

BIO INFORMATICS

The fusion of Biology, cancer science, information Technology is called Bio Informatics.

Aim:- To identify defective genes and cancer Biology

Genetics defects / disorders:-

- Eg:-
- * Autism → mental retardation
 - * Syndromes (Turner)
 - * Dwarfism

These defects are identified in Embryonic condition by using Artificial intelligence and can solved by "gene therapy".

Gene sequence & function, protein structure and function:-

Abnormal gene → Abnormal protein → Abnormal feature

Protein structure:-

→ Units of proteins are Amino acids.

20 Amino acids → 50,000 proteins in human body

→ By the sequence & no. of amino acids these many proteins are formed.

Egs for proteins:- valine, leucine, lysine

There are two types of proteins:-

i) Linear protein → Eg:- Keratin (in hair)

ii) Globular protein → Eg:- Hemoglobin (in Blood)

study of genes in Bio informatics is "Genomics".

study of proteins in Bio informatics is "proteomics".

"Combination of Biology, chemistry, physics, maths, Statistics, computer science, information Technology is called "BIO INFORMATICS"

History of BI:-

This word BI is known at 1991

- * Ancient name of BI is called computational Biology used to analyze Biological data.
- * In 1951 - pauling & corey worked on protein structure. They said protein having a structure called " α -helix".
- * In 1953 - watson & crick discovered the structure of DNA having a structure of "double helix".
- * In 1953 $\rightarrow$ Fredrick sanger discovered the protein structure for this he worked on the protein called "insulin" (controls ~~sg~~ sugar level)
- * In 1955 $\rightarrow$ Dayhoff experimented on protein db.
- * In 1957 $\rightarrow$ Stantly experimented on "DNA cloning".
- * In 1983 $\rightarrow$ karry mullis invented machine called "PCR" (polym ~~er~~ase chain Reaction).
"PCR will multiply the DNA" (copying)
He got nobel prize for this invention
- * In 1981 - IBM invented "personal computer".

* In 1983 - watermann & smith developed
Algorithm for sequence comparison

Eg: Insulin

o — o — o — o — } normal sequence
val lys leu met

o — (o_{tyr}) — o — o — } Abnormal sequence.

* 1985 → microsoft windows

* 1985 → "FAST P" software was developed.
It is a very efficient Algorithm to analyze
Biological data.

* In 1988 → "SWISS PROT" was developed. It is a
database that analyze protein sequences.

* In 1987 → perl (practical Extraction Report
Language) used for Biological analysis.

* In 1987 → NCBI (National centre for
Bio technology information)

Note:-

* 1991 → BLAST slw is a replacement of FAST
P. used for sequence comparison.

* Dr. craig venter sequenced his own genome
He is the head of HGP (Human Genome Project)

Human Genome project (HGP) :-

* This is the outcome of Bio Informatics.

* Genome is a Genetic material which
contains sequences all the genes in our
genome.

* In 1983 - watermann & smith developed
Algorithm for sequence comparison

Eg: - Insulin

o — o — o — o — } normal sequence
val. lys leu met

o — o — o — o — } Abnormal sequence.
 (Trp)

* 1985 → microsoft windows

* 1985 → "FAST P" software was developed.

It is a very efficient Algorithm to analyze
Biological data.

* In 1988 → "SWISS PROT" was developed. It is a
database that analyze protein sequences.

* In 1987 → perl (practical Extraction Report
Language) used for Biological analysis.

* In 1987 → NCBI (National centre for
Bio technology information)

Note:-

* 1991 → BLAST slw is a replacement of FAST
P. used for sequence comparison.

* Dr. coragi venten sequenced his own genome
He is the head of HGP (Human Genome Project)

Human Genome project (HGP) :-

* This is the outcome of Bio Informatics.

* Genome is a Genetic material which
contains sequences all the genes in our
genome.

* Cell - unit of life $\rightarrow$ 37 Trillion cells in human body.

* Each every cell contains a nucleus.

