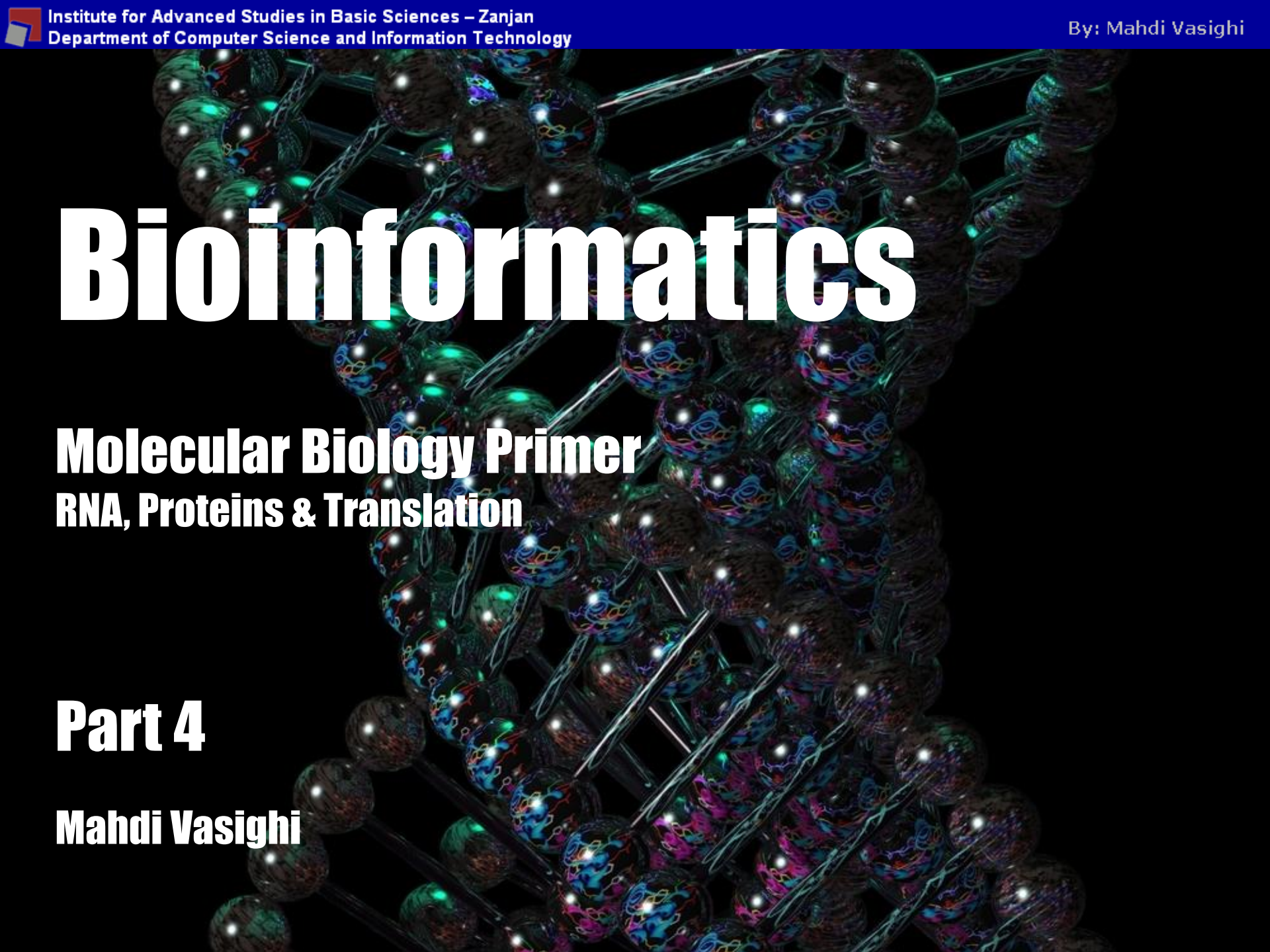


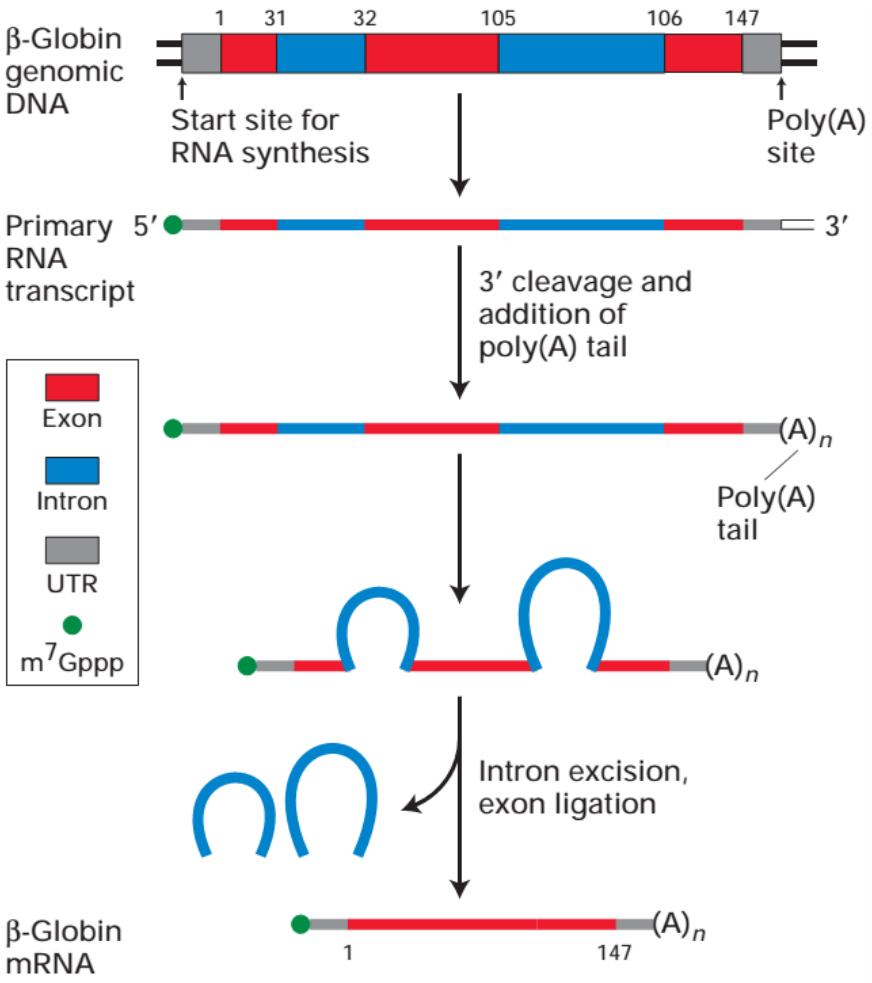
Bioinformatics

Molecular Biology Primer
RNA, Proteins & Translation

Part 4

Mahdi Vasighi

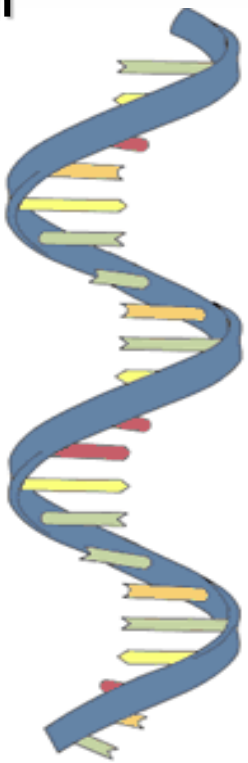
RNA splicing

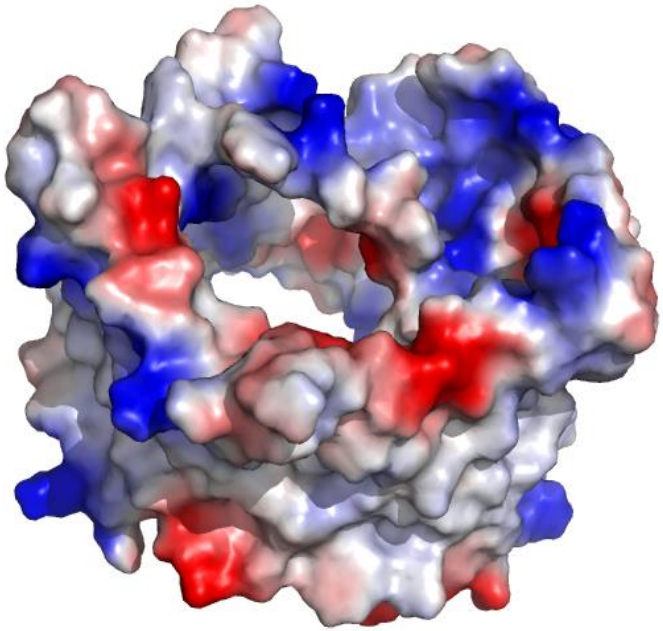


Alternative splicing is an important mechanism for production of different forms of a protein from a single gene, called **isoforms**.

Central dogma of molecular biology

DNA → (Transcription) → RNA → (Translation) → Protein



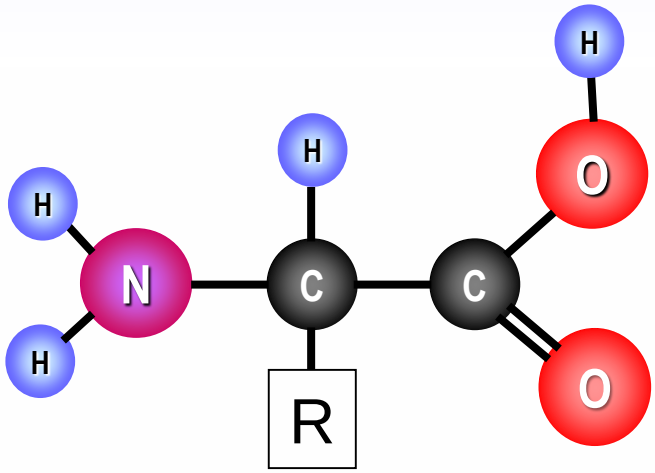


Proteins

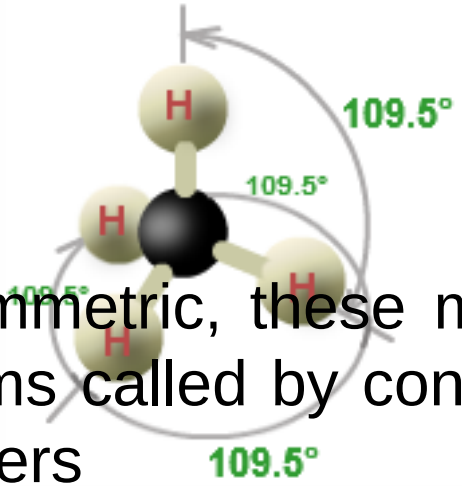
Proteins (polypeptides) are organic compounds made of amino acids arranged in a linear chain and folded into a globular form.

