research articles Journal of proteome research

Comparison of Digestion Protocols for Microgram Quantities of Enriched Protein Samples

W. Judson Hervey, IV,[†,‡] Michael Brad Strader,[†,§] and Gregory B. Hurst*[,†]

Organic and Biological Mass Spectrometry Group, Chemical Sciences Division, Oak Ridge National Laboratory, P. O. Box 2008, Oak Ridge, Tennessee 37831-6131, and Graduate School of Genome Science and Technology, University of Tennessee-Oak Ridge National Laboratory, 1060 Commerce Park Drive, Oak Ridge, Tennessee 37830-8026

Received March 21, 2007

Standard biochemical techniques that are used for protein enrichments, such as affinity isolation and density gradient centrifugation, frequently yield high-nanogram to low-microgram quantities at a significant expenditure of resources and time. The characterization of selected protein enrichments by the "shotgun" mass spectrometry approach is often compromised by the lack of effective and efficient in-solution proteolysis protocols specifically tailored for these small quantities of proteins. This study compares the results of five different digestion protocols that were applied to 2.5 μg portions of protein isolates from two disparate sources: *Rhodopseudomonas palustris* 70*S* ribosomal proteins, and *Bos taurus* microtubule-associated proteins (MAPs). Proteolytic peptides produced according to each protocol in each type of protein isolate were analyzed by one-dimensional liquid chromatography–tandem mass spectrometry (LC–MS/MS). The effectiveness of each digestion protocol was assessed on the basis of three parameters: number of peptide identifications, number of protein identifications, and sequence coverage. The two protocols using a solvent containing 80% acetonitrile (CH_3CN) for trypsin digestions performed as well as, and in some instances better than, protocols employing other solvents and chaotropes in both types of protein isolates. A primary advantage of the 80% CH_3CN protocol is that it requires fewer sample manipulation steps.

Keywords: trypsin • digestion • proteolysis • acetonitrile • mass spectrometry • liquid chromatography • ribosome • microtubule-associated protein

Introduction

The analysis of protein mixtures by mass spectrometry-based techniques[1] has become possible due to a confluence of advancements in analytical instrumentation, multidimensional chromatographic separations, and bioinformatics. One widely used proteomics strategy is to perform qualitative analysis of proteolytic peptides by one-dimensional (1D) or two-dimensional (2D) chromatographic separation coupled via electrospray with tandem mass spectrometry (LC–MS/MS), generating peptide fragmentation spectra that are compared to translated genomic databases.[1,2] This robust approach has been applied, for example, to identifying proteins from complex mixtures,[3] less abundant samples obtained from biochemically enriched multiprotein complexes,[4,5] and affinity isolations of single proteins along with interacting partners.[6–11]

The LC–MS/MS analysis of a protein sample from a biochemical enrichment or affinity isolation presents unique challenges to the workflow of "bottom-up" LC–MS/MS proteomics approaches. The limitation imposed by the small amounts of protein typically available from the biological enrichment requires that all aspects of downstream sample processing be as efficient and effective as possible, including enzymatic proteolysis. Improvements in enzyme activity have been reported when enzymatic reactions are conducted in organic media,[12,13] and more significantly, with trypsin directly in reverse-phase HPLC solvents.[14,15] Recently, our laboratory has demonstrated the effectiveness of conducting trypsin digestions of protein isolates in a solvent containing 80% acetonitrile (CH_3CN), resulting in an increased number of peptide and protein identifications, increased sequence coverage, and more thorough proteolysis for low-microgram to high-nanogram quantities of standard protein mixtures, ribosomes purified by density gradient centrifugation, and an affinity-isolated protein complex.[16]

In the present report, we compare the effectiveness of two digestion protocols performed in 80% CH_3CN[16] with three other protocols commonly used for the digestion of proteins. These protocols include a modified version of the manufacturer's

* Author for correspondence. E-mail, hurstgb@ornl.gov; phone, (865)574-6142; fax, (865)576-8559.

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

‡ Graduate School of Genome Science and Technology, University of Tennessee-Oak Ridge National Laboratory.

§ Present address: Laboratory of Neurotoxicology, National Institute of Mental Health, 10 Center Drive, Bethesda, MD 20892-1262.

Published on Web 07/06/2007

10.1021/pr070159b CCC: $37.00 © 2007 American Chemical Society

instructions from the sequencing-grade modified trypsin,[17,18] a sequential digestion with Lys-C and trypsin,[19,20] and a protocol utilizing the acid-labile surfactant (ALS) RapiGest SF.[21,22] The enzymatic proteolysis of 2.5 μg of protein, a yield representative of biochemical protein enrichment or affinity isolation experiments, was performed on two distinct types of protein isolates: the 70*S* ribosome of *Rhodopseudomonas palustris*,[4] and microtubule-associated proteins (MAPs)[19,23] of the cow, *Bos taurus*, using the digestion protocols described above. The two protocols employing the 80% CH_3CN solvent for trypsin digestions performed as well as or better than the other protocols tested in this study. In as little as 1 h, trypsin digestions in 80% CH_3CN yielded either comparable or more numerous tryptic peptide identifications and increased sequence coverage in both the 70*S* ribosomal isolates and the MAPs.

Experimental Procedures

Reagents obtained from Sigma Chemical Company (St. Louis, MO), were Trizma Hydrochloride, Trizma Base, iodoacetamide (IAA, Ultrapure), hydrochloric acid (molecular biology grade), ammonium bicarbonate (99% purity), and urea (Ultrapure). Other materials used in this study included lyophilized sequencing-grade modified trypsin (Promega, Madison, WI), endoproteinase Lys-C (Roche Applied Science, Penzberg, Germany), formic acid (EM Science SupraPur), calcium chloride (J. T. Baker, 99% purity), HPLC-grade acetonitrile and water (Burdick and Jackson, Muskegon, MI), tris(2-carboxyethyl)-phosphine hydrochloride (TCEP), and micro bicinchoninic acid (BCA) assay reagent (Pierce, Rockford, IL). RapiGest SF was obtained from Waters (Milford, MA). Ultrapure 18 MΩ water used for dialysis of the protein samples was obtained from a Milli-Q system (Millipore, Bedford, MA). HPLC-grade water was used for all sample buffers. Nonstick microcentrifuge tubes were obtained from VWR International (West Chester, PA).

