

The Translation I

NPC, mRNA, tRNA, ORF, starting, capping, charging and tRNA structure (chapter 15)

LMR05.001 – 16

Sulev Kuuse

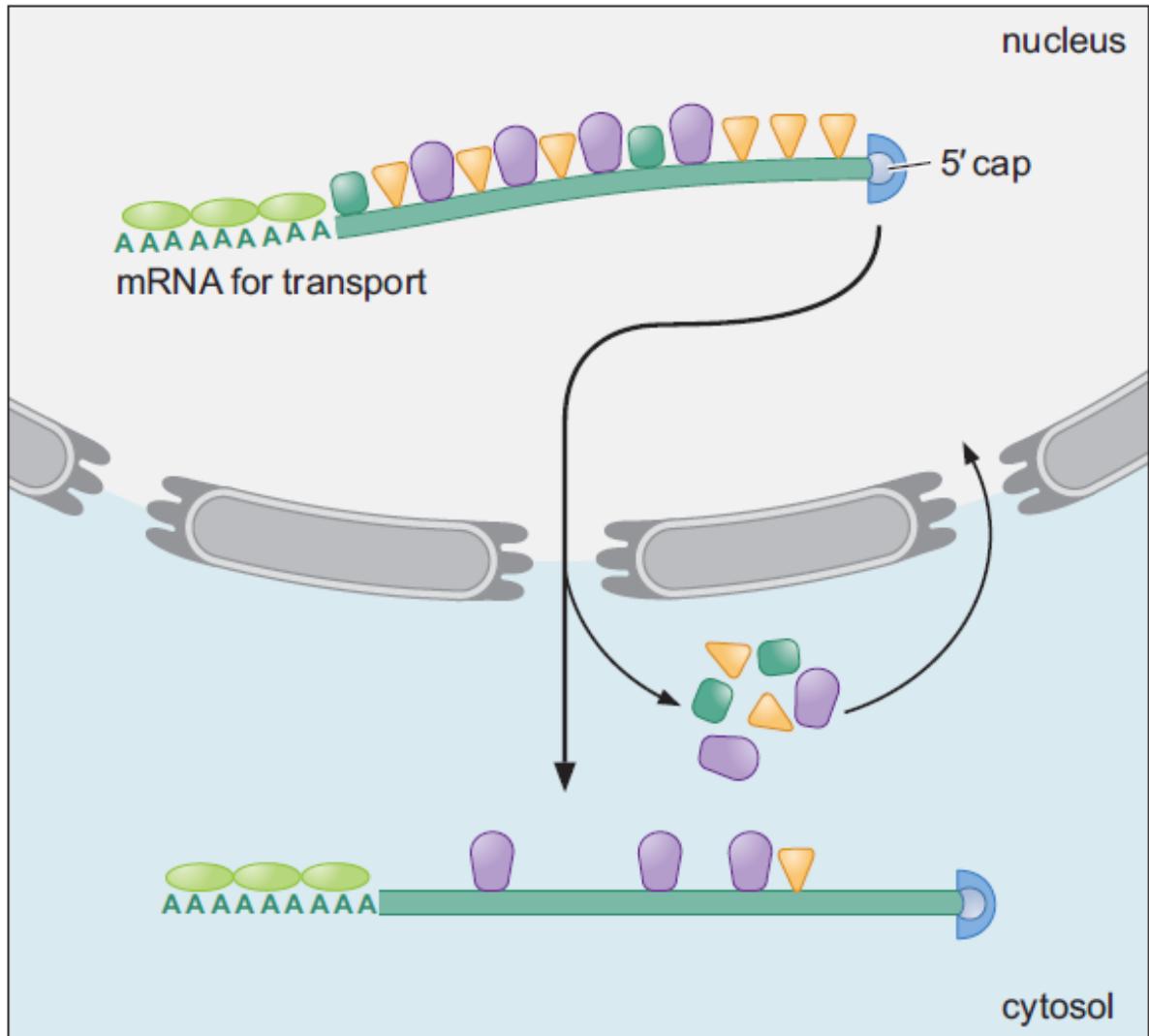
Inst Mol & Cell Biol

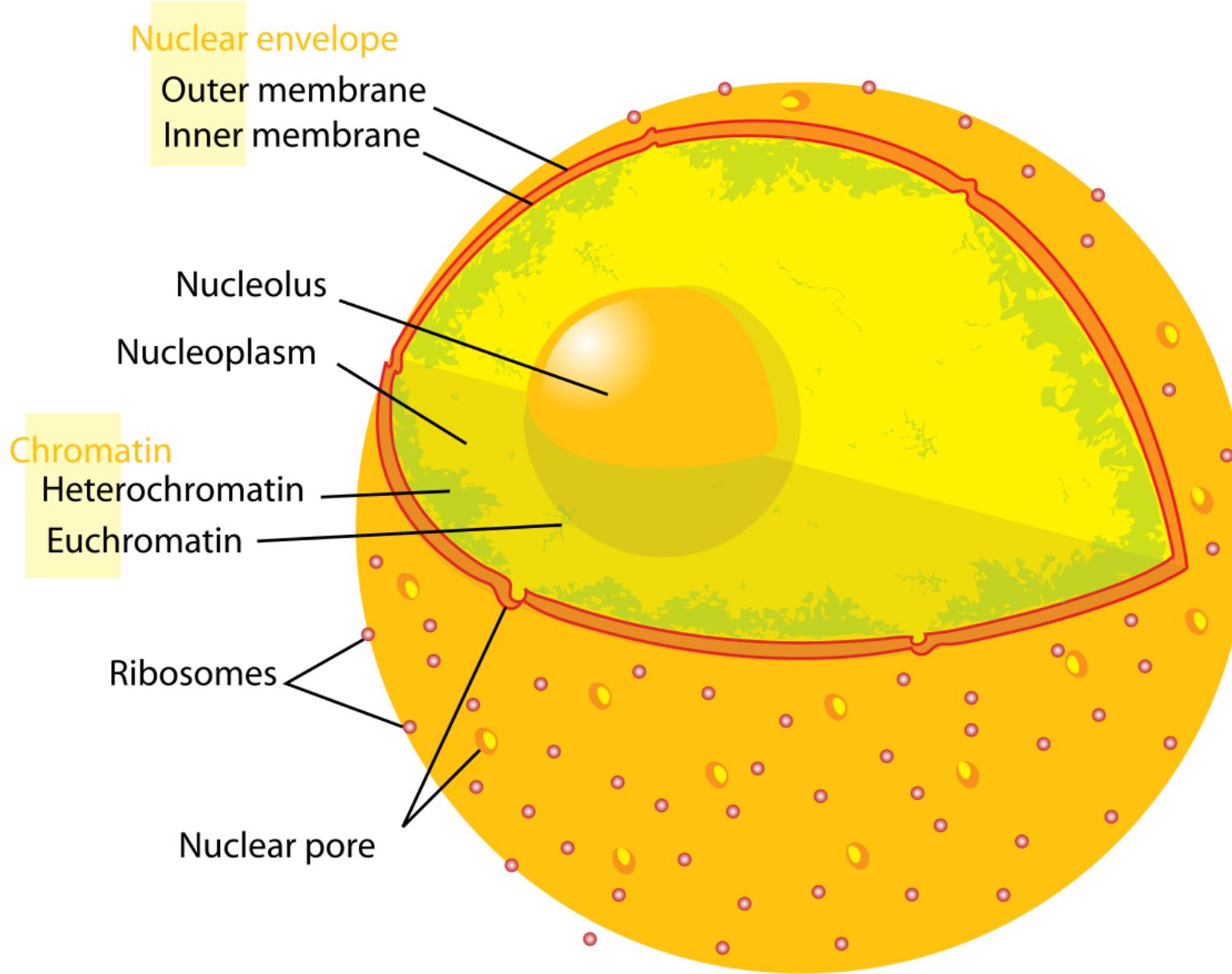
What now?

BEFORE TRANSLATION!

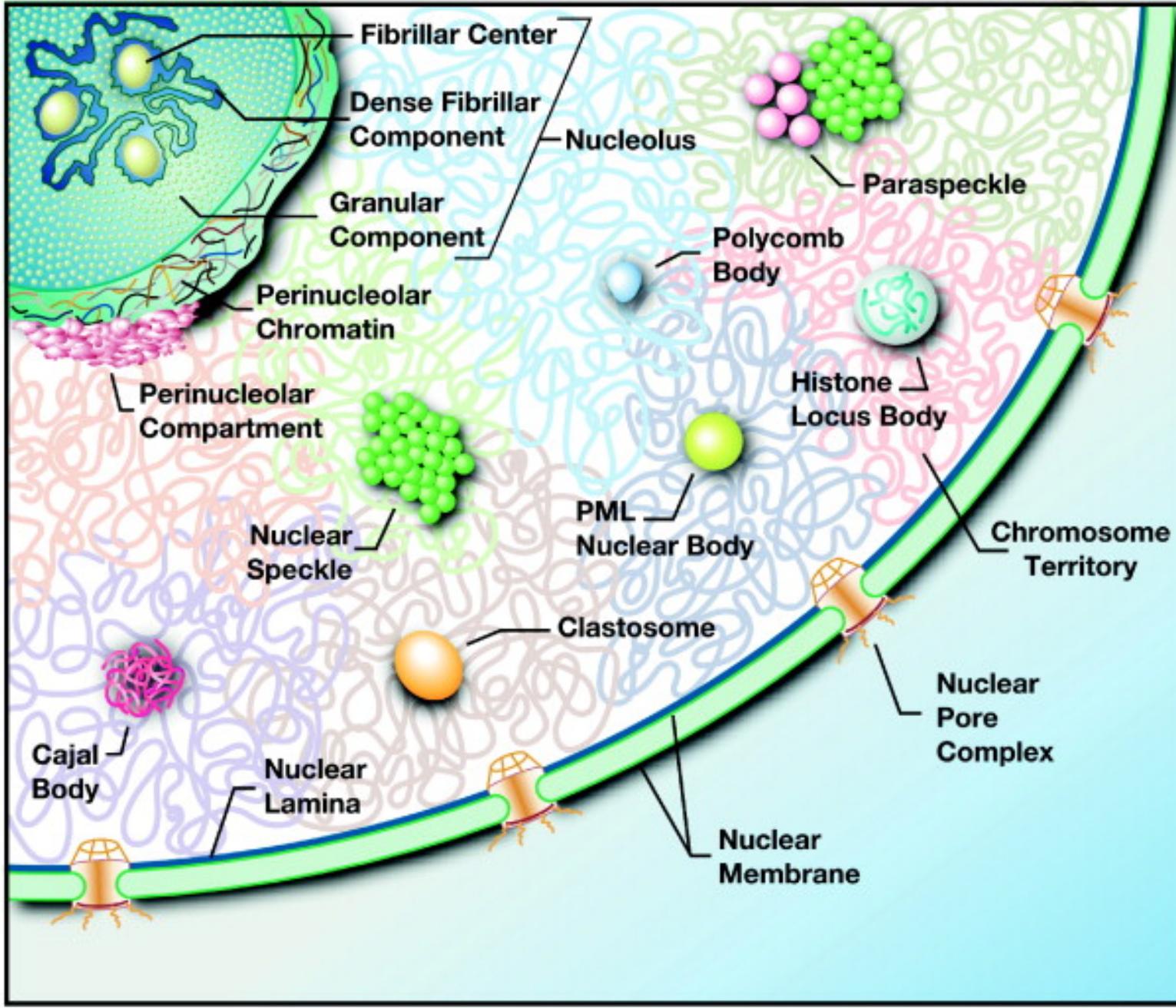
- Fully synthesized, modified, processed, compacted and edited
- RNA must be **transported** to cytoplasm
- Nuclear pore complex

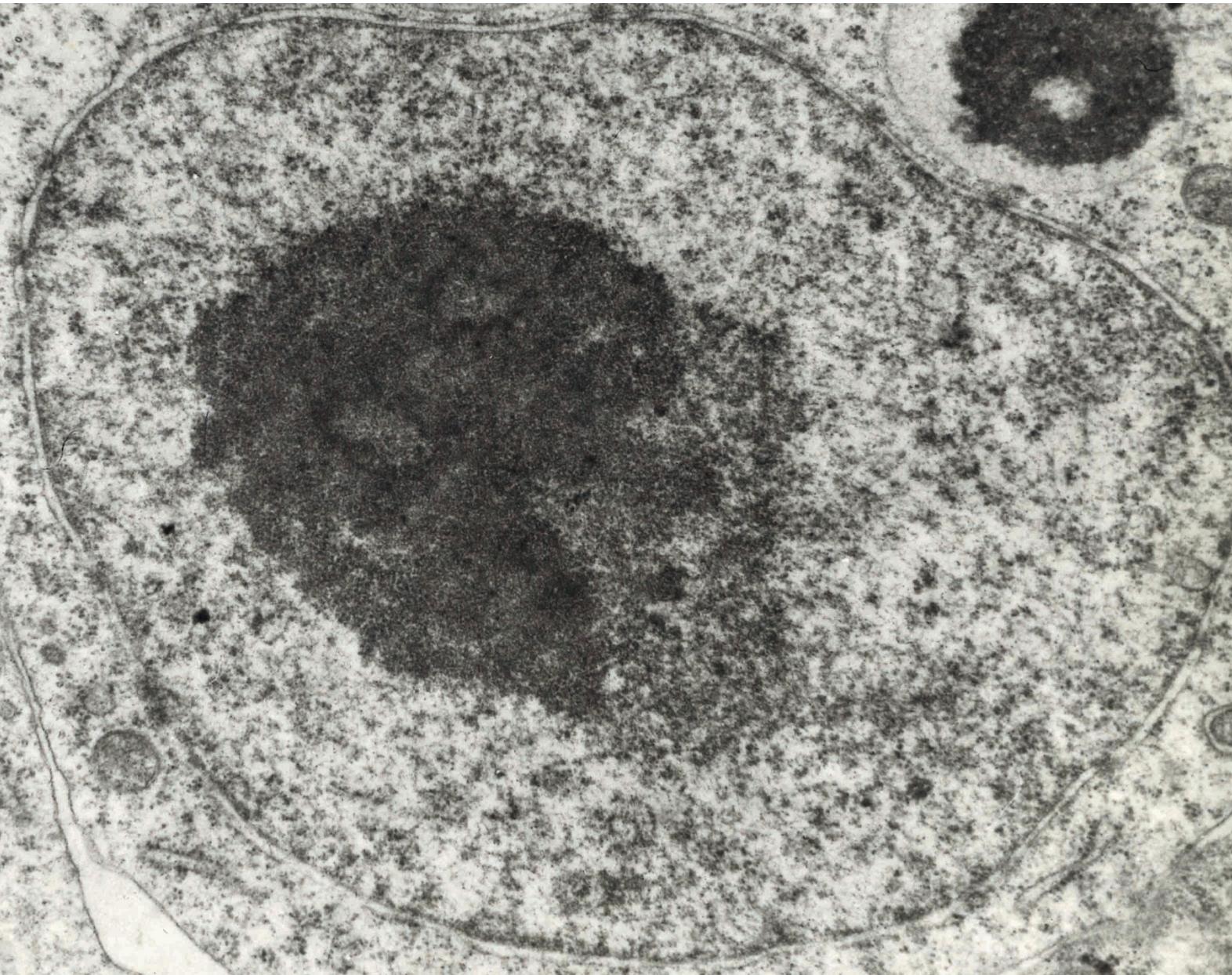
FIGURE 14-33 Transport of mRNAs out of the nucleus. RNA export from the nucleus is an active process, and only certain (appropriate) RNAs are selected for transport. To be selected for transport, the RNA must have the correct collection of proteins bound to it. These will distinguish it from other RNAs, which must be retained in the nucleus or destroyed. Proteins that recognize exon:exon boundaries, for example, indicate that an mRNA that has been appropriately spliced, whereas proteins that bind introns indicate that an RNA that should be retained in the nucleus. Once in the cytosol, some proteins are shed and others are taken on in readiness for translation (Chapter 15).