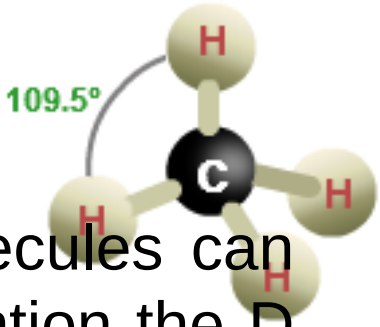
Amino acids are molecules containing:

- a central α-carbon
- an amine group (NH_2)
- a carboxylic acid group (COOH)
- a side-chain (R) that varies between different amino acids.
- a hydrogen atom

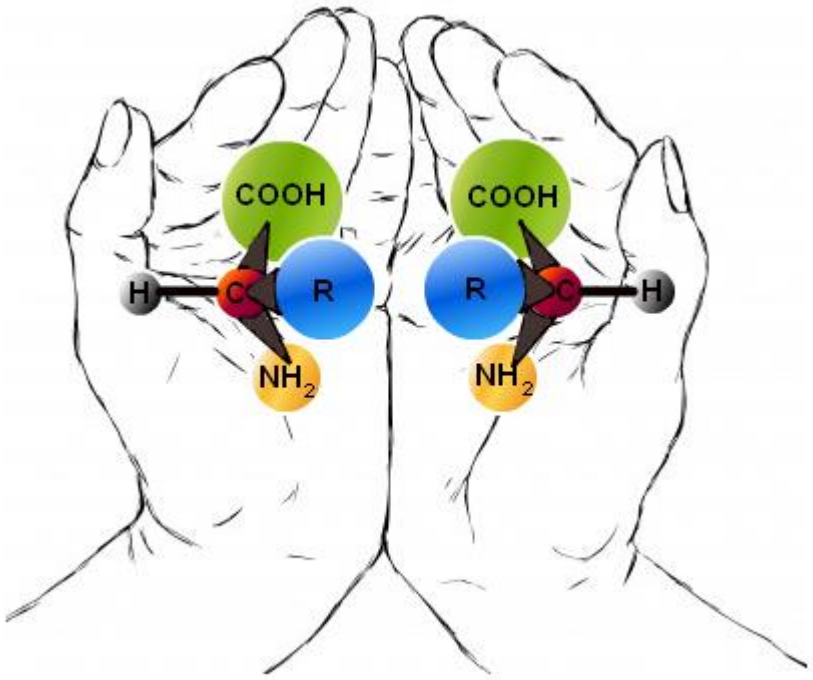


Proteins

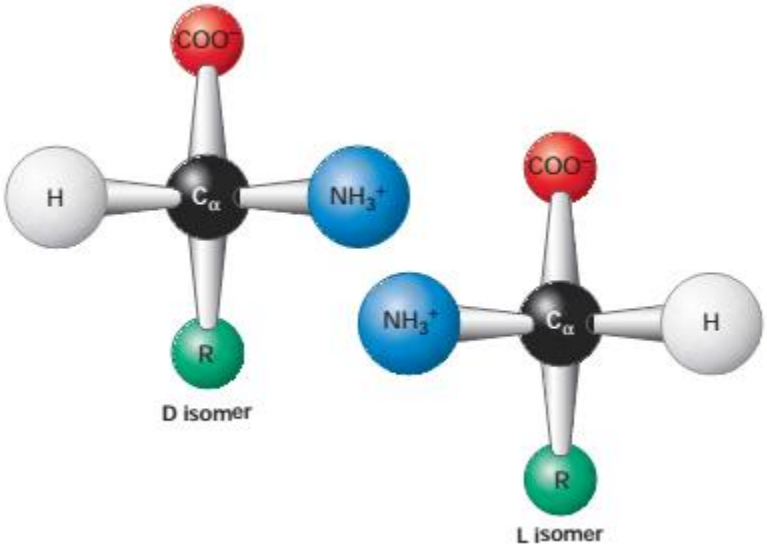




The central α-carbon is asymmetric, these molecules can exist in two mirror-image forms called by convention the D (dextro) and the L (levo) isomers

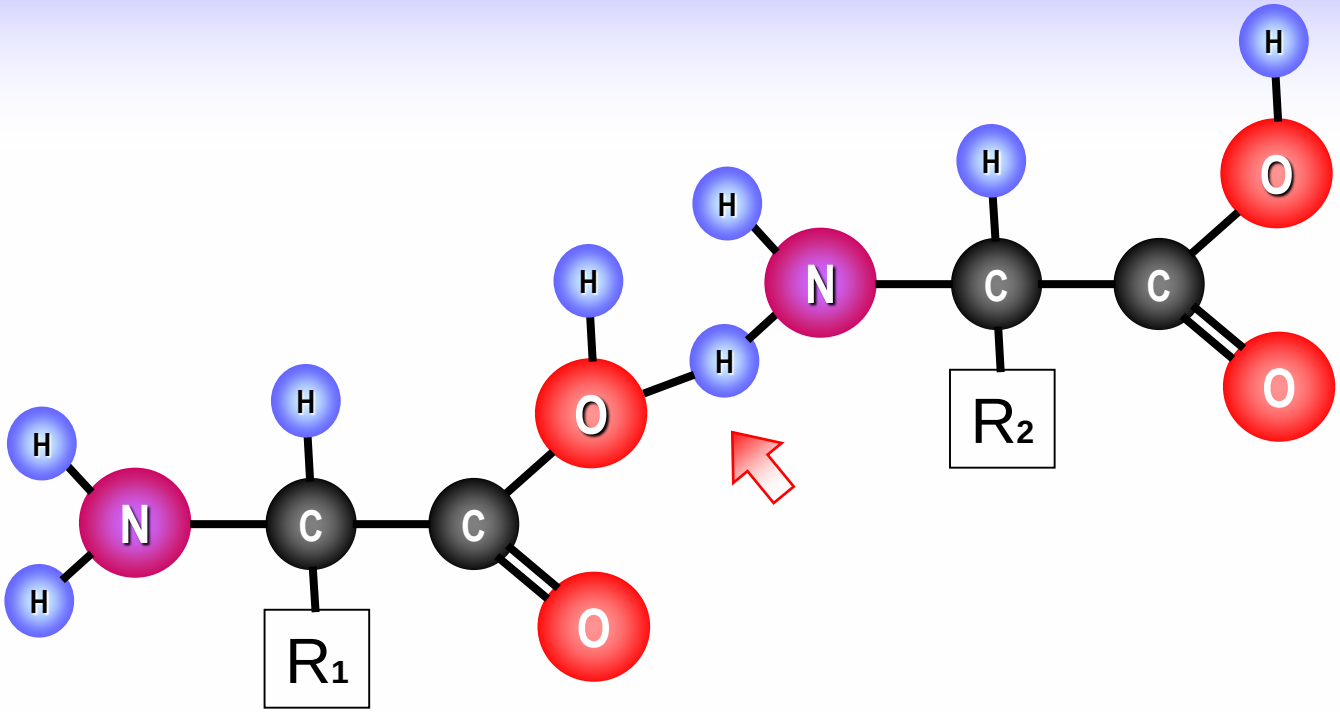


- Chirality and enantiomers
- With rare exceptions, only the L forms of amino acids are found in proteins.



Proteins

Most proteins are linear polymers built from series of different amino acids.



