CHAPTER 13

Viruses Use Cellular Processes to Express Their Genetic Information

Since viruses must use the cell for replication, it is necessary to understand what is going on in the cell and how a virus can utilize these processes. A virus must use cellular energy sources and protein synthetic machinery. Further, many viruses use all or part of the cell's machinery to extract information maintained in the viral genome and convert it to mRNA (the process of **transcription**). While cellular mechanisms for gene expression predominate, virus infection can lead to some important variations. Various RNA viruses face a number of special problems that differ for different viruses. Also, many viruses *modify* or *inhibit* cellular processes in specific ways so that expression of virus-encoded proteins is favored.

To understand how viruses parasitize cellular processes, these processes should be understood. Indeed, the study of virus gene expression has served as a basis for the study and understanding of processes in the cell. All gene expression requires a mechanism for the exact replication of genetic material and the information contained within, as well as a mechanism for "decoding" this genetic information into the proteins that function to carry out the cell's metabolic processes.

Whether prokaryotic or eukaryotic, the cell's genetic information is of two fundamental types: ***cis*-acting genetic elements** or signals, and ***trans*-acting genetic elements**. Genetic elements that act in *cis* work only in the context of the genome in which they are present. These include the following:

1 information for the synthesis of new genetic material using the parental genome as template; and
2 signals for expression of information contained in this material as RNA.

Trans-acting elements are just that information expressed to act, more or less freely, at numerous sites within the cell. Such information includes the genome sequences that are transcribed into mRNA and ultimately *translated* into proteins, as well as the sequences that are transcribed into

RNA with specific function in the translational process: ribosomal RNA (rRNA) and transfer RNA (tRNA). Certain other functional RNA molecules also can be included in this category.

While both prokaryotic and eukaryotic cells utilize DNA as their genetic material, a major difference between eukaryotic and prokaryotic cells is found in the way that the double-stranded (ds) DNA genome is organized and maintained in the cell. Bacterial chromosomes are circular, and whereas they have numerous proteins associated with them at specific sites, genomic DNA can be considered as "free" DNA (i.e., not associated with any chromosomal proteins).

In contrast, eukaryotic DNA is tightly wrapped in protein, mainly histones. Thus, the eukaryotic genome is the protein-nucleic acid complex chromatin. The unique structure of this chromatin and its condensed form of chromosomes, and the ability of these to equally distribute into daughter cells during cell division, are manifestations of the chemical and physical properties of the **deoxyribonucleoprotein** complex.

There are also differences in the way that genetic information is stored in bacterial and eukaryotic chromosomes. In bacterial chromosomes, genes are densely packed and only about 10% to 15% of the total genomic DNA is made up of sequences that do not encode proteins. Nonprotein-encoding sequences include mainly short segments that direct the transcription of specific mRNAs, short segments involved in initiating rounds of DNA replication, and the information-encoding tRNA and rRNA molecules.

In some eukaryotic genomes, on the other hand, 90% or more of the DNA does not encode any stable product at all! Some of this DNA has other functions (such as the DNA sequences at the center and ends of chromosomes) but some of these noncoding DNA sequences appear to be biological "junk" (material accumulated over evolutionary time that appears to have no current function).

Replication of cellular DNA

Despite differences in the nature of bacterial and eukaryotic genomes, the basic process of genomic DNA replication is quite similar. The two strands of cellular DNA are *complementary* in that the sequence of nucleotide bases in one determines the sequence of bases in the other. This follows from the Watson-Crick base-pairing rules. In the process of DNA replication, the following rules are useful to keep in mind:

1 A pairs with T (or U in RNA); G pairs with C.
2 The newly synthesized strand is antiparallel to its template.
3 New strands of DNA "grow" from the 5′ to 3′ direction.

The replication of a DNA molecule using the basic Watson-Crick rules, as well as features near the growing point, are shown in Fig. 13.1. The parental DNA duplex "unwinds" at the growing point (the **replication fork**) and two daughter DNA molecules are formed. Each new daughter contains one parental strand (light line in Fig. 13.1) and one new strand (heavy line). Each base in the new strand and its **polarity** are determined by the three rules just presented.

The replication of DNA can be divided into two phases, *initiation* of a round of DNA replication, which leads to generation of a complete daughter strand, and the *elongation* of this strand following initiation. Each round of DNA replication is initiated with a short piece of RNA that functions as a primer to begin the DNA chain. The first priming reaction takes place at a specific region (or site) on the DNA called the **origin of replication (ori)**. This origin of replication comprises a specific sequence of bases where an **origin-binding protein** interacts and causes local denaturation to allow the replication process to begin.

Following denaturation of the DNA duplex at the ori, there is synthesis of the short RNA primer homologous to a short region of DNA in the origin region. This reaction is mediated by a specific enzyme called **primase**. As the priming reaction proceeds, enzymes and proteins required for unwinding the DNA duplex and maintaining it as single-stranded (ss) material (**topoisomerases**, **helicases**, and **ssDNA-binding proteins**) interact with the denatured "bubble" to keep open the

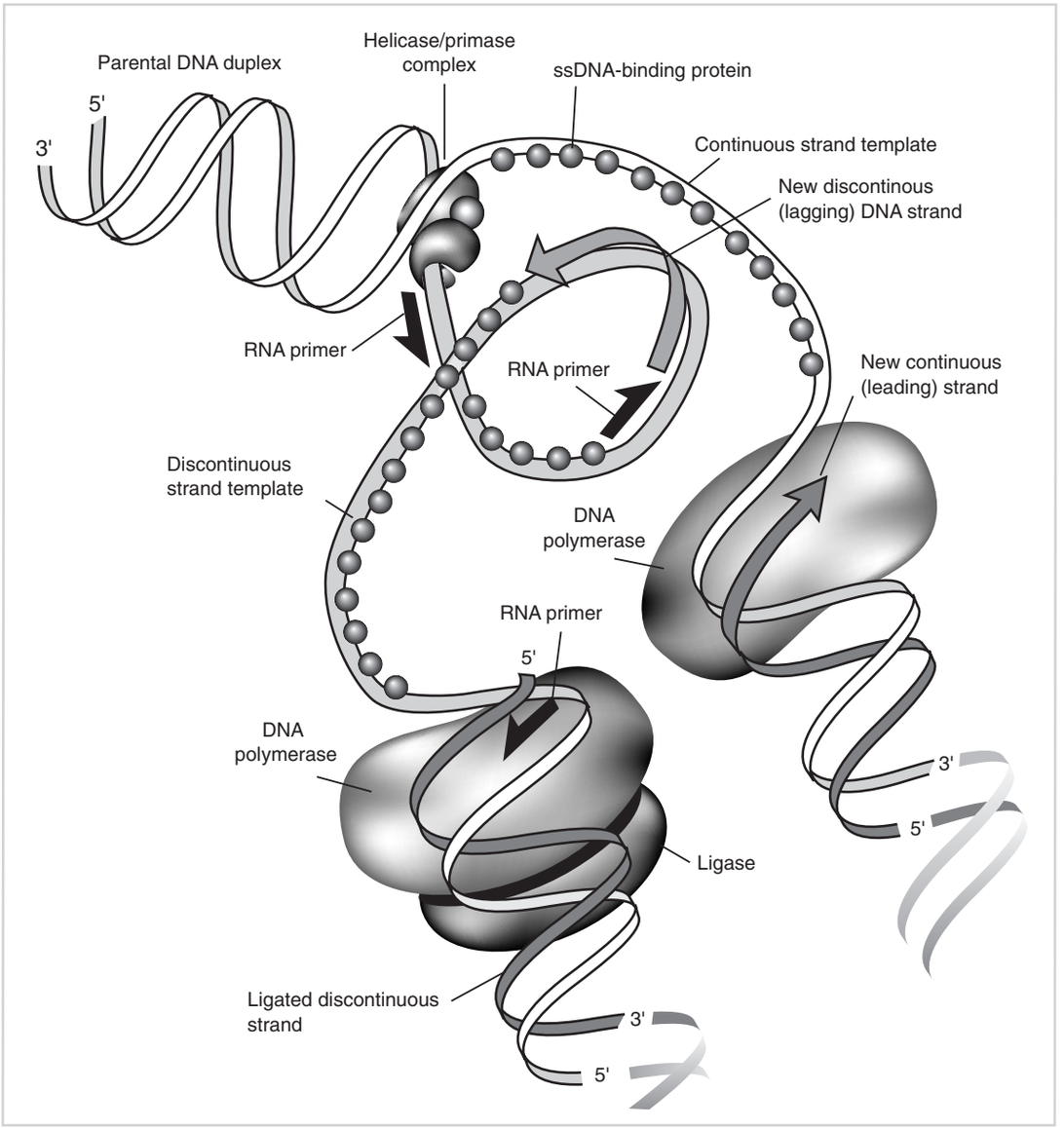


Fig. 13.1 The enzymes and other proteins associated with DNA around a growing replication fork. The process is described in the text. Each new DNA chain must initiate with an RNA primer that forms in the vicinity of the unwinding DNA duplex. The unwinding is mediated by enzymes termed *helicases* that are complexed with primases. One DNA strand grows continuously; this is the "leading strand." Replication on the other ("lagging strand") is discontinuous due to the requirement for DNA synthesis to proceed from 5′ to 3′ antiparallel to the template strand. These discontinuous fragments are also called **Okazaki fragments** after the man who first characterized them. The primers are then removed, the gaps filled in, and the DNA fragments are ligated.

growing fork. **DNA polymerase**, which polymerizes the growing DNA chain by associating with the denatured bubble in the DNA duplex, begins polymerization of new DNA at the 3′ OH of the RNA primer. This new DNA is antiparallel and complementary to the template strand.

