



Application of Proteomics Technologies in Oil Palm Research

Benjamin Yii Chung Lau¹ · Abrizah Othman¹ · Umi Salamah Ramli¹

Published online: 26 October 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Proteomics technologies were first applied in the oil palm research back in 2008. Since proteins are the gene products that are directly correspond to phenotypic traits, proteomic tools hold a strong advantage above other molecular tools to comprehend the biological and molecular mechanisms in the oil palm system. These emerging technologies have been used as non-overlapping tools to link genome-wide transcriptomics and metabolomics-based studies to enhance the oil palm yield and quality through sustainable plant breeding. Many efforts have also been made using the proteomics technologies to address the oil palm's *Ganoderma* disease; the cause and management. At present, the high-throughput screening technologies are being applied to identify potential biomarkers involved in metabolism and cellular development through determination of protein expression changes that correlate with oil production and disease. This review highlights key elements in proteomics pipeline, challenges and some examples of their implementations in plant studies in the context of oil palm in particular. We foresee that the proteomics technologies will play more significant role to address diverse issues related to the oil palm in the effort to improve the oil crop.

Keywords Fatty acids · Oil palm · Quantitative proteomics · *Ganoderma* · Crop improvement

1 Proteomics in Oil Palm Research

Rising number in the world population from 6.6 to 7.6 billion since the past decade has escalated the demand for food. In the context of palm oil, the market has also surged since it provides essential ingredients for food and non-food applications. To address this critical issue, the unequivocal solution is by increasing the yield of palm oil per unit area of land (from the current average production of 3 tons of oil per hectare per year to projected production of 18.5 tons of oil per hectare per year) [1]. To achieve this, the palm oil industry need to adopt policies for oil palm to be grown sustainably to obviate negative environmental impacts. Therefore, it is inevitable that we must address the current issue by improving the yield in existing plantations to increase

global production of palm oil [2]. To meet these challenges, 'omics' technologies have been practiced in the oil palm research to understand the yield-limiting and yield-reducing factors in the effort to enhance the yield and quality of palm oil [1]. However, to this respect, tremendous advances have been made over past years, particularly in oil palm genomics [3–10]. Transcriptomics information generated are simply not sufficient and biologically comprehensive enough to reveal the actual state of the plant system biology at any specific stage and condition. Therefore, proteomics is becoming an increasingly important tool for various applications in the oil palm research to simultaneously explain the oil palm biological processes of interest, such as yield and oil quality, the oil palm growth and development and its natural responses towards environmental stresses like diseases (Fig. 1).

The first step to integrate the proteomics approach into the oil palm research is to generate reference proteome maps for different oil palm species and fruit developmental stages. For that purpose, Lau, Hassan, Daim and co-workers had developed protein extraction protocols tailored for the fruit mesocarp, leaf and root tissues [11–13]. Together with the transcript sequences for *Elaeis guineensis* and *Elaeis oleifera* [7], the proteomics technologies are used to improve the current understanding of the palm oil biosynthesis machinery in

✉ Benjamin Yii Chung Lau
benjamin@mpob.gov.my

Abrizah Othman
abi@mpob.gov.my

Umi Salamah Ramli
umi@mpob.gov.my

¹ Malaysian Palm Oil Board, No. 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia

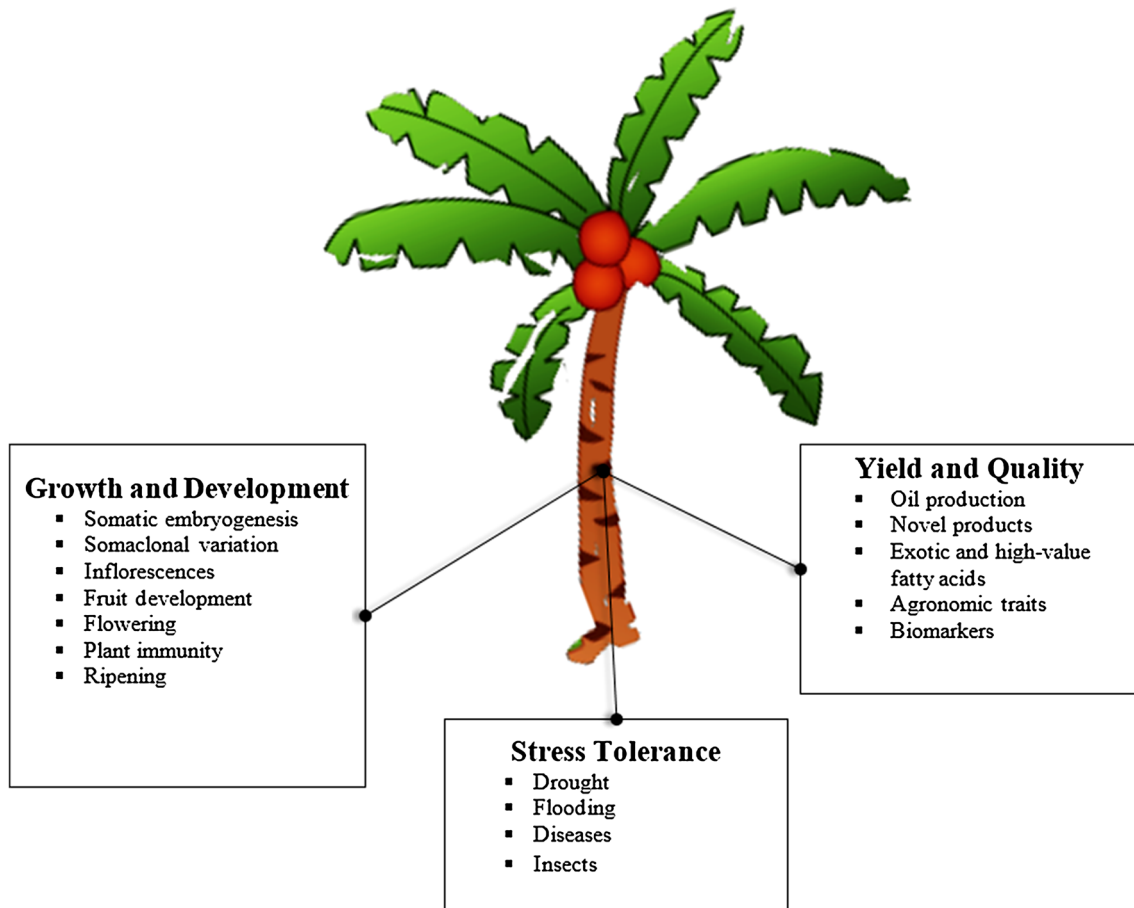


Fig. 1 Current and potential applications using the proteomics technologies

the search for potential protein biomarkers for yield and oil quality. Difference gel electrophoresis (DIGE) analysis was used in several studies to determine the control mechanisms and key proteins that distinguish different agronomic traits related to lipid biosynthesis. For instance, Ooi and co-workers detected 41 unique and differentially expressed proteins in the fruit mesocarp of high- and low-yielding oil palms that were not related to lipid biosynthesis [14]. The findings showed that regulation of lipid biosynthesis involved a myriad of metabolic pathways. The findings were strongly supported by Lau et al. [15, 16] using shotgun quantitative proteomics technologies. They also discovered several differentially accumulated proteins from other metabolic pathways such as glycolysis that would have contributed to the regulation of fatty acid biosynthesis. In a post-translational modification (PTM) study done by Lau and co-workers, selected reaction monitoring (SRM) using targeted peptide fragments corresponded to fatty acid proteins, revealed that phosphorylation could rationalize the differences in the oleic acid content of the studied oil palm species [17].

Protein changes during somatic embryogenesis, which involve cell differentiation, has been studied using gel-based

proteomics to understand molecular events of plant embryo development in vitro [18]. These differentially expressed proteins were identified in the early stage of embryogenesis and they were also stage-specific. Gel-based quantitative proteomics was applied in another study to understand the biological mechanism for the low level of embryogenesis [19]. They discovered three proteins, namely triosephosphate isomerase, L-ascorbate peroxidase and superoxide dismutase as potential protein biomarkers at both protein and transcript levels, respectively.

Proteomics technologies have been actively used to investigate the interaction between oil palm and the pathogenic fungi, *Ganoderma boninense*. Using gel-based proteomics approach, several studies were carried out to determine differentially expressed proteins upon infection of the oil palm root system with *G. boninense*. An optimized protein extraction using phenol/ammonium acetate in methanol was first developed by Al-Obaidi and co-workers to analyze the protein profile of *Ganoderma* species [20]. Al-Obaidi and co-workers had also identified 21 proteins from healthy and *G. boninense*-infected roots that showed differences in their expression profiles [21]. Protein profiling of the infected

root proteins at different time points had indicated that 12 proteins were differentially expressed, 7 days after infection with *G. boninense* [22]. Further comparison between pathogenic and non-pathogenic *Ganoderma* species revealed 24 differentially expressed proteins that corresponded to the *Ganoderma* species inoculations [23]. These proteins could explain the disease susceptibility of the oil palm. In the effort to develop early detection method for *G. boninense*, proteins in the leaf tissues of oil palm were profiled using a combination of gel electrophoresis and shotgun proteomics [24]. Majority of the 51 proteins identified were involved in photosynthesis, carbohydrate metabolism, immunity and defense. Table 1 summarized all the proteins that had been indicated as differentially expressed throughout the various stages of palm oil production and during the *Ganoderma* infection.

2 Challenges in Plant-Based Proteomics

A complete proteomics pipeline is made up from several stages, starting from protein extraction, digestion, separation and quantitation of peptides, and protein identification with mass spectrometry. Proteome of a cell or tissue at any given time is highly dynamic and complex. Combination of different approaches have been employed to study subgroup of proteins and entire proteomes largely because of the proteome complexity and wide dynamic range. Protein and peptide separation techniques have been essentially being scrutinized using two complementary approaches; gel-based and non-gel based (or gel-free or shotgun). These approaches vary in terms of the peptide generation, separation and detection. Ultimately, each of these approaches only covers specific protein subgroups but not the whole proteome. The gel-based approach is the cornerstone of proteomic analysis. Gel electrophoresis is a powerful technique to separate complex protein mixtures to yield qualitative and quantitative high resolution snapshots of intact proteins (two-dimensional) and polypeptides (one-dimensional), resulting in a quick overview of protein isoform varieties and detection of any post-translation modification.

The limitations of the gel-based approach, however, include the inability of the gel technique to resolve hydrophobic proteins or proteins with extreme sizes and isoelectric points. Thus, mass spectrometric analysis of these proteins is often not optimal unless further proteomics tools are employed. Gel-free or shotgun proteomics has been developed continuously mainly because of the need to reduce technical variations for high-throughput workflow that is not achievable using gel electrophoresis. In the gel-free or shotgun approach, liquid chromatography is coupled to an ionization source, which is typically the *nano*electrospray for peptides. Peptide ion separation will occur using only

reversed-phase column or with combination of different columns for high resolution separation. Fragmentation of peptide ions occur normally through collision with gas molecules and the enormous amount of data from the acquired tandem mass spectra are then used for protein database searches using protein search engines such as Mascot for the protein identities.

One of the core challenges in plant proteomics is the low protein concentration specially in circumstances where the amount of tissues are limited, as cell wall and vacuole make up most of the cell mass. Only 1–2% of the total cell volume make up the cytosol, which is the center of majority cellular processes [25]. Specialized procedures are essential to induce plant cell wall disruption to release the proteins. In addition, plant extracts also contain numerous non-proteinaceous compounds such as polyphenols, pigments, polysaccharides, nucleic acids and lipids. Several major crops had been reported to contain high amounts of compounds that interfere with downstream proteomics analysis. For instance, oxidative enzymes (polyphenol oxidase), phenolic compounds, latex and carbohydrates are abundant in banana, *Musa* spp. and stalk tissues [26–28]. These interfering compounds are co-purified with the precipitated proteins, rendering them difficult to solubilize. Solubilization of proteins is decisive in order to resolve them for further downstream analysis using techniques such as Western blot and mass spectrometry.

The existence of a high dynamic range of protein abundances in plant tissues confers additional complications to the protein analyses. For example, 40% of the total protein content of green tissues consists of ribulose-1,5-bisphosphate carboxylase oxygenase or RuBisCo [29] while storage proteins are the most abundant proteins in seeds [25]. The presence of those highly abundant proteins complicates the detection of low abundance proteins by means of protein electrophoresis and mass spectrometry. Normally, these low abundance proteins such as the regulatory proteins are the proteins that we are interested in [30]. Various fractionation techniques have been developed to deal with this wide dynamic range, which can be up to 12 orders of magnitude [31]. They are generally divided into electrophoresis- or chromatography-based fractionations to separate a subset of proteins [32]. For instance, isoelectric focusing that exploits the charge differences of proteins has been utilized to fractionate proteins to capture the less abundant proteins [33, 34]. Other approaches are based on the principle of affinity chromatography such as ATP and metal affinity [35], hydroxyapatite affinity- [36] and immobilized metal affinity-chromatography [37]. The latter technique is extremely useful in the enrichment of phosphorylated proteins in phosphoproteomics study.

Palm oil, which derived from the fruit mesocarp, can comprise up to 90% of the dry weight of oil in the fruit

Table 1 Differentially expressed proteins that have possible involvement in palm oil production and *Ganoderma* disease defence

	Protein	Possible role	References
Oil production	Sucrose synthase	Increase glucose supply for pyruvate and acetyl-CoA precursors	Loei et al. [16], Lau et al. [15]
	α -Amylase isozyme 3D		Loei et al. [16]
	Glyceraldehyde-3-phosphate dehydrogenase	Convert glucose to pyruvate and acetyl-CoA	
	Hexokinase-2		
	Pyruvate kinase isozyme A and G		
	Transaldolase	Cofactor for reductive fatty acid biosynthesis	
	Pyruvate dehydrogenase E1 α -subunit	Convert pyruvate to acetyl-CoA	
	Pyruvate dehydrogenase E1 β -subunit		
	E2 dihydrolipoyllysine-residue acetyltransferase		
	E3, dihydrolipoyl dehydrogenase		
	Fructose-bisphosphate aldolase	Involve in assimilation of carbon	Lau et al. [15]
	Acyl carrier protein (ACP)	Involve directly in various stages of fatty acid biosynthesis	Loei et al. [16]
	3R-hydroxymyristoyl-ACP dehydratase		
	Enoyl-ACP reductase		Loei et al. [16], Lau et al. [15]
	3-Oxoacyl-ACP reductase		
	Acetyl-coenzyme A carboxylase carboxyl transferase subunit β		
	Stearoyl-ACP desaturase		
	Acetyl-coenzyme A carboxylase biotin carboxylase subunit		Lau et al. [15]
	β -Ketoacyl-ACP reductase		
	β -Hydroxyacyl-ACP-dehydrogenase		
	β -Ketoacyl-ACP-synthase		
	Lipase	Hydrolyze fatty acids	
	40S and 60S ribosomal proteins of varying isoforms	Diversion of ATP and NADPH toward lipid biosynthesis	Loei et al. [16]
	Elongation factors		Loei et al. [16], Lau et al. [15]
	Component proteins from photosystem II (PSBA, PSBD, PSBO, PSBQ, PSBQ2)		
	Chlorophyll <i>a-b</i> binding proteins of light-harvesting complex (CB4, CB4A, CB13, CB22)		
	Malate dehydrogenase	Channel ATP to anabolic fatty acid biosynthesis	
	Cytochrome c oxidase subunit 6B		
	Cytochrome <i>b5</i>		
	Cytochrome <i>b-c1</i> complex subunit Rieske		
ATP synthase subunit β			

Table 1 (continued)

