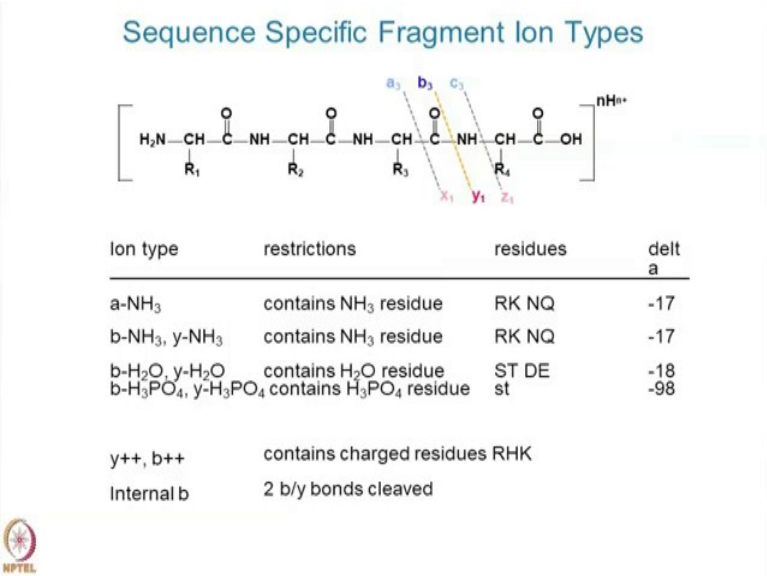
An Introduction to Proteogenomics
Dr. Sanjeeva Srivastava
Dr. Karl Clauser
Department of Biosciences and Bioengineering
Indian Institute of Technology, Bombay
Broad Institute of MIT and Harvard

Lecture - 16
Introduction to Mass Spectrometry based Proteomics – Part III

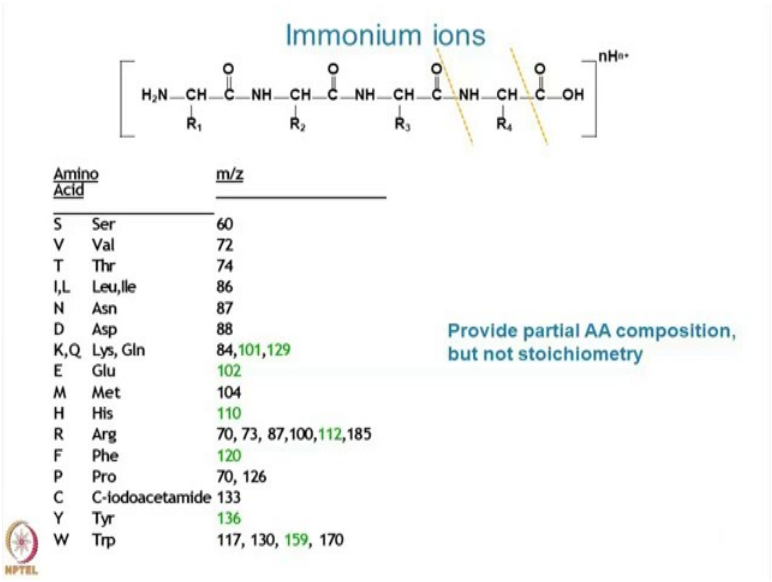
Welcome to MOOC course on Introduction to Proteogenomics. Interpretation of mass spectrometry data is very challenging; there are many softwares which are commercially available or open source softwares which make our job relatively very easy. However, the basic knowledge of mass spectrometry data and its interpretation becomes very crucial. In the last lecture, you were introduced to the basic concepts of mass spectrometry and sample preparation. Today, we have hands on session which aims to provide some basic knowledge required for the manual interpretation of Mass Spectrometry data. Let us welcome Dr. Karl Clauser for today's hands on session.

(Refer Slide Time: 01:21)



Now, other than to say we are going to try to assign fragment ion types and these are the things you should keep in mind for the various fragment ion types.

