Anal. Chem. **2001,** *73,* 2558−2564

Identification of Individual Proteins in Complex Protein Mixtures by High-Resolution, High-Mass-Accuracy MALDI TOF-Mass Spectrometry Analysis of In-Solution Thermal Denaturation/Enzymatic Digestion

Zee-Yong Park and David H. Russell*

The Laboratory for Biological Mass Spectrometry, Department of Chemistry, Texas A&M University, P.O. Box 30012, College Station, Texas 77842-3012

Identification of individual proteins in complex protein mixtures by high-resolution (HR), high-mass-accuracy matrix-assisted laser desorption/ionization (MALDI) time-of-flight mass spectrometry (TOF-MS) is demonstrated for synthetic protein mixtures. Instead of chemical denaturation, thermal denaturation followed by in-solution trypsin digestion is used to achieve uniform digestion of the constituents of the protein mixture. Protein identification is carried out using protein database searches with search scoring systems, which seems more effective than conventional peptide mass mapping without using a scoring system. Identification of individual proteins by MALDI HR-TOF-MS peptide mass mapping dramatically reduces data acquisition/analysis time and does not require special equipment for sample preparation/transfer prior to mass spectral analysis.

The analytical challenge for proteomics is to expand the scope of biological investigations from the analysis of single proteins to simultaneous analysis of all proteins present in the sample.[1−4] Currently, 2D-PAGE/in-gel digestion combined with matrix-assisted laser desorption/ionization (MALDI) peptide mass mapping is widely used for such studies because the mass spectra are relatively simple, the sensitivity (low femtomole) and mass measurement accuracy (<30 ppm) are good, and MALDI tolerates low levels of buffers and some denaturants.[5−11] Tandem mass spectrometric techniques, which typically take more time and effort than MALDI peptide mass mapping, can be also used to increase the confidence level for dubious gel spots.[7,12−15] Although 2D-PAGE/in-gel digestion with mass spectral analysis has demonstrated an outstanding ability in proteomics study, sample preparation for each gel spot is time-consuming and requires extensive sample handling. The entire analysis, from cell lysis to protein database (PDB) searching, requires several days even with simultaneous sample preparation for all the gel spots. There have been several attempts to automate 2D-PAGE/in-gel digestion analysis;[16−18] however, handling of 2D gels and the series of chemical and mechanical treatments required by this procedure represent significant technical obstacles.[2] Hence, several laboratories have initiated programs to develop alternatives to 2D-PAGE/in-gel digestion.[19−23]

Several in-solution digestion methods have been proposed as alternatives to 2D-PAGE/in-gel digestion.[20−23] One motivation for in-solution digestion methods is the compatibility of the process

* Corresponding author: (phone) (979)-845-3345; (fax) (979)-845-9485; (e-mail) Russell@mail.chem.tamu.edu.

(1) Haynes, P. A.; Gygi, S. P.; Figeys, D.; Aebersold, R. *Electrophoresis* **1998**, *19*, 1862−1871.
(2) Quadroni, M.; James, P. *Electrophoresis* **1999**, *20*, 664−677.
(3) Yates, J. R., III. *Trends Genet.* **2000**, *16*, 5−8.
(4) Pandey, A.; Mann, M. *Nature* **2000**, *405*, 837−846.
(5) Rosenfeld, J.; Capdevielle, J.; Guillemot, J. C.; Ferrara, P. *Anal. Biochem.* **1992**, *203*, 173−179.
(6) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850−858.
(7) Shevchenko, A.; Jensen, O. N.; Podtelejnilov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14440−14445.
(8) Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. *Anal. Chem.* **1997**, *69*, 4741−4750.
(9) Jungblut, P.; Thiede, B. *Mass Spectrom. Rev.* **1997**, *16*, 145−162.
(10) Matsui, N. M.; Smith, D. M.; Clauser, K. R.; Fichmann, J.; Andrews, L. E.; Sullivan, C. M.; Burlingame, A. L.; Epstein, L. B. *Electrophoresis* **1997**, *18*, 409−417.
(11) Pu, L.; Foxworth, W. B.; Kier, A. B.; Annan, R. S.; Carr, S. A.; Edmonson, R.; Russell, D. H.; Wood, W. G.; Schroeder, F. *Protein Expression Purif.* **1998**, *13*, 337−348.
(12) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.
(13) Mann, M.; Wilm, M. *Anal. Chem.* **1994**, *66*, 4390−4399.
(14) Yates, J. R., III; Eng, J. K.; McCormack, A. L. *Anal. Chem.* **1995**, *67*, 3202−3210.
(15) Qin, J.; Fenyo, D.; Zhao, Y.; Hall, W. W.; Chao, D. M.; Wilson, C. J.; Young, R. A.; Chait, B. T. *Anal. Chem.* **1997**, *69*, 3995−4001.
(16) Houthaeve, T.; Gausepohl, H.; Ashman, K.; Nilson, T.; Mann, M. *J. Protein Chem.* **1997**, *16*, 343−348.
(17) 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.
(18) Lahm, H.-W.; Langen, H. *Electrophoresis* **2000**, *21*, 2105−2114.
(19) Opiteck, G. J.; Lewis, K. C.; Jorgenson, J. W. *Anal. Chem.* **1997**, *69*, 1518−1524.
(20) Link, A.; Eng, J.; Schieltz, D.; Carmack, E.; Mize, G.; Morris, D.; Garvik, B.; Yates, J. R.. III. *Nat. Biotechnol.* **1999**, *17*, 676−682.
(21) Wall, D. B.; Lubman, D. M.; Flynn, S. J. *Anal. Chem.* **1999**, *71*, 3894−3900.
(22) Wall, D. B.; Kachman, M. T.; Gong, S.; Hinderer, R.; Parus, S.; Misek, D. E.; Hanash, S. M.; Lubman, D. M. *Anal. Chem.* **2000**, *72*, 1099−1111.
(23) Ekstrom, S.; Onnerfjord, P.; Nilsson, J.; Bengtsson, M.; Laurell, T.; Marko-Varga, G. *Anal. Chem.* **2000**, *72*, 286−293.