At the DNA growing point, DNA synthesis is continuous in one direction, but discontinuous in the other direction. Discontinuous or **lagging strand** synthesis occurs because in order to maintain proper polarity, new primer must be placed upstream of the growing point. In other words,

priming must "jump" ahead on the template to continue synthesis of the new DNA strand, because DNA polymerase can only generate newly synthesized product in the 5′ to 3′ direction (reading the template strand 3′ to 5′).

As lagging strand synthesis proceeds, primer RNA must be removed, gaps repaired, and discontinuous fragments ligated together to make the full strand. These final steps require the action of an **exonuclease** for removing RNA primer and **DNA ligase** for linking together the fragments of growing DNA on the lagging strands.

Replication of viral DNA follows the same basic rules as for cellular DNA. Actually, the process for a number of viruses mimics the cellular process exactly. Other DNA-containing viruses are distinguished by special variations on the general theme, and are covered in some detail where particular viruses are discussed.

Depending on the complexity of the virus in question, a few or all of the required enzymes and proteins are supplied by the virus. In the case of herpesviruses, such as HSV, the process is virtually identical to the cellular patterns outlined here, but involves mainly virus-coded proteins. These proteins have clear genetic relationships with cellular enzymes that have the same function. The process for HSV is shown schematically in Fig. 13.2.

Expression of mRNA

The expression of mRNA from DNA involves transcription of one strand of DNA (the mRNA coding strand that is the complementary sense of mRNA). Following initiation of transcription, RNA is polymerized with a DNA-dependent RNA polymerase using Watson-Crick base-pairing rules (except that in RNA, U is found in place of T). Although similar in broad outline, many details of the process differ between prokaryotes and eukaryotes. One major difference is that the bacterial enzyme can associate directly with bacterial DNA and the enzyme itself can form a **pre-initiation complex** and initiate transcription. In eukaryotes, a large number of auxiliary proteins assembling near the transcription start site are required for initiation of transcription, and RNA polymerase can only associate with the template after these proteins associate. The process of transcription termination also differs significantly between the two types of organism.

Eukaryotic transcription

The promoter and initiation of transcription

All nonorganelle transcription occurs in the nucleus. RNA polymerase II (pol II) is "recruited" into the pre-initiation complex formed by association of accessory transcription-associated factors assembling at the site where the transcript is to begin; the process is outlined in Fig. 13.3. Transcription initiates in a "typical" eukaryotic promoter at a sequence of 6 to 10 bases made up on A and T residues (the **TATA box**), which occurs about 25 bases upstream (5′) of where the mRNA starts (**cap site**). The proteins making up this pre-initiation complex make a complex just large enough to reach from this region to the cap site as shown in Fig. 13.3. Formation of the pre-initiation complex around the TATA box can be modulated and facilitated by association of one of a number of **transcription factors** that bind to specific sequences usually upstream of—but within close proximity to—the cap.

These upstream transcription elements can interact with and stabilize the pre-initiation complex because dsDNA is flexible and can "bend" to allow transcription factors to come near to the TATA-binding protein complex; this bending is diagrammed in Fig. 13.4. The whole promoter region (containing the cap site, TATA box, and proximal transcription factor-binding sites) generally occupies the 60 to 120 base pairs immediately upstream (5′) of the transcription start site.

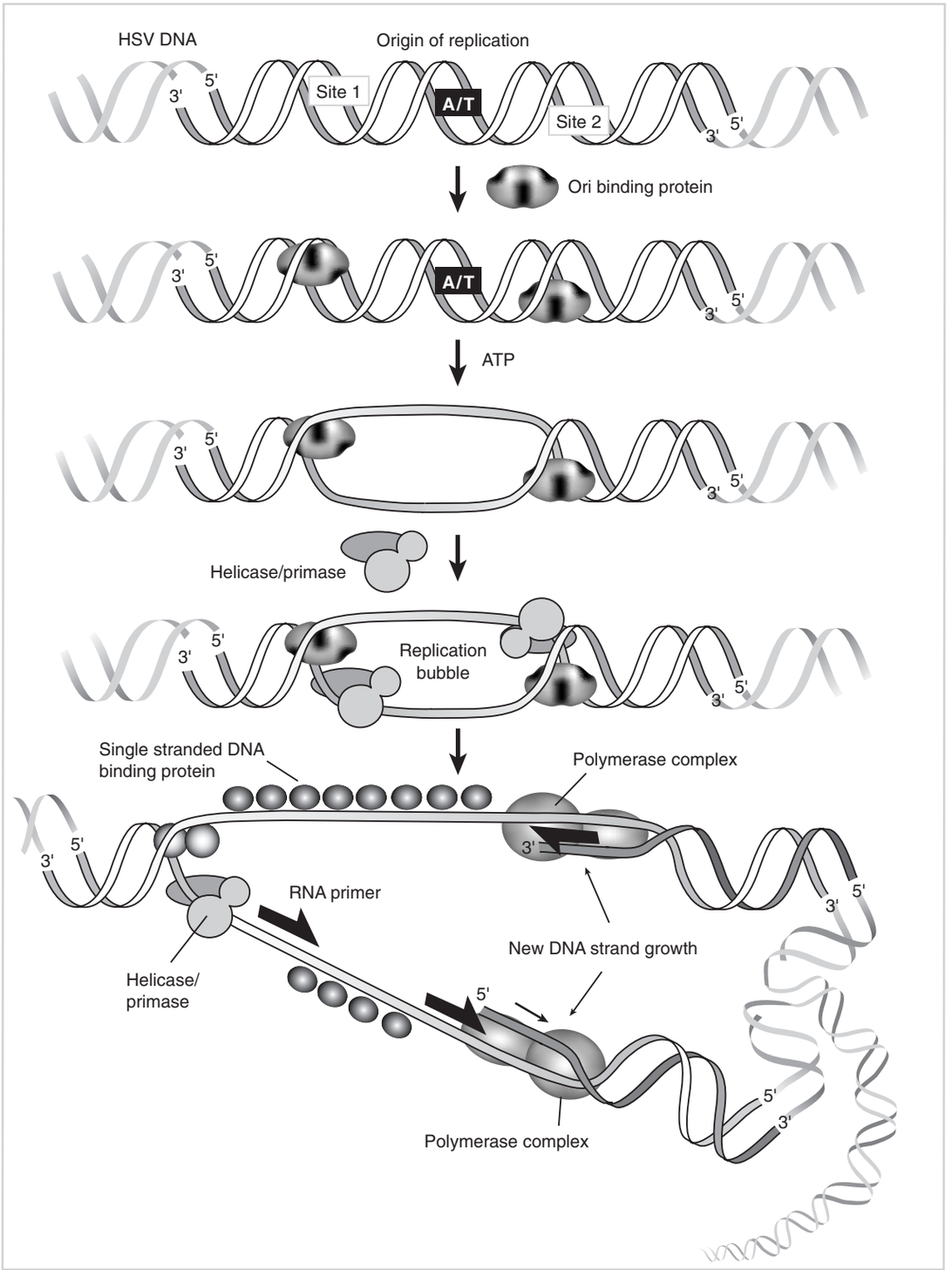


Fig. 13.2 Initiation of HSV DNA replication. This process is virtually identical to that occurring in the cell except that virus-encoded enzymes and proteins are involved. The initial step is denaturation of the DNA at the replication origin with origin binding protein. Following this, the helicase-primase complex and ssDNA-binding proteins associate to allow DNA polymerase to begin DNA synthesis. (Ori, origin of replication; A/T, AT-rich sequence.)

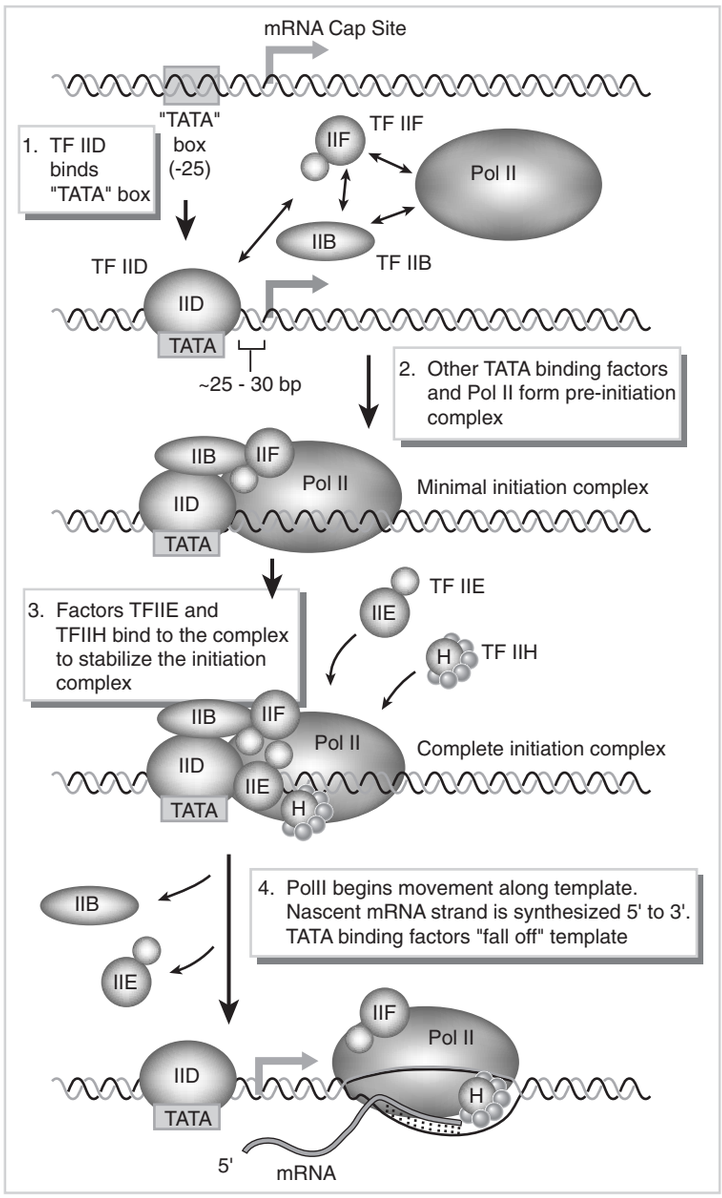


Fig. 13.3 The multistep process of transcription initiation at a eukaryotic promoter. With most promoters, the process begins as shown at the top with assembly of the initiation complex at the TATA box. Upon its full assembly, the DNA template is denatured, and RNA synthesis antiparallel to the template (nonsense strand) is initiated. The relative sizes of the proteins involved show how location of the pre-initiation complex at the TATA box is spaced to allow RNA polymerase to begin RNA synthesis about 25 to 30 bases downstream of it. (TF, transcription factor; Pol II, RNA polymerase II.)

