Biochemistry and Biotechnology Research
Vol. 4(1), pp. 1-16, March 2016
ISSN: 2354-2136
Review

Techniques and applications of proteomics in plant ecophysiology

Fisseha Asmelash

Ethiopian Biodiversity Institute, P.O. Box 30726, Addis Ababa, Ethiopia.

Accepted 1 March, 2016

ABSTRACT

Proteins are the tools that make living machines work. They are the single most important molecules that can tell us a lot about life processes. Proteins have long been studied in biochemistry. However, genome wide study of proteins was only possible in recent years due to technological advances and the availability of sequenced genome. Hence, the genome wide study of proteins is called Proteomics. Proteomics is superior over genomics and other OMICS since it can capture post-translational modifications and other chemical changes which other OMICS cannot. This review article discusses on the detailed techniques in proteomics focusing on plant proteomics. It highlights the status of plant proteomics generally and in Africa specifically. It briefly discusses on the role of proteomics in plant ecophysiology studies and finally comes up with concluding remarks.

Keywords: Amino acids, de-novo sequencing, ecophysiology, mass spectrometer, proteins, proteomics.

E-mail: Fisseha33@gmail.com.

INTRODUCTION

Genes control all metabolic processes by synthesizing enzymes that are proteins (Sinha, 2004). Hence, studying the function of genes rely mainly on the study of proteins. When the study of genome wide proteins (Proteome) takes place, it is known as Proteomics (Liebler, 2002). Proteomics is a relatively new science (Tambor et al., 2010) and it is extremely dynamically evolving (Tambor et al., 2010; Jorrin-Novo, 2014). Proteomics has profound applications in medicine. It also has applications in the food industry. Its application in Plant ecophysiology and breeding is indispensible.

Proteomics is instrumental in the discovery of biomarkers (Proteins) that are associated with diseases (Wulfkuhle and Petricoin, 2003). For instance, Alzheimer's disease and heart disease are known to have characteristic biomarkers that physicians can use for diagnosis (Tyagi et al., 2010). Proteomics also allows the identification of cancer biomarkers such that the detection of cancer at the premalignant stage is made possible. This can ultimately ensure a higher probability for curing the disease (Wulfkuhle and Petricoin, 2003). Therefore, determining the protein(s) implicated in diseases is one of the most important applications of proteomics.

Food products genuineness authentication, nutritional value determination, food allergens quantification, selection of suitable durum wheat cultivars for pasta making and selection and identification of protein markers to select cultivars suitable for flour making are some of the uses of proteomics in food industry (Eldakak et al., 2013).

Proteins are the tools that make living machines work (Adams, 2008). But they need to undergo a number of posttranslational modifications (PTMs) before they are able to manipulate the living machine. A newly synthesized polypeptide chain is converted into a functional protein through series of chemical reactions known as posttranslational modification (Kersten et al., 2006). As of 2015, more than 200 different types of posttranslational modifications have been identified that affect many aspects of cellular functionalities, such as metabolism, signal transduction, and protein stability (Duan and Walther, 2015). Hence, proteins in their native form do not generally control the complex biological functions of cells (Kersten et al., 2006). Likewise, the same PTMs are known to create a mismatch between expressional and functional genes. Meanwhile, for a given gene to be selected for breeding, it should be both

expressional and functional gene for the desired trait (salekdeh et al., 2002).

Therefore, since PTMs can only be studied by proteomics, the role of proteomics in plant ecophysiology and breeding is indispensible. Meanwhile, the focus of this review is to try and illustrate the role of proteomics in plant ecophysiology. In so doing, a general introductory concept on protein biochemistry and proteomics is amicably presented, techniques in proteomics are extensively covered, and the status of the global and African plant proteomics is highlighted.

PROTEINS AND THEIR ROLE IN PLANT PHYSIOLOGY

Protein biochemistry

While nucleic acids are biopolymers that store and transmit the genetic information of cells, proteins are biopolymers that express much of that information (Mathews and Van holde, 2002). The monomeric building blocks of proteins are 20 amino acids (Taiz and Zeiger, 2002; Sinha, 2004; Walker, 2010; Lodish et al., 2013) which when incorporated in to the protein polymer are sometimes called residues (Lodish et al., 2013). These twenty amino acids vary in size, shape, charge and chemical reactivity but have a characteristic structure consisting of central alpha carbon atom (Cα) bonded to four different chemical groups: an amino ($-NH_2$) group, carboxyl or carboxylic acid (COOH) group (hence the amino acid), a hydrogen (H) atom, and one variable group called the side chain or R group (Lodish et al., 2013). For the Nineteen of the amino acids, the amino and carboxyl groups are attached to the carbon atom that is adjacent to the carboxyl group (hence α amino acids) but only proline has the $-NH_2$ group (amino) incorporated into a five-member ring (Figure 1). Therefore, proline is also referred as imino acid (Walker, 2010).

Although the chemistry of amino acids may be more complex, based on the nature of the side chain, there are basically two classes of amino acids; non polar and polar amino acids (Table 1). Amino acids with non polar side chains are called hydrophobic and are poorly soluble in water. Amino acids with polar side chains are called hydrophilic. Hydrophilic amino acids can have +ve charge and are called basic amino acids and when having -ve charge acidic amino acids (Lodish et al., 2013).

When the three-dimensional structures of different protein molecules are compared, it becomes clear that, although the overall conformation of each protein is unique, two regular folding patterns are often found in parts of them. These are the α helix and the β sheet which are otherwise known as protein secondary structures (Figure 2) (Alberts et al., 2002). Meanwhile, protein molecules have four levels of hierarchical structures. These are, primary structure that is the amino acid residue sequence, secondary structure (the local folding), tertiary structure is the three dimensional over all folding and quaternary structure is the association of two or more separate three-dimensional polypeptides forming three dimensional complex (Alberts et al., 2002; Taiz and Zeiger, 2002; Walker, 2010; Lodish et al., 2013). Hence, primary, secondary and tertiary levels of organization involve protein formed from one polypeptide while quaternary level of organization involves protein formed from two or more polypeptides (Walker, 2010) (Figure 2).

Furthermore, studies of the conformation, function, and evolution of proteins have revealed a unit of organization distinct from the four protein structures. This is the protein domain, a substructure produced by any part of a polypeptide chain that can fold independently into a compact, 3D stable structure. The different domains of a protein are often associated with different functions. Small proteins typically consist of only a single domain, while large proteins are formed from several domains linked together by short lengths of polypeptide chains (Alberts et al., 2002). Yet, another substructure is the motif. A motif is a specific combination of secondary structures that has a particular topology and is organized into a characteristic 3D structure. The presence of the same motif in different proteins with similar functions clearly indicates that during evolution these useful combinations of secondary structures have been conserved (Lodish et al., 2013).

The structure and chemistry of each protein has been developed and fine-tuned over billions of years of evolutionary history. Hence, proteins are by far the most structurally complex and functionally sophisticated molecules known (Alberts et al., 2002). The sophisticated functions of proteins include their role as enzymes. Enzymes control almost all life processes hence are called the “agents of life” (Taiz and Zeiger, 2002). All enzymes are proteins, although recently some small ribonucleic acids and protein–RNA complexes have been found to exhibit enzyme-like behaviour in the processing of RNA (Taiz and Zeiger, 2002) (Table 2).

Role of proteins in plant physiology

Plant physiology can be studied at three levels of plants organization viz. at metabolic level, cellular, or whole-plant level (Öpik and Rolfe, 2005). However, whole plant physiological processes of plants can be described only to a limited degree without reference to metabolic and cellular physiology (Öpik and Rolfe, 2005). Cellular and metabolic activities of plants are basically executed by proteins. That is why plant physiology text books, by default, contain much information on the functions of proteins in each of their chapters (Taiz and Zeiger, 2002; Öpik and Rolfe, 2005). Hence, to try and discuss about plant physiological roles of proteins would mean to write a plant physiology text book. So this review can only

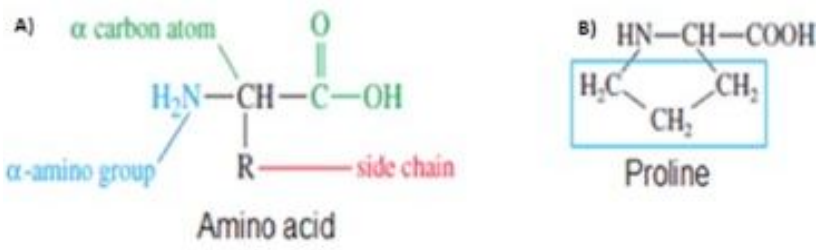

Figure 1. Chemical structure of amino acids (a) and the modification of proline (b).

Table 1. The 20 amino acids with one letter and three letter symbols and with their average residue mass (Based on Taiz and Zeiger, 2002 and Liebler, 2002).

