Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

Proteome profiling using an integrated top-down and bottom-up strategy

Si Wu[1], Errol W. Robinson[1], Natacha Lourette[1], Nikola Tolić[1], Rui Zhao[1], Nathan P. Manes[1], Robert A. Maxwell[1], Ryan D. Estep[2], Scott W. Wong[2], Joshua N. Adkins[1], Richard D. Smith[1], and Ljiljana Paša-Tolić[1]
[1]Pacific Northwest National Laboratory, Richland, WA
[2]Vaccine and Gene Therapy Institute, Oregon Health & Science University, Beaverton, OR

Characterization of protein modification states

Overview

Objective: Comprehensive high-throughput characterization of protein modification states, coordination and dynamics of posttranslational modifications (PTMs)

Methodology

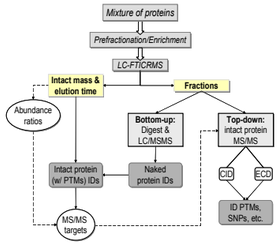
- Intact protein RPLC-FTICR
  - High resolution RPLC
  - 12T FTICR with a novel ICR cell for high resolution and MMA
- Off-line digestion for tentative ID of proteins and PTMs
- Conformation of identifications and localization of PTMs by targeted off-line intact protein MS/MS

Applications

- Identification and quantitation of protein isoforms
- Profiling proteomes at the intact protein level
- Protein degradation pathways

Introduction

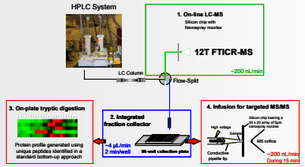
Bottom-up proteomics methods[1] based on the analysis of digested proteins have frequently been used to search for PTMs, but incomplete sequence coverage and lost information on the coordination of modifications are significant obstacles. Top-down[2] proteomics (subjecting intact protein ions to gas-phase dissociation) is well-suited to PTM-discover, but currently incompatible with high throughput analysis. In response to these obstacles, we developed an integrated top-down and bottom-up approach[3,4] facilitated by concurrent LC-MS analysis and fraction collection.



The approach employs high resolution reversed phase LC separations coupled on-line with a 12T Fourier transform ion cyclotron resonance spectrometer to tentatively identify modified proteins, using detected intact protein masses in conjunction with bare protein identifications from the bottom-up data of the same fraction. These identifications are incorporated into a target ion list for subsequent offline gas phase fragmentation (CID and ECD) that uses only an aliquot of the original fraction used for bottom-up analysis.

Methods

Instrumental design for concurrent LC-MS and fractionation



RPLC-FTICR MS

Intact protein MS analyses were performed on a modified Bruker 12T APEX-Q FTICR mass spectrometer that incorporates an ion funnel interface, and various ion transfer, selection and storage multipoles.

Intact protein RPLC was performed at 10,000 psi on a 80 cm × 200 µm column packed in-house with Phenomenex Jupiter particles (C5 stationary phase; 5 µm particle diameter; 300 Å pore size). Mobile phase A was 0.05% trifluoroacetic acid, 0.2% acetic acid, 25% acetonitrile, 5% isopropanol, and 69.75% water while mobile phase B was 0.05% trifluoroacetic acid, 9.95% water, 45% acetonitrile, and 45% isopropanol.

Fractions collected using Triversa Nanomate (Advion Biosciences, Inc.) were reconstituted using sheath solvent (2% acetic acid in 49% aqueous methanol), and analysed using CID (in external accumulation hexapole) and/or ECD.

Bottom-up analysis

50 mM NH_4HCO_3 in 30% (v/v) aqueous acetonitrile was used as the digestion buffer (pH=8.2). Digestion was performed at 37°C overnight. Samples on the plate were evaporated to remove extra organic solvent. The final volume was then adjusted to 15 µl by adding mobile phase A. A Thermo Fisher Scientific LTQ ion trap was used for LC-MS/MS analysis.

Sample preparation

- Mixture of standard proteins: ubiquitin, cytochrome *c*, β-casein, β-lactoglobulin A, β-lactoglobulin B, carbonic anhydrase, myoglobin, and camodulin.
- Proteasome: A catalytic mutant of yeast proteasome was expressed and purified; trypsin-like and caspase activities were knocked out and only chymotrypsin like activities remained.
- Vaccinia: Purified virus (in 1mM Tris Cl, pH 8) was mixed with 8M Urea and 5mM DTT, heated at 55°C for 30 min, and desalted by passing through PD-10 column.