Other control regions or **enhancers** can occur significant distances away from the promoter region. Such enhancers also interact with specific proteins and ultimately act to allow transcription factors to associate with the DNA relatively near the promoter; the process is shown in Fig. 13.4. These enhancers appear to help displace histones from the transcription template and therefore facilitate the rate of transcription initiation from a given promoter.

Unlike the core promoter element itself, however, enhancers serve only to regulate and augment transcription, and the promoter that they act on can mediate measurable transcription in their absence. Enhancers themselves may be subject to modulation of activity by factors stimulating the cell to metabolic activity, such as cytokines and steroid hormones.

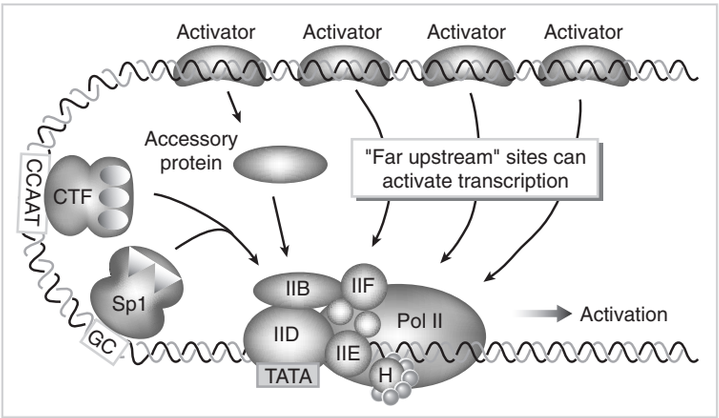


Fig. 13.4 The flexibility of DNA allows transcription factors binding at sites upstream of the TATA box to stabilize formation of the pre-initiation complex. Enhancer elements even further upstream (or in some cases, downstream) can also bind activating proteins that can further facilitate and modulate the process.

Posttranscriptional modification of precursor mRNA

Following initiation of transcription, transcript elongation proceeds. RNA is also modified following initiation by addition of a **cap** at the 5′ end. Capping takes place by the addition of a 7-methyl guanine nucleotide in a 5′-5′ phosphodiester bond to the first base of the transcript. This cap has an important role in initiation of protein synthesis.

Transcription proceeds until the pol II-nascent transcript complex encounters a region of DNA containing sequences providing transcription-termination/polyadenylation signals that occur over 25 to 100 base pairs. A major feature of this region is the presence of one or more **polyadenylation signals**, AATAAA in the mRNA sense strand.

Other short *cis*-acting signals also are present in the polyadenylation region. A specific enzyme (*terminal transferase*) adds a large number of adenine nucleotides at the 3′ end of the RNA just downstream (3′) of the polyadenylation signal as it is cleaved and released from the DNA template. Interestingly, the polymerase itself can continue down the template for a short or a long distance before it finally disassociates and falls off.

In addition to capping and polyadenylation, most eukaryotic mRNAs are spliced. In **splicing**, internal sequences (**introns**) are removed and the remaining portions of the mRNA (**exons**) are religated. Splicing takes place via the action of small nuclear RNA (**snRNA**) in complexes of RNA and protein (ribonucleoprotein) called spliceosomes. The process is complex, but the result is that most mature eukaryotic mRNAs are somewhat or very much smaller than the pre-mRNA precursor or primary transcript.

The generation of mature mRNA in the nucleus is shown diagrammatically in Fig. 13.5. Although splicing is shown to occur after cleavage/polyadenylation in this schematic, the actual process may occur as the nascent RNA chain grows. The maturation of RNA and splicing are shown in a somewhat higher-resolution view in Fig. 13.6.

All modifications occur on the RNA itself: first capping, then cleavage/polyadenylation of the growing RNA chain, then splicing (if any). Thus, almost all eukaryotic mRNAs are capped, polyadenylated, and spliced. Because splicing can occur within or between sequences of mRNA encoding peptides, it can result in the generation of complex "families" of mRNA encoding related or totally unrelated proteins. Some general patterns of splicing known to be important in virus replication are shown in Fig. 13.7a.

Following or during *posttranscriptional* modification, mRNA molecules are transported from the nucleus for release into the cytoplasm where protein synthesis takes place. This mRNA transport can also be regulated during virus infections so that certain spliced mRNAs are either inhibited or facilitated in their passage to the cytoplasm for translation.

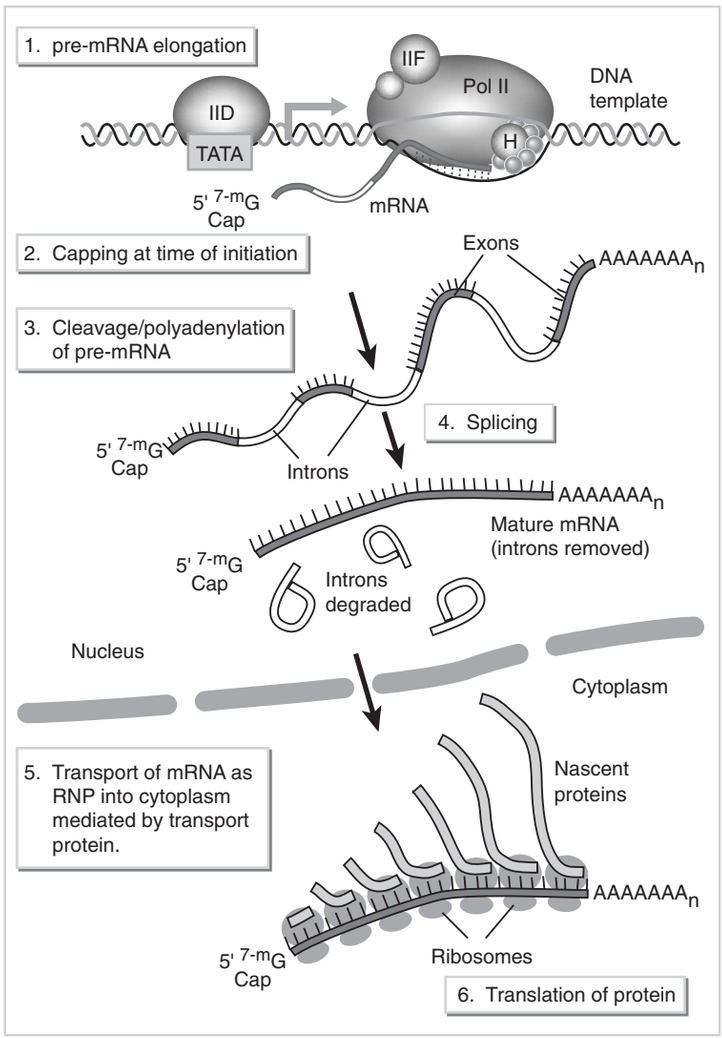


Fig. 13.5 Steps involved in transcription and posttranscriptional modification and maturation of eukaryotic mRNA. The sequence of events is indicated by the numbers 1 through 6. (RNP, ribonucleoprotein; $^{7\text{-m}}$G, 7-methyl guanine.)

Visualization and location of splices in eukaryotic transcripts

Provided a good physical map and cloned copies of the eukaryotic gene encoding a spliced transcript are available, there are a number of techniques for detecting and locating the splice sites in a given transcript; three are shown in Fig. 13.7b. All are based on the fact that when the DNA gene is hybridized to the mature transcript, the introns present in the gene will not be able to hybridize and therefore, must form a single-stranded loop in an otherwise contiguous hybrid.

The unhybridized ssDNA loop can be visualized in the electron microscope using a technique called **R-loop mapping**. R-loop mapping was originally developed as a method of visualizing a DNA-RNA hybrid in a dsDNA molecule by allowing hybridization under conditions where the RNA will displace its cognate DNA strand and anneal to its complementary strand. The displaced DNA will form a loop around the hybrid. Shadowing of the hybrid molecule will form heavy shadows where the nucleic acid is double stranded, and finer shadows where it is single stranded. When such a structure is spread and shadowed for visualization in the electron microscope, the RNA-DNA hybrid can be seen as a region of heavy shadowing connecting a loop of lightly shadowed ssDNA.

