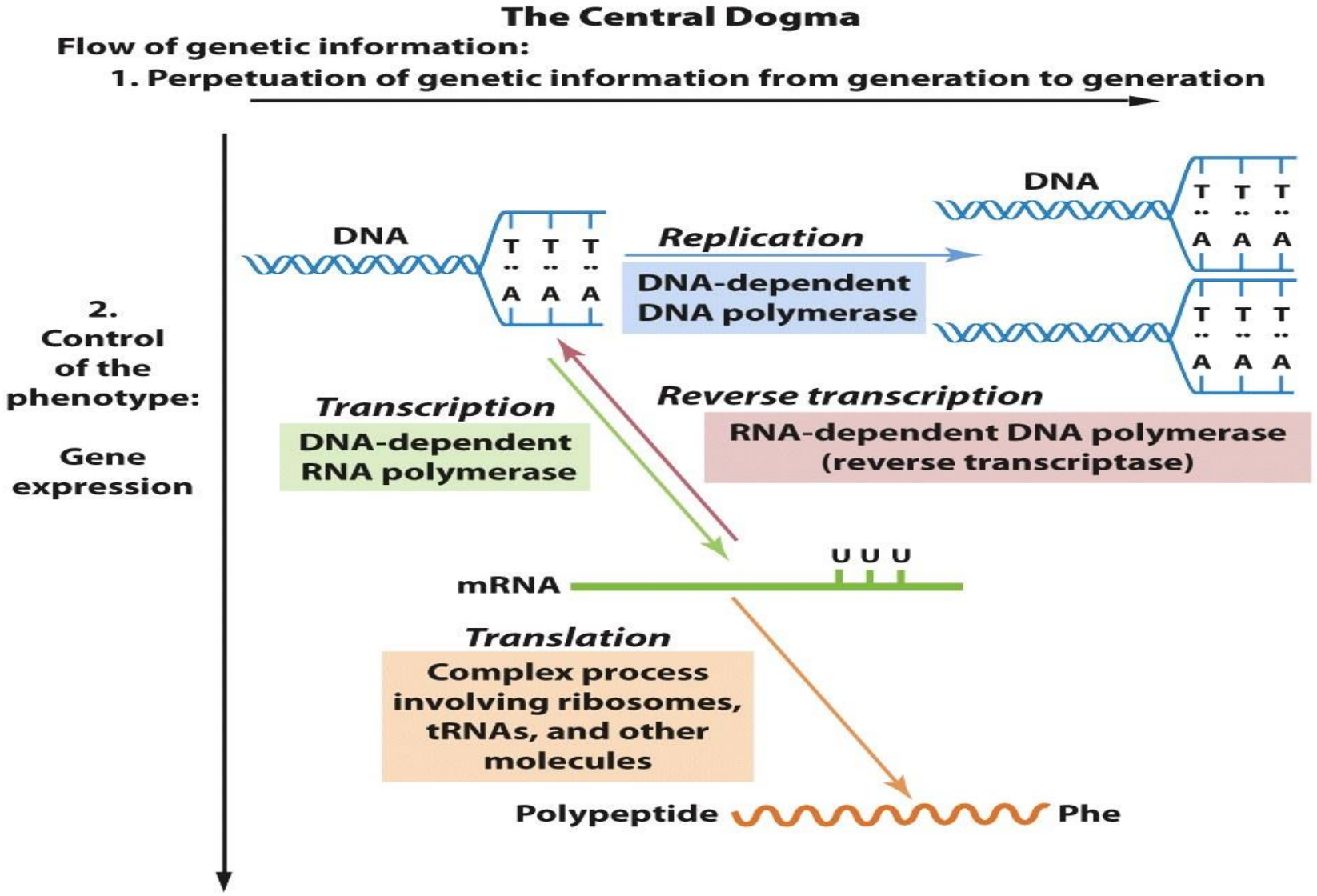
Transcription
RNA Processing

The Central Dogma of life

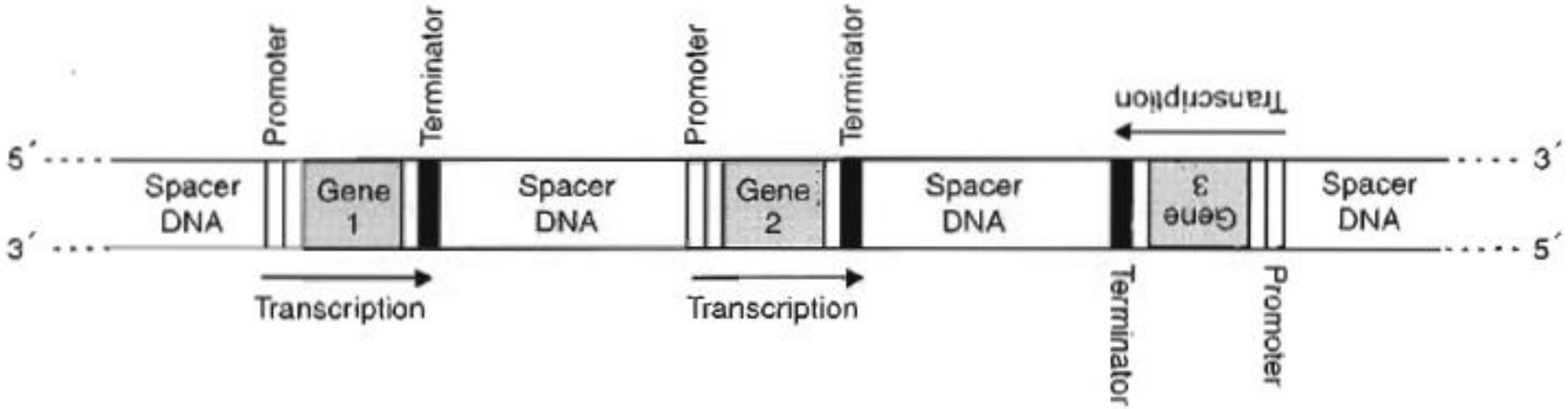


Overview of transcription

- The first stage in the expression of genetic information is transcription of the information in the DNA deoxyribonucleotides sequence into RNA ribonucleotides sequence.
- For any gene, only one strand of the DNA molecule, called the **template strand**, is transcribed by **RNA polymerase**.

- Because RNA polymerase moves in the 3' to 5' direction along the template strand of DNA, the RNA product is antiparallel and complementary to the template.

- RNA polymerase recognizes start signals (**promoters**) and stop signals (**terminators**) for each of the thousands of transcription units in the genome of an organism.



Transcription of several genes on a chromosome

Types of RNA

- RNA molecules play a variety of roles in the cell.

The types of RNA are: -

1- Ribosomal RNA (**rRNA**), which is the most abundant type of RNA in the cell.

2- Transfer RNA (**tRNA**), which is the second most abundant type of RNA.

3- Messenger RNA (**mRNA**) the only type of RNA that is translated, which carries the information specifying the amino acid sequence of a protein to the ribosome.

- The mRNA population in a cell is very heterogeneous in size and base sequence, as the cell has essentially a different mRNA molecule for each of the thousands of different proteins made by that cell.

4- Heterogeneous nuclear RNA (**hnRNA or pre-mRNA**) (the immediate product of gene transcription), which is found only in the nucleus of eukaryotic cells and it represents precursors of mRNA, 75% is degraded in the nucleus and 25% only is processed to mature RNA.

5- Small nuclear RNA (**snRNA**) (RNA molecules with enzymatic activity), which is also only found in the nucleus of eukaryotes, small in size and complexed with proteins (forming ribonucleoproteins).

One of its major functions is to participate in splicing (removal of introns) mRNA.

6- **Micro-RNA**, short, non-coding, ~ 22 nucleotide long, generated by nucleolytic processing of the products of distinct genes or transcription unites, at least some of which control the expression of other genes during development (mature micro RNA molecules can **hybridize** together to form **imperfect** RNA-RNA duplex within the 3' untranslated regions of specific target mRNA causing unexplained gene expression regulation in at least half of the human genes).

7- Small cytoplasmic RNA (**scRNA**), has **catalytic activity** in tRNA processing and acts as signal recognition particle.

8- Small nucleolar (**snoRNA**) acts in **rRNA processing**/ maturation/methylation.

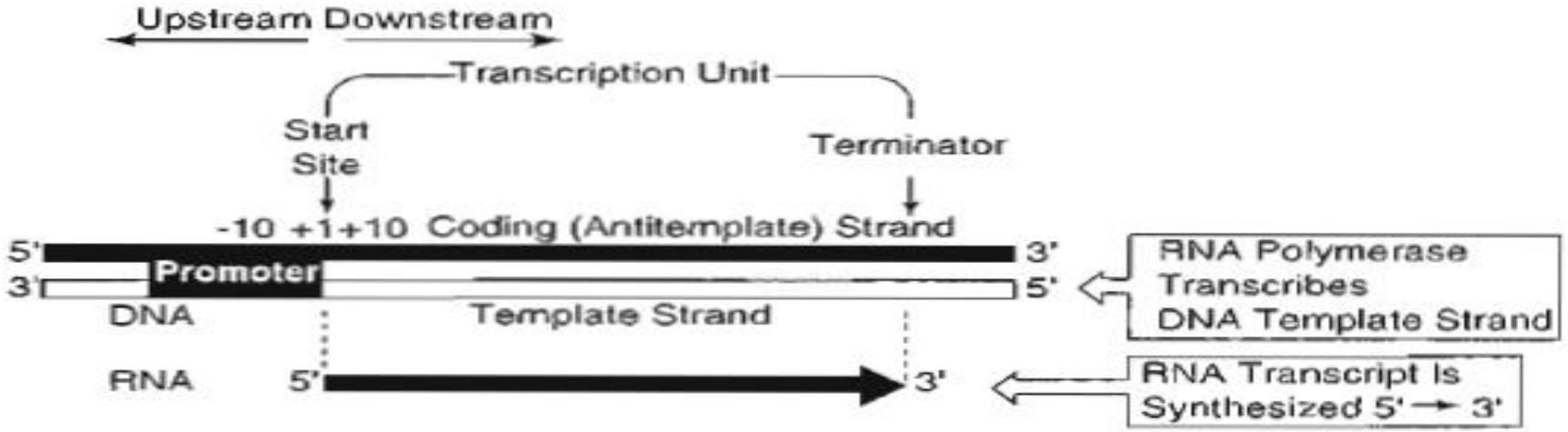
9- Small interfering RNA (**siRNA**) are derived by specific nucleolytic cleavage of larger **double stranded RNAs** to form small 21-25 long products.

