

The Genetic Code

Introduction

- Proteins mostly carry twenty two different types of amino acids but there are only four types of nucleotides (A, U, G, C) in RNA.
- If the information for each amino acid is present in mRNA then a set of nitrogen bases should represent an amino acid, i.e., code for each amino acid.
- A **pair** of nucleotides will produce only **4² = 16 codes**; whereas, the sets of **four** nucleotides will give rise to as many as **256 codes**; therefore, a set of **three** nucleotides would be optimum for making the codes.
- All probable triple alternative combinations of the four bases utilized in DNA or RNA structure give 64 triple codons or possible amino acid names

- A set of experiments by **Dr. Hargobind Khurana** and his colleagues showed this to be true and discovered the codes for a number of amino acids.
- Their pioneering work led to the discovery of the codes for all the amino acids and hence building of what is known as **genetic code dictionary**.
- Codon is a **three nucleotides** sequence in DNA that is copied in mRNA (with U/T replacement), which determines the **type** and **site** of amino acid in the **translated protein**.
- It is read in the 5' → 3' direction in the mRNA.
- It is identified according to base **complementarity** in an anti-parallel manner by the anticodon, present in tRNA (3' → 5' direction).

- Among the 64 codons there are three codons (**UAA, UAG, UGA**) that are called **non-sense** (i.e., **meaning-less**) codons because they do not code for any amino acid but indicate **termination** (stop) of translation of mRNA.
- **UGA** and **UAG** also code for two rare protein amino acids; namely, **selenocysteine** and **pyrrolysine**, respectively.
- The differential usage as stop or meaningful codon depends on need of the organism.
- The remaining 61 codons are divided unevenly on the remaining 20 protein amino acids.

Characters of the genetic code

- **Specific or unambiguous:**
 - Each specific codon indicates a specific amino acid, e.g., UUU encodes phenylalanine only.
 - No two amino acids can share the same codon.
- **Degenerate:**
 - One amino acid may have more than one codon.
 - The 61 codons are not equally divided on the major 20 protein amino acids.
 - Therefore, **arginine** has 6 codons (synonyms or nicknames, CGU, CGC, CGA, CGG, AGA and AGG), whereas, **methionine(AUG)** and **tryptophan** (UGG) has only one.
 - This is why genetic code is said to be **degenerate**

- **Universality:**

- The genetic code is universal ,i.e., specify the same amino acid in all living organisms from viruses, bacteria, plants, insects to mammals.
- Exceptions occur in **mitochondrial** genome where **AUA** specifies **methionine** instead of **isoleucine** and **UGA** specifies **tryptophan** instead of being a stop or **non-sense codon** in other systems.
- Also, **arginine codons** AGA and AGG are stop codons in mitochondrial genome.
- Some other exceptions occur in ciliated protozoa and the single cell plant **acetabularia**.

- **Non-overlapping:**

- The mRNA codons are read in a continuous manner in the 5' → 3' direction without interruption in three-base sequence, i.e., no base functions as a common member of two consecutive codons.
- Therefore, addition or removal of a single base can lead to shifting of the reading frame producing totally different amino acid sequence.
- However, in small viruses, they utilize their genome in every direction possible.
- Therefore, there are overlapping genes but still non-overlapping codons

- **Comma-less:**

- Once the reading is commenced at a specific codon, the sequence is read in a continuous manner without interruption till we reach the stop codon.
- It means that the genetic code is read without any punctuation markers, i.e., there are no codons or nucleotides representing a “**comma**” or “**conjunction**” etc.

- **Colinearity of gene and product:**

- There is a linear correspondence in base sequence in the gene, the mRNA (after splicing) and amino acid sequence of the protein.

1 st nucleotide	2 nd nucleotide				3 rd nucleotide
	U	C	A	G	
U	Phenylalanine Phenylalanine Leucine Leucine	Serine Serine Serine Serine	Tyrosine Tyrosine Stop Stop (Pyrrolysine)	Cysteine Cysteine Stop (Selenocysteine) Tryptophan	U C A G
C	Leucine Leucine Leucine Leucine	Proline Proline Proline Proline	Histidine Histidine Glutamine Glutamine	Arginine Arginine Arginine Arginine	U C A G
A	Isoleucine Isoleucine Isoleucine Methionine	Threonine Threonine Threonine Threonine	Asparagine Asparagine Lysine Lysine	Serine Serine Arginine Arginine	U C A G
G	Valine Valine Valine Valine	Alanine Alanine Alanine Alanine	Aspartate Aspartate Glutamate Glutamate	Glycine Glycine Glycine Glycine	U C A G

Crick's wobble (shaking) hypothesis

- A closer look at the genetic code dictionary reveals that multiple codons for the amino acids generally differ in their third nucleotide.
- The reduced specificity at the third position in the codon is known as **third base degeneracy** or **Wobbling phenomenon** because it does not obey the strict rule of **complementarity**.
- This enables one tRNA with a specific anticodon to complement to several codons for the same amino acid if codon-anticodon binding at the third base is not very strict.
- Wobbling also minimizes the effects of mutations.