Hydrophobic (non polar) R group	Hydrophilic (polar) R group		
	Neutral	Basic ("+"charge)	Acidic ("-" charge)
Alanine [A] (Ala){ 71.08}	Asparagine[N](Asn) {114.10}	Lysine [K] (Lys) {128.17}	Aspartate[D](Asp) {115.09}
Valine [V] (Val) {99.13}	Glutamine[Q] (Gln) {128.13}	Arginine [R](Arg) {156.19}	Glutamate[E](Glu) {129.12}
Leucine [L] (Leu){113.16}	Serine [S] (Ser) {87.08}	Histidine [H](His) {137.14}	
Isoleucine[I](Ile){113.16}	Threonine [T] (Thr) {101.11}		
Proline [P] (Pro) {97.12}	Thyrosine [Y] (Tyr) {97.12}		
Phenylalanine [F](Phe){147.18}			
Tryptophan [W] (Trp) {186.21}			
Methionine [M] (Met) {131.19}			
Cysteine [C] (Cys) {103.14}			
Glysine [G] (Gly) {57.05}			

provide a snapshot of some of the roles of proteins in plant physiology.

The primary plant metabolites (the amino acids themselves, nucleic acids, carbohydrates and lipids) are synthesized by enzymatic actions. The secondary metabolites, which are known to help build plants' defensive mechanism against herbivory, pests and diseases, are also the result of enzymatic actions. All the metabolic activities of plants are possible due to enzymes. Table 2 lists the main chemical reactions that make plant life possible. Meanwhile, one of the roles proteins have in plant physiology is their role as enzymes.

Proteins as photo receptors, membrane proteins, assimilate transporter proteins, signal transduction proteins, resistance proteins, and stress proteins are also some of the roles proteins play in plant physiology. The photosystem (II) and photosystem (I) are protein complexes which are integral components of the thylakoidal membrane of the chloroplast. These two protein complexes are able to trap light so that the light energy is converted to chemical energy through enzymatic action in photosynthesis (Taiz and Zeiger, 2002; Öpik and Rolfe, 2005, Tikkanen, 2009). Light affects photomorphogenesis; regulation of plant growth and morphology by light (Vince, 1960). These include; cell elongation, seed germination, leaf initiation, leaf shape determination and leaf expansion, stem elongation, photoperiodic flowering and dormancy, root elongation, shoot and root geotropism (Vince, 1960) and stomata movements (Öpik and Rolfe, 2005; Taiz and Zeiger, 2010). Morphogenesis is possible due to the

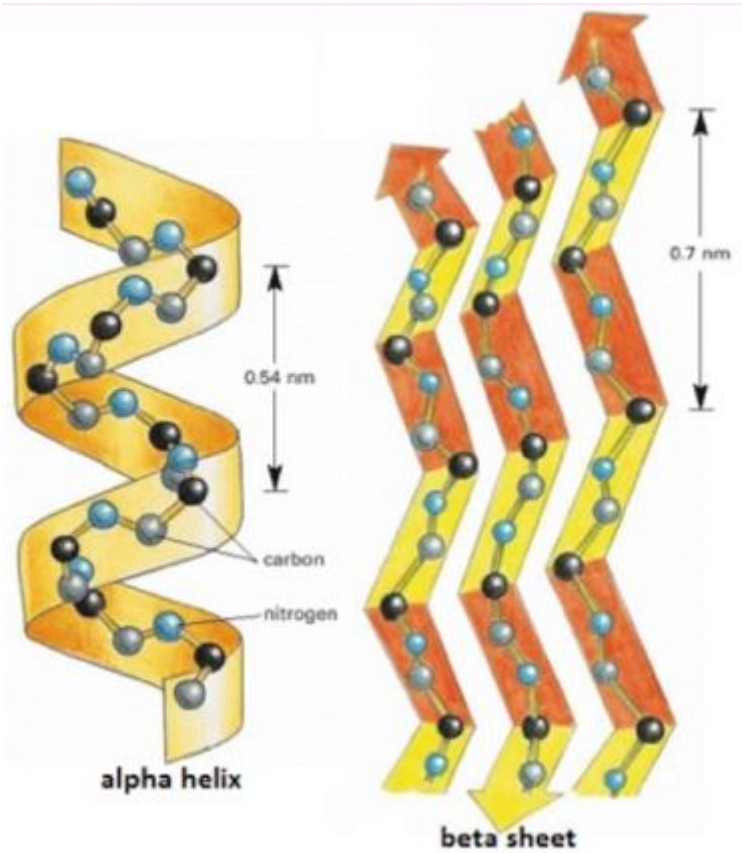

Figure 2. Protein secondary structure; α helix and β sheet (Alberts et al., 2002).

Table 2. Some common types of enzymes (Alberts et al., 2002).

Enzyme	Reaction catalyzed
Hydrolases	General term for enzymes that catalyze a hydrolytic cleavage reaction.
Nucleases	Break down nucleic acids by hydrolyzing bonds between nucleotides.
Proteases	Break down proteins by hydrolyzing bonds between amino acids.
Synthases	General name used for enzymes that synthesize molecules in anabolic reactions by condensing two smaller molecules together.
Isomerases	Catalyze the rearrangement of bonds within a single molecule
Polymerases	Catalyze polymerization reactions such as the synthesis of DNA and RNA
Kinases	Catalyze the addition of phosphate groups to molecules. Protein kinases are an important group of kinases that attach phosphate groups to proteins.
Phosphatases	Catalyze the hydrolytic removal of a phosphate group from a molecule
Oxido-Reductases	General name for enzymes that catalyze reactions in which one molecule is oxidized while the other is reduced. Enzymes of this type are often called *oxidases, reductases,* and *dehydrogenases.*
ATPases	Hydrolyze ATP. Many proteins with a wide range of roles have an energy-harnessing ATPase activity as part of their function, for example, motor proteins such as *myosin* and membrane transport proteins such as the *sodium–potassium pump*

photo receptors; the phytocrome, cytochrome, and zeaxanthin (Taiz and Zeiger, 2002; 2010) which are proteins. Zeaxanthin is a photoreceptor responsible in stomata movement (Taiz and Zeiger, 2010).

Membrane proteins are two types; 1) transport proteins that facilitate the selective passage of ions and other polar molecules, and 2) carrier protein that carry molecules and ions in to the cell membrane (Taiz and Zeiger, 2002). Transport proteins exhibit specificity for the solutes they transport, hence they have great diversity in cells (Taiz and Zeiger, 2002). Experimentally it was observed that in maize root sap the concentration of potassium can be 80 times more than the concentration in the soil solution while in the meantime, the concentration of sodium in the root sap is maintained lower than the soil solution and such selectivity and accumulation of ions is believed to be controlled by proteins (Marschner, 1993). Water movement through cell membranes is known to be selectively facilitated by transport proteins called aquaporins (Taiz and Zeiger, 2002; Kaldenhoff, 2006). Hence, aquaporins facilitate water movement into plant cells. Aquaporins are also believed to reduce mesophyll resistance of leaf to CO_2; hence increase photosynthesis (Kaldenhoff, 2012). Proteins of the vacuole membrane (the tonoplast) also regulate the movement of ions and metabolites between the vacuole and the cytosol where necessary, maintaining the osmotic pressure of vacuoles optimum for enough water gain and cell growth (Taiz and Zeiger, 2002).

Proteins have a role in assimilate transportation in the phloem. Experimental observations showed that carbohydrates injected in to the phloem companion cells do not show cell-to-cell movement but while they were microinjected together with phloem protein they do move thus proteins have role in assimilate transportation (Öpik and Rolfe, 2005). Phloem proteins are also known to function as a plastering agents when the sieve elements are damaged and ruptured; such that assimilate movement presumes (Taiz and Zeiger, 2002; Öpik and Rolfe, 2005). Some phloem proteins are also chaperones which squeeze and reshape large molecules to pass through the plasmodesmata (Öpik and Rolfe, 2005). Evidence also is mounting that some sieve tube sap proteins and RNA are information molecules which prepare the sink for assimilate unloading (Öpik and Rolfe, 2005).

Pathogen recognition takes place due to two relevant groups of host receptors that are both proteins. The first group is a membrane receptor. It is specialized in the recognition of pathogen and builds immunity based on the molecular structure of the pathogen. The second group is mainly the intracellular receptors called resistance proteins (R proteins) that detect pathogen effectors introduced into the host cells to destroy the built immunity (Glowacki et al., 2011). Proteins are known to initiate signal transduction in response to abiotic stresses in yeast and bacteria. Hence, protein homologues are also expected to be present in plants as well (Sanchez-Moreiras and Bonjoch, 2003). Acclimation to stress is also known to be mediated by proteins. During stressful environment plants synthesise stress proteins and proteins that help in the production of products that help the plant resist stress; like proline for water stress, polyamines for antioxidant defence (Bonjoch and Tamayo, 2003; Roveda-Hoyos and Fonseca-Moreno, 2011).

WHAT IS PROTEOMICS?

The terms “proteome” and “proteomics” correspond to “genome” and “genomics” and were first coined by Marc Wilkins and colleagues in the early 1990s (Liebler, 2002). The term “proteome” (PROTEins expressed by genOME) represents the survey of the expression of all proteins in a given time and condition (Hakeem et al., 2012). Proteomics is therefore the study of proteome (Liebler, 2002). Proteomics is the interface between protein biochemistry and molecular biology (Sanchez-Moreiras and Bonjoch, 2003). In more elaborate terms, proteomics is the systematic study of the amounts, modifications, interactions, localization, and function of proteins at the whole-organism, tissue, cellular and sub cellular levels (Lodish et al., 2013) at a given time point (Ghosh and Xu, 2014). According to Tyers and Mann (2003), proteomics is almost everything ‘post-genomic’.

