PROTEOMICS: BROAD STROKES OF EXPRESSIONISM?

With the nucleotide sequences of two Arabidopsis chromosomes already published (Lin et al., 1999; Mayer et al., 1999), and completed sequencing of the remaining three chromosomes forecast by the end of this year, it is clear that the age of plant genomics is upon us. But what’s to be done with all this accumulating information? As exciting as the promise of new biological insight into plant life is, the prospect of making sense of the amassing genomic data can be somewhat daunting. Nearly 8000 structural genes have been identified on Arabidopsis chromosomes 2 and 4 alone. The assignment of function to each of these will be a formidable task.

One avenue to organizing plant genomics data is to draw comparisons from data elaborated from the genomes of other organisms. Indeed, function can be assigned to 50 to 60% of the genes located on the two sequenced Arabidopsis chromosomes on the basis of comparison with known proteins from a variety of systems. A second approach to broadly characterizing the genome has been explored through random mutagenesis. In Arabidopsis, several stock populations have been generated, particularly through insertional mutagenesis with T-DNAs or transposons. The experimental attractiveness of the insertional mutagen is its use as a sequence tag to identify, through polymerase chain reaction technology, any number of mutated genes. Tens of thousands of mutant lines have been generated through insertional strategies, and their usefulness in addressing specific questions of biology has been experimentally verified (see, e.g., Meissner et al., 1999; Parinov et al., 1999).

The use of microarrays to follow genomewide transcriptional activity in response to specific conditions has also been touted as a powerful tool for functional genomics (Schena et al., 1995; see also Bouchez and Höfte, 1998). The advantages of microarrays in testing for multiple physiologic agonists and with regard to mutant and wild-type organisms under a variety of environmental conditions can hardly be overstated. The true reward of having full genomic data, after all, will be in understanding the orchestration of multiple genes in elaborating programs of homeostasis, development, defense, and disease.

But it is also clear that gene transcription, even when defined on a genomewide scale, does not offer the complete molecular picture of the organism that genomics as a science has promised. Specifically, a dependable correlation between mRNA abundance and protein has in fact proven to be lacking in many experimental systems (see, e.g., Gygi et al., 1999). The expression of gene activity refers foremost to protein products, despite the usefulness of any number of libraries of ESTs (i.e., “expressed” sequence tags, generated by reverse transcription of mRNA). And the degree to which proteins—and not their genetic messages—are able to interact, or are prevented from interacting, lies at the center of the signal transduction pathways that integrate all life processes. As opposed to the relatively constant genome sequence that defines the organism as a whole, moreover, the physiologic and developmental state of any particular cell is best depicted by the array of proteins that it contains at any given time.

The complete array of proteins that a cell produces (i.e., its proteome) may be experimentally trickier to define than its genome, but the science of proteomics will surely be one of the greatest dividends to come from genomic sequencing. Access to complete genomic data will mean that even very limited peptide sequence analyses can provide enough information to identify proteins and specify their complete primary structures from public databases. The facile identification of proteins will allow researchers to approach very general issues of biology in ways that would only a decade ago have seemed nothing short of presumptuous. In a single article, for instance, researchers were able to assert that the GroEL chaperonin of Escherichia coli interacts with no fewer than 300 polypeptides, of which over 50 specific GroEL substrates could be identified by sequence (Houry et al., 1999). The power of proteomics is thus clear in that very broad questions, such as how many proteins associate with a given molecule, can be meaningfully posed and answered.

