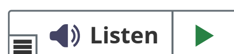
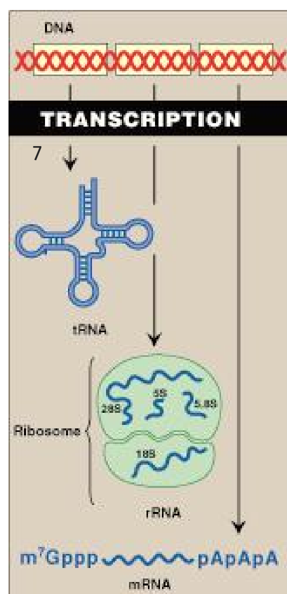
31: RNA Structure, Synthesis, and Processing

Overview



The genetic master plan of an organism is contained in the sequence of deoxyribonucleotides in its DNA. However, it is through ribonucleic acid (RNA), the "working copies" of DNA that the master plan is expressed (Fig. 31.1). The copying process, during which a DNA strand serves as a template for the synthesis of RNA, is called transcription. Transcription produces messenger RNA (mRNA), which is translated into sequences of amino acids (proteins), and ribosomal RNA (rRNA), transfer RNA (tRNA), and additional RNA molecules that perform specialized structural, catalytic, and regulatory functions and are not translated. That is, they are noncoding RNA (ncRNA). Therefore, the final product of gene expression can be RNA or protein, depending upon the gene. (Note: Only ~2% of the genome encodes proteins.) A central feature of transcription is that it is highly selective. For example, many transcripts are made of some regions of the DNA. In other regions, few or no transcripts are made. This selectivity is due, at least in part, to signals embedded in the nucleotide sequence of the DNA. These signals instruct the RNA polymerase (RNA pol) where to start, how often to start, and where to stop transcription. Several regulatory proteins are also involved in this selection process. The biochemical differentiation of an organism's tissues is ultimately a result of the selectivity of the transcription process. (Note: This selectivity of transcription is in contrast to the "all-or-none" nature of genomic replication.) Another important feature of transcription is that many RNA transcripts that initially are faithful copies of one of the two DNA strands may undergo various modifications, such as terminal additions, base modifications, trimming, and internal segment removal, which convert the inactive primary transcript into a functional molecule. The transcriptome is the complete set of RNA transcripts expressed by a genome.

FIGURE 31.1



ic information by transcription.

eukaryotic.) tRNA = transfer RNA; rRNA = ribosomal RNA; mRNA = messenger RNA; nosine-triphosphate cap; pApApA = poly-A tail; p = phosphate.

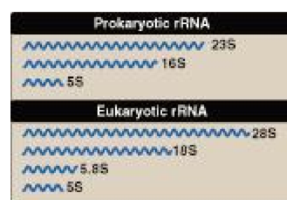
RNA Structure

There are three major types of RNA that participate in the process of protein synthesis: rRNA, tRNA, and mRNA. Like DNA, these RNA are unbranched polymeric molecules composed of nucleoside monophosphates joined together by 3′-to-5′ phosphodiester bonds (see p. 460). However, they differ from DNA in several ways. For example, they are considerably smaller than DNA, contain ribose instead of deoxyribose and uracil (U) instead of thymine (T), and exist as single strands that are capable of folding into complex structures. The three major types of RNA also differ from each other in size, function, and special structural modifications. (Note: In eukaryotes, additional small ncRNA molecules found in the nucleolus [small nucleolar RNA (snoRNA)], nucleus [small nuclear RNA (snRNA)], and cytoplasm [microRNA (miRNA)] perform specialized functions as described on pp. 490, 491 and 525.)

Ribosomal RNA

rRNA are found in association with several proteins as components of the ribosomes, the complex structures that serve as the sites for protein synthesis (see p. 500). Prokaryotic cells contain three distinct size species of rRNA (23S, 16S, and 5S, where S is the Svedberg unit for sedimentation rate that is determined by the size and shape of the particle), as shown in Figure 31.2. Eukaryotic cells contain four nuclear rRNA species (28S, 18S, 5.8S, and 5S) and two rRNA species (12S and 16S) encoded by the mitochondrial DNA. Together, rRNA make up ~80% of the total RNA in the cell. (Note: Some RNA function as catalysts, e.g., an rRNA in protein synthesis [see p. 504]. RNA with catalytic activity is termed a ribozyme.)

FIGURE 31.2

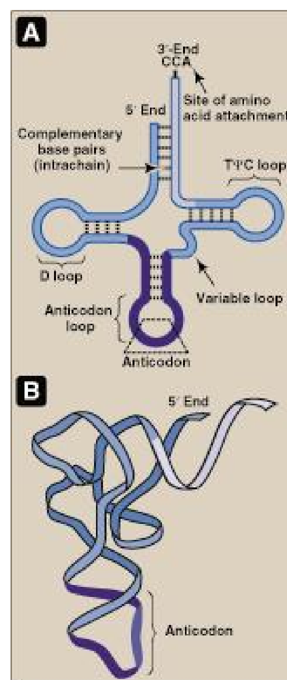


aryotic ribosomal RNA (rRNA).

Transfer RNA

tRNA are the smallest (4S) of the three major types of RNA molecules. There is at least one specific type of tRNA molecule for each of the 20 amino acids commonly found in proteins. Together, tRNA make up ~15% of the total RNA in the cell. The tRNA molecules contain a high percentage of unusual (modified) bases, for example, dihydrouracil (see Fig. 22.2, p. 325), and have extensive intrachain base pairing (Fig. 31.3) that leads to characteristic secondary cloverleaf structure and tertiary structure. Each tRNA serves as an adaptor molecule that carries its specific amino acid, covalently attached to its 3′ end, to the site of protein synthesis. There, it recognizes the genetic code sequence on an mRNA, which specifies the addition of that amino acid to the growing peptide chain (see p. 496). In eukaryotic cells, tRNA are encoded within both the nuclear and mitochondrial chromosomes.

