Proteomic profiling of intact proteins using WAX-RPLC 2-D separations and FTICR mass spectrometry. J Proteome Res 6, 602-610

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · MARCH 2007

Impact Factor: 4.25 · DOI: 10.1021/pr060354a · Source: PubMed

7 AUTHORS, INCLUDING:

David Charles Simpson
Virginia Commonwealth University
13 PUBLICATIONS **387** CITATIONS
SEE PROFILE

Navdeep Jaitly
University of Toronto
38 PUBLICATIONS **2,132** CITATIONS
SEE PROFILE

Richard D Smith
Pacific Northwest National Laboratory
1,132 PUBLICATIONS **46,141** CITATIONS
SEE PROFILE

research articles Journal of proteome research

Proteomic Profiling of Intact Proteins Using WAX-RPLC 2-D Separations and FTICR Mass Spectrometry

Seema Sharma,[‡] David C. Simpson,[†] Nikola Tolić, Navdeep Jaitly, Anoop M. Mayampurath, Richard D. Smith, and Ljiljana Paša-Tolić*

Pacific Northwest National Laboratory, P.O. Box 999, Richland, Washington 99352

Received July 18, 2006

We investigated the combination of weak anion exchange (WAX) fractionation and on-line reversed-phase liquid chromatography (RPLC) separation using a 12 T FTICR mass spectrometer for the detection of intact proteins from a *Shewanella oneidensis* MR-1 cell lysate. This work aimed at optimizing intact protein detection for profiling proteins at a level that incorporates their modification state. A total of 715 intact proteins were detected, and the combined results from the WAX fractions and the unfractionated cell lysate were aligned using LC-MS features to facilitate protein abundance measurements. Protein identifications and post-translational modifications were assigned for ~10% of the detected proteins by comparing intact protein mass measurements to proteins identified in peptide MS/MS analysis of an aliquot of the same fraction. Intact proteins were also detected for *S. oneidensis* lysates obtained from cells grown on ^{13}C-, ^{15}N-depleted media under aerobic and sub-oxic conditions. The strategy can be readily applied for measuring differential protein abundances and provides a platform for high-throughput selection of biologically relevant targets for further characterization.

Keywords: Comparative proteomics • FTICR MS • intact proteins • post-translational modifications

Introduction

Bottom-up proteomics based on either peptide mass finger printing or shotgun proteomics provides excellent proteome coverage;[1−6] however, factors that include incomplete sequence coverage limit the ability to identify protein modifications.[7,8] Gas-phase dissociation techniques for the mass spectrometric characterization of intact proteins as employed for top-down proteomics[9−11] have demonstrated 100% protein sequence coverage and allowed for the identification of protein isoforms, proteolytic processing events, and post-translational modifications (PTMs).[11−14] However, top-down analysis throughput is limited, for example, by fractionation used to reduce sample complexity and by analysis times needed to obtain protein tandem mass spectra of sufficient quality.[15,16] For samples of limited complexity and known composition, such as ribosomal preparations, confident protein identifications can be achieved by interfacing an efficient protein separation to a mass spectrometer capable of high-accuracy intact protein mass measurements (e.g., using mass locking with an internal standard).[17] Similar strategies applied to global protein mixtures result in the detection of many putative intact proteins; however, protein identification poses a challenge as the entire proteome must be considered as potentially being present. For example, Jensen et al. analyzed *Escherichia coli* cells grown on rare isotope-depleted media using an on-line combination of capillary isoelectric focusing (CIEF) with Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS) and detected >500 putative intact proteins, but in general, confident identifications could not be made.[18] In an extension to this work, isotope labeling and CIEF-FTICR MS were used to investigate the effect of a perturbation on protein abundance in *E. coli* cells.[18,19] Protein abundance ratios were readily calculated; however, few identifications could be made with confidence.

An integrated approach that combines complementary information from bottom-up and intact protein analyses has also been adopted for analyzing intact proteins.[20] VerBerkmoes et al. applied such a strategy to *Shewanella oneidensis* cell lysate; strong anion exchange chromatography was used to generate fractions that were analyzed using a bottom-up strategy, as well as analyzed at the intact protein level using FTICR MS.[21] They identified 395 proteins in the bottom-up analysis and 22 intact proteins by matching protein mass measurements against the data obtained from the bottom-up analysis. Millea et al. recently combined multi-dimensional chromatographic separations with simultaneous ESI-TOF MS on-line intact protein analysis and fraction collection of an *E. coli* cytosolic fraction in a so-called "middle-out" strategy.[22] The collected fractions were digested and subsequently analyzed using MALDI-Q-TOF tandem mass spectrometry (MS/MS). At the intact protein level, 75 proteins were detected, and by relating intact protein and peptide data, 46 proteins were identified. In another recent study, Simpson et al. investigated a separation strategy that combined size exclusion chroma-

* Corresponding author. Tel.: (509) 376-8859; fax: (509) 376-2303; e-mail: ljiljana.pasatolic@pnl.gov.

[†] Current address: Seattle Biomedical Research Institute, 307 Westlake Ave. N, Suite 500, Seattle, WA 98109-5219.

[‡] Current address: Center for Advanced Biotechnology and Medicine, 679 Hoes Lane, Piscataway, NJ 08854-5627.

10.1021/pr060354a CCC: $37.00
Published on Web 12/23/2006

tography fractionation with RPLC-FTICR MS and RPLC fractionation with CIEF-FTICR MS, which resulted in the detection of ~472 intact proteins for *S. oneidensis*.[23] An on-line separation will clearly increase the number of putative proteins that can be detected by MS; however, tandem mass spectrometry, which is not as readily compatible with on-line separations, is required for obtaining high-confidence protein identifications.

Reported herein is an investigation into the effectiveness of weak anion exchange fractionation (WAX) combined with on-line RPLC separations using a 12 T FTICR mass spectrometer as a platform suitable for high-throughput profiling of intact proteins. Samples in this investigation were obtained from *S. oneidensis*, an organism of interest because of its potential importance in bioremediation efforts.[24,25] Using this platform, 715 intact proteins were detected with <15 ppm mass measurement accuracy (MMA). Tentative identifications with post-translational modifications (PTMs) were assigned by using protein databases generated from bottom-up analyses of the corresponding fractions. Detected proteins from all the fractions and the lysate were combined to generate 2-D displays that provide the means to profile intact protein abundances in global proteomes. To illustrate the utility of such a platform for comparative proteomics, we obtained intact protein mass measurements of metabolically labeled (i.e., ^{13}C, ^{15}N depletion) *S. oneidensis* cell lysates grown under aerobic versus sub-oxic conditions. Cultivating cells in rare isotope-depleted media allows for improved MMA, sensitivity, and dynamic range.[26] This work is expected to form the basis for subsequent targeted MS/MS[27] and FTICR MS analyses to characterize the most interesting features detected in the initial comparative analysis. Furthermore, the techniques presented in this paper can readily be applied to comparative studies between control samples and those subjected to a biological perturbation, which could lead to interesting insights into the roles of PTMs.

