

Accelerated Articles

Processing Complex Mixtures of Intact Proteins for Direct Analysis by Mass Spectrometry

Fanyu Meng, Benjamin J. Cargile, Steven M. Patrie, Jeffrey R. Johnson, Shaun M. McLoughlin, and Neil L. Kelleher*

Department of Chemistry, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801

For analysis of intact proteins by mass spectrometry (MS), a new twist to a two-dimensional approach to proteome fractionation employs an acid-labile detergent instead of sodium dodecyl sulfate during continuous-elution gel electrophoresis. Use of this acid-labile surfactant (ALS) facilitates subsequent reversed-phase liquid chromatography (RPLC) for a net two-dimensional fractionation illustrated by transforming thousands of intact proteins from *Saccharomyces cerevisiae* to mixtures of 5–20 components (all within ~5 kDa of one another) for presentation via electrospray ionization (ESI) to a Fourier transform MS (FTMS). Between 3 and 13 proteins have been detected directly using ESI-FTMS (or MALDI-TOF), and the fractionation showed a peak capacity of ~400 between 0 and 70 kDa. A probability-based identification was made automatically from raw MS/MS data (obtained using a quadrupole-FTMS hybrid instrument) for one protein that differed from that predicted in a yeast database of ~19 000 protein forms. This ALS–PAGE/RPLC approach to proteome processing ameliorates the “front end” problem that accompanies direct analysis of whole proteins and assists the future realization of protein identification with 100% sequence coverage in a high-throughput format.

Contemporary proteomics using mass spectrometry (MS) mostly involves measurement of tryptic peptides of <3 kDa from abundant and increasingly less abundant proteins from >10⁵ cells.¹ Although the processing of intact proteins (no protease) has yet to be brought to an analogous level, their direct analysis by MS has the potential to characterize a diverse set of biological events

that produce mass discrepancies (Δm 's) between mature proteins and their corresponding open reading frames (ORFs) predicted from raw genomic sequence. Realization of this potential requires a robust protocol for efficient solubilization, fractionation, identification, and characterization of whole proteins.

Two- and even three-dimensional separations have been developed and commercialized involving chromatographic combinations of many types, including anion exchange^{2,3} or size exclusion chromatography (SEC)^{4,5} with subsequent reversed-phase liquid chromatography (RPLC). Isoelectric focusing (IEF) with subsequent RPLC using nonporous silica (NPS) has been reported using electrospray ionization (ESI) and time-of-flight (TOF) MS for protein detection.^{6,7} One-dimensional IEF in capillaries has been coupled to ESI-Fourier transform (FT)MS for protein profiling,⁸ but tandem MS (MS/MS) for direct protein identification has not been achieved in any high-throughput setting to date.

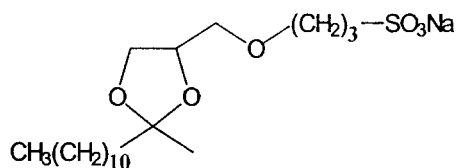
After significant fractionation, intact protein mixtures have been processed further by ESI-FT-MS/MS primarily using off-line sample introduction^{9–12} with direct fragmentation of protein ions

- (2) Holland, L. A.; Jorgenson, J. W. *Anal. Chem.* **1995**, *67*, 3275–3283.
- (3) Isobe, T.; Uchida, K.; Taoka, M.; Shinkai, F.; Manabe, T.; Okuyama, T. *J. Chromatogr.* **1991**, *588*, 115–123.
- (4) Opiteck, G. J.; Jorgenson, J. W. *Anal. Chem.* **1997**, *69*, 2283–2291.
- (5) Opiteck, G. J.; Ramirez, S. M.; Jorgenson, J. W.; Moseley, M. A. *Anal. Biochem.* **1998**, *258*, 349–361.
- (6) Wall, D. B.; Kachman, M. T.; Gong, S. S.; Parus, S. J.; Long, M. W.; Lubman, D. M. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1649–1661.
- (7) Chong, B. E.; Hamler, R. L.; Lubman, D. M.; Ethier, S. P.; Rosenspire, A. J.; Miller, F. R. *Anal. Chem.* **2001**, *73*, 1219–1227.
- (8) Shen, Y.; Berger, S.; Anderson, G.; Smith, R. *Anal. Chem.* **2000**, *72*, 2154–2159.
- (9) Kelleher, N. L.; Taylor, S. V.; Grannis, D.; Kinsland, C.; Chiu, H. J.; Begley, T. P.; McLafferty, F. W. *Protein Sci.* **1998**, *7*, 1796–1801.
- (10) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 806–812.
- (11) Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J. H.; Begley, T. P.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002**, *124*, 672–678.
- (12) Meng, F.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. *Nat. Biotechnol.* **2001**, *19*, 952–957.

* Corresponding author. E-mail: kelleher@scs.uiuc.edu. Fax: 217-244-8068.

(1) Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd. *Anal. Chem.* **2001**, *73*, 5683–5690.

