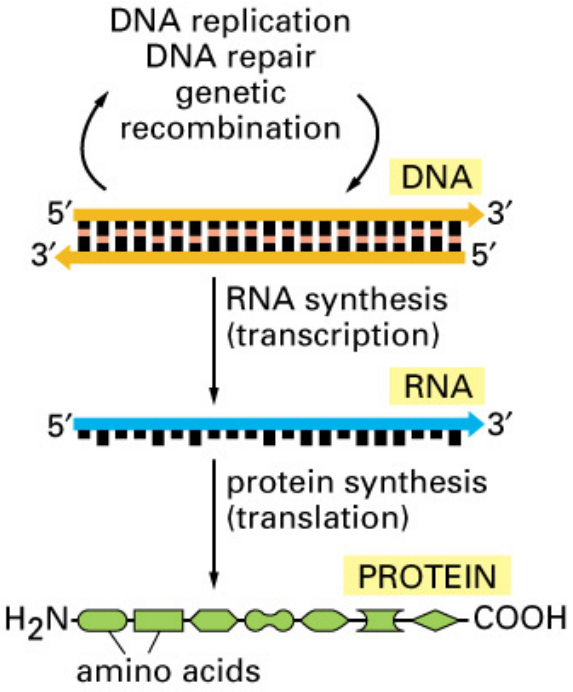
IPAM Cells and Materials:
At the Interface between Mathematics, Biology and Engineering

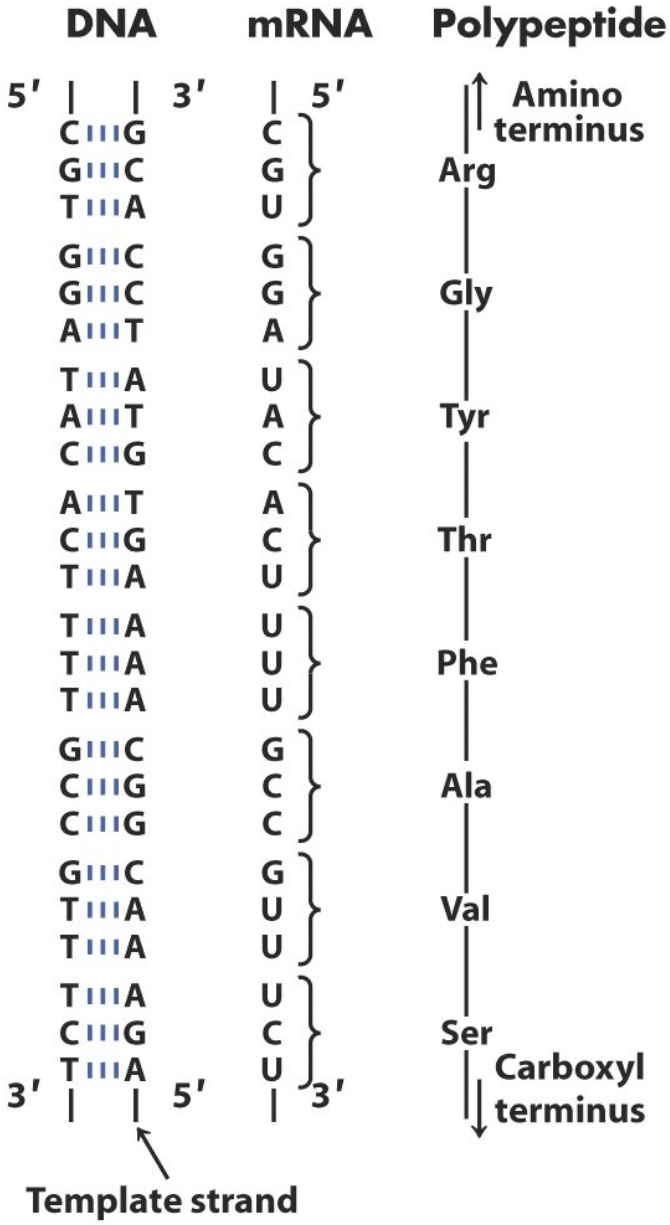
Tutorial 4
Protein Biochemistry 2
Genes to proteins:
Protein synthesis, transport, targeting, and degradation

Dr. Toshikazu Hamasaki

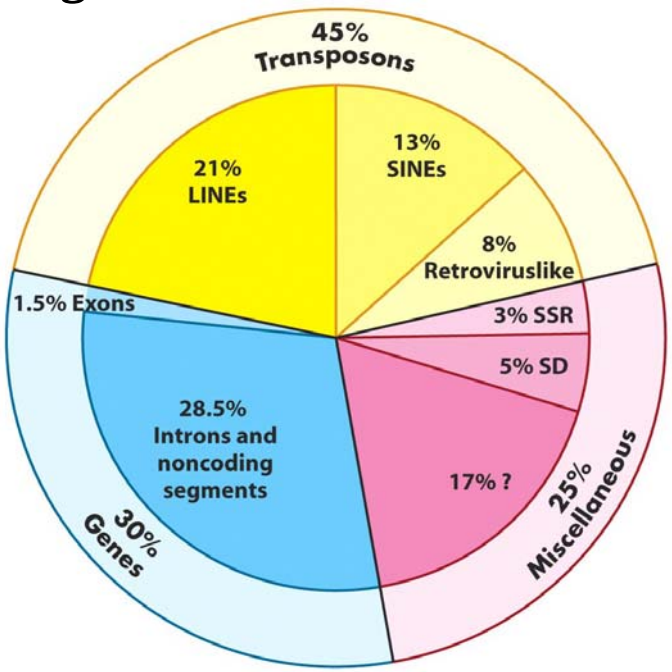
Dept. Bioengineering, UCLA

From DNA to a Protein





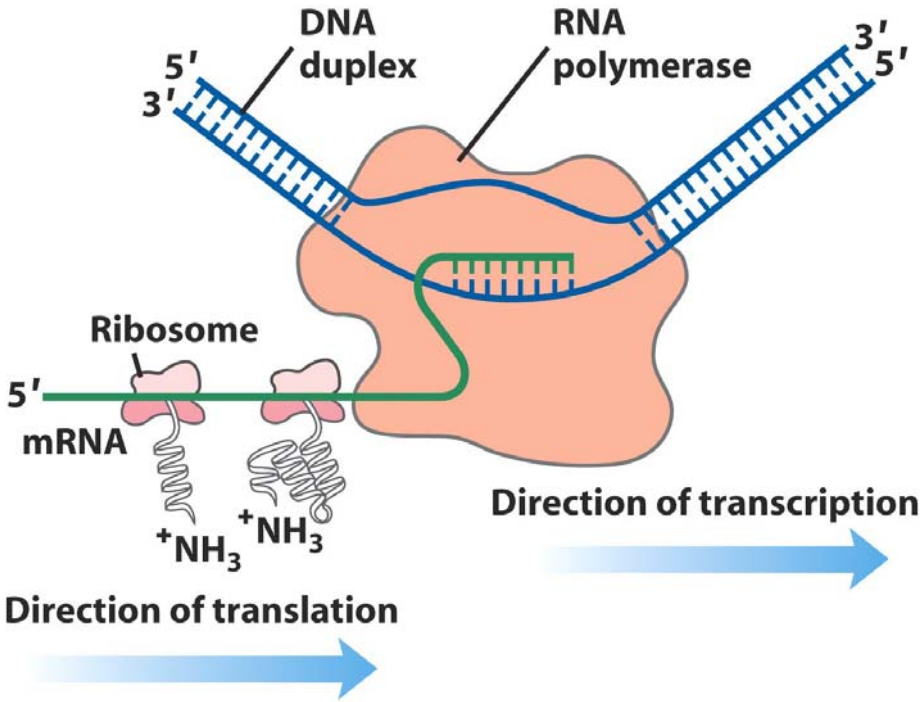
DNA: (genetic codes)
Master Data Storage

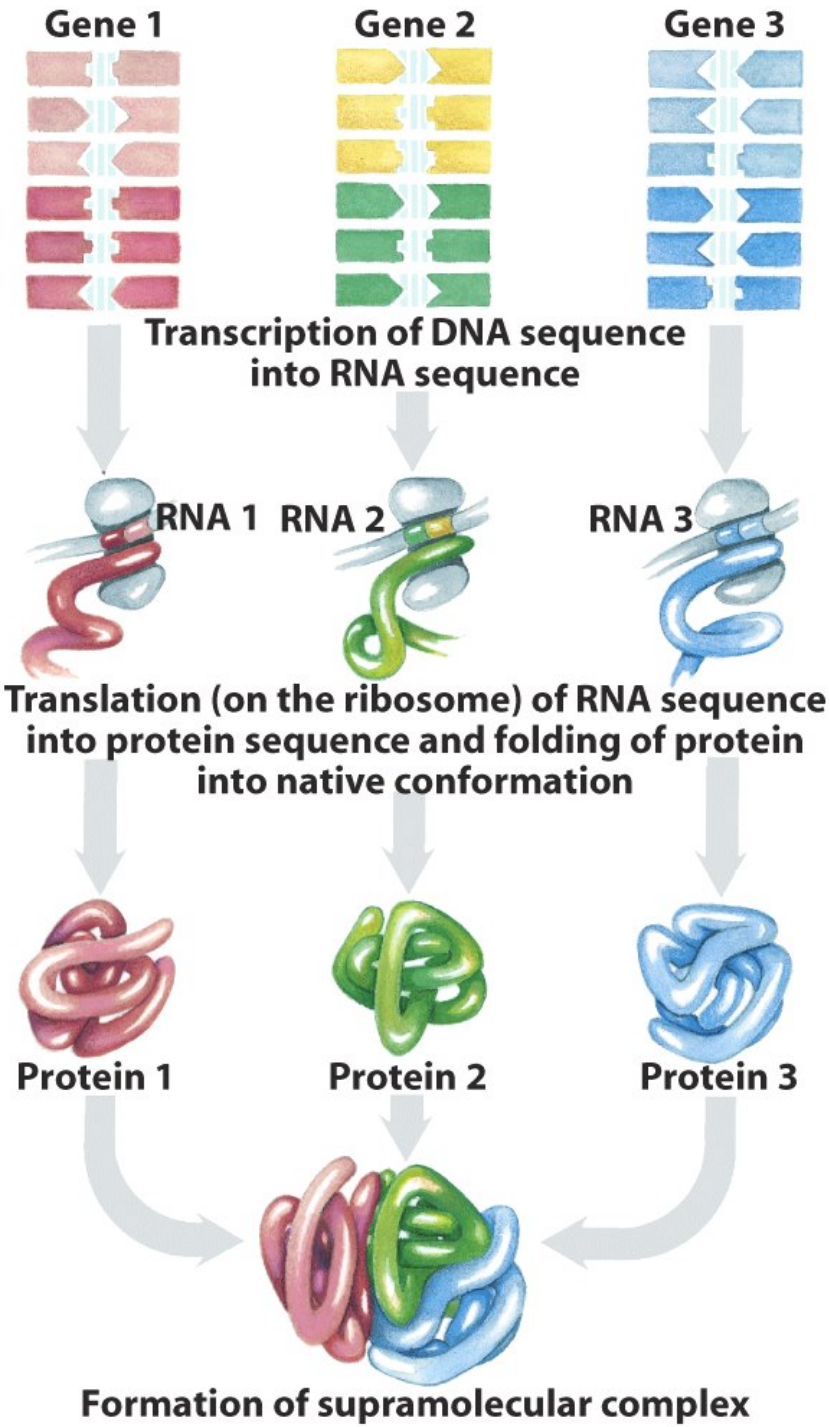


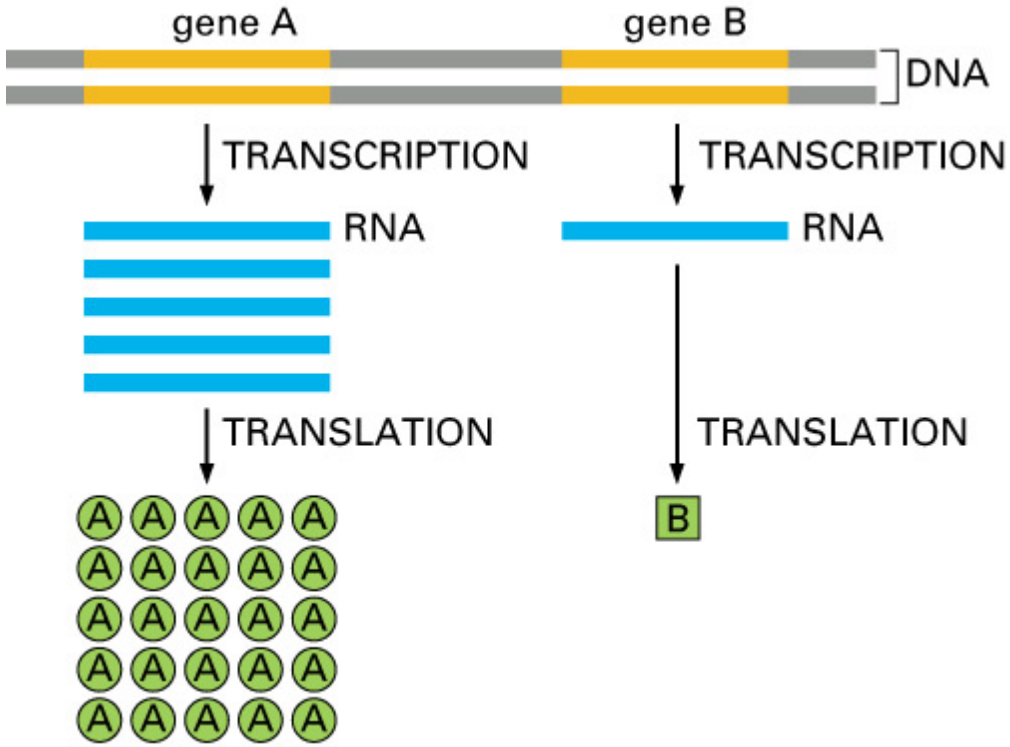
‘coding resion’

‘Direction’ of each fiber is defined.

Each polypeptide chain is a separate gene product.

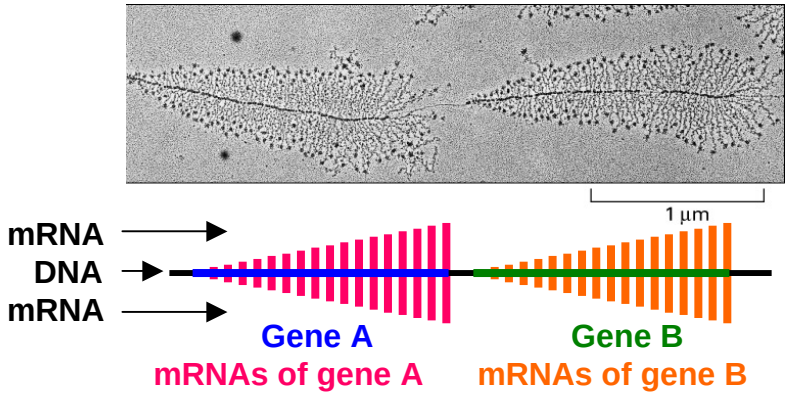






Some proteins are synthesized with amplification; one mRNA can be used to manufacture more than one polypeptide.

Amplification during transcription and/or translation sometimes takes place.



Gene expression and its regulation are not included in this tutorial →
Please take Molecular Biology classes elsewhere

Not all the proteins in a genome are expressed all the time:

Proteins always required for cell to 'LIVE'
'House keeping' proteins
(Cell dies if these proteins are lost)
e.g. **Glycolysis enzymes, Na^+ channel, Na^+/K^+-ATPase**
histone, actin, tubulin, …..
Always produced (genes always expressed) ? feedback reg ?

Proteins produced 'on demand' or ' in response to'.
Hormones (e.g. pituitary hormones), Neurotransmitters
Digestive enzymes

Differentiation, Development (incl. Cell divisions)
Not only more proteins, but also a new set of proteins will be expressed.

Some proteins are made in larger numbers:
Digestive enzymes, Mucins (mucus proteins), Casein (milk protein)
Immunoglobulins, Peptide hormones,
Extracellular matrix proteins (collagen etc)
Keratin (skin, hair, nail)

Proteins in a cell

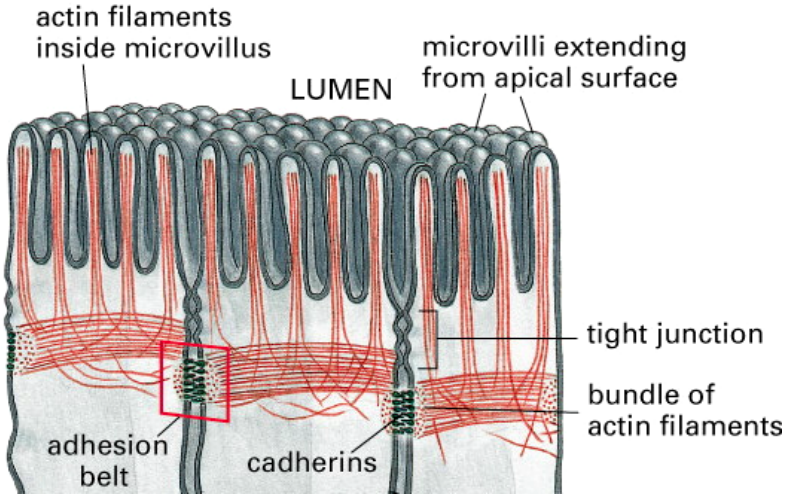
Volume of a prototypical mammalian cell :

$5 \times 5 \times 4\ (\mu m) = 1 \times 10^{-16}\ m^3 = 1 \times 10^{-13}\ l$

Concentration of actin (probably the most abundant protein) in cytoplasm: 100 μM (10^{-4} M)

1 M = 6×10^{23} molecules / *l*

So*, 100 μM = 6×10^{19} / *l



Thus, the amount of actin molecule in the cell would be 10^6
(one actin filament with 1 μm length contains ~450 actins)

The other cytoskeletal protein, tubulin (αβ-tubulin dimer), also the major protein in cells, exists only ~ 10 μM. (10^5 tubulins)
(one microtubule with 1 μm length contains ~220 tubulin dimers)

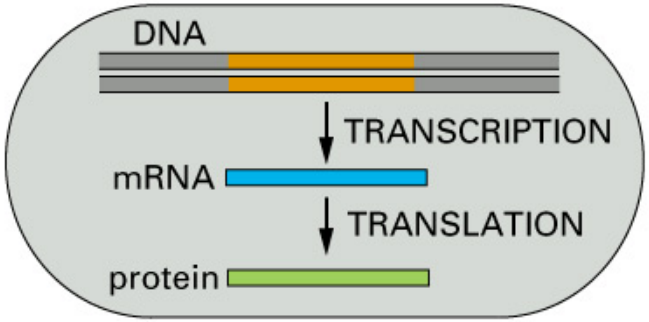
(Normally, cells contain 100s, if not 1000s of these filaments.)

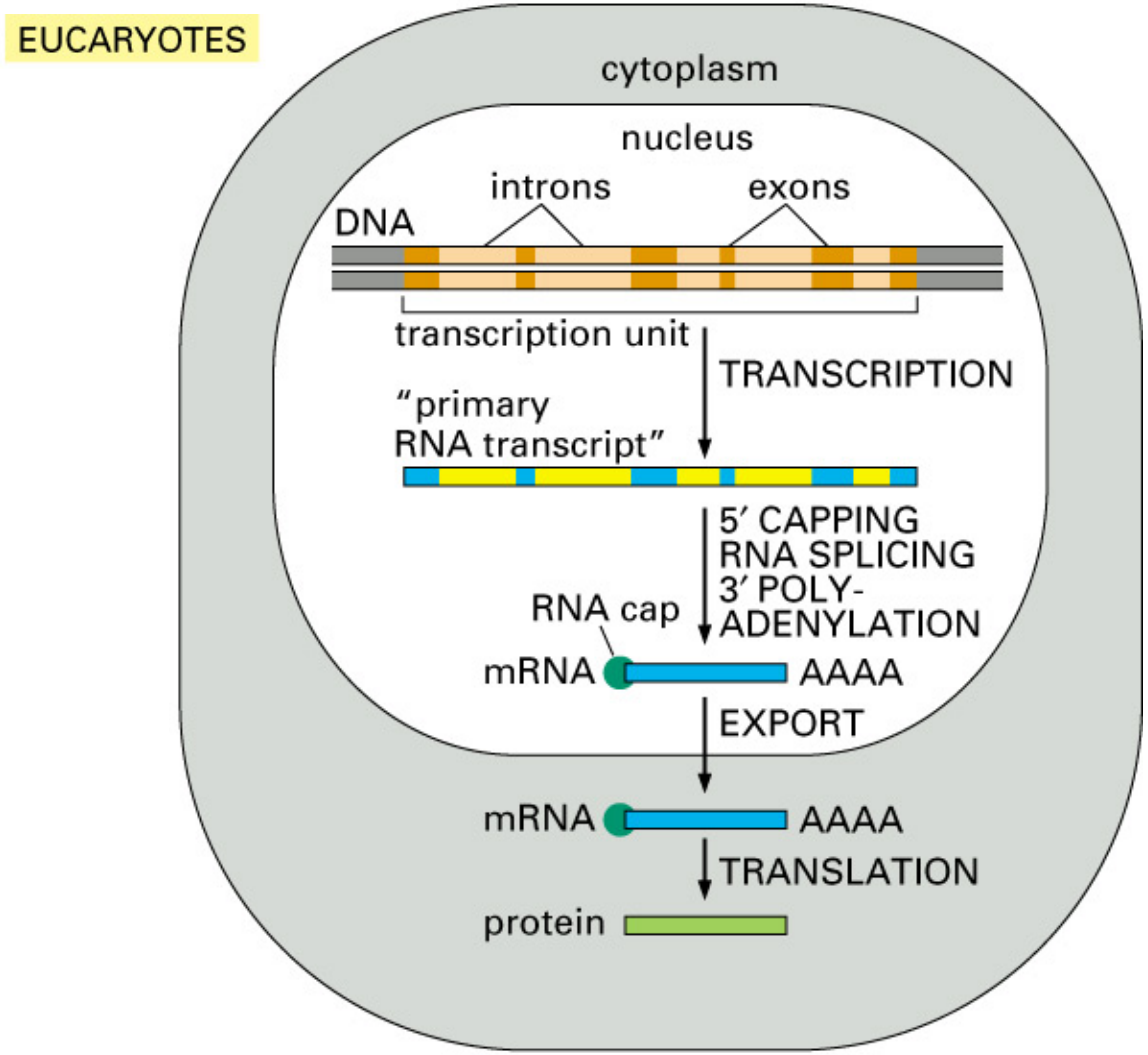
Cells have most of other cytoplasmic proteins in concentration of nM (10^{-9} M) or so (1 nM in the cell would be 60 molecules).

Very Brief Information regarding Transcription

Eukaryoric genes ('transcription unit') include INTRONS. Introns are the nuclear sequences that are usually not used to construct mRNAs. Splicing of the primary RNA transcript (which includes portion from the introns) will yield mature mRNA.

Bacterial (prolaryotic) genes don't have introns, thus transcription and translation are straight forward.





Gene sequence ≠ peptide sequence in a protein

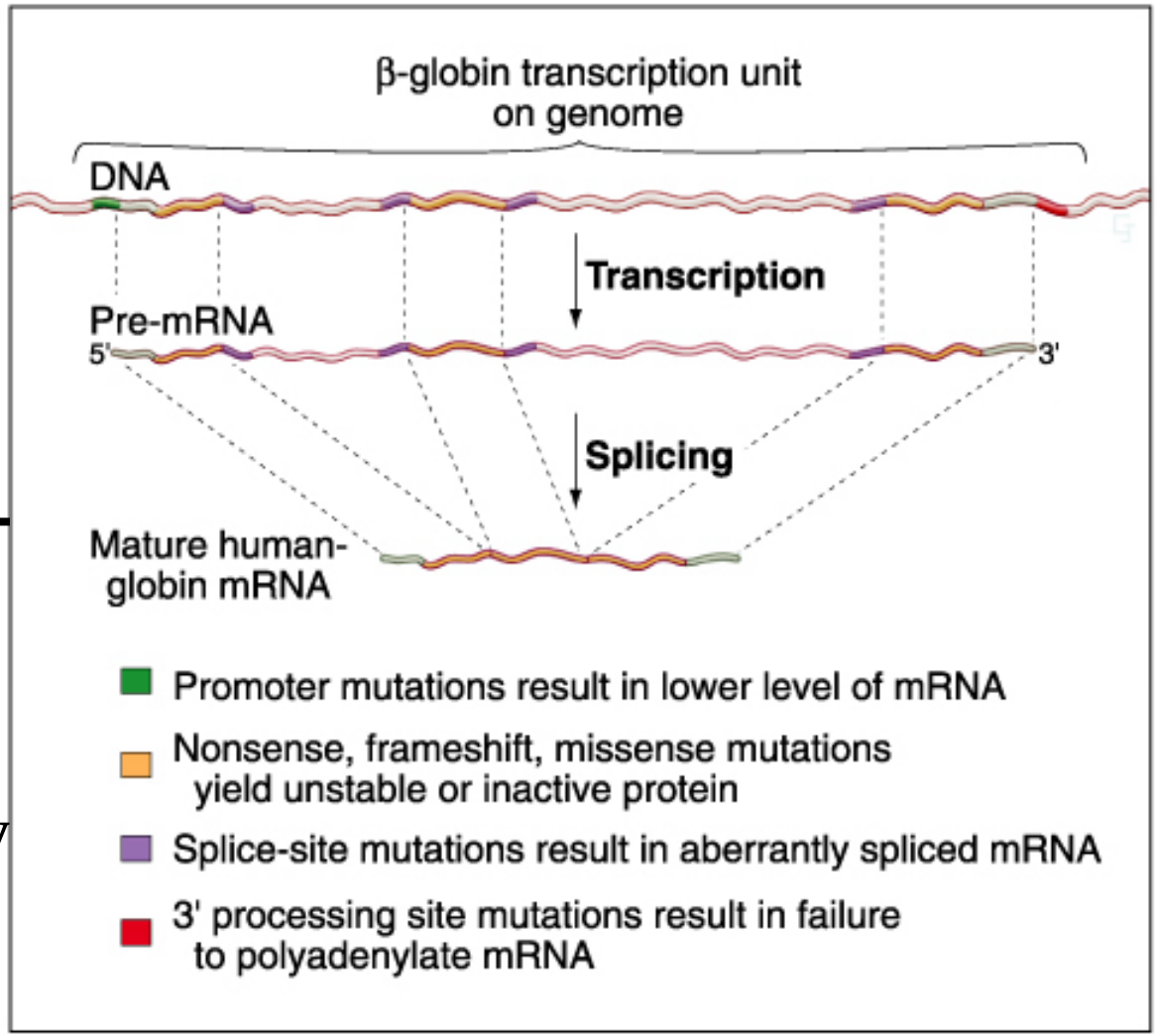
(Nucleotide sequence in a chromosome: GENOMIC SEQUENCE)

β-globin gene ←

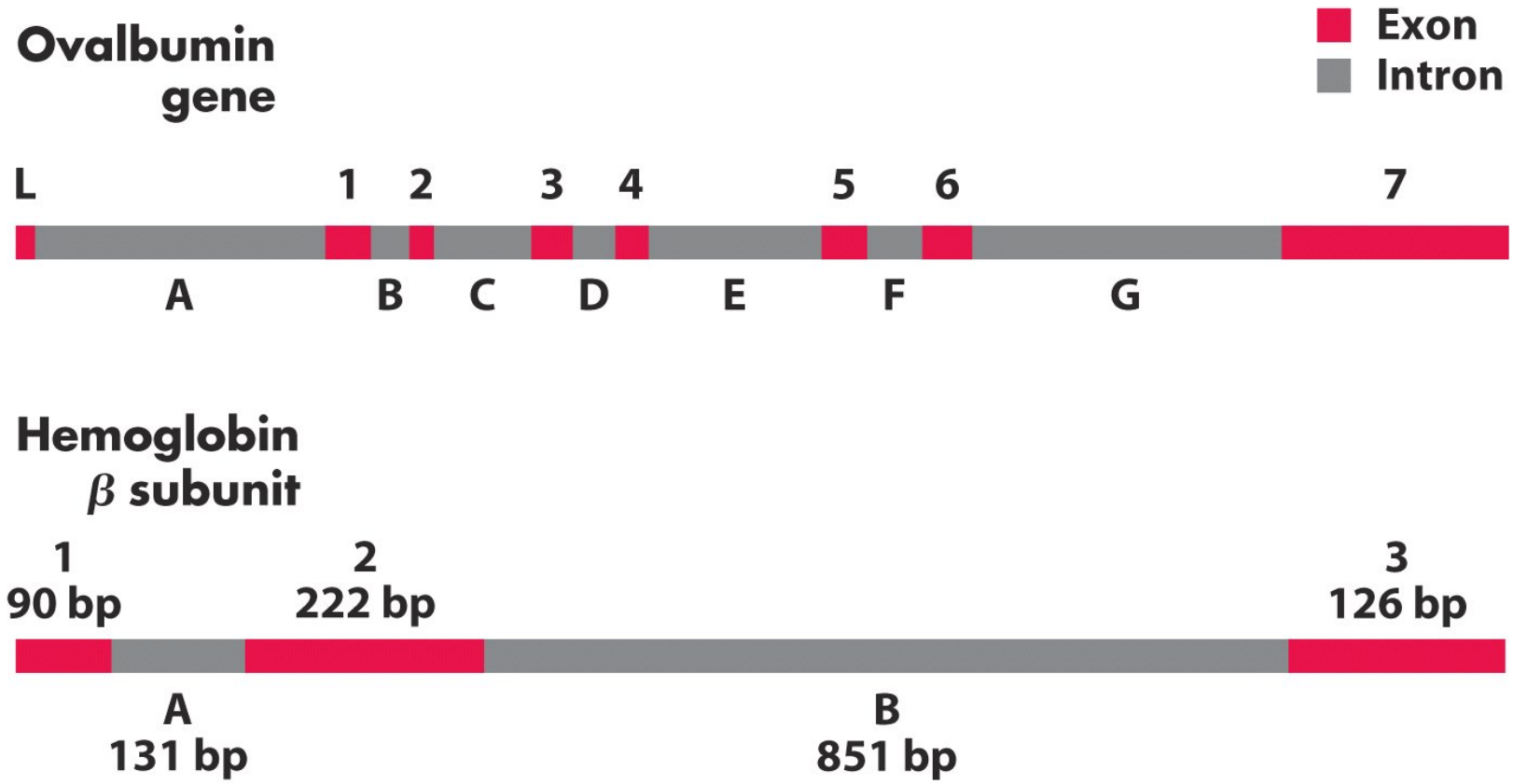
Messenger RNA (mRNA) is the direct template from which β-globin polypeptide chain (protein) is synthesized (Translation) ←