There are four common subdivisions of proteomics. These are (Liebler, 2002):

Mining: the exercise of identifying all (or as many as possible) proteins in a sample.
Protein-expression profiling: comparative analysis of proteins in samples of varying conditions. The variability can be stage of development, exposure to disease, exposure to drugs or chemicals *etc*. Hence, Expression profiling is actually a specialized form of mining.
Mapping of protein modifications: deals with protein posttranslational modification. It is the task of identifying how and where proteins are modified.
Protein-network mapping: the proteomics approach to determining how proteins interact with each other in living systems.

However, according to Lodha et al. (2013), proteomics is simply subdivided in to structural proteomics (in-depth study of the four hierarchical structures of protein) and functional proteomics (studying the functions of proteomes). Tyagi et al. (2010) subdivide proteomics in to structural proteomics, expression proteomics (resembles mining and protein-expression profiling), and interaction proteomics (resembles protein-network mapping). It is still important to know that other authors, e.g. Jorrin-Novo (2014), subdivide proteomics to include more subdivisions viz. descriptive proteomics, comparative

proteomics, posttranslational proteomics, interactomics proteomics, sub-cellular proteomics, proteinomics, and translational proteomics. There is one important discipline in proteomics as well. It is known as quantitative proteomics. Quantitative proteomics is a kind of technique with which the change in abundance of a protein or peptide brought about some altered state is quantified (Wasinger et al., 2013). Hence it is a technique used in protein-expression profiling or comparative proteomics. At this juncture, it is very important to know what translational proteomics is. Translational proteomics deals with facilitating the ways by which proteomics findings are applied in solving the real life problems (Agrawal et al., 2012). Hence translational proteomics is very important since the ultimate goal of science is to solve real life problems and make life easy.

Proteomics comprises well-established methods of classical protein chemistry/biochemistry used before the term “proteome” was coined. These are: protein extraction, protein separation (gel electrophoresis and liquid chromatography), labeling, identification (Western blot), and Edman sequencing (Jorrin-Novo , 2014). As a technique, proteomics is advantaged over other “omics” tools since proteins are the key players in majority of cellular events. In addition to its capability of complementing transcriptome level changes, proteomics can also detect translational and post-translational regulations, thereby providing new insights into complex biological phenomena such as abiotic stress responses in plants (Ghosh and Xu, 2014).

TECHNIQUES IN PROTEOMICS

Any proteomic study like any other laboratory study starts with a sample. From the sample, protein (mixture) will be extracted. The sample can be a cell or tissue. The extracted protein (mixture) will be then separated in to individual protein species. Then, the protein species of interest will be identified and its (their) function(s) determined. This is the common flow of proteomics research.

Protein extraction from samples

Protein extraction from a sample follows the following procedure (Liebler, 2002):

1. The sample is usually pulverized, homogenized, sonicated, or otherwise disrupted to yield a soup that contains cells, subcellular components, and other biological debris in an aqueous buffer or suspension.
2. Then proteins are extracted from this soup by a number of techniques. For proteomic analysis, the objective is to recover as much of the protein as possible with as little contamination by other biomaterials (e.g., lipids, cellulose, nucleic acid, etc.) as possible. This is generally done with the aid of (Liebler, 2002):

a) Detergents: (e.g., SDS, 3-([3-cholamidopropyl]dimethylammonio)-1-propane sulfonate (CHAPS), cholate, Tween), which help to solubilise membrane proteins and aid their separation from lipids.
b) Reductants: (e.g., dithiothreitol [DTT], mercaptoethanol, thiourea), which reduce disulfide bonds or prevent protein oxidation
c) Denaturing agents: (e.g., urea and strong acids), which disrupt protein-protein interactions, secondary and tertiary structures by altering solution ionic strength and PH.
d) Enzymes: (e.g., DNAse, RNAse), which digest contaminating nucleic acids, carbohydrates, and lipids.

Protein identification using Mass Spectrometry (MS) technique

The common flow of protein identification is: 1) the protein of interest, which is, in many cases separated by 2D electrophoresis, is either enzymatically or chemically cleaved, 2) the obtained peptide mixture is analyzed by mass spectrometric techniques, and 3) the obtained peptide mass fingerprint is subsequently compared to “virtual” fingerprints obtained by theoretical cleavage of protein sequences stored in databases and the ones with more number of peptide similarity are retrieved as possible candidate protein (Gevaert and Vandekerckhove, 2000). Before the MS instrument was applied in proteomics, protein sequencing was carried out with Edman technique (Smith, 2001; Steen and Mann, 2004). However, nowadays, this method is very occasionally used in proteomics (Jorrin-Novo, 2014). Edman sequencing relies on a stepwise cleaving (fragmenting) of peptides from the amino terminus using chemicals in a process that may take hours or days. This method needs sample purification, requires much expertise, fails completely if the peptide was acetylated at its amino terminus or otherwise was blocked to Edman reaction, which requires a free amino terminus, and often no sufficiently long and unambiguous peptide sequence could be identified by this method (Steen and Mann, 2004). However MS does not require peptide to be purified to homogeneity, it does not have a problem identifying blocked or modified proteins, it is much more sensitive, and can fragment the peptides in seconds (Steen and Mann, 2004). According to De-Hoffmann and Stroobant (2007), unequalled sensitivity, detection limits, speed and diversity of its applications have raised Mass Spectrometry to an outstanding position among other analytical techniques. Hence, since the start of the application of Mass Spectrometry in protein chemistry, peptide mass fingerprinting (PMF) analysis has become the method of choice in high throughput protein identification (Gevaert and Vandekerckhove, 2000).

The first step in protein identification as indicated earlier is to separate protein species from the extracted protein mixture. This can be done by gel method or gel free method. The gel method refers to sodium dodecyl sulfate one dimensional polyacrylamide gel electrophoresis (SDS-1D-PAGE) or sodium dodecyl sulfate two dimensional Polyacrylamide electrophoresis (SDS-2D-PAGE) while the gel free(also known as LC based) method refers to using liquid chromatography (LC) technique (Jorrin-Novo , 2014). However, SDS-2D-PAGE is mostly applied in proteomics since it has one big advantage; that is, it can serve as image maps to allow investigators to compare changes in the proteome based on changes in the patterns of spots on the gel (Liebler, 2002). But for low abundant protein species, HPLC (high performance liquid chromatography) proves to be better technique (Liebler, 2002).

Once the protein species are separated the protein(s) of interest is (are) hand or machine picked and peptide finger printing carried out using mass spectrometry technique. The most common technique is the matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) where MALDI is the source and TOF is the mass analyzer (Liebler, 2002). MS proteins finger printing can be done or the protein is digested in to peptides and peptides finger printing done. The former approach is known as top down proteomics or protein centric proteomics and the latter one is referred as bottom up proteomics or peptide proteomics (Jorrin-Novo, 2014). Peptide proteomics is the most preferred one due to two reasons: 1) currently, MS instruments are well-suited to the analysis of peptides rather than proteins, and 2) the data obtained from MS analysis of peptides can be taken directly for comparison to protein sequences derived from protein and nucleotide-sequence databases (Liebler, 2002). Peptide proteomics is also possible by obtaining the peptides from Protein mixture directly rather than from separated protein species. When peptide obtained directly from the protein mixture is used for peptide proteomics it is known as shotgun proteomics (Alves et al., 2007).

As has been discussed earlier, protein/protein mixture digestion is an important step in proteomics. Hence, how protein/protein mixture is digested needs to be understood. The ideal protein digestion approach would cleave proteins at certain specific amino acid residues to yield fragments that are most compatible with MS analysis. Specifically, peptide fragments of between 6 and 20 amino acids are ideal for MS analysis and database comparisons. Peptides shorter than 6 amino acids generally are too short to produce unique sequence matches in database searches. On the other hand, it is difficult to obtain sequence information from peptides longer than 20 amino acids in tandem MS analyses. Thus, the objective of protein digestion will be to produce the highest yield of peptides of optimal length for MS analysis (Liebler, 2002).

The digestion of proteins to peptides is done by the use of enzyme proteases or chemicals. Trypsin is the most commonly used enzyme to digest proteins in to peptides. Trypsin cleaves proteins at lysine and arginine residues, unless either of these is followed by a proline residue in the C-terminal direction. The spacing of lysine and arginine residues in many proteins is such that many of the resulting peptides are of a length well-suited to MS analysis. But there are other proteases as well used to digest proteins. These are the Chymotrypsin (cleaving at W-, /Y-, /F-, \P), Glu C/V8 protease (cleaving at E-, /D^{a-}, \P), Lys C (cleaving at /K-, \P) and Asp N (cleaving at /D). Proteins also can be cleaved with some chemicals. The most widely used of these is cyanogen bromide (CNBr), which cleaves proteins at methionine residues (Liebler, 2002).

