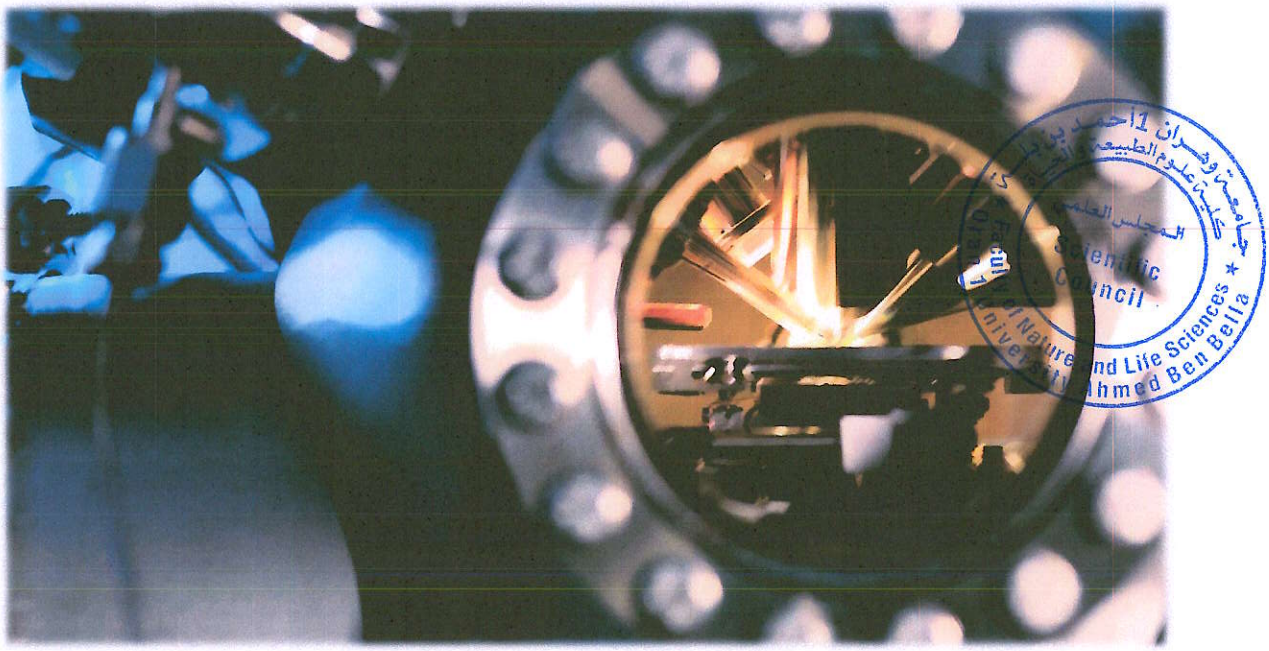
الجمهورية الجزائرية الديمقراطية الشعبية
People's Democratic Republic of Algeria
وزارة التعليم العالي و البحث العلمي
Ministry Of Higher Education and Scientific Research
كلية علوم الطبيعة و الحياة
Faculty Of Natural and Life Sciences

Course Pack: PROTEOMICS.

Biotechnology Department

Master 2 Biochemistry-Immunology



Dr. SEFRAOUI KHELIL Imane,
Master of Conferences

Academic Year: 2023-2024

الجمهورية الجزائرية الديمقراطية الشعبية
People's Democratic Republic of Algeria
وزارة التعليم العالي و البحث العلمي
Ministry Of Higher Education and Scientific Research
كلية علوم الطبيعة و الحياة
Faculty Of Natural and Life Sciences

جامعة وهران 1
Oran University 1
أحمد بن بلة

Course Pack:

PROTEOMICS.

Biotechnology Department

Master 2 Biochemistry-Immunology

Dr. SEFRAOUI KHELIL Imane,

Master of Conferences

Academic Year: 2023-2024

SYNOPSYS

The syllabus for the course on the subject of Proteomics, as presented in the Master's program framework, is outlined below:

Master's Title: Biochemistry-Immunology, Semester: 3

Teaching Unit	VHS	V.H weekly				Factor	Credits	Method of assessment	
	14-16 weeks	C	TD	TP	Other* VHS			Continuous	Exam
UEF2 (O/P)									
Subject 1: Proteomics	**67h30**	**3h00**	**1h30**		**82h30**	**3**	**6**	**X**	**X**

Teaching Objectives: The objective is to grasp the concept of the proteome and its metabolic and pathological variability. Tools for proteome study will be discussed.

Recommended Prerequisite Knowledge: Biochemistry of the TCSNV.

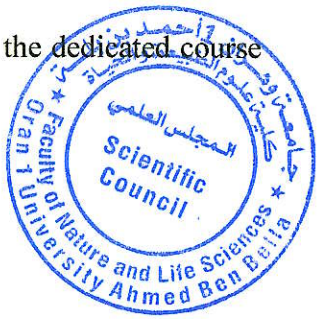
SUMMARY

LIST OF FIGURES

LIST OF TABLES

PREFACE

This course pack covers and details the various chapters of the Proteomics module. This module is a fundamental component of the Master M2 program in Biochemistry-Immunology, launched in 2016 (following harmonization) within the Department of Biotechnology, Faculty of Natural and Life Sciences. I have been in charge of this course since 2020. This course describes the proteomic analysis workflow, mainly focusing on the "Bottom-up" approach. It starts with a general introduction covering concepts of the proteome, historical milestones, and the main stages of proteomics. It then delves into sample preparation for proteomic analysis and explores protein fractionation by gel electrophoresis and liquid Chromatography methods. The course also includes mass spectrometry analysis, addressing basic concepts in MALDI-TOF and ESI MS/MS mass spectrometry, protein identification strategies, and database search for protein. Finally, it covers differential and quantitative proteomic analysis. The lessons are delivered interactively through PowerPoint slides. At the end of each chapter, digital resources containing animated illustrations and explanatory texts are available to students via the dedicated course space on the e-learning platform.



Abstract

Proteomics, the systematic analysis of expressed proteins, describes the technologies collectively used to define the protein complement of the genome or "proteome." These technologies are based on the large-scale screening approach for separating and identifying proteins, simultaneously allowing an integral study of many proteins. Changes in protein expression in variable conditions (Overexpression, silencing, drug influence) can be detected. Proteomics tries to determine all proteins' bio-chemicals and physical properties such as molecular weight (MW), isoelectric point (pI), protein sequences, and three-dimensional structures. Among proteomics analysis techniques, mass spectrometry (MS) linked with two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) or liquid chromatography (LC), along with bioinformatics, are the most efficient technologies used in these studies. Mass spectrometry has become an essential tool for identifying and quantifying proteins due to its high reproducibility, broad range of analyzable masses, and measurement accuracy and precision. This course describes the proteomic analysis workflow, from sample preparation to data analysis, and includes differential and quantitative proteomics.

I. INTRODUCTION

The term "proteome," initially coined in 1994 by Marc Wilkins, has become widely used in scientific writing to define the total set of proteins expressed by a genome in a cell, tissue, or organism at a given time in response to a given environment (Wilkins et al., 1996). This word "hybrid," derived from both "protein" and "genome," describes the complex and dynamic process of protein expression that takes place at different reference points, which can extend from individual cells to entire organisms. Advances in molecular biology have allowed researchers to delve deeper into the practical applications of studying the proteome, which has significant implications for understanding biological processes and developing targeted therapies. It is an essential concept in post-genomic studies, as it encompasses not only the individual proteins generated by genes but also factors such as post-translational modifications and proteolytic cleavage that can create multiple protein variants (Rotilio et al., 2012). Both genetic factors and environmental conditions influence the proteome, representing the protein complement of a genome at a specific point in time. Table 1 includes historical milestones relevant to proteomics, and Box 1 contains more detailed descriptions of the critical analytic techniques used in proteomic studies.

Table 1: Historical Milestones in Proteomics (significant historical occurrences in proteomics).

Milestone	Year	Description	References
Introduction of two-Dimensional gel	1975	The real starting point for proteomics was 1975, when the two-dimensional gel was introduced, and proteins from the bacteria *Escherichia coli* and the mouse were mapped.	(Klose, 1975; O'Farrell, 1975)
Western blotting	1979	Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: analysis of various proteins with specific reactions or ligands.	(Towbin et al., 1979)
Electrospray Ionization (ESI) and Matrix-assisted laser desorption/ionization (MALDI)	1989	Large molecules were first vaporized and ionized using ESI and MALDI. This enabled the transformation of proteins into the gas phase for MS analysis.	(Fenn et al., 1989; Karas & Hillenkamp, 1988)
Algorithms	1990s	Querying databases using different algorithms.	(Bartels, 1990)
LC/MS/MS	1993	Combining microcolumn liquid chromatography (LC) with tandem mass spectrometry improves the ability to manipulate small quantities of peptides in complex mixtures for sequence analysis.	(Arnott et al., 1993)
Proteome and proteomics	1995	The first use of the terms "proteome" and "proteomics" denotes the full complement of an organism's proteins and their study, respectively.	(Wilkins et al., 1996)
DIGE	1997	Difference gel electrophoresis may reproducibly detect differences between two protein samples using just one gel.	(Unlü et al., 1997)
Y2D	1999	Yeast two-hybrid and reverse two-hybrid assays to Identify Interacting Proteins.	(Vidal & Legrain, 1999)
ICAT	1999	Isotopic labeling method in which chemical labeling reagents are used for quantification of proteins.	(Gygi et al., 1999)
MudPIT	2001	Multidimensional protein identification technology uses two-dimensional (2D) chromatography to separate a peptide mixture before analysis by mass spectrometry.	(Washburn et al., 2001)
Protein microarrays	2001	Protein chips are the emerging class of proteomics techniques capable of high-throughput detection from small samples.	(Zhu & Snyder, 2001)
SILAC	2002	The first report was on using SILAC (Stable isotopic labeling by amino acids in cell culture) in proteomic studies.	(Ong et al., 2002)
iTRAQ	2004	Isobaric tagging for relative and absolute quantitation.	(Ross et al., 2004)
Targeted proteomics	2012	Analysis of a preselected group of proteins delivers more precise, quantitative, sensitive data to biologists.	(Picotti & Aebersold, 2012)
Draft map of the human proteome	2014	The issue contains two large-scale proteomics analyses based on publicly available databases, ProteomicsDB and Human Proteome Map.	(Ezkurdia et al., 2014; Kim et al., 2014; Wilhelm et al., 2014)

Proteomics is the systematic study of all or a large subset of the proteins expressed within a cell, tissue, or organism at a given time, including their structure, function, and interactions. This concept was developed as a result of the evolution of technologies that emerged in the 1980s, such as two-dimensional electrophoresis developed in 1975 simultaneously but independently by O'Farrell (O'Farrell, 1975) and Klose (Klose, 1975), which can simultaneously separate hundreds or even thousands of polypeptides and ensure the measurement of protein properties such as molecular weight and isoelectric point improving the identification and characterization of various protein species and isoforms (Gorg, 1991). A modified version of 2DE called difference in-gel electrophoresis (DIGE) allows two different samples (healthy and diseased) to be loaded into the same gel, so it avoids many of the drawbacks of the original method, including inter-gel variation, high labour, and time costs, and a small dynamic range. This makes quantitative proteomics research more sensitive, precise, and reproducible (Larbi & Jefferies, 2009). Meanwhile, 2-DE, MS, and bioinformatic tools are the critical components of an approach termed ''the classical proteomic methodology''(Görg et al., 2009).

The development of soft ionization procedures in mass spectrometry, such as matrix-assisted laser desorption/ionization (MALDI) (Karas & Hillenkamp, 1988) and electrospray ionization (ESI) (Fenn et al., 1989), was revolutionary. Peptide mass fingerprints after enzymatic digestion are ideal for protein identification (Pappin et al., 1993). Confirmation and complementation of identification were obtained by fragmenting the peptides using MS/MS (Dancík et al., 1999; Mann & Wilm, 1994).

Gel electrophoresis has been a widely used technique for protein separation in proteomic analysis. However, chromatographic separation-purification techniques combined with mass spectrometry can provide a more accurate and detailed understanding of proteins' molecular mass and primary sequence (Shi et al., 2004). This technique delivers such precise results that it has become the technique of choice in analytical proteomics. Two approaches have emerged from these developments: bottom-up proteomics and top-down proteomics.

The "bottom-up" approach is the most widely used method in proteomics (Y. Zhang et al., 2013). Bottom-up protein analysis characterizes proteins by analyzing peptides released from them through proteolysis. When bottom-up is performed on a mixture of proteins, it is called "shotgun proteomics" because of its analogy to shotgun genomic sequencing (Yates, 2004).

The "top-down" approach involves the analysis of whole proteins (Catherman et al., 2014). These are identified using MS and MS/MS mass spectra, which can be quite complex due to

the various charge states that protein fragments can adopt. Injecting intact proteins into the mass spectrometer allows for better characterization of post-translational modifications and prevents some inference issues that can occur in peptide-based proteomics. Therefore, this technique is most appropriate for simple and highly purified samples.

This course will describe a typical proteomic analysis workflow consisting of several significant steps, as illustrated in Figure 1.

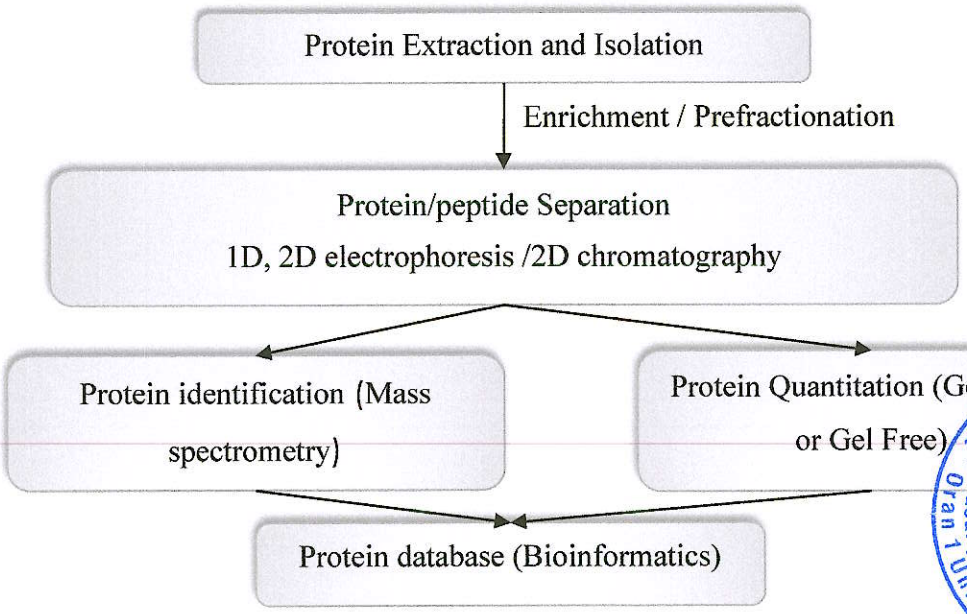


Figure 1. Schematic representation of the significant steps in proteomic analysis.

II. SAMPLE PREPARATION FOR A PROTEOMIC ANALYSIS

Sample preparation is one of the most crucial processes in proteomics research. The results of the experiment depend on the condition of the starting material. One of the most important factors is the nature and type of protein being worked with. Proteins vary in size, charge, sequence, and the presence of specific binding sites. This includes determining whether the protein is soluble or membrane-bound and any unique characteristics in its composition or structure that may impact the purification process, such as carbohydrate or lipid components (Liu et al., 2020). Additionally, it is crucial to ensure that the protein is expressed at a sufficient level for effective purification. The second axis, which determines the protocol to be developed, depends on the study's objectives.

Extraction steps may exploit differences in chemical/structural/functional properties between the target protein and other proteins in the crude mixture (Tab.2). By exploiting tremendous variations in physical and chemical properties among proteins, several different fractionation and chromatographic steps can usually be exploited to design a workable purification scheme (Labrou, 2014).

Table 2. Physicochemical basis of standard protein-separation methods (Labrou, 2014).

Properties	**Methods**
Solubility	- Saline precipitation by Ammonium sulfate "Salting out" - Organic solvents precipitation - Polyethylene glycol
Ionic Charge	- Ion exchange chromatography - Electrophoresis - Isoelectric focusing - Isoelectric precipitation
Polarity	- Reverse phase chromatography (HPLC) - Hydrophobic interaction chromatography
Molecular Size	- Dialysis and ultrafiltration - Gel electrophoresis - Gel filtration /size exclusion chromatography - Ultracentrifugation
Molecular recognition	- Affinity chromatography - Affinity precipitation

Generally, sample preparation aims to create a less complex peptide mixture that is suitable for analysis. It requires pre-fractionation, depletion of most unnecessarily abundant proteins, removal of DNA, lipids, and small metabolites, and sample clean-up from impurities (salts and remaining solid particles). Therefore, typical sample preparation processes for proteomics usually include lysis/homogenization, protein extraction/precipitation, pre-fractionation, and desalting (Fig.2) (Duong & Lee, 2023). Sample preparation is an essential stage affecting the overall efficiency of proteomic studies.

