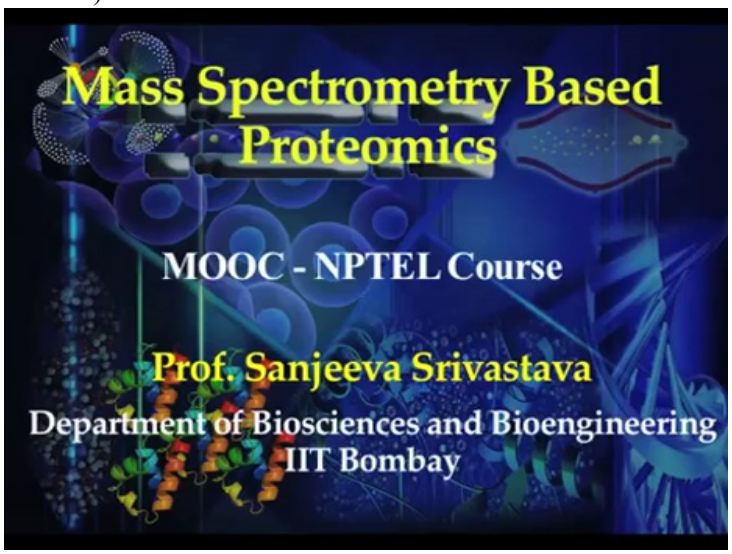
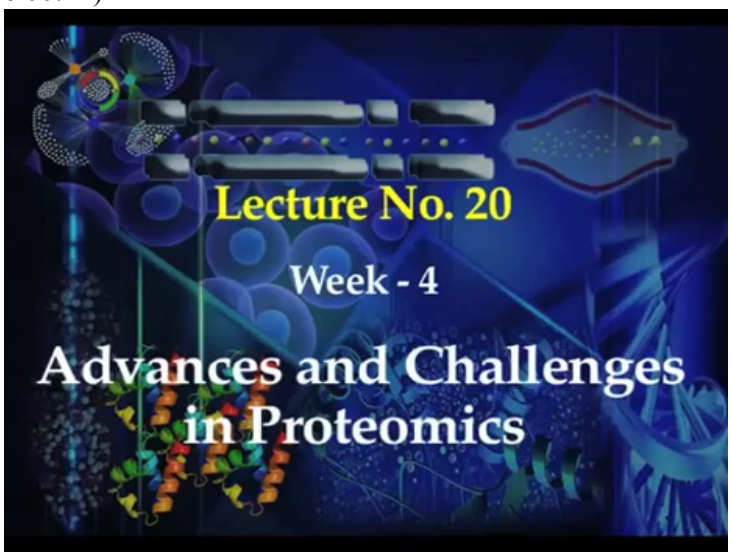
Mass Spectrometry Based Proteomics
Professor Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology, Bombay
Mod 04 Lecture Number 20

(Refer Slide Time 00:10)



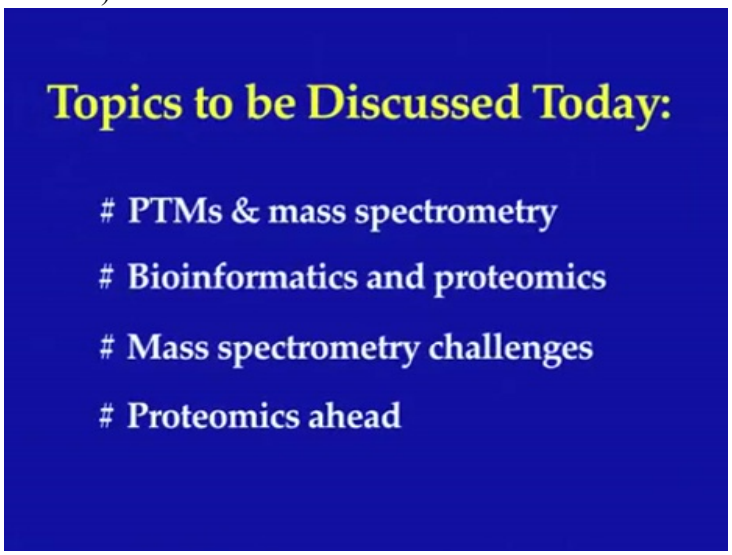
(Refer Slide Time 00:14)



(Refer Slide Time 00:16)

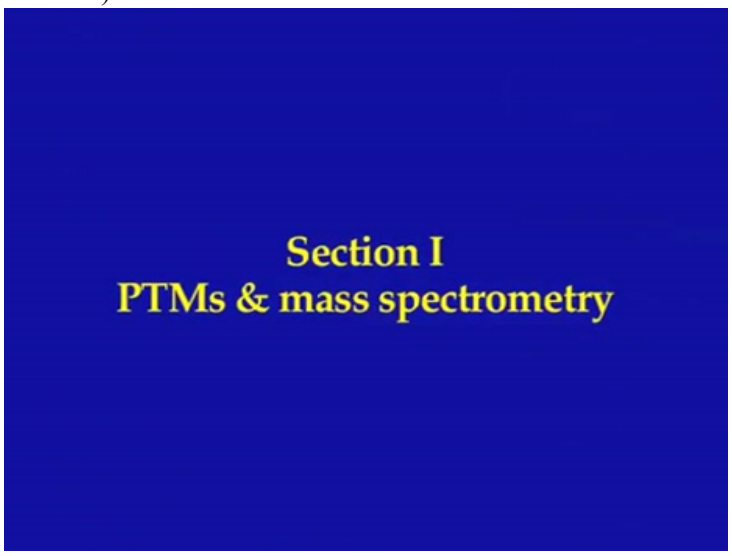
In today's lecture, we will talk about

(Refer Slide Time 00:20)

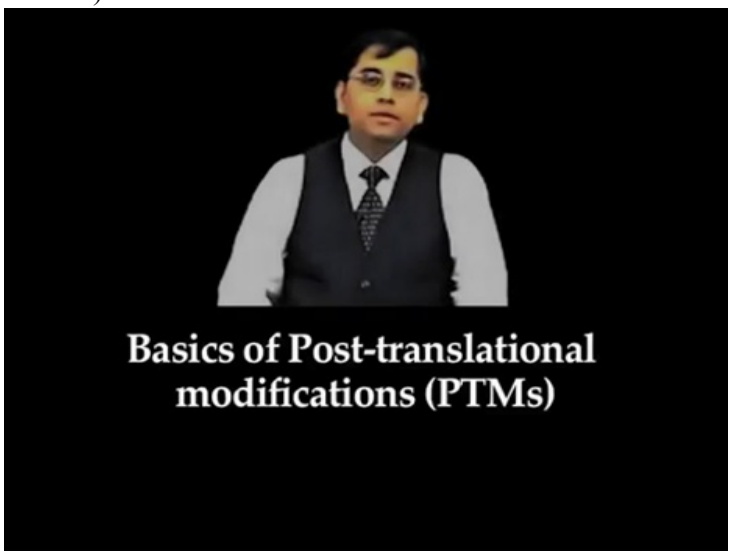


post-translational modifications, structured proteomics, role of bio-informatics, challenges and future directions

(Refer Slide Time 00:33)



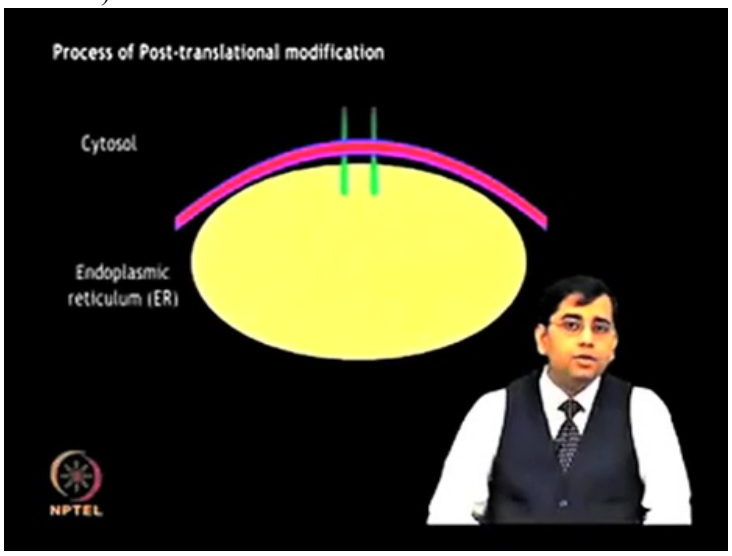
(Refer Slide Time 00:38)



So lets start with PTMs. Post-translational modifications are vital cellular control mechanisms, known as cellular switches that affect protein properties such as protein folding, conformation, activity and functions.

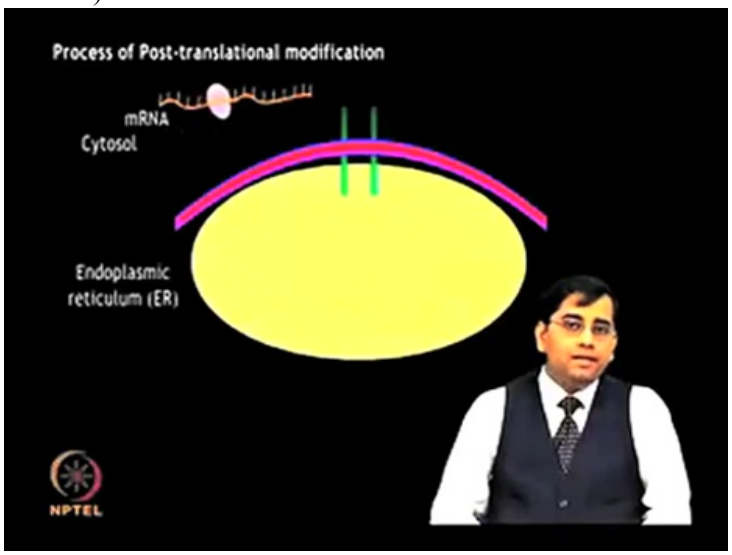
As a result, they play very important role in various diseases.

(Refer Slide Time 01:09)



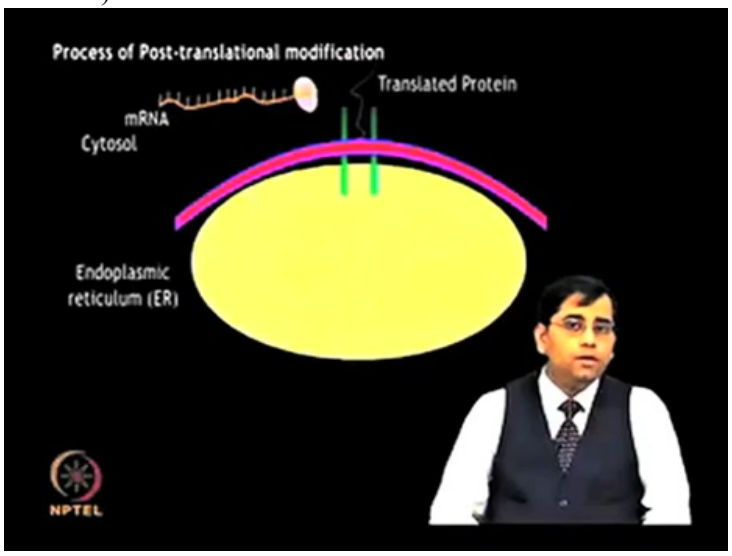
Protein complexity arises due to gene splicing and post-translational modifications.

(Refer Slide Time 01:16)



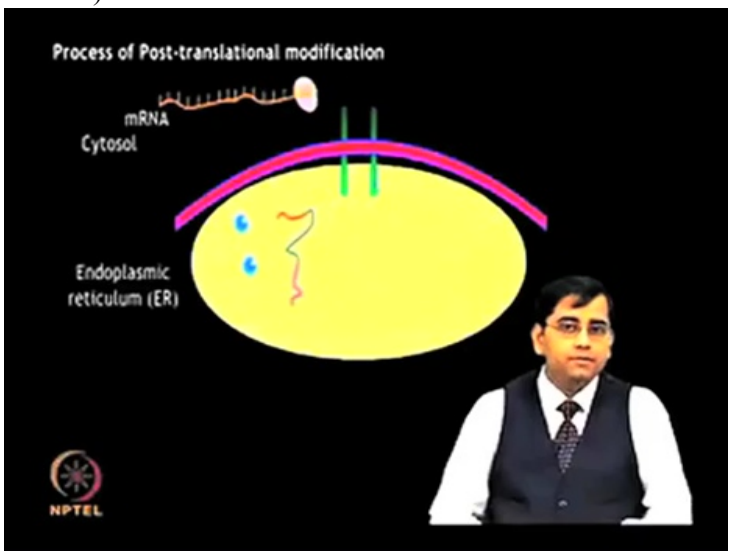
Once the protein is synthesized by the ribosome from its corresponding mRNA in the cytosol,

(Refer Slide Time 01:23)



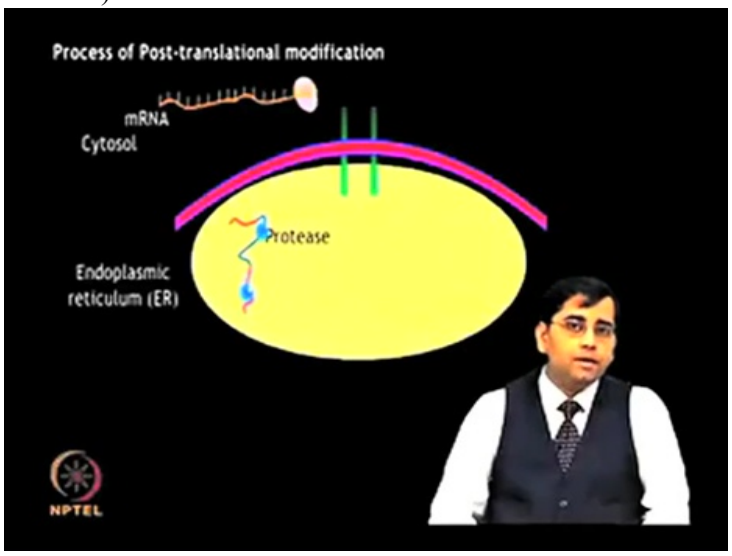
many proteins get directed towards the endoplasmic reticulum for further modification.

(Refer Slide Time 01:32)



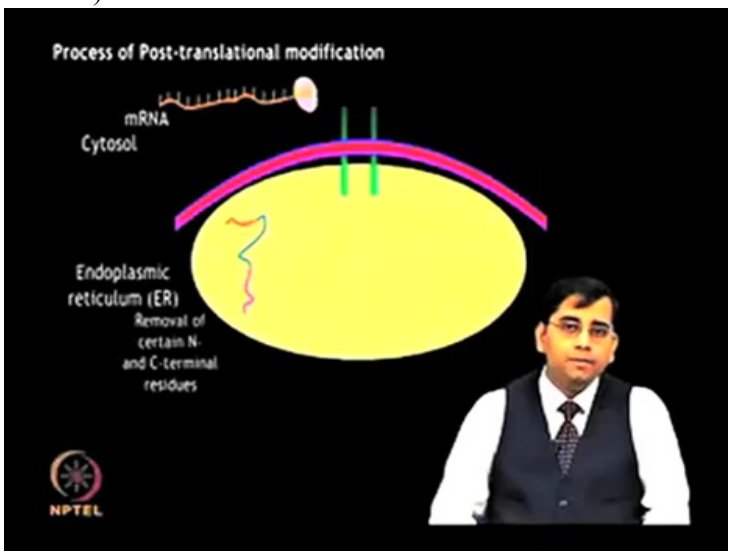
Certain N and C terminal sequences are

(Refer Slide Time 01:35)



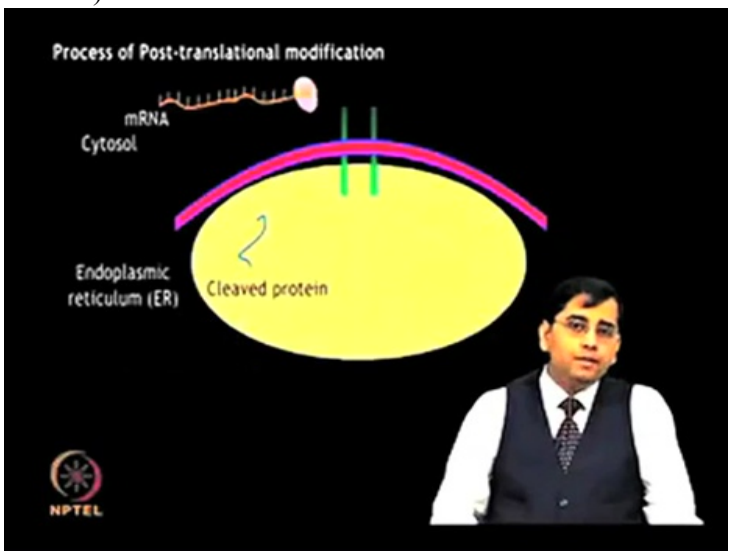
often cleaved in the endoplasmic reticulum after which they are modified by various enzymes

(Refer Slide Time 01:44)



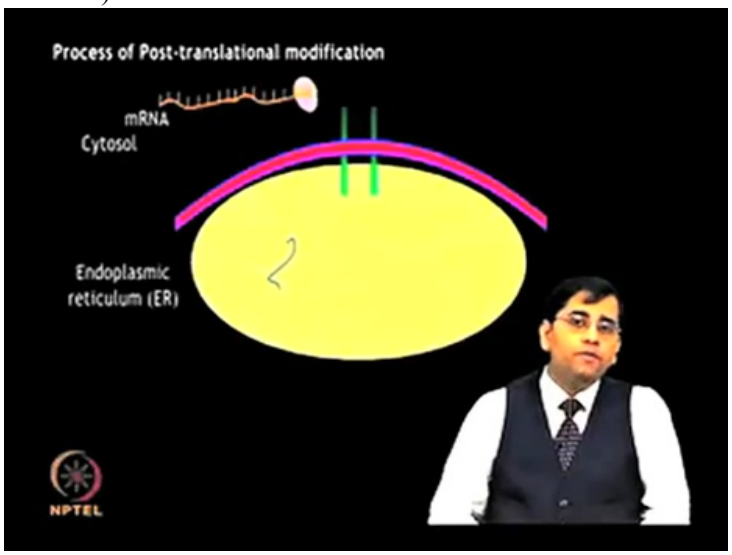
at specific amino acid residues. These modified proteins

(Refer Slide Time 01:50)



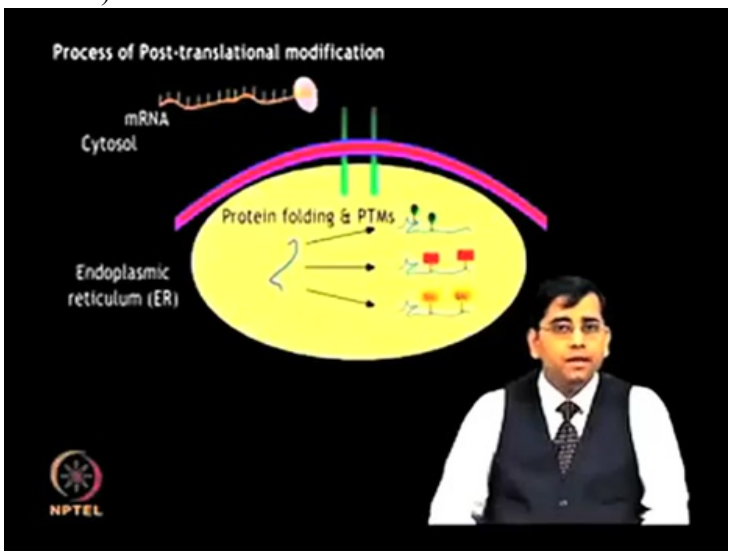
then undergo proper folding to give the functional proteins.

(Refer Slide Time 01:57)



Due to these modifications, the number of proteins are 3 orders of magnitude higher than the total number of genes

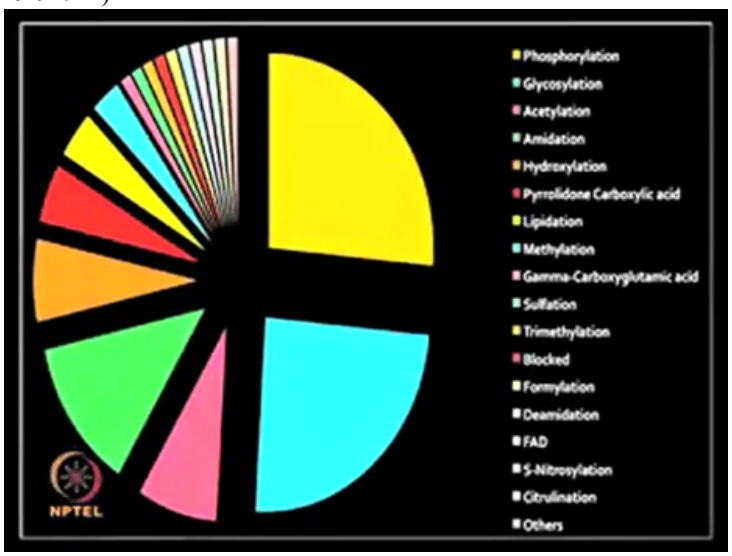
(Refer Slide Time 02:06)



encoded in genome.

