



Plant Proteomics and Systems Biology

3

Flavia Vischi Winck, André Luis Wendt dos Santos,
and Maria Juliana Calderan-Rodrigues

Abstract

Proteome analysis of model and non-model plants is a genuine scientific field in expansion. Several technological advances have contributed to the implementation of different proteomics approaches for qualitative and quantitative analysis of the dynamics of cellular responses at the protein level. The design of time-resolved experiments and the emergent use of multiplexed proteome analysis using chemical or isotopic and isobaric labeling strategies as well as label-free approaches are generating a vast amount of proteomics data that is going to be essential for analysis of protein posttranslational modifications and implementation of systems biology approaches. Through the target proteomics

analysis, especially the ones that combine the untargeted methods, we should expect an improvement in the completeness of the identification of proteome and reveal nuances of regulatory cellular mechanisms related to plant development and responses to environmental stresses. Both genomic sequencing and proteomic advancements in the last decades coupled to integrative data analysis are enriching biological information that was once confined to model plants. Therewith, predictions of a changing environment places proteomics as an especially useful tool for crops performance.

Keywords

Mass spectrometry · Protein · 2-DE · Quantitative · Expression

F. V. Winck (✉)

Institute of Chemistry, University of São Paulo, São Paulo, Brazil

Center for Nuclear Energy in Agriculture, University of São Paulo,
Piracicaba, Brazil
e-mail: winck@cena.usp.br

A. L. W. dos Santos

Laboratory of Plant Cellular Biology, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

M. J. Calderan-Rodrigues

Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany
e-mail: calderan@mpimp-golm.mpg.de

3.1 Introduction

Proteins are fundamental macromolecules that execute a vast number of biological functions in the organisms, ranging from structural activities to exceptionally precise regulatory roles inside the cells. These characteristics intrigued scientists for centuries and the protein sequences started to be detailed in the twentieth century, with the sequencing of the first protein, phenylalanyl chain of insulin 1, in 1951 by Sanger and

Tuppy (1951) using partial hydrolysis techniques. For the first time, they presented a structured amino acid sequence composition encoded by the DNA of the living organisms. Many other proteins were sequenced thereof using biochemical sequencing analysis established by Edman (1949). At the same pace, with the development of ionization methods used for mass spectrometry (MS), specially the electrospray ionization (Fenn et al. 1989), the analysis of intact or digested proteins increased significantly. Several biochemical and molecular properties of thousands of proteins were identified and are currently well known, with the precise description of their amino acid sequence, tridimensional structure, activities, and chemical modifications. The unique functional and molecular characteristic of each protein species is outstandingly provocative and has revealed the enormous complexity of the possible molecular interactions these molecules can perform in a biological system. This also revealed the need for systems analysis in plant species, focusing in quantifying the changes in abundance and modifications of a large number of proteins simultaneously, and if possible, with a cell or tissue spatial and time resolution.

Today, we understand that a single-protein sequence can execute single or multiple biological roles, depending on the chemical modifications these proteins present in their structure and their sub-cellular compartmentalization.

With the vast information about how cellular systems function, it has become imperative since the beginning of the 1990s the study of comprehensive amount of proteins simultaneously in a cell or tissue. The need for the identification of all proteins was urgent, and the term Proteome was, therefore, first mentioned in 1995 by Wilkins et al. (1996), indicating an age of dramatic change on biology scale which has begun with the availability of complete genomic sequences of many organisms. As stated by Wilkins, the proteome is the entire protein complement expressed by a genome. By nature, the proteome is dynamic, being representative of the whole-protein repertoire of a cell or tissue in a certain time and condition. For instance, the proteome is different every period of the lifetime of a plant, or in

response to any challenging environmental stress. The proteome also includes, a priori, all different chemical modifications proteins may present, which are usually present as posttranslational modifications (PTMs) of the synthesized proteins or may be introduced by dynamic processes of chemical modifications of amino acids during cellular signaling events. The alterations on the protein abundance and PTMs are especially important for defining the cell fate in response to changes in environmental conditions, and current proteome research addresses many of these alterations using highly-sensitive analytical tools and sophisticated computational data analysis. The proteome of different species has been explored using different techniques, ranging from the use of two-dimensional polyacrylamide gel electrophoresis (2-DE) and its modified form 2D DIGE (two-dimensional difference gel electrophoresis) (Jorri n-Novo et al. 2019; Martins de Souza et al. 2008), passing through isotope and isobaric labeling (Pappireddi et al. 2019), leading to the massive proteome analysis using shotgun approaches and stable isotope labeling or label-free relative quantification (de Godoy et al. 2008), which further was followed by target approaches (Rodiger and Baginsky 2018), MS imaging (Boughton et al. 2016; Kaspar et al. 2011) and, more recently, single-cell proteomics (Marx 2019).

In the current days, there is an understanding that the identification of all intricated connections of the proteins inside the cells must be revealed. However, to detect the proteome of a cell or to represent it in a systems analysis is not a trivial task, but many efforts have been done to address several technical challenges. The availability of large datasets of plant proteomes in different experimental and field conditions generated by the unprecedented analysis of the proteome of several plant species is opening an exceptional path to the discovery of previous unknown cellular mechanisms and molecular emergent patterns.

In the last 21 years, the analysis of plant or plant-related proteome has increased from four scientific papers a year in 1998 to 1483 scientific papers a year in 2019 according to the PubMed

repository (<https://pubmed.ncbi.nlm.nih.gov/>) queried through the keywords “plant proteomics.” During this period of time, several technical advances have been established for the analysis of plant proteomics, including the optimization of techniques and instrumentation. The proteome analysis in general had begun basically through gel-based approaches, usually with Two-Dimensional Polyacrylamide gel Electrophoresis (2D-PAGE) coupled to peptide sequencing or Mass Spectrometry (MS) (Shevchenko et al. 1996a, b). Further approaches expanded to applications using liquid chromatography (LC) coupled to tandem MS, including shotgun analysis (Neubauer et al. 1998; Ong and Mann 2005), selected/parallel reaction monitoring (SRM/PRM) (Picotti and Aebersold 2012) and target data acquisition (TDA) (Schmidt et al. 2009). Recent applications suggest the combination of the TDA with tandem MS/MS analysis using Data-Dependent Acquisition (DDA) or Data-Independent Acquisition (DIA) as a valuable way to get hypothesis-driven and non-hypothesis-driven quantitative data collection from the same sample (Hart-Smith et al. 2017), even for the analysis of protein post-translational modifications (Pappireddi et al. 2019).

The analysis of the proteome for many different plant species had applied many of the approaches mentioned and increased our knowledge of plant molecular physiology and evolution in an enormous way. The level of details identified by proteomics analysis revealed a large proportion of the repertoire of cellular mechanisms and proteins that define the underlying principles of plant development and metabolism, including seed germination (He and Yang 2013), root growth (Li et al. 2019), stress responses (Kosova et al. 2014), senescence (Kim et al. 2016), light responses (Mettler et al. 2014), among others.

Plant proteome analysis can be performed using many different techniques and computational resources through diverse approaches. Each of these approaches will complement our knowledge of plant phenotypes and influence in the future directions of plant proteomics and systems biology. The information available today and further development of novel techniques for

proteome analysis will definitely pave the way to a more comprehensive understanding of the complex phenomena that take place in plants and how the numerous set of molecular interactions and mechanisms are built and retained in the cells in response to varying environmental and intracellular conditions.

3.2 Research and Technical Approaches

3.2.1 The Gel-Electrophoresis-Based Plant Proteome Analysis

At the end of the 1990s, beginning of 2000s, there was a modest racing into defining the so-called reference gel or reference proteome of an organism or organismal phenotype. Two-dimensional polyacrylamide gel electrophoresis (2-DE) has been used as one of the most powerful techniques of protein separation in a couple of technical steps (O’Farrell 1975). In 2-DE, the proteins contained in a biological sample are usually separated in a polyacrylamide gel by their isoelectric point (pI) and molecular weight (mW). Usually proteins are solubilized by chaotropic chemicals, such as urea and thiourea and solvated by non-ionic detergents. The first dimension of separation occurs with the migration of proteins in their intact form through an electric field, permeating an inert gel matrix soaked with amphoteric substances that form a transient gradient, allowing the proteins to move in these conditions toward the electrodes until they reach to their isoelectric point (pI), where they achieve their minimal solubility, thus high precipitation. In the second dimension of the 2-DE, the proteins separated and immobilized in the first gel dimension are separated by their molecular weight through SDS-PAGE. This will result in a two-dimensional plan with y and x axis, with proteins separated, respectively, by molecular mass (MW) and isoelectric point (pI) visualized as separated gel protein “spots” that can be visualized through dyeing techniques using, for instance, Coomassie Blue or Silver nitrate chemicals (Shevchenko et al.

1996a, b). In an optimal situation, the results of the gel staining will reveal the whole set of proteins from the biological sample analyzed, represented by protein spots. Each spot may contain one or more proteins with the approximate same biochemical characteristics. In proteome approaches, each spot is cut from the gel and digested with specific enzymes (e.g., Trypsin, Lys-C) that are going to generate peptides which can be identified through sequencing or MS analysis. Using gel-based approaches, the identification of the repertoire of proteins from different parts of the plants and those involved in complex developmental processes was possible and contributed to elucidate how proteins interconnect to each other to define a systemic function. The initial efforts into the analysis of the proteome of several important plant species had been done using gel-based approaches, and revealed important aspects of the plant structure and molecular physiology, including the analysis of green and etiolated shoots of rice (Komatsu et al. 1999), *Arabidopsis* seed germination and priming (Gallardo et al. 2001), maize leaves (Porubleva et al. 2001), Medicago mycorrhizal roots (Bestel-Corre et al. 2002), among others.

However, the gel-based proteomics analysis has several limitations and are difficult to be implemented in a large-scale proteomics analysis. Two-dimensional gels are quite laborious and time-consuming to be performed, which renders extra difficulty levels to other limitations in 2-DE gels that include problems with protein spot resolution, gel reproducibility, and low throughput. Usually, in a 2-DE gel, it is possible to separate approximately one thousand proteins simultaneously, but, due to the possible overlap between protein spots with similar mW and pI, the protein separation of some protein spots is not complete.

Nowadays, still, 2-DE is considered a complementary approach for some applications such as subcellular proteomics, analysis of low-complexity samples, and study of protein isoforms and their modifications (Jorin-Novo et al. 2019). The approaches in subcellular proteomics are not going to be addressed here since this topic is presented in a specific chapter of this book.

The 2-DE-based proteomics coupled to protein sequencing or MS techniques has contributed greatly to the development of plant proteomic analysis since the 1990s and has continuously adding important knowledge.

The diversity of 2-DE-based proteomic applications is vast, including analysis of plant structure (Giavalisco et al. 2005), plant–pathogen interactions and molecular signaling (Delaunois et al. 2014), and plant metabolism (Chang et al. 2017), to mention a few examples. Many of the proteomics results (including 2-DE-gel image) were also made available through public repositories, such as GABI Primary Database of Arabidopsis (URL: <https://www.gabipd.org/>).

One important aspect of the use 2-DE-based comparative proteomics resides in the fact that this approach can be implemented for any plant species, no matter if a genome sequence is available to this species. With that in mind, it is of great importance that comparisons between proteome profiles of non-model plants are investigated, elucidating the differences in the cellular responses of the many plant species that are of regionally importance to local agriculture and may not have their genome sequenced in the near future. The 2-DE-based proteomics may be a way to sustain this endeavor.

3.2.2 Mass Spectrometry-Based Proteomics

Among the methods that succeed in answering the increased demand for high-throughput proteomics approaches, the MS is currently the most applied and developed technique. Mass spectrometry is an analytical technique that is applied in many different fields of applications, including elemental analysis, organic and bio-organic analysis, structure elucidation, characterization of ionic species and reactions, and spectral imaging. In all these applications, the use of MS is aiming to identify a compound from the molecular or atomic mass(es) of its constituents (Gross 2017). So, an important aspect of proteome analysis resides in the fact that the chemical species, in

this case proteins or peptides, must be ionized to be analyzed through a mass spectrometer.

From the first attempts to understand electric discharges in gases and charged ions made by Joseph John Thomson with his first instrument that separated ions by mass-to-charge ratio to the most recent mass spectrometers that use high-resolution mass analyzers, there was a huge improvement and expansion of the applications of MS in Biochemistry and Biological Sciences (Gross 2017).

The mass spectrometry is the analytical technique that significantly generated our current substantial amount of plant proteomics data available. The development and evolution of ionization methods for proteins and peptides had contributed drastically to the change in proteomics scale, allowing the easy ionization of peptides and proteins into gas phase. These technical developments were complemented by the co-evolution of the sample preparation for proteomics, types of ionization techniques, and mass analyzers integrated into mass spectrometers, which significantly enhanced proteome coverage, the mass resolution, and throughput obtained for the mass-to-charge measurements. A recent review presented an overview about the protein extraction and preparation for proteomics analysis, discussing many approaches applicable for plant proteomics (Patole and Bindschedler 2019).

For qualitative and quantitative proteome analysis, two types of MS applications have been used more frequently to the identification of proteins: the matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) (Hillenkamp and Karas 1990), and liquid-chromatography (LC) coupled to electrospray ionization mass spectrometry (ESI-MS) (Smith et al. 1990). In the LC-MS analysis, the peptides obtained from the digestion of the total proteome are separated by liquid chromatography prior to the injection into the ion source of the MS instruments, where the peptides are ionized. After that the peptides are analyzed by the instruments resulting in the identification of their mass, and usually thousands of proteins can be identified in a shotgun MS analysis.

The implementation of high-throughput shotgun MS analysis in plant proteomics is usually dependent on the existence of a genome sequence or customized protein sequence databases derived from transcriptomic analyses (RNA-Seq data) for the plant species under investigation. Protein identification has evolved through the development of different computational tools and strategies that allowed the MS data to be analyzed by a series of steps that permit the identification of the best protein hits (or protein groups) in a statistical-based computational analysis. Protein identification is basically performed computationally by comparing the mass of the peptides analyzed (precursor ions) experimentally from the mass spectra data with the mass of the precursor ions generated *in silico* based on the translated genome of the plant species under study. Therefore, the protein identification in this case results from a probability analysis of the best protein hit or group of protein hits. The abundance of the peptide ions analyzed is determined, usually, by the arbitrary signal intensity measures for each peptide ion analyzed in a MS scan or by the counts of MS spectra for a given precursor ion, which gives the information useful for quantitative proteomics, using label-free or labeled approaches.

