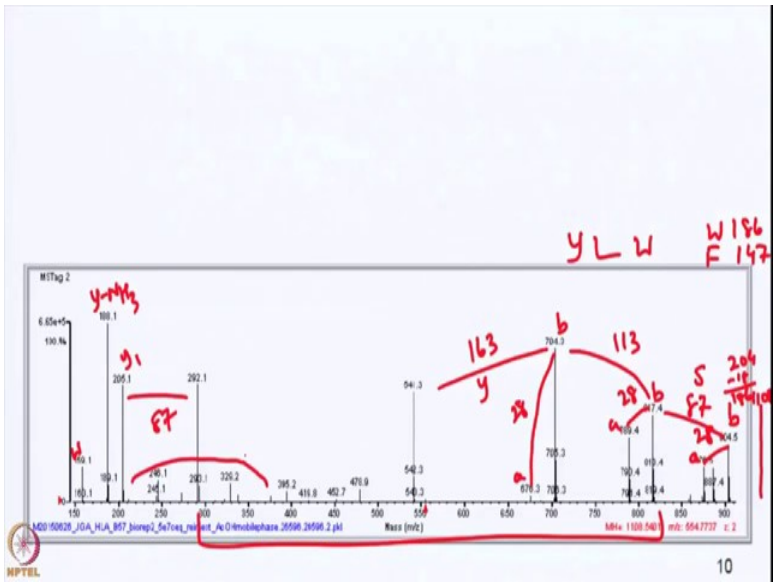
An Introduction to Proteogenomics
Dr. Sanjeeva Srivastava
Dr. Karl Clauser
Department of Biosciences and Bioengineering
Indian Institute of Technology, Bombay
Broad Institute of MIT and Harvard

Lecture - 17
Introduction to Mass Spectrometry based Proteomics – Part IV

Welcome to MOOC course on Introduction to Proteogenomics. In the last hands on session you are introduced to the basics of mass spectrometry data interpretation. This knowledge is very essential while writing the codes for creating softwares, that can analyse the large scale mass spectrometry based data. In today's session Dr. Karl Clauser will show you more complex spectra and make an effort to interpret the data manually. So, let us welcome Dr. Clauser for today's session.

(Refer Slide Time: 01:01)



Alright, let us take a look at the next one there you have here, this one I am going to tell you give you some advance warning that this is not a tryptic peptide ok, this comes from an experiment where were doing immuno peptidomics. So, MHC or what we call HLA peptides: Human Leukocyte Antigen peptides have been isolated. And, this comes from the allele that is B-57 ok, this particular allele has a motif where the c terminus is usually tryptophan or phenylalanine ok. The amino acid mass for tryptophan is 186, phenylalanine is 147 ok, if we add 19 we would get the y 1 ion mass.

So, y 1 if it was a tryptophan would be 205 and if it was a phenylalanine it would be 166 ok. This is consistent here with y 1, if this is tryptophan ok; 159 is also listed on the sheet that I gave you as an immonium ion for tryptophan ok, this is 17 Daltons less. So, that would be y minus ammonia alright, we do not have a whole lot of symmetry here. So, we do not readily have b y pairs do we? Maybe 292 I guess we do have one so, this alright. Then the mass difference between these guys is 28, mass difference between these guys is 28.

So, what ion type is the 904 peak? b or y right, 28 b minus 28 is it a right ok. And, what is the mass gap? Between is 83 87. What amino acid is 87?

Student: Serine .

Serine right? alright, I am going to add an extra peak here, this is 1108 and that mass difference there is 204 minus 18 equal to 186 ok. So, tryptophan is what we say is it that n terminus c terminus sorry alright, this mass gap is 113 so, that would be leucine. What can we do next? This is also 28 right so, it is still b ion and then if we go all the way here that is a long distance, but that is 163.

