Published in final edited form as:
Anal Chem. 2013 April 16; 85(8): 3903–3911. doi:10.1021/ac303289c.

Autophaser: An Algorithm for Automated Generation of Absorption Mode Spectra for FT-ICR MS

David P.A. Kilgour[1,*], **Rebecca Wills**[1], **Yulin Qi**[1], and **Peter B. O'Connor**[1]
[1]Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK

Abstract

Phase correction of Fourier Transform – Ion Cyclotron Resonance (FT-ICR) mass spectrometry data allows the spectra to be presented in absorption mode. Absorption mode spectra offer superior mass resolving power (up to a factor of 2), mass accuracy, and sensitivity over the conventional magnitude mode. Hitherto, the use of absorption mode in FT-ICR mass spectrometry has required either specially adapted instrumentation or a manually intensive process of phase correction or has ignored the potentially significant effects of image charge and the associated frequency shifts. Here we present an algorithm that allows spectra recorded on un-adapted FT-ICR mass spectrometers to be phase corrected, their baseline deviations removed, and then an absorption mode plot presented in an automated manner which requires little user interaction.

INTRODUCTION

Mass resolving power is one of the key performance metrics of mass spectrometers. Increased mass resolving power, and associated benefits such as improved mass accuracy, is particularly beneficial for the study of both complex mixtures (for example natural organic matter or crude oil) and very large molecules (for example bio-macromolecules such as proteins or synthetic polymers). Currently, Fourier Transform - Ion Cyclotron Resonance Mass Spectrometers (FTICR MS)[1, 2] offer the best mass resolving powers and mass accuracies of any type of mass spectrometer but there is still scope for improvement which will allow new problems to be investigated.

Mass accuracy and mass resolving power of FT-ICR MS instruments can be increased by increasing the strength of the magnetic field,[3] but this approach, whilst successful, is highly capital intensive owing to the cost of the superconducting magnets required. Fortunately, other approaches are available which can also bring significant additional improvements in mass resolving power and mass accuracy; for example, recent improvements in FT-ICR cell designs which have shown dramatic performance enhancements over the original cell designs.[4–9]

Perhaps the least well known potential improvement to FT-ICR MS performance is the use of absorption mode spectra in place of the more conventional magnitude mode spectra. Since the development of FT-ICR MS, in the 1970s,[1] it has been recognised that absorption mode spectra would offer up to a two-fold improvement in mass resolution and mass accuracy over magnitude mode spectra.[10–12] Unfortunately, the difficulties involved with producing absorption mode mass spectra have hindered their widespread application in FT-ICR MS.

*Corresponding author. d.p.a.kilgour@warwick.ac.uk.

FT-ICR MS, Absorption mode and the Phasing Problem

The motion of ions in an FT-ICR MS instrument is recorded as a time domain transient. This transient is then Fourier transformed to a frequency spectrum in order to detect the component ion oscillation frequencies. The recorded oscillation frequency of a given ion is a function of its mass-to-charge-ratio (*m/z*), and so the frequency spectrum can be calibrated into a mass spectrum.

Fourier transformation of a time domain ICR signal produces a complex output which can be expressed in either polar (Magnitude and Phase) or Cartesian (Real and Imaginary) terms. Standard trigonometric relationships can be used to convert between these two coordinate system options. The magnitude mode spectrum that is produced by all commercial FT-ICR MS instruments plots the frequency against the magnitude component of the Fourier transformation of the time domain ICR signal. The absorption mode spectrum would plot the frequency against the real part of the Fourier transformation of the time domain ICR signal, if there was no phase shift in the signals of ions of different *m/z*.