Recent analysis using sensitive and efficient mass spectrometers that have high-resolution mass analyzers has revealed the identity of thousands of proteins, increasing the capacity to observe broad mechanisms that take place *intra* and *intercellularly*. However, it is still highly challenging to analyze the whole set of proteins from a cell or tissue. Thus, the plant proteome analysis performed nowadays is still representing a partial view of the total cellular proteome.

Nevertheless, more than 10 years ago, the impact of proteomics analysis to the understanding of plant development and responses to biotic and abiotic stresses and to the identification of the protein repertoire of organelles, tissues and sub-cellular compartments was already clear (Wienkoop et al. 2010). Several initiatives including the creation of LC-MS/MS spectral library repository ProMEX with 116,364 tryptic peptide product ion spectra entries of 13 plant species

(Wienkoop et al. 2012) and the creation of the Multinational Arabidopsis Steering Subcommittee to coordinate Arabidopsis proteome international research were efforts toward the consolidation of databases and data resources that contributed to address some of the main challenges on plant proteomics through the creation of integrated proteome repositories and data analysis platforms, including MASCP Gator (Joshi et al. 2011). Other efforts such as International Plant Proteomics Organization (INPPO) (<http://www.inppo.com/>) were also on place. Currently, some of the most complete information about Arabidopsis proteome are organized in the ProteomicsDB (<https://www.proteomicsdb.org>), which contains MS-based proteomics meta-data and proteomics expression profiles for 30 different tissues, that can be visualized through body maps (Samaras et al. 2020). Another source of plant proteomics data is the PlantPREs (URL www.proteome.ir), an online database of plant proteome related to stresses, containing more than 20,413 protein entries and their expression patterns, extracted from 456 manually curated articles (Mousavi et al. 2016). Alternatively, there is the ATHENA (Arabidopsis THaliana ExpressioN Atlas; http://athena.proteomics.wzw.tum.de:5002/master_arabidopsis-shiny/) database, which has a collection of more than 18,000 proteins and their expression profiles for a set of 30 matching tissues from *Arabidopsis thaliana* (Col-0). It allows the user to explore the comparative expression analysis of proteins in different tissues, to visualize enriched pathways, phosphorylation sites, and similar global gene expression profiles in different tissues of *Arabidopsis thaliana*.

Very recently, a detailed MS-based draft of the *Arabidopsis* proteome was described (Mergner et al. 2020) based on the information retrieved from the proteomics databases ProteomicsDB and ATHENA. In this draft, the molecular data retrieved from more than 18,000 proteins identified revealed that most transcripts and proteins are actually expressed in a non-tissue-specific manner, with only a few transcripts or proteins being expressed in a tissue-specific manner, as it was evidenced for the proteins exclusively identi-

fied in pollen. The authors have found that different tissue types may have distinct quantitative abundance patterns of proteins, showing a positive correlation (Pearson's correlation $r = 0.28-0.7$) between the transcript and protein levels in most tissues (Mergner et al. 2020). However, in recent single-cell proteomics approaches, this behavior was not observed for most of the genes, and there was low correlation of mRNA and protein levels in root hair analysis (Wang et al. 2016), with higher positive correlation of transcript and protein levels found mostly for highly expressed genes.

Some other multi-omics databases and frameworks are operating to integrate functional genomics data providing annotation and visualization options for diverse plant species, including platforms such as ePlant (<http://bar.utoronto.ca/eplant/>), Virtual Plant (<http://virtualplant.bio.nyu.edu/cgi-bin/vpweb/>), MapMan (<https://mapman.gabipd.org/mapmanstore>) (Schwacke et al. 2019), PLAZA (<https://bioinformatics.psb.ugent.be/plaza/>), Plant Metabolic Network (PMN) (<https://plantcyc.org/>), Plant Reactome (<https://plantreactome.gramene.org/>), among others. Right now, some of these platforms contain mostly genomics and transcriptomics datasets, with few proteomics datasets. However, the integration of several layers of plant omics information seems to be in the horizon of these platforms, which is essential for integrative analysis and modeling approaches and the continuous development of systems biology studies in plants (Falter-Braun et al. 2019).

The permanent development and improvement of such databases is essential to guarantee the availability of plant proteomics data for future integrative omics analysis. As the proteomics data is everyday more abundant, it would be important to include omics data of all plant species in these repositories, not only model plants, since the proteomics data is currently the linkage point between the transcriptome information and the cell metabolism.

Quantitative MS-based proteomics seems to be currently the most prominent proteomics approach for investigating plant molecular phenotypes. Some of the most recent approaches for

discovery quantitative proteomics (non-targeted proteomics) include chemical labeling of the proteins or peptides. The most common strategies for relative and quantitative analysis of plant proteomics include one of the following techniques: isobaric tags for relative and absolute quantitation (iTRAQ) (Unwin et al. 2005), tandem mass tags (TMT) (Thompson et al. 2003), metabolic labeling such as stable isotope labeling by amino acids (SILAC) (Ong et al. 2002), and stable isotope ^{15}N (Oda et al. 1999) or label-free quantification (Washburn et al. 2001).

In many different studies performed using shotgun proteomics with LC and tandem MS (MS/MS), the use of data-dependent acquisition (DDA) was the method of choice, combined or not with peptide labeling with tandem mass tag (TMT) for quantitative or absolute quantification in proteomic analysis.

Multiplexing TMT or the use of other peptide labeling techniques, such as isobaric labeling, has been recently applied for the comparative proteome analysis, revealing cellular mechanisms and elucidating the role of candidate genes in diverse cellular responses. The time-resolved analysis of plant processes combined with the quantitative comparative proteomics is revealing novel aspects of plant molecular physiology and will certainly serve as basis for modeling and recognition of regulatory principles of cell response and metabolism.

For instance, in a time-resolved redox proteome analysis using thiol-specific iodo TMTs and LC-MS/MS analysis, the evidences for a fundamental strategy of rapid control of the cell metabolism during seed germination through the modifications of cysteines were confirmed. This biological process seems to be controlled by the function of hundreds of Cys-based redox switches which are operational even when hormonal and genetic programs are not yet functional. With 741 Cys peptides identified and quantified, it was demonstrated that tricarboxylic acid cycle is regulated by thioredoxins (TRX), through the TRX-mediated redox modulation of the activity of succinate dehydrogenase and fumarase. The proteome analysis also indicated that all shifts observed were reductive, indicating the influx of

electrons into the complex cellular thiol redox systems during the early stages of seed germination (Nietzel et al. 2020).

A recent proteome analysis of autophagy-deficient *Arabidopsis* seedlings indicated that this process is a response rapidly activated under diverse stimuli, including microbial elicitors, danger signals, and hormones. Quantitative comparative proteome analysis using TMTs identified more than 11,000 proteins and showed that autophagy is associated with cellular phenotypic plasticity. In autophagy-deficient cells, the plasticity was reduced, and the cell dedifferentiation impaired, indicating that autophagy is an essential mechanism for cells to reprogramming their development in response to several different stimulus, being necessary for wound-induced dedifferentiation and tissue repair (Rodriguez et al. 2020).

