ASMS 2005

AN AUTOMATED SOFTWARE TOOL FOR PROCESSING AND INTERPRETATION OF COMPLICATED INTACT PROTEIN LC/MS DATA SETS

Ignatius J. Kass[1], Hongji Liu[2], Kevin M. Millea[2], John C. Gebler[2], and Scott J. Berger[2]
[1] Waters Corporation, Beverly, MA [2] Life Sciences R&D, Waters Corporation, Milford, MA 01757

OVERVIEW

- Analysis of proteins by electrospray ionization mass spectrometry is complicated by the need to deconvolute multiply charged m/z spectra to produce neutral masses for sample components.
- This deconvolution process becomes significantly more complicated when analyzing data from online LC/MS protein separations.
- TIC information can be used to identify regions of MS data for summation and deconvolution with well resolved protein mixtures, but the approach fails with more complex protein separations.
- In this poster we describe a methodology for automating mass spectral deconvolution of complicated LC/MS data sets.
- The resulting data is output in a tab-delimited text file, and can be represented with intuitive and visually informative displays.
- Proper selection of processing parameters permit accurate intact protein mass determination, and retention of chromatographic profiles of each component.

METHODOLOGY

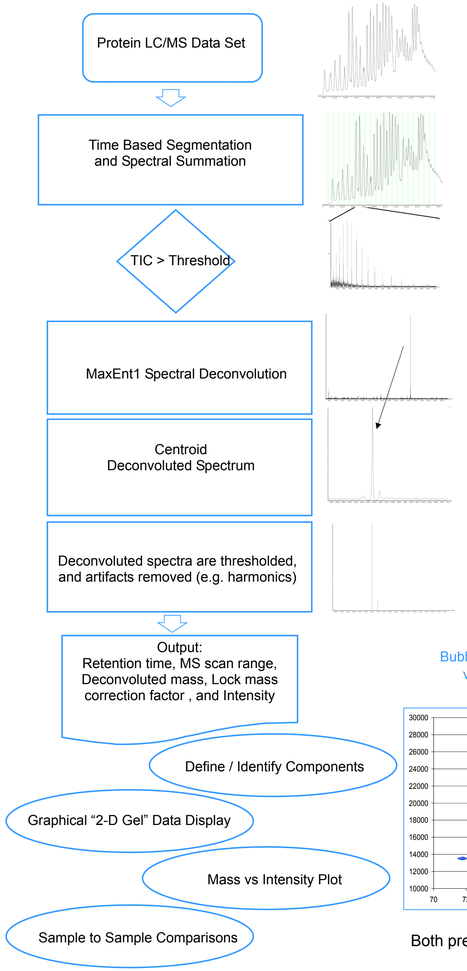
Yeast ribosomal proteins were analyzed by LC/ESI-TOF MS (Waters LCT) as described in work by Liu and coworkers (Ref. 1). LC/MS data was processed using a prototype software program called AutoME (Automated Maximum Entropy) within MassLynx 4.0 software (Waters) using the following conditions. 10 MS scan segments (1 scan/sec) were combined throughout the entire run. Segments containing a total ion intensity over 7000 counts were subjected to processing by the MaxEnt 1 (Ref. 2) spectral deconvolution algorithm (0.75 Da peak width, 3,000 to 45,000 output mass, 1 Da output resolution) until model convergence or 16 iterations was achieved. The resultant deconvoluted spectra was centroided, and components with intensities of greater than 20 counts were recorded to the output data file. Microsoft Excel 2003 was used to generate 2-D Gel bubble plot and Mass Fingerprint displays.

E.coli cytosolic proteins were analyzed by LC/ESI-TOF MS (Waters LCT Premier) under similar conditions, but with a gradient of only 60 min. Spectra (ESI+, V-mode (6,000 resolution), 1Hz, 650-2990 m/z) were processed by the AutoME program using the following conditions: 10 scans/ processing segment, input segment TIC threshold (10,000 counts), MaxEnt1 (0.75 Da width, 3,500-50,000 output mass range, 1 Da resolution, 15 iterations), Data reduction (Harmonic removal with 1 Da tolerance, 100 count minimum component intensity, Top30 peaks recorded/processing segment).

REFERENCES

1) Liu, H., Berger, S.J,, Chakraborty, A.B., et al. (2002) J. Chromatogr B, 782: 267-289.
2) Ferrige, A.G., Seddon, M.J., et al. (1992) Rapid Commun Mass Spectrom 6: 707. and Ferrige, A.G., Seddon, M.J., et al. (1992) Rapid Commun Mass Spectrom 6: 765.

Automated Deconvolution of Protein LC/MS Spectra



Note: key processing parameters in **bold**

Protein LC/MS data sets contain multiply charged m/z spectral patterns, associated with an MS scan number and LC retention time value.

Data can be segmented by summing scans over discrete intervals. A smaller **segment size** retains chromatographic resolution at the cost of increased processing time.

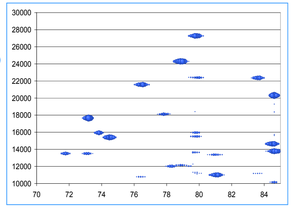
Eliminating processing of segments that have insufficient signal by applying a minimum **summed TIC intensity value** significantly reduces overall analysis time.

Maximum entropy spectral deconvolution approaches are powerful but processor intensive. Increasing the **output mass resolution** for deconvoluted spectra increases processing time, but improves the resolving capability and mass accuracy of output deconvoluted spectra.