- The anticodon of tRNA identifies a number of synonym codons of one amino acid that differ at the 3rd base.
- If the 1st base of the **anti-codon** is C or A, it pairs respectively with G and A only at 3rd base of the codon; but if it is U, it can pair with either G or A as 3rd base of the codon;
- if it is G, it can pair with C or U as 3rd base of the codon; and if it is inosine “I” (nitrogen base as hypoxanthine), it can pair with A, C or U as 3rd base of the codon
- Examples of **wobble codons**: the 2 **arginine codons**, AGA and AGG that bind same **UCU anticodon** and the **three glycine** codons, GGU, GGC and GGA that bind same **CCI anticodon**.
- Wobbling and hence degeneracy limits the types of tRNA required to translate all the amino acids, therefore, only 33tRNAs are required for all the twenty two amino acids (including one ‘AUG’ for initiation).

1st Base of the AntiCodon

Bind With (3rd base of Codon)

C

G

A

U

U

G or A

G

C or U

I

A, C or U

Structure of protein

- The **polypeptide**, **amino acid** and **peptide bond**
- **Proteins**
 - Catalytic/enzymes
 - Structural
- Proteins are **poly amino acids**
- All amino acids contain the amino group(-NH₂) and carboxylic acid (-COOH) that are attached to the α -carbon
- Linkages between the **carboxyl carbon** of one amino acid and the **amino nitrogen** of a second(with the elimination of water) are known as **peptide bond**

- **Protein primary structure** consists of varying combinations of 20 different amino acids in long polypeptide chains.
- **Secondary structure** involves the formation of α -helical forms and β -pleated sheets
- **Tertiary structure** is the overall folding of the protein on itself, which results in the typical three-dimensional arrangement that allows its function.
- **Quaternary structure** exists when several polypeptide chains(subunits) together form an interacting molecule such as hemoglobin.

- **Protein conformation can be**

- **Fibrous** :- long rod water insoluble(structural elements to the animal and human organism)
- **Globular** :- compact functional protein water soluble tertiary and quaternary structure
- **Membrane proteins** are situated in the membranes of cells. They can have different forms, varying with their function.
- **Glycoproteins** are proteins with carbohydrate residues.

Protein Biosynthesis

- Proteins are ***amino acid polymers*** the sequence of which is specified by its mRNA. **Synthesized in cytoplasm, endoplasmic reticulum and/or mitochondria.**
- There are various types of proteins with different structure and function due to the difference in amino acid sequence in their constituting polypeptides.
- Cell utilizes one or more genes to produces each type of polypeptide.
- The prokaryotic protein synthesis has been well studied and has served as a model for the eukaryotic system.
- Now the process and the factors involved in eukaryotic system have also been explained reasonably well,

Cont ...

- Requirements for protein synthesis:
 - **22 amino acids, tRNAs (33type), mRNA, amino-acyl-tRNA synthetases, ribosomes**, several **protein factors** and ATP and GTP for energy
- The process of protein synthesis includes five steps:
 - 1.Activation of amino acids***
 - 2.Initiation***
 - 3.Elongation.***
 - 4. Termination.***
 - 5.Post-translational modification.***

Activation of amino acids:

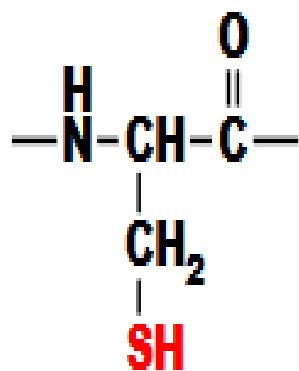
- It is catalyzed by **amino-acyl-tRNA synthetase** that is specialized in binding a specific amino acid to all of the **tRNAs** specific for that amino acid by a high energy bond, as follows. →
- Amino acid + ATP → Amino-acyl-AMP + PPi
2Pi Amino-acyl-AMP + tRNA
Amino-acyl-tRNA + AMP

- The **α -COOH** of the amino acid binds to the **2'-OH** (in class I synthetases) or **3'-OH** (in class II synthetases) group of adenosine of the acceptor arm of the tRNA (i.e., **3'-ACC-**).
- The initiation codon in prokaryotes “AUG” codes for “**Formyl-Methionine**”.
- The formylation of methionine takes place after the amino acid has been attached to tRNA.
- The enzyme **f-Met-tRNA synthetase** is different from the normal “intra-polypeptide” methionine (**Met-tRNA**) synthetase.