FIGURE 31.3



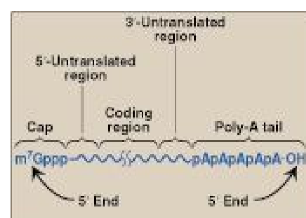
fer RNA (tRNA) secondary structure (cloverleaf). **B:** Folded (tertiary) tRNA structure found
acil; Ψ = pseudouracil; T = thymine; C = cytosine; A = adenine.

> The human mitochondrial chromosome carries 22 tRNA genes. Mutations in these genes can cause human disease. Mutations in the mitochondrial gene for $tRNA^{Lys}$ are associated with myoclonic epilepsy (jerking muscle spasms) with ragged red fibers (MERRF), a disorder that affects skeletal muscle structure and function (myopathy), and with mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS), which affects the brain, nervous system, and muscles. MELAS is also caused by mutations in the mitochondrial $tRNA^{Leu}$ gene.

Messenger RNA

mRNA comprises only ~5% of the RNA in a cell, yet is by far the most heterogeneous type of RNA in size and base sequence. mRNA is coding RNA in that it carries genetic information from DNA for use in protein synthesis. In eukaryotes, this involves transport of mRNA out of the nucleus and into the cytosol. An mRNA carrying information from more than one gene is polycistronic (cistron = gene). Polycistronic mRNA is characteristic of prokaryotes, mitochondria, some viruses, and in chloroplast in plants. An mRNA carrying information from only one gene is monocistronic and is characteristic of eukaryotes. In addition to the protein-coding regions that can be translated, mRNA contains untranslated regions at its 5′- and 3′ ends (Fig. 31.4). Special structural characteristics of eukaryotic (but not prokaryotic) mRNA include a long sequence of adenine (A) nucleotides (a poly-A tail) on the 3′ end of the RNA, plus a cap on the 5′ end consisting of a molecule of 7-methylguanosine attached through an unusual (5′-to-5′) triphosphate linkage. The mechanisms for modifying mRNA to create these special structural characteristics are discussed on pp. 490 and 491.

FIGURE 31.4



tic messenger RNA.

e.

Prokaryotic Gene Transcription

The structure of magnesium-requiring RNA pol, the signals that control transcription, and the varieties of modification that RNA transcripts can undergo differ among organisms, particularly from prokaryotes to eukaryotes. Therefore, the discussions of prokaryotic and eukaryotic transcription are presented separately.

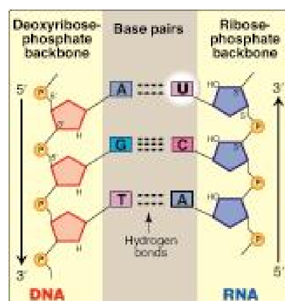
Prokaryotic RNA polymerase

In bacteria, one species of RNA pol synthesizes all of the RNA except for the short RNA primers needed for DNA replication (Note: RNA primers are synthesized by the specialized, monomeric enzyme primase [see p. 466].) RNA pol is a multisubunit enzyme that recognizes a nucleotide sequence (the promoter region) at the beginning of a length of DNA that is to be transcribed. It next makes a complementary RNA copy of the DNA template strand and then recognizes the end of the DNA sequence to be transcribed (the termination region). RNA is synthesized from its 5′ end to its 3′ end, antiparallel to its DNA template strand (see p. 463). The template is copied as it is in DNA synthesis, in which a guanine (G) on the DNA specifies a cytosine (C) in the RNA, a C specifies a G, a T specifies an A, but an A specifies a U instead of a T (Fig. 31.5). The RNA, then, is complementary to the DNA template (antisense, minus) strand and identical to the coding (sense, plus) strand, with U replacing T. Within the DNA molecule, regions of both strands can serve as templates for transcription. For a given gene, however, only one of the two DNA strands can be the template. Which strand is used is determined by the location of the promoter for that gene. Transcription by RNA pol involves a core enzyme and several auxiliary proteins.

Core enzyme

Five of the enzyme's peptide subunits, 2 α, 1 β, 1 β′, and 1 Ω, are required for enzyme assembly (α, Ω), template binding (β′), and the 5′→3′ polymerase activity (β) and together are referred to as the core enzyme (Fig. 31.6). However, this enzyme lacks specificity (i.e., it cannot recognize the promoter region on the DNA template).

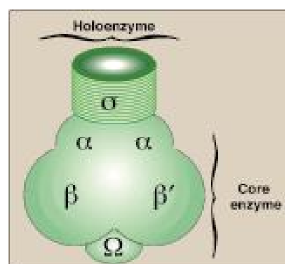
FIGURE 31.5



mentary base pairs between DNA and RNA.

ne; C = cytosine; G = guanine; U = uracil.

FIGURE 31.6



karyotic *RNA polymerase*.

Holoenzyme

The σ subunit (sigma factor) enables RNA pol to recognize promoter regions on the DNA. The σ subunit plus the core enzyme make up the holoenzyme. (Note: Different σ factors recognize different groups of genes, with σ^{70} predominating.)

Steps in RNA synthesis

The process of transcription of a typical gene of *Escherichia coli (E. coli)* can be divided into three phases: initiation, elongation, and termination. A transcription unit extends from the promoter to the termination region, and the initial product of transcription by RNA pol is termed the primary transcript.

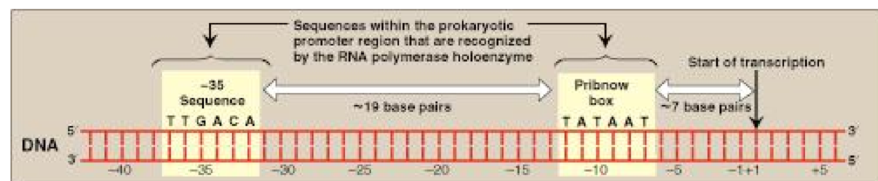
Initiation

Transcription begins with the binding of the RNA pol holoenzyme to a region of the DNA known as the promoter, which is not transcribed. The prokaryotic promoter contains characteristic consensus sequences (Fig. 31.7). (Note: Consensus sequences are idealized sequences in which the base shown at each position is the base most frequently [but not necessarily always] encountered at that position.) Those that are recognized by prokaryotic RNA pol σ factors include the following.

