Challenges and Prospects of Plant Proteomics[1]

Klaas J. van Wijk*

Department of Plant Biology, Emerson Building Third Floor, Cornell University, Ithaca, New York 14853

Proteomics has been defined as "the systematic analysis of the protein population in a tissue, cell, or subcellular compartment." Over the last 2 to 3 years, proteomics has generated a relatively large number of reviews on technical aspect and concepts. This reflects the promise and expectations of proteomics on one hand, and the need for investment in technology and expertise on the other. Proteomics is often associated with two-dimensional electrophoresis (2-DE) and "brute force" identification. Although two-dimensional gels can be informative, proteomics goes far beyond 2-DE gels and brute force identification, as will hopefully become evident from this paper. Plant proteomics is still in its infancy, but is likely to become an active field with a large impact on plant biology.

The first part of this *Update* will focus on general and more technological aspects of proteomics and cover issues such as protein separation, mass spectrometry (MS), bioinformatics, etc. These issues are as much relevant for plants as they are for research in other organisms. The second part of this *Update* focuses specifically on plant proteomics, and will address what has been achieved so far and discuss the questions and issues for plant biology. This will include the issue of homology-based searching with MS data (what can you do without a fully sequenced genome?), the use of expressed sequence tag (EST) databases, gene annotation, and proteomics resources for the plant research community.

Due to limited space, citations are limited to key references. At many locations in this *Update*, reference is made to a *Trends Guide on Proteomics*, edited by Matthias Mann and Walter Blackstock (Blackstock and Mann, 2000). This issue contains high-quality contributions from different experts on many aspects of proteomics, including MS, and provides an excellent starting point for further reading. References to plant proteomics papers have mostly been limited to contributions from the last 3 years; for earlier work, reference is made to an extensive review (Thiellement et al., 1999).

BIOLOGICAL QUESTIONS AND PRACTICAL APPROACHES

The array of proteomics applications varies from straightforward identification of proteins to characterization of posttranslational modifications and protein-protein interactions. These applications can be listed according to increasingly detailed characterizations of the proteome as follows:

(a) "Brute force" identification of proteomes and comparative proteomics with the aim to study differential protein expression. The most widespread techniques for soluble and peripheral membrane proteins is 2-DE, using immobilized pH gradient strips in the first dimension and SDS-PAGE in the second dimension (Fig. 1A). For identification of membrane proteins, chromatography and organic solvent fractionation are most appropriate (Fig. 1E).

(b) Identification of protein-protein interactions and multisubunit complexes using non-denaturing purification techniques, such as non-denaturing gel electrophoresis (e.g. blue-native gel electrophoresis) different types of chromatography, and immunoaffinity purifications (Fig. 1, B–D). Epitope tagging using transgenic organisms followed by affinity purification can be particularly successful to rapidly identify (low abundant) complexes out of total cell extracts.

(c) Analysis of posttranslational modifications, including phosphorylation, lipid-modification, glycosylation, processing, and proteolysis.

(d) Global structural determination of protein complexes, using limited proteolysis, possibly combined with cross-linking or isotope exchange (Bennett et al., 2000).

DEVELOPMENTS AND STRATEGIES IN BIOLOGICAL MS

It is fair to say that proteomics has been driven by rapid advances in biological MS and commercialization of MS equipment, together with increasing amounts of EST and genomic sequence data. The development of commercial, user friendly mass spectrometers with "soft" ionization techniques, such as matrix-assisted laser-desorption ionization (MALDI) and electrospray ionization (ESI) since the early 1990s, has opened the field of MS to "nonexperts." The improved mass accuracy, mass resolution, and sensitivity now allows the rapid identification of

[1] Funding was provided to K.J.v.W. by the Swedish National Research Council (NFR), by the Swedish Agricultural Research Council (SJFR), by the Swedish Strategic Funds (SSF), by the Carl Trygger Foundation, and by Cornell University.

* E-mail kv35@cornell.edu; fax 607–255–5407.

