B.Sc Biotechnology
VIth Semester 2020
Paper name: Genomics and Proteomics
Unit III and IV Reference Notes

Dr. Parikipandla Sridevi
Assistant Professor
Department of Biotechnology
IGNTU, Amarkantak
UNIT III

Introduction to Protein structure, Chemical properties of proteins, Physical interaction that determine the property of proteins

Contents:

1. Introduction to proteins
2. Protein structure
3. Chemical properties of proteins
4. Physical interaction determine the properties of protein

Introduction to protein:

Protein is a macronutrient that is essential to building muscle mass. It is commonly found in animal products, though is also present in other sources, such as nuts and legumes. Chemically, protein is composed of amino acids, which are organic compounds made of carbon, hydrogen, nitrogen, oxygen or sulphur. Amino acids are the building blocks of proteins, and proteins are the building blocks of muscle mass.

Protein Structure:

1) Primary structure - The simplest level of protein structure, primary structure, is simply the sequence of amino acids in a polypeptide chain. For example, the hormone insulin has two polypeptide chains, A and B. The sequence of a protein is determined by the DNA of the gene that encodes the protein (or that encodes a portion of the protein, for multi-subunit proteins). A change in the gene's DNA sequence may lead to a change in the amino acid sequence of the protein. Even changing just one amino acid in a protein’s sequence can affect the protein’s overall structure and function.

2) Secondary structure - A protein’s secondary structure are structures arise from interactions between neighbouring or near-by amino acids as the polypeptide starts to fold into its functional three-dimensional form. Secondary structures arise as H bonds form between local groups of amino acids in a region of the polypeptide chain. The most common types of secondary structures are the helix and the β pleated sheet. Both structures are held in shape by hydrogen bonds, which form between the carbonyl O of one amino acid and the amino H of another.

a. α helix - The carbonyl (C=O) of one amino acid is hydrogen bonded to the amino H (N-H) of an amino acid that is four down the chain. (E.g., the carbonyl of amino acid 1 would form a hydrogen bond to the N-H of amino acid 5). This pattern of bonding pulls the polypeptide chain into a helical structure that resembles a curled ribbon, with each turn of the helix containing 3.6 amino acids. The R groups of the amino acids stick outward from the α helix, where they are free to interact.
molten globule - the initial structure of a protein that is not yet fully folded. It is a state that allows the protein to adopt different conformations before reaching its final, fully folded structure.

The protein
molecular structure of the protein

- **β pleated sheet** - Two or more segments of a polypeptide chain line up next to each other, forming a sheet-like structure held together by hydrogen bonds. The hydrogen bonds form between carbonyl and amino groups of backbone, while the R groups extend above and below the plane of the sheet. The strands of a β pleated sheet may be parallel, pointing in the same direction (meaning that their N- and C-termini match up), or antiparallel, pointing in opposite directions (meaning that the N-terminus of one strand is positioned next to the C-terminus of the other).

3) **Tertiary** - The overall three-dimensional structure of a polypeptide is called its tertiary structure. R group interactions that contribute to tertiary structure include hydrogen bonding, ionic bonding, dipole-dipole interactions, and London dispersion forces – basically, the whole gamut of non-covalent bonds. For example, R groups with like charges repel one another,
while those with opposite charges can form an ionic bond. Disulfide bonds, covalent linkages between the sulfur-containing side chains of cysteines, are much stronger than the other types of bonds that contribute to tertiary structure.

4) **Quaternary** - Some proteins are made up of multiple polypeptide chains, also known as subunits. When these subunits come together, they give the protein its quaternary structure.

### Structure of Proteins

![Structure of Proteins Diagram]

**Chemical Properties Of Proteins:**

1. **Denaturation:** It is a process by a protein lose their original secondary, tertiary or quaternary structure and converts into its simplest form i.e., primary structure due to some external stress or compounds such as strong acid or strong base, a concentrated inorganic salt, an organic solvent (e.g. alcohol or chloroform), radiation or heat. It occurs because the bonds forming secondary or tertiary structure disrupts.
   - **Heat:** Heat can be used to disrupt hydrogen bonds and non-polar hydrophobic interactions. This occurs because heat increase the kinetic energy and cause the molecules to vibrate that bonds disrupt.
   - **Alcohol:** Hydrogen bonding are all disrupted by the addition of alcohol.
   - **Acids and Bases Disrupt Salt Bridges:** Salt bridges result from the neutralization of an acid and amine on side chains. The denaturation reaction on the salt bridge by the addition of an acid results in a further straightening effect on the protein chain.
   - **Heavy Metals Salts:** Heavy metals salt act to denature proteins in much the same manner as acids and bases. They usually contain Hg$^{+2}$, Pb$^{+2}$, Ag$^{+1}$, Ti$^{+1}$, Cd$^{+2}$ etc. Since salts are ionic...
they disrupt salt bridges in proteins. The reaction of a heavy metal salt with a protein usually leads to an insoluble metal protein salt.

2. **Arginine Residue**: The arginine residue of proteins react with alpha or beta-dicarbonyl compounds to form cyclic derivatives. In the reaction with benzil, an iminoimidazolidone derivative is obtained after a benzil acid rearrangement.

3. **Hydrolysis**: Proteins are hydrolyzed by a variety of hydrolytic agents. A. By acidic agents: Proteins, upon hydrolysis with conc. HCl (6–12N) at 100–110°C for 6 to 20 hrs, yield amino acids in the form of their hydrochlorides. B. By alkaline agents: Proteins may also be hydrolyzed with 2N NaOH.

4. **Glutamic and Aspartic Acid Residues**: These amino acid residues are usually esterified with methanolic HCL. There can be side reactions, such as methanolysis of amide derivatives or N,O-acyl migration in serine or threonine residues.

**Physical interactions that determine the properties of proteins:**

1. **EXCLUDED VOLUME**: This is repulsive interaction due to mutual impenetrability between two protein molecules. Excluded volume is usually described as representing approximately four times volume of protein molecule.

2. **ELECTROSTATIC INTERACTIONS**: The electrostatic force is also known as the Coulomb force or Coulomb interaction. It’s the attractive or repulsive force between two electrically charged objects. Like charges repel each other while unlike charges attract each other. Typical charge-charge interactions that favour protein folding are those between oppositely charged R-groups such as K or R and D or E. A substantial component of the energy involved in protein folding is charge-dipole interactions. This refers to the interaction of ionized R-groups of amino acids with the dipole of the water molecule. The slight dipole moment that exists in the polar R-groups of amino acid also influences their interaction with water. It is, therefore, understandable that the majority of the amino acids found on the exterior surfaces of globular proteins contain charged or polar R-groups.

3. **Vander waals forces**: They are attractive forces and active at shorter distances compared to electrostatic forces. It strongly depends on the shape of the interacting forces. They are weak forces. Attractive van der Waals forces involve the interactions among induced dipoles that arise from fluctuations in the charge densities that occur between adjacent uncharged non-bonded atoms. Repulsive van der Waals forces involve the interactions that occur when uncharged non-bonded atoms come very close together but do not induce dipoles. The repulsion is the result of the electron-electron repulsion that occurs as two clouds of electrons begin to overlap.

4. **Hydrogen bonding**: A hydrogen bond is a type of attractive (dipole-dipole) interaction between an electronegative atom and a hydrogen atom bonded to another electronegative atom. This bond always involves a hydrogen atom. Hydrogen bonds can occur between molecules or within parts of a single molecule.
5. **Hydrophobic Forces:** Proteins are composed of amino acids that contain either hydrophilic or hydrophobic R-groups. It is the nature of the interaction of the different R-groups with the aqueous environment that plays the major role in shaping protein structure. The spontaneous folded state of globular proteins is a reflection of a balance between the opposing energetics of H-bonding between hydrophilic R-groups and the aqueous environment and the repulsion from the aqueous environment by the hydrophobic R-groups. The hydrophobicity of certain amino acid R-groups tends to drive them away from the exterior of proteins and into the interior. This driving force restricts the available conformations into which a protein may fold.

**References:**


https://www.slideshare.net/MohamedHassanien/properties-of-proteins

https://brainly.in/question/3646122

Short range interactions, electrostatic forces, Van der waal interactions, hydrogen bonds, Hydrophobic interactions.

