RNA translation - proteosynthesis

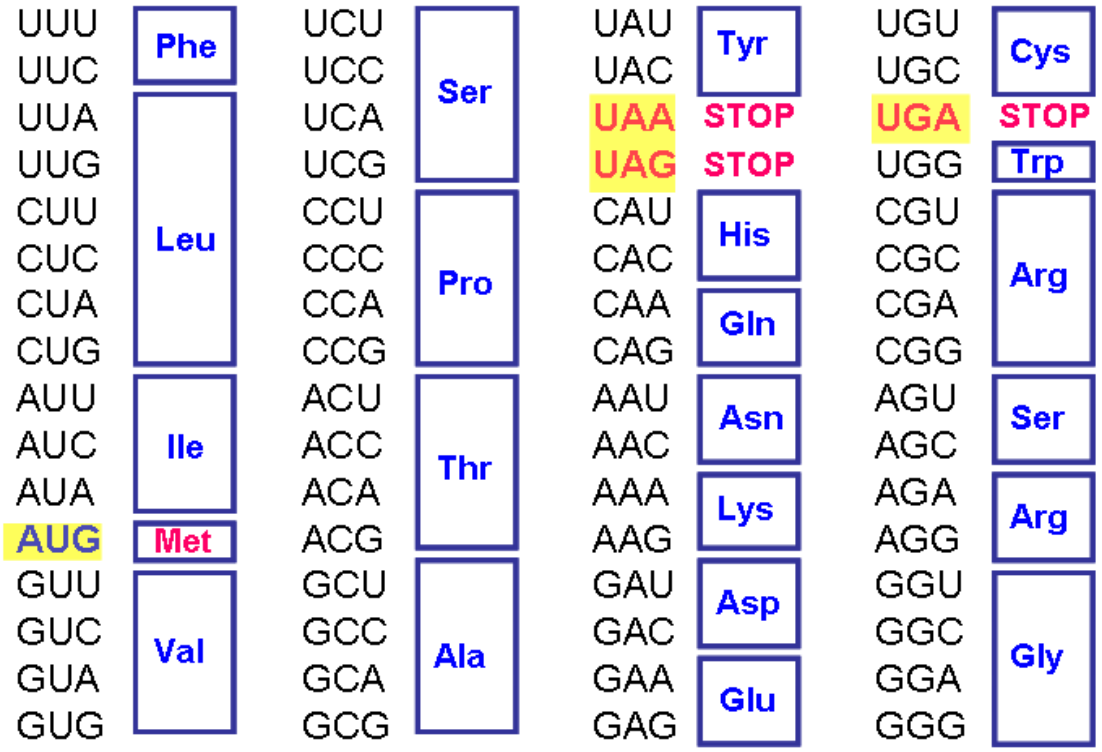
Posttranslational modifications of proteins
Glycoprotein synthesis

Biochemistry
Lecture 14

2008 (J.S.)

The genetic code

Codons in mRNA (read in the direction 5´→ 3´):



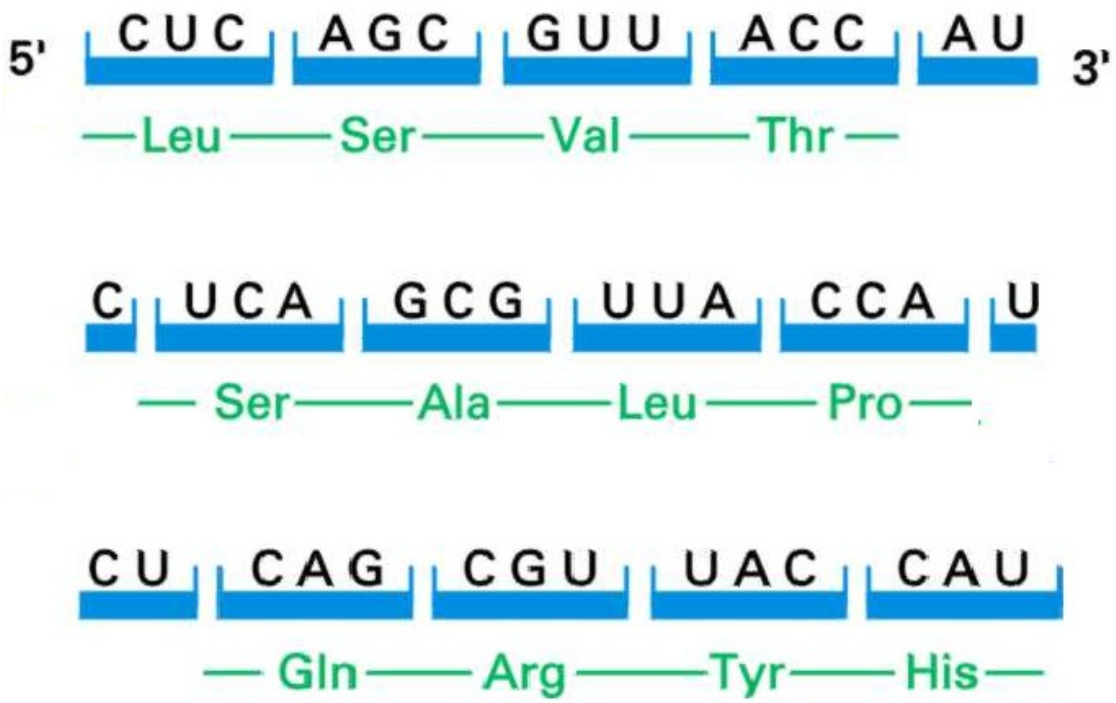
– **Triplets of bases**,
– **unambiguous code**,
– **degeneracy of the code**,
– **non-overlapping**,
– **in a continuous sequence**.

Degeneracy of the genetic code

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

The reading frames

Only one of the three possible frames is the right one,
it begins at the start codon AUG recognized by the Met-tRNA^{Met}



Distinctive codons of human mitochondria

Codon	Standard code	Mitochondrial code
UGA	Stop	Trp
UGG	Trp	Trp
AUA	Ile	Met
AUG	Met	Met
AGA	Arg	Stop
AGG	Arg	Stop

Consequences of mutations

Mutations are structural alterations in the genome that can lead to an alteration in the sequence of a protein translated from it, if they are transcribed into mRNA.

Types of mutations:

– **point mutations,** alterations in a single base, which

- can result in change of the sense (a single amino acid replacement, e.g. GCA→CCA results in Arg→Pro),
- can be neutral, causing no change (e.g. CGA→CGG, both encode Arg),
- can terminate translation, if result in stop-codons (nonsense).

– **insertions** of one or more nucleotides, and

– **deletions** of one or more nucleotides.

The result depends on the number of inserted or deleted nucleotides.

