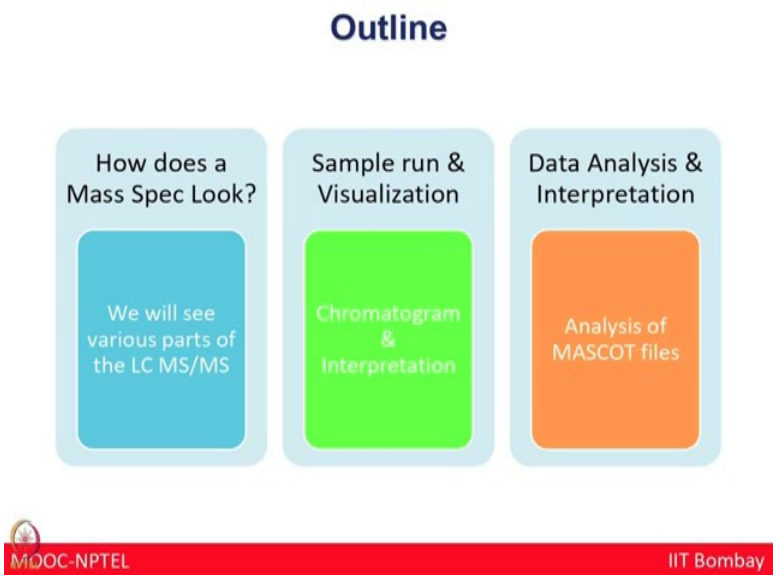
An Introduction to Proteogenomics
Dr. Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology, Bombay

Lecture – 12
Proteomics: Sample Prep & Protein Quantification

Welcome to MOOC course on Introduction to Proteogenomics. In the last module, we covered genomic technologies. In second module, we are talking about proteomics base technologies, and later we are going to talk about various type of big data tools and proteogenomic investigations. So, in second module in the last lecture, I talked to you about an overview of proteomic technologies, the transition from gel based to the mass spectrometry based or gel free proteomics to various type of targeted and functional proteomic based assays.

One technology which has made the huge revolution in the area of proteomics is especially mass spectrometry. So, today's lecture especially I am going to talk to you about MS based proteomics workflow. And here the interest will be also to give you some idea for how to do sample preparation and protein quantification in the laboratory settings.

(Refer Slide Time: 01:19)

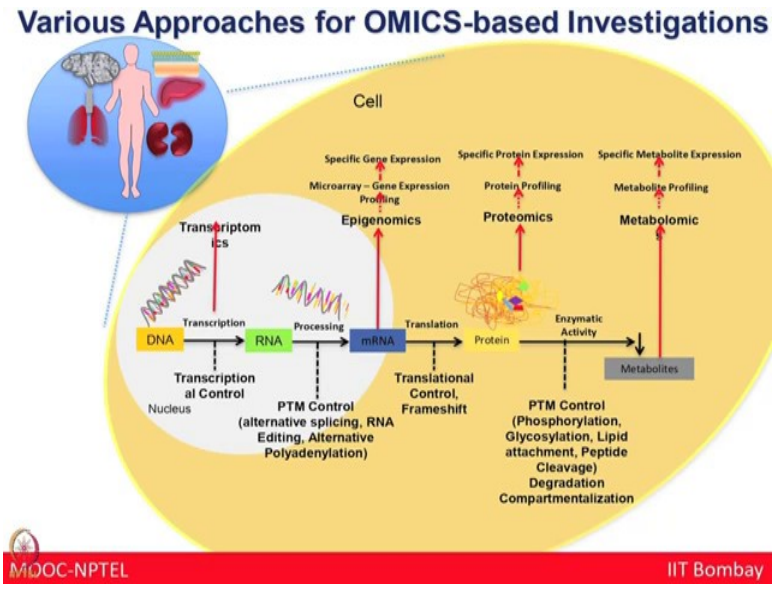


So, first we are going to talk about how a mass spectrometer works, various components of LC and MS/MS, then the sample run and visualization, how to look at the

chromatograms and do data interpretation, and finally, data analysis and interpretation using a free tool MASCOT.

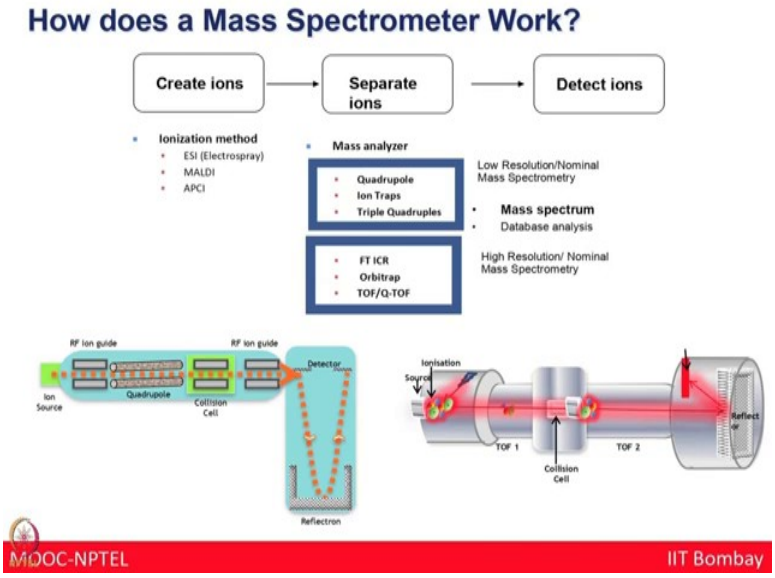
So, just to briefly touch upon what we talked in the previous lecture as well, there are various type of OMICS based investigation are making revolution in the areas of you know various type of clinical problems you really try to understand a system in a much better way. So, from the same we know organism or same clinical sample, now you can look at the information at a DNA level, RNA level, protein level as well as metabolite level. And there are could be variety of ways of doing these investigation to obtain information at genome, transcriptome, proteome and metabolome. And how one could try to put those thing together can really provide a very novel insight to understand a given system.

