Investigative Proteomics: Identification of an Unknown Plant Virus from Infected Plants Using Mass Spectrometry

Bret Cooper
Plant Health Department, Torrey Mesa Research Institute of Syngenta, San Diego, California USA

Donna Eckert, Nancy L. Andon, John R. Yates, III, and Paul A. Haynes
Protein and Metabolite Dynamics Department, Torrey Mesa Research Institute of Syngenta, San Diego, California, USA

We describe the identification of a previously uncharacterized plant virus that is capable of infecting *Nicotiana* spp. and *Arabidopsis thaliana*. Protein extracts were first prepared from leaf tissue of uninfected tobacco plants, and the proteins were visualized with two-dimensional electrophoresis (2-DE). Matching gels were then run using protein extracts of a tobacco plant infected with tobacco mosaic virus (TMV). After visual comparison, the proteins spots that were differentially expressed in infected plant tissues were cut from the gels and analyzed by high performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS). Tandem mass spectrometry data of individual peptides was searched with SEQUEST. Using this approach we demonstrated a successful proof-of-concept experiment by identifying TMV proteins present in the total protein extract. The same procedure was then applied to tobacco plants infected with a laboratory viral isolate of unknown identity. Several of the differentially expressed protein spots were identified as proteins of potato virus X (PVX), thus successfully identifying the causative agent of the uncharacterized viral infection. We believe this demonstrates that HPLC-MS/MS can be used to successfully characterize unknown viruses in infected plants.   (J Am Soc Mass Spectrom 2003, 14, 736–741) © 2003 American Society for Mass Spectrometry

The identification of plant viruses remains cumbersome despite the existence of an abundance of procedures to facilitate the process [1]. Traditional symptom diagnosis and host range studies can help classify some common viruses if a wide range of plants is available. Electron microscopy is also useful [2], but is normally suited for morphological diagnosis such as distinguishing a rod-shaped virus from an icosahedron. Serological, hybridization, and PCR techniques are easy methods available to identify viruses but require advanced knowledge of capsid protein antigenicity or nucleic acid sequence, or require availability of a range of antisera needed to characterize one of many possible viruses [3–5]. Direct sequencing of the virus can also be used and may be the most accurate diagnostic tool, although this requires the initial work of cloning, subcloning, or primer walking. dsRNA analysis is perhaps the most versatile method for identifying RNA viruses since no prior knowledge is needed about the sample [6]. However, this method is insufficient for identifying a true unknown, especially if there are no reference standards available to the user. In practice, since no single method is truly reliable, plant pathologists typically perform many procedures in order to identify an unknown virus. This can be a time-consuming process.

Unlike some other techniques, mass spectrometry offers the promise of identifying an unknown virus without having to perform numerous other experiments. Peptide mass fingerprinting has been shown previously to be successful in the direct identification of purified viral strains [7, 8]. Due to the limitations of peptide mass fingerprinting, we chose to employ an approach based on tandem mass spectrometry (MS/MS) of trypsin-digested individual peptides. In brief, this consisted of a five step process: Protein extracts were prepared from plant leaf samples; proteins were separated using 2-DE; individual proteins were excised and digested with trypsin; peptides were analyzed by HPLC-MS/MS; the resulting tandem mass spectra were analyzed using the program SEQUEST [9, 10]. Using

this approach we first showed that we were able to identify viral proteins in the biological context of a very complex mixture of plant proteins, and then used this approach to characterize a previously unknown viral isolate.

Materials and Methods

Protein Extraction from Plants

Nicotiana tabacum and *N. clevelandii* were inoculated with purified tobacco mosaic *tobamovirus* (TMV) [11] or an unknown virus (described in Results). Leaves were harvested from plants showing disease symptoms and from symptom-free mock-inoculated plants of the same age. Leaves were snap frozen in liquid nitrogen and ground to a fine powder with a mortar and pestle. Protein was precipitated by addition of 10 volumes of ice-cold acetone containing 10% wt/vol trichloroacetic acid and 0.07% vol/vol β-mercaptoethanol. After standing at −20 °C for 45 min, the protein pellet was separated by centrifugation at 30,000 g for 15 min. The pellet was washed three times with 10 volumes of ice-cold acetone containing 10% wt/vol trichloroacetic acid, 0.07% vol/vol β-mercaptoethanol, 1 mM phenyl-methylsulfonylfluoride, and 2 mM EDTA, and lyophilized to dryness.

