14 Transcription: Synthesis of RNA

*Synthesis of RNA from a **DNA template** is called **transcription.** Genes are transcribed by enzymes called **RNA polymerases** that generate a **single-stranded RNA** identical in sequence (with the exception of U in place of T) to one of the strands of the double-stranded DNA. The DNA strand that directs the sequence of nucleotides in the RNA by **complementary base-pairing** is the template strand. The RNA strand that is initially generated is **the primary transcript.** The DNA **template is copied** in the **3′ to 5′** direction, and the **RNA transcript** is **synthesized** in the **5′ to 3′** direction. RNA polymerases differ from DNA polymerases in that they can **initiate** the **synthesis** of new strands in the absence of a primer.*

*In addition to catalyzing the polymerization of **ribonucleotides,** RNA polymerases must be able to recognize the appropriate gene to transcribe, the appropriate strand of the double-stranded DNA to copy, and the **startpoint** of transcription (Fig. 14.1). Specific sequences on DNA, called **promoters**, determine where the RNA polymerase binds and how frequently it initiates transcription. Other regulatory sequences, such as **promoter-proximal elements** and **enhancers,** also affect the frequency of transcription.*

*In **bacteria,** a **single RNA polymerase** produces the primary transcript precursors for all three major classes of RNA: messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA). Because bacteria do not contain nuclei, ribosomes bind to mRNA as it is being transcribed, and protein synthesis occurs simultaneously with transcription.*

***Eukaryotic genes** are transcribed in the nucleus by **three different RNA polymerases,** each principally responsible for one of the major classes of RNA. The primary transcripts are **modified** and **trimmed** to produce the mature RNAs. The precursors of **mRNA** (called **pre-mRNA**) have a **guanosine** "cap" added at the 5′-end and a **poly(A) "tail"** at the 3′-end. **Exons,** which contain the coding sequences for the proteins, are separated in **pre-mRNA** by **introns,** regions that have no coding function. During **splicing reactions,** introns are removed and the exons connected to form the mature mRNA. In eukaryotes, tRNA and rRNA precursors are also modified and trimmed, although not as extensively as pre-mRNA.*

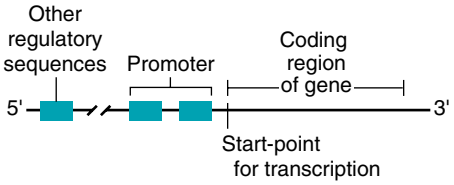


Fig. 14.1. Regions of a gene. A gene is a segment of DNA that functions as a unit to generate an RNA product or, through the processes of transcription and translation, a polypeptide chain. The transcribed region of a gene contains the template for synthesis of an RNA, which begins at the startpoint. A gene also includes regions of DNA that regulate production of the encoded product, such as a promoter region. In a structural gene, the transcribed region contains the coding sequences that dictate the amino acid sequence of a polypeptide chain.

THE WAITING ROOM

Anne Niemick is a 4-year-old girl of Mediterranean ancestry whose height and body weight are below the 20th percentile for girls of her age. She is listless, tires easily, and complains of loss of appetite and shortness of breath on exertion. A dull pain has been present in her right upper quadrant for

The thalassemias are a heterogenous group of hereditary anemias that constitute the most common gene disorder in the world, with a carrier rate of almost 7%. The disease was first discovered in countries around the Mediterranean Sea and was named for the Greek word "thalassa" meaning "sea". However, it is also present in areas extending into India and China that are near the equator.

The thalassemia syndromes are caused by mutations that decrease or abolish the synthesis of the α or β chains in the adult hemoglobin A tetramer. Individual syndromes are named according to the chain whose synthesis is affected and the severity of the deficiency. Thus, in β^0 thalassemia, the superscript 0 denotes none of the β chain is present; in β^+ thalassemia, the + denotes a partial reduction in the synthesis of the β chain. More than 170 different mutations have been identified that cause β thalassemia; most of these interfere with the transcription of β-globin mRNA or its processing or translation.

the last 3 months. Her complexion is slate-gray and she appears pale. Initial laboratory studies indicate a severe anemia (decreased red blood cell count) with a hemoglobin of 6.2 g/dL (reference range, 12–16). A battery of additional hematologic tests shows that Anne has β^+ -thalassemia, intermediate type.

Ivy Sharer, a patient with AIDS (see Chapters 12 and 13), has developed a cough that produces a gray, slightly blood-tinged sputum. A chest radiograph indicates infiltrates in the cavities of both upper lung fields (cavitary infiltrates). A stain of sputum shows the presence of acid-fast bacilli, suggesting a diagnosis of pulmonary tuberculosis caused by *Mycobacterium tuberculosis*.

Amanda Tin picked mushrooms in a wooded area near her home. A few hours after eating one small mushroom, she experienced mild nausea and diarrhea. She brought a mushroom with her to the hospital emergency room. A poison expert identified it as *Amanita phalloides* (the "death cap"). These mushrooms contain the toxin α-amanitin.

Sis Lupus, a 28-year-old computer programmer, notes increasing fatigue, pleuritic chest pain, and a nonproductive cough. In addition, she complains of joint pains, especially in her hands. A rash on both cheeks and the bridge of her nose ("butterfly rash") has been present for the last 6 months. Initial laboratory studies indicate a subnormal white blood cell count and a mild reduction in hemoglobin. Tests result in a diagnosis of systemic lupus erythematous (SLE) (frequently called lupus).

I. ACTION OF RNA POLYMERASE

Transcription, the synthesis of RNA from a DNA template, is carried out by RNA polymerases (Fig. 14.2). Like DNA polymerases, RNA polymerases catalyze the formation of ester bonds between nucleotides that base-pair with the complementary nucleotides on the DNA template. Unlike DNA polymerases, RNA polymerases can initiate the synthesis of new chains in the absence of primers. They also lack the 3′ to 5′ exonuclease activity found in DNA polymerases. A strand of DNA serves as the template for RNA synthesis and is copied in the 3′ to 5′ direction. Synthesis of the new RNA molecule occurs in the 5′ to 3′ direction. The ribonucleoside triphosphates ATP, GTP, CTP, and UTP serve as the precursors. Each nucleotide base sequentially pairs with the complementary deoxyribonucleotide base on the DNA template (A, G, C, and U pair with T, C, G and A, respectively). The polymerase forms an ester bond between the α-phosphate on the ribose 5′-hydroxyl of the nucleotide precursor and the ribose 3′-hydroxyl at the end of the growing RNA chain. The cleavage of a high-energy phosphate bond in the nucleotide triphosphate and release of pyrophosphate (from the β and γ phosphates) provides the energy for this polymerization reaction. Subsequent cleavage of the pyrophosphate by a pyrophosphatase also helps to drive the polymerization reaction forward by removing a product.

RNA polymerases must be able to recognize the startpoint for transcription of each gene and the appropriate strand of DNA to use as a template. They also must be sensitive to signals that reflect the need for the gene product and control the frequency of transcription. A region of regulatory sequences called the promoter, usually contiguous with the transcribed region, controls the binding of RNA polymerase to DNA and identifies the startpoint (see Fig. 14.1). The frequency of transcription is controlled by regulatory sequences within the promoter, nearby the promoter (promoter-proximal elements), and by other regulatory sequences, such as enhancers, that may be located at considerable distances, sometimes thousands of nucleotides, from the startpoint. Both the promoter-proximal elements, and the enhancers interact with proteins that stabilize RNA polymerase binding to the promoter.

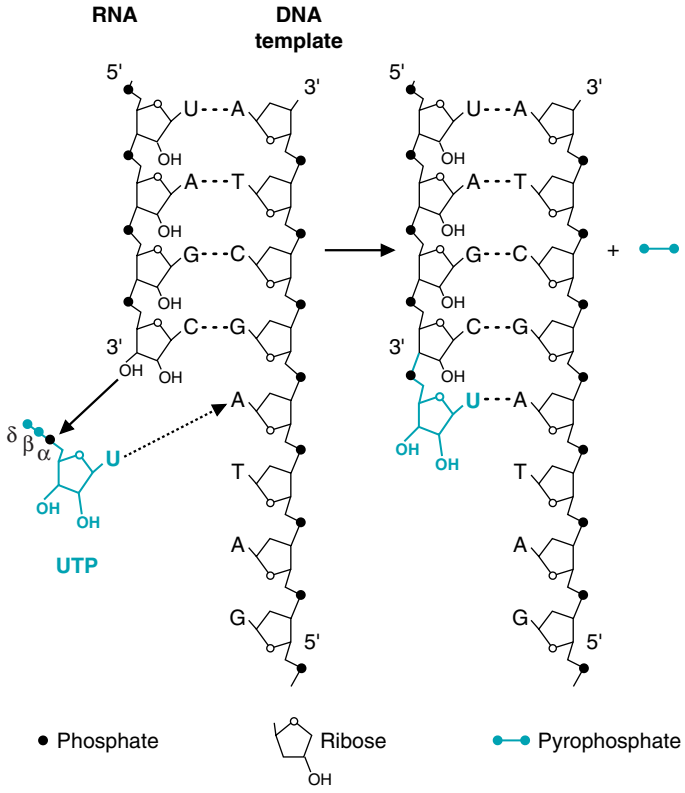


Fig. 14.2. RNA synthesis. The α-phosphate from the added nucleotide (shown in black) connects the ribosyl groups.

II. TYPES OF RNA POLYMERASES

Bacterial cells have a single RNA polymerase that transcribes DNA to generate all of the different types of RNA (mRNA, rRNA, and tRNA). The RNA polymerase of *Escherichia coli* contains four subunits ($\alpha_2\beta\beta'$), which form the core enzyme. Another protein called a σ (sigma) factor binds the core enzyme and directs binding of RNA polymerase to specific promoter regions of the DNA template. The σ factor dissociates shortly after transcription begins. *E. coli* has a number of different σ factors that recognize the promoter regions of different groups of genes. The major σ factor is σ^{70}, a designation related to its molecular weight of 70,000 Daltons.

In contrast to prokaryotes, eukaryotic cells have three RNA polymerases (Table 14.1). Polymerase I produces most of the rRNAs, polymerase II produces mRNA, and polymerase III produces small RNAs, such as tRNA and 5S rRNA. All of these RNA polymerases have the same mechanism of action. However, they recognize different types of promoters.

A. Sequences of Genes

Double-stranded DNA consists of a coding strand and a template strand (Fig. 14.3) The DNA template strand is the strand that is actually used by RNA polymerase during the process of transcription. It is complementary and antiparallel both to the coding (nontemplate) strand of the DNA and to the RNA transcript produced from the template. Thus, the coding strand of the DNA is identical in base sequence and direction to the RNA transcript, except, of course, that wherever this DNA strand contains a T, the RNA transcript contains a U. By convention, the

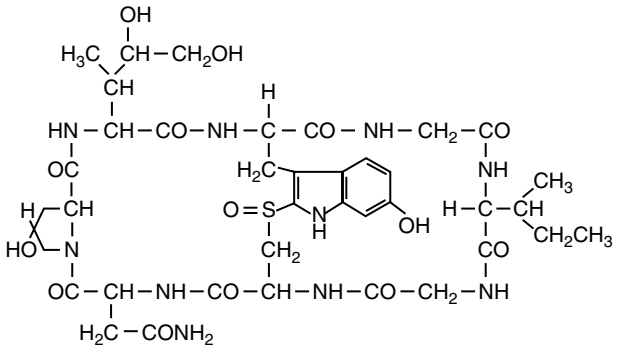
Patients with acquired immune deficiency syndrome (AIDS) frequently develop tuberculosis. After **Ivy Sharer**'s sputum stain suggested that she had tuberculosis, a multidrug antituberculous regimen, which includes an antibiotic of the rifamycin family (rifampin), was begun. A culture of her sputum was taken to confirm the diagnosis.