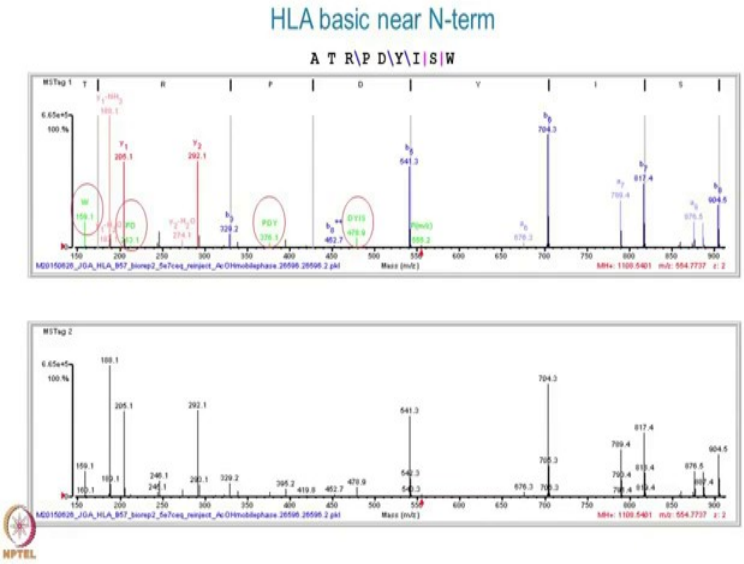
What is 163?

Student: Tyrosine.

Tyrosine.

Tyrosine alright; what else can be? Ok; so, let us see if we keep going the, this was a b-y pair. So, this should probably also be 87 alright so, that much is what I can easily do.

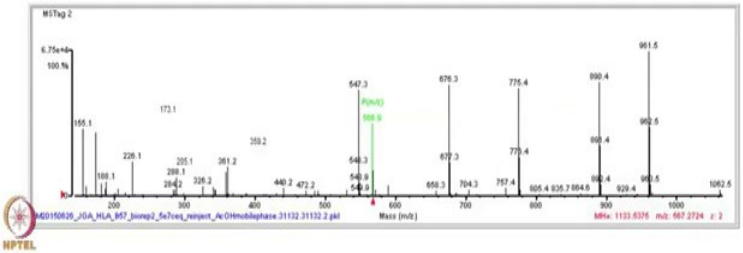
(Refer Slide Time: 05:19)



Let us take a look and see what the best we could expect to do was ok, I said Y L W, this is saying Y S. So, that is correct Y S W and then we have; so, there is longest long enough distances between these things that we do not have individual sequence ok. This b ion here right is in between the R and P ok, some of these other ions here in the spectrum are giving internal ion type fragmentation ok.

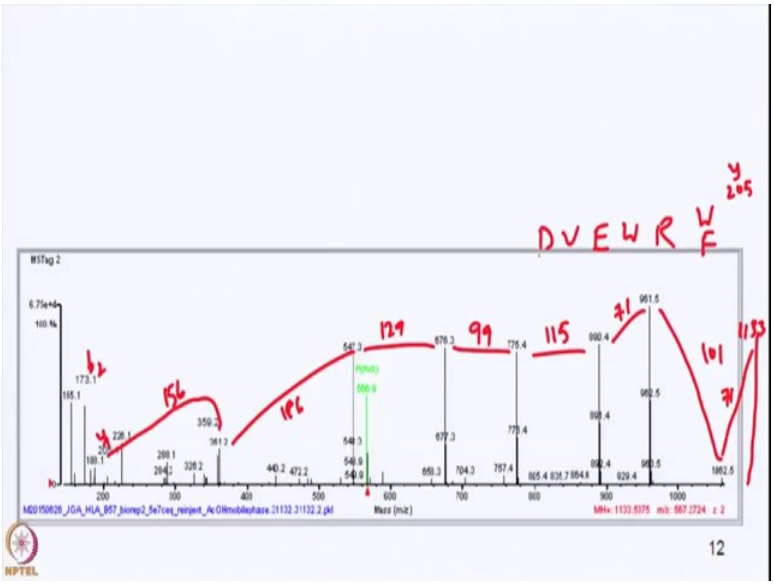
So, this right here saying P D is at 213 and then you can go up to 376 right, that is adding tyrosine, then there is also an ion is fragmenting to give D Y I S. So, that is the extent of sequence that we can easily determine alright, you go on to the next one alright.

(Refer Slide Time: 06:27)



Right so, this one is a good example of why I like to start at high mass, this one is also not a tryptic peptide, I could tell by the file name or it says HLA in it. So, this is again immune and the allele is B 57.