On pages 319–341 of this issue of THE PLANT CELL, Peltrie et al. confront the challenges of proteomics within a plant system. Their particular question, which reflects the broad experimental brush strokes of proteomics, pertains to the enumeration of proteins that can be recovered from the lumen and membrane surfaces of chloroplast thylakoids isolated from pea. In the chloroplast context, the relevance of proteomics, as opposed to the nucleotide sequencing normally associated with genomics, is especially clear: the vast majority of chloroplast proteins are not encoded by the organellar genome but rather by the nucleus. Thus, whereas the cellular proteome will frequently represent less genetic material than the genome, the chloroplasts (and mitochondria) are distinct in containing proteomes that pertain to proteins beyond those encoded by the organellar genome.
IN THIS ISSUE

The investigation of Peltier et al. begins with a reproducible map of the set of proteins under study, a requirement of proteomics in general. Typically, the complex mixture of proteins to be analyzed is resolved on a two-dimensional gel, first according to their isoelectric points (i.e., a measure of their acid/base properties) and then according to their mass. In the present study, the authors were reliably able to visualize over 400 lumenal and thylakoid peripheral protein spots that were estimated to represent over 200 mature proteins.

After the proteins are resolved into spots by two-dimensional electrophoresis, they can be digested into fragments (by an endoprotease such as trypsin) that are then sequenced. It is mass-spectrometric sequencing technology that brings proteomics efforts such as those of Peltier et al. to fruition. Although mass spectrometry (MS) has been commercially available since before the Second World War for the analysis of volatile organic molecules, it is only within the past two decades that matrix-assisted laser desorption/ionization (MALDI) MS and electrospray ionization (ESI) MS technologies were developed so that large complex molecules such as proteins could be analyzed. In the work of Peltier et al., MALDI-MS is highlighted for its ability to fingerprint tryptic digests; ESI tandem MS, like Edman analysis, results in the fragmentation of peptides so that the step-wise loss of amino acid residues can be monitored according to the concomitant loss of mass.

The spectacular power of proteomics is evident in the authors’ approach to the chloroplast. From their pea thylakoid preparations, the sequence identity of over 60 distinct protein species was established. The function or identity of more than half of these is traceable from public databases, but ten of the sequences correspond to entities that have not heretofore been characterized. At this relatively early stage in the history of proteomics, it is thus evident that the broad net that is cast by the relevant methodologies is sure to capture many novel polypeptides.

But what about biological function? Does the authors’ approach tell us anything about thylakoid biology? Although the prime benefits of the broad approach may amount to pointers in the direction of future research, the extensive set of analyzed proteins inevitably reflects the biology of chloroplast genesis and function. As an example, several of the previously unidentified luminal proteins contain signal peptide sequences including a twin arginine motif. The authors have in fact introduced the new protein sequences indicative of translocation by the twin arginine translocation (TAT) pathway into plotting programs that illustrate the consensus signals for translocation into the lumen. Beyond refining a consensus or even probing the rigors of programs that predict the translocation pathways of newly sequenced polypeptides, however, the authors’ analysis again underscores the utility of proteomics in defining the features of a particular set of proteins.

In the present work of Peltier et al., the biological question raised in the proteomics approach pertains to the basis upon which a given set of proteins is directed to an organelle. Clearly, experiments in other systems could just as well question the basis for interactions of proteins with a given photoreceptor, for example, or even with an entire signal cascade. In the near future, as the genomes of plants and other organisms become fully sequenced, proteomics will thus provide fundamental insights into organismal development and homeostasis that the mere knowledge of genome sequence cannot provide.

Harry B. Smith

REFERENCES


Proteomics: Broad Strokes of Expressionism?
Harry B. Smith
*Plant Cell* 2000;12;303-304
DOI 10.1105/tpc.12.3.303

This information is current as of October 20, 2017

| References | This article cites 9 articles, 6 of which can be accessed free at: /content/12/3/303.full.html#ref-list-1 |
| eTOCs | Sign up for eTOCs at: http://www.plantcell.org/cgi/alerts/ctmain |
| CiteTrack Alerts | Sign up for CiteTrack Alerts at: http://www.plantcell.org/cgi/alerts/ctmain |
| Subscription Information | Subscription Information for *The Plant Cell* and *Plant Physiology* is available at: http://www.aspb.org/publications/subscriptions.cfm |

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY