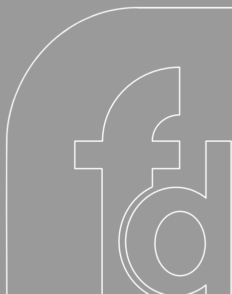
For reprint orders, please contact reprints@future-drugs.com



Understanding protein trafficking in plant cells through proteomics

Songqin Pan, Clay J Carter and Natasha V Raikhel†

The functions of approximately one-third of the proteins encoded by the *Arabidopsis thaliana* genome are completely unknown. Moreover, many annotations of the remainder of the genome supply tentative functions, at best. Knowing the ultimate localization of these proteins, as well as the pathways used for getting there, may provide clues as to their functions. The putative localization of most proteins currently relies on *in silico*-based bioinformatics approaches, which, unfortunately, often result in erroneous predictions. Emerging proteomics techniques coupled with other systems biology approaches now provide researchers with a plethora of methods for elucidating the final location of these proteins on a large scale, as well as the ability to dissect protein-sorting pathways in plants.

Expert Rev. Proteomics 2(5), 781–792 (2005)

CONTENTS

Brief background to protein transport in plants
Need for proteomics approaches
Previous studies & plant proteomic databases
New methods
Localization of organelle proteins by isotope tagging/isotope-coded affinity tag
Expert commentary & five-year view
Key issues
References
Affiliations

†Author for correspondence
University of California, Riverside, Center for Plant Cell Biology, Botany & Plant Sciences, University of California, Riverside, CA 92521, USA
Tel.: +1 951 827 6370
Fax: +1 951 827 2155
nraikhel@ucr.edu

KEYWORDS:
Arabidopsis, ICAT, LOPIT, organelle, protein trafficking, proteome, proteomics, rice, top-down

All eukaryotes must correctly establish and maintain their intracellular compartments as well as the plasma membrane and extracellular spaces. The primary component defining, and giving function to, a given cellular location is its native set of proteins. Thus, it is not surprising that mutations resulting in protein mislocalization, or ones that directly affect protein-sorting pathways, are often detrimental or even lethal (e.g., [1]). Certain pathogens disrupt or even hijack host trafficking mechanisms for their own purposes during infection processes [2].

Like other eukaryotes, the best understood mechanisms of protein localization in plants usually involve a signal derived from a specific sequence of amino acids on the protein to be sorted. With sequencing of the *Arabidopsis* genome complete [3], and rice [4,5] and poplar [6] genomes soon to follow, all proteins in these representative plant species can be predicted to reside in a given subcellular location based upon *in silico* approaches. Although often very effective at predicting the localization of proteins containing canonical sorting signals, current prediction programs (e.g., TargetP, Predotar and iPSORT) often provide erroneous and/or conflicting localizations for other proteins [7–10]. Complicating matters is the fact that start codons from genome sequencing projects are mispredicted with a frequency of approximately 30% [8] and that some proteins can have multiple true localizations (e.g., [11]). Moreover, little direct evidence exists for the ultimate cellular location of most proteins. While much is now understood about the basic pathways for delivery of proteins to various locations in plant cells, knowledge of specific localization for most proteins remains tenuous at best. Proteomics methodologies now allow for the large-scale identification of proteins in a given organelle. Knowing the protein content of a subcellular location can provide valuable insights into the mechanisms of protein transport in plant cells.

Brief background to protein transport in plants

Plants share a number of conserved protein localization sequences and transport mechanisms with other eukaryotes. In particular, similarities include nuclear localization signals (NLSs), cleavable mitochondrial transit peptides, cleavable signal peptides for directing proteins to the endoplasmic reticulum (ER), ER retention signals (e.g. KDEL and HDEL) and signals for lipid attachment (e.g., myristylation and prenylation) [12]. Plants also direct proteins throughout the endomembrane

10.1586/14789450.2.5.781
© 2005 Future Drugs Ltd
ISSN 1478-9450

system through vesicular trafficking. The primary basis for directing vesicles to their target membranes is mediated through soluble *N*-ethylmaleimide sensitive factor attachment protein receptor (SNARE) proteins [13], a mechanism apparently conserved in all eukaryotes. Even so, plants differ from most other eukaryotes in the need to direct proteins to specific subcompartments and membranes of the chloroplast [14] as well as multiple types of vacuoles [15].

Need for proteomics approaches

Classic approaches for the experimental determination of subcellular localization of proteins include biochemical and microscopy-based methodologies. Biochemical approaches primarily rely on cell homogenization followed by centrifugation-based fractionation. Proteins of interest are then detected in various fractions by western blotting and/or enzymatic assays. While effective, these methods require the time-consuming production of antibodies and/or the protein of interest having some sort of measurable activity. Furthermore, identifying the subcellular source of individual fractions is highly dependent upon previously developed markers.

Microscopy-based approaches for the determination of protein localization include immunoelectron microscopy, *in situ* labeling with fluorescently tagged antibodies, and fluorescent protein fusions (e.g., green fluorescent protein [GFP]) coupled with confocal microscopy. Although also effective for the determination of subcellular localization of individual proteins, in general, these methods also depend on the development of good antibodies or the production of multiple transgenic lines of plants, both of which often take a minimum of several months for development. In general, both biochemical and microscopy-based methods are performed on a single protein of interest at a time and can be considered low-throughput methods. However, the high-throughput tagging of *Arabidopsis* proteins with fluorescent markers has recently been described [16,17].

The functions of approximately 35% of the proteins encoded by the *Arabidopsis thaliana* genome are completely unknown, and sequencing of the rice genome is revealing even more complexity [4,5,16,18]. Moreover, many annotations of the remainder of the genome supply tentative functions based solely upon protein homology. Although members of large protein families may share significant identities, their functions can widely vary, (e.g., [19]). Knowing the ultimate localization of these proteins, as well as the pathways used for getting there, may give clues to their functions [8]. This is primarily based on the fact that cells spatially organize their proteins depending upon their function [20]. To address the aforementioned issues, recent years have seen an explosion in the development and employment of organelle proteomics (reviewed in [21–23]). These methods have primarily depended on subcellular fractionation to obtain highly homogenous starting samples followed by mass spectrometry (MS) for protein identification [20]; however, new methods now allow the identification of protein localization without entirely pure samples (e.g., see below; [24]). Current and developing proteomics approaches now allow the large-scale determination of protein localization in plants in a short time frame, and will likely facilitate the identification of presently unknown protein-sorting determinants. This review will provide a brief background of previous organelle proteomic studies in plants, new approaches for tackling these proteomes, and their implications for understanding protein trafficking in plant cells.

Previous studies & plant proteomic databases

In recent years, many studies on specific plant tissues and purified cellular compartments have been reported. Occurring mostly in *Arabidopsis* and rice (although other species have also been examined), these include proteomic studies of the nucleus, plastids, mitochondria, ER, Golgi apparatus, cell wall, peroxisomes, plasma membrane and vacuoles. Online databases containing readily accessible data are shown in TABLE 1. Representative primary publications and/or extensive review articles are also indicated.

Chloroplasts

The chloroplast is probably the best-characterized type of plastid. From the outside working in, it is composed of a highly porous outer envelope membrane followed by a mostly impermeable inner envelope [12]. Inside the inner envelope is the stroma (contains soluble proteins, chloroplast DNA, RNA and metabolites), then the thylakoid membrane (main site of photosynthesis) and the thylakoid space or lumen [12]. Each of these sublocations has its own set of specific functions and corresponding proteins. The vast majority of chloroplastic proteins – predicted to number up to 4000 in total [9,14] – are encoded by the nucleus, and understanding how they reach each sublocale has been a subject of significant study.

It has long been well established that nuclear-encoded proteins destined for the inner portions of the chloroplast (all but outer membrane) generally contain a cleavable N-terminal transit peptide (chloroplast transit peptide [cTP]), resulting in translocation across both the outer and inner envelopes by the translocon at the outer and inner envelope membrane of chloroplasts (Toc and Tic, respectively) import machinery [25]. Stromal proteins then assume their final conformation, while others contain additional information for sorting to other locations (reviewed in [14]). For example, luminal proteins are translocated across the thylakoid membrane by two distinct pathways – TAT and Sec – via secondary cleavable luminal transit peptides (LTP) [14]. While the mechanism(s) for how certain proteins end up in the thylakoid membrane (signal recognition particle pathway) has been elucidated, it is not known how numerous inner envelope and other thylakoid membrane proteins are localized (e.g., spontaneous pathway) [14].

Easily purified, the chloroplast is perhaps the most widely studied plant organelle via proteomics approaches. There are many publications describing proteomic studies on whole and purified subdomains of chloroplasts (extensively reviewed in [9]). Importantly, these examinations have aided in the understanding

Table 1. Plant organelle-specific proteome databases and/or representative publications.

Organelle	Dedicated Organelle Database	URL	Ref.
Plastid	Arabidopsis Plastid Proteome Database (APPD)	http://ppdb.tc.cornell.edu	[33]
Mitochondrion	Arabidopsis Mitochondrial Proteome Database (AMPDB)	www.ampdb.bcs.uwa.edu.au	[7]
Nucleus	Nuclear Localization Signal Database (NLSdb)	http://cubic.bioc.columbia.edu/db/NLSdb	[46]
Nucleolus	Arabidopsis Nucleolar Protein Database (AtNoPDB)	http://bioinf.scri.sari.ac.uk/cgi-bin/atnopdb/home	[50]
Vacuole	-	-	[42–44]
Cell wall	-	-	[55–57]
Plasma membrane	PlantsP (contains data on plasma membrane phosphoproteins)	http://plantsp.sdsc.edu	[59,60]
Endoplasmic reticulum	-	-	[24,62]
Golgi apparatus	-	-	[24,62]
Peroxisome	-	-	[61]
Multiple organelles	Organelle database (includes multiple species) Rice Proteome Database	http://organelledb.lsi.umich.edu http://gene64.dna.affrc.go.jp/RPD/main_en.html	[91] [52,53]

of protein partitioning to, and within, chloroplasts. For example, examinations of pea, spinach and *Arabidopsis* chloroplasts identified many lumen proteins that have a LTP containing a twin-arginine motif that is required for translocation into the thylakoid space via the so-called TAT pathway [26,27]. Recent studies have also specifically touched on the localization of chloroplast membrane proteins and, as expected, most chloroplast integral membrane proteins are predicted to be highly hydrophobic and to contain multiple transmembrane domains [28]. In a separate study, Gomez and colleagues used MS and previous reports to compile experimentally determined cTP cleavage sites for 58 thylakoid integral membrane proteins, and then compared these results with localization program predictions [29]. While some prediction programs, such as TargetP and ChloroP, correctly predicted chloroplast localization for most of these proteins, the actual transit peptide cleavage sites were poorly identified. Interestingly, another large-scale study found that many proteins identified from isolated whole chloroplasts were not predicted to be localized to plastids by current prediction programs [10]. This phenomenon and current issues relating to problems with prediction programs are discussed in detail in [9].

Utilizing another approach, Sun and colleagues examined the physiochemical characteristics of over 800 curated and predicted proteins localized to subdomains of chloroplasts [30]. The authors found that, by total amino acid number, integral inner envelope proteins were an average of two-times larger than luminal proteins and 75% larger than thylakoid membrane proteins. This study also found that integral thylakoid membrane proteins are usually acidic (median isoelectric point [pI] of 5.22), whereas integral inner envelope proteins are highly basic (median pI of 8.7) [30]. Finally, these authors found that the cTPs of thylakoid membrane proteins were 10–13% shorter than integral membrane proteins of the inner membrane and that thylakoid proteins contain much lower cysteine content than those of the inner envelope membrane. Each of these characteristics has been proposed for use in improvement of localization prediction programs.

Several studies have also employed the use of comparative proteomics in wild-type versus mutant lines to understand plastid import machinery [31,32]. For example, Kubis and colleagues found that isoforms (Toc132 and Toc120) of the Toc159 subunit of the Toc protein import complex were apparently specific for import of nonphotosynthetic machinery [32]. It is likely that many reports on proteomics of other organelle transport mutants will follow in the near future.

Importantly, data compiled from multiple proteomic and bioinformatic examinations of plastids can be found at the Plastid Proteome Database (PPDB) [30,33,101]. This searchable database includes data sets from the laboratory of van Wijk as well as numerous external published studies. One point of interest is that, at the time of writing this review, a total of only approximately 600 different *Arabidopsis* plastid proteins are listed in the PPDB, which – although perhaps not comprehensive for all identified proteins from published reports – is far short of the expected total of approximately 3000–4000. It appears that examinations into the remainder of the expected plastid proteome are ripe areas for future studies and should aid in understanding protein trafficking to and within plastids.

Mitochondria

Akin to chloroplasts, mitochondria are also relatively easy to purify. Being a key provider of energy and many primary metabolites, plant mitochondria have been extensively studied via proteomics methods. From the outside in, these organelles contain an outer membrane, a small intermembrane space, an

inner membrane and the soluble matrix [12]. In an attempt to consolidate, and form a coherent picture of all reported mitochondrial proteins, Heazlewood and Millar have developed the Arabidopsis Mitochondrial Protein Database (AMPDB) [7,102]. This searchable database contains data from 19 published proteomics studies on purified mitochondria. To date, over 400 mitochondrial proteins have been verified, and only approximately 70 of these proteins have not been linked to known or putative functions [34].

Proteins localized to plant mitochondria either contain a highly basic, cleavable N-terminal peptide or have internal signals that remain as part of the translocated protein [35]. Translocase of the mitochondrial outer and inner membrane (TOM and TIM, respectively) complexes generally translocate both types of protein across the outer and inner membranes, respectively (reviewed in [35]). While protein import machinery has been extensively studied, comprehensive knowledge of the protein content of mitochondria has lagged behind. A typical program will predict 1500–3000 mitochondrial proteins encoded by the *Arabidopsis* nuclear genome [21]. This appears to be in line with the proposed total of 3000 mitochondrial proteins [36], but unfortunately, it is difficult to build a consensus between programs. For example, TargetP identifies 3182 putative mitochondrial proteins in the *Arabidopsis* genome; however, a consensus of only 1176 proteins is found between TargetP, MitoProtII and iPSORT [7]. Even so, individual prediction programs only identified approximately 40–70% of the experimentally verified mitochondrial proteins. As Heazlewood and colleagues point out, this ambiguity should cause researchers to take care not to read too much into the results from prediction programs [37]. Importantly, having a large number of verified mitochondrial proteins allows a direct comparison and validation of targeting prediction programs. Similar to plastid studies, more in-depth analyses of mitochondria proteomes to identify additional proteins are probably necessary to improve targeting predictions.

