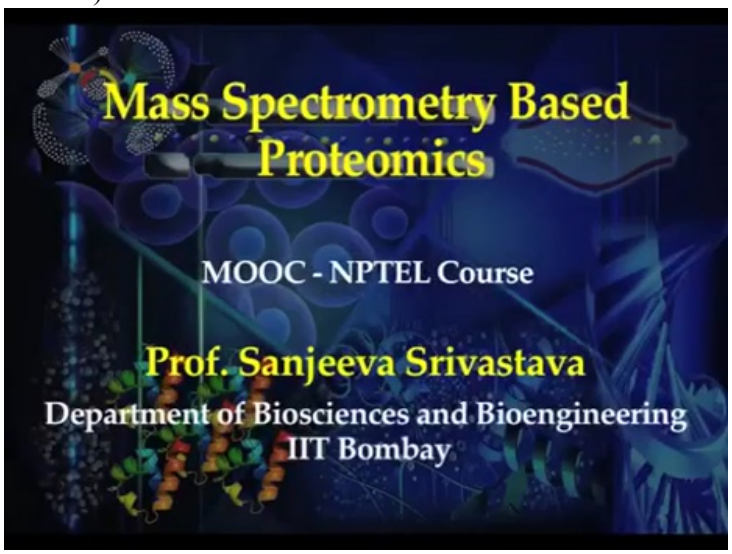
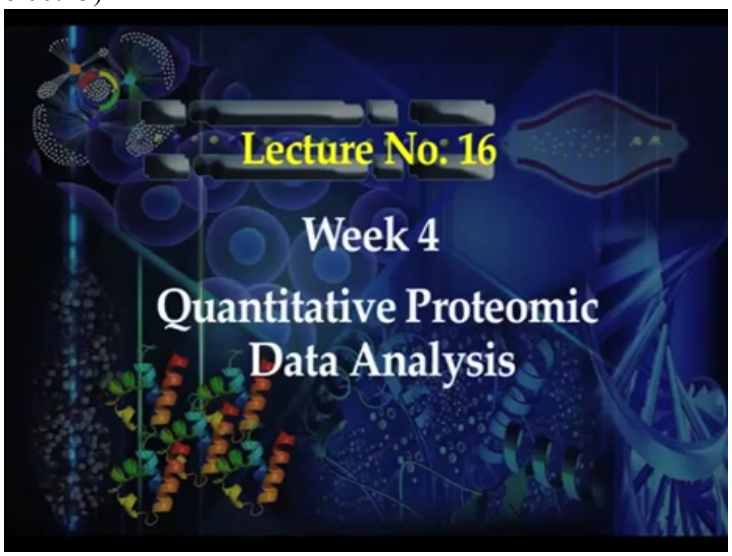
Mass Spectrometry Based Proteomics
Professor Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology, Bombay
Mod 04 Lecture Number 16

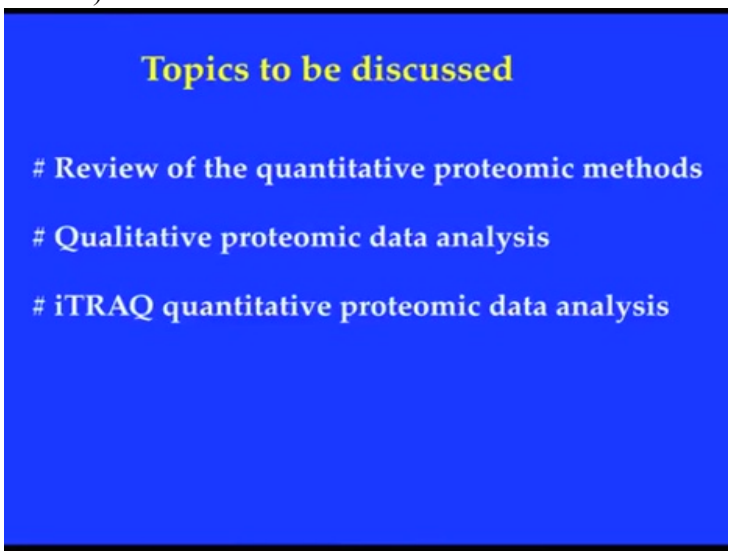
(Refer Slide Time 00:10)



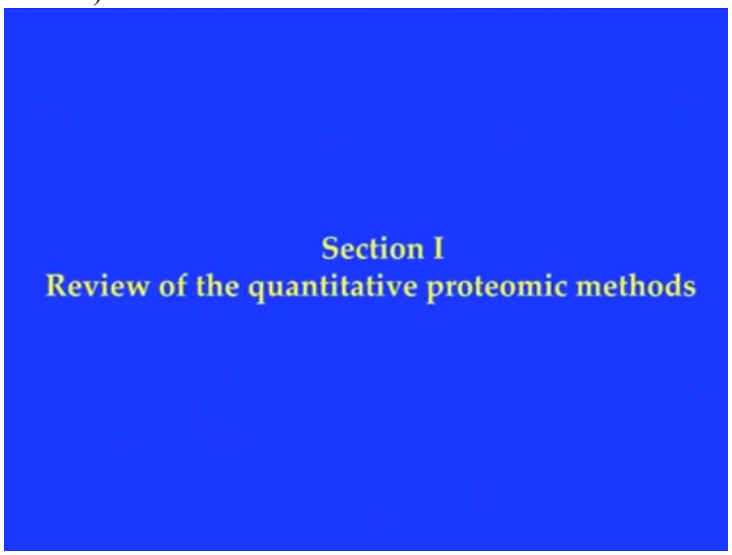
(Refer Slide Time 00:15)



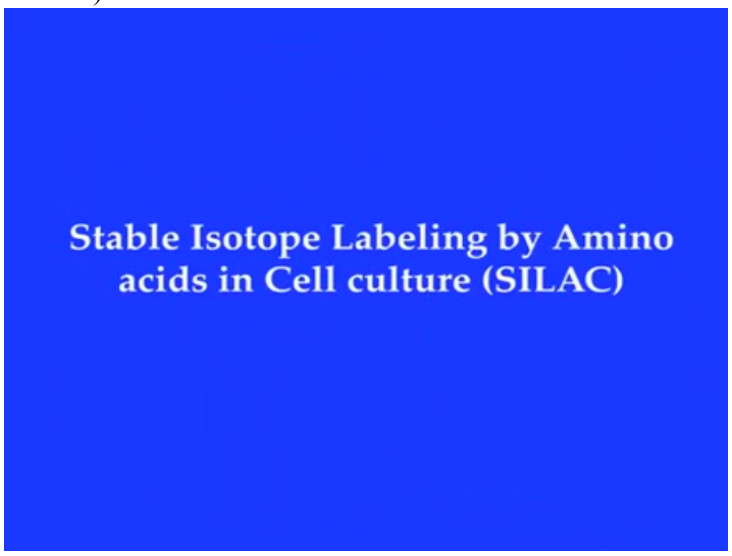
(Refer Slide Time 00:19)



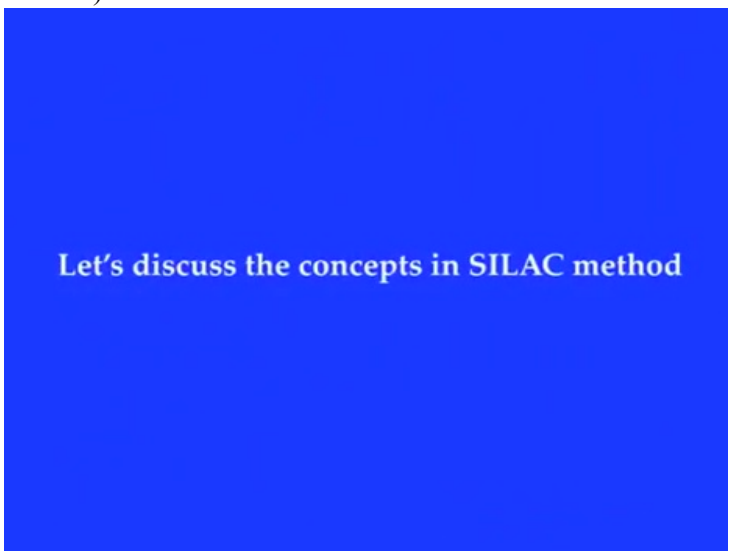
(Refer Slide Time 00:23)



(Refer Slide Time 00:27)



(Refer Slide Time 00:31)

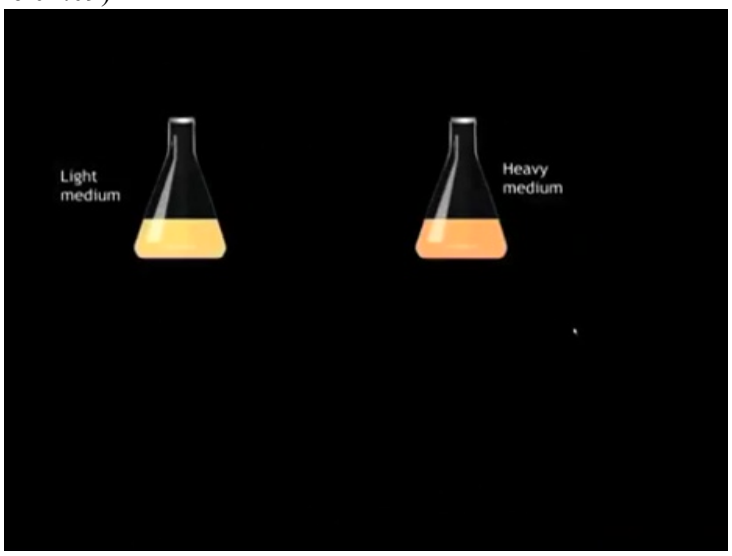


(Refer Slide Time 00:36)

Protein labeling with stable isotopes are effective methods for quantitative proteome profiling using Mass Spectrometry.

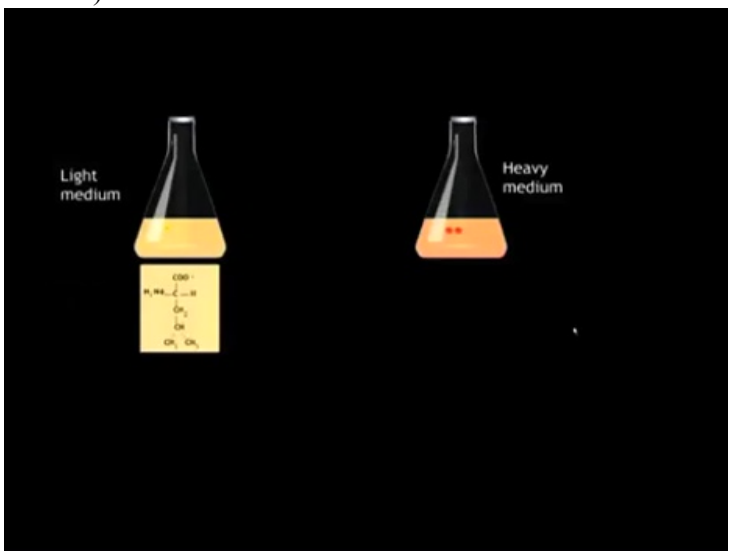
Stable Isotope Labeling by Amino acids in Cell culture or SILAC which is a metabolic labeling strategy to encode the whole cellular proteome is widely used method for the quantitative proteomics.

(Refer Slide Time 01:09)



In the SILAC

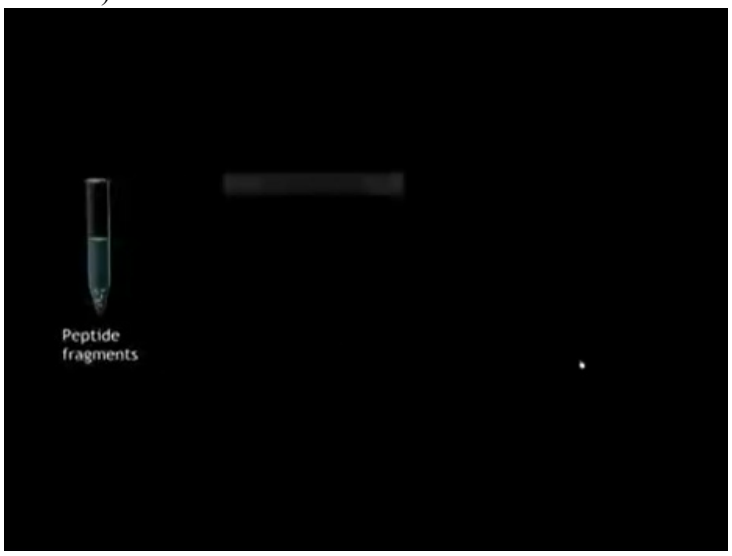
(Refer Slide Time 01:10)



… two group of cells are cultured in media two groups of cells are cultured in media that are identical in all the respects except that one contains a heavy, isotopic analog of an essential amino acid while the other contains the normal, light amino acid

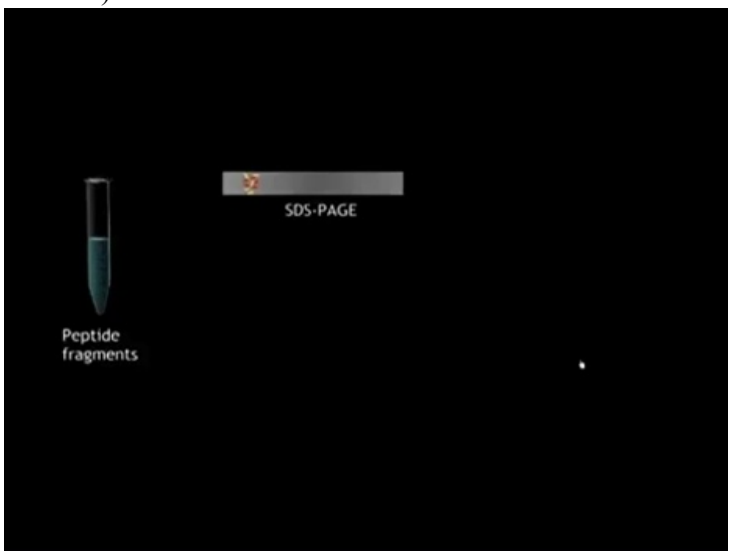
After a number of cell divisions, the grown cells are combined and digested using Trypsin.

(Refer Slide Time 01:43)



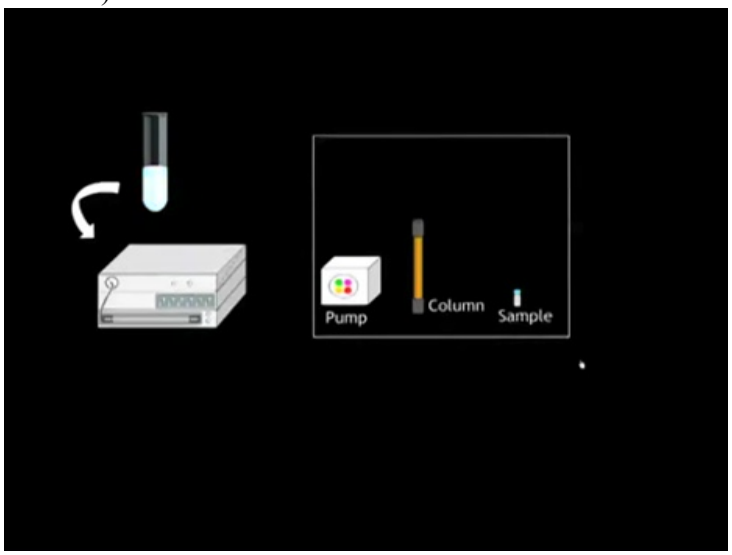
The complex protein mixture is further separated

(Refer Slide Time 01:45)



…by SDS PAGE to simplify the analysis

(Refer Slide Time 01:50)



Further application is carried out by liquid chromatography

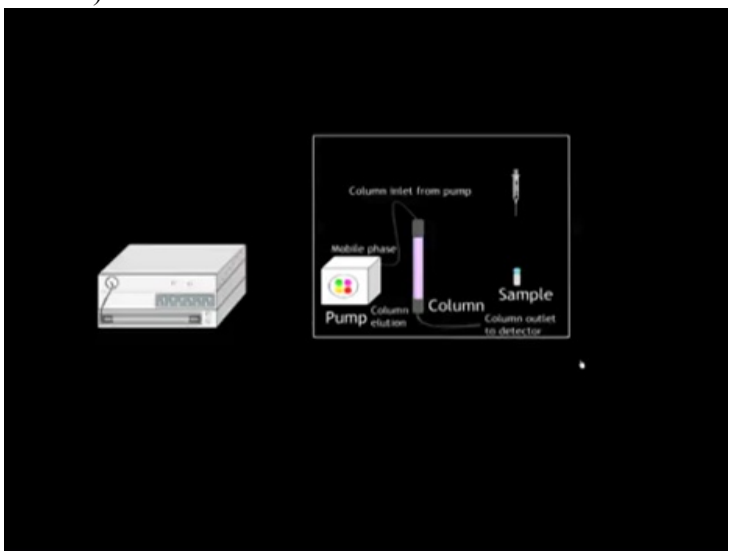
(Refer Slide Time 01:54)

and purified peptide fragments are analyzed

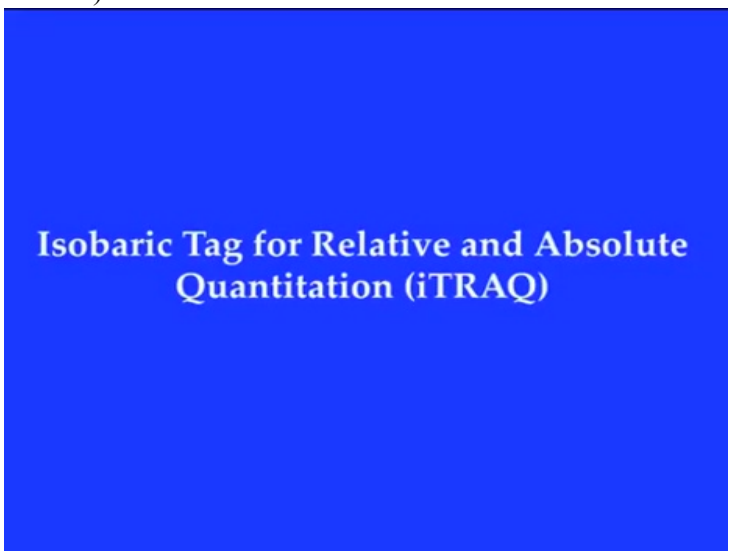
(Refer Slide Time 01:58)

by MS/MS.

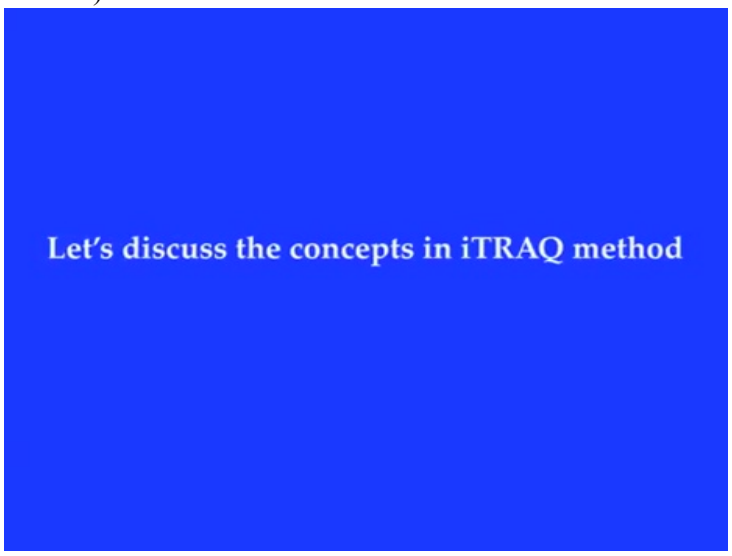
(Refer Slide Time 02:00)



(Refer Slide Time 02:03)



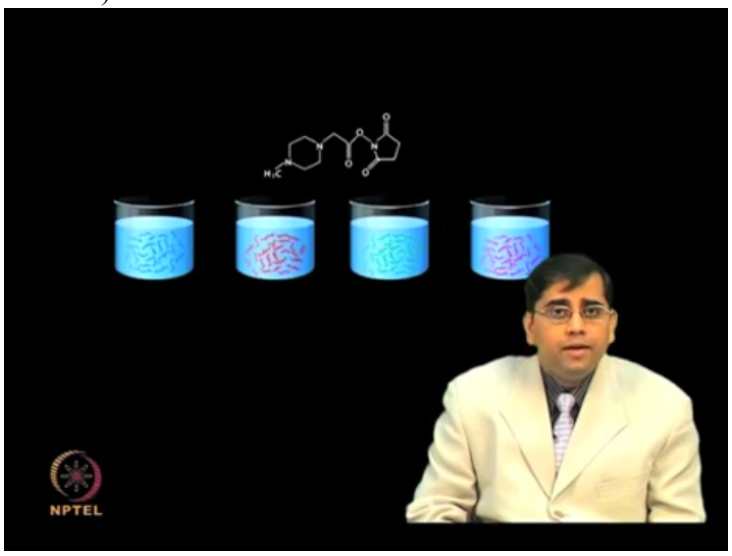
(Refer Slide Time 02:08)



(Refer Slide Time 02:11)

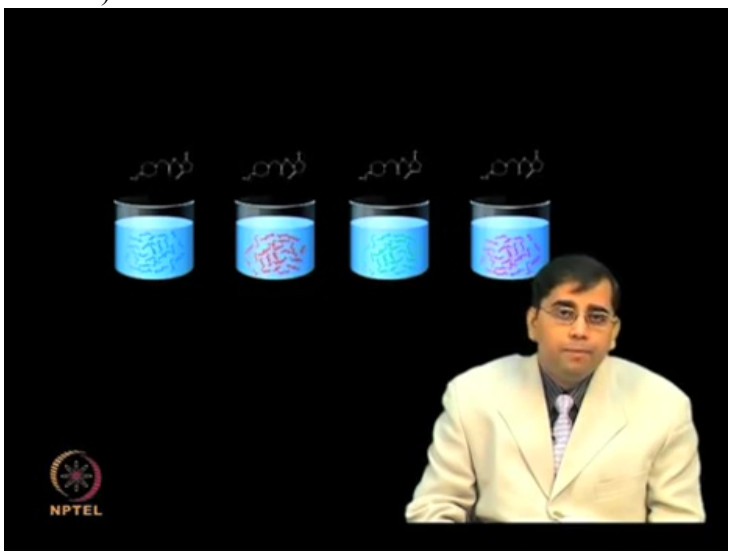
iTRAQ it is a MS based technique for relative and absolute quantitation of protein.

