Introduction to Proteogenomics

Dr. Sanjeeva Srivastava
Mr. Saicharan Ghantasala
Department of Biosciences and Bioengineering
Indian Institute of Technology, Bombay

Supplementary Lecture - 12
Data analysis using Skyline

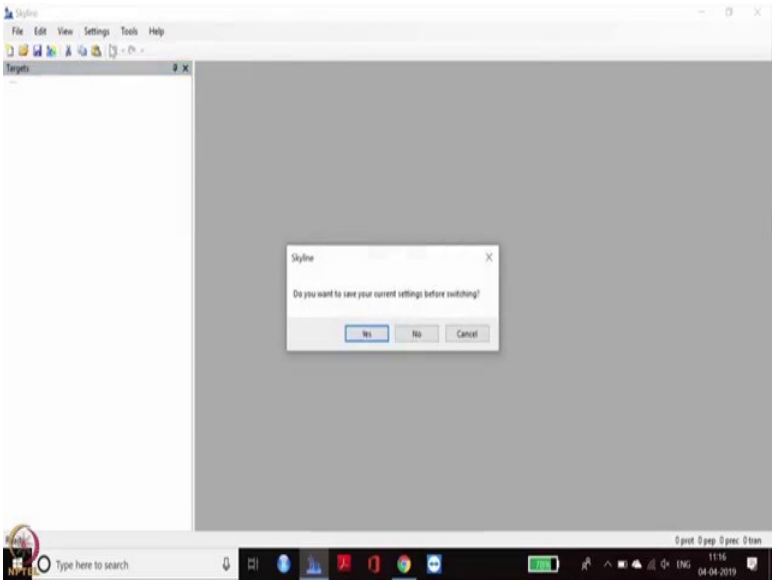
Hello. So, in the previous lecture we took a look at the importance of targeted proteomic approaches in the field of proteomics, the different steps that are required to design a successful targeted proteomics experiment and the principle and basics of selected reaction monitoring approaches. In the lecture I had specified that we would use the software Skyline to analyze a data set. So, today's session would focus on the use of a open source software named Skyline to analyze the data set from a selected reaction monitoring experiment, that was done in our lab. This software was developed by Mr. Brendan Maclean from the lab of Doctor Mike Maccoss at the University of Washington.

The software Skyline has revolutionized the field of targeted proteomics. It is a very simple to use software and there are many tutorials which are available on the website. In order to use Skyline you are requested to register and download the software from the following link. So, using a data set from an experiment that was performed in our lab, we would quickly walk you through the different tabs that are present in the software Skyline and how to look and analyze the data that was generated. So, before we look at Skyline and the different options available I would like to briefly discuss the experimental design.

So, the protein of interest for this particular experiment was apolipoprotein e which was found to be significantly dysregulated in one of the projects. So, for this experiment we had ordered heavy peptides from a vendor and the heavy peptides are nothing, but the same peptides which one wishes to monitor in their sample of interest. But, they differ from the endogenous peptide that is the peptide that is present in the sample. In the presence of a heavy labelled isotope at the c terminal end. So, it is clear that these peptides would contain the same amino acids and hence, their chromatographic behaviour will be exactly similar to the peptide of interest; that is the endogenous peptide or commonly referred to as the light peptide in this case.

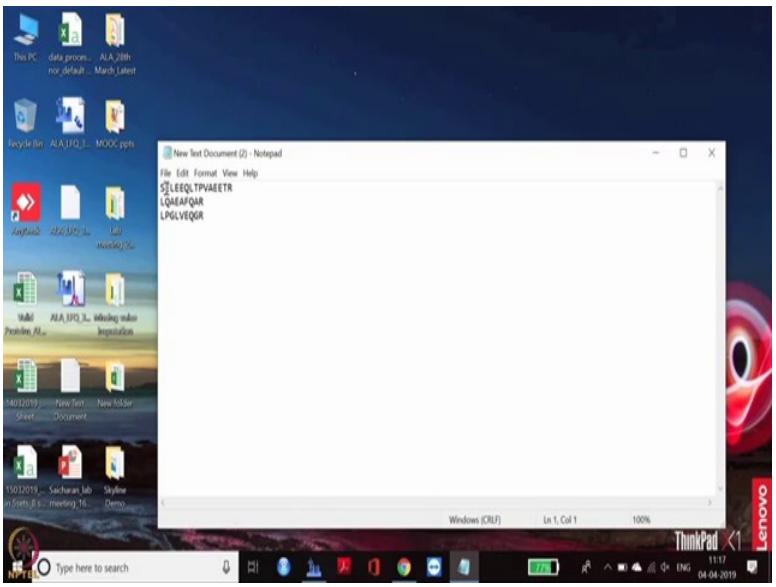
So, the heavy peptide helped us first get an idea of the retention time of our peptide. So, once we knew that the peptide of interest would elude at a particular time, these heavy peptides were spiked in our sample of interest and we monitored three peptides for the protein apolipoprotein-E.