Preparation of Protein Isolates. The 70*S* ribosomal protein complex was enriched from *R. palustris* CGA010 by sucrose density fractionation, as described previously.[4] Microtubule-associated proteins were a gift from the Mitchison laboratory (Department of Systems Biology, Harvard Medical School).[19] Protein isolates were dialyzed for 16 h against Ultrapure water in a 3500 MWCO Slide-A-Lyzer dialysis cassette (Pierce, Rockford, IL), quantified using the micro BCA assay (Pierce), aliquotted into 2.5 μg portions, and concentrated to a volume of approximately 10 μL by centrifugal evaporation in a Speed-Vac (Savant).

Protein Digestion by Enzymatic Proteolysis. Digestions of 2.5 μg portions of 70*S* ribosomal isolates and 2.5 μg portions of MAPs were performed in duplicate according to five different protocols (Figure 1 and Table 1). For all trypsin digestions, trypsin was reconstituted at a concentration of 20 ng/μL in 50 mM Trizma-HCl and 10 mM $CaCl_2$ at pH 7.6 and added to each protein isolate at a trypsin-to-protein ratio of 1:10 (w/w) in a nonstick microcentrifuge tube. Following each trypsin digestion, peptides were concentrated by centrifugal evaporation (SpeedVac) to a volume of ~30 μL, diluted in HPLC solvent A (95% H_2O, 5% CH_3CN, and 0.1% formic acid) to a volume of 123 μL, and stored at −80 °C until LC−MS/MS analysis. Duplicate LC−MS/MS analyses of each duplicate digestion were performed, providing a total of 4 data sets for each combination of sample type and digestion protocol.

Guanidine Hydrochloride Denaturation and 16 h Trypsin Digestion (*Guan-HCl*).[17,18] Protein isolates were suspended in 90 μL of 6 M guanidine hydrochloride, dissolved in 50 mM Trizma-HCl, pH 7.6, heated at 60 °C for 1 h, and cooled to laboratory temperature (approximately 25 °C). Disulfide bonds were reduced with the addition of TCEP to a concentration of 2 mM and incubated at ambient laboratory temperature for 20 min. Cysteine residues were carboxyamidomethylated with the addition of IAA to a concentration of 10 mM and incubated at ambient laboratory temperature for 15 min in the dark. The concentration of guanidine hydrochloride was diluted to 1 M with the addition of 50 mM Trizma-HCl and 10 mM $CaCl_2$ at pH 7.6. Proteolysis was conducted with trypsin for 16 h at 37 °C. Following digestion, samples were concentrated by centrifugal evaporation to a total volume of 100 μL. Solid-phase extraction (SPE) was performed using 100 μL C18 OMIX Tips (Varian), and extracted peptides were concentrated by centrifugal evaporation and dilution in Solvent A as detailed above.

Urea Denaturation, 16 h Lys-C Digestion, and 16 h Trypsin Digestion (*Lys-C/Trypsin*).[19,20] Protein isolates were suspended in 35 μL of 8 M urea dissolved in 100 mM Trizma-HCl, pH 8.5. Disulfide bonds were reduced and cysteine residues alkylated as described in the *Guan-HCl* protocol. In the first proteolysis reaction, 100 ng of Lys-C was added to each protein isolate and incubated at 37 °C for 16 h. Following proteolysis with Lys-C, the urea concentration was diluted to 2 M urea using 100 mM Trizma-HCl at pH 8.5, and proteolysis with trypsin was conducted at 37 °C for 16 h.

Organic Solvent and 1 h Trypsin Digestion (*CH_3CN 1 h*).[16] Protein isolates were suspended in 80% CH_3CN, 20% 50 mM Trizma-HCl, and 10 mM $CaCl_2$ at pH 7.6 to a volume of 100 μL. Disulfide bonds were reduced and cysteine residues alkylated as described in the *Guan-HCl* protocol. Proteolytic digestion was performed with trypsin at 37 °C for 1 h.

Organic Solvent and 16 h Trypsin Digestion (*CH_3CN 16 h*).[16] All steps were performed identically to the *CH_3CN 1 h* protocol except proteolysis, which was extended to 16 h at 37 °C.

Acid-labile Surfactant and 1 h Trypsin Digestion (*RapiGest*).[21,22] Protein isolates were suspended in 90 μL of 0.1% (w/v) RapiGest SF (Waters) and 50 mM NH_4HCO_3 and equilibrated at 37 °C for 2 min. Disulfide bonds were reduced and cysteine residues alkylated as described in the *Guan-HCl* protocol. Proteolytic digestion was performed with trypsin at 37 °C for 1 h. Protein isolates were acidified by addition of 20 μL of 500 mM HCl, incubated at 37 °C for 45 min, and centrifuged at 13 000*g* for 10 min to precipitate the RapiGest SF. The supernatant peptides were removed and concentrated as described above.

Liquid Chromatography and Tandem Mass Spectrometry (LC−MS/MS). Mass spectrometry data were collected using an LCQ Deca XP Plus quadrupole ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) equipped with a nanospray source, coupled with a one-dimensional, reverse-phase HPLC system (LCPackings/Dionex), as described previously.[16] Briefly, 50 μL of each digestion, representing approximately 1.2 μg of digested protein isolate, was loaded onto the reverse-phase HPLC system by a Famos autosampler (LCPackings/Dionex), concentrated on a C18 reverse-phase trapping cartridge (300 μm × 5 mm, 300 Å PepMap, LCPackings/Dionex), and desalted for 10 min with solvent A at a flow rate of 30 μL/min, pumped by a Switchos II unit (LCPackings/Dionex). By means of a switching valve on the Switchos, a lower solvent flow (see below) was then directed in the opposite direction across the trapping cartridge, backflushing the trapped peptides onto a C18 reverse-phase analytical column. The analytical column consisted of an 18 cm length of Jupiter C18 reverse-phase resin with 5 μm diameter particles and 300 Å pore size (Phenomenex)

