Use of Multidimensional Chromatography and Automated Data Processing to Characterize Complicated Protein Mixtures

Scott J. Berger, Hongji Liu, Robert S. Plumb and Steven A. Cohen, Life Sciences R&D, Waters Corporation, 34 Maple St., Milford, MA 01757

ABSTRACT

The analysis of intact protein mixtures by ESI-MS methods is a powerful technique for characterizing cellular complexes and the modification state of component subunits. In this work, yeast ribosomes (approximately 100 protein components) were used as a model complex that was analyzed using multidimensional chromatography (SCX/RPLC) coupled with both offline fraction collection and online mass detection of the intact proteins using an ESI-TOF mass spectrometer. The system is particularly interesting as the extreme basic pI of most ribosomal components prohibits standard 2D gel analysis techniques. The high mass measurement accuracy of the ESI-TOF analyzer permitted tentative identifications of ribosomal subunits, while collected fractions were used to probe the nature of modifications inferred from protein intact mass. In this poster we will also describe the automated application of TIC peak identification and maximum entropy charge state deconvolution algorithms that permitted efficient identification of sample components with minimal human oversight. The potential of this approach for proteomic scale protein and complex analysis will be discussed.

Instruments and Conditions

- Ion exchange
  - Column: Shodex® SP 420N (4.6 x 35 mm, non-porous)
  - Eluent A: 50 mM methylamine, 6M urea, 0.5 mM DTT, pH adjusted to 5.60 with acetic acid, 10% (v/v) acetonitrile
  - Eluent B: Eluent A plus 1M NaCl
  - Flow rate: 0.4 ml/min
  - Gradient: step gradient (0-90% B)

Time	Flow	%A	%B	Curve
0	0.4	100	0	*
20	0.4	92	8	1(Step)
40	0.4	84	16	1
60	0.4	80	20	1
80	0.4	76	24	1
100	0.4	72	28	1
120	0.4	68	32	1
140	0.4	64	36	1
160	0.4	60	40	1
180	0.4	10	90	1
200	0.4	100	0	1

- Reversed phase
  - Column: Symmetry300™ C_4, 3.5 µm (2.1 x 50 cm)
  - Eluent A: 0.1% TFA in water
  - Eluent B: 0.1% TFA in acetonitrile
  - Flow rate: 0.5 ml/min
  - Gradient (linear):
    10-60% B in 18 min (1D)
    20-50% B in 18 min (2D)

- ESI-TOF (Micromass® LCT)
  - Source: 3 kV (capillary), 60V (Sample cone), 3 V (extraction cone), 450 (RF lens)
  - Temperature (°C): 250 (desolvation), 130 (source)
  - Gas flow (L/hr): 100 (cone), 400 (desolvation)
  - MCP detector: 2700 V
  - Max Flight time: 65 µS

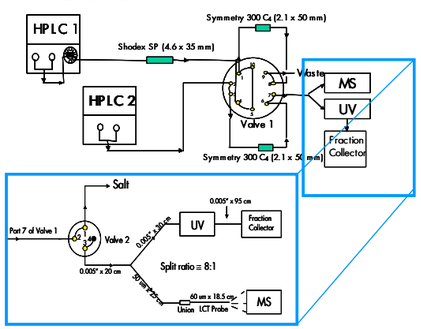


Figure 1: 2-D (SCX/RP) LC/MS system for separation of intact protein mixtures, with exploded diagram of the post-column split MS interface

Intact protein mixtures are applied on to a Shodex SP strong cation exchange column under acidic buffer conditions, and at low salt. Sequential salt steps were used to elute proteins onto alternating Waters Symetry300™ C4 reversed phase columns. Connecting the two columns through a 10-port 2-position regeneration valve permitted development of the RP gradient on the first RP column while loading the next salt fraction on the second RP column.

An automated 4-port 2-position valve is positioned post-column, and is used to divert residual salt present on the RP column during the first few minutes of a second dimension RP analysis. Following the diversion period, column effluent passes through a post column splitter (10% to MS, 90% to UV/fraction collection) which permits identification of both intact protein mass, and subsequent offline protein characterization with digested fractions.

Experimental Rationale

Analyzing intact proteins provides direct information about the extent of co-translational processing and post-translational modifications. However, without prior knowledge, direct identification of a protein (especially highly modified proteins) is often problematic, especially with larger organism databases.

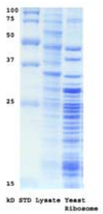
Identification of proteins in complicated mixtures, or mixtures of protein structures with high similarity requires significant resolution, and selectivity in a separation. Thus, a very high resolution 1-D RP-HPLC analysis may provide the theoretical peak capacity for a separation, but not the selectivity to accomplish component resolution. Orthogonal separation schemes provide the power to obtain high peak capacities while increasing selectivity by applying two modes of separation. With this work, our goals were to implement a 2-D chromatographic system that: permits:

- Complicated protein mixtures to be effectively separated
- Highly accurate intact protein mass analysis using an ESI-TOF MS
- Sample collection for confirmation of protein identity, and modification site identification

Yeast Ribsosomal Proteins

Large Subunit
81 ORFs (69 unique)
46 Subunits

Small Subunit
56 ORFs (47 unique)
32 Subunits



Yeast ribosomes were used as a model system to evaluate the 2-D LC/MS analysis system. Ribosomes were purified by differential centrifugation using a variation of published procedures (Ref. 1), and processed by acid extraction (Ref. 2) to remove RNA prior to loading on the SCX column. The above gel demonstrates the relative enrichment of ribosomal proteins relative to an equal loading (10 µg) of yeast cell lysate.