Experimental Procedures

Cell Lysate Preparation. *S. oneidensis* MR-1 cells were cultured in either normal or isotopically-depleted (~99.95% ^{12}C and ~99.97% ^{14}N) Bioexpress media (Cambridge Isotope Laboratories, Andover, MA). Cells were cultured for high biomass using a defined medium in a BioFlo Model 110 fermentor (New Brunswick Scientific, Edison, NJ), which was operated in continuous feed mode as described elsewhere.[25] Dissolved oxygen was carefully controlled, and the carbon source, sodium DL-lactate, was present at 90 mM in the defined medium. Cells were harvested in the steady-state growth phase by centrifugation and were stored at −80 °C until needed. For cell lysis, stored cells were thawed and then suspended in a minimum volume of 20 mM potassium phosphate at pH 7.0 with 5% acetonitrile. Suspended cells were disrupted by bead-beating using a BioSpec Products (Bartlesville, OK) Mini-BeadBeater-8, with 0.1 mm zirconia/silica beads (BioSpec Products). Lysate was drawn off the beads by centrifugation at 14 000 rpm for 10 min. The collected cellular material was subjected to three repetitions of centrifugation at 100 000 rpm for 60 min; the pellet was discarded after each repetition to produce a clarified lysate that could pass through a 0.22 μm filter. Protein concentration in the clarified lysate was determined by using Pierce's (Rockford, IL) Coomassie Plus reagent.

WAX Fractionation of Cell Lysate. WAX fractionation was performed with a PolyLC (Columbia, MD) PolyWAX LP column (5 μm particles, 1000 Å pore size) on a Shimadzu (Kyoto, Japan) HPLC system (DGU-14A degasser, LC-10AD pumps, FRC-10A fraction collector, SCL-10A controller) fitted with a Thermo Separation Products (Rivera Beach, CA) Spectra 100 absorbance detector. The column (100 mm × 4.6 mm) was preceded by a guard column (10 mm × 4 mm) packed with the same material (1.79 mL total column volume). Mobile phase A was composed of 20 mM potassium phosphate (pH 7.0) with 5% acetonitrile, and mobile phase B consisted of 20 mM potassium phosphate and 600 mM potassium chloride (pH 7.0) with 5% acetonitrile. The fractionation started with a 30 min hold at 100% A, continued with a linear gradient to 100% B over 90 min, and finished with a 30 min hold at 100% B. The mobile phase flow rate was 0.5 mL/min, and the injection volume was 5 mL (4.35 mg of protein was injected); absorbance detection was performed at 230 nm, and 12 fractions (6 mL each) were collected in the time frame from 1−145 min. Evaporation and dialysis, performed by using Slide-A-Lyzer (Pierce) dialysis cassettes designed to retain molecules with masses in excess of 3.5 kDa, were used to generate concentrated protein samples. The initial 12 fractions were then combined into six unequal fractions to obtain a sufficient amount of protein in each sample for multiple analyses. The protein concentrations in the unfractionated clarified lysate and in the recovered WAX fractions showed that approximately 55% of the protein that had been injected onto the column was retained at this point. The fraction times for the combined fractions were 1−13, 13−37, 37−61, 61−85, 85−97, and 97−145 min.

RPLC-Ion Trap (IT) MS/MS Analysis of Tryptic Peptides. One aliquot from each WAX fraction and from the unfractionated cell lysate, in each case containing 50 μg of proteins, was digested using trypsin. Protein samples were denatured and reduced by adding solid urea to 7 M, solid thiourea to 2 M, freshly prepared aqueous dithiothreitol solution to 5 mM, and then by incubating at 60 °C for 30 min. For digestion, samples were diluted 10-fold with 100 mM ammonium bicarbonate, calcium chloride was added to 1 mM, and sequencing grade trypsin (Promega, Madison, WI) was added to give a trypsin/protein ratio of 1:50. Digestion was performed at 37 °C and was allowed to proceed overnight. Peptides were then extracted for MS analysis by using Omix tips (Varian, Palo Alto, CA); samples were evaporated to reduce the volume and then adjusted to match the loading mobile phase of the on-line separation.

Peptide RPLC separations were performed using a system similar in principle to that described by Shen et al.[28] The separation was performed under constant pressure conditions at 5000 psi using two ISCO (Lincoln, NE) Model 100 DM high-pressure syringe pumps; the column (60 cm × 150 μm) was packed in-house with Phenomenex (Torrance, CA) Jupiter particles (C_{18} stationary phase, 5 μm particles, and 300 Å pore size). Mobile phase A consisted of 0.05% trifluoroacetic acid (TFA), 0.2% acetic acid, and 99.75% water, while mobile phase B consisted of 0.1% TFA, 9.9% water, and 90% acetonitrile. Approximately 10 μg of peptides was loaded onto the column, and the loading process was allowed to proceed for 20 min. Once loading was complete, the injection loop was taken out of line and the gradient was started. The gradient was produced by adding mobile phase B to a stirred mixing chamber (volume 2.5 mL), which was initially filled with mobile phase A. A split was used to provide an initial flow rate through the column of ~2 μL/min. Electrospray ionization (ESI) using a pulled fused-silica needle was used to interface the RPLC separation to a Finnigan (San Jose, CA) LTQ linear ion trap mass spectrometer.

A maximum of five MS/MS spectra was recorded for peaks found in each survey mass spectrum.

RPLC-FTICR MS Analysis of Intact Proteins. The RPLC system used for on-line intact protein separations was largely the same as that used for peptide separations. Mobile phase A was altered to 0.1% TFA, 25% acetonitrile, and 74.9% water, while mobile phase B remained as 0.1% TFA, 9.9% water, and 90% acetonitrile. The column (80 cm × 75 μm i.d.) was packed with Phenomenex Jupiter particles (C_5 stationary phase, 5 μm particle diameter, and 300 Å pore size). The operating pressure was 8000 psi, and the reduced inner diameter for the column required the loading time to be increased to 45 min (which was later reduced to ~5 min by adding a solid-phase extraction column). Samples were diluted with mobile phase A and clarified by centrifugation (to protect the column from clogging) prior to loading. ESI using a pulled fused-silica needle was used to interface the RPLC separation to a modified Bruker 12 T APEX-Q FTICR mass spectrometer. This mass spectrometer incorporates an electrodynamic ion funnel,[29] quadrupoles for collisional focusing and ion pre-selection, a hexapole for external ion accumulation, and an RF-only ion guide for transferring the ions to the Infinity cell. A three-way pulsed leak valve assembly was used to introduce N_2 gas during the external accumulation event (typically 0.2 s). One mass spectrum was recorded every 2 s, and the average of three mass spectra was used for data analysis.

Data Processing. Peptide RPLC-IT MS/MS data were processed using SEQUEST (ThermoFinnigan, San Jose, CA);[30] no enzyme rules were applied, and identified peptides were filtered according to the rules laid out by Washburn et al.[31] The *S. oneidensis* database used for SEQUEST analysis has previously been described by Romine et al.[32] Provisional databases that contained proteins supported by at least two distinct peptide identifications and in the protein mass range of 5−40 kDa were created for each WAX fraction and the unfractionated cell lysate.

