A UNIFIED ALGORITHM FOR DECONVOLUTION ELECTROSPRAY IONIZATION MASS SPECTRAL DATA

INTRODUCTION

The state of the art in deconvoluting electrospray mass spectral data has changed little in the last 15 years. Following key developments in the 1990s, the maximum entropy (MaxEnt) deconvolution was first applied successfully to electrospray mass spectral data. It has since been used with great success in countless applications to protein analysis. Recent advances in Markov Chain Monte Carlo methods have enabled the development of a new Bayesian algorithm called BayesSpray that retains the power of MaxEnt but which demonstrates its application to peptides, proteins and top-down fragment thereof.

METHODS

Deconvolution using Nested Sampling

Nested sampling is a new inference algorithm by John Skilling specifically designed for large and difficult applications. In mass spectrometry, it is essential to single out mass spectral peaks amidst the vast complexity of the vast array of ions in the mass spectrum. If the likelihood function is highly multimodal, the peak shapes will be diverse, and a standard method such as maximum likelihood estimation could be quite unreliable. Algorithms such as nested sampling can be more efficient as they use nested sampling to explore the phase space to find the most probable solutions.

BayesSpray Results

BayesSpray features a list of masses, intensities, uncertainties and charge state. It uses the same algorithm as MaxEnt 3 to calculate the likelihood, ensuring the quality of the fit to the raw data that has been achieved by the algorithm. It is also possible to extract mock data for individual peaks (or combinations of peaks) allowing users to examine the evidence for each of the peaks.

Human Hemoglobin Variant

Blood collected in EDTA was diluted 500-fold in 50% aqueous acetone containing 2% formic acid. The solution was deproteinized by adding a small amount of water to 2.5% glacial acetic acid and processed by the deproteinization software over the NCI range. 100 Å-130 Å. The data were also processed using Waters' MaxEnt 1 algorithm for comparison.

EoT of Bosnie Ubiquitin

HDMS spectra were acquired on a Waters Synapt G2 HDMS instrument with a modified nanoflow electrospray source. The sample used was native ubiquitin, which is a small ubiquitin-like modifier of 8,000 Da. The sample was infused into the ion source at a flow rate of 100nl/min at a concentration of 100µM. The scan range was 300-1,600 m/z, and the acquisition was performed with a 2-second dwell time. The resulting spectrum was processed by the deconvolution software using the EoT algorithm for comparison.

High Resolution TOF Myoglobin Infusion

An infusion of experiment of 3µg/ml human myoglobin was performed on a Synapt G2 HDMS system. The EoT algorithm was performed with high resolution mass and tuned for ultimate resolution. The resulting spectrum was acquired using a Waters Synapt G2 HDMS instrument.

Gas Phase DHX of Substance P with Ion Mobility

The high mass experiment was performed using a Waters Synapt G2 HDMS instrument. Substrate was synthesized by nanospray infusion of the 3.5 µM solution (4×5) and the resulting spectrum was acquired. The resulting mass spectrum was processed by the deconvolution software using the EoT algorithm for comparison.

RESULTS

Figure 1. Quadrupole electrospray mass spectrometer of human hemoglobin. The mass list shows broadening of the 3+ and 2+ peaks due to the presence of the ε-aminocaproic acid.

Figure 2. A hexapole electrospray mass spectrometer of substance P. The algorithm is in practice stopped when the “maximum likelihood” solution is reached.

Figure 3. Nested sampling of the mass list shows the deconvoluted mass list with proper error bars expressing statistical uncertainties. Like MaxEnt 1 and MaxEnt 3, BayesSpray produces mock data. This is a convenient way for the user to infer masses, with proper error bars expressing statistical uncertainties. In this way, all annotated minor components are present in both spectra, the baseline in the magnified region is consistent with expectations.

Figure 4. Evidence for two triply charged conformers of substance P separated on an ion mobility experiment.

CONCLUSION

BayesSpray is a novel algorithm for deconvolution of electrospray mass spectral data. We have shown results obtained from mass spectra acquired using a range of instruments under diverse operating conditions.

The algorithm provides unified deconvolution of peptide and protein data.

The open concept of a list of masses and intensities accompanied by error bars.

Other valuable diagnostic include the evidence for each of the peaks, information and mock data for each detected peak.

The algorithm is particularly useful for proteins and peptides.

The potential to be used as powerful tools for the MaxEnt 1 and MaxEnt 3, producing significantly cleaner baselines for accurate spectra and improved interpretation of the complex data produced by the top-down fragmentation of large molecules.

REFERENCES


3. www.matrix-science.com

Figure 5. Difference in hydrogen deuterium exchange profile for the deuterated conformers of Substance P. Masses shown in Figure 6. Following BayesSpray deconvolution. The upper and middle plots show the deconvoluted spectra for the mobile and fixed ions respectively. The difference between MaxEnt and BayesSpray reconstructions is evaluated at the most intense mass peak.

Figure 6. Raw myoglobin data at TOF resolution 74,000 (top), BayesSpray data (middle) and BayesSpray reconstruction (bottom). The peak corresponding to the parent mass is totally resolved from the minor species peaks at sodium adduction. The middle plot shows the data produced by BayesSpray, showing that this algorithm maintains both the resolution of TOF and the ability to correctly adduct ions.

Finally, the peak plot in Figure 5 shows a myoglobin spectrum acquired by a Waters Q-TOF instrument operating at TOF resolution 74,000. The plot shows correctly adducted ions for the mobile and fixed ions and is accompanied by error bars. The bottom plot shows the deconvoluted mass list, with proper error bars expressing statistical uncertainties.