Positive Probability Ltd

Note P8: MS Charge Deconvolution – Protein Mixture

Introduction

There are now several methods available for transforming multi-charge data into a zero-charge result. The simplest methods use algebra but these fail on anything but the simplest data because they are unable to take noise and errors into account. Bayesian and entropic data reconstruction methods are much more successful but have the disadvantage that they are time-consuming to compute and are very prone to artefacts since the result must contain the same intensity as that in the data. However, the ***ReSpect***™–based ***Discharge***™ interface is typically about two orders of magnitude faster than other reconstruction methods and is much less prone to artefacts since the result is not forced to have the same intensity as the data. In the example presented here we show the principles behind the PPL methodology.

Data and Data Processing

The data are a mixture of three proteins and all are visually obvious in the baseline corrected spectrum shown in Figure 1 below.

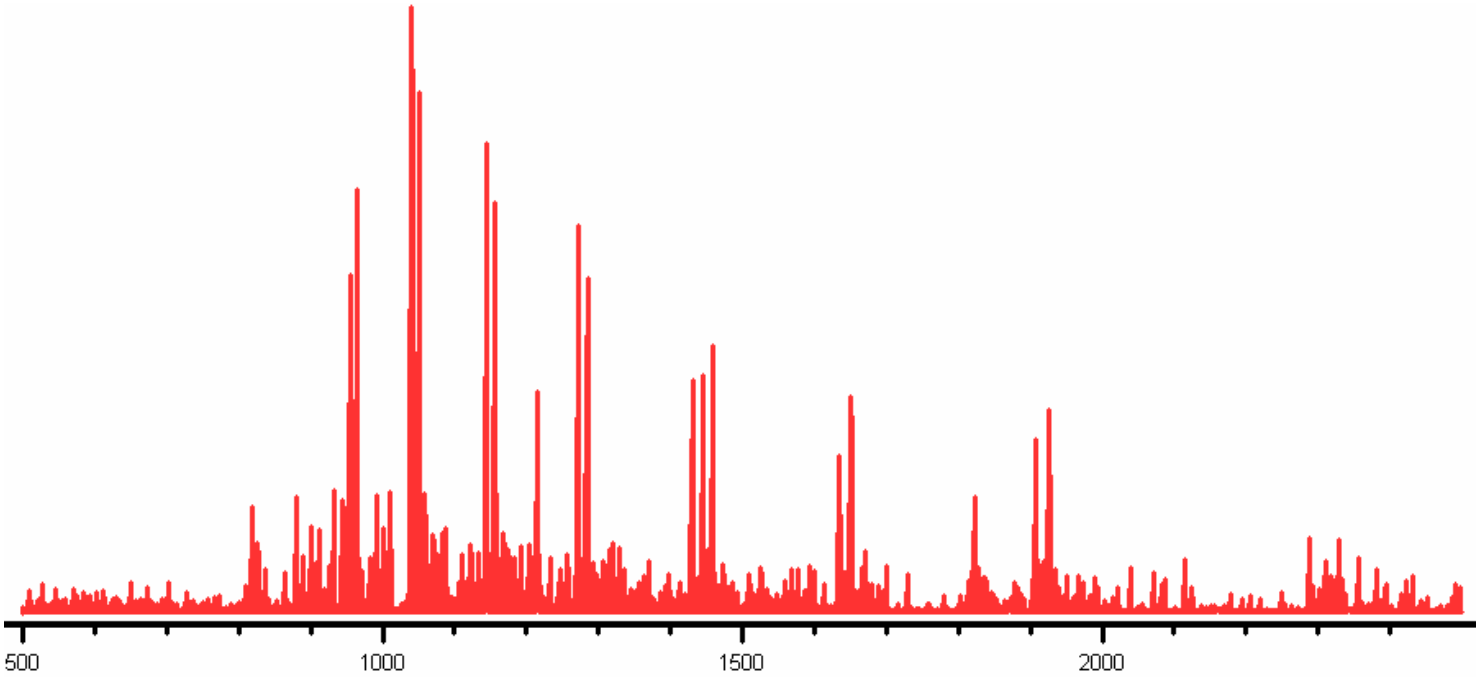

Figure 1. Baseline corrected data.