Intact protein RPLC-FTICR mass spectra were processed using in-house developed software (ICR-2LS). This software (available for download at http://ncrr.pnl.gov/software/) converts isotopic distributions to neutral masses (so-called deisotoping or mass transformation) utilizing an implementation of the THRASH algorithm.[33] Briefly, after the peak picking, an autocorrelation calculation was performed near the most abundant peak to predict the charge state by looking at the frequency of the surrounding peaks. The charge state and m/z value for the most abundant peak were then used to calculate an approximate molecular mass that was in turn used to predict an average molecular formula for a protein using the Averagine algorithm.[34] This molecular formula was then used to calculate a theoretical isotopic distribution using the Mercury algorithm;[35] theoretical and experimental isotopic distributions were then compared to determine an isotopic fit value (i.e., the least-squares error between the theoretical and the experimental data). The charge state, monoisotopic, average, and most abundant molecular masses for the lowest (i.e., best) isotopic fit value were assumed to be correct and are reported. This process was repeated until every isotopic distribution in a spectrum (above a given noise threshold) was processed and reduced to neutral mass.

The resulting mono-isotopic masses were then clustered into unique mass classes (UMCs) based on the neutral mass, charge state, abundance, isotopic fit (i.e., quality of fit between recorded and simulated isotopic pattern), and spectrum number (relating to RPLC retention time). Once the UMCs were defined, constituent mass spectra (members of each UMC) were summed and reprocessed to improve the signal-to-noise (S/N) ratio for low abundance species. To estimate the performance of the system for accurate mass measurement for an LC-MS dataset, we analyzed the mass variation within the UMCs. Figure 1 shows the graphs obtained by varying the clustering parameters used for defining UMCs (isotope fit <0.15, 5−50 members was the most stringent criteria; isotope fit <0.25 and any cluster with more than two members was the most relaxed criteria considered). Two estimators were evaluated; one based on a conservative 95 percentile cutoff of outlier UMCs and the other based on the slope of the rate of change of the UMC mass variation distribution function (i.e., first derivative of the UMC mass variation distribution function). While the latter method appears advantageous since it is invariant of the clustering parameters, the former method (95% cutoff) yields similar results for highly optimized UMCs. Numbers for distinct intact proteins were calculated by counting the UMC clusters separated in mass by more than 200 ppm, as distinct putative proteins. (This was required to compensate for the same isotopic envelope (mass range 5−40 kDa) resulting in neutral masses that differed by ±1 Da, due to the selection of a different isotopic peak as the most abundant isotopic peak within the envelope in consecutive spectra and for different charge states.) UMC masses were also searched against the appropriate provisional protein database (assembled from bottom-up data) for tentative intact protein identifications. Discrepancies between the intact protein masses and the predicted masses for proteins in the provisional databases were used to search for a limited set of protein PTMs (discussed in the following section). Tables of tentative intact protein identifications are provided as Supporting Information.

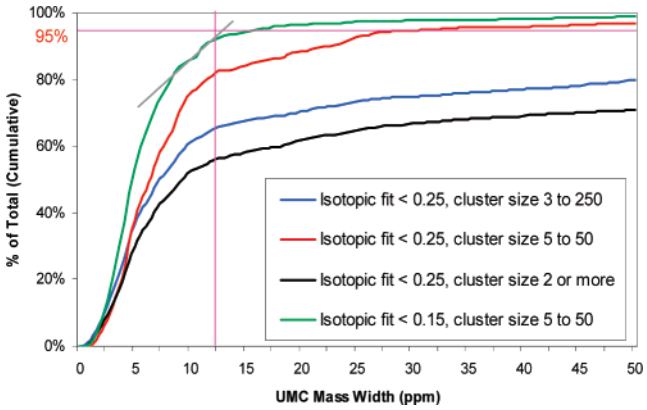

Figure 1. UMC based estimate of MMA for the LC-MS dataset. Two estimators were evaluated; one based on the conservative 95 percentile cutoff of outlier UMCs and the other based on a slope of the rate of change of the UMC mass variation distribution function (i.e., first derivative of the UMC mass variation distribution function). See text for details.

Results and Discussion

A combination of WAX fractionation and on-line RPLC-FTICR MS analysis was utilized to detect intact proteins in the lysate of *S. oneidensis* MR-1 cells grown on normal (i.e., natural isotopic abundance) media as depicted in Figure 2. Extensive ultracentrifugation of the raw lysate was carried out to prevent aggregated proteins from clogging the chromatographic columns. A comparison of calculated isoelectric points (pIs) of

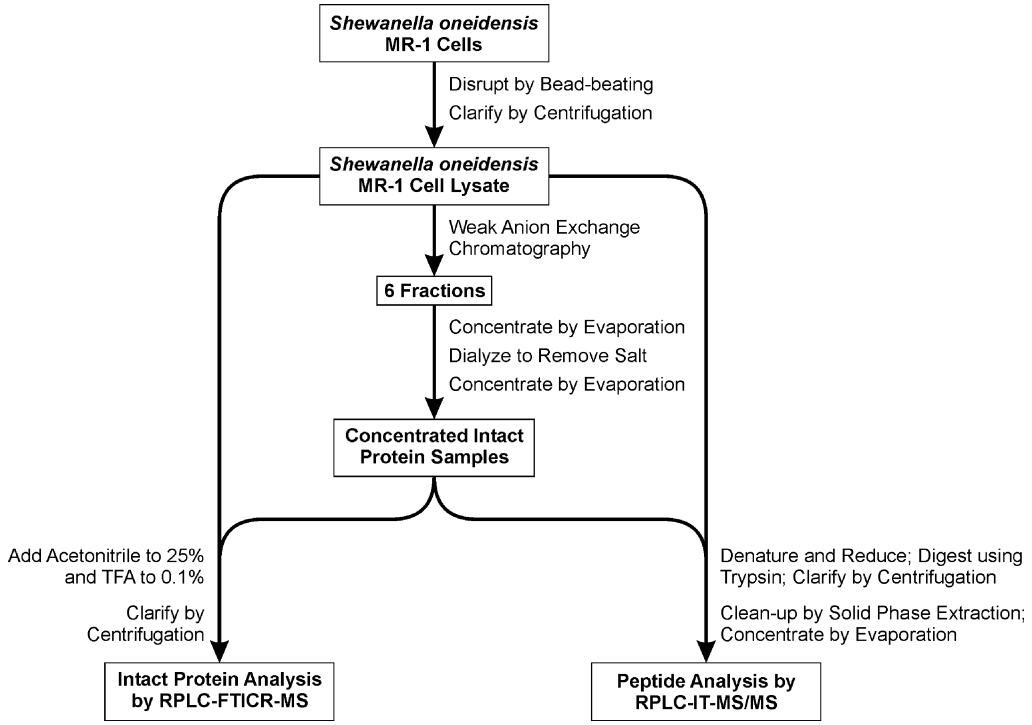

Figure 2. Experimental workflow outlining the preparative steps for generating samples for RPLC-FTICR MS intact protein analysis and RPLC-IT MS/MS peptide analysis.

