A molecular scanner to automate proteomic research and to display proteome images

Binz, Pierre-Alain; Muller, Markus Johann; Walther, Daniel; Bienvenut, Willy-Vincent; Gras, Robin; Hoogland, Christine; Bouchet, Gérard; Gasteiger, Elisabeth; Fabbretti, Robert; Gay, Steven Daryl; Palagi Graf, Patricia; Wilkins, Marc; Rouge, Véronique; Tonella, Luisa [**and 7 more**]

How to cite

BINZ, Pierre-Alain et al. A molecular scanner to automate proteomic research and to display proteome images. In: Analytical chemistry, 1999, vol. 71, n° 21, p. 4981–4988. doi: 10.1021/ac990449e

This publication URL: https://archive-ouverte.unige.ch/unige:80769
Publication DOI: 10.1021/ac990449e

Anal. Chem. **1999,** *71,* 4981−4988

A Molecular Scanner To Automate Proteomic Research and To Display Proteome Images

Pierre-Alain Binz,*,†,‡ Markus Müller,† Daniel Walther,† Willy V. Bienvenut,‡ Robin Gras,† Christine Hoogland,† Gérard Bouchet,† Elisabeth Gasteiger,† Roberto Fabbretti,† Steven Gay,† Patricia Palagi,† Marc R. Wilkins,‡ Véronique Rouge,‡ Luisa Tonella,‡ Salvo Paesano,‡ Gérald Rossellat,‡ Abderrahim Karmime,‡ Amos Bairoch,† Jean-Charles Sanchez,‡ Ron D. Appel,† and Denis F. Hochstrasser‡

Swiss Institute of Bioinformatics, University Medical Center, rue Michel Servet 1, CH-1211 Geneva 4, Switzerland, and Central Clinical Chemistry Laboratory, Geneva University Hospital, rue Micheli-du-Crest 24, CH-1211 Geneva 14, Switzerland

Identification and characterization of all proteins expressed by a genome in biological samples represent major challenges in proteomics. Today's commonly used high-throughput approaches combine two-dimensional electrophoresis (2-DE) with peptide mass fingerprinting (PMF) analysis. Although automation is often possible, a number of limitations still adversely affect the rate of protein identification and annotation in 2-DE databases: the sequential excision process of pieces of gel containing protein; the enzymatic digestion step; the interpretation of mass spectra (reliability of identifications); and the manual updating of 2-DE databases. We present a highly automated method that generates a fully annotated 2-DE map. Using a parallel process, all proteins of a 2-DE are first simultaneously digested proteolytically and electrotransferred onto a poly(vinylidene difluoride) membrane. The membrane is then directly scanned by MALDI-TOF MS. After automated protein identification from the obtained peptide mass fingerprints using PeptIdent software (http://www.expasy.ch/tools/peptident.html), a fully annotated 2-D map is created on-line. It is a multidimensional representation of a proteome that contains interpreted PMF data in addition to protein identification results. This "MS-imaging" method represents a major step toward the development of a clinical molecular scanner.

Today's genome-sequencing projects provide a huge amount of information in the form of nucleotide sequences that are being stored in specific databases. In a first approach, this information can be interpreted in order to obtain the coded amino acid (AA) sequences of all potentially expressed proteins. In their active forms, proteins often differ from the predicted AA sequence as they can be processed or carry posttranslational modifications. Most of these modifications are not predictable from gene sequences. In fact, a single gene sequence may give rise to more than 10 structurally different proteins.[1] As an example, α-1-antitrypsin is known to exist in at least 22 different forms in the human plasma master image in the SWISS-2DPAGE database.[2] This yields by extrapolation between 500 000 and 1 million different protein forms expressed in the human. The description of a proteome,[3] involving the identification of all proteins contained in a biological sample, therefore represents a real experimental challenge. Methods involving high-resolution protein separation, parallelization of sample preparation, and automation of experimental processes and of database comparison, as well as powerful and specific visualization tools, need to be developed and integrated.[4,5]

Identifying a protein from a complex biological sample requires at least three steps. The protein is first isolated. Then very specific experimental attributes, such as peptide mass fingerprinting (PMF) or partial amino acid sequences, are determined. In the third step, identification is attempted by matching these attributes with those computed for all entries in a protein sequence database. The two-dimensional electrophoresis (2-DE) technique is a method of choice to separate with high resolution a large number of proteins in one single procedure, particularly when narrow-range pH gradients are used.[6] It provides a graphical representation of a proteome, where each protein form present in the so-called 2-DE map is represented by a spot or a series of spots and can be described by a p*I*, an apparent molecular weight, and an intensity-related value. Among different methods used routinely, the PMF approach is generally accepted to be currently by far the most effective and rapid way to identify proteins from a 2-DE gel. In this method, proteins are excised and proteolytically

* Corresponding author: (e-mail) Pierre-Alain.Binz@dim.hcuge.ch; (fax) +41-22-3727399.

† University Medical Center.

‡ Geneva University Hospital.

(1) Wilkins, M. R.; Sanchez, J. C.; Williams, K. L.; Hochstrasser, D. F. *Electrophoresis* **1996**, *17*, 830−8.

(2) Hoogland, C.; Sanchez, J. C.; Bairoch, A.; Hochstrasser, D. F.; Appel, R. D. *Nucleic Acids Res.* **1999,** *27*, 289−91.

(3) Wilkins, M. R.; Sanchez, J. C., Gooley, A. A.; Appel, R. D.; Humphery-Smith, I.; Hochstrasser, D. F.; Williams, K. L. *Biotechnol. Gene Eng. Rev.* **1995**, *13*, 19−50.

(4) Williams, K. L.; Hochstrasser, D. F. In *Proteome Research: New Frontiers in Functionnal Genomics*; Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F., Eds.; Springer: Berlin, 1997; pp 1−12.

(5) Hochstrasser, D. F. *Clin. Chem. Lab. Med.* **1998**, *36*, 825−36.

(6) Sanchez, J. C., Hochstrasser, D. H. In *Method in Molecular Biology: 2-D Proteome Analysis Protocol*; Link, A. J., Ed.; Humana Press: Totowa, NJ, 1998; Vol. 112, pp 227−33.

10.1021/ac990449e CCC: $18.00
Published on Web 10/06/1999

digested from protein spots. The resulting peptides are measured by mass spectrometry and then matched against a database of theoretical peptide mass fingerprints deduced from protein sequences. A score is calculated which represents the similarity between the experimental and the theoretical peptide masses. In principle, the protein with the highest score should result in a correct identification.

Various problems emerge with regard to the analysis of very complex biological samples such as human tissue. How can we attain to a reasonable throughput when performing a proteolytic digestion of all proteins after 2-DE? How can we reduce the number of manipulations required for the sample preparation before MS measurement? How can we simultaneously reduce the sizes of the samples and therefore increase their number? How can we handle the huge amount of experimental data and represent the result in a simple and comprehensive way? A number of solutions have been proposed to answer these questions.

Various approaches have been described to automate and accelerate the method. Traini et al.[7] proposed the use of a prototype robotics system to image and to excise a few hundred spots from a stained poly(vinylidene difluoride) (PVDF) blot. The protein samples were then enzymatically digested with an automated liquid handling system. The mass spectra of the peptide mass fingerprints were acquired using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) in automated mode. Proteins were identified using an automated interrogation software. Even though this approach is automated, the time-consuming digestion process is partially sequential and involves expensive sample handling, due to material costs. In addition, since the size of a sample is limited by the size of the excised spot, problems occur when overlapping spots are present on a gel.

To reduce sample handling and to decrease the analyzed sample size to that of the MALDI-TOF MS laser beam impact (a spot of a few tens of a micrometer in diamater), gels or membranes containing peptides or proteins have been used for direct MALDI-TOF MS measurements. Ogorzalek Loo et al.[8] measured protein masses directly from thin-layer isoelectrofocusing gels. Various types of membranes were also used as sample support for peptide or protein mass determinations, such as polyethylene,[9] nonporous polyurethane,[10,11] PVDF (Immobilon PSQ or Trans-Blot),[12−15] or charged membrane Immobilon CD.[12] Use of these sample supports allows the MS instrument to measure spectra separated by distances in the micrometer range. This opens the possibility to scan such a surface and create intensity images using the intensities of the MS signals and, therefore, to localize single peptides or proteins.[13,16,17]

To further increase the throughput of protein identification and to offer a flexible and powerful proteomic visualization tool, we designed a highly automated method that can create a fully annotated 2-D map starting from a 2-DE. This technological tool is called a "molecular scanner". It combines parallel methods for protein digestion and electrotransfers (using the one-step digestion−transfer (OSDT) or the double parallel digestion (DPD) techniques as described by Bienvenut et al.[18]) with peptide mass fingerprinting approaches to identify proteins directly from PVDF membranes, the surface of which is scanned with MALDI-TOF MS. Using a set of dedicated tools it allows us to create, analyze, and visualize a proteome as a multidimensional image. This provides the technological basis for the development of a clinical molecular scanner, which will be adapted and dedicated to medical diagnostics.[19]

EXPERIMENTAL SECTION

Materials and Reagents. IAV-trypsin membranes were prepared as described in Bienvenut et al.[18] Low-range SDS−PAGE standards and Trans-Blot PVDF membrane were purchased from Bio-Rad (Richmond, CA). Trifluoroacetic acid (TFA) and α-cyano-4-hydroxy-*trans*-cinnamic acid (ACCA) were purchased from Sigma (St. Louis, MO). Acetonitrile (AcCN) HPLC grade was purchased from Flucka (Buchs, Switzerland). Methanol (puriss pa) was purchased from Merck (Darmstadt, Germany). High-vacuum grease was purchased from Labofur GmbH (Bern, Switzerland).

Description of the Method. The method can be divided into four main sections (Figure 1).

(A) Separation and Digestion of the Proteins. One-dimensional separation of SDS−PAGE standards and mini 2-DE of human plasma were performed according to Laemmli[20] and Sanchez,[6] respectively. All proteins were proteolytically digested with trypsin and electroblotted onto a PVDF membrane, using OSDT parallel process as described by Bienvenut et al.[18] (Figure 1A). The collecting PVDF membrane thus contained sets of digestion products of all proteins, each of them localized at discrete positions on the surface. IAV−trypsin membrane was prepared as described in Bienvenut et al.[18] Where needed, PVDF membranes were stained with amido black after OSDT.

(B) Acquisition of the Peptide Mass Fingerprinting Data. Matrix solution made of 5 mg/mL ACCA in 50% AcCN−0.1% TFA or of 10 mg/mL ACCA in 70% MeOH was sprayed on the PVDF membrane until the membrane became wet. After air-drying, the membrane was stuck on a modified MALDI sample plate using high-vacuum grease. The stainless steel surface of the MALDI MS sample plate was flattened to allow the deposition of a 4×4 cm^2 PVDF membrane. An array of positions was defined on the membrane. The membrane was then scanned by the MS; i.e., a

(7) Traini, M.; Gooley, A. A.; Ou, K.; Wilkins, M. R.; Tonella, L.; Sanchez, J. C.; Hochstrasser, D. F.; Williams, K. L. *Electrophoresis* **1998**, *19*, 1941−9.
(8) Ogorzalek Loo, R.R; Mitchell, C.; Stevenson, T. I.; Martin, S. A.; Hines, W. M.; Juhasz, P.; Patterson, D. H.; Peltier, J. M.; Loo, J. A.; Andrews, P. C. *Electrophoresis* **1997**, *18*, 382−90.
(9) Blackledge, J. A.; Alexander, A. J. *Anal. Chem.* **1995**, *67*, 843−8.
(10) McComb, M. E.; Oleschuk, R. D.; Manley, D. M.; Donald, L.; Chow, A.; O'Neil, J. D.; Ens, W.; Standing, K. G.; Perreault, H. *Rapid. Commun. Mass. Spectrom.* **1997**, *11*, 1716−22.
(11) McComb, M. E.; Oleschuk, R. D.; Chow, A.; Ens, W.; Standing, K. G.; Perreault, H. *Anal. Chem.* **1998**, *70*, 5142−9.
(12) Schreiner, M.; Strupat, K.; Lottspeich, F.; Eckerskorn, C. *Electrophoresis* **1996**, *17*, 954−61.
(13) Eckerskorn, C.; Strupat, K.; Schleuder, D.; Hochstrasser, D.; Sanchez, J. C.; Lottspeich, F.; Hillenkamp, F. *Anal. Chem.* **1997**, *69*, 2888−92.
(14) Vestling, M. M.; Fenselau, C. *Biochem. Soc. Trans.* **1994**, *22*, 547−51.
(15) Fabris, D.; Vestling, M. M.; Cordero, M. M.; Doroshenko, V. M.; Cotter, R. J.; Fenselau, C. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1051−5.
(16) Caprioli, R. M.; Farmer, T. B.; Gile, J. *Anal. Chem.* **1997**, *69*, 4751−60.
(17) Stoeckli, M.; Farmer, T. B.; Caprioli, R. M. *J. Am. Soc. Mass. Spectrom.* **1999**, *10*, 67−71.
(18) Bienvenut, W. V.; Sanchez, J. C.; Karmime, A.; Rouge, V.; Rose, K.; Binz, P.-A.; Hochstrasser, D. F. *Anal. Chem.* **1999**, *71*, 4800−4807.
(19) Hochstrasser, D. F.; Appel, R. D.; Vargas, R.; Perrier, R.; Vurlod, J. F.; Ravier, F.; Pasquali, C.; Funk, M.; Pellegrini, C.; Muller, A. F.; et al. *MD Comput.* **1991**, *8*, 85−91.
(20) Laemmli, U. K. *Nature* **1970**, *277*, 680−5.

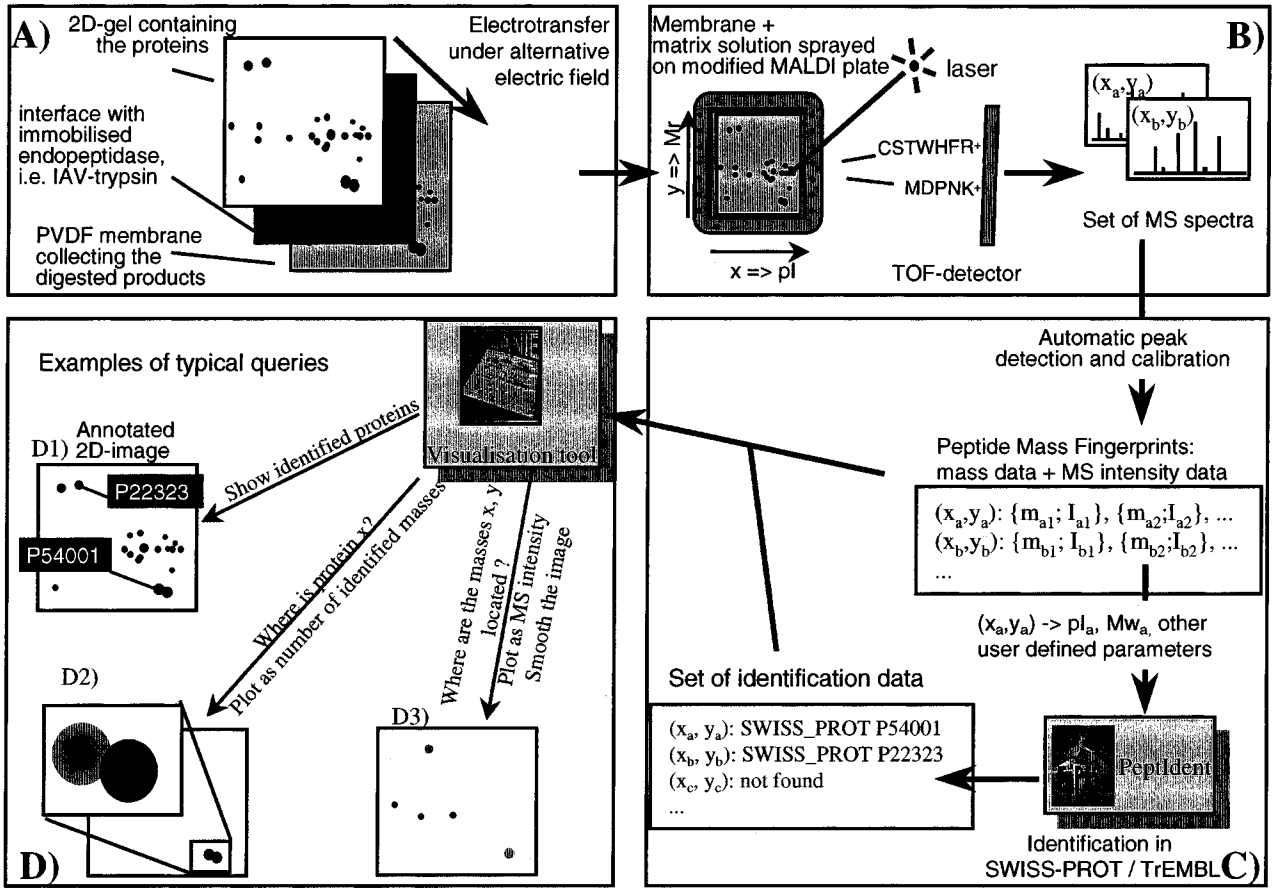

Figure 1. Scheme of the molecular scanner. (A) Parallel digestion and simultaneous electrotransfer of proteins from a 2-D PAGE using the DPD/OSDT method.[18] (B) MALDI-TOF MS scanning of PVDF collecting membrane after spraying with matrix solution. (x_i, y_j) refers to the position where MS spectra were measured on the PVDF membrane. (C) Identification procedure. The peak detection and mass calibration yields sets of PMF. The MS signal measured at each (x_i, y_j) coordinate is represented by its m/z value m_{ix} and its MS intensity I_{ix}. The x_i and y_i values are interpreted as pI and M_r values. The PMF data are submitted to PeptIdent. Identification results are collected together with the PMF data. (D) A visualization tool allows representation of the analyzed data in different forms. Three examples of typical queries and representations are described here. (1) An MS intensity image can be created that contains the identification data as database labels. It is generated in a Melanie readable format. (2) Another option allows one to search for a particular protein and to visualize it as an intensity plot. In this plot, the intensity represents the number of masses identified to belong to the protein at each position. (3) The program further allows one to search for a set of predefined masses and to generate an intensity image where the intensity represents the total intensity of the found mass peaks at each (x, y) position. This image can be smoothed if needed.

mass spectrum was acquired at each position of the array (Figure 1B). The distance between separate MS acquisition on the grid was constant for a given experiment (ranging between 0.2 and 0.5 mm). Mass spectra were acquired on a Voyager Elite MALDI-TOF mass spectrometer (Perseptive Biosystems, Framingham, MA) equipped with a 337-nm nitrogen laser and a delayed extraction device. The accelerating voltage was 18 kV, a delayed extraction parameter of 140 ns was selected, and the m/z value selected for the low-mass gate was generally 850. Laser power was set about 20% above threshold. The diameter of the laser beam on the membrane was about 100 μm. Between 40 and 100 spectra were accumulated, depending on the amount of material analyzed. The set of coordinates of the laser shots on the sample plate and the naming of the MS files were controlled by special software to overcome limitations in the maximum number of coordinates and spectra allowed by the Voyager 4.03 acquisition software.

(C) Processing of the MS Data and Protein Identification. A flexible and interactive tool was developed to automatically treat all MS data consecutively and to perform the various steps of the analysis, starting with peak detection and calibration (Figure 1C). The positions on the sample plate were converted to apparent molecular mass (M_r) and pI values. The PMF data of all spectra, together with the calculated pI and M_r and other user-defined parameters (such as mass tolerance, chemical modifications considered, and species taken into account), were automatically sent over the Internet for protein identification to PeptIdent, a PMF identification tool developed in Geneva[21] and available at the ExPASy server (http://www.expasy.ch[22]).