10.1021/ac001488p CCC: $20.00
Published on Web 04/21/2001

samples with liquid chromatography. In addition, automation of whole protein identification procedures for analyzing complex protein mixtures is more compatible with the in-solution digestion approach.[24] A disadvantage of in-solution methods for analyzing protein mixtures is the range of digestion for different proteins.[23,25,26] That is, some proteins digest efficiently in the absence of denaturing reagents, whereas some proteins digest slowly in the absence of denaturants. Generally speaking, chemical denaturants are not compatible with MALDI MS; however, we have shown that thermal denaturation can be used and is very efficient.[26,27]

An alternative method that appears to be very effective for fully automated protein digestion involves the use of microchip-immobilized enzyme reactors, but this method requires near-complete protein separation and special equipment for sample delivery and digestion.[23] Although preliminary results using this method are quite impressive, the time required for data acquisition/analysis is comparable to 2D-PAGE/in-gel digestion.

We previously reported a thermal denaturation technique which facilitates efficient in-solution digestion regardless of protease sensitivity of the substrate.[26] Here thermal denaturation is used to achieve uniform digestion of all the proteins present in a synthetic protein mixture. The denatured proteins are then enzymatically digested, and the proteolytic peptides are analyzed by high-resolution, high-mass-accuracy MALDI peptide mass mapping without chromatographic separation. Protein mixture identification using high-resolution, high-mass-accuracy MALDI peptide mass mapping can significantly reduce data acquisition/analysis time because constituents are identified from a single mass spectrum.[27,28] Note also that mixture analysis does not require separation of the proteolytic peptides. Previously Mann and co-workers reported on the analysis of a simple protein mixture using in-gel digestion and high-mass-accuracy MALDI peptide mass mapping.[8] Here we demonstrate enhanced mixture analysis capabilities of high-resolution, high-mass-accuracy MALDI peptide mass mapping with database searching for thermally denatured in-solution digestion samples. In-solution digestion results of chemically denatured proteins are also included for comparison purposes. The goal of this research is to develop a rapid method that requires minimal isolation/separation steps for identification of individual proteins present in protein mixtures.

EXPERIMENTAL SECTION

All the proteins used in this study were purchased from Sigma (St. Louis, MO), and all protein digestion was performed using sequencing grade-modified trypsin purchased from Promega (Madison, WI). A synthetic protein mixture system was constructed using 11 different proteins. For example, rabbit phosphorylase (MW, 98 000), bovine serum albumin (MW, 66 000), chicken egg ovalbumin (MW, 44 000), rabbit aldolase (MW, 39 000), bovine carbonic anhydrase (MW, 29 000), horse myoglobin (MW, 17 000), bovine hemoglobin α, β (MW, 15 000), horse cytochrome *c* (MW, 12 000), chicken egg lysozyme (MW, 14 000), and bovine ubiquitin (MW, 8 000) were dissolved in 50 mM ammonium bicarbonate (Sigma) solution. The concentration of each protein was 1−5 μM.

The thermal denaturation and in-solution digestion methods used in this study have been described previously.[26] The synthetic protein mixture was thermally denatured by incubation at 90 °C for 20 min. The sample was then cooled to 37 °C before trypsin was added to initiate the digestion process. The chemical denaturation of the protein mixture was performed using the method described by Ekstrom et al.[23] Higher-order protein structure was destablized with 1 M guanidine hydrochloride (GuHCl), and disulfide bonds were reduced with 5 mM dithiothreitol (DTT) at 70 °C for 20 min. Subsequently the reduced cysteine residues were alkylated with 25 mM iodoacetamide (IAA) for 30 min at room temperature. Higher concentrations of denaturant, reducing, and/or alkylating reagents than the above conditions significantly deteriorate MALDI signals. Thermally and chemically denatured samples were subjected to trypsin digestion using an enzyme ratio of 1:40 (weight of trypsin:weight of total proteins) and a digestion time of 5 h at 37 °C.