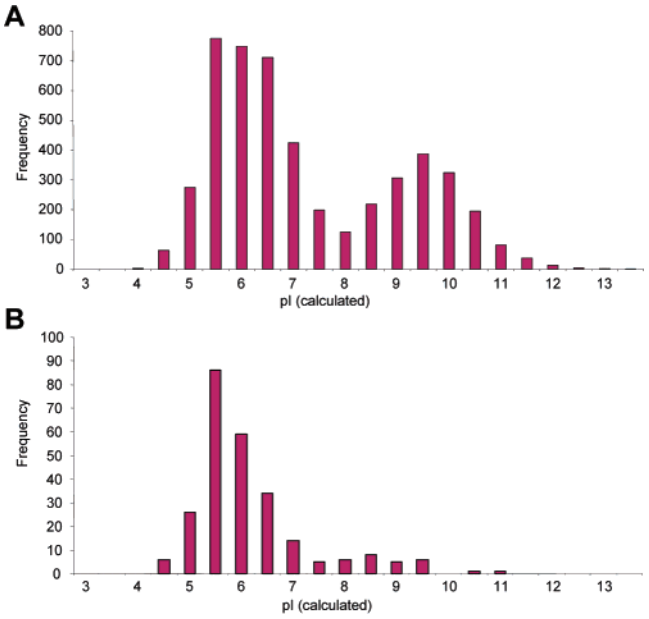

Figure 3. (A) Calculated isoelectric points (pIs) for all proteins contained in the *S. oneidensis* database used for SEQUEST analysis. (B) Calculated isoelectric points (pIs) for all proteins confidently identified by SEQUEST in the RPLC-IT MS/MS analysis of *S. oneidensis* clarified unfractionated cell lysate.

proteins contained in the *S. oneidensis* protein database[32] with proteins identified in the bottom-up analysis of the clarified unfractionated lysate (displayed as histograms in Figure 3) indicated an under-representation of proteins with basic isoelectric points (predominantly membrane-associated proteins).[36] The present work was aimed at profiling global intact protein samples, but as no special efforts were taken to extract membrane proteins, anion exchange fractionation was considered appropriate for lysate fractionation. WAX fractionation was performed at pH 7.0 to maintain protein solubility, which resulted in many proteins not binding to the column; these proteins (which simply flowed through the column) were analyzed along with the other fractions (since the aim here was merely to break the sample into manageable aliquots).

The utility of this strategy for profiling intact proteins is demonstrated by the detection of 715 intact proteins (mass range 5−40 kDa). Figure 4 shows the RPLC-FTICR MS total ion chromatogram and isotopic distributions for some of the proteins detected in the clarified unfractionated cell lysate. The LC-MS features common between each WAX fraction and the unfractionated lysate were used to align the individual 2-D displays using in-house developed software.[37] Combining the results from all WAX fractions and the cell lysate allows the creation of a profile suitable for comparative proteomics studies since it represents all the detected proteins. Also, for the purpose of quantitative measurements, it accounts for proteins being distributed in multiple fractions due to the WAX fractionation variability and low resolution (i.e., moderate overlap between fractions). The combined 2-D display (Master 2-D display) allows easy access to information regarding the fraction number and spectrum number (i.e., RPLC retention time) for each protein. Figure 5 compares the 2-D display reconstructed from the LC-FTICR analysis of the whole cell lysate (Figure 5A) with the 2-D display reconstructed by combining information from the fractions and cell lysate (Figure 5B) and reflects the improved detection of intact proteins following fractionation of the cell lysate. Specifically, the number of detected proteins increased from 195 in the lysate to 715 for the entire experiment (lysate and six WAX fractions).

To increase confidence in protein searching, tentative identifications were made by matching representative UMC masses (obtained by summing the mass spectra over the members of each UMC to increase S/N for low abundance proteins) against

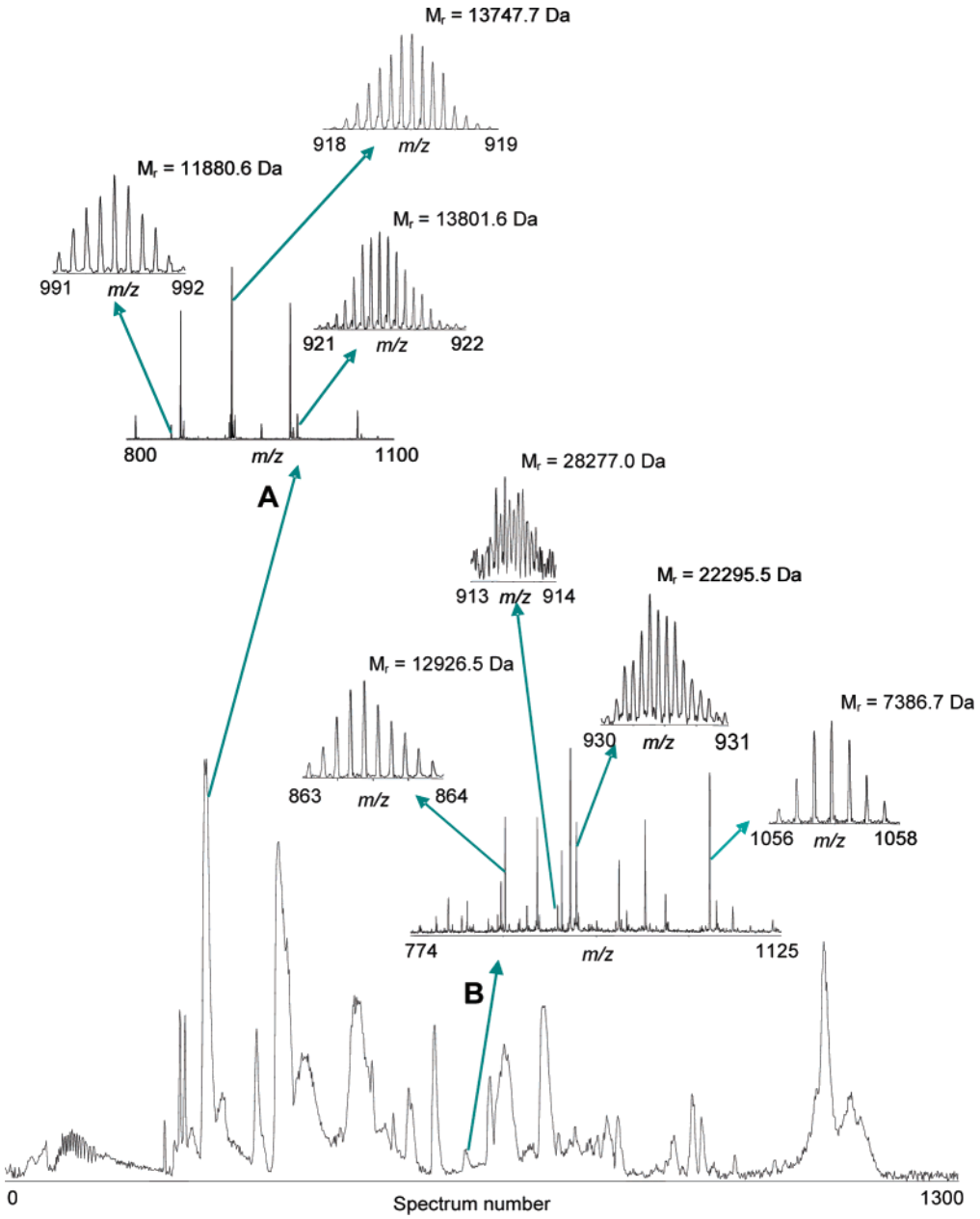

Figure 4. Total ion chromatogram (TIC) obtained for RPLC-FTICR MS intact protein analysis of the unfractionated cell lysate. Inset A shows mass spectrum (*m*/*z* 800−1100) and isotopic distributions of intact proteins (M_r = 11 880.6 Da, M_r = 13 747.7 Da, and M_r = 13 801.6 Da) detected under the most intense TIC peak. Inset B shows mass spectrum (*m*/*z* 774−1125) and isotopic distributions of intact proteins (M_r = 12 926.5 Da, M_r = 28 277.0 Da, M_r = 22 295.5 Da, and M_r = 7386.7 Da) detected under the relatively low-intensity TIC peak. Note that multiple species are detected even under low-intensity TIC peaks.

only those proteins that were identified in the bottom-up analysis of a second aliquot from the same fraction. The MMA <15 ppm used for protein searching was based on the data quality estimation (described in the Data Processing section), which indicated that beyond a certain UMC mass width (12.5 ppm, Figure 1), the number of UMCs that can be defined is independent of the clustering parameters. Notably, relaxing the MMA cutoff did not lead to a significant increase in the number of identified unmodified intact proteins as there is no appreciable increase in the number of UMCs that can be defined beyond 15 ppm. On the contrary, the number of false positives increased dramatically for low MMA (>15 ppm), particularly if multiple post-translational modifications and truncating events were considered.