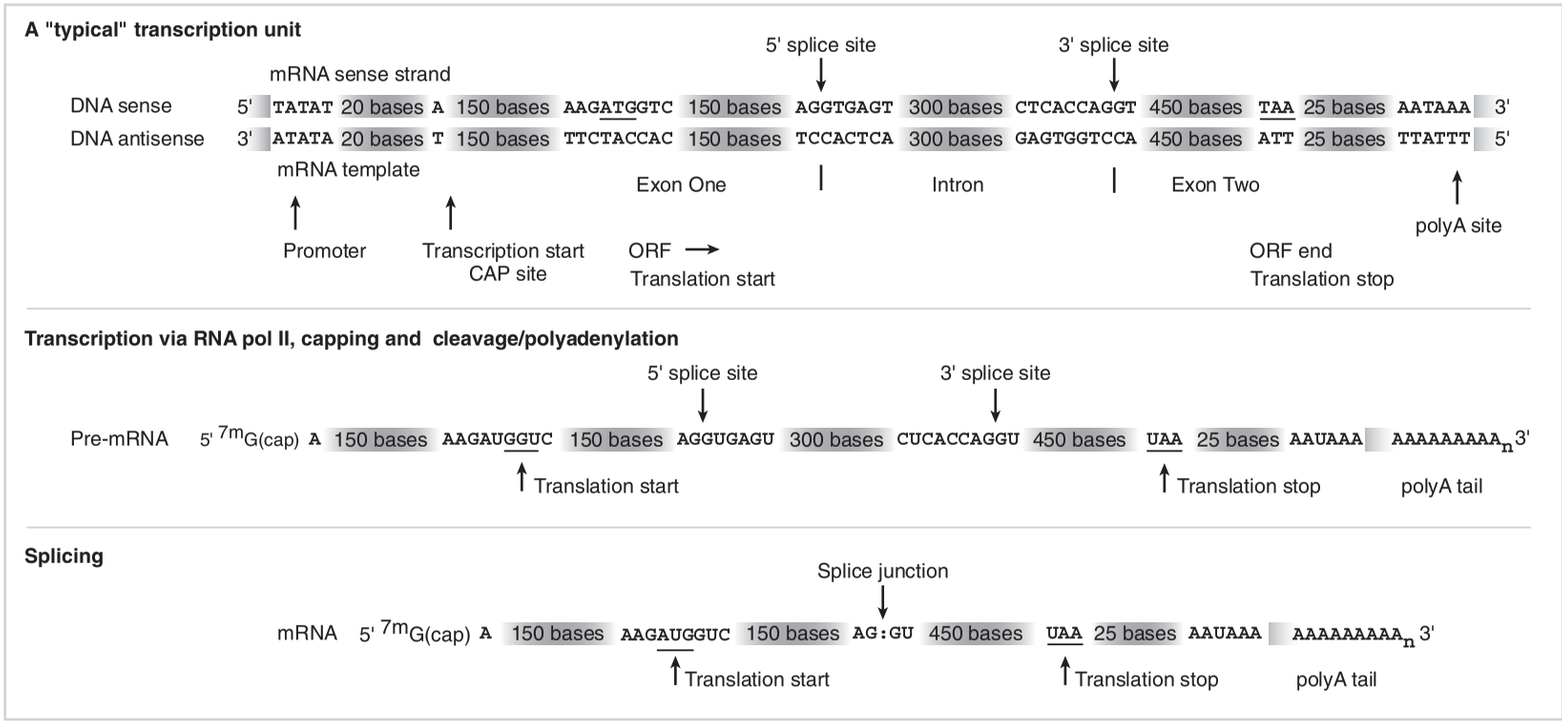


Fig. 13.6 A "high-resolution" example of mRNA processing. The sequence of a hypothetical pre-mRNA transcript is shown. The transcript is capped and polyadenylated, and splicing removes a specific sequence of bases (the intron). This results in the formation of a translational reading frame as shown.

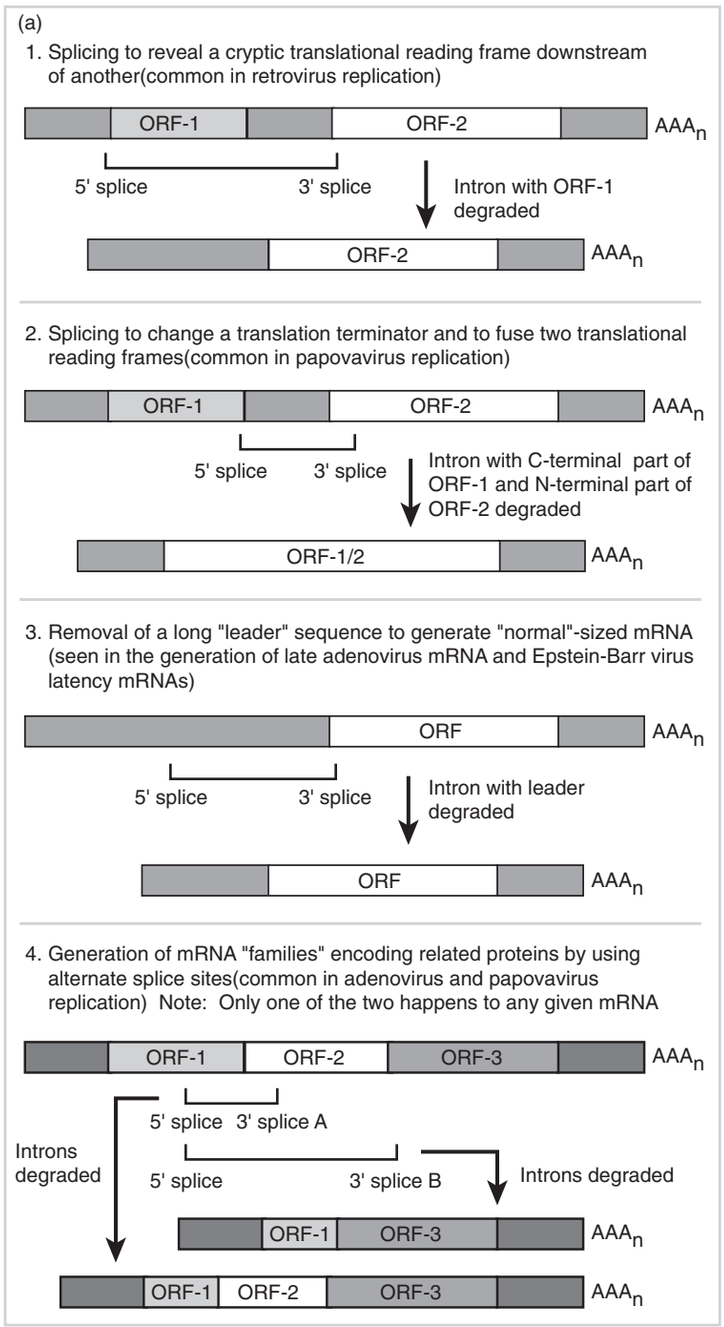


Fig. 13.7 Some splicing patterns seen in the generation of eukaryotic viral mRNAs. *a.* Schematic representation of different splice patterns that have been characterized. *b.* Molecular characterization of spliced transcripts. The formation of a hybrid between a fragment of DNA encoding a transcript and the final, processed mRNA will result in any introns present in the DNA looping out of the hybrid. These can be visualized by electron microscopy, by differential nuclease digestion and gel electrophoresis, or by sequence analysis of the cDNA generated from the transcript and comparison with the DNA sequence of the gene encoding it, as shown in simplified form in Fig. 13.6. *c.* Schematic representation of an electron micrograph of ssDNA introns (arrows) formed by hybridization of adenovirus DNA and late mRNA that has a complexly spliced leader (see Chapter 17). (Based on data in Berget, S. M., Moore, C., and Sharp, P. A. Spliced segments at the 5′-terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America* 1977;74:3171–5.) *d.* Generation of a polymerase chain reaction (PCR) product from HSV latency-associated RNA (LAT) by using primers annealing to regions 5′ and 3′ of an intron. The gene encoding the HSV latency-associated transcript is about 9 kbp long, and there is a 2-kb intron that is located about 600 base pairs 3′ of the transcription start site (see Chapter 18 and Fig. 18.2). PCR amplification of HSV DNA using the first primer set [P1:P(−1)] shown produces a product about 150 nucleotides (nt) long. Amplification using the second primer pair [P2 and P(−2)] will produce a fragment longer than 2000 nucleotides and cannot be seen. Next, LAT RNA from latently infected cells is used as a template for the synthesis of cDNA complementary to it (see Chapter 20). When the first primer pair is used for PCR amplification of LAT cDNA, a product the same size as that formed using genomic DNA as a template is formed. In contrast, however, when primer set 2 is used, the product of the cDNA is only about 160 base pairs long since the 2000-base intron has been spliced out. If the product of PCR primer set 2 were subjected to sequence analysis and compared to the sequence of viral DNA, a discontinuity at the splice sites would be revealed.

For visualization and mapping of splices, ssDNA is hybridized with RNA, and the DNA-RNA duplexes will form heavy shadowed images. Any unhybridized DNA in the interior of a gene will form a single-stranded loop that will shadow lightly, as is shown in Fig. 13.7c. Knowledge of the size of the DNA and the RNA being hybridized allows calculation of where the transcript starts and ends on the DNA strand used for hybridization, and the dimensions of the looped regions provide a measure of the intron's size.