(Refer Slide Time: 02:30)



So, this slide gives you a very broad overview of various approaches which one could take to understand a given system using OMICS-based technologies. Not specifically we are talking about proteomics right now, but this gives you the nutshell of various OMICS technologies which could be used for a given system for better understanding.

(Refer Slide Time: 02:48)



Let us now move onto the mass spectrometers and mass spectrometry based proteomics, how a given mass spectrometer works. So, very broadly we want to talk about any mass spectrometers. First of all you have taken the protein of interest, you have done digestion of that protein, you got the peptides now made. And those peptides are in the solution form which you want to further separate. So, you know a promising way of doing that could be LC MS/MS based approach especially using electrospray ionization as a way to do ionization.

So, if you are linking your protein in the solution form or peptide in a solution form with the liquid chromatography, directly interfacing to the mass spectrometer ESI or electrospray ionization is the right choice. If you want to take now the protein separately, and then you want to add some matrix to that and make it in the solid form, then you can use another type of ionization method which is MALDI or matrix assisted laser desorption ionization.

So, the first goal is to create ions from the peptides of interest either using ESI or MALDI or API or APCI type of ionization methods. Once you have created the ions, now you from the same you know protein you got many peptides made and you have got you know many more ions created from those. So, if let us say you talk about a given sample which is tissue sample you have approximately 5,000 protein. Now, you are able to generate 50,000 peptides from that sample, and those 50,000 peptides now are getting you know millions of ions, and those ions have to move it inside the mass spectrometer

to further get separated which is going to give you the information for the peptide sequence.

So, you want to get as many ions moving it inside the mass analyzers of mass spectrometer as possible. So, there are various physical principles involved to do the separation of these ions. And how effectively you can separate these ions actually dictates and gives you the possibility how well you can identify these proteins eventually. So, there are different type of mass analyzers in play, especially the low resolution based, mass analyzers, or the high resolution based mass analyzers, a quadrupole, ion trap or triple quadrupole could be termed as a low resolution mass analyzers, and Fourier transform ion cyclotron resonance, Orbittrap or you know Q-TOP together could be termed as the high resolution mass analyzers.

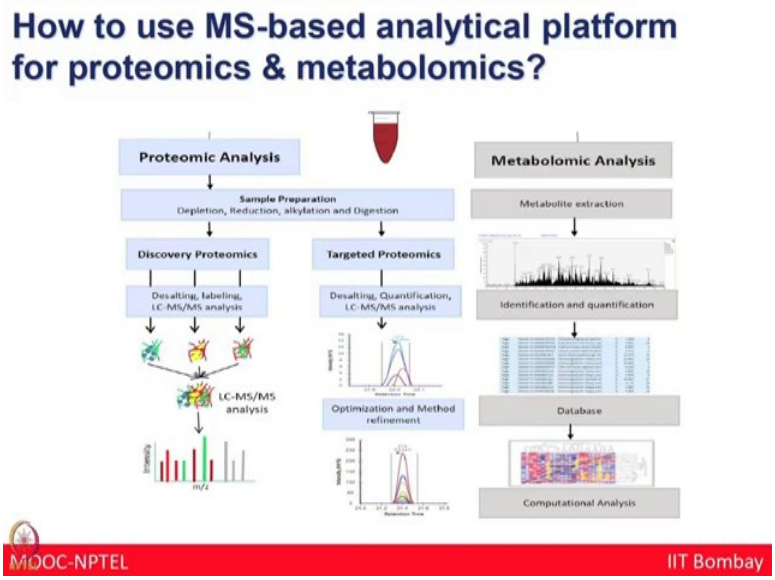
So, these are various types of mass analyzers. Their goal is to separate these ions which are peptide converted to the ions now, in the either based on the you know magnetic electric field or based on the time of flight how much time it takes to travel inside those trough tubes eventually to create to separate these ions based on the master charge ratios. And then finally, you need a detector which could now detect these ions. So, any mass spectrometer will broadly have these three configurations of ionization source, mass analyzers and the detector.

So, I have just shown here you know cartoons of two different type of popular configurations of mass spectrometers. On the left side you can see the Q-TOF configuration where you have the quadrupole and then linked to the time of flight. So, now you got two different mass analyzers linked to each other. This is known as hybrid MS configuration which could be utilized for MS/MS analysis.

So, now, you know you have an ionization source. These ions are passing from the you know quadrupole first, getting fragmented there, and then there again further revolving in the time of flight tube for further separation or you can use MALDI TOF TOF for instance. Now, here you are separating first you know based on the creating ions based on the MALDI ionization source, then separating based on the time of flight tube, selecting the most intense precursors, fragmenting those and then doing the MS/MS analysis using another TOF tube.

So, these are the popular hybrid configurations. More recently there are you know the emergence of newer technology like Orbitrap which even gives you the possibility of looking at tribrid MS configuration and some of those will talk in the subsequent slides. So, question comes how to use mass spectrometry based analytical platforms for proteomics and metabolomics analysis?