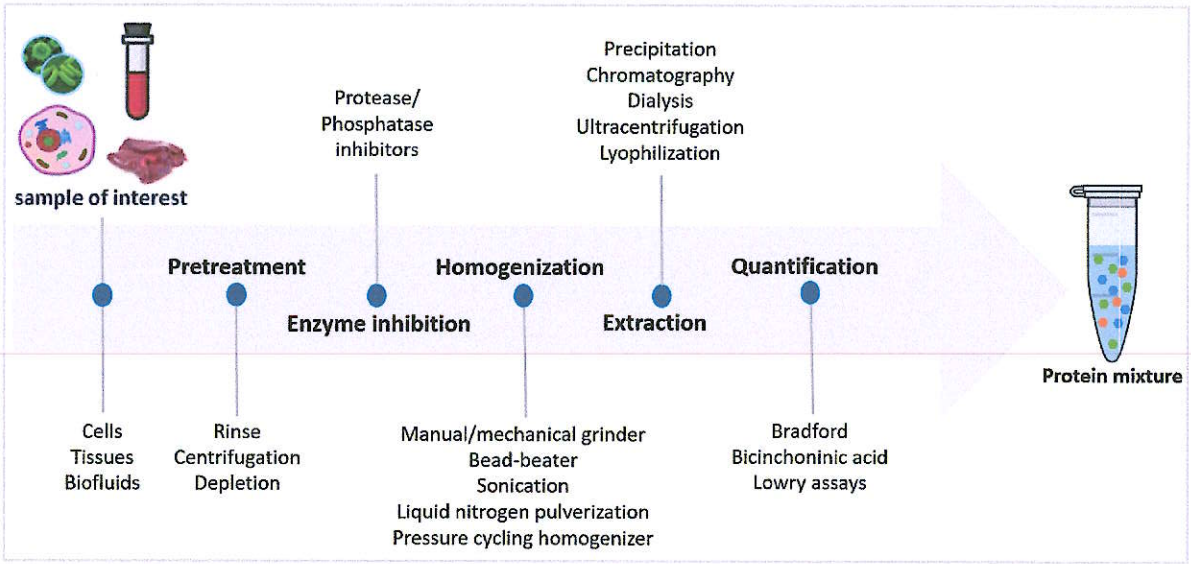


Figure 2. Sample preparation in proteomics (Duong & Lee, 2023).

II. 1 Sample Processing

The first step of each isolation and purification process is disrupting the cells containing the protein. Depending on how fragile the protein is, one of several techniques could be used, including repeated freezing and thawing, sonication, homogenization by high pressure (French press), homogenization by grinding (bead mill), and permeabilization by detergents (e.g., Triton X-100) and/or enzymes (e.g., lysozyme) (Bodzon-Kulakowska et al., 2007). Also, proteases are released during cell lysis, which will start digesting the proteins in the solution. As the protein of interest may be sensitive to proteolysis, it is essential to proceed quickly and conduct many steps at low temperatures to reduce unwanted proteolysis. Alternatively, one or more protease inhibitors can be added to the lysis buffer immediately before cell disruption (Görg et al., 2004). Sometimes, it is also necessary to add DNase to reduce the viscosity of the cell lysate caused

by a high DNA content. Cell debris can be removed by centrifugation so that proteins and other soluble compounds remain in the supernatant.

Due to most proteins' compartmentalization, cell substructures can be separated by a series of runs at different centrifugal forces or in a sucrose/mannitol gradient. This technique allows the separation of different cellular or tissue materials according to their density characteristics. It is useful for concentrating mitochondrial, membrane, nuclear, or other locally abundant proteins (Foster et al., 2006).

II. 2 Precipitation and Differential Solubilization

The most common methods of protein enrichment and purification rely on selective precipitation using acetone, trichloroacetic acid, ethanol, isopropanol, diethylether, chloroform/methanol, ammonium sulfate, polyethylene glycol (PEG), and several commercially available affinity precipitation kits (Bodzon-Kulakowska et al., 2007).

Precipitation using ammonium sulfate $(NH_4)_2SO_4$ is a typical first step in purifying bulk proteins after removing non-desired protein and other molecular contaminants by centrifugation (Wingfield, 2001). The process called "salting out" is shown in Figure 3. Ammonium sulfate is a commonly utilized substance because it is highly soluble in water, largely unaffected by temperature changes, and generally safe for most proteins. Remember that proteins precipitated in their original state by $(NH_4)_2SO_4$ are necessary for their structure and function studies. Furthermore, dialysis can be used to eliminate ammonium sulfate.

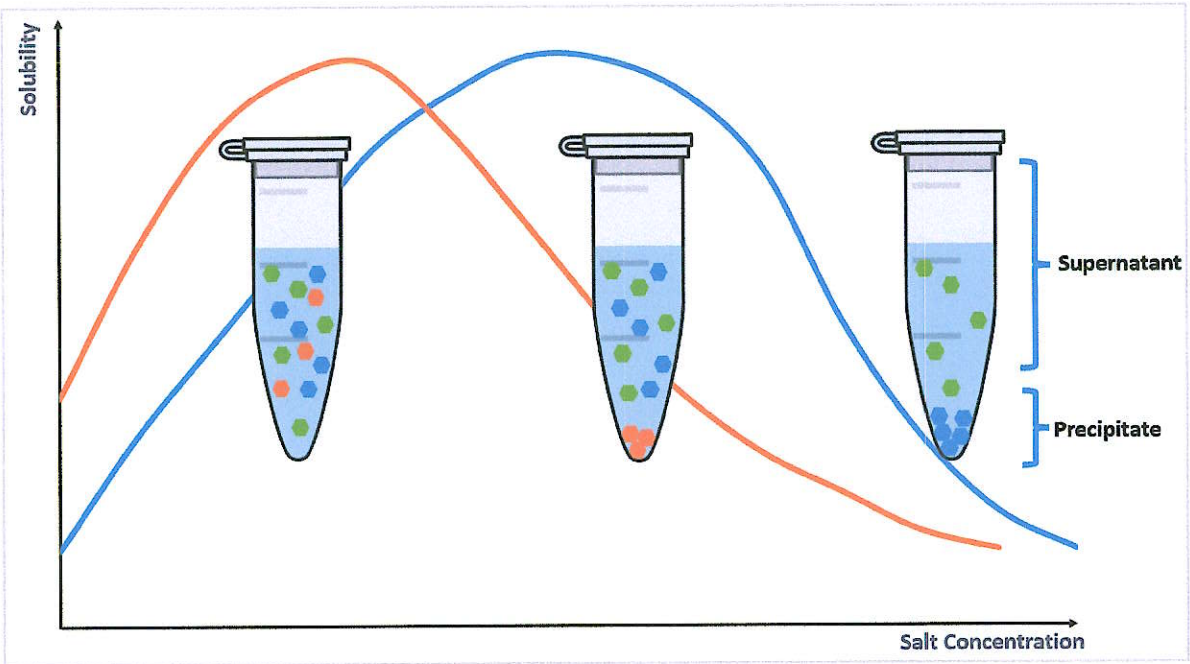


Figure 3. Protein precipitation by Salting out.

The proteins can also be isolated, usually by precipitation with acetone or trichloroacetic acid. Protection against proteolysis is ensured by solubilizing the sample directly in strong denaturants such as 8 M urea, 10% trichloroacetic acid, or 2% SDS and preparing the sample at as low a temperature as possible (Palii et al., 2007).

II. 3 Chromatography

Chromatography is integral to almost all protein purification techniques and is required to distinguish a particular protein from the thousands of other proteins found in cells and tissues. The kind of column and the molecular characteristics of the proteins determine how well they separate in a chromatography column. Proteins can be separated using chromatography in four primary ways: (bio)affinity (AC), ion exchange (IEXC), hydrophobic interaction (HIC), and size-exclusion chromatography (SEC). Ideally, this would involve chromatography with stationary and mobile phase conditions that retain only the protein of interest, eluting all other proteins. In the second step, the elution of the protein of interest is triggered (Labrou, 2014).

The purification technique typically includes one or more chromatographic steps for preparative protein purification. Chromatography's fundamental process involves passing a protein-containing solution down a column filled with chromatography resin chosen to separate proteins according to particular protein characteristics. The conditions needed to elute the protein from the column or the time required to pass the column can be used to distinguish between different proteins based on their interaction with the material in the column. Usually, proteins are detected as they are eluting from the column by measuring the absorbance at 280 nm, at which the aromatic amino acids absorb (Aslam et al., 2017).

II. 3. 1 Size Exclusion Chromatography (also known as Gel Filtration Chromatography)

Proteins can be separated using this method according to their size and shape. The chromatography beads have tiny openings and pores into which proteins smaller than the pore diameter can enter. Larger proteins that can't join the pores flow around the beads and elute faster than small ones that enter the pores. They diffuse out of the pores and enter the rest of the moving solvent before getting trapped again for a short time in more pores. Ultimately, they work through the column and elute at a more significant volume than proteins, which can't enter the pores (Hong et al., 2012). Proteins will, therefore, be separated according to size. The eluate is collected in sequential test tubes or fractions. Figure 4 below shows the pores as channels that go through the bead. The openings in resin beads should be considered pores, not channels.

The stationary phase is typically composed of polymerized agarose (Sepharose), acrylamide beads (Bio-Gel), or dextran polymer (Sephadex) (Hong et al., 2012).

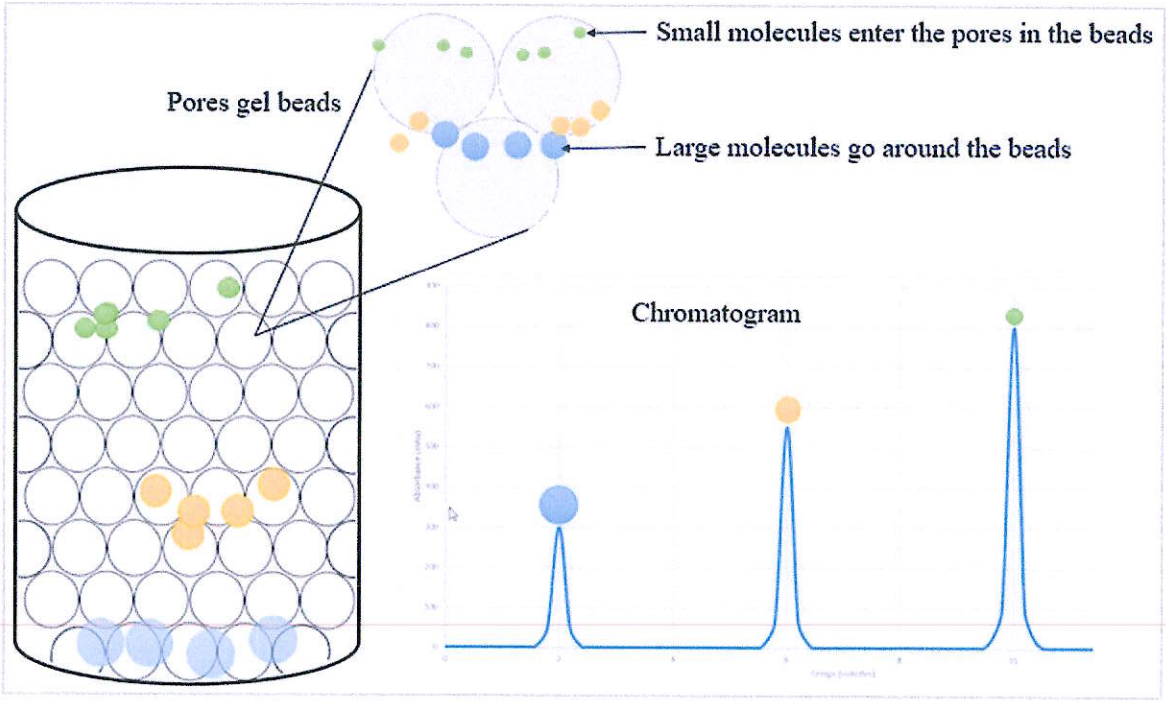


Figure 4. Size Exclusion Chromatography. Inspired by Hong et al. (2012).

Gel filtrations are commonly used as the final purification step (polishing); they are handy for separating unwanted oligomeric formations.

II. 3. 2 Separation based on charge - Ion Exchange Chromatography

The chromatography resin in this type consists of an agarose, acrylamide, or cellulose resin or bead, which is derivatized to contain covalently linked positively or negatively charged groups (Jungbauer & Hahn, 2009). Proteins in the mobile phase will bind through electrostatic interactions to the charged group on the column. In a mixture of proteins, positively charged proteins will bind to a cation exchange resin containing negatively charged groups, like the carboxymethyl group. In contrast, the negatively charged proteins will pass through the column. The positively charged proteins can be eluted from the column with a mobile phase containing either a gradient of increasing salt concentration or a single higher salt concentration (isocratic elution). The most positively charged protein will be eluted at the highest salt concentration. Likewise, negatively charged proteins will bind to an anion exchange resin containing positively charged groups, like the diethylaminoethyl (DEAE), and can be separated analogously.

Proteins can be eluted by adjusting the pH to detach them based on their isoelectric point or increasing the salt concentration to compete with the charged groups on the resin (Staby et al., 2000).

Figure 5 shows an anion exchange column. The beads (blue) contain positively charged functional groups that can bind negative protein (purple) or concentrated regions of negative charge on a protein. Proteins bound through ion-ion interactions can be eluted by increasing the Cl^- concentration stepwise or gradient in the eluting solution. Ion exchange chromatography is a potent tool for protein purification and is frequently used in both analytical and preparative separations.

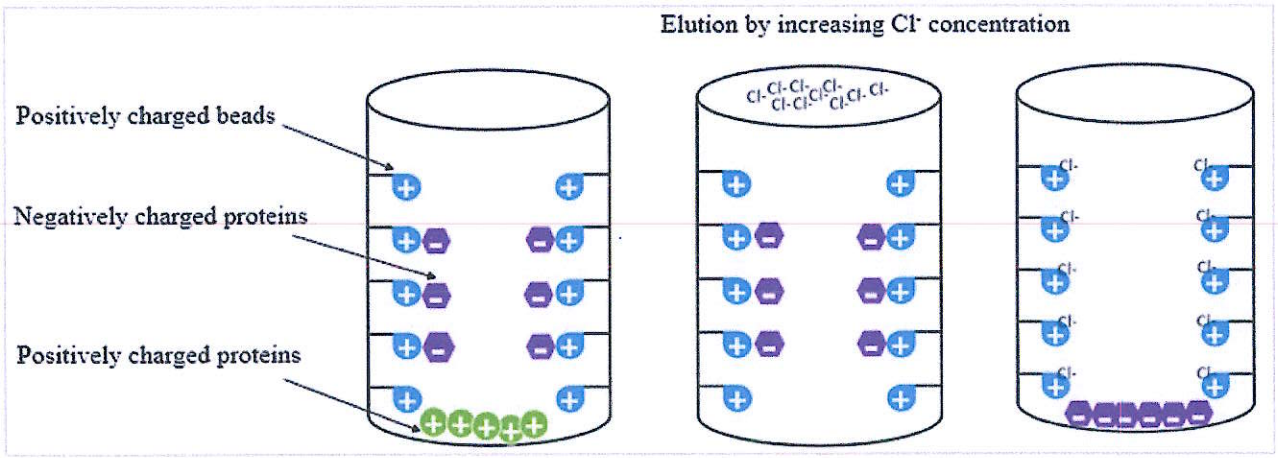


Figure 5. Ion Exchange Chromatography. Inspired by Staby et al. (2000).

II. 3. 3 Affinity Chromatography

In this technique, the chromatography resin is derivatized with a group that binds to a specific site on a protein of interest. This group may bind to the active site of an enzyme or an antibody that recognizes a specific amino acid sequence (an epitope) on a protein (Rodriguez et al., 2020).

Figure 6 illustrates these features of affinity chromatography. This technique relies on specific and reversible interactions between specific compounds (ligands) covalently bound to an inert support, which serves as the stationary phase, and their affinity partner in solution (the analyte or affinity substance). The ligand is attached to a matrix (resin) directly or via a linker. The linker is a polycarbonate chain (C6-C8) between the matrix and the ligand.

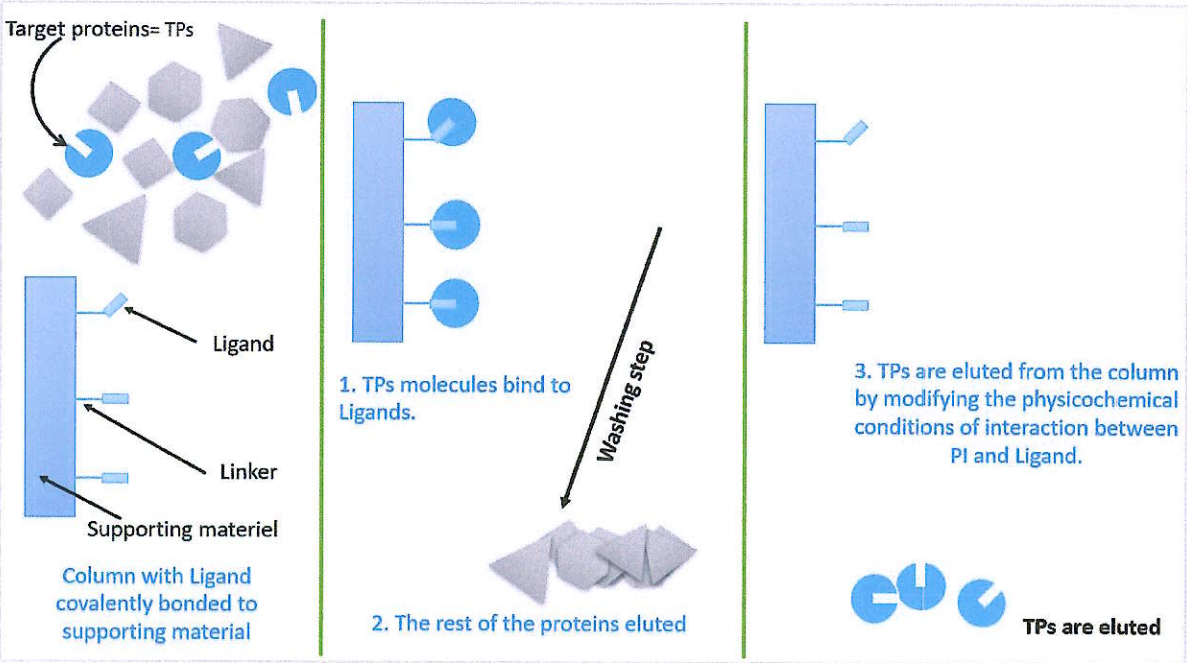


Figure 6. Example of Affinity Chromatography. Inspired by Rodriguez et al. (2020).

