Positive Probability Ltd

Note P9: Artefact-free Results – Method Comparison

Introduction

All data reconstruction methods aim to fit the data to a theoretical model within the noise level. Entropic and Bayesian methods have the fundamental constraint that the result must contain the same intensity as that in the data. Depending on the quality of the data, this can lead to serious artefacts, particularly if the data contains signals that are irrelevant to the problem being solved. ***ReSpect***™ is very different from other methods in that this constraint does not apply and the method will only report features in the result for which there is positive evidence. Therefore, results are much less prone to artefacts.

In this example, we compare the results obtained from PPL's ***ReSpect***™–based ***Discharge***™ interface and a popular maximum entropy method used for performing charge deconvolutions on ESI mass spectrometry data.

Data and Data Processing

The data are a test mixture of 4 purified proteins from various sources – insulin, ubiquitin, cytochrome C and myoglobin. The data are shown in Figure 1 below.

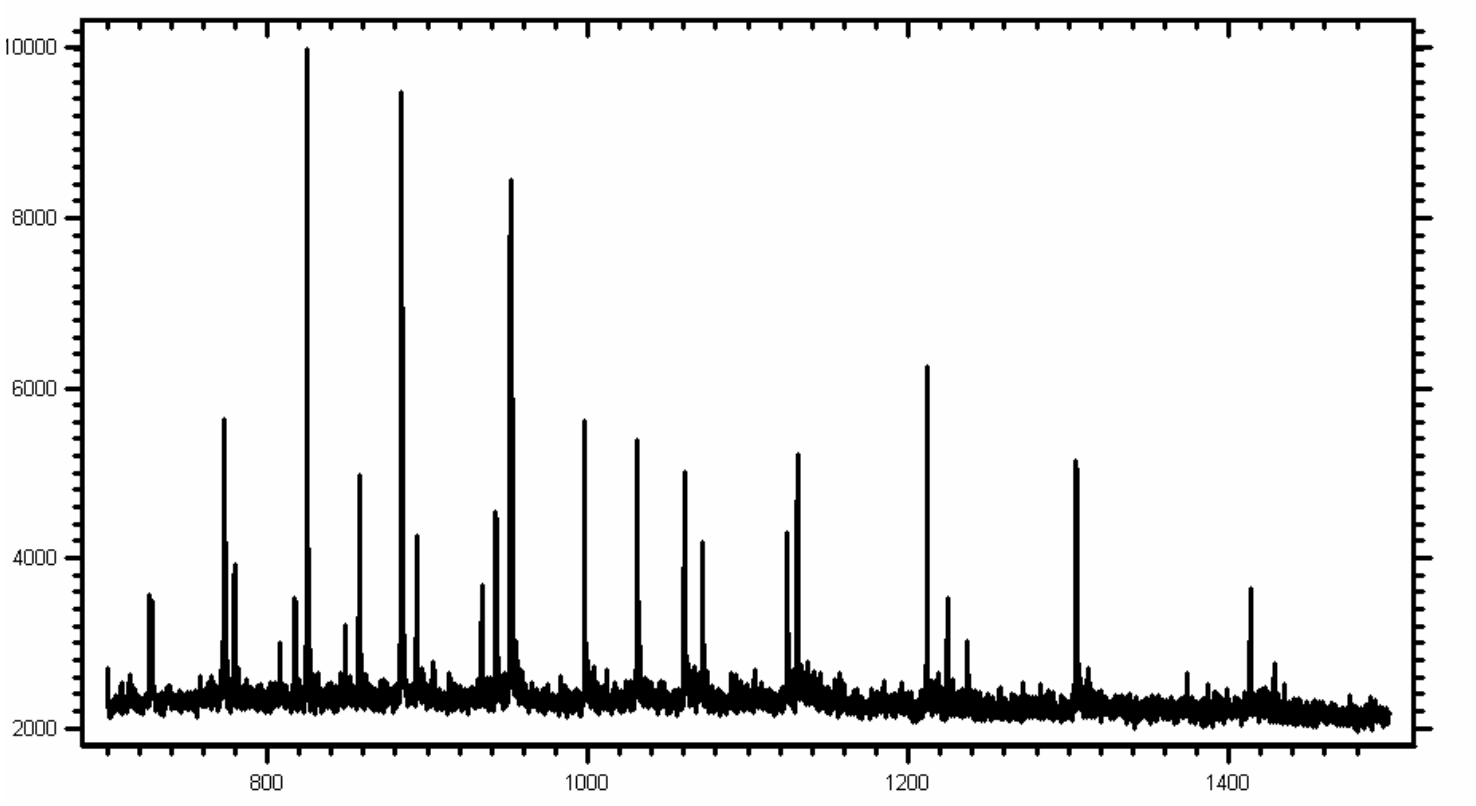

Figure 1. Data – Test mixture of 4 proteins.

All 5 electrospray envelopes are clear in the data.

PPL Processing

The data were baseline corrected using the ***Nadir***™ program. The deconvolution model was generated from the whole baseline corrected spectrum using ***Profile***™. Following deconvolution the resulting peak table was filtered using 1 SD and 68% confidence to remove obvious noise. The ***Discharge***™ program was then used to perform the charge deconvolution to provide a zero-charge result.

Maximum Entropy Processing

The manufacturer's programs were used for baseline correction and the model design. The maximum entropy program was then run to produce a charge deconvolved result. For the manufacturer's program there is no intermediate step and the data are directly transformed to the charge deconvolved result.

Results and Discussion

The ***Discharge***™ and maximum entropy results are shown in Figures 2 and 3 below.