The protein digests were prepared for MALDI using the overlayer sample preparation method.[29−32] This method involves deposition of a thin layer of matrix on to the sample surface using a saturated methanol solution of matrix (α-cyano-4-hydroxycinnamic acid, Sigma). The solution containing analyte is deposited on top of a dried base layer of matrix. The analyte solution, which is a mixture (3:1 = water:methanol) of digested protein samples (1−2 μM in 5 mM ammonium bicarbonate buffer) and the matrix solution (35 mM in methanol), is applied and air-dried. The quality of MALDI time-of-flight (TOF) mass spectra is very dependent on sample preparation technique, especially the matrix/analyte ratio, pH, and sample washing procedure that are used. We generally find that a matrix/analyte (intact proteins) ratio of ∼4000:1 works well; however, the optimum matrix/analyte (M/A) ratio can vary from sample to sample depending upon buffer concentrations and choice of matrix. Generally speaking, fewer digest fragments are observed at low M/A ratios and at relatively high protein concentrations.

MALDI TOF mass analysis was performed by using a Perseptive Biosystems Voyager Elite XL TOF equipped with delayed ion extraction and a pulse nitrogen laser (337 nm). The instrument and general operating procedures have been described previously.[30] All MALDI mass spectra were taken in the reflected mode using delayed extraction tuning procedures described previously.[33] For example, using ion acceleration voltage of 25 kV and a grid voltage of 17.5 kV, the optimum delay time between laser pulse and pulsing in the extraction plate voltage is ∼200 ns. Mass resolution of 10 000−15 000 is routinely achieved in the mass range of digest fragments (1000−5000 u).[30] Signals from 100 laser shots were averaged to increase S/N ratio of each mass spectrum.

(24) Washburn, M. P.; Yates, J. R., III. *Curr. Opin. Microbiol.* **2000**, *3*, 292−297.
(25) Holland, R. D.; Duffy, C. R.; Rafii, F.; Sutherland, J. B.; Heinze, T. M.; Holder, C. L.; Voorhees, K. J.; Lay, J. O., Jr. *Anal. Chem.* **1999**, *71*, 3226−3230.
(26) Park, Z.-Y.; Russell, D. H. *Anal. Chem.* **2000**, *72*, 2667−2670.
(27) Bruce, J. E.; Anderson, G. A.; Wen, J.; Harkewicz, R.; Smith, R. D. *Anal. Chem.* **1999**, *71*, 2595−2599.
(28) Conrads, T. P.; Anderson, G. A.; Veenstra, T. D.; Pasa-Tolic, L.; Smith, R. D. *Anal. Chem.* **2000**, *72*, 3349−3354.
(29) Vorm, O.; Roepstorff, P.; Mann, M. *Anal. Chem.* **1994**, *66*, 3281−3287.
(30) Edmondson, R. D.; Russell, D. H. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 995.
(31) Kussman, M.; Nordhoff, E.; Rahbek Nielsen, H.; Haebel, S.; Rossel Larsen, M.; Jakobsen, L.; Gobom, J.; Mirgorodskaya, E.; Kroll Kristensen, A.; Palm, L.; Roepstorff, P. *J. Mass Spectrom.* **1997**, *32*, 593−601.
(32) Dai, Y.; Whittal, R. M.; Li, L. *Anal. Chem.* **1999**, *71*, 1087−1091.
(33) Barbacci, D. C.; Edmondson, R. D.; Russell, D. H. *Int. J. Mass Spectrom.* **1997**, *165*, 221.

