A UNIFIED ALGORITHM FOR DECONVOLUTING ELECTROSPRAY IONIZATION MASS SPECTRAL DATA

John Skilling[1], Keith Richardson[2], Jeff Brown[2], Iain Campuzano[2], Brian Green[2], Jason Wildgoose[2]
[1]Maximum Entropy Data Consultants Ltd., Kenmare, Ireland; [2]Waters Corporation, Floats Road, Manchester M23 9LZ, UK

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 1980's, maximum entropy (MaxEnt) deconvolution was first applied successfully to electrospray data in 1992[1]. 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 methods, but which can be applied to a wider range of data. We demonstrate its application to peptides, proteins and top-down fragments thereof.

METHODS

Deconvolution Using Nested Sampling

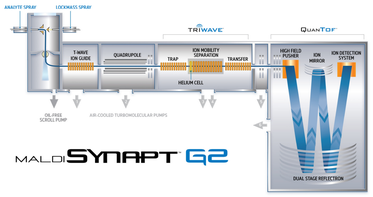
Nested sampling[2] is a new inference algorithm by John Skilling specifically designed for large and difficult applications. In mass spectrometry, iteration is essential because single-pass algorithms are inherently incapable of inferring a spectrum under the nonlinear digital constraint that intensities must all be positive. Nested-sampling iterations steadily and systematically extract information (also known as negative entropy) from the data and yield mass spectra with ever-closer fits.
Although capable of proceeding to a final "maximum likelihood" solution, the algorithm is in practice stopped when it has acquired enough information to define the distribution of spectra that are both intrinsically plausible and offer a probabilistically correct fit to the data. After all, any single solution would be somehow atypical, whereas professional standards demand that results are provided with proper estimates of the corresponding uncertainties.

The immediate output from nested sampling is an ensemble of a hundred or so typical spectra, each in the form of a list of parent masses. These masses have intensities which, when the charge state is unknown, are individually distributed over charge. Just as in statistical mechanics (which helped to inspire nested sampling), the ensemble can be used to define mean properties together with fluctuations. In this way, nested-sampling results can be refined to a list of reliably-inferred masses, with proper error bars expressing statistical uncertainty, and full knowledge of how each mass relates to the data.

Peak Modelling

Mock data is produced from trial masses by convolution with an appropriate (mass dependent) isotope distribution, correction for charge state and, finally, convolution with the instrumental peak shape.

For each dataset, an appropriate model of the instrumental peak shape corresponding to an isotopically pure species was used. A fixed full width at half maximum was used for quadrupole data, whereas a fixed instrument resolution was specified for TOF data.



BayesSpray Results

BayesSpray outputs a list of masses, intensities, uncertainties and reliabilities. Like MaxEnt 1 and MaxEnt 3, BayesSpray produces mock data. This is a convenient way for the user to confirm visually 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 features reported in the final mass list.

Human Hemoglobin Variant

Blood collected in EDTA was diluted 500-fold in 50% aqueous acetonitrile containing 0.2% formic acid. This solution after desalting with cation exchange resin beads (Bio-Rad AG 50W-X8, hydrogen form) was introduced directly into the ESI source of a Waters quadrupole mass spectrometer. Data were acquired over the *m/z* range 930-1180 for 3 minutes and internally calibrated using the multiply charged β-chain ions and processed by the deconvolution software over the *m/z* range 980-1180. The data were also processed using Waters' MaxEnt 1 algorithm for comparison.

ETD of Bovine Ubiquitin

ETD experiments were performed on a Waters Synapt G2 HDMS instrument with a modified nanoflow electrospray source. The reagent used to generate radical anions was nitrosobenzene (*m/z* 107). The ion source polarity and the quadrupole set mass were sequentially switched to deliver cations and singly charged radical anions (generated by glow discharge ionization). ETD fragmentation was performed in the travelling wave Trap cell filled with helium at a pressure of $5x10^{-2}$ mbar. The data were processed using BayesSpray and Waters' MaxEnt 3 algorithm.

