Cellular and Molecular Biology I

BIO-205-6

Camille Goemans

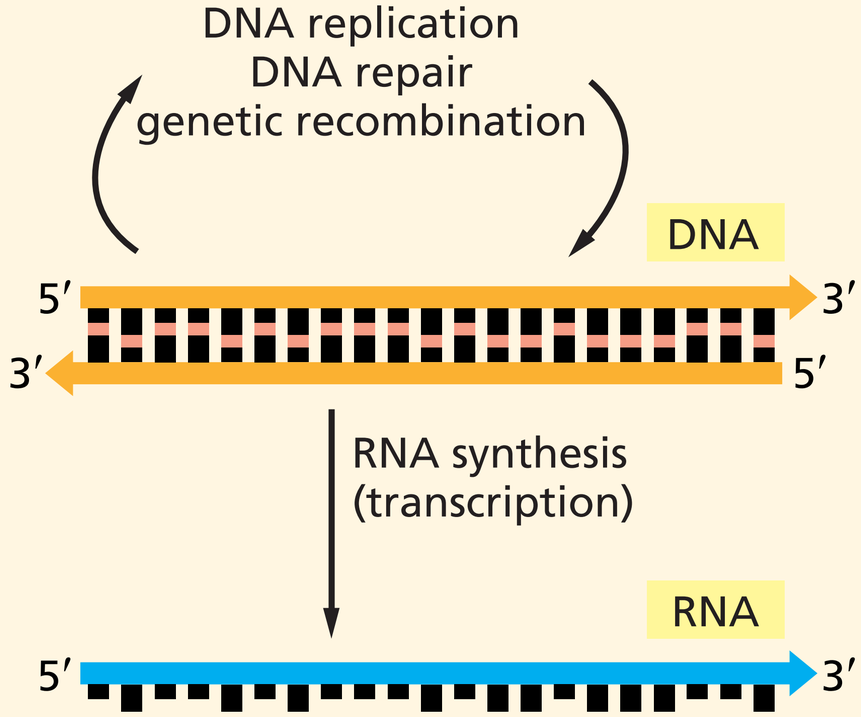
II. DNA to RNA

1. RNA
2. Transcription
3. Transcription initiation
4. **RNA processing**
5. Non-coding RNAs

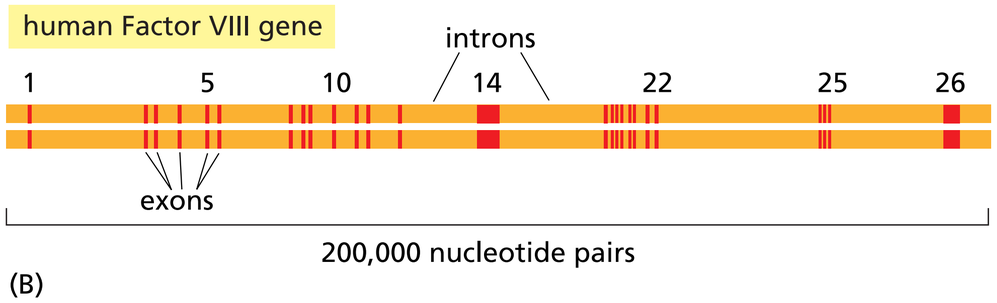
a. RNA capping

b. RNA splicing

c. PolyA tail

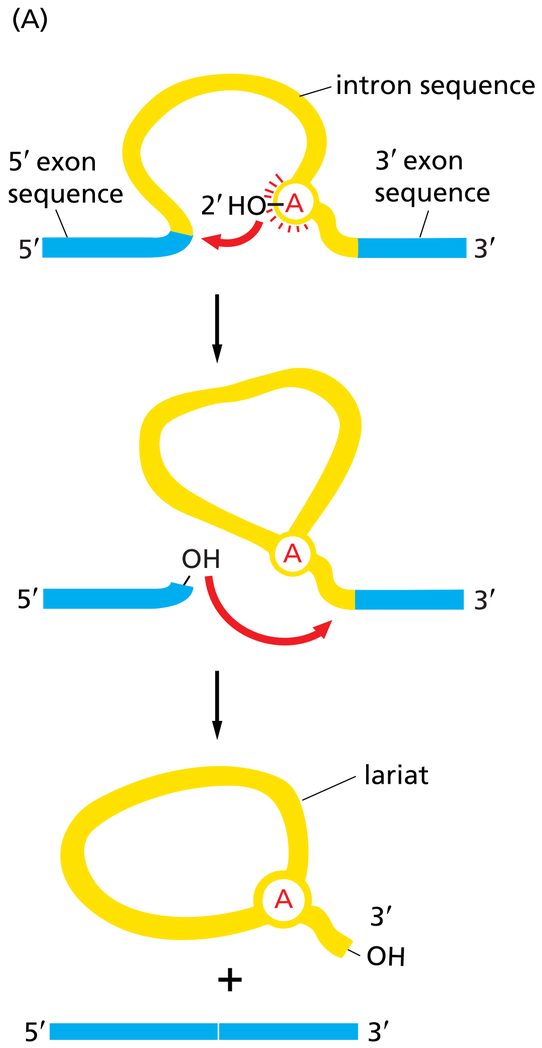


RNA splicing



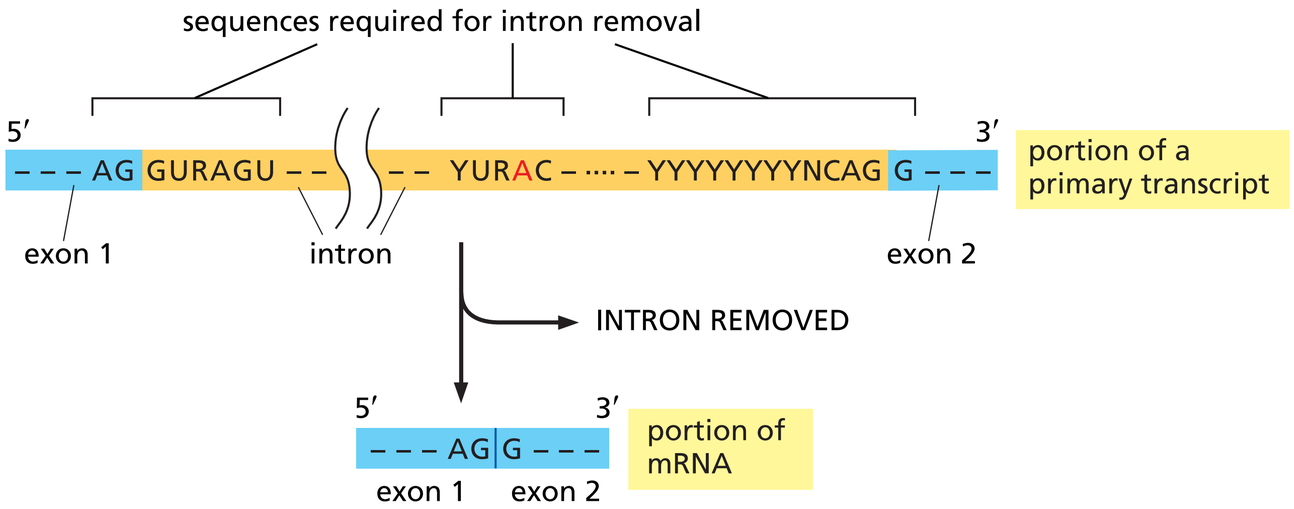
- Protein coding sequence (**exons**) in **eukaryotic** genes is **interrupted** by non-coding sequences (**introns**)
- Both **introns and exons** are transcribed into RNA
- The introns are removed through **RNA splicing**
- Before splicing, the mRNA is called **precursor-mRNA (or pre-mRNA)**

RNA splicing



- Each **splicing event** removes an intron and joins two exons
- The removed part is called a **lariat**
- The **machinery** is **complex** and consists of 5 RNA molecules (**small nuclear RNAs**) called U1, 2, 4, 5, 6
- Each **snRNA** is complexed with at least 7 proteins to form the **small nuclear ribonucleoprotein** or **snRNP**
- The complexity ensures the **accuracy** of the splicing and **flexibility** to accommodate all sequences

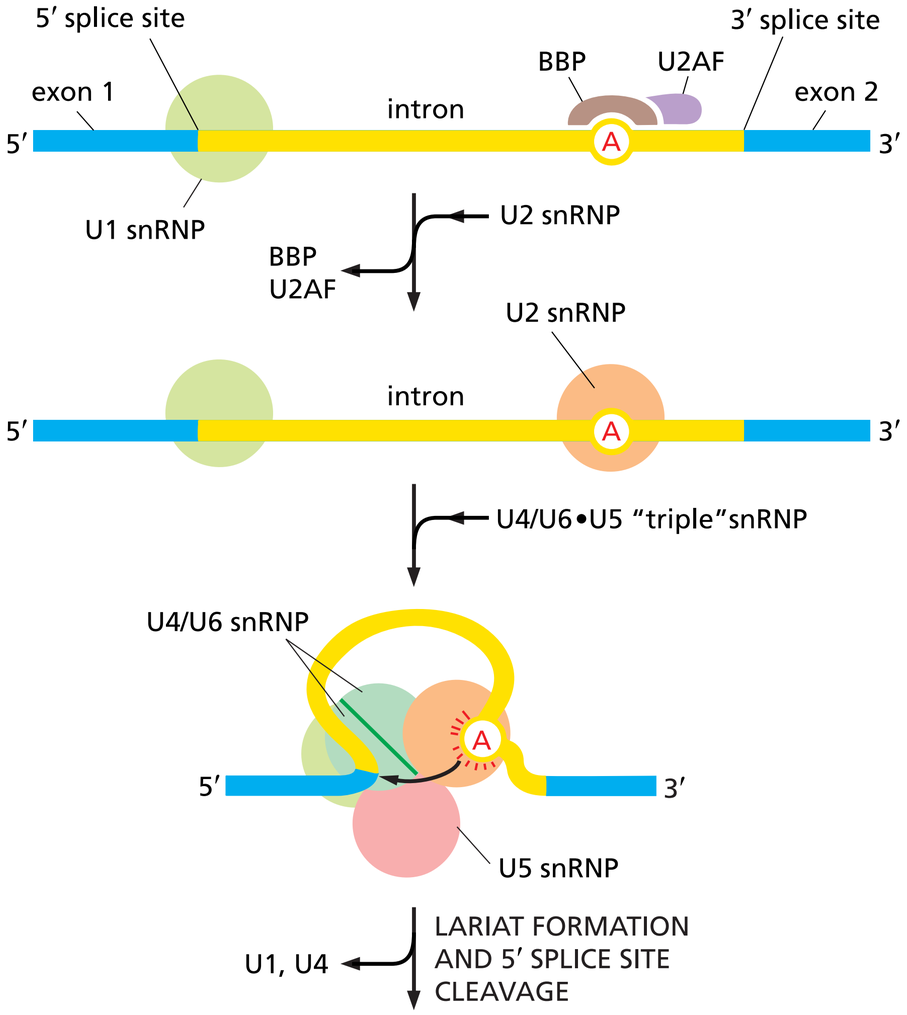
RNA splicing



How does this work?

- The splicing machinery needs to **recognize 3 sequences** (the 5’ splice site, the 3’ splice site and the branch point)
- The 3’ and 5’ splice sites contain **both intron and exon** sequence (junction)
- Sequence recognition happens through **base-pairing** with the snRNAs.
- Sequences are **highly variable** - difficult for scientists to predict the exact sequences

RNA splicing



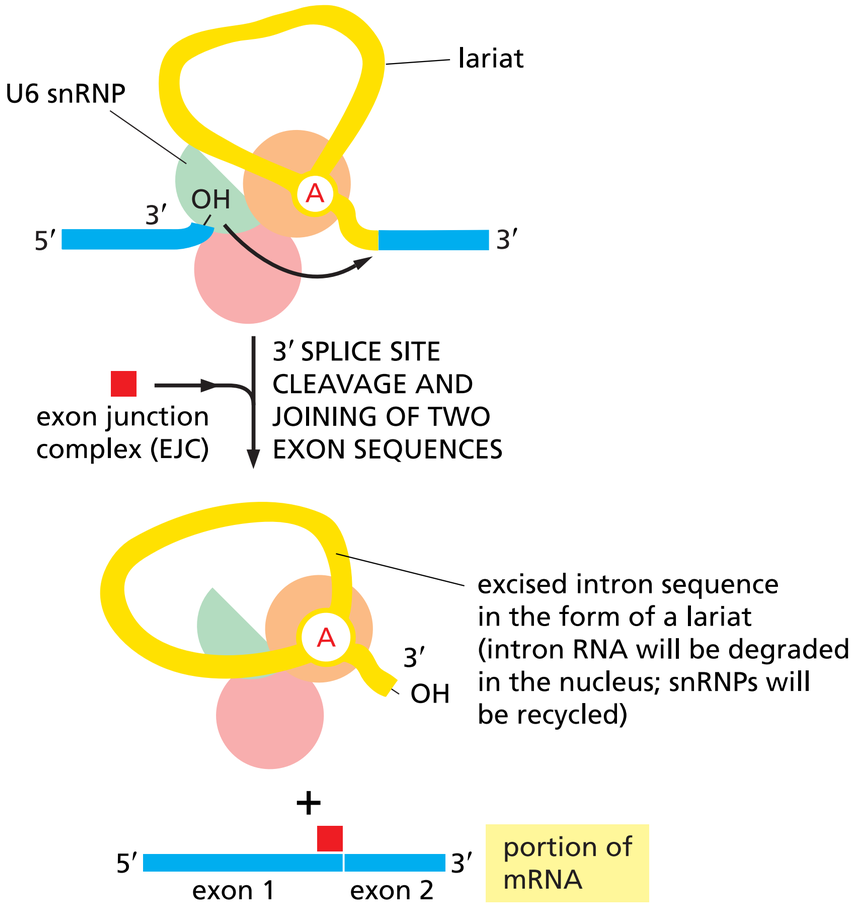
portion of a pre-mRNA transcript

The U1 snRNP forms base pairs with the 5′ splice junction (see Figure 6–29) and the BBP (branch-point binding protein) and U2AF (U2 auxilliary factor) recognize the branch-point site.

The U2 snRNP displaces BBP and U2AF and forms base pairs with the branch-point site consensus sequence.