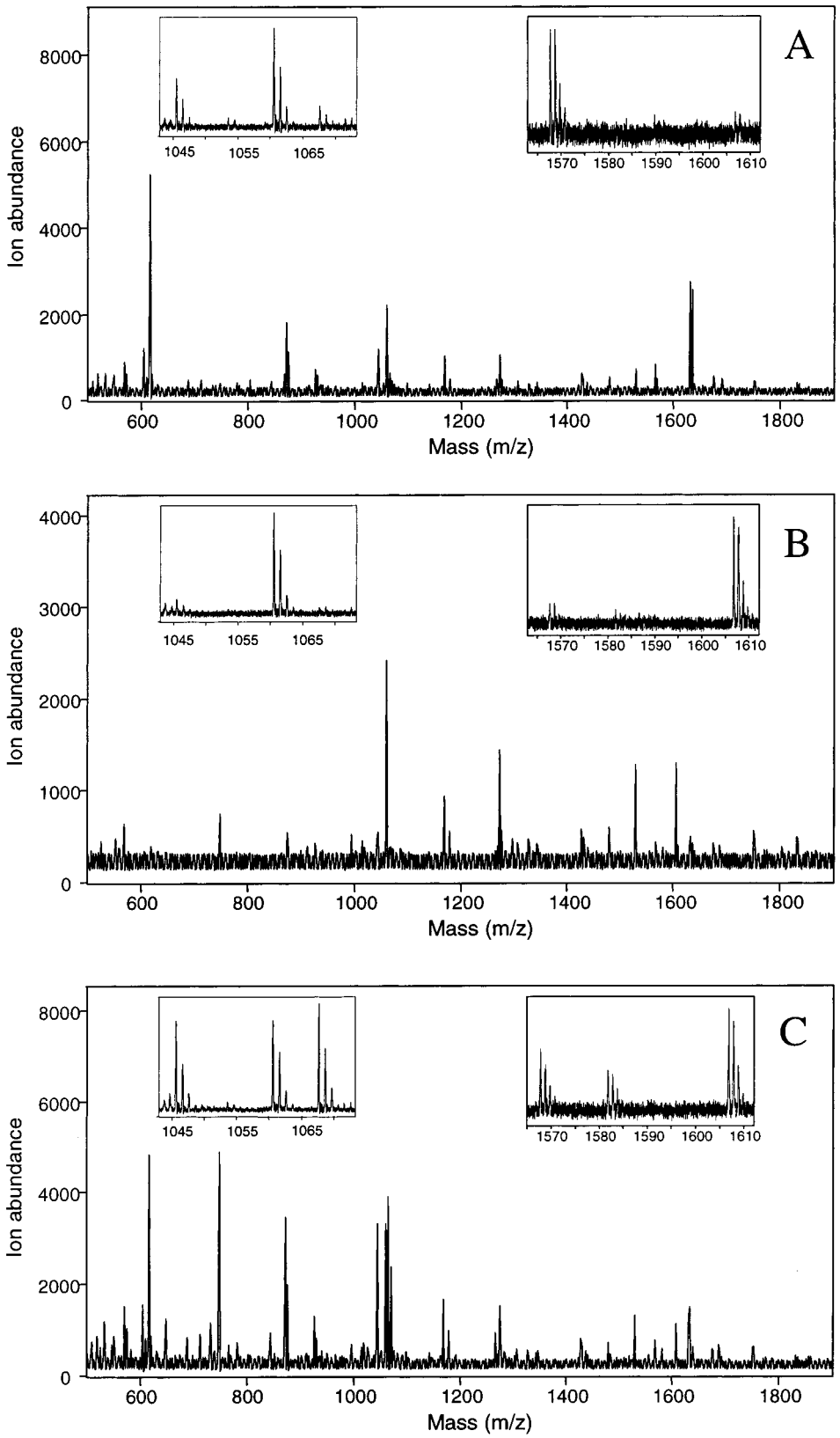

Figure 1. High-resolution MALDI TOF mass spectra of (A) nondenatured, (B) chemically denatured, and (C) thermally denatured tryptic digest of a protein mixture. Inset shows well-resolved monoisotopic distribution of tryptic digest fragments.

All mass spectra were internally calibrated using the matrix (α-cyano-4-hydroxycinnamic acid) dimer signal ($m/z = 379.0930$) and bradykinin $[M + H]^+$ ion signals ($m/z = 1060.5692$). Internal calibration with matrix dimer and bradykinin $[M + H]^+$ ions provides reliable mass accuracy for protein database searches (averages of less than 5 ppm) in broad mass over charge range (500−3500 u). Mass resolution, mass accuracy, and reproducibility of this methodology were described previously.[30]

MS-FIT (University of California, San Francisco, CA) was used to identify digest fragments and to estimate digestion efficiency. The digest fragments were searched against the SwissProt. protein database, and no restrictions were placed on species, molecular weight range, or isoelectric point of the target proteins. Carbamidomethylation of the cysteine residue was considered as a possible modification for chemically denatured protein samples. The mass tolerance or mass error range was kept at 20 ppm for all searches, and the MOWSE scoring scheme was used to give more weight to correct protein matches.[34] The MOWSE scoring system compares the size of the matching digest fragments and the molecular weight of intact proteins so that smaller proteins with larger matching digest fragments give a higher MOWSE score. The so-called "second pass" search scheme developed by Mann and co-workers was combined with MOWSE to identify as many components in the protein mixture samples as possible.[8] Briefly, all mass-assigned ion signals in the mass spectrum are used to search the database at restricted mass tolerance and modification possibilities on the first attempt. The masses that correspond to digest fragments of most highly ranked proteins are eliminated from the list of masses, and the next search is performed with the masses that were not identified in the first search.

