ENHANCING AND AUTOMATING THE MAXIMUM ENTROPY DECONVOLUTION OF PROTEIN SPECTRA ACQUIRED ON HIGH-RESOLUTION TOF MASS SPECTROMETERS

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Keith Richardson[1] ; John Skilling[2] ; Jason Wildgoose[1] ; Iain Campuzano[1] ; Scott Berger[3] ; Robert Bateman[1]
[1]Waters Corporation, Manchester, United Kingdom; [2]Maximum Entropy Data Consultants Ltd, Kenmare, Ireland; [3]Waters Corporation, Milford, MA

INTRODUCTION

Proteins give rise to a series of multiply charged forms during electrospray ionization that must be reduced to a neutral form to allow proper mass measurement and quantitation. Maximum entropy techniques have long been employed to determine the simplest distribution that could give rise to the raw data. Typically a single manually measured peak width is used to process a spectrum containing many charge states of one or more proteins. While this is appropriate for low resolution quadrupole data, we show that it results in damaged reconstructions of high resolution TOF data. We demonstrate that significant benefits, including ease of use and the potential for automation, arise from modelling instrumental and isotopic contributions to peak width correctly.

METHODS

Maximum Entropy

We omit a detailed introduction to maximum entropy methods, but refer the interested reader to the literature[1,2]. However, it is well known that they provide a powerful method for deconvolution of electrospray mass spectra corresponding to mixtures of proteins[3]. Success of the method relies on the quality of the mapping between hypothesized truth and the data domain. This requires a detailed understanding of how the measuring apparatus, in this case a mass spectrometer, responds to various inputs (intact proteins of different masses and charge states).

The algorithm described here is MaxEnt 1 as enhanced in Waters BiopharmaLynx™ software.

Origins of Peak Width

In the type of spectra that we will be discussing, each intact protein is represented by an envelope of peaks corresponding to different charge states. This is illustrated for a single protein (Myoglobin) in figure 1. We shall refer casually to "peak width" throughout this poster. Unless otherwise specified, this can be taken to refer to full width half maximum, standard deviation or any other common measure of peak width.

The width of a peak observed in the electrospray mass spectrum of an intact protein generally depends on the chemical composition of the species involved, the sample history and the type and configuration of the mass spectrometer used.

In this poster we restrict our attention to situations where adducts can be cleanly separated from the unmodified protein after deconvolution. Even after desalting, this becomes impossible at high protein molecular weight. In this molecular weight regime a different approach to modeling peak width is required.

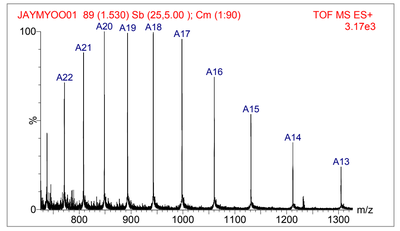

Figure 1. A multiply charged electrospray spectrum of Myoglobin. The numbers above the peaks correspond to charge state.

Modeling Peak Width

In general, the instrumental width for an isotopically pure compound varies with position on the M/Z axis. Two common cases are TOF instruments for which, to a good approximation,

$$W_{TOF}(X) = R^{-1}X$$

where X=M/Z and R is the instrument resolution. In contrast, quadrupole instruments are often configured to have approximately constant peak width

$$W_{QUAD}(X) = C.$$

If the instrument resolution is sufficiently high, or the protein being analyzed has sufficiently low mass (figure 2), individual isotopes can be resolved by deconvolution. In this case the appropriate peak width model, W(X), can be used directly.

When resolution of isotopes is not possible, or is unnecessary, the required peak shape is the convolution of the instrument peak shape with the isotopic profile of the molecule (after adjusting the latter for charge state Z):

$$W^2_{TOTAL}(X) = W^2_{INST}(X) + \left(\frac{W_{ISOT}}{Z}\right)^2.$$

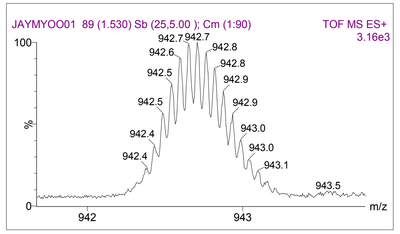

Figure 2. The 18+ Charge state of Myoglobin acquired with TOF resolution 20000. Individual isotopes are visible.