(Refer Slide Time: 03:14)



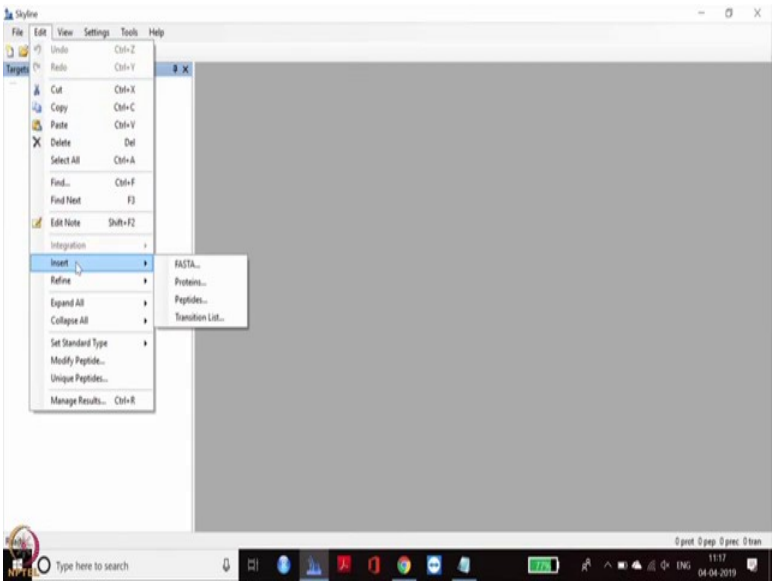
So, let us now look at the software and the different options that are available. The first and the foremost thing that we need to do is go to the default settings and say no. So, that it removes any previous setting which was there in the software, once this is done the particular document is fresh and blank. So, this is where you paste the targets. So, in this particular experiment we had chosen three peptides of interest.

(Refer Slide Time: 03:43)

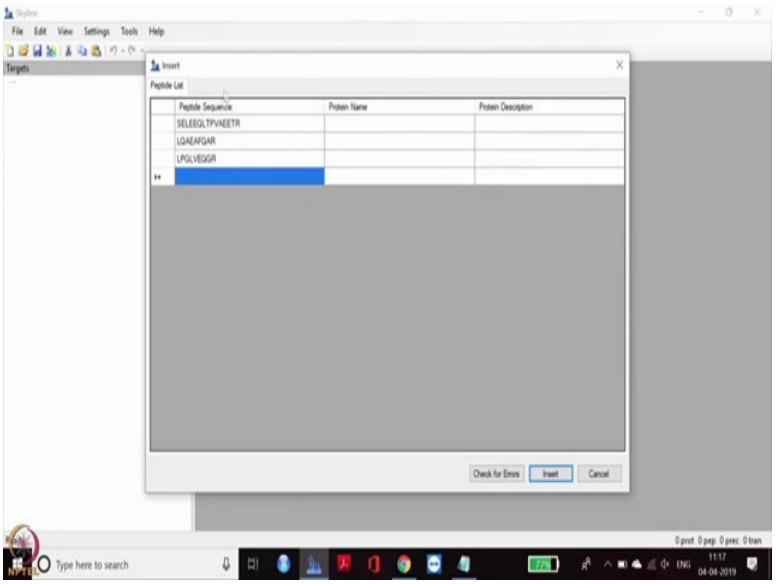


And these were the sequences for the peptides. So, I go to I just copy these three sequences.

(Refer Slide Time: 03:52)

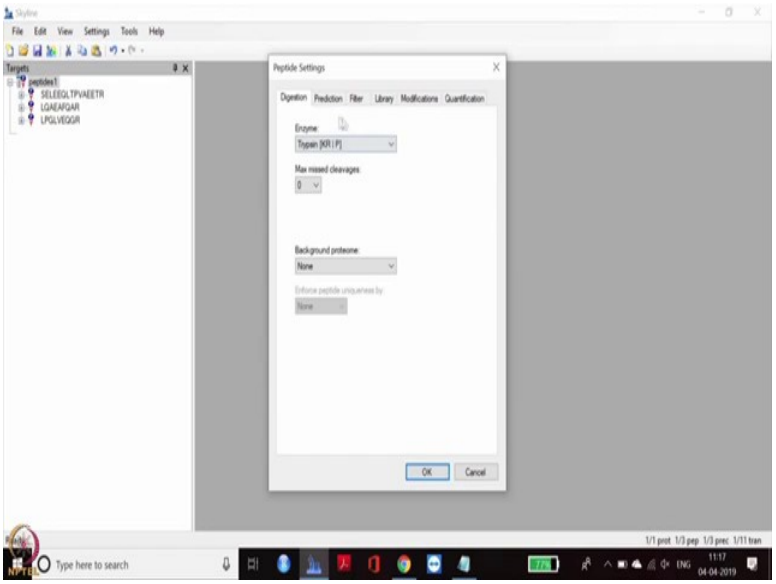


(Refer Slide Time: 04:02)



I go to edit say insert and peptides, I paste these three sequences here. So, what you see is each of these peptide sequence pasted and I say insert. So, to your left you see three peptide sequences for the protein apolipoprotein-E.