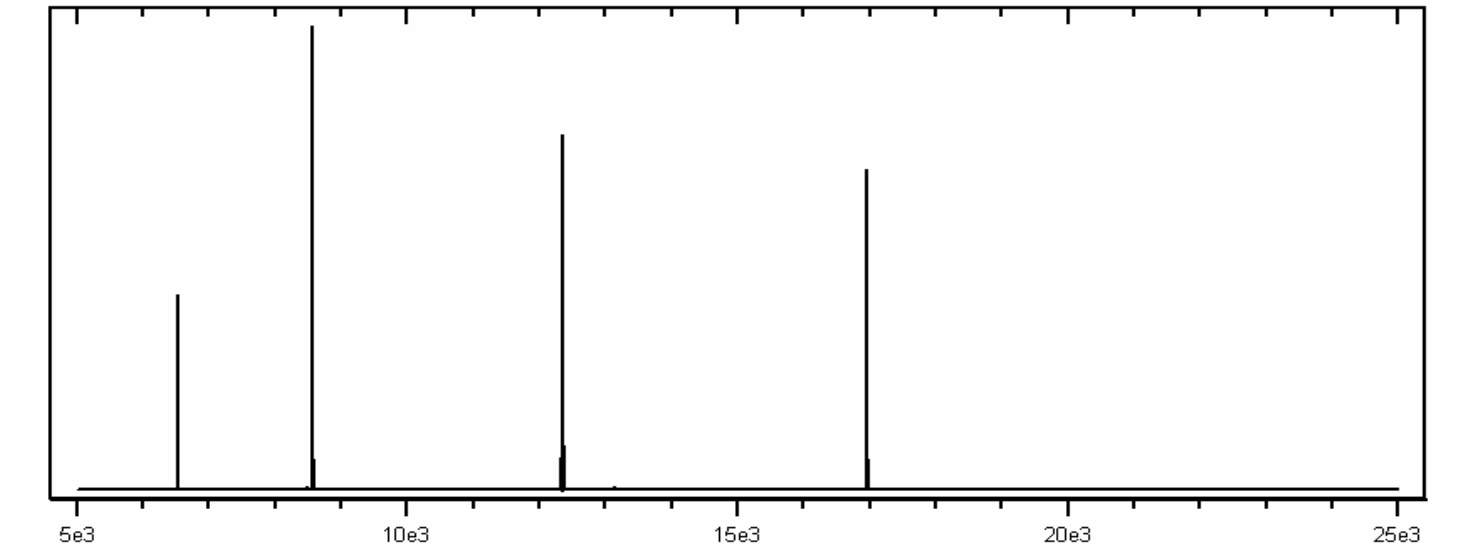

Figure 2. ***Discharge***™ *charge deconvolution.*

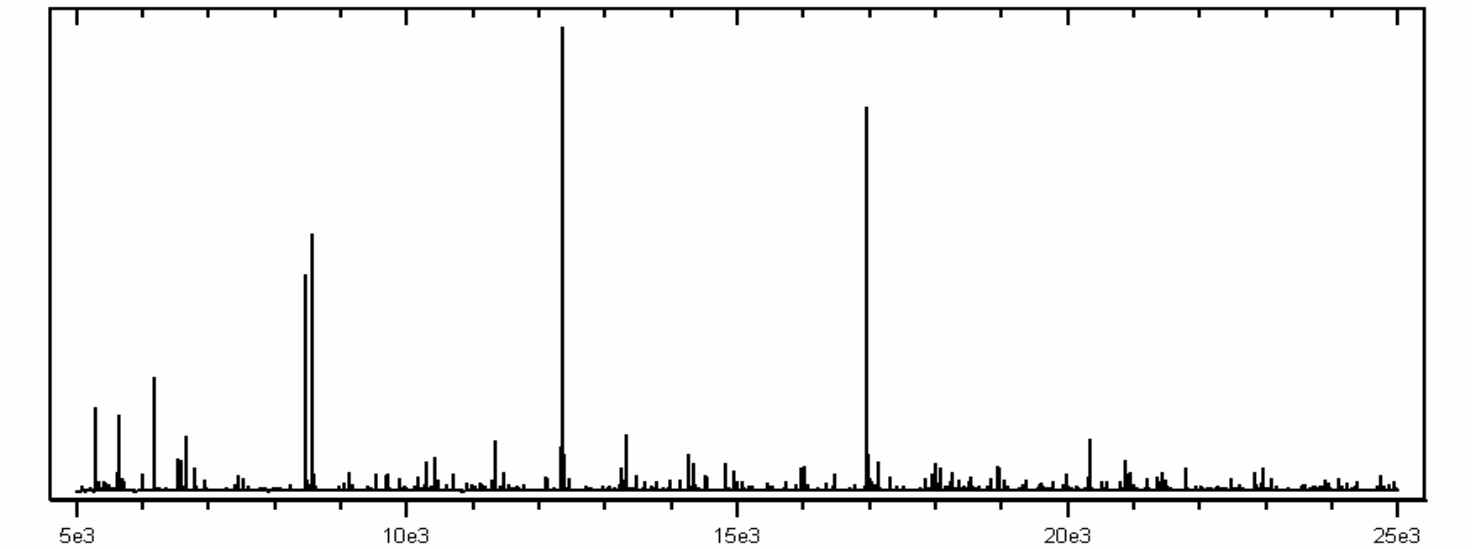

Figure 3. Maximum entropy charge deconvolution

The ***Discharge***™ result contains a total of 11 masses that fit the constraints and allowed errors. For these data the result contains just over half the intensity of that present in the data. This is because almost half the data intensity arises from positive-going noise features. However, these do not fit the constraints and their impact on the result is negligible. In addition, no harmonics (masses at half or twice the main masses) are observed.

The maximum entropy result contains 3742 peaks and its intensity is the same as that in the data. Almost all reconstructed masses are artefacts because of the intensity constraint. The insulin peak at ~6500 is much weaker than many of the artefacts. In addition, there is a very strong harmonic just below 8500 that arises from the intense myoglobin mass at close to 17000.

Zooming to each protein (Figures 4-7) shows how serious the maximum entropy artefacts are and how results can be ambiguous.

Insulin Mass

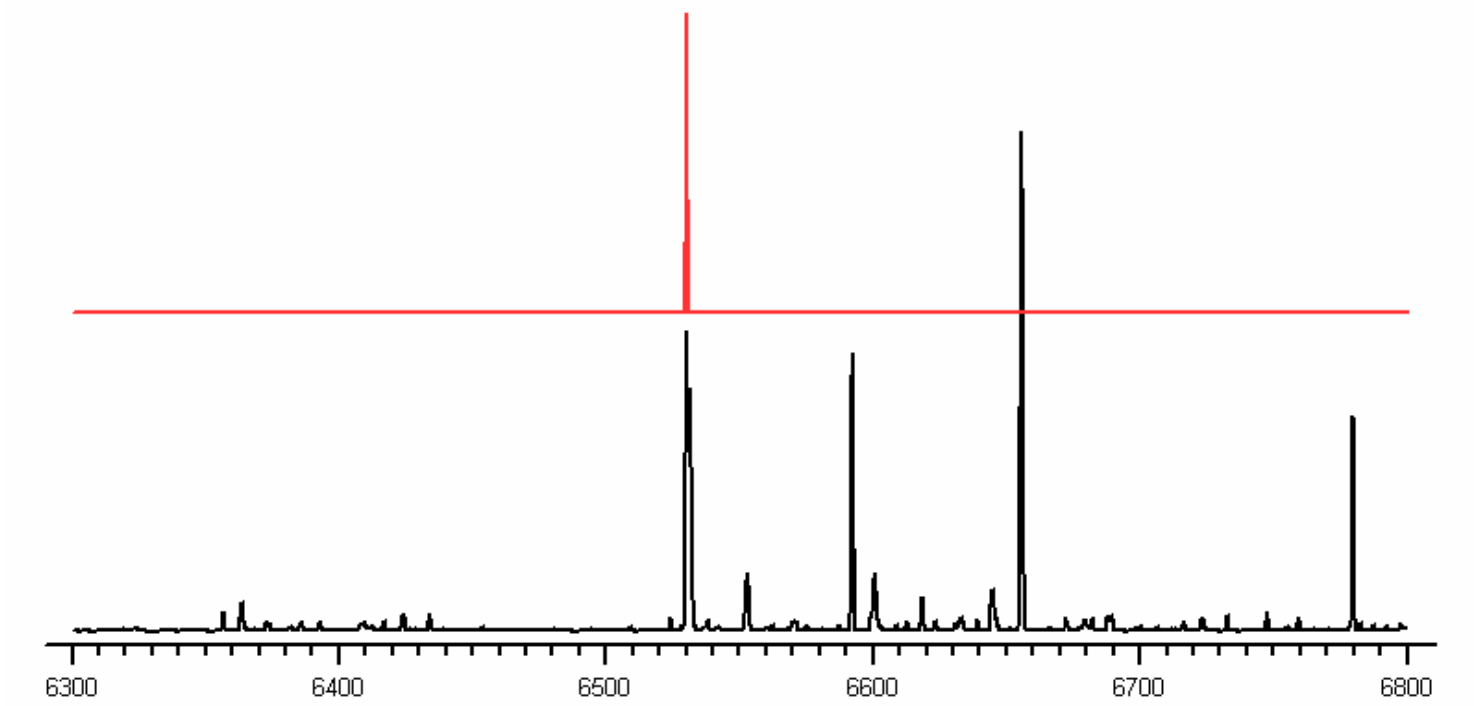

*Figure 4. Insulin mass. **Discharge**™ result (red) and maximum entropy (black).*
In the above diagram, the amplitude of the insulin masses has been scaled to be the same.

This part of the zero-charge spectrum clearly illustrates a serious problem with maximum entropy. In the displayed region there are three artefacts of similar amplitude to the reconstructed insulin mass. In addition, there are a number of weaker artefacts. In the absence of any other information the result would have to be taken at face value and all reconstructed peaks would have to be assumed to be genuine. Indeed, by computing the position of the charges in the data for the three intense artefacts, it is clear that there is no evidence for the reconstructed masses!

On the other hand, the ***Discharge***™ result shows only the insulin mass and there is no evidence in the data for any other masses in the displayed region.