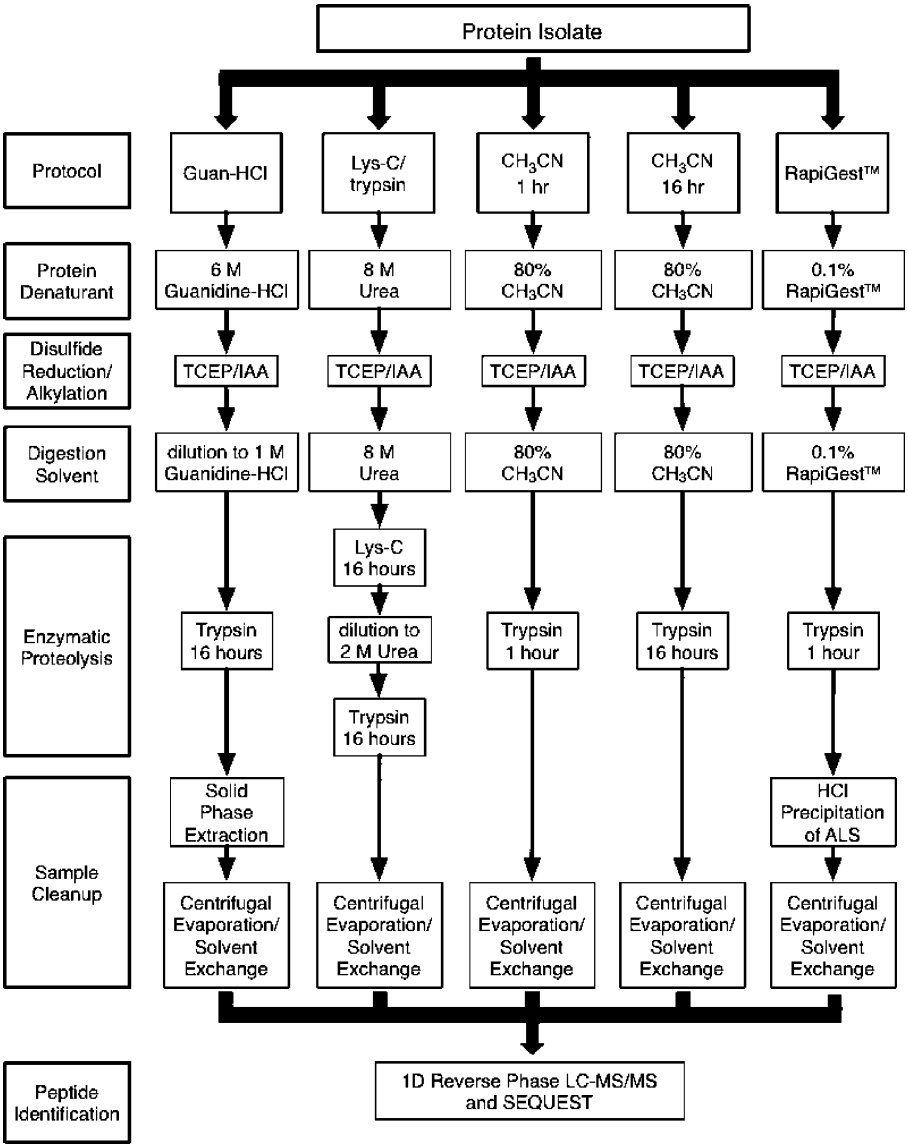

Figure 1. Flowchart representing the digestion protocols compared in this study.

Table 1. Solvents and Incubation Times for the Digestion Protocols

protocol designation	solvent/buffer for trypsin digestion	incubation time at 37 °C
Guan-HCl	1 M guanidine hydrochloride, 50 mM Trizma-HCl, 10 mM $CaCl_2$ pH 7.6	16 h
Lys-C/trypsin	2 M urea, 100 mM Trizma-HCl pH 8.5	32 h
CH_3CN 1 h	80% CH_3CN, 20% 50 mM Trizma-HCl, 10 mM $CaCl_2$ pH 7.6 (v/v)	1 h
CH_3CN 16 h	80% CH_3CN, 20% 50 mM Trizma-HCl, 10 mM $CaCl_2$ pH 7.6 (v/v)	16 h
RapiGest	0.1% RapiGest SF, 50 mM NH_4HCO_3 (w/v)	1 h

packed into a 100 μm i.d. PicoTip Emitter (New Objective) using a pressure cell (New Objective). Peptides were eluted from the reverse-phase column by an Ultimate HPLC pump (LCPackings/Dionex) at a total flow rate of 150 μL/min, with flow splitting provided by a NAN-75 calibrator, for a column flow rate of ~250 nL/min. The gradient elution consisted of a linear ramp over 130 min from 100% solvent A to 50% solvent A/50% solvent B (30% H_2O, 70% CH_3CN, and 0.1% formic acid), followed by a 1 min ramp to, and 19 min hold at, 100% solvent C (5% H_2O, 95% CH_3CN, and 0.1% formic acid). The mass spectrometer was operated in a data-dependent mode, in which the four most intense ions in the precursor ion scan within the mass-to-charge range of 400−2000 Da were subjected to tandem mass spectrometry. Dynamic exclusion[24] was enabled on the instrument with the repeat count set at two, as described previously.[4,16]

Peptide Identification and Data Analysis. Identification of peptides from tandem mass spectra was performed using the SEQUEST algorithm[25] without enzyme specificity. Databases used in the SEQUEST analysis were the predicted proteomes of *R. palustris*[26,27] and *B. taurus*[28] for ribosomal and MAP protein isolates, respectively; FASTA protein sequences for *R. palustris* were obtained from the ORNL microbial genome annotation pipeline,[29] and the protein sequences for *B. taurus* (IPI Cow

3.09)[28] were obtained from the European Bioinformatics Institute (EBI). FASTA protein sequences of common contaminants (e.g., keratins), protein standard mixture components (e.g., myoglobin, hemoglobin), and proteolytic enzymes (trypsin, lys-c) were concatenated to the predicted proteomes of each organism, resulting in a total of 45 843 proteins for *B. taurus* and 4890 proteins for *R. palustris*. A dynamic modification of 57.0 Da was specified for carboxyamidomethylation of cysteine residues.