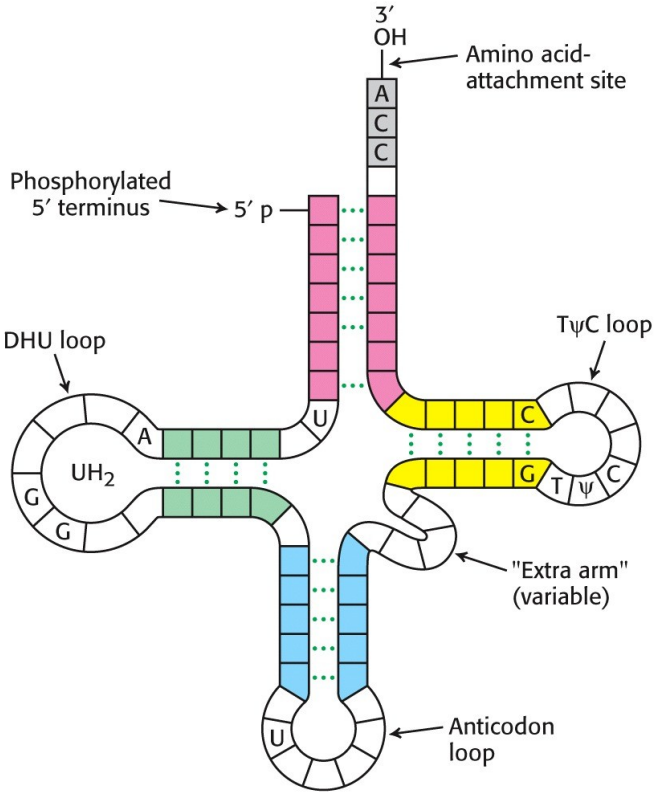
If three nucleotides (or more triplets) are inserted/deleted without a change of the reading frame, polypeptides with inserted/deleted amino acyl residues will be synthesized.

If one or two nucleotides are inserted/deleted, the result is a "frame-shift mutation" that gives nonsense codons, distinct primary structure of proteins, etc.

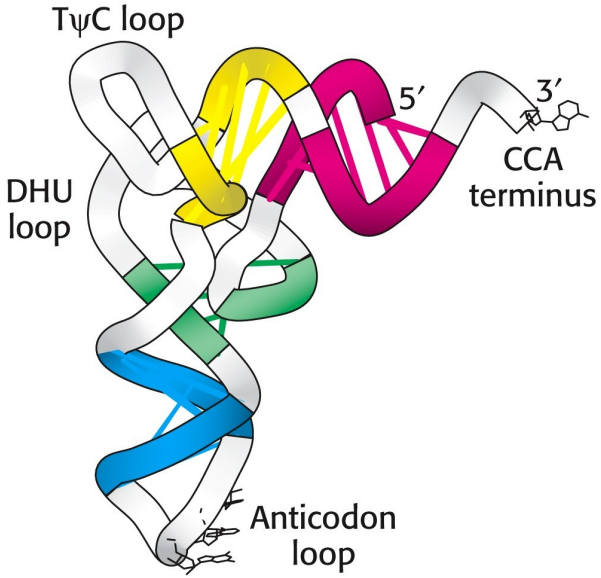
tRNA molecules act as adaptors in translation,

amino acids cannot react directly with bases of nucleotides.

General structure of tRNA molecules

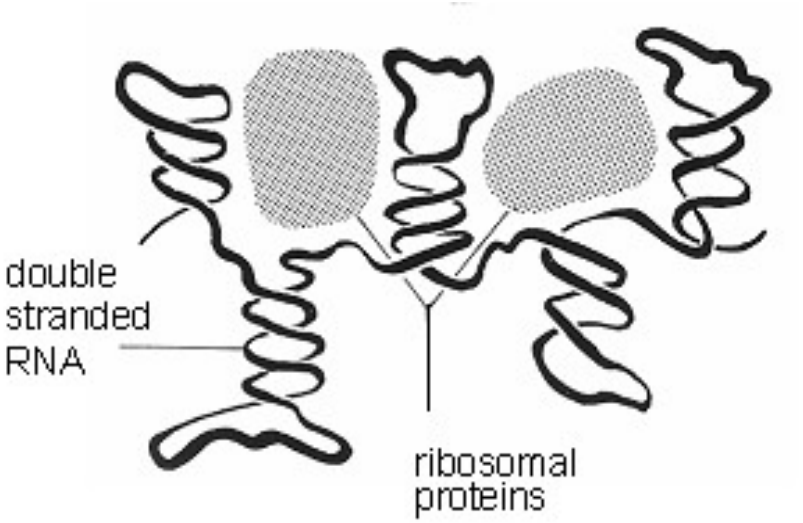


Helix stacking in L-shaped tRNA structure

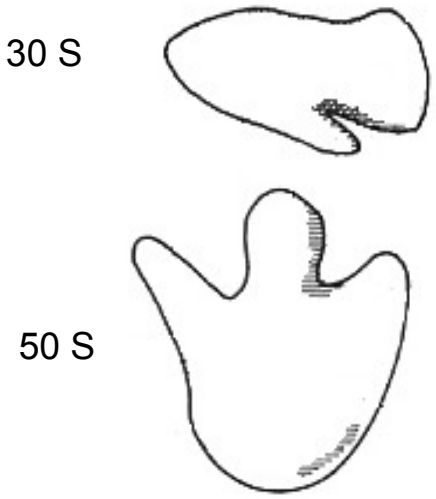


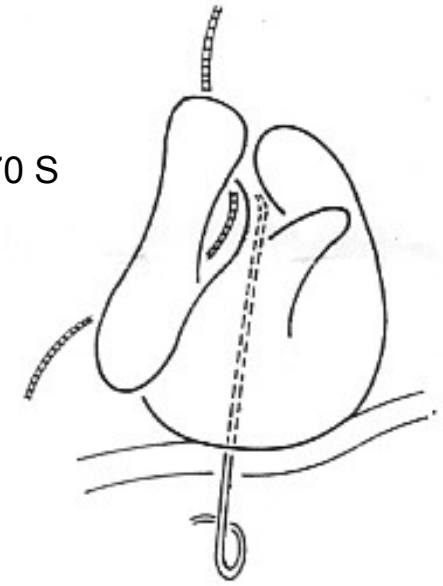
Ribosomes

Schematic representation of the arrangement of ribosomal RNA and ribosomal protein within the ribosome.

