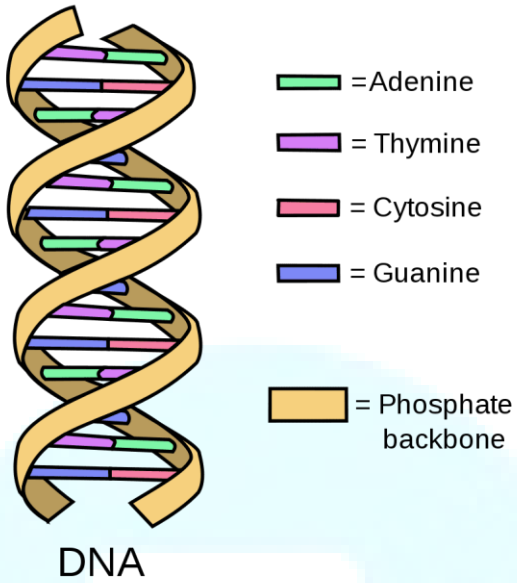
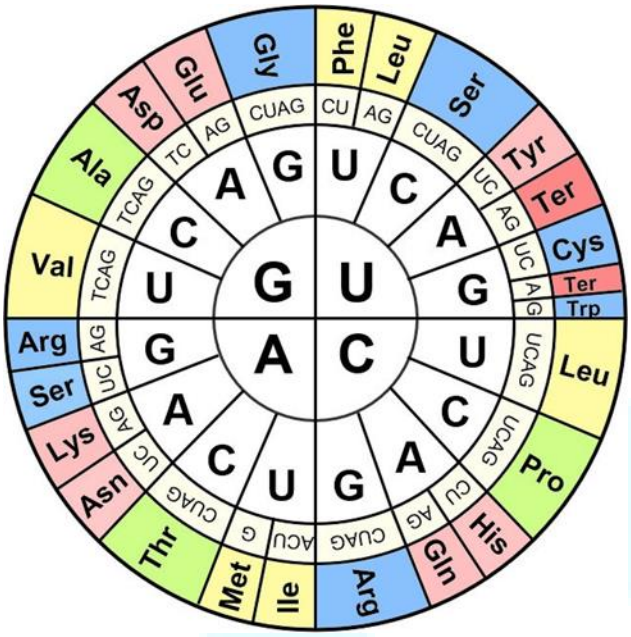
THE GENETIC CODE



- The primary gene action consist of transcription of the genetic information in the DNA into riboploynucleotide sequences of mRNA and the subsquent translation of this information contained in the four letter alphabet A, U, G and C of the mRNA into specific amino acid sequences in the polypeptide chain.
- There are twenty naturally occurring amino acids in proteins.
- Between 1953 and 1960, numerous concepts were forwarded to explain how a four letter alphabet can be translated into a twenty letter alphabet.
- A code of one nucleotide per amino acid is not at all possible since it will account only for four of the twenty naturally occuring amino acids.
- A coding sequence of two nucleotides for one amino acid or a doublet code would produce only 42 or 16 possible coding combinations.
- In order to code for the 20 amino acids, four of these doublet codons will have to code for two amino acids each.
- Such a doublet code would therefore cause an ambiguous determination of at least some of the AA which hardly accords with the precise AA composition and

sequence required in almost in any protein.



- A codon size of three nucleotides for one AA seems more likely since it produces 43 or 64 possible combinations or codons.
- To account for the excess of codons beyond the necessary 20, it is supposed that more than one codon codes for the same AA.
- Thus the genetic code is called degenerate.
- It is also possible that some of the codons in excess of 20 do not code for any of the AA and hence are none – sense codons.
- The first definite coding scheme was the one proposed by George Gamov (1956).
- His coding scheme is known as the diamond code or the overlapping code.
- But it was soon found inadequate, impractical and inconsistent with the facts of protein synthesis and was rejected.
- Later, by 1961, F.H.C. Crick and co workers proposed the concept of genetic code in accordance with the principles of protein synthesis and gene structure.
- One important feature of the genetic code as Crick hypothesized it is that it is comma less i.e. without interspersed nucleotides among the codons.