Results

Proof of principle demonstration

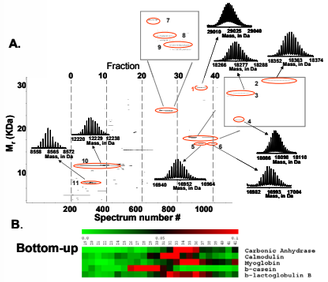


Table 1. Summary of results for standard protein mixture analysis using the integrated approach

Spot in Fig1	Measured MW(Da)	Protein	Modifications	Eluted Fractions	MMA (ppm)
1	29024.7313	carbonic anhydrase II	Removal of N-term Met, N-acetylation	[33,36]	0.01
2	18363.4561	ß-lactoglobulinA	2 disulfide bonds	[35,37]	0.13
3	18277.4323	ß-lactoglobulinB	2 disulfide bonds	[32,34]	0.84
4	18097.4866	Calmodulin	with His-tag	[31,33]	0.86
5	16950.9657	myoglobin	Intact protein	[34,36]	-0.37
6	16993.0104	myoglobin	Lys-acetylation	42	0.46
7	24092.3436	ß-casein, Isoform3	5 phosphorylations, P67H(SAAP) and S122R (SAAP) to Isoform1	26	3.22
8	24023.2204	ß-casein, Isoform2	5 phosphorylations, P67H(SAAP) to Isoform1	27	0.97
9	23983.2346	ß-casein, Isoform1	5 phosphorylations	[28,30]	1.82
10	12229.2211	cytochrome C	Removal of N-term Met, N-acetylation, with Heme	[3,12]	0.15
11	8564.6471	ubiquitin	Intact protein	[4,7]	1.97

Figure 1. Top-down and bottom-up analyses of a standard protein mixture. **A)** 2D display reconstructed from the LC-FTICR data. **B)** Protein elution patterns generated using tryptic peptides identified in each fraction. Tentative identifications of proteins and modified proteins were accomplished by matching bottom-up data with the measured intact protein masses (**Table 1**).

Structural characterization of β-casein

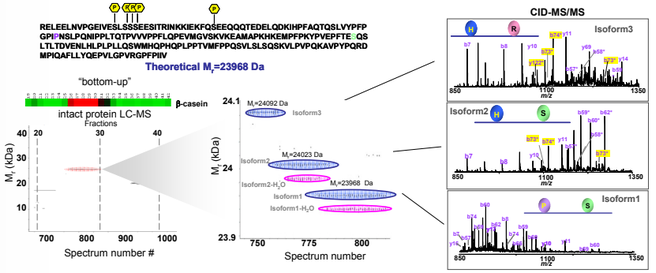
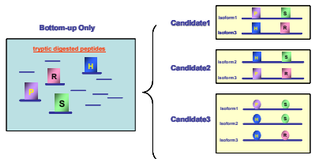


Figure 2. Integrated approach identifies multiple SNPs for bovine β-casein. We identified three phosphorylated isoforms of β -casein that differed in mass by 40 Da and 109 Da, and were assigned as SNP mutants with a Pro to His switch and an additional Ser to Arg replacement, respectively. However, the bottom-up approach, even with 100% sequence coverage, cannot validate the presence of three distinctive isoforms (i.e., distinguish between one isoform with two SNPs and two isoforms each with a single SNP).



Probing interaction partners and substrates of yeast proteasome

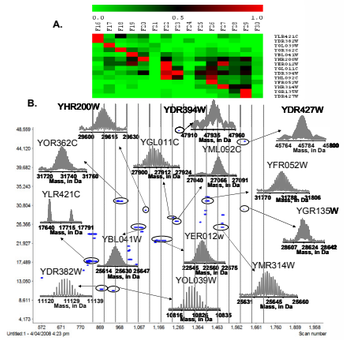


Table 2. Summary of results for yeast proteasome analysis using the integrated approach

ID	UniProt #	Function	Notes	M_{theo}	M_{exp}	MMA (ppm)
1	YLR421C	26S proteasome regulatory subunit RPN13	1 acetylation,1 phosphorylation, loss of N-term 3 residues	17674.62	17674.58	-1.97
	YLR421C	26S proteasome regulatory subunit RPN13	1 acetylation,2 phosphorylation, loss of N-term 3 residues	17754.58	17754.55	-1.97
2	YDR382W	60S acidic ribosomal protein P2-beta	1 phosphorylation	11129.36	11129.34	-1.74
3	YOL039W	60S acidic ribosomal protein P2-alpha	1 phosphorylation	10825.22	10825.19	-2.77
4	YOR362C	Proteasome component C1	loss of N-term Met, Blocking of N-term Thr	N/A	31740.47	N/A
5	YBL041W	Proteasome component C5	loss of N-term 12 residues (signal peptide)	25628.92	25628.91	-0.43
6	YHR200W	26S proteasome regulatory subunit RPN10	loss of N-term Met	29615.84	29615.77	-2.51
7	YER012W	Proteasome component C11	1 acetylation	22558.62	22558.60	-0.92
8	YGL011C	Proteasome component C7-alpha	loss of N-term Met, 1 acetylation	27911.20	27911.16	-1.63
9	YDR394W	26S protease regulatory subunit 6B homolog	1 acetylation	47935.21	47935.19	-0.40
10	YML092C	Proteasome component Y7	loss of N-term Met, 1 acetylation	27072.08	27072.06	-0.51
11	YFR052W	26S proteasome regulatory subunit RPN12	loss of N-term Met	31787.19	31787.16	-0.89
12	YMR314W	Proteasome component PRE5	1 acetylation	25645.24	25645.24	-0.19
13	YGR135W	Proteasome component Y13	loss of N-term Met, 1 acetylation	28624.59	28624.57	-0.71
14	YDR427W	26S proteasome regulatory subunit RPN9	No Modification	45782.85	45782.87	0.61

