

Peter Pristaš

**Molecular
biology**

**RNA structure,
organization and synthesis**

Transcription

Life attributes

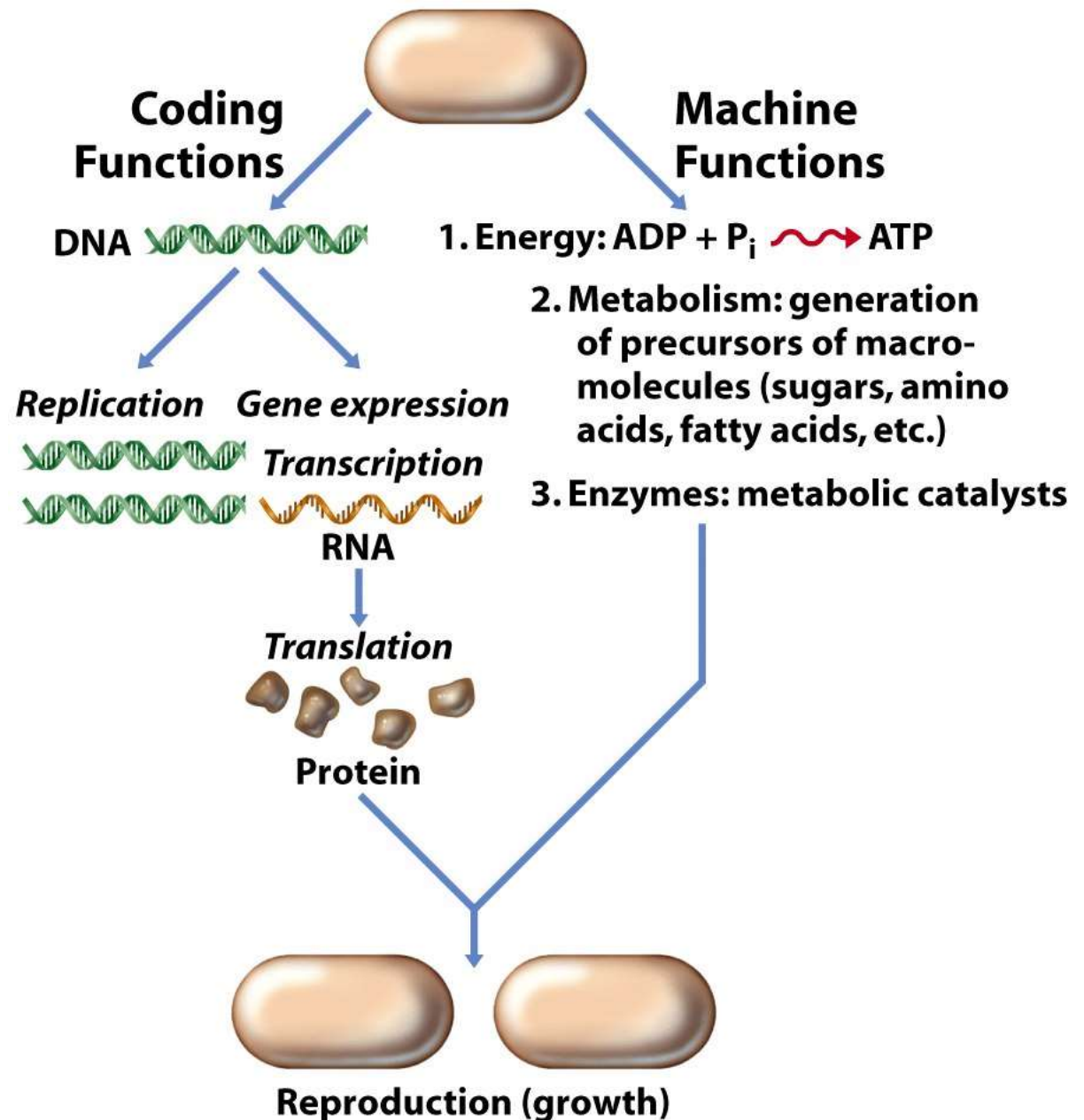


Figure 1-4 Brock Biology of Microorganisms 11/e
© 2006 Pearson Prentice Hall, Inc.

RNA discovery

- DNA and RNA were first described by Friedrich Miescher in 1869. He isolated a phosphorus-containing material from the nuclei of cells found in pus from discarded surgical bandages, and called it “nuclein”. He later found the same material in salmon sperm. Later it was recognized that DNA and RNA are slightly different in structure.

Differences between DNA and RNA

DNA

Usually single molecule

Double stranded

Sugar – deoxyribose

G, A, T, C bases

Stabile

RNA

Multiple molecules

Single stranded

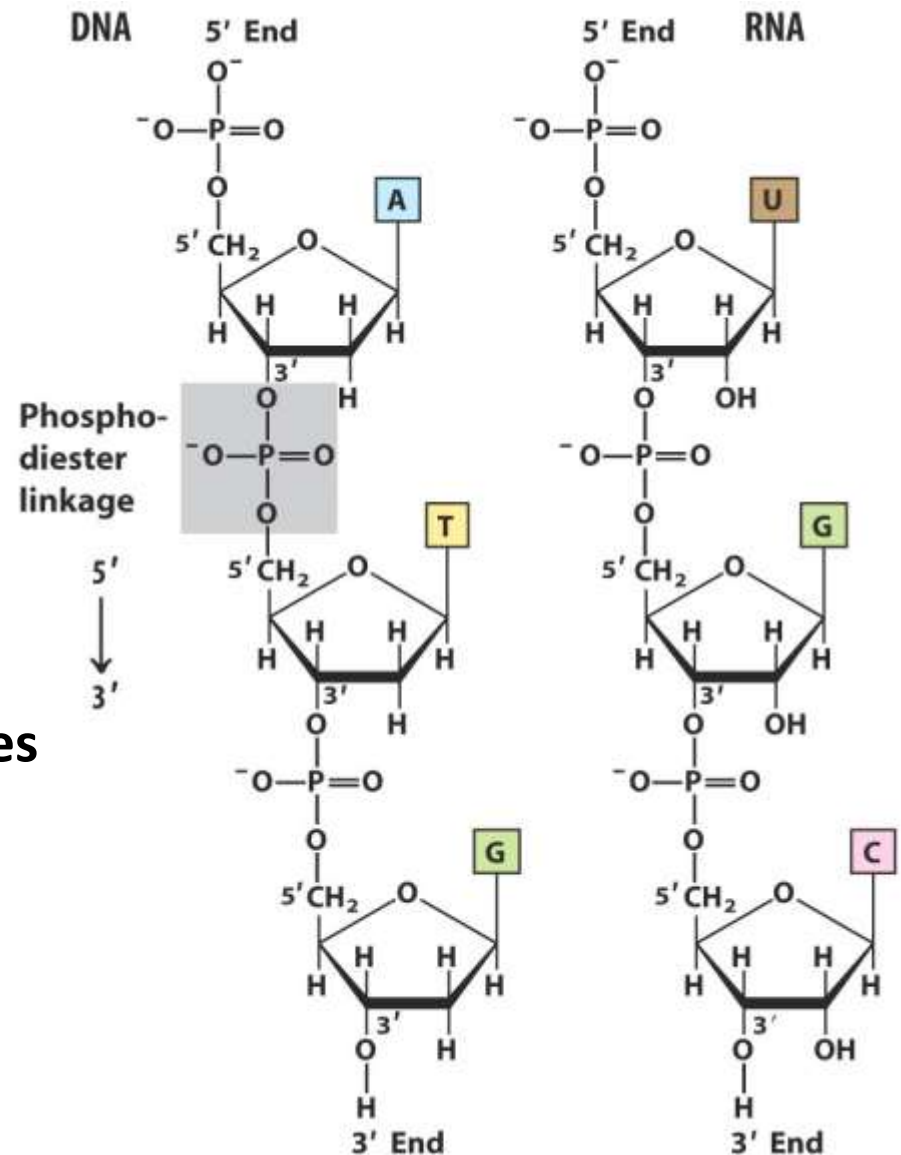
Sugar – ribose

G, A, U, C bases

Labile

What is RNA?

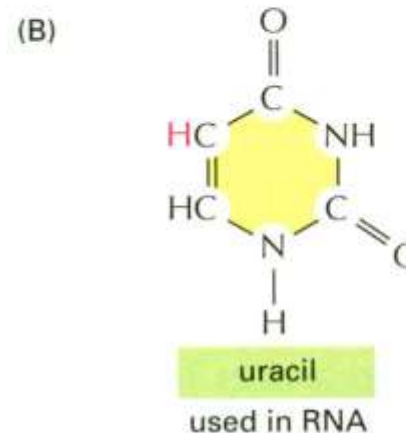
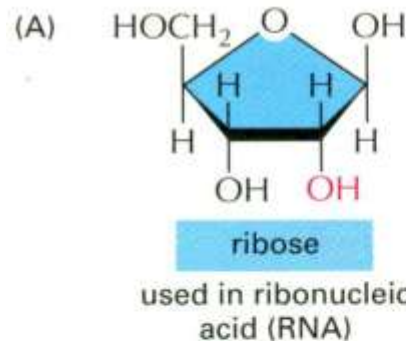
- RNA is a polymer composed of alternating units of ribonucleotides connected through a 3'-5' phosphodiester bond.



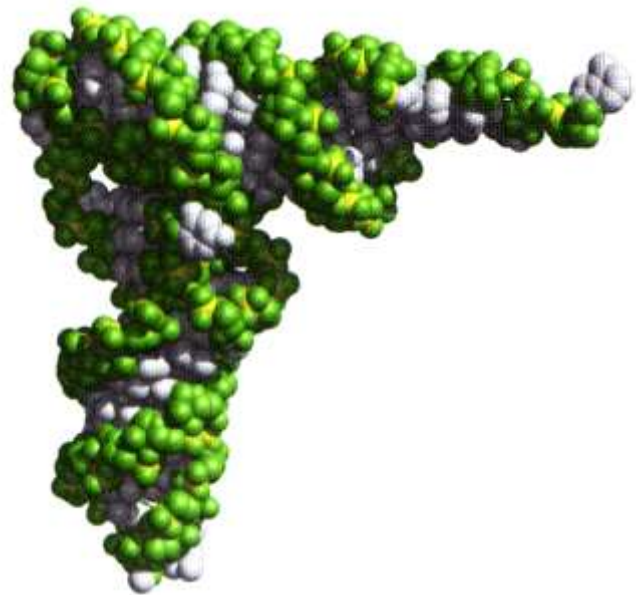
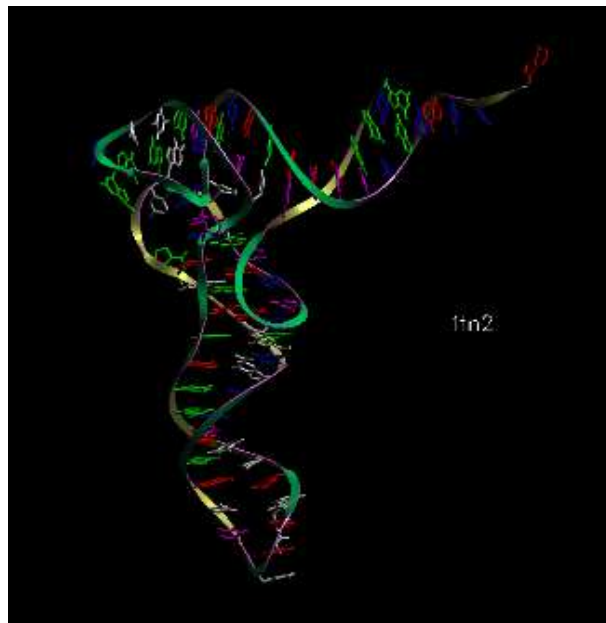
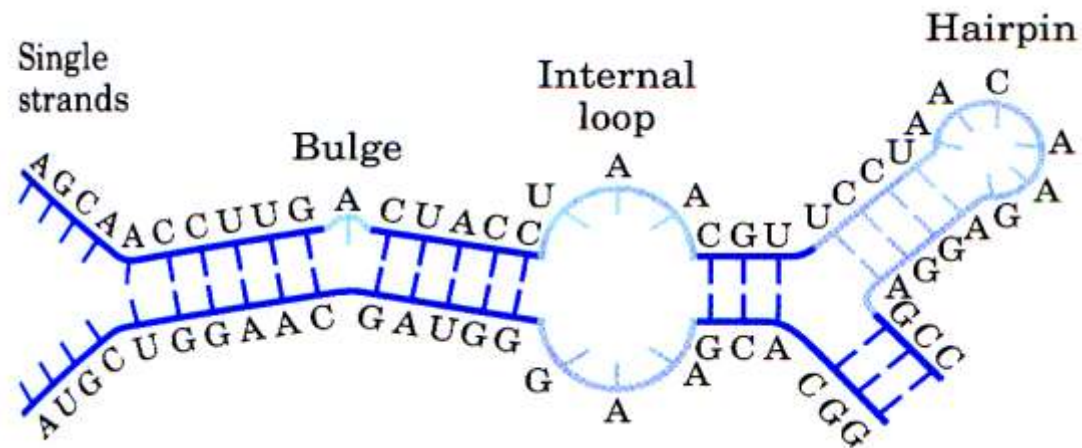
What is RNA

In contrast with DNA,
ribonucleotides contain:

- hydroxyl groups on the 2'-carbon of the ribose sugar
- the base uracil in place of thymine



RNA molecules fold into complex structures



Types of RNA

% total cellular
RNA mass

Ribosomal RNA (rRNA)

85

- the RNA structural component of the ribosome
- in eukaryotes there are 4 major forms: 28S, 18S and 5.8S and 5S
- in prokaryotes there are only 3: 23S, 16S, and 5S

“S” refers to a Svedberg Unit, which is a measure of size based upon the molecular sedimentation rate during ultracentrifugation

Messenger RNA (mRNA)

2

- the RNA that transfers genetic information stored in DNA into a form useable for protein synthesis

Transfer RNA (tRNA)

12

- assists in decoding the information contained within mRNA during translation by recruiting the correct amino acid to the growing peptide chain

Other forms (snRNA, snoRNA)

1

- small nuclear RNAs that participate in RNA processing

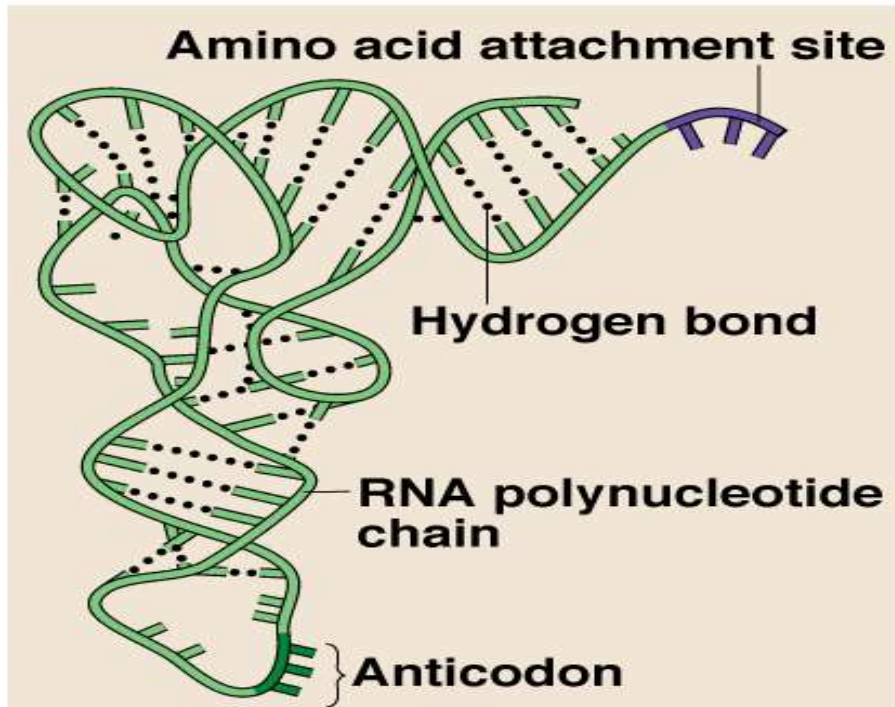
RNA Types

- **tRNA, rRNA, and mRNA participate in the expression of genetic information.**
- **tRNAs link triplet bases in mRNA to specific amino acids.**
- **rRNAs are the major components of ribosomes. Ribosomes facilitate the interaction of proper aa-tRNA with a specific codon and catalyze the addition of amino acids to the growing peptide chain.**
- **mRNAs serve as the intermediary between the gene and the translational machinery.**