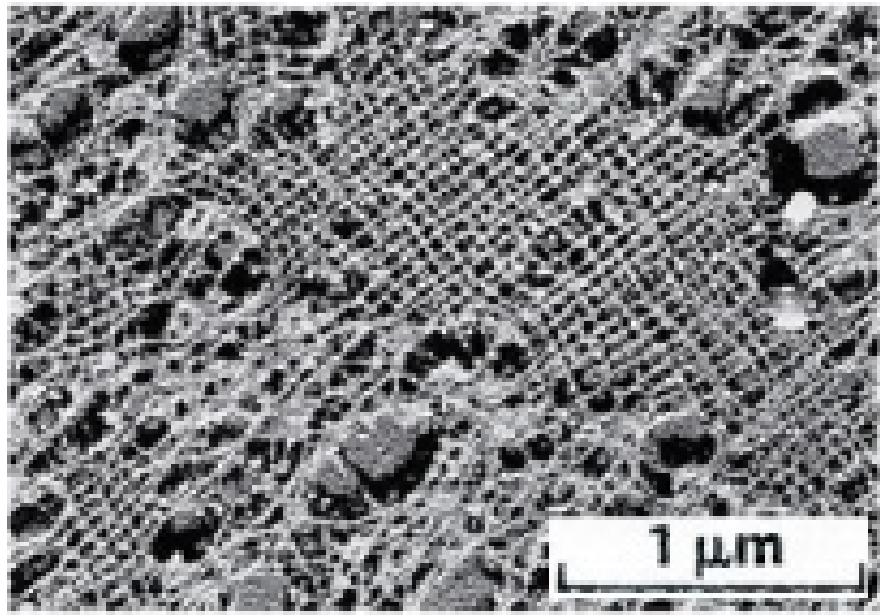


A comprehensive diagram of a human cell nucleus.

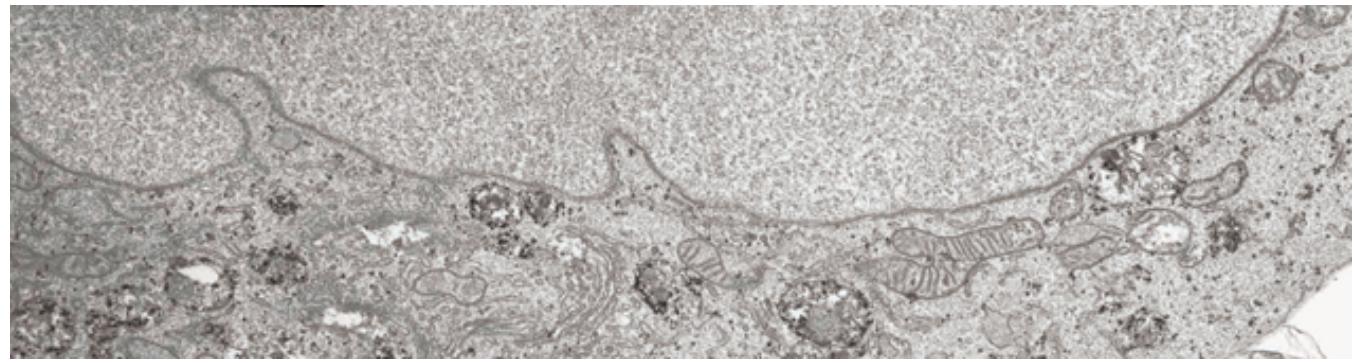




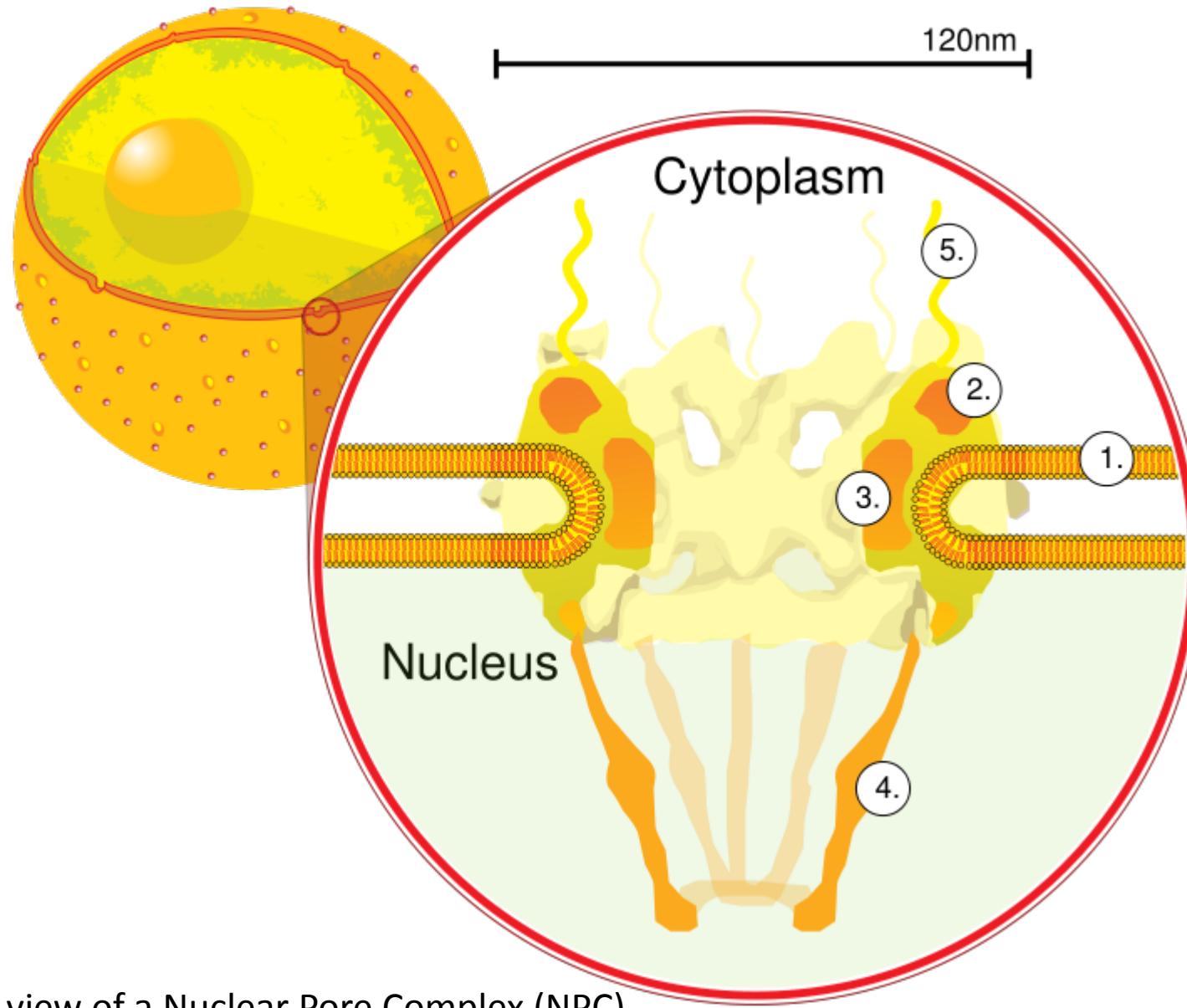
Eucaryotic cell
nucleus with
two outer
membranes.
In the cell nucleus
there is nucleolus
(dark structure).



The SEM picture
from cell nucleus
membrane



The TEM picture from cell - cytoplasma with mitochondrions, Golgi complex,
lysosomes, ribosomes, transport vesicles - below.
Nuclear double membrane
Nucleus with chromatin (top)



Side-view of a Nuclear Pore Complex (NPC).

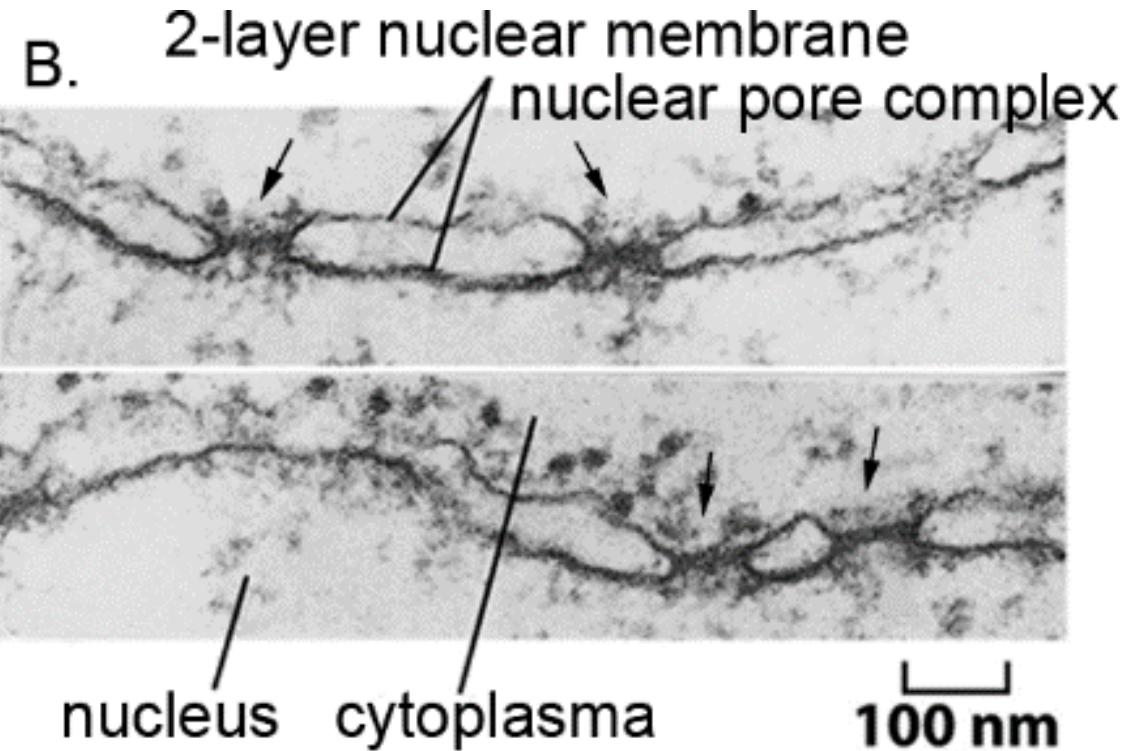
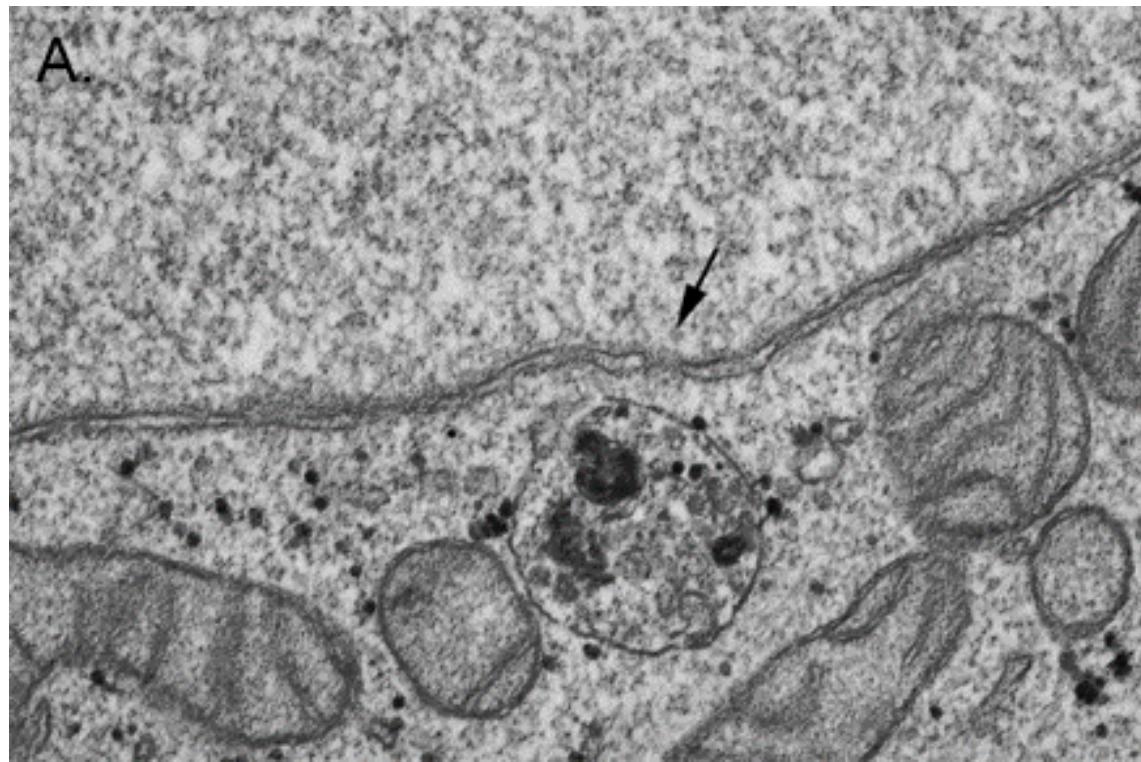
- 1.) Nuclear Envelope
- 2.) Outer Ring of NPC
- 3.) Spokes of NPC
- 4.) Basket of NPC
- 5.) Filaments of NPC.

The molecular mass of the mammalian NPC is about 124 megadaltons (MDa) and it contains approximately 30 different protein components, each in multiple copies. In contrast, the yeast *Saccharomyces cerevisiae* is smaller, with a mass of only 66 MDa.

The entire nuclear pore complex has a diameter of about 120 nanometers in vertebrates.

The diameter of the channel ranges from 5.2 nanometers in humans to 10.7 nm in the frog *Xenopus laevis*, with a depth of roughly 45 nm.

mRNA, which is single-stranded, has a thickness (diameter) of about 0.5 to 1 nm.



A nuclear pore is a part of a large complex of proteins, known as a nuclear pore complex that spans the nuclear envelope, which is the double membrane surrounding the eukaryotic cell nucleus.

There are an average of 1,000 nuclear pore complexes (NPCs) in the nuclear envelope of a vertebrate cell, but it varies depending on cell type and the stage in the life cycle.

The human nuclear pore complex (hNPC) is a 110 MDa structure. The proteins that make up the nuclear pore complex are known as nucleoporins; each NPC contains at least 456 individual protein molecules and is composed of 34 distinct nucleoporin

About half of the nucleoporins typically contain **solenoid protein domains**—either an alpha solenoid or a beta-propeller fold, or in some cases both as separate structural domains.

The other half show structural characteristics typical of "natively unfolded" or intrinsically **disordered proteins**, i.e. they are **highly flexible proteins that lack ordered tertiary structure**.

These **disordered proteins** are the **FG nucleoporins**, so called because their amino-acid sequence contains many **phenylalanine—glycine repeats**.

Nucleoporins are able to transport molecules across the nuclear envelope at a very high rate. A single NPC is able to transport **60,000 protein molecules across the nuclear envelope every minute**.

Nucleoporins mediate transport of macromolecules between the cell nucleus and cytoplasm in eukaryotes.

Certain members of the nucleoporin family form the structural scaffolding of the nuclear pore complex.

However, nucleoporins primarily function by interacting with transport molecules known as **karyopherins**, also known as **Kaps**

These karyopherins interact with nucleoporins that contain repeating sequences of the amino acids phenylalanine (F) and glycine (G) and form **FG peptide repeats**.

FG-repeats are **small hydrophobic segments** that break up long stretches of hydrophilic amino acids.

These flexible parts **form unfolded, or disordered segments without a fixed structure.**

They form a mass of chains which allow smaller molecules to diffuse through,
but exclude large hydrophilic macromolecules.

