

# From Sample to Spray: High Performance Workflow for Top-Down Protein Analysis

<sup>1</sup>Chuck Witkowski, <sup>1</sup>Jeremy Norris, <sup>2</sup>Gary Valaskovic, <sup>3</sup>Top-Down Development Team, <sup>3</sup>Neil Kelleher

<sup>1</sup>Protein Discovery, Inc., Knoxville, TN; <sup>2</sup>New Objective, Inc., Woburn, MA;  
<sup>3</sup>Center for Top-Down Proteomics, University of Illinois, Champagne-Urbana

## Introduction

A new workflow for obtaining high-quality mass spectrometric data on intact proteins from complex biological samples has been developed. The workflow includes sample fractionation using parallel Gel Eluted Liquid Fraction Entrapment Electrophoresis (GELFrEE), advanced polymer phase nanobore liquid chromatography, and high-sensitivity nanoelectrospray ionization coupled to FTMS. This combination provides a powerful front-end solution to generate the highest quality top-down data. When coupled with the most advanced data analysis tools, these tools comprise the most powerful strategy for top-down proteomics and the most comprehensive analyses generated to date.

For fractionation of intact proteins by molecular weight (MW), a sharply improved two-dimensional (2D) separation is presented to drive reproducible and robust fractionation before top-down mass spectrometry of complex mixtures. GELFrEE is implemented by use of tris-glycine and tris-tricine gel systems applied to human cytosolic and nuclear extracts from HeLa S3 cells, to achieve a MW-based fractionation of proteins from 5 to 100 kDa in 1 hour. For top-down tandem mass spectrometry (MS/MS) of the low-mass proteome (5–25 kDa), between 5 and 8 gel-elution GELFrEE fractions are sampled by packed emitter nanospray-LC-MS/MS with 12 or 14.5 tesla Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers. Single injections provide 30-50 detectable proteins, approximately half of which yield automated identifications using the ProSight software platform. Reproducibility metrics of the system are presented, along with comparative analysis of protein targets in mitotic-versus-asynchronous cells. We forward this basic 2D approach to facilitate wider implementation of top-down mass spectrometry and a variety of other protein separation and/or characterization approaches.

**FIGURE 1** WORKFLOW



HeLa Cell Lysate



GELFrEE  
MW-based separation

**Sample Cleanup  
(SDS Removal)**

**ProSightPC™ 2.0**  
Software for Precision Proteomics

Database Search



Online nanoLC MS/MS 12T or 14.5T LTQ-FTMS



Digital Nanospray  
Source

Work flow for Top-Down proteomics. Total protein content from HeLa-S3 nuclei or cytosol is quantified and loaded onto a GELFrEE column. The GELFrEE device separates the protein samples according to molecular weight (MW). Proteins of increasing molecular MW elute into solution-phase fractions.

The solution-phase fractions are cleaned up to remove SDS before re-suspension and injection onto a PicoFrit nano-LC column for MS/MS analysis with a LTQ-FT at either 12 T or 14.5 T. LC-MS/MS files are processed with ProSightPC 2.0, a software suite tailored for Top-Down analysis in a high-throughput setting.

## Methods & Materials

### Samples

- Protein Standards (Sigma) in 0.2% formic acid
- Cytosolic lysate, undigested, reduced, alkylated from in-house cell culture of HeLa S3 (ATCC) cell line

### GELFrEE

- Programmable electrophoretic controller and 8 channel GELFrEE cartridges (Protein Discovery)
- 100  $\mu$ L aliquot of lysate, loaded (0.1% SDS)
- 16 fractions (150  $\mu$ L each), collected over 1 hour
- Aliquots from each fraction were characterized by 1D SDS-PAGE
- SDS removal (post GELFrEE)
- Protein precipitation (cold acetone) of each fraction followed by centrifugation of pellet
- Re-suspension in 40  $\mu$ L of 5% ACN, 0.2% formic acid

### HPLC

- 1D-plus NanoLC (Eksigent)
- A=5% ACN, 0.2% formic acid; B= 95% ACN, 0.2% formic acid
- 300 nL/min. Gradient: 20% B @ 10 min., 60% B @ 45 min., 95% B @ 55 min.
- 10  $\mu$ L autosampler injection
- PicoFrit packed-emitter column (New Objective)
- Porous polymer media (5  $\mu$ m, 300 or 1,000 A), 75  $\mu$ m ID x 10 cm
- 15  $\mu$ m ID emitter

