LECTURE-16

Gel-based electrophoresis

Two-dimensional difference gel electrophoresis (2D-DIGE)

TRANSCRIPT

Welcome to the proteomics course. In previous lectures we talked about 2-DE workflow, how to perform 2-DE step by step. We discussed the protein extraction, we discussed about isoelectric focusing, equilibration, second dimension separation by using SDS-PAGE, finally we talked about different types of staining methods which are available and how to scan and analyse those images. We then talked about different advanced methods which are creative ways of overcoming limitations of gel-based approaches. Now in the same theme today we will talk about another advanced method of gel-based proteomics, difference in gel electrophoresis or DIGE which used for overcoming various limitations of traditional 2-DE.

So the previous lectures focused on gel-based proteomics, 2-DE workflow and the new methods which are used for the proteomics applications.

In today's lecture we will talk about different challenges which are associated with 2-DE and how to overcome those challenges by using another method; 2-D flourescence difference gel electrophoresis (DIGE).

In the previous lecture when we talked about obtaining a good 2-DE gel image, then you may realise that at the end of that experiment all the 2-DE gel images look very good. So I show one 2-DE gel image which appears very good and this is something when everything goes very well. So if your protein separation and staining is fine then at the end you should be able to see good protein separation on the gel. This is a representative 2-DE gel image shown with the human serum sample.

Now, I am showing you another good gel obtained from bacterial sample *B. subtilis*. But often this is not the case. Many times there are different types of issues which could

give rise to very bad images. These issues could be because of the sample preparation method, reagents involved, IEF, various parameters and different types of staining methods being used. So let's talk about some of the issues step by step.

So there are various challenges associated with 2-DE.

Mainly the gel artefacts are major limitations, how to compare 2 gels, how to minimize the variations. Often during the electrophoretic run there are run-to-run variations, then you have different types of technical and biological variations. The image analysis is very challenging process, how to define the spot boundaries, how to extract the intensities and information from those spots. often there's user's bias in doing this type of data analysis. So there are major sources of variations in 2-DE.

I am showing you few problematic gels, how even the small mistakes or issues inherent to you biological samples can give rise to very bad 2-DE gels. For example, I am showing you sample preparation issues. In this gel, as you can see, the TCA-acetone precipitation method was followed but the washing was not performed good, so the TCA remained in the pellet and one can obtain the pattern shown in this gel.

Now if you are preparing the plant protein extract but the solubilization is not sufficient without precipitation, one can see this type of streaking and uneven gel pattern.

Now if you have samples with proteins in abundance such as crude serum. Such sample will show you interference in the focusing strip because of the presence of salt and other interefering components and then you can see these types of bad gels because of either salt or different types of abundant proteins present in the mixture.

Now the sample variation is one problem, often the chemical impurities or the chemical ingradients can also give rise to bad gels such as the chemical impurities shown in this gel, where the urea is impure and you can see the carbamylation trail as shown in this gel.

If TEMED is very old, often very small amount of TEMED is used in making the second dimension gel; SDS-PAGE. So people aften use the bottle for very long time and if the TEMED is very old often you can see bad pattern on the gel something similar to as shown in this image.

Now if your chemical such as Tris is of not very good quality, again that will show some artefacts in the gel.

Streaking is very often seen because of different types of problems associated whether it is coming from sample preparation or whether it is coming from different type of interfering components, often you have nucleic acid or carbohydrates still present in the mixture, the extraction method was not sufficient to eliminate all of those impurities. Many times salt and different types of interfering components are present which interfere in the IEF process and one can see the streaking patterns after staining of the gels.

So as shown in this gel, the presence of salt can be interfering and may result into vertical streaking as shown in the gel. If you are using narrow pH gradient strip such as 4-7 pH gradient strip then the vertical streaking can appear because all the proteins beyond 7 pI will stack together in that region and one can see the vertical streaking.

Now abundant proteins as I briefly talked last time some crude samples contain highly abundant proteins and salts. So abundant proteins are one of the major interfering components which can be easily seen on the gel. If you have serum sample there are many interfering components in the serum, including some abundant proteins such as serum albumin. Now in the plants one can expect Rubisco as a large protein present in the plant leaves. These abundant proteins mask many smaller proteins and create problem in doing the 2-DE because of IEF process and when you stain on the gel you can see different types of streaking as well as the whole region is masked because of the abundance of that protein. So in this gel it is shown that the human serum contains high abundant proteins which is visible in this area. Now there are various ways to overcome these limitations and as in the previous class protein extraction and sample

preparation we have talked about how to overcome these limitations, how to remove the high abundant proteins. So please refer to that lecture and one can overcome these limitations.