	Protein	Possible role	References	
Pathogen defence	Ribulose bisphosphate carboxylase/oxygenase	Reduction in photosynthesis to maintain the growth of cell during infection	Daim et al. [13, 24], Al-Obaidi et al. [20]	
	ATP synthase β subunit	Restrict growth of oil palm cell during infection		
	Oxygen-evolving enhancer protein 1	Reduction in chlorophyll biosynthesis resulted in yellowish marks on infected leaves	Al-Obaidi et al. [20]	
	Malate dehydrogenase	Indication of early infection during plant-phytopathogenic interactions		
	Stromal 70 kDa heat shock-related protein	Indication of stress such as infection		
	Cyclophilin	Defence response		
	Protochlorophyllide reductase B	Unknown		
	Ras-related protein RABB1a	Unknown		
	Putative membrane protein ycf1	Unknown		
	Ferredoxin NADP reductase	Plant defence reaction		Daim et al. [13, 24]
	Maturase k	Reduction in transcription level indicates decrease in metabolites and energy level during infection		
	Hydroxyproline-rich glycoprotein-like	Plant defence reaction		Syahanim et al. [22]
	Putative WAK receptor-like protein kinase	Plant defence reaction		
	Mannose-binding lectin	Plant defence reaction		
	E3 ubiquitin-protein ligase BRE1-like 1	Unknown		
	Putative pumilio homolog 10			
	Cytochrome C oxidase subunit II PS17	Generate energy for early defence mechanism		
	2-Cys peroxiredoxin	Recognition and activation of plant defence through rapid production of reactive oxygen species		
	Superoxide dismutase			
	Ascorbate peroxidase			
	β -1,3-Glucanase	Response to infection		
	Glutathione-S-transferase	Detoxifying contaminants through conjugation		
	Nucleoside diphosphate kinase	Response to infection and involve in early infection signalling event		
Thioredoxin H2	Antioxidant activity against free radicals			
Early flowering protein 1	Unknown			
Ferritin				

mesocarps [14], apart from the plant-based interfering compounds mentioned earlier. Therefore, preparation of proteins from oil palm requires labor intensive workflow in order to be compatible with downstream proteomics analysis. Comprehensive protocols in the preparation of oil palm proteins from fruit mesocarps, young and mature leaves; and roots for gel-based and gel-free proteomics analysis have been developed in recent years. Lau and co-workers had described an approach employing different solvent systems to remove the excessive oils from the oil palm fruit mesocarps prior to protein extraction with phenol [11]. Phenol extraction was also employed in works done by Daim, Al-Obaidi, Silva and

co-workers to investigate proteins from the oil palm leaf [13], *Ganoderma*-infected root [21], callus, embryos and explants [18] using gel electrophoresis-based mass spectrometry analysis. In addition, Daim and co-workers first precipitated the extracted leaf proteins with trichloroacetic acid/acetone before the phenol extraction to improve the number of gel spots and quality [24]. The high quality of proteins extracted with phenol in both works may be attributed to the fact that phenol precipitates only proteins (in phenol phase) and contaminants such as polysaccharides, polyphenol and carbohydrates are removed in the organic phase. In reducing the leaf protein complexity and quality

for two-dimensional gel electrophoresis, Tan and co-workers had applied both trichloroacetic acid/acetone precipitation and polyethylene glycol fractionation [38]. Trichloroacetic acid followed by acetone precipitation was also successfully applied in the extraction of proteins from oil palm root for gel-free mass spectrometric analysis [12].

3 Functional Proteomics Analysis

Early developments in quantitative proteomics were propelled by studies on yeast and mammalian cell lines [39]. Quantitative changes are needed to elucidate changes in protein expression. Staining of proteins on gels with specific stains such as Coomassie is routine to determine their intensities but often than not, this approach is tedious and error-prone. The intensities of the liquid chromatography peak detected using ultraviolet–visible spectrophotometric detector are usually not proportionate to the amount of proteins in a given sample. The reason is that different types of protein absorb different ultraviolet wavelengths which give the chromatograms. For examples, aromatic amino acids (tryptophan, tyrosine and cysteine) absorb ultraviolet wavelength at 280 nm while peptide backbone absorbs ultraviolet wavelength at 215–235 nm. Thus, stable isotope approaches were introduced into mass spectrometry-based proteomics to allow a more accurate and reliable determination of relative variations in peptide abundances. There are several strategies used today in quantitative proteomics and all of these methods have their advantages and disadvantages. Commonly, investigations using quantitative proteome analysis approach rely on non-mass spectrometry-based quantitation techniques [25] such as DIGE and mass spectrometry-based quantitation techniques [40].

Determination of protein ratios using gel-based methods has the potential to be erroneous because of gel-to-gel inconsistencies of the separated protein profiles. DIGE addresses these difficulties and that explain the reason for this technique to be most commonly employed in non-mass spectrometry-based proteomics quantitation, as well as quantitative analysis in plant proteomes. In DIGE, up to two different protein samples and a reference standard (containing equal amounts of both protein samples) are labelled with fluorescent dyes such as CyDyes (Cy3, Cy5, Cy2). The two protein samples are then pooled prior to separation with gel electrophoresis [41, 42]. Protein ratios between the two different samples are calculated by measuring the fluorescence for each protein spot and thus revealing the quantitative data for protein isoforms or differentially regulated proteins [43]. DIGE has been commonly used in plant proteomics studies. For example, in investigations of elicitation effects in plant symbiotes and plant pathogen interactions [44–47] as well as studies on environmental stresses [48–50]. Gomez and

co-workers [51] also demonstrated that DIGE coupled to MALDI-TOF analysis could be used to identify differentially expressed proteins in organisms lacking assembled genomes. Ooi and co-workers had also applied DIGE technique to determine 41 unique differentially accumulated proteins in the oil palm fruit mesocarps at critical oil production stages [14]. Pro-Q Diamond, a fluorescent dye that binds to the phosphate moiety of phosphorylated proteins has also been successfully employed to specifically label and quantify phosphorylated protein isoforms in plant [52–57].

The mass spectrometry-based quantitative methods include both label free quantitation [58, 59] and chemical isotope labelling [60, 61]. Mass spectrometry signals from different liquid chromatography runs are known to be inconsistent due to technical variations for instance, and therefore generate significant error in quantitative proteomics studies. Despite that fact, label-free methods involving liquid chromatography is becoming increasingly prevalent as it circumvents the need for costly protein labelling and is generally suitable for all types of organisms as well as most workflows [43, 62]. Label-free quantitation compares the chromatographic peak areas of extracted ions. Extracted ion chromatograms exploit the additional separation dimensions for higher confidence in the quantitative signals instead of simply comparing the mass spectrometry signals between different analytical liquid chromatography runs. In principal, peptide areas are aligned according to their mass to charge ratio (m/z) and elution time tags in several liquid chromatography runs. The chromatographic peaks are then integrated with peak integration software such as Xalign [63] and Msalign [64]. In order to be able to do that, the liquid chromatography runs must be reproducible, which sometimes can be a challenging task. Reiland and co-workers had used this approach to determine the dynamic regulation of protein phosphorylation in *Arabidopsis* [65].

Spectral counting is an alternative approach that is practical, label-free and measures protein abundance in a semi-quantitative manner [66]. Conversely, this method does not integrate chromatographic peaks nor align the retention time of peptides [67] although it agrees with Extracted Ion Chromatogram peak area measurements [68]. Instead, statistical tools such as G-test and *t*-test are used to count the total number of tandem mass spectra identified for all the peptides from a particular protein to generate the quantitation data [43]. While this method is reproducible, it requires many biological and technical replicates for each sample analyzed. This can be difficult when several experimental conditions and/or time points are analyzed. This approach had been successfully employed to quantify proteins in several studies in plant systems [69–72].

Usage of differential labelling techniques could circumvent these limitations in label-free quantitation engaging liquid chromatography. These approaches rely on the

assumption that both labelled and unlabeled peptides exhibit the same chromatographic and ionization properties but are distinguishable by a mass-shift signature [67]. Specific isotope labelled amino acids (^{13}C or ^{15}N) [73] in the metabolic protein labelling technique known as stable isotope labelling (SIL) with amino acids in cell culture (SILAC), and chemical labels such as in isobaric tags for relative and absolute quantitation (iTRAQ) have been used to quantitate changes in plant proteomes [74, 75]. Labelling methods used in relative quantitation proteomics studies are classed into two categories depending on whether the labels are tagged directly to the peptides or not directly.

Isotope-coded affinity tag (ICAT) was one of the first differential isotope labelling containing a specific chemical reactive group which bound specifically to cysteinyl residues, an isotope mass tag with light or heavy isotopes and a biotin tag for affinity purification [76]. Peptide pairs with 8 kDa mass-shifts are detected during mass spectrometry scans and their ion intensities are relatively compared for quantitation. As tagged cysteine-containing peptides are purified by affinity chromatography, the sample complexity is reduced. However, the obvious disadvantage is that only cysteine-containing peptides are captured by the affinity column. Thus, this impaired the identification and quantitation of proteins with more than one significant peptide as about one in seven proteins do not contain cysteine [67]. A study by Majeran and co-workers had revealed that non-MS-based (2-DE), ICAT and label-free quantitative techniques are complementary [77]. ICAT had been utilized to determine the localization of *Arabidopsis thaliana* organelle proteins [78]. In addition, since ICAT labels specifically to thiol groups, this method has been widely used to study the redox-status of proteins in plants [79, 80].

iTRAQ was developed at first for peptide level labelling [67]. The different between iTRAQ and ICAT is that in ICAT, tagged proteins from different samples are pooled before trypsinization to eliminate vial-to-vial variations. In iTRAQ, the chemical tags label the peptides instead. The iTRAQ isobaric tags have slight differences in their molecular structures and thus generate various fragment ions (also known as reporter ions) in tandem mass spectrometry scans. The overall molecule mass is kept constant at 145 Da (iTRAQ-4plex) and 304 Da (iTRAQ-8plex) by the presence of a mass balance group (carbonyl).

The iTRAQ reagents label the peptide N-terminals and amino groups of lysine side chains. The advantage is that iTRAQ approach allows comparison of four (iTRAQ-4plex) to eight samples (iTRAQ-8plex) in a single experiment. Relative quantitation is ascertained only after peptide fragmentations in MS/MS scans by measuring the intensity of the reporter ions in the mass region of m/z 114–118 and m/z 114–121, for 4plex and 8plex, respectively [43, 67]. iTRAQ method is able to give accurate quantitation spanning two

orders of magnitude for low-complexity samples. However, peptide co-fragmentation happens when two or more closely spaced peptides in MS/MS are selected instead of the single peptide [81, 82]. With a high accuracy mass spectrometer, peptide co-fragmentation effect could be reduced. Tandem mass tags (TMT) are another widely used isobaric tags to label peptides for relative protein quantitation proteomics. As with iTRAQ, the tags share identical chemical structure but have stable isotopes, ^{13}C and ^{15}N incorporated in different combinations in the mass reporter region. The chemical structure of TMTs enable the introduction of five heavy isotopes in the reporter and balancer groups to generate six isobaric tags. Fragmentation of each of the six tags (of a TMT 6-plex, for instant) gives reporter ions at m/z 126, 127, 128, 129, 130 and 131. TMTs react with free amino-terminus peptides and epsilon-amino functions of lysine residues [83].

iTRAQ and TMT reagents have been successfully employed in several quantitative plant proteomics studies. Plant responses towards pathogens had been investigated using this approach [84–87] as well as the signaling role played by trimeric G proteins in plants [88, 89]. Other studies utilized the iTRAQ to investigate the proteomes of grape berries [90] and oil palm mesocarp at different stages of ripening [16]. Quantitative shotgun proteomics using the iTRAQ was also employed to characterize the changes in the *Arabidopsis* phosphoproteome during the *Pseudomonas syringae* pv. *tomato* DC3000 infection [74]. Meanwhile, TMTs have been used mainly in stress-related studies. The TMT quantitative proteomics was used to discover the up- and down-regulation of 63 proteins and 39 proteins, respectively that involved in rice (*Oryza sativa*) cold-responsive pathway. In another study using the TMT tags, significantly differentially expressed proteins were found in the rice shoot after root chilling treatment, which include abscisic acid-responsive and drought-associated proteins. Liu and co-workers had also reported 22 up-regulated proteins involved in the antioxidant defense pathway, cell wall polysaccharide remodeling and cell metabolism process, in response to copper (Cu) stress in cell wall of *Elsholtzia splendens* [91]. Proteome-wide iTRAQ analysis has recently been employed in oil palm studies to reveal differentially expressed proteins involved in important metabolic processes such as fatty acid biosynthesis throughout different fruit developing stages [15, 16]. Table 2 listed some of the key advantages and disadvantages of the proteomics techniques commonly used in crop proteomics research.

4 Data Mining

Model plants are customarily used to investigate the physiological processes of cells, tissues, organelles or whole organisms. Simplicity of study design, biological relevance

Table 2 Comparison of the advantages and disadvantages of the common proteomics strategies used for oil palm and major crop research

Proteomics technology	Concept	Advantages	Disadvantages	Plant	References
Gel-based quantitation	Protein separation according to isoelectric points (pI) and molecular weight	Cheap, quick estimation of protein molecular weight, pI, quantify and post-translational modifications	Laborious, gel to gel inconsistencies, unable to detect low abundance proteins unless silver stain is applied, poor separation for hydrophobic or high pI proteins	Oil palm, banana, date palm, sugarcane	Tan et al. [38], de Carvalho Silva et al. [18], Kumaravel et al. [92], Sghaier-Hammami et al. [93], Daim et al. [13, 24], Hassan et al. [12], Al-Obaidi et al. [20, 21], Amalraj et al. [28]
Fluorescence 2D difference gel electrophoresis (2D-DIGE)	Up to three protein samples can be labelled with fluorescent dyes and pooled for separation. Ratios are measured for quantification	Gel to gel reproducibility as different protein samples are pooled before separation, reduced bias from technical and experimental variations, highly sensitive	Fluorescent dyes are expensive	Oil palm, sugarcane	Rahman et al. [94], Ooi et al. [14]
Pro-Q diamond	Fluorescent dye binds to phosphate group of phosphorylated protein	Binding of dye is highly specific and reproducible	Stain is expensive	Oil palm, rapeseed, rice, maize	Agrawal and Thelen [55], Agrawal and Thelen [54], Chen et al. [95], Luis et al. [96], Lau [97]
Label-based quantitation (ICAT)	Specific chemical group binds to cysteinyl residue before protein digestion. Mass shift of 8 kDa in the MS scan and the ion intensity are compared for relative quantitation	Reduce sample complexity for MS scan as tagged cysteine-containing peptides are purified by affinity chromatography. Vial to vial variation is minimized as tagged proteins from different samples are pooled	Only detect cysteine-containing peptide	Barley	Hagglund et al. [79]
Isobaric tag for relative and absolute quantitation (iTRAQ)	Chemical tags (with different mass) binds covalently to N-terminus and lysine side chain	Peak intensity-based relative and absolute quantitation for up to eight samples can be done in a single LCMS run, thus reduced technical variations	Mass tags are expensive, peptide co-fragmentation occurrence, only quantitate same peptide from different experiments	Oil palm, maize, soybean, grape	Cui et al. [98], Ji et al. [99], Luckner et al. [90], Loei et al. [16], Lau et al. [15]
Tandem mass tags (TMT)	Similar to iTRAQ	Peak intensity-based relative quantitation for up to ten samples can be done in a single LCMS run, thus reduced technical variations	Mass tags are expensive	Rice, grapevine	Lin et al. [100], Wang et al. [101], Rattanakan et al. [102]

Table 2 (continued)

Proteomics technology	Concept	Advantages	Disadvantages	Plant	References
Label-free quantitation	Peptides are separated using C18 based columns and can be coupled to another column such as strong ion exchange for better separation	Label-free approach as the peak intensity are compared for quantitation. Also suitable for all types of organisms. Two-dimensional resolution of peptides and unbiased identification of proteins including low abundance proteins, extreme hydrophobicity or pt	Require many biological and technical replicates for each analyzed sample. May not always be reproducible as it is highly dependent on consistent liquid chromatography runs	Oil palm, soybean, sugarcane, rice	Khan and Komatsu [103], Timabud et al. [104], Wu et al. [105], Heringer et al. [106]

and economics has an important impact on the plant models employed [25]. The green plants or *Viridiplantae* have only 100 species of completed and publicly available plant genomes in 2014 according to CoGepedia (<http://genomerevolution.org>) and plaBi (<http://plabipd.de/>) [107]. Presently, there are over 369,000 known species of flowering plants but the model plants only represent about 0.1% of the known species. The existing plant genomes are the classical *A. thaliana* (thale cress), economically important crop plants such as *Glycine max* (soybean), *Hordeum vulgare* (barley), *Medicago truncatula* (barrel medic), *Populus trichocarpa* (poplar), *Vitis vinifera* (wine grape), *O. sativa* (rice), *Sorghum bicolor* (sorghum) and *Zea mays* (maize), as well as other plants like the *Brachypodium distachyon* (purple false brome) [108]. However, none of these plant genomes are completely annotated [40] as the genome annotation tools remain decidedly lacking [109]. Moreover, only *O. sativa*, *S. bicolor*, *Z. mays* and *B. distachyon* are monocotyledons while the rest of the organisms are dicotyledons. Given that oil palm species is a monocotyledon, from the technical outlook, those plants mentioned earlier are unlikely to be suitable as a model organism in any proteomics study.

