



MOLECULAR BASIS OF INHERITANCE

INTRODUCTION

Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are two types of nucleic acid found in living systems. DNA acts as the genetic material in most of the organisms. RNA though it acts as a genetic material in some viruses, like TMV, QB-bacteriophage but mostly functions as a messenger, adaptor, structural and catalytic molecule.

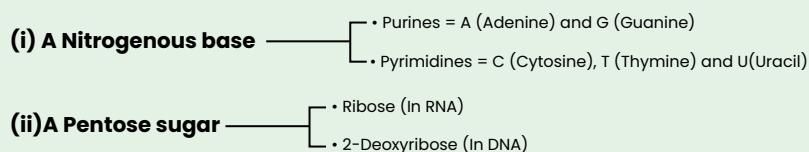
THE DNA

DNA is a long polymer of deoxyribonucleotides. The length of DNA usually defined as number of nucleotides, is also characteristic of an organism.

ORGANISM	DNA CONTENT
Bacteriophage $\emptyset \times 174$	5386 Nucleotides
Bacteriophage lambda	48502 base pairs
<i>Escherichia coli</i>	4.6×10^6 base pairs
Haploid content of human DNA	3.3×10^9 base pairs

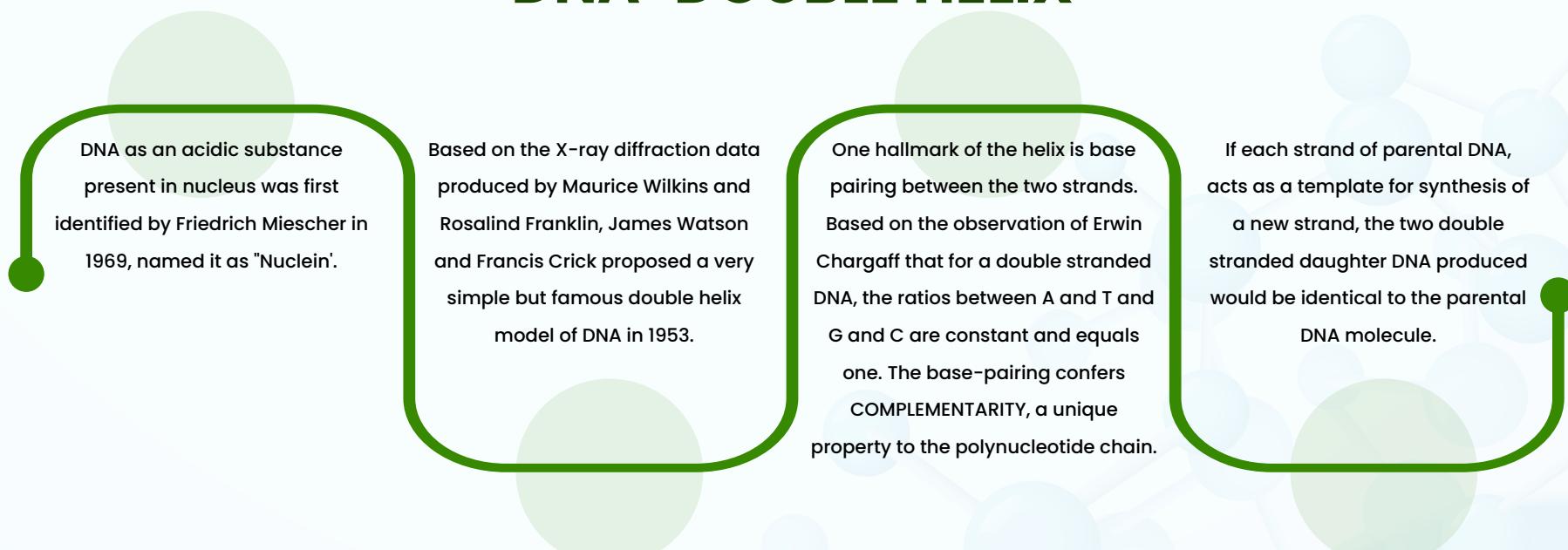
STRUCTURE OF POLYNUCLEOTIDE CHAIN

- The basic unit of a polynucleotide (DNA or RNA) is NUCLEOTIDE – which has 3 components



