Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis

Prof. Arindam Ghosh

Dr. B.C. Roy Multi-Speciality Medical Research Centre

Indian Institute of Technology Kharagpur

Lecture 26 : Overview, Importance, Methods in Proteomics

Namaskar. Hello everyone, welcome back to the lecture series on Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis. We will be starting module 6 and over the next two weeks we will be discussing proteomics alright. So, today's topic is overview of proteomics the importance of proteomics as well as the steps of proteomics and we will be covering those we will be starting proteomic steps and outflows, but we will be first defining what is proteomics, what is the need or utility of proteomics, why do we need to study proteomics at all, what are the types of proteomics alright. Then proteomic methods ok. So, we will be going deep into it alright.

So, strap yourself. So, let us first start with the nomenclature. So, genome ok. So, this let us revise something that you already might be knowing from previous classes alright.

Genome how do you define a genome? Genome can be defined as sum total of all genes in an organism alright and the study that the branch discipline of study that deals with this is actually genomics. So, study of sequence of all genes in an organism right. Similarly, when we are defining transcriptome, transcriptome can be defined what what is transcriptome actually it is a sum total of all RNA transcripts ok in an organism that can that an organism can make in it is life time ok. Not all genes actually transcribe for important products for example, proteins that will be coming very soon. So, transcriptome mix is the study of all these level of RNA transcripts that are produced by many genes in a cell in a given time alright.

So, now if we extrapolate these concepts, it will be very easy for us to define these terms on our own. So, proteome actually properties and activities of all the proteins that an organism can make in it is life time right. So, here basically the word prote is running hand in hand to the word genome, you can just replace that and you will get the definition right. So, protein is actually complementing the genome in this case. So, basically you can say proteomics is the qualitative and quantitative comparison of proteomes under different conditions alright.

We are studying them to unravel many biological processes. So, how can we define right proteomics? You can define if you have understood the concept you can define it in any language you want. One nice way to define it will be it is a scientific discipline concerned with the systematic analysis of proteins present in cells at a given time under given conditions alright. Not I mean under given condition if the condition varies the expression of protein will change and therefore, the output of the proteomics study will change. So, that is important mind it DNA remains same gene remains same.

So, we will be seeing how proteomics the need of proteomics right. So, what does proteomics include? It includes identification, characterization, quantitation of the entire complement of protein in a cell tissue in order to why do you need this in order to understand their function in relation to the life of the cell ok. So, if we need molecular diagnostics we have to understand we have to know proteomics ok. There is no going without proteomics. So, this is the central dogma which you already know right from the first week.

So, in DNA when we are studying everything related to DNA genome when the DNA transcribes RNA messenger RNA transcriptome and finally, the product of the gene transcription when the RNA is translated to form protein and when you are studying that gives rise to proteome alright. So, this is basically the blueprint the genetic library alright this is the gateway the path and finally, this is the execution this is the final result ok you can I mean analyze that using this analogy right. So, why do we need proteomics? Well mainly because of the fact that the level of transcription of a gene is not same as level of expression of that gene means the rate or the amount in which a gene is transcribed to the amount of the protein or the nature of protein that it ultimately translates it is not the same they are drastically different right. One more thing messenger RNA degrades very rapidly and is translated inefficient that is mainly the reason that there is no correlation I mean there is not absolute correlation there is not one to one correlation between the same between transcription and translation. Moreover even after the translation if even if we consider the translation is 100 percent efficient there are many post translational modification which drastically alters the nature and function of the protein there are translocations right.

