Gene Expression

The Flow of Genetic Information from DNA via RNA to Protein

Francis Crick (1956) named the flow of information from DNA → RNA→protein the ***Central Dogma***.

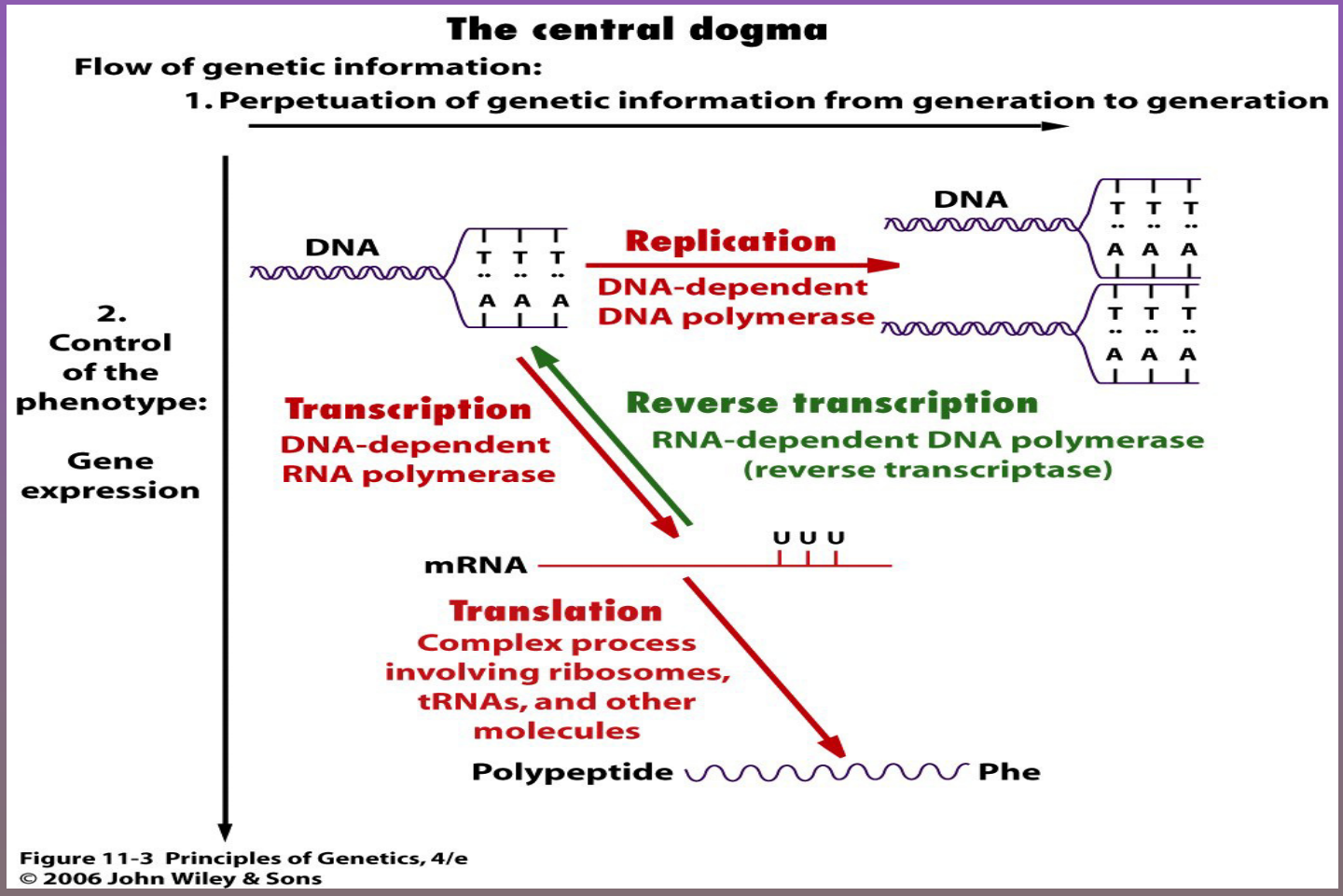


Figure 11-3 Principles of Genetics, 4/e
© 2006 John Wiley & Sons

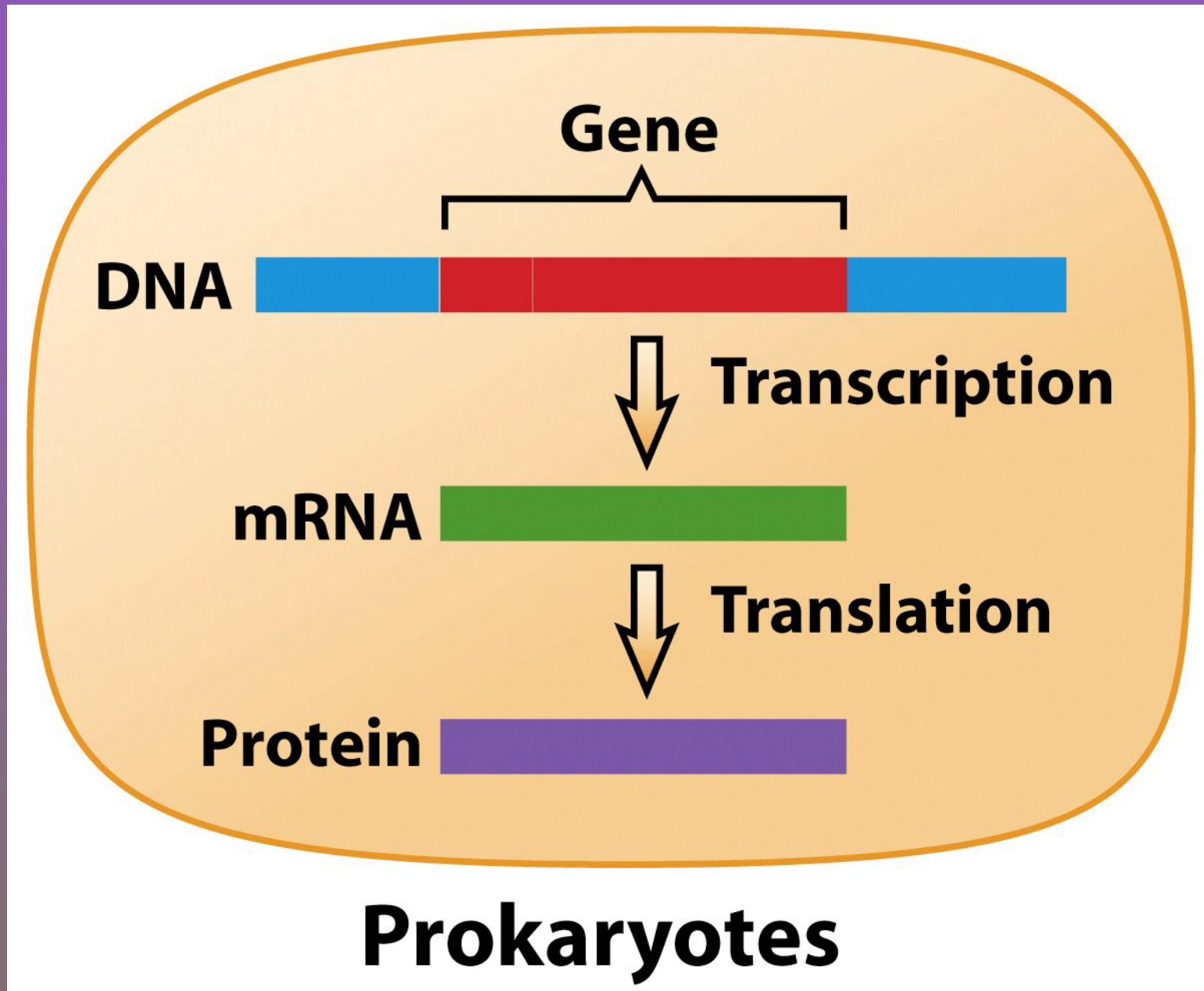


Figure 11-4a Principles of Genetics, 4/e

© 2006 John Wiley & Sons

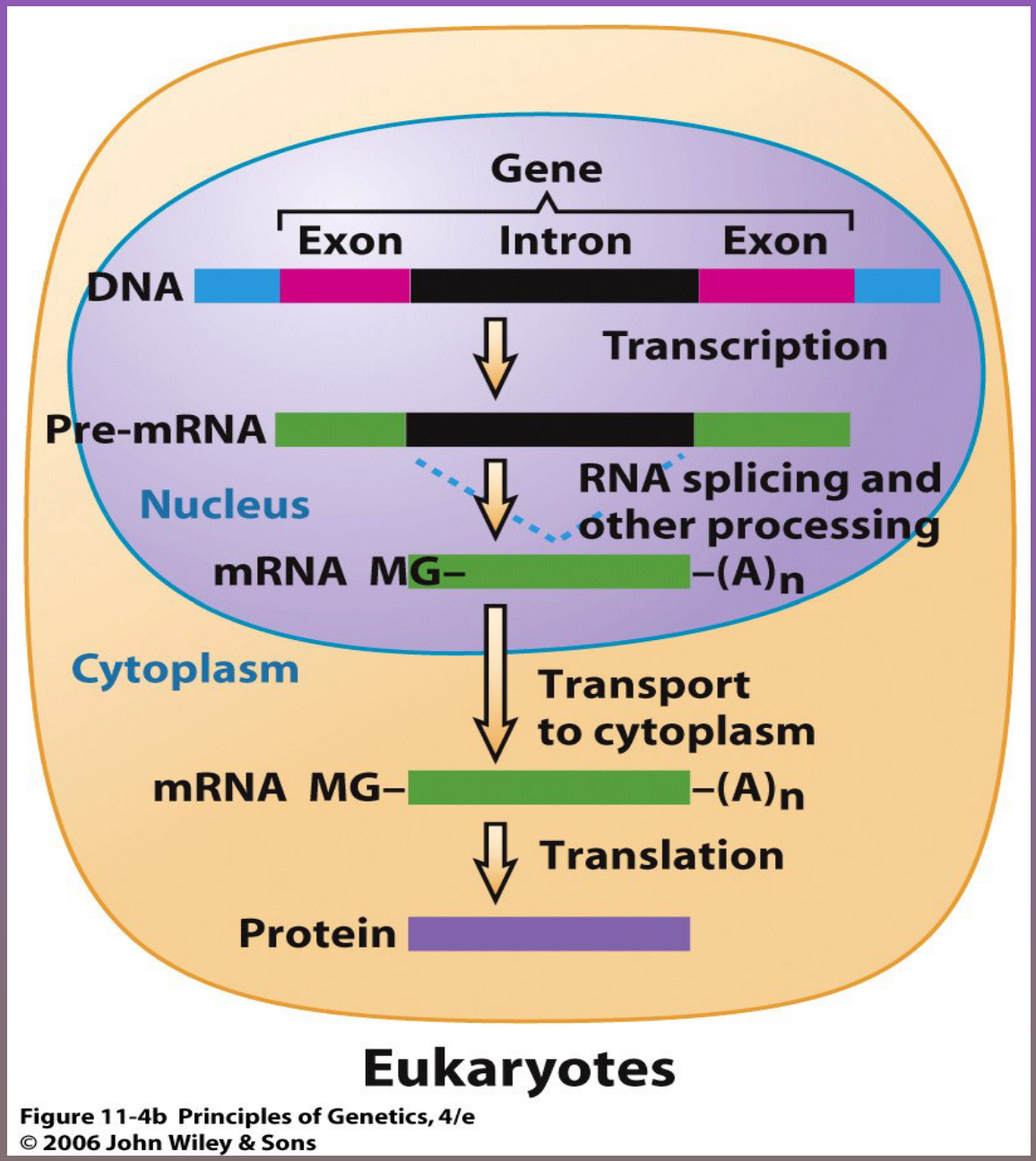