2-D Gel Electrophoresis

Protein powder (250 μg) was solubilized in 350 μl of sample buffer containing 7 M urea, 2 M thiourea, 4% CHAPS and 0.5% ampholytes. A Protean isoelectric focusing unit (BioRad, Hercules, CA) was used with 17 cm IPG strips (BioRad) of pH range 3–10. Strips were rehydrated in the presence of sample under constant low voltage (50 V) for 12 h. Isoelectric focusing was performed for 250 V for 1 h, 500 V for 1 h, 1000 V for 1 h, 4000 V for 3 h, and 8000 V for 4 h, for a total of 45,750 Vh. Following electrophoresis, the proteins in the strips were reduced with DTT and alkylated with iodoacetamide, respectively. Second-dimension electrophoresis was performed on 20 × 25 cm 12% linear gels in a BioRad Protean II XL gel cell (BioRad).

Sample Preparation for LC-MS/MS Analysis

Protein bands from 2-D gels were visualized using silver staining [12, 13], excised using a BioRad spot cutter according to the manufacturer's instruction, and transferred to 96-well plates. The plates were transferred to a Massprep digestion robot (Micromass, Beverley, MA) for destaining [14] and in-gel digestion with trypsin [13]. After digestion, the gel pieces were extracted with 5% formic acid/5% CH_3CN on the Massprep robot. The extracted tryptic peptides were diluted to 80 μl per well with 0.1% formic acid.

High Performance Liquid Chromatography-Tandem Mass Spectrometry

A microbore HPLC system (Surveyor, ThermoFinnigan, San Jose, CA) was modified to operate at 300 nl/min using a simple T-piece flow-splitter. Columns (10 cm x 75 μm i.d.) were prepared by packing 100 Å, 5 μm Zorbax C18 resin (Agilent, Palo Alto, CA) into New Objective's Pico Frits (New Objective, Woburn, MA) columns with integral spray needles at 500 psi pressure. Peptides were eluted by using a gradient from buffer A (5% vol/vol acetonitrile, 0.1% formic acid) to buffer B (90% vol/vol acetonitrile, 0.1% formic acid). Following an initial 10 min wash with buffer A, peptides were eluted using a linear gradient from 0–100% buffer B over a 30 min interval. Samples were introduced onto the analytical column using a Surveyor autosampler (Surveyor, ThermoFinnigan). 100 μl peptide extract samples were first transferred onto a C 18 (300 μm × 5 mm) cartridge (LC Packings, San Francisco, CA) and the eluted peptides were then transferred onto the analytical column using a switching valve. The HPLC column eluent was eluted directly into the electrospray ionization source of a ThermoFinnigan LCQ-Deca ion trap mass spectrometer. Automated peak recognition, dynamic exclusion, and daughter ion scanning of the top two most intense ions were performed using the Xcalibur software as described previously [15, 16]. Spectra were scanned over the range 400–1400 mass units.

Database Searching and Data Interpretation

MS/MS data were analyzed using SEQUEST, a computer program that allows the correlation of experimental data with theoretical spectra generated from known protein sequences [9, 10]. In this work, the criteria we used for a preliminary positive peptide identification for a doubly-charged peptide were a correlation factor (Xcorr) greater than 2.5, a delta cross-correlation factor (dCn) greater than 0.1 (indicating a significant difference between the best match reported and the next best match), a high preliminary scoring, and a minimum of one tryptic peptide terminus. For triply-charged peptides the correlation factor threshold was set at 3.5. All matched peptides were confirmed by visual examination of the spectra, and all spectra were searched against the latest version of the public non-redundant protein database of the National Center for Biotechnology Information (NCBI).