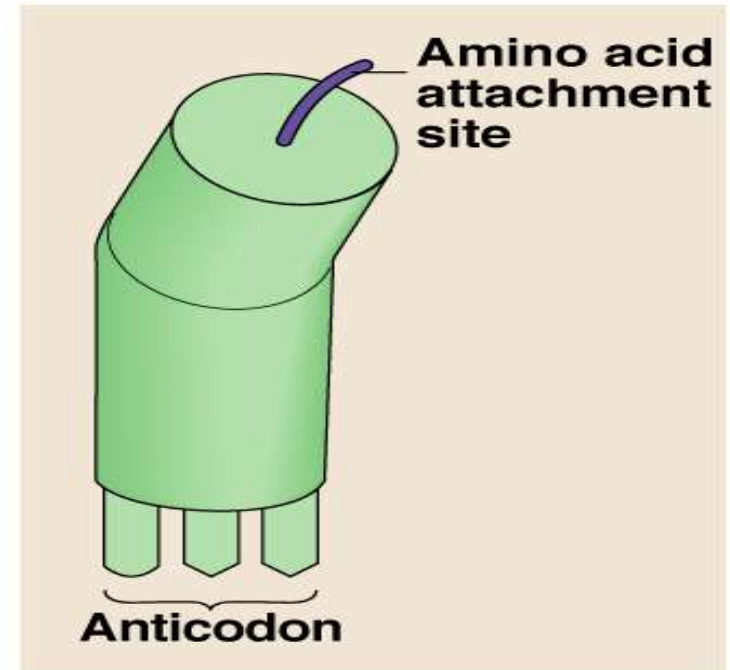
Translation Adapter Molecule

- Generating protein from ribosomes requires change from the nucleic acid to amino acid
- This change is described as translation from the nucleic acid base pair language to the amino acid language
- Crick proposed that some type of adapter molecule was needed to provide the bridge for translation, perhaps a small RNA

tRNA



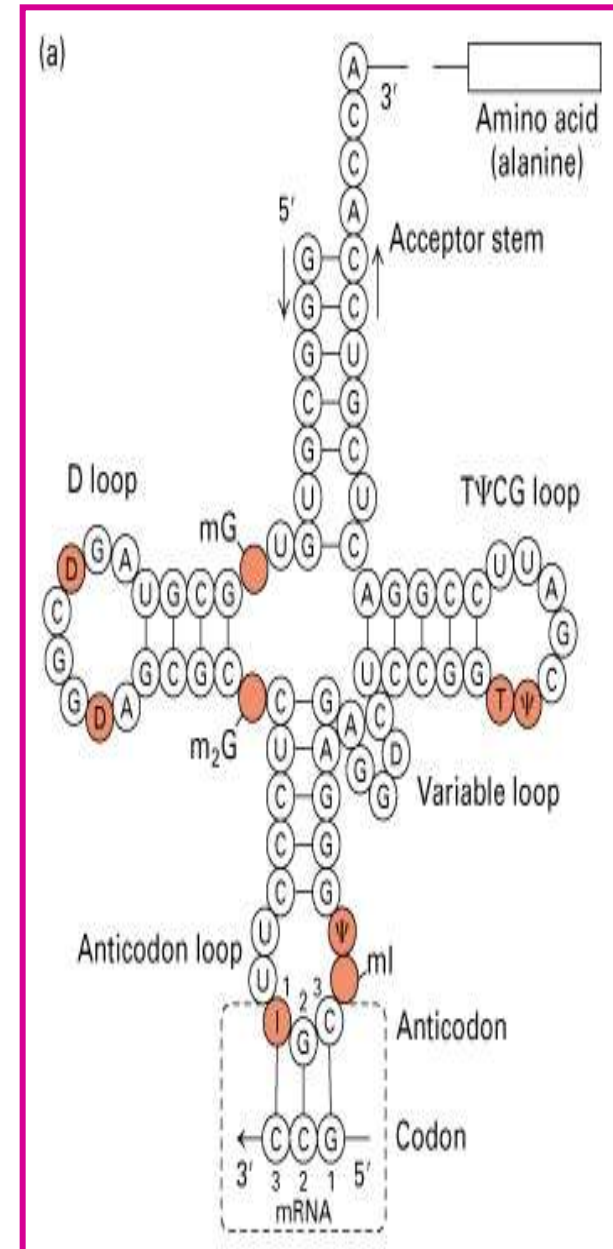
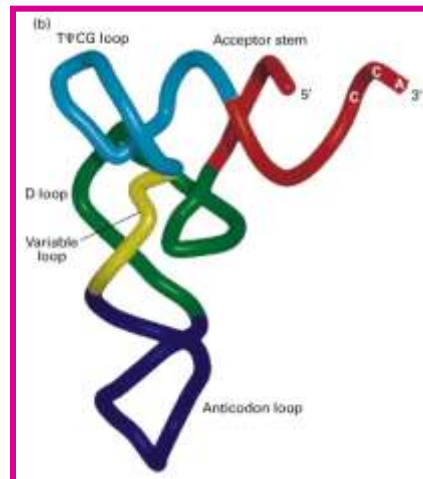
©Addison Wesley Longman, Inc.



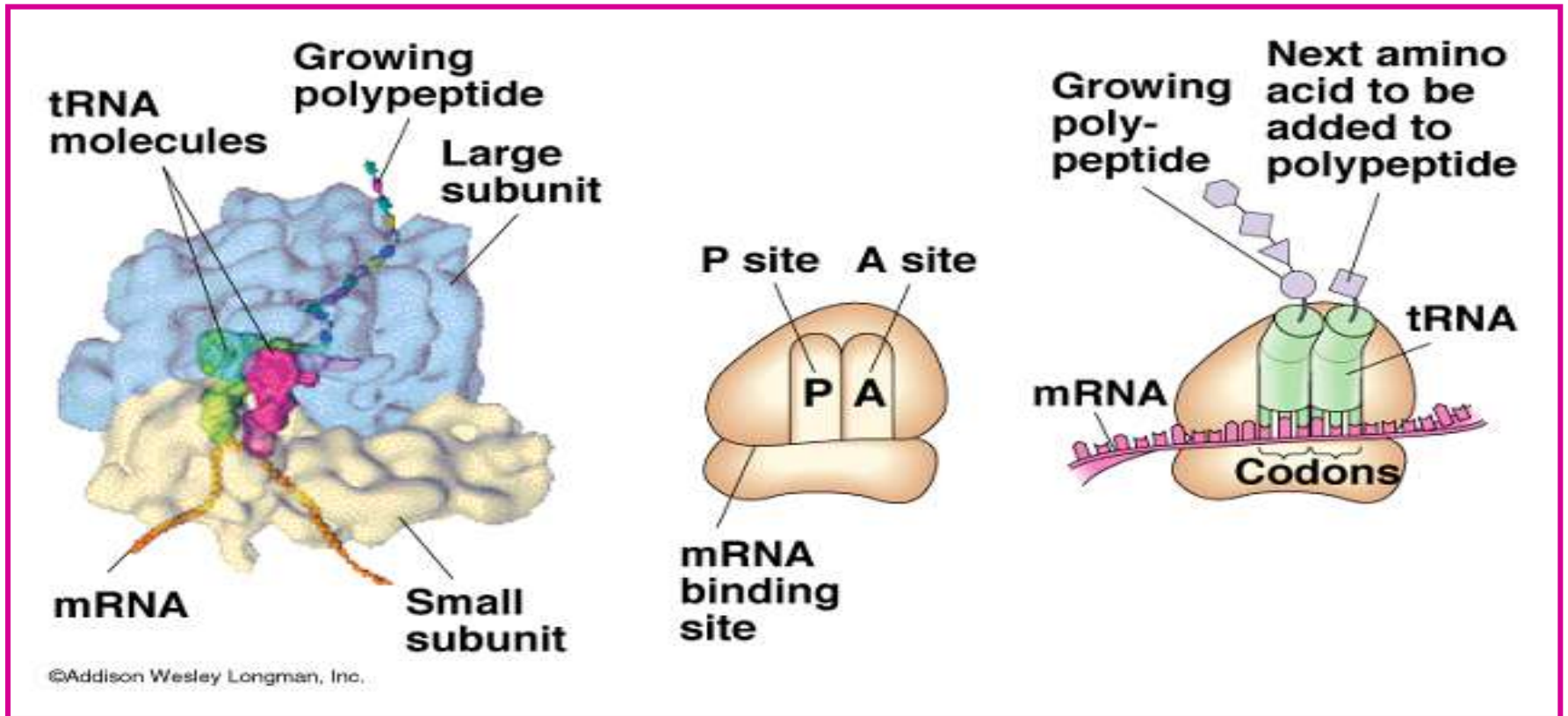
transfer RNA

tRNA structure

- 70-80 nucleotides long
- Post-transcriptional processing
- Bases modifications produce pseudouridine, thymine, inosine
- CCA added to 3' end



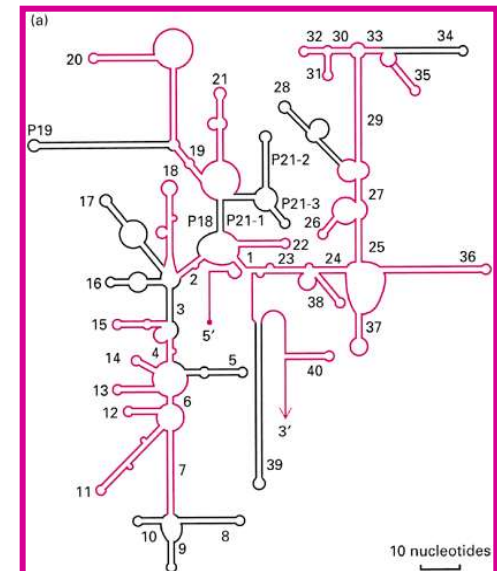
rRNA



ribosomal RNA

Ribosomal RNA

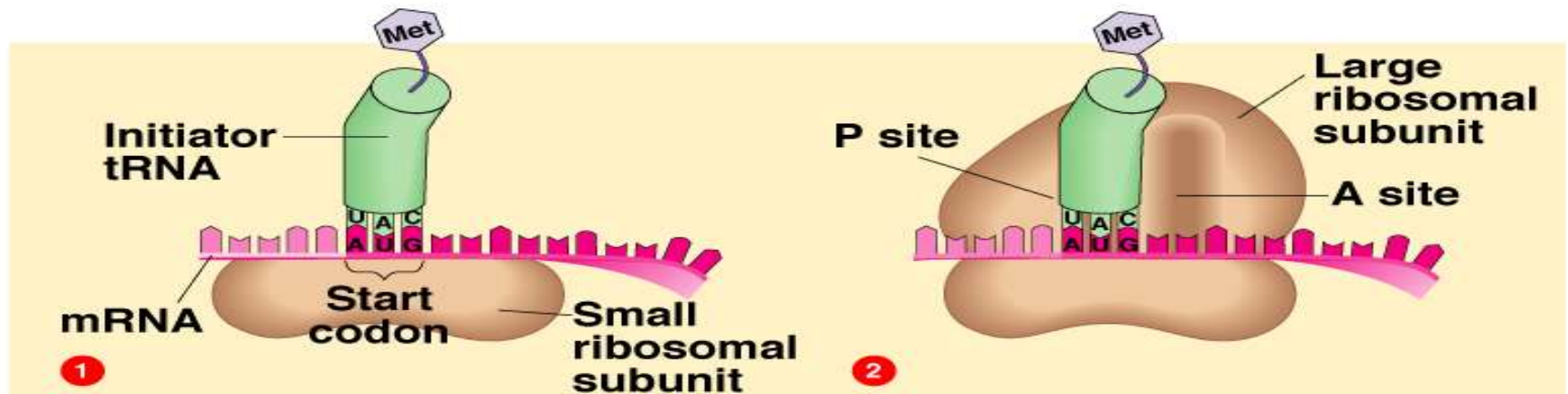
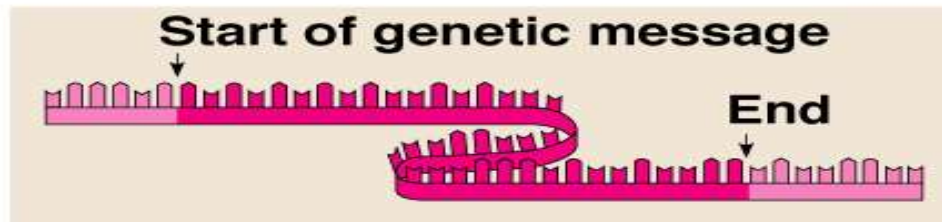
- The two ribosomal subunits both contain ribosomal RNA (rRNA) molecules and a variety of proteins
- rRNAs participate in protein synthesis but do NOT code for proteins
- No translation of rRNA occurs



Information Carrier

- In 1950s and 1960s, the concept that messenger RNA carries information from gene to ribosome developed
- An intermediate carrier was needed as DNA is found in the nucleus, while proteins are made in the cytoplasm
- Some type of molecule must move the information from the DNA in the nucleus to the site of protein synthesis in the cytoplasm

mRNA



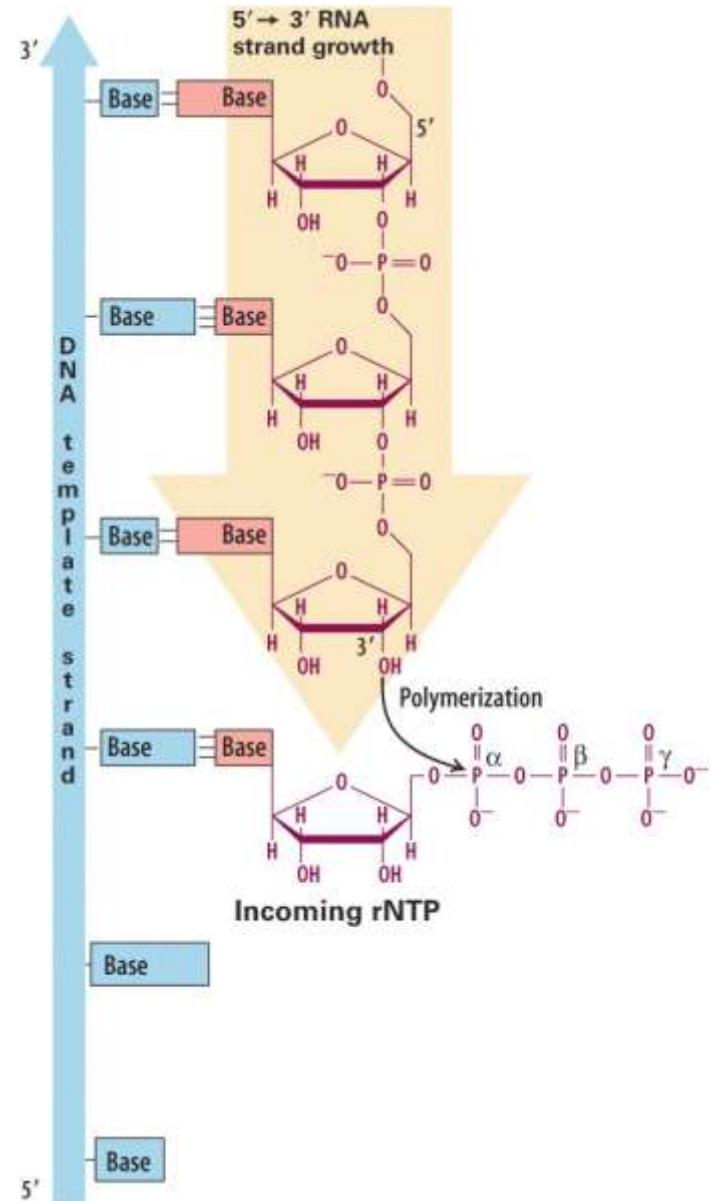
Transcription - RNA synthesis

- Transcription follows the same base-pairing rules as DNA replication
 - Remember **U** replaces T in RNA
 - This base-pairing pattern ensures that the RNA transcript is a faithful copy of the gene
- For transcription to occur at a significant rate, its reaction is enzyme mediated
- The enzyme directing transcription is called **RNA polymerase**

Transcription: RNA Synthesis

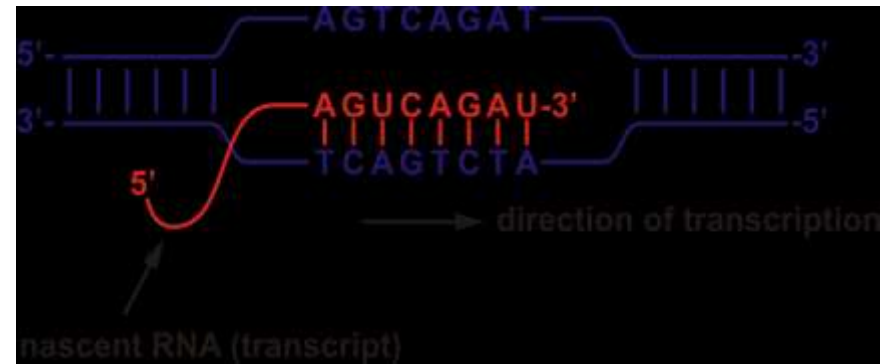
One DNA strand acts as a **template** (read in the 3'→5'), determining the order in which ribonucleoside triphosphate (rNTP) monomers are polymerized to form a complementary RNA strand.

- This polymerization reaction is catalyzed by **RNA polymerase**
- Polymerization involves a **nucleophilic attack** by the 3' oxygen in the growing RNA chain on the α phosphate of the next nucleotide to be added...resulting in a phosphodiester bond (energetically favorable).
- RNA synthesis: 5'→3'

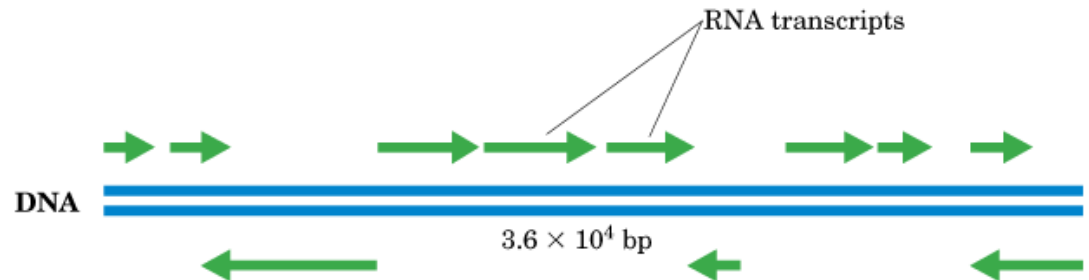


General properties of DNA-dependent RNA polymerases

- (i) **Polarity**: RNA polymerase reads the DNA template in the 3'→5' direction while synthesizing RNA in the 5'→3' direction



- (ii) **DNA template**: Either strand of a DNA double helix can serve as a template for RNA synthesis.



