Mass Spectrometry. An Essential Tool in Proteome Analysis

Jiaxu Li and Sarah M. Assmann*

Department of Biology, The Pennsylvania University, University Park, Pennsylvania 16802

The entire genomic DNA sequence of Arabidopsis will be available by the end of this year. As has been the case for a number of prokaryotic and eukaryotic species whose genomic DNA sequences have been completed, many open reading frames in the Arabidopsis genome will encode proteins of unknown functions. Genome-wide analysis of mRNA expression by microarray approaches can provide important clues about expression patterns and thus the functions of gene products (Somerville and Somerville, 1999). However, for a substantial number of proteins, there may be only a loose correlation between mRNA and protein levels (Abbott, 1999; Gygi et al., 1999). In addition, the functions of proteins depend considerably on post-translational modifications and interaction with other proteins, processes that cannot be deduced from nucleic acid microarray data. Therefore, efficient approaches for identifying proteins, for determining protein expression in different tissues and under different conditions, for identifying post-translational modification of proteins in response to different stimuli, and for characterizing protein interactions will be critical for understanding biological processes in the post-genome era.

Recent innovations in mass spectrometry have significantly improved its application in the study of protein structure and function (for review, see Loo, 1995; Yates, 1998). The value of mass spectrometry to biologists has been established by its effectiveness in characterizing the structure and dynamics of proteins at the femtomole to picomole level. Mass spectrometry can be used to characterize function-critical post-translational modifications, including phosphorylation and glycosylation, to determine the numbers and positions of disulfide bonds in proteins, and to investigate macromolecular complexes, such as protein-ligand, protein-protein, and protein-DNA interactions (Fitzgerald and Siuzdak, 1996; Yates, 1998). When a complete genome sequence is available, mass spectrometric quantitation of the masses of a few tryptic fragments from an unknown protein, followed by the use of algorithms to compare the observed peptide masses against those predicted for the theoretical tryptic fragments of all expressed sequences, will often suffice for exact protein identification (Yates, 1998). This process, known as peptide mass mapping or peptide mass fingerprinting, will be a powerful method for protein identification and expression pattern analysis in Arabidopsis, and can be extended to other plant species (e.g. rice) as soon as genome sequencing is completed.

When a complete genome sequence is not available, amino acid sequencing is required for protein identification. For proteins not amenable to analysis by Edman degradation, tandem mass spectrometric sequencing (Fig. 1) is often employed. Amino acid sequence information provided by mass spectrometric analysis can allow homology searching and cloning or database identification of the corresponding gene (Shevchenko et al., 1997). In this correspondence, we use de novo peptide sequencing of a low abundance broad bean (Vicia faba) protein isolated from two-dimensional gels to illustrate the power of mass spectrometry for protein identification.

Several years ago, using an in-gel kinase assay (Wang and Chollet, 1993), we identified a 48-kD abscisic acid (ABA)-activated and Ca\(^{2+}\)-independent protein kinase (AAPK) from broad bean guard cells (Li and Assmann, 1996). AAPK is a low abundance protein that can only be detected in guard cells. AAPK is activated upon treatment of intact guard cells with ABA, but is not activated by ABA in vitro. The very low amount of protein starting material from a limited quantity of purified guard cells made it unfeasible to purify AAPK by protein purification methods other than two-dimensional gel electrophoresis. However, attempts to isolate AAPK by conventional two-dimensional electrophoresis (isoelectric focusing/SDS-PAGE) based on the method of O’Farrell (1975) were unsuccessful, because the ABA-induced autophosphorylating capability of AAPK, which serves to identify the kinase, was lost. Changes in sample solubilization buffer and detergent (Nonidet P-40, Triton X-100, or CHAPS [3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonic acid]) did not restore AAPK autophosphorylation. Nonequilibrium pH gradient electrophoresis-based protocols for focusing very basic or very acidic proteins were likewise unsuccessful. These results suggest that the autophosphorylation activity of AAPK may be abolished by the detergent present in the sample buffer for the first dimensional electrophoresis or by the isoelectric focusing step. To avoid these problems, a two-dimensional
electrophoresis protocol with non-denaturing PAGE for the first dimension and SDS-PAGE for the second dimension was developed. When proteins separated by this protocol were subjected to an in-gel autophosphorylation assay, a 48-kD Ca\(^{2+}\)-independent 32P-labeled spot was detected from guard cells treated with ABA, but not from control guard cells lacking the ABA treatment (Fig. 2, A and B). These characteristics demonstrated that the protein corresponding to the 48-kD 32P-labeled spot was AAPK. Silver staining of the two-dimensional gels showed that the AAPK spot was separate from other protein spots (Fig. 2, C and D).