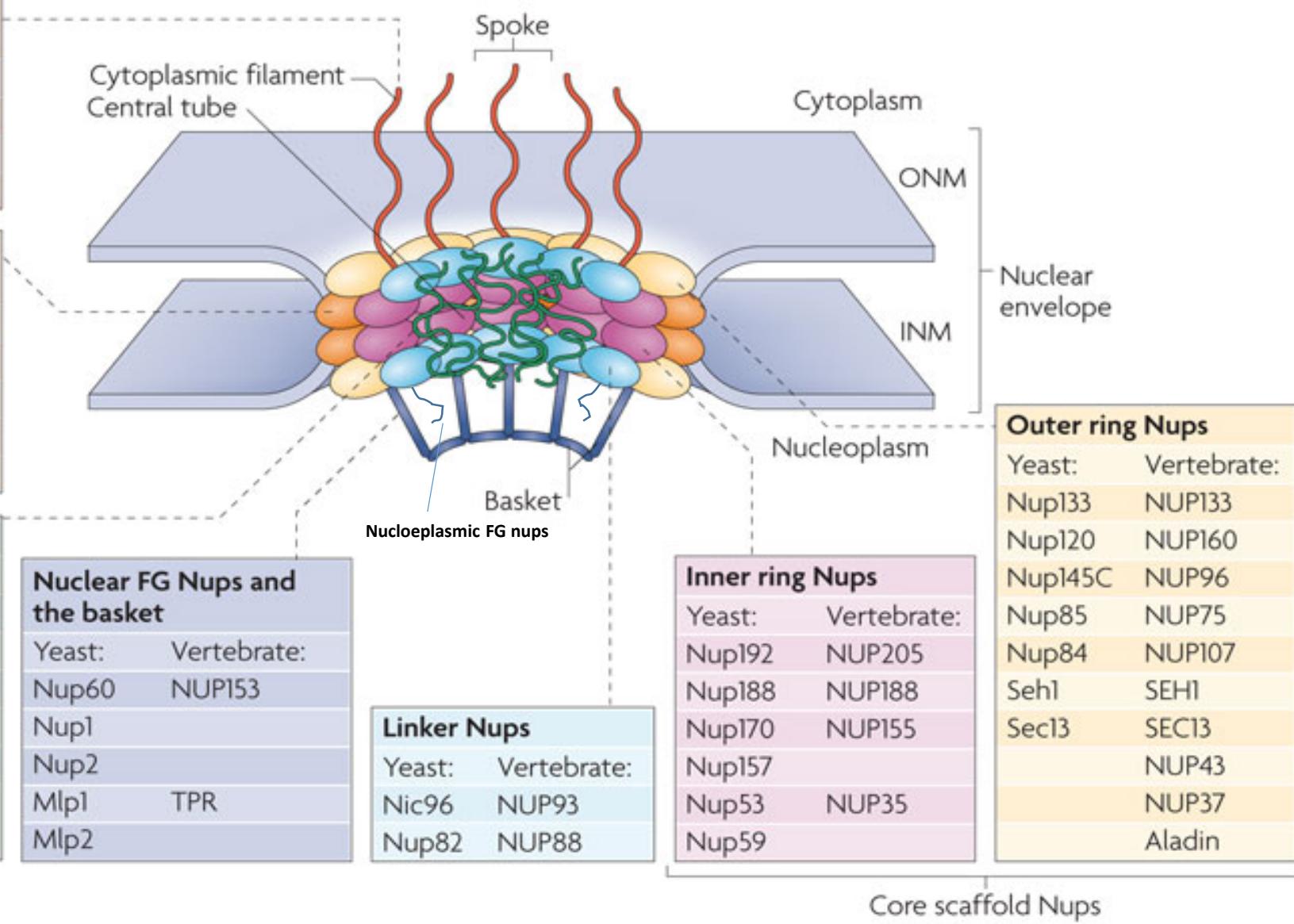
These **large molecules are only able to cross a nuclear pore** if they are accompanied by a **signaling molecule** that temporarily interacts with a nucleoporin's FG-repeat segment.

FG-nucleoporins also contain a globular portion that serves as **an anchor for attachment to the nuclear pore complex.**

Cytoplasmic FG Nups and filaments	
Yeast:	Vertebrate:
Nup159	NUP214
Nup42	NLP1

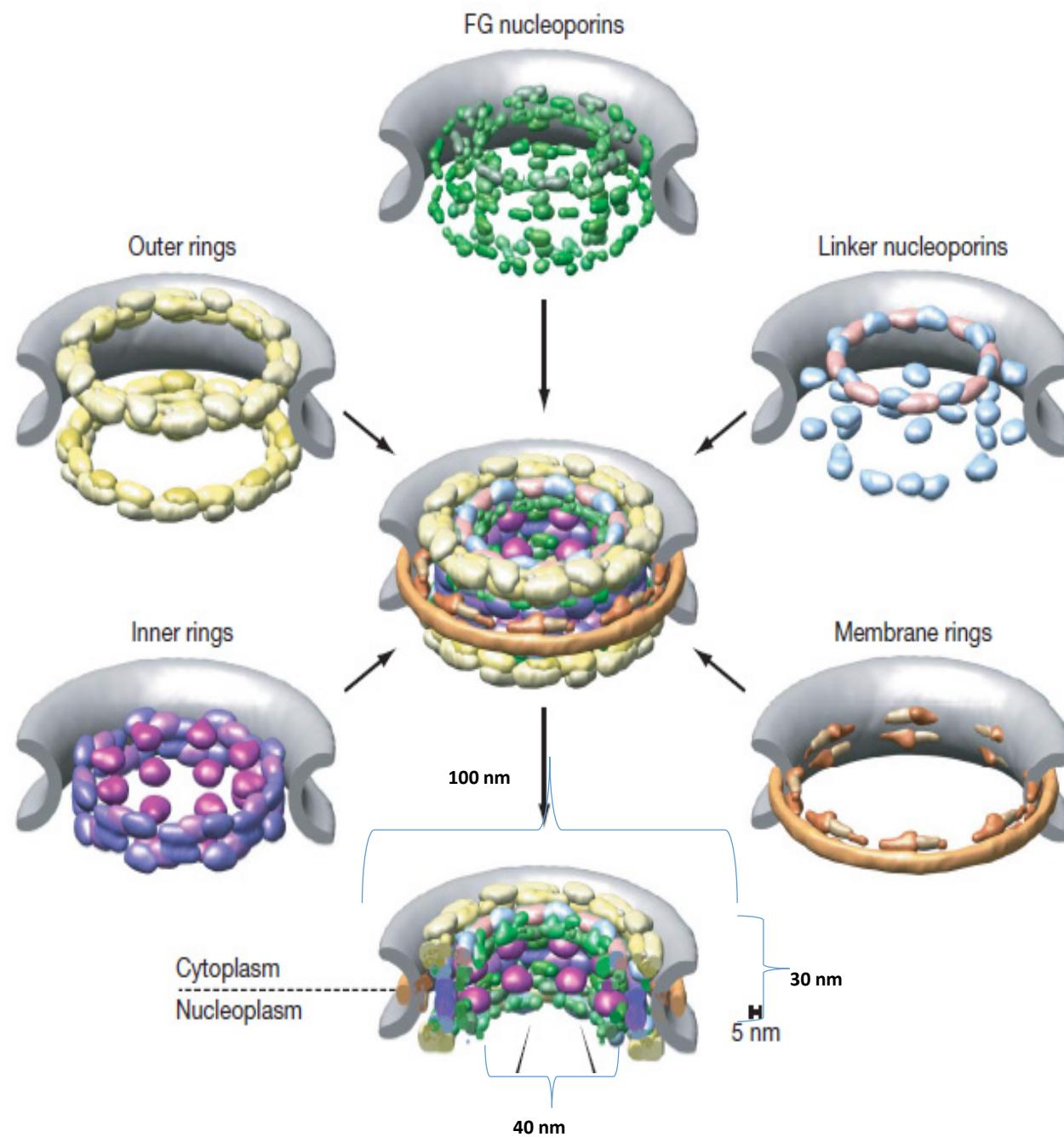
Transmembrane ring Nups	
Yeast:	Vertebrate:
Pom152	GP210
Pom34	
Ndc1	NDC1
	POM121

Central FG Nups	
Yeast:	Vertebrate:
Nup145N	NUP98
Nup116	
Nup100	
Nsp1	NUP62
Nup57	NUP54
Nup49	NUP58 and NUP45

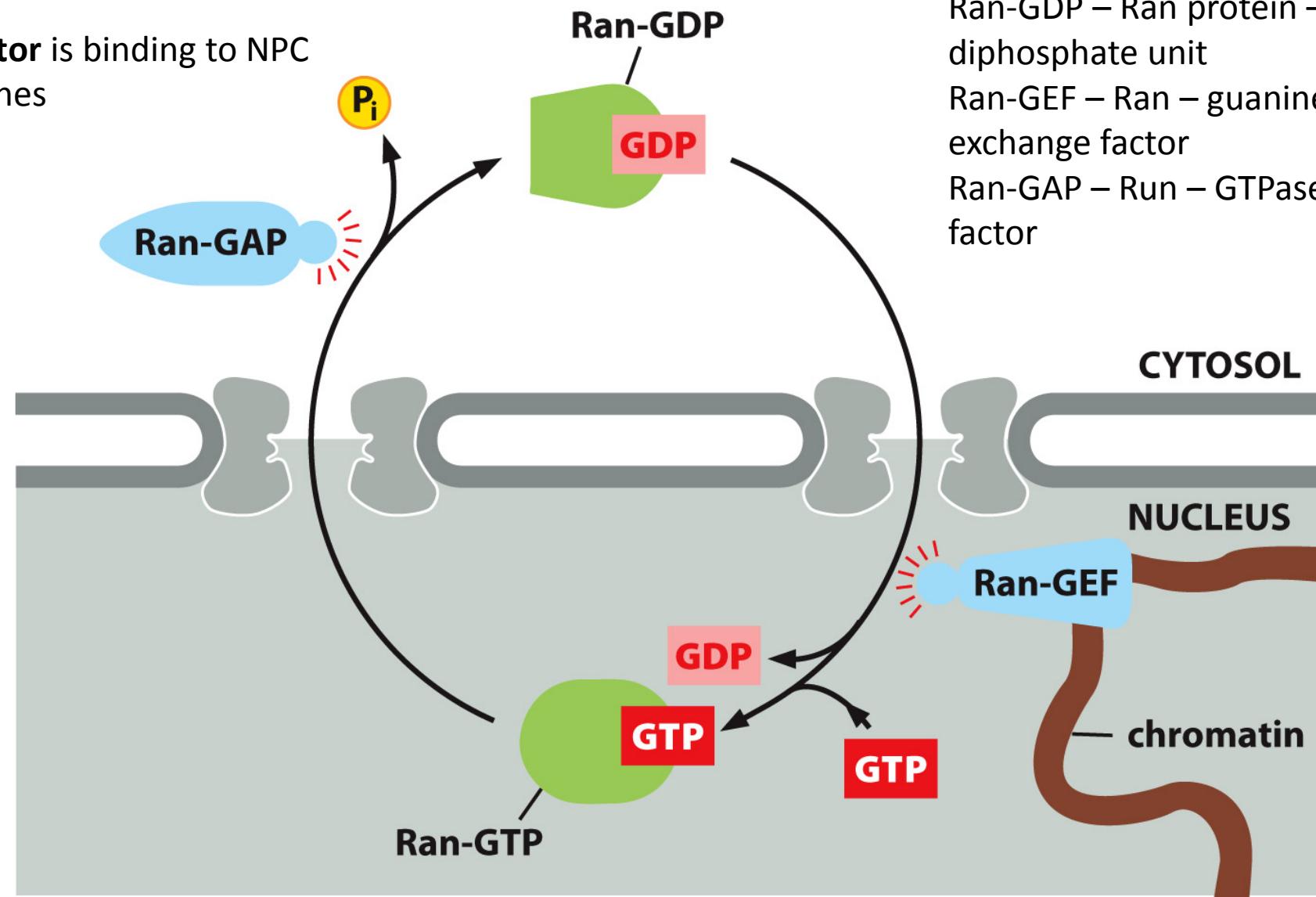


The structure of nuclear pore complex (NPC): Main structure elements are:
Central tube of NPC, spokes of complex and basket of NPC.

The structure of NPC.



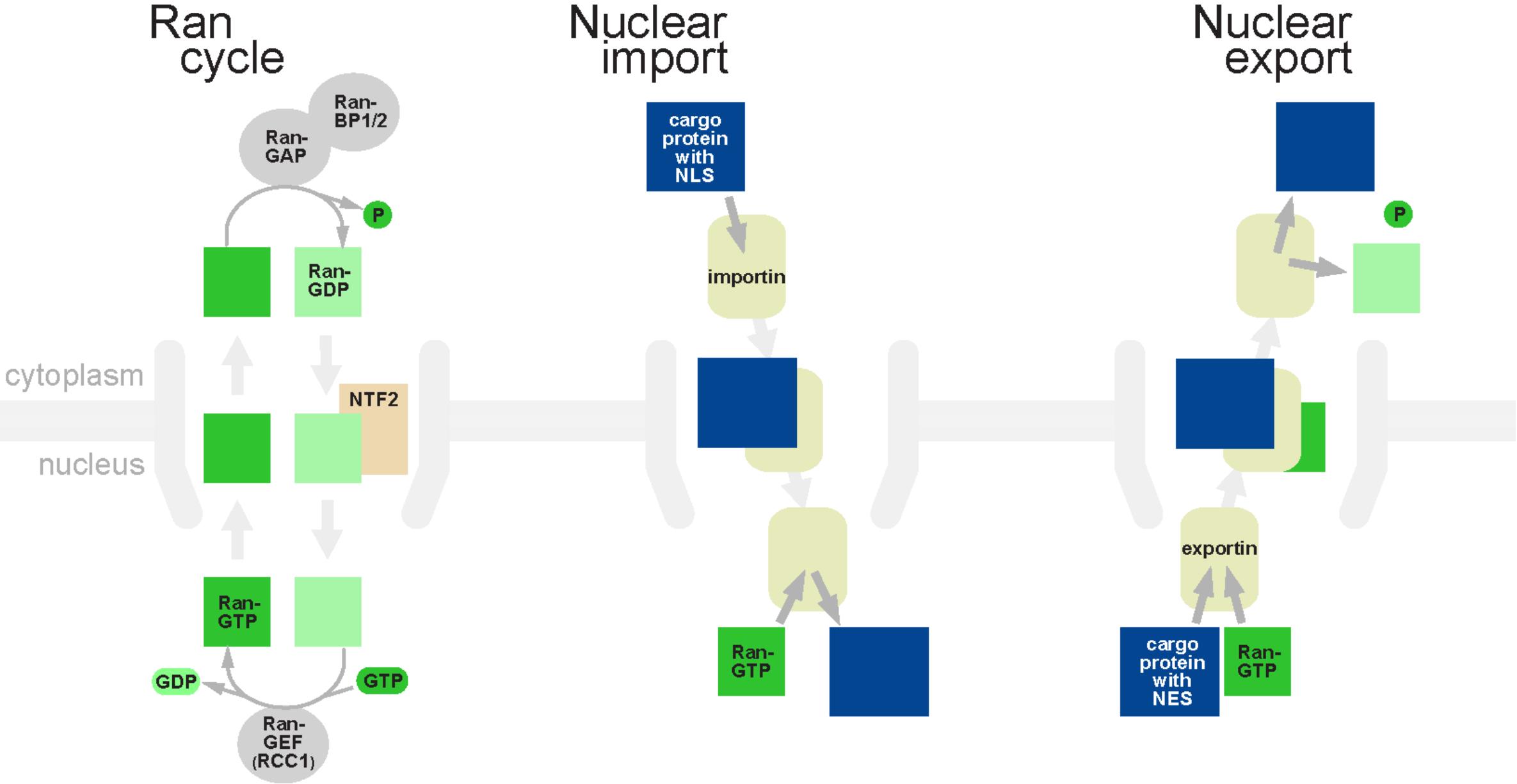
Ran-GDP receptor is binding to NPC
FG nuclear porines



Ran-GDP – Ran protein – guanosine diphosphate unit
Ran-GEF – Ran – guanine nucleotide exchange factor
Ran-GAP – Run – GTPase activating factor

Figure 12-12 Molecular Biology of the Cell 6e (© Garland Science 2015)

The movement of **Ran** (regulator protein) from the cytoplasm to the nucleus and vice versa. Its need the ATP hydrolysis energy.