After protein extraction and IEF process is done one needs to equilibrate the strips. Equilibration process itself can create some problems. For example if you have performed first equilibration with DTT and forgot to do the second equilibration prior to the second dimension separation then you can see some pattern similar to shown here that has been equlibrated only with DTT.

Many times people think that by increasing the time of equilibration probably the gel quality might get better but that's not the case. In this gel it shown that by increasing the equilibration time to 20 min of DTT and 30 min of IAA the gel pattern looks very problematic and that is because of the equilibration excess.

Now, IEF process is very crucial. Many times during the IEF run itself one can monitor the problems if you have a software which shows how the run is progressing. Often if your volatge setting is not correct it is insufficient for the complete focusing, it may result in under-focusing as shown in this image.

If you have too much of voltage then it may result into over-focusing. So optimizing the focusing protocol for appropriate duration and volt-hour is very crucial.

So there are various sources of variations which can be introduced from the biological as well as technical means. There are different biological variations on which there is no control. If you are doing clinical studies even if you have done proper age and gender match still there will be some variations unique to individuals. So those biological variations are very tough to eliminate. But then there are lot of technical variations which can be improved during the sample preparation, IEF settings, equilibration steps. So all of these steps may result into large variations. So in 2-DE when you are separating your control and treatment gels on two different gels and you have so many variations

coming from both biological and technical variants then overall the analysis and the confidence in the data becomes very questionable.

To overcome those limitations the new approach, the advanced electrophoresis method two-dimensional difference gel electrophoresis or 2D-DIGE has emerged.

The 2D-DIGE method was basically aimed to overcome the gel-to-gel variability, which was observed in 2-DE and it allowed the multiplexing of the samples. This method was reported by Unlu *et al.* in 1997 and this method became very user-friendly and thousands of laboratories worldwide are using thistechnology to overcome the gel-to-gel variations encountered by traditional 2-DE. So the DIGE method involves protein labelling and separation techniques. The optical detection of the proteins is done by using the flourescent tags. This method performs linear detection in the wide range of the protein abundance. So overall one expect very sensitive method for giving rise to very large dynamic range of the protein detection by using flourescent tags. This method gives multiplexing and separating proteins on one gel, so it is expected that it will avoid lot of difficulties, which are encountered in the traditional 2-DE gels.

So let's talk about basic principle involved in DIGE technology. So we have to discuss how to do protein labelling then what is the need for doing the internal standards. So let's talk ste by step. As we discussed, this is commonly used protein labelling reagent for DIGE. N-hydroxysuccinimidyl or NHS ester derivatives of cyanine dyes Cy2, Cy3 and Cy5 are used in this method. The Cy3 and Cy5 are used to label the controls and the treatments. One should always do the dye swapping so that there is no dye bias. So if one sample you are adding Cy3 in control and Cy5 in treatment and in other sample you should do the vice versa. So Cy2 remains consistant for labelling the internal standard. Now, for Cy3 and Cy5, these dyes the size and charge are matched so that the labelled samples can co-migrate within the gel.

Now, let's talk about labelling methods in DIGE technology.

So when we were talking about 2-DE workflow in the last class we have looked at different types of staining methods currently available. Such as Coomassie, Silver staining, SYPRO Ruby and in that context I briefly described about Cyanine dyes. So now in DIGE technology those Cyanine dyes are used. There are 3 different Cy dyes as I mentioned, Cy3, Cy5 and Cy2. The excitation and emission wavelength of these three are defined. The Cy3, Cy5 excitation wavelengths are at 540 nm, 620 and 480 nm, respectively. Similarly the emission wavelengths of Cy3, Cy5 and Cy2 are at 590 nm, 680 nm and 530 nm, respectively. So these dyes are photostable, they are pH insensitive and spectrally distinct. By using these dyes the DIGE method can be used.

Now let's talk about different types of labelling methods. In DIGE there are two types of labelling methods which are possible. The most commonly used is lysine labelling or cysteine labelling. Lysine labelling method is also known as minimum labelling method where the charge and the size of dyes are matched. There are no multiple labels and it is known as minimum labelling because it is labelling only 3% of all the proteins. I will give you the reason why that is we are aiming for very small amount of proteome to be labelled and that will be covered in the next slide.