As can be seen, the data are particularly noisy. Therefore, whether the data are deconvolved or simply centroided, many peaks will be found that are noise and unrelated to the three proteins. It is these unrelated features that present algebraic methods with serious problems because they must be assumed to be genuine. The poor quality of the data is clear in the horizontal expansion shown in Figure 2 below.

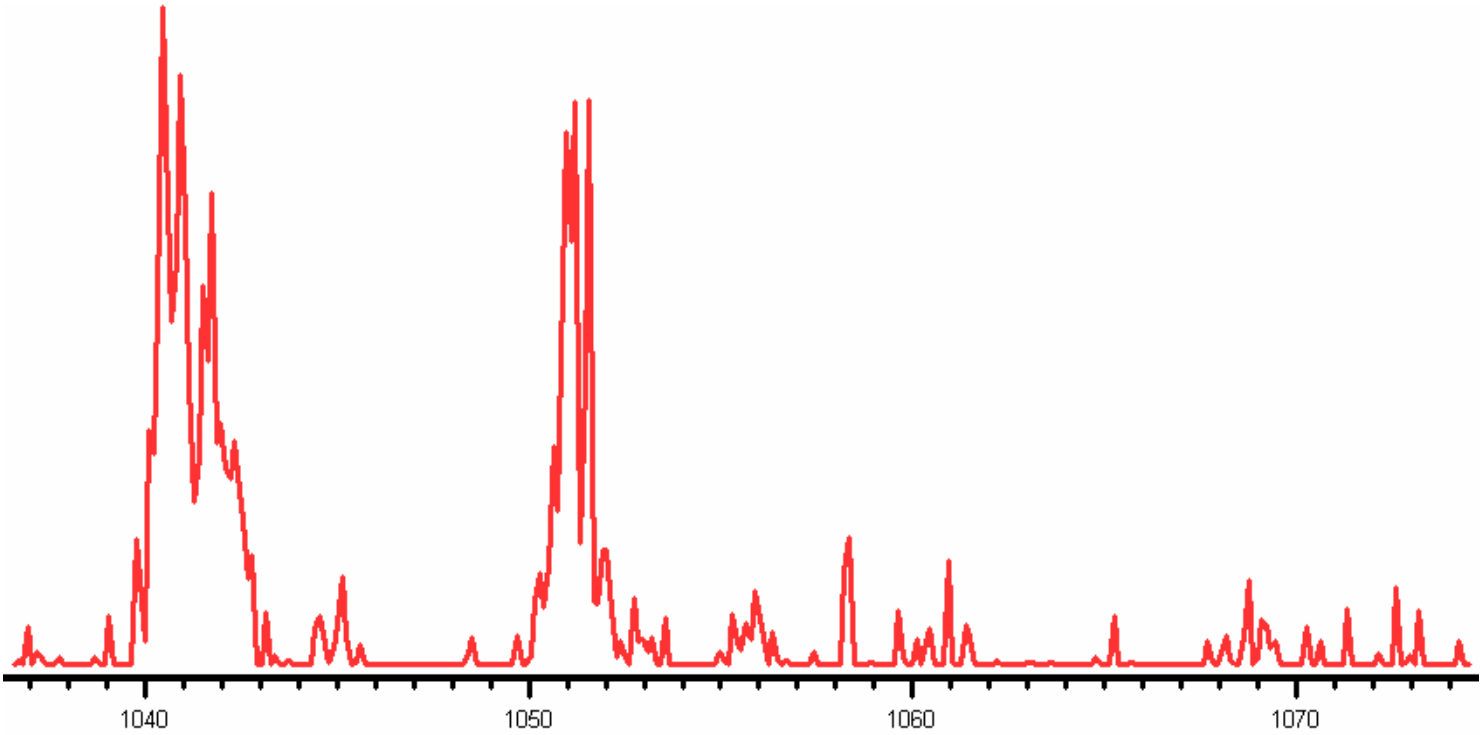

Figure 2. Horizontal expansion showing poor quality of the data.