Unfortunately, the typical mode of operation of FT-ICR MS instruments imparts a large and varying phase shift across the spectrum which prevents the absorption mode spectrum from being calculated as easily as is described above. This phase shift comes from two sources. Firstly, the ions within the FT-ICR cell are excited, prior to detection, by means of an excitation waveform which usually takes the form of a linear frequency sweep; an ion will undergo excitation when its reduced cyclotron frequency matches that of the excitation waveform. Therefore, ions of different *m/z* are sequentially excited at different times during this frequency sweep. Then, secondly, once the excitation waveform has completed, there is a time delay (typically a few milliseconds), before the detection circuitry begins to record, during which all ions will continue to precess around the magnetic field axis. The total phase shift exhibited by an individual ion, ignoring the effect of image charge and other field imperfections, which will be discussed later, can be calculated[13]:

$$\varphi_i = \varphi_0 + \omega_i t_{\mathrm{Total}} \quad (1)$$

where φ_i is the phase shift experienced by an ion of a specific *m/z*, φ_0 is a constant phase shift (corresponding to the angular displacement between the excitation and detection electrodes in the FT-ICR cell), ω_i is the angular frequency of the ion and t_{Total} is the sum of the time from excitation of that ion to the end of the frequency sweep (t_{Excite}) and the delay time between the end of the frequency sweep and the beginning of the transient recording (t_{Delay}), as shown in Eqn (2).

$$t_{\mathrm{Total}} = t_{\mathrm{Excite}} + t_{\mathrm{Delay}} \quad (2)$$

t_{Excite} can be expressed in terms of the parameters of the frequency sweep:

$$t_{\mathrm{Excite}} = \frac{\omega_{\mathrm{Final}} - \omega_i}{R} \quad (3)$$

where ω_{Final} is the final frequency in the frequency sweep and R is the sweep rate. So, combining (2) and (3) into (1):

$$\varphi_i = \varphi_0 + \omega_i \left(\frac{\omega_{\mathrm{Final}} - \omega_i}{R} + t_{\mathrm{Delay}} \right) \quad (4)$$

Or,

$$\varphi_i = \frac{-\omega_i^2}{R} + \omega_i \left(\frac{\omega_{\text{Final}}}{R} + t_{\text{Delay}} \right) + \varphi_0 \quad (5)$$

Eqn (5), the *ab-initio* phase correction function, demonstrates that the phase shift, and hence the phase correction function, should be expected to vary in a quadratic relationship with ICR frequency. This function can be used to calculate the phase shift for all frequencies and hence allow the absorption mode spectrum to be calculated as:

$$\text{Real}_i = \text{Magnitude}_i \times \cos(\theta_i - \varphi_i) \quad (6)$$

As an added complication, the reported phase angle, as returned by the Fourier transformation (θ), will be restricted to values between $-\pi$ and π (and beyond this returns to $-\pi$ again – a feature known as phase wrapping). It is therefore necessary to add an additional 2π onto the reported phase angle for every extra rotation of the ions. As t_{Total} is relatively long compared to $1/\omega_i$, the number of phase wraps across an FT-ICR MS spectrum can be very large – many thousands or tens of thousands of radians. This makes it very difficult to identify the phase correction function and this is why virtually all FT-ICR MS instruments to date have relied on magnitude mode spectra, which do not require the phase shift to be calculated.

Methods for solving the phase shifting and wrapping problems in order to allow the use of absorption mode mass spectra, or small portions of spectra, for FT-ICR MS have been presented previously. Algorithms for producing absorption mode spectra of single peaks;[14] across a narrow *m/z* range;[15] or for a complete spectrum, providing the peaks in the spectrum are sufficiently widely spaced,[16] have been available for more than a decade.

More recently, general methods for producing absorption mode spectra have also been developed.[17–19] These general methods for producing absorption mode spectra from FT-ICR MS data are successful but suffer from specific drawbacks which prevent them from being widely applied. The method of Beu *et al.*[19] requires a specially adapted instrument which can simultaneously excite and detect ions in the FT-ICR cell. The method of Xian *et al.*[17] requires that the FT-ICR mass spectrometer record accurate data about the parameters with which the spectrum was recorded (excitation scan and scan rate parameters and the delay time between the end of excitation and the beginning of the detection phase) in order to provide the data which allows the *ab-initio* phase correction function (Eqn (5)) to be used to produce the absorption mode spectrum, only requiring iterative adjustment of φ_0, the intercept term. Historically, few commercial FT-ICR MS instruments have recorded this data. For example, this information has only been provided on Bruker FT-ICR MS instruments since 2010.