Results and Discussion

Identification of a Known Plant Virus in a Biological Context

A protein extract was prepared from leaf tissue of uninfected tobacco plants (*N. tabacum*) and the proteins were visualized using 2-DE with a pH gradient of 3–10
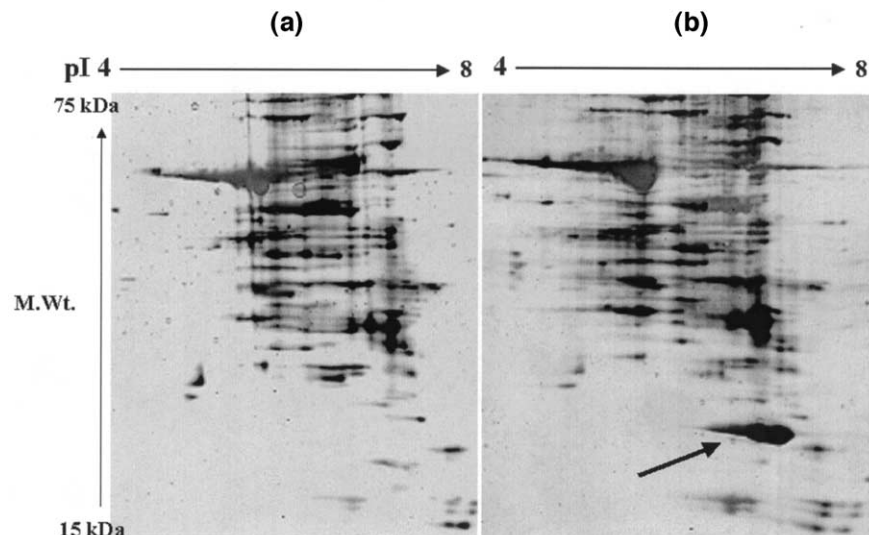
Figure 1. 2-D gel electrophoresis of leaf protein extracts from uninfected *N. tabacum* and *N. tabacum* infected with tobacco mosaic virus. (**a**) Expanded view of an area of detail from leaf protein extract of *N. tabacum* separated using first dimension 17 cm 3–10 p*I* range IPG strips, followed by second dimension 10% linear sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and visualization of proteins with silver staining. (**b**) Detail of the same area from leaf protein extract of *N. tabacum* infected with tobacco mosaic virus, separated under identical conditions. Spots were cut from the gel as indicated in (**b**) and subjected to in-gel digestion with trypsin prior to HPLC-MS/MS analysis. Approximate molecular weight and p*I* range are as indicated.

in the first dimension, as shown in Figure 1a. A similar protein extract was prepared from leaf tissue of tobacco plants infected with TMV. The proteins were separated and visualized under identical conditions as shown in Figure 1b. Several prominent spots that were common between the two gels were first excised and analyzed by HPLC-MS/MS and shown to be abundant proteins of *N. tabacum* such as ATP synthase, glyceraldehyde 3-phosphate dehydrogenase, fructose-bisphosphate aldolase, photosystem II oxygen-evolving complex protein I, and ribulose bisphosphate carboxylase (data not shown). Upon close visual inspection, it was apparent that while the protein expression patterns shown on the two gels are qualitatively similar, there are also numerous subtle differences and several that are very obvious. The single large spot indicated by an arrow in Figure 1b, representing the most obvious difference between the two gels, was then excised from the gel and ana-

lyzed by HPLC-MS/MS. This was found to contain four unique peptides from the coat protein of tobacco mosaic virus, as shown in Table 1. While this information allows us to unambiguously identify the virus responsible for the infection, none of the peptides identified were sufficiently diagnostic to allow identification of the strain of virus involved.