Chart 1



being demonstrated on-line in only two reports using either CE¹³ or LC/FTMS.¹⁴ Such data from direct dissociation of protein ions of >10 kDa (by MS/MS or fragmentation in parallel¹²) can be more specific than peptide mapping for protein identification. This “top down” approach to protein sequence analysis also allows complete protein characterization far more efficiently than current proteomic approaches using tryptic peptides.^{10–12}

In 1994, Rose and Opiteck coupled preparative gel electrophoresis directly with RPLC¹⁵ using native proteins (no detergent). However, the use of a strong, ionic detergent with polyacrylamide gel electrophoresis (i.e., SDS–PAGE) is critical for efficient solubilization and fractionation of proteins based primarily on molecular weight. Such a size-dependent separation as a first dimension of proteome fractionation is attractive given that many protein forms produced from the same gene can reside within a few kilodaltons of one another. These varied forms arise from events such as RNA editing, heterozygous alleles of coding polymorphisms, and of course, covalent modifications to protein primary structure. Prior efforts to remove sodium dodecyl sulfate (SDS) in solution have proven difficult,^{16,17} especially for subsequent MS detection of whole proteins using electrospray ionization. While SDS adducts can be ejected in the gas phase,¹⁸ an “ESI-friendly” replacement for SDS¹⁹ would enable more robust protein processing prior to analysis by RPLC or MS. Building on a seminal 1989 article,²⁰ Waters has developed an acid-labile version of SDS called ALS (for acid-labile surfactant (Chart 1) and demonstrated 1D ALS–PAGE along with sensitivity enhancements for peptide mapping after in-gel digestion.^{21–23}

Here we report an initial demonstration of a proteome-processing approach to address the “front end” problem that accompanies direct analysis of whole proteins. Using ALS in combination with continuous-elution gel electrophoresis and RPLC

(referred to collectively as ALS–PAGE/RPLC), we illustrate sample preparation well-suited for presenting protein mixtures with a defined molecular weight range to a 9.4 T ESI-FT instrument. The ALS–PAGE/RPLC platform was used to fractionate, identify, and characterize intact proteins from cell lysates of *Saccharomyces cerevisiae*.

EXPERIMENTAL SECTION

Cell Growth and Protein Processing. *S. cerevisiae* cells (strain AB972)^{24,25} were grown to mid-log phase in YPD media (Becton Dickinson, MD) and ~1 g of cells (wet mass) were resuspended in 7 mL of lysis buffer (25 mM Tris, 1 mM EDTA, 200 μ M TCEP, 1 μ g/mL chymostatin, 1 μ g/mL leupeptin, 2 μ g/mL aprotinin, pH 8.4, with 1 μ L of DNAase). After lysis by a French press (15 000 psi), the cellular debris was cleared by centrifugation for 30 min (Sorvall, SS-34 rotor; 10000g). The supernatant was dialyzed against 25 mM Tris (pH 8.25) at 4 °C and then stored at –80 °C for later processing.

Three milliliters of the yeast cell extract was loaded on a model 491 Prep Cell (Bio-Rad) following the instructions described in the manual, with 0.1% ALS-I^{21–23} used instead of 0.1% SDS. A 12% T resolving gel was used with a 4% T stacking gel, and 40 fractions with a volume of 6 mL each were collected over 4 h after elution of the dye front. Fractions were either dialyzed against 25 mM NH₄HCO₃ (pH 8.2) or stored directly at –80 °C. For each fraction analyzed, 99% TFA (Sigma-Aldrich) was used to adjust the pH value of the solution to 1.9 (measured by pH meter) before it was stored at room temperature for 2 h. ALS-I degrades into dodeca-2-one and sodium 3-(2,3-dihydroxypropoxy)propanesulfonate with a half-life of 7.6 min²² at pH 1.9.

The fractions were then injected onto a C4 Exterra RPLC column (4.6 \times 100 mm; Waters Corp., Milford, MA) or NPS column (4.6 \times 14 mm; Eichrom Technologies, Darien, IL), washed for 20 min on-column, and eluted with a linear gradient over 20 min with standard solvents (H₂O, CH₃CN) and 0.1% TFA. The RPLC fractions containing ~0.5–10 μ g of total protein were dried down and redissolved in ESI solution (80% CH₃CN; 18% H₂O; 2% HOAc). Solutions of the protein mixtures were analyzed by SDS–PAGE, MALDI-TOF, or ESI-FTMS.

MALDI-TOF. Samples were run on a Voyager-DE-STR (Applied Biosystems) in linear mode with accelerating voltage of 25 kV and delayed extraction for 400 ns. The spectrum shown is the sum of 100 scans, and α -cyano-4-hydroxycinnamic acid was used as the matrix.

ESI-FTMS. Protein solutions were loaded into a microelectrospray assembly,²⁶ terminated with a fused-silica ESI tip with a 20- or 50- μ m i.d. (New Objective, Inc., Cambridge, MA). A syringe pump provided a flow rate of ~300 nL/min, and the typical ESI voltage was ~2.2–2.7 kV.