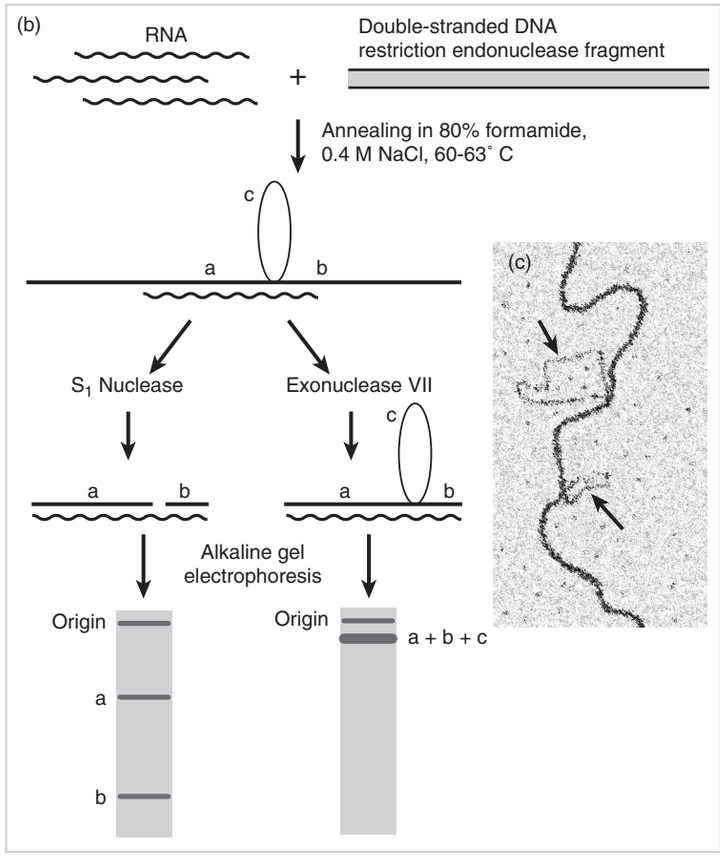


Fig. 13.7 *Continued*

A second method involves the hybridization of radiolabeled DNA with the transcript under study. After the hybrids are formed, the material is divided into two aliquots. One is digested with the **endonuclease** S_1-nuclease. This enzyme is able to cleave randomly within any ssDNA molecule, and will digest all unhybridized DNA whether it is at the end of the probe or present in an unhybridized intron loop.

The second aliquot is digested with an exonuclease, exonuclease VII, which digests ssDNA but can only begin digestion at a free end. Digestion of the hybridized material will result in the ssDNA fragments at the ends of the hybridized duplex being digested, but will leave the intron loops intact.

The two samples are then denatured with alkali, which hydrolyzes the RNA, and the alkali-resistant labeled DNA is fractionated on a denaturing electrophoresis gel. The number of products of endonuclease digestion will be one more than the number of introns in the transcript, and the total size of the fragments will be equal to the total amount of the gene expressed as exons. In contrast, only a single fragment will result from exonuclease digestion, and its size will be equal to the sum of the sizes of the exons and introns.

A complexly spliced transcript cannot be fully analyzed in a single experiment. However, a series of analyses of the products of S_1-nuclease and exonuclease VII digestion of hybrids formed with different portions of the gene encoding a transcript generated by restriction endonuclease digestion will yield a complete picture.

The third, and most detailed and sensitive approach toward characterizing a spliced transcript and its relationship to the DNA encoding it is to carry out comparative sequence analysis of the gene and the cDNA generated from the transcript. This cDNA can be detected by PCR amplifica-

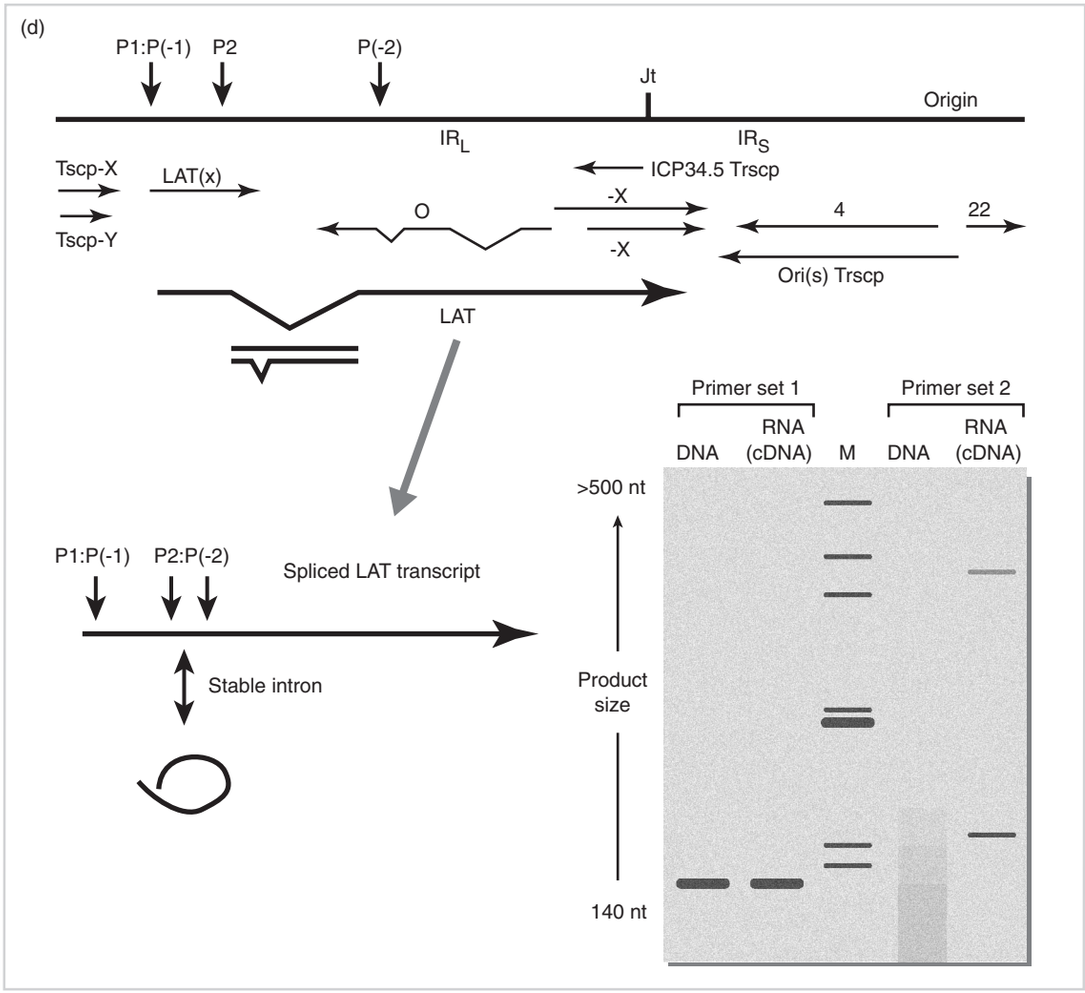


Fig. 13.7 *Continued*

tion of even extremely rare transcripts; therefore, a detailed picture of the splicing patterns of very-low-abundance mRNAs is technically quite feasible. An example of the generation of a PCR-amplified cDNA from a low-abundance latency-associated transcript of HSV is shown in Fig. 13.7d. Latent-phase transcription by herpesviruses is discussed in more detail in Chapter 18.

Prokaryotic transcription

Regions of prokaryotic DNA that are to be expressed as mRNA are very often organized such that a single transcription event results in the production of a single message from which two or more proteins can be translated. The ability of bacterial RNA polymerase to transcribe such mRNA is often controlled by the presence or absence of a DNA-binding protein, called a **repressor**. The DNA sequence to which the repressor can bind is called the **operator** and the genes expressed as a single regulated transcript are called **operons**. This is schematically shown in Fig. 13.8. The operon model for bacterial transcription was first proposed in the early 1960s by Jacob, Monod, and Wollman from their genetic analyses of mutants of *E. coli* unable to grow on disaccharide lactose.

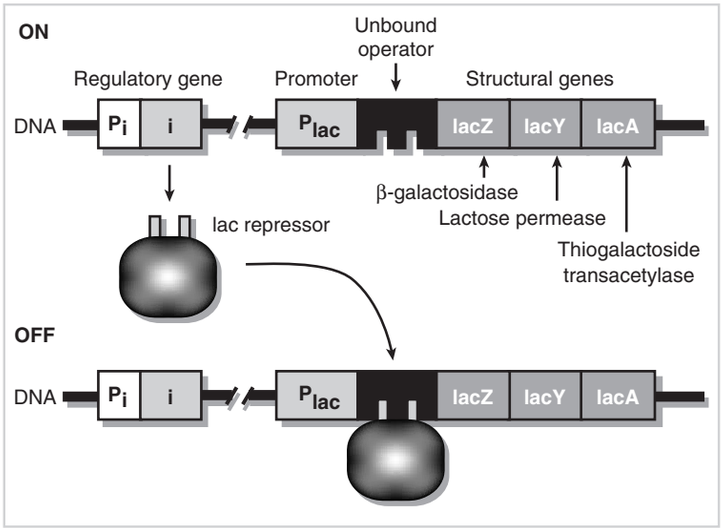


Fig. 13.8 The *E. coli* lac operon. The promoter is always "on," but normally the lac repressor (i) is bound to the operator that blocks transcription. The repressor can be inactivated by addition of lactose. The operator is also sensitive to cAMP levels as explained in the text. All the genes controlled by this operon are expressed as a single mRNA that can be translated into three separate proteins due to internal ribosome initiation.

Since then this operon model has been shown to be valid for a large number of prokaryotic transcriptional units.

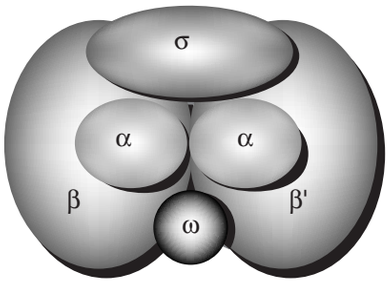
In addition to organization into operons, prokaryotic gene expression differs from that of eukaryotes as a result of a fundamental structural difference between the cells: the lack of a defined nucleus in prokaryotes. In prokaryotic cells, transcription takes place in the same location and at the same time as translation. This coupling of the two events suggests that the most efficient regulation of gene expression in these cells will be at the level of initiation of transcription. The operon model also takes this into account.