Yeast ribosomal proteins are a useful system for evaluation of the 2-D system due to the high complexity (78 subunits (3-43 kD), up to 116 total isoforms), variety of protein modifications (acetylation, methylations, phosphorylations, and proteolytic processing), and high similarity between many subunit isoforms.

Optimising the System

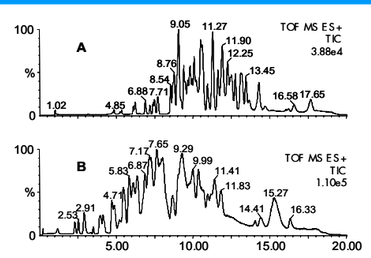


Figure 2: RP protein separations containing TFA (A) yield superior separations to formic acid (B) at the expense of roughly 3-fold MS sensitivity. Comparison of the 1-D RP chromatographic separation of ribosomal proteins (75 µg) using either (A) 0.1% TFA or (B) 2% formic acid as the acidic modifier. A Waters Symmetry300™ C_4 3.5 µm column was developed with a 10-60 %B gradient over 18 minutes and monitored by positive ion ESI-MS.

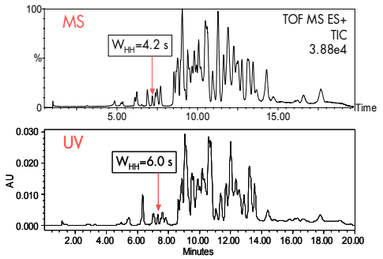


Figure 3: Efficient post-column connections are critical to maintaining chromatographic separations and system sensitivity. Although the post-split flow rate to the LCT ESI-TOF MS is only 10% of the effluent flow from the second dimension RP column, the use of narrow i.d. tubing and low dead-volume connections (Figure 1) minimized band-spread and yielded TIC peak widths narrower than observed with the UV detection/fraction collection pathway.

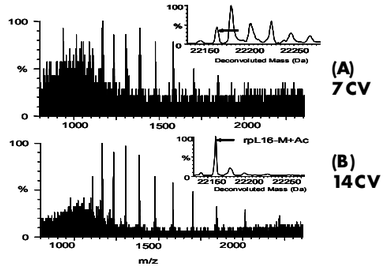


Figure 4: Diversion of residual non-volatile salt on RP from the SCX step elution is critical to obtaining useful MS information on intact proteins. Ribosomal proteins were analyzed by 2-D (SCX/RP) LC/ESI-TOF MS using optimized chromatographic conditions. Diversion of the first 7 column volumes **(A)** of the reversed phase gradient resulted in complex MS spectra as demonstrated with a component eluting at 113.2 minutes into the run (Arrow) with a series of adduct peaks. Increasing the diversion time to 14 column volumes **(B)** produced simplified mass spectra for this component (elution time= 113.4 min) now identifiable as yeast large ribosomal subunit rpL16B (acetylated and lacking the initiating methionine; expected mass 22160.1, observed mass 22160.5). Relative signal intensity is 2-fold higher for mass spectra (B vs. A), and 4-fold greater for the deconvoluted mass spectra (B vs. A, insets).

System Sensitivity

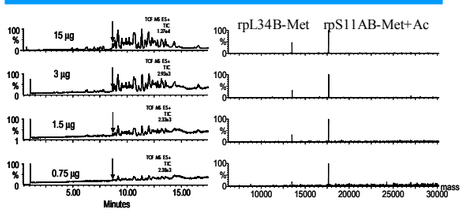


Figure 5: The system demonstrates lower fmol MS sensitivity for intact proteins. System sensitivity was established using injections of 15, 3.0, 1.5 and 0.75 µ ribosomal proteins onto a 1-D RP (50 x 2.1mm Waters Symmetry300™ C_4 column with optimized split/divert valve and TFA modifier. Detection was performed by positive ion ESI-MS-TOF scanning from 500 to 3000 m/z. A minor peak on the TIC (RT= 8.50 min) at the highest injection amount yielded interpretable MaxEnt1™ processed spectra at the lowest level injected. Given a post column split of 1:8, an assuming ~100 equimolar proteins with an average mass of 20 kD, we calculate a MS detection limit of less than 50 fmol per protein.

