Detailed report on the talk on 23rd June 2021

Assignment 2

V Rudra

Btech Biotech semester 5

MRIIRS

**1st presentation by Dr. Kannan Krishnamoorthy**

**evolution of protein sequencing single pure protein to the mixture of proteins in a single cell”.**

In 1944 Dr.Sanger was the first person to work on the genetic problem called diabetes as it wreaked havoc in the older days. Reduced the life expectancy to 4-5 years for the kids who were tested positive for diabetes.

It was a time when people used primitive methods for the testing of proteins. Ninhydrin was used to test the presence of proteins and it was discovered that the all the amino acids are optically active and show the L configuration, only some amino acids showed D configuration.

But even with such problems Dr.Sanger only used paper chromatograph to separate the amino acids or peptides.

It was already known back then that the islet cells of Langerhans are responsible for diabetes but there was no idea how it affected so Banting and Best flushed the proteolytic enzymes from the acni cells surrounding islet cells by treating with secretin. The atrophied islet cell extracts was then injected to the diabetic dogs. Dramatically glucose level came down and dog was active again , thus finding out that pancreas is responsible for the diabetes, so the animal model checked out. The same procedure when tried on juvenile diabetic child and it established the therapeutic use of purified extract from cow pancreas. Thus, a lot of people travelled to Toronto from us and other places of the world to get the medicine for diabetes. During this time Eli Lilly came out with a therapeutic treatment and did not patent it thus making it available to everyone for a cheaper cost, and they didn’t claim it was a cure they stated clearly as their medication was a treatment that one has to continue. Finally in 1951 the structure of insulin came out which was founded by Fredrick Sanger.

Explanation and understanding about the properties of organic compounds were explained by Dr. Kannan Krishnamoorthy. Later on he talked about the paper chromatography and separation using ninhydrin, he explained how butanol acetic acid is used instead of any other salt in paper chromatography as other salts will just evaporate away when heated to 100 degree Celsius.

Fredrick Sanger found out that even though molecular formal was discovered (C(257) H(383) N(65) O(77) S(6)), but there was no point to it as the proper structure was unknown, but by using the acid hydrolysis they confirmed 18 amino acids were present although some amino acids like tryptophan asparagine glutamine cystine are destroyed or modified by this method. Molecular weight was not determined as that depended on the method.

Sangers’ contributions from 1944 to 1955 were

* Labelling amino acids with FDNB
* Demonstrated that Insulin is made up of Two Chains
* Paper Chromatography to identify DNP Amino Acids.
* Breaking intra and inter molecular disulphide bonds with performic acid
* Separating and purifying the two chains and labelling them as A and B
* Focussed on B Chain sequencing first. Tried mild acid hydrolysis create peptide Fragments. Used trypsin enzyme to produce large peptides.
* Separate the mixture of peptides with paper chromatography and elute the bands into test tubes
* Identify the newly created AA and write down possible sequence of peptides
* Solving the jigsaw puzzle from overlapping sequences of purified peptides
* Protein Finger Printing
* Radio labelling of AA
* Established how amino acids are arranged in a specific sequence in Insulin
* Mutation leads to change in amino acid in the sequence.

Strategy of sequencing the b chain of insulin in 1949 by sanger and tuppy

* The first chain, labelled A, had glycine at its terminal, and
* The second, labelled B, had phenylalanyl at its end.
* Progress was hampered by the fact that acid hydrolysis did not produce sufficiently long chain fragments
* Hydrolysis with a proteolytic enzyme produced sufficiently large fragments without affecting their amino acid arrangement
* Within a year Tuppy had identified and determined the sequence of all 30 amino acids in chain B.

Exploiting the power of DNA polymerase sangers method

* Used DNA Polymerase power to add nucleotide or subtract nucleotide (Plus and Minus)
* Editing function of DNA Polymerase switched off(Used Klenow Fragment)
* Judicious use of mixture of radioactive Phosphorous dNTTP's
* Using the right primer to create new strand
* Using gels to run the reaction mixture
* Autoradiography of gel
* Patterns appear as ladder
* Read of the ladder

**Session 2 by Dr.Sanjeev Galande**

**Topic- Proteomics- reinventing the frontier in biology**

Proteomics and genomics have been working together but genomics has gained the upper hand in the past but recently new methods have been invented which are changing the world of proteomics as new methods are being invented to understand different aspects of biology.

Proteomics is a tool in the branch of biology where one uses various methods to understand the complement of proteins, their modifications, and interactions. As human genome project took place the misconception of having millions and billions of genes was cleared. It was discovered that human body has a limited no. of genomes we had 21000 protein coding genes, and some other genes. The misconception of organisms that came before humans were less complex in a genome perspective, for example the number of genomes in humans are compared to the fruit fly were less than double. So the genes are seemingly comparable between 2 organisms that are extremely distant species with a very huge difference of bodily functions, the makeup of the body, the body developmental program and even the morphology. This explained that the number of genes may look very limited coding capacity in terms of genes, what matters is how the information is unpacked or used or translated. Coding genomes is really comparable to a balloon, the balloon will keep growing and growing as air is blown into, similarly, the encoded genes when translated grow greater and greater in size.