(Refer Slide Time 02:21)



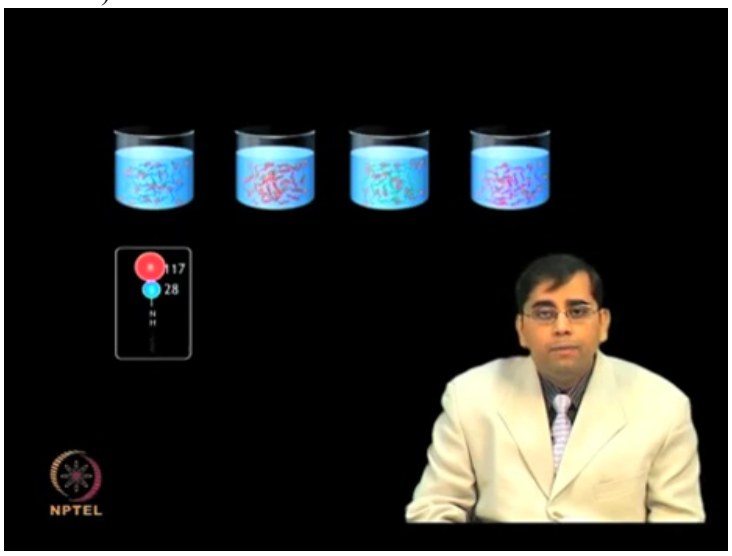
iTRAQ reagents are a set of 4 isobaric amine-specific

(Refer Slide Time 02:26)



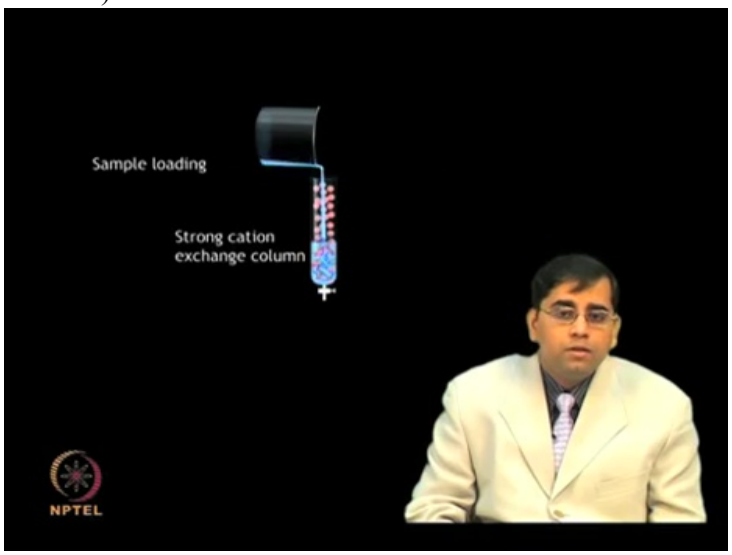
labeling reagents; 114, 115, 116 and 117. An iTRAQ reagent consists of a reporter group

(Refer Slide Time 02:40)



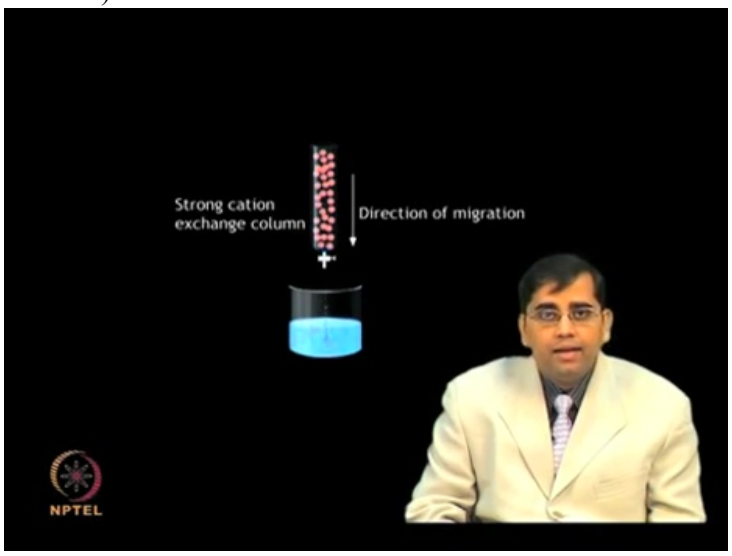
a balancer group and a peptide reactor group

(Refer Slide Time 02:45)



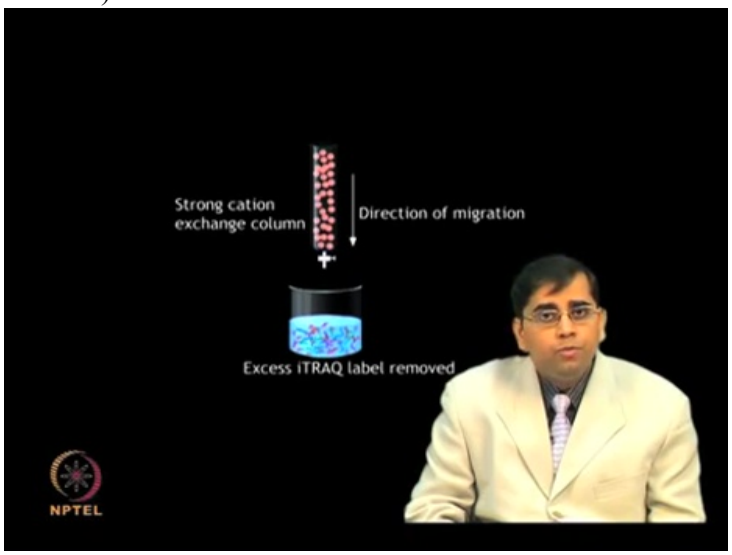
Pooled samples are purified on Strong Cation eXchange SCX column

(Refer Slide Time 02:50)



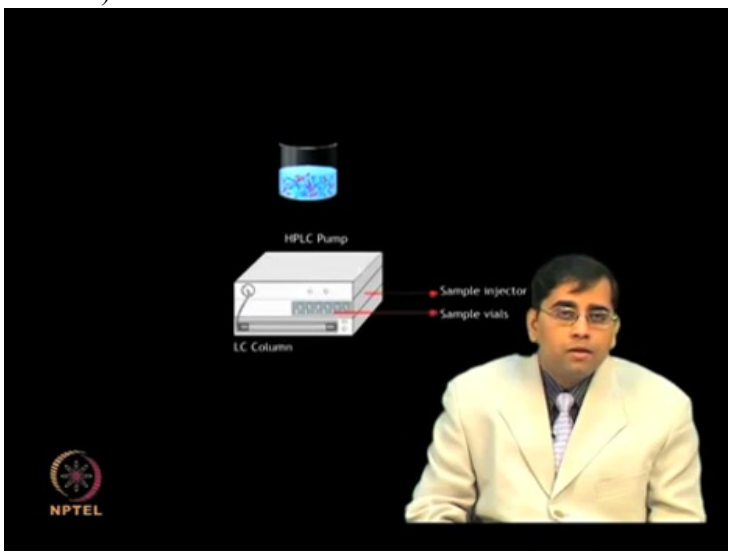
to remove

(Refer Slide Time 02:52)



the excess unbound reagent …

(Refer Slide Time 02:57)

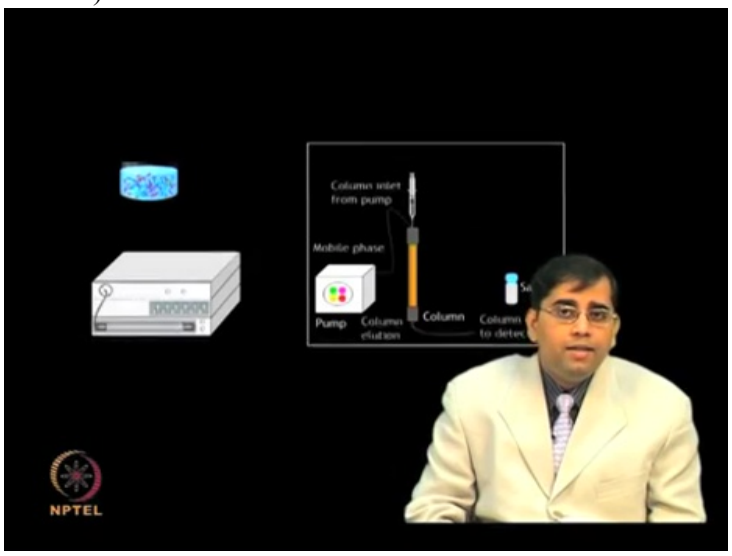


These isobaric labels are detected

(Refer Slide Time 03:00)

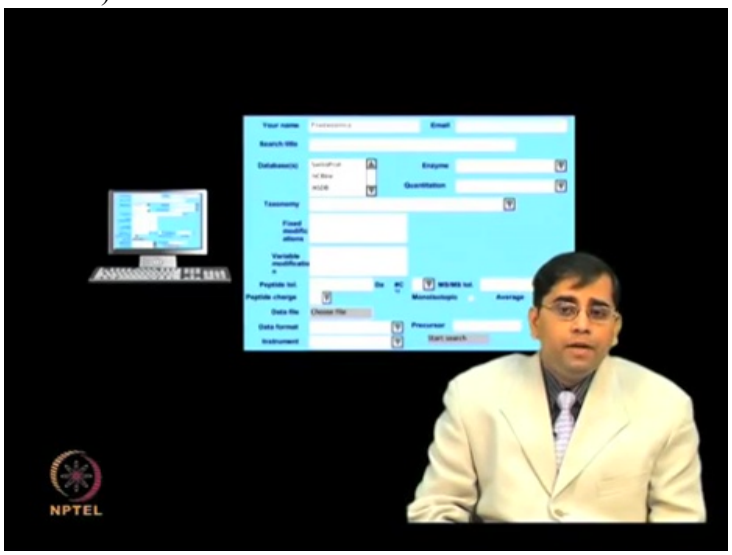
… upon fragmentation and release …

(Refer Slide Time 03:03)



…in mass spectrometry.

(Refer Slide Time 03:07)

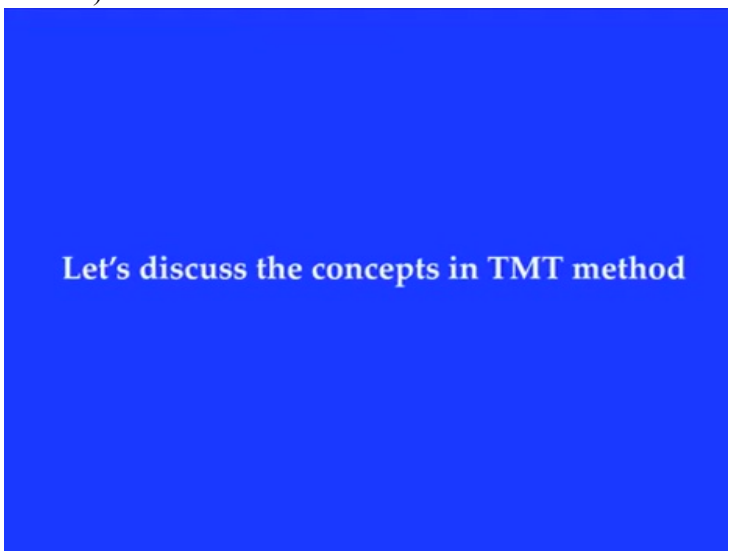


The data obtained from mass spectrometry can be analyzed by using search engines such as MASCOT. The analysis requires inputs regarding the experimental parameters such as enzyme cleavage, modifications, instruments used, peptide tolerance etc.

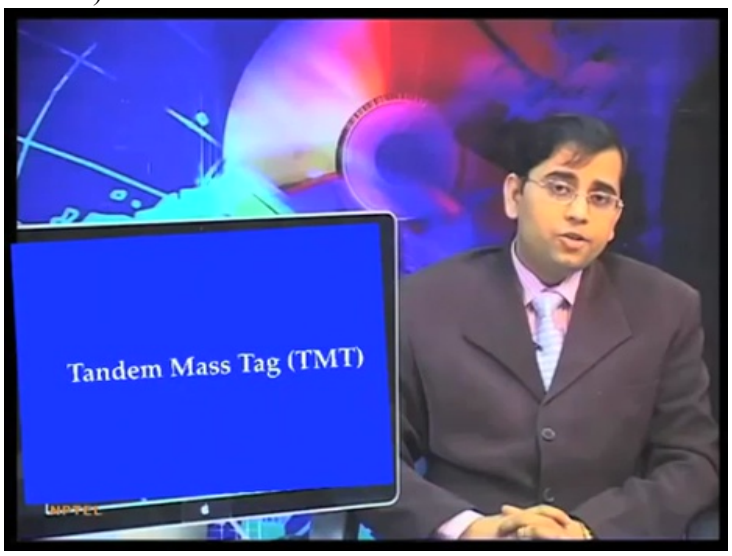
The data files generated from MS is uploaded and the search carried out by employing databases such as NCBI, MSDB and Swiss Prot.

(Refer Slide Time 03:47)

(Refer Slide Time 03:53)



(Refer Slide Time 03:57)



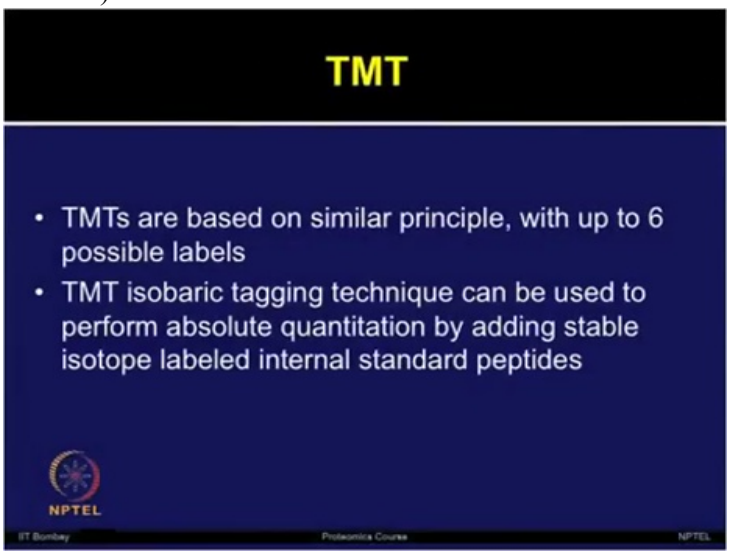
Now let us talk about Tandem Mass Tag or TMT. This method is similar to iTRAQ which we just discussed. TMT is also MS/MS based

(Refer Slide Time 04:10)

quantitative technique which uses the isotopomer labels referred as tandem mass tags.

It also provides the accurate quantification of peptides and proteins. The Tandem Mass Tags have been developed by the Proteome Sciences and currently commercialized by Thermo Fisher. I have given you the reference for the original study on Tandem Mass Tag in the slide.

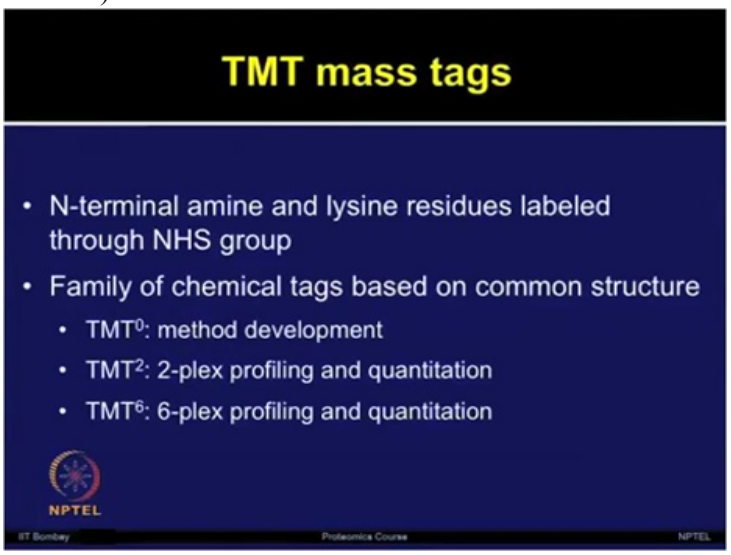
(Refer Slide Time 04:39)



So these Tandem Mass Tags they are based on similar principle of iTRAQ. Here the possibility for multiplexing is up to 6 possible labels. The TMT isobaric tagging technique can be used to perform absolute quantification by adding stable isotope labeled internal standard peptides.

It can be done by comparing the peptides from a target protein to a known amount of labeled standard peptide spiked into a sample. In that way absolute quantification can be obtained.

(Refer Slide Time 05:20)



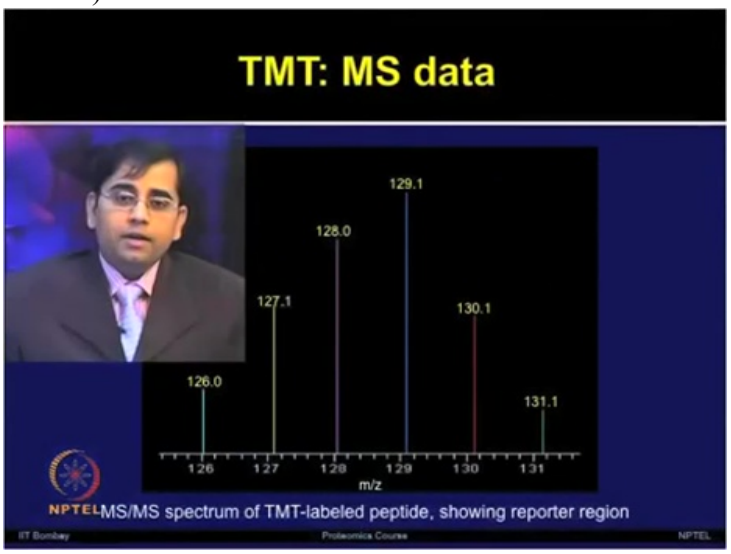
The N-terminal amine and lysine residues are labeled though NHS group. There are family of chemical tags which are based on the common structures. The series of TMT tags available TMT 0, TMT 2-plex, TMT 6-plex.

So these TMTs are innovative set of isobaric mass tags for labeling the proteins and peptides at amine functions and mixing of up to 6 different protein samples are possible.

While duplex and 6-plex labels TMT differ by the number of isotopic substitutions, the TMT 0 is non-isotopically substituted structure that has been produced for only method development.

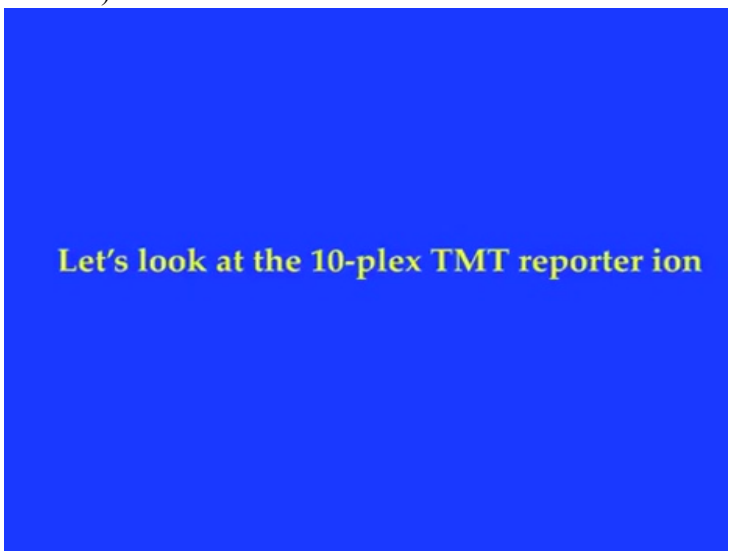
During the MS/MS analysis, the TMT tags give rise to 6 reporter ions from 126 to 131 Dalton; therefore it allows for the relative quantitation

(Refer Slide Time 06:31)

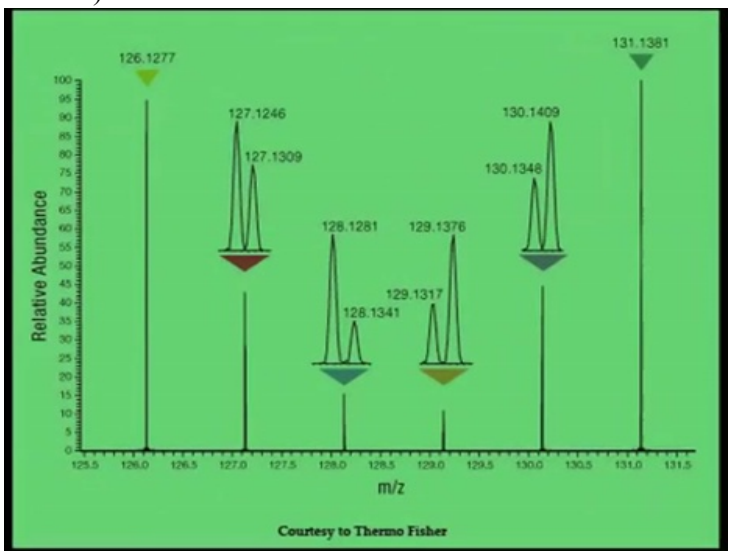


I am showing you one representative MS/MS spectrum of TMT labeled peptide which is showing a reporter region. The relative abundance of target protein or peptide fragment in 6 different samples can be easily measured by comparing these signature mass peaks which are generated by the different mass tags.