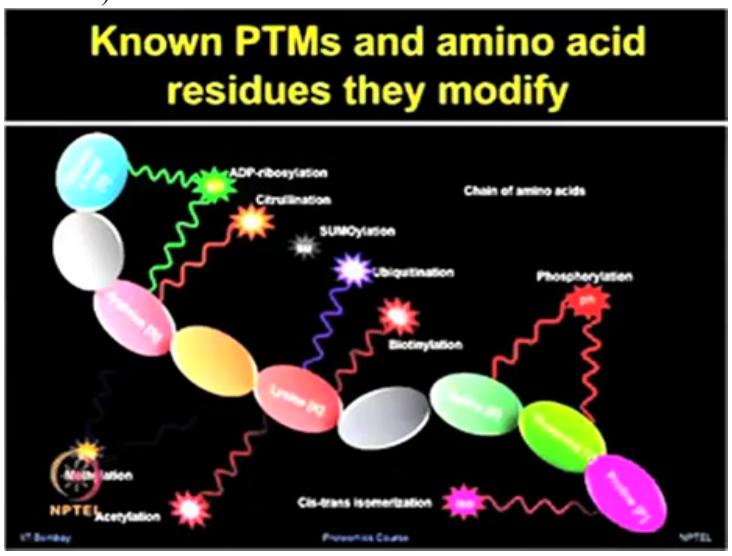
There are several types of post-translational modifications can take place at different amino acid residues.

(Refer Slide Time 02:22)



The most commonly observed post-translational modifications include phosphorylation, glycosylation, methylation as well as hydroxylation and acylation.

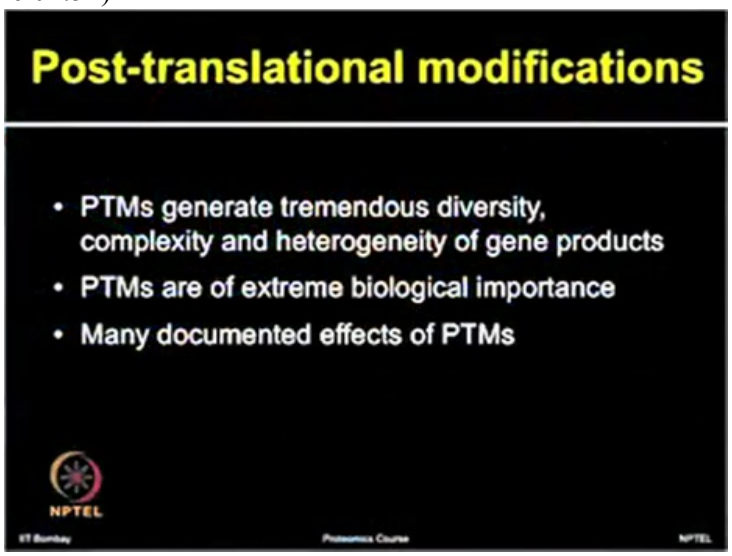
(Refer Slide Time 02:35)



Many of these modifications, particularly phosphorylation, serve as regular mechanisms for protein action.

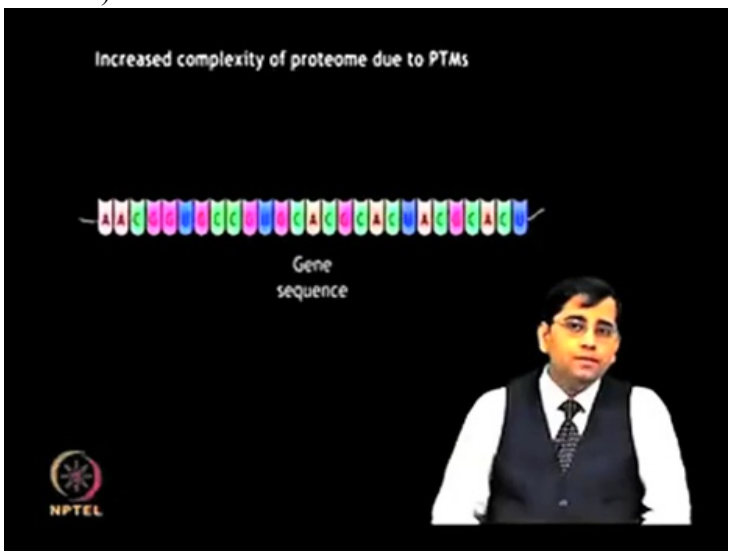
PTMs generate tremendous diversity and are extremely important.

(Refer Slide Time 02:54)



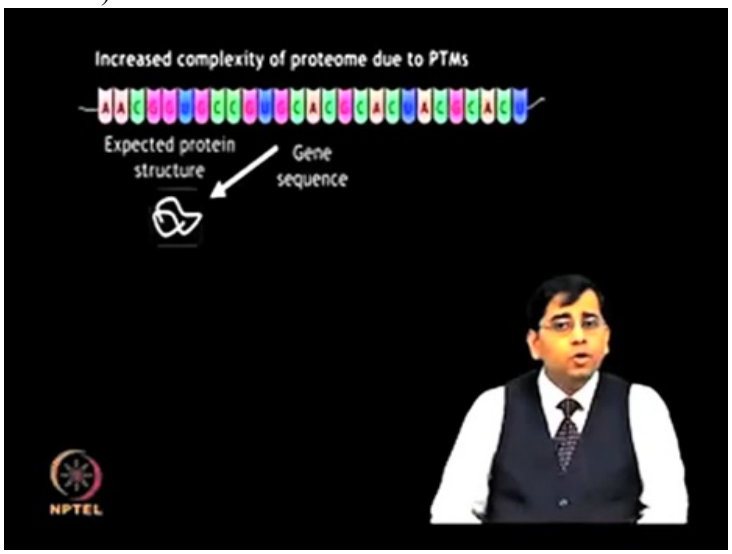
Many documented effects of post-translational modifications include change in enzymatic activity, ability to interact with other proteins, sub-cellular localization and targeted degradation.

(Refer Slide Time 03:15)



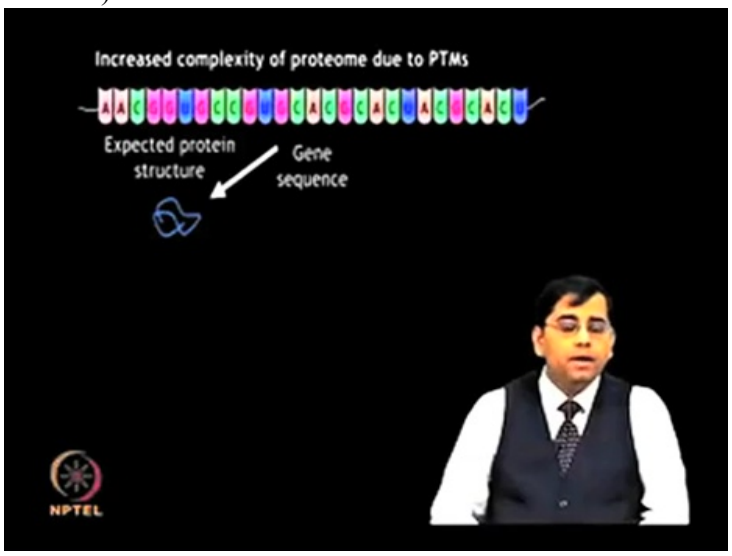
The final structure of functional proteins most often does not correlate directly with the corresponding gene sequence. This is

(Refer Slide Time 03:25)



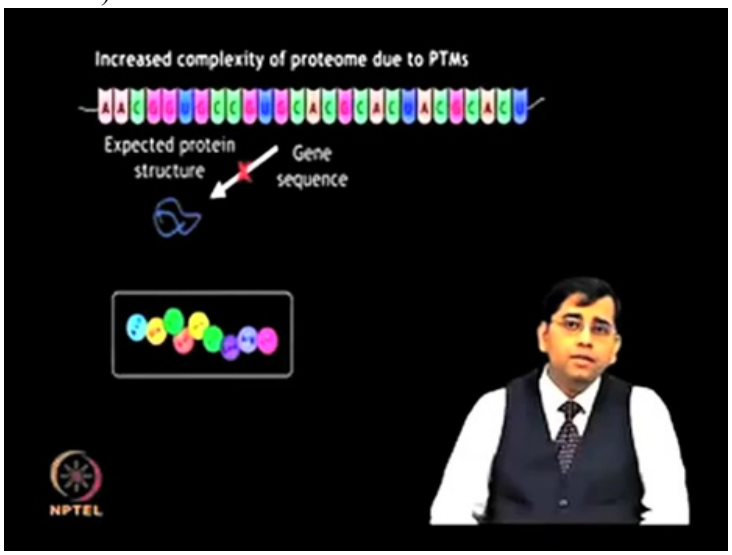
because of PTMs that occur

(Refer Slide Time 03:28)



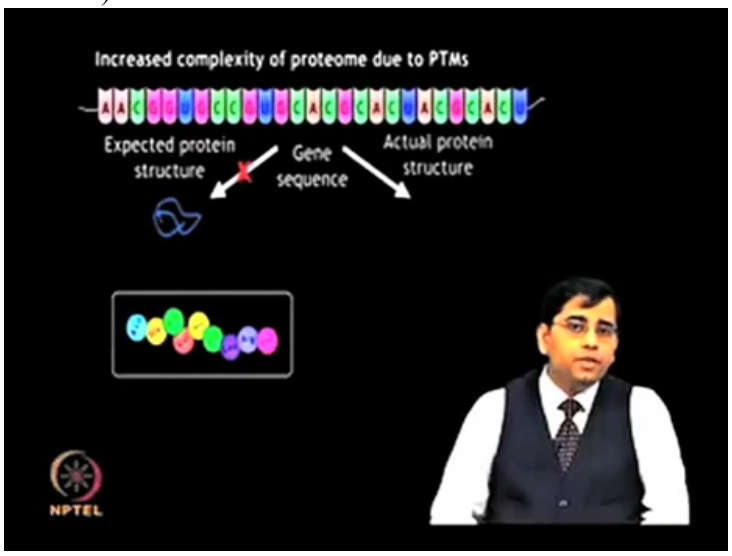
at various amino acid residues in the protein,

(Refer Slide Time 03:32)



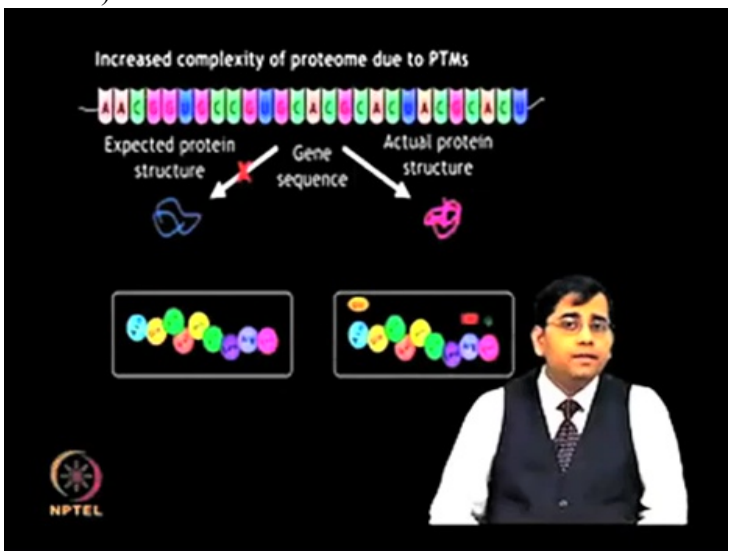
which cause changes in interactions

(Refer Slide Time 03:37)



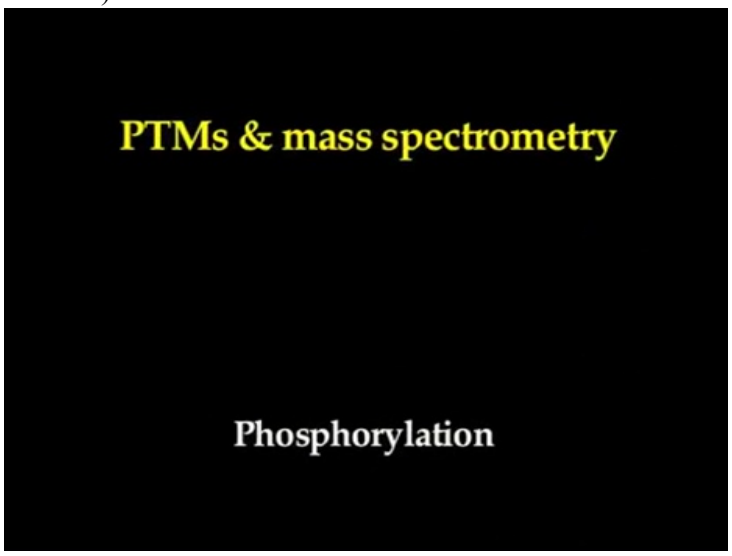
between the amino acid side chains thereby modifying the protein structure.

(Refer Slide Time 03:47)



It further increases the complexity of the proteome as compared to the genome.

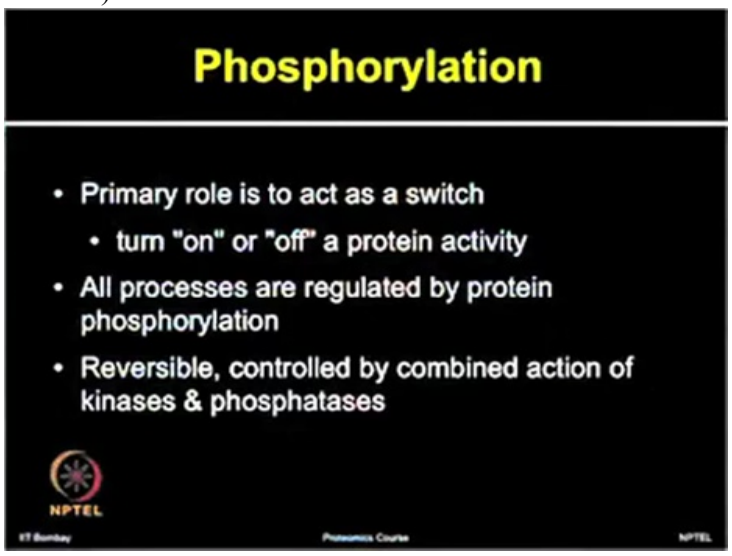
(Refer Slide Time 03:54)



(Refer Slide Time 03:59)

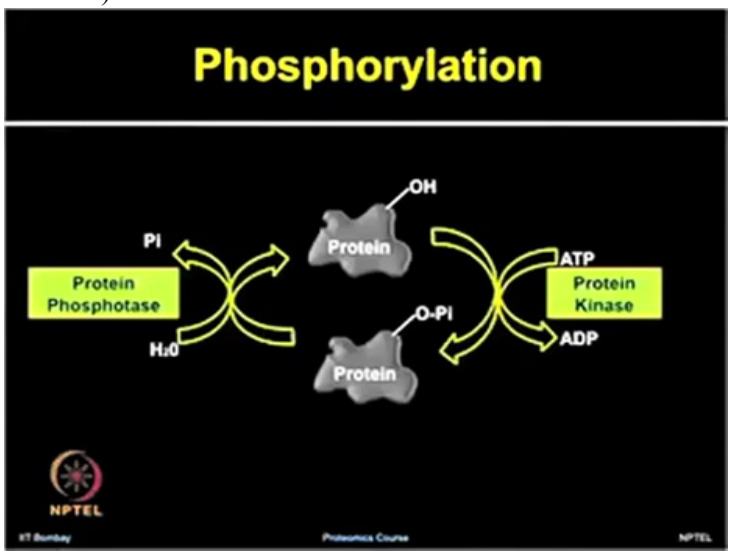
The protein phosphorylation acts as a switch to turn on or turn off the protein activity and governs

(Refer Slide Time 04:08)



wide range of polypeptides from transcription factors, enzymes to cell surface receptors.

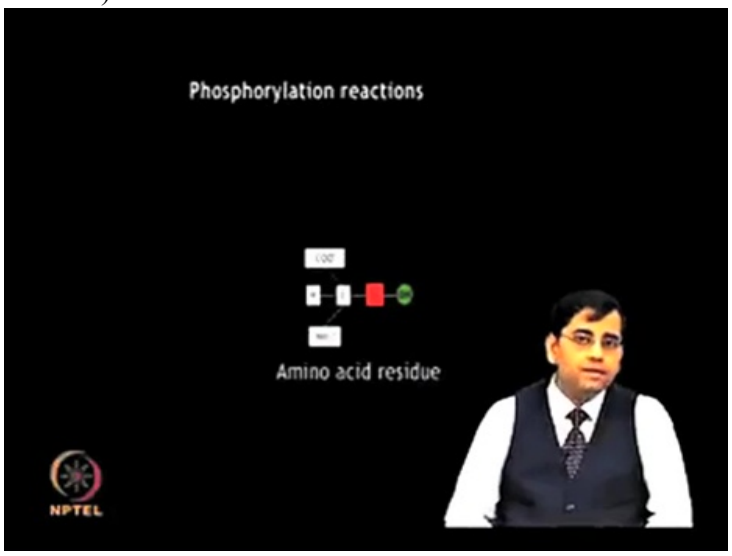
(Refer Slide Time 04:18)



The reversible phosphorylation of proteins catalyzed by kinases and phosphatases regulates important cellular functions.

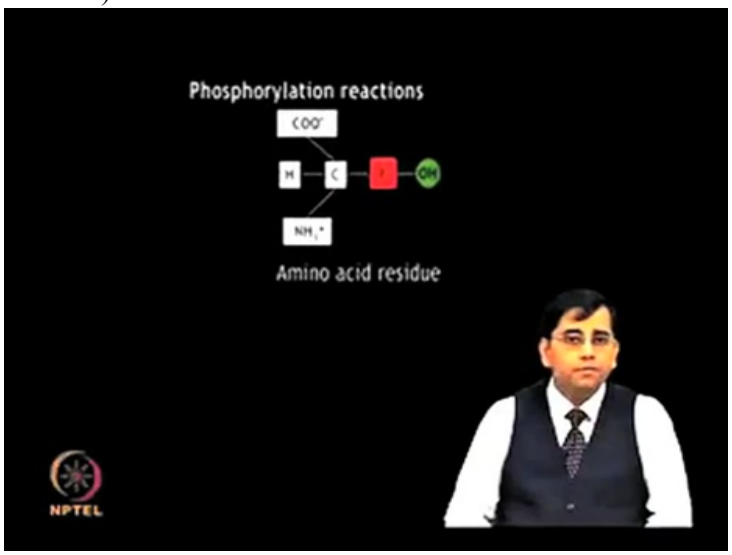
Phosphorylation of amino acid residues is carried out by a class of enzymes known as kinases

(Refer Slide Time 04:38)



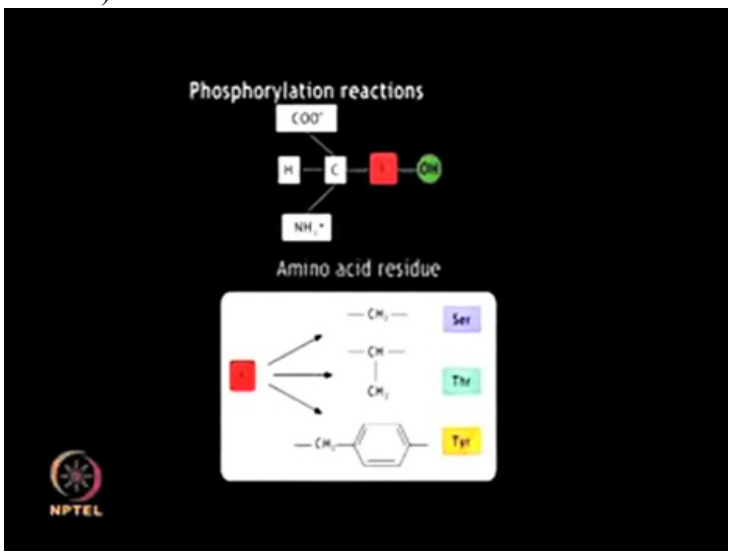
that most commonly modify side chains of

(Refer Slide Time 04:43)



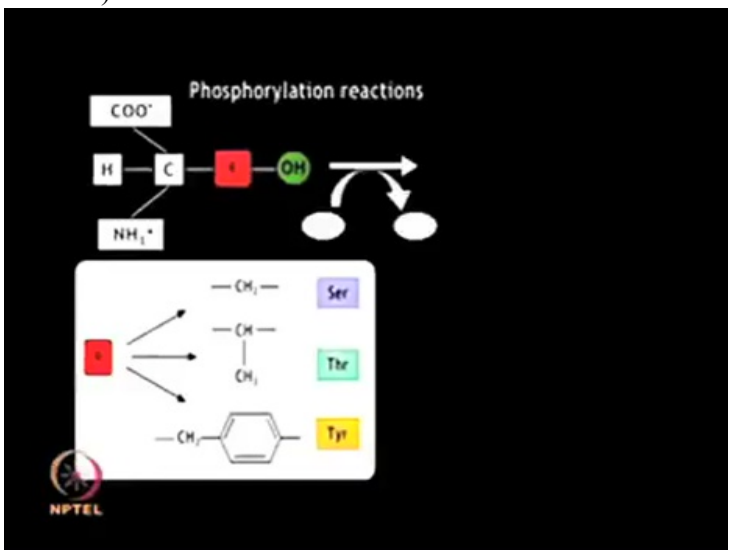
amino acids containing a hydroxyl group.