The Ran-cycle helps to move some proteins to the nucleus (import) or out of nucleus (export).

Transport through the nuclear pore complex

Small particles (< ~30-60 kDa) are able to pass through the nuclear pore complex by **passive diffusion**.

Larger particles are also **able to diffuse passively** through the large diameter of the pore, at rates that decrease gradually with molecular weight.

Efficient passage through the complex requires **several protein factors**, and in particular, **nuclear transport receptors** that bind to cargo molecules and mediate their **translocation across the NPC**, either into the nucleus (importins) or out of it (exportins) and **ATP**.

The largest family of nuclear transport receptors are **karyopherins**, which includes dozens of both importins and exportins; this family is further subdivided to the karyopherin- α and the karyopherin- β subfamilies.



What is translation?

- Process of interpreting mRNA nucleotide sequence into a sequence of amino acids in proteins
- 80% of energy, 50% of dry mass are dedicated to protein synthesis





Lecture XVI, Slide 19, The Translation, amino acids and the structure of proteins

Central dogma

- 1953 Chromosomal DNA functions as the template for RNA molecules

Francis Crick, 1956

central dogma

sup-tRNA - a nonsense suppressor tRNA

Duplication



Transcription



RNA

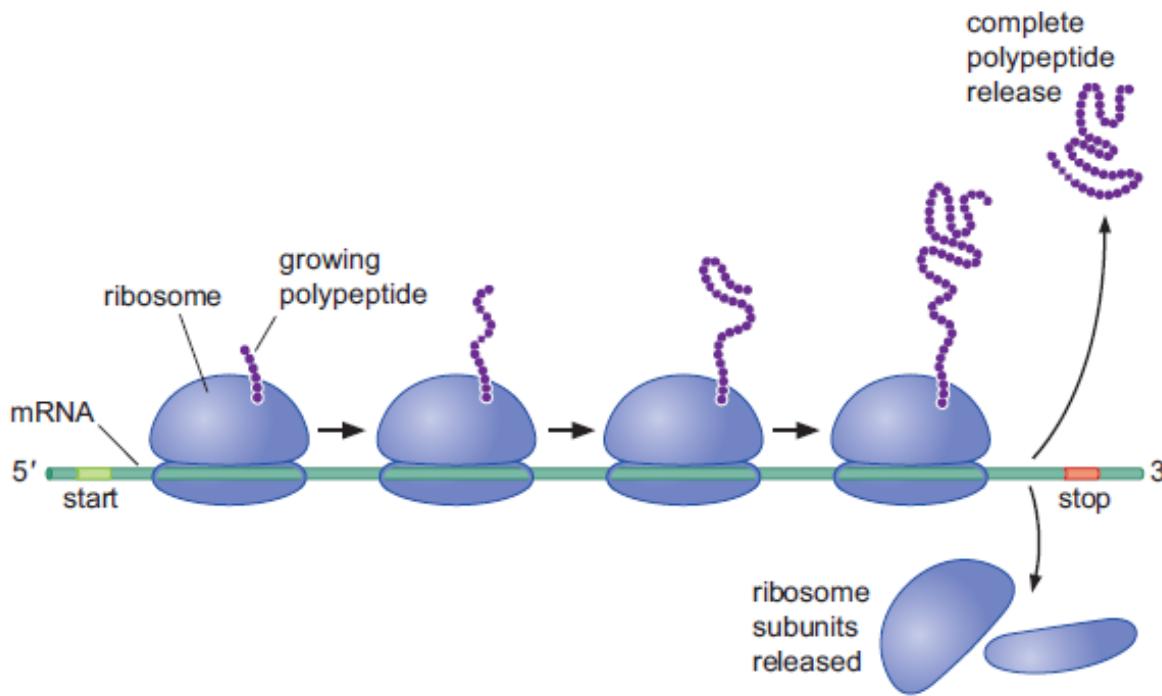
Translation



Protein.

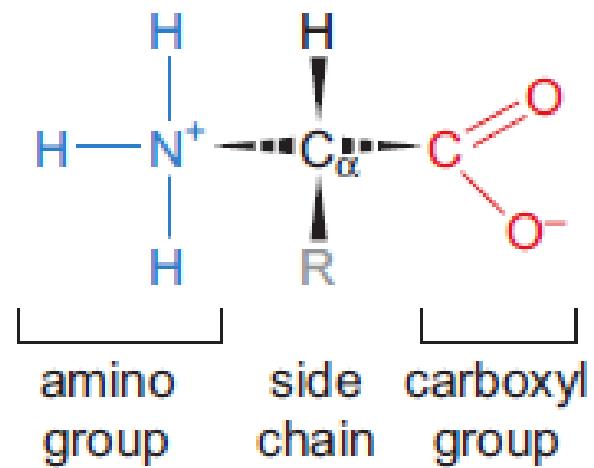
Time for the ribosomes to take over!

FIGURE 2-16 Diagram of a polyribosome. Each ribosome attaches at a start signal at the 5' end of an mRNA chain and synthesizes a polypeptide as it proceeds along the molecule. Several ribosomes may be attached to one mRNA molecule at one time; the entire assembly is called a polyribosome.



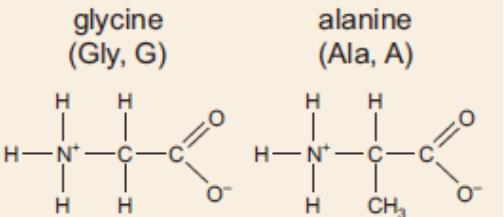
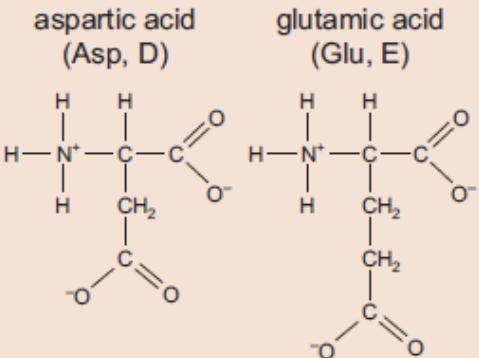
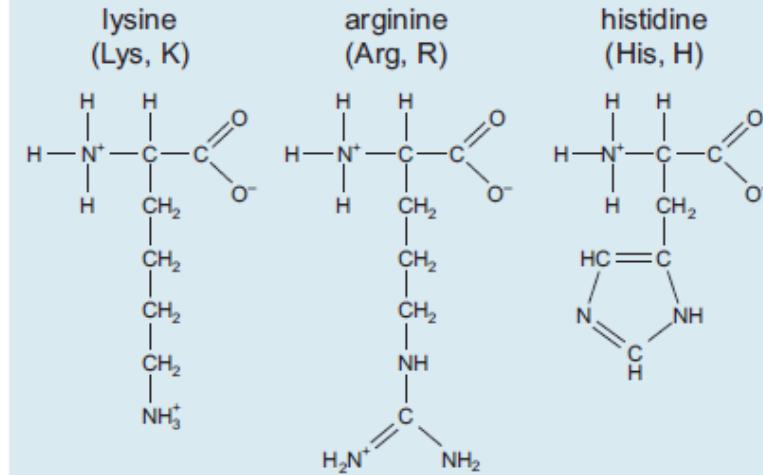
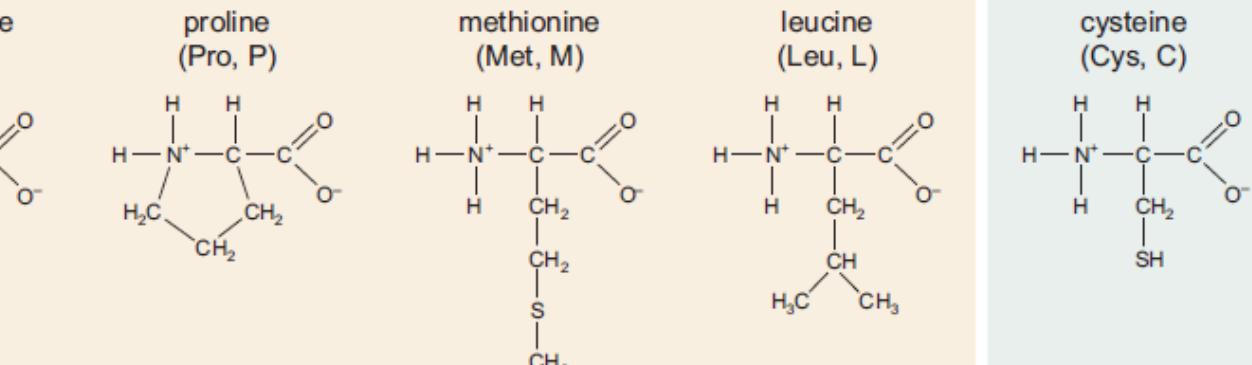
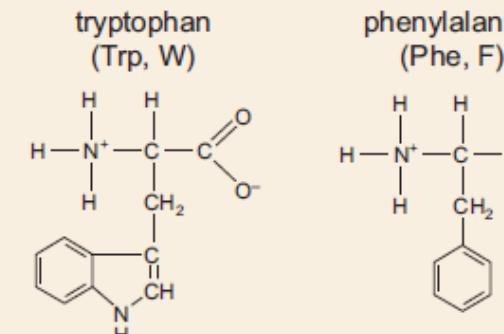
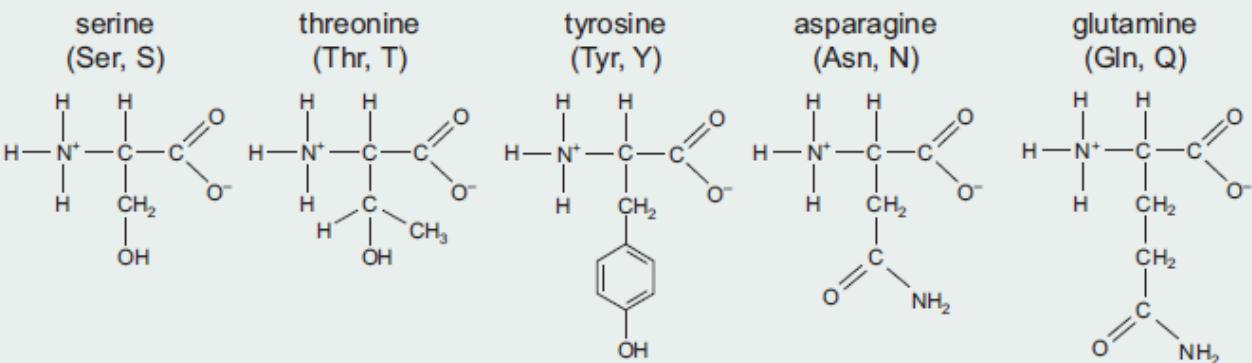
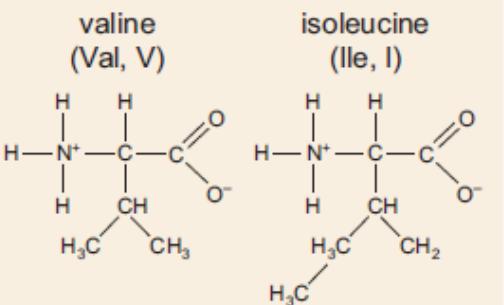


Amino acid structure



- Properties of R group determine:
 - Polarity
 - Size

FIGURE 6-1 Structural features of an amino acid.

neutral-nonpolar amino acids**acidic amino acids****basic amino acids****neutral-polar amino acids**

Peptide bond

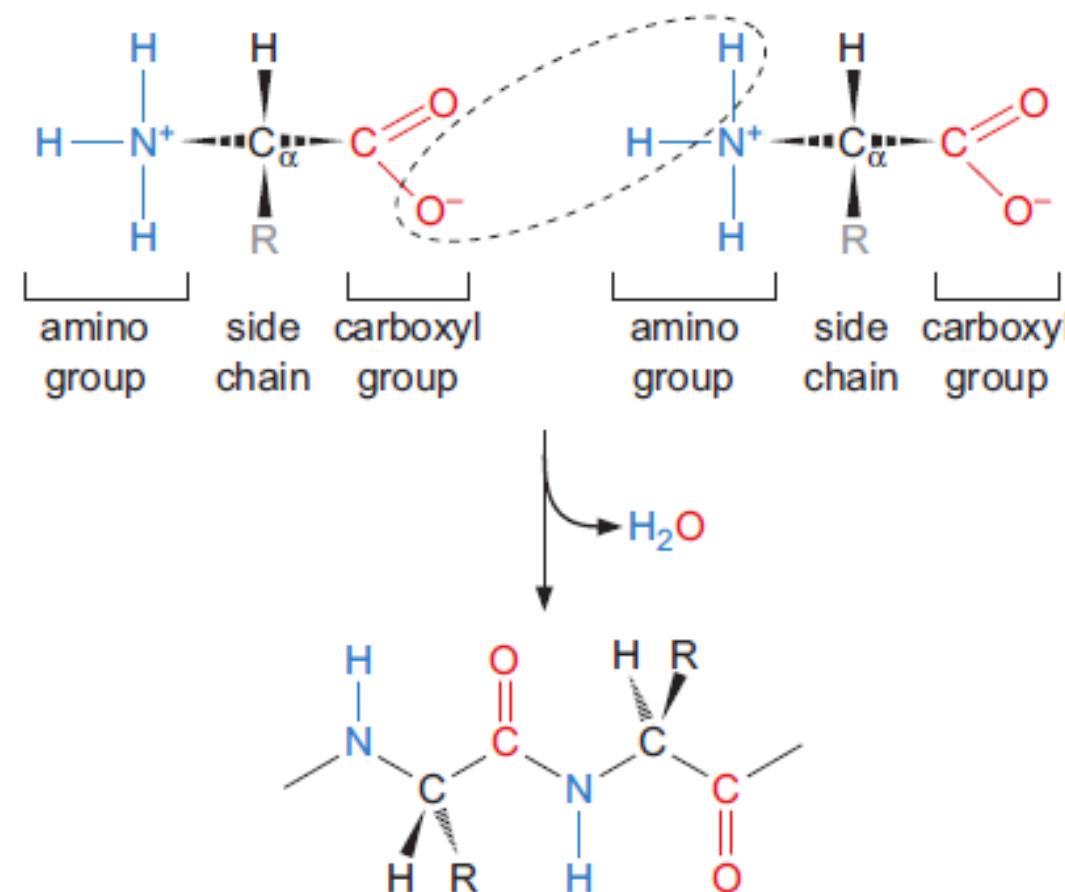


FIGURE 6-3 Peptide-bond formation.