Two different FTMS systems were used in this study. The operation of the 9.4 T (Patrie, unpublished, 2000, NIMFL) actively shielded (Bruker Magnet) and the 9.4 T passively shielded

- (13) Valaskovic, G. A.; Kelleher, N. L.; McLafferty, F. W. *Science* **1996**, *273*, 1199–1202.
- (14) Li, W.; Hendrickson, C. L.; Emmett, M. R.; Marshall, A. G. *Anal. Chem.* **1999**, *71*, 4397–4402.
- (15) Rose, D. J.; Opiteck, G. J. *Anal. Chem.* **1994**, *66*, 2529–2536.
- (16) Davidsson, P.; Westman, A.; Puchades, M.; Nilsson, C. L.; Blennow, K. *Anal. Chem.* **1999**, *71*, 642–647.
- (17) Nilsson, C. L.; Larsson, T.; Gustafsson, E.; Karlsson, K. A.; Davidsson, P. *Anal. Chem.* **2000**, *72*, 2148–2153.
- (18) Fridriksson, E. K.; Baird, B.; McLafferty, F. W. *J. Am. Soc. Mass. Spectrom.* **1999**, *10*, 453–5.
- (19) Loo, R. R. O.; Dales, N.; Andrews, P. C. *Protein Sci.* **1994**, *3*, 1975–1983.
- (20) Jaeger, D. A.; Jamrozik, J.; Golich, T. G.; Clennan, M. W.; Mohebalian, J. *J. Am. Chem. Soc.* **1989**, *111*, 3001–3006.
- (21) Compton, B. J.; Lee, J. J.; Brown, E. K.; Herbert, R. G.; Ding, J.; Livingstone, J.; Bouvier, Ed. S. P.; Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics, Dallas, TX, 1999.
- (22) Ross, A. R. S.; Smith, D.; Langridge, J. I.; Gaskell, S. J. Proceedings of the 48th ASMS Conference on Mass Spectrometry and Allied Topics, Long Beach, CA, 2000.
- (23) Zeller, M.; Katalinic, J. P.; Bouvier, Ed. S. P.; Koenig, S. Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, IL, 2001.

- (24) Cherry, J. M.; Ball, C.; Dolinski, K.; Dwight, S.; Harris, M.; Matese, J. C.; Sherlock, G.; Binkley, G.; Jin, H.; Weng, S.; Botstein, D. <http://genome-www.stanford.edu/Saccharomyces/>.
- (25) Cherry, J. M.; Ball, C.; Weng, S.; Juvik, G.; Schmidt, R.; Adler, C.; Dunn, B.; Dwight, S.; Riles, L.; Mortimer, R. K.; Botstein, D. *Nature* **1997**, *387*, 67–73.
- (26) Emmett, M. R.; Caprioli, R. M. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 605–613.

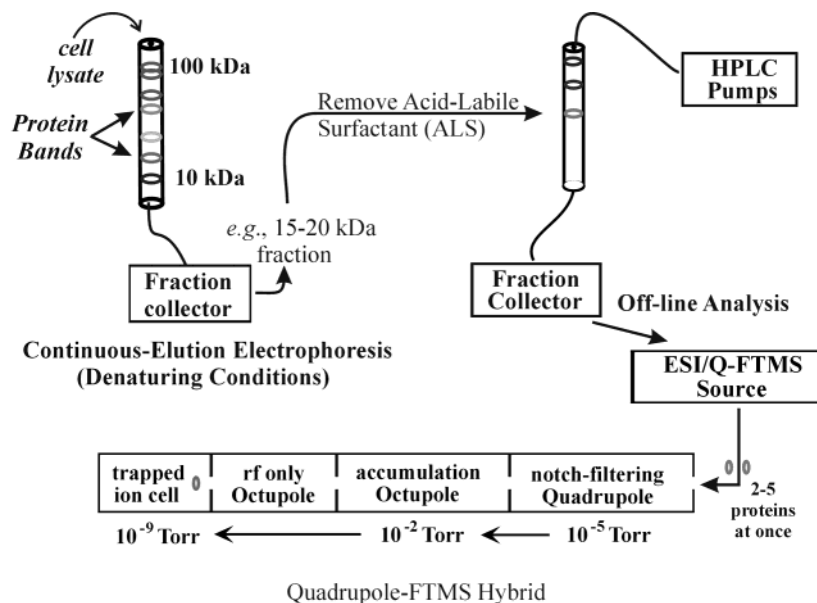


Figure 1. Schematic of ALS–PAGE coupled with RPLC as a proteome fractionation method before characterization of intact protein ions with a 9.4 T quadrupole-FTMS hybrid instrument.

systems are similar and described elsewhere.²⁷ In general, ions from ESI were directed through a heated metal capillary, skimmer, and multiple ion guides into the ion cell ($\sim 10^{-9}$ Torr) of the FTMS. Transients were stored with a MIDAS datastation²⁸ as 512K data sets. Theoretical isotopic distributions were generated using Isopro v3.0²⁹ and fit to experimental data by least squares to assign the most abundant isotopic peak. The mass difference (in units of 1.0024 Da³⁰) between the most abundant isotopic peak and the monoisotopic peak is denoted in *italics* after relative molecular weight (M_r) values. Spectra were calibrated externally using bovine ubiquitin, 8564.64–5.