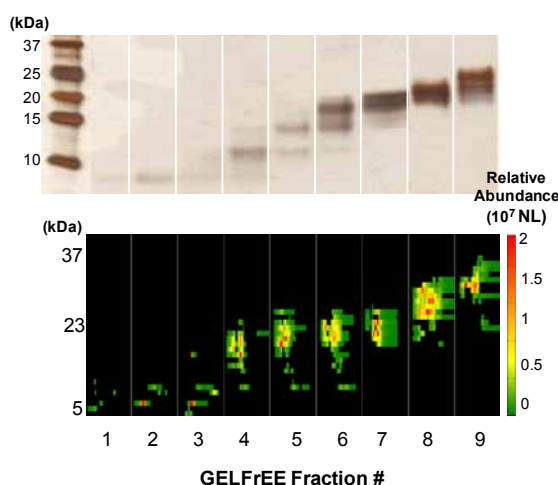
### Mass Spectrometry

- LTQ-FT Ultra (Thermo Fisher) with 12 tesla magnet
- "Low-high" format; full scan acquisition in linear trap, data dependant (top two) MS/MS acquired at high resolution (180,000 RP). Dynamic exclusion with a repeat count of 1, set for 5,000 sec.
- Digital PicoView nanospray source (New Objective)
- Automated emitter rinsing between LC runs
- ESI voltage: 2.5 - 2.8 kV typical

### Data Analysis

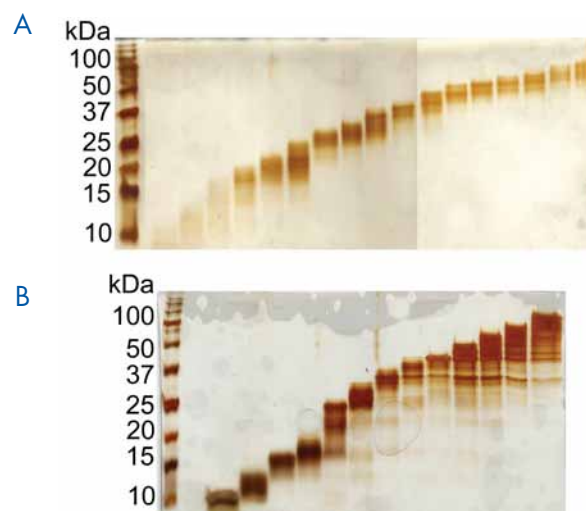
- ProSight PC 2.0 (Thermo Fisher)
- 10 Dalton precursor window
- 10 ppm fragment mass tolerance

**FIGURE 2**



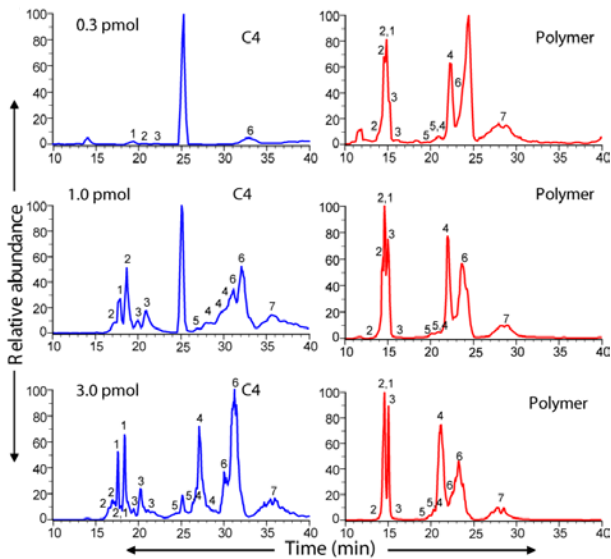
Analytical slab gel and heat map showing masses detected in GELFrEE fractions. The silver-stained slab gel (top) of solution phase GELFrEE fractions shows typical GELFrEE separation (tris-tricine buffer system). After LC-MS of the GELFrEE fractions, the detected masses for each fraction are plotted as a function of LC retention time (bottom). The MS relative abundance is designated with the scale on the right. General correlation can be seen between detected masses and abundances from the MS and the intensities of protein bands detected on the slab gel.

