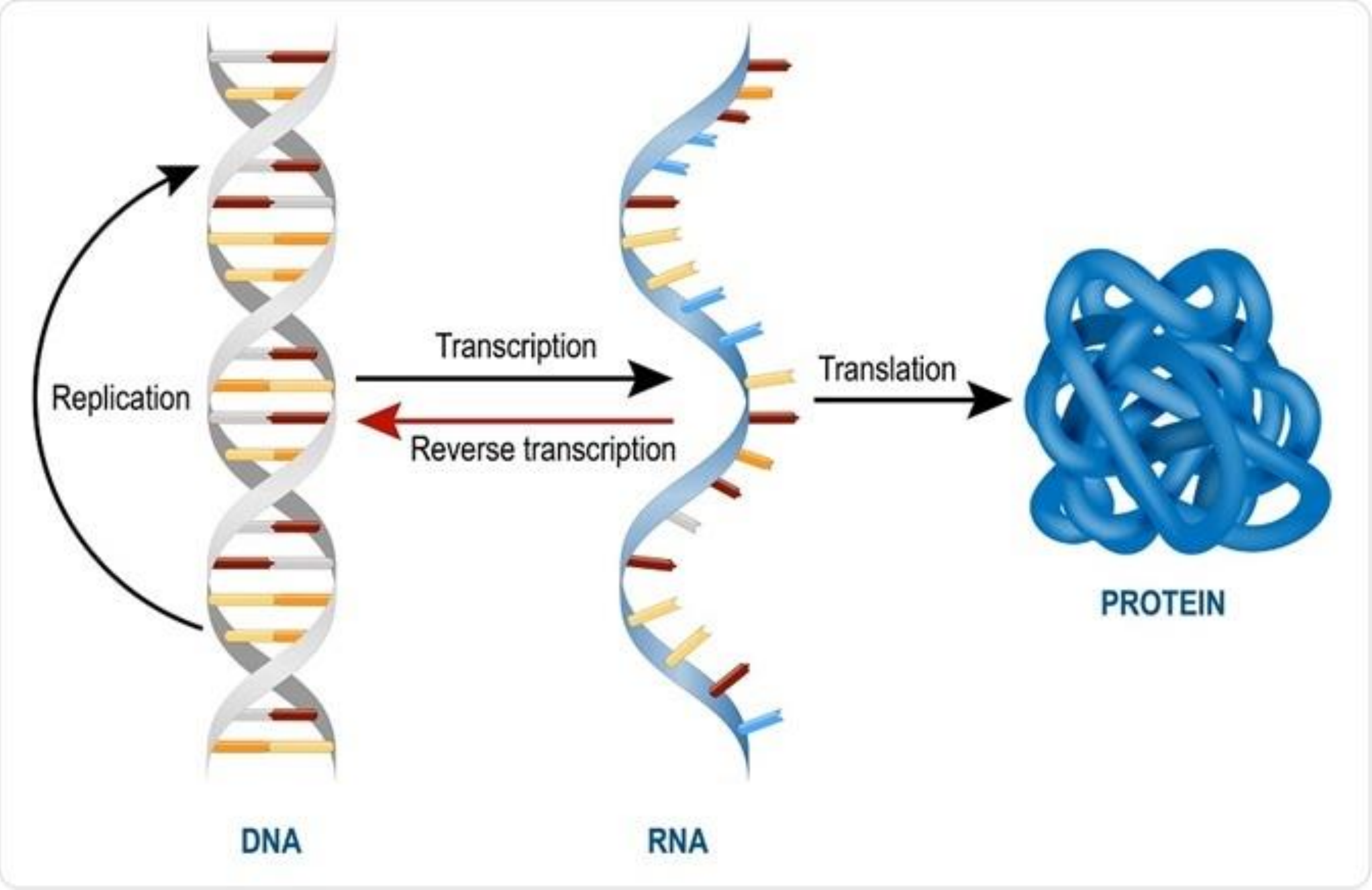
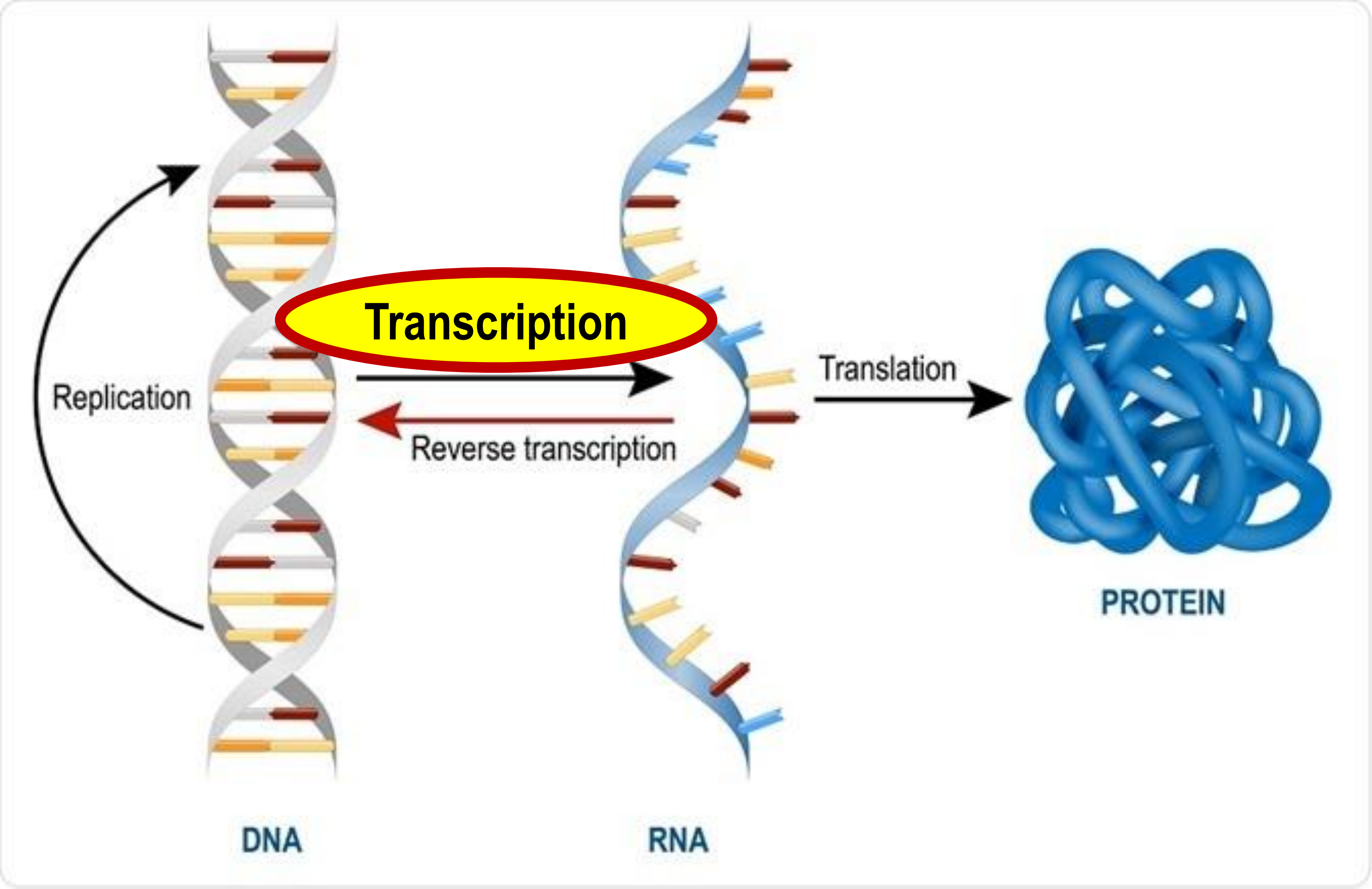
Transcription, translation.

RNDr. Markéta Janatová, Ph.D.
Institute of Medical Biochemistry and Laboratory Diagnostics,
First Faculty of Medicine, Charles University

Central dogma of molecular biology



Central dogma of molecular biology



RNA biosynthesis

= **transcription** of genetic information from

DNA to RNA

- all types of RNA are produced

- according to the complementarity

- building blocks: NTPs (AGCU)

- in a nucleus

Transcription

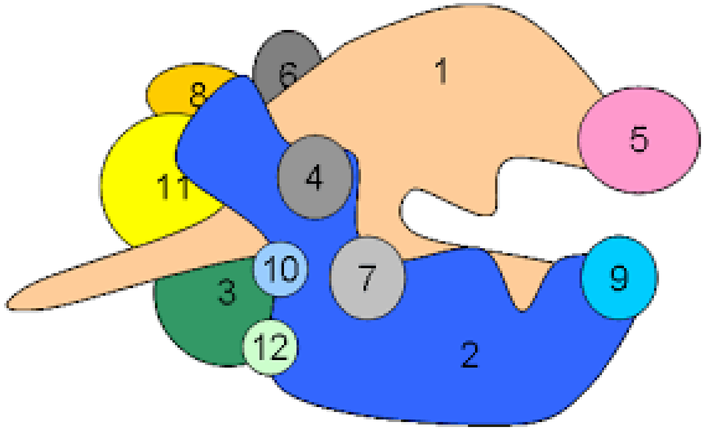
- synthesis of long primary transcripts:

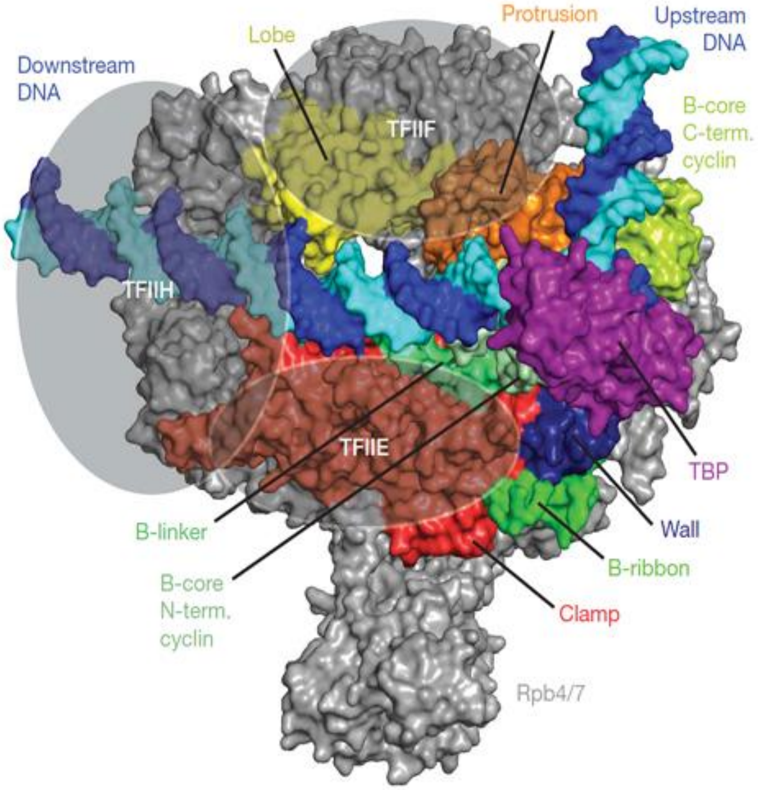
  → **mRNA** – messenger
  → **rRNA** – ribosomal
  → **tRNA** – transfer

- enzyme: **DNA-dependent-RNA-polymerase**

Transcription

- enzyme: **DNA-dependent-RNA-polymerase**





Transcription

- three phases:

1. iniciation – RNA-pol binds DNA via TFs and RNA synthesis is started

2. elongation – synthesis of RNA strand

3. termination – stopping the synthesis and releasing RNA from DNA

Transcription factors

- necessary for transcription initiation

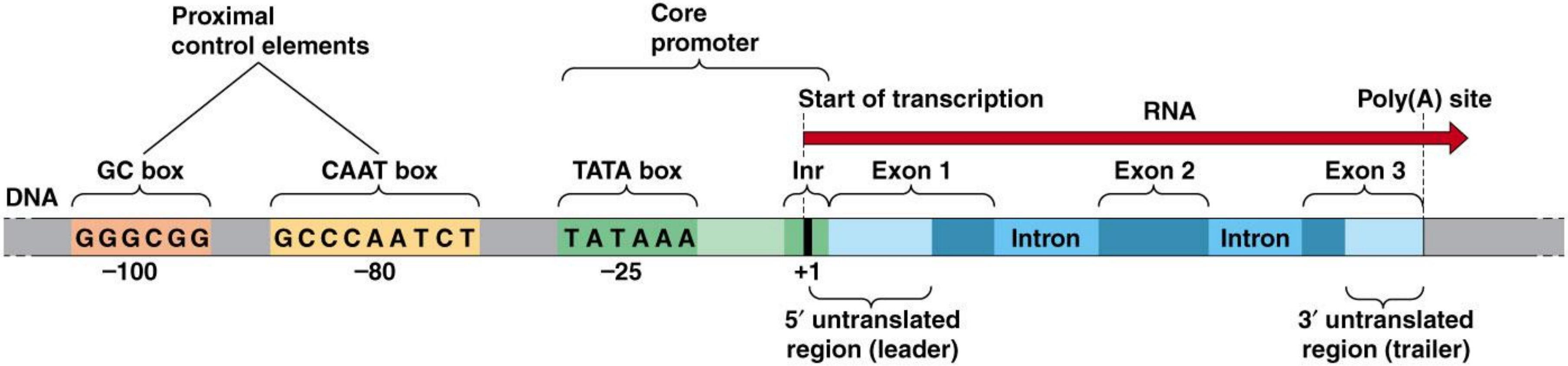
→ recognize a promoter region

→ separate DNA doublehelix

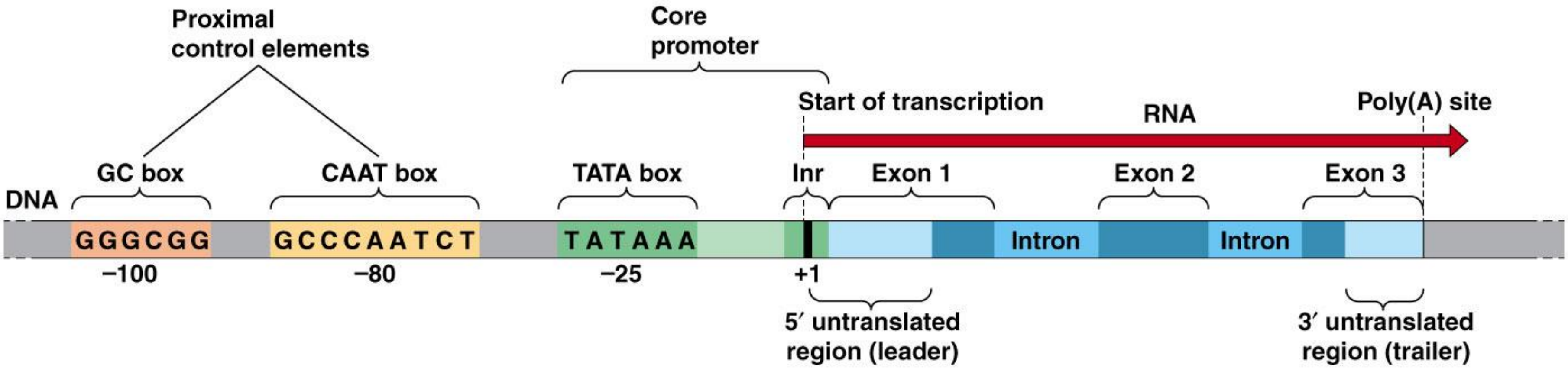
→ activate RNA polymerase

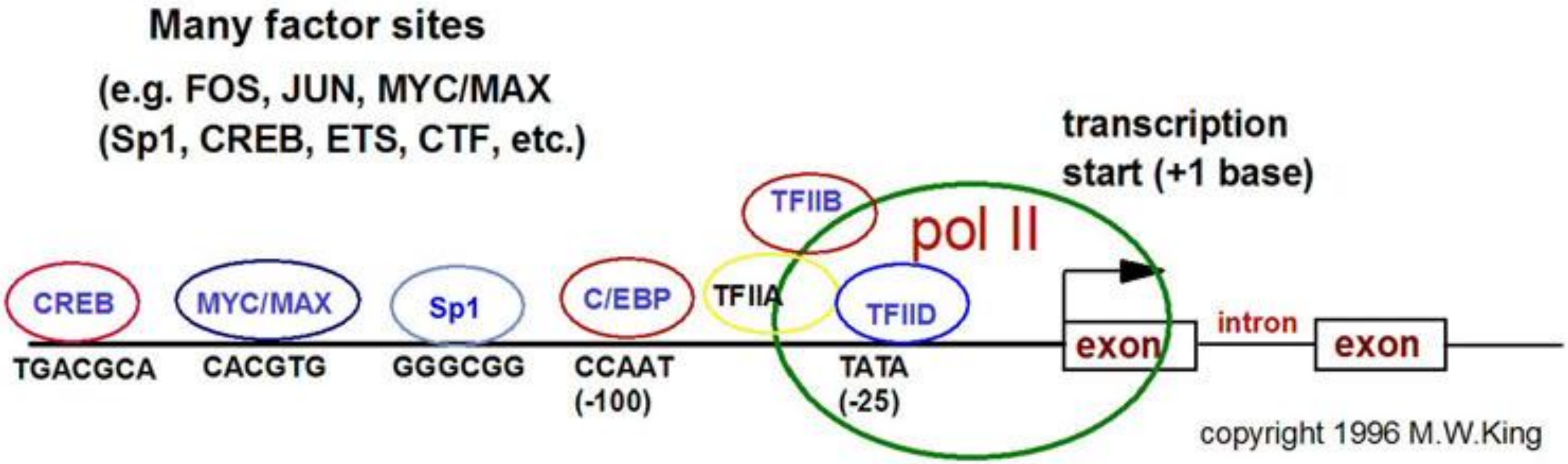
Transcription – promoter

- **promoter contains:** starting nucleotide **A (+1)**
  - **TATA box:** -34_-26, binds TFIID – TBP (TATA binding protein)
  - **CAAT box:** -75_-80, binds TF CTF/NF1
  - **GC box:** -100, more copies, binds TF SP1
  - **Oct box (ATTTGCAT):** binds TF OCT-1/2



Transcription – promoter





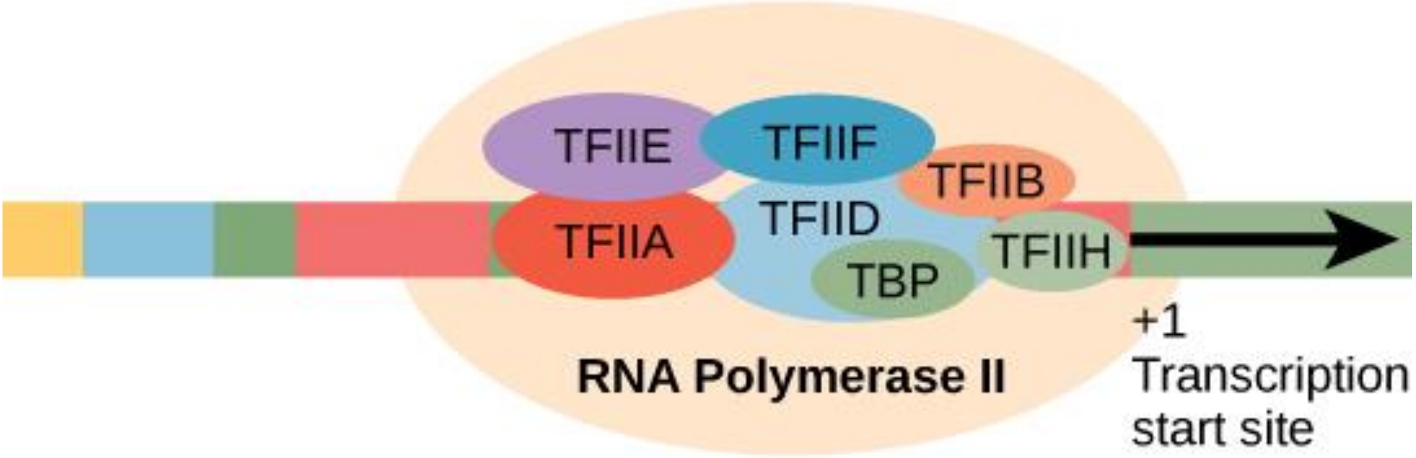
copyright 1996 M.W.King

Transcription – initiation

- **initiation complex of RNA polymerase and TFs**

→ helicase activity – separates DNA

→ proteinkinase activity – activation of polymerase



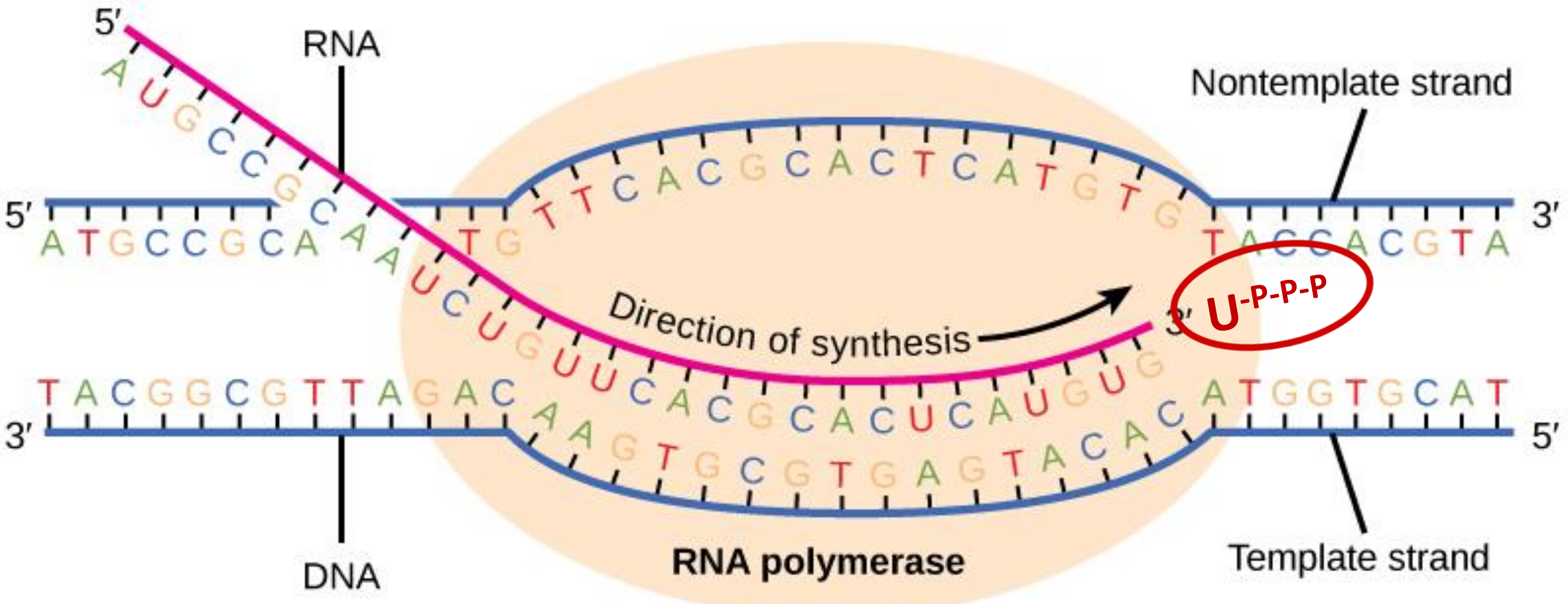
- primer is not needed

Transcription – elongation

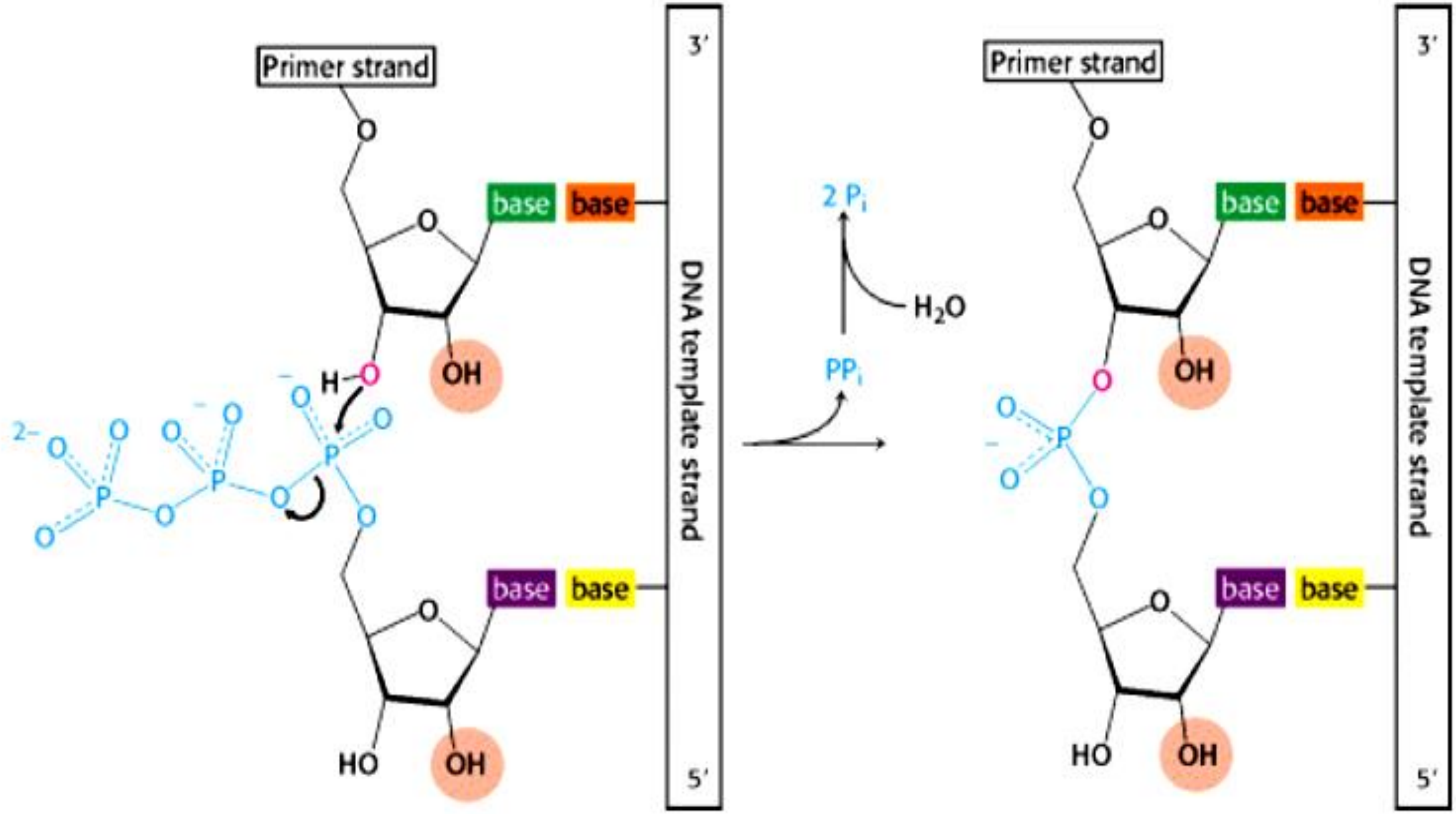
- RNA synthesis according to the complementarity
- **RNA synthesis proceeds in the 5’-3’ direction**
- **RNA-Pol moves in the 3’-5’ direction along the non-coding DNA strand**
- substrates: **NTPs**
- **phosphodiester bond** is formed

Transcription – elongation

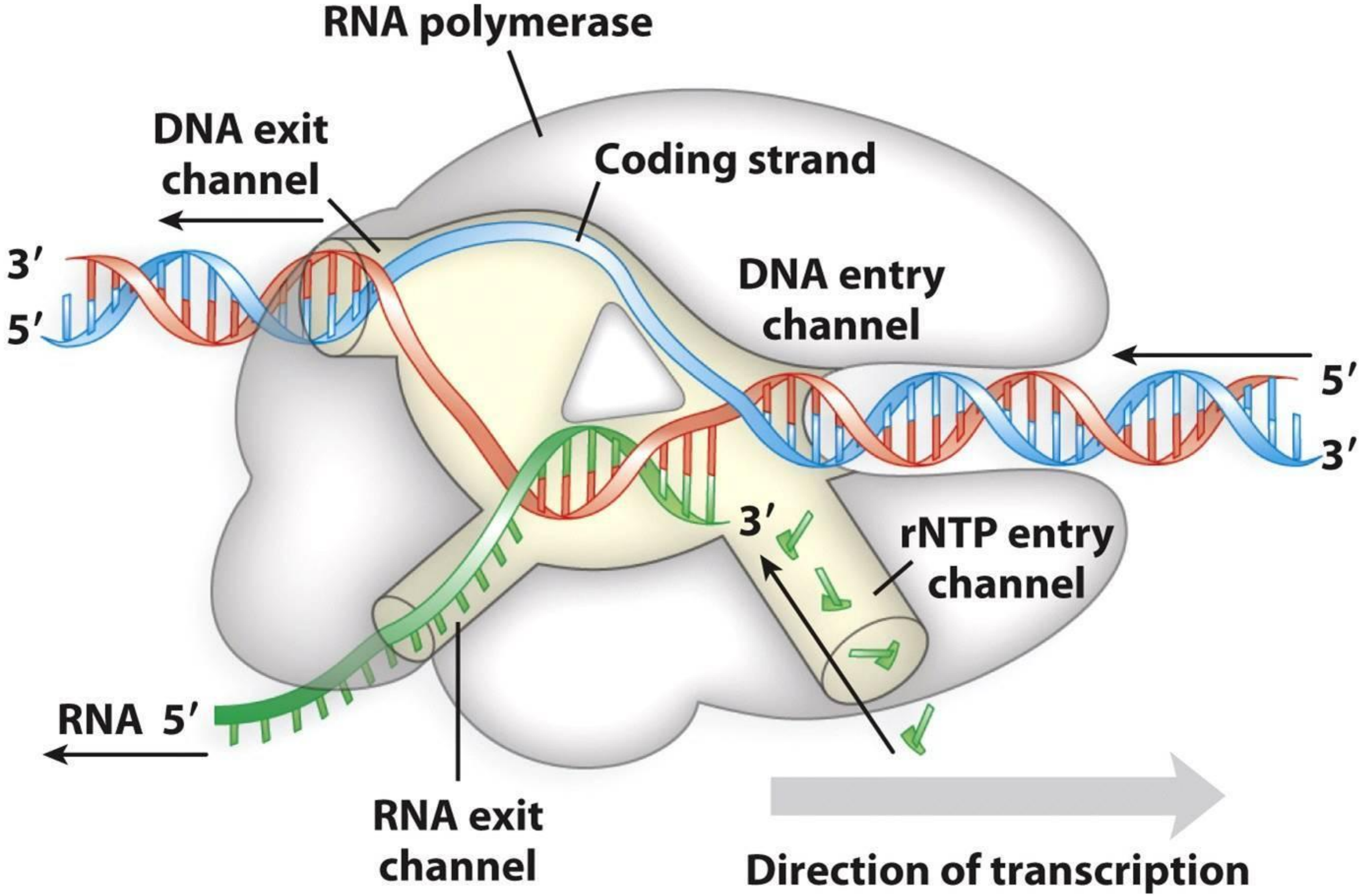
- RNA synthesis according to the complementarity



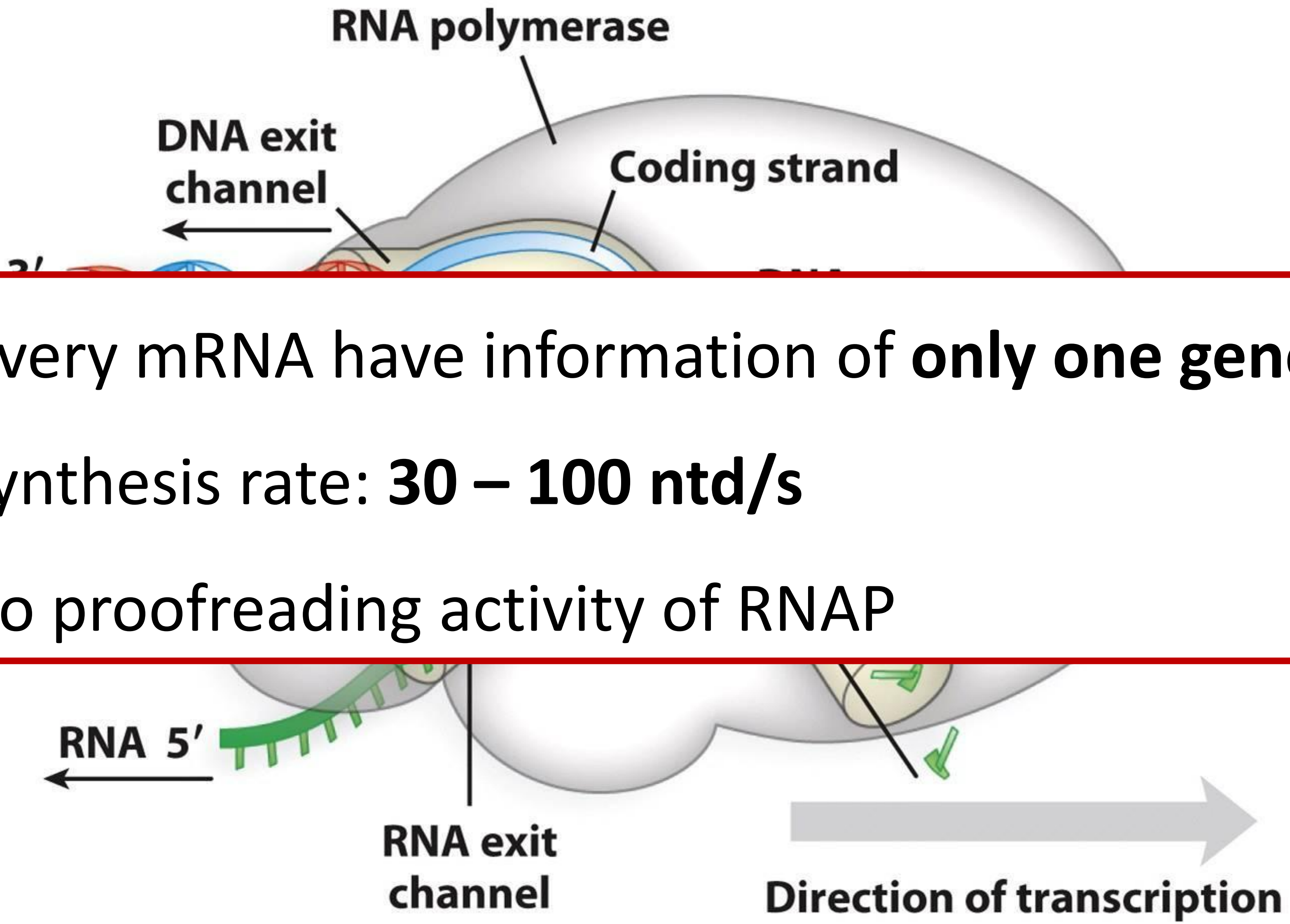
Transcription – elongation



Transcription – elongation



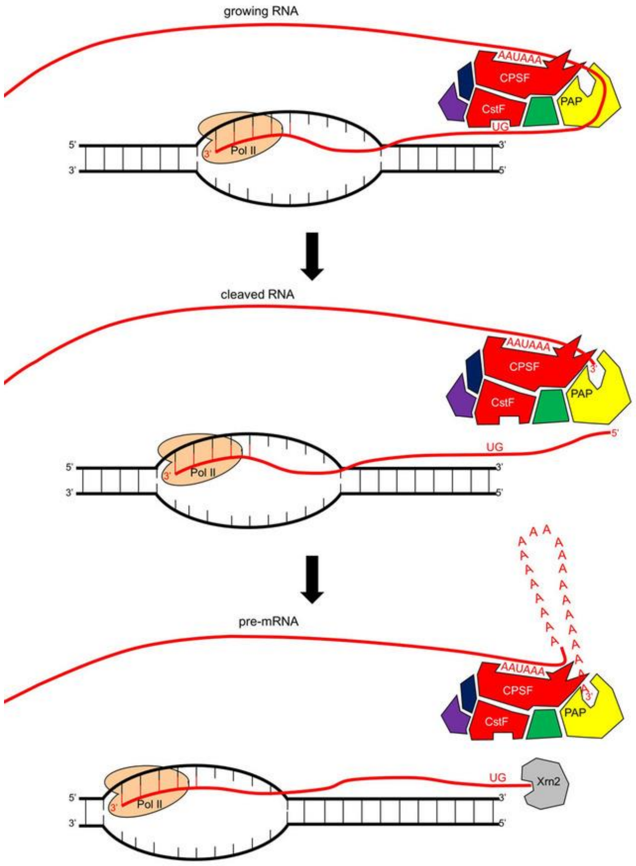
Transcription – elongation



- every mRNA have information of **only one gene**
- synthesis rate: **30 – 100 ntd/s**
- no proofreading activity of RNAP

Transcription – termination

- **polyadenylation signal AAUAAA**
- mRNA is cleaved 10 – 30 nt beyond the signal and released from DNA
- polyadenylation by **poly(A)polymerase**



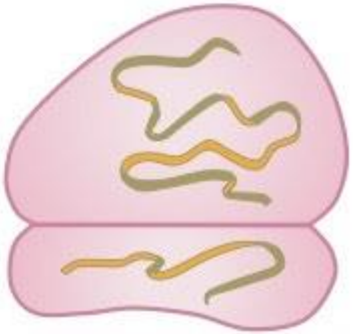
Transcription – result

→ mRNA

Messenger RNA (mRNA)

→ 2-5%

→ rRNA



Ribosomal RNA (rRNA)

→ 70%

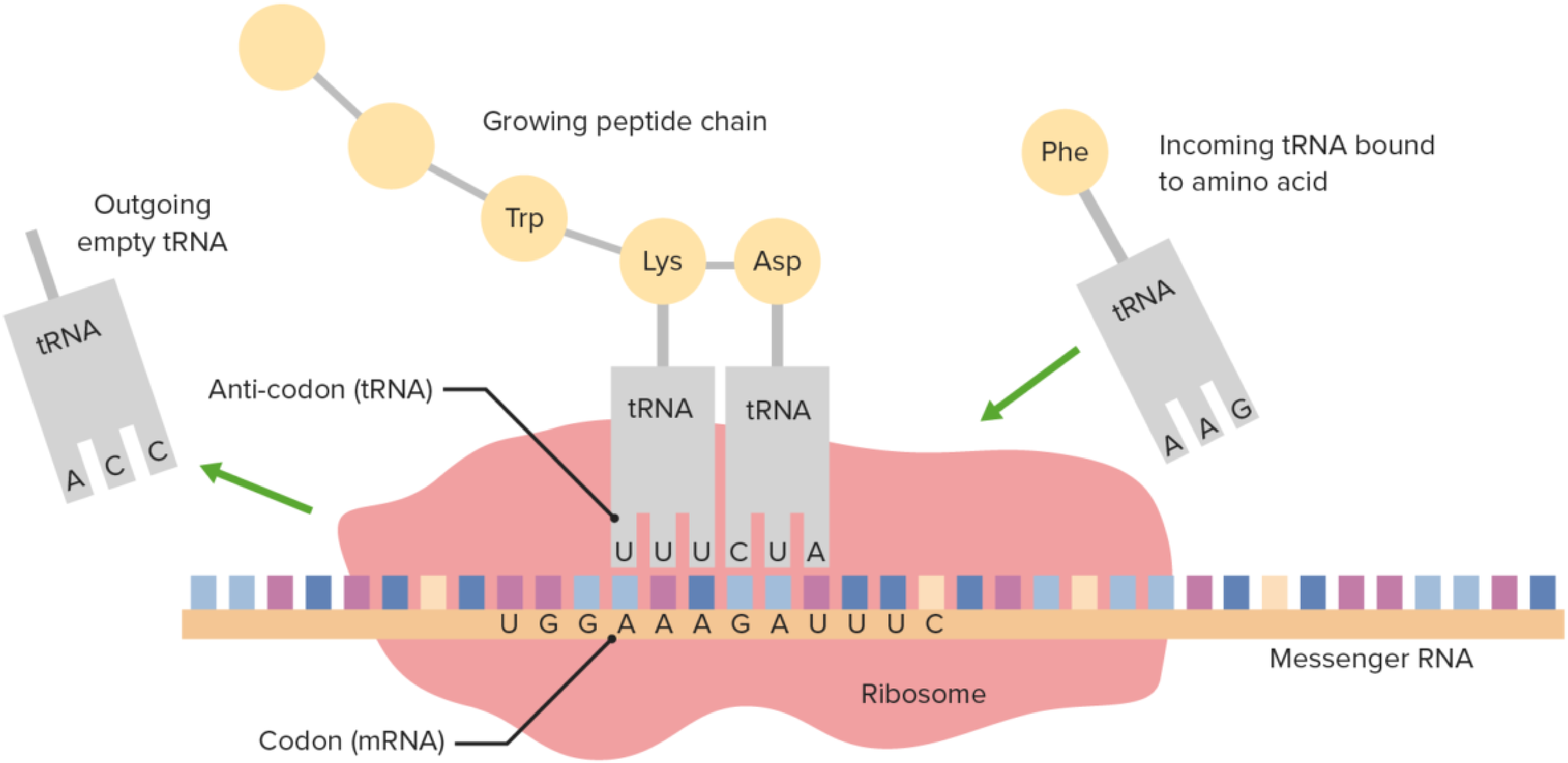
→ tRNA

Transfer RNA (tRNA)

