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 32 : Label-based Protein Quantification Technologies (Part 2)

Hello students. Welcome back to our lecture series on Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis. We are in our module 7 and today's lecture will be discussing will be continuing our lecture which we started in the last class I mean in the part 1 which in which we already covered these topics. And today's class will be extending that concept is a part 2. We will be discussing TMT that is tandem mass tags and SILAC which stands for stable isotope labeling with amino acids in cell culture. We will be focusing on discussion in these two topics for this class and in the next class we will be focusing on label free techniques and thereafter we will be discussing various methods of data acquisition.

Now, this is the slide from the last class and this is a revision slide because you already should know should be able to recall what is quantitative proteomics and what are the utilization of quantitative proteomics. So, I am not going to narrate each of these points because you already know this is the exact slide from the last class. And over there we already discussed the two major phases of quantification are quantitative proteomics are quantification methods of protein and methods of data acquisition. Now, regarding quantification methods we again discuss it can be gel based and gel free and we are here we are discussing isotope labeling methods.

Now, in isotope labeling methods when we are classifying them further there can be two mode of labeling that also we discussed in last class one is in vitro labeling method then one is in vivo labeling method. Now, in the two in vitro labeling method we already discuss isotope coded affinity tag as well as ITRA which stands for isobaric tags for relative and absolute quantitation. And as I told you we will be discussing this one which is an in vitro labeling technique that is TMT and we will be also discussing in vivo one in vivo labeling technique that is SILAC in this class. Now, so let us get right to the topic. So, TMT stands for tandem mass tag all right.

So, whenever you see this acronym in any MCQ just note this is we are tandem mass tags all right. So, what is it? What is tandem mass tag? Basically they are a chemical

level we are discussing isotope labeling techniques. So, the very basic principle will be same we will be labeling something with an isotope in some way all right. And then we will be utilizing those labels to quantify the protein with the help of mass spectrometry ok. This is the basic principle which applies for all of these methods.

These methods have been developed under various situation they have got their pros and cons, but the basic conceptually they are very similar all right. So, if you understand one quite well you will be easily able to explain the other on your own right. So, let us see how well you it is how well easy for you to understand the TMT method. So, it is a chemical label again chemical method all right in vitro. So, we are using chemical non living stuff in order to do this experiment ok that facilitates sample multiplexing very important again keyword we can process multiple sample at once all right.

In mass spectrometry based quantification definitely it is a method which is used in mass spectrometry based quantification for what for quantification identification of biological macromolecules such as protein peptides and nucleic acids. So, this is the basic working definition or the concept of tandem mass tags, but rather than I mean just memorizing the whole thing it is always suggested that student should always note the keywords and frame the definition on their own whenever you are attending any theoretical examination or descriptive answer type examination all right. So, as discussed in last class when we are discussing isobaric mass tag with I mean tags with identical mass there we had two one was eye track and one another is TMT. So, this also falls in the isobaric mass tag category all right. So, which are a set of molecules with the same mass all right, but yet yield reporter regions of differing mass after fragmentation.

So, initially as the tag the mass is same as a tag as a label when they are attached to a protein their mass are same, but after these are fragmented with the which is a part of the analysis system they are initially fragmented and the spectra is analyzed. After the fragmentation that fragment is known as reporter that are different and that actually helps us to quantify to which protein the reporter was originally tagged ok. So, again it is an in vitro labeling technology very important whenever we are facing an MCQ if it appears in vivo TMT will not be the answer all right. In fact, the only answer that is that will be discussed in this whole series is SILAC. So, that will be the only in vivo technique for your interest rather for everything else we are discussing in vitro.

So, ICAT, ITRA, TMT everything is in vitro so developed by thermo scientific all right. So, the good thing is technology can use up to 16 isotope labels to label the amino group of polypeptide. Now, till few years back it was the limitation was up to 10 10 plex now it can move do up to 16 plex. So, technology is improving day by day. So, in few years for night may become 32 all right right.

