Anal. Chem. **2006,** *78,* 125−134

Efficient and Specific Trypsin Digestion of Microgram to Nanogram Quantities of Proteins in Organic−Aqueous Solvent Systems

Michael Brad Strader,[†,‡] David L. Tabb,[§,∥] W. Judson Hervey,[†,⊥] Chongle Pan,[†,⊥] and Gregory B. Hurst*[,†]

Organic and Biological Mass Spectrometry Group, Chemical Sciences Division, and Genome Analysis and Systems Modeling Group, Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, and Graduate School of Genome Science and Technology, University of Tennessee−Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

Mass spectrometry-based identification of the components of multiprotein complexes often involves solution-phase proteolytic digestion of the complex. The affinity purification of individual protein complexes often yields nanogram to low-microgram amounts of protein, which poses several challenges for enzymatic digestion and protein identification. We tested different solvent systems to optimize trypsin digestions of samples containing limited amounts of protein for subsequent analysis by LC−MS−MS. Data collected from digestion of 10-, 2-, 1-, and 0.2-μg portions of a protein standard mixture indicated that an organic−aqueous solvent system containing 80% acetonitrile consistently provided the most complete digestion, producing more peptide identifications than the other solvent systems tested. For example, a 1-h digestion in 80% acetonitrile yielded over 52% more peptides than the overnight digestion of 1 μg of a protein mixture in purely aqueous buffer. This trend was also observed for peptides from digested ribosomal proteins isolated from *Rhodopseudomonas palustris*. In addition to improved digestion efficiency, the shorter digestion times possible with the organic solvent also improved trypsin specificity, resulting in smaller numbers of semitryptic peptides than an overnight digestion protocol using an aqueous solvent. The technique was also demonstrated for an affinity-isolated protein complex, GroEL. To our knowledge, this report is the first using mass spectrometry data to show a linkage between digestion solvent and trypsin specificity.

Mass spectrometry (MS) has become a widely used method for studying proteins, protein complexes, and whole proteomes because of innovations in soft ionization techniques, bioinformatics, and chromatographic separation techniques.[1−7] An example of a high-throughput mass spectrometry strategy commonly used for this purpose is a variation of the "shotgun" approach, involving in-solution digestion of a protein complex followed by one-dimensional (1D) or two-dimensional (2D) liquid chromatography (LC) coupled with electrospray ionization (ESI) MS−MS.[6−8] One of the applications of this method is for characterizing multiprotein complexes by identifying large numbers of proteins in a single data acquisition.[9] Large-scale implementations of this strategy have been reported for yeast and *Escherichia coli*.[10−12] To achieve a goal of characterizing large numbers of protein complexes[13] isolated

* To whom correspondence should be addressed. Phone: (865) 574-6142. Fax: (865) 576-8559. E-mail: hurstgb@ornl.gov.

† Organic and Biological Mass Spectrometry Group, Chemical Sciences Division, Oak Ridge National Laboratory.

‡ Present address: Laboratory of Neurotoxicology, National Institute of Mental Health, Bethesda, MD 20892.

§ Genome Analysis and Systems Modeling Group, Life Sciences Division, Oak Ridge National Laboratory.

∥ Present address: Department of Biomedical Informatics, Vanderbilt University Medical Center 465 21st Ave. S., U9211 MRB III, Nashville, TN 37232-8575.

⊥ University of Tennessee−Oak Ridge National Laboratory.

(1) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64−71.
(2) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1991**, *63*, 1193A−1203A.
(3) Yates, J. R., 3rd; Eng, J. K.; McCormack, A. L.; Schieltz, D. *Anal. Chem.* **1995**, *67*, 1426−1436.
(4) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551−3567.
(5) Tabb, D. L.; McDonald, W. H.; Yates, J. R., 3rd. *J. Proteome Res.* **2002**, *1*, 21−26.
(6) McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R., 3rd. *Anal. Chem.* **1997**, *69*, 767−776.
(7) Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd. *Anal. Chem.* **2001**, *73*, 5683−5690.
(8) Hunt, D. F.; Yates, J. R., 3rd; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233−6237.
(9) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd. *Nat. Biotechnol.* **1999**, *17*, 676−682.
(10) Gavin, A. C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; Hofert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. *Nature* **2002**, *415*, 141−147.
(11) Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; Yang, L.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A. R.; Sassi, H.; Nielsen, P. A.; Rasmussen, K. J.; Andersen, J. R.; Johansen, L. E.; Hansen, L. H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sorensen, B. D.; Matthiesen, J.; Hendrickson, R. C.; Gleeson, F.; Pawson, T.; Moran, M. F.; Durocher, D.; Mann, M.; Hogue, C. W.; Figeys, D.; Tyers, M. *Nature* **2002**, *415*, 180−183.
(12) Butland, G.; Peregrin-Alvarez, J. M.; Li, J.; Yang, W.; Yang, X.; Canadien, V.; Starostine, A.; Richards, D.; Beattie, B.; Krogan, N.; Davey, M.; Parkinson, J.; Greenblatt, J.; Emili, A. *Nature* **2005**, *433*, 531−537.
(13) http://www.ornl.gov/sci/GenomestoLife/index.shtml.

10.1021/ac051348l CCC: $33.50 © 2006 American Chemical Society
Published on Web 11/22/2005

by affinity purification from *Rhodopseudomonas palustris*,[14] an efficient protocol for digesting these complexes is required.

The isolation of individual protein complexes by affinity purification often yields small amounts of protein, complicating enzymatic digestion and peptide identification.[15,16] In many cases, samples may contain less than 100 ng of total protein, resulting in low protein concentrations (≤10 ng/μL) that are unsuitable for efficient enzymatic digestion. One could keep protein concentrations higher by working in smaller volumes, but unless microfluidic techniques are invoked, a few microliters is a lower practical volume limit for conventional pipetting techniques. Because the reaction rate of digestion is proportional to the concentration,[17] dilute protein concentrations may lengthen digestion times. If digestion remains incomplete, large peptide fragments from incompletely digested proteins may not be suitable for identification via MS−MS analysis. For example, these larger peptides may be outside the scan range of the mass spectrometer. Other characteristics of large peptides, such as higher charge states (>3) or multiple missed enzymatic cleavage sites, may place them beyond the default settings for software used for automatic peptide identification from MS or MS−MS data. To promote more complete digestion, it is useful to identify proteolytic digestion methods that maximize peptide yields from samples containing low-microgram to high-nanogram amounts of total protein.

Digestion strategies may employ organic solvents, heat, chaotropes, or surfactants to denature proteins before digestion to render more of the protein's structure accessible to the proteolytic enzyme.[18−21] Thermal denaturation often results in sample loss due to precipitation because many proteins are susceptible to aggregation when treated with heat. Chaotropes and surfactants, on the other hand, can inactivate proteases at high concentrations required for denaturation and therefore require dilution of the sample prior to digestion.[22] Dilution permits proteins to refold, reduces proteolytic activity by decreasing substrate concentration, and can increase surface area available for adsorptive loss of peptides on container walls. In addition, chaotropes and surfactants can compete with peptide ions for adsorption on stationary phases during liquid chromatography and for charge during mass spectrometry.[21,23,24] To avoid this interference, these chemicals can be removed, although this step can lead to additional loss of peptides. An acid-labile surfactant has recently been introduced that decomposes into insoluble degradation products.[25] These degradation products, however, can coprecipitate with hydrophobic peptides and result in sample loss. To minimize the required amount of starting material while maximizing sensitivity in the MS measurement, one must avoid when possible the drawbacks associated with the described digestion strategies.