Protein Ion Isolation and Fragmentation. For MS/MS, ion selection was either a one- or two-stage process. The standard technique of SWIFT is used for the one-stage isolation.³¹ For two-stage isolations, a quadrupole-FTMS hybrid (Q-FTMS) was used to selectively enhance ~ 30 m/z regions.³² Use of the Q-FTMS increases the dynamic range of successful MS/MS experiments by a factor of ~ 30 . In cases where charge states from several proteins appeared in the ~ 30 m/z region selected by the quadrupole filter, a SWIFT was employed to obtain a single charge state for fragmentation by infrared multiphoton dissociation (IRMPD; 75 W, 60–500 ms, 50% power).³³

Automated Data Analysis and Protein Identification. The deconvolution program was adapted from the Zhang and Marshall

algorithm³⁴ with the addition of a prefiltering algorithm to remove background signal from the spectrum. The prefilter uses low-resolution data (32 or 64K) and assumes that the noise intensity in the spectrum will occur with the greatest frequency.³⁵ Thus, the most frequent intensities were subtracted from the spectrum and this reduced spectrum was deconvoluted to yield an output indicating masses of species present and their corresponding m/z values. The “THRASH”³⁰ and database retrieval algorithm (DRA)¹² have been described previously, but several improvements and integration/automation of these programs will be described separately. Detailed ORF predictions for *S. cerevisiae* were acquired from the publicly accessible yeast protein database.²⁴ This database contains nonredundant ORF predictions from several bioinformatic interpretations of raw genomic data. Further we have predicted many protein forms from each gene, for example, by considering putative signal peptides and both Met-kept and Met-cleaved protein forms for each ORF. The database contains $\sim 19\,000$ protein forms. For the DRA program, the mass range was set to ± 2000 Da for intact proteins, and the error tolerance for fragment ion masses was set to ± 1 Da. Automated identifications were assumed correct if the probability of a spurious hit was lower than 0.1% (i.e., P score of < 0.001).¹² The integrity of hits with P scores of > 0.001 was confirmed by manual correlation of the fragmentation patterns with the amino acid sequences of the retrieved proteins.

RESULTS AND DISCUSSION

Performance of the ALS–PAGE/RPLC Fractionation Approach. An overview of the general fractionation and protein-processing platform is shown in Figure 1. Fractionation of 3 mL of whole cell lysate from *S. cerevisiae* using preparative PAGE and the acid-degradable surfactant in place of SDS yielded the separation profile of Figure 2 (only the first 30 fractions shown). ALS in the running buffer did not appreciably change the

- (27) Senko, M. W.; Hendrickson, C. L.; Pasa-Tolic, L.; Marto, J. A.; White, F. M.; Guan, S.; Marshall, A. G. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1824–1828.
- (28) Senko, M. W.; Canterbury, J. D.; Guan, S.; Marshall, A. G. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1839–1844.
- (29) Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229–233.
- (30) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320–332.
- (31) Marshall, A. G.; Wang, T. C. L.; Ricca, T. J. *Am. Chem. Soc.* **1985**, *107*, 7893–7897.
- (32) Hendrickson, C. L.; Quinn, J. P.; Emmett, M. R.; Marshall, A. G. Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, IL, 2001.
- (33) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. *Anal. Chem.* **1994**, *66*, 2809–2815.

- (34) Zhang, Z.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 225–233.
- (35) Ethier, M.; Zidarov, D.; Bertrand, M. Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, IL, 2001.

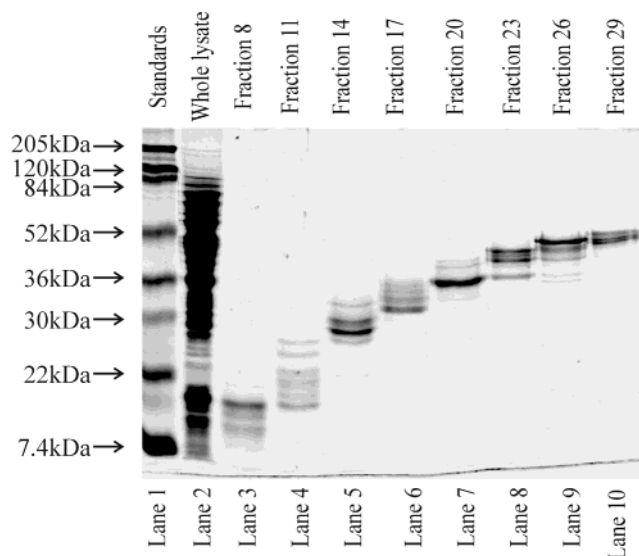


Figure 2. SDS-PAGE analysis of the ALS-preparative PAGE fractions (12% T gel). Lane 1, broad range markers; lane 2, whole cell lysate from *S. cerevisiae*; lanes 3–10, fractions as indicated.

fractionation behavior of the preparative gel, but it did slow the absolute migration times by ~10% relative to the analogous preparative SDS-PAGE run (data not shown). Analysis of the ALS-PAGE fractions by analytical PAGE (Figure 2) indicates a ~5–6-kDa window of proteins for each fraction with substantial changes in the protein content (even abundant ones) in different fractions. The resolution for the ALS-PAGE can be further optimized for small proteins by using a longer resolving gel, a higher %T, or a Tris–tricine buffer system,³⁶ but larger proteins will take even longer to elute out (a 50-kDa protein took ~4 h to elute from the 12% T gel of Figure 2). For the largest proteins, a lower %T gel is more suitable.