Adaptor Hypothesis, discovery of tRNA

RNA templates fold to

- form cavities specific for 20 amino acids

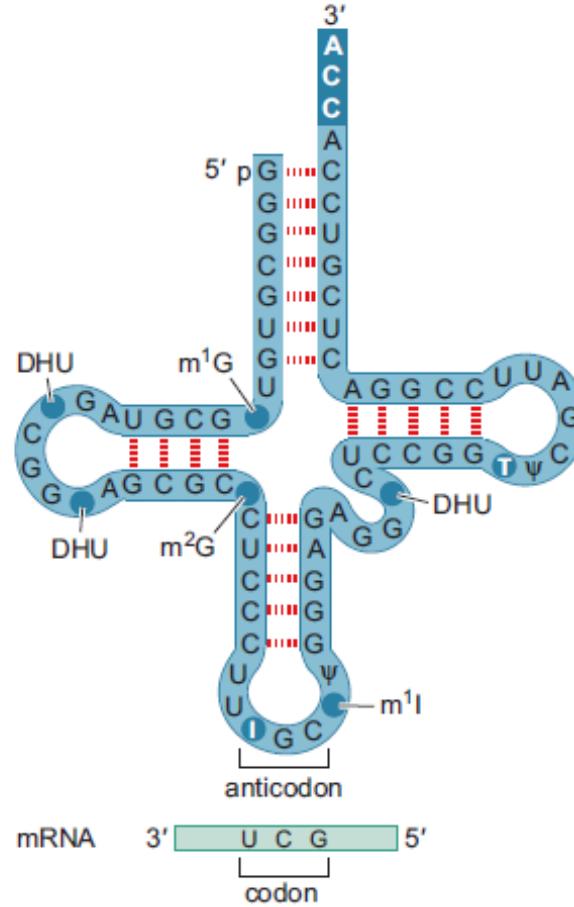
Francis Crick proposed

- the existence of an adaptor molecule

tRNA was discovered

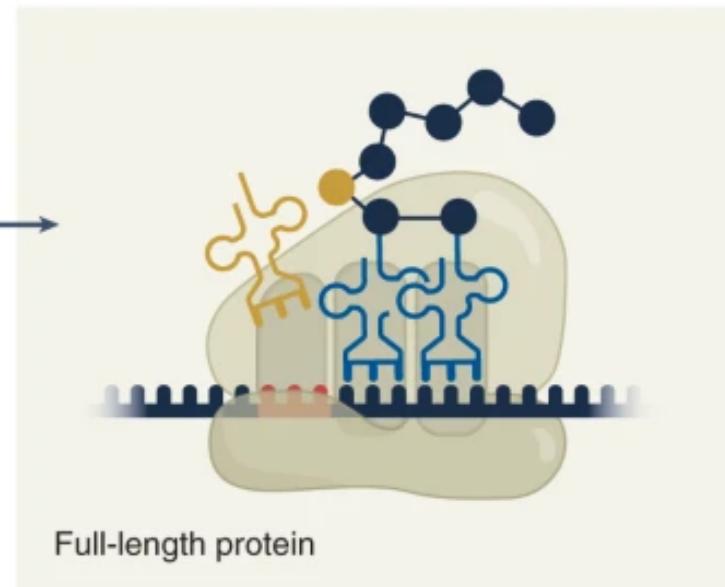
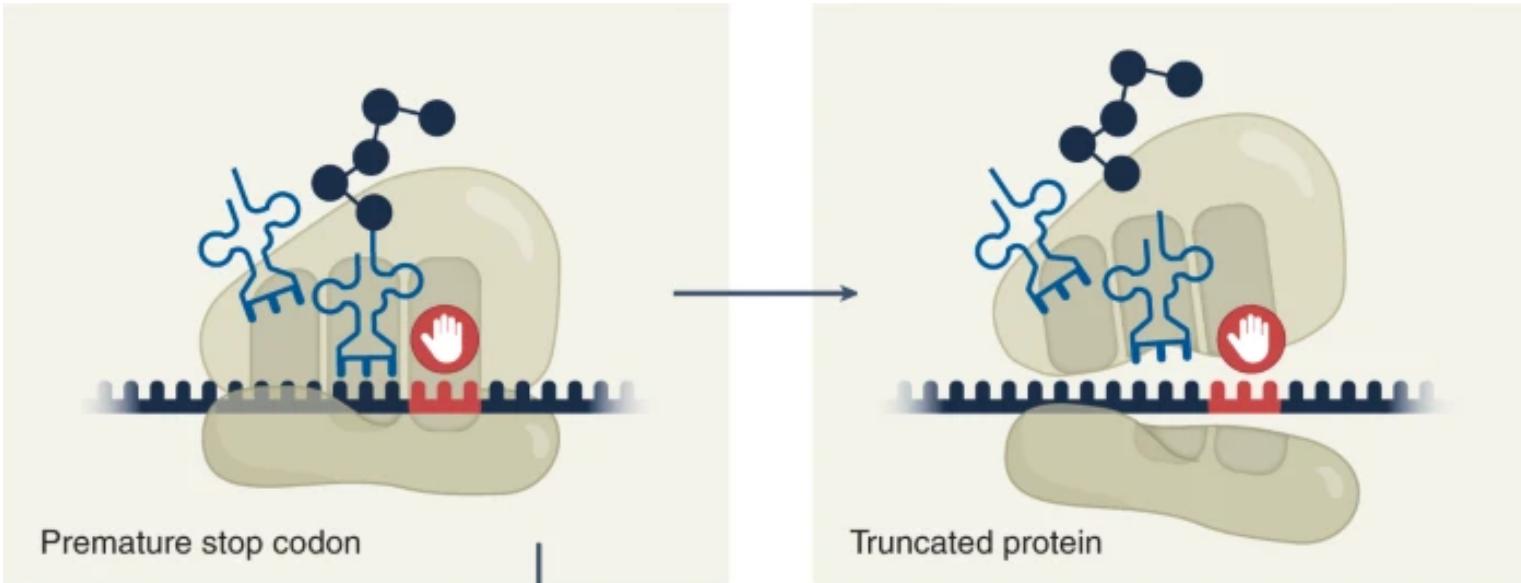
- in late 1950s (1958)

Paul C. Zamecnik (1912 - 2009) and Mahlon B. Hoagland (1921 - 2009) discovered a molecule that is essential for protein synthesis: tRNA.



During their experiments with rat liver cells, Hoagland and Zamecnik noticed that in the presence of ATP, amino acids associate with heat soluble RNA (later named tRNA) and complex of amino acid and tRNA was later called aminoacyl-tRNA

FIGURE 2-14 Yeast alanine tRNA structure, as determined by Robert W. Holley and his associates. The anticodon in this tRNA recognizes the codon for alanine in the mRNA. Several modified nucleosides exist in the structure: ψ = pseudouridine, T = ribothymidine, DHU = 5,6-dihydrouridine, I = inosine, m^1G = 1-methylguanosine, m^1I = 1-methylinosine, and m^2G = N,N-dimethylguanosine.



A nonsense **suppressor tRNA** (sup-tRNA) allows a natural or non-natural amino acid to be assigned to a nonsense codon in mRNA.

What makes translation machinery?

- mRNAs
 - tRNAs
 - Aminoacyl-tRNA synthetases
 - Ribosome
-
- 4 nucleotides - > 20 amino acids. How?

TABLE 2-3 The Genetic Code

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

mRNA contains the message

- Only a portion of mRNA is translated
- Protein-encoding region is called **open reading frame (ORF)**
 - One ORF = one protein
 - Start codon (AUG/GUG/UUG in prokaryotes, always AUG in eukaryotes)
 - Stop codon (UAG/UGA/UAA)

5'-AUG-3'

We can now fully appreciate the origin of the term **open reading frame**. It is a contiguous stretch of codons “read” in a particular frame (as set by the first codon) that is “open” to translation because it lacks a stop codon (i.e., until the last codon in the ORF).



mRNA

The **start codon** has two important functions.

First, it specifies the **first amino acid** to be incorporated into the growing polypeptide chain.

Second, it defines the **reading frame for all subsequent codons**. Because each codon is immediately adjacent to (but not overlapping with) the next codon, and because codons are three nucleotides long, any stretch of mRNA could be translated in three different reading frames (Fig. 15-1)

- Three potential reading frames

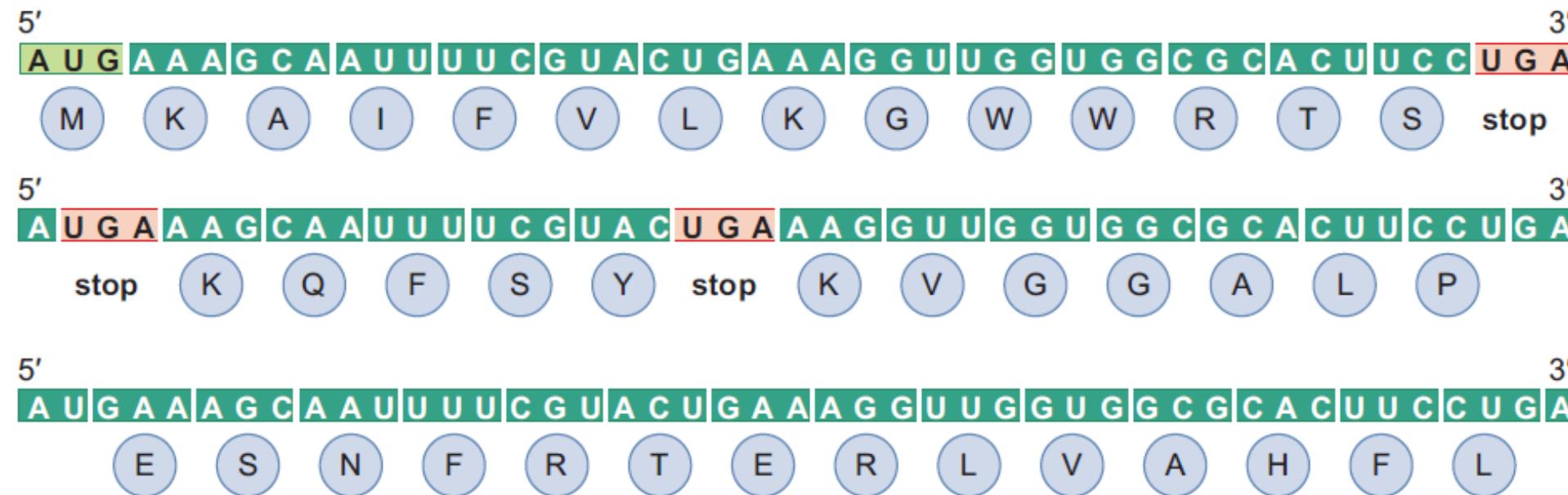


FIGURE 15-1 Three possible reading frames of the *Escherichia coli trp* leader sequence. Start codons are shaded in green, and stop codons are shaded in red. The amino acid sequence encoded by each reading frame is indicated in the single-letter code below each codon.

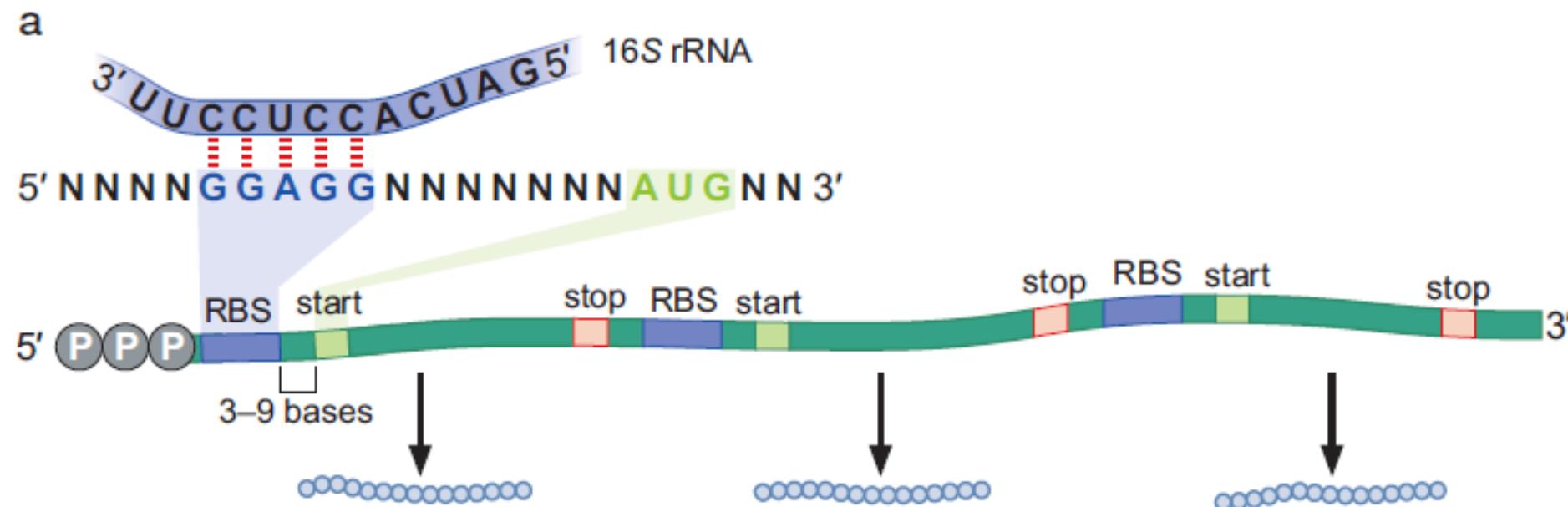
How many ORFs per mRNA?

- Polycistronic mRNA – more than one ORF
 - Prokaryotic mRNAs
 - Often encode proteins that perform related functions
- Monocistronic mRNA – one ORF
 - Eukaryotic mRNAs

Translation starts at the 5' end of the ORF and proceeds one codon at a time to the 3' end. The first and last codons of an ORF are known as the start and stop codons.

How many ORFs per mRNA?

- **Polycistronic mRNA** – more than one ORF



RBS – ribosome binding site or Shine-Dalgarno sequence

FIGURE 15-2 Structure of messenger RNA. (a) A polycistronic prokaryotic message with three ORFs. Each ribosome-binding site is indicated by a purple box labeled RBS. (b) A monocistronic eukaryotic message. The 5' cap is indicated by a "ball" at the end of the mRNA.

How many ORFs per mRNA?

- **Monocistronic mRNA** –one ORF

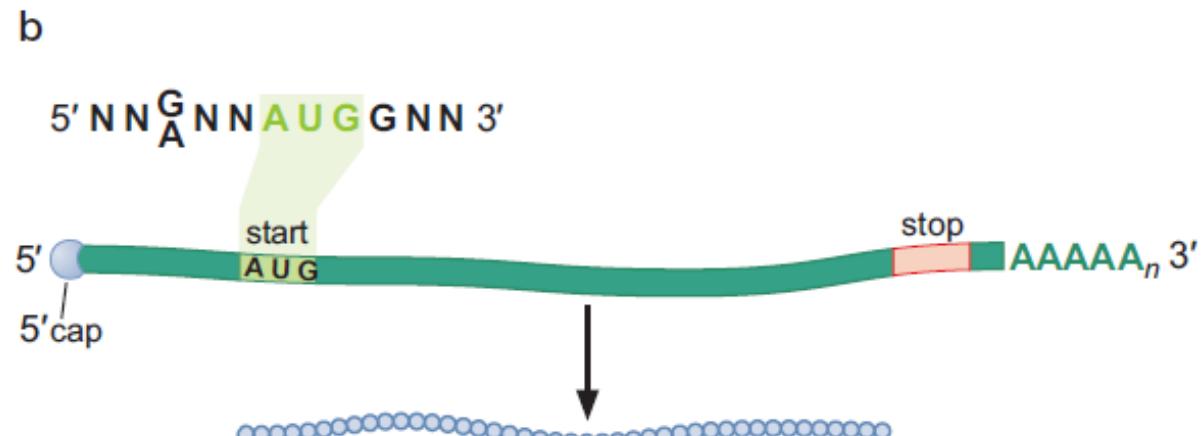


FIGURE 15-2 Structure of messenger RNA. (a) A polycistronic prokaryotic message with three ORFs. Each ribosome-binding site is indicated by a purple box labeled RBS. (b) A monocistronic eukaryotic message. The 5' cap is indicated by a "ball" at the end of the mRNA.

How to attract ribosome to the mRNA?

Prokaryotic mRNA

- Ribosome-binding site (RBS) or Shine-Dalgarno sequence
 - 3-9 nt upstream, complementary to a sequence stretch located near the 3' end of 16S rRNA
 - Prokaryotic RBS: 5'-AGGAGG-3'; the core of region of 16S rRNA: 5'-CCUCCU-3'
 - High complementarity and proper spacing = active translation
- Translational coupling
 - Start codon of the downstream ORF overlaps 3' end of the upstream ORF (5'-AUGA-3') because some prokaryotes ORFs lack the strong RBS
 - Translation of the downstream ORF requires translation of the upstream ORF

How to attract ribosome to the mRNA?