Prokaryotic RNA polymerase

The DNA-dependent RNA polymerase of prokaryotic cells is well studied, especially in *E. coli*. The enzyme shown in Fig. 13.9 contains five subunit polypeptides: two copies of α, one of β, one of β', one of σ, and one of ω. The functions of all the subunits except ω are known quite precisely. The core enzyme, which can carry out nonspecific transcription in vitro, consists of the β' subunit for DNA binding and the two α subunits and the β subunit for initiation of transcription and for interaction with regulatory proteins. The addition of the σ subunit creates the *holoenzyme* that transcribes DNA with great specificity, since this subunit is responsible for correct promoter recognition. It is the holoenzyme that is active in vivo for initiation of transcription.

The prokaryotic promoter and initiation of transcription

The DNA to which the RNA polymerase holoenzyme binds to begin transcription looks very much like its eukaryotic counterpart. Consensus sequences are present at specific locations upstream from the start site of transcription. A sequence with the consensus TATAAT is found at −10 and a sequence TTGACA at the −35 position. The former sequence is often called the **Pribnow box** after its discoverer and is similar in function to the TATA box of eukaryotes. The RNA polymerase holoenzyme binds to the promoter, causing a transcription bubble to form in the DNA. Just as in eukaryotes, transcription begins with a purine triphosphate and chain elongation proceeds in the 5′ to 3′ direction, reading the DNA template from the antisense strand in the 3′ to 5′ direction. The polymerase catalyzes incorporation of about 10 nucleotides into the growing mRNA before



Subunit	Size	#/molecule	Function
α	36.5 kd	2	Chain initiation and interaction with regulatory proteins
β	151 kd	1	Chain initiation and elongation
β'	155 kd	1	DNA binding
σ	70 kd	1	Promoter recognition
ω	11 kd	1	Unknown

Fig. 13.9 The bacterial RNA polymerase molecule. The enzyme is made up of six subunits with different functions. The complete enzyme is called the *holoenzyme*.

the σ subunit dissociates from the complex. Thus, σ is required only for correct initiation and transcription of the RNA chain's first portion.

Control of prokaryotic initiation of transcription

As mentioned earlier, the bacterial RNA polymerase holoenzyme will form a transcription complex and begin to copy DNA, given the presence of a correct promoter sequence. Since the strategy of prokaryotic regulation dictates that gene expression be regulated at the level of this initiation, most **inducible genes** (genes whose expression goes up or down with given cellular conditions) have the general structure of the operon diagrammed in Fig. 13.8. Binding of the repressor protein to the operator sequence of DNA, positioned at or immediately downstream of the initiation site, effectively provides a physical block to progress of the RNA polymerase. The repressor-operator combination acts, in effect, like an "on-off" switch for gene expression, although it should be understood that this binding is not irreversible and that there is some finite chance of transcription taking place even in the "off" state.

Presence of the appropriate inducing molecule, such as the metabolite of lactose responsible for inducing the *lac* operon, will cause a structural change in the repressor such that it can no longer bind to the operator. In cases such as the tryptophan operon, the repressor protein assumes the correct binding conformation only in the presence of the co-repressor (e.g., tryptophan). The overall situation is that regulated prokaryotic gene expression takes place unless it is prevented by the binding of a protein that blocks movement of RNA polymerase.

Enhancement of prokaryotic transcription is also seen. Using the example of operons for the genes required to utilize unusual sugars such as lactose, upregulation of gene expression can be observed. In this case, the response involves a system that can "sense" the amount of glucose presented to the cell and thus the overall nutritional state of that cell. Since the enzymes that metabolize glucose (the glycolytic pathway) are expressed constitutively (unregulated) in most cells, the availability of this sugar is a good signal for the cell to use in regulating the expression of enzymes for the metabolism of other sugars. The level of glucose available to the cell is inversely proportional to

the amount of 3′,5′-cyclic adenosine monophosphate (**cAMP**) within the cell. This nucleotide can interact with a protein called the cyclic AMP receptor protein (CRP). A complex of cAMP-CRP binds to a region of DNA just upstream of the promoter but only in genes that are sensitive to this effect. When the complex binds, the DNA is changed in such a way that the rate of transcription is raised manyfold. If the repressor protein is the "on-off" switch of this gene, then the cAMP-CRP complex is the "volume control" fine-tuning transcription as metabolic need arises. This regulation of the rate of transcription by the level of glucose is called **catabolite repression**.

Termination of transcription

Bacterial RNA polymerase terminates transcription by one of two means: in a ρ-dependent or ρ-independent fashion. The difference between these two involves the response of the system to the **termination factor (ρ factor)** and structural features near the 3′ terminus of the RNA.

In the case of ρ-dependent termination, the mRNA being transcribed contains, near the intended 3′ end, a sequence to which the ρ factor binds. The protein ρ is functional as a hexamer and acts as an ATP-dependent helicase to unwind the product RNA from its template and terminate polymerization. The process is shown in Fig. 13.10.

For ρ-independent termination, the sequence near the intended 3′ terminus of the transcript contains two types of sequence motifs. First, the RNA transcript contains a GC-rich region that can form a base-paired stem loop structure. Immediately downstream from this feature is a U-rich region. The presence of the GC-rich sequence slows progress of the polymerase. The stem loop that forms interacts with the polymerase subunits to further halt their progress. Finally, the AU-rich sequences melt and allow the transcript and template to come apart, terminating transcription. This is shown in Fig. 13.11.

Virus-induced changes in transcription and posttranscriptional processing

Many RNA-containing viruses completely shut down host transcription. Specific details are described in Chapters 15 and 16. The ability of DNA viruses to transcribe predominantly viral transcripts is usually a multistep process with the earliest transcripts encoding genes that serve regulatory functions, causing expression of viral genes to be favored.

Again, mechanisms vary with different viruses. With nuclear-replicating DNA viruses, the process often involves these earliest transcripts being expressed from a viral promoter that has a powerful enhancer, allowing active transcription without extensive modification of the cell. This is followed by changes in the structure of cellular chromatin and increases in viral genomes so that viral transcription templates begin to predominate relatively rapidly.

A major factor in usurpation of the cell's transcriptional capacity by these viruses is the fact that in general, the uninfected cell has much more transcriptional capacity than it is using at the time of infection. Consequently, increases in the availability of viral templates, along with alterations of the host chromatin structure, result in virus-specific transcription predominating.

Some nuclear-replicating viruses also encode regulatory proteins that affect posttranscriptional splicing and transport of transcripts from the nucleus to the cytoplasm. Such alterations in splicing do not affect the basic mechanism of splicing, but can specifically inhibit the generation and transport of spliced mRNA at specific times following infection. This inhibition involves the ability of viral proteins to recognize and modify the activity of spliceosomes. The alteration of splicing and transport of mRNA has especially important roles in aspects of the control of herpesvirus gene expression and in the regulation of viral genome production in lentivirus (retrovirus) infections. Specifics are described in Chapters 18 and 20.

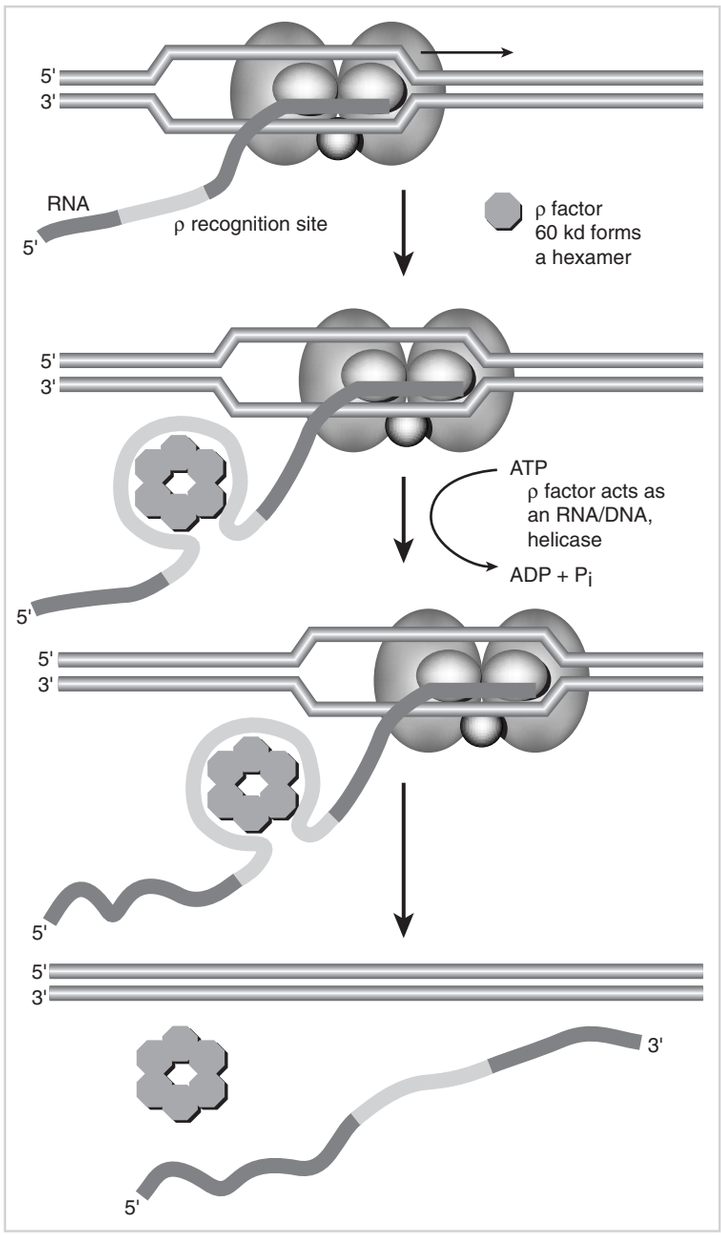


Fig. 13.10 Termination of bacterial transcripts dependent on action of the ρ factor. This factor interacts with a specific recognition site in the nascent transcript.