Complete genome sequences also form the foundation for comprehensive system biology studies by providing the potential of a complete parts list of protein and RNAs of the studied organism [109]. Encouragingly, a comprehensive genome sequencing project led by the Malaysian Palm Oil Board and St. Louis based Orion Genomics, USA for the two key oil palm species, *E. oleifera* and *E. guineensis* was completed in 2010. A total of nearly 35,000 genes were predicted from assembled sequences and transcriptome data of 30 tissue types [7]. Uthapaisanwong and co-workers had also successfully characterized the oil palm chloroplast genome sequence [110]. There are 41,887 non-redundant partial sequences of *E. guineensis* proteins currently available in NCBI protein database (as of 16th November, 2017). This information can significantly support the oil palm proteomics study.

Protein identification and characterization with mass spectrometry efforts could also be significantly ameliorated with the availability of expressed sequence tag (EST) sequences [111]. A large collection of 37,743 *E. guineensis* ESTs had been deposited in the NCBI database. The EST databases are indispensable as those sequence tags can be translated into the six reading frames to identify proteins (homology based) using appropriate software. Nonetheless, the size and quality of EST databases have profound effects on the outcome of the protein identification. The usual limitations of EST databases are that more often than not, most proteins are either not or poorly denoted by short EST sequences that only partly cover the whole protein sequence. Bases misread, insertion or deletion errors during sequencing of ESTs can lead to high error rates (about 0.3%) in EST

sequences, thus reducing the accuracy of peptide matching [112, 113]. Successful protein identifications with only peptide mass fingerprints employing EST databases are not feasible due to the limitations with EST databases. In addition, EST sequences are rarely sufficient in providing significant protein coverage and a satisfactory number of matching peptides [114].

Reliance on complete plant genomes can be lessened as annotating the biological function of proteins can be facilitated by a homology based approach. According to Carpentier et al. [40] and Remmerie et al. [25], there is a requirement for cross-species analysis. When using other species for protein identification with mass spectrometric data, orthologue sequences are preferable as they are more likely to share similar functions [115]. Homologous sequences originate from a sequence in a common ancestor. The sequences are considered different or orthologues when they diverged by a speciation (inter-species) event. Paralogous sequences are sequences that came from a common ancestor and are present in the same genome. However, duplication event (intra-species) transpired in the sequences produce paralogous sequences which may or may not share similar functions [116].

Database dependent- and independent strategies are the two approaches used to execute confident cross-species protein identification. In the former approach, search engines such as Mascot (<http://www.matrixscience.com/>) are used to search peptide sequence data that contains precursor ion mass and a list of product ion masses against a taxonomically confined database [117]. Only a massive amount of peptide masses generated can guarantee its success as this increases the matching probability of several peptides to the homologous protein. This has been demonstrated for pea chloroplast proteins [118] and maize proteins [119]. In a database independent strategy, fragmentation spectra are utilized to obtain de novo peptide sequences [120]. MSBLAST, which uses a combination of BLAST search and peptide de novo sequences, had been adapted for tandem mass spectrometry data to increase the accuracy and hit rate of protein identification [121]. Combination of top-down and bottom-up mass spectrometry is the concept of TBNovo for de novo peptide sequencing to increase sequence coverage [91]. Essentially, tandem mass spectrum from a top-down analysis is utilized as a scaffold while bottom-up tandem mass spectra are aligned to the scaffold. DeepNovo is another model for de novo peptide sequencing which is able to perform complete protein sequence assembly without any reference databases [19]. The deep neural network model learns the features of tandem mass spectra, fragment ions and sequence patterns of peptides to do de novo sequencing. Analyzing modified (in peptide sequences) proteins without completed and annotated genomes has proved to be a daunting effort. The identification of modified peptides and its modification sites

are essentially based on single amino acid identifications and become largely irrelevant when the peptide sequence is not available in any plant database. Application of a de novo sequencing strategy is able to facilitate the identification of modified peptides and may even help to locate the modification site, albeit with requirements for high quality tandem mass spectrometry spectra and certain preferable fragmentation techniques such as electron capture dissociation or electron transfer dissociation [25, 40].

Information on protein localization also helps in understanding the function of proteins and their biological inter-relationships. The Subcellular location database for Arabidopsis proteins (SUBA4) provides the hypothetical localization of many proteins that were identified in various sub-plastidial compartments in *A. thaliana* [122, 123]. LocSigDB is another database that contains 533 protein subcellular locations signals based on 518 experimentally confirmed and published research works [124]. The localization signals are for eight distinct subcellular locations in mainly eukaryotic cell, such as ‘Nuclear localization signal’ and ‘Mitochondrial targeting signal’. Plant Proteome Database (PPDB) was launched in 2004 for *A. thaliana* and maize (*Z. mays*) [125]. PPDB was developed to accommodate plant plastids, but over time, the database expanded to cover the entire proteomes of those two plants. The database consists of cell type-specific proteomes (maize) or specific sub-organelle proteomes such as chloroplasts, thylakoids and nucleoids as well as whole leaf proteome (maize and *A. thaliana*). More than 16,414 *A. thaliana* proteins, prominently from the plastids, have been assigned with subcellular locations. Table 3 listed other open-source software available to analyze large proteomics data.

5 Significance of Post-translational Modifications

Almost all proteins are modified in some way following protein biosynthesis. Many physiological responses result from differential protein modifications rather than changes in protein expression levels. These modifications do not create novel proteins but rather a new ‘protein species’ since the translated protein sequence remains unaltered [146, 147]. The modifications occur through covalent binding of functional groups such as phosphates, sulphates, carbohydrates and lipids [148]. This event, which is known as PTM, is one of the key mechanisms that changes the properties of a protein in cells and greatly enhances the structural diversity and functionality of proteins. This is feasible because PTMs provide a larger repertoire of chemical properties than is possible using the 20 amino acids specified by the genetic code. Protein PTMs could result in alterations in activity, localization, production, interactions with other proteins and

Table 3 Common open-source software to analyze multi-omics dataset

Function	Software	Purpose	Model/algorithm	References
Processing and comparison of raw data	MSight 3.0	Comparison of MS1 feature in two dimensional view and measurement of the distances in <i>m/z</i> and retention time	Not stated	Palagi et al. [126]
	MZmine 2	Data processing and visualization including filtering, peak picking, peak list deisotoping, alignment, gap filling and normalization for statistical analysis	Mexican hat wavelet	Pluskal et al. [127]
Quantification and analytical work	TOPPView	Viewing and inspection of MS and HPLC-MS data and metadata	OpenMS	Sturm and Kohlbacher [128]
	MaxQuant 1.6.2.3	Analyze large-scale mass spectrometric data and supports all labelling-based and label-free quantitation	Comprised of a set of algorithms, including peak detection and scoring of peptides, mass calibration and database searches for protein identity. It also provides statistical summaries	Tyanova et al. [129], Cox and Mann [130]
	Perseus 1.6.1.3	Multivariate quantitative on protein abundance data from mass spectrometry-based analysis including interpreting protein quantitation, interaction and post-translational modification	Normalization, pattern recognition, multiplexhypothesis testing, machine learning module	Tyanova and Cox [131]
Unknown protein sequencing	DeNovoGUI 1.16.2	Visualize and analyse chromatographic features and spectra in batch. Also allows customization of methods, algorithms, filters, detectors and support manual interaction such as peak integration and quantitation	Eclipse Rich Client Platform	Wenig and Odermatt [132]
Targeted protein quantitation	Skyline 4.1	De novo peptide sequencing and validation using BLAST feature	PepNovo (Hidden Markov models)	Muth et al. [133]
		Building selected reaction monitoring/multiple reaction monitoring, parallel reaction monitoring transition list for targeted MS/MS. Also used for quantitative measurements of labelled-based quantitation	MSData component of ProteoWizard data access library	MacLean et al. [134]

Table 3 (continued)

Function	Software	Purpose	Model/algorithm	References
Functional annotation	Database for annotation, visualization and integrated discovery, DAVID 6.8	Provides comprehensive tools for extracting biological meaning behind large list of genes and functional annotations based on GO terms. Also provides visualization of proteins on BioCarta and KEGG pathway maps	DAVID gene functional classification tool and DAVID functional annotation clustering tool	Huang et al. [135]
Protein interaction and pathways	Search tool for the retrieval of interacting genes/proteins, STRING 10.5	Biological database and web resources containing known and predicted protein–protein interaction	Multitude of algorithms	Szklarczyk et al. [136], Szklarczyk et al. [137]
	Reactome 65	Achieve of biological processes and contains tools to visualize, interpret and analyse pathway	Multitude of algorithms	Croft et al. [138], Fabregat et al. [139]
	Kyoto encyclopaedia of genes and genomes, KEGG 87.1	Database resource containing biological system information generated from molecular-level information	Japanese GenomeNet service	Kanehisa and Goto [140], Ogata et al. [141]
	PathVisio 3	For biological pathway editing, visualization and analysis	Open Service Gateway initiative (OSGi)	Kutmon et al. [142]
	Cytoscape 3.6.1	For visualization of complex networks and integrate these networks with other data types such annotation and gene expression profiles	Core software component	Shannon et al. [143]
Protein domain and motif analysis	Pfam 31.0	Identify domain within the protein sequence through multiple sequence alignment for Pfam matches	Hidden Markov models (HMM)	Finn et al. [144]
	InterPro 69.0, InterProScan 5.30–69.0	Classify proteins into families and determine the presence of protein domain. InterProScan scans protein sequence against InterPro's predictive models (from 14 different databases such as PANTHER, SMART TIGRFAMs) to determine the potential functionality		Finn et al. [145]

half-life [149–151]. Modifications are often permanent, but some modifications, such as phosphorylation, are reversible and can be used to switch protein activity ‘on’ and ‘off’ in response to intracellular and extracellular signals. For example, in a signal transduction process, kinase cascades are activated or inactivated through reversible addition and removal of phosphate groups. The esterification of an amino acid side chain through the addition of a phosphate group introduces a strong negative charge, which can subsequently modify the conformation of the protein and alter its stability, activity and potential to interact with other molecules. Genomic sequencing has revealed that protein kinases are probably coded by 2–3% of all eukaryotic genes [152]. PTM is therefore a dynamic phenomenon with a central role in many biological processes. Generally, in regulatory pathways, the status of serine, threonine and tyrosine is regulated by protein kinases and phosphatases [153, 154]. Interference with the activities of the kinases and phosphatases indirectly disrupts these regulatory pathways and may cause disease [155, 156].

The complexity of the proteome is increased significantly by PTMs, particularly in eukaryotes where many proteins exist as a heterogeneous mixture of alternative modified forms. Ideally, it would be possible to catalogue the proteome systematically and quantitatively in terms of the types of PTMs that are present, and specify the modified sites in each case. However, such attempts are complicated by the sheer diversity involved and the transient nature of certain modifications. Every protein could potentially be modified in hundreds of different ways, and might contain multiple modification target sites allowing different forms of modification to take place either singly or in combinations. Thus, it remains the case that most PTMs are discovered unintentionally when individual proteins, complexes, or pathways are studied. It is impossible to predict modifications accurately from the genome sequence. Even when a definitive modification motif is present; it is not necessarily the case that such or any modification will happen.

Until recent years, the analysis of PTMs at the proteomics level has received limited consideration due to the lack of appropriate techniques [148]. However, improved separation methods can resolve different post-translational variants, and gels can be stained with reagents that recognize particular types of modified proteins. Mass spectrometry is at present the method of choice to characterize chemical additions and substitutions. Mass spectrometry analysis can be used to identify peptides carrying chemical adducts and can deduce their positions in the protein sequences.

Signaling proteins and regulatory molecules, which play a vital role in functioning, are typically present in lower amount in the cell. They are also often regulated by phosphorylation. Since the stoichiometry of phosphorylation is usually low, the modified target protein may be present in

limiting amounts and may be difficult to detect and quantify. Ultimately, even if adequate amounts of a particular variant are available, a large quantity of the sample is required for the full characterization of modifications compared to the relatively simple matter of protein identification. Currently, affinity-based techniques are employed to improve the chances of detecting their targets by isolating sub-proteomes with particular types of modification [157].

Investigations into PTMs and differentially expressed proteins are essential to comprehend cellular responses towards changes in environmental conditions [158]. It is clear that plants induce a complex array of pathways and protein phosphorylation cascades during biotic and abiotic stresses [159]. There are over 200 possible PTMs that have been identified and reported [160–162]. Until recently, more than 90,000 of individually modified amino acid residues were found [163], emphasizing the importance of these PTMs in a functional proteome. Phosphorylation is the most common and extensively studied PTM using mass spectrometry approaches [30, 148, 149, 164–170]. The justification is that phosphorylation is one of the primary mechanisms in cellular process regulation [171].

Mass spectrometry-based approach has enabled absolute and relative quantitation of peptides and their PTMs. Internal standard peptides are employed for absolute quantitation for certain proteins and their defined PTMs [172]. Relative quantitation is performed with either the peptide intensity profiling (PIP) or SIL using stable isotope-encoded chemical precursor molecules or alkylating reagents [149, 173].

In an oil palm phosphoproteomics study by Lau et al. [17], 3-enoyl-ACP reductase was deactivated through phosphorylation to direct the metabolic flux towards the production of palmitoyl-ACP during the final phase of the fatty acid biosynthesis. Palmitoyl-ACP is a crucial precursor for the biosynthesis of unsaturated oleic acids. Furthermore, the study discovered that subunit biotin carboxylase of acetyl-CoA carboxylase was also deactivated through phosphorylation at the same phase. The deactivation would have stopped the production of malonyl-ACP, which is the carbon precursor for the initial stage of fatty acid biosynthesis in the oil palm.

5.1 Tracking the Phosphopeptides

A wide array of approaches can be used to scrutinize phosphorylation changes in cell or tissues. Radiolabeling is a classical technique that uses radiolabeled ^{32}P -orthophosphate to detect phosphoproteins. Radioactivity can be very inconvenient, harmful and detrimental in the long term, both to the users and samples [174, 175]. Alternatively, after separation by two-dimensional gel electrophoresis [176], phosphoproteins can be directly visualized on the gel using phosphospecific fluorescent stains and phosphospecific antibodies, which are non-radioactive [52, 53, 56, 177,

178]. Immuno- or Western blot is the most common method used to assess the phosphorylation state of a protein using phosphospecific antibodies (for phosphorylated tyrosine, serine and threonine) transferred from a one-dimensional or two-dimensional gel electrophoresis [179, 180]. In direct staining, phosphospecific stains such as a fluorescent phosphosensor dye, Pro-Q Diamond (Invitrogen) bind directly to the phosphate moiety of phosphoproteins [56, 57, 181]. The advantages of this stain are in its compatibility with other staining methods and the ensuing mass spectrometry analysis. This is particularly crucial when trypsinizations are performed directly on the gel pieces. A similar phosphospecific staining kit called Phos-tag had been used previously in which a Zn^{2+} ion chelator with high selectivity was coupled to a fluorophore [182]. The suitability of these stains in phosphoproteome analysis had been described in previous reports. Agrawal and Thelen [55] identified 70 non-redundant phosphoproteins that belong to the major functional classes from a Pro-Q Diamond stained two-dimensional gel containing rapeseed (*Brassica napus*) proteins. However, while phosphoproteins could be detected, the stains would not indicate the phospho-sites, which is vital in the characterization of phosphorylation events. Special techniques are used to investigate membrane phosphoproteins due to the limitations in two-dimensional gel electrophoresis technique. Integral membrane proteins tend to aggregate during the isoelectric focusing migration and thus, it is not possible to separate them in the second dimension of two-dimensional gel electrophoresis.