Because of the noisy nature of the data, modelling a single peak would be questionable. The modelling was therefore performed over the range m/z 1400-1700. The resulting single model was then used to deconvolve the entire data. The result, shown as a spike plot at 1 standard deviation and 68% confidence is shown in Figure 3 below. A ***ReSpect***™–based charge deconvolution was then performed to produce a zero-charge spectrum for an output range of 5-25 kDa.

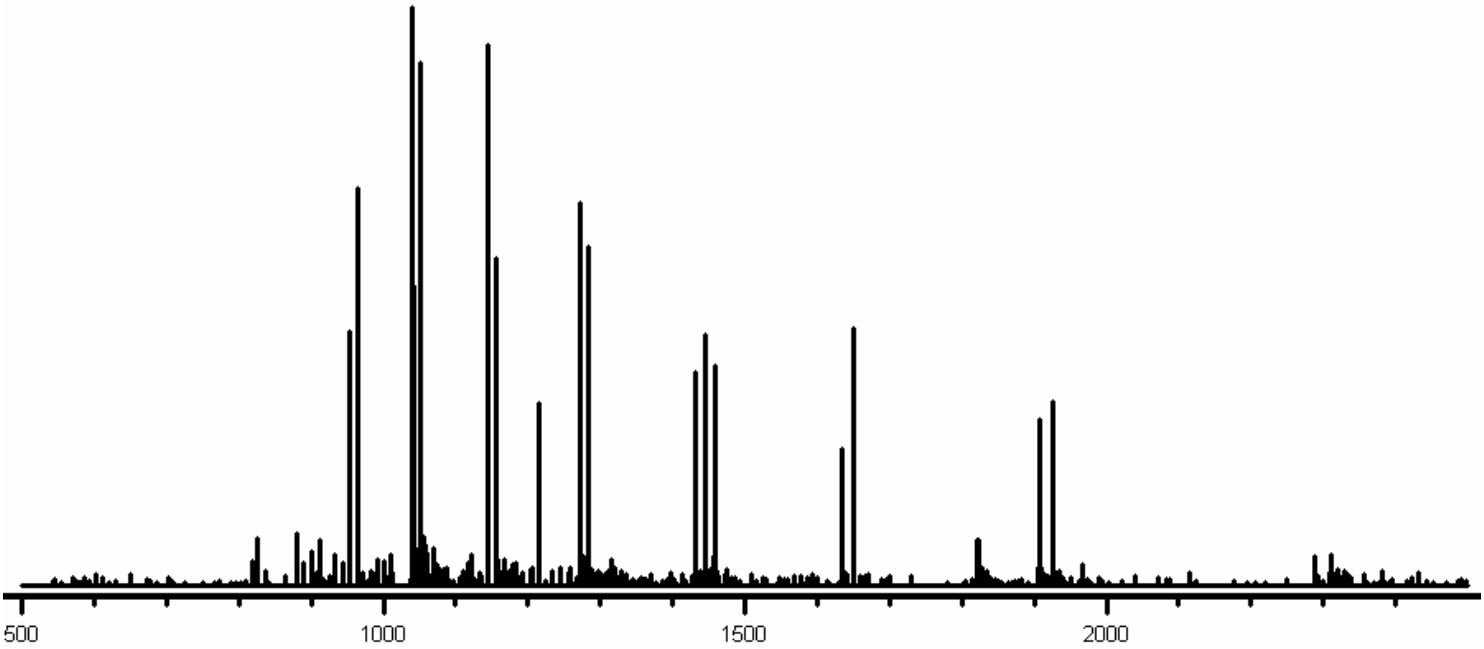

Figure 3. Spike plot of deconvolved result.