Vacuoles

Proteins destined for plant vacuoles are trafficked through a minimum of two distinct pathways [38,39]. Complicating matters is the fact that multiple types of vacuoles exist in plant cells (e.g., protein storage, lytic and central vacuoles) [15,38]. Specifically, soluble proteins trafficked to vacuoles often contain cleavable N- or C-terminal propeptides (NTPP and CTPP, respectively) that are transported via distinct mechanisms [39], although some vacuolar proteins contain currently largely uncharacterized internal sorting determinants [39–41].

Several examinations of the *Arabidopsis* vacuole proteome have been reported in recent years [42–44]. For example, Carter and colleagues identified over 400 soluble and membrane proteins in central vacuoles derived from *Arabidopsis* leaf tissue [44]. These studies found a number of soluble proteins containing putative NTPP and CTPP sorting signals (51 in total); however, many soluble proteins (~90) not containing these sorting determinants were also identified. These included nine proteins containing putative ER localization signals (e.g., KDEL and HDEL). At first glance, these proteins may appear to be contaminants, but it has been reported that ER retention signals can efficiently target some proteins to vacuoles (e.g., [45]). The remainder of proteins not containing known sorting signals likely represents vacuolar residents transported by other mechanisms. To identify other previously unidentified sorting signals, the authors subjected these proteins to numerous bioinformatics analyses; however, no potential signals emerged. Thus, it is probable that other complex sequences or post-translational modifications serve as signals for vacuolar sorting. Of course, the possibility that certain proteins may represent contamination from other compartments cannot be overlooked. The fact that the vacuole serves as a recycling compartment for the cell may complicate studies of this organelle.

As mentioned earlier, SNARE proteins are key mediators of membrane fusion in eukaryotes. In general, a single SNARE protein on a vesicular membrane interacts with three SNAREs on a target membrane to form a SNARE–pin complex, with each SNARE representing one of four basic classes of SNAREs [15]. In a tonoplast-derived (vacuole membrane) fraction, Carter and colleagues identified only four SNARE proteins, with each representing one of the four classes of SNAREs required for membrane fusion [44]. Additionally, accessory proteins of the SNARE–pin complex (e.g., Rabs, *N*-ethylmaleimide-sensitive factor [NSF] and soluble NSF attachment protein [SNAP]) were identified at the tonoplast. Overall, these findings suggested that the major proteins mediating the formation of the tonoplast SNARE–pin complex and membrane fusion at the vacuole were identified. Detailed analyses of other endomembrane components in a similar manner may provide testable hypotheses for understanding membrane trafficking in plant cells.

Nucleus

Plant NLSs are similar to those of other eukaryotes, usually consisting of a series of basic residues preceded by a helix-breaking side chain in monopartite signals or two groups of basic residues separated by 9–12 amino acids in bipartite signals [46]. Surprisingly, relatively few reports exist describing proteomic analyses of plant nuclei [47–51]. A total of 184 proteins were identified in an examination of the *Arabidopsis* nuclear proteome [47], and Pendle and colleagues identified 217 proteins in nucleoli derived from *Arabidopsis* cell culture [49]. These nucleolar proteins have since been deposited in the Arabidopsis Nucleolar Protein Database (AtNoPDB) [50,103]. It should also be noted that a recent analysis of the rice nuclear proteome identified 190 proteins [51], and data concerning this project as well as other organelles can be found at the Rice Proteome Database [52,53,104].

Cytoskeleton

While the cytoskeleton is not an organelle, subcellular organization and trafficking of cargo throughout plant cells is dependent upon it functioning correctly. Perhaps the most important part of the cytoskeleton in this regard for a plant cell

is its network of microtubules. To address this phenomenon, Chuong and colleagues identified 122 tubulin-binding proteins derived from *Arabidopsis* cell-suspension cultures [54]. These proteins fall into numerous functional categories including translation factors, RNA-binding proteins, signaling proteins and metabolic enzymes. This study also strongly supports roles for microtubule networks in the delivery of certain metabolic enzymes to the peroxisome and the movement of mRNAs to various cellular locations for later translation [54].

Other studies

The aforementioned studies are just a few examples of proteomic analyses on specific cellular locations. Other studies have also included the in-depth examinations of the cell wall (e.g., [55–57]), plasma membrane (e.g., [58–60]) and peroxisomes (e.g., [61]), as well as the ER and Golgi apparatus (e.g., [24,62]). Each of these studies will likely aid in the future understanding of protein trafficking in plant cells. Surprisingly, most of these studies – and those described earlier – also found unexpected proteins that would seemingly represent either contamination or mislocalization. Which of these possibilities is the case, or whether these identifications represent the true localization of these proteins remains to be determined. New methods and approaches to overcome this ambiguity are addressed below.

New methods

Historically, the most important step when undertaking any proteomic study of a given organelle is to obtain extremely pure starting samples. To date, complete proteomics of a single cell cannot be performed. However, most individual subcellular compartments can be readily purified – or at least enriched – by various means, which reduces the complexity of proteins in a starting sample and can aid in the identification of low-abundance proteins. The large-scale identification of proteins in a purified organelle can also give important clues as to how proteins are sorted to an organelle.

Identification of protein sorting pathways in nonpurified samples

Modern MS techniques are becoming exceedingly sensitive, with some systems routinely reaching a detection limit of 10 fmol for protein samples [63]. Conventionally, if pure organelles are not obtained, even minor contaminants may lead to the misinterpretation of a protein as being a resident of an organelle. Thus, in general, traditional proteomics approaches are dependent on obtaining extremely pure samples. However, it can be a very difficult task for experimentalists to obtain such pure samples. New techniques are available for organelle proteome profiling when highly pure samples cannot be obtained. For example, the ultimate subcellular localization of membrane proteins has been particularly difficult to predict. In particular, no specific motifs have been identified in membrane proteins residing in various organelles of the endomembrane system. As an initial step to identify these sorting determinants, Dunkley and colleagues recently described a new method termed localization of organelle proteins by isotope tagging (LOPIT) [24].

Localization of organelle proteins by isotope tagging/isotope-coded affinity tag

Quantitative analysis is an essential component of proteomics. It can be used not only for comparison of protein abundance between different samples but also for determination of relative levels of post-translational modifications [64], specificity in protein complexes [65] and dynamic proteomes within a particular cellular compartment [66]. Quantitative proteomics generally employs a stable isotope-labeling strategy, during which all cellular proteins are mass tagged through either *in vivo* metabolic labeling [67] or *in vitro* chemical derivation [68]. The mixture of equal amounts of labeled and control protein samples, followed by proteolysis treatment, results in characteristic isotopic pairs for individual peptide ions based upon labeling methods and peptide amino acid contents [69]. The relative spectral peak intensity of these isotopic pairs can be used for quantitative analyses. Isotope-coded affinity tagging (ICAT) is one of such strategies commonly used in current quantitative proteomics of diverse organisms and tissue/cell types [70–72].

