Chapter 3
Proteomics in Systems Biology

Wan Mohd Aizat and Maizom Hassan

Abstract Proteomics is the study of proteins, the workhorses of cells. Proteins can be subjected to various post-translational modifications, making them dynamic to external perturbation. Proteomics can be divided into four areas: sequence, structural, functional and interaction and expression proteomics. These different areas used different instrumentations and have different focuses. For example, sequence and structural proteomics mainly focus on elucidating a particular protein sequence and structure, respectively. Meanwhile, functional and interaction proteomics concentrate on protein function and interaction partners, whereas expression proteomics allows the cataloguing of total proteins in any given samples, hence providing a holistic overview of various proteins in a cell. The application of expression proteomics in cancer and crop research is detailed in this chapter. The general workflow of expression proteomics consisting the use of mass spectrometry instrumentation has also been described, and some examples of proteomics studies are also presented.

Keywords Expression proteomics · Enzyme · Mass spectrometry · Peptide · Protein · Shotgun proteomics

3.1 Introduction

The term "proteomics" was first coined 20 years ago in an effort to define the total proteins encoded by a given genome [1]. Such powerful term remains influential and has since expanded into various fields of research and organisms. The significance of proteomics comes from how important protein is in living cells. The genome of one organism is always static (unless mutation occurs), and yet proteins are expressed based on tissue types, and their expression may be changed upon stimulation of environmental/external conditions. Several post-translational modifications (PTMs)

W. M. Aizat (✉) · M. Hassan
Institute of Systems Biology, Universiti Kebangsaan Malaysia (UKM), Bangi, Malaysia
e-mail: wma@ukm.edu.my; maizom@ukm.edu.my

W. M. Aizat et al. (eds.), *Omics Applications for Systems Biology*,
Advances in Experimental Medicine and Biology 1102,
https://doi.org/10.1007/978-3-319-98758-3_3

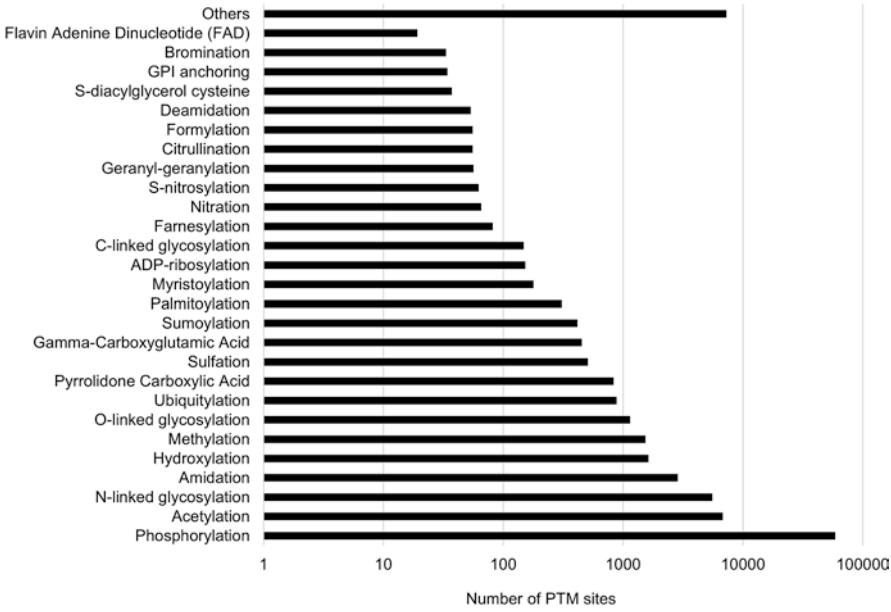

Fig. 3.1 Summary of the number of experimentally observed post-translational modification (PTM) sites documented in the Swiss-Prot database (data obtained from http://selene.princeton.edu/PTMCuration; accessed on July 2018) [3]

(Fig. 3.1) such as phosphorylation, acetylation, glycosylation, amidation, hydroxylation and methylation have also shaped and influenced certain proteins, and hence the level of gene expression may not always correlate with the protein and its activity level [2]. There are estimated more than 200 known PTMs which undeniably increase the proteome complexity of any living being [2, 3]. Furthermore, proteins are known to be the workhorses of cells as they are responsible for various cellular functions such as enzymatic reactions, signaling, gene transcription and translation processes, as well as structural components. This signifies the central role of proteins, and hence the study of proteomics is highly sought for a holistic understanding of cellular regulation.

3.1.1 Different Aspects of Proteomics

Generally, proteomics can be categorised into four distinct study areas, namely, "sequence proteomics", "structural proteomics", "functional and interaction proteomics" and "expression proteomics" (Fig. 3.2) [4]. These different proteomics areas tackle different aspects of protein properties, including primary and three-dimensional structure as well as function and protein abundance, respectively.

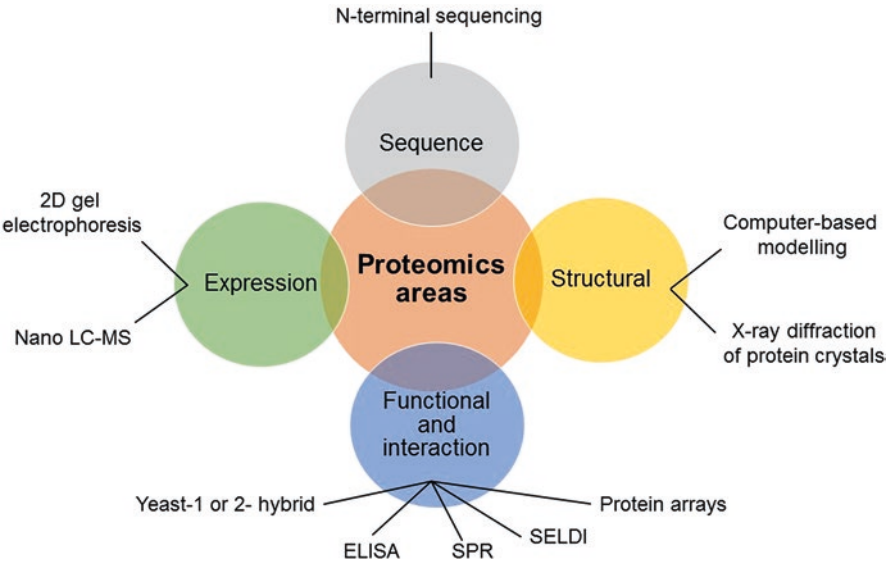

Fig. 3.2 Proteomics can be distinguished into four main aspects which are sequence, structural, functional and interaction as well as expression studies. Each of this aspect is given examples as detailed in text. Surface-enhanced laser desorption/ionisation (SELDI), enzyme-linked immunosorbent assay (ELISA), surface plasmon resonance (SPR), nano-liquid chromatography-mass spectrometry (nano-LC-MS)

Firstly is the "sequence proteomics" in which amino acid sequences in a given protein are determined. Historically, amino acid compositions were elucidated using Edman sequencing [5]. Briefly, this technique uses a chemical called phenyl isothiocyanate and a mild acid hydrolysis to tag and cleave specifically at the N-terminal of the chosen protein (Fig. 3.3). The amino acids "released" from the whole protein will be identified sequentially using chromatographic instruments such as high-pressure liquid chromatography (HPLC) to build the order of the protein sequences. Up till now, this technique is still considered as one of the most sensitive techniques for protein identification and can sequence down to 0.5–1.0 pmole of proteins [4]. However, this procedure requires a non-complex protein mixture as well as laborious and hence not practical to be used for a larger scale of protein identification. This ultimately requires a high-throughput system such as mass spectrometry (MS) which will be discussed later in this chapter.