Addition of organic solvents such as methanol or acetonitrile to buffers can assist in digestion by unfolding and solubilizing proteins.[26,27] Trypsin and some other proteases are resistant to unfolding in organic solvents, retaining activity under conditions that denature other proteins.[28] Organic solvents can be removed after digestion by lyophilization, providing a more efficient "cleanup" step compared to removal of chemical denaturants, etc. Russell et al. have reported that addition of organic solvents can accelerate trypsin reactions for proteolysis-resistant proteins and reduce digestion times to less than 1 h.[19] These digestions, however, were analyzed using matrix-assisted laser desorption/ionization-MS, making the benefits of this digestion technique less clear for the LC−MS−MS analysis of low-microgram and -nanogram amounts of protein. Because addition of organic solvents reduces the potential for sample loss by allowing smaller volumes to be used, obviating chaotrope removal steps, accelerating reaction rates, and enabling protease-resistant proteins to be digested, they seemed an ideal choice for protocols designed to digest samples containing limited amounts of protein, such as those obtained from affinity isolations.

For this report, we systematically tested different solvent systems to optimize trypsin digestion of samples containing low-microgram to high-nanogram amounts of protein. The resulting digestions were analyzed by LC−MS−MS. MS data collected from digestions of samples containing 200 ng−10 μg of protein indicated that an organic−aqueous solvent system containing 80% acetonitrile consistently resulted in the most complete digestion, producing more peptide identifications than several other solvent systems. In addition, the shorter digestion times possible with addition of organic solvents resulted in smaller numbers of semitryptic peptides (peptides resulting from cleavage at one end at a residue other than lysine or arginine) than an overnight digestion protocol using an aqueous solvent. The use of an 80% acetonitrile solvent achieved both more numerous peptide identifications and reduced nonspecific cleavages. This system was then compared with an overnight digestion in aqueous buffer for digestions of protein complexes isolated from *R. palustris* by sucrose density gradient fractionation (the 70S ribosome) and by affinity purification (GroEL).

MATERIALS AND METHODS

Materials. All proteins, salts, buffers, dithiothreitol (DTT), guanidine hydrochloride, trifluoroacetic acid, diethyl pyrocarbonate, phenylmethanesulfonyl fluoride (PMSF), sucrose, and RNase-free DNase I were obtained from Sigma Chemical Co. (St. Louis, MO). RNase Away was obtained from Molecular BioProducts (San Diego, CA). Sequencing-grade trypsin was purchased from

(14) Larimer, F. W.; Chain, P.; Hauser, L.; Lamerdin, J.; Malfatti, S.; Do, L.; Land, M. L.; Pelletier, D. A.; Beatty, J. T.; Lang, A. S.; Tabita, F. R.; Gibson, J. L.; Hanson, T. E.; Bobst, C.; Torres, J. L.; Peres, C.; Harrison, F. H.; Gibson, J.; Harwood, C. S. *Nat. Biotechnol.* **2004**, *22*, 55−61.
(15) Rigaut, G.; Shevchenko, A.; Rutz, B.; Wilm, M.; Mann, M.; Seraphin, B. *Nat. Biotechnol.* **1999**, *17*, 1030−1032.
(16) Gould, K. L.; Ren, L.; Feoktistova, A. S.; Jennings, J. L.; Link, A. J. *Methods* **2004**, *33*, 239−244.
(17) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, 3rd ed.; W. H. Freeman and Co.: New York, 2000; pp 103−131.
(18) Park, Z. Y.; Russell, D. H. *Anal. Chem.* **2000**, *72*, 2667−2670.
(19) Russell, W. K.; Park, Z. Y.; Russell, D. H. *Anal. Chem.* **2001**, *73*, 2682−2685.
(20) Ekstrom, S.; Onnerfjord, P.; Nilsson, J.; Bengtsson, M.; Laurell, T.; Marko-Varga, G. *Anal. Chem.* **2000**, *72*, 286−293.
(21) Doucette, A.; Craft, D.; Li, L. *Anal. Chem.* **2000**, *72*, 3355−3362.
(22) Riviere, R.; Fleming, M.; Elicone, C.; Tempst, P. *Techniques in Protein Chemistry II*; Villafranca, J. J., Ed.; Academic Press: San Diego, CA, 1991.
(23) Ekstrom, S.; Malmstrom, J.; Wallman, L.; Lofgren, M.; Nilsson, J.; Laurell, T.; Marko-Varga, G. *Proteomics* **2002**, *2*, 413−421.
(24) Bosserhoff, A.; Wallach, J.; Frank, R. W. *J. Chromatogr.* **1989**, *473*, 71−77.
(25) Yu, Y. Q.; Gilar, M.; Gebler, J. C. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 711−715.
(26) Griebenow, K.; Kilbanov, A. M. *J. Am. Chem. Soc.* **1996**, *118*, 11695−11700.
(27) Welinder, K. G. *Anal. Biochem.* **1988**, *174*, 54−64.
(28) Simon, L.; Laszlo, K.; Vertesi, A.; Bagi, K.; Szajani, B. *J. Mol. Catal. B: Enzymol.* **1998**, *4*, 41−45.

Table 1. Digestion Solvents

	solvent/buffer[a]	incub. time[b]
1	50 mM Tris-HCl/10 mM $CaCl_2$	*c*
2	50 mM Tris-HCl/10 mM $CaCl_2$	1 h
3	60% acetonitrile/40% 50 mM Tris-HCl/10 mM $CaCl_2$	1 h
4	80% acetonitrile/20% 50 mM Tris-HCl/10 mM $CaCl_2$	1 h
5	60% methanol/40% 50 mM Tris-HCl/10 mM $CaCl_2$	1 h

[a] All buffers were pH 7.6. [b] All incubations were performed at 37 °C. [c] Overnight.

Promega (Madison, WI). Formic acid was obtained from EM Science (affiliate of Merck KgaA, Darmstadt, Germany). HPLC grade acetonitrile and water were used for all LC−MS analyses (Burdick & Jackson, Muskegon, MI). Ultrapure 18-MΩ water obtained from a Millipore Milli-Q system (Bedford, MA) was used for sample buffers. Fused-silica capillary tubing was purchased from Polymicro Technologies (Phoenix, AZ). BCA assay reagent and standards were obtained from Pierce Chemical Co. (Rockford, IL).

Construction of Protein Standard Mixture. Protein standard mixtures were generated using six proteins: bovine serum albumin (MW 69 kDa), yeast alcohol dehydrogenase I (MW 37 kDa), bovine carbonic anhydrase II (MW 29 kDa), horse myoglobin (MW 17 kDa), bovine hemoglobin (MW 15 kDa), and chicken egg lysozyme C (MW 14 kDa). Hemoglobin includes α- and β-polypeptides, and the isomer yeast alcohol dehydrogenase II was found to be a component of yeast alcohol dehydrogenase I, giving a total of eight polypeptides in the mixture. Mixtures contained equal masses of each protein. Each of the proteins was dissolved in 50 mM Tris-HCl/10 mM $CaCl_2$ (pH 7.6) and then combined in equal masses to give 10 μg, 2 μg, 1 μg, and 200 ng of total protein in final volumes of 100 μL of one of the digestion solvents described below.