Ubiquitin Mass

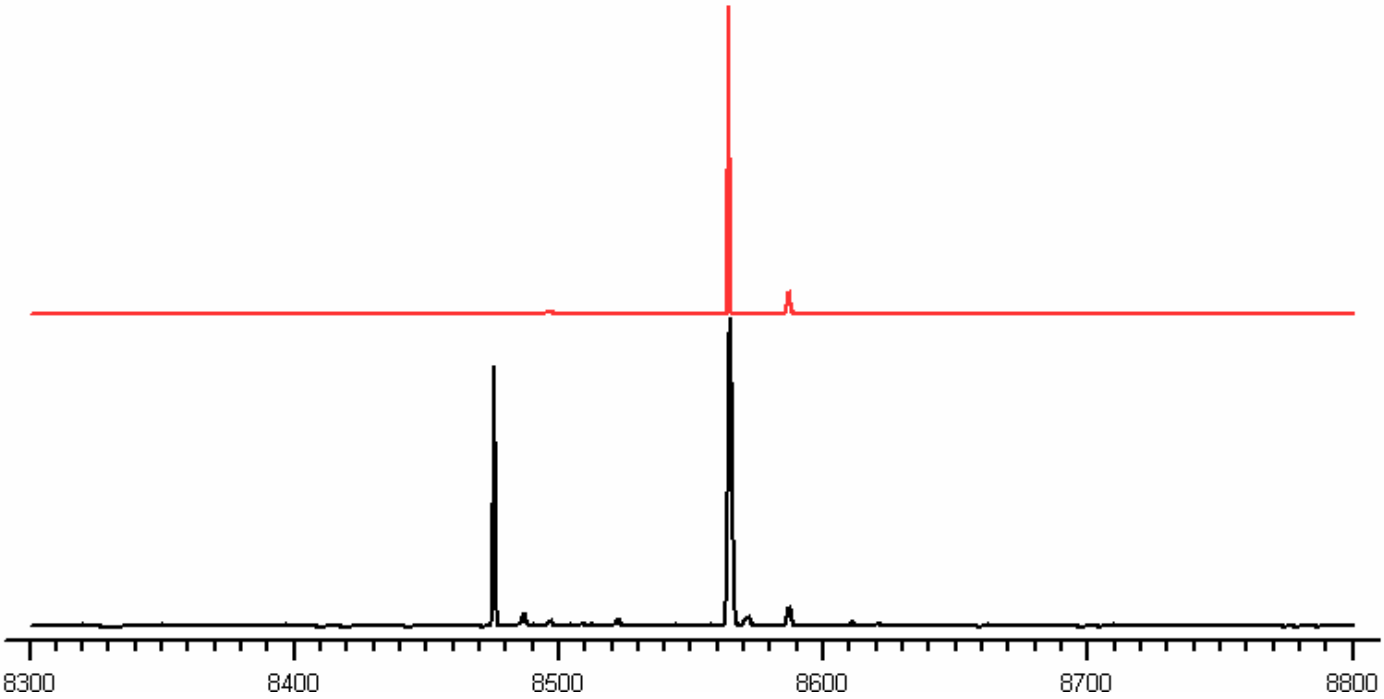

*Figure 5. Ubiquitin mass. **Discharge**™ result (red) and maximum entropy (black).*
In the above diagram, the amplitude of the ubiquitin masses has been scaled to be the same.

The sample contains much more ubiquitin than insulin and the zero-charge result is therefore "cleaner" for both processing methodologies. However, the maximum entropy result shows the presence of an intense artefact at 8475 (M+1) that is a harmonic of the myoglobin mass at 16951. Note that both methods show the presence of a Na adduct.

This illustrates another problem with entropic and Bayesian methods. They are designed to produce, within the noise level, the mathematically most plausible result based on the data, the model and the constraints. The result will also contain the minimum information consistent with the data. However, if one or more charges are absent in a potential electrospray series, these methods will reconstruct the missing charges even if it is unreasonable to do so – this is what mathematical plausibility dictates. The reverse is also true and alternate charges can also fit the data, generating a harmonic at half the true mass, whereas the former can lead to a harmonic at double the true mass. For the data described here, the maximum entropy methodology can fit the intensities in the ubiquitin charge envelope more closely by also fitting a harmonic at half the mass.

The ***ReSpect***™ fitting is subtly different. As stated above, entropic and Bayesian methods generate a result that contains the minimum information consistent with the data. ***ReSpect***™ and its ***Discharge***™ interface generate a result that contains the minimum *features* consistent with the data. Each feature is therefore "tested" to check whether it is making a positive contribution to the result or not. This difference in approach dramatically reduces the potential for harmonic artefacts to be generated.

One might suppose that this means that a monomer will not be reconstructed if its dimer is also present in the data. In this case the envelope intensity profile is uneven and so both masses are required to fit the data correctly.

Cytochrome Mass

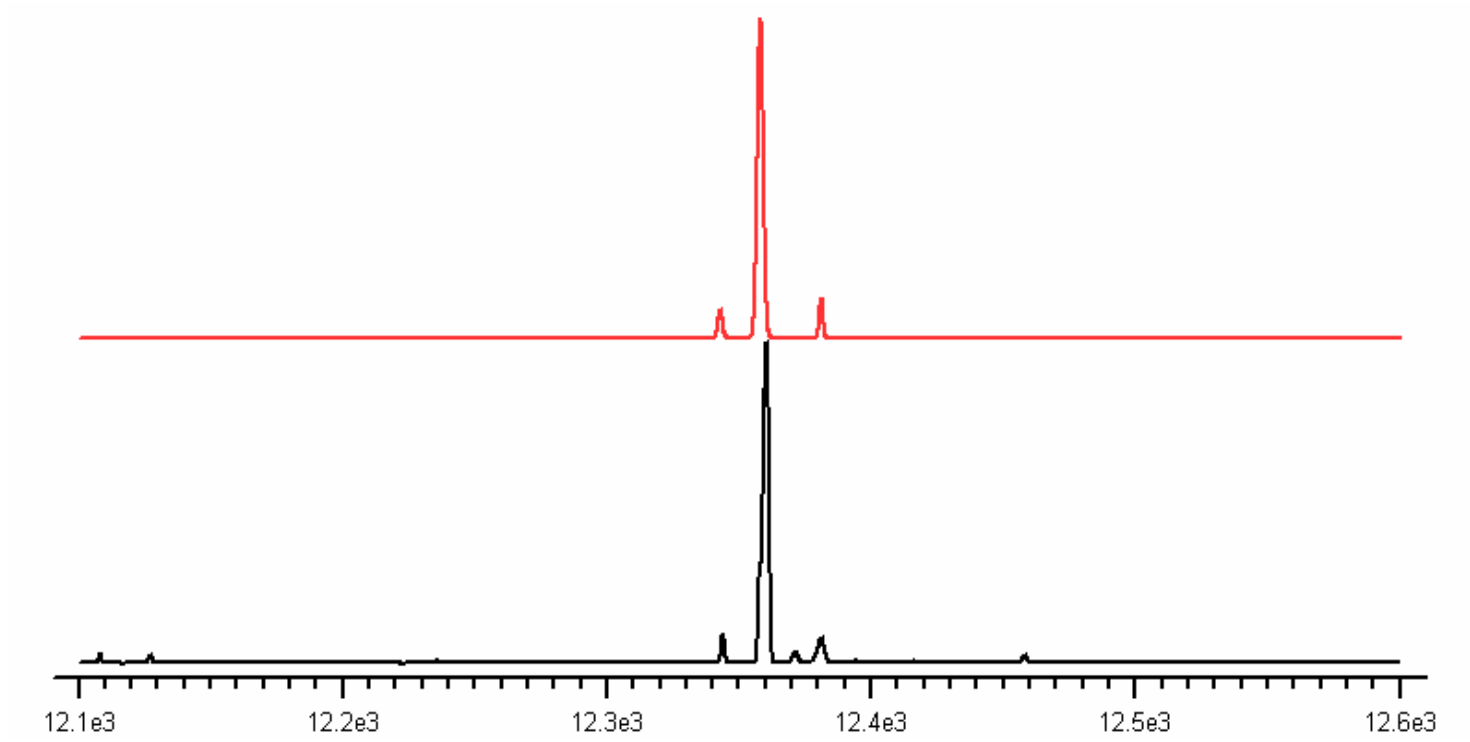

*Figure 6. Cytochrome mass. **Discharge™** result (red) and maximum entropy (black).*
In the above diagram, the amplitude of the cytochrome masses has been scaled to be the same.

