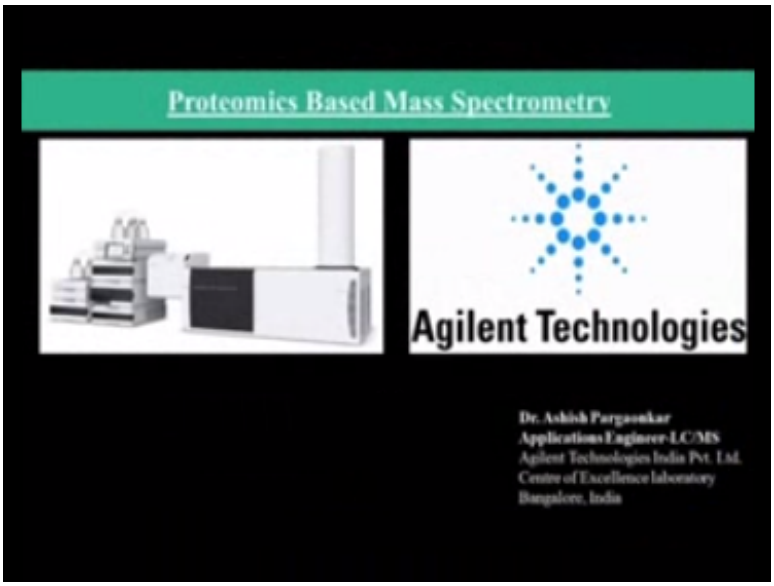
Introduction to Proteomics
Dr. Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology – Bombay

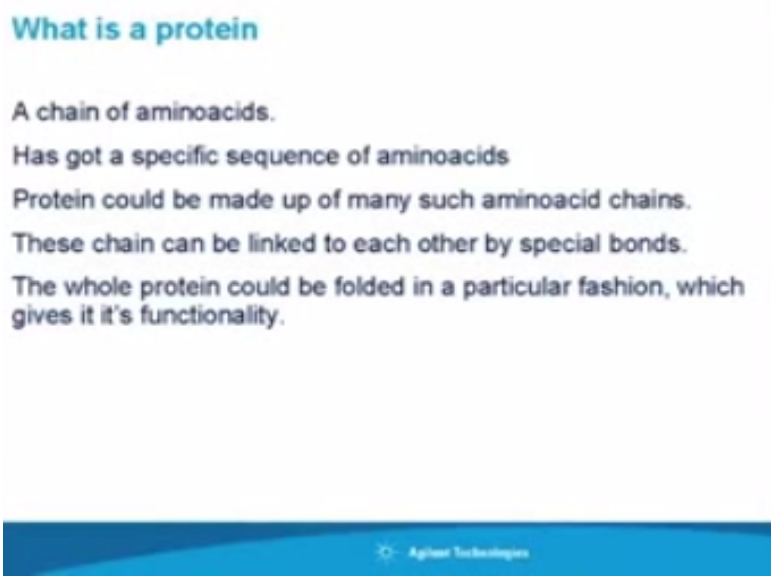
Lecture – 28
Lab Session – Demonstration of Q-TOF MS Technology

(Refer Slide Time: 00:21)



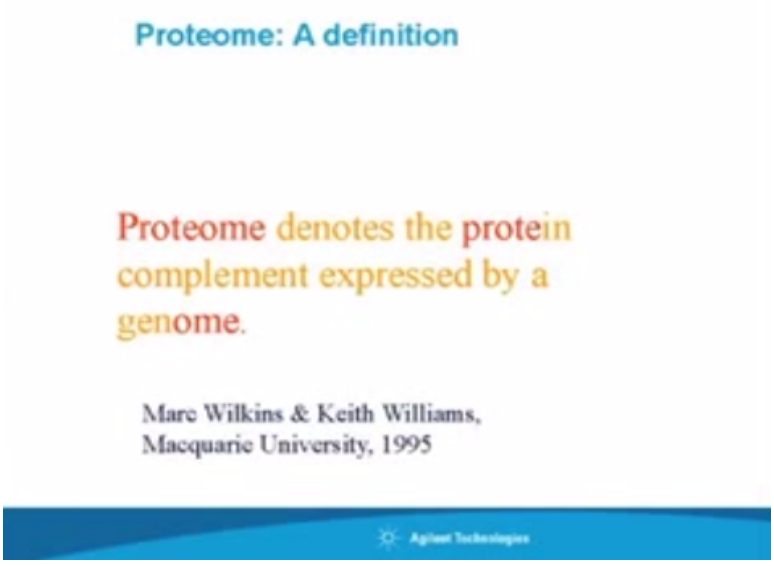
So, good morning everyone. I am Dr. Ashish Pargaonkar and I come from our applications lab which is called Centre of Excellence in Bangalore and I take care of mass spectrometry based proteomics, metabolomics, biopharmaceutical applications, taking mass spectrometry as an approach.

