Positive Probability Ltd

Note P4: Centroiding – The Advantage of Reconstructions

Introduction

Centroiding may be applied to any spectrum or chromatogram but it is applied with fervor in mass spectrometry because the method is fast and it is able to keep up with the vast rate that data may be generated. Even so, traditional centroiding programs are unnecessarily complicated. They are designed to centroid all types of mass spectra in situations from very high to very low S/N and for isolated and overlapped peaks. There are generally numerous parameters for the user to adjust that allow, for example, shoulders to be centroided. Some programs even require different parameters to be set for centroiding shoulders on the left of a peak compared with those on the right!

The biggest problem with these programs is that they perform poorly for low S/N data and some form of smoothing is generally applied first. Information that was present in the data to begin with is lost and attempts to recover it have resulted in highly complicated programs with more options than most users are prepared to be bothered with.

Locator™ is a novel centroiding program that only requires an estimate of the peak width as its input. The program is very tolerant of this input and only a crude estimate is required. ***Locator™*** performs a very rapid reconstruction of the data to reduce the noise without broadening peaks. The peak positions and intensities are automatically extracted from the noise reduced result, taking into account any variation in the noise level across the data. The resulting peak table may then be used as the input for other applications. Computation speed is comparable with traditional methods.

A unique feature is the program is that it provides estimates of the position and intensity errors. Therefore, data thresholding is irrelevant and the peak table may be filtered by selecting a suitable significance and/or confidence level.

Data and Data Processing

The data presented here are the MALDI spectrum of a 5 kDa polymer. The baseline and noise level change substantially across the data. Two experiments were performed. In the first the data were centroided using the program in Data Explorer. The intensity threshold was set at 2% of the base peak area and the resolution was set in accordance with the data peak width. The top 50% of each peak was centroided. In the second experiment the peak width was estimated near the centre of the data. This was used as the only input to the program. The baseline and varying noise level were automatically taken into account during the data reconstruction and generation of the peak table. Peaks with a significance greater or equal to 1 s.d. were retained.

Figure 1 shows the raw data and the baseline corrected result using the ***Nadir™*** baseline correction program. Note the detector dead time and ringing at the beginning of the acquisition. After baseline correction the varying noise level is very apparent.