Proteolytic Digestion of Protein Standard Mixture. The same amount of trypsin (200 ng) was added to each sample for digestion. For 10-μg, 2-μg, 1-μg, and 200-ng samples, the enzyme-to-substrate ratios (w/w) were, therefore, 1:50, 1:20, 1:5, and 1:1, respectively. The protein mixture digestions, performed in triplicate, differed also by the use of different solvent conditions (organic or aqueous) and incubation time for digestion. For all amounts of protein standard mixtures, trypsin digestions were performed in 100 μL of each solvent, using conditions listed in Table 1.

A control sample of 10 μg was digested using the manufacturer's protocol for the trypsin used in these experiments, that included denaturation in 6 M guanidine hydrochloride/50 mM Tris-HCl/10 mM $CaCl_2$ (pH 7.6) for 45 min followed by dilution to 0.5 M guanidine hydrochloride and overnight digestion with 200 ng of trypsin at 37 °C. The resultant peptides from this control digestion were desalted using solid-phase extraction (C_{18} Zip-Tip, Millipore, Billerica, MA).

After digestion, all peptide samples were treated with DTT (20 mM) for 1 h at 37 °C to reduce disulfide bonds. The reduced peptides were lyophilized and resuspended in 100 μL of 95% H_2O/5% acetonitrile/0.1% formic acid. To inhibit further trypsin activity, 2 μL of 10% formic acid was added to each resuspended sample. All samples were stored at −80 °C until analysis.

Isolation and Digestion of *R. palustris* Ribosomal Proteins. 70S ribosomes from *R. palustris* were purified and fractionated using a high-salt sucrose cushion and sucrose density fractionation as previously described.[29] Ribosomal protein extraction and the removal of contaminant rRNA was performed using the acid extraction method.[30] After overnight dialysis in a 3500 molecular weight cutoff dialysis cassette (Slide-A-Lyzer, Pierce) against Ultrapure water, the protein samples were concentrated and then quantitated using the BCA assay according to manufacturer's instructions (Pierce). Ribosomal protein samples were concentrated by solvent evaporation, reconstituted in 50 mM Tris-HCl/10 mM $CaCl_2$ (pH 7.6) and digested with 200 ng of trypsin in solvent systems 1 and 4 (Table 1). Each digestion was performed in duplicate.

Tandem Affinity Purification and Digestion of *R. palustris* GroEL2 Complex. The *R. palustris* wild-type strain (CGA009), harboring the pBBR5-DEST/42 modified Gateway expression plasmid (Invitrogen, Carlsbad, CA) with the GroEL2 open reading frame (ORF) was a generous gift from Dr. Dale Pelletier at the Oak Ridge National Laboratory. The ORF was cloned into the expression plasmid such that V5 and 6xHis affinity tags were fused at the C-terminus of the protein. The presence of both tags allowed the use of a dual-affinity purification strategy to "capture" the GroEL complex, using a strategy similar to tandem affinity purification.[31] The first affinity purification was a Ni−NTA capture in which the C-terminal 6xHis tag chelated the Ni−NTA resin (Qiagen, Valencia, CA). The second affinity purification was a capture using V5 resin (Sigma; anti-V5 antibody conjugated to agarose beads). This is our standard protocol for large-scale isolation of protein complexes from *R. palustris*, in which a large number of strains each bears a plasmid encoding a different affinity-tagged protein.[13]

R. palustris cells harboring the expression plasmid were grown anaerobically and harvested at mid-log phase (OD_{660} ~0.8).[32] Cell pellets were resuspended in NTA binding buffer (50 mM NaH_2PO_4 at pH 8, 300 mM NaCl, 10 mM imidazole, 5 mM ATP, 10 mM $MgCl_2$, 10 mM KCl, 100 μg/mL PMSF, 10 μg/mL leupeptin) and then sonicated with a Cell Disruptor 185 (Amphotech, Beverly, MA) using a series of six 15-s pulses separated by 30-s cooling intervals. Cellular debris was removed in an initial centrifugation at 4 °C using an SS-34 Sorval rotor at 12100*g* for 30 min. The supernatant was centrifuged for an additional 15 min at 23700*g*. The final resulting supernatant was then immediately used in the first stage of the affinity purification of GroEL.

(a) Ni−NTA Capture. After the addition of 100 μL of 50% Ni−NTA bead suspension (previously washed 4×) in NTA binding buffer, the supernatants were incubated on a rotator for 1 h at ambient temperature. The beads were then collected by centrifugation at 425*g*, transferred to new tubes, and washed 4 times with NTA wash buffer (50 mM NaH_2PO_4 at pH 8, 300 mM NaCl, 20 mM imidazole, 5 mM ATP, 10 mM $MgCl_2$, 10 mM KCl).

(29) Strader, M. B.; Verberkmoes, N. C.; Tabb, D. L.; Connelly, H. M.; Barton, J. W.; Bruce, B. D.; Pelletier, D. A.; Davison, B. H.; Hettich, R. L.; Larimer, F. W.; Hurst, G. B. *J. Proteome Res.* **2004**, *3*, 965−978.

(30) Hardy, S. J.; Kurland, C. G.; Voynow, P.; Mora, G. *Biochemistry* **1969**, *8*, 2897−2905.

(31) Puig, O.; Caspary, F.; Rigaut, G.; Rutz, B.; Bouveret, E.; Bragado-Nilsson, E.; Wilm, M.; Seraphin, B. *Methods* **2001**, *24*, 218−229.

(32) Geissler, J. F.; Harwood, C. S.; Gibson, J. *J. Bacteriol.* **1988**, *170*, 1709−1714.

Afterward, bound proteins were eluted from the Ni−NTA beads 4 times with 50 μL of NTA elution buffer (50 mM NaH_2PO_4 at pH 8, 300 mM NaCl, 500 mM imidazole, 5 mM ATP, 10 mM $MgCl_2$, 10 mM KCl). Combined eluents (~200 μL total) were diluted with 400 μL of buffer (5 mM ATP, 10 mM $MgCl_2$, 10 mM KCl) and either stored at −80 °C or immediately used for the second affinity purification step.

(b) V5 Capture. A 100-μL sample of a 50% V5 bead suspension (previously washed in PBS buffer) was added to the combined eluents from the Ni−NTA capture and incubated on a rotator for 1 h at ambient temperature. The beads were then centrifuged at 425*g* and washed 4 times with V5 wash buffer (50 mM Tris-HCl, 10 mM $CaCl_2$ at pH 7.6, 5 mM ATP, 10 mM $MgCl_2$, 10 mM KCl). Afterward, the bound proteins were eluted three times from the V5 beads with 50 μL of V5 elution buffer (80% acetonitrile, 20% 50 mM Tris-HCl, 10 mM $CaCl_2$ at pH 7.6). The combined eluents (~175 μL total) were digested with 200 ng of trypsin in solvent systems 1 and 4 (Table 1). Because of limited sample amounts, only one digestion of GroEL for each of the two solvent conditions was performed. After digestion, peptides were treated with DTT, lyophilized, and resuspended in 100 μL of 95% H_2O/5% acetonitrile/0.1% formic acid. Aliquots at each stage of the affinity purification were analyzed by Western blot using anti-V5 antibodies to determine the purification efficiency.

1D LC−MS−MS Analysis. For all peptide samples, one-dimensional LC−MS−MS experiments were performed with a Famos/Switchos/Ultimate HPLC System (Dionex, Sunnyvale, CA) coupled to an LCQ-DECA XP Plus quadrupole ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) equipped with a nanospray source as previously described.[33] For all 1D LC−MS−MS data acquisition, the LCQ was operated in the data-dependent mode with dynamic exclusion enabled (repeat count 2), where the four most abundant peaks in every MS scan were subjected to MS−MS analysis. Data-dependent LC−MS−MS was performed over a parent m/z range of 400−2000.