An interesting extension of quantitative ICAT techniques in understanding protein cellular localization in plants has recently been demonstrated by Lilley and coworkers [24]. In studies of organelle proteomes, one challenge is that organelle cross-contamination often compromises the confidence in the assignment of individual proteins to a specific organelle, particularly for those proteins whose cellular functions and localization are not known. Since approximately one-third of the total predicted *Arabidopsis* proteins do not have known functions, accurate and highly confident assignment of these proteins to a particular cellular compartment(s) will be very significant in plant biology in understanding potential biologic functions of these proteins. This challenge is even more profound in studies of plant cellular trafficking pathways involving the endomembrane system. Different compartments of the endomembrane system such as the Golgi and ER have very similar physical density that prevent complete purification of each compartment type by gradient centrifugation. To tackle this challenge, the Lilley group developed the LOPIT strategy by employing the quantitative ICAT approach to characterize both Golgi and ER integral membrane proteins from *Arabidopsis* [24]. In their study, iodixanol gradient centrifugation was used to achieve partial purification after crude preparation of a cellular membrane fraction. The relatively enriched Golgi and ER fractions after gradient centrifugation were determined by western blotting of protein markers for either compartment. The fraction enriched in Golgi was labeled with a heavy ICAT tag, and the fraction enriched in ER was labeled with a light ICAT tag. Then the two fractions were mixed together before trypsin treatment followed by affinity purification of tagged peptides, and liquid chromatography (LC) tandem MS (MS/MS) analyses for protein identification and quantification. The analyses can be performed across all gradient fractions in a pair-wise manner. By comparing known protein markers for Golgi or ER, proteins with high heavy/light ICAT ratio were assigned to Golgi and proteins with low heavy/light

ICAT ratio were assigned to ER. For all 28 proteins with known or predicted localization, the LOPIT strategy was able to assign them correctly to Golgi, ER or mitochondria/plastid contaminants. Presumably, this strategy can be scaled up for studies of the entire proteome of a given organelle without the need for complete purification.

Like any other isotopic labeling methods, the LOPIT strategy using ICAT is also limited to a small pool of peptides, the cysteine-containing peptides only. The low codon frequency of cysteine in protein sequence may be a limiting factor for characterization of low-abundance proteins. A recent development of using the small mass reporters, the iTRAQ™ technology, is an excellent addition to the family of isotopic labeling methods [73]. It can be used for all tryptic peptides and, therefore, should significantly improve sequence coverage compared with the ICAT method. MS instruments capable of efficiently scanning low-mass ions are needed in order to employ the iTRAQ technology. Alternatively, other recent studies have also shown that quantitative analyses can be performed with LC/MS without the need for isotopic labeling [74–76]. This direct quantitation by peptide ion signals in LC/MS can apply to entire peptide pools generated from protein digestion with high sequence coverage, and eliminate the need of chemical derivation involved in isotopic labeling. It can be an alternative or complementary approach to ICAT-based strategies.

Coupling of chemical genomics & proteomics

Chemical genomics is an emerging technology that utilizes small organic compounds to perturb or disrupt normal biologic functions [77,78]. In particular, these chemicals can be used to specifically regulate the onset of mutant phenotypes. Several groups have demonstrated the effective use of chemical genomics in plant systems [77,79,80], which can be used as an alternative to conditional mutants. More specifically, Raikhel and colleagues have recently performed large-scale screens to identify compounds that have various affects on the *Arabidopsis* endomembrane system [79,80]. In particular, compounds (known as sorting inhibitors [sortins]) were identified that cause a vacuolar protein, carboxypeptidase Y (CPY), to be secreted [79]. It is thought that CPY is normally sorted to the plant vacuole via a cleavable NTPP (discussed earlier). In addition to CPY, numerous other proteins are specifically secreted in the presence of Sortin1 [UNPUBLISHED DATA]. While multiple pathways exist for protein sorting to the plant vacuole, the receptors and specific steps of these pathways remain essentially unknown. Sortin1 appears to specifically affect only the NTPP trafficking pathway to the vacuole [RAIKHEL & COWORKERS, UNPUBLISHED DATA], and proteomics approaches are now being utilized to identify which proteins are secreted in the presence of Sortin1. Importantly, this approach does not require the isolation of highly purified compartments as only the supernatant fraction of cell cultures need to be subjected to proteomics analyses. This should result in less ambiguous results concerning potential contaminants when studying the endomembrane system in particular.

Top-down mass spectrometry

Currently, the most common approaches to studying plant organelle proteomes utilize bottom-up methodologies, where the fractionated proteins are proteolysed by an amino acid-specific protease such as trypsin that cleaves at the sites C-terminal to arginine or lysine. A set of tryptic peptides is generated to match their mass values to a theoretical genome database to identify proteins in a given sample. Since many proteins' tryptic peptides may have the same or very close mass values (mass degeneracy), this so-called peptide mass fingerprinting (PMF) method may lack specificity in protein identification. Due to the availability of high-end MS instruments that can perform MS/MS scans, it now becomes almost routine that tryptic peptides are selected individually for MS/MS scans to determine their sequences that give high specificity in protein identification [81]. With both methods, cellular proteins are characterized only at the peptide level, either by multiple or even single peptides. The complete primary structure of a native cellular protein requires 100% sequence coverage of tryptic peptides through the entire process of sample preparation and MS analyses, which is improbable but not impossible. Routine bottom-up methods usually provide 15–40% protein sequence coverage, depending on sample complexity [26,82]. Therefore, the majority of information regarding a particular protein molecule is missing in bottom-up proteomics. Since native cellular proteins may be very different from genome-predicted precursors (exact protein products can be manipulated by multiple cellular mechanisms either pre- or post-translationally), it is critical to proteomics studies that intact proteins should be characterized to reveal exact amino acid sequences and molecular structures of cellular proteins in general or in different developmental stages and stress conditions.

Recent improvement in MS technology has allowed characterization of intact native cellular proteins with a top-down method. It has been successfully applied to a study of *Arabidopsis* chloroplast proteome. In this method, a high-resolution (>100,000) and mass accuracy (~1 ppm) MS instrument such as Fourier transform-ion cyclotron resonance (FT-ICR)-MS is generally required to resolve isotopic peaks of protein ion peak clusters and to correctly determine charge states of intact protein ions generated by an electrospray ionization source. When the measured mass of a cellular protein is different from that predicted by genome, the discrepancy can be mapped to an exact site or region of the protein by sequential fragmentation of the protein ions in MS/MS and MS/MS/MS. Accurate mapping of the mass discrepancy can lead to discovery of events in DNA sequencing error, amino acid substitution due to natural mutations, RNA editing and alternative splicing, signal peptide cleavage and, probably most interestingly, post-translational modifications that regulate protein functions in cells. With this top-down method, McLafferty and coworkers have studied 22 proteins fractionated by size-exclusion chromatography (SEC) from chloroplast thylakoid peripheral, thylakoid lumen and stroma fractions [83]. Accurate mass measurement of these proteins revealed clear mass discrepancies for several proteins

compared with the values predicted from the genome sequence. Further characterization of these proteins with sequential fragmentation of MS/MS and MS/MS/MS data discovered correct signal peptide cleavage sites for two proteins (At4g21280 and At4g05180; *in silico* prediction was incorrect), suggesting that the signal peptide prediction program used may not be reliable and needs to be experimentally confirmed. Also, an unusual N-terminal methylation to the first methionine of the rubisco small subunits (At5g38410 and At1g67090) was discovered by the same top-down analyses. However, these differences between mature cellular proteins and their corresponding precursors based on genome prediction were not found by the bottom-up method. The clear advantage of the top-down over the bottom-up method for extensive characterization of molecular structure of native cellular proteins should be important to the studies of cell trafficking systems. Many protein precursors are heavily processed and/or modified when entering the endomembrane system for correct trafficking and translocation. In-depth characterization of these proteins using the top-down method should be a powerful approach to the understanding of cellular mechanisms in the process of endomembrane trafficking.