Rifampin inhibits bacterial RNA polymerase, selectively killing the bacteria that cause the infection. The nuclear RNA polymerase from eukaryotic cells is not affected. Although rifampin can inhibit the synthesis of mitochondrial RNA, the concentration required is considerably higher than that used for treatment of tuberculosis.

Table 14.1 Products of Eukaryotic RNA Polymerases

RNA polymerase I:	RNA
RNA polymerase II:	mRNA
RNA polymerase III:	tRNA + other small RNAs

The mushrooms picked by **Amanda Tin** contained α-amanitin, an inhibitor of eukaryotic RNA polymerases. It is particularly effective at blocking the action of RNA polymerase II. This toxin initially causes gastrointestinal disturbances, then electrolyte imbalance and fever, followed by liver and kidney dysfunction. Between 40 and 90% of the individuals who ingest α-amanitin die within a few days.



α–Amanitin

nucleotide sequence of a gene is represented by the letters of the nitrogenous bases of the coding strand of the DNA duplex. It is written from left to right in the 5′ to 3′ direction.

During translation, mRNA is read 5′ to 3′ in sets of three bases, called codons, that determine the amino acid sequence of the protein (see Fig. 14.3) Thus, the base sequence of the coding strand of the DNA can be used to determine the amino acid sequence of the protein. For this reason, when gene sequences are given, they refer to the coding strand.

A gene consists of the transcribed region and the regions that regulate transcription of the gene (e.g., promoter and enhancer regions)(Fig. 14.4). The base in the coding strand of the gene serving as the startpoint for transcription is numbered +1. This nucleotide corresponds to the first nucleotide incorporated into the RNA at the 5′-end of the transcript. Subsequent nucleotides within the transcribed region of the gene are numbered +2, +3, etc., toward the 3′-end of the gene. Untranscribed sequences to the left of the startpoint, known as the 5′-flanking region of the gene, are numbered –1, –2, –3, etc., starting with the nucleotide (–1) immediately to the left of the startpoint (+1) and moving from right to left. By analogy to a river, the sequences to the left of the startpoint are said to be upstream from the startpoint and those to the right are said to be downstream.

B. Recognition of Genes by RNA Polymerase

For genes to be expressed, RNA polymerase must recognize the appropriate point to start transcription and the strand of the DNA to transcribe (the template strand).

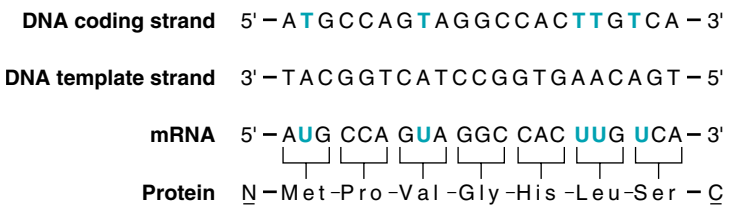


Fig. 14.3. Relationship between the coding strand of DNA, the DNA template strand, the mRNA transcript, and the protein produced from the gene. The bases in mRNA are used in sets of three (called codons) to specify the order of the amino acids inserted into the growing polypeptide chain during the process of translation (see Chapter 15).

The two strands of DNA are antiparallel, with complementary nucleotides at each position. Thus, each strand would produce a different mRNA, resulting in different codons for amino acids and a different protein product. Therefore, it is critical that RNA polymerase transcribe the correct strand.

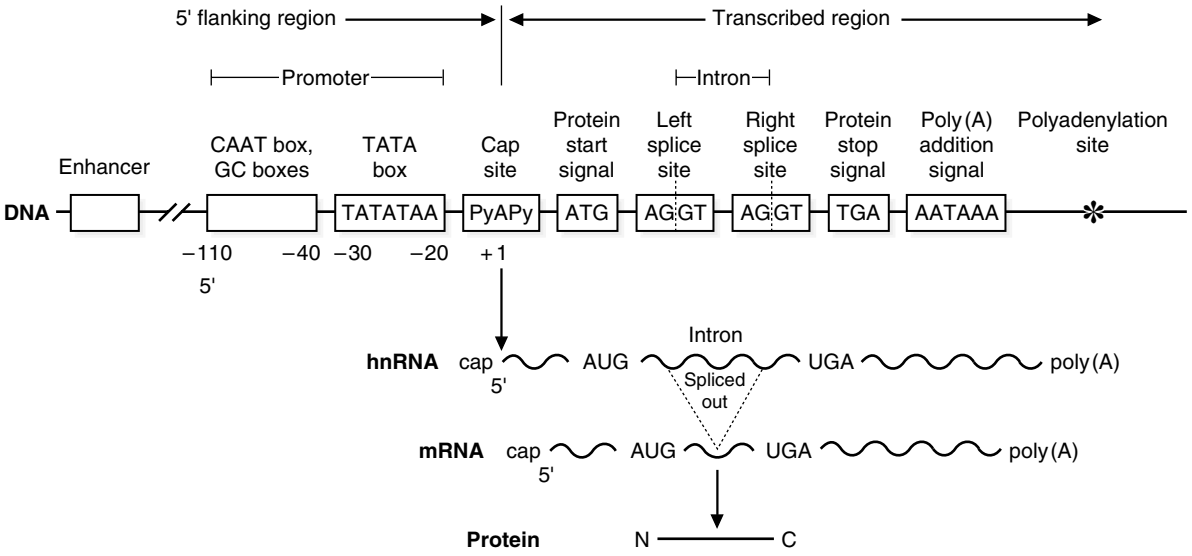


Fig. 14.4. A schematic view of a eukaryotic gene, and steps required to produce a protein product. The gene consists of promoter and transcribed regions. The transcribed region contains introns, which do not contain coding sequence for proteins, and exons, which do carry the coding sequences for proteins. The first RNA form produced is heterogenous nuclear RNA (hn RNA), which contains both intronic and exonic sequences. The hnRNA is modified such that a cap is added at the 5′ end (cap site), and a poly-A tail added to the 3′ end. The introns are removed (a process called splicing) to produce the mature mRNA, which leaves the nucleus to direct protein synthesis in the cytoplasm. Py is pyrimidine (C or T).

RNA polymerase also must recognize which genes to transcribe because transcribed genes are only a small fraction of the total DNA. The genes that are transcribed differ from one type of cell to another and change with changes in physiologic conditions. The signals in DNA that RNA polymerase recognizes are called promoters. Promoters are sequences in DNA (often composed of smaller sequences called boxes or elements) that determine the startpoint and the frequency of transcription. Because they are located on the same molecule of DNA and near the gene they regulate, they are said to be *cis* acting (i.e., *"cis"* refers to acting on the same side). Proteins that bind to these DNA sequences and facilitate or prevent the binding of RNA polymerase are said to be *trans* acting.

C. Promoter Regions of Genes for mRNA

The binding of RNA polymerase and the subsequent initiation of gene transcription involves a number of consensus sequences in the promoter regions of the gene (Fig. 14.5). A consensus sequence is the sequence most commonly found in a given region when many genes are examined. In both prokaryotes and eukaryotes, an adenine- and thymine-rich consensus sequence in the promoter determines the startpoint of transcription by binding proteins that facilitate the binding of RNA polymerase. In the prokaryote *E. coli*, this consensus sequence is TATAAT, which is known as the TATA or Pribnow box. It is centered about –10 and is recognized by the sigma factor σ^{70}. A similar sequence in the –25 region of eukaryotic genes has a consensus sequence of TATA(A/T)A. (The (A/T) in the fifth position indicates that either A or T occurs with equal frequency.) This eukaryotic sequence is also known as a TATA box, but is sometimes named the Hogness or Hogness-Goldberg box after its discoverers. Other consensus sequences involved in binding of RNA polymerase are found further upstream in the promoter region (see Fig. 14.5). Bacterial promoters contain a sequence TTGACA in the –35 region. Eukaryotes frequently have CAAT boxes and GC-rich sequences in the region between –40 and –110. Eukaryotic genes also contain promoter-proximal elements (in the region of –100

Anne Niemick has a β^+ thalassemia classified clinically as β-thalassemia intermedia. She produces an intermediate amount of functional β globin chains (her hemoglobin is 6.2 g/dL; normal is 12–16). β-thalassemia intermedia is usually the result of two different mutations (one that mildly affects the rate of synthesis of β-globin and one severely affecting its rate of synthesis), or, less frequently, homozygosity for a mild mutation in the rate of synthesis, or a complex combination of mutations. For example, mutations within the promoter region of the β-globin gene could result in a significantly decreased rate of β-globin synthesis in an individual who is homozygous for the allele, without completely abolishing synthesis of the protein.

Two of the point mutations that result in a β^+ phenotype are within the TATA box (A→ G or A→ C in the −28 to −31 region). These mutations reduce the accuracy of the startpoint of transcription so that only 20 to 25% of the normal amount of β-globin is synthesized. Other mutations that also reduce the frequency of β-globin transcription have been observed further upstream in the promoter region (−87 C→ G and −88 C→T).

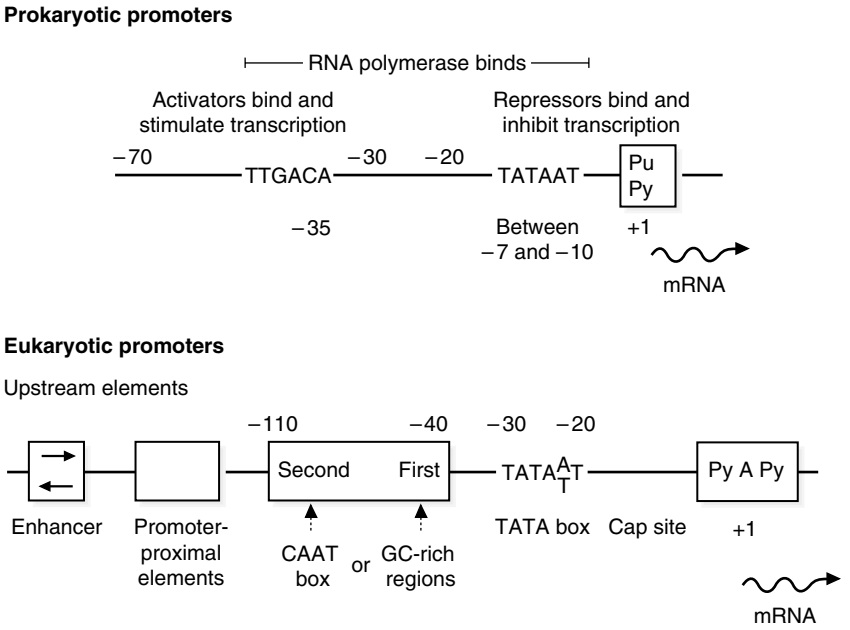


Fig. 14.5. Prokaryotic and eukaryotic promoters. The promoter-proximal region contains binding sites for transcription factors which that can accelerate the rate at which RNA polymerase binds to the promoter. Pu = purine; Py = pyrimidine.

Q: What property of an AT-rich region of a DNA double helix makes it suitable to serve as a recognition site for the startpoint of transcription?

to −200), which are sites that bind other gene regulatory proteins. Genes vary in the number of such sequences present.

In bacteria, a number of protein-producing genes may be linked together and controlled by a single promoter. This genetic unit is called an operon (Fig. 14.6). Proteins bind to the promoter and either inhibit or facilitate transcription of the operon. Repressors are proteins that bind to a region in the promoter known as the operator and inhibit transcription by preventing the binding of RNA polymerase to DNA. Activators are proteins that stimulate transcription by binding within the –35 region or upstream from it, facilitating the binding of RNA polymerase. (Operons are described in more detail in Chapter 16.)