**FIGURE 3**



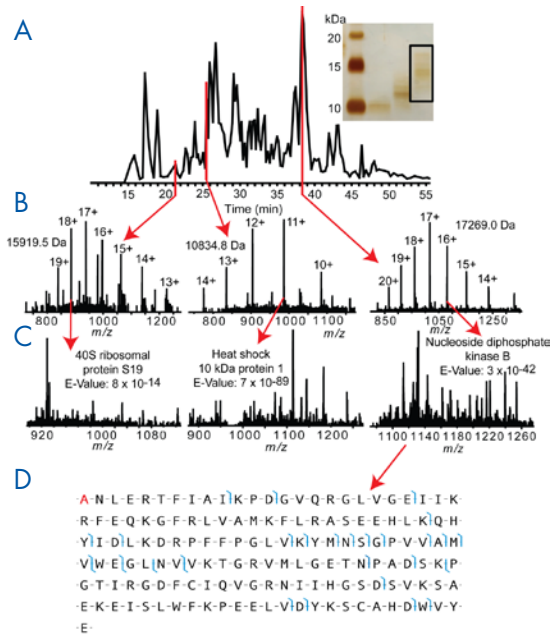
Tris-tricine and tris-glycine GELFrEE runs displayed via slab gels. In A, a slab gel visualization of a tris-tricine GELFrEE fractionation (400  $\mu$ g loaded) is displayed. In B, a slab gel visualization of a tris-glycine GELFrEE fractionation (200 mg loaded) is shown. The tris-tricine run shows a larger number of fractions at lower molecular weight. The samples were nuclear lysates from HeLa-S3 cells.

**FIGURE 4**



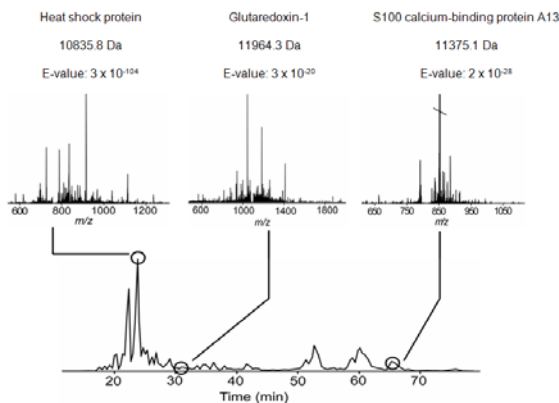
Chromatographic performance of C4 and Polymer packed PicoFrit columns for intact protein analysis. A) Chromatograms for RPLC separation of a seven protein standards mixture of varying concentrations, using C4 (blue) and Polymer (red) (75  $\mu$ m I. D. x 100 mm, 300 A, 5  $\mu$ m stationary phase) PicoFrit columns. Standards used were: 1. Ubiquitin; 2. Cytochrome c; 3.  $\beta$ -lactalbumin; 4. Myoglobin; 5.  $\beta$ -casein; 6. Carbonic Anhydrase; 7. Ovalbumin. Total protein load is shown.

**FIGURE 5**



Examples selected from an LC-MS/MS injection of fraction #3 from a tris-glycine GELFrEE run. A base-peak chromatogram is shown (A), with charge state distributions from selected retention times shown in B. In panel C, abundant charge states (above the arrows) were targeted for fragmentation. Fragmentation mass spectra for each protein are shown along with the corresponding identifications and E-values. A fragmentation map (D) results from the matching fragment ions of nucleoside diphosphate kinase B. The protein is N-terminally acetylated.

**FIGURE 6**



A representative LC-MS/MS run generated from data-dependent (top 2) acquisition. The sample infused was GELFrEE fraction #5 from fractionation of a HeLa nuclear lysate. Data were from a 12 tesla LTQ FT instrument, with identifications resulting from ProSightPC 2.0. All identified proteins are shown in Table 1 (Figure 7).

**Acknowledgements**

The laboratory of NLK was supported by the Packard Foundation, the Sloan Foundation, the National Institutes of Health (GM-067193-07), the UIUC Neuroproteomics Center on Cell-Cell Signaling, and the National Institute on Drug Abuse (P30 DA-018310-06)