Myoglobin Mass

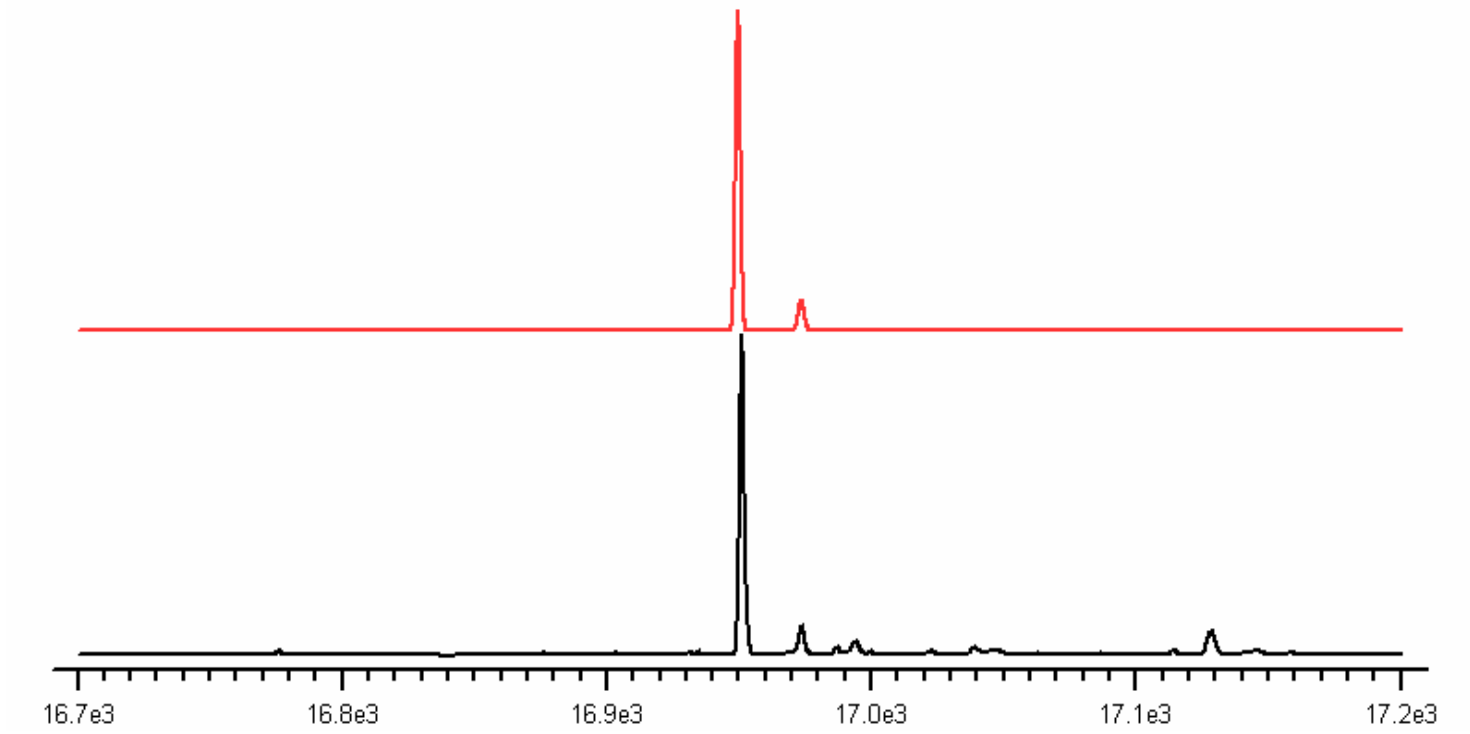

*Figure 7. Myoglobin mass. **Discharge™** result (red) and maximum entropy (black).*
In the above diagram, the amplitude of the myoglobin masses has been scaled to be the same.

The results shown in Figures 6 and 7 show very clean results using the ***Discharge***™ methodology and a Na adduct is seen with both methods. Although not apparent on the zoom scale of Figure 6, the maximum entropy result has an intense partially resolved shoulder on low mass side of the cytochrome peak. Either this or the main peak is therefore artefactual.

The high mass side of the myoglobin peak shows many weak artefacts in the maximum entropy result that are absent in the result from the methodology described here.

Figure 8 below shows which m/z peaks have been used to reconstruct the zero-charge result for the ***ReSpect***™–based methodology. Na adducts are ignored. It is clear that all peaks above the noise level have been used and accounted for. The residual noise is clearly responsible for most of the maximum entropy artefacts.

