LECTURE-09

Sample preparation for proteomics applications

TRANSCRIPT

Welcome to the proteomics course. In today's lecture we will talk about sample preparation for proteomics applications. As you know a very good sample is essential to perform any good experiment and especially when you want to characterize thousands of proteins in a given experiment. It is very essential that you start with a very good sample. An ideal sample preparation step ensures that you have no contamination; you have very good protein yield and no interfering substances are present in your extract. Often each type of biological sample poses its own challenges. For example if your working on bacterial cultures, plants, human samples, different body fluids such as serum, urine, saliva or cerebrospinal fluid. Each sample type bring in its own unique composition and its own challenges. Now when you are performing sample preparation you have to be very cautious about the sample you are processing. You can not follow a generic protocol. You have to optimize the conditions depending upon your sample. An ideal sample preparation will ensure that you have all the protein present in your sample without any contaminations such as nucleic acid, salts and other interfering components.

In today's lecture we will talk about the sample preparation for proteomics application. I will give you a workflow for protein sample preparation such as the first step, how to disrupt the cell, how to lyse them, how to protect the cells during the lysis step so that there is no proteolysis occurring, how to fractionate the samples; often you need to simplify the proteome, you need to ensure that the sample is simple enough to show the good protein throughout the proteome coverage. To obtain a comprehensive proteome coverage often it is important to pre-fractionate your samples. Protein extraction and solubilisation, these are another very common components. Now when we are talking about the sample preparation this is quiet generic for different type of proteomic applications, whether one wants to use for gel-based approaches or gel-free

approaches. Whether you want to 2-DE, DIGE or different types of gel-based applications or you want to perform mass spectrometry and different types of label-free techniques. You have to ensure that you are starting with a very good protein extract. But sample solubilisation and some other components are added more when you are performing gel-based proteomics. So during this lecture I will talk about various types of components which are essential to make a good sample preparation for proteomic applications. I will give you stepwise workflow.

So let's start workflow for proteomic sample preparation.

So the proteome is very complex whether you want to perform your proteomic analysis for the whole organism. It means you want to know all the proteins present in a given organism or in a tissue or in body fluids or in different types of cells. Proteomics can be global or it can be very targeted or expression based. So highly reproducible samples are very important for performing comparative proteomic analysis. If you want to know the difference in your sample as compared to the controls you need to ensure that your sample preparation is very reproducible. If you introduce some artifacts to begin with then obviously you are not going to identify reproducible biological changes.

So let me give you three different terminologies here for proteomic analysis; global proteome analysis, expression proteome analysis and targeted proteome analysis. When I am talking about global proteome analysis it means your aim is to characterize all the proteins present in the given sample. Expression proteome analysis it means you are mainly interested to look for those changes which are due to any chemical or treatment, those are induced either going up or down protein amount is changing. So the protein expression analysis that is most commonly used for various types of clinical and different studies. Targeted proteomic analysis; if you are very focused for a given organelle or a given sample type often you would like to know what is happening in that particular proteome, for eg. mitochondrial proteome. So one need to try different types of strategies when thinking about performing a sample preparation. What is your objective, whether you want to do global profiling or you want to do expression profiling,

in either case you need to extract all your proteins present in that particular target sample. Now when you are looking at targeted proteome analysis you just want to pre-fractionate your sample in such a way that only that particular component is isolated and all the proteins from that organelle or cell are being extracted. Different types of strategies need to be used to perform these types of proteomic analysis.

Now all of this sample processing involves solubilisation, denaturation, reduction and treatment of sample proteins. But you need to involve additional steps depending upon the type of samples and your type of objectives so that the quality of the protein extract can be improved and while you are doing this you have to be very cautious that when you are performing various steps and sequential types of extractions you may also lose a small fraction of the proteins. So one has to be careful when adding various additional steps during the sample preparation.