1D LC−Fourier Transform Ion Cyclotron Resonance (FTICR)-MS Analysis. 1D LC−FTICR-MS experiments were performed with an Ultimate HPLC system coupled with a HiRes-ESI Fourier transform ion cyclotron resonance mass spectrometer (IonSpec, Lake Forest, CA) equipped with a 9.4-T magnet (Cryomagnetics Inc., Oak Ridge, TN). Samples were separated with a Vydac 218MS5.30015 C18 column (300 μm i.d. × 15 cm, 300-Å pore size, 5-μm particles) at a flow rate of 4 μL/min and directly introduced to the FTICR MS with an electrospray source (Analytica, Branford, CT).

Protein Identification from MS Data Analysis. The SEQUEST algorithm was used to match experimental MS−MS spectra with their counterparts predicted from a protein sequence database.[34] An unconstrained database search was employed so that peptides resulting from cleavage at residues other than lysine or arginine at one end (semitryptic peptides) or both ends (nontryptic) could be identified. The sequence database used for searches in this article consisted of two major elements. The 4833 ORFs of the published *R. palustris* database[14] were search targets for the ribosome and GroEL searches but acted as distractors (indicators of false positive identifications) during the protein standard mixture searches. Sequences for the eight proteins in the standard mixture were also included in the database; we added the sequence for alcohol dehydrogenase II because this protein was observed as a component in the alcohol dehydrogenase I standard. The total count of proteins in the protein standard mixture database was 4841, yielding a 0.17% chance of randomly hitting one of the eight standard proteins.

DTASelect assembled, filtered, and compared the identifications from SEQUEST searches on all data sets. This software sorts peptide identifications by the proteins that contain them.[5] A protein in the mixture was considered successfully identified if at least two component peptides passed DTASelect's default SEQUEST score cutoffs. Spectra from singly charged peptides were required to exceed 1.8 in the SEQUEST parameter XCorr, while XCorr values for doubly and triply charged peptides were required to exceed 2.5 and 3.5, respectively.[5] The best matching sequence for each spectrum was required to have an XCorr at least 8% greater than the second best (DeltCN ≥ 0.08).

MS1PeakFinder Algorithm. Software created in the C++ programming language analyzed the mass spectra collected during each LC separation to catalog the observed ions. For each ion, the chromatographic profile was reconstructed from the intensities reported at its m/z through successive scans. Once this list of eluting ions was inferred, the tandem mass spectra were matched to the list, and those that were confidently identified by SEQUEST were flagged.

Scripts in the R statistical environment can be used to visualize these reports.[35] An image called a "matchmap" segregates the ions into three classes: ions observed only during MS scans, ions for which a tandem mass spectrum was collected but not successfully identified, and ions for which a tandem mass spectrum was both collected and confidently identified. These classes were colored yellow, orange, and red, respectively.

RESULTS AND DISCUSSION

The goal of this research was to optimize our digestion protocol through a systematic test of different solvent systems for dilute or limited protein samples, such as those typically obtained through affinity isolation of protein complexes. To this end, we compared digestion of 10-μg, 2-μg, 1-μg, and 200-ng amounts of a protein standard mixture in several solvent systems, including 100% aqueous buffer (50 mM Tris at pH 7.6, 10 mM $CaCl_2$) and the same aqueous buffer with different organic additives, including 60% methanol, 60% acetonitrile, and 80% acetonitrile. To compare digestion efficiencies, we used three criteria. First, we determined the total number of peptide identifications, including the number of semitryptic peptides, individual protein sequence coverage, and peptides per protein for each digested sample. Second, we analyzed the completeness of digestion by analyzing each LC−MS−MS run to identify the ion elution profiles for ions that were not confidently identified. Finally, we employed LC−FTICR-MS to identify incompletely digested ions eluting late in chromatographic runs. The two solvent systems that yielded the most efficient digestion of the protein standard mixture with respect to these criteria were then compared for digesting two biologically

(33) Tabb, D. L.; Narasimhan, C.; Strader, M. B.; Hettich, R. L. *Anal. Chem.* **2005**, *77*, 2464−2474.

(34) Eng, J. K.; McCormack, A. L.; Yates, J. R., 3rd. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.

(35) Hornik, K., The R FAQ, http://www.ci.tuwien.ac.at/~hornik/**R**/; ISBN 3-900051-08-9.

Table 2. Peptide Identifications for Different Quantities of the Protein Standard Mixture[a]

digestion solvent	total protein amount: 0.2 μg peptides identified	0.2 μg tryptic[b] (%)	1 μg peptides identified	1 μg tryptic[b] (%)	2 μg peptides identified	2 μg tryptic[b] (%)	10 μg peptides identified	10 μg tryptic[b] (%)
aqueous overnight	high = 105 94 low = 84	90	high = 130 114 low = 107	82	high = 148 146 low = 127	68	high = 138 137 low = 133	62
aqueous 1 h	high = 95 70 low = 69	91	high = 94 91 low = 87	93	high = 125 111 low = 86	81	high = 123 119 low = 112	70
60% acetonitrile	high = 95 80 low = 64	97	high = 113 100 low = 99	93	high = 165 134 low = 90	96	high = 117 115 low = 111	91
80% acetonitrile	high = 138 132 low = 119	97	high = 197 187 low = 169	97	high = 197 157 low = 157	96	high = 146 141 low = 140	91
60% methanol	high = 67 60 low = 55	97	high = 134 107 low = 90	96	high = 137 137 low = 129	97	high = 120 118 low = 115	91
overnight using chaotrope[c]	n/a[d]		n/a		n/a		high = 41 40 low = 40	68%

[a] Equal masses of each protein were combined to give the corresponding quantities. [b] Fully tryptic peptide percentages for highest value out of three replicates are shown. [c] Digestion protocol included using guanidine hydrochloride prior to digestion and reversed-phase extraction prior to sample injection (see Figure 1). [d] n/a, not analyzed.

relevant complexes from *R. palustris*: the 70S ribosome and GroEL. The 80% acetonitrile system provided results in 1 h that were comparable or superior to those obtained overnight using an aqueous buffer. Because times as short as 5 min have been reported for digestion of myoglobin in 80% acetonitrile,[19] we did not investigate digestion times longer than 1 h in buffers containing methanol or acetonitrile.

Protein Standard Mixture Digestions and LC−MS−MS Analysis. The protein standard mixture used in this study included eight proteins with different proteolytic susceptibilities, to emulate a protein complex isolated by affinity methods. Alcohol dehydrogenase I and II, carbonic anhydrase II, lysozyme C, and hemoglobin α/β chains are amenable to proteolysis, while globular proteins such as myoglobin, stabilized predominantly by a hydrophobic interior, and serum albumin, whose tertiary structure is stabilized by 17 disulfide bonds, are resistant to proteolytic digestion.[18,36]

Using 10-μg samples of the protein standard mixture, we compared a digestion protocol including the chaotrope guanidine hydrochloride and a desalting step against protocols that use no guanidine HCl (see Figure 1). Because sample losses can occur during the dilution and desalting processes, we expected that the number of peptide identifications would be lower for the protocol incorporating a chaotrope. This was indeed the case in our study; the 41 peptide identifications (highest number of three replicates) obtained by the denaturant method were lower than the peptide identification numbers obtained for the other digestion protocols, regardless of the amount of the digested sample (see Table 2).