RESULTS

Figure 1 contains the high-resolution MALDI TOF mass spectra for nondenatured, chemically (1 M guanidine hydrochloride, 5 mM dithiothreitol, 25 mM iodoacetamide) and thermally denatured and digested protein mixture samples. Note that the total number of ion signals in the mass spectra for the thermally denatured samples is higher than that for chemically denatured or nondenatured samples. For example, the thermally denatured sample (76 digest fragments) contains nearly 2 times as many digest fragments as the nondenatured (39 digest fragments) and chemically denatured (26 digest fragments) samples. In addition, the trypsin incubation time required to achieve an acceptable digestion yield is greatly reduced (overnight or days to a few hours) for the thermally denatured sample. Although chemically denatured samples demonstrate a trypsin digestion pattern different from that of the nondenatured samples, there is no noticeable enhancement in the number of digest fragment ions in the mass spectra. Furthermore, less efficient MALDI ionization due to the presence of denaturant and reducing/alkylating reagents may adversely affect the quality of the mass spectral data (see Figure 1B).

In earlier papers, we demonstrated the analytical utility of high-resolution MALDI TOF mass spectrometry and accurate mass measurement for peptide mass mapping.[30,35,36] For example, the peak centroid can be determined more accurately with the narrow peak profiles obtained at high-mass resolution.[30] Another advantage of high-resolution MALDI TOF is that the isotope cluster is well-resolved and all mass calibration is based on monoisotopic mass (all ^{12}C isotope), and the isotope cluster for each ion can be used to screen the data (See insets of Figure 1.). For example, ions that have a correct isotope cluster can be submitted to peak-centroiding, whereas ion signals with an abnormal isotope cluster can be discarded. This is a very effective way to identify ion signals that arise from overlapping similar mass peptide ions or ion signals where poor S/N ratios (<4) limit the mass measurement accuracy.

Table 1. Digest Fragments Identified from a Nondenatured Synthetic Protein Mixture Sample (11 Proteins) by High-Resolution, High-Mass-Accuracy MALDI Peptide Mass Mapping

search score rank	mass submitted	mass error (ppm)	peptide sequence
1	Bovine, Serum Albumin Precursor (8/39 Matches, 11% Seq Coverage)		
	508.2577	11.3	^{229}FGER^{232}
	689.3688	−6.8	^{236}AWSVAR^{241}
	712.3732	−1.4	^{29}SEIAHR^{34}
	927.4950	1.1	^{161}YLYEIAR^{167}
	1305.7232	5.0	^{402}HLVDEPQNLIK^{412}
	1439.8163	2.8	^{360}RHPEYAVSVLLR^{371}
	1479.7977	1.2	^{421}LGEYGFQNALIVR^{433}
	1567.7445	0.8	^{347}DAFLGSFLYEYSR^{359}
2	Chicken, Lysozyme c Precursor (5/39 Matches, 32% Seq Coverage)		
	517.2785	9.8	^{87}TPGSR^{91}
	874.4158	−1.5	^{33}HGLDNYR^{39}
	1045.5445	1.4	^{135}GTDVQAWIR^{143}
	1428.6496	−0.8	^{52}FESNFNTQATNR^{63}
	1675.7980	−2.1	^{116}IVSDGNGMNAWVAWR^{130}
3	Bovine, Hemoglobin β Chain (4/39 Matches, 37% Seq Coverage)		
	1098.5493	−8.3	^{95}LHVDPENFK^{103}
	1177.6701	−2.8	^{8}AAVTAFWGKVK^{18}
	1265.8291	−1.4	^{104}LLGNVLVVVLAR^{115}
	1274.7259	−0.1	^{30}LLVVYPWTQR^{39}
4	Rabbit, Glycogen Phosphorylase (5/39 Matches, 4% Seq Coverage)		
	549.3172	4.2	^{411}FLNR^{414}
	604.3550	15.1	^{244}LWSAK^{248}
	689.3688	9.5	^{571}IHEYK^{575}
	843.5015	8.9	^{353}VLVDLER^{359}
	1053.5706	1.2	^{833}QRLPAPDEK^{841}
5	Bovine, Hemoglobin α Chain (2/23 Matches, 21% Seq Coverage)		
	1529.7395	3.0	^{17}VGGHAAEYGAEALER^{31}
	1833.8894	−1.6	^{41}TYEPHEDLSHGSAQVK^{56}
6	Horse, Cytochrome *c* (2/23 Matches, 17% Seq Coverage)		
	779.4428	−7.9	^{80}MIFAGIK^{86}
	1168.6191	−3.1	^{28}TGPNLHGLFGR^{38}

The mass data of digest fragments from two differently denatured and one nondenatured protein mixture samples were separately entered into the MS-FIT program and searched against the Swissprot. protein database. Proteins identified from the database search are ranked on the basis of MOWSE search score instead of, more commonly used, the number of matching digest fragments. Table 1 shows the search result of the nondenatured protein mixture sample. The actual search results from MS-FIT are somewhat more complicated than Table 1 because identical proteins from different sources are also picked up and ranked high in the search results. The search scores for homologous proteins are slightly lower than the correct match with fewer matching digest fragments. These homologous proteins are not included in Tables 1 and 2. Although the mass errors for most ions are less than 10 ppm, mass errors for some ions exceed 10 ppm (e.g., 10−25 ppm). The high-mass-error ions are attributed to low S/N ratio ions (<4) in the mass spectra. Average mass error of all the digest fragments in the Table 1 is 4.7 ppm.

