The Green proteome: challenges in plant proteomics

Joshua L. Heazlewood*

Joint BioEnergy Institute and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
*Correspondence: jlheazlewood@lbl.gov

An organism is an expression of its underlying molecular composition that reacts and responds to a variety of stimuli. Central to this response are proteins, which undertake a multitude of functional and structural roles throughout an organism. The ability to readily identify and analyze protein populations (proteomics) represents a recent technological advance in the field of biological sciences. Although the term proteomics was coined in the mid 1990's (Wilkins, 2009), the technique was significantly hampered by a lack of cohesive genomics data and advanced instrumentation. The field of proteomics has expanded rapidly in recent years due to the completion of genomes, access and improvements to mass spectrometers and the development of new techniques all of which have increased its role in biological research (Han et al., 2008).

In general the area of plant proteomics suffers from the similar challenges that face other research fields. While the area was energized with the release of the model dicot genome sequence of *Arabidopsis* (Arabidopsis Genome Initiative, 2000) and several years later with the monocot genome of rice (Goff et al., 2002), the development of plant specific procedures for proteomic analyses have been necessary. These have essentially been associated with improvements in sample extraction procedures do deal with recalcitrant tissues and the effects of plant secondary metabolites on both extraction and analysis (Heazlewood and Millar, 2006). A number of successful strategies have been developed to overcome these issues including mass spectrometry compatible extraction procedures designed for plant tissues (Isaacson et al., 2006; Sheoran et al., 2009). In general, the plant proteomics community face many of the same limitations that the proteomics community as a whole faces and will benefit significantly with improvements in the field.

A significant challenge in proteomics when studying plants or any complex biological system is an inability to measure the entire proteome (Ahn et al., 2007). While there have been major advances in instrumentation and sample delivery, the number of proteins that can be reproducibly identified from a single sample analysis is limited to hundreds. Since protein samples derived from whole tissue are likely to contain many thousands of proteins, there are considerable technological advances that will need to be achieved before complete proteomic profiling can occur in plants. Although limiting, a number of approaches have been used to partially overcome these restrictions. This includes sample fractionation and the enrichment of protein subpopulations or compartments prior to sample analysis by mass spectrometry (Eubel et al., 2008; Huang et al., 2009; Hynek et al., 2009; Ferro et al., 2010). When these studies are combined, such approaches begin to create a protein location map of the plant cell and have gone some way to provide overviews of the subcellular proteome (Heazlewood et al., 2007). Advances in instrumentation, sample delivery, and reproducible fractionation will all be essential for limitations in detection and dynamic range to be completely overcome.

A fundamental challenge for proteomics is the ability to deliver large-scale protein quantification and to permit comparative proteomics studies (Schulze and Usadel, 2010). This would enable the examination of global protein changes during processes such as plant development or stress responses and provide a powerful biological application to the technology. Currently, limitations in comparative proteomics are constrained by both an inability to measure the entire protein complement and by technical limitations in quantitation procedures. Initial comparative studies were conducted using 2D-PAGE arrays to compare samples of interest and have now developed to incorporate fluorescent dyes for multiplexing and improved sensitivity (Marouga et al., 2005). Although the approach still suffers from the inherent limitations of 2D-PAGE (Taylor et al., 2011). Techniques involving quantitation by mass spectrometry have been widely embraced as sample analysis has moved away from gel arraying techniques. Early methods involved sample multiplexing with isobaric peptide labels (Zieske, 2006) but have been somewhat superseded by a plethora of label-free techniques (Schulze and Usadel, 2010). Unfortunately there is often an increased need for experimental design and replication to ensure statistical significance when it comes to data analysis and the technique still suffers from sample sensitivity in complex mixtures and dynamic range issues. The utilization of unlabeled targeted approaches (selected reaction monitoring, SRM) greatly improves sample sensitivity by mass spectrometry as well as reducing variation (Lange et al., 2008). While these advantages make the SRM analysis technique very powerful it is still limited by the number of peptides and subsequently proteins that can be monitored during sample analysis (Picotti et al., 2009). Finally, stable isotope labeling with amino acids (SILAC) has been widely employed in non-plant systems to enable reliable protein quantification (Mann, 2006). Since this technique relies on the exogenous addition of labeled amino acids the technique has been difficult to implement in autotrophic plants due to low incorporation rates (Gruhler et al., 2005). While comparative proteomics has been successfully and extensively employed to characterize plants, significant breakthroughs are still required for this technology to enable the complete quantification of an organism's proteome.

