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 33 : Label free methods of protein quantification

Namaskar. Hi students welcome back to module 7 of our lecture series on Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis. We are discussing advanced techniques regarding proteomics and today's topic is label free methods of protein quantification. And this is the ongoing series within the module 7 which is the third part, today is the third part of the series in earlier parts we have discussed all these techniques of label based techniques of protein quantification and today we are discussing the label free technique all right. So, what is a label free technique? From the name itself the thing is quite obvious right. So, label free quantization technique again these are all techniques that are used in mass spectrometry all right mass spectrometry based technique of detection.

So, label free technique is a technique of protein quantification in mass spectrometry to accurately determine the relative protein abundance all right in multiple biological sample without the need of labeling agents very important two keywords relative abundance of protein all right and there is no requirement of any labeling agent all right. So, in contrast to other label based techniques which we discussed in two lectures before this there is no need of any isotope containing compound this does not rely on any stable isotope containing compound to chemically bind and label the protein all right there is no need of that will be using a separate technology itself. So, that we can have idea about the protein abundance relative to one another how much the protein one protein has been expressed more how it has changed etcetera etcetera all right. So, there is no requirement of any isotope based labeling compounds and mind it isotope based labeling compounds are used to chemically bind and label the proteins that we discussed in the preceding two lectures lectures ok.

Now, the thing is there is no limit to the number of samples all right there are limitation we will be discussing that, but as per the technique there is no limit on the number of sample in label free quantification all right. Therefore, in principle it is actually applicable to any kind of sample including material that are not favorable to chemically label all right. For example, either chemical or metabolically labeling is possible for

example, if we get many clinical samples there might be a situation where lab use of label based methods are not feasible in that case or it is expensive all right. So, in that case we can we have got the liberty to choose or to determine the relative expression of the proteins with the help of these techniques that is label free quantification technique. So, it is ideal for large sample analysis in clinical screening when there wherever there is a huge number of sample see the highest amount of label that is possible that we saw in S I L S E as well as in I track and in I cat as well as and you can see all the in tandem mass tags all the techniques that we discussed the maximum was 16 plex up to 16 all right, but in case of label free if the sample is more in number we can still do it in one go with help of these label free techniques .

But the thing is if we are using we are confined to use of only label based methods then we need to divide the sample in small batches then then we need to do. So, label free has got much higher throughput compared to label based techniques whenever there are many many more number of samples all right. So, whenever we are doing a clinical screening there can be hundreds to thousands of samples all right and then label free technique techniques of quantification related quantification takes an edge leads compared to other label based methods. It is also very much relevant for biomarker discovery experiments over there also there are huge number of samples and conditions that we need to analyze in one go ok. So, the basic work flow of label free quantification is almost the same initially we have got our proteins of interest those are big we digest them with the help of proteolytic enzymes like prypsin proteolytic enzyme all right.

So, then they become peptides all right. So, those peptides are then separated or fractionated using liquid chromatography high performance liquid chromatography and tandem mass spectrometry all right. And thereafter this part is actually the same all right in general it is in this part the quantification technique that varies will be discussing what are the principles what are the methods varies. But this is in general the schematic of how a label free quantification happens one method has been shown here that is spectral count we will be discussing that in detail. So, what happens the first step.

So, if we actually break down the schematic into a answer theoretical answer that you need to write step by step in a flow diagram the very first sample is very first step is sample preparation which deals with protein extraction the reduction of protein the alkylation of protein and the digestion of protein basically same right. So, all these steps are needed in sample preparation just like label based methods. Next after preparation separation of sample how are they done preferred method again HPLC. So, separation by liquid chromatography analysis I mean analysis by tandem mass spectrometry and finally, all the data are compared compiled together and then we apply the quantification technique and what what data do we analyze we can analyze peptide we can identify the protein we can quantify the protein and ultimately statistic analysis done. So, data

analysis actually comprises of these things we need to identify and compare the protein we can we need to state what is the change relative change of protein with respect to one sample another.