(Refer Slide Time: 00:41)



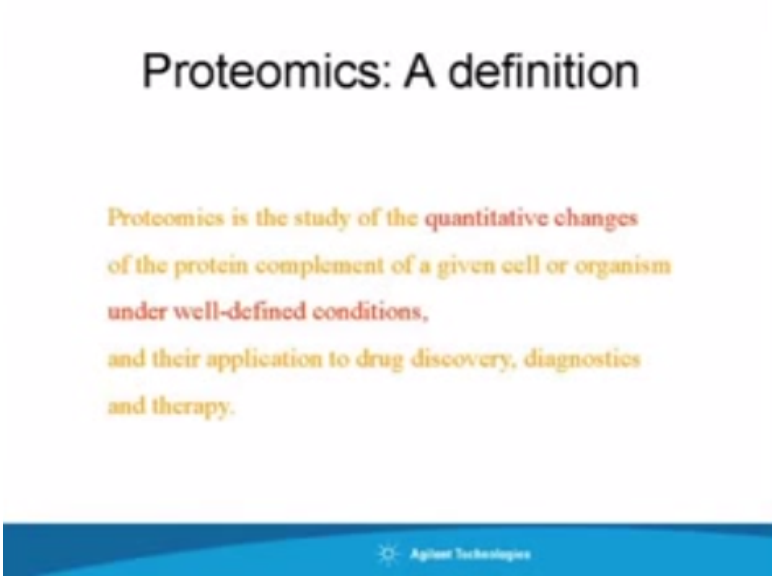
So, basically what is protein and other things, you will know everyone. Like, it is a chain of amino acids, multiple amino acids, polypeptides, they can call it as a peptide, polypeptide and then protein.

(Refer Slide Time: 00:50)



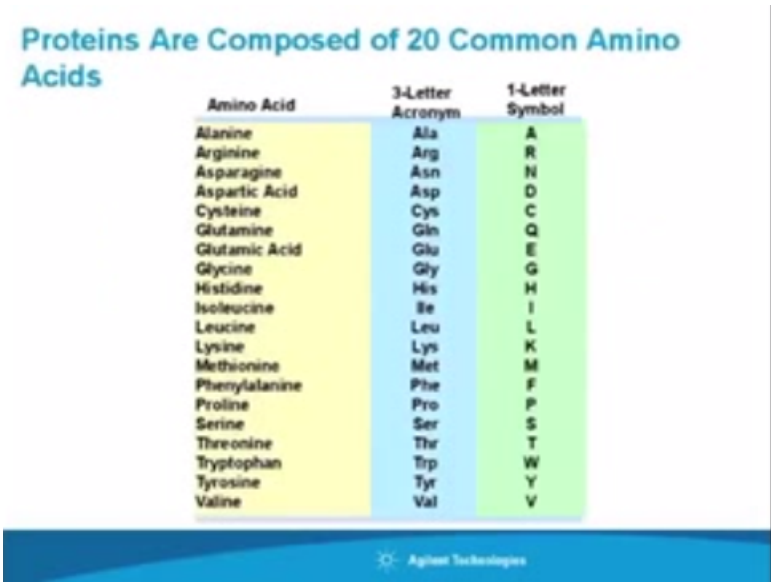
So, proteome denotes the protein compliment expressed by genome.

(Refer Slide Time: 00:57)



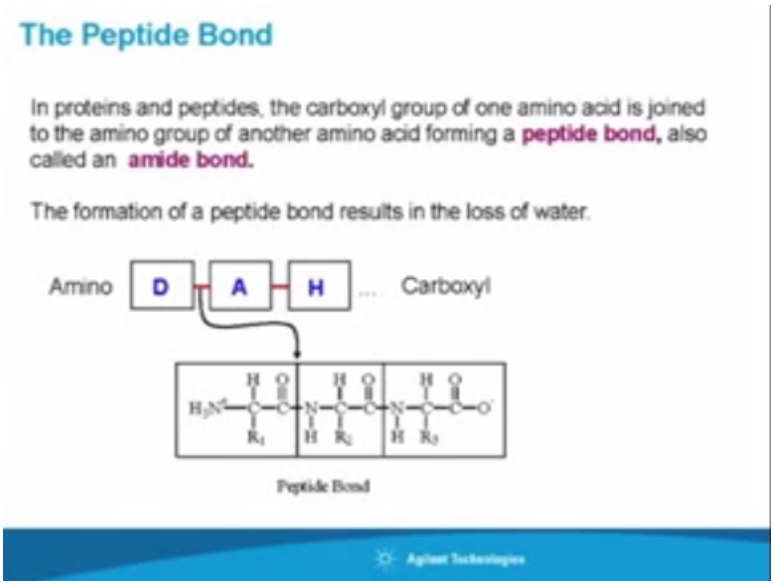
The, proteomics is a study of. So, what is the difference between this term and this term, proteomics, metabolomics, transcriptomics. When it becomes omics, why so omics. So, this term makes it omics. Whenever you measure the quantitative changes of a protein compliment, it becomes protein; of a metabolite compliment, it becomes metabolomics; of RNA transcriptomics; glycan glycomics; lipids lipidomics; anywhere where you measure the changes and you quantify them it becomes omics, okay. So, that is how it is.

(Refer Slide Time: 01:38)



So, then of course there are various proteins composed of 20 common amino acids in different combinations, okay.

(Refer Slide Time: 01:47)



These amino acids are basically linked together by a bond called amide bond or also called as a peptide bond, okay. So, they link together in a peptide bond. All this information is very critical because these are the basis how we interpret in mass spectrometry, okay. So, we must know what is actually happening, okay. If you know that, we very well understand what exactly is happening in mass spectrometry and how to interpret that, alright. So, this is critical.