A more important drawback of the method described by Xian *et al.* is that it does not, in most cases, produce accurately phase corrected absorption mode spectra. This may be because the ion frequencies before and after excitation will differ, as a result of the image charge effect[13, 20] and other field inhomogeneities (including space charge), and the method does not correct for these changes.

The image charge effect will, to a first approximation, cause the frequency of the ion after excitation (ω_i) to shift by a small, but significant, constant frequency ($\Delta\omega_K$), independently of mass (but dependent on charge) as a result of the change in orbital radius in the cell:

$$\omega_i' = \omega_i - \Delta\omega_k \quad (7)$$

Note that this relationship would also be the simplest way of allowing for the frequency shifts which would result from other types of field inhomogeneities, hence this relationship could be generally used to allow for more complex effects than just image charge alone.

The total phase shift exhibited by an individual ion, allowing for frequency shifts associated with image charge, can be calculated as:

$$\varphi_i = \varphi_0 + \omega_i t_{\text{Excite}} + \omega_i' t_{\text{Delay}} \quad (8)$$

So, combining Eqns 3, 7 and 8 gives:

$$\varphi_i = \varphi_0 + \omega_i \frac{\omega_{\text{Final}} - \omega_i}{R} + (\omega_i - \Delta\omega_k) t_{\text{Delay}} \quad (9)$$

Which can be rearranged to:

$$\varphi_i = \varphi_0 + \omega_i \left(\frac{\omega_{\text{Final}} - \omega_i}{R} + (1 - \Delta\omega_K) t_{\text{Delay}} \right) \quad (10)$$

And so,

$$\varphi_i = \varphi_0 + \omega_i \left(\frac{\omega_{\text{Final}}}{R} + (1 - \Delta\omega_K) t_{\text{Delay}} \right) + \frac{\omega_i^2}{R} \quad (11)$$

The exact value of $\Delta\omega_K$ will vary depending on the charge of the ion and the image charge conditions in the cell (which will, in turn depend on the ion population, the cell design and the experimental parameters selected (including the amplitude and duration of the excitation signal which will, in part, determine the change in the ion oscillation radius)).

So, including the effect of image charge, as can be seen in Eqn(11), the phase correction function remains a quadratic function, but the curve may be significantly different from that predicted using Eqn(5). Other electrical field imperfections experienced by the ions in the FT-ICR cell, including those resulting from space charge, could also cause the ICR frequency of the ions to change before or after excitation, but the final phase correction function would remain quadratic in character and we expect that the image charge component may be the most prominent of all the frequency adjusting effects.

As the effect of image charge on the ion frequencies will vary from spectrum to spectrum and is difficult to predict or measure reliably or accurately, the method of Xian *et al.* can only be applicable to specially adapted instruments designed to minimize the effects of image charge, for example by using very large cell designs, such as that used by Xian *et al.*; or experiments where the ion population is kept very low; or experiments where the excitation radius of the ions is minimised.

For example, as can be seen in Figure 1, the phase correction function generated by the method described by Xian *et al.* apparently overestimates the phase shift of ions compared to the phase correction function calculated using the Autophaser algorithm described later in this manuscript. We have ascribed this difference to the fact that the method of Xian *et al.*

does not take into account any potential reduction in the oscillation frequency of the ions, after excitation, due to the image charge effect. This reduction in ion oscillation frequency is significant in the results seen from all FT-ICR mass spectrometers in our study and means that the *ab initio* phase correction function cannot normally be used to phase correct these spectra. An example of the absorption mode spectrum which results from applying the method of Xian *et al.* to an example spectrum is provided in Supplementary Information.

Modelling studies are planned which will investigate whether the image charge effect alone can explain the apparent frequency shifts in the ions or whether other field inhomogeneity or imperfection terms may also be required.