Generally plant proteomic studies have focused on the analysis of multiple organs and cell types. For most plant organs this has been a relatively uncomplicated task and has resulted in an extensive collection of studies analyzing and comparing samples derived from different plant organs (Sheoran et al., 2007; Baerenfaller et al., 2008; Castellana et al., 2008). Consequently, few studies if any have attempted in-depth characterizations of single cell lines. Such an intricate approach will be vital to more completely understand the nuanced role of proteins in plant processes from distinct cell types and even single cells. While genomic and transcriptomics approaches

can rely on amplification to overcome sample abundance limitations, this is currently not possible with proteins. With current available technologies, it is highly challenging to significantly characterize single cells or cell types by traditional mass spectrometry based proteomics methods (Wienkoop et al., 2004). Although there have been improvements in single cell type isolation techniques such as Laser microdissection and mechanical dissection the ability of these techniques to significantly contribute to proteomic analyses are at this stage limited (Kerk et al., 2003; Thome et al., 2006). As a consequence methods for single cell analysis of proteins are employing antibody arrays, cell sorting, and fluorescence microscopy to analyze a targeted set of proteins (Salehi-Reyhani et al., 2011). Given the limitations of resources such as diverse antibody collections within the plant research community, this approach will likely remain poorly utilized by plant researchers. Substantive advances in the sensitivity of mass spectrometers and cell isolation techniques will be required to enable such intricate approaches.

Plant genomes code for tens of thousands of functional gene products, a figure that is to eclipsed when considering the effect of post translational modifications on protein function (Rappsilber and Mann, 2002). The main objective for most plant proteomic studies has been the reliable identification of proteins from samples of interest. While the determination of post translational modifications has been desirable, difficulties in unambiguously determining post translational modifications by mass spectrometry is extremely challenging (Mann and Jensen, 2003). Consequently few large-scale analyses of protein modifications have been undertaken in plants with the exception of protein phosphorylation, where many thousands of phosphorylation sites have been characterized (Heazlewood et al., 2008; Gao et al., 2009). The success of these phosphoproteomic studies highlights the success of enrichment strategies prior to mass spectrometry to characterize post translational modifications (Nakagami et al., 2010). Although, since the enrichment process uncouples post translational modification characterization with routine sample analysis, the process adds a level of complication by not providing protein identification and phosphorylation in a single analysis. While the utilization of high resolution mass spectrometry has provided some increased confidence when assigning other protein modifications (Zybailov et al., 2008), the sheer number of permutations that can occur makes the determination of post translational modifications by mass spectrometry a serious challenge to plant proteomics. The development of novel enrichment strategies, increased sensitivity, and well annotated genomes are likely to enable the further development of this area.

The field of proteomics has witnessed a substantive increase in specialized online resources, tools, and repositories (Deutsch et al., 2008; Vizcaíno et al., 2009), with the discipline of plant sciences being no exception (Weckwerth et al., 2008). Many of these resources and tools are generic to proteomics and include data repositories and utilities developed by research groups for specific analyses (Polpitiya et al., 2008). Nonetheless there is an abundance of plant specific proteomic resources available, with the majority focused on the model plant *Arabidopsis*. These databases cover organelle or subcellular localizations, protein phosphorylation, proteogenomic mapping, protein–protein interaction, plant organ mapping, and organ abundance information (Heazlewood et al., 2007, 2008; Baerenfaller et al., 2008; Castellana et al., 2008; MacLean et al., 2008; Sun et al., 2009; Lalonde et al., 2010). Recently a coordinated effort was made to create an aggregation portal to summarize the varied *Arabidopsis* proteomic resources in a single interface for the research community at large (Joshi et al., 2011). Such an integrated approach likely symbolizes the future of data management and analysis. This resource is significant in that it represents the first example of proteomic data unification by a variety of specialty research groups. The area of plant proteomics is extremely well serviced by online data resources with many services actively interacting and exchanging data. The challenge is to integrate these proteomics data sources with other community plant resources to create a web of interlinked repositories.

Plant research has significantly advanced the field of proteomics by overcoming plant specific challenges and by contributing to the development of proteomic related technologies and analyses. While proteomics research in plants will be greatly supported by general advances in the field, there still remains many specific problems that will ultimately require tailored solutions for plant research. Significant challenges still lie ahead of this field and it is envisioned that future advances in techniques and instrumentation will ultimately deliver the ability to analyze, characterize and quantify the entire protein constituent of a cell.

REFERENCES

Ahn, N. G., Shabb, J. B., Old, W. M., and Resing, K. A. (2007). Achieving in-depth proteomics profiling by mass spectrometry. *ACS Chem. Biol.* 2, 39–52.

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

Baerenfaller, K., Grossmann, J., Grobei, M. A., Hull, R., Hirsch-Hoffmann, M., Yalovsky, S., Zimmermann, P., Grossniklaus, U., Gruissem, W., and Baginsky, S. (2008). Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. *Science* 320, 938–941.