(Refer Slide Time: 02:18)

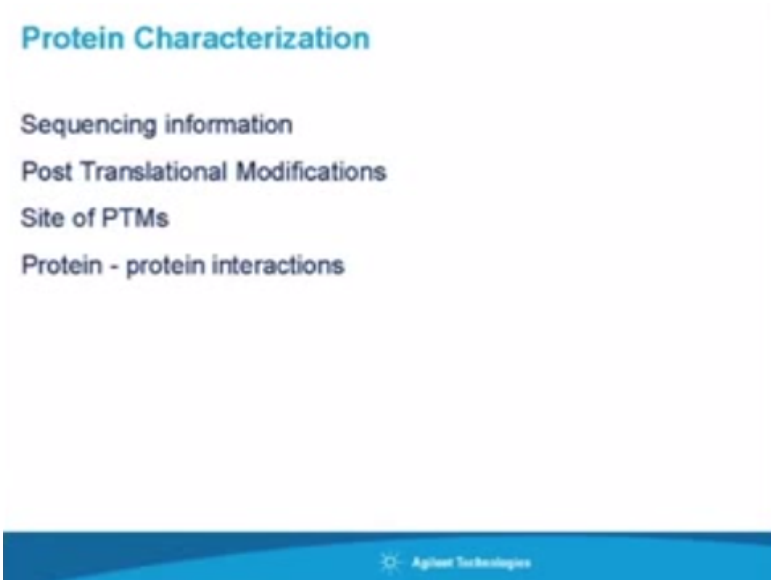
So, what we do exactly there, it is not that complicated. It is quite easy but only thing like there are multiple steps. First step is to identify. You have proteins of two states. You first identify what are those proteins. We have only numbers right now, right. We have to get some identity, okay this protein and all those things then you have to characterize whether the identity is correct or

not and where it is relevant and then you have to quantify, okay.

How much of it, either relatively or absolute, okay. So, for protein identification typically what we follow is the bottom up approach, I will come to that. Peptide mass fingerprinting. You digest a protein with some protease, you get into form of small peptides, you trigger an MS/MS of the small peptides again fragmented. So, you have broken down the protein into small-small pieces, collate it together and submit it into it, so bottom up, okay.

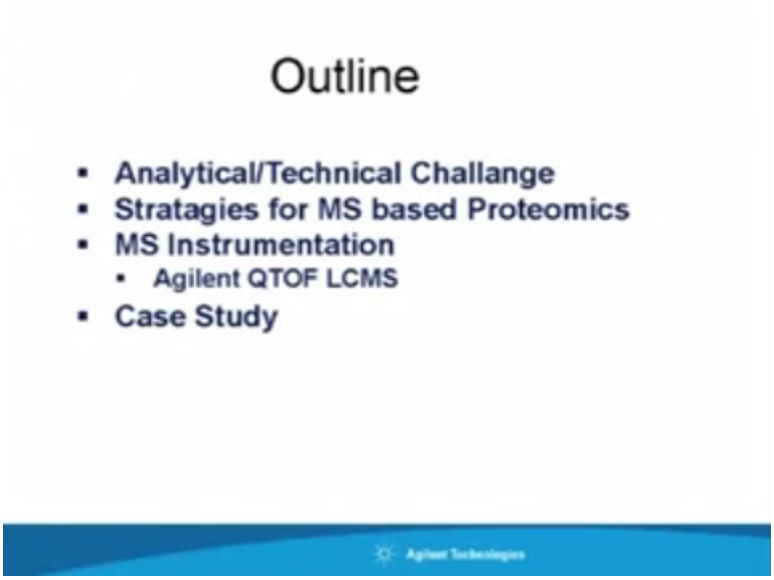
So, you are taking everything together and then submitting into database, so that is how we do it. Then, you get the identity. You have started with some mass number, you have fragmented that another group of mass numbers, you submit into database and get some name what exactly it is, okay. That is what it is.

(Refer Slide Time: 03:29)



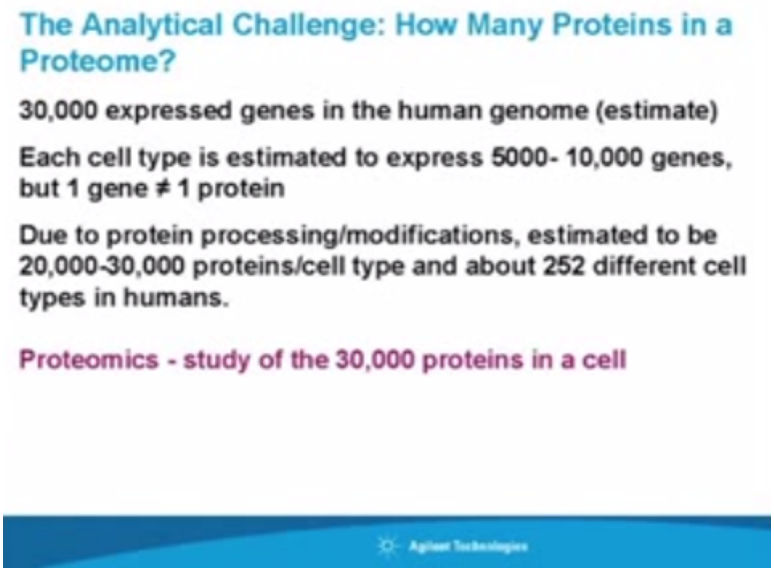
The second characterization is you sequence it; you identify or characterize what are the modifications. The protein can be native form, it can be modified forms and what are the sites, where it is modified. So, that is what the characterization and then interaction and then quantification, either by doing absolute using iTRAQ sort of technology or label free, differential.

(Refer Slide Time: 03:52)



So, what is the analytical challenge. When we say proteomics we have to do, what is the challenge and then how to address that challenge.