(Refer Slide Time 04:48)



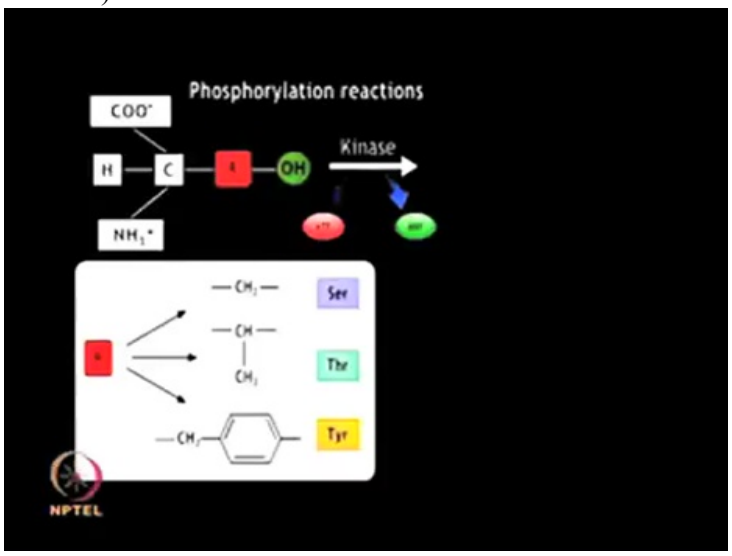
Phosphorylation requires the presence of a phosphate donor molecule such as ATP, GTP or other phoshorylated substrates.

(Refer Slide Time 05:00)



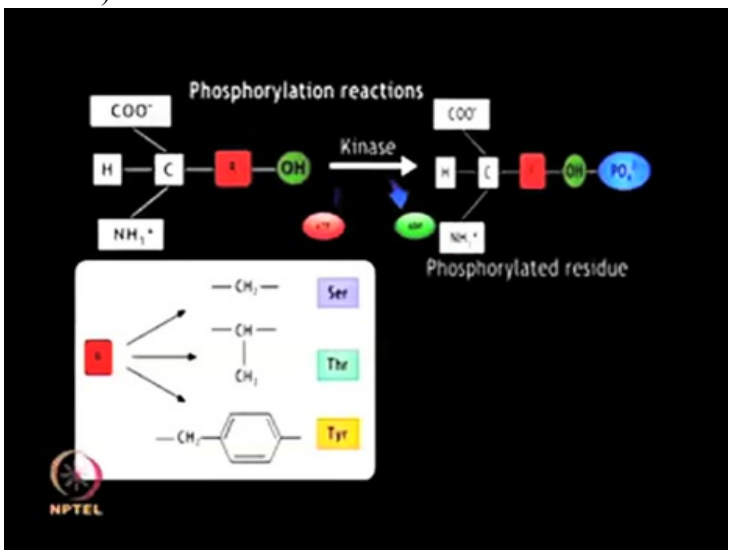
Serine is the most commonly

(Refer Slide Time 05:04)



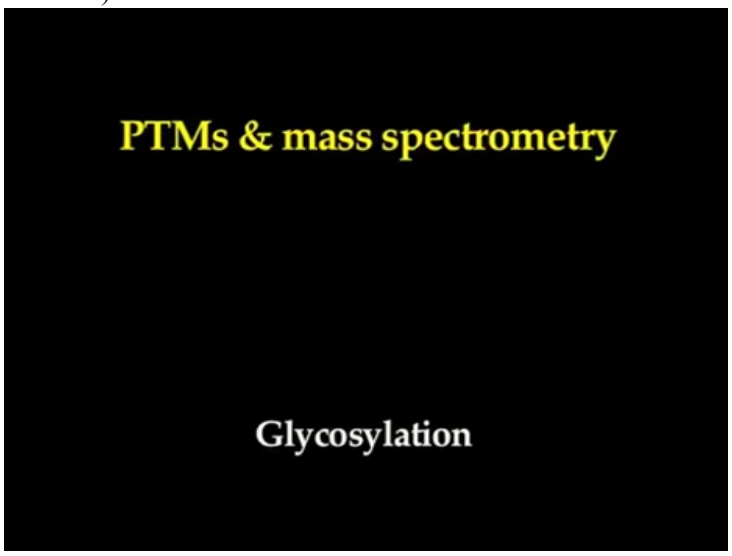
phosphorylated residue followed by threonine and tyrosine. The removal of phosphate groups is carried out by the phosphatase enzyme

(Refer Slide Time 05:15)



and it forms one of the most important mechanisms for protein regulation.

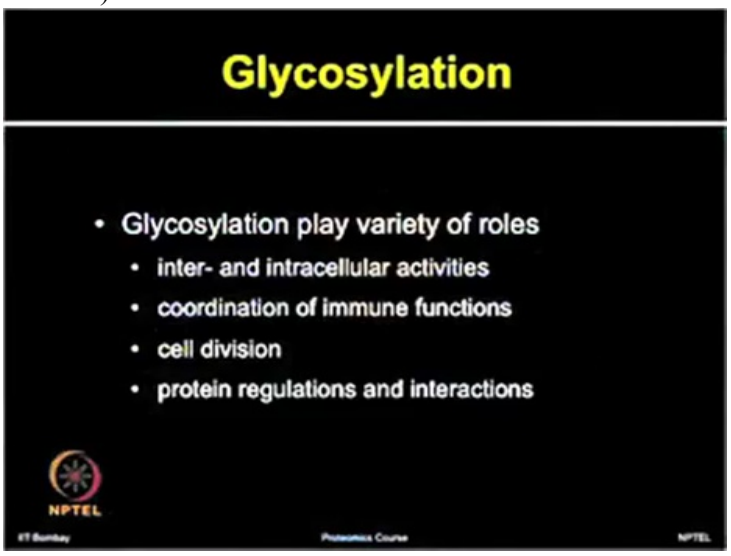
(Refer Slide Time 05:25)



(Refer Slide Time 05:30)

Glycosylation involves linking saccharides to proteins in presence of glycosyl transferases enzymes, giving rise to a glycoprotein.

(Refer Slide Time 05:43)



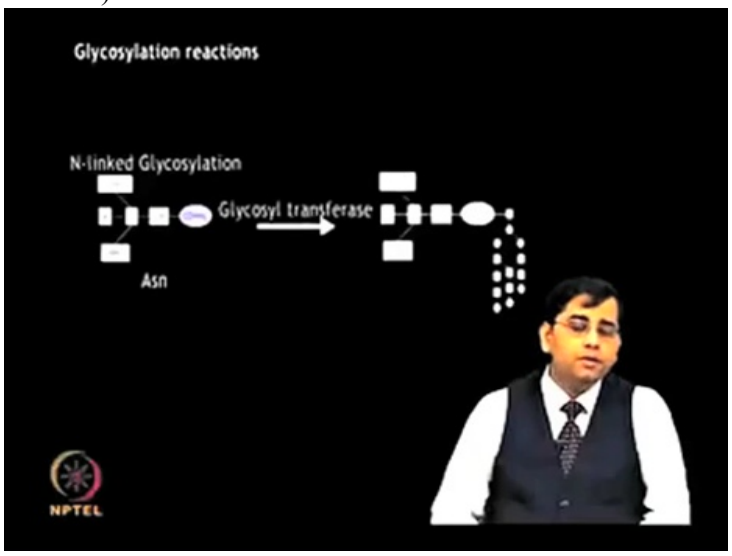
Glycosylation play vital role in various biological functions such as antigenicity of immunological molecules, cell division, protein targeting stability and interactions.

The aberrant glycosylation forms result into various human congenital disorders.

(Refer Slide Time 06:13)

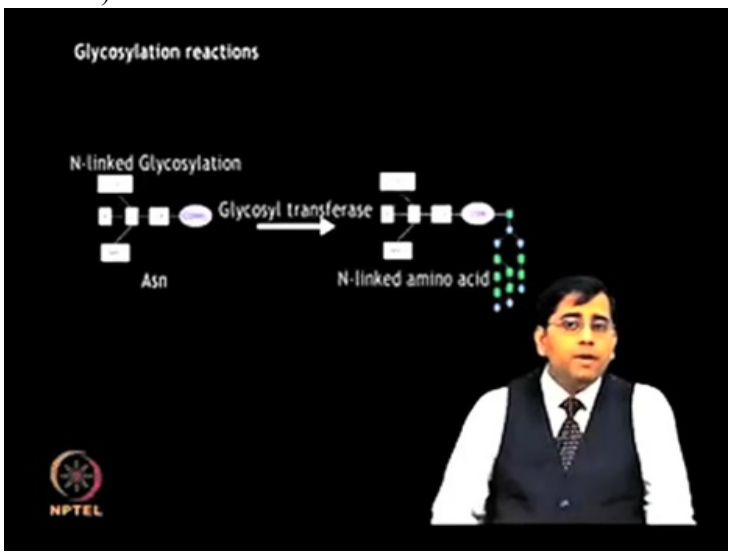
Depending on the linkage between the amino acid and the sugar moiety, there are 4 types of glycosylations;

(Refer Slide Time 06:23)



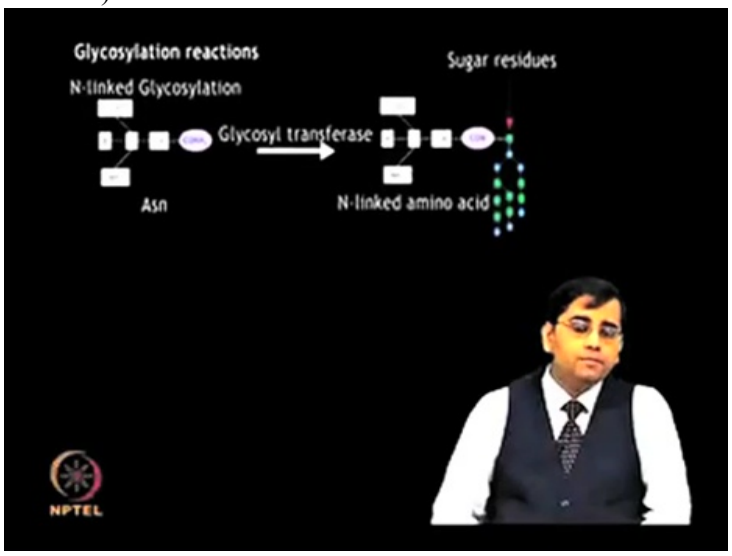
N-linked glycosylation, O-linked glycosylation,

(Refer Slide Time 06:27)

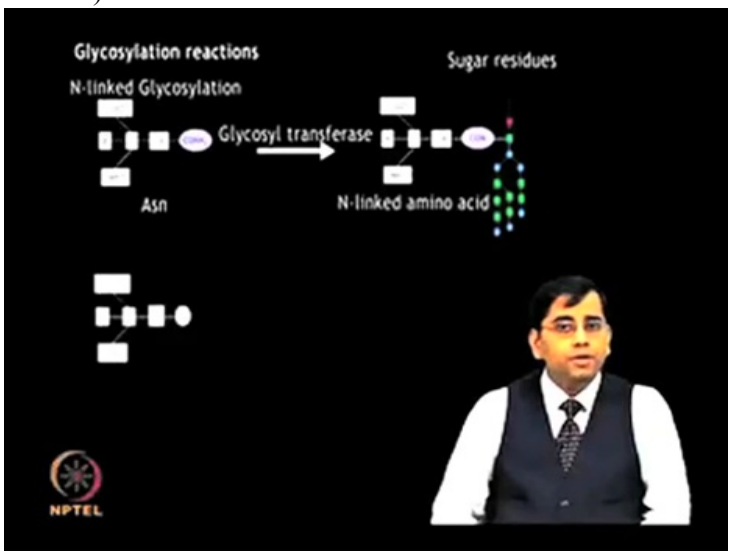


C-ammannosylation and GlycoPhosphatidylInositol anchored (GPI) attachments.

(Refer Slide Time 06:35)

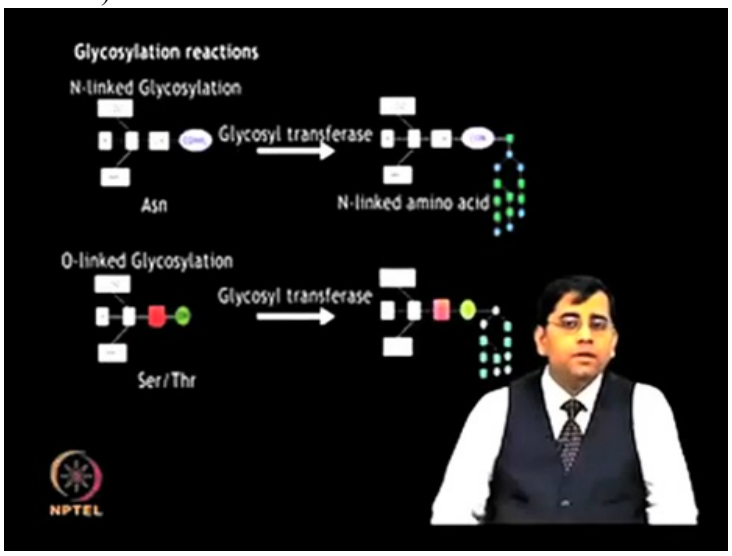


(Refer Slide Time 06:37)



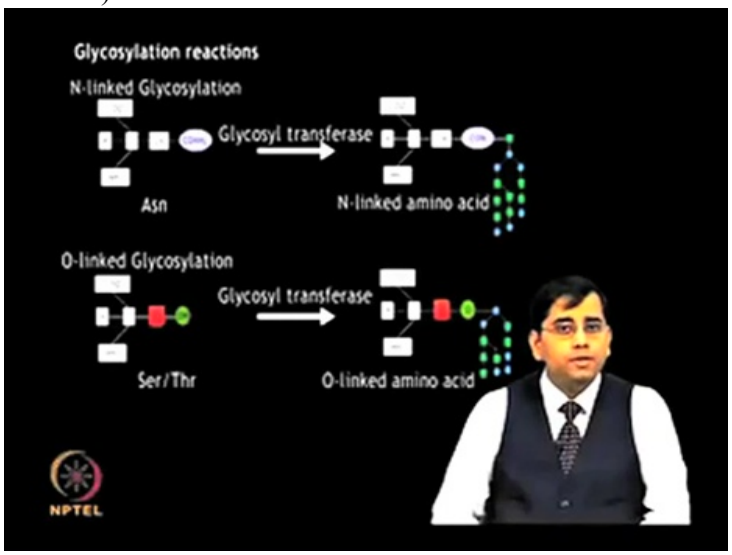
Glycosylation involves the enzymatic addition of saccharide molecules to amino acid side chains.

(Refer Slide Time 06:45)



This can be of two types – N-linked glycosylation,

(Refer Slide Time 06:50)

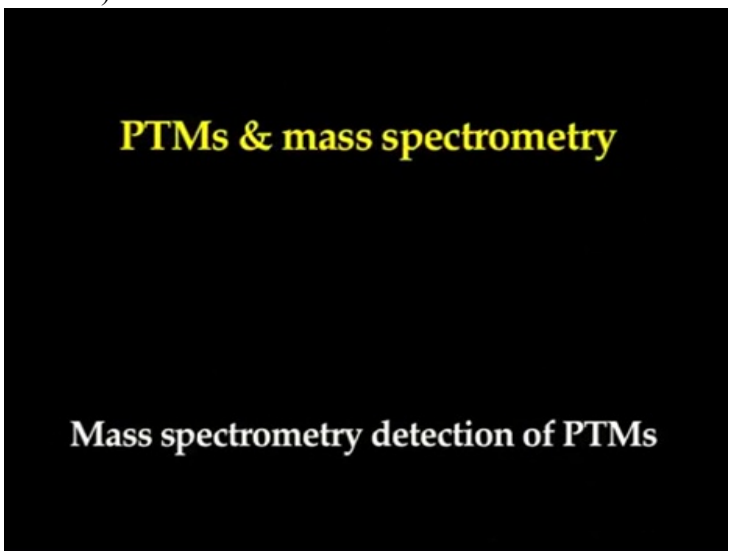


which links sugar residues to the amide group of aspargine and O-linked glycosylation, which links the sugar moieties to the hydroxyl groups of serine or threonine.

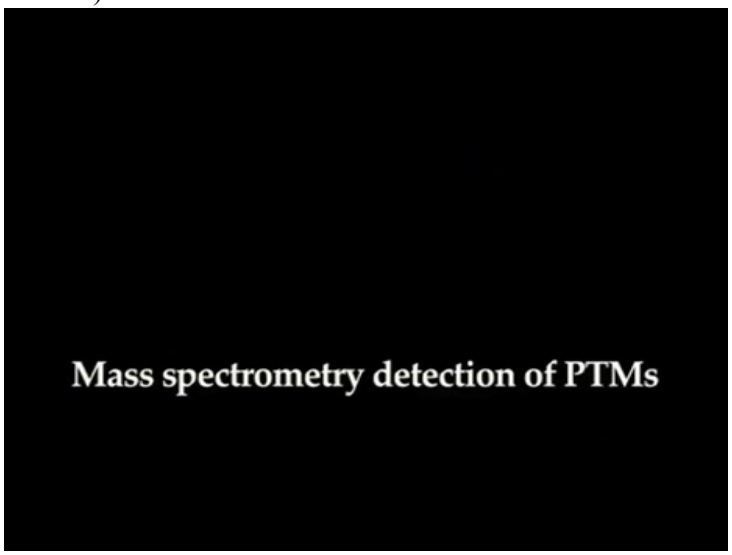
The glycosyl transferase enzymes catalyze these reactions.

Sugar residues that are attached most commonly include galactose, mannose, glucose, N-acetylglucosamine etc.

(Refer Slide Time 07:28)



(Refer Slide Time 07:33)



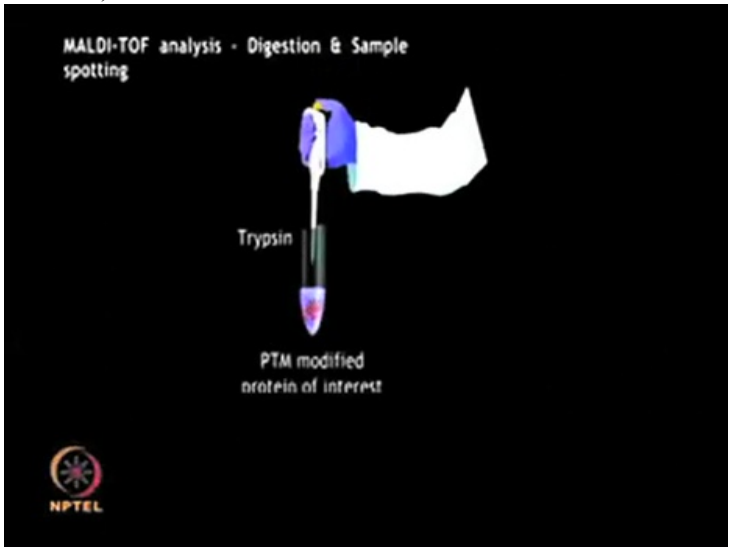
Protein mixture is digested with proteolytic enzyme such as trypsin, and resultant peptides can be analyzed by MALDI-TOF or LC-MS/MS.

The top down mass spectrometry involves analysis of intact proteins using high-resolution Mass Spectrometry techniques.

High-resolution MS platforms such as FTICR-MS, Orbitrap-MS with PTM friendly dissociation techniques such as Electron capture dissociation and Electron Transfer Dissociation ETD are commonly used.

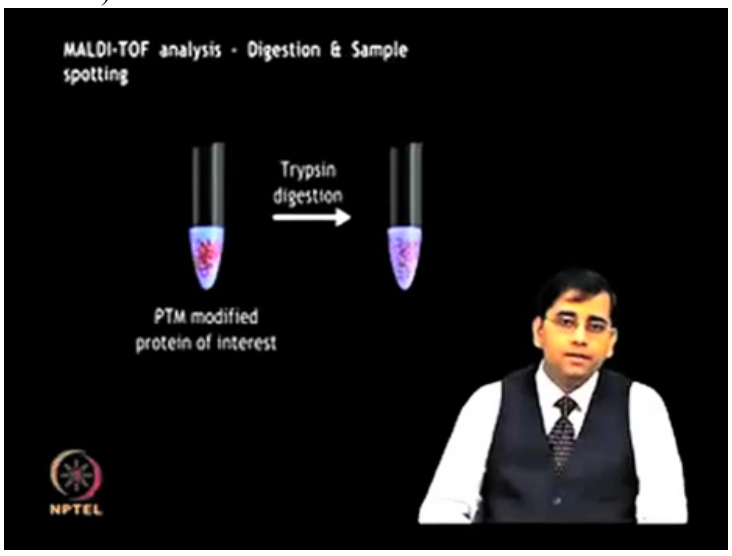
PTMs can be detected by means of mass spectrometry due to the unique fragmentation patterns of phosphorylated serine and threonine residues.