Gas Phase HDX of Substance P with Ion Mobility

The HDX experiment was performed on a Waters Synapt HDMS instrument. Substance-P was ionized by nanospray and the 3^+ precursor (*m/z* 450) was selected by the quadrupole. Ions were passed through the Trap cell where they reacted with the deuterated ammonia reagent. An additional gas inlet needle valve was connected to the Trap gas cell for the introduction of the deuterated ammonia causing a pressure increase of $5x10^{-3}$ mbar. Subsequently, ion mobility measurements were made for the deuterated products of substance-P.

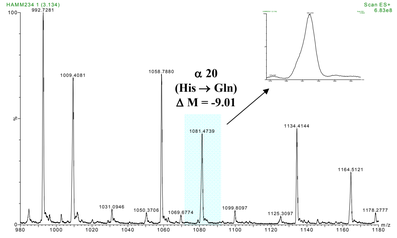


Figure 1. Quadrupole electrospray mass spectrum of human hemoglobin. The inset detail shows broadening of the 20+ alpha chain peak due to the presence of the Le Lamentin variant.

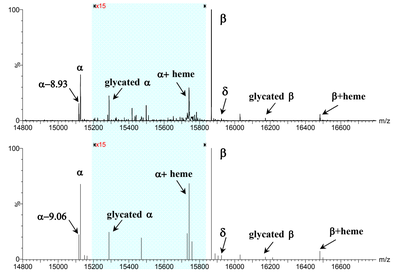


Figure 2. A Hemoglobin Le Lamentin spectrum deconvoluted using both MaxEnt 1 (top) and BayesSpray (bottom). The theoretical mass difference for the α-chain variant is -9.01 Da. The shaded region has been magnified by a factor of 15 for clarity.

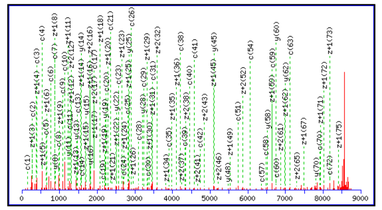
High Resolution TOF Myoglobin Infusion

An infusion experiment of 2pmol/µl horse heart myoglobin at 5µl/min was performed on a Synapt G2 HDMS system.

The Synapt was operating in high resolution mode and tuned for ultimate resolution. The acquisition system employed a prototype 6 GHz mode which, when combined with the ultimate resolution tuning, allowed resolutions of over 70,000 (FWHM) to be achieved with some loss in sensitivity.

RESULTS

Figure 2 demonstrates the ability of BayesSpray to detect minor components such as the δ-chain and glycated α- and β-chains in an infused sample of human hemoglobin. In addition, the unresolved variant α-chain is separated by the algorithm, and the mass difference is measured with an accuracy comparable to the MaxEnt 1 result. Also visible, but not annotated, are sodium and potassium adducts and the variant α-chain plus heme. Note that although the spectra are remarkably similar, and all annotated minor components are present in both spectra, the baseline in the magnified region is considerably cleaner in the new result. This is a typical distinction between MaxEnt and BayesSpray reconstructions of even the most complex intact protein spectra.



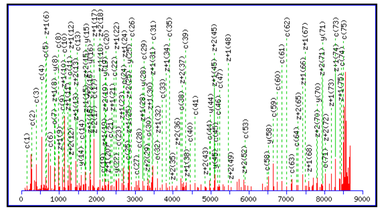


Figure 3. *Processed electron transfer dissociation spectrum of bovine ubiquitin. Precursor ion m/z 714.0^{12+} was selected for fragmentation. The data were processed using both MaxEnt 3 (top) and BayesSpray (bottom). The spectrum was searched and annotated using Mascot.*

Figure 3 shows MaxEnt 3 and BayesSpray reconstructions of top down fragmentation spectra of bovine ubiquitin. The data were searched and annotated using Mascot[3]. Both of the scores reported by the search engine were considerably improved for the BayesSpray reconstruction. The improvement appears to result from the detection, overall, of an additional ten ions by BayesSpray including the contiguous series of ions c_{44} to c_{47} which have masses of around 5000 Da.