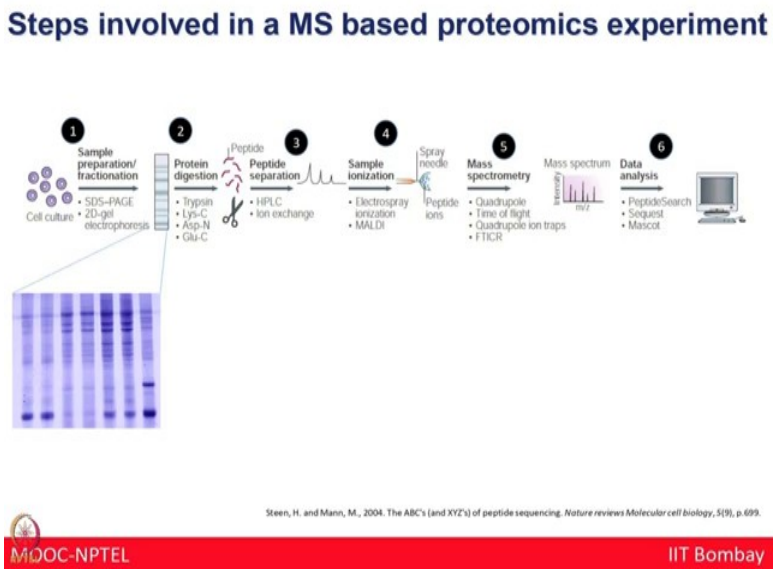
(Refer Slide Time: 07:32)



So, in a very you know kind of busy slide, but you know from the same sample let us say you want to look at the serum proteome and metabolome, from the same sample you can you know deplete the abundant proteins and now process zones for the reduction alkylation and digestion to proceed for the you know using either discovery based workflow or the label free quantification or targeted quantification to look at the entire host proteome or the plasma proteome analysis or you can also from the same sample, you can get the metabolites extracted and then from the mass spectrometry or NMR, you can separate those metabolites identify the part possible metabolites from the same sample.

So, now, you know using these kind of analytical technologies from the same sample you can obtain the picture of all the post proteins and metabolites changes happening in individuals.

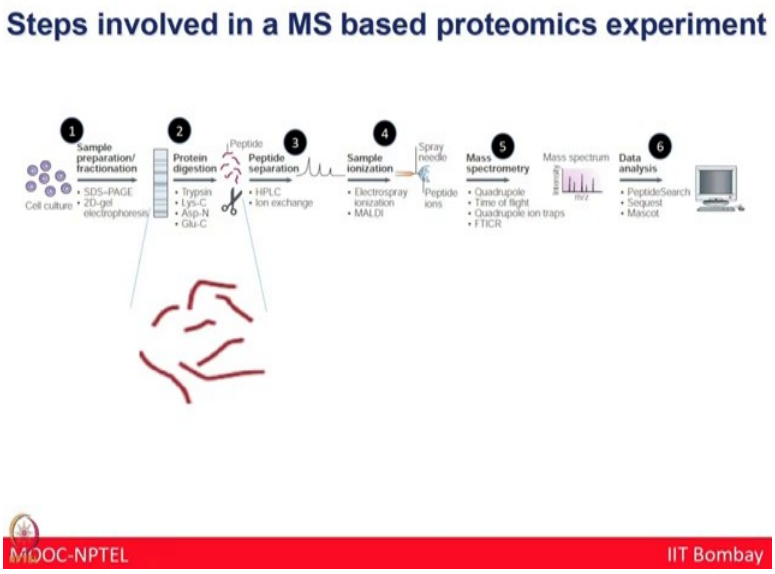
(Refer Slide Time: 08:19)



Let us focus mainly on the proteomics or the proteome right now. And you know very nicely you know summarize in this particular review article, where the ABC and XYZ of the proteins have been described. Just you know let us walk through the procedure and steps involved and described in this article, but you know some of these steps we have shown from our dataset.

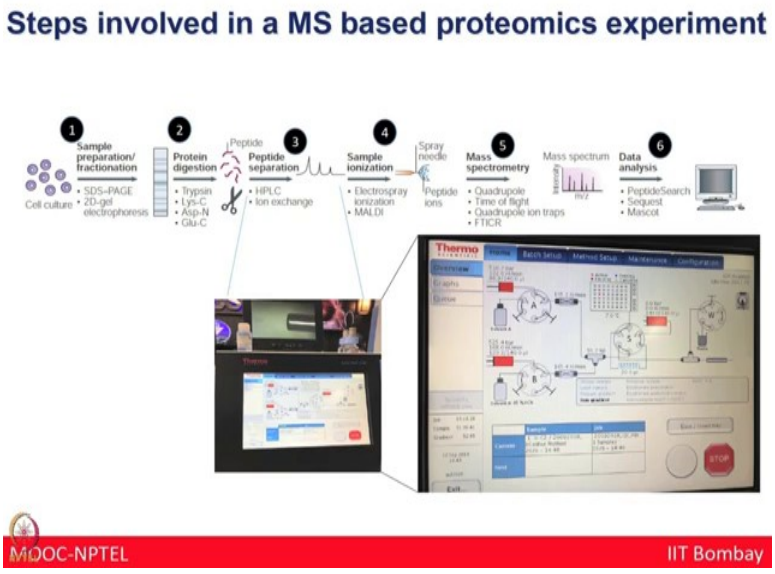
So, let us say you have taken done the protein extraction from a you know cell culture or the tissue. And now you want to further separate those directly either you can use the in gel digestion or use the in solution digestion, so that is shown in the first circle. Here you can separate the complex proteome on the gels to first ensure that you know the protein profile looks good, that the lysate looks good. And then either you can you know cut the entire lane of these protein from the gels, excise all the band for the in gel digestion or go back to the lysate and do the direct in solution digestion of this protein. Good here is to convert the complex proteins to the simplified peptide forms.

(Refer Slide Time: 09:22)



You can use various type of enzymes like trypsin, Lys-C, various type of other you know chymotrypsin enzymes are present which could be help helpful to do digestion of these proteins. Of course, doing you know good protein digestion is very crucial. So, you need to use the you know the right conditions for these enzymatic treatment to happen.