So, protein quantification and finally, whether it is statistically significant or not. So, data analysis comprises of all these three ok. So, how do we do it in label free. So, before knowing what are the methods you should know that in label free quantity proteomic each sample is separately prepared and then subjected to individual LC-MS-MS or LC you can be subjected by double liquid chromatography for the better separation suppose there is a complex mixture it is separated into less complex mixture and then each subtraction can again be subjected to liquid chromatography. So, tandem liquid chromatography and tandem mass spectrometry there is a huge resolution by which you can actually fragment each and every segment.

There after when it is subjected to tandem mass spectrometry on the depending on the peak you know m the mass spectrometry gives the final output in form of spectral peaks ok and the I am that actually is the information about the charge to mass ratio of the ions we are broke breaking down the samples into ionic form gaseous form and a mass spectrometer detects that. So, quantification is based on the comparison of peak intensity of the same peptide alright or the spectral count of the same protein. So, these are the two methods ok. So, again let us discuss them in details label free quantification techniques just as discussed in the last slide are divided into two distinct groups that are used for data extraction. Number one that is known as spectral counting method what does it do remember we are comparing relative abundance or relative change of a protein alright.

So, some same protein might be expressed in a small amount or present in a small amount in one sample and it is present in a different amount in other sample we are comparing the intensity of those two proteins to get a relative change of abundance alright. So, the quantification can be inferred by counting the number of peptides or spectra assigned to a given protein a protein might have a characteristic spectra it is already available in our knowledge database. So, we can count the number of peptides or spectra assigned to a given protein therefore, these these type of method I mean this principle that which uses counting of the number of spectra of a specific protein is known as spectral counting method alright. It is one of the key method that is used in label free quantification technique. And another the second method is known as iron intensity method what happens in iron intensity method when we are using HPLC-MS the quantity values can be measured through extraction of the area of the precursor chromatography peaks precursor iron chromatography peak alright.

So, the area under curve or AUC or the first MS run it gives a signal it gives an area in

the second they are fragmented and further broken down into characteristic spectra. So, that intensity of the specific iron that is given after we have separated it by liquid chromatography and in the first run that area under curve which give that gives us the intensity of the iron. So, that will vary when two separate when the same protein is present in different amount in two sample after they are separated similarly in the first run the area under curve will vary of the same protein ok. So, these are the two method one is spectral counting method and one is the iron intensity method of label free quantification. Now, let us see let us discuss some points key points about spectral counting method.

So, relative protein quantification is achieved by comparing the number of identified tandem mass spectra from the same protein which has been shown to directly correlate with the protein abundance. So, this number of spectra of the protein has found to be directly related to the amount of protein therefore, this method has been adopted. So, days after days group of scientists had observed something right and then they establish that we actually do not need a label just by simply counting the number of spectra we can have a some information about the relative abundance of the protein all right. Next it is possible because of an in so, why does it happen all right it is because the increased amount of protein that is present in the sample when we break down the sample the amount of peptide the proteolytic product that is produced is also more right. So, sample one which has got a small number of protein will have a less number of broken down products or a peptide compared to another sample which has got a bigger amount of protein.

Hence if we digest it in the very first step the amount of product that will go into the HPLC and the MS LC MS chain will be more and that will lead to a higher number of spectral count and that is why it is easier to compare the two sample just by counting the number of spectra. Next so, what happens so, this proteolytic peptide the increase amount of digest usually to increase in the protein sequence coverage all right the number of identified unique peptides and the number of total identified tandem mass spectra for each protein all right. So, so all this I mean it is very easy to visualize more amount of protein we are breaking it and that is leading to a more number of spectra all right. An added advantage being since, the amount of digest I mean the amount of product proteolytic product that is happening the first digest is increasing the one benefit is more sequence coverage is possible with the help of this spectral count method, but mind it we are targeting the same protein all right. So, definitely as just like any other label free technology there are no specialized tools all right.