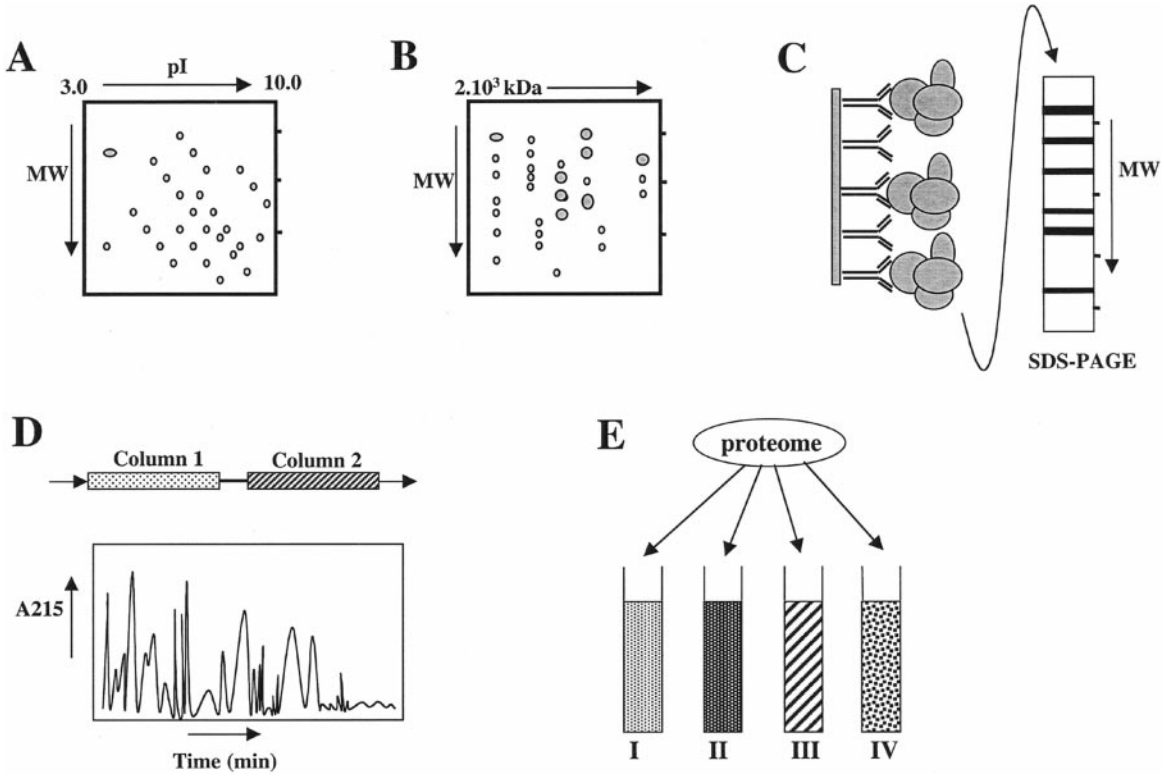

Figure 1. Different strategies for proteome purification and protein separation for identification by MS. A, Separation of individual proteins by fully denaturing 2-DE with immobilized pH gradient strips in the first dimension and SDS-PAGE in the second dimension. B, Separation of protein complexes by non-denaturing 2-DE (such as blue-native-PAGE), followed by SDS-PAGE in the second dimension. C, Purification of protein complexes by immuno-affinity chromatography, followed by SDS-PAGE. D, Multidimensional chromatography and chromatogram using the absorption at 215 nm for detection. E, Organic solvent fractionation for separation of complex protein mixtures of hydrobhobic membrane proteins. The numbers refer to different organic solvent mixtures; common solvents are aceton, isopropanol, chlorophorm, and methanol.

picomol-femtomole amounts of proteins and peptides if matching genomic sequence data are available. Examples of excellent reviews on MS and their application in biology are from Andersen and Mann (2000) and references in Blackstock and Mann (2000). Very useful practical tips and recipes are provided in Rowley et al. (2000).

As already detailed in many reviews, the typical strategy for rapid identification of large numbers of proteins (summarized in Fig. 2) is as follows:

(a) A selected protein separated by electrophoresis or chromatography is digested with a site-specific protease such as trypsin, resulting in a set of peptides. The masses of the peptides are then measured by MALDI-TOF MS, resulting in a list of peptide masses. For each entry in the nucleotide and protein databases, the masses of the predicted tryptic peptides are calculated and compared (within the experimental mass accuracy) with the list of measured peptide masses, using search engines. The correct protein will have a large number of "matching" peptides. Proteins can be identified out of a mixture of two to three different protein species. This method relies on the very high mass accuracy (0.001% for complex mixtures), mass resolution (10.000 full width half height), and sensitivity (femtomol range).

(b) In case a protein cannot be unambiguously identified by MALDI-TOF MS, peptide sequence tags are obtained by ESI tandem MS. The individual peptides are "screened" in a first section of the tandem mass spectrometer and selected peptides are subsequently fragmented along the protein backbone by collision with inert gas molecules (argon or nitrogen). This is termed collision-induced dissociation. The peptide fragments are then separated in a subsequent analyzer to provide amino acid sequence information. The peptide sequence tags and measured ion masses are used to search for proteins in the database employing specialized software. ESI is usually combined with a triple quadrupole analyzer, an ion trap analyzer or with a so-called hybrid of quadrupoles/hexapoles/TOF analyzer.

With the rapid increase in popularity of MS and proteomics, a fairly large assortment of instruments for different budgets and needs have become available (e.g. Micromass, Applied Biosystems, Bruker Daltronics, Sciex, Thermoquest, Kratos, Shimatzu, Pharmacia, and Biotech).