→ 20%

Transcription – result

→ **participating during translation process**



Peptide synthesis

RNA types

- 3 types of RNA and RNA-polymerase:

1. **RNA-polymerase I:** transkription of **rRNA**
2. **RNA-polymerase II:** transkription of **mRNA**
3. **RNA-polymerase III:** transkription of **tRNA**

- each polymerase binds to its specific promoter and TFs

Posttrc. modifications of mRNA

- transcription of structural genes

- long primary transcripts must be modified

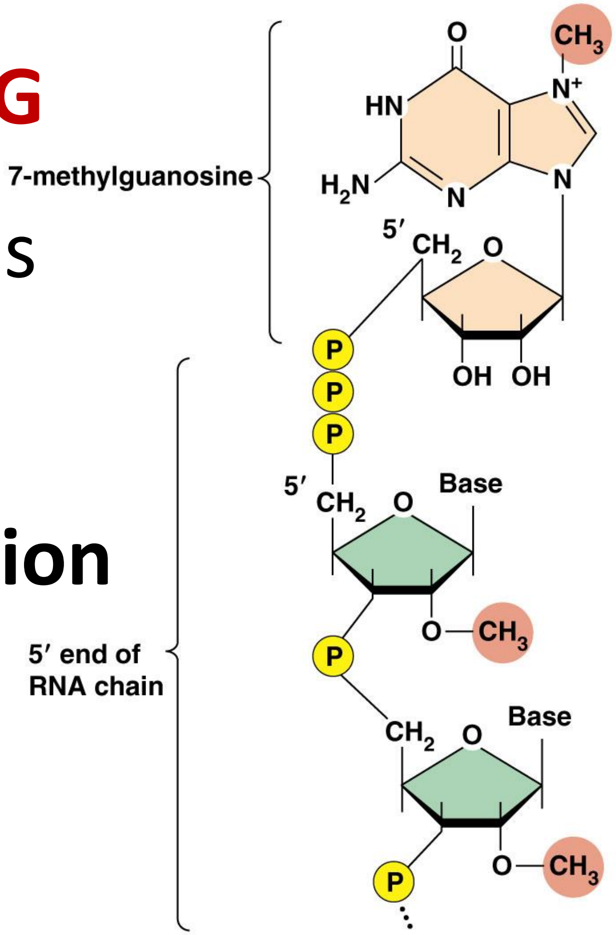
→ capping

→ polyA-tail

→ mRNA splicing

m^7G capping of 5'-end

- **7-methylguanosine** binds by **5'-5' linkage** to 5'-end of mRNA → **m^7G**
- ribose of following nucleotides is methylated
- important for **translation initiation**



© 2012 Pearson Education, Inc.

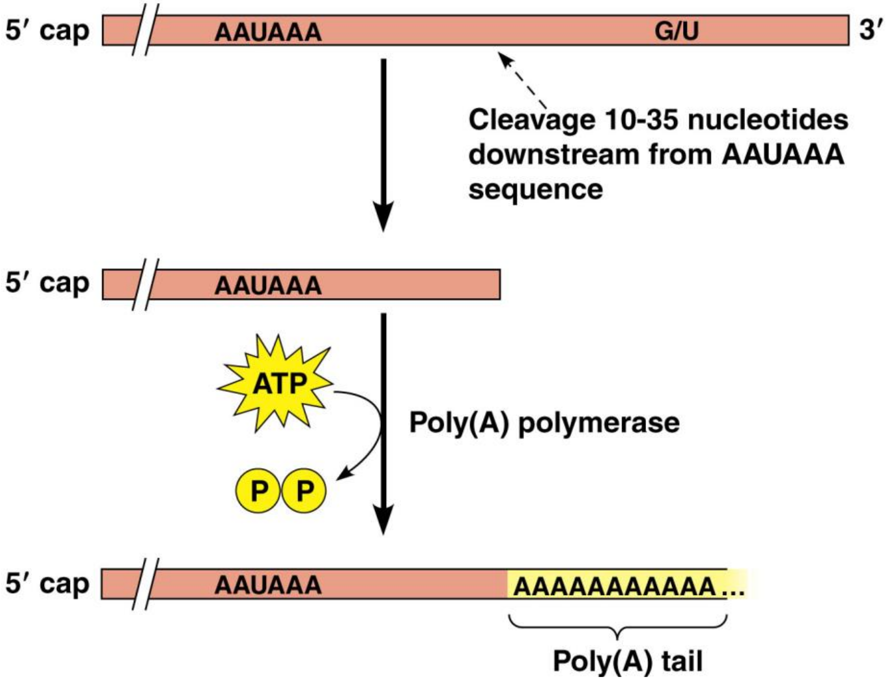
Polyadenylation of mRNA 3’-end

- polyA sequence 50 – 250 nt long at a 3’-end

- is catalyzed by **poly(A)polymerase**

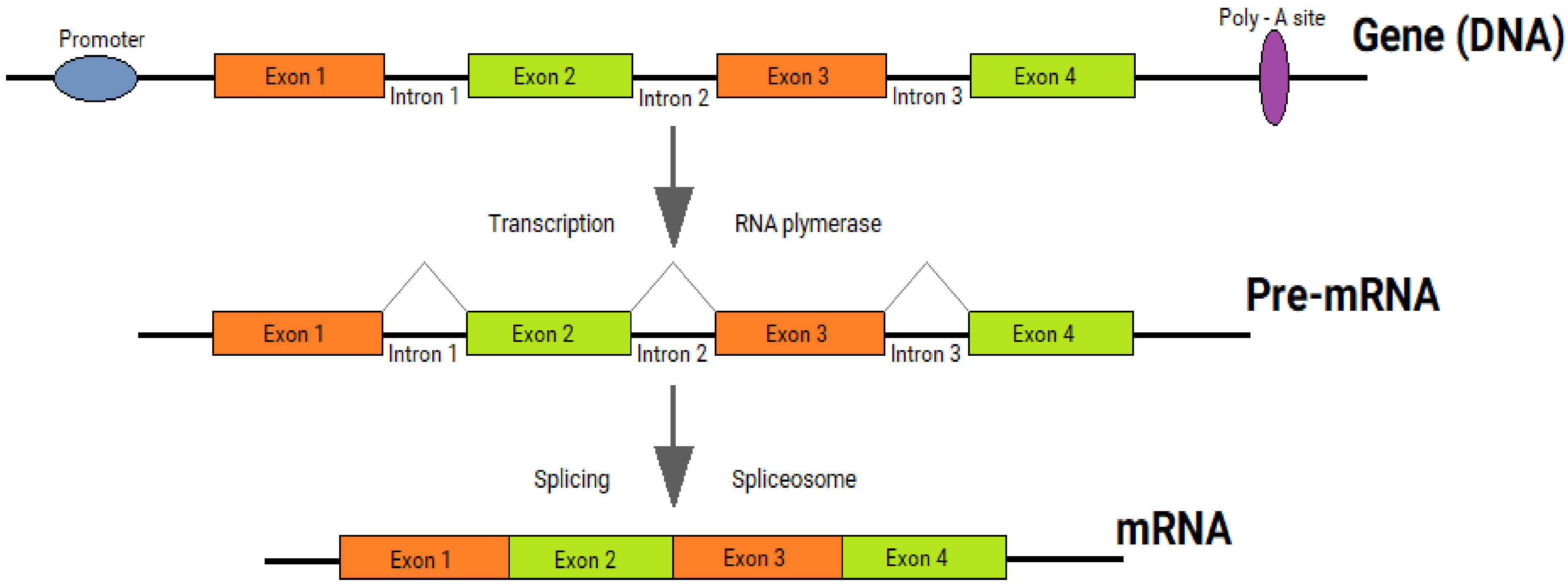
→ mRNA transport into cytoplasm

→ stabilizes 3’-end



mRNA splicing

- RNA contains sequence of **exons** and **introns**



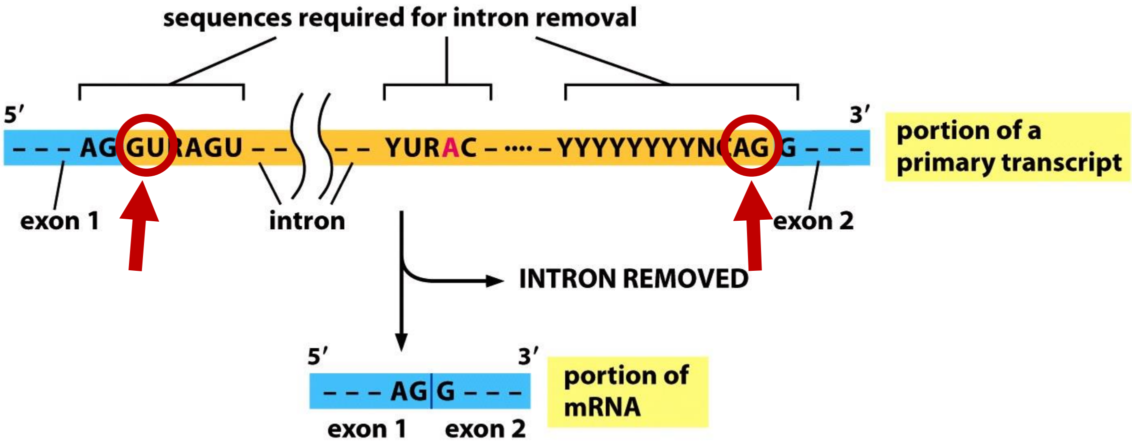
- **introns are much larger!**

mRNA splicing

- RNA contains sequence of exons and introns

- splicing is directed by a primary structure of intron: **5’-GU--…--AG-3’**

→ **5’-splice site, 3’-splice site, branching site (A)**



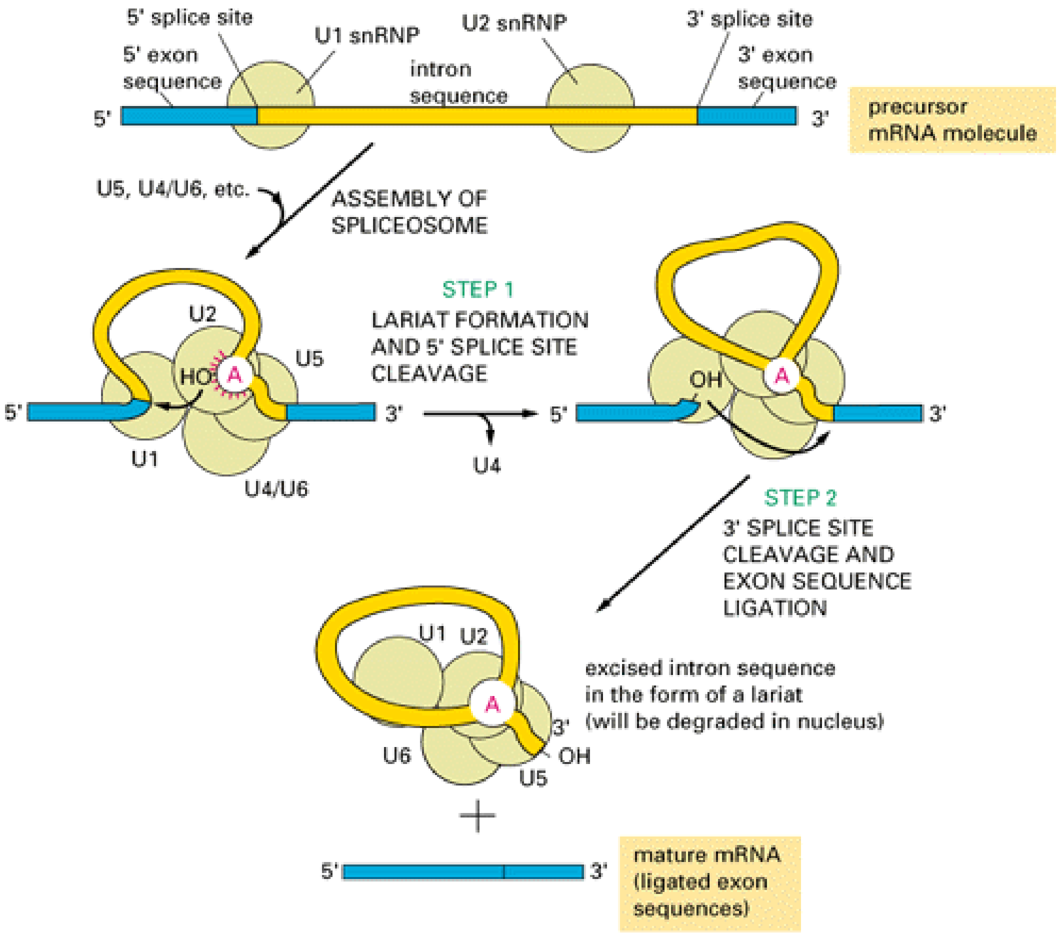
mRNA splicing

- **snRNP** (small nuclear ribonucleoprotein particles)

(U1, U2, U4, U5, U6)

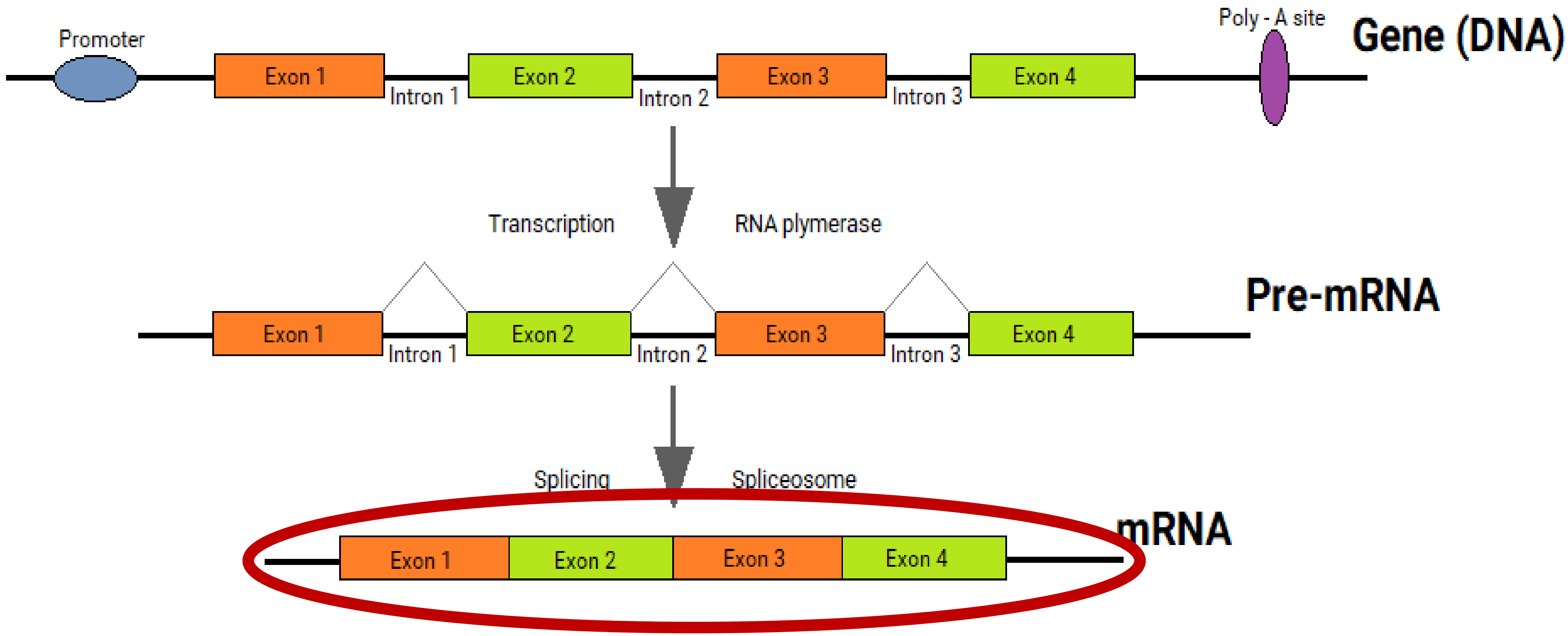
→ **spliceosome**

→ **ribozyme**



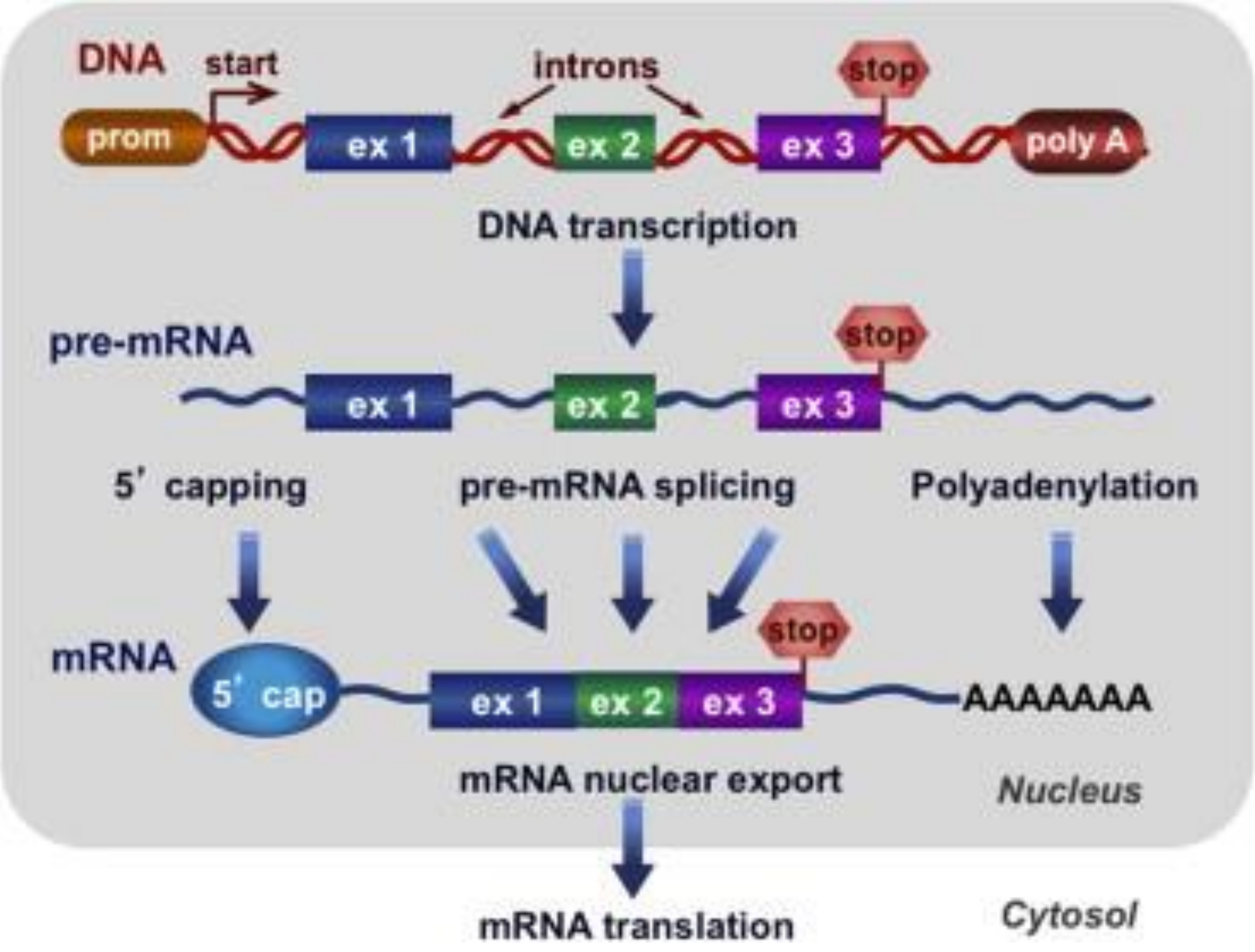
mRNA splicing

- RNA contains sequence of **exons** and **introns**



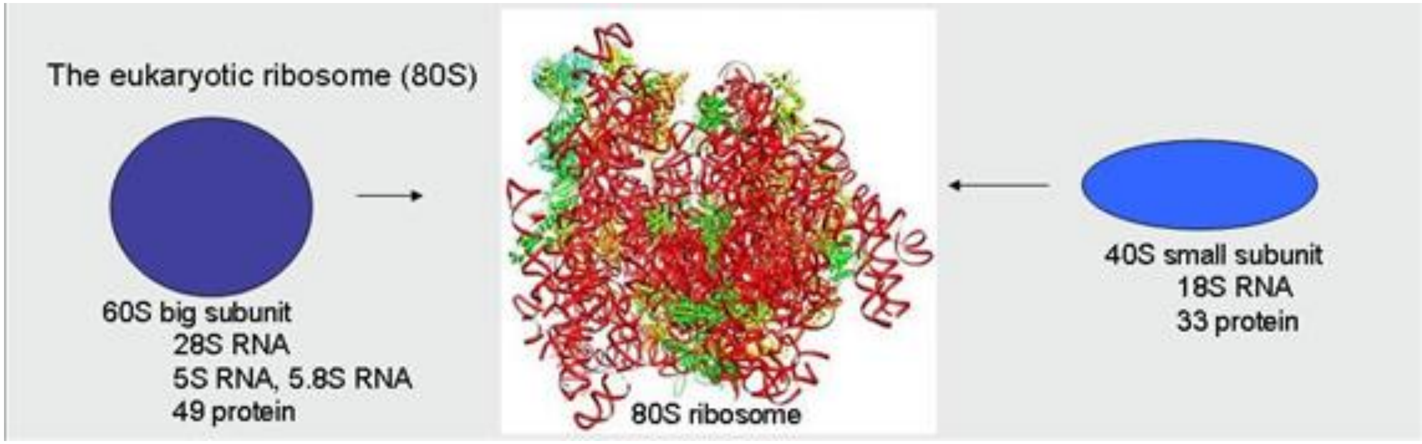
Posttrc. modifications of mRNA

-



rRNA

- component of **ribosome – ribonucleoprotein complex** – protein synthesis via translation

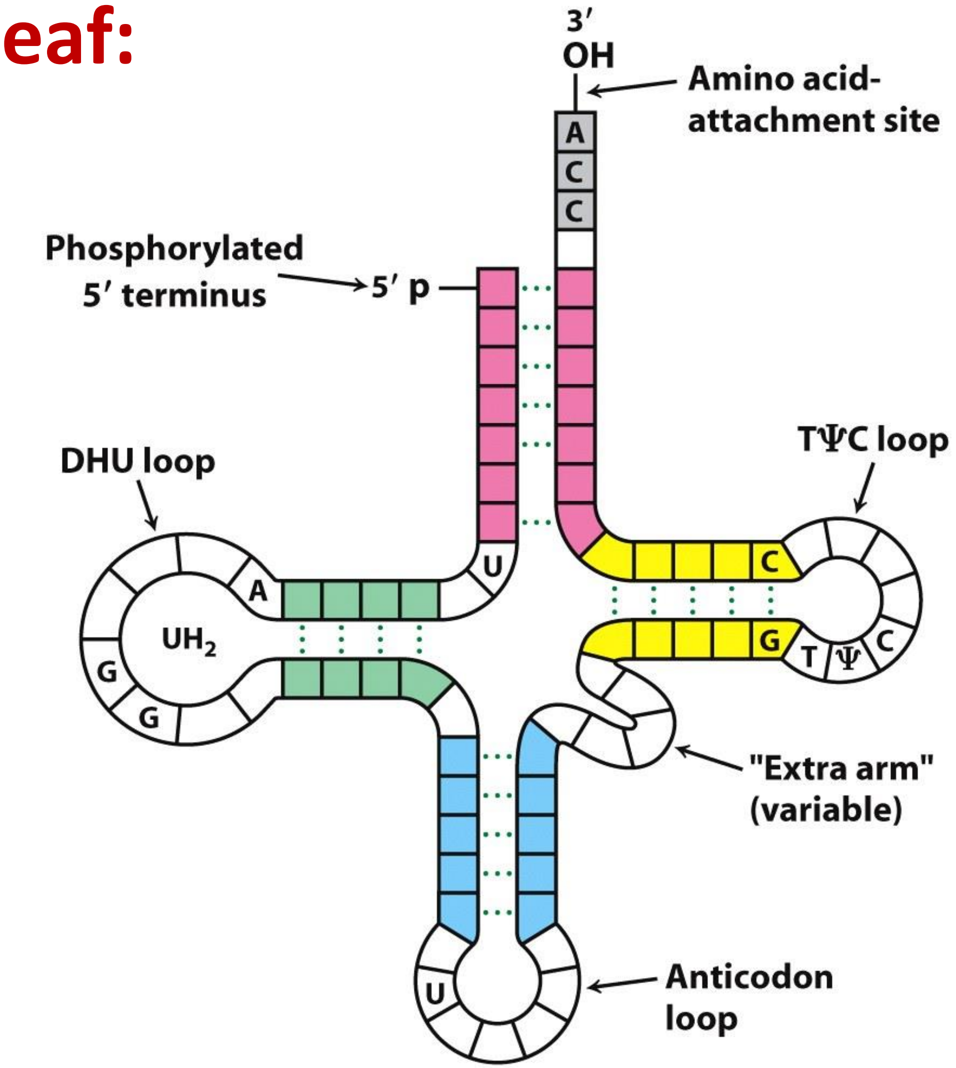


- ***rRNA in ribosome has its own enzymatic peptidyltransferase activity*** → **ribozym**

tRNA

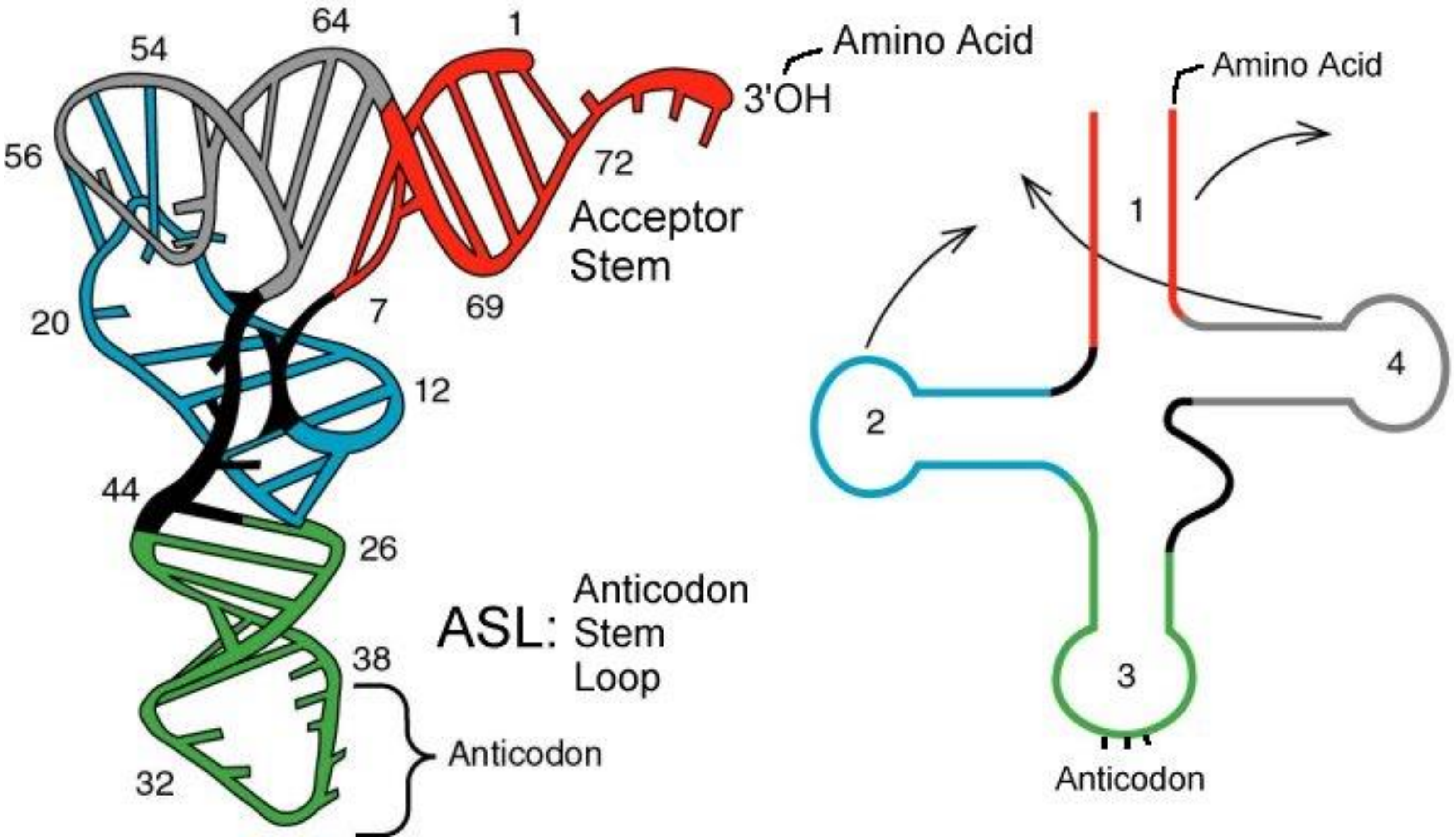
- **2° structure → clover leaf:**

- <u>four loops</u>:
    - acceptor loop
    - pseudouridine loop
    - dihydrouridine loop
    - anticodon loop

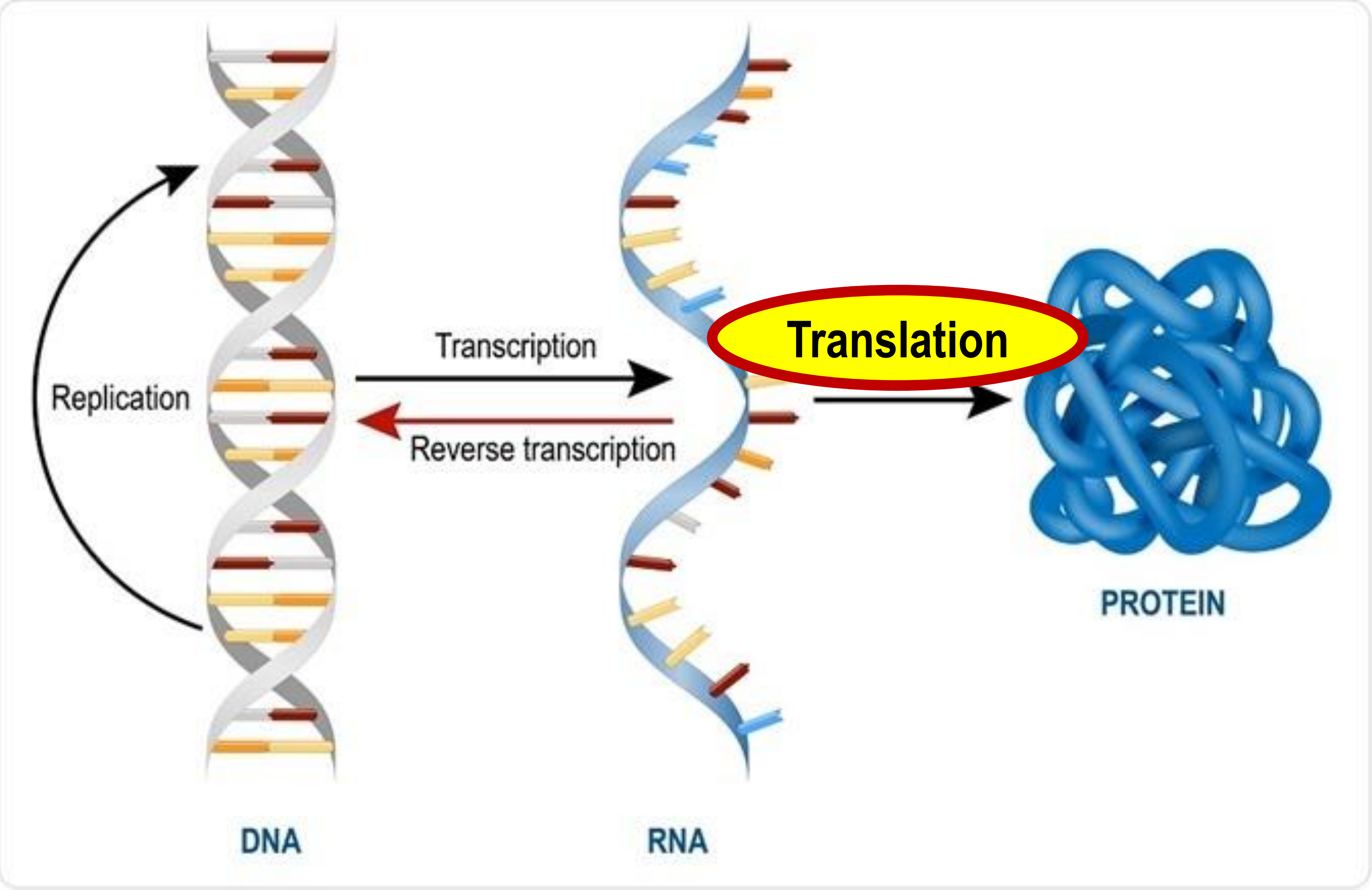


tRNA

- 3° structure → L shape:



Central dogma of molecular biology



Structural genes expression

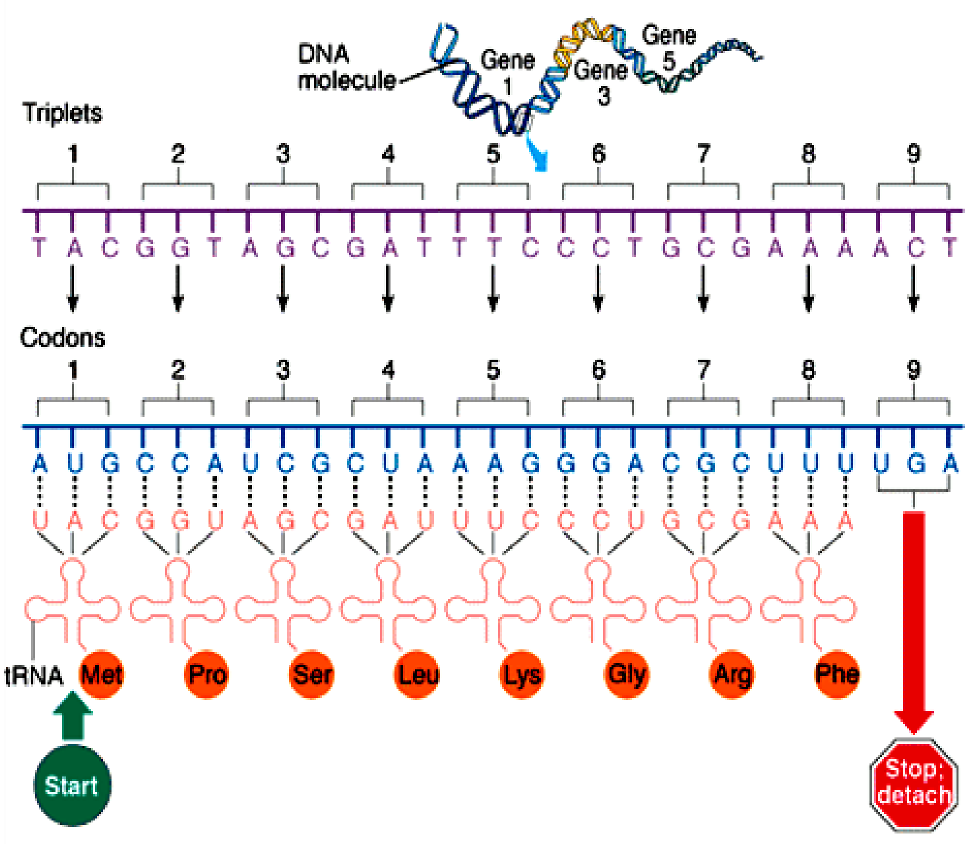
DNA (G, A, C, T)

↓

RNA (G, A, C, U)

↓

20 proteinogenic amino acids

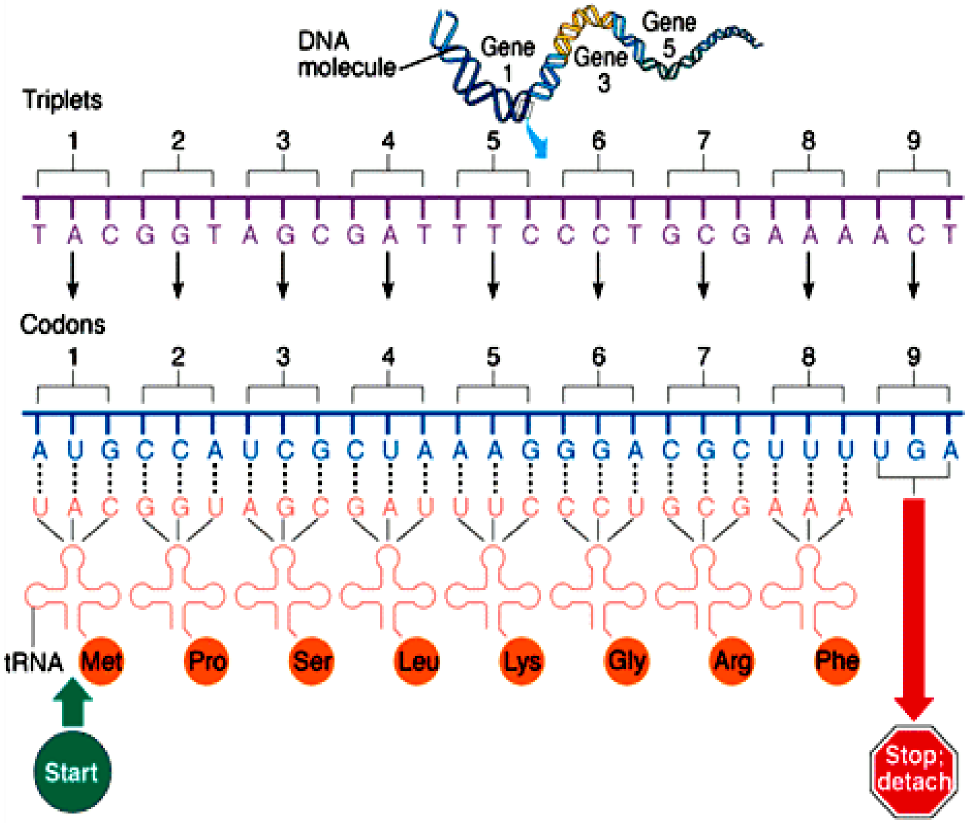


Translation – genetic code

DNA (G, A, C, T)

RNA (G, A, C, U)