↓

mRNA + Substrates + enzyme → complementary DNA (cDNA) (Reverse Transcription)

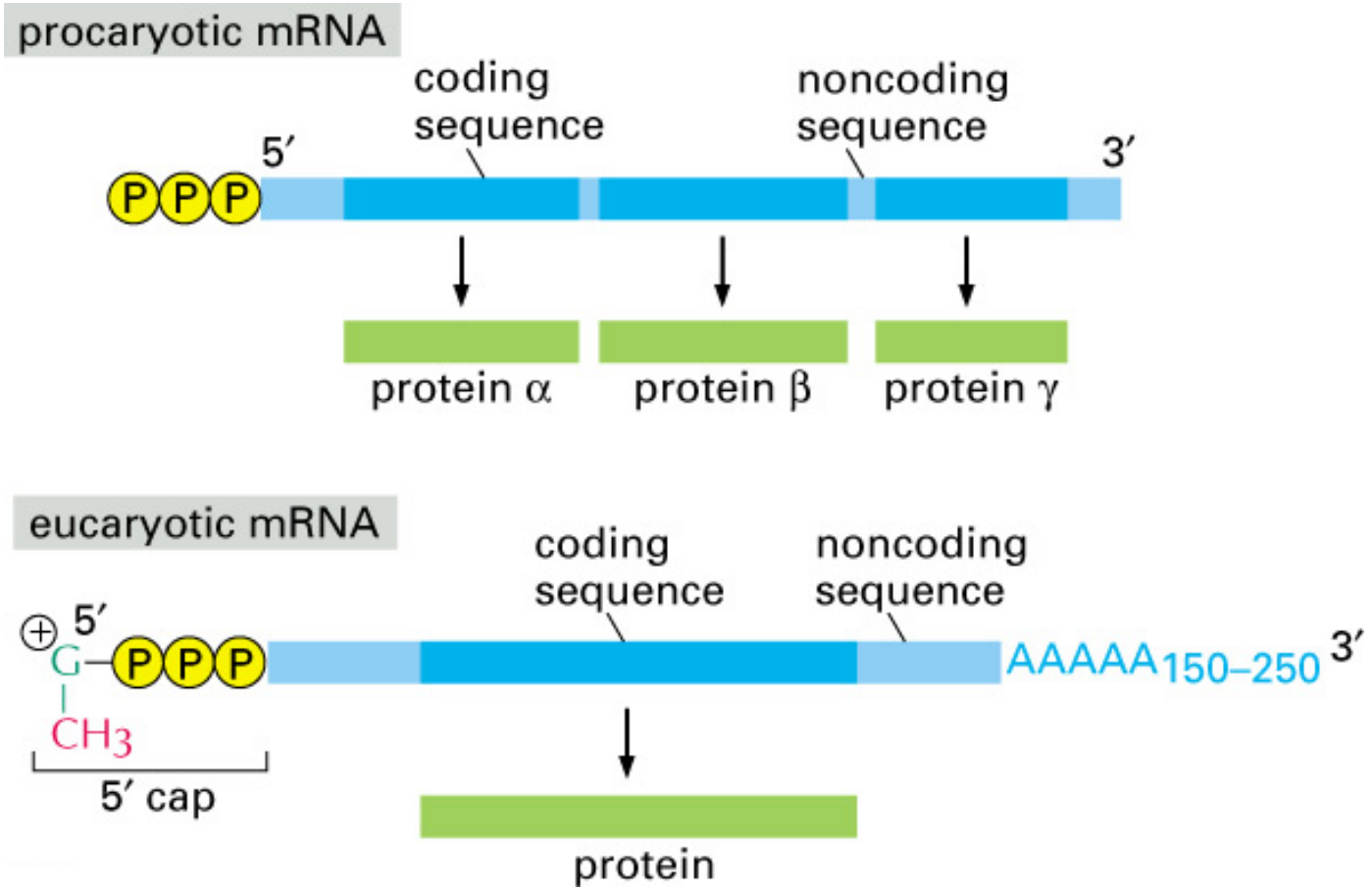


Many genes have larger introns than exons



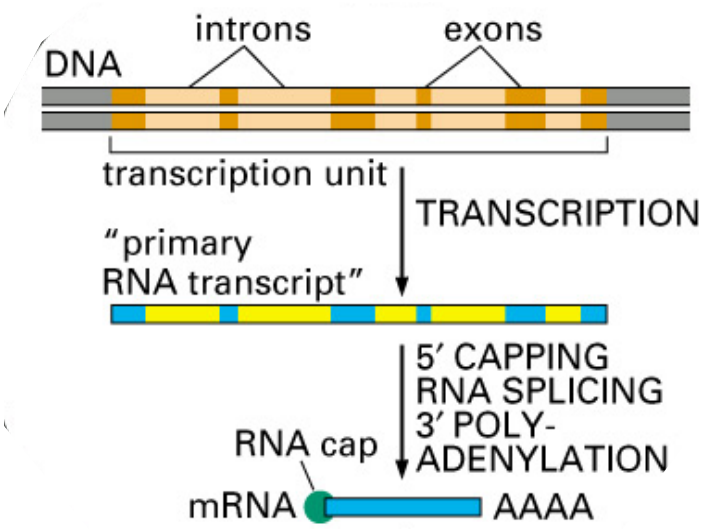
Messenger RNA structure

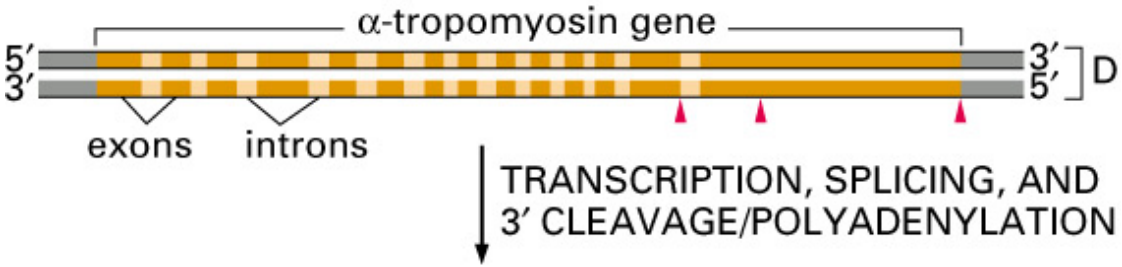
Eukaryotic mRNA has additional 5' CAP and 3' poly-A tail

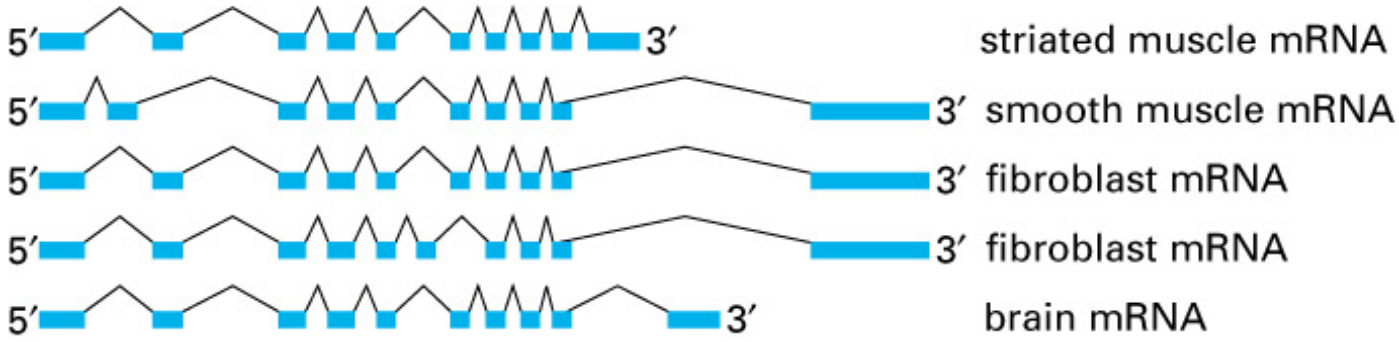


Bacterial mRNA could have one than one coding sequences (for different proteins), whereas each eukaryotic mRNA (almost certainly) codes only one protein.

One gene could code more than one protein (varieties) : ALTERNATIVE SPRICING







mRNA splicing mechanism makes multiple versions (varieties) of mRNA from certain gene (primary RNA transcript). These different mRNAs are translated into different proteins.

Codon; translation from genetic messages (DNA, mRNA) to peptide sequence

Translation from series of 3-letter-word of four-letter language (mRNA) into twenty-letter language (peptide)

First letter of codon (5′ end)

Second letter of codon

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

AUG (Met): Initiation codon
UAA, UAG, UGA: Stop codon

TABLE 27–3 Degeneracy of the Genetic Code

Amino acid	Number of codons	Amino acid	Number of codons
Met	1	Tyr	2
Trp	1	Ile	3
Asn	2	Ala	4
Asp	2	Gly	4
Cys	2	Pro	4
Gln	2	Thr	4
Glu	2	Val	4
His	2	Arg	6
Lys	2	Leu	6
Phe	2	Ser	6

Some amino acids are coded w/ many codons, others a 1~2

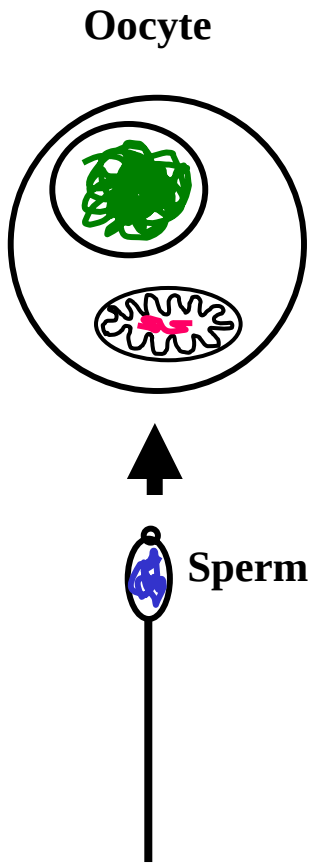
Different DNA nucleotide sequences (Genomic Sequences) could code same protein (amino acid sequence)

Note: Mitochondrion codons are different from main genome

TABLE 1 Known Variant Codon Assignments in Mitochondria

	Codons*				
	UGA	*AUA*	*AGA AGG*	*CUN*	*CGG*
Normal code assignment	Stop	Ile	Arg	Leu	Arg
Animals					
Vertebrates	Trp	Met	Stop	+	+
Drosophila	Trp	Met	Ser	+	+
Yeasts					
Saccharomyces cerevisiae	Trp	Met	+	Thr	+
Torulopsis glabrata	Trp	Met	+	Thr	?
Schizosaccharomyces pombe	Trp	+	+	+	+
Filamentous fungi	Trp	+	+	+	+
Trypanosomes	Trp	+	+	+	+
Higher plants	+	+	+	+	Trp
Chlamydomonas reinhardtii	?	+	+	+	?

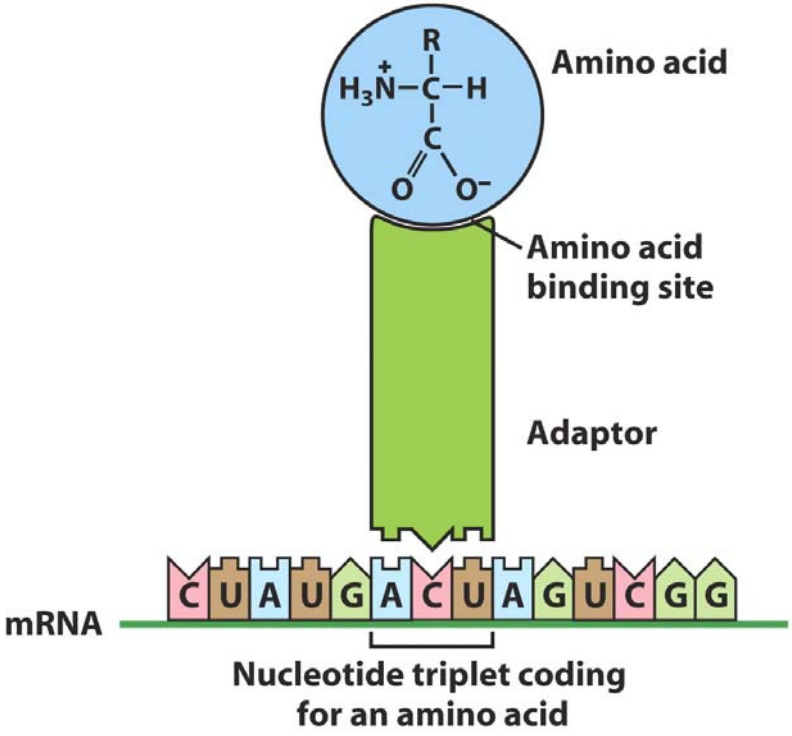
*N indicates any nucleotide; +, codon has the same meaning as in the normal code; ?, codon not observed in this mitochondrial genome.

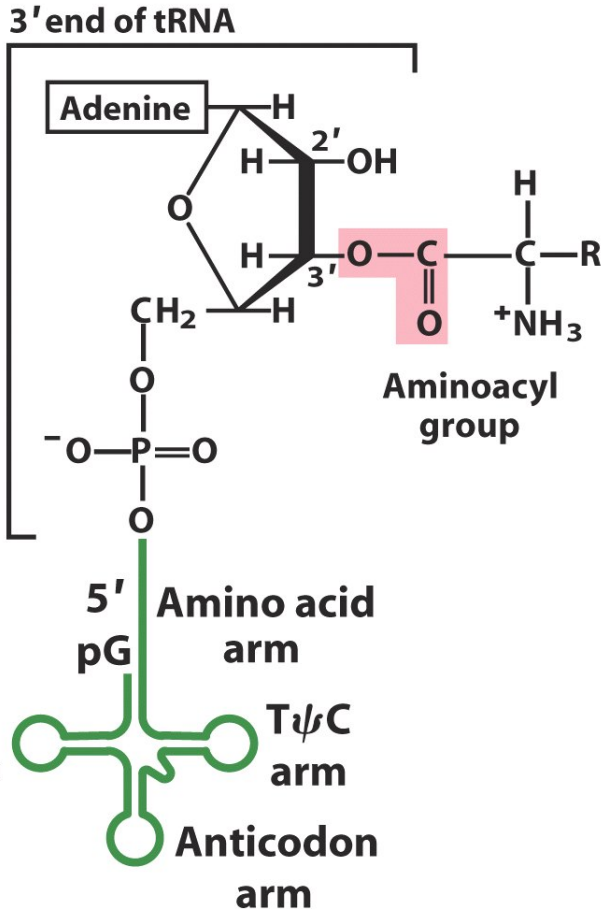


Mitochondrion has their own genes
(genes for some of their own proteins <not all of them>)

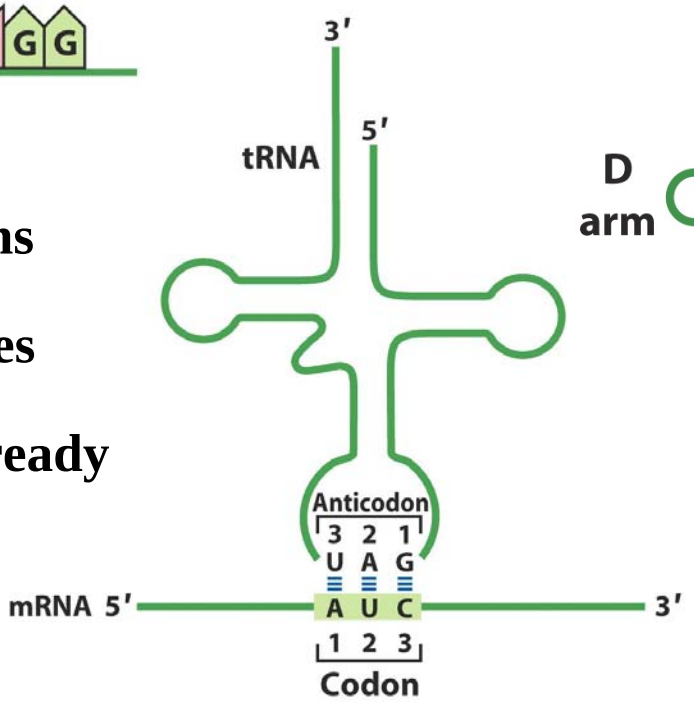
Mitochondrion genes are strictly “Maternal”
So, your genes are come from your Mom’s nuclear genes, your Mom’s mitochondrial genes and your Dad’s nuclear genes.

Aminoacyl transfer RNA (tRNA) : The translators
It has codon message and corresponding amino acid

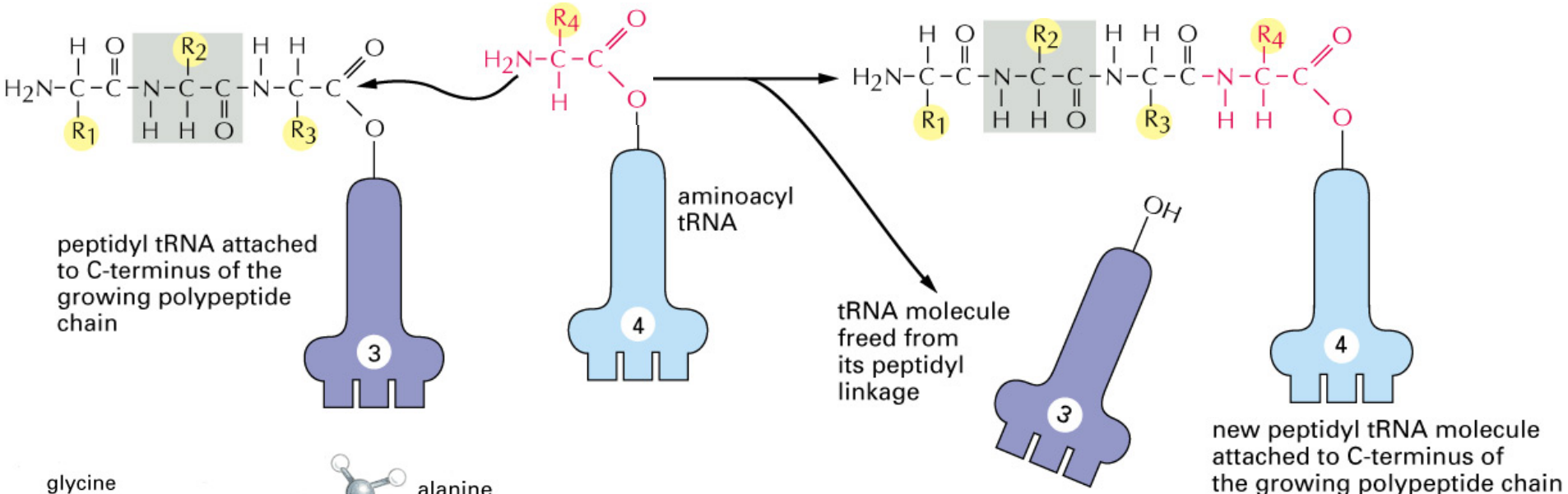


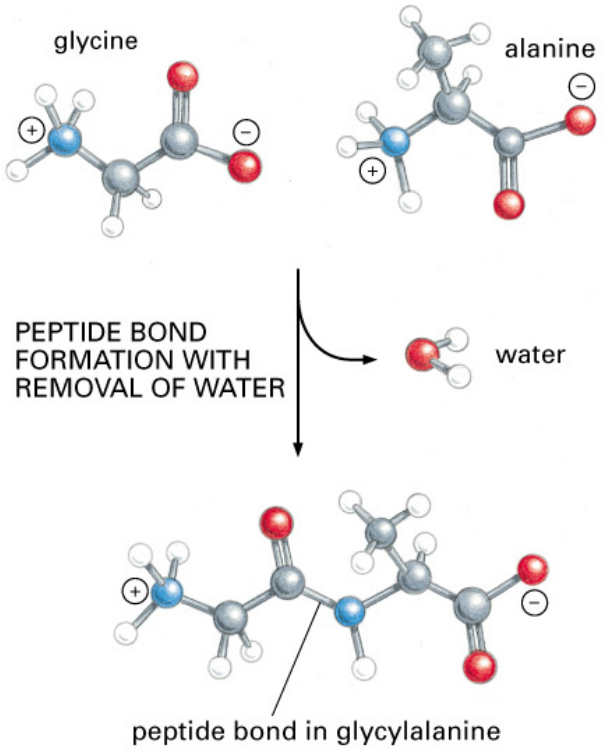


There 61 codon combinations for 20 amino acids ➔
There are 61 different species amino-acyl transfer RNAs.
You need to prepare all 61 ready for protein synthesis.
(AND; enough of each!!)

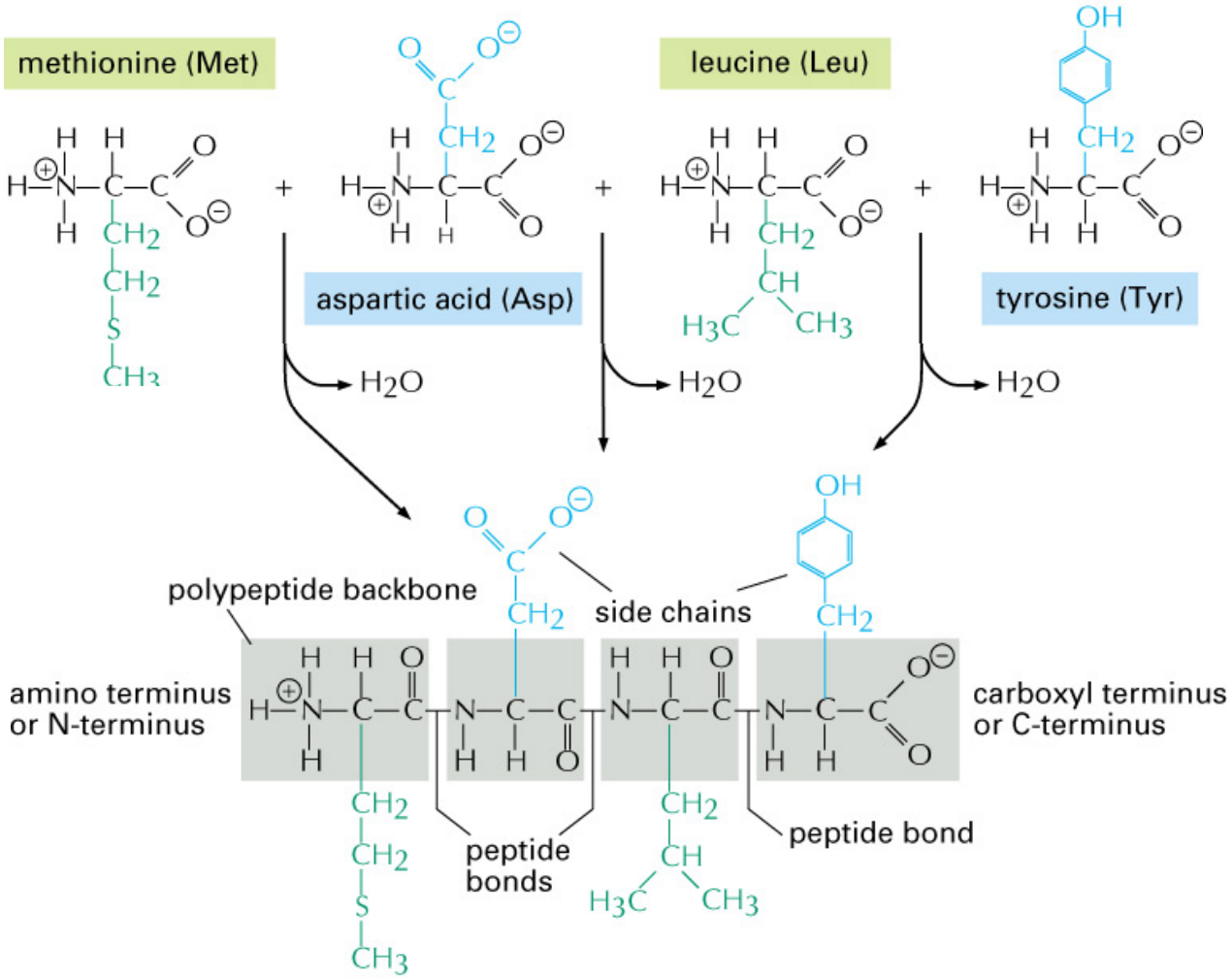


Protein synthesis process (at the first major step) is elongation of polypeptide chain at Carboxi-terminus of the elongating chain .

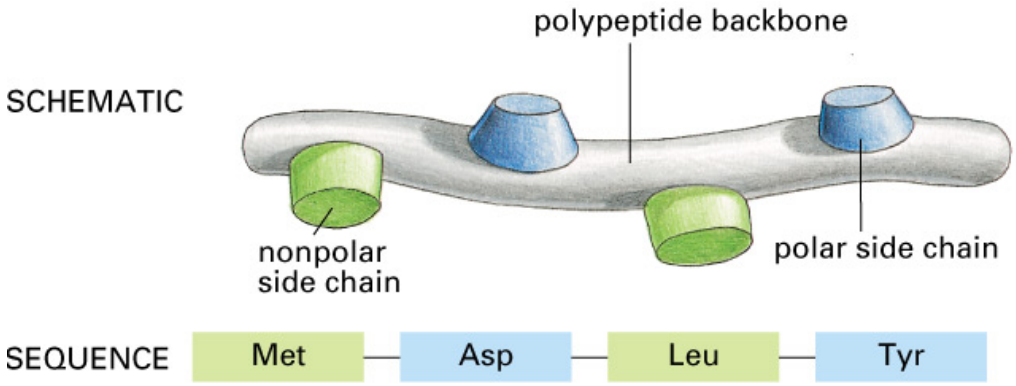




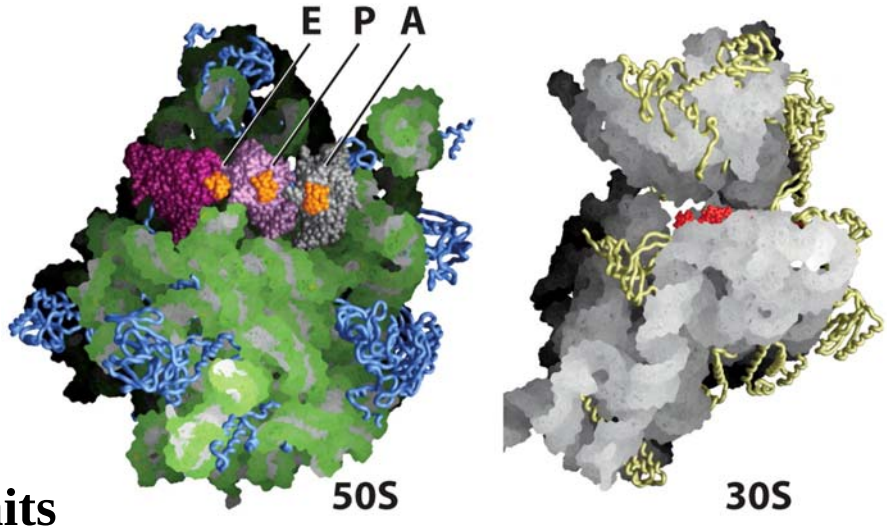
Each <u>specific</u> amino acid (peptide) is added at C-terminus, and condensation (removal of water molecule) takes place to for 'peptide bond'.



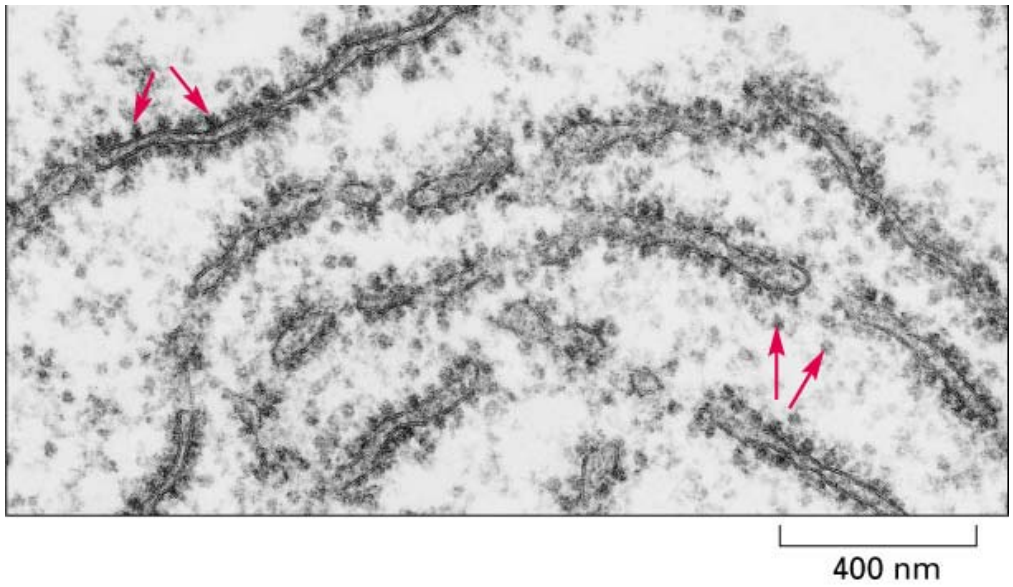
Each protein (polypeptide chain) is synthesized from its amino end (N-end) to C-end, (and the amino acid chain is 'primary' protein structure) according to the code of mRNA (from 5' end to 3' end), and this is done with ribosomes.