- (iii) **Fidelity**: RNA polymerases do not possess 3'→5' proofreading activities:

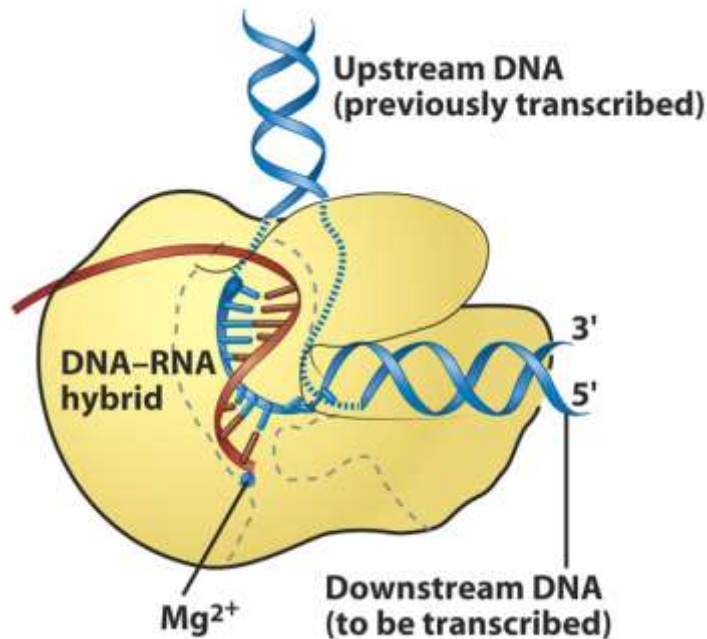
RNA polymerase

error rate : 1 in $10^4 - 10^5$ nt

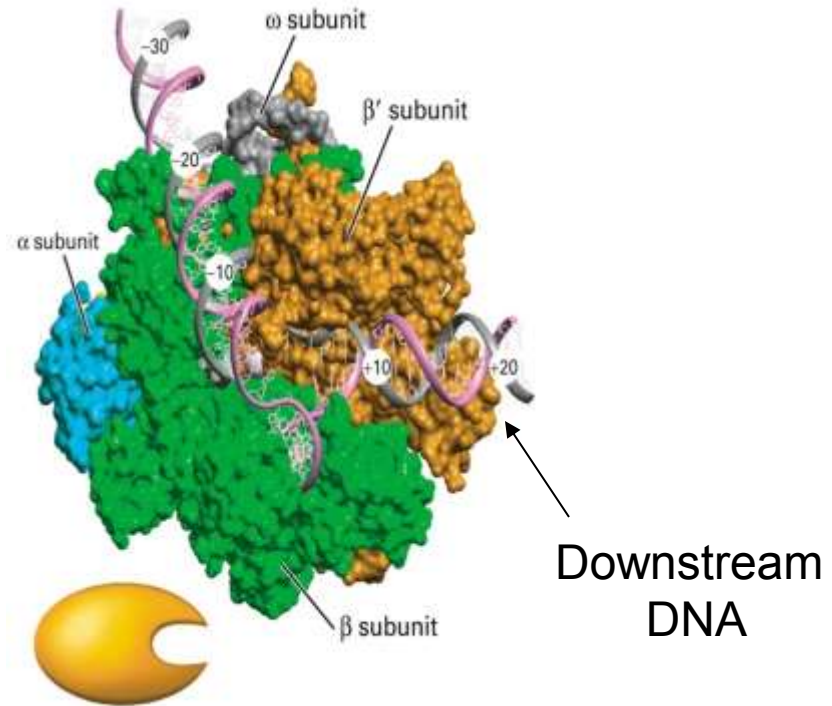
DNA polymerase

error rate : 1 in $10^9 - 10^{10}$ nt

Structure of Prokaryotic RNA Polymerase



Upstream
DNA



-Prokaryotic RNA pol is composed of 5 subunits: 2α , β , β' and ω (**Core enzyme**).

Eukaryotic RNA Polymerases

<u>Polymerase</u>	<u>Products</u>
Polymerase I (Pol I)	rRNAs; 28S, 18S and 5.8S
Polymerase II (Pol II)	mRNA, some small RNAs
Polymerase III (Pol III)	tRNAs and 5S, additional small RNAs

- Unlike the *E. coli* RNA polymerase holoenzyme, each of these require a number of additional proteins called “**transcription factors**” in order to specifically bind to a promoter and initiate transcription.
- **Pol I and III** promoters utilize a small number of ubiquitous transcription factors while **Pol II** uses a large variety of specific ones.

Transcription Phases

Transcription occurs in three phases:

- a. Initiation
- b. Elongation
- c. Termination

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.

(1) Initiation:

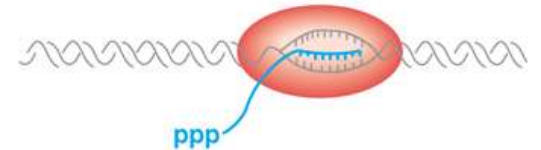
(a) RNA polymerase binds to promoter.



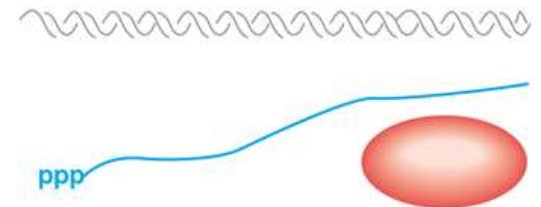
(b) First few phosphodiester bonds form.



(2) Elongation.



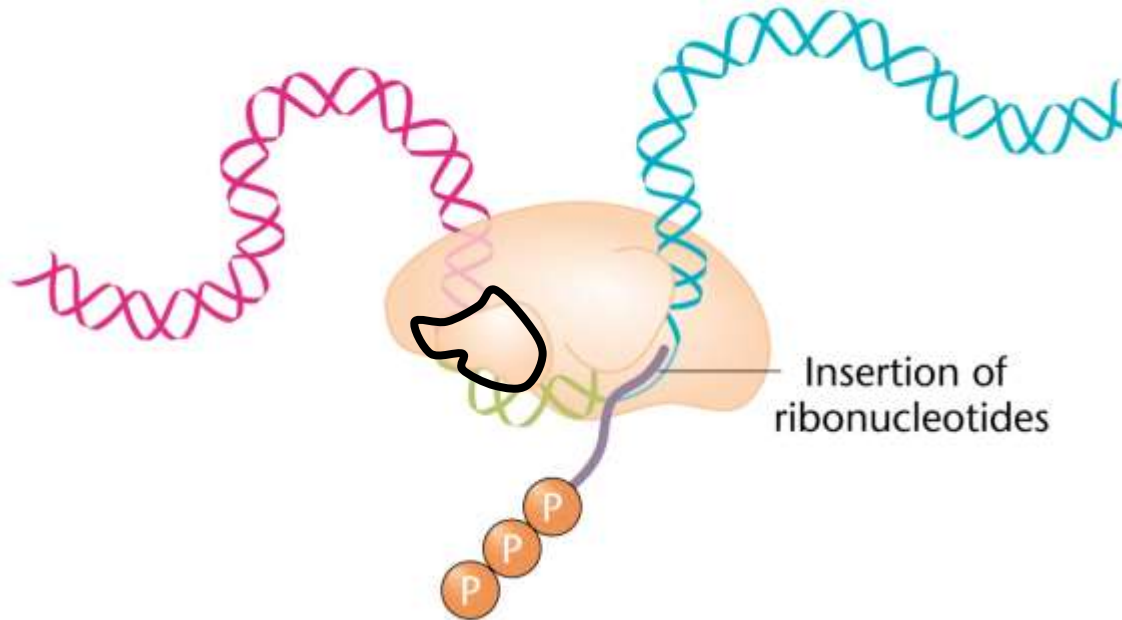
(3) Termination.



Initiation

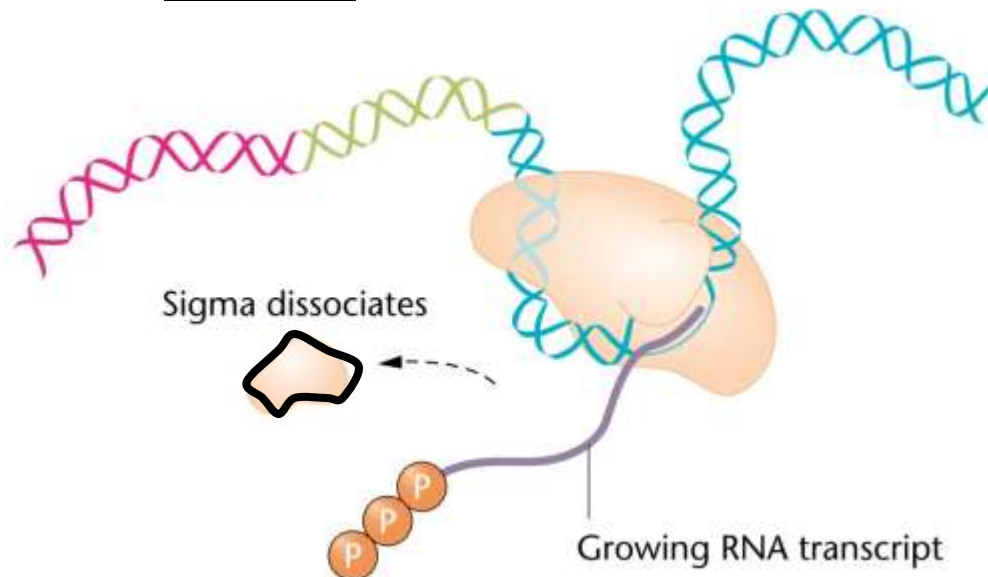
- RNA polymerase recognizes a region, the promoter, which lies just upstream of gene
- Polymerase binds tightly to promoter causing localized separation of the two DNA strands
- Polymerase starts building the RNA chain adding ribonucleotides
- After several ribonucleotides are joined together the enzyme leaves the promoter and elongation begins

(b) Template binding and initiation of transcription



**NO primer
required**

(c) Chain elongation



5' to 3'
50 nt/second

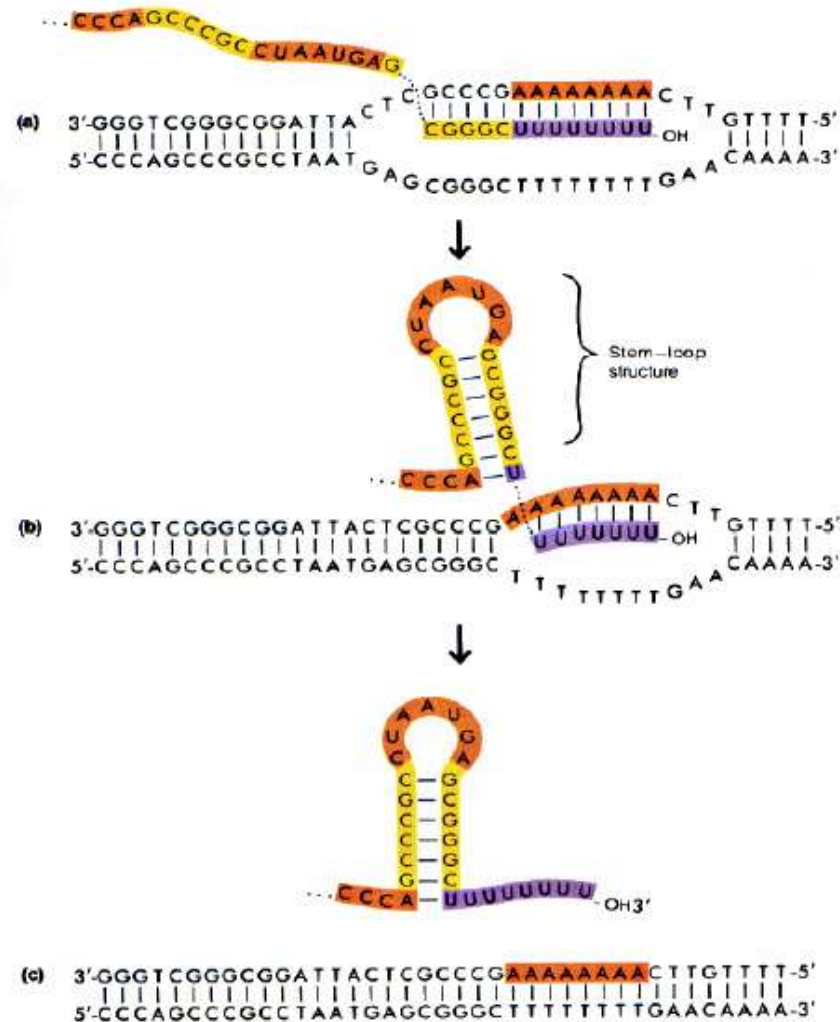
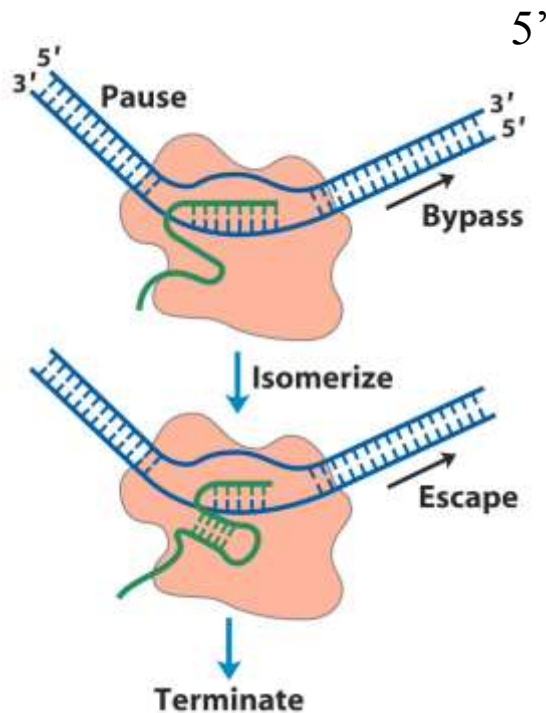
Elongation

- RNA polymerase directs binding of ribonucleotides in the 5' to 3' direction
- Movement of the polymerase along the DNA template causes the “bubble” of separated DNA strands to move also
- As DNA transcription passes, the two DNA strands reform the double helix

Termination

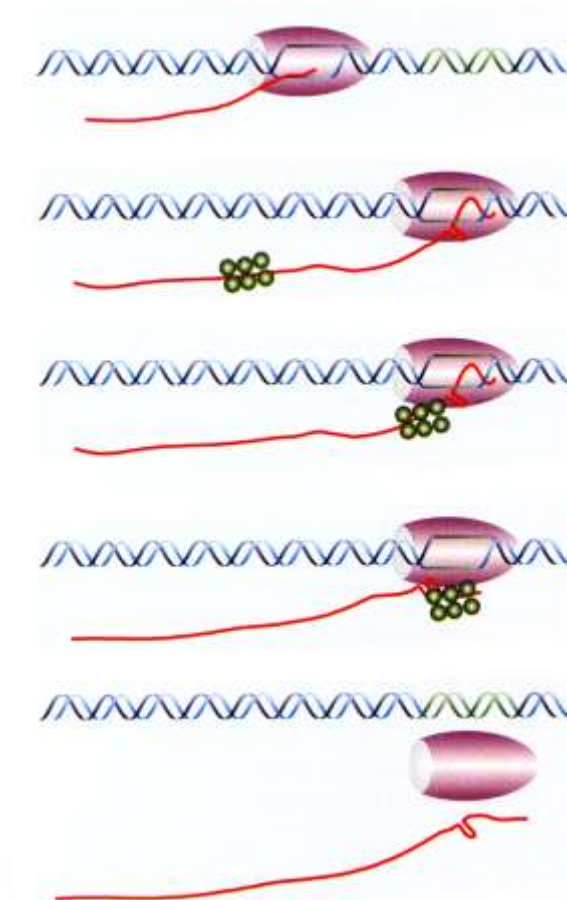
- Analogous to the initiating activity of promoters, there are regions at the other end of genes that serve to terminate transcription
- These terminators work with the RNA polymerase to loosen the association between RNA product and DNA template
- As a result, the RNA dissociates from the RNA polymerase and the DNA and transcription stops

Rho-independent transcriptional termination



Rho-dependent transcriptional termination

1. RNA polymerase transcribes DNA.
2. Rho attaches to its recognition site on RNA.
3. Rho moves along RNA, following the polymerase; RNA polymerase pauses at terminator and rho catches up.
4. Rho unwinds the DNA:RNA hybrid in the transcription bubble.
5. Termination: RNA polymerase, rho, and RNA are released.



Differences Between Transcription and DNA Replication

There are two fundamental differences between transcription and DNA replication

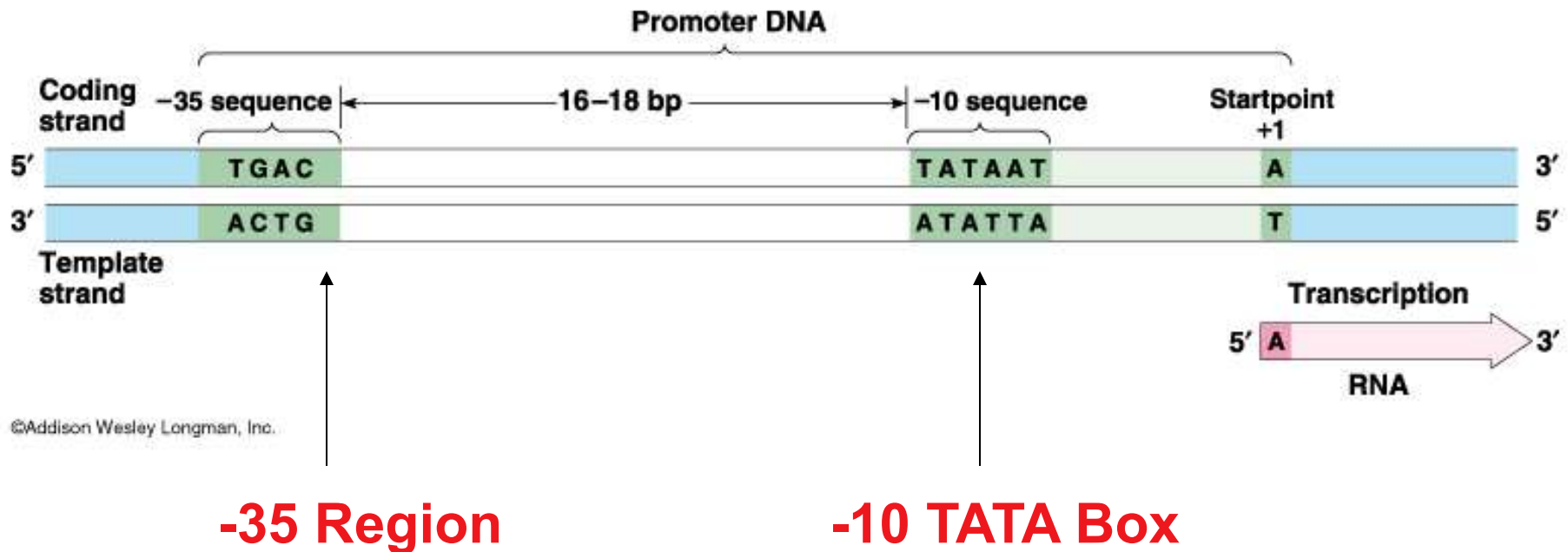
1. RNA polymerase only makes one RNA strand during transcription, it copies only one DNA strand in a given gene
 - This makes transcription **asymmetrical**
 - Replication is **semiconservative**
2. DNA melting is limited and transient during transcription, but the separation is permanent in replication

Transcription Landmarks

- RNA sequences are written 5' to 3', left to right
- Translation occurs 5' to 3' with ribosomes reading the message 5' to 3'
- Genes are written so that transcription proceeds from left to right
- The gene's promoter area lies just before the start area, said to be **upstream** of transcription
- Genes are therefore said to lie **downstream** of their promoters