The final previously published method of producing absorption mode spectra, that of Qi *et al.*,[18, 21, 22] does not require that the instrument on which it was recorded be specially adapted (with simultaneous excitation and detection or an extra-large cell) and can easily allow for frequency shifts during excitation, such as those resulting from the image charge effect. The method produces a general quadratic phase correction function of the type:

$$\varphi_i = a\omega_i^2 + b\omega_i + c \quad (12)$$

where a, b and c are coefficients which are solved for as part of the method and which take into account the image charge effect. It is not necessary to extract exact values for the frequency sweep rate (R), upper frequency limit (ω_{final}), time delay (t_{Delay}), the initial phase shift (φ_0) or the image charge effect ($\Delta\omega_K$): once the appropriate values for a, b and c are found, the general phase correction function (Eqn (7)) can be applied directly to produce absorption mode spectra.

There are, however, potential drawbacks to the method of by Qi *et al.* First, this method requires that the spectrum used to calculate a phase correction function has a sufficient peak density across the whole mass range of interest (a petroleum sample, for example, provides such high peak density over a wide mass range). Furthermore, the process presented by Qi *et al.* can be manually intensive (taking of the order of two hours to phase a single mass spectrum), it requires an expert user and is difficult to automate in practice. These drawbacks may prevent its broad application.

The method presented here, which we call the Autophaser algorithm, offers the advantages of the method of Qi *et al.*, in that the frequency sweep rate, upper frequency limit, time delay and initial phase shift need not be known and the method is not affected by variable effects of image charge, but is automated and can therefore be easily applied to large collections of FT-ICR MS data.

METHODS

The Autophaser algorithm has six steps:

1. Peak Detection
2. Initial fitting to a reduced m/z range (Followed by the Iterative Tune procedure)
3. Extension of the initial fit (Followed by the Iterative Tune procedure)
4. Fitting of the complete spectrum (Followed by the Iterative Tune procedure)
5. Peak symmetry adjustment
6. Baseline correction

Peak detection

It is only for values taken from spectral peaks that the required frequency, magnitude and phase data can be obtained. For other points in the spectrum, the phase value contains no information, as it is based only on noise. Therefore, the Autophaser method requires that a peak detection algorithm be run which extracts the frequency, magnitude, and phase of peaks in the frequency spectrum derived by Fourier transformation of the time domain transient. This data is passed to subsequent components to produce the phase correction function.

Initial fitting

In order to provide an initial estimation of the phase correction function, the algorithm undertakes a brute force fitting on a subset of the complete spectrum. In normal practice, this region has commonly been selected to have to a mass range in the region of ± 40–250 *m/z* around the central mass.

As shown in Eqns (5) and (11), the phase correction function will have a quadratic relationship to the ICR frequency (for a spectrum recorded with a linear frequency sweep excitation or stored waveform inverse Fourier transform (SWIFT) excitation with a quadratic phase shift function[23]), and any quadratic function can be minimally defined by three points. Therefore, candidate phase correction functions for this initial region are generated by automatically selecting three peaks across the region: the largest detected peak (in magnitude mode) in the first, middle and last sections of the spectral portion. The first of these peaks is set to a relative phase wrap, n_0, = 0. Values for the other two peaks (n_1 and n_2) are then scanned through available integer values of n between 0 and 2000. This range for n was empirically determined on data recorded on a number of different FT-ICR MS instruments. A wider range may be needed on other instruments. It should be noted that this method uses a least squares fitting approach and so obviously requires that there be at least three peaks in the initial region.

For each automatically generated value of n_i, the corresponding correction values of φ_i are calculated according to the relationship:

$$\varphi_i = n_i \times 2\pi \quad (13)$$

For each set of n_i, least squares fitting is used to solve for a quadratic relationship (Eqn (12)) between frequency and phase shift and values of a,b and c are produced. From these values, the relative phase shift for all peaks in the region can be calculated and this information is used with Eqn (6) to calculate an absorption mode amplitude for each peak. A figure of merit for the quality of this trial phase correction function can be produced using the following method.