−35 Sequence

A consensus sequence (5′-TTGACA-3′), centered about 35 bases to the left of the transcription start site (see Fig. 31.7), is the initial point of contact for the holoenzyme, and a closed complex is formed. (Note: By convention, the regulatory sequences that control transcription are designated by the 5′→3′ nucleotide sequence on the coding strand. A base in the promoter region is assigned a negative number if it occurs prior to [to the left of, toward the 5′ end of, or “upstream” of] the transcription start site. Therefore, the TTGACA sequence is centered at approximately base −35. The first base at the transcription start site is assigned a position of +1. There is no base designated “0.”)

FIGURE 31.7



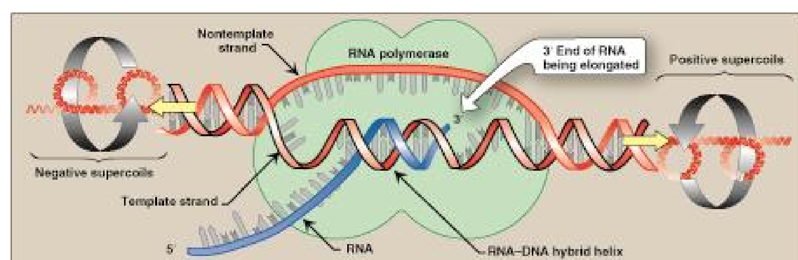
Pribnow box

The holoenzyme moves and covers a second consensus sequence (5′-TATAAT-3′), centered at about −10 (see Fig. 31.7), which is the site of melting (unwinding) of a short stretch (~14 base pairs) of DNA. This initial melting converts the closed initiation complex to an open complex known as a transcription bubble. (Note: A mutation in either the −10 or the −35 sequence can affect the transcription of the gene controlled by the mutant promoter.)

Elongation

Once the promoter has been recognized and bound by the holoenzyme, local unwinding of the DNA helix continues (Fig. 31.8), mediated by the polymerase. (Note: Unwinding generates supercoils in the DNA that can be relieved by DNA topoisomerases [see p. 465].) RNA pol begins to synthesize a transcript of the DNA sequence, and several short pieces of RNA are made and discarded. The elongation phase begins when the transcript (typically starting with a purine) exceeds 10 nucleotides in length. Sigma factor is then released, and the core enzyme is able to leave (clear) the promoter and move along the template strand in a processive manner, serving as its own sliding clamp. During transcription, a short DNA–RNA hybrid helix is formed (see Fig. 31.8). Like DNA pol, RNA pol uses nucleoside triphosphates as substrates and releases pyrophosphate each time a nucleoside monophosphate is added to the growing chain. As with replication, transcription is always in the 5′→3′ direction. In contrast to DNA pol, RNA pol does not require a primer and does not have a 3′→5′ exonuclease domain for proofreading. (Note: Misincorporation of a ribonucleotide causes RNA pol to pause, backtrack, cleave the transcript, and restart. Nonetheless, transcription has a higher error rate than does replication.)

FIGURE 31.8



on of an open initiation complex

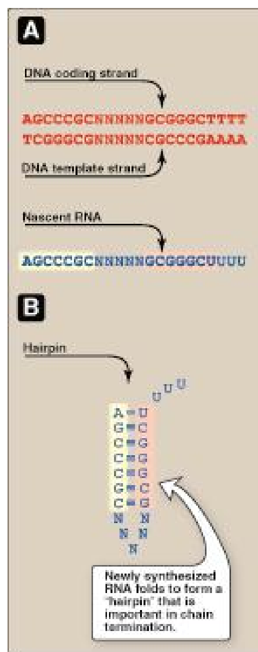
Termination

The elongation of the single-stranded RNA chain continues until a termination signal is reached. Termination can be intrinsic (occur without additional proteins) or dependent upon the participation of a protein known as the ρ (rho) factor.

Rho Independent

For most prokaryotic genes, this type of termination requires a sequence in the DNA template for generating a sequence in the nascent (newly made) RNA that is self-complementary (Fig. 31.9). This allows the RNA to fold back on itself, forming a GC-rich stem (stabilized by hydrogen bonds) plus a loop. This structure is known as a "hairpin." Additionally, just beyond the hairpin, the RNA transcript contains a string of uracil residues (Us) at the 3′ end. The bonding of these Us to the complementary As of the DNA template is weak. This facilitates the separation of the newly synthesized RNA from its DNA template, as the double helix "zips up" behind the RNA pol.

FIGURE 31.9



termination of prokaryotic transcription.

uence generates a self-complementary sequence in the nascent RNA. **B:** Hairpin structure N represents a noncomplementary base; A = adenine, T = thymine; G = guanine; C =

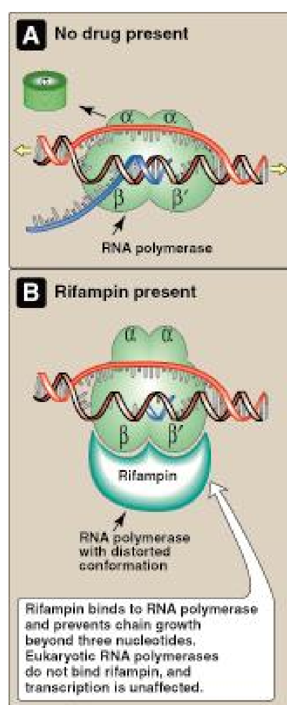
Rho Dependent

This requires the participation of the additional protein rho, which is a hexameric ATPase with helicase activity. Rho binds a C-rich rho utilization (rut) site near the 5′ end of the nascent RNA and, using its ATPase activity, moves along the RNA until it reaches the RNA pol paused at the termination site. The ATP-dependent helicase activity of rho separates the RNA–DNA hybrid helix, causing the release of the RNA.