The second proteomics area is called "structural proteomics". This study concerns protein structural identity to elucidate its putative function. Structural proteomics can be studied using several approaches including computer-based modelling as well as experimental methods such as protein crystallisation [6], nuclear magnetic resonance (NMR) and electron microscopy [6, 7]. X-ray diffraction of protein crystals is one of the most common techniques in elucidating protein structure in this area.

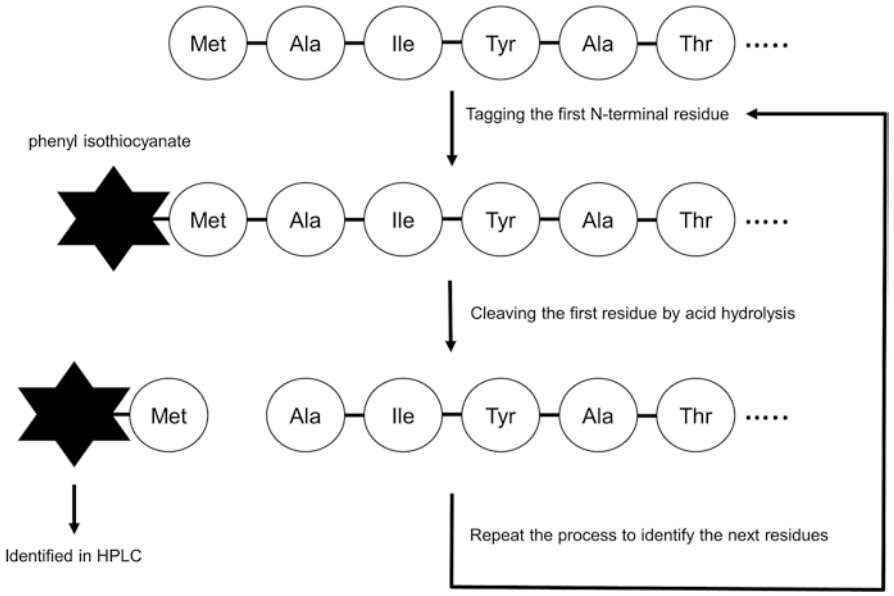

Fig. 3.3 Overall workflow of Edman N-terminal sequencing. The first residue of a protein will be tagged at the N-terminal using phenyl isothiocyanate before cleaved by acid hydrolysis. The identity of the released residue will be determined using analytical instruments such as high-pressure liquid chromatography (HPLC), and the process will be repeated to identify the next residue of the protein, sequentially

Protein functions and activities are studied in the third area of proteomics called "functional and interaction proteomics". Protein function particularly enzymes can be elucidated by examining their reactions in vitro [8]. Other techniques to determine protein function are also available, particularly based on the protein interaction with other proteins, ligands or substrates. Traditionally, techniques such as yeast -one/two-hybrid are a popular tool to elucidate protein-protein/DNA interaction [4]. The introduction of protein microarray experiments has opened a new door for mass characterisation of proteins and protein profiling [9, 10]. Protein post-translational modifications are also able to be detected using certain specific protein arrays [11]. These technologies are based on the interaction of proteins, antibodies and enzymes to other proteins and ligands. One drawback of this approach is it requires known antibodies/enzymes/proteins and hence can be considered as a more targeted proteomics approach.

Last but not least is the "expression proteomics" or also known as "discovery-based proteomics". This approach is useful in elucidating the expression of proteins in a global and untargeted manner. Most proteins in a complex sample can be identified and quantitated to provide an overall protein overview of any experimental samples. This will be highly advantageous for understanding the samples' protein composition as well as finding protein biomarkers. This area of proteomics mainly utilised MS instruments to allow high-throughput analysis of protein/pep-

tide samples. Recently, a proteome map draft of a human has been reported using these instruments which details proteins found in various parts of the body [12, 13]. This suggests that "expression proteomics" is greatly advantageous for a holistic and large-scale study at a system level.

3.1.2 Proteomics in the Context of Systems Biology

As one of the approaches in systems biology, proteomics has been utilised in an integrative approach, combining other omics such as genomics, transcriptomics and metabolomics in an effort to comprehensively understand certain biological questions. Rather than investigating isolated parts of genes/proteins in an organism as what traditional molecular biology have done, a system approach is more useful in characterising the dynamics and structure of a working biological system [14]. This will assist in developing biological models that could be tested upon series of perturbation experiments [15]. Understanding organisms at the proteome level will undeniably contribute to the rationale of the models, considering that proteins are highly modified (due to PTMs) and hence functionally diverse compared to the more static genome.

Due to the nature of systems biology study, not all the different techniques of proteomics can be fully utilised in this area. For example, "sequence proteomics" which is only dedicated for determining protein sequence at a small scale may not be feasible to be used extensively in a larger scale of system studies. On the other hand, "expression proteomics", which can catalogue total quantifiable proteins in any biological system, can be used as the starting platform for a global proteome analysis. Perturbation experiments in any given samples may also be investigated using this approach. One example is using stable isotope labelling by/with amino acids in cell culture (SILAC) which labels specific amino acids with either light or heavy isotopes to investigate the level of protein differences between normal and treated cell lines [16]. Further experimentation can then be employed to characterise any proteins of interest using a more targeted proteomics approach such as in "structural proteomics". Elucidating the protein structure will give an insight to its active sites and how this protein contributes to a given treatment or diseases. Furthermore, protein candidates can be further scrutinised in the "functional and interaction proteomics" approach by finding interacting partners or ligands. Ultimately, these expression, structural and functional protein information can be used to design a workable model for a biological system and hence can be tested in systematic series of perturbation experiments.

3.2 Applications of Proteomics

Proteomics especially the expression proteomics has been applied in various organisms and samples including human cancer and plant crops. These are detailed in the next subsections.

3.2.1 Cancer Proteomics

One of the most studied topics in human is perhaps cancer. Cancer is a complex disease that reflects the genetic as well as protein changes within the cell. Although many effective therapies are present for early detection and diagnosis, cancer remains a major cause of death worldwide, accounting for 8.2 million deaths in 2012 [17]. The most common causes of cancer death are cancers of the lung, liver, stomach, colorectum, breast and oesophagus [18]. In the next two decades, the number of new cancer cases is expected to reach an overwhelming 23.6 million cases [18]. This suggests that more works need to be done to investigate the cause and possible treatment for cancer. Fundamental research particularly at the proteome level will undeniably shed some light into the protein changes that may signal or contribute to the cancer regulation.

