



**FACULTY OF AGRICULTURE SCIENCES AND
ALLIED INDUSTRIES**

(Principles of Biotechnology)

For

M.Sc. Ag (GPB)



RAMA
UNIVERSITY

www.ramau.ac.in

Course Instructor

Dr Shiv Prakash Shrivastav

FASAI(Genetics and Plant Breeding)

Rama University, Kanpur

Introduction to Proteomics

Proteomics is a field of global study of the expression of genetic information at the protein level (proteome). It also deals with assessment of three-dimensional structure of proteins and their interactions.

Human Proteome Organization (HUPO, www.hupo.org) has postulated in 2001 the aims of proteomics in a more precise manner as identification of all proteins coded by human genome (and other genomes, especially those of model organisms) with the subsequent assessment of a) their (protein) expression in different cell types of the organism - expression proteomics, b) protein distribution in subcellular compartments of the organelles, c) post-translational modifications of the proteins, d) protein-protein interactions (b-d constitute the **structural proteomics**) and e) relation between protein structure and function (**functional proteomics**).

The wide range of the definition of proteomics precipitates into an array of distinct sub-specialties, e.g. Clinical proteomics is trying to identify via the analysis of the protein spectrum in the biological samples (urine, saliva, blood) new markers that can be used in future early diagnosis and treatment of a specific disease. The major role in all spheres in early diagnosis and treatment of a given disease is played by bioinformatics (digitalization, 2D picture analysis of the gels, modeling 3D protein structures and the networks of protein interactions, creation, care and development of publicly available databases). One of the interesting proteomic projects is Swedish Human Protein Atlas (<http://www.proteinatlas.org/>), that groups the information about expression of individual proteins in different tissues but its uniqueness lies in the accompanying archive of immunohistochemically stained samples - at the time of writing of this text, there were more than 400,000 available. Even though it may seem that because of the availability of RNA/cDNA microarray chips for analysis of expression on the level of transcripts there is enough detailed information about the currently ongoing status of expression of genetic information, it is necessary to understand that

1. not all mRNAs will be translated into the protein
2. The level of transcription of specific, protein-coding RNA not always corresponds to the level of expression and moreover, to activity of the coded protein due to many factors (mRNA, RNA splicing, posttranslational protein modifications, etc.)

Similar to other "-omics" the development of proteomics was significantly influenced by the recent developments in technology. The real development of the proteomics started only after the classical method of 2D electrophoresis was joined by "soft" ionization techniques of mass spectrometry. These techniques are performed using matrix assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI).

