**GENOMICS**

* **GENOMICS =** Study of structure and functions of a genome of an organism – concerns with the sequencing and analysis of an organism’s genome
* **GENOME =** total DNA content that present within one cell of an organism
* Term “GENOMICS” introduced by – **Thormas Roderich** 🡪 scientific method of mapping, sequencing and analysing and making the use of genetic information for further use in multifarious area 🡪 study of molecular organisation of genomes, their information contents and the gene products they encode
* **Types of Genomics –**

1. Structural genomics
2. Functional genomics
3. Comparative genomics
4. **Structural Genomics**

* Process of finding out the sequences of genome
* Deals with DNA sequencing, sequence assembly, sequence organisation and management.
* Attempts to determine the structure of every protein encoded by the genome, rather than focusing on one particular protein.
* Starting stage of genome analysis i.e. construction of genetic map or sequence maps of high resolution of the organism.
* Includes systematic and determination of 3D structure of proteins found in living cells 🡪 because proteins in every group of individuals vary and so there would also be variations in genome sequences

1. **Functional Genomics**

* Study and understand the function of gene
* Reconstruction of genome sequences (through structural genomics) 🡪 useful to find out the function that the genes do
* Gives an idea of function of all gene sequence and their expression in organism
* Different tools useful for structural genomics – bioinformatics sequences, DNA chips, 2D gels etc.
* Lends support to design experiment to find out the functions that specific genome does
* Strategy – based on systematic study of single gene or protein to all genes 🡪 ∴ large – scale experimental methodologies characterise the functional genomics
* Provide novel information about the genome
* Development of microarray technology and proteomics 🡪 explore the instantaneous events of all the genes expressed in a cell or tissue present at varying environmental conditions like temperature, pH, etc.

1. **Comparative Genomics**

* 1/3rd of the genes encoded on each genome – No predictable or known function. e.g. in *E. coli* K12 about 40 % genes have unknown function
* Function of these genes – predicted by –

1. comparing different genomes
2. transferring functional annotations of protein for better studied organisms to their orthologs (the same gene in different species that connect) as opposed to paralogs i.e., genes related by duplication within the genome from less studied organism

* Includes – analysis of protein sets from completely sequenced genomes.
* General purpose databases + Organisms specific databases – used for comparative genomics

**Methods used for Whole Genome Sequencing**

* SEQUENCING = Determining the order of bases
* WHOLE GENOME SEQUENCING = a laboratory procedure that determines the order of bases in the genome of an organism in one process
* 3 steps –

1. Cloning of the DNA to be sequenced
2. Sequencing reactions + Electrophoretic separations
3. Analysis of ensuing data

* Following are important methods of whole genome sequencing –

1. **Chemical Methods –**

* **Developed by –** Maxam and Gilbert
* **Steps –**

Restriction fragment of DNA – labelled with 32p at either its 5’ or 3’ – using either of the enzymes polynucleotide kinase or terminal transferase

↓

From a restriction map – an enzyme selected to remove a small piece from one end of the molecule leaving just one end labelled

↓

DNA 🡪 chemically cleaved at specific residues in 5 different reactions

↓

Reactions – partially completed 🡪 partial digestion products – separated on a polyacrylamide gel and autoradiographed

↓

Fragments having the labelled terminus seen

1. **Whole Genome Shotgun Sequencing –**

* **Developed by –** Venter and Smith
* **Steps –**

**STEP 1 – Library Construction –**

1. Chromosome isolated from the desired cells following the methods of molecular biology
2. Isolated DNA – randomly fragmented into small pieces using ultrasonic waves
3. Fragments – purified and attached to plasmid vectors
4. Plasmid with single insert – isolated
5. A library of plasmid clones – prepared by transforming *E. coli* strains with plasmid that lacked restriction enzymes

**STEP 2 – Random Sequencing –**

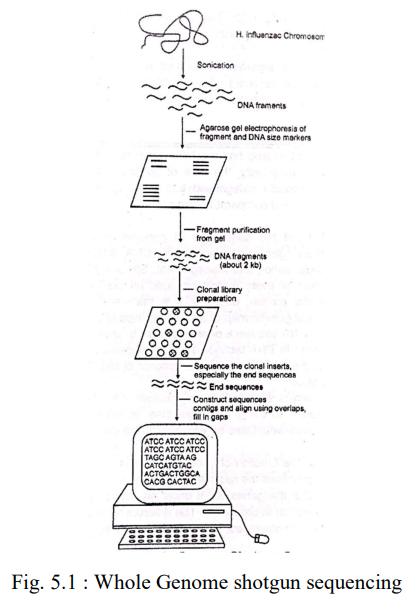
1. DNA – purified from plasmid
2. Thousands of DNA fragments – sequenced using automated sequencer – by using primers labelled with special dyes
3. Normally with universal primers – thousands of templates were used – recognise the plasmid DNA sequences next to bacterial DNA insert
4. Whole genome – sequenced several times

