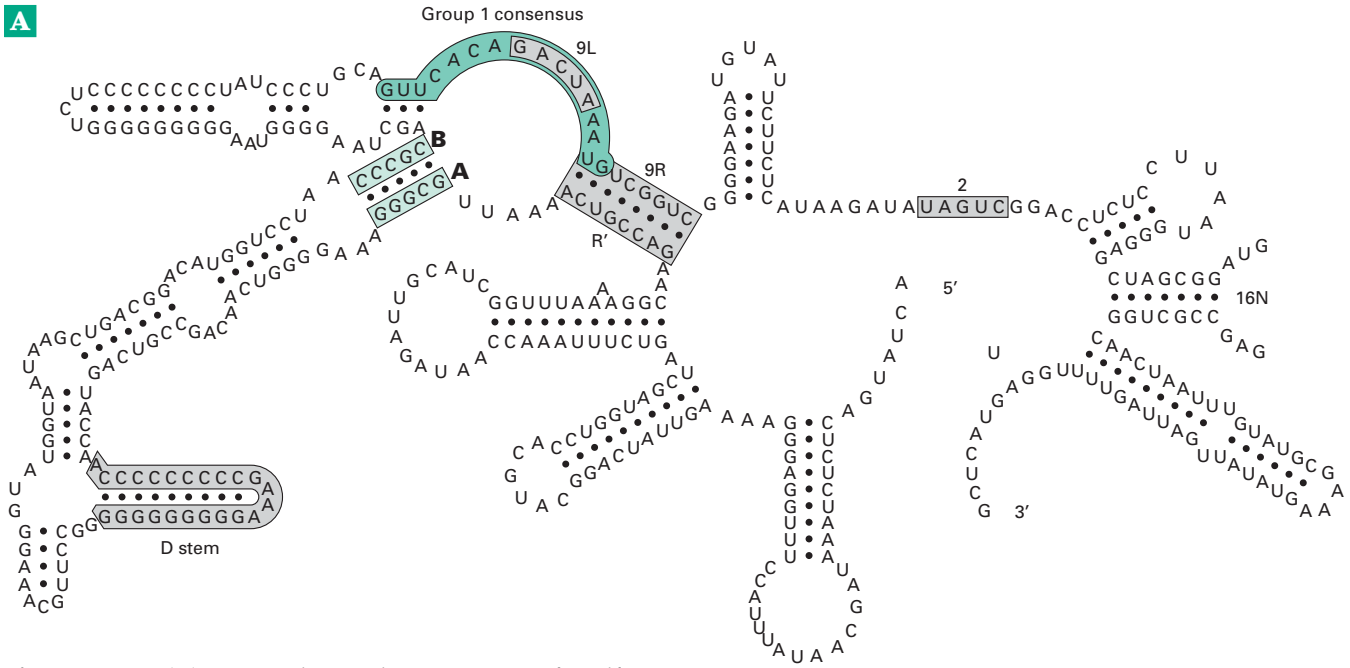
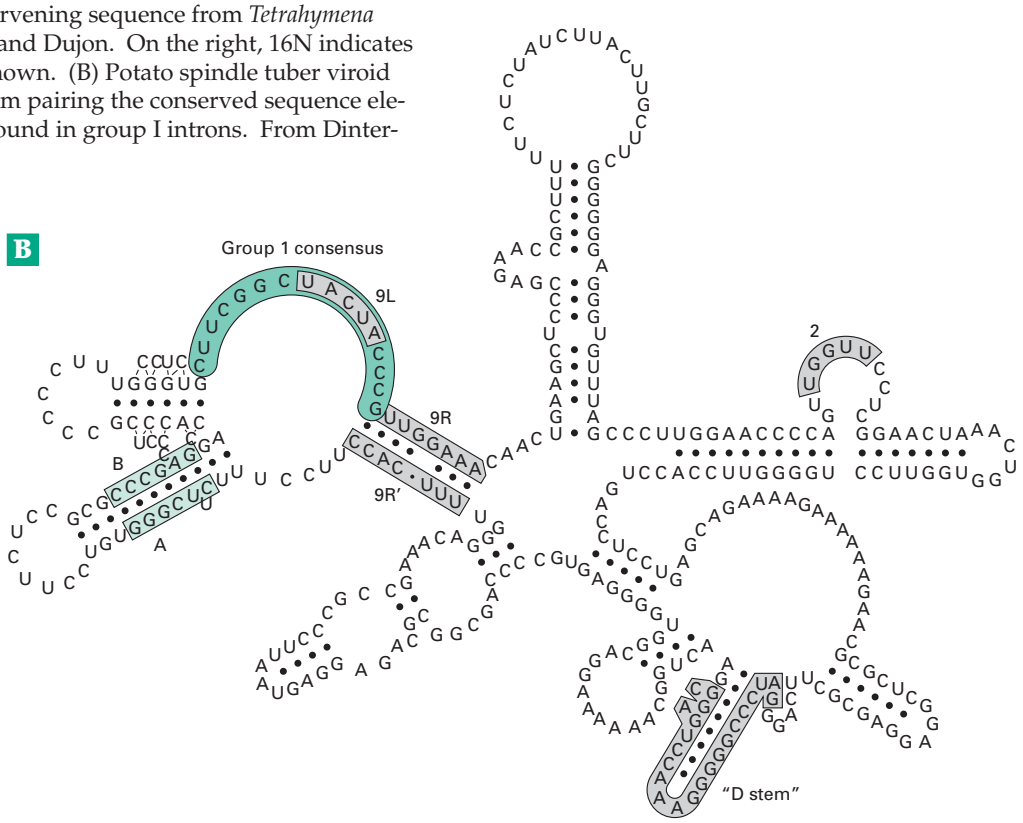


Figure 28-19 (A) Proposed secondary structures of a self-splicing Group I intervening sequence from *Tetrahymena* according to Michel and Dujon. On the right, 16N indicates 16 nucleotides not shown. (B) Potato spindle tuber viroid structure derived from pairing the conserved sequence elements that are also found in group I introns. From Dinter-Gottlieb.[596]



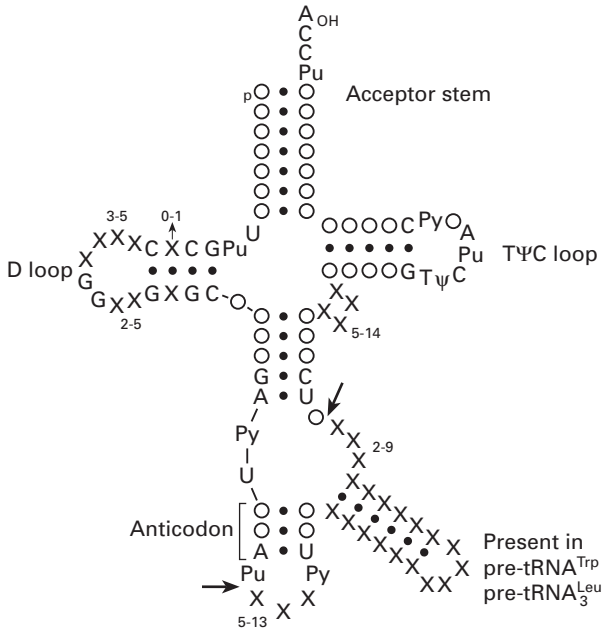


Figure 28-20 Composite structure representing several tRNA precursors arranged in a similar secondary structure (see also Fig. 5-30). The arrows indicate splice points. Variable positions are designated (○) for the mature tRNA and (X) for the intervening sequence and also in loops where insertions or deletions occur. From Ogden *et al.*[603]

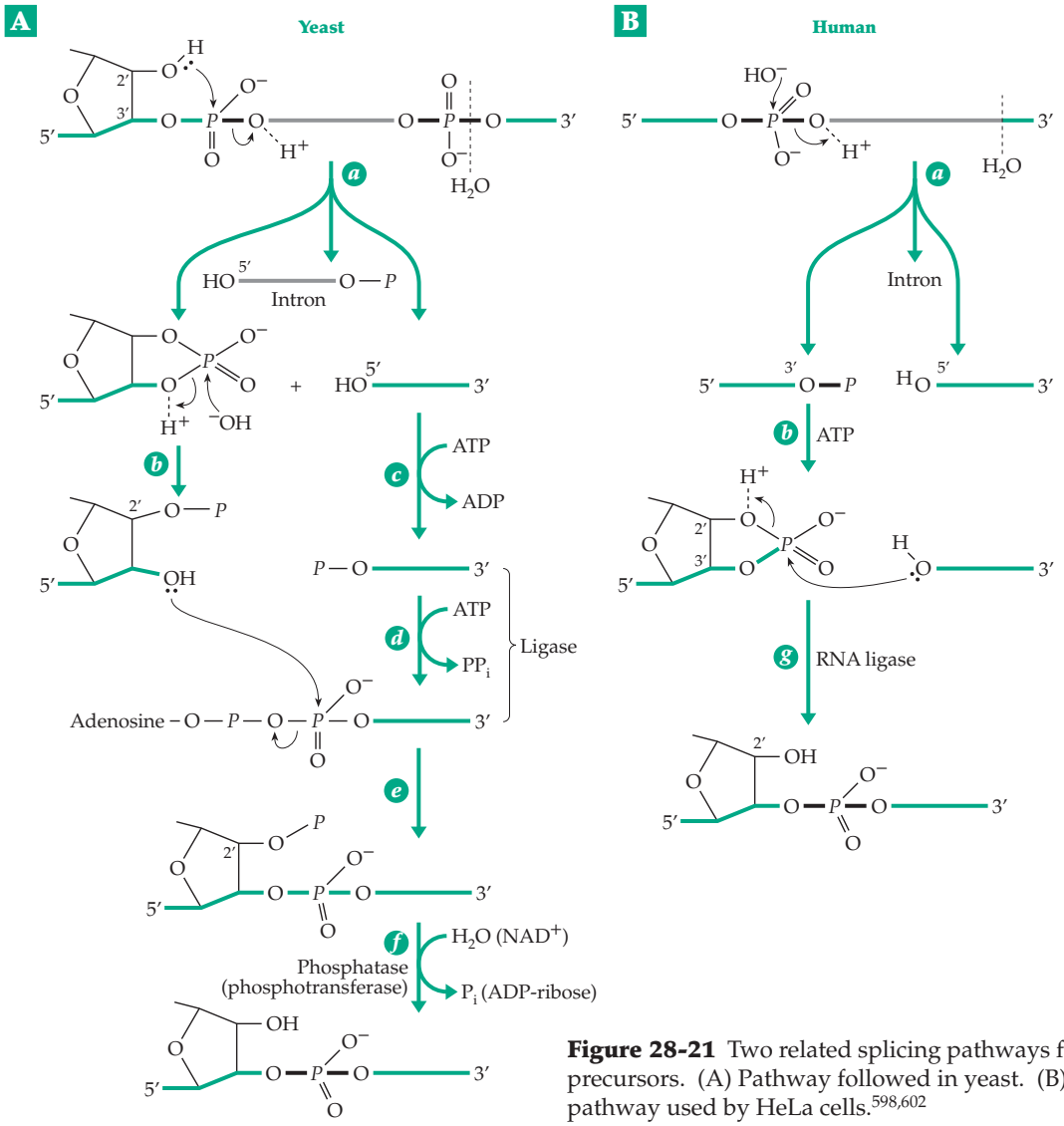


Figure 28-21 Two related splicing pathways for tRNA precursors. (A) Pathway followed in yeast. (B) Related pathway used by HeLa cells.[598,602]

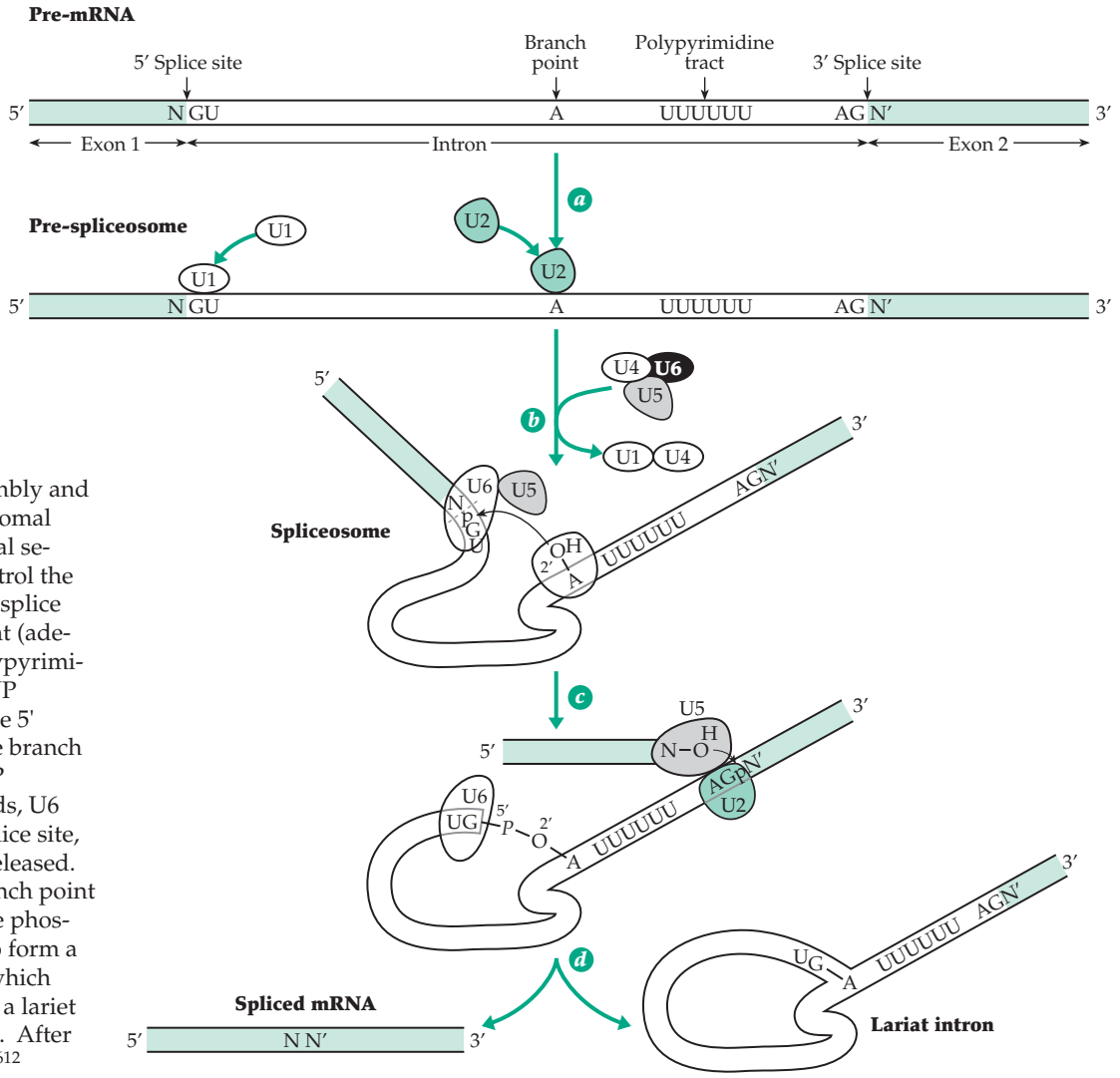


Figure 28-22 Assembly and action of the spliceosomal complex. Four special sequence elements control the process: the 5' and 3' splice sites, the branch point (adenosine A), and a polypyrimidine tract. The snRNP particle U1 locates the 5' splice site and U2 the branch point. The tri-snRNP U4•U6•U5 then binds, U6 recognizing the 5' splice site, and U1 and U4 are released. The 2'-OH of the branch point adenosine attacks the phosphodiester linkage to form a lariet intermediate, which releases the intron in a lariet form in the final step. After Valcárcel and Green.[612]

Relationship to viroids and virusoids. The smallest of viruses are the naked ~250–460 nucleotide single-stranded circular RNA molecules called viroids (Chapter 5).[591–593] Viroids are closely related to group I introns. The conserved paired sequence characteristic of the group I introns are also present in viroids (Fig. 28-19B), and it looks as if viroids may be "escaped introns." Another group of "plant satellite RNAs" or virusoids replicate only with the help of larger RNA viruses. These satellite RNAs are replicated by a rolling-circle mechanism.[593a] The resulting long RNA molecules are self-cleaving, cutting themselves to form the unit length satellite RNAs.[594,595] These are the simplest known self-cleaving RNA molecules. They have been discussed in Chapter 12.

Pre-tRNAs. In the removal of type I introns the formation of specific stem and loop structures directs the splicing reactions (Fig. 28-18).[47,597] Stems and loop structures already exist in tRNA precursors. Cleavage sites are usually located just to the 3' side of the anticodon as is seen in the pre-tRNA molecule and in the composite structure shown in Fig. 28-20. The chemistry of the splicing process[205,598–601] is shown in Fig. 28-21. Cleavage at the two splice sites excises the intron. Then the two ends are rejoined. In yeast cleavage at the 5' end of the intron is by a multimeric enzyme with a pancreatic ribonuclease-type of action that leaves a 2',3'-cyclic phosphodiester (Fig. 28-21A, step *a*) that is opened hydrolytically (step *b*). Another hydrolytic cleavage, catalyzed by a different subunit of the enzyme complex, occurs at the 3' end of the intron. The resulting 5' OH is phosphorylated by GTP (step *c*),[601] and the resulting 5'-phospho group reacts with ATP to displace PP_i and to form a transient adenosine-5'-diphosphate terminus (step *d*). This reacts with the 3' OH formed at the 5' splice site (step *e*) to create a phosphodiester linkage between the two pieces of RNA. A phosphotransferase then removes the superfluous 2' phosphate (step *f*).[601a] Steps *d* and *e* are catalyzed by an **RNA ligase**.[598,599]

In HeLa cells and presumably in normal mammalian nuclei, the initial cleavage at the 5' end of the intron leaves a 3' phospho group (Fig. 28-18B, step *a*), which is cyclized, probably in an ATP-dependent process, to the 2',3'-cyclic phosphate (step *b*). This is ligated to the other piece of RNA by a direct displacement on the cyclic phospho group (step *c*).[602]

The spliceosome. The hnRNA of nuclei, which includes all of the pre-mRNA, is associated with proteins, which sometimes form very large 200S particles.[604] After limited cleavage with nucleases they tend to sediment in the 30S–40S range and to contain a variety of proteins.[605,606] Some of the proteins may have been involved in control of transcription.[606] Others participate in splicing. The smaller snRNP particles then appear to come into the nucleus and displace much, but not all, of the protein present in the pre-mRNA ribonucleoprotein particles.[605]

The 50–60S spliceosome complexes, with their protein and RNA components, are reminiscent of ribosomes. Although smaller in size, they can be vizualized by electron microscopy.[607] Each spliceosome is constructed from the four snRNP particles and additional proteins (Fig. 28-22).[539,608–612] Each spliceosome can accommodate ~500 nucleotides of pre-mRNA. As the pre-mRNA moves through the spliceosome the splice sites, which have only weakly conserved consensus sequences, must be located. The two ends of the introns, which may be much longer than the average 137-nucleotide exon, must be brought close together in the spliceosome.[613] The exact splice sites are usually located by an invariant **GU** at the 5' end and **AG** at the 3' end. The sequence of the first 18 nucleotides of U1 snRNA is largely complementary to that of the 5' splice site, which has the consensus sequence AG:**GU**RAGU, where the colon marks the junction. The **GU** is invariant.

After the U1 snRNP binds to the pre-mRNA (step *a*, Fig. 28-22)[614] the U2 snRNP binds to another almost invariant sequence CUR**A**CU found 20 to 55 nucleotides upstream of the 3' junction.[608,615–617] The A in this sequence becomes a branch point. It is brought close to the 5' splice site with the aid of a preassembled complex of snRNPs U4, U6, and U5. In this complex U4 and U6 are tightly paired, additional proteins are also present,[618–621] and enhancers may be located in adjacent exons.[617] Upon binding of U6 to the 5' splice site, the U1 and U4 snRNPs are released (step *b*, Fig. 28-22) and the 2'-OH of the branch point adenosine attacks the backbone phosphorus atom (step *c*) at the 5' splice junction forming a lariat intermediate. The 3' end created at the 5' junction must now be held and brought close to the 3' splice junction, which is located with the aid of U5 snRNP.[622] The 3' splice junction, utilized in the second splicing step (step *d*, Fig. 28-22) has the consensus sequence (T/C)N(C/T)AG:G.

The first splicing step is dependent upon a divalent metal ion, but the second is not.[623] Both steps appear to be in-line nucleophilic displacement reactions.[624] Additional **splicing factors** are needed for formation of the U4, U6, U5 complex and its function in the second splicing step, which also appears to require ATP.[621,622]

A small fraction of eukaryotic mRNA introns are characterized by **AU** and **AC** (rather than GU and AG) ends. The spliceosomes that act on these introns contain modified snRNAs U4 and U6, which are designated $U4_{atac}$ and $U6_{atac}$. They also require U11 and U12 snRNPs.[534,625,626]

Since pre-mRNAs usually contain many introns, a series of splicing events must occur. These apparently take place consecutively beginning at the 5' end. Similar splicing pathways are followed in yeast, higher plants, insects,[627] and mammals.

Alternative splicing pathways. RNA that contains many introns can undergo splicing in more than one way. Many examples of alternative splicing have been discovered.[612] The mammalian isoenzyme forms of pyruvate kinase called M_1, M_2, L, and R are all tetramers of 60-kDa subunits. The M_1 and M_2 forms are encoded by a single gene. The two mRNAs contain 1593-nucleotide coding regions, which are identical except for a 160-nucleotide sequence that determines the amino acid sequence in a region responsible for intersubunit contact. The difference between the M_1 and M_2 forms involves a choice of two alternative exon regions, one or the other of which is omitted during splicing.[628] Other examples have been found in human collagen,[629] in fibronectin,[630] in neuropeptide formation (Chapter 30),[631] and among human proline-rich salivary proteins,[632] cytoskeletal tropomyosin,[633] platelet-derived growth factor,[634] coagulation factor X,[635] and porphobilinogen deaminase.[636] Alternative splicing is very common in transcripts of viral DNA (Section E).

Alternative splicing could have arisen accidentally, but it is controlled by proteins. Best known is the **alternative splicing factor** (**ASF** or SF2).[612,637] It was first recognized by its function in *Drosophila melanogaster*, where the sex of individuals is determined by alternative splicing of an mRNA.[638] In addition to ASF other serine- and arginine-rich **SR proteins** participate in alternative splice site selection.[612]

Trans splicing. Every mRNA in trypanosomes has, at its 5' end, a short 35-nucleotide sequence that is not encoded in the transcribed gene. It was found that for each mRNA molecule two transcripts are formed and are spliced together, always with the 5' piece from a short SL (spliced leader) gene being joined to each of the others.[639] This trans splicing has since been observed in many plants, animals, and protists.[640]

Among these are the nematode *Caenorhabditis*,[639,641] flatworms,[642] yeast,[643] plant chloroplasts and mitochondria,[640,644] and mammalian cells.[645]

6. Modification and Editing of RNAs

Both mRNA and rRNA undergo rapid methylation of selected residues. About 1–3 internal adenylate residues per kb are methylated at their N^6 positions.[646] These are usually the central adenylates in the sequences GAC or AAC. The methylated sites are not uniformly distributed but are clustered, sometimes in the 3' untranslated ends of the RNA. Many more residues (55 in yeast and ~100 in vertebrates) in rRNA are methylated on selected ribose 2'-OH groups.[173,535c–e,647–648a] About an equal number of uridine residues are converted to pseudouridines (Eq. 28-3). Methylation sites are apparently selected by the fibrillarin-associated snoRNAs U32–U40 (Table 28-3). These **methylation-guide** snoRNAs contain 10- to 14-nucleotide sequences that are complementary to segments of the rRNA that contain the methylation sites, and evidently provide rigid helical regions that are targets for the methylase action. The snoRNAs contain the previously mentioned C and D sequences. Methylation occurs on the ribose of the nucleotide that is base-paired with the fifth nucleotide that is upstream of the D sequence in the snoRNA. Each snoRNA directs methylation of a different ribose.[173] In a similar manner the snoRNAs containing the ACA motif located three nucleotides upstream of their 3' ends appear to direct the conversion of uridines to **pseudouridines**, as in Eq. 28-3.

Editing of RNA transcripts. Interpretation of the genetic code utilized by the mitochondria of trypanosomes and other kinetoplastid protozoa was confounded by the discovery that the DNA sequences of many genes, including the COIII cytochrome oxidase gene present in the kinetoplast maxi circle DNA, do not appear to encode the correct amino acid sequence. In fact, the RNA transcripts are extensively edited, mostly by insertion of multiple uridine residues at many positions[649,650] and by occasional deletions at others. Editing of some mRNAs causes 45% of the message to be rewritten.[651] Several additional kinds of editing were soon discovered, not only in protozoa but also in mammals, plants, and archaea.[652,652a]

A frequent editing change is the hydrolytic deamination of a C to a U residue.[652b] For example, human apolipoprotein B is synthesized in two forms: apoB100, a full-length 512-kDa protein made in the liver and used for transport of cholesterol and triglycerides, and a shorter 241-kDa form, apoB48, used in absorption of dietary lipids (Chapter 21, Section A1). ApoB100 is synthesized from a full-length mRNA, but apoB48 is made according to a shortened mRNA in which a glutamine codon (CAA) has been converted by editing to the translation stop codon UAA.[652,653] A special enzyme deaminates only cytidine 6666 of the mRNA.[654] C to U editing occurs in chloroplasts and mitochondria of plants.[655–657] In *Arabidopsis* mitochondria 456 different C to U conversions have been identified in mRNAs.[656]

Deamination of adenosine residues produces inosine, which occurs in brain mRNA once in ~17,000 ribonucleotides.[658] Some ionotropic glutamate receptors in the brain have subunits translated from inosine-containing mRNAs. A glutamine codon CAG is edited to CIG, an arginine codon. The arginine codon AGA is converted to the glycine codon IGA at another site, ATT is converted to ITT and TAC to TIC. All of these changes affect the properties of the glutamate-activated ion channel.[659] The adenosine deaminases involved in these editing events are usually specific for double-stranded helical segments of RNA, e.g., for the stems in stem–loop structures.[660–662] At least one human adenosine deaminase not only binds to RNA but also contains a DNA-binding domain specific for Z-DNA.[663]

Returning to the trypanosomes and their relatives, mitochondrial RNAs undergo extensive insertion and deletion of U's. The editing site is located by a **guide RNA** (gRNA), which directs the hydrolytic cleavage of the chain and either the addition of U's to the 3' cut end by transfer from UTP or hydrolytic deletion of U's from the 3' cut end. The chain ends are then rejoined by an RNA ligase.[172,664–667c] The functional significance of the editing of kinetoplast mRNA is uncertain. However, at least some of the edited mRNA is translated to give proteins that are presumably used.[668] While trypanosomes usually insert only U's, the slime mold *Physarum polycephalum* may insert dinucleotides such as AA, AU, CU, or GU[669] and may also add nucleotides at the 3'-ends of RNAs.[669a]

A to G editing occurs in RNA of *Drosophila*.[670] Yet another type of editing has been observed in viral RNA from paramyxovirus. The virally encoded RNA polymerase sometimes "stutters" reading the same template base two or more times, with a resulting insertion of a base.[671] Editing of transcripts usually serves an essential biological function, creating the correct sequence for translation of the mRNA and often generating multiple isoforms of proteins.

Finishing the transcripts. Additional modifications must be made to some mRNAs, and there will doubtless be many surprises as the details are worked out. One detail, which was discovered in the 1980s, is the specific function of snRNA U7 in recognition of the 3' end of pre-mRNAs for histones. The U7 RNA apparently base-pairs with a sequence near the 3' end cleavage site, acting as a cutting guide.[47,672,673]

Newly synthesized mRNA emerges from the nuclear pores as nucleoprotein complexes containing as many as ten different proteins. A major component is a 78-kDa polypeptide thought to be associated with the poly(A) tail. These ribonucleoproteins are sometimes stored for long periods of time, for example, in mature seed embryos and in amphibian oocytes.[575a,674] They may also travel rapidly for long distances, e.g., down nerve axons[675] or from cell to cell in plants via transport in the phloem.[676]

E. Transcription of Mitochondrial, Chloroplast, and Viral Genes

In the compact 16-kb chromosomes of mammalian mitochondria, the genes are tightly packed against one another (Fig. 18-3).[677,678] Most genes are transcribed using heavy H strands as templates and specialized bacteriophage-type RNA polymerases encoded in nuclear DNA.[679] A single promoter in the D loop region (Fig. 18-3) is used to make a long transcript from the entire H strand.[680] The transcript is then cut precisely by mitochondrial RNase P at the 5' ends of the tRNAs.[681] Similarly precise cleavage must occur at the 3' ends because there are often no nucleotides or only one to a few nucleotides separating adjacent genes. The nucleases involved differ from those used in nuclear tRNA 3' processing.[682] In animal mitochondria tRNA genes sometimes overlap by one nucleotide. Polyadenylation of the transcripts provides a form of editing that is required to create a UAA translational termination signal to which termination proteins bind.[683] The tRNA genes lack the 3' terminal CCA, which must be added. The lighter L strand has its own promoter, also located in the D loop. Both promoters contain the conserved nonanucleotide sequence 5'-ATATAAGTA. The pre-mRNAs created by these cleavages are not capped but are polyadenylated. Since the promoters are simple and the transcription factors few, mitochondrial transcription is controlled largely by mRNA stability, translation, and posttranslational events.[684]

The 70- to 100-kb circular mitochondrial DNA molecules of yeast and of higher plants contain more genes than do animal mitochondria, but most of the increased size is accounted for by intergenic spacers and by a few long introns. All yeast mitochondrial genes except for that of $tRNA^{Thr}$ are transcribed from a single strand. About 20 different primary transcripts have been identified in *Saccharomyces cerevisiae*. These originate at several points in the genome but always at the sequence 5-ATATAAGTA, the 3′A corresponding to the 5' nucleotide of the transcript.[678] One of these sequences is located at the origin of replication, suggesting the possibility that a normal RNA transcript provides the primer for DNA replication in yeast mitochondria.[677]

The 100- to 160-kb chloroplast genomes (Chapter 23, Section E,2) also have many prokaryotic features. They encode ~50 proteins as well as the tRNAs and rRNAs. Promoter and terminator sequences resemble those of bacteria and protein sequences are often homologous to those in bacteria. This applies, for example, to the α, β, and β′ subunits of RNA polymerase.[685]

1. Viral Transcription and Replication

Because viruses contain small genomes, study of transcription of viral DNA and of replication of RNA viruses has played an important role in helping us to understand transcription in eukaryotes.[47,686–688] An example is the discovery of the virus SV40 enhancer, which has been discussed in Section C,4. Study of viral life cycles is also essential to future progress in fighting viral diseases. Each of the many different viruses has its own often very complex life cycle. Only a few details can be given here. For lucid summaries see Voyles.[259]

Eukaryotic DNA viral genomes, like that of phage λ , usually contain early and late transcriptional units. The small papovaviruses, such as SV40 and polyoma virus, have 5.2 kb genomes. Like the small RNA viruses, they make use of overlapping genes and alternative RNA processing. In SV40 DNA there are two overlapping promoters called **early-early** and **late-early**. The first of these contains a TATA sequence, and both promoters also depend upon a 21-bp repeat segment as well as the SV40 enhancer. At least two proteins, one that binds to the enhancer and one that binds to the 21-bp repeat, are needed for initiation of early transcription. The early-late promoter lacks the TATA sequence but requires the 21-bp repeat and enhancer. A 94-kDa encoded protein called the **large T-antigen** (Chapter 27, Section C,10) is one of the regulators of transcription as well as of DNA replication.[689,690] This protein is also sufficient to transform rodent cells in culture. Although predominantly nuclear it is also inserted into the membrane where it acts as an antigen.

The large icosahedral adenoviruses cause respiratory infections in humans and attack and may cause cancer in many other vertebrate species including birds and amphibians. The 35.9-kb genome of human adenovirus-2 encodes at least 30 proteins, 10 of which appear in the virion. One of these is covalently linked to the 5' end of the DNA. As with smaller DNA viruses extensive use is made of alternative splicing of the transcribed RNA. There are at least six early transcriptional units, each with its own promoter. A variety of mRNAs are created using the various coding segments and a complex array of regulatory sequences control transcription. A 32-kDa phosphoprotein transcription factor designated E1A is encoded by a "pre-early"

gene. It is required along with host-encoded transcription factors for viral transcription.[691,692]

The late region of the adenovirus genome encodes structural proteins for the virus coat. Most of its transcripts begin about 16.5% of the way along the 36.5-kb dsDNA. However, the initial transcripts are cleaved at several different positions to yield a series of different 3'-poly(A)-terminated transcripts. In an exceedingly complex process the transcripts undergo splicing out of genes at their 5'-ends so that the final mRNAs typically code for single proteins. At the 5' ends the cap is joined to short segments from the original mRNA 5' end.[687,688,693]

Viruses SV40, polyoma, and some strains of adenoviruses are oncogenic in some species and cause transformations of cells in culture. Transformed cells always contain integrated viral DNA. That of SV40 can be incorporated at many different sites in the host genome. The integrated DNA does not always include the complete SV40 genome, and parts of the DNA may be inverted, deleted, or scrambled. Integration is not an essential part of the viral life cycle and has no effect on the infective properties of the viruses.[688] Cells transformed by adenoviruses usually also contain only a fragment of the viral genome in their DNA. However, one small set of genes from early region 1A is present in all transformed cells. The encoded proteins appear to be modulators of transcription and may cause cancer by promoting uncontrolled transcription of certain genes.[694]

2. Replication of RNA Viruses

The RNA (+) strands present in many RNA viruses often serve immediately after infection as a messenger RNA. However, replication requires formation of (–) strands of viral RNA from which new (+) strands can be transcribed for assembly into new virus particles. Other RNA viruses contain (–) strands of RNA or double-stranded RNA and, therefore, have significantly different life cycles.[259]

Small RNA viruses. The human polio virus, the common cold virus (rhinoviruses, Fig. 7-15), and other picorna viruses have 7.2- to 7.5-kb genomes with considerable homology (50% between the polio and rhinoviruses) and similar overall structures.[695] The polio genome encodes eight different proteins, one being a small 22-residue peptide that becomes covalently linked to the 5'-end of the RNA through a phosphodiester linkage to the side chain of a tyrosine. Cleavage of this linkage by a host enzyme allows the viral RNA (+) strand, which is polyadenylated at the 3' end, to serve as an mRNA for synthesis of a single large 220-kDa polyprotein. This is cleaved by a host protease at several Gln-Gly bonds to form several proteins. These include the 22-residue RNA-linked peptide, two capsid proteins, and a capsid precursor protein. The latter is cleaved during capsid assembly by a viral protease at an Asn-Ser bond to give two more capsid proteins. These four proteins have masses of 7, 26, 29, and 32 kDa. A viral protease, a large 58-kDa replicase, and a 37-kDa protein of unknown function are also cut from the polyprotein. A host protein initiates cleavage of the polyprotein, but the virally encoded protease later takes over this function.

One of the best understood of the many viral pathogens of plants is the tobacco mosaic virus (Fig. 7-8). Its 6.7-kb positive strand RNA encodes a replicase, coat protein, and at least one other protein.[696]

Influenza viruses. These negative-stranded viruses are classified into types A, B, and C, but it is only type A that infects nonhuman species including birds, horses, pigs, seals, mink, and whales.[697–699] Type A influenza viruses have also caused the great pandemics such as those in 1918–1919 and in 1968. Influenza viruses are surrounded by a lipid bilayer in which the virally encoded **hemagglutinin** and a **neuraminidase** (p. 186) are embedded. The inside of the bilayer is coated with a matrix protein and within this coat eight pieces of RNA of total length 13.6 kb are coiled together with a basic nucleoprotein. Also present are ten molecules each of three other proteins. The eight pieces of RNA vary in length from 900 to 2500 nucleotides. Seven of them encode one each of the seven virion proteins. One encodes an additional nonstructural protein while the smallest piece, using overlapping nucleotide sequences, encodes two nonstructural proteins.[697]

The existence of a fragmented genome evidently underlies the ability of influenza A viruses to undergo rapid changes in antigenic behavior. If a cell is coinfected with two strains of virus, the eight fragments act as independent chromosomes, which can be reassorted into new combinations in the progeny viruses. As a consequence, it is difficult to develop safe, live virus vaccines. A large reservoir of infection among migratory water birds and other animals facilitates the appearance of new strains and their rapid spread throughout the world.[698]

The first step in the replication of influenza viruses, which takes place in the cytoplasm, is the synthesis of (+) strands that can serve both as mRNA for synthesis of proteins and as templates for synthesis of new (–) strands. Three of the capsid proteins form the required RNA polymerase. This "transcriptase" is primed preferentially by 5'-capped 10- to 13-nucleotide segments of RNA that have been cut by a viral nuclease from host mRNAs.[700] The mRNAs made from viral RNA are polyadenylated and are translated by the host cell's ribosomes. However, some transcripts are used as templates to form viral (–) strands, which

are not polyadenylated and which contain uncapped pppA at the 5'-ends.

HIV-1 and other retroviruses. Because of their association with viral oncogenes (Chapter 11) and because of the **human immunodeficiency virus** (HIV-1) and the AIDS epidemic a great deal of attention is focused on retroviruses.[701–701b] Each retrovirus particle contains *two* identical single-stranded (+) RNA molecules, which may be as long as 10 kb. Their unique characteristic is that they induce synthesis of DNA, which must be integrated into the host genome before new viral (+) strands are transcribed. Retroviruses may sometimes cause cancer and may carry oncogenes (Chapter 11). Study of the **Rous sarcoma virus** (RSV), which infects chickens, and of the related **avian myeloblastosis virus** (AMV) and of HIV has revealed a common structure and a complex life style that are largely shared by all known retroviruses.

The organization of retroviruses[687,688,702] always includes a sequence of genes designated *gag* (glycoprotein antigen core proteins), *pol* (polymerase), and *env* (envelope) (Fig. 28-23). These are often followed by an oncogene.[703] In RSV this is the *src* gene (Chapter 11). At each end of the retrovirus gene sequence is a short direct repeat labeled R in Fig. 28-23. In RSV the R sequence is 21 nucleotides in length[688] and in HIV (see Fig. 28-23)[704] it is 98 nucleotides long.[705,706] The 5' end of the viral RNA is capped, and the 3' end is polyadenylated. The dsDNA of the integrated form of the virus (Fig. 28-23) is longer and at each end is bounded by **long terminal repeats** (LTRs). Each LTR consists of a sequence, designated U3, that is present next to R at the 3' end of the viral RNA. In the LTR this is followed by sequence R and then by U5, a unique sequence that came from the 5′ end of the viral RNA. Each ds-LTR begins and ends with a short inversely repeated segment:

5'-TGT —— ACA in RSV
5'-CTG —— CAG in HIV

The integrated provirus is always bounded by a sequence of host DNA that is repeated without inversion at the opposite end. For RSV this is a 5-bp sequence.

The LTRs in RSV are 569 bp in length, and those in HIV are 634 bp in length, 83 bp coming from U5, 98 from R, and 453 from U3.[706] The LTRs themselves often contain promoter and other control elements and even entire genes. The organization of a retrovirus (Fig. 28-23) reflects the complex mode of replication, which is presented in simplified form in Fig. 28-24. The key enzyme is the RNA-directed **reverse transcriptase** (Fig. 27-12).[707–709] The initial synthesis of DNA by this enzyme is primed by a tRNA. RSV uses $tRNA^{Trp}$ and HIV $tRNA^{Lys}$ for this purpose.[710–712] The 3' end of the tRNA, including the nucleotides forming both the acceptor stem and the stem of the TΨC loop (Fig. 5-30), must relinquish its normal base-pairing to form ~18 Watson–Crick base pairs with a **primer binding site** (labeled PB in Fig. 28-24) near the 5' end of the retroviral DNA. Because synthesis of the (–) strand of the retroviral DNA begins so close to the 5' end of the template (Fig. 28-24), only a short piece of DNA, including sequences U5' and R′ complementary to U5 and R, can be formed. For replication to continue RNA must be removed from hybrid regions. This is accomplished by the **RNaseH** activity of the reverse transcriptase.[708,709,713,713a] After removal of the RNA the primer tRNA must undergo a **strand transfer**, in which it shifts from the 5' end of the viral RNA template to the 3' end (Fig. 28-24), utilizing pairing between the right-hand R sequence of the template and the R′ sequence of the growing cDNA copy.[703,711,714] This transfer is sometimes to the second of the pair of identical RNA molecules in the virus, providing a way of increasing diversity by recombination. A second strand transfer of the growing (+) strand is needed to complete the dsDNA, which now contains the two identical LTRs. The 5' and 3' ends of the template RNA are doubtless held close together to facilitate strand transfer.

Integration of the dsDNA into the host DNA can occur at many places. The mechanism of integration probably resembles that used by phage λ (Fig. 27-27) and accounts for the duplication of host sequences at the two ends of the integrated virus. A virally encoded integrase catalyzes the process (see also Chapter 27, Section D,3).[715–717] It is the integrated virus that is transcribed to form new (+) viral RNA strands.

Integrated retroviruses are usually transcribed as full-length RNA copies, which may or may not have introns spliced out. The smaller spliced pieces encode the *env* and other genes such as *src* (Fig. 28-25). The *gag-pol* region is translated as a polyprotein that is cleaved into a number of pieces. These include four proteins of the virus core (encoded by *gag*), the reverse transcriptase with its associated RNaseH, and an **integrase**[718,718a] (all encoded by *pol*).[259] There is also an **aspartic protease** only 99 residues in length within *pol* (Fig. 28-25; Chapter 12, Box 12-C). The promoter and control region for transcription is located in the U3 region and is placed into a position where it can function only upon synthesis of the first LTR. The gene *env* encodes the major viral envelope protein and is translated from a spliced mRNA (Fig. 28-25).

Accessory regulatory genes. HIV and some related retroviruses such as HTLV-1 (which causes rare T-cell leukemias)[719] are distinguished from other retroviruses by a marked increase in the rate of DNA transcription within infected cells as compared with uninfected cells. This is thought to be a result of synthesis of virally encoded proteins that are trans-acting

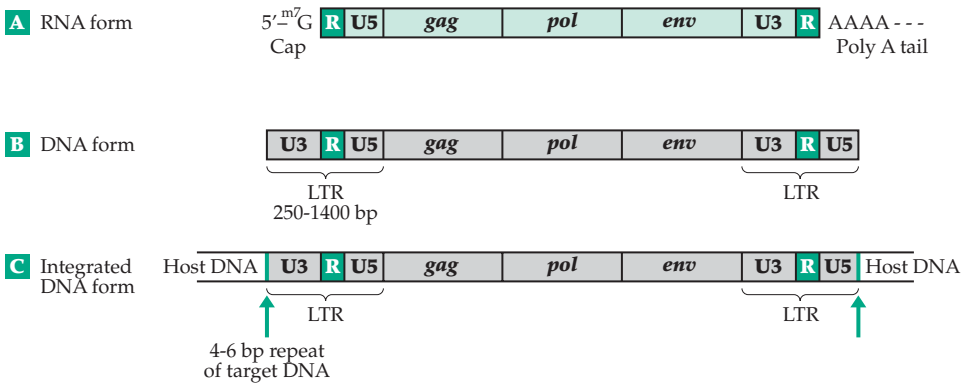


Figure 28-23 Comparison of the forms of a retrovirus. (A) Infective RNA (+) strand. (B) The double-stranded DNA form. (C) The DNA form integrated into the host DNA. LTR, long terminal repeats.

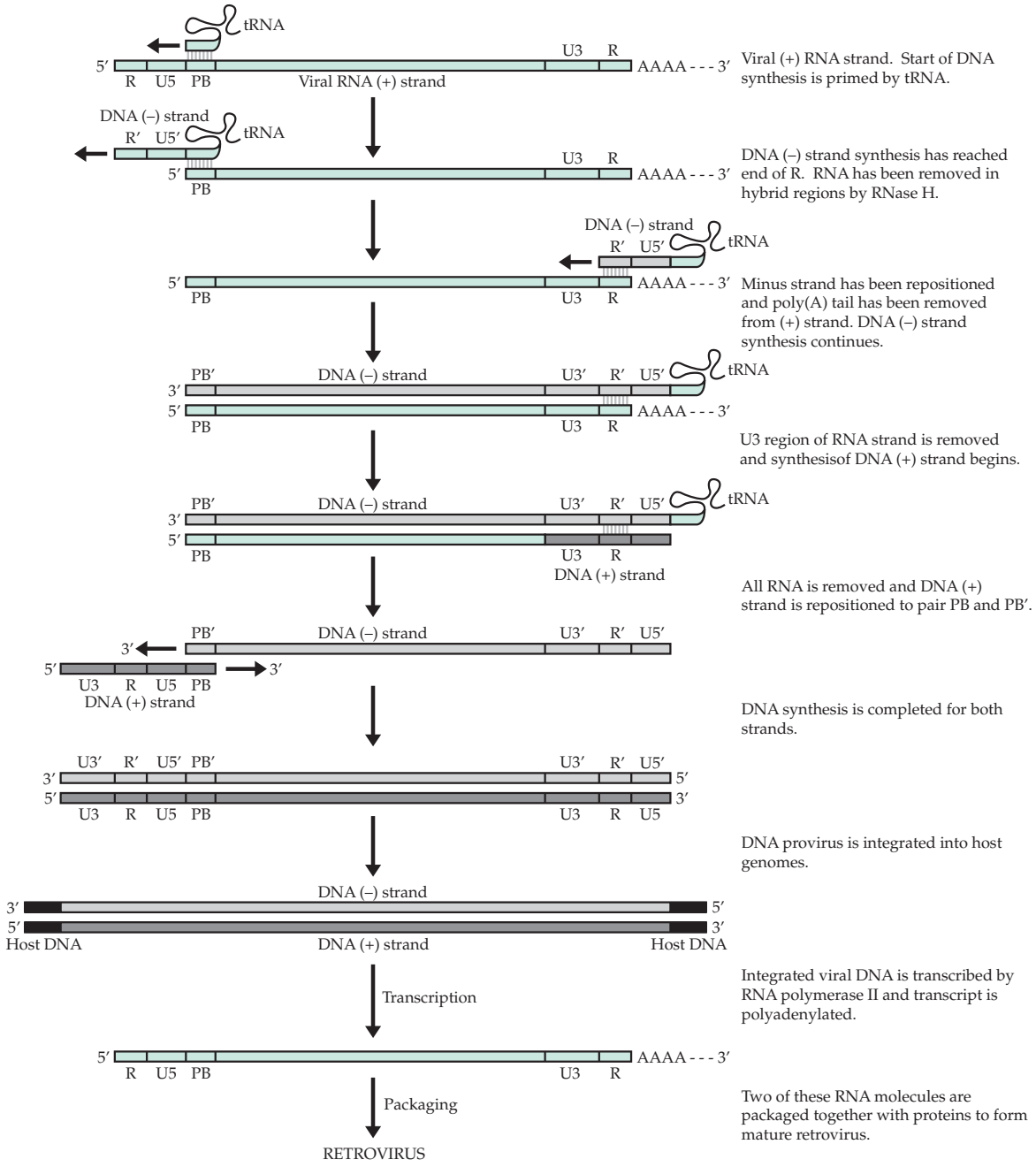


Figure 28-24 Simplified scheme for replication of the RNA genome of a retrovirus. See Sugden.[703] PB, Primer-binding site.

regulators of transcription. The HIV genome map (Fig. 28-26) shows the positions of the nine recognized genes marked in the three reading frames. In addition to the *gag*, *pol*, and *env* genes there are genes for six accessary proteins: Tat, Rev, Vif, Vpr, Vpu, and Nef.[720–722] Transcription is initiated at the promoter in the 3' LTR. This contains a TATAA sequence, an SP1 binding site, and an enhancer that binds transcription factor NF-κB (Fig. 5-40). The full-length 9-kb transcript contains, according to Frankel and Young,[722] the following

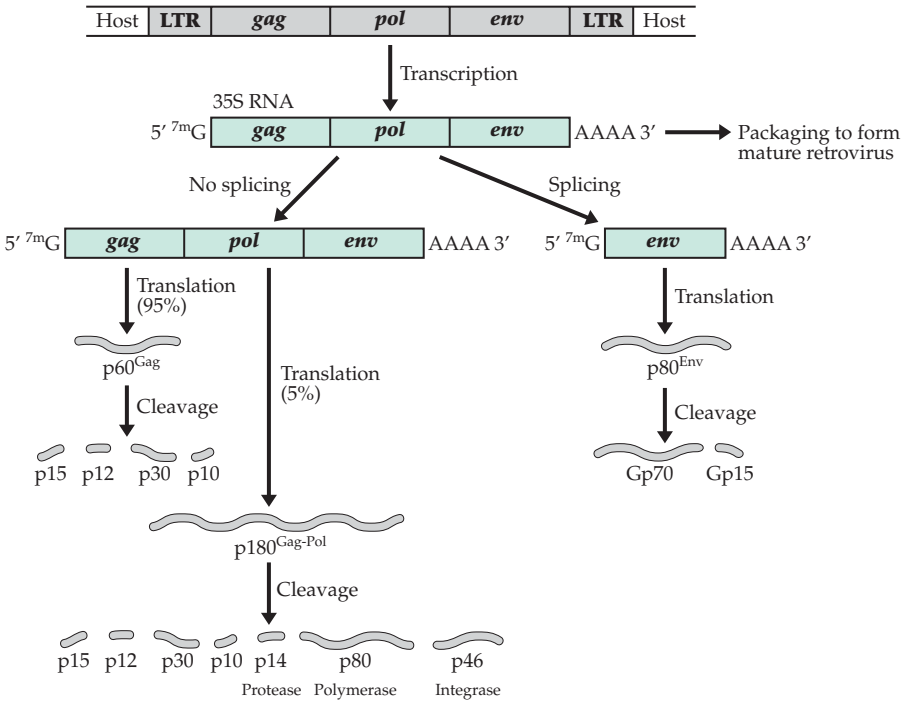


Figure 28-25 Gene expression from a typical retrovirus that has been integrated into a host's genome. This figure illustrates how a variety of proteins are encoded by a single rather short piece of DNA. After Voyles.[259]

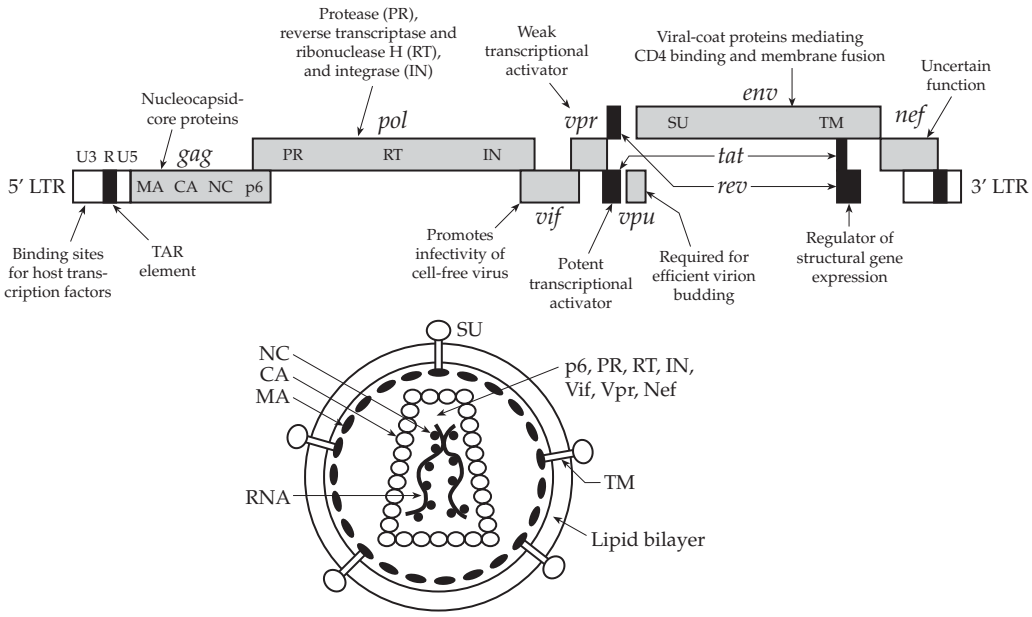


Figure 28-26 Simplified genetic map of the AIDS virus HIV-1. All three reading frames are utilized to encode nine genes, which give rise to 15 proteins. After Frankel and Young.[722]

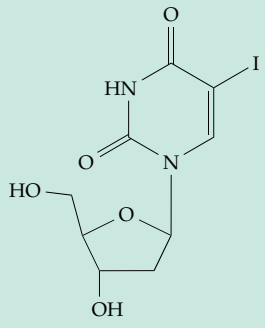
BOX 28-C SYNTHETIC ANTIVIRAL COMPOUNDS

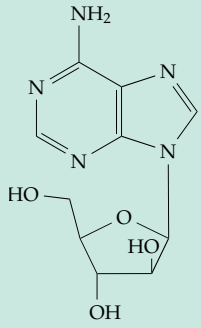
Most bacterial infections can be treated successfully with antibiotics, but the development of satisfactory antiviral agents has been slow. Yet we are susceptible to many dangerous virus diseases, and viruses also take a huge toll among domestic animals and plants.[a] The first antiviral drug, 5-iodo-2'-deoxyuridine (idoxuridine), was introduced in 1962 and was used for 20 years by ophthalmologists to treat serious eye infections by the herpes simplex virus (HSV).

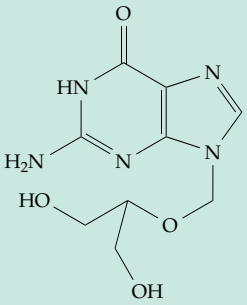
More recently 9-β-D-arabinofuranosyladenine (Ara-A) has become a preferred drug in treatment of ocular herpes infections. This compound, which is a naturally occurring antibiotic, can also be administered intravenously for life-threatening infections such as herpes encephalitis.[a] Ara-A is quite toxic but the guanine derivative 9-(2-hydroxyethoxymethyl) guanine (acyclovir) is less so. Another acyclic 2'-deoxyguanine analog, 9-(1,3-dihydroxypropoxymethyl)-guanine (DHPG), is more soluble, more potent, and has a broader range of effectiveness.[a–c]

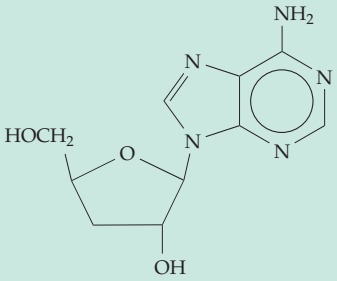
One of the first effective drugs against RNA viruses was ribavirin (1-β-D-ribofuranosyl-1,2,4-triazole-3-carboxamide), an analog of guanosine. It has a broad range of action and is used to treat severe viral pneumonia and bronchitis caused by respiratory syncytial virus (RSV). This common virus usually is mild but may cause death in infants and children. Ribavirin may also be of some value in the treatment of AIDS.

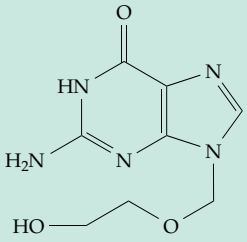
How do these compounds exert their antiviral effects? The nucleoside analogs are usually phosphorylated to the corresponding mono-, di-, and triphosphate derivatives by cellular enzymes. Thus, Ara-A yields Ara-ATP, which inhibits the herpes virus-encoded DNA polymerase. Ara-A may also enter the viral DNA. In addition Ara-A may inhibit

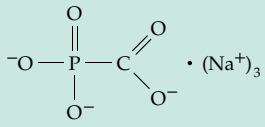

Idoxuridine

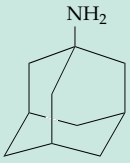

Ara-A

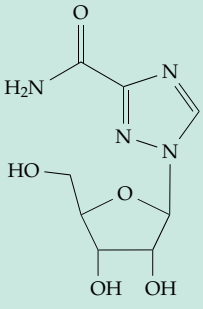

DHPG

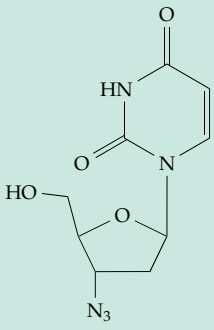

Cordycepin (3'-deoxyadenosine) preferentially inhibits synthesis of ribosomal and tRNA

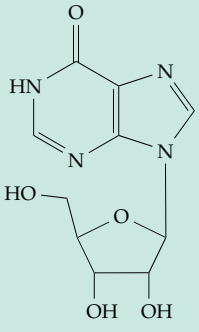

Acyclovir

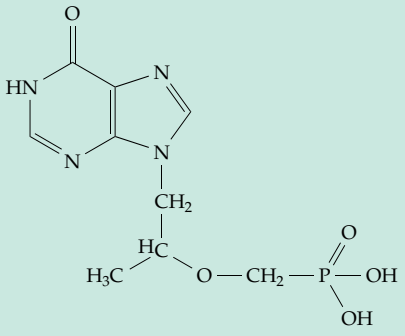

Trisodium phosphonoformate (Foscarnet)


Amantadine


Ribavirin


3′-Deoxy-3′-azidothymidine (AZT)


Dideoxyadenosine


(*R*)-9-(2-Phosphonylmethoxypropyl) adenine

BOX 28-C (continued)

polyadenylation of virally induced mRNA. Some analogs, such as acyclovir, undergo little conversion to the monophosphate by cellular kinases but are efficiently phosphorylated by herpes virus-encoded thymidine kinase. Thus, acyclovir does little damage to uninfected cells.[a] Ribavirin 3'-monophosphate may inhibit IMP dehydrogenase (Fig. 25-16, left) thereby interfering with GTP production. At the same time ribavirin triphosphate competes with GTP to inhibit virally encoded RNA polymerase.[a]

Phosphonoformate is a pyrophosphate analog and inhibits both DNA polymerases and reverse transcriptase. However, toxicity may prevent long-term treatment of AIDS patients. Amantadine has a narrow antiviral specificity. It specifically inhibits initiation of the replication of influenza virus RNA of type A (but not of type B). Active only against retroviruses, 3'-azidothymidine is a reverse transcriptase inhibitor, which acts by a chain termination mechanism. It was synthesized in the early 1960s but only recently has been used in treatment of AIDS victims. More recently a series of 2',3'-dideoxynucleosides, such as dideoxyinosine, have also been used.[d] Acyclic phosphonates, such as phosphonylmethoxypropyladenine, avoid the need for metabolic phosphorylation of the drug.[e]

Development of synthetic antiviral compounds is hardly beyond its infancy.[f] Serious problems must be overcome with most of these compounds. Toxicity (sometimes carcinogenicity), development of resistance by viruses, and enzymatic destruction limit the utility of most drugs. For example, adenosine deaminase destroys Ara-A quite rapidly. With our rapidly advancing knowledge of viral life cycles and protein and nucleic acid structures many new drug targets have been identified.[g] Among the targets for HIV are the reverse transcriptase,[h] protease,[i] and integrase.[j,k] Computer-assisted design, as well as new techniques of synthesis and screening, have allowed development of many non-nucleoside inhibitors.

Oligonucleotide phosphoramidates and other triplex-forming compounds may be designed to bind to specific DNA targets.[l]

---

[a] Robins, R. K. (1986) *Chem. Eng. News* **64**, 28–40
[b] Cheng, Y.-C., Grill, S. P., Dutschman, G. E., Nakayama, K., and Bastow, K. F. (1983) *J. Biol. Chem.* **258**, 12460–12464
[c] Biron, K. K., Fyfe, J. A., Stanat, S. C., Leslie, L. K., Sorrell, J. B., Lambe, C. U., and Coen, D. M. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8769–8773
[d] Sandberg, J. A., and Slikker, W., Jr. (1995) *FASEB J.* **9**, 1157–1163
[e] Tsai, C.-C., Follis, K. E., Sabo, A., Beck, T. W., Grant, R. F., Bischofberger, N., Benveniste, R. E., and Black, R. (1995) *Science* **270**, 1197–1199
[f] Mitsuya, H., Yarchoan, R., and Broder, S. (1990) *Science* **249**, 1533–1544
[g] Richman, D. D. (1996) *Science* **272**, 1886–1888
[h] Althaus, I. W., Chou, J. J., Gonzales, A. J., Deibel, M. R., Chou, K.-C., Kezdy, F. J., Romero, D. L., Palmer, J. R., Thomas, R. C., Aristoff, P. A., Tarpley, W. G., and Reusser, F. (1993) *Biochemistry* **32**, 6548–6554
[i] Rosin, C. D., Belew, R. K., Walker, W. L., Morris, G. M., Olson, A. J., and Goodsell, D. S. (1999) *J. Mol. Biol.* **287**, 77–92
[j] Robinson, W. E., Jr., Reinecke, M. G., Abdel-Malek, S., Jia, Q., and Chow, S. A. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6326–6331
[k] Goldgur, Y., Craigie, R., Cohen, G. H., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H., and Davies, D. R. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13040–13043
[l] Giovannangeli, C., Perrouault, L., Escudé, C., Gryaznov, S., and Hélène, C. (1996) *J. Mol. Biol.* **261**, 386–398

essential regions, whose positions may vary somewhat from one isolate of the virus to another:

A complex pattern of splicing produces more than 30 mRNAs.[723] When viral RNAs are first produced most are doubly spliced, allowing the split genes *tat* and *rev* to be expressed by synthesis of Tat and Rev. The *trans*-activator protein Tat is a small 86-residue cysteine-rich protein, which binds the Tar stem–loop structure and greatly stimulates transcription.[724–725a]

The 116-residue *rev* (regulator of expression of virion genes) gene product is also a transactivator, which is needed for rapid production of singly spliced (4 kb) or unspliced (9kb) *gag-pol* mRNA required for formation of virus structural proteins.[726,727] The effect of Rev is probably on transport from the nucleus rather than on splicing.

Transcription is repressed by the 206-residue N-terminal myristoylated protein, Nef, a phosphoprotein that associates with cytoplasmic membranes.[728–729a] It has been difficult to learn its exact function, but it seems to be required for maintenance of the integrated provirus for long periods of time without extensive replication. Mutations in gene *nef* do not eliminate the ability of the virus to replicate in T lymphocytes and to kill them.

The 23-kDa protein Vif (**viral infectivity factor**) is not needed for growth but is essential for infectivity.[729b] Other genes in HIV are *vpu*, which encodes an 81-residue integral membrane protein (**virion protein U**), and *vpr*, which encodes the 96-kDa **virion protein R**. Several possible functions have been proposed for these small proteins.[722,729b,c]

HIV-1 is a member of the group of slow viruses or **lentiviruses**.[730] Other lentiviruses include the human HIV-2,[731] an immunodeficiency virus that attacks cats causing leukemia,[732] and the human leukemia virus

Position from 5' end	Description
1–55	TAR, a 59-residue stem-loop structure; binding site for Tat
182–199	PB, primer binding site
240–350	Packaging signal. Binds envelope protein NC
248–271	Dimerization site with "kissing loop"
290	Major splice donor site, used to form all spliced mRNAs
1631–1673	Gag-Pol frameshifting region where –1 ribosomal frameshifting occurs to allow 5–10% synthesis of Gag-Pol polyprotein
7362–7596	Rev response element. Binding site for Rev
5358 and 7971	Two major splice acceptor sites; other minor sites are also used
9205–9210	Polyadenylation signal

HTLV-1.[733] Another lentivirus causes two diseases of sheep, **maedi**, a pulmonary disease, and **visna**, a paralytic condition somewhat similar to multiple sclerosis.[734] Because of its slow development there has been doubt as to the cause of AIDS, but there is now little doubt that HIV-1 is the true culprit.

Success in treating AIDS may depend upon better understanding of the complex life cycle of HIV-1,[722,730,735] which is summarized in Fig. 28-27. The cycle begins with the binding of the virion envelope protein to the immunoglobulin-like surface protein **CD4**, which is found principally on the type T4 helper T cells (Chapter 31). Binding of CD4 to the HIV envelope proteins appears to activate the T cells to enter the cell cycle and to take up and integrate the virus. The virus infection destroys these CD4+ lymphocytes with a half-life of less than two days.[735]

A major effort is being made to devise a vaccine against HIV. However, rapid mutation of the *env* gene makes it difficult to accomplish.[736,736a] This high rate of mutation appears to be a result of a high frequency of errors by the HIV reverse transcriptase.[737] There is

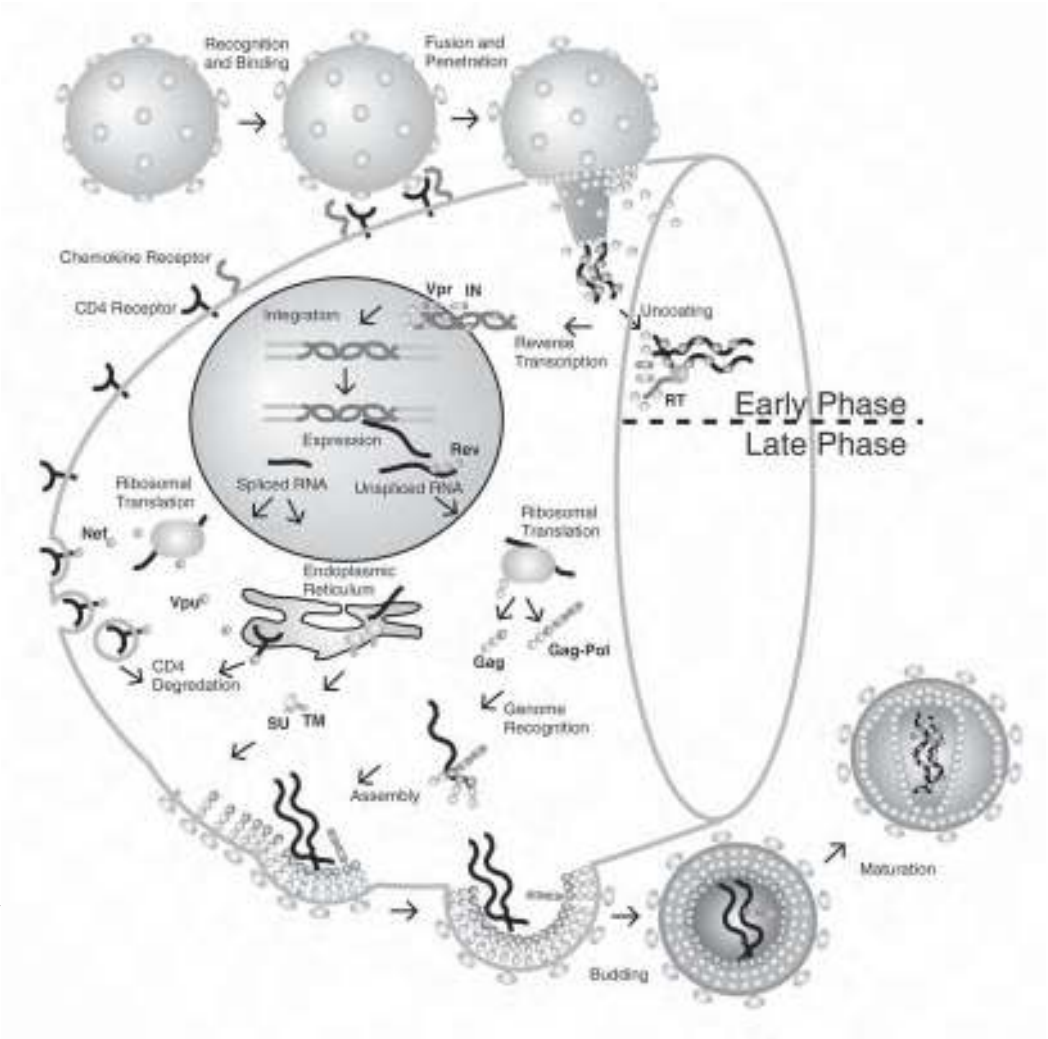


Figure 28-27 General features of the HIV-1 replication cycle. The early phase (upper portion of the diagram) begins with CD4 recognition and involves events up to and including integration of the proviral DNA, and the late phase includes all events from transcription of the integrated DNA to virus budding and maturation. From Turner and Summers.[735]

hope from the fact that some individuals are naturally resistant to HIV infection.[738,739]

3. Retrotransposons

Transposition of DNA, which is discussed in Chapter 27, Section D,4, may seem to be a rare and relatively unimportant event in our body cells. However, transposon DNA accounts for 35% or more of the human genome[740] and apparently plays a major role in evolution. Like other transposons, the DNA sequences known as retrotransposons also move about within DNA. However, they use an indirect mechanism that involves synthesis of mRNA and reverse transcription.[740,741] The reverse transcribed complementary DNA may be inserted back into the genome at new locations. The necessary chemical reactions parallel those involved in the replication of retroviruses (Fig. 28-23, 28-24). Retrotransposens, truncated retrotransposons, and related sequences constitute as much as 16% of the human genome.[741]

There are two classes of retrotrotransposons: those with long terminal repeats (LTRs) and those without (LTRs). The first group is closely related to retroviruses, but its members lack genes for envelope proteins. They do carry *gag* and *pol* genes similar to those of retroviruses (Fig. 28-3). Most retrotransposons are defective and do not move. Over evolutionary time they accumulate in the genome, sometimes to the extent that the genome size grows enormously. This has happened often during the evolution of plants, some of which (e.g., certain lilies) have 40 times more DNA per cell than do humans (Table 1-3).[742] Although most retrotransposons are inactive, some of them occasionally jump to new locations where they may mutate a gene and may sometimes cause disease. However, their major significance is probably in facilitating evolution, perhaps including the formation of new species.[743]

The non-LTR transposons are exemplified by the 6–7 kbp LINES (p. 1539)[741,743] and the short 90–400 bp SINES (p. 1538, Fig. 27-9).[744] Mammalian genomes contain ~50,000 truncated members of the LINE-1 (L1) family and 3000–5000 full-length L1s. Only a few of these are active in our present population. The RNA intermediates that participate in retroposition of LINES are generated by RNA polymerase II, while RNA polymerase III forms the RNA intermediates for propagation of SINES. Participation of these RNAs in trans splicing processes can modify existing genes, contributing to the remodeling of the genome.[744]

Yet another group of mobile elements in the genome are **intein genes**, which encode protein-splicing polypeptides (see Box 29-D). Many inteins also have **homing endonuclease** activity and cleave DNA at specific insertion sequences, initiating incorporation of intein DNA into new locations in the genome.[745] Group II introns, which are found in bacteria and in organelles of fungi and plants, may also act as mobile DNA elements.[746]

References

1. Storz, G. (2002) *Science* **296**, 1260–1262
1a. Zamore, P. D. (2002) *Science* **296**, 1265–1269
1b. Ruvkun, G. (2001) *Science* **294**, 797–799
1c. Cohen, G. N. (1995) *FASEB J.* **9**, 981–982
2. Dickson, R. C., Abelson, J., Barnes, W. M., and Reznikoff, W. S. (1975) *Science* **187**, 27–35
3. Müller-Hill, B. (1996) *The lac Operon. A Short History of a Genetic Paradigm*, de Gruyter, Berlin
3a. Juers, D. H., Heightman, T. D., Vasella, A., McCarter, J. D., Mackenzie, L., Withers, S. G., and Matthews, B. W. (2001) *Biochemistry* **40**, 14781–14794
4. Page, M. G. P., and Rosenbusch, J. P. (1988) *J. Biol. Chem.* **263**, 15906–15914
5. Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) *Science* **271**, 1247–1254
5a. Bell, C. E., and Lewis, M. (2001) *J. Mol. Biol.* **312**, 921–926
6. Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S., and Miller, J. H. (1994) *J. Mol. Biol.* **240**, 421–433
7. Nakanishi, S., Adhya, S., Gottesman, M., and Pastan, I. (1973) *J. Biol. Chem.* **248**, 5937–5942
8. Gilbert, W., and Maxam, A. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3581–3584
9. Simons, A., Tils, D., van Wilcken-Bergmann, B., and Muller-Hill, B. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1624–1628
10. von Hippel, P. H., Bear, D. G., Morgan, W. D., and McSwiggen, J. A. (1984) *Rev. Biochem.* **53**, 389–446
11. Rastinejad, F., Artz, P., and Lu, P. (1993) *J. Mol. Biol.* **233**, 389–399
12. Levandoski, M. M., Tsodikov, O. V., Frank, D. E., Melcher, S. E., Saecker, R. M., and Record, M. T., Jr. (1996) *J. Mol. Biol.* **260**, 697–717
13. Frank, D. E., Saecker, R. M., Bond, J. P., Capp, M. W., Tsodikov, O. V., Melcher, S. E., Levandoski, M. M., and Record, M. T., Jr. (1997) *J. Mol. Biol.* **267**, 1186–1206
14. von Hippel, P. H., and Berg, O. G. (1989) *J. Biol. Chem.* **264**, 675–678
15. Shimamoto, N. (1999) *J. Biol. Chem.* **274**, 15293–15296
16. Harada, K., and Frankel, A. D. (1995) *EMBO J.* **14**, 5798–5811
17. Pace, H. C., Kercher, M. A., Lu, P., Markiewicz, P., Miller, J. H., Chang, G., and Lewis, M. (1997) *Trends Biochem. Sci.* **22**, 334–339
18. Horton, N., Lewis, M., and Lu, P. (1997) *J. Mol. Biol.* **265**, 1–7
19. Slijper, M., Bonvin, A. M. J. J., Boelens, R., and Kaptein, R. (1996) *J. Mol. Biol.* **259**, 761–773
20. Barry, J. K., and Matthews, K. S. (1999) *Biochemistry* **38**, 3579–3590
21. Suckow, J., Markiewicz, P., Kleina, L. G., Miller, J., Kisters-Woike, B., and Müller-Hill, B. (1996) *J. Mol. Biol.* **261**, 509–523
22. Kolata, G. B. (1976) *Science* **191**, 373
23. Sommer, H., Lu, P., and Miller, J. H. (1976) *J. Biol. Chem.* **251**, 3774–3779
24. Müller, J., Oehler, S., and Müller-Hill, B. (1996) *J. Mol. Biol.* **257**, 21–29
25. Lanzer, M., and Bujard, H. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8973–8977
26. McClure, W. R. (1985) *Ann. Rev. Biochem.* **54**, 171–204
27. Auble, D. T., Allen, T. L., and deHaseth, P. L. (1986) *J. Biol. Chem.* **261**, 11202–11206
28. Place, C., Oddos, J., Buc, H., McAllister, W. T., and Buckle, M. (1999) *Biochemistry* **38**, 4948–4957
29. Cheetham, G. M. T., and Steitz, T. A. (1999) *Science* **286**, 2305–2309
30. Brieba, L. G., and Sousa, R. (2000) *Biochemistry* **39**, 919–923
30a. Brieba, L. G., Gopal, V., and Sousa, R. (2001) *J. Biol. Chem.* **276**, 10306–10313
31. Opalka, N., Mooney, R. A., Richter, C., Severinov, K., Landick, R., and Darst, S. A. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 617–622

References

32. Owens, J. T., Chmura, A. J., Murakami, K., Fujita, N., Ishihama, A., and Meares, C. F. (1998) *Biochemistry* **37**, 7670–7675
32a. Vuthoori, S., Bowers, C. W., McCracken, A., Dombroski, A. J., and Hinton, D. M. (2001) *J. Mol. Biol.* **309**, 561–572
32b. Kuznedelov, K., Minakhin, L., Niedziela-Majka, A., Dove, S. L., Rogulja, D., Nickels, B. E., Hochschild, A., Heyduk, T., and Severinov, K. (2002) *Science* **295**, 855–857
33. Brodolin, K., Mustaev, A., Severinov, K., and Nikiforov, V. (2000) *J. Biol. Chem.* **275**, 3661–3666
33a. Ebright, R. H. (2000) *J. Mol. Biol.* **304**, 687–698
33b. Minakhin, L., Bhagat, S., Brunning, A., Campbell, E. A., Darst, S. A., Ebright, R. H., and Severinov, K. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 892–897
33c. Murakami, K. S., Masuda, S., and Darst, S. A. (2002) *Science* **296**, 1280–1284
33d. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., and Darst, S. A. (2002) *Science* **296**, 1285–1290
33e. Vassylyev, D. G., Sekine, S.-i, Laptenko, O., Lee, J., Vassylyeva, M. N., and Borukhov, S. (2002) *Nature (London)* **417**, 712–718
33f. Lin, S., Katilius, E., Haffa, A. L. M., Taguchi, A. K. W., and Woodbury, N. W. (2001) *Biochemistry* **40**, 13767–13773
34. Nudler, E., Avetissova, E., Markovtsov, V., and Goldfarb, A. (1996) *Science* **273**, 211–217
35. Wu, F. Y. H., Huang, W.-J., Sinclair, R. B., and Powers, L. (1992) *J. Biol. Chem.* **267**, 25560–25567
36. Katayama, A., Fujita, N., and Ishihama, A. (2000) *J. Biol. Chem.* **275**, 3583–3592
37. Zaychikov, E., Martin, E., Denissova, L., Kozlov, M., Markovtsov, V., Kashlev, M., Heumann, H., Nikiforov, V., Goldfarb, A., and Mustaev, A. (1996) *Science* **273**, 107–109
38. Mustaev, A., Kozlov, M., Markovtsov, V., Zaychikov, E., Denissova, L., and Goldfarb, A. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6641–6645
39. Buckle, M., Pemberton, I. K., Jacquet, M.-A., and Buc, H. (1999) *J. Mol. Biol.* **285**, 955–964
40. Helmann, J. D., and Chamberlin, M. J. (1988) *Ann. Rev. Biochem.* **57**, 839–872
40a. Chadsey, M. S., and Hughes, K. T. (2001) *J. Mol. Biol.* **306**, 915–929
41. Carmona, M., Claverie-Martin, F., and Magasanik, B. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9568–9572
41a. Wigneshweraraj, S. R., Chaney, M. K., Ishihama, A., and Buck, M. (2001) *J. Mol. Biol.* **306**, 681–701
42. Pérez-Martín, J., and de Lorenzo, V. (1996) *J. Mol. Biol.* **258**, 562–574
43. Marschall, C., Labrousse, V., Kreimer, M., Weichart, D., Kolb, A., and Hengge-Aronis, R. (1998) *J. Mol. Biol.* **276**, 339–353
44. Blaszczak, A., Zylicz, M., Georgopoulos, C., and Liberek, K. (1995) *EMBO J.* **14**, 5085–5093
45. Huang, X., and Helmann, J. D. (1998) *J. Mol. Biol.* **279**, 165–173
46. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 784–788
47. Lewin, B. (2000) *Genes VII*, Oxford Univ. Press, New York
48. Bown, J. A., Owens, J. T., Meares, C. F., Fujita, N., Ishihama, A., Busby, S. J. W., and Minchin, S. D. (1999) *J. Biol. Chem.* **274**, 2263–2270
49. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) *Science* **277**, 1453–1462
50. Severinov, K., Mustaev, A., Severinova, E., Kozlov, M., Darst, S. A., and Goldfarb, A. (1995) *J. Biol. Chem.* **270**, 29428–29432
51. Yang, X., and Price, C. W. (1995) *J. Biol. Chem.* **270**, 23930–23933
52. Traviglia, S. L., Datwyler, S. A., and Meares, C. F. (1999) *Biochemistry* **38**, 4259–4265
53. Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R. L. (1993) *Science* **262**, 1407–1413
54. Ho Jeon, Y., Negishi, T., Shirakawa, M., Yamazaki, T., Fujita, N., Ishihama, A., and Kyogoku, Y. (1995) *Science* **270**, 1495–1497
54a. Zhang, G., and Darst, S. A. (1998) *Science* **281**, 262–266
54b. Wada, T., Yamazaki, T., and Kyogoku, Y. (2000) *J. Biol. Chem.* **275**, 16057–16063
54c. Calles, B., Monsalve, M., Rojo, F., and Salas, M. (2001) *J. Mol. Biol.* **307**, 487–497
54d. Benoff, B., Yang, H., Lawson, C. L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y. W., Berman, H. M., and Ebright, R. H. (2002) *Science* **297**, 1562–1566
55. Bokal, A. J., IV, Ross, W., and Gourse, R. L. (1995) *J. Mol. Biol.* **245**, 197–207
56. Muskhelishvili, G., Buckle, M., Heumann, H., Kahmann, R., and Travers, A. A. (1997) *EMBO J.* **16**, 3655–3665
57. Travers, A., and Muskhelishvili, G. (1998) *J. Mol. Biol.* **279**, 1027–1043
58. Pan, C. Q., Finkel, S. E., Cramton, S. E., Feng, J.-A., Sigman, D. S., and Johnson, R. C. (1996) *J. Mol. Biol.* **264**, 675–695
59. Lamond, A. I. (1985) *Trends Biochem. Sci.* **10**, 271–274
59a. Pemberton, I. K., Muskhelishvili, G., Travers, A. A., and Buckle, M. (2000) *J. Mol. Biol.* **299**, 859–864
60. Cashel, M., and Rudd, K. E. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), pp. 1410–1438, Am. Soc. for Microbiology, Washington, DC
61. Baracchi, E., and Bremer, H. (1988) *J. Biol. Chem.* **263**, 2597–2602
62. Choy, H. E. (2000) *J. Biol. Chem.* **275**, 6783–6789
62a. Barker, M. M., Gaal, T., and Gourse, R. L. (2001) *J. Mol. Biol.* **305**, 689–702
63. Yamagishi, M., Cole, J. R., Nomura, M., Studier, F. W., and Dunn, J. J. (1987) *J. Biol. Chem.* **262**, 3940–3943
64. Singer, P. T., and Wu, C.-W. (1988) *J. Biol. Chem.* **263**, 4208–4214
65. Gelles, J., and Landick, R. (1998) *Cell* **93**, 13–16
66. Bustamante, C., Guthold, M., Zhu, X., and Yang, G. (1999) *J. Biol. Chem.* **274**, 16665–16668
67. Lefèvre, J.-F., Lane, A. N., and Jardetzky, O. (1988) *Biochemistry* **27**, 1086–1094
68. Ellinger, T., Behnke, D., Knaus, R., Bujard, H., and Gralla, J. D. (1994) *J. Mol. Biol.* **239**, 466–475
69. von Hippel, P. H. (1998) *Science* **281**, 660–665
70. Levin, J. R., Blake, J. J., Ganunis, R. A., and Tullius, T. D. (2000) *J. Biol. Chem.* **275**, 6885–6893
71. Yager, T. D., and von Hippel, P. H. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), pp. 1241–1275, Am. Soc. for Microbiology, Washington, DC
72. Barile, M. F., Razin, S., Tully, J. G., and Whitcomb, R. F., eds. (1979) *The Mycoplasmas*, Vol. I, Academic Press, New York
73. Harris, E. H. (1989) *The Chlamydomonas Sourcebook*, Academic Press, San Diego, California
74. Milan, S., D'Ari, L., and Chamberlin, M. J. (1999) *Biochemistry* **38**, 218–225
75. Davenport, R. J., Wuite, G. J. L., Landick, R., and Bustamante, C. (2000) *Science* **287**, 2497–2500
76. Ellinger, T., Behnke, D., Bujard, H., and Gralla, J. D. (1994) *J. Mol. Biol.* **239**, 455–465
76a. Lee, K.-B., Wang, D., Lippard, S. J., and Sharp, P. A. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4239–4244
77. Selby, C. P., and Sancar, A. (1993) *Science* **260**, 53–58
78. Viswanathan, A., You, H. J., and Doetsch, P. W. (1999) *Science* **284**, 159–162
79. Nudler, E. (1999) *J. Mol. Biol.* **288**, 1–12
80. Erie, D. A., Hajiseyedjavadi, O., Young, M. C., and von Hippel, P. H. (1993) *Science* **262**, 867–873
81. Frank-Kamenetskii, M. (1989) *Nature (London)* **337**, 206
82. Krasilnikov, A. S., Podtelezhnikov, A., Vologodskii, A., and Mirkin, S. M. (1999) *J. Mol. Biol.* **292**, 1149–1160
83. Wu, H. (1988) *Cell* **53**, 433–440
84. Giaever, G. N., and Wang, J. C. (1988) *Cell* **55**, 849–856
85. Miller, O. L., Jr. (1973) *Sci. Am.* **228**(March), 34–42
86. Weickert, M. J., and Adhya, S. (1992) *J. Biol. Chem.* **267**, 15869–15874
87. Takeda, Y., Kim, J. G., Caday, C. G., Steers, E., Jr., Ohlendorf, D. H., Anderson, W. F., and Matthews, B. W. (1986) *J. Biol. Chem.* **261**, 8608–8616
88. Matthews, B. W., Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., and Takedo, Y. (1983) *Trends Biochem. Sci.* **8**, 25–29
89. Jana, R., Hazbun, T. R., Fields, J. D., and Mossing, M. C. (1998) *Biochemistry* **37**, 6446–6455
90. Pabo, C. O., and Lewis, M. (1982) *Nature (London)* **298**, 443–447
90a. Bell, C. E., and Lewis, M. (2001) *J. Mol. Biol.* **314**, 1127–1136
91. Jordan, S. R., and Pabo, C. O. (1988) *Science* **242**, 893–899
92. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., and Harrison, S. C. (1988) *Science* **242**, 899–907
93. Wolberger, C., Dong, Y., Ptashne, M., and Harrison, S. C. (1988) *Nature (London)* **335**, 789–795
94. Padmanabhan, S., Jiménez, M. A., González, C., Sanz, J. M., Giménez-Gallego, G., and Rico, M. (1997) *Biochemistry* **36**, 6424–6436
95. Wharton, R. P., and Ptashne, M. (1985) *Nature (London)* **316**, 601–605
96. Wharton, R. P., and Ptashne, M. (1986) *Trends Biochem. Sci.* **11**, 71–73
97. Sauer, R. T., Milla, M. E., Waldburger, C. D., Brown, B. M., and Schildbach, J. F. (1996) *FASEB J.* **10**, 42–48
98. Nooren, I. M. A., Rietveld, A. W. M., Melacini, G., Sauer, R. T., Kaptein, R., and Boelens, R. (1999) *Biochemistry* **38**, 6035–6042
99. He, Y.-y, Stockley, P. G., and Gold, L. (1996) *J. Mol. Biol.* **255**, 55–66
100. Dai, X., Kloster, M., and Rothman-Denes, L. B. (1998) *J. Mol. Biol.* **283**, 43–58
101. Wang, J. C., Barkley, M. D., and Bourgeois, S. (1974) *Nature (London)* **251**, 247–249
102. Horwitz, M. S. Z., and Loeb, L. A. (1988) *Science* **241**, 703–705
103. Arndt, K. T., Boschelli, F., Cook, J., Takeda, Y., Tecza, E., and Lu, P. (1983) *J. Biol. Chem.* **258**, 4177–4183
104. Metzler, W. J., and Lu, P. (1989) *J. Mol. Biol.* **205**, 149–164
105. Borowiec, J. A., and Gralla, J. D. (1986) *Biochemistry* **25**, 5051–5057
106. Ho, Y.-S., Wulff, D. L., and Rosenberg, M. (1983) *Nature (London)* **304**, 703–708
107. Brennan, R. G., Vasu, S., Matthews, B. W., and Otsuka, A. J. (1989) *J. Biol. Chem.* **264**, 5