Proteomics approaches have been increasingly used for differential analysis of various biological samples from cancer patients, including cell lysates, cell secretome, serum, plasma, tumour tissue and body fluids (Table 3.1). This could lead to a better understanding of the molecular basis of cancer pathogenesis, which can spark the discovery of novel cancer-specific biomarkers [19, 20]. The identification of new cancer biomarkers with predictive value is necessary to allow detection and treatment of cancer when it is still curable [20]. In the different types of cancer, the discovery of biomarkers is supposed to improve one or more of the following critical applications: early diagnosis, prognosis and monitoring of disease progression, its response to therapy as well as its recurrence [21]. Table 3.1 summarises several biomarker discoveries in the lung, pancreatic and gastric cancer, using expression proteomics approach.

3.2.2 Crop Proteomics

Changes in global climate behaviour have resulted in the increase of extreme temperature related phenomena including drought, flood, wind, water erosion and storms [32], which in turn influence soil condition [33]. These changes have already negatively affected the production of staple foods, such as maize, wheat, rice and soybean [32, 34]. Every year, more than 50% of yield loss of major crop plants was estimated worldwide due to abiotic stress such as drought, salinity and extreme temperatures [35]. Moreover, as agriculture land becomes less available, farmers are forced to make use of marginal, low-quality soils, which may contain low levels of nutrients [33].

Plant stress response represents an active process that targets at an establishment of novel homeostasis under altered environmental conditions [36]. Elucidation of the molecular mechanisms underlying the plant response to abiotic stress and the development of stress-tolerant plants have received much attention in recent years. Furthermore, understanding the mechanisms through which plant cells tolerate

Table 3.1 Recent studies on cancer biomarkers using expression proteomics approach

Types	Sample	Potential biomarker	References
Lung cancer	Sera	Protein gene product 9.5 (PGP 9.5)	Brichory et al. [22]
	Cell culture	A disintegrin and metalloprotease-17 or ADAM metallopeptidase domain 17 or tumour necrosis factor-α-converting enzyme (ADAM-17) Osteoprotegerin Pentraxin 3 Follistatin Tumour necrosis factor receptor superfamily 1A	Planque et al. [23]
	Serum	Haptoglobin (HP) Apolipoprotein 4	Okano et al. [24]
	Blood plasma	Apolipoprotein E	Rice et al. [25]
	Cell lines	Heat-shock protein 90-beta (Hsp90-beta) Vimentin (VIM)	Zhang et al. [26]
Pancreatic Cancer	Urine	Lymphatic vessel endothelial hyaluronan receptor (LYVE1) Regenerating islet-derived 1 beta (REG1B) Trefoil factor 1 (TFF1)	Radon et al. [27]
	Plasma cells	Anterior gradient homolog 2 (AGR2) Polymeric immunoglobulin receptor (PIGR) Olfactomedin-4 (OLFM4) Syncollin (SYCN) Collagen alpha-1 (VI) chain (COL6A1)	Makawita et al. [28]
	Serum	Cyclin I Rab GDP dissociation inhibitor ß (GDI2) Haptoglobin precursor Serotransferrin precursor	Sun et al. [29]
Gastric cancer	Tissue	Glucose-regulated protein 78 (GRP78) Glutathione S-transferase pi (GST pi) Apolipoprotein AI (Apo AI) Alpha-1 antitrypsin (A1AT) Gastrokine-1 (GKN-1)	Wu et al. [30]
	Tissue	NSP3 Transgelin (SM22-alpha) Prohibitin Heat-shock 27 kDa protein Protein disulphide isomerase A3 Apolipoprotein AI (ApoAI) Alpha-1 antitrypsin (A1AT)	Ryu et al. [31]

these stresses is essential for the improvement of crop tolerance by genetic engineering or genome editing. Proteomics approaches in particular "expression proteomics" have enabled characterisation of target regulatory proteins and biomarker identification to further comprehend the plant physiology and molecular defence under abiotic stresses.

Drought is one of the major abiotic stresses in crops such as wheat and rice. The proteome of two different wheat varieties with different tolerance to drought, Opata M85 (sensitive) and Nesser (tolerant), were evaluated using abscisic acid (ABA) treatment [37]. Abscisic acid is the key phytohormone produced in response to drought and is involved in coordinating various signalling and metabolic pathways during drought stress. Analysis of their root protein profiles showed that abscisic acid affected the expression level of 805 proteins, and several proteins showed variety-specific regulation by abscisic acid, suggesting their role in drought adaptation [37]. Similarly, physiological analysis of leaf and root protein expression from drought-tolerant wild wheat indicated that abscisic acid level was greatly increased in the drought-treated plants, but the increase was greater and more rapid in the leaves than in the roots [38]. Phosphoproteome analysis of seedling leaves from two bread wheat cultivars (Hanxuan 10 and Ningchun 47) subjected to drought stress also found several important regulators of abscisic acid signalling [39]. To unravel the mechanism behind the maize phosphoenolpyruvate carboxylase (PEPC) gene's capability in improving wheat resistance to drought stress, Qin et al. [40] examined proteome changes under drought conditions of two PEPC-containing transgenic wheat lines and the parental control line (Zhoumai19). The expression of several proteins which related to photosynthesis and cytoskeleton synthesis, and also S-adenosylmethionine synthetase, was induced in transgenic wheat under drought stress, thus demonstrating the efficacy of PEPC in crop improvement.

Proteomics studies have also been performed in rice for drought responses. A number of drought-tolerant and drought-sensitive cultivars of *Oryza sativa* L. ssp. *indica* and *O. sativa* L. *japonica* have been examined [41–44]. Several genes and proteins involved in drought response were identified and characterised. These studies have identified a set of proteins that are drought responsive, including 42 in leaf [44], 22 in rice root [43], 31 in peduncle tissue [42] and 53 in leaf [41]. The understanding of drought responses in rice is critical for designing breeding strategies to develop varieties which are more tolerant to water deficit.

Heat stress is also one of the major stresses affecting crop production, and proteomics has been applied to investigate the molecular response of wheat and rice to such stress. In wheat, proteins related to desiccation and oxidative stress [45, 46], photosynthesis, glycolysis, stress defence, heat shock and ATP production [47] were differentially expressed in the tolerant and sensitive cultivar under heat stress treatment. Meanwhile, heat stress induced the increase of small heat shock protein, β-expansins and lipid transfer proteins in the resistant rice cultivar [48]. This suggests that heat induced protein changes related to stress tolerance and biochemical modifications.

Another abiotic stress, high-salinity condition is also being investigated in various proteomics studies. For example, Xu et al. [49] identified 14 proteins involved in rice seed imbibition during salt stress, in which the majority of these proteins were involved in energy supply and storage protein. Meanwhile, several novel salt stress-responsive proteins, including protein synthesis inhibitor I, photosystem II stability/assembly factor HCF136, trigger factor-like protein and cycloartenol-C24-methyltransferase, were upregulated upon salt stress in rice shoot [50]. Jankangram

and colleagues [51] identified ten differential proteins, including gene products involved in photosynthesis, carbon assimilation and the oxidative stress response. They also found that although salinity-sensitive cultivar (Khao Dawk Mali 105) contains elevated transcript level of genes needed for salt tolerance, the post-transcriptional mechanisms controlling protein expression levels were not as efficient as in Pokkali (salinity-tolerant cultivar) [51]. This highlights the importance of studying plant molecular responses at the proteome level, especially during abiotic stresses.