Figure 11-4b Principles of Genetics, 4/e

© 2006 John Wiley & Sons

A two step process:

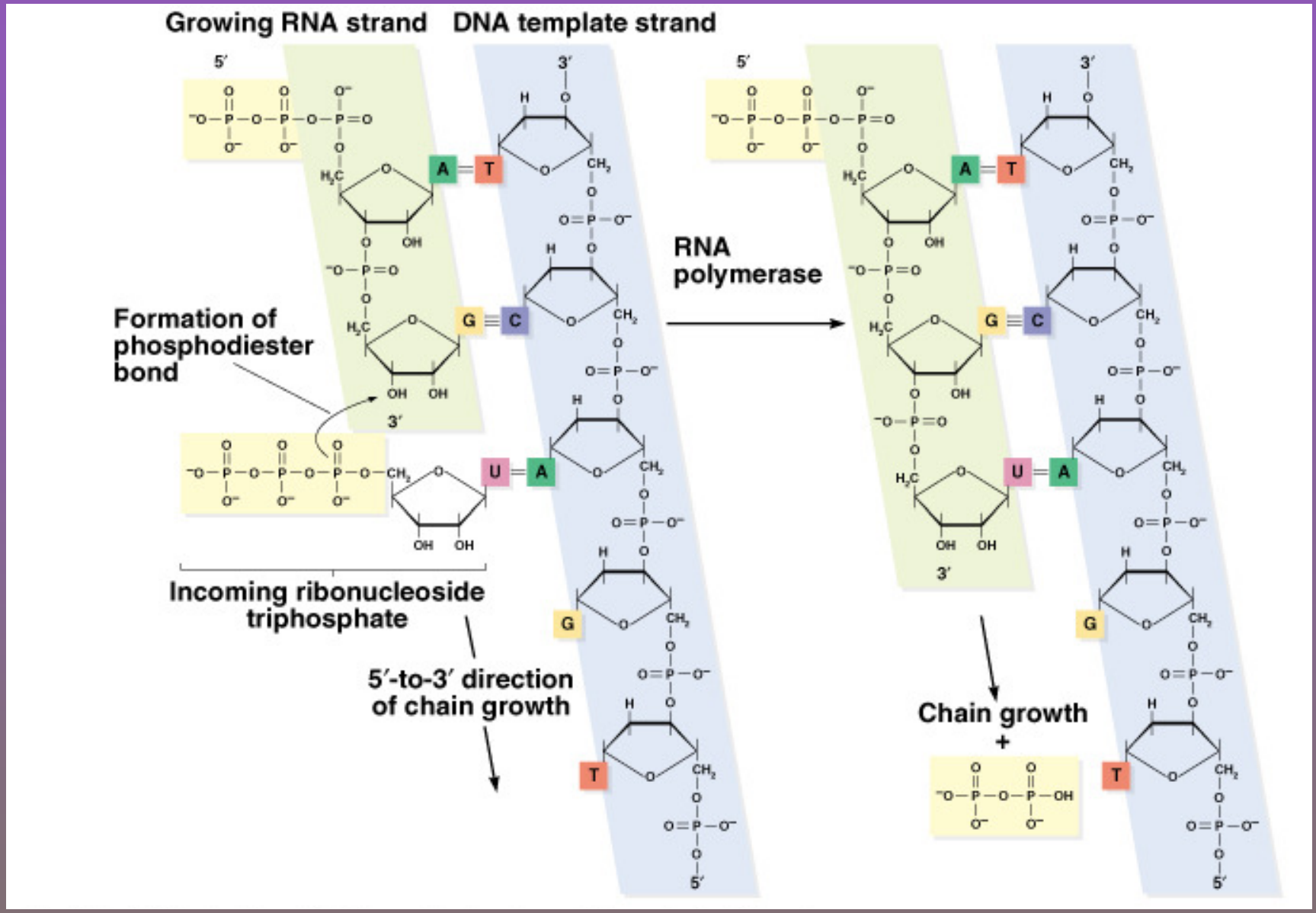
Transcription = synthesis of a single-stranded RNA molecule using the DNA template (1 strand of DNA is transcribed).

Translation = conversion of a messenger RNA sequence into the amino acid sequence of a polypeptide (i.e., protein synthesis)

Both processes generally occur throughout the cell cycle.

Transcription: How is an RNA strand synthesized?