- They form **perfect** RNA-RNA hybrids with their targets anywhere within the length RNA where the complementary sequence exists resulting in reduction of specific protein production because siRNA-mRNA complexes are degraded by nucleolytic machinery (interferes with the expression of specific gene by hybridizing to its corresponding RNA sequence in the target mRNA, then activates degradation of mRNA which can not be translated into proteins).

Transcription: important concepts and terminology

- RNA polymerase locates genes in DNA by searching for promoter regions.
- The promoter is the binding site for transcription factors and RNA polymerase.
- Binding establishes where transcription begins, which strand of DNA is used as the template, and in which direction transcription proceeds.
- RNA polymerase moves along the template strand in the 3' to 5' direction as it synthesizes the RNA product in the 5' to 3' direction using NTPs (ATP, GTP, CTP, UTP) as substrates.
- RNA polymerase does not proofread its work.
- The RNA product is **complementary** and **antiparallel** to the **template strand**.

- The coding (**non-template**) strand is not used during transcription. It is identical in sequence to the RNA molecule, except that RNA contains uracil instead of the thymine found in DNA.
- By convention, the base sequence of a gene is given from the coding strand (5' → 3').
- Transcription ends when RNA polymerase reaches a termination signal.



Transcription of DNA

RNA Polymerases

- There is a single prokaryotic RNA polymerase that synthesizes all types of RNA in the cell.
- The core polymerase has the subunit structure ($\alpha_2\beta\beta`$).
- A protein factor called sigma (σ) is required for the initiation of transcription at the promoter.
- σ factor is released immediately after transcription initiation.

Functions of the subunits:

α: assembly of the tetrameric core

β: ribonucleoside triphosphate binding site (link ribonucleotides together)

β': DNA template binding region

σ (sigma factor): initiation of transcription

- Termination of transcription sometimes requires a protein called rho (**ρ**) factor.
- This enzyme is inhibited by rifampin and actinomycin D.

Promoter "Strength" (activity)

- Affects amount of RNA made, so, it affects level of expression for that gene.
- Not all promoters have same "strength"

- Promoters differ in DNA sequences and "strength"

- RNA polymerase binds differently to different sequences
  "Strong promoters" initiate transcription more often than "weak promoters"

- rRNA has strong promoter: ~1 initiation per second

- lacZ has a weak promoter: ~1 initiation per minute

Eukaryotic RNA polymerases:

- Three types which can be distinguished by the particular types of RNA they produce:

1- **RNA polymerase I** is located in the nucleolus and synthesizes 28S, 18S, and 5.8S rRNAs.

2- **RNA polymerase II** is located in the nucleoplasm and synthesizes hnRNA/mRNA and some snRNA.

3- **RNA polymerase III** is located in the nucleoplasm and synthesizes tRNA, some snRNA, and 5S rRNA.

- Transcription factors (such as **TFIID** for RNA polymerase II) help to initiate transcription.

- The requirements for termination of transcription in eukaryotes are not well understood.

- In addition, RNA polymerase II is inhibited by (**α amanitin**) a toxin from certain mushrooms. It inactivates RNA pol II and can kill a person, while, RNA pol I and III are less affected by toxin

Prokaryotic	Eukaryotic
Single RNA polymerase ($\alpha_2\beta\beta'$)	RNAP 1: rRNA (nucleolus), except 5S rRNA RNAP 2: hnRNA/mRNA and some snRNA RNAP 3: tRNA, 5S rRNA
Requires sigma (σ) to initiate at a promoter	No sigma, but transcription factors (TFIID) bind before RNA polymerase
Sometimes requires rho (ρ) to terminate	No rho required
Inhibited by rifampin → Actinomycin D	RNAP 2 inhibited by α-amanitin (mushrooms) Actinomycin D

Comparison of eukaryotic and prokaryotic RNA polymerases

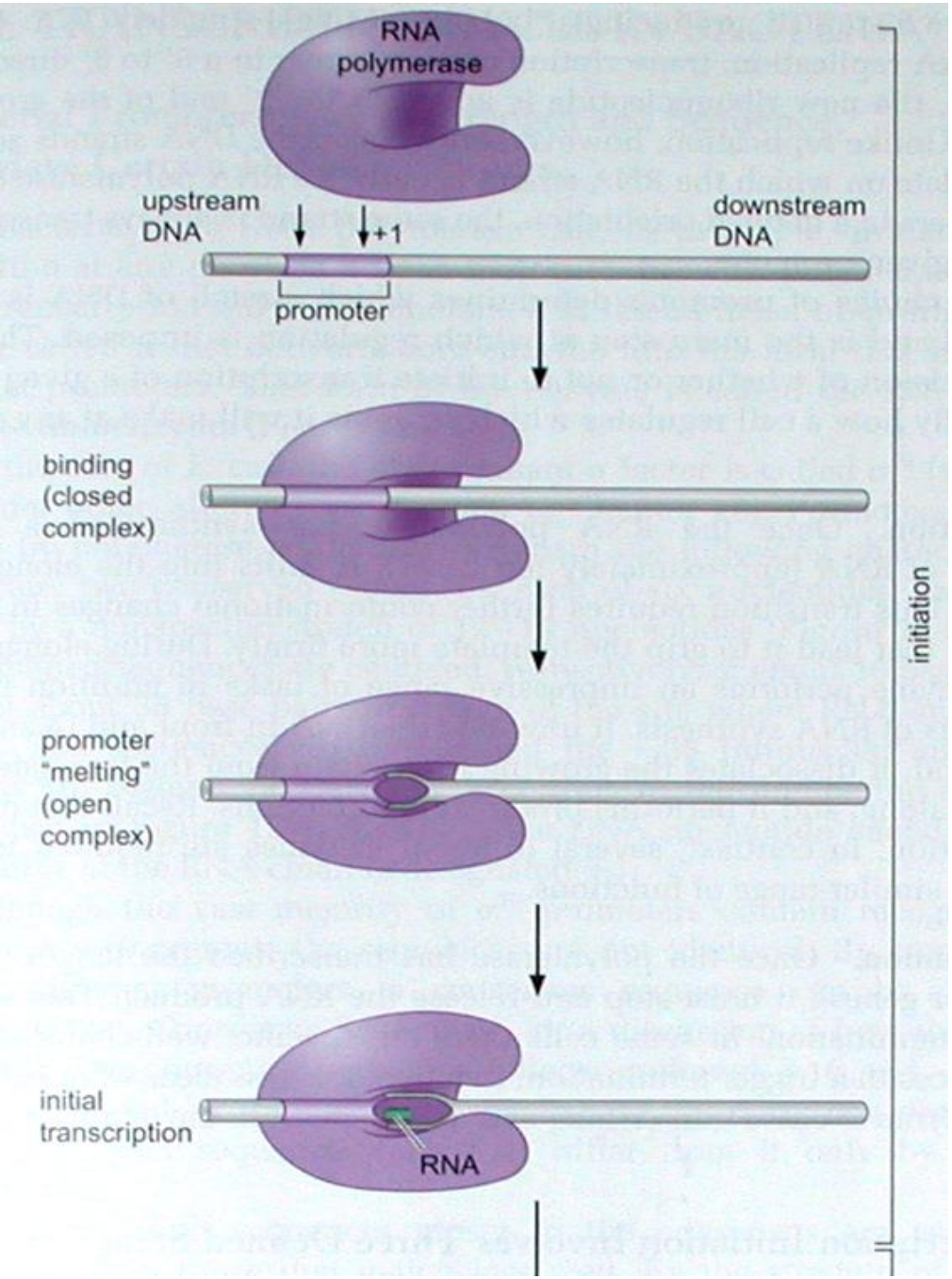
- The following events occur during the expression of a prokaryotic gene:
- With the help of sigma factor, RNA polymerase recognizes and binds to the promoter, region.
- The bacterial promoter contains two "consensus" sequences, called the **Pribnow box [**TATA (TATTAT) box] and the **-35 sequence** (TGTTGACA).

1- The promoter identifies the start site for transcription and orients the enzyme on the template strand.

2- Transcription begins at the + 1 base pair. Sigma factor is released as soon as transcription is initiated.

3- The core polymerase continues moving along the template strand in the 3' to 5' direction, synthesizing the mRNA in the 5' to 3' direction.