- Purines are same in both DNA and RNA, i.e cytosine is common to both DNA and RNA, while in RNA, Uracil is replaced with thymine (5-methyl uracil) in DNA.
- A nitrogenous base is linked to the OH of 1'C pentose sugar through a N-glycosidic linkage to form a NUCLEOSIDE.
- When a phosphate group is linked to OH of 5'C of a nucleoside through phosphoester linkage, a corresponding NUCLEOTIDE is formed.
- Two nucleotides are linked through 3'-5'C phosphodiester linkage to form a DINUCLEOTIDE. More nucleotides in such a manner form polynucleotide.
- A polymer thus formed has at one end a free phosphate moiety at -5' end of sugar and at the other end of polymer the sugar has a free OH of 3'C group.

DNA-DOUBLE HELIX



SALIENT FEATURES OF DOUBLE-HELIX OF DNA

Made up of two polynucleotide chains, where the backbone is constituted by sugar - phosphate and nitrogenous bases project inside

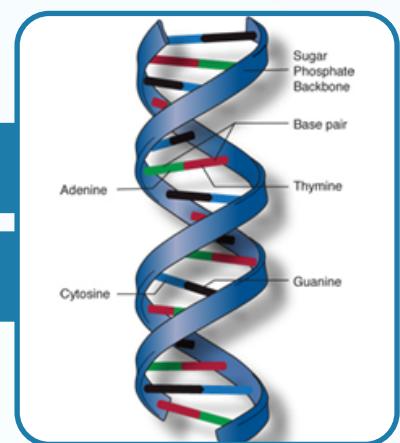
Two chain have anti-parallel polarity, one chain 5' – 3', the other 3' – 5'.

Bases in two strands are paired through hydrogen bonds, forming base pairs. (A = T and G ≡ C). This generates approximately uniform distance between the two strands.

The two chains are coiled in a right handed fashion.

Pitch of the helix = 3.4 nm. Roughly 10 bp in each turn. So, distance between two bp in helix is approx. 0.34 nm. (0.34×10^{-9} m)

The plane of one base pair stacks over the other in double-helix. This in addition to H-bonds, confers stability to the helical structure.



CENTRAL DOGMA OF MOLECULAR BIOLOGY: PROPOSED BY FRANCIS CRICK

States flow of genetic information as DNA → RNA → Protein



In some viruses the flow of information is in reverse direction, i.e. from RNA to DNA.
It is called reverse of central dogma.

PACKAGING OF DNA HELIX

DNA double helix in a typical mammalian cell has 6.6×10^9 bp, the length is approx 2.2 m (6.6×10^9 bp $\times 0.34 \times 10^{-9}$ m/bp), far greater than the dimension of a typical nucleus (approx 10^{-6} m)

In prokaryote (*E. coli*), the DNA being negatively charged is held with some proteins that have positive charges in the NUCLEOID. The DNA in nucleoid is organised in large loops held by proteins

In eukaryotes, it is much more complex.
(i) The positively charged set of basic proteins, HISTONES (rich in lysine and arginine) are organised to form a unit of eight molecules, called HISTONE OCTAMER
(ii) The negatively charged DNA is wrapped around positively charged histone octamer to form a nucleosome. A typical nucleosome contains 200 bp of DNA helix

Nucleosomes constitute the repeating unit of a structure in nucleus called CHROMATIN, a thread like stained bodies seen in nucleus. The nucleosomes in chromatin are seen as 'beads - on string' structure under electron microscope. It is packaged to form chromatin fibers that are further coiled and condensed at metaphase stage to form chromosomes.

Packaging of chromatin at higher level needs non-histone chromosomal (NHC) proteins.