DTASelect (version 2.06)[30] filtered peptide identifications from each search results file[31] according to the default cross-correlation (XCorr) scores (≥1.8, ≥2.5, and ≥3.5 for the 1+, 2+, and 3+ charge state, respectively) and ΔCn ≥ 0.08. Peptide identifications above these score thresholds were assembled by locus into protein identifications, requiring a minimum of two peptide identifications per protein for a positive protein identification. Redundant locus information was discarded by DTASelect. Peptide and protein identification data of replicate chromatographic runs were exported from DTASelect with the organizer.pl Perl script,[32] modified to use the thresholds quoted above, and organized for further analysis in a relational database (Microsoft Access).

Results and Discussion

The characterization of protein enrichments by mass spectrometric identification of peptides depends critically on an effective and efficient proteolysis protocol that is specifically tailored for low-microgram to high-nanogram quantities of protein. This report compares the effectiveness of several protocols for the digestion of protein isolates of this nature. To this end, 2.5 μg portions of two isolates, the *R. palustris* 70*S* ribosomal protein complex and *B. taurus* microtubule-associated proteins (MAPs), were digested according to five different digestion protocols, shown in Figure 1. The five digestion protocols that we studied encompass a variety of protein denaturation methods, digestion solvents (see Table 1), proteolytic enzymes, and incubation periods. These parameters were not comprehensively varied, as described in previous studies;[33] rather, we sought to implement the digestion protocols as described in technical or manufacturer's literature directly. The comparison among the various protocols is based on numbers of identified peptides and proteins, and sequence coverage for proteins; sequence coverage is the fraction of residues in the protein that occur in peptides that have been identified in LC−MS/MS experiments.

The two sets of proteins chosen for this study are distinct in their biological origin, sample complexity, and physical characteristics. First, the 70*S* bacterial ribosome is isolated from a prokaryote, *R. palustris*, whereas MAP isolates are from the cow, a eukaryotic organism. Figure 2 illustrates the differences between the two sets of proteins with respect to sample complexity and the physical characteristics of molecular weight and isoelectric point (p*I*). The more complex MAP isolates cover a wider molecular weight range than ribosomal proteins; isoforms of the lighter tau protein are between 55 and 62 kDa, while the larger MAPs range from 180 to 350 kDa. Additionally, MAP p*I*s range from <5 to >10. Most ribosomal proteins are of low molecular weight, and, with the exception of a few acidic proteins, have high p*I*s, as Figure 2 shows. As the 70*S* ribosome functions within the cytoplasm of the prokaryotic cell, the protein component of the complex is soluble at physiological pH. Thus, the different characteristics of these two protein samples provided a diverse test set for comparing the various digestion protocols.

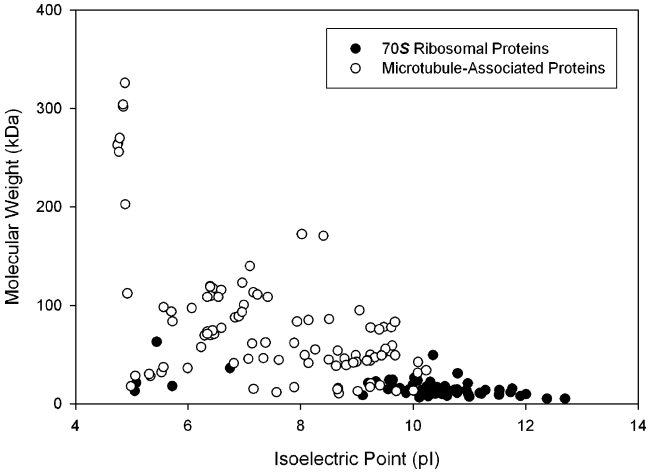

Figure 2. Physical characteristics of proteins annotated as *R. palustris* ribosomal proteins or *B. taurus* MAPs.

***R. palustris* 70*S* Ribosomal Isolates.** The ribosome is a ribonucleic acid−protein complex that functions in the synthesis of proteins. This large macromolecular protein complex is readily sedimented away from other cellular components by sucrose density fractionation. With an estimated 20 000 copies per cell,[34] the 70*S* ribosome of eubacteria is amenable to LC−MS/MS-based approaches due to its intracellular abundance, relative ease of isolation by existing biochemical techniques, and biological significance.[4]

Table 2 shows the average number of peptide identifications from digestions of ribosomal protein isolates. The *CH_3CN 1 h* protocol produced both the most total peptide identifications and the most tryptic peptide identifications of any protocol tested, followed by the *CH_3CN 16 h*, *Guan-HCl*, *RapiGest*, and *Lys-C/trypsin* protocols, in that order. Extension of the digestion in 80% CH_3CN to 16 h resulted in a number of peptide identifications similar to the 1 h digestion. In accord with other studies,[20] the extension of digestion times beyond 1 h provided no increase in the number of peptide identifications. In fact, the *Lys-C/trypsin* protocol, in which proteolysis was conducted for the longest total time (32 h), produced the fewest peptide identifications in this sample. Although the *Guan-HCl* and *CH_3CN 16 h* protocols each had incubation periods of 16 h, approximately 100 more tryptic peptides were detected in the latter (Table 2). It is possible that the *Guan-HCl* and *Lys-C/trypsin* protocols did not yield as many peptides as the *CH_3CN 16 h* protocol due to the combined effects of protein or peptide losses to sample tube surfaces from the chaotrope dilution. Furthermore, it is also possible that SPE was an additional source of peptide loss for the *Guan-HCl* protocol. Base peak chromatograms for one run from each digestion protocol are shown in Figure S1 (Supporting Information).

An examination of the physical characteristics of tryptic peptide identifications from each digestion protocol revealed that the majority of peptide identifications were distributed over a range of 800−1800 Da with p*I* values ranging from 3 to 12, regardless of digestion protocol (Figure S2, Supporting Information).