(Refer Slide Time: 04:20)

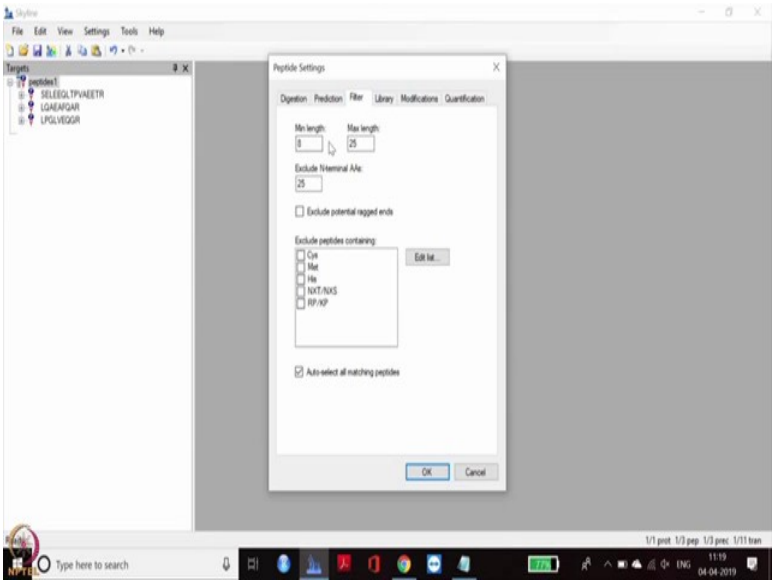


We next look at the settings tab, in the settings tab you need to go to peptide settings; start with the digestion. When you specify the specific parameters, the enzyme used, the maximum mixed cleavages, specify if there is a background proteome. So, in case if you have the

proteome of your organism, you can always integrate that with Skyline and all the process which I showed about pasting the peptide sequence can be avoided.

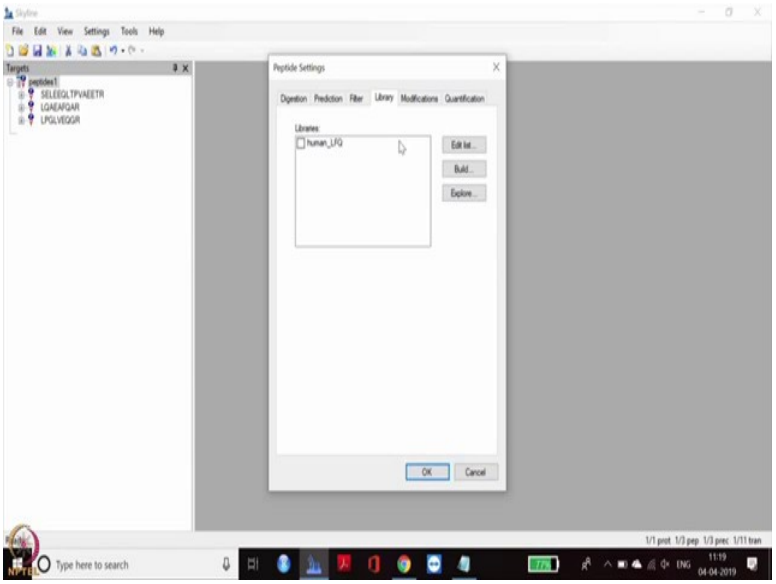
So, once a proteome has been chosen you can simply type the accession number of your protein of interest here and Skyline would generate all the different peptides that result from in-silico digestion of the protein. Once this is done we move to the prediction tab where, if you have a prediction a retention time predictor; you can use this calculator to predict the retention time of your peptide of interest. In this particular experiment we did not spike in anything so, we have left these values default.

(Refer Slide Time: 05:26)



Then we move to the filter tab wherein you specify the length of the peptide that is to be monitored. So, it could be from 8 amino acids to 25 amino acids, though the best peptides usually range from 8 to around 16 or 17 amino acids in length. Then, Skyline also allows you to choose any peptides which you would want to exclude due to their amino acid composition.

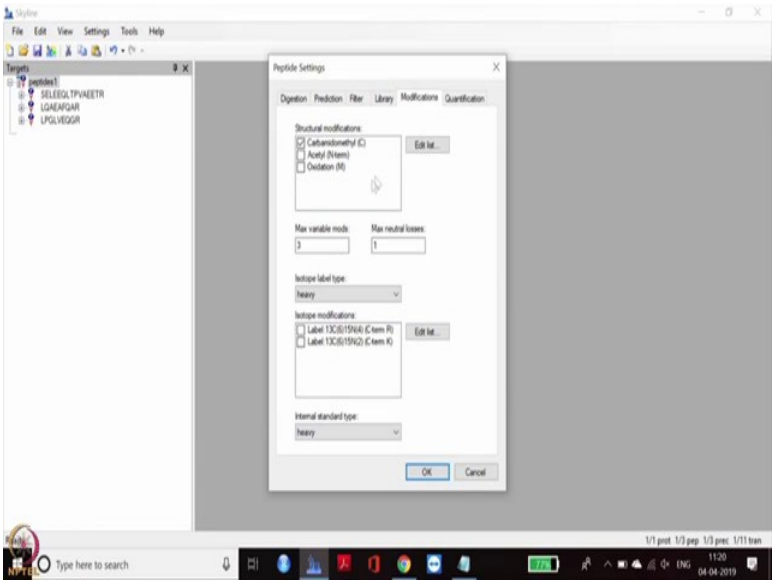
(Refer Slide Time: 05:56)