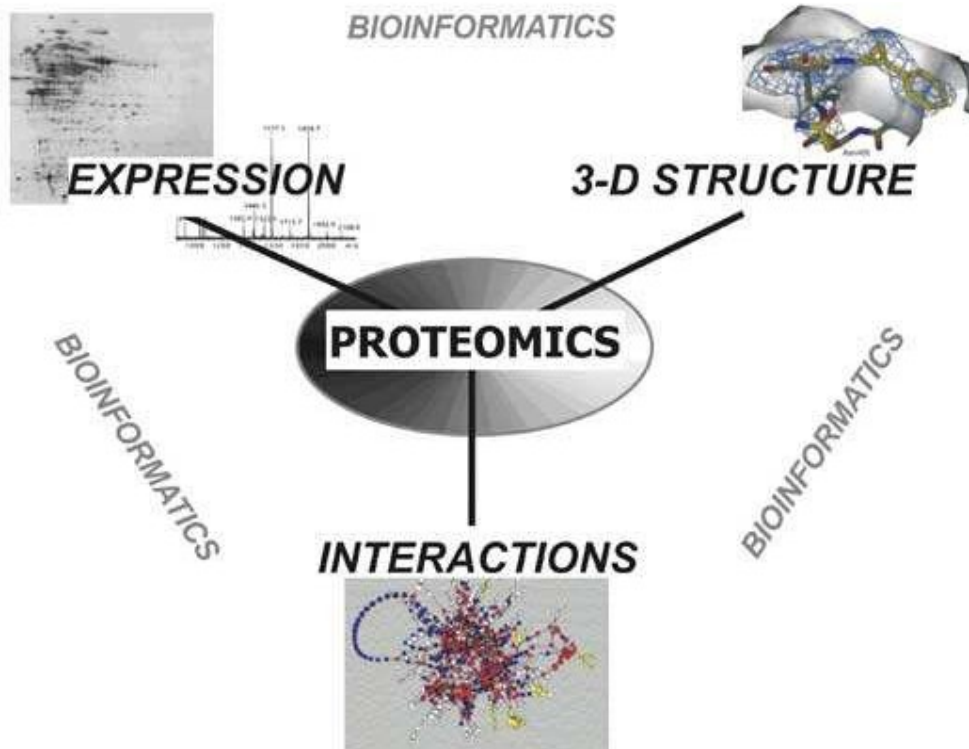


Figure 1: Major proteomics directions

Proteomic methods

1. 2D protein electrophoresis

Nowadays the two-dimensional (2D) protein electrophoresis is a method of choice for protein separation. This method enables to distinguish up to 10 000 proteins. The method is based on two distinct physical and chemical features of proteins: first, the proteins are separated according to their isoelectric point (pI). Isoelectric point is such a pH value, where the overall protein charge equals to zero. This can be obtained by creating a pH gradient in the gel where protein is loaded and electric current is applied. Proteins then migrate towards cathode or anode according to their total charge up to the point where the gel pH equals pI of a given protein. After the separation in the first direction, common electrophoresis on a polyacrylamide gel (PAGE) is applied, but the electric current is applied perpendicular to the original orientation of electrodes. Proteins then migrate in the second direction through gel only according to their size. After both phases of 2D electrophoresis, it is necessary to visualize proteins by one of the staining or labeling methods (chemical or radioactive). The resulting "maps" of proteins can be compared for example in between the experimental and control sample or among the samples from patients with specific disease and their healthy controls and thus identify differentially expressed proteins that can be linked with the pathogenesis of the studied disease. The identity of differentially expressed proteins is verified after "cutting out" the area of the gel with given difference and subsequent analysis using mass spectrometry.

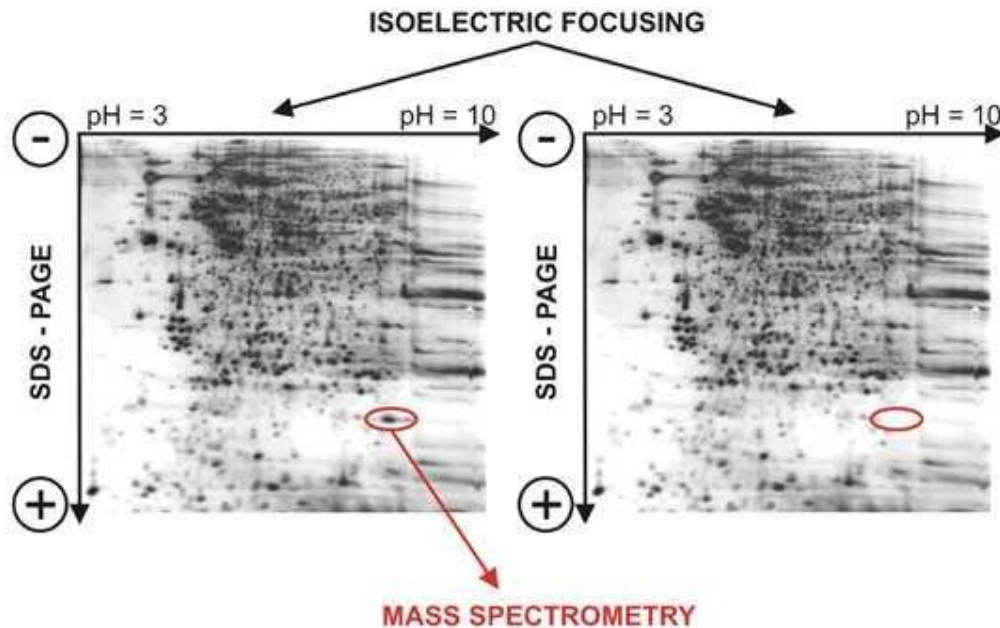


Figure 2. 2D electrophoresis. After isoelectric focusing in pH gradient of the electric field the electrophoretic separation is used in polyacrylamide gel (SDS-PAGE....sodium dodecylsulphate - polyacrylamide gel electrophoresis). Proteins can be visualized after staining (on the picture in Coomassie blue).