Again one gene or one transcript may code for many proteins and those proteins actually function differently the converse is also true one protein can perform can be generated in many processes and one protein may perform many functions ok. We need to study them all right and even one single function can be executed by many proteins. So, you see there is a complex network and if we pin point if we go from end to the beginning I mean if we go from the disease perspective if we pin point a disease we not only we have to locate what is the main genetic abnormality or maybe what is the genetic reason behind it we cannot omit the role of proteins that are actually playing the field role

ultimately the end role in executing all the functions that are leading to the disorders all right. So, if we consider genome versus proteome the basically genome is the blueprint as I told you it is the blueprint which gives us idea about what to do all right, but ultimately what will be done is actually governed by the proteomes right. So, if we go into that is one thing and what is the need of studying genomics proteomics separately there could have been I mean we could have them we could study them horizontally right, but you might find that when we are studying genomics of a certain diseases it appears very simple, but whenever we are going to proteomics it is much complex rightly so, because there are multigenic diseases ok one disease that can be caused by multiple genes ok.

Next whenever that I told you already after even after normal protein is formed it may undergo some modification to form some deformed protein or there may be some alteration in the property ok that is also leading to diseases right. So, in order to understand I mean pinpoint the diagnosis of the disease the nature of the diagnosis in order to I mean identify targets for the drugs we need to study both genomics and proteomics, but for these two for this all right both genes and proteins are very important both genomics and proteomics are must. However, for specific drug targets in almost all bearing few cases without studying proteomics it is very very very difficult even in case of novel drug discovery we have to go for artificial simulation molecular dynamic simulation molecular docking all those things are very essential and for that we need knowledge about proteins ok. So, next we will be discussing the types of proteomics all right. So, types of proteomics can be classified into three categories well proteomics can be classified in multiple ways in later part of our course we will be we will be classifying proteomics in some other way not this there is an approach on top down there is approach on bottom up, but that will be that is actually application based and once we have completed all the applications it will be much easier for you to understand top down and bottom up proteomics, but for this class we will start with these fundamental categories number one is expression proteomics the second variety can be structural proteomics and another is functional proteomics all right.

From the name itself it is easier you can actually try to coin what are these terms ok. In expression proteomics what is expression proteomics it is a definitely a discipline sub discipline or a variety of proteomics that studies changes in protein expression how the protein is expressed both qualitatively and quantitatively under different conditions. So, if we change the condition the protein expression definitely changes depending on what situation the cell is in all right. So, this type of study detects variation in protein expression in different cell such as for example, in hypoxia in a tumor tissue it is subjected to a hypoviric stress or hypoxic stress or in case of any other oxidative stress the tissue is under in a free radical assault anything can lead to change in expression of proteins how the proteins act right. So, this variety of proteomics uses techniques like 2D

gel electrophoresis mass spectrometry to detect and quantify proteins do not worry we will be going in details in all of the methods that are being mentioned here all right.

Next structural proteomics all right. So, it is concerned this discipline is concerned determining the 3D structure of the proteins all right. This uses 3D processes for to view the 3D structure for example, x-ray crystallography nuclear magnetic resonance that is NMR spectroscopy determine the precise arrangement between the atoms and the protein all right. So, those these processes are definite very established biophysical process with these process not only we can study the structure of any protein we can study the structure of any molecule any crystal we will not be diving deep into this just knowing the names is enough for you just know this that structural proteomics can also be used to investigate protein-protein interaction how two proteins are interacting with each other in real time also how proteins are interacting with other molecules such as DNA or RNA. So, this branch or this sub discipline of proteomics is known as structural proteomics.

And lastly we will discuss functional proteomics definitely the name itself is very easy to understand it suggests this branch of proteomics helps us in understanding the functions of proteins and their interaction with other molecules of the cell. It helps us in identifying various protein complexes are interaction that are involved in specific biological processes as well as the individual role of the proteins in these complexes mind it we are focusing on the exact function. Students often find it confusing because I just mentioned structural proteomics can also be used to investigate protein-protein interaction this is just the attachment how they are approaching what are the force fields that are causing any interaction change in interaction, but this is actually the function the functional output how two complexes are related what are they doing individually. So, there is a drastic difference in these three types of proteomics all right. For example, multi enzyme complexes again domain of functional proteomics.