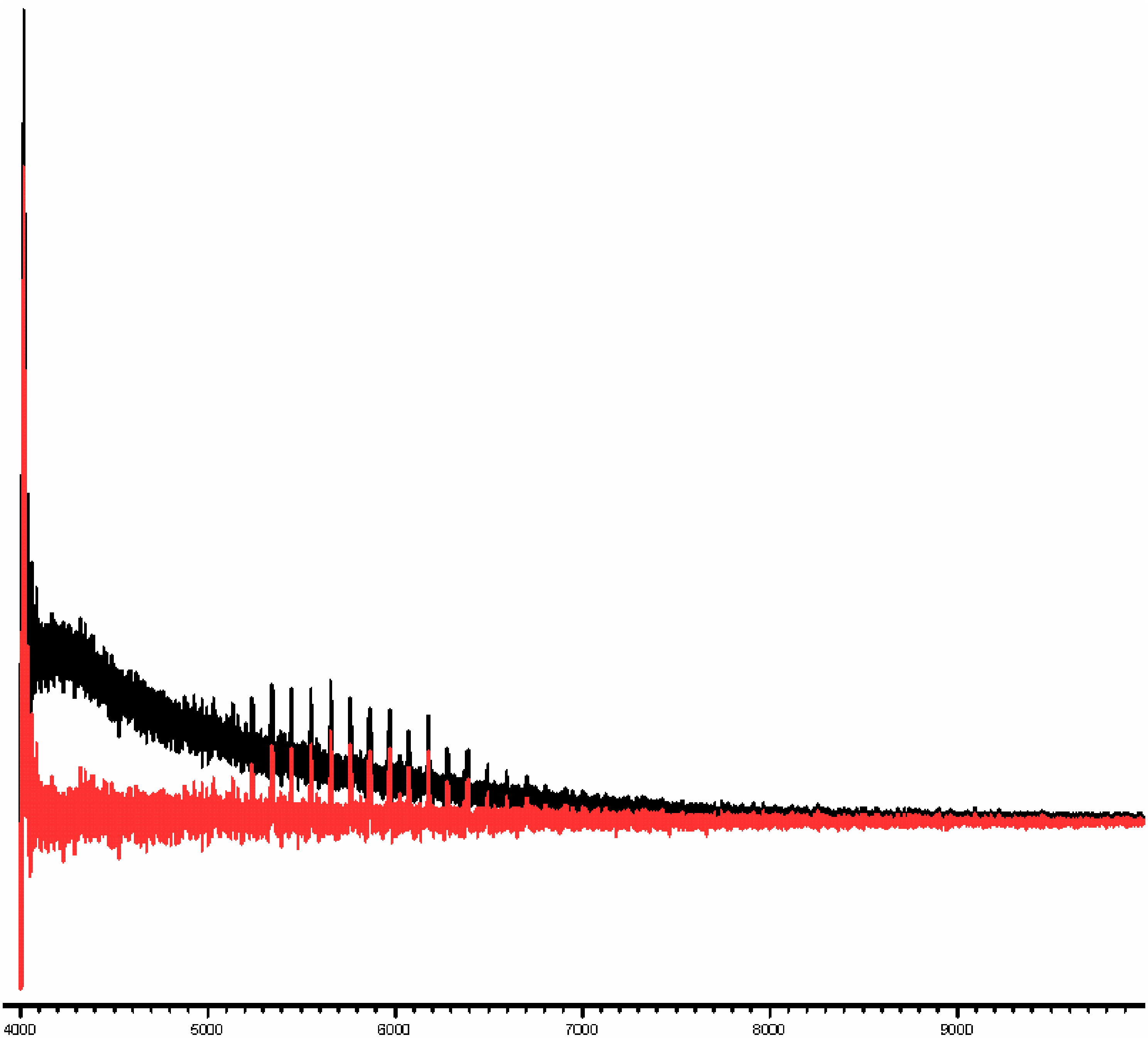

Figure 1. Data (black) and baseline corrected spectrum (red).

Results and Discussion

Conventional centroiding methods require the user to set a threshold and, because the changing noise level is not taken into account, numerous centroids are reported where the noise is high and this is clearly apparent in the centre trace of the Figure 2. The threshold was set at 2% of the maximum peak area so that the genuine weak peaks to the high m/z side of the polymer envelope where retained. Of course, thresholds of this type are arbitrary and several attempts are required to retain the desired information. The result using the data reconstruction approach is noticeably much cleaner, there being many fewer centroids. The weak peaks in the region 7000-8000 are also visually clearer.