Contents

1. Protein-protein interactions
2. Electrostatic forces
3. Van der waal interactions
4. Hydrogen bonds
5. Hydrophobic interactions
6. Conclusions

SHORT-RANGE INTERACTION

Protein-Protein Interaction

Introduction:

Protein-protein interactions (PPIs) handle a wide range of biological processes, including cell-to-cell interactions and metabolic and developmental control. Protein-protein interaction is becoming one of the major objectives of system biology. Non-covalent contacts between the residue side chains are the basis for protein folding, protein assembly, and PPI. These contacts induce a variety of interactions and associations among the proteins. Based on their contrasting structural and functional characteristics, PPIs can be classified in several ways. On the basis of their interaction surface, they may be homo- or hetero oligomeric; as judged by their stability, they may be obligate or non-obligate; as measured by their persistence, they may be transient or permanent. A given PPI may be a combination of these three specific pairs. The transient interactions would form signalling pathways while permanent interactions will form a stable protein complex.

PPIs can

(i) modify the kinetic properties of enzymes;
(ii) act as a general mechanism to allow for substrate channelling;
(iii) construct a new binding site for small effector molecules;
(iv) inactivate or suppress a protein;
(v) change the specificity of a protein for its substrate through interaction with different binding partners;
(vi) Serve a regulatory role in either upstream or downstream level.
Uncovering protein-protein interaction information helps in the identification of drug targets. Studies have shown that proteins with larger number of interactions (hubs) can include families of enzymes, transcription factors, and intrinsically disordered proteins, among others. However, PPIs involve more heterogeneous processes and the scope of their regulation is large. For more accurate understanding of their importance in the cell, one has to identify various interactions and determine the aftermath of the interactions. In recent years, PPI data have been enhanced by guaranteed high-throughput experimental methods, such as two-hybrid systems, mass spectrometry, phage display, and protein chip technology.

Comprehensive PPI networks have been built from these experimental resources. However, the voluminous nature of PPI data is imposing a challenge to laboratory validation. Computational analysis of PPI networks is increasingly becoming a mandatory tool to understand the functions of unexplored proteins. At present, protein-protein interaction (PPI) is one of the key topics for development and progress of modern systems biology.

**Classification of PPI Detection Methods**

Protein-protein interaction detection methods are categorically classified into three types, namely, *in vitro*, *in vivo*, and *in silico* methods. In *in vitro* techniques, a given procedure is performed in a controlled environment outside a living organism. The *in vitro* methods in PPI detection are tandem affinity purification, affinity chromatography, coimmunoprecipitation, protein arrays, protein fragment complementation, phage display, X-ray crystallography, and NMR spectroscopy. In *in vivo* techniques, a given procedure is performed on the whole living organism itself. The *in vivo* methods in PPI detection are yeast two-hybrid (Y2H, Y3H) and synthetic lethality. *In silico* techniques are performed on a computer (or) via computer simulation. The *in silico* methods in PPI detection are sequence-based approaches, structure-based approaches, chromosome proximity, gene fusion, *in silico* 2 hybrid, mirror tree, phylogenetic tree, gene expression approaches.

**SUMMARY OF PPI DETECTION METHODS**

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coimmunoprecipitation (Co-IP)</td>
<td>An immunoprecipitation experiment designed to affinity-purify a bait protein antigen together with its binding partner using a specific antibody against the bait.</td>
</tr>
<tr>
<td>Pull-down Assays</td>
<td>An affinity chromatography method that involves using a tagged or labeled bait to create a specific affinity matrix that will enable binding and purification of a prey protein from a lysate sample or other protein-containing mixture.</td>
</tr>
<tr>
<td><strong>Yeast two-hybrid (Y2H)</strong></td>
<td>Monitor complex formation through transcriptional activation of reporter genes.</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Far-Western Blotting</strong></td>
<td>Similar strategy to Western blotting with one key difference. The antibody probe in a typical Western blotting detection is substituted with an labeled bait protein as the probe.</td>
</tr>
<tr>
<td><strong>Tandem affinity purification-mass spectroscopy (TAP-MS)</strong></td>
<td>TAP-MS is based on the double tagging of the protein of interest on its chromosomal locus, followed by a two-step purification process and mass spectroscopic analysis.</td>
</tr>
<tr>
<td><strong>Protein microarrays</strong></td>
<td>Microarray-based analysis allows the simultaneous analysis of thousands of parameters within a single experiment.</td>
</tr>
<tr>
<td><strong>Bio-Layer Interferometry (BLI)</strong></td>
<td>Change in the number of molecules bound to the biosensor tip causes a shift in the interference pattern that can be measured in real-time.</td>
</tr>
<tr>
<td><strong>Surface Plasmon Resonance (SPR)</strong></td>
<td>SPR angle changes with surface refractive indexes, which is in direct proportion to the molecular mass of the biomolecule attached to the metal surface.</td>
</tr>
</tbody>
</table>

**Computational Analysis of PPI Networks**

A PPI network can be described as a heterogeneous network of proteins joined by interactions as edges. The computational analysis of PPI networks begins with the illustration of the PPI network arrangement. The simplest sketch takes the form of a mathematical graph consisting of nodes and edges [68]. Protein is represented as a node in such a graph and the proteins that interact with it physically are represented as adjacent nodes connected by an edge. An examination of the network can yield a variety of results. For example, neighbouring proteins in the graph probably may share more the same functionality. In addition to the functionality, densely connected subgraphs in the network are likely to form protein complexes as a unit in certain biological processes. Thus, the functionality of a protein can be inferred by spotting at the proteins with which it interacts and the protein complexes to which it resides. The topological prediction of new interactions is a novel and useful option based exclusively on the structural
information provided by the PPI network (PPIN) topology [69]. Some algorithms like random layout algorithm, circular layout algorithm, hierarchical layout algorithm, and so forth are used to visualize the network for further analysis. Precisely, the computational analysis of PPI networks is challenging, with these major barriers being commonly confronted.

(1) the protein interactions are not stable;

(2) one protein may have different roles to perform

(3) two proteins with distinct functions periodically interact with each other.

**Role of PPI Networks in Proteomics**

Predicting the protein functionality is one of the main objectives of the PPI network. Despite the recent comprehensive studies on yeast, there are still a number of functionally unclassified proteins in the yeast database which reflects the impending need to classify the proteins. The functional annotation of human proteins can provide a strong foundation for the complete understanding of cell mechanisms, information that is valuable for drug discovery and development [4]. The increased availability of PPI networks has developed various computational methods to predict protein functions. The availability of reliable information on protein interactions and their presence in physiological and pathophysiological processes are critical for the development of protein-protein-interaction-based therapeutics. The compendium of all known protein-protein interactions (PPIs) for a given cell or organism is called the interactome.

Protein-protein complex identification is the crucial step in finding the signal transduction pathways. Protein-protein complexes mostly consist of antibody-antigen and protease-inhibitor complexes. Crystallography is the major tool for determining protein complexes at atomic resolution. The complete analysis of PPIs can enable better understanding of cellular organization, processes, and functions. The other applications of PPI Network include biological indispensability analysis, assessing the drug ability of molecular targets from network topology, estimation of interactions reliability, identification of domain-domain interactions, prediction of protein interactions, detection of proteins involved in disease pathways, delineation of frequent interaction network motifs, comparison between model organisms and humans, and protein complex identification.

Conclusions

While available methods are unable to predict interactions with 100% accuracy, computational methods will scale down the set of potential interactions to a subset of most likely interactions. These interactions will serve as a starting point for further lab experiments. The gene expression data and protein interaction data will improve the confidence of protein-protein interactions and the corresponding PPI network when used collectively. Recent developments have also led to the construction of networks having all the protein-protein interactions using computational methods for signal transduction pathways and protein. The complete analysis of PPIs can enable better understanding of cellular complex identification.
Electrostatic Force in Protein-Protein Recognition

INTRODUCTION-

Electrostatic interactions are important for understanding molecular interactions, since they are long-range interactions and can guide binding partners to their correct binding positions. To investigate the role of electrostatic forces in molecular recognition, we calculated electrostatic forces between binding partners separated at various distances. The investigation was done on a large set of 275 protein complexes using recently developed DelPhiForce tool and in parallel, evaluating the total electrostatic force via electrostatic association energy. To accomplish the goal, we developed a method to find an appropriate direction to move one chain of protein complex away from its bound position and then calculate the corresponding electrostatic force as a function of separation distance. It is demonstrated that at large distances between the partners, the electrostatic force (magnitude and direction) is consistent among the protocols used and the main factors contributing to it are the net charge of the partners and their interfaces. However, at short distances, where partners form specific pair-wise interactions or de-solvation penalty becomes significant, the outcome depends on the precise balance of these factors. Based on the electrostatic force profile (force as a function of distance), we group the cases into four distinctive categories, among which the most intriguing is the case termed “soft landing.” In this case, the electrostatic force at large distances is favorable assisting the partners to come together, while at short distance it opposes binding, and thus slows down the approach of the partners toward their physical binding.