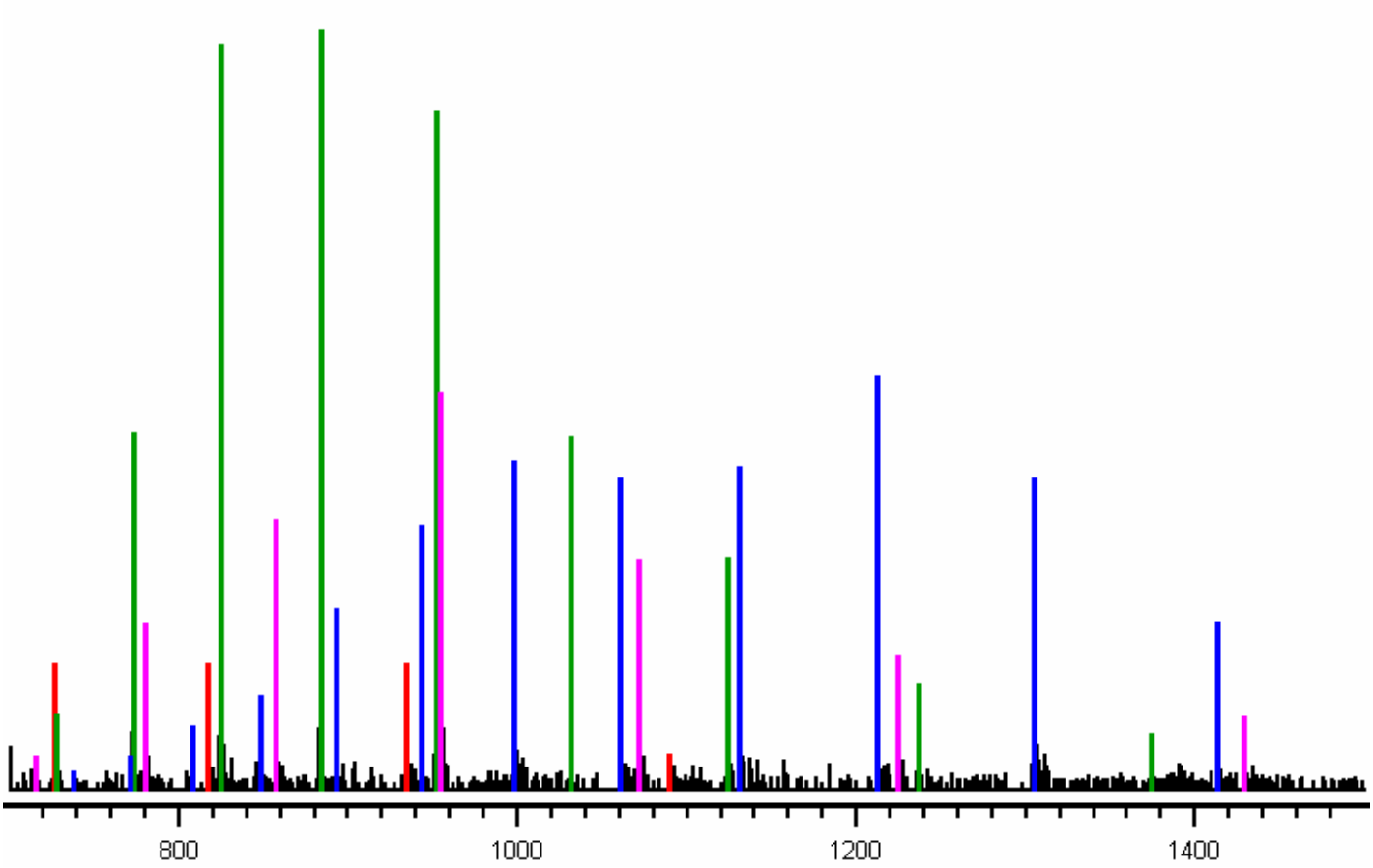

*Figure 8. Evidence in the data (spike plot) for the reconstructed **Discharge**™ masses. Red, insulin; Magenta, ubiquitin; Green, Cytochrome; Blue, myoglobin.*

Another fundamental problem of entropic and Bayesian methods is that the intensity in the result must be the same as that in the data. It follows that there will be an increase in the artefacts by computing the maximum entropy result over narrow output mass ranges. Figures 9-12 compare the maximum entropy results for the whole output mass range from 5-25 kDa and for the ranges displayed in Figures 4-7.

In the figures, the red lines represent the baseline. The black trace is the computation over the range 5-25 kDa and the blue trace is the computation over the displayed region. Note that there are artefacts at the limits of the range for the narrow regions.

Insulin Mass (~6530)

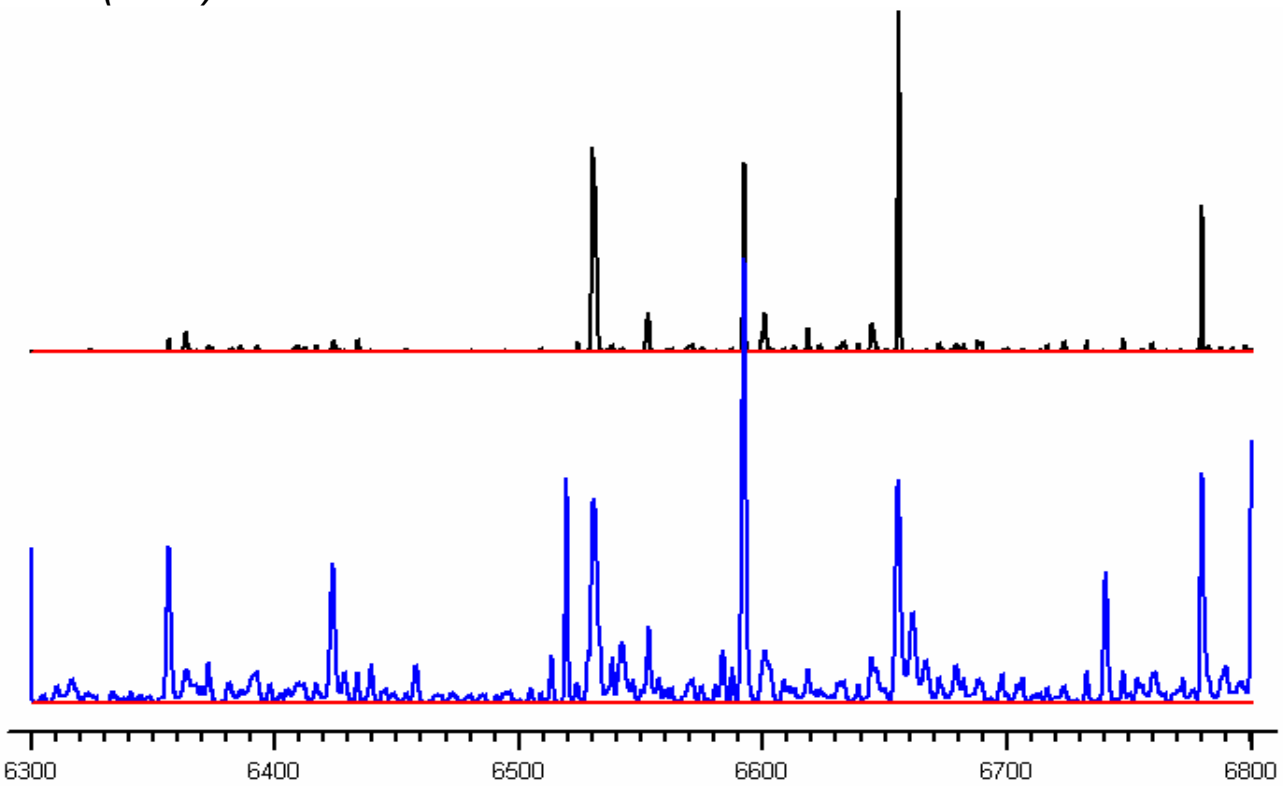

Figure 9. Maximum entropy. Black, 5-25 kDa; Blue, 6.3-6.8 kDa.

Ubiqutin Mass

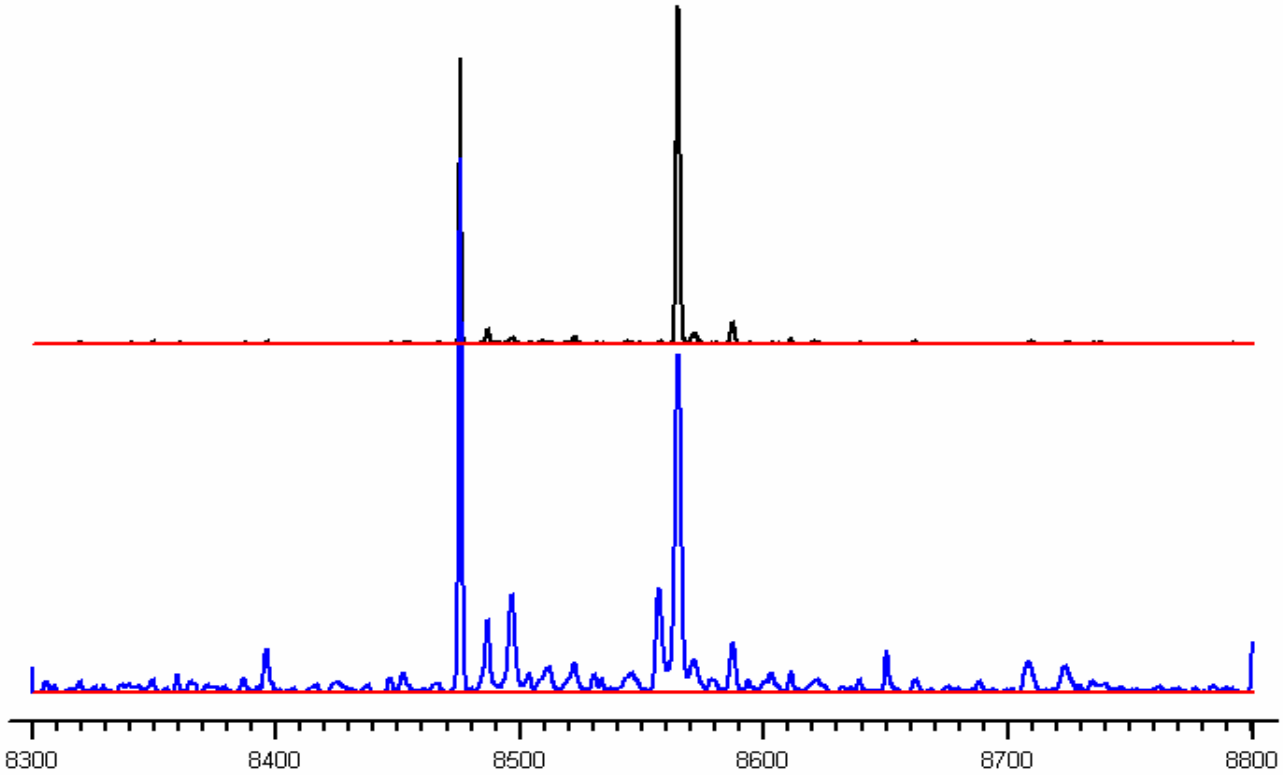

Figure 10. Maximum entropy. Black, 5-25 kDa; Blue, 8.3-8.8 kDa.

