MALDI Quadrupole Time-of-Flight Mass Spectrometry: A Powerful Tool for Proteomic Research

ARTICLE *in* ANALYTICAL CHEMISTRY · JUNE 2000

Impact Factor: 5.64 · DOI: 10.1021/ac9913659 · Source: PubMed

4 AUTHORS, INCLUDING:

Werner Ens
University of Manitoba
144 PUBLICATIONS **5,806** CITATIONS

SEE PROFILE

Kenneth G Standing
University of Manitoba
155 PUBLICATIONS **6,221** CITATIONS

SEE PROFILE

Anal. Chem. **2000,** *72,* 2132−2141

MALDI Quadrupole Time-of-Flight Mass Spectrometry: A Powerful Tool for Proteomic Research

Andrej Shevchenko,*,† Alexander Loboda,‡ Anna Shevchenko,† Werner Ens,‡ and Kenneth G. Standing‡

Peptide and Protein Group, European Molecular Biology Laboratory (EMBL), 69012 Heidelberg, Germany, and Department of Physics, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada

A MALDI QqTOF mass spectrometer has been used to identify proteins separated by one-dimensional or two-dimensional gel electrophoresis at the femtomole level. The high mass resolution and the high mass accuracy of this instrument in both MS and MS/MS modes allow identification of a protein either by peptide mass fingerprinting of the protein digest or from tandem mass spectra acquired by collision-induced dissociation of individual peptide precursors. A peptide mass map of the digest and tandem mass spectra of multiple peptide precursor ions can be acquired from the same sample in the course of a single experiment. Database searching and acquisition of MS and MS/MS spectra can be combined in an interactive fashion, increasing the information value of the analytical data. The approach has demonstrated its usefulness in the comprehensive characterization of protein in-gel digests, in the dissection of complex protein mixtures, and in sequencing of a low molecular weight integral membrane protein. Proteins can be identified in all types of sequence databases, including an EST database. Thus, MALDI QqTOF mass spectrometry promises to have remarkable potential for advancing proteomic research.

Recent advances in mass spectrometric technology together with progress in genomic sequencing and bioinformatics have dramatically changed the position of mass spectrometric methods in a general strategy of molecular and cell biology research. Expanding from its previous role as a "mining tool" for providing protein sequence information,[1,2] mass spectrometry is now becoming deeply integrated into the process of functional characterization of biologically important genes.[3−6] Cutting edge research in the field of functional proteomics, such as dissection of multiprotein complexes,[7−9] quantitative mapping of cellular proteins,[10] or deciphering protein interaction networks,[11] increasingly requires the development of versatile high-throughput techniques for identification of proteins at the femtomole level.

High mass accuracy MALDI peptide mapping is an already recognized high-throughput technique that provides identification of dozens of gel-separated proteins per day at the low-picomole−subpicomole level.[12,13] The technique has nearly always been applied in time-of-flight (TOF) mass spectrometers, where the introduction of delayed extraction (time-lag focusing)[14−17] has increased the accuracy of mass measurements to the tens of parts per million (ppm) level. Such high mass accuracy boosts the specificity of database searching so that unambiguous identification of the protein can usually be achieved if four to six peptide masses are determined with better than 50 ppm accuracy.[18,19] The database searching algorithms are fairly straightforward,[20−24] and the entire analytical approach lends itself to automation, from the

* Corresponding author: Peptide & Protein Group, European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69012 Heidelberg, Germany. (Phone) +49 6221 387 224. (Fax) +49 6221 387 306. (E-mail) shevchenko@EMBL-Heidelberg.de.

† European Molecular Biology Laboratory (EMBL).

‡ University of Manitoba.

(1) Yates, J. R.; Eng, J. K.; McCormack, A. L. *Anal. Chem.* **1995**, *67*, 3202−3210.
(2) Patterson, S. D.; Aebersold, R. *Electrophoresis* **1995**, *16*, 791−814.
(3) Yates, J. R. *J. Mass Spectrom.* **1998**, *33*, 1−19.
(4) Roepstorff, P. *Curr. Opin. Biotechnol.* **1997**, *8*, 6−13.
(5) Haynes, P. A.; Gydi, S. P.; Figeys, D.; Aebersold, R. *Electrophoresis* **1998**, *19*, 1862−1871.
(6) Neubauer, G.; Wilm, M. *Curr. Opin. Mol. Therap.* **1999**, *1*, 695−701.
(7) Neubauer, G.; King, A.; Rappsilber, J.; Calvio, C.; Watson, M.; Ajuh, P.; Sleeman, J.; Lamond, A.; Mann, M. *Nat. Genet.* **1998**, *20*, 46−50.
(8) Zachariae, W.; Shevchenko, A.; Andrews, P. D.; Ciosk, R.; Galova, M.; Stark, M. J.; Mann, M.; Nasmyth, K. *Science* **1998**, *279*, 1216−1219.
(9) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. *Nat. Biotechnol.* **1999**, *17*, 676−682.
(10) Blackstock, W. P.; Weir, M. P. *Trends Biotechnol.* **1999**, *17*, 121−127.
(11) Shevchenko, A.; Zachariae, W.; Shevchenko, A. *Biochem. Soc. Trans.* **1999**, *27*, 549−554.
(12) Traini, M.; Gooley, A. A.; Ou, K.; Wilkins, M. R.; Tonella, L.; Sanchez, J. C.; Hochstrasser, D. F.; Williams, K. L. *Electrophoresis* **1998**, *19*, 1941−1949.
(13) Berndt, P.; Hobohm, U.; Langen, H. *Electrophoresis* **1999**, *20*, 3521−3526.
(14) Brown, R. S.; Lennon, J. J. *Anal. Chem.* **1995**, *67*, 1998−2003.
(15) King, T. B.; Colby, S. M.; Reilly, J. P. *Int. J. Mass Spectrom. Ion Processes* **1995**, *145*, L1−L7.
(16) Vestal, M. L.; Juhasz, P.; Martin, S. A. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1044−1050.
(17) Whittal, R. M.; Li, L. *Anal. Chem.* **1995**, *67*, 1950−1954.
(18) Jensen, O. N.; Podtelejnikov, A.; Mann, M. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1371−1378.
(19) Clauser, K. R.; Baker, P.; Burlingame, A. L. *Anal. Chem* **1999**, *71*, 2871−2882.
(20) Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011−5015.
(21) James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 58−64.
(22) Mann, M.; Højrup, P.; Roepstorff, P. *Biol. Mass Spectrom.* **1993**, *22*, 338−345.
(23) Pappin, D. J. C.; Højrup, P.; Bleasby, A. J. *Curr. Biol.* **1993**, *3*, 327−332.
(24) Yates, J. R.; Speicher, S.; Griffin, P. R.; Hunkapiller, T. *Anal. Biochem.* **1993**, *214*, 397−408.

10.1021/ac9913659 CCC: $19.00
Published on Web 03/25/2000

acquisition of MALDI spectra to the searching against a sequence database.[25]

Despite its power, this technique suffers from some problems. At the subpicomole level (silver-stained protein bands), the efficiency of protein digestion decreases. Very few peptides may then be detectable by MALDI,[26,27] and the resulting low sequence coverage may allow only tentative protein identification. In addition, MALDI peptide mapping is not an effective tool for mining an expressed sequence tag (EST) database,[28] since EST clones are usually too short (~300 bp) and prone to errors and frame shifts to produce a statistically reliable hit. Recently it has also become apparent that MALDI peptide mapping alone may not allow identification of proteins in genomic sequence databases because of insufficient accuracy (40−80%) of the exon−intron prediction algorithms.[29]

Generally speaking, MALDI peptide mapping may operate efficiently when a statistically significant number of peptide peaks can be measured, but it possesses virtually no power to characterize an individual peak in the spectrum. Until recently, characterization of individual peptide ions in MALDI TOF instruments has only been possible with the postsource decay (PSD) technique.[30] However, the complexity of peptide fragmentation patterns, the low yield of fragment ions, and the limited mass accuracy and sensitivity make PSD-based protein identification difficult.[31,32]

It has therefore been suggested to combine MALDI peptide mapping with peptide sequencing by nanoelectrospray tandem mass spectrometry (nanoES MS/MS)[33] in a layered approach.[34] MALDI peptide mapping is implemented as a high-throughput first layer screen, but nanoES MS/MS is then applied if the former method fails to identify the protein. This is an effective procedure, but nanoES MS/MS is labor-intensive and slow, so it represents a serious bottleneck in applications that require high throughput.