Prokaryotic Promoter Lies Just Upstream (5') of Transcribed Region

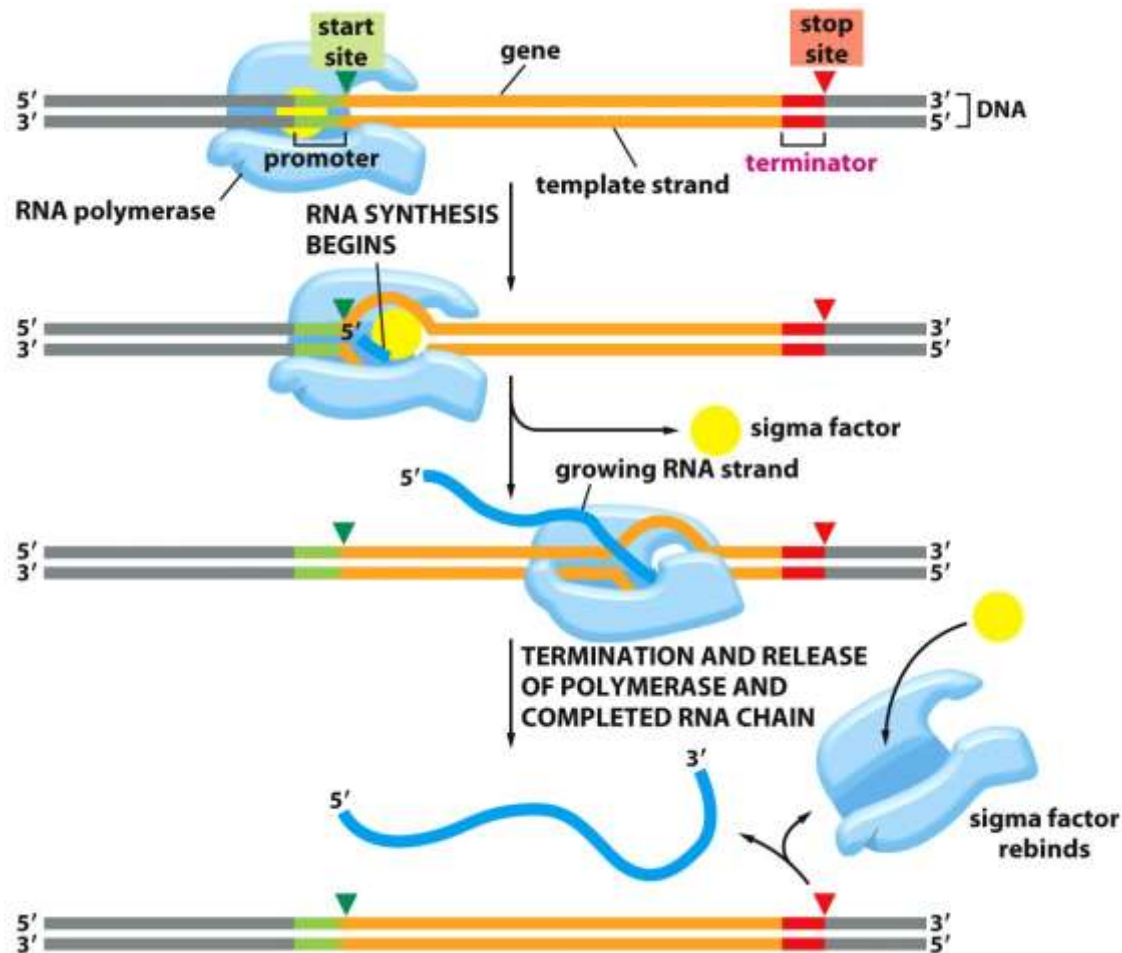
Two Consensus sequences



Prokaryotic promoter elements

	UP element	-35 Region	Spacer	-10 Region	Spacer	RNA start
Consensus sequence	NNAAA ^{AA} _{TT} -A ^T -T ^T TTNNA ^A ANN ^N	TTGACA	N ₁₇	TATAAT	N ₆	+1
<i>rrnB</i> P1	AGAAAATTATTTTAAATTCCT	GTGTCA	N ₁₆	TATAAT	N ₈	A
<i>trp</i>		TTGACA	N ₁₇	TTAACT	N ₇	A
<i>lac</i>		TTTACA	N ₁₇	TATGTT	N ₆	A
<i>recA</i>		TTGATA	N ₁₆	TATAAT	N ₇	A
<i>araBAD</i>		CTGACG	N ₁₈	TACTGT	N ₆	A

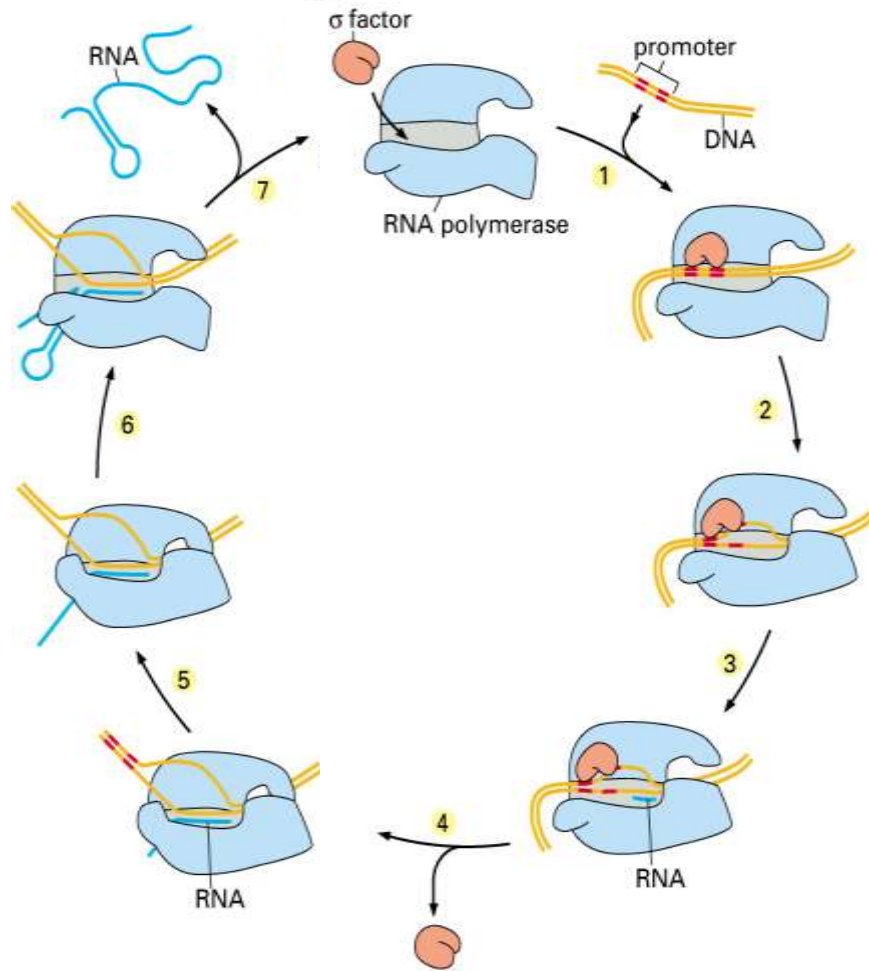
Promoter and terminator sequences of a gene tell the RNA polymerase where to start and stop transcription



Prokaryotic RNA polymerase and transcription

Sigma factor (σ): subunit of RNA polymerase that recognizes and binds to the promoter

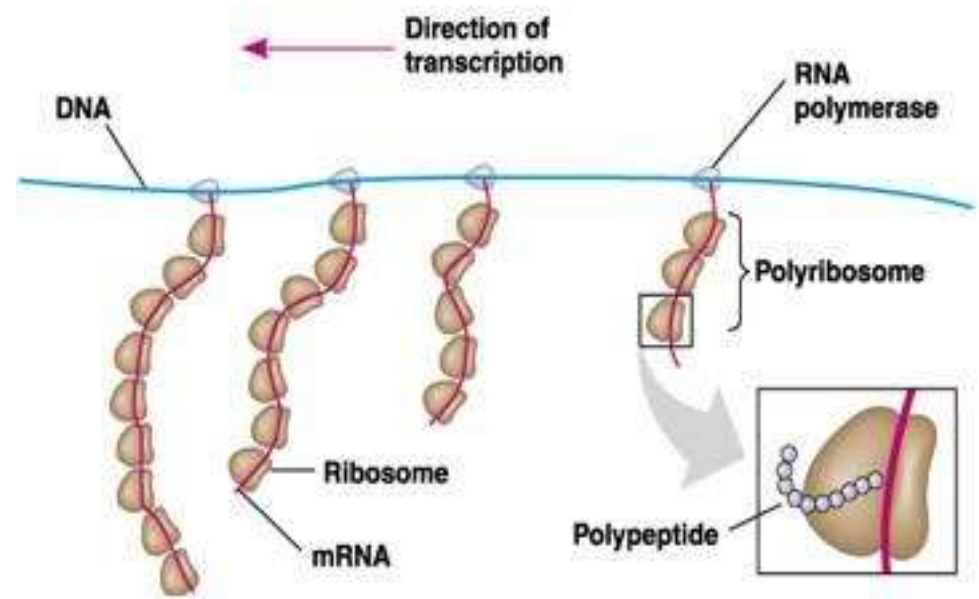
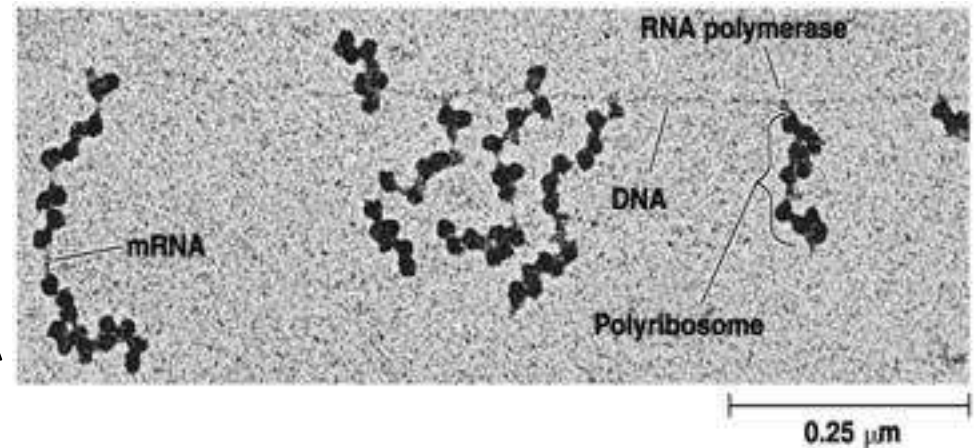
σ + polymerase = **holoenzyme**



1. Holoenzyme is formed and σ factor interacts and binds to the promoter region
2. Polymerase unwinds DNA
3. Transcription begins
4. After ~10 nt are synthesized, σ factor is released and polymerase undergoes conformational change. Elongation mode begins; RNA strand exits polymerase
5. Elongation
6. Polymerase encounters termination signal
7. Full length RNA is released

Simultaneous Transcription of a Gene by Multiple Molecules of RNA Polymerase

Prokaryotes only



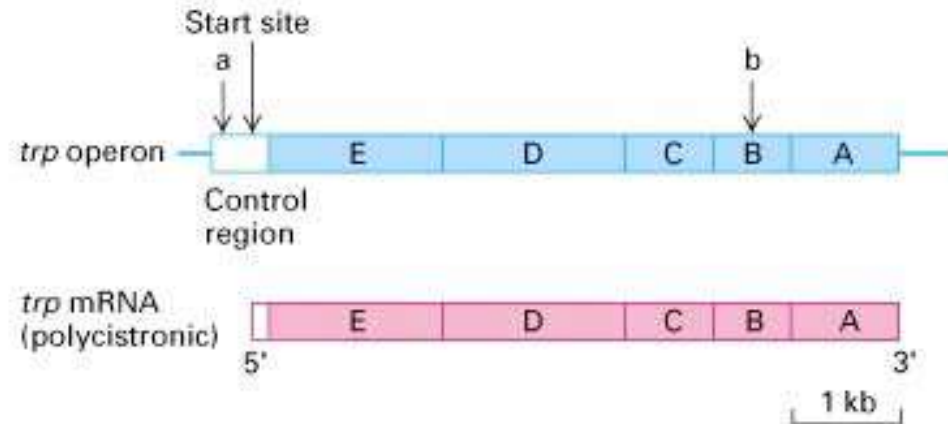
Reprinted with permission from O. L. Miller, Jr., B. A. Hanikalo, and C. A. Thomas, Jr.,
Science 169 (1970): 382. Copyright © 1970 American Association for the Advancement of Science.
©1999 Addison Wesley Longman, Inc.

Gene Structure: Polycistronic vs monocistronic

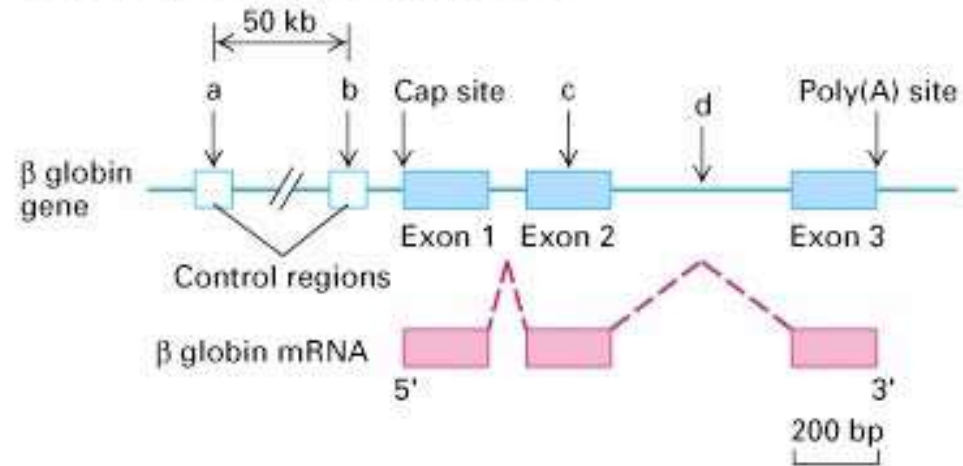
- **Cistron:** an old name for a gene
- **Polycistronic:** 1 promoter directs synthesis of 1 mRNA that can be translated to more than one polypeptide

— Prokaryotic genes

(a) Prokaryotic polycistronic transcription unit



(b) Eukaryotic simple transcription unit



- **Monocistronic:** 1

Transcription in eukaryotes

- Multiple RNA polymerases
- Eukaryotic genes and promoters
- Processing of primary transcript
 - Capping
 - Splicing
 - Polyadenylation

Role of the three RNA Polymerases

RNA Polymerase	Cellular RNAs Synthesized	Mature RNA (Vertebrate)
I	Large rRNA precursor	28S, 18S, and 5.8S rRNAs
II	hnRNAs	mRNAs
	snRNAs	snRNAs
III	5S rRNA precursor	5S rRNA
	tRNA precursors	tRNAs
	U6 snRNA (precursor?)	U6 snRNA
	7SL RNA (precursor?)	7SL RNA
	7SK RNA (precursor?)	7SK RNA

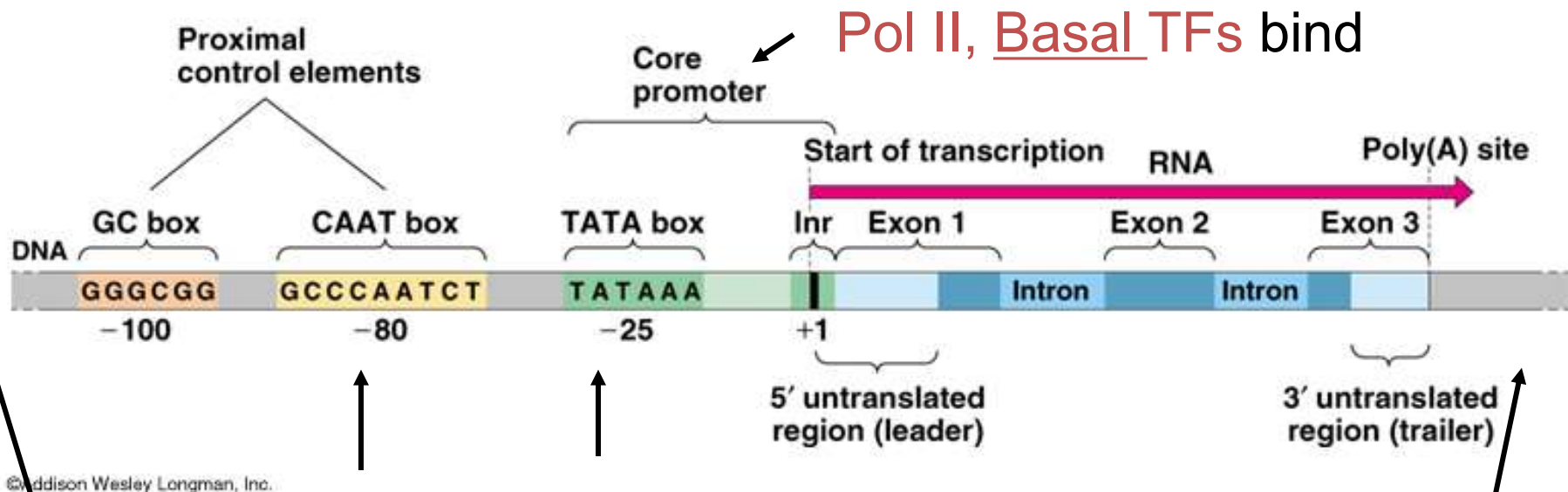
Role of the three RNA Polymerases

TABLE 12-1 The Subunits of RNA Polymerases

Prokaryotic		Eukaryotic		
Bacterial	Archaeal	RNAP I	RNAP II	RNAP III
Core	Core	(Pol I)	(Pol II)	(Pol III)
β'	A'/A''	RPA1	RPB1	RPC1
β	B	RPA2	RPB2	RPC2
α^I	D	RPC5	RPB3	RPC5
α^{II}	L	RPC9	RPB11	RPC9
ω	K	RPB6	RPB6	RPB6
	[+6 others]	[+9 others]	[+7 others]	[+11 others]

Adapted, with permission, from Ebright R.H. 2000. *J. Mol. Biol.* 304: 687–698, Fig. 1, p. 688. © Elsevier.
The subunits in each column are listed in order of decreasing molecular weight.