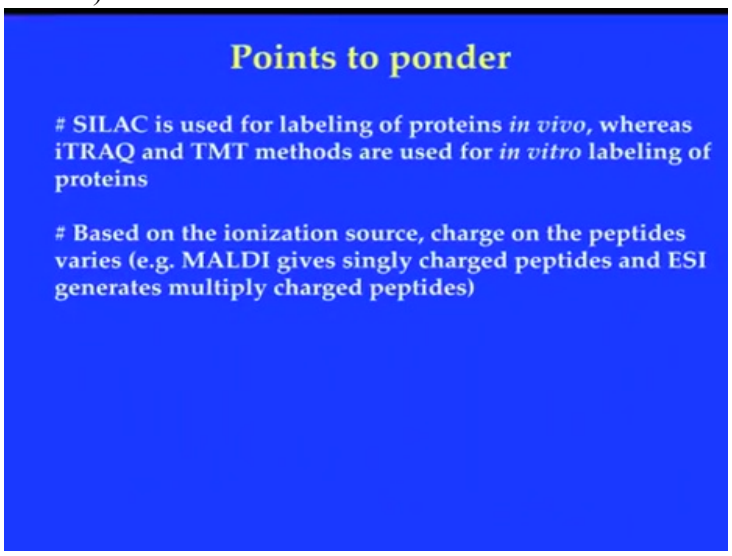
(Refer Slide Time 06:53)



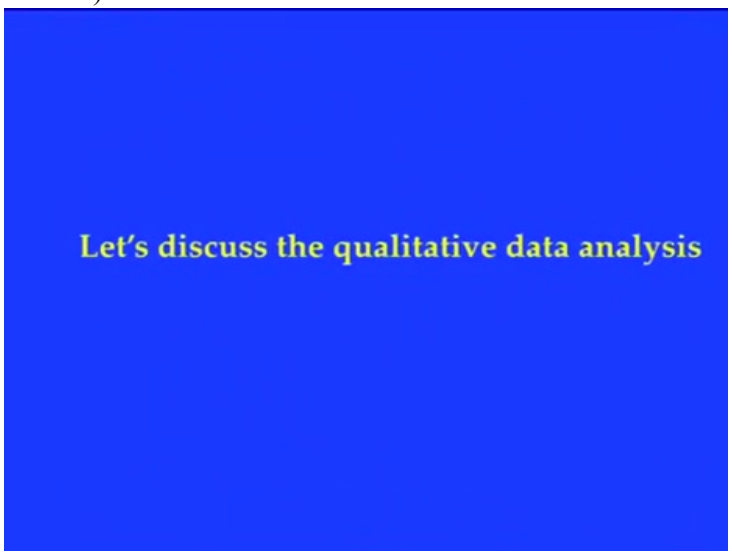
(Refer Slide Time 06:58)



(Refer Slide Time 07:03)

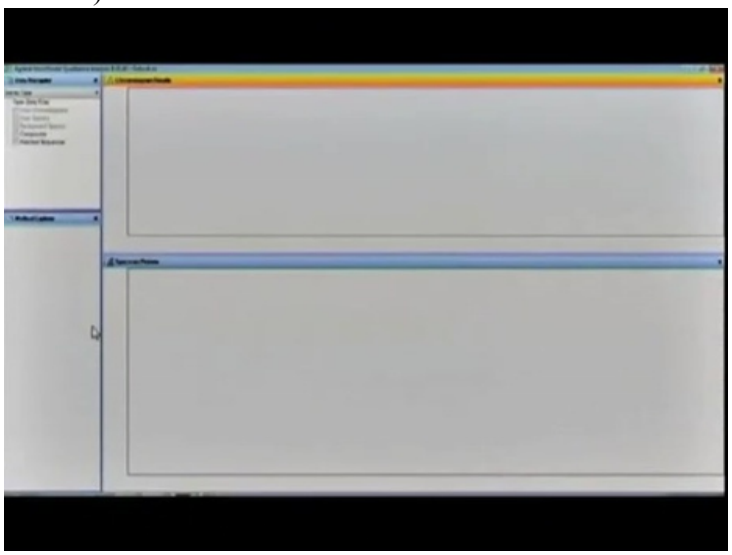


(Refer Slide Time 07:19)



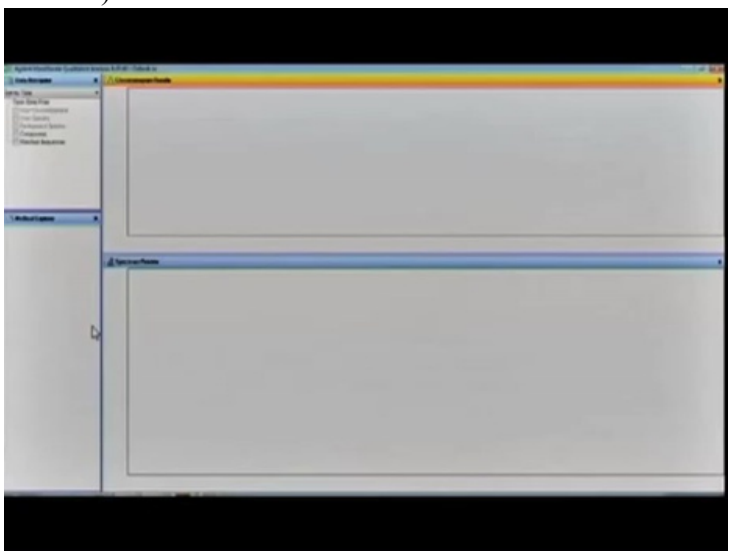
(Refer Slide Time 07:25)

(Refer Slide Time 07:31)



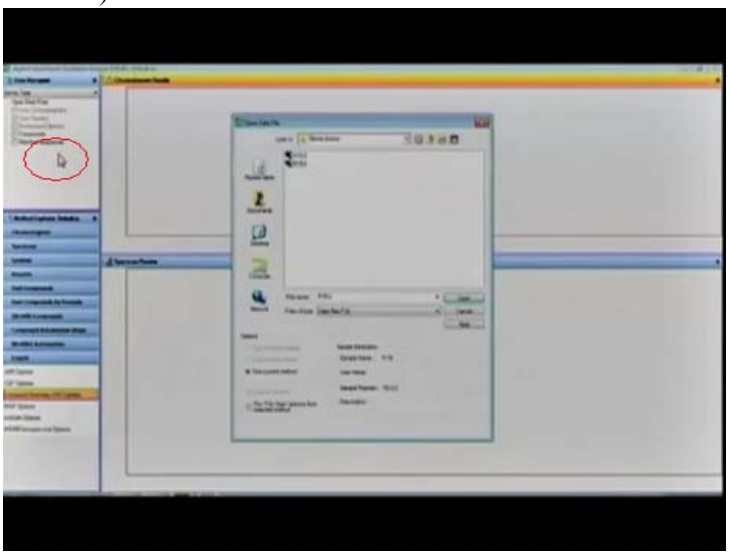
Data Visualization and qualitative analysis; Qualitative analysis, Click on

(Refer Slide Time 07:38)



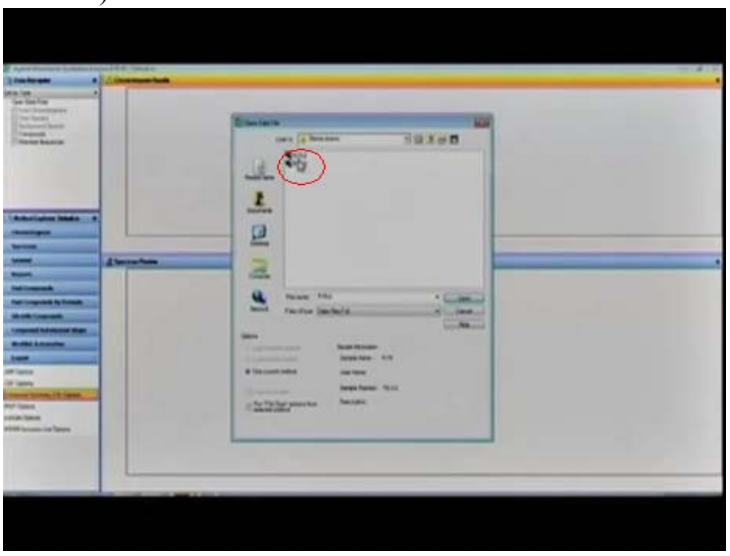
Agilent Qualitative Analysis software icon which will pop up a new window

(Refer Slide Time 07:51)



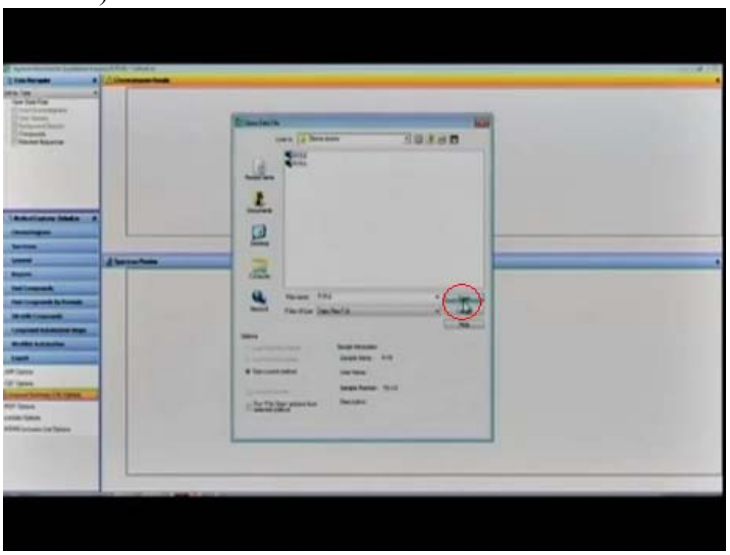
clicking on open a folder or file option we can open a Mass spec

(Refer Slide Time 07:54)



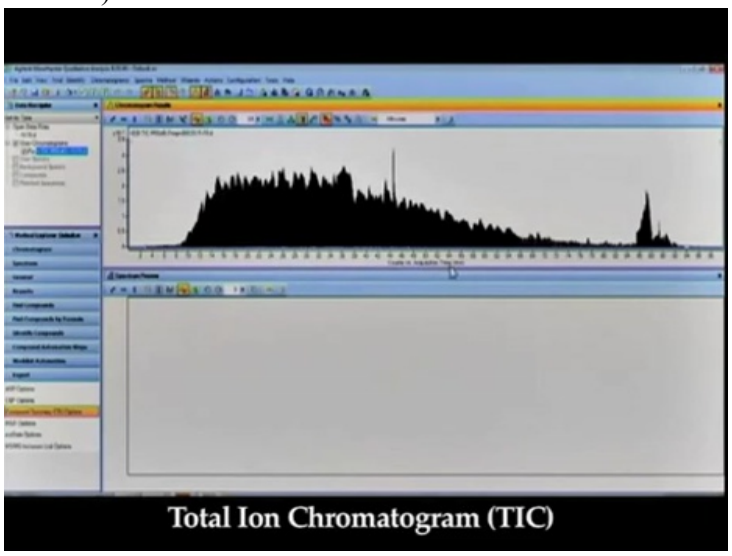
data file of interest.

(Refer Slide Time 07:56)



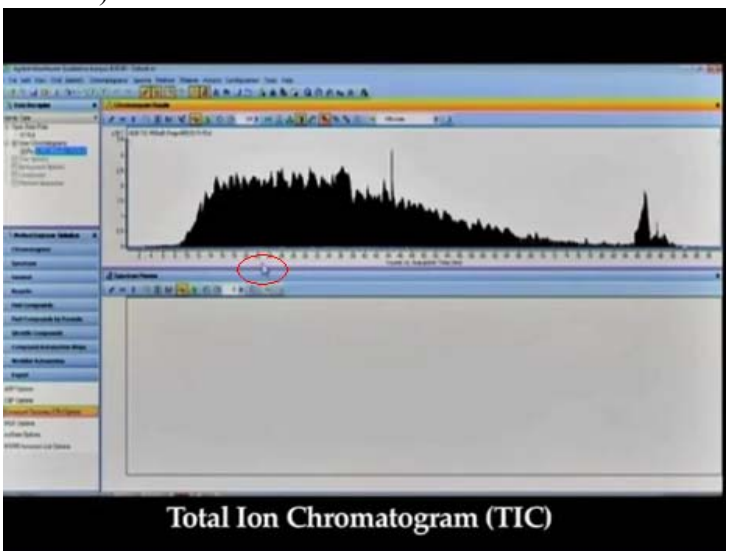
dot (.) d files can be opened

(Refer Slide Time 07:59)



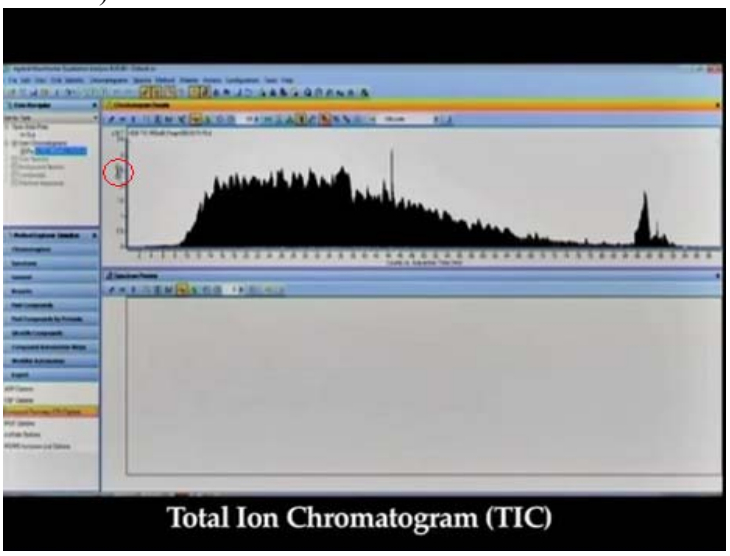
The chromatogram shows

(Refer Slide Time 08:01)



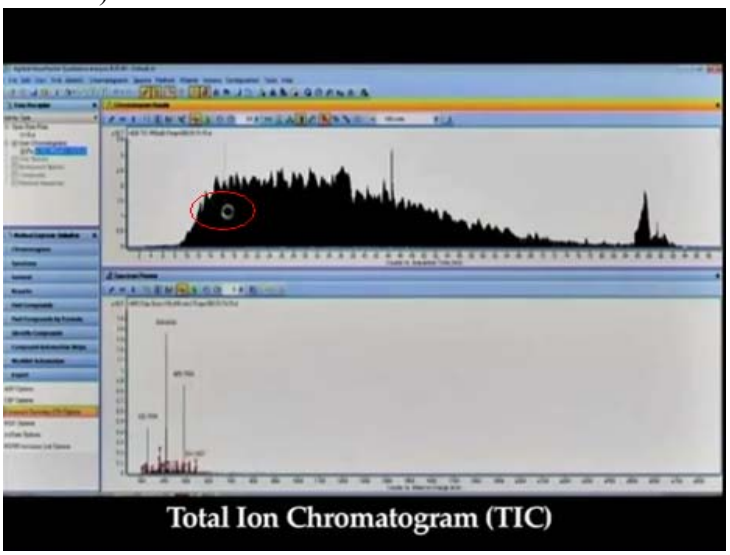
acquisition time period on x axis and

(Refer Slide Time 08:04)



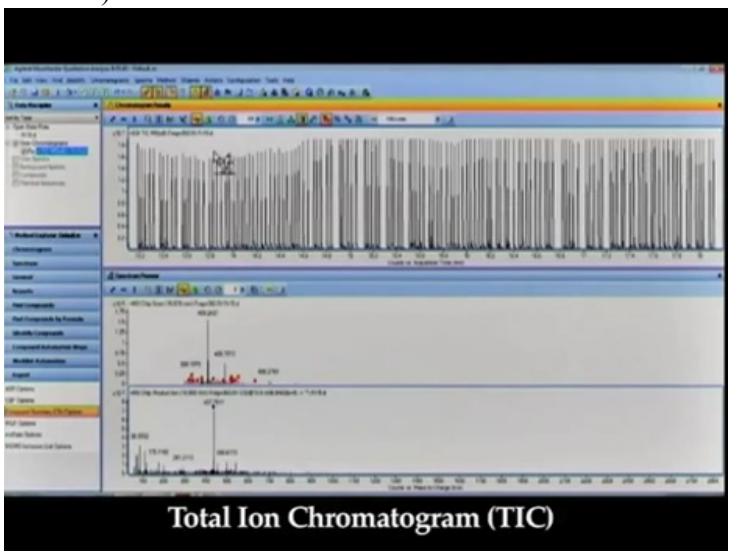
the intensity of eluted peptide ions on the y axis

(Refer Slide Time 08:11)

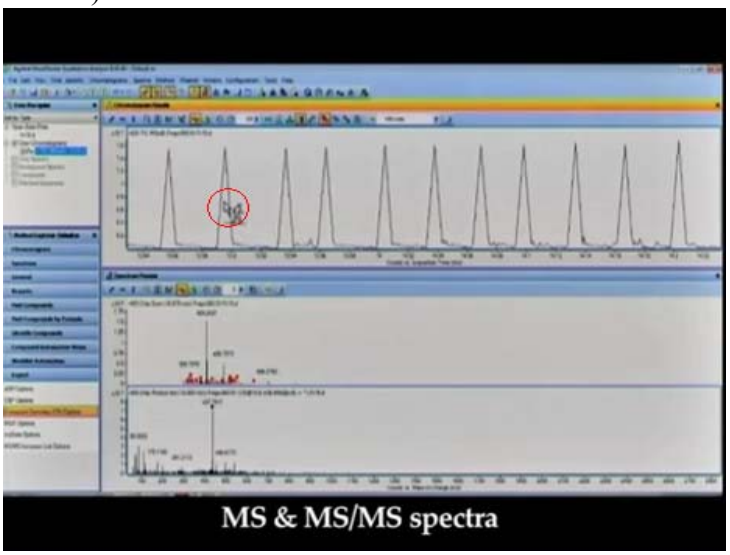


use right click of the mouse to zoom in the chromatogram

(Refer Slide Time 08:14)

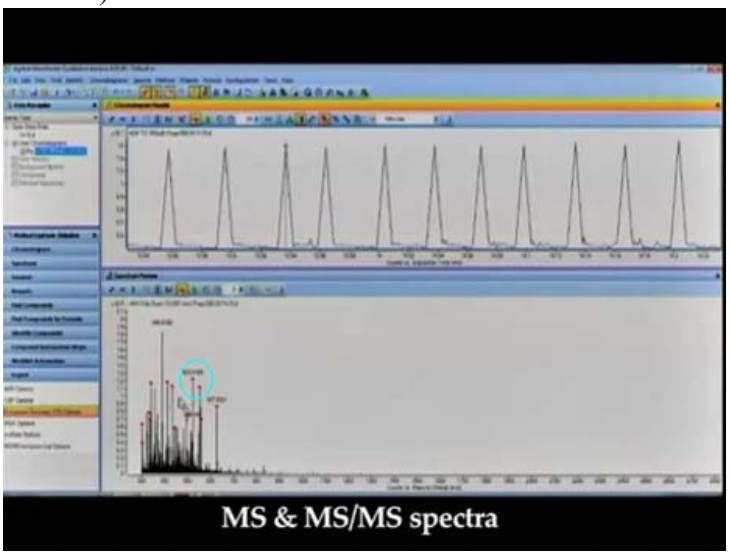


(Refer Slide Time 08:19)



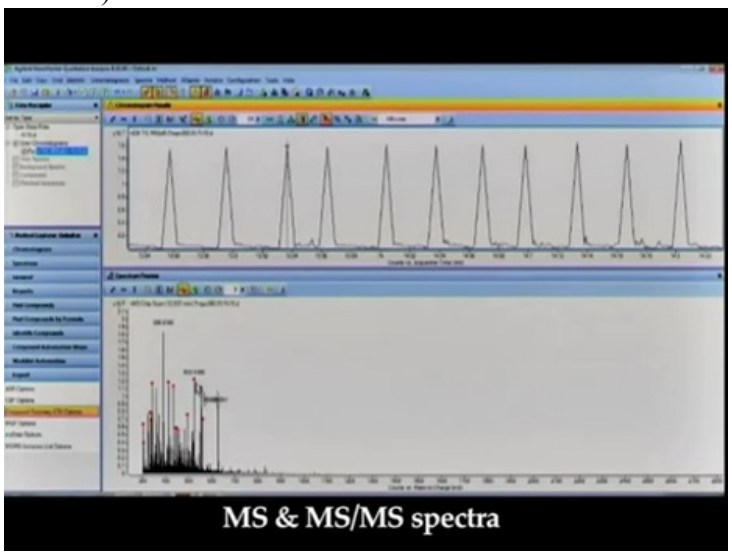
The bigger peaks which we can see are MS Spectra. At the bottom of this window we can see

(Refer Slide Time 08:25)



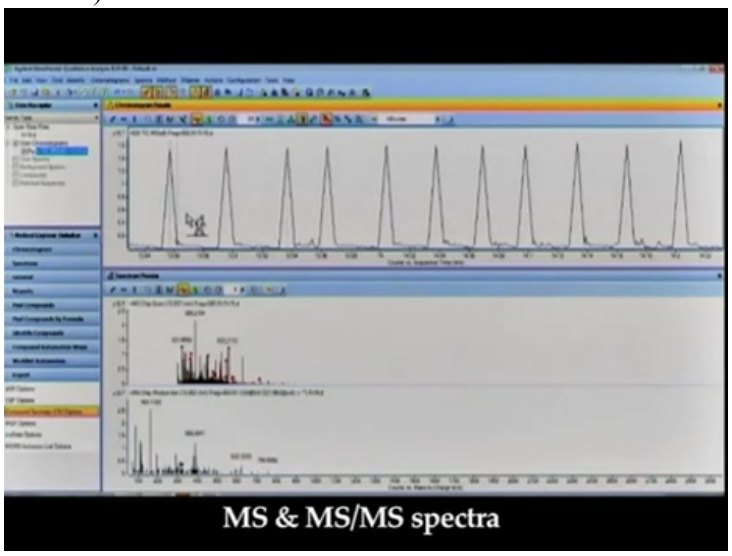
another window which shows different spectral peaks representing the precursor ions with the red color annotation

(Refer Slide Time 08:32)



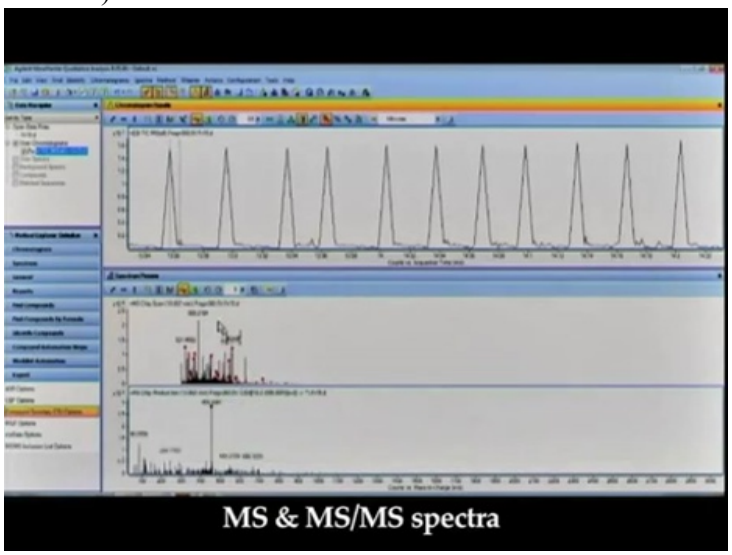
on top of each peak. In the topmost window from one MS spectrum

(Refer Slide Time 08:39)



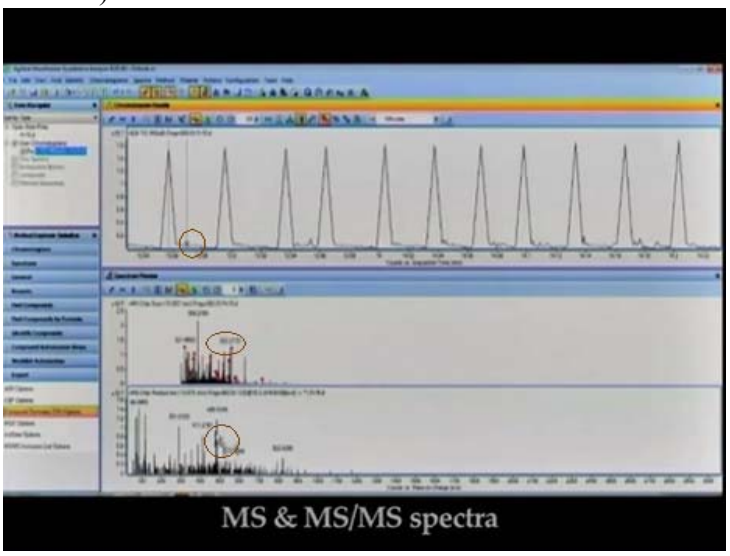
to next MS spectrum

(Refer Slide Time 08:43)



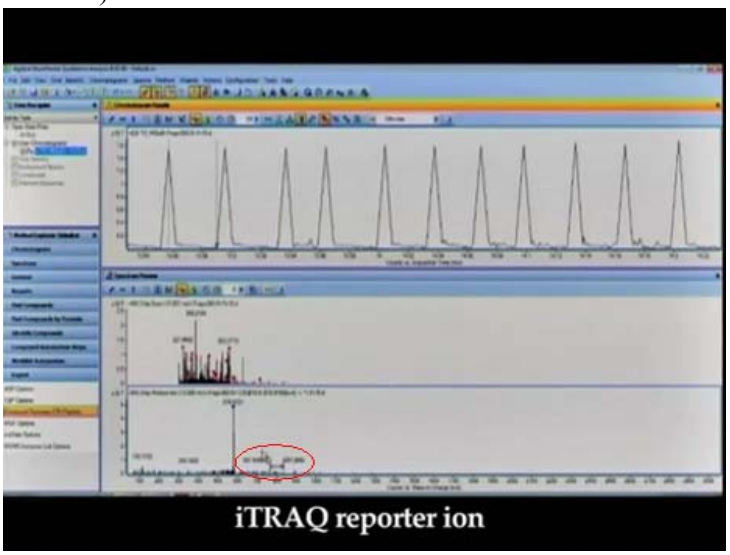
we can see many small peaks which represent MS/MS spectra.