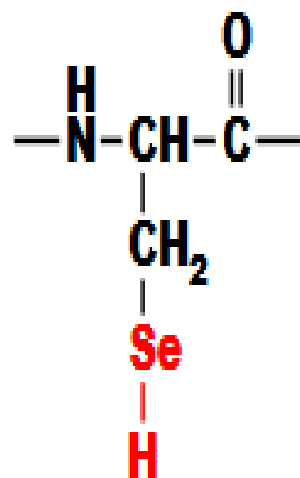
Special case

- **tRNA^{Sec}** for “Seleno-Cysteine” is different and specific for **selenocysteinyI-tRNA**.
- The anticodon of selenocysteinyI-tRNA interacts with a stop codon in the mRNA (UGA) instead of a serine codon.
- The selenocysteinyI-tRNA has a unique structure that is not recognized by the termination machinery and is brought into the ribosome by a dedicated specific elongation factor.

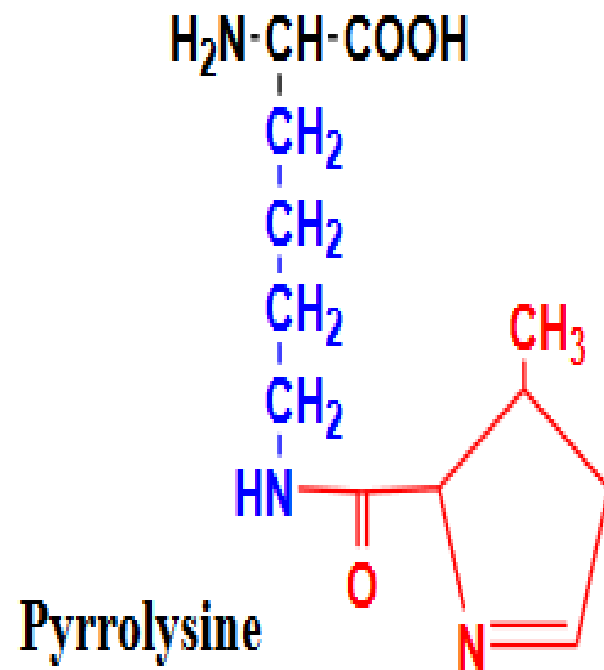
- An element in the 3'-non-translated region of **seleno protein** mRNAs determines whether UGA is read as a stop codon or as **a selenocysteine codon**.
- An example of eukaryotic selenoproteins is the antioxidant **glutathione peroxidase**.
- Similarly, there is a specific tRNA^{Pyl} CUA complementary to UAG stop codon and a specific **aminoacyl-tRNA synthetase** for **pyrrolysine**.



Cysteine



Selenocysteine



Pyrrolysine

- A “Second Genetic Code”:The enzymes **aminoacyl-tRNA synthetases** are specific not only for a single amino acid but for certain **tRNAs** as well.
- The specificity of these enzymes and their importance for protein synthesis has been highlighted by the examples like **f-Met-aminoacyl-tRNA synthetase** and the ***Seleno-Cysteine-aminoacyl-tRNA synthetases***.
- Selecting the specific tRNA in a matrix of dozens of others is as important for the overall fidelity of protein biosynthesis as is distinguishing among amino acids.

- The interaction between **aminoacyl-tRNA synthetases** and **tRNAs** has been referred to as the “**second genetic code,**” reflecting its critical role in maintaining the accuracy of protein synthesis.
- The “**coding**” rules appear to be more complex than those in the “**first**” code.
- The **tRNA amino acid arm** and the **anticodon arm** carry certain specific nucleotides, the pattern of these nucleotides along with the three dimensional structure play an important role in this “**Second Genetic Code**”.