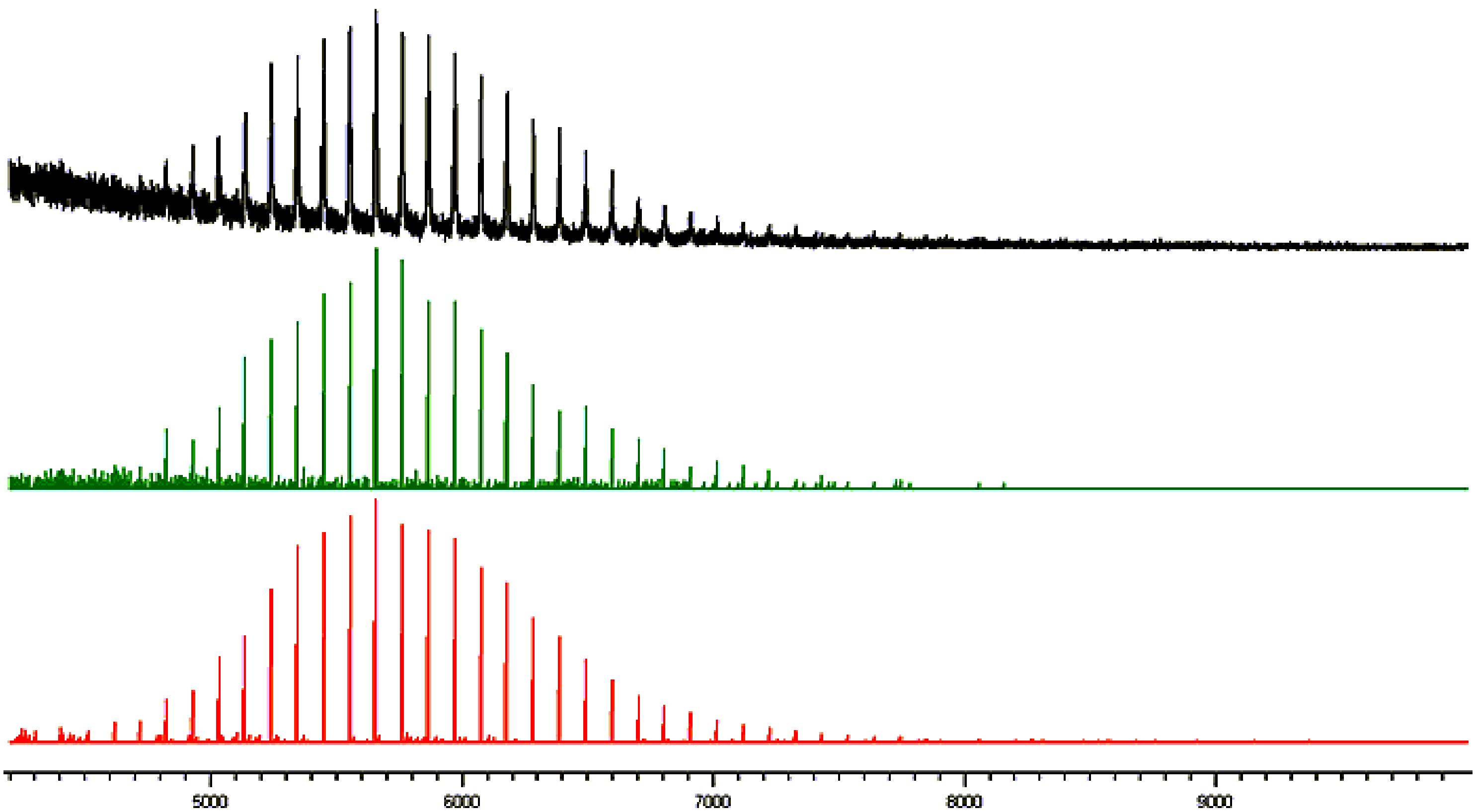

Figure 2. Top: Raw data; Centre: Data Explorer; Bottom: Reconstruction centroids

For the displayed region in Figure 2, Data Explorer detects 310 peaks and noise features because the varying noise level is not taken into account. The ***Locator*™** fast reconstruction method only finds 159 and is much more discerning about the features it detects. The polymer intensity profile is also expected to be smooth and much of the unevenness using Data Explorer is removed in the reconstruction. Figure 3 displays all reconstructed centroids and their 1 SD error bars for the main polymer envelope.