In another approach, the information about the differences between plant tissues (roots, above-ground parts, cauline leaves, 13 stages of flowers, and organs/stages) was investigated through proteome and phosphoproteome analysis performed with LC-MS/MS and iTRAQ labeling of six developmental stages of 16-day-old *Arabidopsis* plants. This work revealed the identity of 2187 proteins and the pattern of protein expression phosphorylated peptides (Lu et al. 2020). The comparative proteomics analysis indicated that reproductive organs expressed, in a more prominent manner, proteins related to translation and metabolic processes, while plant seedlings present a protein repertoire enriched in proteins related to oxidation-reduction and responses to different stresses. In the same work, the transcription factor and transcriptional regulator proteins were identified and a network of putative kinase substrates was generated, reinforcing evidences that bZIP16 transcription factor may be the substrate of Map kinase 6 (MAPK6) during floral development, integrating light and hormone signaling pathways during early seedling development (Lu et al. 2020).

Some plant proteomic approaches also included the implementation of Data-Independent Acquisition (DIA) (Venable et al. 2004), which acquires data from MS1 and MS2 spectra in MS/

MS applications, without pre-selection of precursors. The acquired data is usually analyzed by computational creation of spectral libraries and/or by using these libraries or consulting preexisting libraries for searching characteristic mass spectra (Zhang et al. 2020).

A complementation between discovery and targeted proteomics approaches can be initially addressed by sequential window acquisition of all theoretical mass spectra (SWATH)/DIA (data-independent acquisition) and by comparison of these spectra with constructed spectral libraries. For instance, by developing a mass spectra library, the effects of abscisic acid in *Arabidopsis* were investigated (Zhang et al. 2019). An extensive analysis of *Arabidopsis* proteome performed the quantification of 8793 proteins using a combination of untargeted LC-MS/MS using DDA and DIA and generation of a spectral library. The effects of the hormone abscisic acid (ABA) in *Arabidopsis* proteome was investigated, rendering detailed evidences of the previously described role of oxidative-reduction processes induced by ABA in the cells, and the transitory or gradual response of plant metabolism exposed to ABA with an initial reduction of the metabolic cellular activity and increase in ribosome biogenesis after 2 h of ABA posttreatment followed by an increased metabolism of carbohydrate and sucrose, and reduced metabolism of *N*-acetylglucosamine and macromolecules at 72 h posttreatment (Zhang et al. 2019).

Within an integrative omics analysis, a quantitative label-free proteomics approach investigated the role of autophagy in maize (*Zea mays*). Wild-type maize background plants (W22) were compared with mutant plants for ATG12 protein, one of the proteins identified as AUTOPHAGY-RELATED (ATG) (McLoughlin et al. 2018). The study was conducted in selected two leaves (two and four), which are rapidly expanding sink tissues, in response to short nitrogen stress. The quantitative proteome analysis revealed main biological processes affected in nitrogen starved and non-starved plants, indicating that autophagy is involved in many nutrient recycling mechanisms, but it is also an

important mechanism in the maintenance of normal protein abundance in plants. Other effects of the mutation of *atg12* were observed in the alteration of the abundance of proteins related to biosynthesis of phenylpropanoids, fatty acids, aromatic amino acids, and the positive correlation of the proteome and transcriptome data for genes associated with phenylpropanoid metabolism and glutathione transferase activities. These results suggested that alterations in the autophagic turnover of molecules such as pigments, antioxidants, lipids, among others, are essential for the resulting plant leaf phenotype, even under non-stress, nitrogen-rich conditions (McLoughlin et al. 2018). Further studies of the maize proteome under carbon-stress conditions reinforced the aspects of autophagy as a critical process for proteostasis in plants, likely by the recycling of proteins and organelles. To increase the number of proteins identified, the authors performed two MS runs in a data-dependent mode and another two runs of the same extracts using an exclusion list of the 5000 most prominent peptides from the first analysis, which increased the depth of the proteome analysis of the further runs (McLoughlin et al. 2020). The proteome results of this approach indicated that leaves of maize *atg12* mutant have increased levels of ribosome-associate proteins, and proteins related to redox homeostasis and catabolism of fatty acids, amino acids, small molecules, nucleotides, and glutathione (GSH), suggesting that even under conditions of impaired autophagy, the cells respond partially in a similar way as with full autophagy systems. These effects may be part of a complex compensatory mechanism that may take place in plants serving as alternative for the deficient autophagy (McLoughlin et al. 2020).

The quantitative proteome analysis is significantly contributing to increase our understanding of several complex phenomena in plants. Large studies using time-resolved in-depth quantitative approaches will certainly disclose regulatory aspects of the cellular mechanisms related to the control of cell growth, carbon usage, and stress responses in plants.

3.2.3 Proteomics Analyses of Posttranslational Modifications (PTMs)

With the increased throughput and mass resolution of the current MS instruments, the identification of chemical modifications of peptides or intact proteins occurs through the analysis of posttranslational modifications (PTMs), which do refer to the covalent and generally enzymatic modification of proteins following protein biosynthesis. These modifications generally include potential changes in protein sub-cellular localization, protein stabilization/degradation, enzyme activity, and interactions with protein partners or other biomolecules (Friso and van Wijk 2015; Spoel 2018). The occurrence of PTMs greatly expands proteome complexity and diversity, affecting numerous cellular signaling events and responses to the environment (Friso and van Wijk 2015; Larsen et al. 2006). Based on the type of modification, more than 300 potential PTMs can occur *in vivo* including: (a) reversible/irreversible addition of chemical groups (phosphorylation, acetylation, methylation, and redox-based modifications), (b) reversible addition of polypeptides (ubiquitination, SUMOylation, and other modifications by ubiquitin-like (Ubl) polypeptide), (c) reversible addition of complex molecules (glycosylation, attachment of lipids, ADP-ribosylation, and AMPylation), and (d) irreversible direct modification of amino acids (deamidation, eliminination) or protein cleavage by proteolysis (Larsen et al. 2006; Spoel 2018; Vu et al. 2018). In addition, apart from a single regulatory PTM role, there is also potential crosstalk with other PTMs, making the mechanisms and dynamics of protein modification still more complex (Arsova et al. 2018; Du et al. 2019; Vu et al. 2018).

The study of PTMs is considered technically demanding due to the labile nature, low stoichiometry, and abundance of protein modification when analyzing whole-cell lysates (Larsen et al. 2006; Swaney and Villen 2016). Therefore, a PTM enrichment step for modified peptides is normally necessary before MS analysis (Murray et al. 2012; Swaney and Villen 2016). Dealing

with complex plant proteomes of model and non-model species, it is important to select adequate proteomic approaches aiming the identification of proteins and analysis of their dynamic PTMs (Hu et al. 2015). Traditionally, PTMs have been identified by Edman degradation, amino acid analysis, isotopic labeling, or immunochemistry (Larsen et al. 2006). Nowadays, shotgun proteomics is one of the most widely used approaches to analyze PTMs (Yu et al. 2020). The adoption of large-scale quantitative proteomic approaches including isotope-coded affinity tags (ICATs), tandem mass tags (TMTs), and isobaric tags for relative, and absolute quantitation (iTRAQ) is enabling a more confident identification and multiplexed quantitation of PTMs (Hu et al. 2015; Liu et al. 2019; Murray et al. 2012). Once a PTM site is identified, biological characterization of protein modifications can be addressed with targeted MS-based quantitative approaches such as multiplexed selective reaction monitoring (SRM) or Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra (SWATH MS) (Arsova et al. 2018; Sidoli et al. 2015).