Table 3 lists the number of ribosomal and nonribosomal protein identifications from each digestion protocol. Analysis of variance (ANOVA) indicates that a null hypothesis that the

Table 2. Average Peptide Identifications from *R. palustris* 70*S* Ribosomal Protein Isolates by Digestion Protocol[a]

digestion protocol	tryptic[b] peptides	semi-tryptic peptides[c]	non-specific peptides[d]	total peptide identifications[e]	percent tryptic[f]	3−4 missed cleavages[g]
Guan-HCl	293 ± 19	37 ± 2	0.3 ± 0.5	331 ± 21	88.3−89.0	3 ± 1
Lys-C/trypsin	203 ± 16	36 ± 16	6.5 ± 1.3	245 ± 29	78.8−89.3	14 ± 3
CH_3CN 1 h	391 ± 33	66 ± 9	1.3 ± 1.3	458 ± 26	81.5−87.4	40 ± 8
CH_3CN 16 h	380 ± 49	71 ± 23	2.5 ± 1.7	453 ± 47	75.7−88.1	0
RapiGest	261 ± 27	46 ± 5	0.3 ± 0.5	307 ± 32	84.9−85.4	6 ± 1

[a] Average values with standard deviation are shown from replicate experiments (see text for details). [b] Peptides with two termini resulting from cleavage C-terminal to lysine or arginine residues. [c] Peptides with one terminus resulting from cleavage C-terminal to a lysine or arginine residue. [d] Peptides with neither terminus resulting from cleavage C-terminal to a lysine or arginine residue. [e] Total peptide identifications regardless of enzyme specificity. [f] Percentage of total peptide identifications with two tryptic termini. [g] Tryptic peptides containing either three or four missed trypsin cleavage sites.

Table 3. Protein Identifications from *R. palustris* 70*S* Ribosomal Protein Isolates by Digestion Protocol

digestion protocol	ribosomal/ non-ribosomal[a]
Guan-HCl	46/2
	49/1
	46/2
	46/2
Lys-C/trypsin	41/1
	44/0
	40/1
	40/0
CH_3CN 1 h	47/3
	48/3
	49/4
	49/3
CH_3CN 16 h	45/1
	45/6
	47/4
	47/5
RapiGest	44/2
	40/2
	45/3
	44/4

[a] Based on functional annotation of the protein.

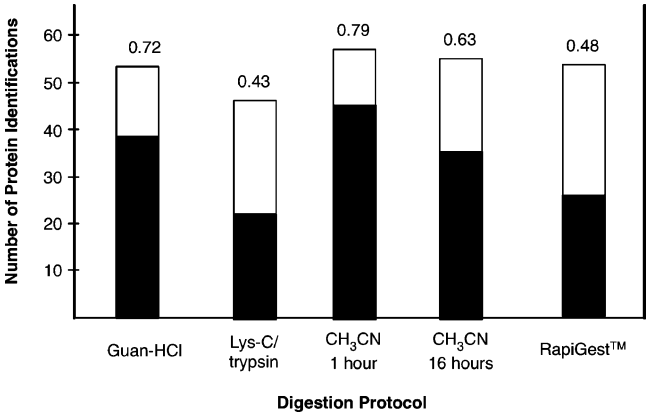

Figure 3. Proportion of identifications from *R. palustris* 70*S* ribosomal protein isolates with average sequence coverage greater than 30% by digestion protocol. The height of each bar indicates the total number of protein identifications from tryptic peptides. The solid portion of each bar indicates the number of proteins with an average sequence coverage greater than 30%; the proportional value is indicated at the top of each bar. Numbers of total protein identifications accumulate all proteins identified in any replicate measurement and, therefore, are larger than any individual measurement listed in Table 3.

five digestion protocols yield similar numbers of protein identifications can be rejected at a >99% confidence level; the *CH_3CN 1 h* protocol led, on average, to identification of the largest number of ribosomal proteins, and the *Lys-C/trypsin* protocol led to the smallest. The 40−49 ribosomal protein identifications, along with a low number of co-sedimenting nonribosomal proteins, are typical of a single LC−MS/MS analysis of tryptic digests of ribosomal proteins isolated by sucrose density fractionation. In our previous report, a single 110-min capillary one-dimensional reverse-phase LC−MS/MS separation of 10 μg trypsin digestions of ribosomal isolates yielded the identification of 186 peptides matching 41 *R. palustris* proteins.[4] Protein identifications from each LC−MS/MS experiment for all digestion protocols are listed in Table S1 (Supporting Information).

Figure 3 compares the digestion protocols in terms of the proportion of identifications from 70*S* ribosomal protein isolates with average sequence coverage greater than 30%. The *CH_3CN 1 h* protocol produced the largest proportion of protein identifications with greater than 30% coverage, followed by the *Guan-HCl*, *CH_3CN 16 h*, *RapiGest*, and *Lys-C/trypsin* protocols, in that order. The *RapiGest* and *Lys-C/trypsin* protocols resulted in ≥30% sequence coverage for less than half of the identified proteins.

Figure S4 in Supporting Information summarizes sequence coverage and physical characteristics for a total of 53 70*S* ribosomal proteins identified via any digestion protocol. The *Guan-HCl* protocol yielded the most protein identifications with the highest sequence coverage followed by the *CH_3CN 1 h* protocol; together, these two protocols provided the highest sequence coverage for 46 of the 53 proteins. Figures S5 and S6 (Supporting Information) show the distribution of sequence coverage values of proteins over 4 p*I* ranges and 3 molecular weight ranges.