Subsequent acidification and fractionation of ALS-PAGE fractions 9 and 17 using RPLC yielded the chromatograms of Figure 3A and C containing 19–26 and 33–38 kDa, respectively, as viewed by SDS-PAGE (e.g., Figure 3B). Assuming the overall peak capacity is the product of those for each separation method,³⁷ the fractionation of proteins of <70 kDa showed an approximate peak capacity of 400 using a relatively large gel and a linear RPLC gradient. The first dimension achieved a size-dependent fractionation under strongly denaturing conditions, with the second dimension desalting and further fractionating mixtures before MS analysis. Many aspects and advantages of combining native PAGE with RPLC were described by Rose and Opiteck.¹⁵ Most apply to our denaturing ALS-PAGE/RPLC approach including the orthogonality of the two separation dimensions and flow rates/analysis times compatible with direct RPLC sampling of the ALS-PAGE effluent. Direct injection of 3 mL of ALS-PAGE fraction 8 onto a RPLC column followed by a 60-min wash at pH 1.9 resulted in a significant loss in chromatographic resolution (data not shown), suggesting that degradation of bulk ALS on-column is not a straightforward option to directly analyze ALS-PAGE effluent by RPLC.

MS Detection of Intact Proteins in Mixtures. Approximately 5–13 proteins could be detected per RPLC fraction using MALDI-TOF. A representative spectrum (Figure 3D) is shown for the ALS-PAGE fraction 9/RPLC fraction 33 (designated 9-33) and contains ~10 proteins within a 6-kDa window. When analyzed on a 9.4 T ESI-FT instrument, fraction 9-33 yielded the raw spectrum of Figure 4C containing four M_r values (20 411.8-13, 20 662.3-13, 21 274.1-13, and 22 167.0-13). Deconvolution of the charge-state distributions gave the output of Figure 4D and reaffirmed that all four of these proteins are within a narrow mass window. Note that the similar species observed do not arise from nonspecific adduction (e.g., the Δm values between the Figure 4C proteins

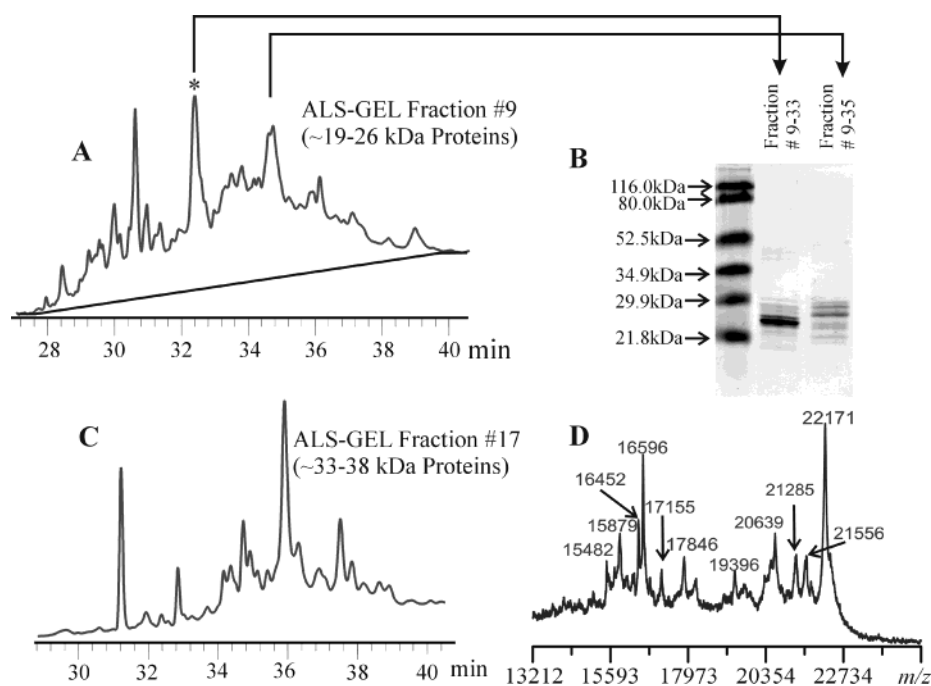


Figure 3. (A) RPLC fractionation (NPS column) of the ALS-PAGE fraction 9 from Figure 2. (B) SDS-PAGE of two RPLC fractions (9-33 and 9-35). (C) RPLC fractionation (C4 column) of ALS-PAGE fraction 17 from Figure 2. (D) Low-resolution MALDI-TOF mass spectrum of the fraction indicated with an asterisk in (A).