Several aspects of plant metabolism and development involve signaling events. Different inorganic or organic compounds can function as signaling molecules that are going to regulate the plant cell behavior. Among the many possible molecular events that may take place in plant signaling mechanisms, the posttranslational modification of proteins is of utmost interest as it controls a vast set of cellular responses, including growth, membrane trafficking, gene expression, degradation, to mention a few.

The use of proteomics to uncover signaling mechanisms is step by step taking advantage of the high resolution and throughput of the MS instruments to uncover the signaling networks of the plant cells. In the current investigations, the analysis of protein targets of the reactive oxygen species (ROS) and reactive nitrogen species (RNS) is gaining more attention due to their promising role in plant growth through the redox homeostasis. The investigation of the cellular role of these chemical species can be done through the analysis of the chemical modifications in the target molecules generated

by the reactions with reactive oxygen and/or nitrogen species.

Among the several possible protein oxidoreduction PTMs that can be identified in plants, the protein nitrosation (or nitrosylation) which represents target protein alterations in response to the function of nitric oxide has gained more attention as central process in defining carbon and nitrogen metabolism in land plants (Nabi et al. 2020) and microalgae under stress conditions (De Mia et al. 2019; Morisse et al. 2014).

Cysteine *S*-nitrosation is a redox-based PTM that mediates major physiological and biochemical effects of nitric oxide (NO). Emerging evidence indicates that protein *S*-nitrosation is ubiquitously involved in the regulation of plant development and stress responses (Feng et al. 2019; Gong et al. 2019). Despite its importance in plants, studies exploring protein signaling pathways that are regulated by *S*-nitrosation during plant development and embryogenesis are still scarce, particularly in non-model plant species. For instance, using redox proteome analysis of *S*-nitrosation, a considerable array of proteins associated with a large variety of molecular functions were identified in Brazilian pine proteome, generating novel insights into the roles of *S*-nitrosation during somatic and zygotic embryo development. Using a method adapted to the Biotin-Switch (Forrester et al. 2009), replacing biotin with the iodo-TMT126 (iodo-Tandem Mass Tag), the occurrence of *in vivo* and *in vitro* *S*-nitrosation was investigated during somatic and zygotic embryo formation of Brazilian pine (*Araucaria angustifolia* (Bertol.) Kuntze), an endangered native conifer of South America.

Previous analyses using physiological, transcriptomic, and quantitative proteomics approaches (dos Santos et al. 2016; Elbl et al. 2015; Silveira et al. 2006) suggested a potential influence of the redox environment and nitric oxide production during Brazilian pine embryo formation. The *S*-nitrosoproteome analyses identified 158 *S*-nitrosylated proteins *in vitro* (i.e., via the incorporation of a NO donor), with 36 proteins detected during seed development (globular embryo until late cotyledonal stage) and 122 proteins detected during somatic embryo formation

(transition from proembryogenic masses to early somatic embryos). This study indicated that most *S*-nitrosylated proteins were involved in metabolism of primary compounds (carbohydrates and nucleic acids) and cellular processes and signaling (turnover of proteins and chaperones). For late-stage embryogenesis, functions associated with *S*-nitrosylated proteins were stress resistance (abiotic stress response), cellular processes (signal translation, chaperones, and protein turnover), and metabolism (energy production, transport, and carbohydrate metabolism). Interestingly, 47 proteins were identified as endogenous *S*-nitrosylated during early and late-stage embryogenesis, suggesting a role of this PTM during Brazilian pine embryo formation. The possibility of generating a stable bond between Cys and iodo-TMT126 enabled the identification of labeled peptides using mass spectrometry and the determination of the position of nitrosylated Cys residues. The identification of the possible cellular biological processes affected in non-model plants by NO reinforce the observations about the myriad of functions regulated by the protein PTMs generated by ROS and RNS, which are likely coordinated by a complex regulatory network (Leon and Costa-Broseta 2020). The cellular effects coordinated by these networks are possibly accomplished by the interplay between redox homeostasis systems which involve cysteine oxidative modifications and the systems of thioredoxins and disulfide reductases, which are expressed by several genes in plants, composing a highly complex regulatory mechanism of cellular response (Navrot et al. 2011).

3.3 Proteomics of Non-model Species

The many challenges already present in the plant proteomics analysis are more pronounced in the investigation of non-model plant species. This is especially true after the past decade, when the number of species with a fully sequenced genome grew substantially, with non-model plant species distinguished by the lack of experimentally validated functional evidence for a great variety of

annotated genes. This scarcity on information may result from long life cycle, large genome size (often polyploid) and recalcitrance to laboratory cultivation, whose characteristics are opposed to model plants, such as *Arabidopsis thaliana*. The study on non-model plants greatly relied on establishing gene and protein sequence homology analyses with model organisms, but this approach is questionable when there is only far phylogenetic relation between the species analyzed. As the more distant organisms are phylogenetically lesser there can be functional extrapolation (Heck and Neely 2020). Plants are physiologically diverse, implying that sometimes homology-based annotation does not accurately indicate the function of groups of genes or their relation to complex cellular mechanisms. Additional co-expression or regression analyses have been shown as powerful tools to be used for the search of functional categories and novel pathways inherent of non-model plant proteomes, which may guide functional investigations in non-model plant species.

Currently, the progress on genome sequencing and annotation, followed by advances in bioinformatics and data integration offer analytical tools similar to the ones used for model plants (Bolger et al. 2017). We are now living the post-model organism era, whose resources have been being well utilized and essential for non-model plants research, such as crops (Heck and Neely 2020). One example is sugarcane, a C4 grass able to accumulate large amounts of sucrose and considered the world's leading biomass crop (Souza et al. 2019). For a long time, rice (also a monocot but with a C3 metabolism) and maize were the "model" organisms phylogenetically closer to sugarcane. The first sugarcane proteomic studies tested different extraction methods (Amalraj et al. 2010), identified specific classes (Cesarino et al. 2012), and provided the characterization of abiotic stress-response (Zhou et al. 2012). Until 2013, proteomes surveys of this crop were scarce (Boaretto and Mazzafera 2013). However, after several perseverant efforts on *Saccharum* spp. genome sequencing and assembly (Boaretto and Mazzafera 2013; Garsmeur et al. 2018; Grativol et al. 2014; Miller et al. 2017; Okura et al. 2016;

Riano-Pachon and Mattiello 2017; Souza et al. 2019; Vettore et al. 2003; Vilela et al. 2017), the catalog of proteins identified by mass spectrometry largely increased for both descriptive and several sorts of biological conditions (for a review see Miller et al. 2017). In this highly polyploid species, the identification of glycoside hydrolases (GH) at the protein level paved the way for target experimentation. Domain prediction and homology investigation indicated that members of the GH families are numerous in sugarcane, but only dedicated extraction and analyses of proteomic pipelines could point to which exact proteins and at which amount were present in different organs and developmental stages (Calderan-Rodrigues et al. 2014, 2016; Fonseca et al. 2018). Combined omics and protein activity essays from these retrieved GHs showed that the cell wall degradation occurring on a particular tissue was timely orchestrated by different proteins that tackled pectin, callose, hemicelluloses, and cellulose at last (Grandis et al. 2019). The GHs identified in this cell disassembly mechanism could be manipulated in a timely controlled fashion to produce plants more amenable to saccharification resulting in augmented ethanol yield. For dedicated functional studies, an alternative that has been used for non-model plants is the transference of the target gene to a model host, allowing a better comprehension of the metabolic changes. This approach was successfully employed for the sugarcane cell wall-related transcription factor SHINE (Martins et al. 2018) and a Dirigent-Jacalin (Andrade et al. 2019).