We do not need any other third party software, we do not need any other interpretation method all right specifically for spectral counting approach it is very easy the built in software itself can do ok. So, it is very inexpensive however, the one important thing is

we still need to normalize the data there might be some error all right this ideally the same amount of sample should be put ok. So, we are considering the same in a same amount of sample some proportion of a protein is more all right therefore, it is digested product therefore, it is number of spectra are more, but if the sample application is erroneous to start with then it will give a false information right. So, therefore, it is very important to normalize each sample for proper statistical analysis all otherwise this method becomes unreliable there are methods of normalization and that is to be taken care of. So, normalization and statistical analysis is very important for counting spectral data set and are very necessary for accurate and reliable detection of protein in protein changes in complex mixture.

So, as and when the number of protein increases we need normalization. For example, there are housekeeping genes that can be seen there are certain specific protein whose concentration per amount of protein remains fixed. So, those spectra can be targeted in both the sample and that can be monitored to normalize both of the samples normalize means the same amount of a housekeeping gene or a target protein that should not have gone a change is targeted. Suppose sample one we apply and theoretically giving an example and that that is not true only for the spectral counting method is true for any proteomic experiment even for western blotting. We apply 5 gram of a sample I am not going into the millimole or the concentration all right.

Suppose the amount of protein amount of protein that is in one sample is more compared to the amount of protein that is in another sample all right. So, what happens if we are studying a specific sub unit of the protein that we need to study in both the groups. So, if the amount of sample is more to start with them that will give a false information all right. So, always need to make sure both the samples are comparable in the baseline. Hence whenever we study or target any specific protein of interest we always normalize it by dividing with the total amount of protein.

Hence we always say this this this per milligram per microgram of protein. Similarly in this case if the application of sample is not controlled we still should have a method. So, that we can detect and normalize the samples so that both of them are comparable at the baseline. So, that the spectral counting method can give us reliable information instead of just getting confused whether the sample application is more or whether if the protein has actually expressed more ok. So, next we come to ion intensity method.

So, what happens very important ion intensity method we are using ESI. So, wherever electro spray ionization method is used for ion fragmentation their ion intensity method is much more used if you are using any label free quantitative technology in protein quantification right. So, signal intensity from ESI this correlates with the ion concentration that is already known. So, what happens the height or area of a peak of a

particular M by Z ratio from a mass spectrum this reflects the number of ions all right. So, what happens in electro spray ionization if you remember.

So, we are concentrating the ions in a jet. So, that the droplet the sample. So, that the liquid breaks up a Rayleigh point and then fragments into ions all right. If I strongly recommend you just go back to the class of electro spray ionization that we discussed in mass spectrometry then it will be much easier for you to understand and visualize the whole thing. So, that height or area of a peak of a particular M by Z ratio of an ion that reflects the number of ion in a mass spectrometer at any given time and this is typically known as ion abundance.

This thing or this index is actually used for ion intensity method in label free quantification. However, this one thing we need to know that this ion abundance cannot be directly used to infer the absolute protein of peptide con concentration all right. The main reason being ionization efficiency. So, not all sample not all sample will have the same rate of being broken down into ion cloud when we are using the electro spray ionization technique. Therefore, it is very important that we should compare again just like it spectral counting method we should compare the ratio of identical proteins because generally identical proteins have got a same propensity or same potential to be ionized.

If you are comparing different proteins one protein may be ionized efficiency means the entire protein is ionized then we get a much higher ion abundance and the second protein does not ionize much therefore, we will get much less ion abundance and then we cannot compare it. However, if you are come again if you are comparing an identical protein in two different samples. So, we are comparing the relative change of a same protein in two samples then we again can use this ion intensity method if you are using electro spray ionization in the mass spectrometer all right. So, do different ionization efficiency of each peptide comparing the ratio of ion abundance between identical peptides are often obtained in different experiment runs can be used to estimate a differential expression. I hope I am clear we cannot compare to different two different proteins because they might have different ionization efficiency.