It would thus be a significant advantage if low-energy collision tandem mass spectra could be acquired in the same mass spectrometer and with the same sample used for MALDI peptide mapping. The division of sample and the extra purification steps required for electrospray measurements would then be eliminated, and much higher throughput of the analysis can be expected. Another important feature of such a technique would be the ability to interrogate any peptide peak observed in a peptide map by tandem mass spectrometry and thus to establish its identity.

Recently a MALDI ion source has been interfaced to an orthogonal injection time-of-flight mass spectrometer[35] and then to a previously developed hybrid quadrupole time-of-flight mass spectrometer (QqTOF)[36] to form a MALDI QqTOF instrument.[37,38] In these devices, a "collisional damping interface"[39] cools the ions produced in the MALDI ion source before they enter the analytical quadrupole Q. The cooled ions can then be transported efficiently through the quadrupoles for measurement of the whole mass spectrum in the TOF section. Alternatively, a precursor ion can be selected in the quadrupole Q and fragmented in the collision cell q. Measurement of the product ions in the TOF section then provides a MS/MS spectrum of the selected precursor.

Thus, both peptide mass mapping and MS/MS measurements can be carried out on the same target and in the same experiment. Moreover, the TOF measurement is almost completely decoupled from the MALDI ion production process, providing high mass accuracy for both precursor and product ions, high resolution, simple selection of precursor ions, precise tuning of the collision energy, and a much simplified calibration procedure.[39,40]

The design and operational principles of the MALDI QqTOF mass spectrometer will be described elsewhere.[41] In the present paper, we focus on the application of MALDI QqTOF mass spectrometry to proteomic research and outline new analytical perspectives made possible by its unique features. Using well-characterized protein standards, as well as proteins that have been isolated in ongoing collaboration projects in EMBL, we demonstrate that MALDI QqTOF mass spectrometry allows facile identification of proteins at the femtomole level in complex mixtures by searching against comprehensive protein and EST databases. We also demonstrate that very hydrophobic proteins known to withstand treatment by trypsin may now be included in the scope of proteomic research. Such proteins can be rapidly identified at the low-picomole level using limited digestion with HCl followed by analysis of the digests by MALDI QqTOF mass spectrometry.

EXPERIMENTAL SECTION

Materials and Reagents. Unless otherwise noted, all chemicals were purchased from Sigma (Sigma Chemicals, St. Louis, MO) and were of analytical grade except for silver nitrate, which was SigmaUltra grade. MilliQ water (Millipore, Bedford, MA) was used to prepare silver and Coomassie staining solutions. For mass spectrometric analysis and preparation of digests, HPLC grade water, methanol, and acetonitrile (LabScan, Dublin, Ireland) were used.

The preparation of membrane proteins from *Saccharomyces cerevisiae* was obtained from Dr. Kai Simons's group (Cell Biology Program, EMBL). The proteins were purified by flotation as

(25) Jensen, O. N.; Mortensen, P.; Vorm, O.; Mann, M. *Anal. Chem.* **1997**, *69*, 1706−1714.
(26) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850−858.
(27) Arnott, D.; O'Connell, K. L.; King, K. L.; Stults, J. T. *Anal. Biochem.* **1998**, *258*, 1−18.
(28) Boguski, M. S. *Trends Biochem. Sci.* **1995**, *20*, 295−296.
(29) Kuster, B.; Mortensen, P.; Mann, M. *Proc. 47th ASMS Conf. Mass Spectrom. Allied Top.*, Dallas TX, 1999; pp 1897−1898.
(30) Spengler, B.; Kirsch, D.; Kaufmann, R.; Jaeger, E. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 105−108.
(31) Griffin, P. R.; MacCoss, M. J.; Eng, J. K.; Blevins, R. A.; Aaronson, J. S.; Yates, J. R. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1546−1551.
(32) Spengler, B. *J. Mass Spectrom.* **1997**, *32*, 1019−1036.
(33) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *66*, 1−8.
(34) Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Boucherie, H.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14440−14445.
(35) Krutchinsky, A. N.; Loboda, A. V.; Spicer, V. L.; Dworschak, R.; Ens, W.; Standing, K. G. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 508−512.
(36) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1015−1024.
(37) Krtutchinsky, A. N.; Loboda, A. V.; Bromirski, M.; Standing, K. G.; Ens, W. *Proc. 46th ASMS Conf. Mass Spectrom. Allied Top.*, Orlando, FL, 1998; p 794.
(38) Loboda, A.; Krutchinsky, A.; Spicer, V.; Ens, W.; Standing, K. *Proc. 47th ASMS Conf. Mass Spectrom. Allied Top.*, Dallas, TX, 1999; pp 1956−1957.
(39) Krutchinsky, A. N.; Chernushevich, I. V.; Spicer, V. L.; Ens, W.; Standing, K. G. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, *569−579*.
(40) Chernushevich, I.; Ens, W.; Standing, K. G. *Anal. Chem.* **1999**, *71*, 452A−461A.
(41) Loboda, A.; Krtutchinsky, A.; Bromirski, M.; Ens, W.; Standing, K. G., submitted to *Rapid Commun. Mass Spectrom.*

described.[42] The protein mixture was redissolved in Laemmli buffer containing 10% SDS, separated by one-dimensional SDS−PAGE, and visualized by Coomassie staining.

A preparation of human membrane-associated proteins was obtained from Dr. Burkhart Schraven's laboratory (Institute for Immunology, University of Heidelberg). The proteins were separated by two-dimensional gel electrophoresis and visualized by Coomassie staining.

A stock solution of bacteriorhodopsin (BR) from *Halobacterium halobium* (Sigma; lot 117H8623) was prepared in formic acid, and the protein content was estimated by amino acid analysis. For in-gel digestion experiments, a separate stock solution of BR was prepared in Laemmli buffer containing 10% SDS. An aliquot was loaded onto a one-dimensional polyacrylamide gel and visualized by Coomassie staining after the gel electrophoresis.

A stock solution of bovine serum albumin (BSA) was also quantified by amino acid analysis.

Enzymatic In-Gel Digestion. In-gel digestion of proteins with trypsin (unmodified, sequencing grade, Boehringer Mannheim, Germany) was performed essentially as described.[26]

In-gel digestion of proteins with the endoproteinase LysC (from *Lysobacter enzymogenes*, sequencing grade, Boehringer Mannheim) was performed using a similar recipe. The enzyme concentration in the digestion buffer was 25 ng/μL, and no $CaCl_2$ was added.

In-Gel Digestion with HCl. A protein band containing 5 pmol of bacteriorhodopsin was excised from the gel and cut into pieces of approximately 1 mm × 1 mm. The gel pieces were incubated for 1 h in 100 mM ammonium bicarbonate/acetonitrile 1:1 (v/v) to remove Coomassie staining and were then washed with water for another 30 min to remove ammonium bicarbonate. The water was discarded, and the gel pieces were shrunk by 50 μL of acetonitrile. The acetonitrile was discarded, and the gel pieces were dried in a vacuum centrifuge. Approximately 30 μL of 6 M HCl containing 10 mM dithiotreitol was added, and the sample was incubated for 45 min at 56 °C. The sample was chilled to room temperature, and the gel pieces were spun in a benchtop centrifuge. A total of 25 μL of water was then added. After a brief shaking, the supernatant was recovered in a separate Eppendorf test tube and immediately dried in a vacuum centrifuge. The gel pieces were twice extracted with water and acetonitrile as described above; then the recovered extracts were pooled and dried in a vacuum centrifuge.

Preparation of Sample Probes. A matrix solution of 160 mg/mL 2,5-dihydroxybenzoic acid (DHB; Sigma) was prepared in a 1:3 mixture of acetonitrile and water. A total of 0.6 μL of the matrix solution was first deposited on the target to form a spot 2−3 mm in diameter. Samples were dissolved in 2−5 μL of 5% formic acid, and typically, a 0.3−0.6-μL aliquot was deposited on top of the matrix spot.

Instrument. All experiments were performed on a prototype MALDI QqTOF mass spectrometer built at the University of Manitoba in collaboration with MDS Sciex (Concord, ON, Canada). Laser pulses were generated by a nitrogen laser (Laser Science Inc.) model 337 ND with laser energy per pulse of 250 μJ. A 0.2-mm core diameter fused-silica optic fiber delivered the laser pulses to the target. The laser beam was focused at the target into a spot about 0.3 mm × 1 mm. A 10-kV acceleration voltage was used in the TOF part of the instrument. The orthogonal injection pulse repetition rate was set to 8.5 kHz.

Acquisition of MS and MS/MS Spectra. Single MS spectra were acquired at a laser repetition rate of 5−10 Hz for time intervals typically less than 1 min with no adjustment of the laser fluence. Tandem mass spectra were acquired at 10−20 Hz, also without adjustments in the laser fluence. Spectra acquisition time varied depending on the ion current but rarely exceeded 5 min. The width of the mass window for Q1 was set to ~2 Da for m/z 500, increasing to 4 Da at m/z 3000. In all experiments, argon was used as a "cooling" gas in Q0 and as a collision gas. The collision energy was set by applying the accelerating voltage at the entrance of the collision cell by the rule 0.6 V/Da and then adjusted manually to obtain a desirable fragmentation pattern.

The instrument was calibrated externally and no postacquisition recalibration of MS and MS/MS spectra was performed.

Protein Identification Using Peptide Mass Maps. Database searching was performed against a comprehensive nonredundant sequence database using PeptideSearch v 3.0 software developed in EMBL.[43] The database currently comprising more than 410 000 sequence entries was loaded in the RAM of a G3 Macintosh computer (Apple Computer Inc., Cupertino, CA). No limitations on protein molecular weights or species of origin were imposed.

Protein Identification Using Tandem Mass Spectra. The m/z of monoisotopic product ions whose signal intensity exceeded 20% of the intensity of the most abundant peak in the spectrum were used for searching databases via the Internet by the MS-Tag program (http://prospector.ucsf.edu/ucsfhtml3.2/mstagfd.htm) at the UCSF Mass Spectrometry Facility server (http://donatello.ucsf.edu). Mass tolerance was set to 0.05 Da for the precursor ion and for product ions. The allowed number of unmatched product ions was initially set to half of the total number of product ions uploaded in the search program and was later adjusted if necessary. Unless stated otherwise, no limitations on protein molecular weight, p*I*, or species of origin were imposed. For searching against an EST database, a three-frame translation of the human EST database was selected.

Specificity of the protease used for protein digestion (trypsin or LysC) was considered when searching against protein and EST databases (unless stated otherwise). In cases where the protein was cleaved by HCl, the enzyme specificity selection was set to "No Enzyme".

RESULTS AND DISCUSSION

Femtomole Sequencing of Proteins. Although 10 000 mass resolution (fwhm) together with subfemtomole sensitivity has already been reported for the MALDI QqTOF instrument,[37] we set out to investigate whether these unique features also lead to facile identification of gel-separated proteins at the femtomole level. To this end, 40 fmol of an in-gel tryptic digest of BSA was analyzed. A peptide map covering more than 12% of the sequence was acquired (Figure 1A). Masses of seven peptide ions matched the masses of BSA tryptic peptides with an average accuracy of 12 ppm. The second position in the list of hits was occupied by eight proteins, each of which had only four peptides matching the

(42) Scheiffele, P.; Rietveld, A.; Wilk, T.; Simons, K. *J. Biol. Chem.* **1999**, *274*, 2038−2044.

(43) Mann, M. In *Microcharacterization of Proteins*; Kellner, R., Lottspeich, F., Meyer, H. E., Eds.; VCH: Weinheim, 1994; pp 223−245.

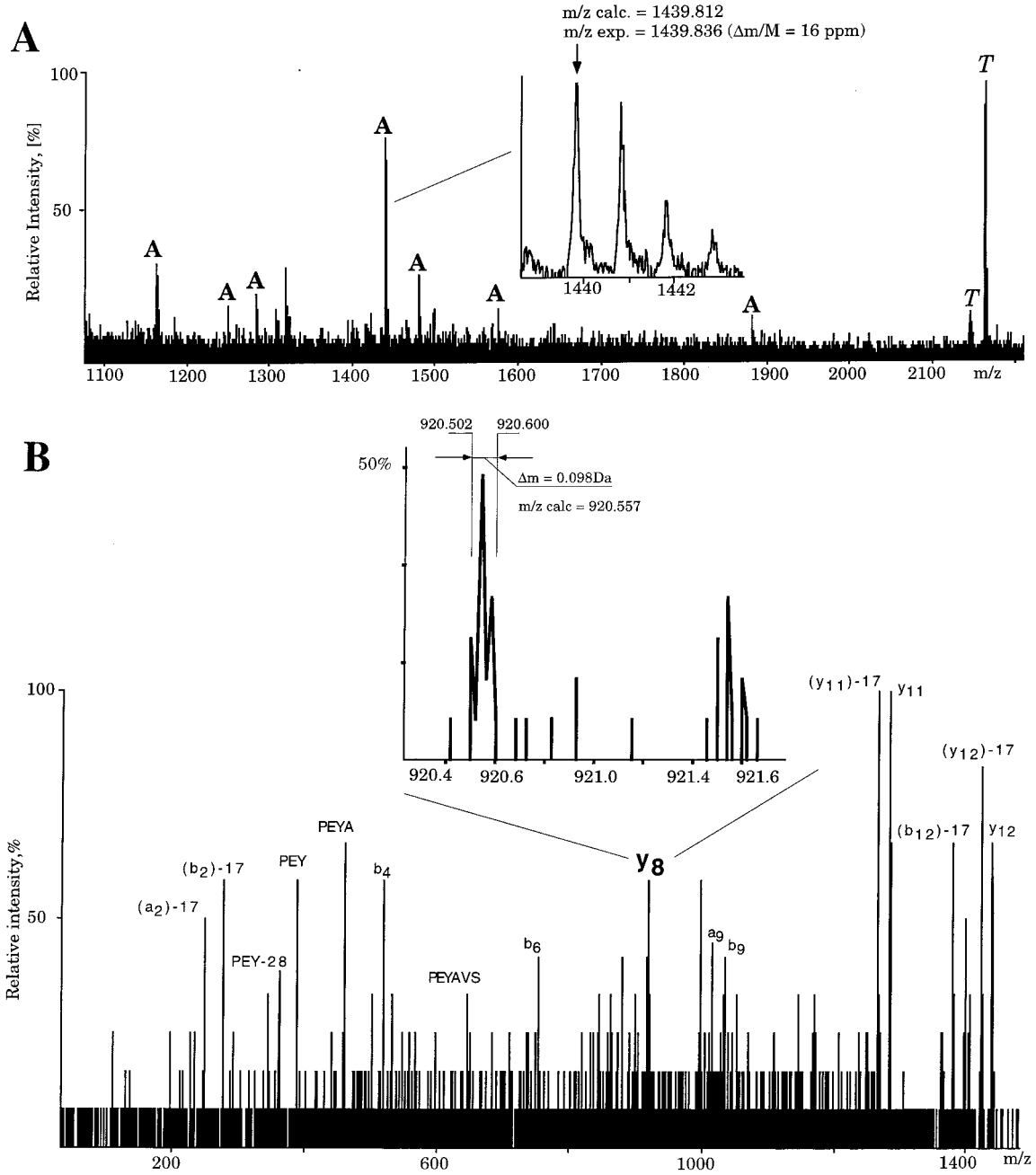

Figure 1. Panel A: spectrum of an in-gel tryptic digest of 40 fmol of BSA. Peaks designated with "A" are BSA tryptic peptides. Autolysis products of trypsin are marked by "T". The inset demonstrates the isotopic cluster of the tryptic peptide RHPEYAVSVLLR. The tandem mass spectrum (panel B) was acquired from this precursor ion. The fragment ions matched to the peptide sequence are designated in panel B using Biemann's nomenclature;[57] the internal fragments are marked by the corresponding sequences. The inset in panel B presents the isotopic cluster of the y_8 fragment ion; 18 ions were detected in its monoisotopic peak.

acquired peptide mass map. Among these proteins, only two were of mammalian origin and the molecular weight of both proteins was higher than 120 000, which is almost 2 times higher than the molecular weight of BSA (~66 000). Thus, despite the relatively low sequence coverage, identification was unambiguous.

In the same experiment, a tandem mass spectrum was acquired from a peptide precursor ion with *m*/*z* 1439.836 (Figure 1B). A total of 22 product ion masses were selected and used for searching a database with the MS-Tag program. The peptide sequence (R)RHPEYAVSVLLR(L), which is present in a number of serum albumins from various species, was retrieved as the top hit. A total of 15 out of the 22 masses of peptide product ions matched the masses of the expected fragments. Although the second listed candidate peptide, (K)DAVISYRHLPLR(M), had precisely the same mass as the BSA peptide, it originated from a microbial protein and only 8 product ions out of the 22 could be matched to its sequence. Thus, even at the low-femtomole level a single tandem mass spectrum acquired on the MALDI QqTOF instrument provided unequivocal identification of the protein and no recourse to peptide mass map data was necessary.

High resolution of the instrument is important for MS/MS identification of proteins at the femtomole level, since ion statistics are usually poor in such MS/MS spectra. Although only a few ions have been detected in a peak (Figure 1B), they were

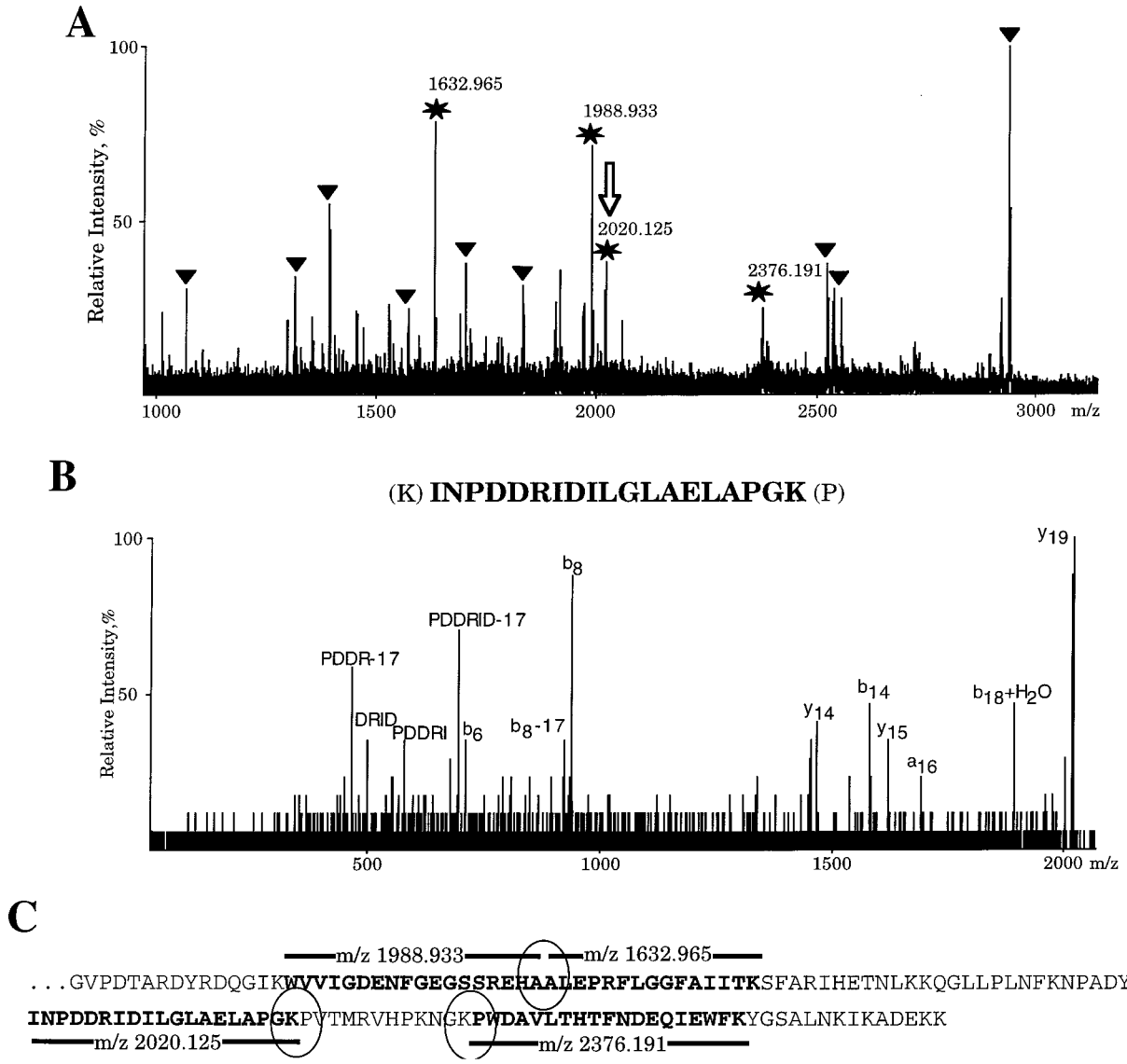

Figure 2. Panel A: spectrum of an in-gel LysC digest of a 90-kDa yeast protein. Masses of peptide ions designated with filled triangles identified the yeast protein Aco1p. Peaks designated with filled asterisks did not match the Aco1p sequence and were subsequently subjected to tandem mass spectrometric sequencing. Panel B: a tandem mass spectrum acquired from the precursor ion with *m/z* 2020.125 (the peak labeled with the arrow in panel A). Database searching using the masses of the product ions hit a peptide from the same protein Aco1p (the peptide sequence above the spectrum). This and the other three peptide ions originated from the C-terminal part of the Aco1p sequence (panel C). The peptide sequences hit are highlighted. Sites of the cleavages unexpected from the specificity of LysC are circled.

"channeled" into a mass window of width less than 0.1 Da. Thus, even for a very weak peak, the approximate mass of its centroid can be used for database searching with a mass tolerance better than ±0.05 Da, which allows very specific database searching. The high resolution of the instrument rescues the information value of spectra with poor ion statistics and represents a significant advantage in protein identification in comparison with PSD.