References

108. Eisenstein, E., and Beckett, D. (1999) *Biochemistry* **38**, 13077–13084
109. Streaker, E. D., and Beckett, D. (1999) *J. Mol. Biol.* **292**, 619–632
109a. Weaver, L. H., Kwon, K., Beckett, D., and Matthews, B. W. (2001) *Protein Sci.* **10**, 2618–2622
110. Dahl, M. K., Degenkolb, J., and Hillen, W. (1994) *J. Mol. Biol.* **243**, 413–424
111. Yoshida, K.-I., Shibayama, T., Aoyama, D., and Fujita, Y. (1999) *J. Mol. Biol.* **285**, 917–929
112. Hinrichs, W., Kisker, C., Düvel, M., Müller, A., Tovar, K., Hillen, W., and Saenger, W. (1994) *Science* **264**, 418–420
113. Kisker, C., Hinrichs, W., Tovar, K., Hillen, W., and Saenger, W. (1995) *J. Mol. Biol.* **247**, 260–280
113a. Schubert, P., Schnappinger, D., Pfleiderer, K., and Hillen, W. (2001) *Biochemistry* **40**, 3257–3263
113b. Lathe, W. C., III, Snel, B., and Bork, P. (2000) *Trends Biochem. Sci.* **25**, 474–479
114. Cohen, G. N. (1986) in *Regulation of Gene Expression* (Booth, I. R., and Higgins, C. F., eds), pp. 1–20, Cambridge Univ. Press, London
115. Cornish, E. C., Argyropoulos, V. P., Pittard, J., and Davidson, B. E. (1986) *J. Biol. Chem.* **261**, 403–410
116. Weiss, D. L., Johnson, D. I., Weith, H. L., and Somerville, R. L. (1986) *J. Biol. Chem.* **261**, 9966–9971
117. Choi, K. Y., Lu, F., and Zalkin, H. (1994) *J. Biol. Chem.* **269**, 24066–24072
118. Xu, H., Moraitis, M., Reedstrom, R. J., and Matthews, K. S. (1998) *J. Biol. Chem.* **273**, 8958–8964
119. Glasfeld, A., Koehler, A. N., Schumacher, M. A., and Brennan, R. G. (1999) *J. Mol. Biol.* **291**, 347–361
120. Charlier, D., Roovers, M., Van Vliet, F., Boyen, A., Cunin, R., Nakamura, Y., Glansdorff, N., and Piérard, A. (1992) *J. Mol. Biol.* **226**, 367–386
121. Van Duyne, G. D., Ghosh, G., Maas, W. K., and Sigler, P. B. (1996) *J. Mol. Biol.* **256**, 377–391
122. Wang, H., Glansdorff, N., and Charlier, D. (1998) *J. Mol. Biol.* **277**, 805–824
123. Ni, J., Sakanyan, V., Charlier, D., Glansdorff, N., and Van Duyne, G. D. (1999) *Nature Struct. Biol.* **6**, 427–432
124. Escolar, L., Pérez-Martín, J., and de Lorenzo, V. (1998) *J. Mol. Biol.* **283**, 537–547
125. White, A., Ding, X., vanderSpek, J. C., Murphy, J. R., and Ringe, D. (1998) *Nature (London)* **394**, 502–506
126. Pohl, E., Holmes, R. K., and Hol, W. G. J. (1998) *J. Biol. Chem.* **273**, 22420–22427
127. Pohl, E., Holmes, R. K., and Hol, W. G. J. (1999) *J. Mol. Biol.* **285**, 1145–1156
127a. Gourley, D. G., Schüttelkopf, A. W., Anderson, L. A., Price, N. C., Boxer, D. H., and Hunter, W. N. (2001) *J. Biol. Chem.* **276**, 20641–20647
128. Postma, P. W. (1986) in *Regulation of Gene Expression* (Booth, I. R., and Higgins, C. F., eds), Cambridge Univ. Press, London
129. Busby, S., and Ebright, R. H. (1999) *J. Mol. Biol.* **293**, 199–213
130. Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991) *Science* **253**, 1001–1007
131. Passner, J. M., and Steitz, T. A. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2843–2847
132. Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. E., and Berman, H. M. (1996) *J. Mol. Biol.* **260**, 395–408
133. Gunasekera, A., Ebright, Y. W., and Ebright, R. H. (1992) *J. Biol. Chem.* **267**, 14713–14720
134. Vossen, K. M., Stickle, D. F., and Fried, M. G. (1996) *J. Mol. Biol.* **255**, 44–54
135. Dalma-Weiszhausz, D. D., and Brenowitz, M. (1996) *Biochemistry* **35**, 3735–3745
136. Roy, S., Garges, S., and Adhya, S. (1998) *J. Biol. Chem.* **273**, 14059–14062
137. Danot, O., Vidal-Ingigliardi, D., and Raibaud, O. (1996) *J. Mol. Biol.* **262**, 1–11
138. Richet, E., and Sogaard-Andersen, L. (1994) *EMBO J.* **13**, 4558–4567
139. Ziegelhoffer, E. C., and Kiley, P. J. (1995) *J. Mol. Biol.* **245**, 351–361
140. Lazazzera, B. A., Beinert, H., Khoroshilova, N., Kennedy, M. C., and Kiley, P. J. (1996) *J. Biol. Chem.* **271**, 2762–2768
141. Popescu, C. V., Bates, D. M., Beinert, H., Münck, E., and Kiley, P. J. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13431–13435
142. Shelver, D., Thorsteinsson, M. V., Kerby, R. L., Chung, S.-Y., Roberts, G. P., Reynolds, M. F., Parks, R. B., and Burstyn, J. N. (1999) *Biochemistry* **38**, 2669–2678
142a. Thorsteinsson, M. V., Kerby, R. L., Youn, H., Conrad, M., Serate, J., Staples, C. R., and Roberts, G. P. (2001) *J. Biol. Chem.* **276**, 26807–26813
143. Greene, J. R., Morrissey, L. M., Foster, L. M., and Geiduschek, E. P. (1986) *J. Biol. Chem.* **261**, 12820–12827
144. Takahashi, K., Kawazoe, Y., Sakumi, K., Nakabeppu, Y., and Sekiguchi, M. (1988) *J. Biol. Chem.* **263**, 13490–13492
145. Lobell, R. B., and Schleif, R. F. (1990) *Science* **250**, 528–532
146. Zhang, X., Reeder, T., and Schleif, R. (1996) *J. Mol. Biol.* **258**, 14–24
147. Niland, P., Hühne, R., and Müller-Hill, B. (1996) *J. Mol. Biol.* **264**, 667–674
147a. Wu, M., and Schleif, R. (2001) *J. Mol. Biol.* **307**, 1001–1009
147b. LaRonde-LeBlanc, N., and Wolberger, C. (2000) *Biochemistry* **39**, 11593–11601
148. Echols, H. (1990) *J. Biol. Chem.* **265**, 14697–14700
149. Rippe, K., von Hippel, P. H., and Langowski, J. (1995) *Trends Biochem. Sci.* **20**, 500–506
150. Blackwood, E. M., and Kadonaga, J. T. (1998) *Science* **281**, 60–63
151. Magasanik, B. (1988) *Trends Biochem. Sci.* **13**, 475–479
152. Rippe, K., Guthold, M., von Hippel, P. H., and Bustamante, C. (1997) *J. Mol. Biol.* **270**, 125–138
153. Flashner, Y., Weiss, D. S., Keener, J., and Kustu, S. (1995) *J. Mol. Biol.* **249**, 700–713
153a. Schulz, A., Langowski, J., and Rippe, K. (2000) *J. Mol. Biol.* **300**, 709–725
154. Brahms, G., Brahms, S., and Magasanik, B. (1995) *J. Mol. Biol.* **246**, 35–42
155. Cullen, P. J., Bowman, W. C., Hartnett, D.-F., Reilly, S. C., and Kranz, R. G. (1998) *J. Mol. Biol.* **278**, 903–914
156. Dworkin, J., Jovanovic, G., and Model, P. (1997) *J. Mol. Biol.* **273**, 377–388
157. Solà, M., Gomis-Rüth, F. X., Serrano, L., González, A., and Coll, M. (1999) *J. Mol. Biol.* **285**, 675–687
158. Makino, K., Amemura, M., Kawamoto, T., Kimura, S., Shinagawa, H., Nakata, A., and Suzuki, M. (1996) *J. Mol. Biol.* **259**, 15–26
159. Shi, L., Liu, W., and Hulett, F. M. (1999) *Biochemistry* **38**, 10119–10125
159a. Georgellis, D., Kwon, O., and Lin, E. C. C. (2001) *Science* **292**, 2314–2316
160. Strohmaier, H., Noiges, R., Kotschan, S., Sawers, G., Högenauer, G., Zechner, E. L., and Koraimann, G. (1998) *J. Mol. Biol.* **277**, 309–316
161. Miyatake, H., Mukai, M., Adachi, S.-i, Nakamura, H., Tamura, K., Iizuka, T., Shiro, Y., Strange, R. W., and Hasnain, S. S. (1999) *J. Biol. Chem.* **274**, 23176–23184
161a. Rusnak, F., and Reiter, T. (2000) *Trends Biochem. Sci.* **25**, 527–529
162. Li, J., Kustu, S., and Stewart, V. (1994) *J. Mol. Biol.* **241**, 150–165
163. Asayama, M., Yamamoto, A., and Kobayashi, Y. (1995) *J. Mol. Biol.* **250**, 11–23
164. Madhusudan, Zapf, J., Hoch, J. A., Whiteley, J. M., Xuong, N. H., and Varughese, K. I. (1997) *Biochemistry* **36**, 12739–12745
165. Eguchi, Y., Itoh, T., and Tomizawa, J.-i. (1991) *Ann. Rev. Biochem.* **60**, 631–652
166. Weintraub, H. M. (1990) *Sci. Am.* **262**(Jan), 40–46
166a. Nellen, W., and Lichtenstein, C. (1993) *Trends Biochem. Sci.* **18**, 419–423
166b. Delihas, N., and Forst, S. (2001) *J. Mol. Biol.* **313**, 1–12
167. Green, P. J., Pines, O., and Inouye, M. (1986) *Ann. Rev. Biochem.* **55**, 569–597
168. Cesareni, G., and Banner, D. W. (1985) *Trends Biochem. Sci.* **10**, 303–306
169. Asano, K., Niimi, T., Yokoyama, S., and Mizobuchi, K. (1998) *J. Biol. Chem.* **273**, 11826–11838
170. Ikenaka, K., Ramakrishnan, G., Inouye, M., Tsung, K., and Inouye, M. (1986) *J. Biol. Chem.* **261**, 9316–9320
171. Andersen, J., Forst, S. A., Zhao, K., Inouye, M., and Delihas, N. (1989) *J. Biol. Chem.* **264**, 17961–17970
171a. Matzke, M., Matzke, A. J. M., and Kooter, J. M. (2001) *Science* **293**, 1080–1083
172. Madison-Antenucci, S., Sabatini, R. S., Pollard, V. W., and Hajduk, S. L. (1998) *EMBO J.* **17**, 6368–6376
173. Nicoloso, M., Qu, L.-H., Michot, B., and Bachellerie, J.-P. (1996) *J. Mol. Biol.* **260**, 178–195
174. Lafontaine, D. L. J., and Tollervey, D. (1998) *Trends Biochem. Sci.* **23**, 383–388
174a. Gerdes, K., Nielsen, A., Thorsted, P., and Wagner, E. G. H. (1992) *J. Mol. Biol.* **226**, 637–649
175. Cohen, J. S., and Hogan, M. E. (1994) *Sci. Am.* **271**(Jun), 76–82
176. Askari, F. K., and McDonnell, W. M. (1996) *N. Engl. J. Med.* **334**, 316–318
176a. Sullenger, B. A., and Gilboa, E. (2002) *Nature (London)* **418**, 252–258
176b. Gitlin, L., Karelsky, S., and Andino, R. (2002) *Nature (London)* **418**, 430–434
177. Wittung-Stafshede, P. (1998) *Science* **281**, 657–658
178. Wagner, R. W. (1994) *Nature (London)* **372**, 333–335
179. Tereshko, V., Gryaznov, S., and Egli, M. (1998) *J. Am. Chem. Soc.* **120**, 269–283
180. Aramini, J. M., Kalisch, B. W., Pon, R. T., van de Sande, J. H., and Germann, M. W. (1996) *Biochemistry* **35**, 9355–9365
181. Mujeeb, A., Reynolds, M. A., and James, T. L. (1997) *Biochemistry* **36**, 2371–2379
182. Branch, A. D. (1998) *Trends Biochem. Sci.* **23**, 45–50
182a. Braasch, D. A., and Corey, D. R. (2002) *Biochemistry* **41**, 4503–4510
183. Platt, T. (1986) *Ann. Rev. Biochem.* **55**, 339–372
184. Yanofsky, C. (1981) *Nature (London)* **289**, 751–758
185. Yanofsky, C. (1988) *J. Biol. Chem.* **263**, 609–612
186. Kuroda, M. I., and Yanofsky, C. (1984) *J. Biol. Chem.* **259**, 12838–12843
187. Baumann, C., Otridge, J., and Gollnick, P. (1996) *J. Biol. Chem.* **271**, 12269–12274
187a. Losick, R., and Sonenshein, A. L. (2001) *Science* **293**, 2018–2019
187b. Valbuzzi, A., and Yanofsky, C. (2001) *Science* **293**, 2057–2059
187c. Yakhnin, A. V., and Babitzke, P. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11067–11072

References

188. Antson, A. A., Dodson, E. J., Dodson, G., Greaves, R. B., Chen, X.-p, and Gollnick, P. (1999) *Nature (London)* **401**, 235 – 242
189. Jensen, K. F., Bonekamp, F., and Poulsen, P. (1986) *Trends Biochem. Sci.* **11**, 362 – 365
190. Hauser, C. A., Sharp, J. A., Hatfield, L. K., and Hatfield, G. W. (1985) *J. Biol. Chem.* **260**, 1765 – 1770
191. Watson, M. D. (1981) *Trends Biochem. Sci.* **6**, 180 – 182
192. Stitt, B. L. (1988) *J. Biol. Chem.* **263**, 11130 – 11137
193. Bear, D. G., and Peabody, D. S. (1988) *Trends Biochem. Sci.* **13**, 343 – 347
194. Nehrke, K. W., and Platt, T. (1994) *J. Mol. Biol.* **243**, 830 – 839
195. Liu, K., and Hanna, M. M. (1995) *J. Mol. Biol.* **247**, 547 – 558
196. Greenblatt, J. (1991) *Trends Biochem. Sci.* **16**, 408 – 411

196a. Toulokhonov, I., Artsimovitch, I., and Landick, R. (2001) *Science* **292**, 730 – 733

196b. Gopal, B., Haire, L. F., Gamblin, S. J., Dodson, E. J., Lane, A. N., Papavinasasundaram, K. G., Colston, M. J., and Dodson, G. (2001) *J. Mol. Biol.* **314**, 1087 – 1095

196c. Zhou, Y., Mah, T.-F., Yu, Y.-T. N., Mogridge, J., Olson, E. R., Greenblatt, J., and Friedman, D. I. (2001) *J. Mol. Biol.* **310**, 33 – 49

197. Zhu, A. Q., and von Hippel, P. H. (1998) *Biochemistry* **37**, 11202 – 11214
198. Kim, D.-E., and Patel, S. S. (1999) *J. Biol. Chem.* **274**, 32667 – 32671

198a. Vincent, F., Openshaw, M., Trautwein, M., Gaskell, S. J., Kohn, H., and Widger, W. R. (2000) *Biochemistry* **39**, 9077 – 9083

198b. Yu, X., Horiguchi, T., Shigesada, K., and Egelman, E. H. (2000) *J. Mol. Biol.* **299**, 1279 – 1287

199. Greenblatt, J., Nodwell, J. R., and Mason, S. W. (1993) *Nature (London)* **364**, 401 – 406
200. O'Hara, B. P., Norman, R. A., Wan, P. T. C., Roe, S. M., Barrett, T. E., Drew, R. E., and Pearl, L. H. (1999) *EMBO J.* **18**, 5175 – 5186

200a. Gopal, B., Papavinasasundaram, K. G., Dodson, G., Colston, M. J., Major, S. A., and Lane, A. N. (2001) *Biochemistry* **40**, 920 – 928

200b. Carlomagno, M. S., and Nappo, A. (2001) *J. Mol. Biol.* **309**, 19 – 28

201. Sozhamannan, S., and Stitt, B. L. (1997) *J. Mol. Biol.* **268**, 689 – 703
202. Dunn, J. J., and Studier, F. W. (1983) *J. Mol. Biol.* **166**, 477 – 535
203. Daniels, C. J., Gupta, R., and Doolittle, W. F. (1985) *J. Biol. Chem.* **260**, 3132 – 3134
204. Lykke-Andersen, J., Aagaard, C., Semionenkov, M., and Garrett, R. A. (1997) *Trends Biochem. Sci.* **22**, 326 – 331
205. Abelson, J., Trotta, C. R., and Li, H. (1998) *J. Biol. Chem.* **273**, 12685 – 12688
206. Starzyk, R. M. (1986) *Trends Biochem. Sci.* **11**, 60
207. Sollner-Webb, B., and Mougey, E. B. (1991) *Trends Biochem. Sci.* **16**, 58 – 62
208. Li, Z., Pandit, S., and Deutscher, M. P. (1999) *EMBO J.* **18**, 2878 – 2885

208a. Nomura, M. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1820 – 1822

209. Perry, R. P. (1976) *Ann. Rev. Biochem.* **45**, 605 – 629
210. King, T. C., Sirdeshmukh, R., and Schlessinger, D. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 185 – 188
211. Denoya, C., and Dubnau, D. (1989) *J. Biol. Chem.* **264**, 2615 – 2624
212. Tscherne, J. S., Nurse, K., Popienick, P., Michel, H., Sochacki, M., and Ofengand, J. (1999) *Biochemistry* **38**, 1884 – 1892
213. Stahl, D. A., Pace, B., Marsh, T., and Pace, N. R. (1984) *J. Biol. Chem.* **259**, 11448 – 11453
214. Raychaudhuri, S., Niu, L., Conrad, J., Lane, B. G., and Ofengand, J. (1999) *J. Biol. Chem.* **274**, 18880 – 18886
215. Niu, L., and Ofengand, J. (1999) *Biochemistry* **38**, 629 – 635
216. Ramamurthy, V., Swann, S. L., Spedaliere, C. J., and Mueller, E. G. (1999) *Biochemistry* **38**, 13106 – 13111
217. Gu, X., Liu, Y., and Santi, D. V. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14270 – 14275
218. Conrad, J., Sun, D., Englund, N., and Ofengand, J. (1998) *J. Biol. Chem.* **273**, 18562 – 18566

218a. Spedaliere, C. J., Hamilton, C. S., and Mueller, E. G. (2000) *Biochemistry* **39**, 9459 – 9465

218b. Arluison, V., Buckle, M., and Grosjean, H. (1999) *J. Mol. Biol.* **289**, 491 – 502

219. Kiss-László, Z., Henry, Y., and Kiss, T. (1998) *EMBO J.* **17**, 797 – 807
220. Segal, D. M., and Eichler, D. C. (1991) *J. Biol. Chem.* **266**, 24385 – 24389
221. Frank, D. N., and Pace, N. R. (1998) *Ann. Rev. Biochem.* **67**, 153 – 180
222. Christian, E. L., and Harris, M. E. (1999) *Biochemistry* **38**, 12629 – 12638

222a. Zuleeg, T., Hansen, A., Pfeiffer, T., Schübel, H., Kreutzer, R., Hartmann, R. K., and Limmer, S. (2001) *Biochemistry* **40**, 3363 – 3369

222b. Cole, K. B., and Dorit, R. L. (2001) *J. Mol. Biol.* **307**, 1181 – 1193

222c. Mikkelsen, N. E., Brännvall, M., Virtanen, A., and Kirsebom, L. A. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6155 – 6160

223. Misra, T. K., and Apirion, D. (1978) *J. Biol. Chem.* **253**, 5594 – 5599
224. Nashimoto, M., Tamura, M., and Kaspar, R. L. (1999) *J. Mol. Biol.* **287**, 727 – 740
225. Deutscher, M. P., and Marlor, C. W. (1985) *J. Biol. Chem.* **260**, 7067 – 7071
226. Robertson, H. D., Altman, S., and Smith, J. D. (1972) *J. Biol. Chem.* **247**, 5243 – 5251
227. Reilly, R. M., and RajBhandary, U. L. (1986) *J. Biol. Chem.* **261**, 2928 – 2935
228. Khorana, H. G., and 19 other authors. (1976) *J. Biol. Chem.* **251**, 565 – 570
229. Ramamoorthy, B., Lees, R. G., Kleid, D. G., and Khorana, H. G. (1976) *J. Biol. Chem.* **251**, 676 – 694
230. Sekiya, T., Contreras, R., Kupper, H., Landy, A., and Khorana, H. G. (1976) *J. Biol. Chem.* **251**, 5124 – 5140
231. Gommers-Ampt, J. H., and Borst, P. (1995) *FASEB J.* **9**, 1034 – 1042
232. Söll, D., and RajBhandary, U. L., eds. (1995) *tRNA Structure, Biosynthesis, and Function*, Am. Soc. for Microbiology, Washington, D.C.
233. Deutscher, M. P., Marlor, C. W., and Zaniewski, R. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6429 – 6430
234. Yue, D., Weiner, A. M., and Maizels, N. (1998) *J. Biol. Chem.* **273**, 29693 – 29700
235. Kline, L. K., and Söll, D. (1982) in *The Enzymes*, Vol. 15 (Boyer, P. D., ed), pp. 567 – __, Academic Press, New York
236. Greenberg, R., and Dudock, B. (1980) *J. Biol. Chem.* **255**, 8296 – 8302
237. Kwong, L. K., Moore, V. G., and Kaiser, I. I. (1977) *J. Biol. Chem.* **252**, 6310 – 6315
238. Becker, H. F., Motorin, Y., Sissler, M., Florentz, C., and Grosjean, H. (1997) *J. Mol. Biol.* **274**, 505 – 518
239. Kambampati, R., and Lauhon, C. T. (2000) *J. Biol. Chem.* **275**, 10727 – 10730
240. Kaiser, J. T., Clausen, T., Bourenkow, G. P., Bartunik, H.-D., Steinbacher, S., and Huber, R. (2000) *J. Mol. Biol.* **297**, 451 – 464
241. Schaeffer, K. P., Altman, S., and Soll, D. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3626 – 3630
242. Reyniers, J. P., Pleasants, J. R., Wostmann, B. S., Katze, J. R., and Farkas, W. R. (1981) *J. Biol. Chem.* **256**, 11591 – 11594
243. Katze, J. R., Gündüz, U., Smith, D. L., Cheng, C. S., and McCloskey, J. A. (1984) *Biochemistry* **23**, 1171 – 1176
244. Slany, R. K., Bösl.M, Crain, P. F., and Kersten, H. (1993) *Biochemistry* **32**, 7811 – 7817
245. Romier, C., Reuter, K., Suck, D., and Ficner, R. (1996) *Biochemistry* **35**, 15734 – 15739
246. Curnow, A. W., and Garcia, G. A. (1995) *J. Biol. Chem.* **270**, 17264 – 17267

246a. Kittendorf, J. D., Barcomb, L. M., Nonekowski, S. T., and Garcia, G. A. (2001) *Biochemistry* **40**, 14123 – 14133

247. Phillipson, D. W., Edmonds, C. G., Crain, P. F., Smith, D. L., Davis, D. R., and McCloskey, J. A. (1987) *J. Biol. Chem.* **262**, 3462 – 3471
248. Kolesky, S., Ouhammouch, M., Brody, E. N., and Geiduschek, E. P. (1999) *J. Mol. Biol.* **291**, 267 – 281
249. Gopal, V., Brieba, L. G., Guajardo, R., McAllister, W. T., and Sousa, R. (1999) *J. Mol. Biol.* **290**, 411 – 431

249a. Liu, C., and Martin, C. T. (2001) *J. Mol. Biol.* **308**, 465 – 475

249b. Severinov, K. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5 – 7

250. Gottesman, M. (1999) *J. Mol. Biol.* **293**, 177 – 180
251. Gottesman, S. (1988) in *Biochemistry*, 2nd ed. (Zubay, G., ed), pp. 1000 – 1007, Macmillan, New York
252. Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, R. T., Ackers, G. K., and Ptashne, M. (1981) *Nature (London)* **294**, 217 – 223
253. Ptashne, M., Johnson, A. D., and Pabo, C. O. (1982) *Sci. Am.* **247**(Nov), 128 – 140
254. Albright, R. A., and Matthews, B. W. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3431 – 3436
255. Honigman, A., Hu, S.-L., Chase, R., and Szbalski, W. (1976) *Nature (London)* **262**, 112 – 116
256. Shimatake, H., and Rosenberg, M. (1981) *Nature (London)* **292**, 128 – 132
257. Maurizi, M. R. (1987) *J. Biol. Chem.* **262**, 2696 – 2793
258. Kastelein, R. A., Remaut, E., Fiers, W., and van Duin, J. (1982) *Nature (London)* **295**, 35 – 41
259. Voyles, B. A. (1993) *The Biology of Viruses*, Mosby, St. Louis, Missouri
260. Brown, D., and Gold, L. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11558 – 11562
261. Schuppli, D., Georgijevic, J., and Weber, H. (2000) *J. Mol. Biol.* **295**, 149 – 154
262. Blumenthal, T., and Carmichael, G. G. (1979) *Ann. Rev. Biochem.* **48**, 525 – 548
263. Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and VandePol, S. (1982) *Science* **215**, 1577 – 1585
264. Mills, D. R., Kramer, F. R., Dobkin, C., Nishihara, T., and Spiegelman, S. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4252 – 4256

264a. Weinzierl, R. O. J. (1999) *Mechanisms of Gene Expression.Structure,Function and Evolution of the Basal Transcriptional Machinery*, Imperial College Press,

265. Langer, D., Hain, J., Thuriaux, P., and Zillig, W. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5768 – 5772
266. Bell, S. D., Kosa, P. L., Sigler, P. B., and Jackson, S. P. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13662 – 13667

266a. Bell, S. D., and Jackson, S. P. (2000) *J. Biol. Chem.* **275**, 31624 – 31629

266b. Thomsen, J., De Biase, A., Kaczanowski, S., Macario, A. J. L., Thomm, M., Zielenkiewicz, P., MacColl, R., and de Macario, E. C. (2001) *J. Mol. Biol.* **309**, 589 – 603

267. Krüger, K., Hermann, T., Armbruster, V., and Pfeifer, F. (1998) *J. Mol. Biol.* **279**, 761 – 771

References

268. Kyrpides, N. C., and Ouzounis, C. A. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8545–8550
268a. Lewis, P. J., Thaker, S. D., and Errington, J. (2000) *EMBO J.* **19**, 710–718
268b. Iborra, F. J., Jackson, D. A., and Cook, P. R. (2001) *Science* **293**, 1139–1142
268c. von Mering, C., and Bork, P. (2002) *Nature (London)* **417**, 797–798
269. Jackson, D. A., Pombo, A., and Iborra, F. (2000) *FASEB J.* **14**, 242–254
270. Workman, J. L., and Kingston, R. E. (1998) *Ann. Rev. Biochem.* **67**, 545–579
271. Hagmann, M. (1999) *Science* **285**, 1200–1203
271a. Jenuwein, T., and Allis, C. D. (2001) *Science* **293**, 1074–1080
271b. Bird, A. (2001) *Science* **294**, 2113–2115
271c. Gamble, M. J., and Freedman, L. P. (2002) *Trends Biochem. Sci.* **27**, 165–167
271d. Burke, L. J., and Baniahmad, A. (2000) *FASEB J.* **14**, 1876–1888
272. Kornberg, R. D. (1999) *Trends Biochem. Sci.* **24**, M46–M49
273. Carey, M., and Smale, S. T. (2000) *Transcriptional Regulation in Eukaryotes*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
273a. Steidl, S., Hynes, M. J., and Brakhage, A. A. (2001) *J. Mol. Biol.* **306**, 643–653
273b. Niehof, M., Kubicka, S., Zender, L., Manns, M. P., and Trautwein, C. (2001) *J. Mol. Biol.* **309**, 855–868
274. Vandromme, M., Gauthier-Rouvière, C., Lamb, N., and Fernandez, A. (1996) *Trends Biochem. Sci.* **21**, 59–64
275. Pennisi, E. (2000) *Science* **288**, 1372–1373
275a. Lo, W.-S., Duggan, L., Emre, N. C. T., Belotserkovskya, R., Lane, W. S., Shiekhattar, R., and Berger, S. L. (2001) *Science* **293**, 1142–1146
275b. Wang, H., Huang, Z.-Q., Xia, L., Feng, Q., Erdjument-Bromage, H., Strahl, B. D., Briggs, S. D., Allis, C. D., Wong, J., Tempst, P., and Zhang, Y. (2001) *Science* **293**, 853–857
275c. Nakayama, J.-i, Rice, J. C., Strahl, B. D., Allis, C. D., and Grewal, S. I. S. (2001) *Science* **292**, 110–113
275d. Jenuwein, T. (2002) *Science* **297**, 2215–2218
275e. Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. S., and Martienssen, R. A. (2002) *Science* **297**, 1833–1837
276. Chandler, S. P., Guschin, D., Landsberger, N., and Wolffe, A. P. (1999) *Biochemistry* **38**, 7008–7018
276a. Burnett, E., Christensen, J., and Tattersall, P. (2001) *J. Mol. Biol.* **314**, 1029–1039
277. Ng, H. H., and Bird, A. (2000) *Trends Biochem. Sci.* **25**, 121–126
278. Alland, L., Muhle, R., Hou, H., Jr., Potes, J., Chin, L., Schreiber-Agus, N., and DePinho, R. A. (1997) *Nature (London)* **387**, 49–55
278a. Arévalo-Rodríguez, M., Cardenas, M. E., Wu, X., Hanes, S. D., and Heitman, J. (2000) *EMBO J.* **19**, 3739–3749
279. Heinzel, T., Lavinsky, R. M., Mullen, T.-M., Söderström, M., Laherty, C. D., Torchia, J., Yang, W.-M., Brard, G., Ngo, S. D., Davie, J. R., Seto, E., Eisenman, R. N., Rose, D. W., Glass, C. K., and Rosenfeld, M. G. (1997) *Nature (London)* **387**, 43–48
279a. Lechner, T., Lusser, A., Pipal, A., Brosch, G., Loidl, A., Goralik-Schramel, M., Sendra, R., Wegener, S., Walton, J. D., and Loidl, P. (2000) *Biochemistry* **39**, 1683–1692
279b. Bernstein, B. E., Tong, J. K., and Schreiber, S. L. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13708–13713
279c. Johnson, C. A., Padget, K., Austin, C. A., and Turner, B. M. (2001) *J. Biol. Chem.* **276**, 4539–4542
280. Bernhard, D., Ausserlechner, M. J., Tonko, M., Löffler, M., Hartmann, B. L., Csordas, A., and Kofler, R. (1999) *FASEB J.* **13**, 1991–2001
280a. Gross, D. S. (2001) *Trends Biochem. Sci.* **26**, 685–686
280b. Smith, J. S., and Boeke, J. D. (2001) *Science* **291**, 608–609
280c. Tanny, J. C., and Moazed, D. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 415–420
280d. Sauve, A. A., Celic, I., Avalos, J., Deng, H., Boeke, J. D., and Schramm, V. L. (2001) *Biochemistry* **40**, 15456–15463
280e. Ghosh, D., Gerasimova, T. I., and Corces, V. G. (2001) *EMBO J.* **20**, 2518–2527
280f. Donze, D., and Kamakaka, R. T. (2001) *EMBO J.* **20**, 520–531
280g. Bell, A. C., West, A. G., and Felsenfeld, G. (2001) *Science* **291**, 447–450
280h. Tijsterman, M., Ketting, R. F., Okihara, K. L., Sijen, T., and Plasterk, R. H. A. (2002) *Science* **295**, 694–697
280i. Ahlquist, P. (2002) *Science* **296**, 1270–1273
280j. Jorgensen, R. A. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11561–11563
281. Kouzarides, T. (2000) *EMBO J.* **19**, 1176–1179
282. Brown, C. E., Lechner, T., Howe, L., and Workman, J. L. (2000) *Trends Biochem. Sci.* **25**, 15–19
283. Trievel, R. C., Rojas, J. R., Sterner, D. E., Venkataramani, R. N., Wang, L., Zhou, J., Allis, C. D., Berger, S. L., and Marmorstein, R. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8931–8936
283a. Sterner, D. E., Belotserkovskaya, R., and Berger, S. L. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11622–11627
283b. Hughes, J. D., Estep, P. W., Tavazoie, S., and Church, G. M. (2000) *J. Mol. Biol.* **296**, 1205–1214
283c. Chen, J., Rappsilber, J., Chiang, Y.-C., Russell, P., Mann, M., and Denis, C. L. (2001) *J. Mol. Biol.* **314**, 683–694
284. Bochar, D. A., Savard, J., Wang, W., Lafleur, D. W., Moore, P., Côté, J., and Shiekhattar, R. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1038–1043
284a. Fry, C. J., and Peterson, C. L. (2002) *Science* **295**, 1847–1848
284b. Aalfs, J. D., and Kingston, R. E. (2000) *Trends Biochem. Sci.* **25**, 548–555
284c. Becker, P. B., and Hörz, W. (2002) *Ann. Rev. Biochem.* **71**, 247–273
285. Kwon, H., Imbalzano, A. N., Khavarl, P. A., Kingston, R. E., and Green, M. R. (1994) *Nature (London)* **370**, 477–481
286. Richard-Foy, H. (1994) *Nature (London)* **370**, 417–418
287. Gavin, I. M., and Simpson, R. T. (1997) *EMBO J.* **16**, 6263–6271
287a. Xue, Y., Canman, J. C., Lee, C. S., Nie, Z., Yang, D., Moreno, G. T., Young, M. K., Salmon, E. D., and Wang, W. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13015–13020
287b. Olave, I. A., Reck-Peterson, S. L., and Crabtree, G. R. (2002) *Ann. Rev. Biochem.* **71**, 755–781
287c. Asturias, F. J., Chung, W.-H., Kornberg, R. D., and Lorch, Y. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13477–13480
288. de la Cruz, J., Kressler, D., and Linder, P. (1999) *Trends Biochem. Sci.* **24**, 192–198
288a. Linder, P., Tanner, N. K., and Banroques, J. (2001) *Trends Biochem. Sci.* **26**, 339–341
288b. Henn, A., Medalia, O., Shi, S.-P., Steinberg, M., Franceschi, F., and Sagi, I. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5007–5012
288c. Story, R. M., Li, H., and Abelson, J. N. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1465–1470
289. Wu, X., Wilcox, C. B., Devasahayam, G., Hackett, R. L., Arévalo-Rodríguez, M., Cardenas, M. E., Heitman, J., and Hanes, S. D. (2000) *EMBO J.* **19**, 3727–3738
289a. Conaway, R. C., Brower, C. S., and Conaway, J. W. (2002) *Science* **296**, 1254–1258
289b. Zheng, N., Schulman, B. A., Song, L., Miller, J. J., Jeffrey, P. D., Wang, P., Chu, C., Koepp, D. M., Elledge, S. J., Pagano, M., Conaway, R. C., Conaway, J. W., Harper, J. W., and Pavietich, N. P. (2002) *Nature (London)* **416**, 703–709
289c. Ottosen, S., Herrera, F. J., and Triezenberg, S. J. (2002) *Science* **296**, 479–481
289d. Gonzalez, F., Delahodde, A., Kodadek, T., and Johnston, S. A. (2002) *Science* **296**, 548–550
289e. Tsuge, T., Matsui, M., and Wei, N. (2001) *J. Mol. Biol.* **305**, 1–9
289f. Goodson, M. L., Hong, Y., Rogers, R., Matunis, M. J., Park-Sarge, O.-K., and Sarge, K. D. (2001) *J. Biol. Chem.* **276**, 18513–18518
289g. Sampson, D. A., Wang, M., and Matunis, M. J. (2001) *J. Biol. Chem.* **276**, 21664–21669
289h. Kito, K., Yeh, E. T. H., and Kamitani, T. (2001) *J. Biol. Chem.* **276**, 20603–20609
289i. Marx, J. (2001) *Science* **292**, 838–839
289j. Buchberger, A., Howard, M. J., Proctor, M., and Bycroft, M. (2001) *J. Mol. Biol.* **307**, 17–24
290. Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R., and Kornberg, R. D. (2000) *Science* **288**, 640–649
291. Kim, W.-Y., and Dahmus, M. E. (1989) *J. Biol. Chem.* **264**, 3169–3176
292. Geiduschek, E. P., and Tocchini-Valentini, G. P. (1988) *Ann. Rev. Biochem.* **57**, 873–914
293. Shaw, P. J., Highett, M. I., Beven, A. F., and Jordan, E. G. (1995) *EMBO J.* **14**, 2896–2906
294. Seither, P., Iben, S., and Grummt, I. (1998) *J. Mol. Biol.* **275**, 43–53
295. Kelly, J. L., and Lehman, I. R. (1986) *J. Biol. Chem.* **261**, 10340–10347
295a. Masters, B. S., Stohl, L. L., and Clayton, D. A. (1987) *Cell* **51**, 89–99
296. Hanley–Bowdoin, L., and Chua, N.-H. (1987) *Trends Biochem. Sci.* **12**, 67–70
297. Myer, V. E., and Young, R. A. (1998) *J. Biol. Chem.* **273**, 27757–27760
298. Fu, J., Gerstein, M., David, P. R., Gnatt, A. L., Bushnell, D. A., Edwards, A. M., and Kornberg, R. D. (1998) *J. Mol. Biol.* **280**, 317–322
299. Jensen, G. J., Meredith, G., Bushnell, D. A., and Kornberg, R. D. (1998) *EMBO J.* **17**, 2353–2358
299a. Todone, F., Weinzierl, R. O. J., Brick, P., and Onesti, S. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6306–6310
299b. Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001) *Science* **292**, 1863–1876
299c. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. (2001) *Science* **292**, 1876–1882
300. Simanek, E. E., Huang, D.-H., Pasternack, L., Machajewski, T. D., Seitz, O., Millar, D. S., Dyson, H. J., and Wong, C.-H. (1998) *J. Am. Chem. Soc.* **120**, 11567–11575
301. Adams, R. L. P., Knowler, J. T., and Leader, D. P. (1992) *The Biochemistry of the Nucleic Acids*, 11th ed., Chapman & Hall, London
302. Dahmus, M. E. (1996) *J. Biol. Chem.* **271**, 19009–19012
303. Corden, J. L., and Patturajan, M. (1997) *Trends Biochem. Sci.* **22**, 413–416
304. Proudfoot, N. (2000) *Trends Biochem. Sci.* **25**, 290–293
304a. Geiduschek, E. P., and Kassavetis, G. A. (2001) *J. Mol. Biol.* **310**, 1–26
304b. Kassavetis, G. A., Letts, G. A., and Geiduschek, E. P. (2001) *EMBO J.* **20**, 2823–2834

References

305. Pombo, A., Jackson, D. A., Hollinshead, M., Wang, Z., Roeder, R. G., and Cook, P. R. (1999) *EMBO J.* **18**, 2241 – 2253
306. Carter, K. C., Bowman, D., Carrington, W., Fogarty, K., McNeil, J. A., Fay, F. S., and Lawrence, J. B. (1993) *Science* **259**, 1330 – 1332
307. Nevins, J. R. (1983) *Ann. Rev. Biochem.* **52**, 441 – 466
308. Leff, S. E., Rosenfeld, M. G., and Evans, R. M. (1986) *Ann. Rev. Biochem.* **55**, 1091 – 1117
309. Ayoubi, T. A. Y., and Van de Ven, W. J. M. (1996) *FASEB J.* **10**, 453 – 460
310. Dantonel, J.-C., Wurtz, J.-M., Poch, O., Moras, D., and Tora, L. (1999) *Trends Biochem. Sci.* **24**, 335 – 339
311. Munoz-Sanjuan, I., Smallwood, P. M., and Nathans, J. (2000) *J. Biol. Chem.* **275**, 2589 – 2597
312. Ince, T. A., and Scotto, K. W. (1995) *J. Biol. Chem.* **270**, 30249 – 30252
312a. Conaway, J. W., Shilatifard, A., Dvir, A., and Conaway, R. C. (2000) *Trends Biochem. Sci.* **25**, 375 – 380
312b. Maniatis, T., and Reed, R. (2002) *Nature (London)* **416**, 499 – 506
312c. Sousa, R. (2001) *Trends Biochem. Sci.* **26**, 695 – 697
313. Arndt, K. T., Styles, C., and Fink, G. R. (1987) *Science* **237**, 874 – 880
314. Moye, W. S., and Zalkin, H. (1987) *J. Biol. Chem.* **262**, 3609 – 3624
315. Shuey, D. J., and Parker, C. S. (1986) *Nature (London)* **323**, 459 – 461
316. Prestridge, D. S. (1995) *J. Mol. Biol.* **249**, 923 – 932
317. Scherf, M., Klingenhoff, A., and Werner, T. (2000) *J. Mol. Biol.* **297**, 599 – 606
318. Kaiser, K., and Meisterernst, M. (1996) *Trends Biochem. Sci.* **21**, 342 – 345
318a. Tsuji, Y., Torti, S. V., and Torti, F. M. (1998) *J. Biol. Chem.* **273**, 2984 – 2992
319. Kadonaga, J. T., Jones, K. A., and Tjian, R. (1986) *Trends Biochem. Sci.* **11**, 20 – 23
320. Kadonaga, J. T., Carner, K. R., Masiarz, F. R., and Tjian, R. (1987) *Cell* **51**, 1079 – 1090
321. Garcia, A., Cereghini, S., and Sontag, E. (2000) *J. Biol. Chem.* **275**, 9385 – 9389
322. Li, B., Adams, C. C., and Workman, J. L. (1994) *J. Biol. Chem.* **269**, 7756 – 7763
323. Fry, C. J., and Farnham, P. J. (1999) *J. Biol. Chem.* **274**, 29583 – 29586
324. Daniel, S., Zhang, S., DePaoli-Roach, A. A., and Kim, K.-H. (1996) *J. Biol. Chem.* **271**, 14692 – 14697
324a. Johnson-Pais, T., Degnin, C., and Thayer, M. J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2211 – 2216
324b. Goodwin, A. J., McInerney, J. M., Glander, M. A., Pomerantz, O., and Lowrey, C. H. (2001) *J. Biol. Chem.* **276**, 26883 – 26892
325. Van Dyke, M. W., Roeder, R. G., and Sawadogo, M. (1988) *Science* **241**, 1335 – 1338
326. Roeder, R. G. (1996) *Trends Biochem. Sci.* **21**, 327 – 335
327. Nikolov, D. B., Chen, H., Halay, E. D., Hoffmann, A., Roeder, R. G., and Burley, S. K. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4862 – 4867
328. Nikolov, D. B., and Burley, S. K. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 15 – 22
328a. Sachs, A. B., and Buratowski, S. (1997) *Trends Biochem. Sci.* **22**, 189 – 192
329. Andel, F., III, Ladurner, A. G., Inouye, C., Tjian, R., and Nogales, E. (1999) *Science* **286**, 2153 – 2156
329a. Gangloff, Y.-G., Romier, C., Thuault, S., Werten, S., and Davidson, I. (2001) *Trends Biochem. Sci.* **26**, 250 – 257
330. Burley, S. K., and Roeder, R. G. (1996) *Ann. Rev. Biochem.* **65**, 769 – 799
331. Green, M. R. (2000) *Trends Biochem. Sci.* **25**, 59 – 63
332. Verrijzer, C. P., and Tjian, R. (1996) *Trends Biochem. Sci.* **21**, 338 – 342
332a. Metzger, D., Scheer, E., Soldatov, A., and Tora, L. (1999) *EMBO J.* **18**, 4823 – 4834
333. Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J., and Dickerson, R. E. (1996) *J. Mol. Biol.* **261**, 239 – 254
334. Parkhurst, K. M., Richards, R. M., Brenowitz, M., and Parkhurst, L. J. (1999) *J. Mol. Biol.* **289**, 1327 – 1341
335. Svejstrup, J. Q., Vichi, P., and Egly, J.-M. (1996) *Trends Biochem. Sci.* **21**, 346 – 350
335a. Holstege, F. C. P., and Young, R. A. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2 – 4
336. Yan, M., and Gralla, J. D. (1997) *EMBO J.* **16**, 7457 – 7467
336a. Mizzen, C. A., and Allis, C. D. (2000) *Science* **289**, 2290 – 2291
337. Nikolov, D. B., Hu, S.-H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N.-H., Roeder, R. G., and Burley, S. K. (1992) *Nature (London)* **360**, 40 – 46
338. Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G., and Burley, S. K. (1995) *Nature (London)* **377**, 119 – 128
338a. Brivanlou, A. H., and Darnell, J. E., Jr. (2002) *Science* **295**, 813 – 818
338b. Smith, R. L., and Johnson, A. D. (2000) *Trends Biochem. Sci.* **25**, 325 – 330
338c. Dunah, A. W., Jeong, H., Griffin, A., Kim, Y.-M., Standaert, D. G., Hersch, S. M., Mouradian, M. M., Young, A. B., Tanese, N., and Krainc, D. (2002) *Science* **296**, 2238 – 2243
339. Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S. C. (1992) *Nature (London)* **356**, 408 – 414
340. Ptashne, M., and Gann, A. (1997) *Nature (London)* **386**, 569 – 577
341. Rodgers, K. K., and Coleman, J. E. (1994) *Protein Sci.* **3**, 608 – 619
342. Lohr, D., Venkov, P., and Zlatanova, J. (1995) *FASEB J.* **9**, 777 – 787
343. Corton, J. C., Moreno, E., and Johnston, S. A. (1998) *J. Biol. Chem.* **273**, 13776 – 13780
344. Ruden, D. M., Ma, J., and Ptashne, M. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4262 – 4266
345. Sigler, P. B. (1988) *Nature (London)* **333**, 210 – 212
346. Ryu, S., and Tjian, R. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7137 – 7142
347. Zawel, L., and Reinberg, D. (1995) *Ann. Rev. Biochem.* **64**, 533 – 561
348. Malik, S., and Roeder, R. G. (2000) *Trends Biochem. Sci.* **25**, 277 – 283
349. Svejstrup, J. Q., Li, Y., Fellows, J., Gnatt, A., Bjorklund, S., and Kornberg, R. D. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6075 – 6078
350. Spåhr, H., Bève, J., Larsson, T., Bergström, J., Karlsson, K.-A., and Gustafsson, C. M. (2000) *J. Biol. Chem.* **275**, 1351 – 1356
350a. Dotson, M. R., Yuan, C. X., Roeder, R. G., Myers, L. C., Gustafsson, C. M., Jiang, Y. W., Li, Y., Kornberg, R. D., and Asturias, F. J. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14307 – 14310
350b. Liu, Y., Ranish, J. A., Aebersold, R., and Hahn, S. (2001) *J. Biol. Chem.* **276**, 7169 – 7175
351. Tan, R. C., Truong, T. N., McCammon, J. A., and Sussman, J. L. (1993) *Biochemistry* **32**, 401 – 403
352. Clarke, M. F., FitzGerald, P. C., Brubaker, J. M., and Simpson, R. T. (1985) *J. Biol. Chem.* **260**, 12394 – 12397
353. Ondek, B., Gloss, L., and Herr, W. (1988) *Nature (London)* **333**, 40 – 45
354. Plon, S., and Wang, J. (1986) *Cell* **45**, 575 – 580
355. Ptashne, M. (1986) *Nature (London)* **322**, 697 – 701
356. Ross, E. D., Keating, A. M., and Maher, L. J., III. (2000) *J. Mol. Biol.* **297**, 321 – 334
357. Udvardy, A. (1999) *EMBO J.* **18**, 1 – 8
357a. Struhl, K. (2001) *Science* **293**, 1054 – 1055
357b. Yie, J., Senger, K., and Thanos, D. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13108 – 13113
358. Werner, M. H., Gronenborn, A. M., and Clore, G. M. (1996) *Science* **271**, 778 – 784
359. Cronk, J. D., Endrizzi, J. A., and Alber, T. (1996) *Protein Sci.* **5**, 1963 – 1972
360. Keller, W., König, P., and Richmond, T. J. (1995) *J. Mol. Biol.* **254**, 657 – 667
361. Hanson, R. W. (1998) *J. Biol. Chem.* **273**, 28543
362. Gardner, K., Moore, T. C., Davis-Smyth, T., Krutzsch, H., and Levens, D. (1994) *J. Biol. Chem.* **269**, 32963 – 32971
363. Lekstrom-Himes, J., and Xanthopoulos, K. G. (1998) *J. Biol. Chem.* **273**, 28545 – 28548
364. Turner, R., and Tjian, R. (1989) *Science* **243**, 1689 – 1694
365. Okahata, Y., Niikura, K., Sugiura, Y., Sawada, M., and Morii, T. (1998) *Biochemistry* **37**, 5666 – 5672
366. Maity, S. N., and de Crombrugghe, B. (1998) *Trends Biochem. Sci.* **23**, 174 – 178
367. Croniger, C., Leahy, P., Reshef, L., and Hanson, R. W. (1998) *J. Biol. Chem.* **273**, 31629 – 31632
368. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) *Science* **240**, 1759 – 1764
369. Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) *Nature (London)* **363**, 38 – 44
370. Chen, L., Glover, J. N. M., Hogan, P. G., Rao, A., and Harrison, S. C. (1998) *Nature (London)* **392**, 42 – 48
371. Mackay, J. P., Shaw, G. L., and King, G. F. (1996) *Biochemistry* **35**, 4867 – 4877
372. Gentz, R., Rauscher, F. J., III, Abate, C., and Curran, T. (1989) *Science* **243**, 1695 – 1699
373. Dang, C. V., McGuire, M., Buckmire, M., and Lee, W. M. F. (1989) *Nature (London)* **337**, 664 – 666
374. Garlatti, M., Tchesnokov, V., Daheshia, M., Feilleux-Duché, S., Hanoune, J., Aggerbeck, M., and Barouki, R. (1993) *J. Biol. Chem.* **268**, 6567 – 6574
375. Umek, R. M., Friedman, A. D., and McKnight, S. L. (1991) *Science* **251**, 288 – 292
376. Liberati, C., di Silvio, A., Ottolenghi, S., and Mantovani, R. (1999) *J. Mol. Biol.* **285**, 1441 – 1455
377. Grandori, C., and Eisenman, R. N. (1997) *Trends Biochem. Sci.* **22**, 177 – 181
378. Noguchi, K., Kitanaka, C., Yamana, H., Kokubu, A., Mochizuki, T., and Kuchino, Y. (1999) *J. Biol. Chem.* **274**, 32580 – 32587
379. Grigorieva, I., Grigoriev, V. G., Rowney, M. K., and Hoover, R. G. (2000) *J. Biol. Chem.* **275**, 7343 – 7350
379a. Otsuki, Y., Tanaka, M., Kamo, T., Kitanaka, C., Kuchino, Y. and Sugimura, H. (2002) *J. Biol. Chem.* In press 206733200
380. Lalli, E., and Sassone-Corsi, P. (1994) *J. Biol. Chem.* **269**, 17359 – 17362
381. Shaywitz, A. J., and Greenberg, M. E. (1999) *Ann. Rev. Biochem.* **68**, 821 – 861
382. De Cesare, D., Fimia, G. M., and Sassone-Corsi, P. (1999) *Trends Biochem. Sci.* **24**, 281 – 284
382a. Goren, I., Tavor, E., Goldblum, A., and Honigman, A. (2001) *J. Mol. Biol.* **313**, 695 – 709
382b. Nagadoi, A., Nakazawa, K.-i, Uda, H., Okuno, K., Maekawa, T., Ishii, S., and Nishimura, Y. (1999) *J. Mol. Biol.* **287**, 593 – 607
382c. Yeagley, D., Agati, J. M., and Quinn, P. G. (1998) *J. Biol. Chem.* **273**, 18743 – 18750

References

382d. Vo, N., and Goodman, R. H. (2001) *J. Biol. Chem.* **276**, 13505–13508
383. Xu, W., Chen, H., Du, K., Asahara, H., Tini, M., Emerson, B. M., Montminy, M., and Evans, R. M. (2001) *Science* **294**, 2507–2511
383a. Martínez-Balbás, M. A., Bauer, U.-M., Nielsen, S., Brehm, A., and Kouzarides, T. (2000) *EMBO J.* **19**, 662–671
383b. Classon, M., Salama, S., Gorka, C., Mulloy, R., Braun, P., and Harlow, E. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10820–10825
383c. Ogawa, H., Ishiguro, K.-i, Gaubatz, S., Livingston, D. M., and Nakatani, Y. (2002) *Science* **296**, 1132–1136
383d. La Thangue, N. B. (2002) *Science* **296**, 1034–1035
383e. Chin, L., Pomerantz, J., and DePinho, R. A. (1998) *Trends Biochem. Sci.* **23**, 291–296
383f. Chang, L., and Karin, M. (2001) *Nature (London)* **410**, 37–40
383g. Weston, C. R., Lambright, D. G., and Davis, R. J. (2002) *Science* **296**, 2345, 2347
383h. Cantley, L. C. (2001) *Science* **292**, 2019–2021
383i. Janknecht, R., and Hunter, T. (1999) *Science* **284**, 443–444
383j. Horvath, C. M. (2000) *Trends Biochem. Sci.* **25**, 496–502
383k. ten Dijke, P., Miyazono, K., and Helden, C.-H. (2000) *Trends Biochem. Sci.* **25**, 64–70
383l. Wu, G., Chen, Y.-G., Ozdamar, B., Gyuricza, C. A., Chong, P. A., Wrana, J. L., Massagué, J., and Shi, Y. (2000) *Science* **287**, 92–97
384. Towle, H. C. (1995) *J. Biol. Chem.* **270**, 23235–23238
384a. Osborne, T. F. (2000) *J. Biol. Chem.* **275**, 32379–32382
384b. Duan, R., Xie, W., Burghardt, R. C., and Safe, S. (2001) *J. Biol. Chem.* **276**, 11590–11598
384c. Hassler, M., and Richmond, T. J. (2001) *EMBO J.* **20**, 3018–3028
385. Treisman, R. (1992) *Trends Biochem. Sci.* **17**, 423–426
386. Ling, Y., West, A. G., Roberts, E. C., Lakey, J. H., and Sharrocks, A. D. (1998) *J. Biol. Chem.* **273**, 10506–10514
387. Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C. F., Trent, J. M., Staudt, L. M., Hudson, J., Jr., Boguski, M. S., Lashkari, D., Shalon, D., Botstein, D., and Brown, P. O. (1999) *Science* **283**, 83–87
388. Kerrigan, L. A., Croston, G. E., Lira, L. M., and Kadonaga, J. T. (1991) *J. Biol. Chem.* **266**, 574–582
388a. Bieker, J. J. (2001) *J. Biol. Chem.* **276**, 34355–34358
389. Mackay, J. P., and Crossley, M. (1998) *Trends Biochem. Sci.* **23**, 1–4
390. Turner, J., and Crossley, M. (1999) *Trends Biochem. Sci.* **24**, 236–241
391. Clarke, N. D., and Berg, J. M. (1998) *Science* **282**, 2018–2022
392. Naltner, A., Ghaffari, M., Whitsett, J. A., and Yan, C. (2000) *J. Biol. Chem.* **275**, 56–62
393. Pizzi, S., Dieci, G., Frigeri, P., Piccoli, G., Stocchi, V., and Ottonello, S. (1999) *J. Biol. Chem.* **274**, 2539–2548
394. Laity, J. H., Dyson, H. J., and Wright, P. E. (2000) *J. Mol. Biol.* **295**, 719–727
395. Klug, A., and Rhodes, D. (1987) *Trends Biochem. Sci.* **12**, 464–469
396. Nolte, R. T., Conlin, R. M., Harrison, S. C., and Brown, R. S. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2938–2943
397. Muro-Pastor, M. I., Gonzalez, R., Strauss, J., Narendja, F., and Scazzocchio, C. (1999) *EMBO J.* **18**, 1584–1597
397a. Molkentin, J. D. (2000) *J. Biol. Chem.* **275**, 38949–38952
398. Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J., and Gronenborn, A. M. (1993) *Science* **261**, 438–446
399. Haas, H., Zadra, I., Stöffler, G., and Angermayr, K. (1999) *J. Biol. Chem.* **274**, 4613–4619
400. Van Winkle, L. J. (1999) *Biomembrane Transport*, Academic Press, San Diego, California
401. Zeng, J., Vallee, B. L., and Kägi, J. H. R. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9984–9988
402. Lawson, C. L., and Carey, J. (1993) *Nature (London)* **366**, 178–182
403. Marsden, I., Jin, C., and Liao, X. (1998) *J. Mol. Biol.* **278**, 293–299
404. Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993) *Nature (London)* **364**, 412–420
405. Pio, F., Kodandapani, R., Ni, C.-Z., Shepard, W., Klemsz, M., McKercher, S. R., Maki, R. A., and Ely, K. R. (1996) *J. Biol. Chem.* **271**, 23329–23337
406. Wasylyk, B., Hagman, J., and Gutierrez-Hartmann, A. (1998) *Trends Biochem. Sci.* **23**, 213–216
406a. Alvarez, B., Martínez-A, C., Burgering, B. M. T., and Carrera, A. C. (2001) *Nature (London)* **413**, 744–747
407. Tisné, C., Delepierre, M., and Hartmann, B. (1999) *J. Mol. Biol.* **293**, 139–150
408. Barnes, P. J., and Karin, M. (1997) *N. Engl. J. Med.* **336**, 1066–1071
408a. Nourbakhsh, M., Kälble, S., Dörrie, A., Hauser, H., Resch, K., and Kracht, M. (2001) *J. Biol. Chem.* **276**, 4501–4508
408b. Tam, W. F., and Sen, R. (2001) *J. Biol. Chem.* **276**, 7701–7704
408c. Schmitz, M. L., Bacher, S., and Kracht, M. (2001) *Trends Biochem. Sci.* **26**, 186–190
409. Malek, S., Huxford, T., and Ghosh, G. (1998) *J. Biol. Chem.* **273**, 25427–25435
410. Chen, F. E., Huang, D.-B., Chen, Y.-Q., and Ghosh, G. (1998) *Nature (London)* **391**, 410–413
411. Shlyakhtenko, L. S., Hsieh, P., Grigoriev, M., Potaman, V. N., Sinden, R. R., and Lyubchenko, Y. L. (2000) *J. Mol. Biol.* **296**, 1169–1173
412. Murphy, F. V., IV, Sweet, R. M., and Churchill, M. E. A. (1999) *EMBO J.* **18**, 6610–6618
413. Sutrias-Grau, M., Bianchi, M. E., and Bernusés, J. (1999) *J. Biol. Chem.* **274**, 1628–1634
414. Yie, J., Merika, M., Munshi, N., Chen, G., and Thanos, D. (1999) *EMBO J.* **18**, 3074–3089
414a. Lewis, R. T., Andreucci, A., and Nikolajczyk, B. S. (2001) *J. Biol. Chem.* **276**, 9550–9557
415. Bergel, M., Herrera, J. E., Thatcher, B. J., Prymakowska-Bosak, M., Vassilev, A., Nakatani, Y., Martin, B., and Bustin, M. (2000) *J. Biol. Chem.* **275**, 11514–11520
415a. Cowell, I. G. (1994) *Trends Biochem. Sci.* **19**, 38–42
415b. Harrington, R. H., and Sharma, A. (2001) *J. Biol. Chem.* **276**, 104–113
416. Zhang, J., Ou, J., Bashmakov, Y., Horton, J. D., Brown, M. S., and Goldstein, J. L., (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3756–3761
416a. Olefsky, J. M. (2001) *J. Biol. Chem.* **276**, 36863–36864
416b. Khorasanizadeh, S., and Rastinejad, F. (2001) *Trends Biochem. Sci.* **26**, 384–390
416c. Valverde, M. A., and Parker, M. G. (2002) *Trends Biochem. Sci.* **27**, 172–173
417. Vázquez, F., Rodríguez-Manzaneque, J. C., Lydon, J. P., Edwards, D. P., O'Mally, B. W., and Iruela-Arispe, M. L. (1999) *J. Biol. Chem.* **274**, 2185–2192
418. Lehmann, M., and Korge, G. (1995) *EMBO J.* **14**, 716–726
418a. Wurtz, J.-M., Guillot, B., Fagart, J., Moras, D., Tietjen, K., and Schindler, M. (2000) *Protein Sci.* **9**, 1073–1084
419. Guschin, D., Chandler, S., and Wolffe, A. P. (1998) *Biochemistry* **37**, 8629–8636
420. Baudino, T. A., Kraichely, D. M., Jefcoat, S. C., Jr., Winchester, S. K., Partridge, N. C., and MacDonald, P. N. (1998) *J. Biol. Chem.* **273**, 16434–16441
420a. Makishima, M., Lu, T. T., Xie, W., Whitfield, G. K., Domoto, H., Evans, R. M., Haussler, M. R., and Mangelsdorf, D. J. (2002) *Science* **296**, 1313–1316
421. Beato, M. (1991) *FASEB J.* **5**, 2044–2051
422. Weatherman, R. V., Fletterick, R. J., and Scanlan, T. S. (1999) *Ann. Rev.Biochem.* **68**, 559–581
423. Di Croce, L., Okret, S., Kersten, S., Gustafsson, J.-Å., Parker, M., Wahli, W., and Beato, M. (1999) *EMBO J.* **18**, 6201–6210
424. Avram, D., Fields, A., Pretty On Top, K., Nevrivy, D. J., Ishmael, J. E., and Leid, M. (2000) *J. Biol. Chem.* **275**, 10315–10322
425. Kliewer, S. A., Lehmann, J. M., and Willson, T. M. (1999) *Science* **284**, 757–760
426. Kosztin, D., Bishop, T. C., and Schulten, K. (1997) *Biophys. J.* **73**, 557–570
427. Driscoll, M. D., Sathya, G., Muyan, M., Klinge, C. M., Hilf, R., and Bambara, R. A. (1998) *J. Biol. Chem.* **273**, 29321–29330
427a. Campbell, R. A., Bhat-Nakshatri, P., Patel, N. M., Constantinidou, D., Ali, S., and Nakshatri, H. (2001) *J. Biol. Chem.* **276**, 9817–9824
427b. An, J., Tzagarakis-Foster, C., Scharaschmidt, T. C., Lomri, N., and Leitman, D. C. (2001) *J. Biol. Chem.* **276**, 17808–17814
428. Rogatsky, I., Trowbridge, J. M., and Garabedian, M. J. (1999) *J. Biol. Chem.* **274**, 22296–22302
429. Kosano, H., Stensgard, B., Charlesworth, M. C., McMahon, N., and Toft, D. (1998) *J. Biol. Chem.* **273**, 32973–32979
429a. Bennett, M. K., and Osborne, T. F. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6340–6344
429b. Luong, A., Hannah, V. C., Brown, M. S., and Goldstein, J. L. (2000) *J. Biol. Chem.* **275**, 26458–26466
430. Clarke, S. D., and Abraham, S. (1992) *FASEB J.* **6**, 3146–3152
431. Lu, C.-A., Lim, E.-K., and Yu, S.-M. (1998) *J. Biol. Chem.* **273**, 10120–10131
432. Cereghino, G. P., and Scheffler, I. E. (1996) *EMBO J.* **15**, 363–374
433. Hong, S.-P., Piper, M. D., Sinclair, D. A., and Dawes, I. W. (1999) *J. Biol. Chem.* **274**, 10523–10532
434. Svaren, J., and Hörz, W. (1997) *Trends Biochem. Sci.* **22**, 93–97
435. Shimizu, T., Toumoto, A., Ihara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y., and Hakoshima, T. (1997) *EMBO J.* **16**, 4689–4697
436. Addess, K. J., Basilion, J. P., Klausner, R. D., Rouault, T. A., and Pardi, A. (1997) *J. Mol. Biol.* **274**, 72–83
437. Wardrop, S. L., Watts, R. N., and Richardson, D. R. (2000) *Biochemistry* **39**, 2748–2758
438. Ke, Y., Sierzputowska-Gracz, H., Gdaniec, Z., and Theil, E. C. (2000) *Biochemistry* **39**, 6235–6242
438a. Meehan, H. A., and Connell, G. J. (2001) *J. Biol. Chem.* **276**, 14791–14796
439. Quinn, J. M., Barraco, P., Eriksson, M., and Merchant, S. (2000) *J. Biol. Chem.* **275**, 6080–6089
440. Hall, D. R., Gourley, D. G., Leonard, G. A., Duke, E. M. H., Anderson, L. A., Boxer, D. H., and Hunter, W. N. (1999) *EMBO J.* **18**, 1435–1446
441. Okamura, H., Hanaoka, S., Nagadoi, A., Makino, K., and Nishimura, Y. (2000) *J. Mol. Biol.* **295**, 1225–1236
442. Ritossa, F. (1962) *Experientia* **18**, 571–573

References

443. Schlesinger, M. J. (1990) *J. Biol. Chem.* **265**, 12111 – 12114
444. Ang, D., Liberek, K., Skowyra, D., Zylicz, M., and Georgopoulos, C. (1991) *J. Biol. Chem.* **266**, 24233 – 24236
445. Morimoto, R. I. (1993) *Science* **259**, 1409 – 1410
446. Giardina, C., and Lis, J. T. (1995) *J. Biol. Chem.* **270**, 10369 – 10372
447. Kimpel, J. A., and Key, J. L. (1985) *Trends Biochem. Sci.* **10**, 353 – 357
448. Craig, E. A., and Gross, C. A. (1991) *Trends Biochem. Sci.* **16**, 135 – 140
449. Welch, W. J. (1993) *Sci. Am.* **268**(May), 57 – 64
450. Kim, K. K., Kim, R., and Kim, S.-H. (1998) *Nature (London)* **394**, 595 – 599
451. Korber, P., Stahl, J. M., Nierhaus, K. H., and Bardwell, J. C. A. (2000) *EMBO J.* **19**, 741 – 748
452. Goldenberg, C. J., Luo, Y., Fenna, M., Baler, R., Weinmann, R., and Voellmy, R. (1988) *J. Biol. Chem.* **263**, 19734 – 19739
453. Harrison, C. J., Bohm, A. A., and Nelson, H. C. M. (1994) *Science* **263**, 224 – 227
454. Larson, J. S., Schuetz, T. J., and Kingston, R. E. (1995) *Biochemistry* **34**, 1902 – 1911
455. Wyman, C., Grotkopp, E., Bustamante, C., and Nelson, H. C. M. (1995) *EMBO J.* **14**, 117 – 123
456. Rouvière, P. E., De Las Penas, A., Mecsas, J., Lu, C. Z., Rudd, K. E., and Gross, C. A. (1995) *EMBO J.* **14**, 1032 – 1042

456a. Dartigalongue, C., Missiakas, D., and Raina, S. (2001) *J. Biol. Chem.* **276**, 20866 – 20875

457. Brevet, A., Plateau, P., Best-Belpomme, M., and Blanquet, S. (1985) *J. Biol. Chem.* **260**, 15566 – 15570
458. Wahab, S. Z., and Yang, D. C. H. (1985) *J. Biol. Chem.* **260**, 5286 – 5289
459. Harnett, S. P., Lowe, G., and Tansley, G. (1985) *Biochemistry* **24**, 2908 – 2915
460. Lee, P. C., Bochner, B. R., and Ames, B. N. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7496 – 7500
461. Guédon, G. F., Gilson, G. J. P., Ebel, J. P., Befort, N. M.-T., and Remy, P. M. (1986) *J. Biol. Chem.* **261**, 16459 – 16465
462. Kodym, R., Calkins, P., and Story, M. (1999) *J. Biol. Chem.* **274**, 5131 – 5137
463. Freestone, P., Nyström, T., Trinei, M., and Norris, V. (1997) *J. Mol. Biol.* **274**, 318 – 324
464. Discher, D. J., Bishopric, N. H., Wu, X., Peterson, C. A., and Webster, K. A. (1998) *J. Biol. Chem.* **273**, 26087 – 26093
465. Sundaresan, M., Yu, Z.-X., Ferrans, V. J., Irani, K., and Finkel, T. (1995) *Science* **270**, 296 – 299

465a. Thannickal, V. J., and Fanburg, B. L. (1995) *J. Biol. Chem.* **270**, 30334 – 30338

466. Burg, M. B., Kwon, E. D., and Kültz, D. (1996) *FASEB J.* **10**, 1598 – 1606
467. Ramagopal, S. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 94 – 98

467a. Dagsgaard, C., Taylor, L. E., O'Brien, K. M., and Poyton, R. O. (2001) *J. Biol. Chem.* **276**, 7593 – 7601

467b. Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J. M., Lane, W. S., and Kaelin, W. G., Jr. (2001) *Science* **292**, 464 – 468

467c. Zhu, H., and Bunn, H. F. (2001) *Science* **292**, 449,451

467d. Lando, D., Peet, D. J., Whelan, D. A., Gorman, J. J., and Whitelaw, M. L. (2002) *Science* **295**, 858 – 861

467e. Srinivas, V., Leshchinsky, I., Sang, N., King, M. P., Minchenko, A., and Caro, J. (2001) *J. Biol. Chem.* **276**, 21995 – 21998

467f. Jaakkola, P., Mole, D. R., Tian, Y.-M., Wilson, M. I., Gielbert, J., Gaskell, S. J., von Kriegsheim, A., Hebestreit, H. F., Mukherji, M., Schofield, C. J., Maxwell, P. H., Pugh, C. W., and Ratcliffe, P. J. (2001) *Science* **292**, 468 – 472

467g. Min, J.-H., Yang, H., Ivan, M., Gertler, F., Kaelin, W. G., Jr., and Pavletich, N. P. (2002) *Science* **296**, 1886 – 1889

467h. Semenza, G. L. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11570 – 11572

467i. Hansson, L. O., Friedler, A., Freund, S., Rüdiger, S., and Fersht, A. R. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10305 – 10309

467j. Stebbins, C. E., Kaelin, W. G., Jr., and Pavletich, N. P. (1999) *Science* **284**, 455 – 461

467k. Baysal, B. E., Ferrell, R. E., Willett-Brozick, J. E., Lawrence, E. C., Myssiorek, D., Bosch, A., van der Mey, A., Taschner, P. E. M., Rubinstein, W. S., Myers, E. N., Richard, C. W., III, Cornelisse, C. J., Devilee, P., and Devlin, B. (2000) *Science* **287**, 848 – 851

468. Kim, M., Thum, K. E., Morishige, D. T., and Mullet, J. E. (1999) *J. Biol. Chem.* **274**, 4684 – 4692
469. Martínez-García, J. F., Huq, E., and Quail, P. H. (2000) *Science* **288**, 859 – 863
470. Wright, C. V. E., Cho, K. W. Y., Oliver, G., and De Robertis, E. M. (1989) *Trends Biochem. Sci.* **14**, 52 – 56
471. Scott, M. P. (1987) *Ann. Rev. Biochem.* **56**, 195 – 227
472. Mihara, H., and Kaiser, E. T. (1988) *Science* **242**, 925 – 927
473. Ford, E., and Hernandez, N. (1997) *J. Biol. Chem.* **272**, 16048 – 16055
474. Gstaiger, M., Georgiev, O., van Leeuwen, H., van der Vliet, P., and Schaffner, W. (1996) *EMBO J.* **15**, 2781 – 2790
475. Lee, L., Stollar, E., Chang, J., Grossmann, J. G., O'Brien, R., Ladbury, J., Carpenter, B., Roberts, S. and Luisi, B. (2001) *Biochemistry* **40**, 6580 – 6588

475a. Lundbäck, T., Chang, J.-F., Phillips, K., Luisi, B., and Ladbury, J. E. (2000) *Biochemistry* **39**, 7570 – 7579

476. Robertson, M. (1988) *Nature (London)* **336**, 522 – 524

476a. Diamond, S. E., and Gutierrez-Hartmann, A. (2000) *J. Biol. Chem.* **275**, 30977 – 30986

477. Tanaka, N., Kato, H., Ishikawa, Y., Hisatake, K., Tashiro, K., Kominami, R., and Muramatsu, M. (1990) *J. Biol. Chem.* **265**, 13836 – 13842
478. Kahl, B. F., Li, H., and Paule, M. R. (2000) *J. Mol. Biol.* **299**, 75 – 89

478a. Miller, G., Panov, K. I., Friedrich, J. K., Trinkle-Mulcahy, L., Lamond, A. I., and Zomerdijk, J. C. B. M. (2001) *EMBO J.* **20**, 1373 – 1382

479. McBryant, S. J., Kassavetis, G. A., and Gottesfeld, J. M. (1995) *J. Mol. Biol.* **250**, 315 – 326
480. Gunnery, S., Ma, Y., and Mathews, M. B. (1999) *J. Mol. Biol.* **286**, 745 – 757
481. Pieler, T., and Theunissen, O. (1993) *Trends Biochem. Sci.* **18**, 226 – 230
482. Miller, J., McLachlan, A. D., and Klug, A. (1985) *EMBO J.* **4**, 1609 – 1614
483. Veldhoen, N., You, Q., Setzer, D. R., and Romaniuk, P. J. (1994) *Biochemistry* **33**, 7568 – 7575
484. Gabrielsen, O. S., and Sentenac, A. (1991) *Trends Biochem. Sci.* **16**, 412 – 416
485. Andrau, J.-C., Sentenac, A., and Werner, M. (1999) *J. Mol. Biol.* **288**, 511 – 520
486. Young, L. S., Dunstan, H. M., Witte, P. R., Smith, T. P., Ottonello, S., and Sprague, K. U. (1991) *Science* **252**, 542 – 546
487. Choder, M., and Aloni, Y. (1988) *J. Biol. Chem.* **263**, 12994 – 13002
488. Lewis, J. D., and Tollervey, D. (2000) *Science* **288**, 1385 – 1389
489. Reines, D., Conaway, J. W., and Conaway, R. C. (1996) *Trends Biochem. Sci.* **21**, 351 – 355
490. Gnatt, A., Fu, J., and Kornberg, R. D. (1997) *J. Biol. Chem.* **272**, 30799 – 30805
491. Izban, M. G., Parsons, M. A., and Sinden, R. R. (1998) *J. Biol. Chem.* **273**, 27009 – 27016
492. Hemming, S. A., and Edwards, A. M. (2000) *J. Biol. Chem.* **275**, 2288 – 2294
493. Nelson, P. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14342 – 14347
494. Botuyan, M. V., Koth, C. M., Mer, G., Chakrabartty, A., Conaway, J. W., Conaway, R. C., Edwards, A. M., Arrowsmith, C. H., and Chazin, W. J. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9033 – 9038
495. Conaway, J. W., and Conaway, R. C. (1999) *Ann. Rev. Biochem.* **68**, 301 – 319
496. Iyer, V., and Struhl, K. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5208 – 5212
497. Guo, Z., and Sherman, F. (1996) *Trends Biochem. Sci.* **21**, 477 – 481
498. Wahle, E., and Keller, W. (1996) *Trends Biochem. Sci.* **21**, 247 – 250
499. Chanfreau, G., Noble, S. M., and Guthrie, C. (1996) *Science* **274**, 1511 – 1514
500. Hirose, Y., and Manley, J. L. (1997) *J. Biol. Chem.* **272**, 29636 – 29642
501. Ohnacker, M., Barabino, S. M. L., Preker, P. J., and Keller, W. (2000) *EMBO J.* **19**, 37 – 47
502. Keller, R. W., Kühn, U., Aragón, M., Bornikova, L., Wahle, E., and Bear, D. G. (2000) *J. Mol. Biol.* **297**, 569 – 583
503. Reeder, R. H., and Lang, W. H. (1997) *Trends Biochem. Sci.* **22**, 473 – 477
504. Syroid, D. E., and Capone, J. P. (1994) *J. Mol. Biol.* **244**, 482 – 493

504a. Ohndorf, U.-M., Steegborn, C., Knijff, R., and Sondermann, P. (2001) *J. Biol. Chem.* **276**, 27188 – 27196

504b. Wang, Z., Bai, L., Hsieh, Y.-J., and Roeder, R. G. (2000) *EMBO J.* **19**, 6823 – 6832

505. Draper, D. E. (1996) *Trends Biochem. Sci.* **21**, 145 – 149
506. Moore, P. B. (1999) *Ann. Rev. Biochem.* **67**, 287 – 300
507. Hermann, T., and Patel, D. J. (1999) *J. Mol. Biol.* **294**, 829 – 849
508. Pan, J., and Woodson, S. A. (1999) *J. Mol. Biol.* **294**, 955 – 965

508a. Zhuang, X., Bartley, L. E., Babcock, H. P., Russell, R., Ha, T., Herschlag, D., and Chu, S. (2000) *Science* **288**, 2048 – 2051

508b. Kent, O., Chaulk, S. G., and MacMillan, A. M. (2000) *J. Mol. Biol.* **304**, 699 – 705

509. Murthy, V. L., Srinivasan, R., Draper, D. E., and Rose, G. D. (1999) *J. Mol. Biol.* **291**, 313 – 327
510. Michel, F., and Westhof, E. (1996) *Science* **273**, 1676 – 1677
511. Gultyaev, A. P., van Batenburg, F. H. D., and Pleij, C. W. A. (1995) *J. Mol. Biol.* **250**, 37 – 51
512. Narlikar, G. J., Bartley, L. E., Khosla, M., and Herschlag, D. (1999) *Biochemistry* **38**, 14192 – 14204
513. Sclavi, B., Sullivan, M., Chance, M. R., Brenowitz, M., and Woodson, S. A. (1998) *Science* **279**, 1940 – 1943
514. Silverman, S. K., and Cech, T. R. (1999) *Biochemistry* **38**, 8691 – 8702
515. Rivas, E., and Eddy, S. R. (1999) *J. Mol. Biol.* **285**, 2053 – 2068
516. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science* **273**, 1696 – 1699
517. Biswas, R., Wahl, M. C., Ban, C., and Sundaralingam, M. (1997) *J. Mol. Biol.* **267**, 1149 – 1156

517a. Deng, J., Xiong, Y., and Sundaralingam, M. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13665 – 13670

517b. Snoussi, K., Nonin-Lecomte, S., and Leroy, J.-L. (2001) *J. Mol. Biol.* **309**, 139 – 153

517c. Schneider, C., Brandl, M., and Sühnel, J. (2001) *J. Mol. Biol.* **305**, 659 – 667