Types of Non-Covalent Interactions

Electrostatic Interaction:
The force between oppositely charged ionic groups on the two protein side chains is known as electrostatic forces (Fig. 5.2a). The force of attraction (F) is inversely proportional to the square of the distance (d) between the charges. For e.g., ionized amino groups (NH₄⁺) on lysine of one protein and as ionized carboxyl group (—COO⁻) of glutamate on the other.
Attraction (F) $\propto \frac{1}{Kd_2}$ [K, dielectric constant] Therefore, the closer the charges would increase attractive forces. This force can be generated by charge transfer reaction in between the antigen and antibody.

**Hydrogen Bonding:**

Another bonding type is hydrogen bond which is reversible bond between hydrophilic groups e.g., NH$_2$, OH and COOH depend on the proximity of two molecules having these groups (Fig. 5.2b). It is a weak bond which involves the release of water between the reacting side chains.

**Vander-Waals Forces:**

Vander-Waals force is greatly influenced by presence of external electron cloud. Ideal gases like hydrogen and nitrogen exhibit this bond between them. The force of attraction (F) is inversely related to the seventh power of distance (d) between two molecules (Fig).

Force (F) $\propto \frac{1}{d^7}$
**Hydrophobic Interaction:**
Hydrophobic, non-polar group’s; like leucine, valine and phenyl-alanine side chain have a tendency to associate in an aqueous environment. When water comes in contact with the hydrophobic molecules, then the hydrophobic bond generates. Increased area of contact between water and hydrophobic molecules decreases entropy and higher the energy state.

![Hydrophobic interaction](image)

**VAN DER WAALS INTERACTIONS INVOLVING PROTEINS**

**INTRODUCTION**
Van der Waals (dispersion) forces contribute to interactions of proteins with other molecules or with surfaces, but because of the structural complexity of protein molecules, the magnitude of these effects is usually estimated based on idealized models of the molecular geometry, e.g., spheres or spheroids. The calculations reported here seek to account for both the geometric irregularity of protein molecules and the material properties of the interacting media. Whereas the latter are found to fall in the generally accepted range, the molecular shape is shown to cause the magnitudes of the interactions to differ significantly from those calculated using idealized models, with important consequences. First, the roughness of the molecular surface leads to much lower average interaction energies for both protein-protein and protein-surface cases relative to calculations in which the protein molecule is approximated as a sphere. These results indicate that a form of steric stabilization may be an important effect in protein solutions. Underlying this behavior is appreciable orientational dependence, one reflection of which is that molecules of complementary shape are found to exhibit very strong attractive dispersion interactions. Although this has been widely discussed previously in the context of molecular recognition processes, the broader implications of these phenomena may also be important at larger molecular separations, e.g., in the dynamics of aggregation, precipitation, and crystal growth.

**Energy in Van der Waals interaction**
Energies associated with van der Waals interactions are quite small. Usually, they are about 2 to 4 kJ/mol per atom pair. When the surfaces of two large molecules come together, a large number of atoms are in Van der Waals contact, and the net effect, summed over many atoms pairs can be substantial.
Macromolecules such as proteins and DNA contain numerous sites of potential van der Waals interactions that the cumulative effect of these small binding forces can be enormous; hence the most stable structure for macromolecules is that where weak interactions are maximized.

**Van der Waals interactions in protein structure**

In addition to hydrogen-bonds and disulfide bonds, protein structure can also be stabilized by Val der waals interactions. In the coiled-coil protein, there are Heptad repeat which form by the side chain interaction between each alpha helix; hepad-repeat is repeated in every 7th residues. If these repeating residues are hydrophobic, such as leucine, van der waals interaction will be formed to stabilize this protein structure.

![Van der Waals interactions](image)

**HYDROGEN BONDS IN PROTEINS**

**INTRODUCTION**

Hydrogen bonds provide most of the directional interactions that underpin protein folding, protein structure and molecular recognition. The core of most protein structures is composed of secondary structures such as α helix and β sheet. This satisfies the hydrogen-bonding potential between main chain carbonyl oxygen and amide nitrogen buried in the hydrophobic core of the protein. Hydrogen bonding between a protein and its ligands (protein, nucleic acid, substrate, effector or inhibitor) provides a directionality and specificity of interaction that is a fundamental aspect of molecular recognition. The energetics and kinetics of hydrogen bonding therefore need to be optimal to allow the rapid sampling and kinetics of folding, conferring stability to the protein structure and providing the specificity required for selective macromolecular interactions.

Some concepts of hydrogen bonds:

- A hydrogen bond is formed by the interaction of a hydrogen atom that is covalently bonded to an electronegative atom (donor) with another electronegative atom (acceptor). Hydrogen bonding confers rigidity to the protein structure and specificity to intermolecular interactions. The accepted (and most frequently observed) geometry for a hydrogen bond is a distance of less than 2.5 Å (1.9 Å) between hydrogen and the acceptor and a donor-hydrogen-acceptor angle of between 90° and 180° (160°).

- During protein folding, the burial of hydrophobic side-chains requires intra-molecular hydrogen bonds to be formed between the main chain polar groups.
• The most stable conformations of polypeptide chains that maximize intrachain hydrogen-bonding potential are α helices and β sheets.

• Specificity in molecular recognition is driven by the interaction of complementary hydrogen-bonding groups on interacting surfaces.

**HYDROPHOBIC INTERACTIONS OF PROTEINS**

**INTRODUCTION**

Proteins fold spontaneously into complicated three-dimensional structures that are essential for biological activity. Much of the driving energy for this folding process comes from the hydrophobic effect, i.e. the removal of nonpolar amino acids from solvent and their burial in the core of the protein. The hydrophobic Interactions are at present attracting the greatest attention of all the various interactions existing between the groups of these macromolecules. “Hydrophobic interaction” is usually understood as the cause of low solubility of nonpolar substances in water, of their hydrophobicity, which results in an enhanced association of these solutes in aqueous solutions. As there are many nonpolar groups in proteins and many of them are clustered together, as if avoiding contact with water, one can suppose that the hydrophobicity of these groups plays an important role in determining the compact state of globular protein just as it does in the case of an oil drop in water.

**HYDROPHIC EFFECT IN PROTEIN FOLDING**

The hydrophobic effect is a major driving force in protein folding. A complete understanding of this effect requires the description of the conformational states of water and protein molecules at different temperatures. Towards this goal, we characterize the cold and hot denatured states of a protein by modelling NMR chemical shifts using restrained molecular dynamics simulations. A detailed analysis of the resulting structures reveals that water molecules in the bulk and at the protein interface form on average the same number of hydrogen bonds. Thus, even if proteins are ‘large’ particles (in terms of the hydrophobic effect, i.e. larger than 1 nm), because of the presence of complex surface patterns of polar and non-polar residues
their behaviour can be compared to that of ‘small’ particles (i.e. smaller than 1 nm). We thus find that the hot denatured state is more compact and richer in secondary structure than the cold denatured state, since water at lower temperatures can form more hydrogen bonds than at high temperatures. Then, using Φ-value analysis we show that the structural differences between the hot and cold denatured states result in two alternative folding mechanisms.

**Hydrophobic interactions**

![Diagram of hydrophobic interactions](Image)

**Conclusions:**
Protein-protein interactions are the physical contacts of high specificity established between two or more proteins molecules as a result of biochemical events steered by interactions that include electrostatic forces, hydrogen bonding and the hydrophobic effect. Protein are vital macromolecules that facilitates diverse biological processes at both cellular processes are regulated via a large number of protein components organised by protein-protein interaction, which refers to intentional physical contacts established between two or more proteins and resulted in specific biochemical events. Such interactions undertake at the core of the entire interatomic system of the living cells, unsurprisingly, specific PPLs are identified with the correlation of multiple diseases.