Antibiotics

Some antibiotics prevent bacterial cell growth by inhibiting RNA synthesis. For example, rifampin (rifampicin) inhibits transcription initiation by binding to the β subunit of prokaryotic RNA pol and preventing chain growth beyond three nucleotides (Fig. 31.10). Rifampin is important in the treatment of tuberculosis. Dactinomycin (actinomycin D) was the first antibiotic to find therapeutic application in tumor chemotherapy. It inserts (intercalates) between the DNA bases and inhibits transcription initiation and elongation in tumor cells.

FIGURE 31.10



ot elongation by RNA polymerase with no drug present. **B:** Inhibition of prokaryotic RNA n (rifampicin).

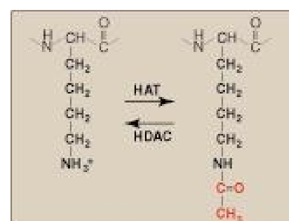
Eukaryotic Gene Transcription

The transcription of eukaryotic genes is a far more complicated process than transcription in prokaryotes. Eukaryotic transcription involves separate polymerases for the synthesis of rRNA, tRNA, and mRNA. In addition, a large number of proteins called transcription factors (TFs) are involved. TFs bind to distinct sites on the DNA within the core promoter region, close (proximal) to it, or some distance away (distal). They are required for both the assembly of a transcription initiation complex at the promoter and the determination of which genes are to be transcribed. (Note: Each eukaryotic RNA pol has its own promoters and TFs that bind core promoter sequences.) For TF to recognize and bind to their specific DNA sequences, the chromatin structure in that region must be decondensed (relaxed) to allow access to the DNA. The role of transcription in the regulation of gene expression is discussed in Chapter 33.

Chromatin structure and gene expression

The association of DNA with histones to form nucleosomes (see p. 473) affects the ability of the transcription machinery to access the DNA to be transcribed. Most actively transcribed genes are found in a relatively decondensed form of chromatin called euchromatin, whereas most inactive segments of DNA are found in highly condensed heterochromatin. The interconversion of these forms is called chromatin remodeling. A major component of chromatin remodeling is the covalent modification of histones (e.g., the acetylation of lysine residues at the amino terminus of histone proteins), as shown in Figure 31.11. Acetylation, mediated by histone acetyltransferases (HATs), eliminates the positive charge on the lysine, thereby decreasing the interaction of the histone with the negatively charged DNA. Removal of the acetyl group by histone deacetylases (HDACs) restores the positive charge and fosters stronger interactions between histones and DNA. (Note: The ATP-dependent repositioning of nucleosomes is also required to access DNA.)

FIGURE 31.11



...ylation of a lysine residue in a histone.

...vides the acetyl group. HAT = histone acetyltransferase; HDAC = histone deacetylase.

Nuclear RNA polymerases

There are three distinct types of RNA pol in the nucleus of eukaryotic cells. All are large enzymes with multiple subunits. Each type of RNA pol recognizes particular genes. (Note: Mitochondria contain a single RNA pol that resembles the bacterial enzyme in its function.)

RNA polymerase I

This enzyme synthesizes the precursor of the 28S, 18S, and 5.8S rRNA in the nucleolus.

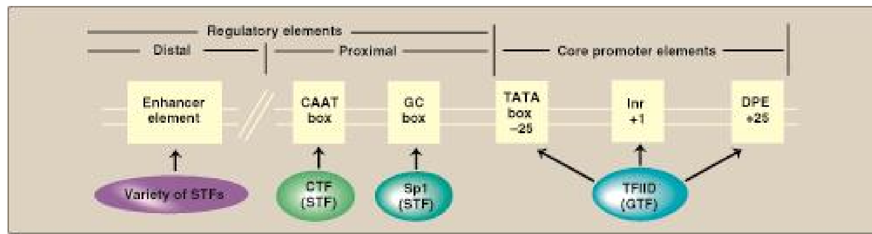
RNA polymerase II

This enzyme synthesizes the nuclear precursors of mRNA that are processed and then translated to proteins. RNA pol II also synthesizes certain small ncRNA, such as snoRNA, snRNA, and miRNA.

Promoters for RNA polymerase II

In some genes transcribed by RNA pol II, a sequence of nucleotides (TATAAA) that is nearly identical to that of the Pribnow box (see p. 485) is found centered ~25 nucleotides upstream of the transcription start site. This core promoter consensus sequence is called the TATA, or Hogness, box. In the majority of genes, however, no TATA box is present. Instead, different core promoter elements such as initiator (Inr) or downstream promoter element (DPE) are present (Fig. 31.12). (Note: No one consensus sequence is found in all core promoters.) Because these sequences are on the same molecule of DNA as the gene being transcribed, they are cis-acting. The sequences serve as binding sites for proteins known as general transcription factors (GTFs), which in turn interact with each other and with RNA pol II.

FIGURE 31.12

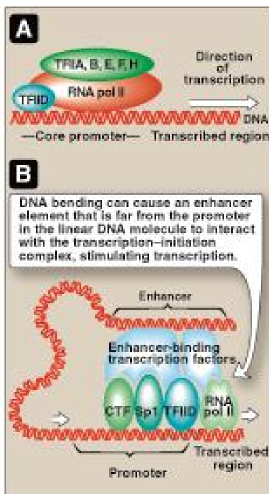


and their trans-acting general and

General transcription factors

GTFs are the minimal requirements for recognition of the promoter, recruitment of RNA pol II to the promoter, formation of the preinitiation complex, and initiation of transcription at a basal level (Fig. 31.13A). GTFs are encoded by different genes, synthesized in the cytosol, and diffuse (transit) to their sites of action, and so are trans-acting. (Note: In contrast to the prokaryotic holoenzyme, eukaryotic RNA pol II does not itself recognize and bind the promoter. Instead, transcription factor IID [TFIID], a GTF containing TATA-binding protein and TATA-associated factors, recognizes and binds the TATA box [and other core promoter elements]. TFIIF, another GTF, brings the polymerase to the promoter. The helicase activity of TFIIH melts the DNA, and its kinase activity phosphorylates polymerase, allowing it to clear the promoter.)