The U4/U6•U5 “triple” snRNP enters the reaction. In this triple snRNP, the U4 and U6 snRNAs are held firmly together by base-pair interactions. Subsequent rearrangements break apart the U4/U6 base pairs, allowing U6 to displace U1 at the 5′ splice junction (see Figure 6–29). This creates the active site that catalyzes the first phosphoryl-transferase reaction.

RNA splicing



Additional RNA–RNA rearrangements create the active site for the second phosphoryl-transferase reaction, which then completes the splice (see Figure 6–25A).

The Exon Junction Complex (EJC) is a multi-protein complex that is deposited on mRNA at the site where two exons are joined together after splicing.

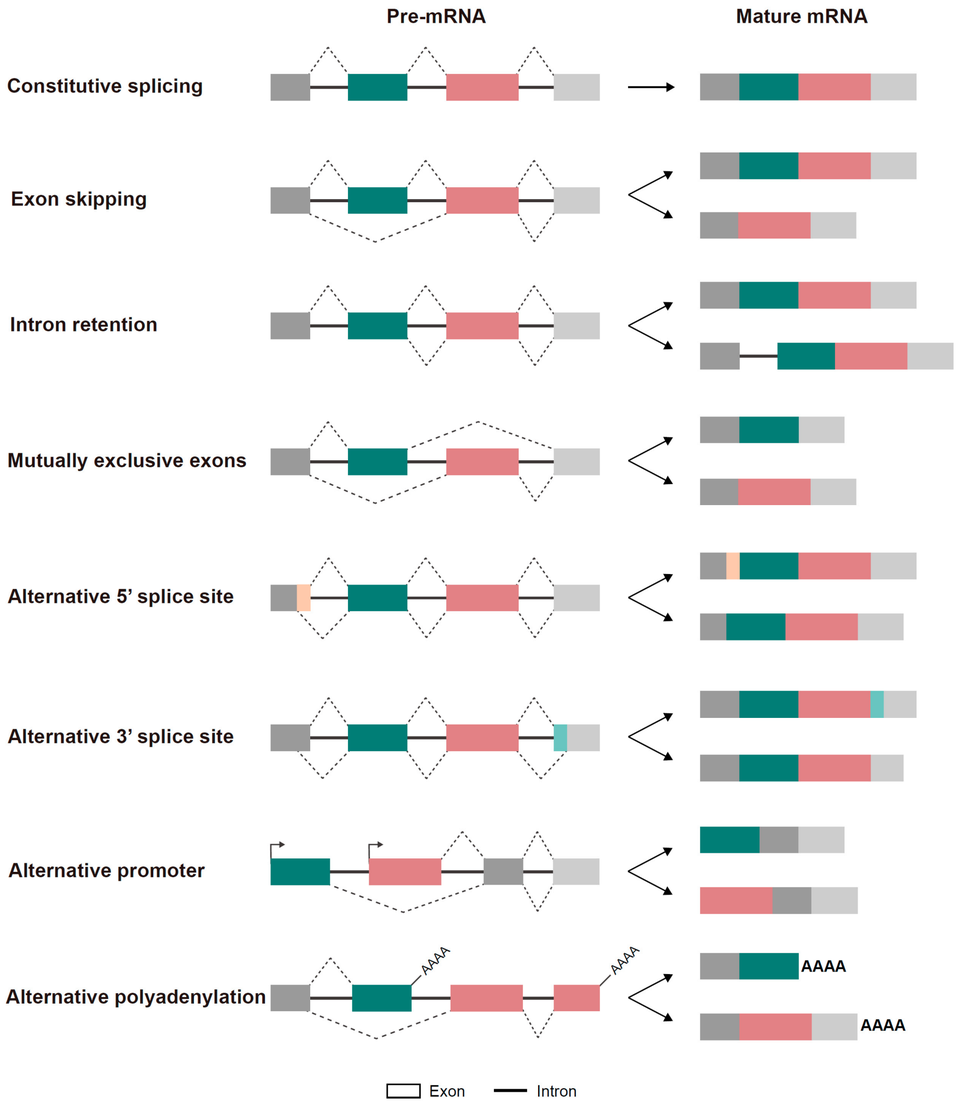
RNA splicing

How does this work?

- **ATP** is used for the assembly and re-arrangements of the spliceosome
- Accuracy is increased because the splicing is **coupled** to transcription (prevents from skipping splice sites)
- Details are **not fully understood**

Concept of alternative splicing

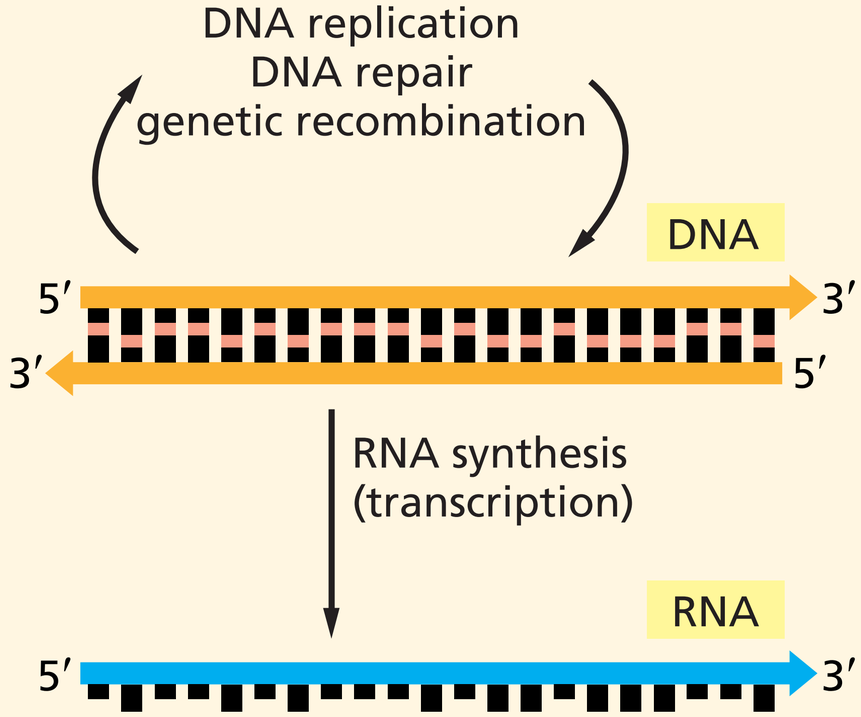
- It facilitates the emergence of **new proteins**
- Provides the ability to synthesize several **different proteins** from the **same gene**



II. DNA to RNA

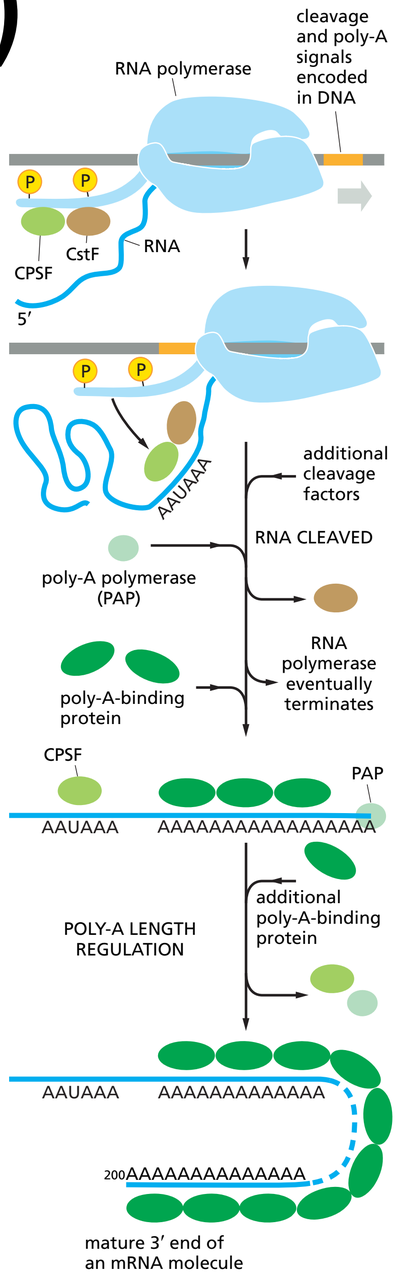
1. RNA
2. Transcription
3. Transcription initiation
4. **RNA processing**
5. Non-coding RNAs

a. RNA capping
b. RNA splicing
c. PolyA tail



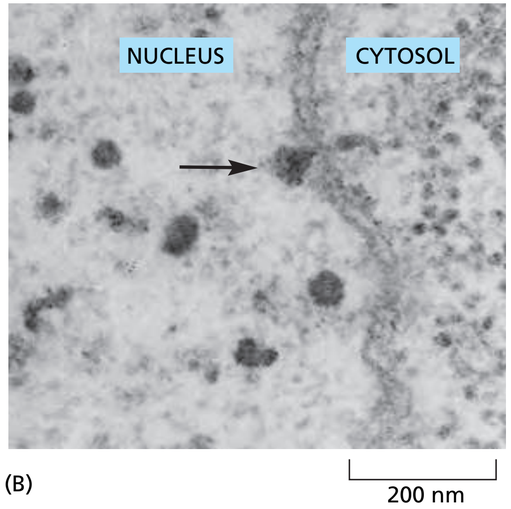
Where to stop transcription (for mRNAs)

- The position of the 3’ end is specified by **signals encoded in the genome** (typically AATAAA)
- These signals are transcribed into RNA and **recognised by RNA-binding-proteins and RNA-processing enzymes**
- **CstF and CPSF** bind their recognition sequence on the emerging RNA molecule
- RNA is **cleaved** from the polymerase
- **PolyA polymerase** adds **~200 A nucleotides** at the end of the sequence, creating the **poly-A tail**
- **poly-A binding proteins** bind to it



Mature mRNAs are exported from the nucleus

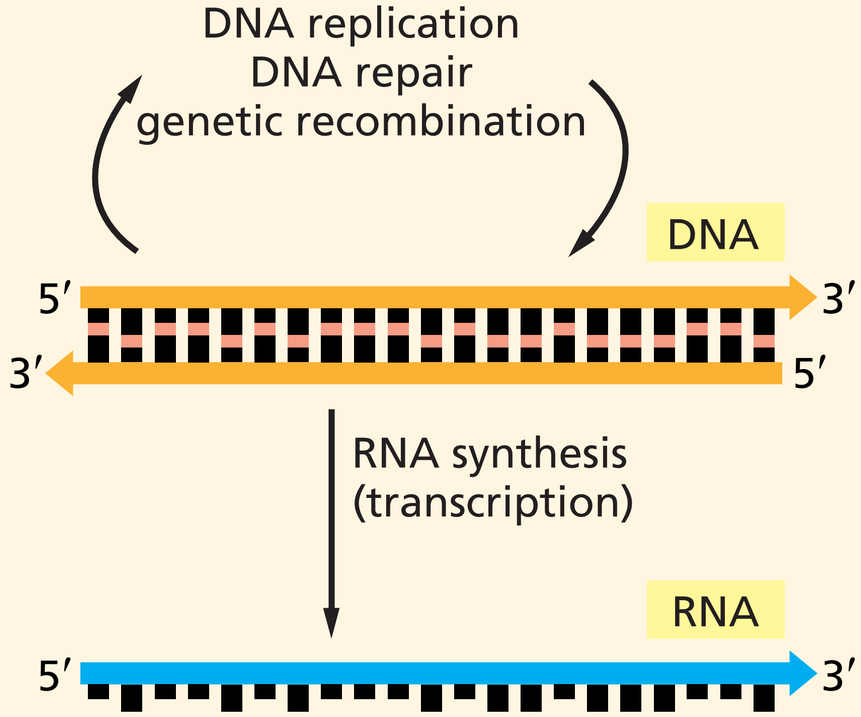
- How to know that they are **mature**?
  - ✓ cap-binding complexes
  - ✓ exon junction complexes
  - ✓ polyA binding proteins



- Processed mRNAs are guided through **nuclear pore complexes** (NPCs), where they are exported