1. Regulated by gene regulatory elements within each gene.
2. RNA is transcribed 5' to 3' from the template (3' to 5').
3. Similar to DNA synthesis, except:
   - ✓ NTPs instead of dNTPs (no deoxy-)
   - ✓ No primer
   - ✓ No proofreading
   - ✓ Adds Uracil (U) instead of thymine (T)
   - ✓ RNA polymerase


Growing RNA strand
DNA template strand
5′
3′
A
T
G
C
U
A
G
T
Formation of phosphodiester bond
Incoming ribonucleoside triphosphate
5′-to-3′ direction of chain growth
RNA polymerase
Chain growth
+
5′
3′

Functions of RNA Polymerase

1. Searches for initiation site (Promoters)

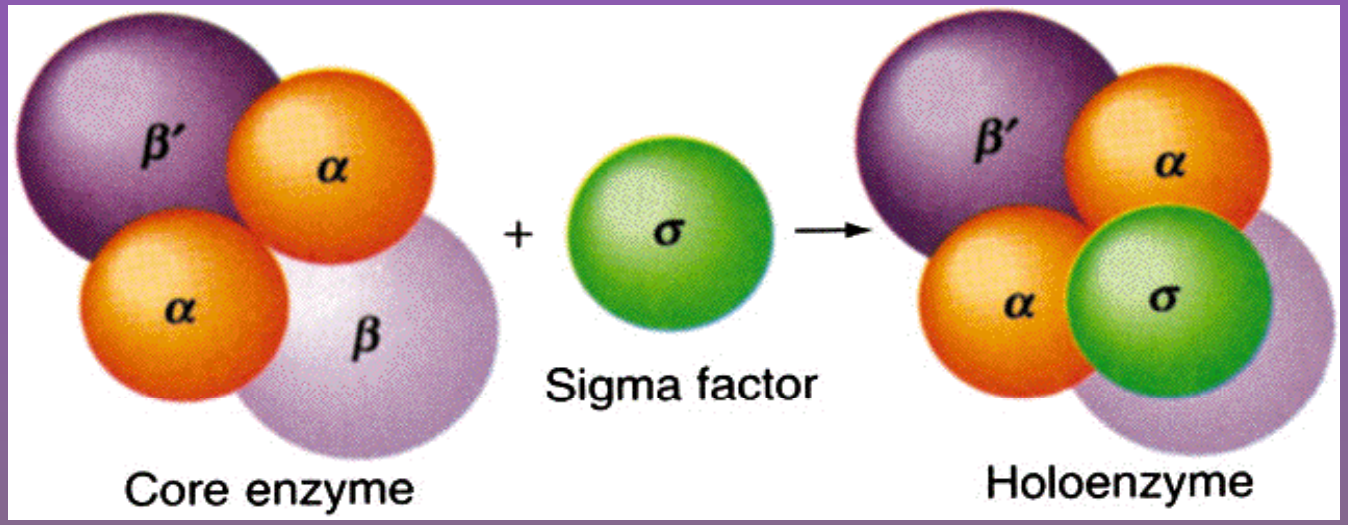
2. Unwinds DNA template

3. Select correct NTP for: base-pairing for initiation and elongation in RNA synthesis

4. Detect termination signals

5. Interact with activator or repressor proteins: controls transcriptional levels

- **In *E. coli*, a single RNA polymerase synthesizes all the three kinds of RNA** (mRNA, rRNA and tRNA).



Sigma factor: recognizes promoters and initiates transcription.

Without sigma, the core enzyme randomly binds DNA but does not transcribe it efficiently.

Different sigma factors recognize different promoter sequences.

Three Steps to Transcription:

1. Initiation
2. Elongation
3. Termination

- ✓ Occur in both prokaryotes and eukaryotes.
- ✓ Elongation is conserved in prokaryotes and eukaryotes.
- ✓ Initiation and termination proceed differently.

Step 1-Initiation, *E. coli* model:

Each gene has three regions:

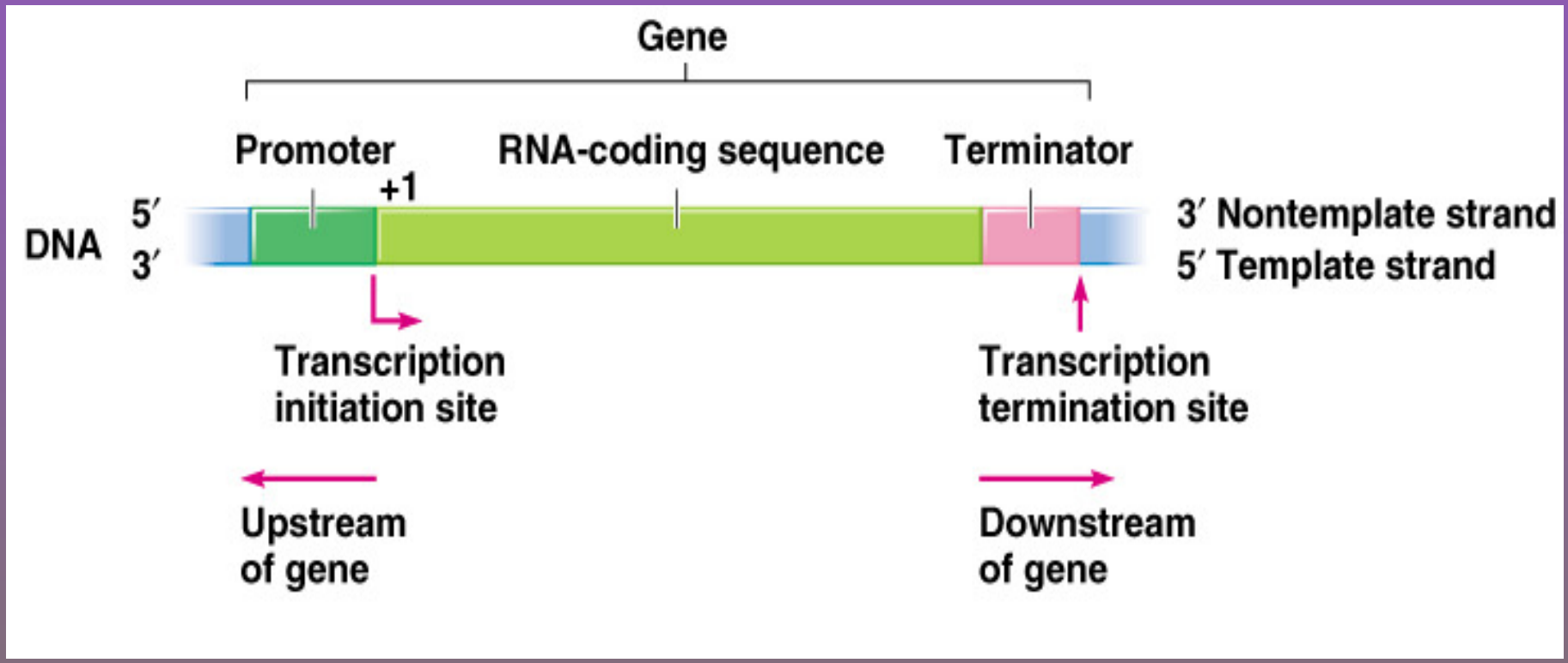
1. 5’ Promoter, attracts RNA polymerase

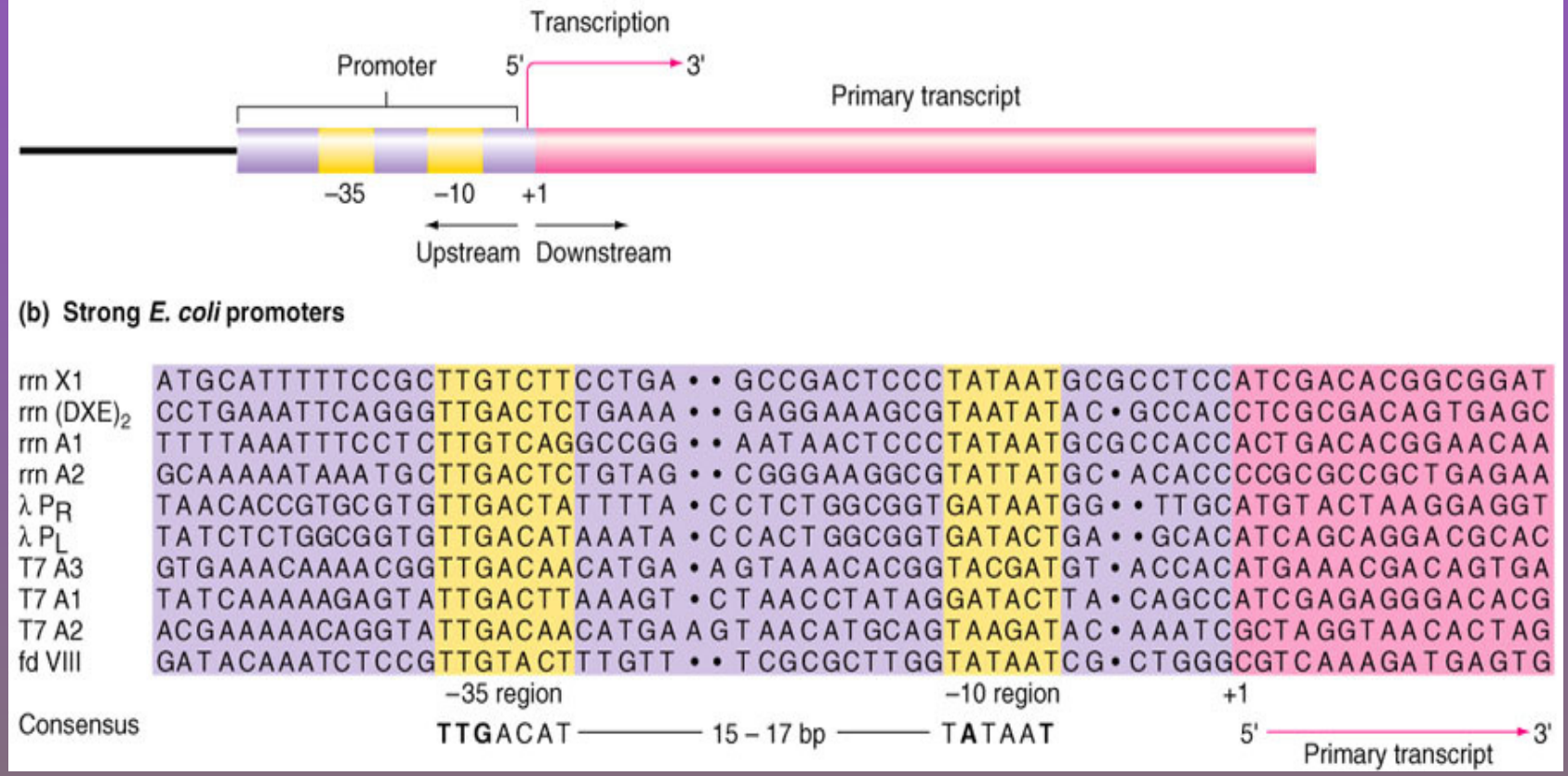
    e.g., -10 bp 5’-TATAAT-3’ (Pribnow box)
    e.g., -35 bp 5’-TTGACA-3’