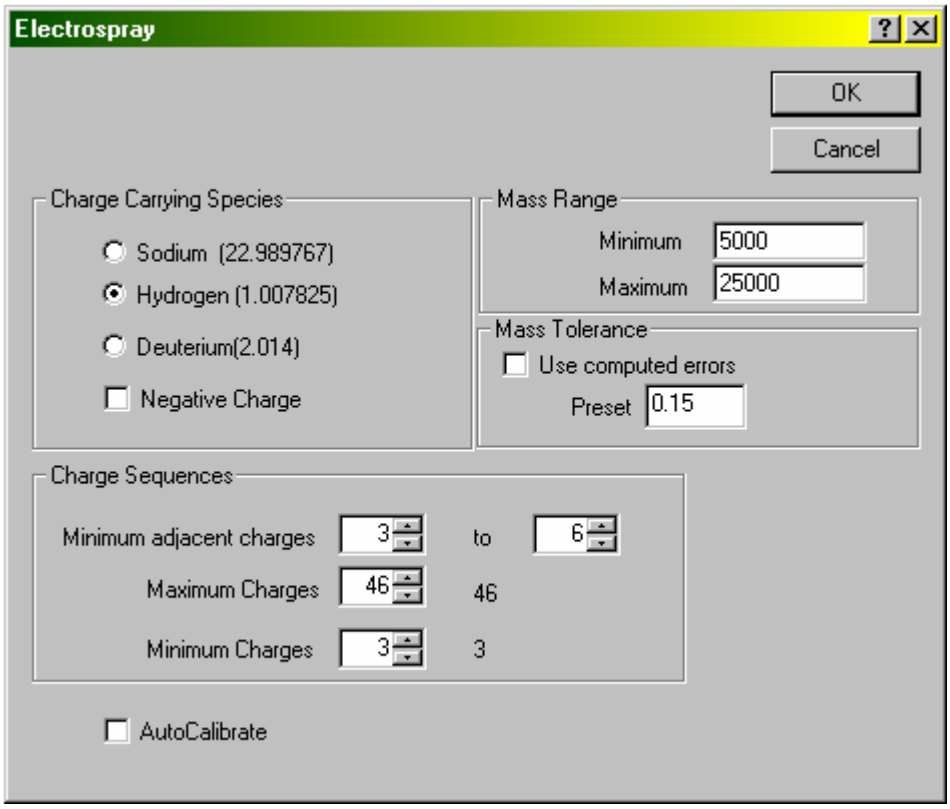

Figure 5. Charge deconvolution input parameters.

The input parameters used for the charge deconvolution (the applied constraints) are shown in Figure 5 on the left. Their meaning is as follows:

Mass Range: The mass range over which the result is to be computed.

Mass Tolerance: The allowed error in Da over the data range. This is a maximum error that incorporates any calibration error.

Charge Sequences: The minimum number of adjacent charges that are required as evidence that a species is present.

The maximum and minimum charge that may be present in the data as calculated from the data and output mass ranges.

Results and Discussion

The zero-charge result, as a spike plot, is shown in Figure 4 below.