(Refer Slide Time: 06:51)



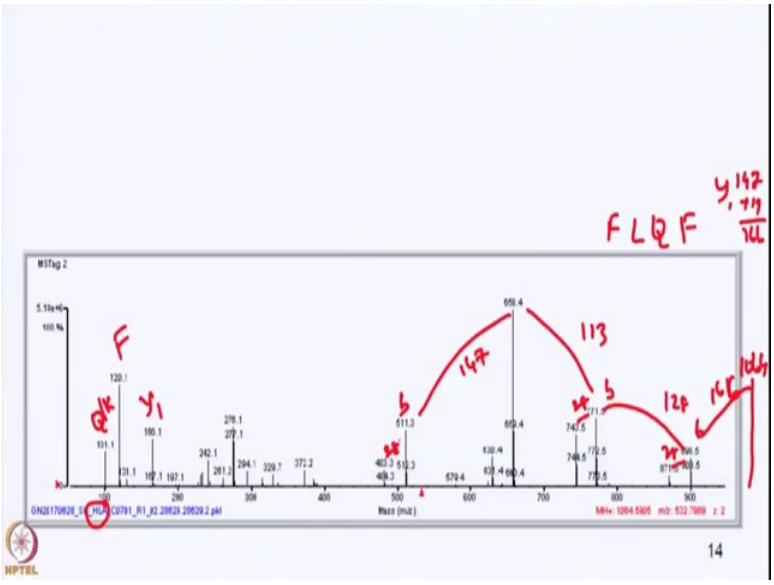
So, it is again something that should end in tryptophan or phenylalanine right here right ok.

Student: Next one is 99.

Ok wait a minute so, that's 71, 99 let us put up this 1133 ok, what can we do this is 101 and 38, 71, ok 547, 53 goes to 600, this is 129 alright. Alright so, we learned this last time that 205, there is y 1 ok so, this is y 1. So, this would be b 2 and that mass difference there is, is that 18 or 28? What is the 18? Ok, b 2 so, that would be the 71 and the 101 and then we should be able to go 71 more or not ok.

So, you want to call that 186 right so, that would be another tryptophan, that would be 156, that would work ok. Arginine, tryptophan, valine, D not enough information to get anything else, alright.

(Refer Slide Time: 09:12)



So, this one is another HLA peptide, but the allele is CO 701, I forget what the c terminus is for that. So, we are going to figure it out. Alright. So, where we start here so, 1064 120 is an immonium ion for phenylalanine, 101 is 4 Q ok, phenylalanine is y 1 ion; should be 147 plus 19 which is 166 ok. So, y 1 for phenylalanine that is a minus 113 gap.

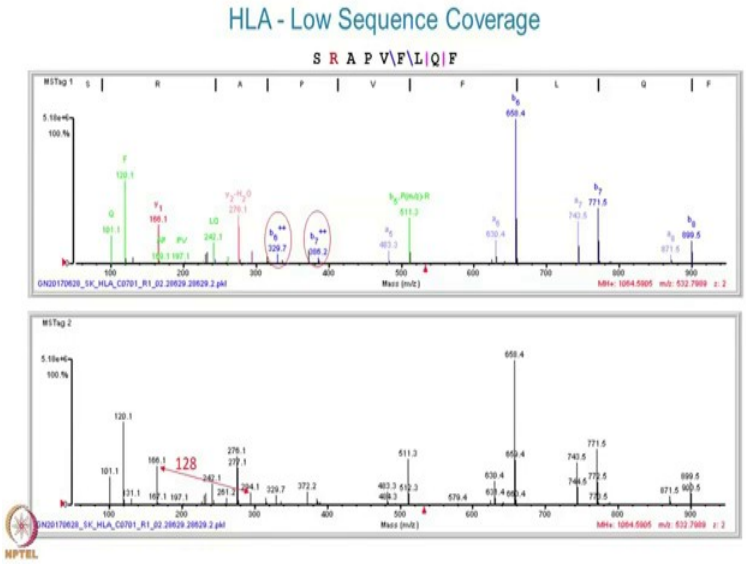
Student: Next to that we have128.

That is 128, yeah.

Student: Sir, before that 147; 658 to 511 is 147.