Eukaryotic mRNA

- 5' cap is required to recruit ribosome
 - Ribosome starts scanning in 5'->3' direction until it comes across first 5'-AUG-3' codon
 - Kozak sequence (5'-G/ANNAAUGG-3')
 - Underlined residues interact with initiator tRNA, not with rRNA
- Poly-A tail

Kozak consensus sequence (Kozak consensus or Kozak sequence) is a nucleic acid motif that functions as the protein translation initiation site in most eukaryotic mRNA transcripts.

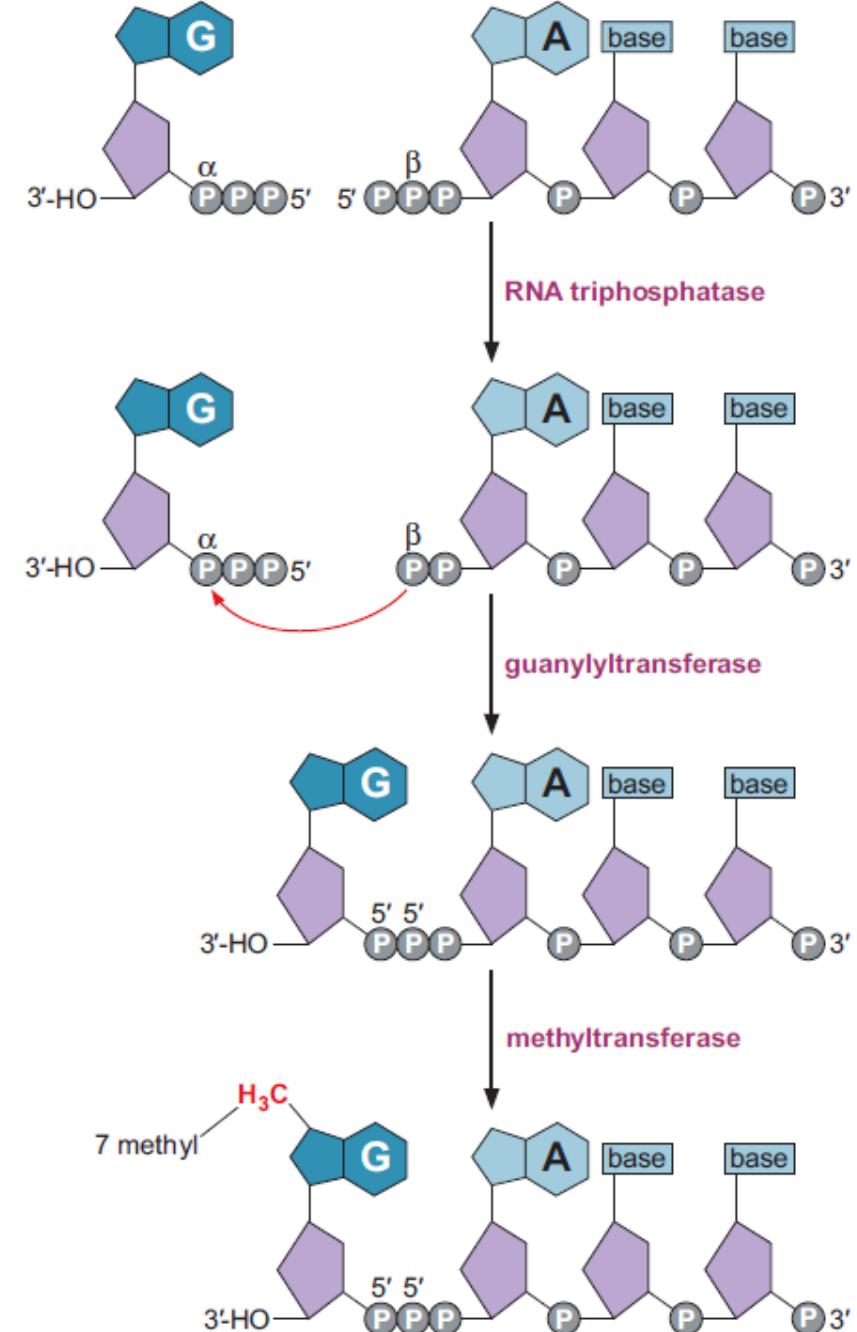
Capping the beginning

- Addition of modified guanine to the 5' end of the RNA
 - Unusual 5'-5' linkage involving three phosphates

1. Phosphate removal
2. GMP addition
3. G methylation

Result is the 5'cap structure ->
ribosome + mRNA ->
translation starting

FIGURE 13-24 The structure and formation of the 5' RNA cap. In the first step, the γ -phosphate at the 5' end of the RNA is removed by an enzyme called RNA triphosphatase (the initiating nucleotide of a transcript initially retains its α -, β -, and γ -phosphates). In the next step, the enzyme guanylyltransferase adds a GMP moiety to the resulting terminal β -phosphatase. This is a two-step process: first, an enzyme-GMP complex is generated from GTP with release of the β - and γ -phosphates of that GTP, and then the GMP from the enzyme is transferred to the β -phosphate of the 5' end of the RNA. Once this linkage is made, the newly added guanine and the purine at the original 5' end of the mRNA are further modified by the addition of methyl groups by methyltransferase. The resulting 5' cap structure subsequently recruits the ribosome to the mRNA for translation to begin (see Chapter 15).



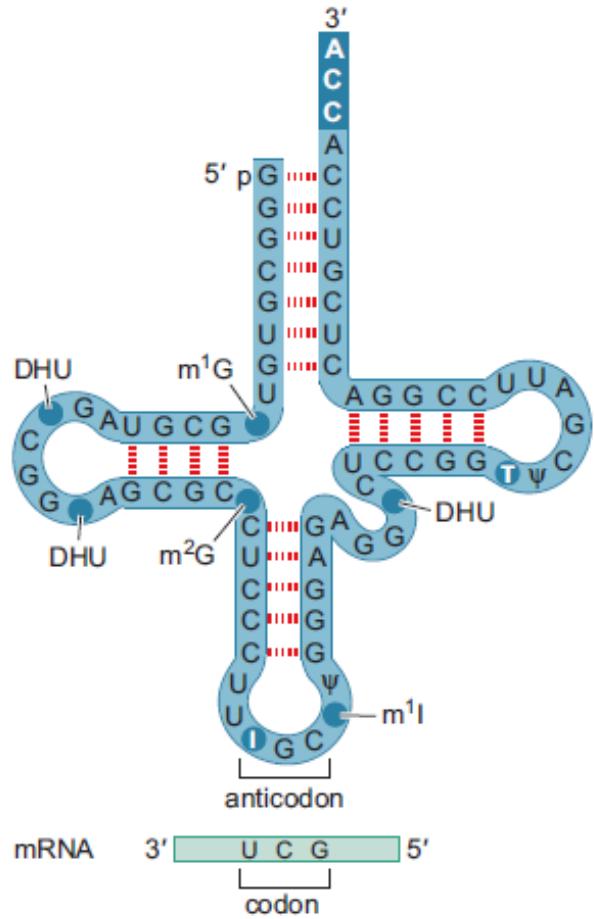


FIGURE 2-14 Yeast alanine tRNA structure, as determined by Robert W. Holley and his associates. The anticodon in this tRNA recognizes the codon for alanine in the mRNA. Several modified nucleosides exist in the structure: ψ = pseudouridine, T = ribothymidine, DHU = 5,6-dihydrouridine, I = inosine, m^1G = 1-methylguanosine, m^1I = 1-methylinosine, and m^2G = *N,N*-dimethylguanosine.

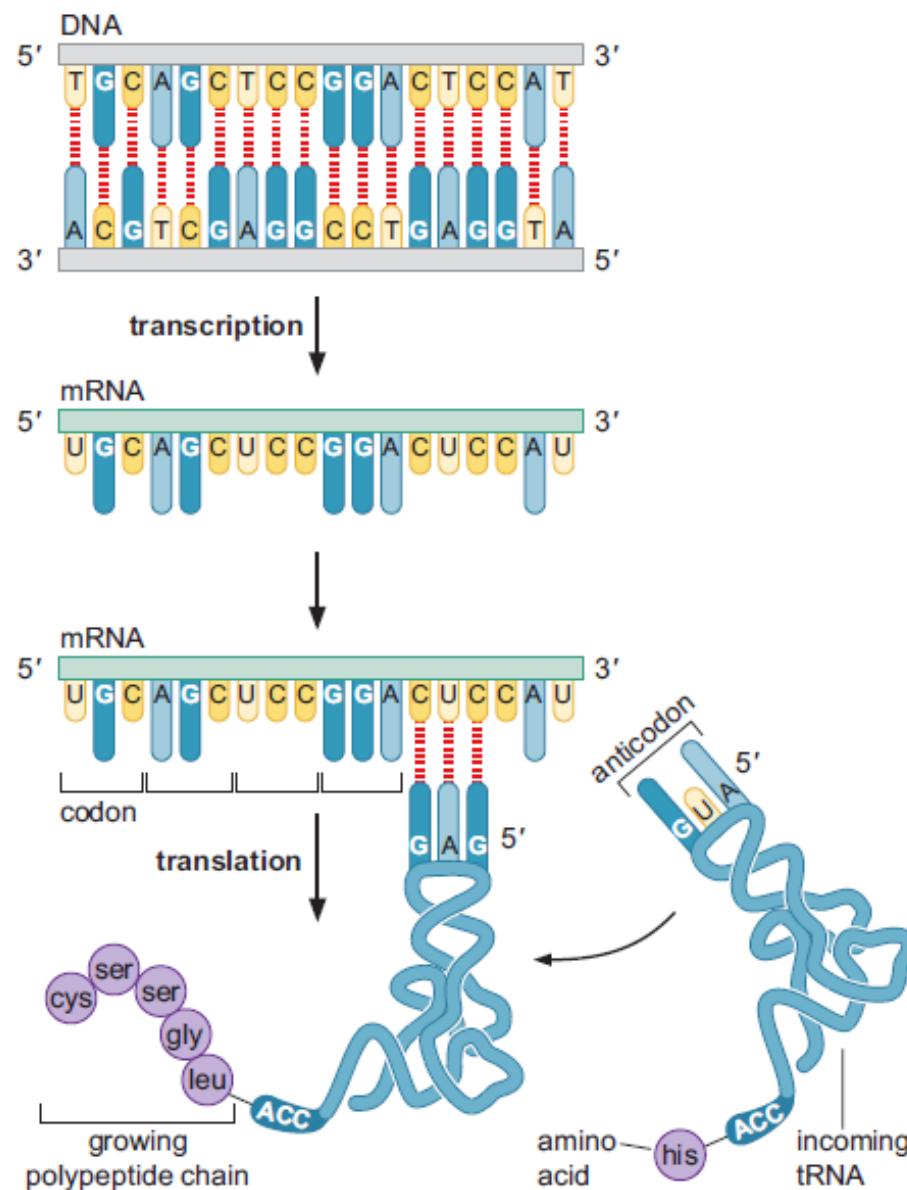


FIGURE 2-15 Transcription and translation. The nucleotides of mRNA are assembled to form a complementary copy of one strand of DNA. Each group of three is a codon that is complementary to a group of three nucleotides in the anticodon region of a specific tRNA molecule. When base pairing occurs, an amino acid carried at the other end of the tRNA molecule is added to the growing protein chain.



tRNA

- 75-95 nucleotides long
- Common features
 - 5'-CCA-3' terminus (via CCA-adding enzyme)
 - Presence of several unusual bases
 - Secondary structure

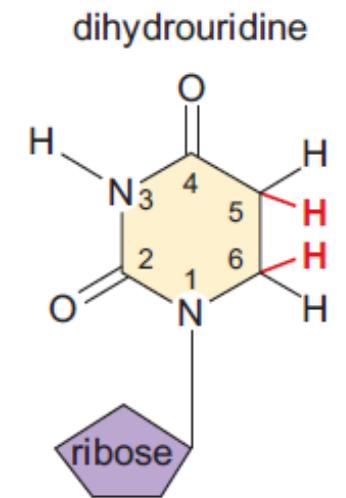
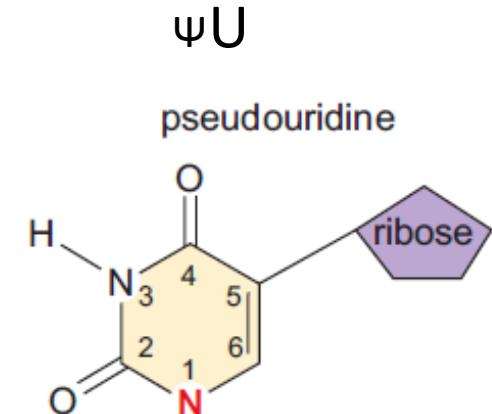
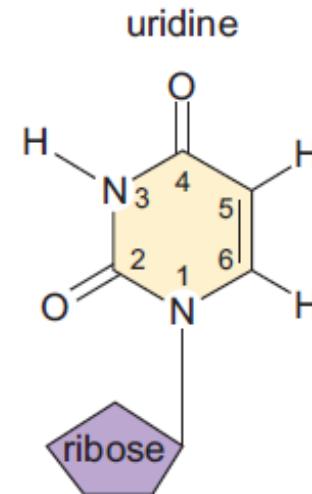
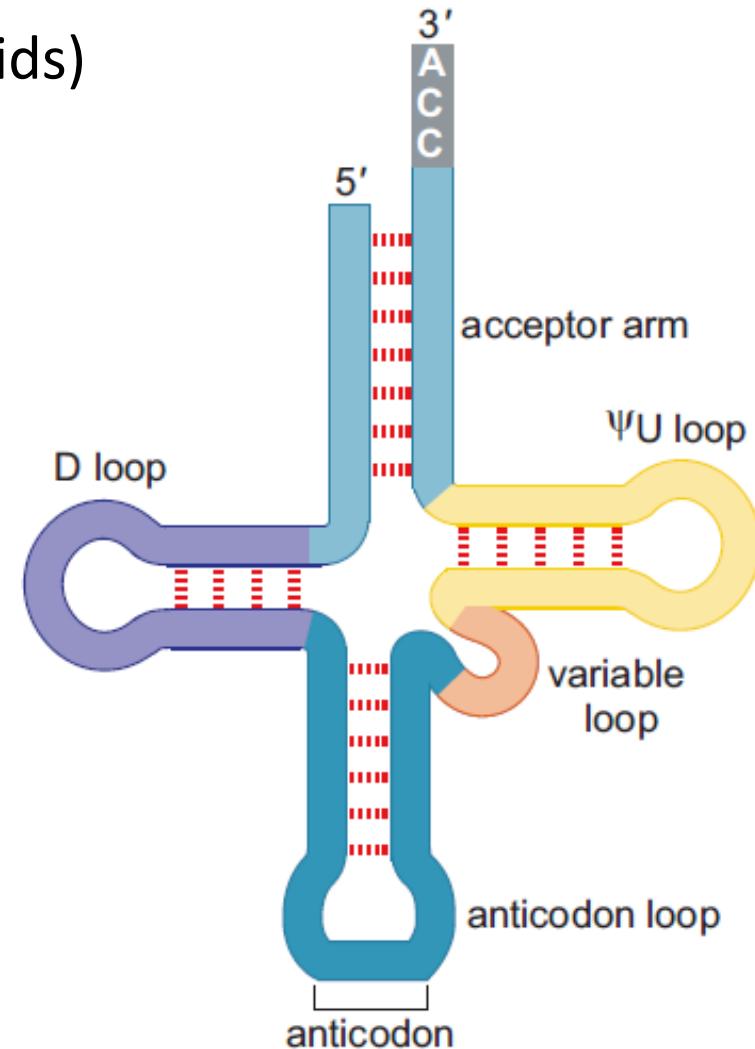


FIGURE 15-3 A subset of modified nucleosides found in tRNA. Uridine and two uridine-related nucleotides are shown.