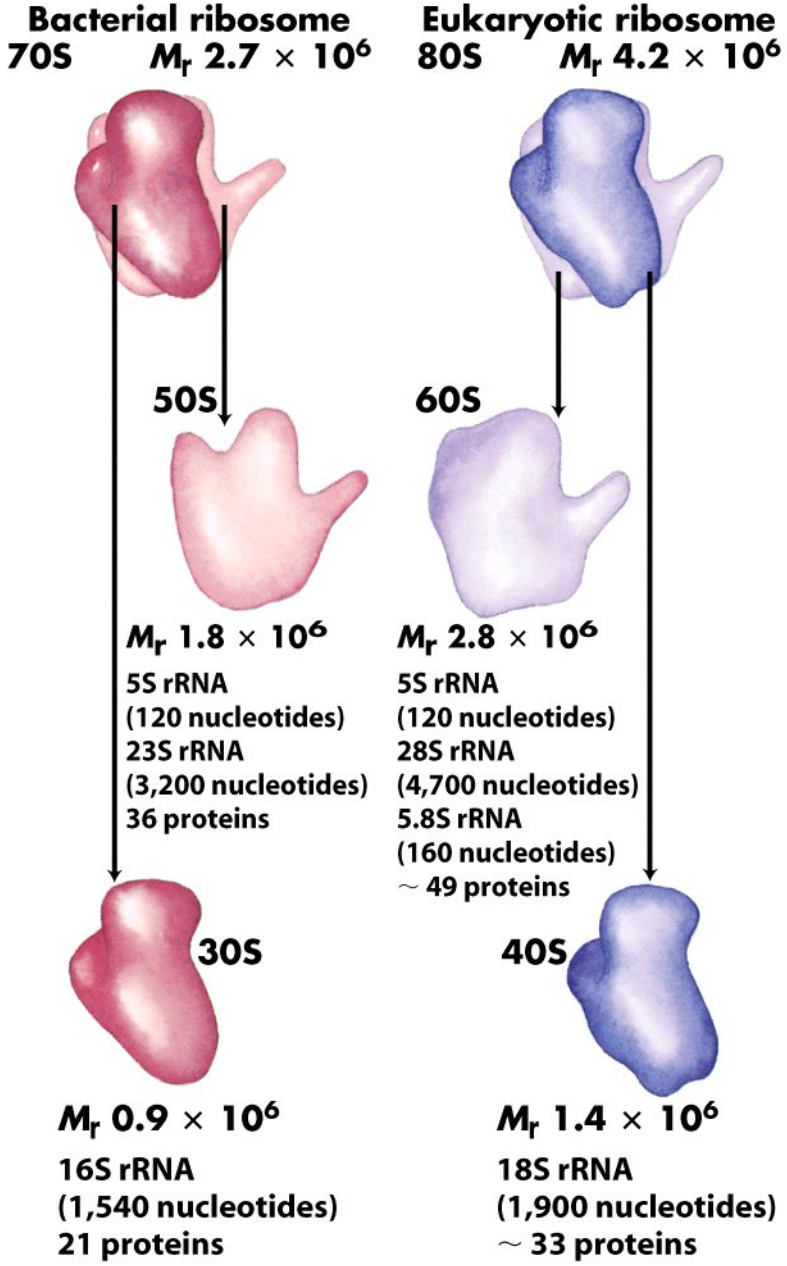


Ribosome (Polyribosome; Polysome)
Polypeptide synthesis factory



2 subunits
each made from many proteins and ribosomal RNA (rRNA)





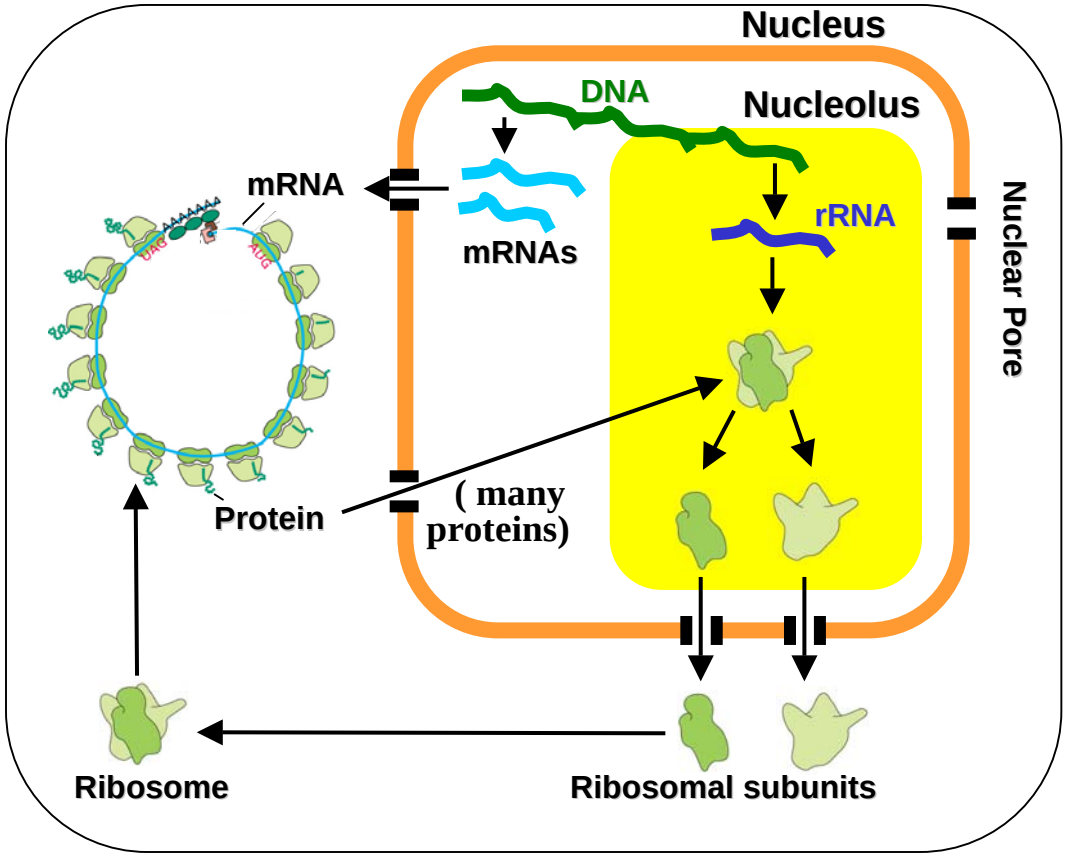
Ribosome Assembly

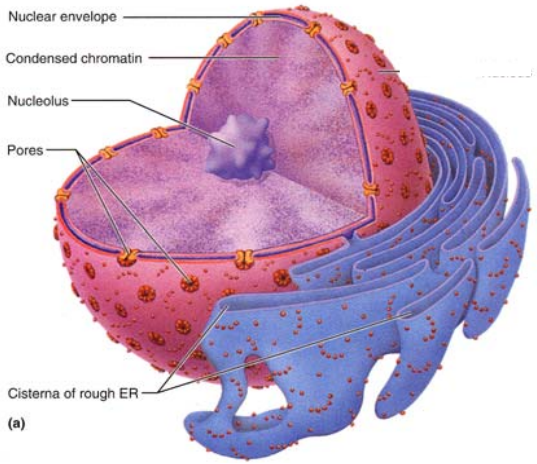
In eukaryotic cells, ribosome subunits are assembled in nucleus, more specifically, in NUCLEOLUS.

Ribosomal proteins are synthesized in cytoplasm, and shipped into nucleolus.

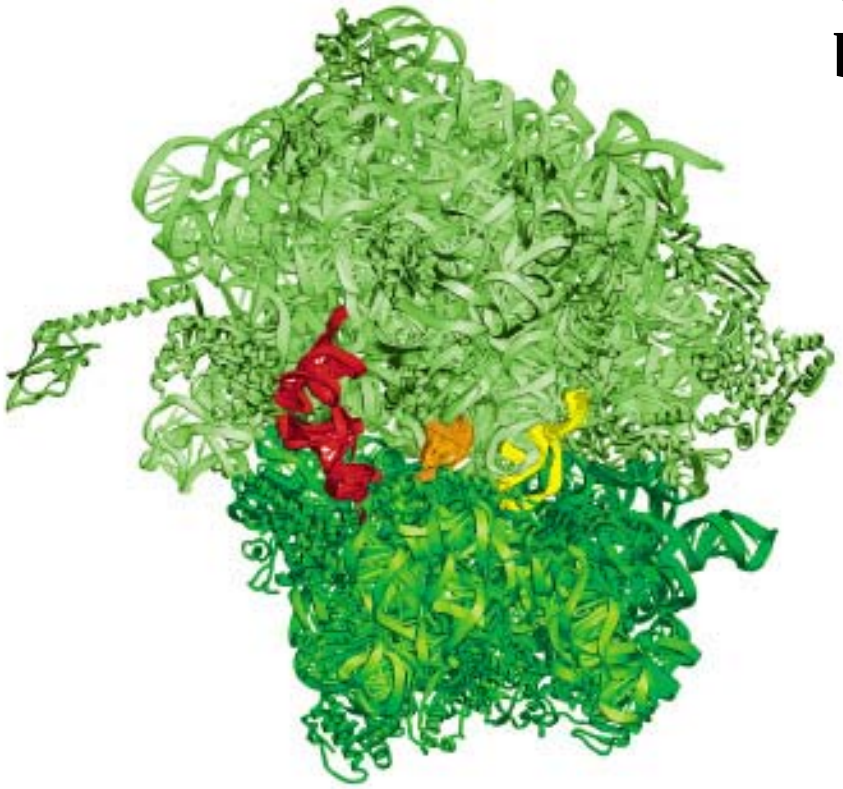
Ribosomal RNAs (rRNAs) are transcribed at site.

Assembled each subunit then will be exported from nucleus into cytoplasm, where they function.

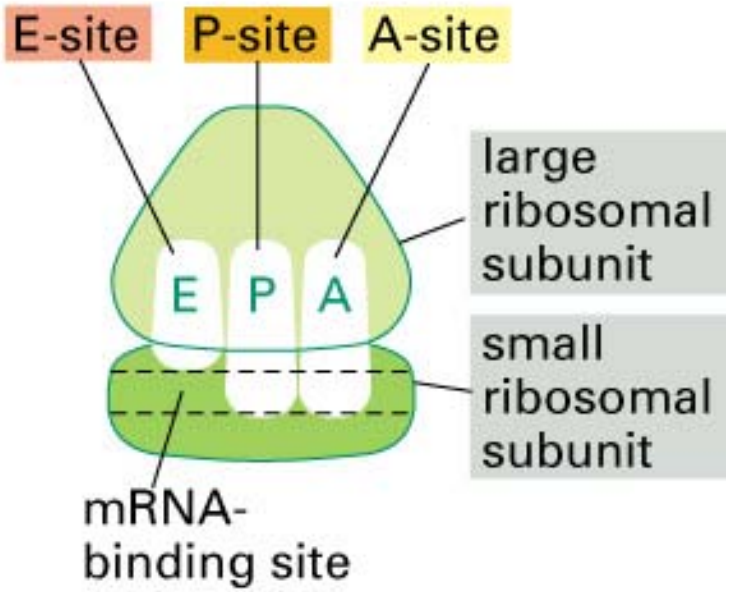




Functionally Important features of Ribosome



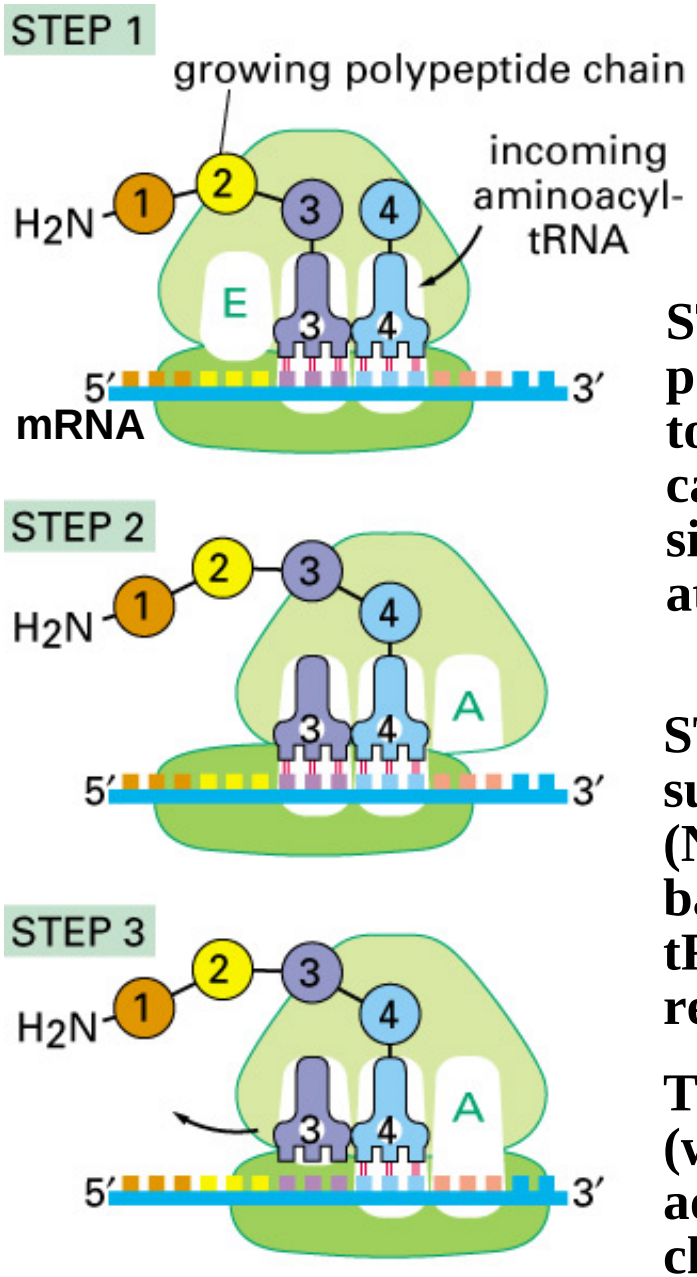
Three Aminoacyl-tRNA binding sites



Translation (or, polypeptide elongation) takes place at ribosome.
Three major distinctive steps of translation:

1. **Initiation**
2. **Elongation**
3. **Termination**

2. Polypeptide (PP) elongation

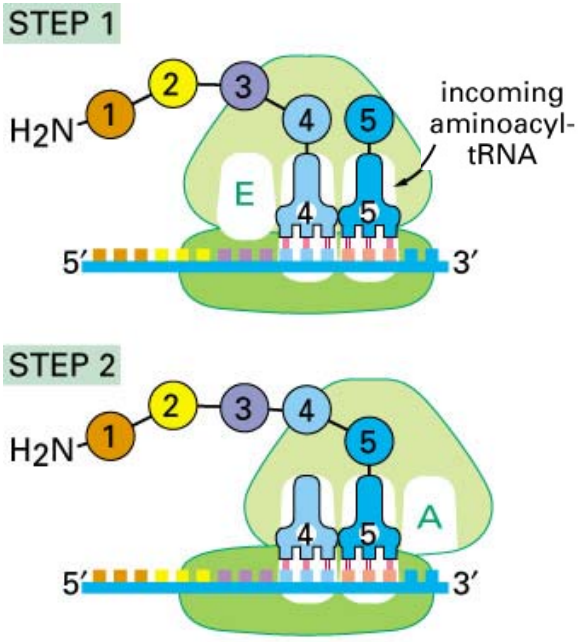


STEP 1: Elongating PP with AA-tRNA attached at its C-end is on ‘P’ site on ribosome. ‘E’ site is free. An another AA-tRNA with matched mRNA codon comes into ‘A’ site.

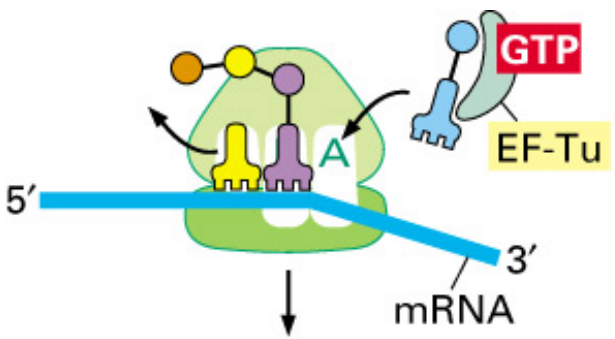
STEP 2: Large ribosome subunit shifts its position, and condensation reaction takes place to attach polypeptide chain into a new peptide came with the AA-tRNA. (Now ‘E” site and ‘P’ site are occupied.) Linkage between previously attached tRNA (3) is broken.

STEP 3: The Large subunit shifts back. (Now ‘A’ site is aligned back with mRNA.) tRNA at ‘E’ site is released from complex.

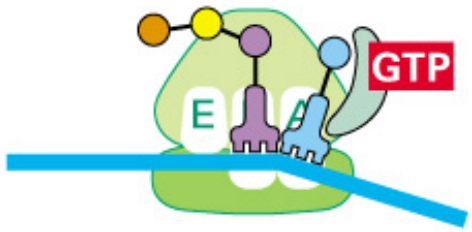
The process repeats (with each amino acid added to the polypeptide chain.)

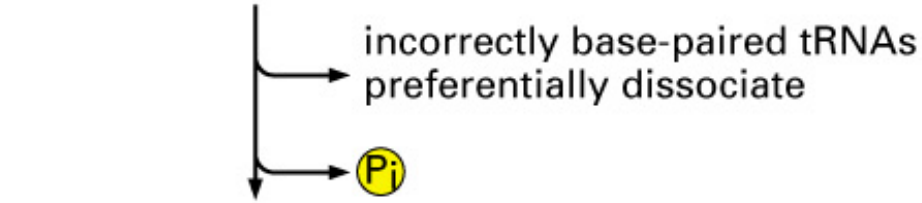


2 GTP hydrolysis requires with one peptide addition (The position shift of ribosomal subunits is called Translocation)

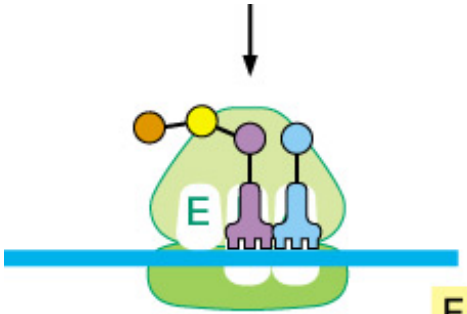


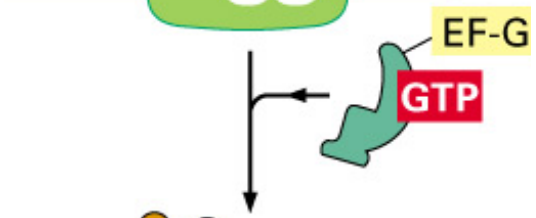
Pr: EF-Tu
Eu: EF-1



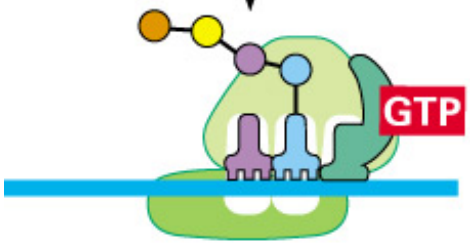


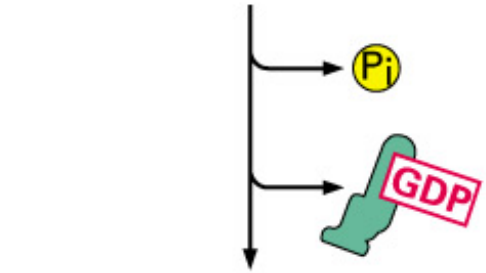
Trials and errors





Pr: EF-G
Eu: EF-2
(A.K.A. Translocase)





Pr: Prokaryotes
Eu: Eukaryotes

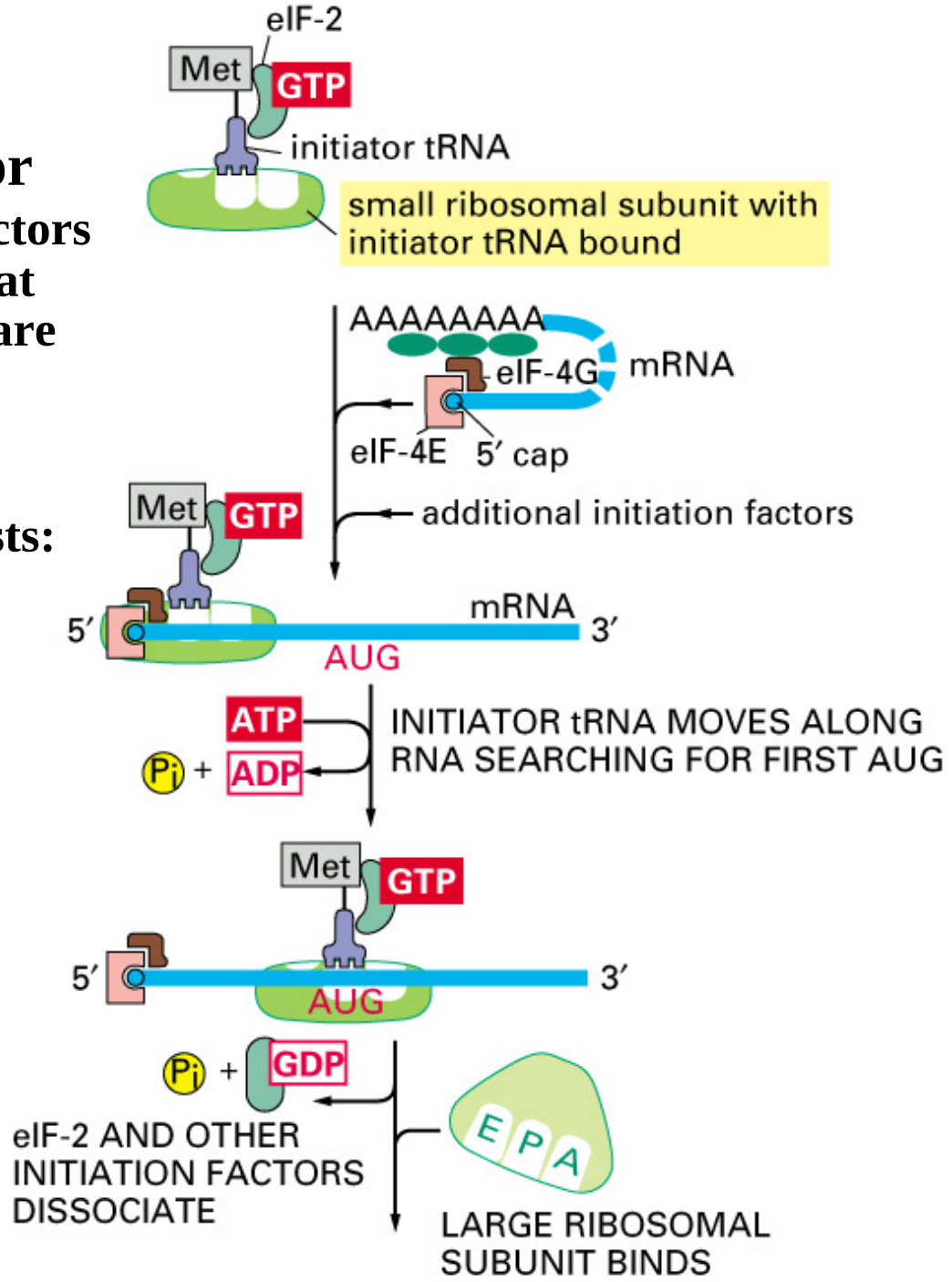
1. Initiation

Initiation begins with binding of special AA-tRNA^{MET} (Initiator tRNA), together with initiation factors to smaller ribosome subunit. (Note at this time two subunits of ribosome are dissociated.)

Initiator tRNA :

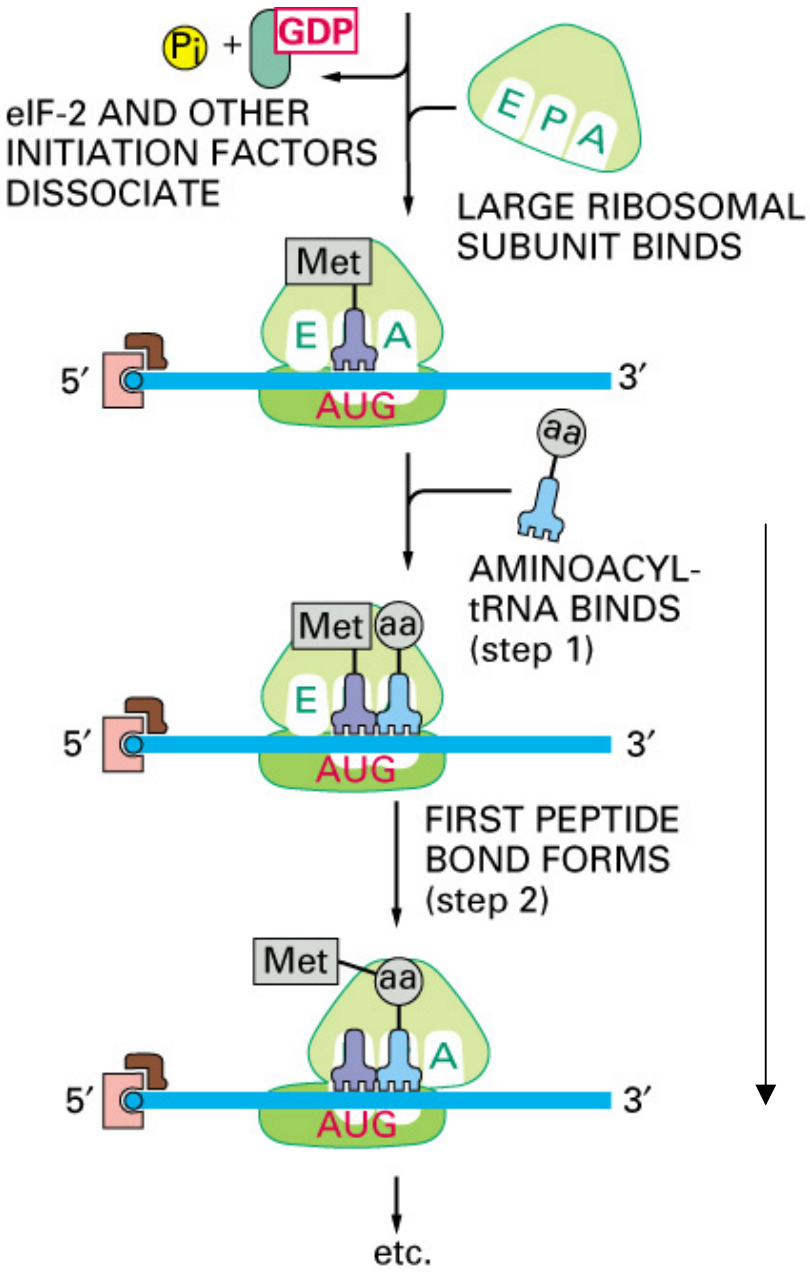
Bacteria, Mitochondria, Chloroplasts:
N-formylmethionyl-tRNA^{fMet} (fMet-tRMA^{fMet})

Eukaryotes:
Special Met-tRNA^{Met}



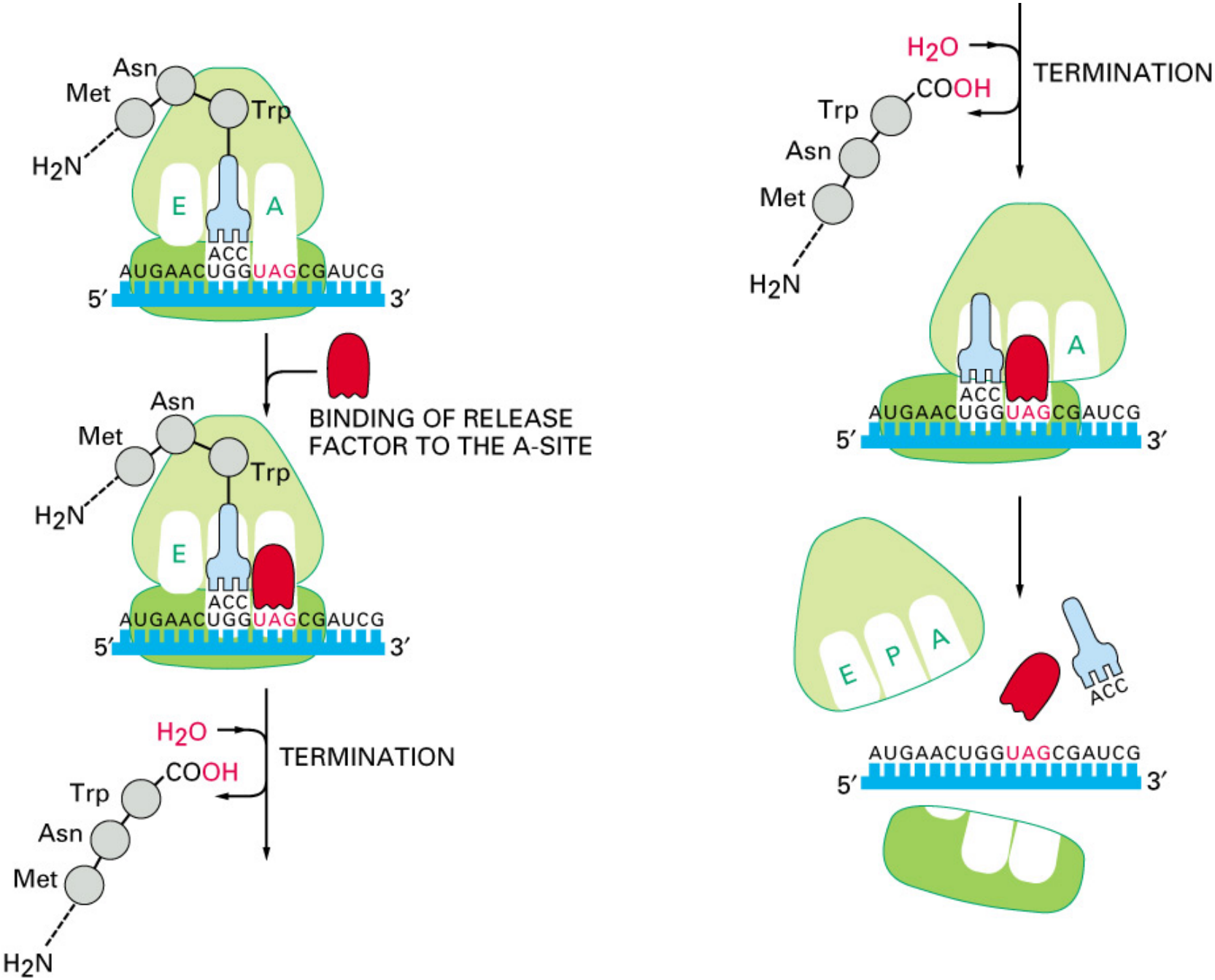
So, The first amino acid in given protein is ALWAYS Methionine !(?)
NOT Necessarily TRUE*!*
Many proteins (especially, secretory proteins) undergo maturation, where Met may be removed.