(Refer Slide Time 08:47)



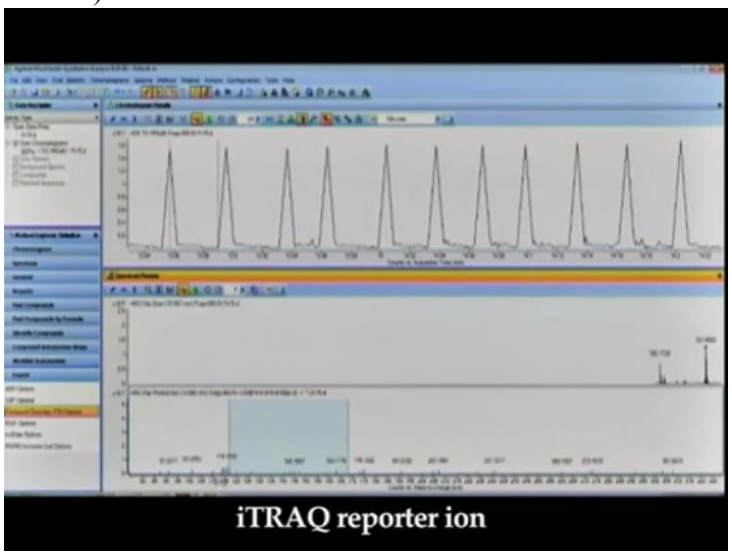
if the peptides are labeled with iTRAQ reagent

(Refer Slide Time 08:51)

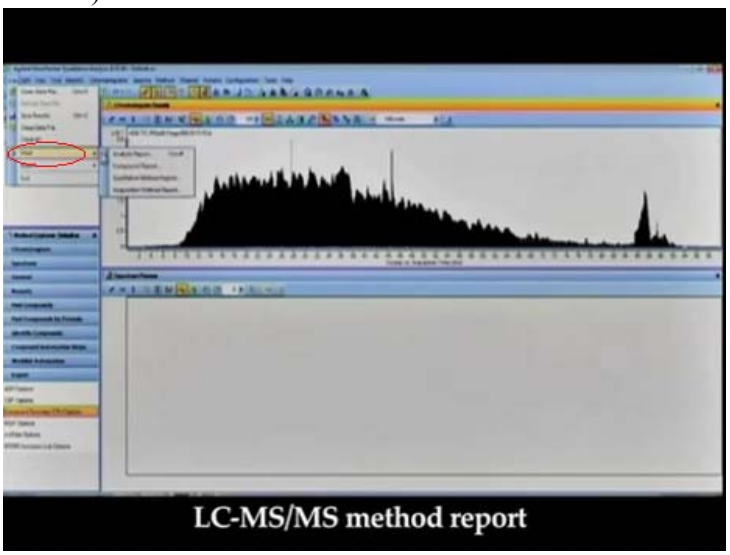


we can see the reporter ions in this MS/MS spectral view window

(Refer Slide Time 09:03)

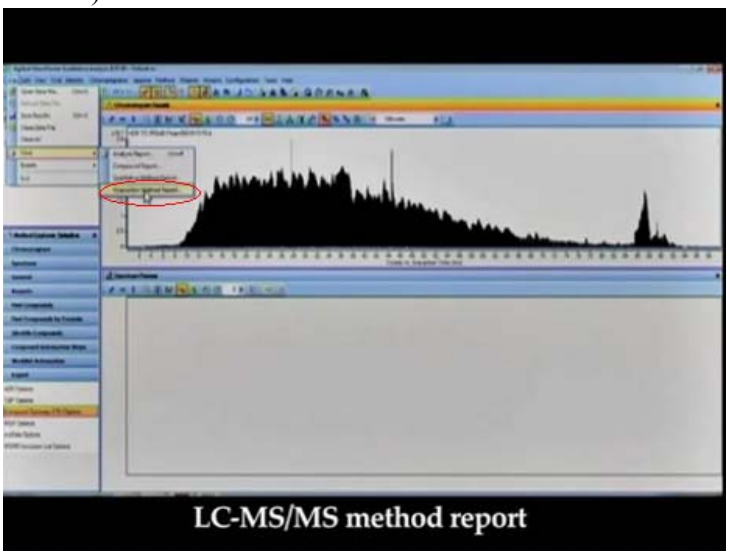


(Refer Slide Time 09:14)



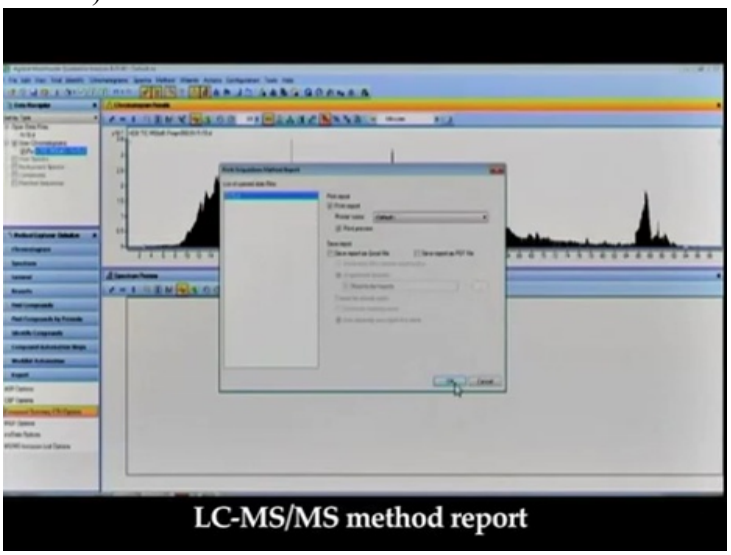
By clicking on the print option

(Refer Slide Time 09:16)



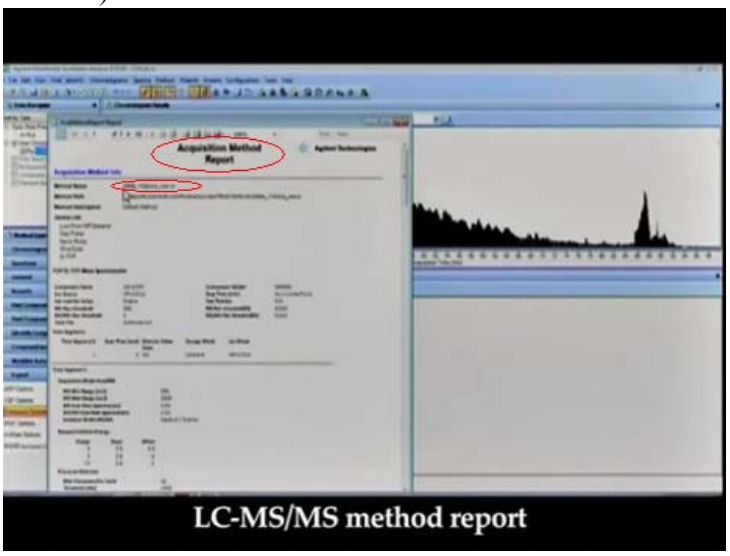
followed by Acquisition method report, we can get the

(Refer Slide Time 09:18)



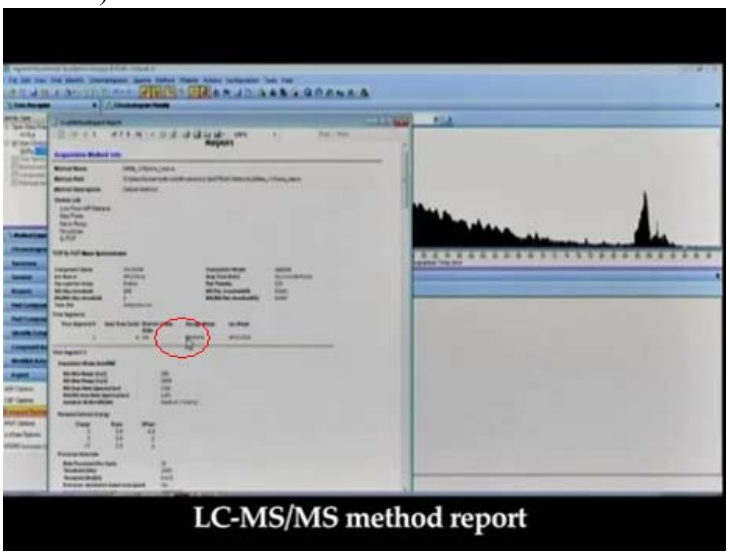
LC MS/MS method which was used for the data acquisition

(Refer Slide Time 09:31)



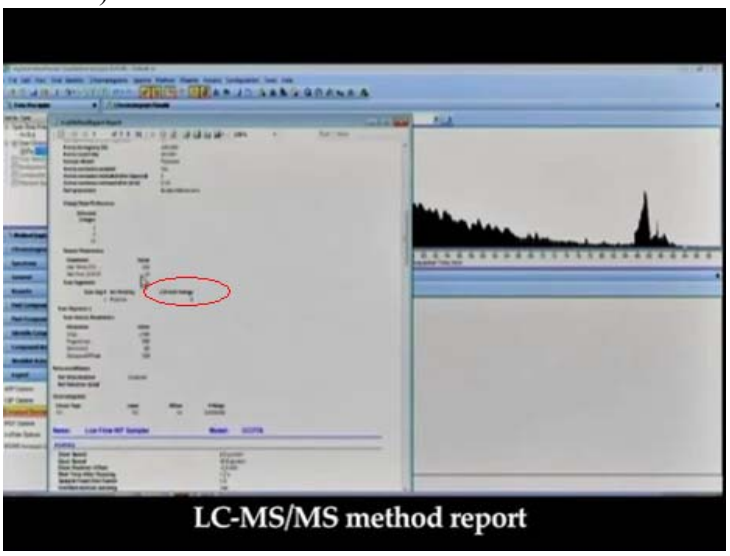
This acquisition method report contains details about method name, data acquisition mode,

(Refer Slide Time 09:36)



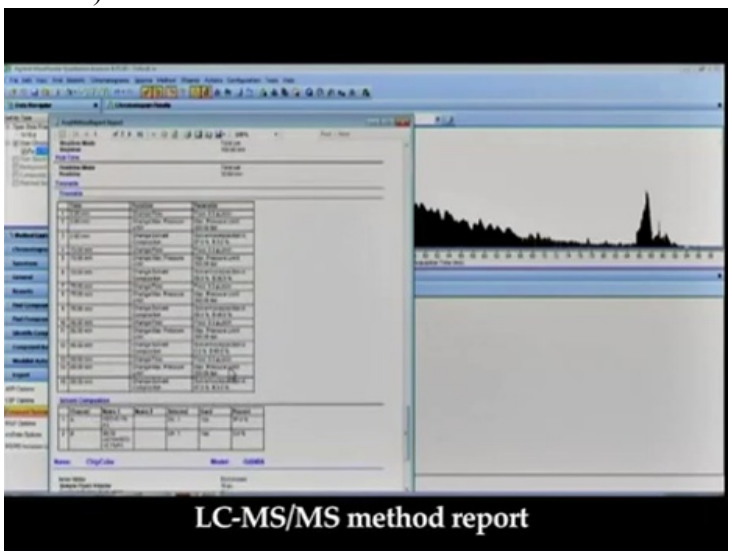
collision energy, gas temperature,

(Refer Slide Time 09:41)



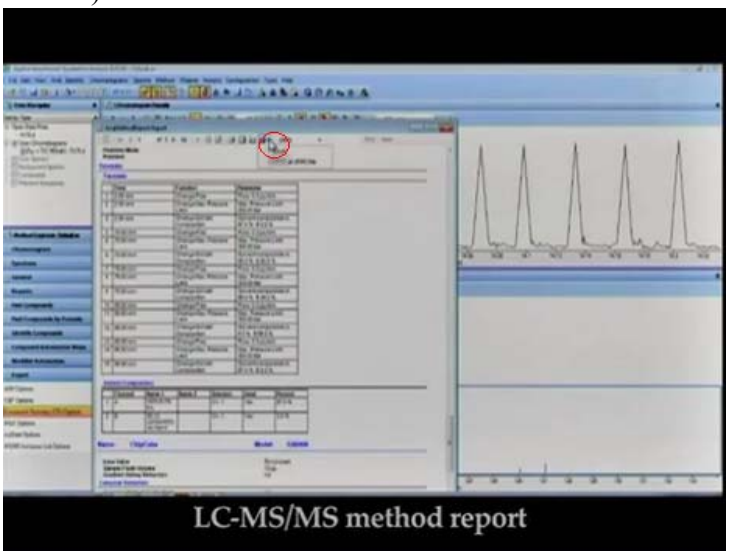
flow rate, Vcap voltage

(Refer Slide Time 09:43)



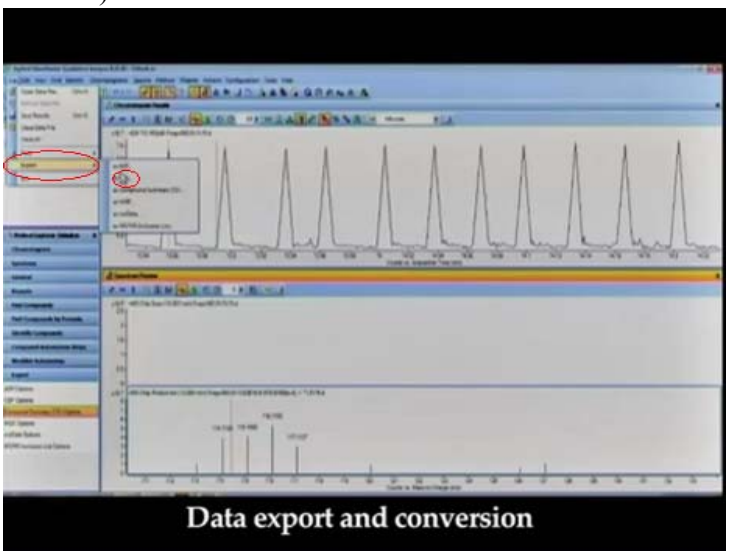
and solvent gradient details. We can save this file either

(Refer Slide Time 09:58)



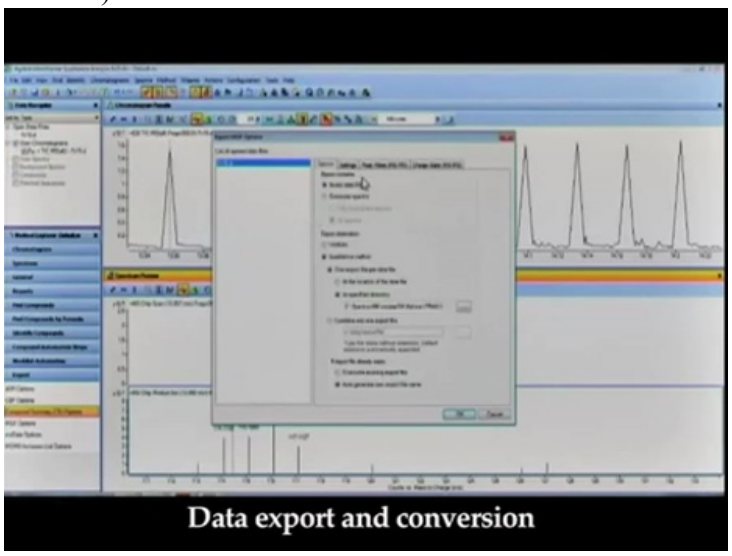
in pdf or dot (.) xi format.

(Refer Slide Time 10:10)



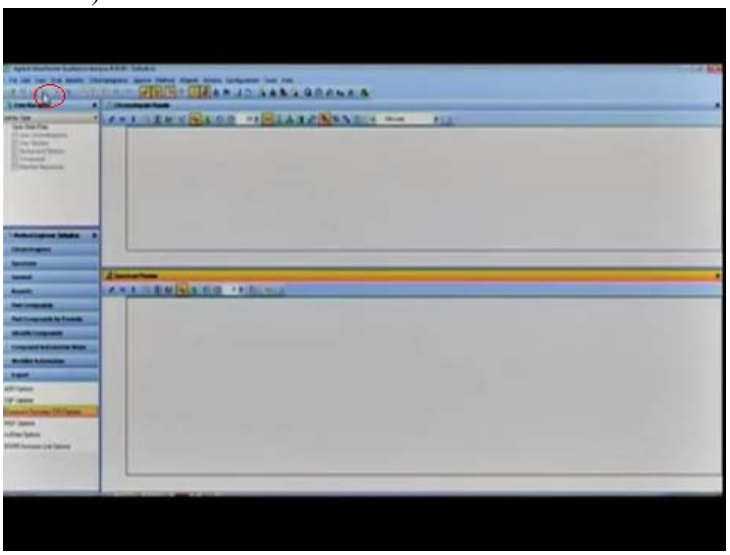
Click on export and you can convert the mass spec data file dot (.) d file into

(Refer Slide Time 10:12)



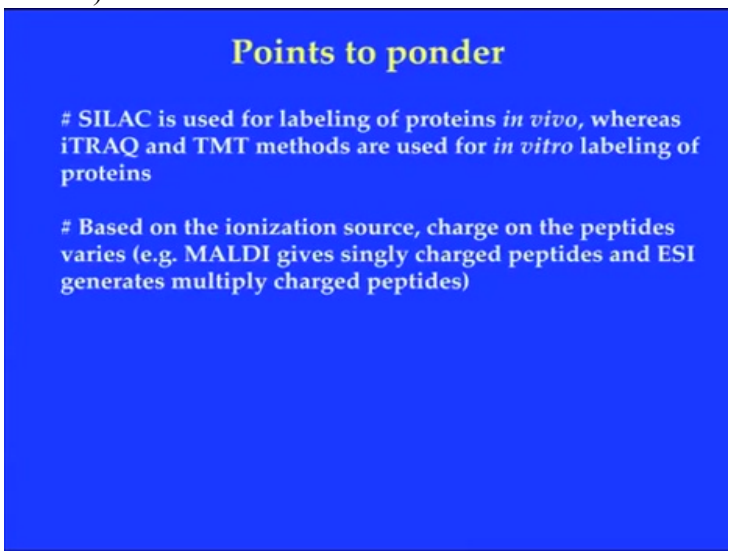
Mascot generate file dot (.) mgf file which can be used for analysis using Mascot.

(Refer Slide Time 10:24)

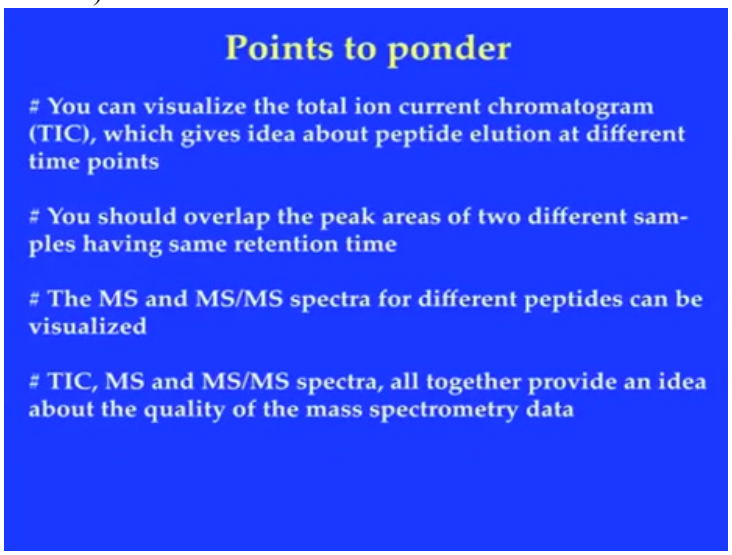


Finally, we can close the file by clicking the folder icon given in the qualitative analysis window.

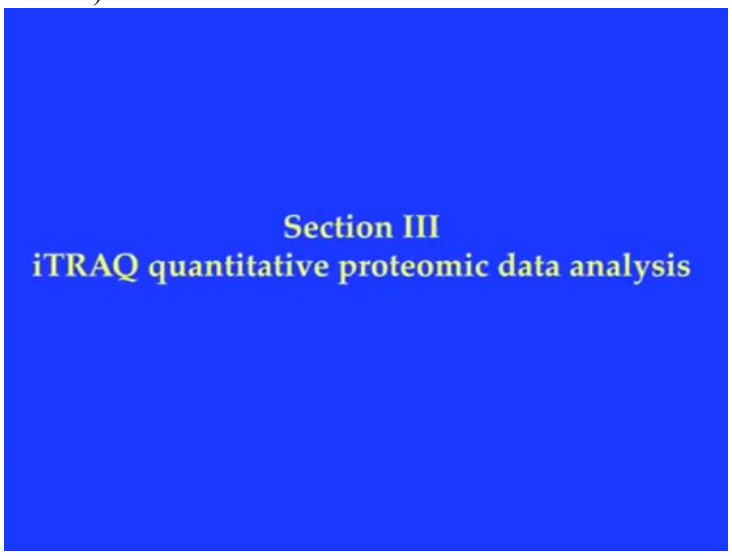
(Refer Slide Time 10:30)



(Refer Slide Time 10:35)



(Refer Slide Time 10:40)



(Refer Slide Time 10:56)

S. No	Software/database	Description	URL site
1	MASCOT	Search engine for protein identification using mass spectrometry data	http://www.matrixscience.com/
2	MS-Fit	Used for mining the sequence of the protein from MS data	prospector.ucsf.edu
3	SEQUEST	Used for interpretation of tandem mass spectra data for protein identification and amino acid sequence	http://fields.scripps.edu/sequest/
4	X!Tandem	Used for protein identification using tandem mass spectra data	http://www.thegpm.org/tandem/index.html
5	Sequit!	*De novo* sequencing of protein using tandem mass spectrum	http://www.sequit.org/
6	MSQuant	Quantitative proteomic information from MS and LC data	http://msquant.sourceforge.net/

Mass spectrometry has succeeded in identifying proteins based on the amino acid sequence information derived from the Tryptic digested peptides.

To analyze the data obtained from the mass spectrometry, sophisticated databases for protein identification were built and PMF, Peptide Mass Fingerprinting, PFF, Peptide Fragmenting Fingerprinting and MS/MS Ion Search are the foremost.

