Applying Automated Maximum Entropy Spectral Deconvolution for the Analysis of Complicated Protein LC/MS Datasets

Scott J. Berger, Ignatius J. Kass, Kevin M. Millea, and John C. Gebler
Life Sciences Research and Development, Waters Corporation, Milford, MA

Overview

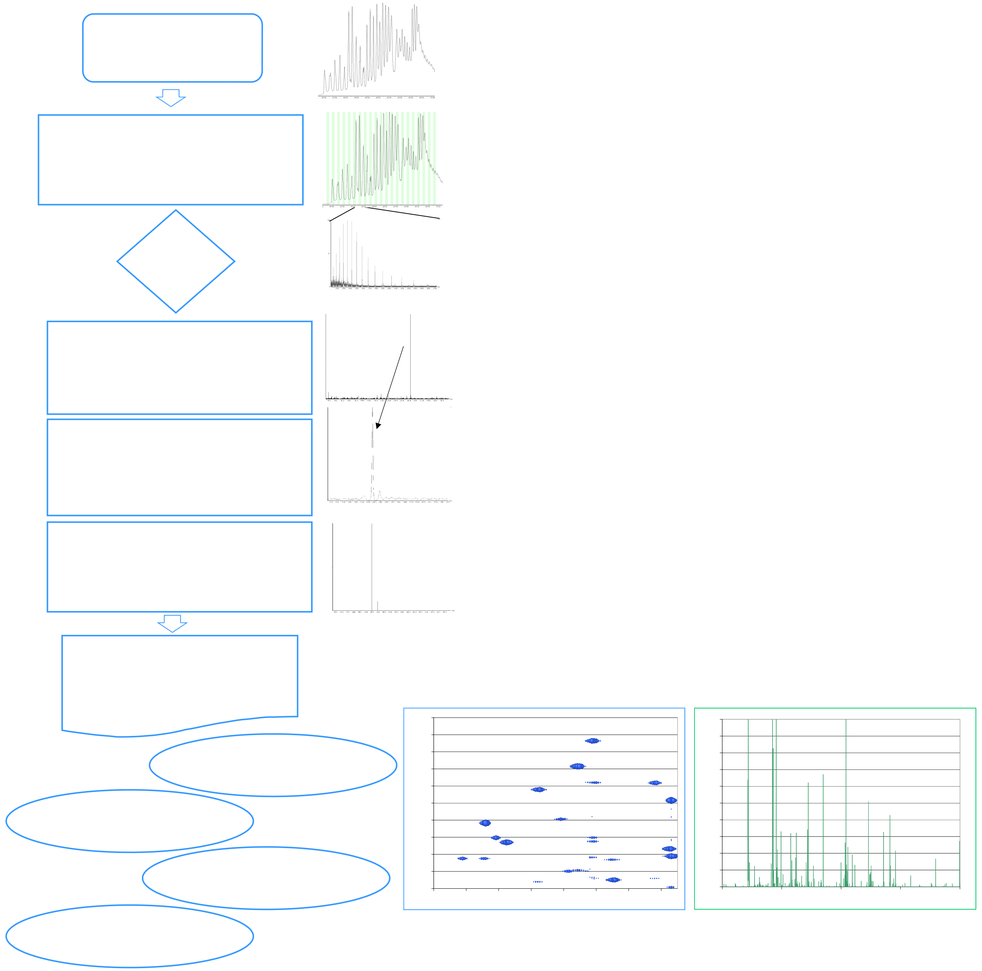
- Analysis of proteins by electrospray ionization mass spectrometry is complicated by the need to deconvolute multiply charged m/z spectra to produce neutral masses for sample components.
- This deconvolution process becomes significantly more complicated when analyzing data from online LC/MS protein 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.

Data Acquisition and Processing

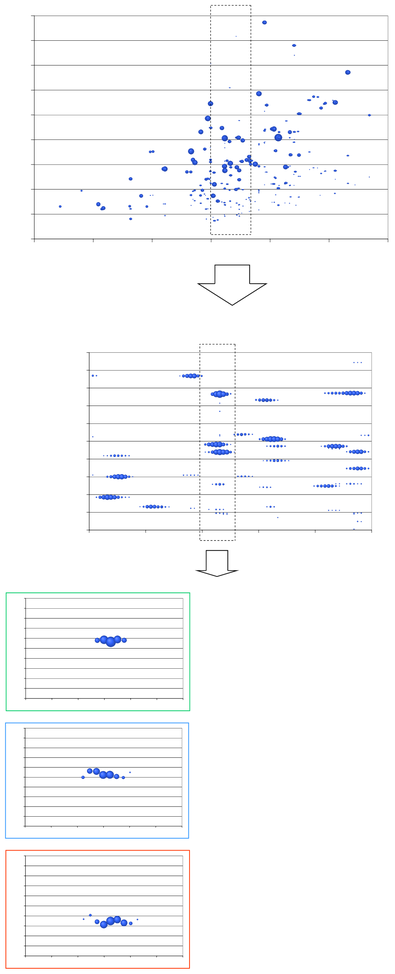
Yeast ribosomal proteins were analyzed by 1-D LC/ESI-TOF MS as described in work by Liu and coworkers (Ref. 1). LC/MS data was processed using a prototype software program called AutoME (Automated Maximum Entropy) using the following conditions. 10 MS scan segments (1 scan/sec) were combined throughout 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 bin size) 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 was used to generate 2-D Gel bubble plot and Mass Fingerprint displays.

E.coli cytosolic proteins were analyzed by 2-D(IEX/RP) LC/ESI-TOF MS as described in work by Millea and coworkers (Ref. 3). AutoME conditions: 10 Scan segments (2 sec/scan), TIC threshold (None), MaxEnt1 (0.75 Da width, 7,000-40,000 output mass range, 3 Da bin size, 10 iterations), 40 count minimum component intensity.

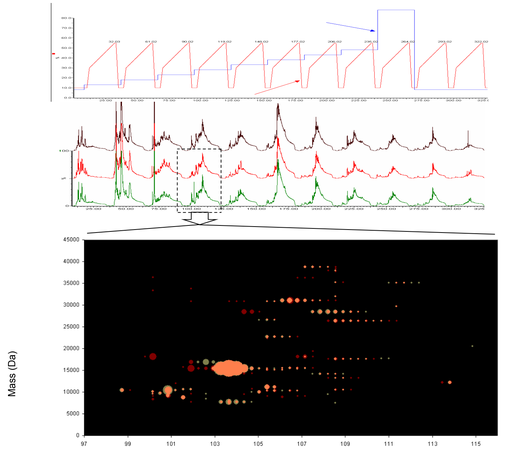
Automated Deconvolution of Protein LC/MS Spectra



Component Level ID of Ribosomal Proteins



Producing Run-Run Comparisons



This 2-D LC(IEX/RP)/ESI-TOF MS analysis of E. coli cytosol was produced using a step gradient of increasing salt in the first dimension, and a series of reversed phase gradients in the second dimension (**Panel A**).

The TIC traces (**Panel B**) of triplicate analyses shows generally reproducible patterns between runs.

The 2-D Gel display (**Panel C**) of two selected runs was generated by producing the individual displays, importing the displays into Adobe Photoshop, and adjusting the transparency of the top-most layer. Green = Run 1 only, Red= Run 2 only, Orange= Signal common to both runs.

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.
- Even complex LC/MS datasets produce data that can be compared between runs in a semi-quantitative manner.