About 1 pmol of AAPK excised from six two-dimensional gels was subjected to tandem mass spectrometry for amino acid sequence analysis (Harvard Microchemistry Facility, Cambridge, MA). Two peptide sequences were obtained from AAPK, both of which turned out to be highly conserved in the PKABA1 class of protein kinases. The founding member of the PKABA1 kinase family is a kinase from wheat whose transcription is induced by ABA (Anderberg and Walker-Simmons, 1992). The AAPK peptide sequence information obtained by mass spectrometry allowed us to clone the AAPK cDNA. Expression in guard cells of GFP-tagged AAPK ultimately allowed us to determine that this kinase mediates ABA-induced stomatal closure via activation of guard cell anion channels (Li et al., 2000).

**Figure 1.** Schematic of peptide sequencing by tandem mass spectrometry. A mass spectrometer consists of an ionization source, a mass analyzer, and a detector. In tandem mass spectrometry, two stages of mass analysis are linked in series. Initially, a protein of interest is digested with a residue-specific protease, e.g. trypsin. A strong electric field is used to nebulize the fluid containing the sample (delivered, e.g. by liquid chromatography) and to charge the peptide, typically by proton attachment or abstraction. This process is termed electrospray ionization. The sample is then delivered to the first mass spectrometer (MS-1), where the peptides (P1–P3) are identified based on their mass to charge ratio (m/z). One peptide from the peptide mixture is selected (P2) and then fragmented by collision with an inert gas such as argon. Fragmentation occurs mainly at the amide bonds of the peptide, resulting in a nested set of peptides differing by the mass of one amino acid. The second mass spectrometer (MS-2) analyzes the m/z ratios of the resulting peptide fragments (F1–F5). By computational assembling of the fragments, the peptide sequence (P2) can be deduced.

**Figure 2.** Isolation of AAPK by two-dimensional gel electrophoresis. Soluble proteins (50 μg) from ABA-treated guard cell protoplasts (B and D) or untreated guard cell protoplasts (A and C) were first separated by non-denaturing PAGE (5%–15% acrylamide gradient). The lanes containing guard cell protoplast proteins were excised from the non-denaturing polyacrylamide gel. After equilibration in SDS-PAGE sample buffer, the excised gel slices were placed horizontally on top of SDS-denaturing gels (12% [w/v] acrylamide) and proteins were resolved vertically by SDS-PAGE. After electrophoresis in the second dimension, proteins in the gels were visualized by silver staining (C and D) or subjected to an in-gel autophosphorylation assay in the presence of EGTA and then visualized by autoradiography (A and B). The positions of molecular mass standards are given on the left in kD. The arrows indicate the position of AAPK.
Plants, like all eukaryotic organisms, have a very large number of protein kinases. The approach described here should allow detection and identification of protein kinases present in limited amount of cells such as isolated bundle sheath cells or dissected tiny tissues such as shoot meristems and root caps. A related approach should be useful for identifying all the proteins in a system of interest. Recently developed techniques (Binz et al., 1999) allow automation of in-gel tryptic digestion of all the proteins in a two-dimensional gel, followed by their transfer to a membrane that can then be scanned by laser mass spectrometry (matrix-assisted laser desorption-ionization mass spectrometry) to obtain diagnostic peptide masses for peptide mass fingerprinting. Thus, mass spectrometry combined with two-dimensional gel electrophoresis will be an important tool for large-scale proteome analysis in the post-genome era.

ACKNOWLEDGMENTS

We are grateful for helpful comments on this manuscript from an anonymous reviewer and from Dr. Amy Harms, Director of the University of Wisconsin Mass Spectroscopy Facility.

Received March 22, 2000; accepted April 13, 2000.

LITERATURE CITED