Genes have multiple splicing events, so large no. of exons and introns go through multiple p and cs we can generate a no. of rearranged genes and therefore different proteins. There’s not just the rearrangement but also chance of forming different complexes so for example protein me complex with B or C, or D so we can imagine a no. of combinations.

Phosphorylation can drive change from one nuclear compartment to another nuclear compartment from inside of the ER, to outside of the ER, from outside to inside from cytoplasm to nucleus. Proteins keep shuttling between the subcellular compartments which is essential for their function. Because if they are in a wrong compartment, it will be like sequestering them.

The biggest change in proteomics was the sensitivity as the zeptomole sensitivity was very high they were able to detect proteins which were in very very small quantities

Challenges in proteomics-

Proteomics: The challenge

The proteome:- dynamic - highly complex-relative protein abundances in a cell can differ by several orders of magnitude

Proteomics aims to analyze the levels and structure of all proteins present in a cell or a tissue including their post-translational modifications

Proteomics approaches include:

1) protein identification

2)protein quantitation or differential analysis

3) protein-protein interactions

4) post-translational modifications

5) structural proteomics

Various types of proteomics-

* Mass spectrometer based proteomics-This area is most commonly associated with proteomics and is a method to determine which proteins are expressed and the amounts of those proteins.
* Array based proteomics-This method uses various arrays to try and define the function of the proteins, regulation levels and interacting partners within the cell.
* Informatics-This is trying to define what information will be needed, stored, accessed and how it can be used to study the proteome of the cell.
* Single cell proteomics - new frontier
* Clinical proteomics - applications

Established & emerging proteomic technologies

Proteome-wide Systems Genetics to identify Functional Regulators of Complex Traits

Proteomic technologies now enable the rapid quantification of thousands of proteins across genetically diverse samples. Integration of these data with systems genetics analyses is a powerful approach to identify new regulators of economically important or disease-relevant phenotypes in various populations. In this review, we summarize the latest proteomic technologies and discuss technical challenges for their use in population studies. We demonstrate how the analysis of correlation structure and loci mapping can be used to identify genetic factors regulating functional protein networks and complex traits. Finally, we provide an extensive summary of the use of proteome-wide systems genetics throughout fungi, plant, and animal kingdoms and discuss the power of this approach to identify candidate regulators and drug targets in large human consortium studies.

**lecture 4- Dr. Jagannath Swaminathan**

**fluorosequencing and the emerging landscape of single molecule protein**

The first question that is asked is why proteomics is hard something like next Gen sequencing for proteomics? rna has only four nucleotide whereas proteins have twenty amino acids and hundreds of post translational modifications. DNA can be amplified whereas proteins it is not possible to amplify proteins. And the final difference is mrna is readily soluble and is easily extractable and the final difference between dna and protein is the dynamic range that is present in the cells in RNA the dynamic range is ten to the power 4 whereas in proteins it can be 10 to the power 9 and by dynamic range we are referring to the expression of least abundant species to the most abundant species.

The problem started with how to take analogies from next Gen DNA sequencing for proteomics. So what was first done was a very parallelisable system just like DNA sequencing in next Gen dna sequencing not applicable for the protein biomolecules, and ven then there was no way of using PCR for protein. So they faced a very big hurdle even though there were method in proteins like optimising and affinity reagents that people use for identifying proteins but it comes with some limitations, because you do not know what the affinity reagent should target, for example if we have new phosphate modification then we have to design a new affinity reagent. The third and final approach is nanopore or tunnelling pores, which is extremely viable but they too come with intrinsically extreme challenges.

A new strategy came up with involves stealing bids from mass spectrometry and some very important old technology called admin degradation for sequencing by degradation.

Fluorosequencing

So in proteins there are so many amino acids that it would be very hard and not very practical to tag each and everyone of them so if we know the positions of few amino acid residues in the peptide backbone, that information is sufficient to identify the protein information. So labeling just for amino acids we can identify up to 95% of the human proteome in almost 20 edman cycles. So we can say that the partial composition of amino acids is enough to make this type of protein inferences.

There are three components to the process of fluoro sequencing technologies-

1. Selective amino acid lableing.
2. Single molecule microscopy
3. edman degradation

They use a process called total internal reflection microscopy, and this method is used to capture signals/Florissant signals at extremely high signal background noise add the glass water interface.

Edmon degradation-this was a process that was invented in 1950s in which a sequential removal of one amino acid per cycle takes place and leaves the rest of the peptide backbone intact.

The peptide at the end terminus reacts with the base and under basic conditions with a molecule called phenyl-isothiocyanate it conjugates to it and under acidic condition the first amino acid pops off. Then they take what is pop off and Rani through fluorescently tagged or stabilized and then run it through HPLC column and the retention time would tell that amino acid what it is. And then if we run it enough number of times we get a denovo sequence. So when we put the immobilised population on the glass slide and these diluted molecules and then when we look at the glass slide under a microscope it would look like stars in the sky. Each shiny sport is a single peptide molecule and they are fluorescently tagged. Once the tagging is completed in different colours, then Edmon degradation takes place.