Isotopic Width

In the above expression for the total observed peak width, the isotopic width W_{ISOT} appears. This is the width, in Daltons, of the isotope distribution for the protein in question. Since the identity of the protein being analyzed is not generally known, we cannot calculate this directly. Fortunately however, the detailed composition of a protein does not have a large effect on the shape of its isotope distribution.

In figure 3 we have plotted the theoretically calculated variance σ^2 (in Da^2) of the isotope distribution of eleven commonly occurring proteins with mass M ranging from just under 3kDa to over 66kDa. The relationship is described to a good approximation by a straight line through the origin:

$$\sigma^2_{ISOT} = VM$$

where the constant V =7.1x10^{-4}Da. The isotopic peak width to use can therefore be determined automatically given the approximate mass of the protein under analysis.

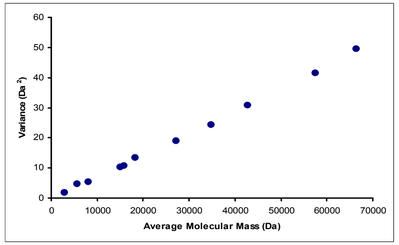

Figure 3. Variance (in Da^2) of the isotope distribution of a variety of proteins as a function of average molecular mass.

Usability and Automation

Data was processed using BiopharmaLynx™. This is a software package designed for automated analysis of TOF peptide mapping and intact protein samples. In order to create an intact protein method, one can specify:

- Instrument resolution
- Lock mass information
- Protein chains that may be present in the sample
- Disulfide bridges
- Fixed and variable modifiers
- Time ranges to process
- Deconvolution parameters

By default, the deconvolution peak width is set using the protein mass and instrument resolution as described above. Where necessary the automatic selection can be overridden with a constant width, or a width that varies according to the user's specification.

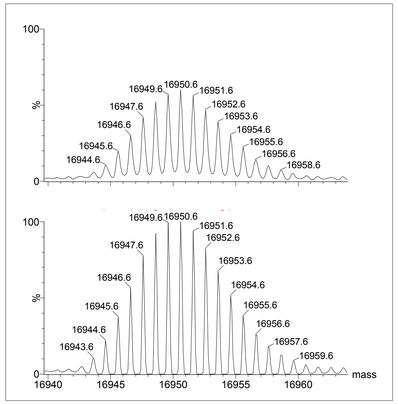

Figure 4. Myoglobin with FWHM resolution 20000 processed using constant peak width (top) and TOF peak width model (bottom).

Once created, a method can be used to automatically process any number of samples. The results appear in a browser which allows detailed comparison of each sample with a designated control sample. Spectra and chromatograms are automatically annotated using the supplied protein sequences.

RESULTS

Improved TOF Peak Width Model

200 fmol/µL Horse Heart Myoglobin in 50:50 ACN/H_20 with 0.2% formic acid was infused at 5 µL/min into a Waters Synapt™ High Definition MS™ system operating in W-optics™ at a full width half maximum (FWHM) resolution of 20000. The experiment was then repeated with the instrument reconfigured to operate at a resolution of 2000. At the higher resolution, individual isotopes are clearly visible in the raw data as shown in figure 2. At low resolution each charge state is a single unresolved peak.

For both datasets, the input spectra were background subtracted using a polynomial order of 25 with 5% of data below the curve. The input m/z range was chosen to be 726-1327 Da/e, including Myoglobin charge states 13+ to 23+. The output mass range was 16800-17100 Da. MaxEnt 1 was allowed to iterate to convergence. Both spectra were processed twice, first with a constant peak width (chosen to be correct near the centre of the spectrum) and again with the correctly varying TOF and isotopic contributions.

Figure 4 shows the results obtained at high resolution. When the TOF model is used, the isotopic peaks are sharper and resolved to baseline.