20 proteinogenic amino acids



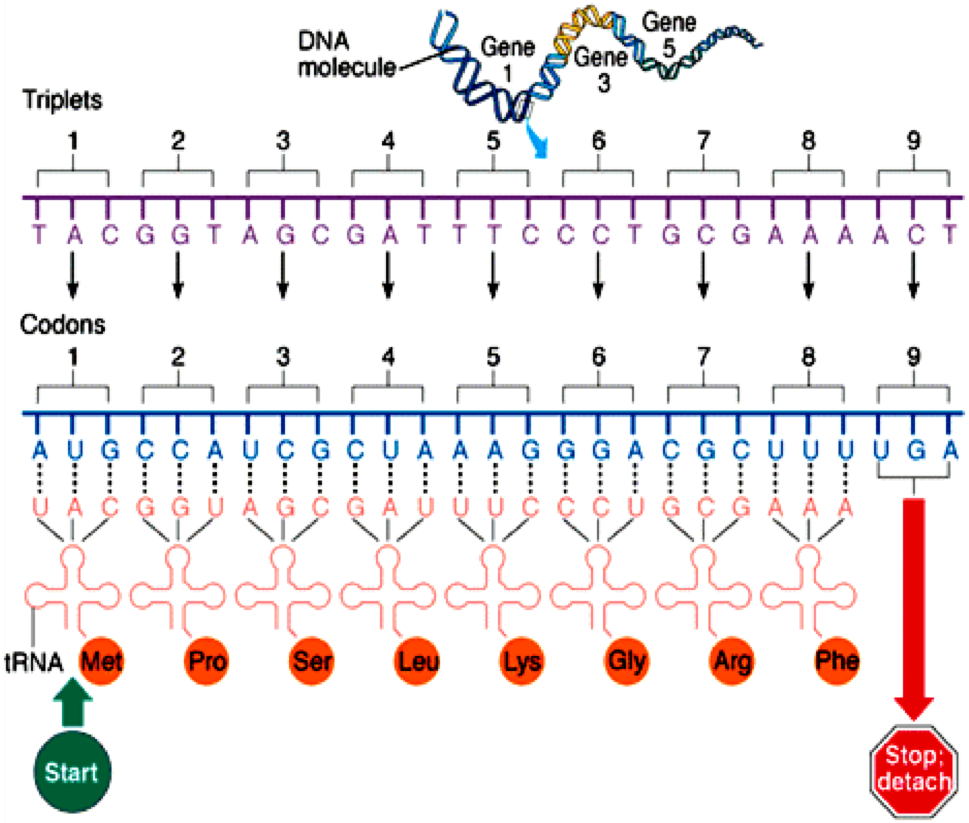
- # of code letters:

Translation – genetic code

DNA (G, A, C, T)

RNA (G, A, C, U)

20 proteinogenic amino acids



- # of code letters: **3 letters and 4 types of ntds**

Translation – genetic code

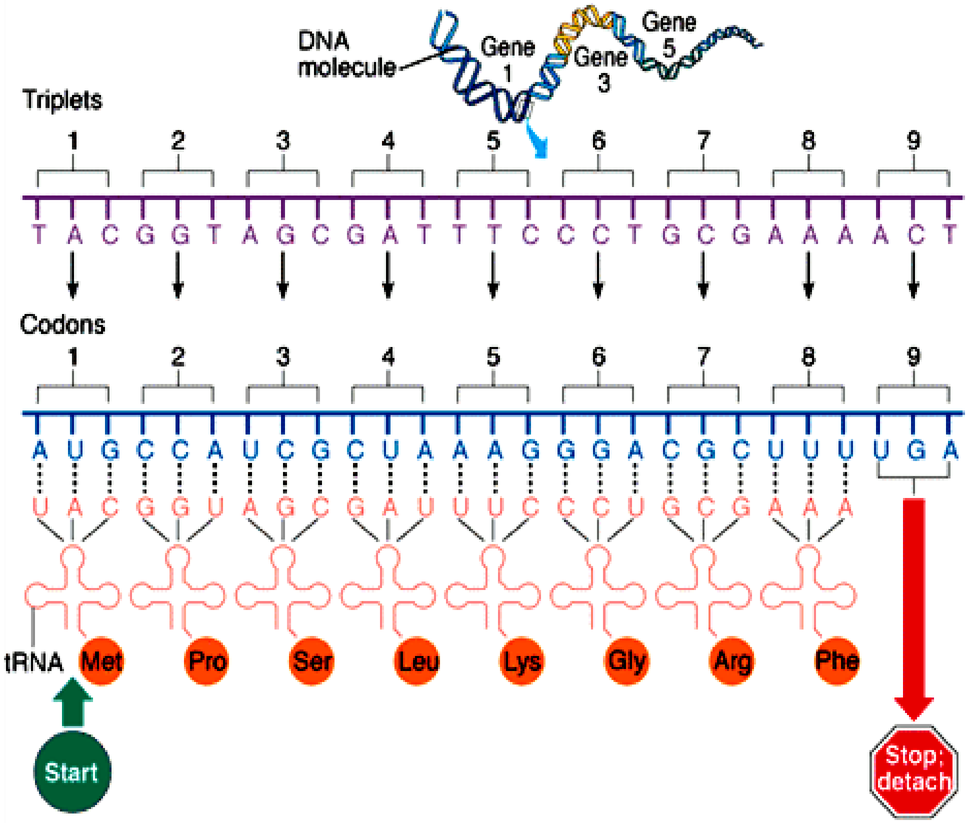
DNA (G, A, C, T)

↓

RNA (G, A, C, U)

↓

20 proteinogenic amino acids

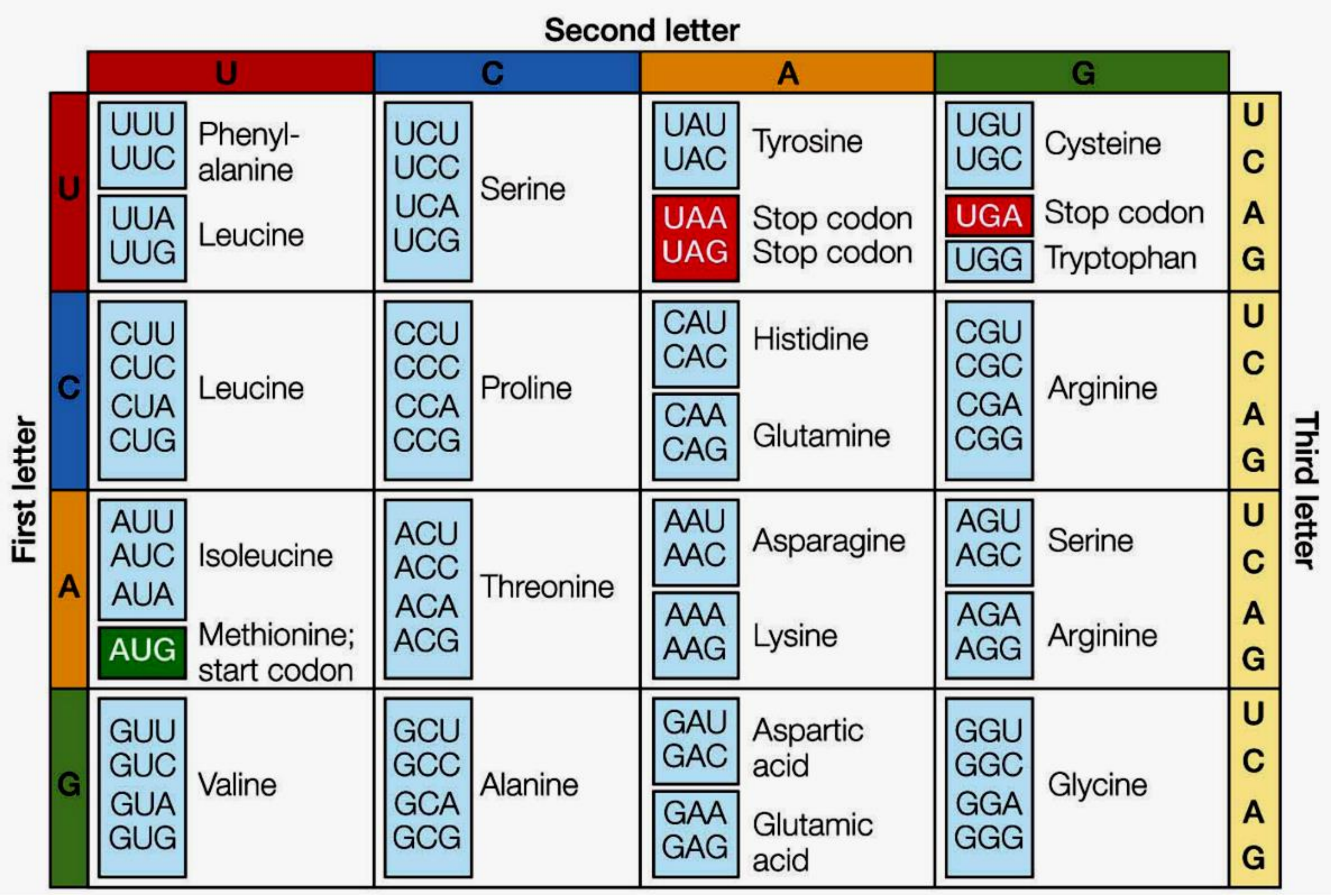


- # of code letters: **3 letters and 4 types of ntds**

→ 4^3 = 64 possibilities

Translation – genetic code

- determines **sequence of AA** in peptide chain

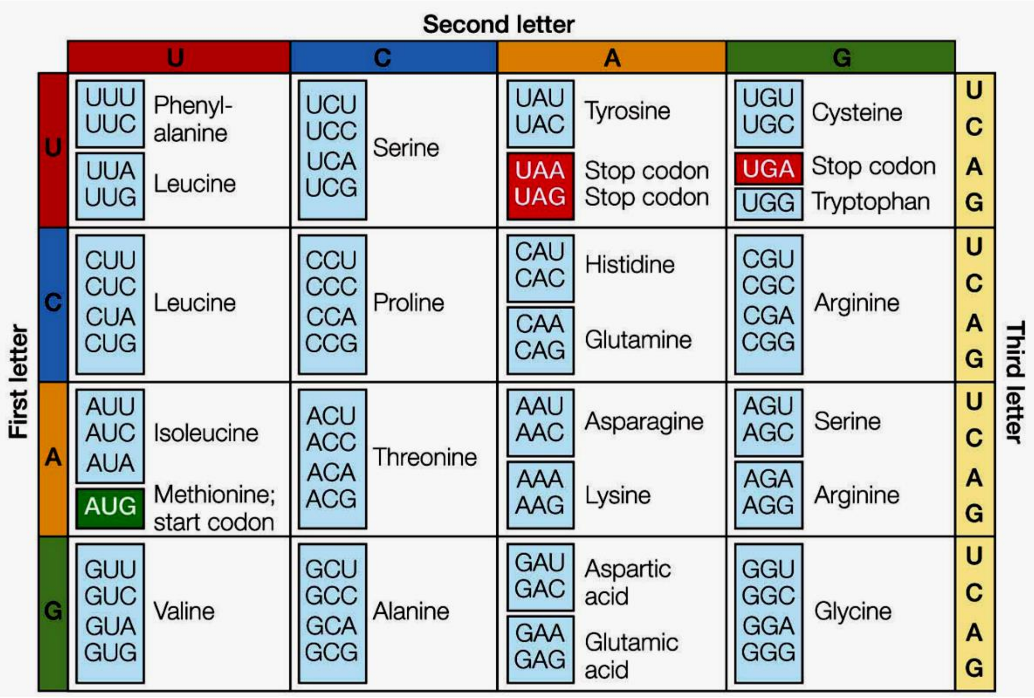


	Second letter								
First letter	U		C		A		G		Third letter
U	UUU	Phenyl-alanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	Leucine	UCA		UAA	Stop codon	UGA	Stop codon	A
	UUG		UCG		UAG	Stop codon	UGG	Tryptophan	G
C	CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lysine	AGA	Arginine	A
	AUG	Methionine; start codon	ACG		AAG		AGG		G
G	GUU	Valine	GCU	Alanine	GAU	Aspartic acid	GGU	Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

Translation – genetic code

- determines **sequence of AA** in peptide chain

DNA sequence:
5′-TTACGAAGGTTATTG-3′



	Second letter								
First letter	U		C		A		G		Third letter
U	UUU UUC	Phenylalanine	UCU UCC UCA UCG	Serine	UAU UAC	Tyrosine	UGU UGC	Cysteine	U C
	UUA UUG	Leucine			UAA UAG	Stop codon Stop codon	UGA	Stop codon	A
							UGG	Tryptophan	G
C	CUU CUC CUA CUG	Leucine	CCU CCC CCA CCG	Proline	CAU CAC	Histidine	CGU CGC CGA CGG	Arginine	U C
					CAA CAG	Glutamine			A G
A	AUU AUC AUA	Isoleucine	ACU ACC ACA ACG	Threonine	AAU AAC	Asparagine	AGU AGC	Serine	U C
	AUG	Methionine; start codon			AAA AAG	Lysine	AGA AGG	Arginine	A G
G	GUU GUC GUA GUG	Valine	GCU GCC GCA GCG	Alanine	GAU GAC	Aspartic acid	GGU GGC GGA GGG	Glycine	U C
					GAA GAG	Glutamic acid			A G

Translation – genetic code

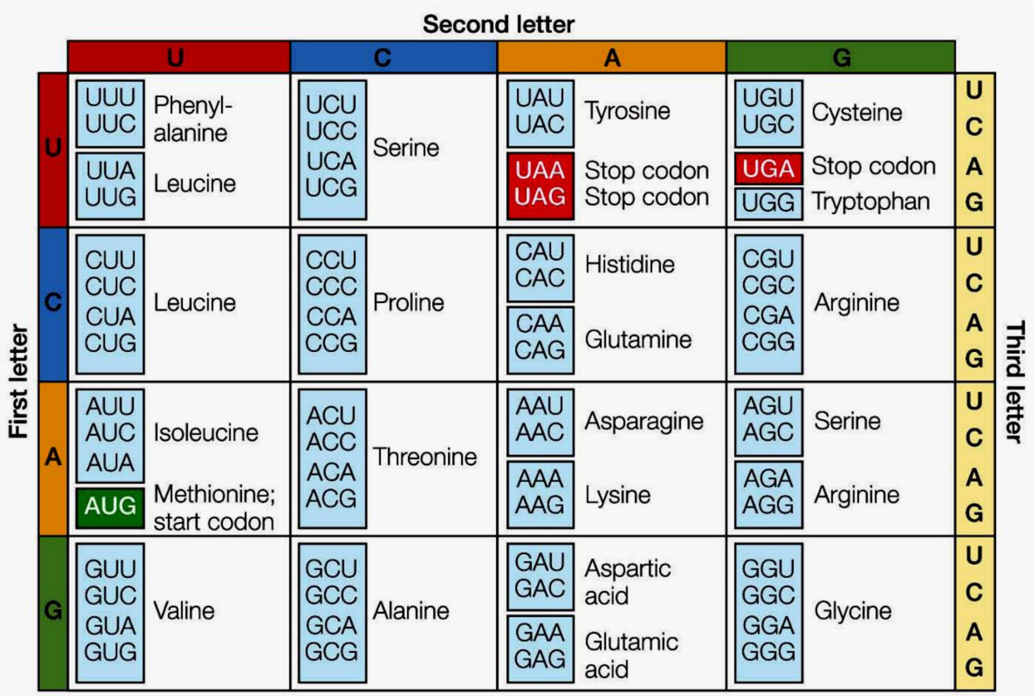
- determines **sequence of AA** in peptide chain

DNA sequence:

5′-TTACGAAGGTTATTG-3′

↓ *Transcription into* **mRNA:**

5′-UUA CGA AGG UUA UUG-3′



	Second letter								
First letter	U		C		A		G		Third letter
U	UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	Leucine	UCA		UAA	Stop codon	UGA	Stop codon	A
	UUG		UCG		UAG	Stop codon	UGG	Tryptophan	G
C	CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lysine	AGA	Arginine	A
	AUG	Methionine; start codon	ACG		AAG		AGG		G
G	GUU	Valine	GCU	Alanine	GAU	Aspartic acid	GGU	Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

Translation – genetic code

- determines **sequence of AA** in peptide chain

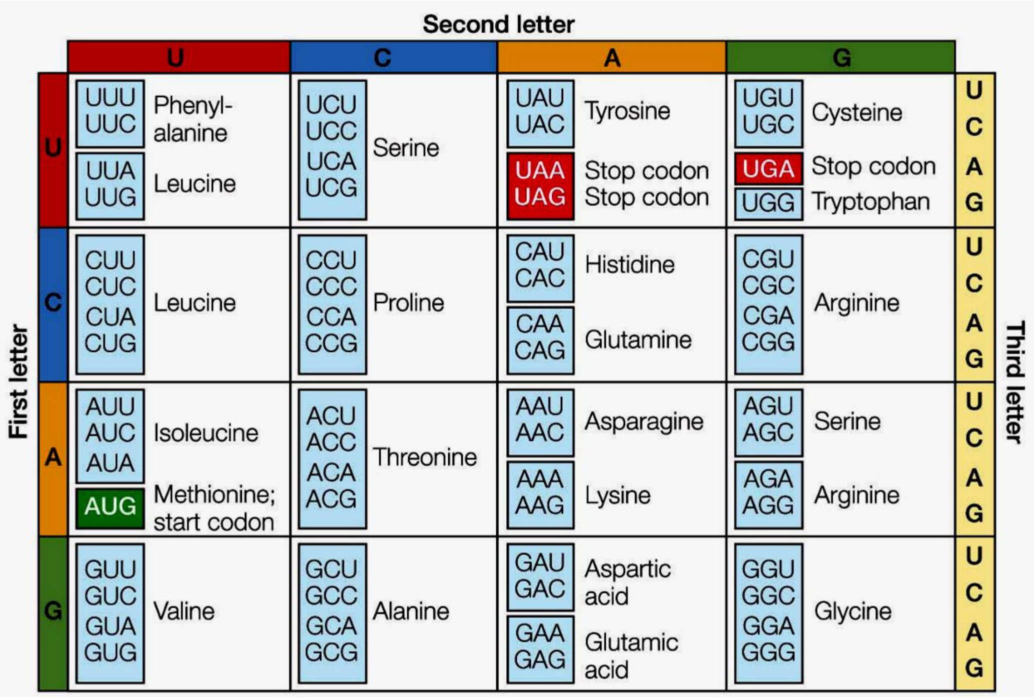
DNA sequence:

5′-TTACGAAGGTTATTG-3′

↓ *Transcription into* **mRNA:**

5′-UUA CGA AGG UUA UUG-3′

Triplet1 Triplet2 Triplet3 Triplet4 Triplet5



First letter	Second letter								Third letter
	U		C		A		G		
U	UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine	U
U	UUC		UCC		UAC		UGC		C
U	UUA	Leucine	UCA		UAA	Stop codon	UGA	Stop codon	A
U	UUG		UCG		UAG	Stop codon	UGG	Tryptophan	G
C	CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine	U
C	CUC		CCC		CAC		CGC		C
C	CUA		CCA		CAA	Glutamine	CGA		A
C	CUG		CCG		CAG		CGG		G
A	AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine	U
A	AUC		ACC		AAC		AGC		C
A	AUA		ACA		AAA	Lysine	AGA	Arginine	A
A	AUG	Methionine; start codon	ACG		AAG		AGG		G
G	GUU	Valine	GCU	Alanine	GAU	Aspartic acid	GGU	Glycine	U
G	GUC		GCC		GAC		GGC		C
G	GUA		GCA		GAA	Glutamic acid	GGA		A
G	GUG		GCG		GAG		GGG		G

Translation – genetic code

- determines **sequence of AA** in peptide chain

DNA sequence:

5′ –TTACGAAGGTTATTG–3′

↓ *Transcription into* **mRNA:**

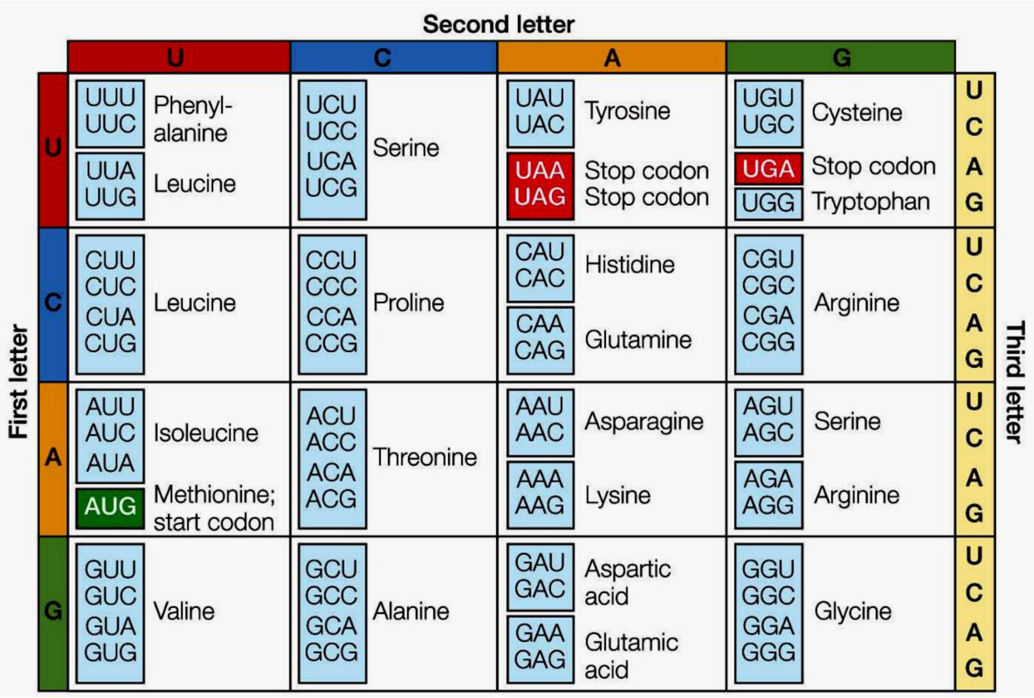
5′ –UUA CGA AGG UUA UUG–3′

Triplet1 Triplet2 Triplet3 Triplet4 Triplet5

↓

Translation into **protein:**

AA1 AA2 AA3 AA4 AA5



	Second letter				
First letter	**U**	**C**	**A**	**G**	**Third letter**
U	UUU UUC Phenylalanine UUA UUG Leucine	UCU UCC UCA UCG Serine	UAU UAC Tyrosine UAA Stop codon UAG Stop codon	UGU UGC Cysteine UGA Stop codon UGG Tryptophan	U C A G
C	CUU CUC CUA CUG Leucine	CCU CCC CCA CCG Proline	CAU CAC Histidine CAA CAG Glutamine	CGU CGC CGA CGG Arginine	U C A G
A	AUU AUC AUA Isoleucine AUG Methionine; start codon	ACU ACC ACA ACG Threonine	AAU AAC Asparagine AAA AAG Lysine	AGU AGC Serine AGA AGG Arginine	U C A G
G	GUU GUC GUA GUG Valine	GCU GCC GCA GCG Alanine	GAU GAC Aspartic acid GAA GAG Glutamic acid	GGU GGC GGA GGG Glycine	U C A G

Translation – genetic code

- determines **sequence of AA** in peptide chain

DNA sequence:

5′-TTACGAAGGTTATTG-3′

↓ *Transcription into* **mRNA:**

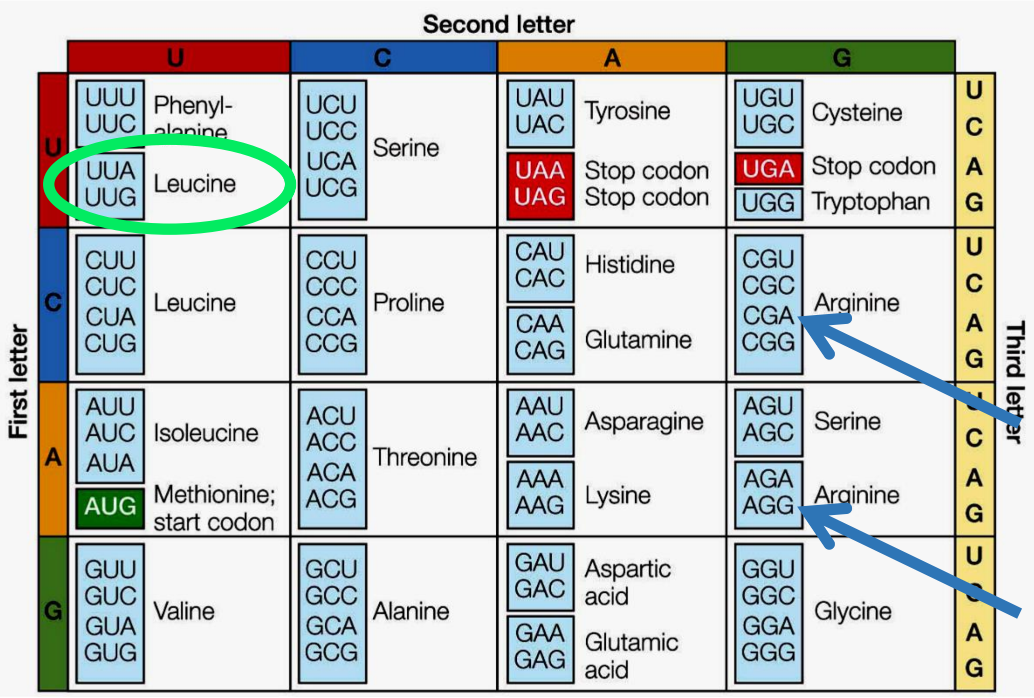
5′-UUA CGA AGG UUA UUG-3′

Triplet1 Triplet2 Triplet3 Triplet4 Triplet5

↓

Translation into **protein:**

N-Leu Arg Arg Leu Leu-C



Second letter

First letter	U	C	A	G	Third letter
U	UUU UUC Phenylalanine UUA UUG Leucine	UCU UCC UCA UCG Serine	UAU UAC Tyrosine UAA Stop codon UAG Stop codon	UGU UGC Cysteine UGA Stop codon UGG Tryptophan	U C A G
C	CUU CUC CUA CUG Leucine	CCU CCC CCA CCG Proline	CAU CAC Histidine CAA CAG Glutamine	CGU CGC CGA CGG Arginine	U C A G
A	AUU AUC AUA Isoleucine AUG Methionine; start codon	ACU ACC ACA ACG Threonine	AAU AAC Asparagine AAA AAG Lysine	AGU AGC Serine AGA AGG Arginine	U C A G
G	GUU GUC GUA GUG Valine	GCU GCC GCA GCG Alanine	GAU GAC Aspartic acid GAA GAG Glutamic acid	GGU GGC GGA GGG Glycine	U C A G

Translation – genetic code

- determines **sequence of AA** in peptide chain

DNA sequence:

5′-TTACGAAGGTTATTG-3′

↓ *Transcription into* **mRNA:**

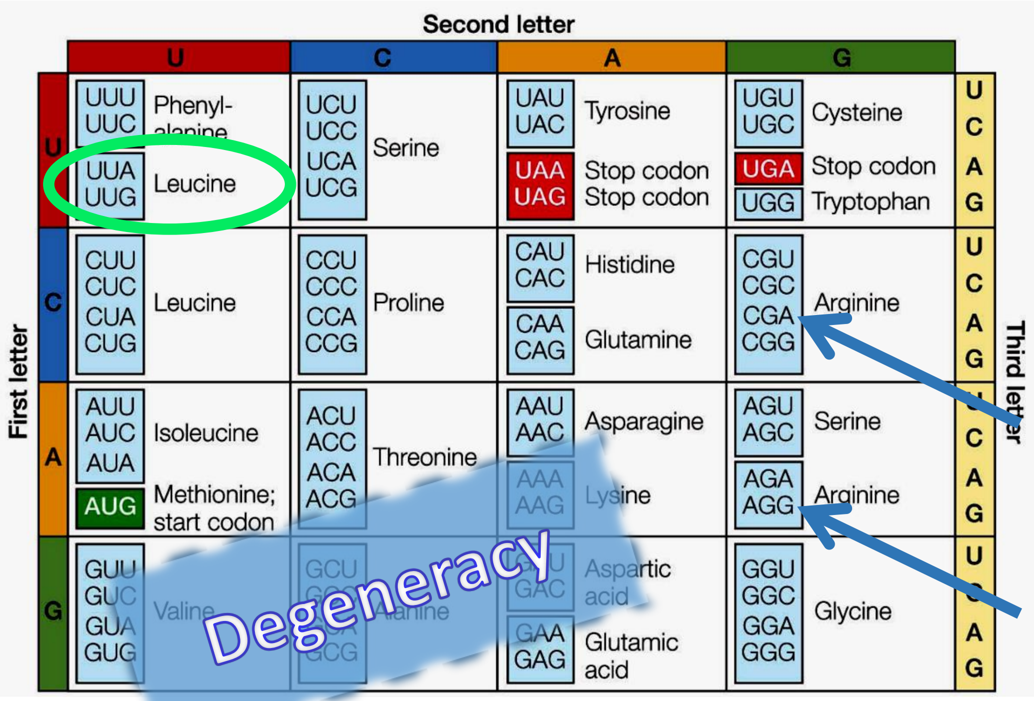
5′-UUA CGA AGG UUA UUG-3′

Triplet1 Triplet2 Triplet3 Triplet4 Triplet5

↓

Translation into **protein:**

N-Leu Arg Arg Leu Leu-C



First letter	Second letter: U		C		A		G		Third letter
U	UUU	Phenyl-alanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	Leucine	UCA		UAA	Stop codon	UGA	Stop codon	A
	UUG		UCG		UAG	Stop codon	UGG	Tryptophan	G
C	CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lysine	AGA	Arginine	A
	AUG	Methionine; start codon	ACG		AAG		AGG		G
G	GUU	Valine	GCU	Alanine	GAU	Aspartic acid	GGU	Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

Degeneracy

Translation – genetic code

- determines **sequence of AA** in peptide chain

→ **triplets:**

Triplet of 4 bases = 4^3 possibilities

→ **degeneracy:** 64 triplets → 20 AAs

→ **unambiguous**

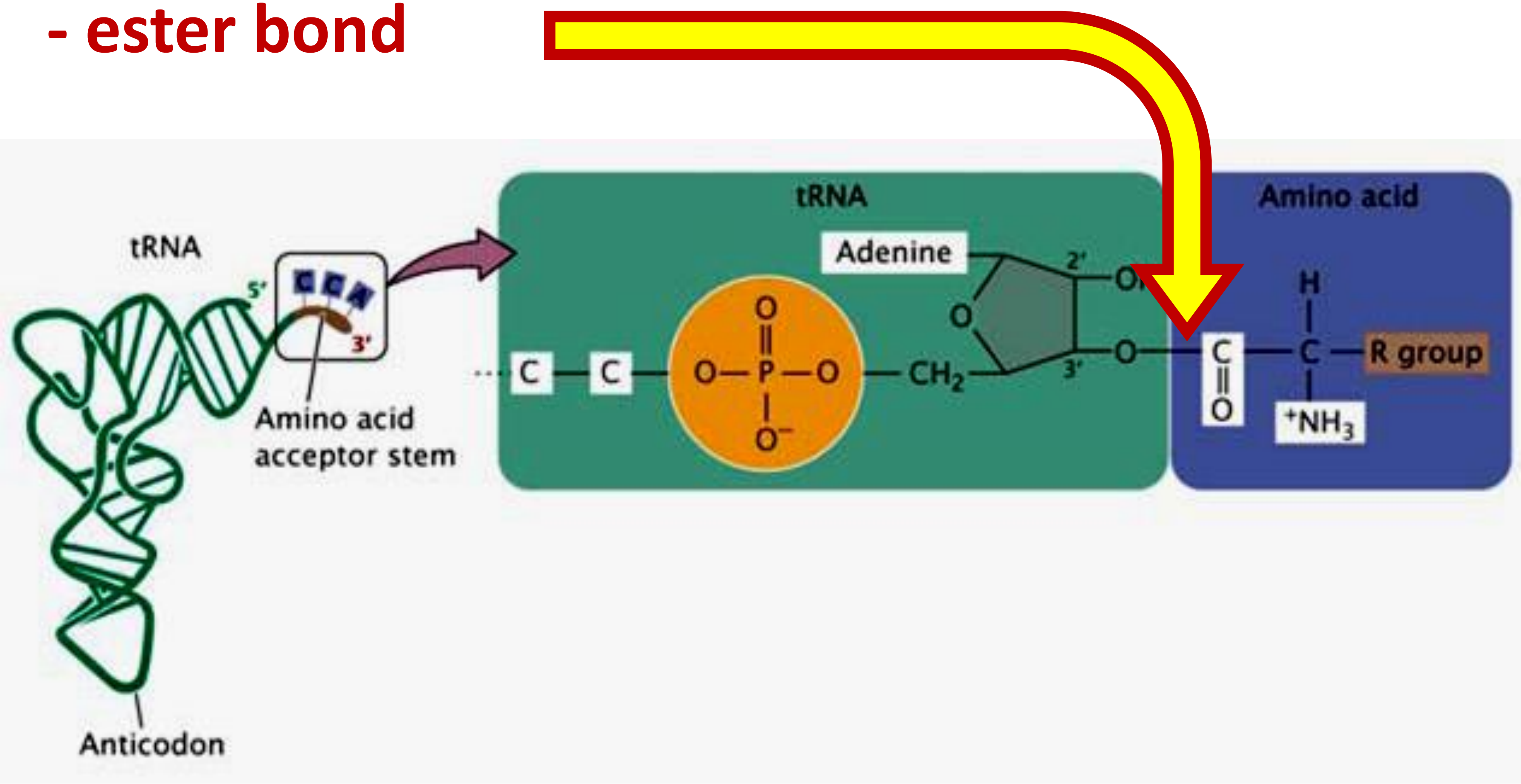
→ **non-overlapping**

→ **universal (almost: except mtch)**

Translation – AA activation

- linkage to tRNA

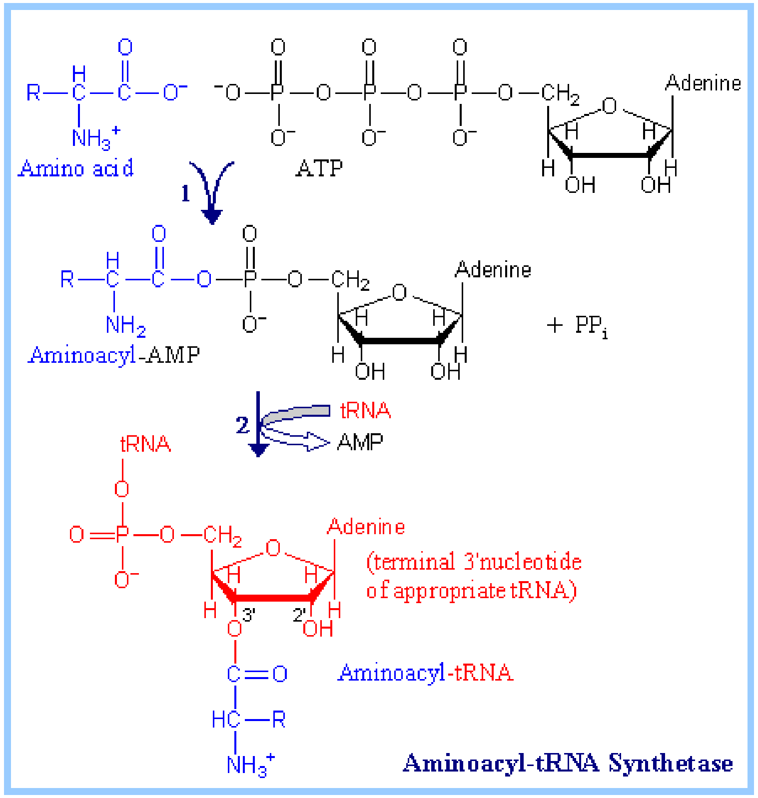
- **ester bond**



Translation – AA activation

- enzymes: aminoacyl-tRNA-synthetases

- two-step, consumption of ATP

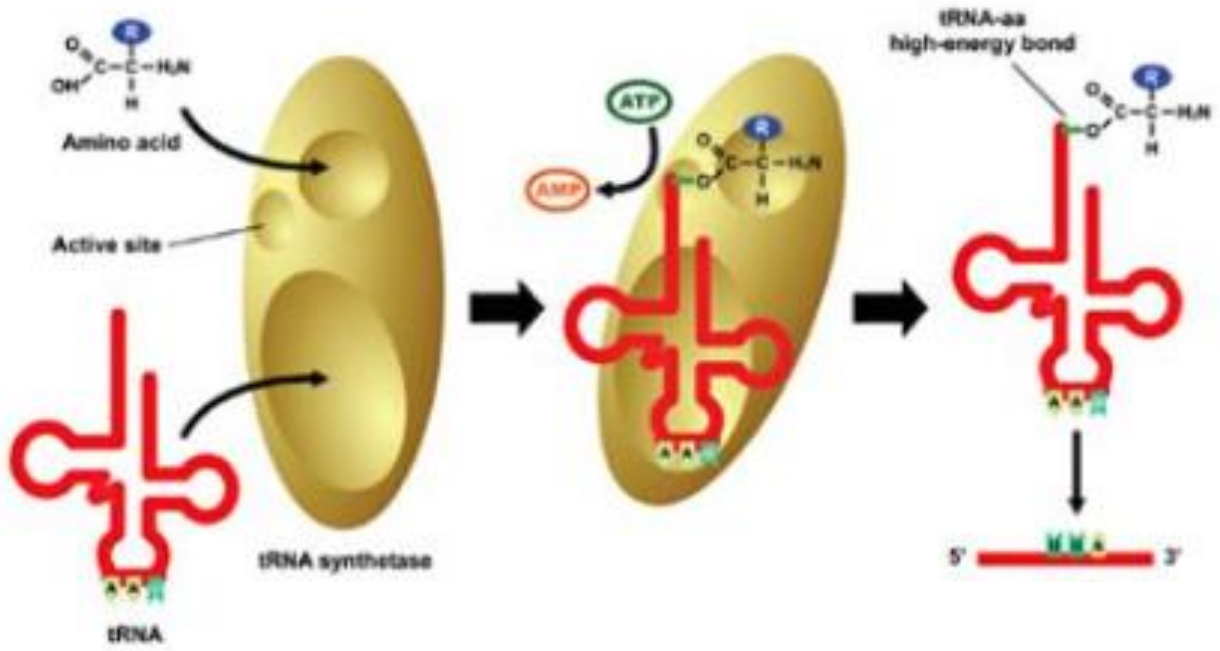


Translation – AA activation

- enzymes: **aminoacyl-tRNA-synthetases**

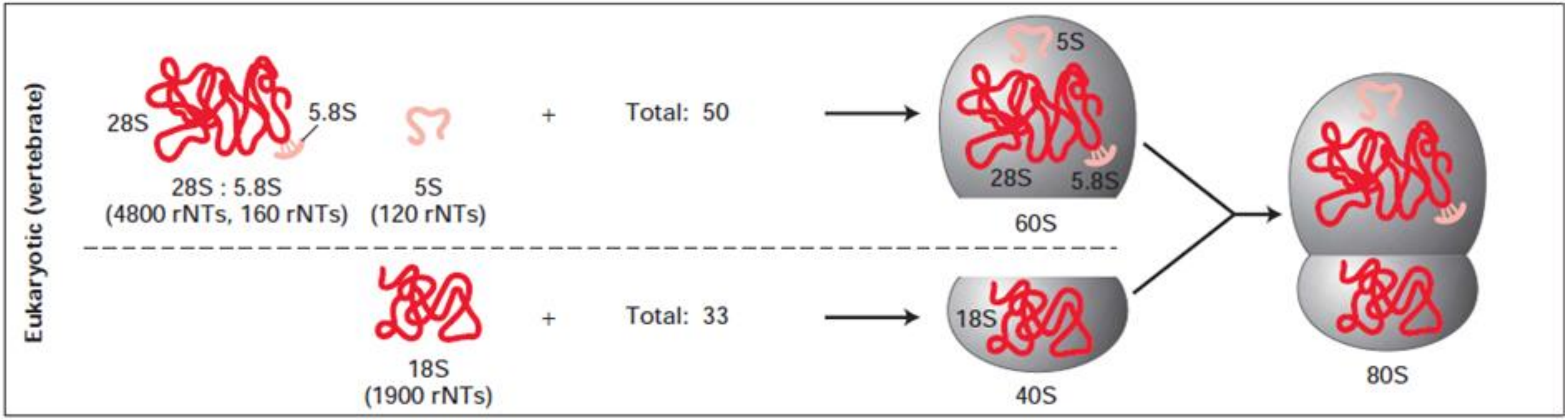
→ energy supply

→ incorporation of a correct AA into peptide



Translation – ribosome

- translation proceeds in **ribosomes** in cytoplasm



Translation – components

- aminoacyl-tRNAs
- ribosomes
- mRNAs
- iniciation, elongation and termination factors
- energy: ATP, GTP

Translation – 3 phases

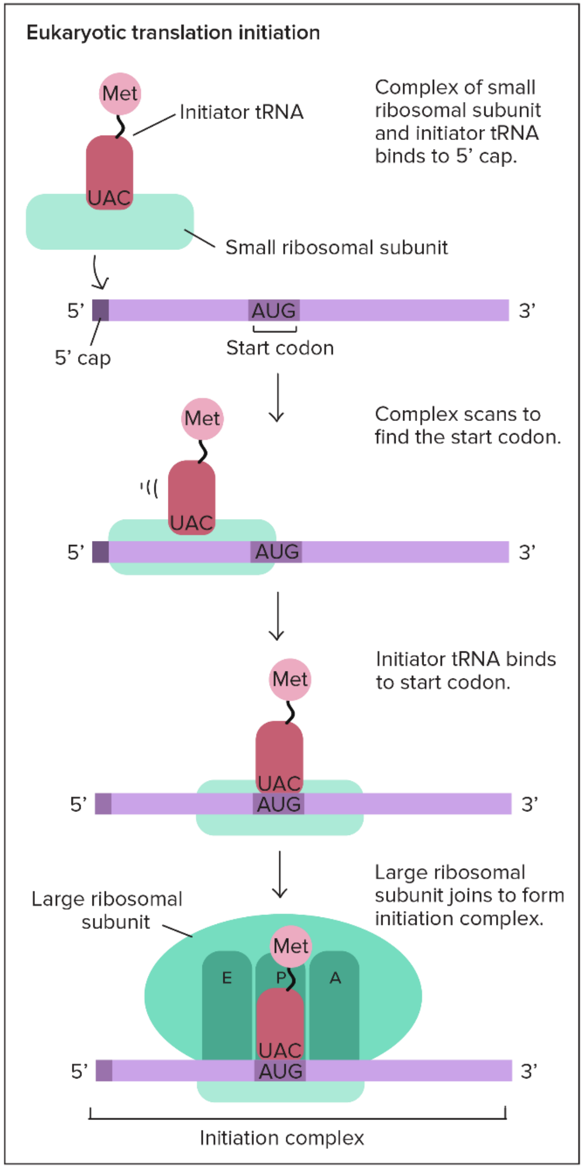
- **initiation**: initiation complex formation →

mRNA binding and iniciation codon finding

- **elongation**: linkage of additional AAs
- **termination**: synthesis stopping

Translation – initiation

- initiation complex formation
- many initiation factors
- energy from ATP and GTP



Eukaryotic translation initiation

Complex of small ribosomal subunit and initiator tRNA binds to 5’ cap.