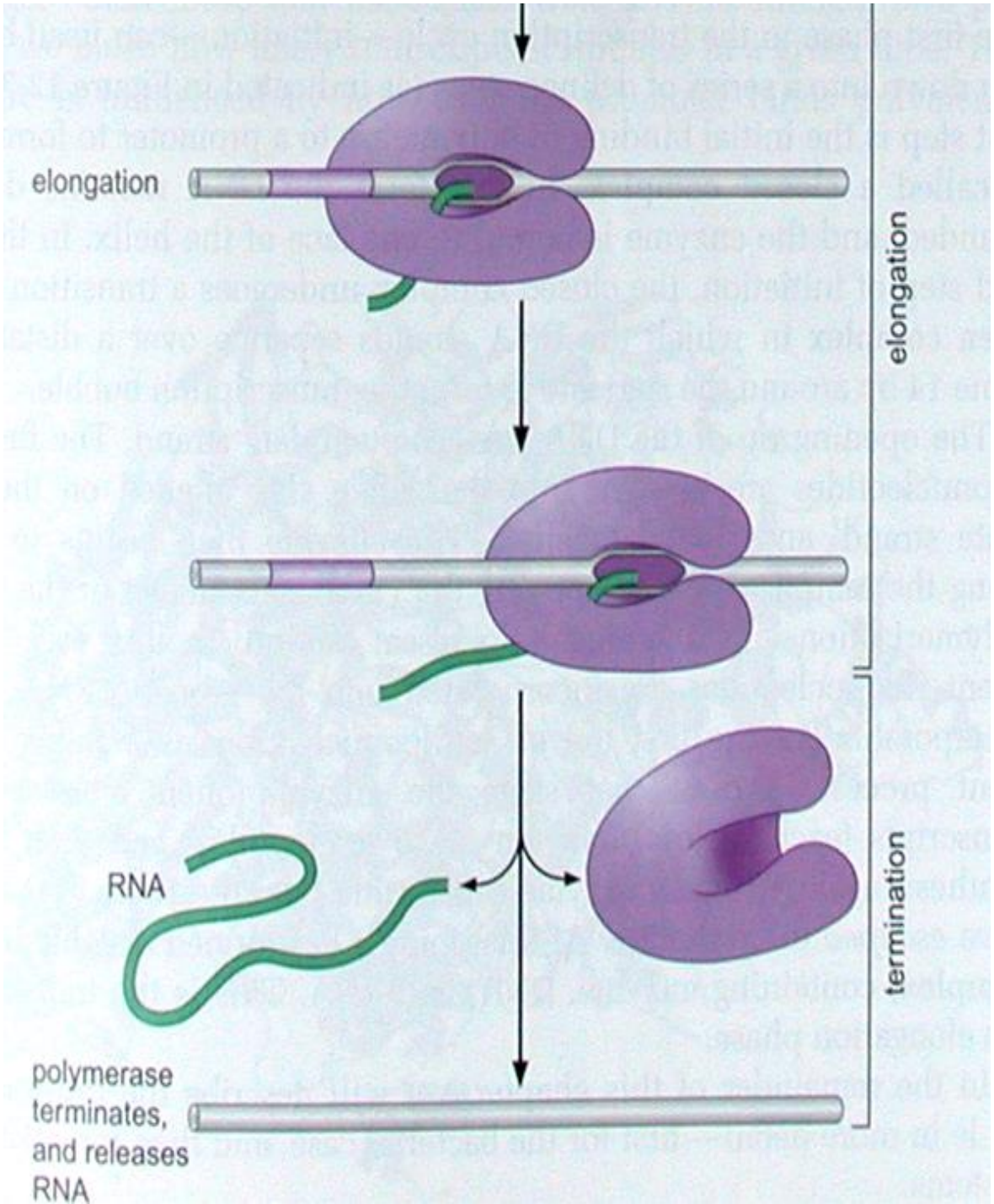
Initiation



A- Binding (closed complex)

B- Promoter “melting” (open complex)

C- Initial transcription



Elongation

Termination

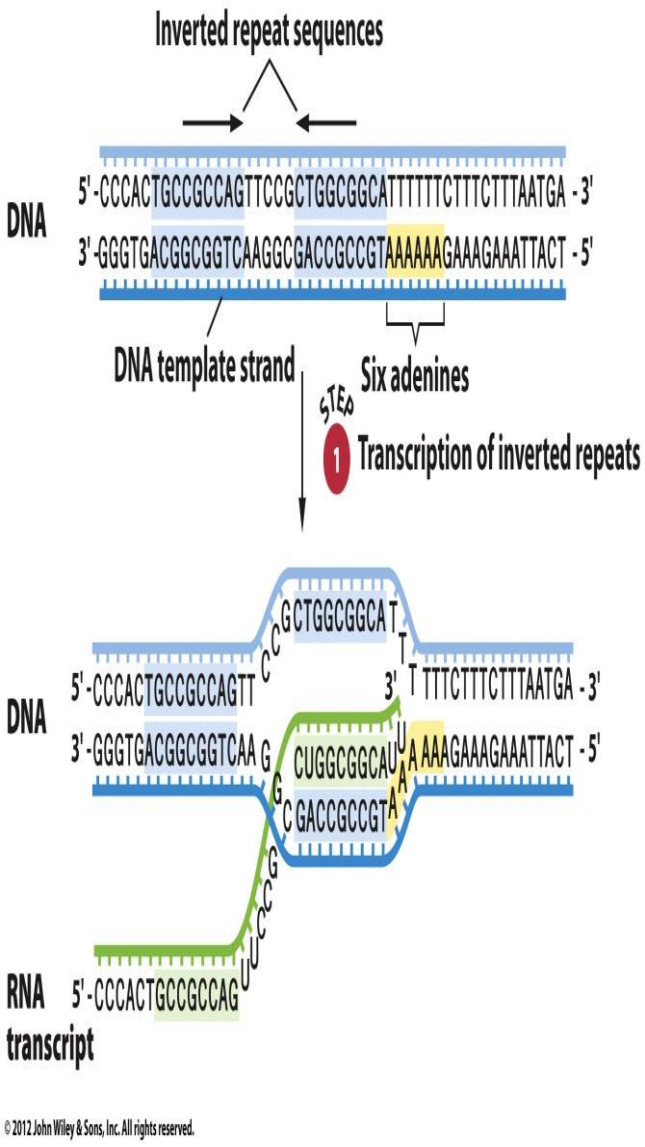
4. RNA polymerase eventually reaches a transcription termination signal, at which point it will stop transcription and release the completed mRNA molecule.

- There are two kinds of transcription terminators commonly found in prokaryotic genes:

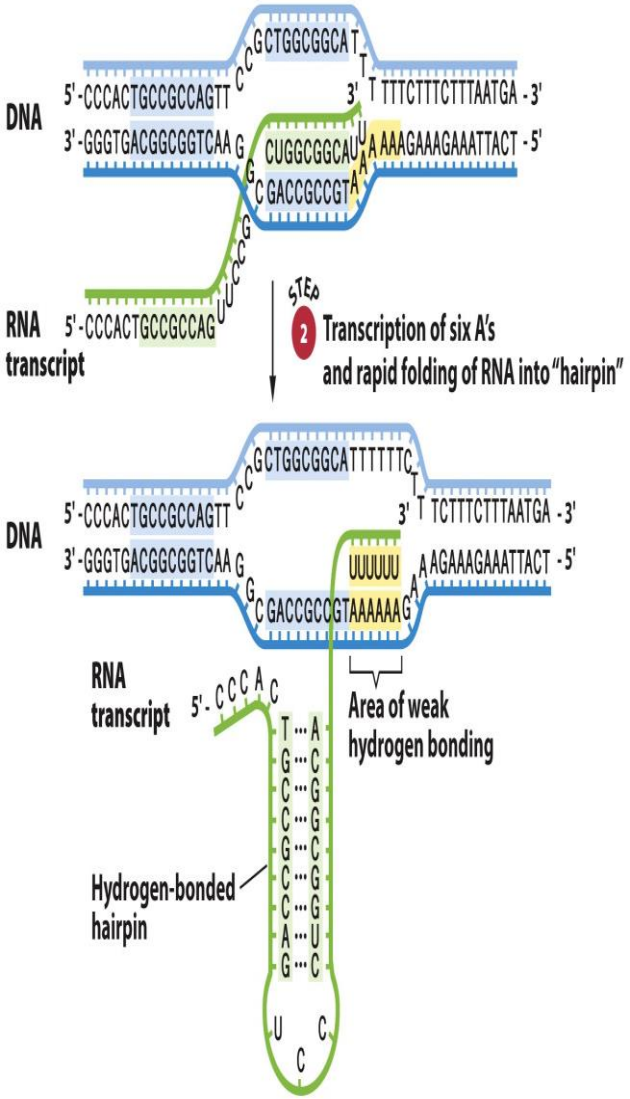
  A- **Rho-independent termination** occurs when the newly formed RNA folds back on itself to form a GC-rich hairpin loop closely followed by **6-8 U** residues. These two structural features of the newly synthesized RNA promote dissociation of the RNA from DNA template.

  B- **Rho-dependent termination** requires participation of rho factor. This protein binds to the newly formed RNA and moves toward the RNA polymerase that has paused at a termination site. Rho then displaces RNA polymerase from the 3' end of the RNA.