Castellana, N. E., Payne, S. H., Shen, Z., Stanke, M., Bafna, V., and Briggs, S. P. (2008). Discovery and revision of *Arabidopsis* genes by proteogenomics. *Proc. Natl. Acad. Sci. U.S.A.* 105, 21034–21038.

Deutsch, E. W., Lam, H., and Aebersold, R. (2008). PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* 9, 429–434.

Eubel, H., Meyer, E. H., Taylor, N. L., Bussell, J. D., O'Toole, N., Heazlewood, J. L., Castleden, I., Small, I. D., Smith, S. M., and Millar, A. H. (2008). Novel proteins, putative membrane transporters, and an integrated metabolic network are revealed by quantitative proteomic analysis of *Arabidopsis* cell culture peroxisomes. *Plant Physiol.* 148, 1809–1829.

Ferro, M., Brugière, S., Salvi, D., Seigneurin-Berny, D., Court, M., Moyet, L., Ramus, C., Miras, S., Mellal, M., Le Gall, S., Kieffer-Jaquinod, S., Bruley, C., Garin, J., Joyard, J., Masselon, C., and Rolland, N. (2010). AT_CHLORO, A comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Mol. Cell. Proteomics* 9, 1063–1084.

Gao, J., Agrawal, G. K., Thelen, J. J., and Xu, D. (2009). P3DB: a plant protein phosphorylation database. *Nucleic Acids Res.* 37, D960–D962.

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W. L., Chen, L., Cooper,

B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296, 92–100.

Gruhler, A., Schulze, W. X., Matthiesen, R., Mann, M., and Jensen, O. N. (2005). Stable isotope labeling of *Arabidopsis thaliana* cells and quantitative proteomics by mass spectrometry. *Mol. Cell. Proteomics* 4, 1697–1709.

Han, X., Aslanian, A., and Yates, J. R. III. (2008). Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* 12, 483–490.

Heazlewood, J. L., Durek, P., Hummel, J., Selbig, J., Weckwerth, W., Walther, D., and Schulze, W. X. (2008). PhosPhAt: a database of phosphorylation sites in *Arabidopsis thaliana* and a plant-specific phosphorylation site predictor. *Nucleic Acids Res.* 36, D1015–D1021.

Heazlewood, J. L., and Millar, A. H. (2006). "Plant proteomics: challenges and resources," in *Plant Proteomics*, ed. C. Finnie (Oxford: Blackwell Publishing), 1–31.

Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Small, I., and Millar, A. H. (2007). SUBA: the *Arabidopsis* subcellular database. *Nucleic Acids Res.* 35, D213–D218.

Huang, S., Taylor, N. L., Narsai, R., Eubel, H., Whelan, J., and Millar, A. H. (2009). Experimental analysis of the rice mitochondrial proteome, its biogenesis, and heterogeneity. *Plant Physiol.* 149, 719–734.

Hynek, R., Svensson, B., Jensen, O. N., Barkholt, V., and Finnie, C. (2009). The plasma membrane proteome of germinating barley embryos. *Proteomics* 9, 3787–3794.

Isaacson, T., Damasceno, C. M. B., Saravanan, R. S., He, Y., Catala, C., Saladie, M., and Rose, J. K. C. (2006). Sample extraction techniques for enhanced proteomic analysis of plant tissues. *Nat. Protoc.* 1, 769–774.

Joshi, H. J., Hirsch-Hoffmann, M., Baerenfaller, K., Gruissem, W., Baginsky, S., Schmidt, R., Schulze, W. X., Sun, Q., van Wijk, K., Egelhofer, V., Wienkoop, S., Weckwerth, W., Bruley, C., Rolland, R., Toyoda, T., Nakagam, H., Jones, A., Briggs, S. P., Castleden, I., Tanz, S., Millar, A. H., and Heazlewood, J. L. (2011). MASCP gator: an aggregation portal for the visualization of *Arabidopsis* proteomics data. *Plant Physiol.* 155, 259–270.

Kerk, N. M., Ceserani, T., Tausta, S. L., Sussex, I. M., and Nelson, T. M. (2003). Laser capture microdissection of cells from plant tissues. *Plant Physiol.* 132, 27–35.

Lalonde, S., Sero, A., Pratelli, R. j., Pilot, G., Chen, J., Sardi, M. I., Parsa, S. A., Kim, D.-Y., Acharya, B. R., Stein, E. V., Hu, H.-C., Villiers, F., Takeda, K., Yang, Y., Han, Y. S., Schwacke, R., Chiang, W., Kato, N., Loqu, D., Assmann, S. M., Kwak, J. M., Schroeder, J., Rhee, S. Y., and Frommer, W. B. (2010). A membrane protein/signaling protein interaction network for *Arabidopsis* version AMPv2. *Front. Physiol.* 1:24. doi: 10.3389/fphys.2010.00024

Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 4, 222.