For the 70*S* ribosomal isolates, the *CH_3CN 1 h* protocol produced the highest number of peptide identifications and ribosomal protein identifications, and resulted in the highest proportion of protein identifications with sequence coverage values greater than 30%. The effectiveness of this protocol on this sample may be due to one or more of the following factors. First, the 80% acetonitrile solvent need not be diluted prior to trypsin digestion, while dilution of the chaotrope in the *Guan-HCl* and *Lys-C/trypsin* protocols is required to avoid inactivating trypsin. Such an increase in reaction volume for the digestion results in exposure to more surface area, which increases the potential for adsorptive loss of proteins or peptides to the walls of a centrifuge tube. Second, it is plausible that trypsin activity is affected in different ways by the various solvents used in these protocols, which span a range of chaotropes, pH values, and water concentrations. The 80% acetonitrile solvent may promote protease activity, as demonstrated in other mixed aqueous−organic solvents.[13,35] Finally, different sample treat-

Table 4. Average Peptide Identifications from *B. taurus* MAP Isolates by Digestion Protocol[a]

digestion protocol	tryptic[b] peptides	semi-tryptic peptides[c]	non-specific peptides[d]	total peptide identifications[e]	percent tryptic[f]	3−4 missed cleavages[g]
Guan-HCl	60 ± 23	5 ± 2	3.0 ± 1.4	67 ± 24	83.7−91.4	8 ± 4
Lys-C/trypsin	296 ± 44	71 ± 18	10.0 ± 4.1	377 ± 65	76.0−81.0	1 ± 1
CH_3CN 1 h	283 ± 137	51 ± 19	5.5 ± 3.0	339 ± 156	72.6−81.0	16 ± 3
CH_3CN 16 h	427 ± 48	75 ± 16	3.3 ± 1.5	505 ± 64	83.1−87.0	4 ± 4
RapiGest	281 ± 73	65 ± 25	0.0 ± 0.0	346 ± 97	79.4−86.3	13 ± 5

[a] Average values with standard deviation are shown from replicate experiments (see text for details). [b] Peptides with two termini resulting from cleavage C-terminal to lysine or arginine residues. [c] Peptides with one terminus resulting from cleavage C-terminal to a lysine or arginine residue. [d] Peptides with neither terminus resulting from cleavage C-terminal to a lysine or arginine residue. [e] Total peptide identifications regardless of enzyme specificity. [f] Percentage of total peptide identifications with two tryptic termini. [g] Tryptic peptide identifications containing either three or four missed trypsin cleavage sites.

ments for downstream compatibility with liquid chromatography are required. While the 80% acetonitrile solvent may easily be removed by centrifugal evaporation, the *Guan-HCl* and *RapiGest* protocols that we followed[4,17,18,21,22] both specify additional sample processing steps (SPE and acid precipitation, respectively) to remove the chaotrope, presenting an additional potential mechanism for loss of peptides.

***B. taurus* Microtubule-Associated Protein (MAP) Isolates.** As a subset of microtubule-binding proteins, MAPs govern the structural dynamics of the cytoskeleton by assembling and stabilizing tubulin heterodimers into microtubules, facilitating essential cellular processes such as mitosis.[34] The subset of characterized MAPs includes the lower molecular weight tau protein and the larger MAP1, MAP2, and MAP3 proteins, with each subspecies present in distinct isoforms as a result of alternative mRNA splicing. The tau and MAP2 proteins are principally expressed in neuronal tissue, while other MAPs have been identified and are distributed throughout other tissues and cell types in the body.[36] The subset of neuronal MAPs are of considerable biological interest, as these proteins have been linked to the formation of cytoskeletal neurofibrillary tangles, a defining characteristic of Alzheimer's disease.[23] The methodology to isolate MAPs from mammalian brain tissue involves at least two cycles of temperature-dependent depolymerization of tubulin proteins followed by phosphocellulose chromatography of the resulting tubulin-enriched fraction, yielding enriched amounts of the higher molecular weight MAPs and some residual amounts of tubulins.[36,37]

Table 4 shows the average number of peptide identifications from digestions of 2.5 μg of MAPs according to the five digestion protocols. The *CH_3CN 16 h* protocol yielded the highest number of both tryptic peptide and total peptide identifications, followed by the *Lys-C/trypsin*, *CH_3CN 1 h*, *RapiGest*, and *Guan-HCl* protocols. (It should be noted that we observed a 20 min delay in detection of major chromatographic peaks during one technical replicate of the *CH_3CN 1 h* protocol, leading to the large standard deviation in the number of peptide identifications.) The numbers of peptide identifications covered a larger range for the 5 digestion protocols applied to the MAPs than for the ribosomal proteins, perhaps reflecting the greater range of concentrations and the total number of MAP proteins. The physical characteristics of tryptic peptide identifications from each digestion protocol did not reveal any distinct, protocol-specific trends, as the majority of peptides ranged from 1300 to 2250 Da and were distributed across a p*I* range from 3.0 to 7.0 (Figure S3, Supporting Information). Likewise, there was no clear trend relating digestion time and number of peptide identifications; the two 16-h digestions had the highest and lowest number of peptide identifications, while the *RapiGest*, *CH_3CN 1 h*, and *Lys-C/trypsin* protocols all yielded similar numbers of peptide identifications. Incubation times of 16 h or longer resulted in fewer missed cleavage sites relative to 1-h incubations. This is significant, as the *Lys-C/trypsin* protocol requires more digestion time and total number of reagents. Digestion of the MAPs according to the *Guan-HCl* protocol resulted in by far the fewest peptide identifications, perhaps due to protein or peptide losses caused by dilution of the guanidine hydrochloride, and SPE, as described above.

Table 5. Protein Identifications from *B. taurus* MAP Isolates by Digestion Protocol

digestion protocol	MAP/ non-MAP[a]
Guan-HCl	9/11
	8/9
	14/13
	13/16
Lys-C/trypsin	18/78
	18/80
	18/61
	17/77
CH_3CN 1 h	18/59
	18/56
	19/3
	14/55
CH_3CN 16 h	19/87
	19/88
	18/80
	18/74
RapiGest	16/43
	19/65
	17/42
	17/52

[a] Based on functional annotation of the protein.

Table 5 lists the numbers of MAP and non-MAP protein identifications from each digestion protocol. ANOVA indicates that a null hypothesis that the five digestion protocols yield similar numbers of protein identifications can be rejected at a >99% confidence level. The various digestion protocols resulted in the identification of 8−19 neuronal MAPs. Protein identifications from each replicate LC−MS/MS experiment for all digestion protocols are listed in Table S2 (Supporting Information). The *Lys-C/trypsin*, *CH_3CN 1 h*, *CH_3CN 16 h*, and *RapiGest* protocols produced comparable numbers of MAP identifications, ranging from 14 to 19, while the number of identifications for the *Guan-HCl* protocol ranged from 8 to 14. The difference in number of identifications of non-MAP proteins differed more drastically by protocol. Excluding the sole outlier replicate of the *CH_3CN 1 h* protocol (see above) yielding 3 non-MAP identifications, the *Guan-HCl* protocol consistently produced significantly fewer non-MAP protein identifications than any of the other protocols.