Then we move to the library tab; so, the use of library is that it would help you in data analysis by comparing it with the library spectra. So, assuming that you have performed a lot of shotgun experiments, after analysis of the shotgun data the files can be used to create a library of all the spectra that were detected in the shotgun approach.

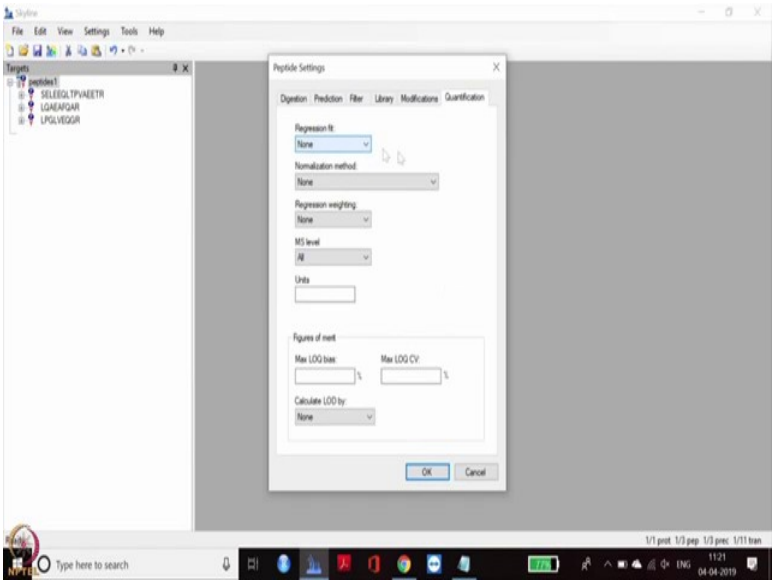
So, once you integrate the library spectra into Skyline and then you generate targeted data, you can always compare the library spectra match. And then Skyline provides a score to the match between the spectra in a targeted experiment to them to that seen in a shotgun experiment. So, you might not be able to see any library here because, you we have just installed the software, but in my case I see a human LFQ library because, I had used it in one of my experiments.

(Refer Slide Time: 07:08)



So, we next move to the modifications tab where in all the specific modifications which might occur during sample preparation can be specified and to this particular tab is of importance here. So, it specifies the isotope label type. So, we have used a heavy peptide as I have mentioned previously and the modification for that was this. So, I have selected this particular modification; however, if there is any modification other than what I have specified, you would always go to the edit list tab and add the modification. All the details for this software have been very well explained in the tutorial section available on the Skyline website.

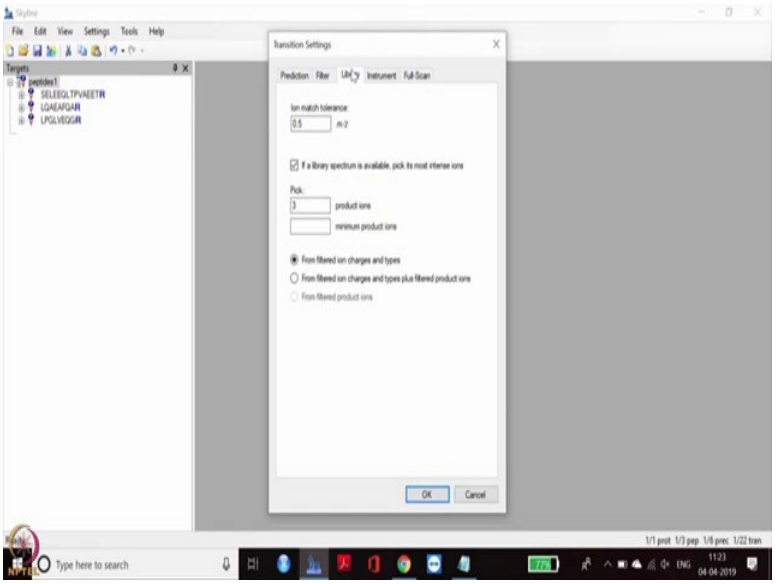
(Refer Slide Time: 08:04)



Then the next tab contains all the quantification related values. In this case our idea was not to quantify, but to just look at the correct peptide in our sample. So, we left all the values as default; once the parameters have been changed you click ok. So, you can see that the peptides now have a blue amino acid here, amino acid labelled in blue. This indicates that the heavy label was present at this site, we again go to the settings tab and we go to the transition settings. So, this is where we generate the different transitions for the peptides and the different charge state of the peptides.