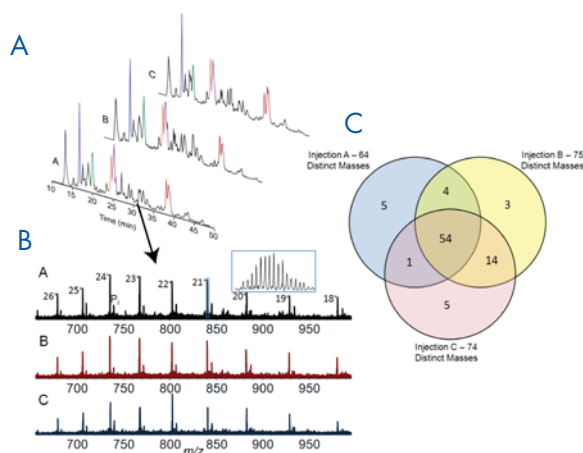
FIGURE 7

Protein Description	Observed Mass (Da)	Calculated Mass (Da)	Mass Difference (ppm)	PTMs	Matching Fragments	E-value	-10 x log (E-value)
Migration-inducing protein 12	4933.5	4933.5	0.2	N-terminal acetylation	32	$1 \times 10^{-35}$	360
40S ribosomal protein S30	6643.9	6643.8	0.1		27	$4 \times 10^{-34}$	334
Metallothionein-2	6932.6	7046.6	N/A*	N-terminal acetylation	16	$5 \times 10^{-19}$	183
Small EDRK-rich factor 2	6953.6	6937.7	N/A*	N-terminal acetylation	9	$1 \times 10^{-6}$	60
Coiled-coil domain-containing protein 72	6972.9	6972.9	-1.1		13	$3 \times 10^{-21}$	205
40S ribosomal protein S28	7935.2	7935.2	0.3	N-terminal acetylation	29	$5 \times 10^{-46}$	453
60S ribosomal protein L38	8138.7	8138.7	-4.5		8	$4 \times 10^{-5}$	74
Oncogene DJ-1 (fragment)	8264.3	8264.3	-0.6		8	$3 \times 10^{-10}$	95
Ubiquitin (fragment)	8845.6	8845.6	0.7		61	$6 \times 10^{-96}$	952
GAPDH (fragment)	8855.6	8855.6	0.5		13	$9 \times 10^{-16}$	159
1-AGP acyltransferase 2 (fragment)	9252.8	9252.8	11.2		9	$1 \times 10^{-6}$	60
ATPase inhibitory factor 1 (fragment)	9510.9	9510.9	0.5		16	$2 \times 10^{-9}$	87
S100 calcium-binding protein A6	10,141.3	10,141.3	1.4	N-terminal acetylation	10	$3 \times 10^{-10}$	95
U6 snRNA-associated SM-like protein LSm8	10,307.3	10,307.3	0.8	N-terminal acetylation	44	$2 \times 10^{-79}$	797
Dynein-associated protein Km23	10,825.7	10,825.7	1.1	N-terminal acetylation	10	$3 \times 10^{-13}$	125
Heat shock 10-kDa protein	10,835.7	10,835.8	-5.7	N-terminal acetylation	79	$3 \times 10^{-104}$	1035
S100 calcium-binding protein A13	11,375.1	11,375.1	0.6	N-terminal acetylation	22	$2 \times 10^{-28}$	277
S100 calcium-binding protein A11	11,757.8	11,757.8	1.3	N-terminal acetylation	29	$2 \times 10^{-17}$	167
S100 calcium-binding protein A16	11,761.0	11,762.0	12.3	N-terminal acetylation	21	$2 \times 10^{-32}$	317
SUMO-1 specific protease 2 (fragment)	11,785.7	11,785.2	32.3	N-terminal acetylation	5	$6 \times 10^{-6}$	52
S100 calcium-binding protein A4	11,794.7	11,859.8	N/A*	N-terminal acetylation	9	$5 \times 10^{-6}$	53
Thioredoxin	11,883.8	11,883.8	-0.5	Methylation	11	$1 \times 10^{-15}$	150
Glutaredoxin-1	11,964.3	11,964.3	1.0	Methylation	15	$6 \times 10^{-14}$	132
Phenylpyruvatetautomerase	12,508.2	12,508.2	1.3		25	$3 \times 10^{-34}$	335
MIF protein	12,538.3	12,551.2	N/A*		8	$1 \times 10^{-5}$	50

\* The observed protein forms were not present in the database; therefore, large mass discrepancies exist between the observed and calculated forms

Proteins identified from the run shown in Figure 6.

FIGURE 8



LC-MS reproducibility. Base-peak chromatograms from technical triplicate GELFrEE-LC-MS injections are shown in A. The largest peaks have been colorized to highlight reproducibility. The intact protein detected at retention time ~30 min is shown for each injection in panel B. Similar charge state distributions are observed for the protein in each injection. A Venn diagram (not to scale) showing overlap of distinct detected masses is presented in C.

## Conclusions

- For fractionation of intact proteins, GELFrEE is robust and general
- GELFrEE enables high loading capacity and a fast (1 hr.) fractionation of complex samples
- Polymer media, packed-emitter nanobore columns offer high performance separations of the low MW proteome (<25 kD)
- 30-50 distinct protein forms are observed in a single LC-MS/MS run
- ProSight PC 2.0 enables the automated identification of intact proteins, including the determination of post-translational modifications
- Extension of LC to higher mass proteins is under development

Originally published at the 2009 Chemical and Pharmaceutical Structure Analysis Symposium in Langhorne, PA

© 2009 New Objective, Inc. All rights reserved.  
All trademarks and registered trademarks are properties of their respective companies.