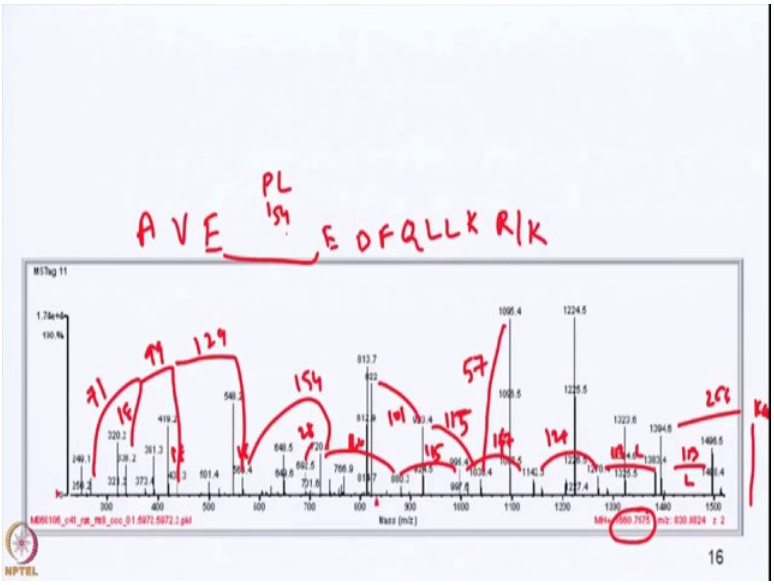
Right here, right there yeah that is 147 that would be that is 161; no, it is 166 which plus 18 right ok. So, again that would be the 147 F, this would be Q, L, F and these would be that would y ion, now that would be b ions. This is 28 right, it is b ion 28, it is b ion alright. What else can we do here? That is again 28 alright; I am going to stop there and let us keep see what the answer says F, L, Q, F ok.

(Refer Slide Time: 12:25)



That is what I had and not enough information to get anything else, the 276 y 2 that is here, this is the 128 ok, these are doubly charged ions here ok. So, this one is not fragmenting very completely alright. So, and then I think pretty sure this is the last one alright.

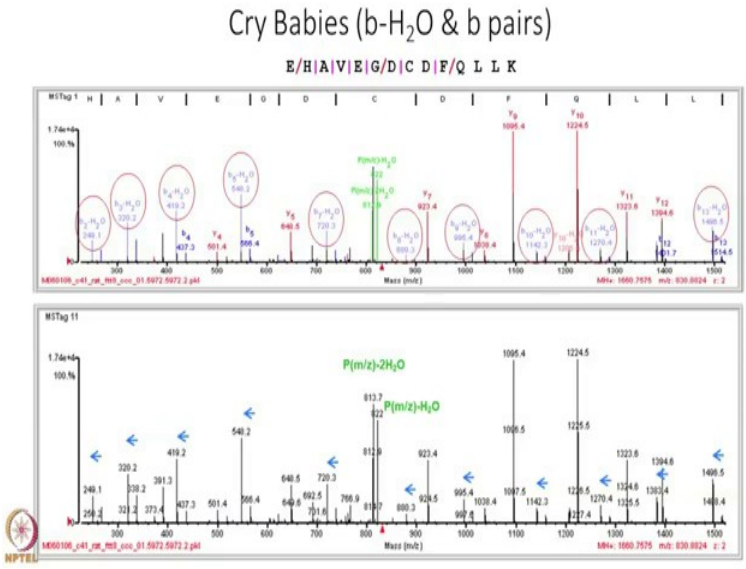
(Refer Slide Time: 13:04)



What is your first impression of that spectrum? Look it looks a lot more complicated than the other ones right ok; so, part of the reason is a little bit longer. So, things are spacings closer together, but there also seems to be bunch of peaks there are close together ok. And, if you look close you could see that these peaks are separated by 18. So, this is 18, 18, 18 that was 28 alright, that is 99, 37, 53, 129.

Do you want 71? This is a tryptic peptide so its arginine or lysine, we do not have down to low mass alright, 266 this is, it is going to be lysine ok. So, I think this is, this lysine, L, L, Q yeah. Well, I did not intend to do that let us just go back here, let us keep going here in 147 F, 566, 566.

(Refer Slide Time: 17:57)



720, it wants to go down to them, alright well pretty close alright that is the end of that. So, hopefully what I have tried to give you is a feel for how incomplete the information is in some of these spectrum ok. Any other questions before we adjourn? Yes.

Student: How do we know that we have band mapping rightly.

How do we know, what?

Student: Sir, it is just like we are taking two peaks finding the distanceif it matches with any amino acid, we are assigning it that.