(D) Analysis of the Results: Creation of Virtual Maps. The identification results of PeptIdent were represented as an annotated image. All outputs of PeptIdent were acquired and stored in a modified format. The program generated a first virtual, annotated "2-D map", a 3-D image where the x and y coordinates related to pI and M_r values, respectively. The z values were represented in gray scale and reflected the intensity of the MS

(21) Binz, P.-A.; Wilkins, M. R.; Gasteiger, E.; Bairoch, A.; Appel, R. D.; Hochstrasser, D. F. In *Microcharacterization of Proteins*, 2nd ed.; Kellner, R., Lottspeich, F., Meyer, H. E., Eds.; Wiley-VCH: New York, 1999; pp 277−300.

(22) Appel, R. D.; Bairoch, A.; Hochstrasser, D. F. *Trends Biochem. Sci.* **1994**, *19*, 258−60.

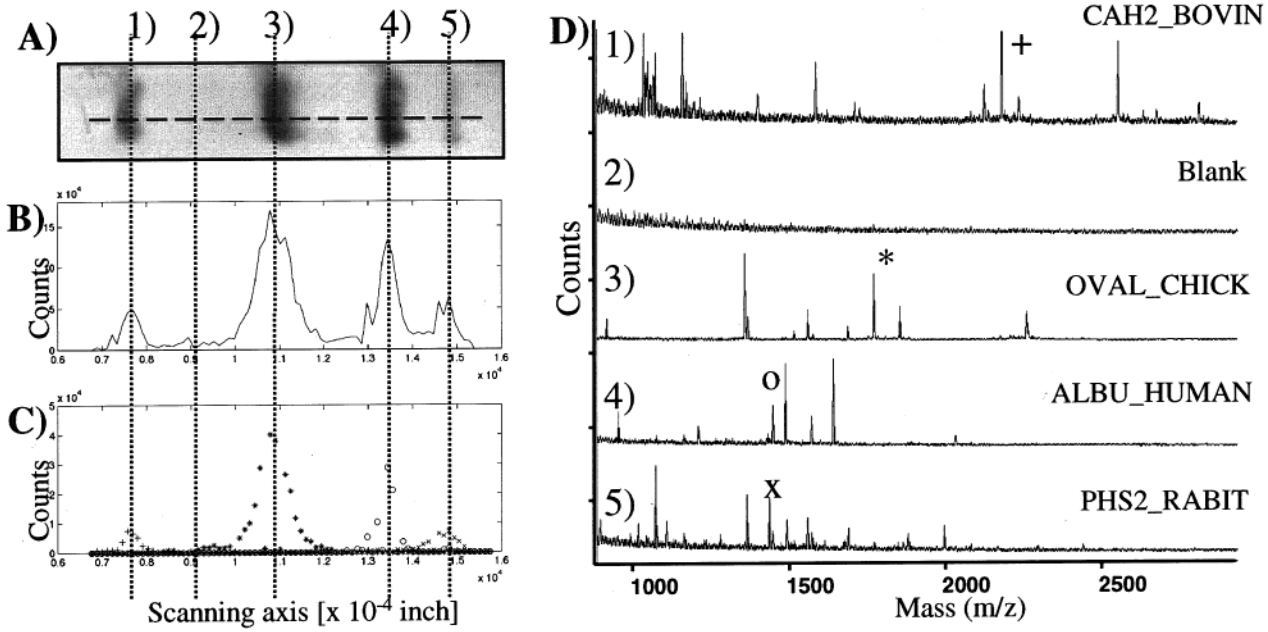

Figure 2. Result of a one-dimensional scan. Low-range standard proteins were separated with 1-DE and then treated with the OSDT procedure. The one-dimensional MS scan was performed on an amido black-stained PVDF membrane along the longitudinal dashed line of the PVDF image A. The plots B and C are MS intensity profiles. They represent the intensity of MS signals (number of counts from the MALDI-TOF MS detector) as a function of the position on the membrane (*x* unit is 10^{-4} in.). In the lined plot, only *m*/*z* values bigger than 1100 were considered. The intensities due to the two internal standards at mass 1498.82 ± 1 and 2095.08 ± 1 Da were excluded. MS spectra measured at the intensity maximums of the plot (positions 1 and 3−5) in (A) and in the background (position 2) in (A) are shown in (D). From these spectra, the four standard proteins could be identified with PeptIdent as labeled with their SWISS-PROT ID identifiers on the respective PMF MS spectra. Plot C is made of single-ion intensity profiles. The selected ions were chosen from the set of peptides specifically matching for one of the four identified proteins. Values of 2198.2 (+), 1774.0 (*), 1440.0 (○), and 1426.7 (×) *m*/*z* were considered with a window of ±1 Da. Matrix solution was 10 mg/mL ACCA in 70% methanol. A total of 110 spectra were accumulated 100 times on a total scanning length of 4.4 cm.

spectra, as defined by the sum of the intensities of the MS signals in the considered MS spectra. The range of m/z values to be considered was predefined. The intensity scale was chosen linear or logarithmic and the image was smoothed in some cases. The image file was stored in a graphical format that can be read by the Melanie 2-DE image analysis software package.[23] The image also contained the identification results, which can be highlighted as labels in Melanie (Figure 1D). The number of distinct attributes contained in the image reflects the number of dimension the image virtually contains. These are as follows: p*I*, M_r, identification labels (SWISS-PROT or TrEMBL AC numbers, ID labels), peptide masses, and MS intensities. Then for all potentially identified proteins, the annotations from PeptIdent (number of missed cleavages, annotated modifications, chemical modifications of Cys and Met residues, peptide sequences) are also available.

From all the data contained in this multidimensional image, the user can choose to filter and visualize only particular aspects (Figure 1D). Proteins or peptides can be searched on the image by filtering part of the total information. Thus, a protein can be visualized by the positions where it has been identified. The *z* intensity can be a binary (black/white for present/absent, respectively) or a gray level. The darkness represents then either the number of peptides found to match the protein in the identification process using PeptIdent or the sum of the MS intensities of the peptide masses matching the queried protein. Instead of searching for a protein, the user can specify and visualize a set of peptide masses. In this case, the image intensity scale can be defined from the number or the MS intensities of the masses detected out of the chosen list (Figure 1D).