The next step in peptide proteomics is to know the mass of the peptides. This is done using mass spectrometer. A mass spectrometer always contains the following elements (De-Hoffmann and Stroobant, 2007); a sample inlet to introduce the compound that is analyzed (for example a liquid chromatograph or a direct insertion probe), an ionization source to produce ions from the sample, one or several mass analyzers to separate the various ions, a detector to 'count' the ions emerging from the last analyzer, and finally a data processing system that produces the mass spectrum in a suitable form. Some mass spectrometers combine the sample inlet and the ionization source and others combine the mass analyzer and the detector.

A mass spectrometer should always perform the following processes (De-Hoffmann and Stroobant, 2007):

1. Produce ions from the sample in the ionization source.
2. Separate these ions according to their mass-to-charge ratio in the mass analyzer.
3. Eventually, fragment the selected ions and analyze the fragments in a second analyzer.
4. Detect the ions emerging from the last analyzer and measure their abundance with the detector that converts the ions into electrical signals.
5. Process the signals from the detector that are transmitted to the computer and control the instrument through feedback.

Ionization of samples is the basic step in mass spectrometry. Ionization is the production of gas phase ions of the compound, for example by electron ionization: $M + e^- \rightarrow M^{*+} + 2e^-$ (De-Hoffmann and Stroobant, 2007). There are several kinds of ionization techniques. Most of these techniques were not effective in proteomics since they generate high energy and result in extensive fragmentation (Steen and Mann, 2004; De-Hoffmann and Stroobant, 2007). This problem was solved by the invention of 'soft' ionization techniques such as matrix-assisted laser desorption/ionization (MALDI) and electro spray ionization (ESI). The latter technique earned its

Table 3. Some of mass analyzers and their principle of separation (De-Hoffmann and Stroobant, 2007).

Type of analyzer	Symbol	Principle of separation
Electric sector	E or ESA	Kinetic energy
Magnetic sector	B	Momentum
Quadrupole	Q	*m/z* (trajectory stability)
Ion trap	IT	*m/z* (resonance frequency)
Time-of-flight	TOF	Velocity (flight time)
Fourier transform ion cyclotron resonance	FTICR	*m/z* (resonance frequency)
Fourier transform orbitrap	FT-OT	*m/z* (resonance frequency)

inventor a share of the Nobel Prize for chemistry in 2002 (Steen and Mann, 2004). Soft ionization transfers highly polar, completely non-volatile molecules with a mass of tens of kDa (such as protein) into the gas phase without destroying them (Steen and Mann, 2004). Therefore MALDI and ESI are the very important components of proteomics techniques (Liebler, 2002; Jorrin-Novo, 2014). The ionized molecule then gets to the mass analyzer. Several types of mass analyzers have been developed. The separation of ions according to their mass-to-charge ratio can be based on different principles (Table 3). All mass analyzers use static or dynamic electric and magnetic fields that can be alone or combined. Most of the basic differences between the various common types of mass analyzer lie in the manner in which such fields are used to achieve separation (De-Hoffmann and Stroobant, 2007). For example, it would be important to see the principle by which a quadruple mass analyzer separates ions. The following two principles are the bases; 1) an ion with higher atomic mass will resist the magnetic field and reaches the detector with less deflection while an ion with lower atomic mass resists the magnetic field less and hence is deflected and may collide with the analyzer, will be neutralized and hence will not be detected. 2) an ion with more charge will be attracted by the magnetic field hence is deflected, may collide with the analyzer, will be neutralized and hence will not be detected. But an ion with less charge is attracted less and deflects less so reaches the detector. Meanwhile, a mass spectrometer detects the mass to the ratio of charge (M/Z) of molecules. Therefore in order to allow all molecules under study to be separated and lastly detected by the detector, collision of ions with the analyzer has to be controlled by oscillating or controlling the magnetic field strength. This magnetic field adjustment process is done in fraction of seconds.

Next to the mass analyzer, the ions will get to the detector. Upon reaching the surface of the detector, the ions will produce current that is proportional to their abundance. The detector detects all the ions and their isotopes. The relative abundance of an ion is depicted on the MS graph as intensity and is represented on the y-axis. Relative abundance of a compound to be analyzed (peptide in proteomics) will be 100% while its fragments will have relative abundance less than 100%. For example the relative abundance of $[H_2O]^+$ in the MS graph below is 100% while the other fragments are less than this. Below Figure 4 shows simple illustration of mass spectrometer by using H_2O as an example and quadruple mass analyzer.

Now the M/Z of the peptide is determined using MS technique what remains is to determine their identity. This is done by searching on the database. Usually the data produced in proteomics is enormous so there are data reduction algorithms and softwares. Likewise, the user begins by selecting the database(s) to be searched. Both protein and/or gene sequence databases may be specified (gene sequences are translated if the latter is selected). An excellent, widely used protein sequence database is the SWISS-PROT database. Other widely used protein sequence databases are the OWL and NCBInr databases. The user then can provide information about the origin of the sample to limit the search to relevant organisms. For example, a sample from wheat proteins can be searched against all organisms, against all plants sequences, against cereals, or most specifically against wheat sequences. Specificity is advantageous because it can limit the number of comparisons to be made with the data and because it can limit the number “false” hits in other organisms. In addition to these features, the user may also enter a molecular-weight range for proteins to be searched. This again limits the number of comparisons to be made (Liebler, 2002).

Next, the user can indicate the enzyme used to cleave the proteins (e.g., trypsin) and specify the possible numbers of “missed cleavages”. These missed cleavages result from incomplete digestion by the enzyme. The matching algorithms thus can generate entries for such peptides, in case they are present in the sample. Finally, the user can specify a number of standard modifications to peptides that can be considered in the matching algorithm. For example, tryptic-digestion protocols usually involve a reduction and alkylation of cysteine thiols with iodoacetamide or iodoacetate, which changes the masses of the cysteine residues within peptides. In addition, free cysteine thiols may undergo modification with acrylamide during SDS-PAGE. The user can also specify common modifications such as phosphorylation, sulfation, glycosylation, and N-terminal modifications. All

these user defined modifications allow the program to generate mass matches for both modified and unmodified versions of the peptides in a database. MS data for both modified and unmodified versions of a particular peptide can thus be matched to a database entry. The user then can enter the measured *m/z* values from the MS data or specify an MS data file to be evaluated automatically. Finally, the user can enter a desired mass tolerance to control how closely the matches between MS *m/z* values and calculated *m/z* values must correspond to be "hits" (Liebler, 2002).

Once the user clicks "Go", the software begins by pre-filtering the database to be used. For example, if wheat was specified as the species to be searched, all non-wheat entries are excluded. If a protein mass range of 2,000 to 100,000 was selected, all proteins with masses outside this range are excluded. Then the remaining sequences in the databases are subjected to a virtual digestion with the enzyme specified. If missed cleavages are allowed, the list of peptides will include those resulting from incomplete digestion. Versions of the peptides bearing the user-specified modifications are also generated. Finally, the entire list of peptides are ranked by mass (or *m/z* values) and each *m/z* signal in each spectrum is then compared to this list. All matches within the user-specified mass tolerances are recorded as "hits" and used for the calculation of scores and identification of corresponding proteins (Liebler, 2002).

Protein identification is built around these essential fact; most peptide sequences of approximately six or more amino acids are largely unique in the proteome of an organism. Put another way, a typical six amino acid peptide maps to a single gene product (Liebler, 2002) and considering all amino acid sequence combinations that are theoretically possible, only a very minor portion of protein sequence is realized in nature, therefore a short peptide sequence is already highly protein specific (Seidler et al., 2010). Thus, if we can obtain the sequence of the peptide or if we can accurately measure its mass, we can identify the protein it came from simply by finding its match in a database of protein sequences (Liebler, 2002; Seidler et al., 2010).

De novo peptide sequencing

Following the above listed steps and searching in to the database there may not be protein match found. In this circumstance denovo sequencing of peptides is mandatory. It should be remembered that genome sequences of most organisms are still unknown (Standing, 2003; Agrawal et al., 2012). Therefore, denovo sequencing is important for those samples with unknown genome sequence (Standing, 2003; Seidler et al., 2010). Even for those samples which have known genome sequence, denovo sequencing is mandatory for the analyses of protein sequence variants or their splice isoforms and for peptides containing non-proteinic or modified amino acids as it is known as peptides with posttranslational modification (Standing, 2003; Seidler et al., 2010).

To carry out denovo peptide sequencing, one has to know three facts. These are peptide nomenclature or otherwise known as the Roepstorff-Fohlmann-Biemann nomenclature, peptide cleavage, and how peptide fragments appear on the MS graph. Figure 3 shows the Roepstorff-Fohlmann-Biemann nomenclature. So when peptides are fragmented they have all the possible fragmented ions of a, b, c and x, y, z to be formed. All these will be analyzed by mass spectrometer to give us an MS graph with all the ions arranged according to their M/Z. However if cleavage takes place only at a designed ion, the MS graph will depict only those ions of desire. As can be noted on Figure 4 for H_2O, all the possible cleavages are seen on the MS graph and the MS graph puts all these ions in an ascending order of M/Z.