Identification of an Unknown Virus Isolate

In 1993, BC was bequeathed a collection of purified plant viruses from a virologist returning to his home country. One tube in the collection contained a precipitate and was labeled TRV 2.7 mg/ml. It was presumed to be tobacco rattle virus (TRV) and was subsequently shown to infect *Nicotiana* spp. In 2000, we showed that this viral isolate would infect *Arabidopsis thaliana* and this virus was subsequently used in a set of gene

Table 1. Virus peptides identified by HPLC-MS/MS

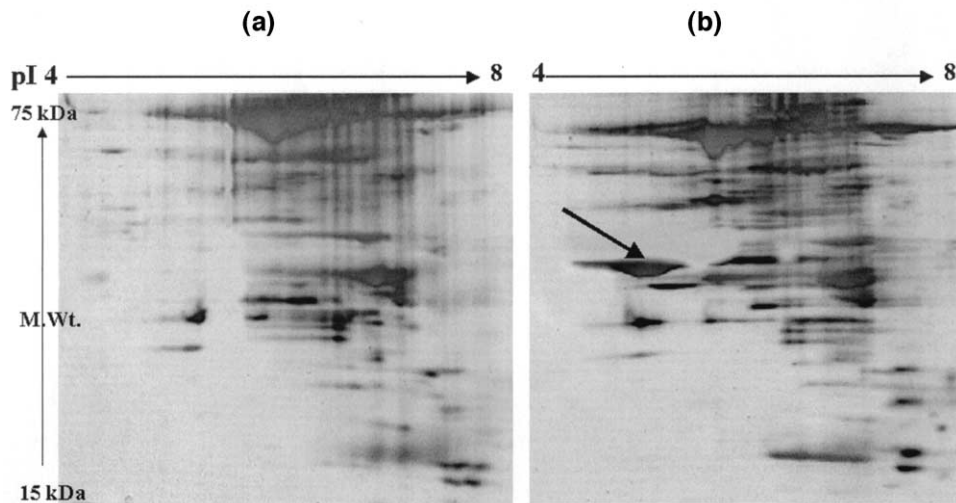| Peptide sequence	Charge state	Xcorr	dCn	Protein header
TMV infected leaf (Figure 1B)				
ALGNQFQTQQAR	2	3.48	0.25	COAT_TMV (P03570)
SAINNLIVELIR	2	3.52	0.15	COAT_TMV (P03570)
PLVTALLGAFDTR	2	4.15	0.35	COAT_TMV (P03570)
RIIEVENQANPTTAETLDATR	2	3.58	0.28	COAT_TMV (P03570)
Unknown viral isolate infected leaf (Figure 2B)				
AQSNDFASLDAAVTR	2	4.11	0.50	COAT_PVX (P07699)
EGLIRPPSEAEMNAAQTAAFVK	3	4.01	0.51	COAT_PVX (P07699)
VPTDTMAQAAWDLVR	2	4.50	0.52	COAT_PVX (P07699)

(a)   **(b)**

Figure 2. 2-D gel electrophoresis of leaf protein extracts from uninfected *N. clevelandii* and *N. clevelandii* infected with unknown virus. (**a**) Expanded view of an area of detail from leaf protein extract of *N. clevelandii* separated using first dimension 17 cm 3–10 p*I* range IPG strips, followed by second dimension 10% linear SDS-PAGE, and visualization of proteins with silver staining. (**b**) Detail of the same area from leaf protein extract of *N. clevelandii* infected with unknown virus, separated under identical conditions. Spots were cut from the gel as indicated in (**b**) and subjected to in-gel digestion with trypsin prior to HPLC-MS/MS analysis. Approximate molecular weight and p*I* range are as indicated.