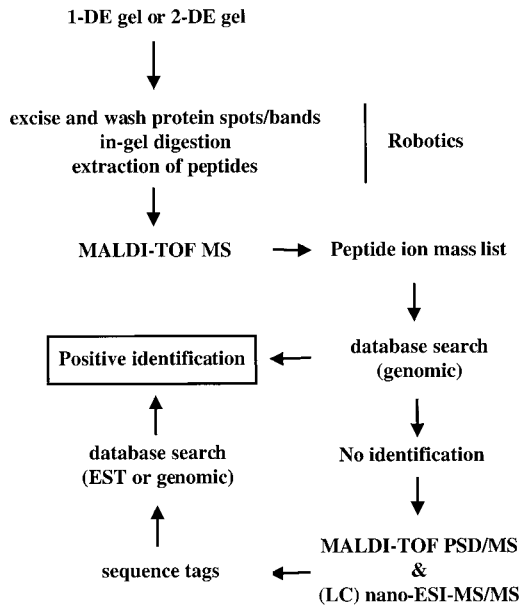

Figure 2. Typical flow chart for the analysis of proteomes by MS. Proteins are separated by 1-DE or 2-DE, after which they are visualized and selected for identification. The protein spots or bands are excised, SDS, stain, and salts are removed, and the proteins are digested by a site-specific protease that is usually trypsin. Peptides subsequently are extracted and a small fraction is used to determine the mono-isotopic peptide ion masses by MALDI-time-of-flight (TOF) MS. The experimental peptide ion masses are then searched against proteins predicted from genomic sequence data. If no positive identification can be achieved, (partial) sequences of the remaining peptides are determined by MALDI-TOF post-source decay (PSD)/MS or ESI-MS/MS. The sequence tags (partial sequence, together with the parent- and fragment ion masses) are used to search against EST or annotated genomic sequence data. In the case of Arabidopsis, this usually results in unambiguous identification.

NEW MASS SPECTROMETERS AND CONFIGURATIONS

Several studies have reported on a successful reversible coupling of a MALDI source to an ESI tandem quadrupole/TOF mass spectrometer (e.g. (Shevchenko et al., 2000). The possibility to routinely use two different ionization techniques (MALDI and ESI) in combination with MS/MS will increase the success rate of protein identification because each ionization method is generally quite selective for different peptides (see below). Using both ionization principles will improve the "coverage" of the proteins and therefore the success of identification. Moreover, MALDI is less sensitive to salts and detergents thus reducing the need for "sample cleanups." In addition, a prototype of a new tandem mass spectrometer (MALDI-TOF-TOF) was reported, possibly improving both speed and sensitivity (Quiniou et al., 2000).

PROTEIN AND PEPTIDE MODIFICATIONS FOR ENHANCED DETECTION AND IDENTIFICATION

A number of protein modifications or derivation have been developed to improve detection and/or identification of peptides in the mass spectrometer. Detection of proteins and peptides in the mass spectrometer requires ionization of the peptides. In general, only a very small fraction of the peptides in any given sample are entering the mass spectrometer (both in MALDI-TOF and ESI-MS) due to incomplete ionization at the inlet of the mass spectrometer. In addition, proteins and peptides are competing with each other during the ionization and extraction process, leading to further loss of specific peptide ions (a process named ion suppression). As an example, phosphorylated peptides in a complex peptide mixture are generally not observed in the mass spectrometer when measuring in (the standard) positive mode, due to the negative charge from the phosphogroup. Purification or enrichment of these phosphopeptides is required for detection, as discussed by Blackstock and Mann (2000). Alternatively, proteins and peptides can be modified to improve detection and identification. Examples of such modifications are Cys alkylation (Sechi and Chait, 1998) and conversion of Lys to homo-Arg for increased sensitivity of tryptic peptide detection by MALDI-TOF MS (Hale et al., 2000). Another more complex modification was reported to help determine the carboxyl termini of proteins (Sechi and Chait, 2000).

THE CENTRAL ROLE OF DATABASE SEARCH ENGINES AND BIOINFORMATICS

MS generally generates a large amount of numerical data and bioinformatics tools therefore are essential to match these MS data to protein, EST, and genome sequence databases. Most search engines have been developed in academic laboratories and some of those have now been commercialized. Examples of useful Web sites and their search engines are www.proteometrics.com, http://prospector.ucsf.edu/, http://195.41.108.38/PepSeaIntro.html, www.mann.embl-heidelberg.de/Services/PeptideSearch/PeptideSearchIntro.html, and www.matrix-science.com/; for a complete listing, see Rowley et al. (2000). Calibration and interpretation of MS spectra is generally carried out using commercial software developed by the vendors of the different mass spectrometers.