(Refer Slide Time 11:30)

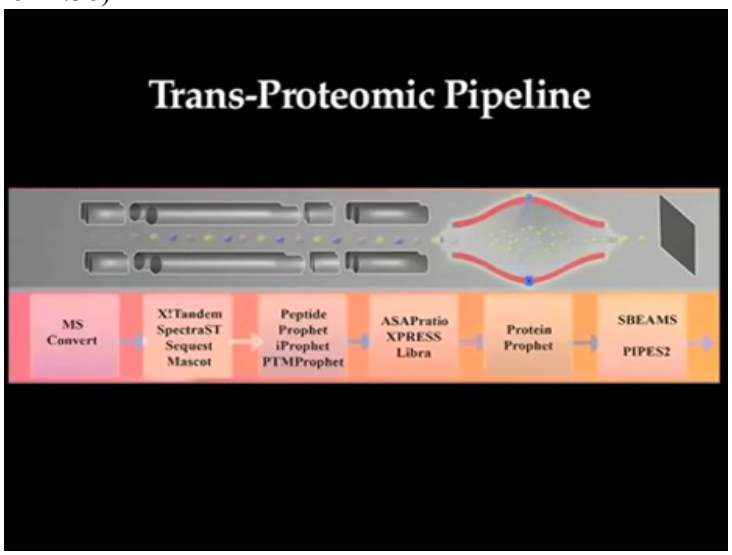
S. No	Software	Description	URL site
1	MapQuant	Used for MS quantification after making two dimensional map	http://arep.med.harvard.edu/MapQuant/
2	XCMS	Used for LC-MS data handling for relative quantization, visualization.	http://metlin.scripps.edu/xcms/
3	MsInspect	Used to combine the LC-MS and LC-MS/MS peptide data and also for peptide array generation	http://proteomics.fhcrc.org/CPL/msinspect/index.html
4	Mzmine	Mainly used for MS and LC-MS data processing purpose	http://mzmine.sourceforge.net/
5	Pep3D	Convert LC-MS or LC-MS/MS data into 2D map as m/z vs time	----------
6	SpecArray	Used for generation of expression peptide arrays from LC-MS data	tools.proteomecenter.org/SpecArray.php
7	Msight	It represent the data from both MS and separation steps such as chromatography, 2De etc	http://web.expasy.org/MSight/

In the last 2 decades, many algorithms, new tools were developed to process the large datasets.

Mascot is the widely used and trusted search engine for mass spectrometry data analysis.

It has comprehensive database covering various organisms and is compatible with all existing mass spectrometry softwares.

(Refer Slide Time 11:36)

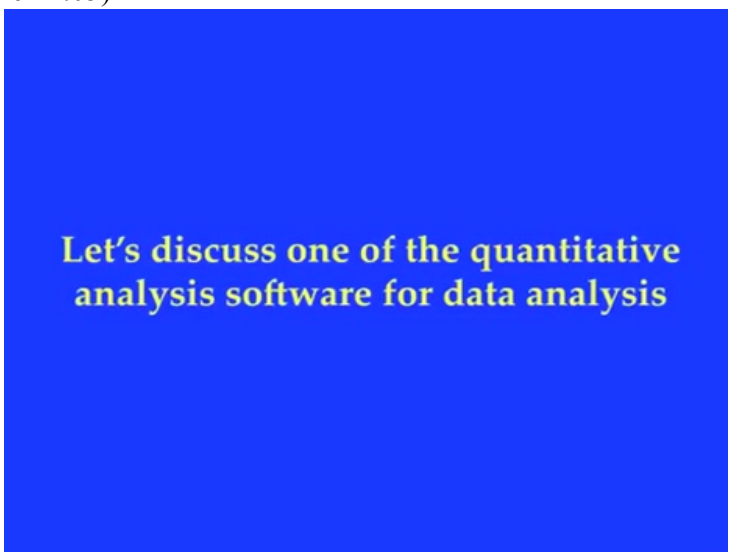


Besides SEQUEST, X! Tandem is also used for MS/MS data analysis.

In addition to vendor-specific tools and databases, there are many online open source data analysis software.

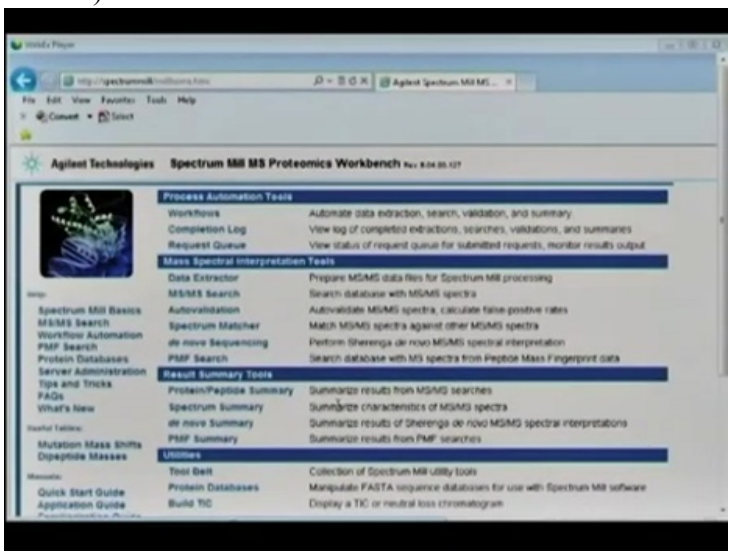
Apart from the softwares mentioned in the table, many open source software were introduced in the recent years such as Trans-Proteomic Pipeline, TPP, OpenMS, MaxQuant and many more.

(Refer Slide Time 12:05)



(Refer Slide Time 12:09)

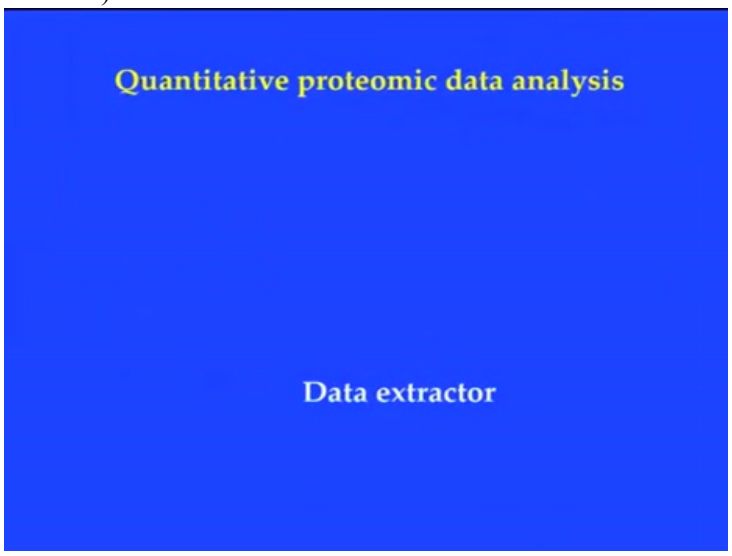
(Refer Slide Time 12:13)



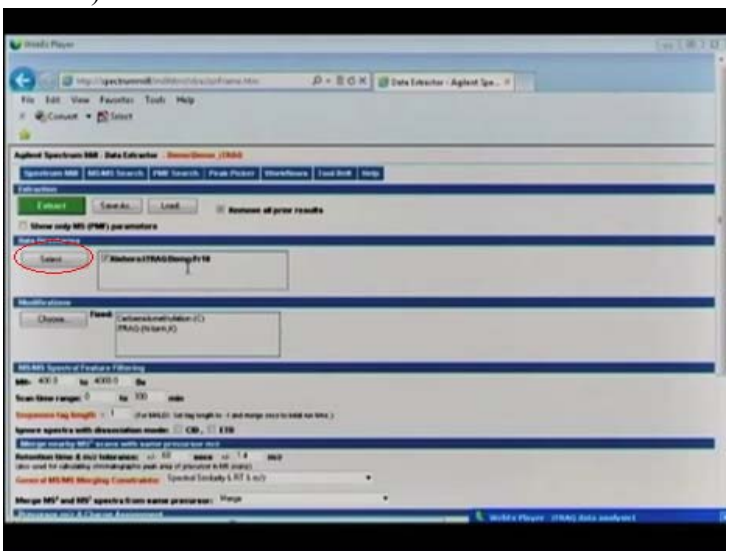
We will focus on the iTRAQ data analysis using Spectrum Mill software.

Spectrum Mill software, Spectrum Mill software is stand alone software used for LC MS/MS data analysis generated using Agilent LC MS/MS mass spectrometry. Here we can see the home page of Spectrum Mill having different tools for data analysis.

(Refer Slide Time 12:38)



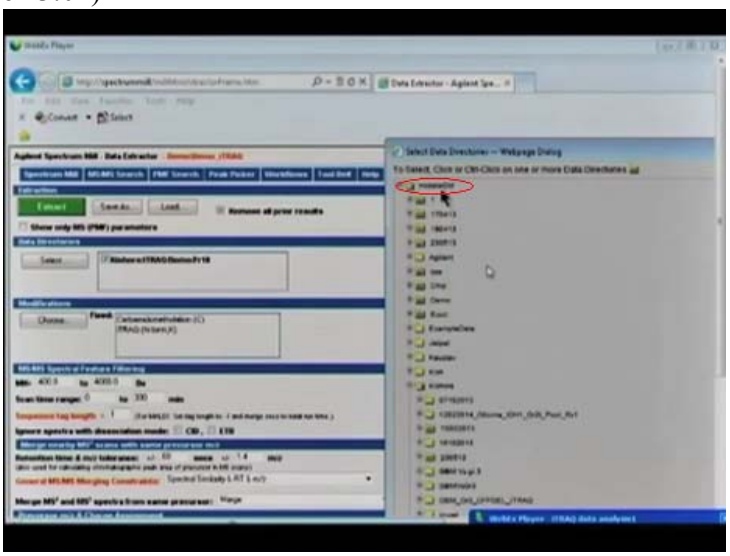
(Refer Slide Time 12:45)



Data extraction

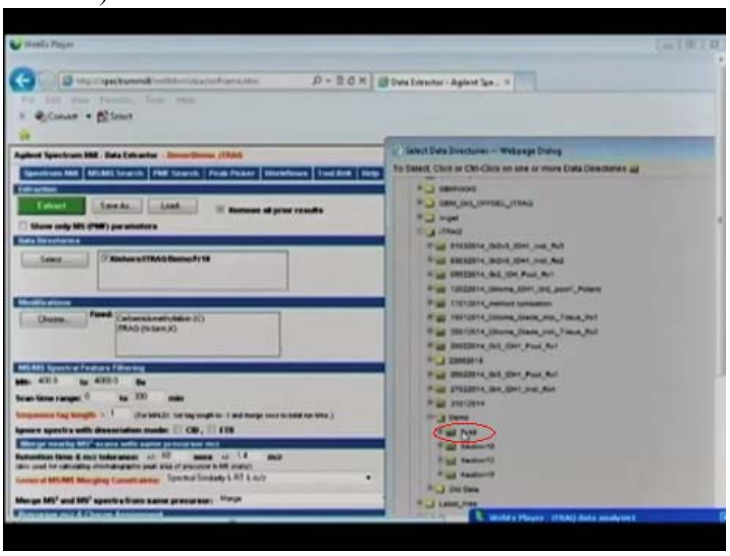
LC MS/MS data analysis can be started using data extractor tool where we can load our data files into the Spectrum Mill and prepare files for MS/MS analysis

(Refer Slide Time 13:04)



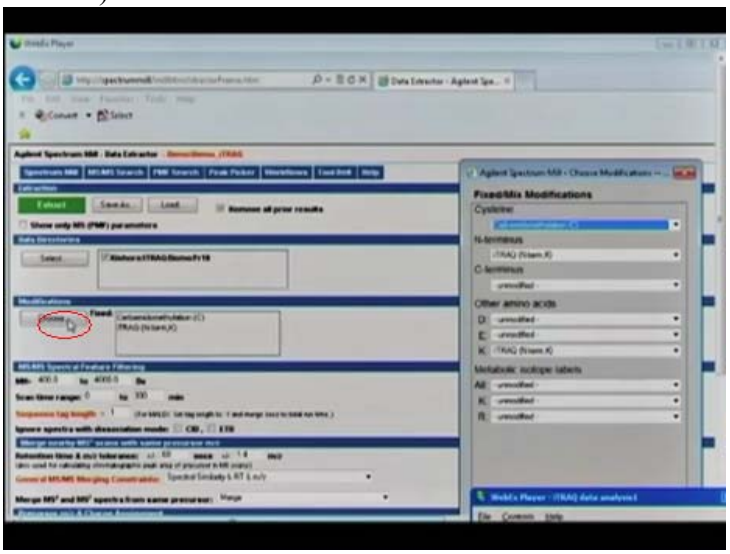
Select the file from the folder named as MSDataSN folder

(Refer Slide Time 13:17)



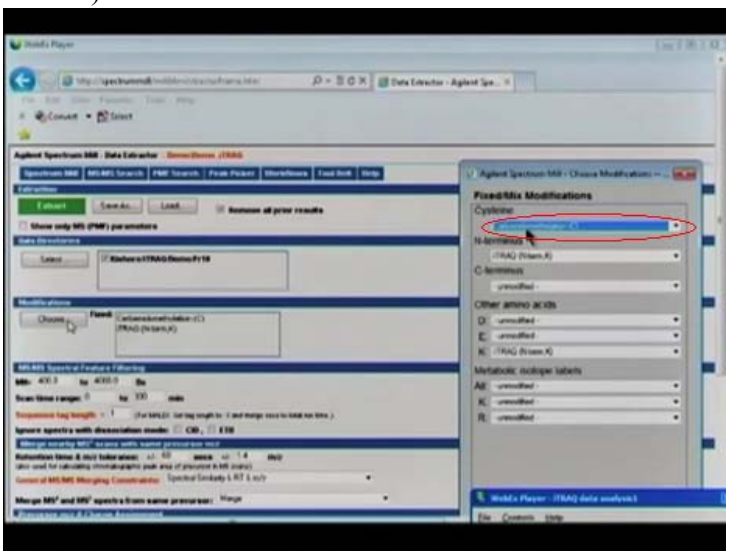
And upload the selected file into the software. As we are analyzing the iTRAQ data analysis

(Refer Slide Time 13:32)



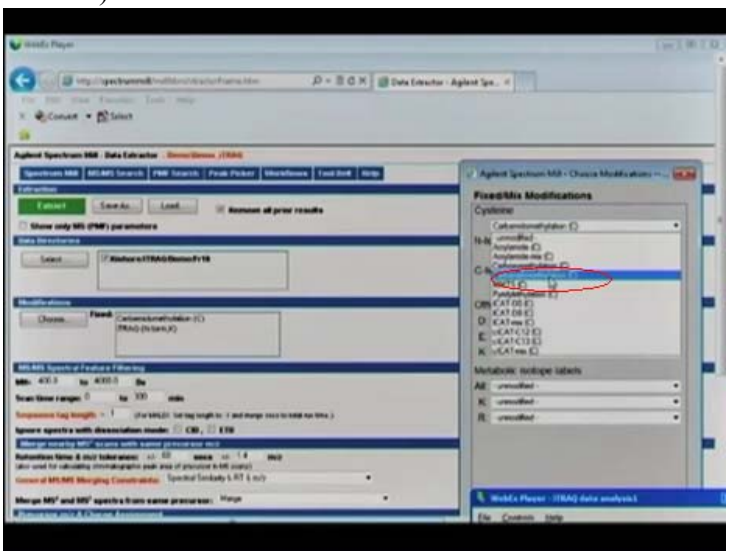
so before extracting the MS/MS data, we need to assign the modifications on the peptide.

(Refer Slide Time 13:42)

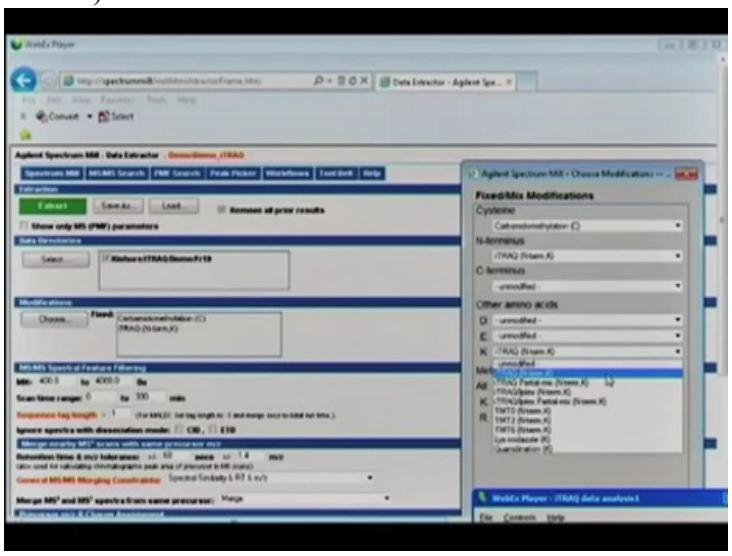


Carbamidomethylation on Cysteine and

(Refer Slide Time 13:47)

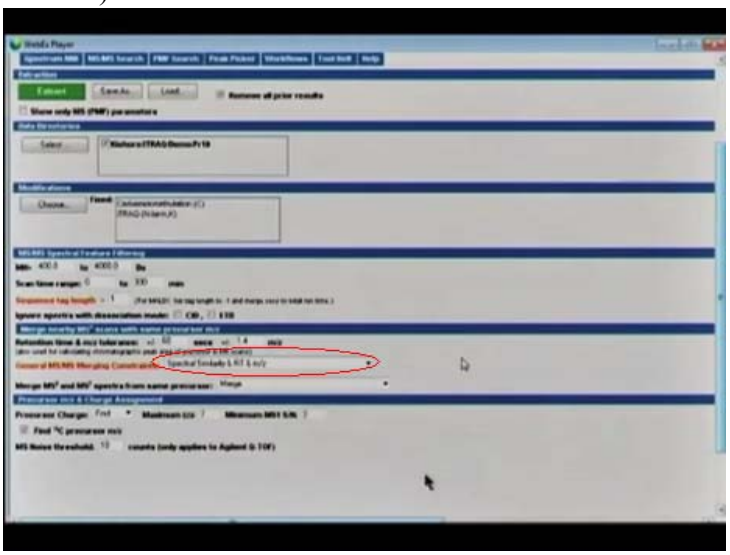


(Refer Slide Time 14:08)



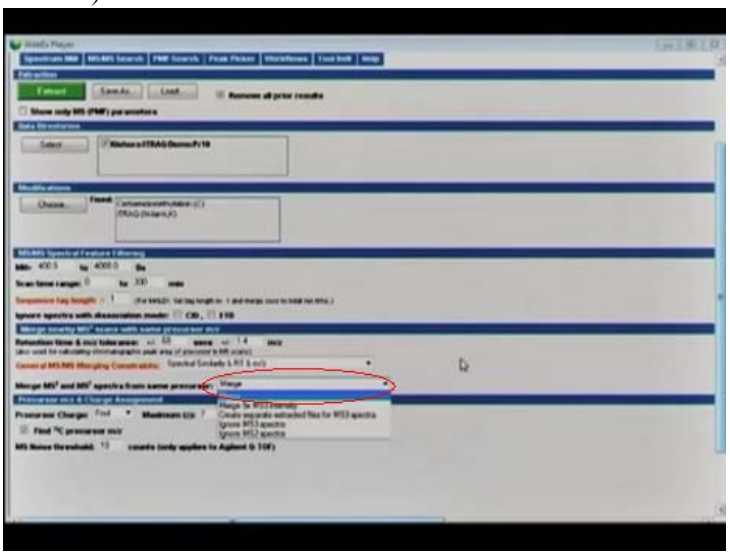
iTRAQ label on N Terminal amino acid and lysine.

(Refer Slide Time 14:33)



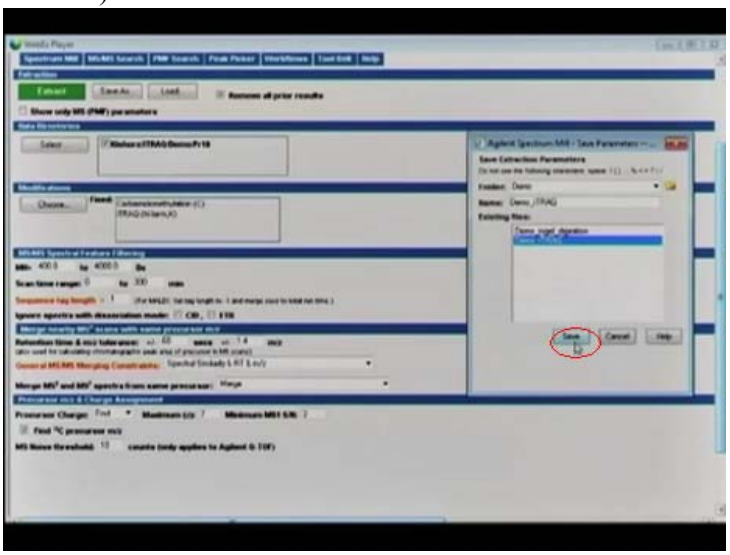
General MS/MS merging constraints should be assigned as spectral similarity and RT and m by z

(Refer Slide Time 14:42)



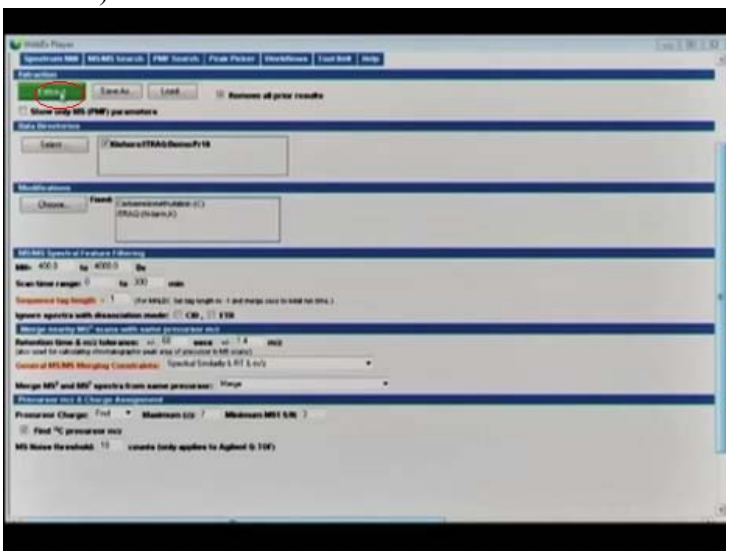
Further merge the MS2 and MS3 spectra from the same precursor.

(Refer Slide Time 15:05)



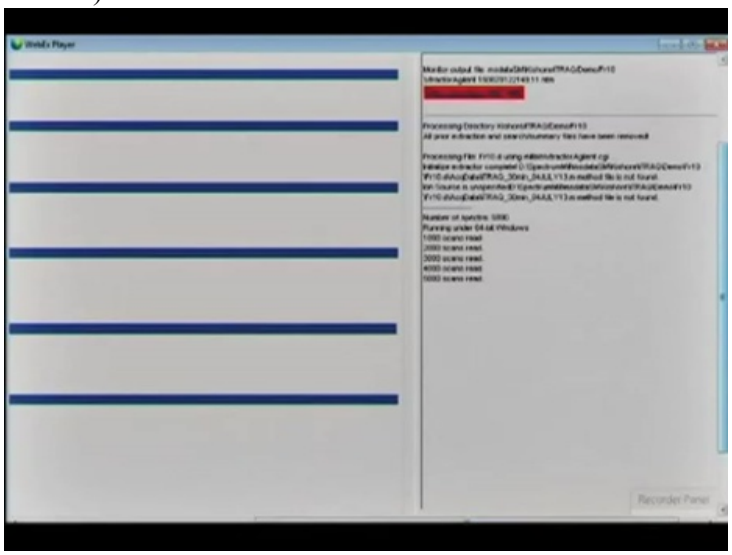
Keep the remaining parameters as default and save the method for future analysis.