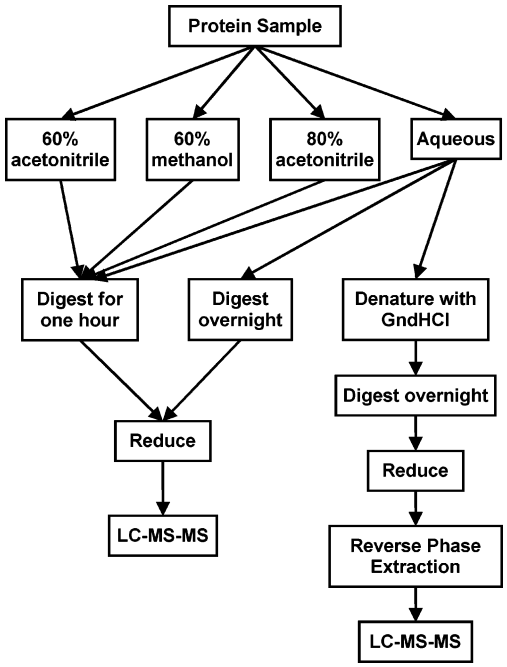

Figure 1. Flowchart representing the solvent systems and digestion protocols used in this study.

Because sample loss would be an even greater concern with lower sample amounts, the denaturant method was not applied to samples containing less than 10 μg of protein.

The 1-h digestion in 80% acetonitrile resulted in the highest number of identifications across the entire range of sample quantities. Table 2 shows the number of peptide identifications for three replicate LC−MS−MS analyses of each protein standard

(36) Slysz, G. W.; Schriemer, D. C. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1044−1050.

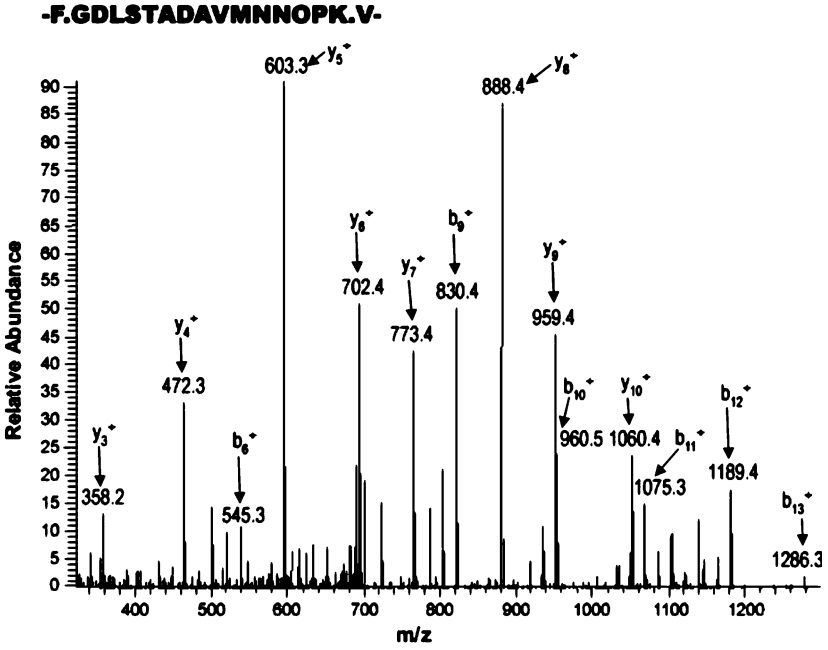

Figure 2. Example MS−MS spectrum identified as a semitryptic peptide. This spectrum represents the doubly charged F.GDLSTADAVMNPK.V, an internal peptide from hemoglobin's β chain. (The periods in the peptide sequence represent protease cleavage points; the actual peptide detected corresponds to the sequence between the periods.) SEQUEST yielded an XCorr of 4.6, an unlikely score for a false positive.

mixture sample. To compare solvent digestion efficiency, we used the value representing the highest number of peptides determined out of the three replicate separations. When 1 μg of protein was digested, the 80% acetonitrile solvent resulted in the identification of 52% more peptides than the overnight aqueous digest. The 200-ng and 10-μg quantities of proteins produced either similar or smaller numbers of peptide identifications than the 1- and 2-μg samples. In the least concentrated sample, digestions with fewer peptides probably resulted from low signal that is characteristic of smaller samples, while the most concentrated sample had a high ratio of substrate to enzyme, resulting in saturation.

A recent report by Olsen et al. suggested that MS data of tryptically digested samples should be analyzed for peptide identifications by configuring search algorithms for strict trypsin specificity.[37] In this configuration, potential peptides resulting from cleavages after residues other than lysine or arginine at one end (semitryptic peptides) or both ends (nontryptic) are not considered in database searches. Their results from in-gel trypsin digests of mouse liver proteins indicated that fully tryptic peptides were exclusively identified except for C-terminal peptides cut by trypsin at either lysine or arginine or semitryptic peptides resulting from in-source fragmentation products of fully tryptic peptides containing internal proline residues.[37] Our results, on the other hand, establish a link between digestion specificity and the solvent in which the digestion is performed; the proportion of fully tryptic peptides depends on the digestion solvent system used. Table 2 demonstrates that trypsin specificity was lowest for digestions performed overnight in the aqueous solvent; for example, 82% of the peptide identifications from the 1-μg overnight digestion in aqueous solvent were fully tryptic, compared to 97% from the 1-h digestion in 80% acetonitrile. From the overnight aqueous digest, the identified peptides included 107 fully tryptic, 23 semitryptic, and 0 nontryptic peptides. This trend was consistent for the other replicates and for the other protein concentrations (data not shown). Manual inspection of the 23 spectra matched to semitryptic peptides from the 1-μg overnight digestions, as exemplified by Figure 2, indicated that these peptides had been successfully identified by SEQUEST. Table 3 lists the 23 semitryptic peptides identified from the 1-μg overnight digests in aqueous solvent. Interestingly, 21 of the 23 had chymotryptic ends (tryptophan, tyrosine, phenylalanine, methionine, leucine, alanine, aspartic acid, and glutamic acid, at a terminal position of the peptide), suggesting that the L-(1-tosylamido-2-phenyl)ethyl chloromethyl ketone treatment of the commercially available trypsin may not have completely inhibited all contaminant chymotrypsin activity. Because the amount of active chymotrypsin is likely very small, the overnight digestion may have allowed sufficient time for chymotrypsin activity while the shorter digestion times possible with addition of organic solvents resulted in more specific proteolysis reactions.[38] This is supported by the fact that a 1-h digest in the same aqueous solvent produced lower proportions of semitryptic peptides than an overnight digestion. If the matches listed in Table 3 were identified to random sequences, there would be a greater probability of obtaining nontryptic identifications rather than semitryptic because a majority of possible sequences are nontryptic. Instead, semitryptic identifications were far more common than nontryptics; only one nontryptic was identified for all solvent systems at 1 μg of protein digested. Among the peptides listed in Table 3, five were identified in at least one of the other digestion conditions. In addition, several were identified with other charge states or were represented by more than one spectrum.

To further examine the effect of digestion conditions, we examined the sequence coverage and numbers of peptides

(37) Olsen, J. V.; Ong, S. E.; Mann, M. *Mol. Cell Proteomics* **2004**, *3*, 608−614.