Initiation:

- Initiation of protein synthesis involves the identification and binding of the mRNA by ribosomes. Requirements of the initiation step are:

- 1. Amino-acyl-tRNA.**
- 2. Ribosome.**
- 3. mRNA.**
- 4. GTP and ATP.**
- 5. At least 10 eukaryotic initiation factors (eIFs).**

A. Formation of the 43S pre-initiation complex:

- Eukaryotic initiation factor-1 (**eIF1**) binds to the 40S subunit of 80S ribosome and dissociates the complete ribosome into its 40S and 60S subunits.
- The initiation factor-3 (**eIF-3**) binding to the liberated 40S subunit prevents reassociation of these subunits

- The amino-acyl-tRNA (met-tRNA) interacts with initiation factor-2 (**eIF-2**), which has a binding site for GTP.
- Association of (GTP + eIF2 + met-tRNA) enables it to bind the 40S-eIF-1-eIF3 complex to give rise to the **43S preinitiation complex**.
- Since the translation always begins at the initiation codon “AUG”, the newly synthesized protein always starts with methionine (**formyl-methionine in prokaryotes**) as the N-terminal amino acid.

B. Formation of the 48S pre-initiation complex:

1. Initiation factor-4 (**eIF-4**) binds the cap of mRNA and activates it (reduces its 5'-end complex secondary structure by ATPase and ATP-dependent helicase activities) to bind the 43S pre-initiation complex to form the 48S complex.
 - Hydrolysis of ATP into ADP + Pi provides the energy for the activation.
2. **This complex scans mRNA for the initiation codon "AUG".**
 - This initiation AUG codon differs from the internal AUG in having a specific sequence around it known as the **Kozak sequence** ("GCC A/G CC AUG G").
 - Thus, anticodon of met-tRNA comes in contact with translation initiation codon in mRNA.
 - The cap and the poly(A) tail work synergistically during this step, through a poly(A) binding protein.

C. Formation of the 80S initiation complex:

- I. **Initiation factor-5 (eIF5)** hydrolyzes GTP on eIF2 into GDP + Pi and activates binding of the 48S pre-initiation complex to 60S ribosomal subunit with release of initiation factors-1, -2, -3 and -4.
 - The complete ribosome contains two amino-acyl-tRNA-binding sites called as the **“P-site”** (peptidyl site, i.e., the preceding amino-acyl-tRNA site carrying one or more amino acids) and the **“A-site”** (i.e., the succeeding amino-acyl-tRNA site).
- II. The **first amino-acyl-tRNA** carrying the first amino acid in the polypeptide chain will be automatically located at P-site and the next amino-acyl-tRNA enters at A-site

Elongation

- I. **An elongation factor-1 α (eEF-1 α)** binds with amino-acyl-tRNA with hydrolysis of GTP into GDP + Pi, forming a complex catalyzed by an active site on ribosome.
 - This complex allows amino-acyl-tRNA to enter at A-site of ribosome with the release of eEF-1 α .
- II. **The free α -NH₂ group** of the new amino acid binds the α -COOH group of the first amino acid (Met) with the transfer of whole peptide chain to tRNA at **A site by peptidyl transferase** (the 28S ribozyme of 60S subunit) with the release of the free tRNA at P-site

- Elongation factor 2 (**eEF-2, translocase**) moves the whole ribosomal complex, along with the newly formed peptidyl-tRNA, by a one-codon distance along the mRNA in 5' to 3' direction.
- Translocation requires hydrolysis of GTP into GDP + Pi and creates an open A-site for entrance of a new amino-acyl-tRNA to recognize a new codon and so on.
- Polypeptide chain is thus increased by one amino acid each time.
- Eukaryotic protein synthesis appears to be slower than the prokaryotic one as the eukaryotic ribosome incorporates **~6 amino acids/second**, whereas, prokaryotic one incorporates **~18 amino acids/second**.
- One mRNA is repeatedly translated into 10⁵ polypeptides before its degradation leading to huge augmentation

Termination

1. When a termination codon in mRNA appears at A-site, there is no tRNA that can recognize it, but there are some auxiliary proteins known as releasing factors (eRFs) that recognize these codons.
2. A releasing factor activates the **peptidyl transferase** to hydrolyze and release the peptide chain on tRNA at P-site.
 - The free tRNA and mRNA are released and the 80S ribosome dissociates into its subunits.
 - This requires hydrolysis of GTP into GDP + Pi.

Post-translation Modification of Proteins:

- The freshly synthesized proteins need to acquire their native three dimensional conformations and have to be modified by covalent linkages / substitutions to become functional.
- Further, they have to be translocated to their destined intracellular locations.
- All these modifications are known as **the post-translational modifications**.
- Below is a brief list of these modifications:

- Proper folding (prevented by molecular chaperones till complete translation), sorting, assembly, targeting, trafficking and aggregation.
- b. Proteolysis of extra-peptide(s) (pre-pro-proteins) and activation.
- c. Glycosylation (N- and O-) that occurs in endoplasmic reticulum and Golgi apparatus.
- d. Hydroxylation of proline and lysine, e.g., in collagen.
- e. Phosphorylation of tyrosine, serine or threonine.
- f. Acetylation such as histones, cross-linking such as collagen, methylation, farnesylation, myristoylation, geranoylation, sulfation, retinoylation, and fatty acylation.
- g. Ubiquitinylation as a marker for degradative fate that is specified by the amino terminal amino acid of the protein. Some amino acids are stabilizing such as serine and methionine, whereas, others are destabilizing and shorten the protein half-life to a few minutes such as arginine and aspartate.