Comprehensive Interpretation of Peptide Mass Maps. Once a protein candidate has been hit by searching a database with a MALDI peptide mass map, a "second pass search" procedure[44] can be applied to assist in identifying the masses of modified peptides. The second pass search routine considers common protein modifications, such as oxidation of methionine residues and acrylamidation of cysteine residues.

However, a number of peptide ions often remain unmatched to the sequence of the identified protein. Unmatched peptide ions can in principle be explained by posttranslational modifications, N- and/or C-terminal truncation, polymorphism in the protein sequence, unspecific cleavage by the protease, etc. However, no definite information regarding the identity of the unmatched peaks can normally be extracted from the analysis of the peptide mass map alone. This is a major limitation and further investigation of the sample by nanoES MS/MS or LC MS/MS is required to address the problem. By contrast, we demonstrate that MALDI QqTOF mass spectrometry can provide a comprehensive interpretation of a peptide map in the course of a single experiment.

A Coomassie-stained band from a 90 kDa yeast protein was in-gel digested with LysC proteinase, and the corresponding peptide mass map was acquired (Figure 2A). Nine peptide ions

(44) Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. *Anal. Chem.* **1997**, *69*,

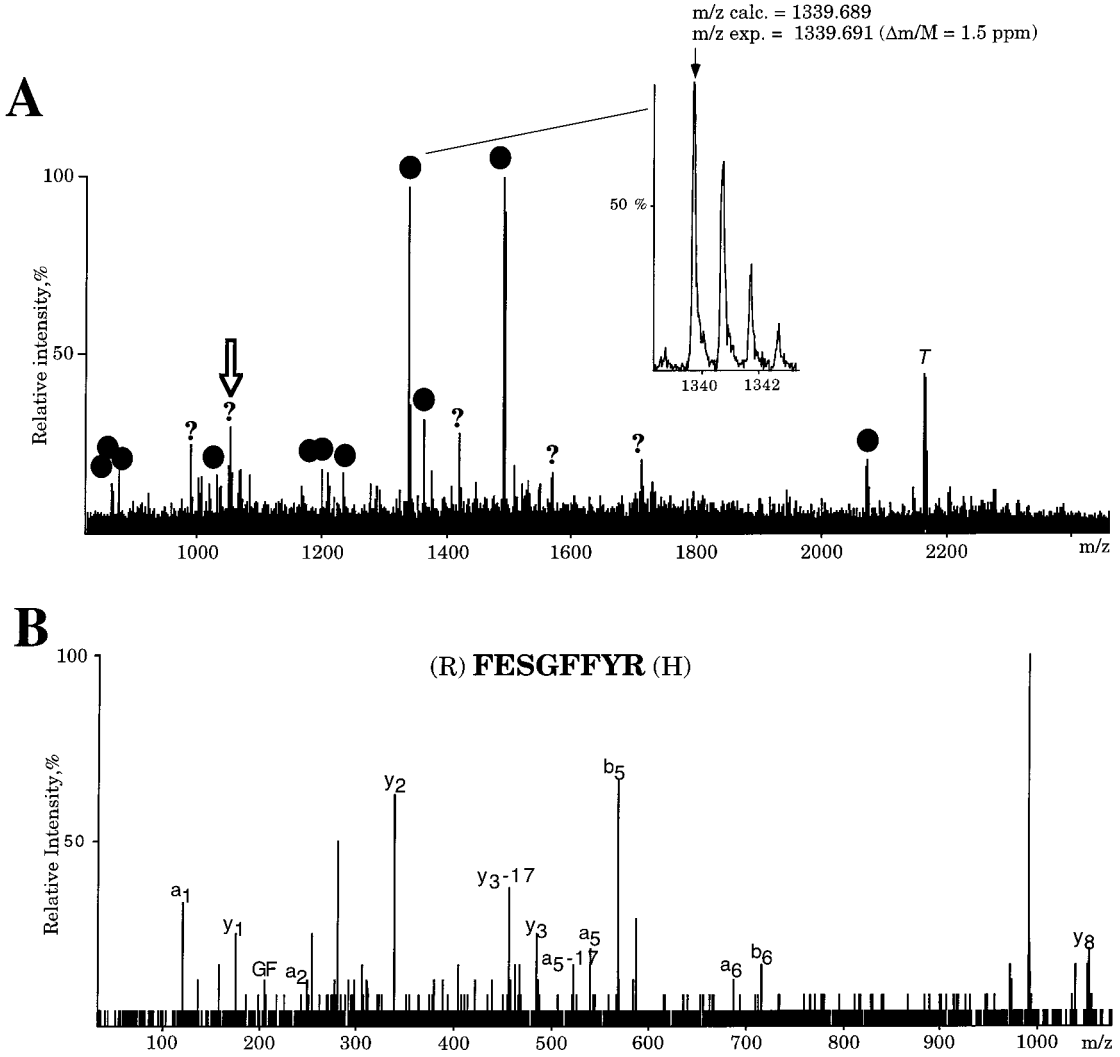

Figure 3. Panel A: spectrum of the in-gel tryptic digest of a 42-kDa yeast protein. Masses of peptide peaks designated with filled circles identified the yeast protein Pda1p. Peptides, which did not match the Pda1p sequence, are designated with a question mark. Database searching with the masses of these ions hit two yeast proteins (Table 1). A tandem mass spectrum (panel B) acquired from the precursor ion *m*/*z* 1052.496 (labeled with an arrow in panel A) identified the protein Ktr1p as a minor component of the mixture.

matched the sequence of yeast aconitate hydratase Aco1p (calculated MW 85 400) with average accuracy of 10 ppm covering 18% of the protein sequence. However, the masses of four prominent peptide ions did not match (Figure 2A). This might indicate either multiple modifications of the protein or another protein present in the same band. However, no reliable hit was produced when the masses matching the aconitase sequence were removed from the searching list and the next iteration of database searching was performed.

Tandem mass spectra were then acquired from these four peptide ions in the same experiment (Figure 2B) and a database was searched without imposing any restrictions on protease specificity. In each case, peptides from Aco1p now appeared at the top of the list of candidates, and in each product ion spectrum, more than half of the fragment ions were matched to the expected fragment pattern.

Close inspection of the peptide sequences provides an explanation for this anomaly, which arises because of assumptions made by the database searching software. Here the software (both PeptideSearch and MS-Fit programs) assumes that LysC protease is unable to cleave the peptide bond ...Lys−Pro... (as is the case for trypsin, another serine protease that cleaves peptide bonds at the C-termini of both Lys and Arg residues). Therefore the search failed to find the peptides (K)INPDDRIDILGLAELAPGK(P) and (K)PWDAVLTHTFNDEQIEWFK(Y) when the enzyme specificity restrictions were applied, since they are followed by or preceded by a proline residue, respectively, in the sequence of Aco1p (Figure 2C).

Two other peptides, (K)WVVIGDENFGEGSSREHA(A) and (A)ALEPRFLGGFAIITK(S) were not found in the original search because they represent N-terminal and C-terminal pieces of a large LysC peptide having *m*/*z* 3602.839 Da (Figure 2C).

Thus, MALDI QqTOF mass spectrometry provided comprehensive characterization of the Aco1p band by establishing the identity of all the prominent peaks in its peptide mass map (Figure 2A).