Proteins

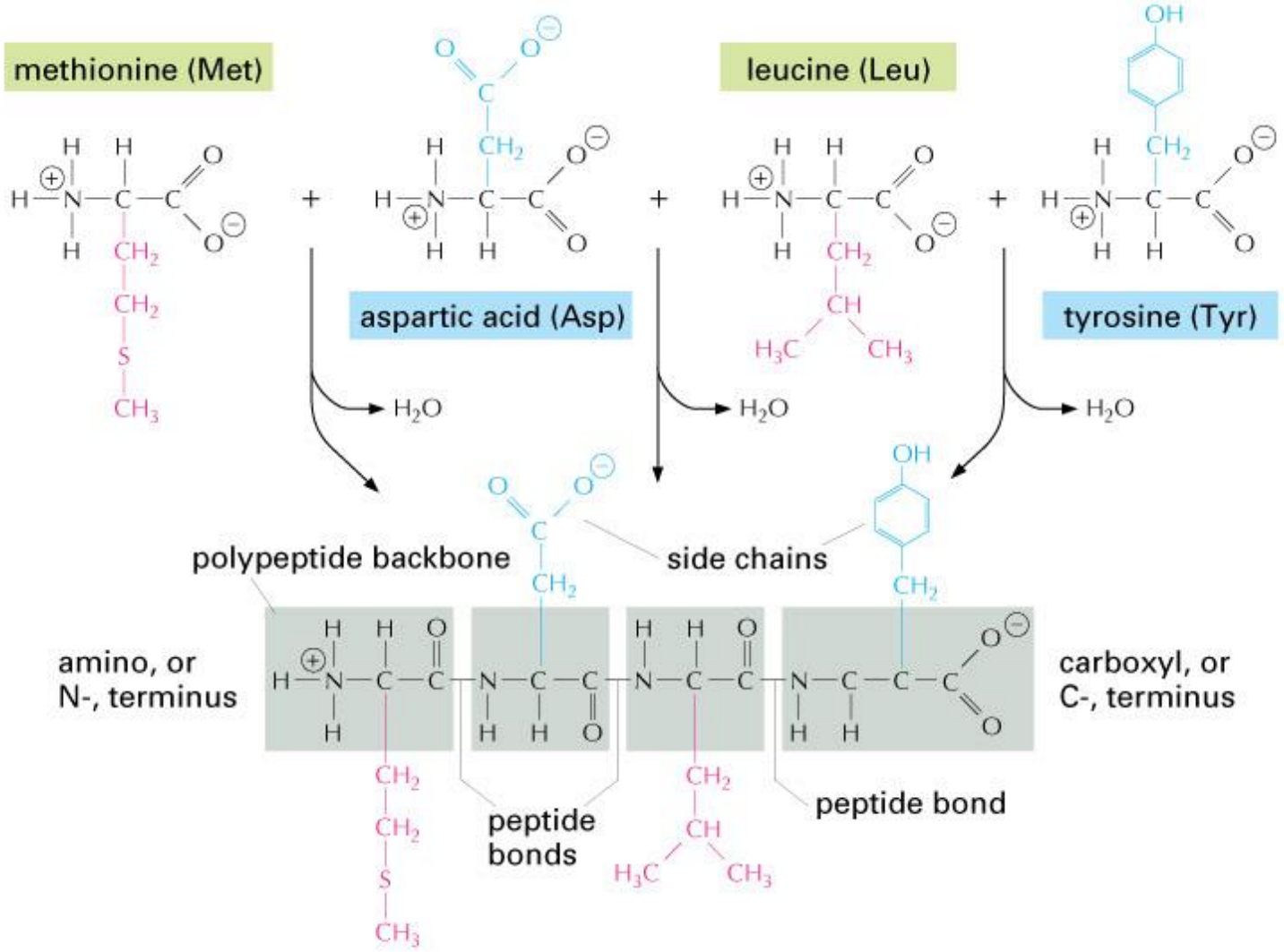
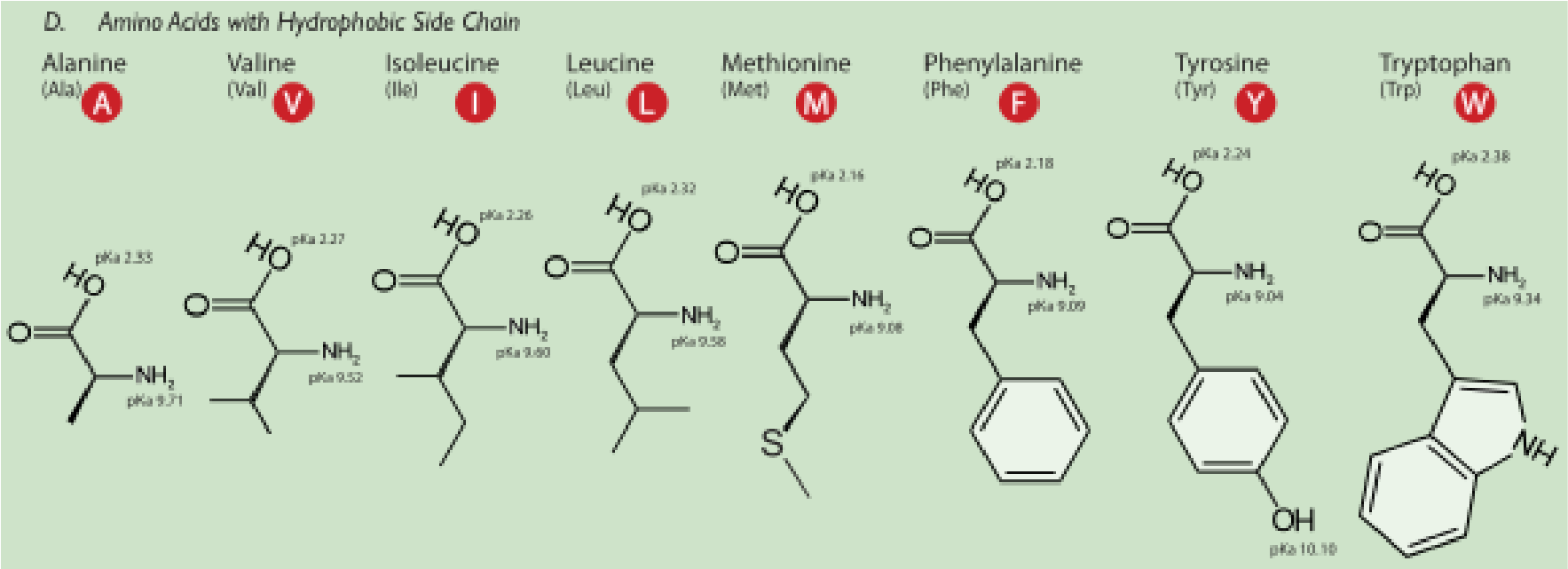
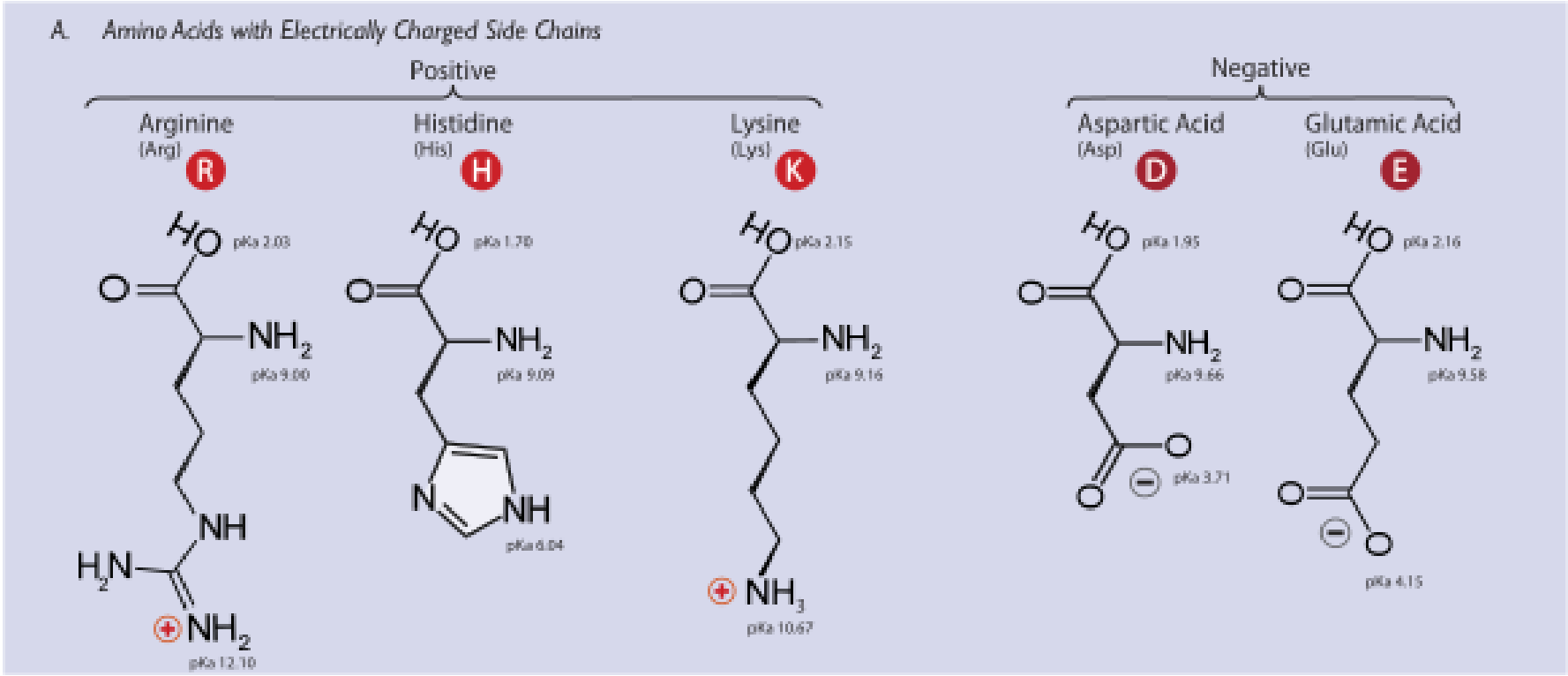


Figure 4-2 Essential Cell Biology, 2/e. (© 2004 Garland Science)

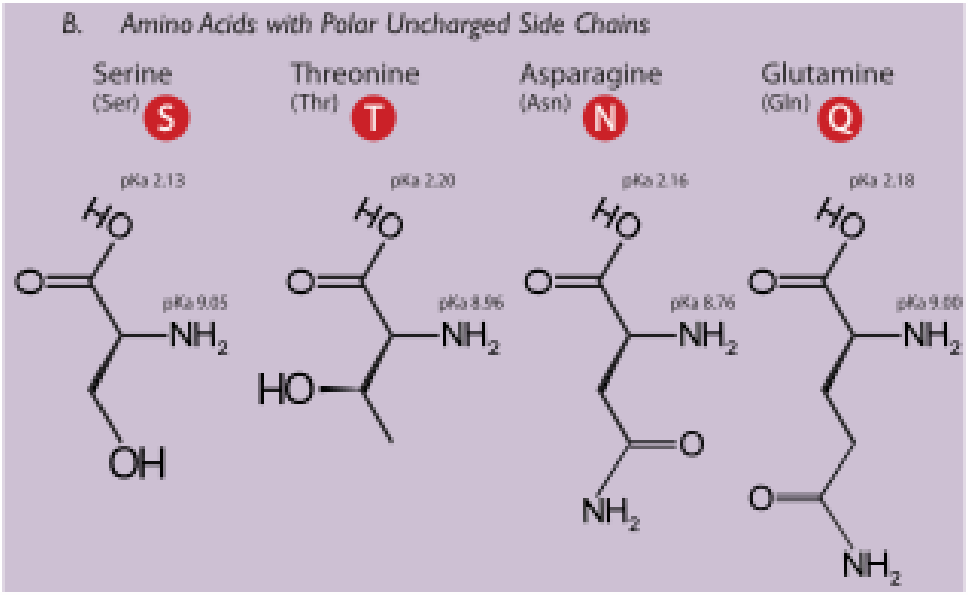
Proteins

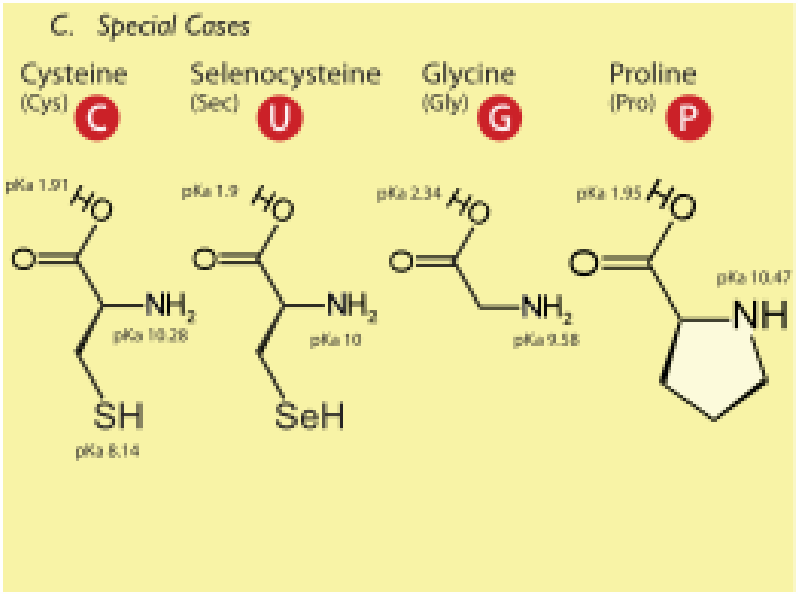

D. Amino Acids with Hydrophobic Side Chain
Alanine (Ala) A
Valine (Val) V
Isoleucine (Ile) I
Leucine (Leu) L
Methionine (Met) M
Phenylalanine (Phe) F
Tyrosine (Tyr) Y
Tryptophan (Trp) W

Proteins


A. Amino Acids with Electrically Charged Side Chains
Positive
Negative
Arginine
(Arg)
R
Histidine
(His)
H
Lysine
(Lys)
K
Aspartic Acid
(Asp)
D
Glutamic Acid
(Glu)
E
pKa 2.03
pKa 9.00
pKa 12.10
pKa 1.70
pKa 9.09
pKa 6.04
pKa 2.15
pKa 9.16
pKa 10.67
pKa 1.95
pKa 9.66
pKa 3.71
pKa 2.16
pKa 9.58
pKa 4.15

Proteins


B. Amino Acids with Polar Uncharged Side Chains
Serine (Ser) S
Threonine (Thr) T
Asparagine (Asn) N
Glutamine (Gln) Q
pKa 2.13
pKa 9.05
pKa 2.20
pKa 8.96
pKa 2.16
pKa 8.76
pKa 2.18
pKa 9.00


C. Special Cases
Cysteine (Cys) C
Selenocysteine (Sec) U
Glycine (Gly) G
Proline (Pro) P
pKa 1.91
pKa 10.28
pKa 8.14
pKa 1.9
pKa 10
pKa 2.34
pKa 9.58
pKa 1.95
pKa 10.47

Proteins

Proteins can be grouped into several broad functional classes:

- Structural proteins, which provide structural rigidity to the cell.
- Transport proteins, which control the flow of materials across cellular membranes.
- Regulatory proteins, which act as sensors and switches to control protein activity and gene function.
- Signaling proteins, including cell surface receptors and other proteins that transmit external signals to the cell interior.
- Motor proteins, which cause motion.