(Refer Slide Time 08:39)



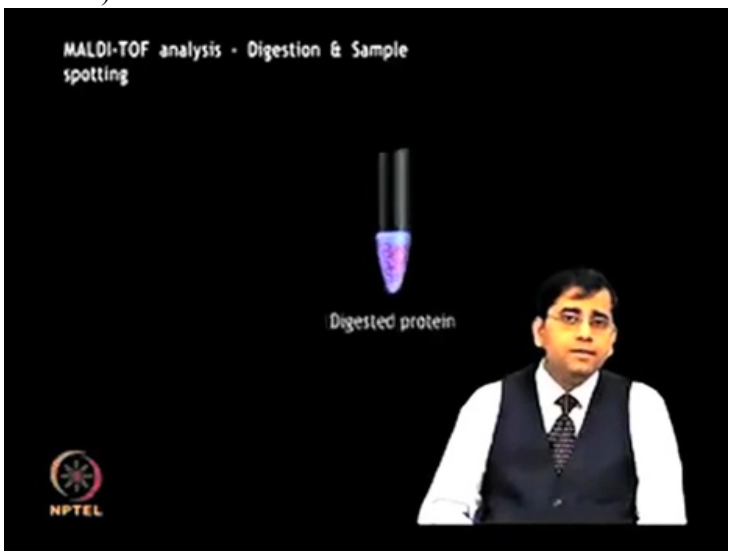
The modified protein of interest is digested into smaller peptide fragments

(Refer Slide Time 08:45)



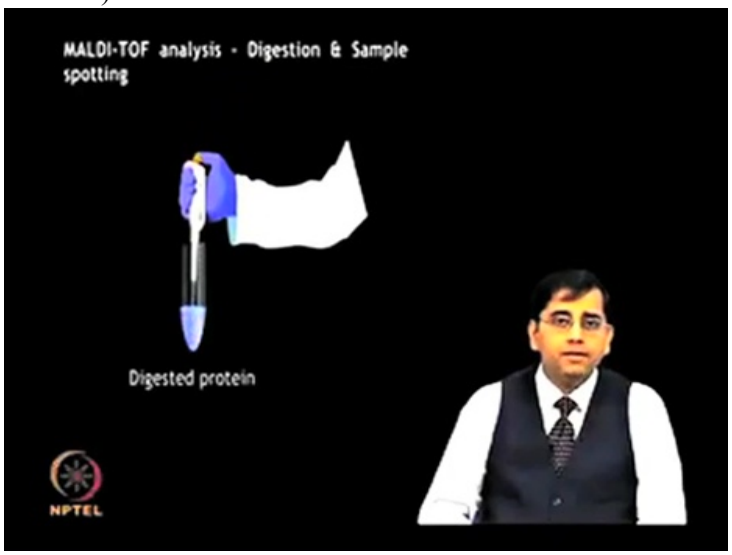
using trypsin. This digest

(Refer Slide Time 08:48)



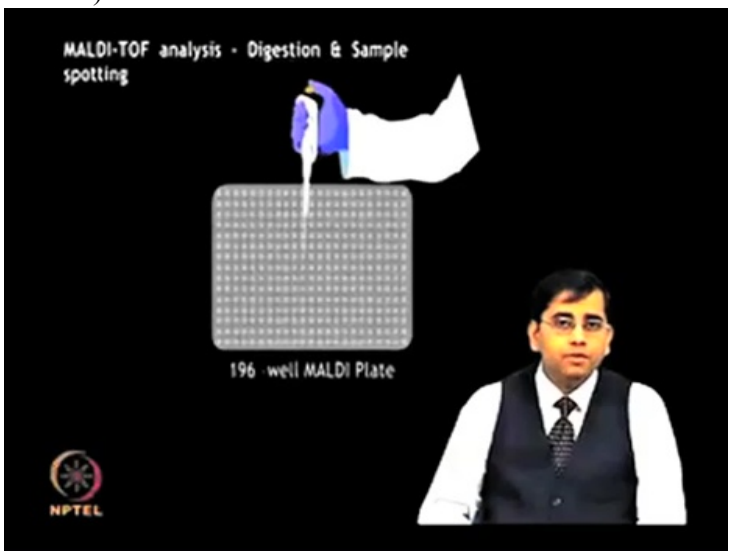
is then mixed with a suitable organic matrix

(Refer Slide Time 08:52)



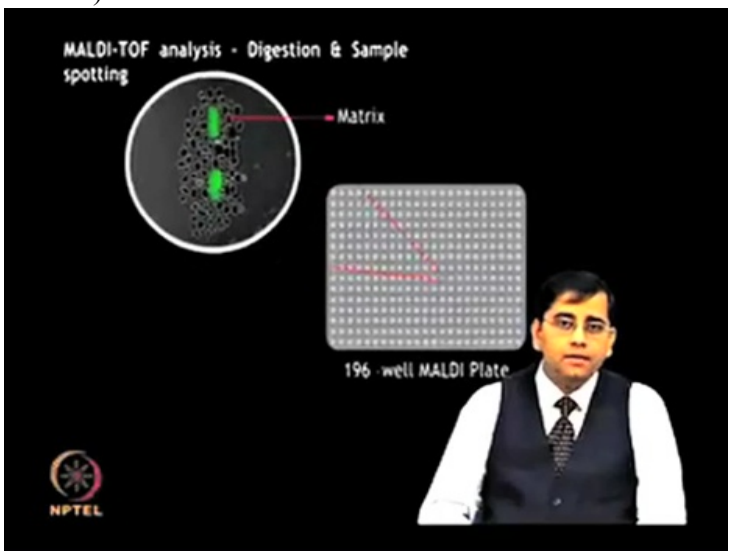
such as

(Refer Slide Time 08:55)



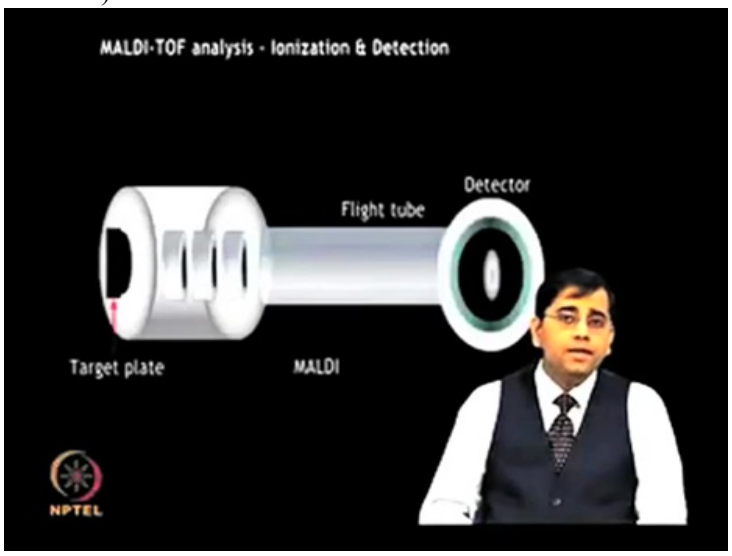
alpha-cyano-4-hydroxycinnamic acid, sinapinic acid etc. and then

(Refer Slide Time 09:01)



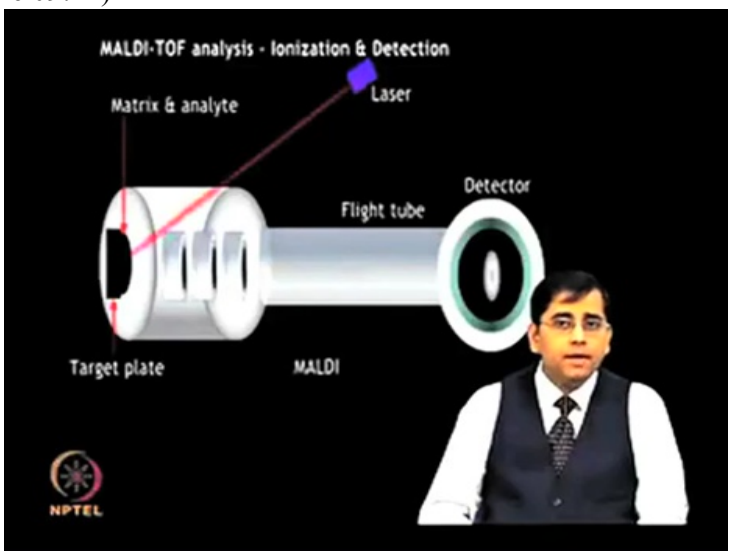
it is spotted on to a MALDI plate.

(Refer Slide Time 09:06)



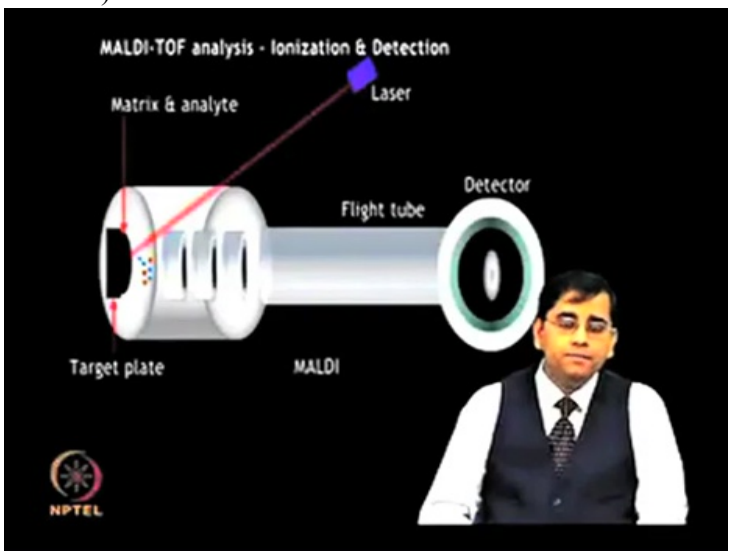
The target plate containing the spotted matrix and analyte is placed in a vacuum chamber with high voltage and short laser pulses are applied. The laser energy gets absorbed by the matrix

(Refer Slide Time 09:24)



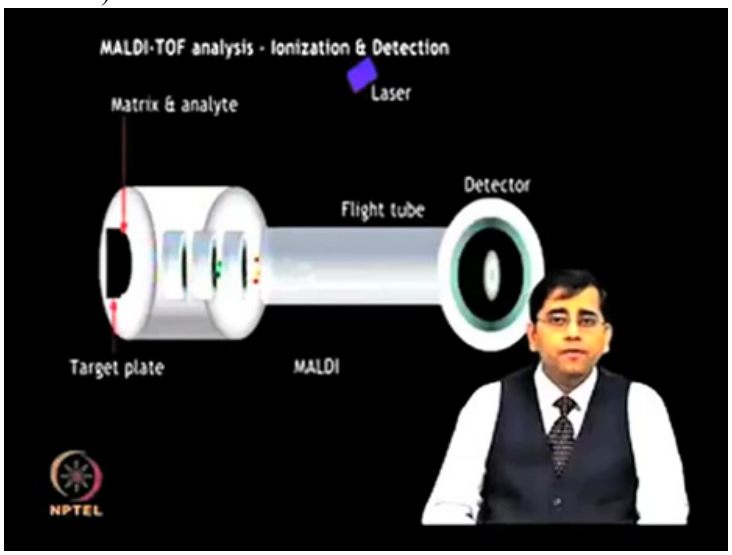
and is transferred to the analyte molecules, which undergo rapid sublimation resulting in

(Refer Slide Time 09:31)



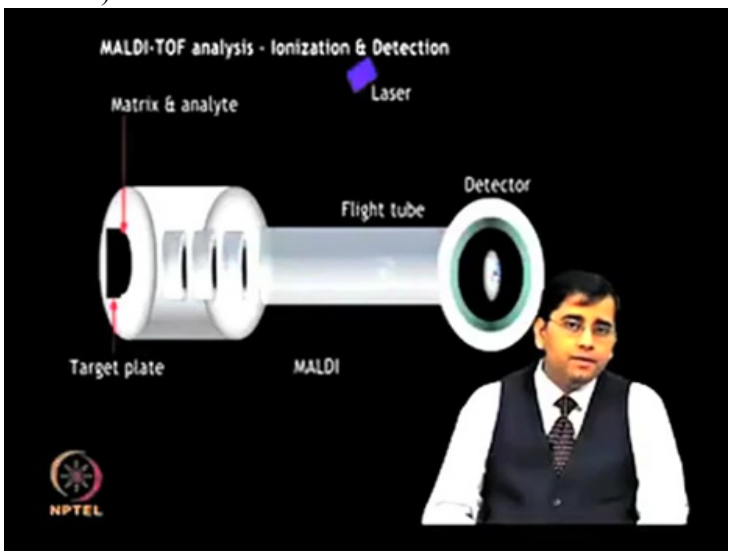
gas phase ions. These ions are accelerated and travel through

(Refer Slide Time 09:38)



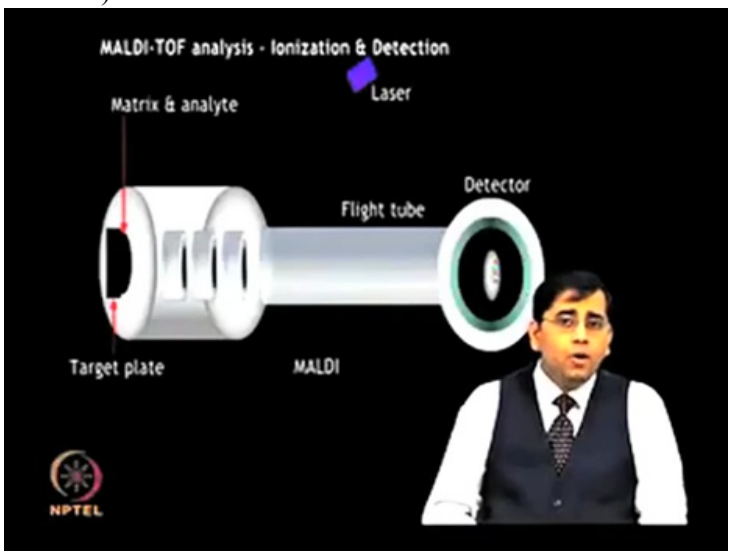
the flight tube at different rates. The lighter ions move rapidly

(Refer Slide Time 09:46)



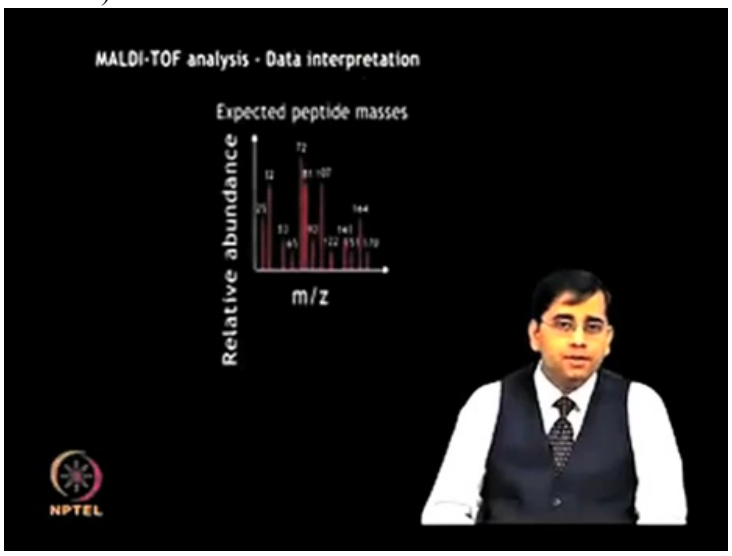
and reach the detector first while the heavier ions

(Refer Slide Time 09:53)



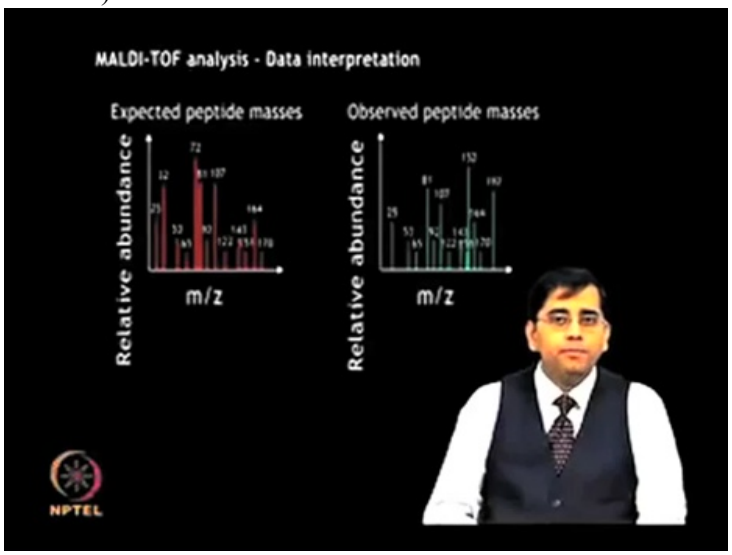
migrate slowly. The ions are resolved and detected on the basis of their m by z ratios and a mass spectrum is

(Refer Slide Time 10:03)



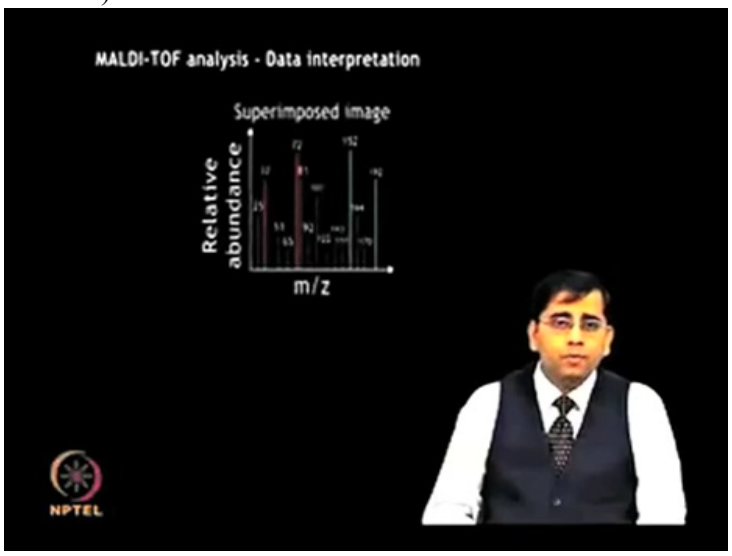
generated. Identification of

(Refer Slide Time 10:06)



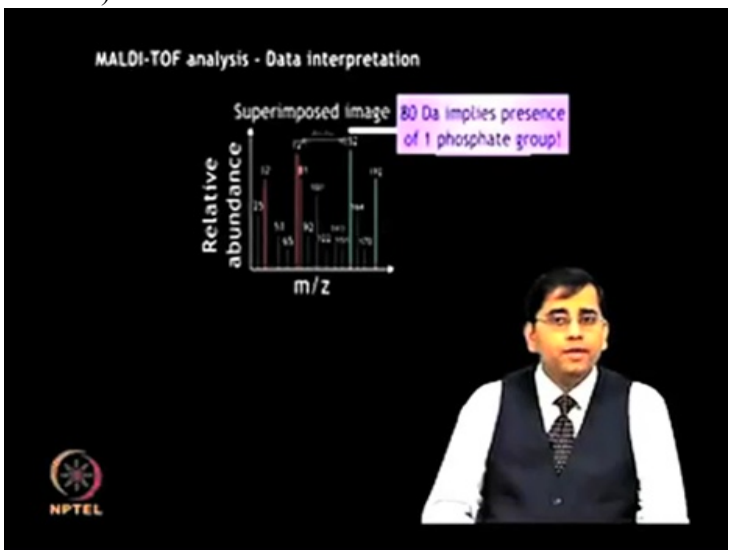
post-translational modifications by MS largely lies in the interpretation of results. Comparison of the list of observed peptide masses

(Refer Slide Time 10:20)



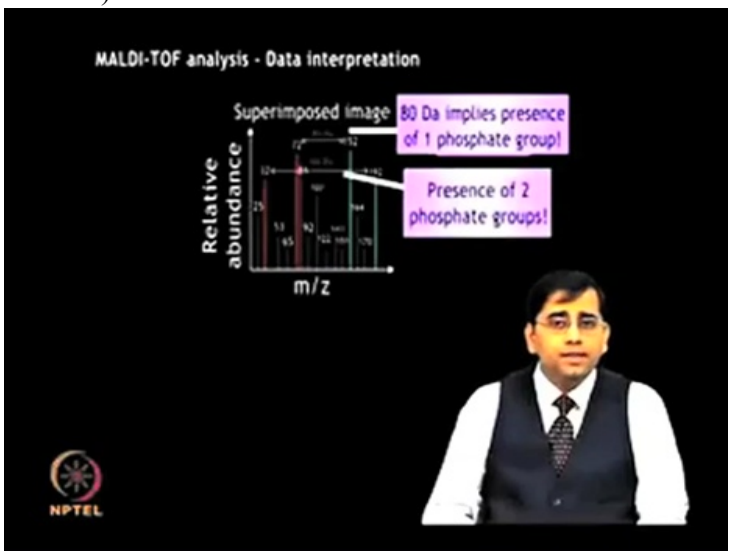
from the spectrum generated with the expected peptide masses enables identification of those peptide fragments that contain

(Refer Slide Time 10:33)



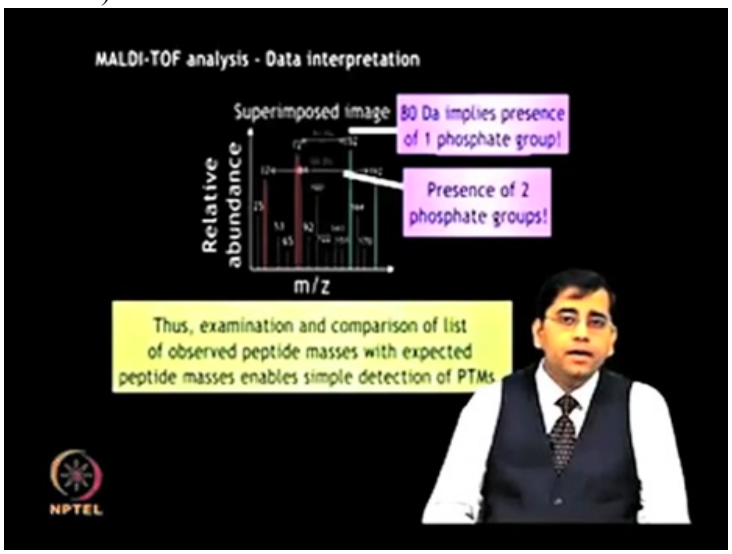
any Post Translational Modifications due to the added mass of a modifying group.