So, we will be next going into I mean going to discuss the basic steps in proteomics workflow. Now, proteomics workflow involves a lot of steps we will be starting our lecture series for the first few classes with the very basic steps all right basic steps. So, see there are actually multiple traditional proteomic techniques for example, ELISA, western blotting those are actually very important and in context of infectious diseases they are a must know area, but in our module of proteomics for the next two classes we will be not discussing those techniques. However, they will be definitely discussed and taught when we were discussing infectious diseases in the later modules all right. So, for now for these two modules we will be discussing several high throughput technologies newer advancements all right, but to understand the newer advancements we must understand the basic steps how we need to go ead with those with understanding those techniques right.

So, the first three steps are definitely in considering any variety of proteomic workflow are sample preparation, then separation and isolation of the proteins and identification and characterization of those proteins all right. Now, these can be done by many methods. Sample preparation we will be discussing definitely if we are concerned about separation isolation of proteins the methods that we will be discussing here in this course for most in that concerns I mean that carries the most importance for you in any type of examine in understanding high throughput proteomics are two gel based approach and chromatography based approach. And if we and we will also discuss the sub varieties that is one dimensional electrophoresis gel electrophoresis are called 1DE, two dimensional gel electrophoresis also known as 2DE all right. In chromatography we will be discussing ion exchange chromatography, size exclusion chromatography, affinity chromatography and liquid chromatography very important all right.

And we will also be discussing HPLC or we will I may have just spoiled your lecture because this was a thing that would have come later, but this is very important for you to know because if you are answering if examiner wants an answer from you if examiner asks you this question what are the techniques for separation isolation of protein it might be so that they are looking for this answer only right. So, HPLC is a definitely modified version of LC, but you never know all right. So, if you answer liquid chromatography definitely right, but satisfying the examiner might be challenging in some cases and also for some multiple choice questions HPLC might become the best answer. When we are studying about protein identification characterization again there are many methods, but we will be mainly focusing on mass spectrometry thereafter introduce some database searching for these proteins we will be discussing DIG differential in gel electrophoresis and also protein microarrays all right. So, in totality this is what we will be discussing in the first few lecture classes ok.

So, we will start with sample preparation. Now sample preparation is actually very crucial step because every in order to prepare protein from any sample in order to carry out proteomic data sample can be anything it can be blood sample it can tissue sample it can be bacterial sample infectious disease sample anything we need to make sure in order to study I mean go for proteomic study you have the right amount of proteins in the sample right. So, how to do it? So, it is a crucial step in proteomics which involves extraction and purification of proteins from biological samples for further analysis. The next step will follow after this step right. So, what do we do generally the extraction of proteins from a cell right every living organism has a cell we need to rupture that cell.

So, in order to rupture that cell it involves combination of physical and chemical method. So, this is basically the goal is there is an intact cell and the cell has many type of protein it can have be mitochondria you can study any cell organ it can have other than the nucleus there are multiple cell organ you can study the cell membrane. So,

whatever be your area of interest you can study Golgi-Wold lysosome whatever you first need to make sure you can isolate the organelle all right. So, the first step is definitely rapturing the cell and making sure you get the right protein. So, that is done by using a combination of physical and chemical methodologies right.

Now, many organic solvents and detergent are commonly used in sample preparation to enhance the protein solubility to disrupt the cellular membranes to release the intracellular proteins all right. For example, tissue division techniques for such as mechanical homogenization there is a mortar and a pestle you can apply mechanical force there is a motor motorized mortar ok. You that is tailor made I mean that fits snugly in the diameter of the glass tube and you can simply put your sample in a test tube and once the motorized mortar is rotating homogenizer you can simply put the test tube in such a way that the samples are mechanically homogenized. Thereafter you need to treat it with lysis buffer there are many steps we are not going into a details, but the basic concept is rapturing of the cell lysis of the cell using mechanical homogenization or ultrasonic sound. For example, sonication again one of very effective method which lysis the membrane and helps in breakdown of tissues and release of proteins right.