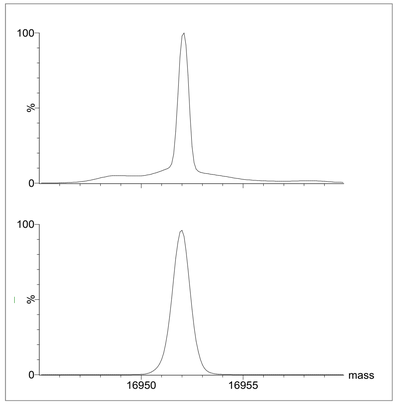

Figure 5. Myoglobin with FWHM resolution 2000 processed using constant peak width (top) and TOF peak width model (bottom).

At low resolution (figure 5), the new result is broader, but use of the TOF model has produced a cleaner result without the unsightly raised baseline. In order to better understand this we extracted an intermediate MaxEnt result with charge states separated out. This is shown in figure 6. With a constant peak width (red trace) the result looks reasonable at intermediate charge states. At low charge states (13^+ to 17^+) the peak width used in deconvolution is too small resulting in broad, low intensity features. At high charge states (21^+ to 23^+) the peak width is too large, producing artificially sharp results. The TOF model (blue trace) gives much more consistent results.

Automated Processing

In a separate experiment designed to measure stability of mass measurement, 1 pmol/µL Human Haemoglobin was introduced into a Waters® Q-Tof™ Premier operating in V-optics™ using a Waters nanoACQUITY™ UPLC system with an overall gradient turn around time of 5 minutes over a period of 12 hours. The instrument FWHM resolution was 10000. An external reference solution was introduced into the mass spectrometer by means of a LockSpray™ reference probe.

The eight resulting haemoglobin datasets were background subtracted using the same parameters as before. The input m/z range was 745-970 Da/e, and the output range was 14800-16800 Da. TOF peak widths were set automatically using the instrument resolution.

Processing of the data was completed in under two and a half minutes. The standard deviation of the eight alpha chain mass measurements was 5.3 ppm, while that of the beta chains was 3.3 ppm.

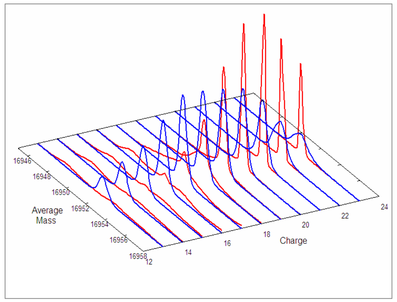

Figure 6. A MaxEnt 1 result with charge state contributions separated out. The red trace is the constant peak width result and the blue trace is the result when the TOF peak width model is used.

DISCUSSION

Setting parameters for deconvolution of intact protein data is often seen as a subjective process. However in the regime discussed here the observed peak width is well understood.

At relatively low molecular weight it is possible to use the instrument contribution alone to obtain a deconvolved spectrum in which each isotope is preserved, even with a wide input m/z range. This is not possible when a constant peak width is used.

As protein molecular weight increases (or instrument resolution decreases), it becomes impossible to separate isotopes. It is then necessary to combine instrumental and isotopic contributions to obtain the correct model for deconvolution.

Automated processing using the improved model along with lock mass correction results in extremely stable mass measurements.

CONCLUSION

- **Observed peak width is well understood for light proteins**
- **Using an appropriate peak width model results in significantly improved deconvolution**
- **Automated processing of batches of intact protein data using a standard method can produce highly reproducible and accurate mass measurements**

References

1. Skilling, *J. Maximum Entropy and Bayesian Methods, J.Skilling (ed), Kluwer Academic Publishers, Dordtrecht*, (1989), pp. 45-52.
2. http://www.maxent.co.uk/documents/MemSys5_manual.pdf
3. A.G. Ferrige, M.J. Seddon, B.N. Green, S.A. Jarvis and J. Skilling, *Rapid Commun. Mass Spectrom.* **6** (1992), pp. 707–711.

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

720002667EN