So, everything needs to be same own since everything remains same and comparable normalized that is the concept then only one change will reflect or change in one parameter will reflect the change in the amount of protein that is expressed in the sample. So, this is basically the ion intensity method mind it again the normalization and everything should be taken care into account that I discussed in so many so much details in the spectral counting method. So, where is it useful? It is useful when applied to high precision mass spectrometer mass spectrometers for detection method for example, in ESI ESI is the mode of ionization and what is the detector we have discussed in detail

the multiple types of detectors one of them is time of flight detector of TOF one of them is Fourier transform ion cyclotron resistance FTICR there are orbitrap mass analyzers we have discussed in detail the time of flight analyzers because that often comes as a short note in theoretical exams, but other than that FTICR and orbitrap mass analyzers also be very high precision result and this ion intensity method can be utilized when we are using ESI mode of ionization come with these type of detectors in a mass spectrometer ok. So, this is all about the label two principles or two methods of label free quantification methods. So, what are its advantages? The advantages are mainly I mean we have we will discuss three major advantages there can be many number one the variability that chemical labeling or tagging may induce is eliminated.

So, one protein. So, if we are using chemical tags two proteins might be tagged in a variable way might these are all experimental variation that can happen alright, but here we are not counting those when we are normalizing we have taken into account that the same protein is undergoing the technical biophysical technique using the same principle and only the change in expression of the protein will give different result. So, experimental variation due to chemical labeling or tagging is eliminated alright. Next one since we are not using any isotope learning labels which are very costly metabolic tags we have to use cell culture very costly compared to I mean here we are not using anything we are just doing the technique without any labels and therefore, it is inexpensive compared to all other chemical or metabolic tag based or label based techniques of protein quantification. And since the number of steps are significantly reduced you saw in label based techniques whether it is tandem mass tag or isotope based tag or isobaric tag we need to tag each and every protein we need to mix those protein again those need to be separated using LCM. So, there are cut down of multiple steps of tagging and labeling and therefore, and the when you are analyzing the same samples by two different method the label free quantification is much easier and much less time consuming compared to the label free I mean label based methods.

So, label free method are much more time consuming compared to label based methods ok, but there are definitely certain disadvantages that we need to be aware of number one the amount I mean the relative ease of measurement alright is easier because tags give us definite signals alright, but here the we are solely relying on spectral count or ion intensity or various of the parameters of the peptide itself which has got their own limitations alright. I mean the phenomena of spectral counting or ion intensity itself is not full proof there might be some error that creeps in mind it these are all developing techniques all advanced techniques high throughput techniques are still in the refining phase. So, these advantages gives us the incentive to go into new much more research and development. So, that we can come up with much better technology. So, mind it with the label free quantification has got some inherent limitation because they rely on the physical process of the mass spectrometry like spectral counting ion intensity relative

abundance.

So, they have got their own limitations, but when you are using label based techniques the visualization is much easier compared to label free techniques. Again the thing is when with other issues associated with label free quantification are the sequence coverage the amount the sequence coverage is much less compared to label based techniques this is one important thing that we need to know because the I told you label free techniques each and every sample should be digested fractionated and then subject to the analysis in mass spectrometer alright. So, complex fractionation prior to the analysis is important as you saw there was one step where you we are using LCLC we are using double HPLC. So, that each and every fraction is completely separated prior to their entry of the sample in a mass spectrometer alright. So, if there are so much complex proteins it is better to use label based methods.

So, so label free methods have got their own limitation label based methods have got their own limitation label free has got some advantage labeling isotope label methods have got their advantage. So, basically all these issues are limitations need to be carefully considered in a setup and the decision should be made which is optimal before designing any specific experiment. This is a general statement this is not the disadvantage this is the general statement because we have discussed what are the advantages and disadvantages of both the techniques. So, considering what is the amount of sample what is our resources how complex is the sample what is the what is the time limit we do have in our hands. So, all these things do dictate what is the amount of what is the mode of protein quantification techniques that we can use alright.

