Protein profiling using WAX-RPLC 2D separation and FTICR-MS intact protein analysis

Seema Sharma, David C. Simpson, Nikola Tolić, Navdeep Jaitly, Anoop Mayampurath, Natacha Lourette, Richard D. Smith, and Ljiljana Paša-Tolić

Pacific Northwest National Laboratory, Richland, WA 99352

Overview

❖ Accurate mass measurements for putative intact proteins from *Shewanella oneidensis* MR-1 cell lysate, using weak anion exchange chromatography fractionation and on-line reversed phase liquid chromatography separations with a 12T FTICR-MS.

❖ 2D displays generated for profiling proteomes at the intact protein level by combining the results from individual fractions.

❖ Applications of intact protein profiling in comparative proteomics illustrated for *S. oneidensis* cells grown under aerobic vs. sub-oxic conditions.

❖ Integration of intact protein and peptide level analyses for tentative identifications of intact proteins and post translational modifications (PTMs) in *S. oneidensis*.

Introduction

Bottom-up proteomics (analyzing peptides that result from protein digestion) has demonstrated capability for broad proteome coverage and good throughput. However, due to incomplete sequence coverage, this approach is not ideally suited for studying modified proteins. The modification complement of a protein can best be elucidated by analyzing the intact protein.

In this work, *S. oneidensis* MR-1 cell lysate was fractionated using weak anion exchange chromatography (WAX) and putative intact proteins were detected using on-line RPLC-FTICR-MS. Tentative identifications for intact proteins were made by searching the putative protein masses against the proteins identified in the corresponding peptide level analysis of aliquots of the same fraction (digested using trypsin and analyzed by RPLC/ion trap (IT)-MS/MS). Modifications allowed in the protein search were oxidation, N-terminal methionine loss and predicted signal peptide loss [1,2].

For the purpose of comparative proteomics, 2D displays for cell lysate fractions were combined to generate a master display that can be used for profiling variations in protein abundances between control samples and those subjected to biological perturbation. In the next stage, detection of proteins/modified proteins of biological interest will be followed by targeted MS/MS characterization using gas phase dissociation techniques (e.g. CID, ECD).

Methods

S. oneidensis cells were disrupted by bead beating with 0.1 mm zirconia/silica beads. The raw lysate was clarified by ultracentrifugation to protect the chromatographic columns from clogging.

WAX chromatography at pH 7.0 was accomplished using a PolyWAX LP column. Fractions collected were dialyzed to remove excess salt and combined to give six fractions. Intact protein RPLC was performed at 8000 psi on a 80 cm × 75 µm column packed in-house with Phenomenex Jupiter particles (C5 stationary phase; 5 µm particle diameter; 300 Å pore size). Mobile phase A was 0.1% trifluoroacetic, 25% acetonitrile, 74.9% water while mobile phase B was 0.1% trifluoroacetic acid, 9.9% water, 90% acetonitrile. A similar system with C18 particles was used for bottom-up proteomics.[3]

Trypsin digests were analyzed using a Finnigan (San Jose, CA, USA) LTQ linear ion trap mass spectrometer. Intact protein MS analyses were performed on a modified Bruker 12T APEX-Q FTICR mass spectrometer that incorporates an ion funnel interface, quadrupoles for collisional focusing and ion pre-selection, a hexapole for external ion accumulation, and RF only ion guide for transferring ions to the infinity cell.

Experimental Scheme

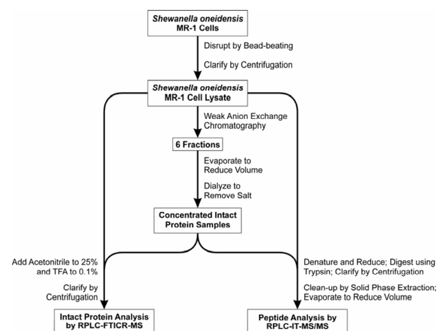

Figure 1: Experimental workflow outlining preparative steps for generating samples for RPLC/FTICR-MS intact protein and RPLC/IT-MS/MS peptide analysis.

Data Analysis

Peptides were identified from the RPLC-IT-MS/MS analysis using SEQUEST and were filtered as suggested by Washburn et al [4]. These were used to generate lists of proteins for the cell lysate and each WAX fraction.

Intact protein mass spectra were processed using in-house developed software to generate lists of neutral monoisotopic masses. The resulting monoisotopic masses were 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 in order 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 2 shows the graphs obtained by varying the clustering parameters used for defining UMCs (isotope fit < 0.15, 5 to 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).