* Nucleus is called master organ of a cell / Brain of a cell.

* Nucleus contains genetic material (DNA) controls all the functions of a cell.

Aim of HGP:-

- 1) To Identify all the genes of human genome. Human genome is containing 25000 genes.
- 2) To determine the sequence of all these genes.

DNA made up of 4 Basis / Nucleotides

There are 4 Nucleotides

They are A $\rightarrow$ Adenine

T $\rightarrow$ Thymine

G $\rightarrow$ Guanine

C $\rightarrow$ Cytosine

These make 25000 genes

- 3) To store these gene sequence in the form of Database
- 4) To improve the tools for data analysis.
- 5) To Address ethical, social, legal issues (ESLI)

Different areas of BI:-

This is divided into 3 main streams

i) Genomics ii) Proteomics iii) Metabolomics

1) Genomics :- focuses on structural & function of genes / gene sequence.

Eg:- trichosis pinna gene. (located in Y chromosome)

- There are 46 chromosomes 22 pairs + X Y
- This gene is present in hair on ear pinna of boy.

ii) proteomics :- focuses on protein sequence. structure & function of proteomics

Eg:- Amino acids $\rightarrow$ Insulin protein

Insulin produced by pancreas $\rightarrow$ "hormone"

iii) metabolomics :- structure & function of metabolites

Eg:- Urea, creatinine

- * By testing urea and creatinine levels of blood we can identify kidney disorder.
- * These levels are 1mg in blood. if it exceeds that is kidney disorders.

Central Dogma of molecular Biology:-

* Dogma means principle.

* This is a Basic life process in all living cell from bacteria to human. is called central Dogma of Molecular Biology.

"Genetic material (DNA) :- It is a db of life which stores all the characteristics and info"

Replication :- It means duplication. The process of synthesis from parent DNA to daughter DNA.

Transcription:- The process of synthesis from daughter DNA to RNA

Translation:- The production of RNA from protein

- * "proteins determines the character"
- * "protein is the practical / functional working form of DNA"

* DNA $\rightarrow$ Deoxy Ribonucleic acid

RNA $\rightarrow$ Ribonucleic acid

parent DNA

$\downarrow$ Replication

daughter DNA

$\downarrow$ Transcription

RNA

$\downarrow$ Translation

Protein

$\rightarrow$ characters

Replication

- * The synthesis of DNA from parental DNA
- * The Enzyme involved here is "polymerase"

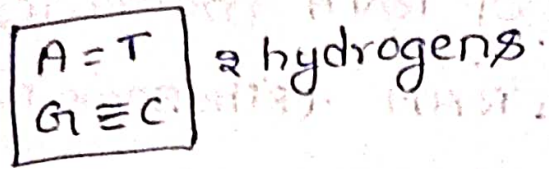


Moving in Opp direc (Antiparallel)
(Double helix)

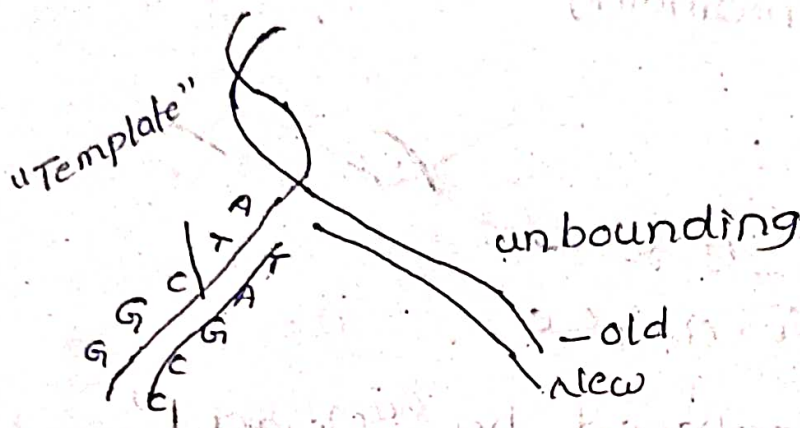
* DNA is a polymer of units of "Nucleotides" (or) N₂ bases

A → Adenine

* Making this Double helix structure DNA follows a rule called complementary base pairing.