The saturation labelling is done by using the cysteine tagging. So amino acid cysteine is targeted to label. This is charge neutral and size matched, multiple labels are used, it is very sensitive and used only for Cy3 and Cy5 labelling. So saturation labelling is not so widely usedbut this is another method one can use for labelling the proteins and it can be used in DIGE methodology.

So let's talk about the popular method for targeting lysine. So why to label only lysine amino acid. Lysine is prefered amino acid as I have showed you in this graph here the distribution of amino acids across the proteins and their percentage distribution. So lysine is prefered amino acid to label and it's frequency is greater than 5% so statistically almost every protein has atleast one lysine residue. There is a good probabilty that your protein will have lysine residue and it will get labelled if you are

targeting the lysine. But if its not the case one can always try the saturation labelling of cysteines.

Now let's talk in more detail about the minimal labelling or lysine labelling. Cy dyes react with the primary amino group of the target proteins and the N-terminal alpha-amino and lysine E-amino groups by the process of nucleophilic substitution. So all the available lysine labelling would create a very hydrophobic environment of the proteins. That was the case observed by the original paper derived by Unlu *et al.* So avoid that hydrophobic protein atmosphere, it is recommended to the minimal labelling of only 3% of the whole proteome so that overall it is not very hydrophobic.

Now in this slide I will describe you the labelling chemistry. So as you can see in the structure here, the stucture of CyDye Flour comprises of a positive charged CyDye Flourophore, a linker and NHS ester reactive group. So at pH 8.5 the NHS ester couples with amino group of lysine in the protein and it forms a covalent bond.

The amino group of lysine residues is positively charged at neutral or acidic isoelectric point. Dye is also positively charged so that the net pI will be unchanged. Therefore Cy3 and Cy5 are going to alter any mass as well as charge overall and they are very well spectrally resolved. So people prefer doing lysine labelling. It is not going to alter overall pI values and mass because both Cy3 and Cy5 will carry the same mass of these labels and there is a good probabilty that your protein will have lysine residues and it will get labelled with the Cyanine dyes.

So how to prepare the sample to perform the DIGE experiment. First of all, you need to reconstitute your dye in dimethylformamide (DMF). The lysis buffer in which your protein sample is dissolved consists of 30 mM Tris, 7 M Urea, 2 M Thiourea, 4% CHAPS, pH 8.5. Please make sure that there is no IPG buffer or carrier ampholyte and there is no reductant added in this lysis buffer. Since you want to perform minimal labelling, you need to add only 400 pmoles of Cy dyes in 50 µg of protein, and during that process the reaction should be performed on ice-water for 30 min. If you go and recollect our previous discussion to visualize the gel with coomassie stain you need more protein.

Even from silver and other stains, you need good amount of protein. But these flourescent dyes are very sensitive. All you are looking for is very small amount of protein. 50 µg of protein is sufficient for performing this experiment. Now, if this reaction of 30 min is done after that the reaction should be quenched with excess of primary amine, the free lysine; 1 µL of 10 mM lysine. So then the reaction will stop and you can proceed with the next step.

So while you are doing this reconstitution of dyes and labelling your protein samples with Cy3 and Cy5 now let's think about Cy2 and the internal standard.

I am taking a very simple case here, where you have 4 healthy controls and 4 patients suffering from a disease. To represent all the proteins present in each of these samples in healthy individuals and diseased patients first you need to make an internal pool which is the representative of all of these 8 samples. You can take 25 µL from each these 8 samples and make to 200 µL of one pooled sample which is your internal pooled standard. So this pooled standard sample can be performed by combining aliquotes of equal amounts of proteins from each sample in the given experiment. Now this pooled sample is labelled with the Cy2 dye so that it can be used as the internal standard.

This internal standard has many advantages. And this is actually one of major success that how limitations of 2-DE have been overcome. Since every protein is present in the internal pool population so that is going to be appearing on your gel. So there's a reference point for each protein already present on your gel whether it's from patient or controls. So gel-to-gel variations are reduced and gel-to-gel matching and comparison is much efficient and easier. Now, this internal standard is creating a reference point for every protein on each gel in the experiment. Later on when a software performs some data analysis this results into accurate spot statistics and more meaningful biological can be interpreted from this data analysis.