**References:**
1. www.hindawai.com
2. www.wikipedia.com
4. www.biologicscorp.com
5. www.sciencedirect.com
INTRODUCTION TO PROTEOMICS.

ANALYSIS OF PROTEOMES

CONTENTS

Proteomes
Types of proteomics
Applications of proteomics
De novo peptide sequencing
Peptide sequencing by Edman degradation
Protein identification by mass spectroscopy
Analysis of proteins

References

PROTEOMICS

Proteomics is the large-scale study of proteins, particularly their function and structure. Proteomics is an excellent approach for studying changes in metabolism in response to different stress conditions.

TYPES OF PROTEOMICS

Based on the protein response under stress conditions proteomics are classified into different groups.

1. Expression proteomics
2. Structural proteomics
3. Functional proteomics
4. Techniques involved proteomics

APPLICATIONS OF PROTEOMICS

ONCOLOGY
To characterize the protein expressions, functions of tumor cells and widely used in biomarker discovery.

**BIO-MEDICAL APPLICATIONS**

- It deals with the fundamentals of the infections origin and their effect on organs.

**AGRICULTURAL APPLICATION**

- Population growth and effect of global climate changes imposing severe limits on the sustainability of agricultural crop production
- Used to know plant-insects interaction

**De Novo Peptide sequencing**

De novo peptide sequencing is the method in which a peptide amino acid sequence is determined from tandem mass spectrometry. In the old days, this was accomplished by the Edman degradation procedure. Analysis by a tandem mass spectrometer is a more common method to solve the sequencing of peptides. De novo sequencing is an assignment of fragment ions from a mass spectrum. Different algorithms are used for interpretation and most instruments come with de novo sequencing programs.

**PEPTIDE SEQUENCING BY EDMAN DEGRADATION**

The method employs a series of chemical reactions to remove and identify the amino acid residue that is at the N-terminus of the polypeptide chain. That is the residue with a free alpha amino group. At the same time, the next residue in the sequence is made available and subjected to the same round of chemical reactions. Reiteration of this process reveals the sequence of the polypeptide.

**PROTEIN IDENTIFICATION BY MASS SPECTROSCOPY**

Mass spectroscopy has become the method of choice for sensitive, reliable and inexpensive protein and peptide identification. With the increasing number of full genome sequences for a variety of organisms and the numerous protein databases constructed there of all the tools necessary for the high through put protein identification with mass spectroscopy are in place. Although the mass spectrometric interpretation algorithms currently in use can clearly produce good results, nearly all MS protein information is based on the characterization of short peptide sequences.
ANALYSIS OF PROTEOMICS

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.

2. **Separation of proteins:** 2D gel electrophoresis is applied for separation of proteins on the basis of their isoelectric points in one dimension.

3. **Identification of proteins:** The separated protein spots on gel are excised and digested in gel by a protease.

REFERENCES

- Wilm M, Mann M: Analytical properties of the nanoelectrospray ion source.

**Determination of sizes (Sedimentation analysis, gel filtration, SDS-PAGE); Native PAGE, Determination of covalent structures – Edman degradation.**

Introduction

An important part of characterizing any protein molecule is to determine its size and shape. Sedimentation and gel filtration are hydrodynamic techniques that can be used for this medium resolution structural analysis. The molecular weight can be determined by combining gradient sedimentation and gel filtration, techniques proposed by Seigel and Monte. Finally, rotary shadowing and negative stain electron microscopy are powerful techniques for resolving the size and shape of single protein molecules and complexes at the nanometre level. A combination of hydrodynamics and electron microscopy is especially powerful.

Most proteins fold into globular domains. Protein folding is driven by the hydrophobic effect, which seeks to minimize contact of polypeptide with solvent. Most proteins fold into globular domains, which have minimal surface area. Peptides from 10-30 kDa typically fold into a single domain. Peptides larger than 50 kDa typically form two or more domains that are independently folded. However, some proteins are highly elongated, either as a string of small globular domains or stabilized by specialized structures such as coiled coils or the collagen triple-helix. The ultimate structural understanding of a protein comes from an atomic level structure obtained
by X-ray crystallography or nuclear magnetic resonance. However, structural information at the nanometre level is frequently invaluable. Hydrodynamics, in particular sedimentation and gel filtration, can provide the structural information, and it becomes even more powerful when combined with electron microscopy (EM).

The interior protein subunits and domains consist of closely packed atom. There are no substantial holes and almost no water molecules in the protein interior. As a consequence, proteins are rigid structures with a Young’s modulus similar to that of Plexiglas. Engineers sometimes categorize biology as a science of “soft weight materials”. This is true of some hydrated gels, but proteins are better thought of as hard dry plastic. This is obviously important for all of biology, to have a rigid material with which to construct the machinery of life. A second consequence of the close packed interior of proteins is that all protein have approximately the same density, about 1.37 g/cm$^3$.

**Sedimentation Analysis**

Sedimentation analysis of macromolecules carried out in analytical ultracentrifuge is a powerful method for the study of proteins, nucleic acids and other polymers and their various complexes. Monitoring sedimentation of macromolecules in the centrifugal field allows their hydrodynamic and thermodynamic characterization in solution, i.e. in native conditions, without interactions with any matrix or surface. This allows direct measurement of molecular weight and sedimentation coefficient of macromolecules, monitoring of sample purity and homogeneity, prediction of size and shape of sedimenting species and, last but not least, study of equilibrium reactions, including determination of their stoichiometry and equilibrium constants.

In the present lecture, we will first focus on the history physical principles of the technique and also on the properties and potential of a modern instrumentation. Two types of experiments performed using analytical ultracentrifuge (i.e. sedimentation velocity and sedimentation equilibrium) will be discussed, together with a brief introduction into sedimentation theory. In the end, sedimentation data analysis will be described and examples of utilization of analytical ultracentrifugation in biomolecule research will be provided.

Combination of new instrumentation and computational software for data analysis has led to major advances in characterization of proteins and their complexes. After temporary silence in the past decades, analytical ultracentrifugation at presence experiences renaissance in proteomic and structural biology research while still being heavily used e.g. for characterization of aggregation of monoclonal antibodies in biopharmaceutical industry.

**Gel Filtration**

Introduction
Gel filtration separates molecules according to differences in size as they pass through a gel filtration medium packed in a column. Unlike ion exchange or affinity chromatography, molecules do not bind to the chromatography medium so buffer composition does not directly affect resolution (the degree of separation between peaks). Consequently, a significant advantage of gel filtration is that conditions can be varied to suit the type of sample or the requirements for further purification, analysis or storage without altering the separation.

Gel filtration is well suited for biomolecules that may be sensitive to changes in pH, concentration of metal ions or co-factors and harsh environment conditions. Separations can be performed in the presence of essential ions or cofactors, detergents, urea, guanidine hydrochloride, at high or low ionic strength, at 37°C or in the cold room according to the requirements of the experiment. Purified proteins can be collected in any chosen buffer.

**Separation by gel filtration**

To perform a separation, gel filtration medium is packed into a column to form a *packed bed*. The medium is a porous matrix in the form of spherical particles that have been chosen for their chemical and physical stability, and inertness (lack of reactivity and absorptive properties). The packed bed is equilibrated with buffer which fills the pores is sometimes referred to as the stationary phase and this liquid is in equilibrium with the liquid outside the particles, referred to as the mobile phase. It should be noted that samples are eluted isocratically, i.e. there is no need to use different buffers during the separation. However, a wash step using the running buffer is usually included at the end of a separation. However, a wash step using the running buffer is usually included at the end of a separation to facilitate the removal of any molecules that may have been retained on the column and to prepare the column for a new run. The following figure shows the most common terms used to describe the separation and illustrates the separation process of gel filtration.
SDS-PAGE

SDS PAGE (SDS Polyacrylamide Gel Electrophoresis)

Two fundamentally different types of gel system exist, non-dissociating (non-denaturing) and dissociating (denaturing). Non-dissociating (non-denaturing) system is designed to separate native protein under conditions that preserve protein function and activity. In contrast, a dissociating system is designed to denature protein into their constituent’s polypeptides and hence examines the polypeptide composition of samples. The denaturing system is probably the most common used and is known as SDS-PAGE.