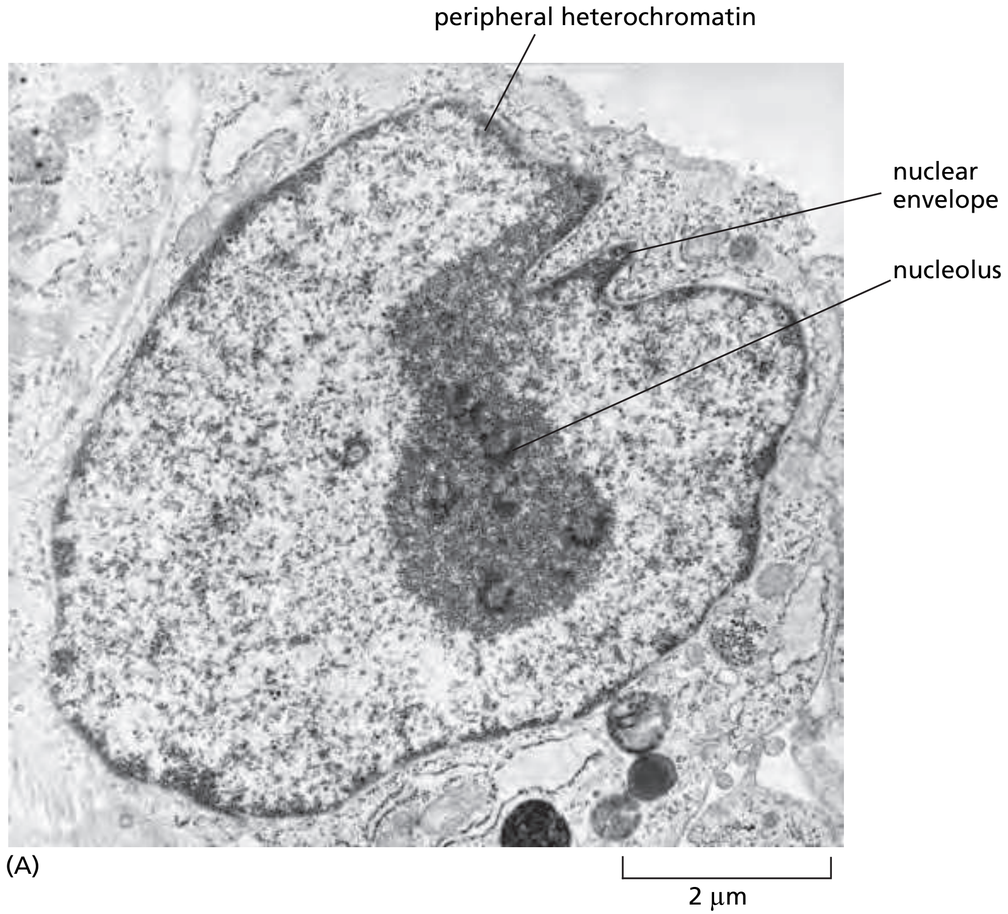
II. DNA to RNA

1. RNA
2. Transcription
3. Transcription initiation
4. RNA processing
5. **Non-coding RNAs**

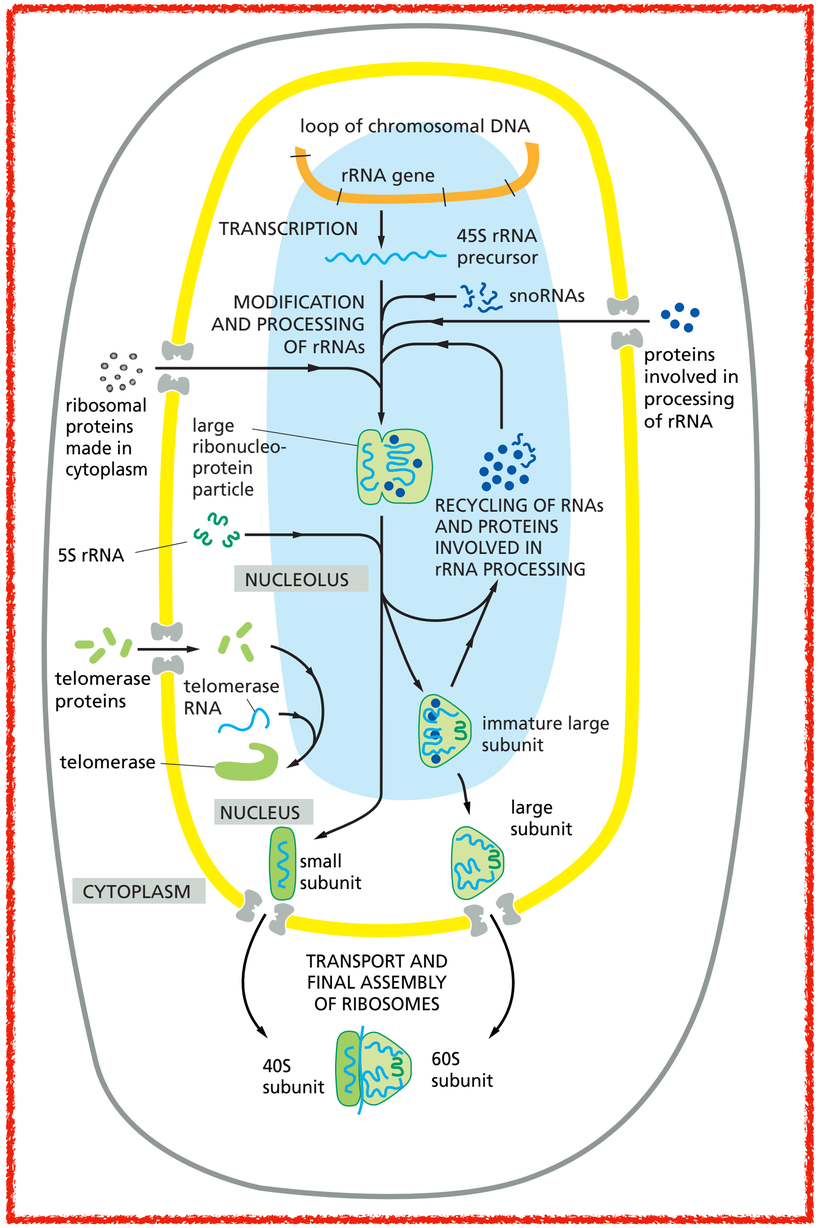


Non-coding RNAs: rRNAs

- 80% of the cellular RNA is **ribosomal RNAs** (rRNAs)
- Produced by **RNA polymerase I** in **eukaryotic** cells
- Does not have a **Ct tail** (which explains the lack of cap and poly-A tail of the rRNAs)
- **Multiple copies** of rRNA genes (~200 in humans) that encode rRNAs to have enough ribosomes (no amplification at the translation level)
- Ribosomes are assembled in the **nucleolus**

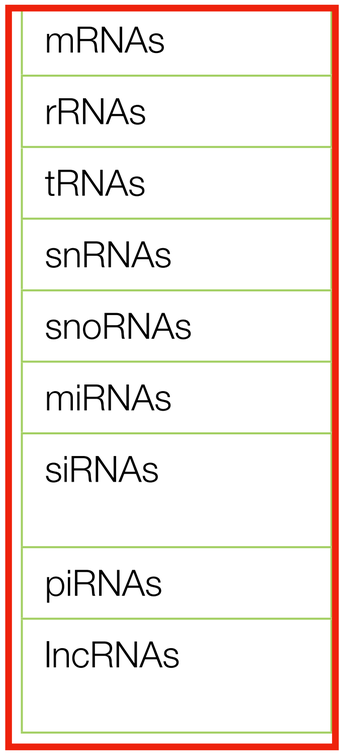


Ribosome assembly



*snoRNA = small nucleolar RNA

Which non-coding RNAs do we already know?



III. RNA to protein

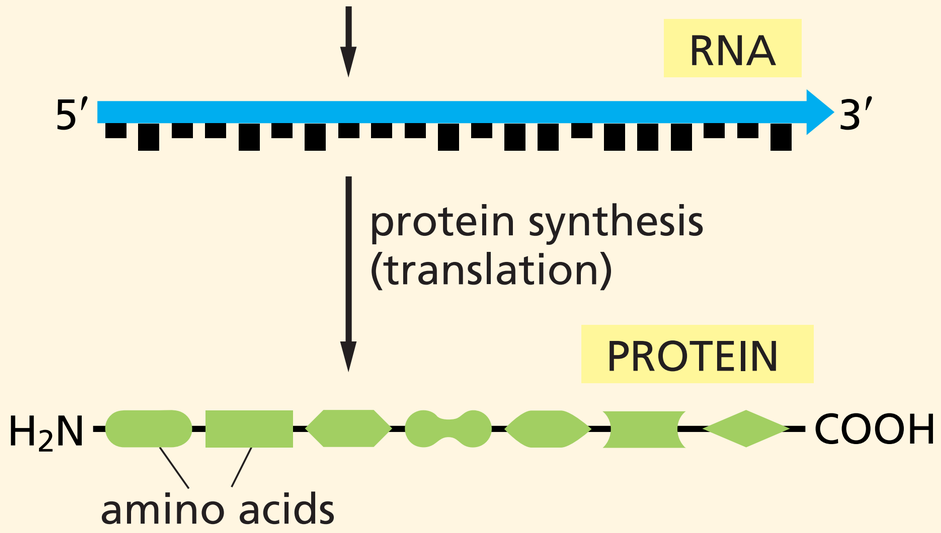
1. Translation

2. Transfer RNA (tRNA)

3. Polypeptide chain

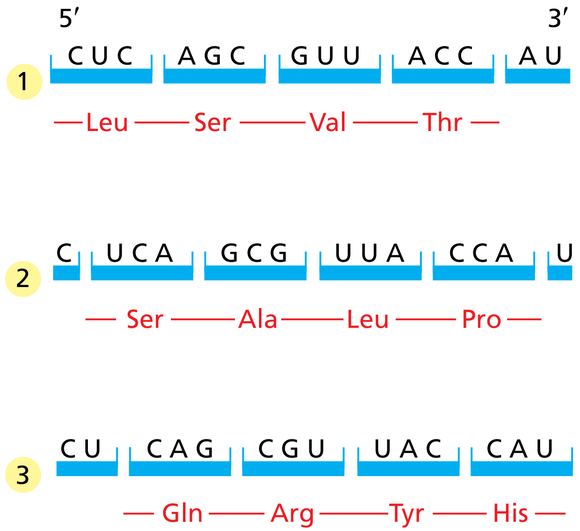
4. Ribosomes

5. Protein folding



Translation

- = conversion of information from **RNA** into **proteins**
- 4 nucleotides and 20 amino-acids — **genetic code**
- Used in **all organisms** (some small variations)
- A sequence has **three reading frames**

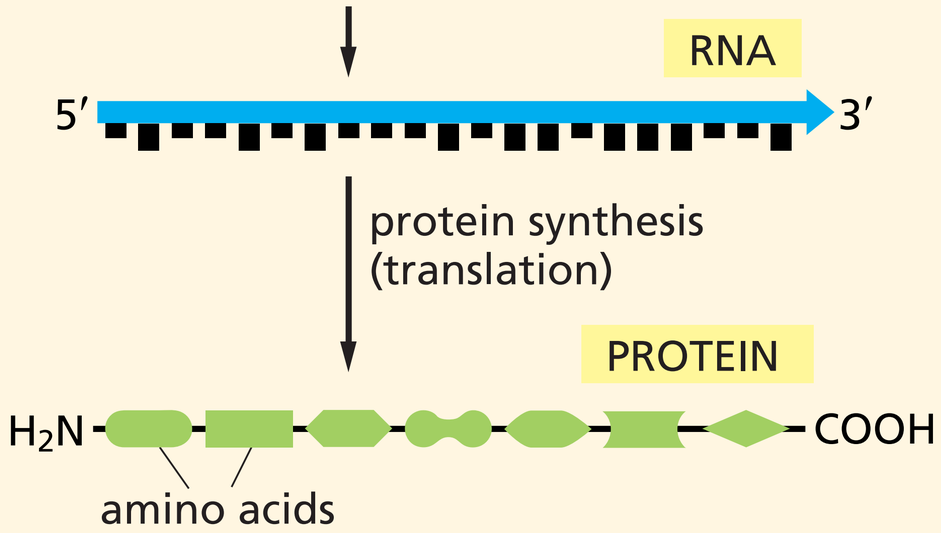


GCA GCC GCG GCU	AGA AGG CGA CGC CGG CGU	GAC GAU	AAC AAU	UGC UGU	GAA GAG	CAA CAG	GGA GGC GGG GGU	CAC CAU	AUA AUC AUU	UUA UUG CUA CUC CUG CUU	AAA AAG	AUG	UUC UUU	CCA CCC CCG CCU	AGC AGU UCA UCC UCG UCU	ACA ACC ACG ACU	UGG	UAC UAU	GUA GUC GUG GUU	UAA UAG UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

The genetic code is redundant (64 possible combinations for 20 amino acids)

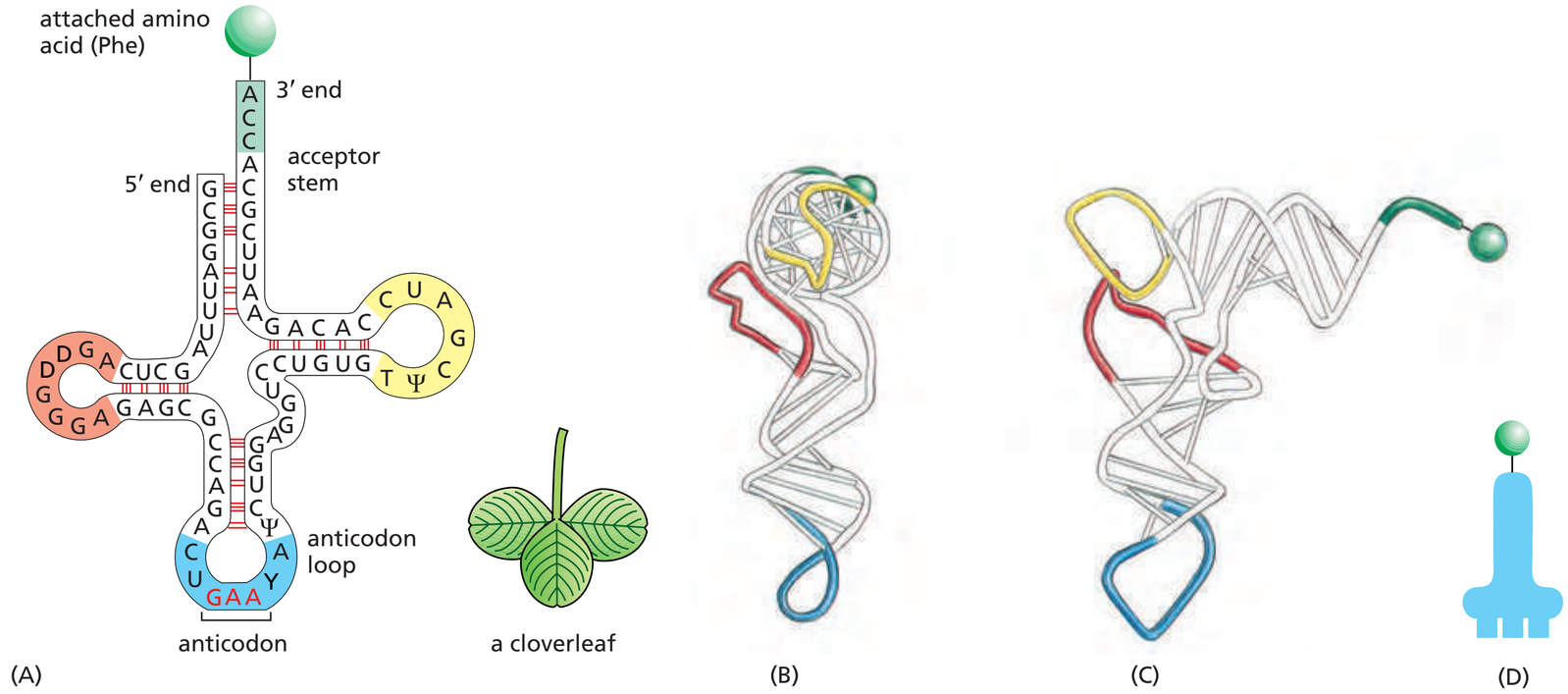
III. RNA to protein

1. Translation
2. **Transfer RNA (tRNA)**
3. Polypeptide chain
4. Ribosomes
5. Protein folding



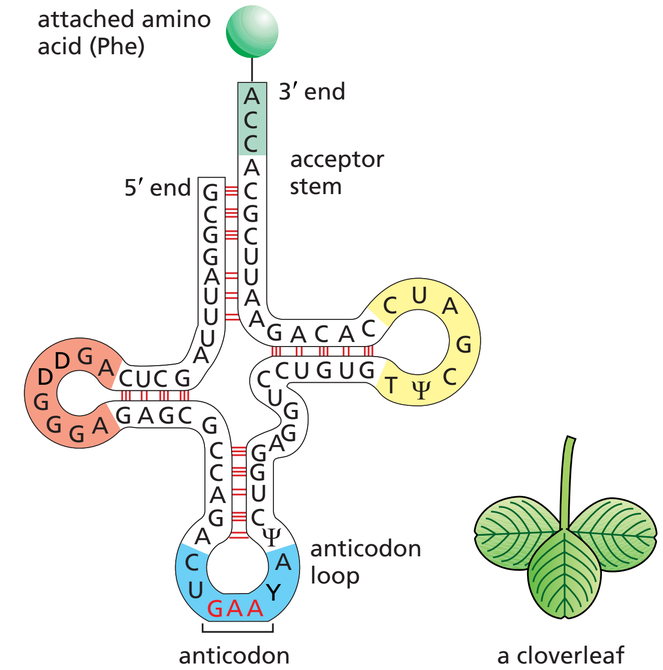
Transfer RNAs (tRNAs)