FIGURE 31.13



eneral transcription factors (TFII) and RNA polymerase II (RNA pol II) at the core oman numeral II denotes a TF for RNA pol II.) **B:** Enhancer stimulation of transcription. ption factor; Sp1 = specificity factor-1.

Regulatory elements and transcriptional activators

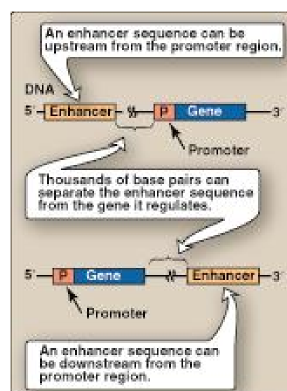
Additional consensus sequences lie upstream of the core promoter (see Fig. 31.12). Those close to the core promoter (within ~200 nucleotides) are the proximal regulatory elements, such as the CAAT and GC boxes. Those farther away are the distal regulatory elements such as enhancers (see d. below). Proteins known as transcriptional activators or specific transcription factors (STFs) bind these regulatory elements. STFs bind to promoter proximal elements to regulate the frequency of transcription initiation and to distal elements to mediate the response to signals such as hormones (see p. 522) and regulate which genes are expressed at a given point in time. A typical protein-coding eukaryotic gene has binding sites for many such factors. STFs have two binding domains. One is a DNA-binding domain, the other is a transcription activation domain that recruits the GTF to the core promoter as well as coactivator proteins such as the HAT enzymes involved in chromatin modification. (Note: The Mediator, a multisubunit coactivator of RNA pol II–catalyzed transcription, binds the polymerase, the GTF, and the STF and regulates transcription initiation.)

Transcriptional activators bind DNA through a variety of motifs, such as the helix-loop-helix, zinc finger, and leucine zipper (see p. 18).

Role of enhancers

Enhancers are special DNA sequences that increase the rate of initiation of transcription by RNA pol II. Enhancers are typically on the same chromosome as the gene whose transcription they stimulate (Fig. 31.13B). However, they can (1) be located upstream (to the 5′ side) or downstream (to the 3′ side) of the transcription start site, (2) be close to or thousands of base pairs away from the promoter (Fig. 31.14), and (3) occur on either strand of the DNA. Enhancers contain DNA sequences called response elements that bind STFs. By bending or looping the DNA, STFs can interact with other TFs bound to a promoter and with RNA pol II, thereby stimulating transcription (see Fig. 31.13B). The Mediator also binds enhancers. (Note: Although silencers are similar to enhancers in that they also can act over long distances, they reduce gene expression.)

FIGURE 31.14



tions of enhancer sequences.

RNA polymerase II inhibitor

α-Amanitin, a potent toxin produced by the poisonous mushroom *Amanita phalloides* (sometimes called the "death cap"), binds RNA pol II tightly and slows its movement, thereby inhibiting mRNA synthesis.

RNA polymerase III

This enzyme synthesizes tRNA, 5S rRNA, and some snRNA and snoRNA.

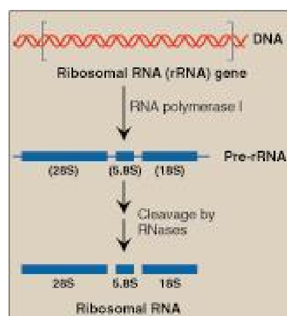
Posttranscriptional Modification of RNA

A primary transcript is the initial, linear, RNA copy of a transcription unit (the segment of DNA between specific initiation and termination sequences). The primary transcripts of both prokaryotic and eukaryotic tRNA and rRNA are posttranscriptionally modified by cleavage of the original transcripts by ribonucleases. tRNA are further modified to help give each species its unique identity. In contrast, prokaryotic mRNA is generally identical to its primary transcript, whereas eukaryotic mRNA is extensively modified both co- and posttranscriptionally.

Ribosomal RNA

rRNA of both prokaryotic and eukaryotic cells are generated from long precursor molecules called pre-rRNA. The 23S, 16S, and 5S rRNA of prokaryotes are produced from a single pre-rRNA molecule, as are the 28S, 18S, and 5.8S rRNA of eukaryotes (Fig. 31.15). (Note: Eukaryotic 5S rRNA is synthesized by RNA pol III and modified separately.) The pre-rRNA are cleaved by ribonucleases to yield intermediate-sized pieces of rRNA, which are further processed (trimmed by exonucleases and modified at some bases and riboses) to produce the required RNA species. (Note: In eukaryotes, rRNA genes are found in long, tandem arrays. rRNA synthesis and processing occur in the nucleolus, with base and sugar modifications facilitated by snoRNA.)

FIGURE 31.15

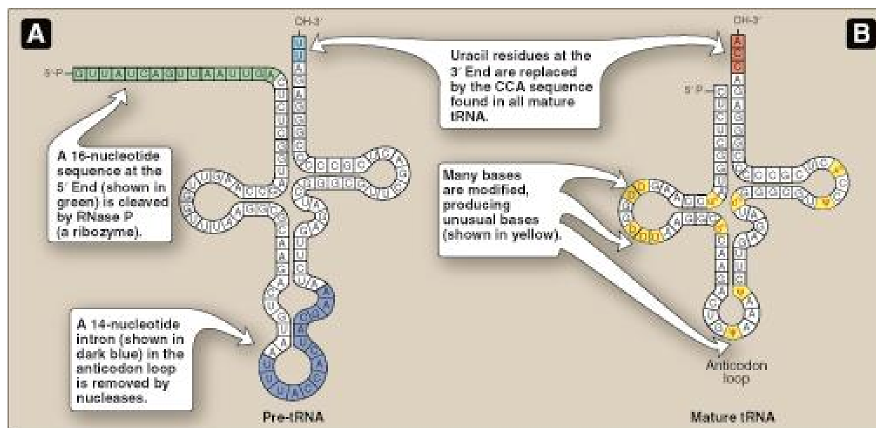


processing of eukaryotic ribosomal RNA by ribonucleases (RNases).