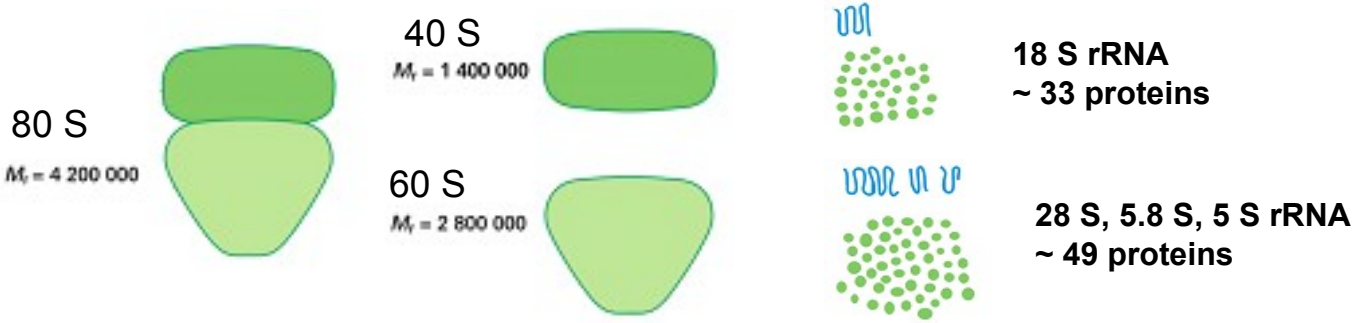


Ribosome of the ***E. coli***



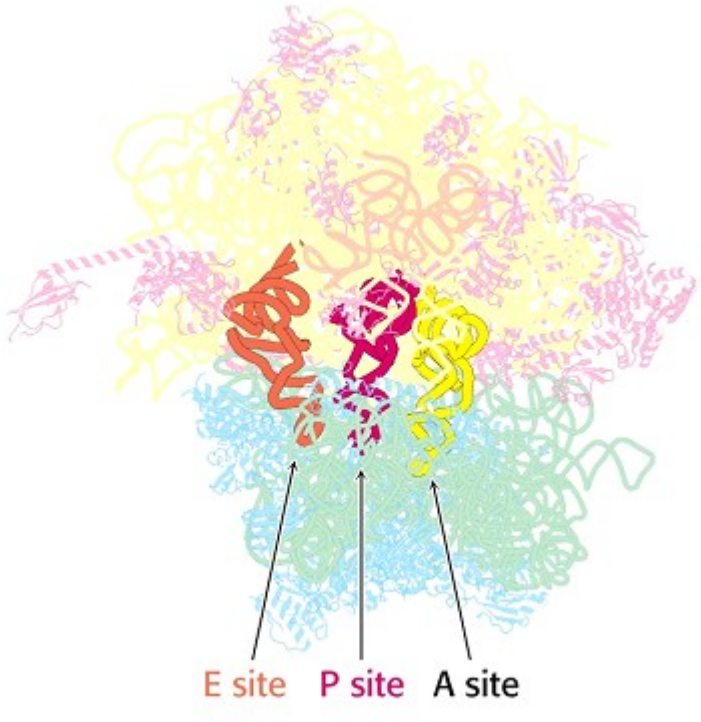


Eukaryotic ribosome

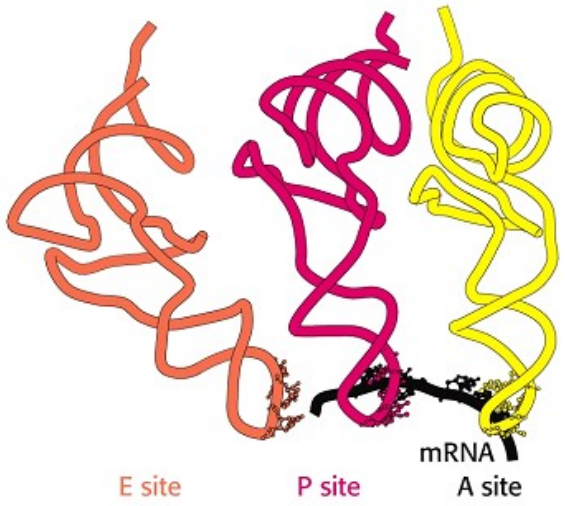


	SMALL subunit	**LARGE** subunit
Sedimentation constant	**40 S**	**60 S**
rRNA types	**18 S** (1874 bases)	**28 S** (4718 bases) **5.8 S** (160 bases) **5 S** (120 bases)
Proteins	$S_1 – S_{33}$	$L_1 – L_{49}$
RNA mass fraction	0.50	0.65

Three binding sites for tRNA

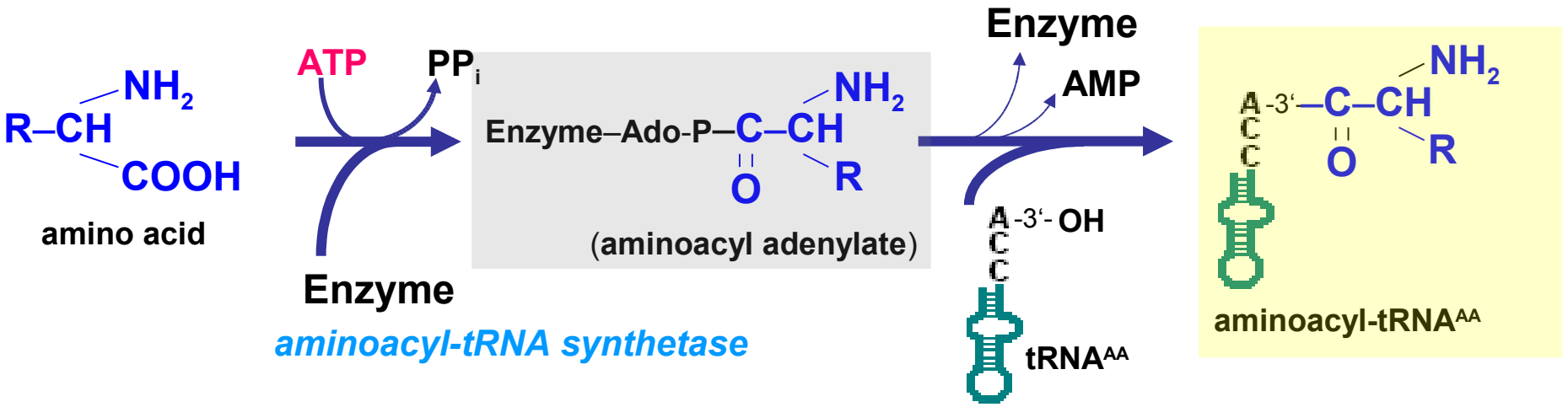


E **exit site**
P **peptide site**
A **amino acid site**



Activation of amino acids

Synthesis of aminoacyl-tRNAs ("charging" of tRNAs)



Aminoacyl-tRNA synthetases (at least 20 distinct enzymes in cells) exhibit the **very high degree of specifity** for amino acids.

The enzyme molecule recognizes both a specific amino acid and a specific tRNA. These enzymes discriminate accurately, the overall rate of occurence of errors in translating mRNA is less than 1 in 10 000.

This high specifity is oft called **the 2nd genetic code**.
It depends on specific location of some bases in tRNA molecules, not on the sole anticodon.

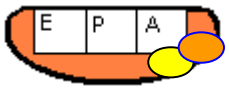
The initiation of protein synthesis

Formation of the 80 S initiation complex

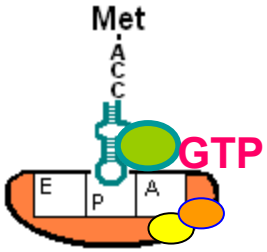
eIFs – eukaryotic initiation factors

Charged Met-tRNA^{Met} binds eIF2 and GTP

Small ribosomal subunit with eIF1A and eIF3 is attached to Met-tRNA

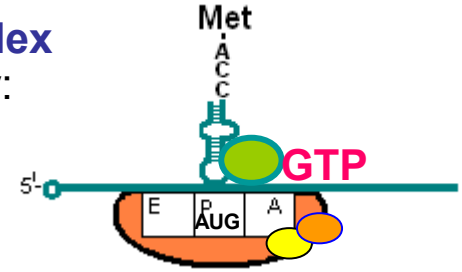


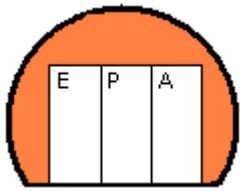
→ **Preinitiation complex**

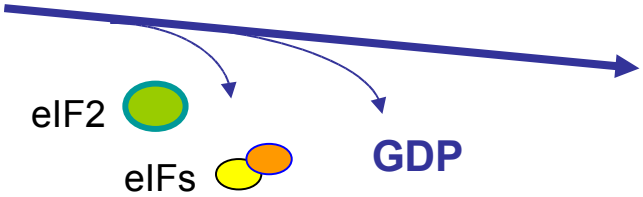


5´-m^{7}Gppp cap of mRNA binds CBP (cap-binding protein) and several eIFs; mRNA binds to preinitiation complex which scans mRNA till reaching the start codon AUG.