The low abundance of phosphorylated proteins in cellular extracts and their relatively low degree of phosphorylation pose major challenges [183–185]. In mass spectrometric analysis, non-phosphorylated peptides often compete with the phosphorylated peptides for ionization. As a result, many phosphoprotein peaks are difficult to detect, either because they have low signal to noise ratio or they are not ionized at all. Therefore, to tackle this obstacle, enrichment techniques, which are commonly applied prior to separation using liquid chromatography, have been used. Immobilized metal affinity chromatography (IMAC) is one of the methods that are used to enrich phosphopeptides from complex mixtures based on affinity of positively charged metal ions (Fe^{3+} , Al^{3+} , Ga^{3+} or Co^{2+}) towards phosphate moieties. Iminodiacetate and nitrilotriacetate are the prototypical metal-binding ligands used in IMAC stationary phases [186–189]. The Fe(III)–NTA complex is perhaps the most frequently utilized to enrich phosphopeptides although the use of other metal–ligand complexes had also been reported [190, 191]. Most recently, Zr(IV)–phosphonate immobilized on various stationary phases had also been employed for phosphopeptide enrichment by several groups [192–196]. The phosphopeptides can be eluted by different salt- and/or pH gradients prior to mass spectrometry analysis. Nonetheless, several

challenges arise when using IMAC. Leaching of ions from the column during enrichment steps, non-specific binding of peptides that contain the acidic amino acids glutamic and aspartic acid and higher specificity for multiply phosphorylated peptides are amongst those complications [186].

Metal oxide affinity chromatography (MOAC) is another valuable technique to isolate phosphopeptides from complex mixtures with high selectivity and recoveries [186, 197–199]. The metal oxides are often more stable at high temperatures and broad pH range [200]. Titanium oxide (TiO_2) is the most popular metal oxide resin used as a selective affinity support to capture phosphorylated peptides [201–206]. At acidic pH, TiO_2 has a positively charged surface [207] that permits very selective enrichment of phosphopeptides from complex samples by their affinities (phosphate groups) toward porous TiO_2 particles (Titansphere) [208]. Water-soluble phosphates are desorbed under alkaline conditions. Strong cation exchange and titanium dioxide-type columns have both been used in phosphopeptide enrichment and SILAC for quantitation to study phosphorylation changes [209].

Technical variations and bias in quantitative analyses are often reported to occur after phosphopeptide enrichment [43]. Nonetheless, successful identification and quantitation of phosphopeptides has been reported using a combination of enrichment strategies and label-free quantitation as in the case of *Arabidopsis* phosphopeptides from a plasma membrane fraction following sucrose treatment [58] and a hypersensitive response study in tomato plants [59]. In addition, iTRAQ labelling has been successfully used to quantify phosphorylated peptides in *Arabidopsis* cells as their defense response to *P. syringae* induction (elicitors) [74]. As a rule of thumb, it is more effective to perform chemical labelling prior to any enrichment strategies due to the fact that enrichment steps confer technical bias in quantitative analyses [43].

Hydrophilic interaction chromatography can also be used as in the pre-separation stage of peptides prior to phosphopeptide enrichment such as IMAC or TiO_2 affinity purifications, in addition to MudPIT LC. Hydrophilic interaction liquid chromatography (HILIC) separates polar biomolecules by the binding of the polar biomolecules to the neutral, hydrophilic stationary phase in hydrophilic interaction chromatography through hydrogen bonds. These bonds can be broken by reducing the organic composition in the mobile phase and the peptides eluted based on their polarities [210].

PTM occurrences can also be detected through neutral loss-triggered tandem mass spectrometry (NLMS³) and SRM approaches. In a phosphoproteomics study, the phosphate group of a phosphopeptide is relatively labile and tend to break away during collision-induced fragmentation, in the form of a phosphoric acid (HPO_3 or H_3PO_4). Hence, the fragmentation of phosphoamino residue-containing (serine,

threonine and tyrosine) precursor ions generates neutral losses of 80 Da (HPO_3) or 98 Da (H_3PO_4) [211]. Usually, the mechanism for loss of H_3PO_4 from phosphoserine and phosphothreonine containing peptide ions is the result of a β -elimination reaction [212]. In a β -elimination reaction, the hydrogen atom on the α -carbon of the phosphorylated amino acid residue is transferred to the phosphate oxygen. As a result, dehydroalanine-(69 Da) or dehydroaminobutyric acid-(83 Da) containing product ions from phosphorylated serine or threonine residues, respectively, and H_3PO_4 are produced. Loss of H_3PO_4 is more dominant for serine phosphorylated peptides and in a lesser extent in threonine phosphorylated peptides. This might be caused by the steric hindrance of the β -methyl group in the side chain of threonine [213]. Tyrosine phosphorylated peptides give a much lower extent of neutral loss and these are in the form of HPO_3 . Mass spectrometry-based strategies such as the NLMS³ and SRM are essentially built on the detection of the characteristic neutral loss generated during the CID fragmentation of phosphoamino-containing peptides. In the NLMS³ mode of operation, the diagnostic neutral loss of H_3PO_4 (98 Da) from the precursor ion in a tandem MS scan automatically triggers the MS³ fragmentation of the neutral loss precursor ion. The aim of the MS³ is to compensate for the lack of sequence-specific information in the MS² spectra of phosphorylation-modified peptides [213] although a study by Villen and co-workers indicated that the collection of MS³ scans did not improve the informative spectra of the peptides identified [214]. MS³ operates in a data-dependent manner, in which the MS³ is triggered by the presence of an intense product ion peak with the mass of a neutral loss. This strategy has been extended to detect phosphopeptides in the oil palm mesocarps by Lau et al. [17]. These neutral loss species from the product ions were calculated on the basis of the product ion mass and charge state, resulting in the neutral product ions of m/z 48.99 or m/z 32.66, relative to the doubly or triply charged phosphorylated product ion. A common problem with a neutral loss scan to detect phosphopeptides is that unassigned peptides may generate ions with a mass similar to the neutral losses as well [214, 215]. There are also instances when a phosphopeptide fails to generate the specified neutral loss and therefore are not detected [216].

SRM is an ideal complementary technique to reliably target and quantitate low abundance phosphopeptides of interest [217–221]. SRM is predominantly performed on a triple quadrupole mass spectrometer as the availability of additional mass filter (third quadrupole, Q3) is exploited to isolate targeted fragment ion for MS². However, in the study by Lau and co-workers, a quadrupole-TOF was used to scan the targeted precursor ion for any loss of neutral loss species (98 Da) instead [17]. Absence of the Q3 mass filter in the quadrupole-TOF implies that only neutral loss species at

98 Da correspond to the loss of H_3PO_4 can be detected after collision-induced fragmentation of the selected precursor ions in Q2. The ideal prerequisites to targeted SRM experiments are the prior knowledge of the primary sequence, type of phosphorylation, sequence motif and predicted fragmentation pathways to identify the potential phosphopeptides (for example, a neutral loss of 98 Da from phosphoserine and phosphothreonine peptides but not phosphotyrosine peptide).

5.2 Prediction of Post-translational Modifications

Automated prediction of PTM sites is one of the main interest areas for bioinformatics investigations. In vivo and in vitro determinations of modified proteins and their PTM sites are not only time-consuming and tedious, but often restricted to the availability and optimization of enzymatic reactions in order to determine the type of modifications and sites [222–224]. Tandem mass spectrometry spectra offer the most informative fingerprints of modified peptides. The spectra encode not only peptide sequences, but also the masses and sequence positions of modifications. For these reasons, computational techniques have been employed to manage the massive amounts of fragmentation spectra, modified protein determination and individual PTM site identification with high accuracy as well as efficiency [225]. The current PTM prediction tools basically are classed into four major groups based on their types of classification schemes [226].

The first group comprises general PTM related resources such as *PROSITE* [227] which predicts types of PTMs based on their sequence pattern consensus. Several signature recognition methods are combined to probe a query protein sequence against observed protein signatures. The *Scansite* tool predicts kinase-specific and signal transduction relevant motifs [228]. Conserved sequence motifs represent imprints of important biochemical properties or biological functions of those proteins.

The second group consists of various neural network prediction tools. These tools cover phosphorylation related prediction servers such as *NetPhos* [229] and *NetPhosK* [222, 230]. *NetPhosK* is the most popular since the server allows a preferred 'threshold' value to be indicated during prediction.

The third group of the prediction tools encompasses different support vector machine based prediction techniques. These methods are constructed on the basis that adjacent residues to the phospho-sites represent the main determinant for kinase specificity [224, 231]. For instant, *PredPhospho* [232] aims to predict phosphorylation sites and the type of kinase that acts at each site. *AutoMotif-Server* [233] also predicts PTM sites in protein sequences using support vector machine classifier with both linear and polynomial kernels. *KinasePhos 2.0* is the web server

Table 4 In silico tools to study some of the common post-translational modifications

PTM type	Prediction method	Tool	Description	References
Phosphorylation	Sequence pattern consensus or weight matrix	PROSITE 20.54	Use a collection of several rules based on profiles and pattern known as ProRule	Sigrist et al. [227, 239]
		Scansite 4.0	Predicts kinase-specific and signal transduction motifs	Yaffe et al. [228], Obenaus et al. [240]
	Artificial neuronal network	NetPhos 3.1 NetPhosK 1.0 Server	Prediction is based on 17 different kinases Produce neural network predictions of kinase specific phosphosites	Blom et al. [229] Hjerrild et al. [230], Martinez-Marquez et al. [220]
	Support vector machine	PredPhospho	Predict phosphorylation sites and type of kinase that acts at each site using consensus sequences	Kim et al. [232]
		AutoMotifServer/AMS 4.0	Predicts 88 different types of the single amino acid PTM sites	Plewczynski et al. [233, 241], Basu and Plewczynski [226]
Machine learning based		KinasePhos 2.0	Identify protein kinase-specific phosphorylation sites based on amino acid residue sequences and coupling patterns	Wong et al. [234]
		Prediction of PK-specific phosphorylation site (PPSP)	Adopt Bayesian decision theory to predict about 70 protein kinase-specific sites	Xue et al. [242]
		Ascore	Use the presence and intensities of site-determining ions to calculate the probability of exact phosphosites	Beausoleil et al. [236]
		PhosphoScore	Similar to Ascore and also considers both the match quality and the normalized intensity of observed peaks compared to a theoretical spectrum	Ruttenberg et al. [237]
		PTMap	Sequence alignment based to identify PTM, polymorphisms and to map sites of unspecified PTMs	Chen et al. [238]
Phosphorylation, acetylation, N-glycosylation		PHOSIDA	List PTM sites associated with particular proteomes. Integrates structural and evolutionary information	Gnad et al. [235]
Glycosylation	Random forest	GPP	Predict the location of N-linked and O-linked glycosylation sites	Hamby and Hirst [243]
		NGlycPred	Discriminate glycosylated from non-glycosylated motif	Chuang et al. [244]
S-Nitrosylation		GlycoMine	Identify C-linked, N-linked, and O-linked glycosylation sites	Li et al. [245]
	Support vector machine	GlycoPP 1.0	Predict potential N- and O-glycosites	Chauhan et al. [246]
	Group-based prediction	GPS-SNO 1.0	Predict putative S-nitrosothiol (SNO) based on a databases of 504 experimentally verified SNO	Xue et al. [247]

Table 4 (continued)

PTM type	Prediction method	Tool	Description	References
Methylation	Information gain feature optimization	PSSMe	Prediction is based on single or combined species-specific models	Wen et al. [248]
Acetylation	Support vector machine	N-Ace	Prediction is based on structural characteristics such as amino acid composition	Lee et al. [249]
	Logistic regression classifiers	LAceP	Integrate different biological features (amino acid physicochemical property, transition probability matrix and position-specific symbol composition) to predict lysine acetylation	Hou et al. [250]

to identify protein kinase-specific phosphorylation sites based on amino acid residue sequences and coupling patterns [234]. *PHOSIDA* is also capable of predicting phosphosites [235].

The final group consists of remaining types of machine learning based PTM prediction tools. *Prediction of PK-specific phosphorylation site (PPSP)* adopted the Bayesian decision theory to predict kinase-specific phosphorylation sites and has been reported to produce precise prediction of the probable phosphorylation sites for about 70 protein kinase groups [226]. *PPSP_balanced* model worked remarkably for all types of protein families. *Ascore*, a probability-based score that was developed by Beausoleil and his colleagues, used the presence and intensities of site-determining ions in tandem mass spectrometry spectra to calculate the probability of exact phosphorylation site localization [236]. *Ascore* re-evaluates the results from search engine on phosphopeptide and designates a confidence value to each of the phosphorylated site. *PhosphoScore* is another algorithm which acts similarly to *Ascore*. *PhosphoScore* considers both the match quality and the normalized intensity of observed spectra peaks compared to a theoretical spectrum. *PhosphoScore* was employed successfully in the studies done by Ruttenberg et al. [237]. *PTMap* is a sequence alignment software used to identify protein PTMs and polymorphisms [238]. The selection of peak, adjustment of inaccurate mass shifts and precise localization of PTM sites are the features that improved searching speed and accuracy of *PTMap*. This software is the first algorithm that contains a scoring system which concentrates on unmatched peaks to eliminate false positives, thus increasing the accuracy and sensitivity of the PTM identifications. Table 4 summarized some of the main *in silico* tools to study several common PTMs.

There are numerous database which provides information on PTMs. *dbPTM* database [251] gathers various information such as the catalytic sites, protein domains and protein variations, in addition to these software or tools. These databases include a majority of experimentally validated PTM sites from SwissProt and *Phospho.ELM*. *Phospho.ELM* comprises over 40,000 amino acid serine, threonine and tyrosine non-redundant phosphorylation sites from vertebrates, *Drosophila melanogaster* and *Caenorhabditis elegans* [252]. Similarly, *PHOSIDA* (<http://www.phosida.com>) comprises more than 80,000 phosphorylated, *N*-glycosylated or acetylated sites from nine different species [235]. For each of the phosphosites, *PHOSIDA* lists matching kinase motifs, predicts secondary structures, conservation pattern, and its dynamic regulation upon stimulus. Unfortunately, none of these species are plants. *PhosphoSitePlus* (<http://www.phosphosite.org>) has 130,000 non-redundant modification sites, primarily on phosphorylation, ubiquitination and acetylation [253].

6 Conclusions

In the post-genomics era for perennial oil crop improvement such as oil palm, it is crucial to first map the entire set of the proteins using the emerging proteomics technologies. Given the lack of protein-level information of the oil palm genes sequenced so far, a systematic effort using the tools of proteomics is essential to elucidating biological functions of interest based on these genomic sequences. Knowing the key controlling mechanisms for metabolic processes such as fatty acid production and plant defense towards pathogen through proteome-wide protein quantitation is significantly important. Subsequent PTM analysis and protein–protein interaction mapping can eventually help to predict the regulatory networks under different planting environments. These information are crucial to strategize breeding programs and to discover biological significant markers for oil palm fruit growth and development, to improve the yield and quality as well as to enhance the plant immunity towards various environmental stresses, in particular, diseases that has obstructed the optimal production of palm oil.