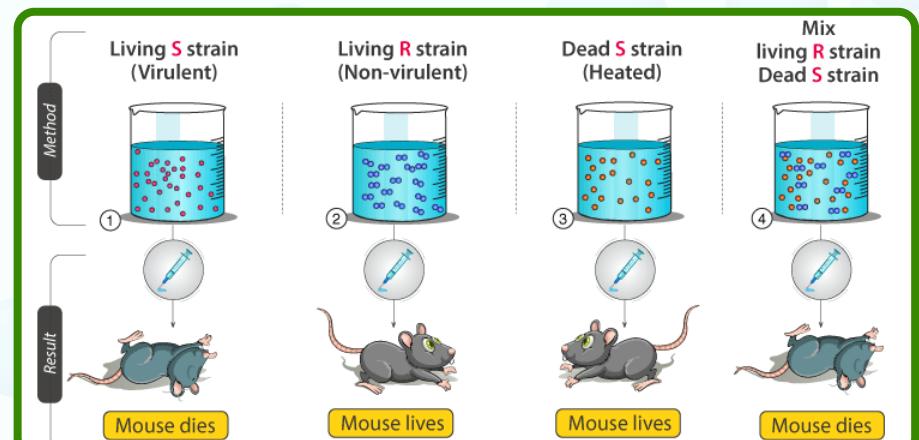
In a typical NUCLEUS (regions of chromatin)

- A. Euchromatin**
 - (i) Loosely packed
 - (ii) Stains light
 - (iii) Transcriptionally active
- B. Heterochromatin**
 - (i) More densely packed
 - (ii) Stains dark
 - (iii) Transcriptionally inactive

1 TRANSFORMING PRINCIPLE

In 1928, Frederick Griffith, in a series of experiments with *Streptococcus pneumoniae*, witnessed a miraculous transformation in Bacteria.

Some of these bacteria produce smooth shiny colonies (S) while others rough colonies (R). S-strain kills mice (virulent) while R-strain do not develop pneumonia. But heat killed S-strain are also non-virulent.



2 BIOCHEMICAL NATURE OF TRANSFORMING PRINCIPLE

Oswald Avery, Colin MacLeod and Maclyn McCarty (1933-44) discovered that DNA of S bacteria caused R bacteria to transform. As protease and RNAase did not affect transformation, but DNase inhibit transformation. They concluded that DNA is the hereditary material. But not all biologists were convinced.

3 THE GENETIC MATERIAL IS DNA

The UNEQUIVOCAL proof that DNA is the genetic material came from the experiments of Alfred Hershey and Martha Chase (1952), on bacteriophages, using radioactive phosphorus ^{32}P and sulphur ^{35}S in separate medium, with *E. coli*.

PROPERTIES OF GENETIC MATERIAL (DNA VERSUS RNA)

A molecule that can act as a genetic material must fulfill the following criteria

Should be able to generate its replica (Replication)

Should be stable chemically and structurally

Should provide the scope for slow mutation required for evolution.

Should be able to express in the form of Mendelian characters

- Both DNA and RNA can direct their duplications.
- The DNA has two complementary strands, if separated by heating can again come together when appropriate conditions are provided.
- 2'-OH group present at every nucleotide in RNA is reactive and makes it easily degradable.
- RNA is also catalytic, hence reactive. Among the two nucleic acids, DNA is a better genetic material.
- Presence of thymine at the place of uracil also confers additional stability to DNA
- Both DNA and RNA are able to mutate. Being unstable, RNA mutate at a faster rate.
- RNA can directly code for the synthesis of proteins and can easily express the characters. DNA, however is dependent on RNA for synthesis of proteins.
- DNA being more stable is preferred for storage of genetic information. For transmission of genetic information RNA is better.

RNA WORLD

RNA was the first genetic material. The essential life processes like metabolism, translation, splicing evolved around RNA. RNA used to act as a genetic material as well as a catalyst, so was reactive and hence unstable. Therefore, DNA has evolved from RNA with chemical modifications that make it more stable.