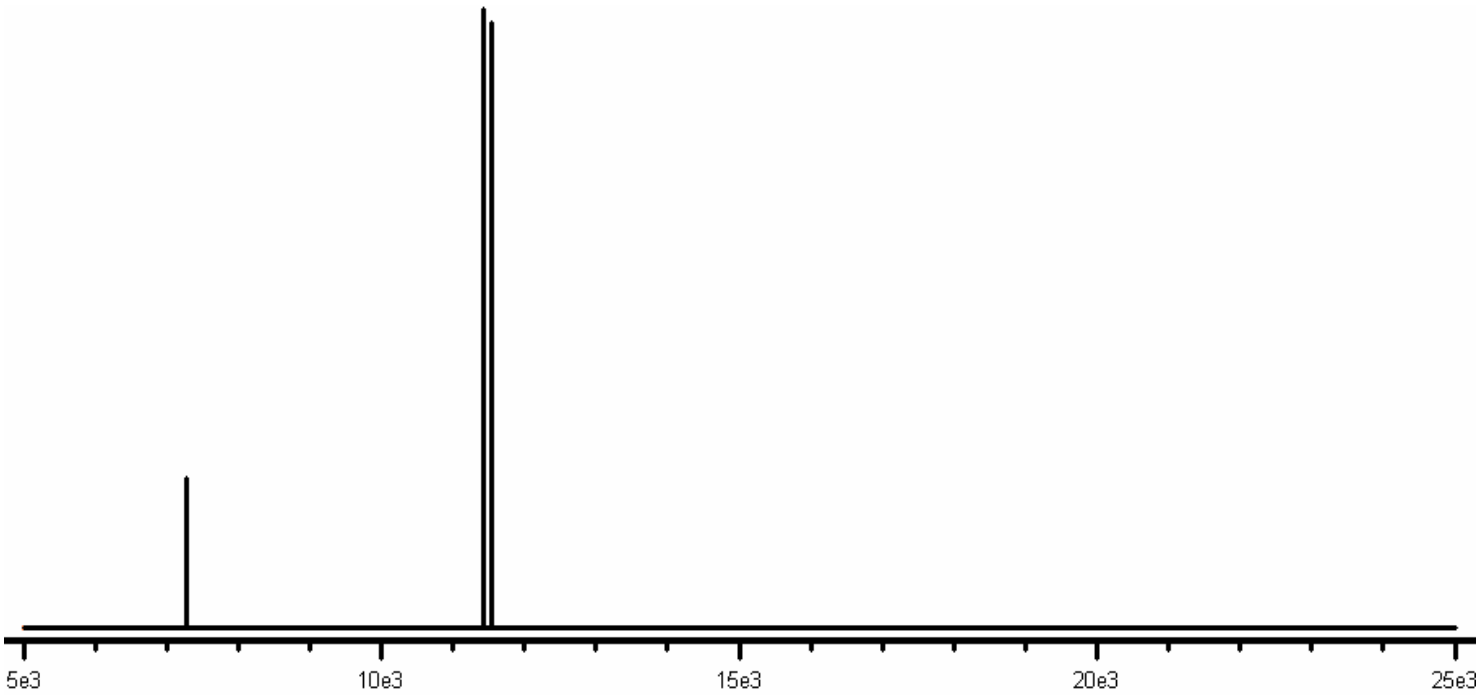

Figure 4. Zero-charge result.

The result is very clean and only 5 masses are reconstructed. Two are very weak and are not apparent in the above display. The reconstructed intensity for the three proteins is 66.3% of the intensity in the data. This means that the intensity that has not been reconstructed represents half the reconstructed intensity. Significantly, other methods would have forced this non-reconstructed intensity into the result and it would have appeared as artefacts that could have confused the interpretation.

The problem is formulated correctly in ***ReSpect***™. Here, the result intensity will nearly always be lower than that in the data because the overriding principle is that ***ReSpect***™ will only reconstruct masses for which there is evidence in the data. Anything else is irrelevant and correctly treated as noise. The fact that only three major masses have been reconstructed is a clear indication that none of the noise features and weak peaks fit the applied constraints – see Figure 5 above.