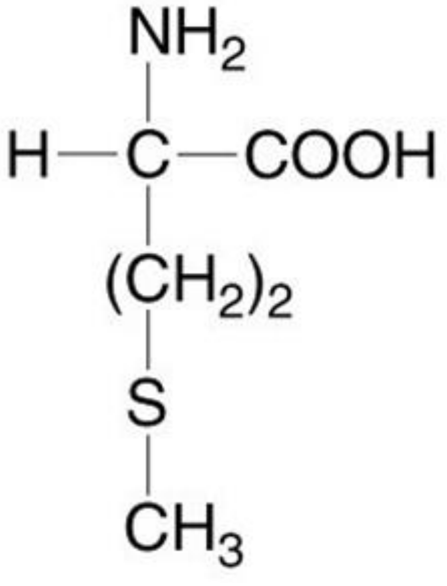
Complex scans to find the start codon.

Initiator tRNA binds to start codon.

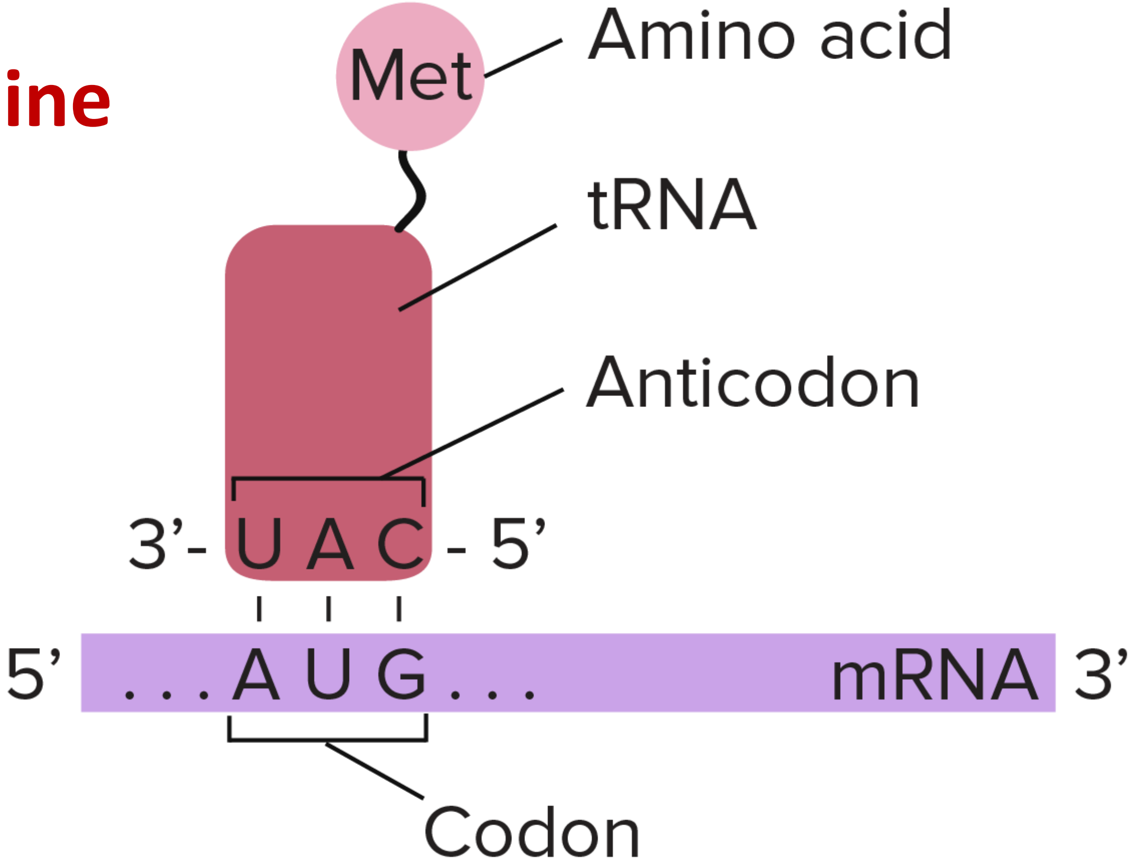
Large ribosomal subunit joins to form initiation complex.