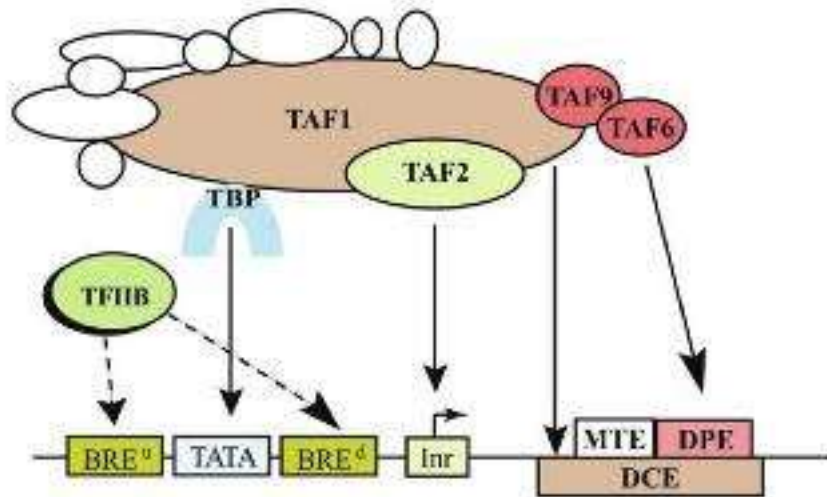
Anatomy of a Eukaryotic Gene



CAAT Box TATA Box

Cis-regulatory Elements may be located thousands of bases away; Regulatory TFs bind.

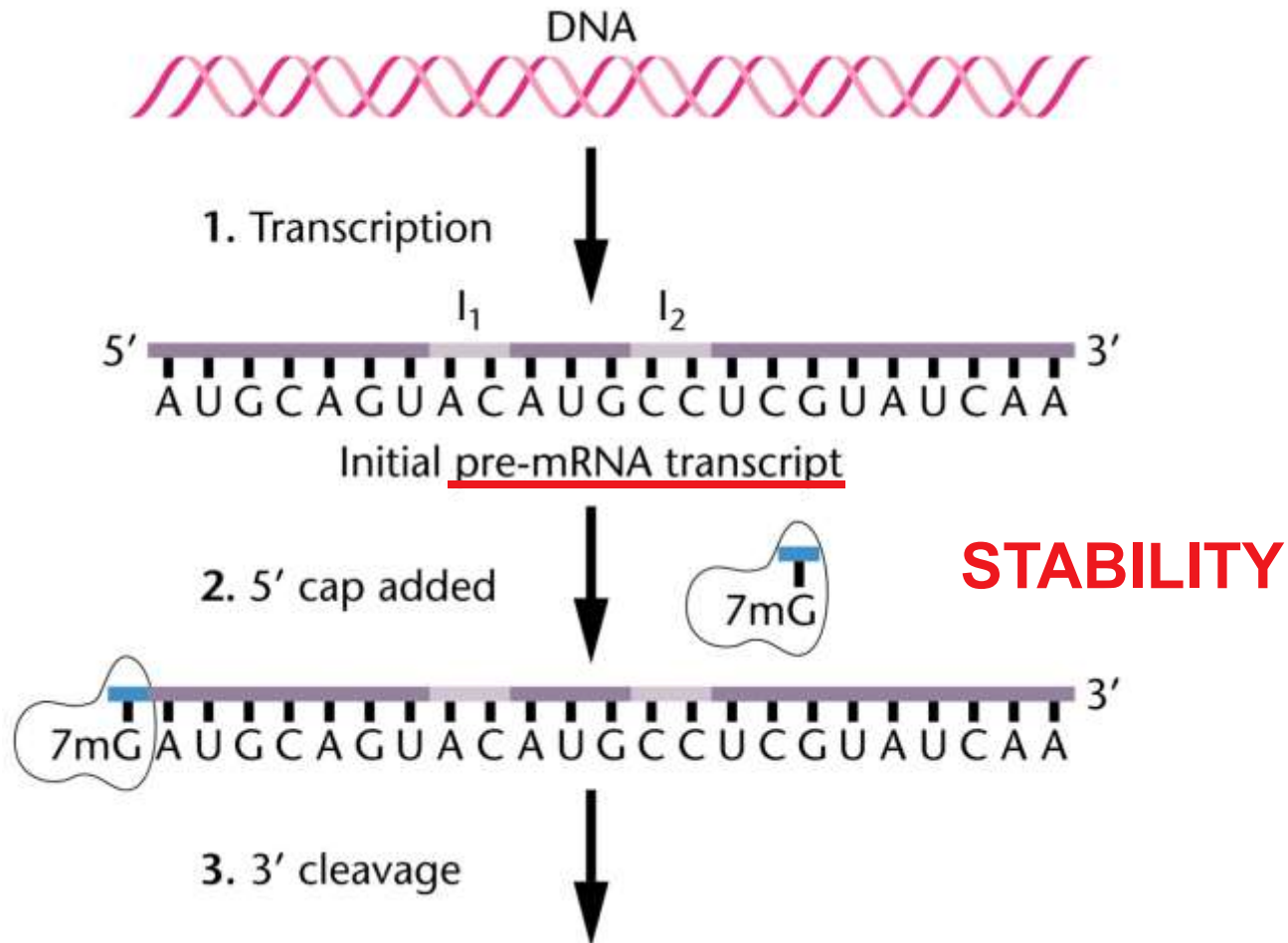
Recognition of core promoter elements by transcription factors

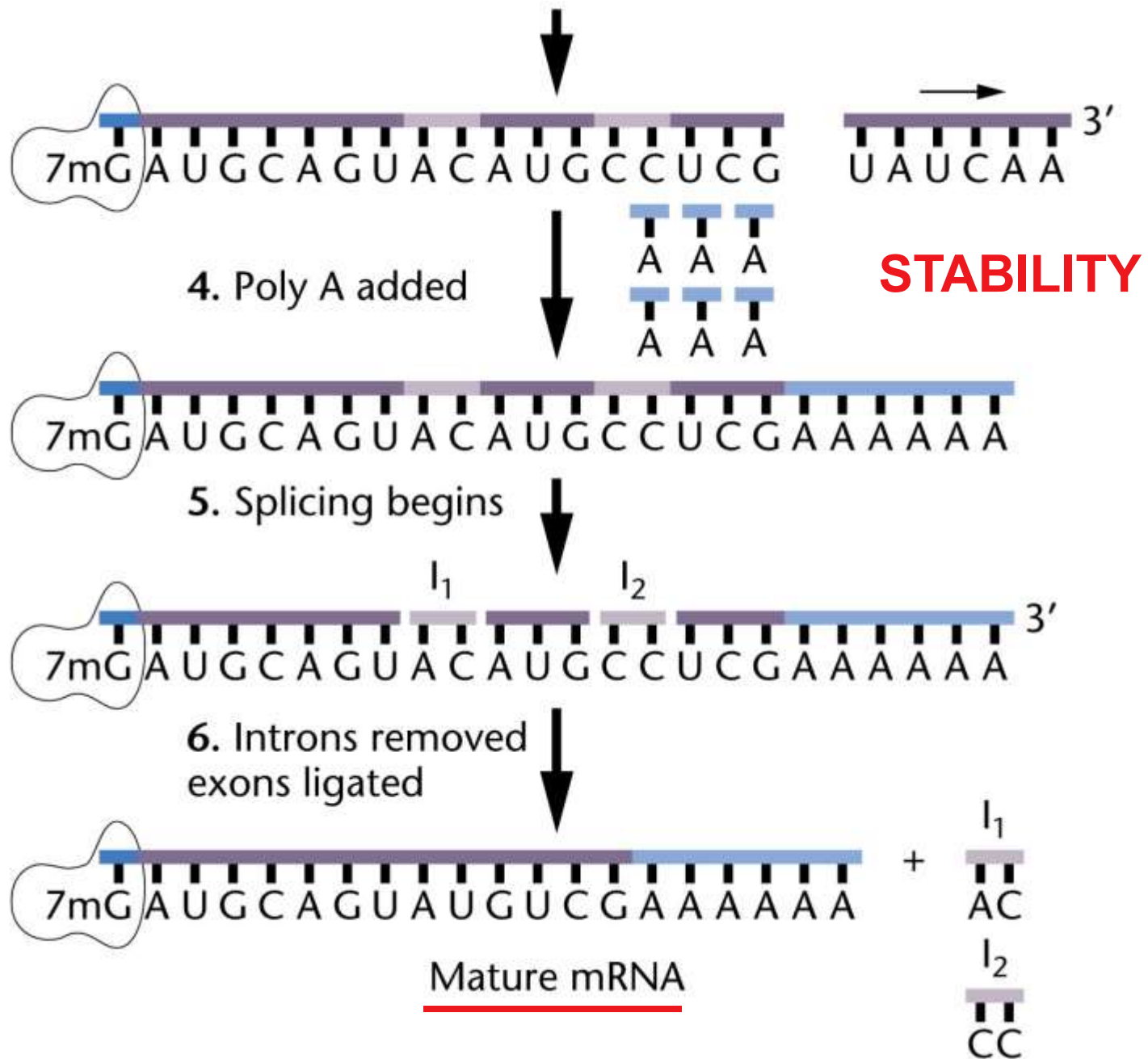


Selectivity of promoters determines which transcription factors are recruited to activate gene transcription.

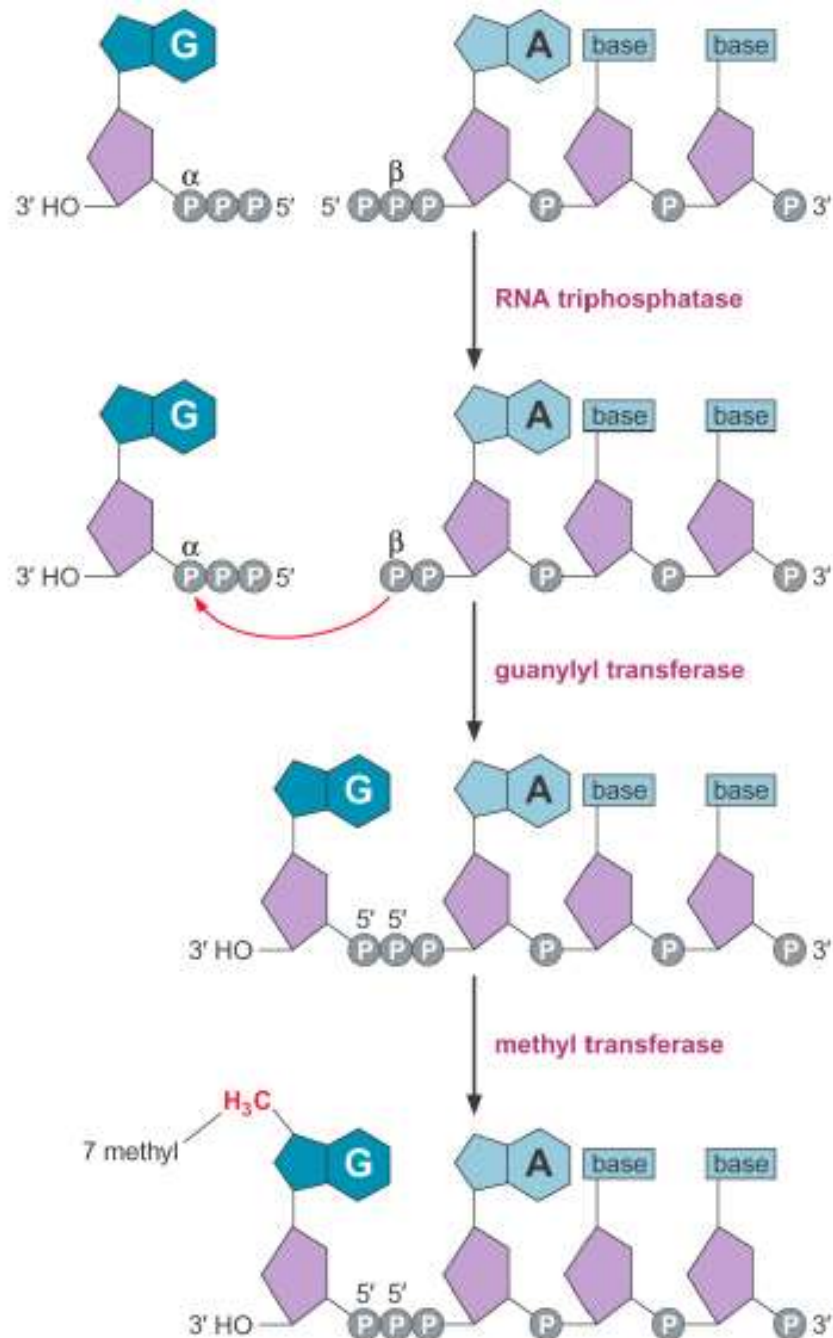
Core Promoter Element	Position	Consensus Sequence (5' to 3')	Bound Protein
BRE ^u	-38 to -32	(G/C)(G/C)(G/A)CGCC	TFIIB
TATA	-31 to -24	TATA(A/T)A(A/T)(A/G)	TBP
BRE ^d	-23 to -17	(G/A)T(T/G/A)(T/G)(G/T)(T/G)(T/G)	TFIIB
Inr	-2 to +5	PyPyAN(T/A)PyPy	TAF1/TAF2
MTE	+18 to +29	C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C)	n.a.
DPE	+28 to +34	(A/G)G(A/T)CGTG	TAF6/TAF9
DCE	3 subelements: +6 to +11 +16 to +21 +30 to +34	core sequence: S _I CTTC S _{II} CTGT S _{III} AGC	TAF1

RNA Processing in Eukaryotes





RNA processing - 5' end capping

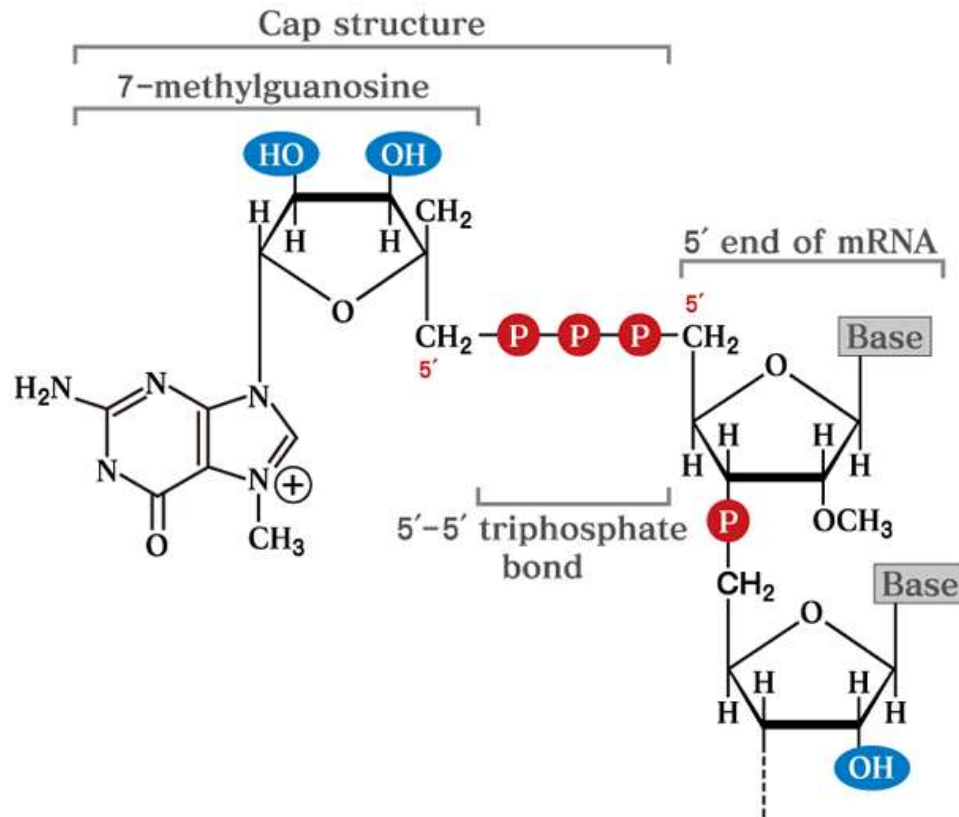


Modification of the 5'-ends of eukaryotic mRNAs is called capping.

The cap consists of a methylated GTP linked to the rest of the mRNA by a 5' to 5' triphosphate "bridge" (Cap Structure). Capping occurs very early during the synthesis of eukaryotic mRNAs, even before mRNA molecules are finished being made by RNA polymerase II.

Capped mRNAs are very efficiently translated by ribosomes to make proteins.

RNA processing - 5' end capping

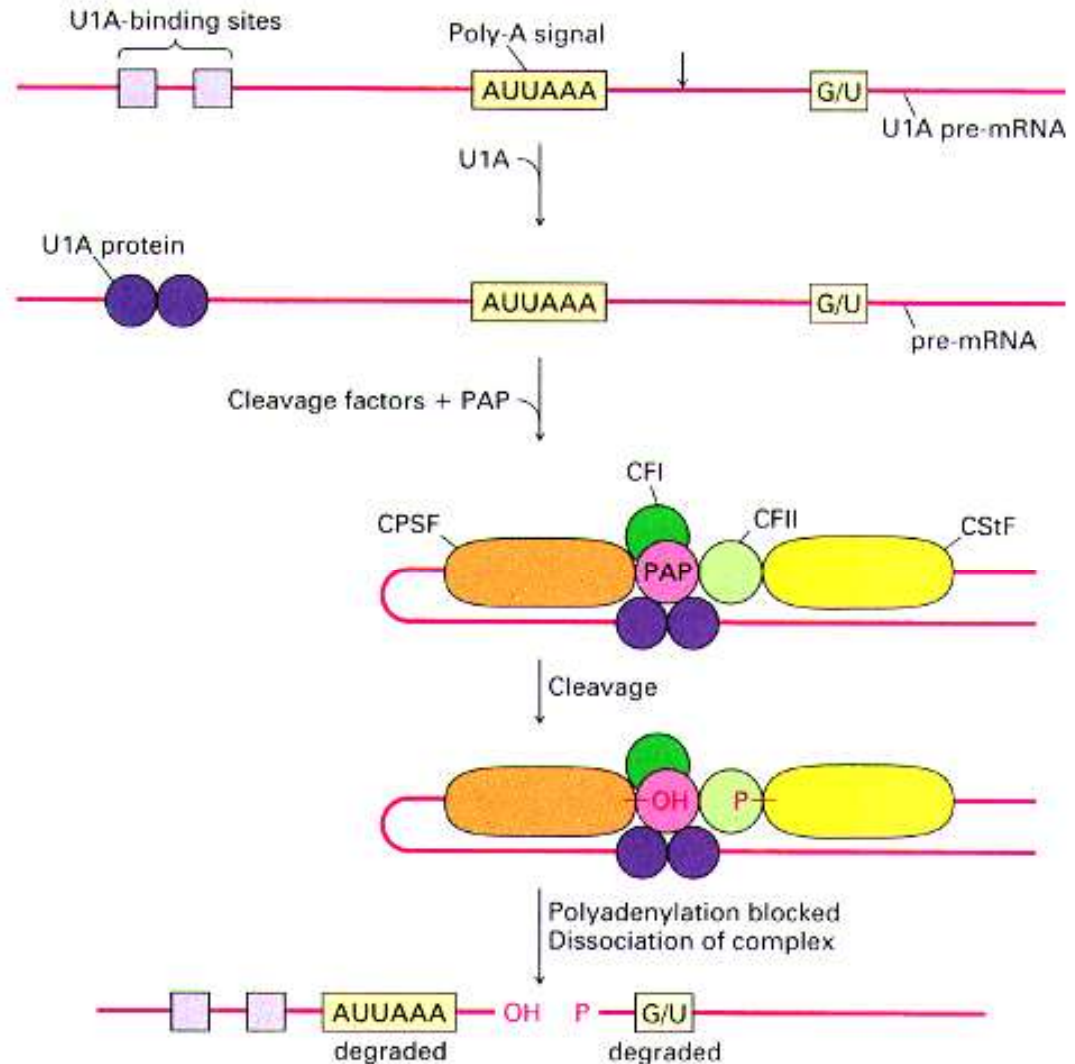


Termination – polyA addition

Eukaryotic protein genes contain a poly-A signal (the sequence AAUAAA about 11-30 nucleotides upstream to where the tail is added) located downstream of the last exon.