(Refer Slide Time 10:40)



In this hypothetical example, two peptide fragments are found to have different m by z values, differing by

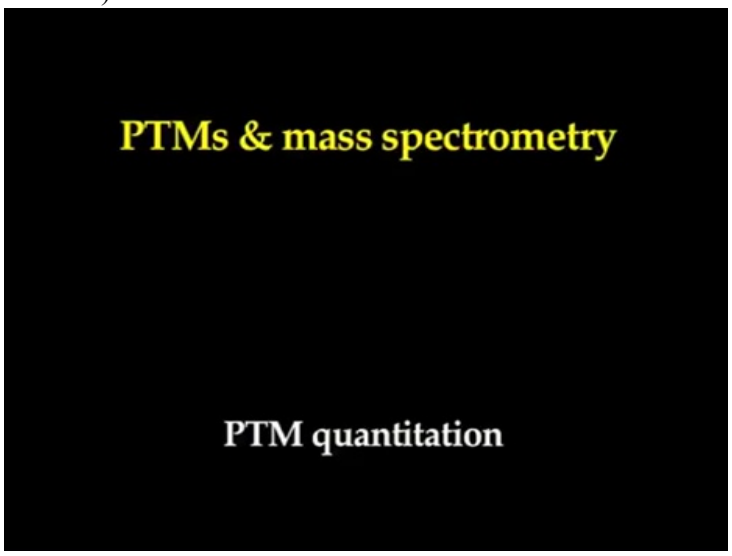
(Refer Slide Time 10:50)



80 Daltons and 160 Daltons.

It is known that the added mass of a phosphate group causes an increase in m/ by of 80 daltons. Therefore, this principle of mass difference enables the detection of modified fragments.

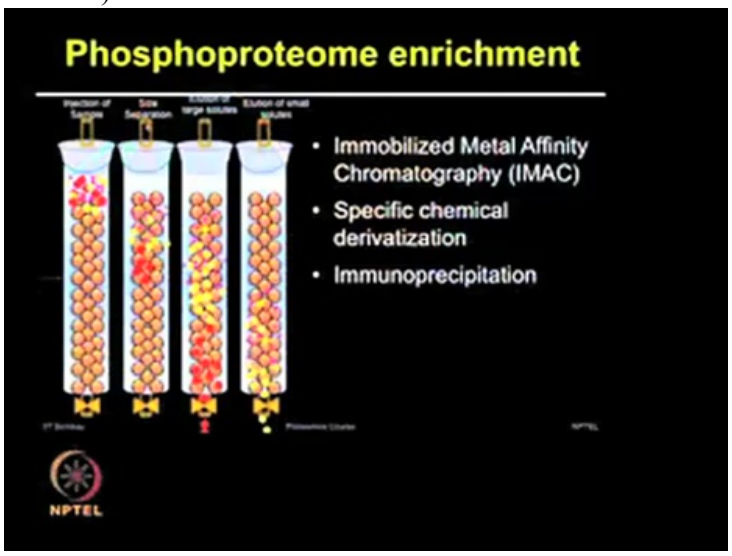
(Refer Slide Time 11:16)



(Refer Slide Time 11:20)

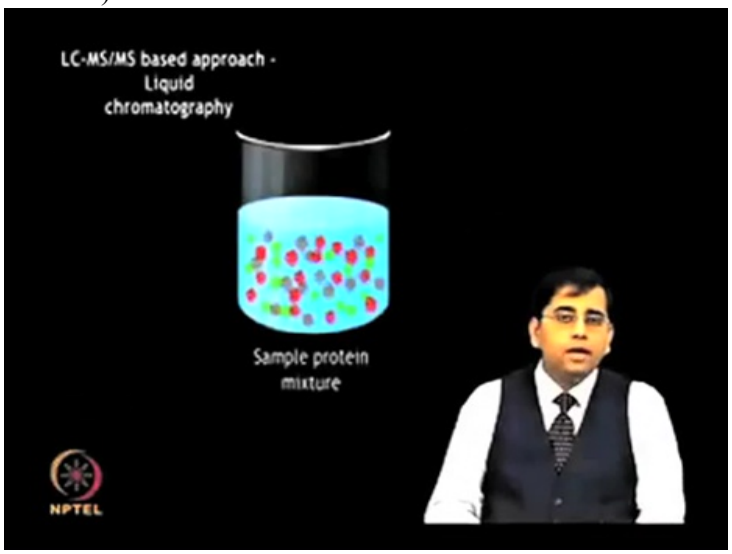
Affinity based enrichments, immuno-purification and metal affinity chromatography are commonly employed for the purification of proteins containing specific PTMs.

(Refer Slide Time 11:38)



Immobilized metal affinity chromatography IMAC and metal oxide affinity resins such as Titanium dioxide, Fe3 O4 are also commonly used for the enrichment of phosphoproteins.

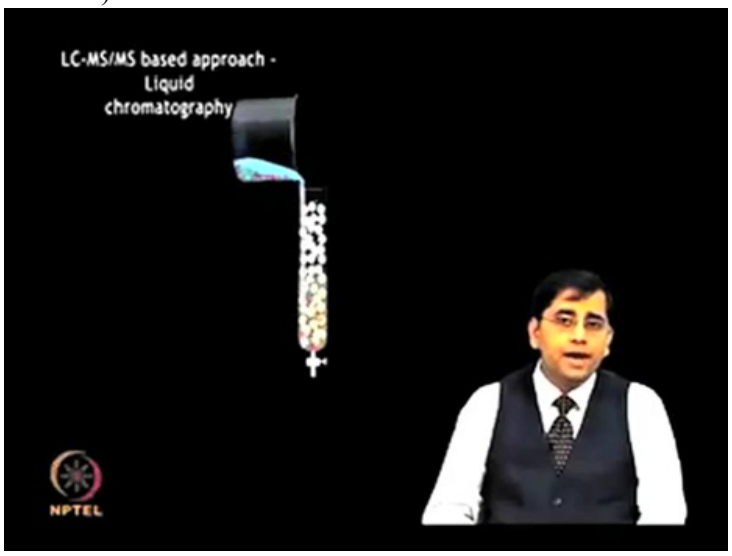
(Refer Slide Time 11:56)



A protein phosphorylation experiment is shown here.

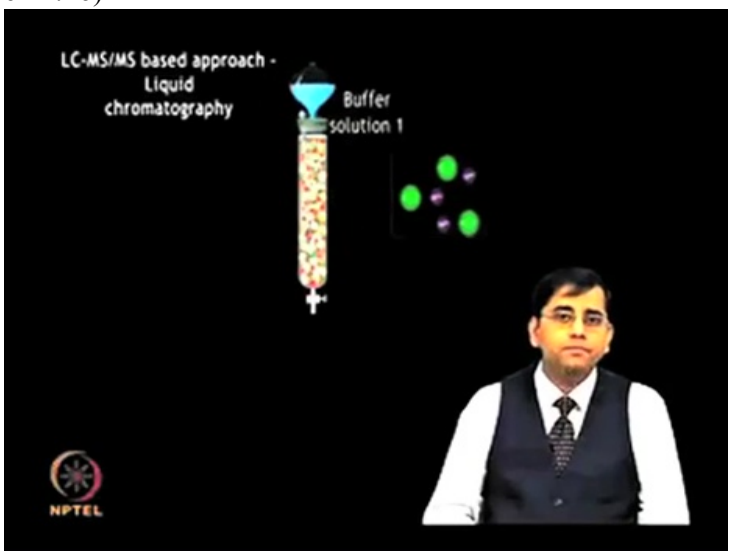
The complex protein sample is loaded onto a miniaturized affinity column,

(Refer Slide Time 12:09)



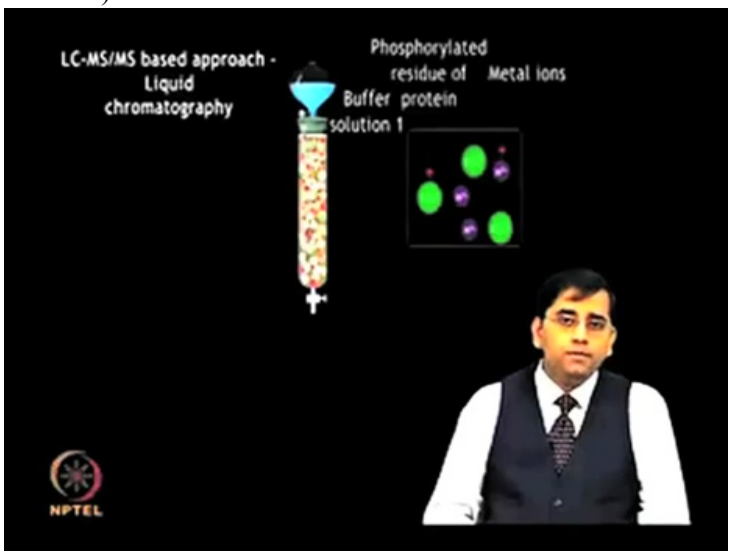
which interacts specifically with proteins having the post-translational modification of interest.

(Refer Slide Time 12:16)



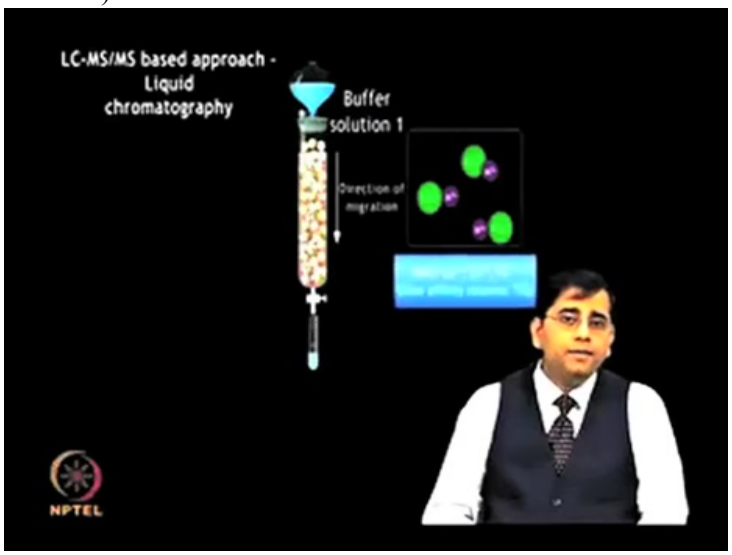
IMAC chromatography columns

(Refer Slide Time 12:20)



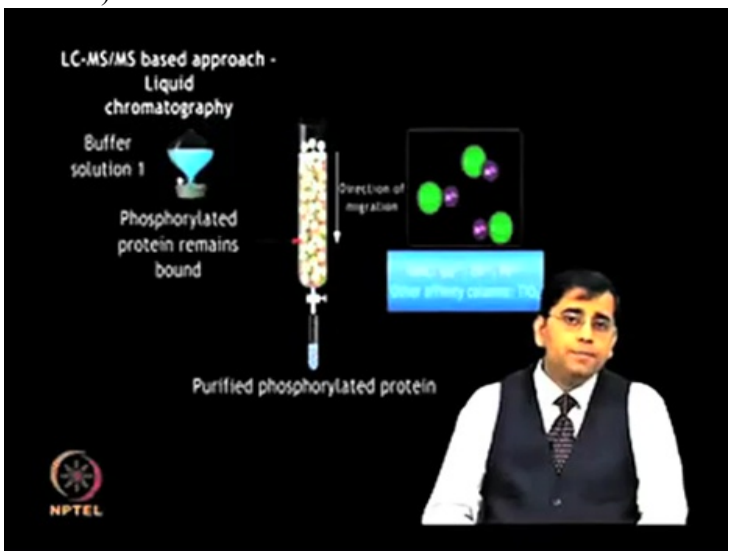
containing ions

(Refer Slide Time 12:25)



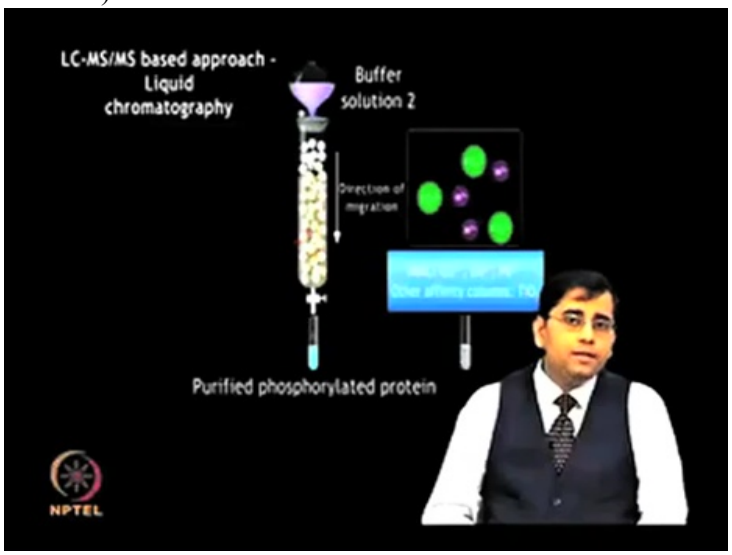
such as zinc, iron, titanium dioxide specifically

(Refer Slide Time 12:27)



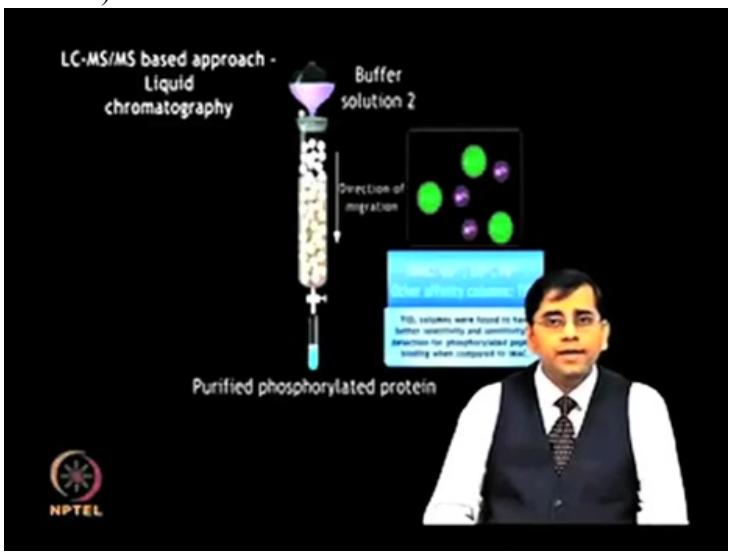
chelate the phosphorylated proteins. The unwanted proteins

(Refer Slide Time 12:34)



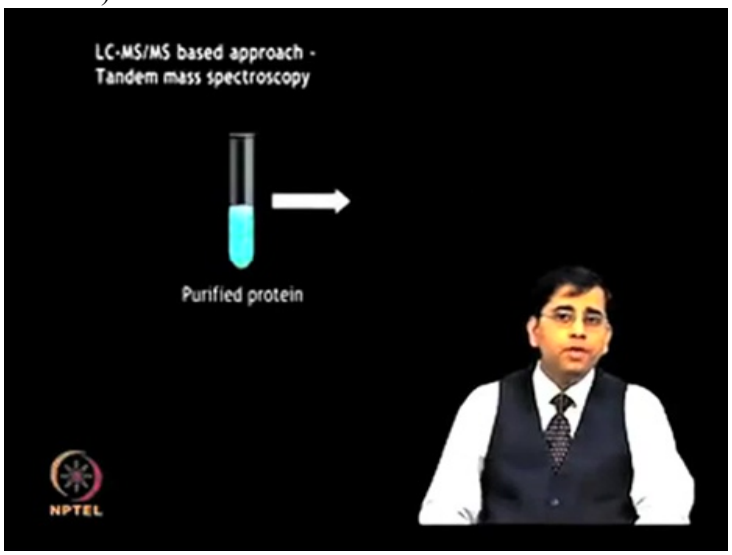
are removed by washing the column with a suitable buffer solution after which the

(Refer Slide Time 12:43)



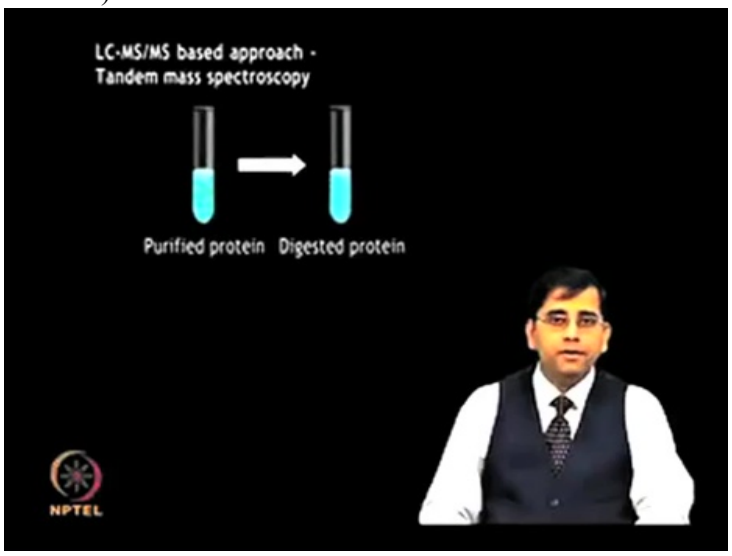
the phosphorylated protein of interest can be eluted out by modifying the buffer solution.