(Refer Slide Time: 03:59)

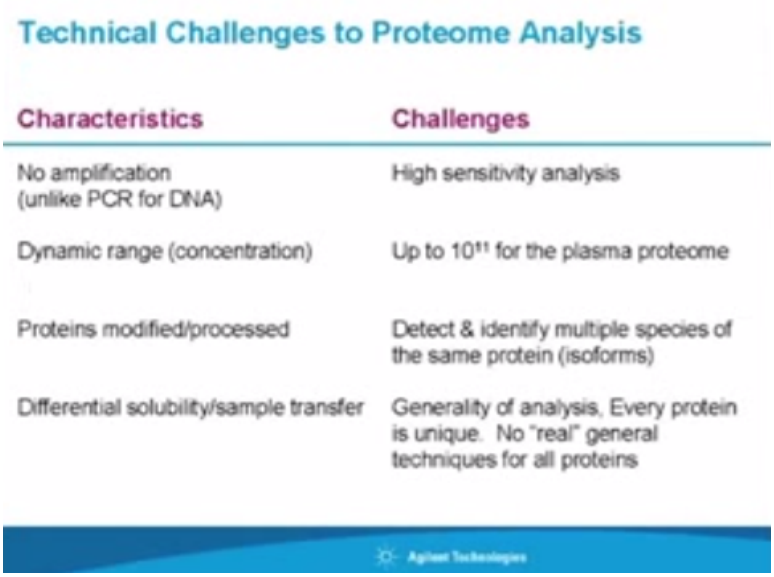


So, the challenge remains the same of what we discussed in the previous presentation also, like there are almost close to 30,000. If we say human, 30,000 expressed genes in human genome and each cell type is estimated to express almost 10,000 genes but not 1 gene is equivalent to 1 protein. 1 gene can express many proteins also. Due to protein processing modifications estimated to be 20,000 to 30,000 proteins per cell type.

There are about close to 250 different cell types. So, you can see the dynamics and the

complexity slowly-slowly it is increasing. So, that is challenge. So, you have to actually typically identify close to 30,000 proteins in one cell. So, you can image, there are 250. So, it is quite complex. How to address that now.

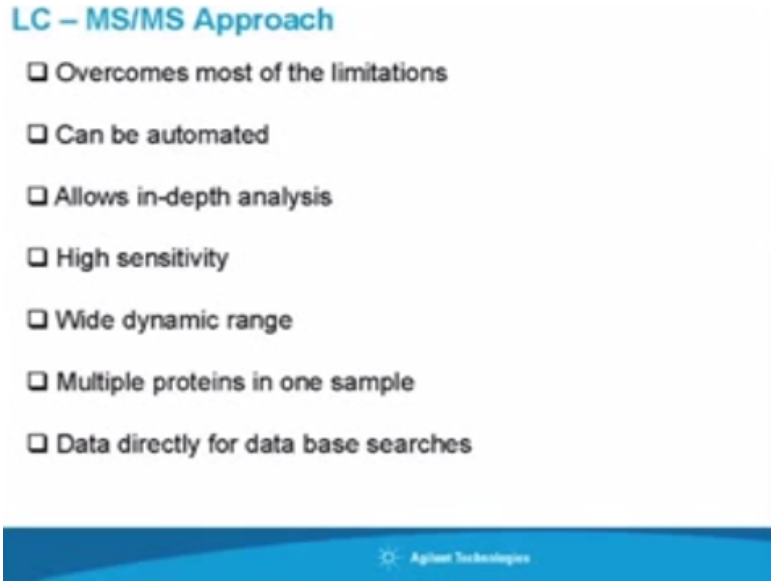
(Refer Slide Time: 04:44)



With comparison to genomic if we say, you cannot amplify proteins, unlike your genes DNA or RNA, you can amplify but you cannot do it. So, you need a high sensitivity analysis. The dynamic range concentration of protein can be having various concentrations. There can be a low abundant protein, there can be a high abundant protein at the same place, same point of time, okay. So, you must have to address this aspect of while analysis.

While following an analytical approach, you must address this aspect also. This is one of our criteria to address this. This is one of the biggest challenge. So, protein modified and processed. So, detect and identify multiple species of same protein. Same protein, various species, different isomass, different modifications, all those things, okay and differential solubility. So, generality of analysis, you cannot generalize for this type of tissue samples. This is the method, it may not be true for other type of analysis, okay. So, that is how it is complex.

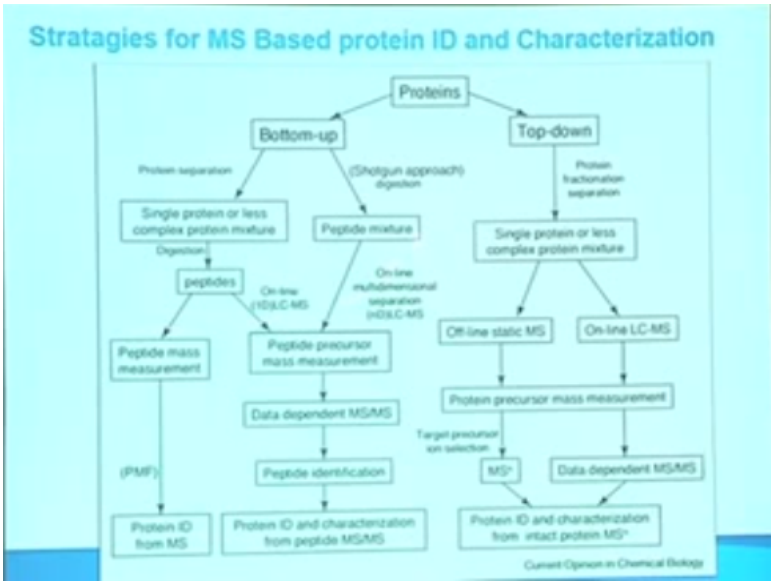
(Refer Slide Time: 05:47)



So, to overcome that we follow the LC-MS/MS which reasonably addresses most of the challenges, okay because it can be first automated. You can do multiple runs and you can automate it. You need not to sit there and do it. You run overnight and morning at least you can compare results and see like what is happening. Allows in-depth analysis, it is highly sensitive, it has wide dynamic range, okay.