While no nuclear-replicating DNA viruses of vertebrates yet characterized encode virus-specific RNA polymerase, at least one, baculovirus, which replicates in insects, does. Further, this is a very common feature in DNA-containing bacteriophages. Indeed, as outlined in Chapter 19, changes in the infected bacteria's polymerase population is the major mechanism for ensuring virus-specific RNA synthesis and the change in types of viral mRNA expressed at different times after infection. This is also seen in the replication of the eukaryotic poxviruses, which replicate in the cytoplasm of the infected host cell, and thus do not have access to cellular transcription machinery (see Chapter 19).

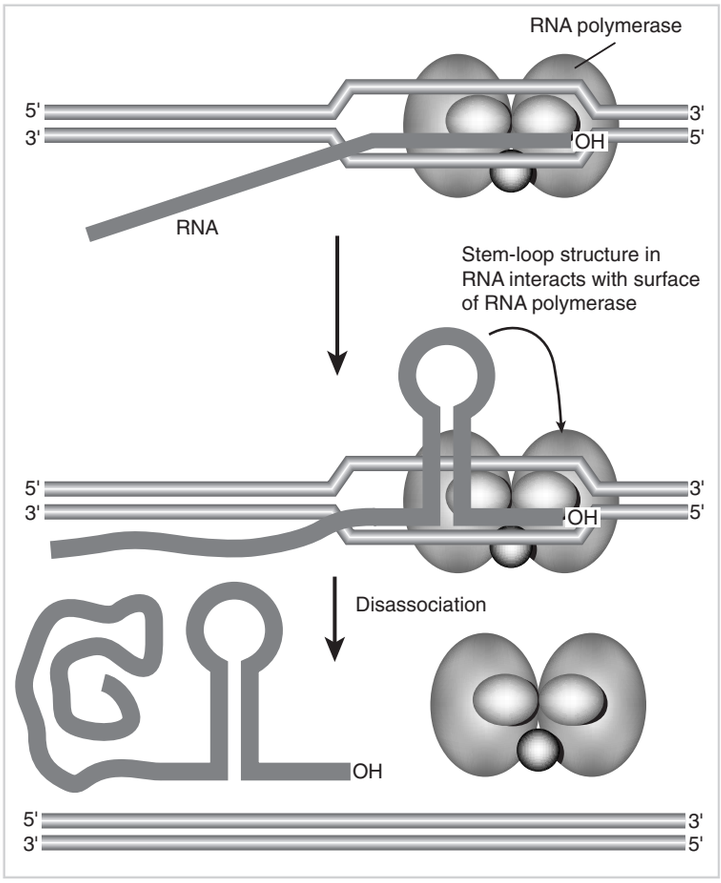


Fig. 13.11 Termination of bacterial transcription independent of the ρ factor. This type of termination, which has some features reminiscent of cleavage/polyadenylation of eukaryotic transcripts, involves the RNA polymerase encountering a specific destabilizing sequence in the DNA template, leading to disassociation of the enzyme from the template.

One other posttranscriptional modification, **RNA editing**, has been observed in the replication of some viruses. RNA editing is an enzymatic process that is commonly seen in the biogenesis of mitochondrial mRNAs. One form of editing is found in the replication of hepatitis delta virus (see Chapter 16). This editing reaction results in the deamination of an adenosine base in the viral mRNA and its conversion to a guanosine, which leads to alteration of a translation signal and expression of a larger protein than is expressed from the unmodified transcript. A second form of RNA editing that occurs as the RNA is expressed is the addition of extra bases to regions of the RNA. This is seen in the replication of some parainfluenza viruses and in Ebola virus.

The mechanism of protein synthesis

Like transcription, the process of protein synthesis is similar in broad outline in prokaryotes and eukaryotes; however, there are significant differences in detail. Some of these differences have important implications in the strategies that viruses must use to regulate gene expression. Viruses use the machinery of the cell for the translation of proteins, and to date, no virus has been characterized that encodes ribosomal proteins or rRNA. However, some viruses do modify ribosome-associated translation factors to ensure expression of their own proteins. A notable example of such a modification is found in the replication cycle of poliovirus described in Chapter 15.

Eukaryotic translation

In a nucleated cell, processed mRNA must be transported from the nucleus. The mRNA does not exist as a free RNA molecule, but is loosely or closely associated with one or a number of RNA-binding proteins that carry out the transport process and may facilitate initial association with the eukaryotic ribosome. This provides yet another point in the flow of information from gene to protein that is subject to modulation or control and thus, is potentially available for viral-encoded mediation.

The features of translational initiation in eukaryotic cells reflect the nature of eukaryotic mRNAs, namely, that they have 5′-methylated caps, that they are translated as monocistronic species, and that ribosomes usually do not bind to the messages at internal sites. Initiation involves assembly of the large (60s) and small (40s) subunits of the ribosome along with the initiator tRNA (met-tRNA in most cases) at the correct AUG codon. These steps require the action of several protein factors along with energy provided by ATP and GTP hydrolysis. The process is shown in Fig. 13.12.

The first phase of this process involves association of the 40s subunit with met-tRNAmet and is carried out by three **eukaryotic translation initiation factors** (eIF-2, eIF-3, and eIF-4C) along with GTP. This complex then binds to the 5′-methylated cap of the mRNA through the action of eIF-4A, eIF-4B, eIF-4F, and **CBP1 (cap-binding protein)** requiring the energy of ATP hydrolysis.

The 40s-tRNA complex then moves in the 5′ to 3′ direction along the mRNA, scanning the sequence for the appropriate AUG that is found within a certain sequence context (the **Kozak sequence**). Movement of the complex requires energy in the form of ATP. Finally, the 60s subunit joins the assembly through the activity of eIF-5 and eIF-6, GTP is hydrolyzed, all of the initiation factors are released, and the ribosome-mRNA complex (now called the 80s initiation complex) is ready for elongation.

The new peptide "grows" from N-terminal to C-terminal and reads the mRNA 5′ to 3′. Translation proceeds to the C-terminal amino acid of the nascent peptide chain, the codon of which is followed by a translation termination codon (UAA, UGA, or UAG). The sequence of bases, starting with the initiation codon, containing all the amino acid codons, and finishing with the three-base termination codon, defines an **open translational reading frame (ORF)**. In mature mRNA, any ORF will have a number of bases evenly divisible by three, but an ORF may be interrupted by introns in the gene encoding the mRNA.

Several ORFs can occur or overlap in the same region of mRNA, especially in viral genomes. Overlapping ORFs can be generated by splicing or by AUG initiation codons being separated by a number of bases not divisible by three. An example might be as follows (where lowercase bases represent those not forming codons):

5′-· · · AUGAAAUGGCCAUUUUAACGA · · · -3′

Translated in "frame 1," the sequence would be read:

AUG AAA UGG CCA UUU UAA

but in "frame 3," it would be read:

augaa AUG GCC AUU UUA A

In such an mRNA, ribosomes might start at one or the other translational reading frame, *but a given ribosome can only initiate translation at a single ORF*. Thus, if the ribosome initiates translation of,