Yeah.

Student: So, there are inclusions like you were working right now, you had a doubt in your mind whether you were band mapping .

Yeah ok.

Student: So, how do we know that we are on the right path and we are doing it correctly.

Ok fine. So, the way that I have said after I have told you this is all the information is and then I tell you magically, this is the answer; your good question would be how do I know magically it is the answer. Well, the spectrum all by itself has no more information, than really we could that we were able to determine. There is uncertainty in what the sequence is, but if the protein is derived from something in the human proteome, the possibilities that would allow for the uncertainty and our interpretation are not present in the proteome. And so, that is why I am describing it as that is the answer ok.

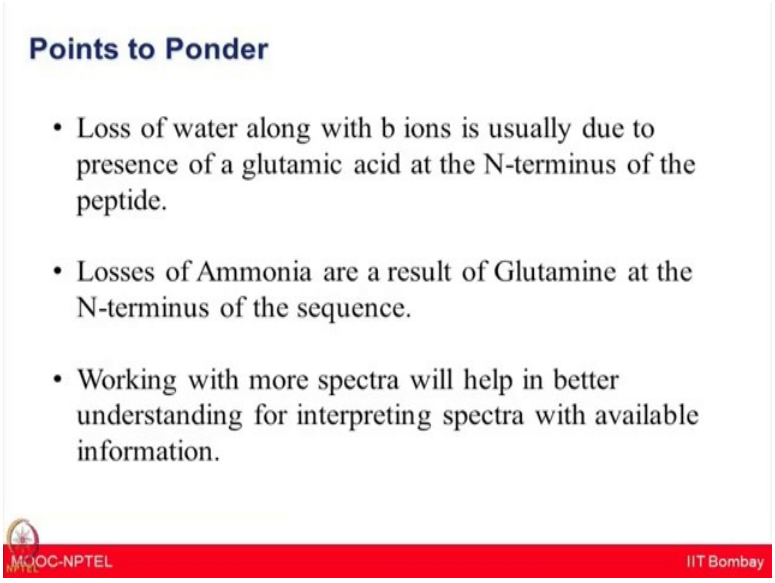
But, let us suppose that after you would had any answer, none of them are in the human proteome; then how do you know what is right? Well, you do not ok; if you have a lot of spectra that you could do good de novo on and they are not in the sequence database, you are looking in; maybe you are looking in the wrong place right. Or, maybe your database is not complete ok, there could be mutations, there could be parts of the genome that are not its part of your database.

Maybe your sample has contaminated with something else, Now there is another aspect to this that I know, but you do not is that; when you get this loss of water happening like this, that often happens when you have a glutamic acid at the N- terminus of the peptide ok. You can get losses of ammonia that look like this when you have a glutamine at the N terminus of a peptide. So, there are you know there is certain aspects of the amino acids have chemistry associated with them, that after you see enough examples starts to become more rule than exception ok. So, there is chemistry behind these things.

Student: What is slash over there?

The slash is I just use nomenclature that indicate what kind of fragmentation is happening between the residues. I use a red slash to be y ion, a blue slash to b ion, a pink vertical bar means there is both, that is just I that is not a universal thing within the field right. But, it allows you when you do a lot of this to just look at the thing and know what is going on.

(Refer Slide Time: 21:45)



I hope today's hands on session helped you to appreciate the complexity of mass spectrometry data resulting from either improper fragmentation or presence of amino acids like proline that creates problem or release of molecules such as water or ammonia due to the presence of amino acids, such as glutamic acid or glutamine at the N terminus. The increasing knowledge of interpretation of data by practicing more examples and even remembering the inputs provided in this session will help you greatly. The way mass spectrometry based, proteomics have really taken big pace for any kind of biological problem; there is no need to look at the complex proteome analysis. If you are able to interpret your mass spec data you can greatly benefit.

I just like to emphasize that sample preparation and data interpretation are the two most crucial steps; you as a participant, as a learner can prepare your own sample. Send your samples to some facilities, where they have the advanced mass spectrometers such as ours at IIT, Bombay. And, then after running those samples you can obtain the raw data as long as you have now ability to interpret your mass spec data, then you are pretty much in a commanding position to get the best from your data set. So, please do go through today's lecture as well as the previous lecture and the hands on sessions in more detail.

Thank you.