Now protein extraction protocols need to ensure that most if not all the proteins in a cell or organelle are extracted. The presence of interfering compounds should be minimized. So if you have optimized a very good protein extraction procedure that should ensure that you have a very wide coverage and that is ultimately going to determine the success of your proteomic experiment.

So the steps involved in ideal protein sample preparation include solubilising the proteins present in a mixture, preventing protein aggregation, denaturing and reducing all the proteins which are present in that mixture, removal of nucleic acid and other contaminants as well as removing salt and some other small interfering components. Again depending upon your sample type you may have to think what different types of components are could be present in that sample type. If you have plant roots you have to get rid of phenolic components. Similarly you have to think specific sample types and what could be the major contaminants present in that sample. For example, serum is rich in lot of salt components. You need to get rid of that. So an ideal protein sample preparation should involve all of these steps as I mentioned previously.

So I am giving you the guidelines for sample preparation. One should start with findings good reference from the literature as a starting point and try to modify the protocol depending upon the objective of that experiment. You need to ensure that you remove the nucleic acids, salts and different particulates. Prepare the samples as freshly as possible and store at -20°C in small aliquots. You should avoid repetitive freezing and thawing steps of the sample.

Why a good protein sample preparation is important? A good sample preparation includes all the proteins present in that mixture and it will provide you high quality data as there will be less interference from the artifacts. A good sample provides reproducible results. You have to perform biological replicates and technical replicates of a given experiment. So a good sample will provide very reproducible results. Once you have optimized a protein extraction protocol then you can apply the same protocol for the large studies. For example, if you are performing a clinical trial study on 200 patients. So once you have optimized the protocol with a small population then the same protocol can be applied for large number of samples. So that is going to ensure the success of the clinical study. Now if you are able to remove lot of contaminants and artifacts present in the sample then your signal to noise ratio will improve. You have to literally see your real signal and do not worry about your background and the noise.

So I am giving you the workflow of protein sample preparation. Let's go step by step.

Sample preparation: The protein extraction should be performed from the source material and then we need to solubilise the protein before doing the analysis. The ideal sample will disrupt all non-covalent bonds present with the proteins and it will remove other interfering components.

So the workflow of the sample preparation could involve cell disruption or cell lysis, protection from proteolysis, sample fractionation, protein extraction and solubilisation, removal of contaminants and quantification. Obviously this workflow can be modified depending upon your sample type and few steps can be moved in that sequence. So, let's follow this workflow and during that workflow I will give you some examples of

different types of samples, how to extract proteins from those and what type of challenges these samples are going to impose.

Cell disruption or Lysis: The lysis is very important step because first of all you would like to break open the cells and remove all the cell components outside so that you can get good protein yield. So cell disruption or lysis is very important and it is often challenging because all the sample types can't be lysed with the same method. So the lysis strategy has to be modified depending upon the sample type.

So why need cell lysis? To facilitate the effective disruption of cells or tissues, to isolate the proteins from intact cells and tissue while avoiding the loss or modification of proteins, to obtain all the proteins which are present in a given sample and to help to maximize the sample recovery and retain the structural integrity. So cell lysis is very important due to all of these factors.

There are different steps involved in cell lysis, you need to disrupt the cell, protect from the proteolysis during the lysis step, homogenize and solubilise your sample. I will describe all of these in more detail.

So cell lysis can be performed with a gentle way or in harsh conditions depending upon the type of cells you want to disrupt. When you are employing gentle disruption method you need to think that you have to break open all the cellular components so your gently disruption should be efficient enough to disrupt the cell. So if you are looking at those cell types which can be easily lysed such as blood cells or culture cells, then you can involve the gentle disruption methods. There are different types of lysis methods available such as a) osmotic lysis in which one can suspend the cell in hypotonic medium, b) detergent lysis; you need to suspend the cells in detergent solution, c) enzymatic lysis; if you are using plant cell one can used cellulase enzyme, if you are using bacterial sample one can use lysozyme. There different types of enzymes present which are used for enzymatic lysis, d) freeze-thaw; this is one of the very commonly used methods for gentle disruption which involves rapid freezing and thawing cycles. You need to cool your sample in a very cold condition like liquid nitrogen and then

immediately put in the boiling water. While, doing this rapid freezing and thawing the cells will break open.