AUTOMATION

The process of spot selection, protein or peptide extraction, MS, and database searching can increasingly be automated, with a throughput of hundreds to thousands of samples per week (Blackstock and Mann, 2000). Automation is most advanced in larger pharmaceutical companies that have the financial resources to develop and build such automated facilities. However, recently many academic institutions are developing proteomics and MS facilities. This will provide many smaller academic laboratories with access to these tools, allowing the individual

researchers to focus on sample preparations and biological questions.

TECHNICAL CHALLENGES, POSSIBILITIES, AND PITFALLS OF PROTEOMICS

Proteomics is still in its infancy and many technical challenges and pitfalls remain. Four central issues of concern are listed below.

Dynamic Resolution

In many cases proteomes contain proteins of relative high and low abundance (low copy number). Molar ratios between abundant and rare proteins of more than 10.000 are not unusual. When highly selective purification methods are used, such as affinity purification, even the very low copy number proteins can be detected. However, if large proteomes consisting of thousands of proteins are analyzed, the dynamic resolution is generally limited and only the most abundant proteins can be detected. The dynamic resolution can be improved by fractionating a proteome into smaller sub-proteomes. In addition, complex proteomes can be analyzed more in-depth by a combination of separation techniques based on different principles, such as multidimensional chromatography (Blackstock and Mann, 2000). Direct coupling of chromatography to the mass spectrometer (online MS) is helpful to improve sensitivity in two ways: (a) through reduction of complexity during the ionization process and thus reducing ion suppression, and (b) drastically increasing peptide concentration during ionization.

Purification of Proteomes

To arrive at meaningful biological results, to address specific questions, and to enhance dynamic resolution, it is often important to obtain a pure (about 95%–99%) proteome. With the high sensitivity of MS, detection of contaminating proteins it is relatively easy. If no verification strategies are in place to recognize potential contaminants (e.g. via N-terminal signal sequences, alternative purification methods, or localization studies), it is possible that these contaminants will be incorrectly assigned to the purified proteome. Thus, even with (or possibly due to) the high sensitivity detection methods, a pure proteome is crucial to obtain meaningful results and reduce time-consuming verification procedures afterward.

Quantification

Many detection methods used in proteomics are not quantitative (such as MS), or only linear over a limited molar range (silver and Coomassie staining). This can often make it difficult to quantify stoichiometry between different proteins or quantitatively determine up- and down-regulation of protein expression. However, several techniques were developed to improve linearity or to introduce internal standards for normalization. Examples are the use of fluorescent dyes (e.g. Sypro Ruby; Patton, 2000) for the detection of protein on SDS-PAGE gels and the derivation and tagging of proteins prior to fractionation and detection by MS (e.g. isotope-coded affinity tags [Gygi et al., 1999]).

Separation, Visualization, and Identification of HydroPhobic Membrane Proteins

It is well known that hydrophobic membrane proteins are much more difficult to handle than hydrophilic proteins. Membrane proteins tend to easily aggregate and adsorb to the surface of tubes and vials. When working with nanomole to femtomole amounts this can lead to significant losses or complete "disappearance" of certain proteins or peptides. In addition, α-helicical transmembrane proteins do not resolve well or not at all on denaturing 2-DE gels, even when specific detergents and different mixtures of chaotropic agents (e.g. urea and thiourea) are used (Santoni et al., 1999). To achieve reproducible separation of α-helical membrane proteins, organic solvent fractionation (Ferro et al., 2000) or reverse phase HPLC is generally required.

PUBLISHED STUDIES IN PLANT PROTEOMICS

The field of plant proteomics is just at the very beginning and lags several years as compared with proteomics of unicellular prokaryotes and eukaryotes. Most of this delay can be explained by the early availability of complete genomic sequences of these unicellular organisms and the reduced complexity of their proteomes. With the completed sequence of the Arabidopis genome and increasing amounts of other plant genome and EST sequence data, it can be expected that plant proteomics will become a very active field.

In 1999, the first review of plant proteomics was published and extensively discussed the plant proteomics literature before 1999 (Thiellement et al., 1999). Most of these studies did not involve MS and therefore were limited to the comparison of expression levels without actual identification of the proteins. In a few cases, limited sets of proteins were identified through Edman sequencing. Many of these studies were focused on the use of 2-DE patterns to identify possible markers for different genotypes and phenotypes and phylogenetic relationships. In this *Update*, we will concentrate mostly on studies from the last 2 to 3 years and refer to Thiellement et al. (1999) for earlier work.

TWO-DIMENSIONAL MAPS OF DIFFERENT PLANT TISSUES

Tsugita and colleagues published several articles in which an attempt was made to map proteomes of different plant tissues from rice and Arabidopsis (Tsugita et al., 1996). The proteins of the different tissues were separated by 2-DE and a small number of abundant proteins were identified by Edman sequencing. At that time, and even today, the technology was/is not sufficient to obtain an analysis of the total proteomes of the different plant tissues at significant depth. Given the current level of technology, only a large-scale proteomics facility with true high-thoughput technology (using multidimensional chromatography and a large no. of state of the art mass spectrometers and robotics) would be able to obtain a more thorough characterization of the complete proteomes of the different tissues. However, there is a general consensus that such analysis is much more insightful after further sub-fractionation according to cell type and subcellular compartments.