Essential Amino Acids

Translation

RNA ⟶ **Protein**
Translation

Translation is the second process of **protein biosynthesis** (part of the overall process of **gene expression**). In translation, Messenger RNA (mRNA) is **decode**d to produce a specific polypeptide according to the rules specified by the **genetic code**.

For translation Three different type of RNA is required:

- messenger RNA (mRNA)
- transfer RNA (tRNA)
- ribosomal RNA (rRNA)

Translation

- **Messenger RNA (mRNA)**

Each 3 Nucleotides is a triplet or CODON (which code for a specific AMINO ACID)

Start codon

mRNA

Stop codon

A	U	G	G	G	C	U	C	C	A	U	C	G	G	C	G	C	A	U	A	A

Codon 1 Codon 2 Codon 3 Codon 4 Codon 5 Codon 6 Codon 7

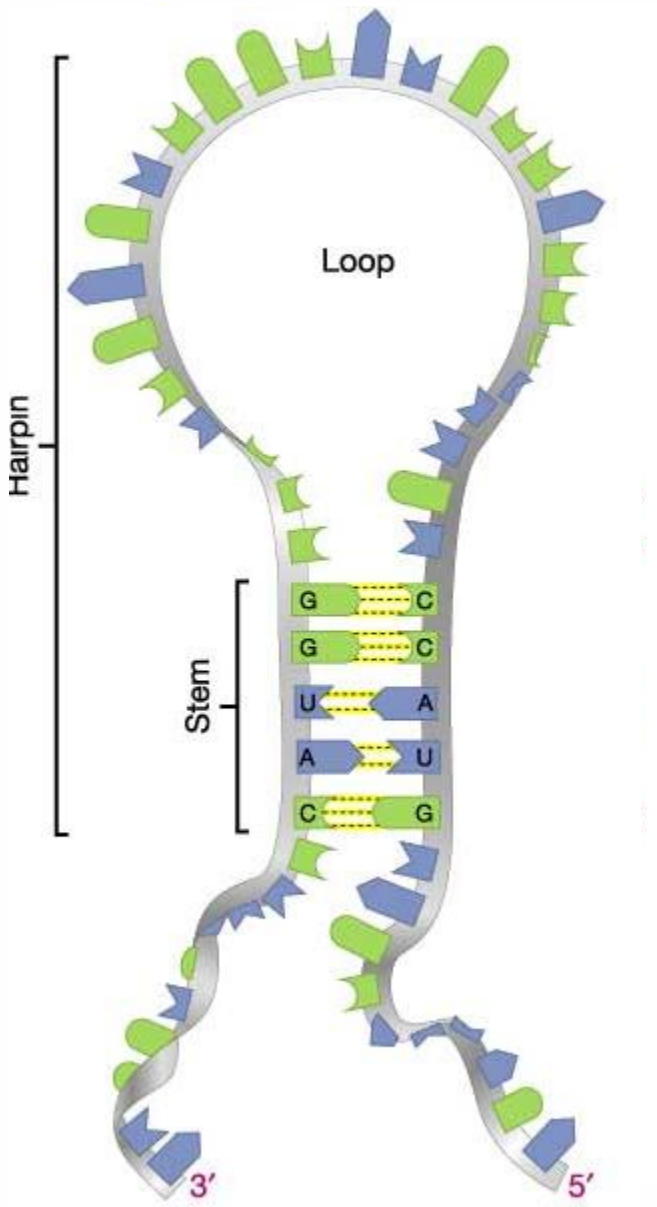
The sequence of codons that runs from a specific start codon to a stop codon is called a **reading frame**.

Translation

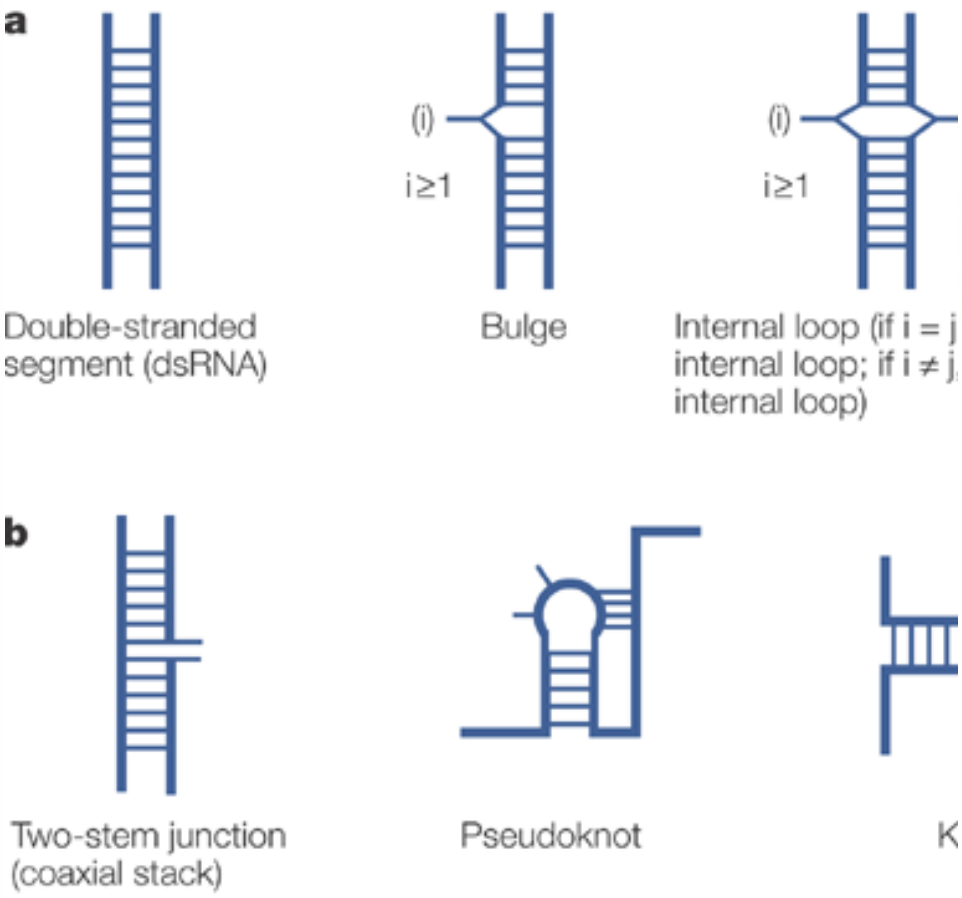
- **RNA secondary structure**

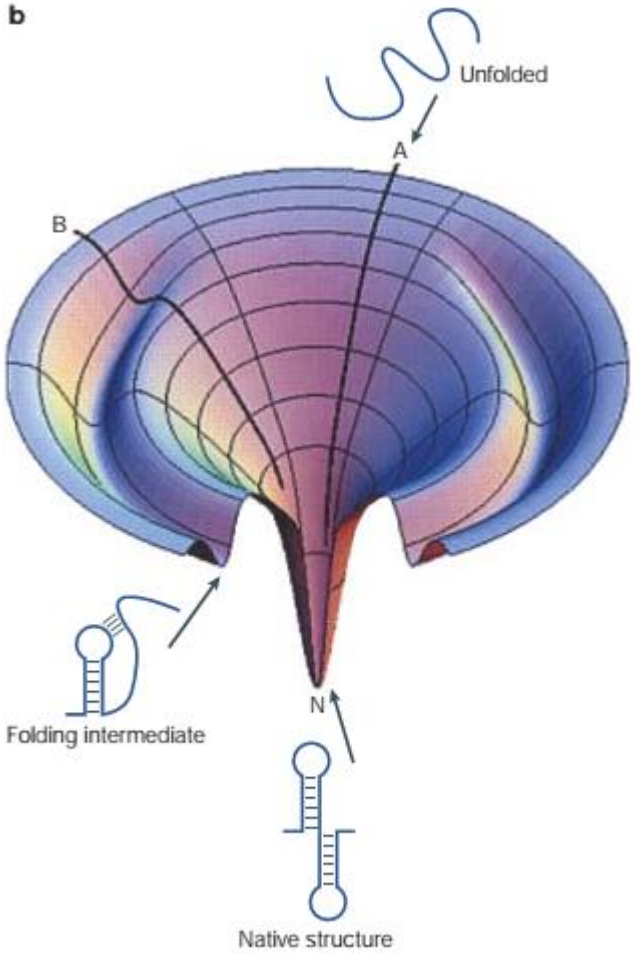
strand folds upon itself to form base pairs so it can have a diverse form of secondary structure.

The secondary structure of biological RNA's can often be uniquely decomposed into stems and loops.