And you can run multiple proteins in one sample at a time. It is in a way compatible to database searches like data as such can be submitted into database and you can get the identity or notations, okay.

(Refer Slide Time: 06:29)



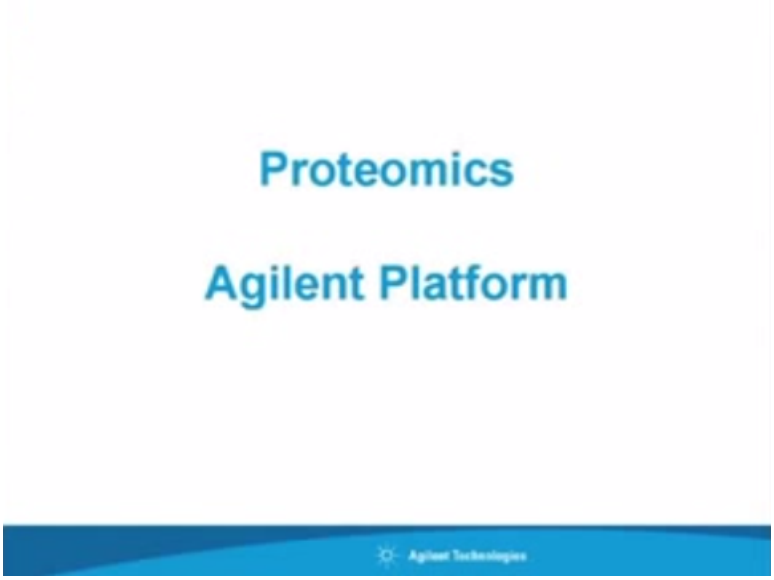
So, there are different strategies of MS based proteomics as we talk this bottom up, you digest. This is most commonly used in (()) (06:37) "voice not clear" type of approach where you have the protein or mixture of protein, you digest it, fragment it, do an MS/MS, hybrid type of mass analyser, okay. When we say hybrid means quadrupole time of flight let us say coupled with collision cell where you can do an MS/MS and get the peptide accurate mass and then you identify the protein, okay.

The other one is topped down where you fragment protein at the intact level itself and get the fragments at intact level itself and then try to identify, that is another approach.

(Refer Slide Time: 07:12)

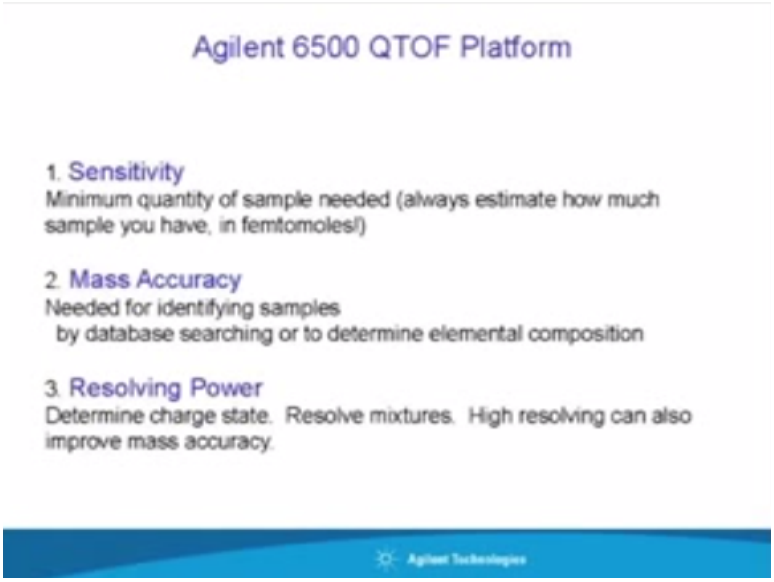
The most popular platforms are either TOF or Tandem TOFs. When I say tandem means quadrupole coupled with time of flight, so quadrupole time of flight tandem. Then are ion traps where you have MS/MS capability, hybrids, okay. Then, magnetic sector. So, there are various types of popular platforms.

(Refer Slide Time: 07:30)



I will just concentrate on something which we have in Agilent portfolio which is being used for proteomics which is quadrupole time of flight, Q-TOF type of instruments.

(Refer Slide Time: 07:38)



Why we use Q-TOF even though we have a triple quad also, but we still say like for proteomics we need a time of flight type of instruments why so, because of three reasons why we need time of flight; one is sensitivity as we understood like there is a challenge that we need to have sensitivity. Then, there are two things which is very critical in the time of flight which probably a quadrupole type of analysers might miss is the mass accuracy and resolution.

(Refer Slide Time: 08:10)

So, typically like in a quadruple time of flight instrument, you have a quadrupole and then a time of flight where the ions will travel like this and time is measured coupled with collision cell where you trigger an MS/MS. When we say what is the word we typically use, LC-MS/MS, okay. So, there is a LC component, liquid chromatography component also coupled to an MS, okay. Why so, because we want to separate.

What exactly liquid chromatography does. It has a stationary phase, reverse phase typically and use separate peptide based on their chemistry, hydrophobicity in this case or different-different chemistries for different-different chromatography, okay. So, you separate your peptides, the outlet of an LC is what liquid but we are saying here that my mass spec cannot see mass as such. It has to be charged. Meaning the language which my LC is talking cannot be understood by mass.