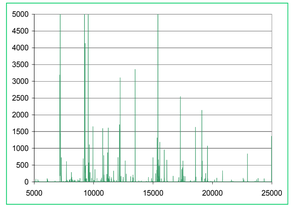
The MaxEnt 1 deconvolution algorithm is iterative, and increasing the **number of iterations** improves the fit of the resulting deconvoluted spectra to the original data and asymptotically improves component quantitation.

Centroiding produces a single data point and composite intensity measurement for each peak in the deconvoluted spectra. Applying a **threshold** to the deconvoluted spectra (absolute or relative to highest peak) eliminates spectral noise from the data. Both processes significantly decrease the size of output data file and simplify analysis of the resulting data set.

2D Gel Display
Bubble plot of RT (x) vs. Mass (y) vs. Intensity (bubble area)



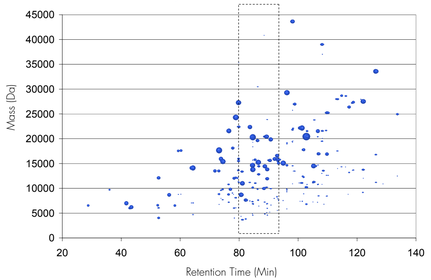
Protein Mass Fingerprint (PMF)
Plot of Mass (in 10 Da bins) vs. Intensity
Whole Run- RT data is ignored

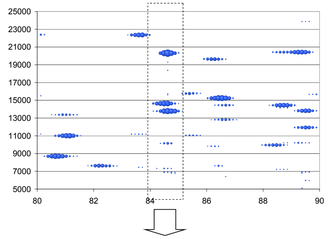


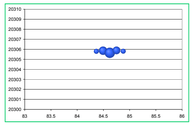
Both presentation formats can be used to generate run-to-run comparisons.

Component Level ID of Ribosomal Proteins

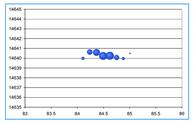
1-D LC/MS Analysis of Yeast Ribosomal Proteins



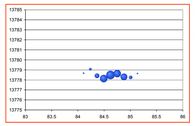




Component: 20305.8 with Peak RT of 84.5 min
Identified over 5 processing segments (50 scans @ 1 Hz)
Summed Intensity of 2805
Mass: 20305.8 (Avg), 20305.8 (Wt Avg)
20305.7 (Min), 20305.9 (Max)
ID: RPL20AB (Processing of N-terminal 3 AA (MYL))
Predicted Mass 20305.6
Difference: 0.18 Da (9 ppm)

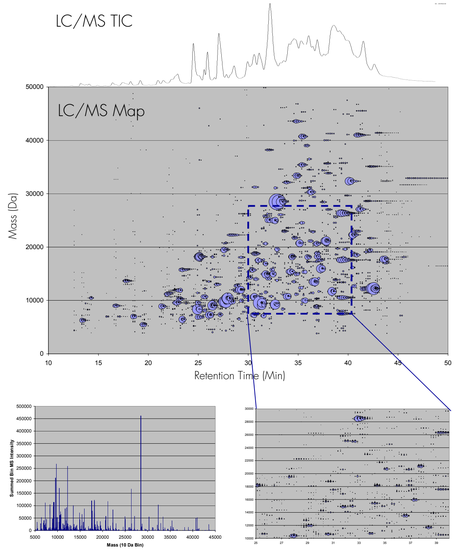


Component: 14640.3 with Peak RT of 84.6 min
Identified over 8 processing segments (80 scans @ 1 Hz)
Summed Intensity of 2500
Mass: 14640.3 (Avg), 14640.3 (Wt Avg)
14640.0 (Min), 14640.6 (Max)
ID: RPL32 (Processing of N-terminal Met)
Predicted Mass 14640.2
Difference: 0.11 Da (8 ppm)



Component: 13778.4 with Peak RT of 84.6 min
Identified over 9 processing segments (90 scans @ 1 Hz)
Summed Intensity of 2720
Mass: 13778.5 (Avg), 13778.4 (Wt Avg)
13778.1 (Min), 13779.1 (Max)
ID: RPL35AB (Processing of N-terminal Met)
Predicted Mass 13778.5
Difference: 0.09 Da (6 ppm)

1D LC/MS Map and PMF of E. *coli* cytosol



The LC/MS analysis of E. coli cytosol provides a more complex sample with components at much greater dynamic range.

TOP: LC/MS Protein Map of E. *coli* Cytosol
LOWER RIGHT: Enlarged image of central map region
LOWER LEFT: Protein Mass Fingerprint (Deconvoluted Mass vs. Intensity) over the entire LC/MS analysis

CONCLUSIONS

- Intact protein LC/MS data sets are amenable to automated data processing and analysis approaches.
- Proper selection of data processing parameters permits intact protein mass data to be measured with high precision and accuracy.
- Processing parameters can be selected such that the underlying chromatographic profiles of individual components are retained or completely ignored.
- Processed data can be generated to compare protein profiles, component intensities, or whole analysis "fingerprints" between samples.
- Intact protein LC/MS analysis can be conducted on a "proteomic" samples with high complexity and a large dynamic range expression.

TO DOWNLOAD A COPY OF THIS POSTER VISIT WWW.WATERS.COM/POSTERS

Lit Code Number 720001219EN