(Refer Slide Time 15:25)



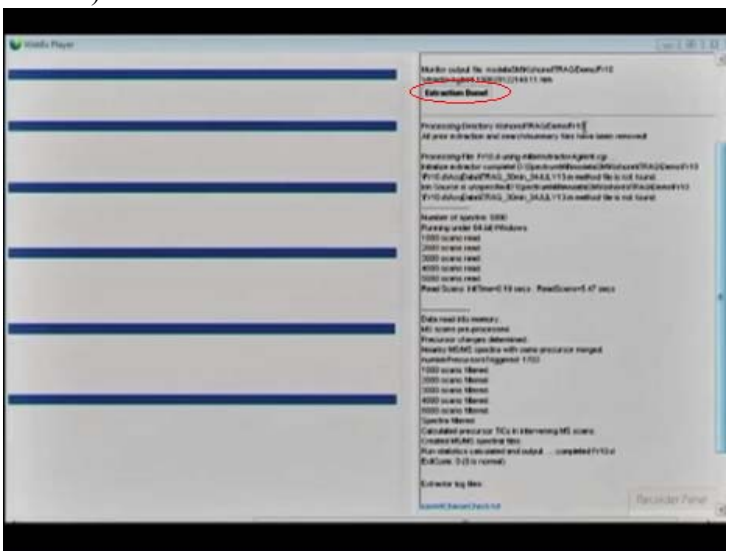
The extraction of MS/MS files

(Refer Slide Time 15:36)



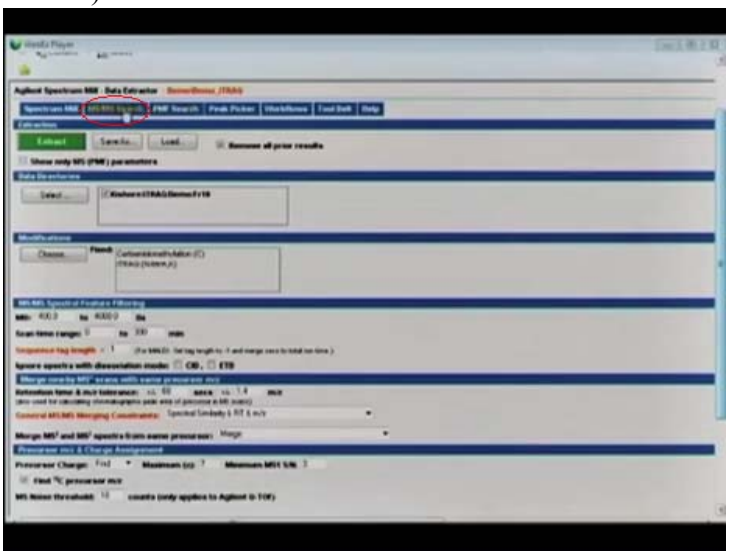
can be monitored in real time right side And once

(Refer Slide Time 16:23)

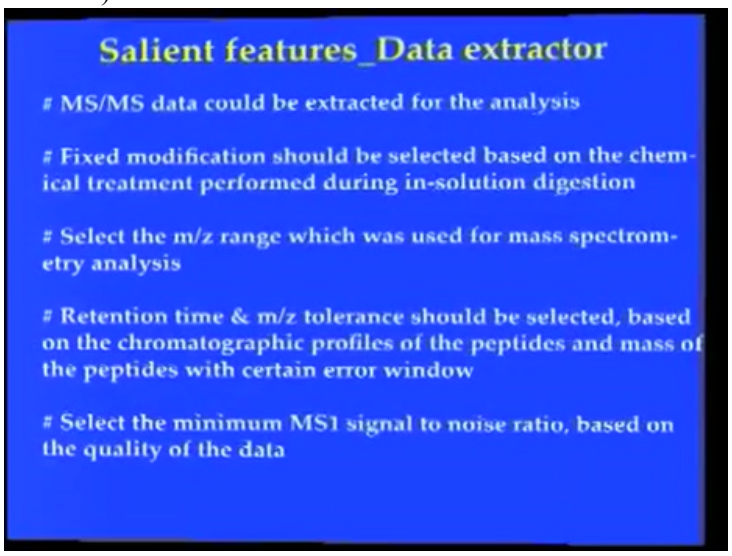


the extraction is done then we can proceed for MS/MS search.

(Refer Slide Time 16:37)

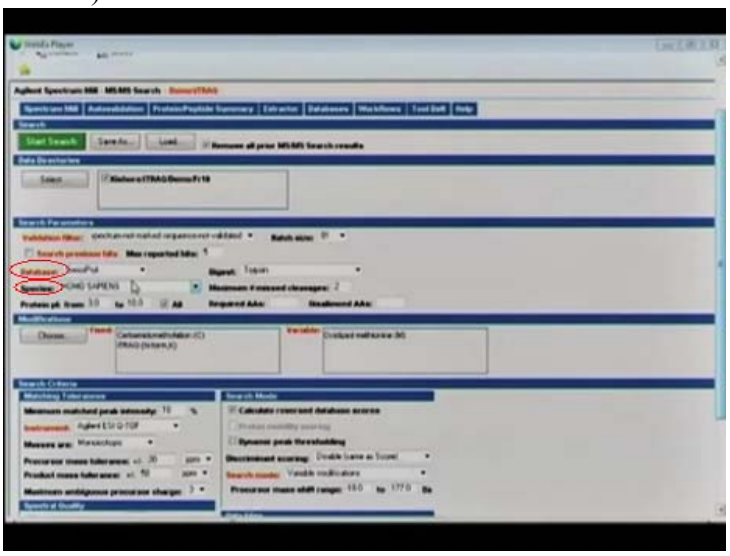


(Refer Slide Time 16:38)



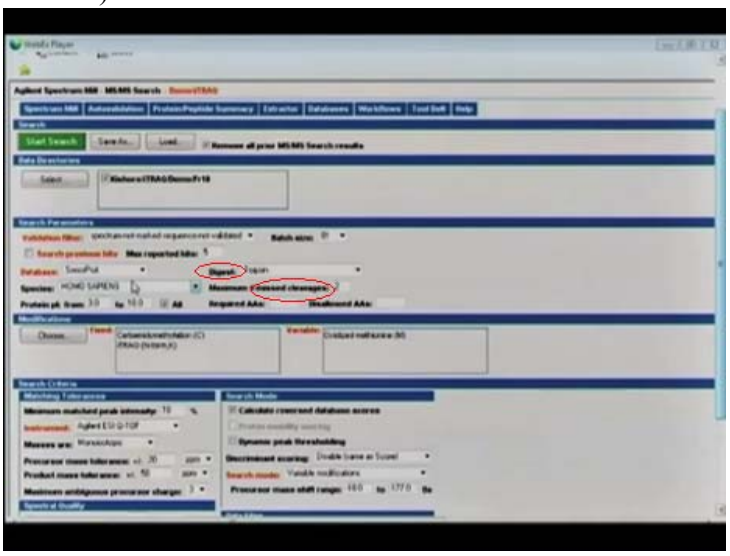
(Refer Slide Time 16:43)

(Refer Slide Time 17:07)



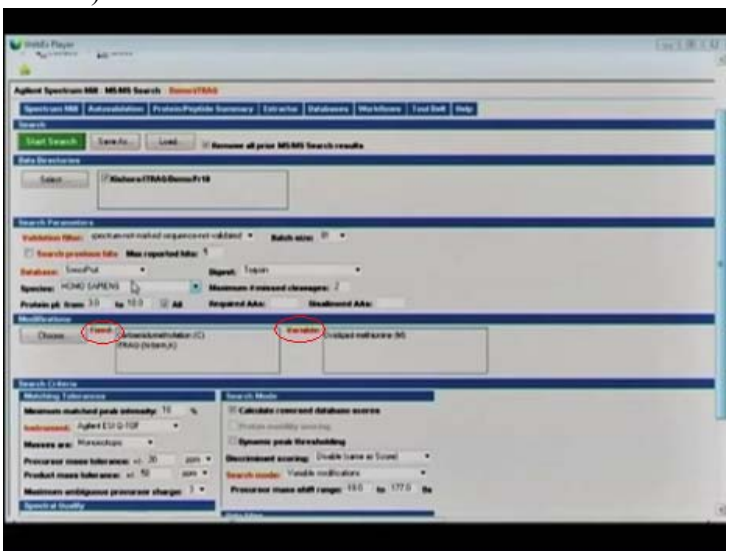
MS/MS ion search is the tool used for data search against the database few important parameters has to be assigned for the analysis such as database, species,

(Refer Slide Time 17:09)



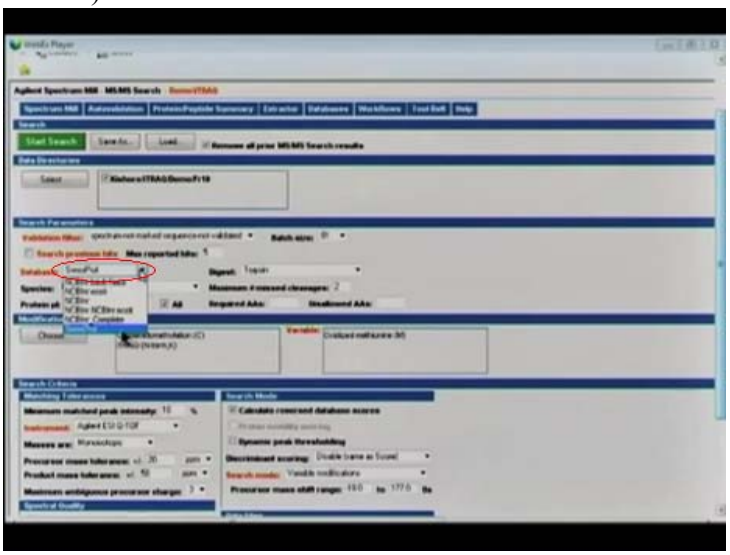
enzyme, used for digestion, missed cleavages,

(Refer Slide Time 17:11)



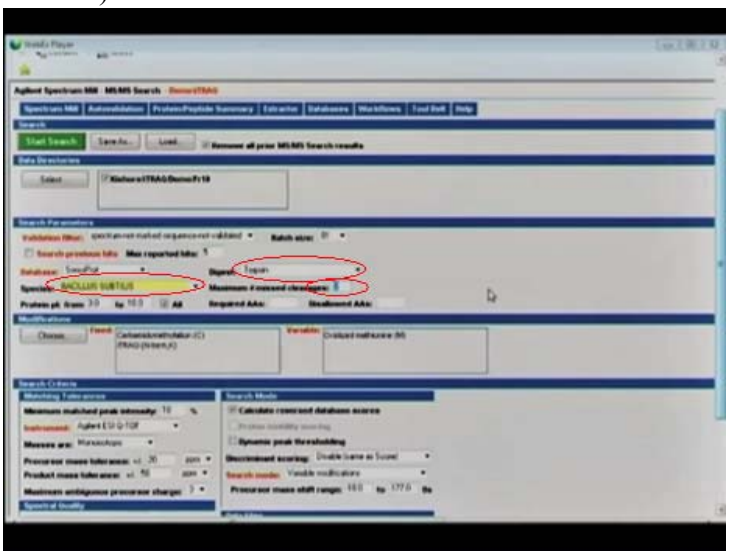
fixed modifications and variable modifications.

(Refer Slide Time 17:19)



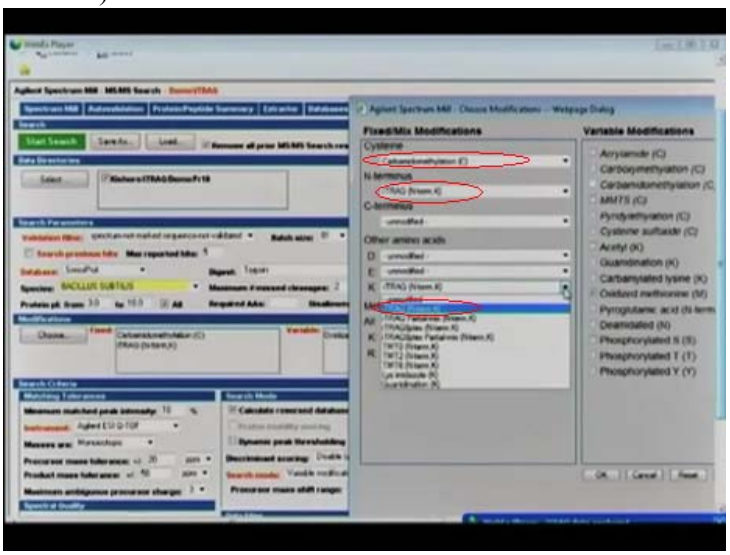
First we select the database as Swiss Prot.

(Refer Slide Time 17:58)



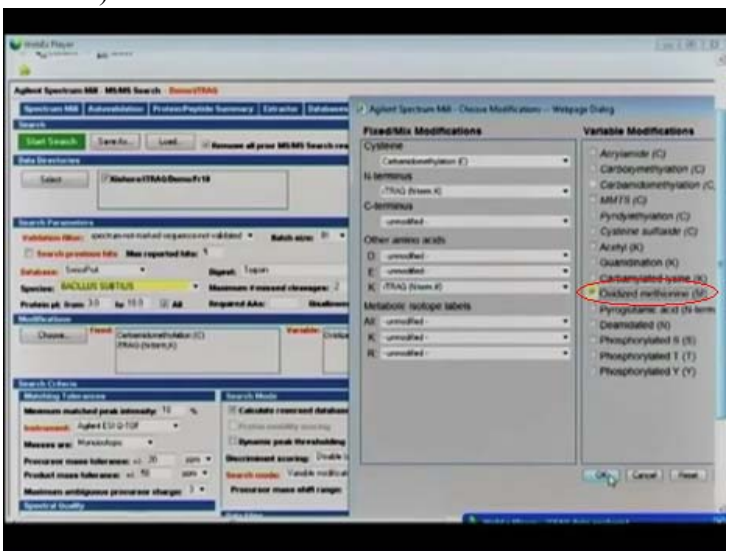
Enzyme as trypsin, Species as Bacillus subtilis, And missed cleavages as 2

(Refer Slide Time 18:20)



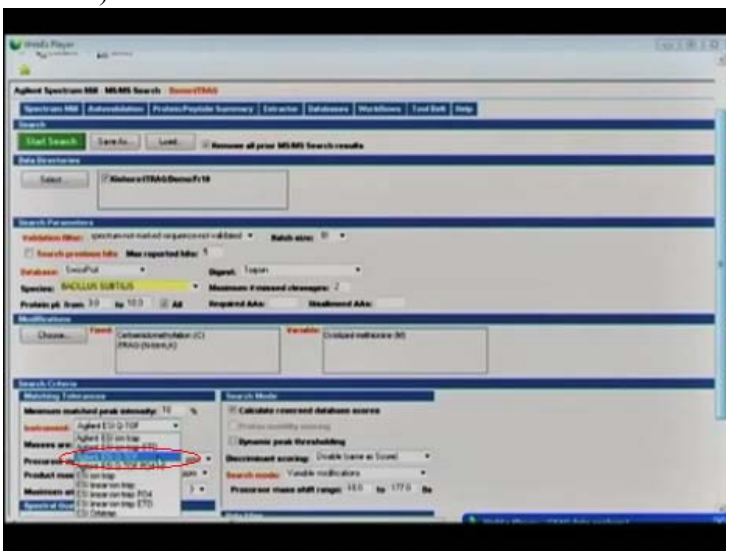
Fixed modification were assigned as Carbamidomethylation on Cysteine and iTRAQ label on N Terminal amino acid

(Refer Slide Time 18:24)



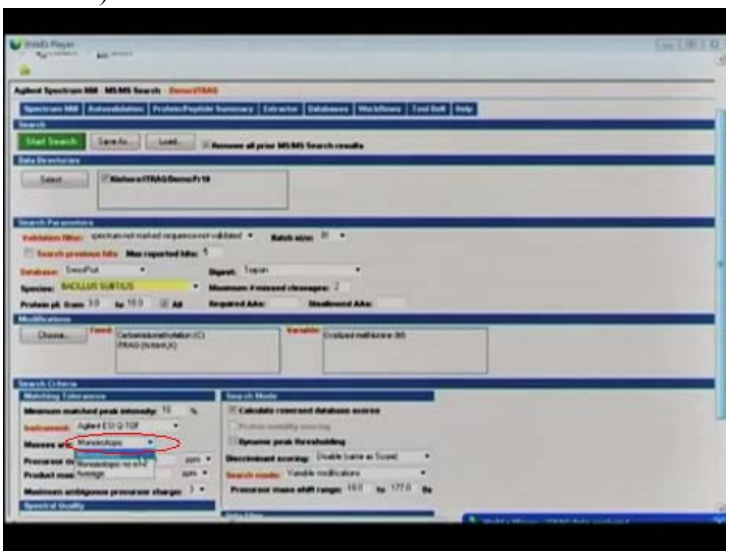
Oxidation on methionine as variable modification

(Refer Slide Time 18:36)



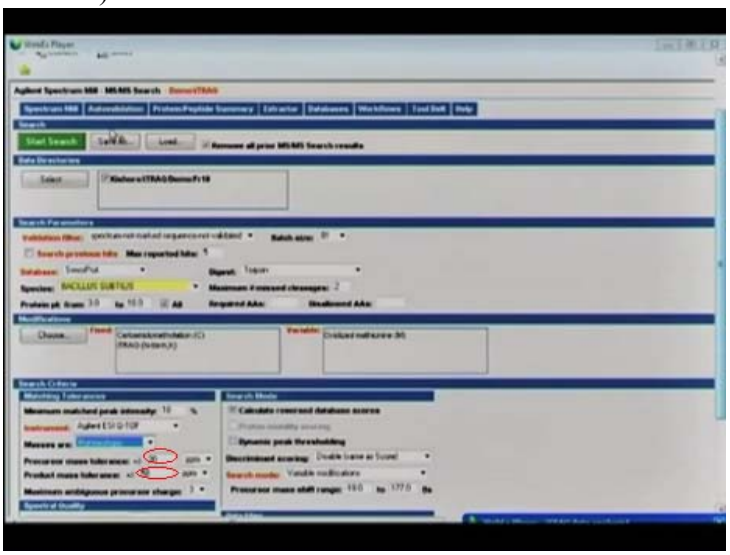
Instrument is selected as Agilent ESI Q-TOF

(Refer Slide Time 18:39)



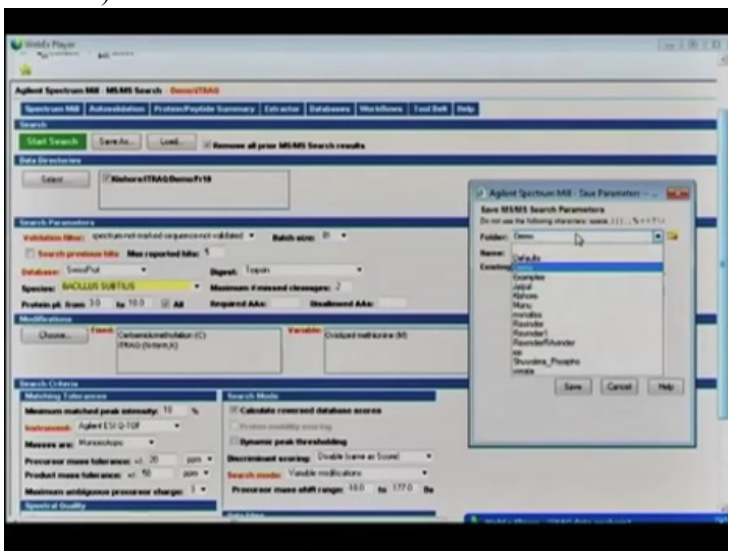
Masses as monoisotopic

(Refer Slide Time 18:46)

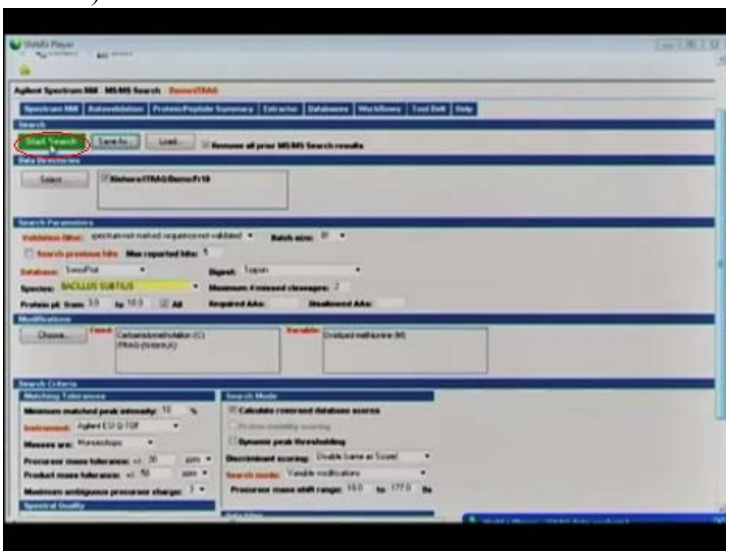


Precursor mass tolerance as 20 ppm and product mass tolerance as 50 ppm which is quite good to proceed for data analysis.

(Refer Slide Time 18:55)



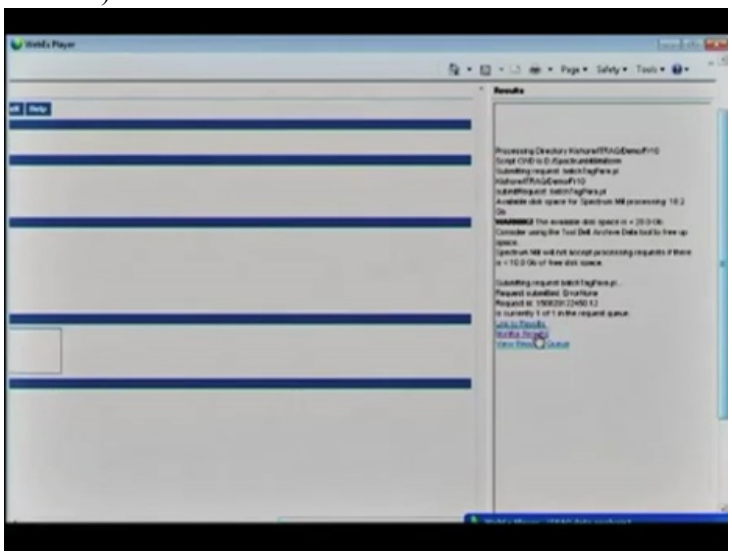
(Refer Slide Time 19:06)



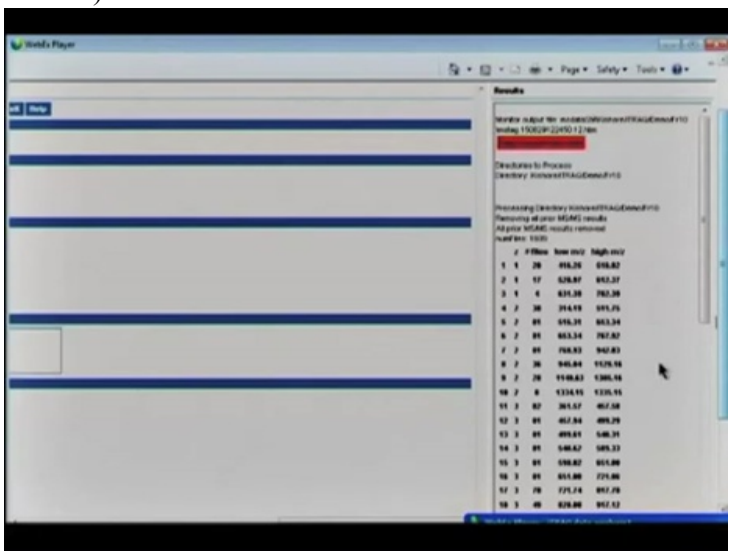
Once all the parameters are assigned

Start search

(Refer Slide Time 19:09)



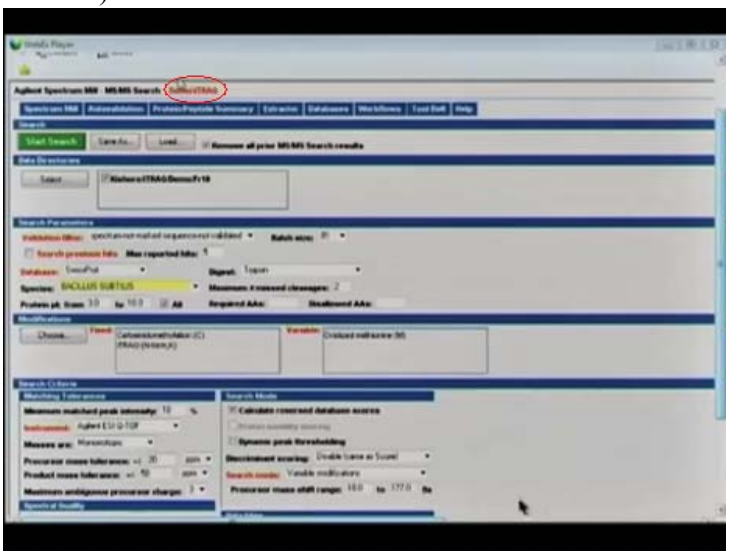
(Refer Slide Time 19:13)



And the search results can be monitored in the real time at the right side tab.