(Refer Slide Time: 01:28)

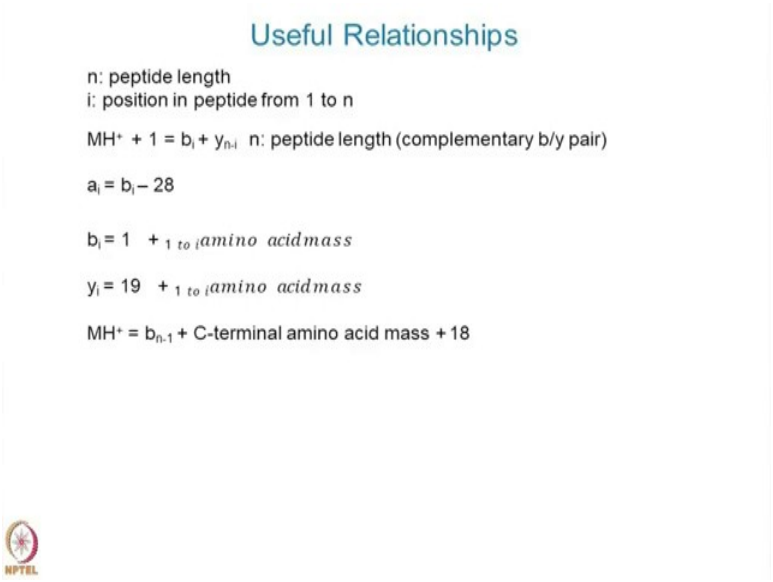


One of the things I did not mention so, far is that there are Immonium ions which give you composition information, but they do not give you stoichiometry. So, if you see a 120 ion in your spectrum that is a good indication that you have phenylalanine, but if you have a really tall 120; it does not mean you have 2 or 3 of them. So, you all you do not have the stoichiometry information. In order to form an Immonium ion, the fragmentation has to occur in two places along the peptide backbone.

And, it includes the side chain that is unique to the amino acid alright, by the techniques that were going to be using in the spectra that are being shown, the Immonium ions that are likely to be present are shown in green. The ones that are shown in black are essentially not; you can only form those if you have a high energy collision type of instrument which is not the same as saying it HCD alright. So, actually from the terminology standpoint that is a distinction between the manufacturers. So, if you have a Q-ToF instrument, the they comes safe from Sciex or from Agilent or Bruker, they might call that fragmentation technique CID.

That fragmentation is exactly the same thing as with thermo calls HCD and the reason is that before thermo created the current instruments, they had ion trap fragmentation that was resonance excitation and they called that CID alright. So, they had to have another name for, did it the same way as other people; ok ,right next, ok.

(Refer Slide Time: 03:21)



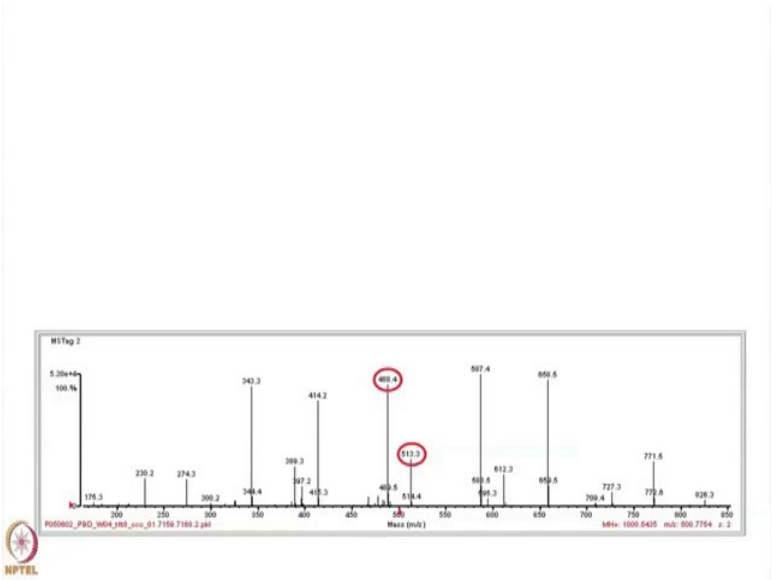
These are the amino acids, if you knew these by hand and by heart it would be easier for you to look at the mass differences, if you had the opportunity to print them that would be helpful. So, these are some of things that is particularly helpful for you to know, that we will make use of in interpreting the spectra ok. So, an a ion is 28 ions less than b ion and if you can decide whether an ion is an a or b or y and if it has something that is 28 Daltons less, then you can probably call it b that is why it is useful alright. If you know your amino acid masses and you just add them all up and add 1 that will give you the mass of the b ion.