Deciphering Complex Protein Mixtures. Comigration of proteins within a single band is a common occurrence when complex protein mixtures are separated by one-dimensional polyacrylamide gel electrophoresis. Although MALDI peptide

Table 1. Candidates for a Minor Component of the Protein Mixture Suggested by Second Pass Searching

Search Parameters
enzyme: trypsin
mass tolerance: 25 ppm
mass is monoisotopic
Cys is CYS_ACETAMIDE. Met is NORMAL

P27810 Alpha-1,2 mannosyltransferase
Gene Name: KTR1
MW: 46 400

m/*z* measd	*m*/*z* calcd	Δ *m*/*z*	sequence
1004.460	1004.447	0.013	(R)EGGFFYER(W)
1052.496	**1052.484**	**0.012**	**(R)FESGFFYR(H)**
1567.669	1567.706	−0.037	(R)SPAYSAYFDYLDR(E)
1709.831	**1709.838**	**−0.007**	**(K)EYEATIPTLWETTR(K)**

these 4 peptides cover 43 out of 393 amino acids (10.94%)

Q06148 Chromosome XII Cosmid 8479
Gene Name: YLR265C
MW: 39 500

m/*z* measd	*m*/*z* calcd	Δ *m*/*z*	sequence
1052.496	**1052.489**	**0.007**	**(R)FEESGELNK(K)**
1548.755	1548.747	0.008	(K)GQQLSDAEWCVKK(I)
1709.831	**1709.823**	**0.008**	**(R)LENFSESEATPEKTK(S)**

these 3 peptides cover 37 out of 342 amino acids (10.82%)

mass mapping is capable of deciphering simple mixtures, its performance depends crucially on the number of peptides originating from each individual protein and on the accuracy of mass measurements.[44] Therefore, spotting of individual proteins in mixtures becomes problematic when only a small number of peaks are detected and consequently a statistically reliable hit cannot be produced.[45] If by chance, the detected peptide masses are found in both candidate proteins, it is not possible to tell which protein is actually present in the mixture or whether both proteins are. In the case study presented below, we demonstrate that MALDI QqTOF mass spectrometry facilitates identification of proteins in very complex mixtures that are hard to decipher using conventional MALDI analysis.

In a peptide mass map of the tryptic digest of a 42-kDa yeast protein, masses of 11 peptides matched the masses of peptides from the α subunit of pyruvate dehydrogenase Pda1p (calculated MW 46 300) with an average accuracy 8 ppm (Figure 3A).

In a second pass search procedure, these masses were eliminated from the search list and the search was repeated under less stringent matching requirements. Two yeast proteins having similar molecular weights were hit (Table 1). Two out of three matching peptide masses for the product of the YLR256C gene can also be matched to the masses of peptides from the protein Ktr1p. To find out which protein was present in the sample, or whether both proteins were, a tandem mass spectrum was acquired from the peptide precursor ion having *m*/*z* 1052.496. Note that this mass occurs in the peptide mass maps expected from both proteins (Table 1).

The MS-Tag program counts the number of the product ions whose masses match those of fragment ions calculated from a given peptide sequence. If a tandem mass spectrum is acquired from a mixture of two different peptides with the same molecular mass, we may expect that the program will include both peptides in the list of candidate sequences. Here database searching was performed under less stringent restrictions than those normally applied, to allow the software to include the peptide in the list of hits even if only a few product ions are matched. In fact, the search produced the peptide sequence (R)FESGFFYR(H) from the Ktr1p protein as a unique hit. The alternative peptide (Table 1) did not appear in the list of candidates even when more than 75% of the masses of the product ions were allowed to mismatch. Thus, the minor component of the mixture was uniquely identified as the Ktr1p protein.

Thus, we have demonstrated that MALDI QqTOF mass spectrometry allows facile identification of individual proteins in mixtures. First, the high mass accuracy minimizes overlap between the lists of masses expected from different protein candidates, making second pass search filtering more efficient. Second, for the same reason, fewer peptides are required to reach certain identification of the protein. Finally, the protein can be reliably identified by a tandem mass spectrum acquired from a single peptide precursor ion.

Identification of Proteins in an Expressed Sequence Tag Database. The expressed sequence tag database has become an important resource for identifying novel genes by mass spectrometry.[46,47] However, protein identification via searching an EST database has almost never been performed by MALDI peptide mass mapping alone. By contrast, a single tandem mass spectrum may suffice to pinpoint the relevant EST clone, so it comes as no surprise that the vast majority of novel proteins in EST databases have been identified by electrospray tandem mass spectrometry. Here we demonstrate that MALDI QqTOF mass spectrometry allows identification of novel proteins by searching an EST database, representing a considerable advantage over conventional MALDI peptide mapping.

A peptide map acquired from in-gel tryptic digest of a 25-kDa human protein contained 18 prominent peptide ions (Figure 4A). Database searching (in which the expected molecular weight of the protein was limited to 100 000) hit a chicken protein (gene name CLE7), having a molecular weight (27 400) similar to the one expected for the human protein. Although 7 matched peptides covered 22% of the CLE7 sequence (Figure 4A), more than 10 peptide ions were still not assigned. We therefore hypothesized that we might be dealing with a human homologue of CLE7. However, MS/MS spectra acquired from the most abundant peaks (in particular, from ions with *m*/*z* 2512.163 and 2640.226) did not match the peptide sequences of the corresponding peptides from CLE7 (data not shown)

Tandem mass spectrum acquired from the ion with *m*/*z* 1083.533 (Figure 4B) did not produce a reliable hit when the search was performed against a protein database. The search was then repeated against a three frame translated human EST database using the same set of data, and some 20 EST clones encoding for the identical peptide sequence (R)NIHSSDWPK(F)

(45) Fountoulakis, M.; Juranville, J. F.; Roder, D.; Evers, S.; Berndt, P.; Langen, H. *Electrophoresis* **1998**, *19*, 1819−1827.
(46) Mann, M. *Trends Biochem. Sci.* **1996**, *21*, 494−495.
(47) Pandey, A.; Lewitter, F. *Trends Biochem. Sci.* **1999**, *24*, 276−280.