So, what are its applications the applications are again a manifold for example, in case of identification the expression profiles of various biological of a same protein across various biological samples might be easily measured since we are talking the same protein different samples in different biological processes will give a same spectra will have same ion intensity right. Where we cannot use label based techniques we discussed for example, in large clinical screening in biomarker discovery for example, certain region cancer biomarker label based techniques are not feasible enough. So, its very label free technologies becomes the method of choice it is important for diagnosing also monitoring certain biological processes or proteomes alright. So, at different stages we can put the sample we can analyze we can halt the progress at different stage or we can take sample which are at different stages of their biological development we can process those samples and use label free techniques to target the same protein and have some idea about them alright. Next studying protein-protein interaction network this is also very much helpful when you are using label free quantification technology ok.

Next this is the entire proteomic workflow. So, you have see here what are now we

initially started with 3 techniques. So, now, the entire picture is with us. So, first protein is purified with the help of chromatography based techniques we have discussed in detail thereafter traditional proteomic techniques can be used rather we have discussed high throughput proteomic techniques. So, ELISA and western blotting are traditional techniques that are good that are mostly used in clinical settings and in centers for easy convenience, but we have also discussed protein microarray in detail which are used for which can be used for detecting multiple samples at once in a very small scenario alright in very small setup. Next characterization of protein results of gel based approaches and we have discussed conventional 2D we have discussed differential gel electrophoresis right.

Mass spectrometry we have discussed in detail regarding the characterization. Next is sequence analysis of protein. So, see regarding sequence analysis you see the one that is mentioned here Edman sequencing Edman degradation technique it it was originally devised and it deals with detection of the N terminal using an reagent phenyl isothiocyanate ok, but this is of historical importance now yes it is used, but it has been largely replaced by mass spectrometry ok. Why this has largely been replaced number one it takes time it needs the I mean more amount of samples yes Edman degradation technique or Edman sequencing is still preferable in small scale lab when amount of peptide sequence is less, but mass spectrometry is much easier we get a characteristic signal of each peptide and they can be easily much to know the sequences. And in fact, in today's date next generation sequencing technology specially the RNA sequencing and the ribosome sequencing gives us an indirect idea about protein sequencing.

Again those are very high throughput technology, but again those are also have some limitation because Edman degradation technique will exactly tell you what are the sequences because this is based on chemical reaction. Mass spectrometry yes with the help of spectral signals with the help of bioinformatic tools we can have some idea, but if we need to maintain a gold standard just like Sanger sequencing over here Edman sequencing can be the best alright. Again next generation sequencing also has got some limitation because they are indirectly telling us the information of the protein sequence by say observing the mRNA sequence which can may not reflect one is to one situation. So, these are the few methods that we discuss about sequence analysis number one is Edman number two mass spectrometry number three next generation sequencing. So, it it is just like gene sequencing right, but we are applying those concept into proteomics, but if we need to mention one method of choice in today's date for sequence analysis the answer will be mass spectrometry alright.

So, we have already discussed sequencing in details because mass spectrometry have covered all the aspects how it is done. Next quantification so, we have discussed quantification in these three classes in details you can see the techniques ICAT, SILAC,

ITRAQ etcetera tandem mass types we have discussed and next the structural analysis that is done by X-ray crystallography NMR spectroscopy those are the methods that we will not be covering in our lecture series, but if we need to know about the structure of a new protein these techniques are very much helpful. So, we come to an end of our mini series within the module 7 that is protein quantification techniques will be touching data acquisition method, shotgun proteomics and all other things that I left behind we also be discussing bio informatics tools of proteomic analysis in the next two remaining lectures of our module on advanced proteomics alright. So, in today's class this is the summary we discuss the label free quantification method how what is the principle, what are the types, how the workflow happens, what are the advantages, disadvantages and applications alright. So, these are my references for today's class and I thank you for your kind attention.