MacLean, D., Burrell, M. A., Studholme, D. J., and Jones, A. M. (2008). PhosCalc: a tool for evaluating the sites of peptide phosphorylation from mass spectrometer data. *BMC Res. Notes* 1, 30. doi: 10.1186/1756-0500-1-30

Mann, M. (2006). Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* 7, 952–958.

Mann, M., and Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* 21, 255–261.

Marouga, R., David, S., and Hawkins, E. (2005). The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Anal. Bioanal. Chem.* 382, 669–678.

Nakagami, H., Sugiyama, N., Mochida, K., Daudi, A., Yoshida, Y., Toyoda, T., Tomita, M., Ishihama, Y., and Shirasu, K. (2010). Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. *Plant Physiol.* 153, 1161–1174.

Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B., and Aebersold, R. (2009). Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* 138, 795–806.

Polpitiya, A. D., Qian, W.-J., Jaitly, N., Petyuk, V. A., Adkins, J. N., Camp, D. G., Anderson, G. A., and Smith, R. D. (2008). DAnTE: a statistical tool for quantitative analysis of-omics data. *Bioinformatics* 24, 1556–1558.

Rappsilber, J., and Mann, M. (2002). What does it mean to identify a protein in proteomics? *Trends Biochem. Sci.* 27, 74–78.

Salehi-Reyhani, A., Kaplinsky, J., Burgin, E., Novakova, M., deMello, A. J., Templer, R. H., Parker, P., Neil, M. A. A., Ces, O., French, P., Willison, K. R., and Klug, D. (2011). A first step towards practical single cell proteomics: a microfluidic antibody capture chip with TIRF detection. *Lab Chip* 11, 1256–1261.

Schulze, W. X., and Usadel, B. (2010). Quantitation in mass-spectrometry-based proteomics. *Annu. Rev. Plant Biol.* 61, 491–516.

Sheoran, I. S., Ross, A. R. S., Olson, D. J. H., and Sawhney, V. K. (2007). Proteomic analysis of tomato (*Lycopersicon esculentum*) pollen. *J. Exp. Bot.* 58, 3525–3535.

Sheoran, I. S., Ross, A. R. S., Olson, D. J. H., and Sawhney, V. K. (2009). Compatibility of plant protein extraction methods with mass spectrometry for proteome analysis. *Plant Sci.* 176, 99–104.

Sun, Q., Zybailov, B., Majeran, W., Friso, G., Olinares, P. D. B., and van Wijk, K. J. (2009). PPDB, the plant proteomics database at Cornell. *Nucleic Acids Res.* 37, D969–D974.

Taylor, N. L., Heazlewood, J. L., and Millar, A. H. (2011). The *Arabidopsis thaliana* 2D gel mitochondrial proteome: refining the value of reference maps for assessing protein abundance, contaminants and post-translational modifications. *Proteomics.* doi: 10.1002/pmic.201000620. [Epub ahead of print].

Thome, M., Skrablin, M. D., and Brandt, S. P. (2006). Tissue-specific mechanical microdissection of higher plants. *Physiol. Plant.* 128, 383–390.

Vizcaíno, J. A., Côté, R., Reisinger, F., M. Foster, J., Mueller, M., Rameseder, J., Hermjakob, H., and Martens, L. (2009). A guide to the proteomics identifications database proteomics data repository. *Proteomics* 9, 4276–4283.

Weckwerth, W., Baginsky, S., van Wijk, K., Heazlewood, J. L., and Millar, A. H. (2008). The multinational *Arabidopsis* steering subcommittee for proteomics assembles the largest proteome database resource for plant systems biology. *J. Proteome Res.* 7, 4209–4210.

Wienkoop, S., Zoeller, D., Ebert, B., Simon-Rosin, U., Fisahn, J., Glinski, M., and Weckwerth, W. (2004). Cell-specific protein profiling in *Arabidopsis thaliana* trichomes: identification of trichome-located proteins involved in sulfur metabolism and detoxification. *Phytochemistry* 65, 1641–1649.

Wilkins, M. (2009). Proteomics data mining. *Expert Rev. Proteomics* 6, 599–603.

Zieske, L. R. (2006). A perspective on the use of iTRAQ™ reagent technology for protein complex and profiling studies. *J. Exp. Bot.* 57, 1501–1508.

Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., and van Wijk, K. J. (2008). Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE* 3, e1994. doi: 10.1371/journal.pone.0001994