So, we need an interface which is called an ion source which we just now see or in this case electrodeionization that is an interface where the liquid is evaporated and only charged (()) (09:24) "voice not clear" in this case peptides gets into the mass spec which is under high vacuum, okay and it travels through the quadrupole. Then, you can trigger an MS/MS and then it can be measured in the time of flight, okay.

Why quadrupole and why time of flight, maybe we will discuss when we actually go into the lab.

You can see it, like there are various modes and then you can realize okay why not only TOF. If TOF can do everything as I told, okay it is accurate mass resolution, then why we need a quadrupole, okay. So, we will see that.

(Refer Slide Time: 09:58)

This is the instrument something like what we have seen in schematics, quadrupole time of flight type of instrument where you have quadrupole as I showed you in the presentation, then you have origin cell, okay and then we have a time of flight, okay. So, quadrupole and time of flight are in tandem, okay. Being in tandem with respect to (()) (11:44) "voice not clear" MS/MS, okay. So, there are two different analysers; quadrupole and time of flight, okay and that is coupled to HPLC (()) (11:52) "voice not clear."

As you know there are thousands of proteins in (()) (11:58 – 12:04 "voice not clear") you digest the protein (()) (12:06) okay and then you do an MS/MS, although things (()) (12:10) "voice not clear" So, you have digested the protein and thousands of proteins becomes ten thousands of peptides and I just subjected to MS. MS is running which is scanning, right. It will have some limitation. You can see up to this, not beyond it or something like that just to make you assume, okay.

It will have some limitation. So, 10,000 peptides coming at a time, it might not see. So, it might see some of them, it might miss most of them (()) (12:43) "voice not clear". To reduce the

complexity, we will use a separation technique before an MS so that your peptide get written based on their hydrophobicity in the stationary phase which is we call it as HPLC (()) (13:02 – 13:08) “voice not clear”

It has some stationary phase particularly on (()) (13:10) “voice not clear” it depends upon what type of (()) (13:13 - 13:24) “voice not clear” or you both the peptides at the same time intervals. So, 10,000 peptides, 2000 will come at 1 units, next 500 will come at 2 units, like that. Then you split all those things and then (()) (13:35) “voice not clear” on your MS, you let us (()) (13:38) “voice not clear” with its full capacity and you will not miss it that is one. You reduce the complexity also.

So, as a HPLC (()) (13:47 – 13:59) “voice not clear” basically two pumps. The purpose of two pumps is to pump the mobile force and (()) (14:05) “voice not clear” in case of peptides it is hydrophobic interaction, okay and if there is no hydrophobicity, it will (()) (14:14) “voice not clear”, okay. We have pump. Then, we have a column compartment where you put the column, basically it is heated (()) (14:25) “voice not clear”

Then, you have other sample. If you recollect (()) (14:32) “voice not clear” it can be automated, okay. So, this is automated. You can put hundreds of sample in there, you define your (()) (14:40) and you in turn carry out your other works and by the time your analysis gets over automated, okay. What I mean to say is if you like put 100 samples and you have run time of the separation, time taken to separate is 13 minutes (()) (14:56).

By the time the number of 100 samples (()) (15:00) “voice not clear” into 300 units, is not or more, is it more than that. What it means, samples might degrade by the time its number come, okay. So, maintain that into constant temperature, you have (()) (15:15) “voice not clear”. So, you can maintain the temperature of the sample in the auto sampler at 4 degrees (()) (15:19 – 15:24) “voice not clear” if you are actually running for (()) (15:26) “voice not clear” and then there is a (()) (15:28) analysis, okay.

So, typically there are two pumps as I told you in the pumps (()) (15:36) “voice not clear”. Pump

A consists of water with (()) (15:40) “voice not clear” okay and pump B consists of (()) (15:42 – 15:47) “voice not clear”, acid to give proteome, okay. Why so, because I know my analytes are peptides which are having many (()) (15:57) “voice not clear” groups which can take proteome and gets charged and I know the basis of (()) (16:03) “voice not clear” I need charge, unless my analytes is ionized, this is blank, it cannot see, okay.

The (()) (16:14) “voice not clear” that you analyse must get charged, okay and (()) (16:20) “voice not clear” your analytes gets ionized with solution, meaning in the solution itself it gets protonated and (()) (16:28) “voice not clear” facilitate the ion. It creates an environment where your ions move into MS suitably broadly I would. The pumps will pump the mobile phase which is (()) (16:42) “voice not clear” let us say.

So, while loading the sample, we will use more of (()) (16:49) “voice not clear” okay, so that your peptide gets retained on the column stationary phase. If this is my stationary phase depending upon hydrophobicity, 10,000 peptides you will find here, you will find there and delicate split across the column, okay. There is no such hydrophobicity and this get retained and you can run for a while with aqua only, so that any salt or other things get washed out and only you get enrichment of your peptides.

Then what you do, you slowly, slowly increase (()) (17:18) “voice not clear” which is also called as organic (()) (17:22) increase slowly-slowly the organic content and reduce the aqueous content, so that the polarity is now changing and depending upon the binding capacity, they will leave off (()) (17:35) “voice not clear”. It means I can measure the peptide equal to the (()) (17:46). Why we need MS, okay. Anyways, it is retaining, everything it is also (()) (17:54) “voice not clear”, then why need (()) (17:57 – 18:04) “voice not clear” what is need of an HPLC.

(()) (18:07) “voice not clear” So, you get the (()) (18:08) “voice not clear” how you see chromatography. I will ask you a question. If give you a chromatograph of an HPLC will you, if there are six gates, what it means. (()) (18:23) “voice not clear” fixed compound. You get two dimension to change your analysis, one is compound, another one is essential time, okay. Based on compound, you can do that, but will it serve my purpose.