(Refer Slide Time 12:51)



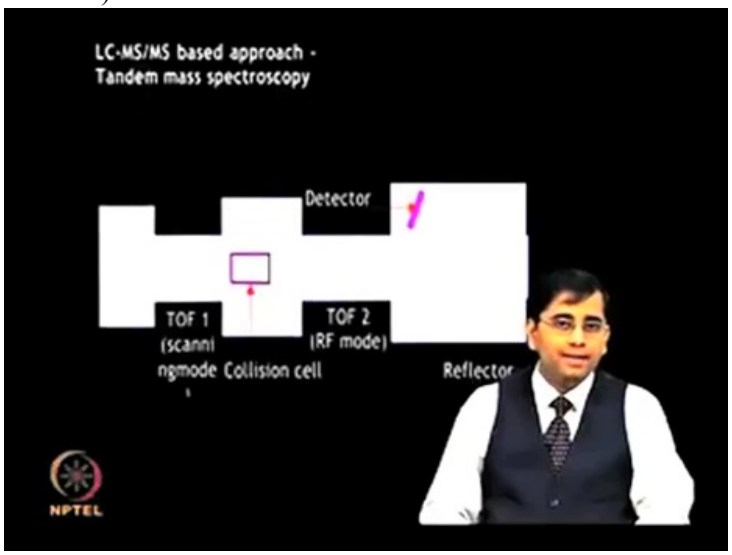
The protein purified by liquid chromatography

(Refer Slide Time 12:55)



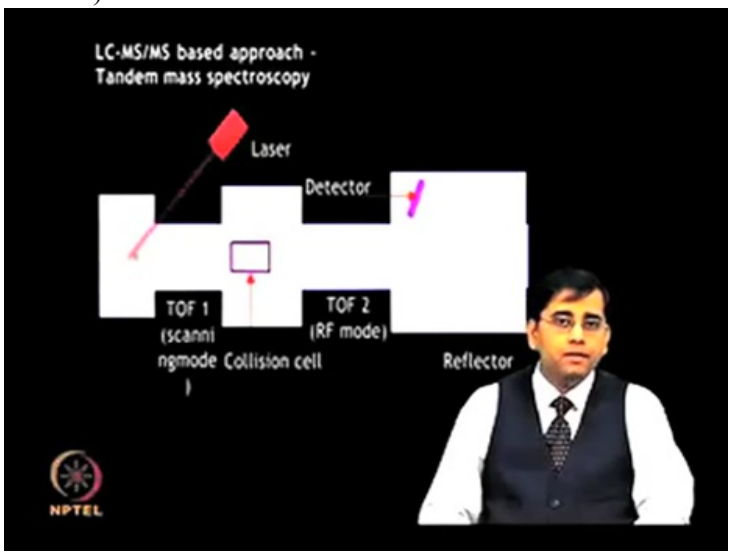
is then subjected to tryptic digestion

(Refer Slide Time 12:59)



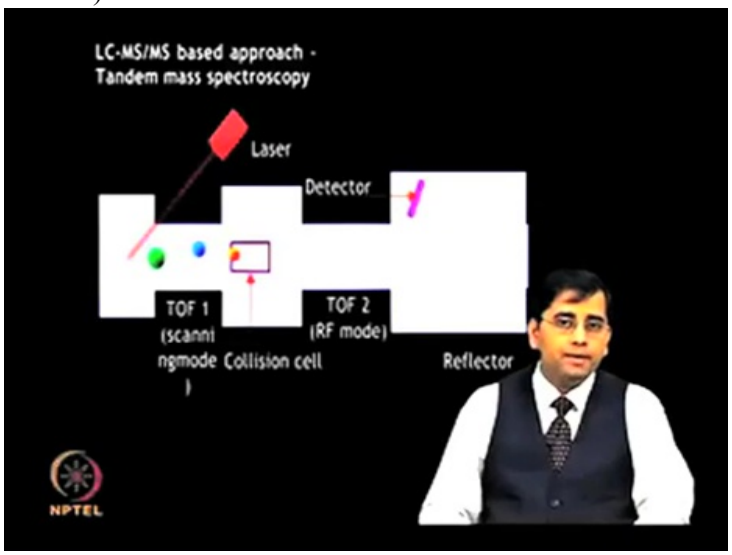
followed by analysis using tandem mass spectrometry. Here I have demonstrated the use of MADI TOF-TOF MS

(Refer Slide Time 13:09)



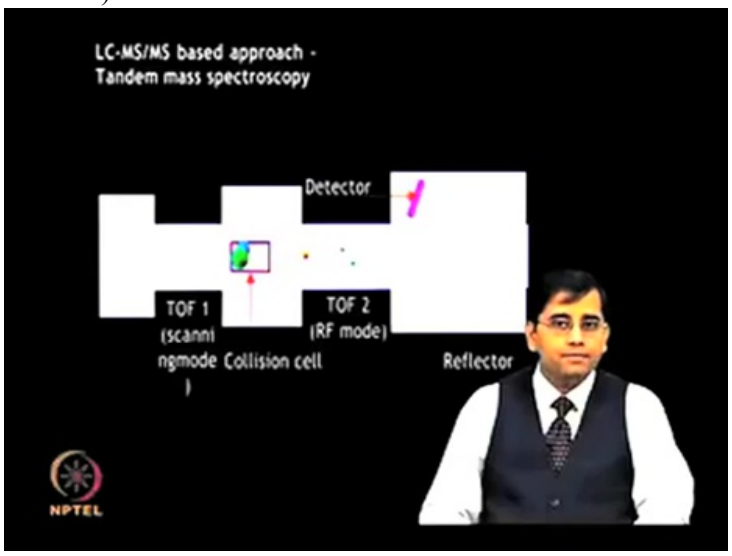
for resolution of the generated ion fragments.

(Refer Slide Time 13:12)



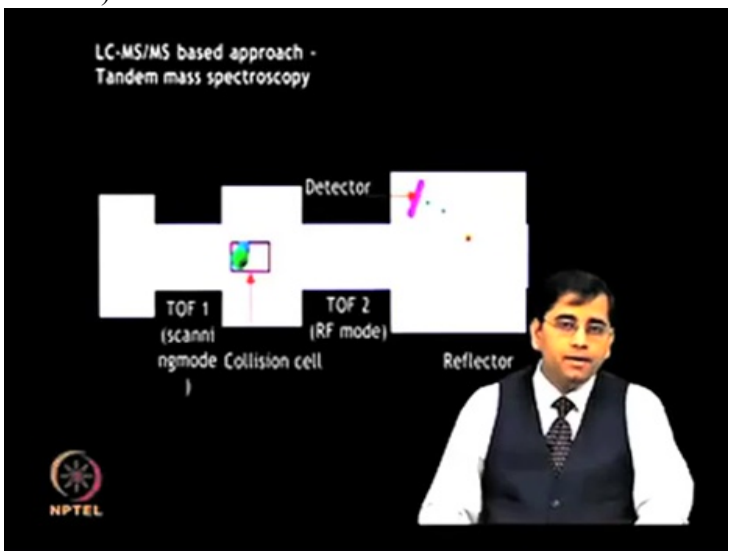
Separation is based on the flight time of the ions and greater resolution is achieved due to the presence of two mass analyzers.

(Refer Slide Time 13:25)



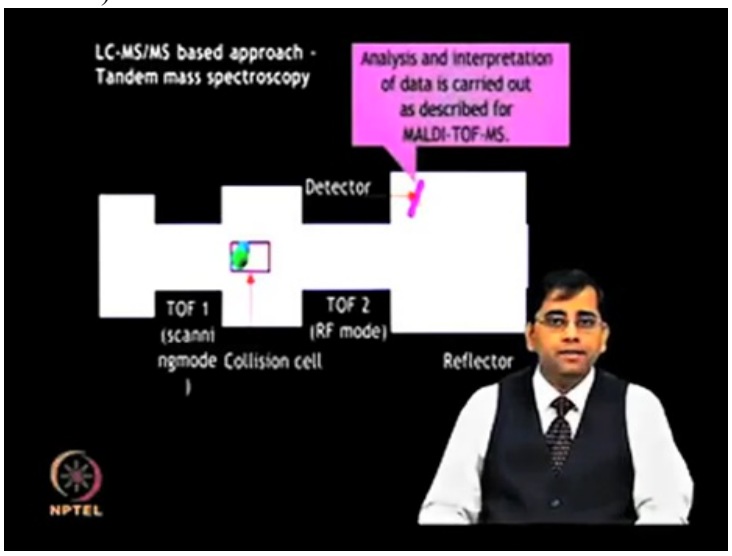
The peptide ion spectrum generated is analyzed by comparing it with the expected spectrum,

(Refer Slide Time 13:36)



thereby allowing determination of modified peptides having

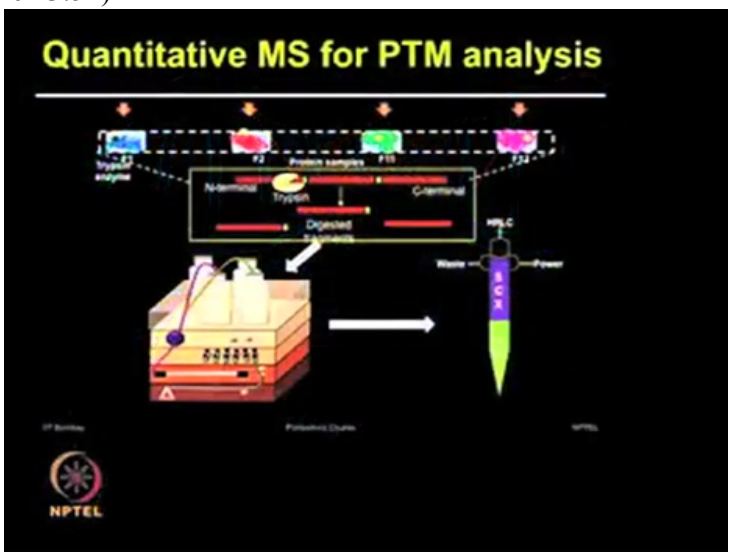
(Refer Slide Time 13:43)



different m by z values.

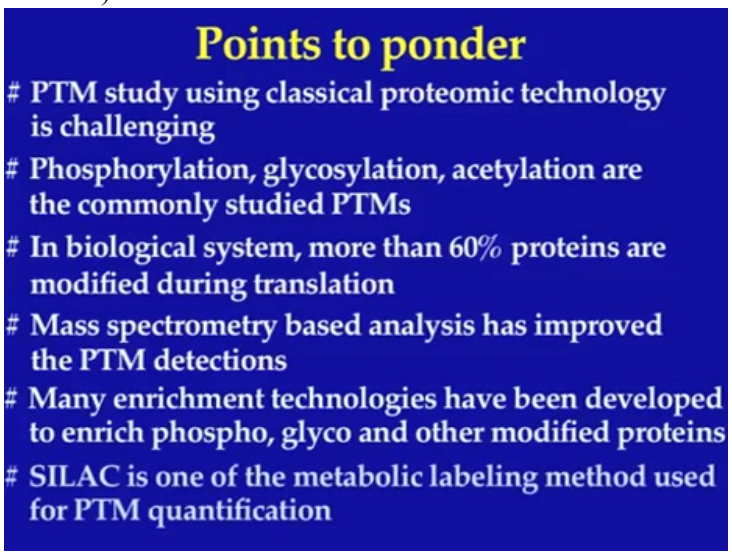
The metabolic labeling methods such as SILAC is used for label-based quantification of PTMs.

(Refer Slide Time 13:54)

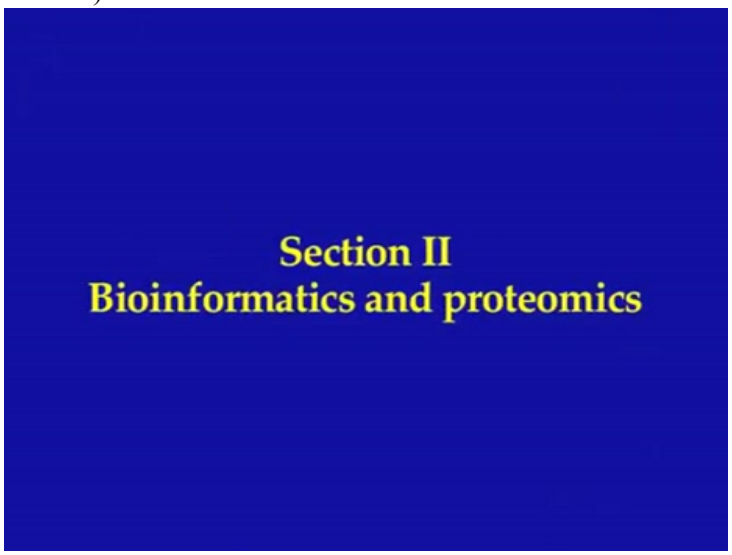


However, this strategy can only be used for the living cells. Other chemical labeling methods such as iTRAQ is also used for PTM analysis.

(Refer Slide Time 14:20)



(Refer Slide Time 14:28)



(Refer Slide Time 14:32)

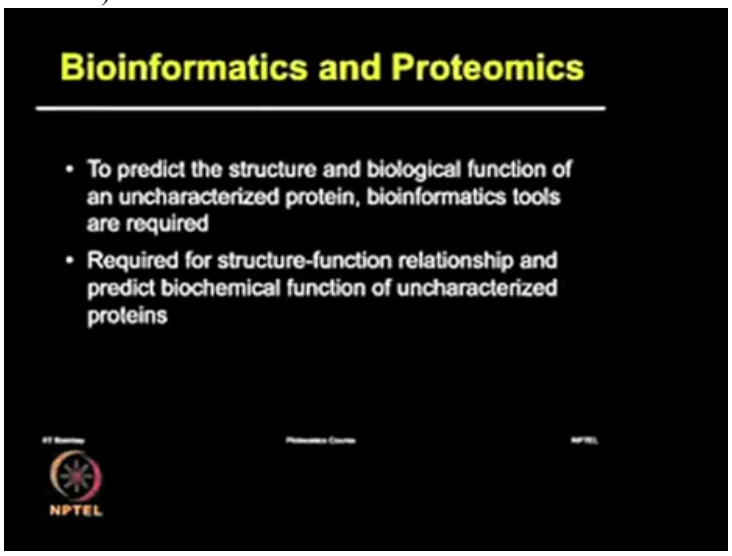
To predict the structture and biological function of an uncharacterized protein, computational methods rely on structural homology of unknown protein from proteins with known structure and biological function.

By relying on such methods for structure-function correlations, it is possible to predict biochemical function of uncharacterized proteins based on structural homology to another protein with a known function

Recent advancement in proteomic and other omics technologies allow large-scale analysis of biological samples, and generate an unprecedented amount of digital data.

In different modules, we have discussed different bioinformatics tools and software for analysing proteome and system level investigation using two-dimensional electrophoresis, mass spectrometry, microarrays and surface Plasmon Resonance.

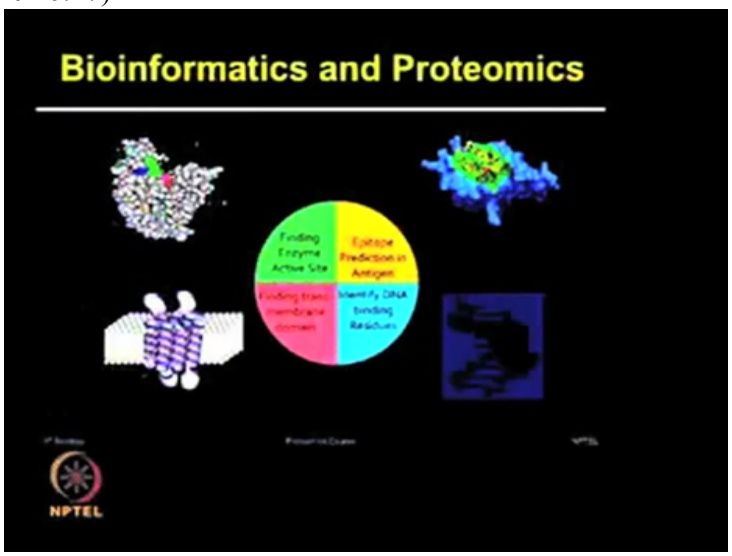
(Refer Slide Time 15:49)



Computational challenges associated with proteomic studies have recently emerged as some of the most critical and limiting factors in this rapidly evolving discipline.

Bioinformatics tools have been widely used for protein sequence analysis. It is also used

(Refer Slide Time 16:17)



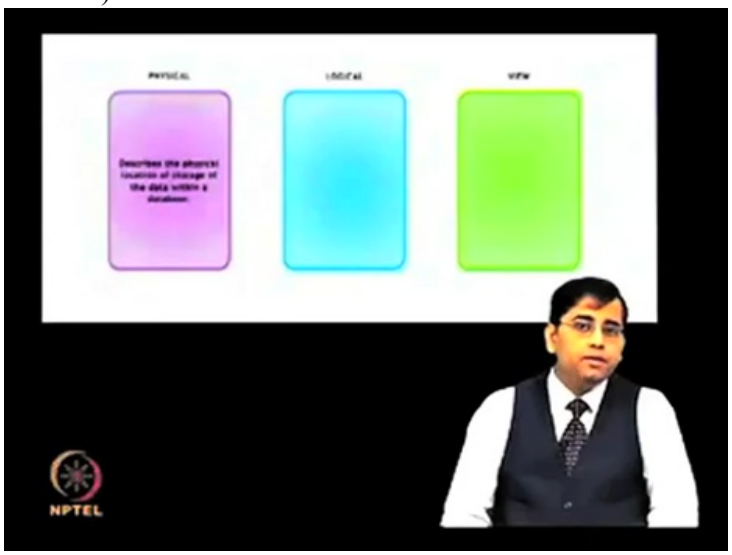
for protein motif detection and Epitope Prediction, Active site determination, determining trans-membrane domains as well as Identification of DNA binding residues.

Database designing is done at various levels

(Refer Slide Time 16:37)

such as Physical, Logical and Conceptual.

(Refer Slide Time 16:44)



At the physical level, the purpose of the database is defined which is

(Refer Slide Time 16:49)

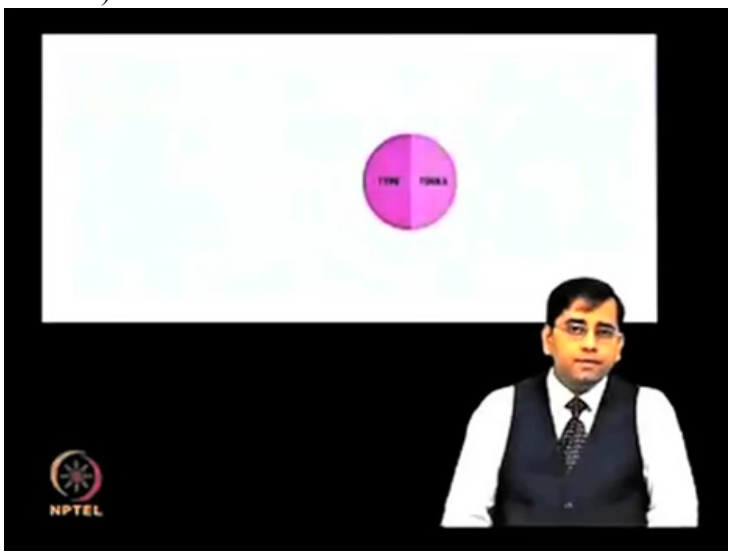
in accordance with the proposed usage.

At the logical level, the tables, attributes of the tables and data-types are defined. At the View level, the views and appearance

(Refer Slide Time 17:06)

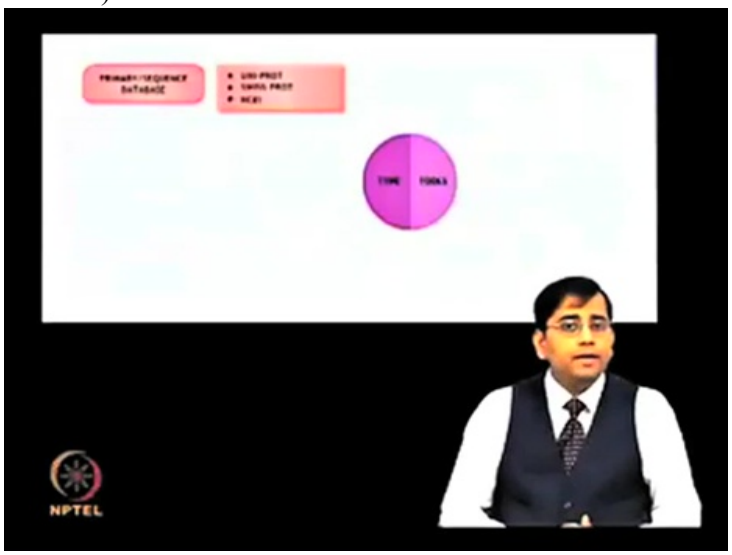
of the database are defined. A typical biological database

(Refer Slide Time 17:12)



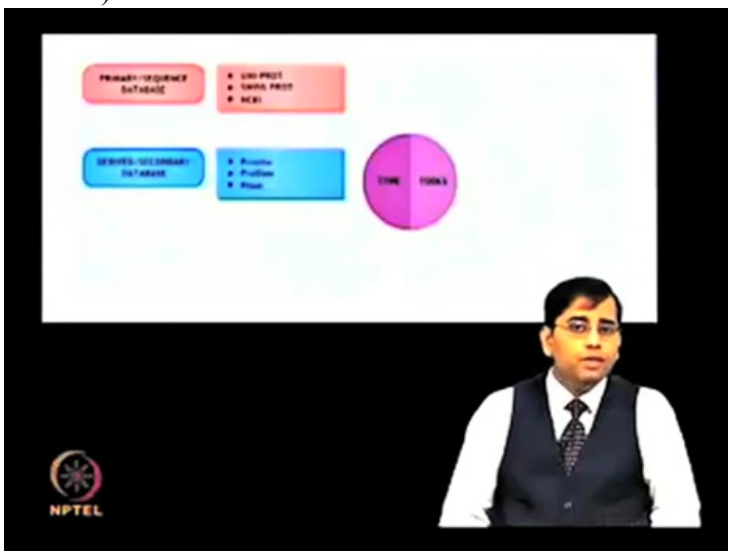
can be characterized by its Type and its Tools.