2. Transcribed sequence, or RNA coding sequence

3. 3’ Terminator, signals the stop point

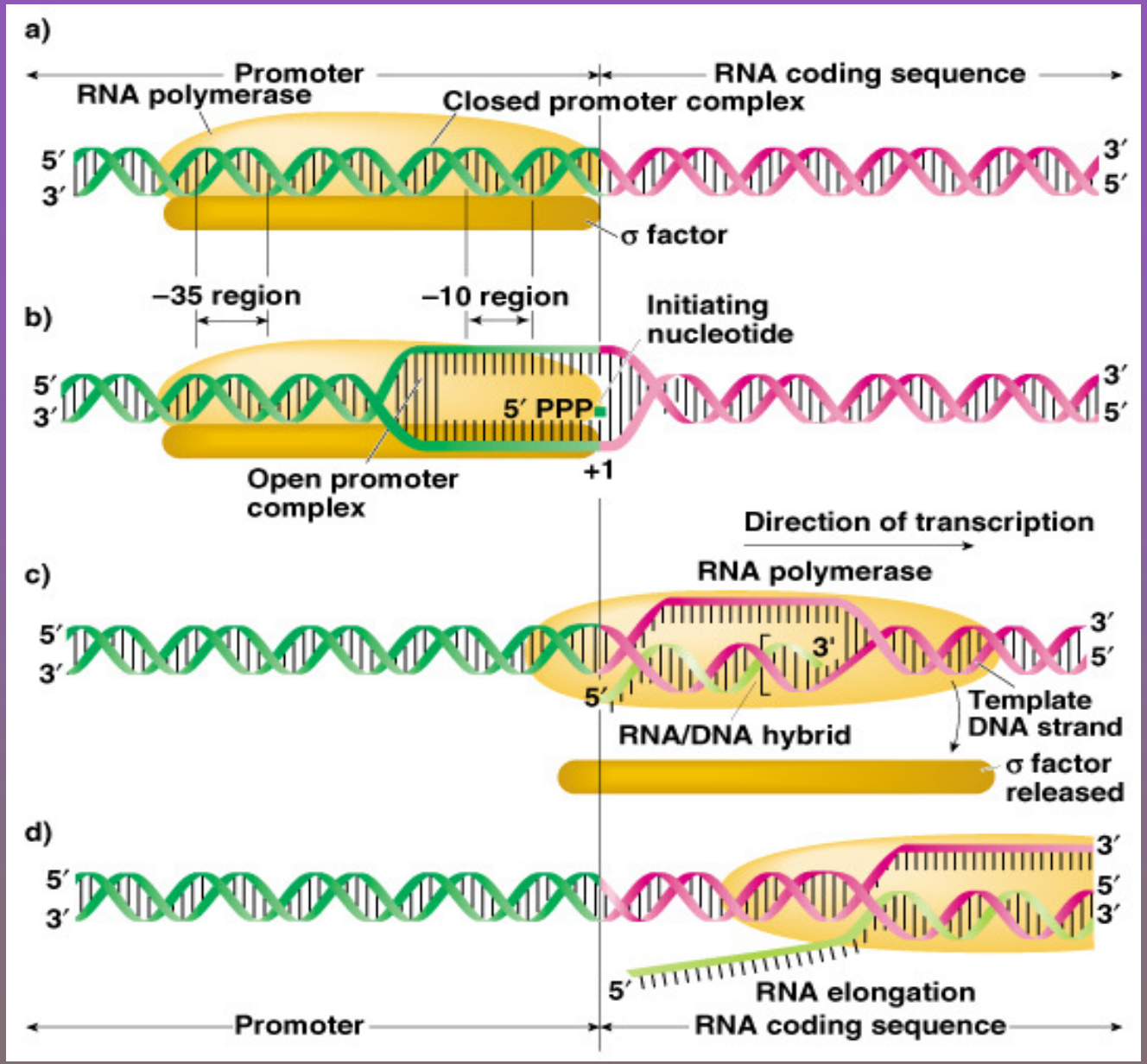
Promoter, RNA-coding sequence, and terminator regions of a gene




Transcription
Promoter
5'
3'
Primary transcript
–35
–10
+1
Upstream
Downstream
(b) Strong E. coli promoters
rrn X1 ATGCATTTTTCCGCTTGTCTTCCTGA••GCCGACTCCCTATAATGCGCCTCCATCGACACGGCGGAT
rrn (DXE)2 CCTGAAATTCAGGGTTGACTCTGAAA••GAGGAAAGCGTAATATAC•GCCACCTCGCGACAGTGAGC
rrn A1 TTTTAAATTTCCTCTTGTCAGGCCGG••AATAACTCCCTATAATGCGCCACCACTGACACGGAACAA
rrn A2 GCAAAAATAAATGCTTGACTCTGTAG••CGGGAAGGCGTATTATGC•ACACCCCGCGCCGCTGAGAA
λ PR TAACACCGTGCGTGTTGACTATTTTA•CCTCTGGCGGTGATAATGG••TTGCATGTACTAAGGAGGT
λ PL TATCTCTGGCGGTGTTGACATAAATA•CCACTGGCGGTGATACTGA••GCACATCAGCAGGACGCAC
T7 A3 GTGAAACAAAACGGTTGACAACATGA•AGTAAACACGGTACGATGT•ACCACATGAAACGACAGTGA
T7 A1 TATCAAAAAGAGTATTGACTTAAAGT•CTAACCTATAGGATACTTA•CAGCCATCGAGAGGGACACG
T7 A2 ACGAAAAACAGGTATTGACAACATGAAGTAACATGCAGTAAGATAC•AAATCGCTAGGTAACACTAG
fd VIII GATACAAATCTCCGTTGTACTTTGTT••TCGCGCTTGGTATAATCG•CTGGGCGTCAAAGATGAGTG
–35 region
–10 region
+1
Consensus
TTGACAT — 15 – 17 bp — TATAAT
5'
3'
Primary transcript

Step 1-Initiation…

1. RNA polymerase combines with sigma factor to create RNA polymerase holoenzyme
2. RNAP scans the DNA looking for promoters.
3. The DNA is still is a double helix (closed complex).
4. RNAP unwinds the DNA resulting in open complex formation.
5. First nucleotides are added to start RNA chain. Transcriptional initiation has occurred!


a)
Promoter
RNA coding sequence
RNA polymerase
Closed promoter complex
5′
3′
3′
5′
σ factor
–35 region
–10 region
b)
Initiating nucleotide
5′
3′
5′ PPP
3′
5′
Open promoter complex
+1
Direction of transcription
c)
RNA polymerase
5′
3′
3′
5′
3′
5′
Template DNA strand
RNA/DNA hybrid
σ factor released
d)
5′
3′
3′
5′
3′
5′
RNA elongation
Promoter
RNA coding sequence