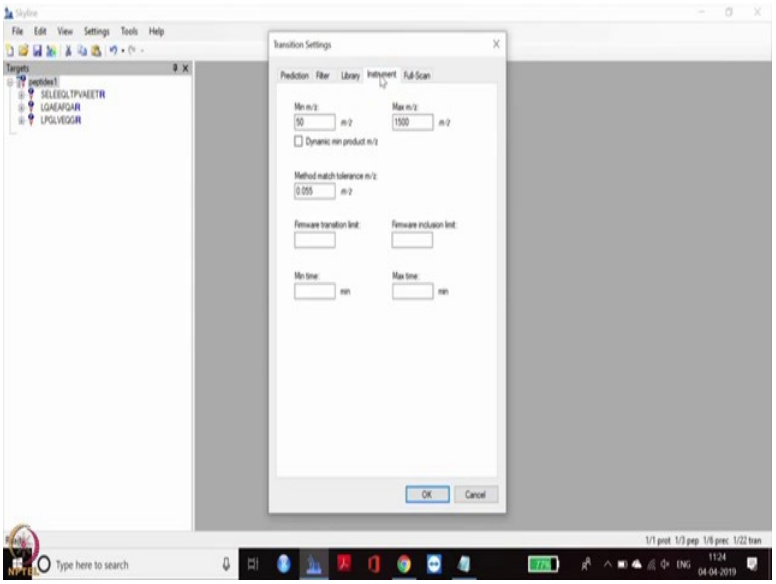
So, we left all the values default here for the prediction tab, then in the filter tab Skyline allows you to specify the precursor charges and the ion charges and the ion types. So, in a typical experiment you would want to monitor precursor charges of 2, 3, 4 and ion charges of 1, 2, 3 and different y ions, b ions and some other ions which you wish to monitor. For our experiment however, we had kept everything very simple; we monitored only a plus 2 precursor charge and then a plus 1 ion charge and only y ions; to not complicate the matters. Here the special ions, you can also specify the special ions.

(Refer Slide Time: 09:50)



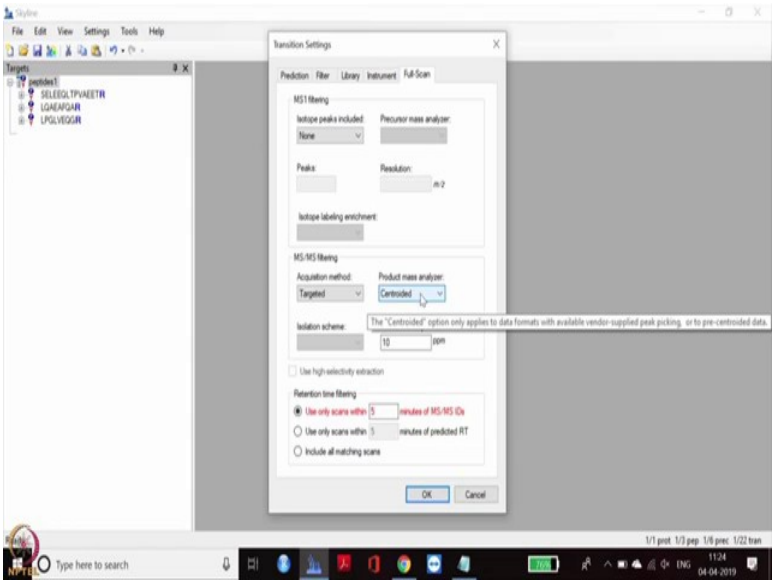
Moving to the library tab, since we do not have a library incorporated we could just deselect this option. In case you have a library which has been incorporated, selecting this would let you choose the peptides and the fragments. So, if you select this, it would give you, you could always specify how many product ions you need to monitor.

(Refer Slide Time: 10:12)



Once that is over we go to the different instrument parameters, we left the parameters as default here and the full scan.

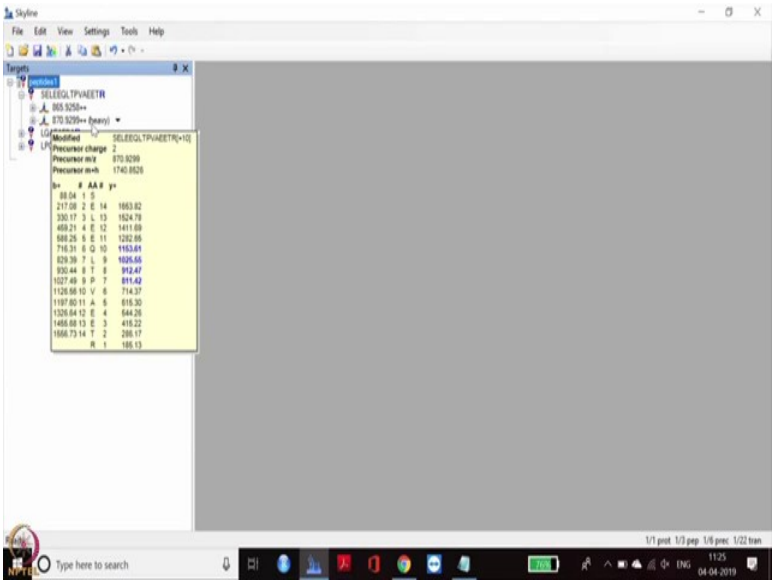
(Refer Slide Time: 10:20)



So, this also is a mall also is default and we clicked ok. So, I would just like to put light on this particular aspect, if the idea is to go for a DIA approach Skyline helps you use this option of DIA. If however, if you wish to do a PRM method you could go to targeted and choose centroided as the product mass analyzer. So, since this method was a SRM method we say none and we click ok.