* This Rule is developed by "Chargaff's Rule".



Based on template's sequence DNA polymerase will synthesize the DNA.

Next, after it separated and again form the daughter DNA using one old strand and one new strand.

* Every dotted DNA consists of one old strand and one new strand. (semi conservative of Replications).

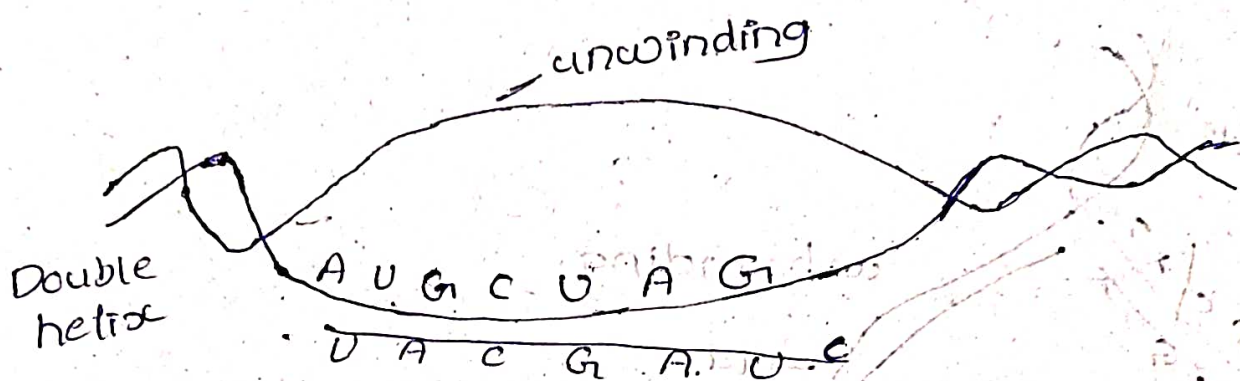
* This process of Replication is done in "nucleus".

Transcription:-

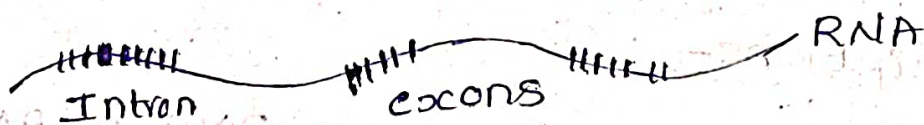
* The synthesis of daughter DNA to RNA is called T.

* The Enzyme involved here is "RNA polymerase".

- * This process is done in nucleus of the cell.
- * Compared to DNA, RNA is highly unstable
mean highly reactive (free) fragile
- * It is a single stranded.
- * RNA means Ribo Nucleus Acid
- * There are three types of RNA molecules.
 - i) m RNA
 - ii) t RNA (transfer)
 - iii) r RNA (ribosome RNA)



- * Here Thymine is replaced by "Uracil"
- * RNA consists of A T U, G, C.
- * The first form of RNA is called primary transcript (or) precursor RNA. This is non functional because it consists of non coding & coding sequences. (non useful)
- * Non coding sequences are called Introns.
- * coding sequences are called exons.



- * All these introns must be removed & all the exons should be connected. This process is called "RNA splicing".

5' cap
Methyl guanine

3' poly A tail

These Methyl guanine & poly A tail are very important to increase the stability of RNA.

Steps of RNA splicing

1) Removal of Introns, joining of exons.

2) 5' end $\rightarrow$ cap was added.

3) 3' poly A tail was added.

Now, This is called as functional RNA.