So, what happens after liquid chromatography and tandem mass spectrometry LC-MS-MS analysis the relative content of protein all different samples can be compared at the same time. So, all 16 different samples can be compared in one go therefore, it is a very high throughput technique where we can analyze multiple sample at once up to 16 different samples. So, hence where it is used. So, TMT is widely used in the field of disease markers screening ok drug action target animal and plant disease resistant or stress resistant mechanism animal and plant development and differentiation mechanism there these are the few common experiments proteomic based experiments in which we can multiplex the experiment. So, that multiple samples can be analyzed in one grow and TMT is very helpful in that regard.

So, this is the basic workflow of TMT schematic I will be again explaining it, but you can just pause for a second observe this diagram and then I will move on to the next slide where I will be describing this technique all right. Note this this is a diagram which is showing up to 10 plex this is a diagram of 10 plex. So, you just read it what it is being told here and then I will explain it again and I will again come back to this diagram ok. So, in case of a 10 plex. So, let us understand with the help of 10 plex 16 plex again works in a similar way.

So, 10 different protein samples we have got 10 different protein samples those are as we already know the standard workflow of proteomics the large proteins need to be degraded or digested to form small peptide right the same thing is happening here, but in case of 16 different samples all of them are denatured reduced alkylated and digested into peptide samples. So, 10 samples we have got now 10 fragmented samples right. What next? These peptide samples are then labeled with the tandem mass tag it is a kit that is available. So, there are 10 different dyes or tags it will be labeled in the kit. So, each and every sample can be labeled with them you just need to mix and incubate standard procedure.

So, after that is done the whole thing is mixed together. So, after the samples are tagged with the TMT labels they are combined into one sample. Then this combined samples are separated using HPRP liquid chromatography can you tell me what is HPRP? Yes, if you have answered it by now you just check with me whether you are right or wrong it is high performance reverse phase liquid chromatography. So, it is an HPLC, but it is in reverse phase we already have discussed what is normal phase and what is reverse phase what is the nature ofmobile phase in normal phase and reverse phase. So, you if you are not able to recall I highly suggest you go andagain reread the chapter on chromatography our video right.

So, after purification these samples are analyzed by HPLC MS ok each sample mind it after purification each sample is analyzed by HPLC mass spectrometry. So, you see

when after we do high performance reverse phase liquid chromatography all the mixed one will give different peaks all right and those can be those fractions are collected and each fractions are then analyzed with the help ofLCMS or HPLC MS right. And lastly data is analyzed to obtain a qualitative and relatively quantitative information of the protein. So, again so, if you now look at the workflow what is happening over here these are the same ok this has been shown with 5 different dye 5 plex this is 10 plex basically all the sample proteins are first degraded into small peptides then all of them are individually labeled with 10 different dye or depending on the kit what are the amount of labels 5 plex or 10 plex then everything is combined into one sample all right. Thereafter the fractionation is done with the help of high performance reverse phase liquid chromatography where we get each and every fraction then each and every fraction is analyzed by LCMS from which get the data analysis all right.

Now, see why do we do this? When we mix the entire thing all right we and if wesubject it to the first wave of mass spectrometry we get a spectrum like this in which all the tags are mixed ok depending on what proteins they were tagged into there can be different type of proteins in which each and every dye or tag has been attached right. But once we fragment them we get different reporter ion signatures all right and those reporter ion signatures give a classic picture and it is very easy if we just corroborate one reporter to back calculate to which protein they were attached using bioinformatic tools ok. So, remember the data analysis is a big part of it just like next generation sequencing where the experiment was one part and then comes data analysis which is also a big part similarly in case of quantitative high throughput proteomics after we get the spectra the spectral analysis is very important all right. Similarly, in case of 16 plex the same thing is happening we suppose there are protein samples from different tissue ok all of them are digested and tagged suppose there are 16 different tags all right. So, we can take 16 different proteins all right.