2. Mass spectrometry (MS)

Mass spectrometry is a method that enables precise measurement of molecular weight of a broad spectrum of substances. As the studied substance has to be intact in gas phase, the use of mass spectrometry for protein analysis (but also of polysaccharides and oligonucleotides) was only enabled by development of "soft" ionization techniques of mass spectrometry, like matrix assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). Protein identification is generally performed in two ways:

1. Protein is digested by trypsin or by other proteolytic enzyme to smaller peptides and their precise molecular weights are measured using MS. The spectrum of those molecular weights is then compared with theoretical spectra that are calculated from protein sequences from available databases (using bioinformatics tools).
2. Tandem MS enables to choose the peptide which is then fragmented by the collision with inert gas. The fragmentation pattern gives either full or partial information about protein sequence that is subjected to the search in databases.

Apart of protein identification, MS is a priceless tool for protein posttranslational modification analysis, because it enables to localize given modifications within the protein and also helps to find out the nature of such modification.

Proteomics in medicine

As stated above, one of the main goals of proteomics is the identification of novel markers that can be used for prediction, prevention, diagnosis, prognosis and therapy optimization in human disease. To use proteomics in real everyday clinical practice it is necessary that the biological material where such marker would be analyzed is easily accessible (blood, urine, saliva, cerebrospinal fluid). The other way to use proteomics is mainly within the experiment,

one of the emerging applications is in the field of therapeutical drug discovery and pharmacoproteomics.

Sources/Other literature

- Molecular Biologist's Guide to Proteomics [Molecular Biologist's Guide to Proteomics](#)

Lecture 1: Introduction and scope of Proteomics

The word “proteome” represents the complete protein pool of an organism encoded by the genome. In broader term, Proteomics, is defined as the total protein content of a cell or that of an organism. Proteomics helps in understanding of alteration in protein expression during different stages of life cycle or under stress condition. Likewise, Proteomics helps in understanding the structure and function of different proteins as well as protein-protein interactions of an organism. A minor defect in either protein structure, its function or alternation in expression pattern can be easily detected using proteomics studies. This is important with regards to drug development and understanding various biological processes, as proteins are the most favorable targets for various drugs. Proteomics on the whole can be divided into three kinds as described below (Fig. 1):