References

517d. Soukup, J. K., Minakawa, N., Matsuda, A., and Strobel, S. A. (2002) *Biochemistry* **41**, 10426–10438
517e. Stormo, G. D., and Ji, Y. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9465–9467
518. Costa, M., and Michel, F. (1997) *EMBO J.* **16**, 3289–3302
519. Nagai, K., Oubridge, C., Ito, N., Avis, J., and Evans, P. (1995) *Trends Biochem. Sci.* **20**, 235–240
520. Bell, S. P., Learned, R. M., Jantzen, H.–M., and Tjian, R. (1988) *Science* **241**, 1192–1197
521. Schmitz, M. L., Maier, U.-G., Brown, J. W. S., and Feix, G. (1989) *J. Biol. Chem.* **264**, 1467–1472
522. Baker, S. M., and Platt, T. (1986) *Cell* **47**, 839–840
523. Read, C., Larose, A.-M., Leblanc, B., Bannister, A. J., Firek, S., Smith, D. R., and Moss, T. (1992) *J. Biol. Chem.* **267**, 10961–10967
524. Sommerville, J. (1985) *Nature (London)* **318**, 410–411
525. Morrissey, J. P., and Tollervey, D. (1995) *Trends Biochem. Sci.* **20**, 78–82
526. Borovjagin, A. V., and Gerbi, S. A. (1999) *J. Mol. Biol.* **286**, 1347–1363
527. Wellauer, P. K., and Dawid, I. B. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 2827–2831
528. Dawid, I. B., and Wellauer, P. K. (1976) *Cell* **8**, 443–448
529. Ghisolfi, L., Joseph, G., Amalric, F., and Erard, M. (1992) *J. Biol. Chem.* **267**, 2955–2959
529a. Bouvet, P., Allain, F. H.-T., Finger, L. D., Dieckmann, T., and Feigon, J. (2001) *J. Mol. Biol.* **309**, 763–775
529b. Allain, F. H.-T., Bouvet, P., Dieckmann, T., and Feigon, J. (2000) *EMBO J.* **19**, 6870–6881
530. Srivastava, M., McBride, O. W., Fleming, P. J., Pollard, H. B., and Burns, A. L. (1990) *J. Biol. Chem.* **265**, 14922–14931
531. Tycowaki, K. T., Shu, M.–D., and Steitz, J. A. (1994) *Science* **266**, 1558–1561
531a. Pederson, T. (2002) *Trends Biochem. Sci.* **27**, 111–112
532. Guthrie, C. (1986) *Trends Biochem. Sci.* **11**, 430–434
533. Birnstiel, M. L., ed. (1988) *Small Nuclear Ribonucleoprotein Particles*, Springer, Vienna/New York
534. Tarn, W.-Y., and Steitz, J. A. (1997) *Trends Biochem. Sci.* **22**, 132–137
535. Séraphin, B. (1995) *EMBO J.* **14**, 2089–2098
535a. Filipowicz, W. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14035–14037
535b. Hüttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J.-P., and Brosius, J. (2001) *EMBO J.* **20**, 2943–2953
535c. Lowe, T. M., and Eddy, S. R. (1999) *Science* **283**, 1168–1171
535d. Xu, Y.-x, Liu, L., Lopez-Estrano, C., and Michaeli, S. (2001) *J. Biol. Chem.* **276**, 14289–14298
535e. Barneche, F., Gaspin, C., Guyot, R., and Echeverría, M. (2001) *J. Mol. Biol.* **311**, 57–73
536. Feeney, R. J., Sauterer, R. A., Feeney, J. L., and Zieve, G. W. (1989) *J. Biol. Chem.* **264**, 5776–5783
537. Hartmuth, K., Raker, V. A., Huber, J., Branlant, C., and Lührmann, R. (1999) *J. Mol. Biol.* **285**, 133–147
537a. Collins, B. M., Harrop, S. J., Kornfeld, G. D., Dawes, I. W., Curmi, P. M. G., and Mabbutt, B. C. (2001) *J. Mol. Biol.* **309**, 915–923
537b. Baulcombe, D. (2002) *Science* **297**, 2002–2003
537c. Hutvágner, G., and Zamore, P. D. (2002) *Science* **297**, 2056–2060
538. Parry, H. D., Scherly, D., and Mattaj, I. W. (1989) *Trends Biochem. Sci.* **14**, 15–19
539. Grabowski, P. J., and Sharp, P. A. (1986) *Science* **233**, 1294–1299
539a. Hannon, G. J. (2002) *Nature (London)* **418**, 244–251
540. Sollner-Webb, B., and Tower, J. (1986) *Ann. Rev. Biochem.* **55**, 801–830
541. Fournier, M. J., and Maxwell, E. S. (1993) *Trends Biochem. Sci.* **18**, 131–135
541a. Saveanu, C., Bienvenu, D., Namane, A., Gleizes, P.-E., Gas, N., Jacquier, A., and Fromont-Racine, M. (2001) *EMBO J.* **20**, 6475–6484
541b. Gadal, O., Strauss, D., Braspenning, J., Hoepfner, D., Petfalski, E., Philippsen, P., Tollervey, D., and Hurt, E. (2001) *EMBO J.* **20**, 3695–3704
541c. Jäkel, S., and Görlich, D. (1998) *EMBO J.* **17**, 4491–4502
542. Beltrame, M., and Tollervey, D. (1995) *EMBO J.* **14**, 4350–4356
543. Hughes, J. M. X. (1996) *J. Mol. Biol.* **259**, 645–654
544. Méreau, A., Fournier, R., Grégoire, A., Mougin, A., Fabrizio, P., Lührmann, R., and Branlant, C. (1997) *J. Mol. Biol.* **273**, 552–571
545. Hengst, J. A., Georgoff, I., Isom, H. C., and Jacob, S. T. (1988) *J. Biol. Chem.* **263**, 19270–19273
546. van Nues, R. W., Rientjes, J. M. J., Morré, S. A., Mollee, E., Planta, R. J., Vanema, J., and Raué, H. A. (1995) *J. Mol. Biol.* **250**, 24–36
546a. Borovjagin, A. V., and Gerbi, S. A. (2000) *J. Mol. Biol.* **300**, 57–74
546b. Peculis, B. A. (2001) *RNA* **7**, 207–219
547. Lygerou, Z., Allmang, C., Tollervey, D., and Séraphin, B. (1996) *Science* **272**, 268–270
547a. Clayton, D. A. (2001) *Nature (London)* **410**, 29, 31
548. Bachellerie, J.–P., Michot, B., Nicoloso, M., Balakin, A., Ni, J., and Fournier, M. J. (1995) *Trends Biochem. Sci.* **20**, 261–264
549. Wang, H., Boisvert, D., Kim, K. K., Kim, R., and Kim, S.-H. (2000) *EMBO J.* **19**, 317–323
550. Reddy, R., Henning, D., Das, G., Harless, M., and Wright, D. (1987) *J. Biol. Chem.* **262**, 75–81
551. Hernandez, N., and Weiner, A. M. (1986) *Cell* **47**, 249–258
552. Mimori, T., Hinterberger, M., Pettersson, I., and Steitz, J. A. (1984) *J. Biol. Chem.* **259**, 560–565
553. Rokeach, L. A., Haselby, J. A., and Hoch, S. O. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4832–4836
554. Deutscher, S. L., and Keene, J. D. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3299–3303
555. Frendeway, D., Dingermann, T., Cooley, L., and Söll, D. (1985) *J. Biol. Chem.* **260**, 449–454
556. Levinger, L., Vasisht, V., Greene, V., Bourne, R., Birk, A., and Kolla, S. (1995) *J. Biol. Chem.* **270**, 18903–18909
557. Pfeiffer, T., Tekos, A., Warnecke, J. M., Drainas, D., Engelke, D. R., Séraphin, B., and Hartmann, R. K. (2000) *J. Mol. Biol.* **298**, 559–565
558. Rossmanith, W. (1997) *J. Mol. Biol.* **265**, 365–371
558a. Doudna, J. A., and Cech, T. R. (2002) *Nature (London)* **418**, 222–228
558b. Xiao, S., Scott, F., Fierke, C. A., and Engelke, D. R. (2002) *Ann. Rev. Biochem.* **71**, 165–189
558c. Houser-Scott, F., Xiao, S., Millikin, C. E., Zengel, J. M., Lindahl, L., and Engelke, D. R. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2684–2689
559. Deutscher, M. D. (1984) *Crit. Revs. Biochem.* **17**, 45–71
559a. Li, F., Wang, J., and Steitz, T. A. (2000) *J. Mol. Biol.* **304**, 483–492
559b. Seth, M., Thurlow, D. L., and Hou, Y.-M. (2002) *Biochemistry* **41**, 4521–4532
560. Grosjean, H., Edqvist, J., Stråby, K. B., and Giegé, R. (1996) *J. Mol. Biol.* **255**, 67–85
561. Constantinesco, F., Motorin, Y., and Grosjean, H. (1999) *J. Mol. Biol.* **291**, 375–392
562. House, C. H., and Miller, S. L. (1996) *Biochemistry* **35**, 315–320
563. Lamond, A. I. (1990) *Trends Biochem. Sci.* **15**, 451–452
564. Schnierle, B. S., Gershon, P. D., and Moss, B. (1994) *J. Biol. Chem.* **269**, 20700–20706
565. Pillutla, R. C., Yue, Z., Maldonado, E., and Shatkin, A. J. (1998) *J. Biol. Chem.* **273**, 21443–21446
566. Terns, M. P., Grimm, C., Lund, E., and Dahlberg, J. E. (1995) *EMBO J.* **14**, 4860–4871
567. Shimba, S., and Reddy, R. (1994) *J. Biol. Chem.* **269**, 12419–12423
568. Booth, B. L., Jr., and Pugh, B. F. (1997) *J. Biol. Chem.* **272**, 984–991
569. Wang, S. P., Deng, L., Ho, C. K., and Shuman, S. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9573–9578
570. Kuge, H., and Richter, J. D. (1995) *EMBO J.* **14**, 6301–6310
571. Hu, G., Gershon, P. D., Hodel, A. E., and Quiocho, F. A. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7149–7154
571a. Hsu, P.-C., Hodel, M. R., Thomas, J. W., Taylor, L. J., Hagedorn, C. H., and Hodel, A. E. (2000) *Biochemistry* **39**, 13730–13736
571b. Martin, G., Keller, W., and Doublié, S. (2000) *EMBO J.* **19**, 4193–4203
571c. Hayes, R., Kudla, J., and Gruissem, W. (1999) *Trends Biochem. Sci.* **24**, 199–202
571d. Gagliardi, D., Perrin, R., Maréchal-Drouard, L., Grienenberger, J.-M., and Leaver, C. J. (2001) *J. Biol. Chem.* **276**, 43541–43547
572. Galili, G., Kawata, E. E., Smith, L. D., and Larkins, B. A. (1988) *J. Biol. Chem.* **263**, 5764–5770
573. Chen, C.-Y. A., and Shyu, A.-B. (1995) *Trends Biochem. Sci.* **20**, 465–470
574. Jacobson, A., and Peltz, S. W. (1996) *Ann. Rev. Biochem.* **65**, 693–739
575. Bashirullah, A., Halsell, S. R., Copperstock, R. L., Kloc, M., Karaiskakis, A., Fisher, W. W., Fu, W., Hamilton, J. K., Etkin, L. D., and Lipshitz, H. D. (1999) *EMBO J.* **18**, 2610–2620
575a. Matsumoto, K., Wassarman, K. M., and Wolffe, A. P. (1998) *EMBO J.* **17**, 2107–2121
575b. Bashirullah, A., Cooperstock, R. L., and Lipshitz, H. D. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7025–7028
576. Raghow, R. (1987) *Trends Biochem. Sci.* **12**, 358–360
577. Hunt, T. (1988) *Nature (London)* **334**, 567–568
578. Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M., and Séraphin, B. (2000) *EMBO J.* **19**, 1661–1671
579. Tharun, S., He, W., Mayes, A. E., Lennertz, P., Beggs, J. D., and Parker, R. (2000) *Nature (London)* **404**, 515–518
579a. Gao, M., Wilusz, C. J., Peltz, S. W., and Wilusz, J. (2001) *EMBO J.* **20**, 1134–1143
580. Phizicky, E. M., and Greer, C. L. (1993) *Trends Biochem. Sci.* **18**, 31–34
581. Cech, T. R., and Bass, B. L. (1986) *Ann. Rev. Biochem.* **55**, 599–629
582. Zaug, A. J., Grosshans, C. A., and Cech, T. R. (1988) *Biochemistry* **27**, 8924–8931
583. Saldanha, R., Mohr, G., Belfort, M., and Lambowitz, A. M. (1993) *FASEB J.* **7**, 15–24
584. Weeks, K. M., and Cech, T. R. (1995) *Biochemistry* **34**, 7728–7738
585. Ho, Y., and Waring, R. B. (1999) *J. Mol. Biol.* **292**, 987–1001
585a. Strauss-Soukup, J. K., and Strobel, S. A. (2000) *J. Mol. Biol.* **302**, 339–358
586. Bevilacqua, P. C., Sugimoto, N., and Turner, D. H. (1996) *Biochemistry* **35**, 648–658

References

587. Golden, B. L., and Cech, T. R. (1996) *Biochemistry* **35**, 3754–3763
588. Nowakowski, J., and Tinoco, I., Jr. (1996) *Biochemistry* **35**, 2577–2585
589. Michel, F., and Ferat, J.-L. (1995) *Ann. Rev. Biochem.* **64**, 435–461
590. Sharp, P. A., and Eisenberg, D. (1987) *Science* **238**, 729–730
590a. Chanfreau, G., and Jacquier, A. (1996) *EMBO J.* **15**, 3466–3476
590b. Swisher, J., Duarte, C. M., Su, L. J., and Pyle, A. M. (2001) *EMBO J.* **20**, 2051–2061
590c. Su, L. J., Qin, P. Z., Michels, W. J., and Pyle, A. M. (2001) *J. Mol. Biol.* **306**, 655–668
590d. Zhang, L., and Doudna, J. A. (2002) *Science* **295**, 2084–2088
591. Diener, T. O. (1984) *Trends Biochem. Sci.* **9**, 133–136
592. Daròs, J. A., Marcos, J. F., Hernández, C., and Flores, R. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12813–12817
593. Feldstein, P. A., Hu, Y., and Owens, R. A. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6560–6565
593a. Navarro, J.-A., and Flores, R. (2000) *EMBO J.* **19**, 2662–2670
594. Forster, A. C., Davies, C., Sheldon, C. C., Jeffries, A. C., and Symons, R. H. (1988) *Nature (London)* **334**, 265–267
595. Haseloff, J., and Gerlach, W. L. (1988) *Nature (London)* **334**, 585–591
596. Dinter-Gottlieb, G. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6250–6254
597. Greer, C. L., and Abelson, J. (1984) *Trends Biochem. Sci.* **9**, 139–141
598. Laski, F. A., Fire, A. Z., RajBhandary, U. L., and Sharp, P. A. (1983) *J. Biol. Chem.* **258**, 11974–11980
599. Phizicky, E. M., Schwartz, R. C., and Abelson, J. (1986) *J. Biol. Chem.* **261**, 2978–2986
600. McCraith, S. M., and Phizicky, E. M. (1991) *J. Biol. Chem.* **266**, 11986–11992
601. Belford, H. G., Westaway, S. K., Abelson, J., and Greer, C. L. (1993) *J. Biol. Chem.* **268**, 2444–2450
601a. Steiger, M. A., Kierzek, R., Turner, D. H., and Phizicky, E. M. (2001) *Biochemistry* **40**, 14098–14105
602. Filipowicz, W., and Gross, H. J. (1984) *Trends Biochem. Sci.* **9**, 68–71
603. Ogden, R. C., Knapp, G., Peebles, C. L., Johnson, J., and Abelson, J. (1981) *Trends Biochem. Sci.* **6**, 154–158
604. Skoglund, U., Andersson, K., Strandberg, B., and Daneholt, B. (1986) *Nature (London)* **319**, 560–564
605. Dreyfuss, G., Swanson, M. S., and Pinol-Roma, S. (1988) *Trends Biochem. Sci.* **13**, 86–91
606. Richter, J. D. (1988) *Trends Biochem. Sci.* **13**, 483–486
607. Müller, S., Wolpensinger, B., Angenitzki, M., Engel, A., Sperling, J., and Sperling, R. (1998) *J. Mol. Biol.* **283**, 383–394
608. Sharp, P. A. (1987) *Science* **235**, 766–771
609. Guthrie, C. (1991) *Science* **253**, 157–163
610. Newman, A. J. (1997) *EMBO J.* **16**, 5797–5800
611. Plessel, G., Lührmann, R., and Kastner, B. (1997) *J. Mol. Biol.* **265**, 87–94
612. Valcárcel, J., and Green, M. R. (1996) *Trends Biochem. Sci.* **21**, 296–301
613. Berget, S. M. (1995) *J. Biol. Chem.* **270**, 2411–2414
614. Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H., and Nagai, K. (1994) *Nature (London)* **372**, 432–438
615. Padgett, R. A., Grabkowski, P. J., Konarska, M. M., and Sharp, P. A. (1985) *Trends Biochem. Sci.* **10**, 154–157
616. Deirdre, A., Scadden, J., and Smith, C. W. J. (1995) *EMBO J.* **14**, 3236–3246
617. Blencowe, B. J. (2000) *Trends Biochem. Sci.* **25**, 106–110
618. Reuter, K., Nottrott, S., Fabrizio, P., Lührmann, R., and Ficner, R. (1999) *J. Mol. Biol.* **294**, 515–525
619. Xie, J., Beickman, K., Otte, E., and Rymond, B. C. (1998) *EMBO J.* **17**, 2938–2946
620. Furman, E., and Glitz, D. G. (1995) *J. Biol. Chem.* **270**, 15515–15522
621. Makarov, E. M., Makarova, O. V., Achsel, T., and Lührmann, R. (2000) *J. Mol. Biol.* **298**, 567–575
622. Jiang, J., Horowitz, D. S., and Xu, R.-M. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3022–3027
623. Sontheimer, E. J., Sun, S., and Piccirilli, J. A. (1997) *Nature (London)* **388**, 801–805
624. Moore, M. J., and Sharp, P. A. (1993) *Nature (London)* **365**, 364–368
625. Tarn, W.-Y., and Steitz, J. A. (1996) *Science* **273**, 1824–1832
626. Will, C. L., Schneider, C., Reed, R., and Lührmann, R. (1999) *Science* **284**, 2003–2005
627. Bailey-Serres, J., Rochaix, J.-D., Wassenegger, M., and Filipowicz, W. (1999) *EMBO J.* **18**, 5153–5158
628. Noguchi, T., Inoue, H., and Tanaka, T. (1986) *J. Biol. Chem.* **261**, 13807–13812
629. Saitta, B., Stokes, D. G., Vissing, H., Timpl, R., and Chu, M.-L. (1990) *J. Biol. Chem.* **265**, 6473–6480
630. Magnuson, V. L., Young, M., Schattenberg, D. G., Mancini, M. A., Chen, D., Steffensen, B., and Klebe, R. J. (1991) *J. Biol. Chem.* **266**, 14654–14662
631. Nawa, H., Kotani, H., and Nakanishi, S. (1984) *Nature (London)* **312**, 729–734
632. Maeda, N., Kim, N.-S., Azen, E. A., and Smithies, O. (1985) *J. Biol. Chem.* **260**, 11123–11130
633. MacLeod, A. R., Houlker, C., Reinach, F. C., Smillie, L. B., Talbot, K., Modi, G., and Walsh, F. S. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7835–7839
634. Collins, T., Bonthron, D. T., and Orkin, S. H. (1987) *Nature (London)* **328**, 621–624
635. Hsu, T.-C., Shore, S. K., Seshsmma, T., Bagasra, O., and Walsh, P. N. (1998) *J. Biol. Chem.* **273**, 13787–13793
636. Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M., and Romeo, P.-H. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6–10
637. Lamond, A. I. (1991) *Trends Biochem. Sci.* **16**, 452–453
638. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999) *Nature (London)* **398**, 579–585
639. Bruzik, J. P., Van Doren, K., Hirsch, D., and Steitz, J. A. (1988) *Nature (London)* **335**, 559–562
640. Bonen, L. (1993) *FASEB J.* **7**, 40–46
641. Hannon, G. J., Maroney, P. A., Yu, Y.-T., Hannon, G. E., and Nilsen, T. W. (1992) *Science* **258**, 1775–1780
642. Davis, R. E., Hardwick, C., Tavernier, P., Hodgson, S., and Singh, H. (1995) *J. Biol. Chem.* **270**, 21813–21819
643. Ghetti, A., and Abelson, J. N. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11461–11464
644. Herrin, D. L., and Schmidt, G. W. (1988) *J. Biol. Chem.* **263**, 14601–14604
645. Eul, J., Graessmann, M., and Graessmann, A. (1995) *EMBO J.* **14**, 3226–3235
646. Vester, B., Nielsen, A. K., Hansen, L. H., and Douthwaite, S. (1998) *J. Mol. Biol.* **282**, 255–264
647. Bachellerie, J.-P., and Cavaillé, J. (1997) *Trends Biochem. Sci.* **22**, 257–261
648. Gaspin, C., Cavaillé, J., Erauso, G., and Bachellerie, J.-P. (2000) *J. Mol. Biol.* **297**, 895–906
648a. Jády, B. E., and Kiss, T. (2001) *EMBO J.* **20**, 541–551
649. Blum, B., Bakalara, N., and Simpson, L. (1990) *Cell* **60**, 189
650. Wilson, M. A., and Pohorille, A. (1996) *J. Am. Chem. Soc.* **118**, 6580–6587
651. Riley, G. R., Corell, R. A., and Stuart, K. (1994) *J. Biol. Chem.* **269**, 6101–6108
652. Hodges, P., and Scott, J. (1992) *Trends Biochem. Sci.* **17**, 77–81
652a. Keegan, L. P., Gallo, A., and O'Connell, M. A. (2000) *Science* **290**, 1707–1709
652b. Gerber, A. P., and Keller, W. (2001) *Trends Biochem. Sci.* **26**, 376–384
653. Navaratnam, N., Fujino, T., Bayliss, J., Jarmuz, A., How, A., Richardson, N., Somasekaram, A., Bhattacharya, S., Carter, C., and Scott, J. (1998) *J. Mol. Biol.* **275**, 695–714
654. Hersberger, M., and Innerarity, T. L. (1998) *J. Biol. Chem.* **273**, 9435–9442
655. Gray, M. W. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8157–8159
656. Giegé, P., and Brennicke, A. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15324–15329
657. Hirose, T., and Sugiura, M. (1997) *EMBO J.* **16**, 6804–6811
658. Paul, M. S., and Bass, B. L. (1998) *EMBO J.* **17**, 1120–1127
659. Maas, S., Melcher, T., Herb, A., Seeburg, P. H., Keller, W., Krause, S., Higuchi, M., and O'Connell, M. A. (1996) *J. Biol. Chem.* **271**, 12221–12226
660. Bass, B. L. (1997) *Trends Biochem. Sci.* **22**, 157–162
660a. Bass, B. L. (2002) *Ann. Rev. Biochem.* **71**, 817–846
661. Lehmann, K. A., and Bass, B. L. (1999) *J. Mol. Biol.* **291**, 1–13
662. Gerber, A., Grosjean, H., Melcher, T., and Keller, W. (1998) *EMBO J.* **17**, 4780–4789
663. Schwartz, T., Rould, M. A., Lowenhaupt, K., Herbert, A., and Rich, A. (1999) *Science* **284**, 1841–1845
664. Kable, M. L., Seiwert, S. D., Heidmann, S., and Stuart, K. (1996) *Science* **273**, 1189–1195
665. Kable, M. L., Heidmann, S., and Stuart, K. D. (1997) *Trends Biochem. Sci.* **22**, 162–166
666. Rusché, L. N., Cruz-Reyes, J., Piller, K. J., and Sollner-Webb, B. (1997) *EMBO J.* **16**, 4069–4081
667. Frech, G. C., Bakalara, N., Simpson, L., and Simpson, A. (1995) *EMBO J.* **14**, 178–187
667a. Müller, U. F., Lambert, L., and Göringer, H. U. (2001) *EMBO J.* **20**, 1394–1404
667b. Grams, J., McManus, M. T., and Hajduk, S. L. (2000) *EMBO J.* **19**, 5525–5532
667c. Schnaufer, A., Panigrahi, A. K., Panicucci, B., Igo, R. P., Jr., Salavati, R., and Stuart, K. (2001) *Science* **291**, 2159–2162
668. Horváth, A., Berry, E. A., and Maslov, D. A. (2000) *Science* **287**, 1639–1640
669. Wang, S. S., Mahendran, R., and Miller, D. L. (1999) *J. Biol. Chem.* **274**, 2725–2731
669a. Cheng, Y.-W., Visomirski-Robic, L. M., and Gott, J. M. (2001) *EMBO J.* **20**, 1405–1414
670. Petschek, J. P., Mermer, M. J., Scheckelhoff, M. R., Simone, A. A., and Vaughn, J. C. (1996) *J. Mol. Biol.* **259**, 885–890
671. Jacques, J.-P., Hausmann, S., and Kolakofsky, D. (1994) *EMBO J.* **13**, 5496–5503
672. Turner, P. (1985) *Nature (London)* **316**, 105–106
673. Mowry, K. L., and Steitz, J. A. (1987) *Science* **238**, 1682–1687
674. Spirin, A. S., and Ajtkhozhin, M. A. (1985) *Trends Biochem. Sci.* **10**, 162–165
675. Tiedge, H., Bloom, F. E., and Richter, D. (1999) *Science* **283**, 186–187
676. Strauss, E. (1999) *Science* **283**, 12–13
677. Tzagoloff, A., and Myers, A. M. (1986) *Ann. Rev. Biochem.* **55**, 249–285

References

678. Mueller, D. M., and Getz, G. S. (1986) *J. Biol. Chem.* **261**, 11756 – 11764
679. Hedtke, B., Börner, T., and Weihe, A. (1997) *Science* **277**, 809 – 811
680. Shadel, G. S., and Clayton, D. A. (1993) *J. Biol. Chem.* **268**, 16083 – 16086
681. Rossmanith, W., Tullo, A., Potuschak, T., Karwan, R., and Sbisà, E. (1995) *J. Biol. Chem.* **270**, 12885 – 12891
682. Mayer, M., Schiffer, S., and Marchfelder, A. (2000) *Biochemistry* **39**, 2096 – 2105
683. Yokobori, S.-i, and Pääbo, S. (1997) *J. Mol. Biol.* **265**, 95 – 99
684. Margossian, S. P., and Butow, R. A. (1996) *Trends Biochem. Sci.* **21**, 392 – 396
685. Ohyama, K., Fukazawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H., and Ozeki, H. (1986) *Nature (London)* **322**, 572 – 574
686. Lewin, R. (1985) *Science* **230**, 55
687. Sambrook, J. (1983) in *Biochemistry* (Zubay, G., ed), Addison-Wesley, Reading, Massachusetts (Chapter 28)
688. Freifelder, D. (1987) *Molecular Biology*, 2nd ed., Jones and Bartlett, Boston, Massachusetts (Chapter 23)
689. Wildeman, A. G. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2123 – 2127
690. Fanning, E., and Knippers, R. (1992) *Ann. Rev. Biochem.* **61**, 55 – 85
691. Nevins, J. R. (1991) *Trends Biochem. Sci.* **16**, 435 – 439
692. Swaminathan, S., and Thimmapaya, B. (1996) *J. Mol. Biol.* **258**, 736 – 746
693. Berget, S. M., Moore, C., and Sharp, P. A. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3171 – 3175
694. Bos, J. L., and van der Eb, A. J. (1985) *Trends Biochem. Sci.* **10**, 310 – 313
695. Hogle, J. M., Chow, M., and Filman, D. J. (1985) *Science* **229**, 1358 – 1365
696. Saito, T., Meshi, T., Takamatsu, N., and Okada, Y. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6074 – 6077
697. Lamb, R. A., and Choppin, P. W. (1983) *Ann. Rev. Biochem.* **52**, 467 – 506
698. Scholtissek, C., and Naylor, E. (1988) *Nature (London)* **331**, 215
699. Pekosz, A., He, B., and Lamb, R. A. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8804 – 8806
700. Ishihama, A., Mizumoto, K., Kawakami, K., Kato, A., and Honda, A. (1986) *J. Biol. Chem.* **261**, 10417 – 10421
701. Coffin, J. M., Hughes, S. H., and Varmus, H. E., eds. (1998) *Retroviruses*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York

701a. Piot, P., Bartos, M., Ghys, P. D., Walker, N., and Schwartländer, B. (2001) *Nature (London)* **410**, 968 – 973

701b. Cohen, J. (2002) *Science* **296**, 2320 – 2324

702. Varmus, H. (1988) *Science* **240**, 1427 – 1434
703. Sugden, B. (1993) *Trends Biochem. Sci.* **18**, 233 – 235
704. Gallo, R. C., and Montagnier, L. (1988) *Sci. Am.* **259**(Oct), 41 – 48
705. Sanchez-Pescador, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempien, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D., and Luciw, P. A. (1985) *Science* **227**, 484 – 492
706. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C., and Wong-Staal, F. (1985) *Nature (London)* **313**, 277 – 284
707. Varmus, H. E. (1985) *Nature (London)* **314**, 583 – 584
708. Vaccaro, J. A., Singh, H. A., and Anderson, K. S. (1999) *Biochemistry* **38**, 15978 – 15985
709. Gabbara, S., Davis, W. R., Hupe, L., Hupe, D., and Peliska, J. A. (1999) *Biochemistry* **38**, 13070 – 13076
710. Litvak, S., Sarih-Cottin, L., Fournier, M., Andreola, M., and Tarrago-Litvak, L. (1994) *Trends Biochem. Sci.* **19**, 114 – 118
711. Litvak, S., and Araya, A. (1982) *Trends Biochem. Sci.* **7**, 361 – 364
712. Fossé, P., Mougel, M., Keith, G., Westhof, E., Ehresmann, B., and Ehresmann, C. (1998) *J. Mol. Biol.* **275**, 731 – 746
713. Gao, H.-Q., Sarafianos, S. G., Arnold, E., and Hughes, S. H. (1999) *J. Mol. Biol.* **294**, 1097 – 1113

713a. Sevilya, Z., Loya, S., Hughes, S. H., and Hizi, A. (2001) *J. Mol. Biol.* **311**, 957 – 971

714. Yu, H., Jetzt, A. E., Ron, Y., Preston, B. D., and Dougherty, J. P. (1998) *J. Biol. Chem.* **273**, 28384 – 28391
715. Bujacz, G., Jaskólski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1995) *J. Mol. Biol.* **253**, 333 – 346
716. Andrake, M. D., and Skalka, A. M. (1996) *J. Biol. Chem.* **271**, 19633 – 19636
717. Greenwald, J., Le, V., Butler, S. L., Bushman, F. D., and Choe, S. (1999) *Biochemistry* **38**, 8892 – 8898
718. Maignan, S., Guilloteau, J.-P., Zhou-Liu, Q., Clément-Mella, C., and Mikol, V. (1998) *J. Mol. Biol.* **282**, 359 – 368

718a. Craigie, R. (2001) *J. Biol. Chem.* **276**, 23213 – 23216

719. Gallo, R. C. (1986) *Sci. Am.* **255**(Dec), 88 – 98
720. Gallo, R. C. (1987) *Sci. Am.* **256**(Jan), 47 – 56
721. Emerman, M., and Malim, M. H. (1998) *Science* **280**, 1880 – 1884
722. Frankel, A. D., and Young, J. A. T. (1998) *Ann. Rev. Biochem.* **67**, 1 – 25
723. Mikaélian, I., Krieg, M., Gait, M. J., and Karn, J. (1996) *J. Mol. Biol.* **257**, 246 – 264
724. Wang, Z., and Rana, T. M. (1998) *Biochemistry* **37**, 4235 – 4243
725. Ivanov, D., Kwak, Y. T., Nee, E., Guo, J., García-Martínez, L. F., and Gaynor, R. B. (1999) *J. Mol. Biol.* **288**, 41 – 56

725a. Jeang, K.-T., Xiao, H., and Rich, E. A. (1999) *J. Biol. Chem.* **274**, 28837 – 28840

726. Cullen, B. R., and Malim, M. H. (1991) *Trends Biochem. Sci.* **16**, 346 – 350
727. Charpentier, B., Stutz, F., and Rosbash, M. (1997) *J. Mol. Biol.* **266**, 950 – 962
728. Barnham, K. J., Monks, S. A., Hinds, M. G., Azad, A. A., and Norton, R. S. (1997) *Biochemistry* **36**, 5970 – 5980
729. Echarri, A., González, M. E., and Carrasco, L. (1996) *J. Mol. Biol.* **262**, 640 – 651

729a. Arold, S. T., and Baur, A. S. (2001) *Trends Biochem. Sci.* **26**, 356 – 363

729b. Pomerantz, R. J. (2002) *Nature (London)* **418**, 594 – 595

729c. Segura-Totten, M., and Wilson, K. L. (2001) *Science* **294**, 1016 – 1017

730. Greene, W. C. (1991) *N. Engl. J. Med.* **324**, 308 – 317
731. Guyadar, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L., and Alizon, M. (1987) *Nature (London)* **326**, 662 – 669
732. Overbaugh, J., Donahue, P. R., Quackenbush, S. L., Hoover, E. A., and Mullins, J. I. (1988) *Science* **239**, 906 – 910
733. Nerenberg, M., Hinrides, S. H., Reynolds, R. K., Khoury, G., and Jay, G. (1987) *Science* **237**, 1324 – 1329
734. Haase, A. T. (1986) *Nature (London)* **322**, 130 – 136
735. Turner, B. G., and Summers, M. F. (1999) *J. Mol. Biol.* **285**, 1 – 32
736. Miedema, F., and Klein, M. R. (1996) *Science* **272**, 505 – 506

736a. Ezzell, C. (2002) *Sci. Am.* **286**(Jun), 40 – 45

737. Roberts, J. D., Bebenek, K., and Kunkel, T. A. (1988) *Science* **242**, 1171 – 1173
738. Hill, C. M., and Littman, D. R. (1996) *Nature (London)* **382**, 668 – 669
739. Cohen, J. (1995) *Science* **270**, 917 – 918
740. Kidwell, M. G., and Lisch, D. R. (1998) *Nature (London)* **393**, 22 – 23
741. Kazazian, H. H., Jr. (2000) *Science* **289**, 1152 – 1153
742. Moffat, A. S. (2000) *Science* **289**, 1455, 1457
743. Ovchinnikov, I., Rubin, A., and Swergold, G. D. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10522 – 10527
744. Perl, A., Colombo, E., Samoilova, E., Butler, M. C., and Banki, K. (2000) *J. Biol. Chem.* **275**, 7261 – 7272
745. Southworth, M. W., Adam, E., Panne, D., Byer, R., Kautz, R., and Perler, F. B. (1998) *EMBO J.* **17**, 918 – 926
746. Strauss, E. (2000) *Science* **289**, 374

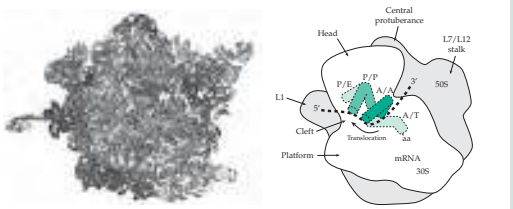
Study Questions

(Additional questions are located on p. 1738.)

1. Describe the sequence of events involved in the initiation of transcription in *E. coli.* As part of your answer, describe those features that must be present in a gene for proper recognition and transcription by the RNA polymerase.

2. How does transcription in eukaryotes differ from that in prokaryotes?

3. In *E. coli* precise spacing between the conserved -35 and -10 (Pribnow) promoter elements has been found to be a critical determinant of promoter strength. What does this suggest about the interaction between RNA polymerase and these conserved sequences in the DNA?

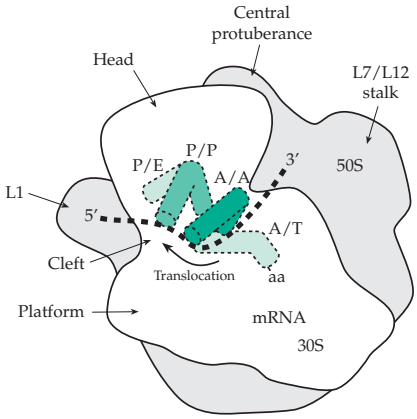


The ~15,000 ribosomes in a bacterial cell synthesize over 4000 proteins following the genetic code in messenger RNA molecules. A ribosome (left) consists of two large subunits, each composed largely of ribosomal RNA, whose folded chains can be seen. About 100 proteins are bound, largely to solvent-exposed surfaces, but with extended "tails" protruding into the ribosome. The messenger RNA (mRNA; right) moves through the ribosome between the large subunits. Amino acids, activated for reaction, are carried into the ribosomes by transfer RNAs (green) which move consecutively from A/T to A/A, A/P, P/P, and P/E sites. They insert their activated amino acids into the growing polypeptide chain in the P site of the 50S subunit. Image of ribosome[33a] courtesy of the authors.

Contents

Ribosomes and the Synthesis of Proteins

29



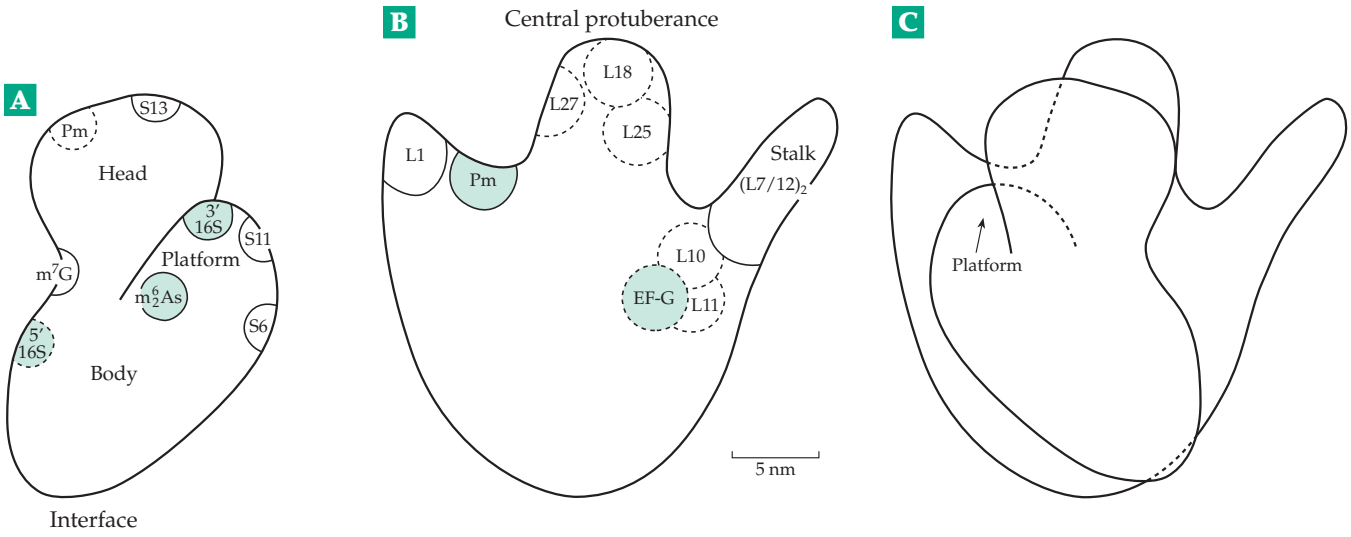
The many thousands of proteins present in each cell are made within the ribosomes, which are able to read each specific mRNA that comes their way. While ribosomes appear as little more than blurred dots in most electron micrographs, the 15,000 ribosomes of one cell of *E. coli* represent one-fourth of the total mass of the cell. Eukaryotic cells contain many times more of these little molecular machines. When ribosomes were first observed in the early 1950s,[1–3] nobody could imagine either their composition or their function. Less than 50 years later (1999) their complete three-dimensional structure was known at nearly atomic resolution, and the function of ribosomes in protein synthesis was quite well understood. However, the structure could not have been obtained without the development of a whole range of new methods.

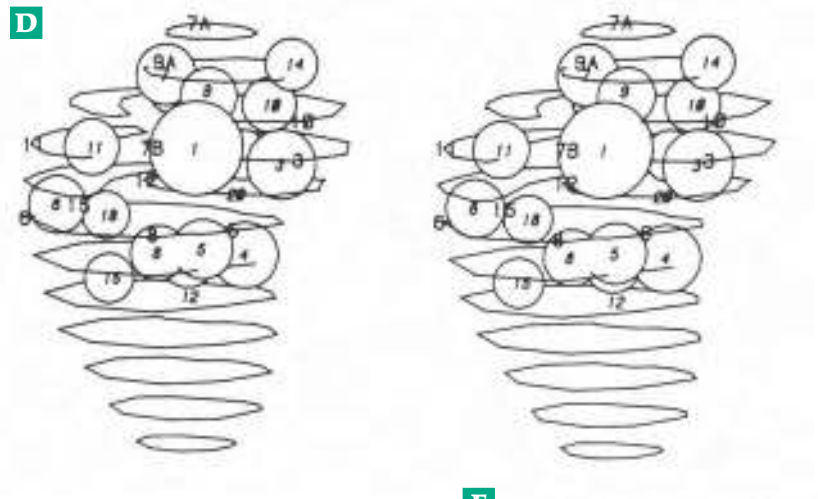
As electron microscopy developed, the fuzzy granules in micrographs assumed well-recognized forms. By the 1970s, the approximate shapes of the two ribosomal subunits were known, locations of several ribosomal proteins had been established, and binding sites of transfer RNAs and other features of ribosomes were being mapped.[2,4–7] The resulting picture of the ribosome structure, which is shown schematically in Fig. 29-1, is quite similar to the present-day view. Later, three-dimensional images were reconstructed from electron micrographs (electron tomography),[8,9] and cryo-electron microscopy provided detailed images at a resolution of ~2 nm.[10]

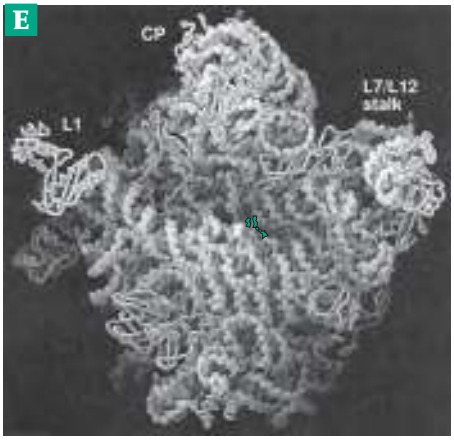
In 1950, when the study of ribosomes began, no methods for determining the sequences of amino acids in proteins or of nucleotides in nucleic acids existed.[11] Sanger published the sequences of the two short chains of insulin in 1953, and the first transfer RNA sequence was published by Holley in 1965.[21] Nevertheless, by 1980 the Wittmanns and coworkers in Berlin had sequenced 53 of the *E. coli* ribosomal proteins[4,22] (Table 29-2), and the three rRNA molecules had also been sequenced.[22a,b,c] In 1950, X-ray crystallography of proteins was still in its infancy; the structure of myoglobin was not determined until 1960. Ribosomal proteins proved hard to crystallize, the first structure being solved in 1980.[23] NMR structural analysis yielded several structures including that of L30 (Fig. 3-25A). In recent years, high-resolution structures of many additional ribosomal proteins in their free forms have been established as have the structures of most of the proteins bound into ribosomes.[24]

The first crystals of bacterial ribosomes in a three-dimensional lattice suitable for study by X-ray diffraction at a resolution of ~1 nm[27]were obtained by Yonath in about 1980.[6,25,26] Now atomic structures are being established at a resolution of 0.3–0.1 nm, or less.[17–19,28–33g] However, such progress would have been impossible without information about ribosomes obtained from improved cryo-electron microscopy,[10,20,33e,34–37a] phylogenetic analysis of ribosomal RNAs,[38–39b] mutational analysis,[40,41] neutron scattering,[42–47] chemical and photochemical crosslinking,[48–53] photoaffinity labeling,[54] immunological labeling,[55] chemical footprinting,[56,57] fluorescence resonance energy transfer (FRET),[58] mass spectrometry,[59,60] and study of the effects of toxic proteins (Box 29-A)[61] and antibiotics (Box 29-B).

Why is the ribosome so large and complex? Aside from the fact that it must form the peptide linkages, it must translate the genetic code in the mRNA into the correct amino acid sequence for each of the thousands of proteins present in the cell. The process takes place







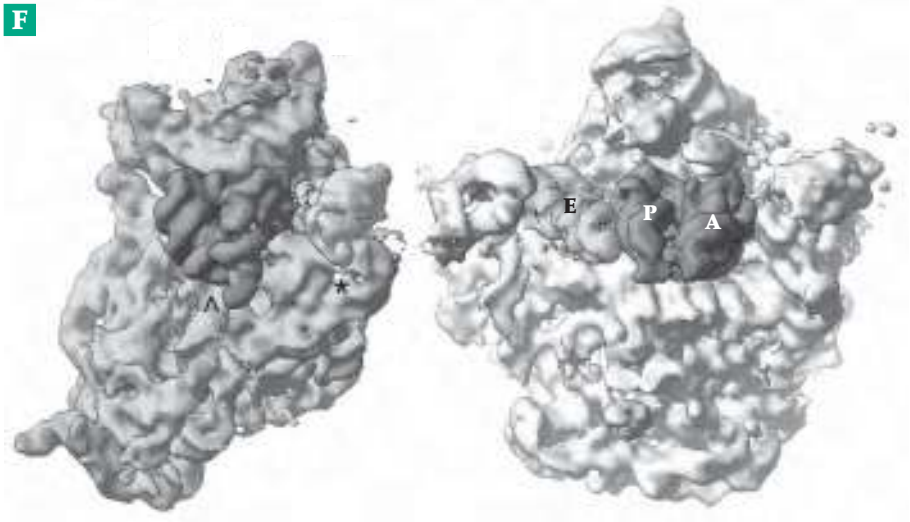
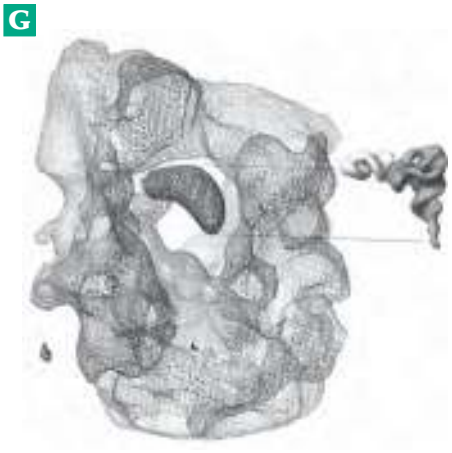
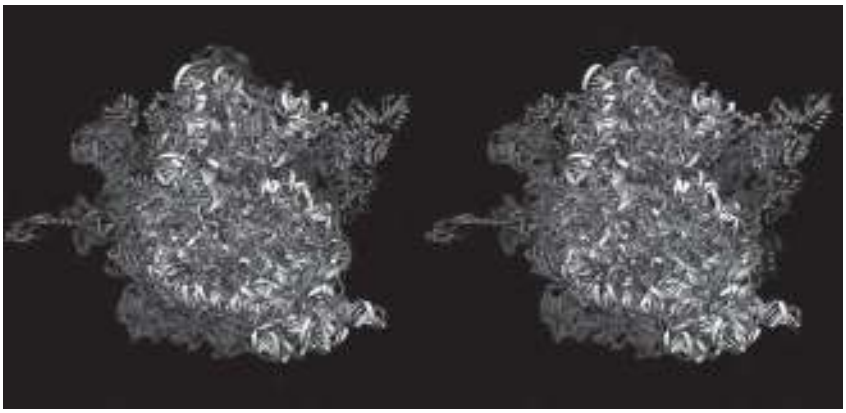
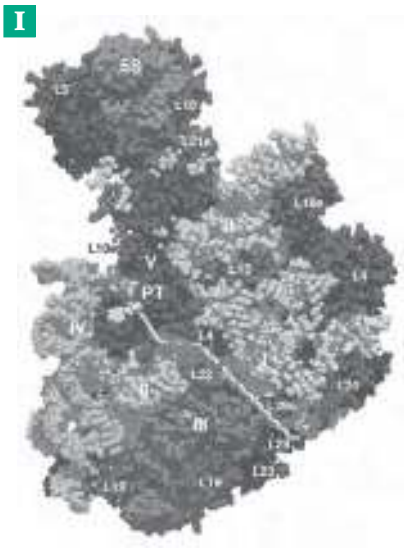
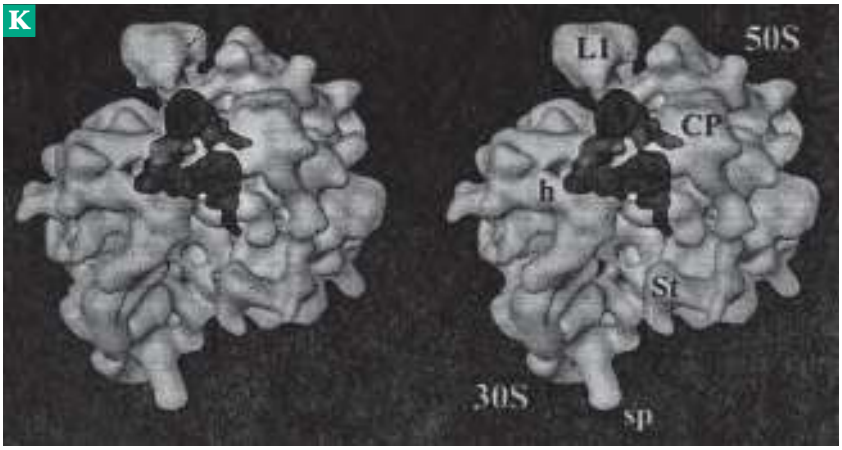


Figure 29-1 (A–C) A 1970s view of a bacterial ribosome achieved by electron microscopy and image reconstruction. These interface views show the surfaces that face each other in the 70S ribosome. Locations marked with dashed lines are on the outer (back) surfaces. From Nagano and Harel.[12] Based on shapes and data of Lake and associates.[13,14] (A) The 30S subunit. Positions of a few proteins located by immunoelectron microscopy and three positions in the 16S RNA are marked. The puromycin binding site labeled Pm was mistakenly thought to be near the peptidyltransferase center. (B) The 50S subunit. Only a few positions of many located by a variety of techniques are marked. EF-G marks the site at which elongation factor G (see Fig. 29-12) binds. A prominent feature is the "stalk", designated St in some of the drawings. It was early shown to be formed by two copies each of the nearly identical L7 and L12. The stalk is flexible and in many circumstances may be folded down or not visible as in the X-ray structures of F–H. (C) The 70S ribosome. The 50S subunit is oriented as in (A) while the 30S subunit has its outer face toward the viewer. (D) Stereoscopic view of a neutron scattering map of the 30S subunit of an *E. coli* ribosome. The proteins studied are represented as spheres with volumes corresponding to those of the anhydrous proteins. Also marked on the map are positions of proteins located by immunoelectron microscopy as mapped by Kahan *et al.*[15] Figure from Ramakrishnan *et al.*[16] Courtesy of V. Ramakrishnan. (E) Model of the 50S subunit from *Haloarcula marismortui.* From









Ban *et al*.[17] Courtesy of T. A. Steitz. The peptidyltransferase center is marked by the green image of the transition state inhibitor shown in Fig. 29-13. (F) Model of three tRNAs bound to a ribosome from *Thermus thermophilus* in the A (aminoacyl), P (peptidyl), and E (exit) sites. These are based on 0.75-nm X-ray data and a number of difference electron density maps. The 3'-CCA end of the A-site tRNA is not modeled but is marked "^". Views are left, facing the inner surface of the 30S subunit; right, facing the inner surface of the 50S subunit. (G) Schematic side view of a ribosome showing a molecule of tRNA bound in the A site between the 30S and 50S subunits. The anticodon of the tRNA is base-paired with mRNA in the "decoding site" on the 30S subunit. The 3'-CCA end with attached aminoacyl group lies in the peptidyltransferase site in the 50S subunit. (F) and (G) are courtesy of Cate *et al*.[18] (H) Stereoscopic view of a model of the 70S ribosome from *T. thermophilus*. The 30S subunit (lighter) is toward the viewer. Courtesy of Harry F. Noller and Albion Baucom. (I) Section through the 0.24 nm-resolution model of the 50S subunit shown in (E). The modeled path of the polypeptide chain through the exit tunnel is marked. Courtesy of Nissen *et al*.[19] (J) Stereo diagram of the relative orientations of the A-, P-, and E-tRNAs and mRNA showing codon–anticodon interactions and the kink between the A and P codons. (H) and (J) courtesy of Yusupov *et al*.[33a] (K) Stereoscopic view of tRNAs in the P site and in an overlapping P/E site as observed by cryo-electron microscopy at a resolution of 0.5 nm in an image of the 70S ribosome at 1.5 nm resolution. The anticodon arms are to the left. Two tRNA molecules are not present simultaneously but their images have been presented together. From Agrawal *et al*.[20] Courtesy of Rajendra Agrawal.

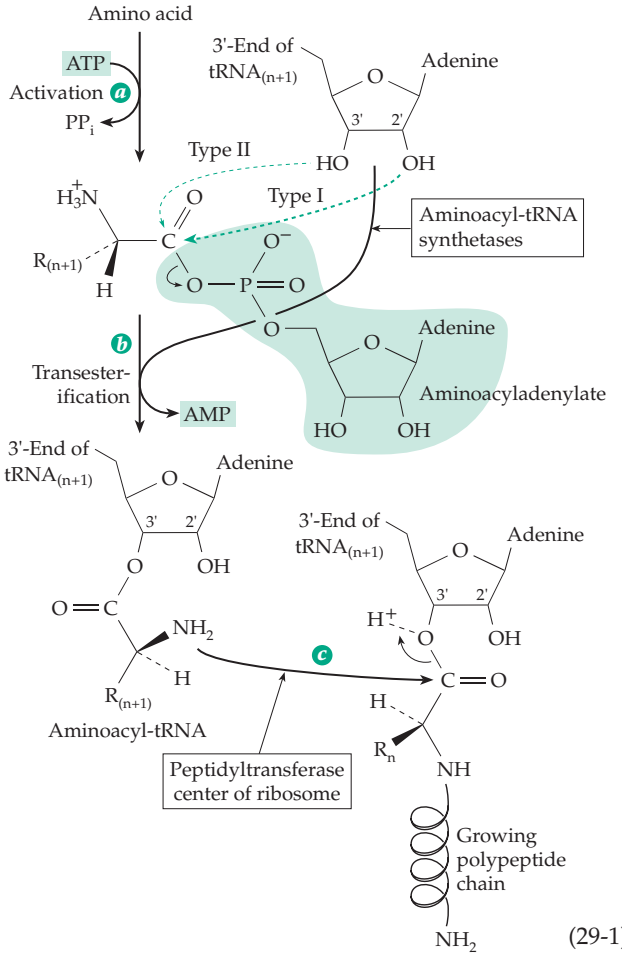
TABLE 29-1
The Composition of Ribosomes

Prokaryotic[a]		Eukaryotic[b]	
Component	Mass, kDa	Component	Mass, kDa
Small (30S) subunit	850	Small (40S) subunit	1440
16S RNA	500	18S RNA	700
Proteins (21)	350 (total)	Proteins (~30)	740
Large (50S) subunit	1450	Large (60S) subunit	2800
23S RNA	950	28S RNA	1700
5S RNA	40	5.8S RNA	51
		5S RNA	39
Proteins (32–34)	460	Proteins (~46)	1010
Complete (70S) ribosome	2300	Complete (80S) ribosome	4240

[a] Data from Wittmann, H. G. (1982) *Ann. Rev. Biochem.* **51**, 155–183. Based on sequences of all components. Presence of spermine, K^+, etc., may add 10%.
[b] Data from Freifelder, D. (1983) *Molecular Biology*, 2nd ed., Jones and Bartlett, Boston, Massachusetts (p. 419) and Mao, H., and Williamson, J. R. (1999) *J. Mol. Biol.* **292**, 345–349.

in several steps: (1) **Initiation** of protein synthesis in bacteria requires binding of the 30S ribosomal subunit to an mRNA molecule, location of the start signal (initiation codon and nearby Shine–Dalgarno sequence, shown in Fig. 29-2 and Eq. 29-8), and binding of the initiator tRNA carrying formylmethionine. The 30S complex must then bind to the 50S ribosomal subunit. (2) **Elongation** of the polypeptide chain in the resulting complete 70S ribosome ensues with binding of the appropriate aminoacyl tRNA to the next codon in the 5' → 3' direction. Base pairs form between the anticodon of the tRNA and the mRNA codon that lies in the **aminoacyl (A) site**; the peptide bond is then formed by the **peptidyltransferase** reaction. This reaction is followed by **translocation**, movement of the initiator tRNA into an exit site at the same time that the second tRNA (together with its mRNA codon and the attached growing peptide chain) moves into the **peptidyl (P) site**. The elongation cycle is repeated until the peptide chain is complete. (3) **Termination** of translation involves release of the completed protein and preparation of the ribosomal subunits for another cycle. The entire process is powered by the hydrolysis of ATP and GTP. The ATP is utilized in a three-step process for the activation of the amino acids, which become linked to the appropriate tRNAs (Eqs. 17-36 and 29-1).[61a] If the inorganic pyrophosphate that is formed is hydrolyzed, two molecules of ATP are required for activation of each amino acid molecule. In addition, at least two molecules of GTP are hydrolyzed to GDP and inorganic phosphate within the ribosome for each peptide linkage formed.

The pairing of codons and anticodons required for insertion of the correct amino acid into the growing polypeptide chain is often referred to as **decoding** of the gene sequence. However, an equally important part of the decoding is the attachment of the correct amino acid to its corresponding **cognate tRNA**. This occurs in the cytoplasm and also in the nucleus.[62] The base pairing of tRNAs and mRNA, which follows, occurs in the **decoding center** on the 30S ribosomal subunit. Both the A and P sites of tRNA-binding and the decoding center are formed by folds of the 16S RNA. The peptide bond formation takes place at the opposite ends of the tRNA molecules in the **peptidyltransferase center** of the 50S subunit (see Figs. 29-1, 29-14). As pointed out in Chapter 12 (top of p. 650), peptidyltransferase is a **ribozyme**. Its active site consists entirely of segments of the 23S RNA (see Fig. 29-14). Another important site is the **GTPase activating center**, at which the 23S



(29-1)

RNA interacts with specific G proteins known as **initiation, elongation,** and **termination factors**.

The proteins in a ribosome may help to hold the RNA into conformations that are correct for its functions. They may also catalyze conformational alterations during the various steps of the translation process. In addition, the proteins may help provide binding sites for substrate molecules and participate in regulatory activities. Both the tRNA **exit (E) site** and the **tunnel** through which the polypeptide chain leaves the ribosome are composed, in part, of ribosomal proteins.

A. The Architecture of Ribosomes

Ribosomes of *E. coli* each have a mass of ~ 2.3 x 10^6 daltons and are ~ 65% RNA and 35% protein. Ribosomes of eukaryotic organisms are larger (~4.3 x 10^6 daltons) and consist of ~50% RNA and 50% protein. Under some conditions such as a low Mg^{2+} concentration complete bacterial ribosomes, called **70S ribosomes**, dissociate into two subunits of unequal size, which are known as **30S** and **50S ribosomal subunits**. The larger 50S subunit is about twice the size of the smaller one (Table 29-1). The small 30S ribosomal subunit contains the 16S rRNA, a chain of ~1500–1700 nucleotides (nt) that, if fully extended, would stretch to a length of over 500 nm. In addition to the highly folded RNA molecule, the 30S subunit contains 21 proteins, each one unique in its amino acid composition and sequence (Table 29-2). Many of these proteins, which are designated S1, S2, S3, etc., are of relatively low molecular mass. Many are strongly basic. They contain numerous lysine and arginine residues, many of which are able to interact with RNA in the ribosome. However, neutral and acidic proteins are also present. The 50S ribosomal subunit contains the ~ 2900 nt 23S rRNA, the ~120 nt 5S RNA, and about 31–34 proteins, two of which (L7 and L12) are present as two copies each. The composition of ribosomes is variable, but most proteins are present in a strict 1:1 ratio. Others may be lacking in some of the ribosomes. Some proteins bind to the ribosomes transiently during their function in protein synthesis as do certain proteins with functions other than protein synthesis. In both subunits the RNA molecules form the internal core. Proteins are largely found on the solvent-exposed surfaces. Some of them form the stalk and other features.[17] They often have globular domains with extended tails that interact with the ribosomal RNA.

Eukaryotic ribosomes are not only larger but also (Table 29-1) contain more protein subunits than do those of bacteria: ~30 for the small subunit and 49 for the large subunit.[63] However, the number of essential proteins may be the same. Both eukaryotic ribosomal proteins and rRNA molecules are larger than those of bacteria. Bacterial ribosomes are ~22 nm in diameter and ~30 nm in the third dimension. Eukaryotic ribosomes are of the order of 1.17 times larger in linear dimensions. Ribosomes of chloroplasts resemble those of eubacteria such as *E. coli* but contain a few more proteins.[64] Mammalian mitochondrial ribosomes also resemble those of bacteria in many respects.[65] However, their RNA chains are shorter and they contain more proteins.[66,66a] The protein content is ~66% compared with ~35% for *E. coli* ribosomes.

1. Ribosomal RNA

The sequences of all three pieces of RNA in the *E. coli* ribosomes are known as are those from many other species. These include eukaryotic mitochondrial, plastid, and cytosolic rRNA. From the sequences alone, it was clear that these long molecules could fold into a complex series of hairpin loops resembling those in tRNA. For example, the 16S rRNA of *E. coli* can fold as in Fig. 29-2A and eukaryotic 18S RNA in a similar way (Fig. 29-4).[38,39,67–69] The actual secondary structures of 16S and 18S RNAs, within the folded molecules revealed by X-ray crystallography, are very similar to that shown in Fig. 29-2A. Ribosomal RNAs undergo many posttranscriptional alterations. Methylation of 2′-hydroxyls and of the nucleic acid bases as well as conversion to pseudouridines (pp. 1638–1641) predominate over 200 modifications, principally in functionally important locations that have been found in human rRNA.[69a]

Chemical modification and crosslinking. Before high-resolution X-ray data were available, two major biochemical approaches were used to deduce the secondary structures of ribosomal RNAs.[38,39] The first was the application of chemical reagents and enzymes that modify the RNA. Crosslinking reagents were used to establish pairs of nucleotides that lie close together in the three-dimensional structure. Cleavage by specific endonucleases was used to establish whether a region of the molecule is double-helical or single-stranded.[68] Nucleases were also used to clip out base-paired fragments, which were separated, denatured, and sequenced. This revealed both hairpin loops and pairings between regions that are far apart in the primary sequence. The ability of nucleic acid bases to undergo specific chemical reactions at positions not involved in base pairing was used to establish whether or not a given base was actually paired.[67,69] Thus, every position in *E. coli* 16S RNA was probed by reactions of dimethylsulfate with adenine at N1 and cytosine at N3, reaction of kethoxal (Eq. 5-16) with guanine at N1 and N2', and by reaction of a carbodiimide with uracil at N3 and with guanine at N1.[67]

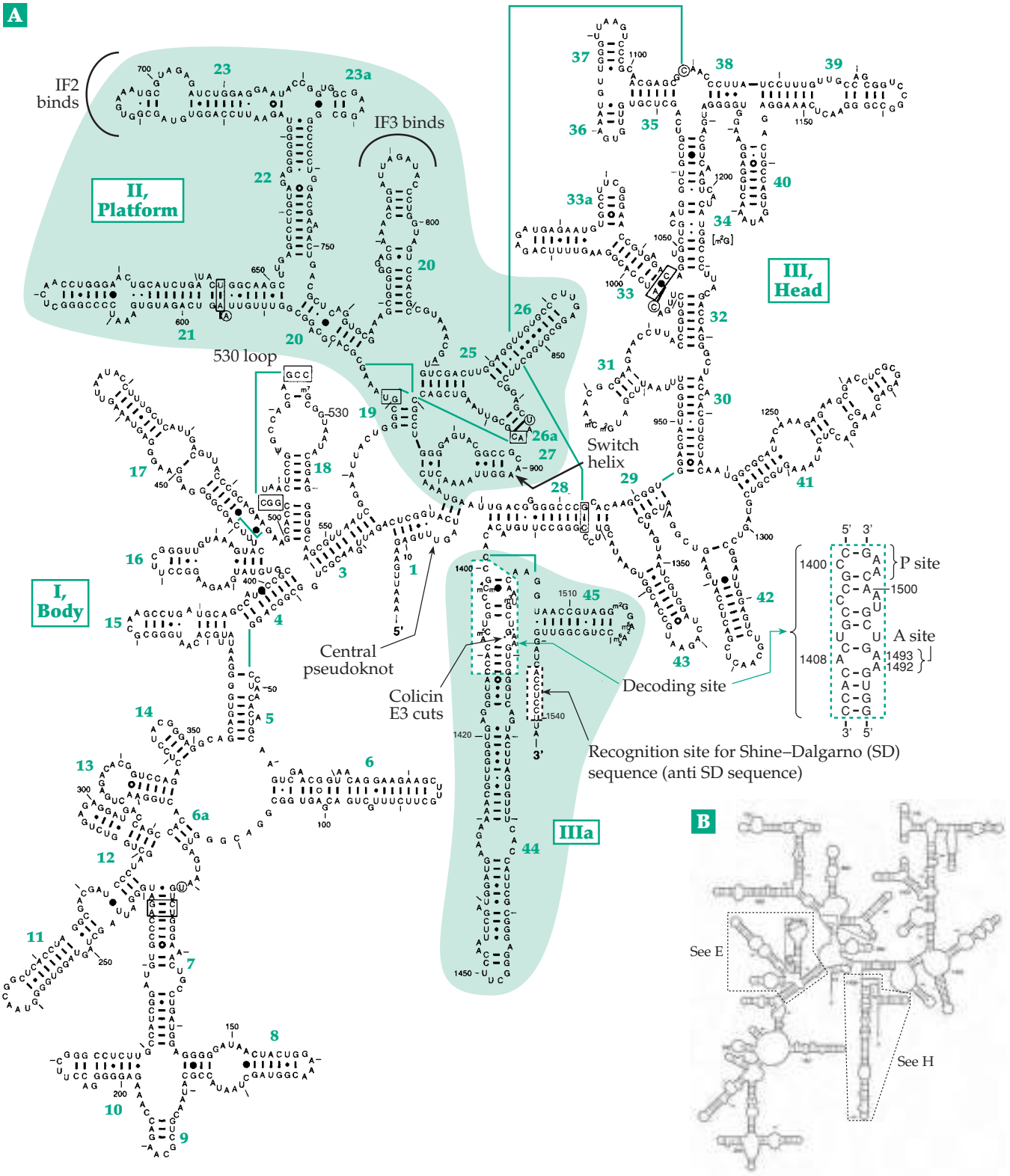
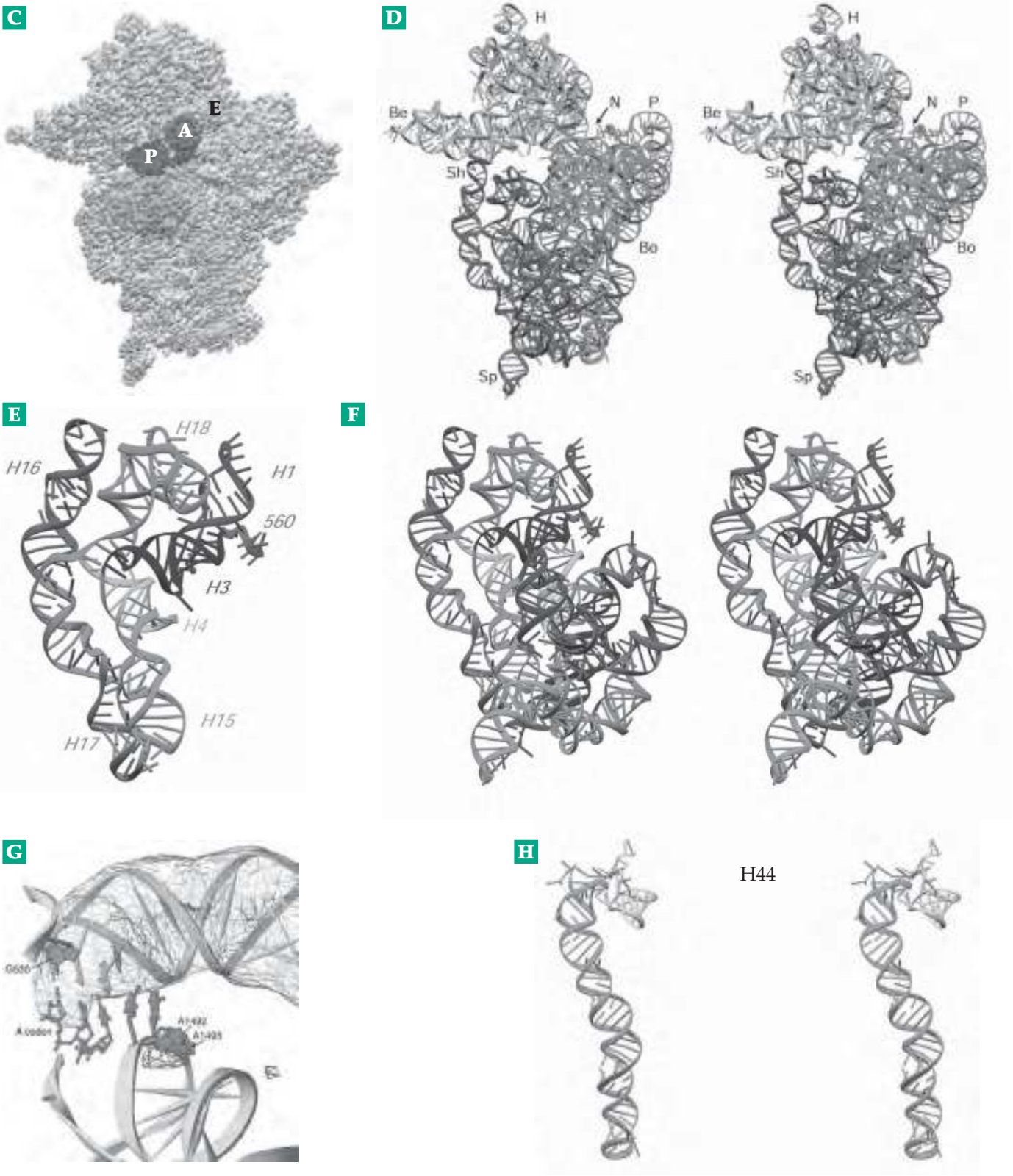


Figure 29-2 (A) Secondary structure model for the 1542-residue *E. coli* 16S rRNA based on comparative sequence analysis.[73a] Dots indicate G•U or A•G pairs; dashes indicate G•C or A•U pairs. Strongly implied tertiary interactions are shown by solid green lines. Helix numbering according to Brimacombe. Courtesy of Robin Gutell. (B) Simplified schematic drawing of type often used. (C) Positions of the A, P, and E sites on the 30S ribosomal subunit from Carter *et al.*[70] (D) Stereoscopic view of the three-dimensional fold of the 16S RNA from *Thermus thermophilus* as revealed by X-ray structural analysis at 0.3 nm resolution. Features labeled are the head (H), beak (Be), neck (N), platform (P), shoulder (Sh), spur (Sp), and body (Bo). (E-H) Selected parts of the 16S RNA. In (E) and (F) the helices are numbered as in (A). (F) and (H) are stereoscopic views. The decoding site



is located at the upper end of helix 44. (G), (H). In (G) the electron density difference observed upon binding of tRNA into the A site is displayed as a Fourier difference map (at 0.7-nm resolution). The molecular model of the tRNA with its anticodon paired with a codon from mRNA is superimposed. Two positions of bases A1492 and A1493 are shown as they are found in the presence and absence of paromomycin. A patch of negative density can be seen near the A1492 and A1493 labels, indicating that these groups may rearrange to interact with the minor groove of the codon–anticodon helix when the A-tRNA is bound. See also p. 1690. Courtesy of Yusupov *et al.*[33a] (D) through (F) and (H) are from Wimberly *et al.*[33] Courtesy of Venki Ramakrishnan.

Phylogenetic comparison. This technique, also called comparative sequence analysis, has proved very powerful.[38,39,39a,b,71,72] An example is illustrated in Fig. 29-3. Here a loop from 23S RNA of *E. coli* is shown and is compared with the sequence of 26S RNA from the fungus *Physarum polycephalum*.[38] Wherever the latter differs from the *E. coli* sequence, the substituted base is indicated in a box. These square boxes, which are concentrated in base-paired regions, indicate compensatory changes for which there is usually a second change that preserves base pairing in a double helical region. The studies also showed clearly that bacterial 16S RNA is homologous with eukaryotic 18S RNA,[68,69] with 17S RNA of dinoflagellates,[73] and also with 12S RNA of human mitochondria. Likewise, 23S RNA of bacteria corresponds to 28S RNA of eukaryotes.

Structural domains in 16S ribosomal RNA. Three major compact structural domains, 5', central, and 3', can be distinguished in 16S RNA.[33] An extended subdomain is also present at the 3' end. These are indicated on Fig. 29-2 as I, II, III, and IIIa. The double helical segments are also numbered. Ribosomal RNA molecules must be folded into compact forms to fit into the envelope of the ribosomal subunits. The individual structural domains form independent globular cores to which several proteins apiece are bound. Domains I, II, and III form the body, platform, and head, respectively (Fig. 29-1A). Relatively minor changes in conformation accompany the incorporation of the rRNA molecules into the ribosomes.[67] On this basis, and taking account of all available data, attempts were made for many years to predict a three-dimensional structure.[12,74–76] One of these[77] is portrayed in Fig. 5-32A as a series of cylinders representing the 45 double-helical segments suggested by the structure of Fig. 29-2A. This can be compared with the X-ray based model shown in Fig. 29-2B.

23S rRNA. The large RNA of the 50S subunit consists of six structural domains.[5] Its secondary structure is shown in Fig. 29-4. As with 16S RNA each domain is tightly folded. However, the domains are interdigitated in such a way that they form a single monolithic structural unit.[17] Nevertheless, there are distinct catalytic sites, as described in Section 4. Like proteins, which are able to undergo conformational alterations that usually involve some rearrangement in their internal hydrogen-bonding patterns, these large RNA molecules may also assume alternative conformations. Conformational changes may involve not only alternative hydrogen bonding patterns but also alternative base-pairing.[77a] Such changes may be essential to the functioning of ribosomes[86] and may also accompany maturation of pre-rRNAs.[87] Eukaryotic 28S RNAs have basically the same structures as the

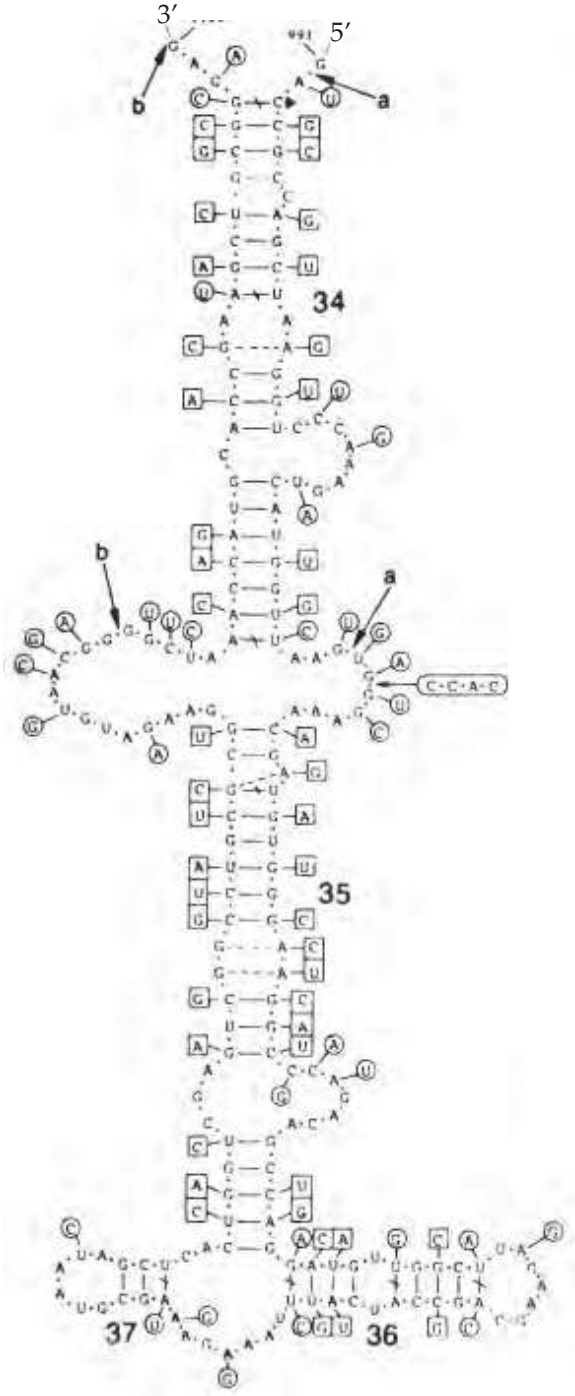


Figure 29-3 Example of phylogenetic comparisons in ribosomal RNA. The diagram shows helices 34–37 of *E. coli* 23S RNA, compared with the corresponding region of *Physarum polycephalum* 26S RNA. The diagram depicts the *E. coli* helices, with changes in the *P. polycephalum* sequences denoted by symbols in boxes on the side. Bases in square boxes are compensating; those in round boxes are mismatching or in single-stranded regions. Solid triangles denote deletions, while bases with arrows indicate insertions. Dotted lines or "crossed-out" base pairs denote modified base-pairing in *P. polycephalum*. The letters "a" and "b" indicate the termini of RNA fragments isolated as a base-paired complex. From Brimacombe.[38]

23S RNAs of bacteria but have been expanded by insertion of additional nucleotides at many places.[10,36,79]

Most of the chemical activity of ribosomes occurs in the interface between the 30S and 50S subunits. Entrance and exit tunnels for both mRNA and the aminoacylated tRNAs are formed between these subunits. The mRNA apparently moves across the platform as the tRNAs move from A to P to E sites experiencing codon selection (decoding) and peptidyltransferase activity. Many loop ends from 16S RNA interact with those of 23S RNA.[41,88]

5S rRNA. This ~120-nucleotide molecule organizes one domain of the 50S ribosomal subunit.[89] Extensive phylogenetic comparisons of 5S RNA sequences led to the secondary structure shown in Fig. 29-5.[90,91] The three-dimensional structure, as seen in a ribosome, is also shown in this figure. Study of base-pairing possibilities suggests that 5S RNA can exist in more than one conformation.[90,92] In a possible second conformation the sequence GUGUGGGG (residues 79–86) pairs in an antiparallel fashion with the sequence CCCCAUGC (residues 35–42), with loss of base pairing in stem 4 (Fig. 29-5). A structure somewhat similar to that of 5S RNA is probable for eukaryotic 5.8S RNA.[94] Nearly a thousand different prokaryotic and eukaryotic 5S RNA sequences have been compared.[95] From them **phylogenetic trees**, which suggest evolutionary pathways between species, have been constructed.[96] Sequences of 16S RNA have been used in a similar way (Fig. 1-5).[97]

2. Ribosomal Proteins

Ribosomal proteins are soluble in concentrated salt solutions. Most of them can be dissolved without damage by buffers containing 2 M LiCl and can then be separated by electrophoresis or ion exchange chromatography and gel filtration (molecular sieving).[98] Although many of them are quite insoluble and are often unstable, all ribosomal proteins of *E. coli* (Table 29-2) have been separated and sequenced, mainly by Wittmann-Liebold and coworkers.[22] The ribosomal proteins of other bacteria usually resemble those of *E. coli*.[99] The more numerous eukaryotic ribosomal proteins have also been isolated and studied individually.[100] Many of these 84 proteins appear to correspond directly in properties and functions to those of *E. coli*.[100a] As with ribosomal RNAs, the sizes of the eukaryotic proteins have been expanded.[101] Mitochondria have their own set of ribosomal proteins, which are more numerous than those of either *E. coli* or yeast.[65–66a,102–102c] Pure individual ribosomal proteins are now produced from the cloned genes as are 16S and 23S ribosomal RNAs.

Most ribosomal proteins are folded into compact forms, much of whose surfaces are accessible to added reagents. However, X-ray structures have revealed that parts of some proteins penetrate deeply into the RNA core.[17] Much of the RNA is also accessible from the outside, and the ribosome contains ~50% of its mass as internal hydration. A ribosome usually contains only one molecule of each kind of protein with the exception of proteins L7 and L12 of the large subunit. There are two of each. Sequencing of the 120-residue proteins from *E. coli* shows that L7 is

TABLE 29-2
Ribosomal Proteins from *E. coli*[a]

Proteins of 30 S Ribosomal subunits			Proteins of 30 S Ribosomal subunits		
Designation	Mass, kDa	Binding[b]	Designation	Mass, kDa	Binding[b]
S1	61.2		L1	24.6	
S2	26.6		L2	29.4	+
S3	25.8		L3	22.3	
S4	23.1	+	L4	22.1	
S5	17.5		L5	20.2	
S6	15.7		L6	18.8	+
S7					
(strain K)	19.7	+	L7	12.2	
(strain B)	17.1	+			
S8	14.0	+	L8		
S9	14.6		L9	15.5	
S10	11.7		L10	17.7	
S11	13.7		L11	14.9	
S12	13.6		L12	12.2	
S13	13.0		L13	16.0	
S14	11.1		L14	13.5	
S15	10.0	+	L15	15.0	
S16	9.2		L16	15.3	+
S17	9.6	+	L17	14.4	+
S18	8.9		L18	12.8	+
S19	10.3		L19	13.0	+
S20	9.6	+	L20	13.4	+
S21	8.4		L21	11.6	
			L22	12.2	
Total mass 350 (strain K)			L23	11.0	+
			L24	11.2	+
			L25	10.7	+
			L26 = S20	9.6	
			L27	9.0	
			L28	8.9	
			L29	7.3	
			L30	6.4	
			L31	7.0	
			L32	6.3	
			L33	6.3	
			L34	5.4	
			Total mass 460[c]		

[a] Molecular masses from Wittmann, H. G. (1982) *Ann. Rev. Biochem.* **51**, 155–183
[b] A plus sign indicates direct binding to ribosomal RNA.
[c] Four copies of L7/L12 are assumed.