With the correct phase correction function (derived from the phase wrap values assigned to the three selected peaks), the intensity of all peaks in the initial test region (discounting harmonic peaks and those from external noise) in the absorption mode spectrum would be maximised. Therefore, in order to understand the quality or success of the phase correction function, a Figure of Merit (*FoM*) is calculated for each peak expressing the accuracy of the phase correction:

$$FoM_{PK} = \cos(\theta_i - \varphi_i) \quad (14)$$

The total FoM (FoM_{Total}) for a given phase correction function, for a spectrum or part of a spectrum containing N peaks is:

$$FoM_{\text{Total}} = \sum_{Pk=0}^{N} FoM_{PK} \quad (15)$$

This FoM_{Total} is recorded (based on all peaks in the initial region) for all tested combinations of n_i for the three test peaks, and the combination which results in the highest FoM_{Total} is used to define the initial phase correction function that is passed onto the next section of the Autophaser algorithm. If no distinct maximum is found then this may indicate that the initial region used was too wide and the relative phase shift across that region exceeded 2000 cycles, or that the region contained a large proportion of low intensity peaks for which spectral noise had degraded the quality of the peak definition (peak maximum frequency and phase). So, the user must either reduce the length of the initial region to assess whether the relative phase shift was too great or adjust the parameters of the peak detection to avoid including poor signal-to-noise peaks.

In the absence of any other information, fitting to the initial region would require four million calculations (n can be any integer value between 0 and 2000 for two peaks, hence the total number of combinations of n_i for the three peaks is 2000^2). However, it is possible to dramatically reduce the numbers of possible valid combinations of n_i for all three peaks because it is known that although the phase correction function is quadratic, the function is actually very close to being a straight line, over a small mass range. Therefore, with n_1 for the first peak held at zero, n_2 for the second peak set at a known value (between 0 and 2000), and the frequency of all three peaks known, linear extrapolation can be used from the first two peaks to define a relatively small potential range of n_3 for the last peak (usually a range of ± 5, around the linear extrapolation of points n_1 and n_2 at the frequency corresponding to n_3). In practice, this method of predicting a small range of values of n_3 for the third peak, for every loop, reduces the total number of calculations done in the initial fitting region by two to three orders of magnitude.

Iterative Tuning Method

The key component in this algorithm is the method by which the correction function is tuned after steps 2, 3 and 4, to improve the quality of the phase correction function.

If one plots peak frequency against FoM_{Pk} for a situation where the phase correction function is not correct, the points will lie on a periodic function of the type:

$$FoM_{Pk_i} = \cos(g\omega_i^2 + h\omega_i + j) \quad (16)$$

Where g, h and j are unknown coefficients, and ω_i is the ICR frequency of peak i. This function has the form of the cosine of a quadratic because the residual error between two quadratics(in this case the true phase correction function and the estimated phase correction function) is, of course, itself a quadratic and the phase wrapping effect adds the cosine term.

If the phase correction function were correct, all points would line along a line at FoM = 1. An example of such a frequency versus FoM_{Pk} plot is presented in Figure 2, showing how the quality of the phase correction functions affects the periodic error function. Points which lie off the line are a result of noise in the original spectrum preventing the true initial peak frequency and/or phase from being accurately recorded.

We use an iterative convergence approach to optimize the phase correction function. In order to improve the quality of an imperfect phase correction function, for example the estimated general phase correction function produced by the initial brute force fitting, the algorithm adjusts the φ_i value of three arbitrary points on the curve (one each at the lowest and highest frequencies to be considered and a third at the middle frequency), by small amounts, and iterates to maximize FoM_{Total}. These three points used for tuning are simply points which lie on the curve and are not actual peak positions. In summary, the workflow for this tuning algorithm is:

- Adjust one of the three points on the initial curve
- Calculate the parameters for the new curve
- Calculate φ_i for all peaks in the region based on this new curve
- Calculate the FoM_{Pk} for each peak (Eqn (14))
- Calculate FoM_{Total} (Eqn (15))
- Iterate to maximize FoM_{Total}

This iteration process does not attempt to search the entire three dimensional data space provided by the variation of the range of potential values of each of the three points. Instead, we have found it sufficient to iterate the three points sequentially, seeking a gross maximum each time, and then to repeat this sequence at increasing levels of accuracy until the results converge upon a consistent maximum value. This greatly reduces the computational requirements of the iteration; from over 1,000,000 calculations to less than 10,000.