(38) Simon, L.; Kotorman, M.; Garab, G.; Laczko, I. *Biochem. Biophys. Res. Commun.* **2001**, *280*, 1367−1371.

Table 3. Semitryptic Peptides from the Overnight Aqueous Digest of a 1-μg Mixture

semitryptic peptide	protein	charge state(s)	spectrum count	X-corr
K.VGGHAAEYGAE.A	bovine hemoglobin α	+1	1	2.4
L.SELSDLHAHK.L	bovine hemoglobin α	+2	1	2.6
A.SHLPSDFTPAVHASLDK.F	bovine hemoglobin α	+3	1	4.3
W.GKVEADIAGHGQEVLIR.L	horse myoglobin	+3	1	4.1
K.GHHEAELKPL.A	horse myoglobin	+3, +2	1, 1	3.6, 2.6
K.YLEFISDAIIHVL.H	horse myoglobin	+2	1	4.8
R.LLVVYPW.T	bovine hemoglobin β	+2	1	3.0
F.GDLSTADAVMNNPK.V	bovine hemoglobin β	+2	2	4.6
L.STADAVMNNPK.V	bovine hemoglobin β	+2	1	3.5
K.LLGNVLVVVL.A	bovine hemoglobin β	+2	2	2.8
K.VVAGVANAL.A	bovine hemoglobin β	+1	1	2.0
K.AVVQDPALKPL.A[a]	bovine carbonic anhydrase	+2	1	3.3
L.ALVYGEATSR.R	bovine carbonic anhydrase	+1, +2	1, 2	2.1, 3.6
W.IVLKEPISVSSQQMLK.F	bovine carbonic anhydrase	+3	1	3.8
R.TLNFNAEGEPELLML.A	bovine carbonic anhydrase	+1, +2	1, 2	2.9, 3.4
L.ANWRPAQPLK.N[a]	bovine carbonic anhydrase	+2	1	2.8
K.SANLMAGHWVAIS.G	yeast alcohol dehydrogenase I	+2	1	2.6
L.GIDGGEGKEELFR.S[a]	yeast alcohol dehydrogenase I	+2	1	3.3
N.GTTVLVGMPAGAK.C[a]	yeast alcohol dehydrogenase I	+2	2	3.8
L.STLPEIYEK.M[a]	yeast alcohol dehydrogenase I	+1	1	1.9

[a] Semitryptic peptides identified in multiple digestion data sets.

Table 4. Protein Sequence Coverage and Peptide Identifications of 1-μg Mixture

	aqueous overnight		aqueous 1 h		60% CH_3CN		80% CH_3CN		60% CH_3OH	
protein	% seq cov	no. of peptides	% seq cov	no. of peptides	% seq cov	no. of peptides	% seq cov	no. of peptides	% seq cov	no. of peptides
yeast alcohol dehydrogenase I	59	26	51	29	40	16	56	25	36	21
yeast alcohol dehydrogenase II	27	10	32	14	14	5	20	10	10	6
bovine serum albumin	23	16	0	0	19	9	70	61	25	17
horse myoglobin	75	24	50	11	70	25	94	30	98	27
bovine hemoglobin α	79	14	62	13	59	13	71	12	54	11
bovine hemoglobin β	73	22	72	17	77	19	62	20	76	23
Chicken lysozyme	38	5	22	2	86	20	86	17	95	16
bovine carbonic anhydrase	59	22	54	14	34	10	72	20	48	17

identified from each protein individually. Table 4 shows these statistics for the replicate with the highest number of peptide identifications for each solvent condition of the 1-μg mixture; the overall trends observed for these data were consistent for other replicates across the entire range of total protein quantities. While most of the proteins yielded similar sequence coverage in each of the digestion conditions, myoglobin, lysozyme, and serum albumin varied considerably in response to the conditions in which they were digested. In the case of the myoglobin's proteolytically resistant structure, an overnight digestion in aqueous solvent resulted in 75% sequence coverage, while a 1-h digestion in either 60% methanol or 80% acetonitrile resulted in nearly 100% sequence coverage. Lysozyme digested poorly in aqueous solvent but digested well in all three organic-containing solvents, most likely because these solvents improved its solubility. In fact, digestion of lysozyme in aqueous solvents was so poor that no identifications were obtained in the mixtures containing 10 and 2 μg of total protein. For serum albumin, a 1-h digestion in 80% acetonitrile resulted in 61 identified peptides, corresponding to 70% of the amino acid sequence. An overnight digestion of the same sample in aqueous buffer resulted in only 16 identified peptides, corresponding to 23% of the serum albumin sequence. Surprisingly, neither 60% acetonitrile nor 60% methanol improved the digestion efficiency of serum albumin substantially. Taken together, the improved sequence coverage for serum albumin suggests that the 80% acetonitrile solvent serves as an excellent denaturant, exposing a larger number of lysine and arginine residues for proteolysis.

Assessing Digestion Completeness by Visualization of LC−MS−MS Runs. We evaluated how well trypsin performed in each solvent by analyzing ion elution profiles from each LC−MS−MS run. Incomplete digestion should result in fewer small peptides and more numerous larger partially digested protein fragments. These large polypeptides can be expected generally to elute later in reversed-phase HPLC gradients due to their larger sizes and hydrophobicities. We visualized the ions observed during each separation using the MS1PeakFinder algorithm developed at ORNL to reveal the retention times and m/z ratios of identifiable and unknown ions for each sample. Figure 3 shows MS1PeakFinder plots for the overnight digestion in aqueous buffer and the 1-h digestion in 80% acetonitrile for the 1-μg protein standard mixture. These plots are not intended to provide intensity or chromatographic peak width information, but rather to indicate whether each parent ion was subjected to CID and whether the tandem mass spectrum was confidently identified by SEQUEST and DTASelect as a peptide. The overall trends observed in these two plots are representative of data across the entire range of protein quantities.