The SDS used in an SDS-PAGE protein identification procedure stands for Sodium Dodecyl Sulphate, an anionic detergent that denatures the proteins and unfolds each polypeptide chain into a linear orientation. The SDS also applies an evenly distributed negative charge to each protein in accordance with its mass.

In SDS-PAGE, the protein mixture is denatured by heating at 100°C in the presence of excess SDS and a reducing reagent (β-mercaptoethanol, TCEP, DTT) is employed to break disulphide bonds. Under these conditions, all reduced polypeptide bind the same amount of SDS on a weight basis (1.4g SDS/g polypeptide) independent of the amino acid composition and sequence of the protein. The SDS-protein complex forms a rod with its length proportional to the molecular weight of the protein. All proteins are now negatively charged with similar charge density and thus can be separated on the basis of their size only.

Native PAGE

Introduction

The technique file describes an optimized method for native polyacrylamide gel electrophoresis (native PAGE) with PhastGel™ gradient 8-25 and PhastGel gradient 10-15 using PhastGel native buffer strips. The method has been optimized using crude protein extracts and commercially available proteins. Therefore, it is generally applicable and offers a good starting point for developing method for specific applications.

Native PAGE is one of the most powerful techniques for studying the composition and structure of native proteins, since both the conformation and biological activity of proteins remain intact during this technique. It is occasionally used for molecular weight (MW) measurements; however SDS-PAGE is easier and in most cases more reliable than native PAGE for this application. Often, it is difficult to find standard proteins that resembled the shape, partial specific volume and degree of hydration as the native protein under investigation.

Gradient gels for native PAGE sharpen the protein bands and allow complex mixtures of proteins to be separated on a single gel. With PhastSystem and PhastGel gradient media, native PAGE is
fast reproducible, and convenient; separation take approximately 60 minutes, are run under exact, programmed conditions, and require no buffer preparation.

**Determination of covalent structures – Edman degradation**

**Introduction**

The strategy of Sanger and colleagues for the sequencing of insulin was to characterize series of small overlapping peptides produced by cleavage of the parent molecule. Determination of the overall amino acid content and the identity of the amino- (N-) terminal residue for each peptide allowed deduction of the sequence of the whole molecule (Sanger, 1959). An alternative approach was that described by Pehr Edman (1950). This allowed determination of extended sequences of peptides or whole proteins. The method employs a series of chemical reactions to remove and identify the amino acid residue that is at the N-terminus of the polypeptide chain, i.e. the residue with a free α-amino group. At the same time, the next residue in the sequence is made available and subjected to the same round of chemical reactions. Reiteration of this process reveals the sequence of the polypeptide. It was later partly automated in an instrument termed a ‘sequenator’.

The Basis of the Method: Peptide sequencing by Edman chemistry may be divided into steps as follows:

**Coupling**

Phenyl isothiocyanate (PITC) reacts with an α-amino group (or in the case of prolyl residue with an imino group) at the N-terminal end of the polypeptide chain, to form a phenylthiocarbamyl derivative of the terminal residue. Basic conditions are required for this reaction. Clearly, a free α-amino group is required for this reaction to occur. Again, a peptide may not have an N-terminus at all: cyclosporin is just one example of a natural peptide that is cyclic.

**Cleavage**

In the presence of strong acid, cleavage occurs at the first peptide bond, giving the peptide (minus the first residue) and the liberated first residue as the anilinothiazolinone (ATZ) form. Once other reactants and products have been washed away, the shortened polypeptide can be taken through another round of coupling and cleavage to release the second residue, and so on in a cyclical fashion. Currently, trifluoroacetic acid (TFA) is used for this cleavage reaction.
Conversion

The ATZ residue is separated from the peptide by extraction in organic solvent (ethyl acetate or chlorobutane), and is then converted to a more stable form to allow better analysis. Conversion to the more stable phenylthiohydantoin (PTH) form is done in aqueous acid (25% TFA, v/v in water).

Analysis of PTH residues

The PTH residue generated by each cycle of Edman chemistry is typically identified by chromatography, originally thin-layer chromatography and latterly reversed-phase high-performance liquid chromatography.

The efficiency of sequencing

The different amino acid residues, being structurally different, react at each stage with different degrees of efficiency. The overall efficiency (‘repetitive yield’) is less than 100 % (usually of the order of 95%), so over the course of a number of cycles the yield of sequence declines, and the degree of stagger, or ‘lag’, gradually increases. At the same time, the amount of background noise increases. The number of cycles at which this occurs may be small or large (50 or more), depending to a great extent on the size and amino acid content of the polypeptide itself, since this
dictates the degree of random acid hydrolysis and other side reactions. The time for one cycle of Edman chemistry has been reduced to 20 minutes. The amount of sample required currently to provide a sequence of a few to 20 or so residues is of the order of 1 Pico mole or less.

The Role of Peptide Sequencing with Edman Chemistry

Originally the aim was to determine the sequence of a protein, wholly or in part. The advent of molecular biology provided an alternative and quicker way to do this, via the sequencing of the respective gene. Partial protein sequence was required for design of the oligonucleotide probes used in the process of gene cloning and for confirmation that isolated clones were indeed the relevant ones. Currently, however, knowledge of the sequence of a gene does not give us complete information on matters that are significant in the function of a protein, such as inter- and intramolecular disulphide bonding patterns or modification and processing events.

Edman chemistry is now a standard method for peptide sequencing. In recent years it has been complemented by a variety of mass spectrometric methods that have been, and continue to be, refined such that they can determine masses of proteins and of peptides derived from them, and can determine peptide sequence from the patterns of fragmentation from peptide into individual amino acid residues.
2D-PAGE, Sample preparation, Solubilisation, Reduction, Resolution, Reproducibility of 2-D PAGE

Contents
1. Introduction
2. Principle
3. Overview
4. Steps in 2D gel electrophoresis
   a. Sample preparation
   b. Isoelectric focusing
   c. SDS-PAGE
   d. Protein detection and image analysis
   e. Protein Identification by Mass Profile Fingerprinting

Introduction
Two-dimensional (2D) gel electrophoresis is a high-resolution technique for the study of proteome. The proteome is the complete set of proteins encoded by a genome, and proteomic analysis consists in profiling the whole proteins expressed in a given cell, tissue, organ, or organism. Proteomic expression has the main purpose of qualitatively and quantitatively comparing proteins expressed under physiological and/or pathological conditions. Although it is not the unique approach used in modern proteomics, two-dimensional electrophoresis (2DE) is
unrivalled allowing simultaneous separation of thousands of proteins and the detection of post-translational modification, not predictable through genome analysis. 2DE combines two physical principles to separate complex protein mixtures: the isoelectric point and the molecular weight. The result is a gel map in which each protein isoform present in the sample can be visualized as a spot, analysed, quantified, and identified by mass spectrometry analysis.

Two dimensional protein electrophoresis or 2DE is an established technique commonly used in purifying and analysing individual proteins from complex biological samples and is currently regarded as the best option for profiling low abundance proteins. By applying this particular technique, the individual proteins are separated from the sample based on their isoelectric points and molecular weights – a process that seem simple enough in theory but is actually more complex in practice.

**Principle**

- In 2D GE proteins are separated as per isoelectric point and protein mass.
- Separation of the proteins by isoelectric point is called isoelectric focusing (IEF). When a gradient of pH is applied to a gel and an electric potential is applied across the gel, making one end more positive than the other.
- At all pH values other than their isoelectric point, proteins will be charged. If they are positively charged, they will be pulled towards the negative end of the gel and if they are negatively charged they will be pulled to the positive end of the gel. The proteins applied in the first dimension will move along the gel and will accumulate at their isoelectric point; that is, the point at which the overall charge on the protein is 0 (a neutral charge).
- In separating the proteins by mass, the gel treated with sodium dodecyl sulphate (SDS) along with other reagents (SDS-PAGE in 1-D). This denatures the proteins (that is, it unfolds them into long, straight molecules) and binds a number of SDS molecules roughly proportional to the protein’s length. Because a protein’s length (when unfolded) is roughly proportional to its mass, since the SDS molecules are negatively charged, the result of this is that all of the proteins will have approximately the same mass-to-charge ratio as each other.
- In addition, proteins will not migrate when they have no charge (a result of the isoelectric focusing step) therefore the coating of the protein in SDS (negatively charged) allows migration of the proteins in the second dimension.
- In the second dimension, an electric potential is again applied, but at a 90 degree angle from the first field. The proteins will be attracted to the more positive side of the gel (because SDS is negatively charged) proportionally to their mass-to-charge ratio.
- The gel therefore acts like a molecular sieve when the current is applied, separating the proteins on the basis of their molecular weight with larger proteins being retained higher in the gel and smaller proteins being able to pass through the sieve and reach lower regions of the gel.