Extension to full spectrum coverage

After the brute force derived initial phase correction function has been tuned using the method described above, the improved phase correction function can be tested against peaks detected in a wider mass range of the spectrum (the initial region plus the same spectral width on either side). The tuning method is used again to provide an improved fit over this extended region. This extension method can be undertaken several times until one or other end of the spectrum is reached, with the correction function being tuned after each extension.

At this point, the correction function can be applied to peaks detected across the complete spectrum and tuned a final time using the tuning function. Once this process is complete, the final phase correction function can be applied to the complete spectrum and this can be used to produce an initial absorption mode spectrum.

Peak Symmetry Adjustment

Because the mass accuracy requirements of FT-ICR MS are high, the initial absorption mode spectrum will probably require further adjustment in order to maximize the potential improvements in mass accuracy. In particular, if the initial spectrum had a large proportion of peaks with a low signal-to-noise ratio (leading to unavoidable errors in accurately determining the phase and frequency at the peak apices) or if the spectrum had a uneven distribution of peaks in the frequency domain (which given that there is an inverse relationship between frequency and mass is likely, if there is a wide mass range of ions in the spectrum), then the initial iterative tuning procedure may not provide an estimate of the phase correction function which is sufficiently accurate. In this case, the phase correction function generated by the tuning method described above will be close, but not close enough, to the actual phase correction function and the effect of this will be seen in slight peak asymmetries in the resultant absorption mode spectrum; one side of each peak may be slightly lower than the other, as shown in Figure 3.

The residual phase correction error (the difference between the estimated phase correction function and the true phase correction function) will remain quadratic because the difference between two quadratics is also quadratic.

These peak asymmetries can be removed and the phase correction function corrected using the following workflow:

- Estimate the average peak width in the spectrum
- For each detected peak in the spectrum, extract a portion of both the magnitude and phase corrected phase arrays (the latter being calculated as ($\theta_i - \varphi_i$: from Eqn (6)), twice the average peak width, centered on the peak.
- Find the optimum phase adjustment value (between $-\pi$ and π) for each peak, ψ_i, which produces the highest minimum value in the Absorption spectrum of each of the extracted peak sections. Record these phase adjustment values as a function of the peak frequency.
- By a least squares method, fit a quadratic to the resultant points, as shown in Figure 4.
- Correct the complete phase corrected phase array using the quadratic generated above ($\theta_i - \varphi_i + \psi_i$)
- Use this new phase corrected array to plot the absorption mode spectrum.

Baseline Correction

Absorption mode spectra have been reported show two types of baseline distortion,[24] specifically a low frequency oscillation (known as baseline roll) and increasing deviations near large peaks, and in particular groups of large peaks. In our experiments, we do not normally see baseline roll (this is being further investigated) however, the baseline distortion close to intense peaks is still present. This baseline distortion could make it harder for peak detection algorithms to detect low level peaks in such a spectrum and it also affects the peak intensity of large peaks measured against simple flat thresholds or baselines. Therefore, it may be advantageous to use some method for removing or ameliorating the effects of the baseline distortion.

Xian *et al.* have previously published a method for baseline subtraction, that was developed specifically for this application.[24] However, we continue to use an older, general purpose algorithm, which, amongst other capabilities (including peak and anomaly detection for chromatographic, spectroscopic and mass spectrometric data), generates an estimation of baseline. In order to produce an estimate of the baseline, it is necessary to identify regions of the spectrum which are not peaks. We identify non-peak regions of the spectrum by a method which is a development of the techniques described by Williams et al.[25] and Friedman.[26] The spectrum is decimated into sections of a length selected by the user (usually between 3 and 15 points, depending on the spectrum); the median value of each section is taken. The difference between the median value of each region is compared to that of the next, and the mean (m) and standard deviation (SD_B) of those differences is calculated. Each section is classified as being accepted as being part of a putative baseline if the difference between the median value of that decimated region and the next is less than ($m \times n.SD_B$), where n is an integer value set by the user. The accepted points are passed through a (normally fifth order) Savitsky-Golay filter[27] to produce a smoothed putative baseline for the decimated spectrum. To allow for the various changes in scale for the peaks, noise and baseline deviation in the uncorrected absorption mode spectrum, three different values of n are used, and the putative baselines of all three values of n are taken onto the next stage: this step is analogous to the variable span concept of Friedman's

supersmoother.[26] Linear interpolation of these decimated arrays is used to generate three arrays (*B1, B2, and B3*), which provide a putative baseline value for every point in the original spectrum. An example portion of a spectrum, showing three different putative baselines, is shown in Figure 5a.