Although the extent of protein modifications in prokaryotes is not as extensive as that observed in eukaryotes, some of the significant PTMs in prokaryotes included for tentative protein identifications were truncation of the initiating N-terminal methionine,[38] processing of the signal peptide sequence, and methionine oxidation (which might be also associated with sample processing). Table 1 summarizes the results obtained from the RPLC-IT MS/MS peptide analysis and RPLC-FTICR MS intact protein analysis of the unfractionated *S. oneidensis* cell lysate and the WAX fractions 1−6. While 447 unique proteins in the 5−40 kDa mass range were identified in the bottom-up analysis, tentative identifications were made for 80 of the 715 intact proteins detected. Overall, the results reflect better proteome coverage in the context of identified proteins

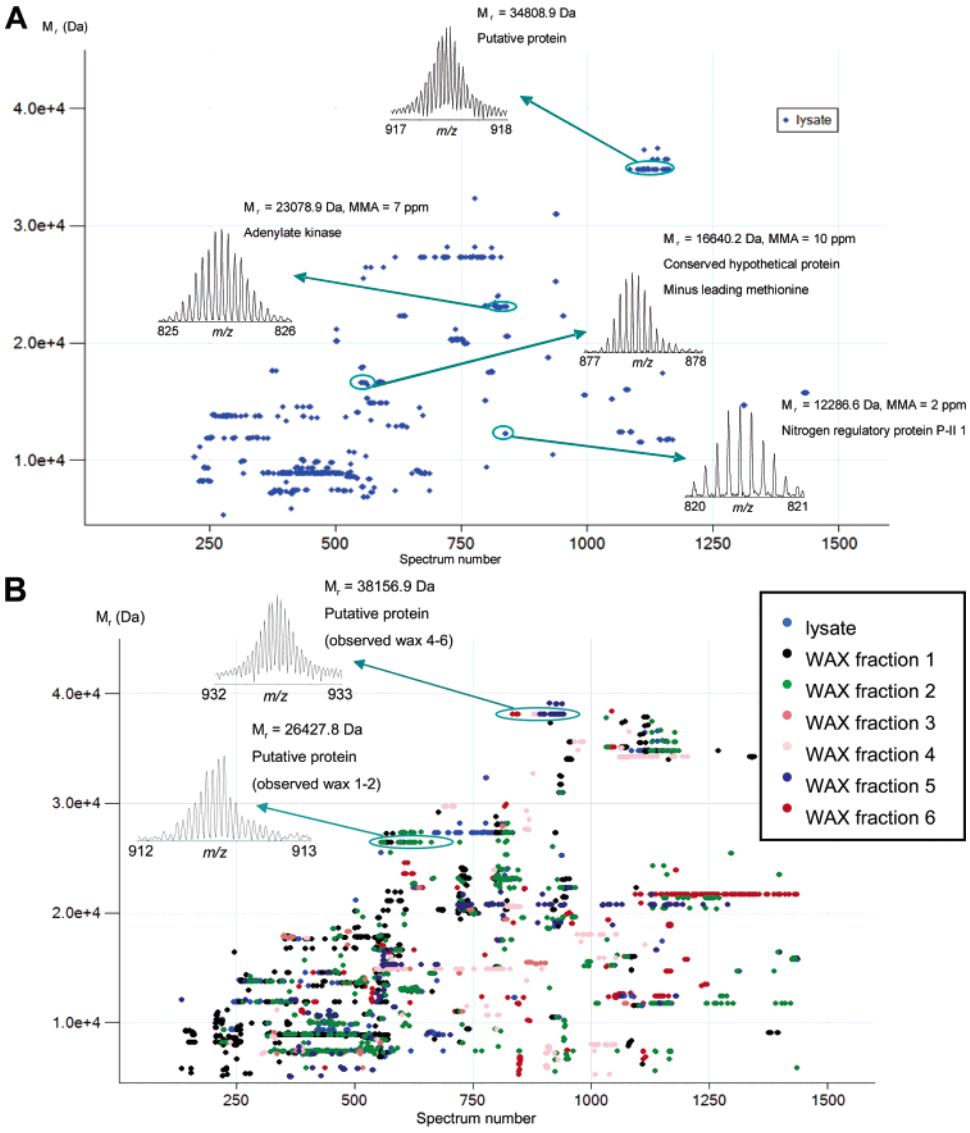

Figure 5. (A) 2-D display (monoisotopic mass vs spectrum number) reconstructed from the RPLC-FTICR MS analysis of unfractionated *S. oneidensis* lysate. Insets show isotopic distributions of a few representative intact proteins detected: M_r = 34 808.9 Da, putative intact protein; adenylate kinase, MMA = 7 ppm; nitrogen regulatory protein P-II 1, MMA = 2 ppm; and conserved hypothetical protein observed with a loss of the leading methionine, MMA = 10 ppm. (B) Combined 2-D display reconstructed from the RPLC-FTICR MS analysis of unfractionated *S. oneidensis* lysate and WAX fractions 1−6. Insets show representative isotopic distributions (M_r = 38 156.9 Da and M_r = 26 427.8 Da) detected in multiple fractions. Intact proteins common between each fraction and the unfractionated lysate were used for alignment of individual 2-D displays to provide the means to profile intact protein abundances in global proteomes.

for the bottom-up approach but also the ability of FTICR MS to detect proteins in more than one modification state.

Throughout the entire experiment, 28% of tentatively identified intact proteins and 15% of proteins identified in the bottom-up analysis exhibited a loss of the N-terminal methionine. Note that the lower percentage in the bottom-up analyses is most likely due to incomplete sequence coverage. Loss of the N-terminal methionine was evidenced at both the peptide and the intact protein level for 10 proteins, including one conserved hypothetical protein (SO1287) where two forms of the protein (with and without the N-terminal methionine) were detected by both approaches.