Now there are different types of challenges being imposed by different type of cells. If the cells are very difficult to lyse then you need to involve rigorous disruption methods such as sonication, French pressure, homogenization or manual grinding. We will talk about different types of cell lysis methods and some of the principles involved as we go along with more specific type of samples. But in all the sample preparations mostly people use manual grinding or homogenization by using bead beater or polytron homogenizers. So these are very commonly used homogenization methods. The manual grinding is performed using motor pestle often very efficient and if you have very less sample and you want to avoid any contamination or you have some samples which are going to pose challenges with motor pestle then you need to use the homogenizers. Then sonication and French pressure are used more when we are applying bacterial, yeast or different types of cells which are difficult to lyse. We will talk about in more detail in more specific example when we talk about extraction of proteins from bacterial samples.

So now I will give you the overview of lysis methods. We have talked about all of these methods briefly. I am giving you an overview now. One can use detergent if your target sample is tissue culture cells this method is going to provide gentle lysis step. Enzymatic lysis is also a gentle lysis method which can be used for plant tissues, bacterial cells and fungal cells. Freeze-thawing can be used for bacterial cells or cultured cells and this is a gently lysis method. French pressure is often applied on bacteria, algae and yeast and is a vigorous lysis method. Glass beads are used with cell suspensions or organism with cell wall and this is a rigorous lysis method. Grinding of solid tissues and microorganisms is another vigorous lysis method. The mechanical homogenization of solid tissues is another vigorous lysis method. Osmotic lysis can be used for blood samples and tissue culture cells which is a gentle lysis method. Sonication can be used for cell suspensions and other bacterial samples, which is

another rigorous lysis method. I hope these methods give you some options to lyse your cells effectively.

Now you have tried to break open your cells and release all the protein components and other cellular components present inside the cell. But often during this process of grinding and lysis you may degrade some of the proteins and you need to ensure that you have to protect your proteins from the proteolytic activities of various enzymes. So the protection of your proteins from the proteolytic step is very important during the protein sample preparation.

So cell lysis will release various proteases which may results into proteolysis. During the sample preparation one can use different types of protease inhibitors which can minimize the artifactual proteolysis.

The effective protease inhibitors contain a mixture of different types of protease inhibitors; irreversible and reversible, which inhibit serine, cysteine and different metalloproteases. This step is more important when you are preparing your samples for gel-based proteomic applications.

Now I will give few specific examples of protease inhibitors such as phenyl methyl sulphonyl fluoride (PMSF) which is very effective against serine and cysteine proteases. It can be inactivated by DTT and it is unstable when you are preparing samples for gel-based applications. Ethylene diamine tetraacetic acid (EDTA) is effective against metalloproteases and also inhibits nucleases. So EDTA can serve both purposes of protecting from proteolysis and partial inhibition of nuclease activity. Ethylene glycol tetraacetic acid (EGTA) is also effective against metalloproteases. When you are making a sample preparation for proteomic applications it is very important that you process the samples in cold conditions to reduce any proteolysis step. These proteolytic artifacts can be minimized by using cold conditions during the grinding as well as during the centrifugation steps.

Now let's talk about sample fractionation. As I mentioned in the beginning depending upon your applications you need to decide a strategy effective way of obtaining all the proteins so that you can study your biological problem in an effective way. For example, if your are looking at a particular organelle, for proteome analysis of a given particular organelle then you need to fractionate your sample or if you are looking for wide proteome coverage but the proteome is posing lot of challenge to provide all the protein mixture separation simultaneously then one need to do fractionation to reduce the complexity of the sample.