So far, the selection of target genes for non-model plants has been performed mostly based on genomic, transcriptomic, or in silico data. Furthermore, the possibility to promptly generate RNAseq data can provide draft proteome databases for MS-identification, which makes the use of proteogenomics a powerful tool to study non-model plant species (Armengaud et al. 2014). One of the frontiers to be crossed relies on a more straightforward use of plant proteomics as an instrument to select the expressing alleles and thus markers more successfully related to phenotype. However, proteomic pipelines did not reach

the same advancement levels as transcriptomic ones, and we urge to have sensitive, high-throughput, and low-cost technologies in order to make this possible (Peace et al. 2019).

The study of non-model plants is extensively laborious and sometimes does not attract as many attention as the possibility to deepen the investigation on known biological pathways by using model plants. However, proteomic investigations of non-model species have brought light to exciting discoveries related to plant metabolic and developmental specificities (Grandis et al. 2019; Sergeant et al. 2019), biomass accumulation (Calderan-Rodrigues et al. 2019), disease responses (Diaz-Vivancos et al. 2006; He et al. 2012; Tahara et al. 2003; Zhang et al. 2015), and several sorts of abiotic stresses (Aghaei and Komatsu 2013; Cia et al. 2018; Huang and Sethna 1991; Wang et al. 2017). Research in this area not only contributed to biological insights from these specific conditions and species but also provided a more complete picture of the whole autotrophic nature, allowing to deepen the evolutionary discussion as well. Once confined to conditional conclusions, non-model plant proteomics can now take a step further, and instead of conflictive, data integration between model and non-model ones and target experimentation will allow these two groups to be considered of the same level. This will attract interest for non-model species research and will fuel a virtuous cycle to provide more and more biological information.

3.4 Future Directions

The recent advances in proteomics analysis of model and non-model plant species have demonstrated that the evolution of techniques and instrumentation had revolutionized how fast we can identify the plant molecular phenotypes. Emergent characteristics have been uncovered by the proteomics, and the proper discovery of many linkages between transcriptomics and other omics through the quantification of proteins is revealing several processes and cellular mechanisms that may play fundamental role in regulating cell growth and metabolism.

It is clear that a big challenge in the proteomics field is the limitation of analyzing, at the same time, all aspects of the plant proteome, specially the regulation of cellular responses, such as the function of transcription factors and regulators, the roles of PTMs, and redox homeostasis in cellular phenotypes. The analysis of how individual cells respond and contribute to an organismic phenotype are revealing how complex the interaction between cells are in defining a phenotype. For instance, there are already indications of molecular similarities between different tissues or parts of a plant but vast differences at the level of cell types or individual cells. The combination of the total proteome analysis of organisms with targeted cell type proteome may drive comprehensive systems analysis of plant phenotypes, mixing the power of holistic analysis with the observation of cell type specificities, specially the regulatory ones.

The analysis of plant phenotypes has been performed using different proteomics approaches, which in many cases integrate cell biology and analytical techniques. This knowledge, mainly based on the application of shotgun mass spectrometry-based proteomics, has brought to light specific quantitative and qualitative information of the molecular physiology of plants.

Nevertheless, plant proteomics at the single-cell level or single-cell type can and will likely contribute to the understanding of the dynamic changes that occur in cells in response to environmental alterations and to regulatory aspects of the plant differentiation or development. It may also disclose the importance, rules, and the degree of heterogeneous responses cell population exert in defining a plant phenotype, elucidating how different mechanisms are coordinated to generate a phenotype from a series of individual cellular responses (Libault et al. 2017).

The use of gel-based proteomics approach is probably taking a more consistent path toward the comparative analysis of subcellular proteomes and targeted analysis of protein complexes, while been extensively used for comparative proteomics analysis of non-model plant species coupled to protein identification by MS.

Time-resolved plant shotgun MS-based proteomics analysis is still the most prominent approach currently applied in the plant proteomic science and will remain like that for the near future, unless a large-scale protein sequencing strategy is developed that could compete or substitute the MS-based proteomics, for some analysis at least.

This technical advance would contribute to drastically expand the number of proteome studies of non-model plant species, bringing more information for a future large comparative systems biology analysis, which may reveal emergent properties of the plant development and responses to environmental changes.