Translation





Nature Reviews | Molecular Cell Biology

Translation

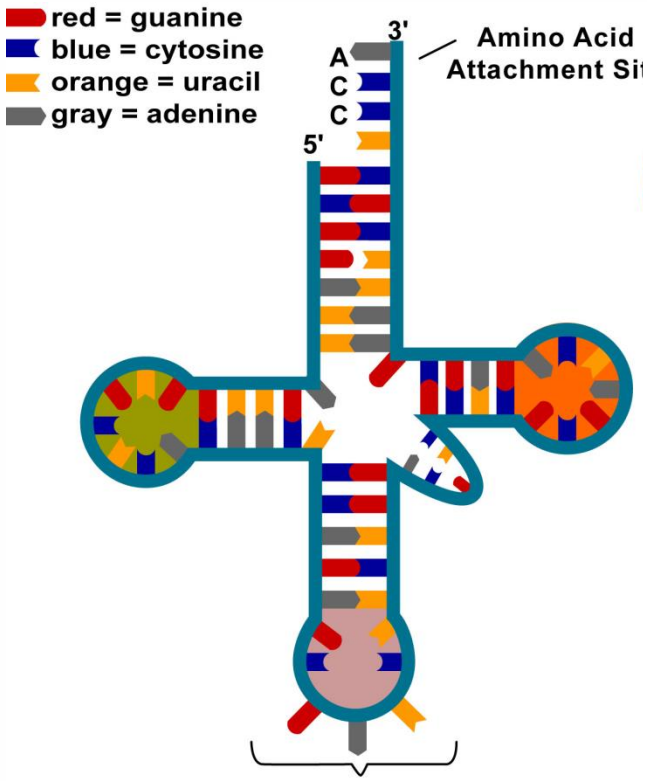
Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots
Discrete Applied Mathematics
Volume 104, Issues 1–3, 15 August 2000, Pages 45–62

Predicting RNA secondary structures from sequence and probing data
Methods
Volume 103, 1 July 2016, Pages 86–98
Advances in RNA Structure

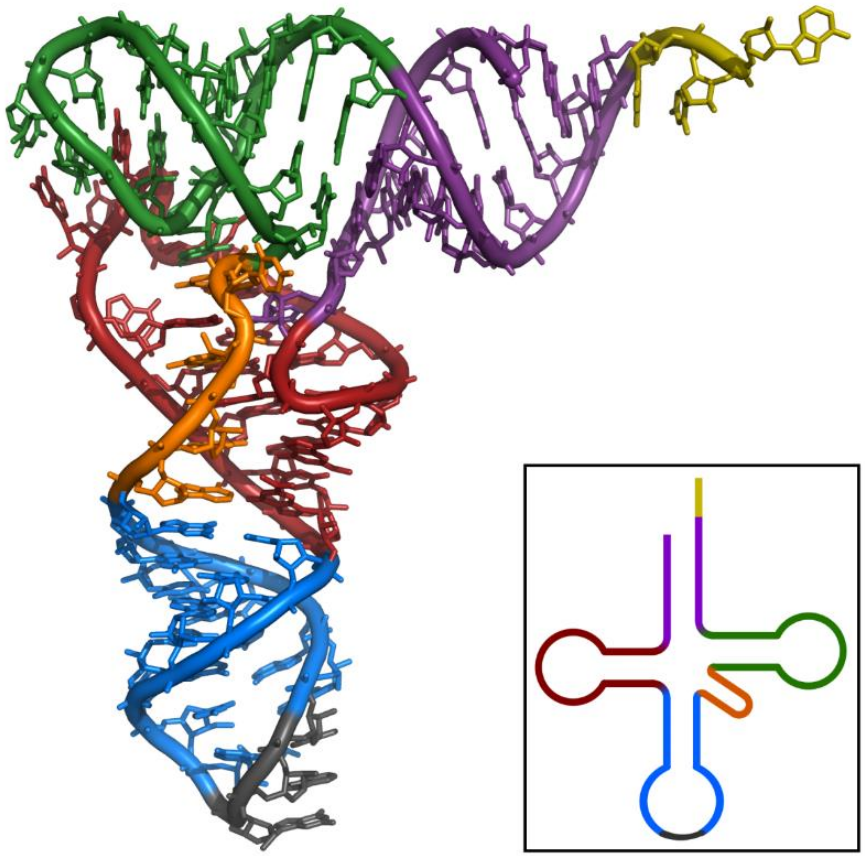
A Dynamic 3D Graphical Representation for RNA Structure Analysis and Its Application in Non-Coding RNA Classification
PLoS ONE
Volume11, issue 5, e0152238.

Translation

- **Transfer RNA** (**tRNA**) is a molecule composed of RNA, typically 73 to 93 nucleotides in length.



2 Dimensional



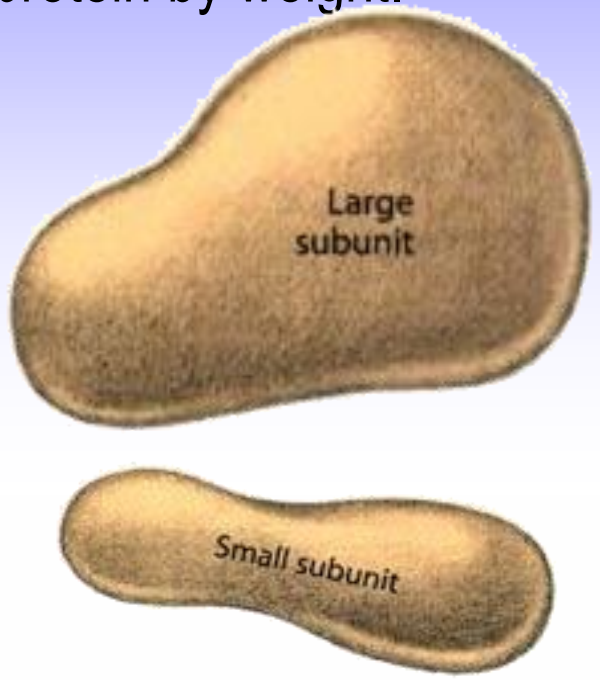
Translation

- **Ribosomal RNA (rRNA)**

Ribosomal ribonucleic acid (rRNA) is the RNA component of the ribosome. It constitutes the predominant material within the ribosome, which is approximately 60% rRNA and 40% protein by weight.

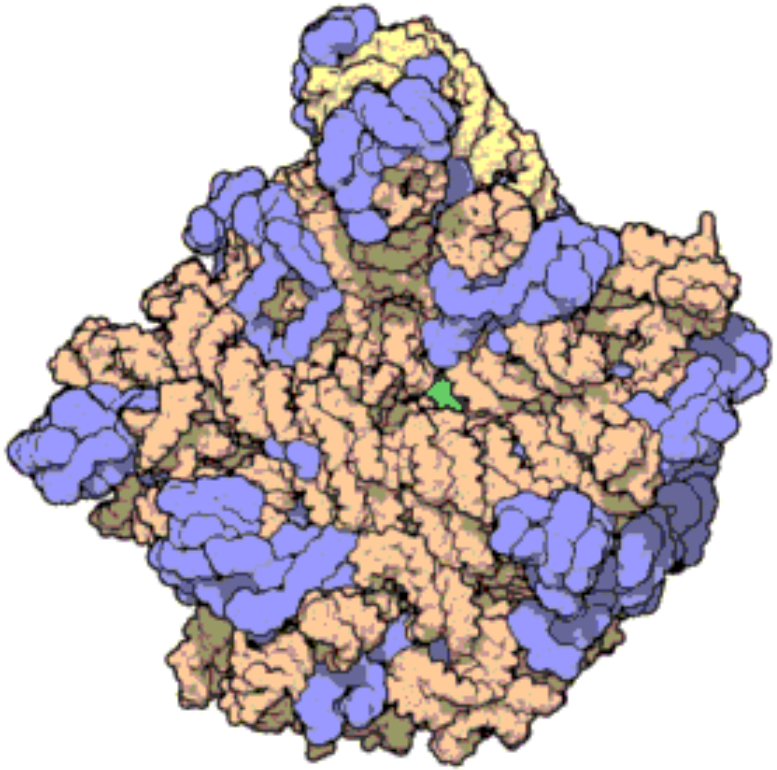
The ribosomal RNAs form two subunits:

- the large subunit (LSU)
- small subunit (SSU)

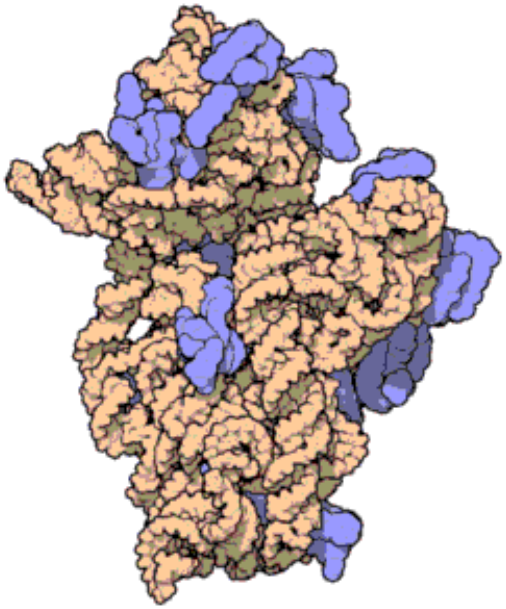


Ribosomal RNA provides a mechanism for decoding mRNA into amino acids and interacts with tRNAs during translation.

Translation



LSU

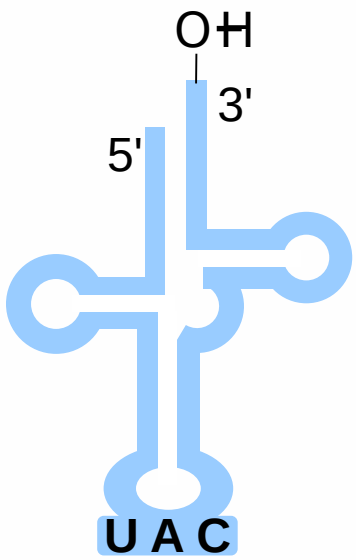


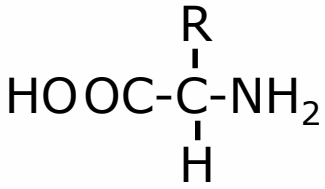
SSU

Translation

Similarly to transcription, translation proceeds in four phases:

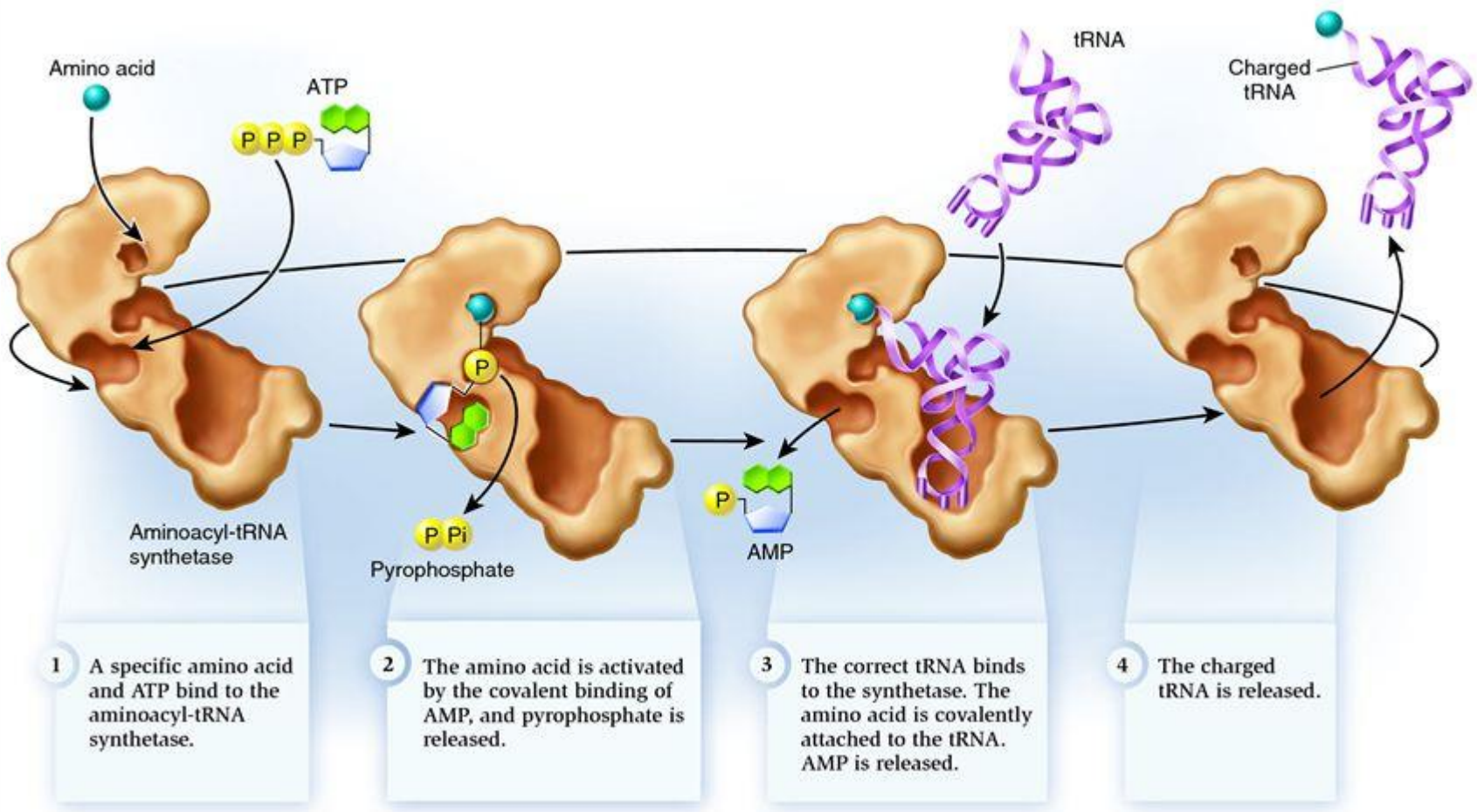
1. **Activation:** the correct amino acid is covalently bonded to the correct transfer RNA (tRNA).