REPLICATION

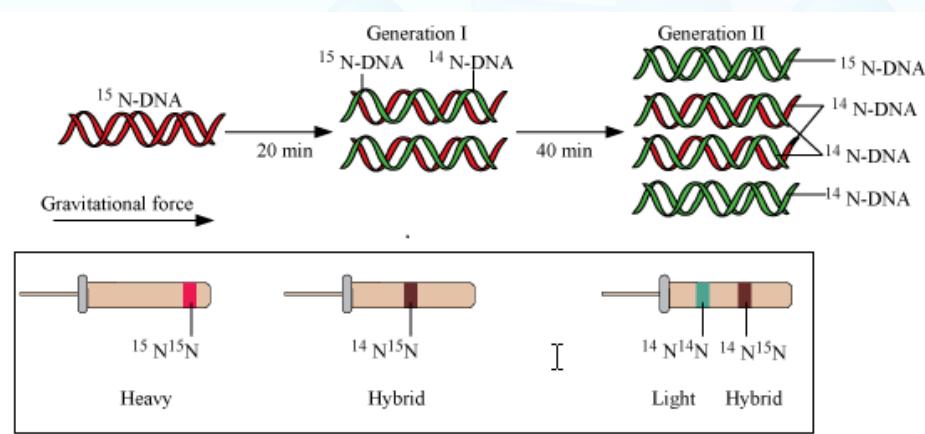
Watson and Crick had immediately proposed a scheme for replication of DNA while proposing the double helix structure of DNA, i.e., the two strands would separate and act as a template for the synthesis of new complementary strands. After completion of replication each DNA molecule would have one parental and one newly synthesised strand, termed as semi-conservative DNA replication

EXPERIMENTAL PROOF

1 Semiconservative DNA replication was shown first in *Escherichia coli*, then in higher organisms like plants and human cells.

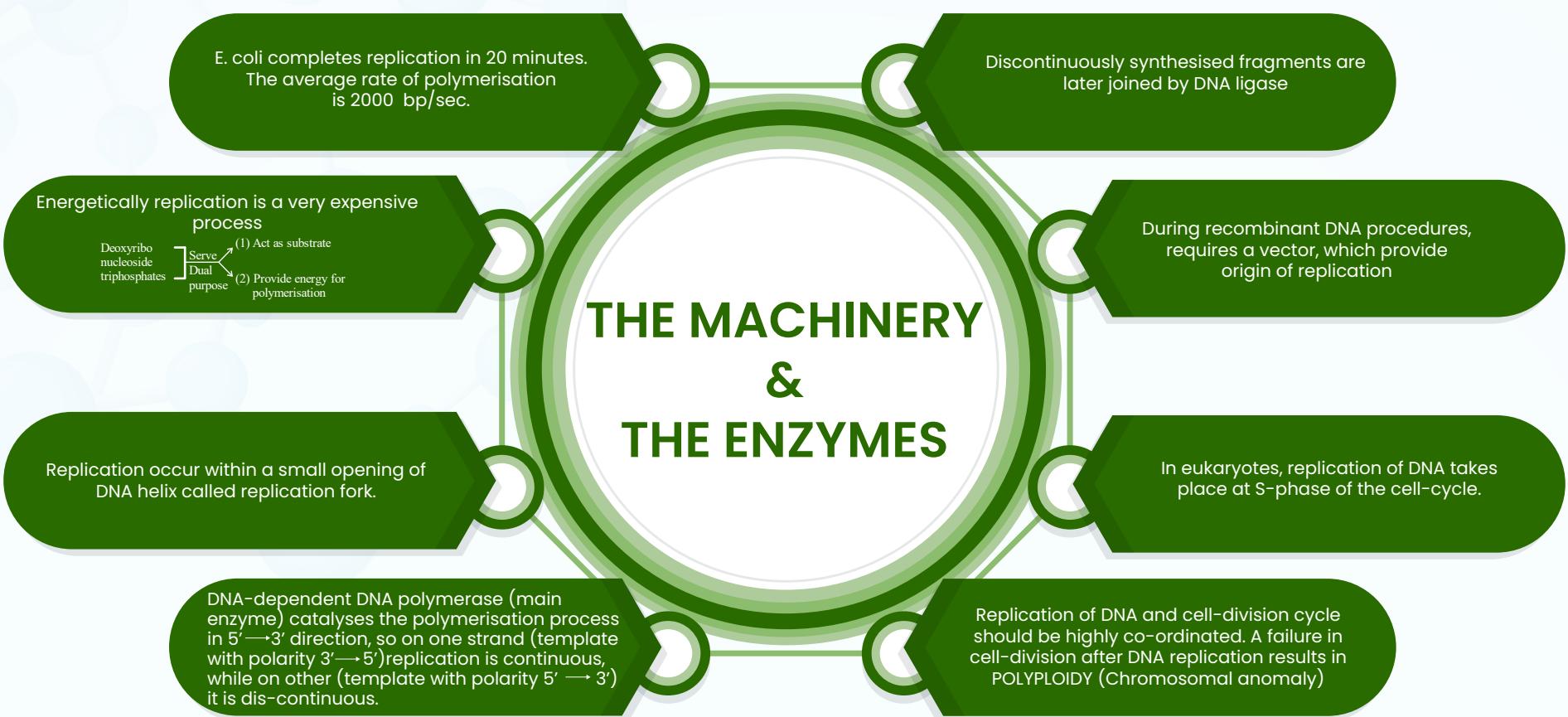
2 Matthew Meselson and Franklin Stahl, performed the experiment (1958) using normal ^{14}N and non-radioactive ^{15}N isotope of Nitrogen as source of NH_4Cl , and centrifugation in a cesium chloride (CsCl) density gradient. The various samples were separated independently on CsCl gradients to measure the densities of DNA. (*E. coli* divides every 20 minutes)