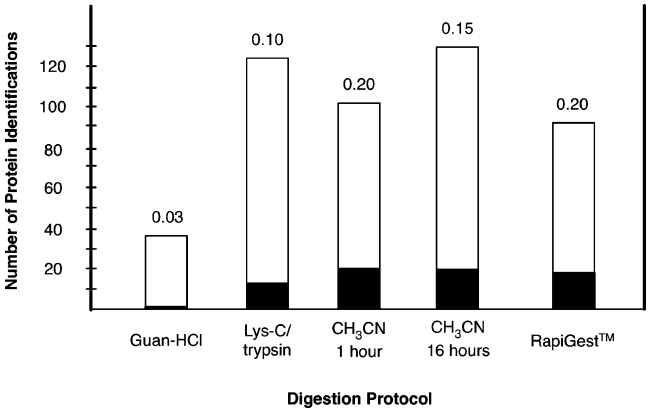

Figure 4. Proportion of identifications from *B. taurus* MAP isolates with average sequence coverage values greater than 30% by digestion protocol. The height of each bar indicates the total number of protein identifications from tryptic peptides by digestion protocol. The solid portion of each bar indicates the number of proteins with an average sequence coverage greater than 30%; the proportional value is indicated at the top of each bar. Numbers of total protein identifications accumulate all proteins identified in any replicate measurement and, therefore, are larger than any individual measurement listed in Table 5.

Figure 4 shows the proportion of protein identifications from MAP isolates with average sequence coverage greater than 30%. Digestions of MAPs according to the *CH_3CN 1 h*, *CH_3CN 16 h*, and *RapiGest* protocols yielded comparable results, with ≥30% sequence coverage for 18−20 MAPs. The *Lys-C/trypsin* protocol, while producing a relatively large number of total protein identifications, yielded only 12 protein identifications at ≥30% sequence coverage. In sharp contrast to the other four protocols, the *Guan-HCl* protocol produced only one protein with a sequence coverage ≥30%. Figure S4 (Supporting Information) provides a detailed analysis of sequence coverage and physical characteristics of 16 unique MAPs collectively identified by all digestion protocols. The *CH_3CN 16 h* or *CH_3CN 1 h* protocol provided the highest average sequence coverage values for 13 of these 16 MAPs.

A previous report of nanospray LC−MS/MS analyses of *B. taurus* MAPs ranged from 147 peptide identifications (64% were fully tryptic peptides) matching 26 protein identifications for a one-dimensional reverse-phase chromatographic separation to 431 peptide identifications (57% were fully tryptic peptides) matching 62 protein identifications for a 3-phase MudPIT separation.[19] The present study is thus comparable to previous results on a comparable analytical platform. For the MAP isolates, the *CH_3CN 16 h* protocol was most effective in terms of both total and tryptic peptide identifications, and yielded numbers of protein identifications and sequence coverage values greater than 30% that were among the highest. The *Guan-HCl* protocol produced the fewest peptide identifications, fewest total protein identifications, and only 1 protein with ≥30% sequence coverage values for MAPs. As discussed above for ribosomal proteins, this protocol has the most sample processing steps, namely, dilution of the guanidine hydrochloride and SPE, which offer the potential for the peptide loss. Digestion protocols conducted for as little as 1 h produced a similar number of peptide identifications as the *Lys-C/trypsin* protocol, suggesting that the combination of two proteases and digestion times beyond 16 h was not advantageous for peptide identifications in this sample.

Conclusions

Protein enrichment strategies often provide yields in the low-microgram to high-nanogram range. The characterization of such protein isolates by LC−MS/MS of peptides requires the digestion of proteins into peptides prior to the measurements of the intact mass and fragmentation pattern of the peptides. This digestion step is frequently taken for granted, especially for the digestion of limited or small quantities of protein, for example, those isolated by affinity purification. Here, we compared the performance of existing protocols for the digestion of small quantities of "real world" protein isolates and further demonstrated the effectiveness of the 80% acetonitrile solvent for trypsin digestions.

In both types of protein isolates examined, protocols where the trypsin digestion was conducted in the 80% acetonitrile solvent performed as well as, and in many instances better than, the other protocols tested. For isolates of the *R. palustris* 70*S* ribosomal protein complex, a 1 h digestion in 80% acetonitrile resulted in the highest number of peptide identifications. A 16 h digestion in 80% acetonitrile yielded in the highest number of peptide identifications for the *B. taurus* MAPs. Digestion in 80% acetonitrile solvent can be advantageous because it eliminates steps in the workflow that can lead to losses of protein or peptides, which becomes a more significant problem for samples that contain small quantities of protein. Chaotrope dilution steps, such as those in the *Guan-HCl* and *Lys-C/trypsin* protocols, increase the surface area to which the sample is exposed inside microcentrifuge tubes, increasing the possibility of adsorptive losses of proteins or peptides. To prevent interference in chromatography, chaotropes are frequently removed from peptide mixtures in a sample cleanup step prior to LC−MS/MS by SPE or precipitation, as in the *Guan-HCl* and *RapiGest* protocols. These steps increase the amount of sample handling, surface area to which the sample is exposed, and exacerbate the potential for peptide loss. The 80% acetonitrile solvent is volatile and may quickly be removed from peptide mixtures by centrifugal evaporation, apparently leading to fewer losses than SPE or precipitation of acid-labile surfactants. Regardless of the type of protein isolates examined, we consistently found the minimization of sample handling steps in the digestion protocols tested to be beneficial. Finally, although not investigated in this study, it is possible that the activity of trypsin differs among the solvents used in the various protocols, which span a considerable range of composition and pH.