H_2O

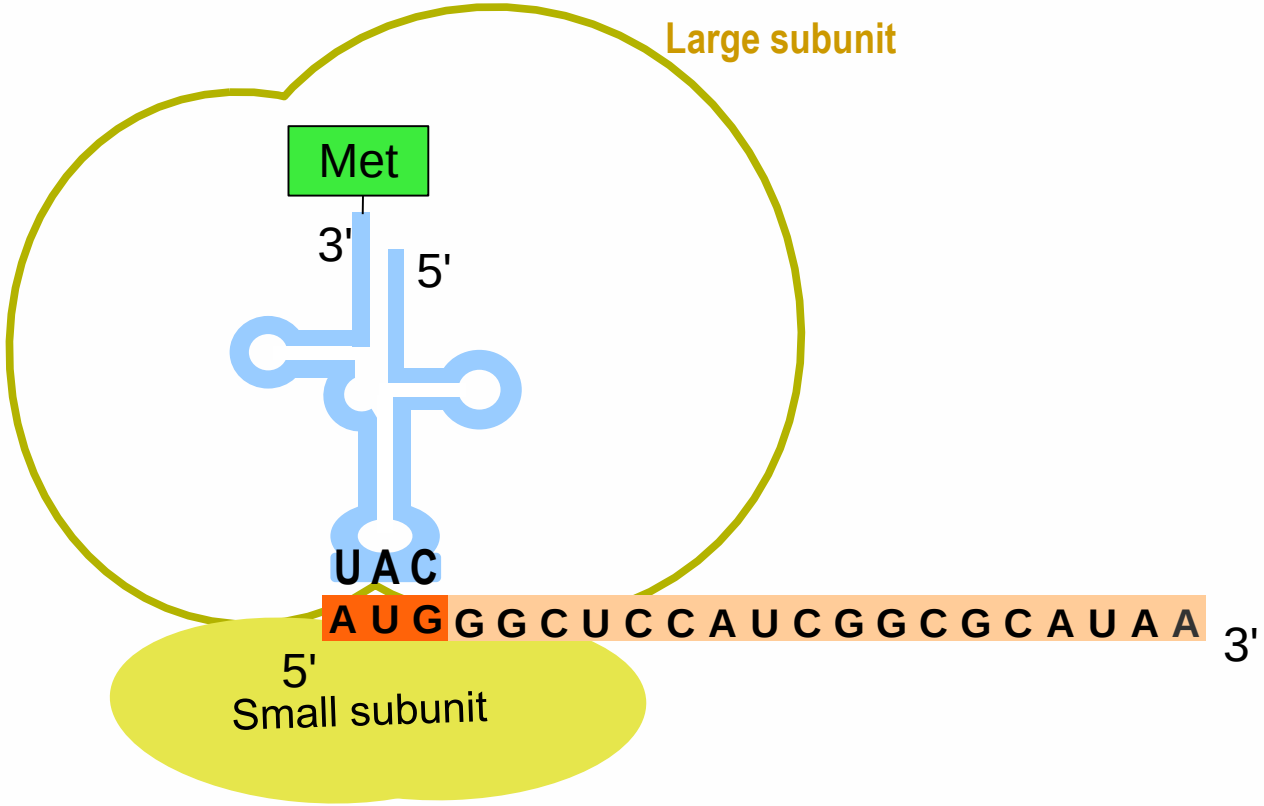
Translation



Translation

Similarly to transcription, translation proceeds in four phases:

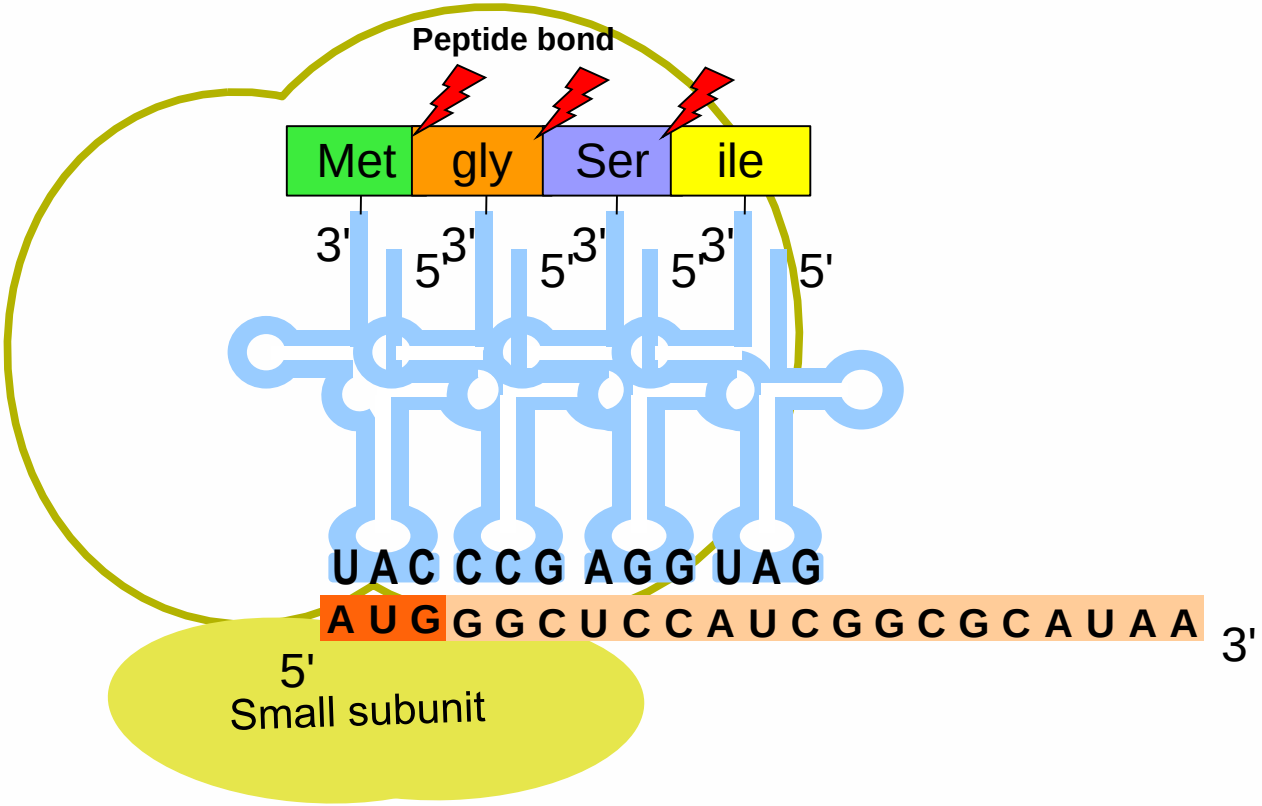
2. **Initiation:** Initiation involves the small subunit of the ribosome binding to the 5' end of mRNA



Translation

Similarly to transcription, translation proceeds in four phases:

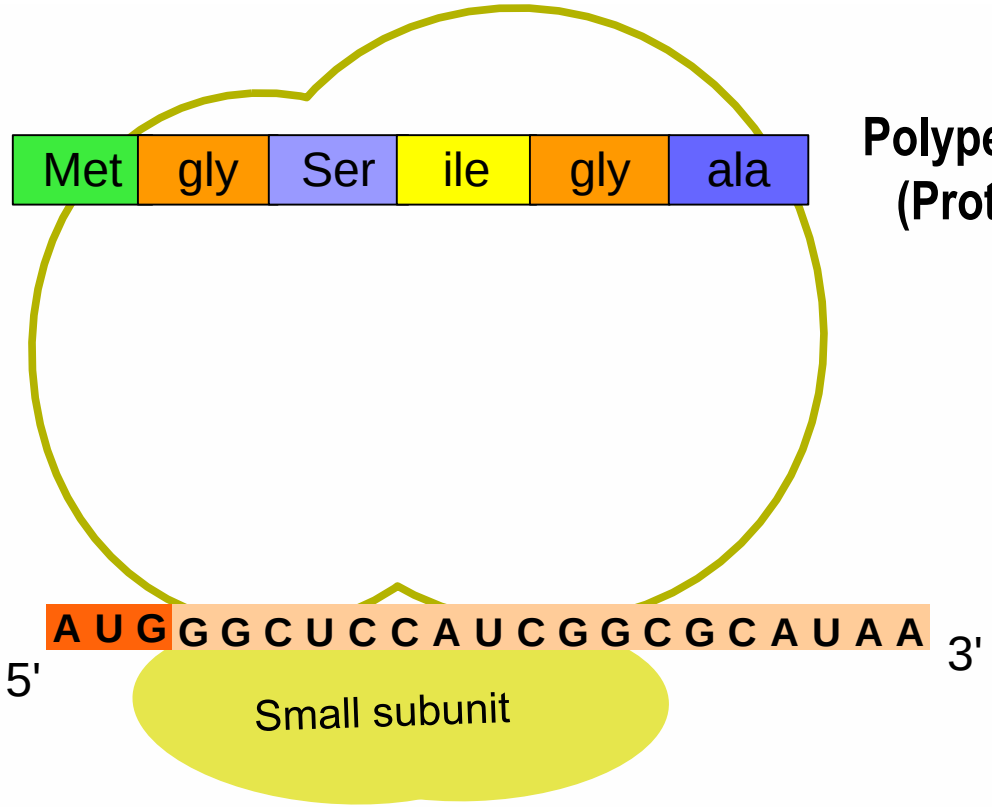
3. **Elongation:** During chain elongation, each additional amino acid is added to the nascent polypeptide chain