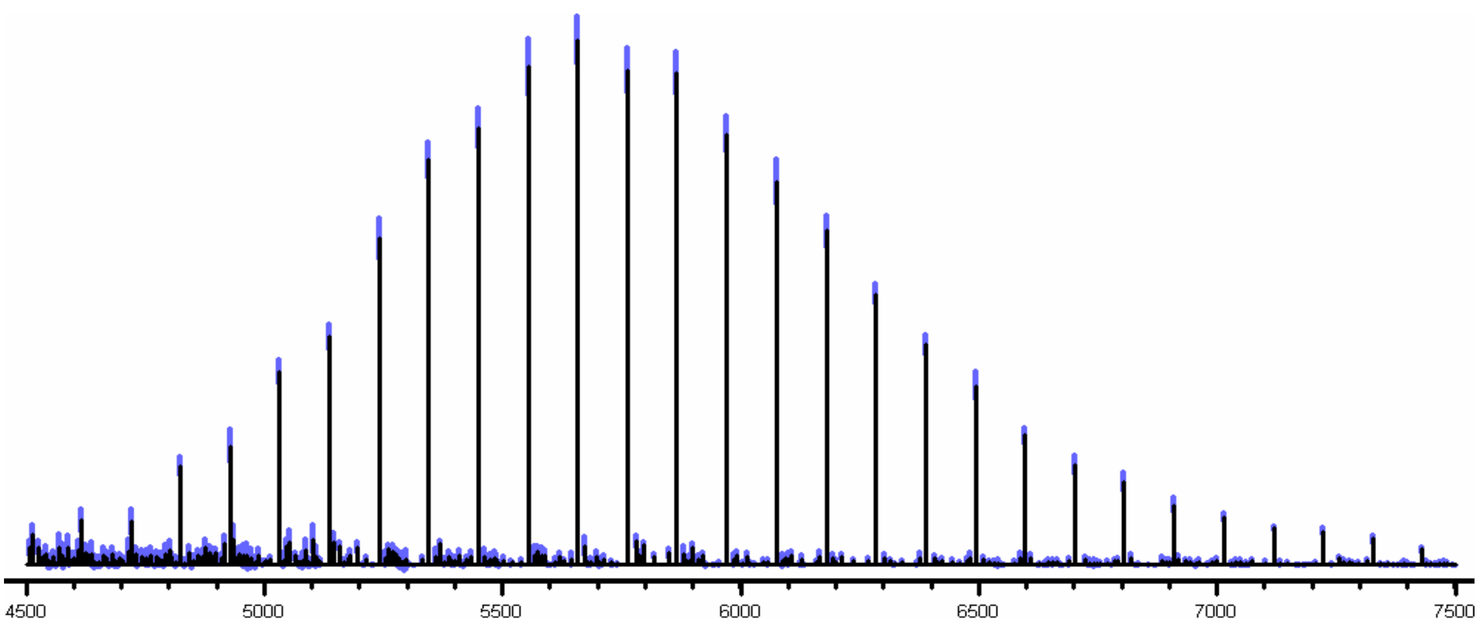

Figure 3. All reconstructed centroids and their 1 SD error bars.

Any conventional thresholding would not take into account the way the noise changes across the data. Figure 4 shows the effect of applying filters of 2 SD 95% confidence. Note how almost all noise centroids are rejected at the low m/z end of the data and yet some are retained at the high m/z end, due to filtering according to the true significance of each feature.

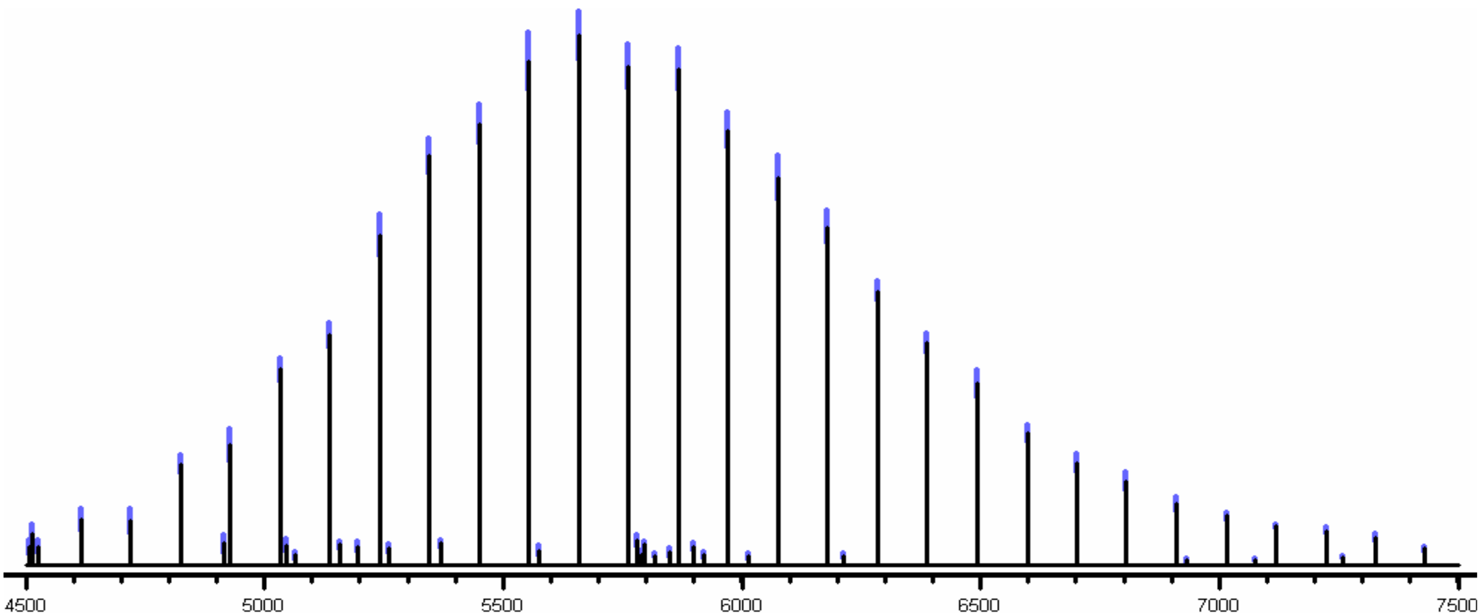

Figure 4. Filtering according to significance and confidence.

Figure 5 shows the evidence in the data for the "noise centroids" near the centre of the polymer envelope. Note that there is positive evidence for them.