2-D LC/MS Data Analysis

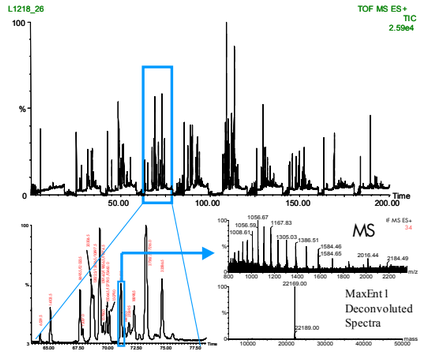
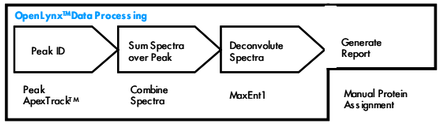


Figure 6: Processing 2-D LC/MS Intact Protein Data. The analysis of 75 µ of ribosomal proteins using the 2-D system produced 157 deconvoluted masses (126 unique masses) ranging from 6629 to 42731 Da. From these 126 unique masses, 91 have been assigned to 67 known ribosomal subunits (85 isoforms) with an average mass accuracy error of ~50 ppm (SD= 25 ppm). Assignments were made using translated yeast ORFs (+/- methionine, +/- acetylation), and known modifications compiled in the Yeast Protein Database (Ref. 3)

Automated Data Analysis



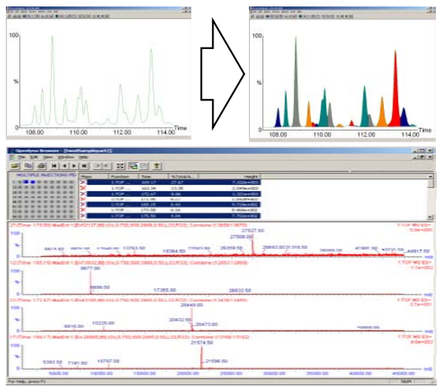
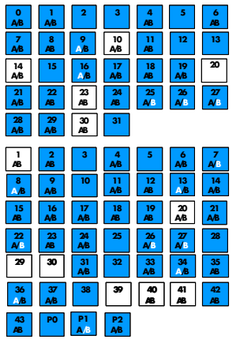


Figure 7: The OpenLynx™ module of MassLynx™ can be used to automate the analysis of intact protein LC/MS datasets. An entire dataset, or subset of data can be selected for automated TIC peak identification using the Apex PeakTrack™ function, the raw MS spectra summed for each peak using the Combine Spectra function, and deconvoluted spectra produced via the MaxEnt1™ deconvolution algorithm. Deconvoluted spectra (or raw spectra, if deconvolution is not selected) can be viewed and formatted in the OpenLynx™ Browser. The deconvolution process is computationally intensive, and depending on processing parameters (m/z range of raw spectra, iterations of processing, output mass range, and output mass bin size) can take hours to days of processing time on a mid-range Pentium III for datasets of ribosomal complexity.

Summary of Ribosomal Data



Small Subunit
33 /47 Isoforms (70%)
27 /32 Subunits (84%)

Large Subunit
52 /69 Isoforms (75%)
40 /46 Subunits (87%)

Observed

Not Observed

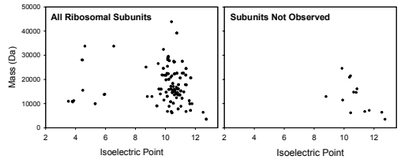


Figure 8: Ribosomal Subunit Coverage. Analysis of 75 µg of an an enriched ribosomal protein fraction permitted assignment of greater than 85% of all known ribosomal subunits with an average mass accuracy error of ~ 50 ppm. In many cases, more than one isoform of a subunit was observed, and several proteins were observed in multiple states of modification (not shown). Analysis of collected fractions at the peptide level are ongoing for confirmation of protein identity, protein modifications, and and identification of modification sites.

The typical subunit not observed during this analysis is smaller and more basic than the average ribosomal protein. Analysis of data from highly resolving 1-D separations (peak capacity ~240 vs. 675 for this 2-D approach) indicate that some of these subunits may not have been eluted of the SCX column under our maximum salt step, or fail to retain on the RP column during step elution. Analysis of peptides from global ribosomal digests, and analysis of the collected fractions should determine whether an unidentified subunit was modified, or not present in the initial sample or was lost during ribosome purification.

CONCLUSIONS

- 2-D (SCX/RP) LC/MS can be applied for the identification and characterization of complicated mixtures of intact proteins.
- The approach described here assigns both protein identification and modifications using the intact mass data, and confirms assignments by further characterization of those proteins from a collected split fraction.
- Data processing for MS spectra (and LC/MS analyses of protein mixtures) can be automated to derive maximum benefits from analytical approaches using MS detection of intact proteins.
- Factoring in the abilities of ESI-TOF MS and maximum entropy spectral deconvolution analysis to process spectra with up to 5 components, this approach currently demonstrates a peak capacity (675 x 5= 3375) capable of analyzing simple proteomes.

REFERENCES

1. Warner, J. R.; Gorenstein, C. In *Methods in Cell Biology*; Prescott, D. M, Ed.; Academic Press: New York, 1978; Vol. XX, pp 45-60.
2. Hardy, S. J. S.; Kurland, C. G.; Voynow, P.; Mora, G. *Biochemistry* **1969**, *8*, 2897-2905.
3. Costanzo, M. C.; et al.. *Nucl. Acids Res.* **2000**, *28*, 73-76.