Lecture 2 Proteomics

Let's reiterate what proteomics is. It is the process of identifying, characterizing, and quantifying all of the proteins found in an organism under different conditions. The two hardest things are to identify and characterize all of the proteins under a certain condition of growth.

If you define that two proteins differ if they have any difference in their amino acid sequence as well as bound molecules, modifications, etc. then the proteome is quite large. One problem is that a single gene can generate multiple proteins through alternative splicing. In addition differential modification, interaction with different molecules, and location may also be considered as making different proteins.

One of the reason that proteomics is so hard is that there are so many different proteins and that they may be expressed at different times in different tissues. This is compounded by the large range in the quantity of proteins going over 10 orders of magnitude. Separating and distinguishing proteins with a single or multiple modifications from the unmodified protein may be difficult. Lastly identifying the interactions between all of the proteins can be very challenging. This is especially true for short lived interactions that may be very important for development.

Approaches

For separation there are limited possibilities. The main two methods are electrophoresis and chromatography. Chromatography may separate proteins easily but is limited by how many proteins it can separate at once and development of techniques can be time consuming. The other approach is to use electrophoresis to separate proteins. Frequently both chromatography and electrophoresis are coupled to help separate very complex mixtures.

Identification of proteins is mostly accomplished today by mass spectrometry. In this method a peptide map or a peptide sequence can be obtained and used to identify a protein. Often this requires a prior knowledge of the possible protein sequences and so often requires the genome to have been sequenced. The mass spectral data is then compared to the predicted fragments and sequences and then identified. Currently, at least two peptides from a single protein need to be identified to confirm the presence of the protein. We will discuss mass spectrometry more later.

Protein-protein interactions are studied by yeast two hybrid assay, protein arrays and pull down experiments. To determine which proteins are coupled to the bait protein often incorporates mass spectrometry. Again we will discuss these methods later.

Electrophoresis

Much of the work to identify proteins in proteomics is done through various electrophoresis processes. There are several electrophoretic methods used in protein studies but the most common for proteomics is SDS gel electrophoresis and isoelectric focusing. These are often combined to increase the resolution of the proteins and is called two dimensional gel electrophoresis. Lastly there is another method that is coming into use known as free flow electrophoresis. There are many other named methods for protein electrophoresis but they all follow these basic methods.

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis. (SDS-PAGE)

SDS-PAGE is the most common method for separating complex mixtures of proteins for analysis. The method separates proteins based on their size. This is accomplished by denaturing the sample an assuming that the protein will take a random orientation, which would be a ball that has a diameter dependent on the number of amino acids present. This then makes the size of the ball proportional to the molecular weight.

To denature proteins we use two compounds. The first is the detergent to denature the protein but SDS has other functions as well. SDS binds to proteins in a set ratio. SDS also adds a negative charge to all of the proteins making them all move in one direction. The addition of SDS in a set ratio gives all proteins roughly the same charge to mass ratio and so will tend to move the same speed in an electric field. The other compound used to denature the proteins is a reductant. We use either β-mercaptoethanol or dithiothreitol for this purpose. The reason we reduce the sample is to eliminate any disulfide bonds that may exist in the protein. These bonds will create loops that will provide structure to the proteins and not allow them to take a random shape. Without the random shape they will not move properly in the gel.

The gel is made out of acrylamide and bis acrylamide. These are polymerized using ammonium persulfate and Tetramethylethylenediamine (TEMED) to polymerize the gel. Together they form a matrix for the proteins to go through. The proteins then act like balls falling through a sieve and the smaller ones go faster and the larger ones more slowly.

The resolution of the gel depends on several factors. Resolution is defined as the ability to distinguish the sizes of two nearly same sized proteins. The size of the gel, the polyacrylamide and cross linker concentration, whether it is a single concentration or a gradient, and the stacking of the samples all affect the resolution. The larger the gel is the better the resolution, usually. The more concentrated the acrylamide the lower the molecular weight that the gel will resolve. The gels separate based on a log scale and so

this adds to the difficulty of using the gels. A gradient gel is one where the concentration of the acrylamide is higher at one end of the gel from the other.