Although very promising, it is clear that the top-down method is still in its early stage of development. It must be understood that, to date, no single MS approach is sufficient for characterization of all cellular proteins of an organelle proteome. Different methods are generally complementary to each other and can maximize coverage of an entire proteome in study (e.g., [33,44]). This likely also holds true between top-down and bottom-up approaches. An obvious obstacle in employing the top-down method is the low efficiency of ionization/detection and fragmentation for intact protein ions compared with the peptide ions. As evidenced in the case study of *Arabidopsis* chloroplast, the top-down method found only 22 protein ions while the bottom-up identified 97 proteins [83]. Large proteins of approximately 45 kDa or larger were not detected at all by the top-down method, although they appeared to be most abundant in the samples [83]. Alternative fragmentation such as electron capture dissociation (ECD) may be complementary to the commonly used collision-activated dissociation (CAD) fragmentation to reveal more comprehensive information for individual cellular proteins [84,85]. Apparently, further refining in these technical areas will be critical in order to make the top-down method a routine approach in proteomics.

Compartment-specific markers to aid in the purification of organelles

As aforementioned, the functions of approximately 35% of the genes in the *Arabidopsis* genome are unknown [3]. To determine their localization, a large-scale effort to tag unknown proteins with fluorescent markers (GFP, yellow fluorescent protein [YFP] and so on) has been undertaken [16,17]. Future studies, including the use of fluorescently tagged organelle-specific marker protein lines, will provide more detailed analyses of the subcellular distribution of the tagged unknown proteins using standard colocalization and fluorescence resonance energy transfer (FRET) approaches. To address this need, a large number of *Arabidopsis* lines expressing known compartment-specific proteins fused with fluorescent proteins are being developed [16,17]. With the purity of starting samples often being of crucial importance, these marker lines will likely provide useful tags when attempting to obtain highly purified organelles for proteomics studies. Moreover, transgenic plant lines tagged with multiple fluorescent markers (e.g., Golgi labeled with GFP, ER with YFP, vacuoles with cyan fluorescent protein, and so on) should provide a useful alternative to antibodies or enzymatic markers in preparing pure samples for proteomics. For example, having compartments that behave similarly during purification labeled with different fluorescent proteins would allow easy determination of purity by following these visible markers.

Expert commentary & five-year view

Large-scale systematic investigations of plant proteomes have only recently emerged and will certainly continue to grow for years to come. In the near future, research focus will likely turn to quantitative and comparative analyses of proteomes, and characterization of phosphoproteomes and cellular protein complexes. To understand cellular mechanisms of trafficking pathways, these proteomics analyses will likely be conducted in the context of chemical treatments or mutant backgrounds that either enhance or inhibit trafficking. A recently reported cellular assay should be a good method to screen for these chemicals or mutants through chemical genomics or mutagenesis [86]. Subsequent proteomic analyses can then identify proteins and/or protein modifications that underline observed cellular trafficking activity. Combined with other systems biology approaches, proteomics will play a major role in dissecting cellular pathways of protein trafficking.

Technology improvement

The recent success in proteomics has been mainly due to the rapid advancement in genomics and MS technology. This is unlikely to change within 5 years. Indeed, MS-based approaches will be irreplaceable for the foreseeable future because it is suitable for studies of proteins of any kind. Due to increased demand, significant improvement in MS technology and system configuration will be expected. These will include quadrupole time-of-flight (qTOF), ion-trap and FT-ICR instruments that are commonly used in proteomics research. A novel instrumentation design introduced to qTOF systems (LC/MS^E scanning) now allows continuous data scanning with minimal information loss, which should be a significant improvement over the currently most common MS/MS methods that often result in some data loss in complex samples [76]. Therefore, the new scanning method should increase sequence coverage for identified proteins by the bottom-up approach, useful for both identification and quantification. Both ion-trap and FT-ICR have also seen improved instrumentation with higher resolution, mass accuracy and speed of analyses. New hybrid systems will be available for ultrahigh resolution and mass accuracy without a magnetic field configuration. This

will significantly reduce cost and simplify operation and maintenance. In fast-paced, high-throughput environments, TOF/TOF systems will see significant improvement over sensitivity and specificity of precursor ion selection for CAD fragmentation. In addition to CAD, alternative fragmentation methods for peptides and proteins are also available and will continue to improve [87,88]. The authors believe that all these improvements in MS technologies will have significant impacts in plant proteomics in coming years.

Protein fractionation or separation prior to MS analysis is an important step in sample preparation to reduce sample complexity. Gel electrophoresis-based protein separation has been the primary method used in current plant proteomics. The handling of gel samples has been proven tedious, time consuming and labor intensive and, therefore, is not cost effective. Furthermore, uncertainty in protein recovery from a gel and efficiency of in-gel digestion and peptide extraction provides additional motivation for new methodology. Nongel-based multidimensional chromatography separation for complex protein samples followed by in-solution digestion will probably offer a more cost-effective approach that can be appreciated by plant scientists in coming years. As an indication of such potential, a newly released 2D protein franctionation system has been successfully applied to fractionating complex protein mixtures isolated from human cancer cells, with a high percentage of yield recovery and convenient sample handling [89]. Importantly, this separation results in the collection of intact proteins, making both bottom-up and top-down analyses possible from the same sample. Further improvement of column surface chemistry is expected and will enhance the system's separation resolution and loading capacity. When coupled with such a powerful protein-separation method, future MS systems should be able to analyze over thousands of proteins in a high-throughput manner to allow plant biologists to have in-depth characterization of plant proteomes. The number of proteins involved in cellular trafficking does likely not exceed more than a few thousand in total and, therefore, a nearly saturated study of plant trafficking systems may be expected in the near future.

Similar to protein separation, the front-end peptide separation of LC/MS/MS analysis is another major target for improvement. Peptide separation by reverse-phase LC is equally, if not more, important than the MS instrument itself. LC configurations can greatly affect analysis time and signal intensity of individual peptides by flow rate and peak resolution. A new generation of ultra-performance (UP)LC will likely replace the current high performance LC instrumentation as a primary LC configuration coupled to mass spectrometers [76]. These new systems are capable of delivering a nonsplit nanoflow at ultrahigh pressure to make it possible to use smaller particle columns for better LC performance. The combination of improved column chemistry and system ability may increase sensitivity and speed of analysis with high reproducibility. For highly complex samples, UPLC can also be configured for 2D-LC separation. As a result, it should be superior to existing systems in analyses of complex samples and low-abundance proteins.