So, after each and everything is mixed and their first LC this is LC is actually the high performance reverse phase liquid chromatography. So, after that fraction they are subjected to tandem mass spectrometry. So, after the first LC it is then subjected to MS MS ok and then we get a signature spectra in which after the first MS the second mass spectrometry gives the classical I mean signature spectrum based on which we can tag each and every reporter and we can have idea where each and every proteins are being how each and every proteins are represented in across various tissues ok. So, we can get idea about just ignore this figure of brain it is just a neural network that is being shown here is not a sample of brain all right. So,brain tissue plasma serum CSS sample can be from anywhere ok if we can analyze all of them together and this technique will give you the same protein how is it differentially and relatively expressed across all of the different samples ok by using the same reporter because the same reporter ion will tag the same protein whether the protein is present in brain whether the protein is present in

plasma or CSS etcetera all right.

And then after another fractionation they will give the idea about the source of the protein whether the protein is from brain or whether the protein is from CSS etcetera etcetera ok. So, what are the nature of the tags in TMT technology? The tag actually is composed of four parts ok it has got if you look into the molecular nature of the tag it is a nice to know area it is not a must know area, but if you are interested you are definitely should read on. So, number one a mass reporter region denoted by m a cleavage cleavable linker region the area which is cleaved known as f a mass normalization region the n based on the signature of which the entire thing is normalized. So, that if the sample amount from 1 is more and the sample amount of 1 is less we can normalize the whole thing. So, that the whole thing can be compared at the base level.

So, importance of normalization region and the protein reactive group which is actually the group which reacts with the protein and helps in the attachment of the tags ok. So, the chemical nature of all tags are identical since it is an isobaric tag mind it is very interesting thing that is going on over here. The chemical nature of the tags are identical, but each contain isotopes substituted at various different positions ok such that the mass reporter and mass normalization region have different molecular masses in each tag ok. So, we know that I have tagged the sample of brain with tag 1, 2, 3 and 4 I have tagged the sample of plasma with 5, 6, 7, 8 I have tagged the sample of serum from 8, 9, 10, 11 or right and then 12, 13 I mean 13, 14, 15, 16 is the sample from say another whole blood. I am compare I can compare any 4 sample or even sample from 4 different individuals right.

Now, since we already know what sample has been tagged with what label now if we can compare the each and every normal if we normalize the entire region we will have definite information regarding all of the 16 areas to which this tags were bound. And it will give simultaneous data about the amount of protein that was expressed in those region. The concept is actually the same everywhere we are doing the same thing we are tagging a protein with the help of an isotope and then we are analyzing the nature of the isotope and the amount of the signal which actually corresponds to the amount of protein present in the tag all right which was tagged. The beauty of this is we can do 16 sample at once. So, this is the leap that has happened over the previous technology which was ICAT and ITRAQ specially where ITRAQ we could use up to 8 right.

So, moving on so, what happens after that the combined MFNR region. So, one mass region reporter region one cleavable linker region or normalization region and one protein reactive group. So, tags have the same total molecular weight and structure. So, that during the chromatography or electrophoretic separation in the single MS mode when we subject them to the first mass spectra MS the molecules level with different

tags are indistinguishable because they have got the same mass and since the structure are also same all right the m by z ratio will be the same upon fragmentation. So, once we fragment them this linker region when they are broken that so, the combined mass is same all right, but they are different the number of amounts of protein that is the amount of the molecule the weight of the normalization or reporter are different, but they are all that 16 tags are designed in such a so, that the combined mass is same I hope you got the point right now.