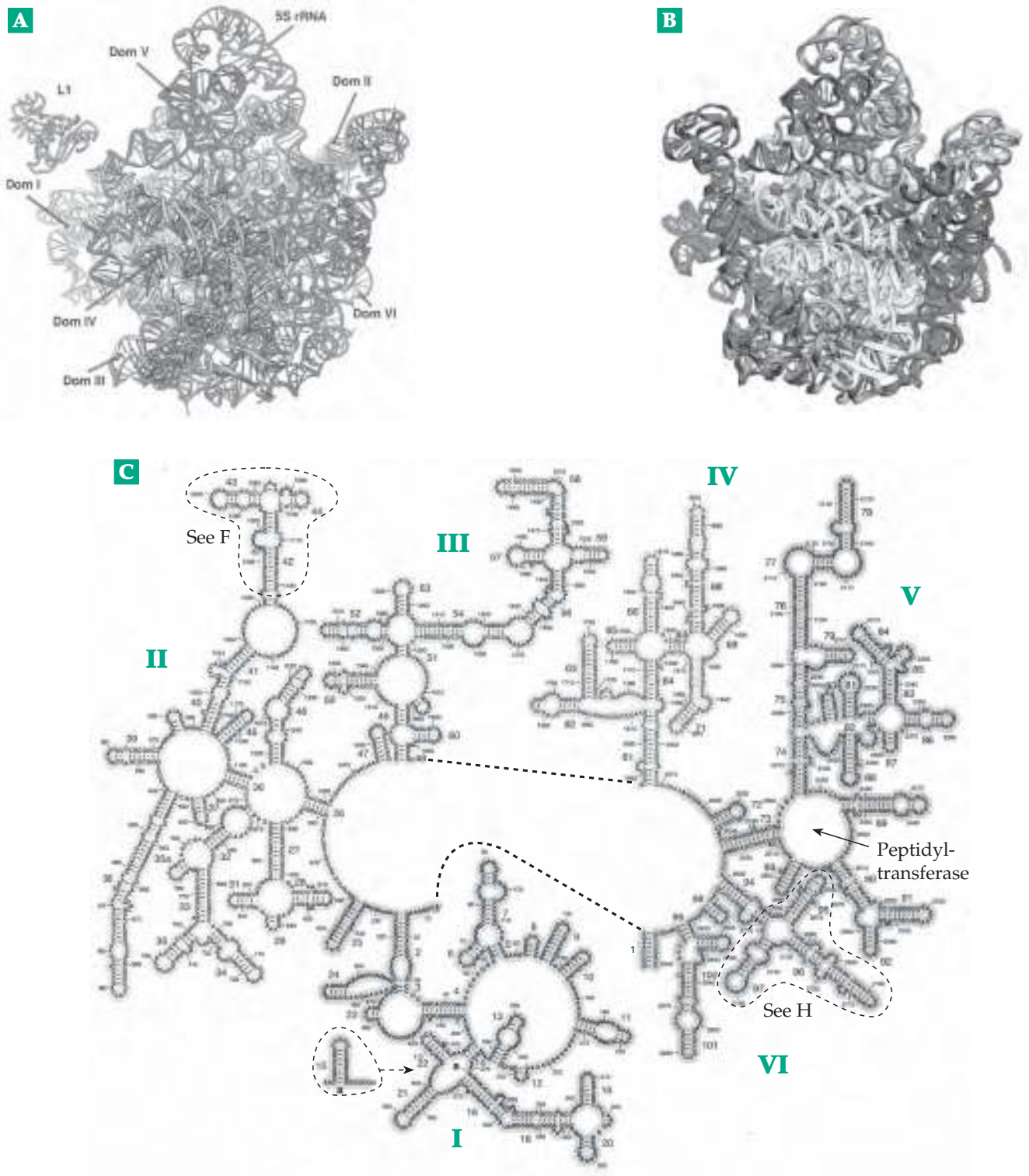
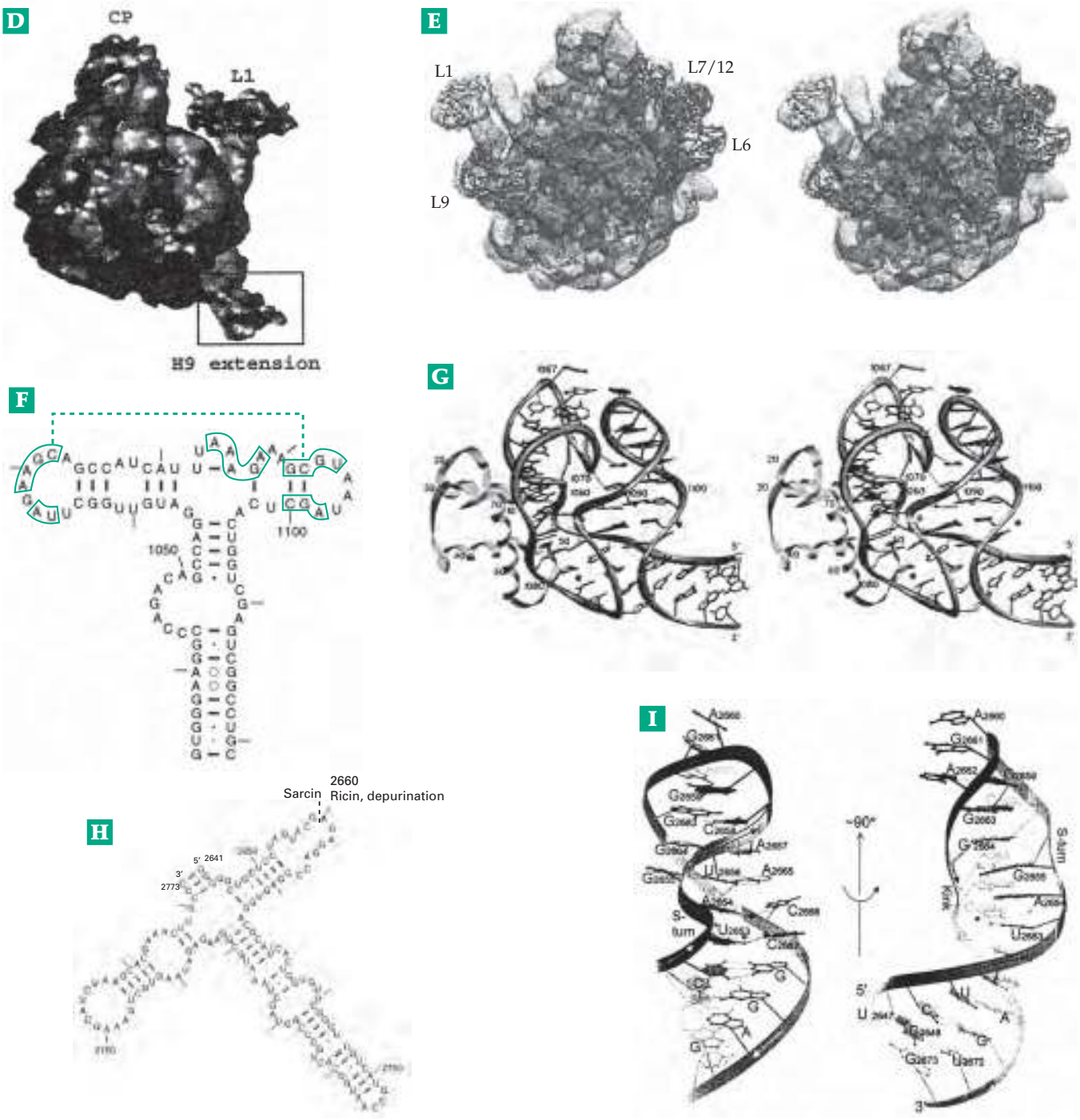


Figure 29-4 Structure of 23S–28S ribosomal RNAs. (A) The three-dimensional structure of RNA from the 50S subunit of ribosomes of *Haloarcula marismortui.* Both the 5S RNA and the six structural domains of the 23S RNA are labeled. Also shown is the backbone structure of protein L1. From Ban *et al.*[17] Courtesy of Thomas A. Steitz. (B) The corresponding structure of the 23S RNA from *Thermus thermophilus*. Courtesy of Yusupov *et al.*[33a] (C) Simplified drawing of the secondary structure of *E. coli* 23S RNA showing the six domains. The peptidyltransferase loop (see also Fig. 29-14) is labeled. This diagram is customarily presented in two halves, which are here connected by dashed lines. Stem-loop 1, which contains both residues 1 and 2000, is often shown in both halves but here only once. From Merryman *et al.*[78] Similar diagrams for *Haloarcula marismortui*[17] and for the mouse[79] reveal a largely conserved structure with nearly identical active sites. (D) Cryo-electron microscopic (Cryo-EM) reconstruction of a 50S subunit of a modified *E. coli* ribosome. The RNA has been modified genetically to have an



approximately 34 nt predicted extension of helix 9 of the 16S RNA (see Fig. 29-2). The helix 9 extension, clearly visible in this image, locates that helix in *E. coli* ribosomes, which have not yet given crystals satisfactory for X-ray investigation. From Matadeen *et al.*[79a] Courtesy of Richard Brimacombe. (E) Stereoscopic interface view of the 50S subunit of an *E. coli* ribosome with atomic structures of ribosomal proteins fitted to the cryo-EM density (semitransparent) of the 50S subunit. Protein structures are displayed as backbone tubes, and rRNA fragments in ball-and-stick format. Courtesy of Mueller *et al.*[37a] (F) The GTPase-activating loop of 23S RNA of *E. coli*. This loop, from domain II, binds to protein L11, which shields nucleotide A1067 from methylation and prevents the binding of the antibiotics thiostrepton (Box 29-B) and micrococcin. Green nucleotides are highly conserved in bacterial, chloroplast, and mitochondrial RNAs. The small loop (1054–1081) containing the thiostrepton-binding site is also part of the binding site for elongation factors EF-Tu and EF-G.[80,81] (G) Stereoscopic view of the 58-nucleotide loop shown in (E) with the associated protein L11. Courtesy of Conn *et al.*[82] (H) Secondary structure of the sarcin/ricin (SR domain) of the *E. coli* 23S RNA.[83–85] The site of hydrolytic cleavage by the ribonuclease sarcin (Box 29-A) is indicated as is the site of depurination catalyzed by the plant toxin ricin (Box 29-A). (I) Three-dimensional structure of the sarcin-ricin loop. The two views are from directions 90° apart. The sites of attack by ribotoxins are at the top. Courtesy of Correll *et al.*[83]

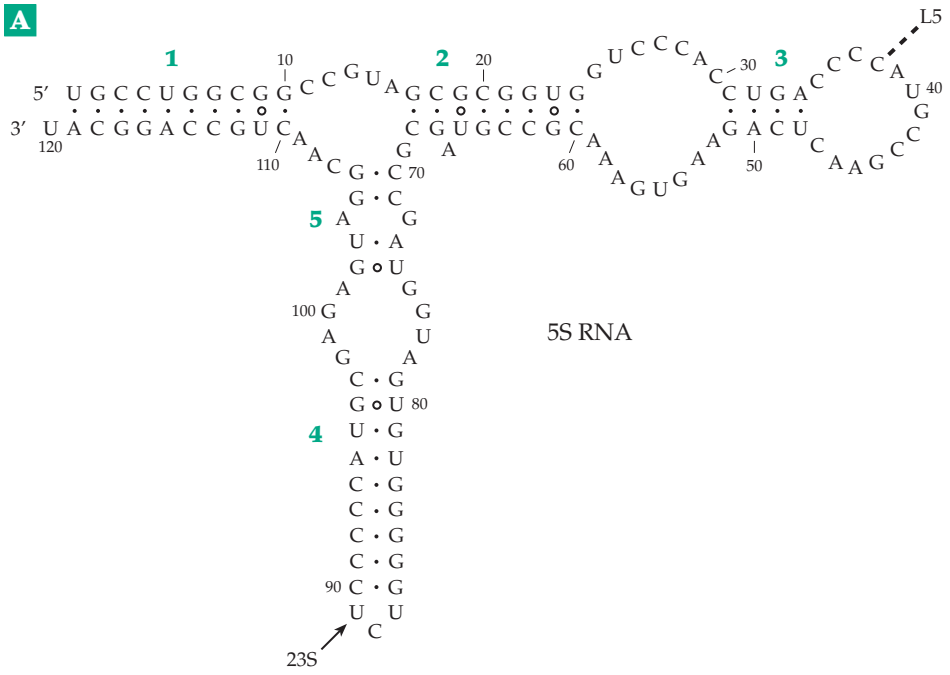


Figure 29-5 (A) Secondary structure of *E. coli* 5S RNA with five universal helical stems (labeled 1–5). This small RNA is found in the central protuberance of the 50S ribosomal subunit. See Fig 29-4A. Photocrosslinking using thiouridine-containing 5S RNA suggested a close proximity of U89 (marked by arrow) with nucleotide 2477 of the 23S RNA in the loop end of helix 89 (Fig. 29-4).[93] (B) Stereoscopic view of the 5S RNA as observed in ribosomes of *Haloarcula marismortui*. From Ban *et al*.[17] Courtesy of Thomas A. Steitz.

N-acetylated L12. Thus, the 50S ribosomal subunit is often described as containing four copies of protein L7/12. They form the flexible stalk seen in Fig. 19-1.

Most ribosomal proteins are rich in lysine and arginine and, therefore, carry a substantial net positive charge. Proteins S20, L7/12, and L10 have over 20% alanine, while L29 is almost as rich in leucine. Proteins S10, S13, L7/L12, L27, L29, and L30 are surprisingly low (<2 mol %) in aromatic amino acids. Proteins S5, S18, and L7 have acetylated N termini while L11, L3, L7/12, L11, L16, and L33 contain methylated amino acids. L11 contains nine methyl groups.[22] Protein S6 is the major phosphoprotein of eukaryotic ribosomes.[103,104] Most ribosomal proteins have no known enzymatic activity. Although often difficult to crystallize, high-resolution three-dimensional structures are known for many free ribosomal proteins.[24] Most of them have shapes resembling those previously found in globular proteins, including DNA-binding proteins. Many have extended "tails" that reach into the interior of the ribosome.[33b] A few seem to assume a defined shape only when packed into a suitable niche in the ribosome. Proteins L7/L12 and the 60-residue L30[105,105a] have similar folding patterns with 2–3 helices and a 3-strand β sheet. The structure of L30 of *E. coli* was deduced[106] by NMR methods (Fig. 3-25) and resembles that from the *Bacillus stearothermophilus* determined by X-ray diffraction.[105,107]

Many specific parts of ribosomal RNA molecules and specific proteins within the intact ribosome were located prior to the determination of high resolution crystal structures. One major approach was the use of **immunoelectron microscopy**. Antibodies to specific ribosomal proteins or to special sites in the RNA were prepared, and electron microscopy was used to map the binding sites of the antibodies on the ribosomal

subunit surfaces.[108,109] In this manner, the locations of numerous proteins in both the 30S and 50S subunits were identified. A few of these are indicated in Fig. 29-1A,B.[5] In several instances more than one distinct antibody binding site was found for a given protein. Pairs of sites were sometimes 8–19 nm apart, suggesting that these proteins assumed an elongated or fibrous conformation. However, X-ray studies have established more compact structures for many of the proteins. Perhaps the ease of denaturation of the proteins led to some errors in localization with antibodies. The X-ray studies have now established exact locations for almost all of the ribosomal proteins. However, the correct identification of each protein involved extensive measurements, many of which were done prior to the availability of the X-ray structures.

A variety of crosslinking reagents have been used to locate the positions of specific proteins within ribosomes. For example, bifunctional compounds may bind covalently to two different SH groups or NH_2 groups.[110,111] Among the many crosslinked protein pairs identified in this way are S5-S8, S7-S9, S6-S18, and S13-S19.[112] Crosslinking experiments on both small and large ribosomal subunits have yielded complex distance maps that helped to establish the packing relationships.[113]

Another important approach has been to isolate ribosomal proteins from bacteria grown in D_2O and then to reconstitute ribosomal subunits with pairs of deuterated proteins. By studying **neutron scattering** the distances between the centers of mass of these pairs could be measured. By triangulation the three-dimensional relationship of the entire group of proteins could be determined. The results of such studies[43,47,114] for the 30S subunit are shown in Fig. 29-1D. Most of the results are in agreement with those obtained by other methods. Neutron scattering from the 50S subunit was investigated by using pairs of protonated proteins in a subunit consisting of otherwise deuterated components. This gives an increase in sensitivity.[47]

3. RNA–Protein Interactions and Assembly of Ribosomes

Within bacterial cells the assembly of ribosomes is coupled to rRNA synthesis and requires only 1–2 minutes.[115] In the laboratory both the 30S ribosomal subunits[116] and the 50S subunits[117–121] of *E. coli* can be completely dissociated into individual protein and RNA molecules and can be reconstituted in a functional form. This is true for both natural 16S or 23S RNA or for RNAs prepared by *in vitro* transcription. In these reassembly experiments, which were pioneered by Nomura,[116] it was found that the order of addition of the protein is important. Some proteins bind directly to ribosomal RNA. For example, S4, S7, S8, S15, S17, and S20 bind directly to 16S RNA.[31] Other proteins bind only after one or more proteins have already bound and the RNA has folded properly to form a structural core (Fig. 29-6A). Domains I, II, and III each form an independent RNA–protein assembly. The lower half of domain I of the 16S RNA, from positions ~60–300, is unreactive toward single-stranded probes and may serve as one core for assembly of the ribosome.[67] Protein S20 binds to the 240–286 stem, which is in this core. Protein S4 also binds directly to 16S RNA in the 5' domain. Proteins S8 and S15 bind in the central domain and S7, which is structurally related to the DNA-binding proteins HU and IHF (Chapter 27),[24,122] binds near the 3' end.

5′- Domain of 16S RNA. The 23-kDa protein S4, one of the largest ribosomal proteins, appears to have an important organizing role for the 5' domain.[123,124] It binds in such a way as to protect sequences 27–47 and 394–556 of the RNA (Fig. 29-2) from chemical modification. The small loop at positions 323–330 is protected in the 30S subunit, and the residues A325, A327, A379, and G331 are universally conserved.[67] The same is true of bulge loop 505–510 and the loop sequence 518–533, which contains 7-methylguanine (m^7G) at position 526. Reconstitution experiments also suggested that S16 binds to S4 as well as to S20. Some mutations in proteins S4 and S5 are associated with reduced fidelity of translation, while others lead to spectinomycin resistance.[134]

Central domain of 16S RNA. Proteins S6, S8, S15, and S18 bind to the central domain II of 16S RNA (Fig. 29-2)[31,31a,67,125–129] and organize the platform region (Fig. 29-6). Protein S8 binds with high affinity to regions 588–606 and 632–651 of helix 21 and plays a key role in ribosomal assembly.[126,130,130a] S15 protects residues 655–672 and 734–751 of helix 22. The region contains functionally important conserved loops at positions 570–571, 766–768, and 811–820 as well as many individual adenines in other locations. S15 binds not only to the 16S RNA but also to the 715 loop of 23S RNA in the large subunit and to its own mRNA.[129] A Mg^{2+}-dependent conformational change in the RNA seems to be important in the assembly of the central domain.[131] S6 and S18 bind to 16S RNA after S15 has bound (Fig. 29-6). Proteins S11 and S21 also bind after S15.[129] S11 binds to the 690 loop of the RNA, as is illustrated in Fig. 29-6B. This loop is conserved in all three phylogenetic domains. Located in the platform of the small subunit, it protrudes into the interface to interact with domain IV of the 23S RNA and is also a site of binding of initiation factor IF3.[133] The mutant A649G in 16S RNA confers resistance to **pactamycin** in *E. coli*. Protein S8 is not only an important structural protein in the central domain but also

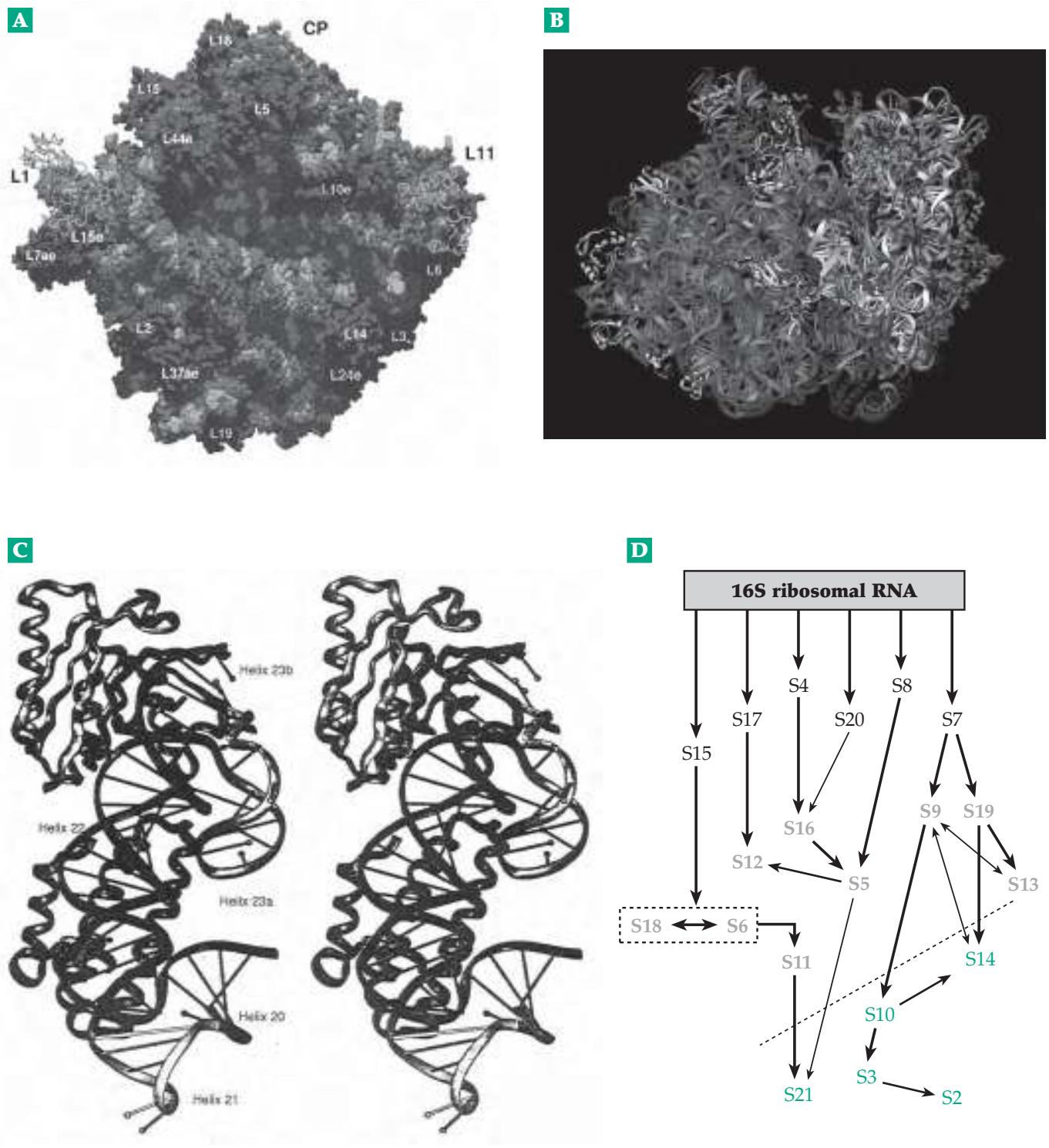
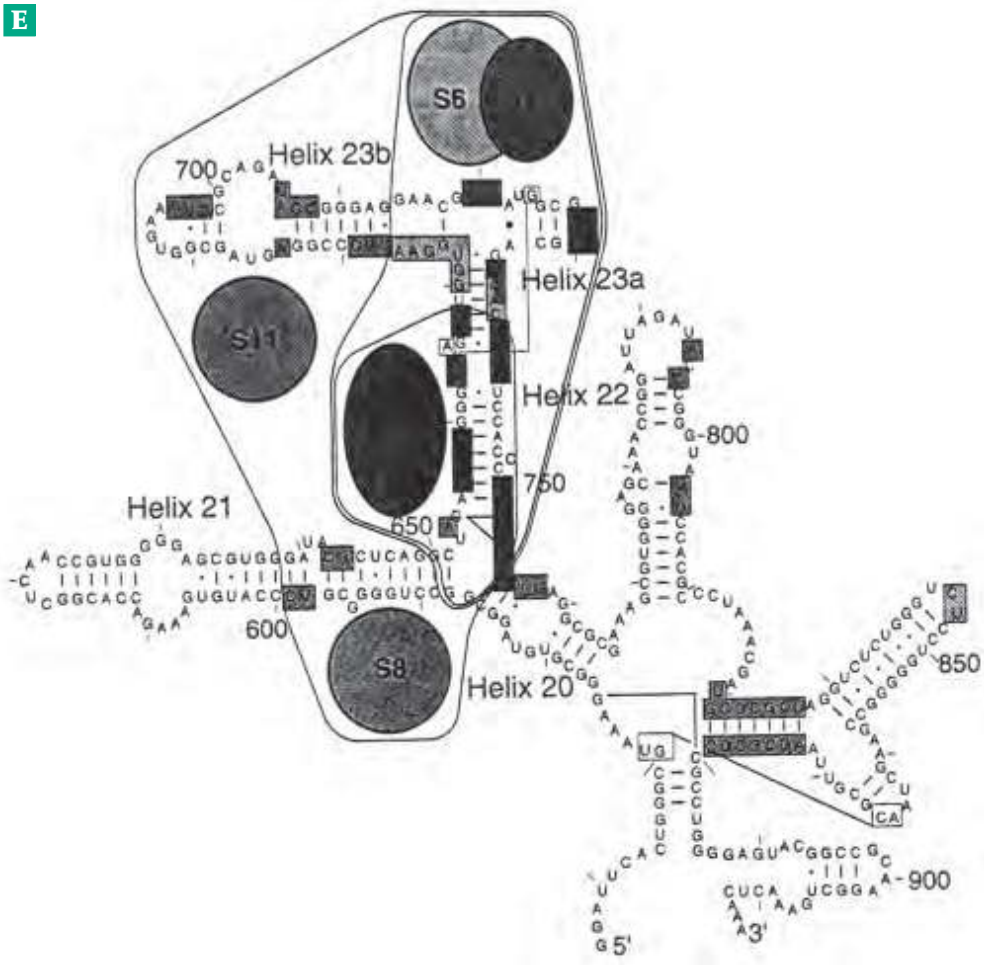
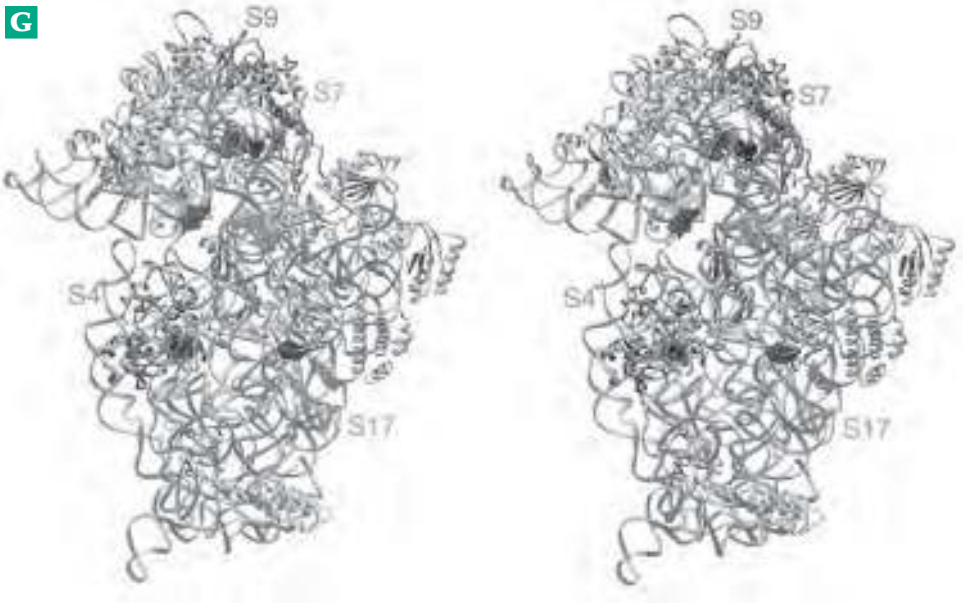


Figure 29-6 Some protein–RNA interactions within the ribosome. (A) A space-filling model of the 23S and 5S RNA with associated proteins from the ribosome of *Haloarcula marismortui*. The CCA ends of bound tRNA molecules in the A, P, and E sites are also included. The view is looking into the active site cleft. The proteins with e after the number are related to eukaryotic ribosomal proteins more closely than to those of *E. coli*.[17] Courtesy of T. A. Steitz. (B) Three-dimensional structure of a 70S ribosome from *Thermus thermophilus*. The 30S subunit is to the right of the 50S subunit. Courtesy of Yusupov *et al*.[33a] (C) Stereoscopic view of the helix 21 to helix 23b region of the 16S RNA with associated proteins S6 (upper left), S18 (upper center, front), and S15 (lower back) from *T. thermophilus*. Courtesy of Agalarov *et al*.[31] (D) Simplified *in vitro* assembly map of the central domain of the 30S bacterial ribosome. Courtesy of Gloria Culver. (E) Contacts of proteins with the central (platform) domain of the 16S RNA component. The sequence shown is that of *Thermus thermophilus*. Courtesy of Agalarov *et al*. (F) Three drawings showing alternative location of the four copies of protein L7/L12. The N-terminal and C-terminal





globular domains (labeled on one of the four molecules) are connected by a flexible region. One pair of L7/L12 molecules retains a fixed position toward the center of the 50S subunit but the C-terminal domains of the other pair are seen to occupy three different positions. Courtesy of Montesano-Roditis *et al.*[146a] (G) Stereoscopic view of the 30S subunit of the *T. thermophilus* ribosome with six positions at which tetracycline binds and one at which the peptide-like antibiotic **edeine** binds. From Pioletti *et al.*[146b]

acts as a translational repressor of its own gene and of other genes of the spectinomycin-resistance operon (see Box 29-B). This operon encodes genes for ten ribosomal proteins, of both the large and small subunits.[126] S8 is a neighbor to proteins S2, S4, S5, S12, S15, and S17.

3′- Domain of 16S RNA. Domain III of 16S RNA binds proteins S2, S3, S7, S9, S10, S13, S14, and S19. Proteins S3 and S14 may be necessary for assembly of ribosomes but may no longer be needed once the 30S subunit has been correctly formed.[135] The largest of the *E. coli* ribosomal proteins is the 556-residue S1. It does not bind tightly and has sometimes been regarded as nonessential. However, mutations in the S1 gene can be lethal, and the protein seems to be essential for both initiation of translation and for elongation of polypeptide chains *in vivo*.[137] S1 behaves as an elongated molecule 22 nm in length[135] and is unusual in having an ~86-residue RNA-binding sequence repeated four times in the central and N-terminal regions. The protein possesses RNA-unwinding activity[137,138] and may employ these four motifs in unwinding mRNA as it enters the ribosome.

23S and 5S RNAs. Reconstitution of the large ribosomal subunit reveals that proteins L3 and L24 act as assembly initiators.[115,118] L1, L9, L20, and several other proteins (Table 29-2) also bind directly and independently to the 23S RNA. Assembly maps similar to that in Fig. 19-6A have been prepared for the 50S subunit.[117]

One of the most prominent features of the 50S subunit is the L1 protuberance, seen on the left side in Fig. 29-6A. This protuberance is formed almost entirely by protein L1, which is one of the largest ribosomal proteins. It binds to the 2105–2184 loop in domain V of the 23S RNA (see Fig. 29-14).[139] L1 has an important regulatory role in bacteria in which it represses translation of its own structural gene by binding to a region in its mRNA close to the Shine–Dalgarno sequence. The polygenic mRNA also carries the code for protein L11.[139] This is one of several examples of such autogenous regulation of translation of ribosomal proteins.[139a] L1 also interacts in the ribosome with the 5S rRNA.[140] The 272-residue L2 also associates directly with 23S RNA and assists in ribosome assembly.[141] Protein L2 is one of the structurally most highly conserved of the ribosomal proteins.[46,142] It binds to the 1794–1865 region of domain IV of 23S RNA. Histidine 229 of this protein may play a functional role in the ribosome. The protein is elongated, and one end contacts 16S RNA.[46,33a] Protein L9 binds to domain V of 23S RNA in the 2100–2190 region. It is an elongated molecule with two globular α/β domains separated by an α helix. This enables it to bind also to domain III, acting as a rigid strut.[143–145]

On the right side of the 50S subunit, as viewed in Fig. 29-1, is the stalk, a pentameric protein complex consisting of two L7/L12 (*E. coli*) or $(L12)_2$ dimers bound to one molecule of L10.[24,146–147] The stalk is not always seen in X-ray structures, e.g., in Fig. 29-6A, and appears to be flexible. In crosslinking experiments the N-terminal domains of L7/L12 can be linked to L10 and also to its neighbor, L11,[82,148,149] which lies in the GTPase-activating center (Fig. 29-4F) at the base of the stalk. However, the C-terminal domains can be crosslinked to three distinctly different locations: to L11 on the platform surface, to L2 and L5 near the peptidyltransferase center, and to S2, S3, and S14 of the head and neck of the 30S subunit.[146] Domain I of 23S RNA, near the 5' end, binds to protein L20.[153]

An independent and essential structural domain of the ribosome is formed around the 5S RNA.[5,108,154–156] Proteins L5, L18, and L25, whose structure is similar to that of glutaminyl-tRNA synthetase,[154] bind specifically to one loop of the 5S RNA.[156a] Furthermore, the L5–L18–L25–5S RNA complex binds the oligonucleotide TCC. This suggests an interaction between the 5S RNA and the TΨC arm of a tRNA molecule bound to the ribosome. In addition, it has been observed that L18 + either L5 or L25 cause 5S RNA to bind to 23S RNA.

Eukaryotic ribosomal proteins. The functions of the 70–80 different eukaryotic ribosomal proteins are less well known than those of *E. coli*. In eukaryotes the assembly of ribosomes begins in the nucleus with binding of proteins to the individual ribosomal RNA precursors (Chapter 28).[121,156b] Significant functional properties that are peculiar to eukaryotic ribosomal proteins include the following: S6 is the site of multiple phosphorylation reactions, which control initiation of protein synthesis.[132,132a,132b] Mammalian S3 may function in the nucleus in DNA repair.[136] Eukaryotic proteins P0, P1, and P2 are homologous to *E. coli* stalk proteins L10, L7, and L12, respectively. Higher eukaryotes possess only one type of P1 and P2,[150] but yeast,[150a,b] maize,[151] and other species have multiple forms. An L7-related protein is also required for a nucleolar function in ribosomal protein synthesis, perhaps as a component of a snoRNP complex (Chapter 28).[152] Rat liver L37 is involved in peptidyltransferase, but sequencing of the 111-residue protein reveals homology with *E. coli* L34 rather than with L16.[157] Proteins L14, L21, L24, L27, L29, and L30 bind to the 5.8S RNA of the large subunit of yeast ribosomes.[158]

Yeast protein L30, which is not homologous to any bacterial protein, controls its own synthesis by a feedback inhibition at the mRNA splicing step. L30 binds to its own pre-mRNA near the 5' splice site, blocking completion of the spliceosome assembly (Chapter 28).[159]

BOX 29-A THE DIPHTHERIA TOXIN AND OTHER RIBOSOME-INACTIVATING PROTEINS

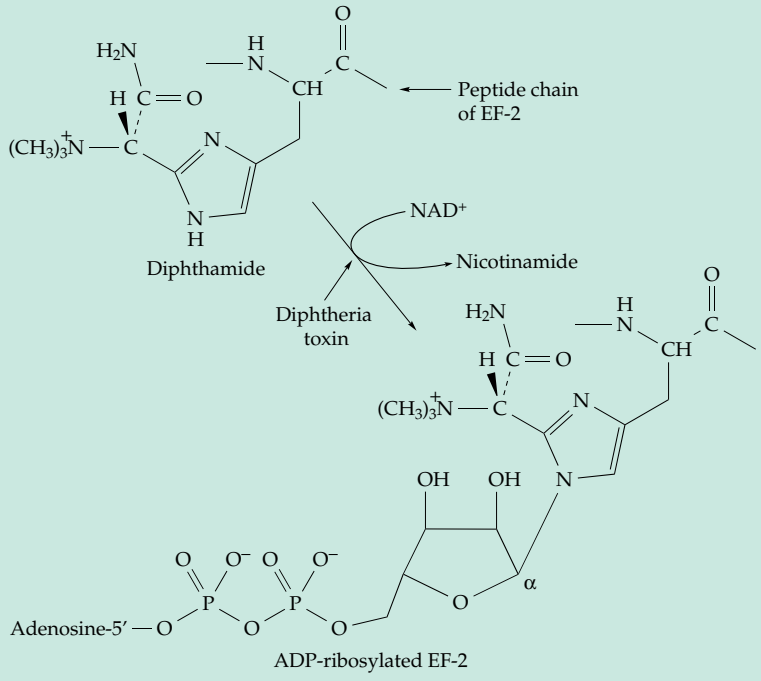
Until a suitable vaccine was developed, an infection by *Corynebacterium diphtheriae* was one of the dread diseases of childhood. Despite the fact that the bacteria caused only superficial membranous lesions in the throat, the patient often died with evident damage to many organs. The cause is a potent heat-labile toxic protein,[a–d] which the bacterium produces when infected by a temperate bacteriophage carrying the *tox* gene and when the inorganic iron of the surroundings has been largely depleted. Diphtheria toxin is a 535-residue protein with a minimum lethal dose (LD_{50}) of only 0.16 mg kg^{-1} for the guinea pig. Tests in cell culture show that the toxin blocks incorporation of amino acids into proteins by inactivation of the eukaryotic elongation factor EF2, which is required for **translocation**, an essential step in protein synthesis in mammalian ribosomes. The toxin acts as an enzyme that transfers (with inversion at the ribose C1) an ADP-ribosyl group from NAD^+ to a side-chain ring nitrogen of the single residue of **diphthamide** in EF2. This modified histidine is found in EF2 and, apparently, in no other protein.[d]

The modified elongation factor reacts normally with GTP, but the complex so formed is unable to participate in translocation. A concentration of the toxin in the cytoplasm of 10^{-8} M is sufficient to promote the fatal reaction. The reaction with diphthamide parallels that of cholera toxin (Box 11-A).

The diphtheria toxin molecule[e,f] consists of three domains, an N-terminal catalytic (C) domain (residues 1–193), a central, largely α-helical, transmembrane (T) domain (residues 205–378), and the C-terminal receptor-binding (R) domain (residues 386–535). Before it enters a cell, the toxin molecule is "nicked" by protease activity between the catalytic and transmembrane domains, a step that is apparently necessary for binding to its receptor, which has been identified as a heparin-binding EGF-like growth factor precursor.[g] The catalytic domain (also called the A fragment) enters the cell through endocytosis from coated pits. Additional proteolytic cleavage, as well as reduction of a disulfide linkage, are required for activity.[f,h] After entering the cytosol the toxin fragment catalyzes inactivation of EF2.[f] A single molecule can kill a cell. The larger 613-residue exotoxin A of *Pseudomonas aeruginosa* catalyzes the same reaction as does diphtheria toxin. It also consists of three domains: a large β-sheet-containing N-terminal region, a central α-helical domain, and a C-terminal domain. The last contains the ADP-ribosyltransferase active site.[a,i]

What is the origin of the *tox* gene, and why is it carried by a virus? Cells do normally contain ADP-ribosyltransferases.[j] The genes for such a protein may have become incorporated into a virus and, after a period of evolution, came to specify the toxic protein.

Another family of toxins attacks ribosomes in a very different way, cleaving ribosomal RNA at specific sites. One of the best known of these is the neurotoxin from *Shigella dysentariae* (Shiga toxin). Like the cholera toxin (Box 11-A) it consists of a single catalytic A subunit and a pentameric ring of B subunits,[k] which binds to specific surface glycolipids.[l] **Verotoxin**, another poison from certain strains of *E. coli*, has a similar structure. Although they have very different effects and there is no detectable similarity in their amino acid sequences, the *structures* of the B pentamers of verotoxin and of the cholera toxin-like heat-labile enterotoxin of *E. coli* are similar.[m,n] The A subunit of Shiga toxin hydrolyzes the *N*-glycosyl linkage of adenine to the ribose ring at position 4324 of 28S ribosomal RNA.[k] A number of plants form very toxic lectins: **ricin** (from castor bean),[o] **viscumin** (from mistletoe), **modeccin**, **abrin**,[p] gelonin,[q] and **volkensin**.[o] The names are derived from the genus names of the plants. All appear to be glycoproteins consisting of two disulfide-linked chains, one of which is a lectin. The



BOX 29-A THE DIPHTHERIA TOXIN AND OTHER RIBOSOME-INACTIVATING PROTEINS (continued)

lectin subunits of ricin[r] and of volkensin[s] bind to galactose residues. The A chains are cytotoxins, which enter cells and, like Shiga toxin, inactivate 60S ribosomal subunits. The 267-residue A chain of ricin is similar to that of a pokeweed viral antigen[t] and of Shiga toxin. It catalyzes the same reaction,[u] the depurination of adenosine 4324. (The pokeweed toxin also catalyzes the corresponding reaction with A2660 of the *E. coli* 23S RNA.[v]) Like the diphtheria toxin these toxic proteins bind to cell surface receptors, are taken up by endocytosis, and are transported through the Golgi to the endoplasmic reticulum. Their structures facilitate uptake but allow them to escape degradation in proteasomes.[w]

Ricin is one of the most toxic substances known. A single molecule can inactivate over 1700 ribosomes per minute and kill the cell.[u] With an LD_{50} of only 1 µg / kg of body weight for many animals, ricin has been used as a poison by assassins. Of more importance is the attempt to couple ricin and related toxins to immunoglobulins to produce **immunotoxins** that will attack cancer cells (Box 31-A). A related goal is to design a potent inhibitor that could serve as an antidote.[x] It is fortunate that most plant seeds do not contain toxins like ricin. Many plants, including such important food grains as wheat and barley, do contain ribosome-inactivating proteins similar to the A chain of ricin. However, the plants lack the B (lectin) subunits and do not enter animal cells.

A group of unusual fungal ribonucleases, which includes **α-sarcin** and **restrictocin**, are produced by *Aspergillus*. The cytotoxic nucleases enter animal cells, where they cut the 28S RNA of ribosomes, specifically on the 3' side of guanosine 4325 in the sarcin / ricin domain (see Fig. 29-4), thereby blocking protein synthesis.[u,y,z] *Staphylococcus aureus* produces a 22-kDa toxic protein thought to be responsible for **toxic shock syndrome**.[aa] Another toxic ribonuclease is **colicin E3** (Box 8-D), which cuts the 16S RNA of *E. coli* after nucleotide 1493 (see Fig. 29-1A).[bb] Colicin D stops protein synthesis by cleavage of four isoaccepting $tRNA^{Arg}$ molecules between positions 38 and 39 in the anticodon loop.[cc]

[a] Han, X. Y., and Galloway, D. R. (1995) *J. Biol. Chem.* **270**, 679–684
[b] Pappenheimer, A. M., Jr. (1977) *Ann. Rev. Biochem.* **46**, 69–94
[c] Pappenheimer, A. M. J. (1993) *Protein Sci.* **2**, 292–298
[d] Ward, W. H. J. (1987) *Trends Biochem. Sci.* **12**, 28–31
[e] Choe, S., Bennett, M. J., Fujii, G., Curmi, P. M. G., Kantardjieff, K. A., Collier, R. J., and Eisenberg, D. (1992) *Nature (London)* **357**, 216–222
[f] Weiss, M. S., Blanke, S. R., Collier, R. J., and Eisenberg, D. (1995) *Biochemistry* **34**, 773–781
[g] Shen, W. H., Choe, S., Eisenberg, D., and Collier, R. J. (1994) *J. Biol. Chem.* **269**, 29077–29084
[h] Blanke, S. R., Huang, K., Wilson, B. A., Papini, E., Covacci, A., and Collier, R. J. (1994) *Biochemistry* **33**, 5155–5161
[i] Douglas, C. M., and Collier, R. J. (1990) *Biochemistry* **29**, 5043–5049
[j] Eide, B., Gierschik, P., and Spiegel, A. (1986) *Biochemistry* **25**, 6711–6715
[k] Kozlov, Y. V., Chernaia, M. M., Fraser, M. E., and James, M. N. G. (1993) *J. Mol. Biol.* **232**, 704–706
[l] Saleh, M. T., and Gariépy, J. (1993) *Biochemistry* **32**, 918–922
[m] Stein, P. E., Boodhoo, A., Tyrrell, G. J., Brunton, J. L., and Read, R. J. (1992) *Nature (London)* **355**, 748–750
[n] Sixma, T. K., Stein, P. E., Hol, W. G. J., and Read, R. J. (1993) *Biochemistry* **32**, 191–198
[o] Lord, J. M., Roberts, L. M., and Robertus, J. D. (1994) *FASEB J.* **8**, 201–208
[p] Tahirov, T. H., Lu, T.-H., Liaw, Y.-C., Chen, Y.-L., and Lin, J.-Y. (1995) *J. Mol. Biol.* **250**, 354–367
[q] Hosur, M. V., Nair, B., Satyamurthy, P., Misquith, S., Surolia, A., and Kannan, K. K. (1995) *J. Mol. Biol.* **250**, 368–380
[r] Weston, S. A., Tucker, A. D., Thatcher, D. R., Derbyshire, D. J., and Pauptit, R. A. (1994) *J. Mol. Biol.* **244**, 410–422
[s] Stirpe, F., Barbieri, L., Abbondanza, A., Falasca, A. I., Brown, A. N. F., Sandvig, K., Olsnes, S., and Pihl, A. (1985) *J. Biol. Chem.* **260**, 14589–14595
[t] Marchant, A., and Hartley, M. R. (1995) *J. Mol. Biol.* **254**, 848–855
[u] Glück, A., and Wool, I. G. (1996) *J. Mol. Biol.* **256**, 838–848
[v] Chan, Y.-L., Sitikov, A. S., and Wool, I. G. (2000) *J. Mol. Biol.* **298**, 795–805
[w] Sandvig, K., and van Deurs, B. (2000) *EMBO J.* **19**, 5943–5950
[x] Yan, X., Hollis, T., Svinth, M., Day, P., Monzingo, A. F., Milne, G. W. A., and Robertus, J. D. (1997) *J. Mol. Biol.* **266**, 1043–1049
[y] Endo, Y., Chan, Y.-L., Lin, A., Tsurugi, K., and Wool, I. G. (1988) *J. Biol. Chem.* **263**, 7917–7920
[z] Nayak, S. K., Bagga, S., Gaur, D., Nair, D. T., Salunke, D. M., and Batra, J. K. (2001) *Biochemistry* **40**, 9115 – 9124
[aa] Blomster-Hautamaa, D. A., Kreiswirth, B. N., Kornblum, J. S., Novick, R. P., and Schlievert, P. M. (1986) *J. Biol. Chem.* **261**, 15783–15786
[bb] Lasater, L. S., Cann, P. A., and Glitz, D. G. (1989) *J. Biol. Chem.* **264**, 21798–21805
[cc] Tomita, K., Ogawa, T., Uozumi, T., Watanabe, K., and Masaki, H. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8278–8283

4. Locating Active Sites in Ribosomes

In early studies, antibodies against haptens covalently linked to ends of the 16S RNA were used to locate the 3' end of 16S RNA on the upper "platform" of the 30S subunit (Fig. 29-1A).[5,160] The 5' terminus was found in the lower body. Two N^6,N^6-dimethyladenosines occur at positions 1518 and 1519, about 25 residues from the 3' end of the 16S RNA (Fig. 29-2). Antibodies were used to locate this position on the lower platform. Likewise, the m^7G at position 526 lies in the "neck" as shown in Fig. 29-1A.[161] Taking into account known protein–RNA interactions, domain II of the 16S RNA was located in the "platform" on the

upper left side of the "body" (viewed from the "outside" as in Fig. 29-1C), while domain III is in the "head." Recent structural studies have confirmed these biochemical localizations.[19,29,33,33a]

The tRNA-binding sites. During protein synthesis tRNA is bound sequentially in at least three places located between the 30S and 50S subunits. These are known as the **A** (**aminoacyl**), **P** (**peptidyl**), and **E** (**exit**) sites. The latter binds deacylated tRNA before it is released from the ribosome.[162,163] Because tRNA is such a large molecule, these sites have subsites in both 30S and 50S ribosomal subunits. When in the P site, a tRNA has its anticodon held firmly and base-paired with a codon in the mRNA in the decoding site of the 30S subunit. The CCA 3'-end with its attached peptidyl chain lies in the 50S subunit at the peptidyltransferase center. A "charged" aminoacyl-tRNA enters the A site, which is close to the 5S RNA in the central protuberance of the large ribosomal subunit, with its anticodon in the decoding site and its aminoacyl group at the peptidyltransferase site.

One end of the P site must be close to the 3' end of the 16S RNA near the two m_2^6A residues (Fig. 29-1A). This conclusion, which was based on photochemical linking of a hypermodified base at position 34 (see Fig. 29-7) in $tRNA^{Val}$ with C-1400 of 16S RNA by cyclobutane dimer formation (Eq. 23-26),[164] has been confirmed by structural studies.[29,33,33d] Investigation of tRNA binding, effects of mutations in ribosomal RNA, and effects of antibiotics pointed to locations of the P and A sites in both ribosomal subunits. These have been located precisely by crystallography. See Fig. 29-1F,J; 29-2C. Residue 6530 together with nucleotides 921–927, 1390–1394, and 1491–1505 of 16S RNA participate in forming the form A and P sites in the decoding center.[33c,d,164a,b,378] The two adjacent adenine rings of A1492 and A1493 swing out from helix 44 (Fig. 29-2; see also Fig. 29-14) to form a major part of the A site. In the 50S subunit the adjacent cytosines C74 and C75 of the CCA 3′-ends of the tRNAs in the A and P sites interact respectively with G2553 of the A loop and G2252 of the P loop (Fig. 29-14B,E).[164b–e] Tetracycline (Fig. 22-7) also binds into the A site (see Box 29-B). It can be photochemically crosslinked to proteins S18 and S4.[167]

The peptidyltransferase site. The position was located by binding of derivatives of the antibiotic **puromycin** (Fig. 29-13). An arylazide derivative of puromycin was photochemically linked (Eq. 23-27) to proteins L23, L18/22, and L15; immunoelectron microscopy, using antibodies to the N^6-dimethyladenosine of puromycin,[165,166] located the binding site adjacent to the central protuberance between the 50S subunit and 30S subunit near S14.[5] 4-Thio-dT-p-C-p-puromycin was photochemically crosslinked to G2553 of the peptidyltransferase A site (see Fig. 29-14). X-ray data provided a precise structure of the peptidyltransferase site (see pp. 1702–1704).[166a] Studies of mutant ribosomes together with affinity labeling and crosslinking experiments pointed to the **peptidyltransferase loop** marked on Fig. 29-4 and further illustrated in Fig. 29-14.[164a,167a,b]

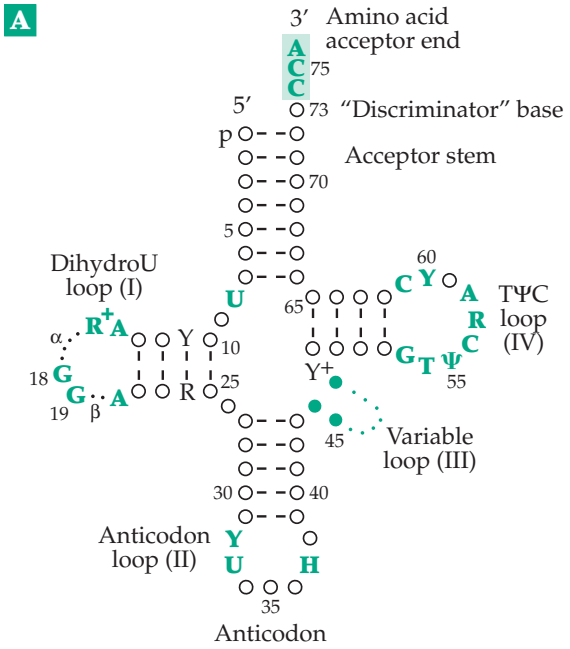
The GTPase-activating center. Also shown in Fig. 29-1C is a site for binding of the elongation factor EF-G (Section C,2). This was located, in part, because the antibiotic **thiostrepton** prevents EF-G from binding to the ribosome. Thiostrepton binds to a complex of protein L11 and a 61-fragment of the 23S RNA (positions 1052–1112; see Fig. 29-4F).[80] Another elongation factor, EF-Tu, also binds at the same site or adjacent to the EF-G site by the head of the small subunit.[168] An additional location of interest is the **polypeptide exit tunnel**, which brings the growing protein chain from the peptidyltransferase site out of the ribosome (Fig. 29-1I).[5,6,33f]

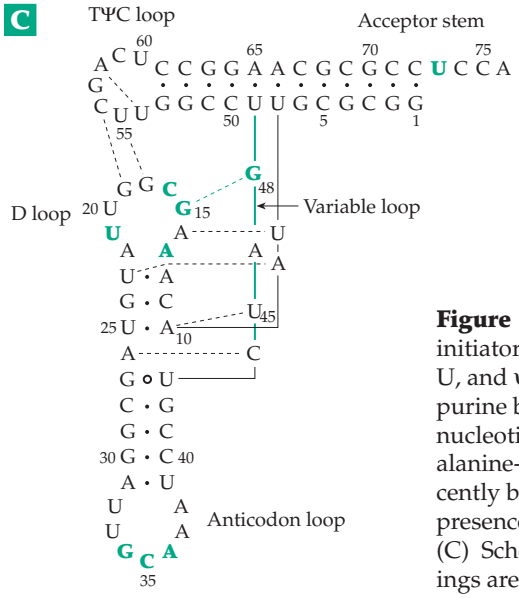
B. Transfer RNAs

The small 4S tRNA molecules have masses of ~26 kDa and consist of 75 ± 5 nucleotides (Figs. 5-30, 5-31, and 29-6). The basic structures are similar in bacteria and eukaryotic cells. The need for "adapters," to carry amino acids to the proper positions along the mRNA template, had been predicted prior to the discovery of tRNA.[169,170] It had been expected that there would be a base sequence constituting an **anticodon**, which would fit against the proper codon at some binding site on the protein-synthesizing machinery. This is just what tRNA molecules do, but their chemistry contained many surprises.

1. Structures of Transfer RNAs

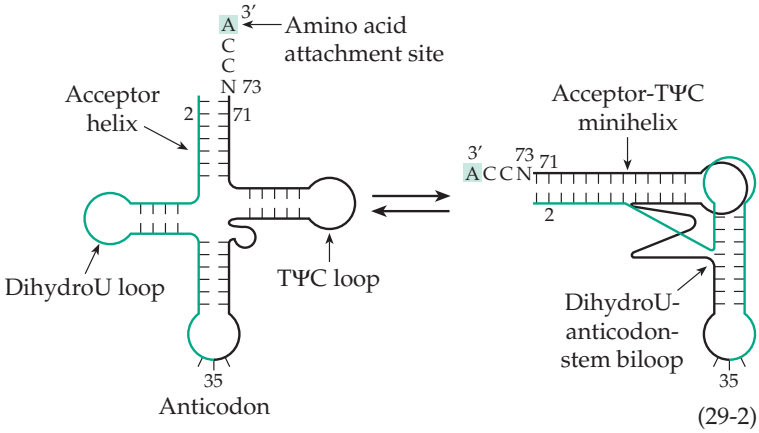
The first surprise was that these molecules are much longer than seems necessary for the formation of adapters. In addition, 10–20% of their bases are modified greatly from their original form.[171] Another surprise was that the anticodons are not all made up of "standard" bases. Thus, hypoxanthine (whose nucleoside is inosine) occurs in some anticodons. Conventional "cloverleaf" representations of tRNA, which display their secondary structures, are shown in Figs. 5-30 and 29-7. However, the molecules usually have an L shape rather than a cloverleaf form (Figs. 5-31 and 29-6),[172] and the L form is essential for functioning in protein synthesis as indicated by X-ray and other data.[173] Three-dimensional structures, now determined for several different tRNAs,[174,175] are all very similar. Structures in solution are also thought to be





similar for the various tRNA molecules.[176,177] One of the four hydrogen-bonded "stems" of a tRNA in the cloverleaf form terminates in the universally conserved CCA-3′ **amino acid acceptor end** (Fig. 29-7), which can carry an esterified amino acid generated as in Eq. 29-1, steps *a*, *b*. The other three stems terminate in loops, which usually contain a large number of modified bases. The modifications may serve to optimize the interaction of the tRNA with other components of the protein-synthesizing machinery.[178] The **dihydroU loop** (loop I) contains 5,6-dihydrouridine in various amounts and in varying positions. The **anticodon loop** (loop II) always contains the anticodon directly opposite the amino acid acceptor end in the cloverleaf drawing. On the 5' side of the anticodon at position 33 there is almost always a U (shaded in Fig. 29-7A) preceded by another pyrimidine. A hydrogen bond from the N3 proton of U-33 and a phosphate oxygen of residue 36 stabilize the U-turn that precedes the anticodon triplet (Fig. 5-31).[177] Next to the 3' side of the anticodon there is usually a **hypermodified** base, such as N^6-(Δ^2-isopentenyl) adenosine (Fig. 5-33) or a more complex derivative.[178] The **variable loop** (loop III) can range between 5 and 21 nt in length.[177,181] The **TψC loop** (loop IV) contains the specific nucleotide sequence for which the loop is named.

Cloverleaf and L forms. Interconversion between the cloverleaf and L forms of tRNA molecules can be pictured as in Eq. 29-2. Notice that in the L



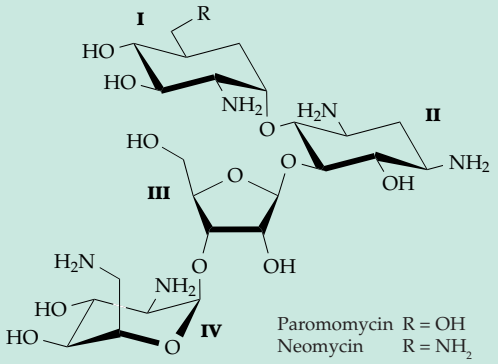
(29-2)

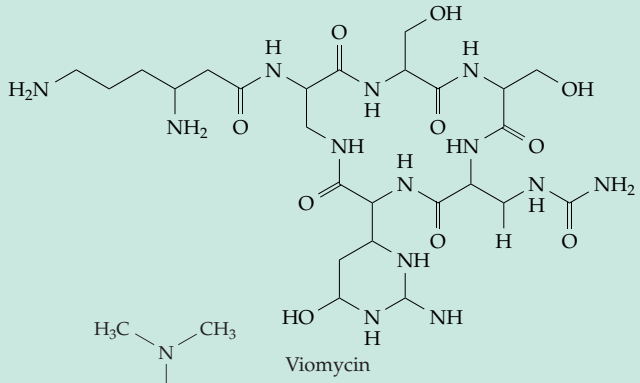
Figure 29-7 (A) Generalized cloverleaf diagram of all tRNA sequences except for initiator tRNAs numbered as in yeast $tRNA^{Phe}$ (Fig. 5-30). Invariant bases: A, C, G, T, U, and ψ; semivariant bases: Y (pyrimidine base), R (purine base), H (hypermodified purine base). The dotted regions (α, β, variable loop) contain different numbers of nucleotides in various tRNA sequences. See Rich.[179] (B) L form of the yeast phenylalanine-specific $tRNA^{Phe}$. The structure is the same as that in Fig. 5-31 but has recently been redetermined at a resolution of 0.20 nm.[175] The new data revealed the presence of ten bound Mg^{2+} ions (green circles) as well as bound spermine (green). (C) Schematic representation of L form of *E. coli* $tRNA^{Cys}$. Some tertiary base pairings are indicated by dashed lines. No modified bases are shown. See Hou *et al.*[180]

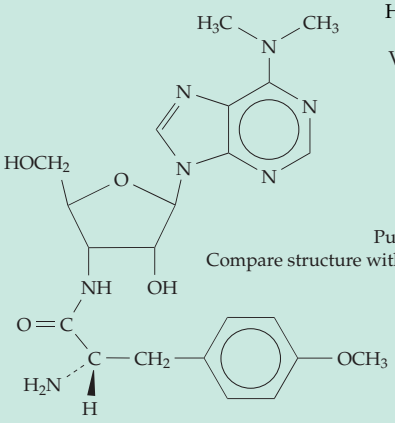
BOX 29-B ANTIBIOTICS THAT INHIBIT PROTEIN SYNTHESIS

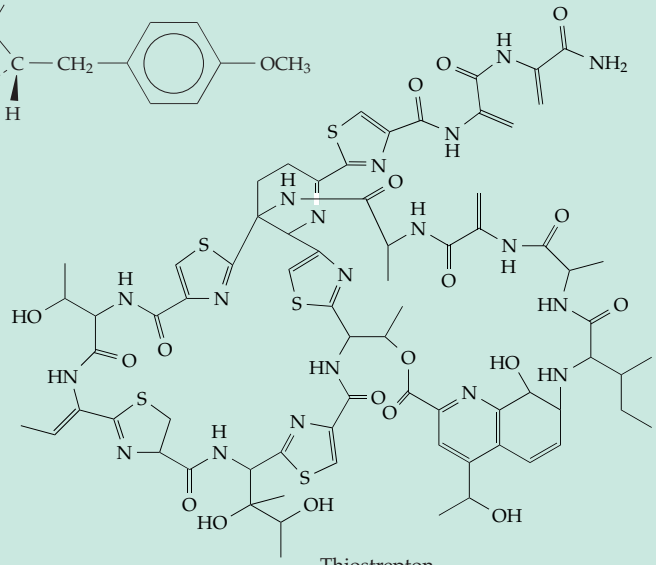
Many of our most effective antibiotics act by blocking protein synthesis on ribosomes.[a–de] The usefulness of many of these remarkable drugs in human medicine depends upon the fact that they inhibit protein synthesis on bacterial 70S ribosomes but do not seriously affect eukaryotic ribosomes. Antibiotics act at a variety of sites involving both ribosomal proteins and rRNA. Some of the resulting points of inhibition of the ribosomal elongation cycle[e] are marked on Fig. 29-12. Binding sites for many antibiotics have been located by footprinting experiments, by studies of mutants, and more recently by X-ray crystallography. Exposure of bacteria to antibiotics often gives rise to antibiotic-resistant mutants. In some of these mutants a specific ribosomal protein has been altered, but in others a specific RNA base has been changed.

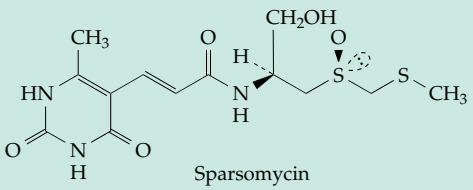
One of the first antibiotics to be studied was **puromycin** (Fig. 29-13), which binds to the 50S subunit and causes premature termination of peptide synthesis. A glance at its structure reveals how it can do this. It resembles in fine detail the 3' end of a tRNA molecule bearing an aminoacyl group. However, it is not an aminoacyl group, and once the growing peptide chain has been transferred onto the puromycin, further chain elongation is impossible. It was shown in 1975 that puromycin could be crosslinked to several different proteins in the ribosome by ultraviolet irradiation.[f,g] L23 of the 50S subunit and S14 of the 30S subunit were labeled most heavily. More recently puromycin, acting on 70S ribosomes, has been shown to label protein S7 and several large subunit proteins near the peptidyltransferase center in the central domain V of the 23S RNA.[h] A puromycin-derived transition state inhibitor has permitted precise identification of the peptidyltransferase site (Fig. 29-13).[i,j] In earlier work immunoelectron microscopy on *N*-bromoacetylpuromycin-labeled 50S subunits had located the site marked Pm in Fig. 29-1A. However, this is quite far from the recently determined location of the peptidyltransferase center,[k]

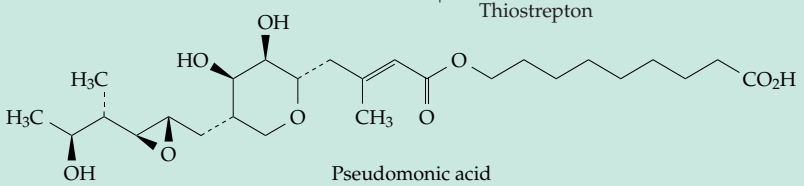












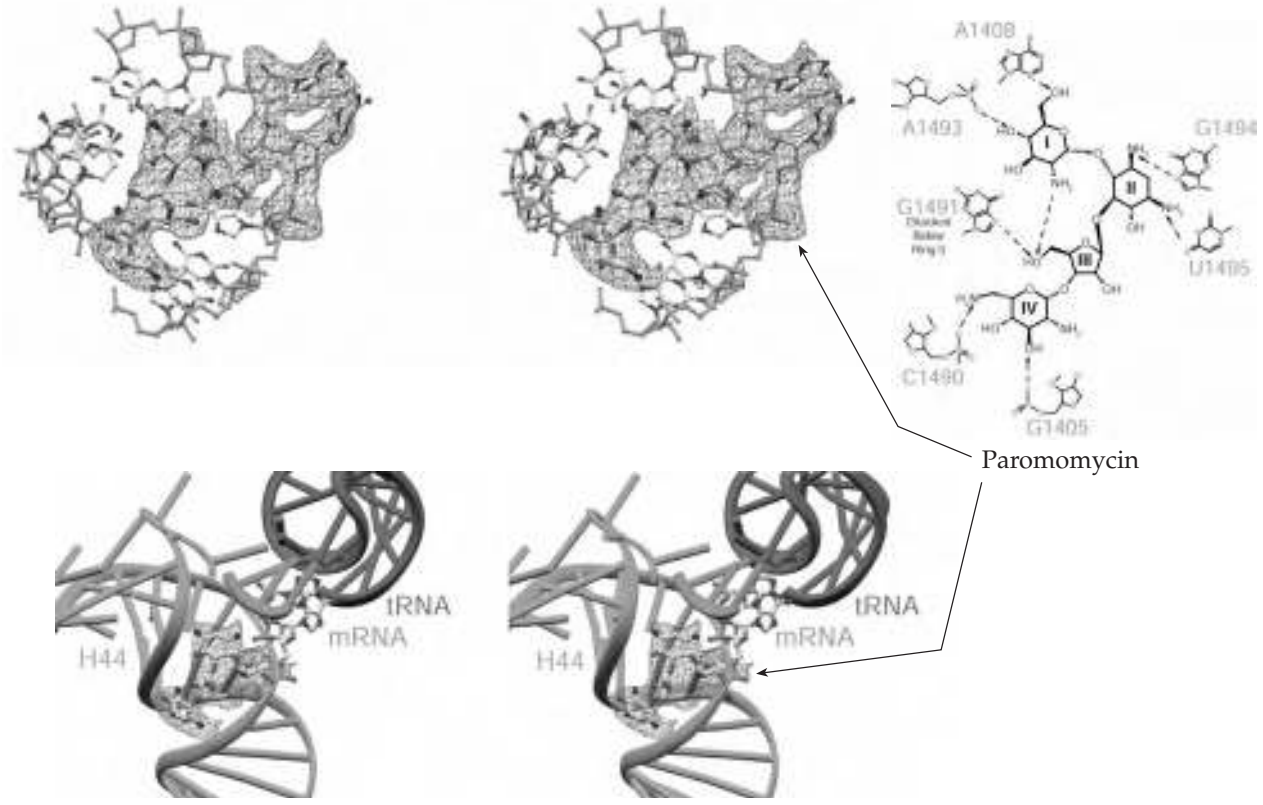
BOX 29-B ANTIBIOTICS THAT INHIBIT PROTEIN SYNTHESIS (continued)

illustrating the difficulty in use of active-site labeling techniques.

A large number of other antibiotics also bind in the vicinity of the peptidyltransferase center (Fig. 29-14).[l,m] Among them are macrolide antibiotics such as erythromycin (Fig. 21-1)[n] and spiramycin,[o,p] chloramphenicol (Fig. 25-10), griseofulvin (Fig. 21-10), streptogramins,[q] oxazolidinones such as linezolid,[r] sparsomycin,[s] and lincomycin. Erythromycin has been very useful in locating the peptidyltransferase center. For example, 23S RNA mutant G2057A, in which A has replaced the normal G, and mutants G2058G and G2058U are resistant to erythromycin.[l,n,t] Mutant C2518U in *Halobacterium halobium* 23S RNA (C2499 in *E. coli*) is among mutants resistant to sparsomycin.[s] Chloramphenicol not only blocks the peptidyltransferase but also causes an accumulation of the compound ppGpp (p. 1715).

The aminoglycoside antibiotics **streptomycin** (Box 20-B),[c,d,u,v] the neomycins,[w] **paromomycin** (see drawing below),[c,x–z] gentamycin,[aa] and kanamycin have one structural unit in common. They often bind to 16S ribosomal RNA in the decoding center. However, they bind in distinctly different ways. Streptomycin causes ribosomes to misread the genetic code[bb,cc] primarily at the first base of a codon. Thus, when poly(U) serves as a messenger RNA, the expected polyphenylalanine product contains 40% isoleucine. When a bacterial population is subjected to the action of any antibiotic, a few mutants are able to grow and survive in the presence of the antibiotic. Streptomycin-resistant mutants of *E. coli* arise at the very low frequency of ~10^{-12}. One of the genes affected (*rpsL*) was mapped at 72 min. Subsequently, it was shown that streptomycin binds to ribosomal protein S12, and that *rpsL* is the gene for this protein. Mutations in the universally conserved 2660 loop of 23S RNA in the sarcin/ricin domain lead to blockage of the elongation cycle. Bacteria containing both a G2661C mutation in their 23S RNA and also a streptomycin resistance mutation in protein S12 lose efficiency in the action of EF-Tu and die. However, they survive in the presence of streptomycin.[dd]

Streptomycin can also be chemically crosslinked to 16S RNA,[ee] and several aminoglycoside antibiotics including streptomycin and spectinomycin bind



Figures are from Carter *et al.*[c]

BOX 29-B (continued)

to and, in footprinting experiments, protect specific nucleotides in 16S RNA.[ff] Streptomycin binds tightly to the upper part of helix 44 of bacterial 16S RNA (Fig. 29-2). This part of the helix is in the region that binds messenger RNA. It also contains parts of the A and P sites of the decoding region of the 30S ribosomal subunit. The same antibiotic binds less tightly to the 915 region in the center of the 16S RNA.[gg,hh] As mentioned above, some streptomycin-resistant mutants become dependent upon the antibiotic and will not grow in its absence. This streptomycin dependence sometimes results from modification in ribosomal protein S4, but the dependence can be suppressed by specific mutations in S5.[ii] It is clear that a single point mutation altering one amino acid is all that is necessary to enormously change the sensitivity of a living organism to a particular toxin, or even to make the organism dependent upon that toxin. Paromomycin also binds to the upper end of helix 44 in the major groove of the RNA (see figure in this box) and close to the streptomycin-binding site.[c,z,jj] Binding distorts the structure of the bulge loop containing adenosines A1492 and A1493, which are markers for the A site of the decoding region. Messenger RNA is also bound at this site as shown in the accompanying figure. Gentamycin also binds in the A site.[aa]

Hygromycin B also binds at the very top of helix 44 blocking the translocation step in the ribosomal cycle.[kk] **Spectinomycin** binds not only to RNA but also to protein S5, as indicated by analysis of resistant mutants. The S5 structural gene *spcA* maps at 64 min, a position in a ribosomal protein operon of the *E. coli* chromosome. The 16S RNA binding site at one end of helix 34 (with protection of G1064 and C1192) is adjacent to S5 as shown by X-ray structural analysis and directed hydroxyl radical probing (see Fig. 29-2).[c] The antibiotic also interferes with the translocation step of polypeptide elongation. **Kasugamycin** inhibits the binding of fMet-tRNA (initiation). In this case, resistant mutants appear in which it is not a protein subunit that has been modified but the 16S RNA. In resistant strains there is less methylation of adenosines 1518 and 1519 (Fig. 29-2) than in normal strains.[ll]

The **tetracyclines** (Fig. 21-10) inhibit the binding of aminoacyl-tRNA at the A site in the 30S ribosomal subunit.[kk] However, this doesn't appear to be a direct effect. Tetracyclines bind to the 16S RNA at two sites. A major site is on helix 34 near the spectinomycin site in the platform region. A second site is on helix 27, the switch helix, which plays a direct role in translocation (see Eq. 29-9).[kk] Although the basis of the inhibition is not clear, there are distinct differences in binding to the 16S bacterial and 18S eukaryotic RNAs that explain the high specificity of the antibiotic toward bacteria.

Another site of antibiotic action is the GTPase-activating center. This center contains a double hairpin structure in the 23S RNA, which binds to protein L11 and the L10•$(L12)_4$ stalk complex. Several proteins, including initiation factor IF1 and the elongation factors EF-Tu, EF-G (bacteria)/EF-2 (eukaryotes), bind to this part of the 50S ribosome (Fig. 29-4). The thiopeptide antibiotics **thiostrepton**, **micrococcin**, and siomycin[t,mm,nn] also bind in this region. Thiostrepton acts by preventing association with the ribosome of an incoming aminoacyl-tRNA as the EF-Tu•GTP complex. Its binding site is primarily in the 23S RNA, but it probably interferes with peptide elongation by interfering with a conformational change in protein L11.[t] A related cyclic peptide (GE2270A) binds to EF-Tu•GDP competing for binding of an aminoacyl-tRNA and blocking the GDP-GTP exchange.[nn] Certain mutations in the EF-Tu protein confer resistance to this antibiotic.[oo] In a similar way kirromycin prevents release of EF-Tu from the ribosome after GTP hydrolysis.[nn,pp,qq]

The binding site of initiation factor IF1 involves both the 30S and 50S ribosomal subunits. The large oligosaccharide antibiotic **evenimicin** protects a specific set of nucleotides in two loops near the peptidyltransferase center (Fig. 29-14).[rr] Erythromycin,[b] other macrolide antibiotics, cycloheximide (Fig. 21-10), and fusidic acid (p. 1266) all prevent translocation by stabilizing the pre-translocation complex.[ss,tt] Fusidic acid may bind to EF-G on the ribosome, preventing an essential conformational change in this G protein.[tt] Fusidic acid also inhibits accumulation of ppGpp. Figure 29-14 shows the locations of some mutations in *E. coli* 23S rRNA that confer resistance to erythromycins and chloramphenicol. Notice that both domains II and V are involved.[n] **Pactamycin** binds to helices 23b and 24a, a binding site for initiation factor IF3.[kk]

Many antibiotics, which inhibit protein synthesis, do not bind to ribosomes but block any of a variety of vital chemical processes needed for growth. Among them are **pseudomonic acid**, which inhibits isoleucyl-tRNA synthetase from many gram-positive bacteria.[uu,vv] **Rapamycin**, best known as an immunosuppressant (Box 9-F), inhibits phosphoinositide-3-kinase and also phosphorylation of the cap-binding protein 4G, a component of the eukaryotic initiation factor complex (Fig. 29-11).[ww] The bacterial enzyme peptide deformylase, which is absent from the human body, has been suggested as a target for design of synthetic antibiotics.[xx]

BOX 29-B ANTIBIOTICS THAT INHIBIT PROTEIN SYNTHESIS (continued)

[a] Gorini, L. (1966) *Sci. Am.* **214**(Apr), 102–109
[b] Narayanan, C. S., and Dubnau, D. (1987) *J. Biol. Chem.* **262**, 1766–1771
[c] Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000) *Nature (London)* **407**, 340–348
[d] Schroeder, R., Waldsich, C., and Wank, H. (2000) *EMBO J.* **19**, 1–9
[de] Schlünzen, F., Zarivach, R., Harms, J., Bashan, A., Tocilj, A., Albrecht, R., Yonath, A., and Franceschi, F. (2001) *Nature (London)* **413**, 814–821
[e] Nierhaus, K. H. (1990) *Biochemistry* **29**, 4997–5008
[f] Olson, H. M., Nicholson, A. W., Cooperman, B. S., and Glitz, D. G. (1985) *J. Biol. Chem.* **260**, 10326–10331
[g] Weitzmann, C. J., and Cooperman, B. S. (1990) *Biochemistry* **29**, 3458–3465
[h] Bischof, O., Kruft, V., and Wittmann-Liebold, B. (1994) *J. Biol. Chem.* **269**, 18315–18319
[i] Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) *Science* **289**, 920–930
[j] Welch, M., Chastang, J., and Yarus, M. (1995) *Biochemistry* **34**, 385–390
[k] Lührmann, R., Bald, R., Stöffler-Meilicke, M., and Stöffler, G. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7276–7280
[l] Garrett, R. (1983) *Trends Biochem. Sci.* **8**, 189–190
[m] Rodriguez-Fonseca, C., Amils, R., and Garrett, R. A. (1995) *J. Mol. Biol.* **247**, 224–235
[n] Douthwaite, S., and Aagaard, C. (1993) *J. Mol. Biol.* **232**, 725–731
[o] Bischof, O., Urlaub, H., Kruft, V., and Wittmann-Liebold, B. (1995) *J. Biol. Chem.* **270**, 23060–23064
[p] Poulsen, S. M., Kofoed, C., and Vester, B. (2000) *J. Mol. Biol.* **304**, 471–481
[q] Porse, B. T., and Garrett, R. A. (1999) *J. Mol. Biol.* **286**, 375–387
[r] Kloss, P., Xiong, L., Shinabarger, D. L., and Mankin, A. S. (1999) *J. Mol. Biol.* **294**, 93–101
[s] Tan, G. T., DeBlasio, A., and Mankin, A. S. (1996) *J. Mol. Biol.* **261**, 222–230
[t] Porse, B. T., Leviev, I., Mankin, A. S., and Garrett, R. A. (1998) *J. Mol. Biol.* **276**, 391–404
[u] Davies, C., Bussiere, D. E., Golden, B. L., Porter, S. J., Ramakrishnan, V., and White, S. W. (1998) *J. Mol. Biol.* **279**, 873–888
[v] Recht, M. I., Douthwaite, S., and Puglisi, J. D. (1999) *EMBO J.* **18**, 3133–3138
[w] Fourmy, D., Recht, M. I., and Puglisi, J. D. (1998) *J. Mol. Biol.* **277**, 347–362
[x] Fourmy, D., Recht, M. I., Blanchard, S. C., and Puglisi, J. D. (1996) *Science* **274**, 1367–1371
[y] Recht, M. I., Douthwaite, S., Dahlquist, K. D., and Puglisi, J. D. (1999) *J. Mol. Biol.* **286**, 33–43
[z] Lynch, S. R., and Puglisi, J. D. (2001) *J. Mol. Biol.* **306**, 1037–1058
[aa] Yoshizawa, S., Fourmy, D., and Puglisi, J. D. (1998) *EMBO J.* **17**, 6437–6448
[bb] Browning, K. S., Maia, D. M., Lax, S. R., and Ravel, J. M. (1987) *J. Biol. Chem.* **262**, 538–541
[cc] Tai, P.-C., Wallace, B. J., and Davis, B. D. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 275–279
[dd] Bilgin, N., and Ehrenberg, M. (1994) *J. Mol. Biol.* **235**, 813–824
[ee] Gravel, M., Melancon, P., and Brakier-Gingras, L. (1987) *Biochemistry* **26**, 6227–6232
[ff] Moazed, D., and Noller, H. F. (1987) *Nature (London)* **327**, 389–394
[gg] Pinard, R., Payant, C., Melancon, P., and Brakier-Gingras, L. (1993) *FASEB J.* **7**, 173–176
[hh] Spickler, C., Brunelle, M.-N., and Brakier-Gingras, L. (1997) *J. Mol. Biol.* **273**, 586–599
[ii] Culver, G. M., Heilek, G. M., and Noller, H. F. (1999) *J. Mol. Biol.* **286**, 355–364
[jj] VanLoock, M. S., Easterwood, T. R., and Harvey, S. C. (1999) *J. Mol. Biol.* **285**, 2069–2078
[kk] Brodersen, D. E., Clemons, W. M., Jr., Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000) *Cell* **103**, 1143–1154
[ll] Vila-Sanjurjo, A., Squires, C. L., and Dahlberg, A. E. (1999) *J. Mol. Biol.* **293**, 1–8
[mm] Porse, B. T., Cundliffe, E., and Garrett, R. A. (1999) *J. Mol. Biol.* **287**, 33–45
[nn] Heffron, S. E., and Jurnak, F. (2000) *Biochemistry* **39**, 37–45
[oo] Zuurmond, A.-M., de Graff, J. M., Olsthoorn-Tieleman, L. N., van Duyl, B. Y., Mörhle, V. G., Jurnak, F., Mesters, J. R., Hilgenfeld, R., and Kraal, B. (2000) *J. Mol. Biol.* **304**, 995–1005
[pp] Mesters, J. R., Zeef, L. A. H., Hilgenfeld, R., de Graaf, J. M., Kraal, B., and Bosch, L. (1994) *EMBO J.* **13**, 4877–4885
[qq] Alexander, C., Bilgin, N., Lindschau, C., Mesters, J. R., Kraal, B., Hilgenfeld, R., Erdmann, V. A., and Lippmann, C. (1995) *J. Biol. Chem.* **270**, 14541–14547
[rr] Belova, L., Tenson, T., Xiong, L., McNicholas, P. M., and Mankin, A. S. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3726–3731
[ss] Johanson, U., AEvarsson, A., Liljas, A., and Hughes, D. (1996) *J. Mol. Biol.* **258**, 420–432
[tt] Laurberg, M., Kristensen, O., Martemyanov, K., Gudkov, A. T., Nagaev, I., Hughes, D., and Liljas, A. (2000) *J. Mol. Biol.* **303**, 593–603
[uu] Yanagisawa, T., Lee, J. T., Wu, H. C., and Kawakami, M. (1994) *J. Biol. Chem.* **269**, 24304–24309
[vv] Sassanfar, M., Kranz, J. E., Gallant, P., Schimmel, P., and Shiba, K. (1996) *Biochemistry* **35**, 9995–10003
[ww] Raught, B., Gingras, A.-C., Gygi, S. P., Imataka, H., Morino, S., Gradi, A., Aebersold, R., and Sonenberg, N. (2000) *EMBO J.* **19**, 434–444
[xx] Hao, B., Gong, W., Rajagopalan, P. T. R., Zhou, Y., Pei, D., and Chan, M. K. (1999) *Biochemistry* **38**, 4712–4719

form the acceptor stem and the TψC arm form a single **acceptor-TψC-minihelix**, while the other two domains fold together to create an **anticodon-dihydrouridine** stem loop.[172] New tertiary interactions, some of which are indicated in Fig. 29-7C, are formed. Mitochondrial tRNAs of metazoa often lack some elements of the cloverleaf. An extreme example is the bovine $mtRNA^{Ser}$, which recognizes AGY codons and completely lacks the dihydrouridine loop. This fact suggests that the L shape of tRNAs cannot be completely invariant.[174,176] As is shown in Fig. 29-7B, divalent metal ions such as Mg^{2+} are bound at discrete sites in tRNA molecules.[171] The tertiary interactions in the "core" of the L form contains several stacked layers of base pairs and triplets (Fig. 29-7C). The top layer is usually the single base 59; below it in succession are the 15:48 pair (see Fig. 29-7), the 21:8:14 triplet, the 13:22 pair (see also p. 231), and then base pairs present in the dihydroU loop. Considerable variation is observed among the different tRNAs.[182,183] The

structural features of this core may also be utilized for recognition by aminoacyl tRNA synthetases.

Initiator tRNAs. While the TψC sequence has been found in all bacterial and most eukaryotic tRNAs examined, it is replaced by UCG in eukaryotic initiator tRNAs. In these tRNAs the preceding two nucleotides, beginning in the stem of loop IV, are also conserved; the complete conserved sequence being GAUCG.[184] Other characteristics of initiator tRNAs are the absence of base-pairing between residues 1, and 72, and the presence of C rather than G at position 1, A rather than G at position 72, and CCU in place of the two dihydroU residues in loop I.[185] Initiator tRNAs of chloroplasts resemble those of bacteria,[186] whereas archaeobacteria have their own unique peculiarities.[187] These include the presence of a hypermodified base known as **archaeosine** (p. 1456) in position 15 of the dihydroU loop.[188,189]

2. Pairing of Codon and Anticodon

Accurate protein synthesis depends upon both correct charging of the tRNAs and correct recognition by an anticodon in the tRNA of the complementary codon in the mRNA. A surprise was the discovery of inosine (I) in anticodons of yeast tRNA (but not in most *E. coli* tRNAs). Another unexpected finding was that fewer than 61 kinds of tRNA exist in a given cell (61 = 64 codons minus three stop codons). Consideration of these matters led Crick, in 1966, to propose the **wobble hypothesis**.[190] According to this proposal the first two bases at the 5' end of the codon (and at the 3' end of the anticodon) must pair in the same ways as do the bases in DNA. However, the third base pair (3' end of the codon and 5' end of the anticodon) is under a less severe steric restriction. That is, there may be some "wobble." Crick suggested the accompanying rule for pairing of the third base. All of the observed

5′-Base in anticodon	Paired 3′-base in codon
G	C or U
C	G
A	U
U	A or G
I	C, A, or U

deviations from the AU, CG pairing of a Watson–Crick helix can be explained in this way. An anticodon with G at the 5' end can pair with codons with either C or U at their 3′ end. Anticodons with C or A at the 5′ end

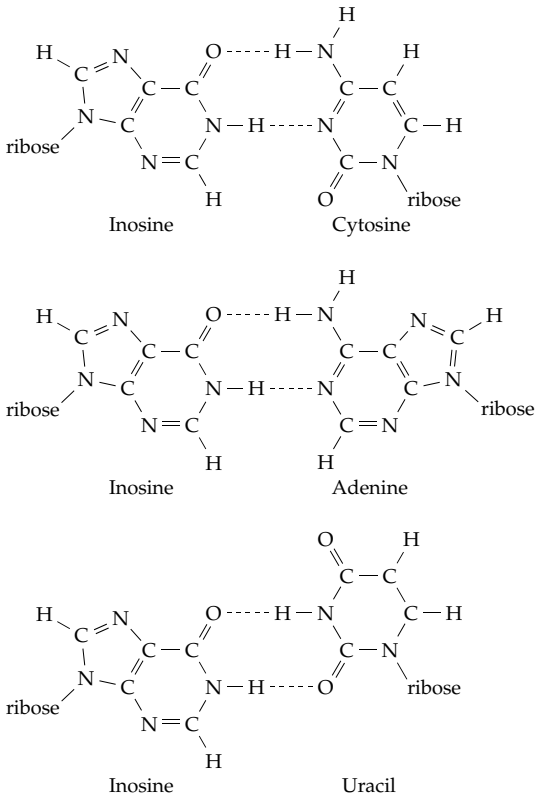


Figure 29-8 Pairing of inosine with cytosine (a Watson–Crick pair) and of inosine with adenine and uracil (wobble pairs).

pair strictly. Anticodons ending with U at the 5′ end can pair with codons containing either A or G in the 3' position. Anticodons with I in the 5' position can recognize codons with any of the three bases in the third position. Comparison with Table 5-5 makes it immediately clear why fewer than 61 anticodons are needed. Many codons represent the same amino acid, and frequently the nature of the base in the 3' position of the codon is immaterial to the meaning of that codon. Thus, there is an economy in using less than the full array of anticodons. Crick showed that his proposal was chemically feasible if the spatial relationships for the wobble pair were allowed to vary from the usual ones in Watson–Crick base pairs. This is illustrated in Fig. 29-8 for binding of inosine to C (a normal Watson–Crick base pair) and to A and to U. Although the word wobble does not convey an exact meaning, the hypothesis has predicted many things correctly. For example, only three tRNAs are required to recognize the six serine codons. In fact, only three are found in *E. coli*.