**STEP 3 – Fragment – Alignments and Gap Closure –**

1. Using special computer programme – compare nucleotide sequence overlaps between fragments 🡪 sequenced DNA fragments clustered + assembled into longer stretches of sequence
2. 2 fragments – joined to form a large stretch of DNA (if the sequences at their ends overlapped and matched) – OVERLAP COMPARISON METHOD
3. Resulted in a set of larger contiguous nucleotide sequence – called **contigs**
4. Contigs – aligned in a proper order to form the completed genome sequence

**STEP 4 – Proof Reading –**

1. Any ambiguities in the sequence – resolved
2. Sequence – checked for the presence of any frame shift mutation 🡪 if so, the mutation is corrected

****

**PROTEOMICS**

* Study of all the proteins produced by a cell
* Identification, analysis and large – scale characterisation of proteome expressed by any cells, tissues and organs under the defined conditions
* **Objectives of proteomics –**

1. To characterise post – transcriptional modifications in protein
2. To prepare 3D map of a cell indicating the exact location of protein

* **PROTEOME =** total protein component in a cell or organism
* **PROTEOMICS =** deals with the study of proteomes
* **Marc Wilkins –** coined the terms ‘proteome’ and ‘proteomics’
* Understanding of alteration in protein expression during different stages of life cycle / under stress condition
* Understanding the structure and function of different proteins as well as protein – protein interactions of an organism
* A minor defect in protein structure / function / alternation in expression pattern – easily detected using proteomics studies 🡪 Important for drug development 🡪 proteins – most favourable targets for various drugs
* **Proteins –**
* Macromolecules + long chains of amino acids
* Amino acid chain – constructed when the cellular machinery of the ribosome translates RNA transcripts from DNA in the cell’s nucleus. The transfer of information within cells commonly follows this path from DNA to RNA to protein.
* **Types of Proteomics –**

1. **Structural Proteomics –**

* Study of structure and nature of protein complexes present in a particular cell organelle
* Mapping out the 3-D structure and nature of protein complexes present specifically in a particular cell organelle
* **AIM –** to build a body of structural information 🡪 help predict the probable structure and potential function for almost any protein – from knowledge of its coding sequence
* Help assembling information about protein – protein interactions + architecture of cells to explain how the expression of certain proteins contributes in cell’s unique characteristics

1. **Functional Proteomics –**

* Use of proteomics techniques – to analyse the characteristics of molecular protein – networks involved in a living cell
* **Recent developments –**

Identification and analysis of molecular protein networks involved in the nuclear pore complex (NPC) in yeast

Helps understand the translocation of molecules from nucleus – to 🡪 cytoplasm and vice versa

1. **Expression Proteomics –**

* Quantitative study of protein expression between samples differing by some variable
* Can compare – Pattern of expression of the complete proteome / its part (sub – proteome) between samples
* Useful in identifying disease specific proteins
* Eg - over expression or under – expression of proteins in cancerous cells and normal cells – taken from a cancer patient and a normal individual, respectively 🡪 can be analysed using various techniques – 2 – dimensional gel electrophoresis, mass spectrometry, microarray, etc. 🡪 can help understand development of cancer 🡪 facilitate development of drugs to treat cancer

**METHODS USED IN PROTEOME ANALYSIS –**

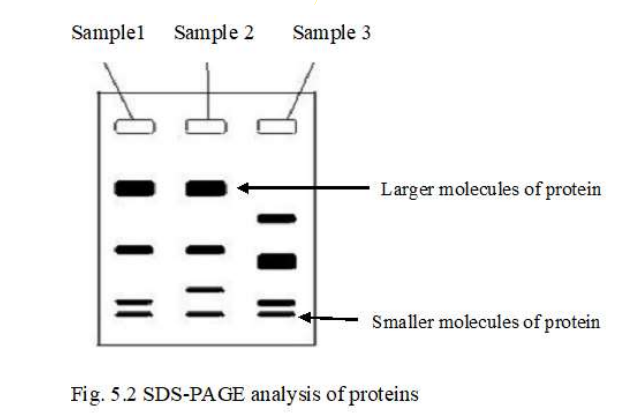
1. **2 – Dimensional Electrophoresis**
2. **SDS – PAGE –**

* Stands for 🡪 “Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis”
* A high – resolution method
* Used universally for analysing the mixture of proteins according to their respective size
* **SDS solubilised in soluble protein** – makes possible the analysis of the other insoluble mixture
* Used to separate – proteins with similar charge : mass ratio (z/m)
* Proteins – first treated with an ionic detergent called sodium dodecyl sulphate (SDS) before electrophoresis (PAGE)
* Identical proteins are denatured by SDS resulting in their sub – units – polypeptide chains get opened and extended
* Molecules separated on the basis of mass – not charge
* Electrophoretic separation is normally used for –