1. Initiation (2)

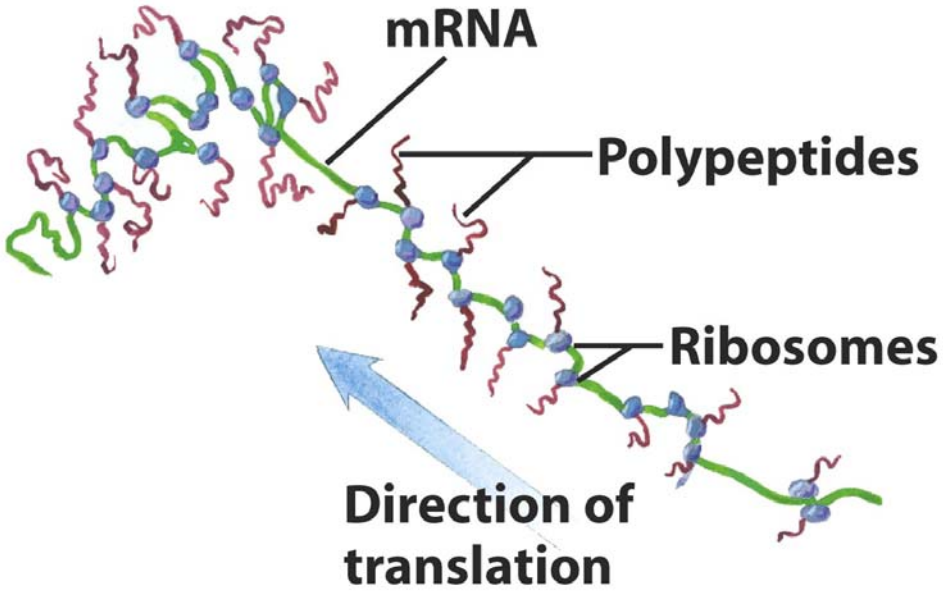


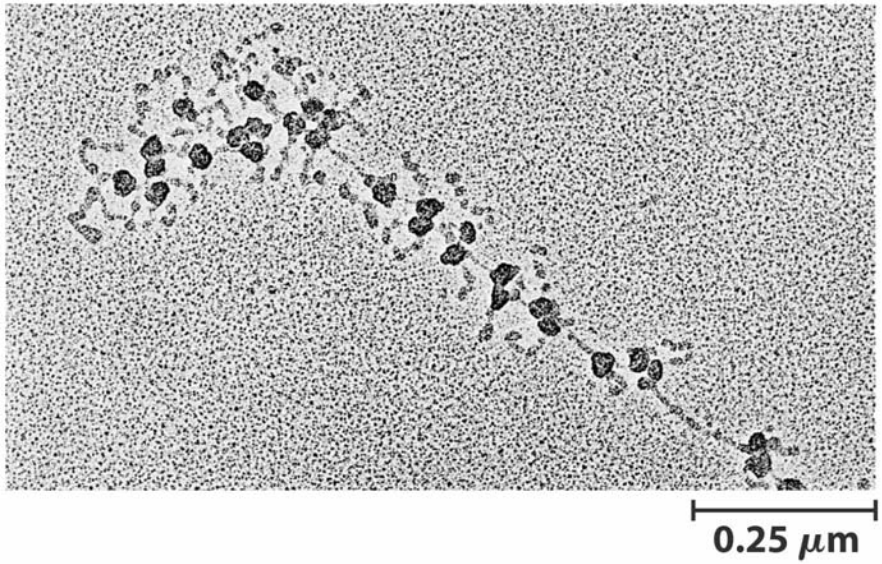
Elongation continues

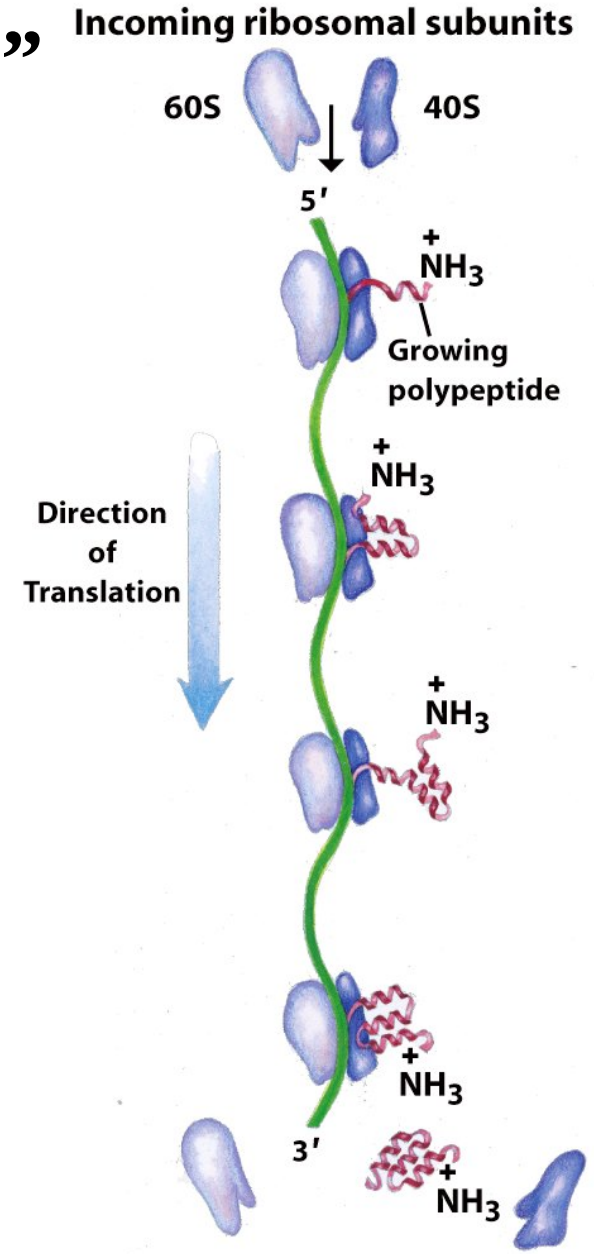
Termination



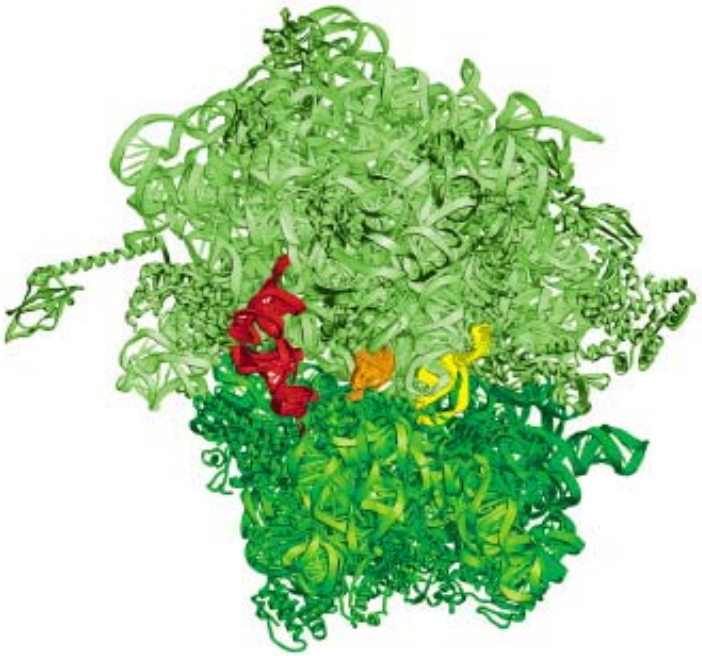
Continuous protein synthesis on one mRNA template: 'Polysome"





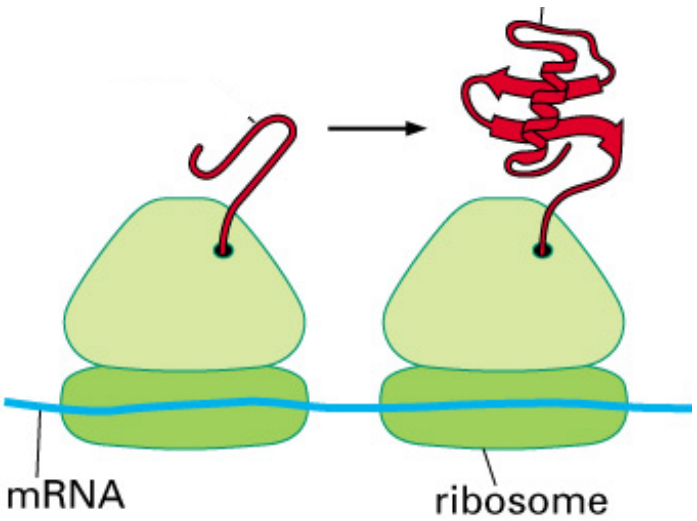


Ribosomes



Q. How many ribosomes in a cell?

A. About 15,000 (in E. coli cell) (Protein synthesis-related proteins makes up about 1/3 of total E. coli proteins!)



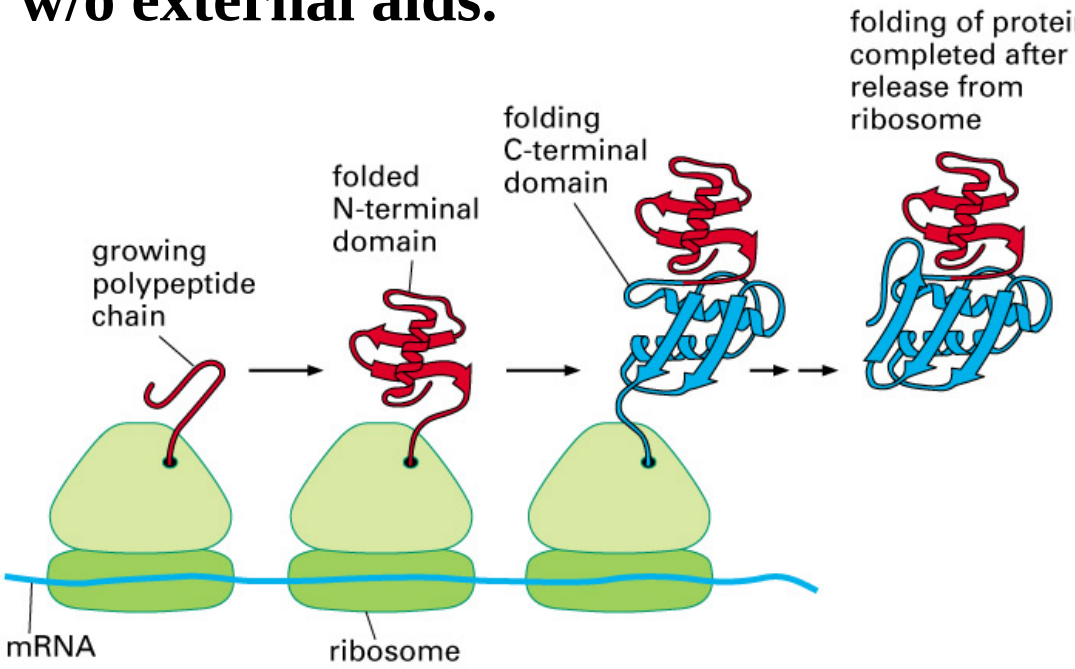
Q. How fast does the peptide elongation undergo?

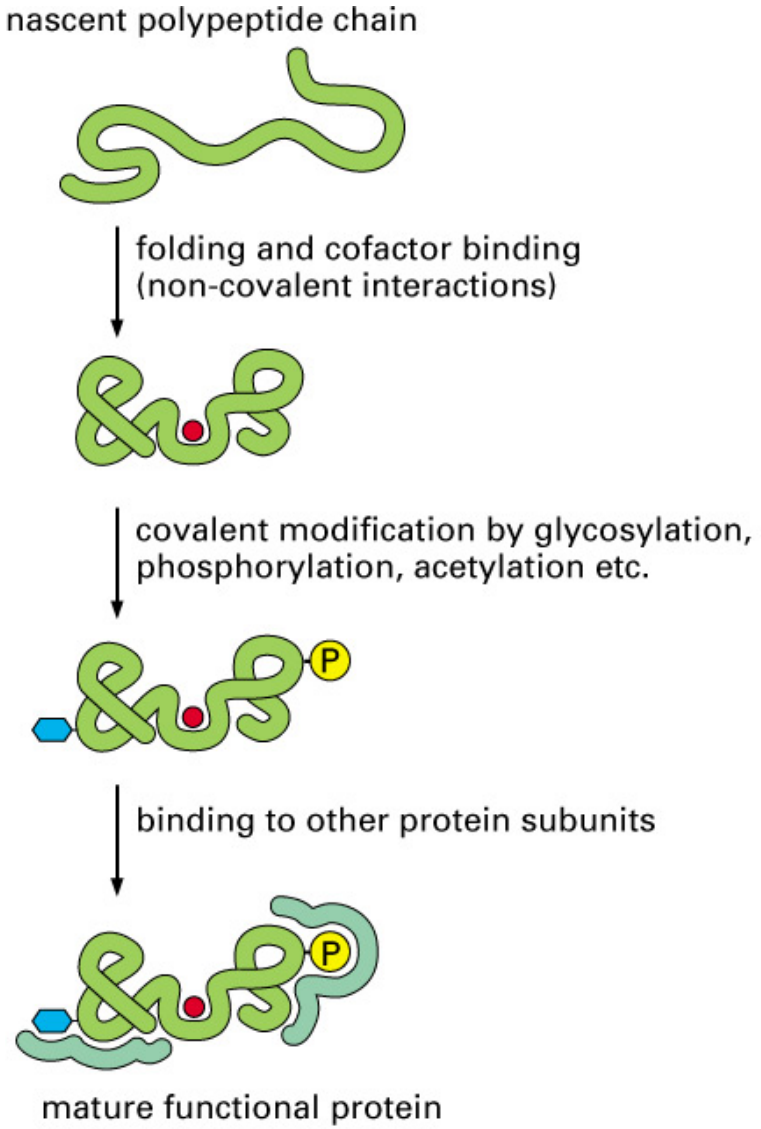
A. ~20 peptides / sec (~2kD / sec) (in E. coli at 37°C; this is incredibly faster than most other cells)

→ It takes minutes to make a protein!

Polypeptide maturation

Each synthesized protein (polypeptide chain) then takes its natural secondary structure (such as α-helix), depending on neighbor amino acids automatically. It then takes larger 3-D structure with or w/o external aids.

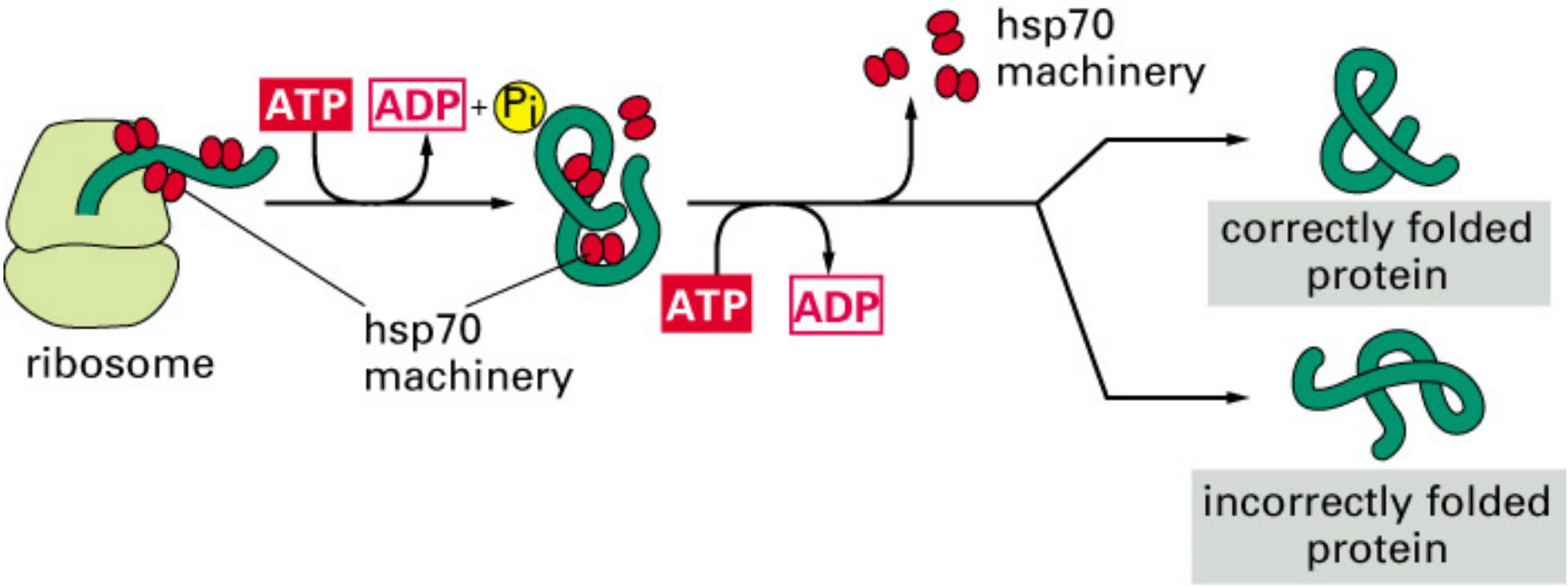




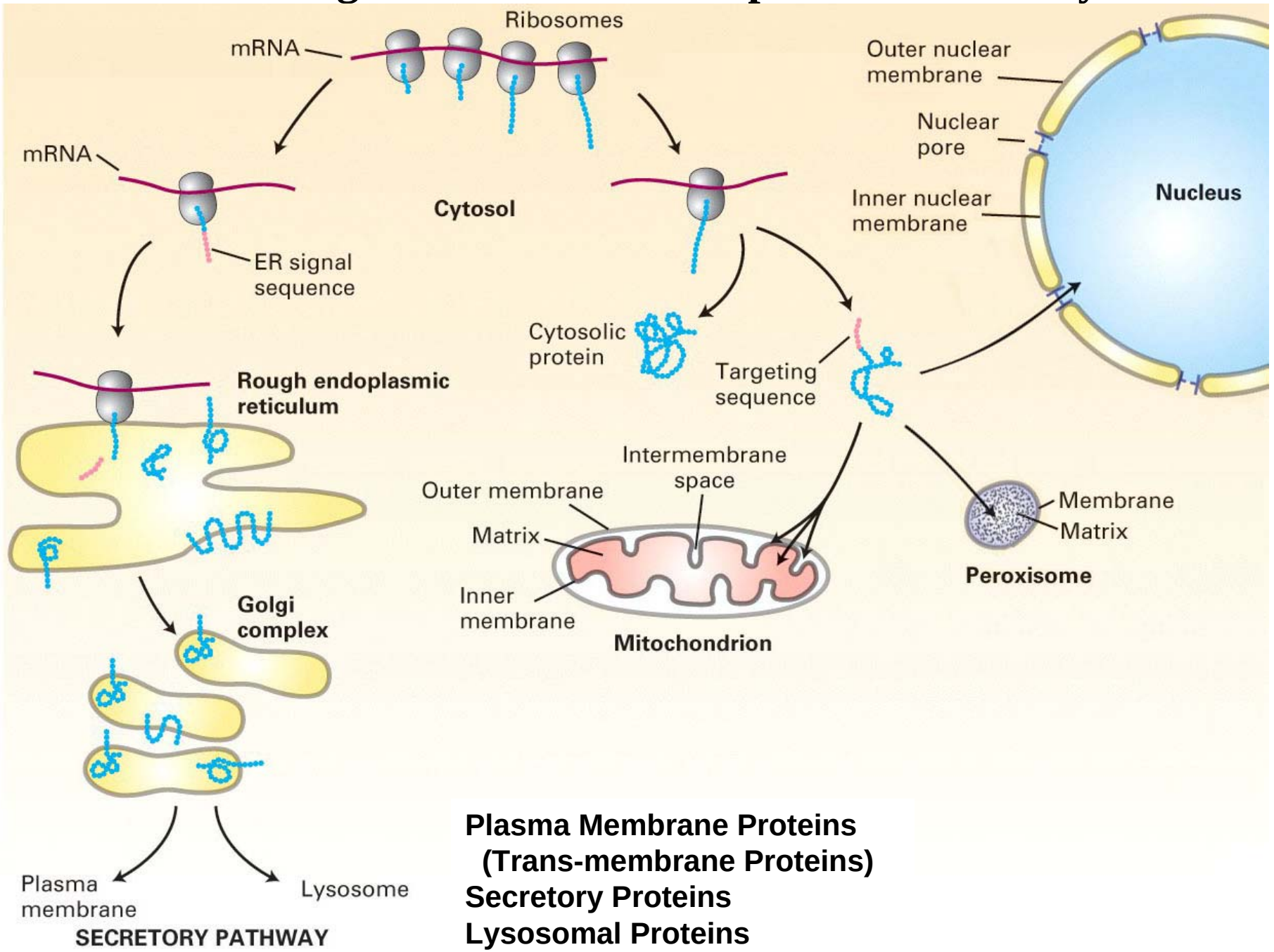
Secondary structure formation takes place as soon as the amino acid chain stretch is formed and out from ribosome.

Polypeptide maturation

Many proteins require external aid (molecular chaperone, chperonin) to for tirtialy 3-D structure.



Overview of sorting of nuclear-encoded proteins in eukaryotic cells



Plasma Membrane Proteins
(Trans-membrane Proteins)
Secretory Proteins
Lysosomal Proteins

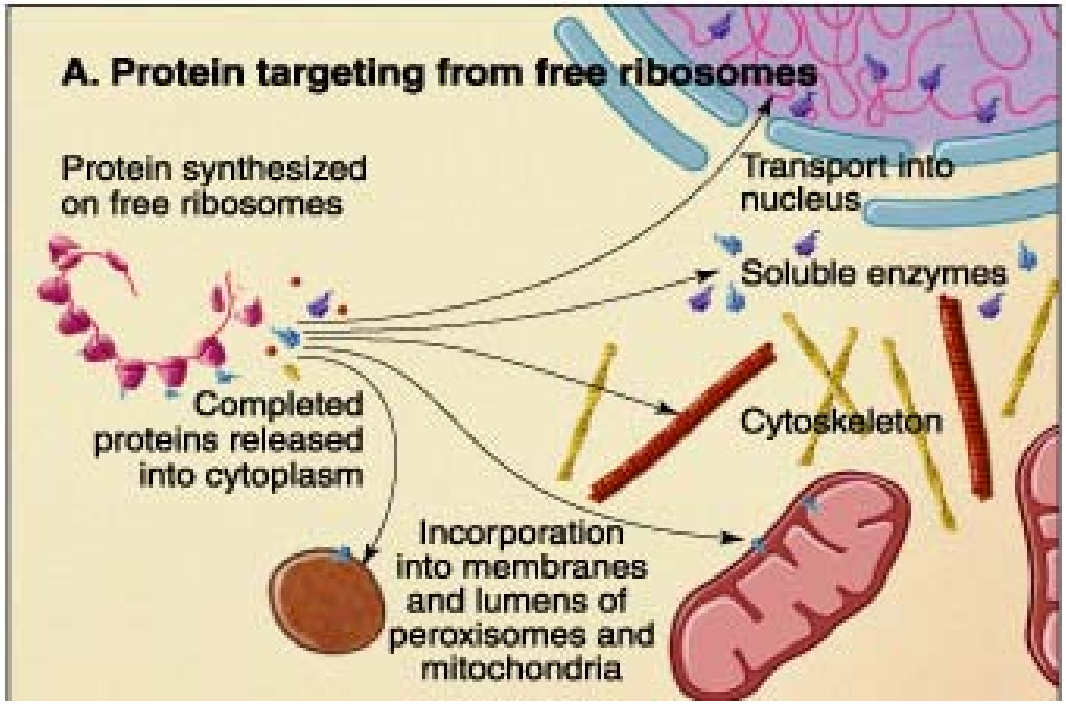
Proteins made with free ribosomes in cytoplasm

Cytoplasmic proteins
- **Soluble enzymes**
  - **Metabolic enzymes**
    - **Glycolytic enzymes**
- **Cytoskeletan Proteins**
  - **Actin, Tubulin, Keratin**
  - **Myosin, Kinesin, Dynein**

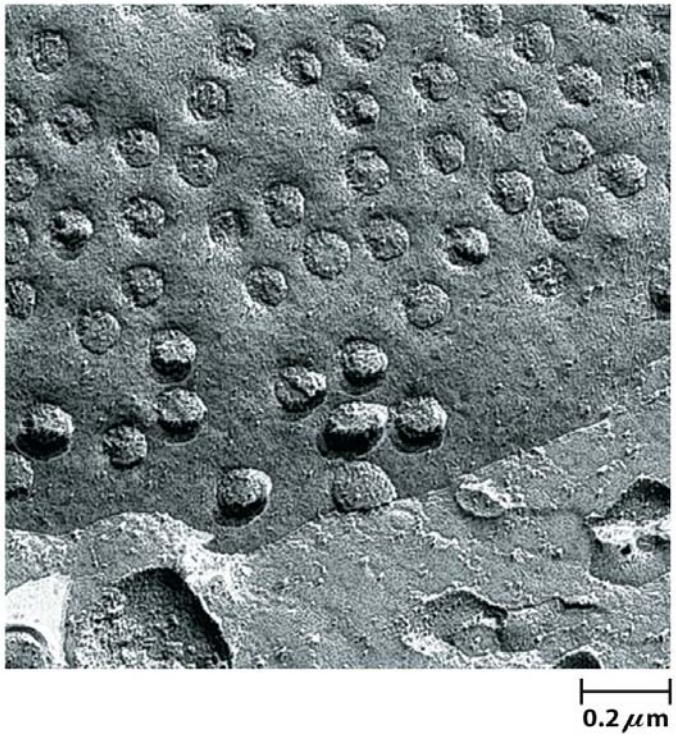
Nuclear Proteins
- **Transported into nucleus w/ specific mechanism**

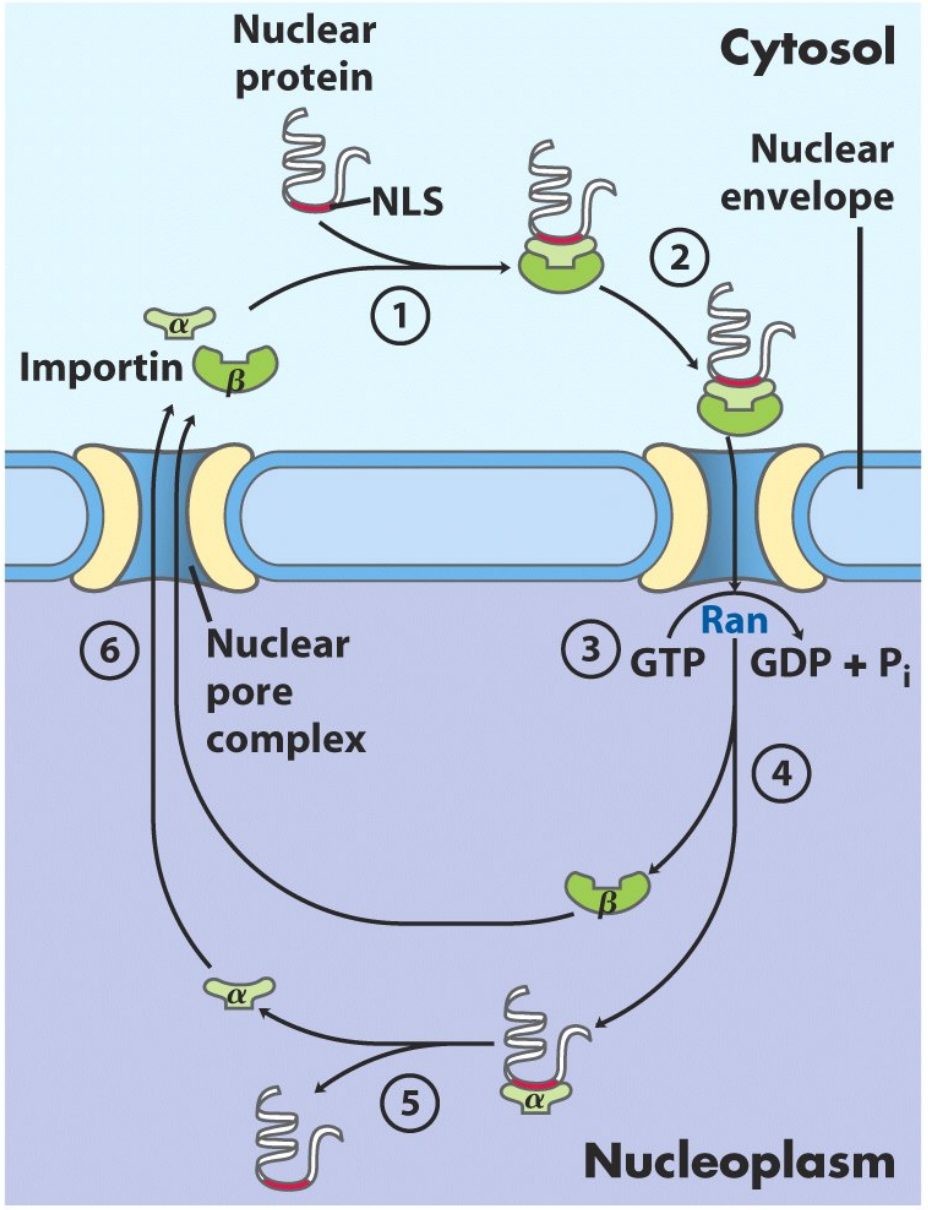
Mitochondrial proteins
- **Transported into mitochondria w/ specific mechanism**

Peroxizomal proteins
- **Transported into peroxizomes w/ specific mechanism**

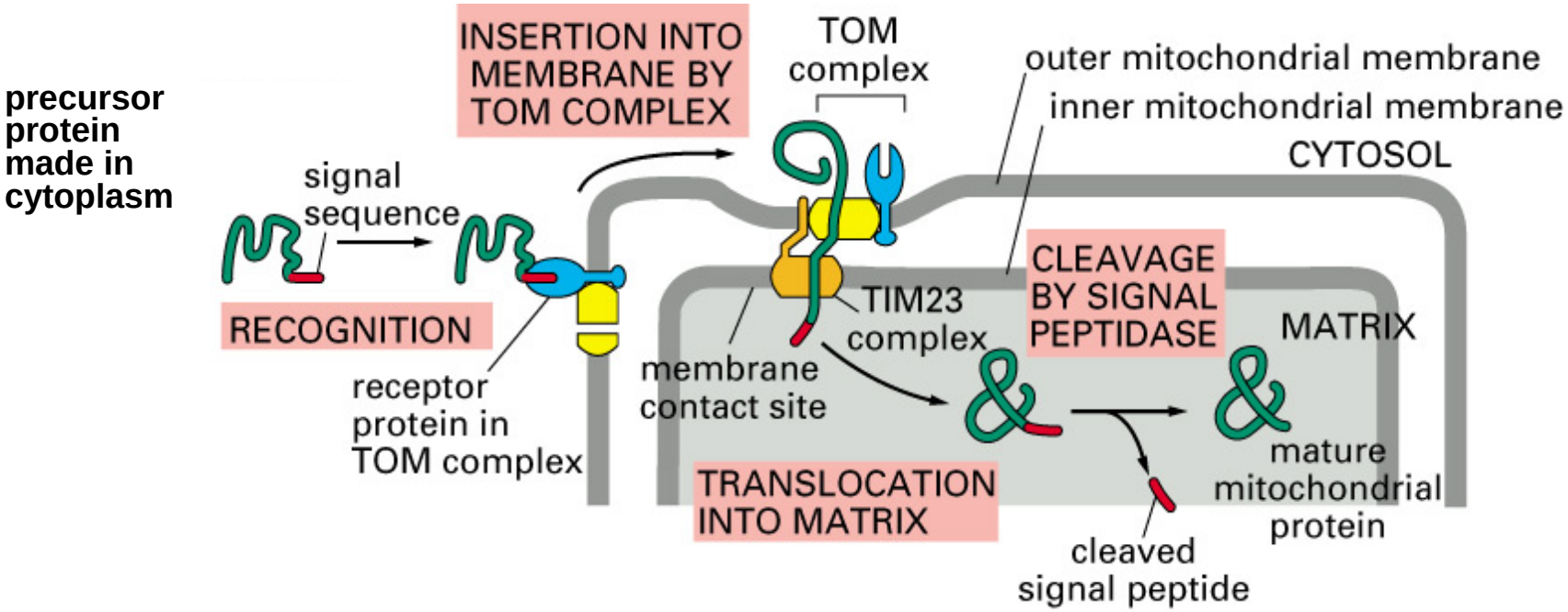


Proteins targeted to nucleus have special sequence, and specific mechanism brings them into nucleus.