Immobilized-Metal Affinity Chromatography (IMAC)

Initially developed for purifying native proteins with an intrinsic affinity to metal ions (Porath et al., 1975), IMAC has proven to be a technology with a wide range of applications. On the chromatographic purification side, the range of proteins was expanded from the primary metalloproteins to antibodies, phosphorylated proteins, and recombinant His-tagged proteins. IMAC is used in proteomics approaches where fractions of the cellular protein pool are enriched and analyzed differentially (phosphoproteome, metalloproteome) by mass spectrometrical techniques.

The Ni-Nitrilotriacetic binds the His tag by chelating the nickel ion with the 6 histidine imidazole groups on the His-tagged protein, as shown in Figure 7. The His tag, probably the most widely used, binds strongly to divalent metal ions such as nickel and cobalt. The protein can be passed through a column containing Ni-nitrilotriacetic. All untagged proteins pass through the column. The protein can be eluted with imidazole, which competes with the imidazole side chain on the His tag for binding to the column, or by a decrease in pH, which decreases the affinity of the tag for the resin. While this procedure is generally used to purify recombinant proteins with an engineered affinity tag (such as a 6xHis tag), it can also be used for natural proteins with an inherent affinity for divalent cations (Block et al., 2009).

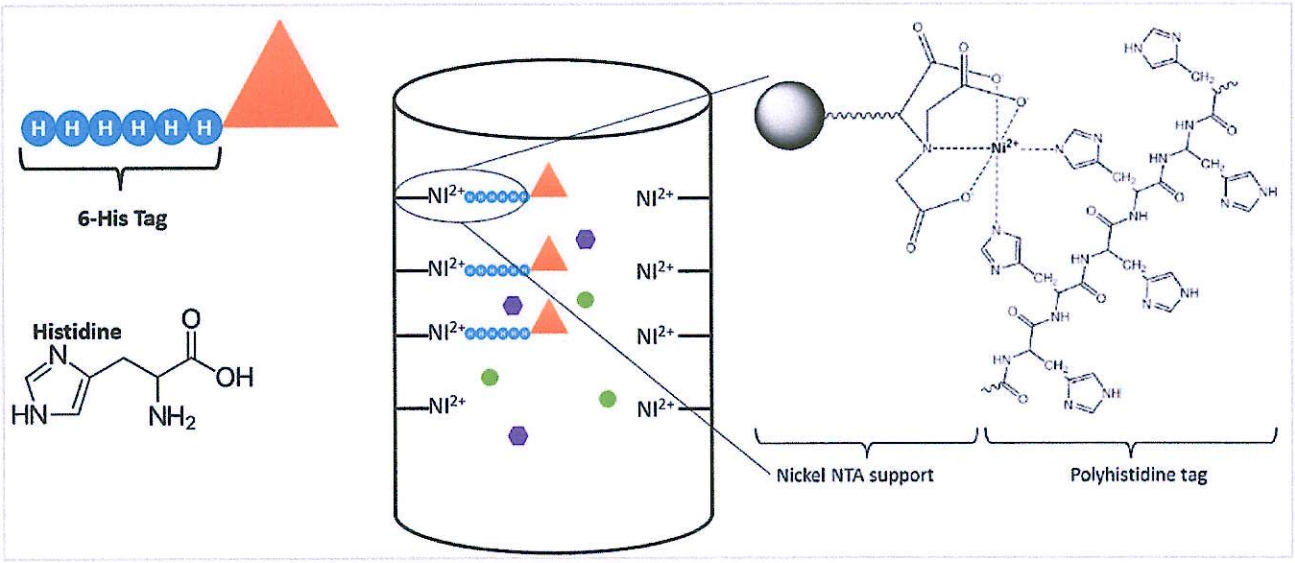


Figure 7. Immobilized-Metal Affinity Chromatography (IMAC). Inspired by Block et al. (2009)

II. 3. 4 Hydrophobic Interaction Chromatography

In Hydrophobic Interaction Chromatography (HIC), a matrix-like silica (which is very polar with exposed OH groups) is modified with ester or ether links connecting the silica surface hydroxyl OHs to nonpolar molecules, typically containing 8 or 18 carbons in the acyl or alkyl chain. Proteins with exposed hydrophobic regions preferentially bind to the beads. Adding high salt concentrations to the aqueous solution enhances the interactions between the proteins and the modified beads, making water more polar. This shifts the equilibrium towards binding the surface-exposed nonpolar regions of the protein to the nonpolar C8 or C18 chains (Queiroz et al., 2001). Proteins are then eluted to increase hydrophobicity by reducing the ionic strength of the buffer, as depicted in Figure 8. The column matrix, shown in grey, has a hydrophobic ligand covalently attached. In high salt conditions (salting out), proteins will bind to the matrix with differing affinity, with more hydrophobic proteins (shown in yellow) binding more tightly than more hydrophilic proteins (shown in blue). When the salt concentration is decreased, more hydrophilic proteins will be released first, followed by more hydrophobic proteins.

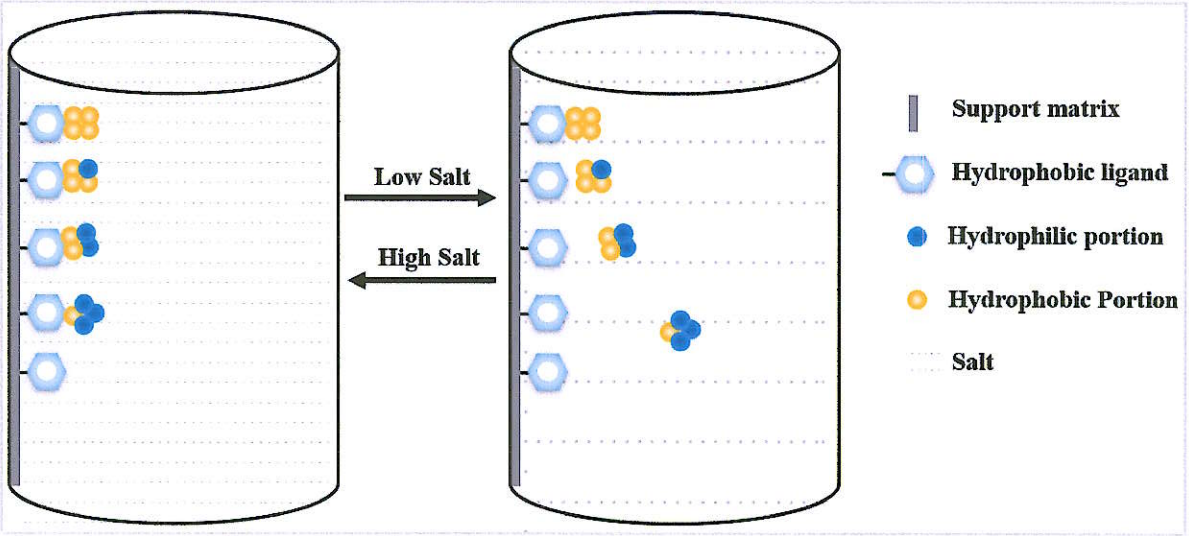


Figure 8: Hydrophobic Interaction Chromatography. Inspired by Queiroz et al., (2001)

II. 3. 5 High-Performance Liquid Chromatography (HPLC)

The initial efforts to separate peptide mixtures using HPLC began in the mid-1970s, around a decade after HPLC was first introduced. It quickly became evident that separating peptides on non-polar stationary phases had significant potential (Sandra et al., 2008). HPLC is a chromatographic technique that uses high pressure to push solutes through the column more quickly than the gravity-driven flow of solvent. The column is packed with tiny, tightly packed beads, which reduces diffusion and significantly enhances resolution. Due to the dense packing of these small beads, the flow would not occur without an external pump. The most common type of HPLC is reversed-phase HPLC (RPLC), where the column's packing material is hydrophobic (Sandra et al., 2008). Proteins are eluted using a gradient of water and increasing concentrations of an organic solvent, such as acetonitrile combined with TFA or formic acid (compatible with MS analysis), with proteins eluting based on their hydrophobicity. Post-HPLC purification, the protein is in a solution with only volatile compounds, making it easy to lyophilize. However, HPLC often denatures the purified proteins, making it unsuitable for proteins that do not refold spontaneously.

II. 4 Protein enrichment

To enrich protein extracts in molecules of interest, it will be necessary to interpose a fractionation step between extraction and separation. This step generally uses approaches derived from classical chromatography and electrophoresis methods. In theory, all the techniques used for protein chromatography can be transposed to form a prefractionation step

in proteomic analysis. The main limitations to their use are scale (proteomics often uses small volumes of valuable biological samples) and the salinity of the buffers used during separation. Protein extracts can be fractionated according to physico-chemical criteria (charge, hydrophilicity, etc.), with each fraction then analyzed individually (Palii et al., 2007). Protein separation from a crude extract can be carried out by reverse phase chromatography, allowing the different fractions to be obtained and analyzed by 2D-PAGE after evaporation of the organic solvent used for elution. Ion exchange chromatography is less suitable due to the use of high-salinity buffers.

Another approach is to select a family of proteins of interest (sub-proteome): a "glyco-proteome" or a "phospho-proteome" (Fíla & Honys, 2012). Micro-purification methods using magnetic beads (Fig.9) also enrich proteins of interest by immunocapture (Fredolini et al., 2016).

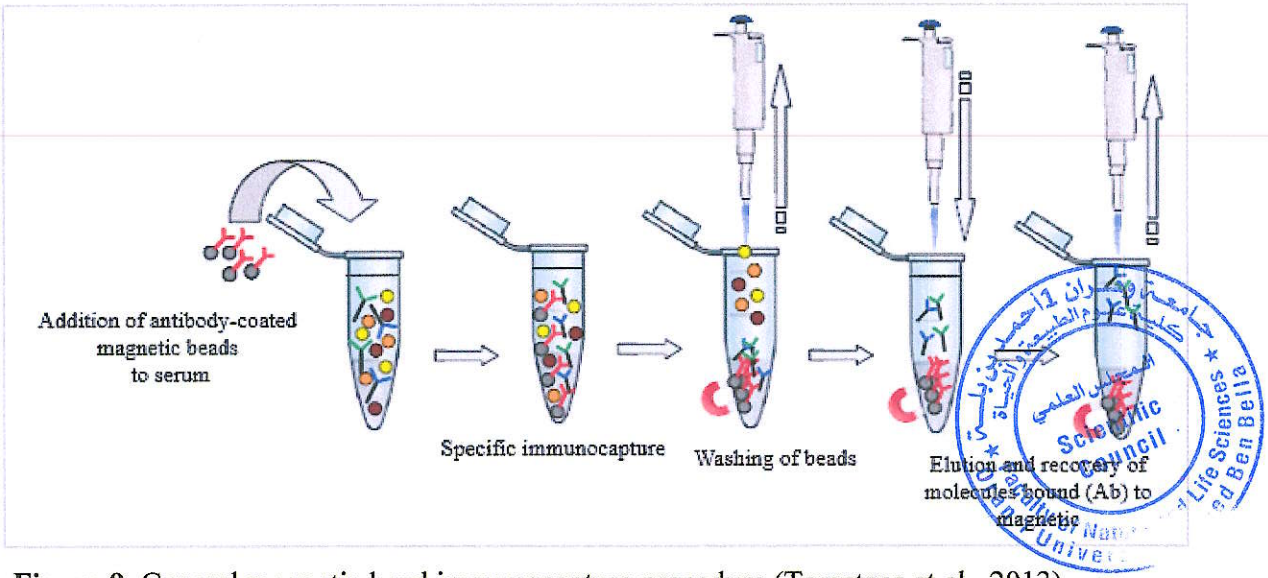


Figure 9. General magnetic bead immunocapture procedure (Torsetnes et al., 2013).

II. 5 Purification assessment

A quantitative system is necessary for measuring the total quantity and concentration of target and total proteins at each stage of protein purification. It is also required to evaluate the target protein's biological activity and overall purity. This strategy aids in directing and refining the purification process under development. Inefficient separation processes can be replaced with ones that generate more protein or preserve biological activity. Therefore, every step in the purification process is quantitatively assessed based on the following parameters: Total protein, Total activity, Specific activity, Yield, and Purification level (Burgess, 2009).

- ***Total Protein (mg)*** is calculated by measuring the concentration in a fraction of the sample and multiplying that by the total sample volume. These days, a Bradford dye-binding assay or a bicinchoninic acid (BCA) assay is most commonly used.

- ***Total Activity (units)*** is measured as the enzyme activity within the assay multiplied by the total volume of the sample. Note that the enzyme unit, or international unit for the enzyme (symbol U), describes the enzyme's catalytic activity. 1 U (µmol/min) is defined as the amount of the enzyme that catalyzes the conversion of one micromole of substrate per minute under the specified conditions of the assay method.

- ***Specific Activity (units/mg)*** is measured by dividing the Total Activity (units) by the Total Protein (mg) to give specific activity as units/mg.

- ***Overall yield (%)*** Yield measures the biological activity retained in the sample after each purification step. The amount in the first step is set to be 100%. All subsequent yield steps will be evaluated using the first purification step. It is calculated by dividing the current step's total activity by the first step's total activity and multiplying it by 100.

- ***The purification level*** evaluates the purity of the protein of interest by dividing the specific activity calculated after each purification step by the specific activity of the first purification step. This is merely setting the initial purity at a value of one and then giving the purity at each step relative to that of the first step. It is essential to make a constant compromise between these two criteria constantly. Indeed, it is crucial not to lose too much biological activity while enriching the fractions obtained with the desired protein. A high level of purification with a low yield leaves little protein for further experimentation.

III. SEPARATION AND ANALYSIS

Both bottom-up and top-down proteomics approaches depend entirely on employed separation technologies to provide large-scale proteome coverage in a given time, accomplish higher analytical throughput, and cover a broad dynamic protein concentration range, including trace amounts of distinct proteins. Two major separation approaches are available in proteomics. Gel-based applications include one-dimensional and two-dimensional polyacrylamide gel electrophoresis. Several gel-free high-throughput screening technologies for protein analysis are equally available, including multidimensional protein identification technology (MudPIT) (Yaoyang Zhang et al., 2013).

III. 1 Electrophoresis

Developed respectively in the mid-1960s and 1970s, one-dimensional and two-dimensional electrophoresis (1DE and 2DE) are the most commonly used techniques for analytical and preparative protein separations. The most common uses are for analytical separations. Electrophoresis refers to the movement of charged particles in an electric field. It is generally performed on an acrylamide gel, on which proteins migrate according to their mass and/or isoelectric point. Also, the movement of a charged protein within a static matrix under an external electric field is influenced by its size and shape (Suárez-Díaz, 2022).

III. 1. 1 Polyacrylamide Gel Electrophoresis (PAGE)

Discovered around the 1950s, the separation medium known as the matrix is primarily composed of polymerized acrylamide (Raymond & Weintraub, 1959). It quickly became widely used in biology due to several advantageous properties: it is optically transparent even under UV light, electrically neutral, and allows control over the gel pore size. Polymerization of acrylamide is initiated by the additions of ammonium persulfate in the presence of tetramethylene-diamine (TEMED), along with a dimer of acrylamide (N, N'-methylene-bis(acrylamide) connected covalently between the amide nitrogens of the acrylamides by a methylene group (Bonaventura et al., 1994; Hwang et al., 2012). The structures of these compounds are shown in Figure 10.

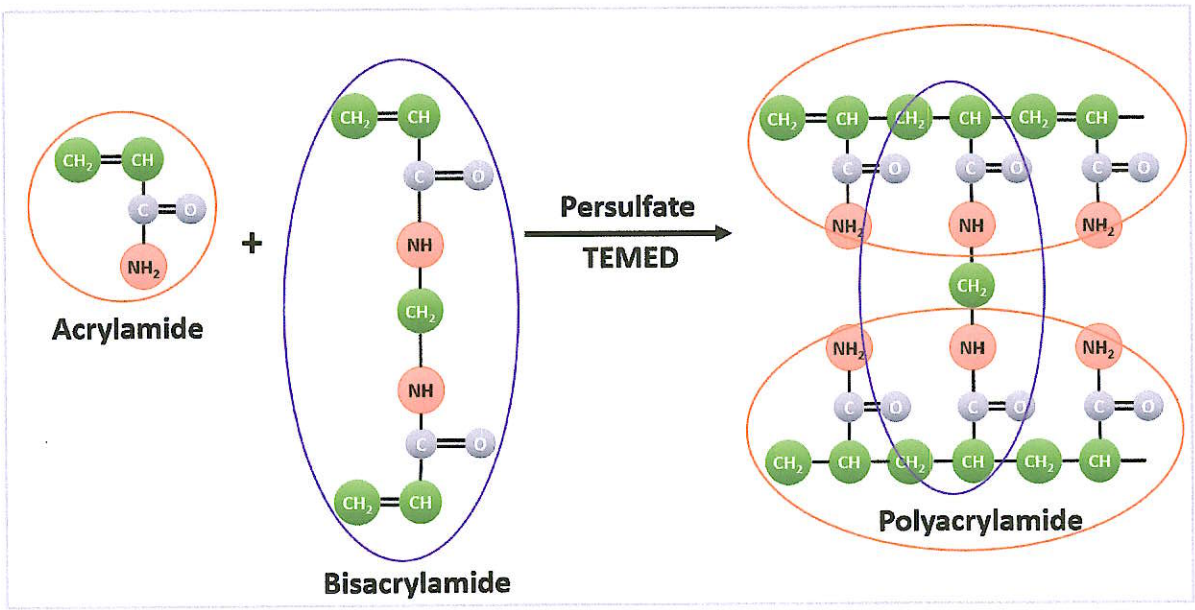


Figure 10. Polymerization of acrylamide. Inspired by Hwang et al. (2012)

The mesh density is low at low acrylamide concentrations, resulting in poor separation of small proteins. This concentration is ideal for separating high-molecular-weight molecules. Conversely, the pore size is so tiny at high acrylamide concentrations that high-molecular-weight molecules can barely move. This concentration is suitable for separating low-molecular-weight molecules (Smith, 1994).