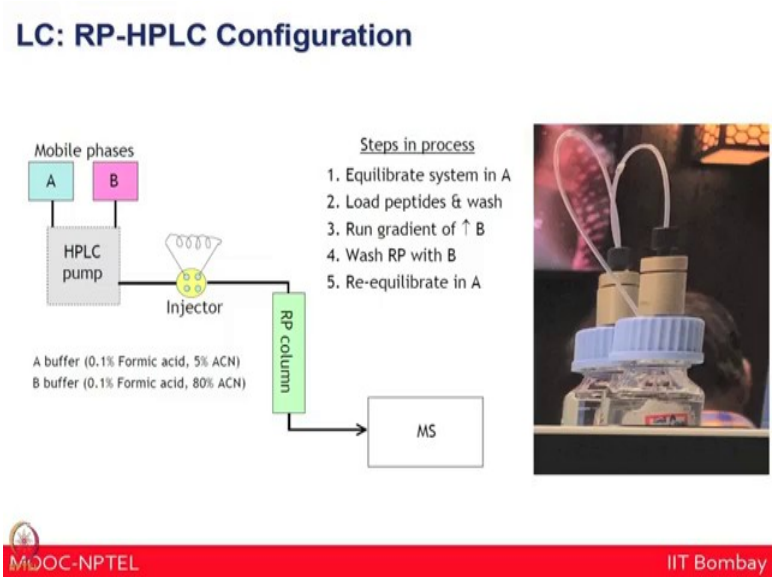
(Refer Slide Time: 09:42)



Once you have done that then you want to now separate the complex peptides based on the different type of chromatography properties and that is where a different type of HPLC configuration, nano LC based on the reverse phase or a strong cation exchange chromatography can be very helpful to now separate these complex peptides to the you know using some principle of chromatography to separate them over the time period.

Once you are now separating these peptides, a popular configuration in the reverse phase HPLC, where you are sequentially ejecting your peptides based on the hydrophobic and hydrophilic interactions. And you are using a gradient of formic acid and acetonitrile to you know once the peptides are bound in the reserved phase C18 material, you are sequentially ejecting them based on this gradient of acetonitrile and formic acid, so that you know not the all the peptides are coming out of the HPLC columns at the same time, but rather you know over the time period you are separating the peptides, and now they can be analyzed using mass spectrometer.

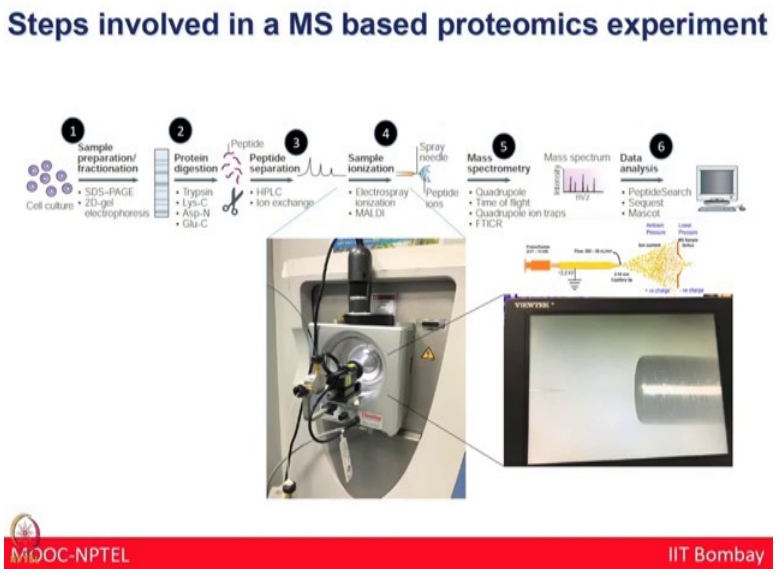
(Refer Slide Time: 10:43)



So, this is actually this configuration is known as online way of you know separating peptides because you have the you know complex peptide sample which is binding to the columns, you know getting it out slowly from the HPLC or nano LC. And then you are you know directly analysing those using electrospray ionisation followed by the mass spectrometry. So, some of these are the you know the details of how one could try to play with the gradients because each sample will have you know different complexity of the hydrophobic, hydrophilic residues.

So, you need to play with for a given sample type that what should be the best gradient to separate your you know complex proteome peptides from that sample type. And we are not talking right on much more detail, but you need to play with this to obtain the good chromatogram of the peptide separation.

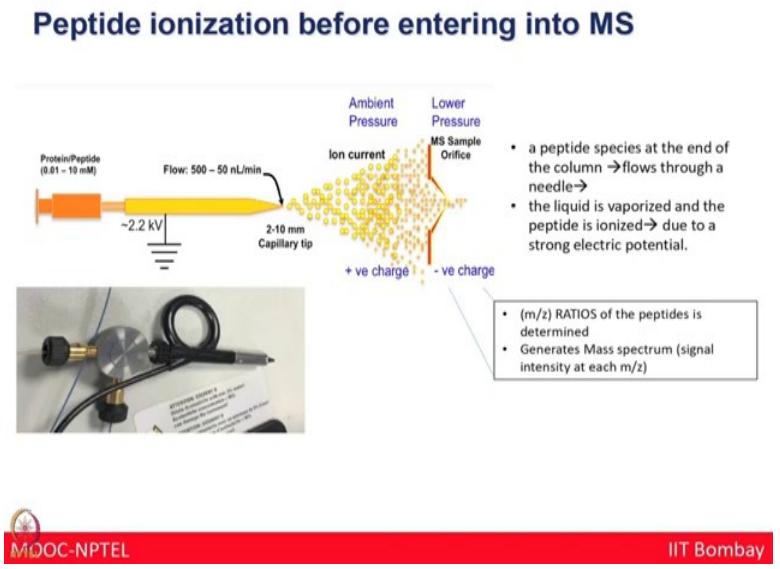
(Refer Slide Time: 11:33)



Once you have optimized a good you know chromatography, now you want to separate these peptides in the liquid form to coming to the ionized forms and that is where the electrospray ionisation or other possibility which we talked earlier MALDI, API, APCI could be utilized as well.