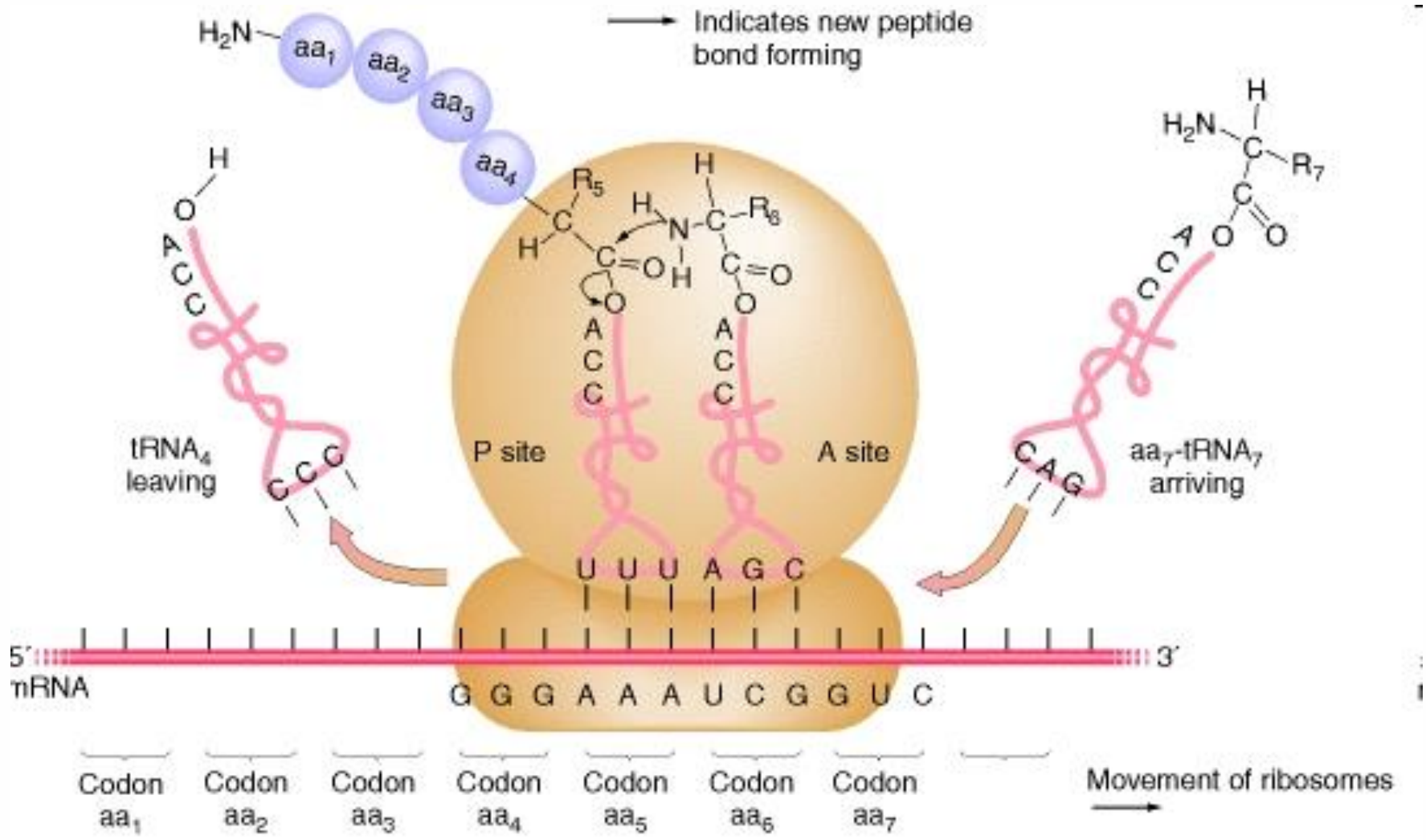
Translation

Similarly to transcription, translation proceeds in four phases:

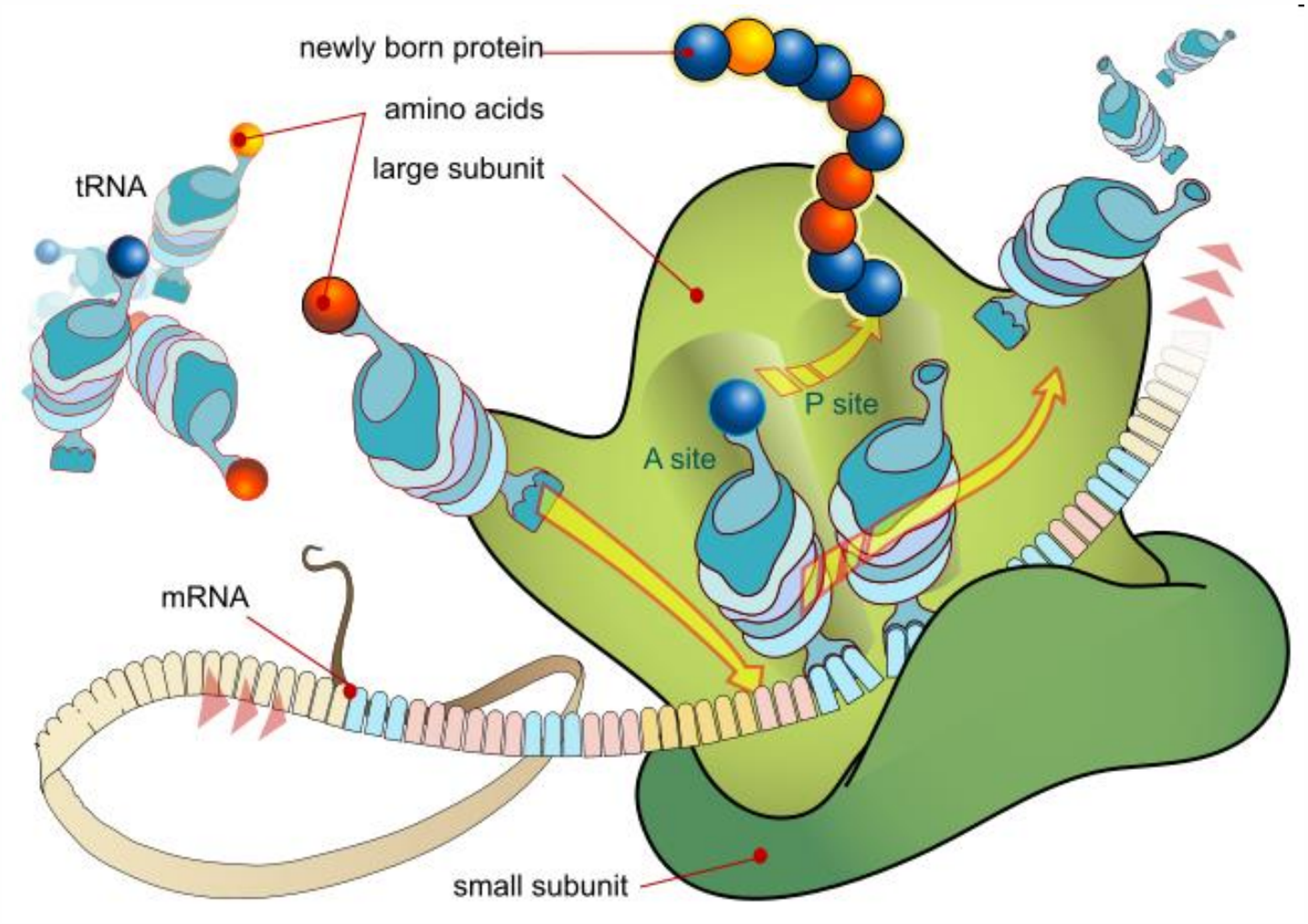
3. **Termination:** No tRNA can recognize or bind to this codon. Instead, the stop codon induces the binding of a release factor protein.



Translation



Translation



Translation

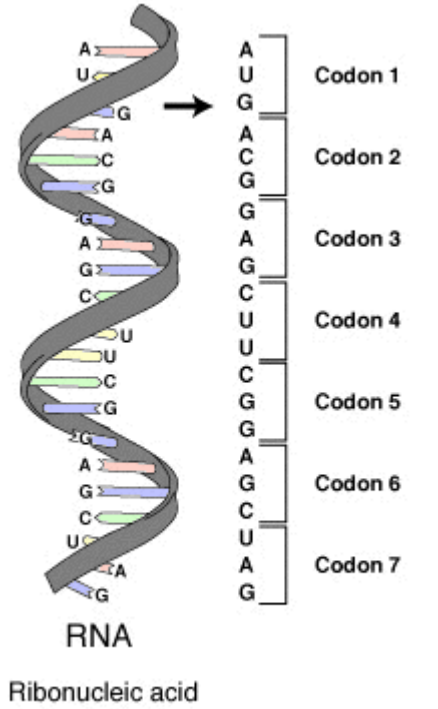
Genetic Code

The genetic code is the set of rules by which information encoded in genetic material (DNA or mRNA sequences) is translated into proteins (amino acid sequences) by living cells.