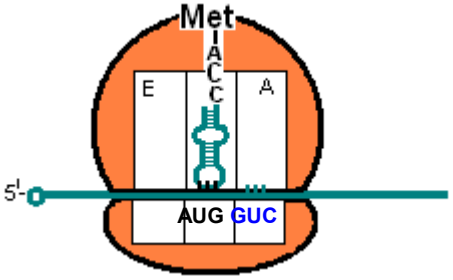
40S initiation complex is formed in this way:





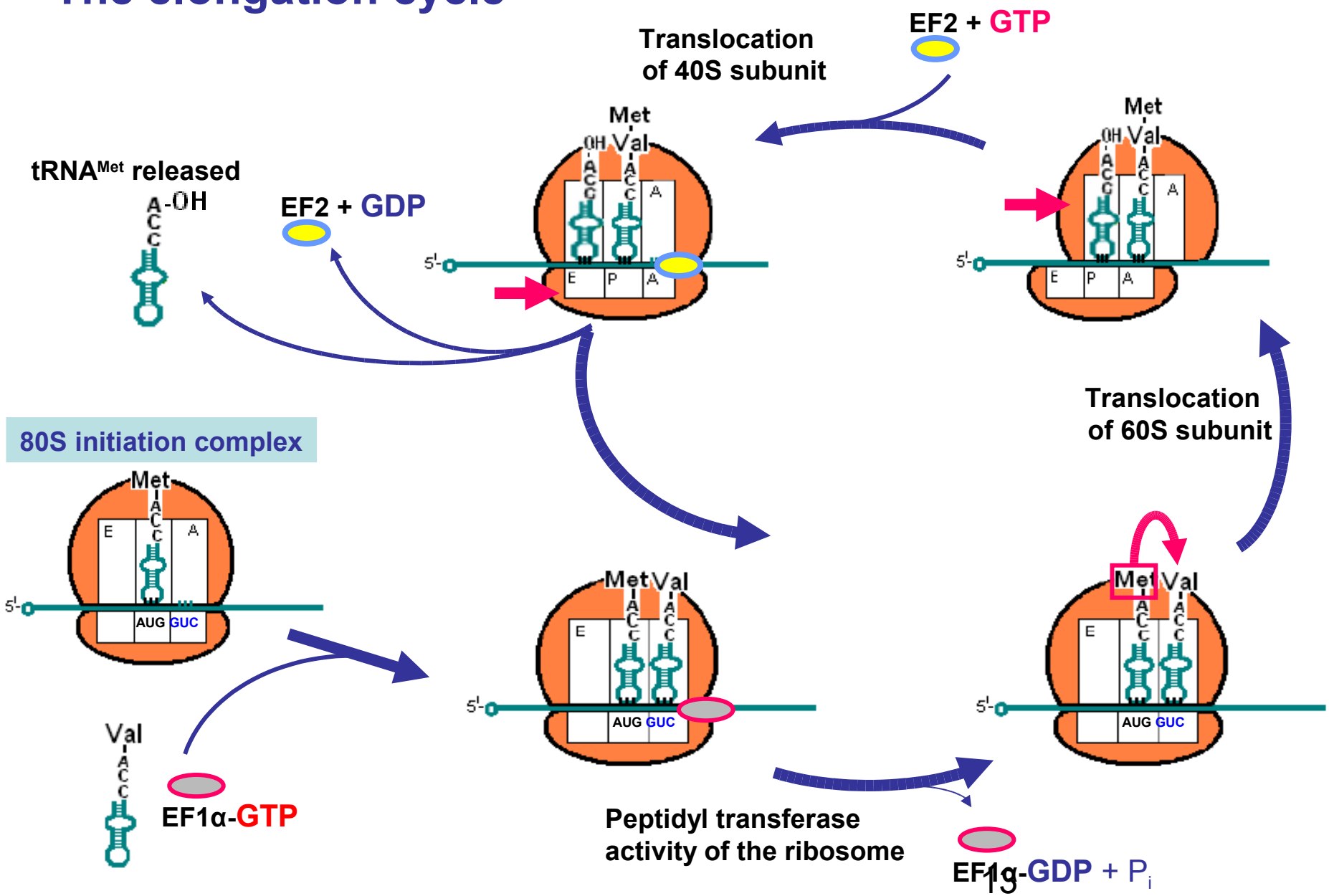


GTP is hydrolyzed, GDP, eIF2, eIFs released and 60S subunit attached giving so 80S initiation complex:.



80S initiation complex

The elongation cycle

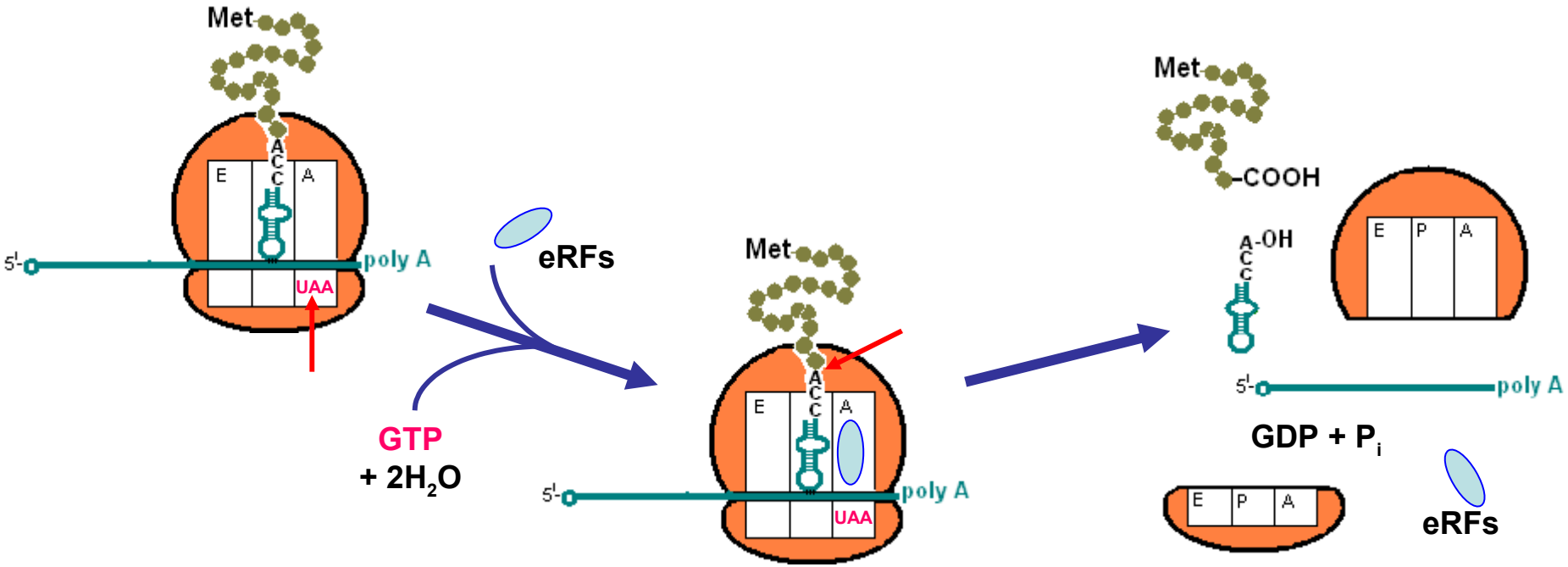


The termination of protein synthesis

Elongation continues as far as the binding site A reaches a **stop codon**.

Releasing factors (eRFs) bind to the site A and ***peptidyl transferase*** catalyzes the hydrolysis of the ester bond between the polypeptide and tRNA.

Ribosomal subunits dissociate, mRNA, "uncharged" tRNA, eRFs and GDP are released.