Because in peptide (()) (18:36) “voice not clear” protein, okay and how many peptides and different peptides from different (()) (18:48) “voice not clear” that is also possible, okay. So, need another dimension to see that, to look beyond the (()) (18:53) “voice not clear” So, for that the third dimension given to that is mass spec which is the mass (()) (19:02) “voice not clear” for mass, okay.

So, when one key of HPLC goes in to the mass spec, it sees beyond that 2 dimension which is marked and I will tell you that (()) (19:12) “voice not clear”, there are 100 more masses, okay within that area. So, as I told like it runs to the column and then it dilutes in the form of liquid like this. It is coming like this. So, right now (()) (19:30) “voice not clear” it is running at 0.4 mL per minute. The flow rate is also (()) (19:34) “voice not clear” with this source only.

I am not going into details (()) (19:44) “voice not clear” this flow particularly. The other flow MS in general, these are not be able to (()) (19:51) “voice not clear”. So, right now initially, when I have to load samples, there is a (()) (19:58) “voice not clear” 3% of organic and 91% of (()) (20:05) “voice not clear” because I have to inject sample. So, it goes in through. So, as you see what comes out here is the stream of liquid but what is measures (()) (20:20) “voice not clear” ions.

So, the (()) (20:23) “voice not clear” is not understood by us right now. So, we need some sort of device which converts the liquid into ions. So, that comes (()) (20:33) interface which is also called ion source or atmospheric pressure ionization technique and what type of atmospheric pressure (()) (20:40) “voice not clear” also called as (()) (20:49) “voice not clear” consisted with the growth rate. Why so, it is something called as (()) (20:54).

So, the stream of liquid is going into the needle and then it is sprayed, why so, because of analytes (()) (21:03) “voice not clear”. So, charge particles or charge analytes are (()) (21:08). So, I want only those parts to get in, not the other (()) (21:13) “voice not clear” and all those things. I want to (()) (21:17) “voice not clear” stream of liquid. Then, there is nebulisation gas which is nitrogen.

Then, that liquid is sprayed like Baygon spray sort of or thing like that because you are pushing for gas around that level so that it get sprayed off. So, you have to split the flow, split the stream of (()) (21:39) "voice not clear" and source created at typically 300 degree centigrade. So, there is a (()) (21:47) "voice not clear" imagine a droplet of liquid with different charged peptide inside that (()) (21:56) "voice not clear" it came out into that high temperature (()) (22:00) will get evaporated slowly, slowly.

The positive charges keep (()) (22:04) "voice not clear". The moment it gets closer, they will repel, they will explode (()) (22:07) "voice not clear" that is a phenomenon in (()) (22:11) "voice not clear" explosion and when the explosion (()) (22:15) "voice not clear" and then gets evaporated, dissolved because (()) (22:22 – 22:27) "voice not clear" otherwise we should not react of something that is (()) (22:30) "voice not clear" the droplet and you see the split. (()) (22:39 – 22:46) "voice not clear".

You see, this is how it is done. Now, one drop has given like this spray and exact at high temperature 300. You can imagine how (()) (22:54) "voice not clear". So, there are different analyser. Quadrupole is one type of mass analyser. So, quadrupole is also called a mass filter analyser because the way it behaves. Quadrupole will form four dots, quadrupole. So, four dots like this. There are four dots like this at some length, okay.

So, what is the logic. It is also called a mass filter why, because it has the ability to filter mass. If you say look for marker 500, it will filter everything away and only allow 500 to get inside the quadrupole and travel through there. Travel (()) (23:51) "voice not clear". These four rods are at different electronic state, like (()) (23:56) "voice not clear" holder. For example, this rod and this rod, alternate rods (()) (24:00) "voice not clear" 500, okay.

So, as a user I will say only 500 but for electronics to understand how much electrical energy it needs to attract mass of 500. So, it will apply that. The 500 ion will come here along with 10,000 other ions, okay. But it has applied efficient energy to capture all the 500. So, what is the outcome. Out of 10,000, maybe 9,000 gets lost, will hit somewhere and get lost. Along with 500,

most ions will (()) (24:35) “voice not clear” come out because of the voltage.

The moment it comes close, the (()) (24:42) “voice not clear”. So, now this instead of attract it will recur 500 and this will attract. It will travel like this. The moment it (()) (24:52) “voice not clear” during this probably the 5000 will get hit here and then get lost. Then, again (()) (24:58) “voice not clear” then it comes here. So, like that it will travel in the quadrupole selectively and other 10,000 will hit somewhere at the other and get lost.

So, in this way, you can filter mass with very-very (()) (25:13) “voice not clear”. Let us understand in general the functionality of why we use this and then applicability and these are the reason specifically because you have experience that is why you are asking. (()) (25:24) “voice not clear” that but most of you may not have experience but at least you should know what exactly quadrupole will do.

So, quadrupole is a mass filter. It is very specific because of this, it is selective and sensitive because that is why it is used for targeted analysis. You have heard the name for triple quad. Quadrupole only (()) (25:49) another quadrupole. Instead of time of flight, you have to put another quadrupole, it becomes triple quad. When you say two mass analysers is a functionality of having MS/MS in the form of follicle cell, you call this as in tandem, tandem MS.

Because they are talking to each other, they are in tandem with MS/MS in between. If there are two different analyser like here in this case quadrupole and time of flight, then it is also called as hybrid mass spec because quadrupole in itself is the mass analyser. Time of flight is mass analyser. Two different mass analyser is coupled with (()) (26:28) “voice not clear” MS/MS, it becomes hybrid mass analyser or hybrid tandem MS.