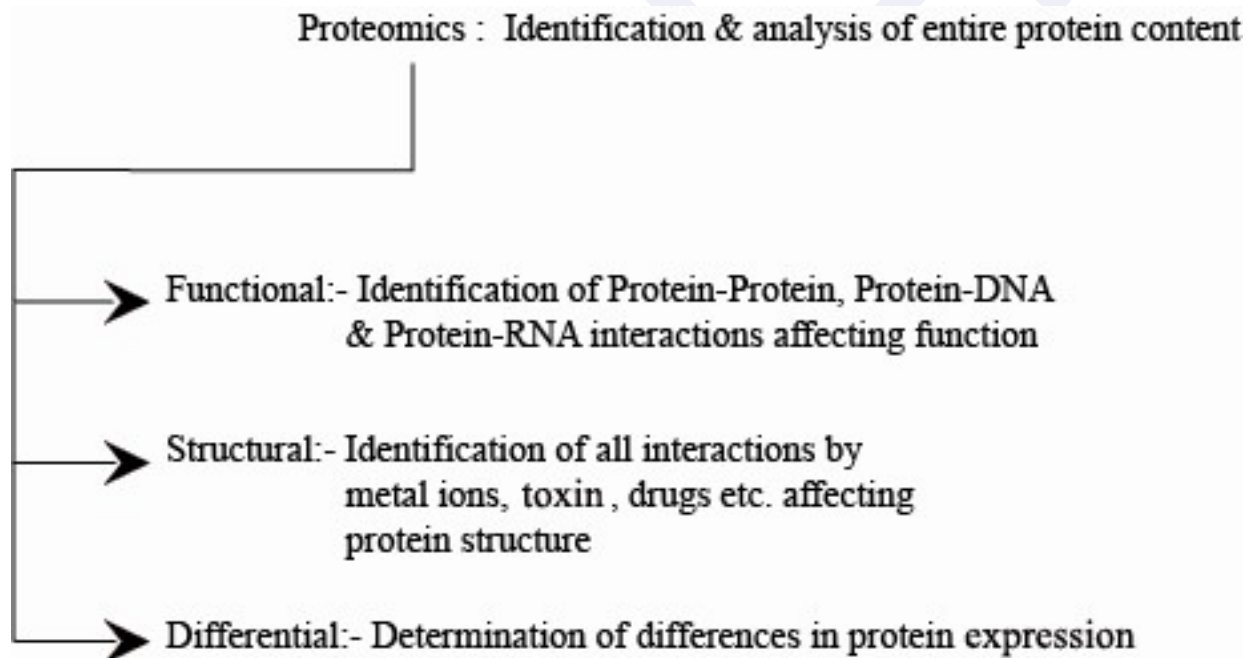


Figure 1: Types of Proteomics: Functional, structural and differential proteomics.

Techniques Involved in Proteomics Study

Some of the very basic analytical techniques are used as major proteomic tools for studying the proteome of an organism. We shall study most of these techniques as we progress in the course. The initial step in all proteomic studies is the separation of a mixture of proteins. This can be carried out using Two Dimensional Gel Electrophoresis technique in which proteins are first of all separated based on their individual charges in 1D. The gel is then turned 90 degrees from its initial position to separate proteins based on the difference in their size. This separation occurs in 2nd dimension hence the name 2D. The spots obtained in 2D electrophoresis are excised and further subjected to mass spectrometric analysis of each protein present in the mixture.

Apart from charge and size of proteins there are number of other intrinsic properties of proteins that can be employed for their separation and detection. One of these techniques is Field Flow Fractionation (FFF) which separates proteins based on their mobility in presence of applied field. The difference in mobility may be attributed to different size and mass of proteins. The applied field can be of many types such as electrical, gravitational, centrifugal etc. This technique helps in determining different components in a protein mixture, different conformations of protein, their interaction with other proteins as well as some organic molecules such as drugs (Fig. 2).

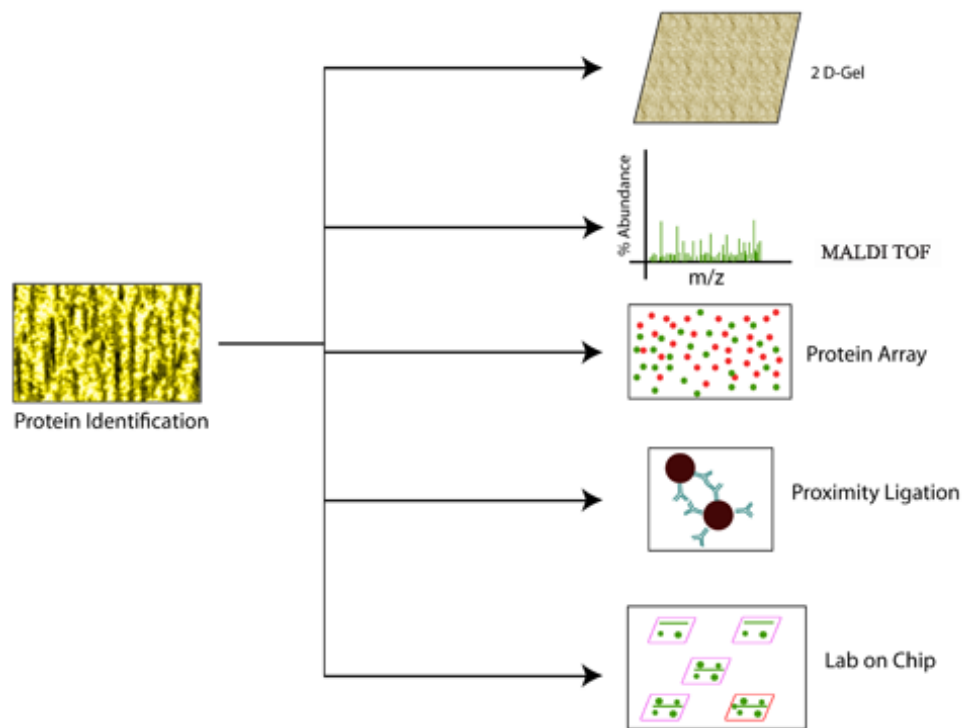


Figure 2. Techniques involved in protein identification during proteomic analysis.