It is also clear that new algorithms need to be developed for improved prediction of protein localization. As previously mentioned, numerous reports have found that some prediction

Key issues

- Many successful large-scale proteomic studies on purified plant organelles have already been reported and have resulted in the identification of several previously unknown protein sorting determinants and pathway components.
- Advances in preparation techniques (e.g., localization of organelle proteins by isotope tagging) means that ultrapure organelles are not absolutely required for effective organellar proteomic studies.
- The coupling of techniques, such as chemical genomics, with proteomics can provide new approaches for identifying protein-sorting pathways.
- Plant lines expressing fluorescent- or epitope-tagged compartment-specific proteins are now available and can provide useful handles for the purification of specific organelles.
- Current protein localization prediction programs need to be refined to reduce the occurrence of incorrect *in silico*-derived localization assignments.
- It is apparent that some protein trafficking mechanisms in plants occur through complex post-transcriptional and post-translational modifications. Top-down identification of protein modifications should aid in the elucidation of protein-sorting pathways.
- A robotic method for high-throughput quantitative proteomics that is suitable for plants is essential for characterization of vacuolar proteomes with mutations in trafficking pathways.
- The commonly used isotopic labeling techniques appear to be challenging in plant proteomics. Several recent studies have pursued a new approach for quantitative proteomics without isotopic labeling. The most recently released ultra-performance liquid chromatography/quadropole time-of-flight system provides a technical platform for quantitative analyses without isotopic labeling, which may become a robotic system suitable for plant quantitative proteomic studies [76].

programs are not particularly reliable. It should be possible to develop or refine current prediction programs using the current large pools of verified organellar proteins identified through proteomic studies. Indeed, Borner and colleagues used proteomics to identify 30 glycosylphosphatidylinositol-anchored proteins (GAPs) in *Arabidopsis* and used the primary sequence of these GAPs to train a search algorithm [90]. This trained algorithm identified an additional 64 putative GAPs not found via previous *in silico* analyses. Furthermore, it is likely that many proteins contain unidentified motifs and/or post-translational modifications required for correct cellular sorting. Identification of these sorting determinants will probably depend on biochemical analyses of individual proteins; however, the use of top-down proteomic methods to obtain 100% protein coverage will be useful in identifying post-translational modifications responsible for protein trafficking.

Acknowledgements

The authors apologize to many authors for the omission of their references owing to restriction in the length of the review and its reference list. Research in the authors' laboratory is supported by grants from the National Science Foundation (MCB 0296080 to Natasha V Raikhel) and the Department of Energy, Division of Energy Biosciences (DE-FG03-02ER15295/A00 to Natasha V Raikhel). Songqin Pan and Clay J Carter contributed equally to the manuscript.

References

Papers of special note have been highlighted as:
- • of interest
- •• of considerable interest

1. Sanderfoot AA, Pilgrim M, Adam L, Raikhel NV. Disruption of individual members of *Arabidopsis syntaxin* gene families indicates each has essential functions. *Plant Cell* 13(3), 659–666 (2001).
2. Collins NC, Thordal-Christensen H, Lipka V *et al.* SNARE-protein-mediated disease resistance at the plant cell wall. *Nature* 425(6961), 973–977 (2003).
3. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana. Nature* 408(6814), 796–815 (2000).
4. Goff SA, Ricke D, Lan TH *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* 296(5565), 92–100 (2002).
5. Yu J, Hu S, Wang J *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296(5565), 79–92 (2002).
6. Brunner AM, Busov VB, Strauss SH. Poplar genome sequence: functional genomics in an ecologically dominant plant species. *Trends Plant Sci.* 9(1), 49–56 (2004).
7. Heazlewood JL, Millar AH. AMPDB: the Arabidopsis Mitochondrial Protein Database. *Nucleic Acids Res.* 33(Database issue), D605–D610 (2005).
   •• **Report of a searchable database containing comprehensive information from 19 proteomic studies on purified mitochondria.**
8. Nair R, Rost B. Mimicking cellular sorting improves prediction of subcellular localization. *J. Mol. Biol.* 348(1), 85–100 (2005).
9. van Wijk KJ. Plastid proteomics. *Plant Physiol. Biochem.* 42(12), 963–977 (2004).
10. Kleffmann T, Russenberger D, von Zychlinski A *et al.* The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. *Curr. Biol.* 14(5), 354–362 (2004).
11. Peeters N, Small I. Dual targeting to mitochondria and chloroplasts. *Biochim. Biophys. Acta* 1541(1–2), 54–63 (2001).
12. *Biochemistry & Molecular Biology of Plants.* Buchanan BB, Gruissem W, Jones RL (Eds), American Society of Plant Physiologists, MA, USA, (2000).
13. Sanderfoot AA, Raikhel NV. Vesicle trafficking. In: *The Secretory System of Arabidopsis, in The Arabidopsis Book.* Somerville CR, Meyerowitz EM (Eds), American Society of Plant Biologists, MD, USA, 24 (2003).
14. Jarvis P, Robinson C. Mechanisms of protein import and routing in chloroplasts. *Curr. Biol.* 14(24), R1064–R1077 (2004).
15. Jurgens G. Membrane trafficking in plants. *Ann. Rev. Cell Dev. Biol.* 20, 481–504 (2004).
16. Tian GW, Mohanty A, Chary SN *et al.* High-throughput fluorescent tagging of full-length *Arabidopsis* gene products in planta. *Plant Physiol.* 135(1), 25–38 (2004).
17. Koroleva OA, Tomlinson ML, Leader D, Shaw P, Doonan JH. High-throughput protein localization in *Arabidopsis* using Agrobacterium-mediated transient expression of GFP–ORF fusions. *Plant J.* 41(1), 162–74 (2005).
18. Kikuchi S, Satoh K, Nagata T *et al.* Collection, mapping, and annotation of over 28,000 cDNA clones from *Japonica* rice. *Science* 301(5631), 376–379 (2003).
19. Dunwell JM, Purvis A, Khuri S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry* 65(1), 7–17 (2004).
20. Dreger M. Subcellular proteomics. *Mass Spectrom. Rev.* 22(1), 27–56 (2003).
21. Millar AH. Location, location, location: surveying the intracellular real estate through proteomics in plants. *Funct. Plant Biol.* 31, 563–571 (2004).
22. Warnock DE, Fahy E, Taylor SW. Identification of protein associations in organelles, using mass spectrometry-based proteomics. *Mass Spectrom. Rev.* 23(4), 259–280 (2004).
23. Tanaka N, Fujita M, Handa H *et al.* Proteomics of the rice cell: systematic identification of the protein populations in subcellular compartments. *Mol. Genet. Genomics* 271(5), 566–576 (2004).
24. Dunkley TP, Watson R, Griffin JL, Dupree P, Lilley KS. Localization of organelle proteins by isotope tagging (LOPIT). *Mol. Cell. Proteomics* 3(11), 1128–1134 (2004).
   •• **Describes a unique application of protein tags to successfully perform proteomics on partially purified organelles.**
25. Keegstra K, Froehlich JE. Protein import into chloroplasts. *Curr. Opin. Plant Biol.* 2(6), 471–476 (1999).
26. Peltier JB, Friso G, Kalume DE *et al.* Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. *Plant Cell* 12(3), 319–341 (2000).
27. Schubert M, Petersson UA, Haas BJ *et al.* Proteome map of the chloroplast lumen of *Arabidopsis thaliana. J. Biol. Chem.* 277(10), 8354–8365 (2002).
28. Ferro M, Salvi D, Brugiere S *et al.* Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana. Mol. Cell. Proteomics* 2(5), 325–345 (2003).
29. Gomez SM, Bil KY, Aguilera R *et al.* Transit peptide cleavage sites of integral thylakoid membrane proteins. *Mol. Cell. Proteomics* 2(10), 1068–1085 (2003).