A viscous protein solution is placed at the top of the gel in a small well formed during the gel's polymerization process. The top and bottom sections of the gel are immersed in reservoirs containing a buffered solution and the corresponding electrode. The proteins move through the hydrated gel when an electric field is applied. The buffer solution's composition in the reservoir and the gel is crucial; its components must not bind to the separated proteins. Furthermore, for native (non-denatured) gels, the medium's pH must ensure that the proteins carry the appropriate charge to migrate in the desired direction (Hwang et al., 2012).

Numerous variations of electrophoresis are commonly used. Gels can be polymerized with or without denaturing agents. Some gels have a continuous gradient of acrylamide concentrations. In other cases, a single slab may consist of two distinct layers polymerized on each other, each with different acrylamide concentrations and pH levels. The top layer is the stacking gel, while the bottom layer is the running gel (Schägger & von Jagow, 1987).

In discontinuous gels, the top part of the gel is a low-concentration acrylamide (2-4%), often in a Tris HCl buffer solution (pH 6.5). The lower part of the gel is 8-15% acrylamide, depending

on the choice of gel, which is selected based on the molecular weight of the proteins to be separated. The upper buffer reservoir contains Tris-buffered with a weak acid such as glycine (pKa2 = 9.6) to the same pH as the running gel (Fig.11). When electrophoresis begins, glycine ions from the upper reservoir (pH 8.8) move into the stacking gel, carrying a partial negative charge. Upon entering the stacking gel (pH 6.8), they become zwitterions with no net charge and stop moving, increasing the gel's electrical resistance due to fewer moving ions. Proteins stack into a thin disc behind the Cl- ions, which move faster due to their high charge density and mobility. At the interface between the stacking and running gels, proteins separate based on the sieving effects of the denser running gel. Glycine eventually enters the running gel, regains its full charge (pH 8.7), overtakes the proteins, and restores the altered charge balance in the stacking gel (Hwang et al., 2012).

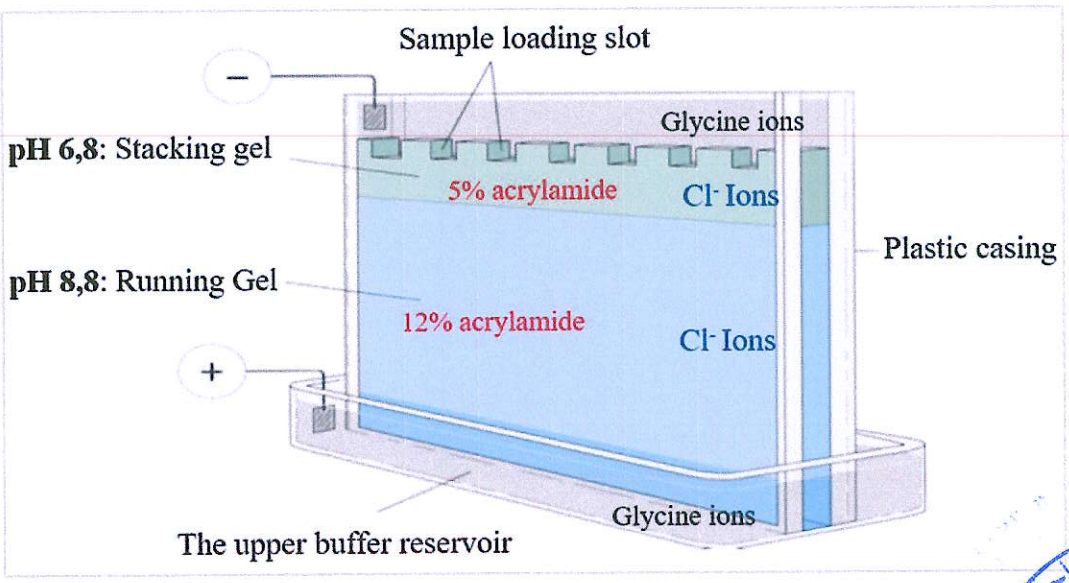


Figure 11. Electrophoresis discontinuous gel and chamber. Inspired by Hwang et al. (2012).

III. 1. 2 Detection of proteins in the gel

Before the proteins are loaded onto the gel, they are mixed with bromophenol blue to stop them from eluting into the lower buffer reservoir. Electrophoretic migrations are carried out until the bromophenol blue tracking dye runs off the gel (Izzo et al., 2006).

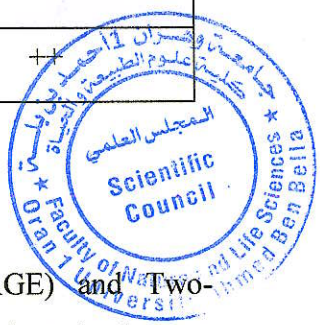
The staining step reveals the proteins separated during electrophoresis. Several staining methods are possible, the three most common being Coomassie blue, silver nitrate, and fluorescence staining (Chevalier et al., 2006).

Each stain has unique characteristics, and the choice depends on the application (Tab.3). Important parameters to consider include detection sensitivity (detection at the µg or ng level), staining homogeneity (depending on the protein type), linearity (proportionality with the amount of protein), and compatibility with subsequent mass spectrometry analysis.

Coomassie Brilliant Blue dye is the most commonly used stain in laboratories. When proteins bind to this dye, there is a corresponding shift in the dye's absorbance properties. The methanol and acetic acid in the dye solution also help fix the gel's proteins, preventing them from diffusing into the solution. After staining, the background dye in the gel is removed with an acetic acid/methanol solution, leaving blue-colored protein bands (1D) or spots (2D). There are several different systems available for the digitization of gel images for analysis of electrophoretic separations of protein (Wang et al., 2023).

Table 3. Parameters to Consider for PAGE Gel Stains.

Stains / Parameters	**Coomassie Brilliant Blue**	**Silver staining**	**Fluorescent dyes**
Sensitivity	+	+++	++
Homogeneity	+++	++	++
Linearity	++	+	+++
MS Compatibility	++	+	++



III. 1. 3 Variations on polyacrylamide gel electrophoresis

Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Two-dimensional gel electrophoresis (2D-PAGE) are two types of gel electrophoresis that are commonly used in proteomics analysis because they solve two problems at once: they fractionate the proteins based on apparent molecular mass (SDS-PAGE) or charge and apparent molecular mass (2D-PAGE) and remove salts, detergents, and other small molecules from the sample (Rabilloud & Lelong, 2011). The advantages of SDS-PAGE are its low cost, ease of use, and great reliability. Despite being higher than size exclusion chromatography, its separation resolution is not as great as 2D-PAGE. However, most proteoform studies favor 2D-PAGE because of its excellent resolution and greatly increased reproducibility since the development of immobilized pH gradient (IPG) strips (Fig.12).

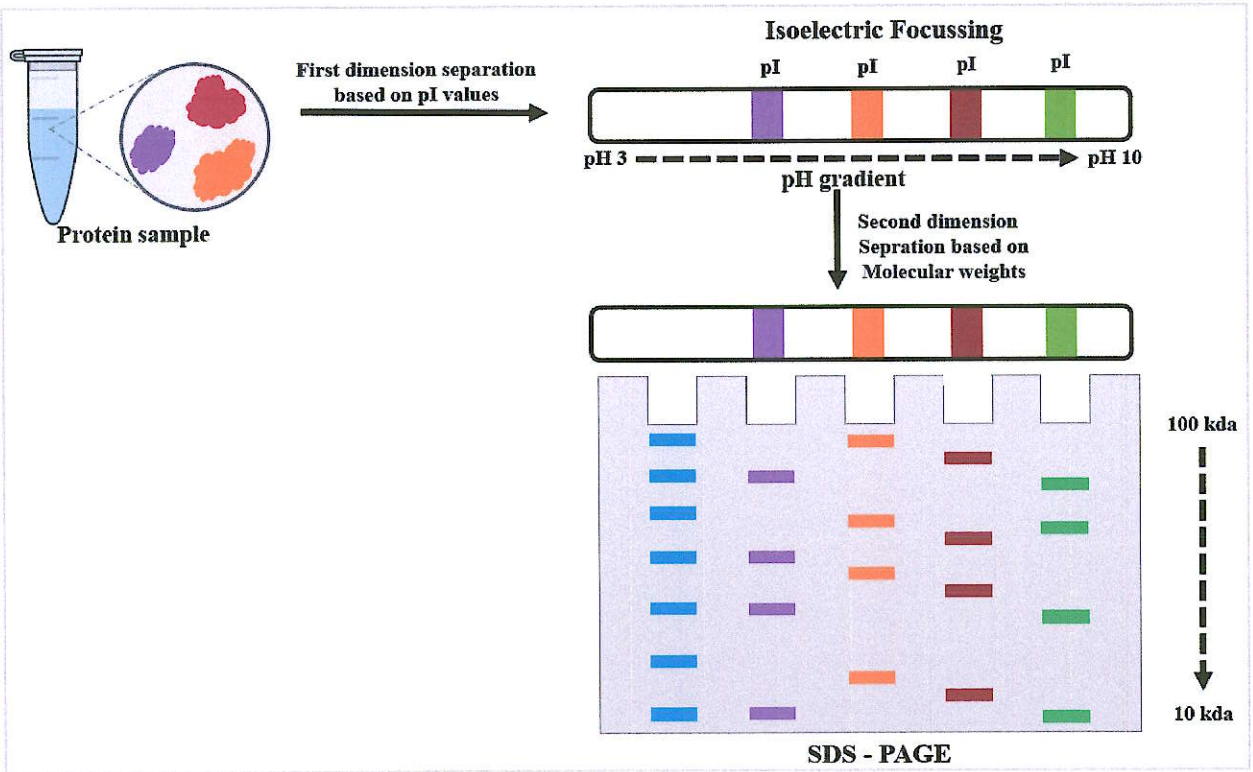


Figure 12. Two-dimensional gel Electrophoresis. Rabilloud & Lelong (2011)

There have been numerous publications and developments in 2D gel electrophoresis since O'Farrell and Klose first devised the technique (Klose, 1975; O'Farrell, 1975), although the principle remains the same. A combination of detergents (such as SDS and CHAPS) and chaotropes (such as urea and thiourea) has generally been used for solubilization and denaturation of protein samples before separation on the 2D gel. Proteins are first separated according to their isoelectric point (pI) by isoelectric focusing (IEF) and secondly according to their molecular mass by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Gorg, 1991). The first-dimension separation was initially performed in polyacrylamide gels cast in narrow tubes with carrier ampholytes. To overcome the limitations of carrier ampholyte-generated pH gradients, immobilized pH gradients (IPGs) were published for the first time in 1982 (Bjellqvist et al., 1982). Since the reactive end is co-polymerized with the acrylamide matrix, extremely stable pH gradients are generated, increasing reproducibility and expanding the basic pH limit of IEF (Görg et al., 1988). The high reproducibility of IPGs makes them useful for developing 2D protein databases (Appel et al., 1999). Proteins with similar pI are separated based on their molecular weight using a sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE) in the second dimension of 2-DE. The electric field strength, sample hydrophobicity, molecule size and shape, ionic strength, and temperature of the electrophoresis

buffer are among the variables influencing protein migration in SDS-PAGE. By employing gradient gels and varying the polyacrylamide fraction, separation can be improved for 2-DE spots containing several proteins with identical molecular weights (Fountoulakis et al., 1998).

III. 1. 4 Differential In Gel Electrophoresis (DIGE)

A modification of two-dimensional (2-D) polyacrylamide gel electrophoresis was described in 1997 to reproducibly detect differences between two protein samples using just one gel (Unlü et al., 1997). This is accomplished by running the two samples on the same 2-D gel, fluorescence imaging the gel post-run to create two images, and then superimposing the images. The two samples are fluorescently tagged with distinct dyes. Regardless of the dye used for tagging, the amine-reactive dyes are made to guarantee that proteins shared by both samples have the same relative mobility.

The CyDye DIGE Three fluorescent cyanine dyes known as fluor minimum dyes (Cy™2, Cy™3, and Cy™5) are used in the 2-D DIGE method for protein labeling (Larbi & Jefferies, 2009). These dyes are matched in mass and charge and have different excitation and emission wavelengths, making them spectrally distinguishable. Therefore, any protein labeled with these dyes will migrate to the same spot on the 2-D gel. This multiplexing methodology minimizes intra-gel variation (Fig.13).

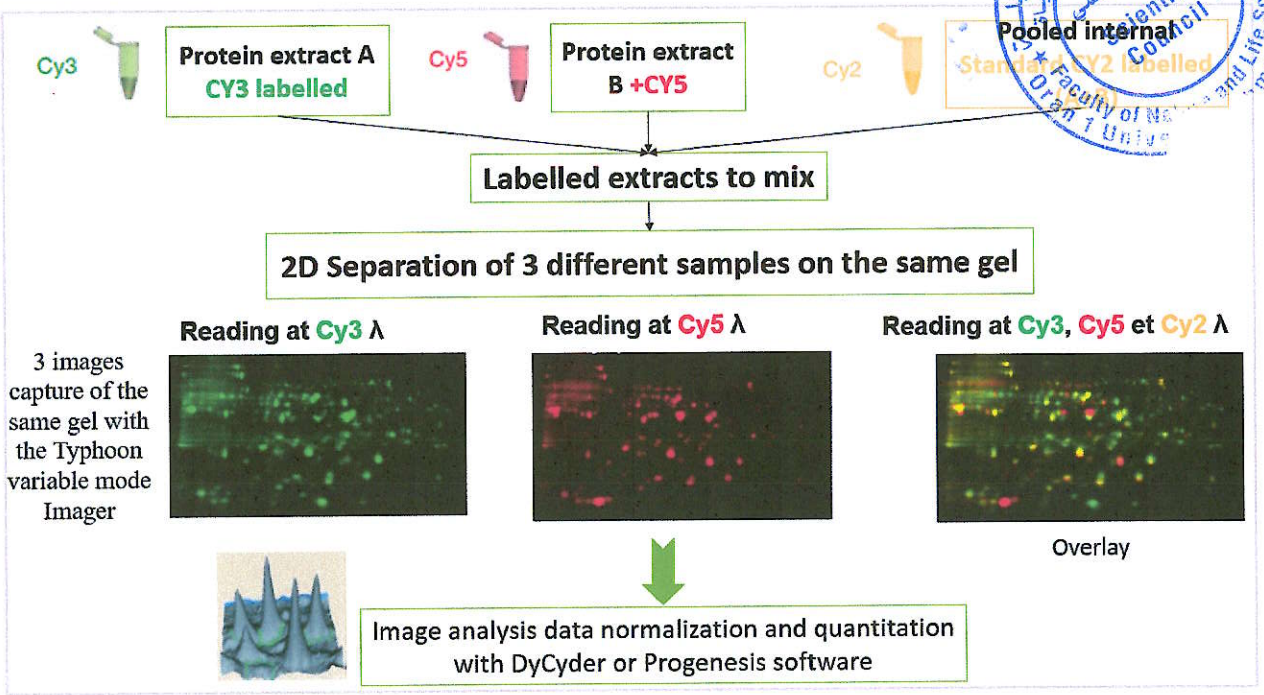


Figure 13. Two-dimensional differential in-gel electrophoresis (2-D DIGE) concept. Inspired by Unlü et al., (1997)

III. 2 Multidimensional Chromatography

While two-dimensional gel electrophoresis (2DE) is a powerful method for protein separation, several limitations have spawned new technologies as alternatives to 2DE. One powerful alternative technique is the so-called MUDPIT (multidimensional analysis of protein identification technology) pioneered by Yates and colleagues (Link et al., 1999).

A complex peptide mixture can be efficiently separated and concentrated before mass spectrometry sequence analysis by utilizing peptides' unique physical properties of charge and hydrophobicity. Because of its large loading capacity, strong cation exchange (SCX) chromatography is usually employed as the major separation technique. On the other hand, reversed-phase (RP) chromatography is a perfect secondary separation technique since it efficiently eliminates salts and may be immediately integrated with mass spectrometry using electrospray ionization. 2D chromatography can be accomplished by either online or offline approach (Fig.14).

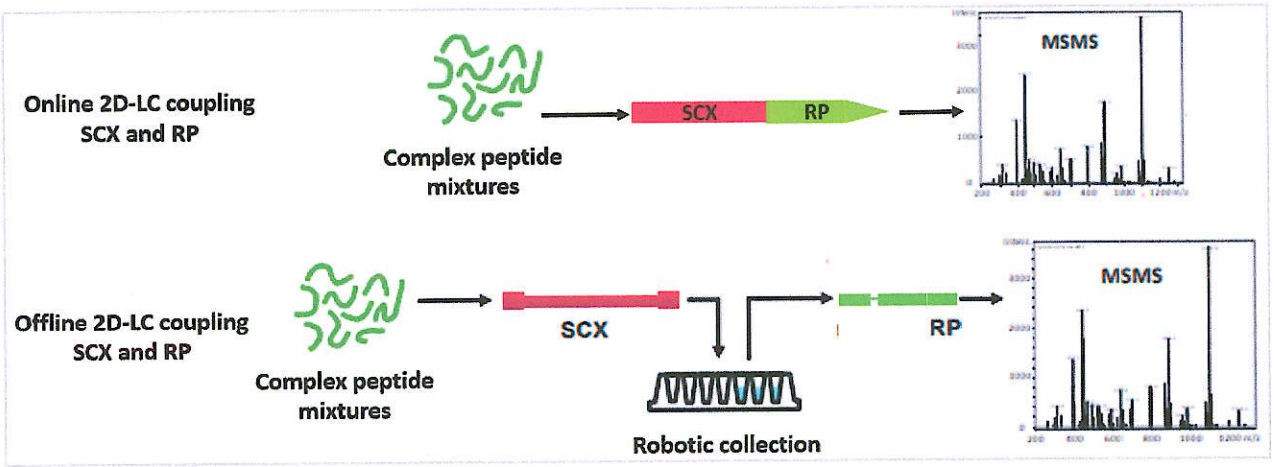


Figure 14. 2D chromatography online and offline approach. Inspired by Link et al. (1999).