expression experiments [17]. Consequently, attempts were made to clone the virus, but TRV degenerate PCR primers were not effective for amplifying TRV genes. In an attempt to determine whether this virus was TRV (before the gene expression manuscript was published), dsRNA analysis of infected plants was performed. Characteristic replicative forms for two TRV genomic RNAs were not present; there was only a single dsRNA with a molecular weight greater than the replicative form for genomic TMV (data not shown). Thus, there was sufficient doubt about the identity of the virus that we were motivated to see if a mass spectrometry based approach could be used to identify the unknown virus.
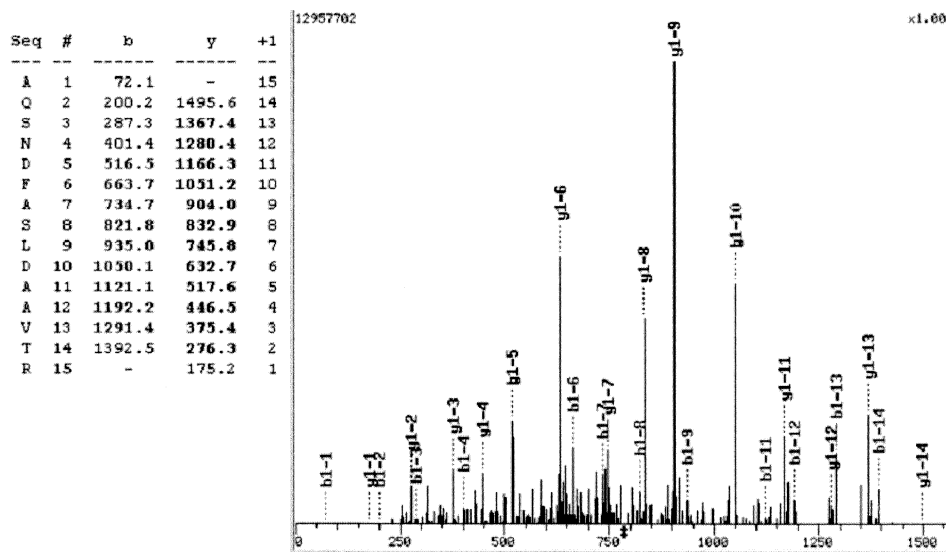
A protein extract was prepared from leaf tissue of uninfected *N. clevelandii* and the proteins were visualized using 2-DE with a pH gradient of 3–10 in the first dimension, as shown in Figure 2a. A protein extract was then prepared from leaf tissue of *N. clevelandii* infected with the virus of unknown origin. The proteins were separated and visualized under identical conditions as shown in Figure 2b. Again, there were several apparent differences between protein expression in the two plant lines. Several excisions were made in the single large spot indicated by an arrow in Figure 2b, representing the most obvious difference between the two gels. Proteins in these samples were analyzed by HPLC-MS/MS, and found to contain three unique peptides of the coat protein of potato virus X *potexvirus*, a single stranded monopartite virus, as shown in Table 1. An example of the tandem mass spectrum of one of the three peptides found, along with the SEQUEST output data for that spectrum, is shown in Figure 3. Again, this information allows us to unambiguously identify the

virus responsible for the infection, although none of the peptides identified were sufficiently diagnostic to allow identification of the strain of virus involved. A host range study or detailed sequence analysis would most likely be necessary to distinguish the exact strain of PVX involved.

Conclusion

Previous reports have demonstrated the successful application of mass spectrometry for the identification of various microorganisms, usually by using a limited proteolytic digestion of whole cells and identifying peptides by MALDI mass spectrometry [18]. A recent study has also detailed a method for rapid identification of viruses and other organisms using a small number of biomarker peptides based on the construction of databases of organism-specific tryptic peptide masses [19]. Mass spectrometry has also been used in the identification of quasispecies of human hepatitis C virus isolates, based on in vitro translation of the amplicons, followed by mass spectrometric analysis of the resulting peptide mix [20]. These methods are typically designed as rapid diagnostic screening tools, and are therefore more suitable for use in partially purified extracts rather than complex biological samples.

In this report we have shown that HPLC-MS/MS of gel-separated proteins can be successfully applied in the identification of different types of viruses in a realistic biological context. In the examples we have presented, it is clear that a high viral load was present along with the plant leaf proteins. It may have been possible to detect viral proteins using one-dimensional

Seq	#	b	y	+1
A	1	72.1	–	15
Q	2	200.2	1495.6	14
S	3	287.3	1367.4	13
N	4	401.4	1280.4	12
D	5	516.5	1166.3	11
F	6	663.7	1051.2	10
A	7	734.7	904.0	9
S	8	821.8	832.9	8
L	9	935.0	745.8	7
D	10	1050.1	632.7	6
A	11	1121.1	517.6	5
A	12	1192.2	446.5	4
V	13	1291.4	375.4	3
T	14	1392.5	276.3	2
R	15	–	175.2	1