Acknowledgements The authors would like to thank the Director-General of the Malaysian Palm Oil Board for permission to publish this article and the funding received for the research projects.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no competing interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Woittiez LS, van Wijk MT, Slingerland M, van Noordwijk M, Giller KE (2017) Yield gaps in oil palm: a quantitative review of contributing factors. *Eur J Agron* 83(Supplement C):57–77. <https://doi.org/10.1016/j.eja.2016.11.002>
2. Kushairi A, Singh R, Ong-Abdullah M (2017) The oil palm industry in Malaysia: thriving with transformative technologies. *J Oil Palm Res* 29(4):431–439. <https://doi.org/10.21894/jopr.2017.00017>
3. Low E-T, Alias H, Boon S-H, Shariff E, Tan C-Y, Ooi L, Cheah S-C, Raha A-R, Wan K-L, Singh R (2008) Oil palm (*Elaeis guineensis* Jacq.) tissue culture ESTs: identifying genes associated with callogenesis and embryogenesis. *BMC Plant Biol* 8(1):62
4. Low E-TL, Rosli R, Jayanthi N, Mohd-Amin AH, Azizi N, Chan K-L, Maqbool NJ, Maclean P, Brauning R, McCulloch A, Moraga R, Ong-Abdullah M, Singh R (2014) Analyses of hypomethylated oil palm gene space. *PLoS ONE* 9(1):e86728. <https://doi.org/10.1371/journal.pone.0086728>
5. Singh R, Low ET, Ooi LC, Ong-Abdullah M, Ting NC, Nagappan J, Nookiah R, Amiruddin MD, Rosli R, Manaf MA, Chan KL, Halim MA, Azizi N, Lakey N, Smith SW, Budiman MA, Hogan M, Bacher B, Van Brunt A, Wang C, Ordway JM, Sambanthamurthi R, Martienssen RA (2013) The oil palm SHELL gene controls oil yield and encodes a homologue of SEED-STICK. *Nature* 500(7462):340–344. <https://doi.org/10.1038/nature12356>
6. Singh R, Low E-TL, Ooi LC-L, Ong-Abdullah M, Nookiah R, Ting N-C, Marjuni M, Chan P-L, Ithnin M, Manaf MAA, Nagappan J, Chan K-L, Rosli R, Halim MA, Azizi N, Budiman MA, Lakey N, Bacher B, Van Brunt A, Wang C, Hogan M, He D, MacDonald JD, Smith SW, Ordway JM, Martienssen RA, Sambanthamurthi R (2014) The oil palm *VIRESCENS* gene controls fruit colour and encodes a R2R3-MYB. *Nat Commun* 5:4106. <https://doi.org/10.1038/ncomms5106>
7. Singh R, Ong-Abdullah M, Low ET, Manaf MA, Rosli R, Nookiah R, Ooi LC, Ooi SE, Chan KL, Halim MA, Azizi N, Nagappan J, Bacher B, Lakey N, Smith SW, He D, Hogan M, Budiman MA, Lee EK, DeSalle R, Kudrna D, Goicoechea JL, Wing RA, Wilson RK, Fulton RS, Ordway JM, Martienssen RA, Sambanthamurthi R (2013) Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* 500(7462):335–339. <https://doi.org/10.1038/nature12309>
8. Singh R, Tan SG, Panandam JM, Rahman RA, Ooi LCL, Low E-TL, Sharma M, Jansen J, Cheah S-C (2009) Mapping quantitative trait loci (QTLs) for fatty acid composition in an interspecific cross of oil palm. *BMC Plant Biol* 9:114–114. <https://doi.org/10.1186/1471-2229-9-114>
9. Ting N-C, Yaakub Z, Kamaruddin K, Mayes S, Massawe F, Sambanthamurthi R, Jansen J, Low LET, Ithnin M, Kushairi A, Arulandoo X, Rosli R, Chan K-L, Amiruddin N, Sriharan K, Lim CC, Nookiah R, Amiruddin MD, Singh R (2016) Fine-mapping and cross-validation of QTLs linked to fatty acid composition in multiple independent interspecific crosses of oil palm. *BMC Genomics* 17:289. <https://doi.org/10.1186/s12864-016-2607-4>
10. Ting N-C, Zaki N, Rosli R, Low E-T, Ithnin M, Cheah S-C, Tan S-G, Singh R (2010) SSR mining in oil palm EST database: application in oil palm germplasm diversity studies. *J Genet* 89(2):135–145. <https://doi.org/10.1007/s12041-010-0053-7>
11. Lau BY, Deb-Choudhury S, Morton JD, Clerens S, Dyer JM, Ramli US (2015) Method developments to extract proteins from oil palm chromoplast for proteomic analysis. *SpringerPlus* 4(1):791. <https://doi.org/10.1186/s40064-015-1576-4>
12. Hassan H, Lau B, Ramli U (2014) Extraction methods for analysis of oil palm leaf and root proteins by two-dimensional gel electrophoresis. *J Oil Palm Res* 26:54–61
13. Daim LD, Ooi TE, Yusof HM, Majid NA, Karsani SA (2015) Optimization of protein extraction and two-dimensional electrophoresis protocols for oil palm leaf. *Protein J* 34(4):304–312. <https://doi.org/10.1007/s10930-015-9626-x>
14. Ooi T, Yeap W, Daim L, Ng B, Lee F, Othman A, Appleton D, Chew F, Kulaveerasingam H (2015) Differential abundance analysis of mesocarp protein from high- and low-yielding oil palms associates non-oil biosynthetic enzymes to lipid biosynthesis. *Proteome Sci* 13(1):28
15. Lau BYC, Morton DJ, Deb-Choudhury S, Clerens S, Dyer JM, Ramli US (2017) Differential expression analysis of oil palm fatty acid biosynthetic enzymes with gel-free quantitative proteomics. *J Oil Palm Res* 29(1):23–34
16. Loei H, Lim J, Tan M, Lim TK, Lin QS, Chew FT, Kulaveerasingam H, Chung MC (2013) Proteomic analysis of the oil palm fruit mesocarp reveals elevated oxidative phosphorylation activity is critical for increased storage oil production. *J Proteome Res* 12(11):5096–5109. <https://doi.org/10.1021/pr400606h>
17. Lau BY, Clerens S, Morton JD, Dyer JM, Deb-Choudhury S, Ramli US (2016) Application of a mass spectrometric approach to detect the presence of fatty acid biosynthetic

- phosphopeptides. *Protein J* 35(2):163–170. <https://doi.org/10.1007/s10930-016-9655-0>
18. de Carvalho Silva R, Carmo LS, Luis ZG, Silva LP, Scherwinski-Pereira JE, Mehta A (2014) Proteomic identification of differentially expressed proteins during the acquisition of somatic embryogenesis in oil palm (*Elaeis guineensis* Jacq.). *J Proteomics* 104(0):112–127. <https://doi.org/10.1016/j.jprot.2014.03.013>
 19. Tan HS, Jacoby RP, Ong-Abdullah M, Taylor NL, Liddell S, Chee WW, Chin CF (2017) Proteomic profiling of mature leaves from oil palm (*Elaeis guineensis* Jacq.). *Electrophoresis* 38(8):1147–1153. <https://doi.org/10.1002/elps.201600506>
 20. Al-Obaidi JR, Saidi NB, Usuldin SR, Hussin SN, Yusoff NM, Idris AS (2016) Comparison of different protein extraction methods for gel-based proteomic analysis of *Ganoderma* spp. *Protein J* 35(2):100–106. <https://doi.org/10.1007/s10930-016-9656-z>
 21. Al-Obaidi JR, Mohd-Yusuf Y, Razali N, Jayapalan JJ, Tey CC, Md-Noh N, Junit SM, Othman RY, Hashim OH (2014) Identification of proteins of altered abundance in oil palm infected with *Ganoderma boninense*. *Int J Mol Sci* 15(3):5175–5192. <https://doi.org/10.3390/ijms15035175>
 22. Syhanim S, Abrizah O, Mohamad Arif AM, Idris AS, Mohd Din A (2013) Identification of differentially expressed proteins in oil palm seedlings artificially infected with *Ganoderma*: a proteomics approach. *J Oil Palm Res* 25(3):298–304
 23. Al-Obaidi JR, Hussin SNIS, Saidi NB, Rahmad N, Idris AS (2017) Comparative proteomic analysis of *Ganoderma* species during in vitro interaction with oil palm root. *Physiol Mol Plant Pathol* 99(Supplement C):16–24. <https://doi.org/10.1016/j.pmpp.2017.02.001>
 24. Daim LDJ, Ooi TE, Ithnin N, Mohd Yusof H, Kulaveerasingam H, Abdul Majid N, Karsani SA (2015) Comparative proteomic analysis of oil palm leaves infected with *Ganoderma boninense* revealed changes in proteins involved in photosynthesis, carbohydrate metabolism, and immunity and defense. *Electrophoresis* 36(15):1699–1710. <https://doi.org/10.1002/elps.201400608>
 25. Carpentier SC, Panis B, Vertommen A, Swennen R, Sergeant K, Renaut J, Laukens K, Witters E, Samyn B, Devreese B (2008) Proteome analysis of non-model plants: a challenging but powerful approach. *Mass Spectrom Rev* 27(4):354–377. <https://doi.org/10.1002/mas.20170>
 26. Gooding PS, Bird C, Robinson SP (2001) Molecular cloning and characterisation of banana fruit polyphenol oxidase. *Planta* 213(5):748–757. <https://doi.org/10.1007/s004250100553>
 27. Wuyts N, De Waele D, Swennen R (2006) Extraction and partial characterization of polyphenol oxidase from banana (*Musa acuminata* Grande naine) roots. *Plant Physiol Biochem* 44(5–6):308–314. <https://doi.org/10.1016/j.plaphy.2006.06.005>
 28. Amalraj RS, Selvaraj N, Veluswamy GK, Ramanujan RP, Muthurajan R, Palaniyandi M, Agrawal GK, Rakwal R, Viswanathan R (2010) Sugarcane proteomics: establishment of a protein extraction method for 2-DE in stalk tissues and initiation of sugarcane proteome reference map. *Electrophoresis* 31(12):1959–1974. <https://doi.org/10.1002/elps.200900779>
 29. McCabe MS, Garratt LC, Schepers F, Jordi WJRM, Stoopen GM, Davelaar E, van Rhijn JHA, Power JB, Davey MR (2001) Effects of PSAG12-IPT gene expression on development and senescence in transgenic lettuce. *Plant Physiol* 127(2):505–516. <https://doi.org/10.1104/pp.010244>
 30. Jensen ON (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* 8(1):33–41. <https://doi.org/10.1016/j.cbpa.2003.12.009>
 31. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC (2000) The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* 21 (6):1104–1115. [https://doi.org/10.1002/\(SICI\)1522-2683\(20000401\)21:6%3C1104::AID-ELPS1104%3E3.0.CO;2-C](https://doi.org/10.1002/(SICI)1522-2683(20000401)21:6%3C1104::AID-ELPS1104%3E3.0.CO;2-C)
 32. van Wijk KJ, Baginsky S (2011) Plastid proteomics in higher plants: current state and future goals. *Plant Physiol* 155(4):1578–1588. <https://doi.org/10.1104/pp.111.172932>
 33. Hey J, Posch A, Cohen A, Liu N, Harbers A (2008) Fractionation of complex protein mixtures by liquid-phase isoelectric focusing. In: Posch A (ed) 2D PAGE: sample preparation and fractionation, vol 424. *Methods in Molecular Biology*™. Humana Press, Totowa, pp 225–239. https://doi.org/10.1007/978-1-60327-064-9_19
 34. Horth P, Miller CA, Preckel T, Wenz C (2006) Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Mol Cell Proteomics* 5(10):1968–1974. <https://doi.org/10.1074/mcp.T600037-MCP200>
 35. Bayer RG, Stael S, Csaszar E, Teige M (2011) Mining the soluble chloroplast proteome by affinity chromatography. *Proteomics* 11(7):1287–1299. <https://doi.org/10.1002/pmic.201000495>
 36. Mamone G, Picariello G, Ferranti P, Addeo F (2010) Hydroxyapatite affinity chromatography for the highly selective enrichment of mono- and multi-phosphorylated peptides in phosphoproteome analysis. *Proteomics* 10(3):380–393. <https://doi.org/10.1002/pmic.200800710>
 37. Aryal UK, Krochko JE, Ross ARS (2011) Identification of phosphoproteins in *Arabidopsis thaliana* leaves using polyethylene glycol fractionation, immobilized metal-ion affinity chromatography, two-dimensional gel electrophoresis and mass spectrometry. *J Proteome Res* 11(1):425–437. <https://doi.org/10.1021/pr200917t>
 38. Tan HS, Liddell S, Ong Abdullah M, Wong WC, Chin CF (2016) Differential proteomic analysis of embryogenic lines in oil palm (*Elaeis guineensis* Jacq.). *J Proteomics* 143:334–345. <https://doi.org/10.1016/j.jprot.2016.04.039>
 39. Schulze WX, Usadel B (2010) Quantitation in mass-spectrometry-based proteomics. *Annu Rev Plant Biol* 61(1):491–516. <https://doi.org/10.1146/annurev-arplant-042809-112132>
 40. Remmerie N, De Vijlder T, Laukens K, Dang TH, Lemièrre F, Mertens I, Valkenburg D, Blust R, Witters E (2011) Next generation functional proteomics in non-model plants: a survey on techniques and applications for the analysis of protein complexes and post-translational modifications. *Phytochemistry* 72(10):1192–1218. <https://doi.org/10.1016/j.phytochem.2011.01.003>
 41. Timms JF, Cramer R (2008) Difference gel electrophoresis. *Proteomics* 8(23–24):4886–4897. <https://doi.org/10.1002/pmic.200800298>
 42. Unlu M, Morgan ME, Minden JS (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18(11):2071–2077. <https://doi.org/10.1002/elps.1150181133>
 43. Bindschedler LV, Cramer R (2011) Quantitative plant proteomics. *Proteomics* 11(4):756–775. <https://doi.org/10.1002/pmic.201000426>
 44. Gerber IB, Laukens K, De Vijlder T, Witters E, Dubery IA (2008) Proteomic profiling of cellular targets of lipopolysaccharide-induced signalling in *Nicotiana tabacum* BY-2 cells. *Biochim Biophys Acta* 1784(11):1750–1762. <https://doi.org/10.1016/j.bbapap.2008.06.012>
 45. Chivasa S, Hamilton JM, Pringle RS, Ndimba BK, Simon WJ, Lindsey K, Slabas AR (2006) Proteomic analysis of differentially expressed proteins in fungal elicitor-treated *Arabidopsis* cell cultures. *J Exp Bot* 57(7):1553–1562. <https://doi.org/10.1093/jxb/erj149>
 46. Amey RC, Schleicher T, Slinn J, Lewis M, Macdonald H, Neill SJ, Spencer-Phillips PTN (2008) Proteomic analysis of a compatible interaction between *Pisum sativum* (pea) and the downy mildew pathogen *Peronospora viciae*. In: Lebeda A,