Translation – initiation

- initiation codon: **AUG**
- first AA: **methionine**



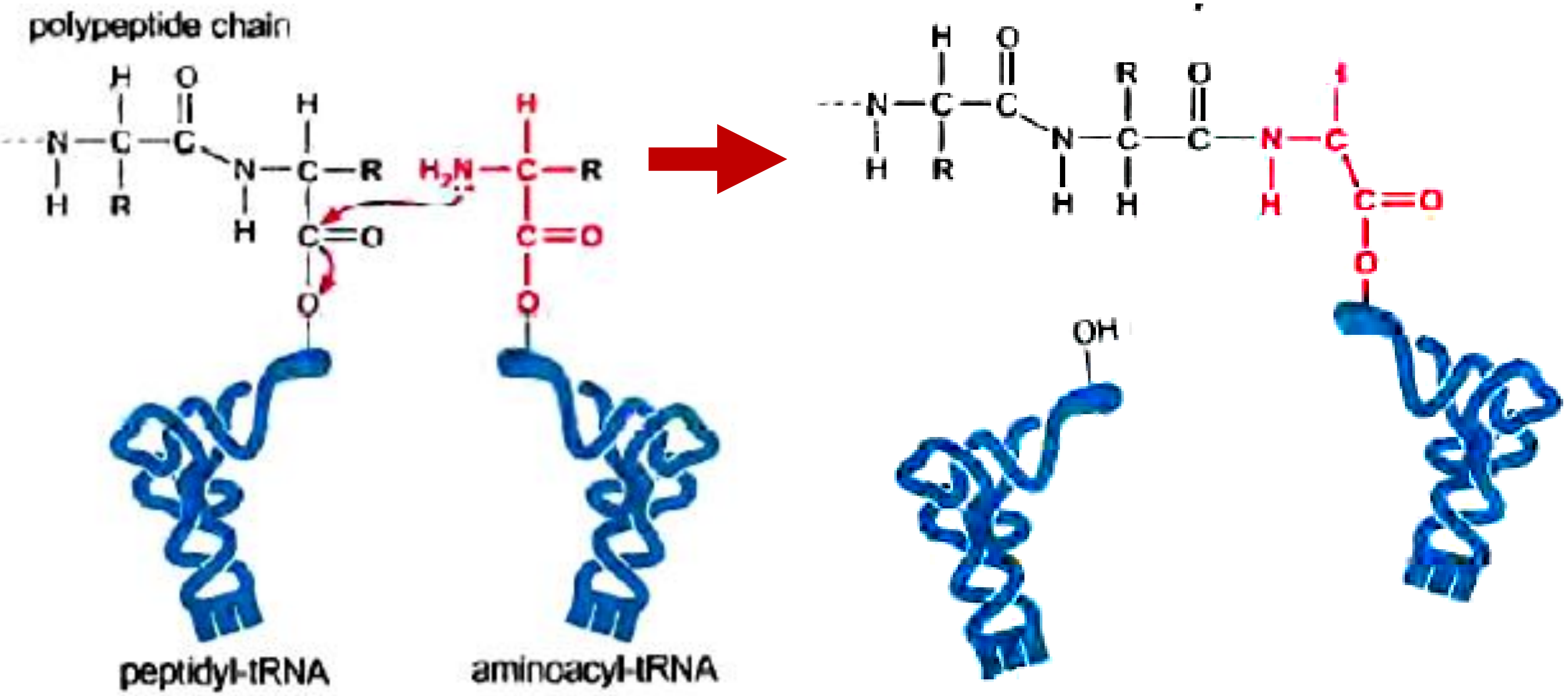
Methionine



Translation – elongation

- elongation of a growing peptide chain

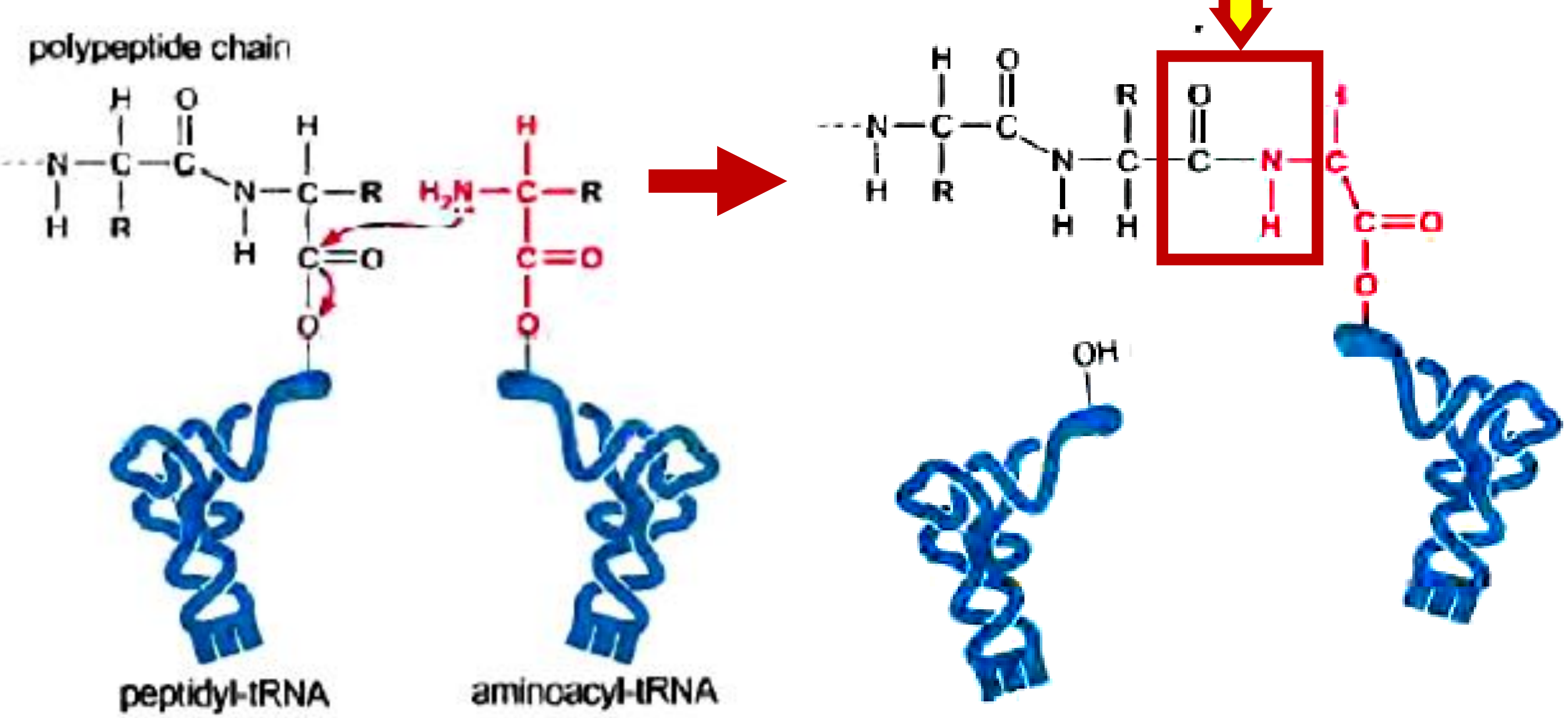
→ **peptide bond** synthesis



Translation – elongation

- elongation of a growing peptide chain

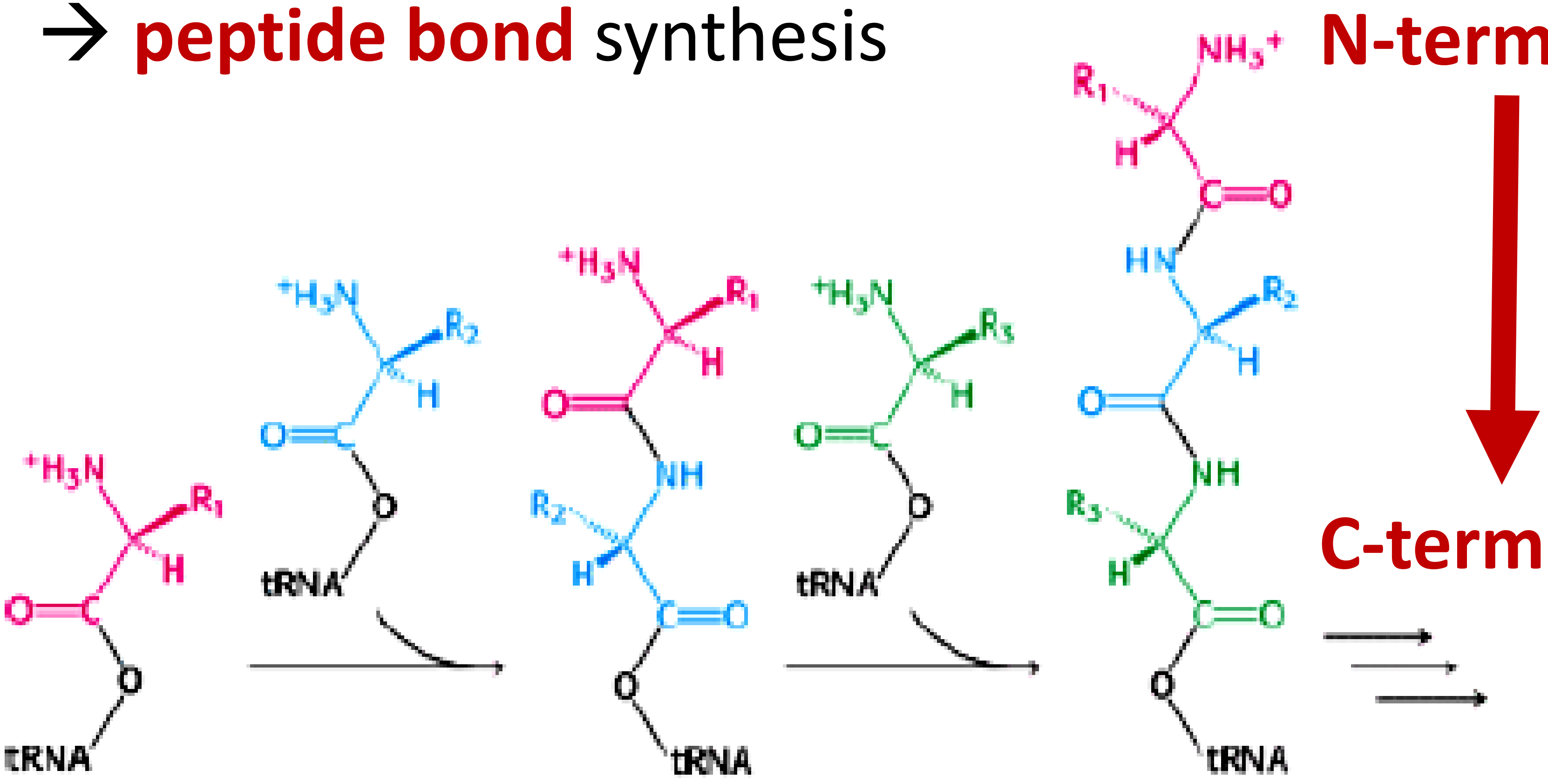
→ **peptide bond** synthesis



Translation – elongation

- elongation of a growing peptide chain

→ **peptide bond** synthesis



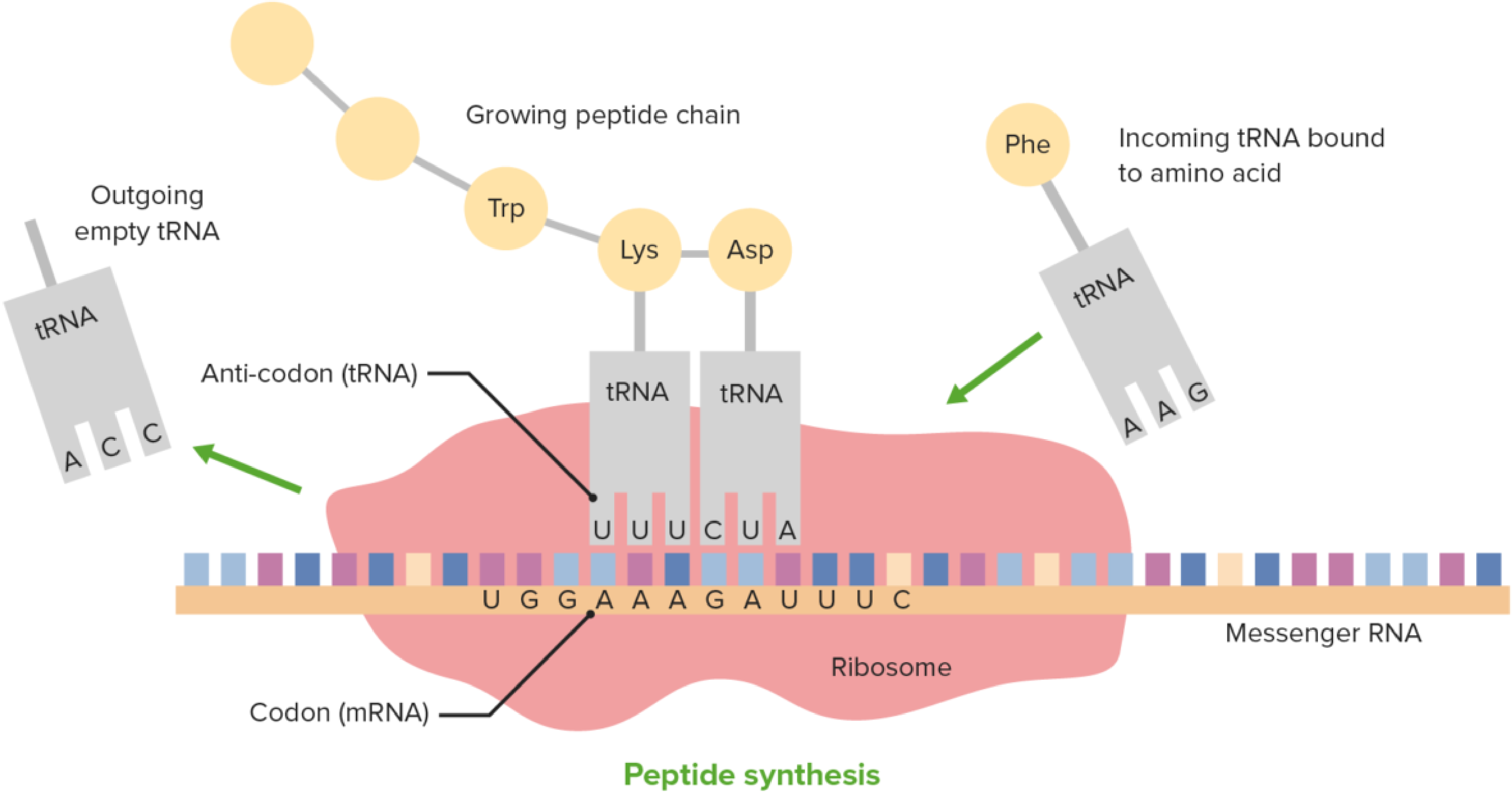
N-terminus

C-terminus

mRNA: 5‘-end → **3‘-end**

Translation – mechanism

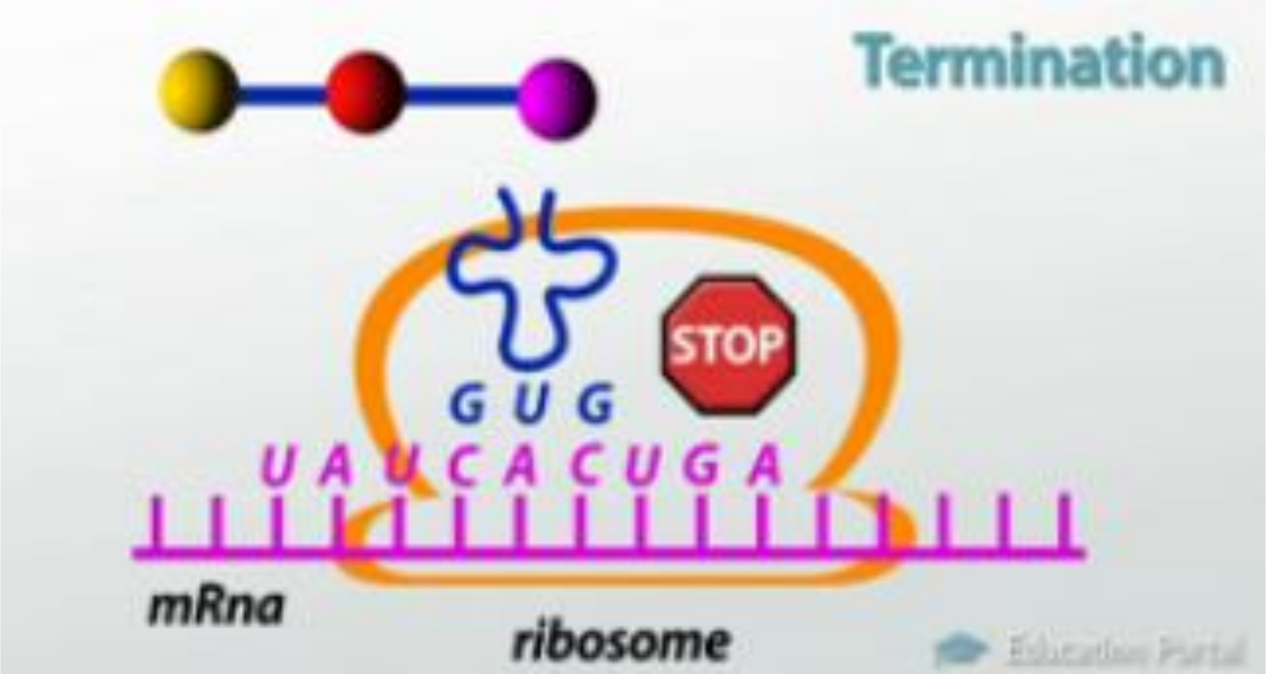
- **ribosome = ribozym;** + energy from GTP



Peptide synthesis

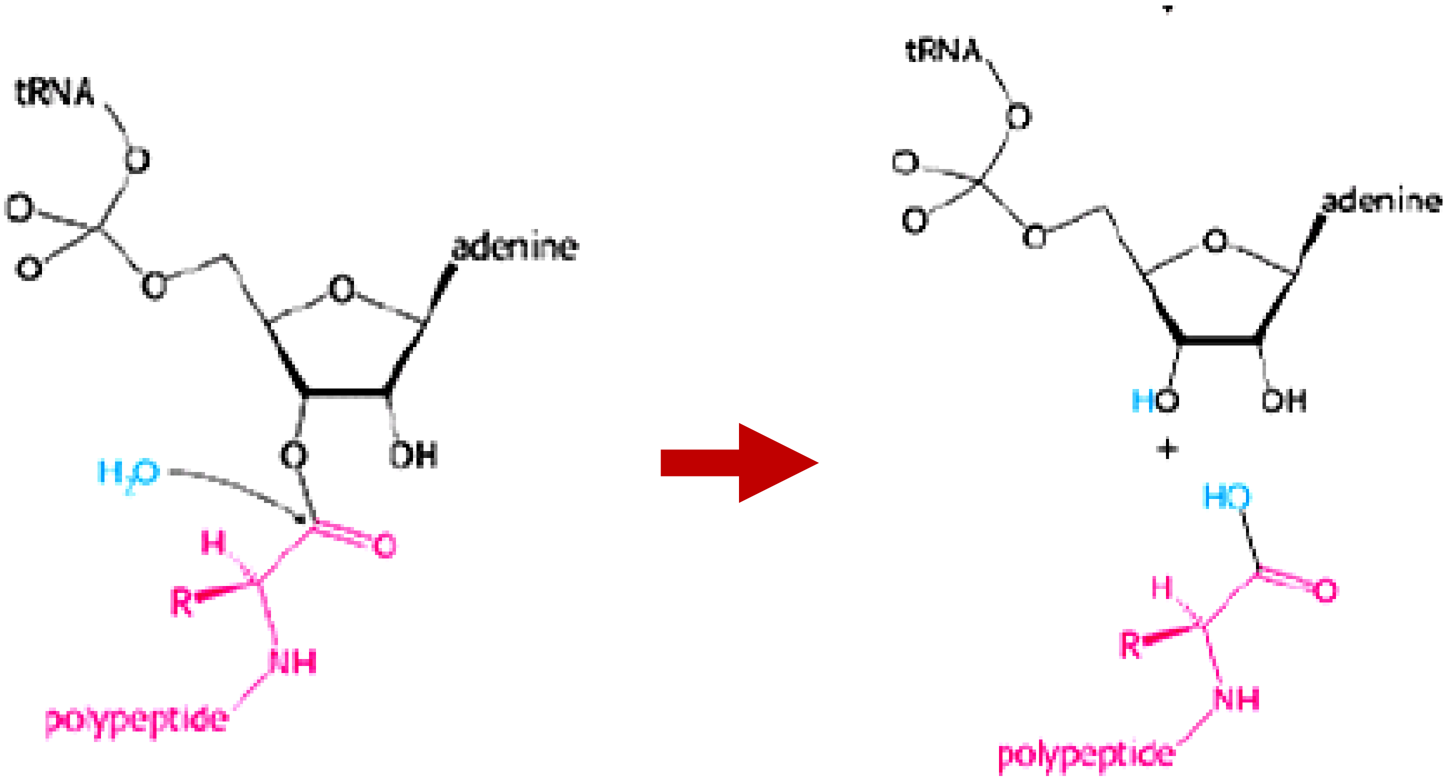
Translation – termination

- termination codons: **UAA**, **UAG**, **UGA**
- releasing factors + GTP

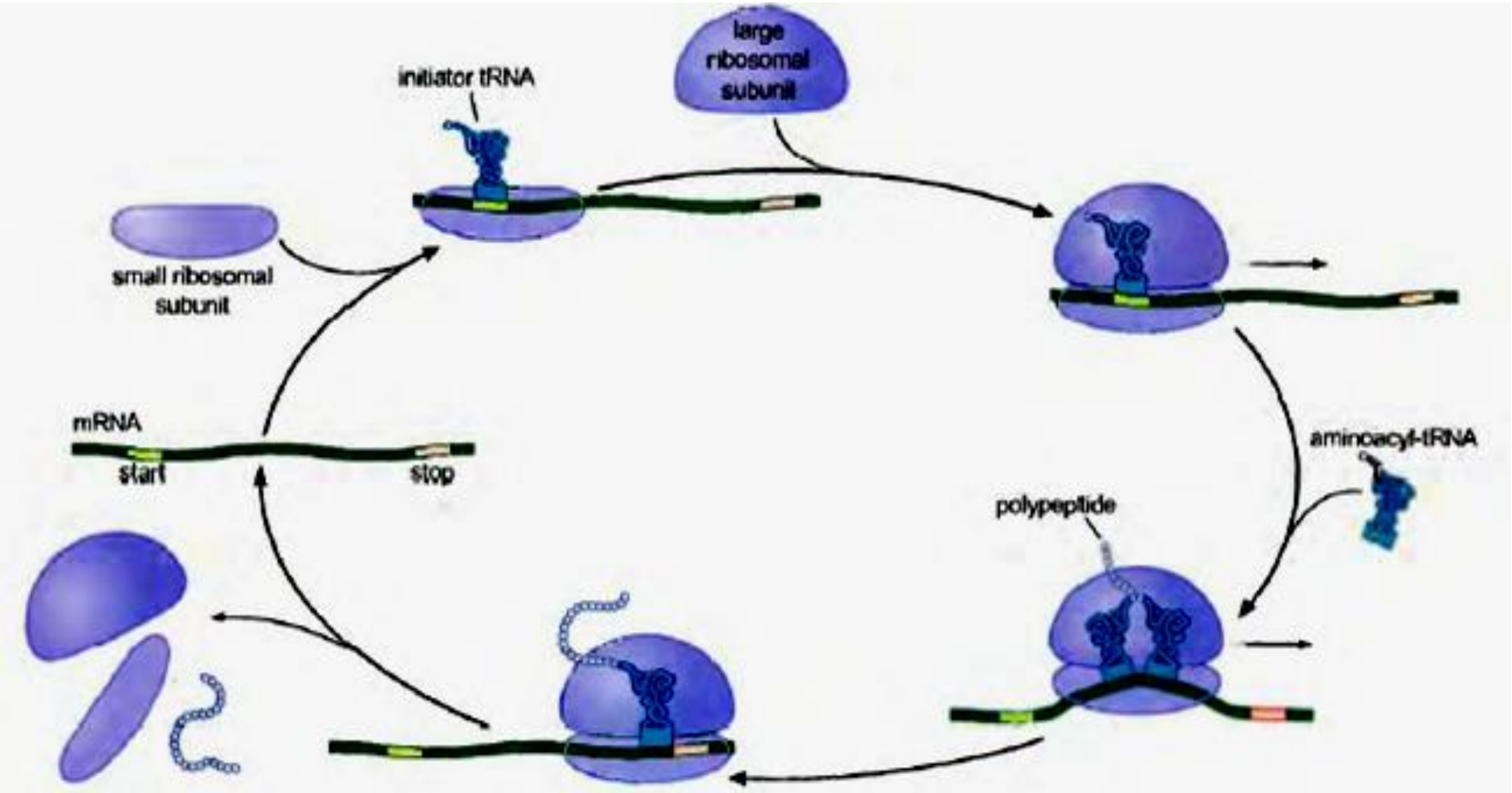


Translation – termination

- termination codons: **UAA**, **UAG**, **UGA**
- releasing factors + GTP



Translation – ribosome cycle

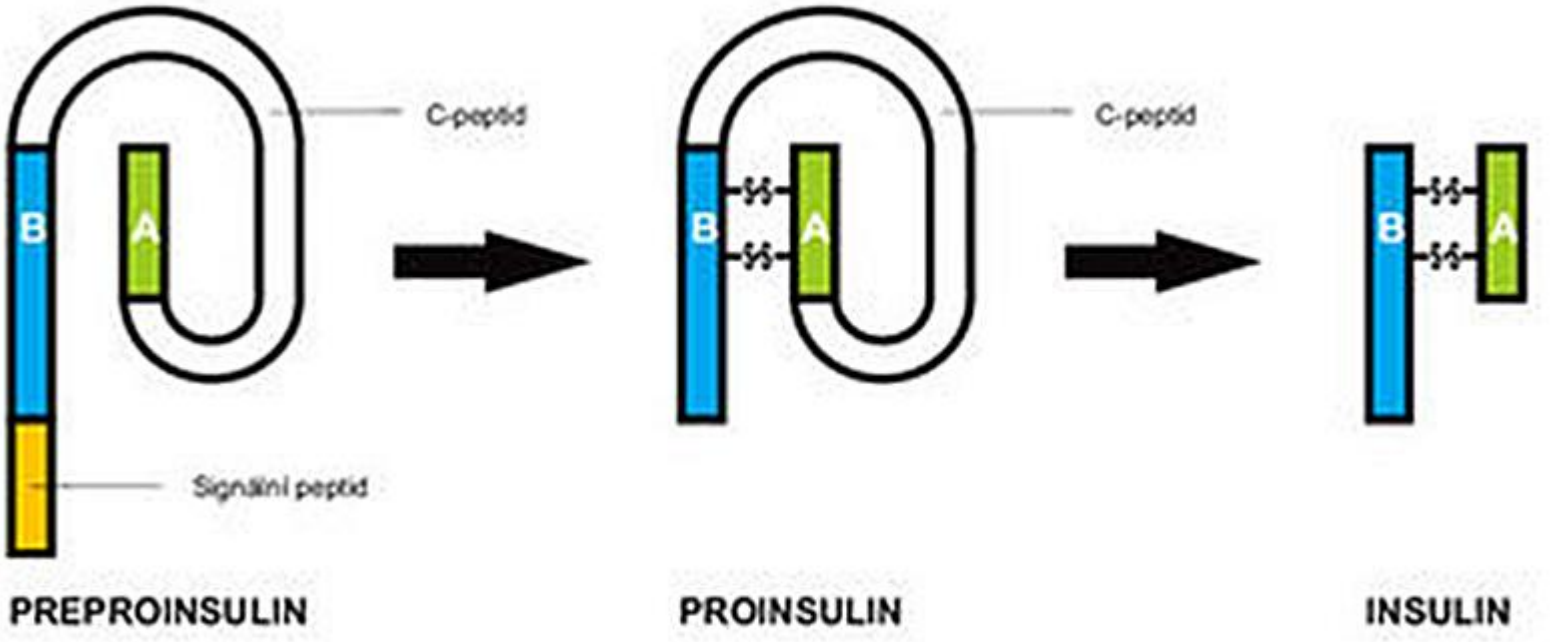


Posttranslation modifications

- synthesis as a precursor molecules
- **covalent linkage of**
    - → functional groups (-OH, -CH_3, …)
    - → whole molecules (glycosylation)
- **proteolytic cleavage**
- disulfide bridges formation
- subunits assembly

Posttranslation modifications

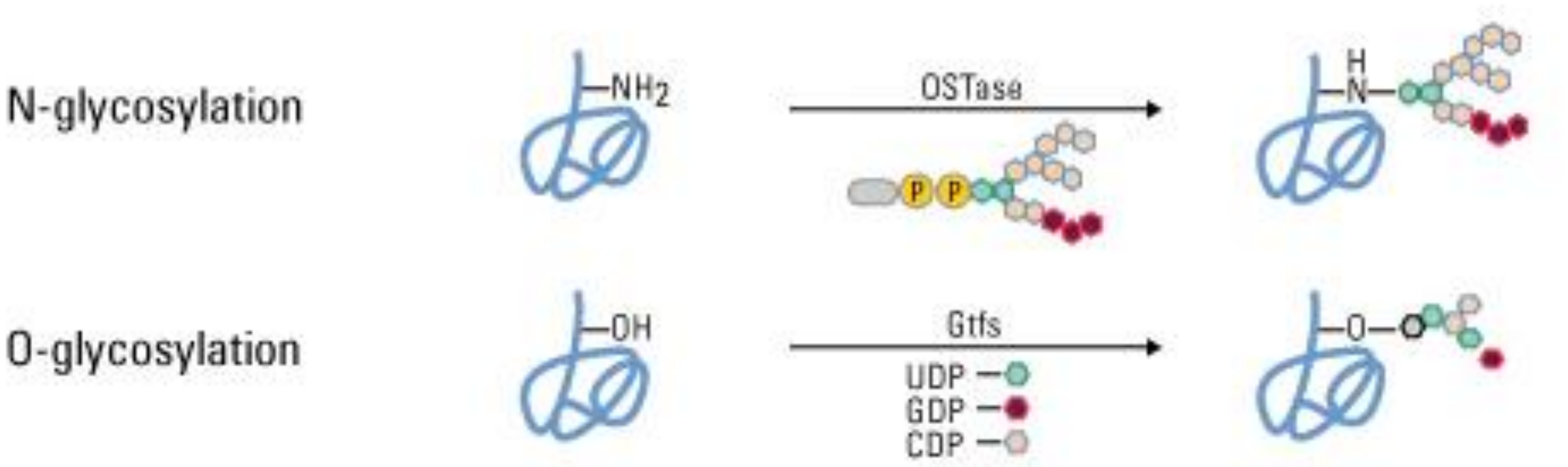
- **proteolytic cleavage**: example – **insulin**



Posttranslation modifications

- **glycosylation**

- proceeds in ER



Posttranslation modifications

- **glycosylation** function:
  - → protection from degradation
  - → extracellular communication
  - → cell-cell adhesion
  - → erythrocyte AB0 system

Posttranslation modifications

- **hydroxylation (-OH)**

→ collagen structure

- and: methylation, phosphorylation, carboxylation, acetylation, …
- further **increase heterogeneity** of produced proteins

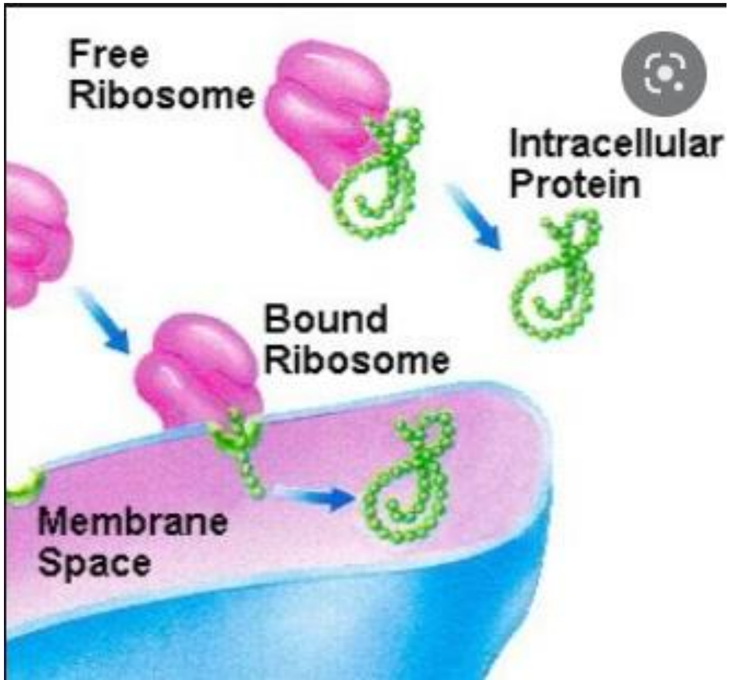
Protein distribution

- free ribosomes:

- → cytoplazmic proteins
- → mitochondrial proteins
- → nuclear proteins

- ribosomes of rough ER:

- → membrane proteins
- → secretory proteins



Thank you for your attention.