Transfer RNA

Both eukaryotic and prokaryotic tRNA are also made from longer precursor molecules that must be modified (Fig. 31.16). Sequences at both ends of the molecule are removed, and, if present, an intervening sequence intron is removed from the anticodon loop by nucleases. Other posttranscriptional modifications include addition of a –CCA sequence by nucleotidyltransferase to the 3′ terminal end of tRNA and modification of bases at specific positions to produce the unusual bases characteristic of tRNA (see p. 324).

FIGURE 31.16



tRNA after posttranscriptional

il), and ^{m}, which means that the base has

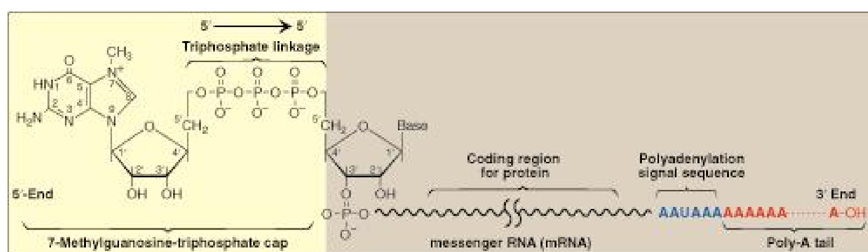
Eukaryotic messenger RNA

The collection of all the primary transcripts synthesized in the nucleus by RNA pol II is known as heterogeneous nuclear RNA (hnRNA). The pre-mRNA components of hnRNA undergo extensive co- and posttranscriptional modification in the nucleus and become mature mRNA. These modifications usually include the following. (Note: Pol II itself recruits the proteins required for the modifications.)

Addition of a 5′ cap

This is the first of the processing reactions for pre-mRNA (Fig. 31.17). The cap is a 7-methylguanosine attached to the 5′ terminal end of the mRNA through an unusual 5′-to-5′ triphosphate linkage that is resistant to most nucleases. Creation of the cap requires removal of the γ-phosphoryl group from the 5′ triphosphate of the pre-mRNA, followed by addition of guanosine monophosphate (from guanosine triphosphate) by the nuclear enzyme guanylyltransferase. Methylation of this terminal G occurs in the cytosol and is catalyzed by guanine-7-methyltransferase. S-Adenosylmethionine (SAM) is the source of the methyl group (see p. 292). Additional methylation steps may occur. The addition of this 7-methylguanosine cap helps stabilize the mRNA and permits efficient initiation of translation (see p. 504).

FIGURE 31.17



ylguanosine cap and polyadenylate

Addition of a 3′-poly-A tail

Most eukaryotic mRNA (with several exceptions, including those for the histones) have a chain of 40 to 250 adenylates (adenosine monophosphates) attached to the 3′ end (see Fig. 31.17). This poly-A tail is not transcribed from the DNA but rather is added by the nuclear enzyme, polyadenylate polymerase, using ATP as the substrate. The pre-mRNA is cleaved downstream of a consensus sequence, called the polyadenylation signal sequence (AAUAAA), found near the 3′ end of the RNA, and the poly-A tail is added to the new 3′ end. The poly-A tail terminates eukaryotic transcription. In addition, it helps stabilize the mRNA, facilitates its exit from the nucleus, and aids in translation. After the mRNA enters the cytosol, the poly-A tail is gradually shortened.

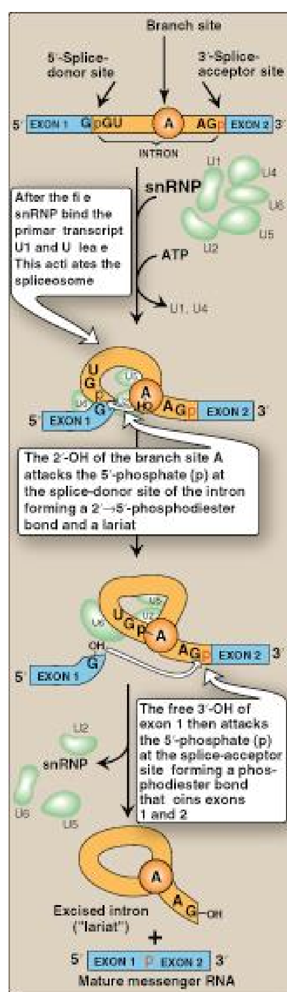
Splicing

Maturation of eukaryotic mRNA usually involves removal from the primary transcript of RNA sequences (introns or intervening sequences) that do not code for protein. The remaining coding (expressed) sequences, the exons, are joined together to form the mature mRNA. The process of removing introns and joining exons is called splicing. The molecular complex that accomplishes these tasks is known as the spliceosome. A few eukaryotic primary transcripts contain no introns (e.g., those from histone genes). Others contain a few introns, whereas some, such as the primary transcripts for the α-chains of collagen, contain >50 introns that must be removed.

Role of small nuclear RNA

In association with multiple proteins, U-rich snRNA form five small nuclear ribonucleoprotein particles (snRNP, or “snurp”) designated as U1, U2, U4, U5, and U6 that mediate splicing. They facilitate the removal of introns by forming base pairs with the consensus sequences at each end of the intron (Fig. 31.18). (Note: In systemic lupus erythematosus, an autoimmune disease, patients produce antibodies against their own nuclear proteins such as snRNP.)

FIGURE 31.18



donor site, and U2 binds the branch A and the 3′ acceptor site. Addition of U4–U6

x.) snRNP = small nuclear ribonucleoprotein particle.

Mechanism

The binding of snRNP brings the sequences of neighboring exons into the correct alignment for splicing, allowing two transesterification reactions (catalyzed by the RNA of U2, U5, and U6) to occur. The 2′-OH group of an A nucleotide (known as the branch site A) in the intron attacks the phosphate at the 5′ end of the intron (splice donor site), forming an unusual 2′→5′ phosphodiester bond and creating a "lariat" structure (see Fig. 31.18). The newly freed 3′-OH of exon 1 attacks the 5′ phosphate at the splice acceptor site, forming a phosphodiester bond that joins exons 1 and 2. The excised intron is released as a lariat, which is typically degraded but may be a precursor for ncRNA such as snoRNA. (Note: The GU and AG sequences at the beginning and end, respectively, of introns are invariant. However, additional sequences are critical for splice site recognition.) After introns have been removed and exons joined, the mature mRNA molecules pass into the cytosol through pores in the nuclear membrane. (Note: The introns in tRNA [see Fig. 31.16] are removed by a different mechanism.)