One very important thing to note if you want to target any specific organelle ok you need to do something known as differential centrifugation very important very important multiple choice question that might prop up any time. So, in order to isolate various cell organelles the method of choice in sample preparation is differential centrifugation. So, what do we do in differential centrifugation we first we take the homogen homogenate or cell lysate in which all the organelles are suspended all right we spin it in centrifuge all right. So, when we spin it in the centrifuge the heaviest organelle will settle down first. So, that is known as pellet all right and there will be a liquid part which is floating above that is known as supernatant all right.

So, what do we do? So, suppose when we spin for 500 or 600 g we get the nuclear pellet ok. So, we take the supernatant now there is a clean supernatant in another tube all right we again centrifuge it using high speed for example, 10000 g now. So, that will precipitate another pellet for example, mitochondria precipitates in 10000 g we have got the supernatant. So, so on and so forth we can do some this serial centrifugation with different speeds. So, that different pellets are achieved and depending on our area of interest we can standardize these experiments and this technique is known as differential centrifugation all right one of the very commonly used techniques in sample preparation preceding all proteomic experiments all right.

So, regarding lysis buffer you need to know that these are the few chemicals that are commonly used. Now lysis buffers are also commercially available under very name for example, RIPA buffer, TRIS buffer right. So, what are the common ingredients? TRIS-

HCL, trimethyl aminomethane that is the form chemical name of TRIS you do not need to know right TRIS-HCL is enough. Salt, sodium chloride, tritonics 100 these are detergent sodium deoxycolate, sodium dodecyl sulfate again detergent, HEPES, tritonics 100 right ethylene diamine tetra acetic acid, glycerol, beta mercaptoethanol, dithiothreatol DTT. So, these are all the ingredients that are commonly used in various lysis buffers in almost all lysis buffers you might find one or more of these ingredients ok.

So, this is how the first step happens you lyse the cell you rupture the cell for sample preparation. Next step is separation and isolation of proteins one I already mentioned differential centrifugation. So, let us focus on gel based approach which is mainly used in high throughput proteomics alright. So, gel based approach let us first discuss one dimensional gel electrophoresis or the traditional gel electrophoresis also known as sodium dodecyl sulfate, polyacrylamide gel electrophoresis or SDS-PAGE. We already you have heard this name when we are discussing the genomic lysens alright PCR how we can resolute those gels ok.

So, what is happening in case of when we are discussing gel electrophoresis in relation to DNA, there different fragments of DNA were loaded in multiple lanes and they were run in an electric field depending on their molecular weight the lightest gel a lightest DNA fragment moved quite forward the most forward in all those sequencing methods for example, Sanger sequencing, Maxam Gilbert sequencing the same phenomena right. In case of proteins again same thing will happen the heaviest one will be the last to move the slowest to move and the fastest one will be the easiest to move. However, in case of protein electrophoresis we are not only banking on their mass, but also the charge alright because protein has got different charge as also different mass. So, we need to make sure that both the properties are taken care of during electrophoresis. So, for that we use sodium dodecyl sulfate it is a very mild detergent what does it do it is used to denature the proteins and impart uniform negative charge that is proportional to their mass alright.

So, this uniform negative charge allows separation of proteins based on only size during electrophoresis alright. So, something that is known as M by Z ratio very very very important. So, M is the mass and Z is the charge and B in gel electrophoresis we separate protein based on their M by Z ratio right. So, that the size determines the M by Z ratio and it is imparted by sodium dodecyl sulfate. Now in one dimensional gel electrophoresis if we just look into the technique briefly this is a gel which is actually casted in a gel casting apparatus there are multiple ingredients that we need to add for example, a TEMED ammonium phosphosulfate APS that we already discussed in pyrosequencing that same APS.

So, polyacrylamide gels so, there are actually two types of gels that we need to cast on

one on top of another right. The lower one that is the resolving gel it has got a pH of 8.8 the solution is made in such a way we need to have tris buffer right. And the upper one the stacking gel the pH is 6.