In eukaryotes, proteins known as general transcription factors (or basal factors) bind to the TATA box and facilitate the binding of RNA polymerase II, the polymerase that transcribes mRNA (Fig. 14.7). This binding process involves at least six basal transcription factors (labeled as TFIIs, transcription factors for RNA polymerase II). The TATA-binding protein (TBP), which is a component of TFIID, initially binds to the TATA box. TFIID consists of both the TBP and a number of transcriptional coactivators. TFIIA and TFIIB interact with TBP. RNA polymerase II binds to the complex of transcription factors and to DNA, and is aligned at the startpoint for transcription. TFIIE, TFIIF, and TFIIH subsequently bind, cleaving adenosine triphosphate (ATP), and transcription of the gene is initiated.

With only these transcription (or basal) factors and RNA polymerase II attached (the basal transcription complex), the gene is transcribed at a low or basal rate.

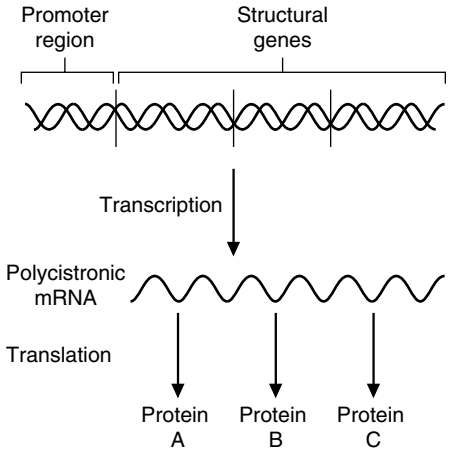


Fig. 14.6. Bacterial operon. A cistron encodes a single polypeptide chain. In bacteria, a single promoter may control transcription of an operon containing many cistrons. A single polycistronic mRNA is transcribed. Its translation produces a number of polypeptide chains.

TFIIH plays a number of roles in both transcription and DNA repair. In both processes, it acts as an ATP-dependent DNA helicase, unwinding DNA for either transcription or repair to occur. Two of the forms of xeroderma pigmentosum (XPB and XPD; see Chapter 13) arise from mutations within two different helicase subunits of TFIIH. TFIIH also contains a kinase activity, and RNA polymerase II is phosphorylated by this factor during certain phases of transcription.

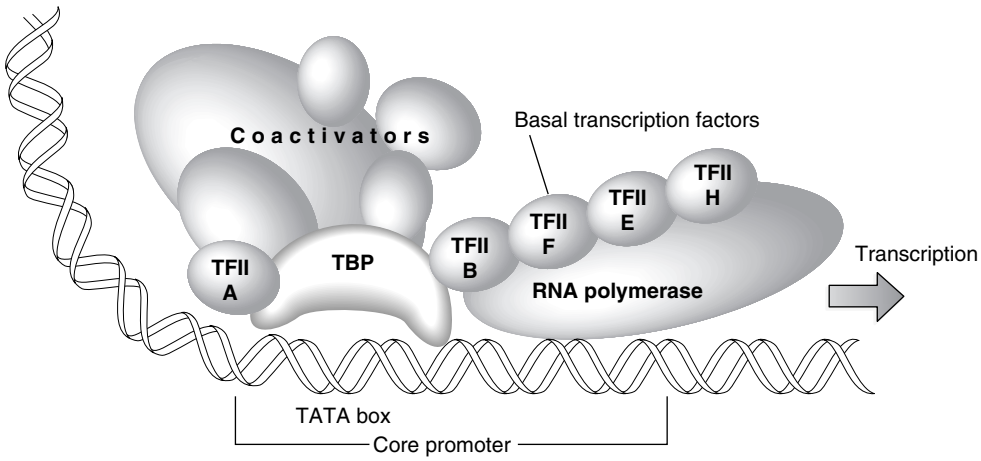


Fig. 14.7. Transcription apparatus. The TATA-binding protein (TBP), a component of TFIID, binds to the TATA box. Transcription factors TFII A and B bind to TBP. RNA polymerase binds, then TFII E, F, and H bind. This complex can transcribe at a basal level. Some coactivator proteins are present as a component of TFIID, and these can bind to other regulatory DNA binding proteins (called specific transcription factors or transcriptional activators).

The rate of transcription can be further increased by binding of other regulatory DNA binding proteins to additional gene regulatory sequences (such as the promoter proximal or enhancer regions). These regulatory DNA binding proteins are called gene-specific transcription factors (or transactivators) because they are specific to the gene involved (see Chapter 16). They interact with coactivators in the basal transcription complex.

A: In regions in which DNA is being transcribed, the two strands of the DNA must be separated. AT base pairs in DNA are joined by only two hydrogen bonds, whereas GC pairs have three hydrogen bonds. Therefore, in AT-rich regions of DNA, the two strands can be separated more readily than in regions that contain GC base pairs.

III. TRANSCRIPTION OF BACTERIAL GENES

In bacteria, binding of RNA polymerase with a σ factor to the promoter region of DNA causes the two DNA strands to unwind and separate within a region approximately 10 to 20 nucleotides in length. As the polymerase transcribes the DNA, the untranscribed region of the helix continues to separate, whereas the transcribed region of the DNA template rejoins its DNA partner (Fig. 14.8). The sigma factor is released when the growing RNA chain is approximately 10 nucleotides long. The elongation reactions continue until the RNA polymerase encounters a transcription termination signal. One type of termination signal involves the formation of a hairpin loop in the transcript, preceding a number of U residues. The second type of

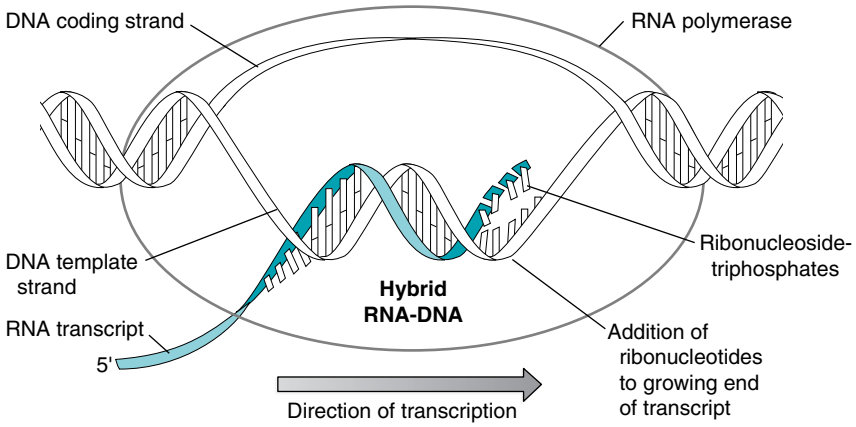


Fig. 14.8. Mechanism of transcription.

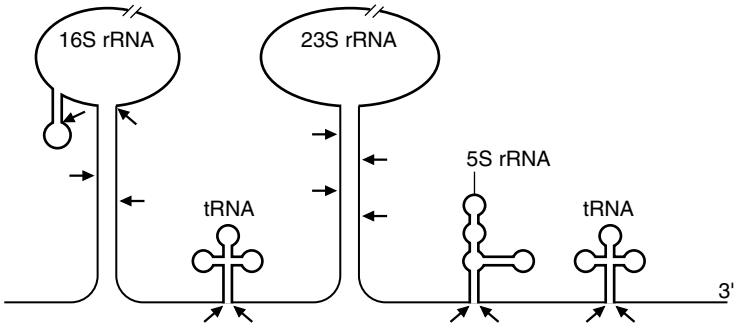


Fig. 14.9. Bacterial rRNA and tRNA transcripts. One large precursor is cleaved (at arrows) to produce 16S, 23S, and 5S rRNA and some tRNAs.

mechanism for termination involves the binding of a protein, the rho factor, which causes release of the RNA transcript from the template. The signal for both termination processes is the sequence of bases in the newly synthesized RNA.

A cistron is a region of DNA that encodes a single polypeptide chain. In bacteria, mRNA is usually generated from an operon as a polycistronic transcript (one that contains the information to produce a number of different proteins). The polycistronic transcript is translated as it is being transcribed. This transcript is not modified and trimmed, and it does not contain introns (regions within the coding sequence of a transcript that are removed before translation occurs). Several different proteins are produced during translation of the polycistronic transcript, one from each cistron (see Fig. 14.6).

In prokaryotes, rRNA is produced as a single, long transcript that is cleaved to produce the 16S, 23S, and 5S ribosomal RNAs. tRNA is also cleaved from larger transcripts (Fig. 14.9). One of the cleavage enzymes, RNase P, is a protein containing an RNA molecule. This RNA actually catalyzes the cleavage reaction.

IV. TRANSCRIPTION OF EUKARYOTIC GENES

The process of transcription in eukaryotes is similar to that in prokaryotes. RNA polymerase binds to the transcription factor complex in the promoter region and to the DNA, the helix unwinds within a region near the startpoint of transcription, DNA strand separation occurs, synthesis of the RNA transcript is initiated, and the RNA transcript is elongated, copying the DNA template. The DNA strands separate as the polymerase approaches and rejoin as the polymerase passes.

One of the major differences between eukaryotes and prokaryotes is that eukaryotes have more elaborate mechanisms for processing the transcripts, particularly the precursors of mRNA (pre-mRNA). Eukaryotes also have three polymerases, rather than just the one present in prokaryotes. Other differences include the facts that eukaryotic mRNA usually contains the coding information for only one polypeptide chain and that eukaryotic RNA is transcribed in the nucleus and migrates to the cytoplasm where translation occurs.

Although each eukaryotic mRNA only codes for one polypeptide chain, a number of proteins can contain different polypeptide chains or have multiple active sites, allowing the protein to catalyze more than one reaction.

The terms hnRNA (heterogeneous nuclear RNA) and pre-mRNA are both used to denote mRNA precursors. The term hnRNA was originally applied to a pool of RNA molecules in the nucleus that were rapidly synthesized and varied greatly in size. These RNA molecules are now known to be the mRNA precursors that vary greatly in size because they contain exons that encode different sizes of polypeptide chains and introns that vary in amount and size.