Steps in Proteomic Analysis

The following steps are involved in analysis of proteome of an organism as shown in Fig. 3:

1. **Purification of proteins:** This step involves extraction of protein samples from whole cell, tissue or sub cellular organelles followed by purification using density gradient centrifugation, chromatographic techniques (exclusion, affinity etc.)
2. **Separation of proteins:** 2D gel electrophoresis is applied for separation of proteins on the basis of their isoelectric points in one dimension and molecular weight on the other. Spots are detected using fluorescent dyes or radioactive probes.
3. **Identification of proteins:** The separated protein spots on gel are excised and digested in gel by a protease (e.g. trypsin). The eluted peptides are identified using mass spectrometry.

Analysis of protein molecules is usually carried out by MALDI-TOF (Matrix Assisted Laser Desorption Ionization-Time of Flight) based peptide mass fingerprinting.

Determined amino acid sequence is finally compared with available database to validate the proteins.

Several online tools are available for proteomic analysis such as Mascot, Aldente, Popitam, Quickmod, Peptide cutter etc.

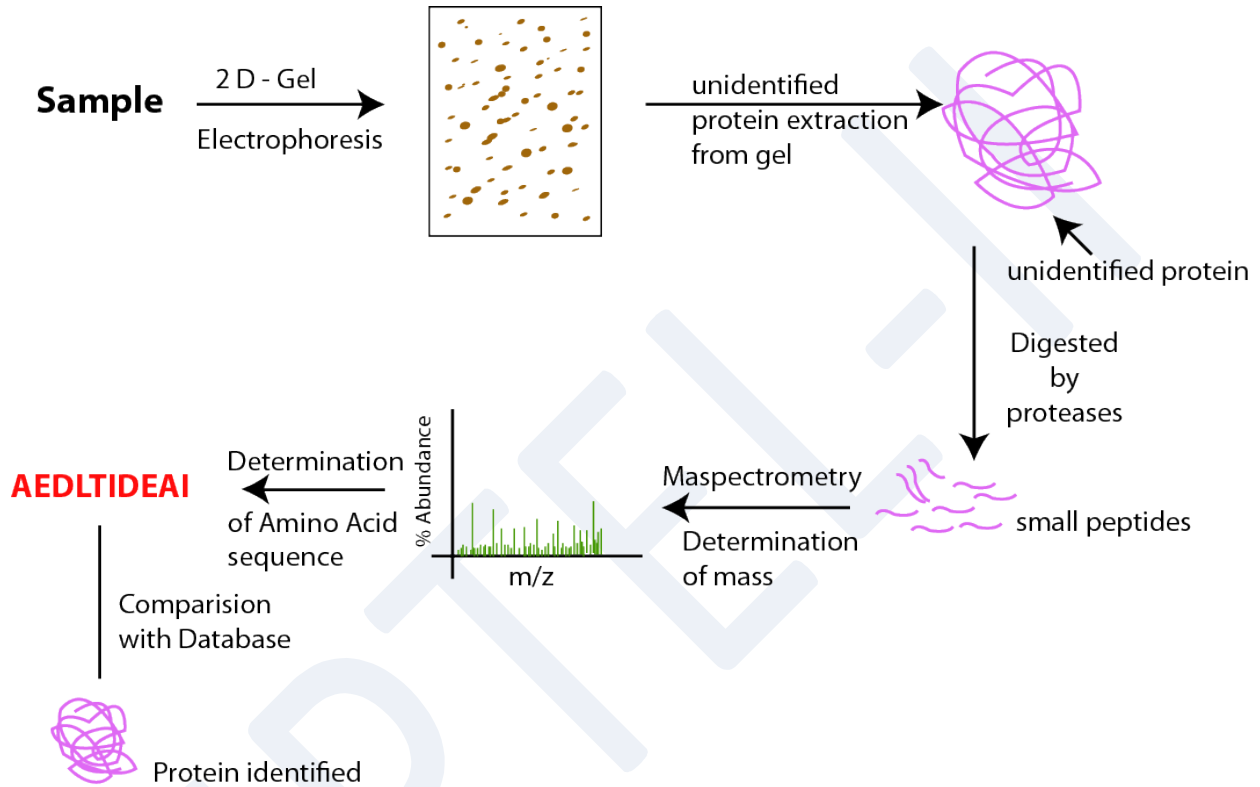


Figure 3. Overview of steps involved in proteomic analysis

Applications of Proteomics:

Proteomics has broad applications in all the aspects of life sciences including several practical applications as drug development against several diseases. Difference in expression protein expression profile of normal and diseased person may be analyzed for target protein. Protein to gene may be predicted (Fig.4). Once protein/gene is identified, function may be predicted. This can help in disease management/drug development.