References

- Aghaei K, Komatsu S (2013) Crop and medicinal plants proteomics in response to salt stress. *Front Plant Sci* 4:8
- Amalraj RS et al (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
- Andrade LM et al (2019) Biomass accumulation and cell wall structure of rice plants overexpressing a Dirigent-Jacalin of sugarcane (ShDJ) under varying conditions of water availability. *Front Plant Sci* 10:65
- Armengaud J et al (2014) Non-model organisms, a species endangered by proteogenomics. *J Proteome* 105:5–18
- Arsova B, Watt M, Usadel B (2018) Monitoring of plant protein post-translational modifications using targeted proteomics. *Front Plant Sci* 9:1168
- Bestel-Corre G et al (2002) Proteome analysis and identification of symbiosis-related proteins from *Medicago truncatula* Gaertn. by two-dimensional electrophoresis and mass spectrometry. *Electrophoresis* 23(1):122–137
- Boaretto LF, Mazzafera P (2013) The proteomes of feedstocks used for the production of second-generation ethanol: a lacuna in the biofuel era. *Ann Appl Biol* 163(1):12–22
- Bolger M et al (2017) From plant genomes to phenotypes. *J Biotechnol* 261:46–52
- Boughton BA et al (2016) Mass spectrometry imaging for plant biology: a review. *Phytochem Rev* 15:445–488
- Calderan-Rodrigues MJ et al (2014) Cell wall proteomics of sugarcane cell suspension cultures. *Proteomics* 14(6):738–749
- Calderan-Rodrigues MJ et al (2016) Cell wall proteome of sugarcane stems: comparison of a destructive and a non-destructive extraction method showed differences in glycoside hydrolases and peroxidases. *BMC Plant Biol* 16:14
- Calderan-Rodrigues MJ et al (2019) Plant cell wall proteomics: a focus on monocot species, brachypodium distachyon, *Saccharum* spp. and *Oryza sativa*. *Int J Mol Sci* 20(8):1975
- Cesarino I et al (2012) Enzymatic activity and proteomic profile of class III peroxidases during sugarcane stem development. *Plant Physiol Biochem* 55:66–76
- Chang TS et al (2017) Mapping and comparative proteomic analysis of the starch biosynthetic pathway in rice by 2D PAGE/MS. *Plant Mol Biol* 95(4–5):333–343
- Cia MC et al (2018) Novel insights into the early stages of ratoon stunting disease of sugarcane inferred from transcript and protein analysis. *Phytopathology* 108(12):1455–1466
- De Mia M et al (2019) Nitric oxide remodels the photosynthetic apparatus upon S-starvation in *Chlamydomonas reinhardtii*. *Plant Physiol* 179(2):718–731
- Delaunoi B et al (2014) Uncovering plant-pathogen crosstalk through apoplastic proteomic studies. *Front Plant Sci* 5:249
- Diaz-Vivancos P et al (2006) The apoplastic antioxidant system in *Prunus*: response to long-term plum pox virus infection. *J Exp Bot* 57(14):3813–3824
- dos Santos AL et al (2016) Quantitative proteomic analysis of *Araucaria angustifolia* (Bertol.) Kuntze cell lines with contrasting embryogenic potential. *J Proteome* 130:180–189
- Du H et al (2019) A new insight to explore the regulation between S-nitrosylation and N-glycosylation. *Plant Direct* 3(2):e00110
- Edman P (1949) A method for the determination of amino acid sequence in peptides. *Arch Biochem* 22(3):475
- Elbl P et al (2015) Comparative transcriptome analysis of early somatic embryo formation and seed development in Brazilian pine, *Araucaria angustifolia* (Bertol.) Kuntze. *Plant Cell Tissue Organ Cult* 120(3):903–915
- Falter-Braun P et al (2019) iPlant Systems Biology (iPSB): an international network hub in the plant community. *Mol Plant* 12(6):727–730
- Feng J, Chen L, Zuo J (2019) Protein S-nitrosylation in plants: current progresses and challenges. *J Integr Plant Biol* 61(12):1206–1223
- Fenn JB et al (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246(4926):64–71
- Fonseca JG et al (2018) Cell wall proteome of sugarcane young and mature leaves and stems. *Proteomics*:18(2)
- Forrester MT et al (2009) Detection of protein S-nitrosylation with the biotin-switch technique. *Free Radic Biol Med* 46(2):119–126
- Friso G, van Wijk KJ (2015) Posttranslational protein modifications in plant metabolism. *Plant Physiol* 169(3):1469–1487
- Gallardo K et al (2001) Proteomic analysis of arabidopsis seed germination and priming. *Plant Physiol* 126(2):835–848

- Garsmeur O et al (2018) A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat Commun* 9(1):2638
- Giavalisco P et al (2005) Proteome analysis of *Arabidopsis thaliana* by two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionisation-time of flight mass spectrometry. *Proteomics* 5(7):1902–1913
- de Godoy LM et al (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* 455(7217):1251–1254
- Gong B et al (2019) Unravelling GSNOR-mediated S-nitrosylation and multiple developmental programs in tomato plants. *Plant Cell Physiol* 60(11):2523–2537
- Grandis A et al (2019) Cell wall hydrolases act in concert during aerenchyma development in sugarcane roots. *Ann Bot* 124(6):1067–1089
- Grativol C et al (2014) Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus *Saccharum*. *Plant J* 79(1):162–172
- Gross JH (2017) *Mass spectrometry: a textbook*. Springer International Publishing, Cham
- Hart-Smith G et al (2017) Improved quantitative plant proteomics via the combination of targeted and untargeted data acquisition. *Front Plant Sci* 8:1669
- He D, Yang P (2013) Proteomics of rice seed germination. *Front Plant Sci* 4:246
- He L et al (2012) Proteomic analysis of the effects of exogenous calcium on hypoxic-responsive proteins in cucumber roots. *Proteome Sci* 10(1):42
- Heck M, Neely BA (2020) Proteomics in non-model organisms: a new analytical frontier. *J Proteome Res* 19(9):3595–3606
- Hillenkamp F, Karas M (1990) Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Methods Enzymol* 193:280–295
- Hu J, Rampitsch C, Bykova NV (2015) Advances in plant proteomics toward improvement of crop productivity and stress resistance. *Front Plant Sci* 6:209
- Huang M, Sethna JP (1991) History dependence of a two-level system. *Phys Rev B Condens Matter* 43(4):3245–3254
- Jorin-Novo JV et al (2019) Gel electrophoresis-based plant proteomics: past, present, and future. Happy 10th anniversary *Journal of Proteomics!* *J Proteome* 198:1–10
- Joshi HJ et al (2011) MASCOP: an aggregation portal for the visualization of *Arabidopsis* proteomics data. *Plant Physiol* 155(1):259–270
- Kaspar S et al (2011) MALDI-imaging mass spectrometry - an emerging technique in plant biology. *Proteomics* 11(9):1840–1850
- Kim J, Woo HR, Nam HG (2016) Toward systems understanding of leaf senescence: an integrated multi-omics perspective on leaf senescence research. *Mol Plant* 9(6):813–825
- Komatsu S, Muhammad A, Rakwal R (1999) Separation and characterization of proteins from green and etiolated shoots of rice (*Oryza sativa* L.): towards a rice proteome. *Electrophoresis* 20(3):630–636
- Kosova K, Vitamvas P, Prasil IT (2014) Proteomics of stress responses in wheat and barley-search for potential protein markers of stress tolerance. *Front Plant Sci* 5:711
- Larsen MR et al (2006) Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *BioTechniques* 40(6):790–798
- Leon J, Costa-Broseta A (2020) Present knowledge and controversies, deficiencies, and misconceptions on nitric oxide synthesis, sensing, and signaling in plants. *Plant Cell Environ* 43:1
- Li L et al (2019) Comparative proteomic analysis provides insights into the regulatory mechanisms of wheat primary root growth. *Sci Rep* 9(1):11741
- Libault M et al (2017) Plant systems biology at the single-cell level. *Trends Plant Sci* 22(11):949–960
- Liu Y et al (2019) Proteomics: a powerful tool to study plant responses to biotic stress. *Plant Methods* 15:135
- Lu J et al (2020) Global quantitative proteomics studies revealed tissue-preferential expression and phosphorylation of regulatory proteins in *Arabidopsis*. *Int J Mol Sci* 21(17):6116
- Martins de Souza D et al (2008) The untiring search for the most complete proteome representation: reviewing the methods. *Brief Funct Genomic Proteomic* 7(4):312–321
- Martins APB et al (2018) Ectopic expression of sugarcane SHINE changes cell wall and improves biomass in rice. *Biomass Bioenergy* 119:322–334
- Marx V (2019) A dream of single-cell proteomics. *Nat Methods* 16(9):809–812
- McLoughlin F et al (2018) Maize multi-omics reveal roles for autophagic recycling in proteome remodelling and lipid turnover. *Nat Plants* 4(12):1056–1070
- McLoughlin F et al (2020) Autophagy plays prominent roles in amino acid, nucleotide, and carbohydrate metabolism during fixed-carbon starvation in maize. *Plant Cell* 32(9):2699–2724
- Mergner J et al (2020) Mass-spectrometry-based draft of the *Arabidopsis* proteome. *Nature* 579(7799):409–414
- Mettler T et al (2014) Systems analysis of the response of photosynthesis, metabolism, and growth to an increase in irradiance in the photosynthetic model organism *Chlamydomonas reinhardtii*. *Plant Cell* 26(6):2310–2350
- Miller JR et al (2017) Initial genome sequencing of the sugarcane CP 96-1252 complex hybrid. *F1000Res* 6:688
- Morisse S et al (2014) Insight into protein S-nitrosylation in *Chlamydomonas reinhardtii*. *Antioxid Redox Signal* 21(9):1271–1284
- Mousavi SA et al (2016) PlantPREs: a database for plant proteome response to stress. *J Proteome* 143:69–72
- Murray CI et al (2012) Identification and quantification of S-nitrosylation by cysteine reactive tandem mass tag switch assay. *Mol Cell Proteomics* 11(2):M111.013441
- Nabi RBS et al (2020) Functional insight of nitric-oxide induced DUF genes in *Arabidopsis thaliana*. *Front Plant Sci* 11:1041
- Navrot N et al (2011) Plant redox proteomics. *J Proteome* 74(8):1450–1462