Cytochrome Mass

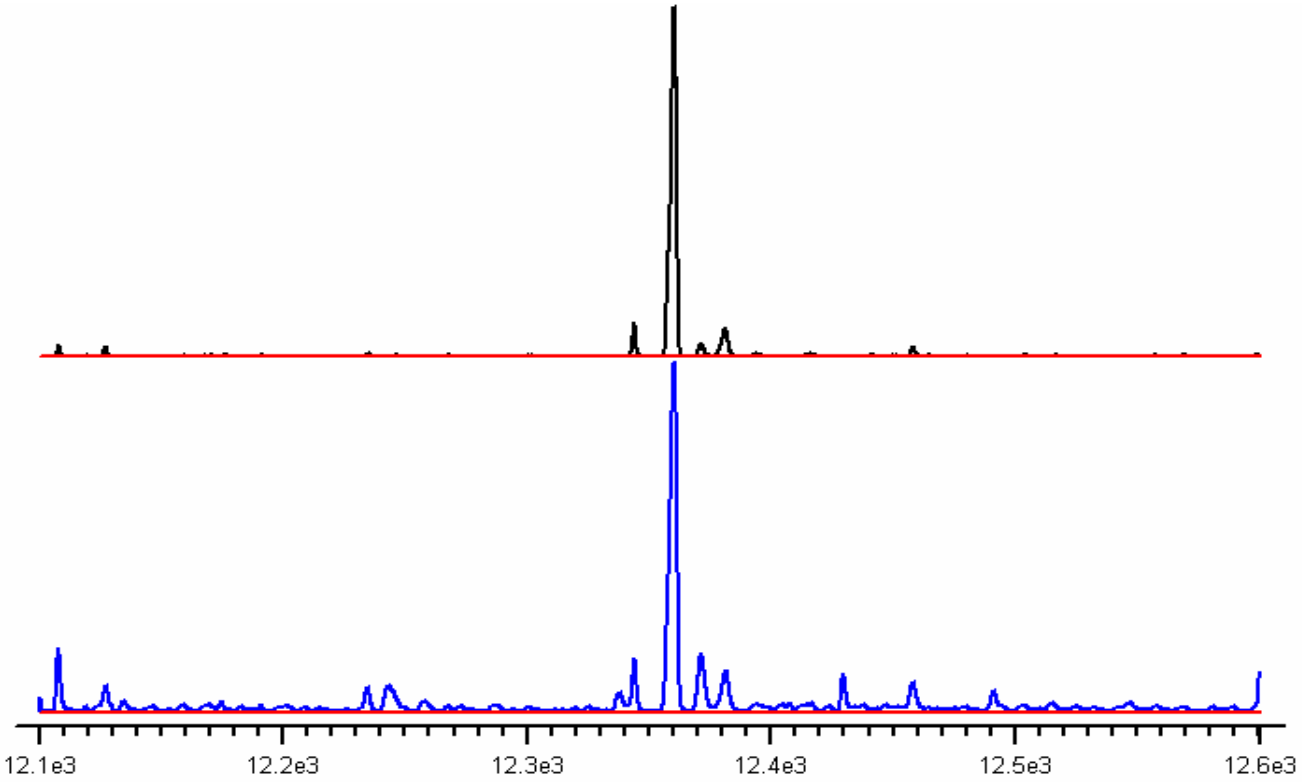

Figure 11. Maximum entropy. Black, 5-25 kDa; Blue, 12.1-12.6 kDa.

Myoglobin Mass

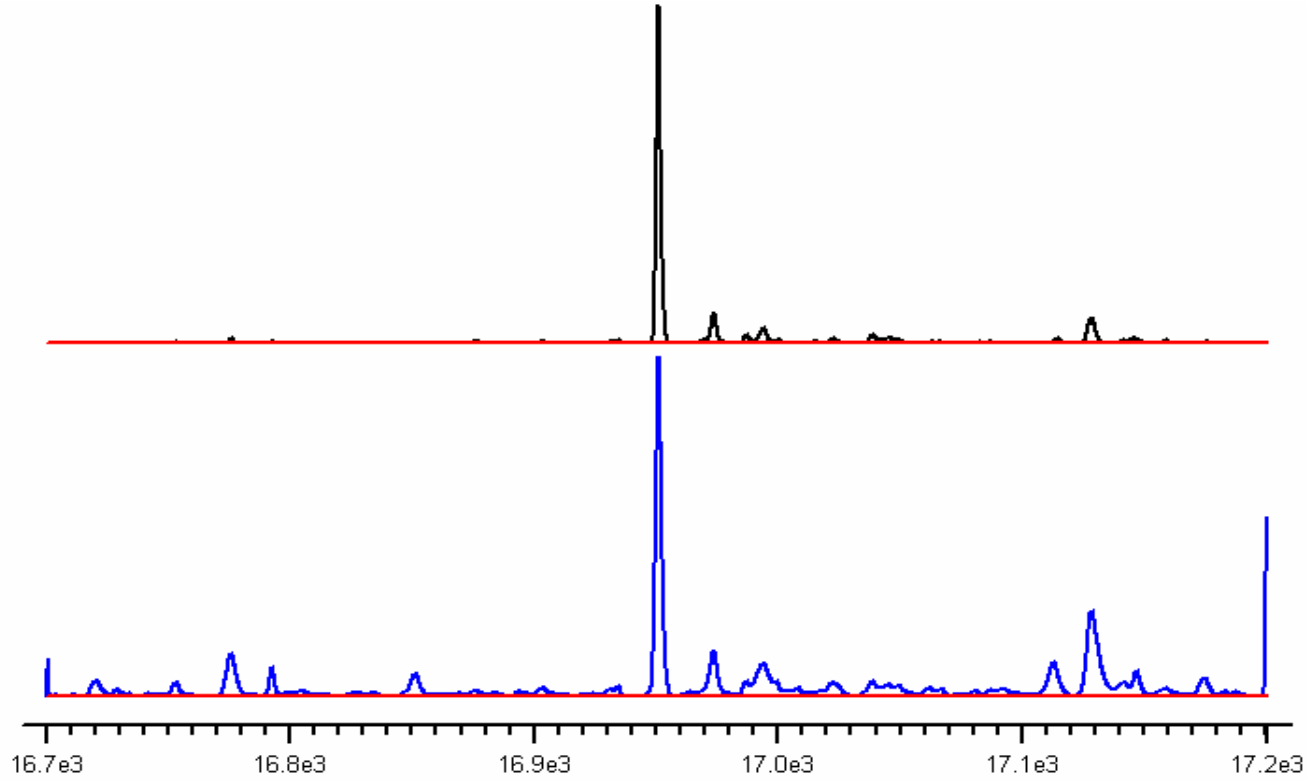

Figure 12. Maximum entropy. Black, 5-25 kDa; Blue, 16.7-17.2 kDa.

The substantially increased level of artefacts is clear in all maximum entropy results. The corresponding ***Discharge™*** results are shown in Figures 13-16. Results have the same amplitude and are overlaid.