- Spencer-Phillips PN, Cooke BM (eds) The downy mildews—genetics, molecular biology and control. Springer, The Netherlands, pp 41–55. https://doi.org/10.1007/978-1-4020-8973-2_5
47. Schenkluhn L, Hohnjec N, Niehaus K, Schmitz U, Colditz F (2010) Differential gel electrophoresis (DIGE) to quantitatively monitor early symbiosis- and pathogenesis-induced changes of the *Medicago truncatula* root proteome. *J Proteomics* 73(4):753–768. <https://doi.org/10.1016/j.jprot.2009.10.009>
 48. Renaut J, Hausman J-F, Wisniewski ME (2006) Proteomics and low-temperature studies: bridging the gap between gene expression and metabolism. *Physiol Plant* 126(1):97–109. <https://doi.org/10.1111/j.1399-3054.2006.00617.x>
 49. Casati P, Zhang X, Burlingame AL, Walbot V (2005) Analysis of leaf proteome after UV-B irradiation in maize lines differing in sensitivity. *Mol Cell Proteomics* 4(11):1673–1685. <https://doi.org/10.1074/mcp.M500173-MCP200>
 50. Zhou S, Sauv  r R, Thannhauser TW (2009) Proteome changes induced by aluminium stress in tomato roots. *J Exp Bot* 60(6):1849–1857. <https://doi.org/10.1093/jxb/erp065>
 51. Gomez A, Lopez JA, Pintos B, Camafeita E, Bueno MA (2009) Proteomic analysis from haploid and diploid embryos of *Quercus suber* L. identifies qualitative and quantitative differential expression patterns. *Proteomics* 9(18):4355–4367. <https://doi.org/10.1002/pmic.200900179>
 52. Gerber IB, Laukens K, Witters E, Dubery IA (2006) Lipopolysaccharide-responsive phosphoproteins in *Nicotiana tabacum* cells. *Plant Physiol Biochem* 44(5–6):369–379. <https://doi.org/10.1016/j.plaphy.2006.06.015>
 53. Barsan C, Sanchez-Bel P, Rombaldi C, Egea I, Rossignol M, Kuntz M, Zouine M, Latch   A, Bouzayen M, Pech J-C (2010) Characteristics of the tomato chromoplast revealed by proteomic analysis. *J Exp Bot* 61(9):2413–2431. <https://doi.org/10.1093/jxb/erq070>
 54. Agrawal G, Thelen J (2009) A high-resolution two dimensional gel- and Pro-Q DPS-based proteomics workflow for phosphoprotein identification and quantitative profiling. In: Graauw M (ed) *Phospho-proteomics*, vol 527. *Methods in Molecular Biology*TM. Humana Press, Totowa, pp 3–19. https://doi.org/10.1007/978-1-60327-834-8_1
 55. Agrawal GK, Thelen JJ (2006) Large scale identification and quantitative profiling of phosphoproteins expressed during seed filling in oilseed rape. *Mol Cell Proteomics* 5(11):2044–2059. <https://doi.org/10.1074/mcp.M600084-MCP200>
 56. Chitteti BR, Peng Z (2007) Proteome and phosphoproteome dynamic change during cell dedifferentiation in *Arabidopsis*. *Proteomics* 7(9):1473–1500. <https://doi.org/10.1002/pmic.200600871>
 57. Steinberg TH, Agnew BJ, Gee KR, Leung WY, Goodman T, Schulenberg B, Hendrickson J, Beechem JM, Haugland RP, Patton WF (2003) Global quantitative phosphoprotein analysis using multiplexed proteomics technology. *Proteomics* 3(7):1128–1144. <https://doi.org/10.1002/pmic.200300434>
 58. Niittyl   T, Fuglsang AT, Palmgren MG, Frommer WB, Schulze WX (2007) Temporal analysis of sucrose-induced phosphorylation changes in plasma membrane proteins of *Arabidopsis*. *Mol Cell Proteomics* 6(10):1711–1726. <https://doi.org/10.1074/mcp.M700164-MCP200>
 59. Stulemeijer IJE, Joosten MHAJ, Jensen ON (2009) Quantitative phosphoproteomics of tomato mounting a hypersensitive response reveals a swift suppression of photosynthetic activity and a differential role for Hsp90 isoforms. *J Proteome Res* 8(3):1168–1182. <https://doi.org/10.1021/pr800619h>
 60. Kito K, Ito T (2008) Mass spectrometry-based approaches toward absolute quantitative proteomics. *Curr Genomics* 9(4):263–274. <https://doi.org/10.2174/138920208784533647>
 61. Chahrouh O, Cobice D, Malone J (2015) Stable isotope labeling methods in mass spectrometry-based quantitative proteomics. *J Pharm Biomed Anal* 113:2–20. <https://doi.org/10.1016/j.jpba.2015.04.013>
 62. Kieffer P, Domm  s J, Hoffmann L, Hausman JF, Renaut J (2008) Quantitative changes in protein expression of cadmium-exposed poplar plants. *Proteomics* 8(12):2514–2530. <https://doi.org/10.1002/pmic.200701110>
 63. Zhang X, Asara JM, Adamec J, Ouzzani M, Elmagarmid AK (2005) Data pre-processing in liquid chromatography–mass spectrometry-based proteomics. *Bioinformatics* 21(21):4054–4059. <https://doi.org/10.1093/bioinformatics/bti660>
 64. Palmblad M, Mills DJ, Bindschedler LV, Cramer R (2007) Chromatographic alignment of LC-MS and LC-MS/MS datasets by genetic algorithm feature extraction. *J Am Soc Mass Spectrom* 18(10):1835–1843. <https://doi.org/10.1016/j.jasms.2007.07.018>
 65. Reiland S, Messerli G, Baerenfaller K, Gerrits B, Endler A, Grossmann J, Gruissem W, Baginsky S (2009) Large-scale Arabidopsis phosphoproteome profiling reveals novel vhl-roplastkinase substrates and phosphorylation networks. *Plant Physiol* 150(2):889–903. <https://doi.org/10.1104/pp.109.138677>
 66. Lilley KS, Dupree P (2007) Plant organelle proteomics. *Curr Opin Plant Biol* 10(6):594–599. <https://doi.org/10.1016/j.pbi.2007.08.006>
 67. Abdallah C, Dumas-Gaudot E, Renaut J, Sergeant K (2012) Gel-based and gel-free quantitative proteomics approaches at a glance. *Int J Plant Genomics* 2012:494572. <https://doi.org/10.1155/2012/494572>
 68. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevensky JR, Resing KA, Ahn NG (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* 4(10):1487–1502. <https://doi.org/10.1074/mcp.M500084-MCP200>
 69. Friso G, Majeran W, Huang M, Sun Q, van Wijk KJ (2010) Reconstruction of metabolic pathways, protein expression, and homeostasis machineries across maize bundle sheath and mesophyll chloroplasts: large-scale quantitative proteomics using the first maize genome assembly. *Plant Physiol* 152(3):1219–1250. <https://doi.org/10.1104/pp.109.152694>
 70. Gammulla CG, Pascovici D, Atwell BJ, Haynes PA (2010) Differential metabolic response of cultured rice (*Oryza sativa*) cells exposed to high- and low-temperature stress. *Proteomics* 10(16):3001–3019. <https://doi.org/10.1002/pmic.201000054>
 71. Lee J, Feng J, Campbell KB, Scheffler BE, Garrett WM, Thiberville S, Stacey G, Naiman DQ, Tucker ML, Pastor-Corrales MA, Cooper B (2009) Quantitative proteomic analysis of bean plants infected by a virulent and avirulent obligate rust fungus. *Mol Cell Proteomics* 8(1):19–31. <https://doi.org/10.1074/mcp.M800156-MCP200>
 72. Zybailov B, Friso G, Kim J, Rudella A, Rodr  guez VR, Asakura Y, Sun Q, van Wijk KJ (2009) Large scale comparative proteomics of a chloroplast Clp protease mutant reveals folding stress, altered protein homeostasis, and feedback regulation of metabolism. *Mol Cell Proteomics* 8(8):1789–1810. <https://doi.org/10.1074/mcp.M900104-MCP200>
 73. Benschop JJ, Mohammed S, O’Flaherty M, Heck AJR, Slijper M, Menke FLH (2007) Quantitative phosphoproteomics of early elicitor signaling in *Arabidopsis*. *Mol Cell Proteomics* 6(7):1198–1214. <https://doi.org/10.1074/mcp.M600429-MCP200>
 74. Jones AME, Bennett MH, Mansfield JW, Grant M (2006) Analysis of the defence phosphoproteome of *Arabidopsis thaliana* using differential mass tagging. *Proteomics* 6(14):4155–4165. <https://doi.org/10.1002/pmic.200500172>
 75. Nuhse TS, Bottrill AR, Jones AM, Peck SC (2007) Quantitative phosphoproteomic analysis of plasma membrane proteins reveals

- regulatory mechanisms of plant innate immune responses. *Plant J* 51(5):931–940. <https://doi.org/10.1111/j.1365-313X.2007.03192.x>
76. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17(10):994–999. <https://doi.org/10.1038/13690>
 77. Majeran W, Cai Y, Sun Q, van Wijk KJ (2005) Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics. *Plant Cell* 17(11):3111–3140. <https://doi.org/10.1105/tpc.105.035519>
 78. Dunkley TPJ, Dupree P, Watson RB, Lilley KS (2004) The use of isotope-coded affinity tags (ICAT) to study organelle proteomes in *Arabidopsis thaliana*. *Biochem Soc Trans* 32:520–523
 79. Hagglund P, Bunkenborg J, Yang F, Harder LM, Finnie C, Svensson B (2010) Identification of thioredoxin target disulfides in proteins released from barley aleurone layers. *J Proteomics* 73(6):1133–1136. <https://doi.org/10.1016/j.jprot.2010.01.007>
 80. Miles GP, Samuel MA, Ranish JA, Donohoe SM, Sperrazzo GM, Ellis BE (2009) Quantitative proteomics identifies oxidant-induced, AtMPK6-dependent changes in *Arabidopsis thaliana* protein profiles. *Plant Signal Behav* 4(6):497–505. <https://doi.org/10.4161/psb.4.6.8538>
 81. Ow SY, Salim M, Noirel J, Evans C, Rehman I, Wright PC (2009) iTRAQ underestimation in simple and complex mixtures: “the good, the bad and the ugly”. *J Proteome Res* 8(11):5347–5355. <https://doi.org/10.1021/pr900634c>
 82. Perkel JM (2009) iTRAQ gets put to the test. *J Proteome Res* 8(11):4885–4885. <https://doi.org/10.1021/pr900897d>
 83. Zhang L, Elias JE (2017) Relative protein quantification using tandem mass tag mass spectrometry. In: Comai L, Katz JE, Mallick P (eds) *Proteomics: methods and protocols*. Springer, New York, pp 185–198. https://doi.org/10.1007/978-1-4939-6747-6_14
 84. Melo-Braga MN, Verano-Braga T, Leon IR, Antonacci D, Nogueira FC, Thelen JJ, Larsen MR, Palmisano G (2012) Modulation of protein phosphorylation, N-glycosylation and Lys-acetylation in grape (*Vitis vinifera*) mesocarp and exocarp owing to *Lobesia botrana* infection. *Mol Cell Proteomics* 11(10):945–956. <https://doi.org/10.1074/mcp.M112.020214>
 85. Marsh E, Alvarez S, Hicks LM, Barbazuk WB, Qiu W, Kovacs L, Schachtman D (2010) Changes in protein abundance during powdery mildew infection of leaf tissues of *Cabernet Sauvignon* grapevine (*Vitis vinifera* L.). *Proteomics* 10(10):2057–2064. <https://doi.org/10.1002/pmic.200900712>
 86. Fan J, Chen C, Yu Q, Brlansky RH, Li Z-G, Gmitter FG (2011) Comparative iTRAQ proteome and transcriptome analyses of sweet orange infected by “*Candidatus Liberibacter asiaticus*”. *Physiol Plant* 143(3):235–245. <https://doi.org/10.1111/j.1399-3054.2011.01502.x>
 87. Mohammadi M, Anoop V, Gleddie S, Harris LJ (2011) Proteomic profiling of two maize inbreds during early gibberella ear rot infection. *Proteomics* 11(18):3675–3684. <https://doi.org/10.1002/pmic.201100177>
 88. Kaffarnik FAR, Jones AME, Rathjen JP, Peck SC (2009) Effector proteins of the bacterial pathogen *Pseudomonas syringae* alter the extracellular proteome of the host plant, *Arabidopsis thaliana*. *Mol Cell Proteomics* 8(1):145–156. <https://doi.org/10.1074/mcp.M800043-MCP200>
 89. Zhao Z, Stanley BA, Zhang W, Assmann SM (2010) ABA-regulated G protein signaling in *Arabidopsis* guard cells: a proteomic perspective. *J Proteome Res* 9(4):1637–1647. <https://doi.org/10.1021/pr901011h>
 90. Lucker J, Laszczak M, Smith D, Lund S (2009) Generation of a predicted protein database from EST data and application to iTRAQ analyses in grape (*Vitis vinifera* cv. Cabernet Sauvignon) berries at ripening initiation. *BMC Genomics* 10(1):50
 91. Liu X, Dekker LJ, Wu S, Vanduijn MM, Luider TM, Tolic N, Kou Q, Dvorkin M, Alexandrova S, Vyatkina K, Pasa-Tolic L, Pevzner PA (2014) De novo protein sequencing by combining top-down and bottom-up tandem mass spectra. *J Proteome Res* 13(7):3241–3248. <https://doi.org/10.1021/pr401300m>
 92. Kumaravel M, Uma S, Backiyarani S, Saraswathi MS, Vaganan MM, Muthusamy M, Sajith KP (2017) Differential proteome analysis during early somatic embryogenesis in *Musa* spp. AAA cv. Grand Naine. *Plant Cell Rep* 36(1):163–178. <https://doi.org/10.1007/s00299-016-2067-y>
 93. Sghaier-Hammami B, Drira N, Jorin-Novo JV (2009) Comparative 2-DE proteomic analysis of date palm (*Phoenix dactylifera* L.) somatic and zygotic embryos. *J Proteomics* 73(1):161–177. <https://doi.org/10.1016/j.jprot.2009.07.003>
 94. Rahman MA, Ren L, Wu W, Yan Y (2015) Proteomic analysis of PEG-induced drought stress responsive protein in TERF1 overexpressed sugarcane (*Saccharum officinarum*) Leaves. *Plant Mol Biol Rep* 33(3):716–730. <https://doi.org/10.1007/s11105-014-0784-3>
 95. Chen X, Zhang W, Zhang B, Zhou J, Wang Y, Yang Q, Ke Y, He H (2011) Phosphoproteins regulated by heat stress in rice leaves. *Proteome Sci* 9(1):37. <https://doi.org/10.1186/1477-5956-9-37>
 96. Luís IM, Alexandre BM, Oliveira MM, Abreu IA (2016) Selection of an appropriate protein extraction method to study the phosphoproteome of maize photosynthetic tissue. *PLoS ONE* 11(10):e0164387. <https://doi.org/10.1371/journal.pone.0164387>
 97. Lau BYC (2015) Proteomic profiling of fatty acid biosynthetic enzymes from oil palm chromoplast. Lincoln University, Lincoln
 98. Cui D, Wu D, Liu J, Li D, Xu C, Li S, Li P, Zhang H, Liu X, Jiang C, Wang L, Chen T, Chen H, Zhao L (2015) Proteomic analysis of seedling roots of two maize inbred lines that differ significantly in the salt stress response. *PLoS ONE* 10(2):e0116697. <https://doi.org/10.1371/journal.pone.0116697>
 99. Ji W, Cong R, Li S, Li R, Qin Z, Li Y, Zhou X, Chen S, Li J (2016) Comparative proteomic analysis of soybean leaves and roots by iTRAQ provides insights into response mechanisms to short-term salt stress. *Front Plant Sci* 7:573. <https://doi.org/10.3389/fpls.2016.00573>
 100. Lin Q, Zhou Z, Luo W, Fang M, Li M, Li H (2017) Screening of proximal and interacting proteins in rice protoplasts by proximity-dependent biotinylation. *Front Plant Sci* 8:749. <https://doi.org/10.3389/fpls.2017.00749>
 101. Wang S, Chen W, Yang C, Yao J, Xiao W, Xin Y, Qiu J, Hu W, Yao H, Ying W, Fu Y, Tong J, Chen Z, Ruan S, Ma H (2016) Comparative proteomic analysis reveals alterations in development and photosynthesis-related proteins in diploid and triploid rice. *BMC Plant Biol* 16(1):199. <https://doi.org/10.1186/s12870-016-0891-4>
 102. Rattanakan S, George I, Haynes PA, Cramer GR (2016) Relative quantification of phosphoproteomic changes in grapevine (*Vitis vinifera* L.) leaves in response to abscisic acid. *Hortic Res* 3:16029. <https://doi.org/10.1038/hortres.2016.29>
 103. Khan MN, Komatsu S (2016) Proteomic analysis of soybean root including hypocotyl during recovery from drought stress. *J Proteomics* 144:39–50. <https://doi.org/10.1016/j.jprot.2016.06.006>
 104. Timabud T, Yin X, Pongdontri P, Komatsu S (2016) Gel-free/label-free proteomic analysis of developing rice grains under heat stress. *J Proteomics* 133:1–19. <https://doi.org/10.1016/j.jprot.2015.12.003>
 105. Wu Y, Mirzaei M, Pascovici D, Chick JM, Atwell BJ, Haynes PA (2016) Quantitative proteomic analysis of two different rice varieties reveals that drought tolerance is correlated with reduced abundance of photosynthetic machinery and increased