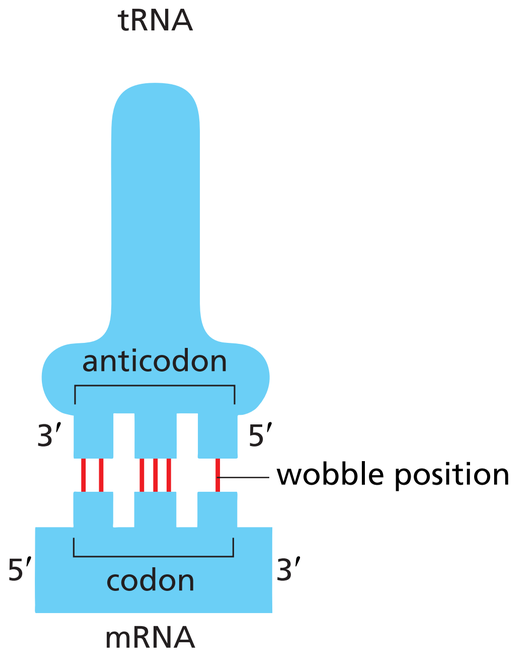
- the **codons** (triplet) is not directly recognised by an amino-acid
- Translation depends on *adaptors* = transfer RNA or **tRNA**
- ~80 nucleotides with **3D structure**



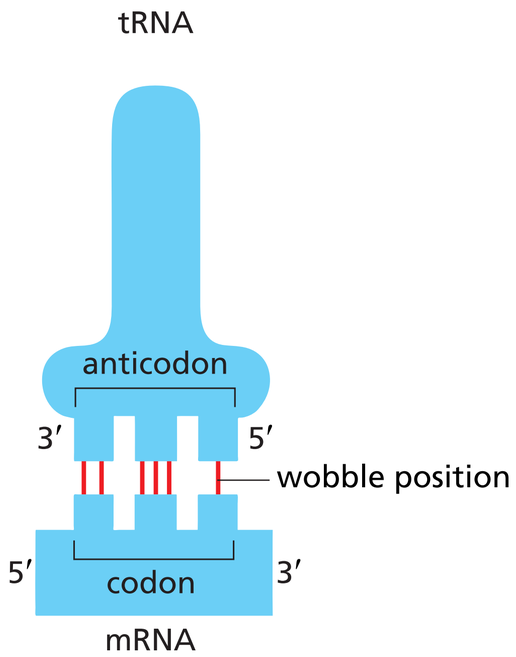
Transfer RNAs (tRNAs)

- **anticodon region:** 3 nucleotides that pair with the complementary codon in the mRNA molecule
- **3'end region:** binds the corresponding amino-acid
- Genetic code is redundant: some amino-acids have **more than one tRNA**; some tRNAs allow a **mismatch** (wobble) at the 3d position





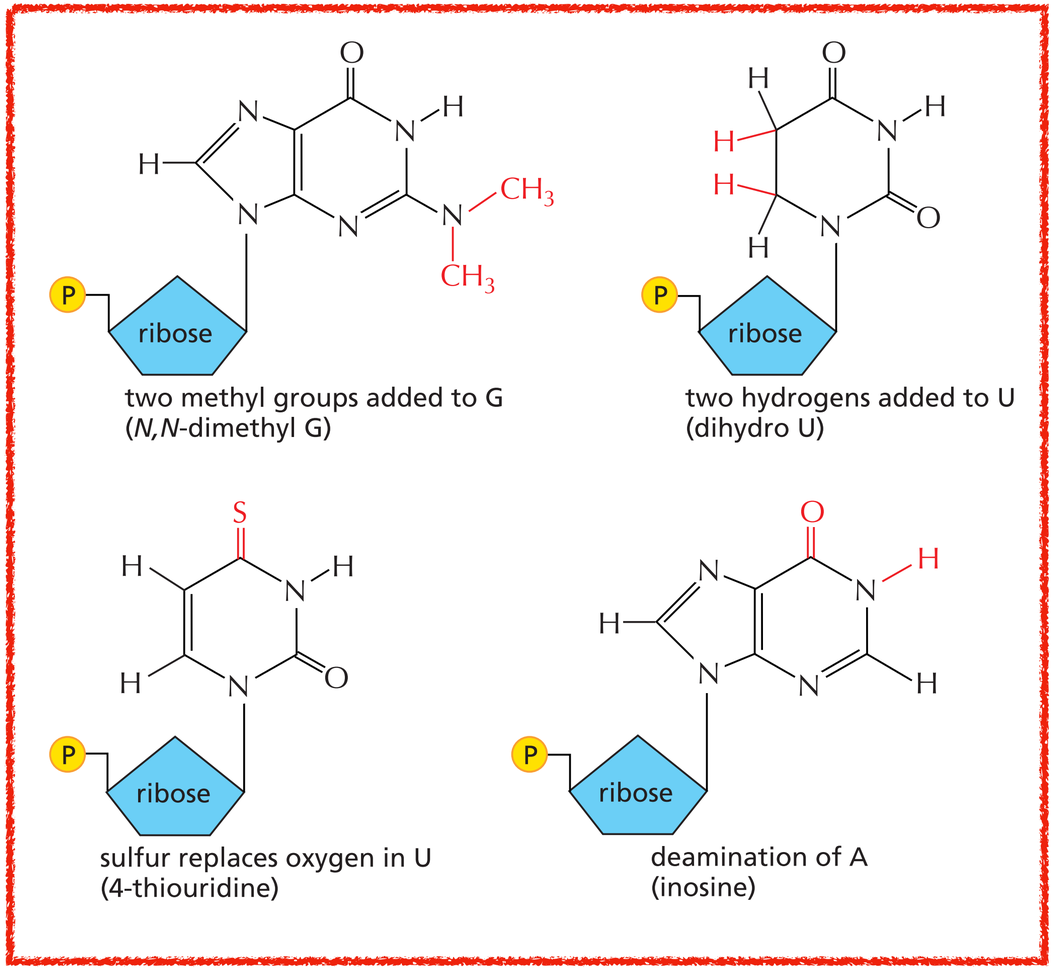
Transfer RNAs (tRNAs)



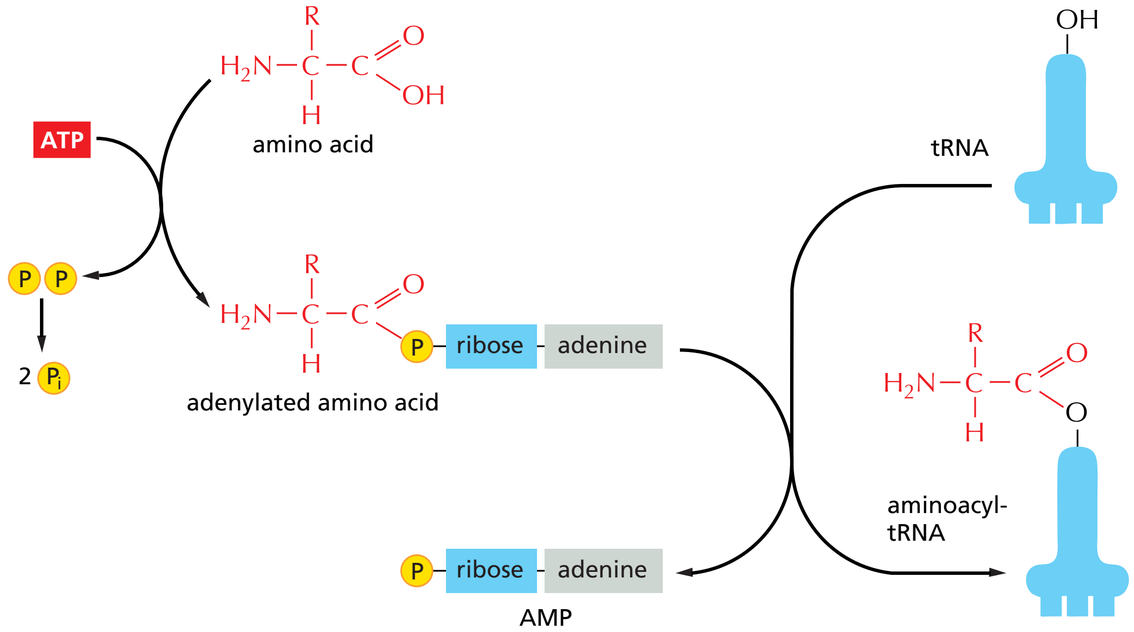
- The **wobble position** during translation refers to the third nucleotide (3′ end) of a codon on the mRNA — and correspondingly, the first nucleotide (5′ end) of the anticodon on the tRNA
- In the genetic code, each amino acid is specified by a 3-base codon (e.g., AUG = methionine)
- The **wobble position** is the **third base** of that codon (for example, the “G” in AUG)
- Base-pairing at this position is **less strict** — it can “wobble.”
- The first two bases of the codon pair **strictly** with the corresponding bases in the tRNA anticodon
- But at the third position, the base-pairing rules are **more flexible**
- This flexibility allows **one tRNA** to recognize **multiple codons** that code for the same amino acid
- It allows the cell to use **fewer tRNAs** (≈ 40) to recognize **61 sense codons**

Transfer RNAs (tRNAs)

- Synthesized by **RNA polymerase III** as larger precursors
- Trimmed to produce mature tRNAs
- Some have introns which must be **spliced out** by a cut-and-paste mechanism (different from mRNA) - this rarely also happens in some prokaryotes
- All tRNAs are **modified chemically**: 1 in 10 nt is an altered version of the initial ribonucleotide (>50 types of modifications are known)

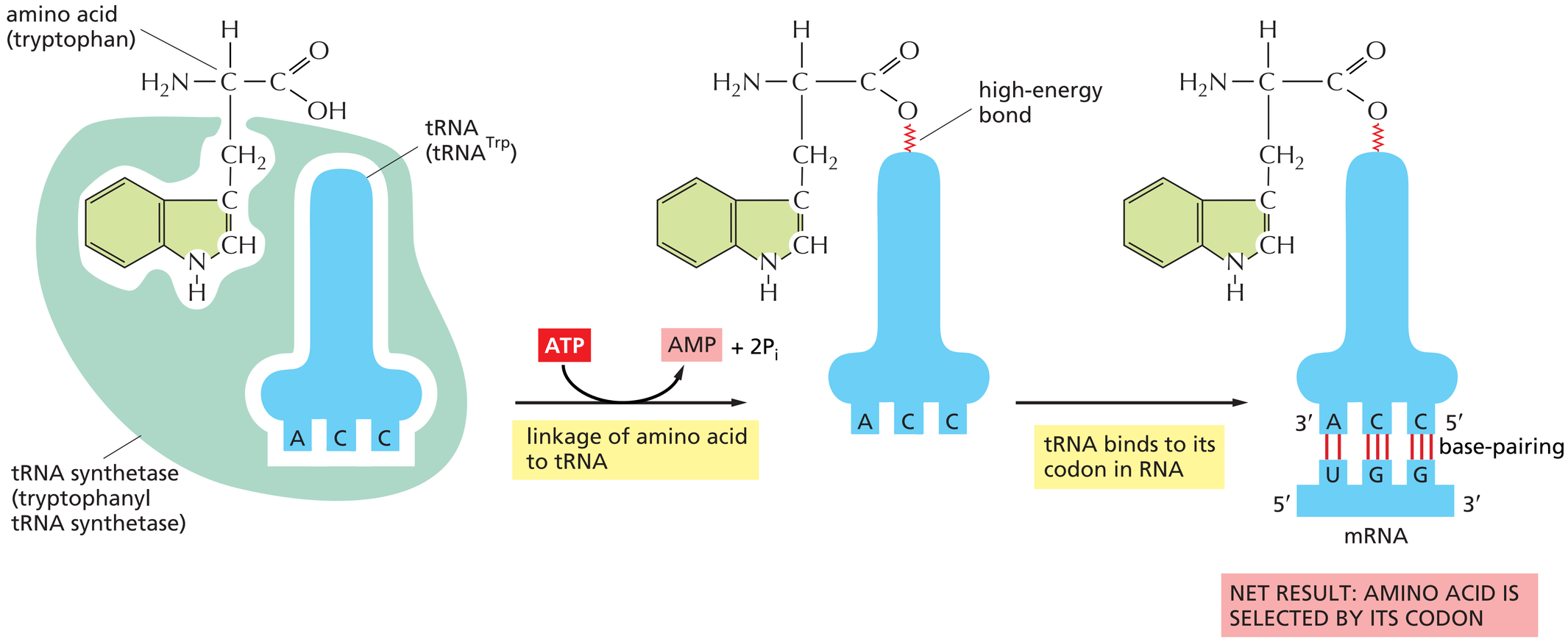


Transfer RNAs (tRNAs)