(34) Pappin, D. J. C.; Hojrup, P.; Bleasby, A. J. *Curr. Biol.* **1993**, *3*, 329−332.
(35) Russell, D. H.; Edmonson, R. D. *J. Mass Spectrom.* **1997**, *32*, 263−276.
(36) Edmondson, R. D.; Russell, D. H. In *Mass Spectrometry of Biological Materials*; Larsen, B. S., McEwen, C. N., Eds.; Marcl Dekker: New York, 1998.

Table 2. Proteins Identified from a Synthetic Protein Mixture (11 Proteins) by High-Resolution MALDI Peptide Mass Mapping

components of protein mixture	non-denatured	chemically denatured	thermally denatured
rabbit phosphorylase			7[a] (76)[b] 6%[c]
bovine serum albumin	8 (82) 11%	3 (82) 6%	12 (82) 15%
rabbit aldolase			5 (39) 13%
chicken ovalbumin			7 (34) 17%
rabbit carbonic anhydrase			4 (22) 29%
horse myoglobin		2 (21) 13%	3 (21) 24%
bovine hemoglobin β	4 (18) 37%	4 (18) 31%	7 (18) 48%
bovine hemoglobin α	2 (14) 21%	2 (14) 21%	3 (14) 30%
chicken lysozyme	5 (19) 32%	5 (19) 34%	7 (19) 48%
horse cytochrome *c*	3 (22) 22%	2 (22) 11%	4 (22) 29%
bovine ubiquitin			2 (12) 19%

[a] Numbers denote matching digest fragments used in the peptide mass mapping. [b] Numbers in parentheses denote theoretical number of digest fragments (MW >500). [c] Amino acid sequence coverage of the protein from matching digest fragments.

High-resolution, high-mass-accuracy MALDI peptide mass mapping identified all 11 proteins from the thermally denatured sample, whereas only 6 proteins are identified for the nondenatured and chemically denatured samples (See Table 2.). Note that in all cases higher amino acid sequence coverages and greater numbers of matching digest fragments are obtained for thermally denatured samples.

DISCUSSION

Our aim is to develop MALDI HR-TOF-MS methods that can be used for rapid identification of individual proteins present in protein mixtures. In this paper, we demonstrate the utility of this method using a synthetic protein mixture that contains proteins that are resistant to digestion and proteins that are readily digested using "in-solution" digest methods. The objective of this investigation is to illustrate that thermal denaturation of the proteins prior to digestion yields more uniform coverage peptide mass maps of the individual proteins. A second objective is to illustrate that peptide mass mapping with HR-TOF-MS and accurate mass measurements (mass errors, <10 ppm) can be used to identify the individual proteins without the use of tandem mass spectrometry peptide sequencing or partial sequencing.

Enzymatic digestion and peptide mass mapping of complex protein mixtures using MALDI TOF-MS for the purpose of protein identification are subject to numerous challenges. For synthetic protein mixtures, the concentrations of individual proteins can be controlled (1−5 μM), and this reduces problems associated with dynamic range, but the problems of identifying low-abundance proteins in the presence of high-abundance proteins will be addressed in a later paper where these same methods are applied to the analysis of whole cell lysates.[37] Another factor to consider is that larger proteins yield a greater number of proteolytic peptides; however, the number of digest fragments does not depend exclusively on the size of the protein (See Table 1,). For example, some proteins maintain rigid higher-order structure and show different sensitivities to digestion.[23,25,26] Although it is generally true that denaturation (both chemical and thermal) of the protein increases the digestion efficiency and reduces the time required to carry out digestion,[23] chemically denatured samples generally require purification steps to eliminate species that reduce the ionization efficiency of MALDI. The benefit of thermal denaturation for enzymatic digestion/peptide mass mapping of single-component protein samples extends to the analysis of protein mixtures. For example, thermal denaturation appears to facilitate efficient digestions of proteins that are insensitive to proteolytic enzymes. Specific examples are myoglobin and ovalbumin, which are both protease resistant, and neither protein is detected in peptide mass maps of nondenatured protein mixture samples. In addition, lysozyme, ubiquitin, aldolase, and serum albumin show drastic enhancements of enzymatic digestion following thermal denaturation.