Therefore, theoretically, if one can fragment peptides on the peptide bonds, the only ions to be depicted on the MS graph will be the "b" and "y" ions. For instance, if a peptide ITHR is fragmented at its peptide bond, $^{y}_{1}$ will be R, $^{y}_{2}$ will be HR, $^{y}_{3}$ will be THR and $^{y}_{4}$ will be ITHR. These ions will appear on the MS graph in an ascending order hence separated by only one peptide mass. So we can sit down and select the ions on the MS graph carefully and subtract ($^{y}_{2}$-$^{y}_{1}$), ($^{y}_{3}$-$^{y}_{2}$) and ($^{y}_{4}$-$^{y}_{3}$). In trypsic peptide, since the trypsin cleaves peptides at the argenin(R) or lysine (L) terminus, determining $^{y}_{1}$ is simple. Then the result obtained from the subtraction of ions will be compared to amino acid masses indicated on Table 1. This way, denovo sequencing is achieved.

This is what is done in peptide sequencing using MS technique. Amino acid residues of peptides are determined by tandem mass spectrometry or otherwise known as MS/MS. In tandem mass spectrometry two or more mass analyzers are used to determine masses of peptides and each amino acid. The most common tandem mass spectrometer in denovo sequencing is the triple quadruple MS/MS instrument (Figure 5). What a triple quadruple MS/MS instrument does is, on the first mass analyzer, by adjusting the magnetic field; it will select a desired peptide to move to the next mass analyzer. Then in the next analyzer, although mass spectrometers must function under high vacuum (low pressure) to rule out collision and fragmentation of ions, in MS/MS there is deliberate collision (De-Hoffmann and Stroobant, 2007). This collision will fragment the peptides along their peptide bond (Standing, 2003; Steen and Mann, 2004). This technique is the so called collision induced desiccation (CID) technique (Liebler, 2002; Standing, 2003; Steen and Mann, 2004). Therefore, in this process, since the peptides fragment predominantly at the peptide bond (although it must be noted that other fragmentations, such as internal cleavages, secondary fragmentations, etc. do occur, thus complicating the mass

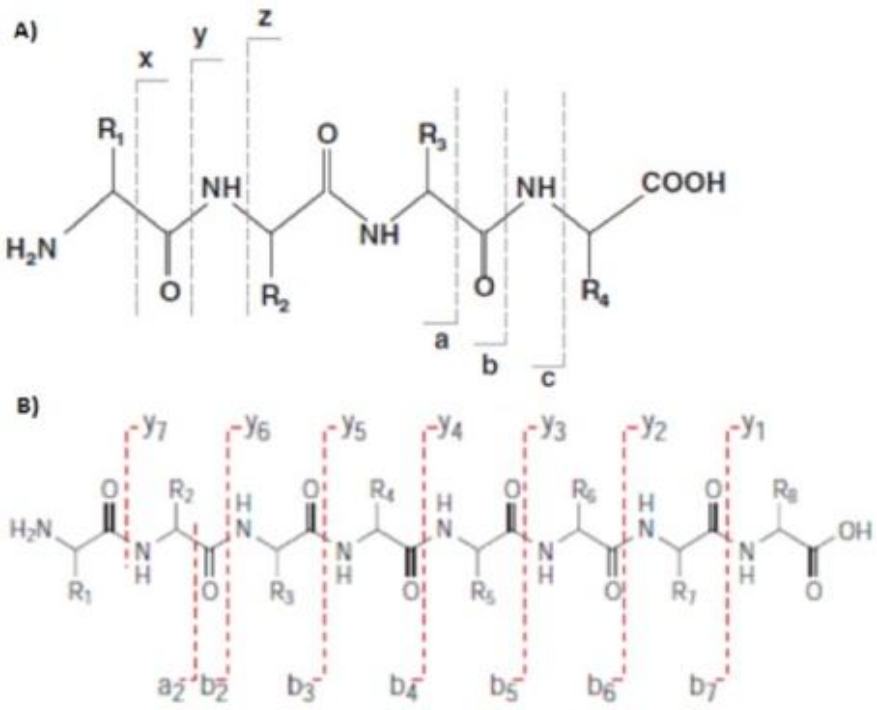

Figure 3. Roepstorff-Fohlmann-Biemann nomenclature. A) a peptide with three amino acids; a, b, and c are ions from the amino terminus and x, y, and z are from the carboxyl terminus (Seidler et al., 2010) and B) a peptide with 8 amino acids; b_2 is at the amino terminus and contains 2 amino acids while y_6 is its carboxyl terminus counterpart containing 6 amino acids (Steen and Mann, 2004).

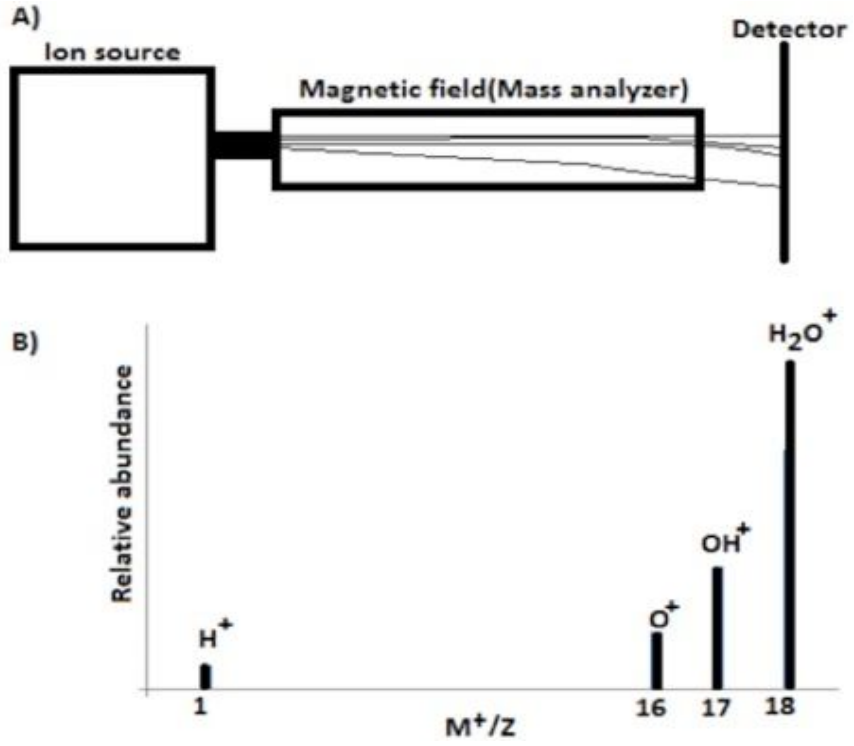

Figure 4. Simple illustration of mass spectrometer and MS graph using H_2O as an example (JEOL, 2006).

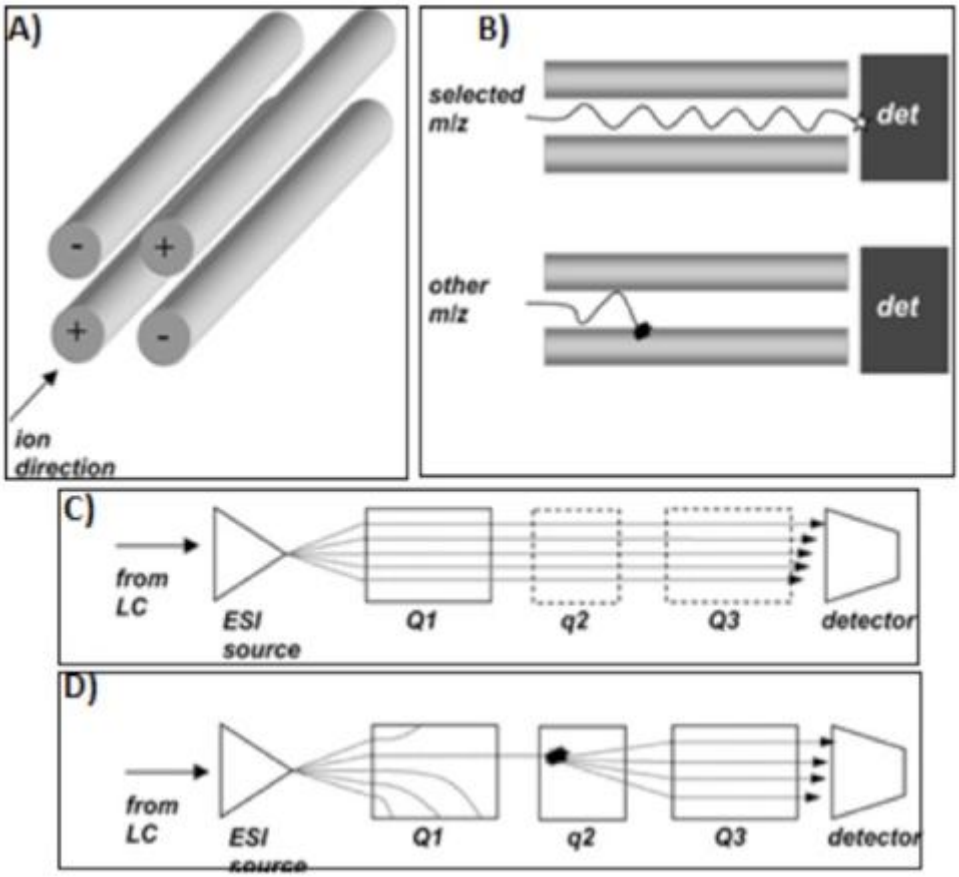

Figure 5. Schematic representation of a triple quadrupole MS instrument. (A) a quadrupole mass analyzer; (B) the trajectories of an ion of the selected *m/z* with that of ions of other *m/z*; (C) operation of the triple quad in full-scan mode; (D) operation of the triple quad in MS-MS mode (Liebler, 2002).

spectrum), the peptide fragments produced and displayed on the MS graph differ by the mass of one amino acid residue. The amino acid sequence can thus be readily deduced (Aitken, 2010).