Signal peptide sequences based on the SignalP Version 3.0 neural network method[39,40] were predicted for 90 of the 447 proteins identified in the bottom-up analysis. Measured intact protein masses were searched against protein masses obtained after deletion of the predicted signal peptide sequence from the genome-derived sequence for these 90 proteins. Figure 6 shows two proteins identified in this manner (SO1638, outer

Table 1. Summary of Results for Proteins Identified in RPLC-IT MS/MS Peptide Analysis and Intact Proteins Detected in RPLC-FTICR MS Analysis of *S. oneidensis* Unfractionated Cell Lysate and WAX Fractions 1−6

protein sample	proteins in provisional bottom-up database	putative intact proteins detected	tentative intact proteins identified (including modified proteins)
whole cell lysate	257	195	27
WAX fraction 1	129	345	29
WAX fraction 2	161	288	29
WAX fraction 3	139	39	5
WAX fraction 4	166	125	11
WAX fraction 5	99	78	7
WAX fraction 6	140	164	12
Total	447	715	84

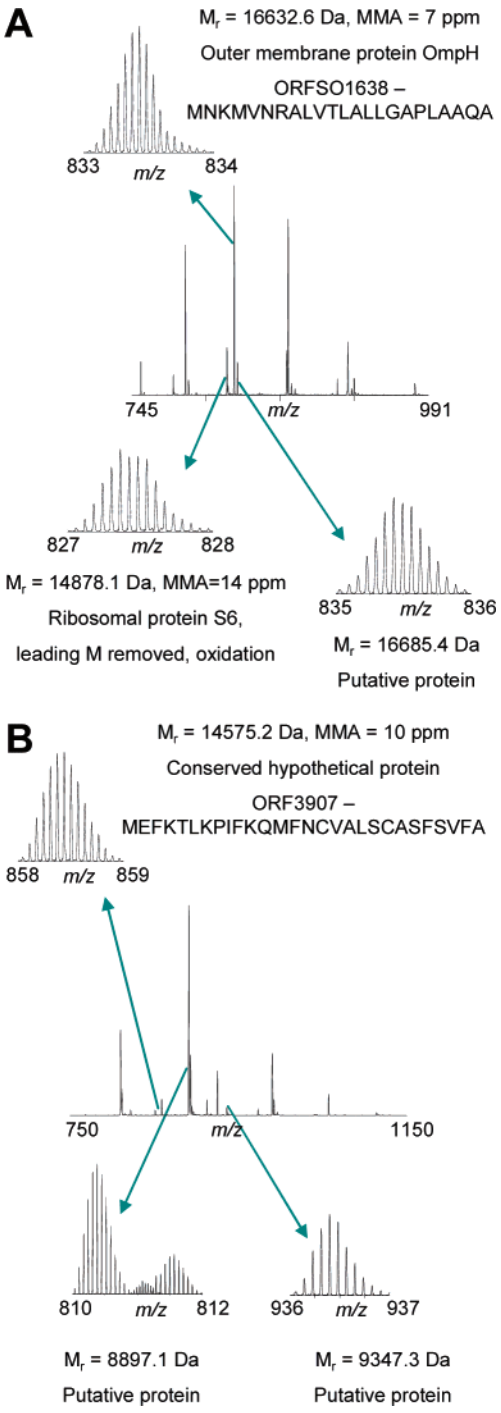

Figure 6. Examples of intact proteins identified with a loss of the predicted signal peptide sequence in RPLC-FTICR MS analysis of the unfractionated cell lysate. (A) Mass spectrum associated with the identification of SO1638 outer membrane protein OmpH observed with loss of the first 23 N-terminal amino acids (M_r = 16 632.6 Da and MMA = 7 ppm) along with coeluting intact proteins (M_r = 14 878.1 Da, ribosomal protein S6 observed with loss of the leading methionine and single oxidation and M_r = 16 685.4 Da, putative intact protein). (B) Mass spectrum associated with the identification of SO3907 conserved hypothetical protein observed with loss of the first 28 N-terminal amino acids (M_r = 14 575.2 Da and MMA = 10 ppm) along with coeluting putative proteins (M_r = 8897.1 Da and M_r = 9347.3 Da).

membrane protein and SO3907, conserved hypothetical protein). After considering peptide MS/MS data for these proteins

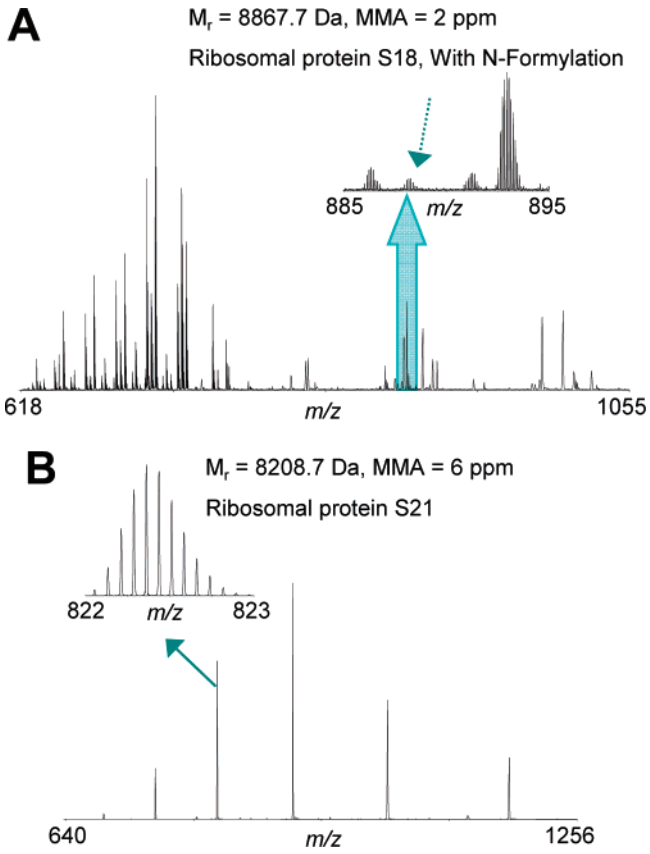

Figure 7. Identification of small intact proteins (<100 amino acids) in WAX fraction 1: (A) mass spectrum (*m/z* 618−1005 and enlarged *m/z* 885−895 region) associated with the identification of low abundance ribosomal protein S18 (SO3928) observed with N-formylation (M_r = 8867.7 Da and MMA = 2 ppm) and (B) mass spectrum (*m/z* 640−1256) associated with the identification of a high abundance ribosomal protein S21 (SO1288) observed in unmodified form (M_r = 8208.7 Da and MMA = 6 ppm). Both these proteins were missing from the provisional databases generated from bottom-up experiments.

and the predicted signal peptide sequences deleted from the genomic sequence, SO1638 could be identified with 77% sequence coverage and SO3907 could be identified with 90% sequence coverage. No peptides from the predicted signal peptide sequence were detected, and partial peptides formed after processing of the signal peptide sequence were detected in both cases.