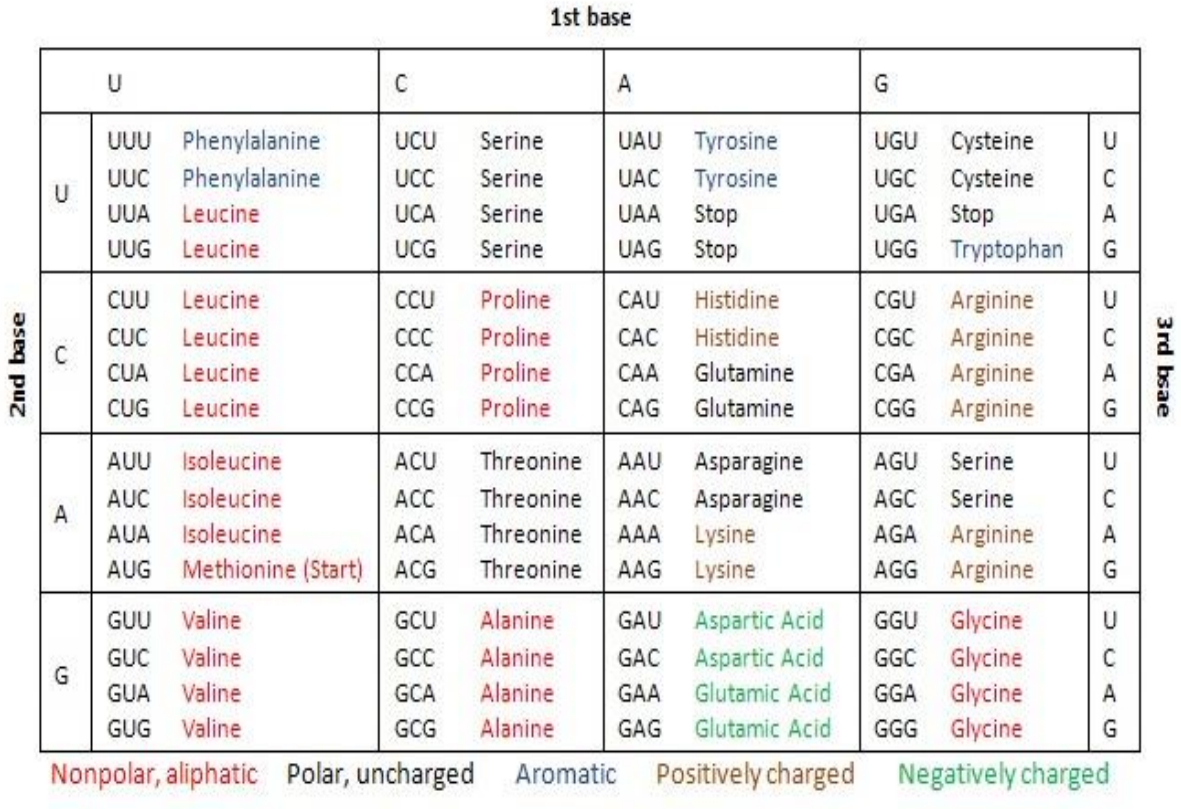
- A comma less codon would have to be read from a particular starting point on the mRNA chain in a 'reading frame' consisting of codons placed side by side, without any interrupting nucleotide sequence.
- Evidence on the length of the codon, the degeneracy of the code and its comma less nature was presented in 1961 by Crick, Barnett, Brenner and Watts – Tobin.
- They studied a number of mutations induced by acridine dyes on the B cistrons of the rII locus of the phage T4.
- Mutations like an insertion (+) or a deletion (-) can bring about a shift in the reading frame in the mRNA by one nucleotide.
- When the deletion or insertion occurred in groups of three, or the original reading frame was re-established.
- However two insertion or two deletion occurring together could not restore the reading frame. These observations strongly supported the triplet nature of the code