So now we have talked about DIGE principle, how to label the sample with Cy dyes, why there is a need for making an internal standard and how to label the internal standard with Cy2 dyes. So let's go to step by step procedure of DIGE method.

So very first thing is you have extracted your protein samples and you need to make your internal standards. Now your control sample, treatment sample and internal standard should be labelled with different Cy dyes. Internal standard is always labelled with Cy2 dye but your control and treatment samples should be dye swapped. Once the labelling is done then the samples should be mixed and all the three labelled samples should be mixed in one combined tube. After that process is done you can take that sample and do the rehydration. We have talked about the rehydration step and the need for it in the previous lecture.

Now this rehydrated strip can be focused in an isoelectric focusing instrument. It is a good idea to avoid any light exposure during rehydration, dye reconstitution process as well as during the IEF run. Once IEF is done equilibration has to be done; first and second equilibrations followed by SDS-PAGE. Overall step as you can see is exactly the same as 2-DE. Only variation here is that all your control and treatment samples have been mixed together and now they are separated on one IPG strip. Now this strip is placed on one SDS-PAGE gel and proteins are separated on one itself whether it's coming from the control or treatments.

Now this one gel can be scanned using flourescent scanner such as one shown here is DIGE imager and Cy3, Cy2 and Cy5 patterns can be obtained. So this slide gives you an overview of the whole DIGE process where one needs to mix Cy2, Cy3 and Cy5 labelled samples in one tube and resolve those on one gel on SDS-PAGE and scan that gel with different excitation and emission wavelengths to obtain 3 gel images derived from the same gel. Here few representative gels are shown obtained from the DIGE experiment for Cy3, Cy5 and Cy2.

The software also generates the combined image obtained from mixing all 3 dyes and the overlapped image, which shows you the abundance of certain proteins. In

overlapped image the red and green are indicative of over- and under-expression, respectively and the white or yellow colour is the representing that there is no change. Although the DIGE image analysis is a software driven process, it is much more automated compared to 2-DE analysis but it requires lot of tools, which can be used for interpretation.

So from your one gel you have obtained 3 images and those can be further analysed by using DIA in the co-detection of 3 images. So I am talking about a software DeCyder which takes these 3 images in to account and makes one image pool by the process of DIA. Now these multiple DIA can be matched in the BVA. So these BVAs can then further be used for obtaining statistical information about these protein changes, whether those are statistically significant or not.

I am showing a representative gel here, where 2 spots are highlighted from disease and healthy controls. So one can look for the abundance of those proteins in 3D views and as shown here, if these proteins are showing higher abundance those are visible in different treatments, whereas there is no protein change happening in the healthy controls.

Now one need to look at the same trend of these 3D views across all the patients. So for example, I have selected here a group of 6 patients and 6 healthy controls. Now this particular protein is showing up-regulation in all the 6 patients and as you can see here and there was no change in the protein level in all the healthy individuals. So by looking at the data carefully patient-wise and the reproducibility pattern, one can be convinced that this protein has some biological relevance and then need to be further identified and characterized.

So after doing the BVA analysis further statistical analysis of different classes can be performed by using EDA or extended data analysis. Now EDA has different types of tools in the software such as principle component analysis (PCA). The principle component is the linear combination of optimally weighted observed variables. Now PCA used to test whether in your protein samples in different patients and controls the

protein expression is uniform across all the multiple samples derived from the same experimental group. So if your results are consistant across the whole population the same protein pattern is consistant in all the controls and in all the patient samples then the spot map from the same experimental group will be located in the same region.

To clarify this point, I will show you this image here. If you have one control and two treatments and by looking at the significant proteins the software can use this PCA tool and give rise to the pattern of healthy and different types of treatment patients. If the grouping is together it means it is highly significant. As you can see here the red dots are showing the controls, while the blue and yellow are showing two different types of patient groups. These three croups are clustered quiet closely. There will be some heterogeneity because of the different clinical sample and patient biological variations involved. But overall it shows that the data is statistically quiet statistically significant in the experimental groups. Now using this software one can obtain further information about different type of discriminants and classes. So the discriminant analysis is used to classify the uniform data for the known classes. One can initially take all the proteins which are significant and try to group those based on the known classes.