In addition to the 80 intact protein identifications discussed previously, we also searched the intact protein data for predicted proteins with less than 100 amino acids that may not be identified by MS/MS analysis since they have fewer tryptic peptides.[32] We identified four additional intact proteins (SO0515, hypothetical protein; SO1288, ribosomal protein S21; SO2176, conserved hypothetical protein; and SO3928, ribosomal protein S18). The mass spectra of species assigned as ribosomal proteins are depicted in Figure 7. Re-evaluating the corresponding peptide MS/MS data revealed that only one fully tryptic peptide for SO3928 and SO1288, and none of the partial and non-tryptic peptides for SO0515 and SO2176 (no fully tryptic peptides had been identified for these), had passed the filtering criteria used in this study. Therefore, characterizing intact proteins might prove to be advantageous for the analysis and confirmation of small proteins. To summarize the tentative intact protein identifications, 30 unique proteins were observed

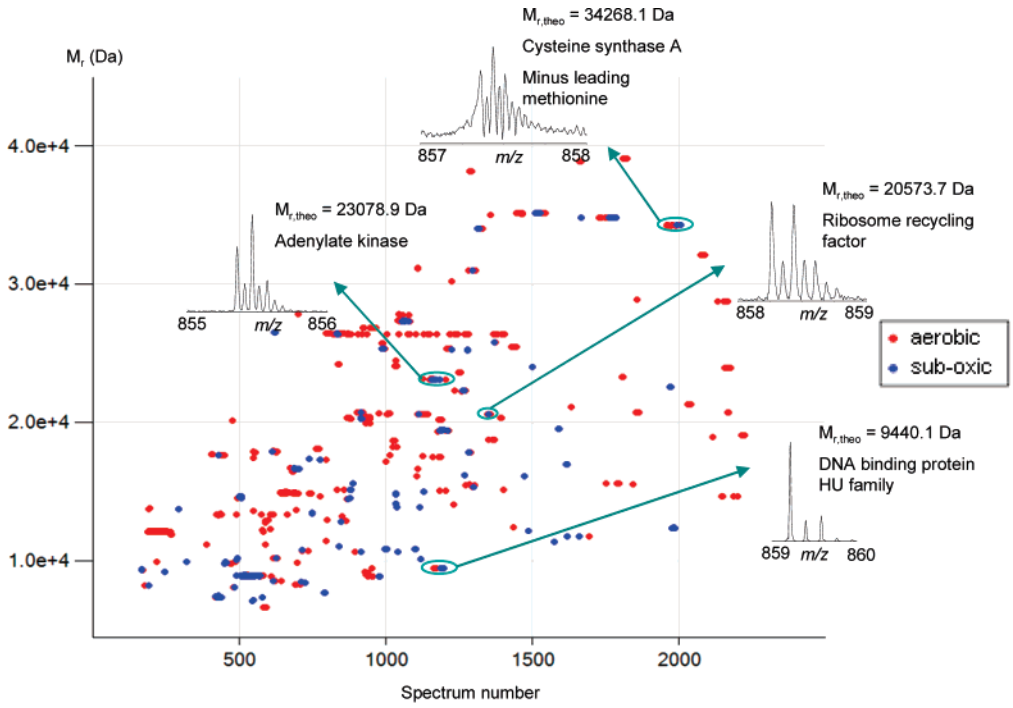

Figure 8. Alignment of the 2-D displays for proteins detected in the RPLC-FTICR MS analysis of *S. oneidensis* lysate from cells grown on ^{13}C-, ^{15}N-depleted media under aerobic conditions (shown in red) and sub-oxic conditions (shown in blue). Insets show examples of intact proteins that were detected under both conditions.

that retained the N-terminal initiating methionine, 20 were observed with a loss of the N-terminal methionine, 17 were observed with a loss of the predicted signal peptide sequence (including those that also indicated a loss of an additional N-terminal amino acid), 13 were oxidized forms of some of these proteins, and four were small proteins (less than 100 amino acids) that were not identified in the bottom-up analysis.

The identification of modified intact proteins is presently limited by the achievable MMA (50−100 ppm),[22,41] which restricts the assignment of post-translational modifications based on MS-only analysis of mixtures with limited complexity (e.g., WAX fractions). The UMC based estimate of <15 ppm for MMA reported herein represents a significant improvement over earlier reports; however, this MMA is still insufficient for considering PTMs such as unpredicted N-terminal processing of the peptide sequence. Therefore, we considered only those N-terminal truncations that corresponded to the loss of a predicted signal peptide sequence.[39,40] One way to achieve better MMA for microbial systems and cultured cells would be to improve the deisotoping of protein isotopic patterns by using samples grown in rare isotope-depleted media, as it allows for better detection and peak picking of the monoisotopic peaks.[26]

Profiling intact proteins can serve as a valuable tool for comparative proteomics as it allows high-throughput detection and can target fractions that contain unknown putative intact proteins of interest for further analysis. In the present study, *S. oneidensis* cells grown in ^{13}C-, ^{15}N-depleted media under aerobic conditions were compared with those grown under sub-oxic conditions. Figure 8 shows the aligned 2-D display for proteins detected in the cell lysate samples for these two cell growth conditions. A total of 196 intact proteins was detected in the lysate obtained from cells grown under aerobic conditions; 120 were detected in the lysate obtained from cells grown under sub-oxic conditions; and 48 intact proteins were common to both profiles. Because of sample limitations, the bottom-up provisional database was created only for the lysate grown under aerobic conditions. Using this database, tentative protein identifications were made for 49 intact proteins in the lysate from aerobic conditions, and 18 identifications were proposed for the lysate obtained from sub-oxic conditions. Seven proteins were tentatively identified as common to both samples. Both the unmodified and the oxidized forms of adenylate kinase (SO2018) were observed under aerobic conditions, while only the unmodified form was detected in the sub-oxic sample, which suggests that the methionine oxidation might not be a sample processing induced artifact.

In the next stages of this investigation, intact protein profiles obtained from fractionation of the two cell lysates will be compared to measure the variation in protein abundances. In addition, MS/MS will be used to characterize the proteins expressed differentially between the two samples.

Conclusion

The present study was aimed at optimizing the detection of intact proteins and facilitating comparative proteomic measurements (i.e., abundance profiling) at the intact protein level. WAX LC fractionation reduced sample complexity in the first dimension, and RPLC provided on-line separation of intact proteins prior to detection with a 12 T FTICR mass spectrometer. This approach resulted in detection of 715 intact proteins throughout the entire experiment. A UMC based estimate of MMA (<15 ppm) for the detected proteins was obtained by analyzing the precision of mass measurements for the whole LC-MS dataset. Using peptide level identifications from each sample as a control, tentative identifications were established for ~10% of the intact proteins detected. The prevalence of modified proteins at the intact protein level was manifested by the assignment of PTMs such as a loss of N-terminal methionine, processing of the predicted signal peptide se-

quence, and methionine oxidation for ~60% of the tentatively identified intact proteins. The use of this strategy to identify intact proteins is expected to further increase as MMA improves as a result of ongoing advances in ICR cell design, deisotoping algorithms, and implementation of automated gain control.

In parallel, we have demonstrated that proteomes can be effectively profiled at the intact protein level. Combined results from all WAX fractions and the cell lysate can be illustrated as a 2-D display, a form that is suitable for comparative abundance measurements since it represents all of the detected proteins, as well as the combined abundances for proteins distributed among more than one fraction. A proof-of-principle intact protein analysis of *S. oneidensis* cell lysates grown on ^{13}C-, ^{15}N-depleted media under aerobic and sub-oxic conditions resulted in the detection of 48 intact proteins common between the two cell states. In the next stages of this investigation, comparative proteomics will be employed for higher throughput selection of biologically interesting candidates as targets for gas-phase MS/MS characterization. The differential expression of unknown intact proteins (and modified proteins) that can readily be obtained by comparative abundance measurements of intact protein profiles can lead to the selection of important markers suitable for further investigation by targeted approaches.

Abbreviations: FTICR MS, Fourier transform ion cyclotron resonance mass spectrometry; MMA, mass measurement accuracy; RPLC, reversed-phase liquid chromatography; UMC, unique mass class; WAX, weak anion exchange chromatography.