From Protein To Gene

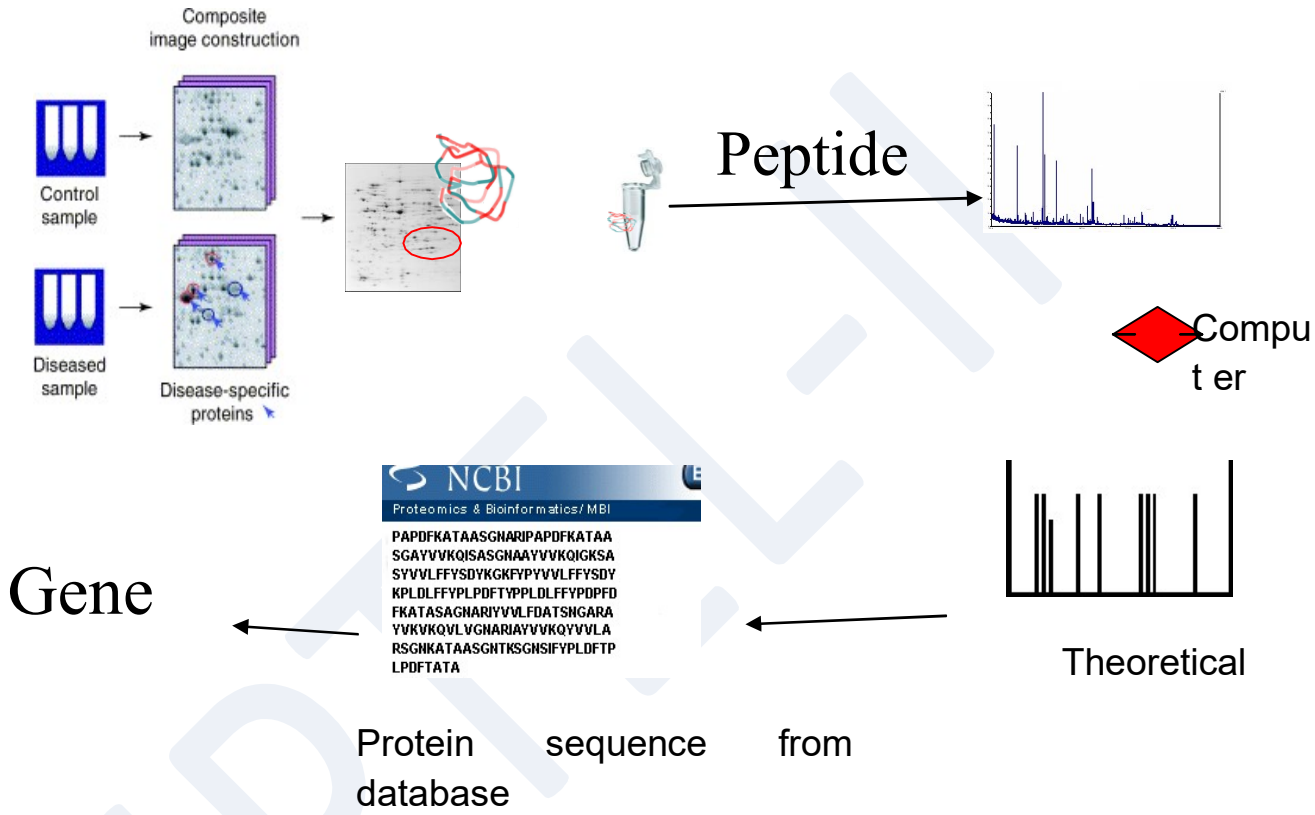


Figure 4 : Proteomic analysis-protein to gene sequence

Whole Genome Sequences of several organisms have been completed but genomic data does not show how proteins function or how these proteins are involved in biological processes. Gene codes for a protein by at several occasion proteins are modified after synthesis (several types of post-translational modifications) for functional diversification. In the current course, we shall study several applications of proteomics in coming lectures.