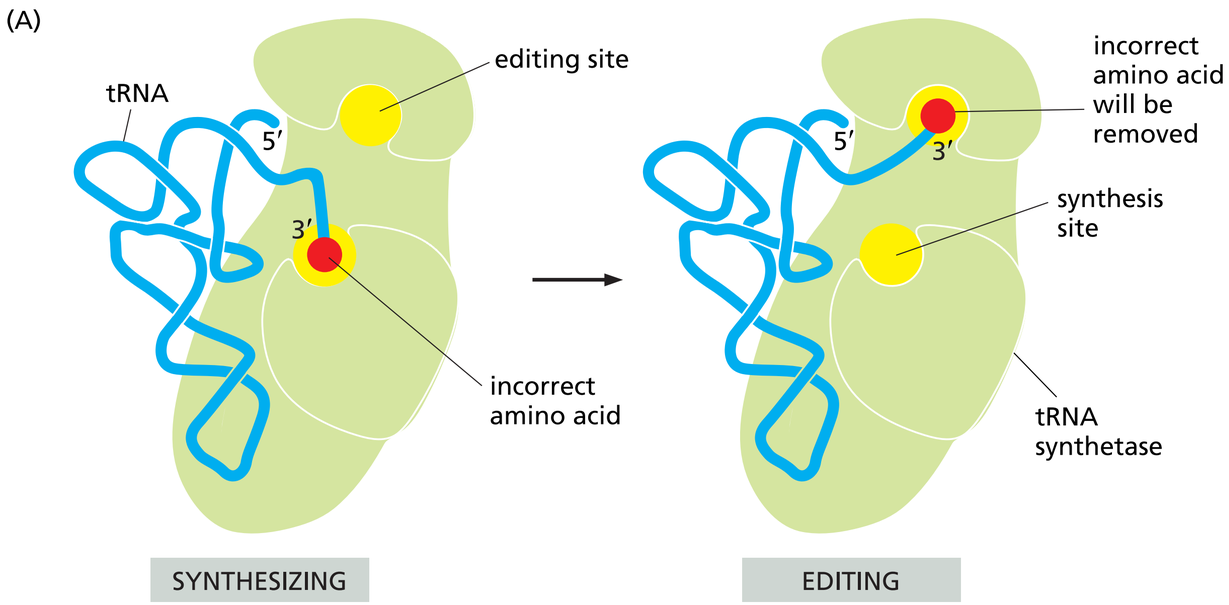


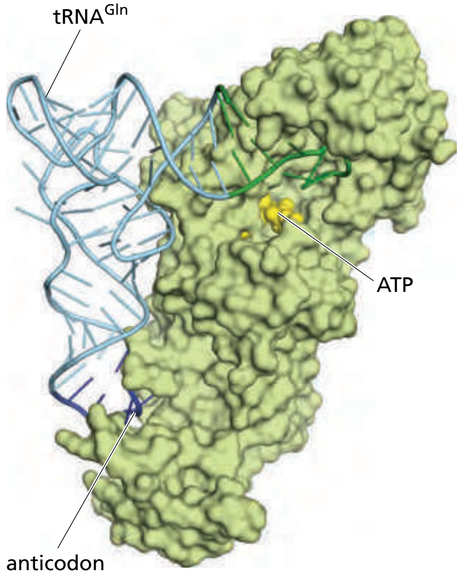
- **amino-acyl-tRNA synthetases** covalently couple each amino acid to their corresponding tRNAs
- reaction coupled with **energy-releasing hydrolysis of ATP**

Transfer RNAs (tRNAs)



Transfer RNAs (tRNAs)



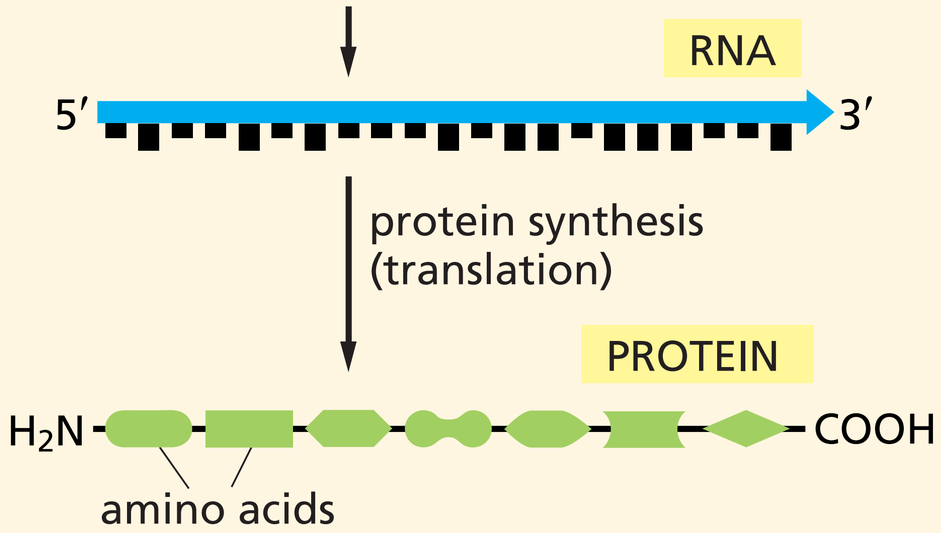


How to ensure that the right amino-acid is selected?

- Correct amino acid has **larger affinity**
- **Larger** amino acid **cannot fit** in the catalytic site of the enzyme
- For similar **size amino acids,** the covalent linkage can happen
- Further tested in the **editing site** - if it fits, it is not the right amino-acid
- The enzymes also need to recognize the **anticodons** of the tRNA

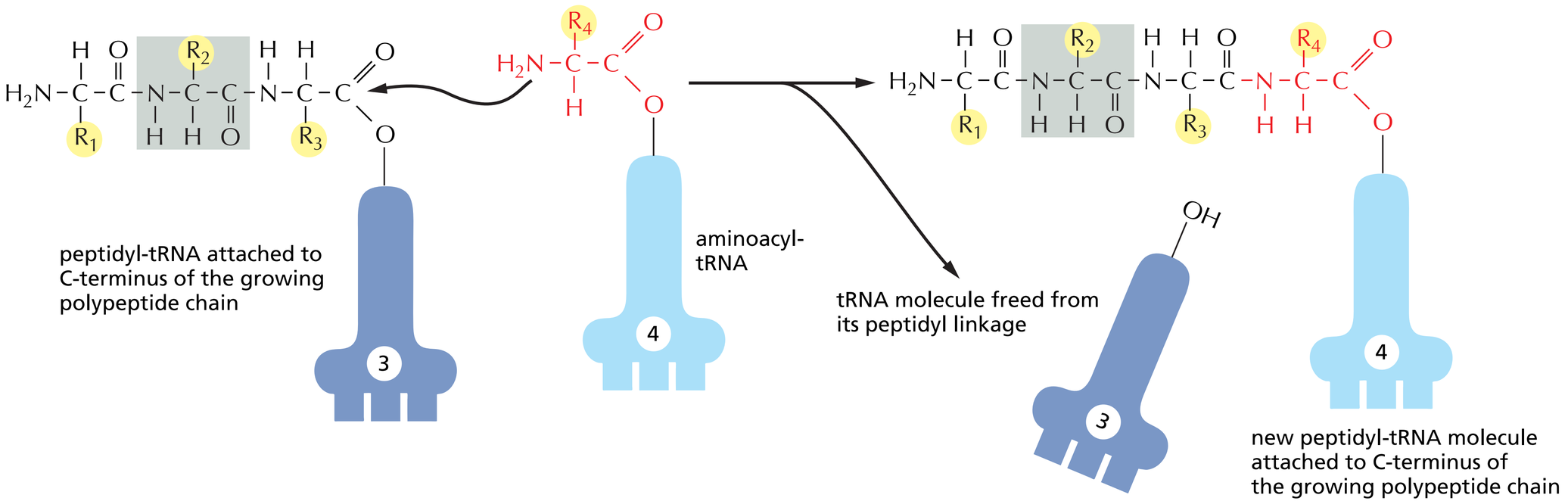
III. RNA to protein

1. Translation
2. Transfer RNA (tRNA)
3. **Polypeptide chain**
4. Ribosomes
5. Protein folding



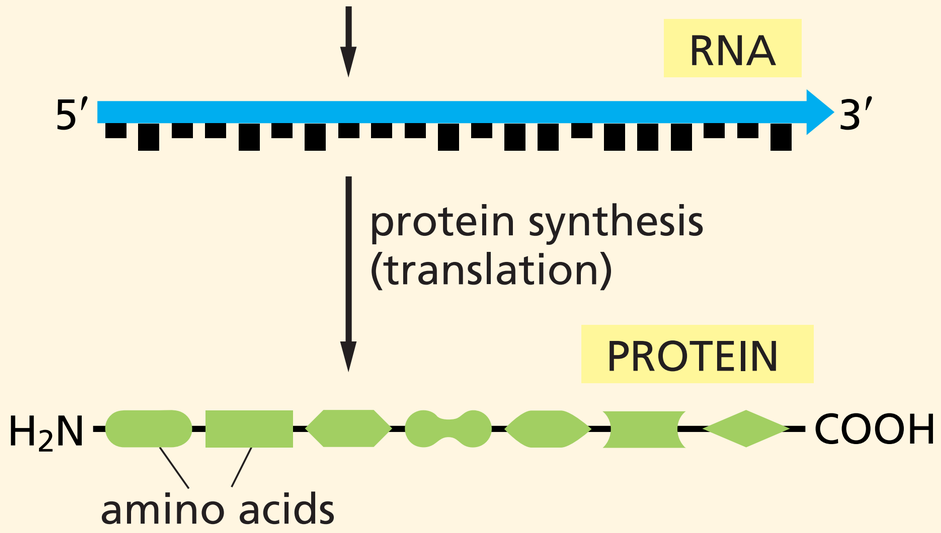
Polypeptide chain

- Formation of a **peptide bond** between the **carboxyl group** at the end of a growing polypeptide chain and a free **incoming amino acid** (assembly from N-terminal to C-terminal)



III. RNA to protein

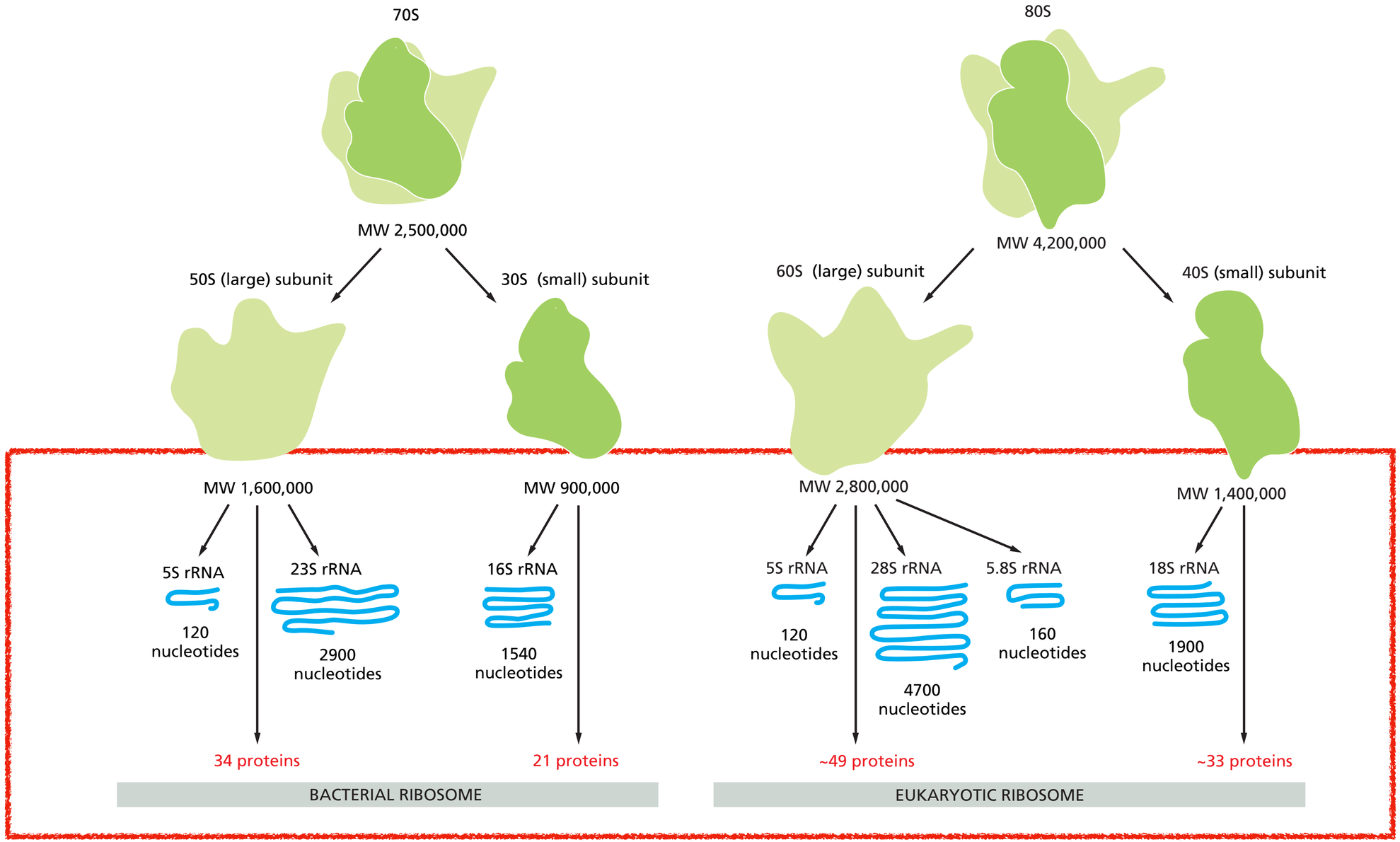
1. Translation
2. Transfer RNA (tRNA)
3. Polypeptide chain
4. **Ribosomes**
5. Protein folding



Ribosomes

- Perform **protein synthesis** (maintain the correct reading frame and ensure accuracy)
- Made of **>50 proteins** and **ribosomal RNAs** (rRNAs)
- **Millions** of ribosomes in the cytoplasm of a eukaryotic cell
- Consist of a **large and small subunit**, each assembled in the **nucleus (nucleolus)**
- The two subunits are **exported to the cytoplasm**, where they **join on** an mRNA molecule