Energy required in protein synthesis

		ATP equivalent
Origin of aminoacyl-tRNA	ATP → AMP + 2 Pi	2
Binding of aminoacyl-tRNA to site A	GTP → GDP + Pi	1
Translocation of ribosome	GTP → GDP + Pi	1
Formation of one peptide bond		**4 ATP**

Velocity of ribosomal protein synthesis

In prokaryotes, approximately hundreds of peptide bonds are formed per second.

In eukaryotes, approximately 1 – 2 peptide bonds per second.

Folding of proteins

A nascent polypeptide chain grows out from the larger ribosomal subunit (about 40 AA residues are hidden in the subunit) following a leader sequence – a signal peptide and the folding of the native three-dimensional conformation may begin.

The details of how proteins make the transitions from unfolded structures to unique conformations in the native forms have not been fully elucidated.

For even a small protein, it would take much too long to fold properly by randomly trying out all possible conformations.

Proteins follow a partly defined folding pathway consisting of **intermediates** between the fully unfolded protein (polypeptide chain) and its native conformation. There is the **tendency to retain partly correct intermediates**, which can interact with one other, leading to increased stabilization (the nucleation-condensation model).

Folding of proteins is supported by **molecular chaperones** that stabilize unfolded or partially folded intermediates, allowing them time to fold properly, and prevent so faulty folding and inappropriate interactions (formation of non-functional structures).

Many of chaperons are so-called "heat shock proteins" (inducible by elevated temperature and various chemicals – conditions that cause unfolding of newly synthesized proteins).

Free ribosomes synthesize proteins that remain within the cell, either within the cytoplasm, or directed to organelles bounded by a double membrane - to the nucleus and mitochondria.

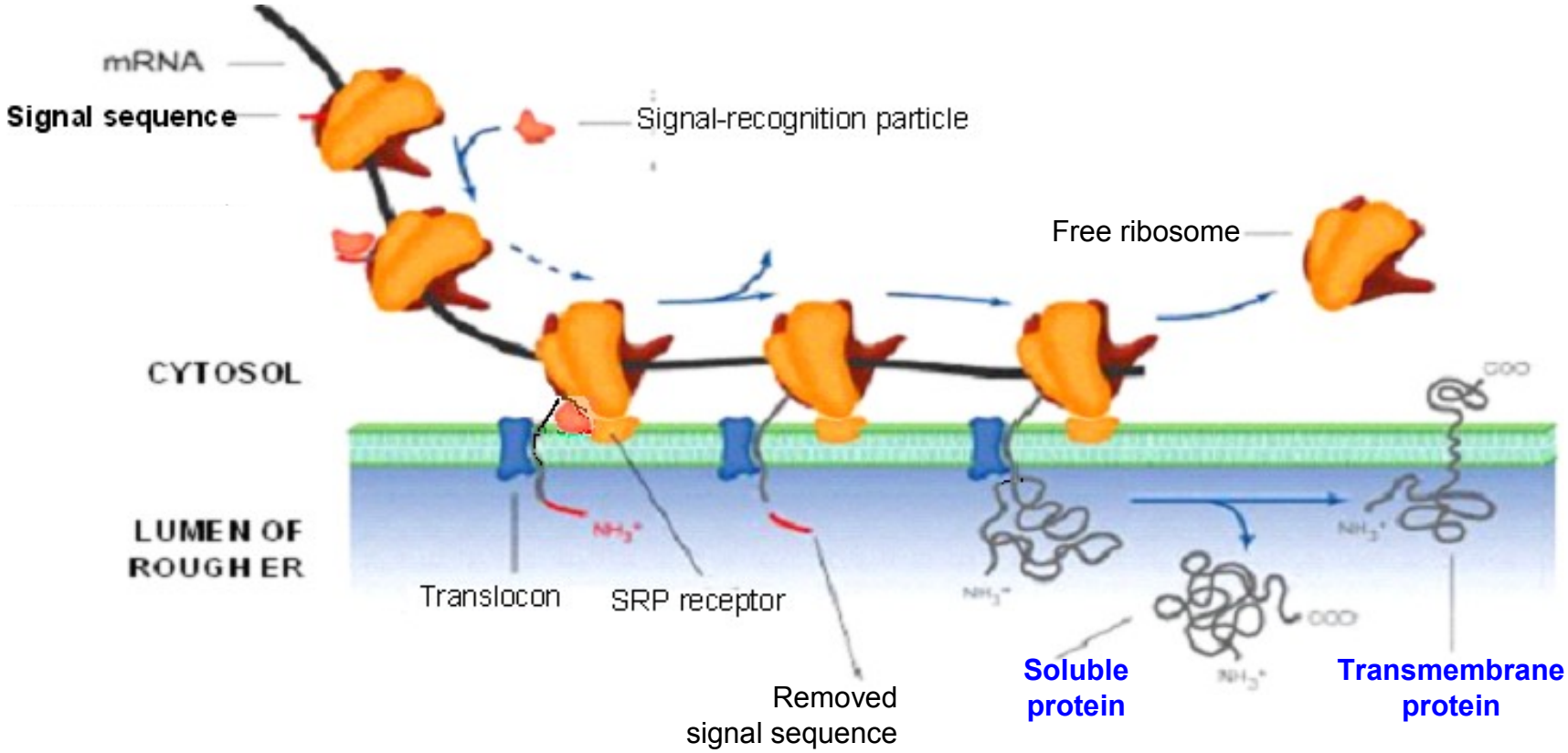
The synthesis of all proteins begins on free ribosomes in the cytoplasm. Protein synthesis continues on free ribosomes in the cytoplasm till an entire **signal sequence** of the nascent chain that directs the ribosome to the endoplasmic reticulum appears in the cytoplasm.

Ribosomes bound to the ER usually synthesize proteins destined to leave the cell (secretory proteins), lysosomal proteins and proteins of the plasma membrane.

Sorting of proteins

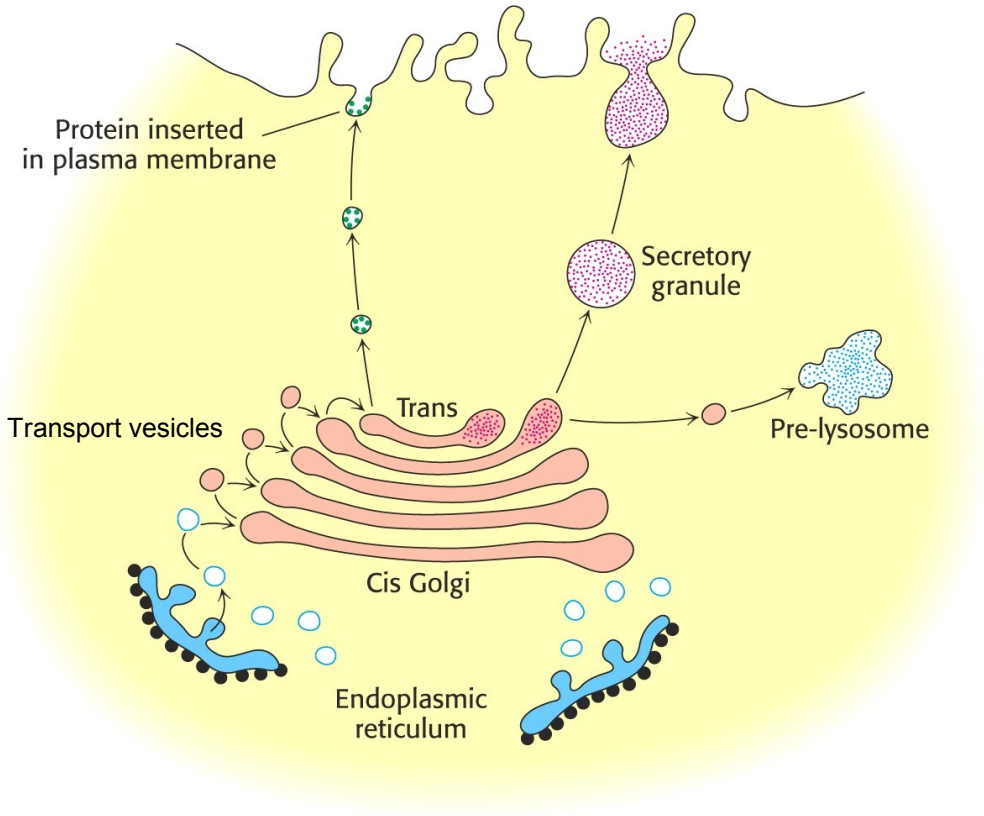
A fundamental component of the sorting system are the **signal sequences** (not always specific) at the N-end of the growing polypeptide chains.
These targeting sequences direct proteins (preproteins) to their destinations and are subsequently removed.