Acknowledgment. This research was funded by the NIH National Center for Research Resources (RR18522). We gratefully acknowledge Dwayne Elias for providing the cell samples; Margaret Romine for signal peptide databases; Christophe Masselon, Anil Shukla, Rui Zhang, and Natacha Lourette for help with instrumentation; and Rui Zhao for assistance in maintaining the RPLC system. Experimental portions of this research were performed in the Environmental Molecular Sciences Laboratory, a U.S. Department of Energy (DOE) national scientific user facility located at the Pacific Northwest National Laboratory (PNNL) in Richland, WA. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.

Supporting Information Available: Tables showing the tentative identifications for intact proteins in *S. oneidensis* unfractionated lysate and WAX fractions 1−6 and the tentative identifications for predicted signal peptides in *S. oneidensis* unfractionated cell lysate and WAX fractions 1−6. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011−5015.
(2) James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 58−64.
(3) Mann, M.; Hojrup, P.; Roepstorff, P. *Biol. Mass Spectrom.* **1993**, *22* (*6*), 338−345.
(4) Pappin, D. J.; Hojrup, P.; Bleasby, A. J. *Curr. Biol.* **1993**, *3*, 327−332.
(5) Yates, J. R., III; Speicher, S.; Griffin, P. R.; Hunkapiller, T. *Anal. Biochem.* **1993**, *214*, 397−408.
(6) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198−207.
(7) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21* (*3*), 255−261.
(8) Cantin, G. T.; Yates, J. R., III. *J. Chromatogr. A* **2004**, *1053*, 7−14.
(9) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265−3266.
(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) Kelleher, N. L. *Anal. Chem.* **2004**, *76*, 196A−203A.
(12) Thomas, C. E.; Kelleher, N. L.; Mizzen, C. A. *J. Proteome Res.* **2006**, *5* (2), 240−247.
(13) Boyne, M. T., II; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. *J. Proteome Res.* **2006**, *5* (*2*), 248−253.
(14) Siuti, N.; Roth, M. J.; Mizzen, C. A.; Kelleher, N. L.; Pesavento, J. J. *J. Proteome Res.* **2006**, *5* (2), 233−239.
(15) Meng, F.; Du, Y.; Miller, L. M.; Patrie, S. M.; Robinson, D. E.; Kelleher, N. L. *Anal. Chem.* **2004**, *76* (10*)*, 2852−2858.
(16) Du, Y.; Parks, B. A.; Sohn, S.; Kwast, K. E.; Kelleher, N. L. *Anal. Chem.* **2006**, *78* (3), 686−694.
(17) Lee, S. W.; Berger, S. J.; Martinovic, S.; Pasa-Tolic, L.; Anderson, G. A.; Shen, Y.; Zhao, R.; Smith, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (9), 5942−5947.
(18) Jensen, P. K.; Pasa-Tolic, L.; Anderson, G. A.; Horner, J. A.; Lipton, M. S.; Bruce, J. E.; Smith, R. D. *Anal. Chem.* **1999**, *71*, 2076−2084.
(19) Paša-Tolić, L.; Jensen, P. K.; Anderson, G. A.; Lipton, M. S.; Peden, K. K.; Martinovic, S.; Tolić, N.; Bruce, J. E.; Smith, R. D. *J. Am. Chem. Soc.* **1999**, *121*, 7949−7950.
(20) Zabrouskov, V.; Giacomelli, L.; van Wijk, K. J.; McLafferty, F. W. *Mol. Cell. Proteomics* **2003**, *2* (12), 1253−1260.
(21) VerBerkmoes, N. C.; Bundy, J. L.; Hauser, L.; Asano, K. G.; Razumovskaya, J.; Larimer, F.; Hettich, R.; Stephenson, J. L., Jr. *J. Proteome Res.* **2002**, *1* (3), 239−252.
(22) Millea, K. M.; Krull, I. S.; Cohen, S. A.; Gebler, J. C.; Berger, S. J. *J. Proteome Res.* **2006**, *5*, 135−146.
(23) Simpson, D. C.; Ahn, S.; Paša-Tolić, L.; Bogdanov, B.; Mottaz, H. M.; Vilkov, A. N.; Anderson, G. A.; Lipton, M. S.; Smith, R. D. *Electrophoresis* **2006**, *27*, 2722−2733.
(24) Lovley, D. R.; Phillips, E. J. P.; Gorby, Y. A.; Landa, E. R. *Nature* **1991**, *350*, 413−416.
(25) Elias, D. A.; Monroe, M. E.; Marshall, M. J.; Romine, M. F.; Belieav, A. S.; Fredrickson, J. K.; Anderson, G. A.; Smith, R. D.; Lipton, M. S. *Proteomics* **2005**, *5* (12), 3120−3130.
(26) Rodgers, R. P.; Blumer, E. N.; Hendrickson, C. L.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (10), 835−840.
(27) Masselon, C.; Paša-Tolić, L.; Toliæ, N.; Anderson, G. A.; Bogdanov, B.; Vilkov, A. N.; Shen, Y.; Zhao, R.; Qian, W. J.; Lipton, M. S.; Camp, D. G., II; Smith, R. D. *Anal. Chem.* **2005**, *77* (2), 400−406.
(28) Shen, Y.; Zhao, R.; Belov, M. E.; Conrads, T. P.; Anderson, G. A.; Tang, K.; Paša-Tolić, L.; Veenstra, T. D.; Lipton, M. S.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 1766−1775.
(29) Shaffer, S. A.; Tang, K.; Anderson, G. A.; Prior, D. C.; Udseth, H. R.; Smith, R. D. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1813−1817.
(30) Eng, J. K.; McCormack, A. L.; Yates, J. R., III. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (*11*), 976−989.
(31) Washburn, M. P.; Wolters, D.; Yates, J. R., III. *Nat. Biotechnol.* **2001**, *19* (3), 242−247.
(32) Romine, M. F.; Elias, D. A.; Monroe, M. E.; Auberry, K. J.; Fang, R.; Fredrickson, J. K.; Anderson, G. A.; Smith, R. D.; Lipton, M. S. *OMICS* **2004**, *8* (3), 239−254.
(33) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (4), 320−332.
(34) Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229−233.
(35) Rockwood, A. L.; Van Orden, S. L.; Smith, R. D. *Anal. Chem.* **1995**, *67*, 2699−2704.
(36) Schwartz, R.; Ting, C. S.; King, J. *Genome Res.* **2001**, *11* (5), 703−709.
(37) Jaitly, N.; Monroe, M. E.; Petyuk, V. A.; Clauss, T.; Adkins, J. N.; Smith, R. D. *Anal. Chem.* **2006**, *78*, 7397−7409.
(38) Giglione, C.; Meinnel, T. *Trends Plant Sci.* **2001**, *6* (12), 566−572.
(39) Nielsen, H.; Engelbrecht, J.; Brunak, S.; von Heijne, G. *Protein Eng.* **1997**, *10* (1), 1−6.
(40) Bendtsen, J. D.; Nielsen, H.; von Heijne, G.; Brunak, S. *J. Mol. Biol.* **2004**, *340* (4), 783−795.
(41) Kelleher, N. L.; Senko, M. W.; Siegel, M. M.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1997**, *8* (4), 380−383.

PR060354A