Proteins containing disulfide bonds, e.g., albumin, lysozyme, and ovalbumin, are identified in the thermally denatured samples without the need of reduction/alkylation of cysteine residues. Only a few digest fragments containing carbamidomethylated cysteine were identified after alkylation of cysteine residues by iodoacetamide. Including these digest fragments in the database search does not significantly affect search results. Tryptic digestion of thermally denatured proteins usually generates a better mass map unless the protein has high content of disulfide bonds relative to its molecular weight. For example, intact bovine insulin has a molecular weight of 5730 and contains three disulfide bonds.[38] Two chains (a and b) of bovine insulin are connected by one intra- and two interchain disulfide bonds, and this kind of detail is not considered in the protein databases. The peptide digestion map of thermally denatured bovine insulin is quite different from that reported in the protein database; consequently, identification by peptide mass mapping even with thermal denaturation is not effective.

Previously, protein identification using mass spectrometry was performed on single-component or simple mixture (3−4 components) fractions from the separation. Since rather complicated protein mixtures (10−15 components) can be directly identified by high-resolution, high-mass-accuracy MALDI peptide mass mapping, we project that such capabilities will reduce the number of samples that must be prepared for subsequent mass spectral analysis by ∼10-fold. The need for complete chromatographic separation of individual protein components from mixture samples is minimized. In addition, protein database search with the MOWSE scoring system can identify four or five proteins in a single search attempt because correct protein matches have much higher (1 or 2 orders of magnitude) MOWSE search scores than false positive matches. In the samples analyzed, 11 proteins were identified in three search attempts. Data analysis time is shortened by fewer search attempts compared to conventional single-component second pass search. Moreover, identification of proteins from a single mass spectrum drastically reduces the amount of data that must be acquired and stored. This reduction

(37) Park, Z.-Y.; Russell, D. H., in preparation.
(38) Voet, D.; Voet, J. G. *Biochemistry*, 2nd ed.; John Wiley & Sons: New York, 1995.

in volume of data becomes even more important as the samples become more complex, e.g., whole cell lysates.[37]

High-accuracy mass measurement greatly reduces the number of protein candidates by limiting the possible amino acid compositions of matching digest fragments.[8,39,40] The MOWSE search score system combined with high-mass-accuracy analysis demonstrates a dramatic increase in discriminating power against false positive matches. Although the MOWSE search score system was not designed for identification of individual proteins present in complex mixtures, we have found it very effective for this type of experiments. Other search engines such as Profound,[41,42] PeptIdent2,[43] and Mascot,[44] which may be more effective for identifying individual proteins in complex mixtures, have been developed. Some comparison studies of single-protein identification results between these peptide mass fingerprinting programs have been reported,[43,45] but a concerted effort to evaluate the performance of individual search programs for protein mixture identification has not been reported. We are currently critically evaluating these programs for direct mixture analysis.

The confidence level of database search results with digest fragment masses alone is a controversial topic because identification is solely based on the mass (m/z). Peptide mass database searches with partial sequence information from chemical modifications (C-terminal modification, alkylation of cysteine residues) or amino acid composition from tandem mass spectrometric techniques provide increased confidence levels,[40,46−49] but higher confidence level analysis requires longer analysis times and additional sample preparation steps. Recently, peptide mass search with N-terminal partial sequence information by rapid chemical degradation has shown drastic increase of confidence level with minimized loss of analysis time and efforts;[50] however, these approaches cannot yet compete with the speed of direct protein mixture identification via high-resolution, high-mass-accuracy MALDI peptide mass mapping. An alternative way to increase the confidence level for peptide mass search results is to employ parallel proteolytic enzymes. Contrary to the in-gel digestion method, introducing a second proteolytic enzyme to increase the confidence levels of the search is not difficult considering that such experiment can be carried out in parallel.

Several experimental factors strongly influence the quality of MALDI TOF mass spectra and search results. First, the presence of chemicals such as buffers, denaturants, and reducing and alkylating reagents seriously influence ionization efficiency of digest fragments. Chemically denatured protein mixtures generate more digest fragments than nondenatured sample, but similar or fewer numbers of ions are always observed in the mass spectra without further purification. Although repeated washings show some enhancement of MALDI signal, the washing alone does not overcome less efficient ionization of digest fragments in the presence of chemical contaminants. Second, MALDI sample preparation methods and M/A ratio can also determine the quality of MALDI TOF mass spectra.[29−32] The overlayer sample preparation appears to give consistent spot-to-spot reproducibility because it forms more homogeneous crystal surfaces than the more commonly used dried droplet method. Depending upon the M/A ratio used, variations in the number of proteins identified and different amino acid sequence coverages are observed (See Table 3.). When concentration information of protein samples is not available, optimum M/A ratio can be determined by parallel analysis of different sample preparations.