Transport of proteins into the endoplasmic reticulum



Signal sequence – usually 15 – 30 hydrophobic AA at the N-end of the nascent chain
Signal-recognition particle (SRP) - six proteins and one RNA molecule (cca 300 nucleotides)
SRP receptor - SRP-R, a docking protein
Translocon - a multisubunit assembly, a protein conducting channel

Protein-sorting pathways



The mannose 6-P signal in the *N*-linked saccharidic component of glycoproteins serves as a chemical marker to target certain lysosomal enzymes to that organelle.

Integral proteins of the ER membrane involve at the carboxylate end the sequence Lys-Asp-Glu-Leu (KDEL).

In the membranes of transport vesicles, there are key membrane proteins called **SNARE proteins** (**s**oluble ***N***-ethylmaleimide-sensitive-factor **a**ttachment protein **re**ceptor) that help draw appropriate membranes together to initiate the fusion process.

These proteins **largely determine the compartment with which a vesicle will fuse**.

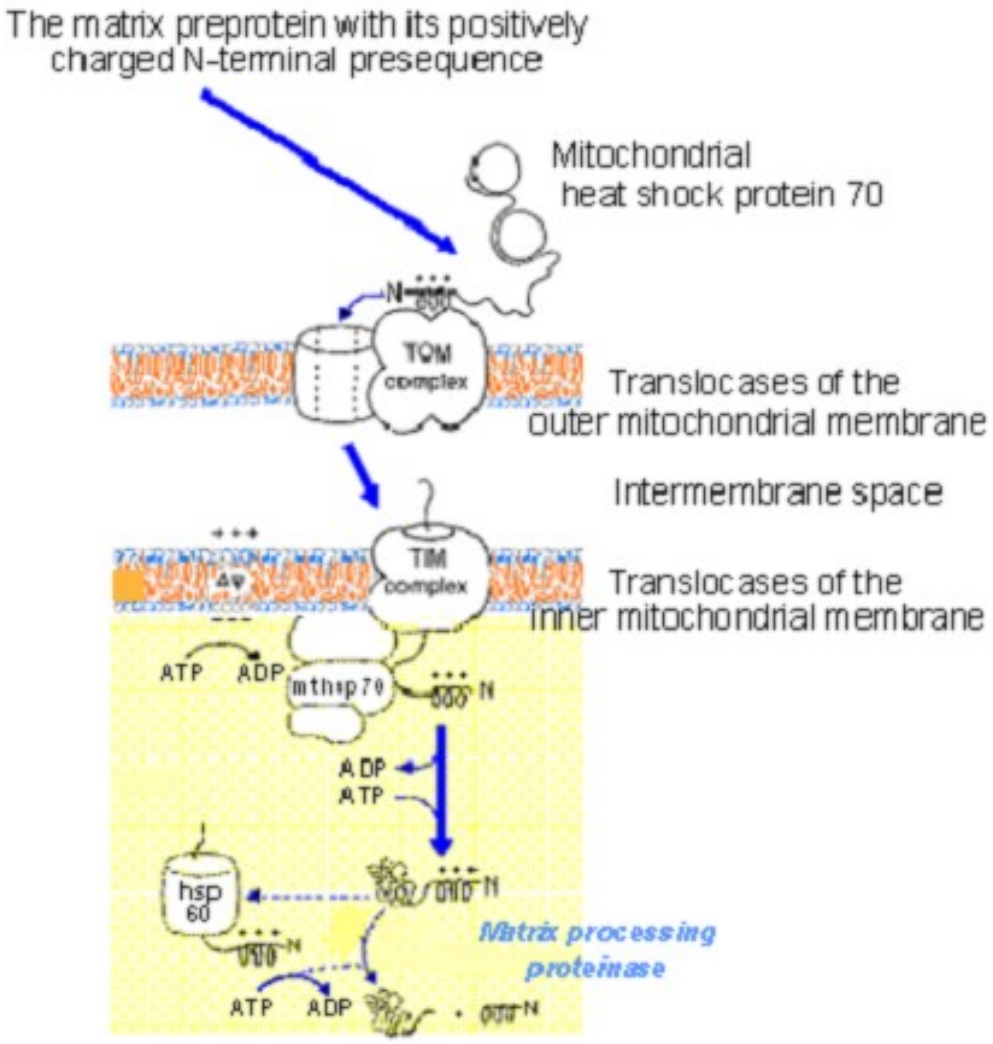
Targeting sequences

Target	Signal
Nucleus	–KKXK or –$(K/R)_2$–$X_{10–12}$–$(K/R)^*$
Peroxisome	–SKL–COO^-
Mitochondrion	*N*-terminal amphipathic helix
Endoplasmic reticulum	–KDEL–COO^-(ER retention)

*The "/" means that either K or R is required.

D = aspartate
E = glutamate
K = lysine
L = leucine
R = arginine
S = serine

Model for the import of nuclear encoded proteins into the mitochondrial matrix



Posttranslational modification of proteins

Many proteins are synthesized from the mRNA template as a precursor molecule, which then must be modified to achieve the active protein.

Hydrolytic clipping of proproteins (prohormones, zymogens, etc.)

Examples: The functional **insulin** molecule is formed by removal of the peptide C that connects the chains A and B in the proinsulin molecule,

inactive proteinases pepsinogen, trypsinogen, proelastase, etc. are activated by hydrolysis of small regions of the polypeptide chains,

proopiomelanocortin may be hydrolyzed to corticotropin, MSH sequences, endorphins and enkephalins.

Many **other posttranslational modifications** of proteins occur:

glycosylation in synthesis of glycoproteins,

hydroxylation of proline residues to hydroxyproline (collagen, elastin) or lysine residues to hydroxylysine (collagen),

γ-**carboxylation** of glutamate residues in some blood-clotting factors,

acetylation (of histones), **methylation**, **prenylation**, etc.