If you add up all your amino acids and add 19 that will give you the mass of the y ion correct and then if you fragment between two amino acids b ion is going to go one way, y ion is go to the other. If you add them back together, you would have the whole peptide right. But, when you fragment; so, I am always going to talk about the parent mass as being singly charged ok, the b ion is singly charged, the y ion is singly charged. So, if you add the two together, you get the MH^+ +1 and those are complementary ions is showing up here. So, b ion plus y ion is the same thing as MH^+ +1. Some software packages they will strip out the charge, I think that is silly because we measure ions, we do not measure masses right.

And, and furthermore if you when you are dealing with high accuracy masses, you have to trust that, they have built to subtract the right amount because it is not 1.0, it is 1.0078

minus the mass of electron and these things matter. So, I think it is easier just to think in terms of ions alright. So, I keep the charge alright, those are the useful things to know. What kind of spectrum would you call this?

(Refer Slide Time: 05:43)



This will be on the quiz, come on.

Student: What do you mean by that?.

If you were to give it a cute little name, what kinda name?

You want to call it a white picket fence; I was going to call it a Royal Enfield. So, this is a relatively beautiful spectrum, it is going to have you can see its got ions, that span the mass range, they look nicely spaced.

Student: Doubly charged ions.

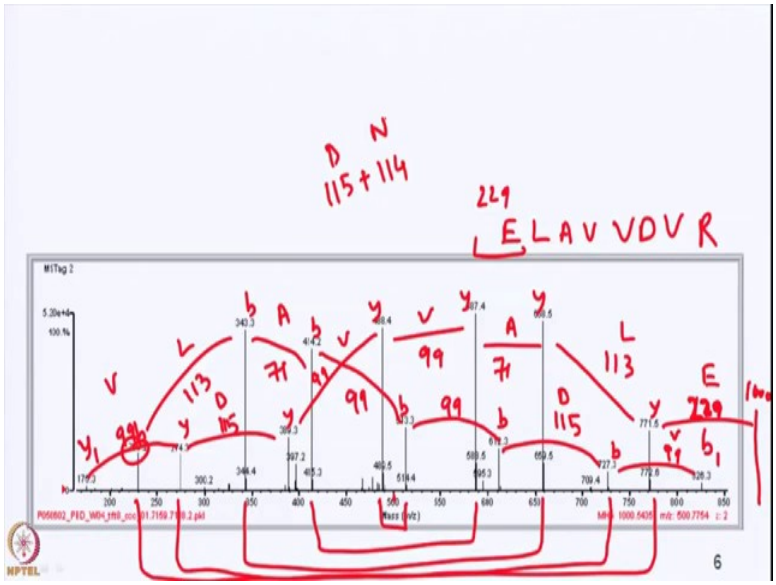
Ok.

Alright. So, before you start doing.

Some math so, the MH plus is a 1000 ok, this right here is at 500 ok. So, if you look, you will see some nice symmetry lysine has a y 1 ion that is 147 ok. Arginine has a y 1 ion that is 175; you should now be able to tell me what the c terminal amino acid of this peptide is; if you are looking at the spectrum alright. So, if this is the where the precursor

mass is and I you can see that there is some symmetry right. So, these two peaks are about equidistant from the center and that so that means, they are probably b y pair ok.

(Refer Slide Time: 07:08)



So, I am just going to do this kind of thing here, then I bet you if you look.

That were the red where the red carrot is.

Correct alright and yet so, the other thing you can do to check, if you cannot really tell the symmetry you can add up the masses, 414 plus 587 should be equal to a 1001.