Figure 3. Top-down and bottom-up analyses of yeast proteasome. **A)** Protein elution patterns generated using tryptic peptides identified in each fraction. **B)** 2D display reconstructed from the LC-FTICR data. Insets show representative neutral mass spectra. Tentative identifications of proteins and modified proteins were accomplished by matching bottom-up data with the measured intact protein masses (**Table 2**).

Intact vaccinia viral proteome profiling

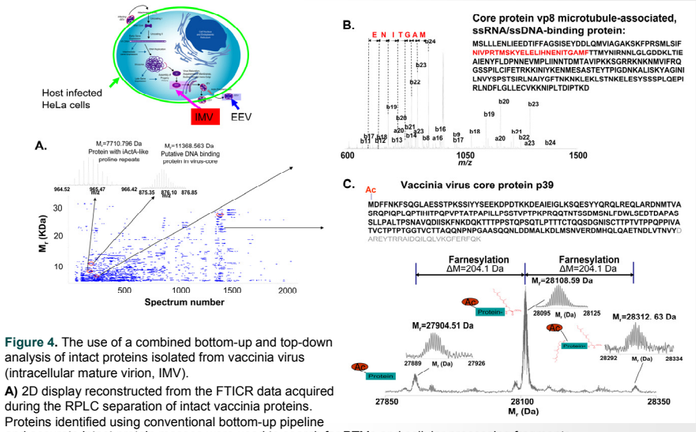


Figure 4. The use of a combined bottom-up and top-down analysis of intact proteins isolated from vaccinia virus (intracellular mature virion, IMV).

A) 2D display reconstructed from the FTICR data acquired during the RPLC separation of intact vaccinia proteins. Proteins identified using conventional bottom-up pipeline and accurate intact protein masses were used to search for PTMs and cellular processing fragments.
B) These tentative protein identifications can be confirmed using offline MS/MS. **C)** The vaccinia virus "39 kDa protein" was found to be farnesylated at differing degrees (none, one, and two farnesylations). This observation suggests a previously unrecognized mechanism of membrane association for the vaccinia virus, which could potentially become a new drug target.

Conclusions

- Characterization of PTMs (e.g., acetylation, phosphorylation) and genetic variants (e.g., SNP isoforms) using the integrated top-down and bottom-up approach was demonstrated with a mixture of standard proteins. Importantly, we validated the unique ability of this approach to identify (and potentially quantify) distinctive protein isoforms (e.g., distinguish between one isoform with two SNPs and two isoforms each with a single SNP in the case of β-casein).
- ~100 putative proteins (M_r>10 kDa) were detected in the intact protein RPLC-FTICR analysis of the yeast proteasome. The integrated approach revealed numerous PTMs, including phosphorylation, methylation, acetylation, and predicted/known signal peptide loss.
- Intact protein RPLC/FTICR-MS analysis allowed detection of ~600 putative proteins (M_r>5 kDa) in the vaccinia virus lysate. The vaccinia virus "39 kDa protein" was found to be farnesylated at differing degrees (none, one, and two farnesylations). This observation suggests a previously unrecognized mechanism of membrane association for the vaccinia virus, which could potentially become a new drug target. The identification of protein PTMs and proteolytic fragments could provide clues to the mechanisms of infection and potential targets for therapeutics.

Acknowledgements

Portions of this research were funded by the NIH National Center for Research Resources (RR18522; RDS) and the National Institute of Allergy and Infectious Diseases, (NIH/DHHS through interagency agreement Y1-AI-4894-01). The authors thank Aleksey Tolmachev, Robby Robinson, Anil Shukla, Yufeng Sheng and Rui Zhang for contributing to the improvement of instrumental capabilities and performance.

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. Aebersold R *et al.*,Mass spectrometry-based proteomics (2003) *Nature*, 6:602-610.
2. Kelleher NL, Top-down proteomics (2004) *Anal. Chem. 76:197A-203A*.
3. Sharma S *et al.*, Proteomic profiling of intact proteins using WAX-RPLC 2-D separations and FTICR mass spectrometry (2007) *J. Proteome Res.* 6:602-610.
4. Smallwood H *et al.*, Identification of a Denitrase Activity in Activated Macrophages Using High-Field LC-FTICR Mass Spectrometry, (2007) *Biochem.* 46:10498-10505.

Contact Information

Si Wu, PhD
Biological Sciences Division
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
e-mail: si.wu@pnl.gov