Ribosomes

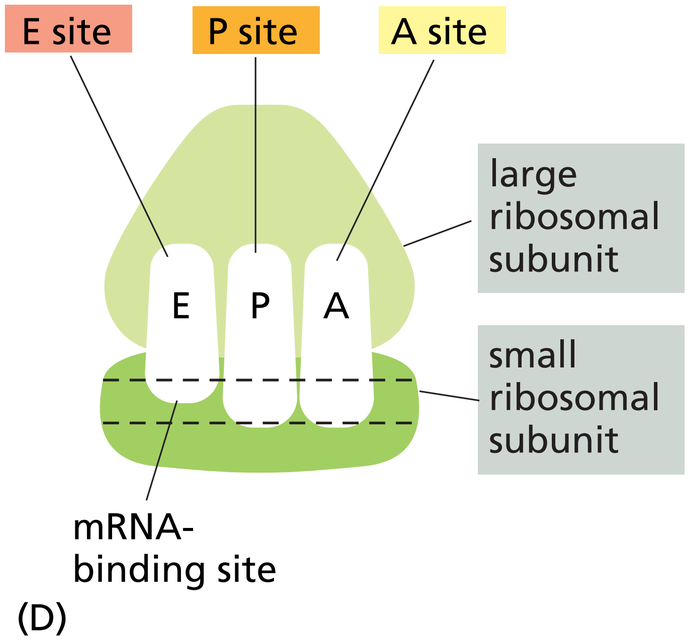

70S
MW 2,500,000
50S (large) subunit
30S (small) subunit
MW 1,600,000
MW 900,000
5S rRNA
120 nucleotides
23S rRNA
2900 nucleotides
16S rRNA
1540 nucleotides
34 proteins
21 proteins
BACTERIAL RIBOSOME
80S
MW 4,200,000
60S (large) subunit
40S (small) subunit
MW 2,800,000
MW 1,400,000
5S rRNA
120 nucleotides
28S rRNA
4700 nucleotides
5.8S rRNA
160 nucleotides
18S rRNA
1900 nucleotides
~49 proteins
~33 proteins
EUKARYOTIC RIBOSOME

Ribosomes

- 4 **binding sites** for RNA:
  - 1 for **mRNA**
  - 3 for **tRNAs** (A, P and E sites)

- a tRNA is kept in A and P sites if its **anticodon** forms base-pairs with a **complementary codon**

- A and P sites are **close enough** to accommodate adjacent codons (maintains the correct reading frame)



(D)

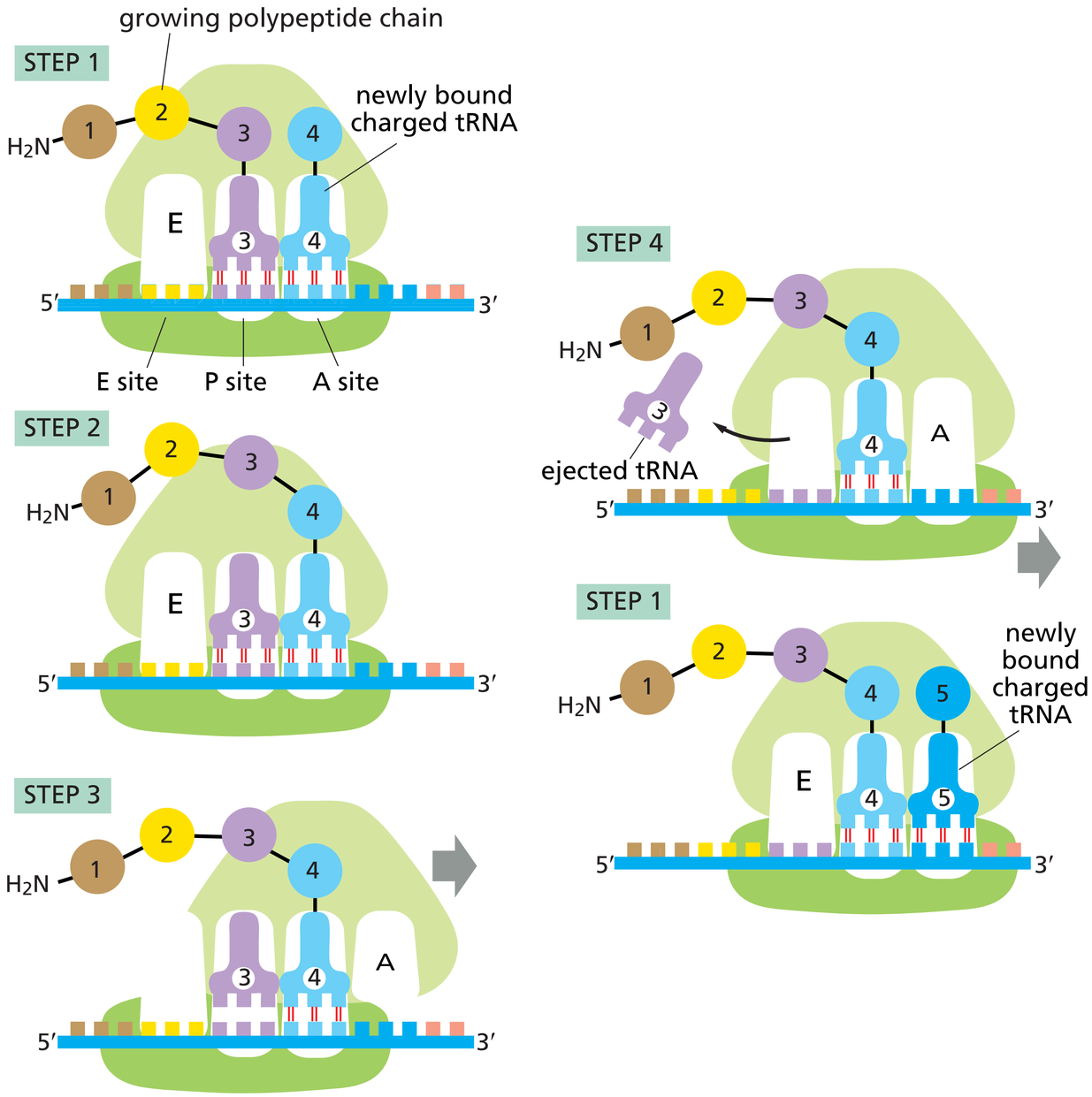
Protein synthesis

Step 1: tRNA binding

Step 2: peptide bond formation by the **peptidyl transferase** in the large subunit

Step 3: large subunit translocation

Step 4: small subunit translocation



Elongation factors

- Make translation more **efficient** and **accurate**
- **EF1** and **EF2** in eukaryotes; **EF-Tu** and **EF-G** in bacteria

Process	Molecule Produced	Machinery	Purpose of Elongation Factors
Transcription	RNA (from DNA)	RNA polymerase	Help RNA polymerase move along DNA and add ribonucleotides efficiently
Translation	Protein (from mRNA)	Ribosome	Help the ribosome add amino acids to a growing polypeptide efficiently

Where to start protein synthesis?

- Important as it sets the **reading frame** for the whole protein
- Starts with at **AUG codon** and a special tRNA to start translation, which carries **methionine**
- All proteins start with **methionine**
- The special tRNA is recognised by **initiation factors**

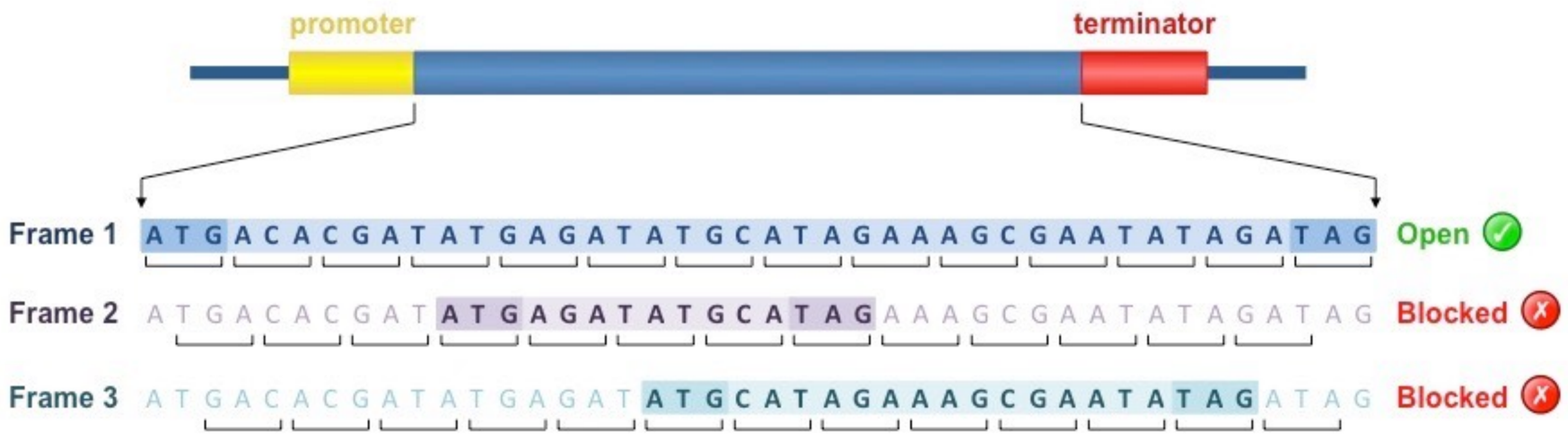
Reading frames

- An **open reading frame (ORF)** is a continuous stretch of codons in DNA or mRNA that can potentially be translated into a protein (no premature stop codons)

- It begins with a **start codon** (usually **AUG**) and ends with a **stop codon** (**UAA**, **UAG**, or **UGA**).

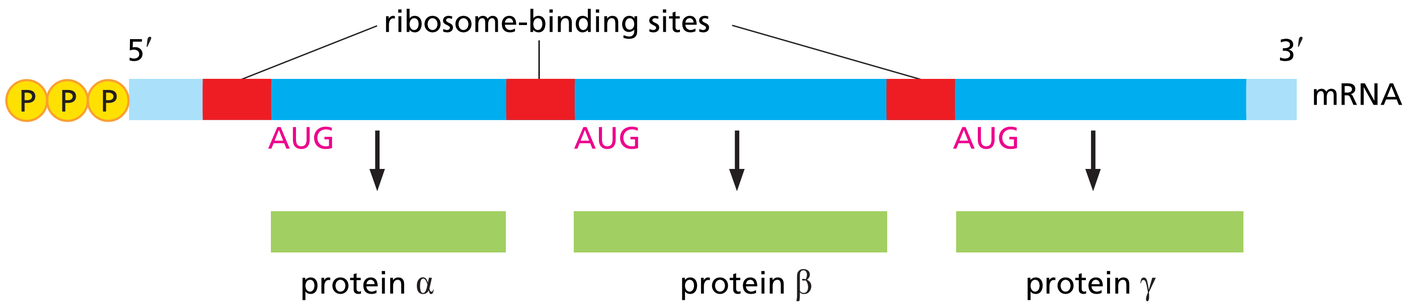
- There are **three possible reading frames** on each strand of DNA (depending on where you start reading).

- For double-stranded DNA, that makes **six possible reading frames** total (three on each strand).



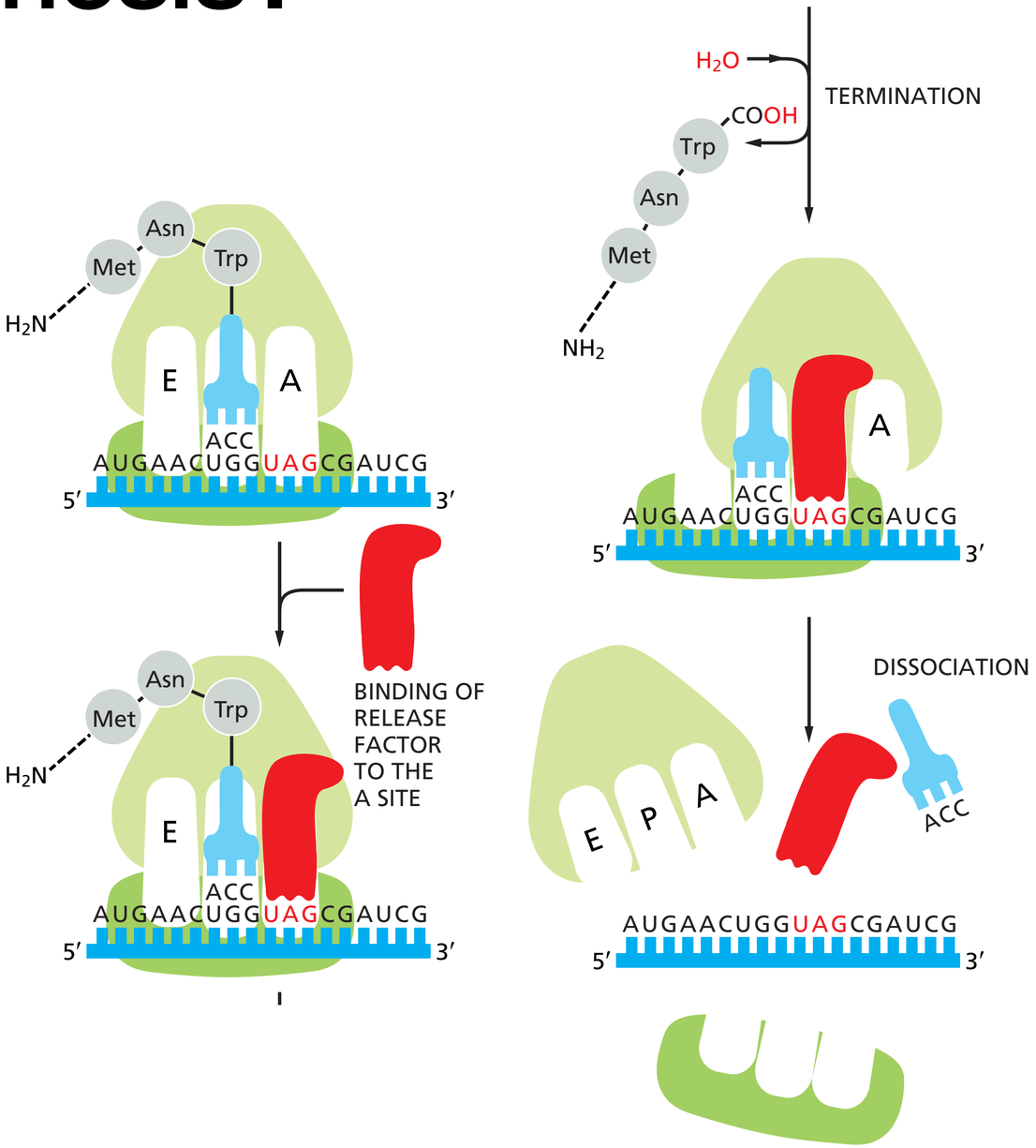
Where to start protein synthesis?