Wild type	CAT	GAT	CAT	GAT	CAT	GAT	CAT	GAT
Insertion	CAT	GGA	TCA	TGA	TCA	TGA	TCA	TGA
Deletion	CAT	GAT	CAT	GTC	ATG	ATC	ATG	ATC
					A			
Both		GGA	TCA	TGA	TCT	GAT	CAT	GAT
			G			A		
3 insertions	CAT	GGA	TCA	TGG	ATC	ATG	CAT	GAT
				G		G		G
3 deletions	CAT		GAC	ATG	ACA	TGA	CAT	GAT	CAT
					T		T		T

- The restoration of the wild type by these various combinations depends on the comma less nature of the triplet code i.e. wild type messages can be restored by changes further along the nucleotide chain because all the intervening nucleotides are evolved in the reading frame.
- This data further suggested that the code is degenerate. A degenerate code is one in which more than one codon specifies the same AA.

- In the cases where the wild type is restored (when the '+' s the '-'s occur together or when three '+' s and three '-' s occurred together) the original reading frame is also restored.
- In 1961 Marshall Nirenberg and J.H. Matthei published results which characterized the first specific coding sequence.
- They try to synthesize artificial mRNA molecules and to use them in an in vitro cell free protein synthesizing system.
- The mRNA molecules were synthesized artificially with the help of the enzyme polynucleotide phosphorylase.
- In contrast to RNA polymerases, the polynucleotide phosphorylase does not require any DNA template.
- As a result, the ribonucleotides are assembled at random, according to relative concentration of the four ribonucleotide diphosphates added to the reaction mixture.
- Nirenberg and Philip Leder in 1964 develop the triplet – binding assay, which led to the specific and precise assignment of the triplets.
- They prepared "three letter" mRNA molecule in all the 64 possible codon arrangements and then formed three letter mRNA – ribosome complexes. To these complexes were added activated AA.
- The amino acid bound by a particular tri nucleotide mRNA sequence is detected by radioactive labelling of only one kind of AA in the mixture and observing whether there is pronounced incorporation of this radioactivity into the ribosomal fraction.
- The specific binding of amino – acyl tRNA molecules to known mRNA ribosome complexes showed the specific assignment between 61 of the 64 triplet codons and 20 amino acids.

- It showed that the code is unambiguous i.e. a single triplet specifies only one AA.



2nd base	1st base: U		1st base: C		1st base: A		1st base: G		3rd base
U	UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine	U
	UUC	Phenylalanine	UCC	Serine	UAC	Tyrosine	UGC	Cysteine	C
	UUA	Leucine	UCA	Serine	UAA	Stop	UGA	Stop	A
	UUG	Leucine	UCG	Serine	UAG	Stop	UGG	Tryptophan	G
C	CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine	U
	CUC	Leucine	CCC	Proline	CAC	Histidine	CGC	Arginine	C
	CUA	Leucine	CCA	Proline	CAA	Glutamine	CGA	Arginine	A
	CUG	Leucine	CCG	Proline	CAG	Glutamine	CGG	Arginine	G
A	AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine	U
	AUC	Isoleucine	ACC	Threonine	AAC	Asparagine	AGC	Serine	C
	AUA	Isoleucine	ACA	Threonine	AAA	Lysine	AGA	Arginine	A
	AUG	Methionine (Start)	ACG	Threonine	AAG	Lysine	AGG	Arginine	G
G	GUU	Valine	GCU	Alanine	GAU	Aspartic Acid	GGU	Glycine	U
	GUC	Valine	GCC	Alanine	GAC	Aspartic Acid	GGC	Glycine	C
	GUA	Valine	GCA	Alanine	GAA	Glutamic Acid	GGA	Glycine	A
	GUG	Valine	GCG	Alanine	GAG	Glutamic Acid	GGG	Glycine	G

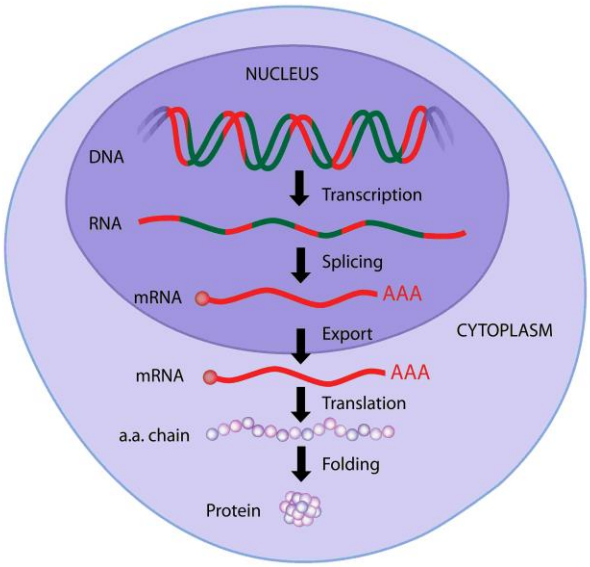
Nonpolar, aliphatic Polar, uncharged Aromatic Positively charged Negatively charged

Properties of the Genetic Code

- The genetic code is composed of nucleotide triplets.
- The genetic code is non overlapping.
- The genetic code is comma-free.
- The genetic code is degenerate
- The genetic code is ordered. (5′ to 3′)
- The genetic code contains start and stop codons.
- The genetic code is universal.

Protein synthesis

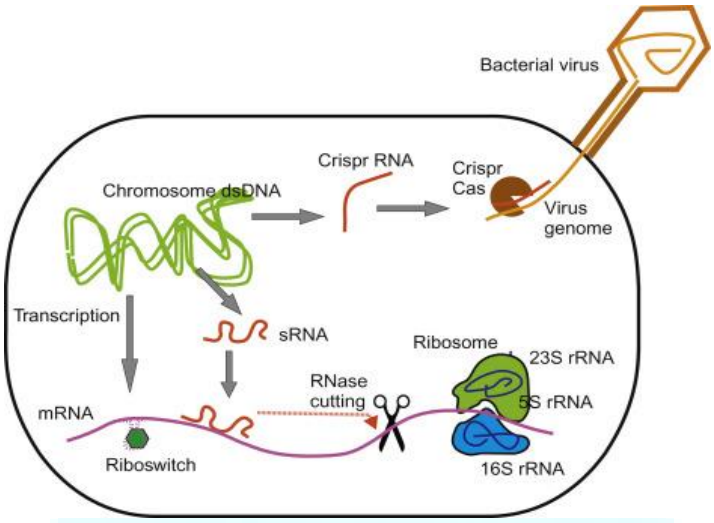
- DNA in eukaryotes is largely restricted to the nucleus and the proteins are synthesized in association with the ribosomes in the cytoplasm. i.e. DNA does not participate directly in the synthesis of proteins.
- The amount of proteins present in a cell is generally proportional to the amount of RNA present and not to the amount of DNA.



- Cellular RNA in eukaryotes is synthesized in the nucleus where the DNA is found. The DNA directed synthesis involves different steps.

1. The transfer of information present in the DNA to a particular group of RNA termed messenger RNA – Transcription
2. The migration of the mRNA from the nucleus to the cytoplasm(in eukaryotes alone)
3. The transfer of information in the form of the sequence of nucleotides in the mRNA into the sequence of amino acids in the polypeptide chain – Translation.