Protein Targeting: Mitochondrial Proteins

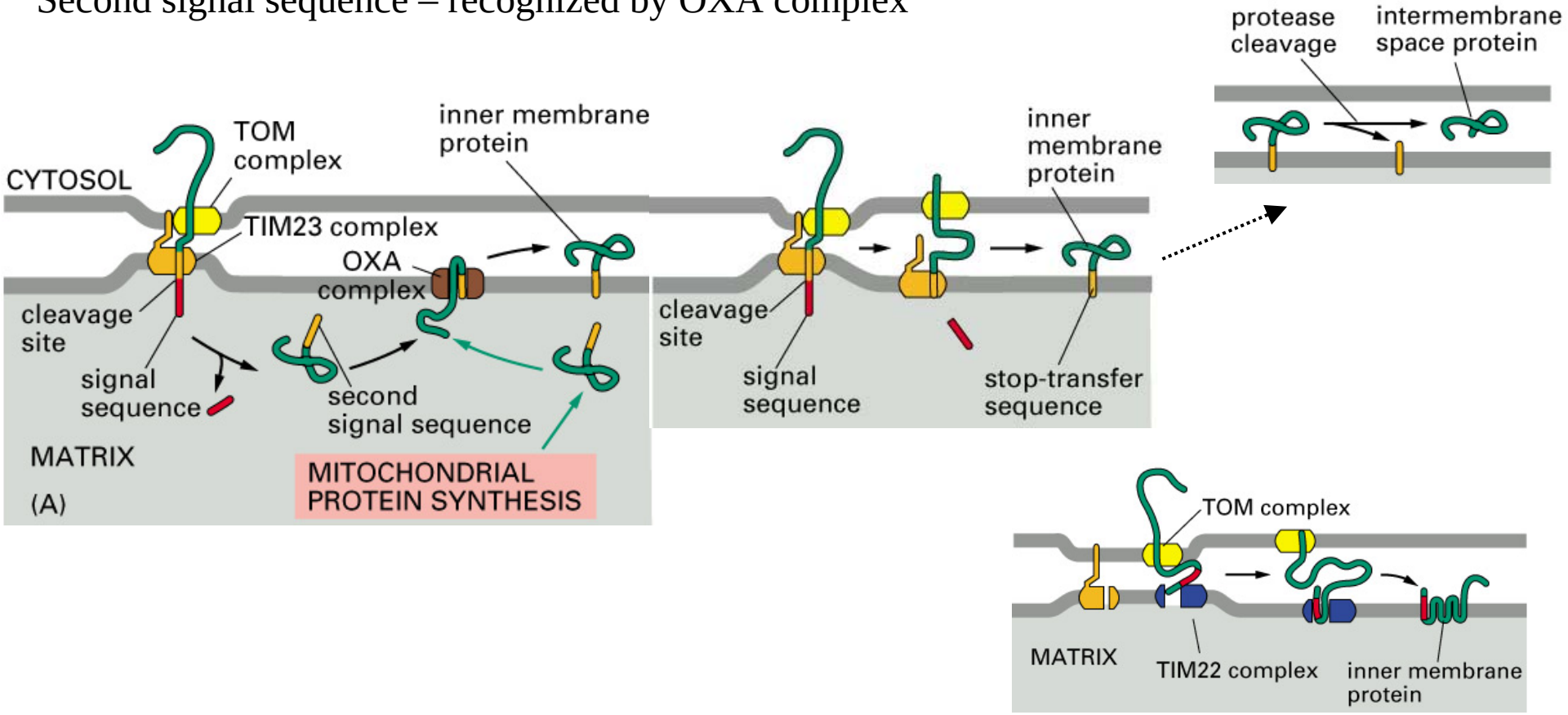


Mitochondorial proteins have special peptide sequence(s) "Signal Sequence" on their N-terminal end, that is recognized by protein import mechanism (protein complex) located on outside on the mitochondorial membranes.
Depending on the location of protein, different signal sequence/import mechanism will be employed.

Protein Targeting: Mitochondrial Proteins

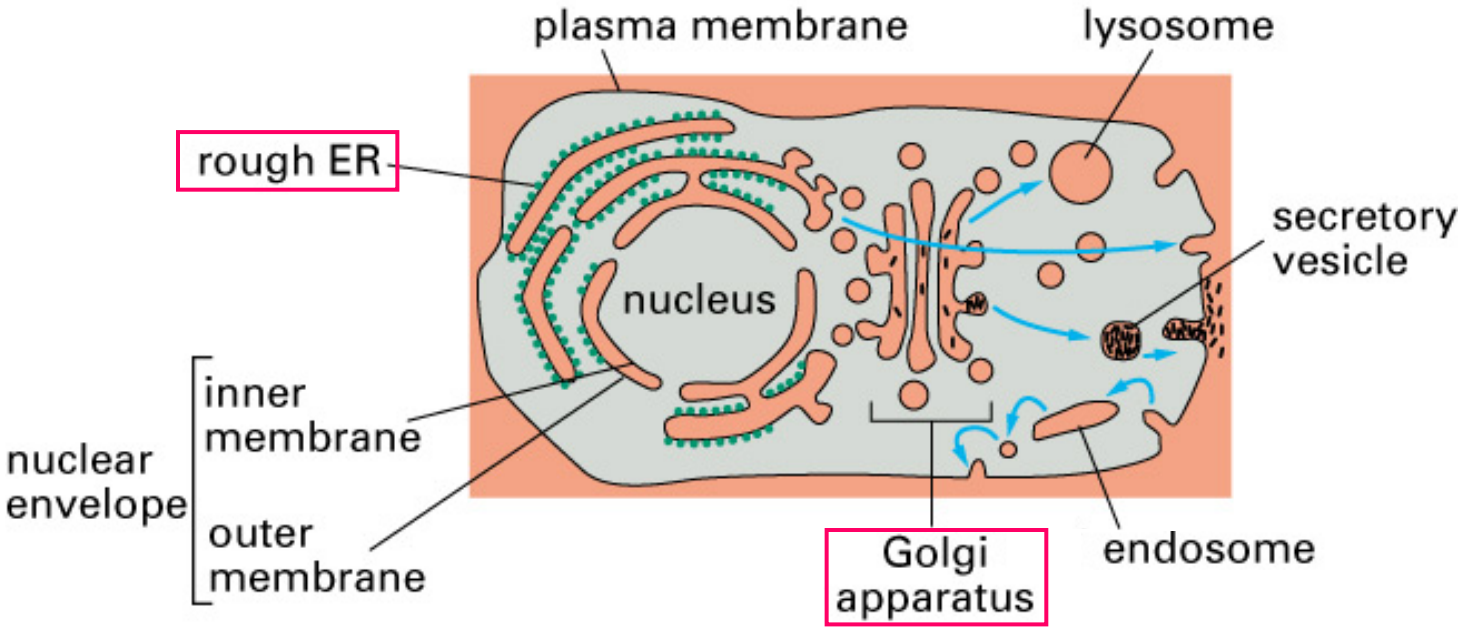
Importing mechanisms:
TOM complex – outer membrane
TIM23 complex, TIM22 complex,
OXA complex **– inner membrane**

Proteins to be imported:
Signal sequence – recognized by TOM (+TIM23) complex,
or TOM/TIM22 complex
Stop-transfer sequence - recognized by TIM23 complex
Second signal sequence – recognized by OXA complex



Protein Synthesis for Secretory Pathway

Organelles involved in Protein Synthesis for secretory proteins (incl. digestive enzymes, immunoglobulins, peptide hormones, peptide neurotransmitters, extracellular matrix proteins), plasma membrane proteins, ER, Golgi, lysosomal proteins



Key Issues:

Targeting signal for rough Endoplasmic Reticulum (rER) entry and the import mechanism
Protein glycosylation
Protein modifications
Protein transport

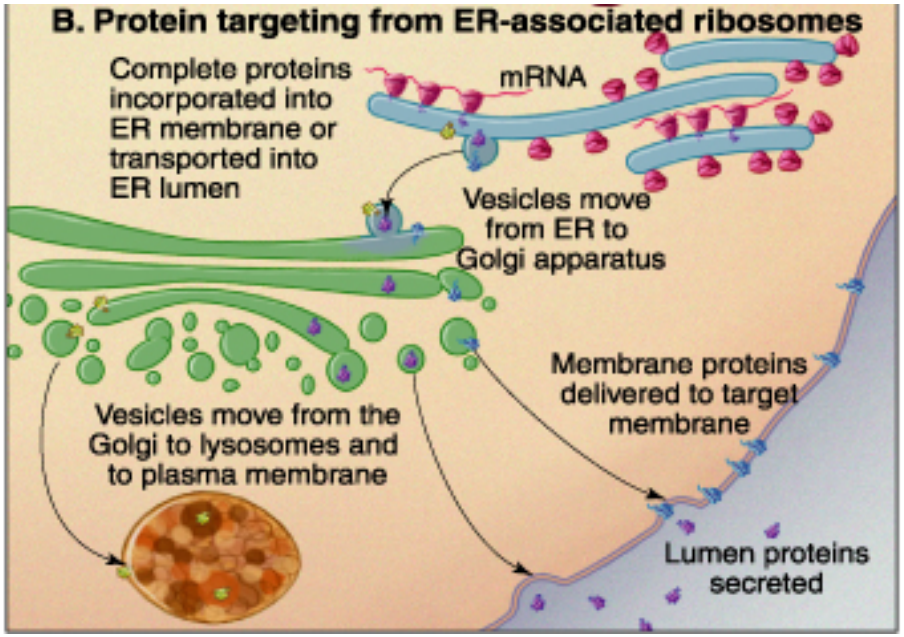
Proteins made with ER-associated ribosomes

ER, Golgi proteins

Plasma Membrane Proteins
- **Channels, pumps, receptors**
- **Adhesion proteins**
- **MHC, Glycocalix**

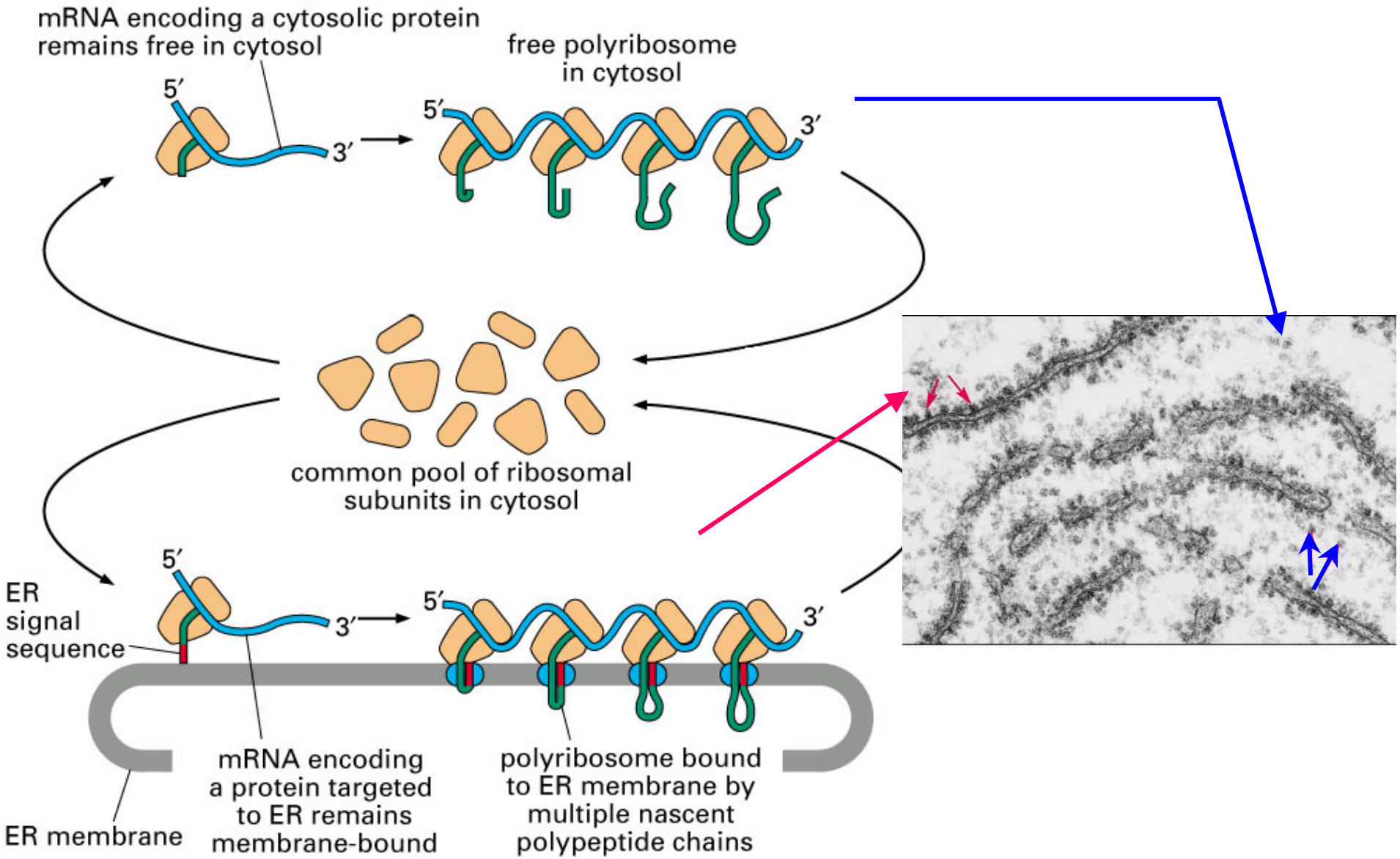
Phagosomal / Lysosomal Proteins
- **Digesting Enzymes**
- **Pumps**

<u>Proteins released to outside the cell</u>
- **Extracellular Matrix Proteins**
  - **Collagen, Fibronectin…**
- **Albumin, Cofactors, Fibrinogen…**
- **Immunoglobulin (IgG, IgM…)**
- **Peptide Hormones, Peptide Neurotransmitters**
  - **Insulin, Growth Hormone**
- **Digestive Enzymes (Zymogens)**
  - **Pepsinogen, Ribonuclease**
- **Mucus proteins (Mucins), Milk proteins**



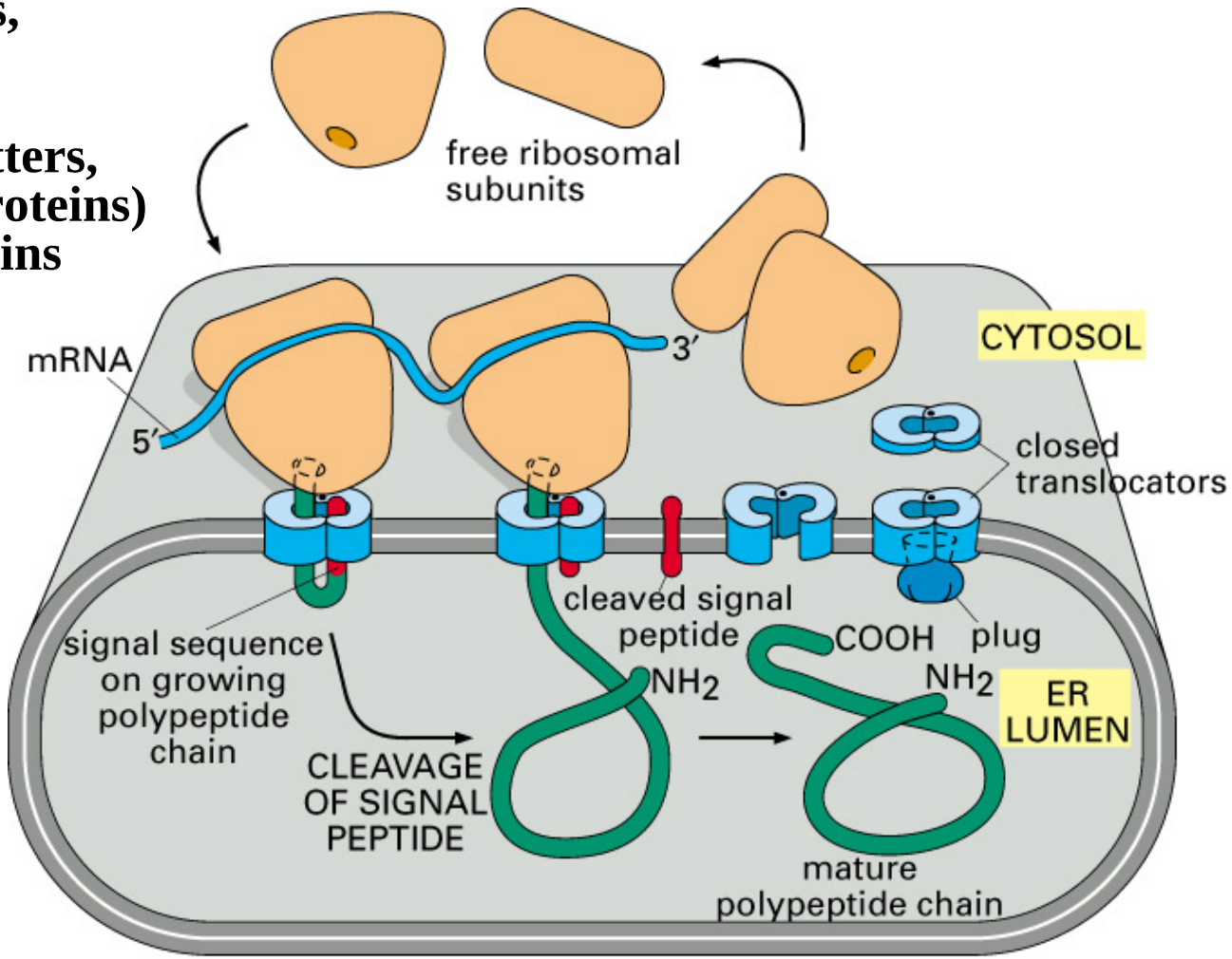
Post-translational modification (Glycosylation)

Same ribosomes (Eukaryotic type ribosomes) are used for either cytoplasmic protein synthesis or protein synthesis on rER

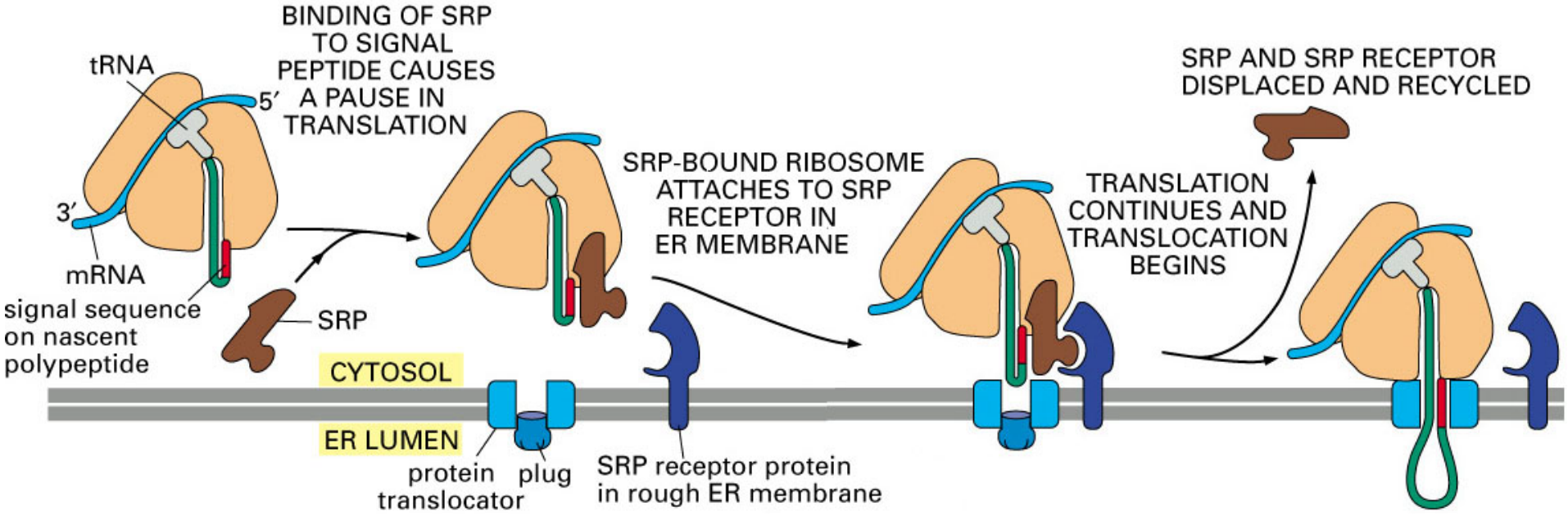


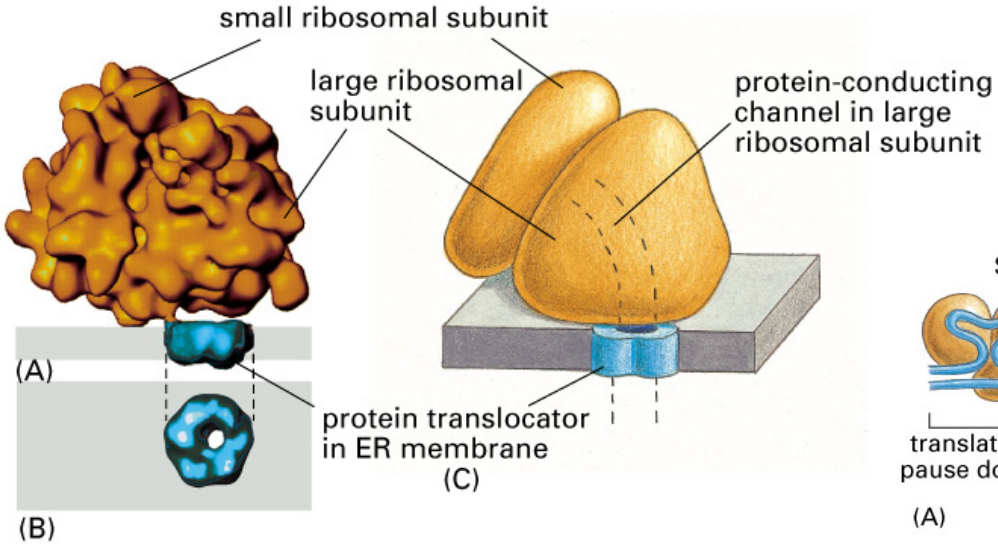
Certain proteins* are translated by ribosomes at cytoplasmic surface of rER; they have specific 'signal sequence (peptide sequence)' at their N-terminus, that will let the synthesized polypeptide to go into the lumen of rER (Cistern) through specific 'pore'.

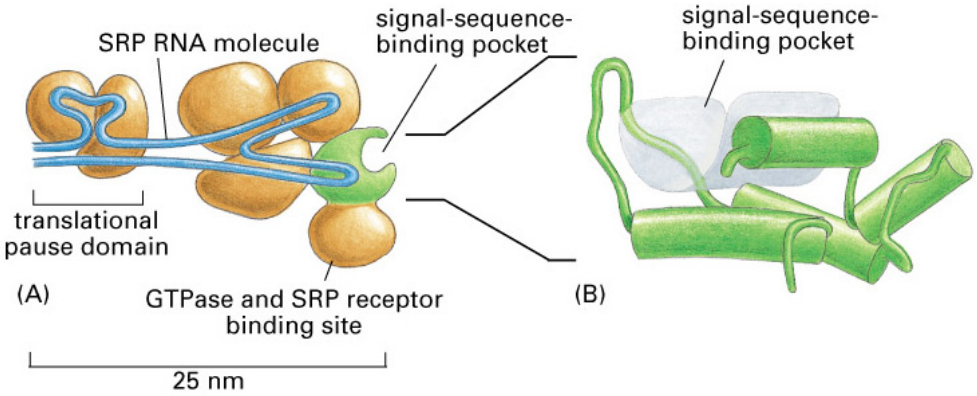
***Secretory proteins**
(incl. Digestive enzymes, immunoglobulins, peptide hormones, peptide neurotransmitters, extracellular matrix proteins)
Plasma membrane proteins
ER, Golgi proteins
Lysosomal proteins



Specific rER-targeted Signal Sequence is used to direct polypeptide into rER



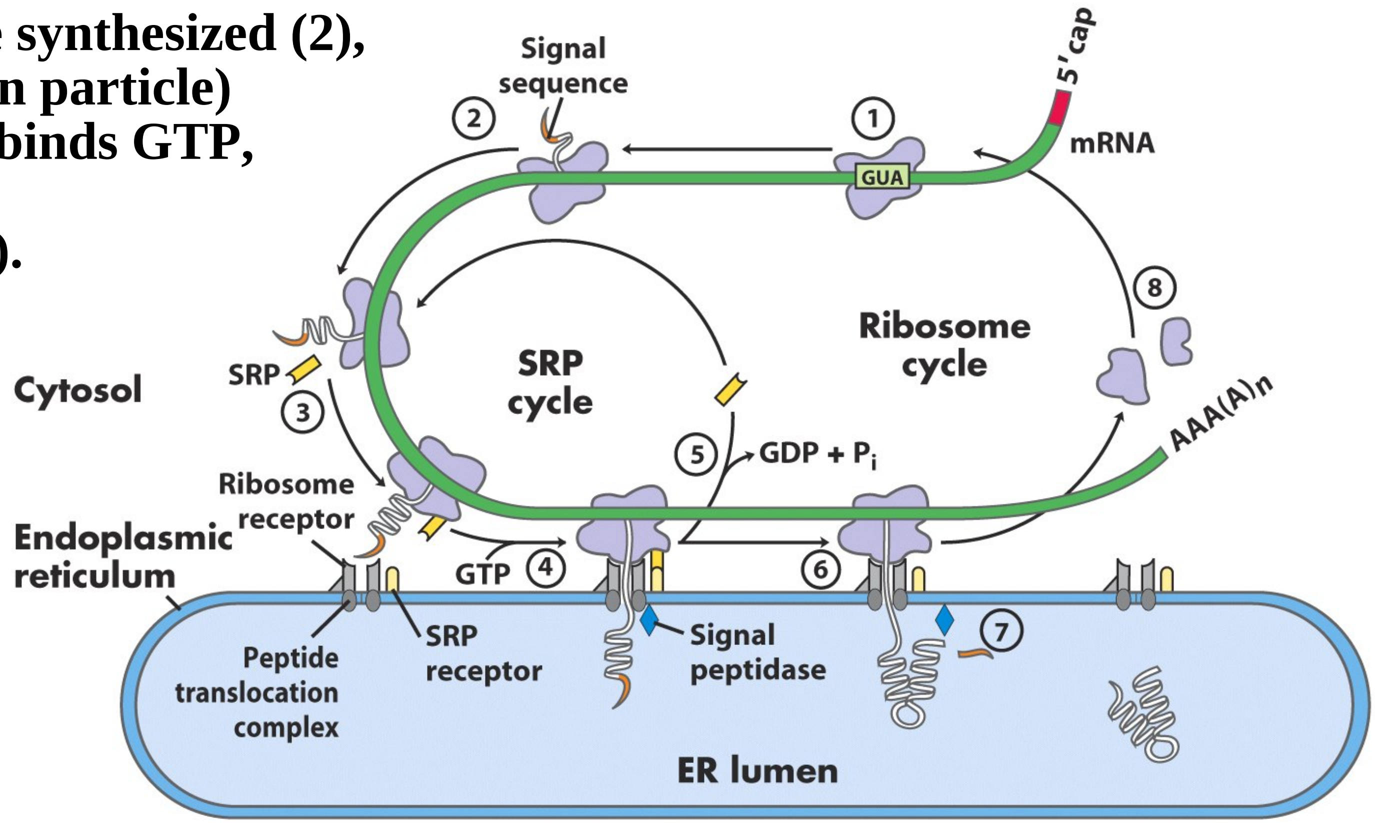




Protein Synthesis at rough-ER (rER)

These proteins that are produced at rER have special amino acid sequence called '<u>signal sequence'</u> at their N-terminus at the synthesis (above).

When the sequence are synthesized (2), SRP (Signal-recognition particle) binds to the ribosome, binds GTP, and halts polypeptide elongation (~70 AA) (3).