- abundance of ClpD1 protease. *J Proteomics* 143:73–82. <https://doi.org/10.1016/j.jprot.2016.05.014>
106. Heringer AS, Reis RS, Passamani LZ, de Souza-Filho GA, Santa-Catarina C, Silveira V (2017) Comparative proteomics analysis of the effect of combined red and blue lights on sugarcane somatic embryogenesis. *Acta Physiol Plant* 39(2):52. <https://doi.org/10.1007/s11738-017-2349-1>
 107. Martinez M (2016) Computational tools for genomic studies in plants. *Curr Genomics* 17(6):509–514. <https://doi.org/10.2174/1389202917666160520103447>
 108. International Brachypodium I (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463(7282):763–768. <https://doi.org/10.1038/nature08747>
 109. Rine J (2014) A future of the model organism model. *Mol Biol Cell* 25(5):549–553. <https://doi.org/10.1091/mbc.E12-10-0768>
 110. Uthapaisanwong P, Chanprasert J, Shearman JR, Sangsrakru D, Yoocha T, Jomchai N, Jantassuriyarat C, Tragoonrun S, Tangphatsornruang S (2012) Characterization of the chloroplast genome sequence of oil palm (*Elaeis guineensis* Jacq.). *Gene* 500(2):172–180. <https://doi.org/10.1016/j.gene.2012.03.061>
 111. Yang L, Luo Y, Wei J, Ren C, Zhou X, He S (2010) Methods for protein identification using expressed sequence tags and peptide mass fingerprinting for seed crops without complete genome sequences. *Seed Sci Res* 20(04):257–262. <https://doi.org/10.1017/S0960258510000243>
 112. Pedretti K, Scheetz T, Braun T, Roberts C, Robinson N, Casavant T (2001) A parallel expressed sequence tag (EST) clustering program. In: Malyskin V (ed) *Parallel computing technologies*, vol 2127. Lecture notes in computer science. Springer, Berlin, pp 490–497. https://doi.org/10.1007/3-540-44743-1_51
 113. Hoff KJ (2009) The effect of sequencing errors on metagenomic gene prediction. *BMC Genomics* 10(1):520. <https://doi.org/10.1186/1471-2164-10-520>
 114. Rowley A, Choudhary JS, Marzioch M, Ward MA, Weir M, Solari RCE, Blackstock WP (2000) Applications of protein mass spectrometry in cell biology. *Methods* 20(4):383–397. <https://doi.org/10.1006/meth.2000.0951>
 115. Gabaldon T (2007) Evolution of proteins and proteomes: a phylogenetics approach. *Evol Bioinform* 1(1):51–61. <https://doi.org/10.4137/ebo.s0>
 116. Fitch WM (1970) Distinguishing homologous from analogous proteins. *Syst Zool* 19(2):99–113. <https://doi.org/10.2307/2412448>
 117. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20(18):3551–3567. [https://doi.org/10.1002/\(sici\)1522-2683\(19991201\)20:18%3C3551::aid-elps3551%3E3.0.co;2-2](https://doi.org/10.1002/(sici)1522-2683(19991201)20:18%3C3551::aid-elps3551%3E3.0.co;2-2)
 118. Peltier J-B, Friso G, Kalume DE, Roepstorff P, Nilsson F, Adamska I, van Wijk KJ (2000) Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. *Plant Cell* 12(3):319–342. <https://doi.org/10.1105/tpc.12.3.319>
 119. Chang WWP, Huang L, Shen M, Webster C, Burlingame AL, Roberts JKM (2000) Patterns of protein synthesis and tolerance of anoxia in root tips of maize seedlings acclimated to a low-oxygen environment, and identification of proteins by mass spectrometry. *Plant Physiol* 122(2):295–318. <https://doi.org/10.1104/pp.122.2.295>
 120. Medzihradsky KF, Chalkley RJ (2015) Lessons in de novo peptide sequencing by tandem mass spectrometry. *Mass Spectrom Rev* 34(1):43–63
 121. Shevchenko A, Sunyaev S, Loboda A, Bork P, Ens W, Standing KG (2001) Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal Chem* 73(9):1917–1926. <https://doi.org/10.1021/ac0013709>
 122. Heazlewood JL, Verboom RE, Tonti-Filippini J, Small I, Millar AH (2007) SUBA: the *Arabidopsis* subcellular database. *Nucleic Acids Res* 35(Database issue):D213–D218. <https://doi.org/10.1093/nar/gkl863>
 123. Hooper CM, Castleden IR, Tanz SK, Aryamanesh N, Millar AH (2017) SUBA4: the interactive data analysis centre for *Arabidopsis* subcellular protein locations. *Nucleic Acids Res* 45(D1):D1064–D1074. <https://doi.org/10.1093/nar/gkw1041>
 124. Negi S, Pandey S, Srinivasan SM, Mohammed A, Guda C (2015) LocSigDB: a database of protein localization signals. *Database*. <https://doi.org/10.1093/database/bav003>
 125. Sun Q, Zybailov B, Majeran W, Friso G, Olinares PDB, van Wijk KJ (2009) PPDB, the plant proteomics database at Cornell. *Nucleic Acids Res* 37(suppl 1):D969–D974. <https://doi.org/10.1093/nar/gkn654>
 126. Palagi PM, Walther D, Quadroni M, Catherinet S, Burgess J, Zimmermann-Ivol CG, Sanchez JC, Binz PA, Hochstrasser DF, Appel RD (2005) MSight: an image analysis software for liquid chromatography-mass spectrometry. *Proteomics* 5(9):2381–2384. <https://doi.org/10.1002/pmic.200401244>
 127. Pluskal T, Castillo S, Villar-Briones A, Orešič M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform* 11(1):395. <https://doi.org/10.1186/1471-2105-11-395>
 128. Sturm M, Kohlbacher O (2009) TOPPView: an open-source viewer for mass spectrometry data. *J Proteome Res* 8(7):3760–3763. <https://doi.org/10.1021/pr900171m>
 129. Tyanova S, Temu T, Cox J (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* 11:2301. <https://doi.org/10.1038/nprot.2016.136>
 130. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26(12):1367–1372. <https://doi.org/10.1038/nbt.1511>
 131. Tyanova S, Cox J (2018) Perseus: a bioinformatics platform for integrative analysis of proteomics data in cancer research. In: von Stechow L (ed) *Cancer systems biology: methods and protocols*. Springer, New York, pp 133–148. https://doi.org/10.1007/978-1-4939-7493-1_7
 132. Wenig P, Odermatt J (2010) OpenChrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data. *BMC Bioinform* 11:405–405. <https://doi.org/10.1186/1471-2105-11-405>
 133. Muth T, Weilnböck L, Rapp E, Huber CG, Martens L, Vaudel M, Barsnes H (2014) DeNovoGUI: an open source graphical user interface for de novo sequencing of tandem mass spectra. *J Proteome Res* 13(2):1143–1146. <https://doi.org/10.1021/pr4008078>
 134. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebner DC, MacCoss MJ (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26(7):966–968. <https://doi.org/10.1093/bioinformatics/btq054>
 135. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA (2007) The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 8(9):R183. <https://doi.org/10.1186/gb-2007-8-9-r183>
 136. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C (2017) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly

- accessible. *Nucleic Acids Res* 45(Database issue):D362–D368. <https://doi.org/10.1093/nar/gkw937>
137. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C (2015) STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43(Database issue):D447–D452. <https://doi.org/10.1093/nar/gku1003>
 138. Croft D, O’Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, Jupe S, Kalatskaya I, Mahajan S, May B, Ndegwa N, Schmidt E, Sharmovsky V, Yung C, Birney E, Hermjakob H, D’Eustachio P, Stein L (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* 39(Database issue):D691–D697. <https://doi.org/10.1093/nar/gkq1018>
 139. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, Milacic M, Roca CD, Rothfels K, Sevilla C, Sharmovsky V, Shorsler S, Varusai T, Viteri G, Weiser J, Wu G, Stein L, Hermjakob H, D’Eustachio P (2018) The reactome pathway knowledge base. *Nucleic Acids Res* 46(D1):D649–D655. <https://doi.org/10.1093/nar/gkx1132>
 140. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1):27–30
 141. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 27(1):29–34
 142. Kutmon M, van Iersel MP, Bohler A, Kelder T, Nunes N, Pico AR, Evelo CT (2015) PathVisio 3: an extendable pathway analysis toolbox. *PLoS Comput Biol* 11(2):e1004085. <https://doi.org/10.1371/journal.pcbi.1004085>
 143. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504. <https://doi.org/10.1101/gr.1239303>
 144. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44(D1):D279–D285. <https://doi.org/10.1093/nar/gkv1344>
 145. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztanyi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto Silvio CE, Wu CH, Xenarios I, Yeh L-S, Young S-Y, Mitchell AL (2017) InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res* 45(Database issue):D190–D199. <https://doi.org/10.1093/nar/gkw1107>
 146. Jungblut P, Thiede B, Zimny-Arndt U, Muller EC, Scheler C, Wittmann-Liebold B, Otto A (1996) Resolution power of two-dimensional electrophoresis and identification of proteins from gels. *Electrophoresis* 17(5):839–847. <https://doi.org/10.1002/elps.1150170505>
 147. Schluter H, Apweiler R, Holzthutter H-G, Jungblut P (2009) Finding one’s way in proteomics: a protein species nomenclature. *Chem Cent J* 3(1):11. <https://doi.org/10.1186/1752-153X-3-11>
 148. Bond AE, Row PE, Dudley E (2011) Post-translation modification of proteins; methodologies and applications in plant sciences. *Phytochemistry* 72(10):975–996. <https://doi.org/10.1016/j.phytochem.2011.01.029>
 149. Mann M, Jensen ON (2003) Proteomic analysis of post-translational modifications. *Nat Biotechnol* 21(3):255–261. <https://doi.org/10.1038/nbt0303-255>
 150. Endler A, Baginsky S (2011) Use of phosphoproteomics to study posttranslational protein modifications in Arabidopsis chloroplasts. In: Jarvis RP (ed) *Chloroplast research in Arabidopsis*, vol 775. *Methods in Molecular Biology*. Humana Press, Totowa, pp 283–296. https://doi.org/10.1007/978-1-61779-237-3_15
 151. Bienvenut WV, Espagne C, Martinez A, Majeran W, Valot B, Zivy M, Vallon O, Adam Z, Meinel T, Giglione C (2011) Dynamics of post-translational modifications and protein stability in the stroma of *Chlamydomonas reinhardtii* chloroplasts. *Proteomics* 11(9):1734–1750. <https://doi.org/10.1002/pmic.201000634>
 152. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferrieria S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Heiner D, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. *Science* 291(5507):1304–1351. <https://doi.org/10.1126/science.1058040>
 153. Adams JA (2001) Kinetic and catalytic mechanisms of protein kinases. *Chem Rev* 101(8):2271–2290. <https://doi.org/10.1021/cr000230w>

154. Johnson SA, Hunter T (2005) Kinomics: methods for deciphering the kinome. *Nat Methods* 2(1):17–25. <https://doi.org/10.1038/nmeth731>
155. Cohen P (2001) The role of protein phosphorylation in human health and disease. *Eur J Biochem* 268(19):5001–5010. <https://doi.org/10.1046/j.0014-2956.2001.02473.x>
156. Lim YP (2005) Mining the tumor phosphoproteome for cancer markers. *Clin Cancer Res* 11(9):3163–3169. <https://doi.org/10.1158/1078-0432.ccr-04-2243>
157. Reinders J, Sickmann A (2005) State-of-the-art in phosphoproteomics. *Proteomics* 5(16):4052–4061. <https://doi.org/10.1002/pmic.200401289>
158. Vissers JPC, Pons S, Hulin A, Tissier R, Berdeaux A, Connolly JB, Langridge JI, Geromanos SJ, Ghaleh B (2009) The use of proteome similarity for the qualitative and quantitative profiling of reperfused myocardium. *J Chromatogr B* 877(13):1317–1326. <https://doi.org/10.1016/j.jchromb.2008.10.024>
159. Emes MJ (2009) Oxidation of methionine residues: the missing link between stress and signalling responses in plants. *Biochem J* 422:e1–e2. <https://doi.org/10.1042/BJ20091063>
160. O'Donovan C, Apweiler R, Bairoch A (2001) The human proteomics initiative (HPI). *Trends Biotechnol* 19(5):178–181. [https://doi.org/10.1016/s0167-7799\(01\)01598-0](https://doi.org/10.1016/s0167-7799(01)01598-0)
161. Prabhakaran S, Lippens G, Steen H, Gunawardena J (2012) Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip Rev* 4(6):565–583. <https://doi.org/10.1002/wsbm.1185>
162. Jensen ON (2006) Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol* 7(6):391–403. <https://doi.org/10.1038/nrm1939>
163. Venne AS, Solari FA, Faden F, Paretto T, Dissmeyer N, Zahedi RP (2015) An improved workflow for quantitative N-terminal charge-based fractional diagonal chromatography (ChaFRADIC) to study proteolytic events in *Arabidopsis thaliana*. *Proteomics* 15(14):2458–2469. <https://doi.org/10.1002/pmic.201500014>
164. Kwon SJ, Choi EY, Choi YJ, Ahn JH, Park OK (2006) Proteomics studies of post-translational modifications in plants. *J Exp Bot* 57(7):1547–1551. <https://doi.org/10.1093/jxb/erj137>
165. Larsen MR, Trelle MB, Thingholm TE, Jensen ON (2006) Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques* 40:790–798. <https://doi.org/10.2144/000112201>
166. Seo J, Lee K-J (2004) Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *J Biochem Mol Biol* 37(1):35–44
167. Ytterberg AJ, Jensen ON (2010) Modification-specific proteomics in plant biology. *J Proteomics* 73(11):2249–2266. <https://doi.org/10.1016/j.jprot.2010.06.002>
168. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villén J, Li J, Cohn MA, Cantley LC, Gygi SP (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci USA* 101(33):12130–12135. <https://doi.org/10.1073/pnas.0404720101>
169. Asara JM, Christofk HR, Freemark LM, Cantley LC (2008) A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* 8(5):994–999. <https://doi.org/10.1002/pmic.200700426>
170. Lu B, Ruse C, Xu T, Park SK, Yates J III (2007) Automatic validation of phosphopeptide identifications from tandem mass spectra. *Anal Chem* 79(4):1301–1310. <https://doi.org/10.1021/ac061334v>
171. Graves JD, Krebs EG (1999) Protein phosphorylation and signal transduction. *Pharmacol Ther* 82(2–3):111–121. [https://doi.org/10.1016/s0163-7258\(98\)00056-4](https://doi.org/10.1016/s0163-7258(98)00056-4)
172. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci* 100(12):6940–6945. <https://doi.org/10.1073/pnas.0832254100>
173. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422(6928):198–207. <https://doi.org/10.1038/nature01511>
174. Bendt AK, Burkovski A, Schaffer S, Bott M, Farwick M, Hermann T (2003) Towards a phosphoproteome map of *Corynebacterium glutamicum*. *Proteomics* 3(8):1637–1646. <https://doi.org/10.1002/pmic.200300494>
175. Su HC, Hutchison CA 3rd, Giddings MC (2007) Mapping phosphoproteins in *Mycoplasma genitalium* and *Mycoplasma pneumoniae*. *BMC Microbiol* 7(1):63. <https://doi.org/10.1186/1471-2180-7-63>
176. O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250(10):4007–4021
177. Patton WF (2002) Detection technologies in proteome analysis. *J Chromatogr B* 771(1–2):3–31. [https://doi.org/10.1016/s1570-0232\(02\)00043-0](https://doi.org/10.1016/s1570-0232(02)00043-0)
178. Boudsocq M, Droillard M-J, Barbier-Brygoo H, Laurière C (2007) Different phosphorylation mechanisms are involved in the activation of sucrose non-fermenting 1 related protein kinases 2 by osmotic stresses and abscisic acid. *Plant Mol Biol* 63(4):491–503. <https://doi.org/10.1007/s11103-006-9103-1>
179. Kauffmann H, Bailey JE, Fussenegger M (2001) Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. *Proteomics* 1:194–199
180. Bockus L, Scofield RH (2009) Phosphoprotein detection on protein electroblot using a phosphate-specific fluorophore. In: Kurien BT, Scofield RH (eds) *Protein blotting and detection*, vol 536. *Methods in Molecular Biology*. Humana Press, Totowa, pp 385–393. https://doi.org/10.1007/978-1-59745-542-8_39
181. Schulenberg B, Aggeler R, Beechem JM, Capaldi RA, Patton WF (2003) Analysis of steady-state protein phosphorylation in mitochondria using a novel fluorescent phosphosensor dye. *J Biol Chem* 278(29):27251–27255. <https://doi.org/10.1074/jbc.C300189200>
182. Nakanishi T, Ando E, Furuta M, Kinoshita E, Kinoshita-Kikuta E, Koike T, Tsunasawa S, Nishimura O (2007) Identification on membrane and characterization of phosphoproteins using an alkoxide-bridged dinuclear metal complex as a phosphate-binding tag molecule. *J Biomol Tech* 18(5):278–286
183. Aebersold R, Goodlett DR (2001) Mass spectrometry in proteomics. *Chem Rev* 101(2):269–296. <https://doi.org/10.1021/cr990076h>
184. Mann M, Ong S-E, Grønborg M, Steen H, Jensen ON, Pandey A (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol* 20(6):261–268. [https://doi.org/10.1016/s0167-7799\(02\)01944-3](https://doi.org/10.1016/s0167-7799(02)01944-3)
185. Simpson RJ (2003) Proteomic methods for phosphorylation site mapping. In: Simpson RJ (ed) *Protein and proteomics. A laboratory manual*. Cold Spring Harbour, New York, pp 597–668
186. Dunn JD, Reid GE, Bruening ML (2010) Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry. *Mass Spectrom Rev* 29(1):29–54. <https://doi.org/10.1002/mas.20219>
187. Sun X, Chiu JF, He QY (2005) Application of immobilized metal affinity chromatography in proteomics. *Expert Rev Proteomics* 2(5):649–657. <https://doi.org/10.1586/14789450.2.5.649>
188. Thingholm TE, Jensen ON, Larsen MR (2009) Analytical strategies for phosphoproteomics. *Proteomics* 9(6):1451–1468. <https://doi.org/10.1002/pmic.200800454>
189. Grimsrud PA, den Os D, Wenger C, Swaney D, Schwartz DL, Sussman D, Ane MR, Coon J-M JJ (2010) Large-scale phosphorylation analysis in *Medicago truncatula* roots provides insight