Once the denovo peptide sequence is determined, identifying the protein will be possible. Considering all amino acid sequence combinations that are theoretically possible, only a very minor portion of protein sequences is realized in nature, and therefore a short peptide sequence is already highly protein-specific. This situation effects that a database-supported, probability-based annotation of peptide MS/MS spectra leads to protein identification at a high level of confidence from fragmentary sequence information (Seidler et al., 2010).

Protein function determination

Protein functions derive from proteins' three dimensional structure, which is in tern specified by its amino acid sequence (Lodish et al., 2013). It is experimentally confirmed that there is a relationship between the amino acid sequence, protein tertiary structure and function of proteins (Alberts et al., 2002; Lodish et al., 2013). The Nobel Prize winning experiment by Christian Anfinsen in 1960 revealed that a protein which is chemically denatured is able to automatically resume its original 3D shape once it is carefully returned to normality (Figure 6). Therefore, having the sequence of the protein is enough to determine its function.

This is when sequence homology is found with a protein of known function, either from the same or different species (Walker, 2010). For example, many key proteins involved in processes such as membrane transport and signal transduction were first identified *in silico*[1] by searching the databases for gene products with sequence similarity to their counterparts (which were well characterized in many cases) in model organisms such as *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans* and various mammalian species (Lam and Blumwald, 2002). This is called homologue protein search.

[1]*In silico* means performed on computer or via computer simulation with models closely reflecting the real world.

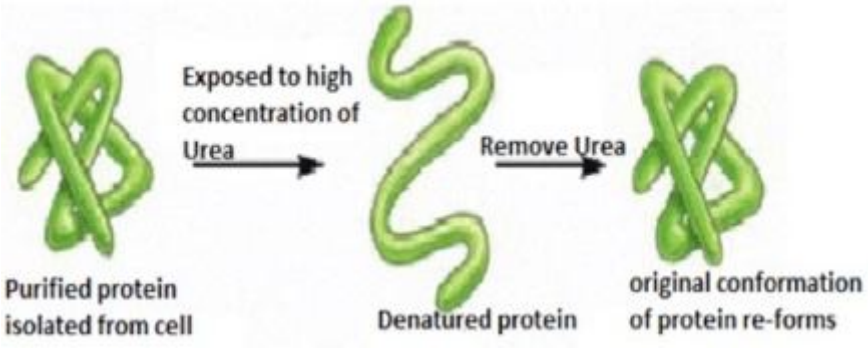

Figure 6. The refolding of a denatured protein. This experiment demonstrates that the conformation of a protein is determined solely by its amino acid sequence (Alberts et al., 2002).

However, this approach does not always work. For example, when the genome of the yeast *Saccharomyces cerevisiae* was completely sequenced in 1996, 6000 genes were identified. Of these, approximately 2000 coded for proteins that were already known to exist in yeast (that is, had been purified and studied in previous years), 2000 had homology with known sequences and hence their function could be deduced by the homology method. But 2000 could not be matched to any known genes, that is, they were 'new', previously undiscovered genes. In these cases, there are a number of other computational methods that can be used to help identify the protein's function (Walker, 2010). These include:

1. Phylogenic profile method: This method aims to identify any other protein(s) that has (have) the same phylogenic profile (that is, the same pattern of presence or absence) as the unknown protein, in all known genomes. If such proteins are found, it is inferred that the unknown protein is involved in the same cellular process as these other protein(s).
2. Method of correlated gene neighbors: If two genes are found to be neighbors in several different genomes, a functional linkage may be inferred between the two proteins.
3. Analysis of fusion: This method is based on the observation that two genes may exist separately in one organism, whereas the genes are fused into a single multifunctional gene in another organism. The existence of the protein product of the fused gene, in which the two functions of the protein clearly interact (being part of the same protein molecule), suggests that in the first organism the two separate proteins also interact.
4. Protein–protein interactions: Given the complex network of pathways that exist in the cell (signaling pathways, biosynthetic pathways, etc.), it is clear that all proteins must interact with other molecules to fulfill their role. Indeed, it is now apparent that proteins do not exist in isolation in the cell; proteins involved in a common pathway appear to exist in a loose interaction, sometimes referred to as a biomodule. Therefore, if one can identify an interaction between our unknown protein and a well characterized protein, it can be inferred that the former has a function somehow related to the latter. The use of protein microarray in this case is instrumental.

However if all the above search methods do not provide the desired data and our peptide sequence is completely unique, the next step is to build the 3D models or determine function from scratch. This procedure is known as *ab initio* modeling or *de novo* modeling (Lee et al., 2009). For detailed concept and methods in *ab initio* modeling readers are directed to (Lee et al., 2009).

Protein microarrays

Protein microarrays are homogeneous or heterogeneous protein micro spots in rows and columns that are bound to (or immobilized at) a solid surface mostly glass (Espina et al., 2003; Walker, 2010; Chen et al., 2014). The immobilized proteins are probed against a labeled protein. Then the complex formation within each micro spot is detected based on fluorescence, chemiluminescence, mass spectrometry, radioactivity or electrochemistry indicating the interaction between various proteins (Chen and Zhu, 2006).

Protein microarrays are of two types; analytical and functional protein microarrays (Chen and Zhu, 2006) or similarly referred as forward phase protein microarrays and reverse phase protein microarrays respectively (Chen et al., 2014). Analytical protein microarray contain a single(homogeneous) immobilized protein to be probed against multiple proteins while functional protein microarray is the reverse; containing multiple (heterogeneous) immobilized proteins to be probed against singe protein (Chen and Zhu, 2006; Chen et al., 2014).

Traditional protein microarray generation starts by in vitro expression of the proteins in cells, followed by

protein purification, transfer, and immobilization onto the microarray solid support. This procedure has to be repeated for each protein separately, an enormous logistic effort in terms of cell culturing and protein purification. Even if the logistic is solved, there are other technological obstacles in protein microarray installation (Kilb et al., 2014). Hence, the generation of whole proteome microarrays is mainly technically demanding and can be realized by well-established research groups or international companies (Walker, 2010; Kilb et al., 2014).

Protein microarrays have similar principle as that of DNA microarrays[2] but differ quite significantly in their application. DNA microarrays are useful in genotyping. Furthermore, their applications to identify transcription factor binding sites and protein-DNA interactions were reported. However, only protein microarrays allow phenotyping, decoding signaling cascades, identifying protein modifications, proofing enzyme substrate specificity, phosphorylation, and binding patterns for many proteins and antibodies in parallel (Kilb et al., 2014). Nowadays, protein microarrays have been widely applied as a promising proteomic technology with great potential for protein expression profiling, biomarker screen, drug discovery, drug target identification and analysis of signaling pathways in health and disease (Chen et al., 2014).

THE GENESIS AND STATE OF PLANT PROTEOMICS

As mentioned earlier, it was Marc Wilkins in 1994 that came up with the word "proteome" while working on the concept as a PhD student (Tyagi et al., 2010). Therefore, the genesis of Proteomics dates back to 1994. However, in practice, plant proteomics and hence, proteomics emerged a bit earlier. In 1989, a paper showing the effects of cold treatment on protein synthesis and mRNA levels in rice leaves was known to have been published (Agrawal et al., 2012). Therefore, we can safely say that plant proteomics started in the late 1980s. Real plant proteomics however, genome wide plant proteins study, was only possible after the genome sequence of a plant, *Arabidopsis thaliana*, was for the first time sequenced in the year 2000.

The inability of genomics to tell us more than proteomics about plant biological processes has fueled the need for plant proteomics over the past years. Although DNA microarray enabled the analysis of genome-wide patterns of gene expression they cannot tell us what that gene does because it is the proteins not the mRNA that provides the majority of plant biological activities (Bonjoch and Tamayo, 2003; Öpik and Rolfe, 2005). Posttranslational modification of proteins, modification of proteins after they come out of the mitochondria is also one of the main reasons why proteomics is superior over genomics (Liebler, 2002). It is posttranslational modifications of proteins that explain why there is high number of unique protein molecules, far exceeding the number of respective genes, particularly in eukaryotes (Tambor et al., 2010).

Only recently, in 2001, it was used to be believed that Edman chemistry was a standard method for peptide sequencing (Smith, 2001). However, the advances in MS, revolutionized peptide sequencing and as a result, has brought about the major progress in proteomics/plant proteomics (Jorrin-Novo, 2014). Moreover, continuous improvements in protein extraction, purification and separation, developments of bioinformatics tools and algorithms for data analysis, protein identification, quantification and characterization have also enabled the advancement of plant proteomics (Jorrin-Novo, 2014).