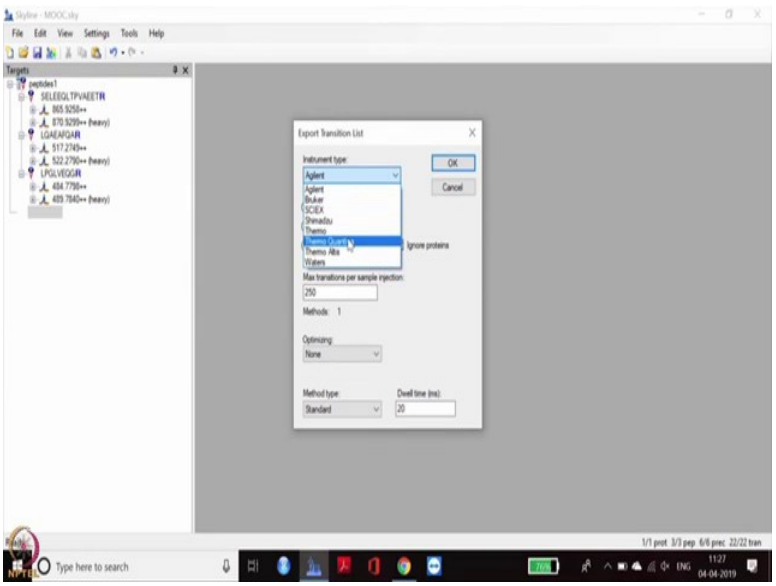
So, now we have successfully made the settings at the peptide level and the transition level. So, if you look at the left side of the of your page, if you see that there is our drop down menu drop down here. When I click on this plus symbol, it gives me two charge states here or two mass to charge ratios here; one is 865.9258, the other is 870.9299.

(Refer Slide Time: 11:41)



So, the one which is of a higher mass to charge is for the heavy label peptide which we had selected and its written ; so, you see that there is heavy in the brackets.

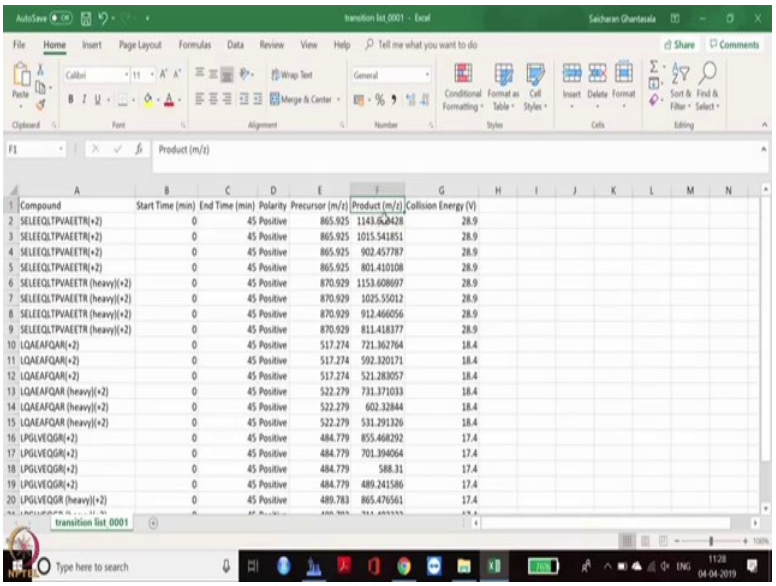
(Refer Slide Time: 11:53)



And, similarly for the other two peptides. On the right hand side of the page the right bottom side you see that one; so, you see that there are a total of 22 transitions coming from 3 peptides and 6 precursors. So, this is the basic information that you see after creating your file.

So, once this file is generated, you need to save the file; let us say save the file as MOOC. Now, once the file has been saved, we can export this as a transition list. So, I specify the instrument type that is TSQ the altis is here and then I say use start and end RTs and say ok. So, default collision energy yes and in this I say transition list.