Step 2-Elongation…

- Elongation is 5’ - 3’
- After 8-9 NTPs have been joined in the growing RNA chain, sigma factor is released and reused for other initiations. Core enzyme completes the transcript
- Core enzyme untwists DNA helix locally, allowing a small region to denature. Newly synthesized RNA forms an RNA-DNA hybrid.
- The region containing RNA polymerase, DNA, and nascent RNA (under progress) is called a *transcript bubble*
- The chain grows at 30–50 nt/second

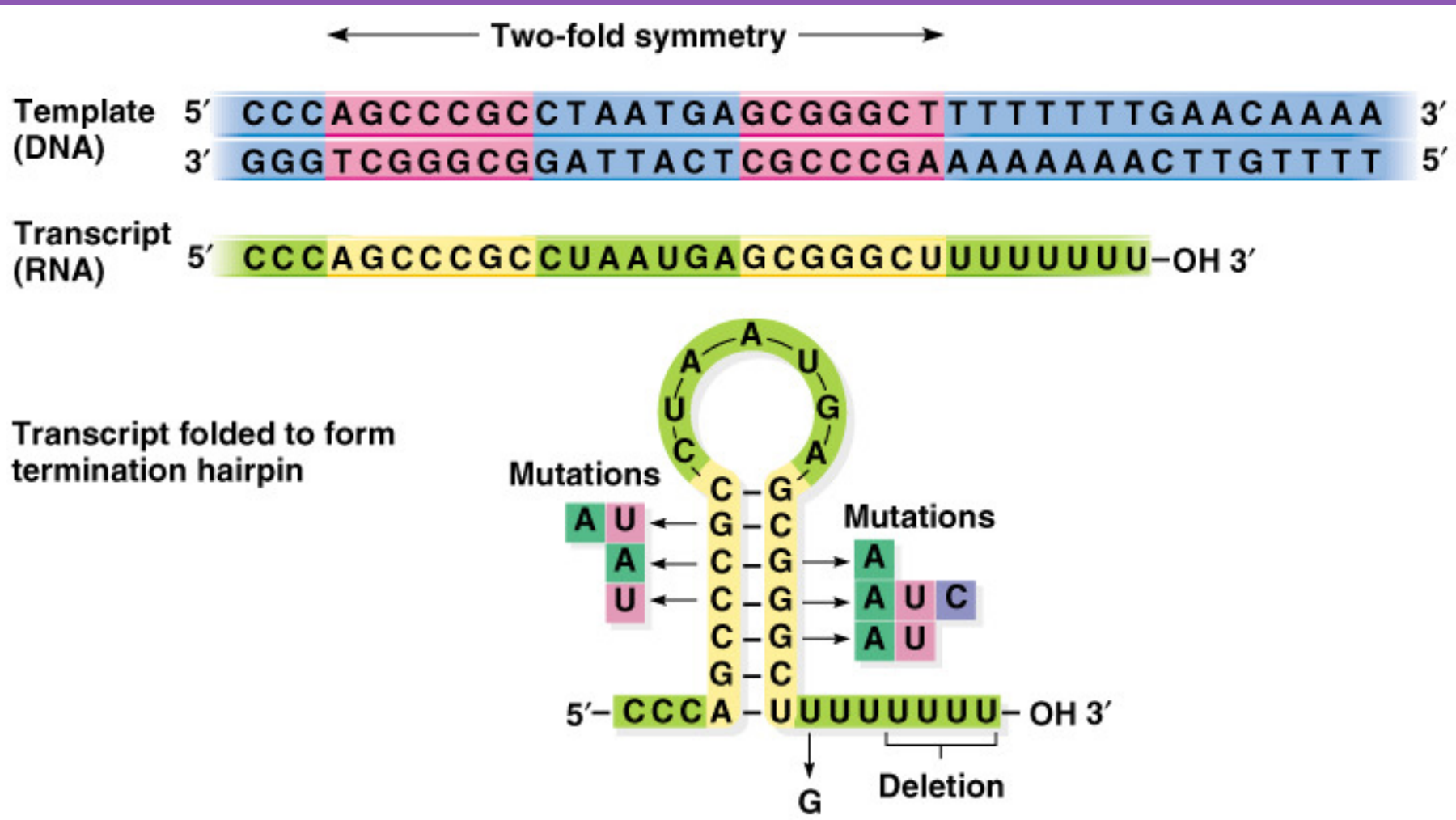
Step 3-Termination

Type I (ρ-independent)

- palindromic sequence having a GC rich region followed by AT rich region at the end of gene.

- When this portion of DNA is transcribed, the RNA transcript forms a *hairpin structure*, followed by a string of Us.

- The hairpin destabilizes the DNA:RNA hybrid leading to dissociation of the RNA from the DNA.


Two-fold symmetry
Template (DNA)
5′ CCCAGCCCGCCTAATGAGCGGGCTTTTTTTTTGAACAAAA 3′
3′ GGGTCGGGCGGATTACTCGCCCGAAAAAAAAACTTGTTTTT 5′
Transcript (RNA)
5′ CCCAGCCCGCCUAAUGAGCGGGCUUUUUUUU–OH 3′
Transcript folded to form termination hairpin
Mutations
Mutations
5′– CCCA – U UUUUUUU – OH 3′
G
Deletion

Type II (ρ-dependent)

• lack the poly (U) region, the protein ρ is required for termination. the rho (ρ) factor binds to the RNA chain and pulls it away from the RNA polymerase and the DNA template, resulting in the termination of transcription.

Transcription in Eukaryotes

- Eukaryotes contain three different RNA polymerases:

1. RNA polymerase I, located in the nucleolus, transcribes the three major rRNAs (28S, 18S, and 5.8S).

2. RNA polymerase II, located in the nucleoplasm, transcribes mRNAs and some snRNAs.

3. RNA polymerase III, located in the nucleoplasm, transcribes tRNAs, 5S rRNA.

- All known eukaryotic RNA polymerases have multiple subunits. An example is yeast RNA pol II with 12 subunits

- In eukaryotes, promoters contain some combination of the following:

a) contain a TATA rich region or **Hogness box** located –25 to -30 from the start of transcription

b) Upstream from the TATA region is a variably located sequence containing the sequence CCAAT (frequently at –75)

c) GC box

d) other sequences located either upstream or downstream that maximize the level of transcription called **enhancers**. They can exert their stimulatory actions over distances of several thousand base pairs.

Production of the mRNA molecule:

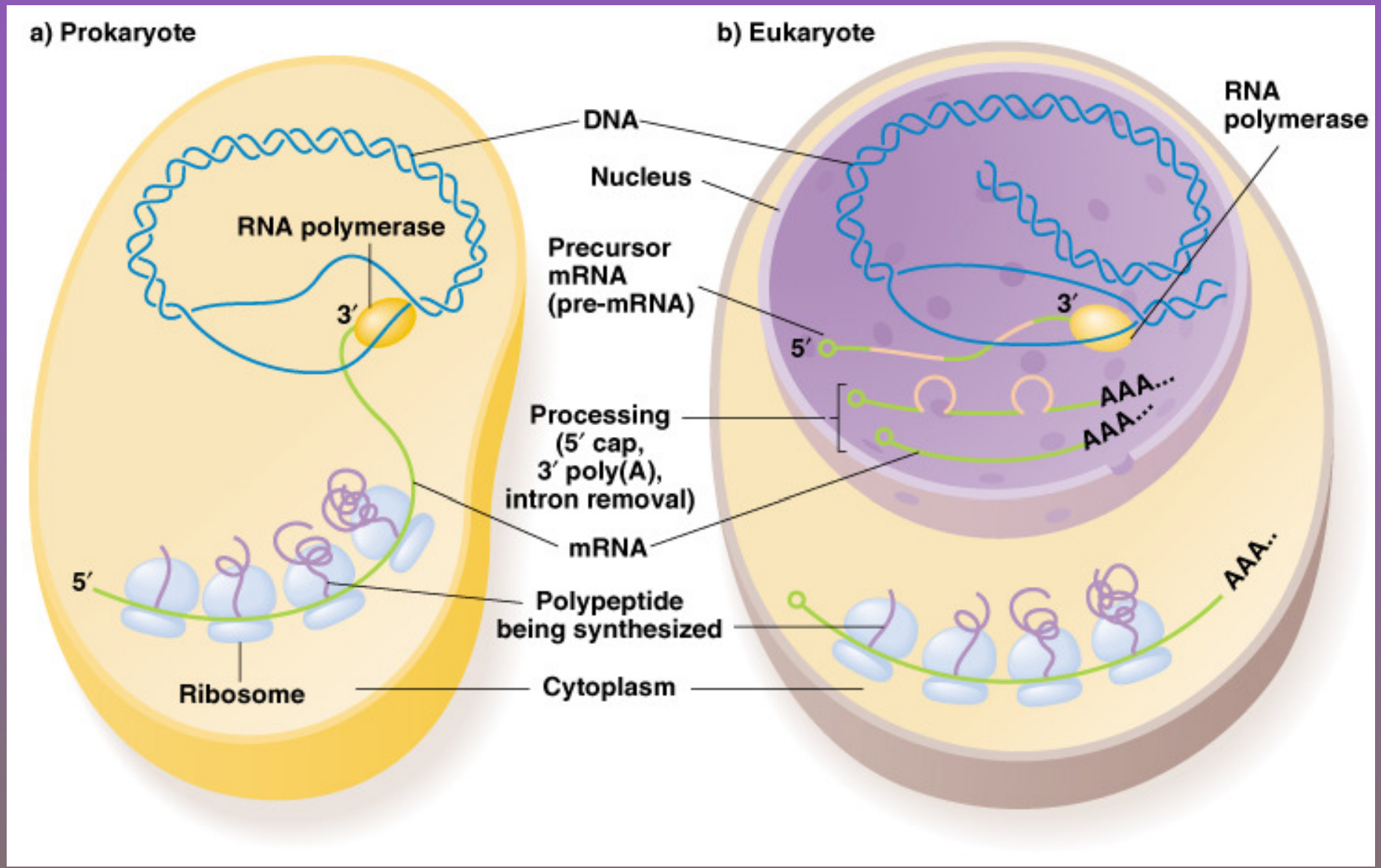
Three main parts:

1. 5’ untranslated region (UTR) leader sequence
2. Coding sequence, specifies amino acids to be translated
3. 3’ untranslated region (UTR) trailer sequence

- **Eukaryotes and prokaryotes produce mRNAs somewhat differently**

  a. **Prokaryotes use the RNA transcript as mRNA without modification. Transcription and translation are coupled in the cytoplasm. mRNA may be polycistronic (encoding several polypeptide).**

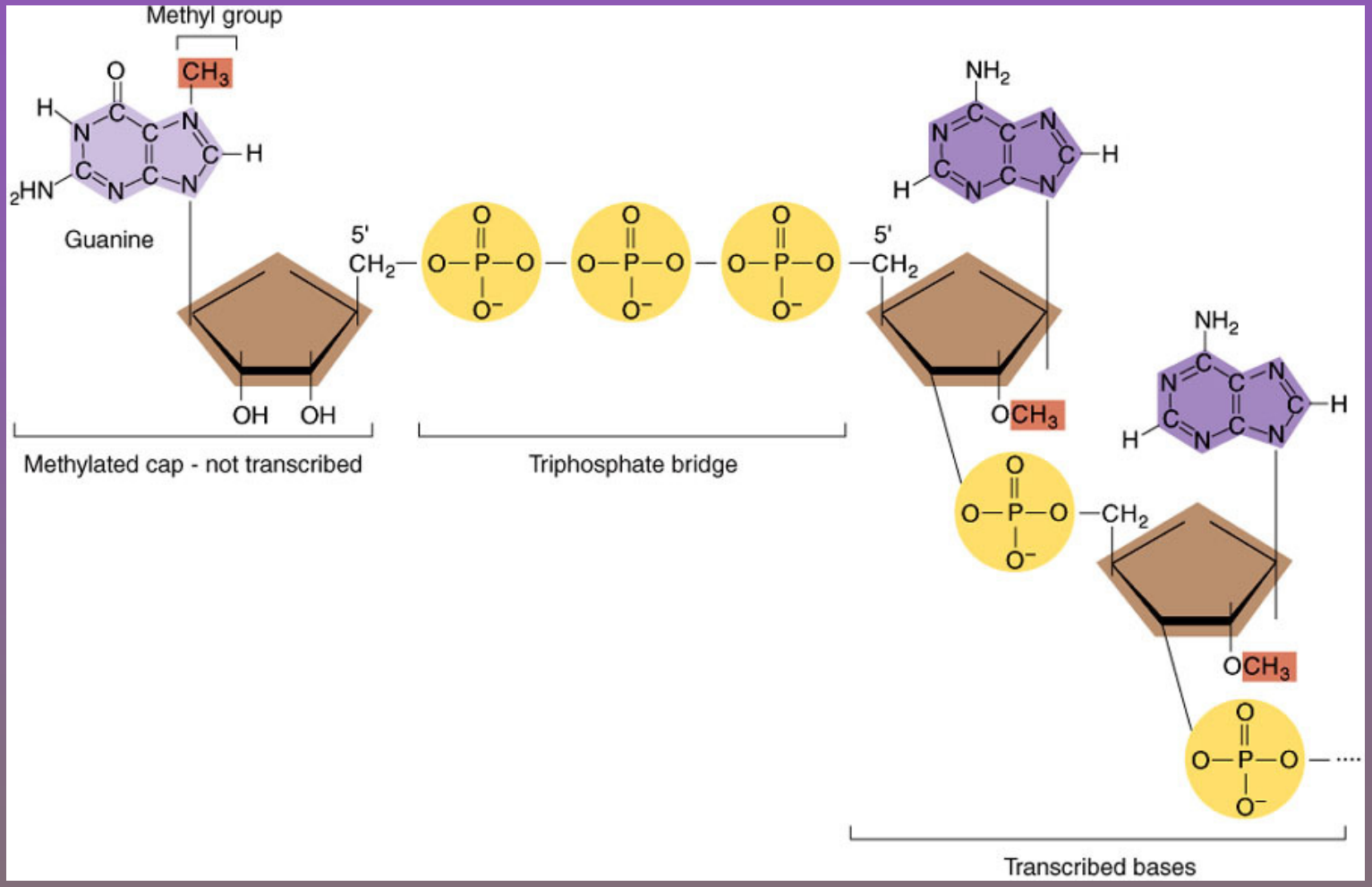
  b. **Eukaryotes modify pre-RNA into mRNA by RNA processing. The processed mRNA migrates from nucleus to cytoplasm before translation. mRNA is monocistronic (encoding a single polypeptide) .**


a) Prokaryote
b) Eukaryote
DNA
RNA polymerase
Nucleus
RNA polymerase
Precursor mRNA (pre-mRNA)
3′
5′
AAA...
AAA...
Processing (5′ cap, 3′ poly(A), intron removal)
mRNA
5′
Polypeptide being synthesized
AAA..
Ribosome
Cytoplasm

Processing of mRNA

The immediate product of transcription is referred to as primary transcript or pre-mRNA or hnRNA (heterogenous nuclear RNA).

1. **Capping of 7-MeG at 5′ end**: a cap of methylated guanosine (7-MeG), is added to the 5′ end of the hnRNA using a 5'-to-5' linkage. The cap is used for ribosome binding to the mRNA during translation initiation.