Student: Ok.

Alright. So, similarly we can go like this ok, those are going to be symmetric; I already told you that if it is a tryptic peptide, y 1 is 175 which would be consistent with the being an arginine at the c terminus right ok. And, if you that y 1 would be symmetric with 826; that means, this should be a b ion ok. Then there is probably a couple others in here so, 230 going all the way.

To 771 ok. So, we can see there is lots of symmetry presents alright, now let us do some arithmetic. My advice whenever you are going to start doing arithmetic is and you have to choose where to start, make one of two choices; start with the big peaks or start it an end.

Ok.

I am going to go with big peaks, but and when you started an end it is helpful to, it is it is nice to have one that you already have an ion type for. One of the things I try to resist doing is starting at the low end, because there is more often going to be multiple things happening in it below the precursor mass. You can have multiply charge things, there is going to be internal ion fragmentation, you usually have a cleaner spectrum up at high mass alright. So now, let us just do some arithmetic. So, this mass gap here is 101, no 58 13 71, this mass gap here is 113.

Then 587 to 513 is 74 that is not an amino acid mass, 488 to 587 is 99 that is an amino acid mass. 488 to 414 is 74 that is not an amino acid mass. Let us go 826 to 520 727 that is 99 ok, 727 to 612 115 ok. 612 to 513 is 99, 513 to 414 is 99 ok , alright, can we go 414 into 353 that is 71 right, 343 down to 230 113. So now, then I would go y 175 to 274 is 99 115 alright, then can I go up here 99 ok. If this is a y ion, this should be a y ion; this should be a y ion, y ion, y ion, y ion, y ion. If this is a b that should be b, b, b, b, b, b ok, valine, oh let us do one other thing.

This is as 100 1000 sorry. Can you read that? That is not a fake peak; we know that that is reasonable because, we measured the precursor mass before we did the MS/MS spectrum. The precursor mass is 1000 alright so, and then if it is 1000, we can go from here with 129 precursor mass is like y ion, it is like the last y ion. So, now we could write down the sequence and let it if we go with y ions, we would write down the sequence backwards ok; arginine already 99 is valine, isoleucine is 113. No wait a minute ok, what leucine is 113, but.

Student: Leucine and Isoleucine.

Yes. So, when you do de novo sequencing usually, if we can tell we just choose leucine. And, I think the simplest reason is that there are 6 codons for leucine and 3 codons for isoleucine. So, if you can tell might was go with the (Refer Time: 13:40) got more cordons right, but at the other day now I can tell that it is leucine rather than isoleucine ok. 115 is aspartic acid, 99 is valine where are we here 71 is alanine, 19. So, there is a bunch of valine is here right, alanine 115 is aspartic acid that is valine and this is glutamic acid 129 is E alright.

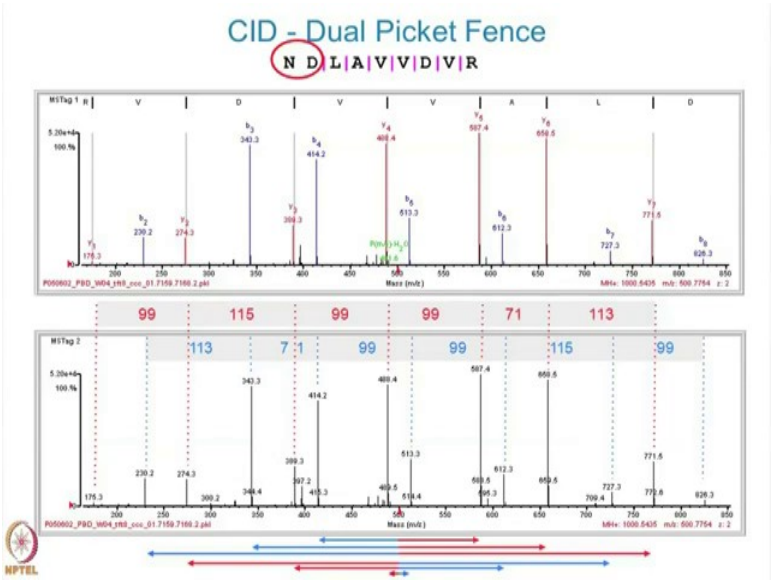
There is a whole lot of junk on the spectrum on the page now, but if I follow it carefully; this is 99 would be valine. And, it would be aspartic acid; it would be another valine and