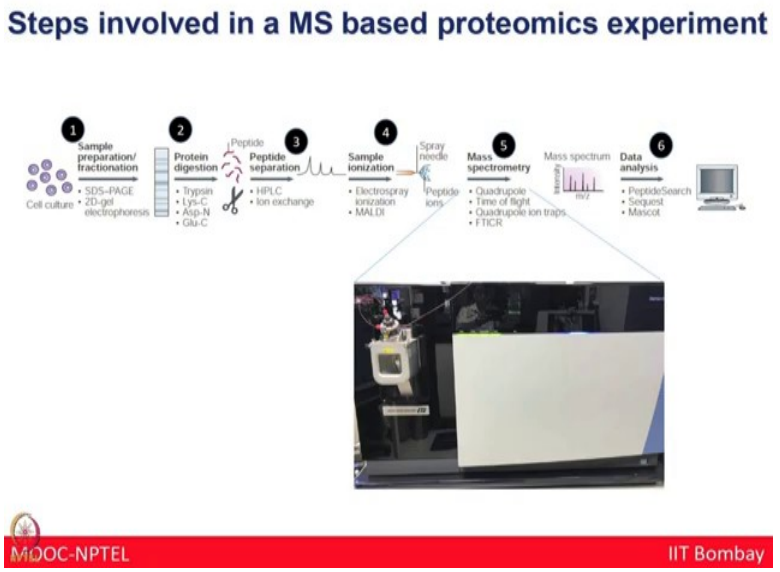
What is shown here that let us say you know from this column, you have now this solution form of the peptides there. Now, you want to convert them to the gaseous ionized forms. So, to do that you are using very high voltage you are also using the pressure differential, and then trying to move all these ions from the electrospray ionisation to the inside the mass analyzers for the further separation.

(Refer Slide Time: 12:18)



So, shown here is much more detail of the process for the electrospray ionisation, where now this you know the nebulization is formed. And you are using the you know the pressure gradient and the charge gradient to ensure that all the peptides in the ionized forms are moving inside the mass analyzers.

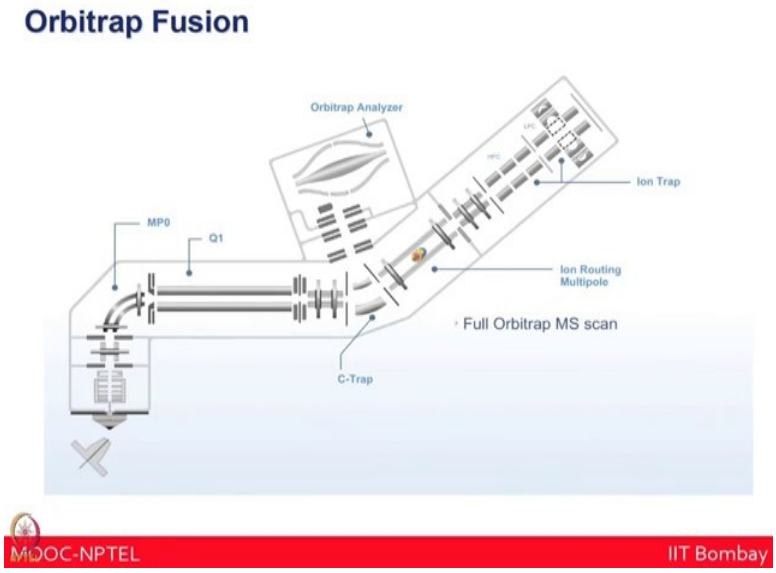
(Refer Slide Time: 12:49)



So, then comes the you know the how to best now separate these ions build on the mass to charge ratios and that is the latest technology like Orbitrap fusion could be very helpful. We have already talked to you about MALDI TOF TOF or Q-TOF kind of configuration which was hybrid technology, but this probe technology is known as the you know tribrid technology. Of course, here you can use any of those configurations of

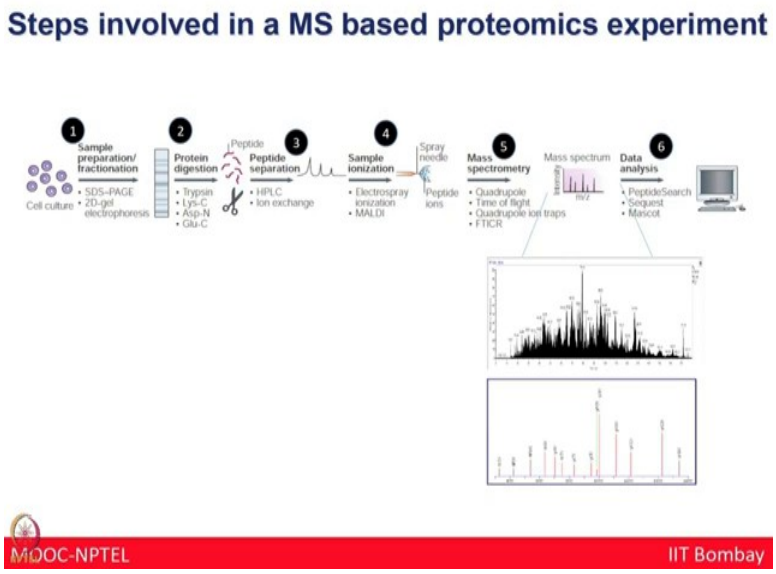
quadrupole, time of flight or even the orbitraps, but I will just now talk to you briefly about orbitrap fusion technology which is we are using it more in our lab and very robust latest advancements.