Why to perform fractionation? The sample fractionation makes it possible to isolate the group of proteins or fractions from a given total proteome. By doing the fractionation you can simplify the analysis of complex protein mixtures. This method can further allow the improved resolution of proteomic data and then when you are obtaining the data from individual fractions and pooling it together it is going to be more informative. When you are applying your protein samples on 2-DE gel or other types of gel-based proteomic techniques this will provide you less crowded protein map. So again fractionation can serve good need in gel-based and gel-free proteomic applications.

So let's talk about different types of fractionation methods available. Simplest for looking for organelle or specific type of proteome, people use different type of centrifugation or ultracentrifugation when you are looking for sub-cellular fractionation of organelles such as mitochondria or chloroplast or cell compartments such as plasma membrane. Different types of chromatography methods can be used for fractionation. If you are interested in looking at all the serum proteins but there are certain proteins which are highly abundant in the serum such as serum albumin protein. So by using affinity chromatography methods, one can remove those highly abundant proteins so that all that proteins in that given mixture can be well resolved. Now affinity chromatography methods can be used for fractionation in different contexts. Sequential extraction, which again going to simplify your proteome, is based on solubility. Different types of chemicals are used so that in a sequential way one can extract the proteins. Few proteins may more soluble in one composition of mixture and other proteins are more

soluble in other solubilisation buffer. By applying different types of chemicals one can obtain the bigger coverage of proteome in doing the sequential extraction. Electrophoresis can be used for fractionation. For example, if you are using gel-free method and you want to directly extract the protein and analyse that using mass spectrometry. So rather than applying the whole sample directly with the liquid chromatography one can first simplify the proteome by isoelectric focusing and doing the IEF process in the liquid phase itself. The liquid phase IEF fractionation can simplify the proteome based on the isoelectric point. In the next lectures when I talk about gel-based and gel-free techniques I will describe you a new technique being used, OFFGEL fractionation which is simplifying the step of isolating the proteins based on isoelectric point in the liquid phase. So one can collect the fractions of different isoelectric point range and then perform the further proteomic analysis.

So we have talked about different types of fractionation strategies and what is the significance of fractionation methods? The foremost important thing is that we would like to increase the proteome coverage. You would like to analyse as many proteins as possible from that given sample. An effective fractionation method separates highly abundant proteins from low abundant proteins of interest and brings them into the dynamic detection range. Often the low abundant proteins could be the proteins of interest which are being masked by the high abundant proteins. If you are able to separate the high abundant proteins then probably you can increase the dynamic detection range and bring everything in the same range. Fractionation can also increase the chances of identifying low abundance proteins or diagnostic or therapeutic interest.

So far we have talked about performing cell lysis, then how to protect proteins from proteolysis. We have then looked at different types of strategies people involve to fractionate the proteome. So now let's talk about protein extraction and solubilisation. This step will be more towards talking about gel-based proteomics where solubilisation will be more important.

After performing the sub-cellular fractionation so that the proteins can be enriched which you are going to be analysis in your experiment. So protein extraction in the aqueous buffer one can follow different types of procedures either use Tris-HCl method followed by desalting, protein precipitation by trichloro acetic acid (TCA), acetone alone or TCA and acetone. I will give more specific composition and recipe when I talk to you about specific types of examples how to perform protein extractions from serum, bacteria and plants. So protein extracts should be soluble and free from protein-protein interactions, protein-DNA or protein-RNA interaction. Similarly there are different types of cellular components present and those should be effectively removed. No metabolites should be interfering in your protein extracts.

Sample solubilisation is important because proteins naturally form complexes with membranes, nucleic acids as well as other proteins. So to avoid all of these issues sample solubilisation is very important.

There are different components being used in sample solubilisation. Let's discuss one by one. First of all let's talk about chaotrops; urea and thiourea. Urea is used as a denaturant which can solubilise and unfold most of the proteins to fully random conformations. Urea is a chaotropic agent, which helps in stabilization of the proteins, unfolding proteins so that all the ionisable groups are exposed to the solution. Thiourea improves solubilisation of membrane proteins more specifically. Mostly both urea and thiourea are mixed together during the solubilisation step.