Gel running times will depend on the size, thickness, buffers, salt concentration of samples, and voltage.

After the gel has been run the proteins need to be stained. There are several stains each with their advantages. Coomassie blue is the most common stain but is limited in its sensitivity. Colloidal coomassie blue gives a more sensitive stain. Several fluorescent dyes are available such as Syrpo organe (1D gels), Sypro red (2E gels), Deep Purple. The dyes can not be used on any type of gel. Sypro organe works well with 1D gels but sypro ruby or deep purple are best for 2 D gels. One problem with sypro ruby is that it contains a heavy metal and so has to be disposed of as a toxic waste which is cumbersome and expensive. On the other hand deep purple is relatively non toxic and can be dumped down the drain.

Isoelectric focusing (IEF)

Isoelectric focusing separates the proteins base on their net charge. These still use polyacrylamide as a matrix but the concentration is low enough to allow most proteins to move freely through the gel. Two types of systems exist. The first is to use tube gels filled with acrylamide and ampholytes which establish the pH gradient in response to the electric field. Tube gels are usually fairly short, 3-5 inches and have a narrow diameter 1 mM common, larger rare. The other system is one which uses ampholytes that are bound to the gel. This is called an emobilized pH gradient. In this pH gradient the gel has to be polymerized after the pH gradient is developed. The immobilized pH gradient has the advantage of no having the anode proteins drifting out of gel after long runs.

The pH gradient can be over a single pH unit or up to 10 pH units. Typical values are pH 5-10 or 3-11. Run times vary from 2-6 hours for tube gels to 6-24 hours for the pH strips. Run times will depend on the salt concentration of the sample, size of the strip, and pH range.

2D Gel Eletrophoresis

Essentially this is just running an IEF gel followed by separating the same sample on an SDS gel. The IEF gel is used as the loading gel for the SDS gel. When running 2D gel the cysteines are often modified to prevent the formation of disulfide bonds during the separation process. From a 2D gel some 700 different spot can be identified. This is not the same as 700 different proteins. Typically one protein can have up to 20 spots with each spot differing due to modification or proteolysis. From most 2D gels only 300-500 proteins can be analyzed due to repeat protein spots, overlapping spots, non-staining spots, etc. Membrane proteins do not separate well in the IEF system and so are not well resolved in the 2D gel. Many membrane proteins do not even enter the 2^{nd} dimension.

Also the basic proteins are difficult to resolve in IEF and some proteins are vary basic and do not even appear on the gels.

Differential Gel Electrophoresis

Differential gel electrophoresis allows the relative quantification of proteins within two samples. Each sample is labeled with either a Cy3 or Cy5 dye. Unlabeled proteins from both samples are mixed and then labeled with Cy2 dye to provide the standard to work from. All three samples are mixed together and then separated by 2D gel electrophoresis. The gel is then scanned for the three dyes and the image analyzed. The integrated intensity will be relative to the protein concentration. By using the mixed sample as a standard the relative concentrations of each protein can then be measured.

Free Flow Electrophoresis

This method is a continuous flow method of electrophoresis which allows the processing of large quantities of materials. It also has the ability to deal with relatively large particles such as organelles, membrane vesicles, and whole cells. The basis of the methods is the same as any electrophoresis method in that we are observing the separation of the sample by an electric field. The unique property is that there is no matrix. Instead we are using laminar flow and surface tension to restrict the diffusion of the samples. Since there is not matrix, recovery of the product is simple. Additionally the method can be used to keep the sample in a functional state.

This technique can be used for IEF in both denaturing and non-denaturing conditions. The system does not use ampholytes as in the other IEF systems but instead uses weak acids with various pKa values to establish the pH gradient. Organelles are separated by zonal electrophoresis also using buffers. And example of this is the isolation of plant leaf mitochondria (Plant J 52:583-394, 2007).

The final method Iso tachophoresis is basically using zones of pH and the electrophoretic mobility of the compounds within these zones. Both this method and zonal are non-equilibrium methods while IEF is an equilibrium method.