Because in this case quadrupole and (()) (26:35) “voice not clear” are also talking to each other with (()) (26:37) “voice not clear” also called as data dependent, data dependent (()) (26:42) “voice not clear”. Now, we need to have two different analysers, okay. So, in our case in unknown scenario, we know that there are different types of protein and I do not know about the protein. So, in first place, I want to establish the identity of that protein.

So, I need accurate mass information. I cannot (()) (27:03) “voice not clear” device which has unit marker resolution. It cannot see beyond (()) (27:09) “voice not clear”, so hence it is not accurate. So, if I (()) (27:14) “voice not clear” data to the database, there will be many (()) (27:17) “voice not clear” of the same mass but this is sequence. It will pull out all the information because it does not have accuracy.

It will see as 500 and 501 only and you will get hand in hand information. For the same mass, we did hundreds of (()) (27:32) “voice not clear”. You started with nowhere and you end up nowhere. So what to do next. I do not know out of just 50 peptide what is correct but the moment you put time of flight, 500 becomes 500.1, 2, 3, 4, 5 and the fragment of 500.1, 2, 3, 4, 5 becomes 10 different fragment of accurate mass.

So, it is absolutely unlikely that a peptide of that sequence will be giving some different MS/MS data with different masses. So, the (()) (28:04) “voice not clear” absolutely perfect. You started with nowhere, you ended somewhere which is reliable, that is why you need time of flight, okay. So, now as an applicable (()) (28:19) apply that. So, broadly the Q-TOF can be run in three different modes, okay.

Modes of applications if I say so, in terms of applications. First is TOF only mode. So, as an applicability, we can run in 50 different modes. First is the TOF only or MS only mode whereas the MS only means quadrupole, collision cell is not behaving as they have to, meaning quadrupole will (()) (28:48) “voice not clear” and collision cell will not dry.

Because ions are going in, they simply pass through quadrupole, simply pass through the collision cell and go to the TOF and TOF will scan. The (()) (29:03) “voice not clear” based on only mass. You got the accurate masses from all the peptide, then you (()) (29:10) “voice not clear” that is one way. Nowadays, mostly none of (()) (29:15) “voice not clear”, okay. Based on only accurate mass, (()) (29:22) “voice not clear”. Then comes the other in this order data dependent approach.

There we take advantage of both (()) (29:34) “voice not clear” quadrupole as well as time of flight. I have some sample (()) (29:38) “voice not clear”, I do not know how many peptides, which peptide is coming at what time and what (()) (29:44) “voice not clear”. I have no clue. I cannot guide my quadrupole to select what and when (()) (29:50) “voice not clear”, so no clue. So, let instrument decide. Depending upon the data as it comes reassign, decides to fragment.

The decision engine of the time of flight will decide what to fragment at what time depending upon the inputs given by the user. So, user will define that look what peptide (()) (30:15) “voice not clear”. Look for only those peptide with some certain (()) (30:20) “voice not clear”, okay. So, based on that it will look for that. In MS only mode let us say, this machine has the capacity to scan (()) (30:32) “voice not clear” that is the range of the machine.

But I know my peptide is (()) (30:42) “voice not clear” cannot go beyond 1200, maximum 1200, okay. So, if my TOF is scanning one spectrum per second, means my cycle time is one second, means it will start from 5. It will come to 3200 and come back. It happens in one second. It will take one second time to scan complete whatever I have given and I know my peptide will not come beyond 1200, meaning I am wasting some time in scanning something which is useless.

Instead of that, I can spend the full time to capture something which is informative. So, I take advantage of quadrupole there to say look for mass from 200 to 1200 only. So, it will filter others as it allowed 1200 to 1200 and that once (()) (31:35) “voice not clear” is applied till 1200 and come back, that is the scan (()) (31:41) “voice not clear”. In data dependent definition the same scan time is spread into two different experiments, experiment 1 is scan first, let everything go into the TOF and in this scan pick up the precursors.

There is (()) (31:57) “voice not clear”. It is based depending upon the data which is coming, I mean threshold (()) (31:04) “voice not clear”, all those filters in place. If it qualifies, let us say 10,000 ions are scanned, out of that 10,000 only 8 or 10 is relevant (()) (32:17). It will pick up both and it will tell the TOF (()) (32:20) “voice not clear” to quadrupole in the same cycle filter only those (()) (32:25) “voice not clear”, okay, that is why it in tandem, it is talking.

It has scanned. It has told that (()) (32:32) “voice not clear” while coming back it will pick up only those ways in the same cycle, fragment those and give it there. So, now the TOF will stitch the information because it has the scanning, it has the (()) (32:44) “voice not clear”. So, it knows that we have the precursor at MS level and it knows what are the fragments of that precursor. So, it will stitch the data, so you will have an MS and MS/MS (()) (32:53) “voice not clear” stitched together and as a user I can see it comfortably on my desktop.

The third one is data directed or target (()) (33:02) “voice not clear” which is possible because we have quadrupole time of flight. When I say targeted means you have to define what are the masses you have fill it. It will trigger MS/MS and give it to you. What it means, it means a quadrupole time of flight device can be used for quantification, okay. Of course, it cannot be that sensitive as the triple quad because of the same reason that the third analyser is scanning, it is not a mass filter like triple quad.

Since it is scanning, it is wasting time. It is scanning something which is not relevant and hence you may lose something, that is why it is less sensitive. It is not that sensitive as a quadrupole.