The development of LC-MS/MS technologies has reduced 2D gel-based separations, but 1D-SDS-PAGE remains a standard separation method for complex protein mixtures based on their molecular weight (Y. Zhang et al., 2013).

III. 3 Proteolytic Digestion

A protein lysate is a highly heterogeneous mixture of proteins with diverse physicochemical properties. Analyzing proteins from their proteolytic peptides avoids challenges with intact protein separation, ionization, and MS characterization. In a typical shotgun proteomics

experiment, proteins are digested using one or several proteases that generally differ in specificity for cleaving the amide bond before or after one or several specific residues. Commonly used proteases and their biochemical specificities and applications are listed in Table 4.

Table 4. Common proteases used for proteomics (Yaoyang Zhang et al., 2013).

Protease	Cleavage Specificity*	Common proteomic usage
Trypsin	-K,R-↑-Z- not -K,R-↑-P-	General protein digestion
Endoproteinase Lys-C	-K-↑-Z-	Alternative to trypsin for increased peptide length; multiple protease digestion; ^{18}O labeling
Chymotrypsin	-W,F,Y-↑-Z- and -L,M,A,D,E-↑-Z- at a slower rate	Multiple protease digestion
Subtilisin	Broad specificity to native and denatured proteins	Multiple protease digestion
Elastase	-B-↑-Z-	Multiple protease digestion
Endoproteinase Lys-N	-Z-↑-K-	Increase peptide length; create higher charge state for ETD
Endoproteinase Glu-C	-E-↑-Z- and 3000 times slower at -D-↑-Z-	Multiple protease digestion; 18O labeling
Endoproteinase Arg-C	-R-↑-Z-	Multiple protease digestion
Endoproteinase Asp-N	-Z-↑-D- and -Z-↑-cysteic acid- but not -Z-↑-C-	Multiple protease digestion
Proteinase K	-X-↑-Y-	Non-specific digestion of membrane-bound proteins
OmpT	-K,R-↑-K,R-	Increased peptide length for middle-down proteomics

*B – uncharged, non-aromatic amino acids (i.e. A, V, L, I, G, S); X – aliphatic, aromatic, or hydrophobic amino acids; and Z – any amino acid.

Trypsin has become the gold standard for protein digestion for shotgun proteomics (Vincent et al., 2020). This is mainly because trypsin is an exact and effective enzyme that is easily accessible and reasonably priced. It also yields short peptides with a basic Lys or Arg at the C-terminus, making them suitable for peptide fragmentation, search algorithm-based identification, and chromatographic separation techniques today. However, only a small percentage of the proteome is studied because most tryptic peptides are too short to be detected by mass spectrometry, with 56% of the resultant peptides having a length of six residues or less (Tsiatsiani & Heck, 2015).

Before protein digestion, DTT and Iodoacetamide (IAM) are used to reduce and alkylate cysteine (Cys) residues involved in disulfide bonds (Fig.15).

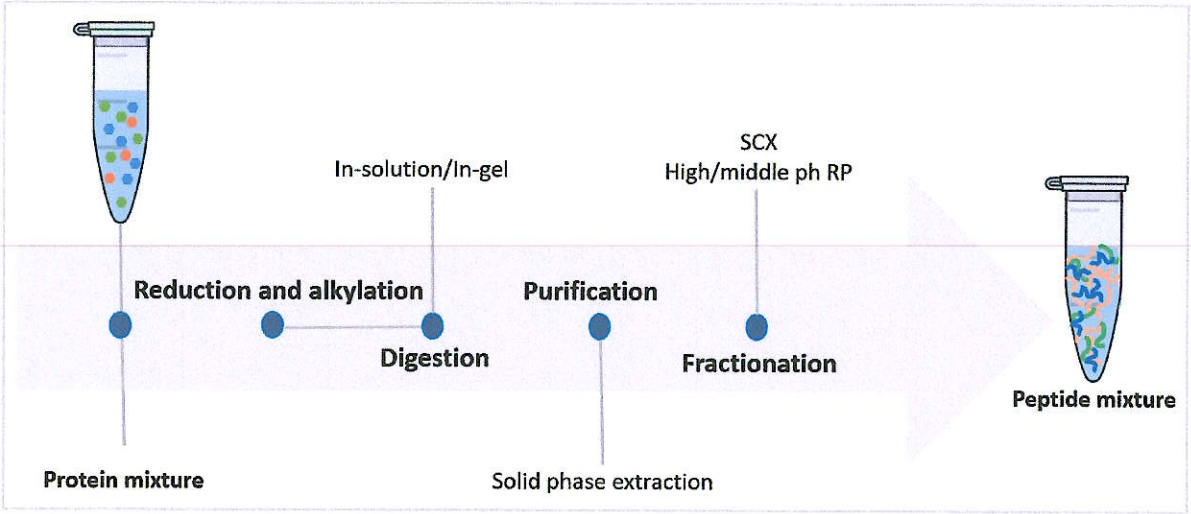


Figure 15. Sample digestion in bottom-up proteomics (Duong & Lee, 2023).

Proteomics began with separating proteins by two-dimensional (2D) gel electrophoresis, followed by *trypsin in-gel digestion*, peptide sequencing, and tandem mass spectrometry (MS/MS) identification. Gel-free or shotgun proteomics quickly gained prominence due to disadvantages associated with a restricted protein dynamic range and inadequate digestion. This method involves *digesting extracted protein mixtures in solution* to produce tens of thousands of peptides, which are subsequently separated (usually via liquid chromatography; LC) and identified using tandem mass spectrometry (MS/MS).

IV. MS ANALYSIS

IV. 1 Principles and Instrumentation

Mass spectrometry has been widely used to analyze biological samples and has become an indispensable tool for proteomics research. It is an advanced technology for complete protein characterization and is extensively used for sequence analysis, studying protein-protein interactions, and identifying post-translational modifications (PTMs) (Aebersold & Mann, 2003).

A mass spectrometer consists of an ion source that converts analyte molecules into gas-phase ions, a mass analyzer that measures the mass-to-charge ratio (m/z) of the ionized analytes, and a detector that registers the number of ions at each m/z value (Fig.16). The real biochemical applications for studying peptides and proteins took off around 1988 with the discovery of "electrospray ionization" (ESI) and "matrix-assisted laser desorption/ionization" (MALDI). The two soft ionization techniques capable of ionizing peptides or proteins revolutionized protein analysis using MS. MALDI seemed as promising as ESI. Fenn and Tanaka jointly received the Nobel Prize in Chemistry in 2002 for "the development of soft desorption and ionization methods for the mass spectrometric analysis of biological macromolecules" (Fenn et al., 1989; Karas & Hillenkamp, 1988). Therefore, ESI ionizes the analytes out of a solution and is readily coupled to liquid-based (chromatographic and electrophoretic) separation tools. MALDI sublimates and ionizes the samples out of a dry, crystalline matrix via laser pulses.

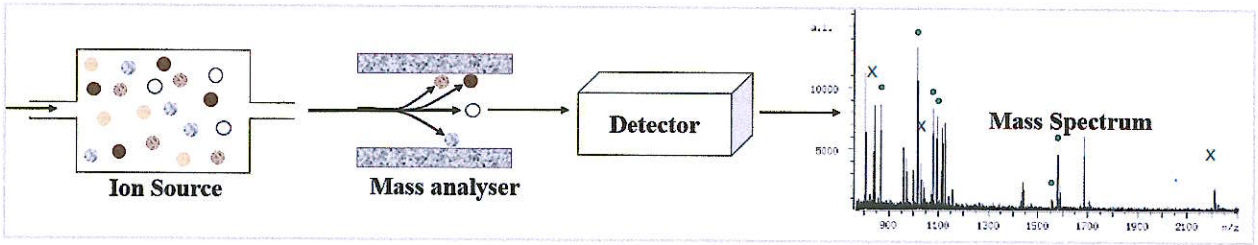


Figure 16. One-stage mass spectrometry (MS) instrumentation. Inspired by Karas & Hillenkamp (1988)

IV. 1. 1 Matrix-Assisted Laser Desorption Ionization (MALDI)

In MALDI, ions are desorbed from the solid phase. A sample is first dissolved in a suitable solvent and mixed with an excessive amount of an appropriate matrix. Subsequently, it is

spotted on a MALDI plate and air-dried (or under a stream of nitrogen gas). Under these circumstances, the sample is co-crystallized with the matrix.

The components in the mixture are brought into the gas phase via a laser beam (usually a nitrogen laser at a wavelength of 337 nm) that hits the sample-matrix crystal, leading to absorption of the laser energy by the matrix and subsequent desorption and ionization of the analytes in the sample (El-Aneed et al., 2009). Three main acidic matrices are commonly employed in proteomic studies, namely: α-Cyano-4-hydroxycinnamic acid (HCCA), 2,5-Dihydroxybenzoic acid (DHB), Sinapinic acid (SA). Figure 17 represents the process by which ions are formed during MALDI-MS.

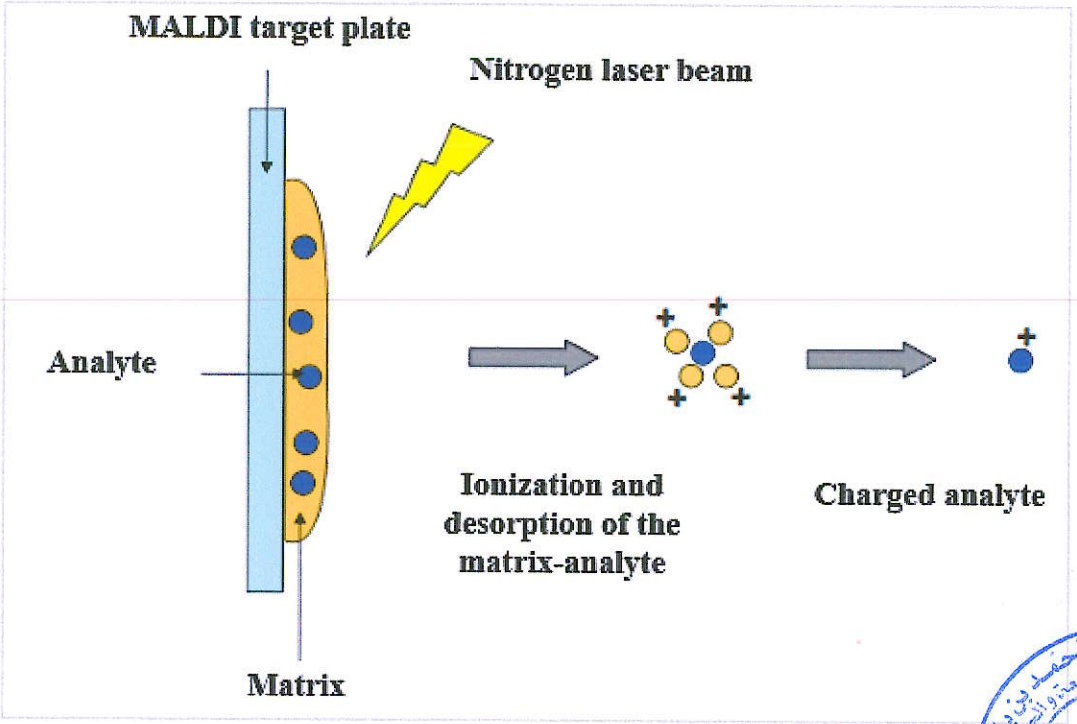


Figure 17. Schematic of the MALDI ionization process. Inspired by El-Aneed et al. (2009).

IV. 1. 2 Electrospray Ionization-Mass Spectrometry (ESI-MS)

In ESI-MS, the sample should be soluble in a preferably polar solvent, which can be infused, under atmospheric pressure, into the ionization source via a thin needle. As the sample is being constantly sprayed, a high electrical potential is applied at the needle (2–5 Kv), providing the electric-field gradient required to produce charge separation at the surface of the liquid (Fig.18). As a result, the liquid protrudes from the capillary tip in what is known as a 'Taylor cone'. These droplets are then electrically driven and vaporized with a warm neutral gas (usually

nitrogen). Under these conditions, the droplets break down, and their size is continuously reduced while they shift inside the source. Eventually, the repulsive forces, also termed the coulombic forces, among the ions on the surface of the shrinking droplets become very high. These forces will ultimately exceed the surface tension of the solvent, resulting in ions that desorb into the gas phase (El-Aneed et al., 2009).

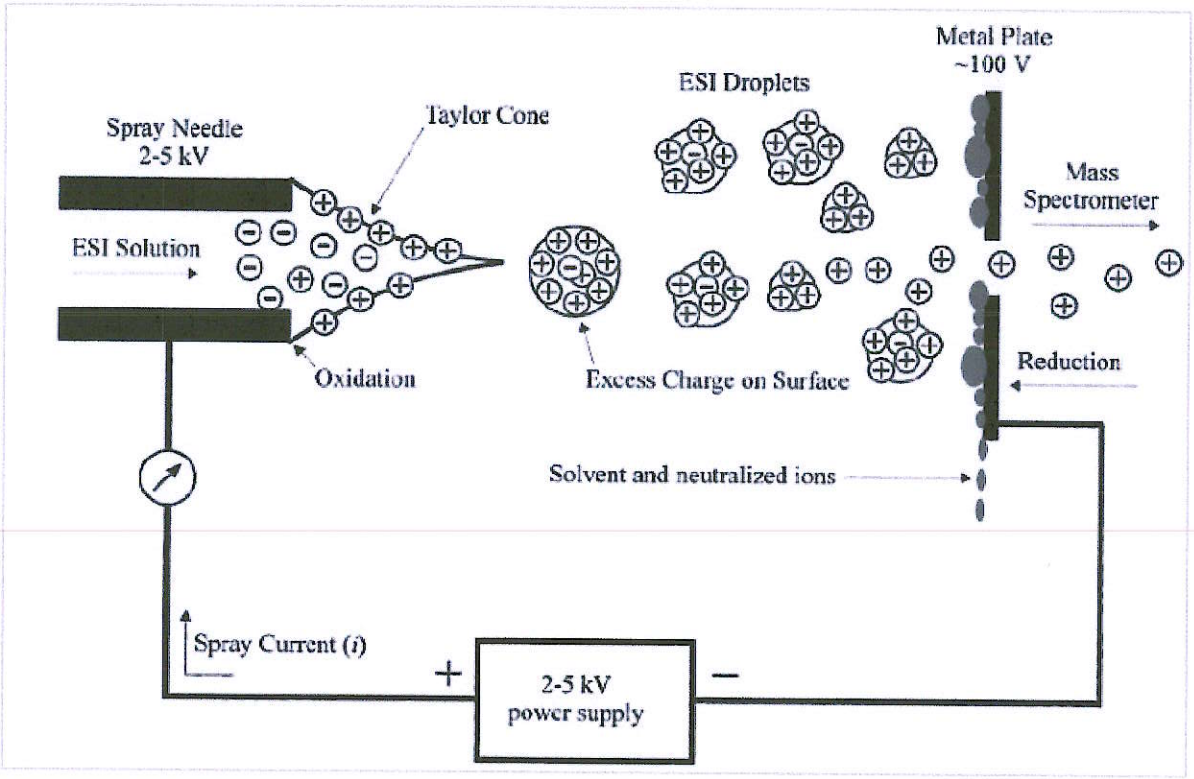


Figure 18. Diagram of the electrospray ionization process (Cech & Enke, 2001).

IV. 1. 3 Mass Analyzer

A mass analyzer is the part of the instrument in which ions are separated based on their m/z values. For proteomics research, four types of mass analyzers are widely used by mass spectroscopists, namely: quadrupole (Q), quadrupole ion trap (QIT), time-of-flight (TOF) mass analyzer (Fig.19), and Fourier-transform ion cyclotron resonance (FTICR) mass analyzer. They vary in their physical principles and analytical performance. 'Hybrid' instruments have been designed to combine the capabilities of different mass analyzers and include the Q–q–Q, Q–q–LIT, Q–TOF, TOF–TOF, and LTQ–FTICR (Han et al., 2008).

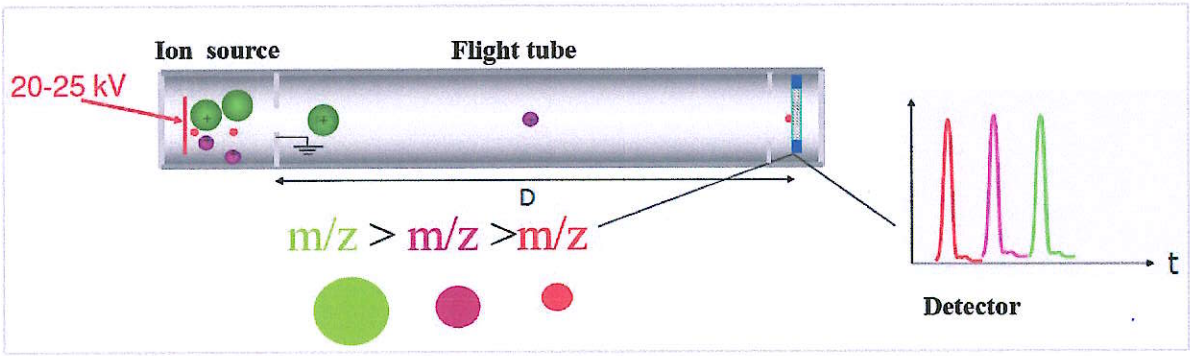


Figure 19. Schematic representation of a conventional ToF (time-of-flight) analyzer. Inspired by Han et al. (2008)

IV. 1. 4 Tandem Mass Spectrometry (MS/MS)

The development of single-stage ESI-MS and MALDI-MS, which generate molecular or quasi-molecular ions and give very little structural information, has inspired most efforts to develop tandem mass spectrometry as a structural tool. Tandem mass spectrometry is the branch of mass spectrometry concerned with the selection of a particular ion (a component of the average mass spectrum, i.e., precursor ion) formed from a molecule or a mixture of molecules and its activation—usually by one or more collisions—to generate characteristic secondary fragment ions. Multiple mass analyzers can be connected in a series so that a specific ion is selected in the first mass analyzer (MS1) and then subjected to collision in the collision cell with a neutral gas, where a fraction of the ion's kinetic energy is transferred into internal energy, effecting its dissociation into various fragment ions. The second analyzer (MS2) separates the product ions before reaching the detector (Shukla & Futrell, 2000). Figure 20 illustrates an MS/MS instrumentation.