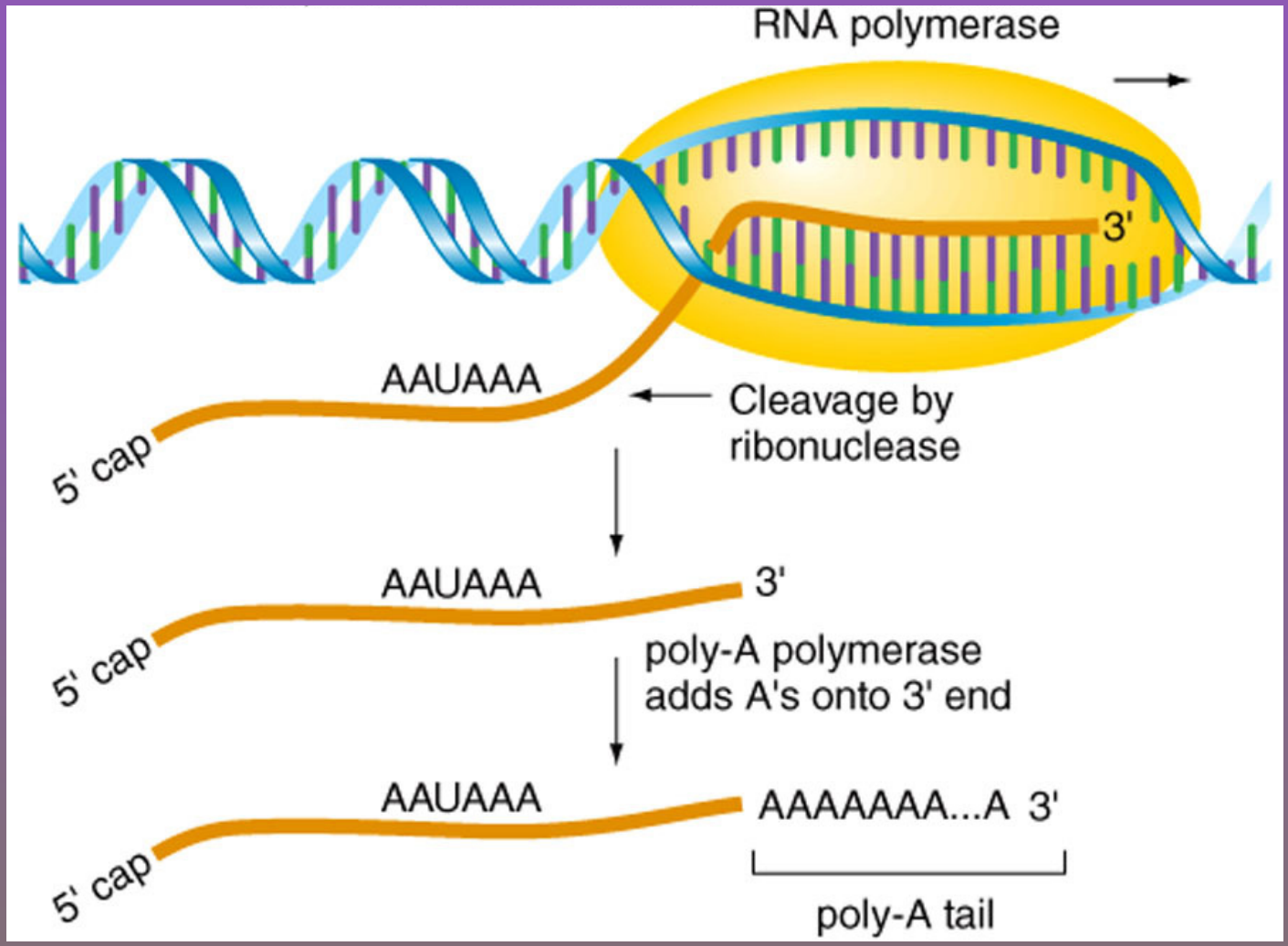

Methyl group
CH3
Guanine
5'
CH2
OH
OH
Methylated cap - not transcribed
Triphosphate bridge
NH2
OCH3
Transcribed bases

2. Addition of poly-A tail at 3′ end:

AAUAAA sequence in the RNA signals a cleavage event in the RNA.

Poly (A) polymerase then adds 150-200 A residues are added to the 3′ end of the mRNA.

The poly-A tail enhances translation efficiency and the stability of the mRNA.


RNA polymerase
3'
AAUAAA
5' cap
Cleavage by
ribonuclease
AAUAAA
3'
5' cap
poly-A polymerase
adds A's onto 3' end
AAUAAA
AAAAAAA...A 3'
5' cap
poly-A tail

3. RNA Splicing:

Eukaryote pre-mRNAs often have intervening introns that must be removed during RNA processing.

intron = non-coding DNA sequences between exons in a gene.

exon = expressed DNA sequences in a gene, code for amino acids.

1. Introns typically begin with a 5'-GU and end with AG-3'.

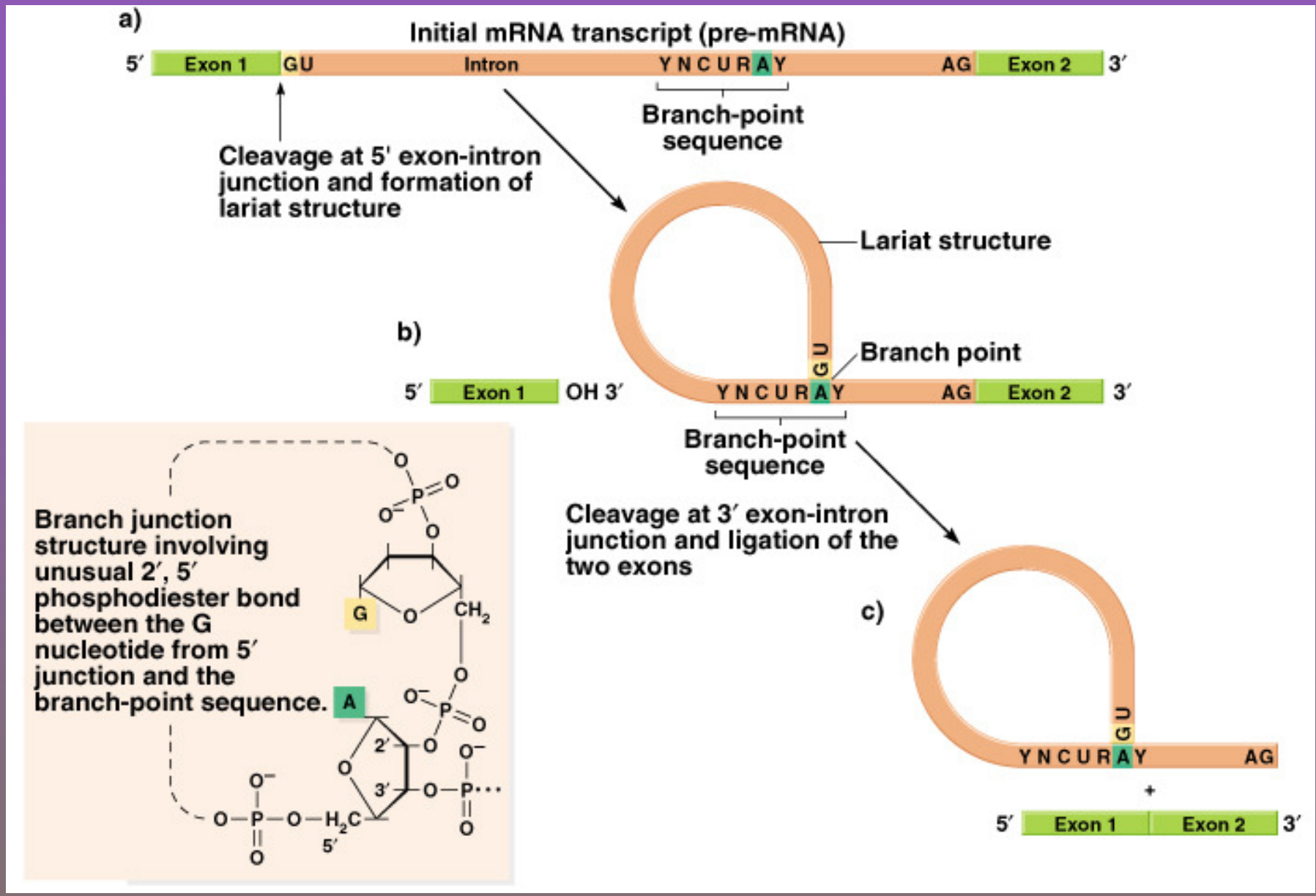
2. Cleavage occurs first at the 5' end of intron 1 (between 2 exons).

3. The now free G joins with an A at a specific branch point sequence in the middle of the intron, using a 2' to 5' phosphodiester bond.