30 Sun Q, Emanuelsson O, van Wijk KJ. Analysis of curated and predicted plastid subproteomes of *Arabidopsis*. Subcellular compartmentalization leads to distinctive proteome properties. *Plant Physiol.* 135(2), 723–734 (2004).
- **Concise analysis of the plastid proteome with implications for protein trafficking.**

31 Kubis S, Baldwin A, Patel R *et al.* The *Arabidopsis* ppi1 mutant is specifically defective in the expression, chloroplast import, and accumulation of photosynthetic proteins. *Plant Cell* 15(8), 1859–1871 (2003).

32 Kubis S, Patel R, Combe J *et al.* Functional specialization amongst the *Arabidopsis* Toc159 family of chloroplast protein import receptors. *Plant Cell* 16(8), 2059–2077 (2004).

33 Friso G, Giacomelli L, Ytterberg AJ *et al.* In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. *Plant Cell* 16(2), 478–499 (2004).

34 Millar AH, Heazlewood JL, Kristensen BK, Braun HP, Moller IM. The plant mitochondrial proteome. *Trends Plant Sci.* 10(1), 36–43 (2005).

35 Wiedemann N, Frazier AE, Pfanner N. The protein import machinery of mitochondria. *J. Biol. Chem.* 279(15), 14473–14476 (2004).

36 Richly E, Chinnery PF, Leister D. Evolutionary diversification of mitochondrial proteomes: implications for human disease. *Trends Genet.* 19(7), 356–362 (2003).

37 Heazlewood JL, Tonti-Filippini JS, Gout AM *et al.* Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell* 16(1), 241–256 (2004).

38 Vitale A, Hinz G. Sorting of proteins to storage vacuoles: how many mechanisms? *Trends Plant Sci.* 10(7), 316–323 (2005).

39 Vitale A, Raikhel NV. What do proteins need to reach different vacuoles? *Trends Plant Sci.* 4(4), 149–155 (1999).

40 Matsuoka K, Neuhaus J-M. *Cis*-elements of protein transport to the vacuoles. *J. Exp. Bot.* 50, 165–174 (1999).

41 Rojo E, Zouhar J, Carter C, Kovaleva V, Raikhel NV. A unique mechanism for protein processing and degradation in *Arabidopsis thaliana. Proc. Natl Acad. Sci. USA* 100(12), 7389–7394 (2003).

42 Szponarski W, Sommerer N, Boyer JC, Rossignol M, Gibrat R. Large-scale characterization of integral proteins from *Arabidopsis* vacuolar membrane by two-dimensional liquid chromatography. *Proteomics* 4(2), 397–406 (2004).

43 Shimaoka T, Ohnishi M, Sazuka T *et al.* Isolation of intact vacuoles and proteomic analysis of tonoplast from suspension-cultured cells of *Arabidopsis thaliana. Plant Cell Physiol.* 45(6), 672–683 (2004).

44 Carter C, Pan S, Zouhar J *et al.* The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. *Plant Cell* 16(12), 3285–3303 (2004).
- **Used complementary proteomics approaches for deep coverage of the *Arabidopsis* vacuole proteome including low-abundant membrane proteins.**

45 Matsushima R, Kondo M, Nishimura M, Hara-Nishimura I. A novel ER-derived compartment, the ER body, selectively accumulates a β-glucosidase with an ER-retention signal in *Arabidopsis. Plant J.* 33(3), 493–502 (2003).

46 Nair R, Carter P, Rost B. NLSdb: database of nuclear localization signals. *Nucleic Acids Res.* 31(1), 397–399(2003).

47 Bae MS, Cho EJ, Choi EY, Park OK. Analysis of the *Arabidopsis* nuclear proteome and its response to cold stress. *Plant J.* 36(5), 652–663 (2003).

48 Calikowski TT, Meulia T, Meier I. A proteomic study of the *Arabidopsis* nuclear matrix. *J. Cell Biochem.* 90(2), 361–378 (2003).

49 Pendle AF, Clark GP, Boon R *et al.* Proteomic analysis of the *Arabidopsis* nucleolus suggests novel nucleolar functions. *Mol. Biol Cell* 16(1), 260–269 (2005).

50 Brown JW, Shaw PJ, Shaw P, Marshall DF. *Arabidopsis* nucleolar protein database (AtNoPDB). *Nucleic Acids Res.* 33(Database issue), D633–D636 (2005).

51 Khan MM, Komatsu S. Rice proteomics: recent developments and analysis of nuclear proteins. *Phytochemistry* 65(12), 1671–1681 (2004).

52 Komatsu S, Kojima K, Suzuki K, Ozaki K, Higo K. Rice Proteome Database based on two-dimensional polyacrylamide gel electrophoresis: its status in 2003. *Nucleic Acids Res.* 32(Database issue), D388–D392 (2004).

53 Komatsu S, Tanaka N. Rice proteome analysis: a step toward functional analysis of the rice genome. *Proteomics* 5(4), 938–949 (2005).

54 Chuong SD, Good AG, Taylor GJ *et al.* Large-scale identification of tubulin-binding proteins provides insight on subcellular trafficking, metabolic channeling, and signaling in plant cells. *Mol. Cell. Proteomics* 3(10), 970–983 (2004).

55 Chivasa S, Ndimba BK, Simon WJ *et al.* Proteomic analysis of the *Arabidopsis thaliana* cell wall. *Electrophoresis* 23(11), 1754–1765 (2002).

56 Boudart G, Jamet E, Rossignol M *et al.* Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: identification by mass spectrometry and bioinformatics. *Proteomics* 5(1), 212–221 (2005).

57 Slabas AR, Ndimba B, Simon WJ, Chivasa S. Proteomic analysis of the *Arabidopsis* cell wall reveals unexpected proteins with new cellular locations. *Biochem. Soc. Trans.* 32(Pt 3), 524–528 (2004).

58 Elortza F, Nuhse TS, Foster LJ *et al.* Proteomic analysis of glycosylphosphatidylinositol-anchored membrane proteins. *Mol. Cell. Proteomics* 2(12), 1261–1270 (2003).

59 Nuhse TS, Stensballe A, Jensen ON, Peck SC. Phosphoproteomics of the *Arabidopsis* plasma membrane and a new phosphorylation site database. *Plant Cell* 16(9), 2394–2405 (2004).
- **Large-scale study of phosphoproteomics of *Arabidopsis* plasma membrane using immobilized metal ion affinity chromatography and 2D liquid chromatography (LC)/tandem mass spectrometry (MS/MS) for phosphorylation mapping.**

60 Marmagne A, Rouet MA, Ferro M *et al.* Identification of new intrinsic proteins in *Arabidopsis* plasma membrane proteome. *Mol. Cell. Proteomics* 3(7), 675–691 (2004).

61 Fukao Y, Hayashi M, Nishimura M. Proteomic analysis of leaf peroxisomal proteins in greening cotyledons of *Arabidopsis thaliana. Plant Cell Physiol.* 43(7), 689–696 (2002).

62 Prime TA, Sherrier DJ, Mahon P, Packman LC, Dupree P. A proteomic analysis of organelles from *Arabidopsis thaliana. Electrophoresis* 21(16), 3488–3499 (2000).

63 Hachey DL, Chaurand P. Proteomics in reproductive medicine: the technology for separation and identification of proteins. *J. Reprod. Immunol.* 63(1), 61–73 (2004).

64 Hegeman AD, Harms AC, Sussman MR, Bunner AE, Harper JF. An isotope labeling strategy for quantifying the degree of phosphorylation at multiple sites in proteins. *J. Am. Soc. Mass Spectrom.* 15(5), 647–653 (2004).