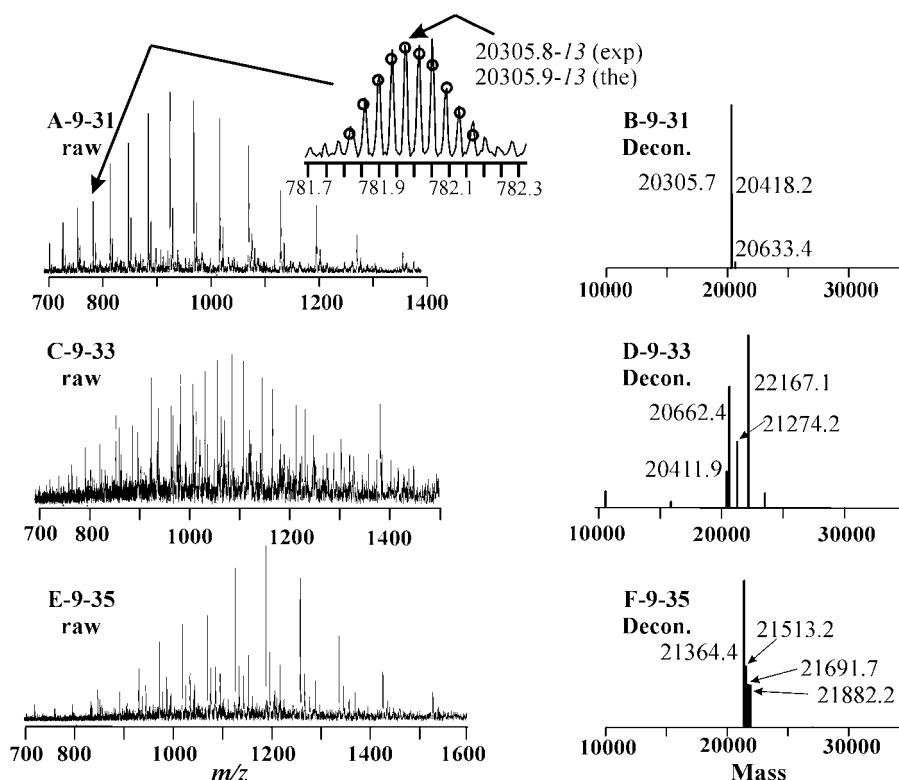


Figure 4. ESI-FT mass spectra of the RPLC fractions collected at 31 (A, 200 scans), 33 (C, 154 scans), and 35 (E, 150 scans) min from one acidified ALS fraction (9). Deconvoluted spectra corresponding to their respective raw data are shown at right; average molecular weight values are indicated.

Table 1. Summarized Output from Database Retrieval Algorithm^a

expt M_r	theor M_r	ΔMW	no. of b/y hits	score ^b	SI ^{b,c}	P score ^{b,d}	Δm mode?
20 292.8	20 700.1	-407.3	6	36	2	0.03	no
20 292.8	20 700.1	-407.3	13	119	2	4e-7	yes

^a Input to database retrieval algorithm (Output from THRASH): 17 fragment ions (all monoisotopic values) for the protein in Figure 4A (20 305.8-13 Da): 3055.72, 7590.18, 4855.84, 2814.53, 2141.10, 7455.10, 4534.50, 2337.23, 4584.36, 6899.81, 9406.30, 3927.06, 7099.94, 9596.19, 10696.7, 10885.8, 12837.7. ^b Defined in ref 12. ^c Specificity index. ^d Probability of a spurious match.

are 250.5, 611.8, and 892.9, which are not multiples of an integer). Similar experiments were performed for samples of different RPLC fractions from the ALS-PAGE fraction 9 (Figure 4A and E). Further, RPLC samples from different ALS-preparative cell fractions yielded intact protein spectra (MALDI-TOF and ESI-FTMS, data not shown) within the size range consistent with their prior analysis by 1D PAGE (e.g., Figure 2). These data verify that the second RPLC dimension can effectively desalt and fractionate proteins previously solubilized in ALS.

The presence of many species within a similar size range has advantages and disadvantages from analytical and biological perspectives. Using ESI, proteins of similar size may have more comparable ionization efficiencies, thus mitigating signal suppression. In this initial study, the apparent ionization efficiencies using MALDI are more equal, allowing twice the number of proteins to be detected (Figure 3D vs Figure 4C). Further optimization of ESI solution and FTMS conditions is being pursued, including use of a Q-FTMS to increase dynamic range. While instrumental

and data-processing software can be optimized for specific M_r ranges, similar charge-state distributions are observed (e.g., Figure 4A, C, and E) and rely heavily on MS/MS to discern whether they arise from the same gene. However, the possibility of measuring many forms of a gene product in the same fraction could provide a valuable "top down" view of posttranslational regulation of protein function.