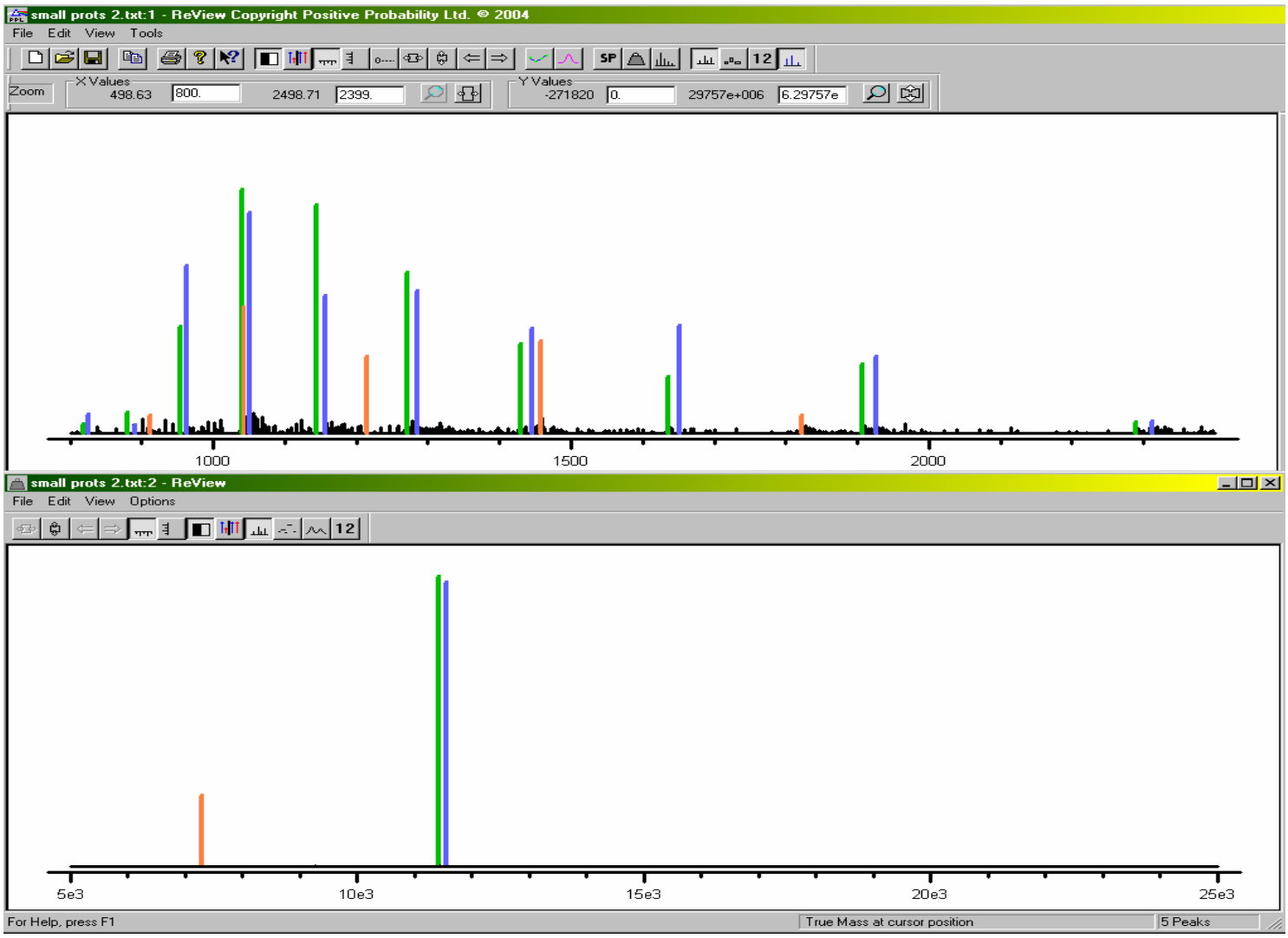

Figure 6. Evidence in the data for the reconstructed masses.

Conclusions

The ***ReSpect***™–based methodology only reconstructs masses for which there is evidence in the data. The results are therefore very clean and unambiguous compared with other methods.