Synthesis of glycoproteins

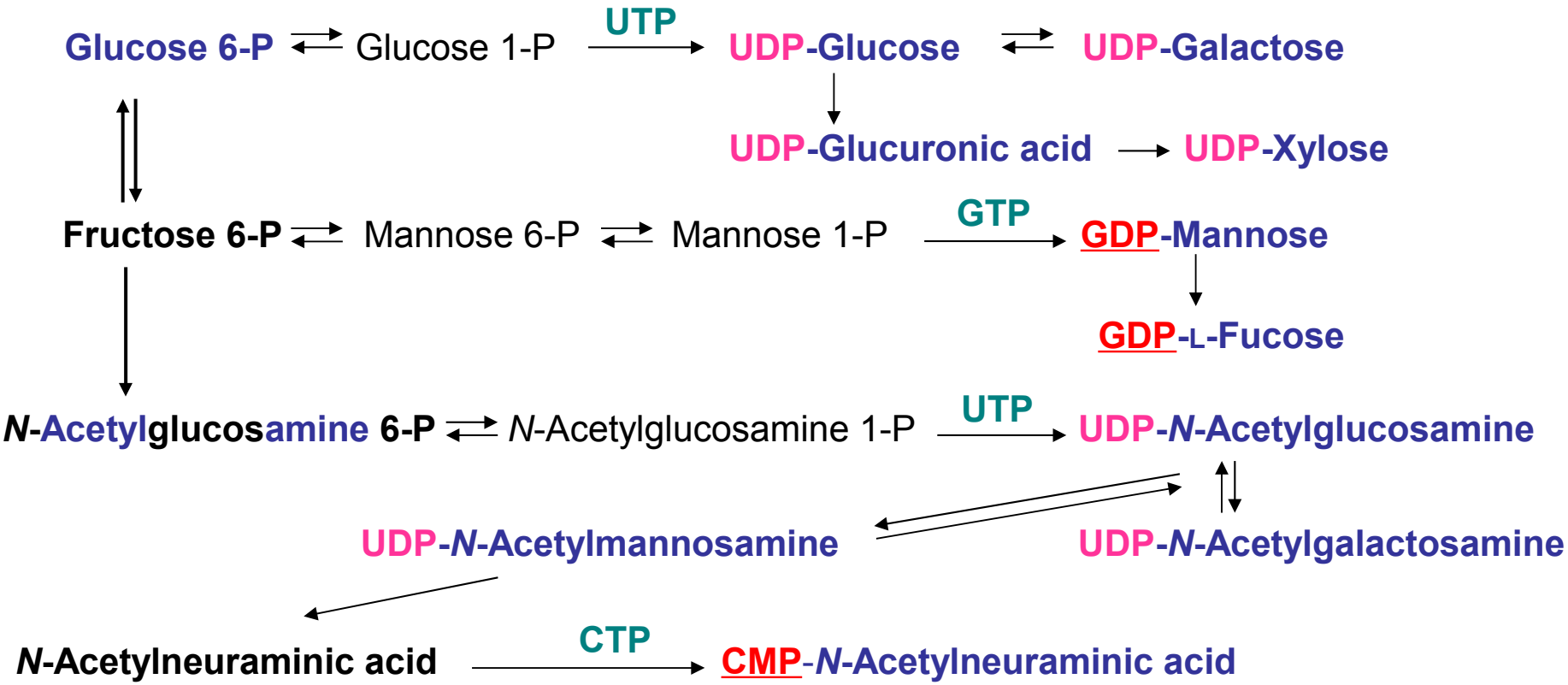
Monosaccharides found in saccharidic components of glycoproteins

		Abbreviation:
Hexoses:	**Glucose**	**Glc**
	Galactose	**Gal**
	Mannose	**Man**
Acetyl hexosamines:	***N*-Acetylglucosamine**	**GlcNAc**
	***N*-Acetylgalactosamine**	**GalNAc**
Pentoses:	**Xylose**	**Xyl**
	Arabinose	**Ara**
Deoxyhexose (Methyl pentose):	**L-Fucose**	**Fuc**
Sialic acids:	***N*-Acetylneuraminic acid** (predominant)	**NeuNAc**

Glycosyl donors in glycoprotein synthesis

Before being incorporated into the oligosaccharide chains, monosaccharides are **activated by formation of nucleotide sugars**.

The glycosyls of these compounds can be transferred to suitable acceptors (catalyzed by appropriate transferases).



O-linked glycosylation

is **posttranslational** process that takes place exclusively in the Golgi complex and which is **direct** – glycosyls from nucleotide sugars (NuDP-glycoses) are transferred to side chains of Ser or Thr residues and elongated by other nucleotide sugars.

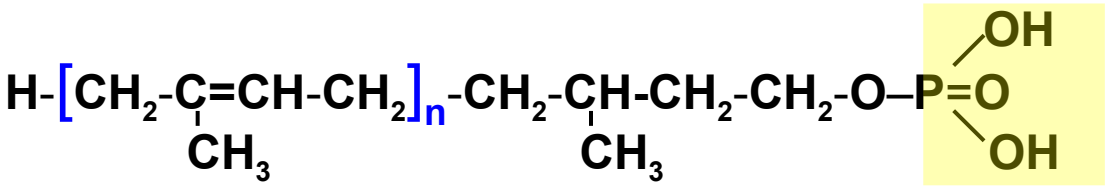
N-glycosylation of "plasma type" glycoproteins

is both **cotranslational and posttranslational** in the ER and continues (as posttranslational one) in the Golgi apparatus.

A large oligosaccharide $(GlcNAc)_2$-$(Man)_9$-$(Glc)_3$ destined for attachment to a *N*-glycosylated protein is **assembled attached to dolichol phosphate** and subsequently transferred *en bloc* to the protein acceptor, to a specific asparagine residue of the growing polypeptide chain.

The large saccharidic component is then "trimmed" and additional sugars are attached in the Golgi complex to form diverse specific patterns.

Dolichol phosphate (anchored in the membrane of endoplasmic reticulum)



n = 18-20

The assembly of dolichol-P-P-$(GlcNAc)_2(Man)_5$

on the cytoplasmic surface of the ER membranes:

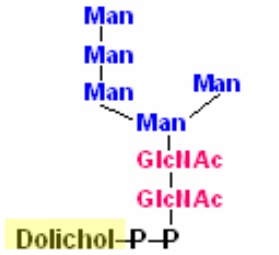
Dolichol-P

↓ 2 UDP-GlcNAc

Dolichol-P-P←**GlcNAc**←**GlcNAc**

↓ 5 GDP-Man

Dolichol-P-P←$(GlcNAc)_2$←$(Man)_5$



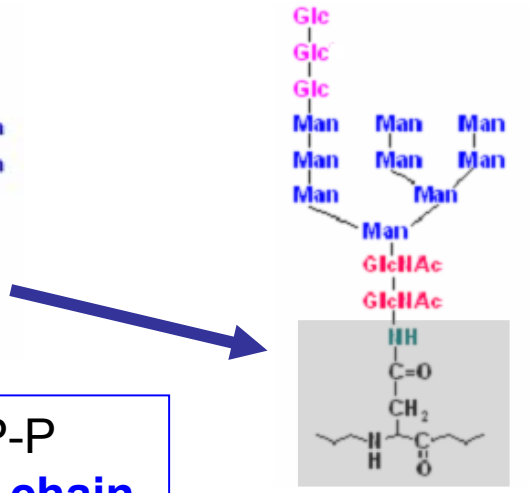
Flip of the polar part to the luminal surface of the ER membrane,
addition of the four external mannose
and **three glucose residues**:

↓ 4 Dolichol-P-Man

Dolichol-P-P←$(GlcNAc)_2(Man)_9$

↓ 3 Dolichol-P-Glc

Dolichol-P-P←$(GlcNAc)_2(Man)_9(Glc)_3$



Transfer of the saccharidic component from dolichol-P-P
to an asparaginyl residue of the growing polypeptide chain.