Direct Fragmentation of Intact Proteins Using a Quadrupole-FTMS Hybrid. The 25⁺ charge state of the protein with an experimental M_r of 20 305.8-13 (Figure 4A) was selected and dissociated using IRMPD (Figure 5), with the THRASH software automatically detecting 22 isotopic distributions corresponding to 17 fragment ion masses over a 3–13-kDa size range (Table 1). These 17 masses were used without human intervention in our DRA software,¹² which yielded the highest scoring identifications shown in Table 1 from two different searches. In the first search, four protein forms corresponded to two nearly identical genes (PIR:S59848 and PIR:S56056) atop the retrieval list with six y-type ion matches within 20 ppm of observed fragment masses (Table 1). The theoretical M_r value from either gene did not match and no b-type ion matched the experimental fragment ions. The second

(36) Baumann, M.; Golabek, A.; Lalowski, M.; Wisniewski, T. *Anal. Biochem.* **1996**, 236, 191–198.

(37) Giddings, J. C. *Anal. Chem.* **1984**, 56, 1258A–1264A.

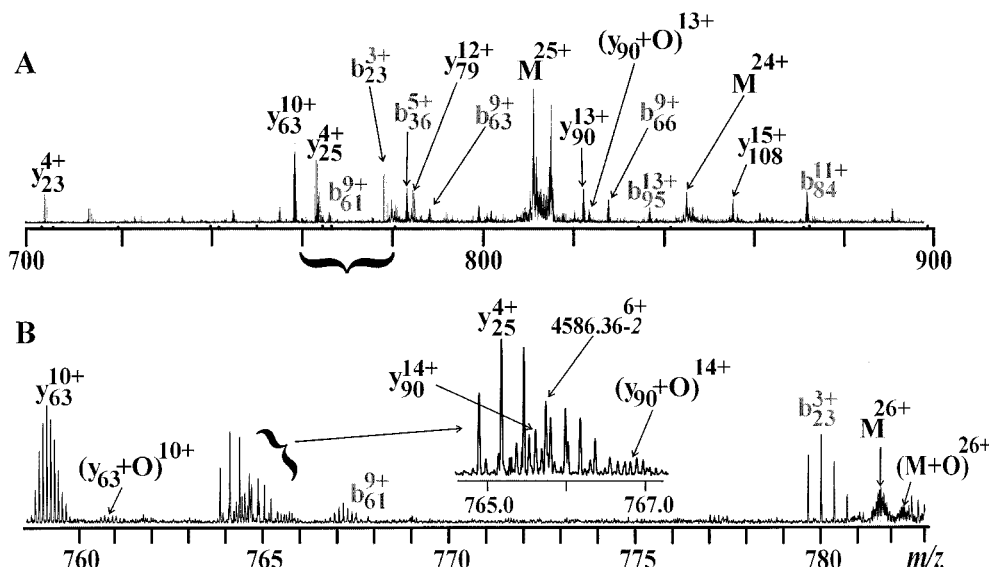


Figure 5. (A) An ESI/Q-FTMS fragmentation spectrum obtained from infrared dissociation of the 25^+ charge state of the 20 305.8-13 component in Figure 4A after 2D separation by ALS-PAGE/RPLC. (B) Expansion of the 760–780 m/z region of the spectrum in (A).

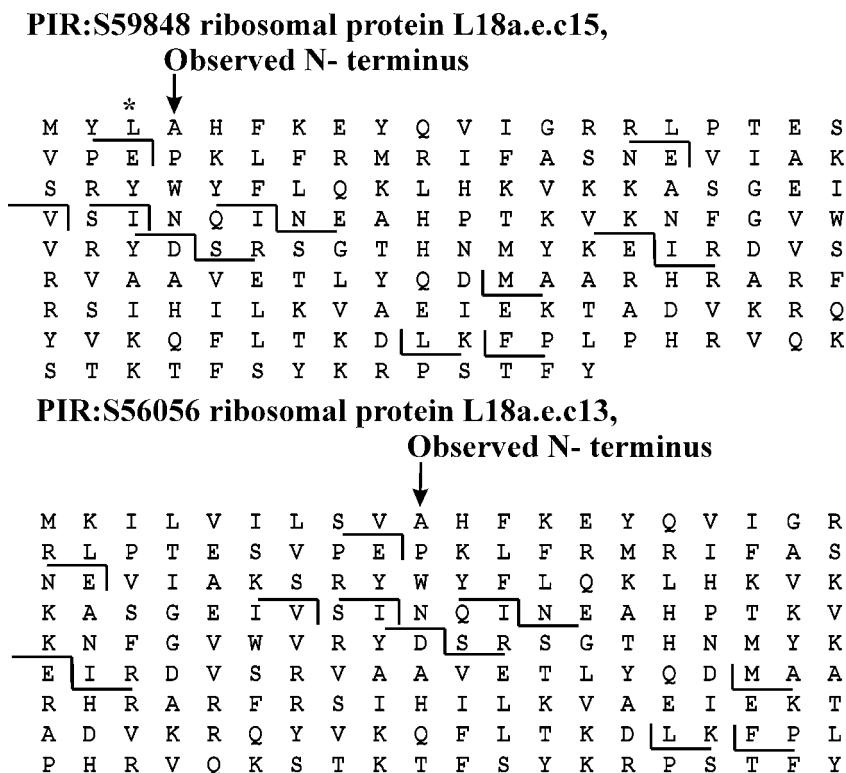


Figure 6. Primary sequence of yeast ribosomal protein S59848 (top) and S56056 (bottom) as found in the annotated database. Arrow indicates observed N-terminus and bars with left- or right-facing lines indicate b- and y-ions, respectively.