The hierarchical clustering is another method which is giving very useful information. For example, in this data we try to look at one control and two different types of treatment groups together. Now on the extreme right side all the green shades are from healthy controls and there are two more population from the patient group which are showing proteins in the red region, which means it has higher protein abundance. So this clustering is showing that the control belongs to one population and these two patient groups derived from the same population but within that population there are certain variations. These items whether proteins, spot maps and experimental groups can be grouped together in the heirarchical way by using this clustering. So I will describe you an animation on how to perform DIGE experiment and you will be able to recall different concepts which we have discussed so far.

Difference in gel electrophoresis:
The pooled internal standard for DIGE is prepared by mixing equal amounts of all samples that are being run in a given experiment. The internal standard prevents problems of gel-to-gel variations.

Dye labeling
Each protein sample as well as the internal standard is labeled with a differently fluorescing cyanine dye, which allows all protein samples to be simultaneously run on a single gel. The dye binds covalently to the e-amino group of lysine residues in proteins.

Mixing and 2DE
The labeled protein samples are mixed and run on a single 2-DE gel. Separation takes place on the basis of isoelectric points of the proteins in first dimension and based on molecular weight of the proteins in the second dimension with the smaller proteins migrating further along the gel.

Gel viewing
The gel containing all the protein samples can be viewed by illuminating it alternately with excitation wavelengths corresponding to the various Cyanine dyes.

Superimposed 2D-DIGE image
Now the superimposed DIGE gel depicting all protein spots of multiple samples can be viewed. Information on molecular weight and pI of proteins can be obtained from the gel. But most importantly by looking at the gel pattern if you are doing the abundance proteomics comparing the expression of two samples different protein expression then that pattern can be immediately observed on the gel with the red and the green patterns. If there is no change it will appear as white.

So after knowing the concepts and after looking at the animation now you should be very confident in overall concepts related to the DIGE experiments. So let me give you a very simple problem. If you want to perform some gel-based proteomics experiment from 250 patients given from the clinical trial, which of the two method will be useful?

Which of these two would be the better technique to separate serum protein samples obtained from 250 patients in a clinical trial? If your answer is 2-DE, then that is not correct. You need some method, which can minimize variations across large population. So you need DIGE.

So let me describe you the above explanation in an animation.

2-DE, although a very useful technique, may not be the best option in this case for analyzing serum proteins from a large number of patients as it would involve running several individual gels which would be a time consuming process. Also variations across the gels would make comparison of results a problem.

Whereas, DIGE is an extremely valuable tool for analysis of a large number of samples simultaneously without having to encounter the problem of gel-to-gel variation. The control and test samples can be differentially labeled using the cyanine dyes and run on a single gel. You can scan the gels by using a DIGE scanner.

Now you have a good clarity that what are the disadvantages of 2-DE and advantages of DIGE. So let's summarize our concepts of 2-DE and DIGE. In 2-DE there are gel-to-gel variations, which can be overcome by DIGE by separating the proteins on one gel. In 2-DE there are large sample variations, whereas in DIGE these variations can be reduced by making an internal standard, which is going to give the reference points for all the proteins. Now in the 2-DE often in the data analysis there is user-bas, how to define the spot boundaries and obtain the statistical information. Whereas in DIGE there is co-detection method involved by the software analysis, which eliminates lot of user-bias of performing this analysis. So overall if you compare 2-DE and DIGE there is more merit to DIGE technique. But not to say that before you want to move to 2D-DIGE approaches first you need to ensure that sample is good, your technique is good and overall you can obtain a very good gel. So first you need to perform a traditional 2-DE, stain the gel with coomassie or silver stain to see if every thing is looking good and every thing is working fine to obtain a very good image. Then you can use the same protein extract to perform the DIGE experiment that will be very cost effective.

In summary in today's lecture we talked about different challenges, which are associated with 2-DE. Then we looked at DIGE workflow, different concepts involved as well as how to do the labeling of the samples, how to make internal standards and then we have seen the steps in animation. We finally compared 2-DE method and DIGE method and I gave an interactivity option to perform an experiment by using 2-DE or DIGE and you gave the right answer of doing the experiment with DIGE method. So overall from this whole experiment, you should be able to conclude that DIGE is better way of separating proteins and it can still provide you very quantitative information in the gel-based proteomics.

I would like to acknowledge Dr. Srinivas and Lalith Kishore for providing some useful information related to 2D-DIGE and we will continue our discussion on 2D-DIGE and data analysis in the next class.