3.3 Proteomics Workflow

Expression proteomics is most useful for cataloguing proteins and finding protein biomarkers in any given samples, especially in the context of systems biology. Hence this subchapter discusses the main experimental design and consideration in the expression proteomics as well as the strategies used to achieve whole proteome analysis.

3.3.1 Research Design and Consideration

Expression proteomics in general may consist of five main stages as illustrated in Fig. 3.4. In brief, after identification of the organism of interest, a suitable protein extraction protocol needs to be established before the protein amount is accurately calculated. Certain amounts of proteins are separated followed by identification using a mass spectrometry (MS) instrument. Raw MS data will then be processed using appropriate software to determine the protein identity.

Protein extraction is one of the most critical steps in a proteomics study. This is because high protein yield and clean samples will generate the best proteome coverage. Protocols can vary between organisms [52, 53] and therefore requires thorough literature search and a few rounds of preliminary experiments to determine the best extrac-

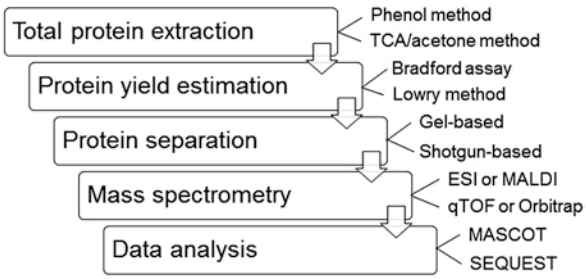

Fig. 3.4 General workflow of an expression proteomics study and examples related to each step (detailed in the text). TCA, trichloroacetic acid; ESI, electrospray ionisation; MALDI, matrix-assisted laser desorption/ionisation; qTOF, quadrupole time-of-flight

tion methods. The two most common methods are phenol and trichloroacetic acid (TCA)/acetone procedures [54, 55]. Determining protein levels in extracted samples is also crucial to evaluate differences between extraction methods and for consistent loading into gels and columns. Protein concentration is often estimated by interpolation of samples' absorbance on a protein standard curve, normally constructed using different concentrations of protein standards such as bovine serum albumin (BSA). A few methods are widely used, namely, Bradford assay and Lowry method [56, 57].

Total protein extracts then need to be first separated to simplify the protein mixture before protein identification can be done. There are two different strategies for this, namely, gel-based and shotgun-based approaches [53, 58, 59]. Gel-based platform employs a two-dimensional gel electrophoresis (2DGE) system which essentially separates the proteins based on their isoelectric point (pI) and molecular weight [60, 61]. Meanwhile, shotgun-based proteomics rely on the resolving power of chromatographic techniques that can separate biomolecules using ion exchange and reversed-phase columns, among others [4]. There are several differences between gel-based and shotgun-based techniques [4, 52]. Firstly, 2DGE is often regarded more laborious as a gel medium is needed to resolve protein spots, whereas chromatographic techniques can often be an on-line procedure without the need of any resolving gel medium beforehand. Secondly, gel-based proteomics require visual inspection and densitometer to quantify protein spot differences between samples [62]. Whereas in shotgun-based approach, quantitative measurement of peptides (corresponding to the proteins) can be achieved through isotope labelling (such as SILAC) or spectral counting for labelled-free approach [63, 64]. Thirdly, protein digestion is required after 2DGE separation for gel-based approach, yet protein mixtures are digested even before the chromatographic run in the shotgun-based workflow. Finally, unlike 2DGE which is not easily automated (although robotic arms for spot picking exist), the chromatography columns often can be coupled with downstream mass spectrometry analysis for peptide identification.

Once proteins are successfully separated and digested, mass spectrometry (MS) will be used to elucidate the molecular mass of these peptides. MS instruments are consisted of three sequentially ordered parts: [1] ion source [2], mass analyser and [3] detector [64–66]. This has been illustrated in Fig. 3.5. Before any peptide samples are able to be measured, they first need to be ionised using the ion source, which can often be either an electrospray ionisation (ESI) or matrix-assisted laser desorption/ionisation (MALDI). Liquid peptide samples separated using liquid chromatography are well suited for ESI and hence a method of choice for shotgun proteomics, whereas single/a few protein samples isolated from gels (SDS-PAGE or 2DGE) are commonly using MALDI for peptide identification. The next part of a MS instrument is the mass analyser which essentially separates the peptide ions before fragmenting them. This generates a profile of peptide ions differing in masses and charges which are denoted as *m/z* values. Most commonly used mass analysers in proteomics studies are quadrupole time-of-flight (qTOF) and Orbitrap, owing to their sensitivity, accuracy and speed [65]. The resulting peptide ions are then measured in the third part of a MS called detector which will supply the data to designated workstations. This generated raw data then need to be thoroughly analysed using appropriate platforms/software.

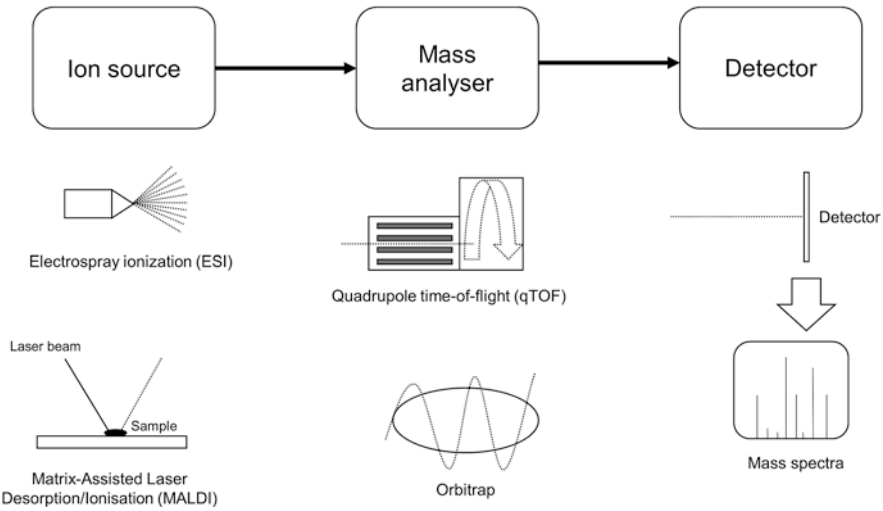

Fig. 3.5 Key components of mass spectrometry (MS) with representative figures of techniques/instruments used in each component. It consists of an ion source which can be either an electrospray ionisation (ESI) or matrix-assisted laser desorption/ionisation (MALDI) type to ionise sample peptides (dotted lines). Once ionised, peptides will be separated based on mass and charge in mass analysers which include quadrupole time-of-flight (qTOF) and Orbitrap. The detector then computes the peptide information (mass spectra) from the mass analysers to be used in subsequent data analysis

There are a number of software available for MS data analysis. Commonly used proteomics software are Mascot (www.matrixscience.com) and SEQUEST (fields.scripps.edu/sequest/) [63]. Other software that are available for proteomics data analysis are listed in Twyman [4] and Rose et al. [53]. These software are mainly differed in their algorithms for protein identification but essentially consist of six main steps as detailed in Twyman [4]. The first two steps involve MS spectral data collection and processing. Then protein sequences from specific databases of organism of interest are theoretically digested using enzymes similarly used for the protein samples earlier. The next steps are processing the theoretical spectra and matching them with the processed MS spectral data. Finally, statistical analysis is required to measure how good is the match between the MS and theoretical spectra for accurate identification of the protein ID. These six main steps are the basis of peptide mass fingerprinting (PMF). De novo protein sequence identification using software such as Peaks and PepNovo is also another alternative [67]. For publication, minimal information about a proteomics experiment (MIAPE, www.psidev.info/node/91) has been established and required in top proteomics journals such as *Proteomics* as well as *Molecular and Cellular Proteomics* [55, 68]. Hence, these requirements, such as depositing data into a public domain [69], need to be followed to ensure successful manuscript revision and publication.