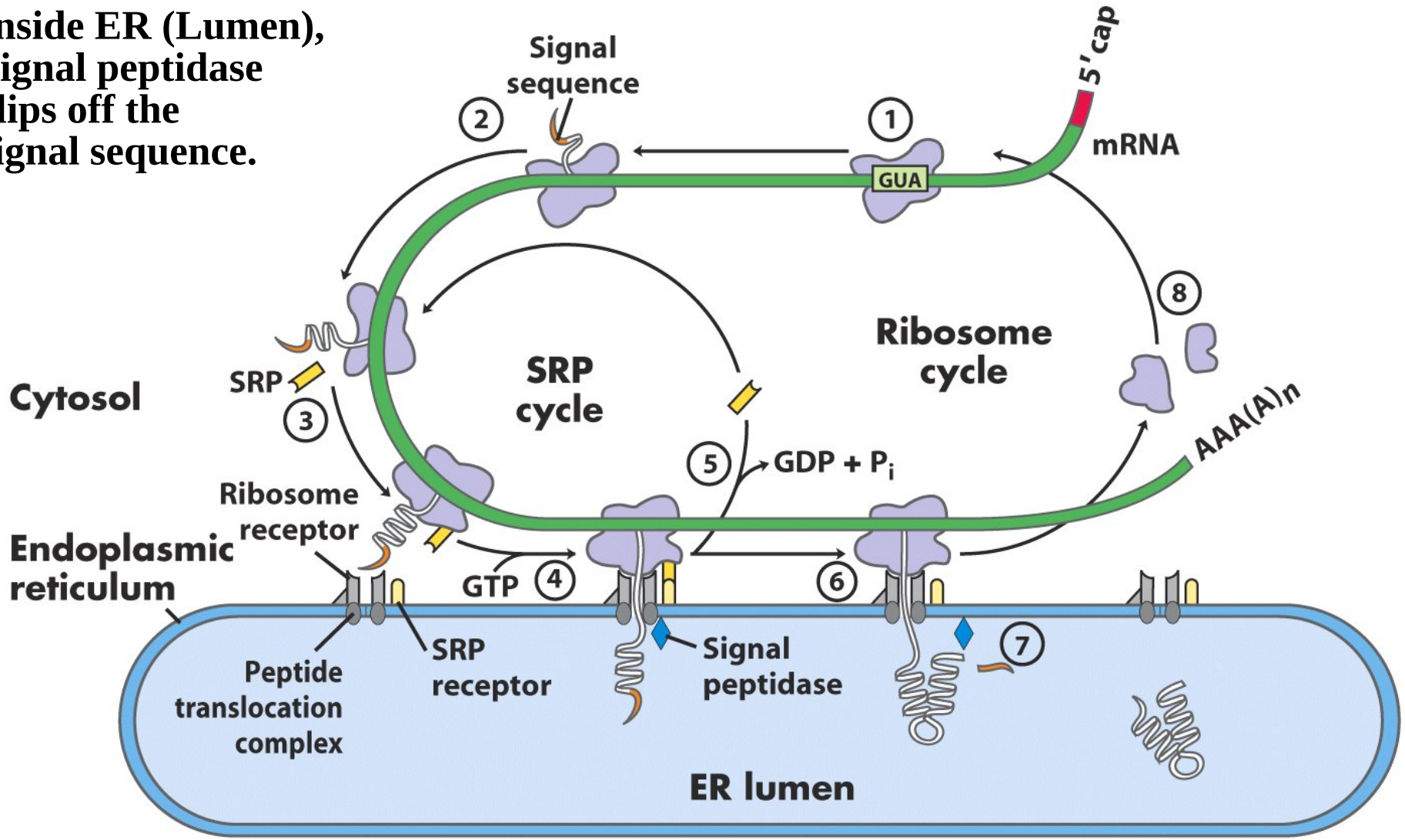


Protein Synthesis at rough-ER (rER)

The complex are now able to bind to Ribosome receptor on the surface of ER (3→4).

The peptide chain is inserted into Peptide translocation complex, and peptide elongation resumes (6). SRP is dissociated from the ribosome and recycled (5).

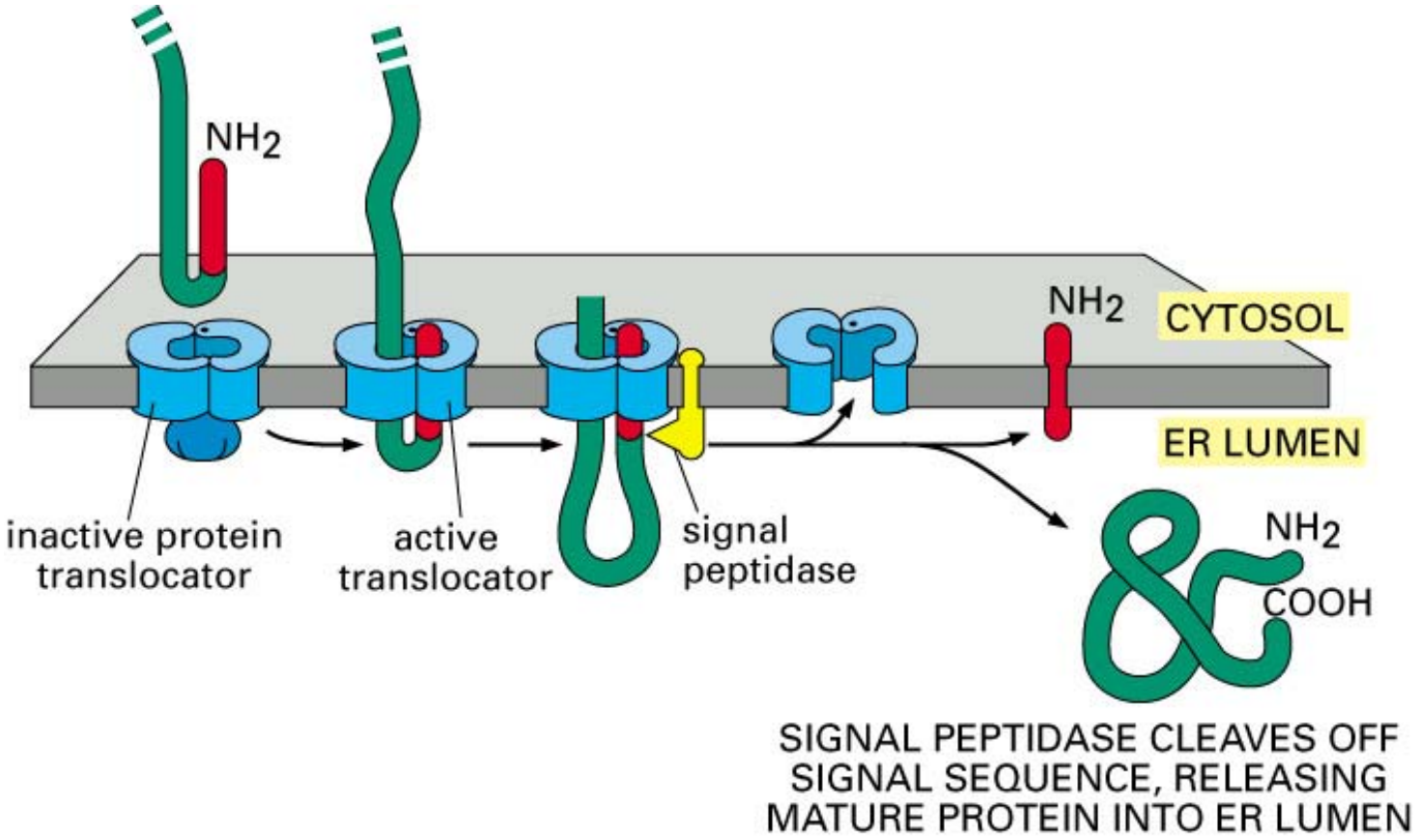
Inside ER (Lumen), Signal peptidase clips off the signal sequence.



Variations in protein translocation across rER membrane

Secretory proteins

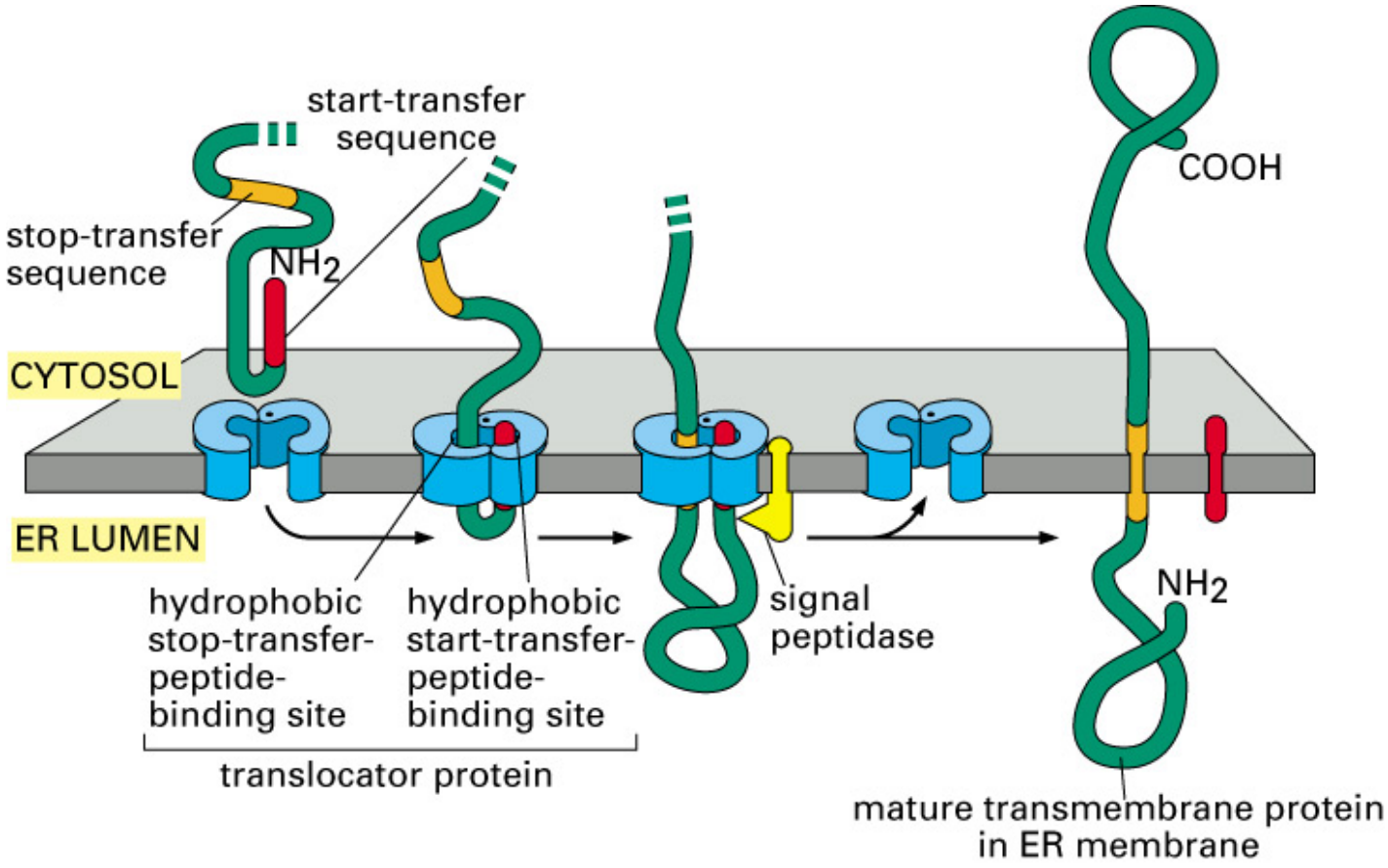
Signal sequence at N-terminus



Variations in protein translocation across rER membrane

Transmembrane proteins

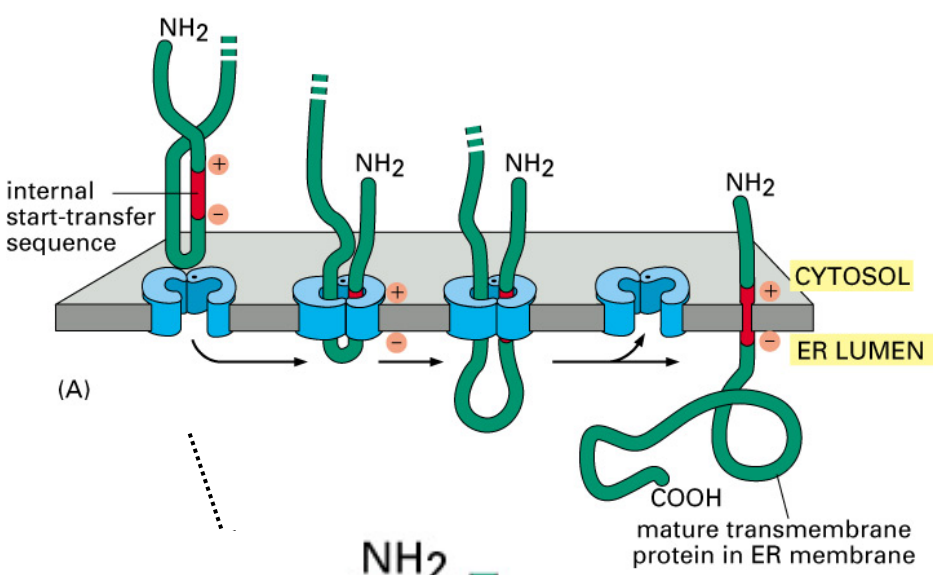
Signal sequence at N-terminus

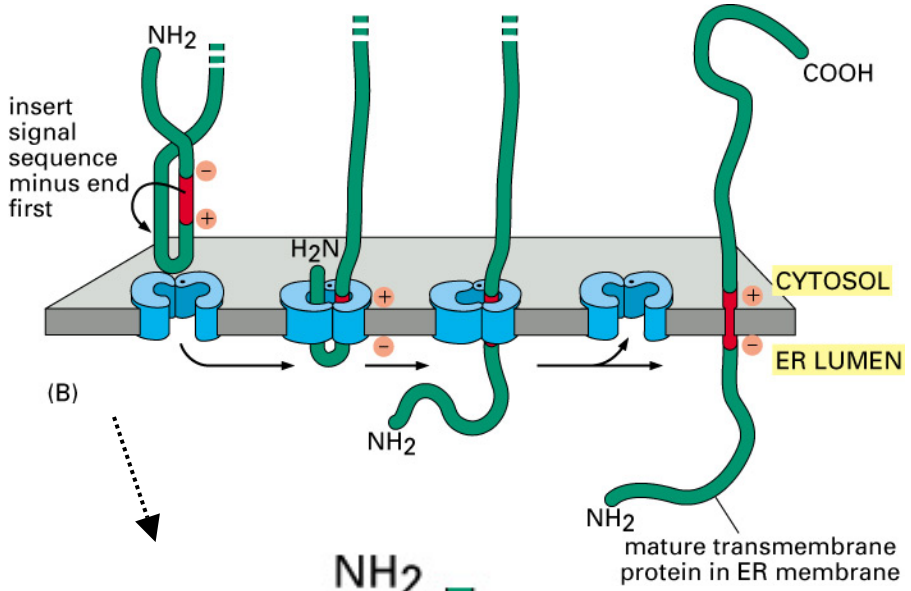


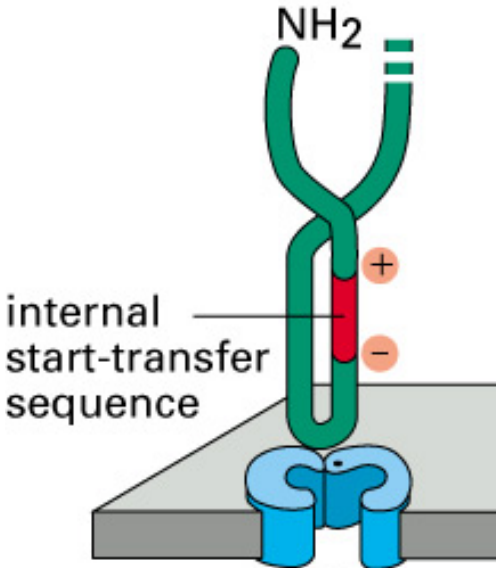
Variations in protein translocation across rER membrane

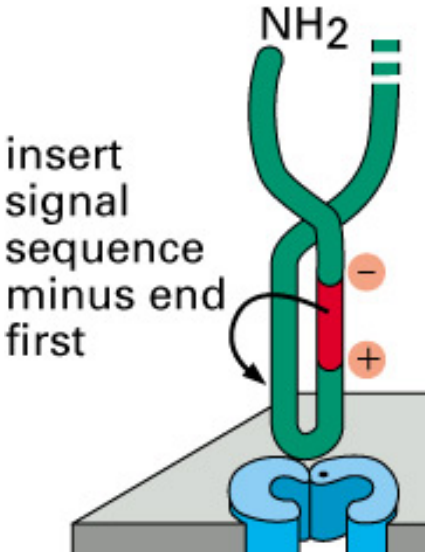
Transmembrane proteins

Signal sequence in the middle



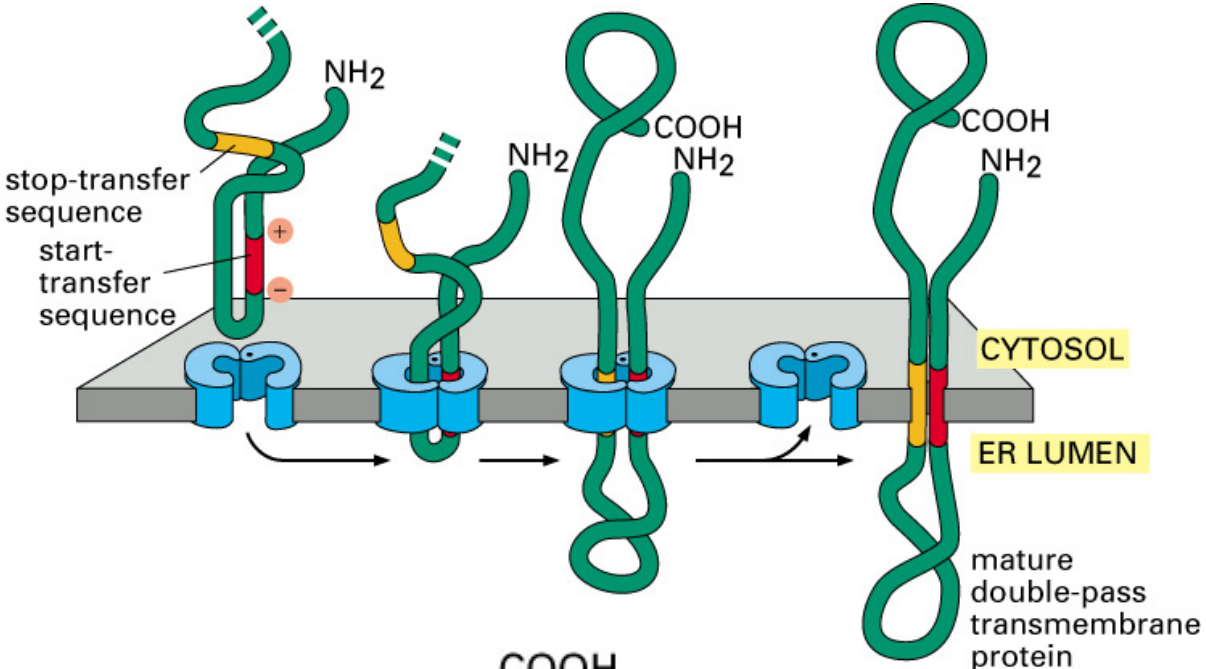


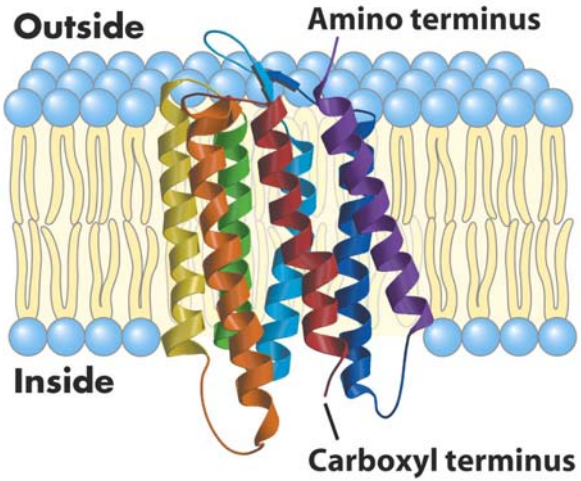


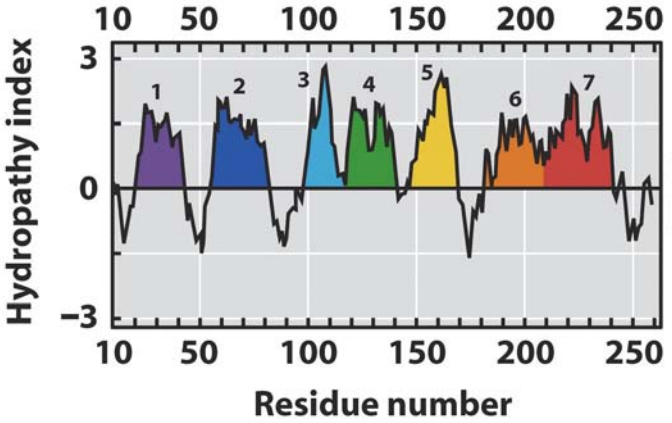


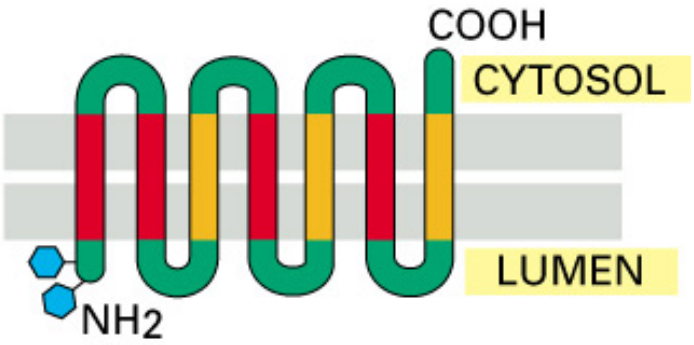
Variations in protein translocation across rER membrane (2)

Multipass transmembrane proteins



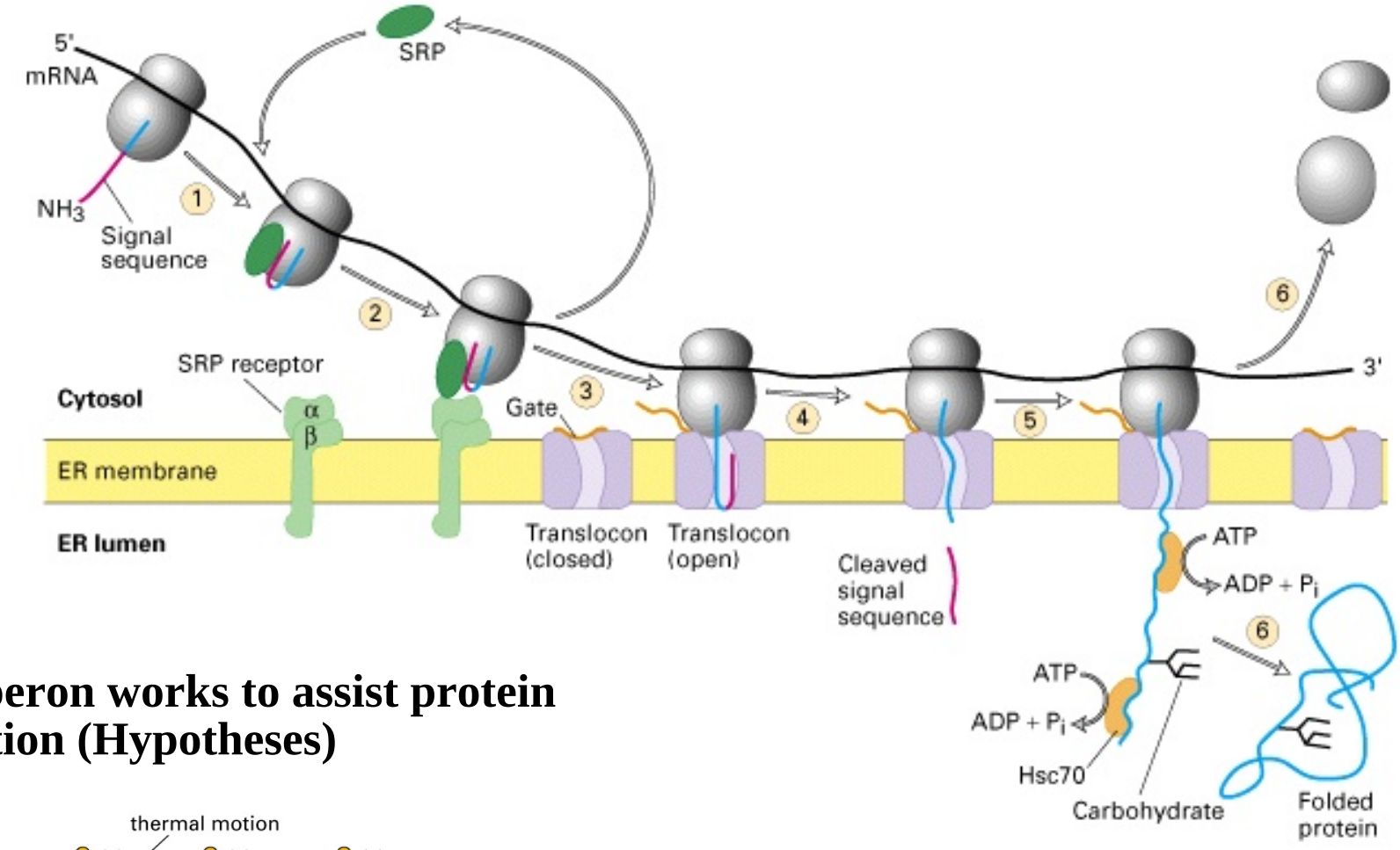




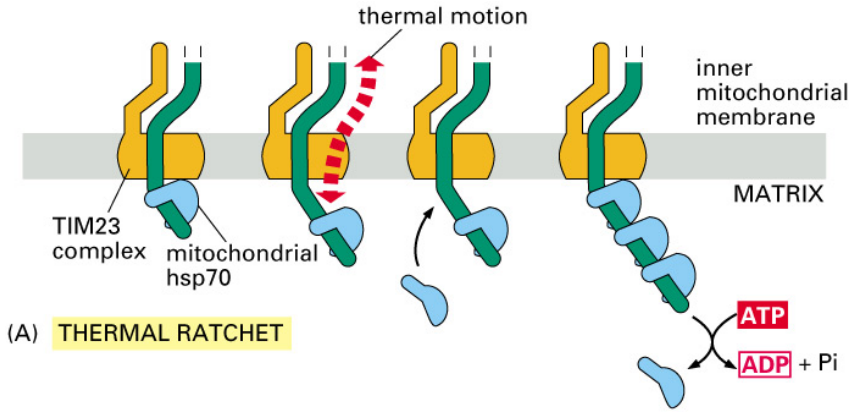


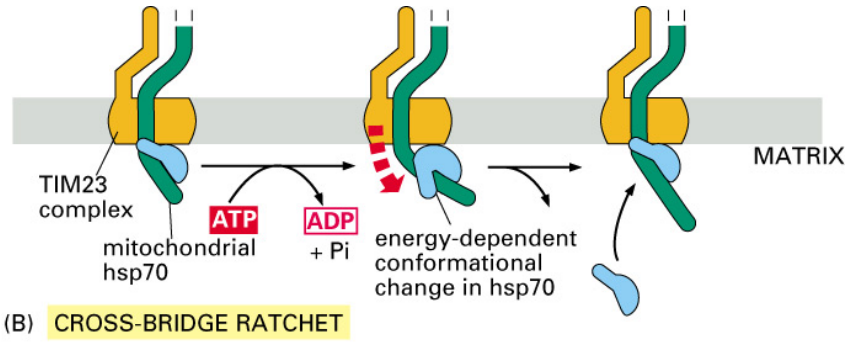
Many plasma-membrane transmembrane proteins have 7 X-membrane passes

Molecular chaperon assists protein translocation



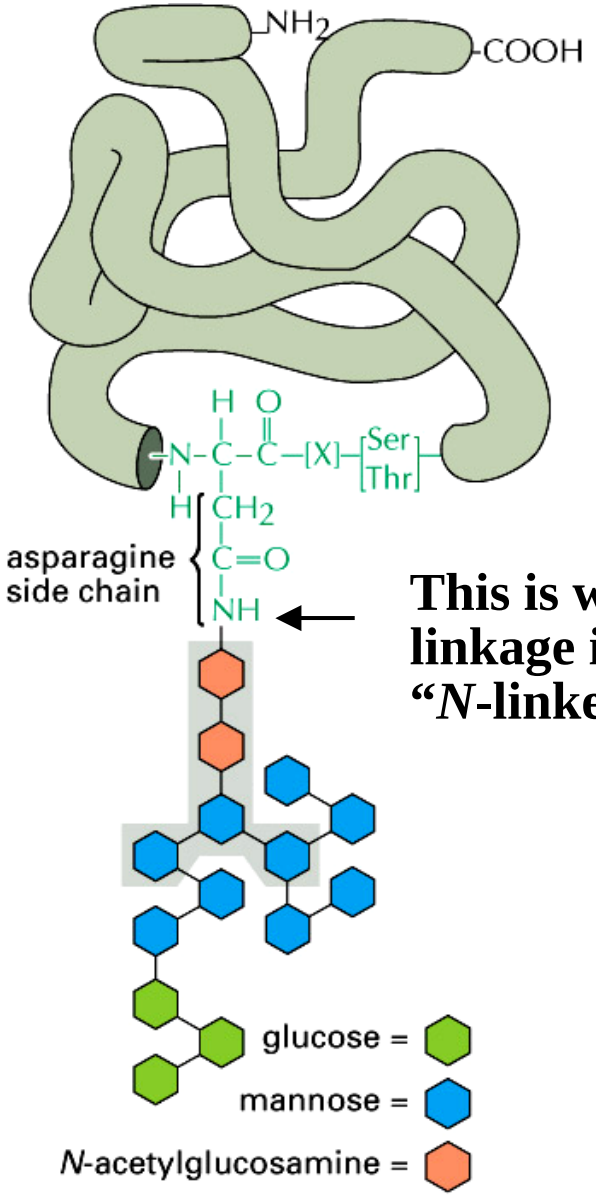
How chaperon works to assist protein translocation (Hypotheses)



