

## Technical Notes

# Proteolysis in Mixed Organic–Aqueous Solvent Systems: Applications for Peptide Mass Mapping Using Mass Spectrometry

William K. Russell, Zee-Yong Park, and David H. Russell\*

The Laboratory for Biological Mass Spectrometry, Department of Chemistry, Texas A&M University, College Station, Texas 77843

**The rate of protein digestion imposes significant limitations on high-throughput protein identification using mass spectrometry. In this report, we demonstrate that proteins are readily digested by trypsin in the presence of organic solvents such as methanol, acetone, 2-propanol, and acetonitrile. The rates of protein digestion in organic solvents, as indicated by the abundances of digest fragment ions in the mass spectrum, are increased relative to aqueous solution. In addition, amino acid coverage for the analyzed proteins increases in the presence of the organic solvents, and proteins that are resistant to proteolysis are readily digested. For example, a 68% amino acid sequence coverage was attained from a tryptic digest of myoglobin in <5 min from an 80% acetonitrile solution, whereas no digest fragments were detected from a 5 min digestion in an aqueous solution. Moreover, the tryptic digestion of a complex protein mixture in an organic–aqueous solvent system showed significantly enhanced digestion for nearly all of the protein components. Enzymatic digestion in an organic–aqueous solvent system is a rapid, simple, and effective peptide mass-mapping technique.**

With the rapid discovery of complete genomes, the challenge of proteomics is to identify and characterize proteins encoded by the genomes as efficiently as possible. Complex mixtures of proteins are typically separated using 1D- or 2D-polyacrylamide gel electrophoresis (PAGE) followed by in-gel proteolytic digestion.<sup>1–7</sup>

The resulting peptides are analyzed with matrix-assisted laser desorption ionization (MALDI) mass spectrometry, and this information is interrogated by searching the protein and nucleic acid databases. Although this procedure has become routine, it is inefficient regarding the time required to prepare and perform gel separation and the percentage of proteins actually visualized. Despite improvements in 2D-gel electrophoresis, very large or very small hydrophobic proteins, as well as proteins with high and low pI, are not easily identified on these gels.<sup>8–11</sup>

Recently, there have been efforts to analyze complex protein mixtures without the use of gel electrophoresis by separating mixtures using liquid chromatography and then proteolytically digesting the fractions.<sup>8,12–14</sup> In-solution digestion is generally less time-consuming, and protein coverage can be more complete than for in-gel digestion. Park et al. showed that thermal denaturation of the proteins prior to digestion greatly aided in-solution digests of soluble, protease-resistant proteins.<sup>15</sup> A similar enhancement was observed in protein mixture samples.<sup>16,17</sup> Such innovations for in-solution digestion do not impact the analysis of membrane proteins, which are insoluble in ordinary aqueous buffer systems. To solubilize these proteins, MALDI-unfriendly reagents, such as detergents, are added to the sample preparation procedure; however, these reagents must be removed from the digest fragments prior to mass spectral analysis.

\* Corresponding author. Phone: 979-845-3345. Fax: 979-845-9485. E-mail: Russell@mail.chem.tamu.edu.

- (1) Rosenfeld, J.; Capdevielle, J.; Guillemot, J. C.; Ferrara, P. *Anal. Biochem.* **1992**, *203*, 173–179.
- (2) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850–858.
- (3) Shevchenko, A.; Jensen, O. N.; Podtelejnikov, 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.
- (4) Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. *Anal. Chem.* **1997**, *69*, 4741–4750.
- (5) Jungblut, P.; Thiede, B. *Mass Spectrom. Rev.* **1997**, *16*, 145–162.
- (6) 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.

- (7) 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.
- (8) 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.
- (9) Nilsson, C. L.; Larsson, T.; Gustafsson, E.; Karlsson, K.-A.; Davidsson, P. *Anal. Chem.* **2000**, *72*, 2148–2153.
- (10) le Coutre, J.; Whitelegge, J. P.; Gross, A.; Turk, E.; Wright, E. M.; Kaback, H. R.; Faull, K. F. *Biochemistry* **2000**, *39*, 4237–4242.
- (11) Washburn, M. P.; Yates, J. R., III. *Curr. Opin. Microbiol.* **2000**, *3*, 292–297.
- (12) Opiteck, G. J.; Lewis, K. C.; Jorgenson, J. W. *Anal. Chem.* **1997**, *69*, 1518–1524.
- (13) Dai, Y.; Li, L.; Roger, D. C.; Long, S. R. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 73–78.
- (14) Wall, D. B.; Lubman, D. M.; Flynn, S. J. *Anal. Chem.* **1999**, *71*, 3894–3900.
- (15) Park, Z.-Y.; Russell, D. H. *Anal. Chem.* **2000**, *72*, 2667–2670.
- (16) Park, Z.-Y.; Russell, D. H. 13th Symposium of the Protein Society, Boston, MA 1999.
- (17) Park, Z.-Y.; Russell, D. H. *Anal. Chem.* **2001**, in press.