3 Taylor and colleagues (1958) used radioactive thymidine and *Vicia faba* (Faba beans), to prove that DNA in chromosomes also replicate semi-conservatively.



Separation of DNA by Centrifugation

THE MACHINERY & THE ENZYMES



TRANSCRIPTION

- It is process of copying genetic information from one strand of DNA into RNA.
- Principle of complementarity governs transcription (except, adenine forms pair with uracil instead of thymine). In transcription only a segment of DNA and only one of the two strands is copied into RNA. Otherwise, one segment of DNA would be coding for two different proteins. Also, the two RNA molecules if produced simultaneously would be complementary to each other, hence would form a double stranded RNA. This would prevent translation.

TRANSCRIPTION UNIT

A transcription unit in DNA primarily has three regions

A PROMOTER	THE STRUCTURAL GENE	A TERMINATOR
------------	---------------------	--------------

- The DNA-dependent RNA polymerase catalyses polymerisation in only one direction 5' → 3'. The strand that has polarity 3' → 5' act as a template. The other strand with polarity 5' → 3' and the sequence same as RNA (except thymine at place of uracil), is referred to as coding strand
- Promoter is located towards 5'-end (upstream) of structural gene (in reference to coding strand). It provides binding site for RNA polymerase
- Terminator is located towards 3'-end (downstream) of coding strands and defines end of transcription.

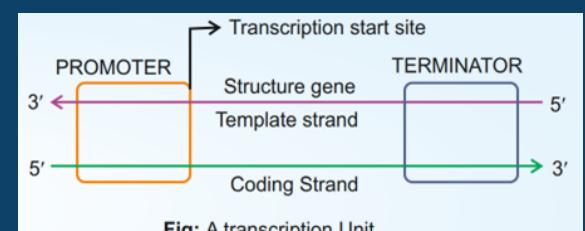


Fig: A transcription Unit

TRANSCRIPTION UNIT AND THE GENE

Genes are located on the DNA which is functional unit of inheritance. DNA sequence coding for tRNA or rRNA also define a gene

Cistron is defined as a segment of DNA coding for polypeptide

The structural gene is monocistronic (mostly in Eukaryotes) or polycistronic (mostly in bacteria or prokaryotes). In Eukaryote genes are split between coding sequences or EXONS, which appear in mature RNA and INTRONS or intervening sequence. The split gene arrangement further complicates the definition of a gene in terms of a DNA segment. Regulatory sequences are defined as regulatory genes, even though they do not code for any RNA or protein.

TYPES OF RNA AND THE PROCESS OF TRANSCRIPTION

A single DNA dependent RNA polymerase catalyses transcription of all three types of RNA (mRNA, tRNA, rRNA) in bacteria.