3.3.2 *Different Strategies for Expression Proteomics*

Both gel-based and shotgun-based approaches (Fig. 3.6) have various improvised strategies to quantify proteins from different samples, often simultaneously. For example, 2D-DIGE (differential gel electrophoresis) method labels different samples with different florescent dyes before 2DGE is performed [62]. Protein abundance will be quantified according to its protein spot signals measured using an imager [62]. On the other hand, shotgun-based approach has also employed similar approach where proteins are labelled to be quantified. Different variations of this approach were reported such as isotope-coded affinity tagging (ICAT), isobaric tags for relative and absolute quantitation (iTRAQ) and SILAC [52]. These techniques used isotopes to label proteins/peptides, and differences between peptides from different samples will be quantitated using appropriate MS analysis. More recently, techniques such as multidimensional protein identification technology

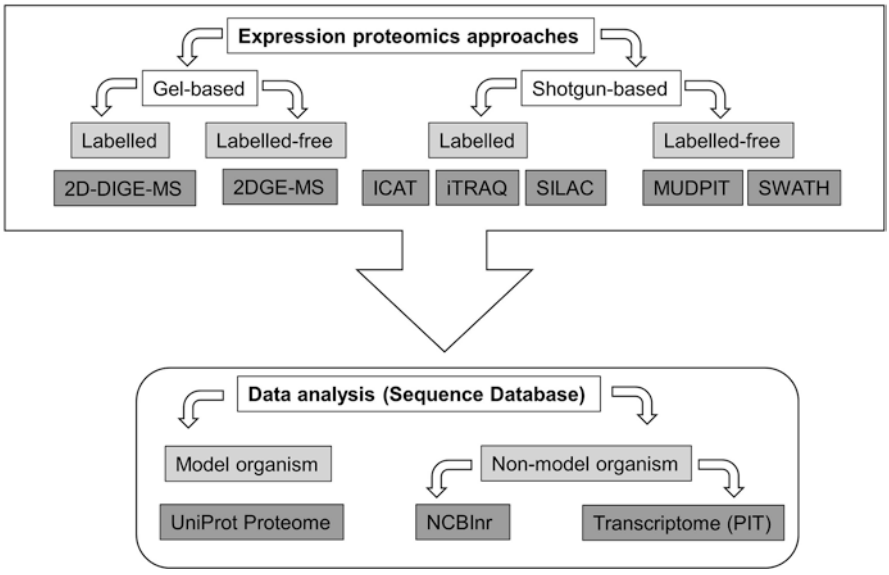

Fig. 3.6 Expression proteomics can be divided either into gel-based or shotgun-based approaches. For labelled and labelled-free gel-based approaches, two-dimensional-differential gel electrophoresis (2D-DIGE) method and 2D gel electrophoresis (2DGE) are, respectively, used which then coupled with mass spectrometry (MS) for protein identification. For labelled shotgun-based approach, isotope-coded affinity tagging (ICAT), isobaric tags for relative and absolute quantitation (iTRAQ) and stable isotope labelling by/with amino acids in cell culture (SILAC) are commonly used, whereas labelled-free utilises either multidimensional protein identification technology (MUDPIT) or sequential window acquisition of all theoretical mass spectra (SWATH-MS). For data analysis, model organism can use its available proteome sequence database from the UniProt website (https://www.uniprot.org/), whereas non-model organism may use either NCBI nonredundant database (NCBInr) or corresponding transcriptome database in a strategy called proteomics informed by transcriptomics (PIT)

(MUDPIT) coupled with Orbitrap [52, 70] and sequential window acquisition of all theoretical mass spectra (SWATH-MS) analysis using TripleTOF technology [71] have been developed and used successively to quantify large number of proteins without labelling. All in all, these various different strategies allow high-throughput and sensitive approaches for quantitative proteomics and have certainly propelled this research area.

However, proteomics approach has often been hindered by the lack of reference database (Fig. 3.6). For model organisms, their complete proteome sequence database can be easily obtained through UniProt database (https://www.uniprot.org/). The availability of a specific and complete protein sequence database is crucial for peptide mass spectra to be analysed and accurately predicted using PMF [4, 68]. This would help protein identification from spectral peptides and hence determine their possible function and biological relevance. PMF without a complete genome/protein database can be a daunting task for any bioinformatics tools as a number of non-significant hits can be generated even when using NCBI nonredundant database (NCBInr) [68]. Therefore, proteomics informed by transcriptomics (PIT) strategy has been introduced for the proteomics analysis of non-model organisms [72, 73]. Using the sequence data information obtained through transcriptomic analysis (detailed in Chap. 2), the protein profile of a given organism can be correctly deduced [72]. Given that a number of herbs and exotic plants for proteomics, which genome sequences are largely unavailable, PIT is the best strategy to be opt for in the future.

3.4 Proteomics and Enzymatic Studies of Kesum

3.4.1 Kesum: A Proteomics Case Study

Proteomics approach has been used in several plant species including model plant Arabidopsis [74] and fruit crops, such as tomato [75, 76] and capsicum [77]. However, proteomics studies in non-model organisms are often hindered by the lack of reference sequence database. While it is still possible to use general public sequence database particularly from NCBI, this often leads to lesser number of identified proteins due to low match hits. Nonetheless, proteomics studies in herbal plant species have been performed in a few species including *Gastrodia elata* orchid [78], *Zanthoxylum nitidum* [79] and *Pueraria radix* [80]. Other non-model species such as *Persicaria minor* Huds. or locally known as kesum, pitcher plant (*Nepenthes* sp.), mahogany (*Swietenia macrophylla* King) seeds and mangosteen (*Garcinia mangostana* L.) fruit are also of interest in this tropical region and currently being investigated using proteomics approach.