(Refer Slide Time: 13:14)



So, in orbitrap fusion technology, you know you are using three different type of mass analyzers. And you are using first the quadrupoles, you also have an orbitrap, and you have an ion trap. So, three type of mass analyzers or three different properties of mass analyzers could be utilized here. And therefore, you have more possibilities of playing with different type of applications in this kind of work flow.

(Refer Slide Time: 13:39)

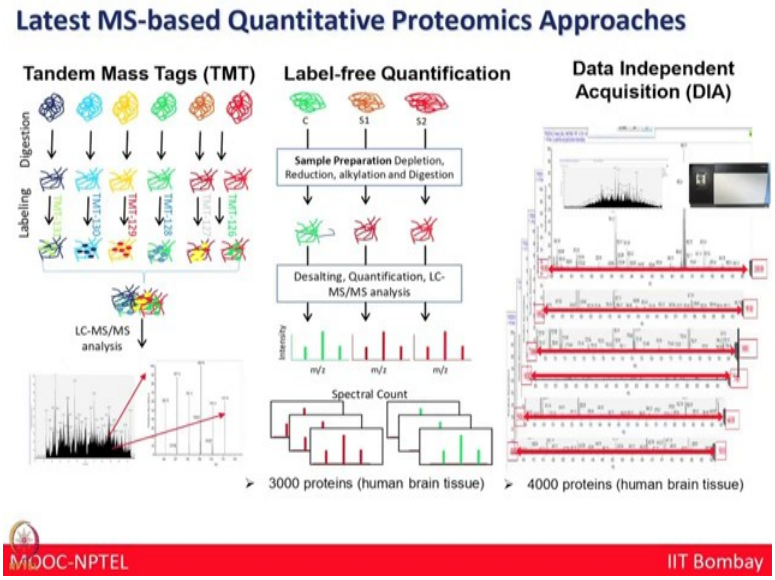


So, from these instruments what you generate, you generate these kind of chromatograms which is shown here in the insights. These kind of you know black forest you can see which shows you the you know the Gaussian distribution of peptites coming out of these columns. And from those you are looking at various you know their ion information or b and y ions, and then try to derive the peptide sequence from this information using various type of databases.

So, we will talk about some of this basics in you know in subsequent part of my lecture, but it gave you the workflow of doing the MS based proteomics is starting from the you know sample prep moving to the protein digestion separating them within the liquid chromatography, ionizing them with aeration sources, separating using different type of hybrid or tribrid configuration and then doing a database search that is a work flow for any type of mass spectrometry based proteomics pyramids.

The latest things for the quantitative proteomics are shown here. I talked to you briefly about iTRAQ technology in the previous lecture, which is isobaric tagging for relative and absolute quantification which labels peptides on their N-terminal side. A similar kind of approach has been used for tandem mass tags or TMT based approach where again you can label the peptides from the N-terminal side. Additionally one could look at the relative abundance of these you know peptides in the label free manner that is known as label free quantification. Or the more recent way of doing the label free quantification is also known as DIA or Data Independent Acquisition.

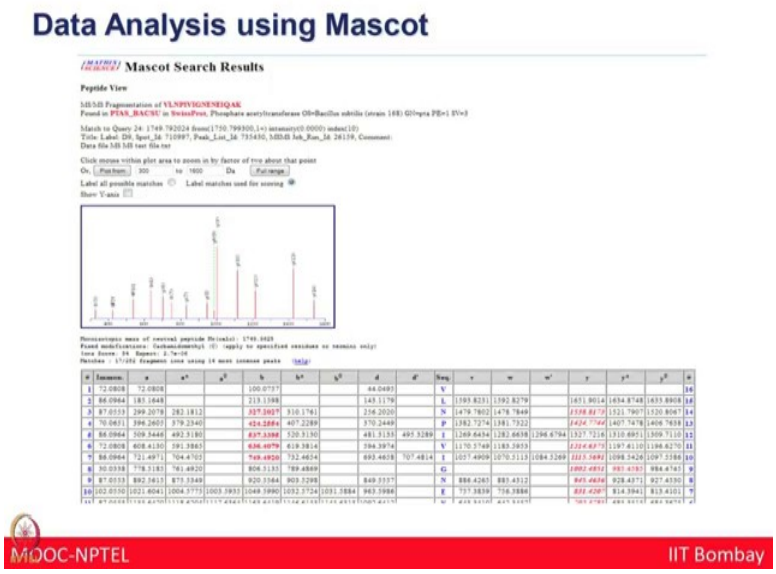
(Refer Slide Time: 15:07)



where you are looking at a specific you know region of m/z from 400 to 500 to up to 1000 region to generate much higher resolution of the spectra. So, these are the latest approaches of doing quantitative proteomics. Of course, as you go along with the workshop, you will be exposed to many of these concepts in more detail. And this kind of trying to set the path for you and giving you the foundation that how different type of technologies are emerging right now which could be utilized for the you know various type of applications eventually. And probably in my next lecture, I will also try to talk to you about the clinical problems in our lab, how we are trying to investigate those using some of these latest tools available in proteomics.

Let us now come and talk about briefly about data analysis aspects, especially you can use one of the free you know tool which is MASCOT where you can so the data file which is generated from the mass spectrometers especially mgf or acs files you can analyze directly in the MASCOT. And you know there are set of the parameters which you have to follow to ensure that you know you are able to search either for MS/MS or peptide mass fingerprint depending on what application you are looking for.