Trimming and final processing of *N*-linked glycoproteins:

$(Glc)_3\,(Man)_9\,(GlcNAc)_2$-protein

– **3** Glc

$(Man)_9\,(GlcNAc)_2$-protein

Trimming

up tp **4** Man removed

$(Man)_5\,(GlcNAc)_2$-protein

High-mannose *N*-linked glycoproteins

glycosylation

– **2** Man

Hybride types

$(Man)_3\,(GlcNAc)_2$-protein

pentasaccharide core region

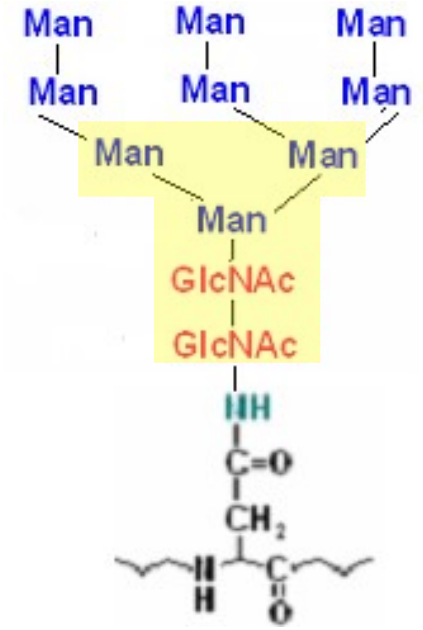
glycosylation

Late phase processing

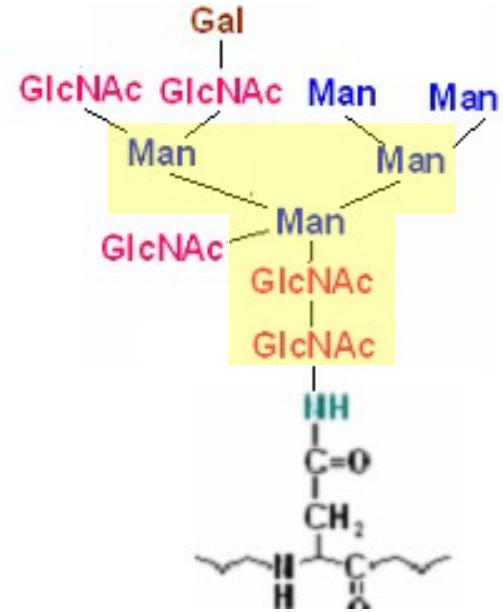
Complex type *N*-linked glycoproteins

Examples of plasma-type (*N*-linked) oligosaccharides

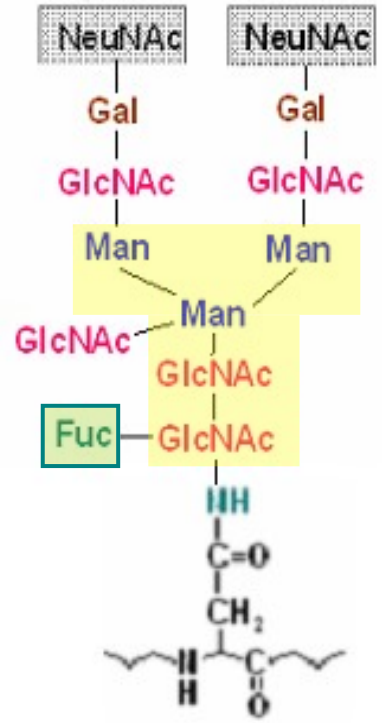
The boxed area encloses the pentasaccharide core region common to all *N*-linked glycoproteins.



High-mannose type
(before trimming to core region)

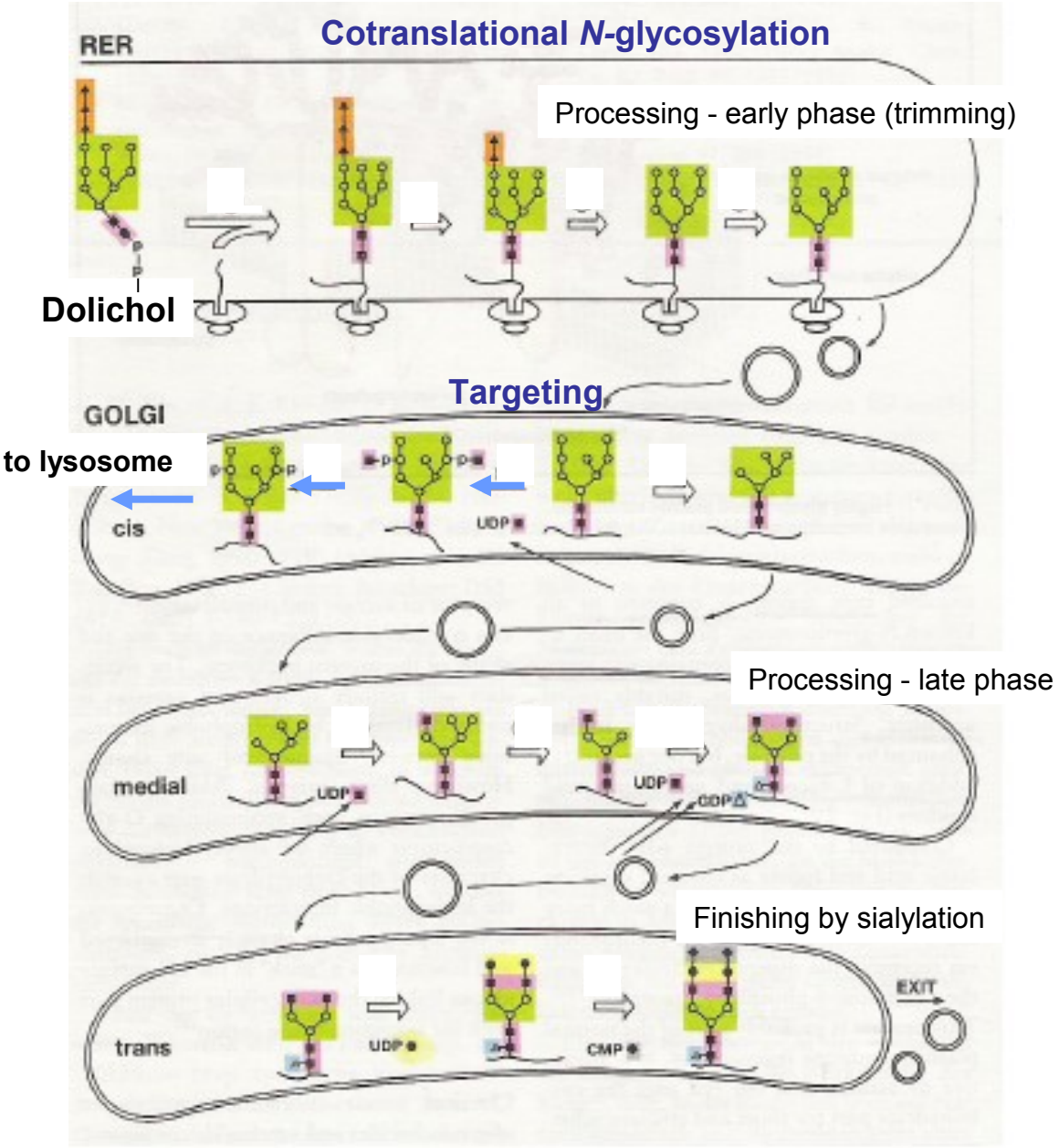


Hybride type
(one antenna)



Complex type
(triantennary chain)

Schematic pathway of oligosaccharide processing in newly synthesized glycoproteins:



The coloured symbols represent

- *N*-acetylglucosamine
- mannose
- glucose
- fucose
- galactose
- *N*-acetylneuraminic acid