Kesum in particular has been used in many traditional cuisine across Southeast Asian countries including Malaysia because of its pungent smell [81]. Moreover, several studies have characterised that the plants contain certain medicinal com-

pounds that exhibit antibacterial and anticancer properties [81]. It is also being used as antidandruff as well as aromatherapy products [82, 83]. Some of the many compounds in kesum can be classified as terpenoids and aliphatic aldehyde compounds which contribute to the smell and taste of this herbal plant [84–86]. In order to investigate the biosynthetic pathways of these compounds, a proteomics analysis employing a shotgun proteomics has been performed in this species. Several proteins have been identified to be differentially expressed upon methyl jasmonate treatment [87] which may be responsible for the induction of different volatile compounds during stress conditions [85, 70]. This herbal proteomics study [87] employed PIT to increase the number of peptide match hits and ultimately assisting the protein identification.

3.4.2 Enzymatic Studies from Kesum

Enzymes of secondary metabolites biosynthetic pathway are attractive targets for development of potential antimicrobial, anticancer drugs and insect-resistant crop plants by deployment of transgenic plant. However, as the numbers of known genes are growing, the elucidation of their functions remains a major bottleneck and lag behind the sequencing capability [88]. Therefore, more studies need to be done to characterise enzymes and their substrates as well as products to understand their activities and mechanistic properties.

Functional characterisation of selected proteins from kesum has also been performed using enzymatic analysis. The research aimed at identifying substrates and products of oxidoreductase enzymes, which are involved in the biosynthetic pathway of monoterpenes and sesquiterpenes in kesum [89–92]. Several buffer compositions containing reducing agents, osmotic reagents, protease inhibitors and phenolic absorbent were employed to select the most suitable extraction buffers for the extraction of selected enzymes [89–92]. These enzymes including terpene alcohol dehydrogenases, terpene aldehyde dehydrogenases and terpene synthases were identified in the cell-free extract of kesum leaves.

References

1. Humphery-Smith I (2015) The 20th anniversary of proteomics and some of its origins. Proteomics 15:1773–1776
2. Banks RE, Dunn MJ, Hochstrasser DF, Sanchez J-C, Blackstock W, Pappin DJ, Selby PJ (2000) Proteomics: new perspectives, new biomedical opportunities. Lancet 356:1749–1756
3. Khoury GA, Baliban RC, Floudas CA (2011) Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep 1:1–5
4. Twyman RM (2013) Principles of proteomics. Garland Science, Abingdon
5. Edman P, Begg G (1967) A Protein Sequenator. Eur J Biochem 1:80–91

6. Sali A, Glaeser R, Earnest T, Baumeister W (2003) From words to literature in structural proteomics. Nature 422:216–225
7. Woolfson M (2018) The development of structural x-ray crystallography. Phys Scr 93:1–32
8. Bisswanger H (2014) Enzyme assays. Perspect Sci 1:41–55
9. LaBaer J, Ramachandran N (2005) Protein microarrays as tools for functional proteomics. Curr Opin Chem Biol 9:14–19
10. Reymond Sutandy FX, Qian J, Chen C-S, Zhu H (2013) Overview of protein microarrays. Curr Protoc Protein Sci 72:1–21
11. Lueong SS, Hoheisel JD, Alhamdani MSS (2014) Protein microarrays as tools for functional proteomics: achievements, promises and challenges. J Proteomics Bioinform 7:1–10
12. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LDN, Patil AH, Nanjappa V, Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK, Marimuthu A, Sathe GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD, Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey G, Mudgal K, Chatterjee A, Huang T-C, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS, Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TSK, Hruban RH, Kerr CL, Bader GD, Iacobuzio-Donahue CA, Gowda H, Pandey A (2014) A draft map of the human proteome. Nature 509:575–581
13. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, 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, Gerstmair A, Faerber F, Kuster B (2014) Mass-spectrometry-based draft of the human proteome. Nature 509:582–587
14. Kitano H (2002) Systems biology: a brief overview. Science 295:1662–1664
15. Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet 2:343–372
16. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, 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 1:376–386
17. www.cruk.org/cancerstats (2014)
18. www.cancer.gov (2014)
19. Savino R, Paduano S, Preianò M, Terracciano R (2012) The proteomics big challenge for biomarkers and new drug-targets discovery. Int J Mol Sci 13:13926–13948
20. Maurya P, Meleady P, Dowling P, Clynes M (2007) Proteomic approaches for serum biomarker discovery in cancer. Anticancer Res 27:1247–1255
21. Sallam RM (2015) Proteomics in cancer biomarkers discovery: challenges and applications. Dis Markers 2015:1–12
22. Brichory F, Beer D, LeNaour F, Giordano T, Hanash S (2001) Proteomics-based identification of protein gene product 9.5 as a tumor antigen that induces a humoral immune response in lung cancer. Cancer Res 61:7908–7912
23. Planque C, Kulasingam V, Smith CR, Reckamp K, Goodglick L, Diamandis EP (2009) Identification of five candidate lung cancer biomarkers by proteomics analysis of conditioned media of four lung cancer cell lines. Mol Cell Proteomics 8:2746–2758
24. Okano T, Seike M, Kuribayashi H, Soeno C, Ishii T, Kida K, Gemma A (2016) Identification of haptoglobin peptide as a novel serum biomarker for lung squamous cell carcinoma by serum proteome and peptidome profiling. Int J Oncol 48:945–952
25. Rice SJ, Liu X, Miller B, Joshi M, Zhu J, Caruso C, Gilbert C, Toth J, Reed M, Rassaei N (2015) Proteomic profiling of human plasma identifies apolipoprotein E as being associated with smoking and a marker for squamous metaplasia of the lung. Proteomics 15:3267–3277