In line with this consensus, several plant proteomics studies have been published recently that have focused on specific subcellular proteomes or protein complexes such as the plasma membrane, roots, mitochondria, and chloroplasts. In addition, a few interesting studies appeared concerning the symbiosis between roots of legumes and nitrogen-fixing bacteria. We will review these contributions briefly.

THE PLASMA MEMBRANE OF TOBACCO AND ARABIDOPSIS

Around 1996, a group of European scientists formed a European Union-supported consortium to study the proteome of the plasma membrane of tobacco and Arabidopsis. This resulted in a number of fairly methodological studies, the construction of 2-DE reference maps (Rouquie et al., 1997; Santoni et al., 1999), and some of the data appeared on a Web site (http://sphinx.rug.ac.be:8080/ppmdb/). A number of plasma membrane-specific proteins were identified and most of these papers highlighted the failure to use 2-DE gels for reproducible and complete mapping of membrane proteins.

ROOT PROTEOMICS OF MAIZE

The pattern of protein synthesis during hypoxic acclimation and anoxia in maize roots was analyzed by incorporation of ^{35}S-Met combined with 2-DE and MS (Chang et al., 2000). This work showed that protein synthesis during acclimation, but not during subsequent anoxia, is crucial for acclimation. This work shows some of the potential and difficulties of proteomics to study up-and down-regulation of protein expression.

CHLOROPLAST PROTEOMICS

A number of proteomics articles concerning subsets of chloroplast proteins have been published. The chloroplast is predicted to contain maximally 2,500 to 3,000 proteins expressed in a wide dynamic range; these studies demonstrate that further purification of these proteins into sub-proteomes and protein complexes is an effective strategy to obtain a more in depth insight (for discussion, see van Wijk, 2000).

Joyard and colleagues used organic solvent fractionation followed by SDS-PAGE to purify and identify integral chloroplast envelope proteins (Ferro et al., 2000). The propensity of hydrophobic proteins to partition in chloroform/methanol mixtures was directly correlated to the ratio between molecular mass and the number of putative transmembrane regions.

The proteins of the 30S and 50S ribosomal subunits in spinach chloroplasts were identified by a combination of 2-DE, chromatography, MS, and Edman sequencing (Yamaguchi and Subramanian, 2000; Yamaguchi et al., 2000). It was concluded that the spinach plastid ribosome comprises 59 proteins, of which 53 are *Escherichia coli* orthologues and six are non-ribosomal plastid-specific proteins (PSRP-1 to PSRP-6). A number of these proteins were shown to be posttranslationally modified. The authors proposed that the PSRPs evolved to perform functions unique to plastid translation and its regulation, including protein targeting/translocation to thylakoid membrane via the plastid 50S subunit. Even though several technical details concerning 2-DE were not described, this is a beautiful example of the strength of proteomics to thoroughly identify very large protein complexes from plants.

A 350-kD ClpP protease complex with 10 different subunits was identified in chloroplasts of Arabidopsis, using blue-native Gel electrophoresis, followed by MALDI-TOF and nano-ESI-MS/MS (Peltier et al., 2001). A new Clp protein was discovered in this complex (ClpS1) that does not belong to any of the known Clp genes families. Several truncations and errors in intron and exon prediction of the annotated Clp genes were corrected using MS data and by matching genomic sequences with cDNA sequences.

The soluble and peripheral proteins in the thylakoids of pea were systematically analyzed by using 2-DE, MS, N-terminal Edman sequencing, and bioinformatics (Peltier et al., 2000). After correcting to eliminate possible isoforms and posttranslational modifications, it was estimated that the thylakoid contains at least 200 to 230 different lumenal and peripheral proteins. Sixty-one proteins were identified, of which about 40% had no annotated function. It is surprising that about 50% of the identified lumenal proteins have a typical twin Argine motif in the lumenal transit peptide, suggesting that they translocate via the so-called TAT pathway, rather than via the Sec translocon (see Keegstra and Cline, 1999).

POSTTRANSLATIONAL MODIFICATIONS

Proteins in plants and other organisms undergo numerous posttranslational modifications, which help to regulate protein function and can alter protein localization. It is well known that several thylakoid proteins are reversible phosphorylated in response to environmental changes. Vener and colleagues used MS to analyze the reversible phosphorylation of surface-exposed hydrophylic loops of the more abundant thylakoid proteins under different physiological conditions (Vener et al., 2000). The reversible phosphorylation of a number of Photosystem II proteins was described. The combination of 2-DE with Triton X-114-based two-phase separation (Sherrier et al., 1999) showed that glycosylphosphatidylinositol-anchored proteins are a relatively abundant class of proteins at the plant plasma membrane and extracellular matrix. One of these was shown to be an arabinogalactan protein, a class of proteins known to be associated with cellular differentiation; theoretical analysis of two additional arabinogalactan protein-like proteins from Arabidopsis. Both papers show how well-designed purification procedures help to identify specific classes of posttranslational modifications.