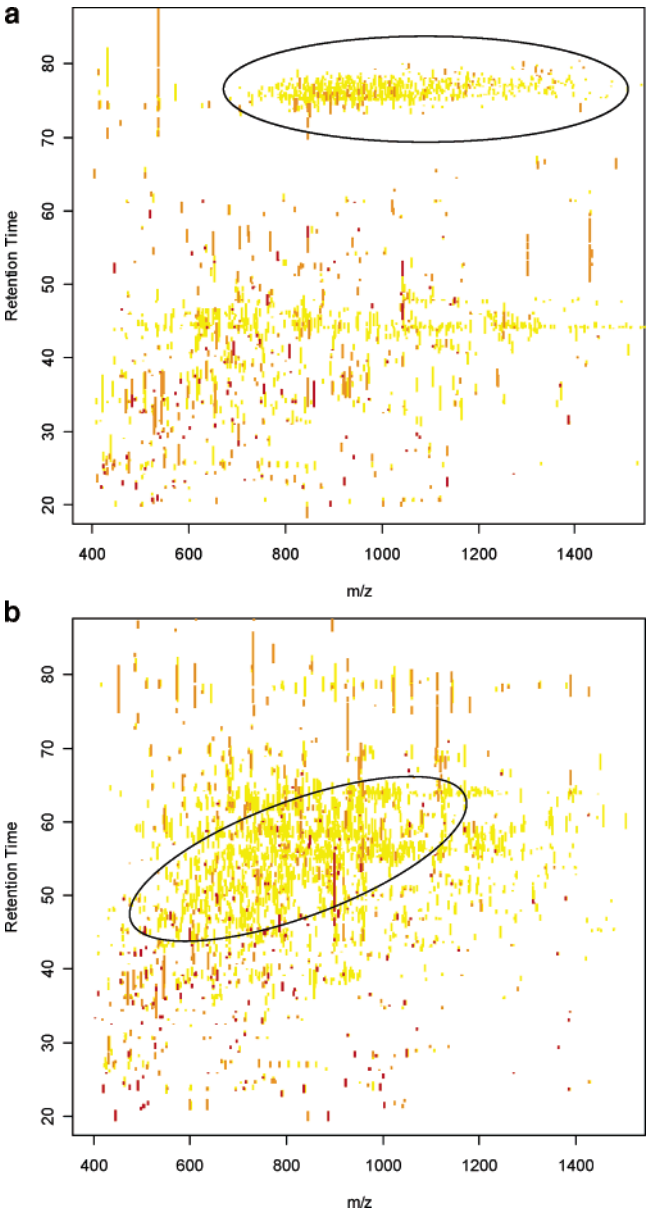

Figure 3. MS1PeakFinder plots for parent ions in LC−MS−MS data of the 1-μg mixture. (a) Overnight digestion in aqueous buffer; (b) 1-h digestion in 80% acetonitrile. The dense patch of ions resulting from incompletely digested proteins eluting between 77 and 82 min is circled in (a). The circled region in (b) indicates a more diffuse region at shorter retention times, corresponding to more complete digestion. Peaks for eluting peptides are red if a tandem mass spectrum resulted in a confident identification, orange if a tandem mass spectrum was acquired but no confident identification was obtained, and yellow if no tandem mass spectra were obtained for the ion.

For both plots, ions corresponding to confident spectral identifications (red peaks) eluted between 20 and 70 min. A dense patch of ions, eluting during a retention time interval of 77−82 min, was seen in the MS1PeakFinder plot representing the overnight digestion (Figure 3a). These ions correspond to peptides that were among the last to elute from the column. A majority of the ions seen in this dense patch did not result in confident identifications and may represent undigested or partially digested proteins that produced ions with more than three charges, preventing identification by SEQUEST as configured for this work. For the 1-h digestion, a less compact patch of species eluted during

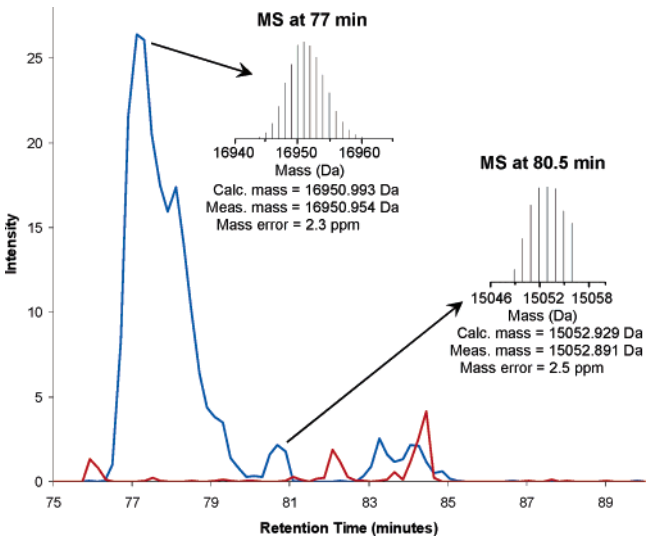

Figure 4. LC−FTICR-MS total ion chromatograms of the 1-μg mixture. The blue trace is from overnight digestion in aqueous buffer; the red trace is from a 1-h digestion in 80% acetonitrile. The insets illustrate the isotopic resolution of nominal masses representing undigested myoglobin and hemoglobin α chain. Peptides were separated using identical chromatographic gradients for both the LC−FTICR-MS and LC−MS−MS measurements shown in Figure 3.

retention times of 45−60 min (Figure 3b). While several red peaks were observed in the 45−60 min retention time interval, a majority of the ions detected during this interval were not confidently identified. Because species producing these unidentified ions eluted sooner than the 77−82 min window observed from the overnight digest, they most likely represent smaller proteolytic fragments that reflect a higher level of digestion. The number of different ions from species eluting during a retention time interval of 77−82 min appeared to decrease with an increasing number of peptide identifications reported in Table 2. For example, the 1-h aqueous digestion produced the fewest peptide identifications and contained 1072 ions detected between 75 and 82 min. On the other hand, only 58 ions were detected during the same retention times for the 80% acetonitrile digestion, for which the largest number of peptide identifications was observed. In fact, the majority of the ions in the 80% acetonitrile digestion were detected during the 20−70-min time interval.

Assessing Digestion Quality by FTICR. To identify species corresponding to the dense patch in Figure 3a eluting between 77 and 82 min, we analyzed protein standard mixtures digested overnight in aqueous buffer or digested for 1 h in 80% acetonitrile using 1D LC−FTICR-mass spectrometry. FTICR allows isotopic resolution of multiply charged ions, allowing the determination of molecular mass to better than 10 ppm accuracy.[39] By knowing the accurate molecular masses, we were able to determine that species eluting during the time interval corresponding to the dense patch of unidentified ions in the LC−MS−MS results shown in Figure 3a represented intact undigested proteins from the mixture and partially digested protein fragments with charge states too high to be identified by SEQUEST as configured for this work. Figure 4 shows the LC−FTICR total ion chromatograms of the mixture digested overnight in aqueous buffer and digested for 1 h in 80% acetonitrile. The C18 reversed-phase column, LC

(39) Bogdanov, B.; Smith, R. D. *Mass Spectrom. Rev.* **2005**, *24*, 168−200.

Table 5. Peptide and Protein Identifications for *R. Palustris* 70S Ribosomal Proteins

	ribosomal proteins					
	0.2 μg		1 μg		GroEL[a]	
digestion solvent	peptide IDs	ribosomal protein IDs	peptide IDs	ribosomal protein IDs	peptide IDs	protein IDs
aqueous[b]	22	9	102	31	57	2
80% acetonitrile[c]	36	13	144	39	54	3

[a] We identified both gene products, GroEL1 and GroEL2, from each digestion method. The third identified protein from the GroEL digestion in 80% acetonitrile was apolipoprotein *N*-acyltransferase, with two peptides identified (4.3% sequence coverage). [b] Overnight digestion in 50 mM Tris buffer (pH 7.6)/10 mM $CaCl_2$. [c] One-hour digestion in 80% acetonitrile.

instrumentation, and gradients used for these experiments were identical to those used for the LC−MS−MS analyses described above, which were performed using the quadrupole ion trap mass spectrometer for detection. As illustrated in the insets in Figure 4, intact myoglobin (most abundant isotope mass [MAIM][29] measured 16 950.954 Da, calculated MAIM 16 950.993 Da, mass error 2.3 ppm) and hemoglobin α chain (measured MAIM 15,052.891 Da, calculated MAIM 15 052.929 Da, mass error 2.5 ppm) were identified in the 75−82-min retention time window for the overnight aqueous digestion. In addition, several unidentified species, probably partially digested protein fragments, in the 27−29 kDa range were observed (data not shown). We did not identify any intact proteins in the 75−82-min retention time window for the 80% acetonitrile digestion shown in Figure 4, and consistent with the LC−MS−MS results, overall ion signal was lower in this retention time window. These data collectively show that trypsin cuts more efficiently in the 80% acetonitrile digestion.