26. Zhang W, Li Y, Yang S, Li W, Ming Z, Zhang Y, Hou Y, Niu Z, Rong B, Zhang X (2013) Differential mitochondrial proteome analysis of human lung adenocarcinoma and normal bronchial epithelium cell lines using quantitative mass spectrometry. Thorac Cancer 4:373–379
27. Radon TP, Massat NJ, Jones R, Alrawashdeh W, Dumartin L, Ennis D, Duffy SW, Kocher HM, Pereira SP, Guarner L (2015) Identification of a three-biomarker panel in urine for early detection of pancreatic adenocarcinoma. Clin Cancer Res 21:3512–3521
28. Makawita S, Smith C, Batruch I, Zheng Y, Rückert F, Grützmann R, Pilarsky C, Gallinger S, Diamandis EP (2011) Integrated proteomic profiling of cell line conditioned media and pancreatic juice for the identification of pancreatic cancer biomarkers. Mol Cell Proteomics 10:1–21
29. Sun Z-L, Zhu Y, Wang F-Q, Chen R, Peng T, Fan Z-N, Xu Z-K, Miao Y (2007) Serum proteomic-based analysis of pancreatic carcinoma for the identification of potential cancer biomarkers. Biochimica Biophys Acta, Proteins Proteomics 1774:764–771
30. Wu J-Y, Cheng C-C, Wang J-Y, Wu D-C, Hsieh J-S, Lee S-C, Wang W-M (2014) Discovery of tumor markers for gastric cancer by proteomics. PLOS ONE 9:e84158
31. Ryu J-W, Kim H-J, Lee Y-S, Myong N-H, Hwang C-H, Lee G-S, Yom H-C (2003) The proteomics approach to find biomarkers in gastric cancer. J Korean Med Sci 18:505–509
32. FAO (2015) Climate change and food security: risks and responses. Food and Agriculture Organization of the United Nations (FAO). www.fao.org
33. Barkla BJ (2016) Identification of abiotic stress protein biomarkers by proteomic screening of crop cultivar diversity. Proteomes 4:26
34. Kang Y, Khan S, Ma X (2009) Climate change impacts on crop yield, crop water productivity and food security – a review. Prog Nat Sci 19:1665–1674
35. Qin F, Shinozaki K, Yamaguchi-Shinozaki K (2011) Achievements and challenges in understanding plant abiotic stress responses and tolerance. Plant Cell Physiol 52:1569–1582
36. Kosová K, Vítámvás P, Urban MO, Klíma M, Roy A, Prášil IT (2015) Biological networks underlying abiotic stress tolerance in temperate crops—a proteomic perspective. Int J Mol Sci 16:20913–20942
37. Alvarez S, Roy Choudhury S, Pandey S (2014) Comparative quantitative proteomics analysis of the ABA response of roots of drought-sensitive and drought-tolerant wheat varieties identifies proteomic signatures of drought adaptability. J Proteome Res 13:1688–1701
38. Liu H, Sultan MARF, li Liu X, Zhang J, Yu F, Xian Zhao H (2015) Physiological and comparative proteomic analysis reveals different drought responses in roots and leaves of drought-tolerant wild wheat (*Triticum boeoticum*). PLOS ONE 10:e0121852
39. Zhang M, Lv D, Ge P, Bian Y, Chen G, Zhu G, Li X, Yan Y (2014) Phosphoproteome analysis reveals new drought response and defense mechanisms of seedling leaves in bread wheat (*Triticum aestivum* L.). J Proteome 109:290–308
40. Qin N, Xu W, Hu L, Li Y, Wang H, Qi X, Fang Y, Hua X (2016) Drought tolerance and proteomics studies of transgenic wheat containing the maize C4 phosphoenolpyruvate carboxylase (PEPC) gene. Protoplasma 253:1503–1512
41. Maksup S, Roytrakul S, Supaibulwatana K (2014) Physiological and comparative proteomic analyses of Thai jasmine rice and two check cultivars in response to drought stress. J Plant Interact 9:43–55
42. Muthurajan R, Shobbar Z-S, Jagadish S, Bruskiewich R, Ismail A, Leung H, Bennett J (2011) Physiological and proteomic responses of rice peduncles to drought stress. Mol Biotechnol 48:173–182
43. Rabello AR, Guimarães CM, Rangel PH, da Silva FR, Seixas D, de Souza E, Brasileiro AC, Spehar CR, Ferreira ME, Mehta Â (2008) Identification of drought-responsive genes in roots of upland rice (*Oryza sativa* L.). BMC Genomics 9:485
44. Salekdeh GH, Siopongco J, Wade LJ, Ghareyazie B, Bennett J (2002) Proteomic analysis of rice leaves during drought stress and recovery. Proteomics 2:1131–1145
45. Kumar RR, Singh GP, Goswami S, Pathak H, Rai RD (2014) Proteome analysis of wheat (*Triticum aestivum*) for the identification of differentially expressed heat-responsive proteins. Aust J Crop Sci 8:973

46. Laino P, Shelton D, Finnie C, De Leonardis AM, Mastrangelo AM, Svensson B, Lafiandra D, Masci S (2010) Comparative proteome analysis of metabolic proteins from seeds of durum wheat (cv. Svevo) subjected to heat stress. Proteomics 10:2359–2368
47. Wang X, Dinler BS, Vignjevic M, Jacobsen S, Wollenweber B (2015) Physiological and proteome studies of responses to heat stress during grain filling in contrasting wheat cultivars. Plant Sci 230:33–50
48. Mu Q, Zhang W, Zhang Y, Yan H, Liu K, Matsui T, Tian X, Yang P (2017) iTRAQ-based quantitative proteomics analysis on rice anther responding to high temperature. Int J Mol Sci 18:1811
49. Xu E, Chen M, He H, Zhan C, Cheng Y, Zhang H, Wang Z (2017) Proteomic analysis reveals proteins involved in seed imbibition under salt stress in rice. Front Plant Sci 7:2006
50. Li X-J, Yang M-F, Zhu Y, Liang Y, Shen S-H (2011) Proteomic analysis of salt stress responses in rice shoot. J Plant Biol 54:384
51. Jankangram W, Thammasirirak S, Jones MG, Hartwell J, Theerakulpisut P (2011) Proteomic and transcriptomic analysis reveals evidence for the basis of salt sensitivity in Thai jasmine rice (*Oryza sativa* L. cv. KDML 105). Afr J Biotechnol 10:16157–16166
52. Jorrín-Novo JV, Pascual J, Sánchez-Lucas R, Romero-Rodríguez MC, Rodríguez-Ortega MJ, Lenz C, Valledor L (2015) Fourteen years of plant proteomics reflected in Proteomics: moving from model species and 2DE-based approaches to orphan species and gel-free platforms. Proteomics 15:1089–1112
53. Rose JKC, Bashir S, Giovannoni JJ, Jahn MM, Saravanan RS (2004) Tackling the plant proteome: practical approaches, hurdles and experimental tools. Plant J 39:715–733
54. Carpentier SC, Witters E, Laukens K, Deckers P, Swennen R, Panis B (2005) Preparation of protein extracts from recalcitrant plant tissues: an evaluation of different methods for two-dimensional gel electrophoresis analysis. Proteomics 5:2497–2507
55. Jorrín-Novo JV, Maldonado AM, Echevarría-Zomeño S, Valledor L, Castillejo MA, Curto M, Valero J, Sghaier B, Donoso G, Redondo I (2009) Plant proteomics update (2007–2008): second-generation proteomic techniques, an appropriate experimental design, and data analysis to fulfill MIAPE standards, increase plant proteome coverage and expand biological knowledge. J Proteome 72:285–314
56. Redmile-Gordon MA, Armenise E, White RP, Hirsch PR, Goulding KWT (2013) A comparison of two colorimetric assays, based upon Lowry and Bradford techniques, to estimate total protein in soil extracts. Soil Biol Biochem 67:166–173
57. Okutucu B, Dınçer A, Habib Ö, Zıhnıoglu F (2007) Comparison of five methods for determination of total plasma protein concentration. J Biochem Biophys Methods 70:709–711
58. Martínez-Esteso MJ, Martínez-Márquez A, Sellés-Marchart S, Morante-Carriel JA, Bru-Martínez R (2015) The role of proteomics in progressing insights into plant secondary metabolism. Front Plant Sci 6:504
59. Finoulst I, Pinkse M, Van Dongen W, Verhaert P (2011) Sample preparation techniques for the untargeted LC-MS-based discovery of peptides in complex biological matrices. Biomed Res Int 2011:245291
60. Cañas B, Piñeiro C, Calvo E, López-Ferrer D, Gallardo JM (2007) Trends in sample preparation for classical and second generation proteomics. J Chromatogr A 1153:235–258
61. Neverova I, Van Eyk JE (2005) Role of chromatographic techniques in proteomic analysis. J Chromatogr B 815:51–63
62. Issaq HJ, Veenstra TD (2008) Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. BioTechniques 44:697
63. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics 4:1487–1502
64. Zhu W, Smith JW, Huang C-M (2009) Mass spectrometry-based label-free quantitative proteomics. Biomed Res Int 2010:840518