**Overview**
Steps in 2D-Gel Electrophoresis

1. Sample preparation
2. Isoelectric focusing (first dimensions)
3. SDS-PAGE (second dimension)
4. Visualisation of proteins spots
5. Identification of protein spots

Sample Preparation
Appropriate sample preparation is absolutely essential for good 2-D results. The optimal procedure must be determined empirically for each sample type. Ideally, the process will result in the complete solubilisation, disaggregation, denaturation, and reduction of the proteins in the sample.

In order to characterize specific proteins in a complex protein mixture, the proteins of interest must be completely soluble under electrophoresis conditions. Different treatments and conditions are required to solubilize different types of protein samples; some proteins are naturally found in complexes with membranes, nucleic acids, or other proteins, some proteins form various non-specific aggregates, and some proteins precipitate when removed from their normal environment. The effectiveness of solubilisation depends on the choice of cell disruption method, protein concentration and dissolution method, choice of detergents, and composition of sample solution.
To fully analyse all intracellular proteins, the cells must be effectively disrupted. Choice of disruption method depends on whether the sample is from cells, solid tissue, or other biological material and whether the analysis is targeting all proteins or just a particular subcellular fraction.

Proteases may be liberated upon cell disruption. Proteolysis greatly complicates analysis of the 2-D result, thus the protein sample should be protected from proteolysis during cell disruption and subsequent preparation.

If only a subset of the proteins in a tissue or cell type is of interest, prefractionation can be employed during sample preparation. If proteins from one particular subcellular compartment (e.g. nuclei, mitochondria, plasma membrane) are desired, the organelle of interest can be purified by differential centrifugation or other means prior to solubilisation of proteins for 2-D electrophoresis.

Precipitation of the proteins in the sample and removal of interfering substances are optional steps. The decision to employ these steps depends on the nature of the sample and the experimental goal. Precipitation procedures, which are used both to concentrate the sample and to separate the proteins from potentially interfering substances. Removal techniques, which eliminate specific contaminants (salts, small ionic molecules, ionic detergents, nucleic acids, polysaccharides, lipids, and phenolic compounds) might have on the 2-D result if they are not removed.

The composition of the sample solution is particularly critical for 2-D because solubilisation treatments for the first-dimension separation must not affect the protein pI, nor leave the sample in a highly conductive solution. In general, concentrated urea as well as one or more detergents are used.

**Methods of cell disruption**

Cell disruption should be performed at cold temperatures. Keep the sample on ice as much as possible and use chilled solutions.

Proteases may be liberated upon cell disruption, thus the protein sample should be protected from proteolysis if one of these methods to be used. It is generally preferable to disrupt the sample material directly into a strongly denaturing lysis solution, in order to rapidly inactivate proteases and other enzymatic activities that may modify proteins. Cell disruption is often carried out in an appropriate solubilisation solution for the proteins of interest.

**Gentle lysis methods**

These methods are generally employed when the sample of interest consists of easily lysed cells (such as tissue culture cells, blood cells, some microorganisms). Gentle lysis methods can also be employed when only one particular subcellular fraction is to be
analysed. For example, conditions can be chosen in which only cytoplasmic proteins are released, or intact mitochondria or other organelles are recovered by differential centrifugation. Sometimes these techniques are combined (e.g. osmotic lysis following enzymatic treatment, freeze-thaw in the presence of detergent).

**More vigorous lysis method**

These methods are employed when cells are less easily disrupted, i.e. cells in solid tissues or cells with tough cell walls. More vigorous lysis methods will result in complete disruption of the cells, but care must be taken to avoid heating or foaming during these procedures.

**Protection against proteolysis**

When cells are lysed, proteases are often liberated or activated. Degradation of proteins through protease action greatly complicates the analysis of 2-D electrophoresis results, so measures should be taken to avoid this problem. If possible, inhibit proteases by disrupting the sample directly into strong denaturants such as 8 M urea, 10% TCA, or 2% SDS (34-38). Proteases are less active at the lower temperatures, so sample preparation at as low a temperature as possible is recommended.

**Precipitation procedures**

Protein precipitation is an optional step in sample preparation for 2-D electrophoresis. Precipitation, followed by resuspension in sample solution, is generally employed to selectively separate proteins in the sample from contaminating species such as salts, detergents, nucleic acids, lipids, etc., that would otherwise interfere with the 2-D result. Precipitation followed by resuspension can also be employed to prepare a concentrated protein sample from a dilute source (e.g. plant tissues, urine).

**Removal of contaminantants that affect 2-D results**

Non-protein impurities in the sample can interfere with separation and subsequent visualisation of the 2-D result, so sample preparation can include steps to rid the sample of these substances.

Salt contamination is the most frequent cause of insufficient focusing of the protein spots!

- **Protease inhibitors**
  - PMSF (Phenylmethylsulfonyl Fluoride), Pefabloc, EDTA, leupeptin, Aprotinin, Pepstatin.
- **Contaminant removal**
  - Filtration, Centrifugation, Chromatography, Solvent Extraction.

Sample Solubilisation

Much like regular old’ SDS-PAGE, those tissue or blood samples for 2DE need to be processed and solubilized before they can be loaded into the IEF gel. In an ideal world, all proteins would
solubilise immediately with no qualitative or quantitative changes. In the real world, not all proteins are equal, and this first step immediately tilts 2DE towards detecting highly soluble and abundant proteins. Don’t reach for Tween and Triton, however; solubilisation requires IEF-compatible lysis reagents, such as electrically neutral detergents (e.g. CHAPS) and chaotropes like urea.

Obnoxiously enough, urea can interfere with downstream steps! As a result, this step requires a great deal of sample-specific optimization to avoid bias and contamination.

- **Protein Solubilisation**
  - 8 M Urea (neutral chaotrope)
  - 4% CHAPS (zwitterionic detergent)
  - 2-20 mM Tris base (for buffering)
  - 5-20 mM DTT (to reduce disulphides)
  - Carrier ampholytes to IPG buffer (up to 2% v/v) to enhance protein solubility and reduce charge-charge interactions.

2. **Isoelectric Focusing**

Now, as the sample is solubilised, time to load it on an IEF gel where, similar to SDS-PAGE, the proteins will be pushed through the acrylamide gel by an electric field. Where IEF gets more exciting is that the gel incorporates a pH gradient, and each protein moves only until it reaches its isoelectric point (pI). The pI is the pH where a protein has no net charge, meaning the field has no effect and the protein stays put, focusing tightly into a band within 0.01 pH unit of its pI.

As straightforward as this sound, IEF throws several wrenches into the scientific works. First, proteins become less soluble and can even precipitate out as they move closer to their pI, especially in low-salt, IEF-friendly buffers. Second, IEF gels and buffers interfere with sample prep for mass spectrometry (MS) and can be difficult to stain for analysis. This means IEF must almost always be done first so that SDS-PAGE can make the sample MS-compatible. Finally, it’s very easy to contaminate your sample with keratin, so this step requires gloves, diligent depilation, and working behind a “sneeze shield.”

- A protein with a net negative charge will migrate toward the anode, becoming less negatively charged until it also reaches zero net charge.
- If a protein should diffuse away from its pI, it immediately gains charge and migrate back. This is the focusing effect which allows proteins to be separated on the basis of very small charge differences.
- The resolution is determined by the slope of the pH gradient and the electric field strength so, IEF is therefore performed at high voltages (typically in excess of 1000 V).
- When the proteins have reached their final positions in the pH gradient, there is very little ionic movement in the system, resulting in a very low final current.
- The original method for first-dimension IEF depended on ampholyte-generated pH gradients in cylindrical polyacrylamide gels cast in glass rods or tubes. Now it is replaced by DryStrip gels.
- **Isoelectric Point (pI):** pH at which a protein has a neutral charge; loss or gain of protons H+ in a pH gradient (in a pH below their pI, they carry a net negative charge)
• Requires very high voltages (10000 V)
• Requires a long period of time (10 h)
• Degree of resolution determined by slope of pH gradient and electric field strength.
• Uses ampholytes to establish pH gradient.
• IPG strips:

An immobilised pH gradient (IPG) is made by covalently integrating acrylamide and variable pH ampholytes at time of gel casting. Stable gradients

1. SDS-PAGE

After separating the proteins based on their pI, you need to separate them according to their molecular weights. This can be done by applying an equilibration step to the strip containing the separated proteins (to reduce any disulphide bonds that may have formed during IEF phase) and treating the proteins with SDS. Once this is done, the proteins will be aligned along two axes – isoelectric point on one end and molecular weight on the other.