Table 3. Effects of Matrix-to-Analyte Ratio (M/A) of MALDI Sample Preparation on the Identification of Chemically Denatured Synthetic Protein Mixture Sample

components of protein mixture	M/A ≈ 4000	M/A ≈ 1600
bovine serum albumin	6[a] (82)[b] 9%[c]	7 (82) 12%
chicken egg ovalbumin	3 (34) 10%	7 (34) 24%
horse myoglobin	4 (21) 24%	5 (21) 35%
bovine hemoglobin β	4 (18) 38%	
bovine hemoglobin α	2 (14) 19%	
chiken lysozyme	5 (19) 34%	4 (19) 23%

[a] Numbers denote matching digest fragments used in the peptide mass mapping. [b] Numbers in parentheses denote theoretical number of digest fragments (MW >500). [c] Amino acid sequence coverage of the protein from matching digest fragments.

CONCLUSIONS

Thermal denaturation/in-solution digestion with high-resolution, high-mass-accuracy (average error of ∼5 ppm) MALDI peptide mass mapping is a rapid, versatile protein identification tool for mixture analysis. This method can be combined with powerful chromatographic methods; however, the high-resolution separation capabilities of mass spectrometry minimize requirements for sample handling/purification. Using MALDI HR-TOF-MS, a whole cell protein identification procedure can be carried out in a single day. Automation of separation and sample preparation protocols, which can be carried out in parallel, could greatly increase sample throughput. Palmblad and Smith have described similar high-resolution/high-mass accuracy using FT-MS.[27,28,51] The overall complexity of FT-MS, such as the requirement for six stages of differential pumping, is far greater than for TOF-MS. Another disadvantage for high-throughput applications of FT-MS is the time requirements for acquisition of high-

(39) Biemann, K. *Methods Enzymol.* **1990**, *193*, 295−305.
(40) Clauser, K. R.; Baker, P.; Burlingame, A. L. *Anal. Chem.* **1999**, *71*, 2871−2882.
(41) Eriksson, J.; Chait, B. T.; Fenyo, D. *Anal. Chem.* **2000**, *72*, 999−1005.
(42) Zhang, W.; Chait, B. T. *Anal. Chem.* **2000**, *72*, 2482−2489.
(43) Gras, R.; Muller, M.; Gasteiger, E.; Gay, S.; Binz, P.-A.; Bienvenut, W.; Hoogland, C.; Sanchez, J.-C.; Bairoch, A.; Hochstrasser, D. F.; Appel, R. D. *Electrophoresis* **1999**, *20*, 3535−3550.
(44) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551−3567.
(45) Tang, C.; Zhang, W.; Fenyo, D.; Chait, B. T. *48th ASMS Conference*; Long Beach, CA, 2000.
(46) Sechi, S.; Chait, B. T. *Anal. Chem.* **1998**, *70*, 5150−5158.
(47) Kuster, B.; Mann, M. *Anal. Chem.* **1999**, *71*, 1431−1440.
(48) Goodlett, D. R.; Bruce, J. E.; Anderson, G. A.; Rist, B.; Pasa-Tolic, L.; Fiehn, O.; Smith, R. D.; Aebersold, R. *Anal. Chem.* **2000**, *72*, 1112−1118.
(49) Kosaka, T.; Takazawa, T.; Nakamura, T. *Anal. Chem.* **2000**, *72*, 1179−1185.
(50) Hoving, S.; Munchbach, M.; Schmid, H.; Signor, L.; Lehmann, A.; Staudenmann, W.; Quadroni, M.; James, P. *Anal. Chem.* **2000**, *72*, 1006−1014.
(51) Palmblad, M.; Wetterhall, M.; Markides, K.; Hakansson, P.; Bergquist, J. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1029−1034.

resolution spectra, e.g., hundreds per spectra is typical, but the requirement for signal-averaging to achieve adequate S/N ratios increases the time by at least 10-fold. The time requirements can also be reduced by operating at more modest mass resolution. For example, Bruce et al. demonstrated mass resolution of ~20 000, which is equivalent to that for high-performance TOF-MS with data acquisition times of 2−3 min. The utility of FT-ICR for direct mixture analysis is also limited by the maximum ion densities that can be accumulated in the ion cell; e.g., 10^6-10^8 ions depend on the cell geometry and magnetic field strength.[52] The limitations of ion density can be reduced by acquiring the mass spectra in small segments, e.g., 100−200 m/z ranges;[52] however, this procedure compounds the problem of long signal acquisition cycles.

(52) Guan, S. H.; Marshall, A. G.; Scheppele, S. E. *Anal. Chem.* **1996**, *68*, 46−71.

Clearly, the rapid analysis times of TOF-MS are highly attractive for high sample throughput application. Such analysis capability could be of even greater significance if our efforts to increase the mass resolution of TOF-MS from ~10 000−20 000 to >100 000 are successful.

ACKNOWLEDGMENT

This project is funded by the U. S. Department of Energy, Division of Chemical Sciences, Basic Energy Sciences (DE-FG03-95ER14505), the National Science Foundation (CHE-9629966), and the National Institute of Environmental Health Sciences (P30-ES09106).

Received for review December 19, 2000. Accepted March 13, 2001.

AC001488P