RESULTS

Representation of the Analysis of a One-Dimensional Scan of 1-DE. To set up the various experimental parameters of the method, we have performed a number of analyses on a protein mixture of molecular weight standards separated on SDS−PAGE and treated by the DPD or OSDT method. The selected collecting membrane was PVDF. The membranes were initially stained with amido black to visualize the positions of the peptide fingerprinting bands (Figure 2A). Matrix solution was sprayed on the whole surface of the membrane. About 1.5 mL was used to spray a 4.4 × 0.5 cm^2 PVDF membrane. The volume of matrix solution effectively deposited on the membrane was estimated to be 1−2 μL/mm^2. After air-drying, the membrane was scanned in one dimension with MALDI-TOF MS. The summed intensity of the detected MS signals, for a given mass range, was plotted against the axis coordinate along the membrane (Figure 2B). The intensity of the MS spectra obtained from the stained membrane varied along the scanning axis. The positions of the four maximum intensities on the MS profiles correlated with the positions of the four stained bands. The MS profiles revealed distinctly resolved bands, thus suggesting a conserved separation of the peptide fingerprints during the DPD or OSDT step and during matrix deposition. The peptide containing areas are separated by blank areas, showing no MS intensity (position 2 in Figure 2A and MS spectrum in Figure 2D). No significant broadening of the band

(23) Appel, R. D.; Palagi, P. M.; Walther, D.; Vargas, J. R.; Sanchez, J. C.; Ravier, F.; Pasquali, C.; Hochstrasser, D. F. *Electrophoresis* **1997**, *18*, 2724−34.

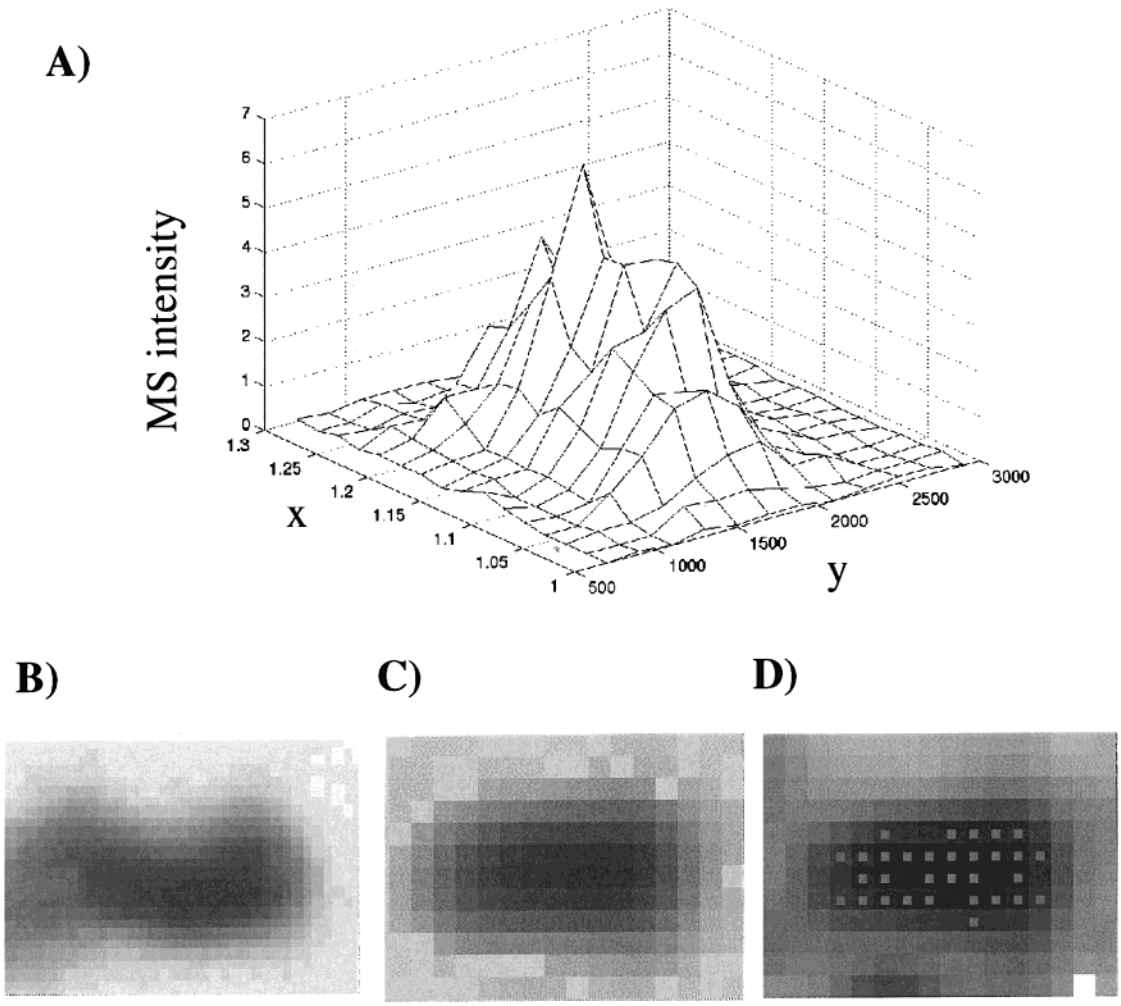

Figure 3. Two-dimensional MS scan of 1-DE: the soybean trypsin inhibitor band. From the same membrane as in Figure 2, a piece of 1.1 × 0.9 cm^2 was cut around the soybean trypsin inhibitor band and sprayed with a 10 mg/mL ACCA solution in 70% methanol. An array of 16 × 12 points was defined around the center of the band, with distance between spots of 500 μm. (A) 3-D MS intensity profile. All *m*/*z* higher than 1100 Da were considered to create the smoothed image. (B) amido black-stained image. (C) MS intensity image. (D) MS intensity image, plotted in a logarithmic scale. The white dots represent the positions where the ITRA_SOYBN was unambiguously identified with a minimum of 5 *m*/*z* matching values with PeptIdent.

was observed in comparison to the corresponding undigested electrotransferred stained protein bands. This would suggest that the peptides are not diffusing significantly during the digestion, during the transfer process, and on the membrane upon matrix apposition. From this membrane, protein identification was performed from each MS spectrum at maximum of MS intensity (Figure 2D). All four proteins (positions 1, 3, 4, and 5 in Figure 2A) could be unambiguously identified in SWISS-PROT using PeptIdent.

Different matrix solvents and apposition techniques of the matrix solution were compared. As an example, the methanol-containing matrix solution wetted the surface of the PVDF in a much more homogeneous manner than the acetonitrile-containing solution. More intense MS spectra and more homogeneous MS profiles were obtained with the methanolic matrix solution (data not shown).