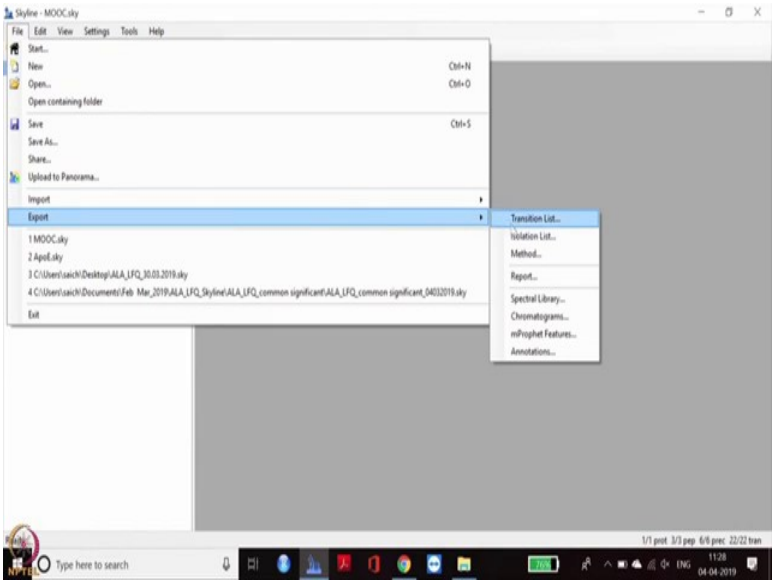
(Refer Slide Time: 13:20)



Compound	Start Time (min)	End Time (min)	Polarity	Precursor (m/z)	Product (m/z)	Collision Energy (V)
SELEEQLTPVAEETR(+2)	0	45	Positive	865.925	1143.62428	28.9
SELEEQLTPVAEETR(+2)	0	45	Positive	865.925	1015.541851	28.9
SELEEQLTPVAEETR(+2)	0	45	Positive	865.925	902.457787	28.9
SELEEQLTPVAEETR(+2)	0	45	Positive	865.925	801.410108	28.9
SELEEQLTPVAEETR (heavy)(+2)	0	45	Positive	870.929	1153.608697	28.9
SELEEQLTPVAEETR (heavy)(+2)	0	45	Positive	870.929	1025.55012	28.9
SELEEQLTPVAEETR (heavy)(+2)	0	45	Positive	870.929	912.466056	28.9
SELEEQLTPVAEETR (heavy)(+2)	0	45	Positive	870.929	811.418377	28.9
LQAEAFQAR(+2)	0	45	Positive	517.274	721.362764	18.4
LQAEAFQAR(+2)	0	45	Positive	517.274	592.320171	18.4
LQAEAFQAR(+2)	0	45	Positive	517.274	521.283057	18.4
LQAEAFQAR (heavy)(+2)	0	45	Positive	522.279	731.371033	18.4
LQAEAFQAR (heavy)(+2)	0	45	Positive	522.279	602.32844	18.4
LQAEAFQAR (heavy)(+2)	0	45	Positive	522.279	531.291326	18.4
LPGLVEQGR(+2)	0	45	Positive	484.779	855.468292	17.4
LPGLVEQGR(+2)	0	45	Positive	484.779	701.394064	17.4
LPGLVEQGR(+2)	0	45	Positive	484.779	588.31	17.4
LPGLVEQGR(+2)	0	45	Positive	484.779	489.241586	17.4
LPGLVEQGR (heavy)(+2)	0	45	Positive	489.783	865.476561	17.4

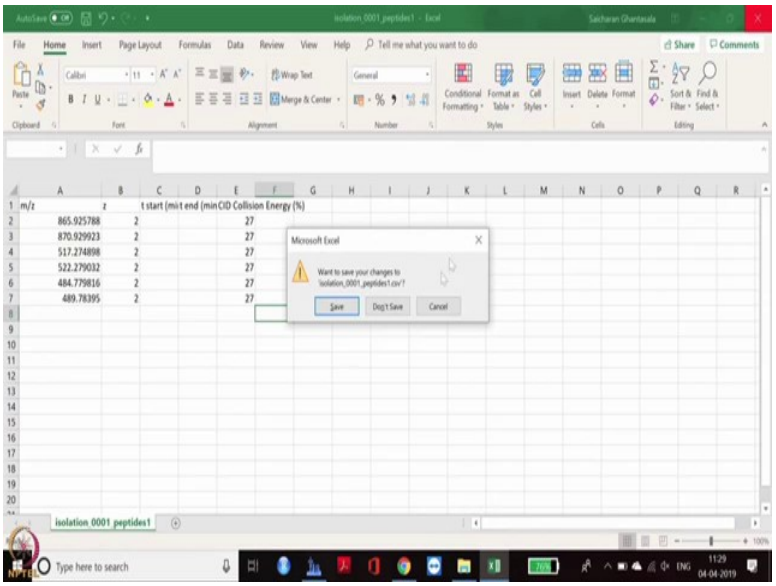
So, now Skyline has actually exported the entire list, if you can see this; as you can see here Skyline has exported each of these peptides and its corresponding transitions. So, the precursor is the Q 1 value, the product is the Q 3 value, the collision energies, the start and the end times for the experiment.

(Refer Slide Time: 13:49)



So, this was the transition list; however, if you export this method as an isolation list then the resulting file is that for a PRM.