Analysis of the 70S Ribosome and GroEL Complexes from *R. palustris*. One of our major goals is to develop an efficient protocol for high-throughput MS analysis of protein complexes from microbes.[13,40] Currently, we use dual-affinity purification to isolate protein complexes, a method that results in variable amounts of protein for each purified complex. Because limited sample amounts can complicate enzymatic digestion and protein identifications, we have focused on tailoring digestion protocols for samples, obtained through dual-affinity purification, that are destined for MS analysis. The intent of this report is to describe a protocol for protein digestion that maximizes the number of protein identifications from MS. Our results indicate that the 80% acetonitrile solvent system was most suitable for digesting a protein standard mixture under conditions similar to those imposed by our overall workflow. Biological samples, however, are often significantly more complex than our model mixture, containing a large number of proteins at different concentrations, with different proteolytic susceptibilities and solubilities, and varying amounts and types of impurities. With this in mind, we chose the 70S ribosome (purified by sucrose density fractionation)[29] and dual-affinity purified GroEL from *R. palustris* as representative protein complexes to test the effectiveness of the 80% acetonitrile digestion protocol for biologically relevant samples. As a control, we compared the 80% acetonitrile digestions with our standard protocol for affinity isolation samples, which is overnight digestion in aqueous buffer.

Because the 70S ribosome is composed of over 50 proteins, it contains a wider variety of subunits than most complexes that would be obtained from affinity isolation procedures and thus offers a significant challenge for our approach. Table 5 shows the peptide and protein identifications obtained from digesting 1 μg and 200 ng of the *R. palustris* 70S ribosome. For both sample amounts, the 80% acetonitrile digestions gave more peptide and protein identifications than the overnight aqueous digestions. For example, 39 proteins were identified for a 1-h digestion of 1 μg of ribosomes compared to 30 proteins for the overnight digestion in aqueous solvent. The identifications obtained for the 1-h digestion corresponded to 72% of the proteins comprising the 70S ribosome. In contrast to the protein standard mixtures, no semitryptic identifications were obtained for either digestion of the ribosome samples; only fully tryptic peptides were identified. Although these samples contained 6−7 times as many protein sequences as the protein standard mixture, the 80% acetonitrile digest continued to outperform the overnight aqueous digest. The lower numbers of peptide and protein identifications observed for the smaller 200-ng sample size suggest that other aspects of our workflow, such as the LC−MS−MS analysis, are limiting factors for sensitivity as sample sizes decrease. Indeed, a more comprehensive characterization of the ribosomal proteins requires larger amounts of protein (and replicate analyses) than used in the current work.[29] However, the increased peptide and protein identifications obtained here by using the 80% acetonitrile digestion solvent for both the 1-μg and 200-ng sample sizes suggest that this solvent represents a significant step toward the use of lower amounts of protein for characterization of ribosomes or other large protein complexes.

Table 5 also shows the peptide and protein identifications for dual-affinity purified GroEL. In the two digestions, we identified both gene products of GroEL genes expressed by *R. palustris* (GroEL1 and GroEL2). These data indicate that a 1-h digest in 80% acetonitrile produced results comparable to the overnight aqueous digests. These results suggest that GroEL may be fairly amenable to proteolysis and therefore the potential advantages of the organic solvent were not needed to maximize peptide yields. The overnight aqueous digest, however, yielded many semitryptic peptides; 40% of the 51 peptide identifications were of this class. In contrast, 25% of the identifications were semitryptic for the 1-h digestion with acetonitrile. The identification of semitryptic peptides from the digestion in acetonitrile may have occurred due to endogenous proteases from *R. palustris*, as this sample would have contained the highest complement of impurities of the systems studied. While the results for GroEL do not show an increase in

(40) Buchanan, M. V.; Larimer, F. W.; Wiley, H. S.; Kennel, S. J.; Squier, T. J.; Ramsey, J. M.; Rodland, K. D.; Hurst, G. B.; Smith, R. D.; Xu, Y.; Dixon, D.; Doktycz, M. J.; Colson, S.; Gesteland, R.; Giometti, C.; Young, M.; Giddings, M. *Omics* **2002**, *6*, 287−303.

the peptides identifiable for these proteins by use of organic solvents, other biological complexes may contain components that are more proteolytically resistant and that require the organic additive to serve as a denaturant.

CONCLUSION

Our results demonstrate that effective trypsin digestions of high-nanogram to microgram amounts of proteins can be performed in 1 h using 80% acetonitrile as the solvent. This protocol does not require the chemical denaturing steps necessary for protein digestion methods that require chaotropes or surfactants, followed by dilution or purification steps that can collectively lead to sample loss and diminished sensitivity. As the data from the dual-affinity purified GroEL attest, however, the use of acetonitrile in rapid digestions may not improve peptide recovery relative to traditional overnight digests when the target proteins are amenable to digestion. The benefits appear to be most pronounced in protease-resistant proteins such as bovine serum albumin and chicken lysozyme. Large protein complexes, such as the ribosome, still require tens of micrograms of material for complete characterization, although we have presented evidence that digestion in 80% acetonitrile allows detection of more ribosomal proteins than the aqueous digestion.

A further benefit of employing 80% acetonitrile in digestion buffers is the reduction of nonspecific cleavage. Because overnight aqueous digestions can result in increased numbers of semitryptic peptides, as observed for the protein standard mixture and GroEL samples, database searching with no protease specificity becomes necessary to maximize peptide identification. These searches, however, generate larger numbers of candidate sequences, take more time to run, and yield higher false positive rates. The use of 80% acetonitrile can produce substantially higher proportions of fully tryptic peptides than overnight digestions. As a result, these spectra can be identified by more efficient searches that consider only tryptic peptides.

This digestion method opens new opportunities to efficiently detect and characterize proteins in amounts typically obtained from affinity isolation of biological complexes. First, the ability to minimize sample loss and improve the number of peptide identifications will ultimately allow more sensitive detection of less abundant components from biological complexes. Second, the 1-h incubation time possible with this method is a distinct improvement for high-throughput applications. The results presented here suggest further investigation of at least two aspects of protocols for digestion in organic-containing solvents. Some proteins are likely to precipitate in solutions with high organic content; in fact, organic precipitation is an accepted method for purifying proteins. The ability of the protocols presented here for detecting such organic-insoluble proteins should be explored. Also, while we did not test the use of 80% acetonitrile to digest larger sample amounts ($\geq$20 μg), it is likely that these digestion conditions will be applicable to more abundant samples simply by increasing the volume of solvent. This study, therefore, presents a digestion method that potentially can be adapted to a variety of different types of samples.

ACKNOWLEDGMENT

We thank Caroline S. Harwood, Department of Microbiology, University of Iowa, for generously contributing the *R. palustris* strain used in this study. We thank the John Yates laboratory at The Scripps Research Institute for DTASelect and Contrast software. W.J.H. and C.P. acknowledge support from the University of Tennessee—Oak Ridge National Laboratory Graduate School of Genome Science and Technology. We thank William Hayes McDonald for technical advice and critical evaluation of the manuscript. This research was funded by the U.S. DOE Office of Biological and Environmental Sciences Genomes:GTL Program. Oak Ridge National Laboratory is operated and managed by the University of Tennessee—Battelle, L.L.C. for the U.S. Department of Energy under Contract DE-AC05-00OR22725.

Received for review July 28, 2005. Accepted October 21, 2005.

AC051348L