GENE REGULATION IN PROKARYOTES

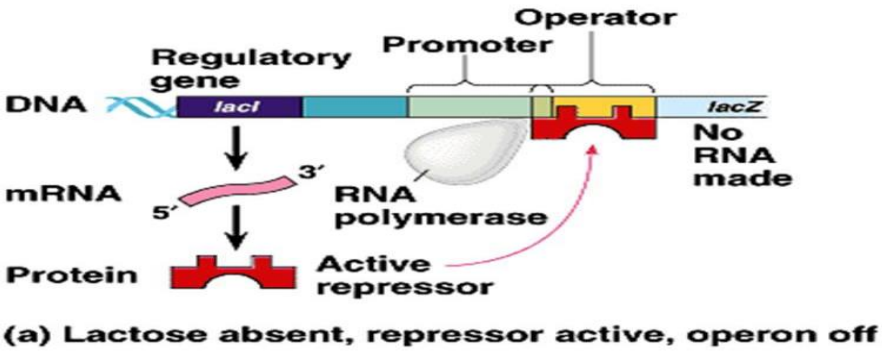


- In a cell, most of the gene products are needed only under certain environmental conditions.
- Constitutive synthesis of such a gene product would clearly be wasteful.
- The evolution of regulatory mechanisms that would provide for the synthesis of such gene products only when and where they are needed would clearly endow the organisms possessing these regulatory mechanisms a selective advantage over organisms lacking these mechanisms.
- Enzymes that are involved in catabolic pathways such as those involved in galactose, lactose and arabinose utilization are necessary only when the substrate is present.
- Such genes are inducible genes which are activated by the presence of the substrate or its derivative (inducer).
- Similarly, enzymes involved in anabolic reactions need not be synthesized if the end product is present in supra optimal concentrations in the cell.
- Such repressible genes are turned off in presence of the specific substance (repressor).

- They are activated (de-repressed) only in the absence of that particular molecule.
- Induction and suppression of genes can be accomplished by essentially the same mechanism known as the operon concept, which was described by Francois Jacob and .Jaques Monod in 1961.
- They distinguished two types of regulatory sequences in DNA - sequences that code for trans-acting products and cis-acting sequences that funcction exclusively within the DNA.
- Jacob and Monod proposed that the trans acting element known as the regulatory gene codes for a protein -the repressor/co repressor which binds to the cis-acting element, the operator.
- The operator is always located contiguous to the structural genes whose expression it controls.
- When the repressor is bound to the operator, transcription of the structural genes cannot occur as RNA polymerase is prevented from binding to the promoter adjacent to the operator.
- The complete contiguous unit consisting of the structural genes, operator, promoter and the regulator gene are called an operon.
- Since the product of the regulatory gene, the repressor, acts by shutting off the transcription of structural genes, the operon model of Jacob and Monod is referred to as negative control system.

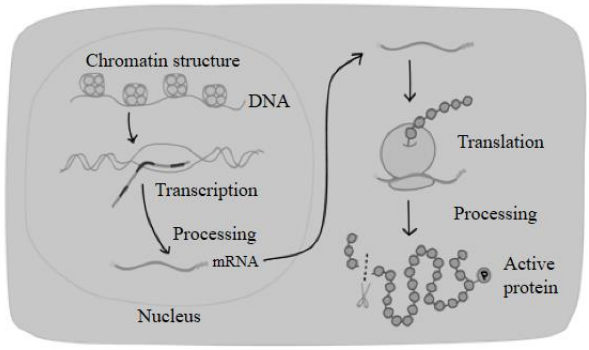
The best-known example for an inducible operon is the lac operon concerned with the synthesis of enzymes involved in the utilization of lactose.

- The three enzymes involved in lactose metabolism are β galactosidase (involved in the breaking down of lactose into glucose and galactose), lac permease (transporting lactose into the cell) and trans acetylase (transfers an acetyl group from the acetyl coA to lactose), coded by z, y, and a genes respectively.
- The three genes are contiguous and are always transcribed together into a polycistronic mRNA.



(a) Lactose absent, repressor active, operon off

GENE REGULATION IN EUKARYOTES



- Gene expression in eukaryotes is more complicated than in prokaryotes because of they are compartmentalized by an elaborate system of membranes.
- This compartmentalization subdivides the cells into separate organelles the most conspicuous one being the nucleus.
- Eukaryotic cells also possess organelles such as mitochondria, chloroplasts and endoplasmic reticulum each performing a different function.
- This subdivision of eukaryotic cells into organelles physically separates the events of gene expression.
- The primary event, the transcription occurs in the nucleus.

- RNA transcripts are also modified in the nucleus by capping, polyadenylation and the removal of introns.
- The resulting messenger RNAs are then exported to the cytoplasm where they are translated by the ribosomes.
- This physical separation of the events of gene expression makes it possible for regulation to occur indifferent places.
- Regulation can occur in the nucleus at either the DNA or RNA level, or in the cytoplasm at either the RNA or polypeptide level.