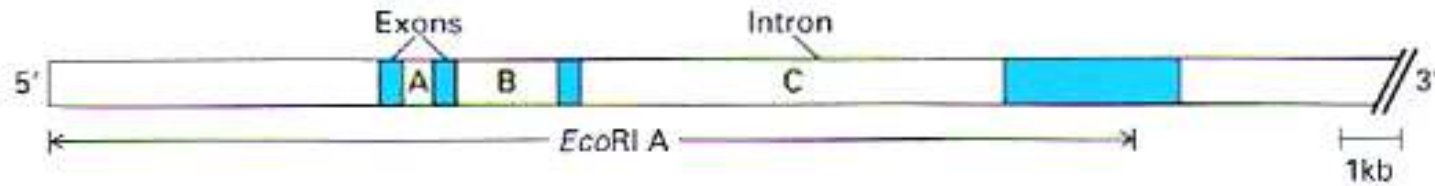
Transcription often terminates at 0.5 - 2 kb downstream of the poly-A signal.

AAUAAA is recognized by an endonuclease that cuts the RNA, allowing the tail to be added by a specific enzyme: polyA polymerase.

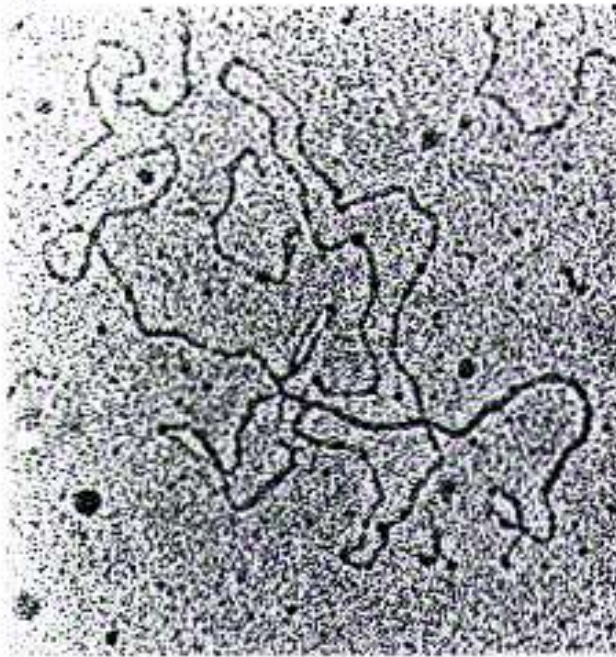


Pre-mRNA has introns

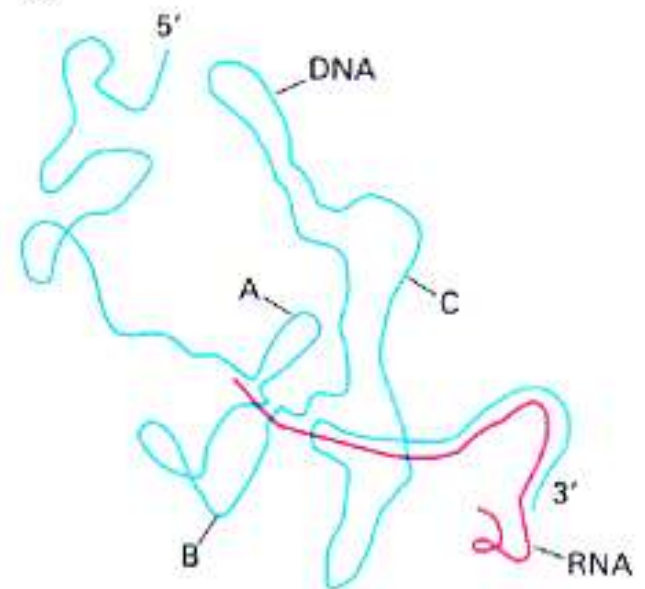
(a)



(b)

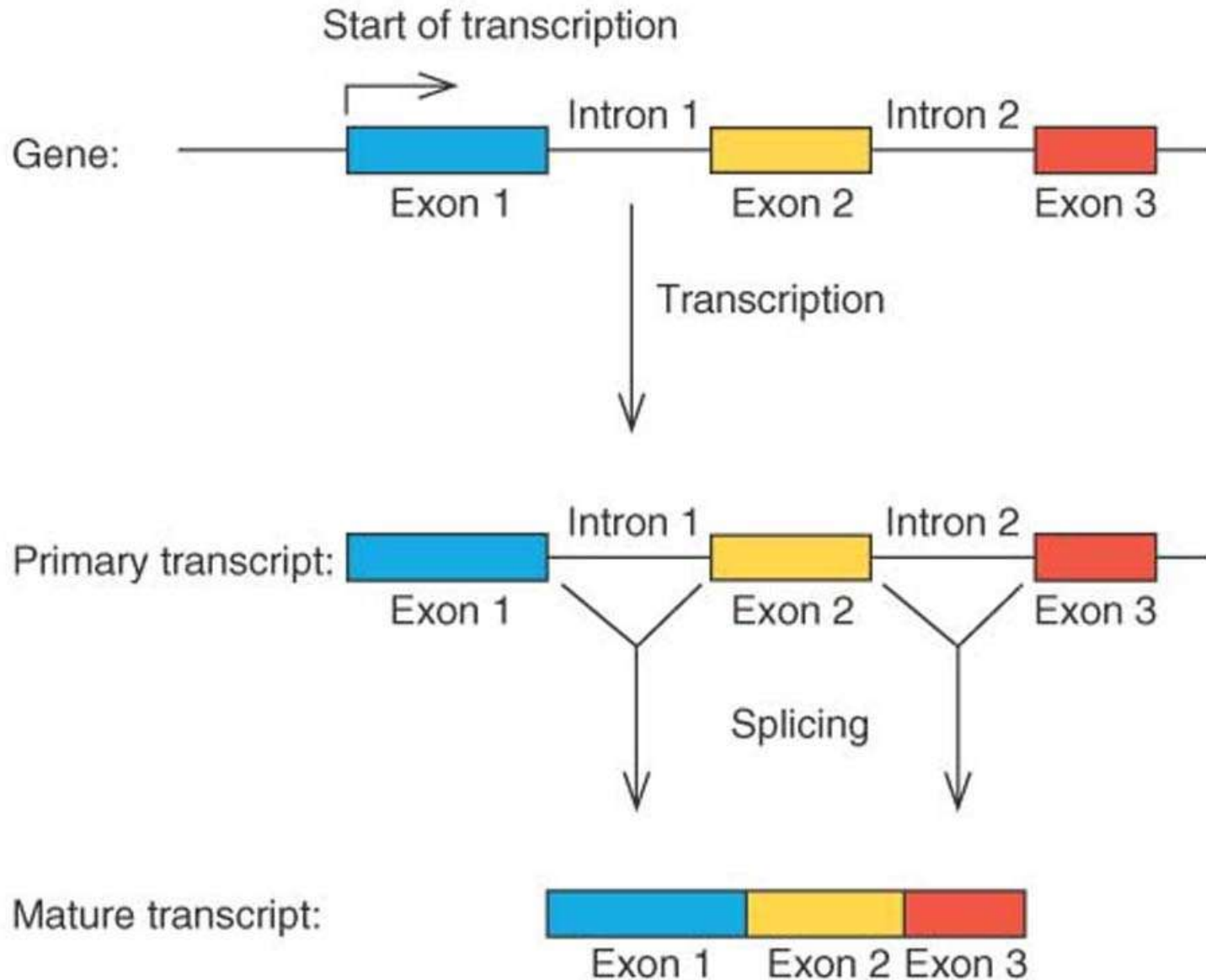


(c)

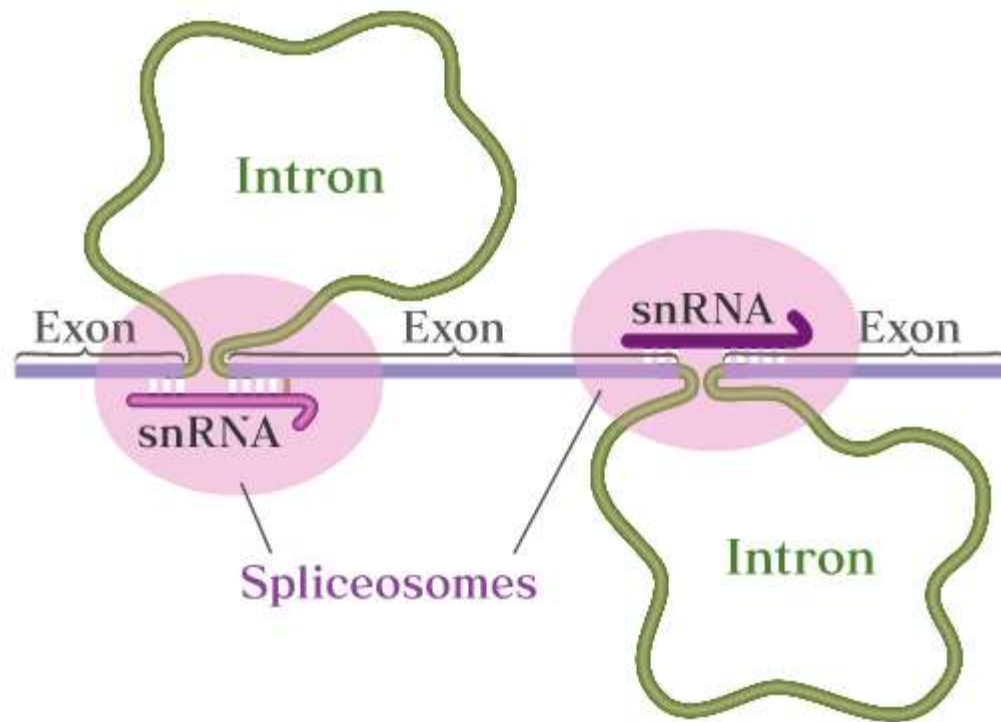


Eukaryotic genes contain introns

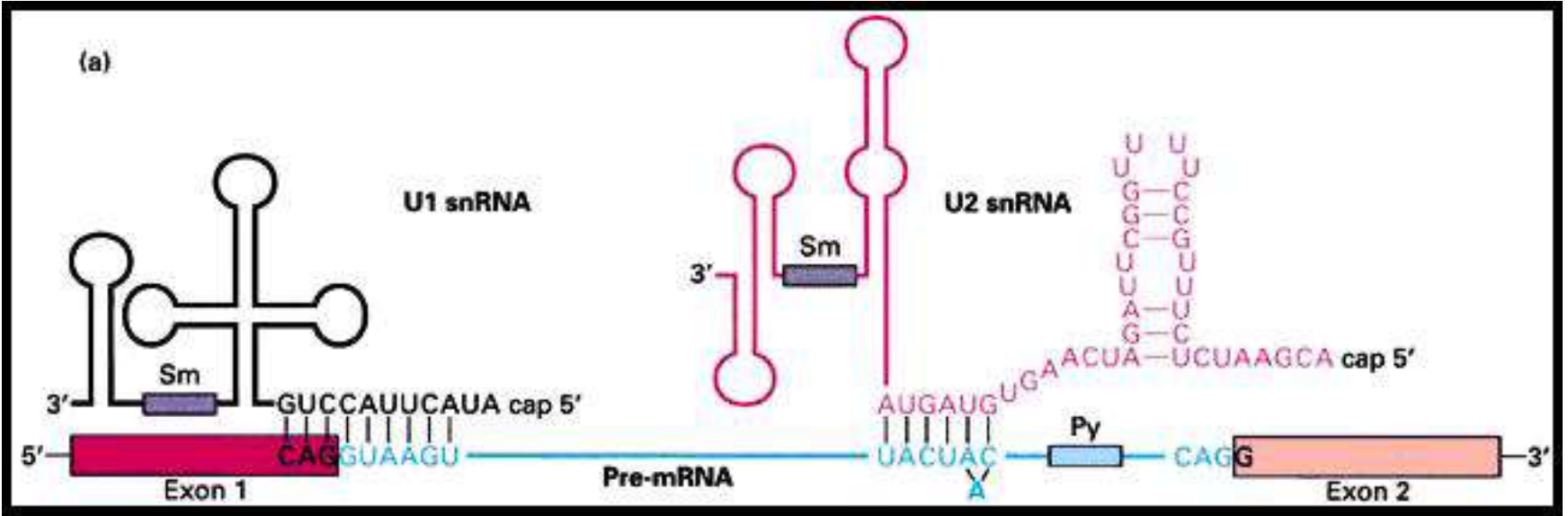
Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



Introns are removed by a process called splicing

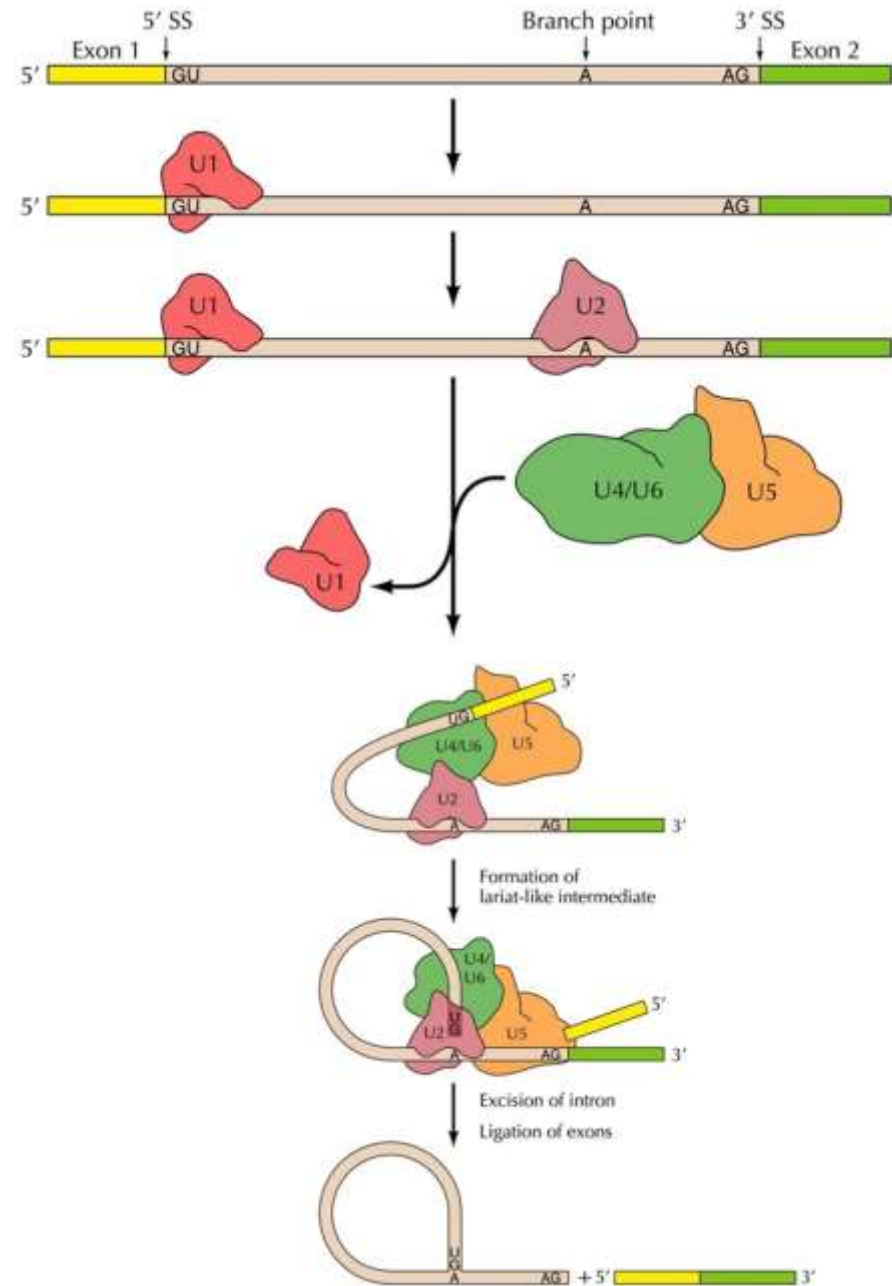


Splicing includes multiple proteins and small nuclear RNAs called snRNAs



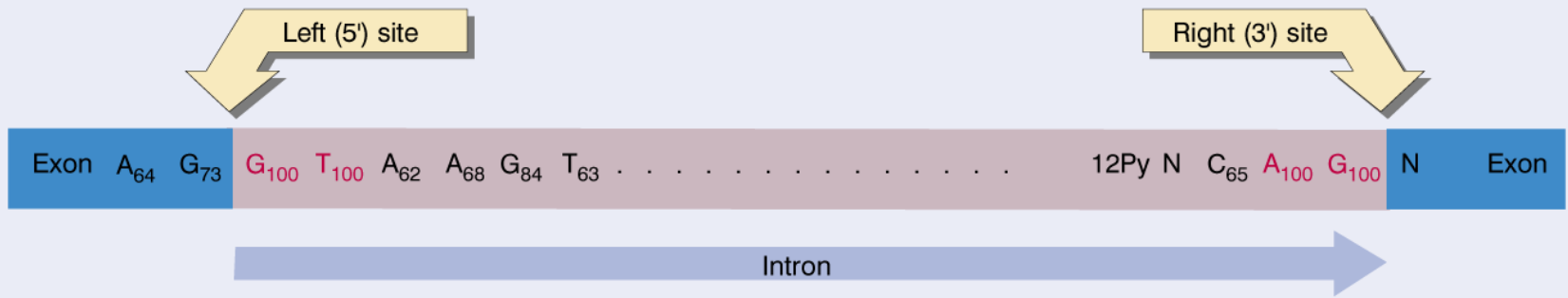
How snRNAs work. Here two snRNAs are shown forming partial hybrids with a pre-mRNA. U1 is forming a hybrid at the junction of the 5' exon and U2 is forming a partial hybrid with a sequence near the 3' exon

Splicing includes multiple proteins and small nuclear RNAs called snRNAs



The splicing complex recognizes semiconserved sequences

Figure 22.3 The ends of nuclear introns are defined by the GT-AG rule.