There are different types of detergents which are also used in solubilisation such as SDS. SDS is very efficient in solubilising hydrophobic proteins. But due to its anionic nature it limits it's effectiveness for the conventional proteomic analyses. SDS, anionic detergent, is not compatible with isoelectric focusing. So if you are preparing your protein preparation to perform 2-DE, SDS should be avoided from the sample solubilisation. If your objective is extract the proteins and separate them on SDS-PAGE then SDS is very useful. When you want to do 2-DE, DIGE or other advanced gel-based

proteomic application where you cannot use SDS; zwitterionic and non-ionic detergents are used for such applications.

CHAPS, a zwitterionic detergent is one of the most commonly used detergents used in protein solubilisation. When your objective is to perform 2-DE experiment. It prevents non-specific aggregations through the hydrophobic interactions and it helps in sample solubilisation. Depending upon on your sample type, different types of detergents could be useful, in few cases ASB14 and sulfo-betaine detergents are better solubilizing agents. You also have options of using neutral detergents such as NP-40 although they are less commonly used. So one can not provide you a list of most effective solubilisation agents. No single zwitterionic or non-ionic detergent can completely solubilise all the proteins. So depending upon your sample type and if you know that your sample is rich in specific type of proteins you need to try different type of detergents.

Now let's talk about reductants. In the solubilisation reducing agents cleave the disulphide bonds, which are present between and within the protein chains and it prevents the disulphide bonds formation. Most commonly used reductants are dithiotreitol (DTT) and ß-mercaptoethanol. These are used for the reduction of disulphide bonds which are present in the proteins. Tributylphosphine (TBP) is one of the non-ionic reducing agents, another commonly used reducing agents when the aim is to increase the solubility of the proteins. Often TBP is used in the 2-DE experiments.

If your aim is to perform isoelectric focusing from your samples, the solubilising agent should include the carrier ampholytes or IPG buffers. These are added to the sample solutions prior to the IEF which we will talk in the next lectures where I describe you the different steps involved in performing a gel-based proteomics experiment. The ampholytes posses charge-to-charge interactions, they minimize the protein aggregation and enhance the protein solubility. Different buffers or bases are added which sometimes minimize the proteolysis and helps in the complete solubilisation of the proteins.

If your aim is to perform a 2-DE experiment the sample solution involves chaotropic denaturing agent such as 8 M Urea and 2 M Thiourea. Detergents such as CHAPS are most commonly used. It could be between 2 to 4 %. Commonly used reductant include 2-100 mM DTT or ß-mercaptoethanol and carrier ampholytes in the concentration of 0.5% Bio-Lyte.

The sample solution components ensure that the protein solubility is good during the extraction and protein separation. A typical sample solution for the gel-based 2-DE application includes 8 M Urea, 2 M Thiourea, 4% CHAPS, 2% IPG buffer, 40 mM DTT as well as few other components depending upon your sample type.

So, as I am giving you an overview of how to prepare a very good sample, let me give you examples of different challenges being imposed by different sample types.

If you are using the tissue culture grown cells you have to grow them in a medium which is rich in different components including salt and serum proteins. So one needs to get rid of those components if you want to perform a good sample preparation from the tissue culture cells. If you are interested in plant cells to extract the proteins; those are very hard tissues and there are many interfering substances present in there such as phenolics and other salts. Now you need to get rid of those interfering substances. Fungal cells such as yeast or other type of fungus; if you are interested in performing proteomic applications on these samples you need to break open very tough cells. So the proteolysis problem also occurs in these samples. The bacterial cells have high ratio of nucleic acids to proteins and cell lysis is also very tedious. Body fluids such as cerebrospinal fluid (CSF) are very dilute. So if you want to perform proteomic experiment on CSF you need to concentrate your samples. Body fluids such as serum is very rich in abundant proteins as well as salt. So you need to get rid of those abundant proteins such as serum albumin protein and remove the interfering salts.