Analysis

• SDS-PAGE is an electrophoretic method for separating polypeptides according to their molecular weights. The technique is performed in polyacrylamide gels containing sodium dodecyl sulphate (SDS).
• SDS is an anionic detergent. SDS masks the charge of the proteins themselves net negative charge per unit mass.
• Besides SDS, a reducing agent such as DDT is also added to break any disulphide bonds present in the proteins.
• When proteins are treated with both SDS and a reducing agent, the degree of electrophoretic separation within a polyacrylamide gel depends largely on the molecular weight of the protein.
• In fact, there is an approximate linear relationship between the logarithm of the molecular weight and the relative distance of migration of the SDS-polypeptide complex.
• Separation of proteins on the basis of MW, not pI
• Requires modest voltages (200 V)
• Requires a shorter period of time (2h)
• Presence of SDS is critical to disrupting structure and making mobility ~1/MW.
• Degree of resolution determined by %acrylamide & electric field strength.
• Detection is done with the help of the following:
  o Coomassie Stain (100ng to 10μg protein)
  o Silver Stain (1ng to 1μg protein)
  o Fluorescent (Sypro Ruby) Stain (1ng & up)
4. Protein detection and image analysis

• This step plays a crucial role, as

  o Only what is detected can further be analysed and

  o Quantitative variations observed at this stage are the basis to select the few spots of interest, in comparative studies, that will be the only ones processed for further analysis with mass spectrometry.

• Detection with organic dyes can be summarized in one single process, colloidal Coomassie Blue staining, which has really become a reference standard.

• Although the sensitivity is moderate and homogeneity are good and compatibility with mass spectrometry is excellent.

• Silver staining is much more sensitive but less homogeneous, because of its delicate mechanism, and its compatibility with mass spectrometry is problematic.

• The consequence of the presence of formaldehyde at the image development step, formaldehyde-free silver staining protocols have been recently proposed.

• Protein detection by fluorescence give good sensitivity and also good compatibility with mass spectrometry.

• Other modes of detection are environment-sensitive probes, noncovalent binding and covalent binding.

• The use of chemically related, reactive fluorescent probes differing mainly by their excitation and emission wavelengths allows to perform multiplexing of samples on 2D gels.
• This multiplexing process solves in turn two difficult problems in the comparative analysis of gel images, namely the assignment of small positional differences and taking into account moderate quantitative changes.

**Reproducibility of 2-D PAGE**

• 2D gel-based proteomics is also widely used in bacterial proteomics, when the complexity of the sample is low enough.

• Theoretical proteome of *B. subtilis* showing the distribution of all 4100 predicted proteins according to their isoelectric points and molecular weights.

• Cytoplasmic proteins were harvested from *B. subtilis* wild type cells grown in Belitsky minimal medium at an OD500 of 0.4 and separated by 2-D GE.

**2D gels in post-translational modifications**

• 2D gels are also very appropriate when post-translational modifications are studied.

• Many post translational modifications do alter the pI and/or the MW of the proteins and thus induce position shifts in 2D gels. This is true for example for phosphorylation, glycosylation, but also more delicate modifications such as glutathionylation, or more forgotten modifications such as protein cleavage.

• The proteomics analysis reported here shows that a major cellular response to oxidative stress is the modification of several peroxiredoxins. An acidic form of peroxiredoxins appeared to be systematically increased under oxidative stress conditions due to post transcriptional modifications.

• 2D GE also used in immunoproteomics, where it is the immune response of patients that is probed at a proteomic level.

• 2D GE maps of proteins from Chlamydia trachomatis were probed with sera from 17 seropositive patients with genital inflammatory disease.

• Immunoblot patterns (comprising 28 to 2 spots, average 14.8) were different for each patient; however, antibodies against a spot-cluster due to the chlamydia-specific antigen outer membrane protein-2 (OMP2) were observed in all sera.

5. Protein Identification by Mass Profile Fingerprinting

• Due to the high resolution of 2D gels, very simple and cheap MS process can be used to identify a protein from a 2D gel.

• We can identify proteins at the sub-microgram level without sequence determination by chemical degradation.
The protein, usually isolated by one- or two-dimensional gel electrophoresis, is digested by enzymatic or chemical means and the masses of the resulting peptides are determined by mass spectrometry.

The resulting mass profile, i.e., the list of the molecular masses of peptides produced by the digestion, serves as a fingerprint which uniquely defines a particular protein.

This fingerprint may be used to search the database of known sequences to find proteins with a similar profile. This provides a rapid and sensitive link between genomic sequences and 2D gel electrophoresis mapping of cellular proteins.

**Protein identification using mass spectrometry**

**Introduction**

Since its invention in 1905, mass spectrometry (MS) has become a widely established technique for analysing chemical structures in quantities down to trace levels.

Due to a lack of suitable ionization techniques for high mass biomolecules, proteins remained inaccessible to MS analysis for decades. Since the introduction of soft ionization techniques such as Matrix Assisted Laser Desorption Ionization (MALDI) and Electrospray Ionization (ESI), MS at the end of the 1980s [1, 2] protein analysis by mass spectrometry underwent a rapid phase of development.

In parallel, an increasing number of full genome sequences for a variety of organisms are now available and numerous protein databases were constructed from this information.

Well annotated, high-quality protein databases built the ground on which high-throughput protein identification with mass spectrometry can be performed.

The modular arrangement of different types of mass analysers in combination with MALDI- or ESI has resulted in a wide variety of different mass spectrometric instrumentation (e.g., MALDI-TOF, ESI-Q-TOF, ESI-ion trap, MALDI-TOF/TOF, ESI-FT-ICR, etc.).

All of these MS techniques allowed the determination of the primary structure of a protein, though they always required additional sample preparation techniques. Furthermore, the analysis of posttranslational modifications such as phosphorylation or glycosylation has become possible.

Modern mass spectrometers now combine attributes like high sensitivity, mass accuracy, mass resolution, and rapid analysis as well as sophisticated data handling in a system-dependent manner.
In addition to these technical aspects in mass spectrometry, greatly improved sample separation and preparation techniques have also lead to enhanced sensitivity.

The quantification of chemically or metabolically labelled proteins is yet another focus of interest in mass spectrometry. Despite these advances current MS approaches still have limitations and are therefore subjected to further development. The aim of this paper is therefore to highlight the different mass spectrometric techniques currently applied in proteome research by giving a brief overview of methods for identification of posttranslational modifications and discussing their suitability of strategies for protein quantification.

**Protein identification using mass spectrometry:** A method overview analysers. Ions are generally detected by secondary electron multipliers (SEM) or by micro channel plate (MCP) detectors. Usually, the detector enables the mass spectrometer to generate an analog signal, by producing secondary electrons, which are further amplified.

The analog signal from the detector is finally digitized and processed by a computer. Several additional designs and applications for ion detection are in use, e.g., photon-sensitive detectors but are beyond the scope of this review.

Giving a brief overview of methods for identification of posttranslational modifications and discussing their suitability of strategies for protein quantification.

The primary structure of a protein, though they always required additional sample preparation techniques. Furthermore, the analysis of posttranslational modifications such as phosphorylation or glycosylation has become possible.

Modern mass spectrometers now combine attributes like high sensitivity, mass accuracy, mass resolution, and rapid analysis as well as sophisticated data handling in a system-dependent manner. In addition to these technical aspects in mass spectrometry, greatly improved sample separation and preparation techniques have also lead to enhanced sensitivity. The quantification of chemically or metabolically labelled proteins is yet another focus of interest in mass spectrometry (see previous chapter). Despite these advances current MS approaches still have limitations and are therefore subjected to further development. The aim of this paper is therefore to highlight the different mass spectrometric techniques currently applied in proteome research by giving a brief overview of methods for identification of posttranslational modifications and discussing their suitability of strategies for protein quantification.