- In **bacteria**, no cap but a **specific ribosomal binding site** (called the **Shine-Dalgarno** sequence)
- This sequence is recognised by the **16S rRNA**, which **positions** the ribosomes properly to read the first AUG
- Bacterial mRNAs are often **polycistronic** (i.e. they contain multiple genes that will produce different proteins)

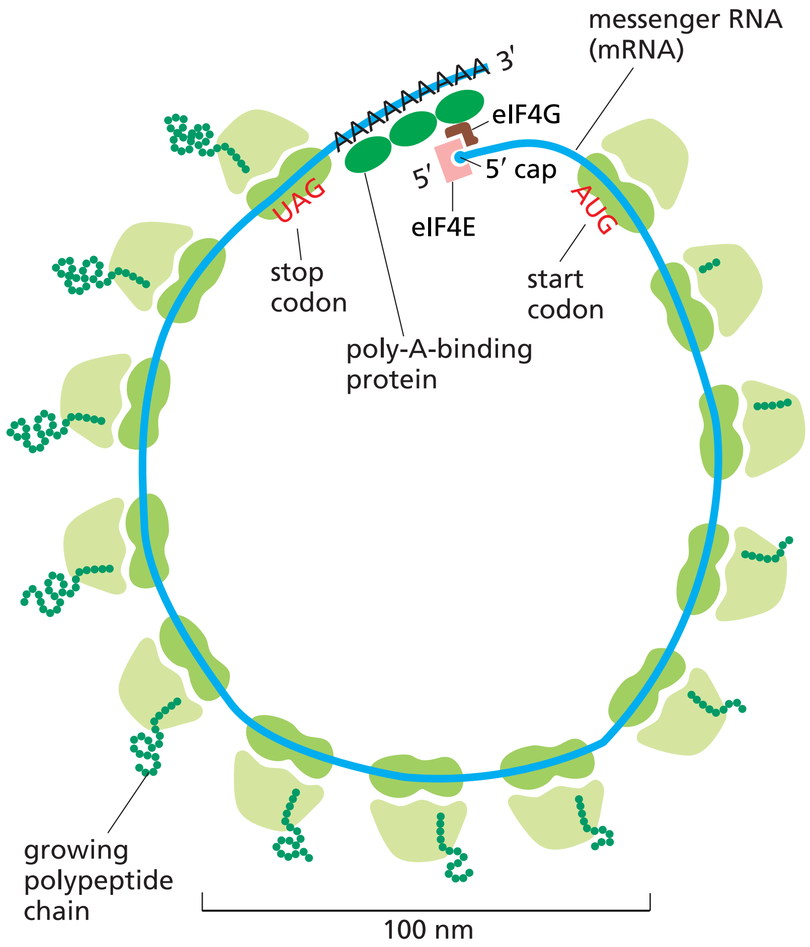


Where to stop protein synthesis?

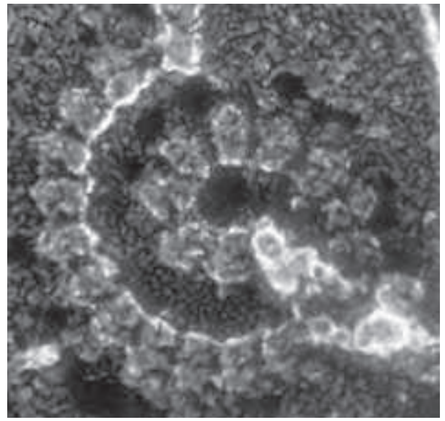
- The end of the sequence is signalled by **stop codons** (UAA, UAG or UGA)
- No **tRNA** and amino-acid but signal to the ribosome to **stop translation**
- Binding of **release factors**



Proteins are made on polyribosomes (=polysomes)

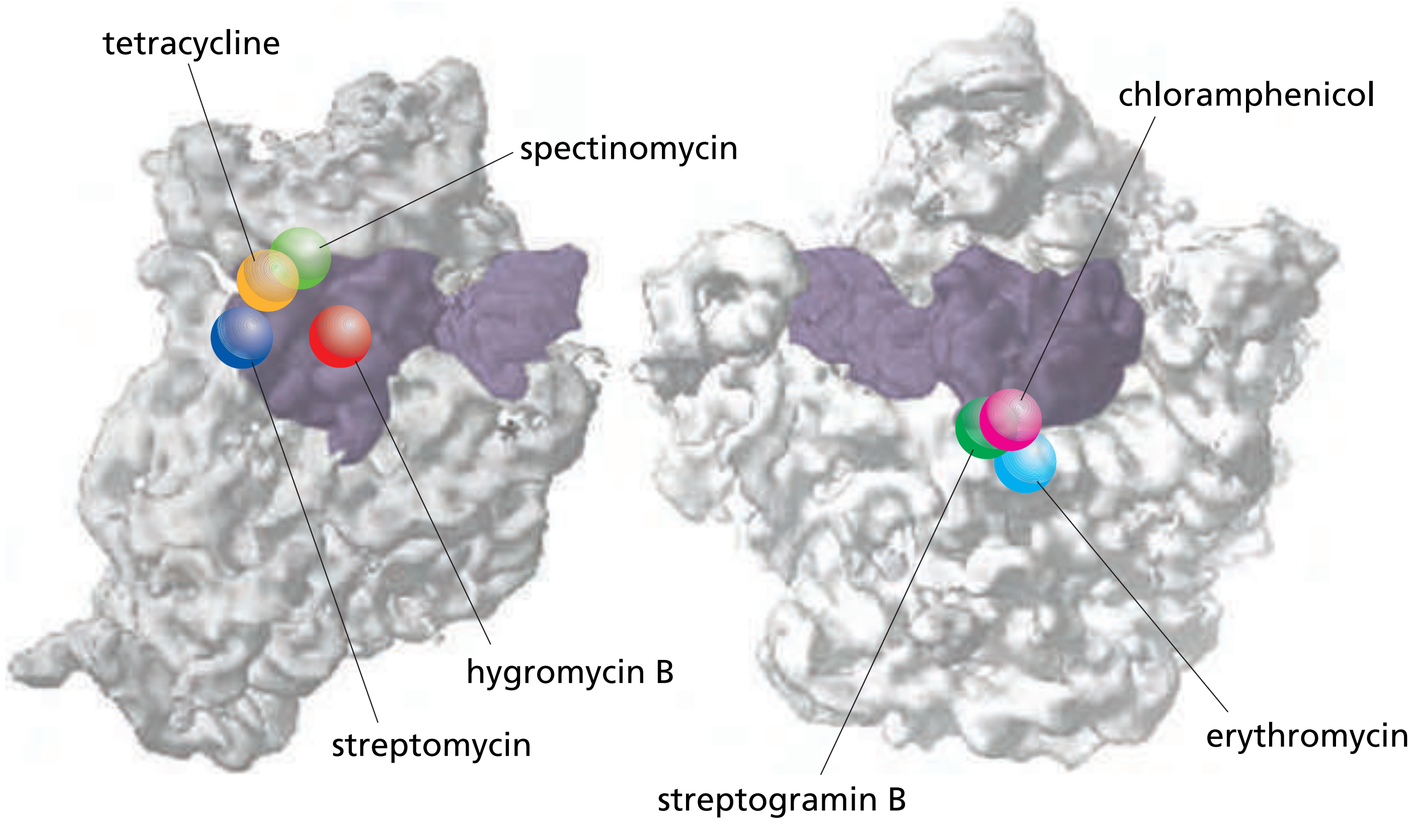


(A)



(B)

Bacterial synthesis as a target for antibiotics

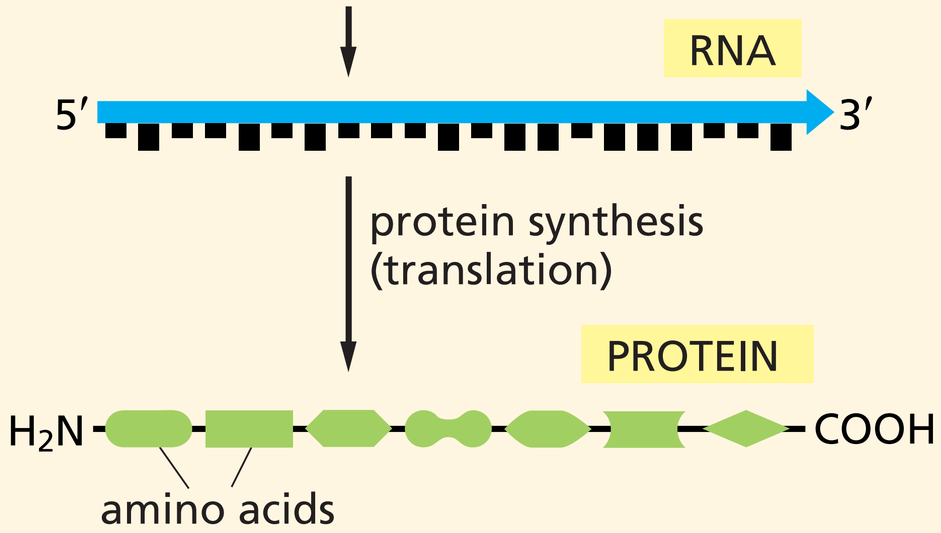


small ribosomal subunit

large ribosomal subunit

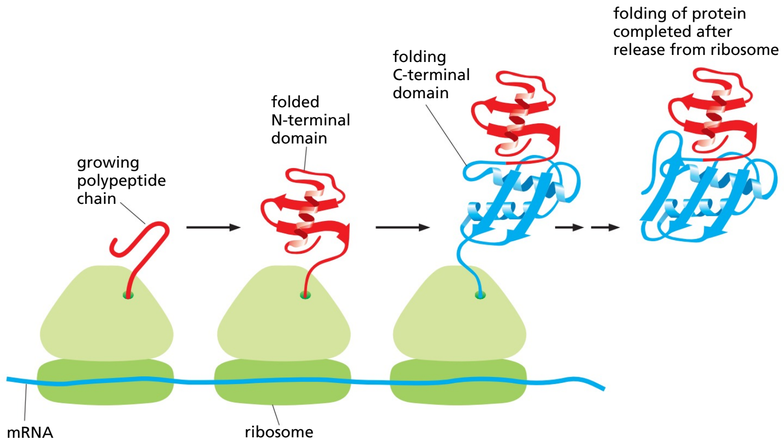
III. RNA to protein

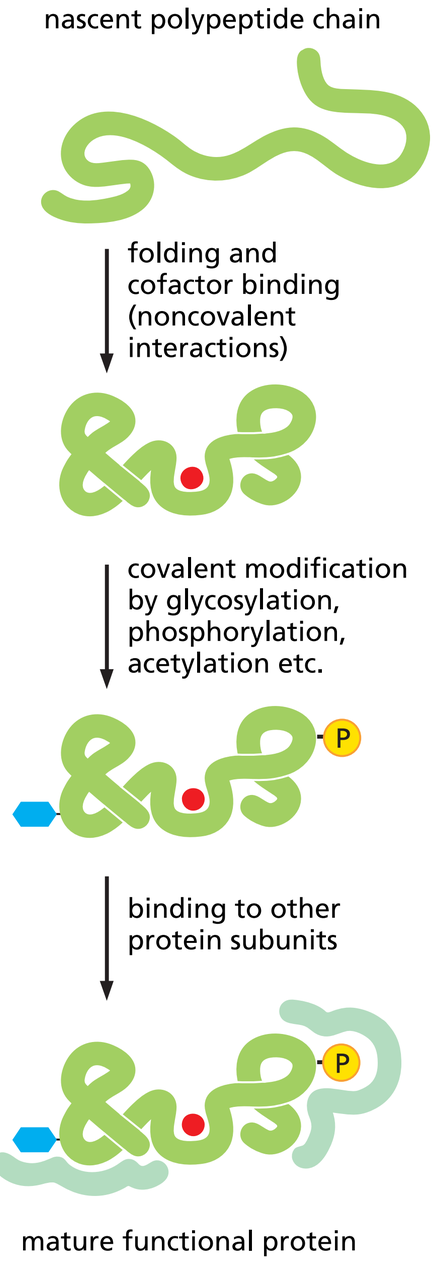
1. Translation
2. Transfer RNA (tRNA)
3. Polypeptide chain
4. Ribosomes
5. **Protein folding**



Protein folding

- **Polypeptide chains** must reach their **3D structure** to be active as **proteins**
- **All information** needed is part of the **sequence**
- The protein core is typically **hydrophobic**
- Conformation with **lowest free energy**
- For some proteins, folding occurs as soon **as the chain emerges from the ribosome**