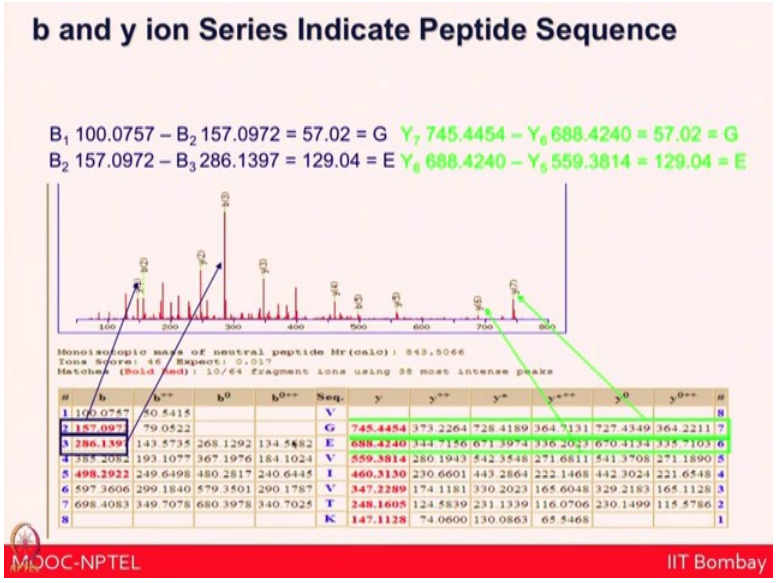
(Refer Slide Time: 16:18)



But what you know you look from this kind of data set is essentially that are you able to obtain the you know good number of b and y ions which are generated from the fragmentation. So, when you are doing the collision inside for the MS/MS analysis, during the fragmentation of the peptide bond of the CO-NH bond, the b and y ions are generated. The b ions are retained more towards the amino terminal and the y ions are retained more towards the carboxyl terminal. And therefore, for a given peptide bond when it breaks you are generating both b and y ions, but looking at the carboxyl or the amine terminal, you can now look at you know are they close to the you know carboxyl terminal for the y ions or close to the amine terminal for the b ion series.

So, when you are using a limelight trypsin, you are doing more you know digestion at the carboxyl terminal of lysine and arginine, therefore, you have much more higher abundance of the y ions. So, software gives you the pattern and those which passes the threshold are shown in the red colour. So, the y ions now you can see here in this particular one most of the y ions are red in colour, whereas, the b ions are very few which are red in colour.

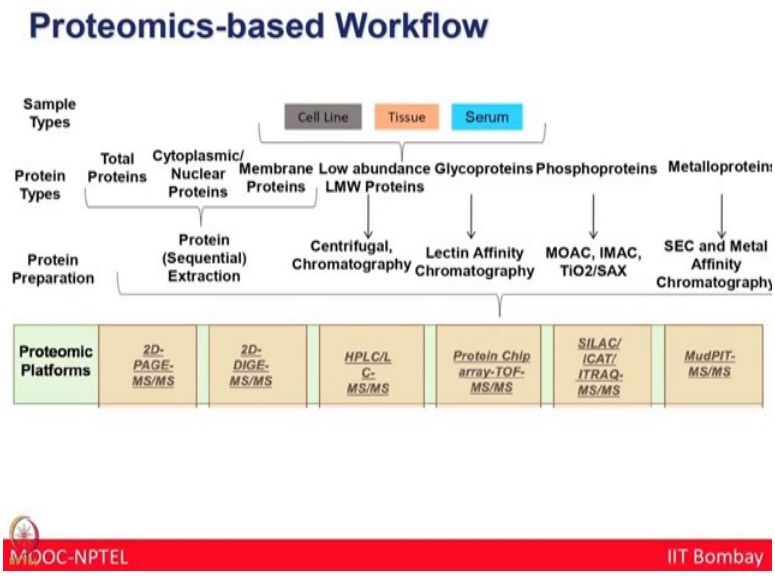
(Refer Slide Time: 17:31)



How you can use this information to generate the you know the peptide sequence, very briefly if you are looking at this particular spectrum and looking at b and y ion information. So, the difference of b1 and b2 is let us say 57 in this case which is the number specific for the glycine amino acid or when you look at the difference of b2 and b3 it gives you 129 which is specific molecular weight for the glutamic acid. Now, if you are looking from the y series y7 minus y6 it gives you exactly same 57 difference so that confirms that from the n terminal or c terminal, when you are looking at y7 and y6 or corresponding b1 and b2 ions, it gives you the same amino acid.

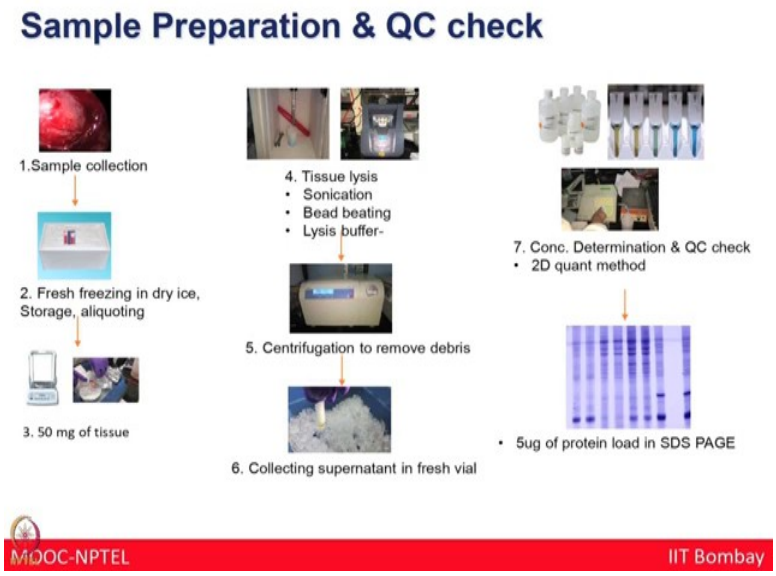
Or now you can go to the next amino acid and you say y6 minus y5 also gives me the same difference, it means the these amino acids are correct for this peptide sequence. So, in this manner, you can manually derive the amino acids and make the peptide sequence for a given stretch of sequence. This is just you know some you know basic understanding for you how the peptide sequences are being derived. Of course, you can you know do it manually, but if you have to do for thousands of you know peptides have millions of peptides, then you need to use these software to obtain these information and much more high throughput way.