SEQUEST Search Result:
Peptide = R.AQSNDFASLDAAVTR, Protein Header = COAT_PVX
Xcorr = 4.11, delta Xcorr = 0.50, peptide mass = 1566.67,
charge state = +2

Figure 3. An example tandem mass spectrum (MS/MS) of a peptide identified from potato virus X. The area highlighted in Figure 2b was cut from the gel and digested with trypsin. Extracted peptides were analyzed by HPLC-MS/MS. This spectrum represents one example of a peptide identified by HPLC-MS/MS and SEQUEST database searching of the resulting spectra as being from potato virus X (see also Table 1). The SEQUEST output data for this spectrum is also indicated.

protein electrophoresis as a comparative tool, or even direct digestion of the protein extract and analysis of all the peptides generated [21]. However, the use of 2-DE as a separation tool allowed us to clearly distinguish viral proteins that could then be analyzed by HPLC-MS/MS. The analysis of protein spots from the 2-DE gels using HPLC-MS/MS enabled us to clearly identify the unknown viral isolate, without any a priori knowledge of the causative agent other than the fact that it appeared to be a plant virus. In addition, the use of tandem mass spectrometry of individual peptides in conjunction with database searching of MS/MS spectra allowed us to identify a number of different peptides from each viral strain. We believe that, especially in the case of samples present in a complex biological matrix, this gives us greater statistical confidence in our data than would be generated using a peptide mass finger-printing approach.

Here we have first presented a proof-of-concept experiment to show that a plant virus can indeed be successfully identified in protein extracts from leaf tissue of experimentally infected plants. We then showed that a virus that had previously been tentatively identified as a TRV was in fact a strain of PVX. We were also unable to find any evidence that the unknown virus was contaminated with any other plant virus, a possibility that had been considered as one possible explanation for the unusual properties of the observed viral infection.

In conclusion, we believe that with further development, a method such as the one we have presented here may be of significant value in the field of plant pathology and plant disease diagnosis, both for confirming expected results and for generating unexpected data as shown in the example we have presented here. The method retains the versatility of dsRNA analysis for identifying mixed infections while maintaining the specificity of identification associated with serological testing. The method will also become even more valuable in the diagnosis of plant bacterial and fungal infections as more genomic sequence data is deposited in the public domain. This demonstrates yet again the power of mass spectrometry as a versatile analytical problem-solving tool.

Acknowledgments

References

1. Matthews, R. E. F. In *Plant Virology*; Matthews, R. E. F. Ed.; Academic Press: San Diego, 1991; pp 12-53.
2. Desjardins, P. R.; Ban, N.; Mathews, D. M.; Kitasako, J. T.; Dodds, J. A.; McPherson, A. Electron Microscopy of Satellite

Tobacco Mosaic Virus Crystals: Metal-Coated, Negatively Stained, and Stereo Pairs. *J. Electron. Microscop. (Tokyo)* **2000,** *49,* 509–514.

3. Clark, M. F.; Adams, A. N. Characteristics of the Microplate Method of Enzyme-Linked Immunosorbent Assay for the Detection of Plant Viruses. *J. Gen. Virol.* **1977,** *34,* 475–483.

4. Sanchez-Navarro, J. A.; Canizares, M. C.; Cano, E. A.; Pallas, V. Simultaneous Detection of Five Carnation Viruses by Non-Isotopic Molecular Hybridization. *J. Virol. Methods* **1999,** *82,* 167–175.

5. Pappu, S. S.; Brand, R.; Pappu, H. R.; Rybicki, E. P.; Gough, K. H.; Frenkel, M. J.; Niblett, C. L. A Polymerase Chain Reaction Method Adapted for Selective Amplification and Cloning of 3' Sequences of Potyviral Genomes: Application to Dasheen Mosaic Virus. *J. Virol. Methods* **1993,** *43,* 267.

6. Dawson, W. O.; Dodds, J. A. Characterization of Sub-Genomic Double-Stranded RNAs from Virus-Infected Plants. *Biochem. Biophys. Res. Commun.* **1982,** *107,* 1230–1235.