Acknowledgment. The authors acknowledge Dr. Ryoma Ohi, the laboratory of Dr. Timothy J. Mitchison, Department of Systems Biology, Harvard Medical School, and Dr. W. Hayes McDonald (ORNL) for the MAP protein samples used in this study. We also thank the members of laboratory of Dr. Dale A. Pelletier of the Biological and Environmental Sciences Directorate of Oak Ridge National Laboratory for assistance with the growth of *R. palustris* cell cultures and laboratory pedagogy. W.J.H. gratefully acknowledges Christine T. Shook of the University of Tennessee Health Science Center, formerly of the OBMS group, for several invaluable discussions and the UT-ORNL Graduate School of Genome Science and Technology for support. Research sponsored by the Office of Biological and Environmental Research, U.S. Department of Energy, under contract No. DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC.

Supporting Information Available: Figures S1−S6 and Tables S1 and S2. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212−217.
(2) McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R., III. *Anal. Chem.* **1997**, *69*, 767−776.
(3) Washburn, M. P.; Wolters, D.; Yates, J. R., III. *Nat. Biotechnol.* **2001**, *19*, 242−247.
(4) 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.
(5) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III. *Nat. Biotechnol.* **1999**, *17*, 676−682.
(6) Rigaut, G.; Shevchenko, A.; Rutz, B.; Wilm, M.; Mann, M.; Seraphin, B. *Nat. Biotechnol.* **1999**, *17*, 1030−1032.
(7) 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.
(8) 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.
(9) Gould, K. L.; Ren, L.; Feoktistova, A. S.; Jennings, J. L.; Link, A. J. *Methods* **2004**, *33*, 239−244.
(10) Butland, G.; Peregrin-Alvarez, J. M.; Li, J.; Yang, W. H.; Yang, X. C.; Canadien, V.; Starostine, A.; Richards, D.; Beattie, B.; Krogan, N.; Davey, M.; Parkinson, J.; Greenblatt, J.; Emili, A. *Nature* **2005**, *433*, 531−537.
(11) Gingras, A. C.; Aebersold, R.; Raught, B. *J. Physiol. (London)* **2005**, *563*, 11−21.
(12) Klibanov, A. M. *Nature* **2001**, *409*, 241−246.
(13) Gupta, M. N.; Roy, I. *Eur. J. Biochem.* **2004**, *271*, 2575−2583.
(14) Welinder, K. G. *Anal. Biochem.* **1988**, *174*, 54−64.
(15) Russell, W. K.; Park, Z. Y.; Russell, D. H. *Anal. Chem.* **2001**, *73*, 2682−2685.
(16) Strader, M. B.; Tabb, D. L.; Hervey, W. J.; Pan, C.; Hurst, G. B. *Anal. Chem.* **2006**, *78*, 125−134.
(17) Promega Usage Information for Sequencing-Grade Modified Trypsin (http://www.promega.com/tbs/9piv511/9piv511.pdf).
(18) Riviere, L. R.; Tempst, P. Enzymatic Digestion of Proteins in Solution. In *Current Protocols in Protein Science*; Coligan, J. E., Ed.; Wiley: Hoboken, NJ, 1995; pp 11.1.1−11.1.19.
(19) McDonald, W. H.; Ohi, R.; Miyamoto, D. T.; Mitchison, T. J.; Yates, J. R. *Int. J. Mass Spectrom.* **2002**, *219*, 245−251.
(20) Klammer, A. A.; MacCoss, M. J. *J. Proteome Res.* **2006**, *5*, 695−700.
(21) Yu, Y. Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S.; Gebler, J. C. *Anal. Chem.* **2003**, *75*, 6023−6028.
(22) Yu, Y. Q.; Gilar, M.; Gebler, J. C. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 711−715.
(23) Dehmelt, L.; Halpain, S. *GenomeBiology* **2004**, *6*, 204.
(24) ThermoFinnigan Technical Support Bulletin 105 (http://www.thermo.com/eThermo/CMA/PDFs/Articles/articlesFile_21286.pdf).
(25) Yates, J. R., III; Eng, J. K.; McCormack, A. L.; Schieltz, D. *Anal. Chem.* **1995**, *67*, 1426−1436.
(26) VerBerkmoes, N. C.; Shah, M. B.; Lankford, P. K.; Pelletier, D. A.; Strader, M. B.; Tabb, D. L.; McDonald, W. H.; Barton, J. W.; Hurst, G. B.; Hauser, L.; Davison, B. H.; Beatty, J. T.; Harwood, C. S.; Tabita, F. R.; Hettich, R. L.; Larimer, F. W. *J. Proteome Res.* **2006**, *5*, 287−298.
(27) 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.
(28) IPI-International Protein Index: Statistics Web page, http://www.ebi.ac.uk/IPI/IPIcow.html.
(29) ORNL Genome Annotation Pipeline for *R. palustris*, http://www.ornl.gov/sci/GenomestoLife/index.shtml.
(30) Tabb, D. L.; McDonald, W. H.; Yates, J. R., III. *J. Proteome Res.* **2002**, *1*, 21−26.
(31) McDonald, W. H.; Tabb, D. L.; Sadygov, R. G.; MacCoss, M. J.; Venable, J.; Graumann, J.; Johnson, J. R.; Cociorva, D.; Yates, J. R., III. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2162−2168.
(32) Haynes, P. A.; Miller, S.; Radabaugh, T.; Galligan, M.; Breci, L.; Rohrbough, J.; Hickman, F.; Merchant, N. *J. Biomol. Tech.* **2006**, *17*, 97−102.
(33) Lundell, N.; Schreitmuller, T. *Anal. Biochem.* **1999**, *266*, 31−47.
(34) Wolfe, S. L. *Introduction to Cell and Molecular Biology*; Wadsworth: Belmont, CA, 1995.
(35) Batra, R.; Gupta M. N. *Biotechnol. Lett.* **1994**, *16*, 1059−1064.
(36) Olmsted, J. B. *Annu. Rev. Cell Biol.* **1986**, *2*, 421−457.
(37) Hyman, A.; Drechsel, D.; Kellog, D.; Salser, S.; Sawin, K.; Steffen, P.; Wordeman, L.; Mitchison, T. *Methods Enzymol.* **1991**, *196*, 478−485.

PR070159B