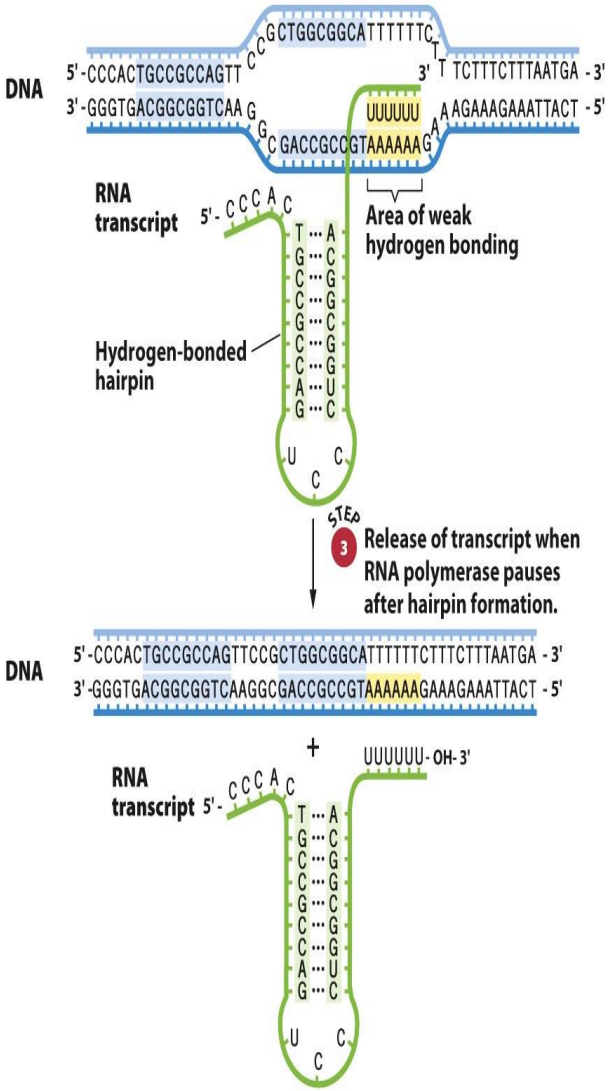
Transcription termination



A



B

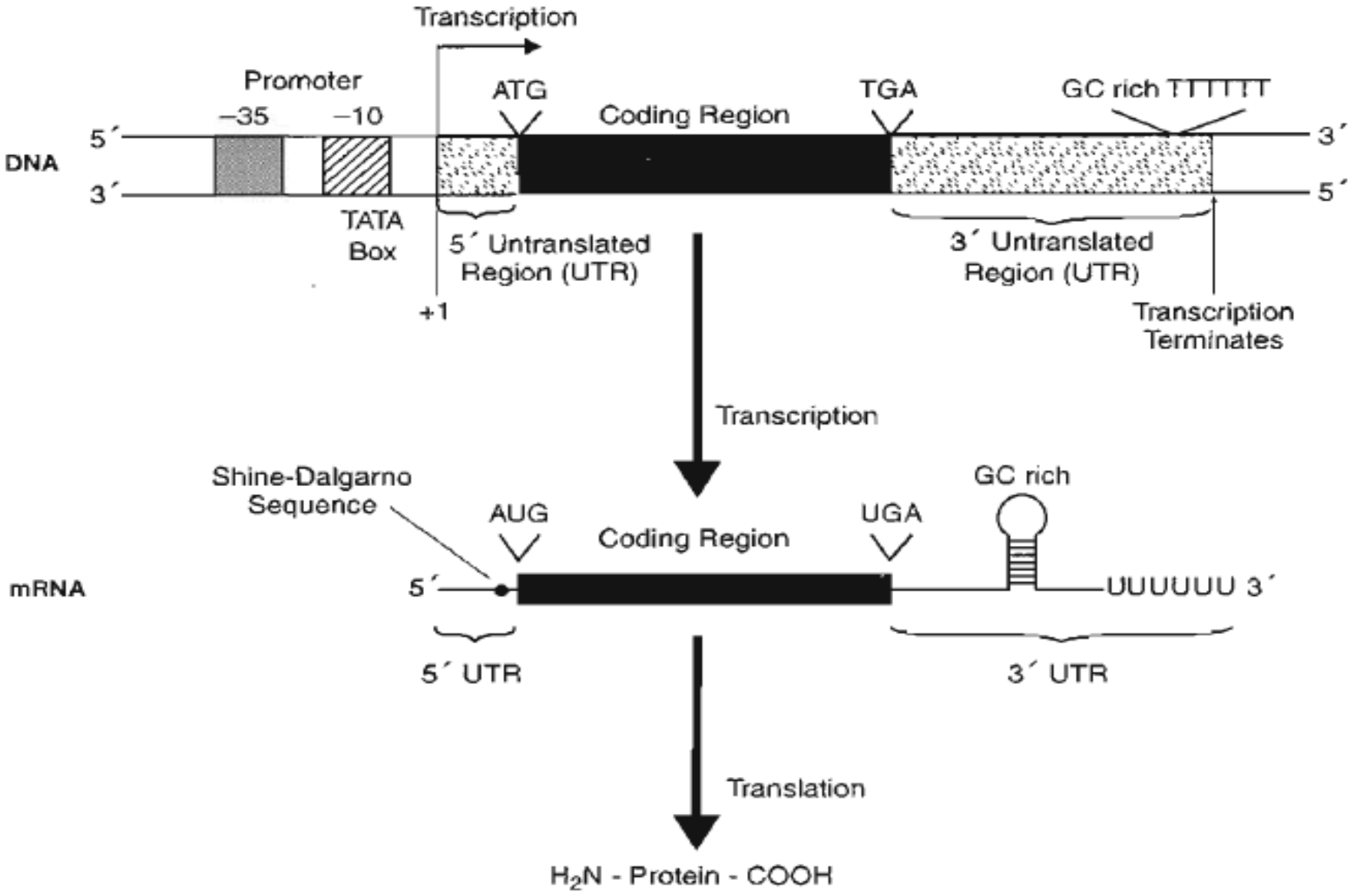


C

5. Transcription and translation can occur simultaneously in bacteria because there is no processing of prokaryotic mRNA (generally no introns), ribosomes can begin translating the message even before transcription is complete.

- Ribosomes bind to a sequence called Shine-Dalgarno sequence in the 5' untranslated region (UTR) of the message
- Protein synthesis begins at an AUG codon at the beginning of the coding region and continues until the ribosome reaches a stop codon at the end of the coding region.

6. The ribosome translates the message in the 5' to 3' direction, synthesizing the protein from amino terminus to carboxyl terminus.

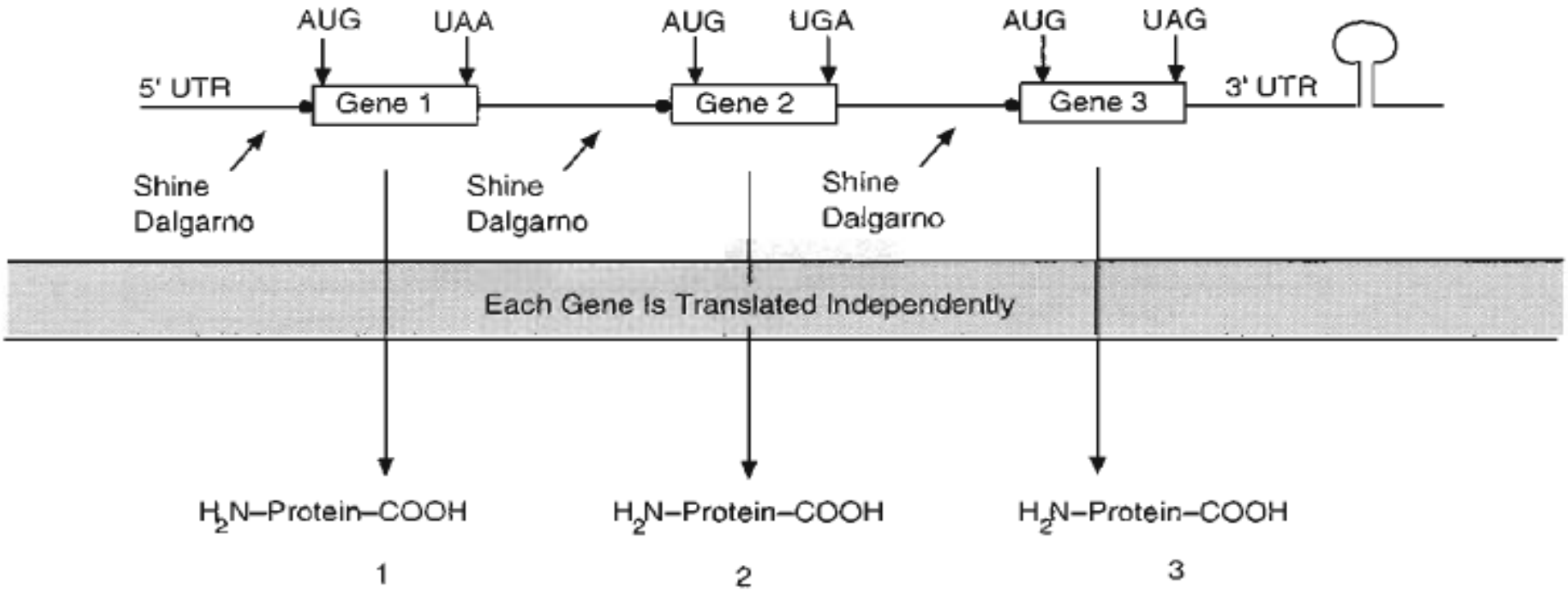


A prokaryotic transcription unit.

- The mRNA produced by the gene shown above is a **monocistronic** message. That is, it is transcribed from a single gene and codes for only a single protein.

- The word **cistron** is another name for a gene. Some bacterial operons produce **polycistronic** messages. In these cases, related genes grouped together in the DNA are transcribed as one unit.

- The mRNA in this case contains information from several genes and codes for several different proteins



Prokaryotic polycistronic message codes for several different proteins

Production of eukaryotic mRNA

- In eukaryotes, most genes are composed of coding segments (**exons**) interrupted by noncoding segments (**introns**).

- Both exons and introns are transcribed in the nucleus.

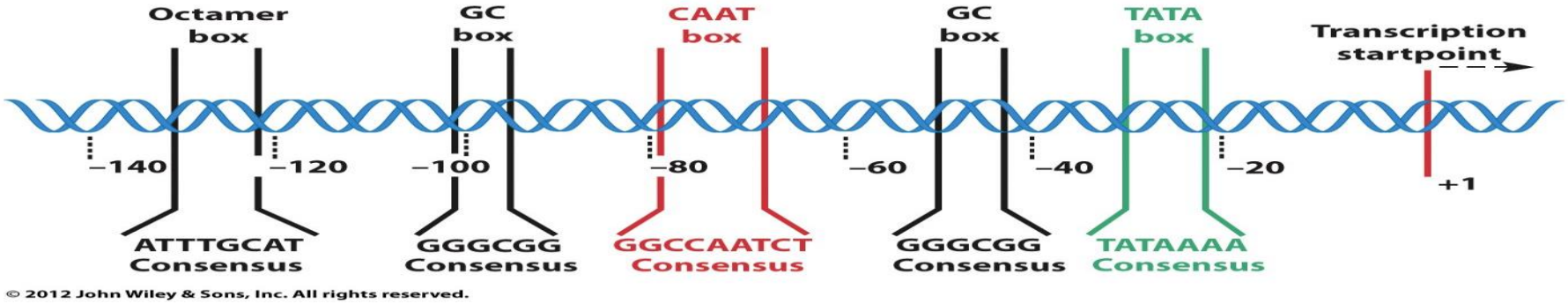
- Introns are removed during processing of the RNA molecule in the nucleus.

- In eukaryotes, all mRNA is **monocistronic**.
- The mature mRNA is translated in the cytoplasm.

- Transcription of a typical eukaryotic gene occurs as follows:

1. With the help of proteins called transcription factors, RNA polymerase II recognizes and binds to the promoter region. The basal promoter region of eukaryotic genes usually has **two consensus sequences** called the **TATA box** (also called **Hogness box**) and the **CAAT box**.

2. RNA polymerase II separates the strands of the DNA over a short region to initiate transcription and read the DNA sequence. The template strand is read in the 3' to 5' direction as the RNA product (the **primary transcript**) is synthesized in the 5' to 3' direction.



A Typical RNA Polymerase II Promoter

Class II promoters (most similar to bacterial promoters)

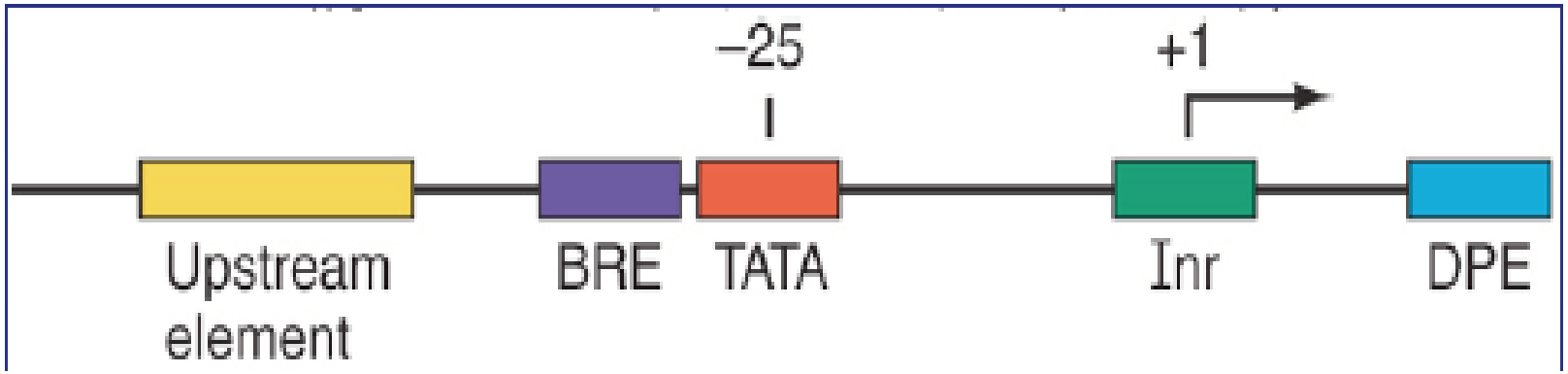
- Common type of promoter (most genes use this)
- Many variations, but "consensus" has a "Core" + "Upstream" Core (3 elements):

1. "TATA box" (5'-TATA-3')
2. TFIIB recognition element (BRE)
3. Initiator box (Inr) with an "A" at +1, most common

- Downstream promoter element (DPE, less common)
- Core promoter is recognized by general TFs that associate with RNA pol to form a preinitiation complex at great majority of promoters
- At least one of these elements is missing in most promoters
  e.g., highly expressed specialized genes tend to have TATA boxes, but promoters for housekeeping genes tend to lack them

- Upstream elements: quite varied in number and can be orientation-independent (but relatively position-dependent) & recognized by other TFs (relatively gene-specific) that participate in initiation at smaller sub-sets of promoters.

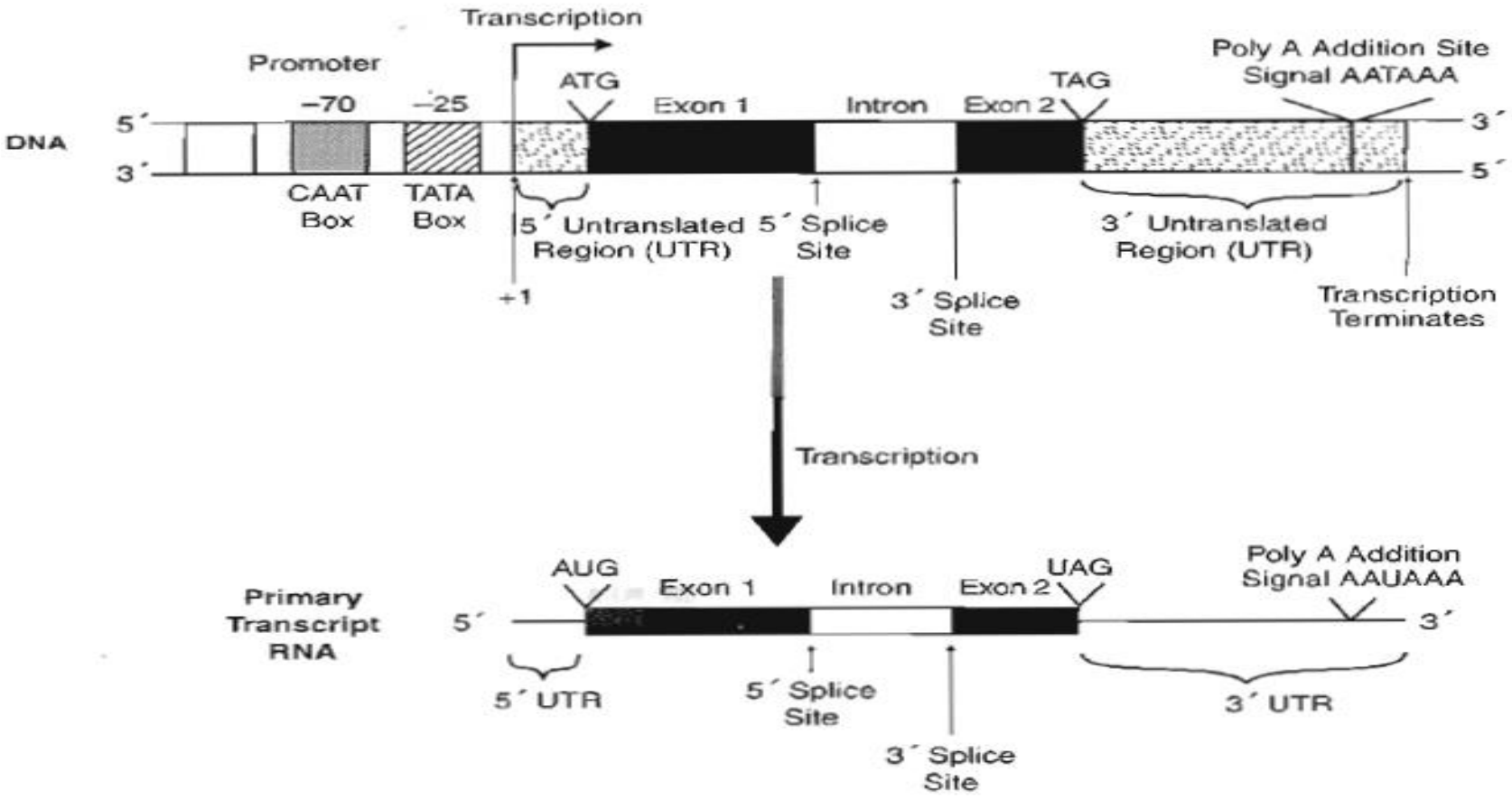
1. GC box (GC rich)
2. CAAT box (5’-CCAAT-3)



TFIIB recognition element; Inr: initiator box; DPE: downstream promoter element

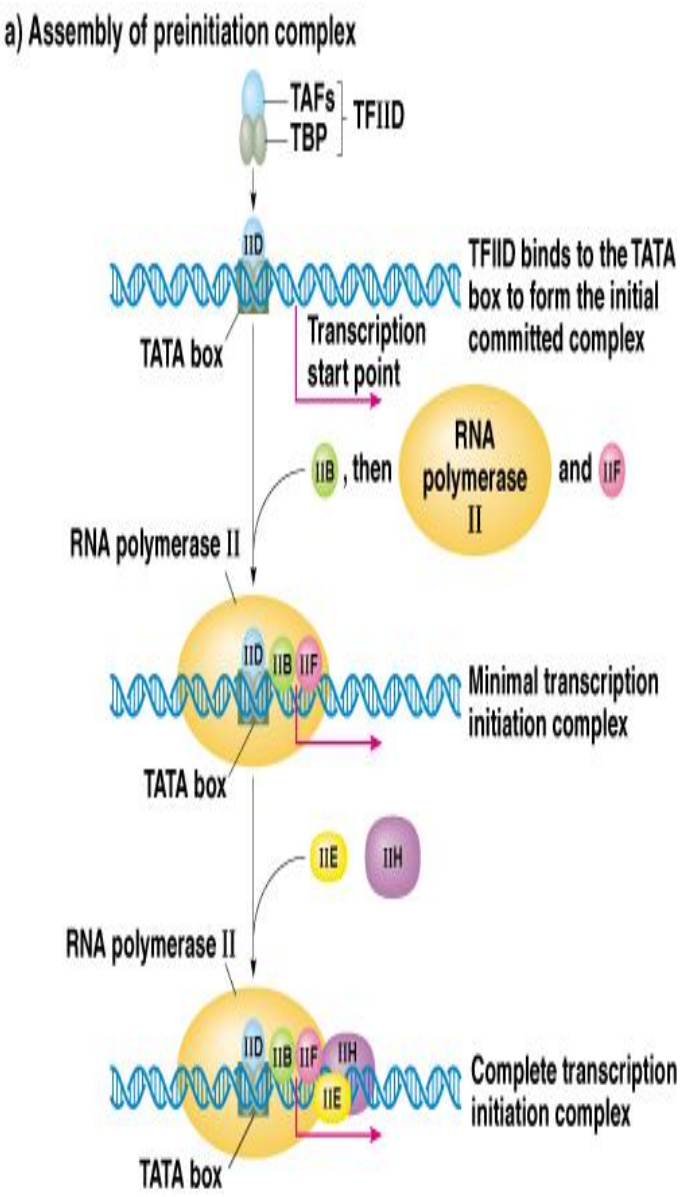
3. RNA polymerase II ends transcription when it reaches a termination signal.

- These signals are not well understood in eukaryotes.



A eukaryotic transcription unit

Order of binding is: IID + IIA + IIB + RNA poly. II + IIF +IIE +IIH



1- TBP in TFIID binds to the TATA box
2- TFIIA and TFIIB are recruited with TFIIB binding to the BRE
3- RNA Pol II-TFIIF complex is then recruited
4- TFIIE and TFIIH then bind upstream of Pol II to form the pre-initiation complex
- Promoter melting using energy from ATP hydrolysis by TFIIH)
- Promoter escapes after the phosphorylation of the C-terminal domain tail

Table 11.3

General Transcription Initiation Factors

Factor	Subunits	Size (kDa)	Function
TFIID-TBP	1	27	TATA box recognition, positioning of TATA box DNA around TFIIB and Pol II
TFIID-TAF_{II}s	14	15–250	Core promoter recognition (non-TATA elements), positive and negative regulation
TFIIA	3	12, 19, 35	Stabilization of TBP binding; stabilization of TAF–DNA binding
TFIIB	1	38	Recruitment of Pol II and TFIIF; start-site recognition for Pol II
TFIIF	3	156 total	Promoter targeting of Pol II
TFIIE	2	92 total	TFIIH recruitment; modulation of TFIIH helicase ATPase, and kinase activities; promoter melting
TFIIH	9	525 total	Promoter melting; promoter clearance via phosphorylation of CTD

Eukaryotic Transcription: DNA-Directed RNA Synthesis

consists of

Initiation

requires

Chromatin remodeling, binding of general transcription factors and *RNA polymerase* to core promoter sites up- or downstream of the coding region

which is facilitated by

Specific transcription factors bound to enhancer sequences

Elongation

requires

Local unwinding of the DNA helix

followed by

Synthesis of a 5'→3' RNA transcript coded for by the DNA template read in the 3'→5' direction

Termination

requires

A termination signal sequence

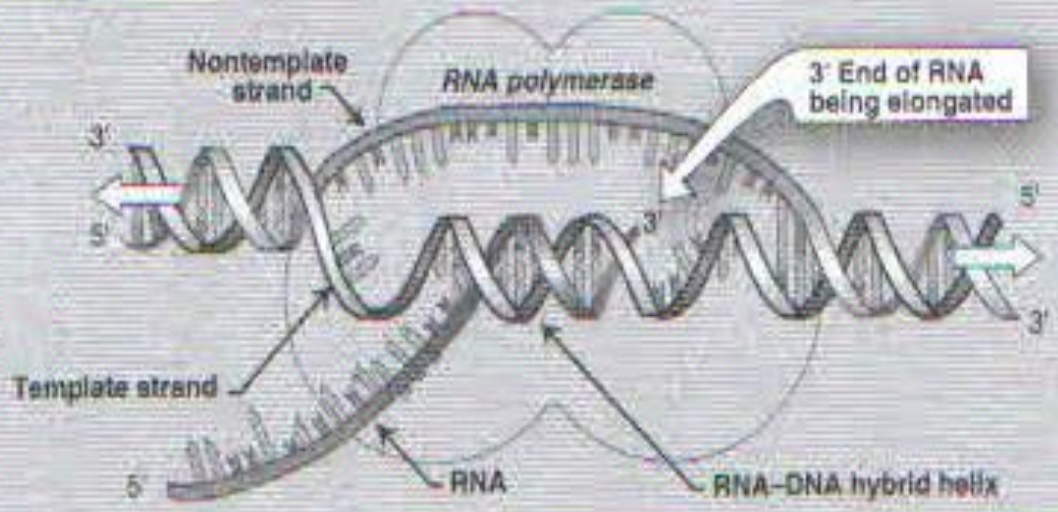
results in

Release of *RNA polymerase* and newly synthesized transcript from DNA

Co- and post-transcriptional modification

for example

- Splicing of pre-mRNA to remove non-coding introns and join exons
- Cleavage, trimming and base/sugar modification in pre-rRNA
- Trimming, addition of 3'-CCA, and base modification in pre-tRNA.
- Addition of a 3' poly-A "tail" and a 5'-7-methyl guanosine "cap" to pre-mRNA.



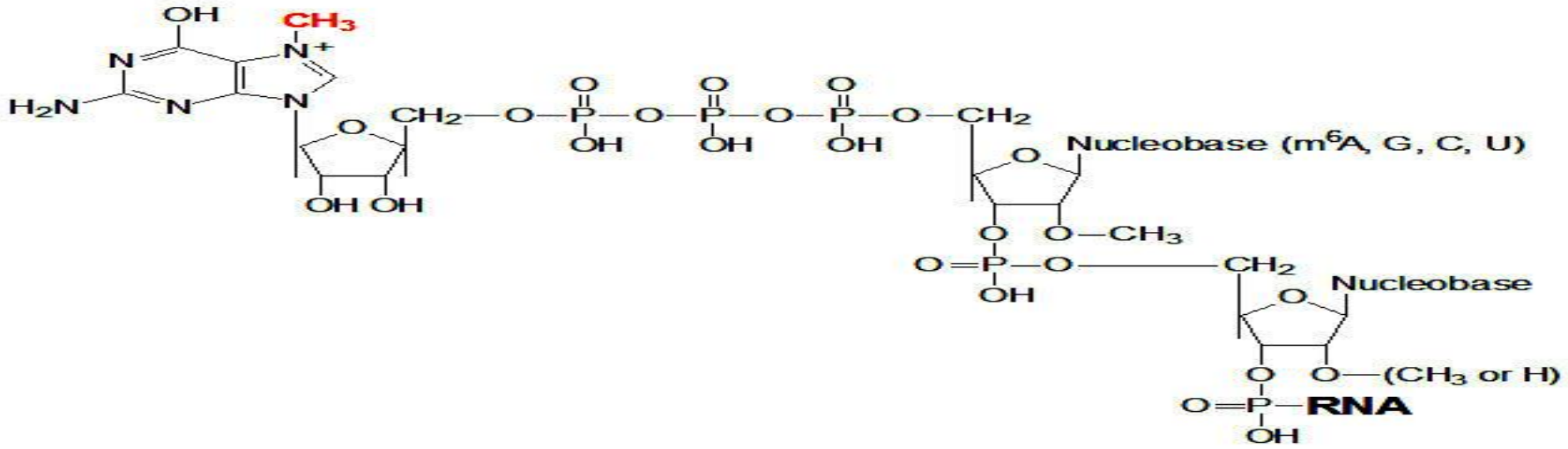
Transcription enhancers and silencers

- Both are binding sites for transcription factors (TF's)
- **Enhancers**: "non-promoter DNA elements that stimulate transcription" They interact with general transcription factors to promote formation of pre-initiation complex to increase the amount of Transcription from a nearby promoter (core + upstream elements)
- **Silencers**: Decrease amount of Transcription from nearby promoters
- Initially Defined as being "Position and orientation independent"
- Found upstream, within, or downstream of genes, they function in either orientation (not always true)
- Sometimes a DNA element can act as an enhancer or a silencer depending on what is bound to it.

Posttranscriptional processing of RNAs:

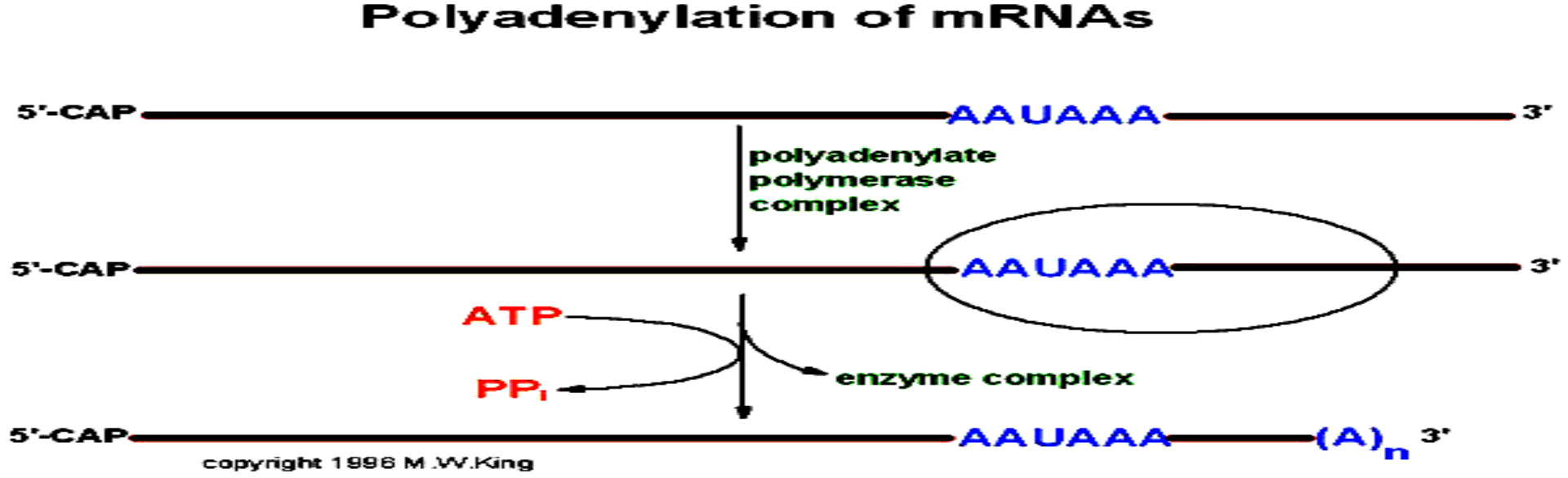
- Bacterial rRNAs and tRNAs undergoes no additional processing, after being transcribed they are immediately ready for use in translation.
- Translation of bacterial mRNAs can begin even before transcription is completed due to the lack of the nuclear-cytoplasmic separation that exists in eukaryotes and to afford a unique opportunity for regulating the transcription of certain genes.
- An additional feature of bacterial mRNAs is that most are **polycistronic** which means that multiple polypeptides can be synthesized from a single primary transcript.
- Polycistronic mRNAs are very rare in eukaryotic cells but have been identified.
- In addition, several viruses encode **polycistronic** RNAs.

- In contrast to bacterial transcripts, eukaryotic RNAs (all 3 classes) undergo post-transcriptional processing.
- All 3 classes of RNA are transcribed from genes that contain introns.
- The sequences encoded by the intronic DNA must be removed from the primary transcript prior to the RNAs being biologically active, this process of intron removal is called **RNA splicing,** additional processing occurs to mRNAs, the 5' end of all eukaryotic mRNAs are capped with a unique 5'→5' linkage to a **7-methyl GTP**.
- The capped end of the mRNA is thus, protected from **exonucleases** and more importantly is recognized by specific proteins of the translational machinery.
- The capping process occurs after the newly synthesizing mRNA is around **20–30** bases long.



Structure of the 5'-cap of eukaryotic mRNAs

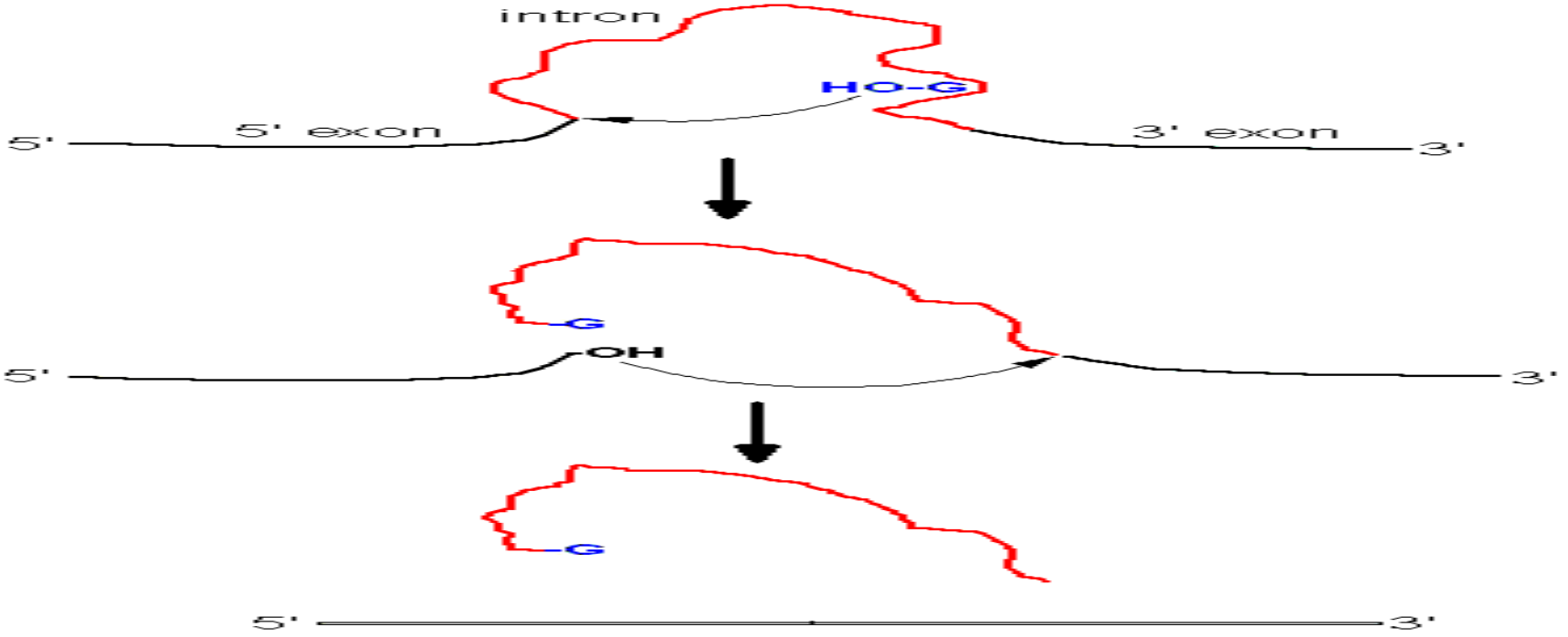
- Messenger RNAs also are polyadenylated at the 3' end.
A specific sequence, AAUAAA, is recognized by the endonuclease activity of by **polyadenylate polymerase** which cleaves the primary transcript approximately 11–30 bases 3' of the sequence element.
- A stretch of 20–250 A residues is then added to the 3' end by the **polyadenylate polymerase** activity.