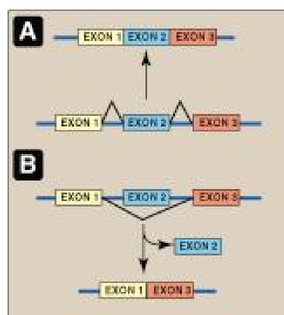
Effect of splice site mutations

Mutations at splice sites can lead to improper splicing and the production of aberrant proteins. It is estimated that at least 20% of all genetic diseases are a result of mutations that affect RNA splicing. For example, mutations that cause the incorrect splicing of β-globin mRNA are responsible for some cases of β-thalassemia, a disease in which the production of the β-globin protein is defective (see p. 39). Splice site mutations can result in exons being skipped (removed) or introns retained. They can also activate cryptic splice sites, which are sites that contain the 5′ or 3′ consensus sequence but are not normally used.

Alternative splicing

The pre-mRNA molecules from >90% of human genes can be spliced in alternative ways in different tissues. Because this produces multiple variations of the mRNA and, therefore, of its protein product (Fig. 31.19), it is a mechanism for producing a large, diverse set of proteins from a limited set of genes. For example, the mRNA for tropomyosin (TM), an actin filament–binding protein of the cytoskeleton (and of the contractile apparatus in muscle cells), undergoes extensive tissue-specific alternative splicing with production of multiple isoforms of the TM protein.

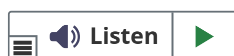
FIGURE 31.19



patterns in eukaryotic messenger RNA (mRNA).

) of exon 2 from the mRNA in panel B results in a protein product that is different than e mRNA in panel A.