So, after we cleave them definitely n and r the sum of n and r in all the 16 tags will be different all right. Therefore, the sequence information obtained from the fragmentation of the peptide backbone which is obtained simultaneously they will be analyzed because the reporters are different. So, after the tags are fragmented then when it is subjected to another mass spectrometry MS they will give us the signature mass spectrum mass spectra which is characteristic for each and every reporter ion of each 16 tag and that will give us an idea about what was the concentration of protein in the parent sample ok . So, quantification of label peptides so, the TMT tags actually now comes the software department how do we do it that nature of the TMT tag structures can be accessed in public these are all available in public domain . In the website unimod you need to create an account you can just browse right now unimod.

org you will see a proteomic website is opening over there you first need to register and then you need to log in with your laboratory id definitely you need to register first and then you will be able to access all these information . Also mass spectrometry software like mascot all right can accurately calculate the masses of these tags see mass spectrometry mascot is a very popular software that is used in mass spectrometry and latest version of mascot possess the ability to perform quantification of TMT and other isobaric mass tag. For example, the mass tags of eye track can also be quantified using mascot software. So, basically it eliminates the need of any other supplementary software because suppose we have procured a mass spectrometer from company X and we have procured another lab has procured another mass spectrometer from company Y each and every developer will have their proprietary software that will work in a different way. The use of TMT nullifies that because we have got a software that is available in public domain and universally which is able to analyze and quantify the signal and which will not mandate us to procured the proprietary software from those companies ok.

So, what are the advantages very high sensitivity all right low abundance of protein can be detected. So, a very small amount of protein can be detected by use of a TMT when you are using TMT in mass spectrometry. Strong separation ability can separate acid alkaline protein less than 10 k d to greater than 200 k d. So, in extremes of ranges and even insoluble proteins they can also be now separated using this tandem mast tag technology. And definitely it can it does have a wide scope of application for example,

identification of many unknown protein for example, membrane protein, nuclear protein, extracellular protein everything we can target with the help of tandem mast tags.

And again see we have said 10 it is actually 16, 16 samples up to 16 samples the 10 plexis utilized I mean routinely because the cost has come down really compared to when it was discovered, but now at in today's date up to 16 samples can be analyzed at the same time and which is specially suitable for different protein samples and it is a protein sample and processing I mean instead of processing one single protein or sample multiple times we can multiplex the assay multiple samples can be analyzed together at once. So, how does it compare to iTRAQ that weread in our last class iTRAQ was also advancement over ICAT right. So, basically it is better why because iTRAQ with iTRAQ we can level up to 8 samples and TMT with the help of tandem mast tag and today's date it can level up to 16 differentsamples right 16 different isotopes can bedesigned I mean 16 different isotope base tags are used in TMT. Another thing the tags in TMT actually are very small compared to iTRAQ the labels of iTRAQ are actually large. So, what happens if themolecular weight increases the hydrophobicity of level protein or peptide also increases and that creates a problem alright.

When the sample is less hydrophobic the identification technique and quantification procedure becomes easier, but in case of iTRAQ the since the tags are larger it is a problem compared to TMT if we compare one is to one side by side comparison then it is a disadvantage of iTRAQ that the tags are relatively large right. And one very important thing is the nature of chemical that is used for iTRAQ the labeling the base medium is isopropyl alcohol or isopropanol whereas, in tandem mast tags it is acetonitrile. If you have paid attention to the class on chromatography and reverse phase chromatography I told you must have noticed acetonitrile is the main ingredient of mobile phase of many proteomic experiment most reverse phase chromatography experiment that we do in proteomics the mobile phase is made up of acetonitrile and water and therefore, TMT exhibits a much better compatibility compared to iTRAQ alright. And at the same mass spectrometric resolution TMT quantifies a greater number of peptides finalizing 6 samples this is a technical I mean thing if we are combining 6 samples. So, you see if we are combining 6 samples the logic might be or the argument may be.

So, if we have iTRAQ or if you have less than 8 samples what to choose. So, do we choose TMT or do we choose iTRAQ right because if it is more than 8 sample there is no other option we need to go for TMT right. So, what if it is less? So, it has been seen that when analyzing 6 samples TMT quantifies greater number of peptides ok. We can achieve a better result better picture separation using tandem mass tags ok. We can achieve the better quantification the amount of protein will be denoted much better with the help of TMT compared to iTRAQ.