(Refer Slide Time: 19:03)



So, coming to the proteomics based workflows again you can use variety of sample type you know because the technologies are very robust, you can use any kind of sample from the you know tissue, serum, cell culture, any of the you know bio specimen you can utilize. And then you know after doing the you know good protein preparation which you know for which you have you need to know your system very well. You need to do the you know good way of protein extraction and separation. Then you can use various platform for investigation of different type of problems, especially you can use the mass spectrometry based approaches or 2D dye each followed by mass spectrometry or you can use quantitative mass spectrometry or you can use shortgun proteomics or you can also use complementary technologies like microarrays and SPR for different investigations.

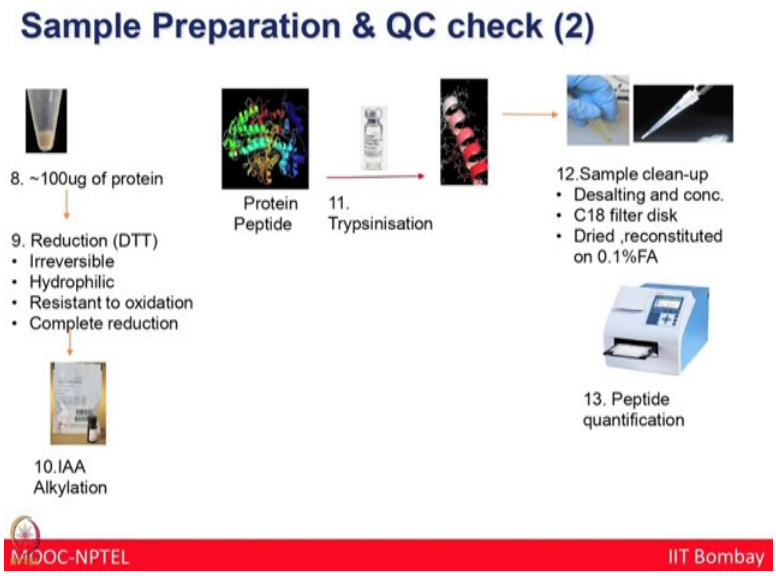
(Refer Slide Time: 19:43)



Let us talk more of the practical way how to do some of these things. Especially let us start with the sample collection. So, you know collecting samples for the biological you know bio specimens are not easy, especially it requires ethics approval to start a given study from the clinical samples. And then you need to you know work under supervision of these conditions and pathologist to obtain the right sample types, immediately they have to be flash frozen and is stored in the right condition, and then immediately you have to do the protein extraction to take the your proteome analysis forward.

So, a lot happens on the sample collection and the each sample processing level. Of course, you need a minimum amount of tissue to do this kind of processing let us say around 30 to 50 milligram of tissue. You do lysis, you can do sonication you know in different way the bead beating or lysis buffer addition. To do lysis centrifuge to remove the debris, then you can collect the supernatant which could be further you know utilize for doing the quality control checks, QC checks as well as to determine the protein quantification values for your further experiments.

(Refer Slide Time: 20:45)



So, once you have done those you know kind of you know basic idea for protein quality and protein quantity, then you want to now you know take let us say 100 microgram of protein for further digestion. To do the digestion before you start you know treating with enzyme like trypsin, you know to first do the reduction to make sure that all the disulfide bonds are broken, and then you are doing the alkylation to make sure that they are not reforming these bonds again back, then doing trypsinisation and then you are going further for the sample cleanup steps if the peptides are not digested, but because of the interference from the buffers and the salt which were used you want to remove them and get rid of them. So, you can pass them from the C18 filter papers or the you know Ziptips, and then you can only elute out the clean peptides for further investigation.

So, we are going to do these kind of you know basic experiments in the lab to show you how you know you can take any kind of complex bio specimen do the protein extraction you know then digestion, of course, in the process you have to do the protein quantification and the quality check on the gels to ensure that your process is fine, and then only you are ready to do any kind of mass spectrometry based analysis.

So, in the summary, I tried to give you a brief idea how a mass spectrometer works, and how a mass spectrometry based proteomics workflow could be utilized, you know which involves series of this steps from the sample preparation to the looking at you know the digestion, followed by separating using chromatography, and then utilizing complex mass spectrometers for further separation and protein identification or you can also use different type of tags or no tag even label free to the do the protein quantification.

Most of this information which we are talking is heavily used in many laboratories. And I will continue my discussion about proteomics workflow and how it could be utilized for clinical problems in my next lecture, which will be the third lecture of the basic proteomics site and then we are going to jump into the workshop where many of our you know the expert colleagues are going to talk to you about various tools and new software's and you know new ways of looking at the proteome analysis in a very very high throughput manner. I hope you know some of these basic lectures are giving you the right foundation to build your concepts and now you can take it forward as we go along in the hands on session to build it further.

Thank you.

(Refer Slide Time: 23:19)