search was performed in “ Δm mode”, which allowed the program to consider simple mass discrepancies (i.e., only one source of a Δm within the protein) by calculating the Δm between the observed M_r and the theoretical M_r for a particular protein being considered. Using this Δm option, seven previously unmatched fragment ions now matched the b_{23} , b_{36} , b_{61} , b_{63} , b_{66} , b_{84} , and b_{95} ions from the S59848 gene product (or that from its duplicate gene, S56056) with a probability score of 4×10^{-7} (Table 1).¹²

Given both the M_r and fragment ion information, we have observed an N-terminally truncated product from gene S59848 or its duplicate, S56056 (Figure 6 top and bottom, respectively).

The removal of three or nine residues from S59848 (–407.3 Da) or S56056 (–996.7 Da), respectively, both result in a theoretical M_r of 20305.9-13 matching the observed value within 10 ppm. Possible explanations of this Δm include a sequence error or a protease in the cell (biologically relevant) or in the lysate (artifactual) cleaving these residues prior to our analysis. For S56056, no signal peptide sequence motif was found to explain a loss of nine residues. For S59848, it contains a 5' intron (like 67% of other yeast ribosomal proteins^{38,39}) which interrupts the codon for the Leu (Figure 6, asterisk) directly adjacent to the Ala we observed as the N-terminus (Figure 6, top left). Upon inspection,

no basis for an incorrectly predicted donor or acceptor site during RNA splicing was found. A homologue in *Schizosaccharomyces pombe* was observed to begin with an Ala residue analogous to the N-terminal Ala observed in this work.⁴⁰

Outlook. The ALS–PAGE/RPLC method should be applicable to membrane proteins soluble in ALS. Initial analysis of 22- and 46-kDa membrane proteins with six and eight transmembrane helices, respectively, by ALS–PAGE showed no loss in resolution versus SDS–PAGE (data not shown). Further, RPLC protocols using formic acid/2-propanol mobile phases could process complex mixtures of membrane proteins effectively.⁴¹ For mammalian proteome projects, the number of protein forms per sample will rise by at least two to three, placing a premium on mass spectral sensitivity, dynamic range, and MS/MS capabilities. The use of a Q-FTMS for enhanced “top down” characterizations of wild-type proteins will allow identification of simple sequence changes, such as sequence cleavages, PTMs, or point mutations. For robust identification of protein forms containing multiple Δm 's, long sequence tags produced by electron capture dissociation⁴² of Q-FTMS enriched ion populations is a viable approach for top down proteomics in mammalian systems. The subsequent challenge is to explain the absolute mass values of the fragment and protein ions to within 1 Da.

-
- (38) 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.
- (39) Planta, R. J.; Mager, W. H. *Yeast* **1998**, *14*, 471–477.
- (40) Appel, R. D.; Bairoch, A.; Hochstrasser, D. F. *Trends Biochem. Sci.* **1994**, *19*, 258–260. <http://www.expasy.ch/cgi-bin/sprot-search-ac?P05732>.
- (41) Whitelegge, J. P.; Gundersen, C. B.; Faull, K. F. *Protein Sci.* **1998**, *7*, 1423–1430.
- (42) Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1774–1779.

CONCLUSION

Use of an acid-labile substitute for SDS during preparative PAGE facilitates subsequent desalting and further fractionation of intact proteins using RPLC. This 2D proteome processing is well suited to presenting modest mixtures of intact proteins for MS. Further, use of a size-dependent fractionation as a first dimension is compelling given that many protein forms within ~5 kDa of one another can emanate from the same gene (an emerging theme for eukaryotic proteome projects.) Future efforts will involve continuous sampling of ALS–PAGE effluent by RPLC, increasing the overall resolving power, and implementation of nanospray robotics for sample-dependent MS/MS analysis of intact proteins using ESI-Q-FTMS. Such advances will establish an efficient system for identification and characterization of intact proteins on a large scale.

ACKNOWLEDGMENT

The acid-labile analogue of SDS was a generous gift from Edward Bouvier and Reb Russell of the Waters Corp. The authors thank John Quinn, Mark Emmett, Chris Hendrickson, and Alan Marshall for support with 9.4 T Q-FTMS instrumentation (National Science Foundation; National High-Field FT-ICR MS facility, NIMFL, NSF CHE-94-13008). N.L.K. received support from the University of Illinois (Critical Research Initiative), the National Institutes of Health (K22-AI01748-02), the Research Corp., and the Burroughs Wellcome and Searle Foundations. B.J.C. received a NSF Graduate Fellowship. J.R.J. received summer support from the Colgate-Palmolive Co.

Received for review January 23, 2002. Accepted April 22, 2002.

AC020049I