A. Synthesis of Eukaryotic mRNA

In eukaryotes, extensive processing of the primary transcript occurs before the mature mRNA is formed and can migrate to the cytosol, where it is translated into a protein product. RNA polymerase II synthesizes a large primary transcript from the template strand that is capped at the 5′ end as it is transcribed (Fig. 14.10). The transcript also rapidly acquires a poly(A) tail at the 3′ end. Pre-mRNAs thus contain untranslated regions at both the 5′ and 3’ ends (the leader and trailing sequences, respectively). These untranslated regions are retained in the mature

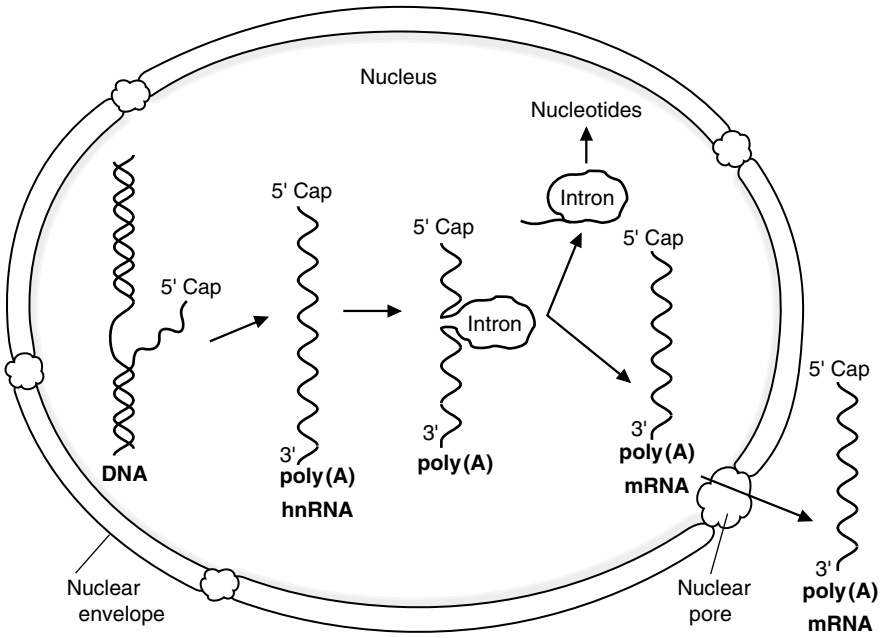


Fig. 14.10. Overview of mRNA synthesis. Transcription produces hnRNA from the DNA template. hnRNA processing involves addition of a 5′-cap and a poly(A) tail and splicing to join exons and remove introns. The product, mRNA, migrates to the cytoplasm, where it will direct protein synthesis.

mRNA. The coding region of the pre-mRNA, which begins with the start codon for protein synthesis and ends with the stop codon, contains both exons and introns. Exons consist of the nucleotide codons that dictate the amino acid sequence of the eventual protein product. Between the exons, interspersing regions called introns contain nucleotide sequences that are removed by splicing reactions to form the mature RNA. The mature RNA thus contains a leader sequence (that includes the cap), a coding region comprising exons, and a tailing sequence that includes the poly(A) tail.

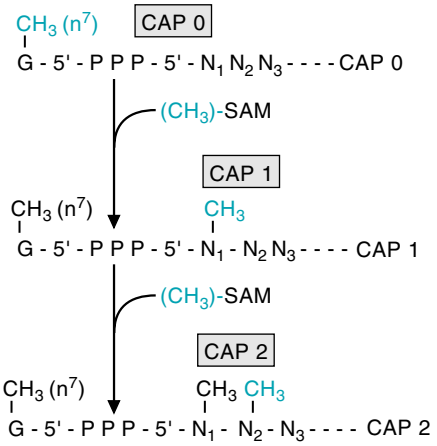
This mature mRNA complexes with the poly(A) binding protein and other proteins. It travels through pores in the nuclear envelope into the cytoplasm. There it combines with ribosomes and directs the incorporation of amino acids into proteins.

1. TRANSCRIPTION AND CAPPING OF mRNA

"Capping" of the primary transcript synthesized by RNA polymerase II occurs at its 5′-end as it is being transcribed (Fig.14.11). The 5′- terminal, the initial nucleotide of the transcript, is a pyrimidine with three phosphate groups attached to the 5′- hydroxyl of the ribose. To form the cap, the terminal triphosphate loses one phosphate, forming a 5′-diphosphate. The β-phosphate of the diphosphate then attacks the α-phosphate of GTP, liberating pyrophosphate, and forming an unusual 5′ to 5′ triphosphate linkage. A methyl group is transferred from S-adenosylmethionine (SAM), a universal methyl donor, to position 7 of the added guanine ring. Methylation also occurs on the ribose 2′-hydroxyl group in the terminal nucleotide to which the cap is attached, and sometimes the 2′-hydroxyl group of the adjacent nucleotide ribose. This cap "seals" the 5′ end of the primary transcript and decreases the rate of degradation. It also serves as a recognition site for the binding of the mature mRNA to a ribosome at the initiation of protein synthesis.

Once SAM (S-adenosylmethionine) donates its methyl group, it must be regenerated by reactions that require the vitamins folate and B_{12}. Thus formation of mRNA is also one of the processes affected by a deficiency of these vitamins.

There are three different types of methyl caps, shown in blue: CAP0 refers to the methylated guanosine (on the nitrogen at the seven position, N^7) added in the 5′ to 5′ linkage to the mRNA; CAP1 refers to CAP0 with the addition of a methyl to the 22′ carbon of ribose on the nucleotide (N_1) at the 5′ end of the chain; and CAP2 refers to CAP1 with the addition of another 2′ methyl group to the next nucleotide (N_2). The methyl groups are donated by S-adenosylmethionine (SAM).



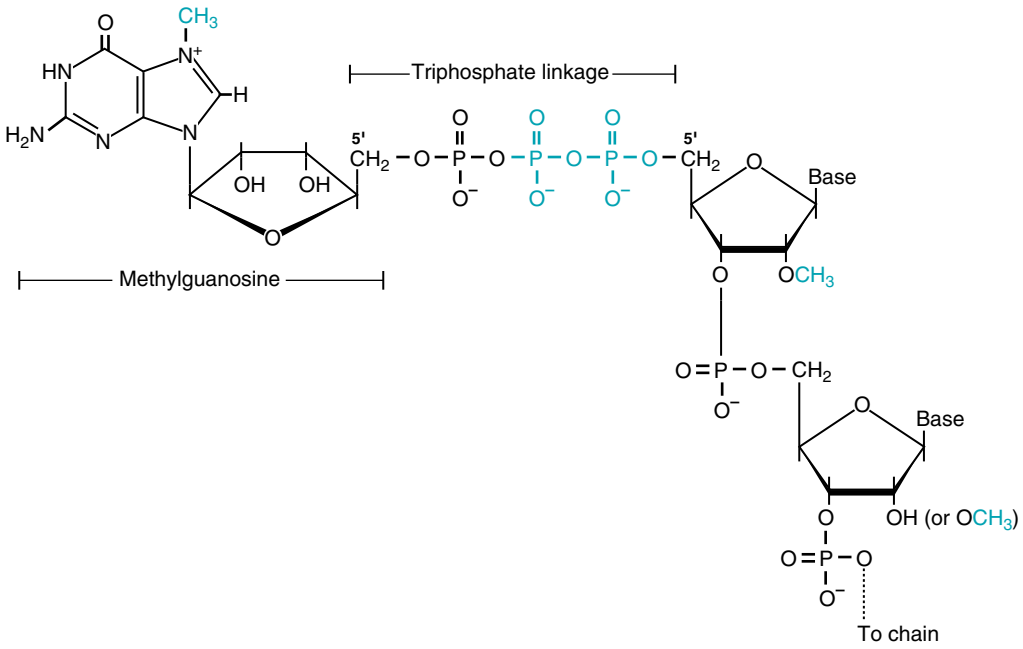


Fig. 14.11. The cap structure in eukaryotic mRNA. The phosphates in blue originated from the original RNA transcript; the phosphate in black comes from GTP.

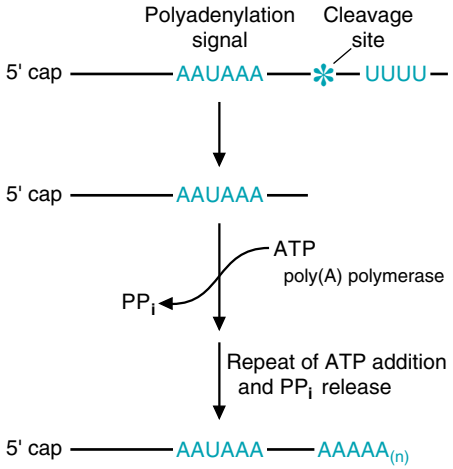


Fig. 14.12. Synthesis of the poly(A) tail. As RNA polymerase continues to transcribe the DNA, enzymes cleave the transcript (hnRNA) at a point 10–20 nucleotides beyond an AAUAAA sequence, just before a run of Us (or Gs). Approximately 250 adenine nucleotides are then added to the 3′-end of the hnRNA, one at a time, by poly(A) polymerase.

2. ADDITION OF A POLY(A) TAIL

After the RNA polymerase transcribes the stop codon for protein translation, it passes a sequence called the polyadenylation signal (AAUAAA) (Fig. 14.12). It continues past the polyadenylation signal until it reaches an unknown, and possibly unspecific, termination signal many nucleotides later. However, as the primary transcript is released from the RNA polymerase elongation complex, an enzyme complex binds to the polyadenylation signal and cleaves the primary transcript approximately 10 to 20 nucleotides downstream, thereby forming the 3′ end. After this cleavage, a poly(A) tail that can be over 200 nucleotides in length is added to the 3′-end. Thus, there is no poly(dT) sequence in the DNA template that corresponds to this tail; it is added posttranscriptionally. ATP serves as the precursor for the sequential addition of the adenine nucleotides. They are added one at a time, with poly(A) polymerase catalyzing each addition. The poly(A) tail is a protein binding site that protects the mRNA from degradation.

Within a few days of initiation of treatment for tuberculosis, laboratory staining results of **Ivy Sharer**'s sputum confirmed the diagnosis of pulmonary tuberculosis caused by *M. tuberculosis*. Therefore, the multidrug therapy, which included the antibiotic rifampin, was continued. Rifampin binds to the RNA polymerases of several bacteria. *M. tuberculosis* rapidly develops resistance to rifampin through mutations that result in an RNA polymerase that cannot bind the complex structure. Simultaneous treatment with a drug that works through a different mechanism decreases the selective advantage of the mutation and the rate at which resistance develops.

The presence of a poly (A) tail on eukaryotic mRNA allows this form of RNA to be easily separated from the more abundant rRNA. After extracting all of the RNA from a cell, the total RNA is applied to a column of beads to which oligo-dT has been covalently attached. As the mRNA flows through the column, its poly (A) tail will base pair with the oligo-dT, and the mRNA will become bound to the column. All other types of RNA will flow through the column and not bind to the beads. The bound mRNA can then be eluted from the column by changing the ionic strength of the buffer.

3. REMOVAL OF INTRONS

Eukaryotic pre-mRNA transcripts contain regions known as exons and introns. Exons appear in the mature mRNA; introns are removed from the transcript and are not found in the mature mRNA (see Fig. 14.10). Introns, therefore, do not contribute to the amino acid sequence of the protein. Some genes contain 50 or more introns. These introns are carefully removed from the pre-mRNA transcript and the exons spliced together, so that the appropriate protein is produced from the gene.

The consensus sequences at the intron/exon boundaries of the pre-mRNA are AGGU (AGGT in the DNA). The sequences vary to some extent on the exon side of the boundaries, but almost all introns begin with a 5′ GU and end with a 3′ AG (Fig. 14.13). These intron sequences at the left splice site and the right splice site are, therefore, invariant. Because every 5′ GU and 3′ AG combination does not result in a functional splice site, clearly other features within the exon or intron help to define the appropriate splice sites. These features, which are currently being identified, involve the presence of positive- and negative-acting *cis*-regulatory sequences within the intron.

A complex structure known as a spliceosome ensures that exons are spliced together with great accuracy (Fig. 14.14). Small nuclear ribonucleoproteins

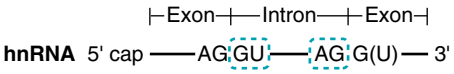


Fig. 14.13. Splice junctions in hnRNA. The intron sequences shown in blue dashed boxes are invariant. They always appear at this position in introns. The sequences on the exon side of the splice sites are more variable.

Anne Niemick has β⁺-thalassemia (enough of the β-chain is produced to maintain blood hemoglobin levels above 6.0 g/dL). One mutation resulting in β⁺-thalassemia is a point mutation (AATAAA → AACAAA) that changes the sequence in hnRNA at the polyadenylation signal site from AAUAAA to AACAAA. Homozygous individuals with this mutation produce only one-tenth the amount of normal β-globin mRNA.