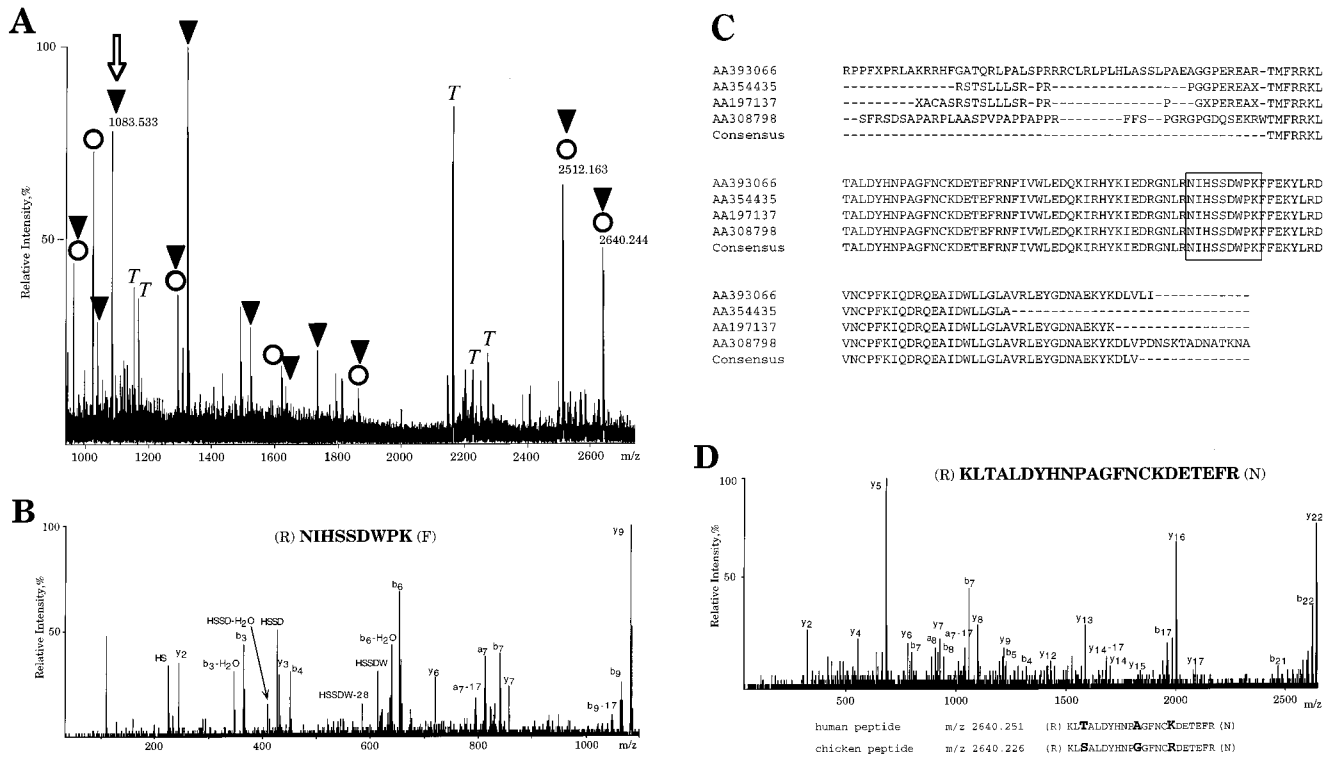

Figure 4. Panel A: spectrum of a tryptic in-gel digest of a 25-kDa human protein. The peptide peaks designated with open circles matched the masses of peptides from the avian protein CLE7. Panel B: tandem mass spectrum acquired from the peptide ion *m*/*z* 1083.533 (designated with an open arrow in panel A). Search of an EST database with the masses of the product ions hit the sequence (R)NIHSSDWPK(F) Panel C: consensus protein sequence composed from aligned sequences of human EST clones containing the peptide hit (boxed). Ten peptide ions (designated with filled triangles in panel A) were retrospectively matched to the consensus sequence. Panel D: tandem mass spectrum of the precursor ion with *m*/*z* 2640.244. Sequences of corresponding peptides from human and avian proteins are pasted below the spectrum (see text for details).

were hit. Masses of 15 out of 20 product ions were matched to the masses of the fragment ions calculated from this sequence.

To further verify the hit, we selected four EST clones that encoded for the longest protein sequence (Figure 4C). Protein sequences were obtained by translation of the EST sequences in the same frame in which the peptide NIHSSDWPK had been hit, and these were aligned using Clustal X software. Although the protein sequences deduced from selected EST clones appeared to vary at the N- and C-termini, all four clones contained an identical 12.7-kDa stretch of consensus noninterrupted sequence. When the peptide mass map was screened against the consensus sequence, 10 peptides were matched with accuracy on average better than 16 ppm (Figure 4A). The ions *m*/*z* 2512.163 and 2640.244 were now matched to the consensus sequence, and the suggested sequences of the human peptides were fully consistent with the tandem mass spectra (Figure 4D). Three amino acid residues appeared to be different in the peptide from human protein compared with the corresponding peptide from the avian protein CLE7, although both peptides have almost identical masses.

When the present paper was in preparation, a full length sequence of the human protein was submitted to a sequence database (acc no. AAD34094, June 1999). All major peaks observed in the peptide map could be assigned to the full length sequence of the protein (Figure 4A).

The interactive acquisition−database searching approach offered by MALDI QqTOF mass spectrometry significantly improves certainty of protein identification in an EST database. If searching against an EST database with MS/MS data retrieves a number of candidate sequences having a similar score, a peptide map of the digest can be screened against a three frame translation of the sequences of EST inserts (or against a consensus sequence of combined EST clones). Given a good selection of MALDI peptide maps (more than 10 peptide ions are usually observed from 1 pmol amount of the in-gel digested protein), it is reasonable to anticipate that more peptides can be matched to the same EST sequence. The match can then be confirmed by a tandem mass spectrum acquired from a candidate precursor ion. The high mass accuracy peptide map previously acquired in the same experiment allows us to concentrate our efforts on promising candidate precursor ions rather than to acquire a larger number of spectra from a random selection of precursors.

It is important to note that identification in protein and EST sequence databases can now be performed using a single technique in the course of a single experiment. Considering the continuing progress in EST sequencing efforts and in the development of software, we may anticipate that protein identification in an EST database can be performed with a sensitivity and throughput similar to that in a protein sequence database.

An Alternative Approach for Protein In-Gel Cleavage. In-gel[26] or in situ (on the surface of PVDF or nitrocellulose membranes)[48] digestion with trypsin is probably the most widely used protein cleavage method in proteomic research.

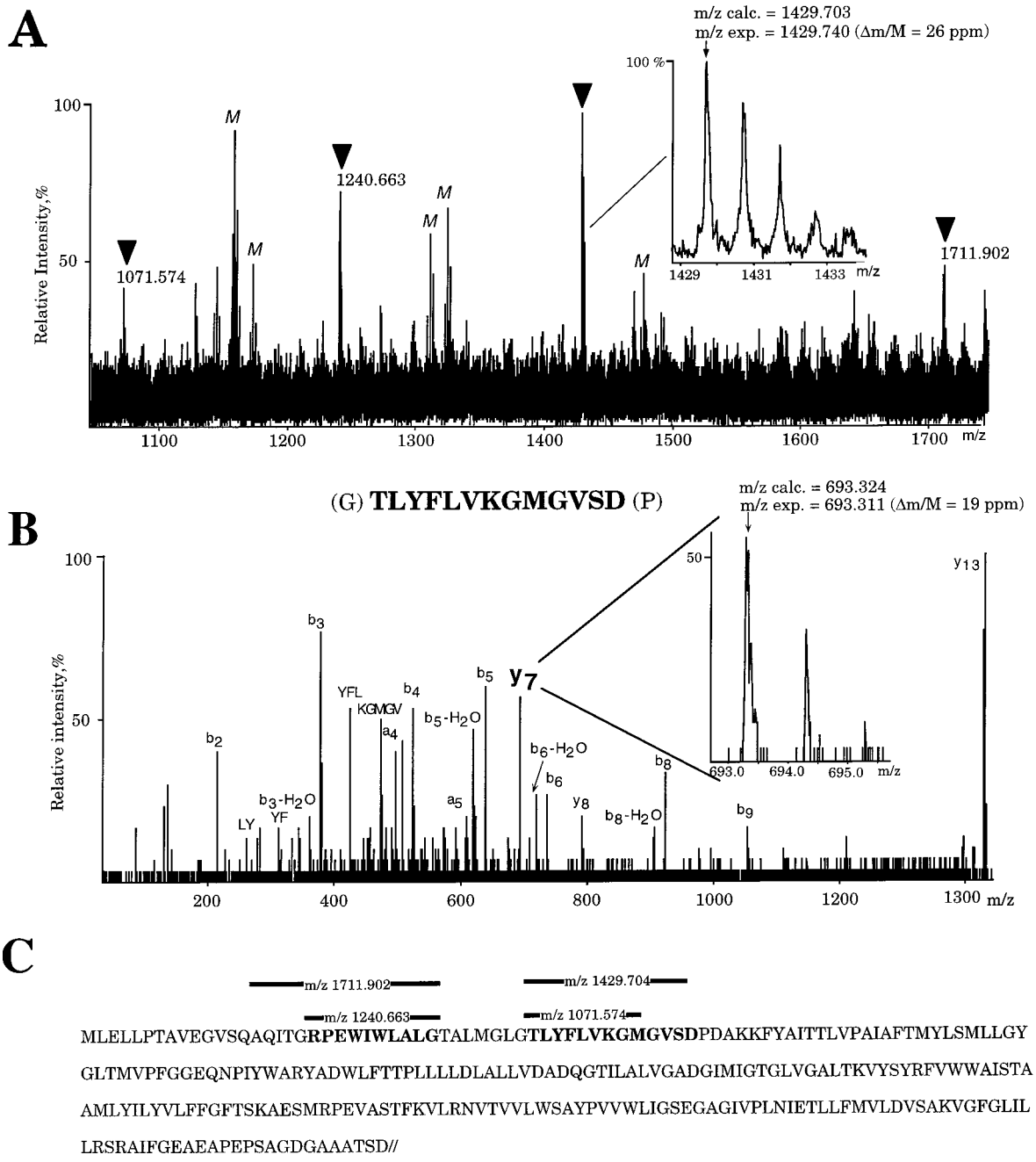

Figure 5. Panel A: spectrum of the limited in-gel cleavage of 5 pmol of BR with 6 M HCl. Tandem mass spectra were acquired from each precursor ion designated with filled triangles in panel A and independently identified BR upon searching a database. Peaks designated with "M" originate from the matrix. Panel B: tandem mass spectrum of the ion with *m/z* 1429.703. The identified peptides are highlighted in the BR sequence (panel C).