RNA polymerase binds to promoter and initiates transcription. It uses nucleoside triphosphates as substrate and polymerises in a template depended fashion following the rule of complementarity. It somehow also facilitates opening of the helix and continues elongation.

Only a short stretch of RNA remains bound to the enzyme. Once the polymerase reaches the terminator region, the nascent RNA and RNA polymerase falls off. This results in termination of transcription.

RNA polymerase is only capable of catalysing the elongation process. It associates transiently with initiation factor and termination factor to initiate and terminate the transcription respectively.

In bacteria, mRNA does not require any processing, so transcription and translation are coupled. In Eukaryotes, there are two additional complexities.



There are at least three RNA polymerase in the nucleus (in addition to RNA polymerase found in organelles) and a clear cut division of labour.

ENZYME	FUNCTIONS
RNA pol-I	Transcribes 28S, 5.8S and 18S rRNA
RNA pol-II	Transcribes mRNA precursor i.e. hnRNA
RNA pol-III	Transcribes 5S rRNA, tRNA and SnRNAs

The primary transcript (hnRNA) is subjected to SPLICING, where INTRONS are removed and EXONS are joined in a defined order. The hnRNA undergoes additional processing called capping and tailing to form mRNA. In CAPPING an unusual nucleotide (methyl guanosine triphosphate) is added to 5' -end of hnRNA. In TAILING, adenylate residues (200-300) are added to 3' -end in a template independent manner. Fully processed hnRNA is called mRNA that is transported out of the nucleus for TRANSLATION. The split- gene arrangement represent probably an ancient feature of the genome. The presence of introns is reminiscent of antiquity, and the process of splicing represents the dominance of RNA-world.

GENETIC CODE

George Gamow, suggested, that the genetic code should be triplet.

Chemical method developed by Har Gobind Khorana was instrumental in synthesising RNA molecules with defined combinations of bases. Marshall Nirenberg's cell-free system for protein synthesis finally helped the code to be deciphered. Severo Ochoa enzyme (Polynucleotide phosphorylase) was helpful in polymerising RNA with defined sequences in a template independent manner (enzymatic synthesis of RNA)

Salient Features of Genetic Code:

- 01 The codon is Triplet. 61 codons code for amino acids and 3 codons are stop codons.
- 02 The code is DEGENERATE, i.e., some amino acids are coded by more than one codon.
- 03 The codon is read on mRNA in contiguous fashion i.e., there are no punctuations.
- 04 The code is nearly universal. (eg : UUU codes for phenylalanine from bacteria to humans). Exception = some variations have been found in mitochondrial codons and in some protozoans.
- 05 AUG has dual function. It codes for methionine and act as initiator codon.
- 06 UAA, UAG and UGA are stop terminator codons.

MUTATIONS AND GENETIC CODE

Insertion or deletion of one or two bases changes the reading frame from the point of insertion or deletion and called frame shift mutations.

Insertion or deletion of three or its multiple base, insert or delete in one or multiple codon hence one or multiple amino acids, and reading frame remains unaltered from that point onwards.

tRNA-The Adapter Molecule

Francis Crick postulated the presence of an adapter molecule, that would read the code and bind to specific amino acid. The tRNA, then called sRNA (soluble RNA) was known before genetic code was postulated. tRNA has an anti-codon loop that has bases complementary to the code, and it also has an amino acid acceptor end to which it binds to amino acids. tRNAs are specific for each amino acid. For initiation, there is another specific tRNA that is called initiator tRNA. There are no tRNAs for stop codons.

Secondary structure of tRNA looks like a cloverleaf, though the actual structure is a compact molecule which looks like inverted L.