Representation of the Analysis of a Two-Dimensional Scan from a Single Band of 1-DE. Similarly, Figure 3 shows the result of a 2-D scan and its interpretation performed on a single protein band. Low-range SDS standards were separated on 1-DE, processed with the OSDT method, and the PVDF was stained with amido black. A 0.8 × 0.6 cm^2 piece of membrane containing the digested soybean trypsin inhibitor was scanned with a resolution of 0.5 mm. The amido black-stained image of the band after OSDT on a PVDF membrane (Figure 3B) was compared with a MS intensity image calculated from all MS spectra (Figure 3C and D), where only *m*/*z* values higher than 1100 Da were considered (there are disturbing matrix signals below 1100 Da). A 3-D plot (Figure 3A) representing the absolute intensities as function of the *x*, *y* position on the membrane was created with Matlab 5.2 (MathWorks, Inc., 24 Prime Parkway, Natick MA), and a MS intensity image was calculated under the same conditions (Figure 3C). The MS intensity profile followed relatively smoothed curves, suggesting a relatively homogeneous quality of the matrix crystallization. This approach highlighted the possibility of creating intensity profiles with MS intensity values and describing spot areas. Addition of a matrix solution did not seem to lead to any significant diffusion of peptides on the matrix surface. The sensitivity of such an MS staining was equal or better than that of amido black staining of the transferred protein. The intensity

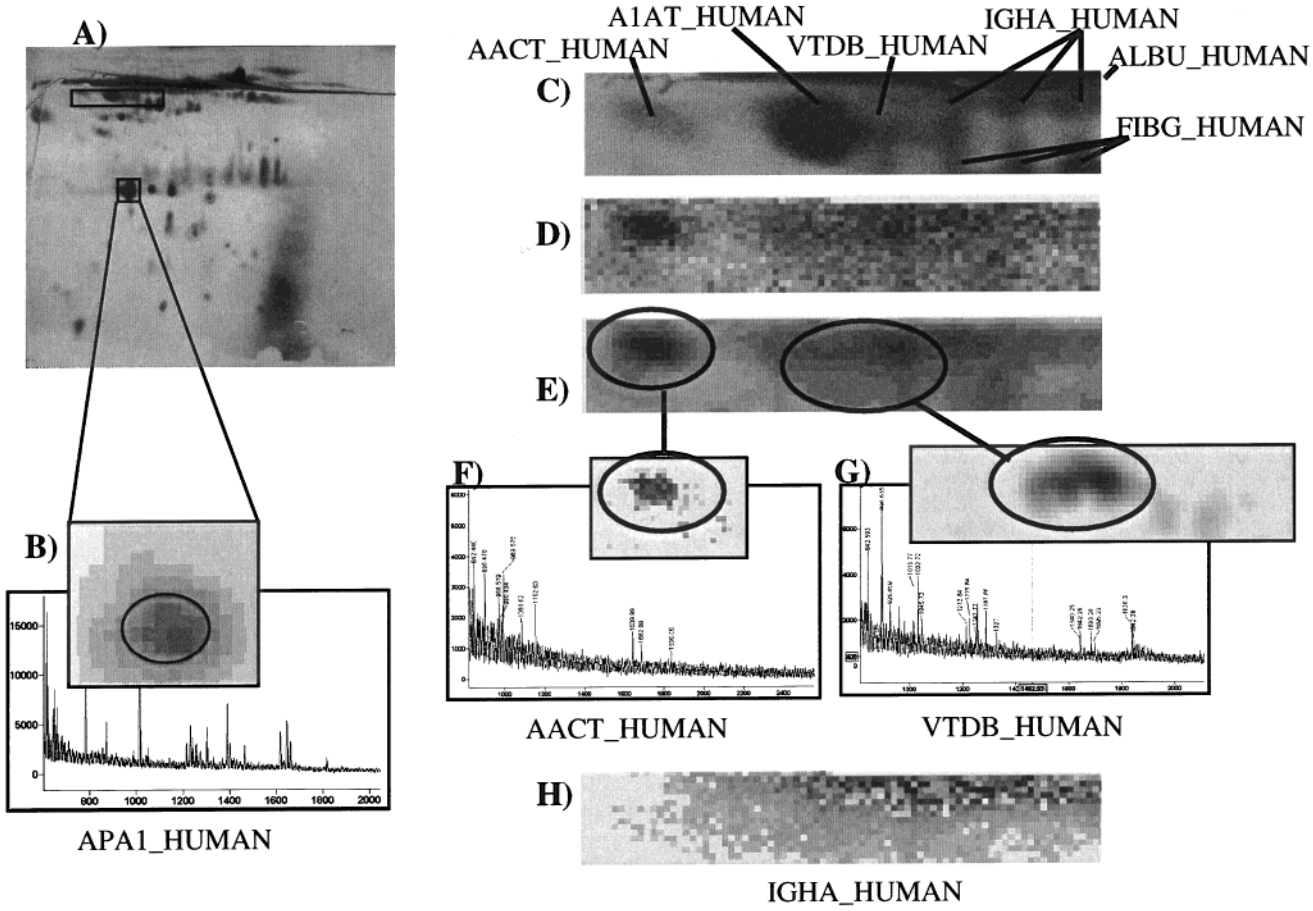

Figure 4. 2-D scan of a plasma mini-2-DE after OSDT. Possibilities of extracting proteomic information from a 2-D MS scan. A 250-μg sample of human plasma was separated on mini-2-DE. Proteins were digested and transferred on PVDF using the OSDT procedure. (A) Image of the membrane stained with amido black. (B) Smoothed MS intensity image of the region where the apolypoprotein-A1 (SWISS-PROT AC P02647) was identified. The circle indicates the size of the spot visible by amido black staining. This shows that MS intensities are still detected where amido black staining is blank. (C) Enlargement of the image A, showing the positions corresponding to the proteins identified on the area. The labels are the SWISS-2DPAGE ID names: AACT_HUMAN (α-antichymotrypsin, AC P01011), VTDB_HUMAN (vitamin D binding protein, AC P02774), ALBU_HUMAN (serum albumin, AC P02768), A1AT_HUMAN (α-1-antitrypsin, AC P01009), FIBG_HUMAN (fibrinogen γ α-chain, AC, P02679), IGHA_HUMAN (immunoglobulin α chain, AC P99002). (D) Raw MS intensity image of the region zoomed from the amido black image. (E) Same image, but smoothed. (F) In this MS image, the intensity is related to the number of MS signals used to identify the query protein, i.e., AACT_HUMAN (SWISS-PROT AC P01011). One of the MS spectra of the spot is also shown. (G) Smoothed MS image, where the intensity represents the summed intensity of the MS signals used to identify VTDB_HUMAN (SWISS-PROT AC P02774) at each pixel. One of the MS spectra allowing the identification of the protein is shown. (H) MS intensity image of the same region, where only MS signals belonging to immunoglobulin α chain (IGHA_HUMAN) peptides are considered.

images obtained from these values show that the experimental noise, which may be due to inhomogeneity of the matrix crystallization quality on the membrane or to the low doping of the matrix by the analytes, was limited. From the 192 acquired spectra the soybean trypsin inhibitor (SWISS-PROT ID, ITRA_SOYBN; AC, P01070) could be unambiguously identified 31 times (white dots in Figure 3D) with a minimum of 5 matched m/z values, representing a contiguous region in the center of the 2-D image.

Identification by a Two-Dimensional Scan of Human Plasma Proteins Separated by 2-DE. From a human plasma sample separated on a mini-2-DE and transferred using the OSDT method, we have cut a section of the collecting PVDF membrane around the amido black-stained spot of the apolipoprotein-A1 (SWISS-PROT AC P02647). The smoothed MS intensity image is shown in Figure 4B. From the 195 MS files measured from pixels at 380-μm steps, 77 yielded the apolipoprotein-A1 as the identified protein with a minimum of 5 peptide masses matched. The image shows that MS intensities are detected around the area corresponding to the amido black-stained visible surface. The observed signal corresponds to the most intense peaks of the spectra from which apolipoprotein A1 was identified, thus suggesting that "MS staining" is more sensitive than the amido black staining method.

Various representations from a more complex protein mixture are shown in Figure 4. Another section of the amido black-stained PVDF membrane, obtained after OSDT treatment of a mini-2-DE gel from a human plasma sample, was also scanned with the MALDI-TOF MS. Its size was about 1.83×0.37 cm^2 (Figure 4A and C). A total of 890 spectra were measured at 400-μm steps. The chosen section of the scanned PVDF membrane contained a set of overlapping spots and trains of spots, as deduced from the known repartition of identified proteins in the human plasma image in the SWISS-2DPAGE database (http://www.expasy.ch/cgi-bin/map2/big?PLASMA_HUMAN). It also contained contami-

nation from the adjacent and very abundant albumin, centered above the upper right corner of the excised PVDF surface. In addition, a probably high number of proteins, whose sequences are unknown in databases, were also present in this sample. The MS intensity image reveals a continuous background of MS signals, represented by a gray background (Figure 4D,E). This suggests that a lot of peptide material is measured on the whole surface and that the protein spots are not isolated entities. The analysis tool allowed, however, filtering of this complex feature and gave the possibility of extracting spots corresponding to single proteins. Protein spots can therefore be isolated from chemical noise. As examples, parts F and G of Figure 4 show two different regions of the image from which the α-1-antichymotrypsin (SWISS−PROT ID, AACT_HUMAN; AC, P01011) and the vitamin D binding protein (SWISS−PROT ID, VTDB_HUMAN; AC, P02774) were identified and visualized as two isolated spots, respectively. The figure shows two possibilities of representing specific "intensity values". In the first possibility, (the AACT_HUMAN spot in Figure 4F), the intensity of each pixel is proportional to the absolute number of peptides identified for a given protein. In Figure 4F, the protein is AACT_HUMAN. This intensity is somehow related to the confidence of identification. In the second possibility, i.e., the VTDB_HUMAN spot in Figure 4G, the intensity is proportional to the sum of the MS intensities of the peptides peaks identified for a given protein. In Figure 4G, the protein is VTDB_HUMAN. The intensity was then smoothed. Here the intensity is more related to the protein concentration than the previous one. Therefore, we have graphically extracted the contribution of the two proteins from the total MS intensity image shown in Figure 4C and D. Some proteins are highly abundant, and are present in multiple forms, such as the immunoglobulin α chains. They are detected on a large part of the area, thus yielding chemical noise for other proteins (Figure 4H). A number of proteins were clearly identified from this sample, and their relative positions on the membrane correlated with those identified in the human plasma master gel in the SWISS-2DPAGE database (Figure 4C).

DISCUSSION

The technique presented, known as a molecular scanner, provides a powerful tool for proteomics research. First, it is a high-throughput method dedicated to protein identification using peptide mass fingerprinting or other methods in the future and applied to the entire 2-DE. It uses a parallel method of protein digestion. Thus, in one experimental step, thousands of proteins can be chemically processed or digested simultaneously, under identical experimental conditions. The obtained sample can be directly used for MS measurements. This method limits losses of material caused by sample manipulation. The size of each MS sample is reduced to the size of the laser beam used in the MALDI-TOF MS, i.e., about 10^{-2} mm^2. A single protein 2-DE spot can therefore be represented by more than 100 spectra. Second, the PMF analysis is fully automated and can be modularly modified at any step, i.e., choice of the peak detection algorithm, of the calibration procedure, of the masses considered for identification, of the arguments sent to PeptIdent, and of the image representation.

The molecular scanner provides virtual images which can be considered as graphical projections of an automatically generated proteomic database. The database can be searched by protein identifiers (i.e., protein "name") or by mass-related identification results. The user can choose to visualize a single protein by searching the positions where the protein has been identified. As the position of a set of masses can be searched, a protein can be visualized as a function of the number and/or the intensity of MS signals matched by PeptIdent for this protein. Where a protein yields a train of spots on a 2-DE gel, the spot corresponding to one particular form of the protein can be isolated by searching a specific peptide mass in the spectra. This allows the systematic analysis of posttranslational modifications. In this respect, all PeptIdent results could be used as input data for a characterization step using FindMod.[24] FindMod is a tool that interprets unannotated MS signals for a given protein and PMF data. It looks, by mass difference, for the occurrence of posttranslational modifications using a set of intelligent rules as well as for potential amino acid substitutions. It can therefore be systematically linked to PeptIdent; i.e., after the identification step, it helps to further characterize and discriminate all spots of a train. In the future, different potential posttranslational modifications will be automatically highlighted in various colors on the image obtained by this scanner.

The high resolution obtained by the MS scanning becomes particularly useful when overlapping spots occur. This can be interpreted as a mixture of proteins. Reconstitution of intensity envelopes from peptide mass fingerprinting allows us to discriminate the two or more overlapping spots. Then one or the other spots can be visualized by choosing the peptide masses specific to this particular protein form in order to create an image or they can be represented by different coloring systems.

As an additional feature, neither the gel nor the PVDF membrane needs to be chemically stained. The MS intensity acts as a "coloring" agent. Since spots can be localized, the image can therefore be compared, aligned, and matched with other gel images or PVDF image stained with conventional methods.