A convenient approach to solubilizing proteins is to change the solvent system;<sup>18–21</sup> this has the added potential of accelerating proteolytic digestion by denaturing the protein.<sup>22,23</sup> Trypsin maintains its proteolytic activity in a variety of organic solvent systems, even at high concentration levels.<sup>23–27</sup> The use of enzymes that retain high activity in organic or mixed solvents, combined with the denaturing and increasing solubility of substrate proteins in organic solvents, could significantly impact high-throughput analysis of proteins via in-solution digestion.<sup>23,27</sup>

Digesting proteins in mixed-solvent systems such as methanol–water, acetonitrile–water, 2-propanol–water, or acetone–water for the purpose of protein identification and mapping is shown to be more efficient than conventional approaches. For example, we are able to achieve greater amino acid coverage of the target protein in shorter periods of time and minimize the need for detergents or chemical denaturants. This is the first report of utilizing proteolytic enzymes in organic solvent media for rapid protein identification purposes using mass spectrometry.

## EXPERIMENTAL SECTION

All proteins were purchased from Sigma (St. Louis, MO), and modified porcine trypsin (sequencing grade) was purchased from Promega (Madison, WI). Methanol, 2-propanol, acetonitrile, and acetone were of HPLC grade or higher. A synthetic protein mixture system was constructed using 11 proteins: rabbit phosphorylase (MW, 98 kDa), bovine serum albumin (MW, 66 kDa), chicken egg ovalbumin (MW, 44 kDa), rabbit aldolase (MW, 39 kDa), bovine carbonic anhydrase (MW, 29 kDa), horse myoglobin (MW, 17 kDa), bovine hemoglobin  $\alpha$ ,  $\beta$  (MW, 15 kDa), horse cytochrome *c* (MW, 12 kDa), chicken egg lysozyme (MW, 14 kDa), and bovine ubiquitin (MW, 8 kDa). The proteins were dissolved in 50 mM ammonium bicarbonate (Sigma) solution having  $\sim 1\text{--}3\ \mu\text{M}$  concentrations of each protein. Trypsin solutions were prepared as directed by the supplier and added to the protein mixture in a molar ratio of 1:50. The digest mixture was sampled and analyzed by mass spectrometry over a 5-h period of time.

The cytochrome *c* (18  $\mu\text{M}$ ) digest was performed in 7 mM ammonium bicarbonate buffer, pH  $\sim 8$  with concentrations of organic solvent as high as 80%. Myoglobin (13  $\mu\text{M}$ ) was digested in a solution of 80% acetonitrile and 25 mM ammonium bicarbonate buffer, pH  $\sim 8$  (final concentration) and was compared to a control sample digested in 25 mM ammonium bicarbonate. Trypsin was added to the protein(s) of interest in a molar ratio of 1:50. The reaction was incubated in a water bath at 37 °C, and aliquots were sampled over time.

High-resolution MALDI time-of-flight (HR-MALDI-TOF) mass spectral analysis was performed by taking 2  $\mu\text{L}$  of the digest