- Neubauer G et al (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat Genet* 20(1):46–50
- Nietzel T et al (2020) Redox-mediated kick-start of mitochondrial energy metabolism drives resource-efficient seed germination. *Proc Natl Acad Sci U S A* 117(1):741–751
- O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250(10):4007–4021
- Oda Y et al (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* 96(12):6591–6596
- Okura VK et al (2016) BAC-pool sequencing and assembly of 19 Mb of the complex sugarcane genome. *Front Plant Sci* 7:342
- Ong SE, Mann M (2005) Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* 1(5):252–262
- Ong SE et al (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1(5):376–386
- Pappireddi N, Martin L, Wuhr M (2019) A review on quantitative multiplexed proteomics. *ChemBioChem* 20(10):1210–1224
- Patole C, Bindschedler LV (2019) Chapter 4 - Plant proteomics: a guide to improve the proteome coverage. In: Meena SN, Naik MM (eds) *Advances in biological science research*. Academic Press, New York, NY, pp 45–67
- Peace CP et al (2019) Apple whole genome sequences: recent advances and new prospects. *Hortic Res* 6:59
- Picotti P, Aebersold R (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* 9(6):555–566
- Porubleva L et al (2001) The proteome of maize leaves: use of gene sequences and expressed sequence tag data for identification of proteins with peptide mass fingerprints. *Electrophoresis* 22(9):1724–1738
- Riano-Pachon DM, Mattiello L (2017) Draft genome sequencing of the sugarcane hybrid SP80-3280. *F1000Res* 6:861
- Rodiger A, Baginsky S (2018) Tailored use of targeted proteomics in plant-specific applications. *Front Plant Sci* 9:1204
- Rodriguez E et al (2020) Autophagy mediates temporary reprogramming and dedifferentiation in plant somatic cells. *EMBO J* 39(4):e103315
- Samaras P et al (2020) ProteomicsDB: a multi-omics and multi-organism resource for life science research. *Nucleic Acids Res* 48(D1):D1153–D1163
- Sanger F, Tuppy H (1951) The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *Biochem J* 49(4):463–481
- Schmidt A, Claassen M, Aebersold R (2009) Directed mass spectrometry: towards hypothesis-driven proteomics. *Curr Opin Chem Biol* 13(5–6):510–517
- Schwacke R et al (2019) MapMan4: a refined protein classification and annotation framework applicable to multi-omics data analysis. *Mol Plant* 12(6):879–892
- Sergeant K et al (2019) The dynamics of the cell wall proteome of developing alfalfa stems. *Biology (Basel)* 8(3):60
- Shevchenko A et al (1996a) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc Natl Acad Sci U S A* 93(25):14440–14445
- Shevchenko A et al (1996b) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 68(5):850–858
- Sidoli S et al (2015) Sequential window acquisition of all theoretical mass spectra (SWATH) analysis for characterization and quantification of histone post-translational modifications. *Mol Cell Proteomics* 14(9):2420–2428
- Silveira V et al (2006) Polyamine effects on the endogenous polyamine contents, nitric oxide release, growth and differentiation of embryogenic suspension cultures of *Araucaria angustifolia* (Bert.) O. Ktze. *Plant Sci* 171(1):91–98
- Smith RD et al (1990) New developments in biochemical mass spectrometry: electrospray ionization. *Anal Chem* 62(9):882–899
- Souza GM et al (2019) Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. *GigaScience* 8(12):giz129
- Spoel SH (2018) Orchestrating the proteome with post-translational modifications. *J Exp Bot* 69(19):4499–4503
- Swaney DL, Villen J (2016) Proteomic analysis of protein posttranslational modifications by mass spectrometry. *Cold Spring Harb Protoc* 2016(3):pdb.top077743
- Tahara ST, Mehta A, Rosato YB (2003) Proteins induced by *Xanthomonas axonopodis* pv. *passiflorae* with leaf extract of the host plant (*Passiflorae edulis*). *Proteomics* 3(1):95–102
- Thompson A et al (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 75(8):1895–1904
- Unwin RD et al (2005) Quantitative proteomic analysis using isobaric protein tags enables rapid comparison of changes in transcript and protein levels in transformed cells. *Mol Cell Proteomics* 4(7):924–935
- Venable JD et al (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods* 1(1):39–45
- Vettore AL et al (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Res* 13(12):2725–2735
- Vilela MM et al (2017) Analysis of three sugarcane homo/homeologous regions suggests independent polyploidization events of *Saccharum officinarum* and *Saccharum spontaneum*. *Genome Biol Evol* 9(2):266–278

- Vu LD, Gevaert K, De Smet I (2018) Protein language: post-translational modifications talking to each other. *Trends Plant Sci* 23(12):1068–1080
- Wang H, Lan P, Shen RF (2016) Integration of transcriptomic and proteomic analysis towards understanding the systems biology of root hairs. *Proteomics* 16(5):877–893
- Wang Y et al (2017) Proteomic analysis of *Camellia sinensis* (L.) reveals a synergistic network in the response to drought stress and recovery. *J Plant Physiol* 219:91–99
- Washburn MP, Wolters D, Yates JR III. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19(3):242–247
- Wienkoop S, Baginsky S, Weckwerth W (2010) *Arabidopsis thaliana* as a model organism for plant proteome research. *J Proteome* 73(11):2239–2248
- Wienkoop S et al (2012) ProMEX - a mass spectral reference database for plant proteomics. *Front Plant Sci* 3:125
- Wilkins MR et al (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 13:19–50
- Yu F et al (2020) Identification of modified peptides using localization-aware open search. *Nat Commun* 11(1):4065
- Zhang CX, Tian Y, Cong PH (2015) Proteome analysis of pathogen-responsive proteins from apple leaves induced by the alternaria blotch *alternaria alternata*. *PLoS One* 10(6):e0122233
- Zhang H et al (2019) *Arabidopsis* proteome and the mass spectral assay library. *Sci Data* 6(1):278
- Zhang F et al (2020) Data-independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. *Proteomics* 20(17–18):e1900276
- Zhou G et al (2012) Proteomic analysis of osmotic stress-responsive proteins in sugarcane leaves. *Plant Mol Biol Report* 30(2):349–359