7. She, Y. M.; Haber, S.; Seifers, D. L.; Loboda, A.; Chernushevich, I.; Perreault, H.; Ens, W.; Standing, K. G. Determination of the Complete Amino Acid Sequence for the Coat Protein of Brome Mosaic Virus by Time-of-Flight Mass Spectrometry. Evidence for Mutations Associated with Change of Propagation Host. *J. Biol. Chem.* **2001,** *276,* 20039–20047.

8. Lewis, J. K.; Bendahmane, M.; Smith, T. J.; Beachy, R. N.; Siuzdak, G. Identification of Viral Mutants by Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** *95,* 8596–8601.

9. Yates, J. R., III; Eng, J. K.; McCormack, A. L.; Schieltz, D. A Method to Correlate Tandem Mass Spectra of Modified Peptides to Amino Acid Sequences in the Protein Database. *Anal. Chem.* **1995,** *67,* 1426–1436.

10. Eng, J.; McCormack, A. L.; Yates, J. R., III. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Mass Spectrom.* **1994,** *5,* 976–989.

11. Cooper, B.; Dodds, J. A. Differences in the Subcellular Localization of Tobacco Mosaic Virus and Cucumber Mosaic Virus Movement Proteins in Infected and Transgenic Plants. *J. Gen. Virol* **1995,** *76* (Part 12) 3217–3221.

12. Blum, H.; Beier, H.; Gross, H. J. Improved Silver Staining of Plant Proteins, RNA, and DNA in Polyacrylamide Gels. *Electrophoresis* **1987,** *8,* 93–99.

13. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass Spectrometric Sequencing of Proteins from Silver Stained Polyacrylamide Gels. *Anal. Chem.* **1996,** *68,* 850–858.

14. Gharahdaghi, F.; Weinberg, C. R.; Meagher, D. A.; Imai, B. S.; Mische, S. M. Mass Spectrometric Identification of Proteins from Silver-Stained Polyacrylamide Gel: A Method for the Removal of Silver Ions to Enhance Sensitivity. *Electrophoresis* **1999,** *20,* 601–605.

15. Haynes, P. A.; Fripp, N.; Aebersold, R. Identification of Gel Separated Proteins by Liquid Chromatography-Electrospray Tandem Mass Spectrometry: Comparison of Methods and Their Limitations. *Electrophoresis* **1998,** *19,* 939–945.

16. Andon, N. L.; Hollingworth, S.; Koller, A.; Greenland, A. J.; Yates, J. R., III; Haynes, P. A. Proteomic Characterization of Wheat Amyloplasts Using Identification of Proteins by Tandem Mass Spectrometry. *Proteomics* **2002,** *2,* 1156–1168.

17. Whitham, S. A.; Quan, S.; Chang, H.-S.; Cooper, B.; Estes, B.; Zhu, T.; Wang, X.; Hou, Y.-M. Diverse RNA Viruses Elicit the Expression of Common Sets of Genes in Susceptible Arabidopsis thaliana Plants. *Plant J* **2002,** *32,* 1–13.

18. Yao, Z. P.; Afonso, C.; Fenselau, C. Rapid Microorganism Identification with On-Slide Proteolytic Digestion Followed by Matrix-Assisted Laser Desorption/Ionization Tandem Mass Spectrometry and Database Searching. *Rapid Commun. Mass Spectrom.* **2002,** *16,* 1953–1956.

19. Yao, Z. P.; Demirev, P. A.; Fenselau, C. Mass Spectrometry-Based Proteolytic Mapping for Rapid Virus Identification. *Anal. Chem.* **2002,** *74,* 2529–2534.

20. Ayers, M.; Siu, K.; Roberts, E.; Garvin, A. M.; Tellier, R. Characterization of Hepatitis C Virus Quasispecies by Matrix-Assisted Laser Desorption/ Ionization Time-of-Flight (Mass Spectrometry) Mutation Detection. *J. Clin. Microbiol.* **2002,** *40,* 3455–3462.

21. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III. Direct Analysis of Protein Complexes Using Mass Spectrometry. *Nat. Biotechnol.* **1999,** *17,* 676–682.