TRANSLATION

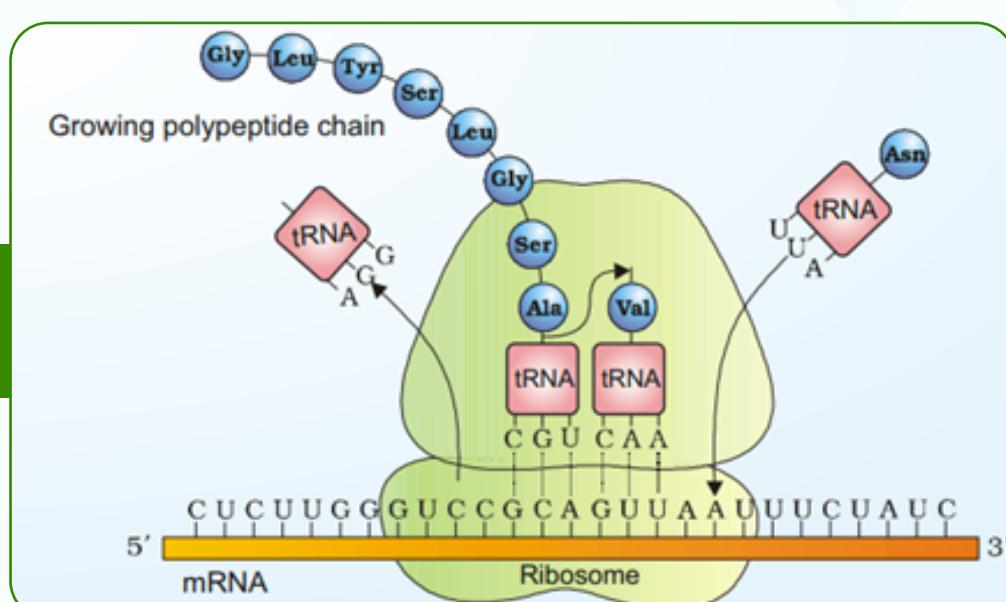
Translation refers to the process of polymerisation of amino-acids to form a polypeptide. The order and sequence of amino-acids are defined by the sequence of bases in the mRNA.

In the first phase amino acids are activated in the presence of ATP and linked to their cognate tRNA by a process called charging of tRNA, or aminoacylation of tRNA. Protein synthesis takes place on the ribosomes

Ribosomes consist of structural RNAs and about 80 different proteins. It has two subunits. When the small sub unit encounters an mRNA, the process of translation begins. There are two sites in the large subunit, for subsequent amino acids to bind and thus, be close enough to each other for the formation of a peptide bond by the catalyst (23S rRNA in bacteria is the enzyme-ribozyme). Presence of catalyst would enhance the rate of peptide bond formation.

A translational unit in mRNA is flanked by a start codon (AUG) and the stop codon. The untranslated additional sequence on mRNA are called untranslated regions, (UTRs), present at both 5'-end (before start codon) and at 3'-end (after stop codons). UTRs are required for efficient translation.

The ribosome moves from codon to codon along the mRNA. Amino acids are added one by one and translated into polypeptide sequences. At the end, a release factor binds to the stop codon, terminating translation and releasing the complete polypeptide from the ribosome.



REGULATION OF GENE EXPRESSION

Gene expression results in formation of a polypeptide. It can be regulated at several levels. In Eukaryotes, the regulation could be exerted at

Transcriptional level
(Formation of primary transcript) **1**

Processing level
(Regulation of Splicing) **2**

Transport of mRNA from nucleus to cytoplasm **3**

Translational level **4**

Metabolic, physiological or environmental conditions regulate expression of genes.

Development and differentiation of embryo into adult organisms are also a result of the coordinated regulation of expression of several set of genes.

In prokaryotes, control of the rate of transcriptional initiation is the predominant site for control of gene expression.

THE *Lac* OPERON

Francois Jacob and Jacque Monod were the first to elucidate a transcriptionally regulated system, the *lac operon* (*lac* refers to lactose), a polycistronic structural gene regulated by a common promoter and regulatory gene, called operon.

Regulation of *lac* operon by repressor is negative regulation however *lac* operon is under control of positive regulation as well.

Lac operon consists of one regulatory gene (*i*) and three structural genes (*z*, *y* and *a*). *i* gene (*i* refers inhibitor) codes for repressor, *z*-for β -galactosidase (β -gal), *y*-for permease and gene *a* codes for transacetylase. All three gene products in *lac* operon are needed for metabolism of lactose.

Regulatory proteins can act both positively (activators) and negatively (repressors).