Next we move on to SILAC or stable isotope labeling with amino acid in cell culture ok. This is the abbreviation is SILAC SILAC. So, what is it? It is again a stable isotope labeling technique where the keyword is metabolic labeling. It is a definite it is a technique of isotope based labeling method in quantity proteomic that uses metabolic labeling. How is it different from all these definition in vivo and by now you must have guessed why it is different because we are using live cell living organism I am living cell culture ok live cells I would not say organism it is a live cell which is you.

So, definitely this is therefore, this is actually classified lightly as in vivo we are not using non living chemicals ok. So, this technology enables accurate and replicable determination and measurement of relative differential alterations in protein abundance ok. Now, what happens actually? It involves replacement of light or normal normal isotopes proteins have got carbon and nitrogen right in amino acid they are replaced with their heavy version. So, 12 carbon is replaced with 13 carbon 14 nitrogen is replaced with 15 and 15 all right during protein translation in vivo all right. So, when a physiological process is occurring then the labeling is happening all right we are not labeling in vitro with the help of chemical experiment all right.

We give the condition we provide the condition to the cell and when protein synthesis happens the metabolically those heavy amino acids are engulfed and then proteins are tagged all right. So, again they have got a wider application comprehensive identification characterization and quantification of protein in complex organism mechanism. Basically when the mechanism is not clear how the protein is been incorporated what are the various pathways then by labeling the proteins at various stages with the help of silac and then analyzing them will give us a much better picture what is happening in a feasible for example, in a signals signaling system mechanism detection method of choice will be this all right. So, the regarding the principle this technique relies on cell culture all right where stable isotope labeled amino acid for natural amino acids in the proteome during the process of protein metabolism. So, whenever in cell culture we are providing the media the culture media which has got essential amino acid like leucine lysine methionine as well as non essential like arginine.

These are indispensable for cultured cell lines these amino acids that are present in the media for replication for the cell growth all right. So, we use this area we target this area where we replace the carbon and nitrogen of these amino acids with their heavy version all right. So, the these amino acids utilized in silac labeling. So, what happens the most utilized isotopically labeled amino acid in silac experiment a 13 carbon and 15 nitrogen labeled as KNR ok . A specific proteolytic enzyme trypsin we already know right cleave the carboxylside of KNR residue.

So, KNR stands for lysine and arginine. So, mind it. So, the trypsin will cleave the amino acid at lysine arginine residues and then in the next I mean whenever the cell is actually incubating the media for 2 3 generation passages they will incorporate those areas and the trypsin will cleave on those areas and then ultimate n tax will be analyzed. So, therefore, the combination of silac labeling with KNR and tryptic digestion allows for overall proteomic quantification. So, what is happening 2 groups of identical cells are cultured in the same complete medium except that one medium is normal common KNR common lysine arginine without any labeling that is referred to as the light medium all right. Whereas, in another medium it contains heavy labeled lysine arginine and those have got names K 4 R 6 K 8 R 10 etcetera those are known as heavy medium those are available commercially. So, so these cells are grown in the one grown heavy medium and one is grown in light medium all right.

So, after 5 6 generation the cell repeats itself will culture ok. So, if you incubate one cell in a plate or in a petri dish and if you cover it with a media the the it will repeat all right it will grow it will cover the entire surface and then again we need to scrape of the cell we need to centrifuge you need to split the cell that is also next passaging we need give it into another medium then again the cell divide. So, this is the standard way of cell culture that is already happening with the help of normal media here in one line of cell in one I mean condition or one group we are exposing it to a different media which has got heavy labeled amino acids. So, after 5 6 generation after 5 6 such passages in cell culture the heavy labeled amino acid are fully incorporated in the proteome. So, all proteins will have heavy amino acid and what will happen definitely the mass the amount the weight of the protein even if the nature are same for same identical protein.