Insulin Mass

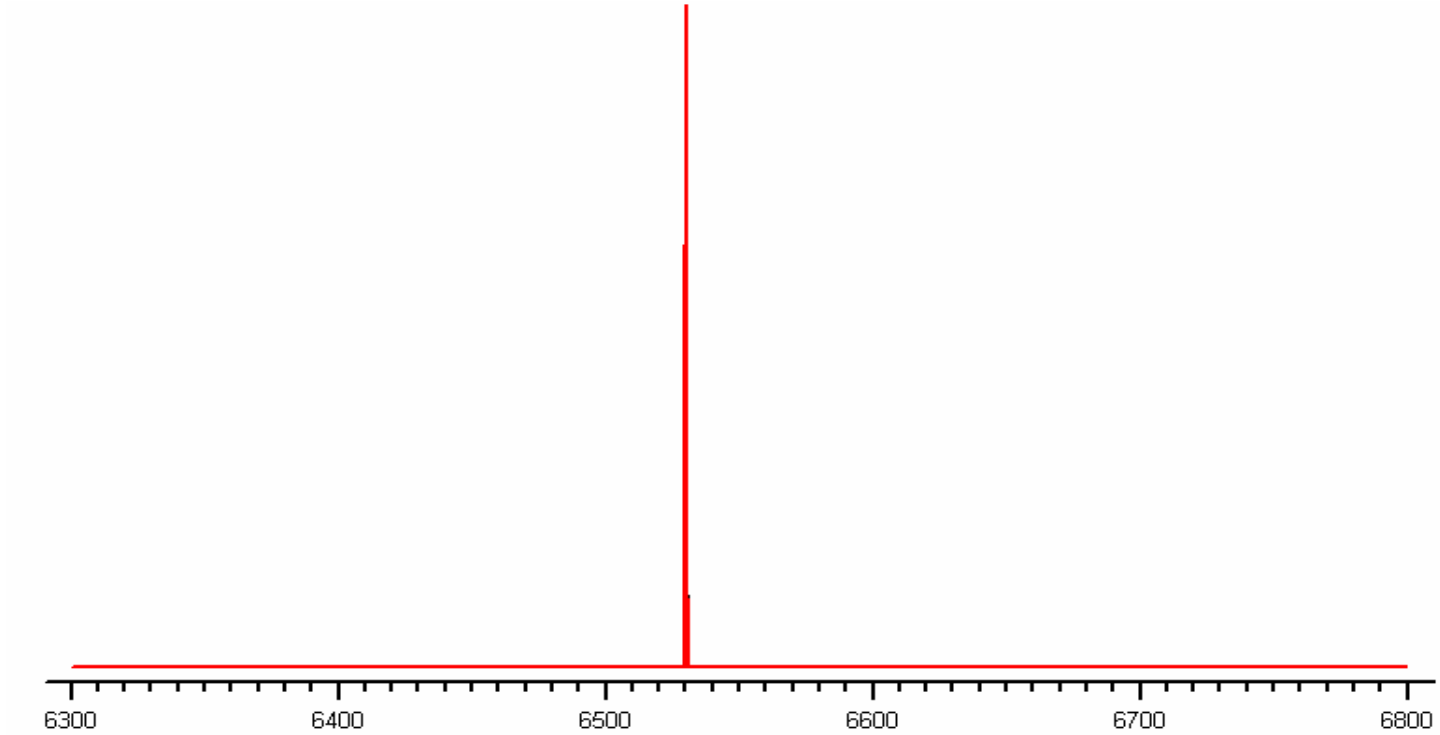

Figure 13. ***Discharge™****. Black, 5-25 kDa; Red, 6.3-6.8 kDa.*

Ubiquitin Mass

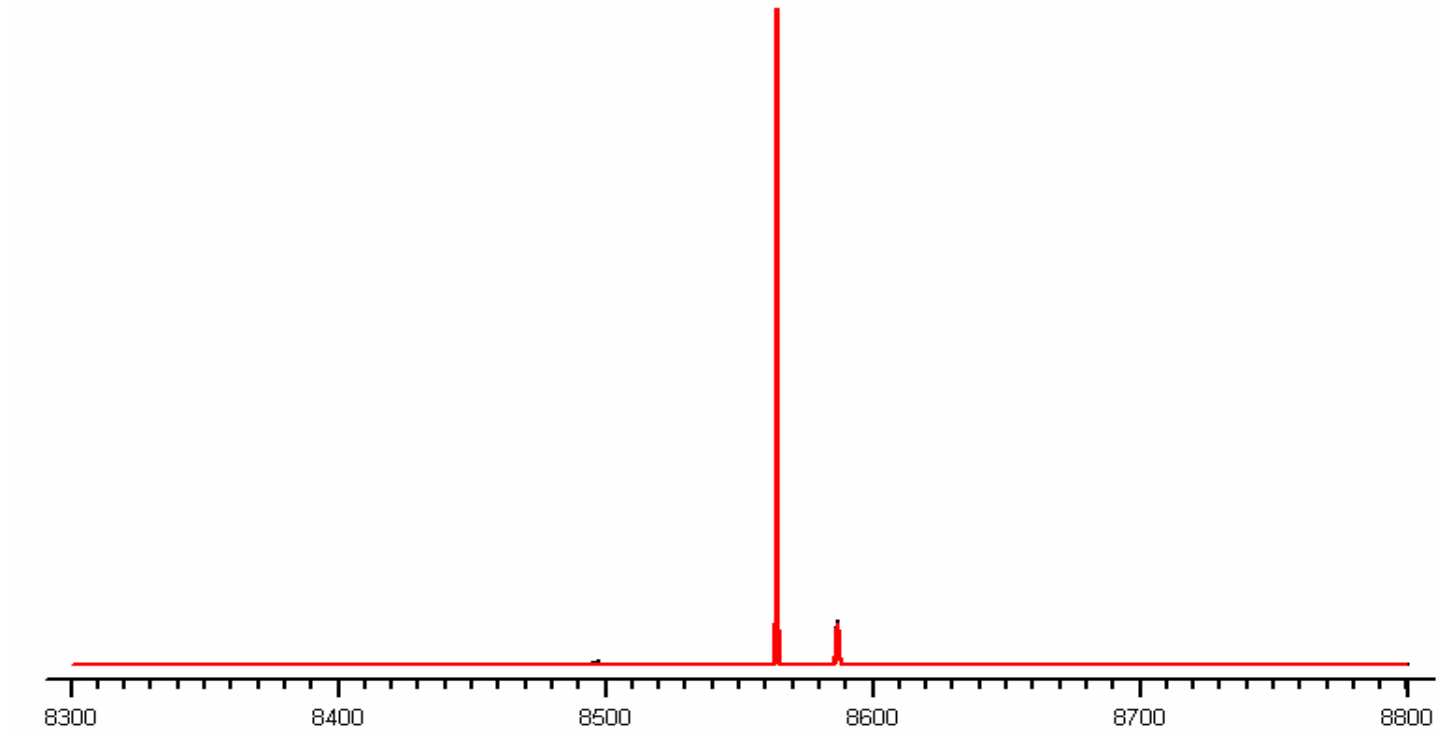

Figure 14. ***Discharge™****. Black, 5-25 kDa; Red, 8.3-8.8 kDa.*

Cytochrome Mass

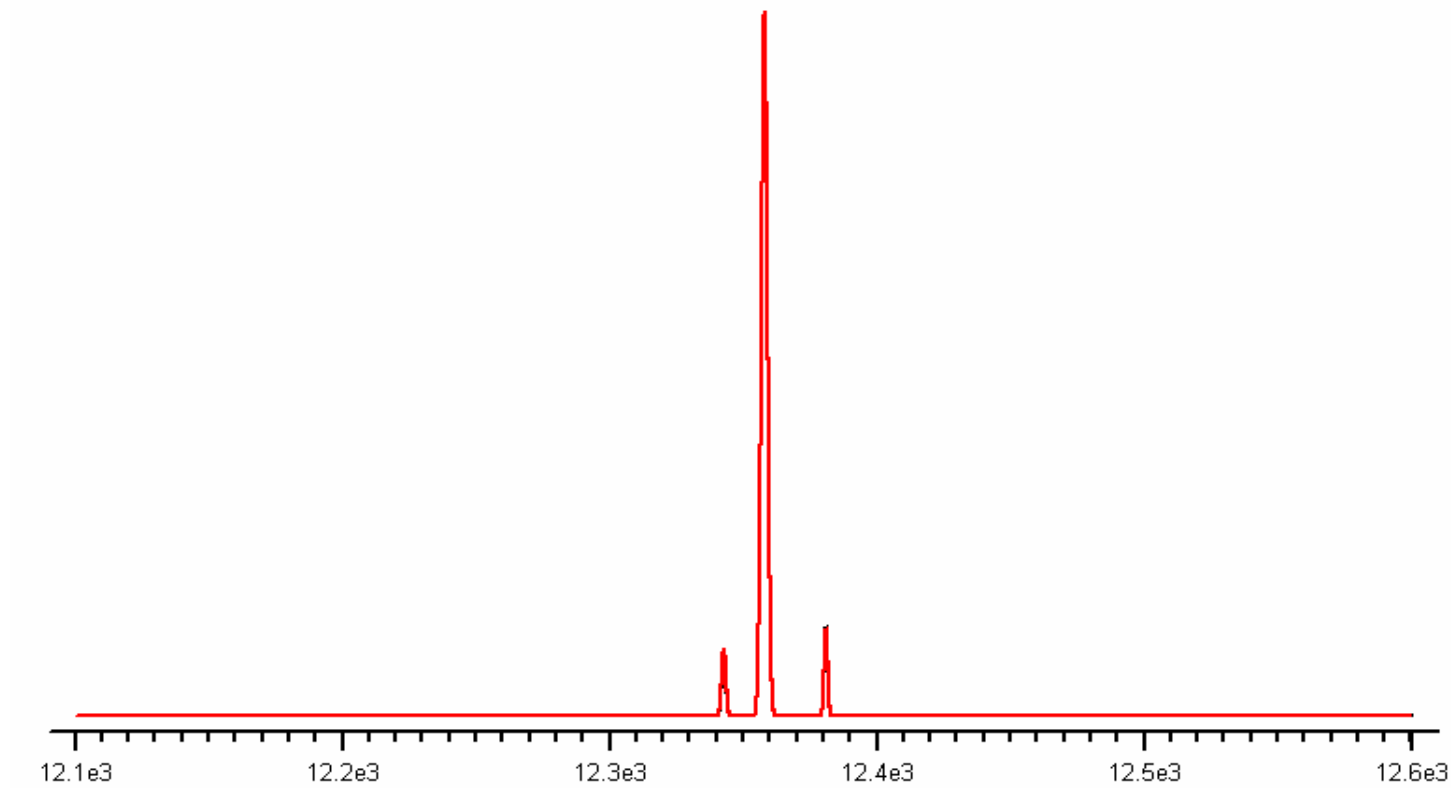

Figure 15. ***Discharge™****. Black, 5-25 kDa; Red, 12.1-12.6 kDa.*

Myoglobin Mass

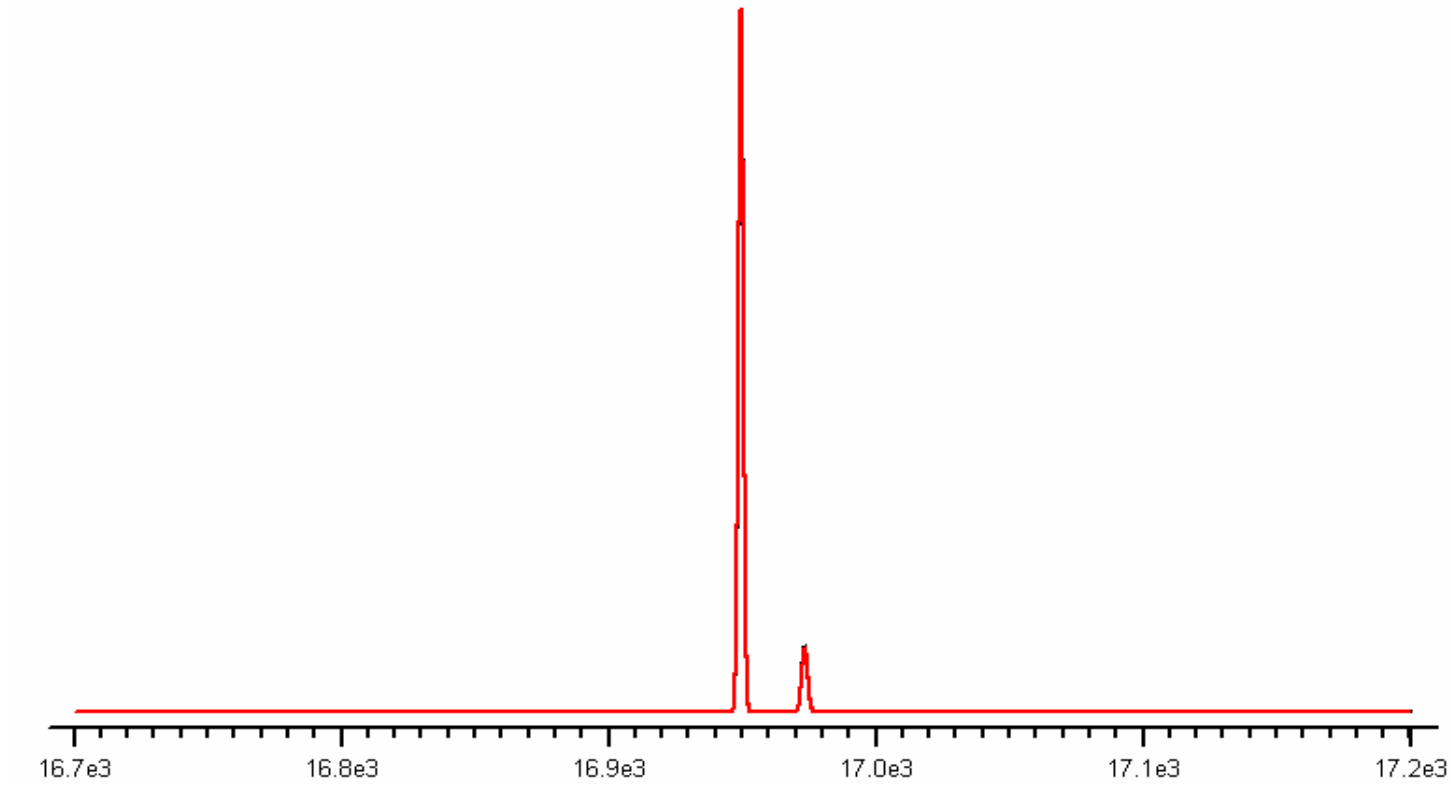

Figure 16. ***Discharge™****. Black, 5-25 kDa; Red, 16.7-17.2 kDa.*

The ***Discharge***™ results are virtually identical regardless of the output mass range for the computation and the only way to see this clearly is by overlaying the results for Figures 13-16. By using inputs that are not very sensible it is naturally possible for harmonic peaks to be generated because there will be evidence in the data for them if the main mass is outside the reduced output mass range. For example, if an unrealistically low number of adjacent charges are selected for the ubiquitin mass, then the even charges for the myoglobin will be interpreted as a plausible mass to account for more of the data. However, using realistic inputs will always reduce the potential for artefacts and harmonics.

This example has highlighted some of the problems with entropic methods. The ***ReSpect***™–based methodology and its ***Discharge***™ interface have been specifically designed to minimise these problems so that clean, unambiguous zero-charge results are obtained.

Conclusions

The ***ReSpect***™–based methodology has the following benefits over maximum entropy methods:

1. Artefacts and harmonics are much reduced and frequently absent.
2. By not forcing the result to have the same intensity as that in the data, zero-charge results are virtually independent of the output mass range over which the result is computed.
3. There is always positive evidence in the data for any reconstructed zero-charge mass.