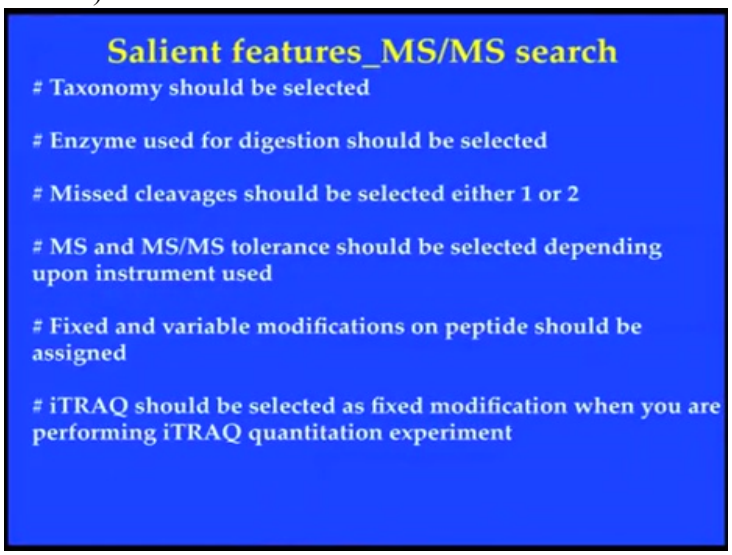
Once the MS/MS search is done, we can validate the data for statistically analysis using

(Refer Slide Time 20:21)

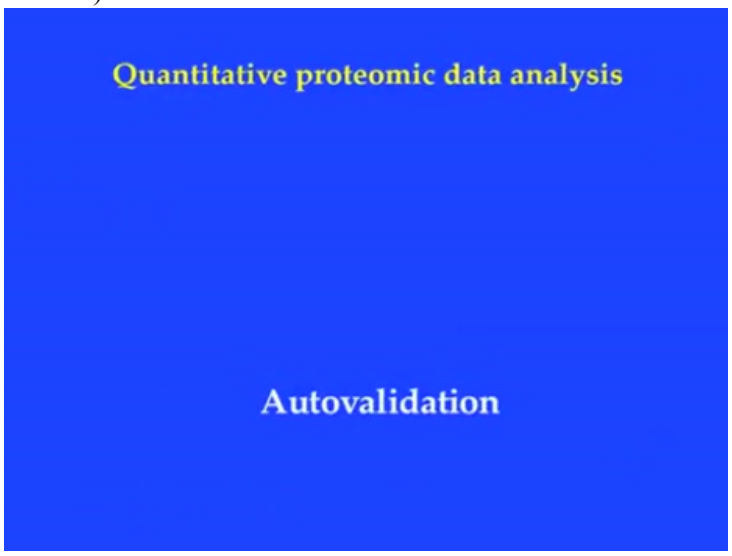


Auto-validation tab

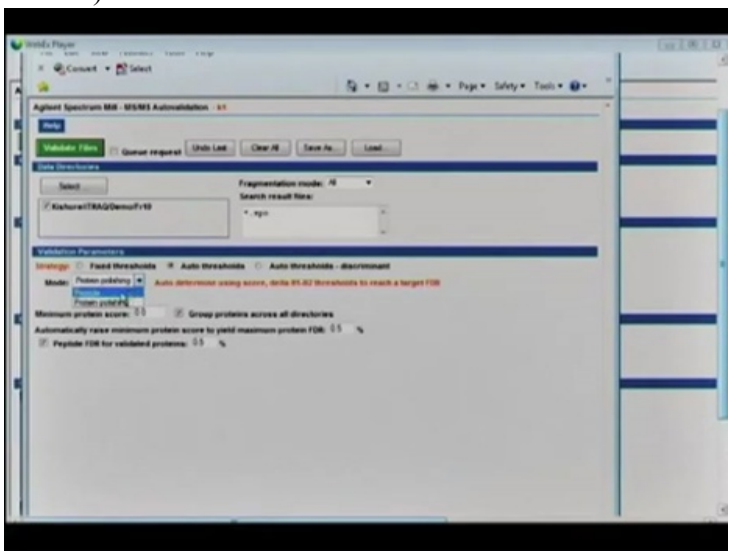
(Refer Slide Time 20:22)



(Refer Slide Time 20:27)



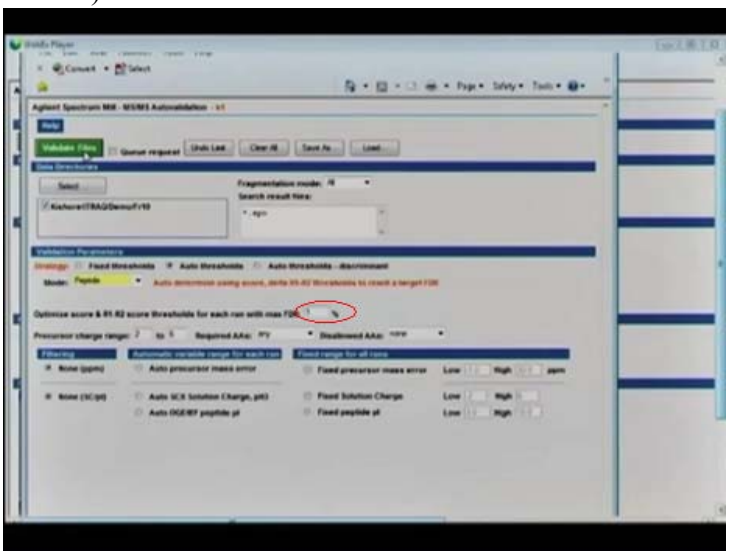
(Refer Slide Time 20:50)



Auto-validation, in Auto-validation both peptide and protein validation is needed to be performed

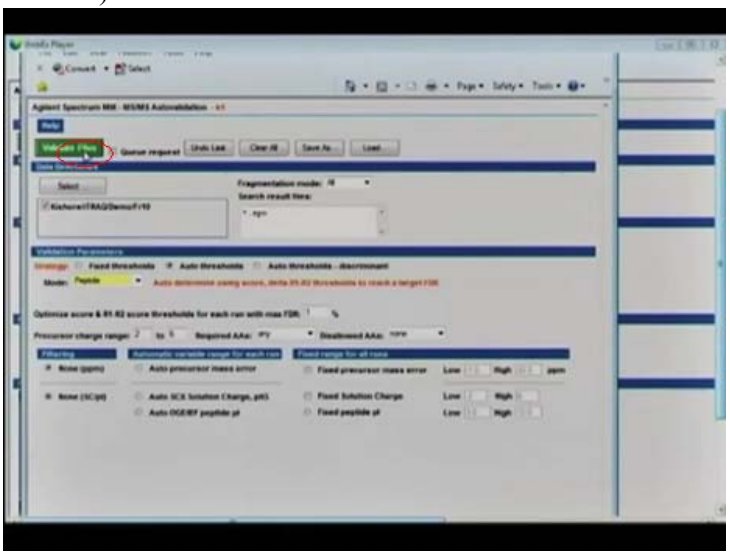
First we will do the peptide validation by selecting peptide in mode tab

(Refer Slide Time 20:58)



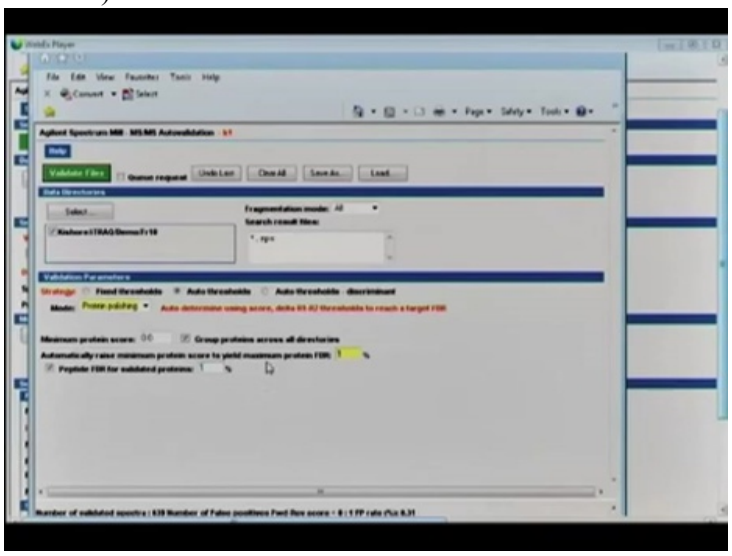
1% and FDR and remaining parameters are set as default.

(Refer Slide Time 21:05)



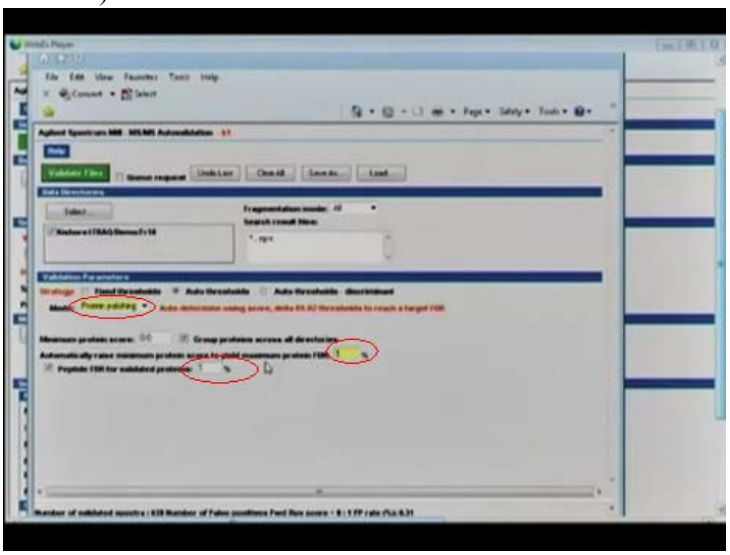
Click on Validate the files

(Refer Slide Time 21:40)



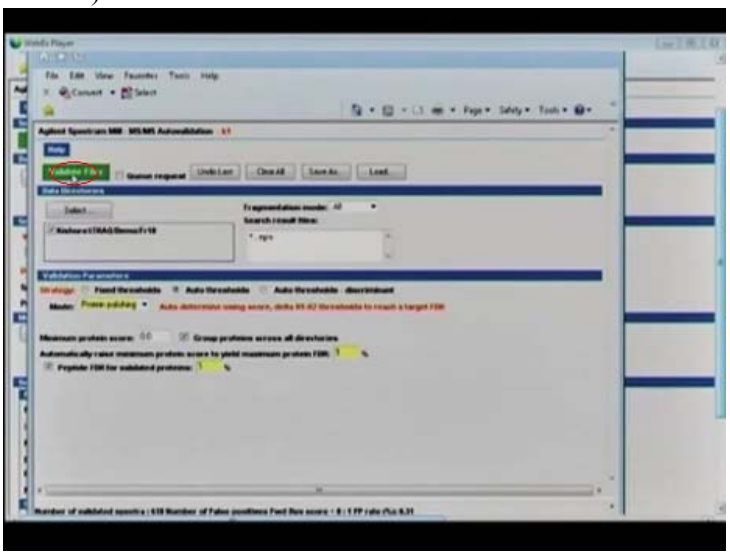
Similarly protein validation is done by selecting protein polishing in mode tab

(Refer Slide Time 21:42)



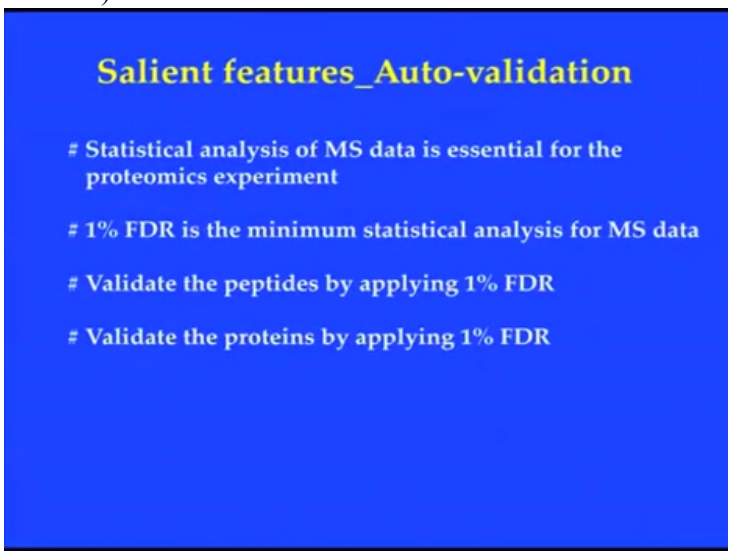
1% FDR is assigned and remaining parameters are set as default.

(Refer Slide Time 21:43)

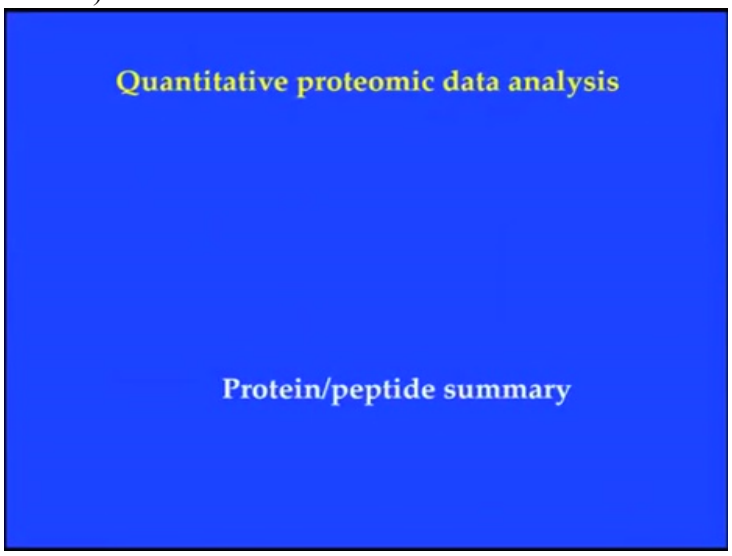


Now validate files

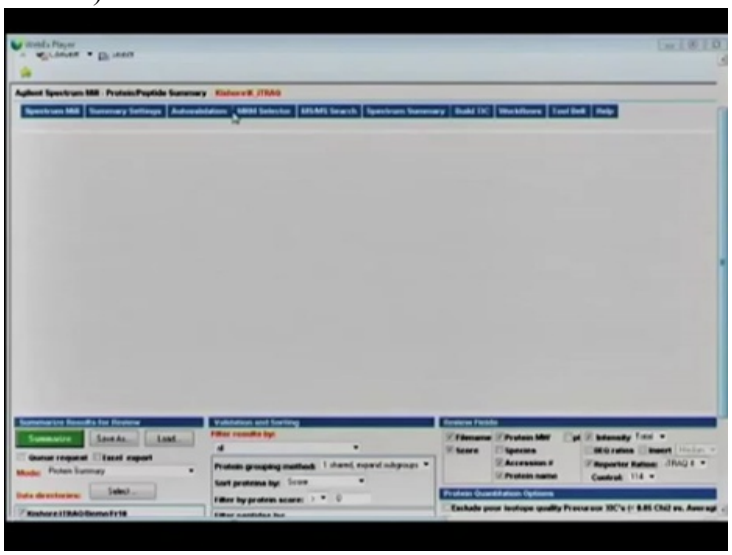
(Refer Slide Time 21:52)



(Refer Slide Time 21:57)

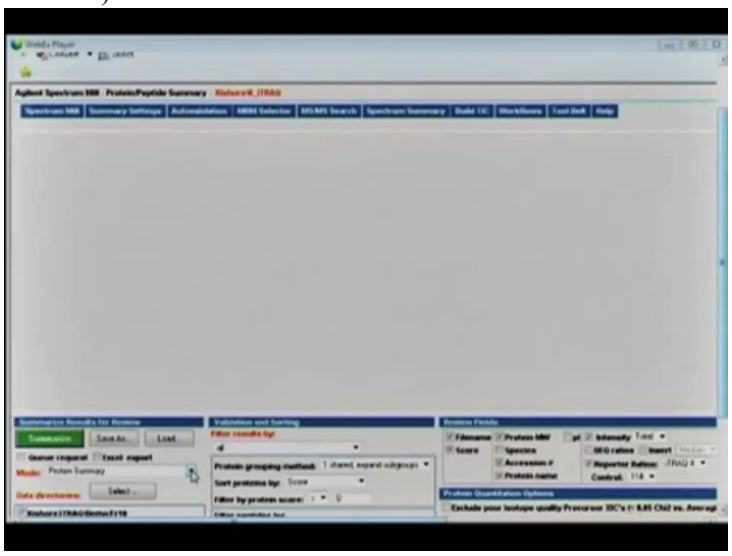


(Refer Slide Time 22:05)



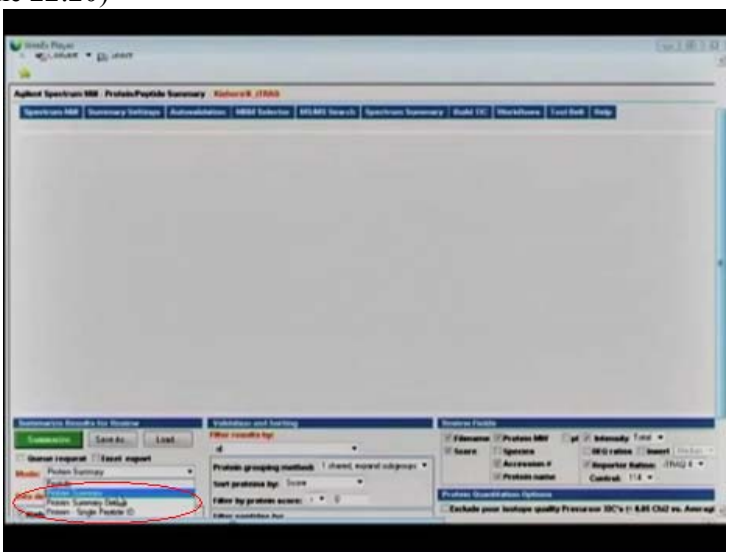
Protein peptide summary

(Refer Slide Time 22:10)



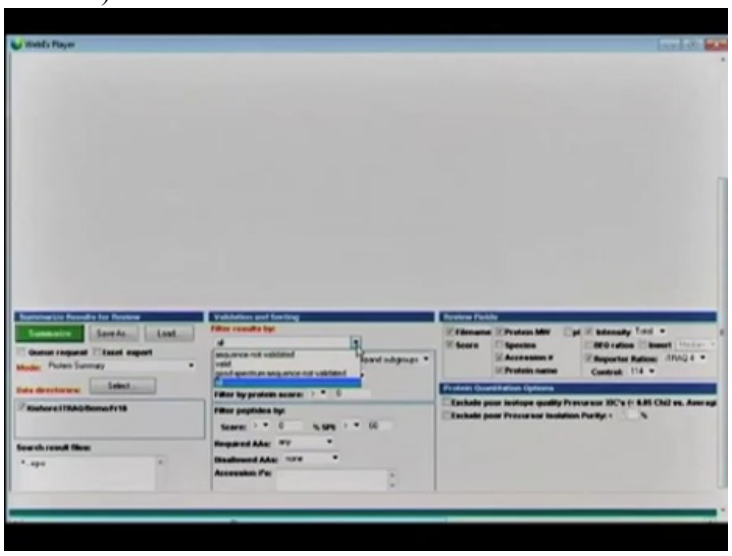
Protein peptide summary is the tool to visualize the results obtained from the MS/MS analysis.

(Refer Slide Time 22:20)



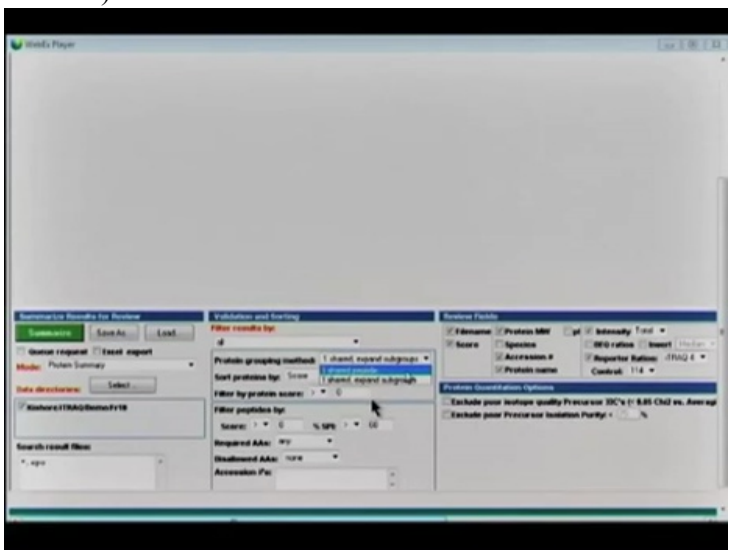
In the mode tab, we have different options to display the results

(Refer Slide Time 22:39)

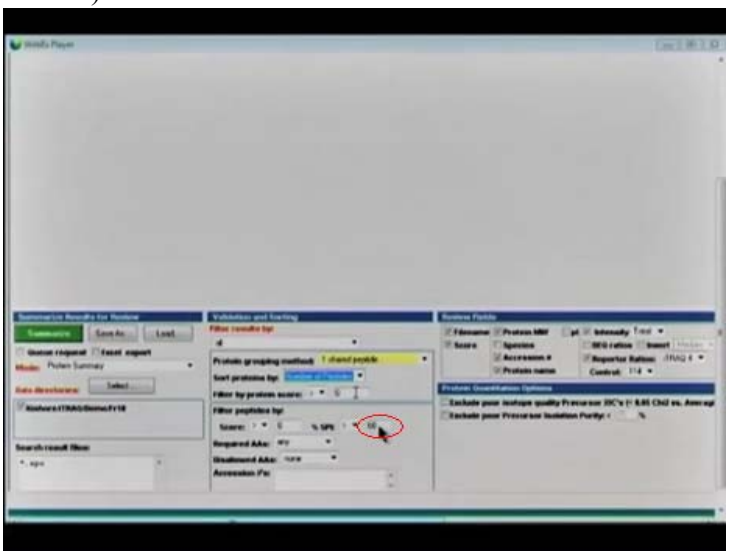


We need to assign few parameters prior to displaying the results such as Click valid in filter result tabs to display only the statistically significant protein list.