The matter is made more complex by the fact that U34 in the first (5') anticodon position of tRNA is usually modified postranscriptionally.[190a] For

example, it is usually converted to the 2-thio-5-CH_3 or 2-thio-5-CH_2-NH_2^+-CH_3 derivative in anticodons recognizing A in the 3' position of a codon. For recognition of other bases the 5–OH, $-OCH_3$, or $-OCH_2COO^-$ derivative is usually present. Yokoyama *et al*. attributed the selectivity to the stabilization of the C2' endo form of the ring in the former group.[191]

3. Aminoacylation of tRNAs

Discrimination between some pairs of tRNAs depends entirely on the anticodon sequence. For example, $tRNA^{Met}$ contains the anticodon CAU. That for a minor $tRNA^{Ile}$ is the same except that the cytosine has been posttranscriptionally modified by covalent linkage of a molecule of lysine via its ε-amino group to C2 of the cytosine. The latter base (**lysidine**) is correctly recognized by *E. coli* isoleucyl-tRNA synthetase; but, if the cytosine is unmodified, it is aminoacylated by methionyl-tRNA synthetase.[192] In most instances the acceptor specificity, or **tRNA identity**, is not determined solely by the anticodon sequence. Thus, when a methionine initiator tRNA was modified to contain a tryptophan anticodon, it was only partially charged with tryptophan *in vivo*. However, when A73 of the methionine tRNA was also converted to G73, only tryptophan was inserted.[193] Nucleotide 73 (Fig. 29-7) is sometimes called the **discriminator nucleotide**.[194–196] It is A in methionine and leucine tRNAs,[197] G in tryptophan tRNAs, and C in histidyl RNAs.[198] The tRNA features needed to establish its identity are sometimes referred to as its **identity-determinant set**.[196,198a,b] This includes the anticodon and other features needed for recognition by the aminoacyl-tRNA synthetases that "charge" the tRNAs with aminoacyl groups.[199–204] For example, for *E. coli* $tRNA^{Val}$ the recognition determinants are A35 and C36 of the anticodon, A73, G20, G45, and a regular A-RNA acceptor helix.[205] All known mature tRNAs contain a 3'-CCA end on which the aminoacylation occurs. Nevertheless, alterations in this sequence still allow correct aminoacylation of some tRNAs.[206]

The aminoacyl-tRNA synthetases (amino acid: tRNA ligases) join amino acids to their appropriate transfer RNA molecules for protein synthesis. They have the very important task of selecting both a specific amino acid and a specific tRNA and joining them according to Eq. 29-1.[175,195,207] These reactions represent the first step in the decoding of mRNAs. Organisms usually contain one aminoacyl-tRNA synthetase for each of the 20 amino acids. Each synthetase must select a specific amino acid and a correct tRNA for that amino acid. The same enzyme transfers an activated amino acid to all of the **isoacceptor tRNAs** specific for a given amino acid. Some aminoacyl-tRNA synthetases attach the aminoacyl group to the 2'–OH of the tRNA substrate, some to the 3'–OH. The chemical mechanism is the same in both cases.

Structures. Aminoacyl-tRNA synthetases vary in size, the subunit masses ranging from 37- to 110-kDa (329–951 residues). There are monomeric species, dimers, tetramers, and $(\alpha\beta)_2$ mixed tetramers.[207,208] Sequence comparisons, together with X-ray structural investigations, have shown that the enzymes can be classified into two groups, each containing ten enzymes.[207,209,209–209b] **Class I aminoacyl-tRNA synthetases** share two consensus sequence motifs: HIGH and KMSKS. Their ATP-binding active sites are in a Rossman fold nucleotide-binding domain (Fig. 2-13).[210] The KMSKS sequence parallels the Walker sequence found in various nucleotide-binding proteins including ATP synthase.[211–213] The actual sequences vary considerably, e.g., the KMSKS sequence for a tyrosyl-tRNA synthetase is actually KFGKT.[211]

Class II aminoacyl-tRNA synthetases contain a different set of three "signature sequences," two of which form an ATP-binding catalytic domain. The active site structure is built on an antiparallel β sheet and is surrounded by two helices (Fig. 29-9). Each class contains subgroups with inserted loops that form other domains. In the following tabulation the reference numbers refer to three-dimensional structural studies.

Class I
Glu,[209] Gln,[218–220] Arg,[221]
Tyr,[222] Trp,[212]
Ile,[223] Leu,[224] Val, Cys, Met[210,225,226]

Class II
His,[227–229] Pro, Ser,[230,231] Thr
Asp,[232,233] Asn,[234]Lys,[217,235]
Phe,[236] Ala, Gly,[204,237]

The 37-kDa 334-residue subunits of the dimeric type I tryptophanyl-tRNA synthetase[238] are the smallest known; the largest bacterial synthetase is an alanine-specific type II tetramer with 95-kDa 875-residue subunits.[239] Gene deletions show that a much smaller core, comparable in size to that of the tryptophanyl-tRNA synthetase, is needed for amino acid activation. The synthetases share little sequence homology except for a short 11-residue part of the adenylate binding site near the N terminus.[240,241] Some of the synthetases contain bound zinc ions.[225,242]

Recognition of cognate tRNAs. Many attempts have been made to learn what part or parts of tRNA molecules are involved in recognition by aminoacyl-tRNA synthetases. Nucleotide sequences of isoacceptor tRNAs have been compared. Chemically modified and fragmented tRNA molecules have been studied, and many mutant tRNAs have been made. These

have often been mutants of suppressor tRNAs that place specific amino acids such as phenylalanine or alanine into a peptide at a termination codon, often the termination codon UAG (see Section C4). An alternative approach is to synthesize DNA templates, which can be transcribed *in vitro* by phage T7 RNA polymerase to give mutant tRNAs,[243] or to make such tRNAs by solid-phase chemical synthesis.[244] Although these contain no modified bases, they serve as substrates for the aminoacyl-tRNA synthetases.

The results of these efforts show that no method of tRNA recognition is universal.[244a] In some cases, e.g., for methionine- or valine-specific tRNAs, the synthetase does not aminoacylate a modified tRNA if the anticodon structure is incorrect. Although the anticodon is 7.5 nm away from the CCA end of the tRNA, the synthetases are large enzymes. Many of them are able to accommodate this large distance between a recognition site and the active site (Fig. 29-9A). For some other tRNAs the anticodon is not involved in recognition.[245] For yeast $tRNA^{Phe}$ residues in the stem of the dihydrouridine loop and at the upper end of the amino acid acceptor stem seem to be critical.[241]

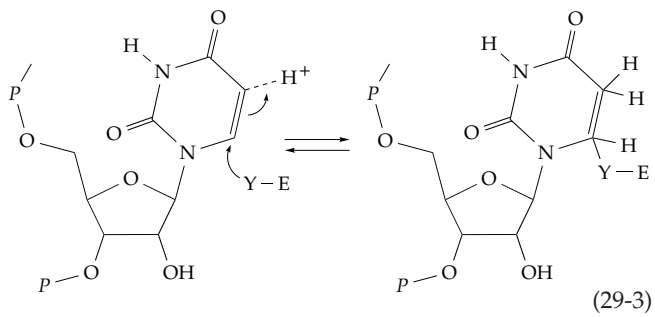
For some other tRNAs only the acceptor helix is essential for recognition. Change of one base-pair, the pair G3 • U70 (a "wobble" pair) of an *E. coli* $tRNA^{Ala}$ • mRNA complex to the unnatural A3 • U70, prevents aminoacylation. Conversely, a G3 • U70 pair formed with tRNAs specific for other amino acids causes them to become substrates for the alanyl-tRNA synthetase.[241,246] Even a shortened tRNA minihelix consisting of a 7-bp acceptor stem, 6-nucleotide loop, and ACCA 3' end is a substrate for this enzyme.[247,247a] A seryl-tRNA synthetase depends upon recognition of two base pairs in the acceptor stem.[248] Synthetic DNA oligomers with sequences corresponding to those of *E. coli* $tRNA^{Phe}$ or $tRNA^{Lys}$, and with either deoxythymidine or deoxyuridine in the positions occupied by ribouridine in the tRNAs, are also substrates for the synthetases. The affinity and reaction rates are somewhat decreased, but the ribose 2'-OH is not essential for recognition.[249]

Mechanisms of reaction. Activation of an amino acid occurs by a direct in-line nucleophilic displacement by a carboxylate oxygen atom of the amino acid on the α phosphorus atom of MgATP to form the aminoacyl adenylate (Eq. 29-1, step *a*). For yeast phenylalanyl-tRNA synthetases the preferred form of MgATP appears to be the β,γ-bidentate (Λ screw sense) complex (p. 643).[250] This is followed by a second nucleophilic displacement, this one on the C=O group of the aminoacyl adenylate by the –OH group of the tRNA (Eq. 29-1, step *b*; Fig. 29-9C). A conformational change in the protein may be required to permit dissociation of the product, the aminoacyl-tRNA. In the complex of a class I synthetase with aminoacyl adenylate and tRNA the 3' CCA acceptor end of the tRNA is straight, but in a class II synthetase it is bent. In the two classes of synthetase the tRNAs approach the enzyme in a mirror-symmetric fashion. The 2'-OH of the terminal ribose is positioned to attack the carbonyl of the aminoacyl adenylate in class I enzymes, while the 3'-OH is positional for the attack in class II enzymes.[207]

The three-dimensional structure of *E. coli* glutaminyl-tRNA synthetase is shown in Fig. 29-9A. The active site lies against a β sheet in a structure similar to the nucleotide binding domain of a dehydrogenase (Fig. 2-13). The site of binding of a ATP is marked in Fig. 29-9A. The details of this binding to tyrosine-tRNA synthetase have been studied intensively.[251–255] Binding of the tRNA substrates is less well understood. A large series of mutants involving 40 basic residues were prepared by Bedouelle and Winter[256]; study of these mutant enzymes, together with computer-assisted modeling, led to a proposed structure for a transition state for a complex with tRNA as is shown in Fig. 29-9B. Kinetic studies of heterodimers prepared from mutant and normal enzyme confirm that both subunits of the dimeric enzyme interact with the tRNA.

The active site of a type II synthetase is shown in Fig. 29-9C,D.[217,217a] The expected movement of electrons in the reaction with ATP is illustrated by the green arrow in D. Both metal ions and active-site protein groups may participate as is also proposed for another type II enzyme.[229]

Some data suggested that a transient covalent linkage of tRNA to the synthetases may form through addition of a nucleophilic group of the enzyme to the 6' position of the uracil (or 4-thiouracil) present in position 8 of all tRNAs (Eq. 29-3).[257] The two isoacceptors $tRNA^{Tyr}$ species in *E. coli* contain 4-thiouracil at this position. The C=C bond in this base can be saturated by sodium borohydride reduction, which was found not only to prevent the covalent interaction with the enzyme but also to prevent aminoacylation of the tRNA. However, Eq. 29-3 probably describes a side reaction irrelevant to tRNA function.

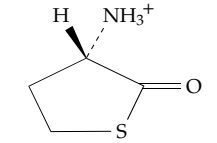


(29-3)

Correcting errors. Much attention has been devoted to "proofreading" or "editing" activities of

amino acid-tRNA synthetases (see p. 482). For the majority of the enzymes proofreading is not needed.[209a] Thus, tyrosyl-tRNA synthetase (Fig. 29-9B) mistakenly chooses phenylalanine instead of tyrosine only 5 in 10^4 times, apparently a tolerable rate of error. This enzyme, as well as a tryptophan-specific synthetase,[238] depends largely upon differences in the Gibbs energy of binding to select the correct substrate. However, the discrimination between valine and isoleucine by isoleucyl-tRNA synthetase poses a more difficult problem. It is apparently solved, in part, by a "double sieve" editing mechanism,[223,224,258,259] which is described briefly on p. 482. In the first sieve competitors that are larger than the substrate or are differently shaped are excluded by steric repulsion from binding in the active site. Isoleucyl-tRNA synthetase doesn't convert leucine into an aminoacyladenylate, but it does act on the smaller valine. However, most of the resulting enzyme-bound valyl-adenylate is hydrolyzed to valine and AMP before it can be transferred to $tRNA^{Ileu}$. It shifts into an editing site, which is too small for the isoleucyl-adenylate, in effect passing through a second sieve.[259a] Some of the activated valine is transferred to tRNA and is removed in a second editing reaction.[223] Misactivation of threonine or some other amino acids by valyl-tRNA synthetase is corrected in an analogous fashion.[259b,c]

A similar editing process prevents isoleucyl-, leucyl-, and methionyl-tRNA synthetases from attaching L-homocysteine to tRNAs.[260–263] In this case, instead of hydrolysis the editing site catalyzes conversion of the homocysteinyl-adenylate into homocysteine lactone. Naturally occurring mutations in tRNA molecules can sometimes have serious consequences. For example, a human mutation is responsible for a fragile mitochondrial isoleucine tRNA and serious cardiomyopathy and opthalmophegia (see also Box 18-B).[263a]



L-Homocysteine lactone

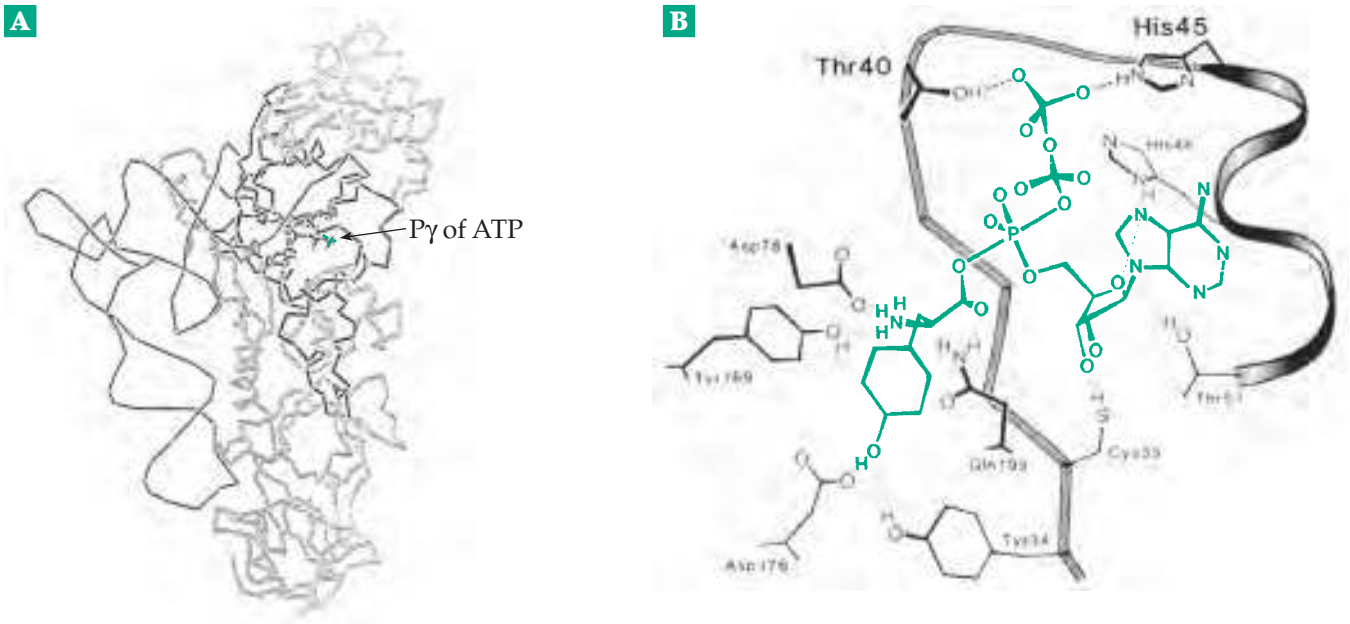


Figure 29-9 Selected views of aminoacyl-tRNA synthetase structure and action. (A) Alpha-carbon trace of the type I *E. coli* glutaminyl-tRNA synthetase. The phosphate backbone of $tRNA^{Gln}$ is shown in black; ATP is shown in the active-site cleft. The canonical dinucleotide fold domain near the N terminus is shaded. Two structural motifs (black), proposed to link the active site with regions of the protein–RNA interface involved in tRNA discrimination, are indicated. The α helix (top) connects tRNA recognition in the minor groove of the acceptor stem with binding of the ribose group of ATP. The large loop (center) connects anticodon recognition by the two β-barrel domains (bottom) with sequences flanking the MSK sequence motif, which interacts with the phosphates of ATP. From Perona *et al.*[214] Courtesy of Thomas A. Steitz. (B) The active site structure of tyrosyl-tRNA synthetase from *Bacillus stearothermophilus* with a penta-coordinate transition state structure modeled.[215] From Leatherbarrow *et al.*[216] (C) Schematic representation of the active site of the lysyl-tRNA synthetase showing potential hydrogen bonding interaction in the ternary complex with lysine and ATP. The invariant motif 2 Arg 262 plays a key role in the recognition of the lysine carboxylate and the ATP α phosphate, while the invariat motif 2 Arg 480 binds the

Proofreading involves kinetic as well as thermodynamic considerations.[264–266] The relative rates of a hydrolytic reaction and the competing activating reaction must always be considered. These ratios can be strongly affected by conformational changes, which may occur in several steps (see also Section C,2).

Essential modification reactions of aminoacyl-tRNAs. In bacteria the initiator tRNA needed to start the synthesis of a polypeptide is initially aminoacylated by methionine, but the methionyl-tRNA^{fMet} must then be *N*-formylated by transfer of a formyl group from N^{10}-formyltetrahydrofolate (Fig. 15-18; Eq. 29-4).[267,268] In gram-positive bacteria and in archaea, mitochondria, and chloroplasts the glutamine-specific tRNA^{Gln} is charged with glutamate to form Glu-tRNA^{Gln}. The latter is converted by action of an ATP-dependent amidotransferase (see Eq. 24-22) to the necessary Gln-tRNA^{Gln} (Eq. 29-5).[268a] In a similar way, RNA^{Asn} in some organisms is charged with aspartate, then converted by transamidation to Asp-tRNA^{Asn} (Eq. 29-6).[207,267,269] An important reaction, that occurs in all kingdoms of life, is the charging of the special tRNA^{Sec} with serine and conversion of the product into selenocysteinyl-tRNA^{Sec} (p. 827; Eq 29-7).[267,270]

$$\text{tRNA}^{fMet} \xrightarrow[\text{ATP}]{\text{Met}} \text{Met-tRNA}^{fMet} \xrightarrow[\text{THF}]{\text{Methyl-THF}} \text{fMet-tRNA}^{fMet} \quad (29\text{-}4)$$

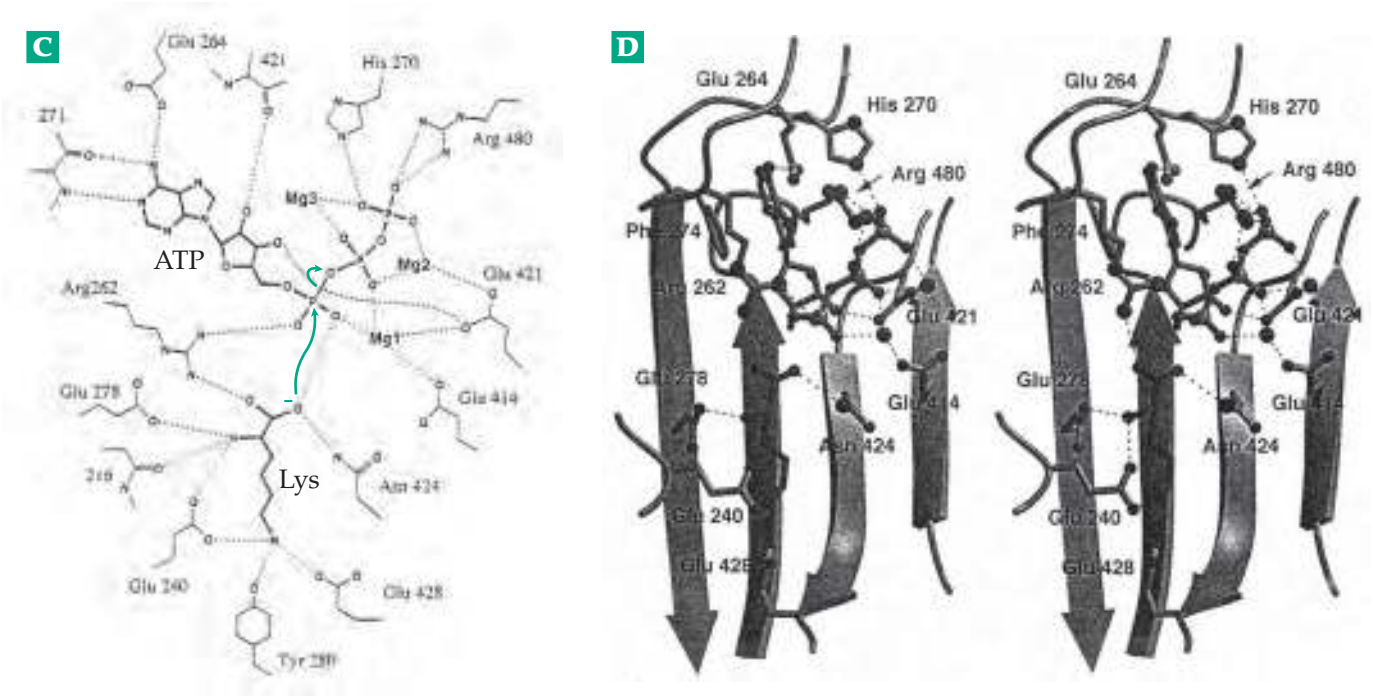
$$\text{tRNA}^{Gln} \xrightarrow[\text{ATP}]{\text{Glu}} \text{Glu-tRNA}^{Gln} \xrightarrow[\text{ATP}]{\text{Gln}} \text{Gln-tRNA}^{Gln} \quad (29\text{-}5)$$

$$\text{tRNA}^{Asn} \xrightarrow[\text{ATP}]{\text{Asp}} \text{Asp-tRNA}^{Asn} \xrightarrow[\text{ATP}]{\text{Gln}} \text{Asn-tRNA}^{Asn} \quad (29\text{-}6)$$

$$\text{tRNA}^{Sec} \xrightarrow[\text{ATP}]{\text{Ser}} \text{Ser-tRNA}^{Sec} \xrightarrow[\text{(p. 827)}]{\text{Se-}P} \text{Selenocysteinyl-tRNA}^{Sec} \quad (29\text{-}7)$$

The introduction of selenocysteine into proteins at selected stop codons using this tRNA is described in Section C,5.

Examination of the complete genome sequences of methanogens revealed an apparent lack of cysteinyl-tRNA synthetase. However, prolyl-tRNA synthetase does correctly aminoacylate the tRNAs for both proline and cysteine in these archaeobacteria.[271–272a]



γ phosphate of the ATP. A number of conserved residues in the motif 2 loop (residues 264–271) assume an ordered conformation only upon ATP binding. The positions of the Mg^{2+} sites are indicated. (D) View of the active site of the type II lysyl-tRNA synthetase showing the conformations of the substrates lysine and ATP before the first step of the reaction takes place. The ATP molecule is located on one side of the central β sheet of the C-terminal domain, with the adenine ring sandwiched between a conserved phenylalanine (Phe 274) and the motif 3 arginine residue (Arg 480). The pyrophosphate moiety is bent toward the adenine placing the α phosphate in the correct position for nucleophilic attack of the lysine carboxylate oxygen. The hydrogen bonding and electrostatic interactions between the substrates and some of the key residues, including the invariant motif 2 arginine (Arg 262), are shown. The three Mg^{2+} ions (green) involved in catalysis are included. (C) and (D) are from Desogus *et al.*[217]

Additional functions of aminoacyl-tRNA synthetases. The primary function of these enzymes in protein synthesis is well known, but they have a whole range of other activities.[273,274] In *E. coli* the large alanyl-tRNA synthetase can repress transcription of its own gene by binding to a palindromic sequence in the control region of the gene.[275] Expression of some genes, such as that for threonyl-tRNA synthetase, is regulated at the translational level.[273,274,276] In mammalian cells the formation of the threonine-specific synthetase appears to be regulated by a phosphorylation–dephosphorylation mechanism.[277] Other synthetases participate in mitochondrial RNA splicing[278] and in aminoacylation of tRNA-like 3' ends of viral genomes (see Fig. 28-24) and of N termini of certain proteins.[279] For example, an arginyl group may be transferred onto the N terminus of a protein, marking it for rapid degradation.[280] Under conditions of apoptosis, tyrosyl tRNA synthetase is hydrolytically cleaved to form two different cytokines.[274] Phenylalanyl-tRNA synthetase is a DNA-binding protein.[281] Within the nucleus newly synthesized tRNAs are checked before being exported to the cytoplasm. Only tRNAs with mature 5' and 3' ends are exported. In both *Xenopus* oocytes[281a] and in *S. cere- visiae*[281b] the tRNAs are also tested prior to export, using aminoacyl-tRNA synthetases, to ensure that they are functional.

Many proteins have structures related to those of aminoacyl-tRNA synthetases.[282,283] For example, asparagine synthetase A functions via an aspartyl-adenylate intermediate (Chapter 24, Section B), and its structure resembls that of aspartyl-tRNA synthetase.[284] The *his G* gene of histidine biosynthesis (Fig. 25-13) encodes an ATP phosphoribosyltransferase with structural homology to the catalytic domain of histidyl-tRNA synthetase.[284] The reason is not clear, but some aminoacyl-tRNA synthetases, especially the histidyl-tRNA synthetase, are common autoantigens for the inflammatory disease **polymyositis**.[285,286]

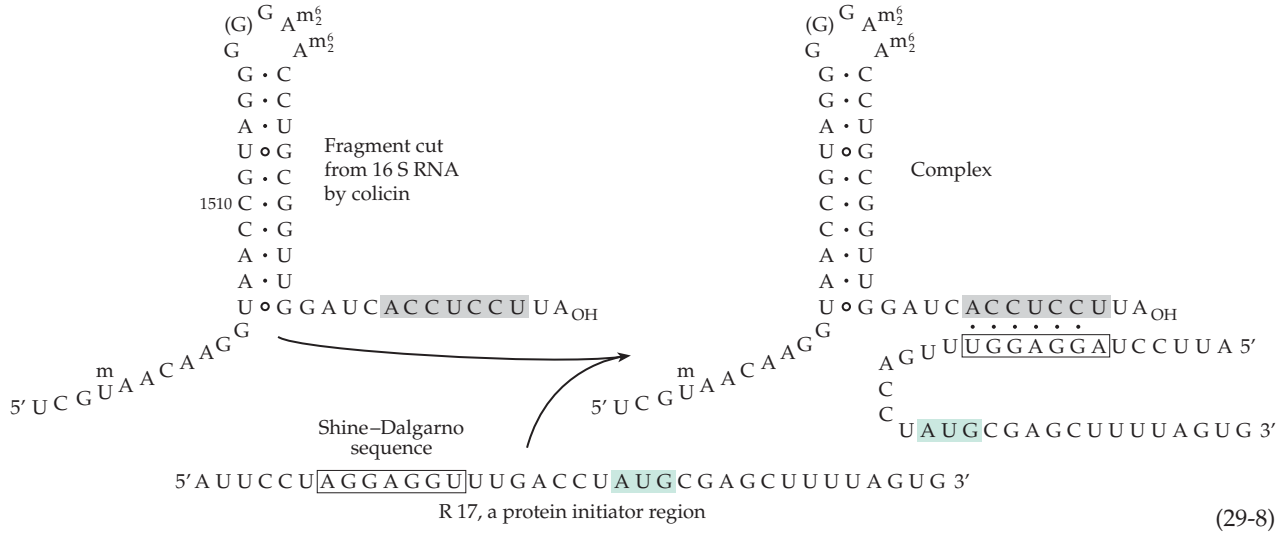
C. Protein Synthesis: The Ribosome Cycle

Initiation (Figs. 29-10 and 29-11), elongation (Fig. 29-12), and termination are three distinct steps in the synthesis of a protein. A variety of specialized proteins are required for each stage of synthesis. Their sequential interaction with ribosomes can be viewed as a means of ensuring an orderly sequence of steps in the synthesis cycle. The rate of protein formation will depend upon the concentrations of amino acids, tRNAs, protein factors, numbers of ribosomes, and kinetic constants. The formation of specific proteins can also be inhibited by **translational repressors**, proteins that compete with ribosomes for binding to target mRNAs.[287]

1. Initiation

For most polypeptide chains initiation begins with one of the three **initiation codons**, most commonly the methionine codon AUG. When properly placed in an mRNA chain, GUG may also serve as a bacterial initiation codon. In such cases, it codes for methionine rather than for valine. Occasionally UUG, AUU, ACG, and perhaps other codons can initiate translation.[288,289] This is less frequent in eukaryotes than in bacteria. The sequence of bases preceding the initiation codon must also be important for recognition of the "start" signal.

In *E. coli* polypeptide chains are always initiated with the amino acid ***N*-formylmethionine**. Some bacteria can apparently live without the ability to formylate methionyl-tRNA,[290] but most eubacteria as well as mitochondria and chloroplasts use formylmethionine for initiation. In a few cases, both among bacteria and eukaryotes, initiation can sometimes occur with other amino acids.[291] The first step is the alignment of the proper initiation codon correctly on



(29-8)

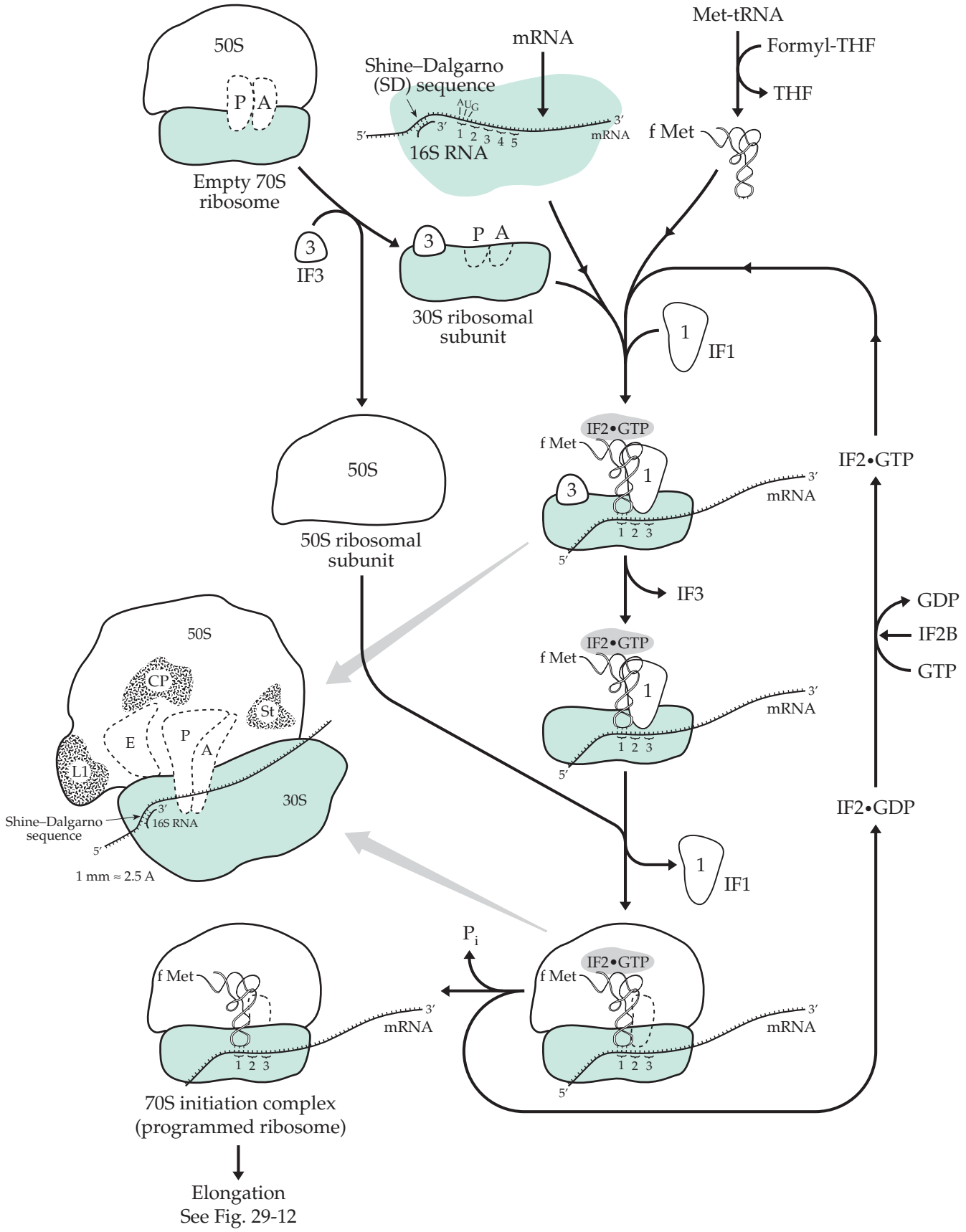


Figure 29-10 Initiation of protein synthesis on bacterial ribosomes. Images are not drawn to scale. Some details are indicated on the larger scale image at the left.

the ribosome and the binding to it of a molecule of initiator tRNA carrying *N*-formylmethionine.[268,292,293] The process by which this occurs is relatively complex, partly because it is essential for the ribosomes to distinguish the true initiation codon from the many AUG codons in internal positions in the message. In bacteria recognition of the initiation codon is assisted by base pairing between the conserved sequence ACCUCCU at positions 1534–1540 at the 3' end of the 16S RNA (Fig. 29-2A) and the complementary **Shine–Dalgarno sequence** AGGAGGU, which is found near the 5' end of most mRNA molecules.[53] This is illustrated in Eq. 29-8 for a messenger RNA in the form of the A protein initiator region from the R17 phage RNA.[294,295] Ribosomal protein S1 also seems to be required for this binding.[135,295a]

Prokaryotic initiation factors. In addition to the ribosomal proteins, the initiation factors **IF1**, **IF2**, and **IF3**, whose molecular masses are 9.5, 9.7, and 19.7 kDa, respectively,[70,296] are essential. They coordinate a sequence of reactions that begins with the dissociation of 70S ribosomes into their 30S and 50S subunits. Then, as is shown in Fig. 29-10, the mRNA, the initiator tRNA charged with formylmethionine, the three initiation factors, and the ribosomal subunits react to form 70S **programmed ribosomes**, which carry the bound mRNA and are ready to initiate protein synthesis. IF2 is a specialized G protein (Chapter 11), which binds and hydrolyzes GTP. It resembles the better known elongation factor EF-Tu (Section 2). The ~172-residue IF3 consists of two compact α/β domains linked by a flexible sequence, which may exist as an α helix.[296a–298] Its C-terminal domain binds to the central domain of the 16S RNA near nucleotides 819–859 (Fig. 29-2). When bound it protects nucleotides in the 690 loop from chemical modification[297] and induces a conformational change in the loop.[297a] Binding of IF3 prevents association of the 30S and 50S subunits, assuring the cell of a supply of free 30S subunits for translational initiation. It also promotes the binding of the other two factors: IF1 and IF2.[299] Binding of IF2, as its GTP complex, stimulates the binding of fMet-tRNA in the adjacent P site.[300] Another function of IF2, in cooperation with IF1, may be to remove peptidyl-tRNAs with short polypeptide chains under conditions in which such peptidyl-tRNAs accumulate to abnormal levels.[301] However, the order of binding, which is implied in Fig. 29-10, has been hard to establish.

IF1, which is essential to the viability of bacteria, binds and partially occludes the A site of the ribosome, preventing the initiator fMet-tRNA from incorrectly occupying the 30S A site.[70,296a,302] Binding of IF1 also causes the functionally important bases A1492 and A1493 of 16S RNA (Fig. 29-2) to be flipped out of helix 44 and to bind to pockets in IF1. This induces further long-range conformational changes.[70] It has also been hard to establish whether the charged initiator tRNA binds into the P site before or after the mRNA binds to the 30S subunit. Some evidence supports the latter possibility,[296] which is indicated in Fig. 29-10. In any case, an important step is the specific base-pairing of the initiator tRNA with the first AUG start codon. IF3 seems to be essential for this pairing process, which establishes the correct reading frame for translation of the genetic message.[303] A proofreading function at this step is often attributed to IF3.[304] Intact ribosomes bind charged tRNAs tightly in the P site. Perhaps the initial binding to the 30S subunit is loose enough to allow the mRNA, which ties itself to the 3' end of the 16S RNA via Eq. 29-3, to move back and forth until the correct reading frame is located. Then a conformational change occurs and locks the initiator tRNA in place. This change also weakens the binding of IF3, which dissociates from the complex, allowing the 50S subunits to rejoin the complex. The ribosome-binding domain of IF3 is homologous to spliceosome protein U1A (Chapter 28).[304]

The hydrolysis of GTP during initiation is essential as is shown by the fact that 5'-guanylmethylene diphosphonate, a GTP analog containing a methylene bridge between the terminal and central phosphorus atoms (see p. 558), can substitute for GTP in all steps up to and including the binding of the 50S ribosome. However, it cannot function in the final step because it cannot undergo hydrolysis. Why is GTP hydrolysis needed? It may provide energy for the conformational rearrangement of ribosomal components, or it may simply be required for release of the IF2•GDP complex. For example, IF2•GTP may bind to the ribosome with a high affinity, but IF2•GDP only weakly. Remember that G-proteins exist in at least two conformations, one stabilized by GDP and another by GTP (Chapter 11). When the hydrolysis of the bound GTP is incorporated into a reaction sequence, it provides a Gibbs energy change that may be needed to drive the reactions. In this case, it ensures that the charged initiator tRNA is firmly bound and ready to initiate translation.

Some information about spatial arrangements of the ribosomal proteins involved in initiation was provided by the fact that antibodies against proteins S19 and S21 block the formation of a complex with fMet-tRNA, while antibodies against S2, S18, and S20 block the binding of IF3. Crosslinking experiments showed that IF2 and S19 are close together and that IF3 is close to S12 (Fig. 29-1A).

Initiation of protein synthesis in eukaryotes. Most eukaryotic mRNAs have a 5' cap (p. 1642) and lack a Shine–Dalgarno sequence. Otherwise, initiation follows a pattern similar to that in bacteria but more complex.[305–308] There are at least ten eukaryotic

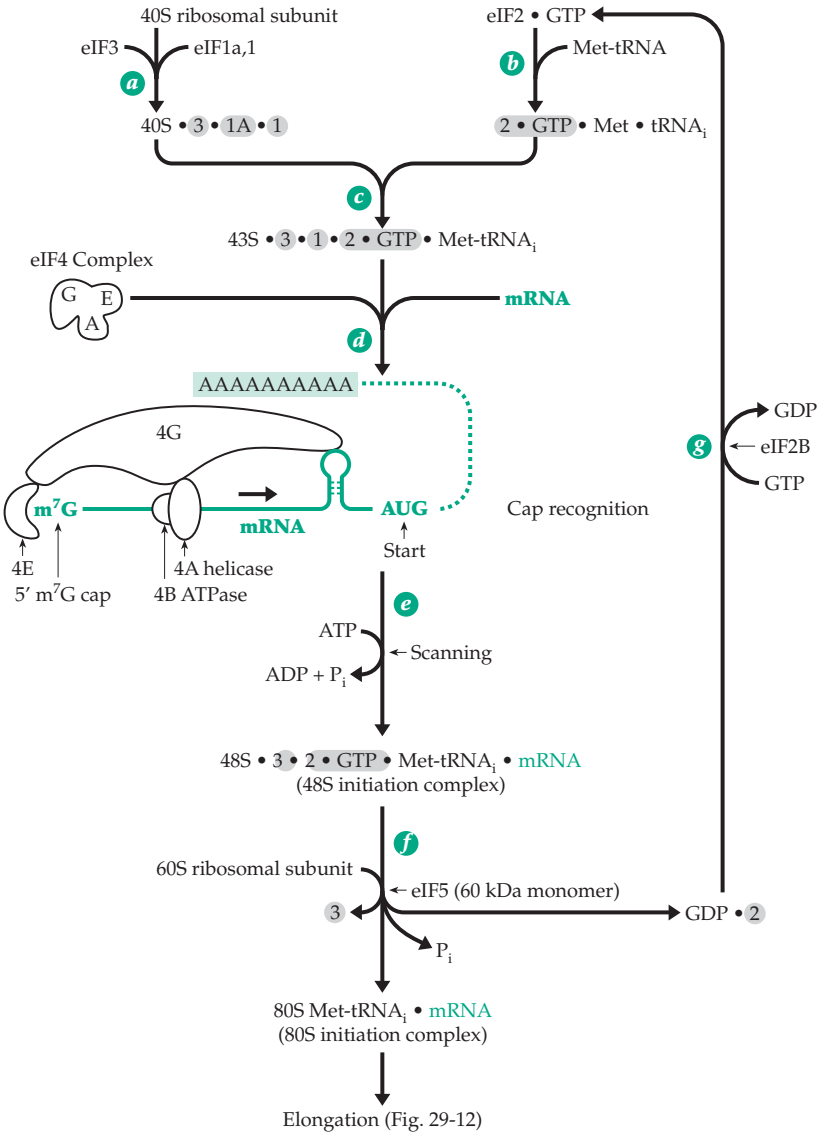


Figure 29-11 Initiation of eukaryotic protein synthesis. 1, 2, 3 = eIF1,2,3

initiation factors (eIFs), some composed of several peptides.[309] Hydrolysis of both ATP and GTP is required to form the initiation complex. Cap-binding proteins help to locate the 5' end, but the first initiation codons occur at greatly varying distances from the cap. Ribosomes apparently conduct a systematic scanning beginning at the 5' end cap of the mRNA to locate the first initiation codon.[305,310,311] Initiation of translation in eukaryotes is also often subject to controls that are more complex than those in bacteria.[308,312–314a] At least 25 polypeptides are involved.[315] Specific functions of only a few of these are described here. Functions of some are unknown or uncertain, and new proteins such as the stimulatory factor 4H[316,317] have been discovered recently.

The first initiation step is the dissociation of idle 80S ribosomes into their 40S and 60S subunits. This depends upon the ~700-kDa eIF3, a complex of 5–11 peptides of mass 30 to 170 kDa each, which binds to the 40S subunit (Fig. 29-11, step *a*).[306,311,318–320] In a separate reaction (step *b*) the charged initiator tRNA (Met-tRNA$_i$) is bound by the G protein eIF2,[321–324] an αβγ mixed trimer whose α subunit not only binds GTP but is also the site of regulation by a phosphorylation–dephosphorylation mechanism.[325,326] As is indicated in Fig. 29-11, GDP-containing eIF2, released from ribosomes upon formation of the initiation complex, does not bind the charged initiator tRNA. The bound GDP must first be exchanged for GTP, a process that requires a five-subunit guanine nucleotide exchange factor eIF2B (Fig. 29-11, step *g*).[327–328a] However, if the α subunit of eIF2 becomes phosphorylated, the nucleotide exchange and consequently the initiation of protein synthesis, is retarded.[329–330c] In reticulocytes protein synthesis stops rapidly if there is a deficiency of heme. This appears to occur via a **heme-sensitive eIF2α kinase** whose catalytic activity is inhibited by heme.[322,327,331] The [NAD$^+$]/[NADH] ratio may also be a factor in controlling the nucleotide-exchange GEF-catalyzed reaction.[332]

The ternary complex of eIF2 with GTP and Met-tRNA$_i$ binds to the 40S complex (Fig 29-11, step *c*). If the ribo-trinucleotide AUG, the initiation codon, is added to this complex, it is converted to a 43S initiation complex.[321] However, with natural mRNAs several additonal protein factors as well as ATP are needed (Fig. 29-11, step *d*). Factor 4F (eIF4F) is a large complex of several components known as factors 4A, 4B, 4E, and the large 220-kDa 4G (formerly EIF4γ or p220).[333–334a] Factor 4G is a multifunctional adapter or scaffold that apparently organizes the complex and coordinates various control mechanisms.[335–338] Factor 4E is a cap-binding protein, which recognizes and binds tightly to the ^{7}mG cap present on most mRNAs.[306,339–341a] It also binds to 4G. Factor 4A, an ATPase, acts together with 4B as an RNA helicase to unwind the mRNA and remove hairpin loops[342,343] during the scanning to locate the initiation codon (step *e*). Kozak suggested that the 43S ribosome scans from the cap at the 5' end and stops at the

first initiation codon, which is usually the AUG found within the sequence (A/G)NNAUGG.

AUG codons in other positions, known as **internal ribosome entry sites** (IRES),[311,344–347] and, more rarely, non-AUG codons can also initiate translation with lower efficiency.[348] Thus, mechanisms exist for synthesis of small amounts of proteins of varying lengths and of proteins that are encoded in any one of the three reading frames.[305,349–350a] Even circular RNAs can serve as mRNAs by this mechanism.[351] It is significant that, as shown in Fig. 19-11, factor 4G, the large subunit of eIF4, also binds to the poly(A) tail present on the 3' terminus of most mRNAs. This binding, which seems to be essential for rapid initiation,[352–354] is mediated by yet another protein, the **poly(A)-binding protein**. The importance of this protein in the human body is emphasized by its identification as **ataxin-2**, the protein defective in type 2 spinocerebellar ataxia (see Table 26-4).[354] The significance to the regulation of initiation is not clear, but the poly(A) binding may favor reuse of the mRNA, which may be translated repeatedly under conditions of rapid growth.

The last initiation step (step *f*, Fig. 29-11) is the reaction of the 60S ribosomal subunit with the 48S initiation complex to form the 80S initiation complex. Initiation factors 3, 4C, the eIF2•GDP complex, and inorganic phosphate are all released in this process, which is promoted by IF5. This monomeric ~60-kDa protein[355,356] also stimulates conversion of the GTP bound to IF2 into GDP and P_i. IF5 is unique as the only known protein containing **hypusine**, N^{ε} -(4-amino-2-hydroxybutyl)lysine, a posttranslationally modified lysine. It occurs only at position 50 in the 17-kDa protein.[356–358] Hypusine is not present in eubacteria but is essential for viability of both eukaryotes and archaeobacteria[358] and is present within an invariant 12-residue sequence.

2. Elongation of Polypeptide Chains

Once the initiating fMet-tRNA of bacteria or the eukaryotic Met-$tRNA_i$ is in place in the P site of a ribosome and is paired with the initiation codon in the mRNA, peptide chain growth can commence. Amino acid residues are added in turn by insertion at the C-terminal end of the growing peptide chain. Elongation requires three processes repeated over and over until the entire peptide is formed.

1. Codon-specific binding of a charged tRNA bearing the next amino acid at the A site (decoding).
2. Formation of the peptide bond. This process transfers the growing peptide chain from the tRNA in the P site onto the aminoacyl-tRNA in the A site.
3. Translocation of the peptidyl tRNA from the A site to the P site. This process also involves movement of the used tRNA from the P site into the exit site and simultaneous movement of the mRNA to bring the next codon into place in the A site. Both the release from the A site and translocation require energy. This is provided by the hydrolysis of GTP, one molecule for each of the two processes.[359,360]

The elongation cycle for *E. coli* is shown in Fig. 29-12. That for eukaryotic ribosomes is similar except that 40S and 60S subunits are involved in formation of the complete 80S ribosome.

Codon-specific binding of an aminoacyl-tRNA (decoding). The binding of an aminoacyl-tRNA to the A site of the 70S or 80S initiation complex depends upon a protein called **elongation factor Tu** (**EF-Tu** or **eIF1** in eukaryotes), which is present as a mixed dimer with a second protein, **EF-Ts**. In *E. coli* EF-Ts is a stable 35-kDa protein, while Tu is a 43-kDa soluble protein present in a large excess over Ts. Tu is one of the most abundant soluble proteins in bacterial cells and accounts for about 5% of the total protein. Most of the tRNAs in a bacterial cell are present as complexes with Tu. Tu may also have functions other than in protein synthesis and is found associated with the plasma membrane as well as with ribosomes.

Factor Tu is a G protein. It not only carries the aminoacyl-tRNAs into the A site on ribosomes, as shown in Fig. 29-12, but also binds and hydrolyzes GTP during the elongation cycle.[361–368] Factor Ts is a nucleotide exchange factor that catalyzes the exchange of GDP bound to Tu for GTP.[369] This is shown in Fig. 29-12 (steps *a* and *b*). The GTP/GDP-binding site of EF-Tu is located in the N-terminal portion. Eukaryotic eEF-Tu is also called EF1α or EF-TA. As isolated from various sources it has a molecular mass ranging from 50 to 53 kDa. Like the bacterial counterpart, it is abundant.[368,370–373] Like bacterial EF-Tu it exists largely as a complex with a more abundant nucleotide exchange factor EF1β. The complex tends to be bound to actin filaments.[373,374] Fungal eEF-Tu contains mono-, di-, and trimethylated lysine at up to 16 positions.[371]

EF-Tu will bind to any aminoacylated tRNA other than $tRNA^{f\text{-}Met}$, the initiator $tRNA$[374a] (step *c*, Fig. 29-12), and carry it to the ribosome (step *d*), where it binds into the A site. There it is selected if it forms a proper base pair with the mRNA codon in the A site or is rejected if it does not. This decoding process involves both an initial step and a proofreading step. The aminoacyl-tRNA binds both to the decoding site in the 16S RNA and to the peptidyltransferase site in the 23S RNA. (See discussions on p. 1687.) The decoding site is on the platform at the upper end of helix 44 (Fig. 29-2). Nucleotide G1401 plays a crucial role.[375] When one of the isoacceptor species of *E. coli* $tRNA^{Val}$ is irradiated with ultraviolet light, the

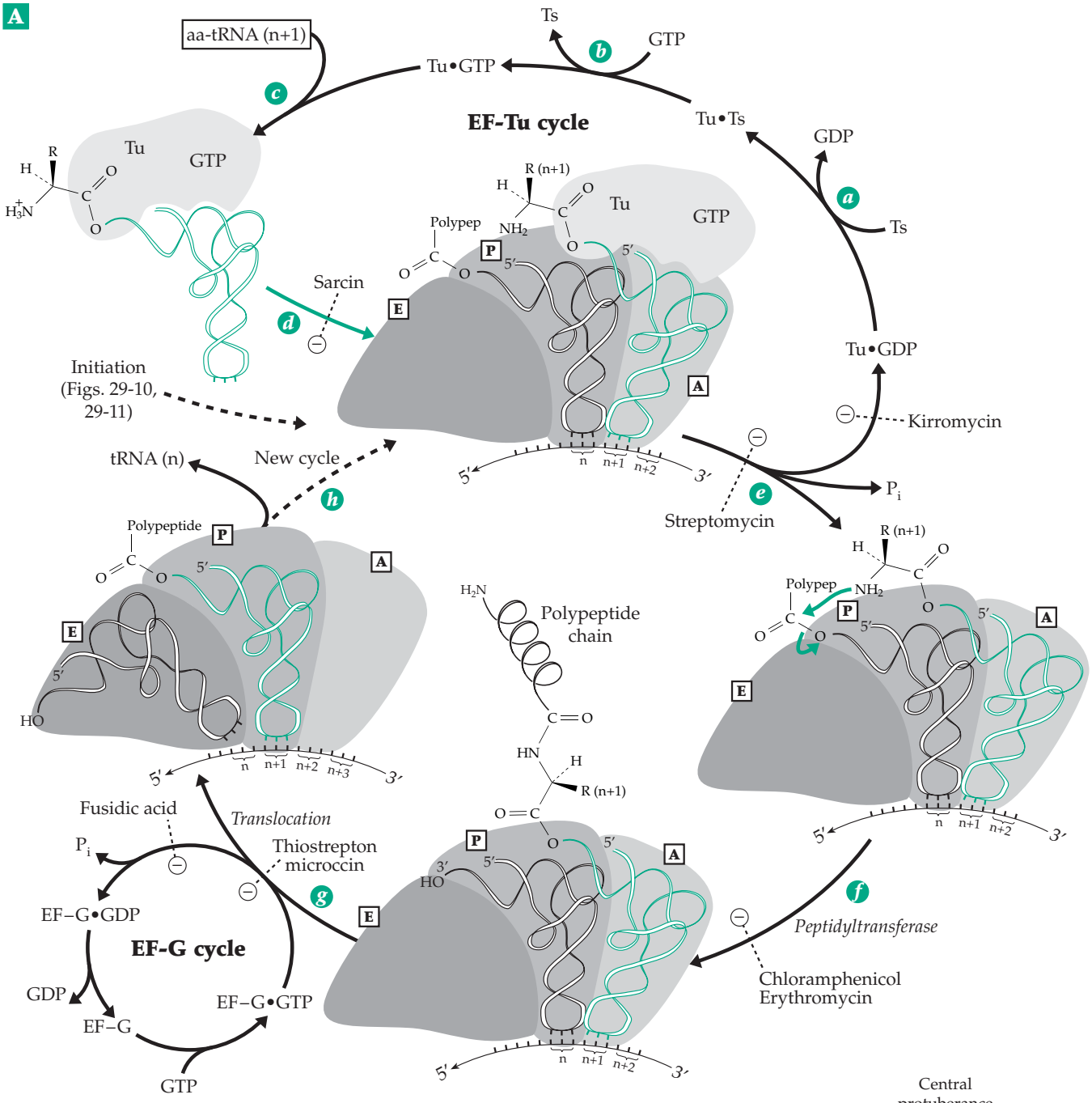
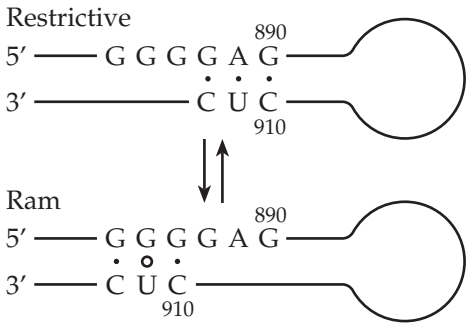


Figure 29-12 (A) Classic version of the polypeptide elongation cycle. The green color traces the pathway of an incoming tRNA carrying a new aminoacyl group. The decoding center is at the lower end of the A and P sites, while the peptidyltransferase is at the upper edge. The drawing is schematic, and the orientations of the tRNAs in the three sites are not pictured correctly. Here Tu=EF-Tu and G=EF-G. (B) Path of transfer RNA through ribosome.

5-(carboxymethoxy)uridine at position 34 in the anticodon becomes crosslinked specifically with C-1400 of the 16S rRNA in the 30S subunit.[5,376] This nucleotide lies in the deep cleft, in the decoding region, between the neck and the platform of this ribosomal subunit (Fig. 29-1). Various crosslinking and protection experiments[377,378] show that other helix 44 residues bind the tRNAs in both the P and A sites. A1492 and A1493 form part of the A site, while C1400 is in the P site.[378,379] Also strongly affecting tRNA binding and decoding is the nearby **switch helix** in the 900 region of the 16S RNA. This helix readily undergoes a shift between two hydrogen-bonded configurations (Eq. 29-9).[378,378a,380]



(29-9)

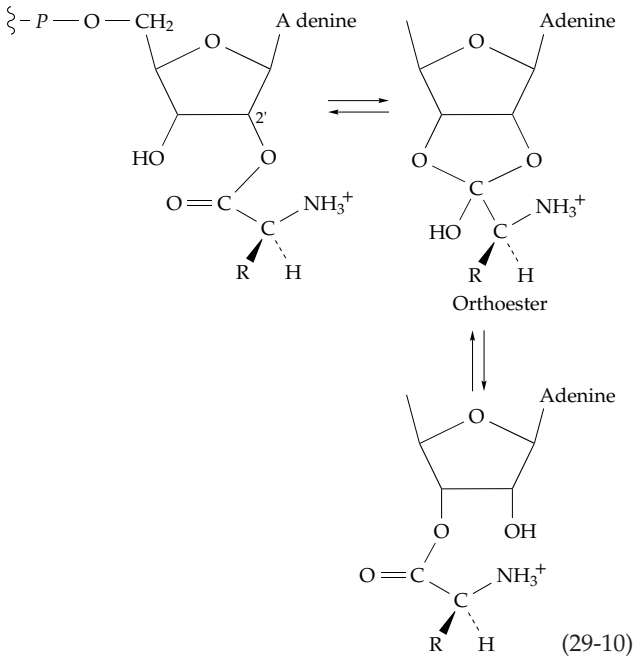
Judging by the effects of mutations in 16S RNA or in proteins S5 and S12 that favor one or the other conformation, the restrictive conformation gives a greater fidelity in translation than the "ram" (ribosomal ambiguity) conformation.[380] This loop is near the central pseudoknot in the 16S RNA and is involved in binding S5 and S12 as well as streptomycin (Box 29-B), all of which affect fidelity of protein synthesis. As mentioned on p. 1687 the adenine rings of residues A1492 and A1493 move out to interact with the CCA-3' ends of the tRNA (Fig. 29-14).

The location of binding of EF-Tu on ribosomes has been established directly by cryo-EM. It binds both to the L7/L12 stalk and to the body of the 50S ribosomal subunit.[381] The other end of the P site is at the peptidyltransferase locus and has been photochemically labeled by azide derivatives of aminoacyl groups bound to a tRNA.[382] The labeling is primarily in the 50S subunit of *E. coli* ribosomes and involves the central loop of domain V (residues 2043–2625) of the 23S RNA. Residues U2584 and U2585 are major sites of crosslinking (see Fig. 29-14). The presence of nearby sites of mutation leading to resistance to chloramphenicol or erythromycin[383,384] (Box 29-B) served to confirm the central loop as part of the peptidyltransferase. Domain II of 23S RNA is also involved, and there is evidence that the unique sequence UGG at positions 807–809 may also interact with the CCA end of tRNA in the P site.[361]

Bound Tu leaves a "footprint" at positions 2655 and 2661 in the sarcin/ricin loop of domain VI (Fig. 29-4H) when chemical probes are applied.[385] From a thermodynamic viewpoint, the hydrolysis of GTP to GDP and P_i during the functioning of EF-Tu is unnecessary, but it appears to drive a conformational change needed to bring the reacting groups together or as part of a proofreading mechanism.[385-385d] The hydrolysis of GTP appears to follow codon–anticodon recognition between the tRNA and mRNA in the A site (see Figs. 29-2G and 29-14). EF-Tu • GDP has a greatly reduced affinity for an aminoacyl-tRNA and dissociates, leaving the latter firmly bound into the A site.[385b–d] Simonson and Lake proposed that binding of a tRNA into the A site is *preceded* by binding into a **D site**. After the initial binding the anticodon of the tRNA wing from the D site into the A site as a result of a conformational rearrangement of the base stacking within the tRNA.[385e]

The peptidyltransferase reaction. It has been difficult to establish whether the 2' or the 3' of the terminal adenosine of tRNA carries the activated aminoacyl or peptidyl group. Rapid equilibration between the two via an orthoester may occur (Eq. 29-10), and EF-Tu of *E. coli* binds to either the 2' or 3' isomer.[386,386a] However, reaction of the 3'-aminoacyl-tRNA at the peptidyltransferase site is probable.[387,387a,b]

The peptidyltransferase reaction resembles that of the proteases (Chapter 12, pp. 649, 650), with a tetrahedrally bonded intermediate probable (Fig 29-13A). As is shown on pp. 649–650, the catalytic acid has been proposed to be the N3 atom of adenosine 2486 (2451 in *E. coli*) in the *H. marismortui* 23S RNA. This is in the central loop of domain V (Fig. 29-14). However, replacement of A2451 with G, U, or C did



(29-10)

not totally destroy peptidyltransferase activity in *Thermus aquaticus*, nor did replacement of the essential G2447 with A, U, or C. Polacek *et al.* suggest that the ribosome may need only to hold the reacting aminoacyl and peptidyl groups attached to tRNA in the correct positions for reaction and that no other catalysis is necessary.[387c] However, A2451G, U, or C mutant ribosomes have very low activity and organisms with such mutations are often not viable.[33f,167b] The pH rate profile for peptidyltransferase activity indicates a pK_a of ~7.5 in the RNA.[33f,387d] This is consistent with the view that A2451 may serve as a catalytic base. There has been considerable discussion about the pK_a. Can it be assigned to A2451? As mentioned on pp. 751–753, many enzymes have a broad pH region of maximum velocity over which catalytic groups of quite different microscopic pK_as (pp. 305–307) may function. For ribosomal RNAs, as for proteins, tight bonding between ionized groups in a substrate–catalyst complex may lock in an overall protonation state of the macromolecule. However, a proton may jump from one group to another within the complex (e.g., as in Fig. 29-14D) to provide a set of tautomeric species in a pH-independent equilibrium. Among these some will be on the catalytic pathway. One may arise by deprotonation of the reacting $-NH_2$ group of the aminoacyl-tRNA (Eq. 29-1, step *c*). Conformational changes,[387e] which may be induced by proton movements, may also be encompassed within the array of pH-independent equilibria.

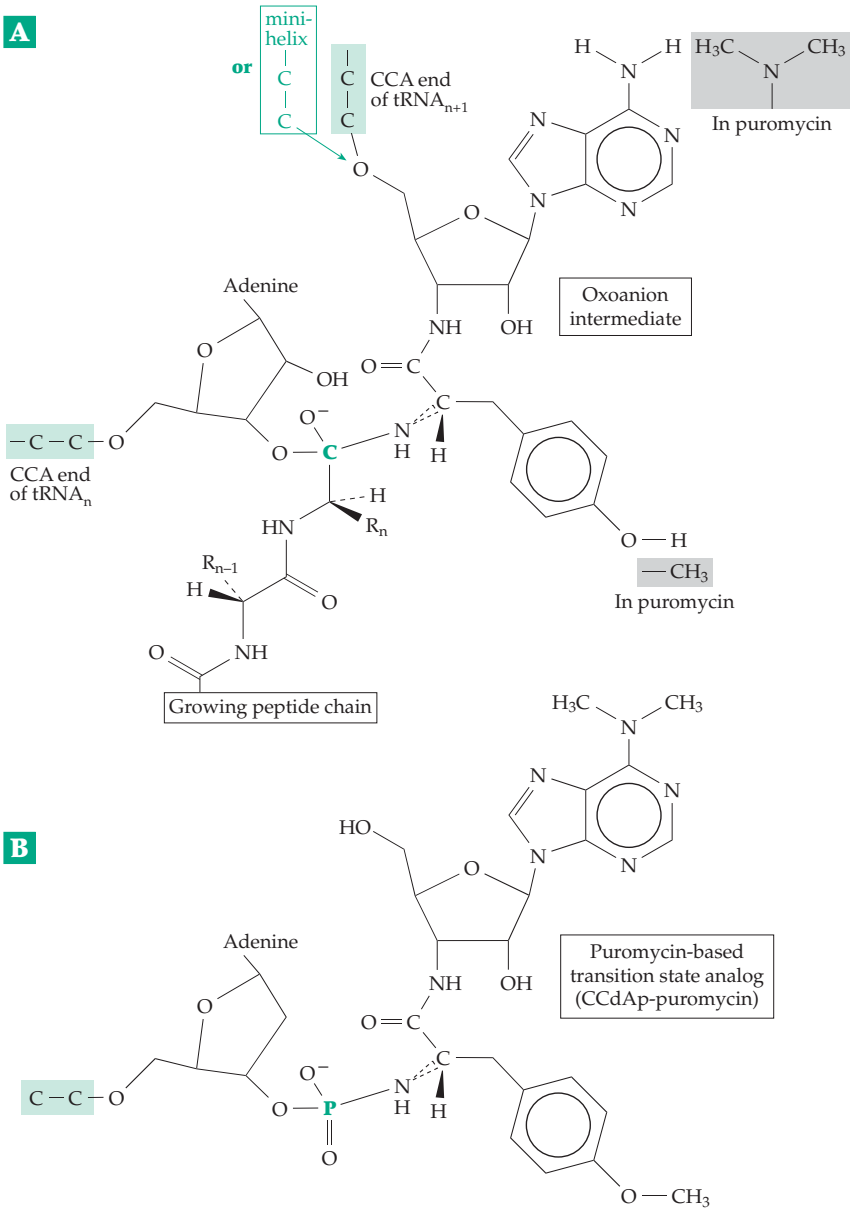


Figure 29-13 (A) Structure of expected intermediate with tetrahedral C-atom in peptidyltransferase reaction with a tRNA, with a minihelix analog, or with the antibiotic puromycin. (B) Transition-state (or bisubstrate) analog formed with puromycin and a mimic of the CCA end of a tRNA. See Box 29-B.

A careful stereochemical analysis has led to the conclusion that for all of the different aminoacyl groups to be able to react in the same way at the peptidyltransferase site and to all generate trans amide linkages, the torsion angles ϕ and ψ of the resulting peptide must be approximately those of an α helix.[388] Thus, the peptide emerging from the ribosome exit tunnel may be largely helical.

Elongation factor EF-G and translocation. The third step in the elongation sequence on ribosomes (Fig. 29-12, step *g*) depends upon **EF-G**, a monomeric GTP-binding protein with a sequence homologous with that of other members of the G protein family. It apparently utilizes the Gibbs energy of hydrolysis of GTP to GDP to drive translocation of the peptidyl-tRNA from the A site to the P site (Fig. 29-12) and of the previously utilized (deacylated) tRNA to the exit site. EF-G binds to the 50S ribosomal subunit at the base of the L7 / 12 stalk as indicated in Fig. 29-1.[392,393] It competes with EF-Tu, which binds in nearly the same location.[5] EF-G is a large five-domain GTPase. Domain 1 contains the GTPase site and resembles other G proteins, and domain 2 has some similarity to the

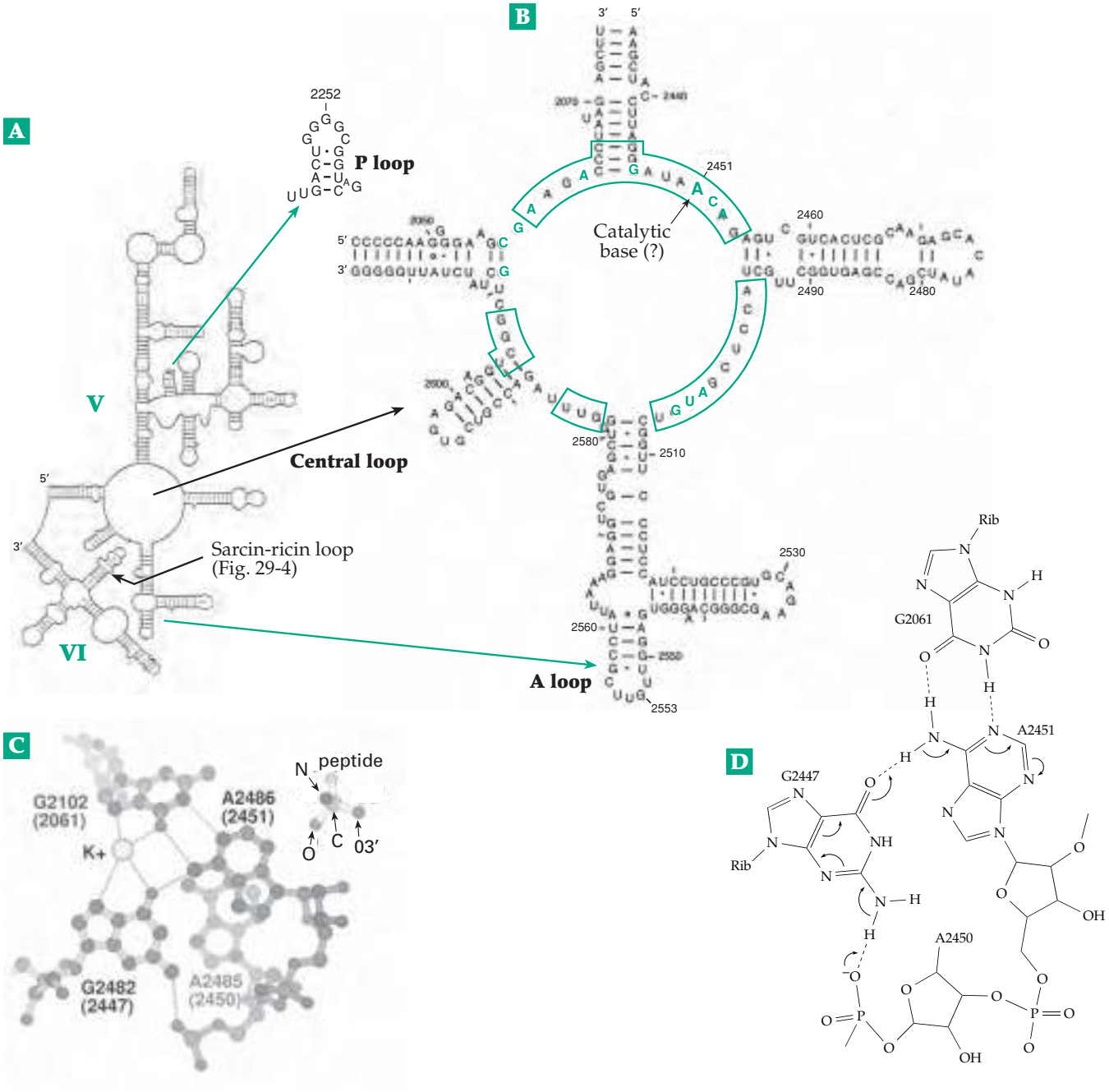
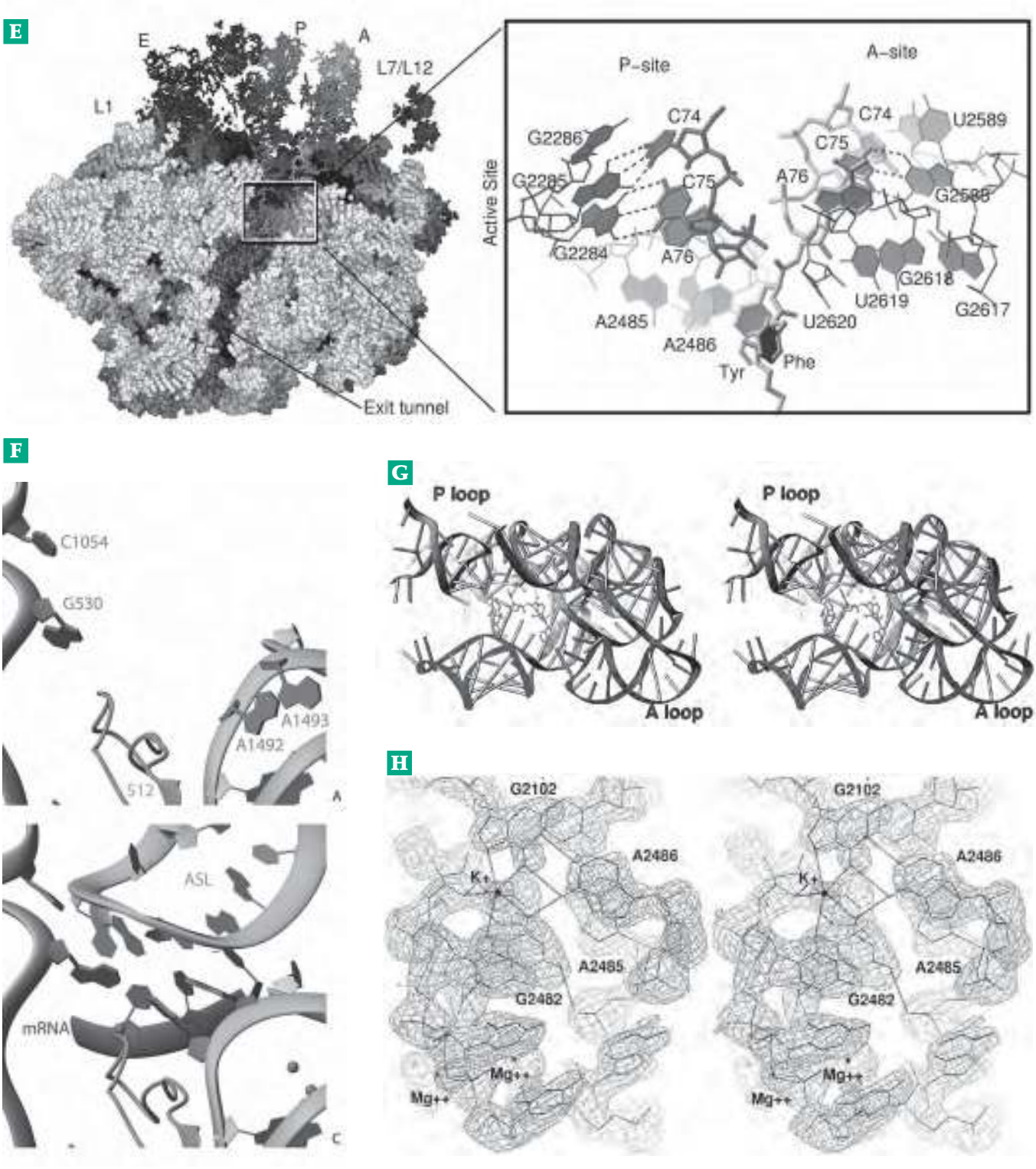


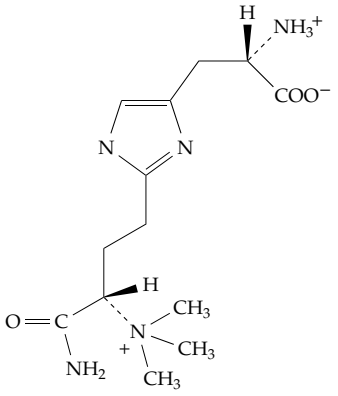
Figure 29-14 The ribosomal peptidyltransferase center in the 23S RNA. (A) Secondary structure map for domains V and VI of *E. coli* 23S RNA. From Samaha *et al.*[389] (B) Sequences of the P loop and A loops and of the central loop of domain V and of the 23S RNA of *Halobacterium halobium* with numbering according to the *E. coli* sequence. Sequences within the green boxes are highly conserved in eubacteria, chloroplast, and mitochondrial RNAs. See Kloss *et al.*[390] and Garret.[80] Sites of mutations that confer resistance to erythromycin (G2057, A2058, and C2611) and chloramphenicol (G2057, G2447, A2451, C2452, A2503, and U2504) are indicated. See Douthwaite and Aagaard.[383] (C) A skeletal representation of the peptidyltransferase active site which is depicted more completely in the stereoscopic view in (H). Included is the peptidyl group on the 2'-end of a tRNA. (D) Schematic region of domain V showing the hydrogen-bonding interactions of the catalytic base A2486 (2451, *E. coli*) with neighboring bases and also locations of nearby K^+ and Mg^{2+} ions. (C, D, G, H) are from Nissen *et al.*[19] (E) Interaction of the CCA 3'-ends of ribosome-bound tRNAs (at top) with the large ribosomal subunit. This is a cutaway view with tRNAs in the A, D, and E sites. The ribosome is sliced to show the acceptor ends of the tRNAs in the entrance to the peptide tunnel. Enlarged section shows interactions of the CCA 3'-ends in the P and A sites. The CCA end of a deacylated tRNA is shown in the P site, while a peptidyl-tRNA analog (CCA puromycin–phenylalanine–caproic acid–biotin) is shown in the A site. Bases of the 23S rRNA are numbered as in *H. marismortui*. Bases A2486 and U2620 (corresponding to A2451 and U2585) are closest to the newly formed peptide bond. From Schmeing *et al.*[33f] (F) Discrete states of the A site of the 30S subunit, as deduced from four different crystal structures. The tRNA, anticodon stem-loop (ASL), A-site mRNA codon, P-site mRNA, protein S12, and



important bases involved in conformational changes are shown. A few elements of the 16S RNA such as helix 44 (lower right), 530 loop (mid and lower left), and helix 34 (upper left) are also shown. At the top, the native 30S subunit. A1492 and A1493 have been stacked in the interior of H44 and G530 is in the *syn* conformation. C1054 is shown in the upper left corner. Below, when the codon and cognate tRNA-ASL bind in the A site, A1492 and A1493 flip out to monitor the codon-anticodon interaction, and G530 switches to the *anti* conformation to interact with A1492, the anticodon in the second position, and the codon in the third. Two Mg^{2+} ions are visible near the region vacated by A1492 and A1493 in the interior of helix 44, and one is located close to the ribose of the codon in the wobble position. From Ogle *et al.*[33d] Courtesy of Venki Ramakrishnan. (G) Three-dimensional structure of the active-site region of domain V including the P loop[391] and A loop and active-site region. (H) Three-dimensional structure of the active-site region showing the hydrogen-bonding interactions of the catalytic base A 2486 (2451, *E. coli*) with neighboring bases and also locations of nearby K^+ and Mg^{2+} ions. (C, G, H) are from Nissen *et al.*[19]

corresponding domain in EF-Tu. GTP hydrolysis appears to induce within EF-G a conformational change that is coupled directly to a rapid mechanical movement within the ribosome.[394,394a] The large domain 4 protrudes from the molecule and by its shape mimics a molecule of tRNA, and the complete EF-G molecule bears a striking resemblance to an EF-Tu•aminoacyl-tRNA complex.[395,396] It leaves chemical footprints around position 1067 (the thiostrepton-binding site) in domain II of the 23S RNA and in the universally conserved loop around position 2660.[385]

Factor EF-G from eukaryotes (eEF2) is similar to the bacterial protein, but its interaction with the larger eukaryotic ribosomes seems to be more complex. For example, interaction with the ribosomal stalk is more extensive.[37] EF2 contains a single modified histidine called **diphthamide**.[397] This amino acid is not found in other proteins but is always present in eukaryotic EF2 and also in EF-G from archaeobacteria. It is the site of modification by diphtheria toxin (Box 29-A).



Diphthamide

The ribosomal translocation process is quite complex. As the tRNAs move from A to P to E sites on the 16S RNA platform, the mRNA must also move in discrete single-codon steps. The acceptor stems of the tRNAs in the A and P sites must react at the appropriate times in the peptidyltransferase center. Study of protection from chemical probes suggests that tRNAs sometimes lie with the anticodon loop in the A site of the small ribosomal subunit, while the acceptor stem is in the P site of the large subunit (an A/P site as illustrated in Fig. 29-12B). Each aminoacyl-tRNA enters as a complex with EF-Tu and may initially bind with its anticodon in the A site and the acceptor stem with attached EF-Tu in a transient T site, the composite state being A/T. After loss of EF-Tu the acceptor stem can move into the A site to give an A/A state. The peptidyltransferase reaction itself necessarily involves movement at the acceptor stems by 0.1 nm or more. However, additional movement of ~1 nm is needed to move the two tRNAs into states A/P and P/E, respectively. Movement of the mRNA then moves the anticodon loops of the peptidyl-tRNA into the P/P state and of the deacylated tRNA completely into the exit site (Fig. 29-12B).[86,397a] Translocation may occur at different times in the 30S and 50S subunits. The pathway of the mRNA through the ribosome is known[397b,c] and is indicated approximately in Fig. 29-12B.

EF-G seems to be the motor protein that drives translocation in the 30S subunit. When it is not attached to a ribosome, the EF-G•GTP complex is very stable, but in its functioning location GTP is hydrolyzed rapidly. This occurs prior to translocation[398,398a] and presumably causes an internal alteration in the ribosome that energizes it for the translocation step. G proteins usually undergo large changes in conformation when GTP is hydrolyzed (Chapter 11). A very large change is observed for EF-Tu, but such a change has not been seen for EF-G. However, large conformational changes in the ribosome do evidently accompany translocation.[399] The hydrolytic activity of EF-G is stimulated strongly by its binding to the L7/L12 stalk proteins.[400] Eukaryotic EF2 like EF-G binds to the stalk proteins (P1, P2, P0 complex) and also to domains of 28S rRNA equivalent to the bacterial 1067 and sarcin / ricin loops.[401] However, EF-G and EF2 are not functionally interchangeable.

Translocation occurs slowly even in the absence of GTP. However, it is greatly enhanced by GTP hydrolysis.[394a,402] Even empty ribosome, without tRNAs, but in the presence of GTP and EF-Tu and EF-G, hydrolyze GTP. The ribosome may sequentially bind EF-Tu, then EF-G, oscillating between two differing states.[403] The movement of mRNA through the ribosome has been plotted using a variety of immunochemical, crosslinking, and chemical footprinting methods.[52,404–407]

A third elongation factor, eEF3, which is an ATPase, is required by yeast and fungi.[408–410] The 1044-residue yeast protein may be required for ATP-dependent release of deacylated tRNA from the exit site.

Polyribosomes. Under suitable conditions ribosomes isolated from cells are found to sediment together in clusters, often of five or more. These **polyribosomes** (or **polysomes**), which can be seen in electron micrographs (Fig. 28-5), are held together by chains of mRNA. Polyribosomes arise because a single mRNA molecule is being translated by several ribosomes at once. As the 5' terminus of the mRNA emerges from one ribosome, it may soon combine with another and initiate translation of a second peptide chain, etc. The length of the mRNA determines how many ribosomes are likely to be associated in a polyribosome.

Rates of synthesis of ribosomes and of proteins. In a rapidly growing yeast cell with a generation time of ~100 min there are nearly 200,000 ribosomes. Almost 200 new ribosomes must be formed in one minute. Each

of the 150 tandemly repeated ribosomal RNA genes must be transcribed into the 4560 nucleotides of one ribosome in less than one minute. The ~150 nuclear pores must import nearly 1000 ribosomal proteins per minute and must export ~25 ribosomal subunits per minute.[410a] The ribosomes that are formed can at 37°C add 14–17 amino acids per second to a growing polypeptide chain,[410b,c] while eukaryotic ribosomes can add 2–4 amino acids per second.[410d,e]

3. Termination of Polypeptide Synthesis

A ribosome faithfully translates the genetic message, adding amino acids to the polypeptide chain until a stop codon is reached. Then a **termination** or **release factor** acts, probably by binding directly to the stop codon on the mRNA in the A site.[411–413b] In *E. coli* termination factor **RF1**, a 47-kDa protein, recognizes UAA or UAG, while **RF2**, a very similar protein,[414,415] recognizes UAA or UGA. There are several hundred molecules per cell of these release factors. They not only recognize the stop codons but also catalyze the hydrolytic removal of the peptidyl chain from the tRNA in the ribosomal P site. They bind into the A site, where they may interact with mRNA bases in addition to those of the stop codon.[415,416] Hydrolytic release of the polypeptide chain from the tRNA in the P site may represent a change in specificity of the peptidyltransferase center induced by binding of a release factor. Genes are often terminated by a succession of two stop codons. Thus, there is a safety factor that prevents translation from continuing in case the first stop codon is missed. An example is provided by the *I* gene of the *lac* operon of *E. coli* (Fig. 28-2), which has a second stop codon in phase with the TGA codon marked in the figure and located five codons further "downstream" (to the right). A third release factor **RF3**[416] is a GTP-binding protein resembling EF-G. It is not essential to life for *E. coli*, but it accelerates the release of RF1 or RF2 and is needed for rapid growth.[417,418] Eukaryotes contain one release factor **eRF1**, which recognizes all three termination codons, and a second release factor **eRF3**, which binds and hydrolyzes GTP.[413,413b,418a,419]

Just as elongation factor EF-G mimicks the aminoacyl-tRNA•Tu complex, release factors RF1 and RF2, in their shapes, mimick molecules of tRNA.[419a] One domain of human eRF1 has an anticodon-recognition domain and a conserved GGQ sequence in a second domain, which mimicks the amino-acceptor arm of tRNA.[419,419b] Mutations in either eRF1 or eRF3 affect translational accuracy and may allow "read-through" of stop codons. In yeast (*S. cerevisiae*) a 685-residue subunit of eRF3 has an N-terminal domain, that like the human **prion protein** (Box 29-E) is capable of being transformed into a self-seeding amyloid-like conformation. In the yeast the formation of amyloid aggregate leads to depletion of the termination factor and increased readthrough.[420–422]

Recycling factors. Even though release factors remove the completed polypeptide chain, a ribosome is not ready for reuse until the deacylated tRNA in the P site is removed and the mRNA is released. This depends upon **ribosome recycling factor** (**RRF**) together with EF-G. The recycling factor is also a tRNA mimic.[419a,423-424b] It may bind into the empty A site, and in an action similar to that of the translocation step of elongation remove the P site tRNA.[417,423–425] However, probing with hydroxyl radicals indicates a different mode of binding.[425a]

4. Preventing and Correcting Errors in Translation on the Ribosome

The wrong amino acid is inserted into most positions in a protein about one time in ~10^4, a frequency[361,426,427] of ~10^{-4}. However, in *E. coli* misreading of certain codons is observed more often. For example, AAU (Asn) is read as AAA (Lys) with a frequency[428] of ~5 x 10^{-3}. Misreading also depends upon adjacent codons, i.e., the codon context.[429] Having all of the tRNAs charged with the correct amino acids, as discussed in Section B3, is a first essential for accurate translation. A second is finding the correct location of the initiation codon and binding of the aminoacyl-initiator tRNA into the P site. The decoding process by which the correct aminoacyl-tRNA is brought into the A site is still not fully understood. It has often been proposed that (as in DNA replication; Chapter 27, Section C,2) the fidelity of this process depends upon two consecutive recognition steps.[265,266] The first is the binding of the complex of EF-Tu•GTP and the charged tRNA to the ribosome. The second may be associated with the conformational change that locks the aminoacyl-tRNA into the A site and perhaps sends to the peptidyltransferase center a signal that the correct codon-anticodon pairing has been achieved. Some checking is done in the first step. For example, many of the 380 possible mischarged forms of aminoacyl-tRNAs that may have escaped previous proofreading steps (pp. 1695–1696) are rejected because they bind too loosely or too tightly to EF-Tu.[429a,b] Codon–anticodon base pairing may also be checked in the P site[429c] after translocation. The P site is buried deep in a cleft in the RNA of the large subunit. It is designed to hold the mRNAs in a kinked conformation (Fig. 29-1E) with the codons in the A and P sites oriented differently. Some tRNA residues required for high-fidelity participate in imposing this geometry. Mutants in either of the major rRNAs or in tRNAs can lead to loss of fidelity in base pairing and sometimes to excessive frame-shifting.[429c] During the

proofreading process a mispaired aminoacyl-tRNA may be allowed to dissociate and be replaced by a new one. Certain mutations, such as those in ribosomal protein S12 that lead to streptomycin resistance, cause greatly increased fidelity of protein synthesis. However, these mutations slow bacterial growth,[426,430] perhaps because some misreading is necessary for synthesis of minor essential proteins.