(Refer Slide Time 17:19)



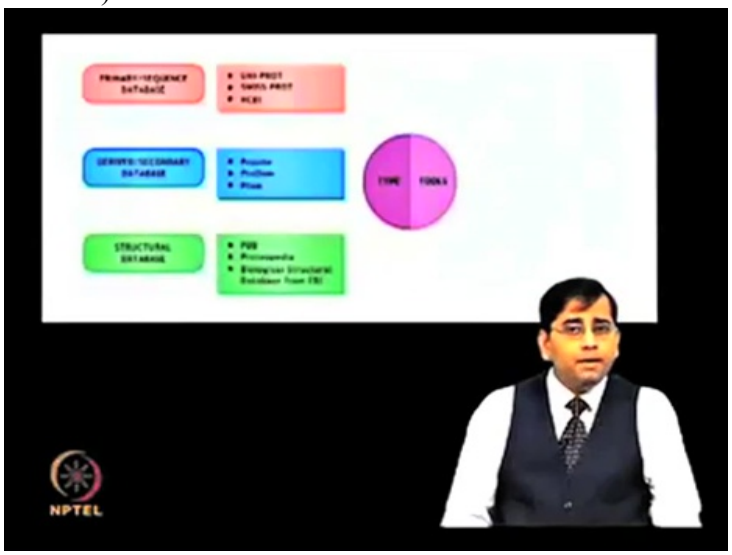
The Type defines the category of data that it includes, such as

(Refer Slide Time 17:25)



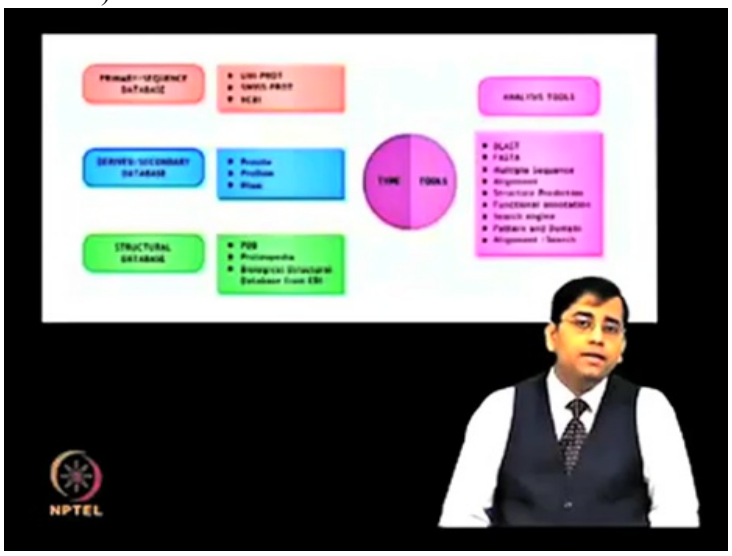
sequence, domains or structure. This implies

(Refer Slide Time 17:31)



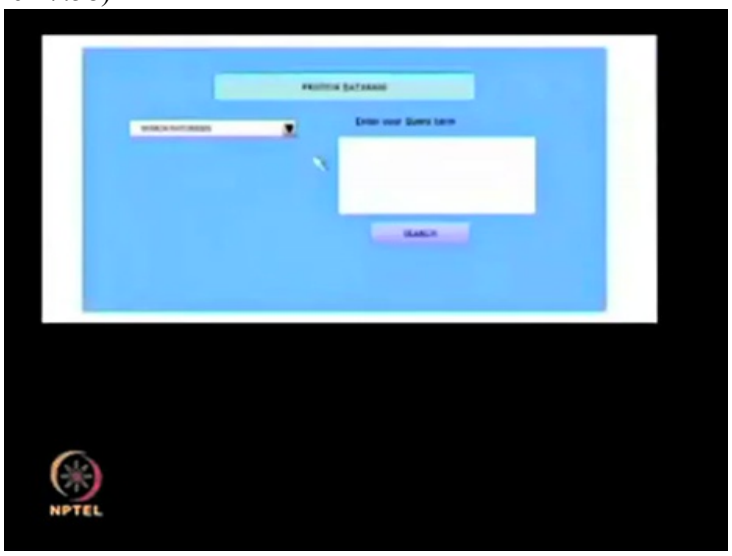
that the particular database, its most prominent feature includes either

(Refer Slide Time 17:38)



sequences, domains or structure and it is particularly used for their analysis. The analysis tools defines the platforms that the site will provide for gaining an insight into the protein data.

(Refer Slide Time 17:58)



For extracting the protein information from a database,

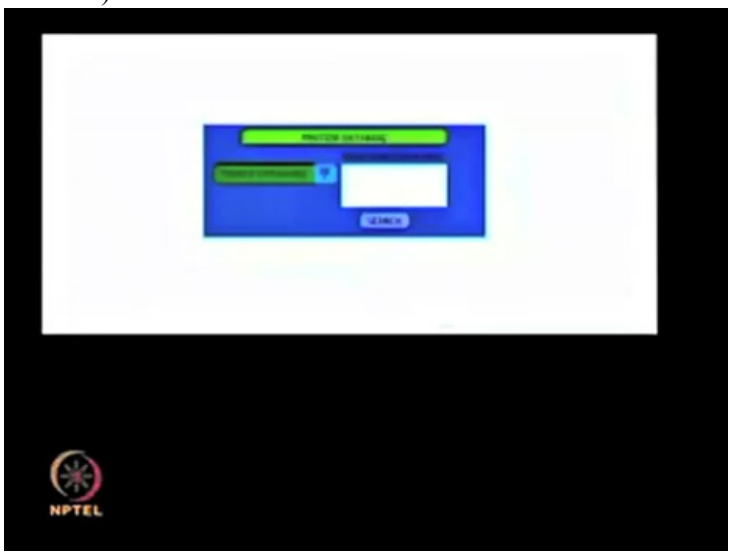
(Refer Slide Time 18:02)

users can give a variety of

(Refer Slide Time 18:05)

input terms. These can be: Unique ID, Molecular Name, Amino-acid sequence, Keyword, Literature, Gene, Taxonomy etc.

(Refer Slide Time 18:21)



Once the user submits the query, the output can be of multiple formats.

The generalized information that users can obtain from protein databases is

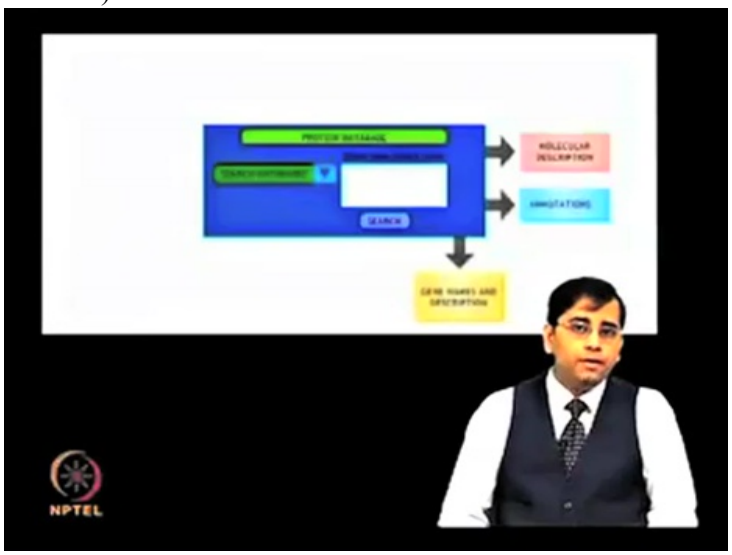
(Refer Slide Time 18:35)

General Description of the protein molecule,

(Refer Slide Time 18:40)

The generalized information that users can obtain

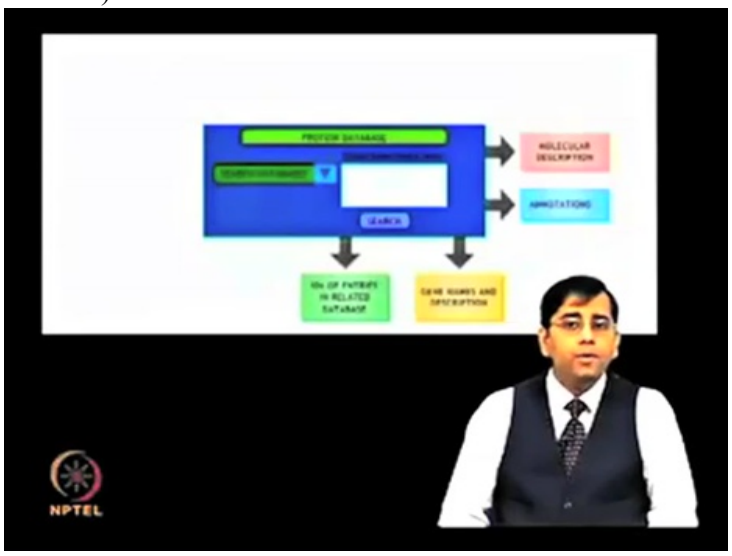
(Refer Slide Time 18:44)



from protein databases is

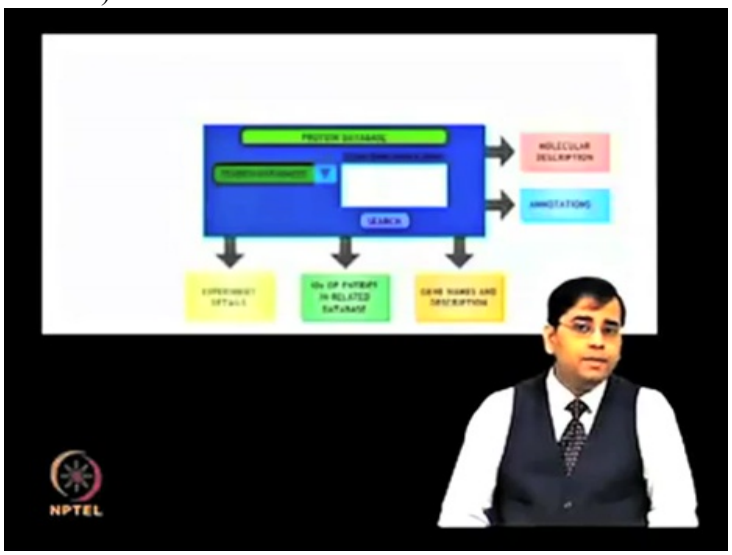
General Description of the protein molecule,

(Refer Slide Time 18:51)



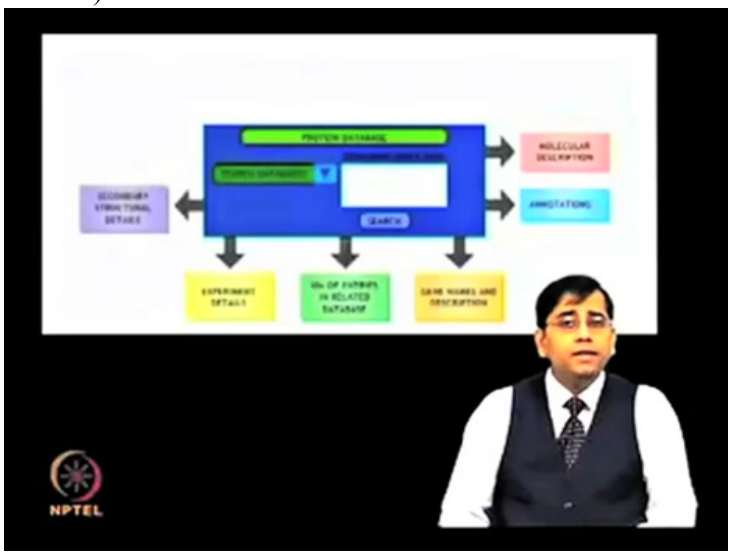
annotations of the protein, name and description of the gene that transcribes them

(Refer Slide Time 18:57)



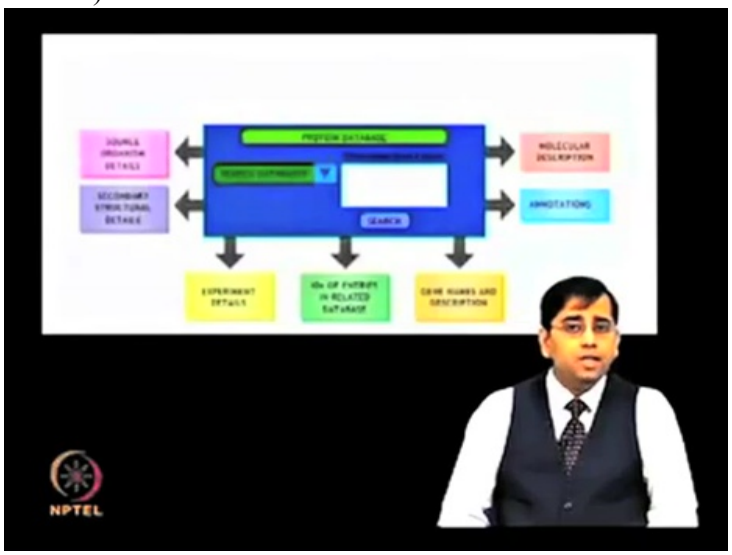
, ID of the same protein in other relevant databases,

(Refer Slide Time 19:06)



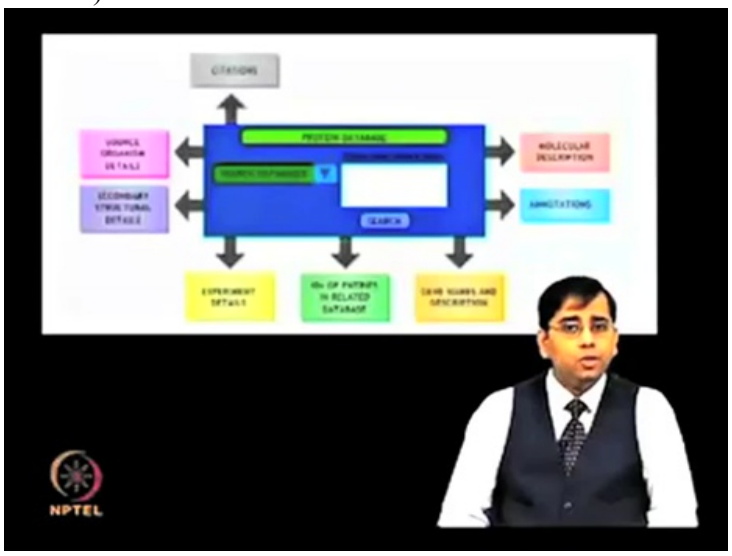
Details of the experiment conducted for characterizing proteins,

(Refer Slide Time 19:11)



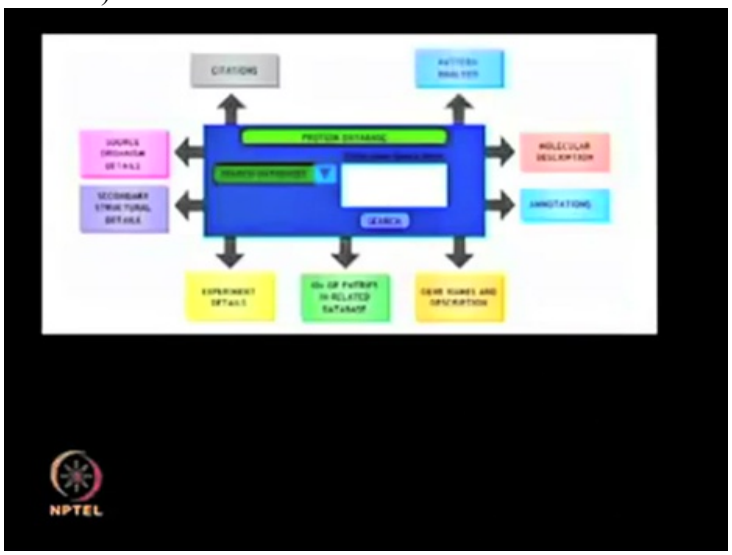
Details of the Protein's secondary structures,

(Refer Slide Time 19:19)



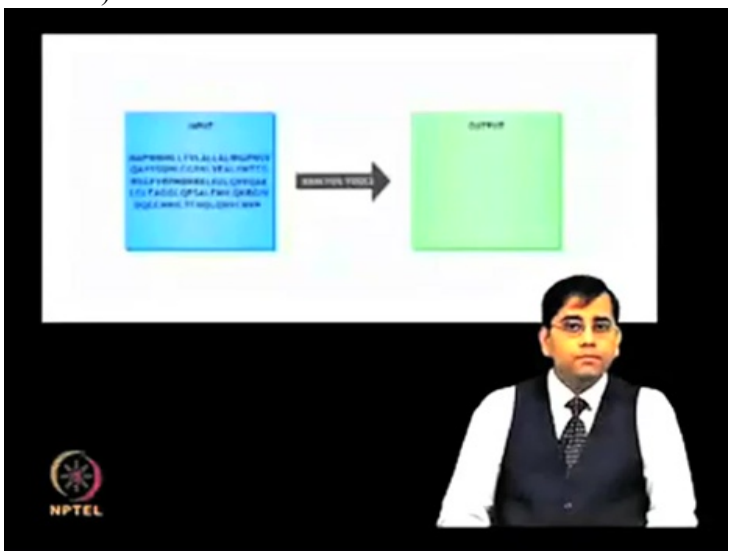
Details of the organism which was used as a source for obtaining the protein

(Refer Slide Time 19:26)



and Citations of research conducted

(Refer Slide Time 19:30)



Database analysis tools

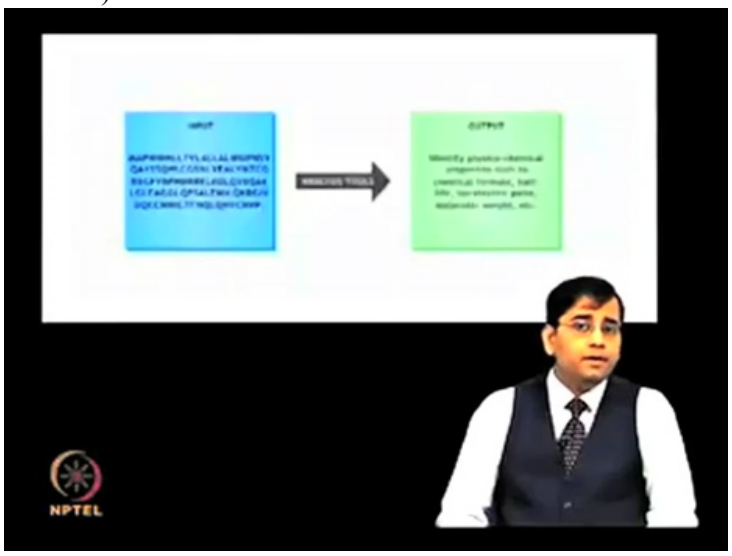
Different kinds of analysis can be conducted on a given protein sequence. The query can be the protein name, sequence or any other identifier of the protein.

Various kinds of results output can be obtained. Identity of protein from sequence, identify physico-chemical properties,

(Refer Slide Time 19:58)

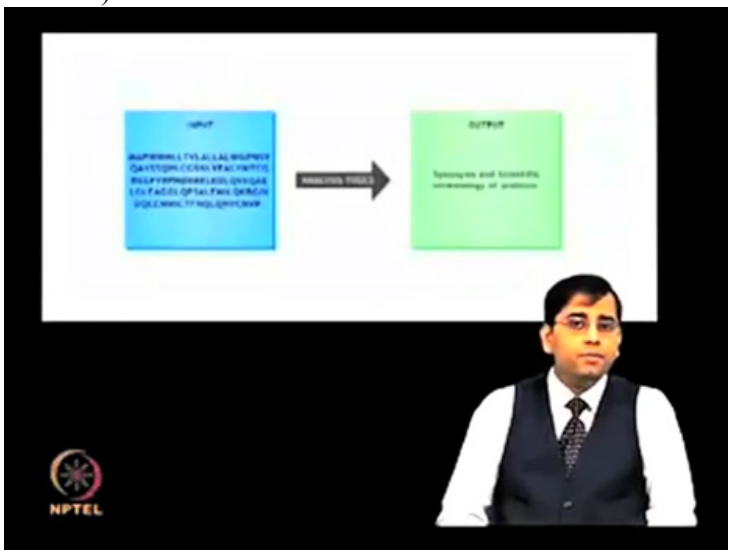
molecular weight, iso-electric point, sequence tag information;

(Refer Slide Time 20:06)



Similarly search algorithms such as versions of BLAST, FASTA and Multiple Sequence Alignment; Finding conserved and variable domains in the protein to study its evolutionary relationships with other proteins; Molecular modeling and visualization tool, Secondary and tertiary structure prediction and structural analysis,

(Refer Slide Time 20:37)



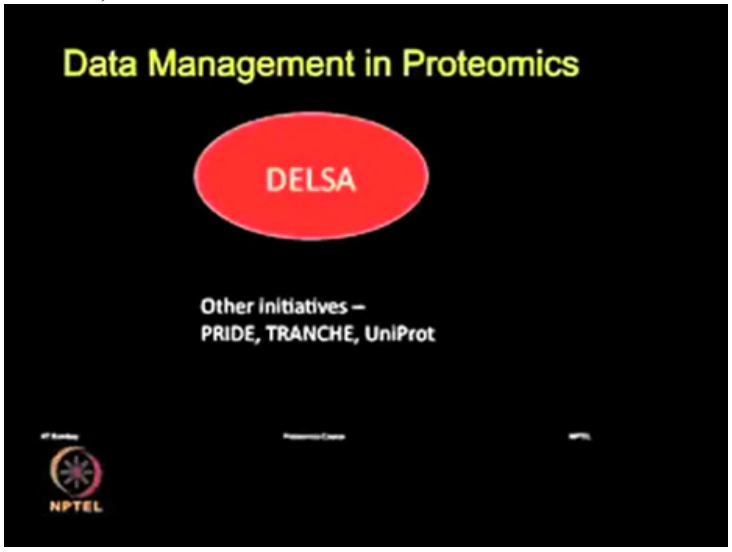
Biological text analysis such as bio-medical acronyms, gene-protein synonyms etc.