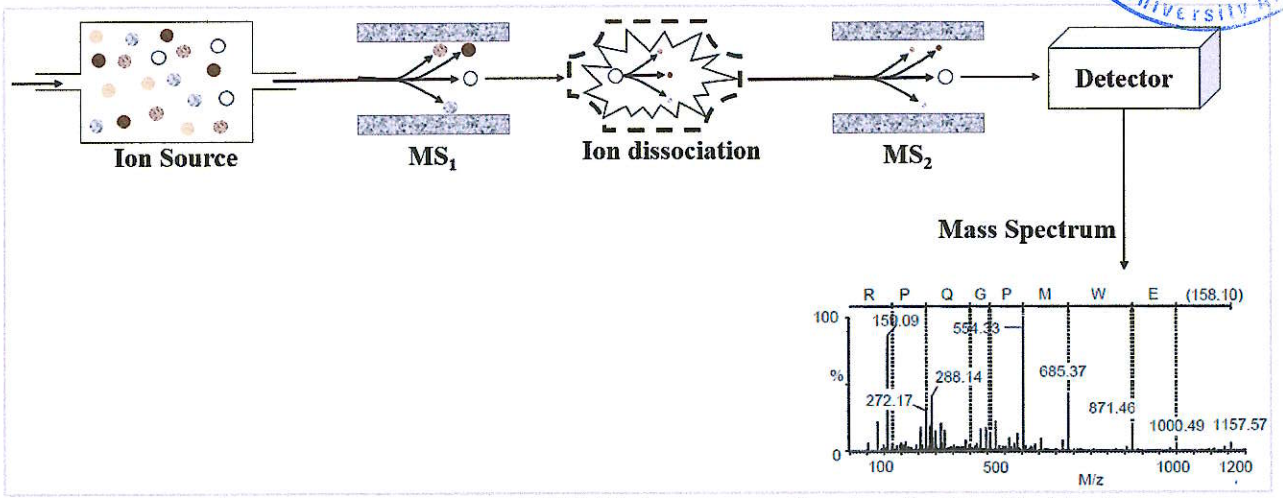


Figure 20. Tandem mass spectrometry (MS/MS) instrumentation. Inspired by Shukla & Futrell (2000)

IV. 2 Protein Identification by Mass Spectrometry

The two main ionization methods (ESI and MALDI) are used for charging and transferring peptides into the gas phase in shotgun proteomics. ESI has been generally coupled with RPLC separations and MALDI with gel-based separations. MALDI is often performed by transferring proteins or peptides from a gel to a support or substrate. However, methodologies were developed to perform MALDI directly from 2D gels to simplify the process and improve reproducibility (Yaoyang Zhang et al., 2013). MALDI-MS typically analyzes relatively simple peptide mixtures, whereas integrated liquid-chromatography ESI-MS systems (LC-MS) are preferred for analyzing complex samples (Fig.21 A & B).

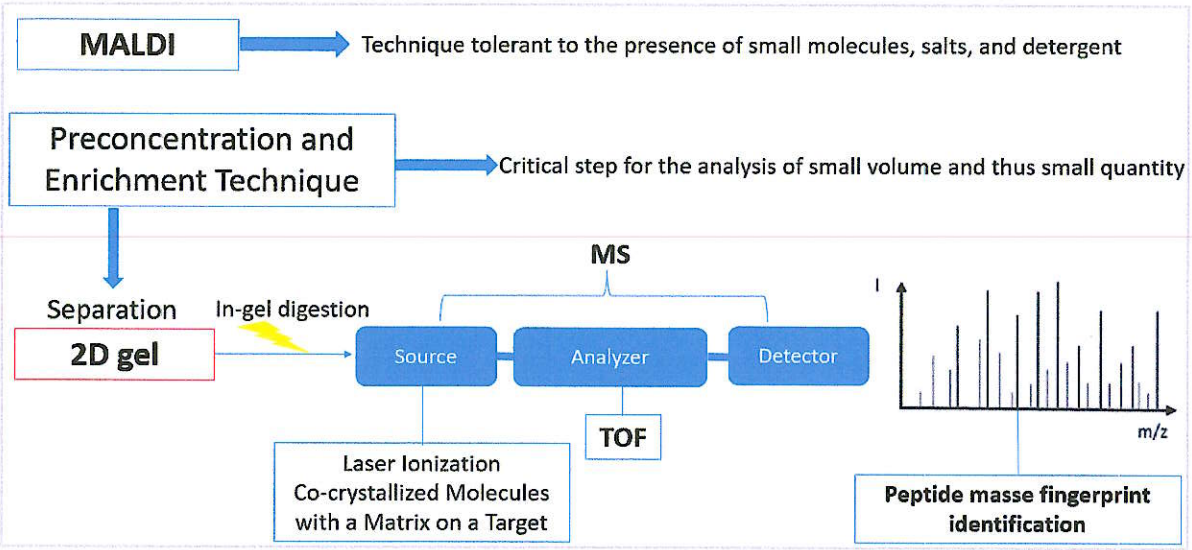


Figure 21 A. Schematic diagram of MALDI-TOF. Inspired by Yaoyang Zhang et al. (2013)

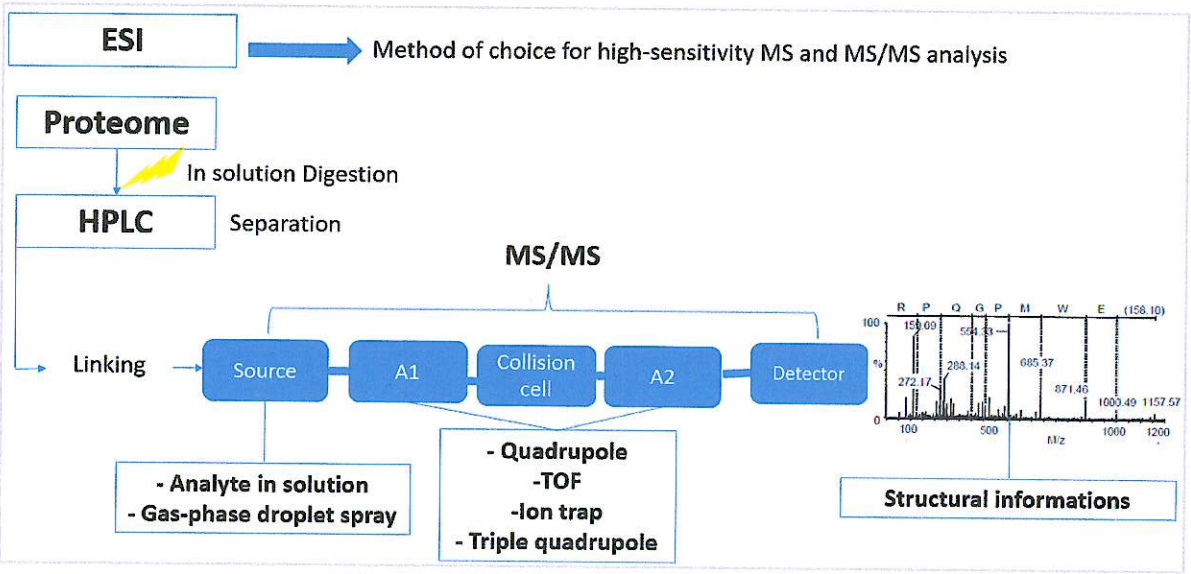


Figure 21 B. Schematic diagram of ESI MS/MS. Inspired by Yaoyang Zhang et al. (2013)

IV. 2. 1 Peptide Mass Fingerprint (PMF)

There has also been a significant crossover of separation and ionization methods in both the application and development of new ionization strategies. Matrix-assisted laser desorption-ionization-time-of-flight (MALDI-TOF) mass spectrometry (MS) is most often employed as the initial technique in a hierarchical approach to protein identification (Aebersold & Mann, 2003). Direct analysis of the m/z values of peptides resulting from a digest yields a set of peptide masses, which can be regarded as a unique mass map, a so-called peptide mass fingerprint (PMF) (Blueggel et al., 2004).

These peptide masses are then used to search against a computer-generated list formed from the simulated digest of a protein database with the same enzyme (Fig.22).

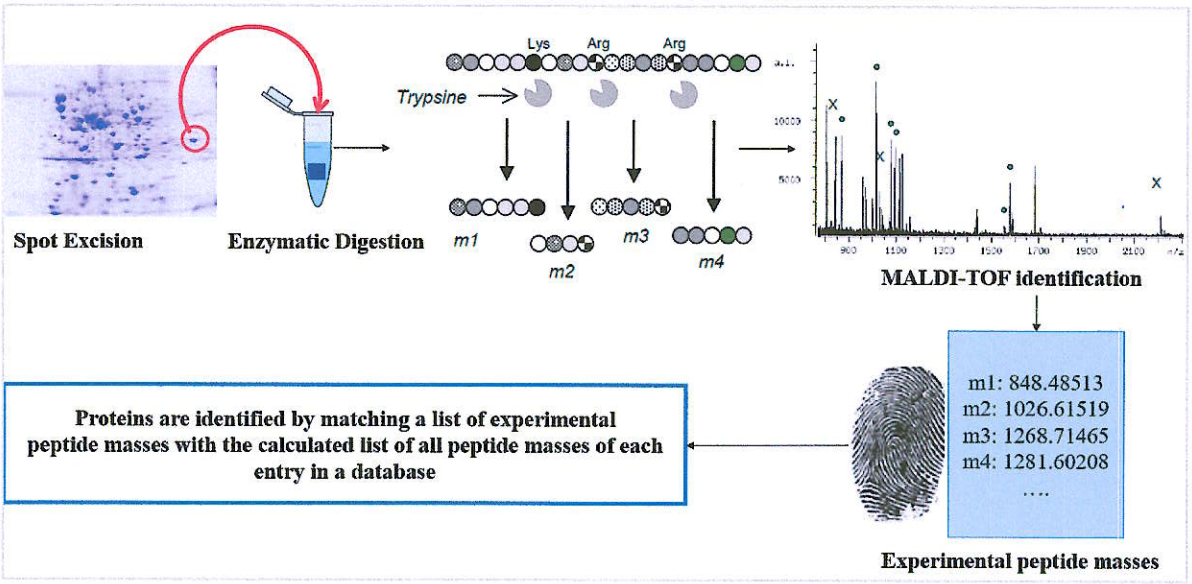


Figure 22. Protein identification by Peptide-mass fingerprinting. Inspired by Blueggel et al. (2004)

Numerous online services are available for protein identification (Tab.5). The ExPASy service of the Swiss Institute of Bioinformatics (http://www.expasy.ch) provides the largest selection of Internet tools and the annotated SwissProt sequence database. Additional navigation entry points, particularly for MS analyses, are CompSearch, ProteinProspector, and Prowl (Vihinen, 2001).

Table 5. Peptide-mass fingerprinting softwares and corresponding URL Addresses.

Program Name	URL
ProFound	http://prowl.rockefeller.edu/profound_bin/WebProFound.exe
Mascot	http://www.matrixscience.com
MS-Fit	http://prospector.ucsf.edu/
Mowse	http://srs.hgmp.mrc.ac.uk/cgi-bin/mowse
MSA	http://www.molgen.mpg.de/~mass-spec/ASMS-Conference/msaProgram.htm
PepMapper	http://wolf.bms.umist.ac.uk/mapper/
Peptident	http://us.expasy.org/tools/peptident.html
PeptideSearch	http://www.mann.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html

MS tools on the web can be grouped as methods for the analysis of protein molecular weight calculation and protein sequence tags (PeptideSearch, TagIdent, PeptideMass); peptide mass fingerprinting (MassSearch, MS-Fit, PeptideSearch, Profound, MS-Digest, Mascot, MSProduct, PeptIdent); fingerprinting with sequence tag information (MSEdman, MOWSE, Mascot); peptide sequence tags (MS-Tag, PepFrag, PeptideSearch, MS-Seq); amino acid composition (AACompIdent, PropSearch, MultiIdent, AACompSim, MS-Comp); and other tools for interpretation of MS data (Amino Acid Information, Compute pI/MW, MSDigest, PeptideMass, GlycoMod, MS-Isotope, FindMod) (Vihinen, 2001).

IV. 2. 2 Protein Identification Using Tandem Mass Spectrometry

Several peptide masses from the same protein are required for unambiguous identification at the mass accuracy achieved with the MALDI-TOF mass spectrometers frequently used for peptide mass measurement (10-100 ppm). In contrast, the amino acid sequence of even a relatively small peptide can uniquely identify a protein. Large-scale protein identification, therefore, critically depends on tandem mass spectrometry for generating sequence-specific spectra for peptides (Nesvizhskii, 2007). Typically, but not exclusively, these mass spectrometers are used in conjunction with ESI. Fragment ion spectra are generated by collision-induced dissociation (CID). The peptide ion to be analysed is isolated and fragmented in a collision cell, and the fragment ion spectrum is recorded. The low-energy CID spectra of peptides generated by ESI-MS/MS are relatively simple to interpret, and a straightforward nomenclature for annotating the MS spectra has been adapted (Wells & McLuckey, 2005). CID fragmentation breaks peptide bonds (between the C-terminal of one amino acid and the N-terminal of the following amino acid), thereby forming b and y ions (Fig.23). The mass difference between two consecutive peaks in the spectrum is therefore equal to the mass of one amino acid.

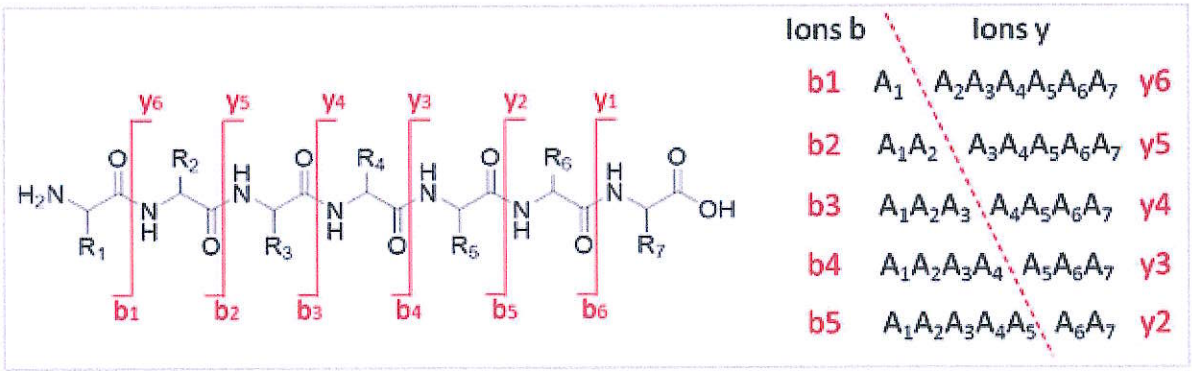


Figure 23. Peptide fragment ion nomenclature (Biemann, 1990).

A. Peptide fragments fingerprint

The other type of MS data is generated by tandem mass spectrometry (MS/MS). Many different types of mass spectrometers can select and isolate peptides as precursor ions and subject the ions to collision-induced (CID) or post-source decay (PSD). The masses of resulting peptide fragments are measured. As fragmentation occurs mainly in the peptide bond of the amino acid chain backbone, a ladder of sequence ions is generated. The resulting peptide fragmentation masses differ by the mass of amino acid residues, thus allowing stretches of the peptide sequence to be deduced. Therefore, the following peptide fragmentation data is called peptide fragmentation fingerprint (PFF) (Blueggel et al., 2004).

Peptide fragmentation spectra can be correlated to protein sequences in a database. This strategy is called peptide fragmentation fingerprinting, similar to peptide mass fingerprinting. The proteins in these databases are digested *in silico* according to the digestion rules of the enzyme used, and the peptides are then fragmented *in silico* according to the fragmentation mode used. The masses of the experimental peptides and fragments are then compared to those of the theoretical peptides and fragments (Fig.24) using a search algorithm. In contrast to peptide maps, which represent an overview of the whole protein, peptide fragmentation provides primary structure information on a single peptide. The more fragments a peptide is identified with, the more robust the identification (Aebersold & Mann, 2003).

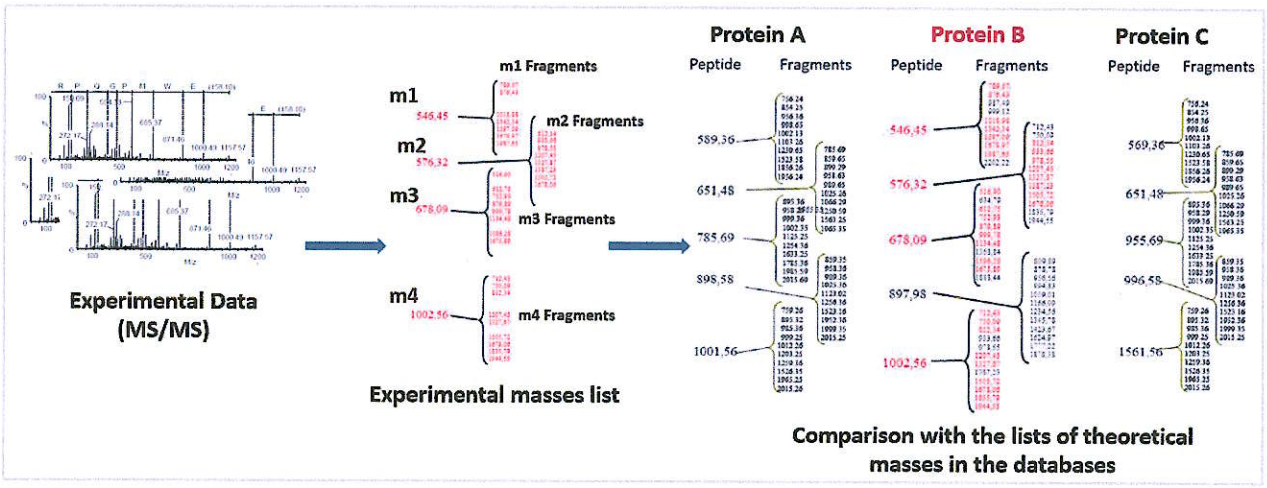


Figure 24. Schematic representation of database search by peptide fragment fingerprinting.

Table 6 displays the commonly used software tools for identifying peptides and proteins through peptide fragment fingerprinting (PFF) (Blueggel et al., 2004).

Table 6. Peptide fragmentation fingerprinting softwares and corresponding URL addresses.

Program Name	**URL**
Mascot	http://www.matrixscience.com/
MS-Seq	http://prospector.ucsf.edu/
MS-Tag	http://prospector.ucsf.edu/
PepFrag	http://prowl.rockefeller.edu/prowl/pepfragch.html
SEQUEST	http://fields.scripps.edu/sequest/

B. *De Novo* sequencing and homology searching:

In proteomics, *de Novo* sequencing is the process of deriving peptide sequences from tandem mass spectra without the assistance of a sequence database (Ma & Johnson, 2012). Unlike the strategy of identification through peptide fragment fingerprinting, which involves comparing experimental masses to theoretical masses, *de Novo* sequencing entails directly interpreting a fragmentation spectrum to deduce the primary sequence of the peptide. Although powerful, *de novo* sequencing often can only determine partially correct sequence tags because of imperfect tandem mass spectra. However, these sequence 'tags' can be searched in a sequence database to identify the exact or homologous peptide (Fig.25). Homology searches are particularly useful for studying organisms whose genomes have not been sequenced. The peptide sequence tags must be compared to protein sequences contained within a database (Shevchenko et al., 2001). For example, this comparison is carried out through a similarity search using the MS BLAST algorithm (Basic Local Alignment Search Tool). Thus, if a protein in the database contains peptides similar to the *de novo* sequence tags, the studied protein is highly likely to be an ortholog of that database protein. Despite not obtaining the complete protein sequence, functional information is thus obtained through homology.

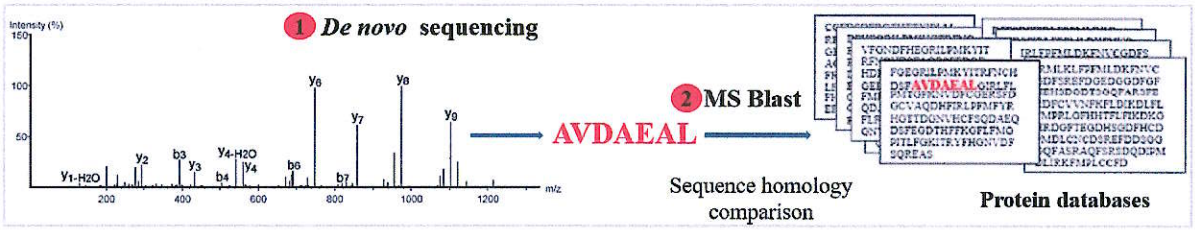


Figure 25. MS/MS spectrum through *de novo* sequencing with an MS BLAST search in a protein database.

V. QUANTITATIVE PROTEOMICS (DIFFERENTIAL QUANTITATION)

The first technique for quantifying the proteome was 2-DE, a conventional proteomic instrument. To offer relative quantification, the staining densities of proteins are compared over different gels. 2-DE does, however, have its limitations. First, proper quantification is hampered by the co-appearance of protein spots and has a limited resolution for many proteins that comprise the whole proteome. Moreover, 2-DE is inappropriate for membrane proteins since it reacts poorly with hydrophobic proteins. In samples with a high dynamic range, like plasma, where protein expression can change by up to 12 orders of magnitude, it can also not analyze low-abundance proteins. 2-DE quantification only picks up on extreme differences and yields imprecise estimates of quantity changes. Additionally, 2-DE requires much work and effort because it only permits individual gel. However, Difference gel electrophoresis (DIGE) technology provides a robust quantitative component to proteomics experiments involving two-dimensional (2D) gel electrophoresis (Friedman, 2007).

MS-based methods have mainly replaced 2-DE quantification. The isotope labeling approach allows the mixing multiple samples at different experimental stages. Quantification strategies use stable isotopes (2H, 13C, 15N, and 18O) for sample labeling, which is then mixed at a 1:1 ratio with an unlabeled sample and analyzed simultaneously. The absolute or relative protein abundance can be obtained by measuring the intensities of different isotope-coded peaks distinguished by mass spectrometry(Pappireddi et al., 2019) (Fig.26).

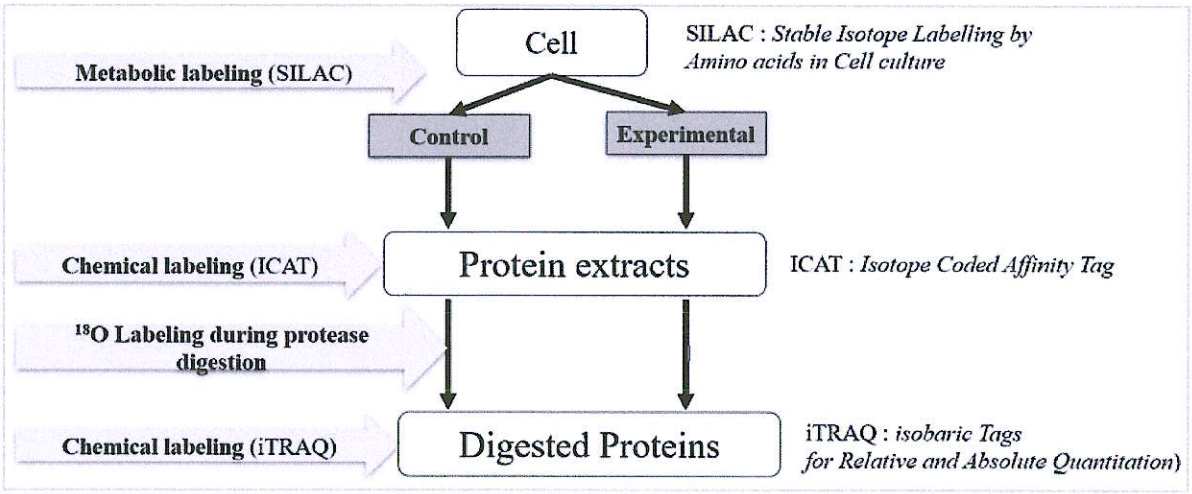


Figure 26. Schematic diagram of Isotope labeling approaches.

V. 1 Metabolic labeling

Stable isotope labeling of amino acids in cell culture (SILAC) has emerged as a popular alternative in which only selected amino acids are labeled, typically arginine and lysine. Following cleavage with trypsin, all peptides contain at least one labeled amino acid (Ong et al., 2002). No chemical labeling or affinity purification steps are performed, and the method is compatible with virtually all cell culture conditions, including primary cells (Fig.27). SILAC requires living cells but may be advantageous over other quantitative proteomics techniques whenever cell culture is used.

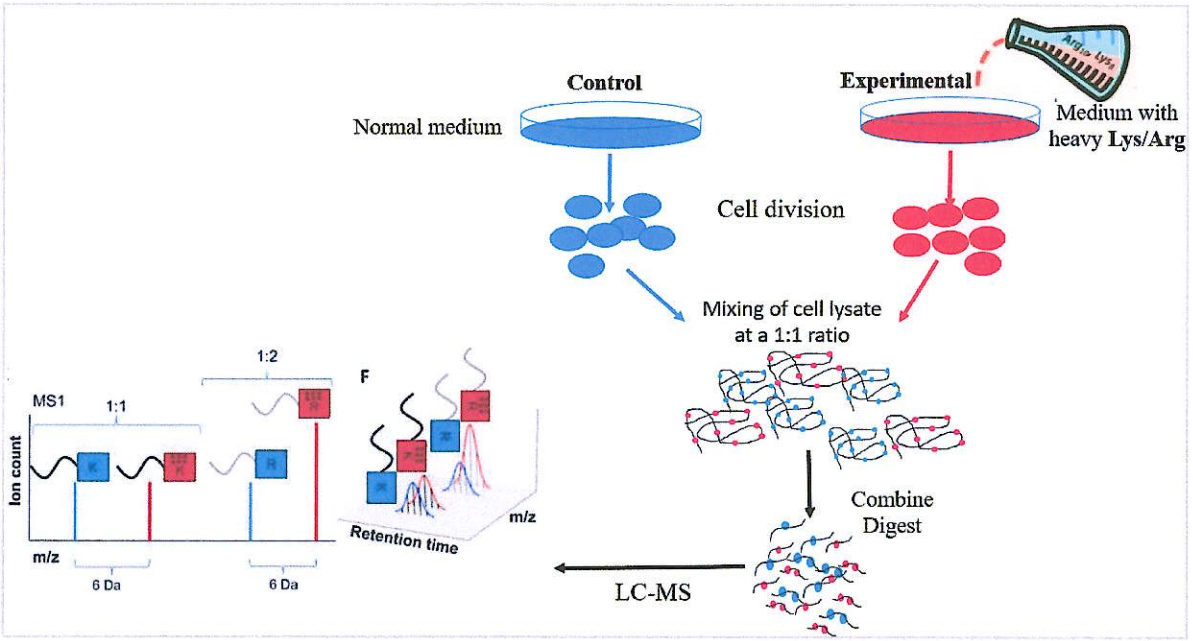


Figure 27. Stable Isotope Labelling by Amino acids in Cell culture (Mass spectrum taken from Ref (Pappireddi et al., 2019).

V. 2 Chemical labeling

In 1999, Aebersold and co-workers (Gygi et al., 1999) introduced the isotope-coded affinity tag (ICAT) method for relative quantitation of protein abundance. In this approach, an isotopically labeled affinity reagent is attached to particular amino acids in all proteins in the population. After protein digestion into peptides, the labeled peptides are affinity-purified using the newly incorporated affinity tag, thereby simplifying the peptide mixture while incorporating the isotopic label.

While MS1-based isotope labeling offers exquisite quantification for more abundant peptides, it suffers from a lack of multiplexing capability because, as the number of samples increases, so does the complexity of the MS1 spectrum. Multiplexed proteomics based on isobaric mass tags promises to overcome or mitigate these limitations. A popular version of isobaric mass tags is the iTRAQ reagent (Ross et al., 2004), which has expanded to incorporate up to eight reporter mass ions. As a result, the run time needed to evaluate several samples can be decreased. This is especially important for biological research involving evaluating several variables or time points, such as signaling networks.

Isobaric mass tags are all the same mass, and only upon fragmentation are the different mass tags observed. The isobaric mass tag consists of an amine-specific reactive group, a balancer group, and a reporter mass group. The amine-specific reactive group targets the peptide amino-termini and lysine side chains (Fig.28).

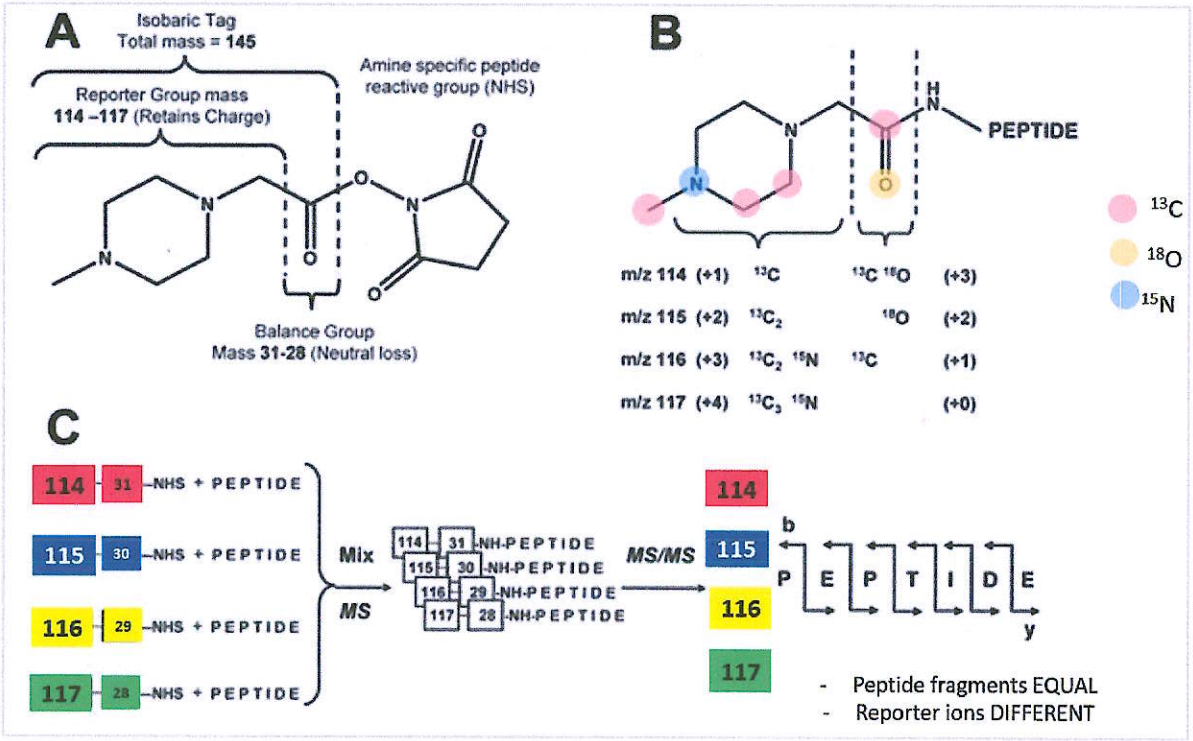


Figure 28. The multiplex isobaric tags (Ross et al., 2004).

A: Chemical structure of the different iTRAQ tags 114, 115, 116, and 117.
B: Chemical structure of the tag after attachment to the peptide.
C: Reporter-Balance-Peptide INTACT (4 samples identical m/z)

The structure, shown in Figure 29, can encode up to four different conditions. An 8-plex iTRAQ is also commercially available (Pappireddi et al., 2019).

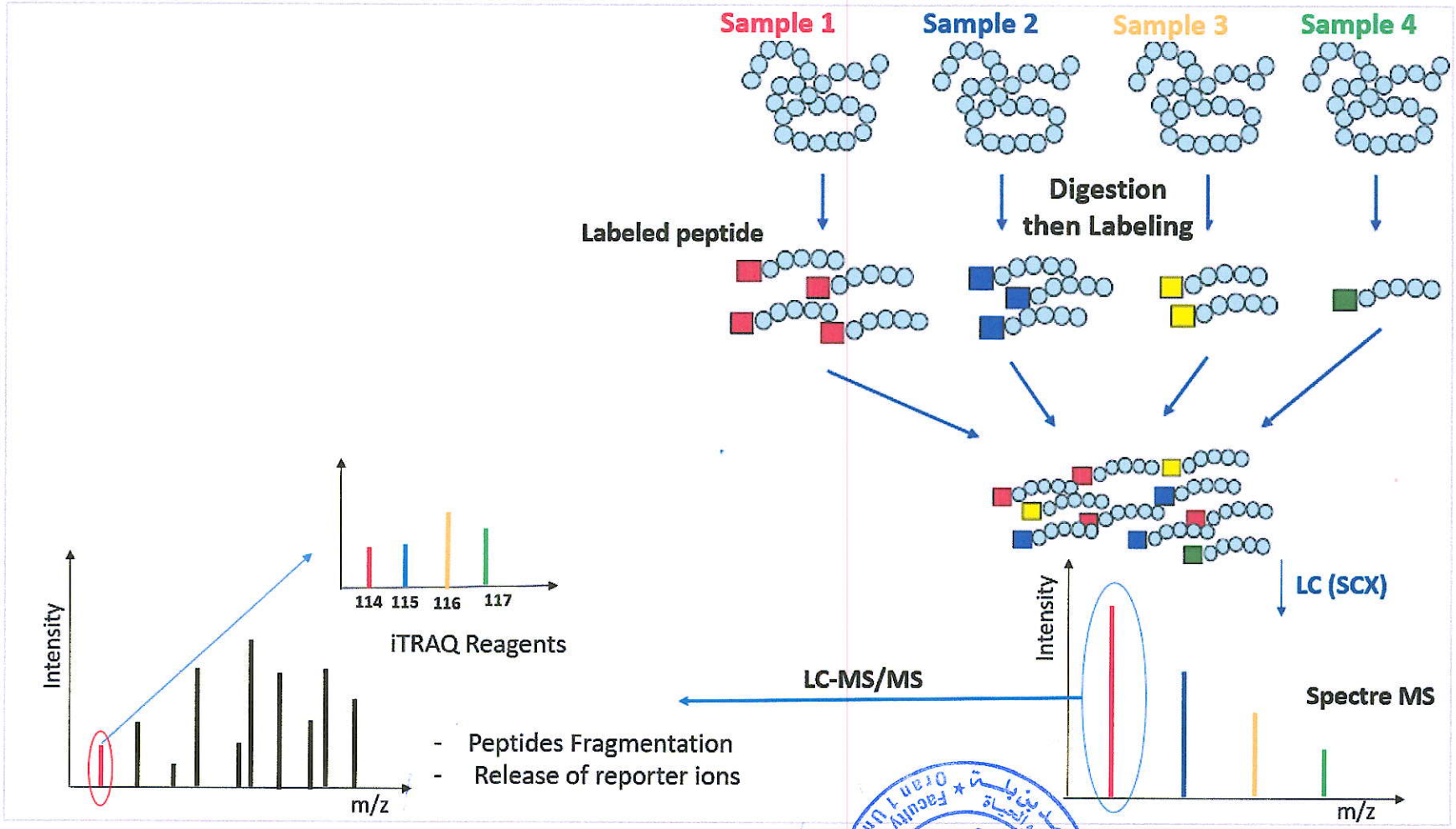


Figure 29. Schematic diagram of Itraq 4-Plex. Inspired by Pappireddi et al. (2019)

BIBLIOGRAPHIC REFERENCES

Aebersold, R., & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, *422*(6928), 198-207. https://doi.org/10.1038/nature01511

Appel, R. D., Bairoch, A., & Hochstrasser, D. F. (1999). 2-D databases on the World Wide Web. *Methods Mol Biol*, *112*, 383-391. https://doi.org/10.1385/1-59259-584-7:383

Arnott, D., Shabanowitz, J., & Hunt, D. F. (1993). Mass spectrometry of proteins and peptides: sensitive and accurate mass measurement and sequence analysis. *Clin Chem*, *39*(9), 2005-2010.

Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., & Rasool, M. H. (2017). Proteomics: Technologies and Their Applications. *J Chromatogr Sci*, *55*(2), 182-196. https://doi.org/10.1093/chromsci/bmw167

Bartels, C. (1990). Fast algorithm for peptide sequencing by mass spectroscopy. *Biomed Environ Mass Spectrom*, *19*(6), 363-368. https://doi.org/10.1002/bms.1200190607

Biemann, K. (1990). Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol*, *193*, 886-887. https://doi.org/10.1016/0076-6879(90)93460-3

Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Görg, A., Westermeier, R., & Postel, W. (1982). Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J Biochem Biophys Methods*, *6*(4), 317-339. https://doi.org/10.1016/0165-022x(82)90013-6

Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J., & Schäfer, F. (2009). Immobilized-metal affinity chromatography (IMAC): a review. *Methods Enzymol*, *463*, 439-473. https://doi.org/10.1016/s0076-6879(09)63027-5

Blueggel, M., Chamrad, D., & Meyer, H. E. (2004). Bioinformatics in proteomics. *Curr Pharm Biotechnol*, *5*(1), 79-88. https://doi.org/10.2174/1389201043489648

Bodzon-Kulakowska, A., Bierczynska-Krzysik, A., Dylag, T., Drabik, A., Suder, P., Noga, M., Jarzebinska, J., & Silberring, J. (2007). Methods for samples preparation in proteomic research. *J Chromatogr B Analyt Technol Biomed Life Sci*, *849*(1-2), 1-31. https://doi.org/10.1016/j.jchromb.2006.10.040

Bonaventura, C., Bonaventura, J., Stevens, R., & Millington, D. (1994). Acrylamide in Polyacrylamide Gels Can Modify Proteins during Electrophoresis. *Analytical Biochemistry*, *222*(1), 44-48. https://doi.org/https://doi.org/10.1006/abio.1994.1451

Burgess, R. R. (2009). Chapter 4 Preparing a Purification Summary Table. In R. R. Burgess & M. P. Deutscher (Eds.), *Methods in Enzymology* (Vol. 463, pp. 29-34). Academic Press. https://doi.org/https://doi.org/10.1016/S0076-6879(09)63004-4

Catherman, A. D., Skinner, O. S., & Kelleher, N. L. (2014). Top Down proteomics: facts and perspectives. *Biochem Biophys Res Commun*, *445*(4), 683-693. https://doi.org/10.1016/j.bbrc.2014.02.041

Cech, N. B., & Enke, C. G. (2001). Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev*, *20*(6), 362-387. https://doi.org/10.1002/mas.10008

Chevalier, F., Centeno, D., Rofidal, V., Tauzin, M., Martin, O., Sommerer, N., & Rossignol, M. (2006). Different impact of staining procedures using visible stains and fluorescent

dyes for large-scale investigation of proteomes by MALDI-TOF mass spectrometry. *J Proteome Res*, *5*(3), 512-520. https://doi.org/10.1021/pr050194n

Dancík, V., Addona, T. A., Clauser, K. R., Vath, J. E., & Pevzner, P. A. (1999). De novo peptide sequencing via tandem mass spectrometry. *J Comput Biol*, *6*(3-4), 327-342. https://doi.org/10.1089/106652799318300

Duong, V. A., & Lee, H. (2023). Bottom-Up Proteomics: Advancements in Sample Preparation. *Int J Mol Sci*, *24*(6). https://doi.org/10.3390/ijms24065350

El-Aneed, A., Cohen, A., & Banoub, J. (2009). Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Applied Spectroscopy Reviews*, *44*(3), 210-230. https://doi.org/10.1080/05704920902717872

Ezkurdia, I., Vázquez, J., Valencia, A., & Tress, M. (2014). Analyzing the first drafts of the human proteome. *J Proteome Res*, *13*(8), 3854-3855. https://doi.org/10.1021/pr500572z

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science*, *246*(4926), 64-71. https://doi.org/10.1126/science.2675315

Fíla, J., & Honys, D. (2012). Enrichment techniques employed in phosphoproteomics. *Amino Acids*, *43*(3), 1025-1047. https://doi.org/10.1007/s00726-011-1111-z

Foster, L. J., de Hoog, C. L., Zhang, Y., Zhang, Y., Xie, X., Mootha, V. K., & Mann, M. (2006). A mammalian organelle map by protein correlation profiling. *Cell*, *125*(1), 187-199. https://doi.org/10.1016/j.cell.2006.03.022

Fountoulakis, M., Juranville, J. F., Röder, D., Evers, S., Berndt, P., & Langen, H. (1998). Reference map of the low molecular mass proteins of Haemophilus influenzae. *Electrophoresis*, *19*(10), 1819-1827. https://doi.org/10.1002/elps.1150191046

Fredolini, C., Byström, S., Pin, E., Edfors, F., Tamburro, D., Iglesias, M. J., Häggmark, A., Hong, M. G., Uhlen, M., Nilsson, P., & Schwenk, J. M. (2016). Immunocapture strategies in translational proteomics. *Expert Rev Proteomics*, *13*(1), 83-98. https://doi.org/10.1586/14789450.2016.1111141

Friedman, D. B. (2007). Quantitative proteomics for two-dimensional gels using difference gel electrophoresis. *Methods Mol Biol*, *367*, 219-239. https://doi.org/10.1385/1-59745-275-0:219

Gorg, A. (1991). Two-dimensional electrophoresis. *Nature*, *349*(6309), 545-546. https://doi.org/10.1038/349545a0

Görg, A., Drews, O., Lück, C., Weiland, F., & Weiss, W. (2009). 2-DE with IPGs. *Electrophoresis*, *30 Suppl 1*, S122-132. https://doi.org/10.1002/elps.200900051

Görg, A., Postel, W., & Günther, S. (1988). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, *9*(9), 531-546. https://doi.org/10.1002/elps.1150090913

Görg, A., Weiss, W., & Dunn, M. J. (2004). Current two-dimensional electrophoresis technology for proteomics. *Proteomics*, *4*(12), 3665-3685. https://doi.org/10.1002/pmic.200401031

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., & Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*, *17*(10), 994-999. https://doi.org/10.1038/13690

Han, X., Aslanian, A., & Yates, J. R. (2008). Mass spectrometry for proteomics. *Current Opinion in Chemical Biology*, *12*(5), 483-490. https://doi.org/https://doi.org/10.1016/j.cbpa.2008.07.024

Hong, P., Koza, S., & Bouvier, E. S. (2012). Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *J Liq Chromatogr Relat Technol*, *35*(20), 2923-2950. https://doi.org/10.1080/10826076.2012.743724

Hwang, A. C., Grey, P. H., Cuddy, K., & Oppenheimer, D. G. (2012). Pouring and running a protein gel by reusing commercial cassettes. *J Vis Exp*(60). https://doi.org/10.3791/3465

Izzo, V., Costa, M. A., Di Fiore, R., Duro, G., Bellavia, D., Cascone, E., Colombo, P., Gioviale, M. C., & Barbieri, R. (2006). Electrophoresis of proteins and DNA on horizontal sodium dodecyl sulfate polyacrylamide gels. *Immun Ageing*, *3*, 7. https://doi.org/10.1186/1742-4933-3-7

Jungbauer, A., & Hahn, R. (2009). Chapter 22 Ion-Exchange Chromatography. In R. R. Burgess & M. P. Deutscher (Eds.), *Methods in Enzymology* (Vol. 463, pp. 349-371). Academic Press. https://doi.org/https://doi.org/10.1016/S0076-6879(09)63022-6

Karas, M., & Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*, *60*(20), 2299-2301. https://doi.org/10.1021/ac00171a028

Kim, M. S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., Thomas, J. K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabuddhe, N. A., Balakrishnan, L., Advani, J., George, B., Renuse, S., . . . Pandey, A. (2014). A draft map of the human proteome. *Nature*, *509*(7502), 575-581. https://doi.org/10.1038/nature13302

Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, *26*(3), 231-243. https://doi.org/10.1007/bf00281458

Labrou, N. E. (2014). Protein Purification: An Overview. In N. E. Labrou (Ed.), *Protein Downstream Processing: Design, Development and Application of High and Low-Resolution Methods* (pp. 3-10). Humana Press. https://doi.org/10.1007/978-1-62703-977-2_1

Larbi, N. B., & Jefferies, C. (2009). 2D-DIGE: comparative proteomics of cellular signalling pathways. *Methods Mol Biol*, *517*, 105-132. https://doi.org/10.1007/978-1-59745-541-1_8

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., & Yates, J. R., 3rd. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*, *17*(7), 676-682. https://doi.org/10.1038/10890

Liu, S., Li, Z., Yu, B., Wang, S., Shen, Y., & Cong, H. (2020). Recent advances on protein separation and purification methods. *Advances in Colloid and Interface Science*, *284*, 102254. https://doi.org/https://doi.org/10.1016/j.cis.2020.102254

Ma, B., & Johnson, R. (2012). De novo sequencing and homology searching. *Mol Cell Proteomics*, *11*(2), O111.014902. https://doi.org/10.1074/mcp.O111.014902

Mann, M., & Wilm, M. (1994). Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem*, *66*(24), 4390-4399. https://doi.org/10.1021/ac00096a002

Nesvizhskii, A. I. (2007). Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol Biol*, *367*, 87-119. https://doi.org/10.1385/1-59745-275-0:87

O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem*, *250*(10), 4007-4021.

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*, *1*(5), 376-386. https://doi.org/10.1074/mcp.m200025-mcp200

Palii, C., Laura, M., Montreuil, J., & Artenie, V. (2007). L'APPROCHE PROTEOMIQUE ET SES APPLICATIONS. *Analele Ştiinţifice Ale Universităţii Alexandru Ioan Cuza din Iaşi,Sectiunea II A : Genetica si Biologie Moleculara*, *8*.

Pappin, D. J., Hojrup, P., & Bleasby, A. J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol*, *3*(6), 327-332. https://doi.org/10.1016/0960-9822(93)90195-t

Pappireddi, N., Martin, L., & Wühr, M. (2019). A Review on Quantitative Multiplexed Proteomics. *Chembiochem*, *20*(10), 1210-1224. https://doi.org/10.1002/cbic.201800650

Picotti, P., & Aebersold, R. (2012). Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods*, *9*(6), 555-566. https://doi.org/10.1038/nmeth.2015

Porath, J., Carlsson, J., Olsson, I., & Belfrage, G. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature*, *258*(5536), 598-599. https://doi.org/10.1038/258598a0

Queiroz, J. A., Tomaz, C. T., & Cabral, J. M. (2001). Hydrophobic interaction chromatography of proteins. *J Biotechnol*, *87*(2), 143-159. https://doi.org/10.1016/s0168-1656(01)00237-1

Rabilloud, T., & Lelong, C. (2011). Two-dimensional gel electrophoresis in proteomics: a tutorial. *J Proteomics*, *74*(10), 1829-1841. https://doi.org/10.1016/j.jprot.2011.05.040

Raymond, S., & Weintraub, L. (1959). Acrylamide gel as a supporting medium for zone electrophoresis. *Science*, *130*(3377), 711. https://doi.org/10.1126/science.130.3377.711

Rodriguez, E. L., Poddar, S., Iftekhar, S., Suh, K., Woolfork, A. G., Ovbude, S., Pekarek, A., Walters, M., Lott, S., & Hage, D. S. (2020). Affinity chromatography: A review of trends and developments over the past 50 years. *J Chromatogr B Analyt Technol Biomed Life Sci*, *1157*, 122332. https://doi.org/10.1016/j.jchromb.2020.122332

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., & Pappin, D. J. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*, *3*(12), 1154-1169. https://doi.org/10.1074/mcp.M400129-MCP200

Rotilio, D., Della Corte, A., D'Imperio, M., Coletta, W., Marcone, S., Silvestri, C., Giordano, L., Di Michele, M., & Donati, M. B. (2012). Proteomics: bases for protein complexity understanding. *Thromb Res*, *129*(3), 257-262. https://doi.org/10.1016/j.thromres.2011.12.035

Sandra, K., Moshir, M., D'Hondt, F., Verleysen, K., Kas, K., & Sandra, P. (2008). Highly efficient peptide separations in proteomics Part 1. Unidimensional high performance liquid chromatography. *J Chromatogr B Analyt Technol Biomed Life Sci*, *866*(1-2), 48-63. https://doi.org/10.1016/j.jchromb.2007.10.034

Schägger, H., & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem*, *166*(2), 368-379. https://doi.org/10.1016/0003-2697(87)90587-2

Shevchenko, A., Sunyaev, S., Loboda, A., Shevchenko, A., Bork, P., Ens, W., & Standing, K. G. (2001). Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal Chem*, *73*(9), 1917-1926. https://doi.org/10.1021/ac0013709

Shi, Y., Xiang, R., Horváth, C., & Wilkins, J. A. (2004). The role of liquid chromatography in proteomics. *J Chromatogr A*, *1053*(1-2), 27-36.

Shukla, A. K., & Futrell, J. H. (2000). Tandem mass spectrometry: dissociation of ions by collisional activation. *J Mass Spectrom*, *35*(9), 1069-1090. https://doi.org/10.1002/1096-9888(200009)35:9<1069::Aid-jms54>3.0.Co;2-c

Smith, B. J. (1994). SDS Polyacrylamide Gel Electrophoresis of Proteins. In J. M. Walker (Ed.), *Basic Protein and Peptide Protocols* (pp. 23-34). Humana Press. https://doi.org/10.1385/0-89603-268-X:23

Staby, A., Jensen, I. H., & Mollerup, I. (2000). Comparison of chromatographic ion-exchange resins: I. Strong anion-exchange resins. *Journal of Chromatography A*, *897*(1), 99-111. https://doi.org/https://doi.org/10.1016/S0021-9673(00)00780-9

Suárez-Díaz, E. (2022). The Electrophoretic Revolution in the 1960s: Historical Epistemology Meets the Global History of Science and Technology. *Ber Wiss*, *45*(3), 332-343. https://doi.org/10.1002/bewi.202200024

Torsetnes, S. B., Nordlund, M. S., Paus, E., Halvorsen, T. G., & Reubsaet, L. (2013). Digging deeper into the field of the small cell lung cancer tumor marker ProGRP: a method for differentiation of its isoforms. *J Proteome Res*, *12*(1), 412-420. https://doi.org/10.1021/pr300751j

Towbin, H., Staehelin, T., & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A*, *76*(9), 4350-4354. https://doi.org/10.1073/pnas.76.9.4350

Unlü, M., Morgan, M. E., & Minden, J. S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis*, *18*(11), 2071-2077. https://doi.org/10.1002/elps.1150181133

Vidal, M., & Legrain, P. (1999). Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res*, *27*(4), 919-929. https://doi.org/10.1093/nar/27.4.919

Vihinen, M. (2001). Bioinformatics in proteomics. *Biomol Eng*, *18*(5), 241-248. https://doi.org/10.1016/s1389-0344(01)00099-5

Vincent, D., Savin, K., Rochfort, S., & Spangenberg, G. (2020). The Power of Three in Cannabis Shotgun Proteomics: Proteases, Databases and Search Engines. *Proteomes*, *8*(2). https://doi.org/10.3390/proteomes8020013

Wang, Q., Chang, N., Moomaw, B., Medberry, S., & Gallagher, S. R. (2023). Overview of Digital Electrophoresis Analysis. *Curr Protoc*, *3*(1), e651. https://doi.org/10.1002/cpz1.651

Washburn, M. P., Wolters, D., & Yates, J. R., 3rd. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*, *19*(3), 242-247. https://doi.org/10.1038/85686

Wells, J. M., & McLuckey, S. A. (2005). Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol*, *402*, 148-185. https://doi.org/10.1016/s0076-6879(05)02005-7

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J. H., Bantscheff, M., . . . Kuster, B. (2014). Mass-spectrometry-based draft of the human proteome. *Nature*, *509*(7502), 582-587. https://doi.org/10.1038/nature13319

Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., & Williams, K. L. (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev*, *13*, 19-50. https://doi.org/10.1080/02648725.1996.10647923

Wingfield, P. (2001). Protein precipitation using ammonium sulfate. *Curr Protoc Protein Sci*, *Appendix 3*, Appendix 3F. https://doi.org/10.1002/0471140864.psa03fs13

Yates, J. R., 3rd. (2004). Mass spectral analysis in proteomics. *Annu Rev Biophys Biomol Struct*, *33*, 297-316. https://doi.org/10.1146/annurev.biophys.33.111502.082538

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., & Yates, J. R., III. (2013). Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews*, *113*(4), 2343-2394. https://doi.org/10.1021/cr3003533

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., & Yates, J. R., 3rd. (2013). Protein analysis by shotgun/bottom-up proteomics. *Chem Rev*, *113*(4), 2343-2394. https://doi.org/10.1021/cr3003533

Zhu, H., & Snyder, M. (2001). Protein arrays and microarrays. *Curr Opin Chem Biol*, *5*(1), 40-45. https://doi.org/10.1016/s1367-5931(00)00170-8