tRNA has a cloverleaf structure

- Principal features of the tRNA are:
 - **Acceptor stem** (the site of attachment of the amino acids)
 - **Ψ U loop** (psii-uridine or pseudouridine - the modified base is often found within the sequence 5'-T Ψ UCG-3')
 - **D loop** (its name from the characteristic presence of dihydrouridines in the loop)
 - **Anticodon loop** (contains the anticodon, a three-nucleotide-long sequence that is responsible for recognizing the codon by base pairing with the mRNA. The anticodon is always bracketed on the 3' end by a purine and on its 5' end by uracil)
 - **Variable loop** (sits between the anticodon loop and the Ψ U loop and, as its name implies, varies in size from 3 to 21 bases)

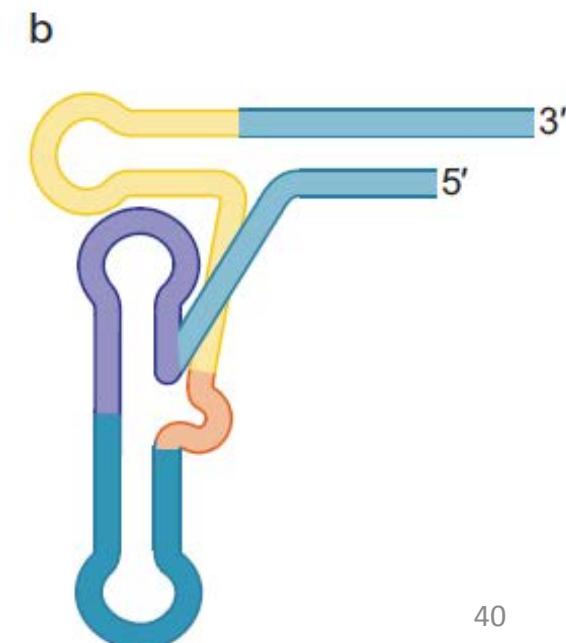
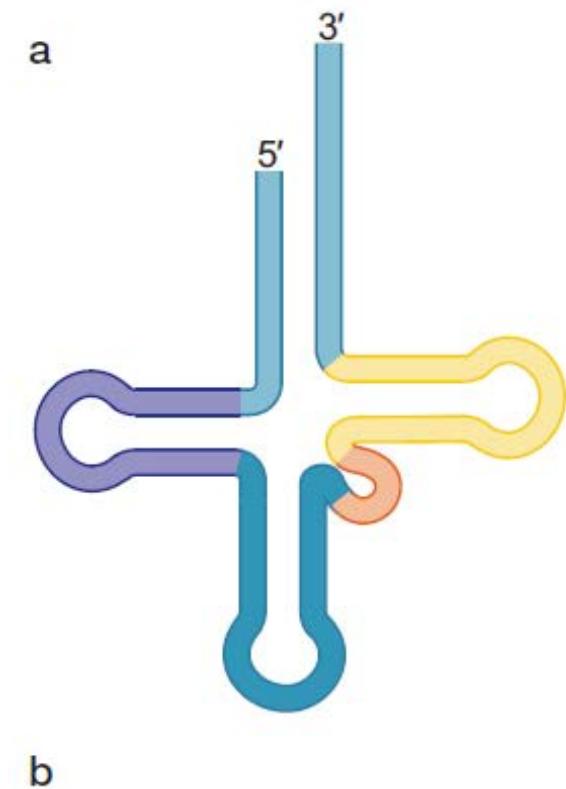
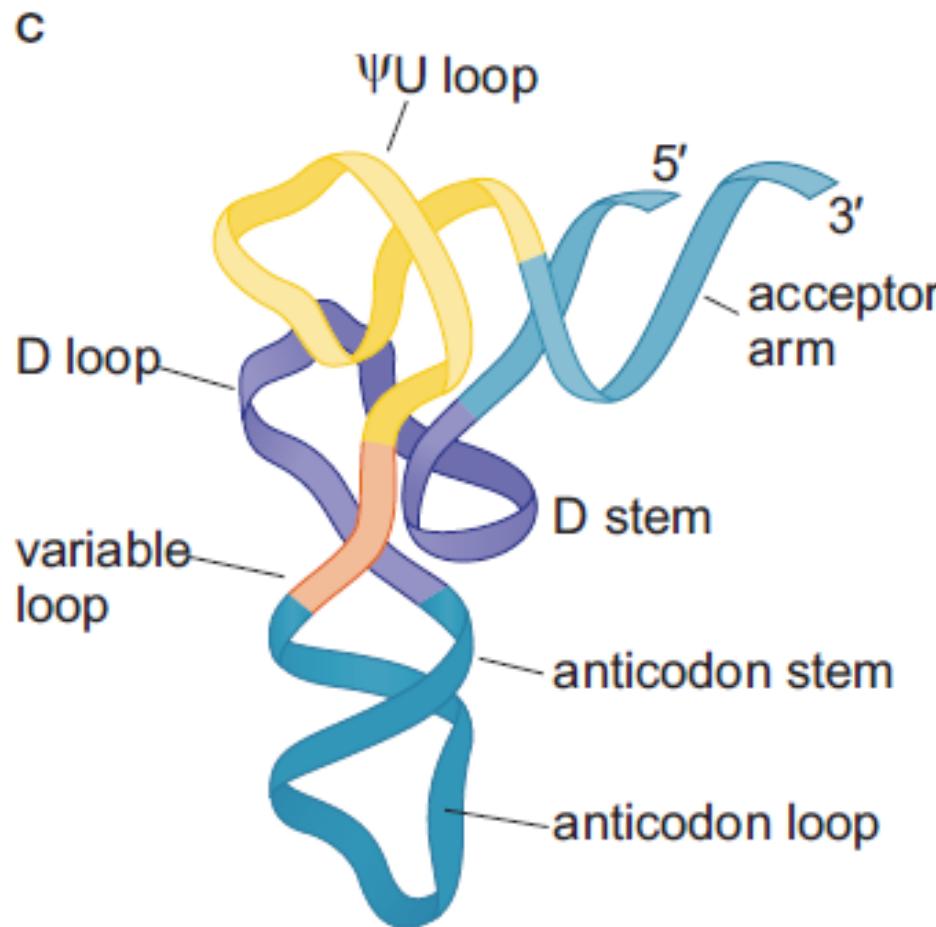
FIGURE 15-4 Cloverleaf representation of the secondary structure of tRNA. In this representation of a tRNA, the base pairings between different parts of the tRNA are indicated by the dotted red lines.



From cloverleaf to L-shaped

A – cloverleaf
B- L-shaped
C – ribbon structure

FIGURE 15-5 Conversion between the cloverleaf and the actual 3D structure of a tRNA. (a) Cloverleaf representation. (b) L-shaped representation showing the location of the base-paired regions of the final folded tRNA. (c) Ribbon representation of the actual folded structure of a tRNA. Note that although this diagram illustrates how the actual tRNA structure is related to the cloverleaf representation, a tRNA does not attain its final structure by first base pairing and then folding into an L shape.



Nothing happens without a charge

Adenylylation, more commonly known as AMPylation, is a process in which an adenosine monophosphate (AMP) molecule is covalently attached to the amino acid side chain of a protein

- **tRNAs need to be charged**, ie. amino acid attached to them
 - High energy acyl linkage between the carboxyl group of the amino acid and the 2'- or 3'-OH group of the tRNA CCA terminus
- **Two-step process by aminoacyl-tRNA synthetases**
 - **Adenylylation** – amino acid reacts with ATP (AMP is added, pyrophosphate released) **Adenylylation refers to transfer of AMP, as opposed to adenylation, which would indicate the transfer of adenine**
 - **tRNA charging** – adenylylated amino acid reacts with tRNA, release of AMP
tRNA molecules to which an amino acid is attached are said to be charged, and tRNAs that lack an amino acid are said to be uncharged.

Charging the tRNA, step I

First step is **adenylylation** in which the amino acid reacts with ATP to become adenylylated with the concomitant release of pyrophosphate.

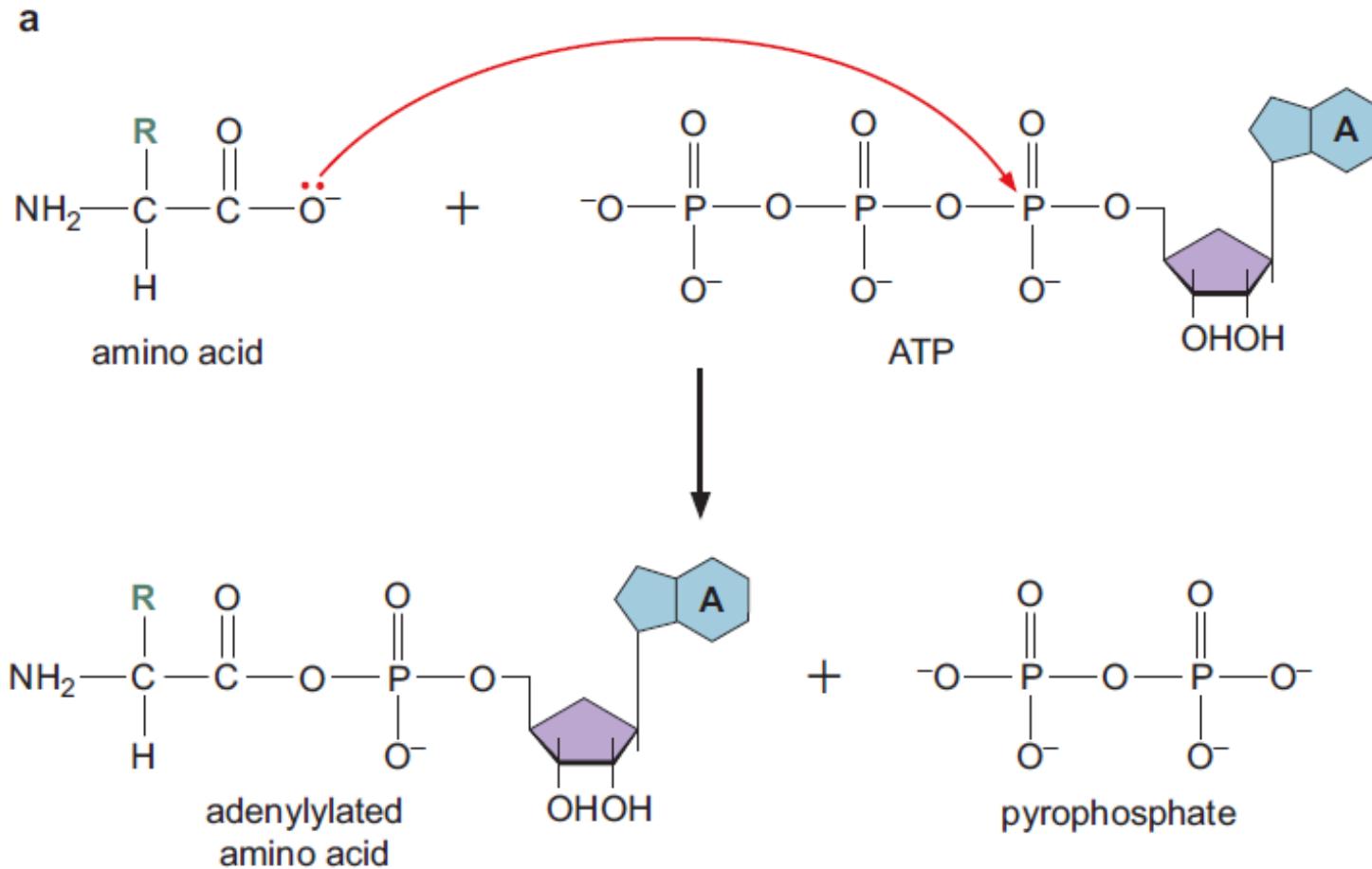
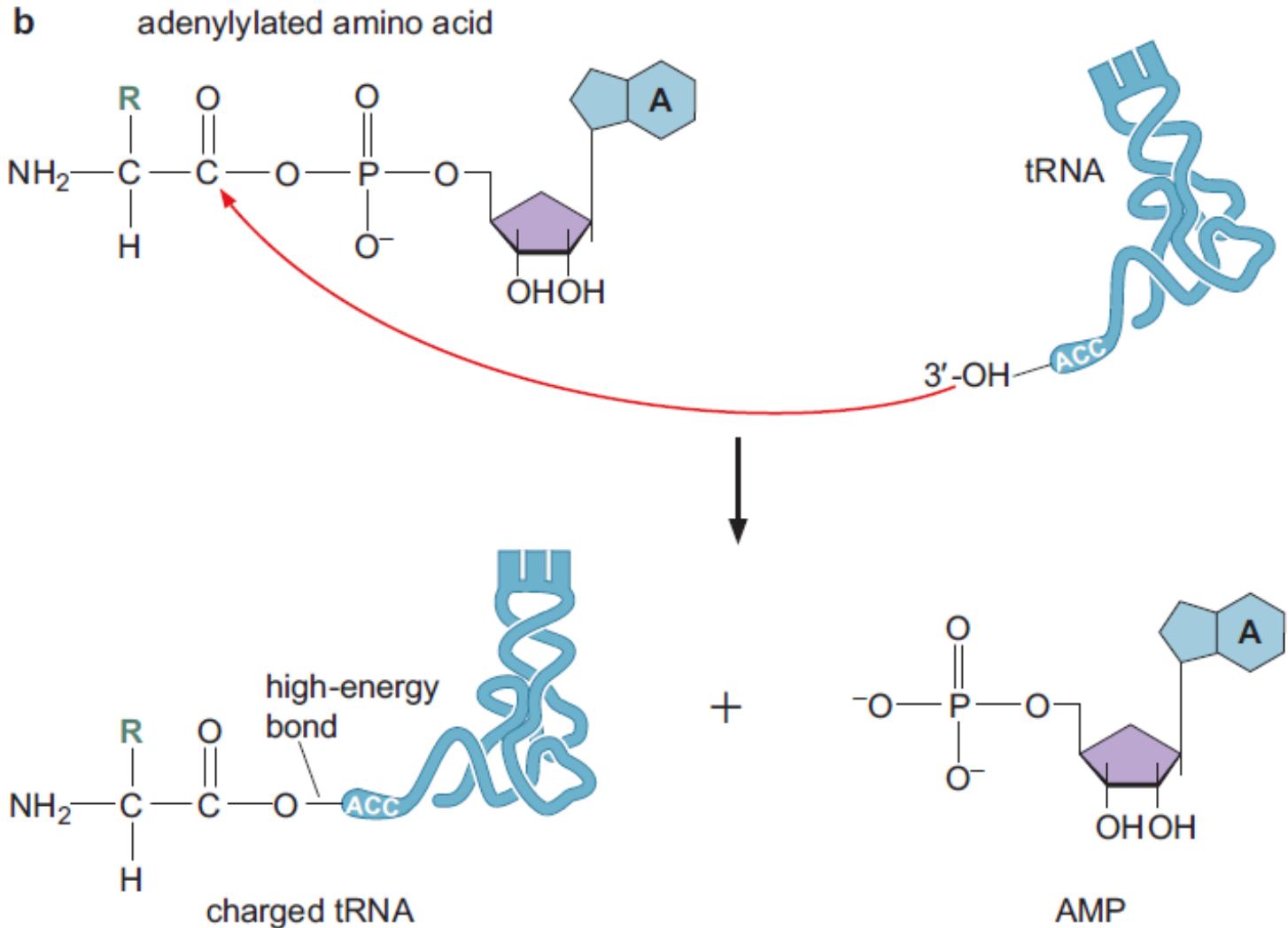


FIGURE 15-6 The two steps of aminoacyl-tRNA charging. (a) Adenylylation of amino acid. (b) Transfer of the adenylylated amino acid to tRNA. The process shown is for a class II tRNA synthetase (which attaches the amino acid to the 3'-OH).

Charging the tRNA, step II



Step two is **tRNA charging** in which the adenylylated amino acid, which **remains tightly bound to the synthetase**, reacts with tRNA. This reaction results in **the transfer of the amino acid to the 3' end of the tRNA via the 2'- or 3'- hydroxyl** and the **release of AMP**.