The journey and status of plant proteomics have been reviewed by (Van-Wijk, 2001; Agrawal et al., 2012), and more extensively, by Jorrin-Novo et al. (2014). In 1999, Thiellement and colleagues published the first review of plant proteomics and extensively discussed the plant proteomics literature before 1999 (Van-Wijk, 2001). Accordingly, until 1999, plant proteomics studies did not involve MS and therefore were limited to the comparison of expression levels without actual identification of the proteins. In a few cases, limited sets of proteins were identified through Edman sequencing (Van-Wijk, 2001). Agrawal et al. (2012) also indicated as of 2000, the number of plant proteomics reports published were very minimum. But after 12 years, in 2012, there were several books on plant proteomics and several special issues on plant proteomics published by the journals Proteomics, Phytochemistry, and Journal of Proteomics (Agrawal et al., 2012). Moreover, there are several review series in plant proteomics (Agrawal et al., 2012). These indicated the very fast progression in plant proteomics. However, one important aspect of plant proteomics is still way behind. That is the translational plant proteomics. For instance, as of 2012, there were no reported success stories whereby the knowledge gained directly from proteomics studies has been applied to improve crop stress tolerance (Agrawal et al., 2012).

Status of plant proteomics in Africa

To find Information about the state of proteomics particularly plant proteomics in Africa is a very daunting task. This quite well could be due to the limited number of

[2] Contains thousands of “spots”. Each spot contains many copies of the same DNA sequence that uniquely represents a gene. RNA is extracted from the sample, and is labeled with different dyes during the synthesis of cDNA by reverse transcriptase. Then the cDNA is hybridized onto the microarray slide, where each cDNA molecule representing a gene will bind to the spot containing its complementary DNA sequence. The microarray slide is then excited with a laser at suitable wavelengths to detect the dyes; in other words, the comparative expression of genes. In so doing identification of the responsible genes is possible.

plant proteomics works that maybe available. According to Ndimba and Thomas (2008), as of 2008, there were no articles published on proteomics in all of Africa. Current data also show that proteomics and specifically plant proteomics is at the inception stage in Africa. For instance, the research outputs by Centre for Proteomic and Genomic Research (CPGR), South African genomics and proteomics center, are minimal and none deal on plant proteomics (http://www.cpgr.org.za/research-outputs/ cited on 28/02/2016). However, a more comprehensive review on proteomics and specifically plant proteomics developments in Africa over the past eight years, after 2008, need to be intensively reviewed.

THE APPLICATION OF PROTEOMICS IN PLANT ECOPHYSIOLOGY

Plants unlike animals are sessile and are constantly bombarded with changes in their environment (Esmon et al., 2005). Temperature fluctuations, poor light and low water content in the soil are just few of the factors to which plants must be able to respond continuously (Esmon et al., 2005). Moreover, plants must respond to physical forces of nature such as gravity or touch stimulation (Esmon et al., 2005). Plant ecophysiology deals with understanding and studying plants physiological adaptive responses to these myriads of environmental conditions.

As has been discussed earlier, proteins have plethora of roles in plant physiology. Meanwhile, studying proteins directly is undoubtedly the preferred methodology in plant ecophysiological studies. Most importantly, post-translational modifications, protein-protein interactions and enzymatic activities which are the direct effectors of plants ecophysiological responses can only be studied by proteomics and not genome expression studies (Kottapalli et al., 2009). For instance, phosphorylation, a posttranslational modification, is well known to trigger either activation or inactivation of stress signaling cascade in plants (Sanchez-Moreiras and Bonjoch, 2003). The phosphorelation of photosystem (II) protein complexes is also known to be an excellent mechanism protecting it from photo damage when exposed for prolonged excess light (Tikkanen, 2009). Several studies have, meanwhile, proven that the changes in gene expression at transcript level do not often correspond with the changes at protein level. Therefore, investigation of changes in plant proteome is highly important since proteins, unlike transcripts, are direct effectors of plant environmental responses (Kosova et al., 2011). For instance, a large number of genes that are implicated in the drought response have already been identified but determining which are most useful for breeding drought resistant crop varieties has been almost impossible due to lack of the proteomics (Deeba et al., 2012). Proteins also regulate transcription and translations; hence they regulate plant environmental response even at transcript levels (Kosova et al., 2011).

Biotic stresses are of great importance in plant ecophysiology. To gain an in depth understanding of plant-pathogen interactions, numerous studies have been carried out from the plant as well as from the pathogen perspectives. These studies reveal that plant-pathogen interaction results from precise communication between the plant and the invading pathogen. Hence, if compatible communication takes place, plants are incapable of mounting effective anti infectious defense responses, allowing the pathogens to complete their life cycle. If incompatible communication takes place, plants trigger a series of complex defense responses against pathogenic interactions to forestall pathogen growth. Understanding these plant-pathogen interactions is hardly possible by conventional biochemical and genetic experimental methods. Currently, proteomics provides a comprehensive insight to understand the intricate plant-pathogen interactions (Lodha et al., 2013). Meanwhile a pathogen disguises the host to forge a compatible communication. Therefore through proteomics, how a pathogen disguised the host can be detected and hence, possibly corrected. Proteomics is also very vital to know the protein dynamics of plants at different levels of disease stages thereby enabling the identification of protein markers at the early stage of the disease development. This will ensure the arrest of disease spread as potential disease plant(s) can be culled out early. In such a manner, Sghaier-Hammami et al. (2012) employed proteomics to try to identify brittle leaf disease biomarkers for the economically important Tunisian date palm.

Therefore, proteomics has irreplaceable role in plant ecophysiology since; it enables us understand plant functions better, it enables us select plant species or ecotypes with the best desirable abiotic stress tolerant genotypes, it most importantly, helps us understand the pathogen-disease interaction better than other methods so that disease resistance of plants can be manipulated, it enables us identify early plant disease biomarkers for better control of disease spread. Detailed review on the proteomics of plants abiotic stresses and that of biotic stresses is available in Roveda-Hoyos and Fonseca-Moreno (2011), Kosova et al. (2011), Hakeem et al. (2012) and Lodha et al. (2013), respectively.

THE SHORTCOMINGS OF PROTEOMICS

The fact that proteomics is still capital and knowledge intensive technology is one of the shortcomings of proteomics (Jorrin-Novo et al., 2014). The nature of proteins, lack of ample DNA sequence data and other technical factors add up to the impediments in proteomics.

Proteins, if compared to nucleic acids, are molecular

entities of the greatest difficulty to work with. Hence, their dynamism and variability is more complex than what was thought at the early days of proteomics (Jorrin-Novo, 2014). Good sample quality is one of the critical factors for successful proteomic experiments and is challenging to obtain from plant tissues. An enriched level of proteases and oxidative enzymes in plant tissues make it extremely difficult to extract stable protein mixtures. Moreover, secondary metabolites produced in plant cells often interfere with subsequent protein fractionation and downstream analyses. Hence it is notoriously difficult to extract complete and representative protein population from plant tissues. Additional hindrance comes from the cell wall that is difficult to fragment (Ghosh and Xu, 2014). Proteomics is possible as long as genome sequence data is available. The number of fully sequenced genomes however is small (Agrawal et al., 2012) hindering the scope of proteomics. Technical drawbacks associated with 1-D and 2-D electrophoresis, mass spectrometer and protein microarrays are also the other technical challenges in proteomics.

CONCLUSION

Proteomics and plant proteomics are very young science and can still be very expensive. However, considering the myriads of superiority it has over other molecular studies, its economical implications maybe counterbalanced with its potential in novelizing biotechnology. From the scantly available information, it can be concluded that Africa is lagging behind in plant proteomics. This should not continue in the future. Considering the huge biological resource Africa has, the adoption and the adaption of plant proteomics shall not be considered as a luxury but a necessity. Africa needs plant proteomics to bioprospect the various plant species it has in the wild. Since plant biodiversity is the backbone of most African countries, thus halting some of the devastating pathogenic diseases of this resource may be one of the potential roles of plant proteomics in Africa.

REFERENCES

Adams J, **2008**. The proteome: Discovering the structure and function of proteins. Nat Educ, 1(3):1-6.

Agrawal GK, Pedreschi R, Barkla BJ, Bindschedler LV, Cramer R, Sarkar A, Renaut J, Job D, Rakwal R, **2012**. Translational plant proteomics: A perspective. J Proteomics, 7(5):4588–4601.

Aitken A, **2010**. Mass spectrometric techniques. K. Wilson and J. Walker (eds) Principles and Techniques of Biochemistry and Molecular Biology (7th edn). New York: Cambridge University Press. pp: 352-379.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P, **2002**. Molecular Biology of the Cell. 4th edn. New York: Garland Science

Alves P, Arnold RJ, Novotny MV, Radivojac P, Reilly JP, Tang H, **2007**. Advancement in protein inference from shotgun proteomics using peptide detectability. Pacific Symp Biocomput, 12: 409-420

Bonjoch NP, **Tamayo** PR, **2003**. Protein content quantification by Bradford method. Manuel J. Reigosa Roger (eds), Handbook of plant ecophysiology techniques. Kluwer academic publishers. pp 283-296.