A strictly hypothetical way in which an alteration in hydrogen bonding could be used to signal the achievement of correct base pairing is illustrated in Fig. 29-15. As indicated by the curved arrows, the approach of a negatively charged group could induce an electron pair to move from the ring nitrogen on the right side. If the pairs of hydrogen bonds were correct, a concerted flow of electrons could take place across the base pair and out into group H-Y and beyond through the postulated tautomeric chain. If the base pair were not correctly formed, the signal could not be transmitted, except during an occasional mispairing with a minor tautomer. Note that another reciprocal electron transfer in the opposite direction to that shown in the figure is also possible through the same base pair. Similar tautomeric shifts are possible for all legitimate base pairs. Initiation of a signal of the type shown could also occur by the addition of some nucleophile to a purine or pyrimidine ring, e.g., to C-6 of the cytosine ring as in Eq. 29-3. In ribosomes such electronic signals could be passed in turn through each of the base pairs involved in codon–anticodon recognition and also through other base pairs formed within loops of ribosomal RNA. If group H-Y is connected by a suitable chain of hydrogen bonds that passes through the active site of the peptidyltransferase, coupling between the recognition signal and the formation of

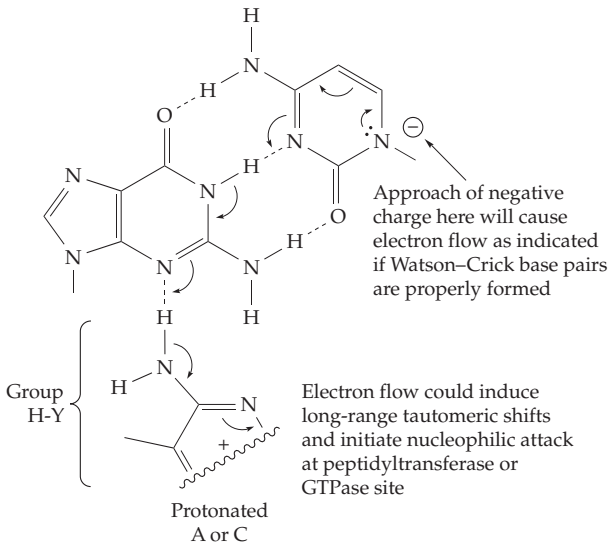


Figure 29-15 Hypothetical scheme by which an electronic signal might be sent through a base pair to initiate the peptidyltransferase reaction. See also Metzler.[431]

the transition state might be accomplished. Since changes in hydrogen bonding can trigger conformational alterations, the sensing of correct hydrogen bonding could increase the rate of the peptidyltransferase reaction as has been observed experimentally.[266] Base pairing in both the A and P sites may be sensed in similar ways. The observation that rRNA residues hydrogen-bond with groups in the minor grooves of base pairs[33d] seems to be consistent with the proposal of Fig. 29-15.

Codon usage. The usage of codons in specific mRNAs is not random.[432] For example, in a glyceraldehyde-3-phosphate dehydrogenase gene of yeast >96% of the 1004 codons make use of only 25 of the 61 possible coding triplets. Genes undergoing most rapid transcription are more highly biased toward these 25 than are other genes.[433] Many other evolutionary factors have affected usage. These include the need for translational accuracy.[432]

Dealing with lost peptidyl-tRNAs and broken transcripts. Many problems arise during protein synthesis. For example, a peptidyl-tRNA may become detached from a ribosome. In *E. coli* this seems to happen most frequently with peptidyl-tRNA^{Lys}. A 193-residue **peptidyl-tRNA hydrolase** is essential for life![434–436] It releases the tRNA for reuse, recycling all peptidyl-tRNAs other than formylmethionyl-tRNA. Perhaps the enzyme is essential because detached peptidyl-tRNAs are toxic, but it is more likely to be to avoid a shortage of free tRNA^{Lys}.[434]

If a ribosome starts to synthesize a protein using a broken mRNA, it will reach the end of the mRNA but will not find a termination codon. The peptidyl-tRNA will eventually fall off, but the ribosome will be stalled temporarily. Eukaryotes try to prevent this problem by sending only intact mRNAs to ribosomes,[436a,b] but bacteria have a ribosome rescue system that also tags the partially formed protein on the stalled ribosome for rapid proteolytic degradation.[437–440c] Bacteria synthesize a special 362-residue RNA that resembles a tRNA but also contains a short mRNA-like module that codes for the 11-residue peptide tag AANDENYALAA. This hybrid tRNA-mRNA, which is designated tmRNA (or *ssrA* RNA), mimicks tRNA^{Ala} and is recognized and charged by alanyl-tRNA synthetase. The resulting tmRNA^{Ala} binds into the A site of the stalled ribosome, undergoes the peptidyltransferase and translocation steps (Fig. 29-16). The old mRNA is released, the mRNA-like sequence of tmRNA becomes seated, and translation of the new tail sequence follows. The tail sequence is similar to C-terminal sequences that are known to mark other proteins for rapid proteolytic degradation. An associated protein SmpB is also required for functioning of the tmRNA system.[438a,b]

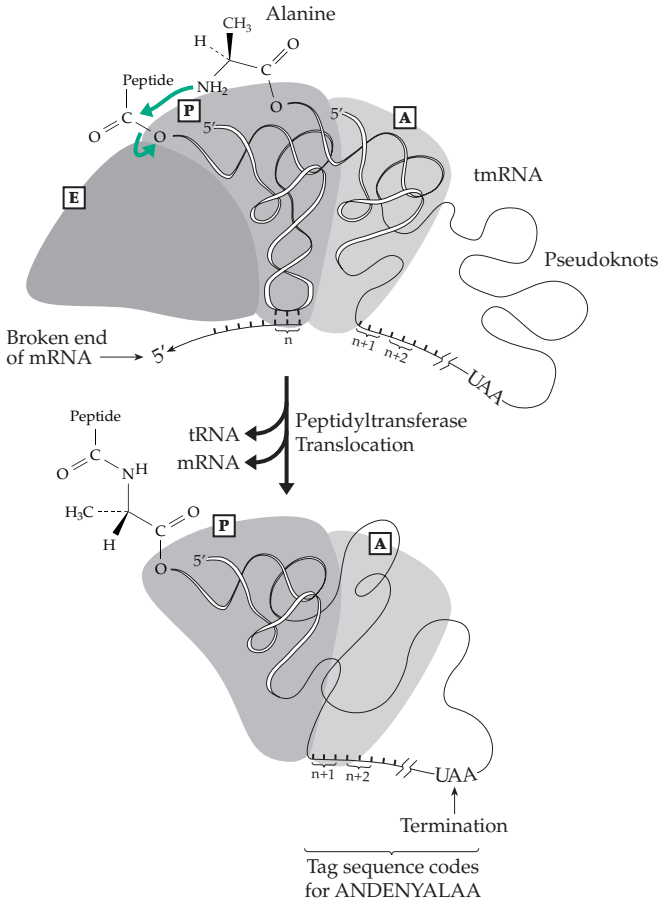


Figure 29-16 Schematic diagram of the tmRNA structure and its function in the rescue of ribosomes stalled at the end of a messenger RNA that has been broken and has lost its in-frame termination codon. After it binds into the ribosomal A site the tmRNA, which has been charged with alanine, undergoes the peptidyltransferase reaction and translocation to the P site. Then it lays down its mRNA-like coding sequence, which is used by the ribosome to add ten more amino acids to form the 11-residue C-terminal degradation signal AANDENYALAA. This induces rapid degradation of the imperfect protein that has been formed.[436a]

5. Suppressor Genes

The suppression of nonsense mutations by suppressor genes has been discussed in Chapter 26. The chemical nature of these genes was discerned, in part, from experiments involving transfer of suppressor gene *supF(su3)* into the DNA of a bacteriophage. This DNA was found to specifically hybridize with a minor transfer RNA, $tRNA_1^{Tyr}$. Subsequent investigation showed that *sup F* is a structural gene for this tRNA, and that in it the normal 5'-GUA-3' (Tyr) anticodon has been replaced with CUA. The latter can pair with the chain termination codon 5'-UAG-3' (the *amber* codon) permitting the ribosome to insert tyrosine at the site of chain termination signals introduced in *amber* mutations. It may seem puzzling that a tRNA, which prevents chain termination, does not prevent synthesis of other essential proteins within the bacterium. However, suppression is typically less than 30% efficient. Hence, many protein chains terminate normally. Since two chain termination signals are often present in a gene, most protein synthesis in the presence of the small amount of suppressor tRNA present is concluded normally. Premature chain termination caused by selected *amber* mutations will be partially inhibited, permitting the cell to make enough of the missing proteins to survive. The nucleotide sequence of a further mutated *supF* tRNA and of its longer precursor is shown in Fig. 28-10. Several other suppressor genes have also been identified as specific tRNA structural genes.[441]

A suppressor of frame-shift mutations in *Salmonella* is a tRNA containing at the anticodon position the nucleotide quartet CCCC instead of the usual CCC triplet anticodon.[442,443] It has eight unpaired bases in the anticodon loop instead of the usual seven. Other frame-shift repressor tRNAs have been identified in *E. coli*,[444] *Salmonella*, and yeast.[445] Not all suppressor genes encode tRNAs. For example, a UGA suppressor from *E. coli* is a mutant 16S rRNA from which C1054 has been deleted.[446] A general nonsense suppressor in yeast is homologous to yeast elongation factor EF-1α as well as to *E. coli* EF-Tu.[447]

Among other suppressor genes present in eukaryotic organisms[448] are mammalian genes encoding serine tRNAs that are *opal* (UGA) suppressors. These and other eukaryotic suppressor tRNAs have specific and important normal functions in cells. For example, a specific kinase phosphorylates the *opal* suppressor seryl-tRNA to its phosphoseryl derivative.[448] This suppressor tRNAs may sometimes be responsible for introducing phosphoserine at specific positions in proteins. An *opal* suppressor is also used for the introduction of selenocysteine. An *amber* suppressor is used by some methane-forming Archaea to introduce **pyrrolysine** into specific sites in methyltransferases. In pyrrolysine the epsilon amino group of lysine is joined by an amide linkage to a derivative of pyrroline-5-carboxylate (p. 1374).[448a,b]

Selenocysteine (Sec) Selenocysteine is incorporated into a small number of proteins in species from all three kingdoms of life by a suppressor $tRNA^{Sec}$ that reads certain UGA codons, which are marked as representing selenocysteine.[449,450] The selenocysteinyl-tRNA is made from a seryl-tRNA (Eq. 29-7) as described further in Chapters 16 and 24. In *E. coli* selenocysteine is present in three proteins, all formate dehydrogenases. The archaeon *Methanococcus jannaschii* contains genes for seven selenocysteine-containing proteins. Only one Sec-containing protein has been found in the nematode *Caenorhabditis elegans* and none in the yeast

Saccharomyces cerevisiae. However, there are at least 14 in the human body.[451,452] One of these, selenoprotein P, contains ten selenocysteine residues.[452,453] Products of four special genes are needed for incorporation of selenocysteine into *E. coli* proteins.[454,455] *Sel C* encodes the special $tRNA^{Sec}$, which becomes charged with selenocysteine.[456] *Sel D* encodes selenophosphate synthetase and *Sel A* selenocysteine synthetase (Eq. 29-7). *Sel B* encodes a special elongation factor, which resembles EF-Tu but has an extra domain that binds to an mRNA segment known as the **SECIS** (selenocysteine insertion sequence).[457–461a] The SECIS sequence follows the 3' end of the UGA termination codon. It is a 40-nt segment that is able to form a stem-loop structure. However, in archaea and in eukaryotes the SECIS sequence lies at the end of the selenoprotein gene in the 3' nontranslated region. It may be some distance away and may function by a foldback mechanism. It recodes the entire message, acting on any in-frame UGA codon.[461] In mammals a special SECIS-binding protein SBP2 is also required.[462]

Expanding the genetic code. Suppressor tRNAs can also be created artificially and are being used in protein engineering. *Amber*, *ochre*, or *opal* chain termination mutations can be introduced readily at many points in a protein (Chapter 26). Suppressor tRNAs can be made that will then place any one of the possible amino acids into most of the mutated positions.[463] Synthetic amino acids not normally found in proteins can also be incorporated using such tRNAs.[464–467] The TAG(UAG) *amber* stop codon is often used together with a genetically engineered tRNA. In early experiments these techniques were used to create hundreds of mutant forms of the *lac* repressor protein (see Chapter 28, p. 1606). Since then a variety of additional approaches have been explored. Transfer RNAs have been engineered to recognize four-base codons such as AGGU and CGGG.[468,469] Organisms such as *Micrococcus luteus*, in which not all of the available triplet codons have been utilized, allows development of a mutation system using an unassigned codon rather than a stop codon.[468] A general method for site-specific incorporation of any amino acid or amino acid analog requires a suppressor tRNA that is not aminoacylated by any aminoacyl-tRNA synthetase present within the host cell, and also an aminoacyl-tRNA synthetase that acts only on the suppressor tRNA and no other tRNA in the cell.[470] Several such systems are being developed.[470–472] Another idea is to utilize a 65th codon–anticodon pair, one depending upon a new synthetic nucleoside that can be incorporated into mRNA.[473]

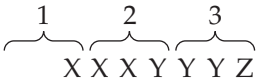
Another possible application of suppressor genes is *in vivo* suppression of undesirable termination codons. An example comes from a β^0 thalassemia caused by mutation of lysine codon CAG to UAG. By changing the anticodon of a human $tRNA^{Lys}$ gene to an *amber* suppressor, premature termination of globin chains was suppressed in an *in vivo* test.[474,475]

6. Read-Through Translation and Ribosomal Frameshifting

If termination codons are not recognized efficiently by termination factors, synthesis continues past the termination codons and new longer protein chains are made. This **read-through translation**[429,476] may sometimes be accidental, but it is also used by cells to form several important proteins. For example, the 14-kDa coat protein of bacteriophage Qβ is elongated by read-through during translation of the RNA about 4% of the time. This produces a 38-kDa protein known as A_1, which has an extra 200 amino acid residues at the C terminus and is essential for formation of infectious virus particles.

A similar situation is met with retroviruses (Chapter 28) whose coat protein gene *gag* is fused to the reverse transcription polyprotein precursor gene *pol* (Fig. 28-26).[477] In fact, the polyprotein gene overlaps the 3' end of the *gag* gene. Read-through translation allows synthesis of the polyprotein about 5% of the time. However, the *pol* gene is written in a different reading frame: −1 with respect to the *gag* gene. For example, in the HIV genome (Fig. 28-26) the *pol* gene begins at nucleotide 1638 in the −1 reading frame with respect to the *gag* gene. In both the Rous sarcoma virus and HIV a polyprotein that is formed contains peptide sequences from both *gag* and *pol* genes.[478,479] This fusion of the two proteins is accomplished by a frameshift, which occurs on the ribosome as it operates in the region of overlap of the *gag* and *pol* genes. This mechanism allows synthesis of relatively small amounts of the enzymes encoded by *pol* but large amounts of structural proteins encoded by *gag*.[480] Many other examples of essential ribosomal frameshifting are known.[429,477,481–483] For example, the gene for *E. coli* ribosomal release factor RF2 has a UGA termination codon at position 26, but the coding sequence for the protein continues in the +1 frame.[484]

Ribosomal frameshifting can be accounted for by more than one mechanism. It can occur when a four-base anticodon is present in a suppressor tRNA. It can result from incorrect base pairing. If a tRNA slips over by one nucleotide, a single base in the mRNA can be left unpaired with the reading frame being shifted +1. However, most frameshifts are in the –1 direction and occur at specific locations in the mRNA, i.e., they are **programmed frameshifts**.[484a] These often occur at "slippery sites"[485] including the following mRNA sequences in which three codons are marked:



This sequence is followed closely by an element of secondary structure, most often a pseudoknot.[486–492] Eukaryotic frameshifts are almost always in the –1 direction, the exception being found in the mammalian mRNA for **antizyme**, a negative regulator of ornithine decarboxylase (Chapter 24, p. 1382).[493] The frame-shift occurs at an initially in-frame termination codon (UGA), which is followed by a pseudoknot. Most translation is terminated at this stop codon, but frameshifting, which is induced by a high polyamine concentration, allows read-through and synthesis of the antizyme protein. In rare cases frameshifting may lead to **translational bypass** of some codons on the mRNA. Such a case is found in a bacteriophage T4 mRNA for which the *E. coli* ribosomes bypass 50 nucleotides in order to complete the synthesis of a

BOX 29-C NONRIBOSOMAL PEPTIDE SYNTHESIS

Many small biologically active peptides, including hormones and some antibiotics, are synthesized on ribosomes as precursor proteins, which are cut into small pieces and may then be modified in a variety of ways. However, many other peptides including many antibiotics are made without use of ribosomes by large polyfunctional synthetases. The first of these, gramicidin S synthetase, was described by Lipmann and coworkers in 1971.[a] It is discussed on p. 994 as is the mechanism of synthesis. It is now recognized that these enzymes are modular and have much in common with fatty acid synthetases (Fig. 17-12 and p. 1186) and polyketide synthetases (Fig. 21-11). They are able to link not only the amino acids found in proteins but also modified and unusual amino acids. They may also join one or more α-hydroxy acids to a peptide to form a depsipeptide, and they may contain modules that carry out modification reactions such as methylation, acylation, or glycosylation.[b–e] Because of their modular nature they are attractive proteins for genetic engineering.[b,f,g]

Each synthetase module contains three active site domains: The **A domain** catalyzes activation of the amino acid (or hydroxyacid) by formation of an aminoacyl- or hydroxyacyl-adenylate, just as occurs with aminoacyl-tRNA synthetases. However, in three-dimensional structure the A domains do not resemble either of the classes of aminoacyl-tRNA synthetases but are similar to luciferyl adenylate (Eq. 23-46) and acyl-CoA synthetases.[h] The **T-domain** or **peptidyl carrier protein domain** resembles the acyl carrier domains of fatty acid and polyketide synthetases in containing bound phosphopantetheine (Fig. 14-1). Its –SH group, like the CCA-terminal ribosyl –OH group of a tRNA, displaces AMP, transferring the activated amino acid or hydroxy acid to the thiol sulfur of phosphopantetheine. The **C-domain** catalyzes condensation (peptidyl transfer). The first or **initiation module** lacks a C-domain, and the final **termination module** contains an extra termination domain. The process parallels that outlined in Fig. 21-11.[i]

A few of the products of nonribosomal peptide synthesis are gramicidin S (Fig. 2-4), enniatins, bacitracins, and tyrocidines (p. 994),[b,e] vancamycin (Box 20-H),[j] actinomycin (Box 28-A),[k] the siderophore yersiniabactin,[l] surfactin (Fig. 2-4),[m,n] and cyclosporin (Box 9-F).[o,p] The δ -(L-α-aminoadipoyl)-L-cysteinyl-D-valine synthetase,[q,r] which forms the precursor to penicillin and cephalosporins (Box 20-G), also belongs to this group of enzymes as do synthetases that make cyclooctadepsipeptides with antihelminthic activity[s] and many other compounds.[t–v]

[a] Lipmann, F. (1971) *Science* **173**, 875–884
[b] Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5848–5853
[c] Linne, U., and Marahiel, M. A. (2000) *Biochemistry* **39**, 10439–10447
[d] Guenzi, E., Galli, G., Grgurina, I., Pace, E., Ferranti, P., and Grandi, G. (1998) *J. Biol. Chem.* **273**, 14403–14410
[e] Belshaw, P. J., Walsh, C. T., and Stachelhaus, T. (1999) *Science* **284**, 486–489
[f] Symmank, H., Saenger, W., and Bernhard, F. (1999) *J. Biol. Chem.* **274**, 21581–21588
[g] de Ferra, F., Rodriguez, F., Tortora, O., Tosi, C., and Grandi, G. (1997) *J. Biol. Chem.* **272**, 25304–25309
[h] Pfeifer, E., Pavela-Vrancic, M., von Döhren, H., and Kleinkauf, H. (1995) *Biochemistry* **34**, 7450–7459
[i] Cane, D. E., Walsh, C. T., and Khosla, C. (1998) *Science* **282**, 63–68
[j] Trauger, J. W., and Walsh, C. T. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3112–3117
[k] Pfennig, F., Schauwecker, F., and Keller, U. (1999) *J. Biol. Chem.* **274**, 12508–12516
[l] Miller, D. A., and Walsh, C. T. (2001) *Biochemistry* **40**, 5313–5321
[m] Weinreb, P. H., Quadri, L. E. N., Walsh, C. T., and Zuber, P. (1998) *Biochemistry* **37**, 1575–1584
[n] Reuter, K., Mofid, M. R., Marahiel, M. A., and Ficner, R. (1999) *EMBO J.* **18**, 6823–6831
[o] Lawen, A., and Traber, R. (1993) *J. Biol. Chem.* **268**, 20452–20465
[p] Hoffmann, K., Schneider-Scherzer, E., Kleinkauf, H., and Zocher, R. (1994) *J. Biol. Chem.* **269**, 12710–12714
[q] Shiau, C.-Y., Byford, M. F., Aplin, R. T., Baldwin, J. E., and Schofield, C. J. (1997) *Biochemistry* **36**, 8798–8806
[r] Kallow, W., Kennedy, J., Arezi, B., Turner, G., and von Döhren, H. (2000) *J. Mol. Biol.* **297**, 395–408
[s] Weckwerth, W., Miyamoto, K., Iinuma, K., Krause, M., Glinski, M., Storm, T., Bonse, G., Kleinkauf, H., and Zocher, R. (2000) *J. Biol. Chem.* **275**, 17909–17915
[t] Milne, J. C., Roy, R. S., Eliot, A. C., Kelleher, N. L., Wokhlu, A., Nickels, B., and Walsh, C. T. (1999) *Biochemistry* **38**, 4768–4781
[u] Gaitatzis, N., Kunze, B., and Müller, R. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11136–11141
[v] Gewolb, J. (2002) *Science* **295**, 2205–2207

topoisomerase subunit.[494–495a] Ribosomal protein L9 may play a role in the bypass process.

7. RNA Viruses

The RNA-containing bacteriophages are convenient sources of relatively simple mRNA molecules, whose sequences can be studied.[496] The genetic information for these viruses is carried by RNA molecules consisting of only 3500–4500 nucleotides and which may contain only four genes (p. 247). The RNA from phages f2, R17, MS2, and the more distant Qβ have been studied intensively.[483,497]

Parts of the 2569-nucleotide sequence for the RNA of phage MS2[498] are shown in Fig. 29-17. The 5' end (upper left center) still bears the triphosphate group of the initiating GTP. Following a number of hairpin loops there is a ribosome-protected region, which begins with the initiation codon GUG for the A protein

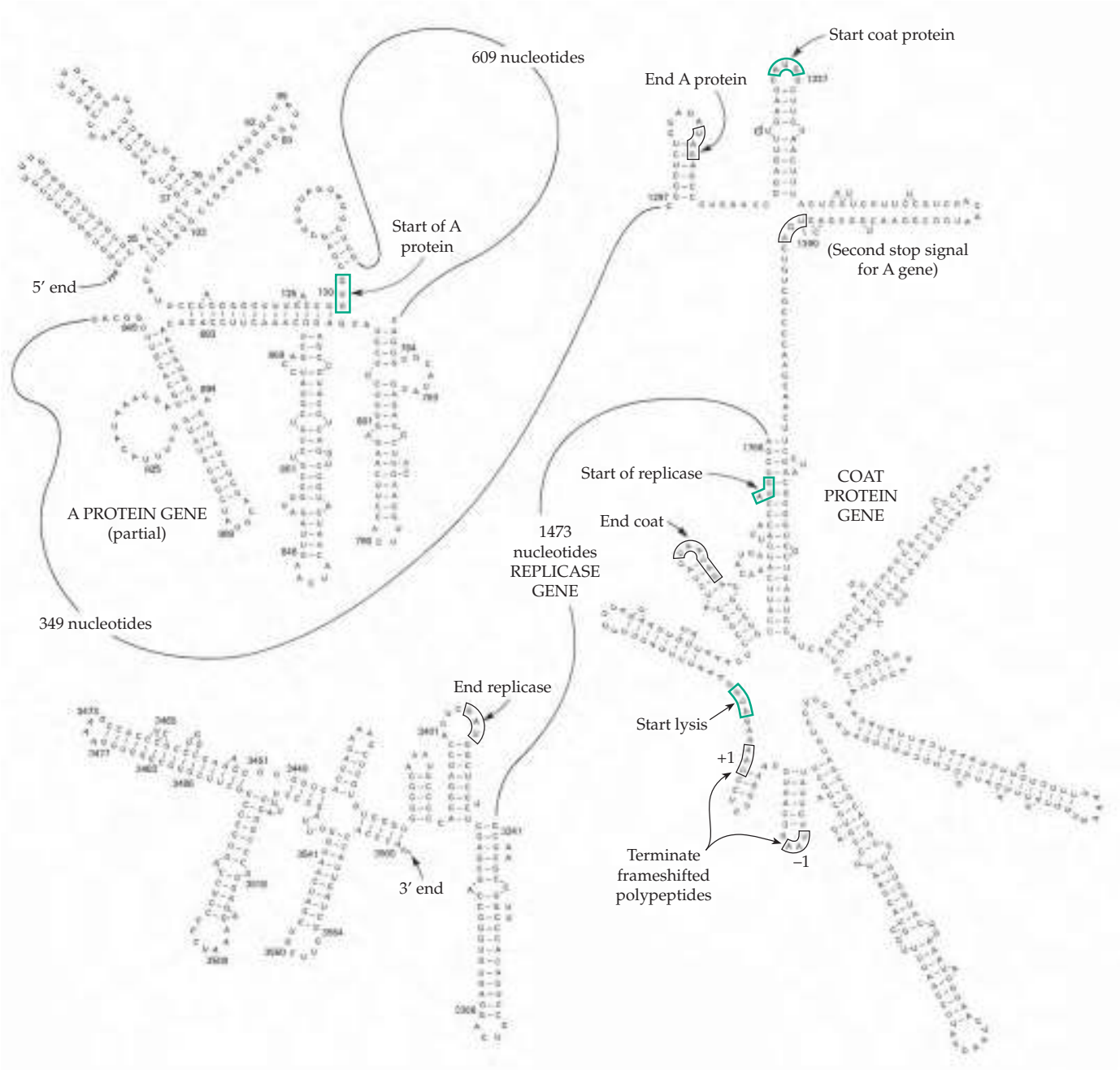


Figure 29-17 Partial sequence and secondary structure model of RNA of bacteriophage MS2. The initiation and termination codons for each of the three genes (A protein, coat protein, and replicase) are enclosed in boxes as is the second stop signal that is in-frame for the A protein gene but out-of-frame for the coat protein gene. The entire coat protein gene is shown but less than one-third of the entire sequence is given. From W. Fiers and associates.[499–501]

(enclosed in a box). Here is some of the first direct evidence ever obtained that GUG as well as AUG is a biologically important initiation codon. Following the initiation codon the nucleotide sequence codes exactly for the established amino acid sequence of the protein. The termination codon UAG is also enclosed in a box in the figure. Following this is a short intergenic region, which includes one side of a hairpin loop with the initiator codon AUG for the next gene at the end. The nucleotide sequence following this codes exactly for the experimentally established sequence end of the coat protein.[499] One other feature of the sequence shown is the UGA termination codon in a box shortly after the beginning of the coat protein gene (at position 1390). This termination signal is out of phase with the initiator codon AUG; hence, it does not represent a termination point for the coat gene. However, it is in phase with the UAG termination codon for the A protein. In the presence of various host amber (UAG) suppressor genes, the A protein is elongated and terminated at this UGA codon.

The coat gene, containing only 390 nucleotides, is shown in its entirety. The secondary structure proposed resembles a flower.[499] The gene ends with a double stop signal UAAUAG. Following an intergenic sequence of 36 nucleotides the long replicase gene starts with an AUG codon. It ends at position 3395 leaving an untranslated segment of 174 nucleotides at the 3' end.

Initially it was thought that MS2 RNA contained only the three genes mentioned in the preceding paragraphs, but later it was found to have an additional gene required for lysis of the host cell.[483] The initiation codon for this gene begins at position 1678 (Fig. 29-17) in the +1 reading frame. There is a UAA stop just two codons before this in the same frame and another UAA stop codon in the −1 frame beginning at position 1652. As a result of these stop codons any reading frameshift during synthesis of the coat protein yields mistranslated proteins that are terminated at these codons. There is no Shine–Dalgarno sequence to bind ribosomes for initiation in this region, but because the initiation codon for the lysis (L) protein is nearby, reinitiation occurs and the L protein is made in the relatively small amounts needed. This arrangement permits efficient use of the RNA by making use of overlapping genes. It also ensures that enough coat protein has been synthesized to make new virus particles before the L protein accumulation causes lysis.[483]

Many viral RNAs that are formed within eukaryotic cells lack a 5' cap. They depend upon internal ribosomal entry sites (IRESs). This has been studied most with picorna viruses.[338,502,503] These viruses not only initiate translation at discrete sites in uncapped RNA but carry out a proteolytic cleavage of initiation factor 4G (Fig. 29-11), which seems to be necessary for initiation of viral-RNA translation.[338,504] The IRES region of hepatitis C viral RNA contains a complex pseudoknotted secondary structure that is necessary for initiation.[346,505,506] Cryo-EM reveals a pronounced change in the 40S ribosomal subunit structure when the viral IRES binds.[346] Some RNA viruses of plants have complex secondary structures in the untranslated 3' region that promote efficient initiation of translation.[506a]

8. Other Functions of Ribosomes

In addition to making proteins, ribosomes also participate in regulatory mechanisms that influence the entire cell. One such mechanism is seen in the **stringent response**.[507–510] Many amino acid-requiring auxotrophs of *E. coli* and other bacteria, when deprived of an essential amino acid, respond by decreasing their production of ribosomal RNA, ribosomal proteins, purine nucleoside triphosphates, lipids, and other essential materials. However, mutations in the gene *rel* (relaxed) lead to continued production of rRNA even in the absence of an essential amino acid. (The stringent response is "relaxed.") It was observed that the **guanosine polyphosphates ppGpp** and **pppGpp**, originally termed MS or "magic spot" compounds, accumulate in stringent (rel^+) strains to a concentration of ~1 mM but not in relaxed (rel^-) strains. Guanosine polyphosphates are synthesized on the ribosomes by transfer of a pyrophospho group from ATP (Eq. 29-11):

$$\text{ATP} + \text{GDP(GTP)} \longrightarrow \text{ppGpp(pppGpp)} + \text{AMP} \qquad (29\text{-}11)$$

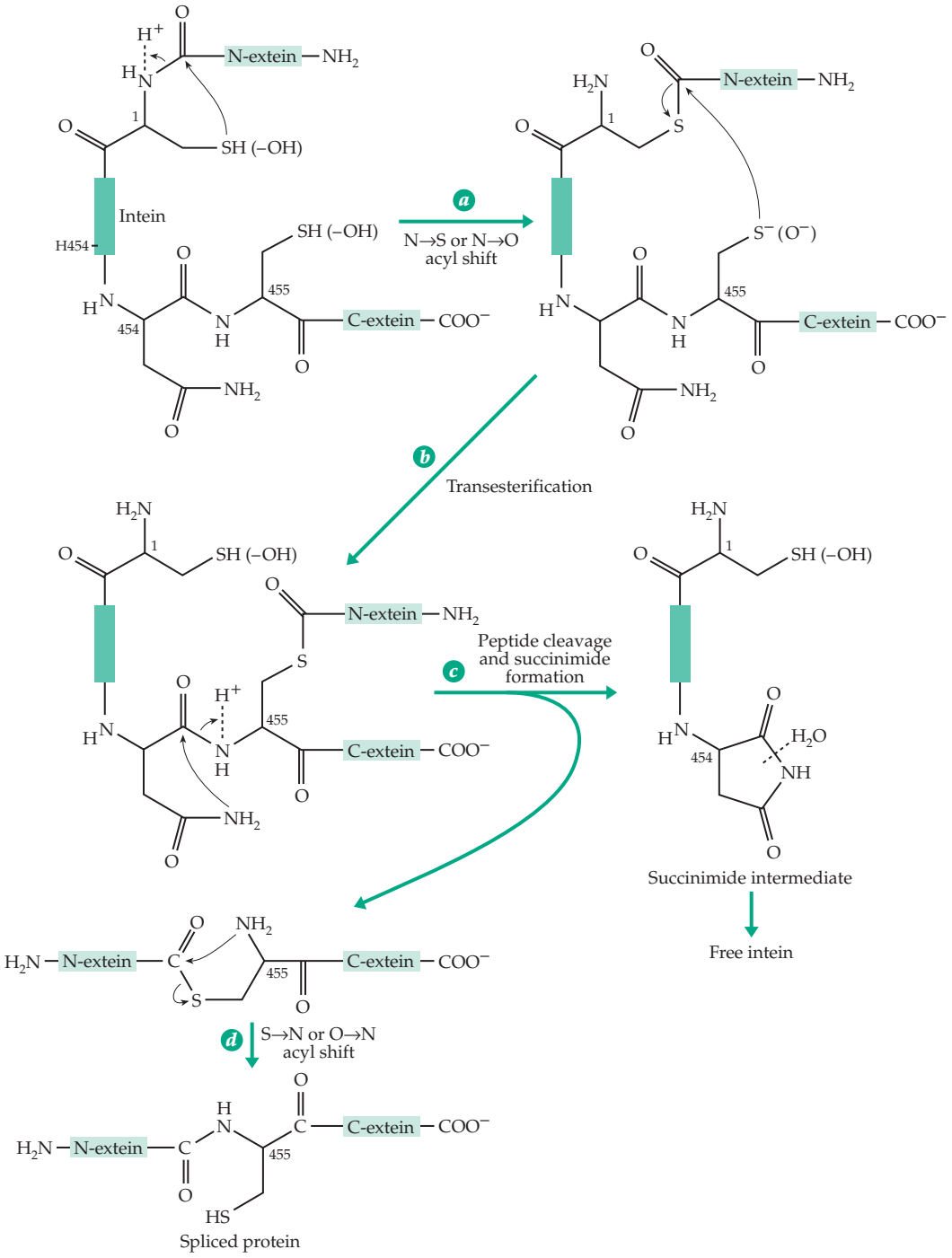
The reaction is catalyzed by the 84-kDa ppGpp synthetase (**stringent factor**), which is encoded by the *rel* gene and is present only in stringent strains.[511–512a] It binds to ribosomes and becomes active only if mRNA is bound to the ribosomes and if codon-selected uncharged tRNA is present in the A sites. A second ppGpp synthetase (PSII) is encoded by gene **spoT**, which also codes for a ppGpp hydrolase.[510] The presence of an uncharged tRNA in the ribosomal A site is expected during amino acid starvation. The stringent factor competes with elongation factor EF-G for its ribosomal site.[513]

The most important effect of accumulating ppGpp may be to bind to an allosteric site on RNA polymerase.[509] The ppGpp-polymerase complex appears to be inefficient in initiating transcription of genes for rRNA, other stable RNAs, and ribosomal proteins. However, it stimulates expression of various amino acid biosynthetic genes and catabolic genes, perhaps via the "discriminator sequence" (Chapter 28; p. 1608). This is not the only effect of ppGpp. The fidelity of translation is decreased when amino acid concentrations fall and

BOX 29-D PROTEIN SPLICING, INTEINS, AND HOMING ENDONUCLEASES

Like self-splicing RNAs, which excise introns from their chains, a few proteins are able to splice out segments of their own chains as **inteins**. The surrounding protein sequences can be referred to as **exteins**. Over 100 self-splicing proteins are known. They are found in all kingdoms of life.[a–d] The inteins, which are excised, are typically 50 kDa in size but range from ~360 to over 500 residues.

The mechanism of splicing is related to the chemistry of pyruvoyl enzyme activation (Eq. 14-41), succinimide formation from asparagine residues (Eq. 2-24), and protein carboxymethylation (Box 12-A). The intein always contains serine or cysteine in its N-terminal (1)-position and asparagine in its C-terminal position. The latter is always followed by cysteine, serine, or threonine in the N-terminal



BOX 29-D (continued)

position of the C-extein. The penultinate residue in the intein is usually (~90 %) histidine, which is thought to play a catalytic role. Other residues in the catalytic domains, which form the ends of the inteins, may also participate in catalysis.

One of the first inteins discovered was found in the 119-kDa precursor to a subunit of a vacuolar ATPase of yeast.[a,c] In this 50-kDa intein Thr 72, His 75, and His 197 may have catalytic functions.[d] The intein is spliced out to form the 69-kDa subunit. The splicing mechanism, which is illustrated for this intein, is shown in the accompanying equations.[b,d–g] Step *a* is an N→S or N→O acyl shift. This is followed by transesterification (step *b*) which involves either thioesters (as illustrated) or oxygen esters. Formation of a succinmide intermediate (step *c*) releases the intein and the spliced protein. The latter must undergo an S→N or O→N acyl shift (step *d*), and the succinimide in the extein must be hydrolyzed to complete the process.

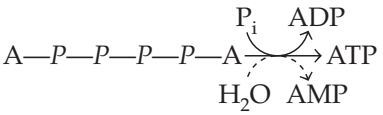
Why do cells ever splice proteins? It isn't clear. However, a curious fact is that many inteins are **homing endonucleases**.[h–k] The genes for these nucleases are often present in introns in mRNA, and the homing endonuclease often cuts DNA in such a way as to initiate movement of its own gene (Chapter 27). The endonuclease itself is found in the center of the intein between the two end domains, which contain the catalytic centers for the splicing reaction.

A few cases are known in which proteins undergo *trans* splicing. For example, the *dnaE* gene of *Synechocystis*, which codes for DNA polymerase III, is actually two partial genes that are 745 kb apart and on opposite strands of the DNA. One of the partial genes codes for a protein containing the N-terminal splice site for an intein, and the other gene codes for a polypeptide containing the C-terminal splice site. Evidently the two splicing domains associate and then catalyze the splicing sequence in the usual way. Split inteins have become very useful in protein engineering because they can be used to join various polypeptide sequences.[k–m] They have also provided an efficient system for purification of specific proteins.[b,n]

[a] Cooper, A. A., and Stevens, T. H. (1995) *Trends Biochem. Sci.* **20**, 351–356
[b] Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F. B., and Xu, M.-Q. (1996) *J. Biol. Chem.* **271**, 22159–22168
[c] Chong, S., Williams, K. S., Wotkowicz, C., and Xu, M.-Q. (1998) *J. Biol. Chem.* **273**, 10567–10577
[d] Poland, B. W., Xu, M.-Q., and Quiocho, F. A. (2000) *J. Biol. Chem.* **275**, 16408–16413
[e] Clarke, N. D. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11084–11088
[f] Xu, M.-Q., and Perler, F. B. (1996) *EMBO J.* **15**, 5146–5153
[g] Shao, Y., Xu, M.-Q., and Paulus, H. (1996) *Biochemistry* **35**, 3810–3815
[h] Chuprina, V. P., Heinemann, U., Nurislamov, A. A., Zielenkiewicz, P., and Dickerson, R. E. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 593–597
[i] Gimble, F. S., and Wang, J. (1996) *J. Mol. Biol.* **263**, 163 – 180
[j] Pietrokovski, S. (1998) *Protein Sci.* **7**, 64–71
[k] Perler, F. B. (1999) *Trends Biochem. Sci.* **24**, 209–211
[l] Martin, D. D., Xu, M.-Q., and Evans, T. C., Jr. (2001) *Biochemistry* **40**, 1393–1402
[m] Otomo, T., Ito, N., Kyogoku, Y., and Yamazaki, T. (1999) *Biochemistry* **38**, 16040–16044
[n] Evans, T. C., Jr., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L., Benner, J., Liu, X.-Q., and Xu, M.-Q. (2000) *J. Biol. Chem.* **275**, 9091–9094

ribosomal action slows. However, ppGpp apparently binds to the ribosome and slows the binding of the aminoacyl-tRNA•EF-Tu•GTP complex. This allows more time for rejection of mispaired tRNAs and increases the fidelity of translation.[514] Under conditions of nutrient starvation the accumulating ppGpp may promote enzymatic degradation of unneeded proteins[512a,b] and may also induce programmed cell death.[515]

Another "alarmone" that regulates both transcription and DNA replication and other cell functions is diadenosine tetraphosphate (Ap_4A). Effects of Ap_4A and related compounds have been discussed in Chapter 28 (p. 1635). These compounds affect many biological events including replication, growth, and differentiation.[516] However, the synthesis of Ap_4A is a reaction not of ribosomes but of an aminoacyl-tRNA synthetase. An enzyme-bound aminoacyl adenylate carries out adenylylation of ATP rather than aminoacylation of tRNA, especially when Zn^{2+} is present. Ap_4A is abundant in blood platelets, where it is stored in dense granules.[517] Both Ap_2A and AP_3A accumulate as gran- ules in myocardial tissues,[518] and Ap_5A and Ap_6A are also present in adrenal chromaffin cells, in blood plate- lets, and in synaptic vesicles.[519] These compounds are catabolized by hydrolases or in lower eukaryotes by phosphorylases. For example, Ap_4A may be converted into ATP + AMP or converted into ATP and ADP (Eq. 29-12).[516,520]



(29-12)

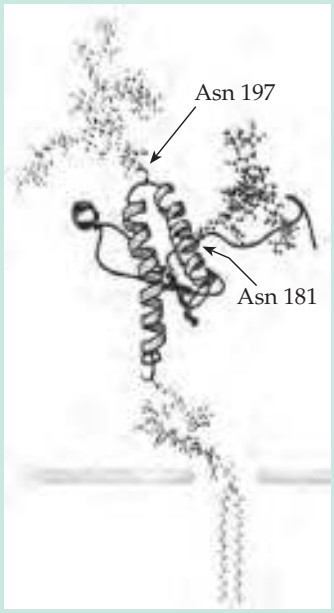
A quite different role of ribosomes is to regulate the life span of certain mRNA molecules. The best studied example is the mRNA for the microtubule

BOX 29-E PRIONS AND AMYLOID DISEASES

The mysterious prions (proteinaceous infective agents), which are described briefly on p. 248, are under intensive investigation. Prion diseases affect fewer than one in 100,000 persons, but there is fear of a possible epidemic. Furthermore, there is a close relationship of prions to a large family of **amyloid diseases**. The most frequent of these is Alzheimer disease, which is estimated to affect one-third of people over 85 years of age in the United States.[a,b]

Prion diseases include **scrapie** of sheep and goats, **bovine spongiform encephalopathy** (**BSE** or mad cow disease), **chronic wasting disease** (**CWD**) of deer and elk, and the human diseases **kuru**, **Creutzfeldt-Jakob disease** (**CJD**), **Gerstmann-Sträussler-Scheinker syndrome** (**GSS**),[c] and **fatal familial insomnia** (**FFI**).[a,d–f] The diseases have a variety of symptoms that include dementia, ataxia (loss of muscular coordination), insomnia, and behavioral problems. All involve some loss of neurons, which may or may not be indicated by a sponge-like appearance of the brain. A characteristic feature of prion disease is the appearance of **amyloid (starch-like) plaques**, which consist of fibrils of insoluble protein.[a] Exhaustive attempts failed to identify a virus particle or an associated DNA or RNA. On this basis, Stanley Prusiner suggested that the diseases are transmitted by pure proteins.[a] All of the diseases seem to involve the same protein, which is known as the **prion protein** (**PrP**). It is encoded by a single-copy gene on human chromosome 20.[g] The amino acid sequence of the C-terminal region of PrP is highly conserved among all animals. However, there are more than 20 known human genetic variants, and a second prion protein has been found in mice.[h] The function of the normal cellular prion protein (**PrP^C**) is unknown, but it appears to be a copper ion carrier, which may be essential to proper synaptic function.[i,j] "Knockout mice" lacking PrP are resistant to prion disease[j,k] but may not be completely healthy.

If it were not for the diseases, PrP^C might be viewed as just another cell surface glycoprotein. Determination of its three-dimensional structure has been difficult, but use of NMR spectroscopy and modeling has given a nearly complete picture, which is shown in the accompanying drawing.[d,l–n] The 250-residue (~220 residues after removal of N- and C-terminal signal sequences) has a long N-terminal tail, a glycosylated globular domain, and a C terminus that is anchored in the outer membrane of neurons by a glycosylphosphatidylinositol (GPI) anchor similar to that shown in Fig. 8-13. The globular domain contains three α helices, a small β sheet, and two glycosylation sites. These last carry



typical N-linked, branched sialic acid-containing oligosaccharides with a total of 52 or more sugar residues. The N-terminal 120 amino acid residue "tail" appears to be largely unstructured. However, it contains five octapeptide repeats with the consensus sequence PHGGGWGQ, each able to bind one Cu^{2+} or Mn^{2+} ion.[d,i,o,p]

How can this ordinarily harmless protein become a killer? The prion is a 20- to 30-kDa hydrophobic particle, which is thought to arise from PrP^C by a conformational alteration in which the α helices are largely changed into a β structure. The new conformer is often designated PrP^{Sc} or PrP-res. The latter abbreviation arises from the fact that native PrP^C can be completely hydrolyzed to small fragments by proteinase K, but PrP^{Sc} contains a 142-residue extremely resistant core (residues 90–231), which is not hydrolyzed and is over 80% β sheet.[q] Evidently the PrP^{Sc} form is able to associate to form a "seed" that, when conditions are favorable, can induce the conformational change in other molecules spreading the PrP^{Sc} form throughout the brain and even into tissues of the immune system.[r] With prion diseases and other amyloid diseases the body may be able to fight off the process by normal proteolytic turnover of the prion protein.

About 85% of all cases of prion disease are **sporadic CJD**. These are thought to arise by spontaneous conversion of PrP^C to PrP^{Sc}. Inherited (familial) forms of CJD, GSS, and FFI are also known. A series of point mutations as well as expansion of the octapeptide repeats[s] account for the various diseases, which have an autosomal dominant inheritance. At least 23 pathogenic mutations have been

BOX 29-E (continued)

reported.[t,u] The point mutations occur at several locations, some of them adjacent to the glycosylation sites.[m] These mutant proteins may be more readily converted to the less soluble PrP^{Sc} type structure, initiating the disease process. However, a mutant with a stop codon (TAG) in the place of the tyrosine 145 codon loses its C-terminal anchor and is degraded rapidly in the proteasomal pathway.[t]

The infectious forms of prion diseases are more puzzling. They account for less than one percent of all cases. Attention was first focused on kuru, a disease of the Fore people of New Guinea. In earlier times they practiced a ritualistic cannibalism of brain tissue that apparently propagated the disease, which is now nearly extinct. Of present concern are over 100 cases of a "new variant" form of CJD, some involving teenage persons and young adults, which have been reported in Europe.[u,v] This disease may have originated in sheep, then jumped to cattle, where it was spread by the ingestion of prion-contaminated meat and bone meal.[a] In addition, more than 120 cases of CJD have arisen from injection of prion-contaminated human growth hormone. Other cases have been traced to contaminated surgical instruments, to tissue grafts, and to use of contaminated human pituitary gonadotrophin.[a]

A hard-to-understand aspect of the "protein-only" theory of prion diseases is the existence of various "strains" of prion proteins. These do not involve differences in amino acid sequence but differences in the conformations of the PrP^{Sc} forms and in the glycosylation patterns.[d,m,w] How can there be several different conformations of the same protein, all of which seed the conversion of normal PrP into differing insoluble forms? In spite of this puzzle, support for the explanation of strain differences comes from a yeast prion system, which involves transcription termination factor eRF3.[x–z] In this system, which involves a prion whose insoluble form can be redissolved by guanidine hydrochloride,[aa] differing strains have also been described.[y,bb,cc] Nevertheless, the presence of the various strains of animal prions, as well as observed vaccination of inbred mice against specific strains,[dd] may be more readily understood if the disease is transmitted by an unidentified virus rather than by a pure protein.[r,u,ee,ff] In fact, the diseases have not been successfully transmitted by truly virus-free proteins synthesized from recombinant DNA.[ee]

What are the prospects for a cure for prion diseases? Several compounds show some effect in slowing accumulation of amyloid plaques,[d,v,gg,hh] but suitable drugs have not been developed. Prevention is the best cure, but more needs to be known about the basic biology of the disease transmission before effective strategies for prevention can be developed.[u]

What is the nature of the insoluble forms of the prion protein? They are hard to study because of the extreme insolubility, but the conversion of α helix to β sheet seems to be fundamental to the process and has been confirmed for the yeast prion by X-ray diffraction.[ii] It has been known since the 1950s that many soluble α-helix-rich proteins can be transformed easily into a fibrillar form in which the polypeptide chains are thought to form a β sheet. The chains are probably folded into hairpin loops that form an antiparallel β sheet (see Fig. 2-11).[jj–ll] For example, by heating at pH 2 insulin can be converted to fibrils, whose polarized infrared spectrum (Fig. 23-3A) indicates a **cross-β structure** with strands lying perpendicular to the fibril axis.[jj,mm] Many other proteins are also able to undergo similar transformation. Most biophysical evidence is consistent with the cross-β structure for the fibrils, which typically have diameters of 7–12 nm.[ii,ll,nn] These may be formed by association of thinner 2 to 5 nm fibrils.[oo] However, β-helical structures have been proposed for some amyloid fibrils[pp] and polyproline II helices for others.[qq]

A wide range of human diseases involving amyloid deposits are known. These include not only the prion diseases and the neurodegenerative diseases, Alzheimer, Parkinson, and the polyglutamine repeat diseases (Table 26-4),[rr,ss] but also **systemic amyloidoses**.[tt] Among the latter are deposits of transthyretin,[uu] the 37-residue **amylin** that develops in the β cells of the pancrease in type II diabetes,[vv] mutant forms of lysozyme,[ww] and of β2 microglobulin,[xx] and gelsolin.[yy] A serum protein amyloid P, a calcium-binding protein, is usually also a component of amyloid deposits.[zz]

[a] Prusiner, S. B. (2001) *N. Engl. J. Med.* **344**, 1516–1526
[b] Manuelidis, L., Fritch, W., and Xi, Y.-G. (1997) *Science* **277**, 94–98
[c] Tagliavini, F., Lievans, P. M.-J., Tranchant, C., Warter, J.-M., Mohr, M., Giaccone, G., Perini, F., Rossi, G., Salmona, M., Piccardo, P., Ghetti, B., Beavis, R. C., Bugiani, O., Frangione, B., and Prelli, F. (2001) *J. Biol. Chem.* **276**, 6009–6015
[d] Caughey, B. (2001) *Trends Biochem. Sci.* **26**, 235–242
[e] Prusiner, S. B., ed. (1999) *Prion Biology and Diseases*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
[f] Prusiner, S. B. (1996) *Trends Biochem. Sci.* **21**, 482–487
[g] Prusiner, S. B. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13363–13383
[h] Moore, R. C., and 20 other authors. (1999) *J. Mol. Biol.* **292**, 797–817

BOX 29-E PRIONS AND AMYLOID DISEASES (continued)

[i] Kramer, M. L., Kratzin, H. D., Schmidt, B., Römer, A., Windl, O., Liemann, S., Hornemann, S., and Kretzschmar, H. (2001) *J. Biol. Chem.* **276**, 16711–16719

[j] Yokoyama, T., Kimura, K. M., Ushiki, Y., Yamada, S., Morooka, A., Nakashiba, T., Sassa, T., and Itohara, S. (2001) *J. Biol. Chem.* **276**, 11265–11271

[k] Weissmann, C., and Aguzzi, A. (1999) *Science* **286**, 914–915

[l] Zahn, R., Liu, A., Lührs, T., Riek, R., von Schroetter, C., Garcia, F. L., Billeter, M., Calzolai, L., Wider, G., and Wüthrich, K. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 145–150

[m] Rudd, P. M., Wormald, M. R., Wing, D. R., Prusiner, S. B., and Dwek, R. A. (2001) *Biochemistry* **40**, 3759–3766

[n] Viles, J. H., Donne, D., Kroon, G., Prusiner, S. B., Cohen, F. E., Dyson, H. J., and Wright, P. E. (2001) *Biochemistry* **40**, 2743–2753

[o] Prince, R. C., and Gunson, D. E. (1998) *Trends Biochem. Sci.* **23**, 197–198

[p] Brown, D. R., Hafiz, F., Glasssmith, L. L., Wong, B.-S., Jones, I. M., Clive, C., and Haswell, S. J. (2000) *EMBO J.* **19**, 1180–1186

[q] Baskakov, I. V., Aagaard, C., Mehlhorn, I., Wille, H., Groth, D., Baldwin, M. A., Prusiner, S. B., and Cohen, F. E. (2000) *Biochemistry* **39**, 2792–2804

[r] Balter, M. (1999) *Science* **286**, 660–662

[s] Narwa, R., and Harris, D. A. (1999) *Biochemistry* **38**, 8770–8777

[t] Zanusso, G., Petersen, R. B., Jin, T., Jing, Y., Kanoush, R., Ferrari, S., Gambetti, P., and Singh, N. (1999) *J. Biol. Chem.* **274**, 23396–23404

[u] Almond, J., and Pattison, J. (1997) *Nature (London)* **389**, 437–438

[v] Thompson, C. (2001) *Nature (London)* **409**, 660–661

[w] Hill, A. F., Desbruslais, M., Joiner, S., Sidle, K. C. L., Gowland, I., Collinge, J., Doey, L. J., and Lantos, P. (1997) *Nature (London)* **389**, 448–450

[x] Patino, M. M., Liu, J.-J., Glover, J. R., and Lindquist, S. (1996) *Science* **273**, 622–626

[y] Liebman, S. W., and Derkatch, I. L. (1999) *J. Biol. Chem.* **274**, 1181–1184

[z] Zhou, P., Derkatch, I. L., Uptain, S. M., Patino, M. M., Lindquist, S., and Liebman, S. W. (1999) *EMBO J.* **18**, 1182–1191

[aa] Wickner, R. B., Edskes, H. K., Maddelein, M.-L., Taylor, K. L., and Moriyama, H. (1999) *J. Biol. Chem.* **274**, 555–558

[bb] Sparrer, H. E., Santoso, A., Szoka, F. C., Jr., and Weissman, J. S. (2000) *Science* **289**, 595–599

[cc] Chien, P., and Weissman, J. S. (2001) *Nature (London)* **410**, 223–227

[dd] Manuelidis, L. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2520–2525

[ee] Manuelidis, L., Sklaviadis, T., Akowitz, A., and Fritch, W. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5124–5128

[ff] Manuelidis, L. (1997) *Annales De L'Institut Pasteur* **8**, 311–326

[gg] Tagliavini, F., and 20 other authors. (1997) *Science* **276**, 1119–1122

[hh] Supattapone, S., Nguyen, H.-O. B., Cohen, F. E., Prusiner, S. B., and Scott, M. R. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14529–14534

[ii] Balbirnie, M., Grothe, R., and Eisenberg, D. S. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2375–2380

[jj] Burke, M. J., and Rougvie, M. A. (1972) *Biochemistry* **11**, 2435–2439

[kk] Bradbury, E. M., Brown, L., Downie, A. R., Elliott, A., Fraser, R. D. B., Hanby, W. E., and Macdonald, T. R. R. (1960) *J. Mol. Biol.* **2**, 276–286

[ll] Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., and Blake, C. C. F. (1997) *J. Mol. Biol.* **273**, 729–739

[mm] Bouchard, M., Zurdo, J., Nettleton, E. J., Dobson, C. M., and Robinson, C. V. (2000) *Protein Sci.* **9**, 1960–1967

[nn] Hörnberg, A., Eneqvist, T., Olofsson, A., Lundgren, E., and Sauer-Eriksson, A. E. (2000) *J. Mol. Biol.* **302**, 649–669

[oo] Chiti, F., Taddei, N., Bucciantini, M., White, P., Ramponi, G., and Dobson, C. M. (2000) *EMBO J.* **19**, 1441–1449

[pp] Lazo, N. D., and Downing, D. T. (1998) *Biochemistry* **37**, 1731–1736

[qq] Blanch, E. W., Morozova-Roche, L. A., Cochran, D. A. E., Doig, A. J., Hecht, L., and Barron, L. D. (2000) *J. Mol. Biol.* **301**, 553–563

[rr] Koo, E. H., Lansbury, P. T., Jr., and Kelly, J. W. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9989–9990

[ss] Kaytor, M. D., and Warren, S. T. (1999) *J. Biol. Chem.* **274**, 37507–37510

[tt] Hardy, J., and Gwinn-Hardy, K. (1998) *Science* **282**, 1075–1079

[uu] Sebastiao, M. P., Lamzin, V., Saraiva, M. J., and Damas, A. M. (2001) *J. Mol. Biol.* **306**, 733–744

[vv] Moriarty, D. F., and Raleigh, D. P. (1999) *Biochemistry* **38**, 1811–1818

[ww] Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C. F., and Pepys, M. B. (1997) *Nature (London)* **385**, 787–793

[xx] Chiti, F., Mangione, P., Andreola, A., Giorgetti, S., Stefani, M., Dobson, C. M., Bellotti, V., and Taddei, N. (2001) *J. Mol. Biol.* **307**, 379–391

[yy] Robinson, R. C., Choe, S., and Burtnick, L. D. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2117–2118

[zz] Lambowitz, A. M., and Perlman, P. S. (1990) *Trends Biochem. Sci.* **15**, 440–444

proteins tubulin.[521–523] Accumulating α and β tubulin subunits act in a feedback loop to induce the degradation of their mRNA. They do this by binding to the N-terminal sequence MREI of β tubulin as it is emerging from ribosomes. This binding allosterically activates an endonuclease that cuts the polysome-bound mRNA. A similar mechanism for tubulin mRNA may involve the MREC N-terminal sequence of that protein.

An unexpected finding was that **phosphatidylserine synthase** of *E. coli* is tightly bound to ribosomes.[524] This enzyme, which incorporates serine into phospholipids according to step *h* of Fig. 21-4, is responsible for synthesis of the principal membrane lipid of *E. coli*. Its localization on ribosomes may be linked to the joint regulation of the synthesis of proteins and lipids.

D. Processing, Secretion, and Turnover of Proteins

The concepts of processing and turnover of proteins have been introduced in Chapter 10, and many details have been presented in other chapters. However, as we complete our discussion of protein synthesis, it is appropriate to discuss processing further. As

polypeptide chains leave the ribosomes via the exit channels, they may follow several different paths. They may enter the cytosol and fold quickly into a compact form. This may require only a few seconds, whereas the translation process in the ribosome may take many seconds. The folding will therefore be **cotranslational**.[525] Depending upon the N-terminal signal peptide the protein may later unfold and pass through a membrane pore or **translocon** into the endoplasmic reticulum (ER), a mitochondrion, chloroplast, or peroxisome. Wherever it is, it will be crowded together with thousands of other proteins. It will interact with many of these, and evolution will have enabled some of these to become chaperones (discussed in Chapter 10).[526]

A single rapidly growing cell of *E. coli* may contain ~2.3 x 10^6 soluble polypeptide chains of ~2600 different types with an average length of ~317 residues and average mass of ~35 kDa.[527] These are chaperoned in a variety of ways,[528,529] not only in the cytoplasm but in the periplasm (see p. 364).[530,531] The three chaperones **trigger factor** (TF), DnaK, and GroEL participate in folding newly formed proteins.[525a,b] TF is a prolyl-*cis*–*trans* isomerase (Box 9-F), which associates with the large ribosomal subunit with a 1:1 stoichiometry.[525b–d] DnaK and related chaperones hold and protect newly formed polypeptides in extended conformations, while the GroES-GroEL chaperonins assist folding within their internal cavities (Box 7-A).[525b,e–g] **Prefoldin**, a 90-kDa complex,[525b] has a special function in chaperoning microtubule subunits. A Type II chaperonin also assists the folding of actin and tubulin.[525h] Chaperones assist not only in the folding of proteins but also in translocation into the ER (e.g., by the Hsp 70 homolog BiP), and into mitochondria and other organelles.[532,533] **Co-chaperones** are additional proteins that act as selective agents to direct proteins to a particular chaperone. For example, the DnaJ protein is a scanning factor for the Hsp 70 chaperone DnaK. As is described on p. 518, it catalyzes ATP-dependent association of a substrate with the binding cavity of DnaK.[533] A chaperone whose function has long remained elusive is an abundant secreted glycoprotein known as **clusterin.**[534] It seems to have a protective function in protection against stress.

How are the possible choices for newly formed proteins made? Much seems to depend upon the amino acid sequences at the ends of the polypeptide chains. As they emerge from a ribosome, some N-terminal signal sequences bind to recognition proteins. One such protein labels the ends of proteins destined for secretion into the vesicles of the ER. This protein ensures that the protein end binds to the **signal recognition particle** (SRP), enters a translocon pore, and undergoes cotranslational passage into the periplasmic space in bacteria or the ER in eukaryotes. Cotranslational modification reactions also occur both in the cytosol and in the ER vesicles. These too influence the choice of destinations as do additional signal sequences in the polypeptide chains. Proteins may be directed to the various organelles, to residence in membranes,[535] or to secretion into the external medium. It was somewhat surprising to discover that under some circumstances most newly synthesized proteins are degraded in proteasomes.[536] Cotranslational degradation of proteins with imperfect ends may account for some of this.[537–539] In addition, imperfect proteins that are retained in the ER may be sent back into the cytoplasm for degradation in proteasomes.[540]

1. Cotranslational and Posttranslational Processing

The modifications that lead to the presence of fully functional proteins in their proper locations begin while peptide chains are still emerging from the ribosomes.[541] In bacteria and in eukaryotic organelles the first of these modifications is hydrolytic removal of the *N*-formyl group by an Fe^{2+}-dependent **deformylase** leaving the N-terminal methionine.[542–543] Deformylases are present in eukaryotes as well as in bacteria, making the deformylase a less attractive target for antibiotic design than has sometimes been proposed.[544,544a] When the chain is only 20–30 residues long, the terminal methionine that remains after deformylation may be removed by a ribosome-associated methionine aminopeptidase.[545] The methionine is usually removed if followed by P, G, A, S, or T and is retained if followed by K, R, L, I, F, or N. With other amino acids removal is variable.[546] A ribosome-bound *N*-acetyltransferase may acetylate the N terminus either before or after removal of Met.[547] Approximate rules for eukaryotic cells are[541]:

N-terminal D, E, N	Acetylation without removal of Met
N-terminal P, V, C	Removal of Met; no acetylation
N-terminal G, A, S, T	Removal of Met followed by acetylation
Other N termini	No modification

An example is provided by actin, which contains acetyl-Met-Asp, acetyl-Met-Gln or acetyl-Met-Cys-Asp at the N terminus immediately after synthesis. Then, within ~15 min the acetyl-Met is cleaved off, and the next terminal residue is acetylated.[548] *N*-Acylation of nascent peptides by fatty acyl groups can also occur cotranslationally. For example, 14-carbon myristoyl groups are added in amide linkage to the N-terminal glycines of many cellular and virally encoded proteins.[535,549,550] This may take place on the ribosomes,[551]

but it is often not clear whether the modification is cotranslational or posttranslational. The same may be said of many other "posttranslational" alterations, many of which may begin on a nascent polypeptide chain. Fatty acyl groups (mainly palmitoyl) may form thioester linkages with cysteine side chains.[552] This often occurs near the C terminus (see p. 559). Other C-terminal modifications include prenylation (p. 559)[553] and attachment to diacylglycerols via thioester linkages to cysteine (pp. 402, 428)[553a] or to glycosylphosphatidylinositol glycan anchors (Fig. 8-13; Eq. 20-23).

The addition of an N-terminal myristoyl group to a protein causes a relatively permanent alteration as does methylation of histidine, lysine, or arginine side chains.[554,555] So do hydroxylation, vitamin K-dependent carboxylation (Eq. 15-55), and many other alterations. In contrast, glycosylation, phosphorylation, and sulfation produce reversible alterations. Sometimes, as in the conversion of proenzymes to active enzymes, a modification step is used to generate a catalytic activity. In other instances, as in the processing of glycoproteins in the Golgi, the major function of the modification reaction seems to be one of directing a protein to the correct intracellular location.

2. Forming Functional Proteins

Proenzymes and other precursor proteins are often almost totally inactive until they are activated by some alteration that occurs when they reach their destination in a cell or in the body. Cleavage of the polypeptide chains of proenzymes, covalent attachment of coenzymes,[556,557] oxidation (Eq. 16-57) or halogenation (Eq. 25-6) of tyrosine or tryptophan[558] side chains, and oxidation of cysteine in the sequence CTPSR to formylglycine (in sulfatase formation; Eq. 12-44)[559] are only a few of many modifications needed to form functional proteins. Sometimes, as in activation of chymotrypsinogen, a single simple modification creates the active protein. In other cases modification may be quite complex. For example, although many polypeptide antibiotics are formed by nonribosomal synthesis (Box 29-C), some are created on ribosomes and may require extensive subsequent alteration. An example is **microcin**, a 69-residue peptide antibiotic formed by some strains of *E. coli*. Eight Gly-Cys and Gly-Ser pairs in a pre-microcin chain are cyclized to thiazole and oxazole rings. Then the 69-residue antibiotic is cut out from the precursor and secreted into the medium.[560,561] The 22-residue antibiotic **epidermin** is one of a family of **lantibiotics** that contain lanthionine as a characteristic component. Biosynthesis involves dehydration of serine and threonine residues, sulfide (thioether) bridge formation, oxidative decarboxylation, and removal of a leader peptide.[562]

3. Translocation through Membranes

The processes by which proteins are selected for secretion into the periplasmic space of bacteria or into the vesicles of the endoplasmic reticulum of eukaryotic cells are similar and have been discussed in Chapter 10 (pp. 519–521). However, some details are still being worked out. The first step in translocation is binding of the N terminus of a protein that is emerging from a ribosome to the **signal recognition particle**.[563,564] The core of this particle has a universally conserved structure consisting of two proteins and an RNA molecule.[565–571b]

Bacteria	Eukaryotes
4.5S RNA (~114 nt)	7S RNA (~295 nt)
Protein Ffh	Protein SRP54
Protein Ftsy (SRP receptor)[572]	Protein SRα

All of these proteins are GTPases, and in eukaryotes SRα is associated with a third protein SRβ, which is also a GTPase.[573] Either protein Ffh or SRP54 recognizes the N terminus of the protein that is to be translocated, chaperoning it to the receptor Ftsy[574] or SRα, where it may be anchored to the translocating pore (**translocon**). Hydrolysis of GTP by both proteins accompanies the recognition process. The eukaryotic SRP is more complex, containing six proteins and a larger RNA than in bacteria.[568,575–576] One domain of the 7S RNA is homologous to the bacterial 4.5S RNA, while an additional domain is closely related in its sequence to that of the highly repetitive *Alu* sequences in DNA (Fig. 27-9). However, the significance of this similarity is unclear.

In eukaryotic cells binding of SRP54 induces a transient retardation of translocation, an **elongation arrest**, while the SRP complex binds to its receptor SRα. This 70-kDa peripheral membrane protein is tightly associated with the 30-kDa integral membrane protein SRβ. Binding to this receptor leads the nascent polypeptide chain from the ribosome directly into the Sec61 translocon,[33f,576a] which consists of α, β, and γ subunits and a central aqueous pore. The ribosome apparently also becomes bonded firmly to the translocon until synthesis of the polypeptide chain has been completed.[573,577,578] The translocon complex also contains additional components[563,577,579] including the **leader peptidase** (signal peptidase, p. 620)[563,580–582a] and the **oligosaccharyltransferase** of ER membranes (Eq. 20-21).[563,583] The latter transfers an oligosaccharide from a lipid carrier onto certain asparagine side chains of polypeptides entering the ER. This and other glycosylation reactions help to keep the polypeptide moving to its correct destination, whether it be in some membrane surface, a lysosome or other organelle, or a secretion vesicle (Chapter 20). Furthermore,

some proteins are translocated by a mechanism that doesn't depend upon SRP but utilizes a different complex, which consists in yeast of proteins Sec62, Sec63, Sec71, Sec72, and Kar2 (the chaperone BiP present in the ER lumen).[577]

Translocation of most bacterial proteins occurs posttranslationally rather than cotranslationally.[584] After recognition by SRP the polypeptide chains are transferred to chaperone complexes.[585] Some proteins are escorted to the folding compartment of the GroEL-GroES chaperonin (Box 7-A). Those that are to be secreted are often chaperoned by the protein **SecB**.[586] Genetic analysis shows that for *E. coli* the secretion of many proteins requires the products of genes *secA*, *secB*, *secD*, *secE*, *secF*, *secG*, and *secY*. Gene *secA* encodes a 92-kDa cytoplasmic ATPase protein (**SecA**), SecB is a 64-kDa homotetramic chaperone that prevents folding of preproteins prior to export, and SecY is a 42-kDa integral membrane protein.[587,588]

A complex of the three transmembrane proteins **SecYEG** forms the translocation channel in *E. coli* membranes. **SecY** and **SecE** are essential components in most bacteria and are homologous to components of the eukaryotic **Sec61p** translocon complex.[589] From their sequences SecY, SecE, and SecG are predicted to have ten, three, and two transmembrane sequences, respectively. Additional accessory proteins in the complex are designated SecD, SecF, and yajC. Their functions are uncertain, and they are not essential for transport. The driving force for translocation is provided by the peripheral ATPase Sec A.[588,590] For many proteins the signal sequence, which is usually positively charged, stays on the negatively charged cytoplasmic surface of the membrane, while SecA in an ATP-dependent process pushes a loop of protein through the membrane. Hydrolysis of a second ATP molecule seems to be required to release SecA, allowing it to reload with ATP and to assist the next 20–30 residue polypeptide sequence to be translocated.[591,592] Surprisingly, SecA is also an ATP-dependent RNA helicase.[590] The significance of its apparent ability to translocate along either an RNA or a polypeptide chain is unclear. The protonmotive force provided by the membrane potential is another important factor in the translocation of many proteins.[591,593–596] Yet another factor is the lipid composition of the membrane. Non-bilayer lipids seem to be required for efficient transport.[597]

As is mentioned in Chapter 10, bacteria have additional mechanisms of polypeptide transport. A recently recognized Sec-independent pathway is used by *E. coli* and many other bacteria to secrete proteins that contain the twin-arginine motif RRX$\phi\phi$, where ϕ is a hydrophobic amino acid, in their N-terminal signal sequences. Proteins encoded by genes *tatABC* are required by this **Tat pathway**.[594,598,599] Related pathways have been identified both in mitochondria and in chloroplasts. Small peptides may pass out of the periplasmic space through the porins in the outer membrane of gram-negative bacteria. However, larger proteins require conduit molecules such as the **TolC** channel-tunnel, which directly connects an inner membrane translocon with a channel in the outer membrane of *E. coli* cells.[600,601]

Eukaryotic cells also have additional transport mechanisms. One of these is an ABC transporter (p. 417) known as the **transporter associated with antigen processing** (**TAP**). It carries small polypeptides generated by proteasomes from the cytosol into the ER for export and binding to **MHC Class I** molecules and subsequent presentation to the immune system (Fig. 31-15).[540,602]

4. Translocation into Organelles

Most of the proteins of mitochondria are encoded in nuclear DNA and are synthesized on cytoplasmic ribosomes. Mitochondria do not utilize proteins homologous to those of the bacterial Sec system but have their own set of transport proteins.[603–605] These proteins, which include an outer membrane complex (Tom) and an inner membrane complex (Tim), are discussed in Chapter 18 (see Fig. 18-4). Perhaps these specialized mitochondrial proteins are needed because transport into the mitochondrial matrix is in an opposite direction to the transport out through bacterial membranes.

The transport of proteins into chloroplasts also occurs by more than one mechanism. An SRP-dependent pathway may be needed only for insertion of proteins into membranes.[594] Other proteins, among which are the 23-kDa and 16-kDa photosystem II proteins (Chapter 23), enter by a pathway related to the Tat pathway of bacteria. In thylakoids this pathway is directly dependent upon the large pH difference (Δ pH) across the thylakoid membrane. In contrast to the bacterial Sec pathway, the Δ pH pathway seems to be able to transport completely folded proteins.

Proteins destined for peroxisomes have their own targeting signals. One of these (**PTS1**) is the sequence SKL at the C terminus. A second signal (**PTS2**) is an N-terminal nonapeptide (R / K)(L / V / I) X_5(H / Q) (L / A).[606,607]

5. Membrane Proteins

Some proteins enter membranes immediately after synthesis. The translocon channel is not required. However, in *E. coli* an additional protein **YidC** is needed.[603] Homologs of this protein are found in mitochondria (**Oxa1** protein) and in thylakoid membranes of chloroplasts (**Alb3** protein).[608] These proteins may function in cotranslational insertion. If a protein carries a

positively charged N-terminal region, it will tend to stick to the negatively charged cytoplasmic surface of a cell membrane. This "positive inside" rule (p. 401)[609] is strong for bacterial proteins but somewhat weaker for eukaryotic cells. A second topological rule is that hydrophobic segments of proteins will be attracted to membrane surfaces and can enter the membrane (perhaps via translocon pores) as loops (Fig. 29-18).[596] Passage of the loop out through the membrane will be facilitated if negatively charged groups are present in the loop and are acted upon by the membrane potential (pp. 401,402).[595,610,611] If the entire polypeptide chain follows the loop out through the membrane, the protein will be anchored to the inside of the cell membrane with its C terminus outside. On the other hand, if the C terminus also has a positively charged cluster nearby, a membrane associated **leader peptidase** (or signal peptidase) may cut the loop past the signal sequence on the outside of the membrane leaving the bulk of the protein with its N terminus outside (Fig. 29-18B). How are **polytopic** integral membrane proteins with multiple cytoplasmic and external loops formed? Hydrophobic signal sequences are not always at the N terminus of a polypeptide chain. Suitable **internal signal sequences** may be found in the sequences that form the transmembrane helices, e.g., those present in the many 7-helix receptors found in a membrane. This suggests the possibility that successive loops may be translocated. If the N terminus is allowed to pass through a translocon in one of the steps, the topology of Fig. 11-6 or of Fig. 23-41 will result.[611]

Genetic methods have also been applied to study the insertion of coat subunits of phage M13 into the plasma membrane of *E. coli*.[612–614] The subunits are stored in the plasma membrane waiting to form a cylindrical shell about a viral DNA molecule as it is extruded from the bacterium.[612] The rod-like subunits (Fig. 7-7) have their N termini in the periplasmic space and their C termini in the cytoplasm. Each end carries a cluster of electrically charged residues, mostly negative at the N terminus and positive at the C terminus. Insertion into the membrane occurs only if the membrane has its normal membrane potential with a positive external surface charge and a negative internal charge, complementary to the charges on the coat subunit. Insertion does not occur unless the leader peptide with its positively charged N-terminal cluster and the C-terminal positive cluster are both present.[610,615] This suggests insertion by the loop mechanism of Fig. 29-18A. Genetic studies of the *E. coli* leader peptidase revealed that this protein also has an internal signal sequence, which becomes inserted into the membrane and which is not cleaved.[612,616,617] The final orientation of the mature enzyme is indicated in Fig. 29-18C.

Targeting of proteins to specialized domains of a membrane are less well understood. These include caveolae and lipid rafts, domains that are high in cholesterol and sphingolipids and which function in endocytosis and in cell signaling. A recent proposal is that proteins with hydrophobic surfaces needed in these domains become coated with a lipid "shell" before entering the membrane.[617a]

6. Secretion of Proteins

Cells continuously secrete materials via small cytoplasmic vesicles, which in eukaryotes arise largely from the Golgi apparatus (pp. 425–427; Fig. 20-8). The vesicles of this **constitutive pathway** may have diameters of ~50 nm. They carry phospholipids, proteins, and other constituents for incorporation into the plasma membrane of the cell.[618,619] In addition, there are

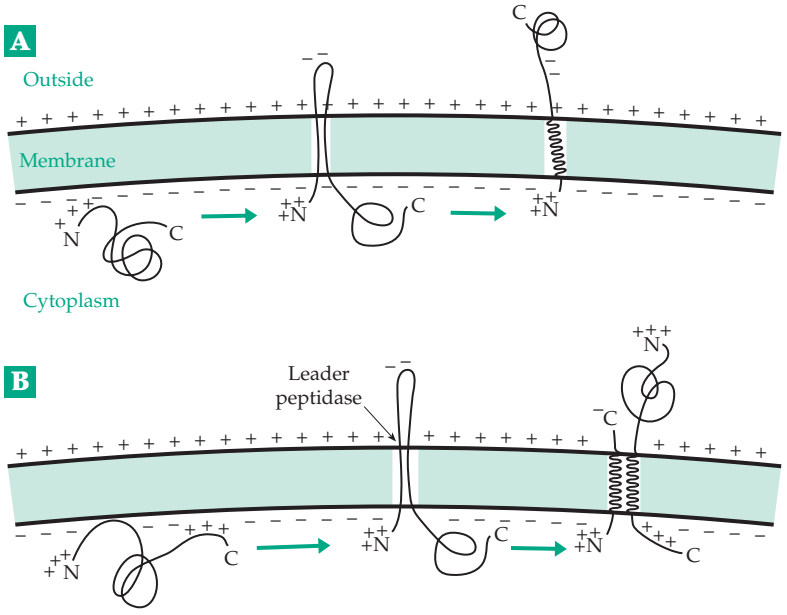


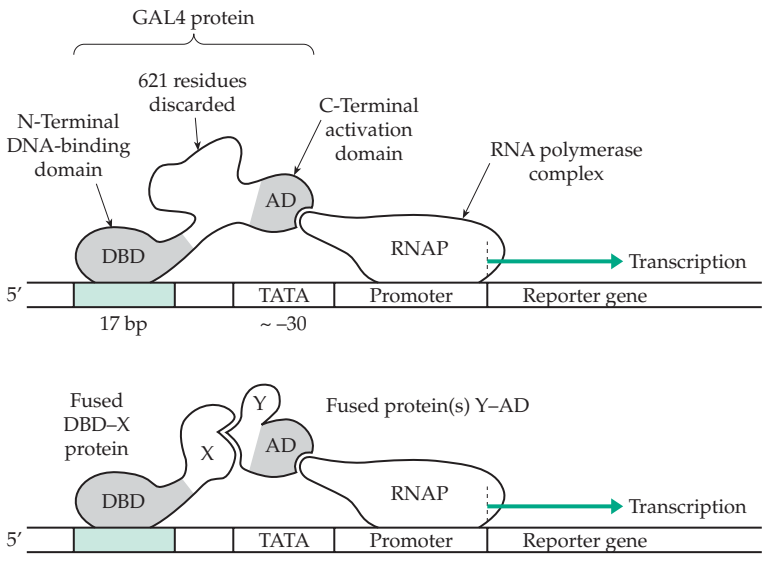
Figure 29-18 (A) Proposed mechanism for insertion of a loop of polypeptide chain through a translocon pore in a membrane with the positively charged N terminus anchored to the negatively charged inner membrane surface. (B) Cleavage of a polypeptide loop formed as in (A) by a leader peptidase to give a polypeptide chain anchored by a positively charged cluster near its C terminus. (C) Membrane topology of the *E. coli* leader peptidase. The active site is in the periplasmic domain. See Tschantz *et al.*[580]

BOX 29-F THE YEAST TWO-HYBRID SYSTEM FOR IDENTIFYING PROTEIN–PROTEIN INTERACTIONS

Many techniques including ultracentrifugation, chemical crosslinking, and X-ray crystallography are used to identify interactions between proteins. However, for study of the entire proteome new approaches are needed. One new technique that has already been widely applied is the yeast two-hybrid system.[a–f] In its original form[a,b] a transcriptional activator is utilized together with a reporter gene, e.g., a fused *GAL1– lac2* gene that when transcribed yields β-galactosidase. This enzyme then cleaves a chromogenic substrate (see p. 1494) to give a blue color. The 881-residue transcription factor GAL4 (p. 1630) is often used in two-hybrid systems. It binds upstream of the TATA sequence in the promoter regions of genes that code for enzymes of galactose catabolism. The N-terminal **DNA-binding domain** of *GAL4* binds to a specific 17-bp palindromic sequence in the DNA, while the acidic C-terminal **activation domain** activates transcription by interacting with the RNA polymerase complex bound to the promoter (see figure). The GAL4 activator seems to be quite flexible and is able to activate transcription, even if the distance from its binding site on DNA to the transcription initiation site is varied considerably. The two-hybrid system was constructed by cloning separately the pieces of DNA that code for residues 1–147 of the DNA-binding domain of the GAL4 protein (DBD in the figure) and for residues 768–881 of the activation domain (AD). The intervening nucleotides coding for the remaining 620 residues of GAL4 are discarded. The two-hybrid system tests whether a protein X, sometimes called the **bait**, binds or otherwise interacts strongly with another protein (Y, often called the **prey**) or with a series of other proteins (Y_1, $Y_2 \cdots Y_n$). To carry out the test the gene for protein X is fused with that for the DNA-binding domain of GAL4. When expressed in a living yeast cell the hybrid protein DBD–X will be formed. The gene fusion must be in-frame to ensure a correct structure for the X portion. Likewise, genes for protein Y, or for a series Y_1, $Y_2 \cdots Y_n$, are fused in-frame to the gene fragment carrying the GAL4 activation domain. Y–AD hybrids will be formed. The test is made using a strain of yeast in which the GAL4 gene has been replaced with a hybrid *GAL1– lacZ* reporter gene. If both DBD–X and Y–AD are present, and if they interact strongly (bind tightly), transcription of the reporter gene will be activated, and a blue colony will grow from the yeast cell. It is useful to create the hybrids with protein X, or proteins X_1 $X_2 \cdots$ (the baits) in a haploid strain of one of the two yeast mating types *MAT*a or *MAT*α (pp. 20, 1574) and the proteins Y (the prey) in a strain of the other mating type. Mating of the two strains will produce diploid cells that express both the DBD–X and Y–AD hybrids.

Many variants of the two-hybrid system have been devised.[d] For example, a green fluorescent protein reporter can be used.[g] Because significant biological protein–protein interactions often require that three or more proteins interact,[d] hybrid systems involving more than two proteins have been developed. Two-hybrid systems for bacteria have also been devised.[h] A virtue of the two-hybrid methods is that they work with undenatured, if not totally natural, proteins. This is in contrast to widely used methods that involve separation of denatured proteins on gels or columns.

The most popular two-hybrid systems utilize microarrays.[d,f] In the simplest approach hybrid DBD–X is tested against a library of Y–ADs



Top: The yeast GAL4 protein interacts with the RNA polymerase complex to activate transcription of a suitable reporter gene. Bottom: Two hybrid proteins, one containing the DNA-binding domain of GAL4, fused to protein X and the other containing protein Y fused to the activation domain of GAL4, are present in a cell. If X and Y bind strongly to each other, activation domain AD will be held close to the RNA polymerase and will activate transcription.