It is worth noting here that BayesSpray is capable of the correct interpretation of resolved species with the same unit mass and charge state. This is increasingly important given the high resolution afforded by modern instrumentation.

Figure 4 shows the isotopes of the ion mobility separated conformers of undeuterated substance P displayed in Waters DriftScope software.

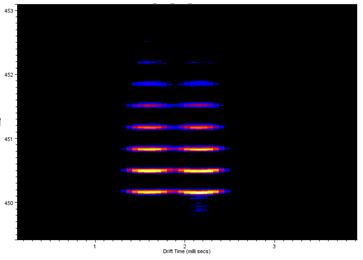


Figure 4. Evidence for two triply charged conformers of substance P separated in an ion mobility experiment.

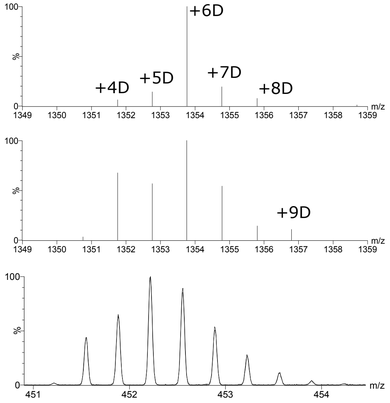


Figure 5. Difference in hydrogen deuterium exchange profile for the deuterated conformers of Substance P shown in Figure 4. following BayesSpray deconvolution. The upper and middle plots show the deconvoluted spectra for the more mobile and less mobile conformers respectively. The bottom plot shows the superimposed raw data and BayesSpray mock data corresponding to the second conformer.

The deconvoluted results for the two spectra obtained from the deuterated sample by summing over these peaks are shown in the top two plots in Figure 5. The striking difference in the pattern of deuterium uptake appears to support the observation of two distinct conformers in the ion mobility experiment.

Finally, the top plot in Figure 6 shows a myoglobin spectrum acquired by a Waters Synapt G2 instrument operating at TOF resolution 74,000. The peak corresponding to the parent mass is followed by a regular series of smaller peaks corresponding to sodium adducts. The middle plot shows the mock data produced by BayesSpray, showing that it can reproduce both the high resolution isotopic structure and the pattern of sodium adduction. The bottom plot shows the deconvoluted mass list.

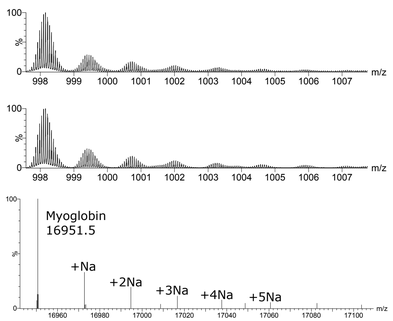


Figure 6. Raw myoglobin data at TOF resolution 74,000 (top), BayesSpray mock data (middle) and BayesSpray reconstruction (bottom).

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 output consists of a list of masses and intensities accompanied by error bars.**
- **Other available diagnostics include the evidence (which is used in model comparison), information and mock data for each detected peak.**
- **BayesSpray appears to be at least as powerful as MaxEnt 1 and MaxEnt 3, producing significantly cleaner baselines for protein spectra and improved interpretation of the complex data produced by the top down fragmentation of large molecules.**

References

1. A.G. Ferrige, M.J. Seddon, B.N. Green, S.A. Jarvis and J. Skilling, Disentangling electrospray spectra with maximum entropy, *Rapid Commun. Mass Spectrom* **6** (1992), pp. 707–711.
2. John Skilling. Nested sampling. In R. Fischer, R. Preuss, and U. von Toussaint, editors, Bayesian inference and maximum entropy methods in science and engineering, AIP Conference Proceedings 735, pages 395–405, 2004.
3. www.matrix-science.com

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