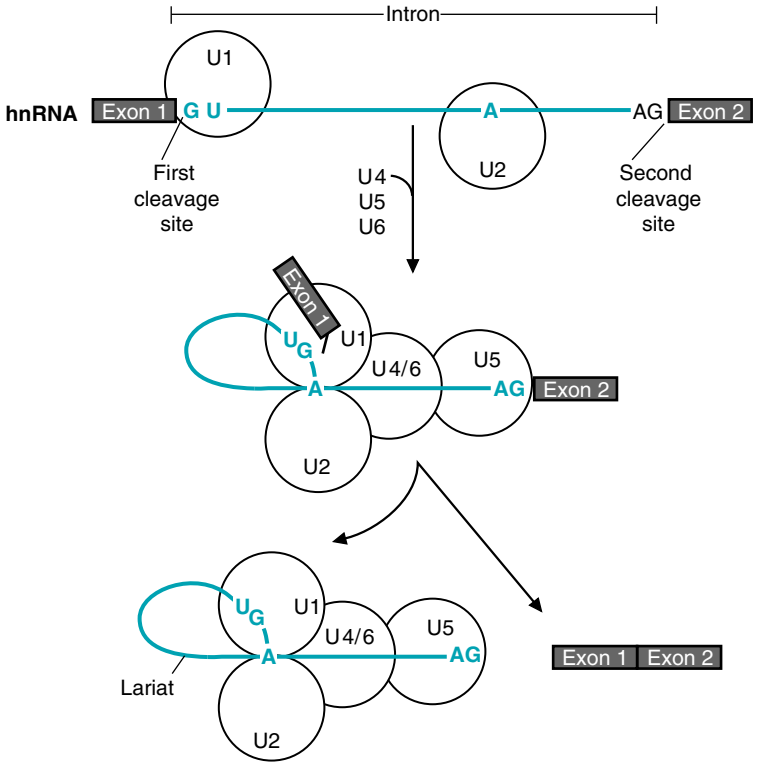


Fig. 14.14. Splicing process. Nuclear ribonucleoproteins (snurps U_1 to U_6) bind to the intron, causing it to form a loop. The complex is called a splicesosome. The U1 snurp binds near the first exon/intron junction, and U2 binds within the intron in a region containing an adenine nucleotide residue. Another group of snurps, U4, U5, and U6, binds to the complex, and the loop is formed. The phosphate attached to the G residue at the 5′-end of the intron forms a 2′–5′ linkage with the 2′-hydroxyl group of the adenine nucleotide residue. Cleavage occurs at the end of the first exon, between the AG residues at the 3’ end of the exon and the GU residues at the 5′ end of the intron. The complex continues to be held in place by the spliceosome. A second cleavage occurs at the 3′-end of the intron after the AG sequence. The exons are joined together. The intron, shaped like a lariat, is released and degraded to nucleotides.

Some types of β^0-thalassemia (little or none of the hemoglobin β chain produced) are caused by homozygous mutations in the splice junction sequences at intron/exon boundaries. In some individuals, an AT replaces a GT in the gene at the 5′ end of the first or second intron. Mutations also occur within the splice junction sequences at the 3′-end of introns (GT at the donor site 5′-end and AG at the acceptor site 3′-end). Mutations at either site totally abolish normal splicing and result in β^0 thalassemia.

(snRNPs), called "snurps," are involved in formation of the spliceosome. Because snurps are rich in uracil, they are identified by numbers preceded by a U.

Exons frequently code for separate functional or structural domains of proteins. Proteins with similar functional regions (e.g., ATP or NAD binding regions) frequently have similar domains, although their overall structure and amino acid sequence is quite different. A process known as exon shuffling has probably occurred throughout evolution, allowing new proteins to develop with functions similar to those of other proteins.

B. Synthesis of Eukaryotic rRNA

Ribosomal RNAs (rRNAs) form the ribonucleoprotein complexes on which protein synthesis occurs. In eukaryotes, the rRNA gene exists as many copies in the nucleolar organizer region of the nucleus (Fig. 14.15, circle 1). Each gene produces a large, 45S transcript that is cleaved to produce the 18S, 28S, and 5.8S rRNAs. Approximately 1,000 copies of this gene are present in the human genome. The genes are linked in tan-

Systemic lupus erythematous is an autoimmune disease characterized by a particular spectrum of autoantibodies against many cellular components, including chromatin, ribonucleoprotein, and cell membrane phospholipids. In this disorder, the body makes these antibodies against its own components. snRNPs are one of the targets of these antibodies. In fact, snRNPs were discovered as a result of studies using antibodies obtained from patients with SLE.

Tests were performed on **Sis Lupus's** blood to detect elevated levels of antinuclear antibodies (including antibodies to DNA, antibodies to histone, antibodies to ribonucleoproteins, and antibodies to nuclear antigens). The tests were strongly positive and, in conjunction with her symptoms, led to a diagnosis of systemic lupus erythematosus (SLE).

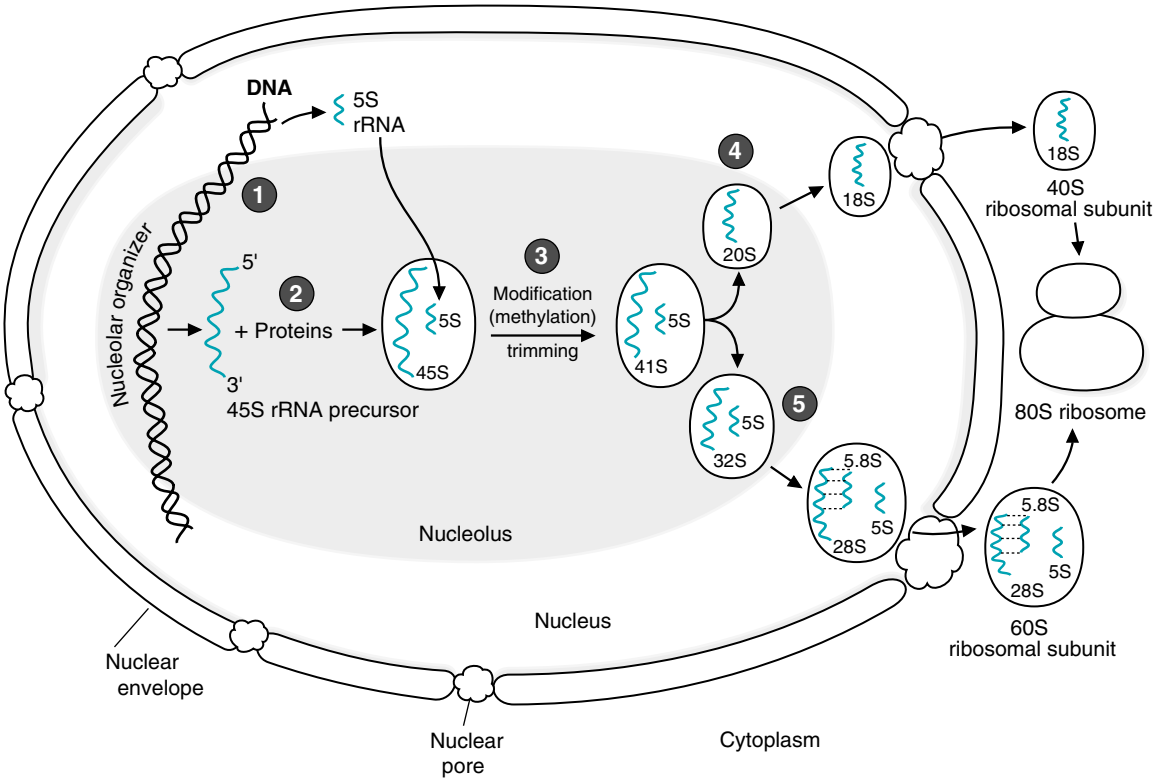


Fig. 14.15. rRNA and ribosome synthesis. The 5S rRNA is transcribed in the nucleoplasm and moves into the nucleolus. The other rRNAs are transcribed from DNA and mature in the nucleolus, forming the 40S and 60S ribosomal subunits, which migrate to the cytoplasm.

dem, separated by spacer regions that contain the termination signal for one gene and the promoter for the next. Promoters for rRNA genes are located in the 5′-flanking region of the genes and extend into the region surrounding the startpoint. rRNA genes caught in the act of transcription by electron micrographs show that many RNA polymerase I molecules can be attached to a gene at any given time, all moving toward the 3′ end as the 45S rRNA precursors are synthesized.

As the 45S rRNA precursors are released from the DNA, they complex with proteins, forming ribonucleoprotein particles that generate the granular regions of the nucleolus (see Fig. 14.15, circle 2). Processing of the transcript occurs in the granular regions. 5S rRNA, produced by RNA polymerase III from genes located outside the nucleolus in the nucleoplasm, migrates into the nucleolus and joins the ribonucleoprotein particles.

One to two percent of the nucleotides of the 45S precursor become methylated, primarily on the 2′-hydroxyl groups of ribose moieties (see Fig. 14.15, circle 3). These methyl groups may serve as markers for cleavage of the 45S precursors and are conserved in the mature rRNA. A series of cleavages in the 45S transcripts occur to produce the mature rRNAs (Fig. 14.16).

In the production of cytoplasmic ribosomes in human cells, one portion of the 45S rRNA precursor becomes the 18S rRNA that, complexed with proteins, forms the small 40S ribosomal subunit (Fig. 14.15, circle 4). Another segment of the precursor folds back on itself and is cleaved, forming 28S rRNA, hydrogen-bonded to the 5.8S rRNA. The 5S rRNA, transcribed from nonnucleolar genes, and a number of proteins complex with the 28S and 5.8S rRNAs to form the 60S ribosomal subunit (Fig. 14.15, circle 5). The ribosomal subunits migrate through the nuclear pores. In the cytoplasm, the 40S and 60S ribosomal subunits interact with mRNA, forming the 80S ribosomes on which protein synthesis occurs.

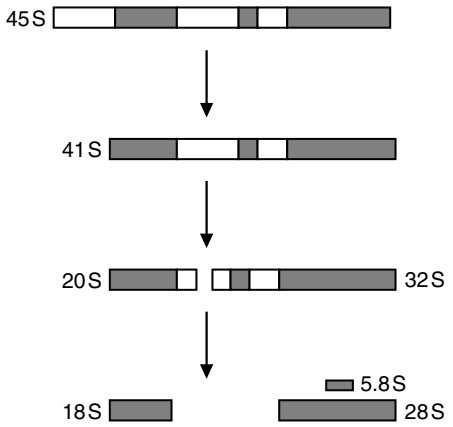


Fig. 14.16 Maturation of the 45S rRNA precursor. The clear regions are removed, and the shaded regions become the mature rRNAs. (The 5S rRNA is not produced from this precursor.)

Some rRNA precursors contain introns within the regions that become the mature rRNAs. These introns are removed by the rRNA precursors themselves, which undergo self-splicing reactions. The rRNA precursors that catalyze their own splicing are known as ribozymes (see Chapter 12, III.A).