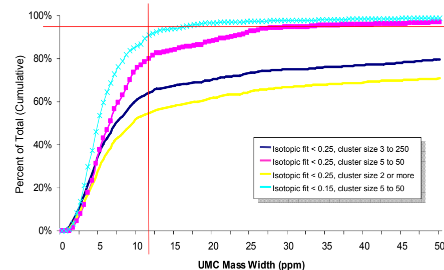

Figure 2: To estimate mass measurement accuracy for an LC-MS dataset we analyzed the mass variation of the UMCs. Two estimators were evaluated; one based on a conservative 95-percentile cut-off of outlier UMCs and the other based on a slope of the rate of change of the UMC mass variation distribution function (i.e., first derivative of the UMC mass variation distribution function). While the latter method appears advantageous (since it is invariant of the clustering parameters), the former method (95% cut-off) yields similar results for highly optimized UMCs.

Results

2D displays reconstructed from RPLC-FTICR-MS analysis

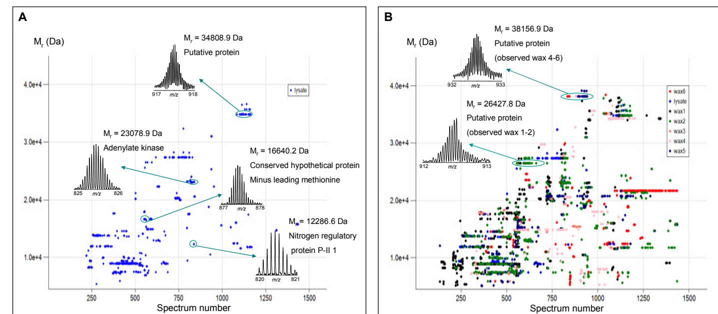

Figure 3: (A) 2D display reconstructed from the RPLC-FTICR-MS intact protein analysis of unfractionated *S. oneidensis* lysate. (B) 2D display for RPLC-FTICR-MS intact protein analysis of unfractionated *S. oneidensis* lysate combined with WAX fractions, reflecting the increased detection of putative intact proteins following fractionation of the cell lysate. (The number of putative intact proteins increased from 195 to 715 following the fractionation of cell lysate). The UMCs common to different fractions and the lysate were used to align the individual 2D displays using in-house developed software. The combined 2D display allows easy access to information regarding the WAX fraction and LC retention time associated with each protein.

Summary of results

Table 1: Summary of the results for proteins (M_r < 40 kDa) identified in RPLC-IT-MS/MS peptide analysis; UMC clusters detected and tentatively identified in RPLC/FTICR-MS intact protein analysis. 195 putative proteins were detected in the whole cell lysate alone while 715 putative intact proteins were detected in the whole cell lysate plus six WAX fractions. As expected, the number of tentative identifications for intact proteins also increased from 27 to 84 following the fractionation of the cell lysate. Overall, the results reflect better proteome coverage in the context of identified proteins for the bottom up approach, but also indicate the ability of FTICR-MS to detect proteins in more than one modification states.

Shewanella oneidensis sample	Multipeptide (≥ 2) proteins (RPLC/IT-MS/MS) less than 40kDa	UMC clusters (RPLC/FTICR-MS)	Tentative protein identifications (RPLC/FTICR-MS) (15 ppm MMA)
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 unique proteins	**447**	**715**	**84**

Examples showing the identification of modified proteins

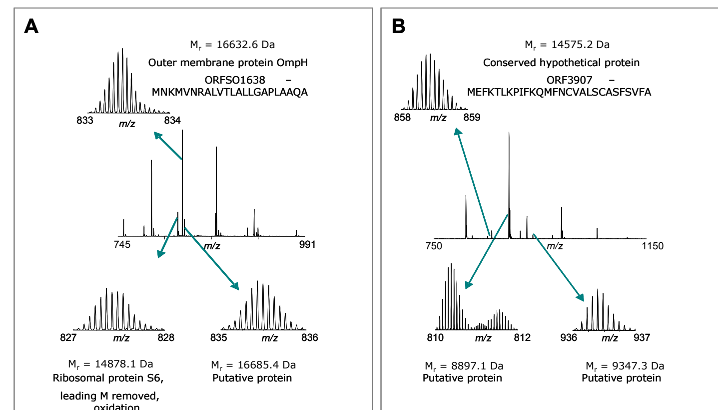

Figure 4: Examples of identified modified intact proteins in the RPLC-FTICR-MS analysis of a cell lysate. (A) Mass spectrum associated with the identification of SO1638 outer membrane protein OmpH. (B) Mass spectrum associated with the identification of SO3907 conserved hypothetical protein. For both examples, tryptic partial peptides supporting signal peptide losses were detected, while no peptides constituting the predicted signal peptide sequence were observed in the peptide MS/MS analyses. Signal peptide losses were observed for 17 of the 84 intact proteins tentatively identified through the entire experiment. (Signal peptide losses were predicted by the SignalP version 3.0 neural network method [1,2] for 95 of the 447 proteins identified in the corresponding bottom up analysis.)