Database Mining in Proteomics and Visualization Tools

Collective improvement in any research field can be accelerated by sharing scientific data among different research groups across the world, seeing as it allows other researchers to

access, validate and reanalyze one's findings and correlate the results with their own observations.

(Refer Slide Time 21:21)

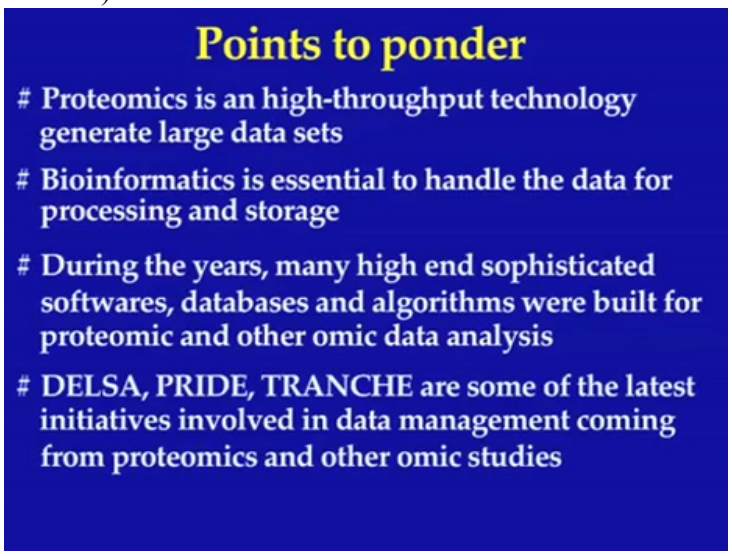


Data management is critical when using high-throughput proteomic techniques. Several Internet databases have been established to collect the proteomic datasets.

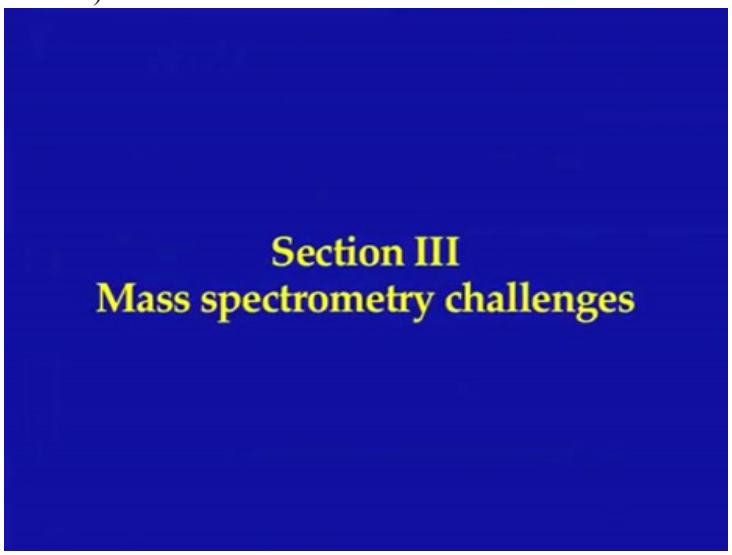
Data-Enabled life Sciences' Alliance (DELSA) is a timely and important initiative to create a common data bank where, on one hand we can access the huge data set generated by the various research groups worldwide, on the other hand we can also deposit our datasets, which may be useful for a wide range of researchers working in similar fields.

At present, the broad field of DELSA encompasses biological sciences, ecology, environmental sciences, evolution, genomics and proteomics, computer sciences, cyber infrastructure, management, health sciences, and policies for global distribution.

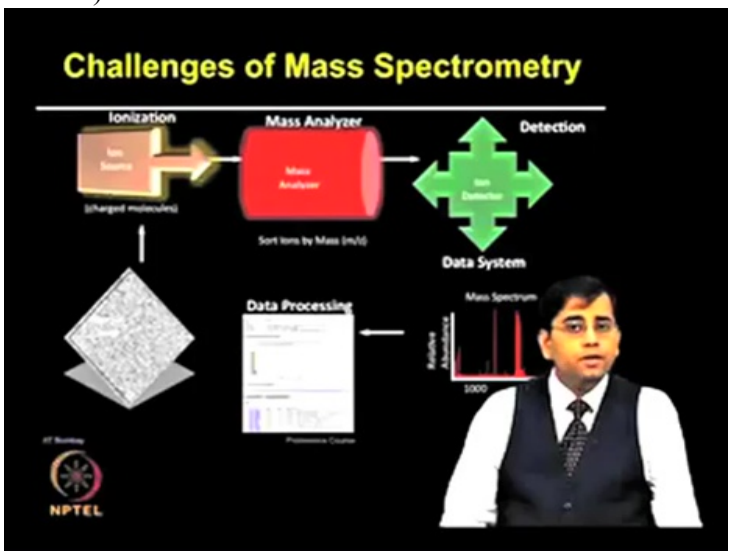
(Refer Slide Time 22:44)



(Refer Slide Time 22:50)



(Refer Slide Time 22:57)



MS-based proteomics encounters the following biological problems while analysis of huge number. of proteins; fragile nature of proteins, substantial loses occur during sample collection and processing steps, presence of multiple isoforms of single protein, wide dynamic range of protein concentrations in biological fluids, presence of high-abundance proteins masking low-abundance marker proteins

Additionally, technological limitations associated with most of the commonly used MS-based approaches include typical dynamic ranges of only 10 to the power 2 to 10 to the power 4, inadequate coverage of whole proteome unless sample is fractionated extensively, low-throughput and issues of robustness and cost, over fitting the data, machine fluctuation, instrument noise and contaminants in spectrum and lack of standard procedure for analysis and interpretation of MS and MS/MS spectrum.

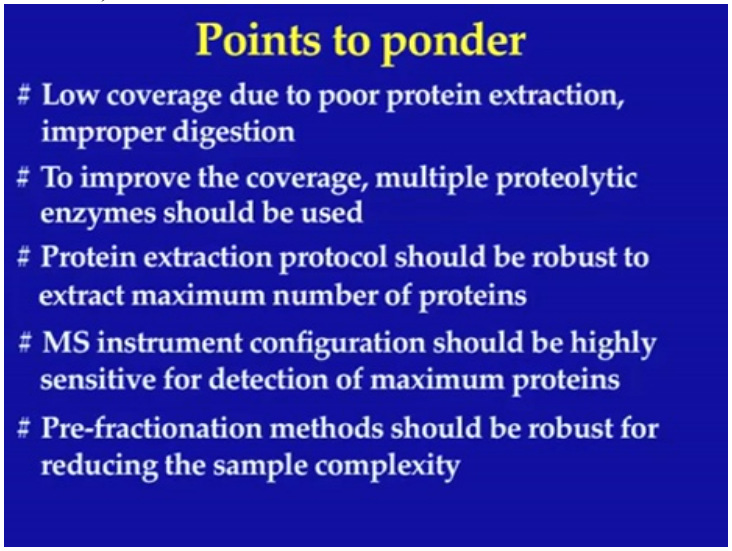
To overcome the technological challenges, different novel and amalgamated approaches have emerged in last few years.

The most promising advancements includes large scale quantitative proteomics, Culture-Derived Isotope Tags and Super-SILAC based technologies

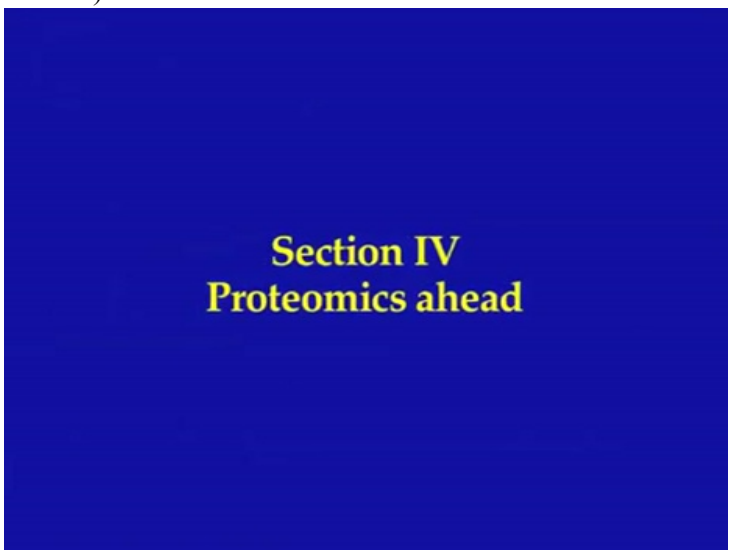
Multiplexing; Tandem Mass Tags (TMT) and iTRAQ 8-plexing, Quantitative accuracy: label-free LC-MS/MS, Low sample consumption and large-scale analysis, Chip-based and

Nano-LC-MS, Sensitive quantitation of proteins within complex mixtures, biomarker discovery, multiple reaction monitoring MRM MS, Large-scale biomarker discovery etc.

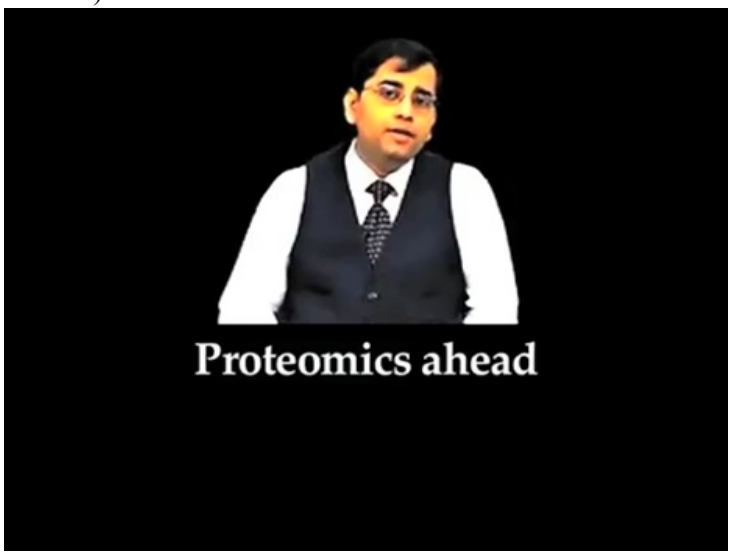
(Refer Slide Time 25:49)



(Refer Slide Time 25:53)



(Refer Slide Time 26:00)



Impending future of this promising research area will highly depend on the collaborative initiatives at a global level and establishment of effective data repositories accessible to the proteomic researchers across the world.

(Refer Slide Time 26:20)

In 2010 Human Proteome Organization (HUPO) has launched a global Human Proteome Project HPP. This project is designed to map the entire human proteins encoded by the genome.

Let us now discuss some of the targetted focused initiative.

(Refer Slide Time 26:44)

The Human Liver Proteome Project, this is the first initiative of the human proteome project for human organ, tissues with an intention of generation of comprehensive protein atlas of the liver and international liver tissue network, collection and distribution of normal liver samples and validation of new discoveries.

(Refer Slide Time 27:19)

Human Plasma Proteome Project, analysis of the proten constituents of human plasma and serum.

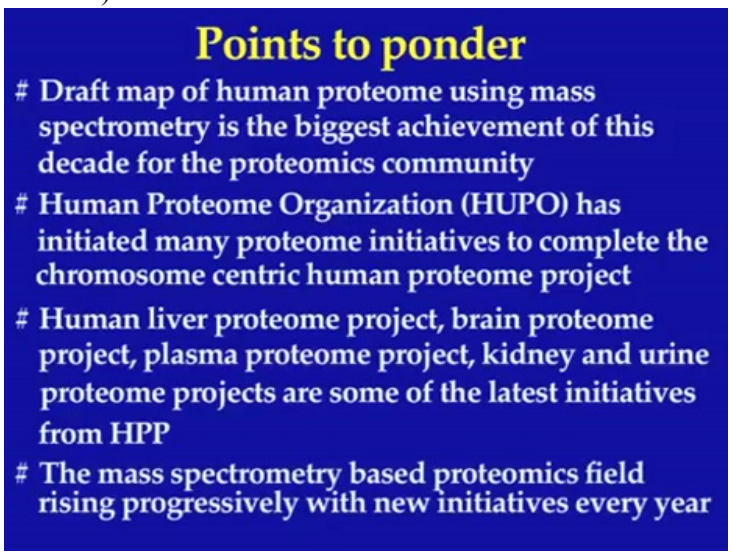
(Refer Slide Time 27:31)

Human Brain Proteome Project BPP focusses on the revolution of the brain related proteomics alteration, focusing on understanding neuro-degenerative diseases, aging and identification of prognostic and diagnostic biomarkers.

(Refer Slide Time 27:53)

Human Kidney and Urine Proteome Project aims to understand kidney functions, mechanism of chronic kidney diseases at a protein level and discover biomarkers and target molecules for due thereupatics of kidney diseases.

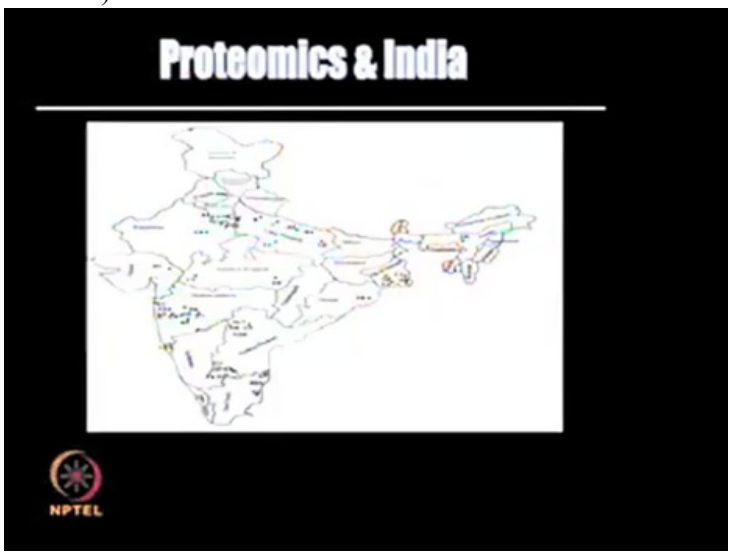
(Refer Slide Time 28:22)



(Refer Slide Time 28:29)

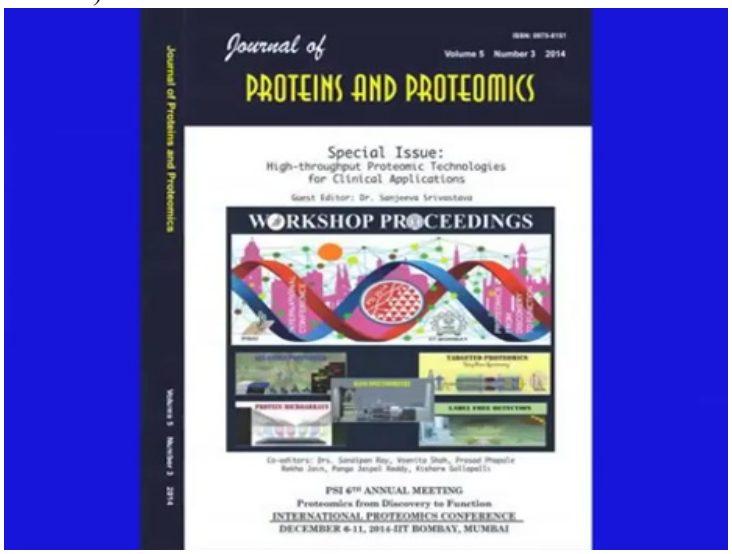
Over the last decade proteomics research is progressing in different regions of India with a considerable interest.

(Refer Slide Time 28:36)

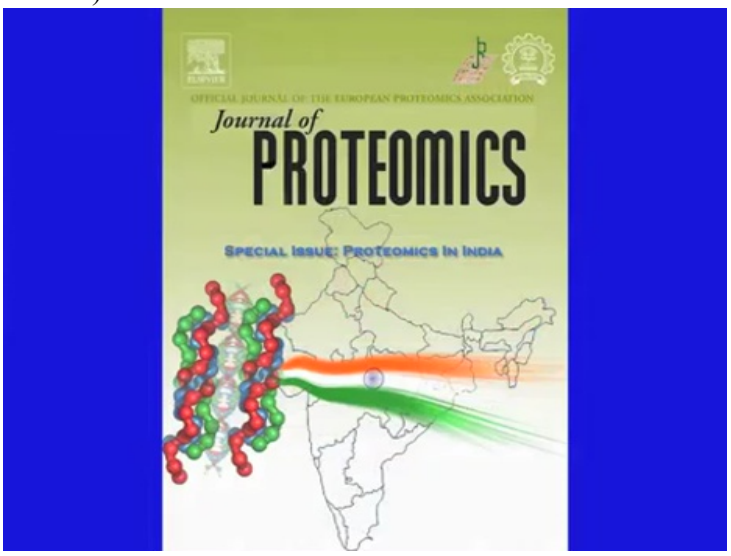


India is playing an increasingly significant role in global genomics and proteomics Research and Development, as evident from recent publications and patents.

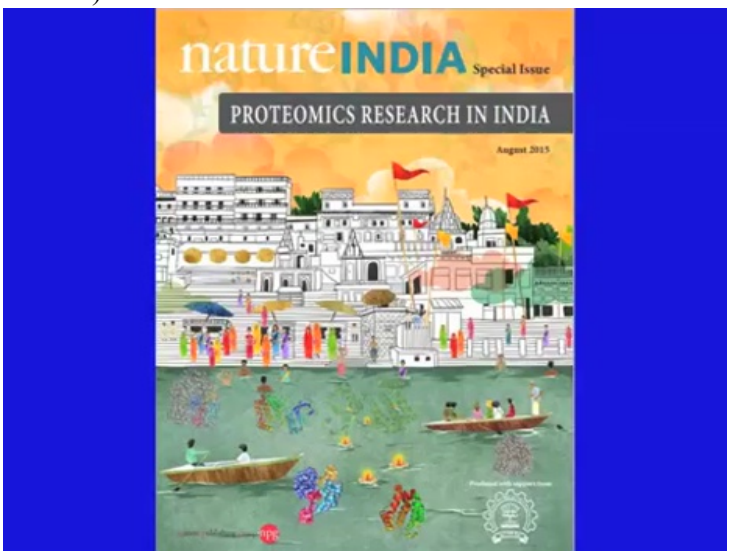
(Refer Slide Time 28:57)



(Refer Slide Time 28:59)



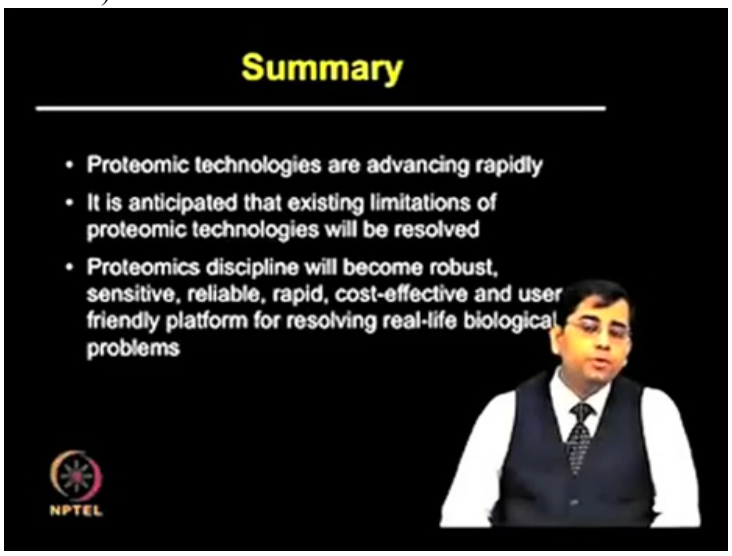
(Refer Slide Time 29:02)



Indian government is also supporting the basic and applied proteomics as well as other omics-based research and multiple national and international funding agencies are providing investments on existing and new research projects.

Considering the space of emerging proteome-level research, it can be anticipated that in coming future some amicable solutions for the existing limitations associated with

(Refer Slide Time 29:42)



the burgeoning field of proteomics will come forward through world-wide research initiatives and this discipline will become more robust, sensitive, reliable, rapid, cost-effective and user-friendly for resolving real-life biological problems.

(Refer Slide Time 30:06)

Hope this course has given you foundation for proteomic concepts and enthused you for research in proteomics area. Thank you.

(Refer Slide Time 30:22)