However, tryptic digestion usually requires hours to complete. It is also not efficient for very hydrophobic proteins and very basic proteins, since too long and too short peptides are generated, respectively. Recently Gobom et al.[49] suggested performing vapor-phase cleavage of gel-separated proteins by hexafluorobutyric acid (HFBA), an approach pioneered by Tsugita et al.[50] However, ~50 pmol of gel-separated proteins were required for identification, and protein modifications, such as desamidation of glutamine and asparagine amino acid residues, may impair the accuracy of the peptide sequence determined.

Here we demonstrate that any chemical cleavage method that is able to generate a mixture of 1−3-kDa peptides may work equally well for protein identification, when applied in combination with MALDI QqTOF mass spectrometry.

As a test protein we used bacteriorhodopsin of molecular weight 28 000, whose sequence comprises seven transmembrane domains. BR is resistant to conventional digestion with trypsin, so only one peptide was observed in its in-gel tryptic digest performed overnight (data not shown). It would therefore be

(48) Aebersold, R. H.; Leavitt, J.; Saavedra, R. A.; Hood, L. E.; Kent, S. B. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6970−6974.
(49) Gobom, J.; Mirgorodskaya, E.; Nordhoff, E.; Højrup, P.; Roepstorff, P. *Anal. Chem* **1999**, *71*, 919−927.
(50) Tsugita, A.; Takamoto, K.; Kamo, M.; Iwadate, H. *Eur. J. Biochem*. **1992**, *206*, 691−696.

difficult if not impossible to identify the protein by conventional MALDI peptide mapping.

A spectrum of limited in-gel digest of 5 pmol of BR with 6 M HCl is presented in Figure 5A. Note that it was not necessary to digest proteins in the vapor phase. Vapor-phase cleavage is only required to generate ladders of C-terminally truncated peptides and it severely limits the yield of digestion products.[49,50] Tandem mass spectra were acquired from four peptides (Figure 5B), and each spectrum independently and unambiguously pinpointed the corresponding BR peptides upon searching against a protein database (Figure 5C).

It is known, however, that protein hydrolysis with strong mineral acids may result in modifications or destruction of certain amino acid residues. Our preliminary experiments performed on model proteins suggested that tryptophan residues survive mild acidic treatment, but desamidation of asparagine and glutamine amino acid residues occurred to a variable extent.[51] Complete desamidation could be taken into account by modifying the database searching software. Special settings, accounting for conversion of Asn to Asp and of Gln to Glu, might result in a much improved score.

Thus, we have demonstrated that mild digestion of a protein with HCl can generate peptides suitable for protein identification in less than 1 h. Furthermore, acid is a "universal digester", and peptides can be generated from proteins extremely resistant toward cleavage by proteolytic enzymes. Proteins could be identified by this method using the tandem mass spectrum acquired from a single precursor ion and identification of gel-separated proteins is possible at the low-picomole level.

Interactive Approach to Data Acquisition and Database Searching. MALDI QqTOF mass spectrometry changes the entire approach to characterization of complex protein digests. When LC MS/MS or nanoES MS/MS analysis is performed, the sample is continuously consumed during the analysis. Thus, it is usually beneficial to acquire as much MS/MS data as possible before the sample is exhausted and to proceed with interpretation of spectra and database searching later. Although recently the sensitivity and speed of nano-ES MS/MS-based[36,52] and LC MS/MS-based[53,54] protein identification have been increased by employing hybrid quadrupole TOF instruments, the approach suffers from several inherent problems. If the protein mixture contains a major component, peptide peaks originating from this protein would be sequenced preferentially. Also real-time interpretation of tandem mass spectra is difficult, and therefore, it is hard to tell whether the fragmented peptide originates from an already identified protein, in which case the acquisition of its tandem mass spectrum would likely be redundant.

With MALDI QqTOF mass spectrometry, the acquisition of both MS and MS/MS spectra can be halted at any time without affecting the results of the analysis. Therefore a detailed and accurate peptide map can be acquired first and immediately used for searching a database. Once the protein has been identified, repetitive iteration of second pass search filtering and database searching helps to identify additional components (if present) in the mixture. The throughput of the analysis is not much affected, since searching a database with the peptide mass map takes only a few seconds when the database is already loaded into the computer RAM. Intelligent interrogation of a peptide map navigates the subsequent analysis by tandem mass spectrometry and allows us to concentrate our efforts on those peaks whose identity is not established. In turn, this enables efficient consumption of the sample, increases the sensitivity, and minimizes the redundancy of the acquired data. The technique therefore promises higher fidelity in the analysis of complex protein mixtures, in protein quantification,[55] and in identification of posttranslational modifications.

High throughput and high specificity of the protein characterization previously achieved by the combination of peptide mass mapping and tandem mass spectrometry in a layered approach[34,56] appears to be an inherently "built-in" principle of MALDI QqTOF mass spectrometry. The demonstrated versatility of the technique may enable high-throughput identification of a broad scope of proteins in all types of sequence databases, and thus may eliminate a major obstacle in the exploration of proteomes of higher eukaryotic organisms, including mammals.

ACKNOWLEDGMENT

The work at Manitoba was supported by grants from the Natural Sciences and Engineering Research Council of Canada and from MDS Sciex. We are grateful to Victor Spicer and to James McNabb for excellent technical support. The authors thank Dr. M. Wilm from EMBL and Dr. W. Zachariae from MPI for Molecular Cell Biology and Genetics for critical reading of the manuscript and useful discussions, and M. Bagnat and Dr. K. Simons (EMBL) for excellent collaboration.

Received for review November 29, 1999. Accepted February 7, 2000.

AC9913659

(51) Shevchenko, A.; Shevchenko, A.; Loboda, A, unpublished.

(52) Wilm, M.; Neubauer, G.; Taylor, L.; Shevchenko, A.; Bachi, A. In *Proteome and Protein Analysis*; Kamp, R. M., Kyriakidis, D., Choli-Papadopoulou, T., Eds.; Springer: Berlin, Heidelberg, New York, 1999; pp 65−79.

(53) Figeys, D.; Lock, C.; Taylor, L.; Aebersold, R. *Rapid Commun Mass Spectrom* **1998**, *12*, 1435−44.

(54) Moseley, M. A. In *Mass Spectrometry in Biology and Medicine.*; Burlingame, A. L., Carr, S. A., Baldwin, M. A., Eds.; Humana Press: Totowa, NJ, 1999, pp 179−198.

(55) Oda, Y.; Huang, K.; Cross, F.; Cowburn, D.; Chait, B. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6591−6596.

(56) Shevchenko, A.; Wilm, M.; Vorm, O.; Jensen, O. N.; Podtelejnikov, A. V.; Neubauer, G.; Mortensen, P.; Mann, M. *Biochem. Soc. Trans.* **1996**, *24*, 893−896.

(57) Biemann, K. *Biomed. Environ. Mass Spectrom.* **1988**, *16*, 99−111.