65 Blagoev B, Kratchmarova I, Ong SE *et al.* A proteomics strategy to elucidate functional protein–protein interactions applied to EGF signaling. *Nature Biotechnol.* 21(3), 315–318 (2003).

66 Andersen JS, Lam YW, Leung AK *et al.* Nucleolar proteome dynamics. *Nature* 433(7021), 77–83 (2005).

67 Pan S, Gu S, Bradbury EM, Chen X. Single peptide-based protein identification in human proteome through MALDI-TOF MS coupled with amino acids coded mass tagging. *Anal. Chem.* 75(6), 1316–1324 (2003).

68 Gygi SP, Rist B, Gerber SA *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* 17(10), 994–999 (1999).

• **Describes development of an isotope-coded affinity tagging-based MS technology for protein identification and quantification with cysteine-containing peptides.**

69 Zhu H, Pan S, Gu S, Bradbury EM, Chen X. Amino acid residue specific stable isotope labeling for quantitative proteomics. *Rapid Commun. Mass Spectrom.* 16(22), 2115–2123 (2002).

70 Shiio Y, Donohoe S, Yi EC *et al.* Quantitative proteomic analysis of Myc oncoprotein function. *Embo. J.* 21(19), 5088–5096 (2002).

71 Prokai L, Zharikova AD, Stevens SM Jr. Effect of chronic morphine exposure on the synaptic plasma-membrane subproteome of rats: a quantitative protein profiling study based on isotope-coded affinity tags and liquid chromatography/mass spectrometry. *J. Mass Spectrom.* 40(2), 169–175 (2005).

72 Li C, Hong Y, Tan YX *et al.* Accurate qualitative and quantitative proteomic analysis of clinical hepatocellular carcinoma using laser capture microdissection coupled with isotope-coded affinity tag and two-dimensional liquid chromatography mass spectrometry. *Mol. Cell. Proteomics* 3(4), 399–409 (2004).

73 Ross PL, Huang YN, Marchese JN *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3(12), 1154–1169 (2004).

74 Rojo E, Martin R, Carter C *et al.* VPEγ exhibits a caspase-like activity that contributes to defense against pathogens. *Curr. Biol.* 14(21), 1897–1906 (2004).

75 Chelius D, Zhang T, Wang G, Shen RF. Global protein identification and quantification technology using two-dimensional liquid chromatography nanospray mass spectrometry. *Anal. Chem.* 75(23), 6658–6665 (2003).

•• **Describes a 2D-LC/MS/MS method for large-scale protein identification and quantification without the need for isotopic labeling.**

76 Silva JC, Denny R, Dorschel CA *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. *Anal. Chem.* 77(7), 2187–2200 (2005).

77 Zhao Y, Dai X, Blackwell HE, Schreiber SL, Chory J. SIR1, an upstream component in auxin signaling identified by chemical genetics. *Science* 301(5636), 1107–1110 (2003).

78 Blackwell HE, Zhao Y. Chemical genetic approaches to plant biology. *Plant Physiol.* 133(2), 448–455 (2003).

79 Zouhar J, Hicks GR, Raikhel NV. Sorting inhibitors (sortins): chemical compounds to study vacuolar sorting in *Arabidopsis. Proc. Natl Acad. Sci. USA* 101(25), 9497–9501 (2004).

80 Surpin M, Rojas-Pierce M, Carter C *et al.* The power of chemical genomics to study the link between endomembrane system components and the gravitropic response. *Proc. Natl Acad. Sci. USA* 102(13), 4902–4907 (2005).

81 Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 422(6928), 198–207 (2003).

•• **Comprehensive review on MS technology and its broad applications in functional proteomics.**

82 Andersen JS, Svensson B, Roepstorff P. Electrospray ionization and matrix assisted laser desorption/ionization mass spectrometry: powerful analytical tools in recombinant protein chemistry. *Nature Biotechnol.* 14(4), 449–457 (1996).

83 Zabrouskov V, Giacomelli L, Van Wijk KJ, McLafferty FW. A new approach for plant proteomics: characterization of chloroplast proteins of *Arabidopsis thaliana* by top-down mass spectrometry. *Mol. Cell. Proteomics* 2(12), 1253–1260 (2003).

84 Ge Y, Lawhorn BG, ElNaggar M *et al.* Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.* 124(4), 672–678 (2002).

85 Nielsen ML, Savitski MM, Zubarev RA. Improving protein identification using complementary fragmentation techniques in Fourier transform mass spectrometry. *Mol. Cell. Proteomics* 4(8), 1180–1188 (2005).

86 Kim JY, Rim Y, Wang J, Jackson D. A novel cell-to-cell trafficking assay indicates that the KNOX homeodomain is necessary and sufficient for intercellular protein and mRNA trafficking. *Genes Dev.* 19(7), 788–793 (2005).

87 Breuker K, Oh H, Lin C, Carpenter BK, McLafferty FW. Nonergodic and conformational control of the electron capture dissociation of protein cations. *Proc. Natl Acad. Sci. USA* 101(39), 14011–14016 (2004).

88 McFarland MA, Chalmers MJ, Quinn JP, Hendrickson CL, Marshall AG. Evaluation and optimization of electron capture dissociation efficiency in Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* 16(7), 1060–1066 (2005).

89 Wall DB, Kachman MT, Gong S *et al.* Isoelectric focusing nonporous RP HPLC: a two-dimensional liquid-phase separation method for mapping of cellular proteins with identification using MALDI-TOF mass spectrometry. *Anal. Chem.* 72(6), 1099–1111 (2000).

90 Borner GH, Lilley KS, Stevens TJ, Dupree P. Identification of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis.* A proteomic and genomic analysis. *Plant Physiol.* 132(2), 568–577 (2003).

91 Wiwatwattana N, Kumar A. Organelle DB: a cross-species database of protein localization and function. *Nucleic Acids Res.* 33(Database issue), D598–D604 (2005).

• **Introduction of a useful multiorganism organellar protein database.**

Websites

101 The Plastid Proteome Database http://ppdb.tc.cornell.edu (Viewed September 2005)

102 Arabidopsis Mitochondrial Protein Database www.ampdb.bcs.uwa.edu.au (Viewed September 2005)

103 Arabidopsis Nucleolar Protein Database http://bioinf.scri.sari.ac.uk/cgi-bin/atnopdb/home (Viewed September 2005)

104 Rice Proteome Database http://gene64.dna.affrc.go.jp/RPD/main_en.html (Viewed September 2005)

Affiliations

- *Songqin Pan, PhD*
  Academic Coordinator, WM Keck Proteomics Laboratory, Center for Plant Cell Biology, Botany & Plant Sciences, University of California, Riverside, CA 92521, USA
  Tel.: +1 951 827 7114
  Fax: +1 951 827 2155
  sqpan@ucr.edu
- *Clay J Carter, PhD*
  Assistant Professor, Department of Biology, University of Minnesota, Duluth, MN 55812, USA
  Tel.: +1 218 726 7347
  Fax: +1 218 726 8142
  cjcarter@d.umn.edu
- *Natasha V Raikhel, PhD*
  Distinguished Professor, Center for Plant Cell Biology, Botany & Plant Sciences, University of California, Riverside, CA 92521, USA
  Tel.: +1 951 827 6370
  Fax: +1 951 827 2155
  nraikhel@ucr.edu