- into in vivo kinase activity in legumes. *Plant Physiol* 152(1):19–28. <https://doi.org/10.1104/pp.109.149625>
190. Nuhse TS, Stensballe A, Jensen ON, Peck SC (2003) Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol Cell Proteomics* 2(11):1234–1243. <https://doi.org/10.1074/mcp.T300006-MCP200>
 191. Posewitz MC, Tempst P (1999) Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem* 71(14):2883–2892. <https://doi.org/10.1021/ac981409y>
 192. Dong J, Zhou H, Wu R, Ye M, Zou H (2007) Specific capture of phosphopeptides by Zr⁴⁺-modified monolithic capillary column. *J Sep Sci* 30(17):2917–2923. <https://doi.org/10.1002/jssc.20070.0350>
 193. Feng S, Ye M, Zhou H, Jiang X, Jiang X, Zou H, Gong B (2007) Immobilized zirconium ion affinity chromatography for specific enrichment of phosphopeptides in phosphoproteome analysis. *Mol Cell Proteomics* 6(9):1656–1665. <https://doi.org/10.1074/mcp.T600071-MCP200>
 194. Wei J, Zhang Y, Wang J, Tan F, Liu J, Cai Y, Qian X (2008) Highly efficient enrichment of phosphopeptides by magnetic nanoparticles coated with zirconium phosphonate for phosphoproteome analysis. *Rapid Commun Mass Spectrom* 22(7):1069–1080. <https://doi.org/10.1002/rcm.3485>
 195. Yu L-R, Zhu Z, Chan KC, Issaq HJ, Dimitrov DS, Veenstra TD (2007) Improved titanium dioxide enrichment of phosphopeptides from HeLa cells and high confident phosphopeptide identification by cross-validation of MS/MS and MS/MS/MS spectra. *J Proteome Res* 6(11):4150–4162. <https://doi.org/10.1021/pr070152u>
 196. Zhou H, Xu S, Ye M, Feng S, Pan C, Jiang X, Li X, Han G, Fu Y, Zou H (2006) Zirconium phosphonate-modified porous silicon for highly specific capture of phosphopeptides and MALDI-TOF MS analysis. *J Proteome Res* 5(9):2431–2437. <https://doi.org/10.1021/pr060162f>
 197. Heintz D, Wurtz V, High AA, Van Dorsselaer A, Reski R, Sarnighausen E (2004) An efficient protocol for the identification of protein phosphorylation in a seedless plant, sensitive enough to detect members of signalling cascades. *Electrophoresis* 25(7–8):1149–1159. <https://doi.org/10.1002/elps.200305795>
 198. Schmidt A, Csaszar E, Ammerer G, Mechtler K (2008) Enhanced detection and identification of multiply phosphorylated peptides using TiO₂ enrichment in combination with MALDI TOF/TOF MS. *Proteomics* 8(21):4577–4592. <https://doi.org/10.1002/pmic.200800279>
 199. Hsu J-L, Wang L-Y, Wang S-Y, Lin C-H, Ho K-C, Shi F-K, Chang I-F (2009) Functional phosphoproteomic profiling of phosphorylation sites in membrane fractions of salt-stressed *Arabidopsis thaliana*. *Proteome Sci* 7(1):42. <https://doi.org/10.1186/1477-5956-7-42>
 200. Nawrocki J, Dunlap C, McCormick A, Carr PW (2004) Part I. Chromatography using ultra-stable metal oxide-based stationary phases for HPLC. *J Chromatogr A* 1028(1):1–30. <https://doi.org/10.1016/j.chroma.2003.11.052>
 201. Ikeguchi Y, Nakamura H (1997) Determination of organic phosphates by column-switching high performance anion-exchange chromatography using on-line preconcentration on titania. *Anal Sci* 13(3):479–483. <https://doi.org/10.2116/analsci.13.479>
 202. Ikeguchi Y, Nakamura H (2000) Selective enrichment of phospholipids by titania. *Anal Sci* 16(5):541–543. <https://doi.org/10.2116/analsci.16.541>
 203. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jørgensen TJD (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* 4(7):873–886. <https://doi.org/10.1074/mcp.T500007-MCP200>
 204. Pinkse MWH, Uitto PM, Hilhorst MJ, Ooms B, Heck AJR (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-nanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* 76(14):3935–3943. <https://doi.org/10.1021/ac0498617>
 205. Sano A, Nakamura H (2004) Chemo-affinity of titania for the column-switching HPLC analysis of phosphopeptides. *Anal Sci* 20(3):565–566. <https://doi.org/10.2116/analsci.20.565>
 206. Sano A, Nakamura H (2004) Titania as a chemo-affinity support for the column-switching HPLC analysis of phosphopeptides: application to the characterization of phosphorylation sites in proteins by combination with protease digestion and electrospray ionization mass spectrometry. *Anal Sci* 20(5):861–864. <https://doi.org/10.2116/analsci.20.861>
 207. Kosmulski M (2002) The significance of the difference in the point of zero charge between rutile and anatase. *Adv Colloid Interface Sci* 99(3):255–264. [https://doi.org/10.1016/s0001-8686\(02\)00080-5](https://doi.org/10.1016/s0001-8686(02)00080-5)
 208. Thingholm TE, Larsen MR, Ingrell CR, Kassem M, Jensen ON (2008) TiO₂-based phosphoproteomic analysis of the plasma membrane and the effects of phosphatase inhibitor treatment. *J Proteome Res* 7(8):3304–3313. <https://doi.org/10.1021/pr80099y>
 209. Olsen JV, Blagoev B, Gnäd F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127(3):635–648. <https://doi.org/10.1016/j.cell.2006.09.026>
 210. McNulty DE, Annan RS (2008) Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol Cell Proteomics* 7(5):971–980. <https://doi.org/10.1074/mcp.M700543-MCP200>
 211. Moon JH, Shin YS, Kim MS (2009) Utility of reaction intermediate monitoring with photodissociation multi-stage (MSn) time-of-flight mass spectrometry for mechanistic and structural studies: phosphopeptides. *Int J Mass Spectrom* 288(1–3):16–21. <https://doi.org/10.1016/j.ijms.2009.07.008>
 212. DeGnove J, Qin J (1998) Fragmentation of phosphopeptides in an ion trap mass spectrometer. *J Am Soc Mass Spectrom* 9(11):1175–1188. [https://doi.org/10.1016/s1044-0305\(98\)00088-9](https://doi.org/10.1016/s1044-0305(98)00088-9)
 213. Boersem PJ, Mohammed S, Heck AJ (2009) Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom* 44(6):861–878. <https://doi.org/10.1002/jms.1599>
 214. Villen J, Beausoleil SA, Gygi SP (2008) Evaluation of the utility of neutral-loss-dependent MS3 strategies in large-scale phosphorylation analysis. *Proteomics* 8(21):4444–4452. <https://doi.org/10.1002/pmic.200800283>
 215. Yocum AK, Chinnaiyan AM (2009) Current affairs in quantitative targeted proteomics: multiple reaction monitoring—mass spectrometry. *Brief Funct Genomics Proteomics* 8(2):145–157. <https://doi.org/10.1093/bfgp/eln056>
 216. Cox DM, Zhong F, Du M, Duchoslav E, Sakuma T, McDermott JC (2005) Multiple reaction monitoring as a method for identifying protein posttranslational modifications. *J Biomol Tech* 16(2):83–90
 217. Domanski D, Murphy LC, Borchers CH (2010) Assay development for the determination of phosphorylation stoichiometry using multiple reaction monitoring methods with and without phosphatase treatment: application to breast cancer signaling pathways. *Anal Chem* 82(13):5610–5620. <https://doi.org/10.1021/ac1005553>
 218. Fan J, Mohareb F, Jones AM, Bessant C (2012) MRmaid: the SRM assay design tool for *Arabidopsis* and other species. *Front Plant Sci* 3:164. <https://doi.org/10.3389/fpls.2012.00164>

219. Lange V, Picotti P, Domon B, Aebersold R (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol*. <https://doi.org/10.1038/msb.2008.61>
220. Martinez-Marquez A, Morante-Cariel J, Selles-Marchart S, Martinez-Esteso MJ, Pineda-Lucas JL, Luque I, Bru-Martinez R (2013) Development and validation of MRM methods to quantify protein isoforms of polyphenol oxidase in loquat fruits. *J Proteome Res* 12(12):5709–5722. <https://doi.org/10.1021/pr4006712>
221. Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, White FM (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc Natl Acad Sci* 104(14):5860–5865. <https://doi.org/10.1073/pnas.0608638104>
222. Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4(6):1633–1649. <https://doi.org/10.1002/pmic.200300771>
223. Xue Y, Zhou F, Zhu M, Ahmed K, Chen G, Yao X (2005) GPS: a comprehensive www server for phosphorylation sites prediction. *Nucleic Acids Res* 33(suppl 2):W184–W187. <https://doi.org/10.1093/nar/gki393>
224. Zanzoni A, Carbajo D, Diella F, Gherardini PF, Tramontano A, Helmer-Citterich M, Via A (2011) Phospho3D 2.0: an enhanced database of three-dimensional structures of phosphorylation sites. *Nucleic Acids Res* 39(suppl 1):D268–D271. <https://doi.org/10.1093/nar/gkq936>
225. Yoo P, Ho Y, Zhou B, Zomaya A (2008) SiteSeek: post-translational modification analysis using adaptive locality-effective kernel methods and new profiles. *BMC Bioinform* 9(1):272
226. Basu S, Plewczynski D (2010) AMS 3.0: prediction of post-translational modifications. *BMC Bioinform* 11(1):210
227. Sigrist CJA, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P (2002) PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinform* 3(3):265–274. <https://doi.org/10.1093/bib/3.3.265>
228. Yaffe MB, Leparo GG, Lai J, Obata T, Volinia S, Cantley LC (2001) A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat Biotechnol* 19(4):348–353. <https://doi.org/10.1038/86737>
229. Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 294(5):1351–1362. <https://doi.org/10.1006/jmbi.1999.3310>
230. Hjerrild M, Stensballe A, Rasmussen TE, Kofoed CB, Blom N, Sicheritz-Ponten T, Larsen MR, Brunak S, Jensen ON, Gammeltoft S (2004) Identification of phosphorylation sites in protein kinase A substrates using artificial neural networks and mass spectrometry. *J Proteome Res* 3(3):426–433. <https://doi.org/10.1021/pr0341033>
231. Zhou S, Shoelson SE, Chaudhuri M, Gish G, Pawson T, Haser WG, King F, Roberts T, Ratnofsky S, Lechleider RJ, Neel BG, Birge RB, Fajardo JE, Chou MM, Hanafusa H, Schaffhausen B, Cantley LC (1993) SH2 domains recognize specific phosphopeptide sequences. *Cell* 72(5):767–778
232. Kim JH, Lee J, Oh B, Kimm K, Koh I (2004) Prediction of phosphorylation sites using SVMs. *Bioinformatics* 20(17):3179–3184. <https://doi.org/10.1093/bioinformatics/bth382>
233. Plewczynski D, Tkacz A, Wyrwicz LS, Rychlewski L (2005) AutoMotif server: prediction of single residue post-translational modifications in proteins. *Bioinformatics* 21(10):2525–2527. <https://doi.org/10.1093/bioinformatics/bti333>
234. Wong Y-H, Lee T-Y, Liang H-K, Huang C-M, Wang T-Y, Yang Y-H, Chu C-H, Huang H-D, Ko M-T, Hwang J-K (2007) KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Res* 35(suppl 2):W588–W594. <https://doi.org/10.1093/nar/gkm322>
235. Gnad F, Gunawardena J, Mann M (2011) PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res* 39(suppl 1):D253–D260. <https://doi.org/10.1093/nar/gkq1159>
236. Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* 24(10):1285–1292. <https://doi.org/10.1038/nbt1240>
237. Ruttenberg BE, Pisitkun T, Knepper MA, Hoffert JD (2008) PhosphoScore: an open-source phosphorylation site assignment tool for MSn data. *J Proteome Res* 7(7):3054–3059. <https://doi.org/10.1021/pr800169k>
238. Chen Y, Chen W, Cobb MH, Zhao Y (2009) PTMap—a sequence alignment software for unrestricted, accurate, and full-spectrum identification of post-translational modification sites. *Proc Natl Acad Sci USA* 106(3):761–766. <https://doi.org/10.1073/pnas.0811739106>
239. Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* 38(Database issue):D161–D166. <https://doi.org/10.1093/nar/gkp885>
240. Obenaus JC, Cantley LC, Yaffe MB (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res* 31(13):3635–3641
241. Plewczynski D, Basu S, Saha I (2012) AMS 4.0: consensus prediction of post-translational modifications in protein sequences. *Amino Acids* 43(2):573–582. <https://doi.org/10.1007/s00726-012-1290-2>
242. Xue Y, Li A, Wang L, Feng H, Yao X (2006) PPSp: prediction of PK-specific phosphorylation site with Bayesian decision theory. *BMC Bioinform* 7:163–163. <https://doi.org/10.1186/1471-2105-7-163>
243. Hamby SE, Hirst JD (2008) Prediction of glycosylation sites using random forests. *BMC Bioinform* 9:500–500. <https://doi.org/10.1186/1471-2105-9-500>
244. Chuang G-Y, Boyington JC, Joyce MG, Zhu J, Nabel GJ, Kwong PD, Georgiev I (2012) Computational prediction of N-linked glycosylation incorporating structural properties and patterns. *Bioinformatics* 28(17):2249–2255. <https://doi.org/10.1093/bioinformatics/bts426>
245. Li F, Li C, Wang M, Webb GI, Zhang Y, Whisstock JC, Song J (2015) GlycoMine: a machine learning-based approach for predicting N-, C- and O-linked glycosylation in the human proteome. *Bioinformatics* 31(9):1411–1419. <https://doi.org/10.1093/bioinformatics/btu852>
246. Chauhan JS, Bhat AH, Raghava GPS, Rao A (2012) GlycoPP: a webserver for prediction of N- and O-glycosites in prokaryotic protein sequences. *PLoS ONE* 7(7):e40155. <https://doi.org/10.1371/journal.pone.0040155>
247. Xue Y, Liu Z, Gao X, Jin C, Wen L, Yao X, Ren J (2010) GPS-SNO: computational prediction of protein S-nitrosylation sites with a modified GPS algorithm. *PLoS ONE* 5(6):e11290. <https://doi.org/10.1371/journal.pone.0011290>
248. Wen P-P, Shi S-P, Xu H-D, Wang L-N, Qiu J-D (2016) Accurate in silico prediction of species-specific methylation sites based on information gain feature optimization. *Bioinformatics* 32(20):3107–3115. <https://doi.org/10.1093/bioinformatics/btw377>
249. Lee TY, Hsu JB, Lin FM, Chang WC, Hsu PC, Huang HD (2010) N-Ace: using solvent accessibility and physicochemical properties to identify protein N-acetylation sites. *J Comput Chem* 31(15):2759–2771. <https://doi.org/10.1002/jcc.21569>
250. Hou T, Zheng G, Zhang P, Jia J, Li J, Xie L, Wei C, Li Y (2014) LAceP: lysine acetylation site prediction using logistic regression

- classifiers. *PLoS ONE* 9(2):e89575. <https://doi.org/10.1371/journal.pone.0089575>
251. Lee TY, Huang HD, Hung JH, Huang HY, Yang YS, Wang TH (2006) dbPTM: an information repository of protein post-translational modification. *Nucleic Acids Res* 34(Database issue):D622–D627. <https://doi.org/10.1093/nar/gkj083>
252. Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ, Diella F (2011) Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res* 39(suppl 1):D261–D267. <https://doi.org/10.1093/nar/gkq1104>
253. Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, Latham V, Sullivan M (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* 40(Database issue):D261–D270. <https://doi.org/10.1093/nar/gkr1122>