Splicing of RNAs

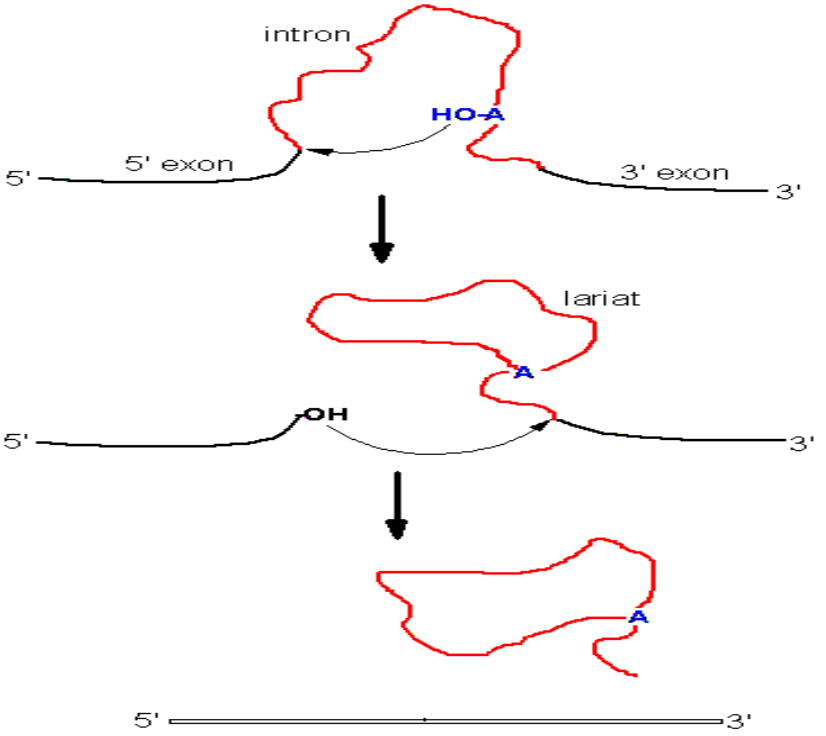
- There are several different classes of reactions involved in intron removal.
- The 2 most common are the group I and group II introns.
  Group I introns are found in nuclear, mitochondrial and chloroplast rRNA genes, group II in mitochondrial and chloroplast mRNA genes.
- Many of the group I and group II introns are self-splicing.
- Group I introns require an external guanosine as a cofactor.

- The 3'–OH of the guanosine nucleotide acts as a nucleophile to attack the 5'–phosphate of the 5' nucleotide of the intron.
- The resultant 3'–OH at the 3' end of the 5' exon then attacks the 5' nucleotide of the 3' exon releasing the intron and covalently attaching the two exons together.
- The 3' end of the 5' exon is termed the splice donor site and the 5' end of the 3' exon is termed the splice acceptor site.



Splicing by group 1 introns

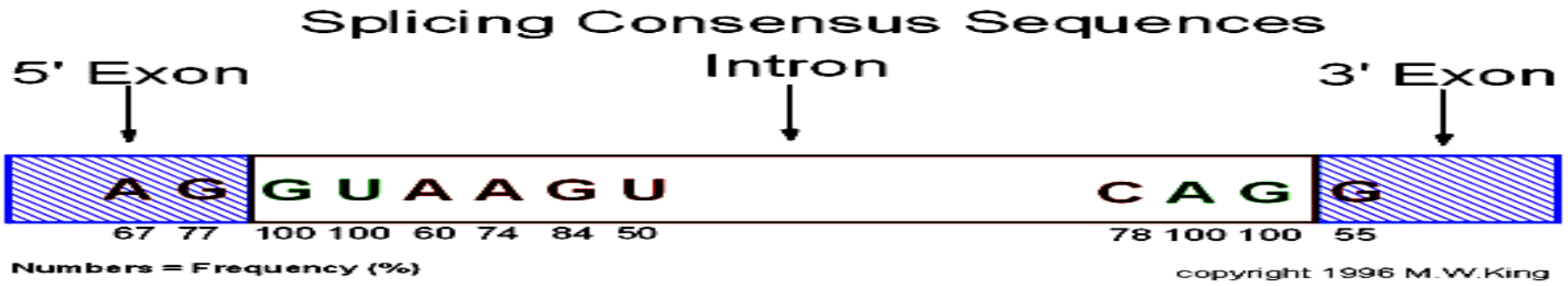
- Group II introns are spliced similarly except that instead of an external nucleophile, the 2'–OH of an adenine residue within the intron is the nucleophile.
- This residue attacks the 3' nucleotide of the 5' exon forming an internal loop called a lariat structure.
- The 3' end of the 5' exon then attacks the 5' end of the 3' exon as in group I splicing releasing the intron and covalently attaching the two exons together.



Splicing by group 2 introns

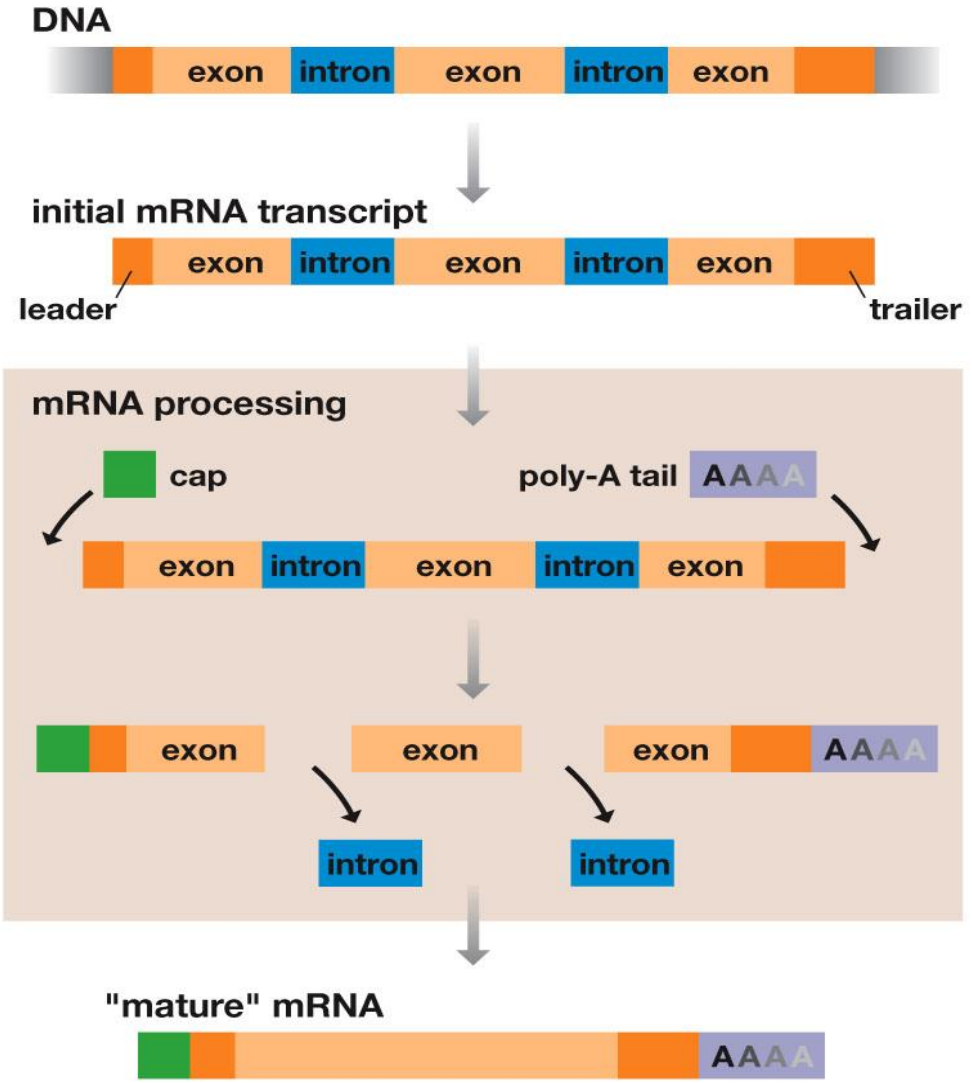
- The third class of introns is also the largest class found in nuclear mRNAs, that undergoes a splicing reaction similar to group II introns in that an internal lariat structure is formed.
- However, the splicing is catalyzed by specialized RNA–protein complexes called small nuclear ribonucleoprotein particles (snRNPs).
- The RNAs found in snRNPs are identified as U1, U2, U4, U5 and U6.
- Analysis of a large number of mRNA genes has led to the identification of highly conserved consensus sequences at the 5' and 3' ends of essentially all mRNA introns.
- The U1 RNA has sequences that are complimentary to sequences near the 5' end of the intron, its binding allows distinguishing the GU at the 5' end of the intron from other randomly placed GU sequences in mRNAs.