8 and where the lanes are actually there. So, we need to load the samples in the stacking gel lanes they stack they form a uniform stack at the interface and then we start the electric current and then the gel is run ok. So, this is how it happens after the thing is loaded we connect the apparatus to a power supply this is a vertical gel electrophoresis apparatus vertical right. We all we have also seen apparatus where the DNA are running like this from left to right or that is horizontal gel electrophoresis apparatus. For example, agarose gel electrophoresis always preferred is mostly preferred in horizontal electrophoresis apparatus this is vertical gel electrophoresis apparatus alright. So, SDS proteins are then separated in a polyacto-almonite gel matrix using electric field and you already know the drill the smaller proteins move quickly through the gel while larger protein moves at a slower pace and ultimately we get something like this.

We can also load molecular weight marker which are mixture of proteins which have got known bands I mean known fragments. So, we already know that if we so, for example, there are 12 bands and we already know that the lower band is supposed to be 20 Kg per kilo Dalton that is the molecular weight of protein that unit and the upper band is 25 Kg and we already know the fifth band for example, is 37 Kg and the sixth band is 50 Kg. So, if there we have some unknown protein which we run in parallel and if they land in between we can easily understand for example, in this case the unknown protein is 40 kilo Dalton. So, it is very quick process to separate as well as understand the characteristic of the protein. Two dimensional electrophoresis takes it to one step forward where the whole thing is actually the first step is known as isoelectric focusing.

So, in the first step the proteins are separate on the basis of their isoelectric point that is p i in which the I mean isoelectric point means thus pH at which the protein has no charge. The buffer is made in such a way so that the first the proteins will there is a gradient and the proteins will be separated based on their isoelectric pH. And we actually then run an electrophoresis in a direction which is perpendicular to the first direction. So, now, in the first step if it is the isoelectric focusing that is happening the second step it will be run parallel. So, in the second step is SDS page where the proteins are separated based on their molecular weight right.

So, this separation by two different parameters allows the better differentiation of proteins because if you are running in one dimension there might be protein which has got same molecular weight, but they might have different isoelectric pH. And if you run only isoelectric focusing there might be proteins they have got same isoelectric pH, but they might different molecular weight. So, this differentiation is very important.

Therefore, 2D gel electrophoresis has got much more resolution than 1D new eventually electrophoresis, but it has also got some limitations that we cannot visualize very larger very low proteins all right. So, this is how the two dimensional gel electrophoresis looks like and in reality these are the real world two dimensional gel electrophoresis outcome graphs that we are electrophorogram where we can see.

For example, if we are analyzing two different to similar proteins are the different condition from the dots we can see that in the first diagram compared to the first diagram the second diagram these boxes this has been over expressed in the control and under expressed in the sample these few dots have been over under expressed in the control and over expressed again we can see these spots and based on these spots we can get very good information about different proteins all right and their pH. The next step is chromatography based approach all right. So, separation isolation of proteins we are done with the gel based approach. So, what is chromatography based approach? Chromatography is basically an important biophysical technique that enables separation identification of and purification of components for qualitative and quantitative analysis right almost similar, but how is it different? First of all we do not need any current in electrophoresis we need a current electric field right. So, here the concept is a sample mixture a mixture of different ingredients that we need to separate is separated into various components as they move through different phases.

For example, I dissolve a material in one ingredient and I pass the material through another ingredient and the material will react differently. This is the concept of stationary phase and mobile phase that is very much used in chromatography. So, the ingredient the thing which is dissolved in is known as mobile phase and the phase through which the mobile phase is passed is known as stationary phase right. And the different interaction the different types of interaction with the stationary phase cause them to travel at different varying rates. It is very easy because I think in school level and even in your basic science undergraduate experiment you might have done this where spot of black ink, this is the paper chromatography.

If you dip it in a mixture of water any organic solvent after sometime when water I mean seeps up or soaks the whole paper you might see it will separate different dyes because each dye has got a different migration speed alright. So, this is the very basic paper chromatography experiment based on which further concepts will be discussed. Now, let us first understand liquid chromatography. So, liquid chromatography is actually separation technique in which the mobile phase is a liquid right. This is a very basic definition the mobile phase has to be liquid.