C. Synthesis of Eukaryotic tRNA

A transfer RNA has one binding site for a specific sequence of three nucleotides in mRNA (the anticodon site) and another binding site for the encoded amino acid. tRNAs thus ensure that the genetic code is translated into the correct sequence of

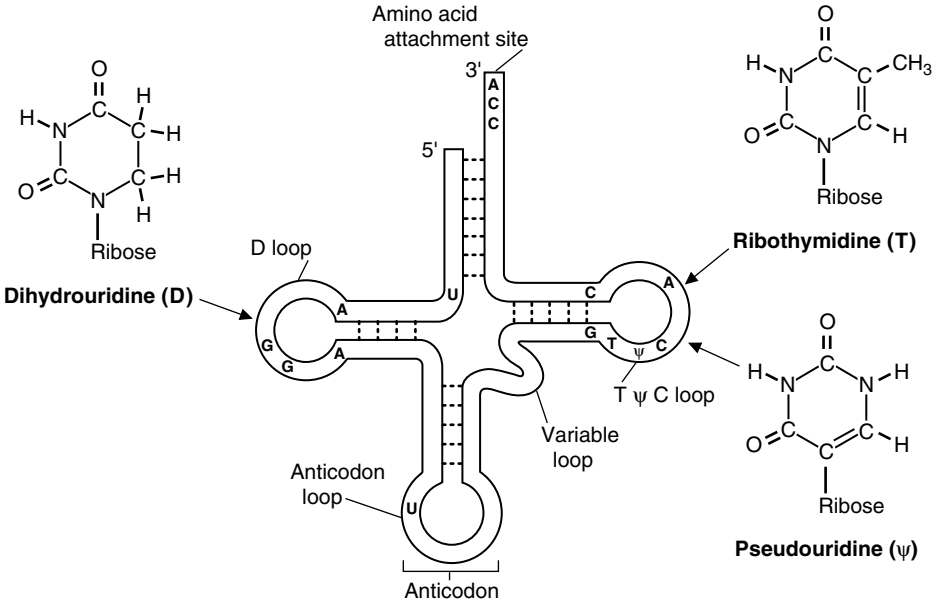


Fig. 14.17. The tRNA cloverleaf. Bases that commonly occur in a particular position are indicated by letters. Base-pairing in stem regions is indicated by lines between the strands. The locations of the modified bases dihydrouridine (D), ribothymidine (T), and pseudouridine (Ψ) are indicated.

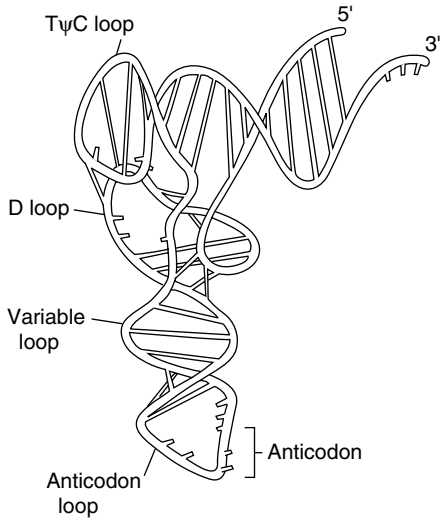


Fig. 14.18. The three-dimensional folding of tRNA. (Reprinted with permission from Kim SH, et al. Science 1974;185:436. Copyright 1974 American Association for the Advancement of Science.)

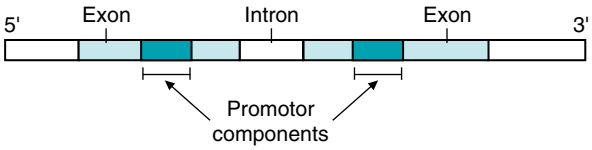


Fig. 14.19. Promoter for tRNA transcription. The segments of the genes from which the mature tRNA is produced are indicated by the light blue color. The two regions of the promoter lie within these segments, and are indicated by the dark blue color.

amino acids. At least 20 types of tRNAs occur in cells, one for every amino acid that is incorporated into growing polypeptide chains during the synthesis of proteins. tRNAs have a cloverleaf structure that folds into a three-dimensional L shape, and contains a number of bases that are modified posttranscriptionally (Fig. 14.17). The loop closest to the 5′-end is known as the D-loop because it contains dihydrouridine (D). The second, or anticodon, loop contains the trinucleotide anticodon that base-pairs with the codon on mRNA. The third loop (the TΨC loop) contains both ribothymidine (T) and pseudouridine (Ψ). A fourth loop, known as the variable loop because it varies in size, is frequently found between the anticodon and TΨC loops. Base-pairing occurs in the stem regions of tRNA, and a three-nucleotide sequence (e.g., CCA) at the 3′-end is the attachment site for the specific amino acid carried by each tRNA. Different tRNAs bind different amino acids. The three-dimensional structure of tRNA has been determined and is shown in Figure 14.18. tRNA is produced by RNA polymerase III, which recognizes a split promoter within the transcribed region of the gene (Fig. 14.19). One segment of the promoter is located between +8 and +19. A second segment is 30 to 60 base pairs downstream from the first.

tRNA precursors of approximately 100 nucleotides in length are generated. (Fig. 14.20, circle 1). The pre-tRNA assumes a cloverleaf shape and is subsequently cleaved at the 5′- and 3′-ends (see Fig. 14.20, circle 2). The enzyme that acts at the

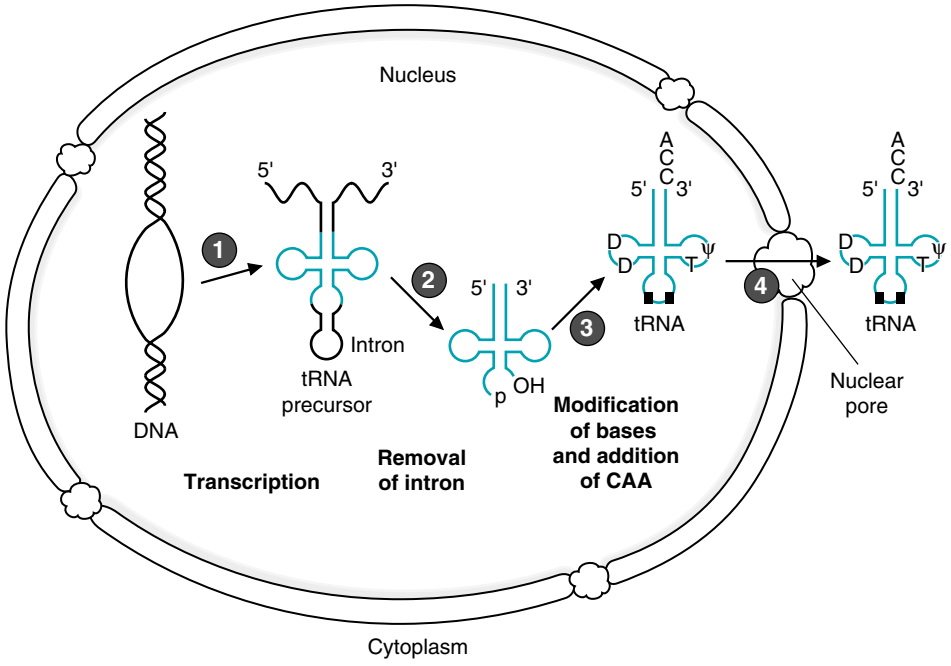


Fig. 14.20. Overview of tRNA synthesis. D, T, Ψ, and ■ indicate modified bases.
D = dihydrouracil, T = ribothymine, ψ = pseudouridine, ■ = other modified bases

5′-end is RNase P, similar to the RNase P of bacteria. Both enzymes contain a small RNA (M1) that has catalytic activity and serves as an endonuclease. Some tRNA precursors contain introns that are removed by endonucleases. To close the opening, a 2′- or 3′-phosphate group from one end is ligated to a 5′-hydroxyl on the other end by an RNA ligase.

The bases are modified at the same time the endonucleolytic cleavage reactions are occurring (see Fig. 14.20, circle 3). Three modifications occur in most tRNAs: (1) Uracil is methylated by *S*-adenosylmethionine (SAM) to form thymine; (2) one of the double bonds of uracil is reduced to form dihydrouracil;, and (3) a uracil residue (attached to ribose by an *N*-glycosidic bond) is rotated to form pseudouridine, which contains uracil linked to ribose by a carbon–carbon bond. (see Fig. 14.17). Other, less common but more complex, modifications also occur and involve bases other than uracil. Of particular note is the deamination of adenosine to form the base inosine.

The final step in forming the mature tRNA is the addition of a CCA sequence at its 3′-end (see Fig 14.20, circle 4). These nucleotides are added one at a time by nucleotidyltransferase. The tRNA then migrates to the cytoplasm. The terminal adenosine at the 3′-end is the site at which the specific amino acid for each tRNA is bound and activated for incorporation into a protein.

V. DIFFERENCES IN SIZE BETWEEN EUKARYOTIC AND PROKARYOTIC DNA

A. Human Cells Are Diploid

Except for the germ cells, most normal human cells are diploid. Therefore, they contain two copies of each chromosome, and each chromosome contains genes that are alleles of the genes on the homologous chromosome. Because one chromosome in each set of homologous chromosomes is obtained from each parent, the alleles can be identical, containing the same DNA sequence, or they can differ. A diploid human cell contains 2,000 times more DNA than the genome of the bacterium in the haploid *E. coli* cell (approximately 4×10^6 base pairs).

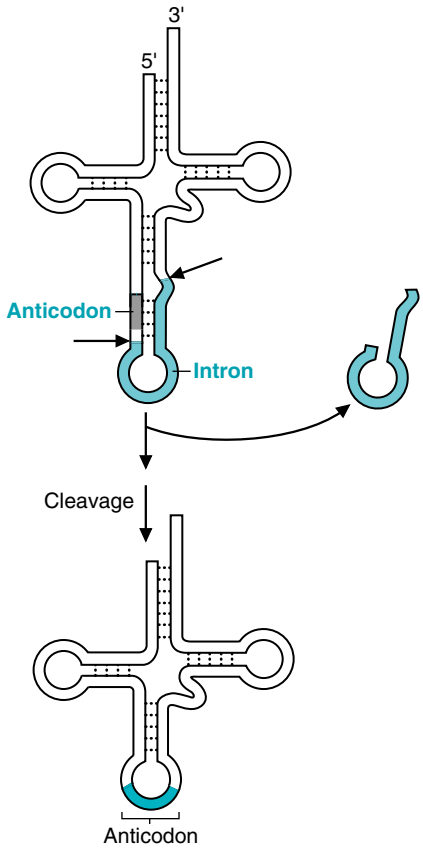
B. Human Genes Contain Introns

Eukaryotic introns contribute to the DNA size difference between bacteria and human cells. In eukaryotic genes, introns (noncoding regions) occur within sequences that code for proteins. Consequently, the primary transcript (heterogeneous nuclear RNA or hnRNA) averages roughly 10 times longer than the mature mRNA produced by removal of the introns. In contrast, bacterial genes do not contain introns.

C. Repetitive Sequences in Eukaryotic DNA

Although diploidy and introns account for some of the difference between the DNA content of humans and bacteria, a large difference remains that is related to the greater complexity of the human organism. Bacterial cells have a single copy of each gene, called unique DNA, and they contain very little DNA that does not produce functional products. Eukaryotic cells contain substantial amounts of DNA that does not code for functional products (i.e., proteins or rRNA and tRNA). In addition, some genes that encode functional products are present in multiple copies,

The removal of introns in pre-tRNA has been most extensively studied in yeast. The intron, less than 20 nucleotides long, is located within the anticodon loop at the 3′-end. Removal of the intron by endonucleases leaves the two halves of the tRNA, which remain held together by hydrogen bonds between base pairs in the stem regions. The opening is closed by an RNA ligase.



Q: Calculate the number of different proteins, 300 amino acids in length, which could be produced from the *E. coli* genome (4×10^6 base pairs of DNA).