For example, the membrane protein those two cells will have the same protein cell membrane, but the amino acid that is incorporated in forming the new cell membrane will be different and that will lead to a increase in molecular weight that will lead to discernible mass shape between the two groups of identical proteins. So, mind now we have got two protein just like a previous concept. So, we have got two proteins identical proteins which are labeled with two different tags in vitro the nature of the tags were different. So, by this I mean the analyzing different tags we are able to target the proteins here the tags has been done in vivo all right and again these tags are varying in mass which can be detected by mass spectrometry. However, there is no observed despite in the other chemical properties because it is same we just have just changed the amino acid by labeling them with heavy carbon and nitrogen.

So, in the next step the two cells are harvested in equal proportion followed by protein extraction subsequent equal mixing in 1 is to 1 mass ratio. So, we are harvesting we are mixing the cell lysate we are mixing all of them and then we are subjecting them to again digestion by trypsin followed by subsequent analysis using liquid chromatography and

tandem mass spectrometry. So, after the analysis done the intensity ratio of introduced isotope labeled peptides to unlabeled peptides all right if you just compare the spectra intensity of the level we could determine the quantification of identical proteins under two or three treatment condition. Suppose this one the protein is expressed in one way the second condition the proteins are expressed in different way. Now since we are targeting we are exposing these two proteins in the same media we have got the same property, but labeled in a different way we can now by targeting the same protein, but these proteins having different tags under mass spectrometry we can have idea what was the amount of the protein in the parent sample.

The concept of isotope labeling and its subsequent analysis is basically the same in all type of isotope labeled methods different labels at a different protein give a different unique mass spectra which helps us to analyze, but the way of tagging is different. So, what are the applications of SILAC distinguish quantitative variation in protein expression across various samples monitoring quantitative differences at protein level under different conditions. So, different experimental condition when we are giving stress or with the help of any drug we are discussing treatment we are giving any sort of chemical or stress in vivo hypoxic stress what proteins are expressed differently they can be studied. By differentially expressed in proteins in organelles such as nucleus and nucleola. So, basically we are comparing the same protein which is expressed differentially across various systems.

So, it be an intra sample or in a same tissue between various organelles so on and so forth. Protein turnover again can be monitored very important investigate dynamic changes in the PTMs that is post translational modification can be studied with the help of SILAC. And also distinguish protein protein interaction and network just because the things are happening in vivo we will be easily able to identify by characterizing a tag what proteins are coming together how are they interacting. So, since it is a in vivo experiment we can demonstrate very easily what are the physiological changes that are happening what receptors are interacting together what how they are giving a unique mass spectral tag when labeled with this stable isotopes in vivo. Thus they help by interacting they can directly discriminate them from unleveled environmental contaminants.

So, they will have I mean environmental contamination experimental contamination will have no effect in this because the only way this proteins have been leveled that is in vivo unless the cell divides unless the cell culture is viable tagging will not happen. So, what are the advantage again high labeling efficiency very important as high as 99 percent and it remains unaffected by lysate and other factors all right. So, any experimental manual contamination manual error will not have any effect on this unless the cell culture totally fails all right. Next high sensitivity the amount of sample

requirement is small only 10 microgram of protein per sample very high precision all right. Since multiple samples are mixed digested simultaneously and subsequently identify this sequential approach ensures consistent treatment of the samples minimizing the influence of experimental procedure as I told you human error the effect of human error can be nullified all right .

In other techniques if you just there is a problem in petting the amount of tag if it is added more it can affect the labeling all right not here. Here the whole thing is been controlled in vivo the amount of protein that is being incorporated will complete depend on the physiology not our reagent pipetting. Equipment variation will have no effect therefore, enhancing the precision and reproducibility. So, if with the help of SILSC we can establish the amount of protein that has been expressed more from one sample to another it is much convincing compared to other methods of isotope based labeling.

So, that brings us to the end of this class. So, in today's class we have discussed two techniques one is tandem mass tag on in vitro isotope labeling techniques. We have discussed the principle workflow nature of tag the quantification software the advantages and how it is compared to iTRAQ which we read in the previous class. We also discussed stable isotope labeling technique with amino acid in cell culture SILAC we have also discussed and we have discussed principle workflow applications and advantages. So, these are my references for today's class and I thank you for your kind attention.