then another valine and alanine leucine and glutamic acid. And, now if we try to read off the b ions; let us see if we get the same thing. So, we got leucine, no, we have leucine, alanine, valine, another valine, aspartic acid D and then valine and then we would go all the way to here in the arginine. So, it looks like there is something that is not quite right. So, what do we what did we do wrong? So, I said E L A V and it oh is it going to be I got a partly backwards. So, V D V D V V A L E so, the E is not right should be N D.

Student: E N D.

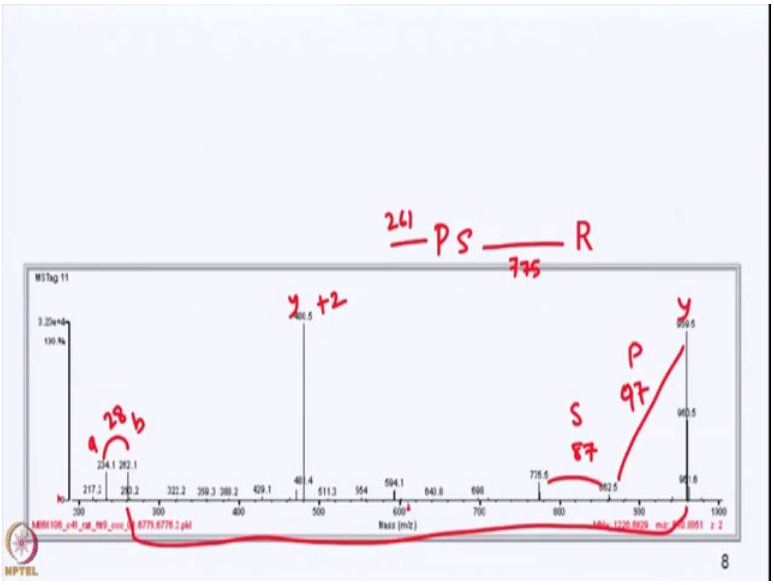
Then why is that alright, the E how did I have the, oh that is so, I can do my bad math ok So, this is this is 229 ok, it is now 129. So, we need right that make that matches our this b 1 ion here ok. So, we have some combination of amino acids that adds up to 229 and there is a couple of ways you can do that, you can do 115 plus 114 and that is D and N.

(Refer Slide Time: 16:45)



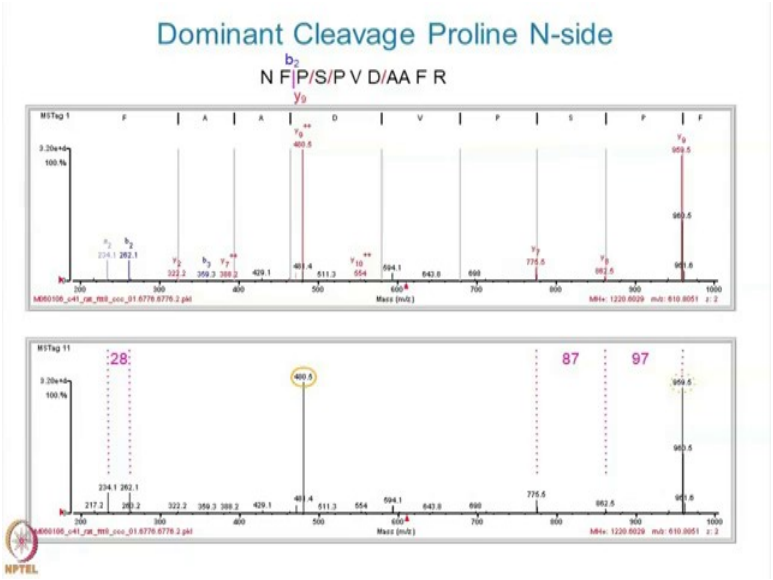
So, if we did now take a look at the way that this spectrum is labeled, we do not have fragmentation between N and D. So, the this is this would start at 230 and then this last gap here is 129. So, this is the sequence is complete with exception, we do not know the order of N and D. So, from a pure *de novo* standpoint we do not know whether it is N D or D N ok. But, if we then add the requirement to this peptide has to come from the human proteome, turns out one of the only one of those two choices is present right and so, it is the N D part alright.