1. gel acts as **molecular sieves** 🡪 hence separates the molecules on the basis of their size
2. gel **suppresses conventional currents produced by small temperature gradient** 🡪 improves the resolution

* Polyacrylamide gel is used due to –

1. chemically inert
2. stable over a wide range of pH, temperature
3. transparent
4. Better for size fraction of proteins

****

Proteins are denatured 🡪 have negative charge with a uniform charge to mass ratio (z/m) when treated with SDS

↓

Proteins migrates towards anode at alkaline pH through PAGE gel during electrophoresis

↓

Smaller polypeptides moves faster followed by the larger polypeptides

↓

Intrinsic charge on proteins masked – Separation based on size

* **Applications of SDS – PAGE –**
* Molecular weight of the separated protein – analysed by comparing the molecular weight of the standard protein and its mobility
* Analysis of a complex mixture of proteins 🡪 resolution is improved by the initial movement through a stacking gel 🡪 Final bands in the separating gel are sharper and focused in better way

1. **IEF –**

* Biomolecule like proteins have electric charge which depends on molecule to molecule and conditions of medium (pH of buffer in which dissolved).
* Due to the differences in amino acid composition 🡪 proteins have net charge or iso-electric points (no charge) as a given pH of buffer.
* **Application –** Separation of atmospheric substances such as proteins which differ in their isoelectric points
* **Isoelectric point** – pH value at which the net charges on molecules are zero

Ampholytes (i. e. complex mixture of synthetic polyamino – polycarboxylic acids) – Introduced into gel to create the pH gradient (wide range from 3 to 10 / narrow range of 7 to 8)

↓

Potential difference applied across the gel

↓

Molecule having difference in isoelectric points by a little as 0.01 pH unit can be separated.

Proteins migrate depending on their charge – until they reach a region which pH corresponds to respective isoelectric points at which pH proteins possess no net charge and hence got focused

1. **Mass spectrometry**

* Employed fixed magnetic and electric field – to separate ions of different mass and energy
* 2 – dimensional electrophoresis 🡪 more powerful when coupled with mass spectrometry
* The unknown protein spot – cut from gel 🡪 cleaved by trypsin digestion into fragments 🡪 analysed by mass spectrometer 🡪 mass of fragments is plotted
* This mass finger print – used to estimate the probable amino acid composition of each fragment and tentatively identify the protein
* **Applications of MS –**

1. Study of proteome and its charges
2. Study of covalent modification of proteins which can affect their activity
3. Identification of unknown compounds
4. Determination of structural and chemical properties of compounds when present in small amount (10-4 -10-8 g)

* **STEPS OF MS –**

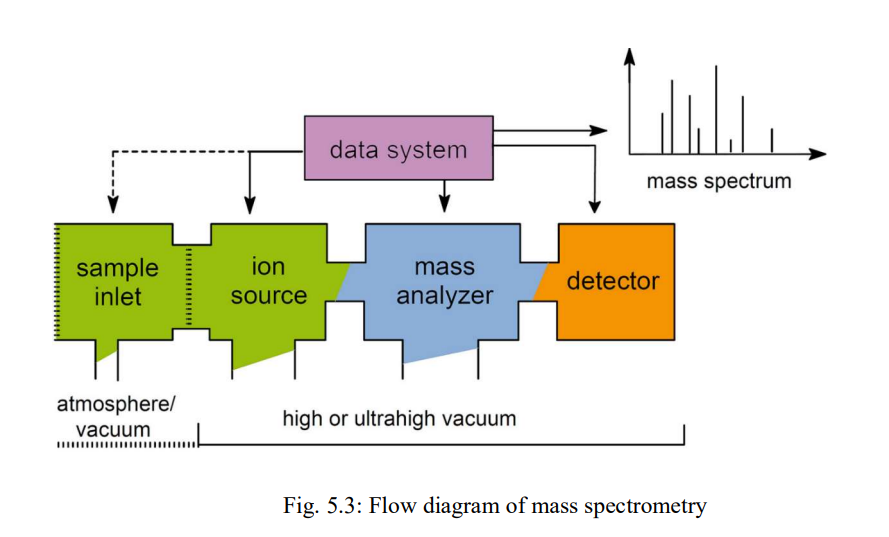
Production of ions of the material in sample

Their separation on the basis of their mass change (m:e)

Determination of relative abundance of each ion

* **3 Components of MS –** source of ion + an analyser + a detector
* Does not directly measure the molecular mass 🡪 but detects m:e ratio.
* Mass is measured in terms of Dalton (Da).

One Dalton = 1/12th mass of a single atom of isotonic carbon (13 C)

****