BOX 29-F THE YEAST TWO-HYBRID SYSTEM FOR IDENTIFYING PROTEIN–PROTEIN INTERACTIONS (continued)

prepared by random cleavage of DNA of known sequence. For example, the entire genome of the gastric pathogenic bacterium *Helicobacter pylori* was cut into ~1000 nucleotide pieces. These were cloned into plasmids in *E. coli*, then into yeast. Over 10 million *E. coli* clones provided a final two million independent yeast colonies, which carried the Y–ADs (prey). The genomic DNA fragments were also used to prepare 285 DBD–X (baits) from 261 genes. In a series of two-hybrid screening tests more than 1200 different protein–protein interactions connecting 47% of the proteome were detected.[i]

The complete yeast (*S. cerevisiae*) has been probed using at least two large-scale two-hybrid investigations. Uetz *et al.*[j] generated a large set of ~6000 genetically engineered yeast colonies, each one expressing just one of the possible Y–AD hybrid proteins (prey) derived from the ~6000 gene products identified in the yeast genome. These strains were distributed into microtiter plates and were individually crossed with 192 strains of yeast, each of which expressed a single BDB–X hybrid. This simple automated array screening identified 281 interacting pairs. In a second approach, the cells producing the Y–AD prey hybrids were mixed to give a single library. This was then screened against nearly all of the possible DBD–X hybrids in a large-scale automated procedure. The two approaches together detected 957 probable interactions involving 1004 different proteins.[j] In an independent study, using similar approaches but different cloning vehicles, Ito *et al.*[k,l] identified 4549 two-hybrid interactions among 3278 proteins. Of these 841 interactions were judged to be most relevant (core). Surprisingly, only 135 were identical to those found by Uetz *et al.* The significance and possible reasons for this disparity have been discussed.[f,j,l]

Interpretation of results of these studies is still difficult. Results of two-hybrid methods become more useful if they can be coordinated with other approaches. For example, computational methods can predict interactions from genome sequences alone.[m,n,o] More than 45,000 interactions have been predicted among yeast proteins.[m] Reliable identification of such motifs as DNA-binding domains and Ca^{2+}- binding domains can complement two-hybrid analysis.[n] The yeast genome is predicted to contain 162 coiled-coil sequences and at least 213 unique interactions between them.[o] Examination of sequences of protein families in the Protein Data Bank (PDB) led to prediction of 8151 interactions of 664 types between protein families in yeast.[p]

Improved experimental procedures of other types can also complement two-hybrid methods. Among these are formaldehyde crosslinking with immunoprecipitation,[q] methods that couple mass spectrometry and crosslinking,[r] and detection of intermolecular nuclear Overhauser enhancements in protein–protein complexes.[s] Phage display methods (see Fig. 3-16) have been developed as another method of detecting protein–protein interactions[e] as has fluorescence resonance energy transfer (FRET; p. 1291).[d] Evanescent wave methods, e.g., surface plasmon resonance (Box 3-F) are increasingly being used to quantify protein–protein interactions. These may be combined with single-hybrid methods in inexpensive and rapid micro-devices.[d]

[a] Fields, S., and Song, O.-K. (1989) *Nature (London)* **340**, 245–246
[b] Chien, C.-t, Bartel, P. L., Sternglanz, R., and Fields, S. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9578–9582
[c] Finley, R. L., Jr., and Brent, R. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12980–12984
[d] Mendelsohn, A. R., and Brent, R. (1999) *Science* **284**, 1948–1950
[e] Allen, J. B., Walberg, M. W., Edwards, M. C., and Elledge, S. J. (1995) *Trends Biochem. Sci.* **20**, 511–516
[f] Oliver, S. (2000) *Nature (London)* **403**, 601–603
[g] Shioda, T., Andriole, S., Yahata, T., and Isselbacher, K. J. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5220–5224
[h] Joung, J. K., Ramm, E. I., and Pabo, C. O. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7382–7387
[i] Rain, J.-C., and 12 other authors. (2001) *Nature (London)* **409**, 211–215
[j] Uetz, P., and 19 other authors. (2000) *Nature (London)* **403**, 623–627
[k] Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T., Nishizawa, M., Yamamoto, K., Kuhara, S., and Sakaki, Y. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1143–1147
[l] Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569–4574
[m] Marcotte, E. M., Pellegrini, M., Ng, H.-L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999) *Science* **285**, 751–753
[n] Gallet, X., Charloteaux, B., Thomas, A., and Brasseur, R. (2000) *J. Mol. Biol.* **302**, 917–926
[o] Newman, J. R. S., Wolf, E., and Kim, P. S. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13203–13208
[p] Park, J., Lappe, M., and Teichmann, S. A. (2001) *J. Mol. Biol.* **307**, 929–938
[q] Orlando, V. (2000) *Trends Biochem. Sci.* **25**, 99–104
[r] Bennett, K. L., Kussmann, M., Björk, P., Godzwon, M., Mikkelsen, M., Sorensen, P., and Roepstorff, P. (2000) *Protein Sci.* **9**, 1503–1518
[s] Clore, G. M. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9021–9025

regulated pathways for storage and release by exocytosis of hormones, neuropeptides, and neurotransmitters. The last are secreted from both small (~50 nm diam.) synaptic vesicles and larger **dense core vesicles** (>100 nm diam.),[620,621] which are discussed in Chapter 30, Section B,8. In every case specialized proteins (discussed on pp. 427 and 521 and in Chapters 20 and 30) are involved.[622–624]

7. Protein Folding

Before a protein can function its polypeptide chain must fold into its own native tertiary structure.[624a] This folding is influenced by many surrounding proteins, by the state of glycosylation of side chains,[625] and by other posttranslational modifications, by the presence of cis amide linkages in unfolded or folded forms (pp. 82, 83; Box 9-F),[626] and by possibility for formation of disulfide bridges (pp. 521, 522).[626a–628] The prediction of the folding pattern of proteins from the amino acid sequence remains a major goal of protein chemistry. In principle, a protein fold can be predicted from the DNA sequence of the genes, with proper allowance for effects of posttranscriptional modification. This goal once seemed intractable, but two things have provided new hope. (1) The speed and power of computers is still increasing. This not only has allowed more rapid calculations but also has led to improvement in experimental methods. (2) Methods for studying folding, which include mass spectroscopy,[629] NMR spectroscopy,[630–633] and optical methods,[634,635] have become more rapid and more sensitive.[636] As a consequence, we have an abundance of new data.

Anfinsen, in 1963, proposed that the three-dimensional structure of a protein is in its lowest Gibbs energy state when present in its natural environment.[637,638] However, there is a problem with this suggestion (the "Levinthal paradox"; p. 82). Even if a polypeptide chain occupies only two of the lowest energy regions of the Ramachandran diagram (Fig. 2-9), a 100-residue protein would have ~10^{30} possible conformations. If a folding protein checked all of these conformations at a realistic rate of ~10^{11} s^{-1}, it would take ~10^{11} years to fold. Furthermore, in a test tube of protein, which would contain at most 10^{18} molecules, each of the molecules would probably have a different conformation.[639] In fact, most proteins fold reliably to the same final structure in less than a second, and some in a millisecond.[639–643] It is also true that proteins, under altered solvent conditions, can misfold into totally "incorrect" structures.[644] Most can assume an amyloid structure under some conditions (Box 29-E). One clear conclusion is that folding is not totally random but follows a **folding pathway,** which is dictated by the sequence. Nevertheless, experimental data indicate that there is an ensemble of related structures at each stage of folding.[631]

We cannot answer the question posed by Anfinsen's hypothesis. Does the native state have a minimum value of the Gibbs energy? Nevertheless, it is observed that proteins usually behave as if folded, unfolded forms are in a true thermodynamic equilibrium, and that this equilibrium is attained rapidly. The difference ΔG between a folded and a denatured protein is only 21–63 kJ mol^{-1}, which shows that folded proteins are only marginally more stable than are unfolded polypeptide chains.[645] The value of ΔG of unfolding as a function of temperature T is given by Eq. 29-13, where $\Delta H(T)$ and ΔC_p are the changes in enthalpy and heat capacity upon unfolding.[645,646]

$$\Delta G\,(T) = \Delta H_m\,(1-T/T_m) - \Delta C_p\,[(T_m-T) - T\,\ln(T/T_m)] \qquad (29\text{-}13)$$

ΔH_m is the enthalpy change at T_m, the midpoint of the thermal unfolding curve (the "melting" temperature). The temperature of maximum stability T_s occurs when $\Delta S = 0$ (Eq. 29-14). T_s is usually between –10°C and

$$T_s = T_m \exp(-\Delta H_m / [T_m \Delta C_p]) \qquad (29\text{-}14)$$

35°C. For staphylococcal nuclease (Fig. 12-29) T_s = 18°C and T_m = –19°C and +57°C, i.e., *the protein is denatured by either cooling below 18°C or heating above 57°C*, a behavior that is common for many proteins. Cold denaturation is observed whenever the unfolded state has a higher heat capacity than does the folded state.[647]

Can we predict the Gibbs energy of unfolding from the protein sequence? To do this it is necessary to utilize experimental data on known proteins to obtain a series of terms (Eq. 29-15) that can be summed to give the $\Delta G°_u$, the standard Gibbs energy of unfolding.

$$\Delta G°_u = \Delta G°_{charge} + \Delta G°_{hyd} + \Delta G°_{conf} + \Delta G°_{vW} + \Delta G°_{Hbond} \qquad (29\text{-}15)$$

The terms refer to summations of all of the charge–charge, hydrophobic, configurational, van der Waals, and hydrogen-bonding interactions between both main-chain and side-chain atoms.[638,646,648] Such computations are formidable and are uncertain, especially for electrostatic (charge–charge and hydrogen-bonding) interactions.[649–652] Both the folded and denatured state must be considered,[653] as must the heat capacities[654] and configurational entropies.[655]

While we tend to think of proteins as having fixed structures, conformational changes are basic to life. Many proteins are very flexible and in part disordered.[655a] At the same time proteins can be misfolded leading to amyloid formation (Box 29-E) and to other diseases.[655b–e]

8. Completing the Cycle: Proteolytic Degradation of Proteins

Like all other body constituents proteins must undergo breakdown as well as synthesis. Regular turnover of all proteins is essential, and defects in the process may lead to amyloid deposits (Box 29-E). Turnover and degradation of proteins depends upon a variety of proteases, many of which are discussed in Chapter 12, Section C. Because of the changing needs of body cells, specific proteins turn over at widely varying rates. Some enzymes, hormones, and regulatory proteins have half-lives of only a few minutes while others may function for months or years. How can the regulation of the breakdown of thousands of different proteins be controlled? The answer seems to lie in the amino acid sequence. Just as the sequence determines the location that a protein occupies in the cell and its folding pattern, it also determines the turnover rate.

Much of the breakdown takes place in the cytosol in proteasomes (Box 7-A) and is controlled by the ubiquitin system (Box 10-C), which selects the proteins for degradation; control of the system is quite complex.[656–661] One aspect depends upon the N-terminal amino acid of the substrate protein. Defective proteins often have N-terminal destabilizing amino acids such as phenylalanine, leucine, aspartic acid, lysine, and arginine (p. 527). If an internal lysine is also present, the protein may be conjugated with ubiquitin and degraded rapidly. Many metabolic processes, such as the cell cycle (Fig. 11-15), are controlled by protein degradation.[538,662] In some cases an arginine residue is transferred onto the N terminus of a protein by an **arginyl-tRNA protein transferase**. This creates a better substrate for ubiquitination and rapid degradation.[280,663] In other cases proteolytic cleavage uncovers an arginine or other destabilizing residue and speeds hydrolysis.[538] While ubiquitination often initiates the degradation of proteins, it also helps to direct proteins to specific locations within a cell.[663a] Rapid degradation of a ubiquitinated protein may require hydrolytic **deubiquitination** by a metalloprotease, which is a subunit of the 26S proteasome lid (Box 7-A). This allows the ubiquitin to be recycled and also directs the deubiquitinated protein into the proteasome.[663b]

As mentioned on p. 1854, an important function of proteasomes is formation of short antigenic peptides for use by the immune system.[664] Inhibition of proteasome activity reduces or prevents antigen presentation (Chapter 31).[665,666] In this immune surveillance system mature proteins of host cells are cut up and checked for self-identity. The checking also includes the rapidly degraded imperfect proteins and foreign proteins from invading organisms or viruses.[667,667a]

Lysosomes, which contain more than 50 proteases, lipases, glycosidases, and other hydrolases, also play a major role in protein degradation.[668] Their importance is emphasized by the range of lysosomal deficiency diseases (Table 20-1).[668] Lysosomes also function in the process of **autophagy**, by which cells can sacrifice a whole section, organelles and all, by walling off a large vesicle or **autophagosome** and fusing it with a lysosome.[669] In such a way a tadpole can resorb its tail while becoming a frog. We have now come full circle: our proteins have been converted back to the amino acids and other small molecules derived from them. The amino acids can be reutilized or can be catabolized, depending upon the needs of the organism.

E. Proteomics

The vast amount of data on protein structures and improved methods of predicting structures[670–672] have led to development of new areas of science variously designated as genomics, proteomics, transcriptomics,[673,674] and bioinformatics.[675–677] These fields encompass all of the methods for sequence determination, observation of gene expression, protein synthesis by cells, and mathematical analysis of resulting data. Proteomics includes new approaches to polypeptide separation[678] and identification,[678,679] sequencing at the attomole level,[680–681] and comparison of sequences between species.[682–684] Protein separation by liquid chromatography,[685] capillary electrophoresis,[686] or two-dimensional gel electrophoresis[687,688] can be followed by mass spectrometry of intact proteins or of proteolytic fragments.[689,690] Microarrays on proteome chips can be used to observe production of thousands of proteins simultaneously.[691–694] Structural genomics centers have been established for rapid determination of protein structures using NMR[695,696] and X-ray methods.[697–699] If each center determines 200 or more structures per year there will soon be 16,000 new structures, enough to allow us to predict much about all the rest.[699] Then we can study all the important remaining details for millions and millions of proteins.

Current efforts to understand the structures, conformational movements, and functions of these molecules range from the classification of nearly 10,000 different protein folds[700] to investigation of the dynamics of single protein molecules.[701] Well known motifs such as β sheets and α helices are studied with the goal of more accurate predictions of structure and better understanding of interactions between proteins in solution and in membranes. For example, one natural topology is the β barrel, which may contain 8 to 22 strands (e.g., see Fig. 8-20). These cylindrical proteins are abundant in outer membranes of gram-negative bacteria.[702–704] The partial electrical changes at the edges of the β sheets (see Fig. 2-11,B) may interact to help stabilize the barrels. In contrast, soluble β barrel-containing proteins are designed to avoid edge-to-

edge interactions, which could cause aggregation of the proteins.[704] A recently discovered membrane-protein motif is an α **barrel**, which is composed of 12 α helices stacked side-by-side with side-chain groups fitting together in a knobs-in-holes fashion.[705] An example is the TolC protein of *E. coli*. A trimer of 428-residue subunits forms a long cylinder, which is a 12-stranded β barrel at one end and an α barrel at the other.[705] More common are transmembrane α helices, many of which are present in 7-helix receptors (e.g., Fig. 11-6). Relatively accurate prediction methods are now available for these structures,[706,707] but there are still uncertainties about mechanisms of transmission of signals across the membranes.

Predictions of structures of more complex proteins from their amino acid sequences presents a major challenge.[708] Assignment of domains within the protein is a first step.[709–712] Regions of probable helix or β-strand structure can be recognized but it is difficult to predict the exact lengths of the helices and the structures of connecting loops and strands. These depend upon many factors including the possible formation of ion pairs[713] and of locks at the ends of strands created by van der Waals interactions.[714] There are also circular proteins.[715] Composite structures such as those of silks (Box 2-B) have surprising properties. Both silkworm,[716,717] and spider silks[718,719] undergo marked changes in properties upon spinning of the random coil forms of the proteins found in silk glands into the drawn fibers.

Whether we discuss silk, proteins embedded in membranes, or soluble complexes of cytosolic proteins, we must ask questions about interactions. A first step is to identify interactions[720–730] among proteins either *in vitro* or in living cells.[731] Proteomic methods, which include the yeast two-hybrid method (Box 29-F), are widely used for this purpose. It is possible to identify large sets of interacting proteins, to identify disease states, to observe effects of drugs, and to compare metabolism among species.

References

1. Nomura, M. (1990) *Trends Biochem. Sci.* **15**, 244–247
2. Lake, J. A. (1981) *Sci. Am.* **245**(Aug), 84–97
3. Schachman, H. D., Pardee, A. B., and Stanier, R. Y. (1952) *Arch. Biochem. Biophys.* **38**, 245–260
4. Wittmann, H. G. (1983) *Ann. Rev. Biochem.* **52**, 35–65
5. Prince, J. B., Gutell, R. R., and Garrett, R. A. (1983) *Trends Biochem. Sci.* **8**, 359–363
6. Yonath, A., Leonard, K. R., and Wittmann, H. G. (1987) *Science* **236**, 813–816
7. Lake, J. A. (1985) *Ann. Rev. Biochem.* **54**, 507–530
8. Frank, J., Verschoor, A., Wagenknecht, T., Radermacher, M., and Carazo, J.-M. (1988) *Trends Biochem. Sci.* **13**, 123–127
9. Baumeister, W., and Steven, A. C. (2000) *Trends Biochem. Sci.* **25**, 624–631
10. Morgan, D. G., Ménétret, J.-F., Radermacher, M., Neuhof, A., Akey, I. V., Rapoport, T. A., and Akey, C. W. (2000) *J. Mol. Biol.* **301**, 301–321
11. De Chadarevian, S. (1999) *Trends Biochem. Sci.* **24**, 203–206
12. Nagano, K., and Harel, M. (1987) *Prog. Biophys. and Mol. Biol.* **48**, 67–101
13. Lake, J. A. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1903–1907
14. Lake, J. A., and Strycharz, W. A. (1981) *J. Mol. Biol.* **153**, 979–992
15. Kahan, L., Winkelmann, D. A., and Lake, J. A. (1981) *J. Mol. Biol.* **145**, 193–214
16. Ramakrishnan, V., Capel, M., Kjeldgaard, M., Engelman, D. M., and Moore, P. B. (1984) *J. Mol. Biol.* **174**, 265–284
17. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science* **289**, 905–920
18. Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., and Noller, H. F. (1999) *Science* **285**, 2095–2104
19. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) *Science* **289**, 920–930
20. Agrawal, R. K., Penczek, P., Grassucci, R. A., Burkhardt, N., Nierhaus, K. H., and Frank, J. (1999) *J. Biol. Chem.* **274**, 8723–8729
21. Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., and Zamir, A. (1965) *Science* **147**, 1462–1465
22. Wittmann, H. G. (1982) *Ann. Rev. Biochem.* **51**, 155–183

22a. Brownlee, G. G., Sanger, F., and Burell, B. G. (1967) *Nature (London)* **215**, 735–736

22b. Brosius, J., Palmer, M. L., Kennedy, P. J., and Noller, H. F. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4801–4805

22c. Brosius, J., Dull, T. J., and Noller, H. F. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 201–204

23. Leijonmarck, M., Eriksson, S., and Liljas, A. (1980) *Nature (London)* **286**, 824–826
24. Ramakrishnan, V., and White, S. W. (1998) *Trends Biochem. Sci.* **23**, 208–212
25. Yonath, A. (1984) *Trends Biochem. Sci.* **9**, 227–230
26. Yonath, A., and Wittmann, H. G. (1989) *Trends Biochem. Sci.* **14**, 329–335
27. Yonath, A., Glotz, C., Gewitz, H. S., Bartels, K. S., von Böhlen, K., Makowski, I., and Wittman, H. G. (1988) *J. Mol. Biol.* **203**, 831–834
28. Moore, P. B. (2001) *Biochemistry* **40**, 3243–3250
29. Schluenzen, F., Tociij, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. (2000) *Cell* **102**, 615–623
30. Culver, G. M., Cate, J. H., Yusupova, G. Z., Yusupov, M. M., and Noller, H. F. (1999) *Science* **285**, 2133–2135
31. Agalarov, S. C., Prasad, G. S., Funke, P. M., Stout, C. D., and Williamson, J. R. (2000) *Science* **288**, 107–112

31a. Recht, M. I., and Williamson, J. R. (2001) *J. Mol. Biol.* **313**, 35–48

32. Pennisi, E. (1999) *Science* **285**, 2048–2051
33. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) *Nature (London)* **407**, 327–339

33a. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. D., and Noller, H. F. (2001) *Science* **292**, 883–896

33b. Ramakrishnan, V., and Moore, P. B. (2001) *Curr. Opin. in Struct.Biol.* **11**, 144–154

33c. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4899–4903

33d. Ogle, J. M., Brodersen, D. E., Clemons, W. M. J., Tarry, M. J., Carter, A. P., and Ramakrishnan, V. (2001) *Science* **292**, 897–902

33e. Mathews, M. B., and Pe'ery, T. (2001) *Trends Biochem. Sci.* **26**, 585–587

33f. Moore, P. B., and Steitz, T. A. (2002) *Nature (London)* **418**, 229–235

33g. Pioletti, M., Schlünzen, F., Harms, J., Zarivach, R., Glühmann, M., Avila, H., Bashan, A., Bartels, H., Auerbach, T., Jacobi, C., Hartsch, T., Yonath, A., and Franceschi, F. (2001) *EMBO J.* **20**, 1829–1839

34. Malhotra, A., Penczek, P., Agrawal, R. K., Gabashvili, I. S., Grassucci, R. A., Jünemann, R., Burkhardt, N., Nierhaus, K. H., and Frank, J. (1998) *J. Mol. Biol.* **280**, 103–116
35. Liljas, A. (1999) *Science* **285**, 2077–2078
36. Dube, P., Bacher, G., Stark, H., Mueller, F., Zemlin, F., van Heel, M., and Brimacombe, R. (1998) *J. Mol. Biol.* **279**, 403–421
37. Gomez-Lorenzo, M. G., Spahn, C. M. T., Agrawal, R. K., Grassucci, R. A., Penczek, P., Chakraburtty, K., Ballesta, J. P. G., Lavandera, J. L., Garcia-Bustos, J. F., and Frank, J. (2000) *EMBO J.* **19**, 2710–2718

37a. Mueller, F., Sommer, I., Baranov, P., Matadeen, R., Stoldt, M., Wöhnert, J., Görlach, M., van Heel, M., and Brimacombe, R. (2000) *J. Mol. Biol.* **298**, 35–59

38. Brimacombe, R. (1984) *Trends Biochem. Sci.* **9**, 273–277
39. Noller, H. F. (1984) *Ann. Rev. Biochem.* **53**, 119–162

39a. Gutell, R. R., Noller, H. F., and Woese, C. R. (1986) *EMBO J.* **5**, 1111–1113

39b. Pace, N. R., Olsen, G. J., and Woese, C. R. (1986) *Cell* **45**, 325–326

References

40. Dragon, F., Payant, C., and Brakier-Gingras, L. (1994) *J. Mol. Biol.* **244**, 74–85
41. Lee, K., Varma, S., SantaLucia, J., Jr., and Cunningham, P. R. (1997) *J. Mol. Biol.* **269**, 732–743
42. Capel, M. S., Kjeldgaard, M., Engelman, D. M., and Moore, P. B. (1988) *J. Mol. Biol.* **200**, 65–87
43. Moore, P. B. (1988) *Nature (London)* **331**, 223–227
44. Svergun, D. I., Koch, M. H. J., Pedersen, J. S., and Serdyuk, I. N. (1994) *J. Mol. Biol.* **240**, 78–86
45. Wadzack, J., Burkhardt, N., Jünemann, R., Diedrich, G., Nierhaus, K. H., Frank, J., Penczek, P., Meerwinck, W., Schmitt, M., Willumeit, R., and Stuhrmann, H. B. (1997) *J. Mol. Biol.* **266**, 343–356
46. Willumeit, R., Forthmann, S., Beckmann, J., Diedrich, G., Ratering, R., Stuhrmann, H. B., and Nierhaus, K. H. (2001) *J. Mol. Biol.* **305**, 167–177
47. Svergun, D. I., and Nierhaus, K. H. (2000) *J. Biol. Chem.* **275**, 14432–14439
48. Herwig, S., Kruft, V., Eckart, K., and Wittmann-Liebold, B. (1993) *J. Biol. Chem.* **268**, 4643–4650
49. Urlaub, H., Kruft, V., Bischof, O., Müller, E.-C., and Wittmann-Liebold, B. (1995) *EMBO J.* **14**, 4578–4588
50. Wang, R., Alexander, R. W., VanLoock, M., Vladimirov, S., Bukhtiyarov, Y., Harvey, S. C., and Cooperman, B. S. (1999) *J. Mol. Biol.* **286**, 521–540
51. Culver, G. M., Heilek, G. M., and Noller, H. F. (1999) *J. Mol. Biol.* **286**, 355–364
52. Wower, J., Kirillov, S. V., Wower, I. K., Guven, S., Hixson, S. S., and Zimmermann, R. A. (2000) *J. Biol. Chem.* **275**, 37887–37894
53. Shapkina, T. G., Dolan, M. A., Babin, P., and Wollenzien, P. (2000) *J. Mol. Biol.* **299**, 615–628
54. Juzumiene, D. I., Shapkina, T. G., and Wollenzien, P. (1995) *J. Biol. Chem.* **270**, 12794–12800
55. Montesano-Roditis, L., Glitz, D. G., Perrault, A. R., and Cooperman, B. S. (1997) *J. Biol. Chem.* **272**, 8695–8703
56. Laughrea, M., and Tam, J. (1992) *Biochemistry* **31**, 12035–12041
57. Hüttenhofer, A., and Noller, H. F. (1994) *EMBO J.* **13**, 3892–3901
58. Czworkowski, J., Odom, O. W., and Hardesty, B. (1991) *Biochemistry* **30**, 4821–4830
59. Urlaub, H., Thiede, B., Müller, E.-C., Brimacombe, R., and Wittmann-Liebold, B. (1997) *J. Biol. Chem.* **272**, 14547–14555
60. Benjamin, D. R., Robinson, C. V., Hendrick, J. P., Hartl, F. U., and Dobson, C. M. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7391–7395
61. Glück, A., and Wool, I. G. (1996) *J. Mol. Biol.* **256**, 838–848

61a. Zamecnik, P. (1984) *Trends Biochem. Sci.* **9**, 464–466

62. Schimmel, P., and Wang, C.-C. (1999) *Trends Biochem. Sci.* **24**, 127–128
63. Freifelder, D. (1983) *Molecular Biology*, 2nd ed., Jones and Bartlett, Boston, Massachusetts (p. 419)
64. Yamaguchi, K., and Subramanian, A. R. (2000) *J. Biol. Chem.* **275**, 28466–28482
65. O'Brien, T. W., Fiesler, S. E., Denslow, N. D., Thiede, B., Wittmann-Liebold, B., Mougey, E. B., Sylvester, J. E., and Graack, H.-R. (1999) *J. Biol. Chem.* **274**, 36043–36051
66. Koc, E. C., Burkhart, W., Blackburn, K., Koc, H., Moseley, A., and Spremulli, L. L. (2001) *Protein Sci.* **10**, 471–481

66a. Koc, E. C., Burkhart, W., Blackburn, K., Moyer, M. B., Schlatzer, D. M., Moseley, A., and Spremulli, L. L. (2001) *J. Biol. Chem.* **276**, 43958–43969

67. Moazed, D., Stern, S., and Noller, H. F. (1986) *J. Mol. Biol.* **187**, 399–416
68. Choi, Y. C. (1985) *J. Biol. Chem.* **260**, 12769–12772
69. Rairkar, A., Rubino, H. M., and Lockard, R. E. (1988) *Biochemistry* **27**, 582–592

69a. Decatur, W. A., and Fournier, M. J. (2002) *Trends Biochem. Sci.* **27**, 344–351

70. Carter, A. P., Clemons, W. M., Jr., Brodersen, D. E., Morgan-Warren, R. J., Hartsch, T., Wimberly, B. T., and Ramakrishnan, V. (2001) *Science* **291**, 498–501
71. Leontis, N. B., and Westhof, E. (1998) *J. Mol. Biol.* **283**, 571–583
72. Schnare, M. N., Damberger, S. H., Gray, M. W., and Gutell, R. R. (1996) *J. Mol. Biol.* **256**, 701–719
73. Herzog, M., and Maroteaux, L. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8644–8648

73a. Gutell, R. R. (1994) *Nucleic Acids Res.* **22**, 3502–3507

74. Brimacombe, R., Atmadja, J., Stiege, W., and Schüler, D. (1988) *J. Mol. Biol.* **199**, 115–136
75. Malhotra, A., and Harvey, S. C. (1994) *J. Mol. Biol.* **240**, 308–340
76. Stern, S., Weiser, B., and Noller, H. F. (1988) *J. Biochem. Biophys. Methods* **204**, 447–481
77. Mueller, F., and Brimacombe, R. (1997) *J. Mol. Biol.* **271**, 524–544

77a. Lodmell, J. S., and Dahlberg, A. E. (1997) *Science* **277**, 1262–1267

78. Merryman, C., Moazed, D., Daubresse, G., and Noller, H. F. (1999) *J. Mol. Biol.* **285**, 107–113
79. Larsson, S. L., and Nygård, O. (2001) *Biochemistry* **40**, 3222–3231

79a. Matadeen, R., Sergiev, P., Leonov, A., Pape, T., van der Sluis, E., Mueller, F., Osswald, M., von Knoblauch, K., Brimacombe, R., Bogdanov, A., van Heel, M., and Dontsova, O. (2001) *J. Mol. Biol.* **307**, 1341–1349

80. Garrett, R. (1983) *Trends Biochem. Sci.* **8**, 189–190
81. Porse, B. T., Cundliffe, E., and Garrett, R. A. (1999) *J. Mol. Biol.* **287**, 33–45
82. Conn, G. L., Draper, D. E., Lattman, E. E., and Gittis, A. G. (1999) *Science* **284**, 1171–1174
83. Correll, C. C., Wool, I. G., and Munishkin, A. (1999) *J. Mol. Biol.* **292**, 275–287
84. Macbeth, M. R., and Wool, I. G. (1999) *J. Mol. Biol.* **285**, 965–975
85. Chan, Y.-L., Sitikov, A. S., and Wool, I. G. (2000) *J. Mol. Biol.* **298**, 795–805
86. Wilson, K. S., and Noller, H. F. (1998) *Cell* **92**, 337–349
87. Klein, B. K., Staden, A., and Schlessinger, D. (1985) *J. Biol. Chem.* **260**, 8114–8120
88. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. D., and Noller, H. F. (2001) *Science* **1**, 1–20
89. Khaitovich, P., and Mankin, A. S. (1999) *J. Mol. Biol.* **291**, 1025–1034
90. Christensen, A., Mathiesen, M., Peattie, D., and Garrett, R. A. (1985) *Biochemistry* **24**, 2284–2291
91. Romby, P., Westhof, E., Toukifimpa, R., Mache, R., Ebel, J.-P., Ehresmann, C., and Ehresmann, B. (1988) *Biochemistry* **27**, 4721–4730
92. Leontis, N. B., and Moore, P. B. (1986) *Biochemistry* **25**, 3916–3925
93. Dontsova, O., Tishkov, V., Dokudovskaya, S., Bogdanov, A., Döring, T., Rinke-Appel, J., Thamm, S., Greuer, B., and Brimacombe, R. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4125–4129
94. Lee, K. M., and Marshall, A. G. (1986) *Biochemistry* **25**, 8245–8252
95. Funari, S. S., Rapp, G., Perbandt, M., Dierks, K., Vallazza, M., Betzel, C., Erdmann, V. A., and Svergun, D. I. (2000) *J. Biol. Chem.* **275**, 31283–31288
96. Garrett, R. A. (1985) *Nature (London)* **318**, 233–235
97. Olsen, G. J., and Woese, C. R. (1993) *FASEB J.* **7**, 113–123
98. Dijk, J., and Littlechild, J. A. (1979) *Methods Enzymol.* **59**, 481–502
99. Kruft, V., and Wittmann-Liebold, B. (1991) *Biochemistry* **30**, 11781–11787
100. Bielka, H., ed. (1982) *The Eukaryotic Ribosome*, Springer, Berlin

100a. Wool, I. G., Chan, Y. L., and Gluck, A. (1995) *Biochem. Cell Biol.* **73**, 933–947

101. van Beekvelt, C. A., Kooi, E. A., de Graaff-Vincent, M., van't Riet, J., Venema, J., and Raué, H. A. (2000) *J. Mol. Biol.* **296**, 7–17
102. Goldschmidt-Reisin, S., Kitakawa, M., Herfurth, E., Wittmann-Liebold, B., Grohmann, L., and Graack, H.-R. (1998) *J. Biol. Chem.* **273**, 34828–34836

102a. Saveanu, C., Fromont-Racine, M., Harington, A., Ricard, F., Namane, A., and Jacquier, A. (2001) *J. Biol. Chem.* **276**, 15861–15867

102b. Graack, H.-R., Bryant, M. L., and O'Brien, T. W. (1999) *Biochemistry* **38**, 16569–16577

102c. Suzuki, T., Terasaki, M., Takemoto-Hori, C., Hanada, T., Ueda, T., Wada, A., and Watanabe, K. (2001) *J. Biol. Chem.* **276**, 33181–33195

103. Chan, Y.-L., and Wool, I. G. (1988) *J. Biol. Chem.* **263**, 2891–2896
104. Andres, J. L., Johansen, J. W., and Maller, J. L. (1987) *J. Biol. Chem.* **262**, 14389–14393
105. Wilson, K. S., Appelt, K., Badger, J., Tanaka, I., and White, S. W. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7251–7255

105a. Leijonmarck, M., Appelt, K., Badger, J., Liljas, A., Wilson, K. S., and White, S. W. (1988) *Proteins* **3**, 243–251

106. Stern, S., Wilson, R. C., and Noller, H. F. (1986) *J. Mol. Biol.* **192**, 101–110
107. van de Ven, F. J. M., and Hilbers, C. W. (1986) *J. Mol. Biol.* **192**, 389–417
108. Nag, B., Tewari, D. S., Sommer, A., Olson, H. M., Glitz, D. G., and Traut, R. R. (1987) *J. Biol. Chem.* **262**, 9681–9687
109. Wittmann, H.-G. (1976) *Eur. J. Biochem.* **61**, 1–13
110. Brockmöller, J., and Kamp, R. M. (1988) *Biochemistry* **27**, 3372–3381
111. Xiang, R. H., and Lee, J. C. (1989) *J. Biol. Chem.* **264**, 10542–10546
112. Pohl, T., and Wittmann-Liebold, B. (1988) *J. Biol. Chem.* **263**, 4293–4301
113. Yeh, Y.-C., Traut, R. R., and Lee, J. C. (1986) *J. Biol. Chem.* **261**, 14148–14153
114. Capel, M. S., Engelman, D. M., Freeborn, B. R., Kjeldgaard, M., Langer, J. A., Ramakrishnan, V., Schindler, D. G., Schneider, D. K., Schoenborn, B. P., Sillers, I.-Y., Yabuki, S., and Moore, P. B. (1987) *Science* **238**, 1403–1406
115. Liiv, A., Tenson, T., and Remme, J. (1996) *J. Mol. Biol.* **263**, 396–410
116. Nomura, M. (1973) *Science* **179**, 864–873
117. Herold, M., and Nierhaus, K. H. (1987) *J. Biol. Chem.* **262**, 8826–8833
118. Liiv, A., and Remme, J. (1998) *J. Mol. Biol.* **276**, 537–545
119. Green, R., and Noller, H. F. (1999) *Biochemistry* **38**, 1772–1779
120. Khaitovich, P., Tenson, T., Kloss, P., and Mankin, A. S. (1999) *Biochemistry* **38**, 1780–1788
121. Michael, W. M., and Dreyfuss, G. (1996) *J. Biol. Chem.* **271**, 11571–11574
122. Wimberly, B. T., White, S. W., and Ramakrishnan, V. (1997) *Structure* **5**, 1187–1198

References

123. Davies, C., Gerstner, R. B., Draper, D. E., Ramakrishnan, V., and White, S. W. (1998) *EMBO J.* **17**, 4545–4558
124. Markus, M. A., Gerstner, R. B., Draper, D. E., and Torchia, D. A. (1998) *EMBO J.* **17**, 4559–4571
125. Agalarov, S. C., Zheleznyakova, E. N., Selivanova, O. M., Zheleznaya, L. A., Matvienko, N. I., Vasiliev, V. D., and Spirin, A. S. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 999–1003
126. Nevskaya, N., Tishchenko, S., Nikulin, A., Al-Karadaghi, S., Liljas, A., Ehresmann, B., Ehresmann, C., Garber, M., and Nikonov, S. (1998) *J. Mol. Biol.* **279**, 233–244
127. Kean, J. M., and Draper, D. E. (1985) *Biochemistry* **24**, 5052–5061
128. Svensson, P., Changchien, L., Craven, G. R., and Noller, H. F. (1988) *J. Mol. Biol.* **200**, 301–308
129. Serganov, A., Bénard, L., Portier, C., Ennifar, E., Garber, M., Ehresmann, B., and Ehresmann, C. (2001) *J. Mol. Biol.* **305**, 785–803
130. Kalurachchi, K., and Nikonowicz, E. P. (1998) *J. Mol. Biol.* **280**, 639–654

130a. Tishchenko, S., Nikulin, A., Fomenkova, N., Nevskaya, N., Nikonov, O., Dumas, P., Moine, H., Ehresmann, B., Ehresmann, C., Piendl, W., Lamzin, V., Garber, M., and Nikonov, S. (2001) *J. Mol. Biol.* **311**, 311–324

131. Orr, J. W., Hagerman, P. J., and Williamson, J. R. (1998) *J. Mol. Biol.* **275**, 453–464
132. Hei, Y.-j, Pelech, S. L., Chen, X., Diamond, J., and McNeill, J. H. (1994) *J. Biol. Chem.* **269**, 7816–7823

132a. Martin, K. A., Schalm, S. S., Richardson, C., Romanelli, A., Keon, K. L., and Blenis, J. (2001) *J. Biol. Chem.* **276**, 7884–7891

132b. Martin, K. A., Schalm, S. S., Romanelli, A., Keon, K. L., and Blenis, J. (2001) *J. Biol. Chem.* **276**, 7892–7898

133. Morosyuk, S. V., Cunningham, P. R., and SantaLucia, J., Jr. (2001) *J. Mol. Biol.* **307**, 197–211
134. Davies, C., Bussiere, D. E., Golden, B. L., Porter, S. J., Ramakrishnan, V., and White, S. W. (1998) *J. Mol. Biol.* **279**, 873–888
135. Subramanian, A. R. (1984) *Trends Biochem. Sci.* **9**, 491–494
136. Kim, J., Chubatsu, L. S., Admon, A., Stahl, J., Fellous, R., and Linn, S. (1995) *J. Biol. Chem.* **270**, 13620–13629
137. Sorensen, M. A., Fricke, J., and Pedersen, S. (1998) *J. Mol. Biol.* **280**, 561–569
138. Laughrea, M., and Tam, J. (1991) *Biochemistry* **30**, 11412–11420
139. Nikonov, S., Nevskaya, N., Eliseikina, I., Fomenkova, N., Nikulin, A., Ossina, N., Garber, M., Jonsson, B.-H., Briand, C., Al-Karadaghi, S., Svensson, A., AEvarsson, A., and Liljas, A. (1996) *EMBO J.* **15**, 1350–1359

139a. Zengel, J. M., and Lindahl, L. (1994) *Prog. Nucleic Acid Res. Mol. Biol.* **47**, 331–370

140. Yeh, L.-C. C., and Lee, J. C. (1995) *J. Mol. Biol.* **246**, 295–307
141. Nag, B., Tewari, D. S., Etchison, J. R., Sommer, A., and Traut, R. R. (1986) *J. Biol. Chem.* **261**, 13892–13897
142. Nakagawa, A., Nakashima, T., Taniguchi, M., Hosaka, H., Kimura, M., and Tanaka, I. (1999) *EMBO J.* **18**, 1459–1467
143. Hoffman, D. W., Cameron, C. S., Davies, C., White, S. W., and Ramakrishnan, V. (1996) *J. Mol. Biol.* **264**, 1058–1071
144. Kuhlman, B., Yang, H. Y., Boice, J. A., Fairman, R., and Raleigh, D. P. (1997) *J. Mol. Biol.* **270**, 640–647
145. Lieberman, K. R., Firpo, M. A., Herr, A. J., Nguyenle, T., Atkins, J. F., Gesteland, R. F., and Noller, H. F. (2000) *J. Mol. Biol.* **297**, 1129–1143
146. Dey, D., Bochkariov, D. E., Jokhadze, G. G., and Traut, R. R. (1998) *J. Biol. Chem.* **273**, 1670–1676

146a. Montesano-Roditis, L., Glitz, D. G., Traut, R. R., and Stewart, P. L. (2001) *J. Biol. Chem.* **276**, 14117–14123

146a. Montesano-Roditis, L., Glitz, D. G., Traut, R. R., and Stewart, P. L. (2001) *J. Biol. Chem.* **276**, 14117–14123

146b. Pioletti, M., Schlünzen, F., Harms, J., Zarivach, R., Glühmann, M., Avila, H., Bashan, A., Bartels, H., Auerbach, T., Jacobi, C., Hartsch, T., Yonath, A., and Franceschi, F. (2001) *EMBO J.* **20**, 1829–1839

147. Wahl, M. C., Bourenkov, G. P., Bartunik, H. D., and Huber, R. (2000) *EMBO J.* **19**, 174–186
148. Porse, B. T., Leviev, I., Mankin, A. S., and Garrett, R. A. (1998) *J. Mol. Biol.* **276**, 391–404
149. Holmberg, L., and Noller, H. F. (1999) *J. Mol. Biol.* **289**, 223–233
150. Gonzalo, P., Lavergne, J.-P., and Reboud, J.-P. (2001) *J. Biol. Chem.* **276**, 19762–19769

150a. Zurdo, J., González, C., Sanz, J. M., Rico, M., Remacha, M., and Ballesta, J. P. G. (2000) *Biochemistry* **39**, 8935–8943

150b. Guarinos, E., Remacha, M., and Ballesta, J. P. G. (2001) *J. Biol. Chem.* **276**, 32474–32479

151. Szick-Miranda, K., and Bailey-Serres, J. (2001) *J. Biol. Chem.* **276**, 10921–10928
152. Dunbar, D. A., Gragon, F., Lee, S. J., and Baserga, S. J. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13027–13032
153. Rohl, R., and Nierhaus, K. H. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 729–733
154. Stoldt, M., Wöhnert, J., Görlach, M., and Brown, L. R. (1998) *EMBO J.* **17**, 6377–6384
155. Lu, M., and Steitz, T. A. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2023–2028
156. Chan, Y. L., Lin, A., McNally, J., and Wool, I. G. (1987) *J. Biol. Chem.* **262**, 12879–12886

156a. Stoldt, M., Wöhnert, J., Ohlenschläger, O., Görlach, M., and Brown, L. R. (1999) *EMBO J.* **18**, 6508–6521

156b. Johnson, A. W., Lund, E., and Dahlberg, J. (2002) *Trends Biochem. Sci.* **27**, 580 – 585

157. Lin, A., McNally, J., and Wool, I. G. (1983) *J. Biol. Chem.* **258**, 10664–10671
158. Lee, J. C., Henry, B., and Yeh, Y.-C. (1983) *J. Biol. Chem.* **258**, 854–858
159. Mao, H., and Williamson, J. R. (1999) *J. Mol. Biol.* **292**, 345–359
160. Stoffler-Meilicke, M., Stoffler, G., Odom, O. W., Zinn, A., Kramer, G., and Hardesty, B. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5538–5542
161. Trempe, M. R., Ohgi, K., and Glitz, D. G. (1982) *J. Biol. Chem.* **257**, 9822–9829
162. Hausner, T.-P., Geigenmüller, U., and Nierhaus, K. N. (1988) *J. Biol. Chem.* **263**, 13103–13111
163. Prince, J. B., and Garrett, R. A. (1982) *Trends Biochem. Sci.* **7**, 79
164. Denman, R., Nègre, D., Cunningham, P. R., Nurse, K., Colgan, J., Weitzmann, C., and Ofengand, J. (1989) *Biochemistry* **28**, 1012–1019

164a. Noller, H. F. (1993) *FASEB J.* **7**, 87–89

164b. Vila-Sanjurjo, A., and Dahlberg, A. E. (2001) *J. Mol. Biol.* **308**, 457–463

164c. Samaha, R. R., Green, R., and Noller, H. F. (1995) *Nature (London)* **377**, 309–314

164d. Porse, B. T., Thi-Ngoc, H. P., and Garrett, R. A. (1996) *J. Mol. Biol.* **264**, 472–483

164e. Green, R., Samaha, R. R., and Noller, H. F. (1997) *J. Mol. Biol.* **266**, 40–50

165. Lührmann, R., Bald, R., Stöffler-Meilicke, M., and Stöffler, G. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7276–7280
166. Olson, H. M., Nicholson, A. W., Cooperman, B. S., and Glitz, D. G. (1985) *J. Biol. Chem.* **260**, 10326–10331

166a. Green, R., Switzer, C., and Noller, H. F. (1998) *Science* **280**, 286–289

167. Goldman, R. A., Hasan, T., Hall, C. C., Strycharz, W. A., and Cooperman, B. S. (1983) *Biochemistry* **22**, 359–368

167a. Porse, B. T., and Garrett, R. A. (1995) *J. Mol. Biol.* **249**, 1–10

167b. Thompson, J., Kim, D. F., O'Connor, M., Lieberman, K. R., Bayfield, M. A., Gregory, S. T., Green, R., Noller, H. F., and Dahlberg, A. E. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9002–9007

168. Langer, J. A., and Lake, J. A. (1986) *J. Mol. Biol.* **187**, 617–621
169. Ibba, M., Becker, H. D., Stathopoulos, C., Tumbula, D. L., and Söll, D. (2000) *Trends Biochem. Sci.* **25**, 311–316
170. Weisblum, B. (1999) *Trends Biochem. Sci.* **24**, 247–250
171. Yue, D., Kintanar, A., and Horowitz, J. (1994) *Biochemistry* **33**, 8905–8911
172. Schimmel, P. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4521–4522
173. Senger, B., Aphasizhev, R., Walter, P., and Fasiolo, F. (1995) *J. Mol. Biol.* **249**, 45–58
174. Frazer-Abel, A. A., and Hagerman, P. J. (1999) *J. Mol. Biol.* **285**, 581–593
175. Jovine, L., Djordjevic, S., and Rhodes, D. (2000) *J. Mol. Biol.* **301**, 401–414
176. Auffinger, P., and Westhof, E. (1999) *J. Mol. Biol.* **292**, 467–483
177. Perreau, V. M., Keith, G., Holmes, W. M., Przykorska, A., Santos, M. A. S., and Tuite, M. F. (1999) *J. Mol. Biol.* **293**, 1039–1053
178. Wilson, R. K., and Roe, B. A. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 409–413
179. Rich, A. (1978) *Trends Biochem. Sci.* **3**, 34–37
180. Hou, Y.-M., Westhof, E., and Giegé, R. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6776–6780
181. Himeno, H., Yoshida, S., Soma, A., and Nishikawa, K. (1997) *J. Mol. Biol.* **268**, 704–711
182. Christian, T., Lipman, R. S. A., Evilia, C., and Hou, Y.-M. (2000) *J. Mol. Biol.* **303**, 503–514
183. Hamann, C. S., and Hou, Y.-M. (2000) *J. Mol. Biol.* **295**, 777–789
184. Drabkin, H. J., and RajBhandary, U. L. (1985) *J. Biol. Chem.* **260**, 5580–5587
185. Woo, N. H., Roe, B. A., and Rich, A. (1980) *Nature (London)* **286**, 346–351
186. Calagan, J. L., Pirtle, R. M., Pirtle, I. L., Kashdan, M. A., Vreman, H. J., and Dudock, B. S. (1980) *J. Biol. Chem.* **255**, 9981–9984
187. Kuchino, Y., Ihara, M., Yabusaki, Y., and Nishimura, S. (1982) *Nature (London)* **298**, 684–685
188. Bai, Y., Fox, D. T., Lacy, J. A., Van Lanen, S. G., and Iwata-Reuyl, D. (2000) *J. Biol. Chem.* **275**, 28731–28738
189. Watanabe, M., Nameki, N., Matsuo-Takasaki, M., Nishimura, S., and Okada, N. (2001) *J. Biol. Chem.* **276**, 2387–2394
190. Crick, F. H. C. (1966) *J. Mol. Biol.* **19**, 548–555

190a. Inagaki, Y., Kojima, A., Bessho, Y., Hori, H., Ohama, T., and Osawa, S. (1995) *J. Mol. Biol.* **251**, 486–492

191. Booth, V., Keizer, D. W., Kamphuis, M. B., Clark-Lewis, I., and Sykes, B. D. (2002) *Biochemistry* **41**, 10418 – 10425
192. Auld, D. S., and Schimmel, P. (1995) *Science* **267**, 1994–1996
193. Pak, M., Willis, I. M., and Schulman, L. H. (1994) *J. Biol. Chem.* **269**, 2277–2282
194. Li, S., Pelka, H., and Schulman, L. H. (1993) *J. Biol. Chem.* **268**, 18335–18339
195. Hong, K.-W., Ibba, M., Weygand-Durasevic, I., Rogers, M. J., Thomann, H.-U., and Söll, D. (1996) *EMBO J.* **15**, 1983–1991

References

196. Nureki, O., Niimi, T., Muramatsu, T., Kanno, H., Kohno, T., Florentz, C., Giegé, R., and Yokoyama, S. (1994) *J. Mol. Biol.* **236**, 710–724
197. Tocchini-Valentini, G., Saks, M. E., and Abelson, J. (2000) *J. Mol. Biol.* **298**, 779–793
198. Yan, W., and Francklyn, C. (1994) *J. Biol. Chem.* **269**, 10022–10027
198a. Shimada, A., Nureki, O., Goto, M., Takahashi, S., and Yokoyama, S. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13537–13542
198b. Xu, F., Jiang, G., Li, W., He, X., Jin, Y., and Wang, D. (2002) *Biochemistry* **41**, 8087–8092
199. Normanly, J., and Abelson, J. (1989) *Ann. Rev. Biochem.* **58**, 1029–1049
199a. Sampson, J. R., DiRenzo, A. B., Behlen, L. S., and Uhlenbeck, O. C. (1989) *Science* **243**, 1363–1366
200. Saks, M. E., Sampson, J. R., and Abelson, J. N. (1994) *Science* **263**, 191–197
201. McClain, W. H. (1993) *FASEB J.* **7**, 72–78
202. Ribas de Pouplana, L., and Schimmel, P. (2001) *J. Biol. Chem.* **276**, 6881–6884
203. Sherman, J. M., Thomann, H.-U., and Söll, D. (1996) *J. Mol. Biol.* **256**, 818–828
204. Arnez, J. G., Dock-Bregeon, A.-C., and Moras, D. (1999) *J. Mol. Biol.* **286**, 1449–1459
205. Horowitz, J., Chu, W.-C., Derrick, W. B., Liu, J. C.-H., Liu, M., and Yue, D. (1999) *Biochemistry* **38**, 7737–7746
206. Liu, M., and Horowitz, J. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10389–10393
207. Arnez, J. G., and Moras, D. (1997) *Trends Biochem. Sci.* **22**, 211–216
208. Schimmel, P. (1991) *Trends Biochem. Sci.* **16**, 1–3
209. Nureki, O., Vassylyev, D. G., Katayanagi, K., Shimizu, T., Sekine, S.-i, Kigawa, T., Miyazawa, T., Yokoyama, S., and Morikawa, K. (1995) *Science* **267**, 1958–1965
209a. Liu, J., Ibba, M., Hong, K.-W., and Söll, D. (1998) *Biochemistry* **37**, 9836–9842
209b. Ribas de Pouplana, L., and Schimmel, P. (2001) *Trends Biochem. Sci.* **26**, 591–596
210. Zelwer, C., Risler, J. L., and Brunie, S. (1982) *J. Mol. Biol.* **155**, 63–81
211. Xin, Y., Li, W., and First, E. A. (2000) *Biochemistry* **39**, 340–347
212. Ilyin, V. A., Temple, B., Hu, M., Li, G., Yin, Y., Vachette, P., and Carter, C. W., Jr. (2000) *Protein Sci.* **9**, 218–231
213. First, E. A., and Fersht, A. R. (1995) *Biochemistry* **34**, 5030–5043
214. Perona, J. J., Rould, M. A., and Steitz, T. A. (1993) *Biochemistry* **32**, 8758–8771
215. Fersht, A. R., Leatherbarrow, R. J., and Wells, T. N. C. (1986) *Trends Biochem. Sci.* **11**, 321–325
216. Leatherbarrow, R. J., Fersht, A. R., and Winter, G. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7840–7844
217. Desogus, G., Todone, F., Brick, P., and Onesti, S. (2000) *Biochemistry* **39**, 8418–8425
217a. Onesti, S., Desogus, G., Brevet, A., Chen, J., Plateau, P., Blanquet, S., and Brick, P. (2000) *Biochemistry* **39**, 12853–12861
218. Rould, M. A., Perona, J. J., Söll, D., and Steitz, T. A. (1989) *Science* **246**, 1135–1142
219. Rould, M. A., Perona, J. J., and Steitz, T. A. (1991) *Nature (London)* **352**, 213–218
220. Arnez, J. G., and Steitz, T. A. (1994) *Biochemistry* **33**, 7560–7567
221. Cavarelli, J., Delagoutte, B., Eriani, G., Gangloff, J., and Moras, D. (1998) *EMBO J.* **17**, 5438–5448
222. Brick, P., Bhat, T. N., and Blow, D. M. (1988) *J. Mol. Biol.* **208**, 83–98
223. Nureki, O., Vassylyev, D. G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T. L., Schimmel, P., and Yokoyama, S. (1998) *Science* **280**, 578–582
224. Cusack, S., Yaremchuk, A., and Tukalo, M. (2000) *EMBO J.* **19**, 2351–2361
225. Fourmy, D., Mechulam, Y., and Blanquet, S. (1995) *Biochemistry* **34**, 15681–15688
226. Mechulam, Y., Schmitt, E., Maveyraud, L., Zelwer, C., Nureki, O., Yokoyama, S., Konno, M., and Blanquet, S. (1999) *J. Mol. Biol.* **294**, 1287–1297
227. Arnez, J. G., Harris, D. C., Mitschler, A., Rees, B., Francklyn, C. S., and Moras, D. (1995) *EMBO J.* **14**, 4143–4155
228. Qiu, X., Janson, C. A., Blackburn, M. N., Chhohan, I. K., Hibbs, M., and Abdel-Meguid, S. S. (1999) *Biochemistry* **38**, 12296–12304
229. Arnez, J. G., Augustine, J. G., Moras, D., and Francklyn, C. S. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7144–7149
230. Biou, V., Yaremchuk, A., Tukalo, M., and Cusack, S. (1994) *Science* **263**, 1404–1410
231. Cusack, S., Yaremchuk, A., and Tukalo, M. (1996) *EMBO J.* **15**, 2834–2842
232. Eiler, S., Dock-Bregeon, A.-C., Moulinier, L., Thierry, J.-C., and Moras, D. (1999) *EMBO J.* **18**, 6532–6541
233. Archontis, G., Simonson, T., and Karplus, M. (2001) *J. Mol. Biol.* **306**, 307–327
234. Berthet-Colominas, C., Seignovert, L., Härtlein, M., Grotli, M., Cusack, S., and Leberman, R. (1998) *EMBO J.* **17**, 2947–2960
235. Commans, S., Lazard, M., Delort, F., Blanquet, S., and Plateau, P. (1998) *J. Mol. Biol.* **278**, 801–813
236. Reshetnikova, L., Moor, N., Lavrik, O., and Vassylyev, D. G. (1999) *J. Mol. Biol.* **287**, 555–568
237. Logan, D. T., Mazauric, M.-H., Kern, D., and Moras, D. (1995) *EMBO J.* **14**, 4156–4167
238. Xu, Z.-J., Love, M. L., Ma, L. Y. Y., Blum, M., Bronskill, P. M., Bernstein, J., Grey, A. A., Hofmann, T., Camerman, N., and Wong, J. T.-F. (1989) *J. Biol. Chem.* **264**, 4304–4311
239. Jasin, M., Regan, L., and Schimmel, P. (1983) *Nature (London)* **306**, 441–447
240. Webster, T. A., Lathrop, R. H., and Smith, T. F. (1987) *Biochemistry* **26**, 6950–6957
241. Schimmel, P. (1989) *Biochemistry* **28**, 2747–2759
242. Liu, J., Lin, S.-x, Blochet, J.-E., Pézolet, M., and Lapointe, J. (1993) *Biochemistry* **32**, 11390–11396
243. Sampson, J., and Uhlenbeck, O. C. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1033–1037
244. Ogilvie, K. K., Usman, N., Nicoghosian, K., and Cedergren, R. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5764–5768
244a. Hendrickson, T. L. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13473–13475
245. Alexander, R. W., and Schimmel, P. (1999) *Biochemistry* **38**, 16359–16365
246. McClain, W. H., Chen, Y.-M., Foss, K., and Schneider, J. (1988) *Science* **242**, 1681–1684
247. Francklyn, C., and Schimmel, P. (1989) *Nature (London)* **337**, 478–481
247a. Sardesai, N. Y., Green, R., and Schimmel, P. (1999) *Biochemistry* **38**, 12080–12088
248. Rogers, M. J., and Söll, D. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6627–6631
249. Khan, A. S., and Roe, B. A. (1988) *Science* **241**, 74–79
250. Connolly, B. A., Von der Haar, F., and Eckstein, F. (1980) *J. Biol. Chem.* **255**, 11301–11307
251. Fersht, A. R. (1987) *Biochemistry* **26**, 8031–8037
252. Leatherbarrow, R. J., and Fersht, A. R. (1987) *Biochemistry* **26**, 8524–8528
253. Ferscht, A. R. (1988) *Biochemistry* **27**, 1577–1580
254. Fersht, A. (1985) *Enzyme Structure and Mechanism*, 2nd ed., Freeman, New York
255. Xin, Y., Li, W., and First, E. A. (2000) *J. Mol. Biol.* **303**, 299–310
256. Bedouelle, H., and Winter, G. (1986) *Nature (London)* **320**, 371–373
257. Starzyk, R. M., Koontz, S. W., and Schimmel, P. (1982) *Nature (London)* **298**, 136–140
258. Fersht, A. R. (1977) *Enzyme Structure and Mechanism*, 1st ed., Freeman, San Francisco, California (p. 283)
259. Fersht, A. R. (1998) *Science* **280**, 541
259a. Mursinna, R. S., Lincecum, T. L., Jr., and Martinis, S. A. (2001) *Biochemistry* **40**, 5376–5381
259b. Tardif, K. D., Liu, M., Vitseva, O., Hou, Y.-M., and Horowitz, J. (2001) *Biochemistry* **40**, 8118–8125
259c. Tardif, K. D., and Horowitz, J. (2002) *Nucleic Acids Res.* **30**, 2538–2545
260. Gao, W., Goldman, E., and Jakubowski, H. (1994) *Biochemistry* **33**, 11528–11535
261. Jakubowski, H. (1995) *J. Biol. Chem.* **270**, 17672–17673
262. Jakubowski, H. (1996) *Biochemistry* **35**, 8252–8259
263. Serre, L., Verdon, G., Choinowski, T., Hervouet, N., Risler, J.-L., and Zelwer, C. (2001) *J. Mol. Biol.* **306**, 863–876
263a. Kelley, S. O., Steinberg, S. V., and Schimmel, P. (2001) *J. Biol. Chem.* **276**, 10607–10611
264. Yarus, M. (1992) *Trends Biochem. Sci.* **17**, 171–174
265. Farabaugh, P. J., and Björk, G. R. (1999) *EMBO J.* **18**, 1427–1434
266. Rodnina, M. V., and Wintermeyer, W. (2001) *Trends Biochem. Sci.* **26**, 124–130
267. Ibba, M., Curnow, A. W., and Söll, D. (1997) *Trends Biochem. Sci.* **22**, 39–42
268. Wallis, N. G., Dardel, F., and Blanquet, S. (1995) *Biochemistry* **34**, 7668–7677
268a. Horiuchi, K. Y., Harpel, M. R., Shen, L., Luo, Y., Rogers, K. C., and Copeland, R. A. (2001) *Biochemistry* **40**, 6450–6457
269. Curnow, A. W., Hong,, K-w., Yuan, R., Kim, S.-i, Martins, O., Winkler, W., Henkin, T. M., and Söll, D. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11819–11826
270. Böck, A., Forchhammer, K., Heider, J., and Baron, C. (1991) *Trends Biochem. Sci.* **16**, 463–467
271. Yarus, M. (2000) *Science* **287**, 440–441
272. Stathopoulos, C., Li, T., Longman, R., Vothknecht, U. C., Becker, H. D., Ibba, M., and Söll, D. (2000) *Science* **287**, 479–482
272a. Stathopoulos, C., Kim, W., Li, T., Anderson, I., Deutsch, B., Palioura, S., Whitman, W., and Söll, D. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14292–14297
273. Wool, I. G. (1996) *Trends Biochem. Sci.* **21**, 164–165
274. Martinis, S. A., Plateau, P., Cavarelli, J., and Florentz, C. (1999) *EMBO J.* **18**, 4591–4596
275. Putney, S. D., and Schimmel, P. (1981) *Nature (London)* **291**, 632–635
276. Romby, P., Caillet, J., Ebel, C., Sacerdot, C., Graffe, M., Eyermann, F., Brunel, C., Moine, H., Ehresmann, C., Ehresmann, B., and Springer, M. (1996) *EMBO J.* **15**, 5976–5987
277. Gerken, S. C., and Arfin, S. M. (1984) *J. Biol. Chem.* **259**, 11160–11161
278. Lambowitz, A. M., and Perlman, P. S. (1990) *Trends Biochem. Sci.* **15**, 440–444
279. Abramochkin, G., and Shrader, T. E. (1996) *J. Biol. Chem.* **271**, 22901–22907
280. Li, J., and Pickart, C. M. (1995) *Biochemistry* **34**, 139–147
281. Dou, X., Limmer, S., and Kreutzer, R. (2001) *J. Mol. Biol.* **305**, 451–458
281a. Lund, E., and Dahlberg, J. E. (1998) *Science* **282**, 2003–2004
281b. Sarkar, S., Azad, A. K., and Hopper, A. K. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14366–14371

References

282. Schimmel, P., and Ribas De Pouplana, L. (2000) *Trends Biochem. Sci.* **25**, 207–209
283. Kisselev, L., Frolova, L., and Haenni, A.-L. (1993) *Trends Biochem. Sci.* **18**, 263–267
284. Sissler, M., Delorme, C., Bond, J., Ehrlich, S. D., Renault, P., and Francklyn, C. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8985–8990
285. Raben, N., Nichols, R., Dohlman, J., McPhie, P., Sridhar, V., Hyde, C., Leff, R., and Plotz, P. (1994) *J. Biol. Chem.* **269**, 24277–24283
286. Ge, Q., Trieu, E. P., and Targoff, I. N. (1994) *J. Biol. Chem.* **269**, 28790–28797
287. Winter, R. B., Morrissey, L., Gauss, P., Gold, L., Hsu, T., and Karam, J. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7822–7826
288. Peabody, D. S. (1989) *J. Biol. Chem.* **264**, 5031–5035
289. Gupta, K. C., and Patwardhan, S. (1988) *J. Biol. Chem.* **263**, 8553–8556
290. Newton, D. T., Creuzenet, C., and Mangroo, D. (1999) *J. Biol. Chem.* **274**, 22143–22146
291. RajBhandary, U. L. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1325–1327
292. Li, S., Kumar, N. V., Varshney, U., and RajBhandary, U. L. (1996) *J. Biol. Chem.* **271**, 1022–1028
293. Takeuchi, N., Kawakami, M., Omori, A., Ueda, T., Spremulli, L. L., and Watanabe, K. (1998) *J. Biol. Chem.* **273**, 15085–15090
294. Shine, J., and Dalgarno, L. (1975) *Nature (London)* **254**, 34–38
295. Steitz, J. A., and Jakes, K. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4734–4738

295a. van Dieijen, G., Zipori, P., van Prooijen, W., and van Duin, J. (1978) *Eur. J. Biochem.* **90**, 571–580

296. Wu, X.-Q., Iyengar, P., and RajBhandary, U. L. (1996) *EMBO J.* **15**, 4734–4739

296a. Roll-Mecak, A., Shin, B.-S., Dever, T. E., and Burley, S. K. (2001) *Trends Biochem. Sci.* **26**, 705–709

297. Biou, V., Shu, F., and Ramakrishnan, V. (1995) *EMBO J.* **14**, 4056–4064

297a. Morosyuk, S. V., Cunningham, P. R., SantaLucia, J., Jr. (2001) *J. Mol. Biol.* **307**, 197–211

298. Moreau, M., de Cock, E., Fortier, P.-L., Garcia, C., Albaret, zC., Blanquet, S., Lallemand, J.-Y., and Dardel, F. (1997) *J. Mol. Biol.* **266**, 15–22
299. Yu, N.-J., and Spremulli, L. L. (1997) *Biochemistry* **36**, 14827–14835
300. La Teana, A., Pon, C. L., and Gualerzi, C. O. (1996) *J. Mol. Biol.* **256**, 667–675
301. Karimi, R., Pavlov, M. Y., Heurgué-Hamard, V., Buckingham, R. H., and Ehrenberg, M. (1998) *J. Mol. Biol.* **281**, 241–252
302. Dahlquist, K. D., and Puglisi, J. D. (2000) *J. Mol. Biol.* **299**, 1–15
303. Meinnel, T., Sacerdot, C., Graffe, M., Blanquet, S., and Springer, M. (1999) *J. Mol. Biol.* **290**, 825–837
304. Garcia, C., Fortier, P.-L., Blanquet, S., Lallemand, J.-Y., and Dardel, F. (1995) *J. Mol. Biol.* **254**, 247–259
305. Kozak, M. (1983) *Microbiol. Rev.* **47**, 1–45
306. Moldave, R. (1985) *Ann. Rev. Biochem.* **54**, 1109–1149
307. Lorsch, J. R., and Herschlag, D. (1999) *EMBO J.* **18**, 6705–6717
308. Rhoads, R. E. (1999) *J. Biol. Chem.* **274**, 30337–30340
309. Rhoads, R. E. (1993) *J. Biol. Chem.* **268**, 3017–3020
310. Browning, K. S., Maia, D. M., Lax, S. R., and Ravel, J. M. (1987) *J. Biol. Chem.* **262**, 538–541
311. Pestova, T. V., and Hellen, C. U. T. (1999) *Trends Biochem. Sci.* **24**, 85–87
312. Altmann, M., and Trachsel, H. (1993) *Trends Biochem. Sci.* **18**, 429–432
313. Sonenberg, N., Hershey, J. W. B., and Mathews, M. B., eds. (2000) *Translational Control of Gene Expression*, Cold Spring Harbor Lab. Press, Plainview, New York
314. Carey, M., and Smale, S. T. (2000) *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques*, Cold Spring Harbor Lab. Press, Plainview, New York

314a. Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I., and Hellen, C. U. T. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7029–7036

315. Jackson, R. J. (1998) *Nature (London)* **394**, 829–831
316. Richter-Cook, N. J., Dever, T. E., Hensold, J. O., and Merrick, W. C. (1998) *J. Biol. Chem.* **273**, 7579–7587
317. Richter, N. J., Rogers, G. W., Jr., Hensold, J. O., and Merrick, W. C. (1999) *J. Biol. Chem.* **274**, 35415–35424
318. Melander, Y., Holmberg, L., and Nygård, O. (1997) *J. Biol. Chem.* **272**, 3254–3258
319. Korneeva, N. L., Lamphear, B. J., Hennigan, F. L. C., and Rhoads, R. E. (2000) *J. Biol. Chem.* **275**, 41369–41376

319a. Browning, K. S., Gallie, D. R., Hershey, J. W. B., Hinnebusch, A. G., Maitra, U., Merrick, W. C., and Norbury, C. (2001) *Trends Biochem. Sci.* **26**, 284

320. Valásek, L., Phan, L., Schoenfeld, L. W., Valásková, V., and Hinnebusch, A. G. (2001) *EMBO J.* **20**, 891–904
321. Raychaudhuri, P., and Maitra, U. (1986) *J. Biol. Chem.* **261**, 7723–7728
322. Proud, C. G. (1986) *Trends Biochem. Sci.* **11**, 73–77
323. Gil, J., Esteban, M., and Roth, D. (2000) *Biochemistry* **39**, 7521–7530
324. Sudhakar, A., Krishnamoorthy, T., Jain, A., Chatterjee, U., Hasnain, S. E., Kaufman, R. J., and Ramaiah, K. V. A. (1999) *Biochemistry* **38**, 15398–15405
325. Kimball, S. R., Fabian, J. R., Pavitt, G. D., Hinnebusch, A. G., and Jefferson, L. S. (1998) *J. Biol. Chem.* **273**, 12841–12845
326. De Haro, C., Méndez, R., and Santoyo, J. (1996) *FASEB J.* **10**, 1378–1387
327. Rose, D. W., Welch, W. J., Kramer, G., and Hardesty, B. (1989) *J. Biol. Chem.* **264**, 6239–6244
328. Williams, D. D., Pavitt, G. D., and Proud, C. G. (2001) *J. Biol. Chem.* **276**, 3733–3742

328a. Asano, K., Krishnamoorthy, T., Phan, L., Pavitt, G. D., Hinnebusch, A. G. (1999) *EMBO J.* **18**, 1673–1688

329. Mueller, P. P., Grueter, P., Hinnebusch, A. G., and Trachsel, H. (1998) *J. Biol. Chem.* **273**, 32870–32877
330. Qiu, H., Dong, J., Hu, C., Francklyn, C. S., and Hinnebusch, A. G. (2001) *EMBO J.* **20**, 1425–1438

330a. Wang, X., Paulin, F. E. M., Campbell, L. E., Gomez, E., O'Brien, K., Morrice, N., and Proud, C. G. (2001) *EMBO J.* **20**, 4349–4359

330b. Wang, X., Li, W., Williams, M., Terada, N., Alessi, D. R., and Proud, C. G. (2001) *EMBO J.* **20**, 4370–4379

330c. Cho, S., and Hoffman, D. W. (2002) *Biochemistry* **41**, 5730–5742

331. Chen, J.-J., and London, I. M. (1995) *Trends Biochem. Sci.* **20**, 105–108
332. Dholakia, J. N., Mueser, T. C., Woodley, C. L., Parkhurst, L. J., and Wahba, A. J. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6746–6750
333. Gingras, A.-C., Raught, B., and Sonenberg, N. (1999) *Ann. Rev. Biochem.* **68**, 913–963
334. Herbert, T. P., Kilhams, G. R., Batty, I. H., and Proud, C. G. (2000) *J. Biol. Chem.* **275**, 11249–11256

334a. Karim, M. M., Hughes, J. M. X., Warwicker, J., Scheper, G. C., Proud, C. G., and McCarthy, J. E. G. (2001) *J. Biol. Chem.* **276**, 20750–20757

335. Dever, T. E. (1999) *Trends Biochem. Sci.* **24**, 398–403
336. De Gregorio, E., Preiss, T., and Hentze, M. W. (1999) *EMBO J.* **18**, 4865–4874
337. Korneeva, N. L., Lamphear, B. J., Hennigan, F. L. C., Merrick, W. C., and Rhoads, R. E. (2001) *J. Biol. Chem.* **276**, 2872–2879
338. Lamphear, B. J., Kirchweger, R., Skern, T., and Rhoads, R. E. (1995) *J. Biol. Chem.* **270**, 21975–21983
339. McCubbin, W. D., Edery, I., Altmann, M., Sonenberg, N., and Kay, C. M. (1988) *J. Biol. Chem.* **263**, 17663–17671
340. Keiper, B. D., Lamphear, B. J., Deshpande, A. M., Jankowska-Anyszka, M., Aamodt, E. J., Blumenthal, T., and Rhoads, R. E. (2000) *J. Biol. Chem.* **275**, 10590–10596
341. Rao, G. N. (2000) *J. Biol. Chem.* **275**, 16993–16999

341a. Niedzwiecka, A., Stepinski, J., Darzynkiewicz, E., Sonenberg, N., and Stolarski, R. (2002) *Biochemistry* **41**, 12140–12148

342. Naranda, T., Strong, W. B., Menaya, J., Fabbri, B. J., and Hershey, J. W. B. (1994) *J. Biol. Chem.* **269**, 14465–14472
343. Dominguez, D., Altmann, M., Benz, J., Baumann, U., and Trachsel, H. (1999) *J. Biol. Chem.* **274**, 26720–26726
344. Chappell, S. A., Edelman, G. M., and Mauro, V. P. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1536–1541
345. Hentze, M. W. (1997) *Science* **275**, 500–501

345a. Fernandez, J., Yaman, I., Mishra, R., Merrick, W. C., Snider, M. D., Lamers, W. H., and Hatzoglou, M. (2001) *J. Biol. Chem.* **276**, 12285–12291

346. Spahn, C. M. T., Kieft, J. S., Grassucci, R. A., Penczek, P. A., Zhou, K., Doudna, J. A., and Frank, J. (2001) *Science* **291**, 1959–1962
347. Hudder, A., and Werner, R. (2000) *J. Biol. Chem.* **275**, 34586–34591
348. Kozak, M. (1997) *EMBO J.* **16**, 2482–2492
349. Geballe, A. P., and Morris, D. R. (1994) *Trends Biochem. Sci.* **19**, 159–164
350. Cigan, A. M., Feng, L., and Donahue, T. F. (1988) *Science* **242**, 93–97

350a. Galy, B. (2001) *Trends Biochem. Sci.* **26**, 220

351. Chen, C.-y, and Sarnow, P. (1995) *Science* **268**, 415–417
352. Otero, L. J., Ashe, M. P., and Sachs, A. B. (1999) *EMBO J.* **18**, 3153–3163
353. Le, H., Browning, K. S., and Gallie, D. R. (2000) *J. Biol. Chem.* **275**, 17452–17462
354. Kozlov, G., Trempe, J.-F., Khaleghpour, K., Kahvejian, A., Ekiel, I., and Gehring, K. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4409–4413
355. Das, S., Maiti, T., Das, K., and Maitra, U. (1997) *J. Biol. Chem.* **272**, 31712–31718
356. Joe, Y. A., and Park, M. H. (1994) *J. Biol. Chem.* **269**, 25916–25921
357. Park, M. H., Wolff, E. C., and Folk, J. E. (1993) *Trends Biochem. Sci.* **18**, 475–479
358. Joao, H. C., Csonga, R., Klier, H., Koettnitz, K., Auer, M., and Eder, J. (1995) *Biochemistry* **34**, 14703–14711
359. Chetverin, A. B., and Spirin, A. S. (1982) *Biochim. Biophys. Acta.* **683**, 153–179
360. Gnirke, A., Geigenmüller, U., Rheinberger, H.-J., and Nierhaus, K. H. (1989) *J. Biol. Chem.* **264**, 7291–7301
361. Thompson, R. C. (1988) *Trends Biochem. Sci.* **13**, 91–93
362. Abel, K., and Jurnak, F. (1996) *Structure* **4**, 229–238
363. Jurnak, F. (1985) *Science* **230**, 32–36

References

364. Sprinzl, M. (1994) *Trends Biochem. Sci.* **19**, 245–250
365. Krásny, L., Mesters, J. R., Tieleman, L. N., Kraal, B., Fucík, V., Hilgenfeld, R., and Jonák, J. (1998) *J. Mol. Biol.* **283**, 371–381
366. Song, H., Parsons, M. R., Rowsell, S., Leonard, G., and Phillips, S. E. V. (1999) *J. Mol. Biol.* **285**, 1245–1256
367. Krab, I. M., and Parmeggiani, A. (1999) *Biochemistry* **38**, 13035–13041
368. Andersen, G. R., Thirup, S., Spremulli, L. L., and Nyborg, J. (2000) *J. Mol. Biol.* **297**, 421–436
369. Jiang, Y., Nock, S., Nesper, M., Sprinzl, M., and Sigler, P. B. (1996) *Biochemistry* **35**, 10269–10278
370. Uetsuki, T., Naito, A., Nagata, S., and Kaziro, Y. (1989) *J. Biol. Chem.* **264**, 5791–5798
371. Riis, B., Rattan, S. I. S., Clark, B. F. C., and Merrick, W. C. (1990) *Trends Biochem. Sci.* **15**, 420–424
372. Cavallius, J., and Merrick, W. C. (1998) *J. Biol. Chem.* **273**, 28752–28758
373. Edmonds, B. T., Bell, A., Wyckoff, J., Condeelis, J., and Leyh, T. S. (1998) *J. Biol. Chem.* **273**, 10288–10295
374. Condeelis, J. (1995) *Trends Biochem. Sci.* **20**, 169–170
374a. Pleiss, J. A., and Uhlenbeck, O. C. (2001) *J. Mol. Biol.* **308**, 895–905
375. Cunningham, P. R., Nurse, K., Weitzmann, C. J., Nègre, D., and Ofengand, J. (1992) *Biochemistry* **31**, 7629–7637
376. Gornicki, P., Nurse, K., Hellmann, W., Boublik, M., and Ofengand, J. (1984) *J. Biol. Chem.* **259**, 10493–10498
377. Weller, J., and Hill, W. E. (1994) *J. Biol. Chem.* **269**, 19369–19374
378. Lynch, S. R., and Puglisi, J. D. (2001) *J. Mol. Biol.* **306**, 1023–1035
378a. Ericson, G., Minchew, P., and Wollenzien, P. (1995) *J. Mol. Biol.* **250**, 407–419
379. VanLoock, M. S., Easterwood, T. R., and Harvey, S. C. (1999) *J. Mol. Biol.* **285**, 2069–2078
380. Lodmell, J. S., and Dahlberg, A. E. (1997) *Science* **277**, 1262–1267
381. Stark, H., Rodnina, M. V., Rinke-Appel, J., Brimacombe, R., Wintermeyer, W., and van Heel, M. (1997) *Nature (London)* **389**, 403–406
382. Hall, C. C., Smith, J. E., and Cooperman, B. S. (1985) *Biochemistry* **24**, 5702–5711
383. Douthwaite, S., and Aagaard, C. (1993) *J. Mol. Biol.* **232**, 725–731
384. Villsen, I. D., Vester, B., and Douthwaite, S. (1999) *J. Mol. Biol.* **286**, 365–374
385. Moazed, D., Robertson, J. M., and Noller, H. F. (1988) *Nature (London)* **334**, 362–364
385a. Rodnina, M. V., Pape, T., Fricke, R., Kuhn, L., and Wintermeyer, W. (1996) *J. Biol. Chem.* **271**, 646–652
385b. Piepenburg, O., Pape, T., Pleiss, J. A., Wintermeyer, W., Uhlenbeck, O. C., and Rodnina, M. V. (2000) *Biochemistry* **39**, 1734–1738
385c. Knudsen, C., Wieden, H.-J., and Rodnina, M. V. (2001) *J. Biol. Chem.* **276**, 22183–22190
385d. Vogeley, L., Palm, G. J., Mesters, J. R., and Hilgenfeld, R. (2001) *J. Biol. Chem.* **276**, 17149–17155
385e. Simonson, A. B., and Lake, J. A. (2002) *Nature (London)* **416**, 281–285
386. Hecht, S. M., Tan, K. H., Chinault, A. C., and Arcari, P. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 437–441
386a. Sprinzl, M., Kucharzewski, M., Hobbs, J. B., and Cramer, F. (1977) *Eur. J. Biochem.* **78**, 55–61
387. Taiji, M., Yokoyama, S., and Miyazawa, T. (1985) *Biochemistry* **24**, 5776–5780
387a. Wagner, T., and Sprinzl, M. (1983) *Biochemistry* **22**, 94–98
387b. Limmer, St., Vogtherr, M., Nawrot, B., Hillenbrand, R., and Sprinzl, M. (1997) *Angew. Chem. Int. Ed. Engl.* **36**, 2485–2489
387c. Polacek, N., Gaynor, M., Yassin, A., and Mankin, A. S. (2001) *Nature (London)* **411**, 498–501
387d. Katunin, V. I., Muth, G. W., Strobel, S. A., Wintermeyer, W., and Rodnina, M. V. (2002) *Molecular Cell* **10**, 339–346
387e. Bayfield, M. A., Dahlberg, A. E., Schulmeister, U., Dorner, S., and Barta, A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10096–10101
388. Lim, V. I., and Spirin, A. S. (1986) *J. Mol. Biol.* **188**, 565–577
389. Samaha, R. R., Green, R., and Noller, H. F. (1995) *Nature (London)* **377**, 309–314
390. Kloss, P., Xiong, L., Shinabarger, D. L., and Mankin, A. S. (1999) *J. Mol. Biol.* **294**, 93–101
391. Gregory, S. T., and Dahlberg, A. E. (1999) *J. Mol. Biol.* **285**, 1475–1483
392. Agrawal, R. K., Penczek, P., Grassucci, R. A., and Frank, J. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6134–6138
393. Wilson, K. S., and Noller, H. F. (1998) *Cell* **92**, 131–139
394. Savelsbergh, A., Matassova, N. B., Rodnina, M. V., and Wintermeyer, W. (2000) *J. Mol. Biol.* **300**, 951–961
394a. Katunin, V. I., Savelsbergh, A., Rodnina, M. V., and Wintermeyer, W. (2002) *Biochemistry* **41**, 12806–12812
395. Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B. F. C., and Nyborg, J. (1995) *Science* **270**, 1464–1472
396. Nyborg, J., Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Clark, B. F. C., and Reshetnikova, L. (1996) *Trends Biochem. Sci.* **21**, 81–82
397. Van Ness, B. G., Howard, J. B., and Bodley, J. W. (1980) *J. Biol. Chem.* **255**, 10710–10716
397a. Nierhaus, K. H., Wadzack, J., Burkhardt, N., Jünemann, R., Meerwinck, W., Willumeit, R., and Stuhrmann, H. B. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 945–950
397b. Frank, J., and Agrawal, R. K. (1998) *Biophys. J.* **74**, 589–594
397c. Culver, G. M. (2001) *Structure* **9**, 751–758
398. Rodnina, M. V., Savelsbergh, A., Katunin, V. I., and Wintermeyer, W. (1997) *Nature (London)* **385**, 37–41
398a. Agrawal, R. K., Linde, J., Sengupta, J., Nierhaus, K. H., and Frank, J. (2001) *J. Mol. Biol.* **311**, 777–787
399. Gabashvili, I. S., Agrawal, R. K., Grassucci, R., Squires, C. L., Dahlberg, A. E., and Frank, J. (1999) *EMBO J.* **18**, 6501–6507
400. Mohr, D., Wintermeyer, W., and Rodnina, M. V. (2000) *EMBO J.* **19**, 3458–3464
401. Uchiumi, T., Hori, K., Nomura, T., and Hachimori, A. (1999) *J. Biol. Chem.* **274**, 27578–27582
402. VanLoock, M. S., Agrawal, R. K., Gabashvili, I. S., Qi, L., Frank, J., and Harvey, S. C. (2000) *J. Mol. Biol.* **304**, 507–515
403. Mesters, J. R., Potapov, A. P., de Graaf, J. M., and Kraal, B. (1994) *J. Mol. Biol.* **242**, 644–654
404. Montesano-Roditis, L., and Glitz, D. G. (1994) *J. Biol. Chem.* **269**, 6458–6470
405. Bhangu, R., and Wollenzien, P. (1992) *Biochemistry* **31**, 5937–5944
406. Graifer, D. M., Juzumiene, D. I., Karpova, G. G., and Wollenzien, P. (1994) *Biochemistry* **33**, 6201–6206
407. Sergiev, P. V., Lavrik, I. N., Wlasoff, V. A., Dokudovskaya, S. S., Dontsova, O. A., Bogdanov, A. A., and Brimacombe, R. (1997) *RNA* **3**, 464–475
408. Qin, S., Moldave, K., and McLaughlin, C. S. (1987) *J. Biol. Chem.* **262**, 7802–7807
409. Triana-Alonso, F. J., Chakraburtty, K., and Nierhaus, K. H. (1995) *J. Biol. Chem.* **270**, 20473–20478
410. Kambampati, R., Pellegrino, C., Paiva, A., Huang, L., Mende-Mueller, L., and Chakraburtty, K. (2000) *J. Biol. Chem.* **275**, 16963–16968
410a. Warner, J. R. (1999) *Trends Biochem. Sci.* **24**, 437–440
410b. Schleif, R., Hess, W., Finkelstein, S., and Ellis, D. (1973) *J. Bacteriol.* **115**, 9–14
410c. Andersson, D. I., Bohman, K., Isaksson, L. A., and Kurland, C. G. (1982) *Mol. Gen. Genet.* **187**, 467–472
410d. Dintzis, H. M. (1961) *Proc. Natl. Acad. Sci. U.S.A.* **47**, 247–261
410e. Goustin, A. S., and Wilt, F. H. (1982) *Biochim. Biophys. Acta.* **699**, 22–27
411. Tate, W. P., and Brown, C. M. (1992) *Biochemistry* **31**, 2443–2450
412. Caskey, C. T. (1980) *Trends Biochem. Sci.* **5**, 234–237
413. Stansfield, I., Jones, K. M., and Tuite, M. F. (1995) *Trends Biochem. Sci.* **20**, 489–491
413a. Wilson, K. S., Ito, K., Noller, H. F., and Nakamura, Y. (2000) *Nature Struct. Biol.* **7**, 866–870
413b. Kisselev, L. L., and Buckingham, R. H. (2000) *Trends Biochem. Sci.* **25**, 561–566
414. Tate, W. P., Hornig, H., and Luhrmann, R. (1983) *J. Biol. Chem.* **258**, 10360–10365
415. Poole, E. S., Brown, C. M., and Tate, W. P. (1995) *EMBO J.* **14**, 151–158
416. Zhang, S., Rydén-Aulin, M., and Isaksson, L. A. (1998) *J. Mol. Biol.* **284**, 1243–1246
417. Pavlov, M. Y., Freistroffer, D. V., MacDougall, J., Buckingham, R. H., and Ehrenberg, M. (1997) *EMBO J.* **16**, 4134–4141
418. Pavlov, M. Y., Freistroffer, D. V., Heurguè-Hamard, V., Buckingham, R. H., and Ehrenberg, M. (1997) *J. Mol. Biol.* **273**, 389–401
418a. Ito, K., Frolova, L., Seit-Nebi, A., Karamyshev, A., Kisselev, L., and Nakamura, Y. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8494–8499
419. Velichutina, I. V., Hong, J. Y., Mesecar, A. D., Chernoff, Y. O., and Liebman, S. W. (2001) *J. Mol. Biol.* **305**, 715–727
419a. Nakamura, Y., Ito, K., and Ehrenberg, M. (2000) *Cell* **101**, 349–352
419b. Song, H., Mugnier, P., Das, A. K., Webb, H. M., Evans, D. R., Tuite, M. F., Hemmings, B. A., and Barford, D. (2000) *Cell* **100**, 311–321
420. Wickner, R. B., Edskes, H. K., Maddelein, M.-L., Taylor, K. L., and Moriyama, H. (1999) *J. Biol. Chem.* **274**, 555–558
421. Sparrer, H. E., Santoso, A., Szoka, F. C., Jr., and Weissman, J. S. (2000) *Science* **289**, 595–599
422. King, C.-Y. (2001) *J. Mol. Biol.* **307**, 1247–1260
423. Selmer, M., Al-Karadaghi, S., Hirokawa, G., Kaji, A., and Liljas, A. (1999) *Science* **286**, 2349–2352
424. Yoshida, T., Uchiyama, S., Nakano, H., Kashimori, H., Kijima, H., Ohshima, T., Saihara, Y., Ishino, T., Shimahara, H., Yoshida, T., Yokose, K., Ohkubo, T., Kaji, A., and Kobayashi, Y. (2001) *Biochemistry* **40**, 2387–2396
424a. Kim, K. K., Min, K., and Suh, S. W. (2000) *EMBO J.* **19**, 2362–2370
424b. Karimi, R., Pavlov, M. Y., Buckingham, R. H., and Ehrenberg, M. (1999) *Mol. Cell. Biol.* **3**, 601–609
425. Rolland, N., Janosi, L., Block, M. A., Shuda, M., Teyssier, E., Miège, C., Chéniclet, C., Carde, J.-P., Kaji, A., and Joyard, J. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5464–5469
425a. Lancaster, L., Kiel, M. C., Kaji, A., and Noller, H. F. (2002) *Cell* **111**, 129–140