Applications to comparative proteomics studies

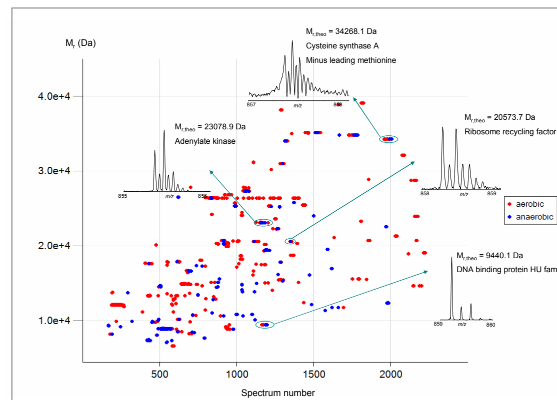

Figure 5: Alignment of the 2D displays for UMCs detected in the RPLC-FTICR-MS analysis of *S. oneidensis* lysate from cells grown on ^{13}C, ^{15}N depleted media under aerobic conditions (shown in red) and sub-oxic conditions (shown in blue). 196 putative proteins were detected (29 tentatively identified) in the lysate obtained from cells grown under aerobic conditions; 120 were detected (18 tentatively identified) in the lysate obtained from cells grown under sub-oxic conditions; 48 putative proteins and 7 tentatively identified proteins were found to be common between the two profiles.

Conclusions and Future Directions

❖ WAX fractionation and RPLC/FTICR-MS analysis allowed the detection of 715 putative intact proteins in the *S. oneidensis* cell lysate. Using peptide level identifications from each sample as a control, tentative identifications were established for ~10% of the putative intact proteins. Protein modifications considered for the tentative identifications were: oxidation, N-terminal methionine loss and predicted signal peptide loss.

❖ Protein profiles were presented as 2D displays with results from *S. oneidensis* cell lysate combined with the those from the WAX fractions. The combined 2D display allows easy access to information regarding the fraction number and LC retention time for each detected protein and enables subsequent targeted LC-MS/MS analyses.

❖ On comparing profiles generated for *S. oneidensis* cell lysate grown on ^{13}C, ^{15}N depleted media under aerobic vs. sub-oxic conditions, 48 putative proteins were found to be common (with 7 tentative identifications). In addition, cultivating cells in rare isotope depleted media allows for improved accuracy, sensitivity, dynamic range and detection limits.

❖ As a next step, novel targeted LC-MS/MS will be used to characterize the most interesting LC-MS features detected in the initial comparative analysis.

❖ Improvements in deisotoping algorithms, cell design and implementation of automated gain control (AGC) are expected to enhance the utility of proteomics measurements at the intact protein level.

Acknowledgements

This research was funded by the NIH National Center for Research Resources (RR18522). We gratefully acknowledge Dwayne Elias for providing the cell samples, Anil Shukla and Rui Zhang for help with instrumentation, and Rui Zhao for help 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, Washington. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under Contract No. DE-AC05-76RLO 1830.

References

1. Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., Protein Eng. 1997, 10, 1-6.
2. Bendtsen, J. D., Nielsen, H., von Heijne, G., Brunak, S., J. Mol. Biol. 2004, 340, 783-795.
3. Shen, Y., Zhao, R., Belov, M. E., Conrads, T. P., Anderson, G. A., Tang, K., Paša-Tolić, L., Veenstra, T. D., Lipton, M. S., Udseth, H. R., Smith, R. D., Anal. Chem. 2001, 73, 1766-1775.
4. Washburn, M. P., Wolters, D. A., Yates, J. R., III, Nat. Biotechnol., 2001, 19, 242-247.

Contact Information

Ljiljana Paša-Tolić, Ph. D.
Biological Sciences Division, K8-98
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
e-mail: ljiljana.pasatolic@pnl.gov
Phone: 509 376 8859