Chapter Summary

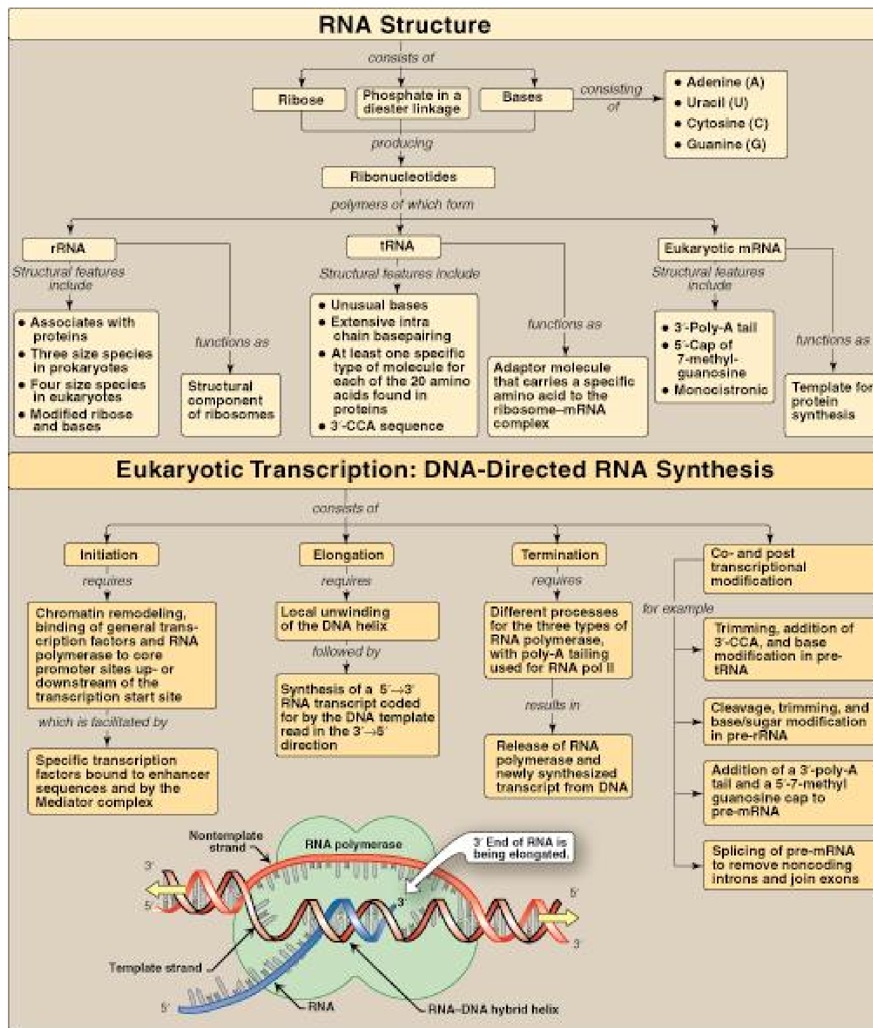


- Three major types of RNA participate in the process of protein synthesis: **rRNA**, **tRNA**, and **mRNA**. RNA differ from DNA by containing **ribose** instead of deoxyribose and **U** instead of T. **rRNA** is a component of the **ribosomes**. **tRNA** serves as an **adaptor** molecule that carries a specific amino acid to the site of protein synthesis. **mRNA** (coding RNA) carries genetic information from DNA for use in protein synthesis.
- The process of RNA synthesis is called **transcription**. The enzyme that synthesizes RNA, **RNA pol**, uses **ribonucleoside triphosphates** as substrates for **5′→3′ polymerase activity**. In both prokaryotes and eukaryotes, RNA pol does not require a primer.
- In **prokaryotic** cells, the **core RNA pol enzyme** has five subunits (2 α, 1 β, 1 β′, and 1 Ω). The core enzyme requires an additional subunit, **sigma (σ) factor**, to recognize the nucleotide sequence (**promoter** region) in DNA. This region contains **consensus sequences** that are highly conserved and include the –**10 Pribnow box** and the –**35 sequence**. Another protein, **rho** (**ρ**), is required for **termination** of transcription of some genes.
- In the **eukaryotic** cell nucleus, there are three distinct types of RNA pol. **RNA pol I** synthesizes the precursor of rRNA in the nucleolus. **RNA pol II** synthesizes the precursors for mRNA and some ncRNA, and **RNA pol III** synthesizes the precursors of tRNA and 5S rRNA. Core **promoters** for genes transcribed by **RNA pol II** contain **cis-acting** consensus sequences, such as the **TATA** (**Hogness**) **box**, which serve as binding sites for **trans-acting GTFs**. Upstream of these are **proximal** regulatory elements, such as the CAAT and GC boxes, and **distal** regulatory elements, such as **enhancers**. **STFs** (transcriptional activators) and **mediator complex** bind these elements and regulate gene expression. Eukaryotic transcription requires that the **chromatin** be relaxed (decondensed) in a process known as **chromatin remodeling**.
- **A primary transcript** is a linear copy of a **transcription unit**, the segment of DNA between specific initiation and termination sequences. Prokaryotic mRNA is generally identical to its primary transcript, whereas eukaryotic **pre-mRNA** is extensively modified co- and posttranscriptionally. For example, a **7-methylguanosine cap** is attached to the 5′ end of the mRNA through a 5′-to-5′ linkage. A **long poly-A tail** is attached by polyadenylate polymerase to the 3′ end of most mRNA. Most eukaryotic mRNA also contains i**ntervening sequences** (**introns**) that must be removed for the mRNA to be functional. Their removal, as well as the joining of **expressed sequences** (**exons**), requires a **spliceosome** composed of **"snurps"** that mediate the process of **splicing**. Eukaryotic mRNA is **monocistronic**, containing information from just one gene, whereas prokaryotic mRNA is **polycistronic (**Fig. 31.20**)**.

FIGURE 31.20

Key concept map for RNA structure and synthesis.

rRNA = ribosomal RNA; tRNA = transfer RNA; mRNA = messenger RNA.



Study Questions

Choose the ONE best answer.

31.1. An 8-month-old male with severe anemia is found to have β-thalassemia. Genetic analysis shows that one of his β-globin genes has a mutation that creates a new splice acceptor site 19 nucleotides upstream of the normal splice acceptor site of the first intron. Which of the following best describes the new messenger RNA molecule that can be produced from this mutant gene?

A. Exon 1 will be too short.

B. Exon 1 will be too long.

C. Exon 2 will be too short.

D. Exon 2 will be too long.

E. Exon 2 will be missing.

Correct answer = D. Because the mutation creates an additional splice acceptor site (the 3′ end) upstream of the normal acceptor site of intron 1, the 19 nucleotides that are usually found at the 3′ end of the excised intron 1 lariat can remain behind as part of exon 2. The presence of these extra nucleotides in the coding region of the mutant messenger RNA (mRNA) molecule will prevent the ribosome from translating the message into a normal β-globin protein molecule. Those mRNA for which the normal splice site is used to remove the first intron will be normal, and their translation will produce normal β-globin protein.

31.2. A 4-year-old child who easily tires and has trouble walking is diagnosed with Duchenne muscular dystrophy, an X-linked recessive disorder. Genetic analysis shows that the patient's gene for the muscle protein dystrophin contains a mutation in its promoter region. Of the choices listed, which of the following would be the most likely to be defective due to this mutation?

A. Initiation of dystrophin transcription

B. Termination of dystrophin transcription

C. Capping of dystrophin messenger RNA

D. Splicing of dystrophin messenger RNA

E. Tailing of dystrophin messenger RNA

Correct answer = A. Mutations in the promoter typically prevent formation of the RNA polymerase II transcription initiation complex, resulting in a decrease in the initiation of messenger RNA (mRNA) synthesis. A deficiency of dystrophin mRNA will result in a deficiency in the production of the dystrophin protein. Capping, splicing, and tailing defects are not a consequence of promoter mutations. They can, however, result in mRNA with decreased stability (capping and tailing defects) or an mRNA in which exons have been skipped (lost) or introns retained (splicing defects).

31.3. A mutation to which of the following sequences in eukaryotic messenger RNA (mRNA) would most likely affect the process by which the 3′-end polyadenylate (poly-A) tail is added to the mRNA?

A. AAUAAA

B. CAAT

C. CCA

D. GU… A … AG

E. TATAAA

Correct answer = A. An endonuclease cleaves mRNA just downstream of this polyadenylation signal, creating a new 3′ end to which polyadenylate polymerase adds the poly-A tail using ATP as the substrate in a template-independent process. CAAT and TATAAA are sequences found in promoters for RNA polymerase II. CCA is added to the 3′ end of pre–transfer RNA by nucleotidyltransferase. GU…A…AG denotes an intron in eukaryotic pre-mRNA.

31.4. Which of the following protein factors identifies the promoter of protein-coding genes in eukaryotes?

A. Pribnow box

B. Rho

C. Sigma

D. TFIID

E. U1

Correct answer = D. The general transcription factor TFIID recognizes and binds core promoter elements such as the TATA-like box in eukaryotic protein-coding genes. These genes are transcribed by RNA polymerase II. The Pribnow box is a cis-acting element in prokaryotic promoters. Rho is involved in the termination of prokaryotic transcription. Sigma is the subunit of prokaryotic RNA polymerase that recognizes and binds the prokaryotic promoter. U1 is a ribonucleoprotein involved in splicing of eukaryotic pre-mRNA.

31.5. What is the sequence (conventionally written) of the RNA product of the DNA template sequence, GATCTAC, also conventionally written?

Correct answer = 5′-GUAGAUC-3′. Nucleic acid sequences are conventionally written 5′ to 3′. The template strand (5′-GATCTAC-3′) is used as 3′-CATCTAG-5′. The RNA product is complementary to the template strand (and identical to the coding strand), with U replacing T.