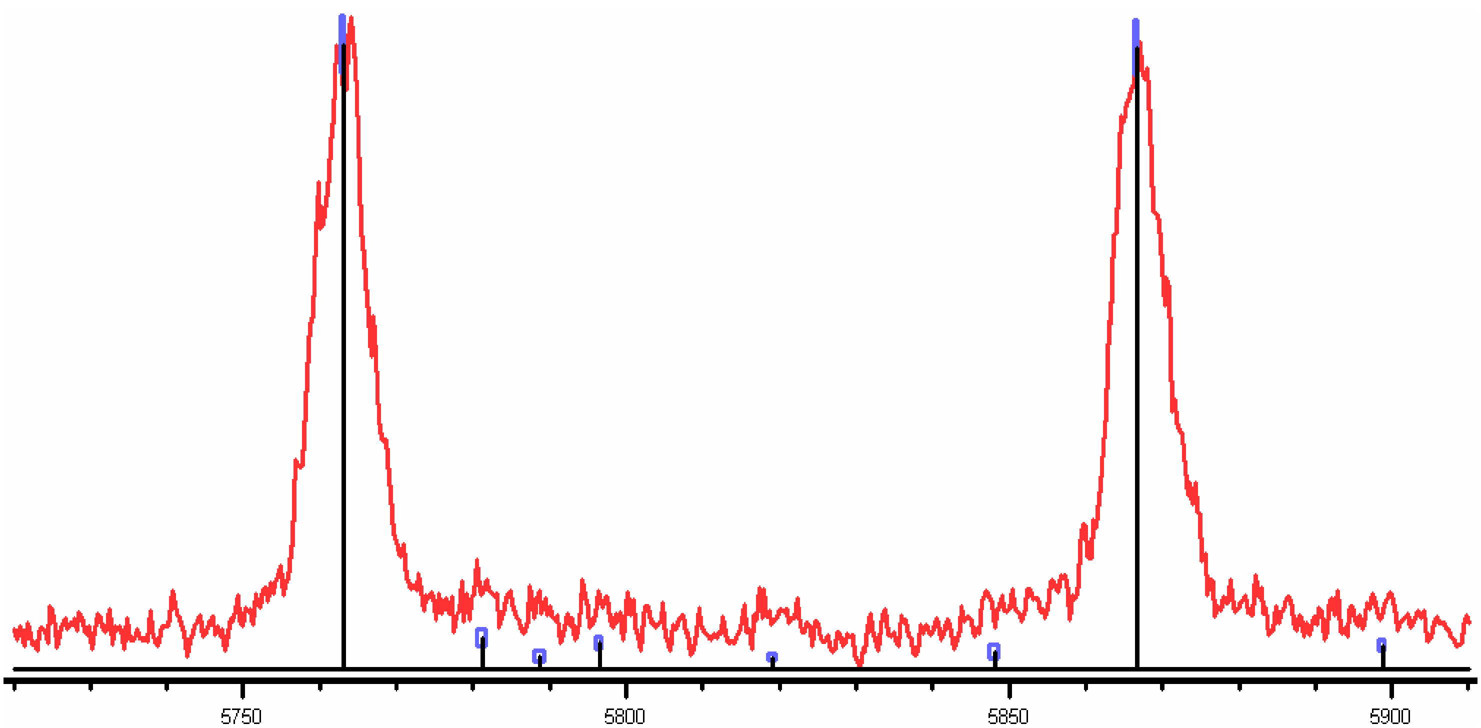

Figure 5. Evidence in the data for the "noise centroids".

Of course, there is no benefit in reconstruction methods if results are not improved. For simple data like the example discussed here, further processing that would make full use of the errors is not necessary. However, many data would require additional processing that would rely on the errors of each peak in the peak table; for example a charge deconvolution. Here, the mass difference between each pair of peaks should be constant and the found variation is a measure of the quality of the two methods. The mass differences for the 29 peaks from m/z 4600-7600 were recorded and their mean and standard deviations calculated. The figures are shown in the table below.

Comparison of Quantified Results

	Mean difference (Da)	Standard deviation (Da)
Data Explorer	104.23	1.46
Reconstruction	104.25	0.48

The standard deviation using the fast reconstruction methodology is only 1/3rd that using the traditional method.

Conclusions

This example demonstrates that:

1. The novel fast reconstruction centroiding program produces a smooth intensity profile much closer to the expectation than traditional centroiding.
2. Mass accuracy, in this case, is improved by a factor of 3.
3. Accounting for the variable noise level produces much cleaner results.
4. Traditional thresholding is redundant.