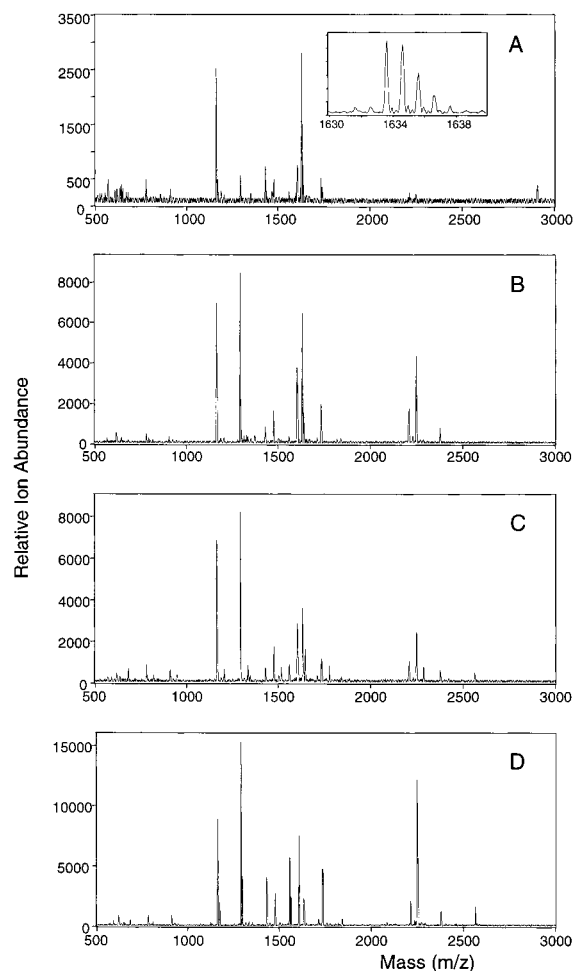


Figure 1. Tryptic digest (1-h) of cytochrome *c* in (A) water, (B) 60% methanol, (C) 40% acetone, and (D) 40% acetonitrile.

sample and diluting it with 18  $\mu\text{L}$  of water. The resulting solution was mixed 1:1 with 15 mM  $\alpha$ -cyano-4-hydroxycinnamic acid matrix. Samples were spotted using the overlay method.<sup>28–30</sup> A Perceptive Biosystems Voyager Elite XL TOF described elsewhere was used for the analysis.<sup>29,31</sup> Samples were externally calibrated. Database mass searching was performed using MS-Fit, and database, SwissProt., as described elsewhere.<sup>15,17</sup>

## RESULTS

Cytochrome *c* is very susceptible to in-solution enzymatic digestion.<sup>15,16</sup> To evaluate the effects of organic solvents on the digestion and peptide mass mapping, cytochrome *c* was digested with trypsin in methanol–water, acetonitrile–water, 2-propanol–water, or acetone–water solutions. Figure 1 contains the HR-MALDI-TOF mass spectra of the digest after a 1-h incubation with trypsin at 37 °C. The optimum organic concentration (based on the yield of peptide ions) for the mixed solvent system for cytochrome *c* was found to be 60/40 for methanol–water (Figure 1B), 40/60 for acetone–water (Figure 1C), and 40/60 for aceto-

- (18) Yerushalmi, H.; Lebediker, M.; Schuldiner, S. *J. Biol. Chem.* **1995**, *270*, 6856–6863.  
 (19) Molloy, M. P.; Herber, B. R.; Williams, K. L.; Gooley, A. A. *Electrophoresis* **1999**, *20*, 701–704.  
 (20) Gomez-Puyou, A. *Biomolecules in Organic Solvents*; CRC Press: Boca Raton, FL, 1992.  
 (21) Bhardwaj, S.; Day, R. A. *Anal. Lett.* **1998**, *31*, 2625–2634.  
 (22) Fink, A. L.; Painter, B. *Biochemistry* **1987**, *26*, 1665–1671.  
 (23) Welinder, K. G. *Anal. Biochem.* **1988**, *174*, 54–64.  
 (24) Zaks, A.; Klibanov, A. M. *J. Biol. Chem.* **1988**, *263*, 8017–8021.  
 (25) Klibanov, A. M. *Trends Biotechnol.* **1997**, *15*, 97–101.  
 (26) Simon, L. M.; Laszlo, K.; Vertesi, A.; Bagi, K.; Szajani, B. *J. Mol. Catal. B: Enzym.* **1998**, *4*, 41–45.  
 (27) Houen, G.; Sando, T. *Anal. Biochem.* **1991**, *193*, 186–190.

- (28) Vorm, O.; Roepstorff, P.; Mann, M. *Anal. Chem.* **1994**, *66*, 3281–3287.  
 (29) Edmondson, R. D.; Russell, D. H. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 995.  
 (30) Dai, Y.; Whittal, R. M.; Li, L. *Anal. Chem.* **1999**, *71*, 1087–1091.  
 (31) Barbacci, D. C.; Edmondson, R. D.; Russell, D. H. *Int. J. Mass Spectrom.* **1997**, *165*, 221.

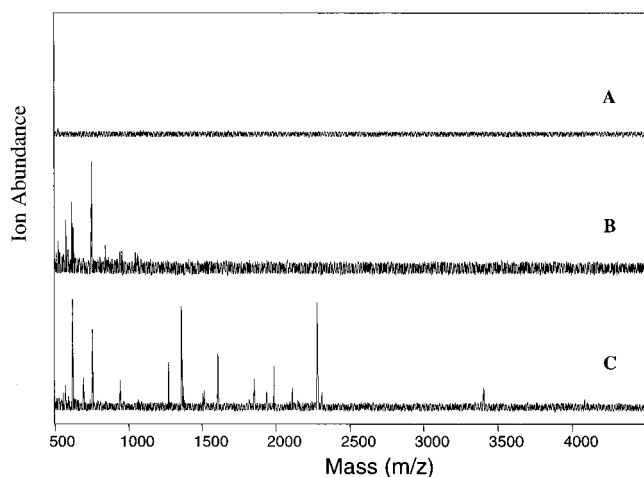


Figure 2. (A) Myoglobin digest in water after 5 min, (B) myoglobin digest in water after 1 h, and (C) myoglobin digest in 80% acetonitrile after 5 min.

Table 1. Yield of an Aqueous Solution Digest Compared to One in a Methanol–Water Solution

protein	amino acid coverage, %	
	digest, water	digest, methanol–water
rabbit phosphorylase	0	14
bovine serum albumin	13	17
bovine hemoglobin $\alpha$	21	41
bovine hemoglobin $\beta$	30	46
rabbit aldolase	9	7
chicken ovalbumin	0	17
rabbit carbonic anhydrase	0	8
horse myoglobin	0	40
horse cytochrome <i>c</i>	30	44
chicken lysozyme	29	43
bovine ubiquitin	0	85

nitrile–water (Figure 1D). The digest fragments for the aqueous solvent system cover 58% of the entire amino acid sequence, whereas sequence coverages of 66%, 79%, and 72% were obtained for the methanol–water, acetone–water, and acetonitrile–water solutions, respectively.

Globular proteins, such as myoglobin, are known to be somewhat resistant to proteolytic digestion.<sup>15,16,32</sup> Figure 2 contains the HR-MALDI-TOF mass spectrum for a 5-min digest of myoglobin in a solution containing 80% acetonitrile (Figure 2C) compared to a digest in an aqueous solution. The peptide map covers 104 of the 153 (68%) amino acids possible in the acetonitrile solution. In contrast, no peptide fragments are detected in the aqueous digest (Figure 2A). Even after 1 h, the myoglobin digest yielded relatively few peaks (Figure 2B).

A mixture of 11 proteins in a methanol–water solution was digested and compared with a digest of the same proteins for an aqueous solution (Table 1). All of the proteins in the methanol–water digest solution were identified, as compared to only 6 of the 11 proteins in the aqueous case. The amino acid coverage was generally higher in the methanol–water digest, as well.

## DISCUSSION

The time required for 1D- and 2D-PAGE gels separation and problems associated with the visualization of proteins provides

ample justification for the development of alternative analysis methods.<sup>8,33,34</sup> We have found that protein mass mapping from organic–aqueous solvent mixtures is possible, and the digestion takes place at a highly accelerated rate. The simplicity and applicability of the technique make it especially attractive for proteomic studies.

For single protein samples, digestion in mixed solvents yields much higher ion signals when compared to digestion in aqueous solution. The amino acid sequence coverages were also higher in the mixed-solvent digestion system. There are some differences in the peptide maps generated by each of the three different organic solvents as well as the aqueous solution case. After protein database searching, most of these unique masses were identified as digest fragments containing missed cleavage sites. For example, trypsin did not cleave between contiguous lysine groups (...XXXXKK or ...XXXXKKK) in the presence of high concentrations of methanol, even after prolonged periods of time. The reason(s) for missed cleavages at hydrophilic sites is unclear, but this could be the result of a reduction of solvation at these sites in the presence of organic solvents. Additional experiments to probe this abnormality are currently underway.

Myoglobin and chicken ovalbumin are resistant to proteolysis unless they are denatured, either chemically or thermally;<sup>15,16,32</sup> however, in the presence of mixed solvents, both of these proteins are readily digested, which suggests that the solvent system induces significant structural changes. Initially we were surprised to find that a difficult-to-digest protein such as myoglobin was efficiently digested after only 1 h. We then performed a series of time-course studies and found that myoglobin was efficiently digested after only 5 min. For example, using an 80% acetonitrile solution, myoglobin yielded 15 peptide ion signals corresponding to 68% of the amino acid coverage. The aqueous digest requires ~14.5 h to yield similar amino acid coverage (52%, data not shown). We attribute the accelerated digestion to denaturation of the protein (in this case, myoglobin or ovalbumin), but we cannot rule out the effects of increased solubility of the peptide fragments.

The number of digest fragments detected for an 11-component protein mixture in mixed solvent were more than adequate for protein identification using HR-MALDI-TOFMS and database searching (hemoglobin  $\beta$  had the lowest number of mass peaks, 5). Conversely, we could only identify 6 of the 11 proteins in digests performed using aqueous solution, and aqueous digest generally resulted in lower amino acid coverages (see Table 1). The average increase in amino acid coverage is 21%, with the highest increase being from 0% to 85% (ubiquitin).

The activity of trypsin toward digestion of cytochrome *c* was monitored while increasing the fraction of organic solvent and measuring the abundances of peptide fragments in the HR-MALDI-TOF mass spectrum. Under conditions used for this work, trypsin maintains proteolytic activity in mixed-solvent systems up to 80% organic, in agreement with findings of Simon et al.<sup>26</sup> For example, Simon and co-workers investigated the enzymatic activities of the hydrolytic enzymes trypsin, chymotrypsin, carboxypeptidase, and lipase in an organic–aqueous solvent mixture

(33) Shen, Y.; Xiang, F.; Veenstra, T. D.; Fung, E. N.; Smith, R. D. *Anal. Chem.* **1999**, *71*, 5348–5353.

(34) Link, A.; Eng, J.; Schieltz, D.; Carmack, E.; Mize, G.; Morris, D.; Garvik, B.; Yates, J. R., III. *Nat. Biotechnol.* **1999**, *17*, 676–682.

and reported that the activity of trypsin remained fairly constant in a variety of solvents, including acetonitrile, in concentrations up to 80% organic. We see an apparent increase in the "activity" of trypsin, that is, we detect more digest fragments in shorter periods of time. Whether this is an actual increase in the activity or a combination of other factors, such as solubility of the substrate, conformation of the substrate, or an opening of the active site, is currently being evaluated. We should point out that we reacted trypsin with proteins, whereas Simon et al. reacted trypsin with the peptide mimic *N*-benzoyl-L-arginine ethyl ester.<sup>26</sup> The size of this common substrate is small, and any change in conformation under different solvent conditions is almost negligible. On the other hand, the proteins that were investigated here are relatively large, and significant conformational changes have been reported in organic solvent systems.<sup>35,36</sup> We also note that Welinder's observations are consistent with ours.<sup>23</sup>

A similar serine protease, subtilisin, has been thoroughly studied in various organic solvent systems.<sup>24,25,37-41</sup> These studies show that organic solvent induces significant structural changes at the secondary structure level,<sup>39,40</sup> but catalytic activity of subtilisin was still observed.<sup>24,41</sup> Similarly, the organic solvent does not seem to negatively effect trypsin; if it does, the denaturation occurs over a time period that is longer than the time of the

experiment. The overall specificity of the trypsin, with the exception of miss cleavages, as noted above, seems to be unaltered by the solvent systems used in this study, that is, all of the masses corresponded to peptide fragments with arginine or lysine at the C-terminus.

## CONCLUSIONS

In this report, we show that proteolysis in mixed organic-aqueous solvent systems is highly efficient in terms of both rate of digestion and amino acid sequence coverage. Efficient digestion is crucial to the successful identification of individual proteins in mixtures. The higher digestion yield for each protein correlates to more digest fragments, thereby increasing the confidence level of database search results. The application of this technique to water-insoluble proteins, such as membrane proteins, could prove to be of great benefit. Moreover, one could imagine modifying Welinder's experiment to separate complex protein mixtures using reverse-phase HPLC and digesting the fractions immediately without the removal of the solvent.<sup>23</sup> The fractions could then be directly analyzed by MALDI- or ESI-MS. Similarly, the works of Welinder and Simon suggest that other proteases could show similar enhancement of digestion in mixed-solvent systems.<sup>23,26</sup>

## ACKNOWLEDGMENT

This project was 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 November 13, 2000. Accepted March 5, 2001.

AC001332P

(35) Kamatari, Y. O.; Konno, T.; Kataoka, M.; Akasaka, K. *J. Mol. Biol.* **1996**, *259*, 512-523.

(36) Kamatari, Y. O.; Ohji, S.; Konno, T.; Seki, Y.; Soda, K.; Kataoka, M.; Akasaka, K. *Protein Sci.* **1999**, *8*, 873-882.

(37) Zaks, A.; Klibanov, A. M. *J. Am. Chem. Soc.* **1986**, *108*, 2767-2768.

(38) Zaks, A.; Klibanov, A. M. *J. Biol. Chem.* **1988**, *263*, 3194-3201.

(39) Griebenow, K.; Klibanov, A. M. *J. Am. Chem. Soc.* **1996**, *118*, 11695-11700.

(40) Griebenow, K.; Klibanov, A. M. *Biotechnol. Bioeng.* **1997**, *53*, 351-362.

(41) Castro, G. R. *Enzyme Microb. Technol.* **1999**, *25*, 689-694.