1. Proper folding (prevented by molecular chaperones till complete translation), sorting, assembly, targeting, trafficking and aggregation.
2. Proteolysis of extra-peptide(s) (pre-pro-proteins) and activation.
3. Glycosylation (N-and O-) that occurs in endoplasmic reticulum and Golgi apparatus.
4. Hydroxylation of proline and lysine, e.g., in collagen.
5. Phosphorylation of tyrosine, serine or threonine.

1. Acetylation such as histones, cross-linking such as collagen, methylation, farnesylation, myristoylation, geranoylation, sulfation, retinoylation, and fatty acylation.

2. Ubiquitinylation as a marker for degradative fate that is specified by the amino terminal amino acid of the protein.

- Some amino acids are stabilizing such as serine and methionine, whereas, others are destabilizing and shorten the protein half-life to a few minutes such as arginine and aspartate.

Regulation of protein synthesis

- Cellular stresses:
 - During stresses -such as amino acids and glucose starvation, viral infection, protein misfolding, serum deprivation, hyperosmolality and heat shock, the protein synthesis is slowed down or inhibited.
 - The inhibition is brought about by the inactivation of **eIF2**. The initiation factor is under regulatory control of at least 4 different protein kinases that phosphorylate it resulting in its inactivation.

- **Inducers:**

- Since a high level of protein synthesis is required during cell division, the mitogenic factors should act as inducers of protein synthesis.
- A number of endocrine stimuli also enhance the protein synthesis.
- Insulin and mitogenic growth factors induce **phosphorylation** of **eIF4** leading to enhancement of initiation.
- This results in the posttranscriptional increase in protein synthesis in liver, adipocytes and muscles.

- **Regulatory mRNA-binding proteins:**

- Specific RNA-binding proteins that bind to sequences in either the 5'-non-translated region (5'-UTR) or 3'-UTR of certain mRNAs regulate their rate of translation.
- Examples are mRNAs for proteins concerned with iron metabolism.
- When iron levels are low the rate of synthesis of the membrane transferrin receptor mRNA increases so that cells bind plasma transferrin and transfer more iron.
- In the 3'-UTR of the receptor mRNA there is specific iron response element that make hair-pin loop structures recognized by a specific cytoplasmic iron response element-binding protein

- This binding happens when cellular iron concentration is low and act to stabilizes mRNA to be available for translation.
- When cellular iron level is high, this binding is prevented by iron binding into iron response element-binding protein leading to unstable **transferrin receptor mRNA** that is degraded and no protein is synthesized.
- Oppositely, translation of the ferritin mRNA into the cellular iron sequestering protein, ferritin, is prevented through interaction of a specific iron response element-binding protein with a specific iron response element at its 5'-UTR.

- High cellular iron concentration binds into and releases the iron response element-binding protein allowing a high rate of mRNA translation.
- Low cellular iron allow the iron response element-binding protein binding and prevents translation of the ferritin mRNA

- **Viral and bacterial proteins:**

- Some viral proteins prevent association of the initiation factors eIF2, eIF4 and mRNA,
- e.g., the lysogenic phage infected Diphtheria exotoxin ADP-ribosylates eEF2 and inactivates it.

- **Effect of antibiotics on protein synthesis:**

- Many antibiotics selectively inhibit protein synthesis in bacteria (but also mitochondria, leading to fatigue and complications) because of the distinction between eukaryotic and prokaryotic ribosomal system

Antibiotic Inhibitors of Protein Synthesis

- **Aminoglycosides** (streptomycin, gentamycin and amikacin), they bind to the 23S rRNA and disturb mRNA binding to the ribosome.
- **Tetracyclines** prevent the binding of amino-acyl-tRNA to the A site.
- **Chloramphenicol** inhibits peptidyltransferase.
- 4. Puromycin has a structural similarity with tyrosyl-tRNA and causes premature release of the peptide in both eukaryotes and prokaryotes. 5. Cycloheximide inhibits peptidyl transferase in eukaryotes only. 6. Ricin, a toxin in castor beans inactivates by cleavage the eukaryotic 28S rRNA.