This is one of the most used methods in proteomics to separate various complex mixture of proteins based on their size, charge or chemical affinity towards any substrate or

stationary phase right. This is an example of a mobile phase or liquid that is passed through a column alright the circular packing of chemical. So, when we are passing the liquid the material in the column is reacting differently with the materials that are dissolved in this green liquid and by the time the liquid is coming out they will be separated because each and every ingredient of the liquid is acting interacting differently with the stationary phase alright. So, be very careful I mean be very clear about the concept that liquid chromatography is the chromatography subtype where the mobile phase is liquid. Now, it can be of many types I mean just mention three different types of interaction that a sample can have with the substrate.

So, let us discuss them for example, size exclusion chromatography right. So, these type of chromatography where there is a column is also known as column chromatography right. So, liquid chromatography can be carried out in a column it can also be carried out in a plane ok. Anyway so, size exclusion chromatography here the stationary phase consist of for example, beads porous resins ok. So, the molecules that are dissolved will be interacting differently with the stationary phase how since it is a size exclusion chromatography we are trying to separate two molecules one is bigger in size and one is smaller in size.

So, as the molecule that are dissolved in a liquid mind it the mobile phase has to be liquid because we are discussing liquid chromatography here right. So, as they are passing so, what happens the smaller so, these porous resins have got pores in them and they can actually trap the smaller molecules actually enter the pores and they get trapped they cannot come out whereas, the larger molecules they do not enter the pore and they can simply bypass through the sides and they are coming out first right. So, when we are collecting them differentially first we will see the bigger molecules that are coming out and then we will see slowly the smaller molecules are coming out. So, after they are dissolved we can treat them with another buffer that is known as elution buffer e l u t i u n elution buffer then the smaller particles can come out on their own or automatically depending on the gravity anything right. So, this size exclusion chromatography ion exchange chromatography similar concept where the chromatographic technique is used to purify proteins based on their charges.

So, now, we are separating two different type of charged proteins. So, one protein might be positively charged and one protein might be negatively charged. We will be designing the wall of the column in such a way so, that it attracts opposite charges. For example, here we see the wall is positively charged. So, definitely it will only attract the negatively charged proteins they will be tightly attached to the walls and the because negatively charged molecules are immobilized to the cation surfaces right cation surfaces are immobilized in the column and only we will be getting the positively charged proteins all right.

So, next we can so, after the positively charged proteins are separated now we can add specific elution buffers that will again displace them from the attached cations. Now, as you have understood we can design in such a way so, that the immobilized anions are present on the wall. So, depending on type of columns or depending on the type of stationary phase ion exchange chromatography can be classified into cation exchange or anion exchange chromatography. Next affinity chromatography again it is using the property of interaction of the target analyte with the immobilized receptors on the column right. So, this affinity chromatography separates protein based on the interaction with the immobilized ligand.

Now, here we see the ligand for example, for easy understanding we have designed in a hexagonal fashion. So, only the desired proteins that have strong interaction with the ligand will get attached and the other proteins will be coming out and finally, when you are eluting those target proteins that were bound with the ligand can be separated ok. So, these are three types of liquid chromatography. Well in modern day this thing I mean this liquid chromatography has been greatly improved and it utilizes very small packaging particles that are tightly packed in a column. This there is relatively very high pressure and the solvent that is mobile phase is pushed through the column using pump and they can be easily separated in small amount of time and it can be detected using computer that is known as chromatogram.

The final outcome of all chromatography in which we are getting curves how we can separate is known as chromatogram alright, but we will be stopping our discussion here because in the next class we need to will be discussing this high performance liquid chromatography in details. So, in summary we have discussed various nomenclatures and definition, we have discussed the utility of proteomics, we have discussed the genome versus proteomes, we have discussed types of proteomics how we can classify proteomics based on the different types and methods, we have also discussed steps in proteomic workflow where we have discussed the gel based and chromatography based methods. So, these are my references for today's class and I thank you for your patient hearing. Thank you.