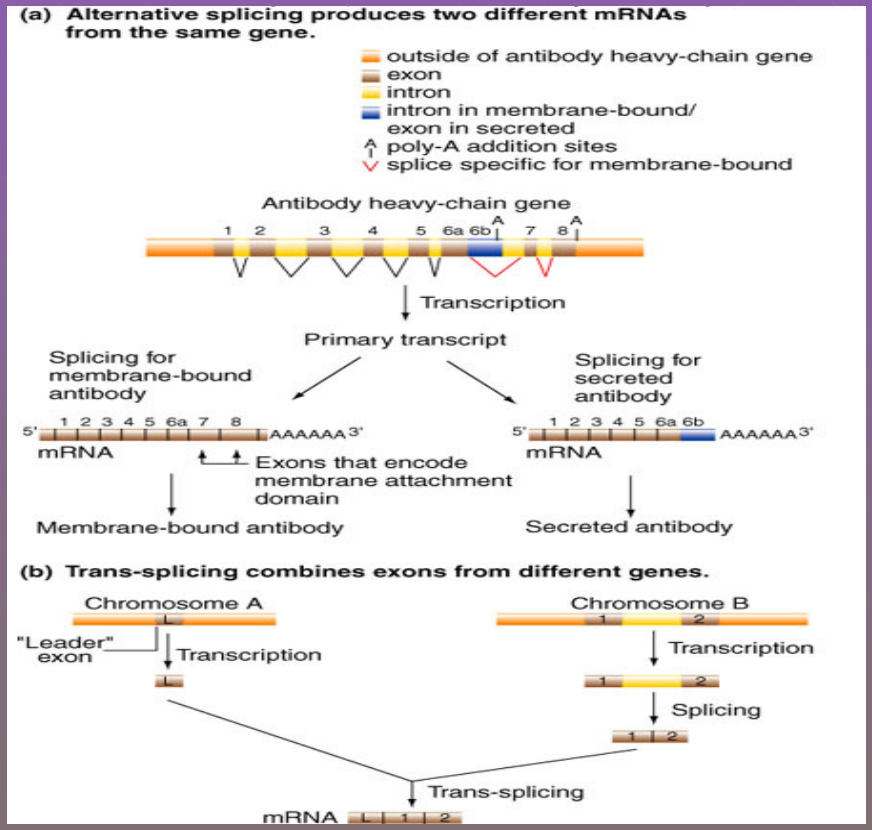
4. Intron forms a lariat-shaped structure. Lariat is excised, and the exons are joined to form a spliced mRNA.

5. Splicing is mediated by splicosomes, complexes of small nuclear RNAs (snRNAs) and proteins, that cleave the intron at the 3' end and join the exons.

6. Introns are degraded by the cell.


a)
Initial mRNA transcript (pre-mRNA)
5′
Exon 1
GU
Intron
Y N C U R A Y
AG
Exon 2
3′
Branch-point sequence
Cleavage at 5' exon-intron junction and formation of lariat structure
Lariat structure
b)
Branch point
5′
Exon 1
OH 3′
Y N C U R A Y
AG
Exon 2
3′
Branch-point sequence
Cleavage at 3′ exon-intron junction and ligation of the two exons
c)
Y N C U R A Y
AG
5′
Exon 1
Exon 2
3′
Branch junction structure involving unusual 2′, 5′ phosphodiester bond between the G nucleotide from 5′ junction and the branch-point sequence.
G
A

Sometimes two or more functional mRNA are developed from a single hnRNA due to difference in splicing pattern, called as **alternative splicing**. Thus, due to alternative splicing multiple protein products can be produced from a single gene.



Characteristics of the genetic code (written as in mRNA, 5' to 3'):

1. The code is triplet: Each mRNA codon (word) that specifies a particular amino acid in a polypeptide chain consists of three nucleotides (letters).

2. The code is non-overlapping: The mRNA encoding one protein is read in successive groups of three nucleotides.

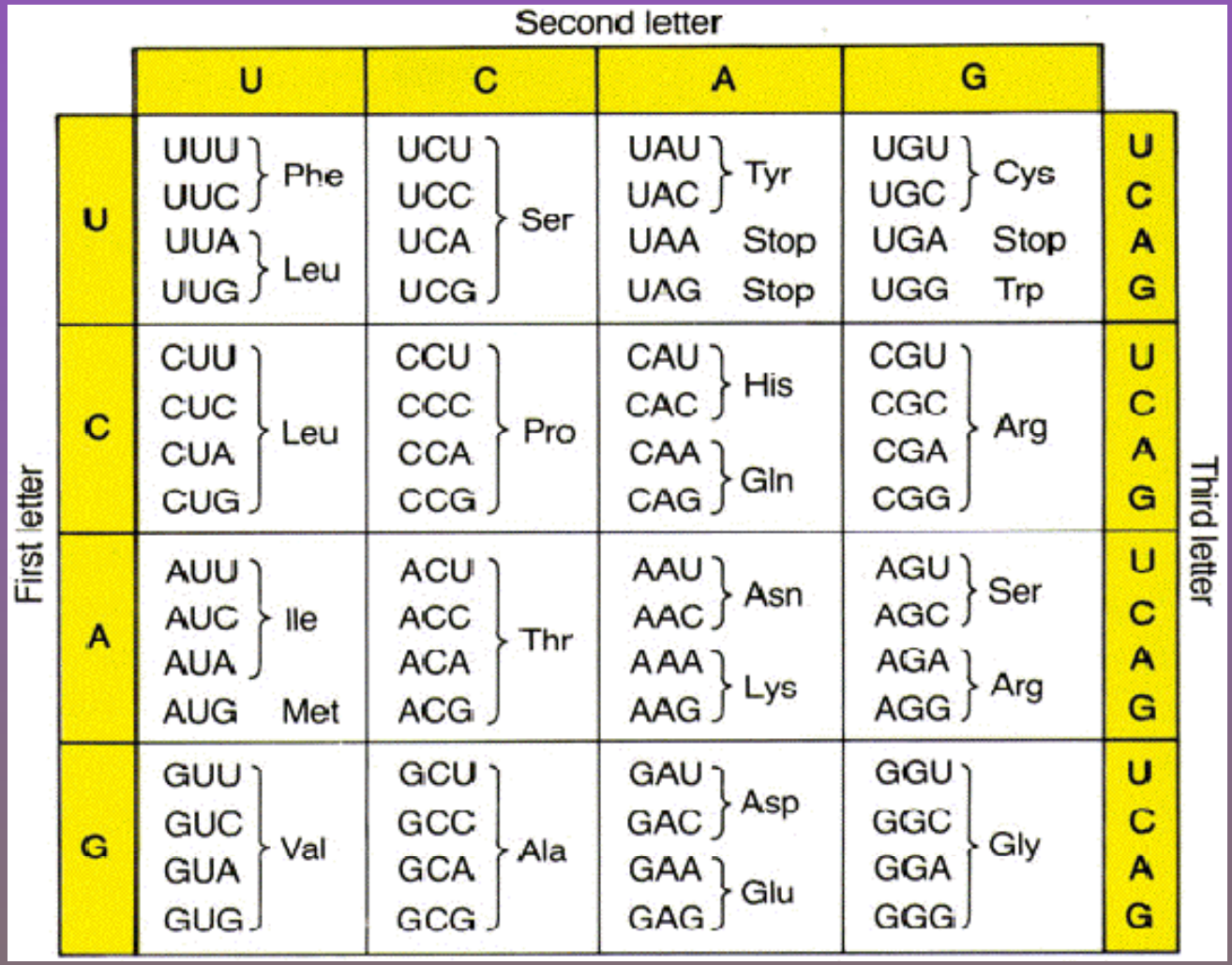
3. Code is comma free: mRNA is read continuously, 3 bases at a time without skipping bases

4. Code is almost universal: Most codons have the same meaning in different organisms.

5. The code is degenerate: More than one codon code for the same amino acids. Only two amino acids; tryptophan and methionine are coded by one codon each.

6. The genetic code is unambiguous: each triplet codon is believed to have only one meaning.

7. Code has start and stop signals: AUG codes for Met and is the usual start signal. UAA, UAG, and UGA are stop codons and specify the end of translation of a polypeptide.


Second letter
U
C
A
G
First letter
Third letter
U
UUU Phe
UUC
UUA Leu
UUG
UCU Ser
UCC
UCA
UCG
UAU Tyr
UAC
UAA Stop
UAG Stop
UGU Cys
UGC
UGA Stop
UGG Trp
C
CUU Leu
CUC
CUA
CUG
CCU Pro
CCC
CCA
CCG
CAU His
CAC
CAA Gln
CAG
CGU Arg
CGC
CGA
CGG
A
AUU Ile
AUC
AUA
AUG Met
ACU Thr
ACC
ACA
ACG
AAU Asn
AAC
AAA Lys
AAG
AGU Ser
AGC
AGA Arg
AGG
G
GUU Val
GUC
GUA
GUG
GCU Ala
GCC
GCA
GCG
GAU Asp
GAC
GAA Glu
GAG
GGU Gly
GGC
GGA
GGG

Wobble hypothesis:

- ✓ Proposed by Francis Crick in 1966.
- ✓ Occurs at 3’ end of codon/5’ end of anti-codon.
- ✓ There are 61 codons which code for amino acids.
- ✓ the number of tRNA molecules which act as anticodon is always much lesser than the number of codons
- ✓ Hence, the wobble would allow the anticodon of a single tRNA type to pair with more than one codon in mRNA, often without changing the amino acid

	5′ anti-codon	3′ codon
G	pairs with	U or C
C	pairs with	G
A	pairs with	U
U	pairs with	A or G
I (Inosine)	pairs with	A, U, or C

I = post-transcription modified purine