SYMBIOTIC INTERACTION BETWEEN N-FIXING BACTERIA AND LEGUMES

The symbiosis between N-fixing bacteria and legumes results in formation of root nodules and is very important in agriculture. Two different groups reported protein expression studies in nitrogen fixing root nodules of soybean (Panter et al., 2000) and white sweet clover (*Melilotus alba*; Natera et al., 2000), respectively. Soybean peribacteroid membrane proteins were isolated from nitrogen-fixing root nodules and subjected to N-terminal sequencing. Sequence data from 17 putative peribacteroid membrane proteins were obtained. Six of these proteins are homologous to proteins of known function.

2-DE was also used to identify differentially expressed proteins during the symbiotic interaction between the bacterium *Sinorhizobium meliloti* strain 1021 and white sweet clover. The aim was to characterize novel symbiosis proteins and to determine how the two symbiotic partners alter their respective metabolisms as part of the interaction. Proteome maps from control white sweet clover roots, wild-type nodules, cultured *S. meliloti*, and *S. meliloti* bacteroids were generated and compared. Over 250 proteins were induced or up-regulated in the nodule, compared with the root, and over 350 proteins were down-regulated in the bacteroid form of the rhizobia, compared with cultured cells. N-terminal amino acid sequencing and MS were used to assign putative identity to nearly 100 nodule, bacterial, and bacteroid proteins. Bacteroid cells showed down-regulation of several proteins involved in nitrogen acquisition, indicating that the bacteroids were nitrogen proficient.

Both studies are excellent examples of the potential of proteomics in plant-symbiont interactions.

LIMITATIONS AND POSSIBILITIES OF PLANT PROTEOMICS WITH NON-SEQUENCED PLANT SPECIES

The identification and characterization of proteins by Edman sequencing and MS is greatly accelerated by the availability of genomic or EST sequences. The genome of Arabidopsis is the first plant genome that is (virtually) completely sequenced and publicly available (Initiative, 2000). Thus, from a technical perspective, Arabidopsis is clearly the most optimal organism for a proteomics approach. However, Arabidopsis is a dicotyledon, and is not an agricultural crop species, nor is it a legume or tree. Thus, although an excellent model system for the understanding of many basic biological processes, Arabidopsis has clear limitations. Therefore, the obvious question is to what extent are proteomics projects feasible with non-sequenced plant species. There are two options: (a) If EST sequences are available, proteins can be identified by sequence tags obtained by ESI-MS/MS or using MALDI-TOF PSD. However, this is strongly dependent on the size and the quality of the EST database. Typical shortcomings of EST databases are that many proteins are not or only poorly (by short ESTs only partly covering the protein) represented in EST databases and that the error rate in EST sequences can be quite significant. Peptide mass fingerprints alone are seldom used successfully to identify proteins from EST databases because ESTs are generally too short to obtain significant protein coverage and sufficient number of matching peptides (Rowley et al., 2000); or (b) homology-based searching, preferentially using (interpreted) sequences obtained from MS/MS, MALDI-TOF PSD, or Edman sequencing. Plant proteins generally share a significant amount of identity/homology between each other. This makes it feasible to search the database directly with ion masses generated by MS, but generally it is better to first translate the MS data into amino acid sequences and then use FASTA or similar search engines. Direct searches with peptide masses (peptide mass finger prints) are typically less useful because they require a 100% match between the experimental and homologous peptide sequences. MALDI-TOF MS alone is only successful when a very large amount of peptide masses are generated, thereby improving the chance to match several peptides to the homologous protein. This was demonstrated for pea chloroplast proteins (Peltier et al., 2000) and for maize proteins (Chang et al., 2000) searched against the complete plant database.