- The U2 RNA also recognizes sequences in the intron, in this case near the 3' end.
- The addition of U4, U5 and U6 RNAs forms a complex identified as the spliceosome (snRNA plus ~40 proteins) that removes the intron and joins the two exons together.
- U7 is involved in the production of the correct 3‘ ends of histone mRNA which lacks poly (A) tail.



- An additional mechanism of intron removal is the process of tRNA splicing.
- These introns are spliced by a specific splicing endonuclease that involves a cut-and-paste mechanism.

- In order for tRNA intron removal to occur the tRNA must first be properly folded into its characteristic cloverleaf shape.
- Misfolded precursor tRNAs are not processed which allows the splicing reaction to serve as a control step in the generation of mature tRNAs.



RNA processing: pre-mRNA → mRNA

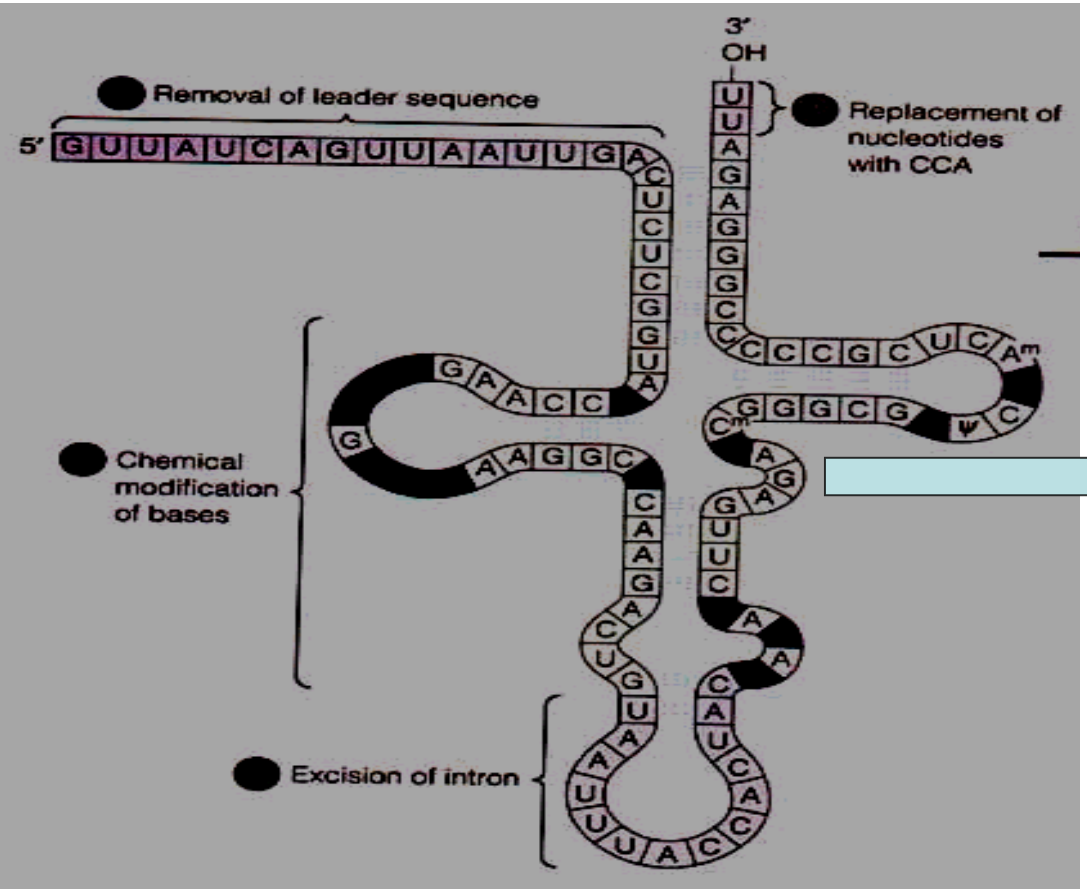
Modifications of tRNA

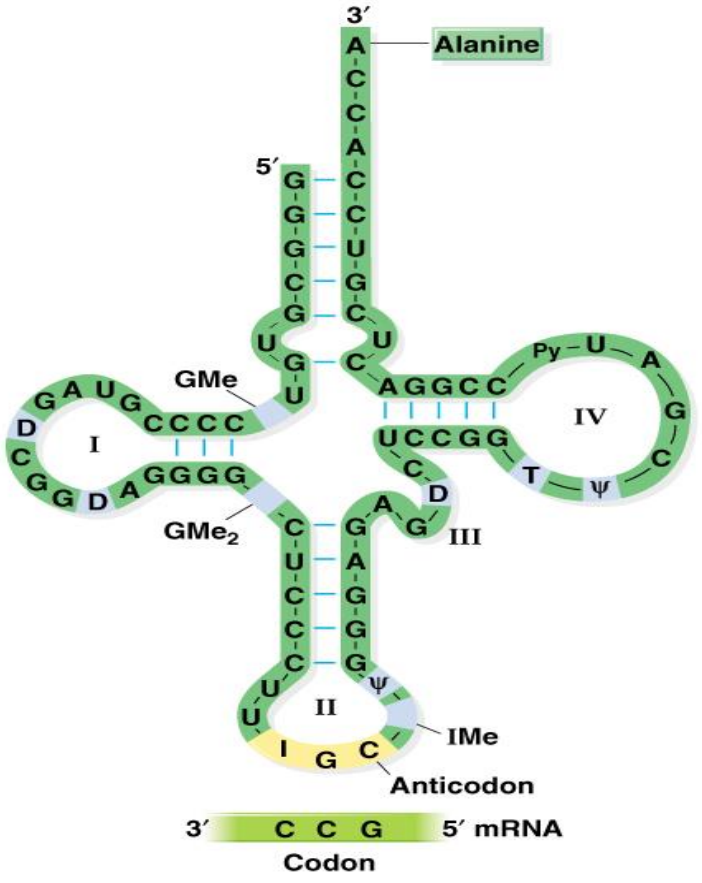
1- Removal of 5' extrasequence

2- Addition of:

- CCA at 3' end
- Anticodon loop

3- Methylation of some bases





rRNA is used to construct ribosomes

- Eukaryotic ribosomal RNA is transcribed in the nucleolus by RNA polymerase I as a single piece of 45S RNA, which is subsequently cleaved to yield 28S rRNA, 18S rRNA, and 5.8S rRNA.
- RNA polymerase III transcribes the 5S rRNA unit from a separate gene. The ribosomal subunits assemble in the nucleolus as the rRNA pieces combine with ribosomal proteins.
- Eukaryotic ribosomal subunits are 60S and 40S. They join during protein synthesis to form the whole 80S ribosome.

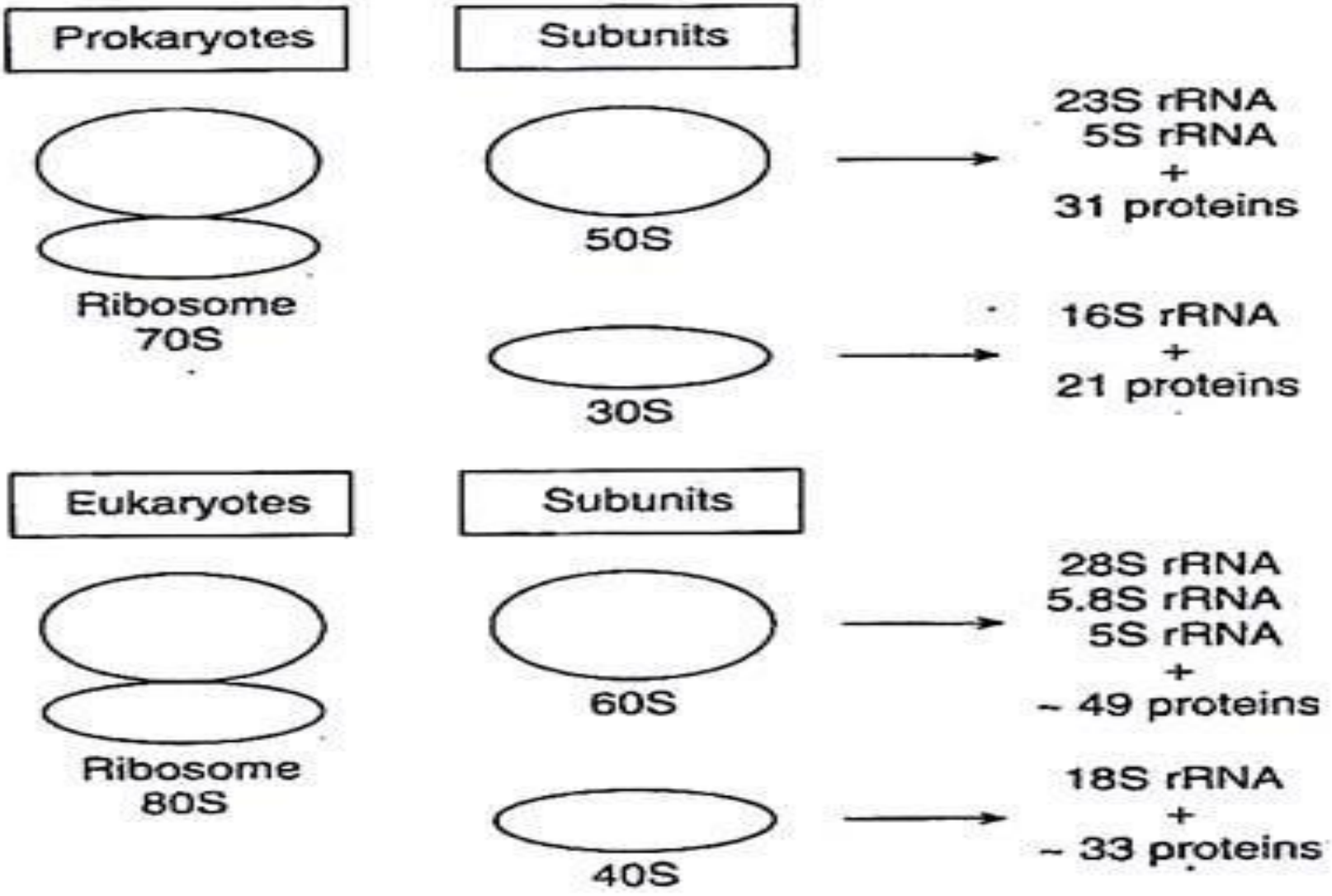


Fig. 2.47: Composition of typical prokaryotic and eukaryotic ribosomes