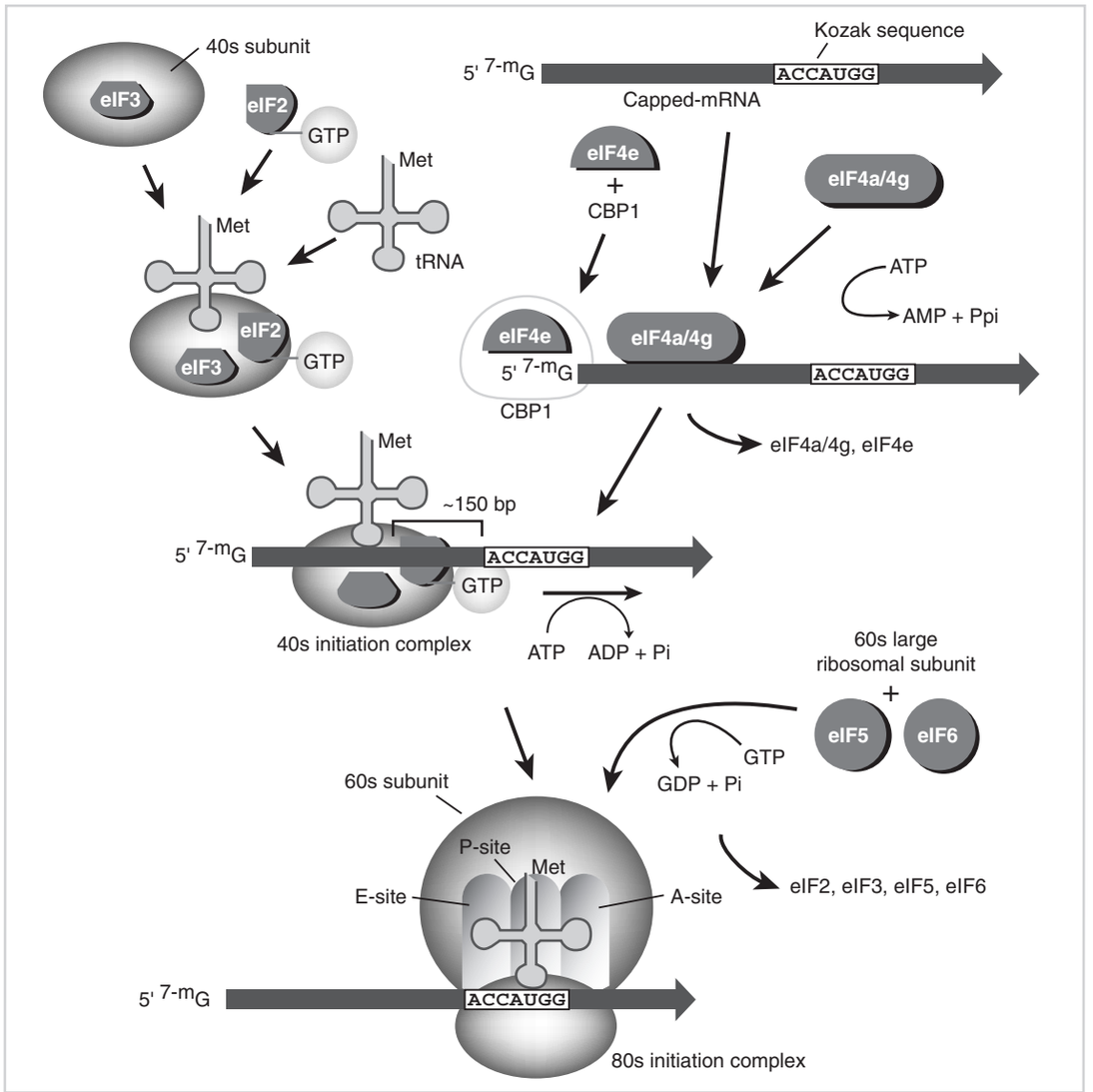


Fig. 13.12 Initiation of eukaryotic translation. Note the initiation complex contains the 40s ribosomal subunit and must interact with the 5′ end of the mRNA molecule via the cap structure or an equivalent. The 60s subunit only becomes associated with the complex at the Kozak (or equivalent) sequence. The ribosome dissociates back into the two subunits at the termination of translation. This means that internal initiation, especially if an upstream open reading frame has been translated, is impossible or at least extremely rare. (Pi, inorganic phosphate; Ppi, pyrophosphate.)

say, the second ORF, this is because it has missed the start of the first one, and if it has started at the first one, it cannot read any others on the mRNA. In other words, when a eukaryotic ribosome initiates translation at an ORF, it continues until a termination signal is encountered. Translation termination results in the ribosome falling off the mRNA strand, and any other potential translational reading frames downstream of the one terminated are essentially unreadable by the ribosome.

This simply means that a eukaryotic mRNA molecule containing multiple translational reading frames in sequence will not be able to express any beyond the first one translated (or, possibly two, if the ribosome has a "choice") from the 5′ end of the transcript. Any ORFs downstream of these are considered hidden or **cryptic ORFs**.

This property of eukaryotic translation has important implications both in the effect of splicing on revealing "cryptic" or hidden translational reading frames, and in the generation of some eukaryotic viral mRNAs.

Prokaryotic translation

Prokaryotic messages have three structural features that differ from eukaryotic versions. First, mRNA is not capped and methylated at the 5′ end. Second, mRNA may be translated into more than one protein from different coding sequences and is, thus, **polycistronic mRNA**. Finally, ribosome attachment to mRNA in prokaryotes occurs at internal sites rather than at the 5′ end. In addition, prokaryotic mRNAs are transcribed and translated at the same time and in the same place in the cell (**coupled transcription/translation**).

Features of prokaryotic translation reflect these structural and functional differences. Initiation, shown in Fig. 13.13, begins with the association of **initiator tRNA** (*N*-formyl-methionine-tRNA,

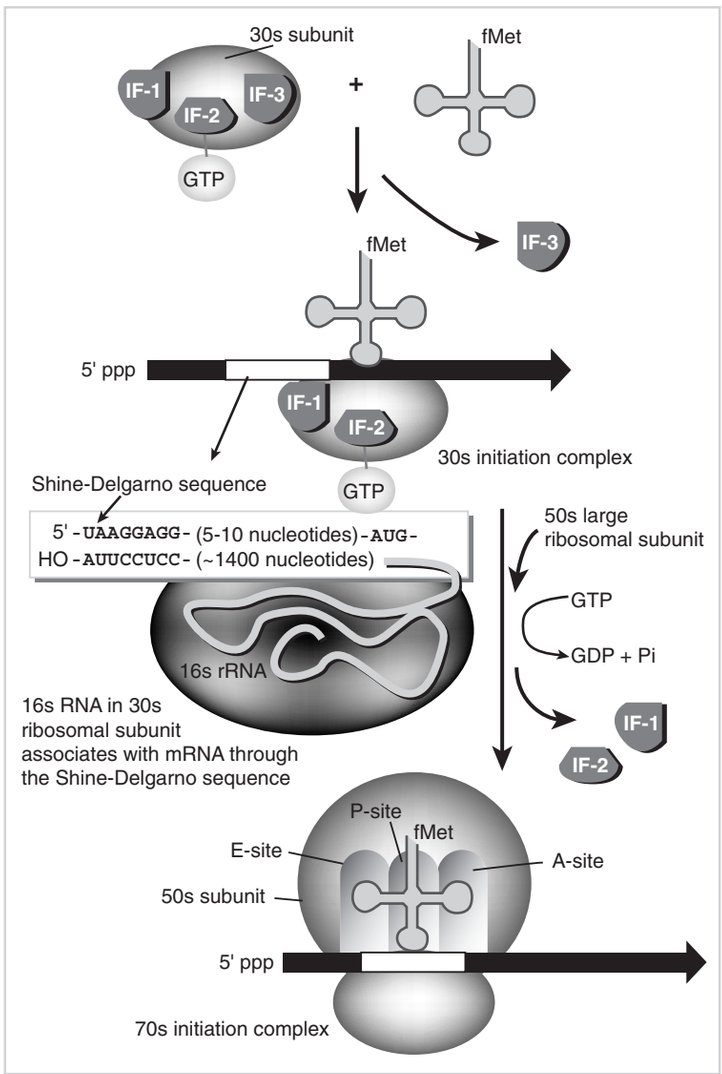


Fig. 13.13 Initiation of translation of a prokaryotic mRNA. This can occur anywhere there is a Shine-Dalgarno sequence in the mRNA since the 30s ribosome associates with the mRNA at that site by virtue of the presence of a complementary sequence in the 3′ end of the ribosomal RNA. (fMet, formylmethionine; Pi, inorganic phosphate.)

or Fmet-tRNA) with the small (30s) ribosomal subunit, together with mRNA through the action of three factors (IF-1, IF-2, and IF-3), along with GTP. The complex that forms involves direct binding of the 30s subunit with its Fmet-tRNA to the AUG that initiates translation of the ORF.

This AUG is defined by the presence of a series of bases (called the **Shine-Dalgarno sequence**) in the mRNA upstream from the start codon that is complementary to the 3′ end of the 16s rRNA in the 30s ribosomal subunit. The large (50s) ribosomal subunit now binds, accompanied by GTP hydrolysis and release of factors, to form the 70s initiation complex. From this point, elongation and termination occur in much the same manner as seen in eukaryotic cells.

Virus-induced changes in translation

Many viruses specifically alter or inhibit host cell protein synthesis. The ways they accomplish this vary greatly, and are described in Part IV where the replication cycle of specific viruses is covered in detail. Some viruses, notably retroviruses and some RNA viruses, can also *suppress* the termination of translation at a specific stop codon. The mechanism for such suppression may involve the ribosome actually skipping or jumping a base at the termination signal. When this happens, the translational reading frame being translated is shifted by a base or two. Other modes of suppression are not so well characterized, but all involve the mRNA at the site of suppression having a unique structure that facilitates it. This suppression is not absolute, but occurs with either high or low frequency resulting in a single mRNA translational reading frame being able to encode multiple, related proteins.

QUESTIONS FOR CHAPTER 13

1 A given mRNA molecule has the following structure. What is the maximum number of amino acids that the final protein product could contain?

Cap-300 bases-AUG-2097 bases-UAA-20 bases-AAAA

2 Assume the following sequence of bases occurs in an open reading frame (ORF) whose reading frame is indicated by grouping the capitalized bases three at a time:

. . . AUG . . . (300 bases) . . . CGC AAU ACA UGC CCU
ACC AUG AAU AAU ACC UAA gguaaaug . . .

What effect might deletion of the fourth A in the above strand of mRNA have on the size of a protein encoded by this ORF?

3 Both prokaryotic and eukaryotic cells transcribe mRNA from DNA and translate these mRNAs into proteins. However, there are differences between the two kinds of cells in the manner in which mRNAs are produced and utilized to program translation. In the table below, indicate which of the features applies to which kind of mRNA. Write "Yes" if the feature is true for that kind of mRNA or "No" if it is not true.

Continued

Feature	Eukaryotic mRNA	Prokaryotic mRNA
The small ribosomal subunit is correctly oriented to begin translation by association with the Shine-Dalgarno sequence.		
Open reading frames generally begin with an AUG codon.		
The 5′ end of the mRNA has a methylated cap structure covalently attached after transcription.		
During protein synthesis, an open reading frame can be translated by more than one ribosome, forming a polyribosome.		
Termination of transcription may occur at a site characterized by the formation of a GC-rich stem loop structure just upstream from a U-rich sequence.		

4 Which of the following statements is/are true regarding the primer for most DNA replication?
- **a** It is degraded by an exonuclease.
- **b** It is made up of ribonucleic acid.
- **c** It is synthesized by a primase.

5 All of the following are characteristics of *eukaryotic* mRNA, except:
- **a** A 5′-methylated guanine cap.
- **b** Polycistronic translation.
- **c** Polyadenylated 3′ tail.
- **d** Nuclear splicing of most mRNAs.
- **e** The use of AUGs instead of ATGs.

6 In what cellular location would one find viral glycoproteins being translated?

7 What is the *minimum size* of a viral mRNA encoding a structural protein of 1100 amino acids?