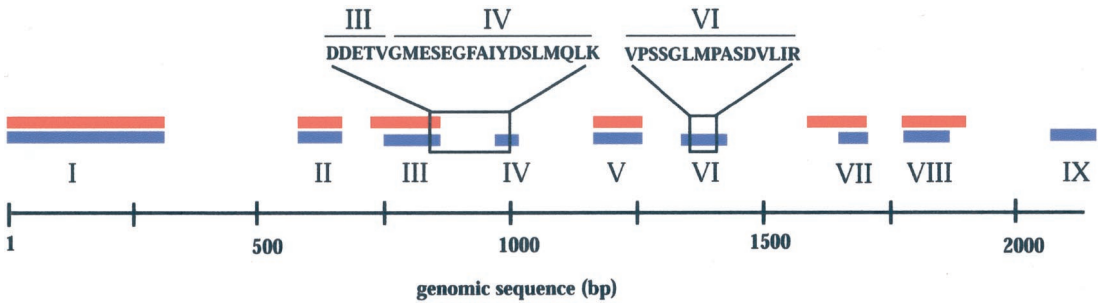

Figure 3. Example of typical errors in annotation of a gene in Arabidopsis. Correction of annotation of the Clp3 gene on chromosome I of Arabidopsis (accession no. 2342674). The boxes in red represent the annotated exons and the boxes in blue represent the correct exons. The misassigments were corrected by matching of the raw genomic sequence with overlapping ESTs (accession nos. 8684305, 8685348, 8686535, 8680144, 8734821, and others). Two matching protein sequences obtained by nano-ESI-MS/MS are indicated and confirm the new annotation.

ANNOTATION OF THE ARABIDOPSIS GENOME

The current generation of MS data search engines can search against protein sequence data, ESTs, and annotated genomic sequences. They are not able to search against the raw genomic sequence but instead have to rely on predicted protein sequences (but see Kuster et al., 1999), which would not matter if gene annotation was without errors. However, we observed that about 40% of proteins identified in the chloroplast of Arabidopsis have at least one significant misassignment. This became evident because a number of sequence tags obtained by MS/MS did not or only partially matched the predicted protein sequence. By systematic matching of ESTs against the genomic sequences we could arrive in many cases at the correct protein sequence (see also Mann and Pandey, 2001). An example for Arabidopsis is shown in Figure 3 (see also Peltier et al., 2001). Misassignment was found to affect up to 50% of the predicted protein sequence, varying from small to very large errors in intron-exon boundary prediction, to incorrect prediction of the start Met or erroneous prediction of the C terminus. Misassignment of the N terminus often affects as well the predicted localization (see Peltier et al., 2001). If the misassignment is very significant, the success rate of protein identification by MS will be diminished (but not if the raw sequence can be searched directly), and will also show as a large inconsistency between the experimental and theoretical mass and pI.

COMPLEMENTION OF PLANT PROTEOMICS WITH OTHER FUNCTIONAL GENOMICS TOOLS

In our view, proteomics will be most useful when combined with other functional genomics tools and approaches. A combination of microarray and proteomics analysis will indicate whether gene regulation is controlled at the level of transcription, or translation and protein accumulation. Protein function can be further studied by a combination of reverse and forward genetics and proteomics, as has already been demonstrated in yeast and *E. coli*.

About 40% of the predicted proteins in the Arabidopsis genome have no assigned function. Although reverse genetics will help to determine such functions, redundancy, lethality, and strong phenotypes can often prevent obtaining any insight in gene function. Most proteins have a transient or stable interaction with other proteins and the determination of these interacting proteins often can help to obtain more insight in gene function. Epitope tagging of transgenes, followed by affinity purification, will be very useful to identify (low abundant) protein complexes. For more high-throughput analysis, multidimensional purification schemes can possible be used for rapid identification of protein-protein interactions.

DEVELOPMENT OF PROTEOMICS RESOURCES FOR THE PLANT RESEARCH COMMUNITY

2-DE gel protein reference maps of sub-proteomes of different plant species are expected to become a central tool for organizing and understanding plant proteomes. A few Web sites with small, organized 2-DE databases are already available (http://sphinx.rug.ac.be:8080/ppmdb/index.html, http://www.biokemi.su.se/chloroplast/, and http://www.expasy.ch/ch2d/). Reference 2-DE maps will be used to follow differential protein expression and post-translational modifications.

As was mentioned above, proteomics is well suited to determine interaction between pairs of proteins but also to identify multisubunit complexes. A protein-protein interaction database for plant proteins could be a very useful tool for the plant research community.

Many universities and research institutes are investing in proteomics and MS facilities. Such facilities will undoubtedly be important to accelerate plant proteomics and will allow individual researchers to focus on sample preparation and biological question, rather than becoming MS experts themselves.

ACKNOWLEDGMENTS

I gratefully acknowledge Giulia Friso, Jean-Benoit Peltier, Jimmy Ytterberg, and Lisa Giacomelli for numerous stimulating discussions and for critically reading the manuscript. I apologize to many colleagues whose contributions could not be cited in this *Update* due to space restrictions.

Received April 2, 2001; accepted April 2, 2001.