(Refer Slide Time: 17:37)



So, let us go to the next one. So, this one is going to be easy because, there is not very much to work with alright. So, you take a look at the spectrum, there is not a lot of look at the symmetry.

(Refer Slide Time: 17:58)



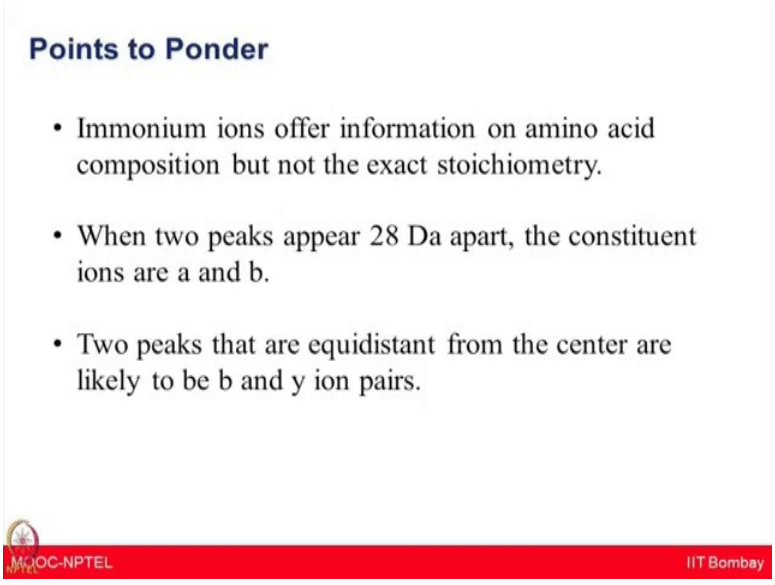
But, there if you look at this is to this is doubly charged 2 times 480 would give you 960 right, there is 2 charges on this, there is only 1 charge on this alright. What else have we got? This is 1220, is there any symmetry here? 262 to 959 that should be a b y pair I think right, 221 and then the mass difference between these two is 26 is 28, alright. What

is the ion type of the peak at 262? This was the one of the things I gave you of useful information, that the difference between a b ion and a ion is 28.

So, we could assign that as a and then as a b ion, if this one is a b and then has to be y. What else can we do here? So, then that would mean that this is y doubly charged, then if we take a look at this mass gap that is 97 and that mass difference 25 is 87 ok, that is serine, proline alright; alright there is not all lot else here. So, we do not have down to low mass ok. But if we so, at some point there is some sequence here that is P S and then that the distance here is something that adds up to 261. And, then the distance out to here would be 775, this is this one comes from a peptide experiment and it.

So, there should be either an arginine or a lysine here ok that is about all the information there is in a spectrum. There is the P S part, that is about all the information there was, that information all together is enough to produce that one answer as a peptide from the human proteome.

(Refer Slide Time: 21:03)



In today's hands on session, you were introduced to the fine points of mass spectrometry based data interpretation. The spectra they are having all the information, but how to get the most relevant information? Think about if you are locating even the post translational modification such as phosphorylation or even glycosylation, the data interpretation becomes very crucial and very challenging. However, in today's hands on session Dr.

Clauser tried to explain you how to get the best interpretation from these complex spectra and even to glean into the PTM based analysis.

As it was discussed in today's lecture, the b ions differ from the a ions by a mass of 28 Daltons. This observation helps in better understanding and differentiating various fragments. In the next session, you will be provided with comprehensive information on interpretation of further more complex spectra.

Thank you.