Complexity of genes

- Splicing in some genes seems straightforward such as glo

```

ACAUUUGCUUCUGACACAACUGUGUUCACUAGCAACCUCAAACAGACACCAUGGUGCACCUGACUCCUGAGGAGAAG
                                (met) val his leu thr pro glu glu lys
                                1           5
UCUGCCGUUACUGCCCUGUGGGGCAAGGUGAACGUGGAUGAAGUUGGUGAGGCCUGGGCAGGUUGGUUAUAAG
ser ala val thr ala leu trp gly lys val asn val asp glu val gly gly glu ala leu gly arg
10           15           20           25           30
GUUACAAGACAGGUUUAAGGAGACCAAUAGA AACUGGGCAUGUGGAGACAGAGAAGACUCUUGGGUUUCUGAUAGGC

ACUGACUCUCUCUGCCUAUUGGUCUAUUUUCCACCCUUAGGCUGCUGGUGGUCUACCCUUGGACCCAGAGGUUCUUU
                                leu leu val val tyr pro trp thr gln arg phe phe
                                31           35           40
GAGUCCUUUGGGGAUCUGUCCACUCCUGAUGCUGUUUUGGGCAACCCUAAGGUGAAGGCUCUAGGCAAGAAGUG
glu ser phe gly asp leu ser thr pro asp ala val met gly asn pro lys val lys ala his gly lys lys val
45           50           55           60           65
CUCGGUGCCUUUAGUGAUGGGCUGGCUCACCUGGACAACCUAAGGGCACC UUUGCCACACUGAGUGAGCUGCAC
leu gly ala phe ser asp gly leu ala his leu asp asn leu lys gly thr phe ala thr leu ser glu leu his
70           75           80           85           90
UGUGACAAGCUGCACGUGGAUCCUGAGAACUUCAGGUGAGUCUAUGGGACCCUUGAUGUUUUCUUUCCCUUCUUU
cys asp lys leu his val asp pro glu asn phe arg
95           100           104
UCUAUGGUUAAGUUC AUGUCAUAGGAAGGGGAGAAGUAACAGGGUACAGUUUAGA AUGGGAAACAGACGAAUGAUUG
CAUCAGUGUGGAAGUCUCAGGAUCGUUUUAGUUUCUUUUUUUUGCUGUUCAUAA CAUUGUGUAUAACAAAAGGAAAU
AUCUCUGAGAUACAUAUAGUAACUAAAAAAACUUUACACAGUCUGCCUAGUACAUAUACU AUUUGGAUAUAUUGUG
UGCUUAUUUGCAUAUAUACAUAUUCUCCUACUUUAUUUUUCUUUUUUUUUUAUUGAUACAUAUAUCAUAUAUUUAUG
GGUUAAGUGUAAUGUUUUAAUUGUGUACACAUAUUGACCAAUUCAGGGUAAUUUUGCAUUUGUAAUUUUAAAAAU
GCUUUCUUUUUAAUAUACUUUUUUUGUUAUCUUAUUUCUAUAUACUUUCCUAUUCUCUUUUCUUCAGGGCAAUAUGA
UACAAGUAUACUGCCUUCUUUGCACCAUUCUAAGAAUAACAGUGAUAAUUUCUGGGUUAAGGCAAUAGCAAUUUU
CUGCAUUAUUUUUUUGCAUAUAAAUUGUAACUGAUGUAAGAGGUUUCAUUUGCUAAUAGCAGCUACA AUCCAG
CUACCAUUCUGCUUUUUUUUUAUGGUUGGGAUAAGGCUGGAUUAUUCUGAGUCCAAGCUAGGCCUUUUGCUAAUCAU
GUUUAUACCUCUUAUUCUUCUCCACAGCUCUUGGGCAACGUGCUGGUCUGUGUGCUGGCCCAUACAUUUGGCAAA
                                leu leu gly asn val leu val cys val leu ala his his phe gly lys
                                105           110           115           120
GAAUUCACCCACCAUGCAGGCGGCCUAUCAGAAAGUGGUGGCGUGGUGGCUAAUGCCUGGCCACAAGUAU
glu phe thr pro val gln ala ala tyr gln lys val val ala gly val ala asn ala leu ala his lys tyr
125           130           135           140           145
CACUAAGCUCGCUUUCUUGCUGUCCA AUUUUUAUUAAGGUUCCUUUGUCCCUAAGUCCAACUACUAAACUGGGGG
his TER
AUUUUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUAUUAAAAAA CAUUUAUUUUAUUUGC

```

For other genes splicing is much more complex

- Fibrillin is a protein that is part of connective tissue. Mutations in it are associated with Marfan Syndrome (long limbs, crowned teeth elastic joints, heart problems and spinal column deformities. The protein is 3500 aa, and the gene is 110 kb long made up of 65 introns.
- Titin has 175 introns.
- With these large complex genes it is difficult to identify all of the exons and introns.

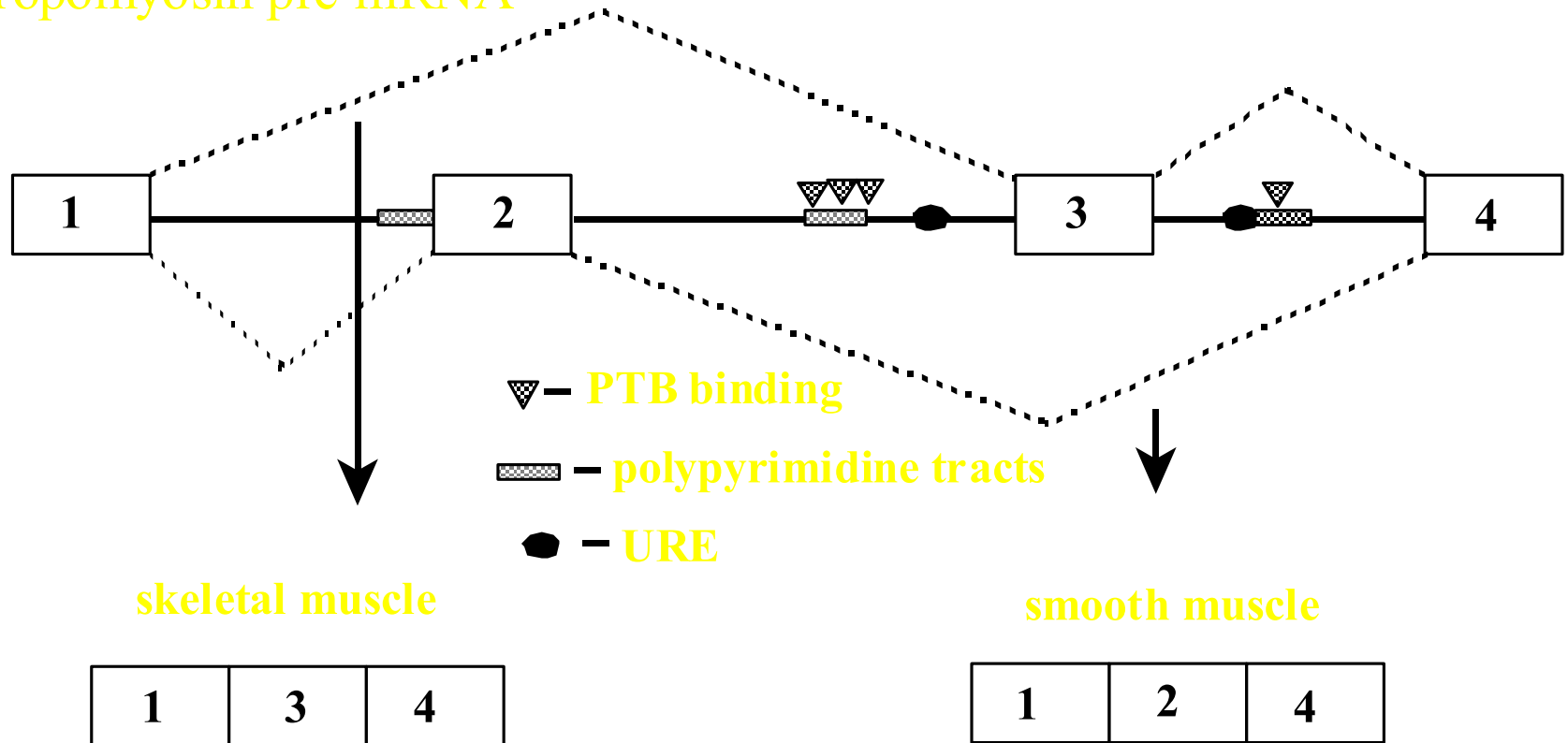


Alternative RNA splicing

- Shortly after the discovery of splicing came the realization that the exons in some genes were not utilized in the same way in every cell or stage of development. In other words exons could be skipped or added. This means that variations of a protein (called isoforms) can be produced from the same gene.

Alternative splicing of a tropomyosin

α -tropomyosin pre-mRNA



Alternative RNA splicing

- Shortly after the discovery of splicing came the realization that the exons in some genes were not utilized in the same way in every cell or stage of development. In other words exons could be skipped or added. This means that variations of a protein (called isoforms) can be produced from the same gene.

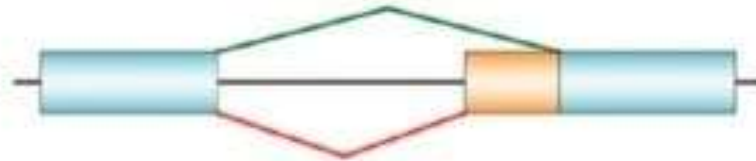
Alternative Splicing

b

Exon skipping/inclusion



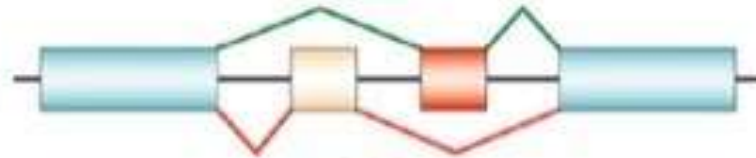
Alternative 3' splice sites



Alternative 5' splice sites



Mutually exclusive exons



Intron retention



Genome sequence analysis

Introns are abundant

94% human, 85% fruitfly, 95% nemtode, 95% plant genes

Human genome (25,000 genes)

**Only 2 times as many genes as in fruitflies (13,600)
or in nematodes (19,000)**

Alternative splicing is abundant

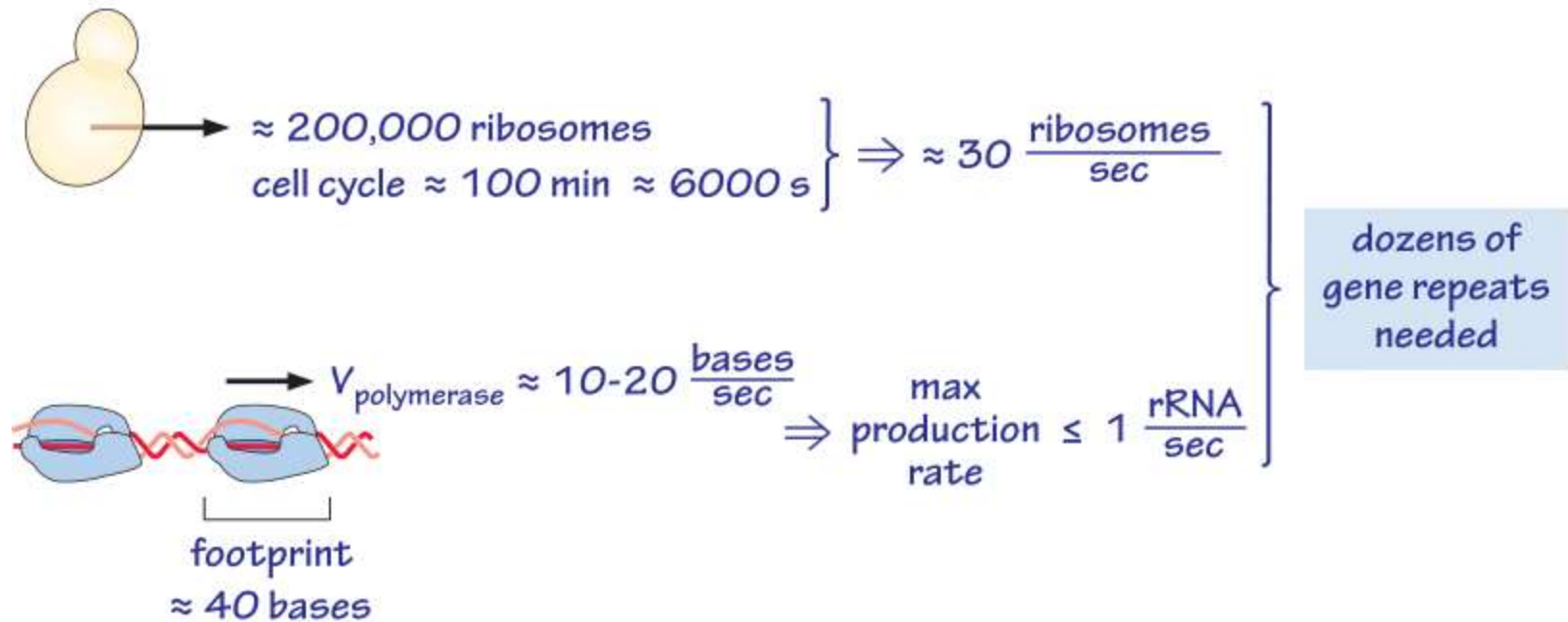
~75% of human gene transcripts are alternatively spliced

Number of rRNA and tRNA Genes

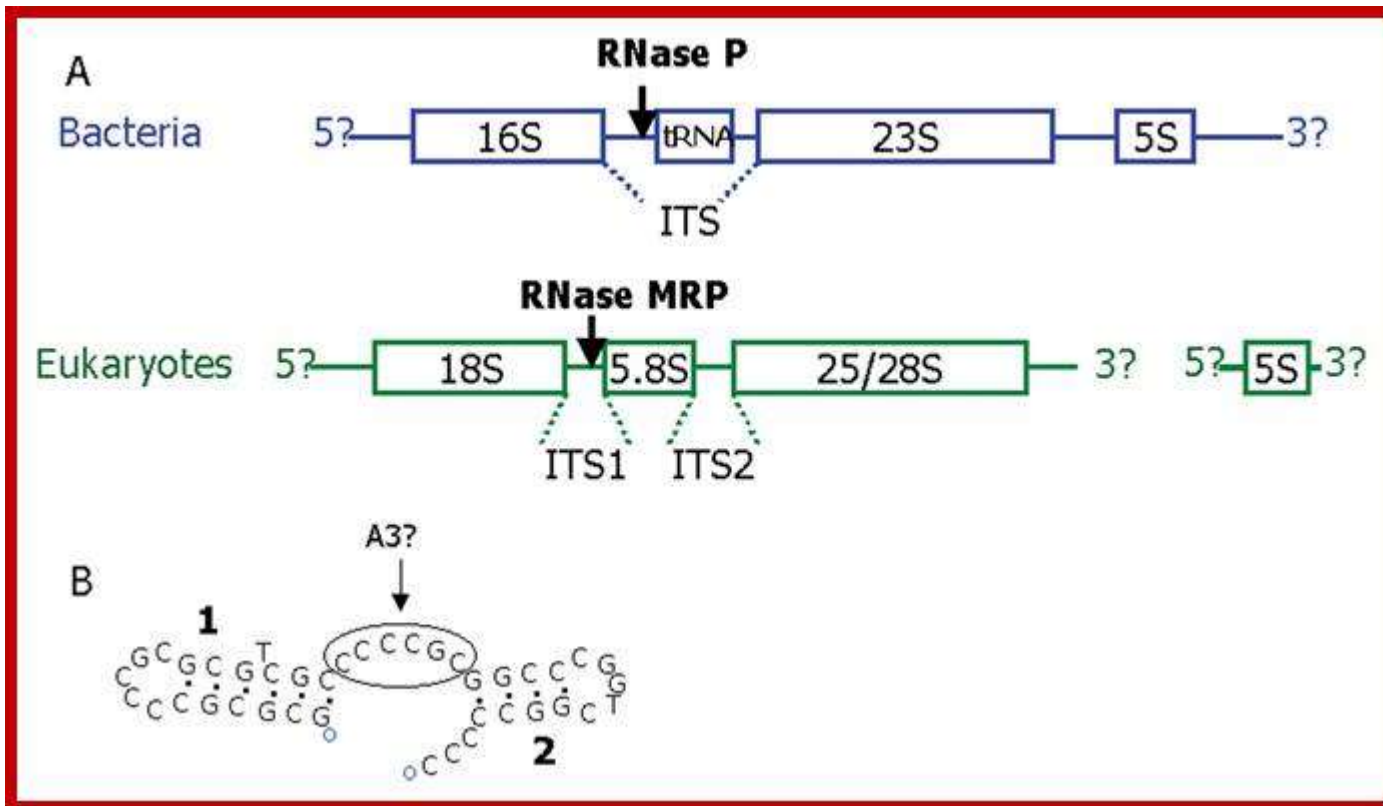
Species	16s/23s/5s or 18s/28s-5.8s rRNA genes	5s rRNA genes	tRNA genes
E.coli	7	Embedded in rRNA end segment	~60
S.cerviciae	140	140	250-400
Dictyostelium discoideum	180	180	>850
D.melanogaster: XY: XX:	150 250	150 150	850 850
X.laevis	450-600	24000	1150
Homo sapiens	280	2000	>2000

Number of rRNA and tRNA Genes

how many rRNA genes are needed?