A number of differences between eukaryotes and prokaryotes affect the processes of replication, transcription, and translation, in addition to the content of their DNA. Eukaryotic DNA is complexed with histones, and prokaryotic DNA is not. In eukaryotic cells, the process of transcription, which occurs in the nucleus, is separated by the nuclear envelope from the process of translation (protein synthesis from the mRNA template), which occurs in the cytoplasm. Because prokaryotes lack nuclei, the processes of transcription and translation occur simultaneously. Transcription of bacterial DNA requires only one promoter per operon. In contrast, human DNA requires one promoter for each gene.

Complexity may explain some of the differences between the DNA content of bacteria and humans. But an extension of this line of reasoning would lead to the conclusion that frogs are more complex than humans, because frogs have 8 feet of DNA per diploid nucleus compared to the 6 feet in a human cell. Logic, or perhaps vanity, suggests that the amount of DNA per cell does not necessarily reflect the complexity of the organism. One of the features of frog DNA that may explain its length is that frogs have more repetitive DNA than humans. More than 75% of the frog genome is in the moderately and highly repetitive category, whereas only about 35% of the human genome is repetitive.

A: Four million base pairs contain $4 \times 10^6/3$ or 1.33 million codons. If each protein contained approximately 300 amino acids, *E. coli* could produce about 4,000 different proteins ($1.33 \times 10^6/300$).

called highly repetitive or moderately repetitive DNA. Approximately 64% of the DNA in the human genome is unique, consisting of DNA sequences present in one or a very few copies in the genome (Fig. 14.21). Unique DNA sequences are transcribed to generate mRNA, which is translated to produce proteins.

Highly repetitive DNA consists of sequences approximately 6 to- 100 base pairs in length that are present in hundreds of thousands to millions of copies, clustered within a few locations in the genome (see Fig. 14.21). It occurs in centromeres (which join sister chromatids during mitosis) and in telomeres (the ends of chromosomes). This DNA represents approximately 10% of the human genome. It is not transcribed.

Moderately repetitive DNA is present in a few to tens of thousands of copies in the genome (see Fig. 14.21). This fraction constitutes approximately 25% of the human genome. It contains DNA that is functional and transcribed to produce rRNA, tRNA, and also some mRNA. The histone genes, present in a few hundred copies in the genome, belong to this class. Moderately repetitive DNA also includes some gene sequences that are functional but not transcribed. Promoters and enhancers (which are involved in regulating gene expression) are examples of gene sequences in this category. Other groups of moderately repetitive gene sequences that have been found in the human are called the Alu sequences (approximately 300 base pairs in length). Alu

Alu sequences in DNA were named for the enzyme Alu (obtained from **A**rthrobacter **lu**teus), that which is able to cleave them. Alu sequences make up 6 to 8% of the human genome. In some cases of familial hypercholesterolemia, homologous recombination is believed to have occurred between two Alu repeats, resulting in a large deletion in the low- density lipoprotein (LDL) receptor gene. The LDL receptor mediates uptake of the cholesterol-containing LDL particle into many cell types and, in the absence of functional LDL receptors, blood cholesterol levels are elevated. Patients who are homozygous for this mutation may die from of cardiac disease as early as in their second or third decade of life.

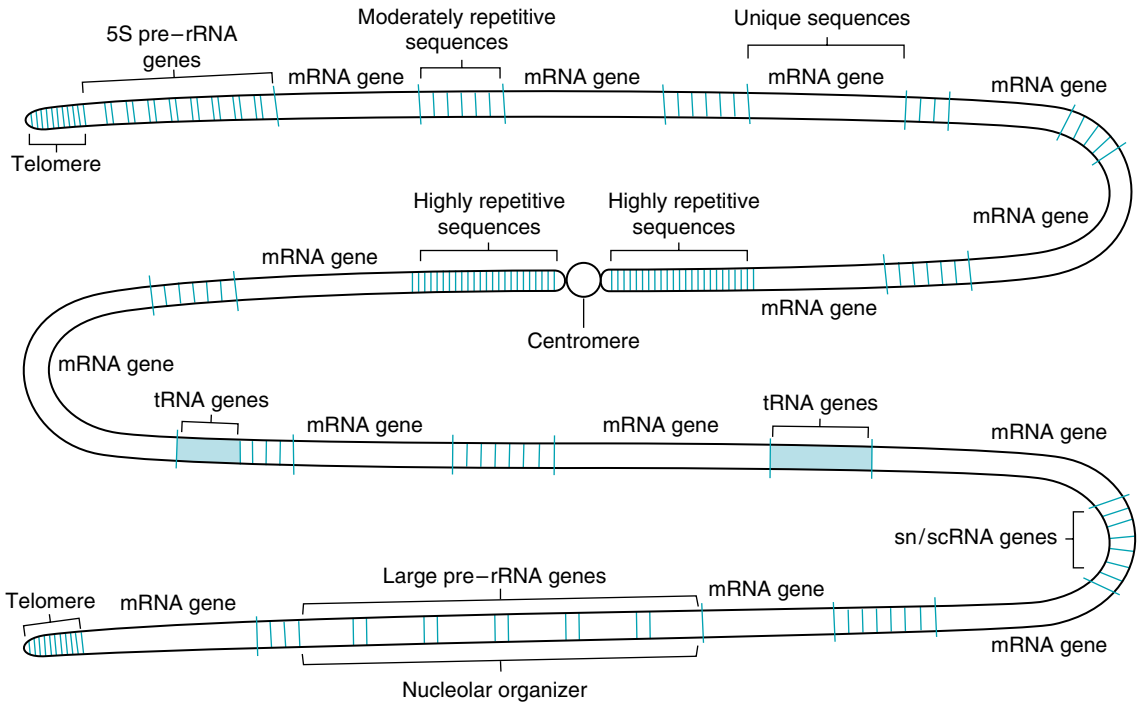


Fig. 14.21. Distribution of unique, moderately repetitive, and highly repetitive sequences in a hypothetical human chromosome. Unique genes encode mRNA. These genes occur in single copies. The genes for the large rRNA and the tRNA precursors occur in multiple copies that are clustered in the genome. The large rRNA genes form the nucleolar organizer. Moderately repetitive sequences are dispersed throughout the genome, and highly repetitive sequences are clustered around the centromere and at the ends of the chromosome (the telomeres). From Wolfe SL. Mol Cell Biol 1993:761.

Table 14.2. Differences between Eukaryotes and Prokaryotes

	Eukaryotes (human)	Prokaryotes (E. coli)
Nucleus	Yes	No
Chromosomes		
Number	23 per haploid cell	1 per haploid cell
DNA	Linear	Circular
Histones	Yes	No
Genome		
Diploid	Somatic cells	No
Haploid	Germ cells	All cells
Size	3×10^9 base pairs per haploid cell	4×10^6 base pairs
Genes		
Unique	64%	100%
Repetitive		
Moderately	25%	None
Highly	10%	None
Operons	No	Yes
mRNA		
Polycistronic	No	Yes
Introns (hnRNA)	Yes	No
Translation	Separate from transcription	Coupled with transcription

sequences are also examples of SINEs (**S**hort **IN**terspersed **E**lements). The LINE sequences (**L**ong **IN**terspersed **E**lements) are 6,000 to 7,000 base pairs in length. The function of the Alu and LINE sequences has not been determined.

Major differences between prokaryotic and eukaryotic DNA and RNA are summarized in Table 14.2.

LINE (**L**ong **IN**terspersed **E**lements) make up about 5% of the human genome. In some patients with hemophilia (a disease in which blood does not clot normally), a LINE sequence has been inserted into exon 14 of the gene for Factor VIII, a protein of the blood-clotting system. The insertion of the LINE sequence leads to the production of a nonfunctional protein.

CLINICAL COMMENTS

Anne Niemick. Patients with β^+-thalassemia who maintain their hemoglobin levels above 6.0 to 7.0 g/dL are usually classified as having thalassemia intermedia. In the β-thalassemias, the α chains of adult hemoglobin A ($\alpha_2\beta_2$) continue to be synthesized at a normal rate. These chains accumulate in the bone marrow, where the red blood cells are synthesized in erythropoiesis (generation of red blood cells). The accumulation of α chains diminishes erythopoiesis, resulting in an anemia. Individuals who are homozygous for a severe mutation require constant transfusions.

Individuals with thalassemia intermedia, such as **Anne Niemick,** could have inherited two different defective alleles, one from each parent. One parent may be a "silent" carrier, with one normal allele and one mildly affected allele. This parent produces enough functional β-globin, so no clinical symptoms of thalassemia appear. (However, they generally have a somewhat decreased amount of hemoglobin, resulting in microcytic hypochromic red blood cells.) When this parent contributes the mildly defective allele and the other heterozygous parent contributes a more severely defective allele,

The mutations that cause the thalassemias affect the synthesis of either the α- or the β-chains of adult hemoglobin, causing an anemia. They are classified by the chain affected (α- or β-) and by the amount of chain synthesized (0 for no synthesis and + for synthesis of some functional chains). They are also classified as major, intermediate, or minor, according to the severity of the clinical disorder. β-thalassemia major (also called homozygous β-thalassemia) is a clinically severe disorder requiring frequent blood transfusions. It is caused by the inheritance of two alleles for a severe mutation. In β-thalassemia intermedia, the patient exhibits a less severe clinical phenotype and is able to maintain hemoglobin levels above 6 g/dL. It is usually the result of two different mild mutations or homozygosity for a mild mutation. β-thalassemia minor (also known as β-thalassemia trait) is a heterozygous disorder involving a single mutation that is often clinically asymptomatic.

During embryonic and fetal life, the β-chain is replaced by the ϵ and γ chains. As a result, patients with severe mutations in the α-chain tend to die in utero, whereas those with mutations in the β chains exhibit symptoms postnatally, as hemoglobin F is normally replaced with adult hemoglobin A.

thalassemia intermedia occurs in the child. The child is thus heterozygous for two different defective alleles.

Ivy Sharer. Ivy Sharer was treated with a multidrug regimen for tuberculosis because the microbes that cause the disease frequently become resistant to the individual drugs. Rifampin in combination with the drug isoniazid (which affects metabolism of vitamin B_6 in the pathogenic bacteria) is usually effective, but months of treatment are required.

Just as bacteria can become resistant to drugs, so can HIV. A great concern to physicians treating patients with AIDS is the appearance of resistant strains of HIV-1 in patients taking a single drug, such as ZDV (zidovudine), for 6 months or more. Therefore, multidrug regimens are used that include other nucleoside reverse transcriptase inhibitors, such as dideoxyinosine (didanosine, formerly called ddI) and dideoxycytidine (zalcitabine, formerly called ddC). The multidrug therapy often includes nonnucleoside inhibitors (e.g., efavirenz), which acts allosterically on reverse transcriptase, and protease inhibitors (e.g., indinavir), which prevent the HIV polyprotein from being cleaved into its mature products (see Biochemical Comments).

Recent studies have indicated that a failure to properly dispose of cellular debris, a normal byproduct of cell death, may lead to the induction of autoantibodies directed against chromatin in patients with SLE. Normal cells have a finite lifetime, and are programmed to die (apoptosis) through a distinct biochemical mechanism. One of the steps in this mechanism is the stepwise degradation of cellular DNA (and other cellular components). If the normal intracellular components are exposed to the immune system, auto-antibodies against them may be generated. The enzyme in cells that degrades DNA is deoxyribonuclease I (DNase I), and individuals with SLE have reduced serum activity levels of DNase I compared with individuals who do not have the disease. Through an understanding of the molecular mechanism whereby autoantibodies are generated, it may be possible to develop therapies to combat this disorder.