FIGURE 15-6 The two steps of aminoacyl-tRNA charging. (a) Adenylylation of amino acid. (b) Transfer of the adenylylated amino acid to tRNA. The process shown is for a class II tRNA synthetase (which attaches the amino acid to the 3'-OH).

tRNA synthetases

- Each of the 20 amino acids is added to the tRNA by a single, dedicated **tRNA synthetase**
 - One synthetase recognizes and charges more than one tRNA (**isoaccepting tRNAs**)
- An aminoacyl-tRNA synthetase can never attach more than one kind of amino acid to a given tRNA!

Two classes of synthetases

- Class I
 - Generally monomeric, adds amino acid to 2'-OH of the tRNA
- Class II
 - Generally di/tetrameric, adds amino acid to 3'-OH of the tRNA

TABLE 15-1 Classes of Aminoacyl-tRNA Synthetases

Class II	Quaternary Structure	Class I	Quaternary Structure
Gly	($\alpha_2\beta_2$)	Glu	(α)
Ala	(α_4)	Gln	(α)
Pro	(α_2)	Arg	(α)
Ser	(α_2)	Cys	(α_2)
Thr	(α_2)	Met	(α_2)
His	(α_2)	Val	(α)
Asp	(α_2)	Ile	(α)
Asn	(α_2)	Leu	(α)
Lys	(α_2)	Tyr	(α)
Phe	($\alpha_2\beta_2$)	Trp	(α)

Adapted, with permission, from Delarue M. 1995. *Curr. Opin. Struct. Biol.* 5: 48–55, Table 1. © Elsevier.

Class I enzymes are generally monomeric, whereas class II enzymes are dimeric or tetrameric, with residues from two subunits contributing to the binding site for a single tRNA. α and β refer to subunits of the tRNA synthetases, and the subscripts indicate their stoichiometry.

Tough life of an aminoacyl-tRNA synthetase

- Aminoacyl-tRNA synthetases have two crucial challenges
 - They must **recognize correct tRNAs for a particular amino acid**
 - They must **charge all isoaccepting tRNAs with correct amino acid**

How does tRNA synthetase recognize correct tRNA?

- Two distant sites
 - **Acceptor stem (discriminator base)**

The **acceptor stem** is an especially important determinant for the specificity of tRNA synthetase recognition

- **Anticodon loop**

The **anticodon** dictates the amino acid that the tRNA is responsible for incorporating into the growing polypeptide chain

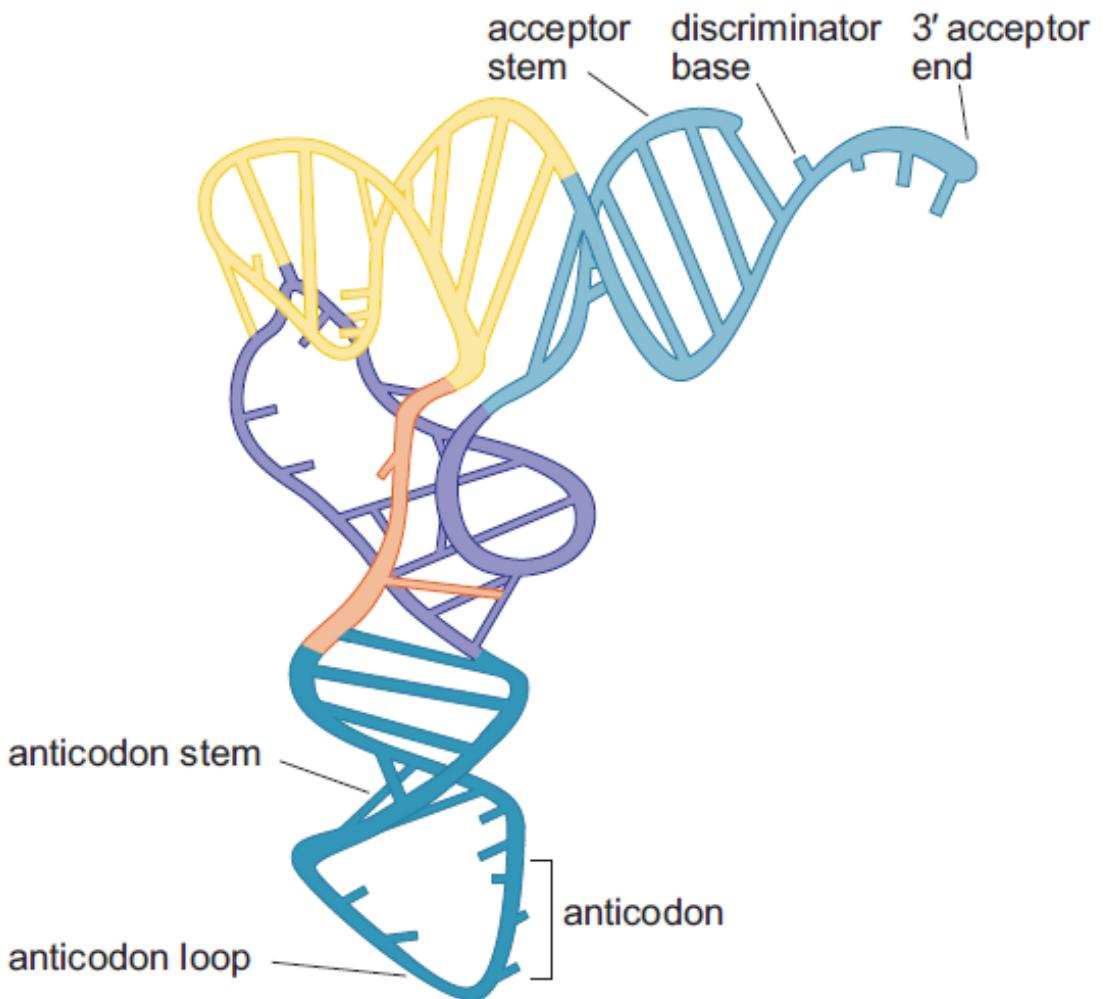


FIGURE 15-7 Structure of tRNA: elements required for aminoacyl synthetase recognition.

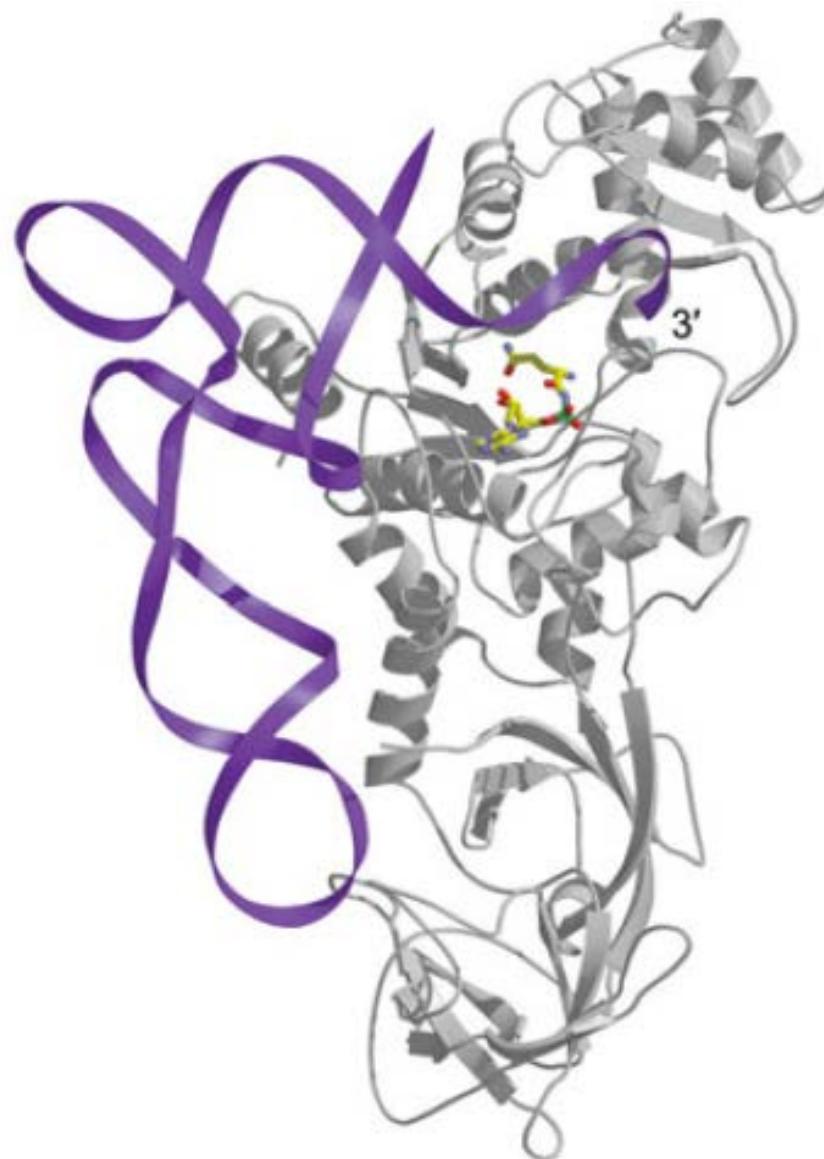
How does tRNA synthetase recognize correct tRNA?

Aminoacyl-tRNA formation is very accurate

The reason for this is the **relatively small size of amino acids** and, in some cases, **their similarity**.

Despite this challenge, the frequency of mischarging is very low; typically, less than 1 in 1000 tRNAs is charged with the incorrect amino acid.

FIGURE 15-8 Cocrystal structure of glut-aminyl aminoacyl-tRNA synthetase with tRNA^{Gln}. Enzyme (gray); tRNA^{Gln} (purple). The yellow, red, and green molecule is glutaminyl-AMP. Note the proximity of this molecule to the 3' end of the tRNA and the points of contact between the tRNA and the synthetase. (Rath V.L. et al. 1998. *Structure* 6: 439–449.) Image prepared with MolScript, BobScript, and Raster3D.



Why can't only the anticodon loop be used for tRNA discrimination?

- Up to six codons for one amino acid (serine for example)
 - Serine – 5'-AGC-3' and 5'-UCA-3', they are completely different from one another

Charging is complex but accurate

- Less than 1 in 1000 tRNAs are charged incorrectly
- **Catalytic pocket**
 - Discriminates between correct and incorrect amino acid
- **Editing pocket**
 - Isoleucyl-tRNA synthetase can remove too small an amino acid
- Discriminate twice: initial binding (1:100) and editing (1:100)

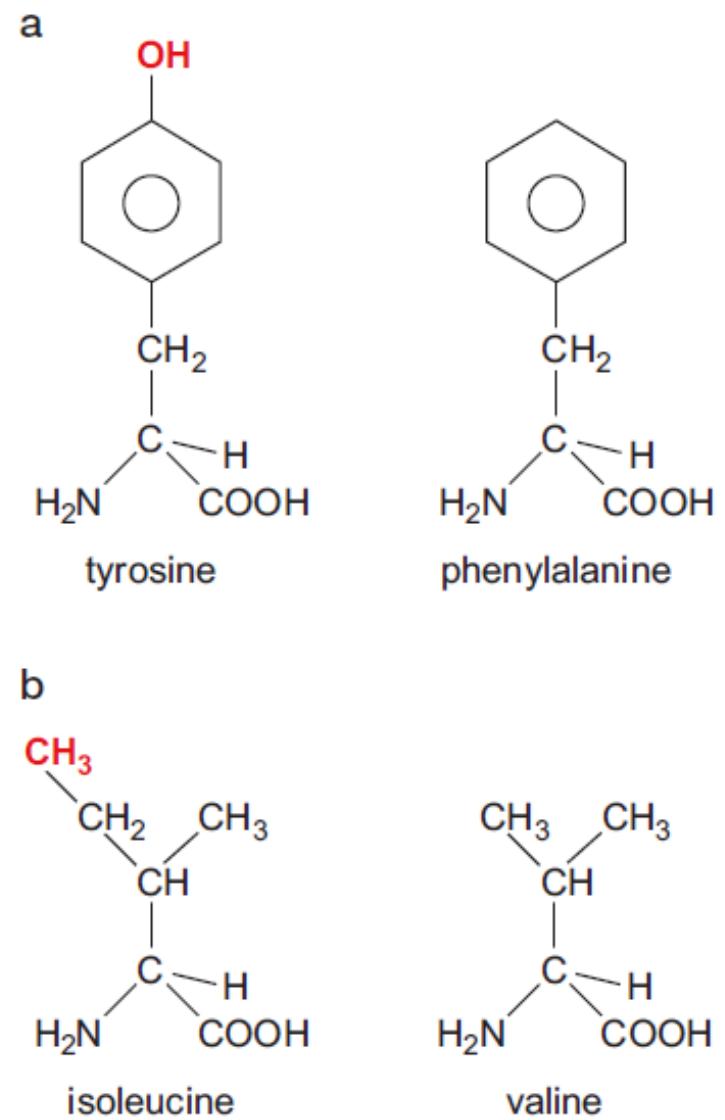


FIGURE 15-9 Distinguishing features of similar amino acids.

Ribosome is blind, must therefore trust others

- Ribosome cannot distinguish between correctly and incorrectly charged tRNAs, only checks for proper codon-anticodon interaction
- Translation machinery relies on the high fidelity of the aa-tRNA synthetases to ensure accurate decoding

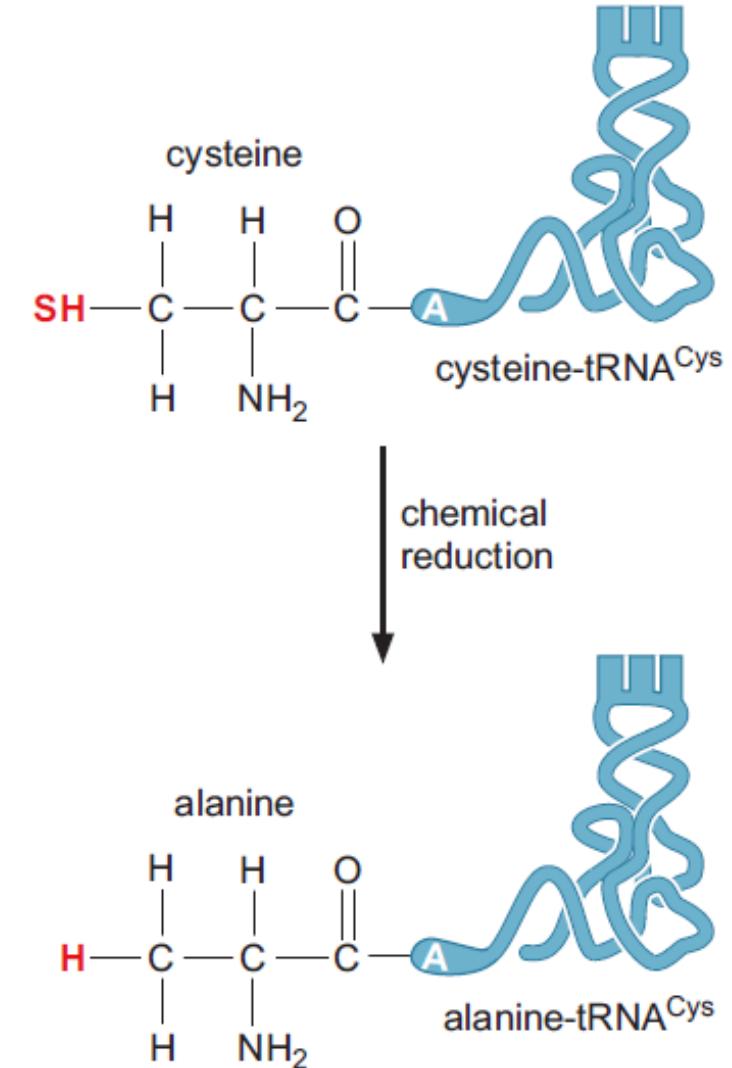


FIGURE 15-10 Chemical reduction of cysteine-tRNA^{Cys} to alanine-tRNA^{Cys}.