65. Domon B, Aebersold R (2010) Options and considerations when selecting a quantitative proteomics strategy. Nat Biotechnol 28:710–721
66. Lin D, Tabb DL, Yates JR (2003) Large-scale protein identification using mass spectrometry. Biochimica Biophys Acta, Proteins Proteomics 1646:1–10
67. Romero-Rodríguez MC, Pascual J, Valledor L, Jorrín-Novo J (2014) Improving the quality of protein identification in non-model species. Characterization of *Quercus ilex* seed and *Pinus radiata* needle proteomes by using SEQUEST and custom databases. J Proteome 105:85–91
68. Jorrín-Novo JV (2015) Scientific standards and MIAPEs in plant proteomics research and publications. Front Plant Sci 6:473
69. Taylor CF, Paton NW, Lilley KS, Binz P-A, Julian RK, Jones AR, Zhu W, Apweiler R, Aebersold R, Deutsch EW (2007) The minimum information about a proteomics experiment (MIAPE). Nat Biotechnol 25:887–893
70. Webb KJ, Xu T, Park SK, Yates JR (2013) Modified MuDPIT separation identified 4488 proteins in a system-wide analysis of quiescence in yeast. J Proteome Res 12:2177–2184
71. Rosenberger G, Koh CC, Guo T, Röst HL, Kouvonen P, Collins BC, Heusel M, Liu Y, Caron E, Vichalkovski A, Faini M, Schubert OT, Faridi P, Ebhardt HA, Matondo M, Lam H, Bader SL, Campbell DS, Deutsch EW, Moritz RL, Tate S, Aebersold R (2014) A repository of assays to quantify 10,000 human proteins by SWATH-MS. Sci Data 1:1–15
72. Evans VC, Barker G, Heesom KJ, Fan J, Bessant C, Matthews DA (2012) *De novo* derivation of proteomes from transcriptomes for transcript and protein identification. Nat Methods 9:1207–1211
73. Mudenda L, Pierlé SA, Turse JE, Scoles GA, Purvine SO, Nicora CD, Clauss TRW, Ueti MW, Brown WC, Brayton KA (2014) Proteomics informed by transcriptomics identifies novel secreted proteins in *Dermacentor andersoni* saliva. Int J Parasitol 44:1029–1037
74. Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, Zimmermann P, Grossniklaus U, Gruissem W, Baginsky S (2008) Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. Science 320:938–941
75. Faurobert M, Mihr C, Bertin N, Pawlowski T, Negroni L, Sommerer N, Causse M (2007) Major proteome variations associated with cherry tomato pericarp development and ripening. Plant Physiol 143:1327–1346
76. Rocco M, D'Ambrosio C, Arena S, Faurobert M, Scaloni A, Marra M (2006) Proteomic analysis of tomato fruits from two ecotypes during ripening. Proteomics 6:3781–3791
77. Aizat WM, Able JA, Stangoulis JCR, Able AJ (2013) Proteomic analysis during capsicum ripening reveals differential expression of ACC oxidase isoform 4 and other candidates. Funct Plant Biol 40:1115–1128
78. Zeng X, Li Y, Ling H, Chen J, Guo S (2018) Revealing proteins associated with symbiotic germination of *Gastrodia elata* by proteomic analysis. Bot Stud 59:8
79. Lu Q, Zhang ZS, Zhan RT, He R (2018) Proteomic analysis of *Zanthoxylum nitidum* seeds dormancy release: influence of stratification and gibberellin. Ind Crop Prod 122:7–15
80. Kim Y, Chung WS, Jang HJ (2018) Proteins isolated of *Pueraria radix* possible to cause allergenic react with immunoglobulin E in human sera. Mol Cell Toxicol 14:233–239
81. Vikram P, Chiruvella KK, Ripain IHA, Arifullah M (2014) A recent review on phytochemical constituents and medicinal properties of kesum (*Polygonum minus* Huds.). Asian Pac J Trop Biomed 4:430–435
82. Bunawan H, Talip N, Noor NM (2011) Foliar anatomy and micromorphology of *Polygonum minus* Huds. And their taxonomic implications. Vascular 5:5–10
83. Azlim Almey A, Ahmed Jalal Khan C, Syed Zahir I, Mustapha Suleiman K, Aisyah M, Kamarul Rahim K (2010) Total phenolic content and primary antioxidant activity of methanolic and ethanolic extracts of aromatic plants' leaves. Int Food Res J 17:1077–1084
84. Baharum SN, Bunawan H, Ghani M a A, Mustapha WAW, Noor NM (2010) Analysis of the chemical composition of the essential oil of *Polygonum minus* Huds. Using two-dimensional gas chromatography-time-of-flight mass spectrometry (GC-TOF MS). Molecules 15:7006–7015

85. Khairudin K, Sukiran NA, Goh H-H, Baharum SN, Noor NM (2014) Direct discrimination of different plant populations and study on temperature effects by Fourier transform infrared spectroscopy. Metabolomics 10:203–211
86. Goh HH, Khairudin K, Sukiran NA, Baharum SN, Normah M (2015) Metabolite profiling reveals temperature effects on the VOCs and flavonoids of different plant populations. Plant Biol 18:130–139
87. Aizat WM, Ibrahim S, Rahnamaie-Tajadod R, Loke K-K, Goh H-H, Noor NM (2018) Extensive mass spectrometry proteomics data of *Persicaria minor* herb upon methyl jasmonate treatment. Data Brief 16:1091–1094
88. Amin SR, Erdin S, Ward RM, Lua RC, Lichtarge O (2013) Prediction and experimental validation of enzyme substrate specificity in protein structures. Proc Natl Acad Sci 110:E4195–E4202
89. Hassan M, Maarof ND, Ali ZM, Noor NM, Othman R, Mori N (2012) Monoterpene alcohol metabolism: identification, purification, and characterization of two geraniol dehydrogenase isoenzymes from *Polygonum minus* leaves. Biosci Biotechnol Biochem 76:1463–1470
90. Ahmad-Sohdi NAS, Seman-Kamarulzaman A-F, Mohamed-Hussein Z-A, Hassan M (2015) Purification and characterization of a novel NAD (P)+-farnesol dehydrogenase from *Polygonum minus* leaves. PLOS ONE 10:e0143310
91. Seman-Kamarulzaman A-F, Mohamed-Hussein Z-A, Ng CL, Hassan M (2016) Novel NAD+-farnesal dehydrogenase from *Polygonum minus* leaves. Purification and characterization of enzyme in juvenile hormone III biosynthetic pathway in plant. PLOS ONE 11:e0161707
92. Nik-Abdul-Ghani N-R, Mohamed-Hussein Z-A, Hassan M (2017) Citral dehydrogenase involved in geraniol oxidation pathway: purification, characterization and kinetic studies from *Persicaria minor* (*Polygonum minus* Huds.). J Plant Biochem Biotechnol 27:1–12