References

426. Gast, F.-U., Peters, F., and Pingoud, A. (1987) *J. Biol. Chem.* **262**, 11920–11926
427. Ibba, M., and Söll, D. (1999) *Science* **286**, 1893–1897
428. Parker, J., Johnston, T. C., Borgia, P. T., Holtz, G., Remaut, E., and Fiers, W. (1983) *J. Biol. Chem.* **258**, 10007–10012
429. Precup, J., and Parker, J. (1987) *J. Biol. Chem.* **262**, 11351–11355
429a. LaRiviere, F. J., Wolfson, A. D., and Uhlenbeck, O. C. (2001) *Science* **294**, 165–168
429b. Ibba, M. (2001) *Science* **294**, 70–71
429c. Stahl, G., McCarty, G. P., and Farabaug, P. J. (2002) *Trends Biochem. Sci.* **27**, 178–183
430. Kurland, C. G. (1987) *Trends Biochem. Sci.* **12**, 169–171
431. Metzler, D. E. (1977) *Biochemistry; The Chemical Reactions of Living Cells*, Academic Press, New York (pp. 914–915)
432. Oresic, M., and Shalloway, D. (1998) *J. Mol. Biol.* **281**, 31–48
433. Bennetzen, J. L., and Hall, B. D. (1982) *J. Biol. Chem.* **257**, 3026–3031
434. Heurgué-Hamard, V., Mora, L., Guarneros, G., and Buckingham, R. H. (1996) *EMBO J.* **15**, 2826–2833
435. Schmitt, E., Mechulam, Y., Fromant, M., Plateau, P., and Blanquet, S. (1997) *EMBO J.* **16**, 4760–4769
436. Fromant, M., Plateau, P., Schmitt, E., Mechulam, Y., and Blanquet, S. (1999) *Biochemistry* **38**, 4982–4987
436a. van Hoof, A., Frischmeyer, P. A., Dietz, H. C., and Parker, R. (2002) *Science* **295**, 2262–2264
436b. Maquat, L. E. (2002) *Science* **295**, 2221–2222
437. Keiler, K. C., Waller, P. R. H., and Sauer, R. T. (1996) *Science* **271**, 990–994
438. Karzai, A. W., and Sauer, R. T. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3040–3044
438a. Roche, E. D., and Sauer, R. T. (2001) *J. Biol. Chem.* **276**, 28509–28515
438b. Gillet, R., and Felden, B. (2001) *EMBO J.* **20**, 2966–2976
439. Nameki, N., Tadaki, T., Muto, A., and Himeno, H. (1999) *J. Mol. Biol.* **289**, 1–7
440. Barends, S., Wower, J., and Kraal, B. (2000) *Biochemistry* **39**, 2652–2658
440a. Barends, S., Karzai, A. W., Sauer, R. T., and Wower, J. (2001) *J. Mol. Biol.* **314**, 9–21
440b. Zwieb, C., Guven, S. A., Wower, I. K., and Wower, J. (2001) *Biochemistry* **40**, 9587–9595
440c. Stagg, S. M., Frazer-Abel, A. A., Hagerman, P. J., and Harvey, S. C. (2001) *J. Mol. Biol.* **309**, 727–735
441. Littauer, U. Z., and Inouye, H. (1973) *Ann. Rev. Biochem.* **42**, 439–470
442. Riddle, D. L., and Carbon, J. (1973) *Nature New Biol.* **242**, 230–234
443. Magliery, T. J., Anderson, J. C., and Schultz, P. G. (2001) *J. Mol. Biol.* **307**, 755–769
444. Murgola, E. J., Prather, N. E., Mims, B. H., Pagel, F. T., and Hijazi, K. A. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4936–4939
445. Bossi, L., and Smith, D. M. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6105–6109
446. Murgola, E. J., Hijazi, K. A., Göringer, H. U., and Dahlberg, A. E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4162–4165
447. Surguchov, A. P. (1988) *Trends Biochem. Sci.* **13**, 120–123
448. Hatfield, D. (1985) *Trends Biochem. Sci.* **10**, 201–204
448a. Srinivasan, G., James, C. M., and Krzycki, J. A. (2002) *Science* **296**, 1459–1462
448b. Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. A., and Chan, M. K. (2002) *Science* **296**, 1462–1466
449. Engelberg-Kulka, H., and Schoulaker-Schwarz, R. (1988) *Trends Biochem. Sci.* **13**, 419–421
450. Stadtman, T. C. (1996) *Ann. Rev. Biochem.* **65**, 83–100
451. Kryukov, G. V., Kryukov, V. M., and Gladyshev, V. N. (1999) *J. Biol. Chem.* **274**, 33888–33897
452. Lescure, A., Gautheret, D., Carbon, P., and Krol, A. (1999) *J. Biol. Chem.* **274**, 38147–38154
453. Saito, Y., Hayashi, T., Tanaka, A., Watanabe, Y., Suzuki, M., Saito, E., and Takahashi, K. (1999) *J. Biol. Chem.* **274**, 2866–2871
454. Low, S. C., and Berry, M. J. (1996) *Trends Biochem. Sci.* **21**, 203–208
455. Suppmann, S., Persson, B. C., and Böck, A. (1999) *EMBO J.* **18**, 2284–2293
456. Ioudovitch, A., and Steinberg, S. V. (1999) *J. Mol. Biol.* **290**, 365–371
457. Forchhammer, K., Leinfelder, W., and Böck, A. (1989) *Nature (London)* **342**, 453–456
458. Kromayer, M., Wilting, R., Tormay, P., and Böck, A. (1996) *J. Mol. Biol.* **262**, 413–420
459. Liu, Z., Reches, M., and Engelberg-Kulka, H. (1999) *J. Mol. Biol.* **294**, 1073–1086
460. Rother, M., Wilting, R., Commans, S., and Böck, A. (2000) *J. Mol. Biol.* **299**, 351–358
461. Low, S. C., Grundner-Culemann, E., Harney, J. W., and Berry, M. J. (2000) *EMBO J.* **19**, 6882–6890
461a. Nasim, M. T., Jaenecke, S., Belduz, A., Kollmus, H., Flohé, L., and McCarthy, J. E. G. (2000) *J. Biol. Chem.* **275**, 14846–14852
462. Copeland, P. R., Fletcher, J. E., Carlson, B. A., Hatfield, D. L., and Driscoll, D. M. (2000) *EMBO J.* **19**, 306–314
463. Normanly, J., Masson, J.-M., Kleina, L. G., Abelson, J., and Miller, J. H. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6548–6552
464. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989) *Science* **244**, 182–188
465. Anthony-Cahill, S. J., Griffith, M. C., Noren, C. J., Suich, D. J., and Schultz, P. G. (1989) *Trends Biochem. Sci.* **14**, 400–403
466. Mendel, D., Ellman, J. A., Chang, Z., Veenstra, D. L., Kollman, P. A., and Schultz, P. G. (1992) *Science* **266**, 1798–1802
467. Chung, H.-H., Benson, D. R., and Schultz, P. G. (1993) *Science* **259**, 806–809
468. Moore, B., Persson, B. C., Nelson, C. C., Gesteland, R. F., and Atkins, J. F. (2000) *J. Mol. Biol.* **298**, 195–209
469. Hohsaka, T., Ashizuka, Y., Sasaki, H., Murakami, H., and Sisido, M. (1999) *J. Am. Chem. Soc.* **121**, 12194–12195
470. Kowal, A. K., Köhrer, C., and RajBhandary, U. L. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2268–2273
471. Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) *Science* **292**, 498–500
472. Döring, V., Mootz, H. D., Nangle, L. A., Hendrickson, T. L., de Crécy-Lagard, V., Schimmel, P., and Marlière, P. (2001) *Science* **292**, 501–504
473. Bain, J. D., Switzer, C., Chamberlin, A. R., and Benner, S. A. (1992) *Nature (London)* **356**, 537–539
474. Temple, G. F., Dozy, A. M., Roy, K. L., and Kan, Y. W. (1982) *Nature (London)* **296**, 537–540
475. Ho, Y.-S., and Kan, Y. W. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2185–2188
476. Ryoji, M., Hsia, K., and Kaji, A. (1983) *Trends Biochem. Sci.* **8**, 88–90
477. Hatfield, D., and Oroszlan, S. (1990) *Trends Biochem. Sci.* **15**, 186–190
478. Jacks, T., and Varmus, H. E. (1985) *Science* **230**, 1237–1242
479. Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988) *Nature (London)* **331**, 280–283
480. Varmus, H. (1988) *Science* **240**, 1427–1434
481. Alam, S. L., Atkins, J. F., and Gesteland, R. F. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14177–14179
482. Gesteland, R. F., and Atkins, J. F. (1996) *Ann. Rev. Biochem.* **65**, 741–768
483. Kastelein, R. A., Remaut, E., Fiers, W., and van Duin, J. (1982) *Nature (London)* **295**, 35–41
484. Dayhuff, T. J., Atkins, J. F., and Gesteland, R. F. (1986) *J. Biol. Chem.* **261**, 7491–7500
484a. Harger, J. W., Meskauskas, A., and Dinman, J. D. (2002) *Trends Biochem. Sci.* **27**, 448–454
485. Spanjaard, R. A., and van Duin, J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7967–7971
486. Klovins, J., and van Duin, J. (1999) *J. Mol. Biol.* **294**, 875–884
487. Barak, Z., Lindsley, D., and Gallant, J. (1996) *J. Mol. Biol.* **256**, 676–684
487a. Chamorro, M., Parkin, N., and Varmus, H. E. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 713–717
487b. Chen, X., Chamorro, M., Lee, S. I., Shen, L. X., Hines, J. V., Tinoco, I., Jr., and Varmus, H. E. (1995) *EMBO J.* **14**, 842–852
487c. Chen, X., Kang, H., Shen, L. X., Chamorro, M., Varmus, H. E., and Tinoco, I., Jr. (1996) *J. Mol. Biol.* **260**, 479–483
488. Giedroc, D. P., Theimer, C. A., and Nixon, P. L. (2000) *J. Mol. Biol.* **298**, 167–185
489. Marczinke, B., Fisher, R., Vidakovic, M., Bloys, A. J., and Brierley, I. (1998) *J. Mol. Biol.* **284**, 205–225
490. Marczinke, B., Hagervall, T., and Brierley, I. (2000) *J. Mol. Biol.* **295**, 179–191
491. Liphardt, J., Napthine, S., Kontos, H., and Brierley, I. (1999) *J. Mol. Biol.* **288**, 321–335
492. Kim, Y.-G., Su, L., Maas, S., O'Neill, A., and Rich, A. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14234–14239
493. Matsufuji, S., Matsufuji, T., Wills, N. M., Gesteland, R. F., and Atkins, J. F. (1996) *EMBO J.* **15**, 1360–1370
494. Adamski, F. M., Atkins, J. F., and Gesteland, R. F. (1996) *J. Mol. Biol.* **261**, 357–371
495. Herr, A. J., Gesteland, R. F., and Atkins, J. F. (2000) *EMBO J.* **19**, 2671–2680
495a. Herr, A. J., Wills, N. M., Nelson, C. C., Gesteland, R. F., and Atkins, J. F. (2001) *J. Mol. Biol.* **311**, 445–452
496. Weissmann, C., Billeter, M. A., Goodman, H. M., Hindley, J., and Weber, H. (1973) *Ann. Rev. Biochem.* **42**, 303–328
497. Beekwilder, J., Nieuwenhuizen, R., Poot, R., and van Duin, J. (1996) *J. Mol. Biol.* **256**, 8–19
498. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G., and Ysebaert, M. (1976) *Nature (London)* **260**, 500–507
499. Min Jou, W., Haegeman, G., Ysebaert, M., and Fiers, W. (1972) *Nature (London)* **237**, 82–88
500. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G. H., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, C., and Ysebaert, M. (1976) *Nature (London)* **260**, 500–507
501. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Merregaert, J., Min Jou, W., Raeymakers, A., Volckaert, G., Ysebaert, M., Van de Kerckhove, J., Nolf, F., and Van Montagu, M. (1975) *Nature (London)* **256**, 273–278
502. Jackson, R. J., Howell, M. T., and Kaminski, A. (1990) *Trends Biochem. Sci.* **15**, 477–483
503. Pilipenko, E. V., Gmyl, A. P., Maslova, S. V., Belov, G. A., Sinyakov, A. N., Huang, M., Brown, T. D. K., and Agol, V. I. (1994) *J. Mol. Biol.* **241**, 398–414
504. Gan, W., and Rhoads, R. E. (1996) *J. Biol. Chem.* **271**, 623–626

References

505. Kieft, J. S., Zhou, K., Jubin, R., Murray, M. G., Lau, J. Y. N., and Doudna, J. A. (1999) *J. Mol. Biol.* **292**, 513–529
506. Lott, W. B., Takyar, S. S., Tuppen, J., Crawford, D. H. G., Harrison, M., Sloots, T. P., and Gowans, E. J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4916–4921
506a. Guo, L., Allen, E., and Miller, W. A. (2000) *RNA* **6**, 1808–1820
507. Kaplan, S., Atherly, A. G., and Barrett, A. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 689–692
508. Cozzone, A. J. (1981) *Trends Biochem. Sci.* **6**, 108–110
509. Baracchini, E., and Bremer, H. (1988) *J. Biol. Chem.* **263**, 2597–2602
510. Murray, K. D., and Bremer, H. (1996) *J. Mol. Biol.* **259**, 41–57
511. Metzger, S., Dror,, IB, Aizenman, E., Schreiber, G., Toone, M., Friesen, J. D., Cashel, M., and Glaser, G. (1988) *J. Biol. Chem.* **263**, 15699–15704
512. Hernandez, V. J., and Cashel, M. (1995) *J. Mol. Biol.* **252**, 536–549
512a. Barker, M. M., Gaal, T., Josaitis, C. A., and Gourse, R. L. (2001) *J. Mol. Biol.* **305**, 673–688
512b. Gottesman, S., and Maurizi, M. R. (2001) *Science* **293**, 614 – 615
513. Wagner, E. G. H., and Kurland, C. G. (1980) *Biochemistry* **19**, 1234–1240
514. Nègre, D., Cortay, J.-C., Donini, P., and Cozzone, A. J. (1989) *Biochemistry* **28**, 1814–1819
515. Marianovsky, I., Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (2001) *J. Biol. Chem.* **276**, 5975–5984
516. Swarbrick, J. D., Bashtannyk, T., Maksel, D., Zhang, X.-R., Blackburn, G. M., Gayler, K. R., and Gooley, P. R. (2000) *J. Mol. Biol.* **302**, 1165–1177
517. Kim, B. K., Zamecnik, P., Taylor, G., Guo, M. J., and Blackburn, G. M. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11056–11058
518. Luo, J., Jankowski, J., Knobloch, M., van der Giet, M., Gardanis, K., Russ, T., Vahlensieck, U., Neumann, J., Schmitz, W., Tepel, M., Deng, M. C., Zidek, W., and Schlüter, H. (1999) *FASEB J.* **13**, 695–705
519. Ingram, S. W., Stratemann, S. A., and Barnes, L. D. (1999) *Biochemistry* **38**, 3649–3655
520. Cartwright, J. L., and McLennan, A. G. (1999) *J. Biol. Chem.* **274**, 8604–8610
521. Yen, T. J., Machlin, P. S., and Cleveland, D. W. (1988) *Nature (London)* **334**, 580–585
522. Hunt, T. (1988) *Nature (London)* **334**, 567–568
523. Cleveland, D. W. (1988) *Trends Biochem. Sci.* **13**, 339–343
524. Louie, K., and Dowhan, W. (1980) *J. Biol. Chem.* **255**, 1124–1127
525. Fedorov, A. N., and Baldwin, T. O. (1997) *J. Biol. Chem.* **272**, 32715–32718
525a. Patzelt, H., Rüdiger, S., Brehmer, D., Kramer, G., Vorderwülbecke, S., Schaffitzel, E., Waitz, A., Hesterkamp, T., Dong, L., Schneider-Mergener, J., Bukau, B., and Deuerling, E. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14244–14249
525b. Hartl, F. U., and Hayer-Hartl, M. (2002) *Science* **295**, 1852–1858
525c. Stoller, G., Rücknagel, K. P., Nierhaus, K. H., Schmid, F. X., Fischer, G., and Rahfeld, J.-U. (1995) *EMBO J.* **14**, 4939–4948
525d. Maier, R., Scholz, C., and Schmid, F. X. (2001) *J. Mol. Biol.* **314**, 1181–1190
525e. Richardson, A., Landry, S. J., and Georgopoulos, C. (1998) *Trends Biochem. Sci.* **23**, 138–143
525f. Keskin, O., Bahar, I., Flatow, D., Covell, D. G., and Jernigan, R. L. (2002) *Biochemistry* **41**, 491–501
525g. Flaux, J., Bertelsen, E. B., Horwich, A. L., and Wüthrich, K. (2002) *Nature (London)* **418**, 207–211
525h. Llorca, O., Martín-Benito, J., Grantham, J., Ritco-Vonsovici, M., Willison, K. R., Carrascosa, J. L., and Valpuesta, J. M. (2001) *EMBO J.* **20**, 4065–4075
526. Ruddon, R. W., and Bedows, E. (1997) *J. Biol. Chem.* **272**, 3125–3128
527. Ellis, R. J. (2000) *Trends Biochem. Sci.* **25**, 210–212
528. Sakikawa, C., Taguchi, H., Makino, Y., and Yoshida, M. (1999) *J. Biol. Chem.* **274**, 21251–21256
529. Ma, J., Sigler, P. B., Xu, Z., and Karplus, M. (2000) *J. Mol. Biol.* **302**, 303–313
530. Schäfer, U., Beck, K., and Müller, M. (1999) *J. Biol. Chem.* **274**, 24567–24574
531. Jones, C. H., Danese, P. N., Pinkner, J. S., Silhavy, T. J., and Hultgren, S. J. (1997) *EMBO J.* **16**, 6394–6406
532. Frydman, J., and Höhfeld, J. (1997) *Trends Biochem. Sci.* **22**, 87–92
533. Rüdiger, S., Schneider-Mergener, J., and Bukau, B. (2001) *EMBO J.* **20**, 1042–1050
534. Wilson, M. R., and Easterbrook-Smith, S. B. (2000) *Trends Biochem. Sci.* **25**, 95–98
535. Zha, J., Weiler, S., Oh, K. J., Wei, M. C., and Korsmeyer, S. J. (2000) *Science* **290**, 1761–1765
536. Schubert, U., Antón, L. C., Gibbs, J., Norbury, C. C., Yewdell, J. W., and Bennink, J. R. (2000) *Nature (London)* **404**, 770–774
537. Turner, G. C., and Varshavsky, A. (2000) *Science* **289**, 2117–2120
538. Scheffner, M., and Whitaker, N. J. (2001) *Nature (London)* **410**, 882–883
539. Wickner, S., Maurizi, M. R., and Gottesman, S. (1999) *Science* **286**, 1888–1893
540. Bonifacino, J. S. (1996) *Nature (London)* **384**, 405–406
541. Arfin, S. M., and Bradshaw, R. A. (1988) *Biochemistry* **27**, 7979–7984
542. Rajagopalan, P. T. R., Grimme, S., and Pei, D. (2000) *Biochemistry* **39**, 779–790
542a. Deng, H., Callender, R., Zhu, J., Nguyen, K. T., and Pei, D. (2002) *Biochemistry* **41**, 10563–10569
543. Solbiati, J., Chapman-Smith, A., Miller, J. L., Miller, C. G., and Cronan, J. E., JR. (1999) *J. Mol. Biol.* **290**, 607–614
544. Giglione, C., Serero, A., Pierre, M., Boisson, B., and Meinnel, T. (2000) *EMBO J.* **19**, 5916–5929
544a. Serero, A., Giglione, C., and Meinnel, T. (2001) *J. Mol. Biol.* **314**, 695–708
545. Boissel, J.-P., Kasper, T. J., and Bunn, H. F. (1988) *J. Biol. Chem.* **263**, 8443–8449
546. Ben-Bassat, A., and Bauer, K. (1987) *Nature (London)* **326**, 315
547. Polevoda, B., and Sherman, F. (2000) *J. Biol. Chem.* **275**, 36479–36482
548. Strauch, A. R., and Rubenstein, P. A. (1984) *J. Biol. Chem.* **259**, 7224–7229
549. Farazi, T. A., Waksman, G., and Gordon, J. I. (2001) *Biochemistry* **40**, 6335–6343
550. Utsumi, T., Sato, M., Nakano, K., Takemura, D., Iwata, H., and Ishisaka, R. (2001) *J. Biol. Chem.* **276**, 10505–10513
551. Wilcox, C., Hu, J.-S., and Olson, E. N. (1987) *Science* **238**, 1275–1278
552. Towler, D. A., Gordon, J. I., Adams, S. P., and Glaser, L. (1988) *Ann. Rev. Biochem.* **57**, 69–99
553. Wu, Z., Demma, M., Strickland, C. L., Radisky, E. S., Poulter, C. D., Le, H. V., and Windsor, W. T. (1999) *Biochemistry* **38**, 11239–11249
553a. Matsuyama, S., Yokota, N., and Tokuda, H. (1997) *EMBO J.* **16**, 6947 – 6955
554. Aletta, J. M., Cimato, T. R., and Ettinger, M. J. (1998) *Trends Biochem. Sci.* **23**, 89–91
555. Zobel-Thropp, P., Gary, J. D., and Clarke, S. (1998) *J. Biol. Chem.* **273**, 29283–29286
556. Chapman-Smith, A., and Cronan, J. E., Jr. (1999) *Trends Biochem. Sci.* **24**, 359–363
557. Campeau, E., and Gravel, R. A. (2001) *J. Biol. Chem.* **276**, 12310–12316
558. Craig, A. G., Jimenez, E. C., Dykert, J., Nielsen, D. B., Gulyas, J., Abogadie, F. C., Porter, J., Rivier, J. E., Cruz, L. J., Olivera, B. M., and McIntosh, J. M. (1997) *J. Biol. Chem.* **272**, 4689–4698
559. Dierks, T., Lecca, M. R., Schlotterhose, P., Schmidt, B., and von Figura, K. (1999) *EMBO J.* **18**, 2084–2091
560. Stubbe, J. (1996) *Science* **274**, 1152–1153
561. Kelleher, N. L., Hendrickson, C. L., and Walsh, C. T. (1999) *Biochemistry* **38**, 15623–15630
562. Kupke, T., Kempter, C., Gnau, V., Jung, G., and Götz, F. (1994) *J. Biol. Chem.* **269**, 5653–5659
563. Rapoport, T. A. (1992) *Science* **258**, 931–935
564. Corsi, A. K., and Schekman, R. (1996) *J. Biol. Chem.* **271**, 30299–30302
565. Freymann, D. M., Keenan, R. J., Stroud, R. M., and Walter, P. (1997) *Nature (London)* **385**, 361–364
566. Montoya, G., Svensson, C., Luirink, J., and Sinning, I. (1997) *Nature (London)* **385**, 365–368
567. Peluso, P., Herschlag, D., Nock, S., Freymann, D. M., Johnson, A. E., and Walter, P. (2000) *Science* **288**, 1640–1643
568. Batey, R. T., Rambo, R. P., Lucast, L., Rha, B., and Doudna, J. A. (2000) *Science* **287**, 1232–1239
569. Batey, R. T., Sagar, M. B., and Doudna, J. A. (2001) *J. Mol. Biol.* **307**, 229–246
570. Diener, J. L., and Wilson, C. (2000) *Biochemistry* **39**, 12862–12874
571. Walter, P., Keenan, R., and Schmitz, U. (2000) *Science* **287**, 1212–1213
571a. Peluso, P., Shan, S.-o, Nock, S., Herschlag, D., and Walter, P. (2001) *Biochemistry* **40**, 15224–15233
571b. Pool, M. R., Stumm, J., Fulga, T. A., Sinning, I., and Dobberstein, B. (2002) *Science* **297**, 1345–1348
572. de Leeuw, E., te Kaat, K., Moser, C., Menestrina, G., Demel, R., de Kruijff, B., Oudega, B., Luirink, J., and Sinning, I. (2000) *EMBO J.* **19**, 531–541
573. Fulga, T. A., Sinning, I., Dobberstein, B., and Pool, M. R. (2001) *EMBO J.* **20**, 2338–2347
574. Bibi, E., Herskovits, A. A., Bochkareva, E. S., and Zelazny, A. (2001) *Trends Biochem. Sci.* **26**, 15–16
575. Lütcke, H. (1995) *Eur. J. Biochem.* **228**, 531–550
575a. Huang, Q., Abdulrahman, S., Yin, J., and Zwieb, C. (2002) *Biochemistry* **41**, 11362 – 11371
575b. Hainzl, T., Huang, S., and Sauer-Eriksson, A. E. (2002) *Nature (London)* **417**, 767 – 771
576. Weichenrieder, O., Wild, K., Strub, K., and Cusack, S. (2000) *Nature (London)* **408**, 167–173
576a. Beckman, R. (2001) *Cell* **107**, 361–72
577. Young, B. P., Craven, R. A., Reid, P. J., Willer, M., and Stirling, C. J. (2001) *EMBO J.* **20**, 262–271
578. Römisch, K. (2001) *Trends Biochem. Sci.* **26**, 13
579. Beswick, V., Baleux, F., Huynh-Dinh, T., Képès, F., Neumann, J.-M., and Sanson, A. (1996) *Biochemistry* **35**, 14717–14724
580. Tschantz, W. R., Paetzel, M., Cao, G., Suciu, D., Inouye, M., and Dalbey, R. E. (1995) *Biochemistry* **34**, 3935–3941
581. Chatterjee, S., Suciu, D., Dalbey, R. E., Kahn, P. C., and Inouye, M. (1995) *J. Mol. Biol.* **245**, 311–314

References

582. VanValkenburgh, C., Chen, X., Mullins, C., Fang, H., and Green, N. (1999) *J. Biol. Chem.* **274**, 11519–11525
582a. Weihofen, A., Binns, K., Lemberg, M. K., Ashman, K., and Martoglio, B. (2002) *Science* **296**, 2215 – 2218
583. Suzuki, T., Yan, Q., and Lennarz, W. J. (1998) *J. Biol. Chem.* **273**, 10083–10086
584. Netzer, W. J., and Hartl, F. U. (1997) *Nature (London)* **388**, 343–349
585. Wickner, W., and Leonard, M. R. (1996) *J. Biol. Chem.* **271**, 29514–29516
586. Kim, J., Miller, A., Wang, L., Müller, J. P., and Kendall, D. A. (2001) *Biochemistry* **40**, 3674–3680
587. Verner, K., and Schatz, G. (1988) *Science* **241**, 1307–1313
588. van der Does, C., Manting, E. H., Kaufmann, A., Lutz, M., and Driessen, A. J. M. (1998) *Biochemistry* **37**, 201–210
589. Collinson, I., Breyton, C., Duong, F., Tziatzios, C., Schubert, D., Or, E., Rapoport, T., and Kühlbrandt, W. (2001) *EMBO J.* **20**, 2462–2471
590. Sianidis, G., Karamanou, S., Vrontou, E., Boulias, K., Repanas, K., Kyrpides, N., Politou, A. S., and Economou, A. (2001) *EMBO J.* **20**, 961–970
591. Kluger, R., and Smyth, T. (1981) *J. Am. Chem. Soc.* **103**, 1216–1218
592. Yahr, T. L., and Wickner, W. T. (2000) *EMBO J.* **19**, 4393–4401
593. Driessen, A. J. M. (1992) *Trends Biochem. Sci.* **17**, 219–223
594. Dalbey, R. E., and Robinson, C. (1999) *Trends Biochem. Sci.* **24**, 17–22
595. Delgado-Partin, V. M., and Dalbey, R. E. (1998) *J. Biol. Chem.* **273**, 9927–9934
596. Gafvelin, G., Sakaguchi, M., Andersson, H., and von Heijne, G. (1997) *J. Biol. Chem.* **272**, 6119–6127
597. Rietveld, A. G., Koorengevel, M. C., and de Kruijff, B. (1995) *EMBO J.* **14**, 5506–5513
598. Jongbloed, J. D. H., Martin, U., Antelmann, H., Hecker, M., Tjalsma, H., Venema, G., Bron, S., van Dijl, J. M., and Müller, J. (2000) *J. Biol. Chem.* **275**, 41350–41357
599. Sambasivarao, D., Turner, R. J., Simala-Grant, J. L., Shaw, G., Hu, J., and Weiner, J. H. (2000) *J. Biol. Chem.* **275**, 22526–22531
600. Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., and Hughes, C. (2000) *Nature (London)* **405**, 914–919
601. Buchanan, S. K. (2001) *Trends Biochem. Sci.* **26**, 3–6
602. Reits, E. A. J., Vos, J. C., Grommé, M., and Neefjes, J. (2000) *Nature (London)* **404**, 774–778
603. Samuelson, J. C., Chen, M., Jiang, F., Möller, I., Wiedmann, M., Kuhn, A., Phillips, G. J., and Dalbey, R. E. (2000) *Nature (London)* **406**, 637–640
604. Tokatlidis, K., and Schatz, G. (1999) *J. Biol. Chem.* **274**, 35285–35288
605. Schatz, G. (1998) *Nature (London)* **395**, 439–440
606. McNew, J. A., and Goodman, J. M. (1996) *Trends Biochem. Sci.* **21**, 54–58
607. Subramani, S. (1996) *J. Biol. Chem.* **271**, 32483–32486
608. Hell, K., Neupert, W., and Stuart, R. A. (2001) *EMBO J.* **20**, 1281–1288
609. von Heijne, G. (1989) *Nature (London)* **341**, 456–458
610. Kuhn, A., Wickner, W., and Kreil, G. (1986) *Nature (London)* **322**, 335–339
611. Bibi, E. (1998) *Trends Biochem. Sci.* **23**, 51–55
612. Dalbey, R. E., Kahn, A., and Wickner, W. (1987) *J. Biol. Chem.* **262**, 13241–13245
613. Yamane, K., Ichehara, S., and Mizushima, S. (1987) *J. Biol. Chem.* **263**, 2358–2362
614. Meijer, A. B., Spruijt, R. B., Wolfs, C. J. A. M., and Hemminga, M. A. (2001) *Biochemistry* **40**, 5081–5086
615. Soekarjo, M., Eisenhawer, M., Kuhn, A., and Vogel, H. (1996) *Biochemistry* **35**, 1232–1241
616. Moore, K. E., and Miura, S. (1988) *J. Biol. Chem.* **263**, 11575–11583
617. von Heijne, G., Wickner, W., and Dalbey, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3363–3366
617a. Anderson, R. G. W., and Jacobson, K. (2002) *Science* **296**, 1821–1825
618. Thomas-Reetz, A. C., and De Camilli, P. (1994) *FASEB J.* **8**, 209–216
619. Rothman, J. E., and Wieland, F. T. (1996) *Science* **272**, 227–234
620. Nicholls, D. G. (1994) *Proteins, Transmitters and Synapses*, Blackwell Scientific Publications, Oxford
621. Bean, A. J., Zhang, X., and Hökfelt, T. (1994) *FASEB J.* **8**, 630–638
622. Huttner, W. B., Gerdes, H.-H., and Rosa, P. (1991) *Trends Biochem. Sci.* **16**, 27–30
623. Guo, W., Grant, A., and Novick, P. (1999) *J. Biol. Chem.* **274**, 23558–23564
624. Peters, C., Bayer, M. J., Bühler, S., Andersen, J. S., Mann, M., and Mayer, A. (2001) *Nature (London)* **409**, 581–588
624a. Pain, R., ed. (2000) *Mechanisms of Protein Folding*, 2nd ed., Oxford Univ. Press, London
625. Holst, B., Bruun, A. W., Kielland-Brandt, M. C., and Winther, J. R. (1996) *EMBO J.* **15**, 3538–3546
626. Pirkl, F., and Buchner, J. (2001) *J. Mol. Biol.* **308**, 795–806
626a. Gilbert, H. F. (1997) *J. Biol. Chem.* **272**, 29399–29402
627. Qiao, Z.-S., Guo, Z.-Y., and Feng, Y.-M. (2001) *Biochemistry* **40**, 2662–2668
628. Tu, B. P., Ho-Schleyer, S. C., Travers, K. J., and Weissman, J. S. (2000) *Science* **290**, 1571–1573
629. Miranker, A., Robinson, C. V., Radford, S. E., and Dobson, C. M. (1996) *FASEB J.* **10**, 93–101
630. Balbach, J., Steegborn, C., Schindler, T., and Schmid, F. X. (1999) *J. Mol. Biol.* **285**, 829–842
631. Arrington, C. B., Teesch, L. M., and Robertson, A. D. (1999) *J. Mol. Biol.* **285**, 1265–1275
632. Rumbley, J., Hoang, L., Mayne, L., and Englander, S. W. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 105–112
633. Steegborn, C., Schneider-Hassloff, H., Zeeb, M., and Balbach, J. (2000) *Biochemistry* **39**, 7910–7919
634. Arai, M., Ikura, T., Semisotnov, G. V., Kihara, H., Amemiya, Y., and Kuwajima, K. (1998) *J. Mol. Biol.* **275**, 149–162
635. Panick, G., Malessa, R., Winter, R., Rapp, G., Frye, K. J., and Royer, C. A. (1998) *J. Mol. Biol.* **275**, 389–402
636. Radford, S. E. (2000) *Trends Biochem. Sci.* **25**, 611–618
637. Anfinsen, C. B. (1973) *Science* **181**, 223–230
638. Dill, K. A. (1990) *Biochemistry* **29**, 7133–7155
639. Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M., and Karplus, M. (2000) *Trends Biochem. Sci.* **25**, 331–339
640. Baker, D. (2000) *Nature (London)* **405**, 39–42
641. Clarke, D. T., Doig, A. J., Stapley, B. J., and Jones, G. R. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7232–7237
642. Nölting, B., Golbik, R., Neira, J. L., Soler-Gonzalez, A. S., Schreiber, G., and Fersht, A. R. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 826–830
643. Zhou, Y., and Karplus, M. (1999) *J. Mol. Biol.* **293**, 917–951
644. Damaschun, G., Damaschun, H., Gast, K., and Zirwer, D. (1999) *J. Mol. Biol.* **291**, 715–725
645. Pace, C. N. (1990) *Trends Biochem. Sci.* **15**, 14–17
646. Rees, D. C., and Robertson, A. D. (2001) *Protein Sci.* **10**, 1187–1194
647. Andersen, N. H., Cort, J. R., Liu, Z., Sjoberg, S. J., and Tong, H. (1996) *J. Am. Chem. Soc.* **118**, 10309–10310
648. Swint-Kruse, L., and Robertson, A. D. (1995) *Biochemistry* **34**, 4724–4732
649. Yang, A.-S., and Honig, B. (1993) *J. Mol. Biol.* **231**, 459–474
650. Yang, A.-S., and Honig, B. (1994) *J. Mol. Biol.* **237**, 602–614
651. Pace, C. N. (2001) *Biochemistry* **40**, 310–313
652. Sippl, M. J. (1996) *J. Mol. Biol.* **260**, 644–648
653. Shortle, D. (1996) *FASEB J.* **10**, 27–34
654. Häckel, M., Hinz, H.-J., and Hedwig, G. R. (1999) *J. Mol. Biol.* **291**, 197–213
655. Makhatadze, G. I., and Privalov, P. L. (1996) *Protein Sci.* **5**, 507–510
655a. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) *Biochemistry* **41**, 6575 – 6582
655b. Dobson, C. M. (1999) *Trends Biochem. Sci.* **24**, 329 – 332
655c. Stevens, F. J., Pokkuluri, P. R., and Schiffer, M. (2000) *Biochemistry* **39**, 15291 – 15296
655d. Ellis, R. J., and Pinheiro, T. J. T. (2002) *Nature (London)* **416**, 483 – 484
655e. Oxenoid, K., Sönnichsen, F. D., and Sanders, C. R. (2001) *Biochemistry* **40**, 5111 – 5118
656. Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P. M., Huibregtse, J. M., and Pavletich, N. P. (1999) *Science* **286**, 1321–1326
657. Reyes, J. C. (2001) *Trends Biochem. Sci.* **26**, 18–20
658. Johnston, S. C., Riddle, S. M., Cohen, R. E., and Hill, C. P. (1999) *EMBO J.* **18**, 3877–3887
659. Whitby, F. G., Masters, E. I., Kramer, L., Knowlton, J. R., Yao, Y., Wang, C. C., and Hill, C. P. (2000) *Nature (London)* **408**, 115–120
660. Joazeiro, C. A. P., and Hunter, T. (2000) *Science* **289**, 2061–2062
661. Satoh, K., Sasajima, H., Nyoumura, K.-i, Yokosawa, H., and Sawada, H. (2001) *Biochemistry* **40**, 314–319
662. Byrd, C., Turner, G. C., and Varshavsky, A. (1998) *EMBO J.* **17**, 269–277
663. Davydov, I. V., and Varchavsky, A. (2000) *J. Biol. Chem.* **275**, 22931–22941
663a. Marx, J. (2002) *Science* **297**, 1792 – 1794
663b. Verma, R., Aravind, L., Oania, R., McDonald, W. H., Yates, J. R., III, Koonin, E. V., and Deshaies, R. J. (2002) *Science* **298**, 611 – 615
664. Kuttler, C., Nussbaum, A. K., Dick, T. P., Rammensee, H.-G., Schild, H., and Hadeler, K.-P. (2000) *J. Mol. Biol.* **298**, 417–429
665. Fineschi, B., and Miller, J. (1997) *Trends Biochem. Sci.* **22**, 377–382
666. Cascio, P., Hilton, C., Kisselev, A. F., Rock, K. L., and Goldberg, A. L. (2001) *EMBO J.* **20**, 2357–2366
667. Schild, H., and Rammensee, H.-G. (2000) *Nature (London)* **404**, 709–710
667a. Cyr, D. M., Höhfeld, J., and Patterson, C. (2002) *Trends Biochem. Sci.* **27**, 368 – 375
668. Dell'Angelica, E. C., Mullins, C., Caplan, S., and Bonifacino, J. S. (2000) *FASEB J.* **14**, 1265–1278
669. Jentsch, S., and Ulrich, H. D. (1998) *Nature (London)* **395**, 321–323
670. Wolf, Y. I., Grishin, N. V., and Koonin, E. V. (2000) *J. Mol. Biol.* **299**, 897–905
671. Simons, K. T., Strauss, C., and Baker, D. (2001) *J. Mol. Biol.* **306**, 1191–1199
672. Takada, S. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11698–11700
673. Abbott, A. (1999) *Nature (London)* **402**, 715–720

References

674. Pradet-Balade, B., Boulmé, F., Beug, H., Müllner, E. W., and Garcia-Sanz, J. A. (2001) *Trends Biochem. Sci.* **26**, 225–229
675. Attwood, T. K., and Parry-Smith, D. J. (1999) *Introduction to Bioinformatics*, Longman, Green, New York
676. Spengler, S. J. (2000) *Science* **287**, 1221–1223
677. Attwood, T. K. (2000) *Science* **290**, 471–473
678. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9390–9395
679. MacBeath, G., and Schreiber, S. L. (2000) *Science* **289**, 1760–1763
680. Horn, D. M., Zubarev, R. A., and McLafferty, F. W. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10313–10317
680a. Rajagopol, I., and Ahern, K. (2001) *Science* **294**, 2571 – 2573
681. Miyashita, M., Presley, J. M., Buchholz, B. A., Lam, K. S., Lee, Y. M., and Vogel, J. S. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4403–4408
682. Mann, M., and Pandey, A. (2001) *Trends Biochem. Sci.* **26**, 54–61
683. Baldauf, S. L., Roger, A. J., Wenk-Siefert, I., and Doolittle, W. F. (2000) *Science* **290**, 972–977
684. Gerstein, M. (1997) *J. Mol. Biol.* **274**, 562–576
685. Gómez, S. M., Nishio, J. N., Faull, K. F., and Whitelegge, J. P. (2002) *Molecular & Cellular Proteomics* **1**, 46 – 59
686. Michels, D. A., Hu, S., Schoenherr, R. M., Eggertson, M. J., and Dovichi, N. J. (2002) *Molecular & Cellular Proteomics* **1**, 69 – 74
687. Smolka, M., Zhou, H., and Aebersold, R. (2002) *Molecular & Cellular Proteomics* **1**, 19 – 29
688. Herbert, B. R., Harry, J. L., Packer, N. H., Gooley, A. A., Pedersen, S. K., and Williams, K. L. (2001) *Trends in Biotechnology* **19**, S3 – S9
689. Ge, Y., Lawhorn, B. G., ElNaggar, M., Strauss, E., Park, J.-H., Begley, T. P., and McLafferty, F. W. (2002) *J. Am. Chem. Soc.* **124**, 672 – 678
690. Griffin, T. J., and Aebersold, R. (2001) *J. Biol. Chem.* **276**, 45497 – 45500
691. Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R. A., Gerstein, M., and Snyder, M. (2001) *Science* **293**, 2101 – 2105
692. Kodadek, T. (2002) *Trends Biochem. Sci.* **27**, 295 – 300
693. Smith, G. P., Patel, S. U., Windass, J. D., Thornton, J. M., Winter, G., and Griffiths, A. D. (1998) *J. Mol. Biol.* **277**, 317 – 332
694. Blagoev, B., Pandey, A. (2001) *Trends Biochem. Sci.* **26**, 639 – 641
695. Yee, A., and 23 other authors. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1825 – 1830
696. Riek, R., Fiauz, J., Bertelsen, E. B., Horwich, A. L., and Wüthrich, K. (2002) *J. Am. Chem. Soc.* **124**, 12144 – 12153
697. Tugarinov, V., Muhandiram, R., Ayed, A., and Kay, L. E. (2002) *J. Am. Chem. Soc.* **124**, 10025 – 10035
698. Jhoti, H. (2001) *Trends in Biotechnology* **19**, S67 – S71
699. Thornton, J. (2001) *Trends Biochem. Sci.* **26**, 88 – 89
700. Koonin, E. V., Wolf, Y. I., and Karev, G. P. (2002) *Nature (London)* **420**, 218 – 223
701. Mallis, R. J., Brazin, K., Jez, J. M., Wilson, E. K., Dieckmann, G. R., Robic, S., and Harrahy, J. (2001) *Trends Biochem. Sci.* **26**, 642 – 643
702. Zhang, C., and Kim, S.-H. (2000) *J. Mol. Biol.* **299**, 1075 – 1089
703. Wimley, W. C. (2002) ***Protein Sci.*** 11, 301 – 312
704. Richardson, J. S., and Richardson, D. C. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2754 – 2759
705. Calladine, C. R., Sharff, A., and Luisi, B. (2001) *J. Mol. Biol.* **305**, 603 – 618
706. Chen, C. P., Kernytsky, A., and Rost, B. (2002) *Protein Sci.* **11**, 2774 – 2791
707. Chen, C. P., and Rost, B. (2002) *Protein Sci.* **11**, 2766 – 2773
708. Baker, D., and Sali, A. (2001) *Science* **294**, 93 – 96
709. Marsden, R. L., McGuffin, L. J., and Jones, D. T. (2002) *Protein Sci.* **11**, 2814 – 2824
710. Jiang, W. Baker, M. L., Ludtke, S. J., and Chiu, W. (2001) *J. Mol. Biol.* **308**, 1033 – 1044
711. Taylor, W. R. (2002) *Nature (London)* **416**, 657 – 660
712. Apic, G., Gough, J., and Teichmann, S. A. (2001) *J. Mol. Biol.* **310**, 311 – 325
713. Lee, K. K., Fitch, C. A., and Garcia-Moreno, E. B. (2002) *Protein Sci.* **11**, 1004 – 1016
714. Berezovsky, I. N., and Trifonov, E. N. (2002) *J. Mol. Biol.* **307**, 1419 – 1426
715. Trabi, M., and Craik, D. J. (2002) *Trends Biochem. Sci.* **27**, 132 – 138
716. Shao, Z., and Vollrath, F. (2002) *Nature (London)* **418**, 741
717. Asakura, T., Ashida, J., Yamane, T., Kameda, T., Nakazawa, Y., Ohgo, K., and Komatsu, K. (2001) *J. Mol. Biol.* **306**, 291 – 305
718. van Beek, J. D., Hess, S., Vollrath, F., and Meier, B. H. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10266 – 10271
719. Vollrath, F., and Knight, D. P. (2001) *Nature (London)* **410**, 541 – 548
720. Sprinzak, E., and Margalit, H. (2001) *J. Mol. Biol.* **311**, 681 – 692
721. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., and Bork, P. (2002) *Nature (London)* **417**, 399 – 403
722. Gavin, A.-C., and 37 other authors. (2002) *Nature (London)* **415**, 141 – 147
723. Stagljar, I., and Fields, S. (2002) *Trends Biochem. Sci*. **27**, 559 – 563
724. Tong, A. H. Y., Drees, B., Nardelli, G., Bader, G. D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., Paoluzi, S., Quondam, M., Zucconi, A., Hogue, C. W. V., Fields, S., Boone, C., and Cesareni, G. (2002) *Science* **295**, 321 – 324
725. Cornish-Bowden, A., and Cárdenas, M. L. (2001) *Trends Biochem. Sci.* **26**, 463 – 465
726. Grant, S. G. N., and Husi, H. (2001) *Trends in Biotechnology* **19**, S49 – S54
727. Natsume, T., Nakayama, H., and Isobe, T. (2001) *Trends in Biotechnology* **19**, S28 – S33
728. Ray, P., Pimenta, H., Paulmurugan, R., Berger, F., Phelps, M. E. Iyer, M., and Gambhir, S. S. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3105 – 3110
729. Florens, L., and 17 other authors. (2002) *Nature (London)* **419**, 520 – 526
730. Koller, A., Washburn, M. P., Lange, B. M., Andon, N. L., Deciu, C., Haynes, P. A., Hays, L. Schieltz, D., Ulaszek, R., Wei, J., Wolters, D., and Yates, J. R., III. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11969 – 11974
731. Remy, I., and Michnick, S. W. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7678 – 7683

Study Questions for chapters 28 and 29

1. Describe the role of sigma factors (σ) in transcription by prokaryotic RNA polymerases. What is the effect of the release of σ from the holoenzyme once transcription has been initiated. How would a mutation that prevents a σ factor from dissociating from core RNA polymerase affect the rate of transcription?

2. Explain how histidine biosynthesis is controlled in *E. coli*, a bacterium that has no *his* repressor.

3. One mechanism of transcriptional control in prokaryotes, especially of several operons controlling the biosynthesis of amino acids, is **attenuation**. Briefly describe the mechanism of attenuation. How does the supply of amino acid in the cell affect the process?

4. Is transcription attenuation likely to be an important mechanism of transcriptional regulation in eukaryotic cells?

5. Discuss two main DNA-recognition motifs found in eukaryotic transcription factors. Describe their structures, indicate how they bind to DNA, and discuss how each specifically recognizes its DNA binding site.

Study Questions

6. How can a DNA enhancer sequence located as many as several thousand base pairs from a gene transcription start site influence transcription even if its orientation is reversed?

7. Some eukaryotic DNA viruses code for two or more mRNA transcripts of differing lengths from the same region on the DNA. Suggest an explanation. How do you expect the two translation products of these mRNAs to differ?

8. High salt concentrations weaken the interaction of histones with DNA but have little effect on the binding of many regulatory proteins. Explain this observation in terms of the modes of interaction of the two types of protein.

9. Discuss the changes that must be made in a typical eukaryotic structural gene to allow its protein product to be synthesized in bacteria.

10. List the different types of covalent modification that may be made to tRNA. To ribosomal RNA. To messenger RNA.

11. List various small RNAs and their functions within cells.

12. Some amino acids utilize only one codon of the 64 in the genetic code. Other amino acids use as many as six codons (Tables 5-5, 5-6). What advantages to a cell is provided by utilization of several codons for a single amino acid?

13. In what ways is the genetic code not quite "universal?" What is meant by "editing" of mRNA?

14. Why is it necessary to have "adapters" in the form of tRNAs to read the genetic code during translation?

15. Most nonsense suppressor genes are mutants of tRNA genes. In view of this fact, how can cells survive the presence of such mutations?

16. Explain how the protein synthesizing machinery is able to differentiate the initiation AUG codon from an internal AUG (methionine) codon in prokaryotes. How is this accomplished in eukaryotes?

17. The amino acid sequence of a mature protein sometimes differs from that deduced from the DNA nucleotide sequence of the structural gene for that protein. Discuss three ways by which this may occur.

18. Write out in detail, using structural formulas, the chemical mechanism of synthesis of an aminoacyl-tRNA and of incorporation of the aminoacyl group into a peptide chain being formed by a ribosome.

19. a) Calculate the minimum number of ATP equivalents consumed in the biosynthesis of a 300-amino acid *E. coli* protein, having the N-terminal sequence Ala-Ser-Val-Tyr, from the free amino acids.

    b) Much of this energy involves hydrolysis of GTP. What is the role of this GTP hydrolysis in protein synthesis?

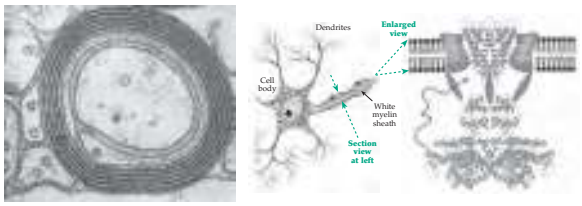
20. How do the polypeptide products produced in the presence of (a) puromycin and (b) streptomycin differ from polypeptides synthesized in the absence of these inhibitors? Explain your answer.

21. What is the significance to protein synthesis of each of the following?
    Shine–Dalgarno sequence
    Signal recognition particle
    proteasome

22. How can useful antibiotics that act on ribosomes kill bacteria but not people?

23. Compare termination of translation in bacteria and in eukaryotes.

24. List some types of error that are likely to be made during protein synthesis. What mechanisms have cells developed to deal with these?

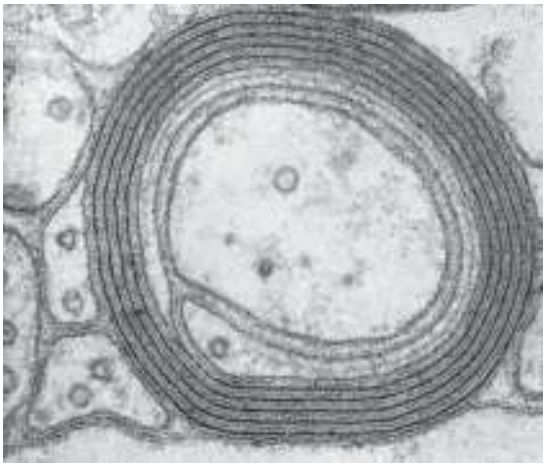


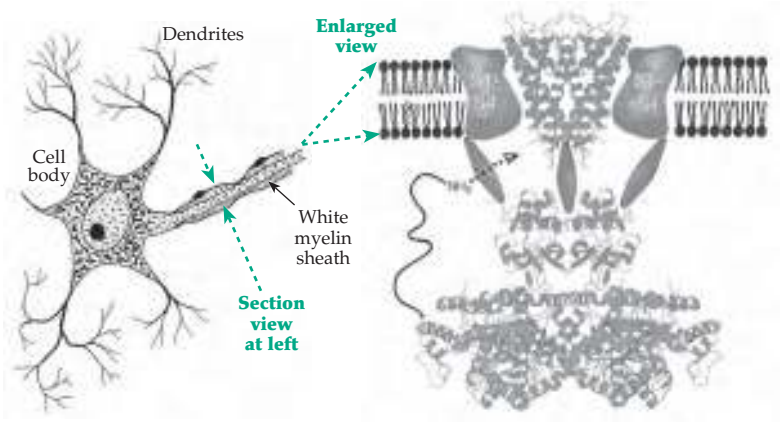
Center: Diagram of the cell body of a neuron with dendrites and a short section of its long myelinated axon (see Fig. 30-8). Left: Electron micrograph of a thin section through an axon showing the myelin sheath formed by the wrapping of the plasma membrane of a neuroglial cell around the axon (see p. 390 and Fig. 30-9). Right: Model of a voltage-regulated K^+ channel in the cell membrane of an axon. The pore, which is formed from four α-subunits, is represented by that of the bacterial pore shown in Fig. 8-21. Also shown are an inner cytoplasmic activation gate consisting of four β-subunits, which are proposed to form ball-and-chain devices that can close the pores in response to voltage changes. From Zhou *et al.* See Fig. 30-18.

Contents

Chemical Communication Between Cells

30





The regulation of growth and metabolism of complex multicellular organisms depends heavily upon chemical messages sent between cells. This includes secretion of hormones into the circulatory system,[1–3] chemical transfer of information through communicating cell junctions, and passage of signals between neurons in the brain. This chapter deals with these matters and also with communication between organisms, i.e., with the biochemistry of ecological relationships. Embryonic development and differentiation of tissues also require communication between cells as does the functioning of the immune system. These topics are considered in Chapters 31 and 32.

A. The Hormones

The term hormone has traditionally been applied to substances synthesized in and secreted by one tissue and which act to influence distant target organs or tissues. However, many peptide hormones also act as neurotransmitters, passing across very short gaps between cells. In addition, many chemical messengers, including the peptide growth factors, act more locally. Looking at lower invertebrates as simple as *Hydra*, we find peptides resembling our own hormones and neurotransmitters. These are secreted by neuroendocrine cells of *Hydra* and diffuse throughout the body. In higher animals hormones regulate the concentrations of nutrients such as glucose and of ions such as Ca^{2+} and phosphate in the blood. They control the volume and osmotic pressure of body fluids, as well as digestion, growth, reproduction, and responses to stress.

1. Receptors, Feedback Loops, and Cascades

Every hormone must have one or more receptors, most of which are proteins. These may be found embedded in the outer surface of the plasma membrane, in the cytoplasm, or in the cell nucleus. Binding of a hormone to its receptor often elicits both a rapid response and a slower one. For example, we have seen that glucagon, adrenaline, and vasopressin bind to cell surface receptors and promote the synthesis of cyclic AMP (Fig. 11-4). The cAMP induces rapid chemical modifications of many proteins. Some of these may diffuse into the nucleus and affect transcription of genes, a slower response. Insulin (Chapter 11, Section G) also exerts both rapid and slower responses.

Receptor types. Many different kinds of protein can serve as hormone receptors. Some of these are discussed in Chapter 11. The most abundant are the G protein-coupled 7-helix receptors[4–5c] such as that of a β adrenergic receptor pictured in Fig. 11-6. Glucagon, adrenaline, ACTH, and gastrin are a few of the hormones that bind to receptors of this type. Similar receptors respond to light (rhodopsin; Chapter 23) and over 1000 different 7-helix receptors respond to smell and taste. The G proteins and their controlling cycles, Eq. 11-10,[5–7] have also been considered in Chapter 11. The reality of the dissociation and reassociation of the α and βγ subunits in response to binding of a hormone has been demonstrated in living cells by the use of fluorescence resonance energy transfer (FRET).[8] Not all receptors activate G proteins. One large group of membrane-associated receptors have single transmembrane helices but require dimerization to be effective.

The bacterial chemoreceptor (Figs. 11-8 and 19-5) has a very small ligand-binding domain and a larger internal domain that activates a histidine kinase. Many growth-factor receptors, including the insulin receptor (Figs. 11-11, 11-12), have internal domains with protein tyrosine kinase activity.

In contrast, steroid hormones, thyroxine, and retinoids bind to internal receptors. In 1968, Gorski *et al.*[9] and Jensen *et al.*[10,11] proposed independently that steroid hormone receptors in the cytoplasm bind incoming steroid molecules and after an "activation" step carry the hormone into the nucleus, where the hormone–receptor complex would bind at many sites in the chromatin inducing transcription of selected genes.[12] Doubt has been cast on the assumption that the steroid hormone receptors must bind hormone initially in the cytoplasm. However, the role of steroid receptors in regulating transcription is well established (see discussion in Chapter 22, Section E,5; Chapter 28, Section C,6).

Feedback loops. Maintenance of a steady state within an organism depends upon numerous negative-feedback loops. Hormones assist in adjusting reaction rates to maintain a steady state when conditions are changed. For example, blood glucose rises after a meal. This increase is sensed in the pancreatic beta cells (pp. 998, 999), which release insulin. The released insulin promotes uptake of glucose by cells and its conversion into glycogen and lipid stores. When the glucose level falls, inhibitory mechanisms that decrease insulin release are allowed to operate.

Similar regulatory loops can be traced for nearly all hormones. Sometimes they involve several stages and involve sensing devices in the central nervous system. In such cases neural impulses stimulate the **hypothalamus** of the brain (Fig. 30-1) to release **neurohormones**, which travel to the anterior lobe of the pituitary gland. The pituitary, in turn, releases hormones such as **corticotropin** (adrenocorticotropic hormone, **ACTH**), which stimulate the adrenal cortex to release its hormones. The latter exerts feedback inhibition upon the hypothalamus to decrease the secretion of ACTH by the pituitary. Steroids also participate in feedback loops to the hypothalamus.[13] Using ^{3}H-labeled hormones or fluorescent analogs, it has been possible to locate specific brain cells sensitive to a given hormone by autoradiography.[14]

A characteristic of hormonal effects is that they are seldom unique, and are often balanced by counteracting effects of other hormones. For example, both glucagon and adrenaline promote the release of glucose from liver glycogen into the bloodstream. The glucocorticoids stimulate the rate of production of glucose from other body constituents (Chapter 11). Growth hormone tends to increase glucose levels by inhibiting utilization of sugar by tissues. On the other hand, insulin acts to promote uptake of glucose by tissues and a more efficient utilization. The thyroid

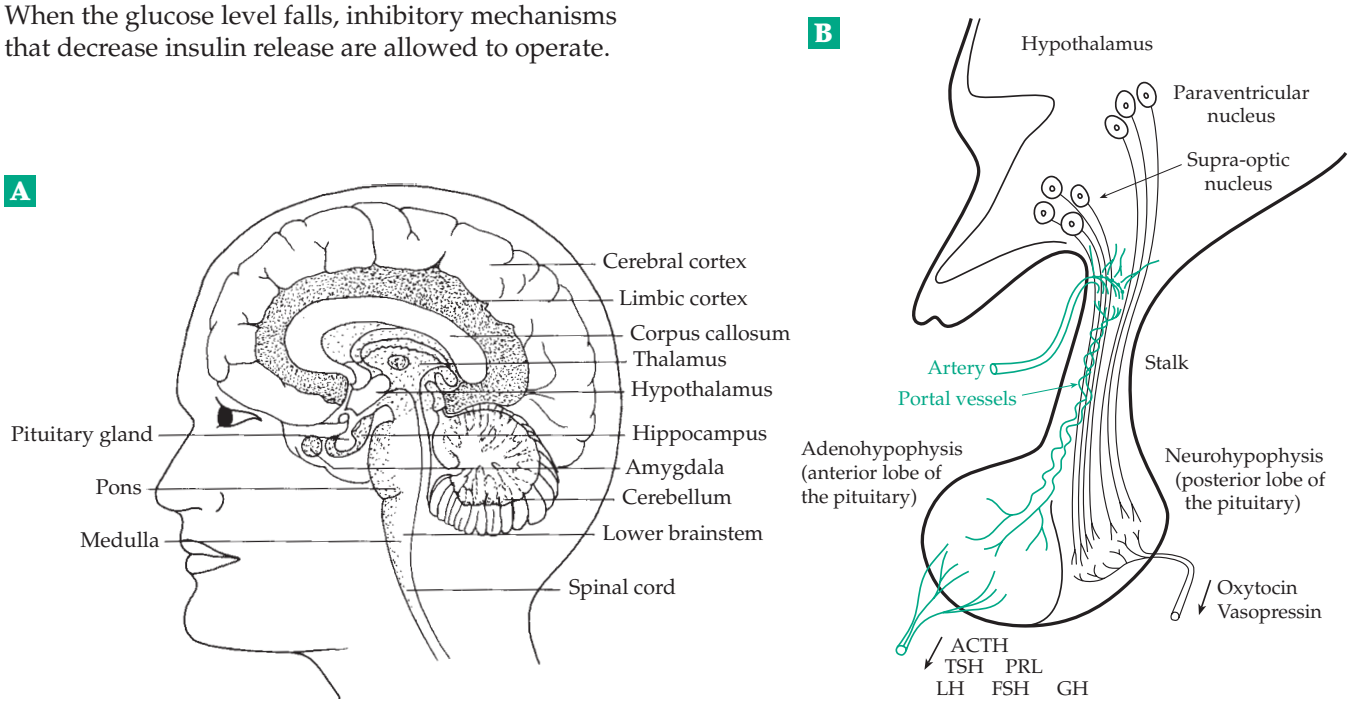


Figure 30-1 (A) Median sagittal section of the human brain. From Maya Pines.[15] (B) Drawing illustrating the synthesis of peptide hormones in the hypothalamus and transport via portal blood vessels into the anterior lobe of the pituitary gland or via nerve tracts into the posterior lobe.[16]

hormone increases the overall rate of metabolism of cells and also tends to promote a decrease in blood glucose.

Signaling cascades. As we have seen in Chapter 11, hormones frequently elicit the synthesis of second messengers such as cAMP, inositol phosphates, or diacyglycerol. This not only provides amplification of the initial hormonal signal but also allows a single hormone to control a "domain" of many metabolic processes. Each of these processes, in turn, can influence others. Many processes are affected by several different hormones and by more than one second messenger. Since we are far from knowing how many hormones exist and how many second messengers are released, the network of regulatory interactions within cells may be one of overwhelming complexity. An abbreviated version of a mitogen-activated kinase (MAP kinase) cascade[4,17] is shown in Fig. 11-13. These cascades are not only initiated but also are propagated by a series of phosphorylation reactions catalyzed by more than 1000 protein kinases encoded by the human genome.[18] Together with more than 500 protein phosphatases, which are often joined together as a bifunctional protein (p. 545),[19,20] they form a complex branching network of interactions.[21,22] These help to control responses not only to hormones but also to varying metabolite concentrations and physical stimuli.

Second messengers as hormones. The same compounds that serve as intracellular second messengers sometimes act as hormones. Tomkins suggested[23] that cAMP and some other small molecules serve as "symbols" indicating a metabolic need. For example, in bacteria ppGpp (p. 1715) serves as a symbol of nitrogen or amino acid deficiency. In cells ranging from those of bacteria to animals, cAMP is a symbol for carbon-source starvation. In *E. coli* cAMP levels increase during carbon-source starvation and stimulate the initiation of transcription of many bacterial operons (Chapter 28). In *Dictyostelium discoideum* (Box 11-C) cAMP is released by cells, when substrate depletion occurs. In this instance, the cyclic nucleotide acts as a hormone transmitting a signal to other cells.

Whereas cAMP is sometimes used by lower organisms as a hormone, its metabolic lability makes it unsuitable for higher animals. Thus, in our bodies the hormones glucagon and adrenaline carry a message to cell surfaces, where binding to receptors stimulates cAMP production. This, in turn, leads to mobilization of metabolic stores such as those of glycogen and triglycerides, just as if these cells had also been subjected to acute starvation. Tomkins suggested that hormones are produced by "sensor" cells in direct contact with environmental signals and travel to and activate more sequestered "responder" cells. The picture can be generalized further by realizing that neurotransmitters are largely derivatives of amino acids. These amino acids may have originally served as intracellular symbols reflecting changes in environmental amino acid concentration but were later utilized in short-range intercellular communication within the nervous system.

The vertebrate hormones. The principal established vertebrate hormones are listed in Tables 30-1 and 30-2. Also given are references to other parts of the text, where specific hormones are discussed. The hormones can be divided into four groups on the basis of chemical structure: (1) peptides and proteins, (2) derivatives of the aromatic amino acids, (3) steroids and prostaglandins, and (4) volatile compounds such as NO and CO. The most numerous are the peptide hormones, many of which also act as neurotransmitters. Peptide hormones, e.g., those with insulin-like effects, function in all phyla of the metazoa, and hormone-like molecules are found in bacteria.[24]

2. Hormones of the Pituitary Gland (Hypophysis) and Hypothalamus

Connected to the brain by a stalk (Fig. 30-1), the pituitary gland releases at least ten peptide or protein hormones that regulate the activity of other **endocrine** (hormone-producing) glands in distant parts of the body. The pituitary is composed of several distinct parts: the anterior lobe (**adenohypophysis**), a thin intermediate portion (**pars intermedia**), and a posterior lobe (**neurohypophysis**). Each has its own characteristic endocrine functions.

The anterior lobe of the pituitary secretes a series of ten or more peptide hormones ranging in size from the ~20-residue β-melanotropin to the ~200-residue growth hormone (somatotropin). Several of these contain a common heptapeptide unit, which is marked in green in the following structure:

Ac•Ser•Tyr•Ser•Met•Glu•His•Phe•Arg•Trp•Gly•Lys•Pro•ValNH_2

Structure of α-melanotropin from pig, beef, and horse

Not only this heptapeptide but also the entire amino acid sequence of **α-melanotropin** is found within the sequence of **corticotropin** (Fig. 30-2), which has an additional 29 amino acids at the C-terminal end.[25] The same heptapeptide was also found in the **lipotropins**. The explanation is, in part, that several of these hormones arise from a single 31-kDa precursor protein called **prepro-opiomelanocortin**.[25,26] It contains an N-terminal signal sequence that is removed shortly after synthesis, as well as pairs of adjacent basic residues (Arg-Arg, Arg-Lys, Lys-Arg and Lys-Lys) at a number of places (Fig. 30-2). After removal of the

signal sequence, further cleavage is thought to occur within the secretory vesicles by proteases, which cut either on the carboxyl side of these basic pairs or between them.[25,27,28] The same precursor is made in both anterior and intermediate lobes and is rapidly cut to ACTH, β-lipotropin, and an N-terminal part. In the intermediate lobe the ACTH is then cleaved at the Lys-Lys and Arg-Arg pairs to form αMSH and another peptide called corticotropinlike intermediate lobe peptide (CLIP). Beta lipotropin is degraded rapidly in the intermediate lobe and more slowly in the anterior lobe to γ-lipotropin and the opioid peptide β-endorphin

TABLE 30-1
Peptide and Protein Hormones of Vertebrates

Source and name of hormone	No. residues	Principal site of action	References Chapter, Section
A. Pituitary gland (hypophysis)			
1. Adenohypophysis (anterior portion)			
Corticotropin (ACTH)[a]	39	Adrenal cortex, adipose tissue	Fig. 2-4
β-Melanotropin (β melanocyte-stimulating hormone, β-MSH)[a]	18–22	Skin	Fig. 30-2
β-Lipotropin (β-LPH)[a]	91	Precursor of β-MSH and β-endorphin	Fig. 30-2
γ-Lipotropin (γ-LPH)[a]	58	Precursor of β-MSH	Fig. 30-2
β Endorphin[a]	31	Brain	
Somatotropin (growth hormone, GH)	~200	All tissues	
Prolactin (mammotropin)	~200		
Thyrotropin (thyroid-stimulating hormone, TSH)[b]		Thyroid	Ch 25, B2
Follitropin (follicle-stimulating hormone, FH)[b]		Ovaries, testes	
Lutropin (luteinizing hormone, ICSH or LH)[b]		Ovaries, testes	
2. Pars intermedia (intermediate portion)			
α-Melanotropin (α-melanocyte-stimulating hormone, αMSH)[a]	13	Skin	Fig. 30-2, pp. 1742, 1748
3. Neurohypophysis (posterior portion)			
Oxytocin (ocytocin)	9	Uterus, mammary glands	Fig. 2-4
Vasopressin (antidiuretic hormone)	9		Fig. 2-4
B. Pancreas			
Insulin	51	All cells	Fig. 7-17, Ch 11, G
Glucagon	29	Liver, adipose tissue	Ch 11, D
C. Ovary (corpus luteum)			
Relaxin	—	Pelvic ligaments	p. 1746
D. Thyroid			Ch 25, B2
Calcitonin (thyrocalcitonin)	32	Bones, kidney	Box 6-D
E. Parathyroid			
Parathyrin (parathyroid hormone)	84	Bones, kidney	Box 6-D
F. Kidney			
Erythropoietin		Bone marrow	
Renin (an enzyme)		Adrenal cortex	p. 621; Box 22-D

[a] Arise by cleavage of pro-opiomelanocortin.
[b] Related two-subunit (αβ) proteins with a common β subunit for these three hormones, for human chorionic gonadotropin (hCGH), and for mitogen-regulated protein (proliferin).

(Section B,10). Precursor proteins have been identified for many other peptide hormones, even those with very short chains.[29–30a] Proteolytic cleavages and other processing reactions occur within the secretory pathways of organisms from yeast to humans.[30b,c]

Many pituitary hormones have a pyroglutamate (5-oxoproline) residue at the N terminus (e.g., see Fig. 2-4). This presumably arises by attack of the terminal $-NH_2$ group on the amide carbon of an N-terminal glutamine side chain with displacement of NH_3 (Eq. 10-10).[29] The C terminus is often an amide of the carboxyl group with ammonia, which usually arises from a peptide chain containing one additional glycine residue at the C terminus (Eq. 10-11). The processing of peptide hormones doesn't end with their synthesis. They are usually degraded quickly or are converted into derivatives with weaker hormonal activity.

Pituitary growth hormone and related hormones. The pituitary growth hormone (**somatotropin**)[31] and **prolactin** are 22- to 23-kDa proteins, which share homology also with human **placental lactogen**, a lactogenic hormone secreted by the placenta,[32] and with a growth factor called **mitogen-regulated protein** (or proliferin).[33–35] The polypeptide chain of the 191-residue porcine[36] and human[37] growth hormone folds into an antiparallel four helix bundle (similar to that in Fig. 2-22) but with two long irregular connecting strands. The high degree of homology among somatotropins of many other species indicates a near identity of three-dimensional structures. However, biological function is species-specific. Humans and monkeys respond only to growth hormone from primates. The interaction of an aspartate side chain at position 171 with arginine 43 of the receptor protein may account for some of this specificity.[38] The receptor is a member of a large superfamily of receptor proteins with single transmembrane helices and extracellular domains similar to that of tissue factor (Fig. 12-18) and also, in some respects, to immunoglobulin domains.[39,40] Receptors for growth hormones bind in specific ways to two molecules of receptor protein.[41,42]

Human growth hormone produced in bacteria is used to help very short children to grow. Bovine growth hormone produced in bacteria is used to increase milk production from cows.[43] However, this use may be damaging to the cows.[44] Some humans produce too much growth hormone, often as a result of tumors in the pituitary. The resulting condition of **acromegaly**[45] causes excessive bone growth and many other problems. Growth hormone has a broad range of other effects, e.g., mimicking the action of insulin.[46]

Lactogenic hormones also have 4-helix bundle structures, and the prolactin receptor structure resembles that of growth hormone as well as those of a large cytokine family.[47,48] Prolactin affects placental development during pregnancy. However, during the latter half of pregnancy the placenta has a dominant endocrine effect, synthesizing both progesterone and the placental lactogen.[32,49]

The pituitary glycoprotein hormones. The thyroid-stimulating hormone **thyrotropin** (TSH), together with **folitropin** (FH) and **lutropin** (LH; Table 30-1), form a family of related ~28-kDa dimeric glycoproteins in which each subunit has a three-loop structure stabilized by a characteristic "cystine knot."[50] Also included in the family is the placental **chorionic gonadotropin**,[51,52] which is found only in human beings and a few other species. LH has a central role in promoting both spermatogenesis and ovulation by stimulating synthesis of steroid hormones in the testes and ovary, respectively.[53] Human chorionic gonadotropin (hCG) is also essential for maintenance of pregnancy and acts by stimulating the ovaries to secrete required steroid hormones.

All of these glycoprotein hormones are αβ dimers, and within a single species the subunits of TSH, FH, LH, and CG are identical. However, the β subunits are all different.[54,55] In the human there are at least six genes or pseudogenes for the hCG β chain in a cluster that also contains a single LH β chain gene.[54] The hormones undergo glycosylation and sulfation in the Golgi before secretion.[56] The hormones bind to 7-helix receptors, which are coupled to formation of cAMP or inositol trisphosphate.[57] Mutations in LH may cause male infertility,[58] while mutations in the corresponding receptor may cause male precocious puberty.[59]

Hypothalamic releasing hormones. As was mentioned previously, the anterior lobe of the pituitary

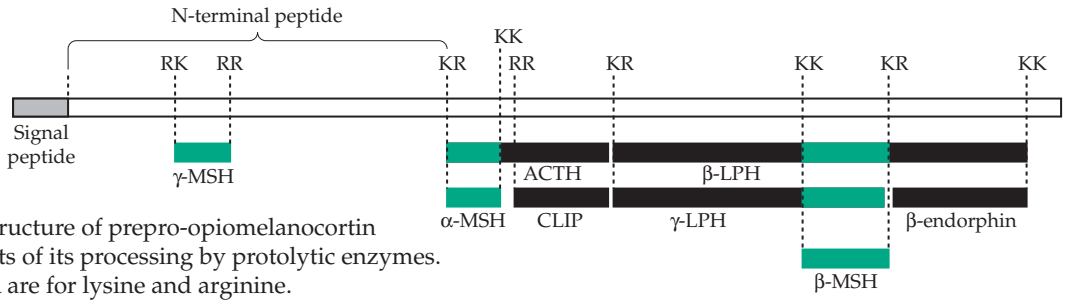


Figure 30-2 Schematic structure of prepro-opiomelanocortin and of some of the products of its processing by protolytic enzymes. The abbreviations K and R are for lysine and arginine.

releases its hormones in response to at least nine neurohormones known as **releasing hormones** or releasing factors.[16,60,61] They are secreted in minute quantities by the hypothalamus into a special portal vein that carries them directly to the pituitary where they exert their effects (Fig. 30-1B). As is indicated in Fig. 2-4 and in Table 30-3, several releasing factors are small peptides, but others are quite large. **Thyrotropin-releasing hormone** (THR, thyroliberin)[62] is a tripeptide; but human **growth hormone-releasing hormone** (somatoliberin) is a 44-residue peptide. Both are synthesized as larger proteins, which are cleaved and processed to form the mature C-amidated hormones.[63,64] **Corticotropin-releasing hormone** (CRH; CRF; corticoliberin), the 41-residue ACTH-releasing factor, is also cut from the much larger prepro-CRF.[65,66] Release of both LH and FSH is stimulated by a single **gonadotropin-releasing hormone** (GnRH).[5a,10,67,68] The releasing factors bind to 7-helix G-protein coupled receptor.[68] Both the releasing factor and gonadotropin are released into the appropriate parts of the bloodstream in a pulsatile fashion emphasizing the neural origin of their release.[69,70]

The hypothalamus also synthesizes **release-inhibiting factors**.[67] One of these, **somatostatin** (Table 30-3), inhibits release of somatotropin, thus counteracting the effect of the growth hormone releasing hormone. Somatostatin acts both in the pituitary and also in the pancreas, where it inhibits the release of both insulin and glucagon.[71,72] The result is a lowering of blood glucose. This suggested a new approach to the treatment of diabetes. However, because of the many other effects of somatostatin[73] and its rapid degradation it has not been useful clinically. Nevertheless, hundreds of analogs of somatostatin have been synthesized, some of which may be of practical

TABLE 30-2
Nonpeptide Vertebrate Hormones

Type, source, and name of hormone	Principal site of action	References Chapter, Section
A. Amino acid derivatives		
1. Thyroid		Ch 25,B,2
Thyroxine and triiodothyronine	Most cells	
2. Adrenal medulla		Ch 11
Adrenaline, noradrenaline (epinephrine, norepinephrine)	Most cells	
3. Pineal gland		
Melatonin	Melanophores	Fig. 25-12
4. Nerves and other cells		
Serotonin (5-hydroxytryptamine)	Arterioles, central nervous system	Fig. 25-12
B. Steroids and prostaglandins		Ch 22, F
1. Testes		
Testosterone	Most cells	Fig. 22-11
2. Ovaries		
Estrogen (estradiol-17β)	Most cells	Fig. 22-11
3. Corpus luteum		
Progesterone	Uterus, mammary glands	Ch 12; Ch 28
4. Adrenal cortex		Fig. 22-11
Corticosterone, cortisol	Most cells	Fig. 22-11
Aldosterone	Kidney	Fig. 22-11
5. Various tissues		
Prostaglandins	Smooth muscle	Ch 21, D
C. Volatile hormones		Fig. 21-8
1. Nitric oxide, NO	Endothelium, brain	
2. Carbon monoxide, CO	Brain	

value.[67] The biological activity of somatostatin resides largely in the sequence FYKT at positions 6–10, a sequence that is thought to form a beta turn (Fig. 30-3). Much of the rest of the molecule can be left off and the disulfide bridge moved up as far as positions 6–11 with retention of high potency. Human somatostatin is synthesized initially as a 116-residue precursor.[74]

A 56-residue peptide, which is formed from the 10-kDa precursor to GnRH, inhibits secretion of prolactin.[75] Inhibition of FSH release is accomplished by feedback inhibition. Hormones known as **inhibins** are produced in the gonads and act to inhibit release of FSH from the pituitary.[76]

Vasopressin and oxytocin. In contrast to the large peptide hormones made in the anterior lobe of the pituitary are **vasopressin** and **oxytocin**, which are secreted from the neurohypophysis, the posterior lobe.[60] The neurohypophysis consists of neural tissue, whose secretions are directly controlled by the central nervous system. In fact, the cell bodies of the secretory neurons are located in specific nuclei of the hypothalamus (Fig. 30-1B). About 4000 vasopressin-secretory neurons and a similar number of oxytocin neurons are present in the neurohypophysis of the rat.[16] Vasopressin is a major regulator of blood volume and pressure,[77] and its secretion is influenced by stress. It increases the water permeability of the kidney collecting duct cells by inducing translocation of aquaporin proteins from intracellular storage vesicles into the apical plasma membrane.[78] Vasopressin binds to

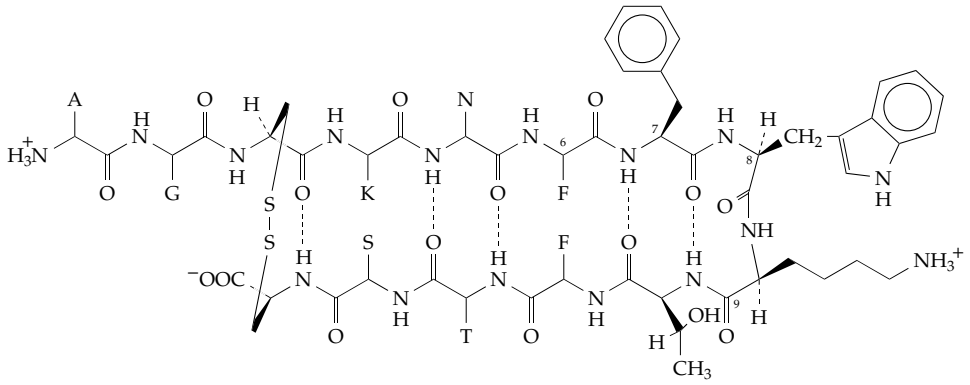


Figure 30-3 Possible secondary structure of somatostatin with a beta turn at residues 7–10 and a disulfide bond between positions 3 and 14. The true conformation would have the plane of the beta sheet puckered and twisted.

TABLE 30-3
Releasing and Inhibiting Hormones from the Hypothalamus

Name	Number of amino acid residues	Sequence[a]
Thyrotropin-releasing hormone (thyroliberin, TRH)	3	*p*EHP–NH_2
Gonadotropin-releasing hormone (GnRH, LH- and FSH-releasing hormone)	10	*p*EHWSYGLRPG–NH_2
GH-releasing hormone (somatoliberin)	44	
Corticotropin-releasing hormone (corticoliberin, CRH)	41	
MSH-releasing factor (melanoliberin)[b]	5	CYIQNC └S—S┘
Somatostatin (GH-release-inhibiting hormone)	14	AGCKNFFWKTFTSC └——S—S——┘
MSH-release-inhibiting factor[b]	3	PLG
Prolactin-releasing factor		
Dopamine (prolactin-release-inhibiting factor)		

[a] Standard one-letter abbreviations are used. *p*E is pyroglutamyl (5-oxoprolyl) and –NH_2 at the right indicates a C-terminal *carboxamide*.
[b] Ring and tail fragments of oxytocin.