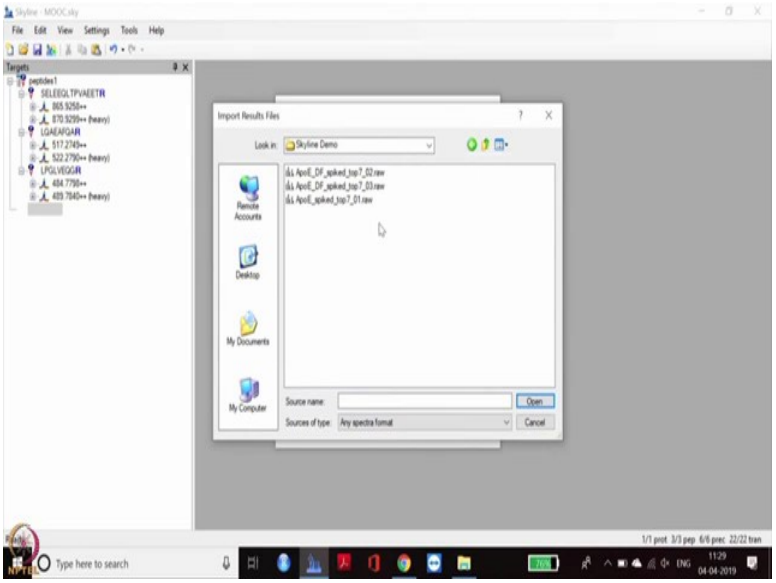
(Refer Slide Time: 14:09)



As you can see there is no mention of the sequence information, all that the method says is the mass to charge which is the Q 1 value; the charge state of the peptide, the parent ion, the collision energy that is to be applied. So, for a PRM experiment all the data will be captured that is all the fragments which get generated after fragmentation of the parent ion will get

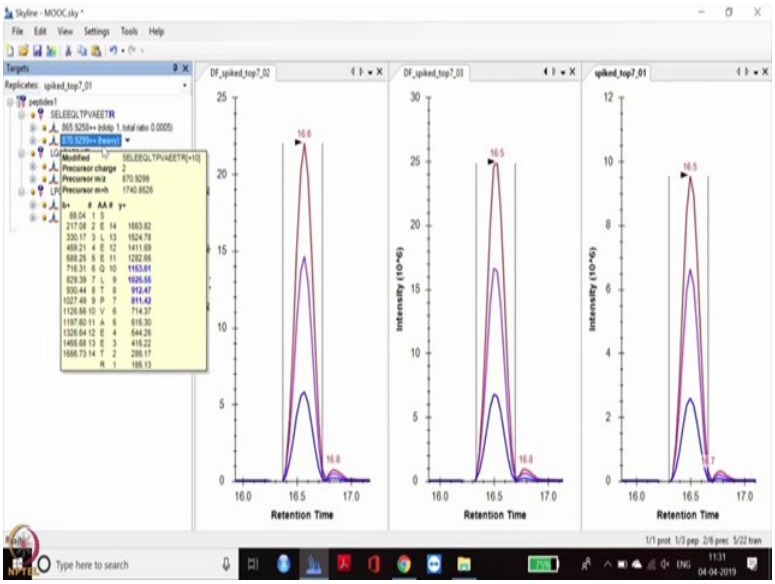
captured. So, assuming that the experiment has been performed, I now quickly import the data which was generated from this experiment, the results I say ok.

(Refer Slide Time: 14:54)



And, these are the same files which were run in triplicates I say ok.

(Refer Slide Time: 15:03)



So, Skyline has now imported all the files and if you see this let us go to a different view for this I go to tiled. So, as you can see each of these samples you see the intensity for the peptides in each of these samples. So, for the peptide 1 you see that the total ratio is 0.0005 which means that the intensity of this particular endogenous peptide as compared to the

heavy peptide is the ratio of 0.0005. So, the peak so, currently I have selected the information for the heavy peptide. And when I click on the sequence per se, what we see is this is the synthetic peptide that is the heavy peptide and this somewhere here is the endogenous peptide.

So, if I zoom in I will get to see the endogenous peptide peaks as well. So, this is the endogenous peptide and this is the heavy peptide. So, the next step in the data analysis involves looking at the transitions which have been detected and the and deleting the ones which have not been detected. So, in this particular case I have not been able to see the transition for E 6; so, I just remove that. Here I do not see the transition for Y 6 I remove that. So, this is the very basic information that can be provided to you about the use of Skyline.

So, this was a very basic session on Skyline and the different capabilities that the software has. I am sure that you are now interested to explore this software even further and get a much better and a comprehensive idea of the power of targeted proteomics. In addition to Skyline, there are many other vendor specific softwares which can do the same job.

We request you to take a look at the different software options that are available, try out a few test datasets and get a better understanding of targeted proteomics experiment. I am sure that now you are interested in getting a much better understanding of targeted proteomics approaches and how they are likely to help you better your research. We wish you all the very best, in exploring this exciting new field of proteomics.

Thank you.

(Refer Slide Time: 18:09)