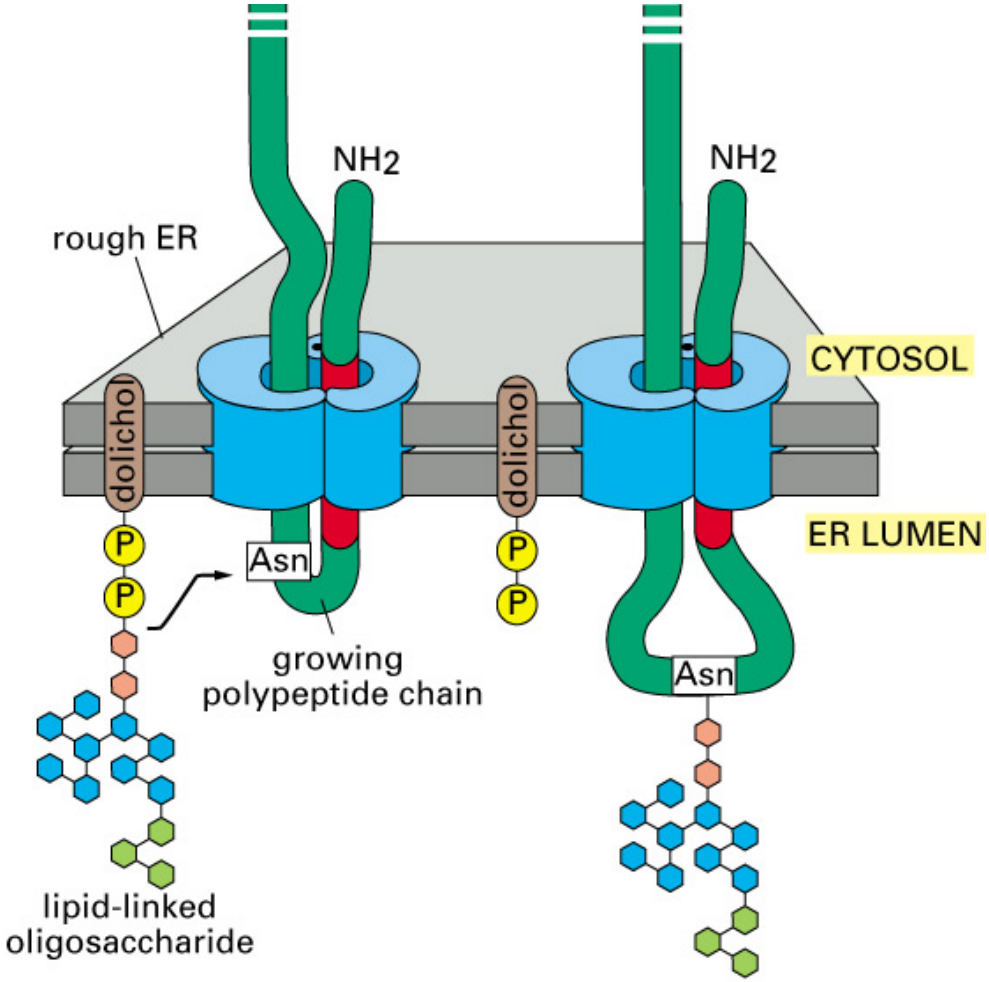


(Most of) Proteins are Glycosylated

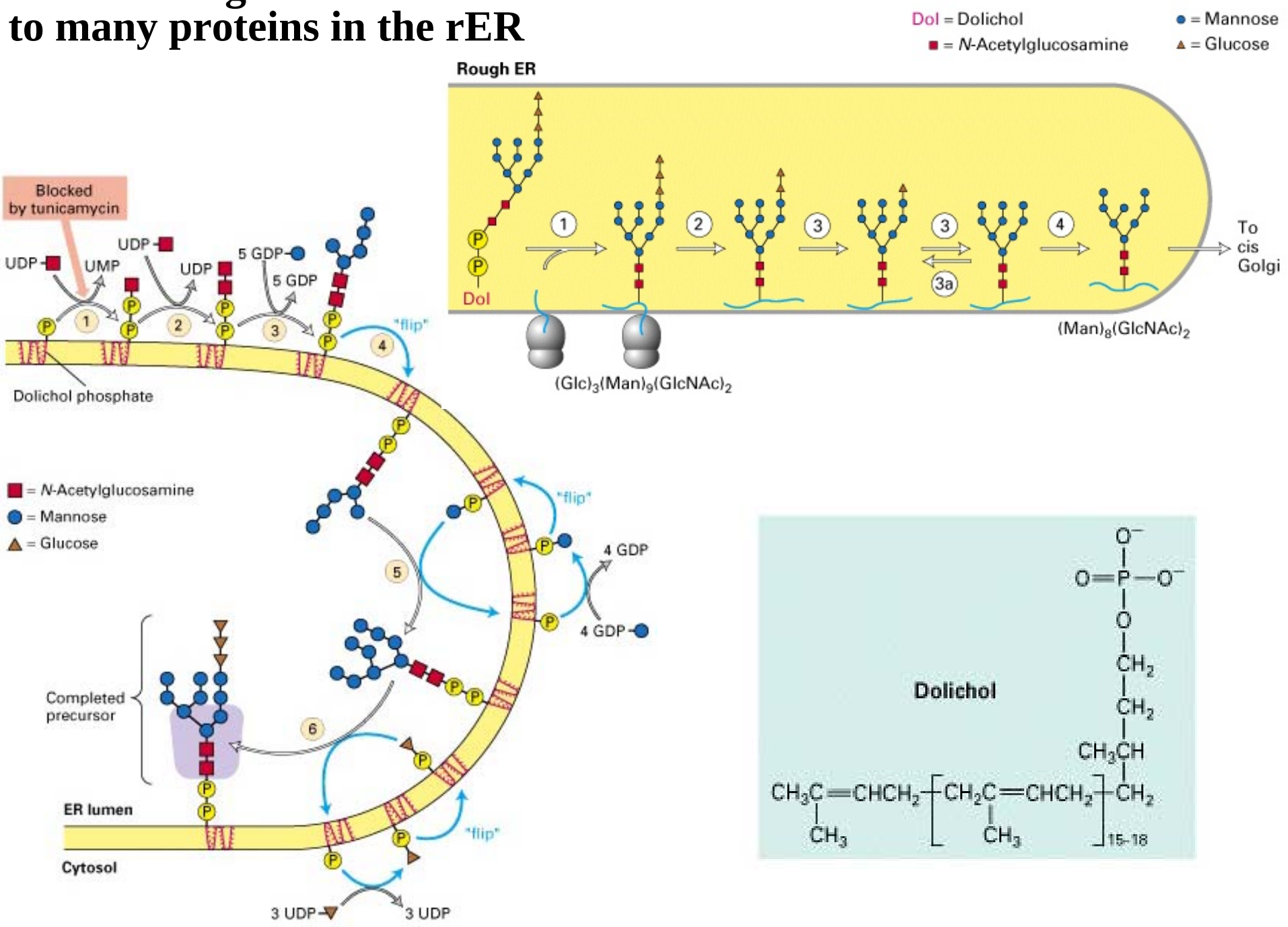
Addition of common *N*-linked Oligosccharide to Asn



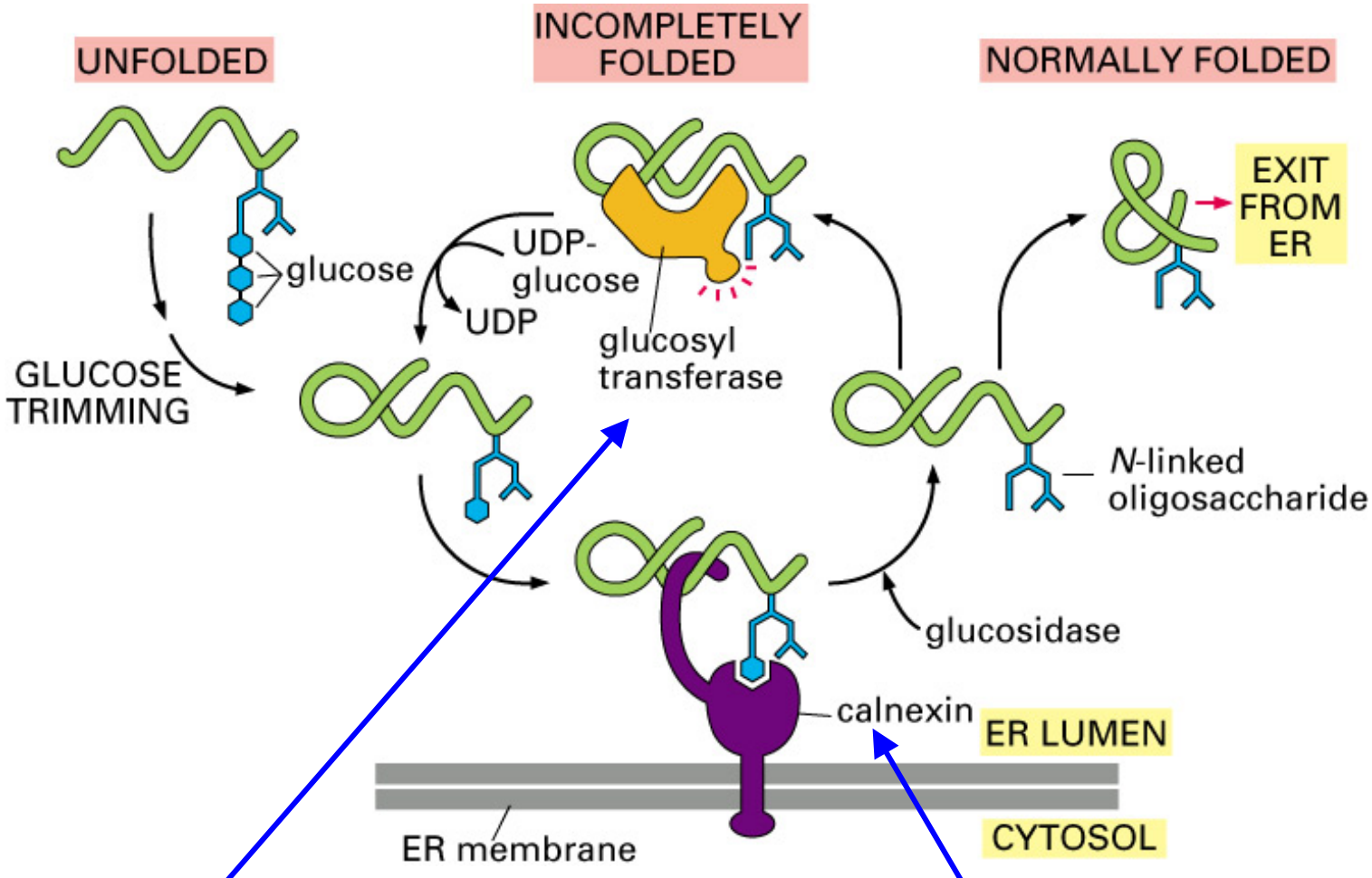
This is why the linkage is called "*N*-linked".



N-linked oligosaccharide is added to many proteins in the rER



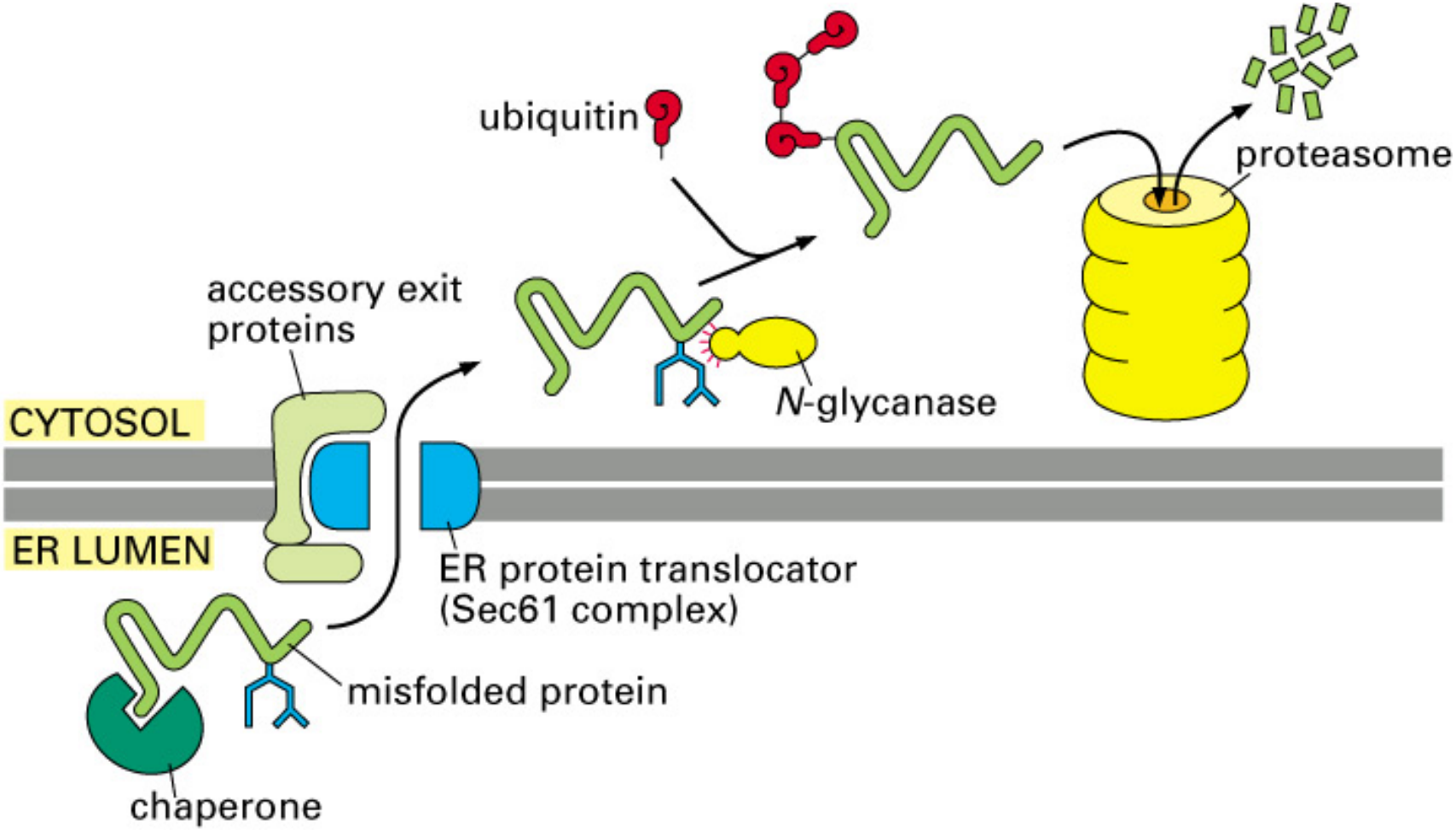
Checking the state of protein folding



Calnexin holds the glucose, so as the protein.

If the protein is not properly folded, glycosyl transferase reattach glucose at the end of the oligosaccharide, and the protein is remained in ER (until the proper job is done).

Misfolded proteins are degraded at Cytoplasm.

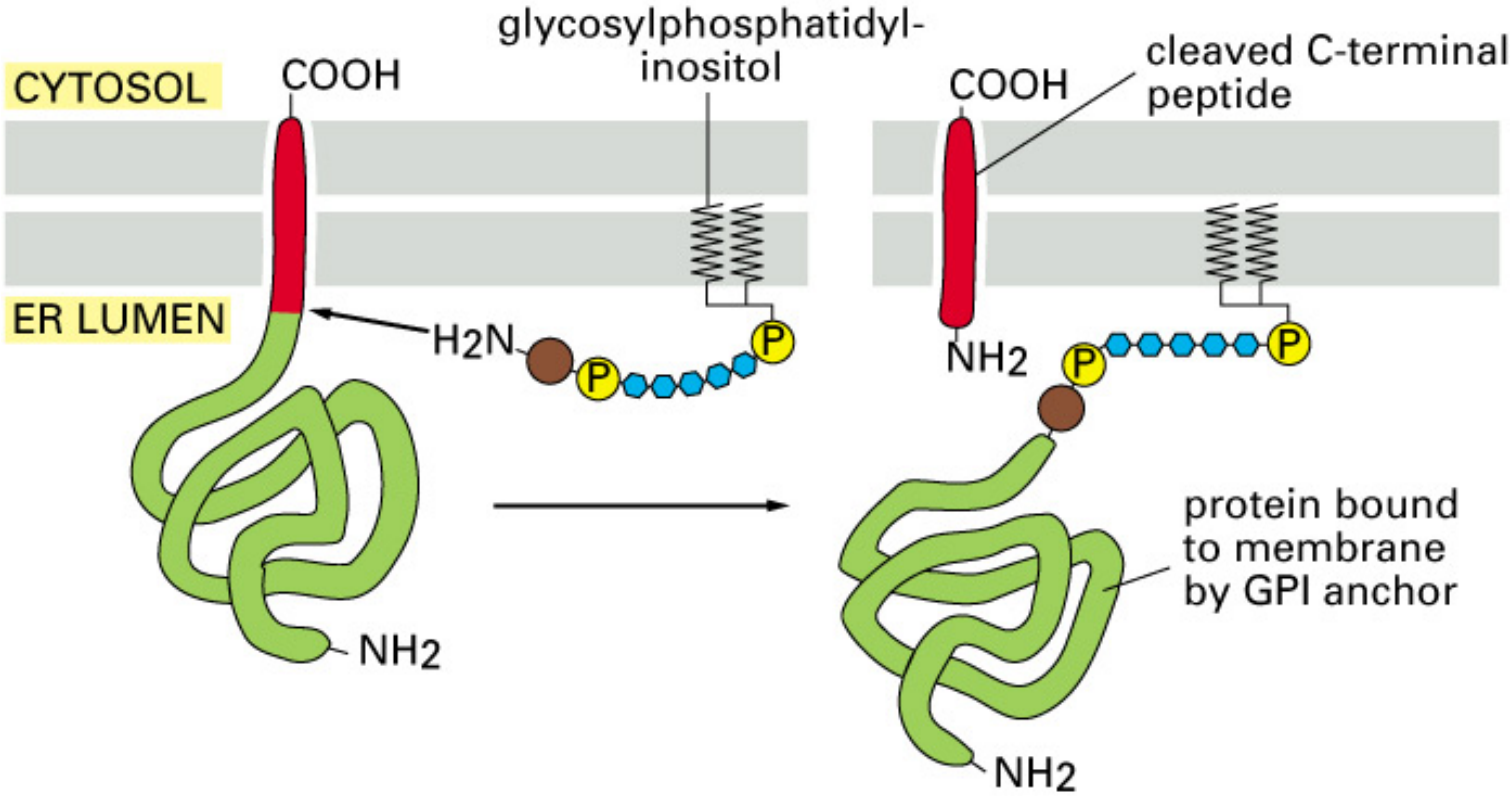


Many plasma membrane proteins become anchored via GPI into plasma membrane.

GPI: Glycosylphosphatidylinositol.

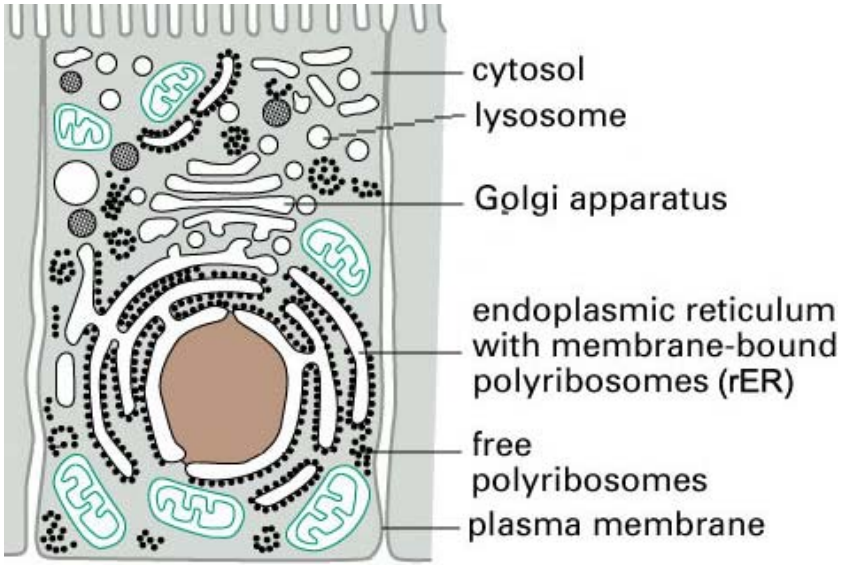
These proteins face EXTERNALLY on the plasma membrane.

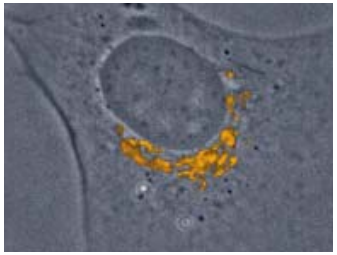
The anchor is cleavable (at plasma membrane by an enzyme).

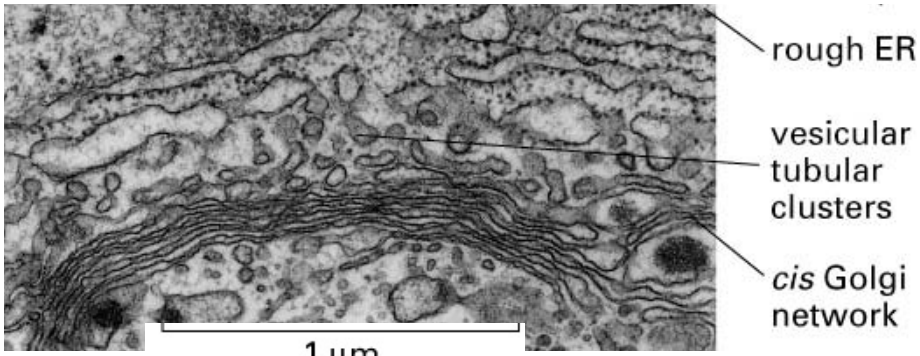


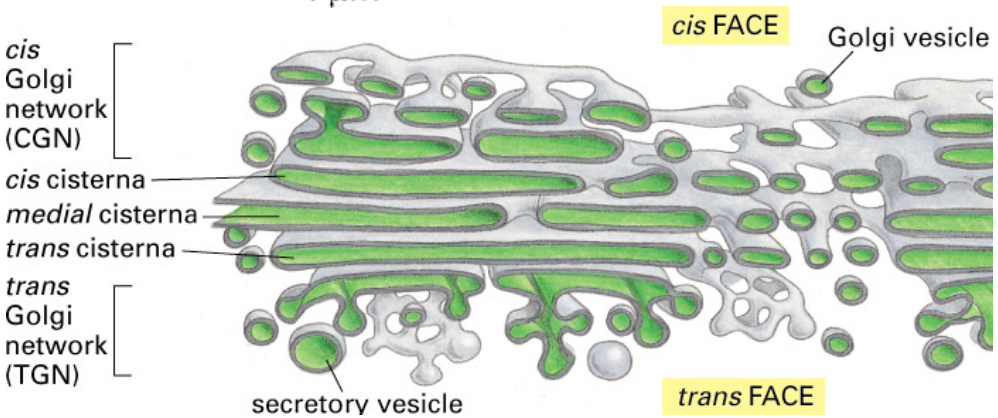
GPI –anchored proteins are accumulated at lipid raft.

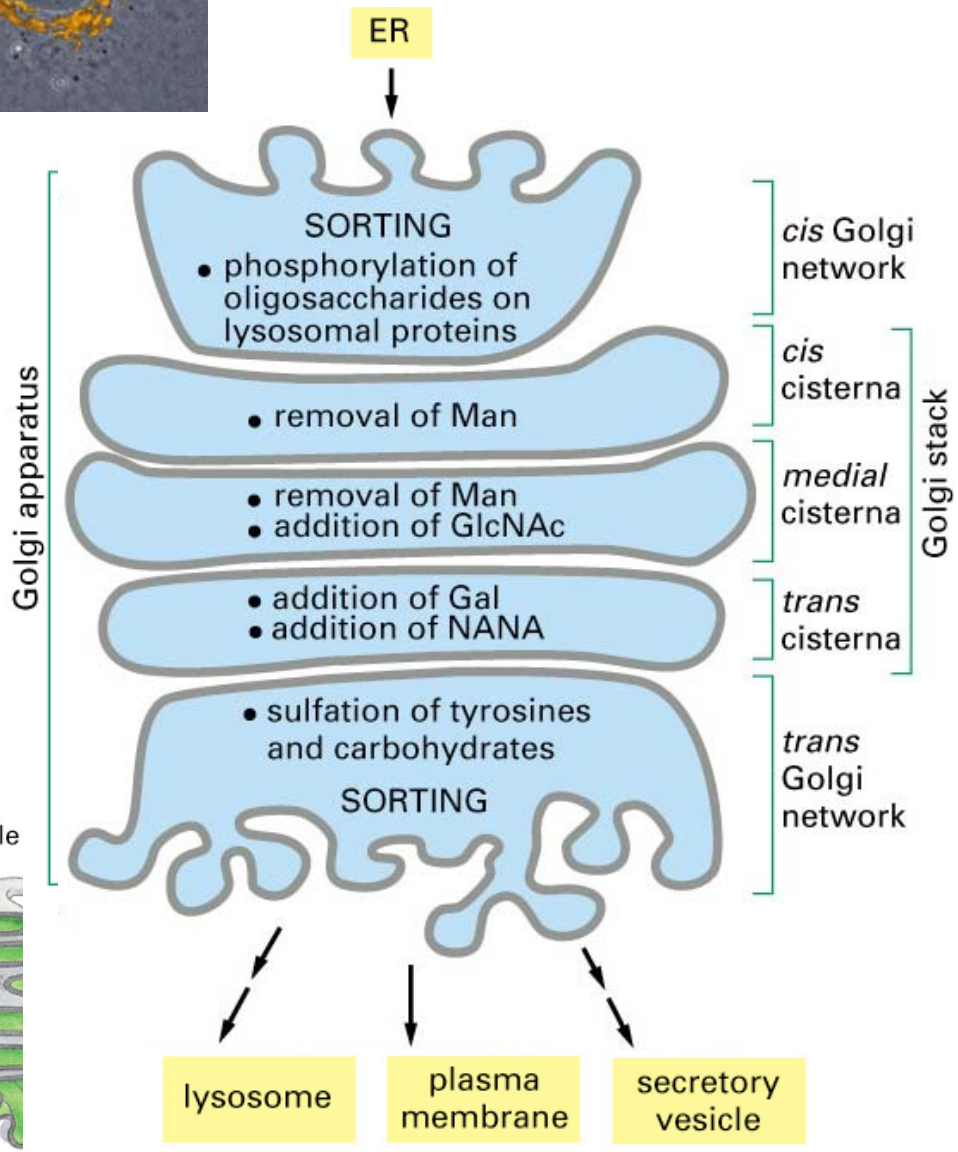
Proteins synthesized at rER undergo maturation at Golgi apparatus



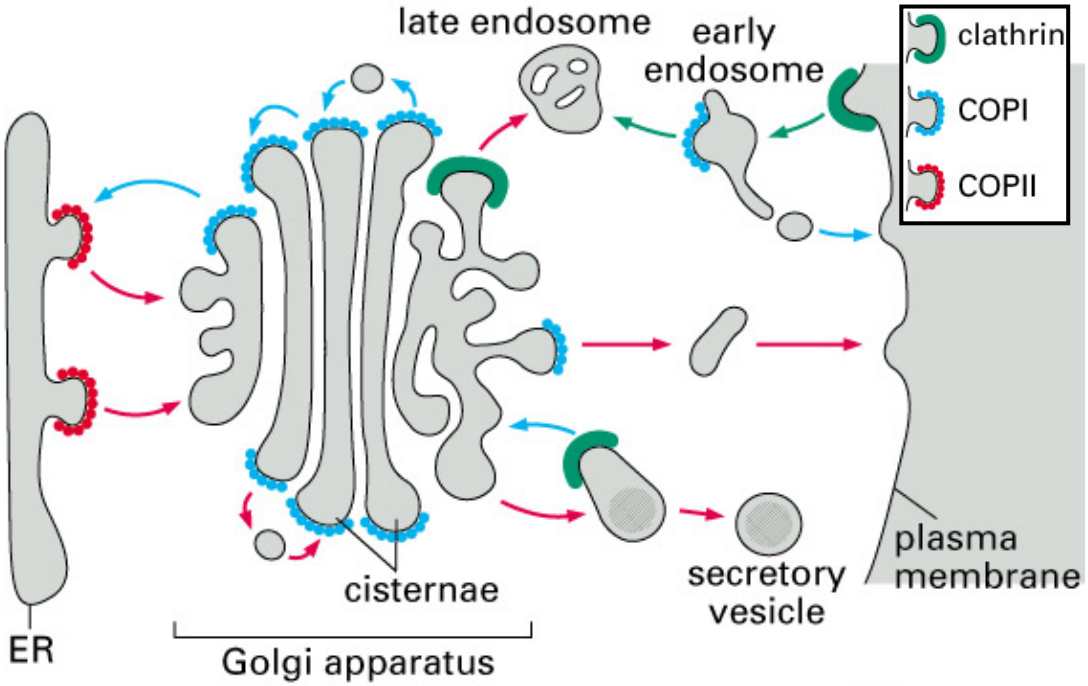


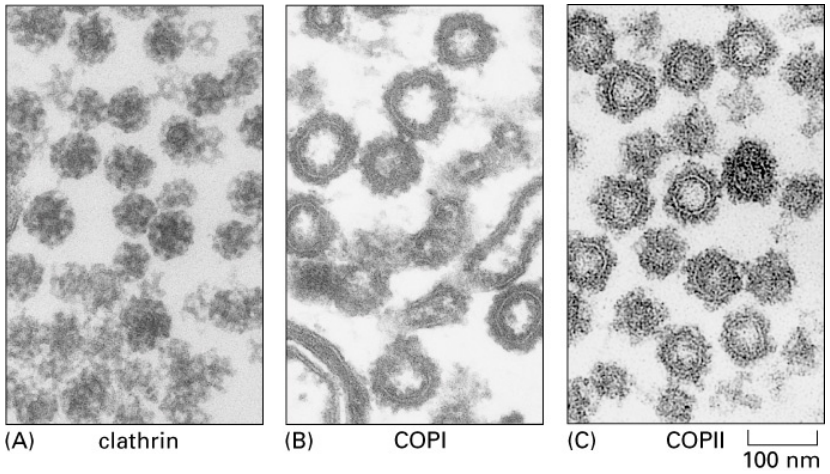






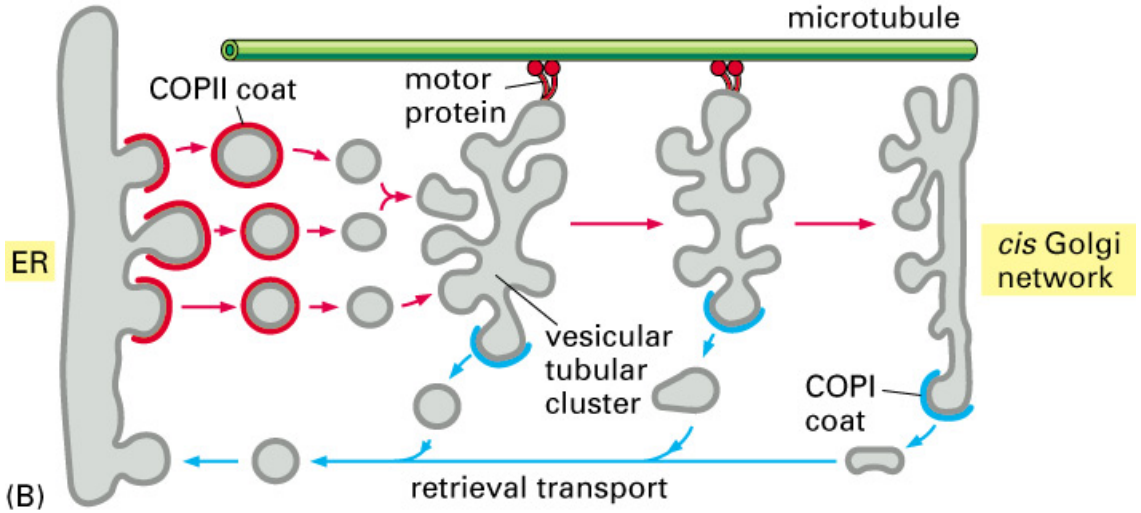
Coated vesicles transfer contents b/w ER, Golgi, and other membranes





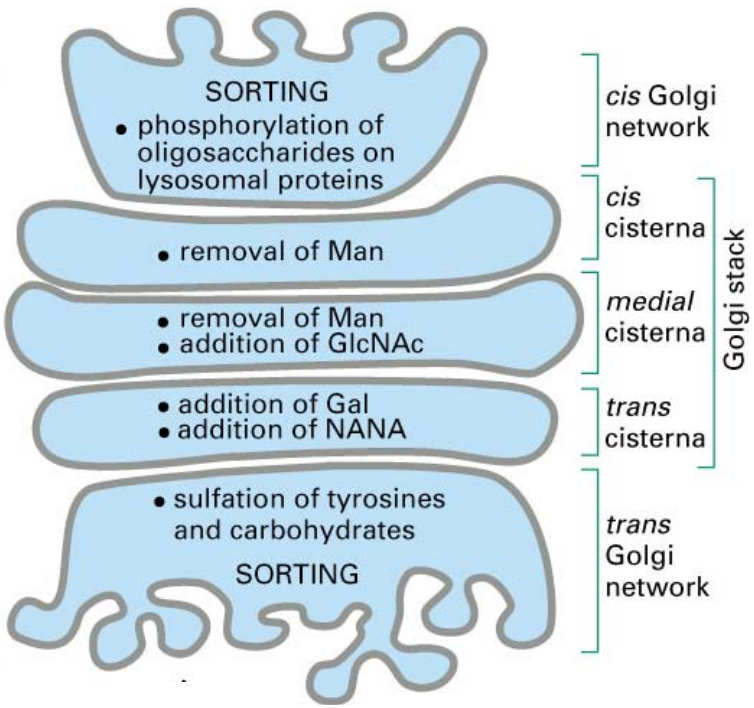
Only correctly folded proteins are transferred from rER to Golgi.