As for chemical staining methods, the intensity of the MS signals are proportional neither to the amount of protein loaded nor to the amount of amino acid contained in the different spots. This relies on the desorption process and on the ionization yields. Thus, the intensity of the MS signals only partially correlates with the intensity of an amido black staining (see Figure 3 of the accompanying paper[18]) or with the absolute amount of material. In the 1-D scan (Figure 2), the SDS gel was loaded with 1 μg of each protein. Therefore, no estimation of protein amount can be deduced from a single MS image. However, comparative studies may be performed between several MS images in cases where identical spots are compared.

The illustrated experiments gave a preliminary idea of the sensitivity of the method. In 1-D experiments, we have loaded 1 μg of each molecular weight standard, which corresponds to about 10−33 pmol of proteins (see Figure 2). The size of the bands, visible on the control PVDF membrane, i.e., membrane obtained without protein digestion, and stained with amido black, were about 15 mm^2. All proteins could be identified very clearly. The sensitivity here was 66 ng/mm^2, respectively, of 0.66−2 pmol/mm^2. As the area of a protein spot on a mini-2-D gel covers about

(24) Wilkins, M. R.; Gasteiger, E.; Gooley, A. A.; Herbert, B. R.; Malloy, M. P.; Binz, P.-A.; Ou, K.; Sanchez, J.-C.; Bairoch, A.; Williams, K. L.; Hochstrasser, D. F. *J. Mol. Biol.* **1999**, *289*, 645−657.

this size, one could extrapolate that the detection limit for a clear identification lays around the low-picomole range today, with no optimization of the method. Another experiment was performed by loading 0.2 μg of each molecular weight standard. All proteins could be identified, but on a fewer number of pixels (not shown). Moreover, one single MS measurement covers about 10^{-2} mm^2. Although the efficiency of peptide extraction, cocrystallization, and ionization processes is not known, every identification was performed on about 10 fmol of initial protein. Note that these calculations are to give a rough estimate of the current sensitivity of the identification process. It is also to be noticed that spectra measured at positions neighboring pixels of identified proteins still contain detectable peaks. The sensitivity of peptide peak detection is therefore higher than the sensitivity of protein identification.

An interesting observation has to be mentioned. When spectra measured after an in-gel digestion and after a membrane scan are compared, differences are noticeable in term of peptides detected. Even if most of the mass signals are present in both cases, the intensity of the signals can differ strongly. Some signals are present in the in-gel spectrum and absent in the scanned spectrum.[18] Unfortunately, there is no evident relationship between the presence and intensity of a signal and a physicochemical property of the corresponding peptides. However, most of the peptides detected with the two methods are perfectly digested peptides and generally cover similar sequence percentage. More detailed studies will have to be undertaken.

As MS information represents an additional specific property for a spot, two images containing MS information could be aligned by matching their mass spectra and/or their resulting identifications in the Melanie 2-DE image analysis software. This procedure could replace and/or confirm manual and software-based alignment of matched gel images.

The development of the molecular scanner required us to develop and to integrate high-throughput methods for sample preparation and analysis. Specific bioinformatics tools had to be created as well. The molecular scanner was designed as a set of interconnecting modules, which can be exchanged and modified in a very flexible manner. It can therefore easily be adapted for improvements and modifications. The current bottleneck of this technique is the time necessary to scan the membrane with the mass spectrometer. Without optimization, the MS scanning time of a 4 $\times$ 4 cm^2 surface is about 55 h at a 0.4-mm resolution with 64 laser shots/position. This means that a full 16 $\times$ 16 cm^2 membrane would require, under the same conditions, more than 36 days of continuous measurements and about 40 Gb of memory to archive the raw data. As people tend to stretch the p*I* axis using narrow pH gradient strips in the first dimension, this would increase the separation power of the protein spots, but increase the measurement time needed. To accelerate the acquisition rate of the MS spectra, limited currently by the 3-Hz frequency of the laser and by a fixed number of laser shots per pixel, one should at least be able to software-control the number of required laser shots, i.e., to skip acquisition when spectra are empty or where the signal-to-noise ratio is above a given threshold. This may gain a factor of 2−5. Due to ion statistics, it is difficult to reduce drastically the number of laser shoots per pixel. As the detector is inactive at least 99% of the time, the acquisition frequency should be increased, either by a increase of the laser repetition rate or by the use of multiple lasers at neighboring positions on the membrane. As time is required to allow relaxation of the crystals between two laser shoots, there is a physical limitation of the pulse rate alone. As the specificity of a protein identification strongly depends on the mass accuracy, efforts can be also focused on the comparison of mass patterns in neighboring pixels. Finally, as MS technology develops, we anticipate that the full scan of a 10 $\times$ 10 cm^2 mini-2-DE gel will be performed in a few hours.

CONCLUSION

In medicine, the development of computer-assisted tomography methods allowed visualization of the complexity of the human body as a volume of anatomically related organs and tissues. The cellular components of a tissue can today be described using immunohistology and immunocytology. There is an obvious need to describe the protein content of a cell or of a biological fluid. The molecular scanner allows us to analyze many proteins in such a complex system. It reports, at the molecular dimension, the complexity of protein content. The presentation of a proteome as a searchable database, which can be visualized as user-defined 3-D images, provides a powerful tool for comparative analysis in proteomics. The method, initially starting from 2-DE separation of proteins, can be adapted to other fields such as protein chips or other multidimensional separation methods. It can also be applied in clinical diagnostics where modifications occurring to proteins, i.e., mutations and changes in posttranslational modifications, have to be monitored. These changes may be observed as changes in the PMF patterns, although they may not influence the migration of the protein itself. In addition to the presented approach, high-throughput MS/MS sequencing methods[25,26] or chip technology could represent complementary features. They are yielding additional information and provide a huge amount of data to be analyzed through visualization methods, such as the one proposed here. Finally, this technique allows combination with additional types of analysis. The same surface can be reused for new analysis, such as an MS scan under different conditions or with another laser, i.e., an IR laser.[13] In the case of particularly interesting spots, one can use the known coordinates of the location of the spot to perform additional chemistry on this particular area. The spot of interest can similarly be cut using a dedicated excision system to be submitted to further analyses, such as MS/MS. The molecular scanner is therefore a tool that can be fully integrated in any more general proteomic analysis process.

ACKNOWLEDGMENT

We are deeply grateful to the Helmut Horten Foundation for its financial support. This work was also supported by the Swiss National Fund for Scientific Research (Grants 32-49314.96 and 31-52974.97) and the Montus Foundation. We are also very thankful to Dr. Keith Rose for his technical and critical help.

Received for review April 27, 1999. Accepted September 2, 1999.

AC990449E

(25) Gevaert, K.; De Mol, H.; Verschelde, J. L.; Van Damme, J.; De Boeck, S.; Vandekerckhove, J. *J. Protein Chem.* **1997**, *16*, 335−42.

(26) Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. *Nature* **1996**, *379*, 466−9.