Chen B, Liu Y, Shen Y, Xia Z, He G, Liang S, **2014**. Protein microarrays in proteome wide applications. J Proteom Bioinformat, 12: 1-6.

Chen SC, **Zhu** H, **2006**. Protein microarrays. BioTechniques, 40(4):423-429.

Deeba F, Pandey AK, Ranjan S, Mishra A, Singh R, Sharma YK, Shirke PA, Pandey V, **2012**. Physiological and proteomic responses of cotton (*Gossypium herbaceum* L.) to drought stress. Plant Physiol Biochem, 53: 6-18.

De-Hoffmann E, **Stroobant** V, **2007**. Mass Spectrometry: Principles and Applications (3rd edition). England: John Wiley & Sons Ltd.

Duan G, **Walther** D, **2015**. The roles of post-translational modifications in the context of protein interaction networks. PLoS Computational Biol, 11(2):1-23.

Eldakak M, Milad SIM, Nawar AI, Rohila JS, **2013**. Proteomics: a biotechnology tool for crop improvement. Front Plant Sci, 4:35.

Esmon CA, Pedmale UV, Liscum E, **2005**. Plant tropisms: providing the power of movement to a sessile organism. Int J Dev Biol, 49: 665-674.

Espina V, Mehta AI, Winters ME, Calvert V, Wulfkuhle J, Petricoin EF, Liotta LA, **2003**. Protein microarrays: Molecular profiling technologies for clinical specimens. Proteomics, 3:2091-2100.

Gevaert K, **Vandekerckhove** J, **2000**. Protein identification in proteomics. Electrophoresis, 21:1145-1154.

Ghosh D, **Xu** J, **2014**. Abiotic stress responses in plant roots: a proteomics perspective. Front Plant Sci, 5(6):1-13.

Glowacki S, Macioszek VK, Konowicz AK, **2011**. R proteins as fundamentals of plant innate immunity. Cell Mol Biol Lett, 16: 1-24.

Hakeem KR, Chandna R, Ahmad P, Iqbal M, Ozturk M, **2012**. Relevance of proteomic investigations in plant abiotic stress physiology. J Integrative Biol, 16(11):621-635.

JEOL, **2006**. Mass Spectrometer: A short explanation for absolute novice.USA: JEOL incorporated.

Jorrin-Novo JV, **2014**. Plant Proteomics Methods and Protocols.In: Jesus V. Jorrin-Novo et al. (eds.), Plant Proteomics: Methods and Protocols. Springer Science+Business Media. pp: 3-13.

Jorrin-Novo JV, Pascual J, Lucas RS, Romero-Rodriguez C, Ortega MR, Lenz C, Valledor L, **2014**. Fourteen years of plant proteomics reflected in “Proteomics”: Moving from model species and 2-DE based approaches to orphan species and gel-free platforms. Proteomics, 15(5-6):1089-112.

Kaldenhoff R, **2006**. Besides water: Functions of plant membrane intrinsic proteins and aquaporins. Progress Bot, 67: 206-218.

Kaldenhoff R, **2012**. Mechanisms underlying CO_2 diffusion in leaves. Curr Opin Plant Biol, 15: 276–281.

Kersten B, Agrawal GK Iwahashi H, Rakwal R, **2006**. Plant phosphoproteomics: A long road ahead. Proteomics, 6: 5517–5528.

Kilb N, Burger J, Roth G, **2014**. Protein microarray generation by *in-situ* protein expression from template DNA. Eng Life Sci, 14: 352–364.

Kosova K, Vitamvas P, Prasil IT, Renaut J, **2011**. Plant proteome changes under abiotic stresss - Contribution of proteomics studies to understanding plant stress response. J Proteom, 74:1301– 1322.

Kottapalli KR, Rakwal R, Shibato J, Burow G, Tissue D, Burke J, Puppala N, Burow M, Payton P, **2009**. Physiology and proteomics of the water-deficit stress response in three contrasting peanut genotypes. Plant Cell Environ, 32: 380-407.

Lam BCH, **Blumwald** E, **2002**. Domains as functional building blocks of plant proteins. Trend Plant Sci, 7(12):544-549.

Lee J, Wu S, Zhang Y. **2009**. Ab Initio Protein Structure Prediction. In: Rigden DJ (eds): From Protein Structure to Function with Bioinformatics. Springer Science + Business Media. pp: 3-19.

Liebler DC, **2002**. Introduction to Proteomics: Tools for the New Biology. New Jersey: Humana Press.

Lodha TD, Hembram P, Tep N, Basak J, **2013**. Proteomics: A successful approach to understand the molecular mechanism of plant-pathogen interaction. Am J Plant Sci, 4: 1212-1226.

Lodish H, Berrk A, Kaiser CA, Krieger M, Bretscher A, Ploegh H, Amon A, Scott, MP, **2013**. Molecular cell biology (7th edn). New York: W.H. Freeman and company.

Marschner H, **1993**. Mineral nutrition of higher plants. 2nd edn. Harcourt Brace & Company, Publishers. Academic Press.

Mathews CK, **Van holde** KE, **2002**. Biochemistry. Second edition. California: The Benjamin/Cummings Publishing Company.

Ndimba BK, **Thomas** LA, **2008**. Proteomics in South Africa: Current status, challenges and prospects. Biotechnol J, 3: 1368–1374.
Öpik H, **Rolfe** S, **2005**. The Physiology of Flowering Plants. 4th edn. New York: Cambridge University Press.
Roveda-Hoyos G, **Fonseca-Moreno** LP, **2011**. Proteomics: a tool for the study of plant response to abiotic stress. Agronomia Colombiana, 29(2): 221-230.
Salekdeh GH, Siopongco J, Wade LJ, Ghareyazie B, Bennett J, **2002**. A proteomic approach to analyzing drought-and salt- responsiveness in rice. Field Crop Res, 76:199-219.
Sanchez-Moreiras AM, **Bonjoch** NP, **2003**. Two-dimensional electrophoresis. Stress proteins. In: Manuel J. Reigosa Roger (eds), Handbook of plant ecophysiology techniques. Kluwer academic publishers. pp: 297-334.
Seidler J, Zinn N, Boehm ME, Lehmann WD, **2010**. De novo sequencing of peptides by MS/MS. Proteomics, 10: 634–649
Sghaier-Hammami B, Saidi MN, Castillejo MA, Jorriin-Novo jV, Namsi A, Drira N, Gargouri-Bouzid R, **2012**. Proteomics analysis of date palm leaves affected at three characteristic stages of brittle leaf disease. Planta, 236(5):1599-613.
Sinha RK, **2004**. Modern plant physiology. UK: Alpha science international limited.
Smith JB, **2001**. Peptide Sequencing by Edman Degradation. Encyclopaedia of Life Sciences. pp: 1-3.
Standing KG, **2003**. Peptide and protein de novo sequencing by mass spectrometry. Curr Opin Structure Biol, 13: 595–601.
Steen H, **Mann** M, **2004**. The ABC'S (and XYZ'S) of peptide sequencing. Nat Rev, 5: 699-711.
Taiz L, **Zeiger** E, **2002**. Plant Physiology. 3rd Edition. Massachusetts, Sunderland: Sinauer Associates.
Taiz L, **Zeiger** E, **2010**. Plant Physiology. 5th Edition. Massachusetts, Sunderland: Sinauer Associates.
Tambor V, Fucikova A, Lenco J, Kacerovsky M, Rehacek V, Stulik J, Pudil R, **2010**. Application of Proteomics in Biomarker Discovery: a Primer for the Clinician. Physiol Res, 59: 471-497.
Tikkanen M, **2009**. Thylakoid protein phosphorelation in regulation of photosynthesis. PhD thesis. University of Turku, Finland.
Tyagi S, Raghvendra, Singh U, Kalra T, Munjal K, Vikas, **2010**. Practical applications of proteomics-a technique for large scale study of proteins: an overview. Int J Pharmaceut Sci Rev Res, 3(1):87-90.
Tyers M, **Mann** M, **2003**. From genomics to proteomics. Nature, 422(6928):193-197.
Van-Wijk KJ, **2001**. Challenges and prospects of plant proteomics. Plant Physiol, 126:501-507.
Vince D, **1960**. Photomorphogenesis in plant stem. Biol Rev, 39:506-536
Walker J. **2010**. Protein structure, purification, characterisation and function analysis. In: Wilson K, Walker J (eds), Principles and Techniques of Biochemistry and Molecular Biology. (7th edn). New York: Cambridge University Press. pp: 300-351.
Wasinger VC, Zeng M, Yau Y, **2013**. Current status and advances in quantitative proteomic mass spectrometry. Int J Proteom: 1-12.
Wulfkuhle JD, Liotta LA, Petricoin EF, **2003**. Proteomic application for early detection of cancer. Nat Rev, 3:267-275.

Citation: Asmelash F, 2016. Techniques and applications of proteomics in plant ecophysiology. Biochem Biotechnol Res, 4(1): 1-16.