Amanda Tin. The toxin α-amanitin is capable of causing irreversible hepatocellular and renal dysfunction through inhibition of mammalian RNA polymerases. Fortunately, **Amanda Tin's** toxicity proved mild. She developed only gastrointestinal symptoms and slight changes in her hepatic and renal function, which returned to normal within a few weeks. Treatment was primarily supportive, with fluid and electrolyte replacement for that lost through the gastrointestinal tract. No effective antidote is available for the *A. phalloides* toxin.

Sis Lupus. SLE is a multisystem disease of unknown origin characterized by inflammation related to the presence of autoantibodies in the blood. These autoantibodies react with antigens normally found in the nucleus, cytoplasm, and plasma membrane of the cell. Such "self" antigen–antibody (autoimmune) interactions initiate an inflammatory cascade that produces the broad symptom profile of multiorgan dysfunction found in **Sis Lupus.**

Pharmacological therapy for SLE, directed at immunosuppression, includes high doses of corticosteroids, which suppress the immune response. In refractory cases that do not respond to corticosteroids, cytotoxic drugs inhibiting the synthesis of antibody-producing cells by the bone marrow are also used to cause immunosuppression.

BIOCHEMICAL COMMENTS

Production of the Virus That Causes AIDS. AIDS is caused by the human immuno-deficiency virus (HIV). Two forms of the virus have been discovered, HIV-1, which is prevalent in industrialized countries, and HIV-2, which is prevalent in certain regions of Africa. Eight to ten years or more can elapse between the initial infection and development of the full-blown syndrome.

Proteins in the viral coat bind to membrane receptors (named CD4) of helper T lymphocytes, a class of cells involved in the immune response. Subsequently, conformational changes occur that allow the viral coat proteins to bind to a chemokine coreceptor in the cell membrane. The lipid in the viral coat then fuses with the cell membrane, and the viral core enters the cell, releasing its RNA and enzymes (including the reverse transcriptase) by a process called "uncoating." Reverse transcriptase uses the viral RNA as a template to produce a single-stranded DNA copy, which then serves as a template for synthesis of a double-stranded DNA. An integrase enzyme, also carried by the virus, enables this DNA to integrate into the host cell genome as a provirus (Fig. 14.22).

The reverse transcriptase that copies the viral genome has a high error rate because it does not have proofreading capability. Because incorporation of mismatched bases leads to evolution of the virus, new populations develop rapidly in response to changes in the environment. Mutations in the viral reverse transcriptase gene cause the enzyme to become resistant to drugs such as ZDV. Current vaccines developed experimentally against the gp120 surface protein may not be effective because this protein also mutates. As a consequence of the rapid mutation rate of HIV, current treatments for AIDS have limited effectiveness, and a cure has proved elusive.

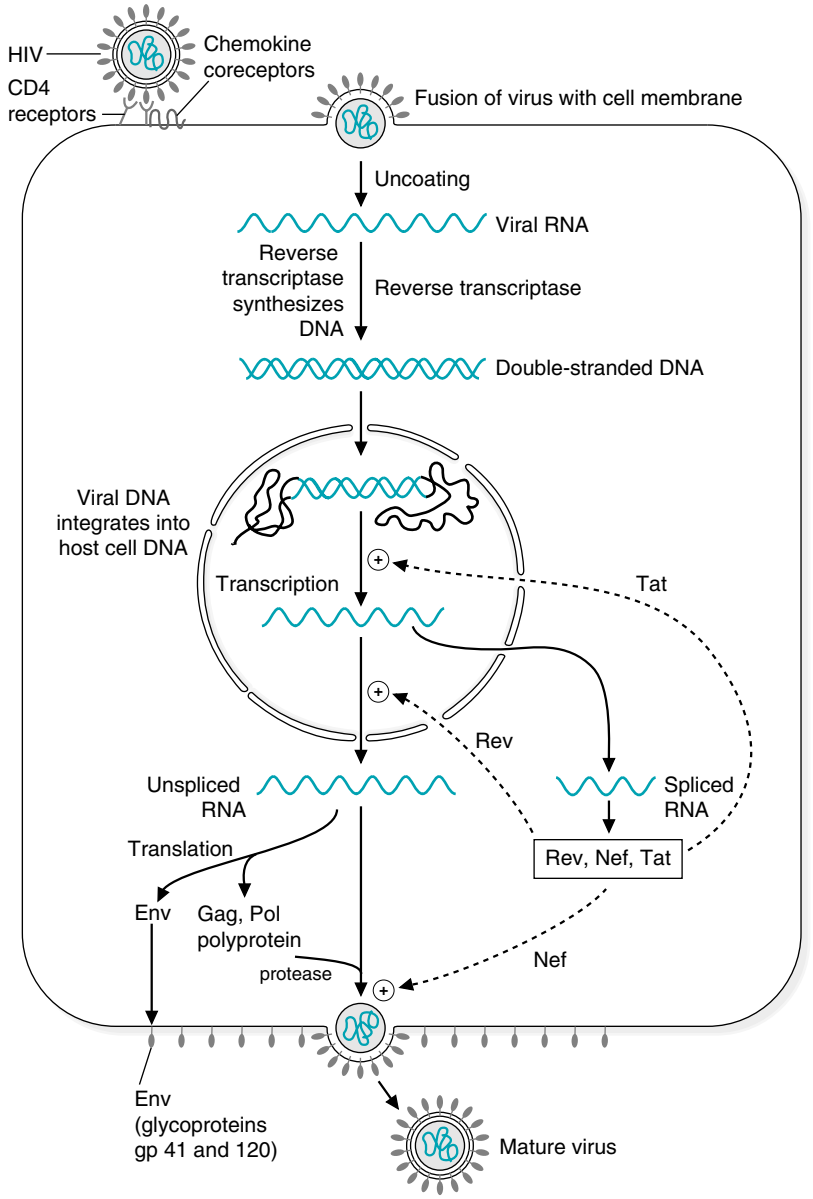


Fig. 14.22. Infection of a host cell by HIV. The HIV virus particle binds to the CD4 receptor and a chemokine coreceptor in the host cell membrane. The virus enters the cell and uncoats, releasing its RNA and proteins. The viral enzyme reverse transcriptase produces a double-stranded DNA copy that is integrated into the host cell genome. HIV is now a provirus. Transcripts of the viral DNA are spliced and translated to produce the proteins Tat, Rev, and Nef. Tat stimulates transcription of the viral DNA, and Rev causes the viral RNA transcripts to leave the nucleus unspliced. The unspliced RNA serves as the viral genome and also codes for the proteins of the viral core and envelope. The envelope proteins (gp41 and gp120, which are derived from the env protein) enter the cell membrane. The viral core proteins are synthesized as a polyprotein, which is cleaved by a protease as the viral particles form and bud from the cell membrane. The particles carry membrane lipid as a coat that contains gp41 and gp120. Nef indirectly aids in the assembly of viral particles. Pol is the reverse transcriptase produced from the viral RNA. ⊕ = stimulates.

In the initial stage of transcription of the provirus, the transcript is spliced, and three proteins, Nef, Tat, and Rev, are produced. Tat stimulates transcription of the viral genes. As Rev accumulates, it allows unspliced viral RNA to leave the nucleus and to produce proteins of the viral envelope and viral core, including reverse transcriptase. Two of the envelope glycoproteins (gp41 and gp120, which are derived from the env gene product) form a complex that embeds in the cell membrane. The other proteins, which are translated as a polyprotein and cleaved by the viral protease, combine with the full-length viral RNA to form core viral particles, which bud from the cell membrane. Thus, the virus obtains its lipid coat from the host cell membrane, and the coat contains the viral proteins gp41 and gp120. These surface proteins of the virus bind to CD4 receptors on other human helper T lymphocytes, and the infection spreads.

In an uninfected person, helper T lymphocytes usually number approximately 1,000/mL. Infection with HIV causes the number of these cells to decrease, which results in a deficiency of the immune system. When the number of T lymphocytes drops below 200/mL, the disease is in an advanced stage, and opportunistic infections, such as tuberculosis, occur. Although macrophages and dendritic cells lack CD4 receptors, they can also become infected with HIV and can carry the virus to the central nervous system.

The most effective means of combating HIV infection involves the use of drugs that inhibit the viral reverse transcriptase or the viral protease. However, these drugs only hold the infection at bay; they do not effect a cure.

Drugs currently used to treat AIDS act on the viral reverse transcriptase or the protease (see Fig. 14.22). The nonnucleoside drugs (e.g., efavirenz) bind to reverse transcriptase and inhibit its action. The nucleoside analogs (e.g., ZDV) add to the 3′ end of the growing DNA transcript produced by reverse transcriptase and prevent further elongation. The protease inhibitors (e.g., indinavir) bind to the protease and prevent it from cleaving the polyprotein.

Suggested References

A more complete coverage of transcription can be found in:

Lewin B. Chapter 5: mRNA. In: Genes VII. Oxford: Oxford University Press, 2000:119–138.

Lewin B. Chapter 9: mRNA and Transcription. In: Genes VII. Oxford: Oxford University Press, 2000:232–272.

Lewin, B. Chapter 20: Initiation of transcription. In: Genes VII. Oxford: Oxford University Press, 2000:649–684.

Tjian R. Molecular machines that control genes. Sci Am 1995;272:54–61.

References covering diseases discussed in this chapter:

Clumeck N. Choosing the best initial therapy for HIV-1 infection. N Engl J Med 1999;341:1925–1926. [See also the accompanying articles.]

Frankel AD, Young JAT. HIV-1: Fifteen proteins and an RNA. Annu Rev Biochem 1998;67:1–25.

Weatherall DJ, Clegg JB, Higgs DR, Wood WG. The hemoglobinopathies. Vol. III. In: Scriver CR, Beaudet AL, Sly WS, Valle B., eds. The Metabolic and Molecular Bases of Inherited Disease. 8th Ed. New York: McGraw Hill, 2001: 4571–4636.

Walport MJ. Lupus, DNase and defective disposal of cellular debris. Nature Ggenetics 2000;25:135–136.

REVIEW QUESTIONS—CHAPTER 14

1. The short transcript AUCCGUACG would be derived from which of the following template DNA sequences? (Note all sequences are written from 5′ to 3′)

   (A) ATCCGTACG
   (B) CGTACGGAT
   (C) AUCCGUACG
   (D) TAGGCATGC
   (E) GCATGCCTA

2. Given that the LD_{50} (the dose at which 50% of the recipients die) of amanitin is 0.1 mg per kg body weight, and that the average mushroom contains 7 mg amanitin, how many mushrooms must be consumed by Amanda Tin (50 kg body weight) to be above the LD_{50}?

   (A) 1
   (B) 2
   (C) 3
   (D) 4
   (E) 5

3. Which of the following eukaryotic DNA control sequences does not need to be in a fixed location, and is most responsible for high rates of transcription of particular genes?

   (A) Promoter
   (B) Promoter-proximal element
   (C) Enhancer
   (D) Operator
   (E) Splice donor site

4. Which of the following is true of both eukaryotic and prokaryotic gene expression?

   (A) After transcription, a 3′ poly A tail and a 5′ cap are added to mRNA.
   (B) Translation of mRNA can begin before transcription is complete.
   (C) mRNA is synthesized in the 3′ to 5′ direction.
   (D) RNA polymerase binds at a promoter region upstream of the gene.
   (E) Mature mRNA is always precisely co-linear to the gene from which it was transcribed.

5. In a segment of a transcribed gene, the nontemplate strand of DNA has the following sequence:
   5′...AGCTCACTG...3′
   What will be the corresponding sequence in the RNA produced from this segment of the gene?

   (A) CAGUGAGCU
   (B) AGCUCACUG
   (C) CAGTGAGCT
   (D) UCGAGUGAC
   (E) GTCACTCGA