Lactose is the substrate of β -galactosidase and it regulates switching on/off of operon, so called INDUCER. Regulation of *lac* operon is regulation of enzyme synthesis by its substrate.

Each operon has its specific operator and specific repressor.

The repressor of operon is synthesised (all the time- constitutively from the '*i*' gene)

A very low level of expression of *lac* operon has to be present in the cell all the time otherwise lactose cannot enter the cells. Lactose or allolactose is the inducer of *lac* operon. Glucose or galactose cannot act as inducers for *lac* operon.



HUMAN GENOME PROJECT - (HGP)

Launched in 1990, a 13 year project was co-ordinated by U.S. department of energy and National Institute of Health, Wellcome Trust (UK), Japan, France, Germany, China participated. It was completed in 2003.

Human genome has approximately 3×10^9 bp and the cost of sequencing in the beginning was US\$3 per bp, i.e. 9 billion US dollars. HGP lead to the rapid development of a new area in biology called bioinformatics

Many non-human model organisms like bacteria, Yeast, *Caenorhabditis elegans*, *Drosophila*, plant (rice and *Arabidopsis*) have also been sequenced.

METHODOLOGIES

Expressed sequence tags (ESTs): Focused on identifying all genes that expressed as RNA.

Sequence annotation: Blind approach of sequencing the whole genome containing coding and non coding sequences; needing vectors like BAC (Bacterial artificial chromosomes) and YAC (Yeast Artificial Chromosomes).

SALIENT FEATURES OF HUMAN GENOME

Human genome contains 3164.7 million bp. Average gene consist of 3000 bases.

Largest gene DYSTROPHIN has 2.4 million bases. Total genes estimated at 30,000 .

Almost 99.9 % nucleotide bases exactly same in all people. Functions unknown for 50% discovered genes. Less than 2% genome codes for protein. Chromosome 1 has most genes (2968) and Y-chromosome has the fewest (231).

At 1.4 million locations single base DNA differences (SNPs – single nucleotide polymorphism, snips) occur. SNPs can help in tracing human history.

• The fragments were sequenced using automated DNA sequencer that worked on the principle of a method developed by Frederick Sanger.

• The sequence of chromosome-1 was completed only in May 2006. This was the last of the 24 human chromosomes (22 autosomes and X and Y) to be sequenced

DNA FINGERPRINTING

99.9 % base sequence among humans is same. 0.1 % differences in sequence of DNA make every individual unique in their phenotype. DNA fingerprinting involves identifying difference in repetitive DNA, a small stretch of DNA repeated many times, called satellite DNA. Depending on base composition (A:T or G:C rich), length of segment and number of repetitive units, the satellite DNA is classified into micro-satellites and minisatellites. They do not code for any proteins. They form large portion of human genome and show high degree of polymorphism and form the basis of DNA fingerprinting. Since DNA from every tissue (blood, hair-follicle, skin, bone, saliva, sperm) show same degree of polymorphism, they have forensic application.

Allelic sequence variation has traditionally been described as a DNA polymorphism, if more than one variant (allele) at a locus occurs in human population with a frequency greater than 0.01.

DNA fingerprinting has much wider application in determining population and genetic diversities. Currently, many different probes are used to generate DNA fingerprints.

Polymorphisms are inheritable from parent to child so DNA finger printing solves paternity disputes.

The technique of DNA finger printing was initially developed by Alec Jeffreys.

Steps = Isolation of DNA (I) → Digestion by restriction endonuclease (II) → Separation of DNA fragments by electrophoresis (III)

Detection of hybridised DNA Fragments by AUTORADIOGRAPHY (VI) ← Hybridisation using labelled VNTR probe (V) ← Blotting on nitrocellulose or Nylon membrane (IV)

SIGNIFICANCE

- VNTR are called mini-satellite, a small DNA sequence arranged tandemly in many copies. The size of VNTR varies from 0.1 to 20 kb. So after hybridisation with VNTR probe, the autoradiogram gives many bands of differing sizes. These bands give a characteristic pattern for an individual DNA. It differs from individual to individual in a population except in MONOZYGOTIC twins.
- The sensitivity of the technique has been increased by use of polymerase chain reaction (PCR).