General technical considerations-

Mass spectrometry is a highly sensitive and accurate method for the determination.

Primary structure of a protein, though they always required additional sample preparation techniques. Furthermore, the analysis of posttranslational modifications such as phosphorylation or glycosylation has become possible.
Modern mass spectrometers now combine attributes like high sensitivity, mass accuracy, mass resolution, and rapid analysis as well as sophisticated data handling in a system-dependent manner.

In addition to these technical aspects in mass spectrometry, greatly improved sample separation and preparation techniques have also lead to enhanced sensitivity.

The quantification of chemically or metabolically labelled proteins is yet another focus of interest in mass spectrometry. Despite these advances current MS approaches still have limitations and are therefore subjected this paper is therefore to highlight the different mass spectrometric techniques currently applied in proteome research by giving a brief overview of methods for identification of posttranslational modifications and discussing their suitability of strategies for protein quantification.

MALDI – Source and sample introduction-

For this ionization technique, the purified analyte is generally dissolved in a matrix solution, spotted onto a solid target and co-crystallized with the matrix.

The matrix, which typically contains a UV sensitive aromatic compound, is used to facilitate UV-laser energy-absorption and energy-transfer. The irradiated area of the crystals and the analyte embedded therein are vaporized by the laser energy uptake. Although the mechanism of ion formation during the MALDI process is still a matter of some controversy the efficiency of ionization and the initial ion velocity can be controlled by the choice of matrix or the composition of the analyte sample. Typical matrix compounds include 2,5-dihydroxybenzoic acid (DHB), 3,5-dimethoxy-4-hydroxy-cinnamic acid (sinapinic acid), and L-cyano-4-hydroxy-cinnamic acid (HCCA). The analyte molecules are normally ionized by simple protonation, leading to the formation of the typical singly charged [M+H] type species (where M is the mass of the analyte molecule). Trace contaminations of earth alkali metals in the matrix will especially generate [M+X]+ ions (where X = Li, Na, K, etc.). Once the ions are vaporized, they are accelerated in an electric field and different mass analyzers can be used to measure their m/z. The most commonly used instrument type is the MALDI-TOF-MS design whose performance has dramatically improved due to the introduction of delayed ion extraction and reflectron technology.

The MALDI evaporation process generates ions with an initial velocity distribution, which normally causes low resolution due to start-time errors.
This effect is compensated with delayed ion extraction by the use of a two-stage acceleration field in combination with a delay time resulting from appropriate acceleration voltages following the laser pulse.

MALDI-TOF instruments are capable of analyzing intact proteins and complex peptide mixtures since they have an almost unlimited mass range that can be analyzed within their flight tube. The MALDI technique generates singly-charged molecules with a typical detection limit in the low femtomol range.

MALDI has long been considered a ‘soft’ ionization technique that apparently generates almost exclusively intact ions. In fact, a significant degree of metastable decay occurs after ion acceleration which is used in reflectron TOF or in modern TOF-TOF analysers for simple post-source decay (PSD) analysis. Such an analysis provides some structural information about an analyte ion, which can be used for the interpretation of the mass spectrum and the identification of the analyte molecule.

**Protein identification using mass spectrometry: A method overview.**

ESI – Source and sample introduction-

The introduction of charged molecules into the mass spectrometer with ESI sources is carried out using different quantities of aqueous sample under atmospheric pressure conditions.

In nanoelectrospray (nanoES) technology, for example, only a few microliters of sample are needed for spraying from the highly charged (up to 3,000 V) tip of a metal coated glass needle to the inlet of the mass spectrometer.

The finely field, which helps to accelerate the charged droplets and to form a constant spray of 20–200 nL/min.

Evaporation of the solvent, which is normally supported by a dry gas, decreases the droplet size and thus increases the surface charge density, finally releasing solvent-free ionized analyte molecules.

Here, organic solvents, e.g., 2-propanol or ace-tonitrile, facilitate the evaporation process and enhance the formation of a stable spray. The resulting ions are directed into an orifice and focused stepwise under increasing vacuum conditions by electrostatic lenses to form an ion beam. The ESI technique generates primarily multiply charged molecules. It has been demonstrated that the maximum charge states and charge state distributions of ions generated by electrospray ionization are influenced by solvents that are more volatile than water. Protein identification using mass spectrometry: A method overview analyzers. Ions are generally detected by secondary electron multipliers (SEM) or by microchannel plate (MCP) detectors. Usually, the detector enables the mass spectrometer to generate an analog signal, by producing secondary electrons, which are further amplified. The analog signal from the detector is finally
digitized and processed by a computer. Several additional designs and applications for ion detection are in use, e.g., photon-sensitive detectors but are beyond the scope of this review.

**Analysis of proteins and peptides by mass spectrometry**

In this section, the most widely used modern mass spectrometry techniques for identification of proteins and peptides will be described. At present, the typical approach for analysing proteins is to gather protein spots from 2-D gels, to convert them into peptides, obtain sequence tags of the peptides, and then identify the corresponding proteins from matching sequences in a database. The procedure for a successful protein identification is thereby arranged in a hierarchy of methods depending on the degree of protein sample complexity.

De novo sequencing using mass spectrometric data:
De novo is Latin which means “over again” or “anew”. The de novo peptide sequencing is a method for peptide sequencing performed without prior knowledge of the amino acid sequence.

This method can obtain the peptide sequences without a protein database, which can overcome the limitations of database-dependent methods like peptide mass fingerprinting (PMF).

It uses computational approaches to deduce the sequence of peptide directly from the experimental MS/MS spectra. It can be used for un-sequenced organisms, antibodies, peptides with posttranslational modifications (PTMs), and endogenous peptides. In this method, the peptide is fragmented along the peptide backbone and the resulting fragment ions are measured to produce spectra.

There are 3 ways to break bonds to form peptide fragment: alkyl carbonyl (CHR-CO), peptide amide bond (CO-NH), and amino alkyl bond (NH-CHR). Therefore, it can form 6 types of
fragmentation ions, including the N-terminal charged fragment ions which are classed as a, b, or c, and the C-terminal charged ones which are classed as x, y, or z.

And because the peptide amide bone (CO-NH) is the most vulnerable, the most common peptide fragments observed in low energy collisions are a, b and y ions.

De novo methods use the knowledge of the fragmentation methods employed in the MS. CID, Collision induced dissociation, also known as collisionally activated dissociation, is the most common form of fragmentation. In this method the ions can obtain high kinetic energy and
collide with neutral molecules. Some of the kinetic energy is converted into internal energy which leads to bond breakage and the fragmentation of the molecules into smaller fragments.

This method results in the formation of b and y series ions from the precursor ion. The Electron capture dissociation (ECD) and Electron transfer dissociation (ETD) have been implemented in the recent mass spectrometer. In these methods, ions are fragmented after reaction with electrons. After fragmentation, it forms c and z type ions through cleavage of the peptide bond between the amino group and alpha carbon.

The mass can usually uniquely determine the residue. The main principle of de novo sequencing is to use the mass difference between two fragment ions to calculate the mass of an amino acid residue on the peptide backbone. For example, the mass difference between the y7 and y6 ions in the following figure is equal to 101, which is the mass of residue T.

Thus, if one can identify either the y-ion or b-ion series in the spectrum, the peptide sequence can be determined. However, the spectrum obtained from the mass spectrometry instrument does not tell the ion types of the peaks, which need an expert or a computer algorithm to figure out.

There are couples of software packages used for de novo sequencing, such as PEAKS, Lutefisk, PepNovo, SHERENGA, etc. But there are some notes you have to mind. y and b ion fragments which contain the amino acid residues R,K,Q, and N may appear to lose ammonia (-17). Y and b ion fragments which contain the amino acid residues S, T, and E may appear to lose water (-18).

De novo sequencing can identify previous unknown peptide sequences. In addition, it can search for posttranslational modifications or for identifications of mutations by homology-based software. However, de novo sequencing will not be able to derive a complete sequence or will have uncertainty in a portion of the derived sequence. And sometimes it can be difficult to determine the directionality of a sequence. Low mass accuracy fragment ion measurements cannot distinguish between lysine and glutamine (differ by 0.036 Da) nor between phenylalanine and oxidized methionine (differ by 0.033 Da).