(Refer Slide Time 22:50)

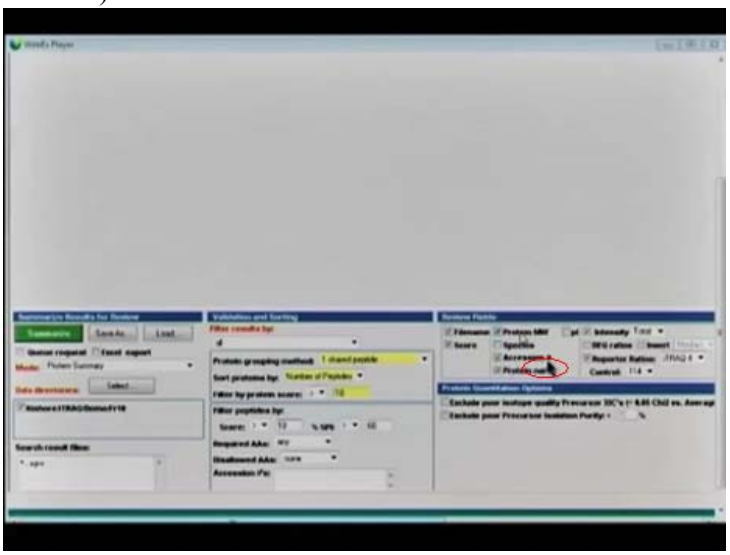


(Refer Slide Time 23:00)



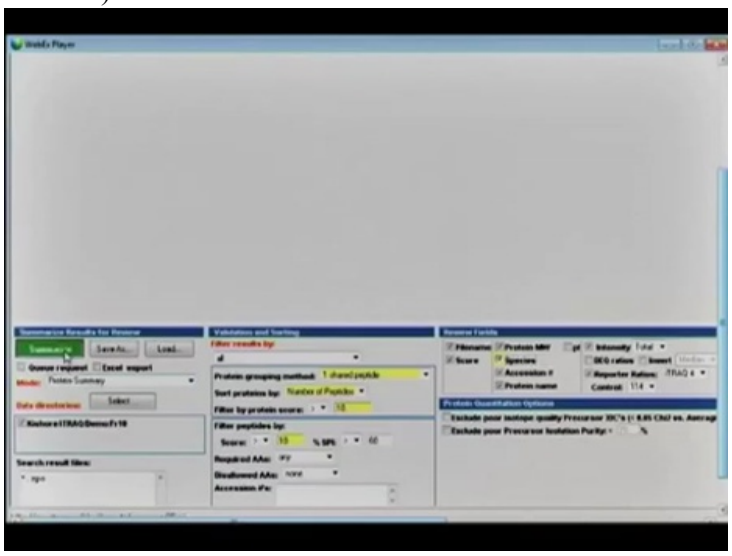
SPI was assigned as 60 which is an optimum value for MS/MS analysis

(Refer Slide Time 23:23)



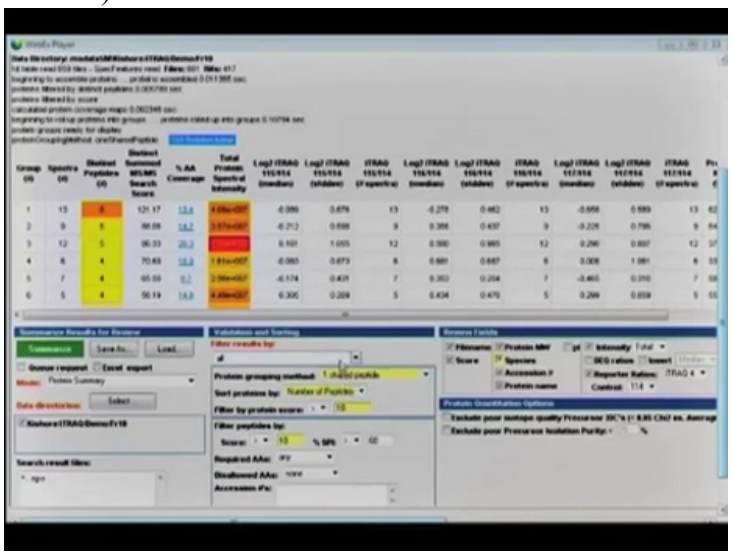
In the review field we can select species, molecular weight, score, peptide information, Reporter ion density, session number and many other parameters

(Refer Slide Time 23:34)



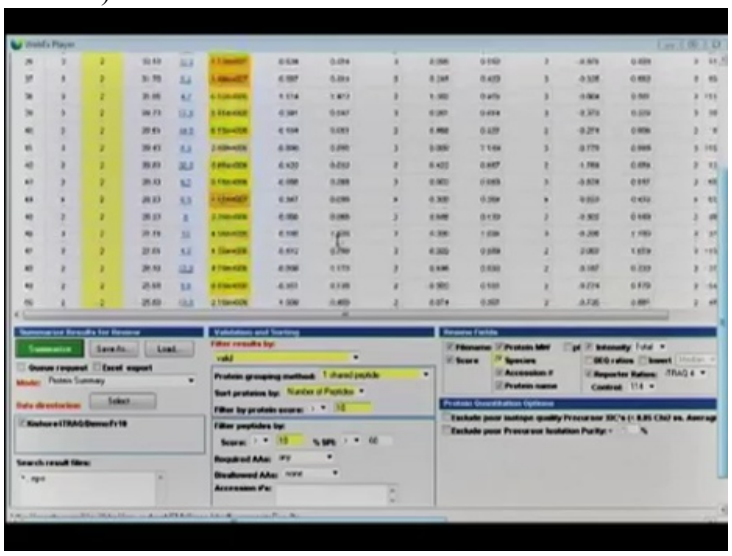
Keep remaining parameters as default

(Refer Slide Time 23:47)

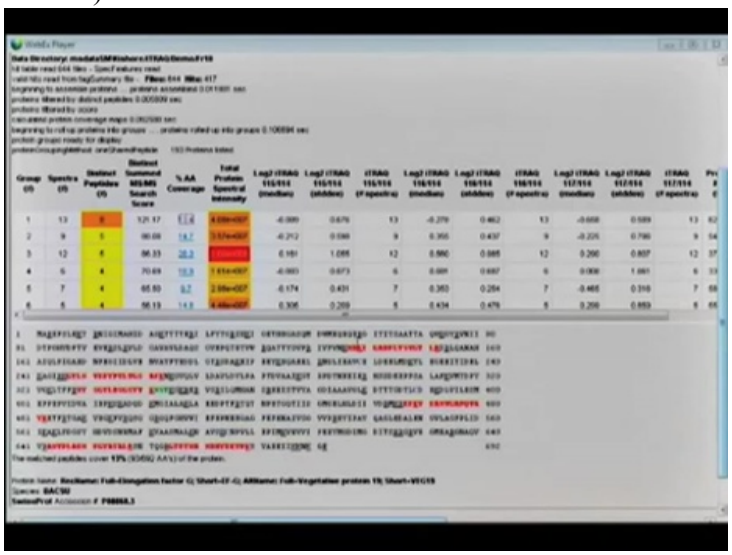


and summarize the results. It will show the number of proteins identified or quantified from the analyzed sample

(Refer Slide Time 23:59)

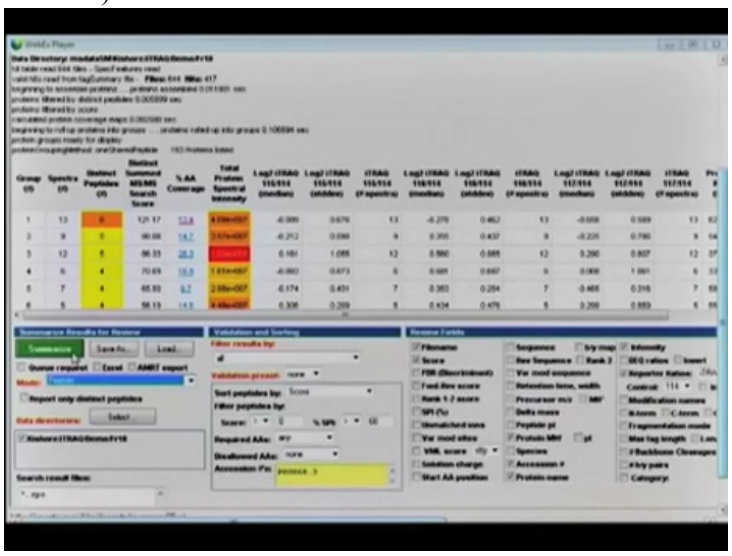


(Refer Slide Time 24:15)



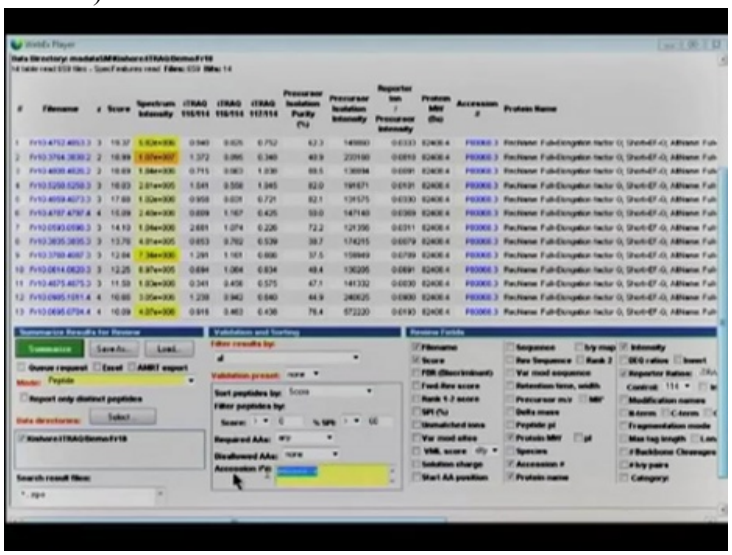
in the coverage tab the highlighted sequence will represent the peptides detected in mass spectrometry and remaining sequences will appear in black color.

(Refer Slide Time 24:47)



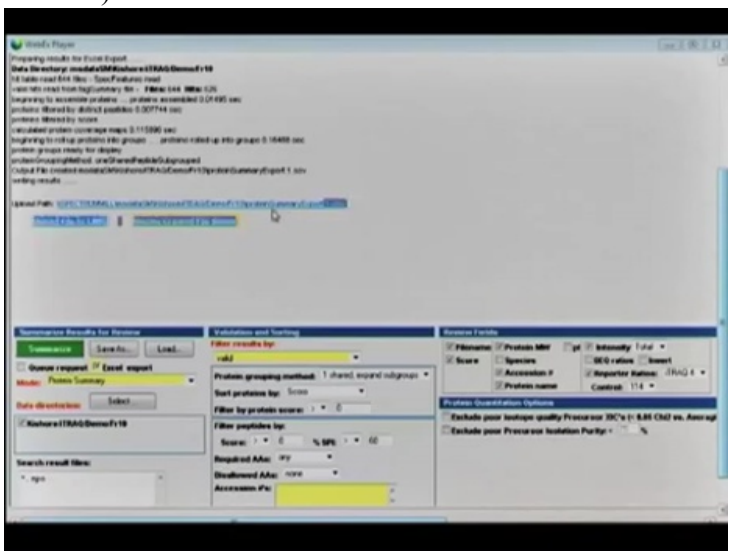
It also displays the results with all the selected parameters

(Refer Slide Time 24:50)

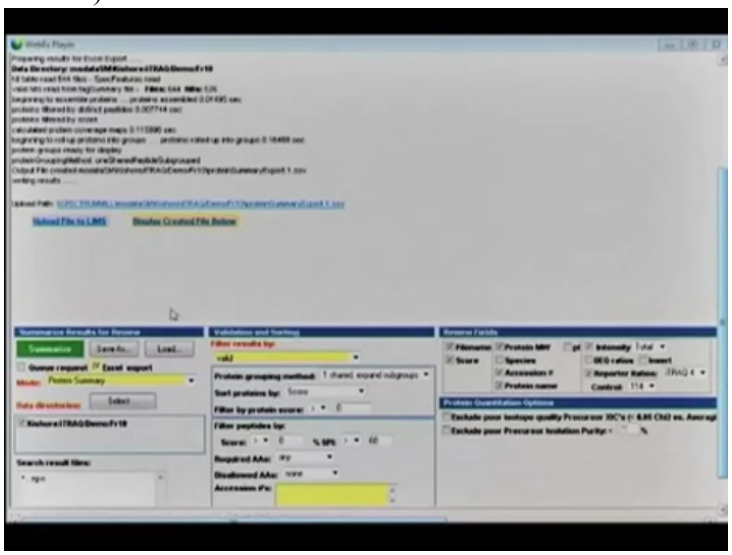


such as protein name, species, unique peptides, score and intensity of the reporter ion, accession number and many more. Further we can summarize the results based on the peptides and we can export or save the results for further analysis.

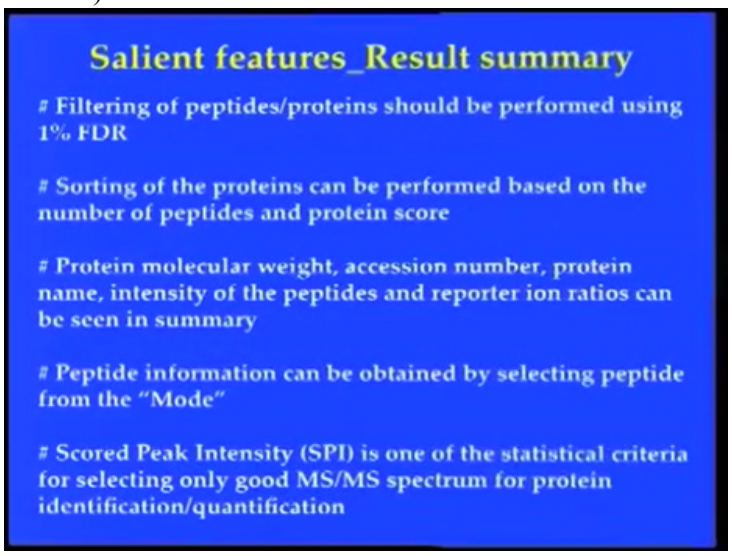
(Refer Slide Time 25:29)



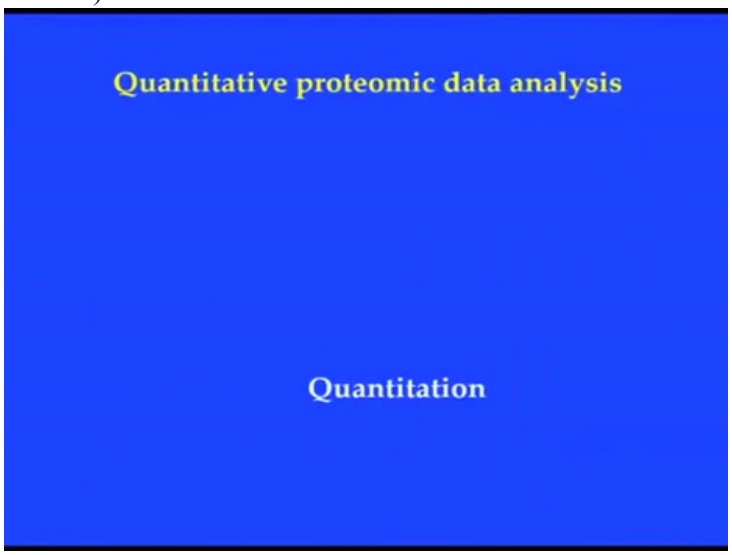
(Refer Slide Time 25:41)



(Refer Slide Time 25:42)

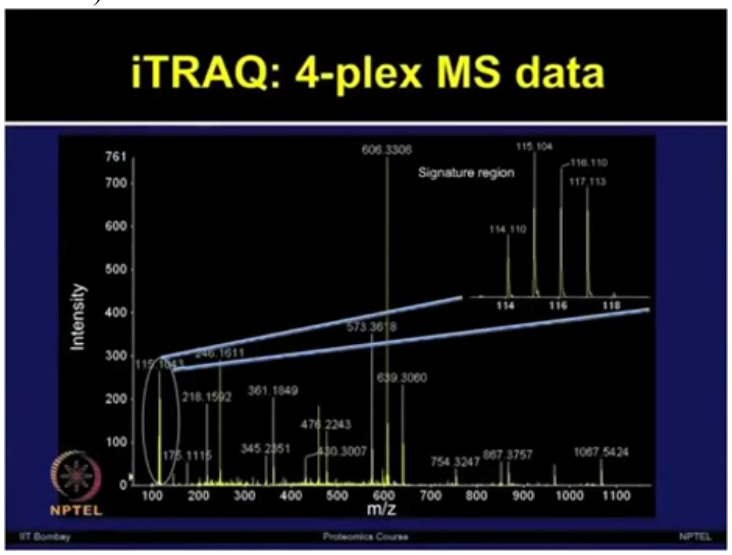


(Refer Slide Time 25:48)



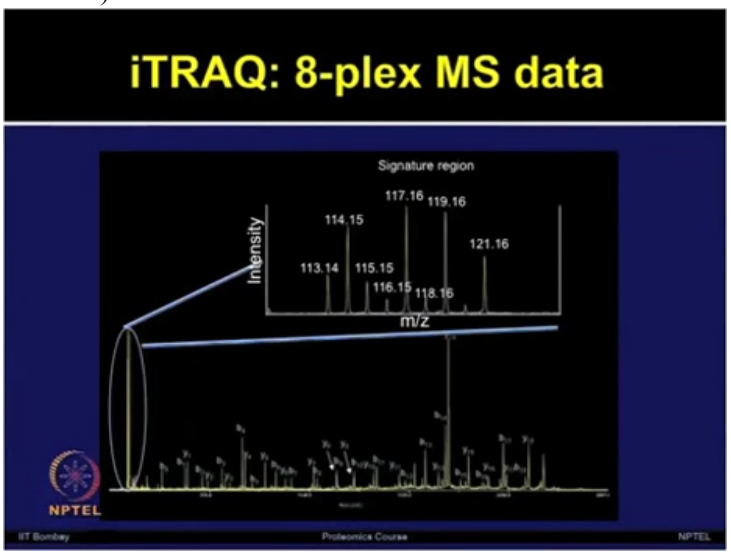
(Refer Slide Time 25:53)

(Refer Slide Time 25:59)



This is the representative spectra for the 4-plex iTRAQ experiment. The MS data is shown and MS/MS spectrum is showing the reporter region, the signature of these 4-flex iTRAQ labeled peptides; 114, 115, 116 and 17

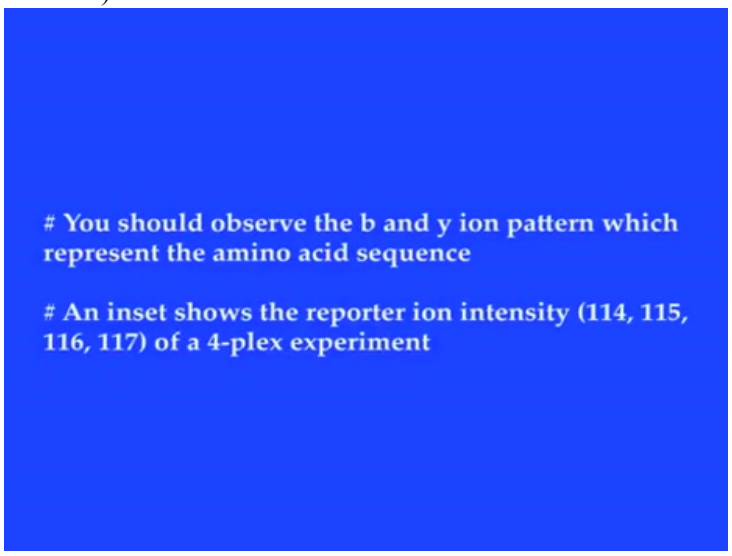
(Refer Slide Time 26:22)



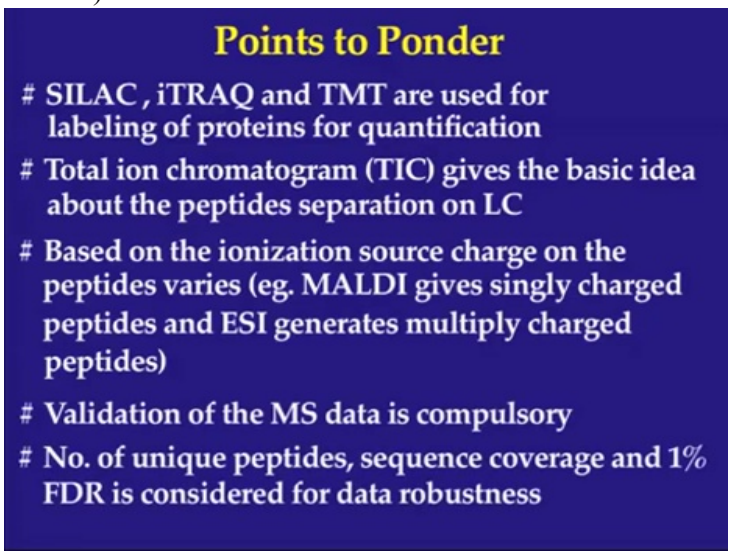
Now we will have a look on the 8-Plex MS data

So in MS/MS spectrum, now here we are showing the reporter region signature of 8-plex iTRAQ region showing 113.14, 114.15, 115.15, 116.15, 117.16, 118.16, 119.16 and 121.16 reporter ions.

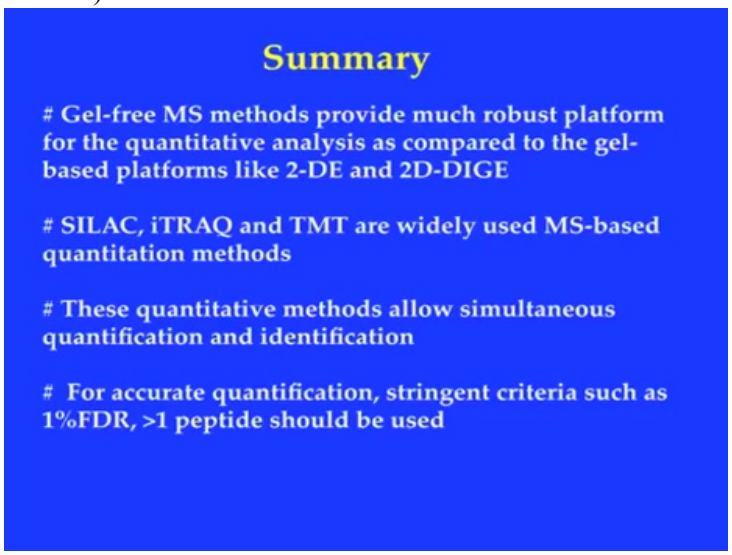
(Refer Slide Time 26:58)



(Refer Slide Time 27:18)



(Refer Slide Time 27:24)



(Refer Slide Time 27:34)