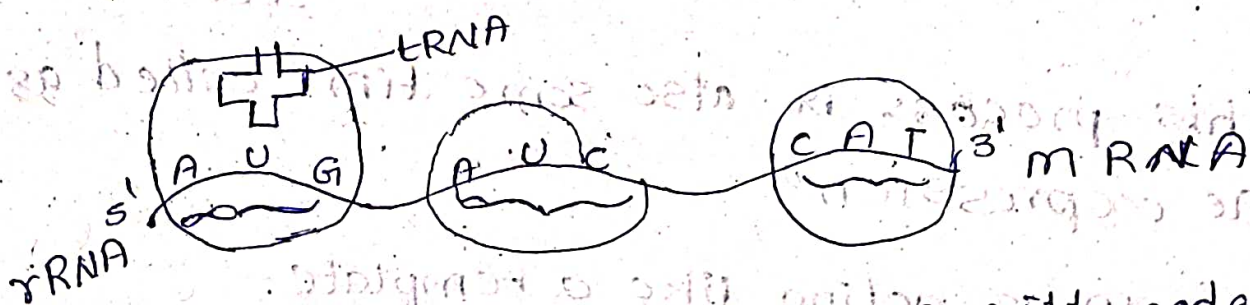
Translation:-

* Translation is also takes place in cytoplasm of the cell.

* The synthesis of protein from RNA is called Translation.

* The cell organelle that are involved in protein synthesis is "Ribosomes" (protein factories)

* After splicing mRNA (messenger RNA) is formed.



* Here 3 $\times$ 2 basis of RNA will code for one Amino acid

* eg:- AUG $\rightarrow$ methionine, eg: UAC $\rightarrow$ Tryptophan

eg:- CAU $\rightarrow$ cysteine.

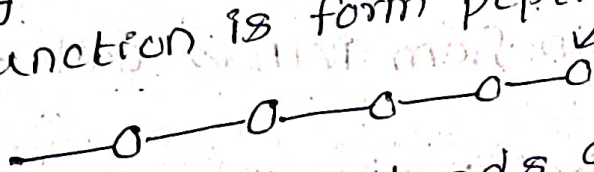
* This 3 N₂ basis of RNA will code for Amino acids This is called "Genetic code"

* Ribosomes are consists of RNA.

* tRNA transfers the Amino Acid from Amino Acids pool to protein cell's sight



By Peptidyl transferase present in Ribosome function is form peptide bonds.



* Many peptide bonds are together called Polypeptide chains

* The characters / feature are decided by the protein by this process eg:- height, colour, IA

* So this process is also some time called as "gene expression"

* The mRNA is acting like a template.

* AUG is called as start codon

Protein Structure:-

It can be studied at 4 levels

i) Primary Structure:-

* A Linear seq of AA which are joined by peptide bond

* Proteins are function less (non-functional)

ii) Secondary Structure:-

The primary str further folds to form secondary structure



In this it have hydrogen bonds and peptide bond.

It is two types

i) α -helix:- It is imparted from soft nature of protein



→ helical coiled

ii) β -pleated sheet



→ much harder than α -helix

* Very few proteins are functional in their secondary structure. (%10)

Eg:- Keratin

iii) Tertiary Structure:-



- * It is a compact & globular structure
- * Majority of proteins in nature are functional in their tertiary structure

* tertiary structure stabilized by 4 bonds

1. Hydrogen bonds

2. Hydrophobic bonds

3. Ionic bonds

4. Disulfide bonds :- The sulphur contained enzyme involved in this
eg:- cysteine

* It form 3-D structure in this

eg:- hormones, enzymes, Myoglobin
(protein of muscles)

iv) Quaternary structure

* It not exhibited by all proteins. It is exhibited by special proteins called oligomeric proteins

* proteins are divided into 2 based on polypeptide

i) Monomeric :- It made up of single peptide chain. ~~~~~ eg:- Myoglobin

ii) Oligomeric :- It made up of > 1 peptide chain
eg:- Hemoglobin ~~~~~

↓
Structure is

* One molecule of Hb is carrying 4 molecules of oxygen

* Most of proteins are monomeric in our body.