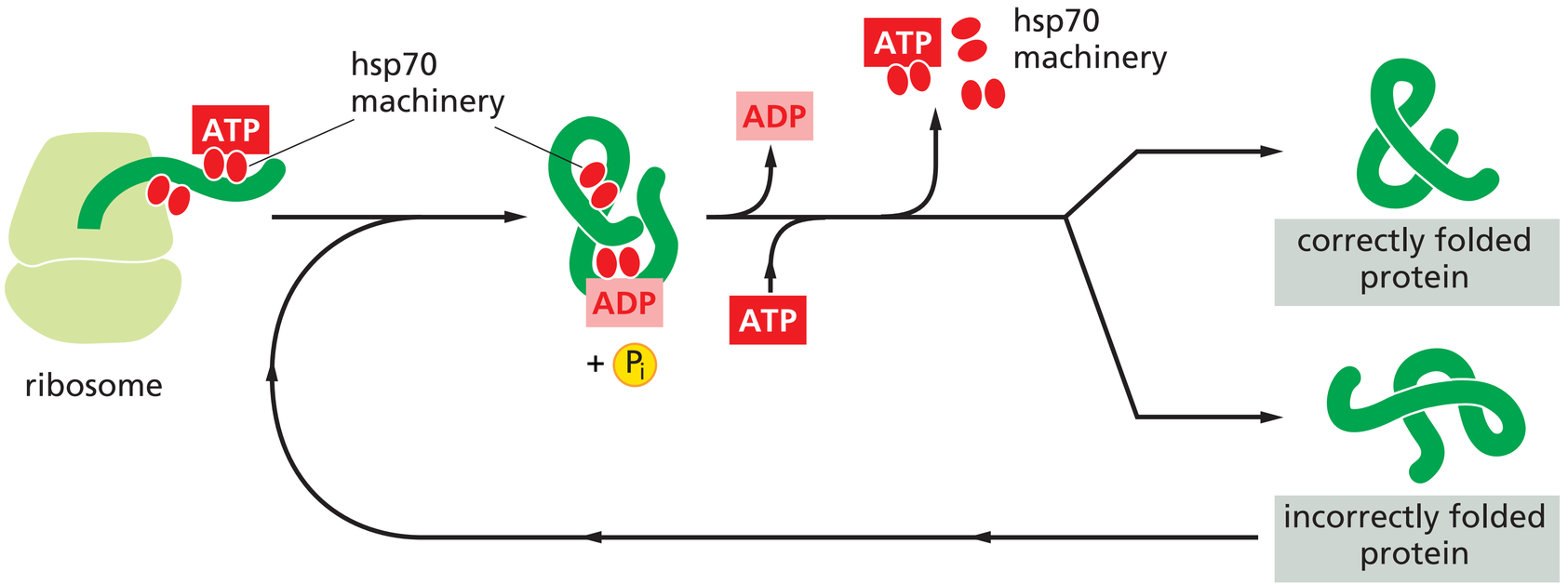




Molecular chaperones

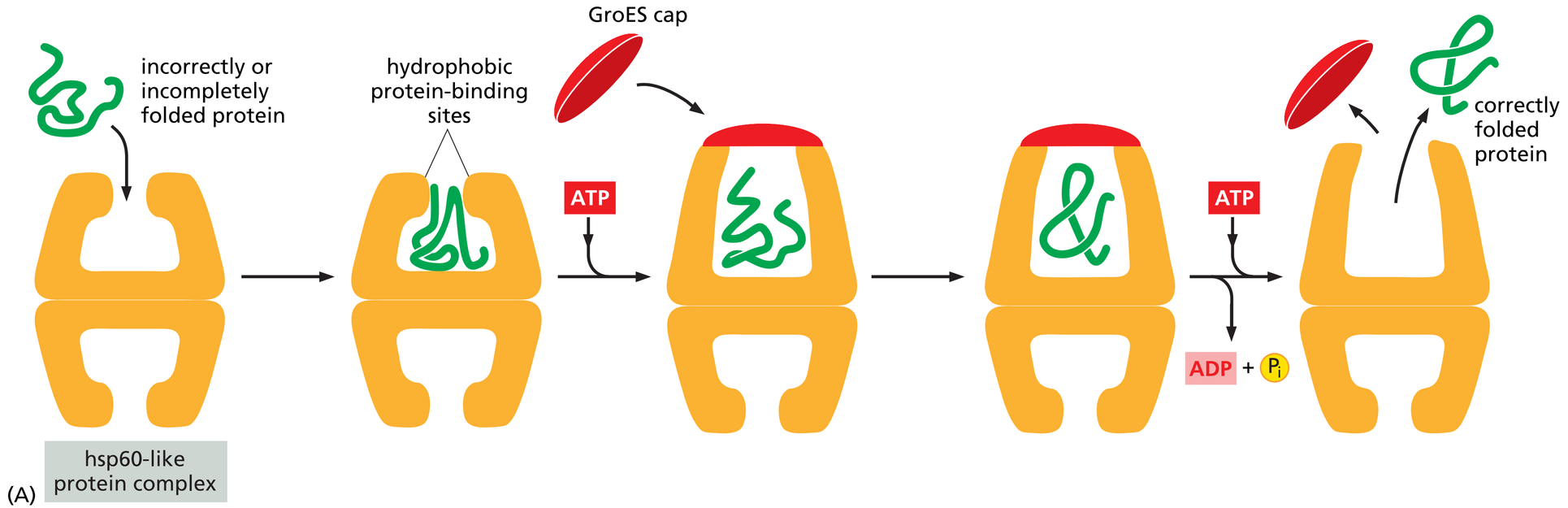
- All the information needed for proteins to get to their **3D structure** is encoded in their **AA sequence**
- Yet, some proteins require the help of **molecular chaperones** to properly get to their folding and would otherwise aggregate
- Chaperones recognise incorrect folding by exposure of **hydrophobic surfaces**
- They are also called **heat shock proteins** (hsp), as they were discovered because they are overexpressed during heat shock
- Major families are **Hsp60** and **Hsp70** + associates
- Use **ATP** to bind and release unfolded proteins

Hsp70



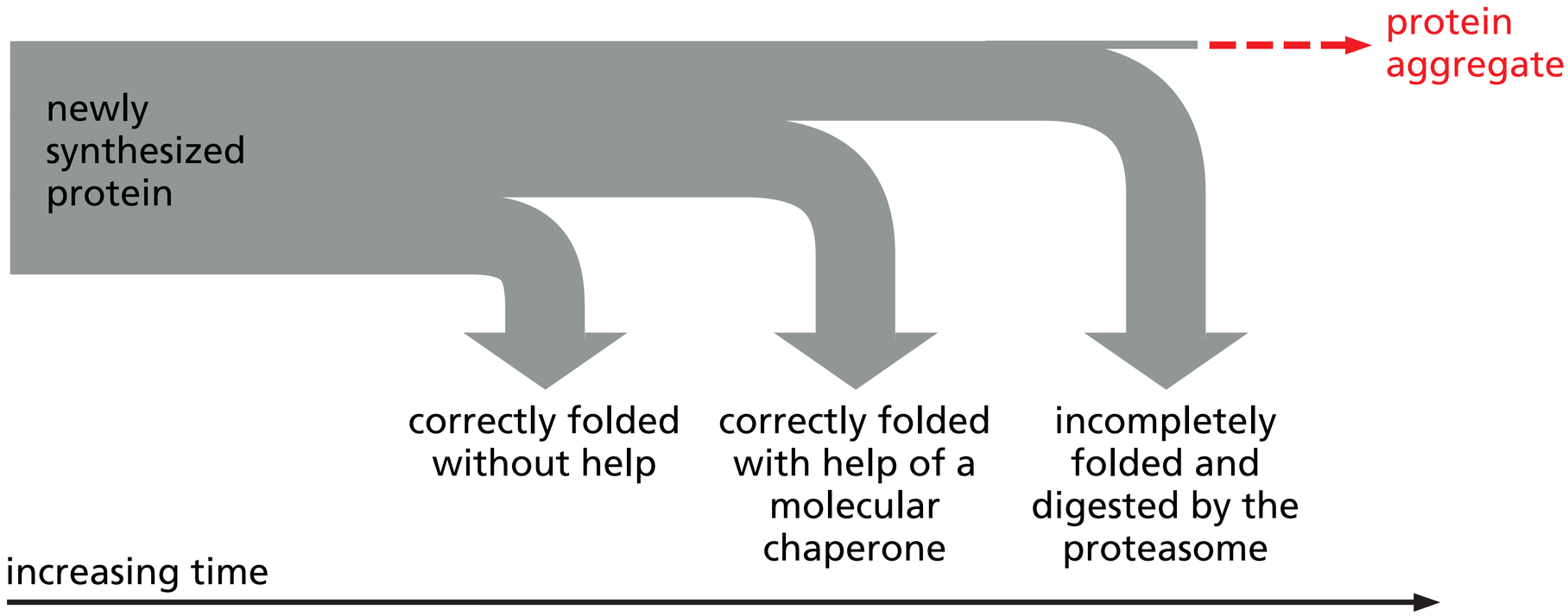
- Acts **early** in the protein life
- **Binding and release** of the substrate

Hsp60



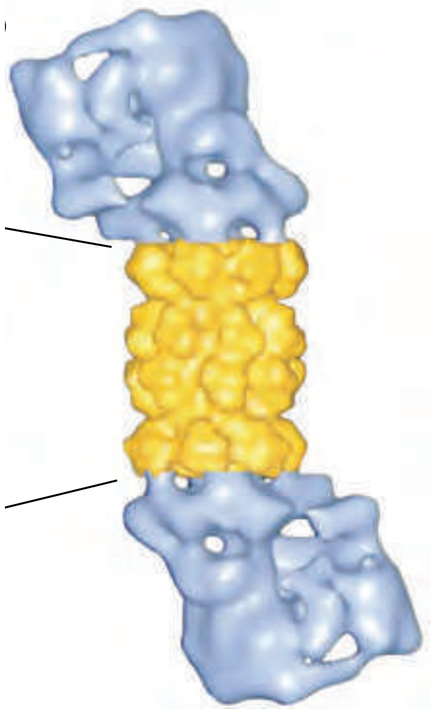
- Acts when the **whole protein** is synthesized
- **Isolation chamber** for the folding process

Quality control choices



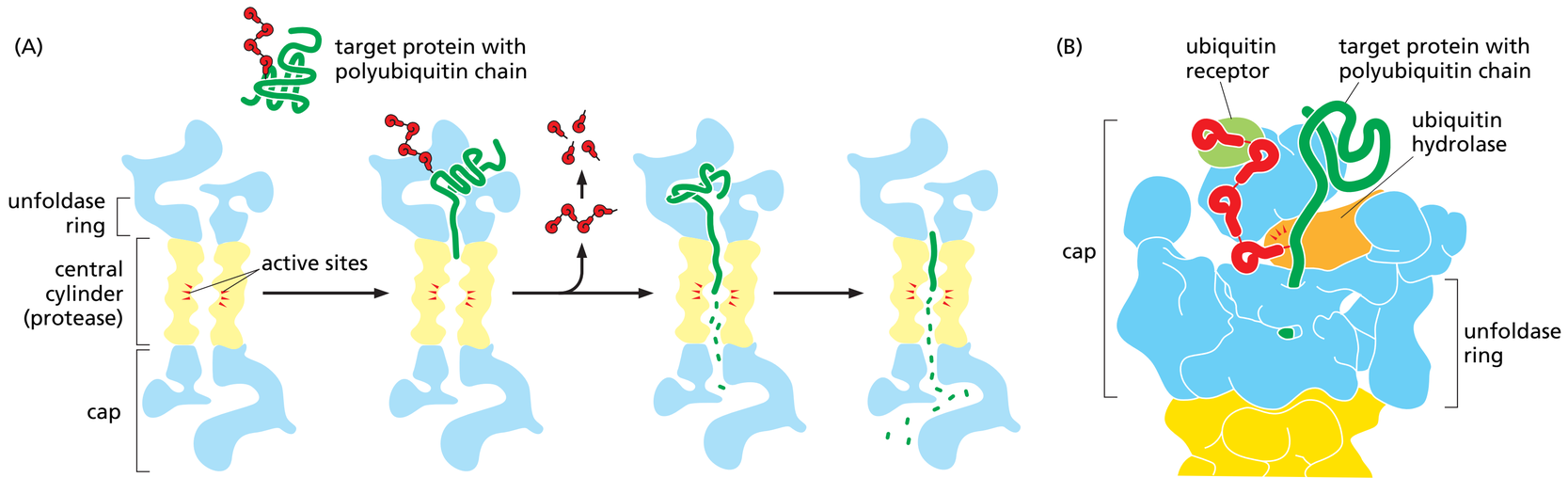
Proteasome (only in eukaryotes)

- Large **protein complex**
- Destroys **aberrant proteins**
- Present in **many copies** in the cytosol and nucleus

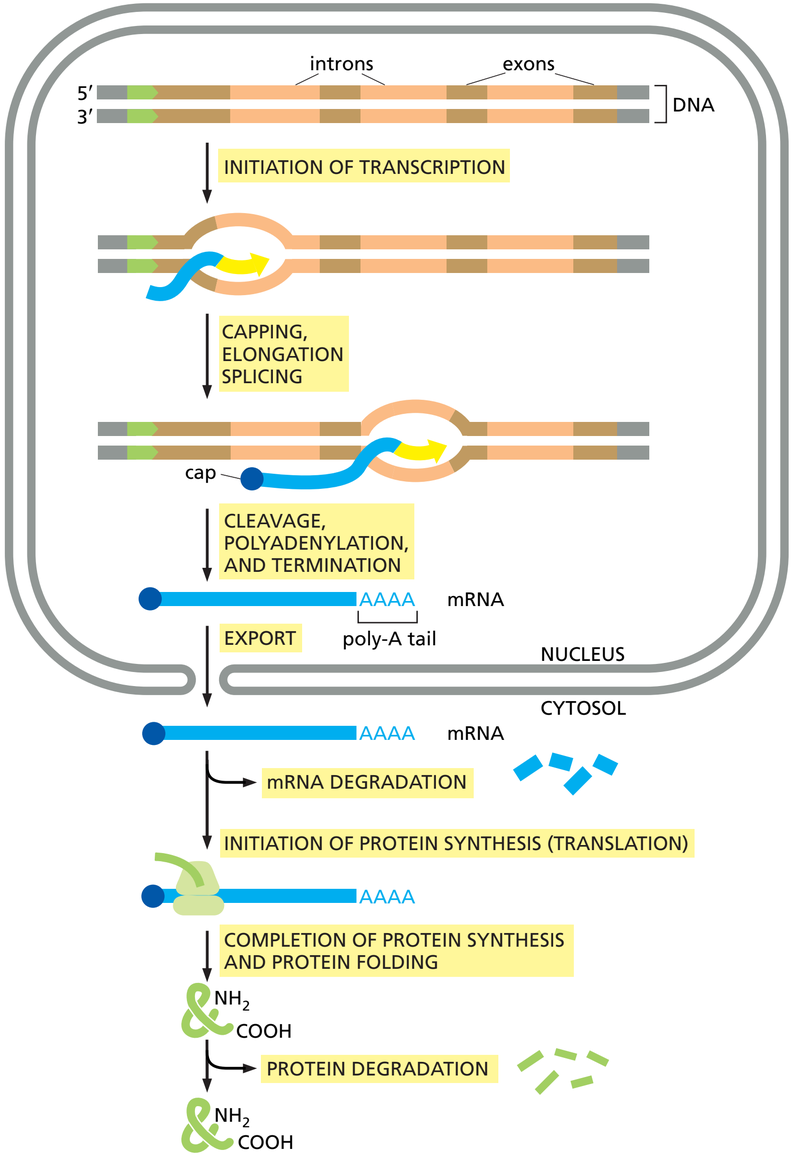


Proteasome (only in eukaryotes)

- Central **cylinder**, with some proteins acting as **proteases**
- Target proteins are **unfolded** as they move through the cap (made of **unfoldases**), exposing them to proteases
- Destruction mark is the covalent attachement of the small protein **ubiquitin**
- Also allows to ensure the **short lifetime** of certain normal proteins (e.g. cell cycle)



Recap: from DNA to proteins



See you next week!