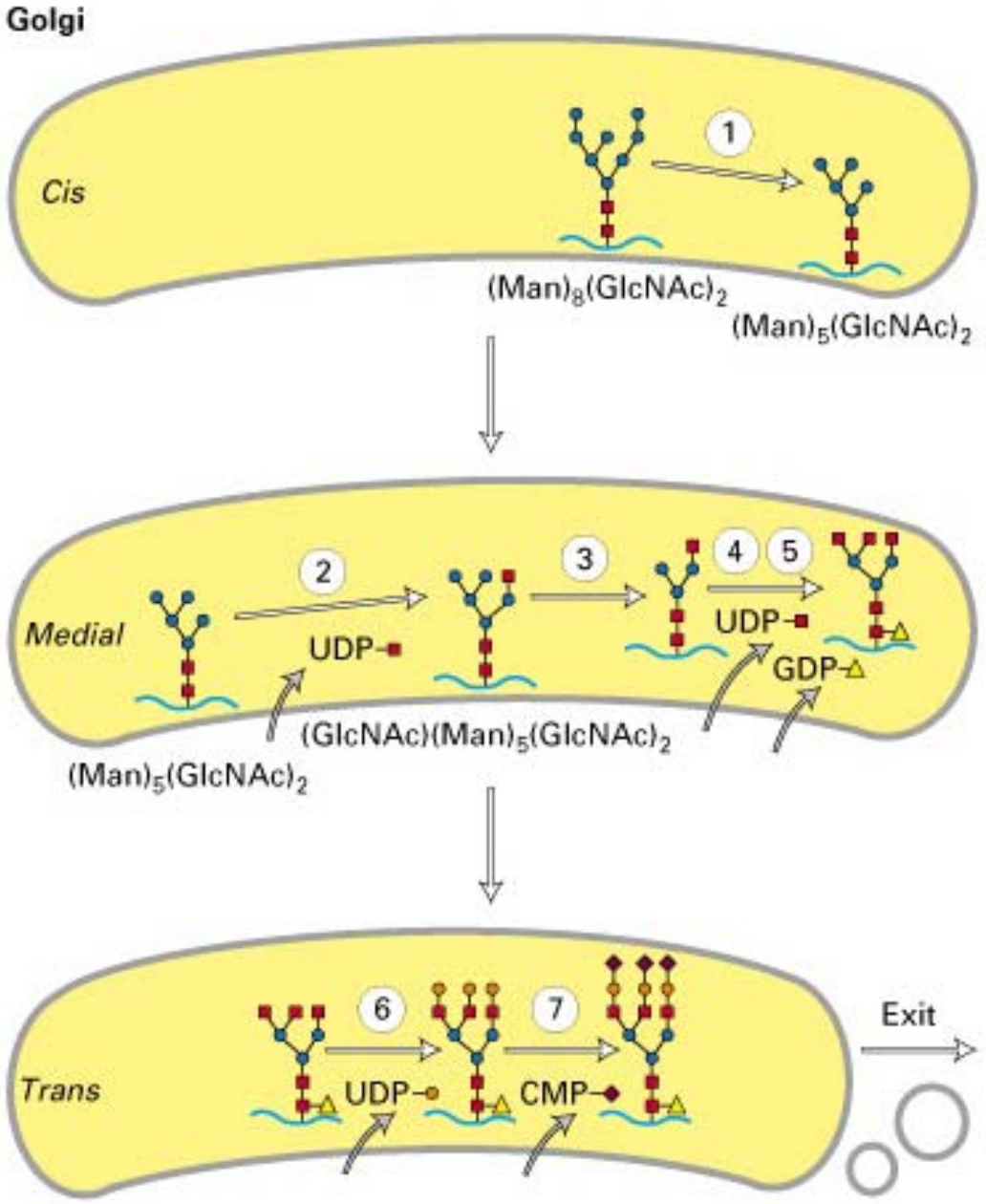
(rER) Protein segregation → vesicle formation (w/COPII) → coat removal → vesicle fusion to form Vesicular Tubular Cluster



Modifications to *N*-linked oligosaccharides are completed in the Golgi complex

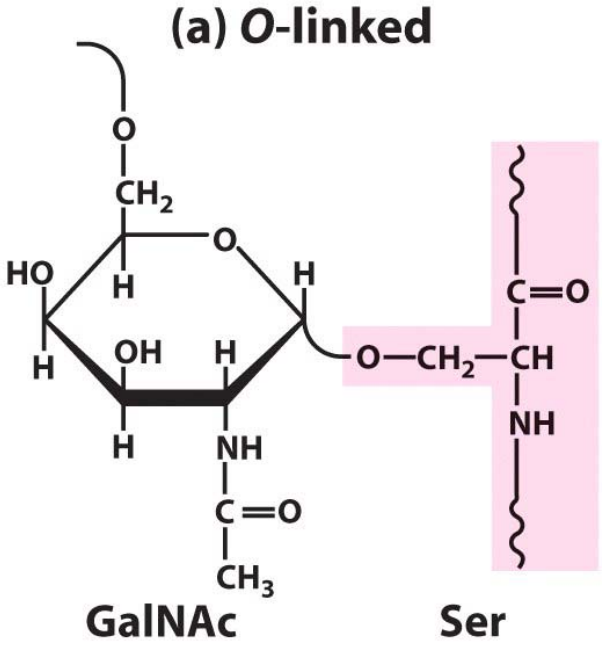
Oligosaccharides may promote folding and stability of glycoproteins

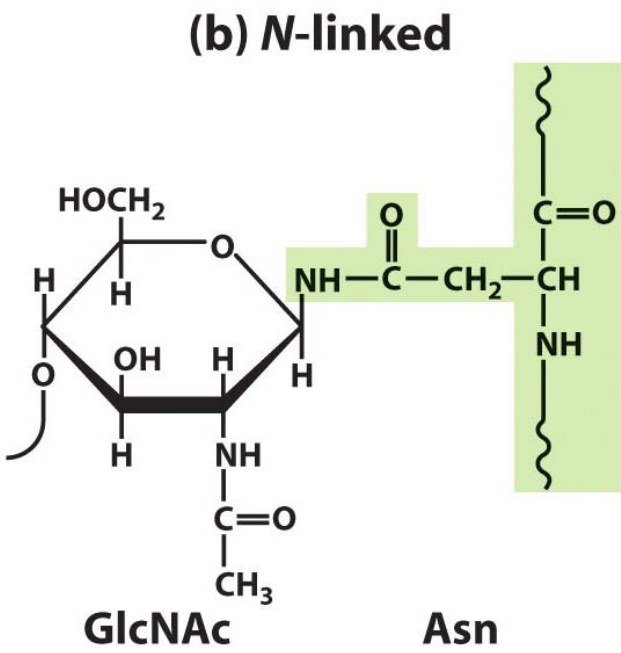


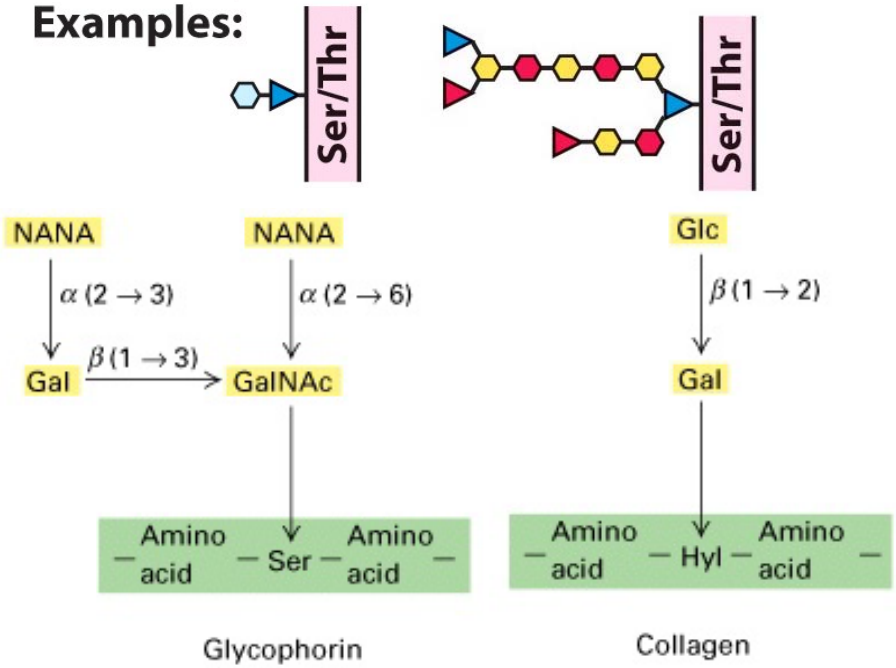


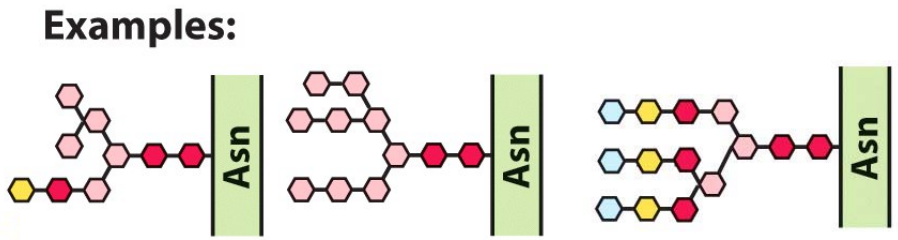
■ = *N*-acetylglucosamine
● = Mannose
▲ = Glucose
● = Galactose
△ = Fucose
◆ = *N*-acetylneuraminic acid

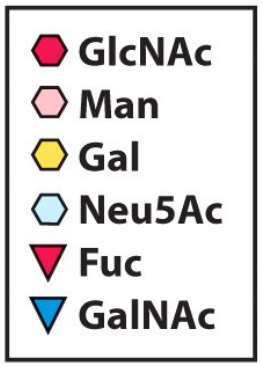
Some proteins also modified w/ *O*-linked oligosaccharide.



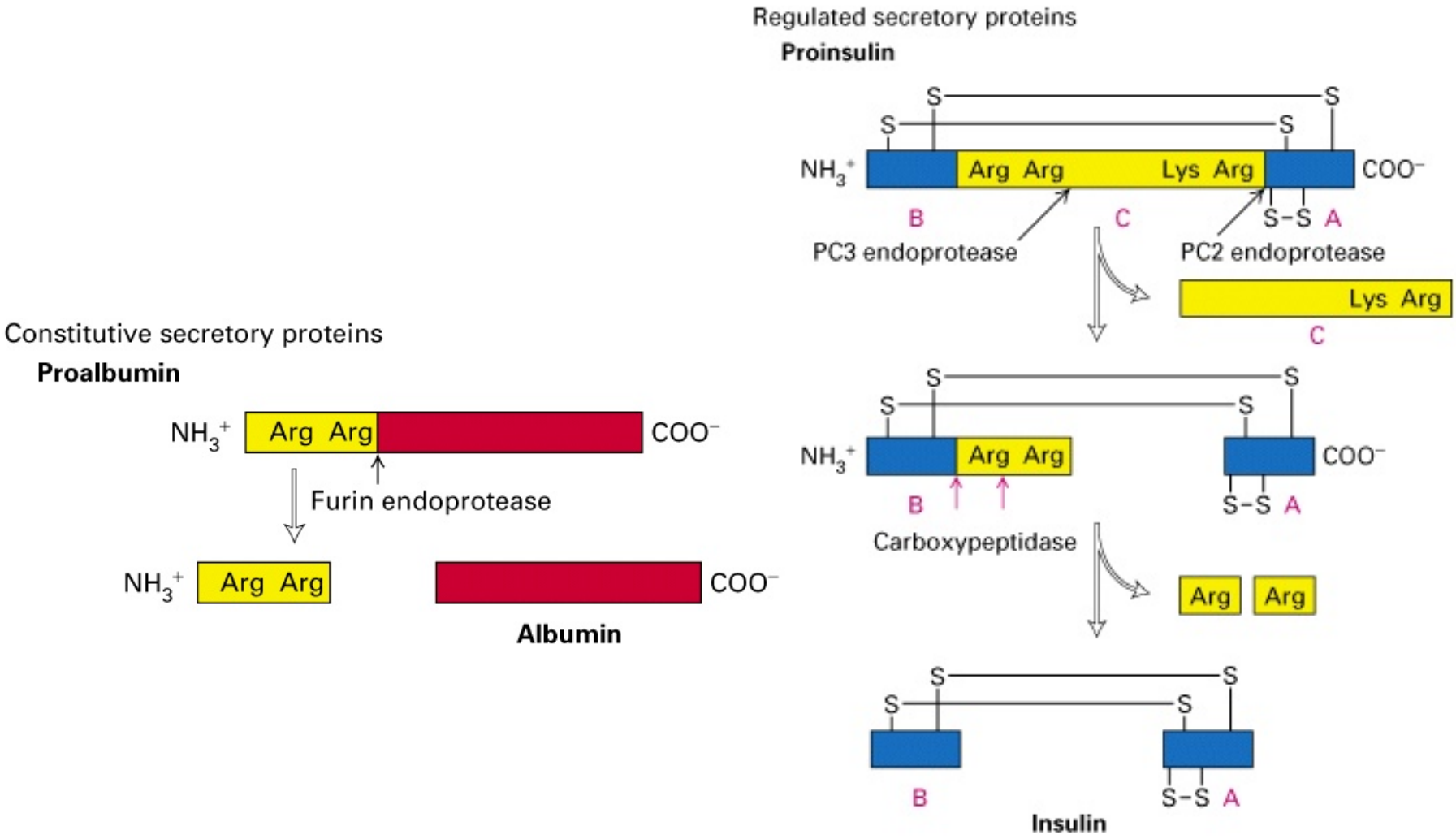




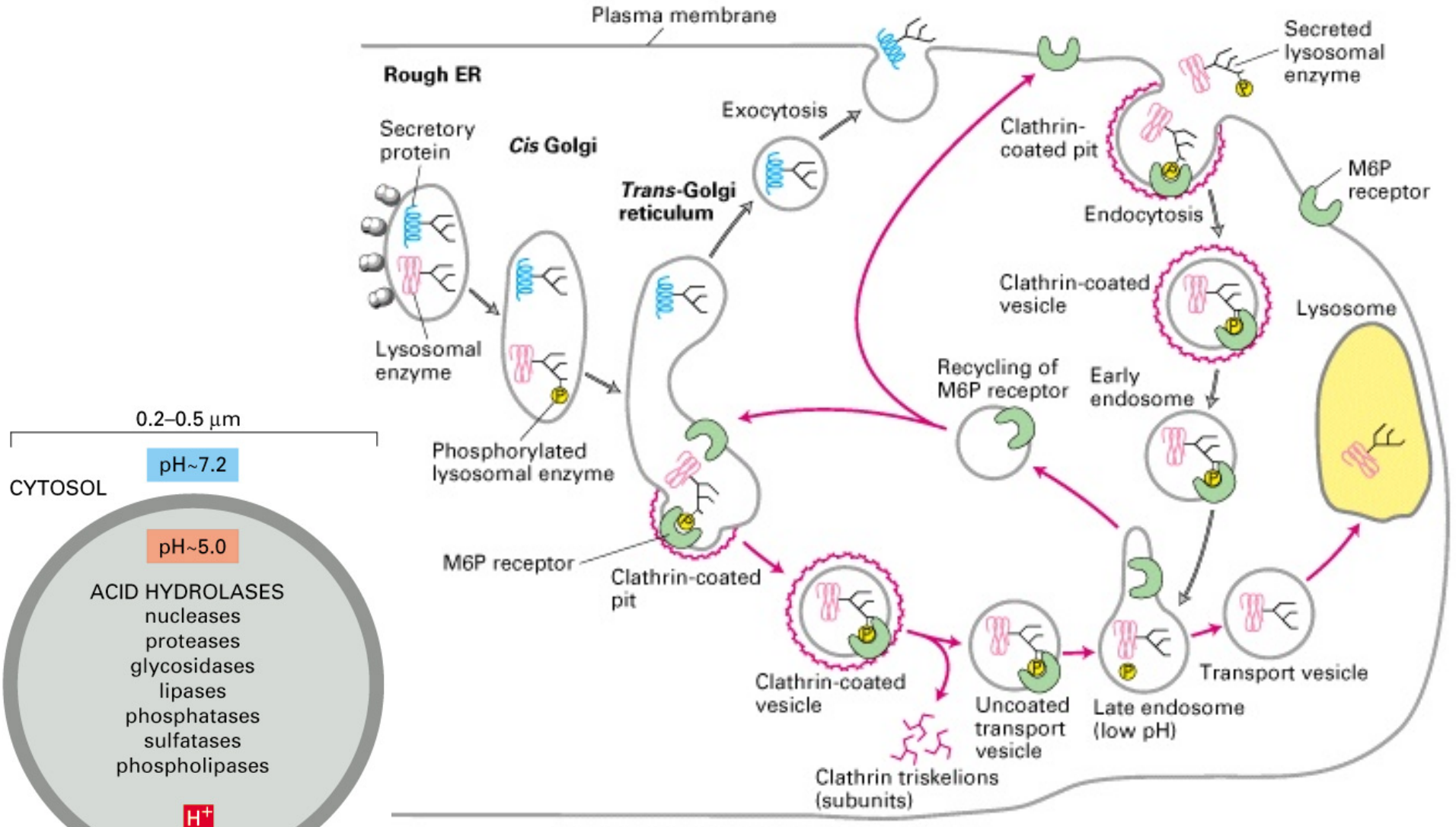




Proteolytic processing in maturation

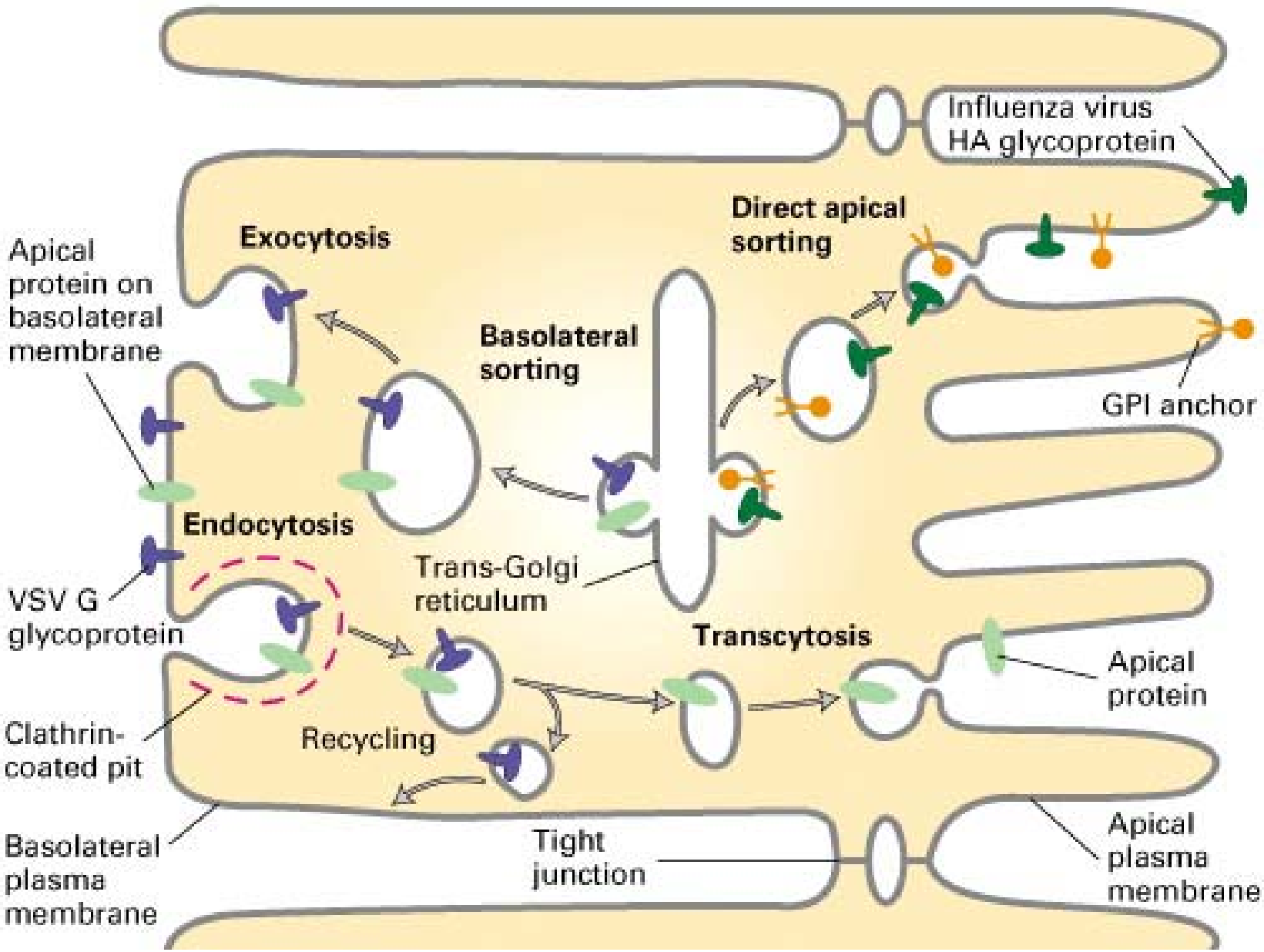


Mannose 6-phosphate residues : lysosomal Target Signal

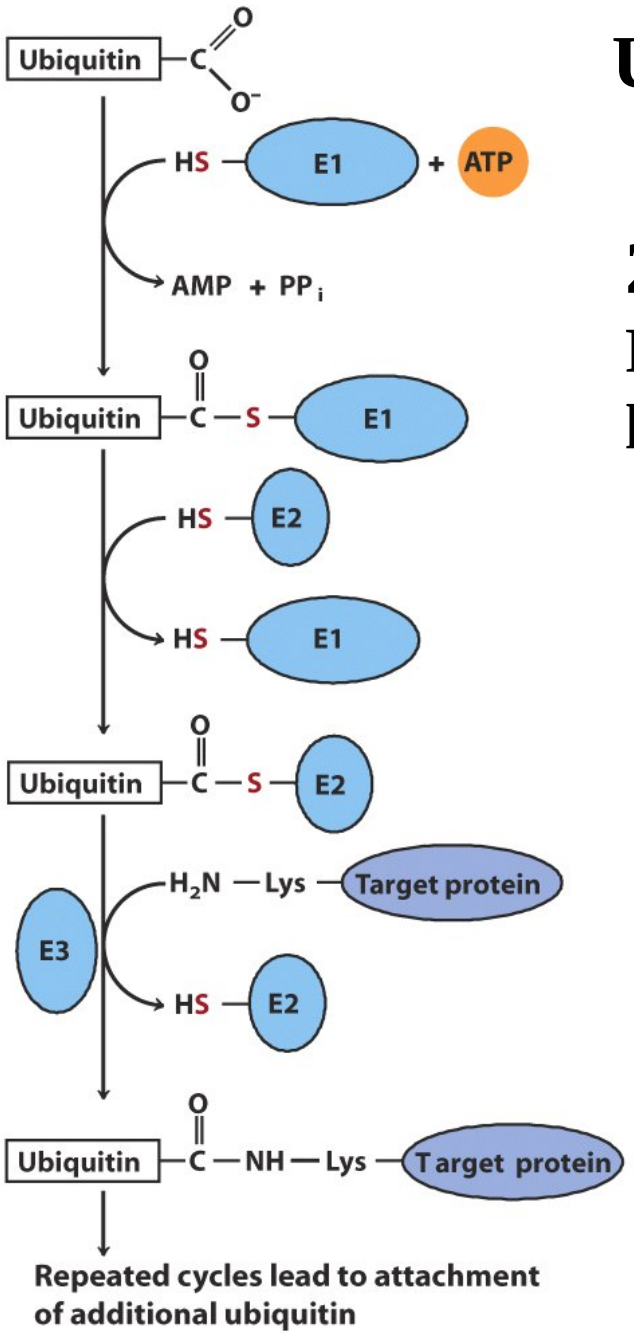


lysosome

Some proteins are sorted from the Golgi complex to the apical or basolateral plasma membrane



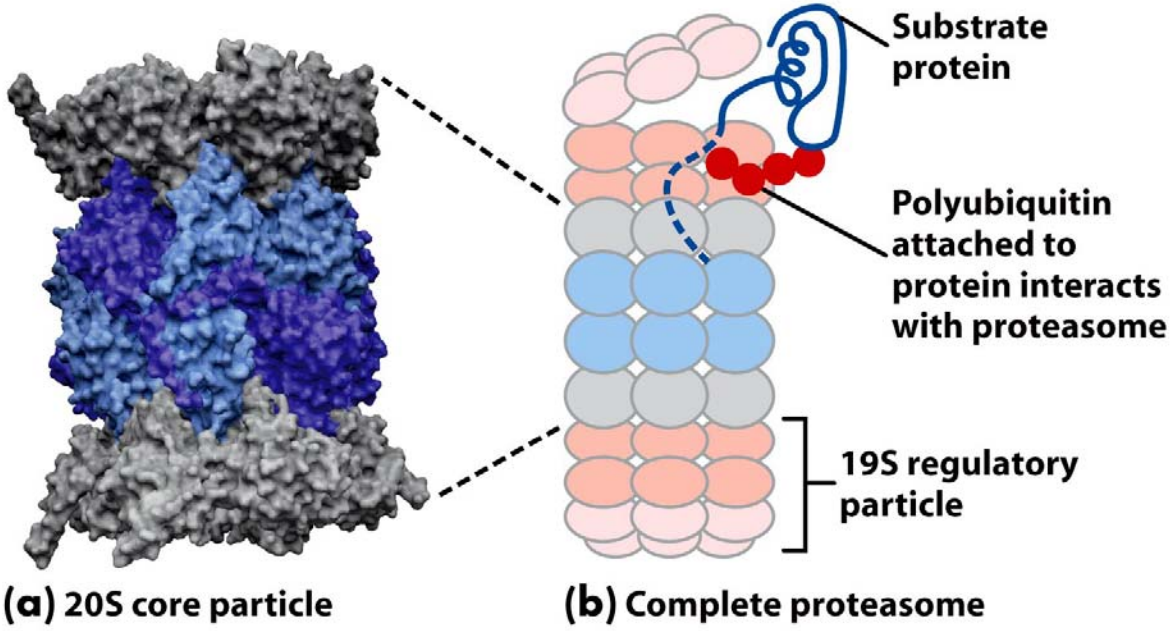
Protein Degradation Mechanism



Ubiquitin : Highly conserved small protein (76 AAs)

26S Proteosome

Protein degradation copmlex located within cytoplasm



Protein lifespan

No convincing generalized theory exist. (or there is no such rules)

(In general), isolated, highly purified proteins are more likely unstable than those in assembled complex.

In the native environment, there must be many molecules (from other proteins to ions, smaller organic molecules etc) that contribute protein stabilities.

Note: We (microtubule biochemists) know that tubulin (which is very unstable as unassembled state; requires molecular chaperon for correct folding) molecules can be repaired to correct its shape by chaperon.

TABLE 27–9 Relationship between Protein Half-Life and Amino-Terminal Amino Acid Residue

Amino-terminal residue	*Half-life**
Stabilizing	
Met, Gly, Ala, Ser, Thr, Val	>20 h
Destabilizing	
Ile, Gln	~30 min
Tyr, Glu	~10 min
Pro	~7 min
Leu, Phe, Asp, Lys	~3 min
Arg	~2 min

*Half-lives were measured in yeast for the β-galactosidase protein modified so that in each experiment it had a different amino-terminal residue. (See Chapter 9 for a discussion of techniques used to engineer proteins with altered amino acid sequences.) Half-lives may vary for different proteins and in different organisms, but this general pattern appears to hold for all organisms.