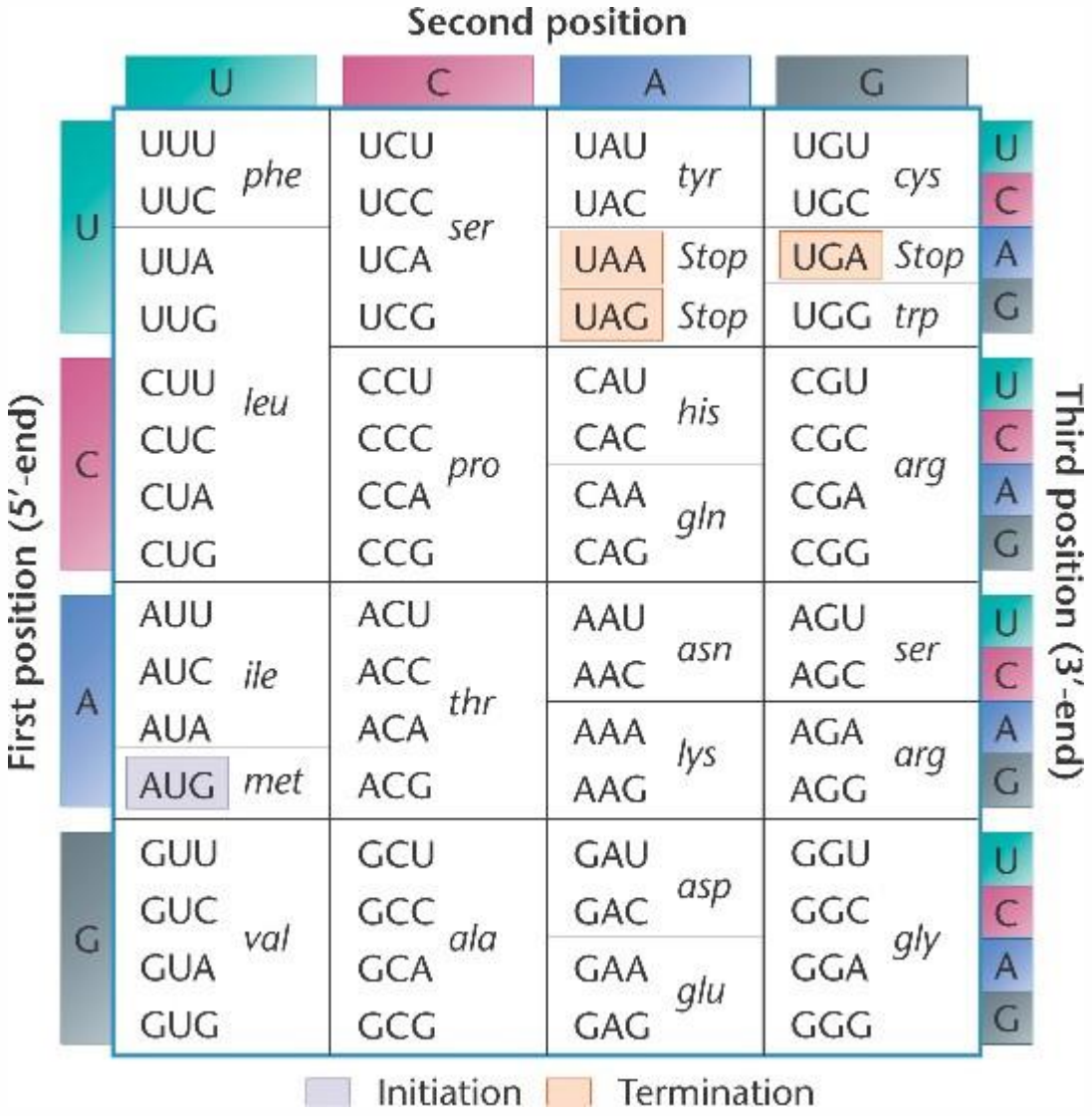
The code defines how sequences of three nucleotides, called **codons**, specify which amino acid will be added next during protein synthesis.

- 4 different base
- 3 position for each Codon
- $4^3 = 64$ different combination
- There is only 20 amino acids!

Degeneracy



Translation



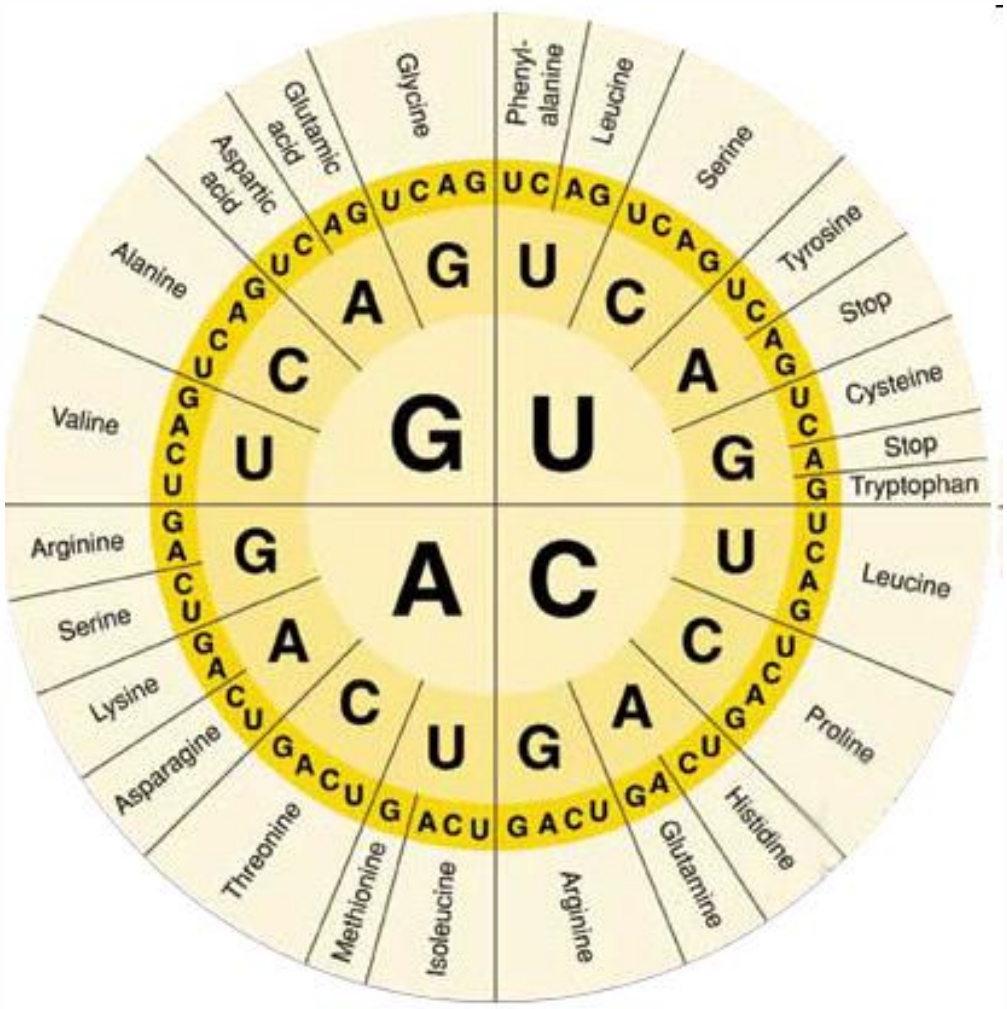
Translation

First position (5′-end)	Second position: U	Second position: C	Second position: A	Second position: G	Third position (3′-end)
U	UUU phe	UCU ser	UAU tyr	UGU cys	U
	UUC phe	UCC ser	UAC tyr	UGC cys	C
	UUA leu	UCA ser	UAA Stop	UGA Stop	A
	UUG leu	UCG ser	UAG Stop	UGG trp	G
C	CUU leu	CCU pro	CAU his	CGU arg	U
	CUC leu	CCC pro	CAC his	CGC arg	C
	CUA leu	CCA pro	CAA gln	CGA arg	A
	CUG leu	CCG pro	CAG gln	CGG arg	G
A	AUU ile	ACU thr	AAU asn	AGU ser	U
	AUC ile	ACC thr	AAC asn	AGC ser	C
	AUA ile	ACA thr	AAA lys	AGA arg	A
	AUG met	ACG thr	AAG lys	AGG arg	G

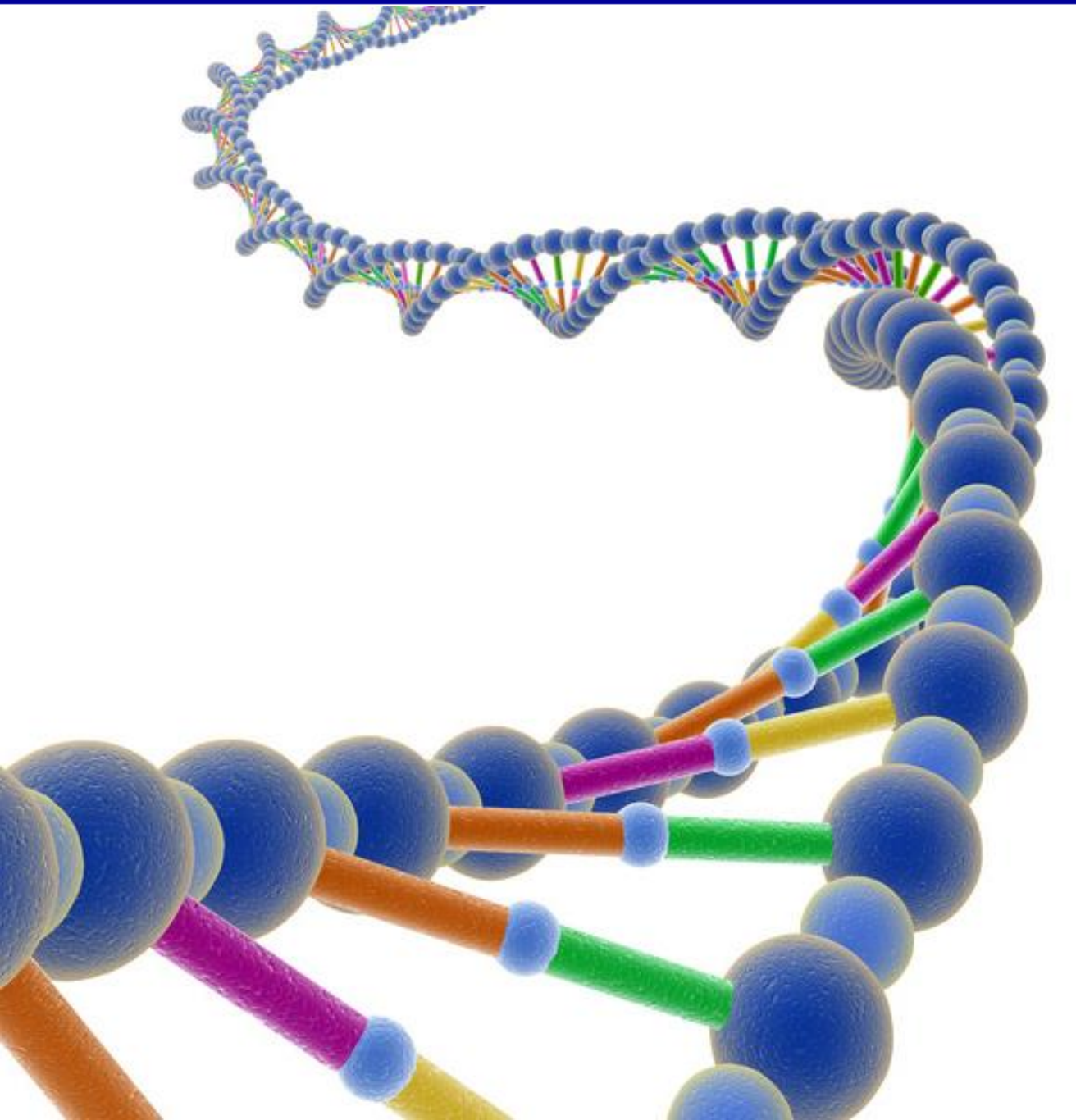
Codon	Universal Code	Unusual Code*	Occurrence
UGA	Stop	Trp	*Mycoplasma, Spiroplasma,* mitochondria of many species
CUG	Leu	Thr	Mitochondria in yeasts
UAA, UAG	Stop	Gln	*Acetabularia, Tetrahymena,* Paramecium, etc.
UGA	Stop	Cys	*Euplotes*

*"Unusual code" is used in nuclear genes of the listed organisms and in mitochondrial genes as indicated.
SOURCE: S. Osawa et al., 1992, *Microbiol. Rev.* **56**:229.

Translation



By: Mahdi Vasighi



Conditional Essential Amino Acids