So although the sample diversity is very much, samples are very complex I will still try to take 3 representative examples. 1) serum sample obtained from human. 2) the bacterial

sample from *Bacillus* species. 3) the plant leaf sample. By showing you the protein extraction and solubilisation methods for these 3 different types of samples I will try to give you the diversity and different types of methods being used to perform various types of proteomic applications.

So I will first start with the serum proteome analysis. But before we talk about how to perform serum proteome analysis it means all the proteins present in the human serum. First of all how to obtain these samples, how to store these samples, how to minimize various types of pre-analytical variations that is one of the very important considerations.

So before we talk about how to really process the sample let's talk about different types of clinical issues involved in these samples for the pre-analytical factors.

Proteomics aims for the simultaneous analyses of thousands of proteins of a given clinical sample; whether it's serum, urine, saliva, CSF or tissue. The impact of pre-analytical factors which occur prior to the point of actual sample analysis is very high for proteome-scale clinical studies. The pre-analytical factors could be due to biological variations or technical artifacts.

Your studies could be influenced due to intrinsic or extrinsic factors. The intrinsic influences include gender, age and ethnicity. The extrinsic influences include diet, medication, smoking and alcohol consumption etc. So when you are designing a clinical study you need to ensure that you have no bias due to the intrinsic factors. You should try to segregate the population with different groups of age and gender. In the discovery phase try to minimize these kinds of variations and try to perform your analysis with a narrow range of age group and different gender groups in the same ethnicity. But when you want to validate your samples then you need to extend your analysis to different age groups, ethnicity and gender. Try to avoid the extrinsic influences such as diet, smoking, alcohol, drug medication. These are going to alter the proteome and your discovery process will be influenced by these factors. So in the large cohort of patients

the number of non-disease related biological effects will influence the proteome changes induced due to the diseases. So the study design should be designed to match the age, gender and minimize the other influences without any bias. Often it is very useful to involve a statistician before you are designing these types of experiments and thinking about different pre-analytical factors before you perform your experiment often going to determine how successful your analysis is going to be.

Now we have looked at different types of biological pre-analytical factors. Then there are different types of technical artifacts, how to collect the samples, how to process the samples and how to store those. The sample collection mode; the gross effects of factors such as the patient posture and tourniquet application time are very important. Sample container types; when you are collecting samples such as serum or plasma, they exhibit differences as a result of coagulation, specifically the removal of fibrinogen.

For sample collection and handling procedure one has to pay attention. The collection and handling procedure of bio-fluids will affect the sensitivity, selectivity and reproducibility of the experiments. Collection tubes in which you are collecting your serum samples often is going to influence your analysis, if you are using different types of tube materials. The shedding components of the tubes or adsorption of serum to the tubes will in some way influence the proteome analysis. The CSF measurement of different proteins such as ß-amyloid and tau proteins differ when collected in tubes of different materials and the affects were lowest in the polystyrene tubes. So it is very important to understand that avoid different types of sample-tubes being used for the collection of the biological samples.

Sample storage is another crucial factor. Whether you are storing your clinical samples at -20°C or -80°C how quickly have you saved your clinical samples or how much delay was there before the sample was collected and it was stored. All of these small variations influence the sample analysis and the proteome analysis later on. So avoid multiple freeze-thawing of your samples, store the samples in small aliquots so that you don't have to freeze-thaw the whole sample together, avoid very long term storage or

storage at improper temperature, try to use fresh samples as long term stored sample as progressive degradation of serum proteins may occur.

So we will continue our discussion on how to perform the protein extraction and proteome analysis of serum and different types of biological samples in the next lecture. In summary today you have learnt about how to perform cell lysis, how to protect the proteins from proteolysis by adding various types of protease inhibitors, sample, fractionation methods, how to use protein extraction and solubilisation for effective protein solubilisation and now are discussing about specific examples. We will continue our discussion in the next lecture on serum proteome analysis followed by bacterial and plant protein and proteome analysis.