LITERATURE CITED

Andersen JS, Mann M (2000) Functional genomics by mass spectrometry. FEBS Lett **480:** 25–31

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815

Bennett KL, Matthiesen T, Roepstorff P (2000) Probing protein surface topology by chemical surface labeling, crosslinking, and mass spectrometry. Methods Mol Biol **146:** 113–131

Blackstock W, Mann M, editors (2000) Proteomics: A *Trends* Guide. Elsevier, Amsterdam

Chang WW, Huang L, Shen M, Webster C, Burlingame AL, Roberts JK (2000) Patterns of protein synthesis and tolerance of anoxia in root tips of maize seedlings acclimated to a low-oxygen environment, and identification of proteins by mass spectrometry. Plant Physiol **122:** 295–318

Ferro M, Seigneurin-Berny D, Rolland N, Chapel A, Salvi D, Garin J, Joyard J (2000) Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. Electrophoresis **21:** 3517–3526

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol **17:** 994–999

Hale JE, Butler JP, Knierman MD, Becker GW (2000) Increased sensitivity of tryptic peptide detection by MALDI-TOF mass spectrometry is achieved by conversion of lysine to homoarginine. Anal Biochem **287:** 110–117

Keegstra K, Cline K (1999) Protein import and routing systems of chloroplasts. Plant Cell **11:** 557–570

Kuster B, Mortensen P, Mann M (1999) Identifying proteins in genome databases using mass spectrometry. *In* Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics. Dallas, TX, pp 1897–1989

Mann M, Pandey A (2001) Use of mass spectrometry-derived data to annotate nucleotide and protein sequence databases. Trends Biochem Sci **26:** 54–61

Natera SH, Guerreiro N, Djordjevic MA (2000) Proteome analysis of differentially displayed proteins as a tool for the investigation of symbiosis. Mol Plant Microbe Interact **13:** 995–1009

Panter S, Thomson R, de Bruxelles G, Laver D, Trevaskis B, Udvardi M (2000) Identification with proteomics of novel proteins associated with the peribacteroid membrane of soybean root nodules. Mol Plant Microbe Interact **13:** 325–333

Patton WF (2000) A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. Electrophoresis **21:** 1123–1144

Peltier J, Ytterberg J, Liberles DA, Roepstorff P, van Wijk KJ (2001) Identification of a 350 kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of *Arabidopsis thaliana*. J Biol Chem **276:** 16318–16327

Peltier JB, Friso G, Kalume DE, Roepstorff P, Nilsson F, Adamska I, van Wijk KJ (2000) Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. Plant Cell **12:** 319–342

Quiniou ML, Yates AJ, Langridge-Smith PR (2000) Laser photo-induced dissociation using tandem time-of-flight mass spectrometry. Rapid Commun Mass Spectrom **14:** 361–367

Rouquie D, Peltier JB, MarquisMansion M, Tournaire C, Doumas P, Rossignol M (1997) Construction of a directory of tobacco plasma membrane proteins by combined two-dimensional gel electrophoresis and protein sequencing. Electrophoresis **18:** 654–660

Rowley A, Choudhary JS, Marzioch M, Ward MA, Weir M, Solari RC, Blackstock WP (2000) Applications of protein mass spectrometry in cell biology. Methods **20:** 383–397

Santoni V, Rabilloud T, Doumas P, Rouquie D, Mansion M, Kieffer S, Garin J, Rossignol M (1999) Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels. Electrophoresis **20:** 705–711

Sechi S, Chait BT (1998) Modification of cysteine residues by alkylation: a tool in peptide mapping and protein identification. Anal Chem **70:** 5150–5158

Sechi S, Chait BT (2000) A method to define the carboxyl terminal of proteins. Anal Chem **72:** 3374–3378

Sherrier DJ, Prime TA, Dupree P (1999) Glycosylphosphatidylinositol-anchored cell-surface proteins from Arabidopsis. Electrophoresis **20:** 2027–2035

Shevchenko A, Loboda A, Ens W, Standing KG (2000) MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. Anal Chem **72:** 2132–2141

Thiellement H, Bahrman N, Damerval C, Plomion C, Rossignol M, Santoni V, de Vienne D, Zivy M (1999) Proteomics for genetic and physiological studies in plants. Electrophoresis **20:** 2013–2026

Tsugita A, Kamo M, Kawakami T, Ohki Y (1996) Two-dimensional electrophoresis of plant proteins and standardization of gel patterns. Electrophoresis **17:** 855–865

van Wijk KJ (2000) Proteomics of the chloroplast: experimentation and prediction. Trends Plant Sci **5:** 420–425

Vener AV, Harms A, Sussman MR, Vierstra RD (2001) Mass spectrometric resolution of reversible protein phosphorylation in photosynthetic membranes of Arabidopsis thaliana. J Biol Chem **276:** 6959–6966

Yamaguchi K, Subramanian AR (2000) The plastid ribosomal proteins. Identification of all the proteins in the 50 s subunit of an organelle ribosome (chloroplast). J Biol Chem **275:** 28466–28482

Yamaguchi K, von Knoblauch K, Subramanian AR (2000) The plastid ribosomal proteins. Identification of all the proteins in the 30 s subunit of an organelle ribosome (chloroplast). J Biol Chem **275:** 28455–28465