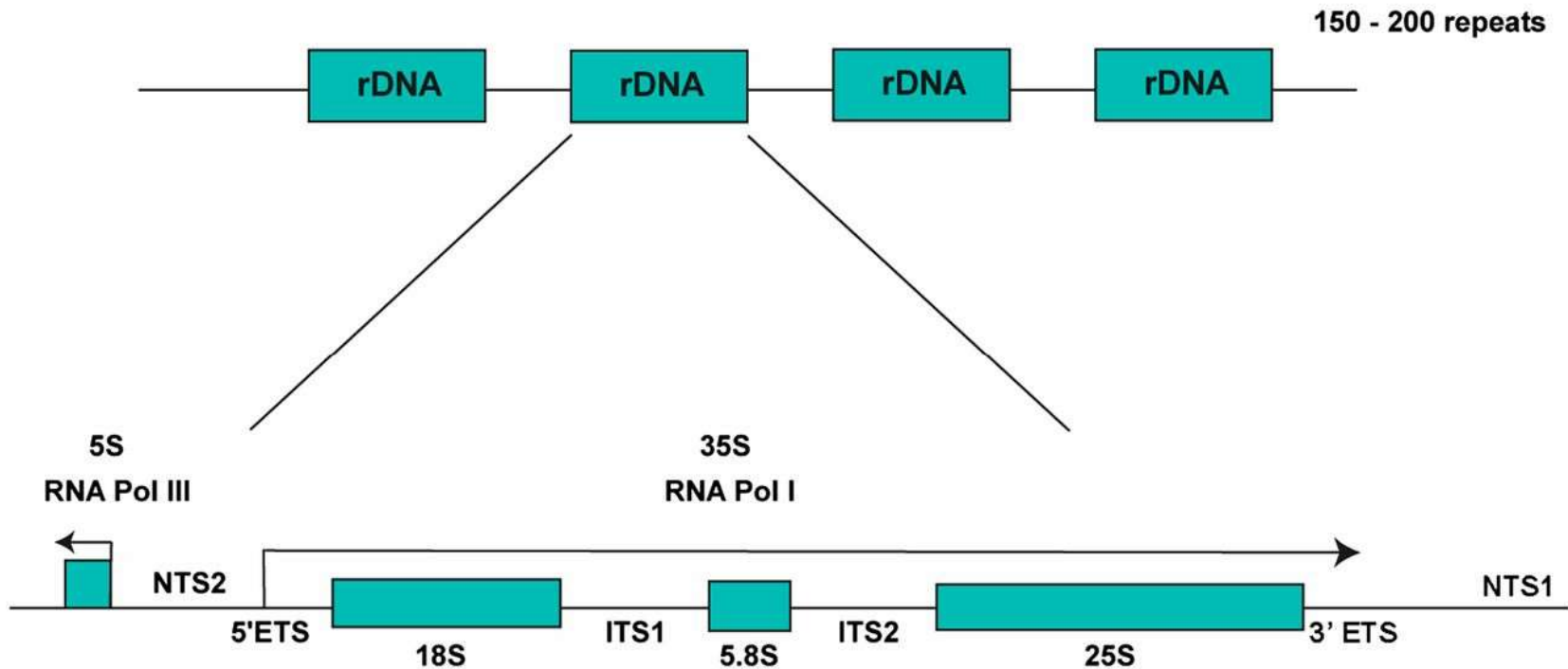
Organization of rRNA genes



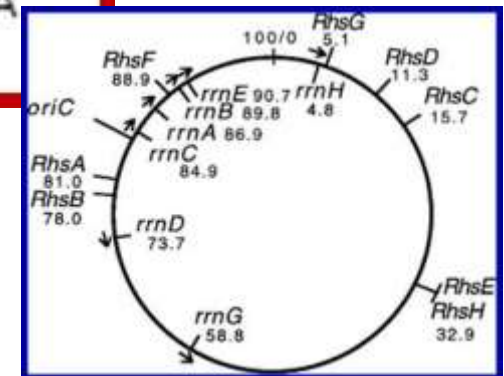
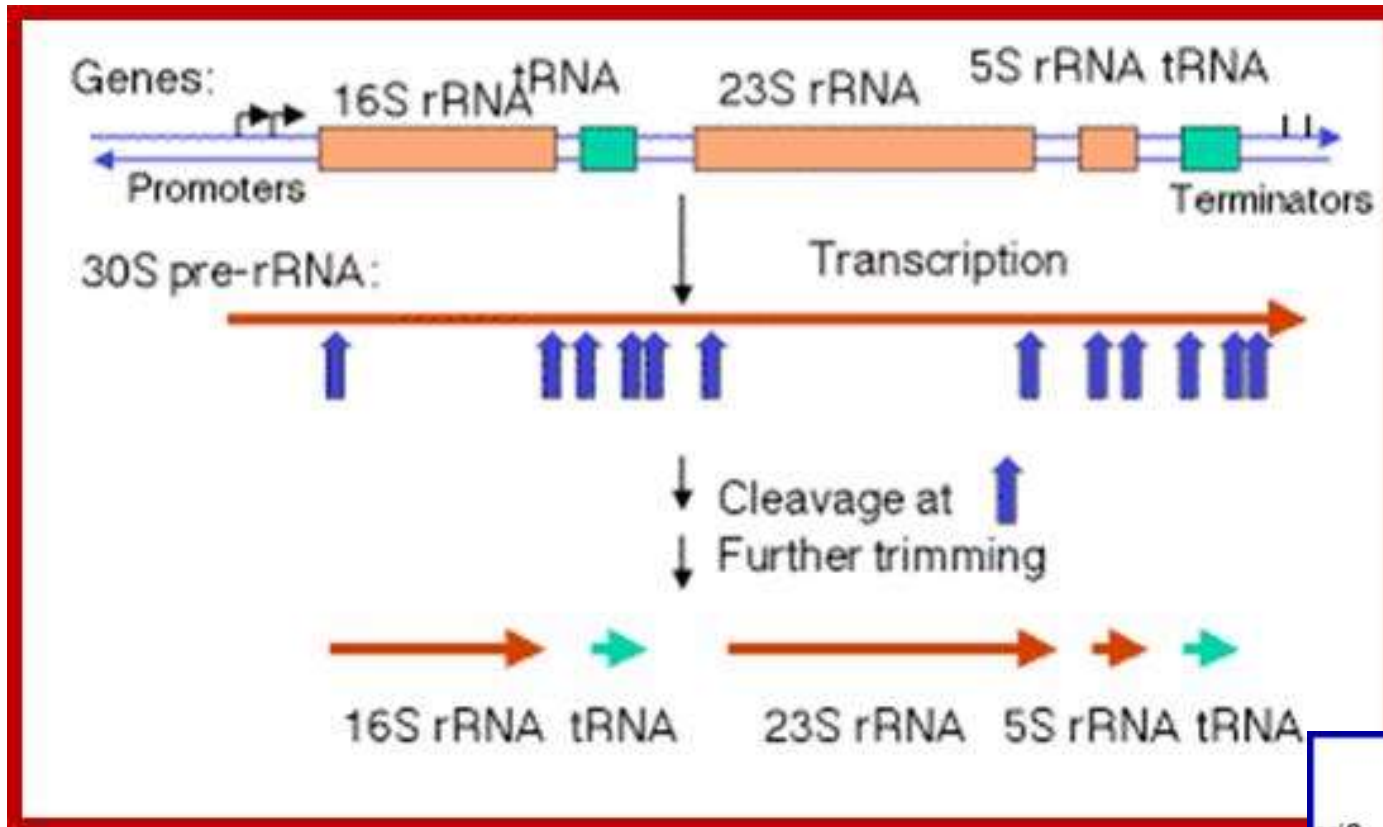
the 5S, 5.8S and 28S in mammals, 25S in plants

Organization of rRNA genes

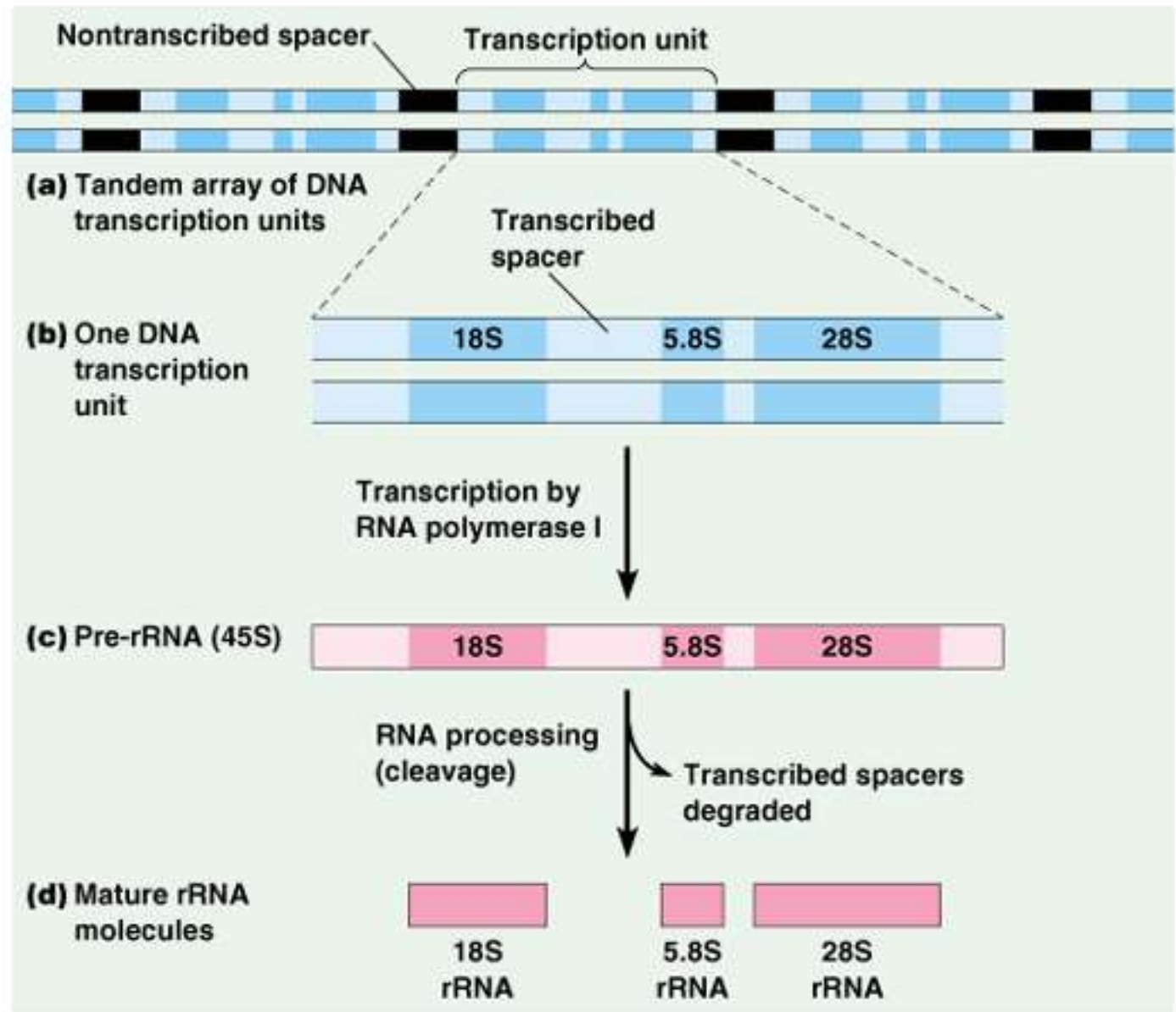
S. cerevisiae Chromosome XII



rRNA operon in Escherichia coli

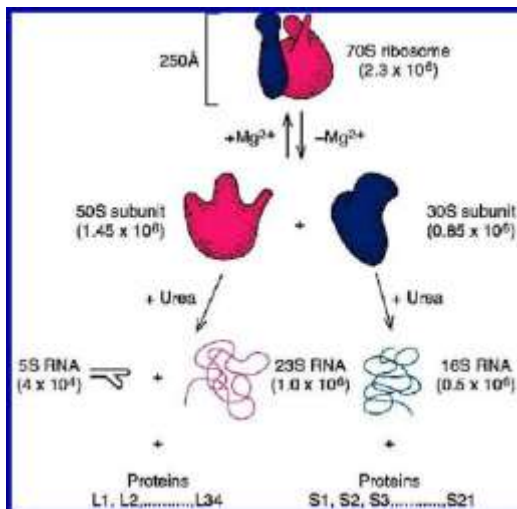


Mechanism of rRNA splicing



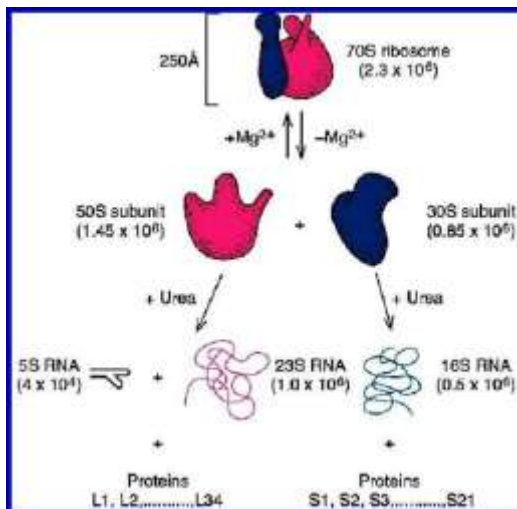
Mechanism of rRNA splicing

Most of the RNAs transcribed are precursor RNAs and they have to be processed to their functional state. Ribosomal RNAs are synthesized as large precursors and as they are synthesized they are subjected to processing, which involves, methylation (modifications) and pseudouridylation of bases at specific sites by using secondary structural conformations; then cleaving the precursor transcript at specific positions. Methylases, methylate 4-O-guanine bases (GC pairs) and 2-O- group of ribose sugar. Methylation is very important in the binding of riboproteins to rRNA and also for the folding of rRNA as and when proteins bind to form a ribosomal structure.



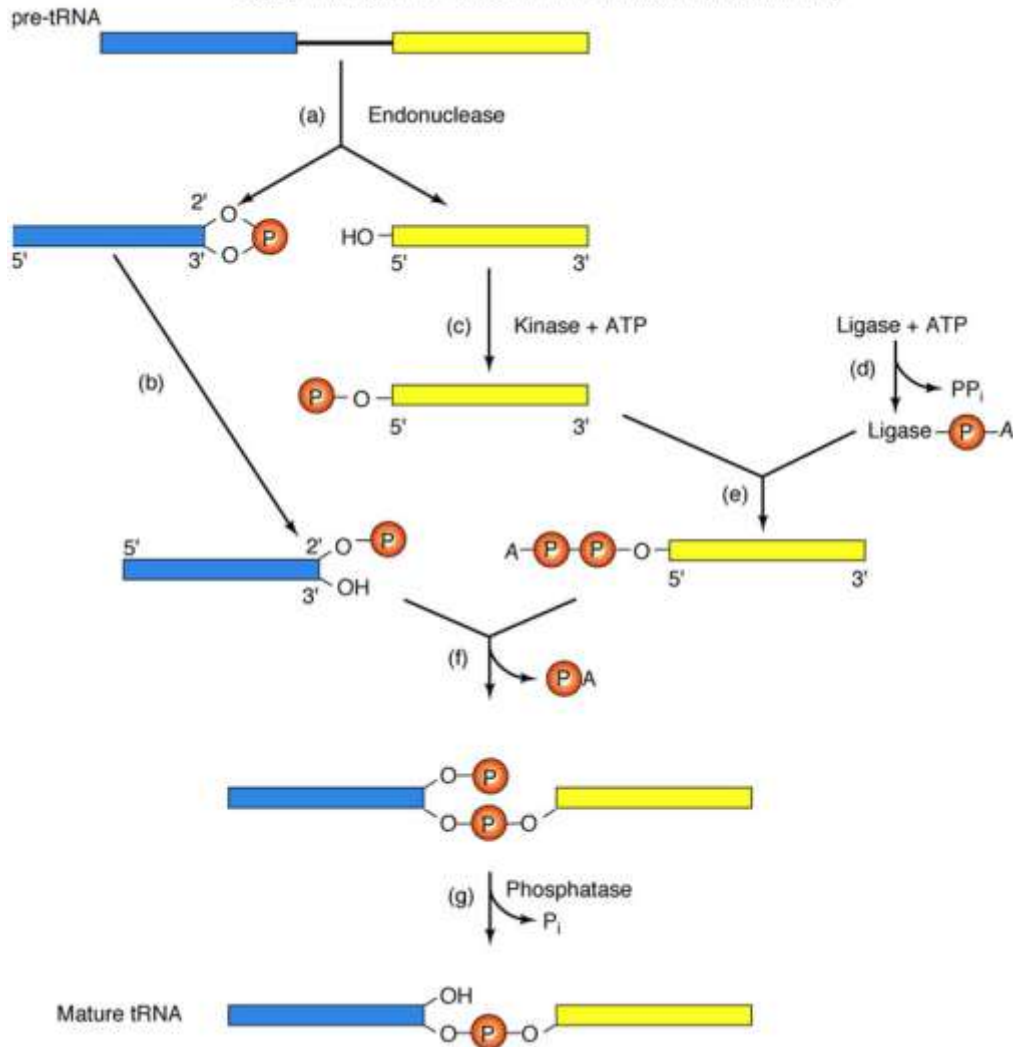
Mechanism of rRNA splicing

Most of the RNAs transcribed are precursor RNAs and they have to be processed to their functional state. Ribosomal RNAs are synthesized as large precursors and as they are synthesized they are subjected to processing, which involves, methylation (modifications) and pseudouridylation of bases at specific sites by using secondary structural conformations; then cleaving the precursor transcript at specific positions. Methylases, methylate 4-O-guanine bases (GC pairs) and 2-O- group of ribose sugar. Methylation is very important in the binding of riboproteins to rRNA and also for the folding of rRNA as and when proteins bind to form a ribosomal structure.



Mechanism of tRNA splicing

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



C



Mechanism of tRNA splicing