The spectrum is statistically analysed in order to determine the standard deviation (SD_s) of the intensity difference between any point and the next point. Points in the original spectrum (S_i) are classed as being part of the baseline (i.e. not peaks) if they are less than the minimum of the three putative baseline predictions for that point, plus SD_s (Eqn(12)):

$$S_i \geq (\min[B1_i, B2_i, B3_i] + SD_s) : \text{True=point in baseline} \quad (17)$$

Ignoring points identified by Eqn(12) as being in peaks, the spectrum is re-decimated (again, by a user controlled parameter – usually between 3 and 15), as are the putative baseline arrays *B1, B2* and *B3*. A hybrid baseline array is generated comprising, for each spectral section, the portion of putative baseline arrays *B1, B2* and *B3* which has the lowest error relative to the equivalent peak-free spectral portion. This hybrid baseline array is passed through a second Savitsky-Golay filter (usually third order) in order to produce the final predicted baseline as shown in Figure 5b. The final predicted baseline can then be subtracted from the uncorrected absorption mode spectrum in order to produce the baseline corrected absorption mode spectrum as shown in Figure 5c.

RESULTS

The Autophaser Algorithm produces absorption mode spectra which are comparable to those produced by the manual method of Qi *et al*.[18] and which show the expected improvements in resolving power (√3 – 2 fold), signal to noise ratio (√2 fold), and mass accuracy, over the same spectra when processed in magnitude mode.[17, 21] An example spectrum phased by both methods is provided in the Supplementary Information.

We have applied the Autophaser algorithm to mass spectra of many different sample types with positive results. For example, a clear benefit can be seen comparing the mass accuracy of assigned peaks from the analysis of Talisker single malt whisky in magnitude mode and absorption mode, generated using the Autophaser algorithm. The whisky was diluted 1:100 in 50% methanol/water and analysed on a 12T Bruker Daltonik Solarix FT-ICR MS. The most intense 2000 peaks were assigned in the magnitude mode spectrum and in the absorption mode, both with and without background subtraction, using the KMD algorithm.[28, 29] In the magnitude mode spectrum, the average mass error was 0.080ppm ± 0.104 (standard deviation), in the absorption mode spectrum without baseline correction the mass error was 0.065ppm ± 0.081 and in the absorption mode with baseline correction, the mass error was 0.066ppm ± 0.084. The slight degradation in mass accuracy after baseline subtraction was expected and supports the conclusion that in future, it may be better not to undertake baseline subtraction in order to allow simple peak detection algorithms to work. Instead, it may be preferable to use improved peak detection algorithms which are capable of detecting peaks and identifying the true peak amplitude against a moving background rather than relying on a simple linear threshold method of peak detection. The baseline detection algorithm described in this manuscript can also be used to detect peaks simply by using Eqn(17) and selecting and fitting as peak regions those portions of the spectrum which return a false value.

Regarding the potential of absorption mode spectra to improve the assignment of peaks with a low signal to noise ratio, we are using the Autophaser algorithm to help localize the site of binding between a ruthenium compound and bovine insulin in an ongoing study. Figure 6

shows the major steps in the progress of the phase correction of a ruthenium bound insulin spectrum. Experimental parameters for this spectrum are provided in the Supplementary Information.

As has been noted in previous studies, the absorption mode spectrum allows more peaks to be both detected and assigned, increasing the amount of information that can be taken from a mass spectrum. For example, in Figure 7, the isotope distribution patterns of two sets of peaks in the collisionally induced dissociation mass spectra of insulin, after reaction with a ruthenium compound, are shown. In Figure 7a, the peaks in the unphased spectrum have a barely discernible isotope spacing pattern, making it difficult to confidently calculate the charge state and therefore accurately measure the mass of the peak. In Figure 7b, the absorption mode spectrum has been phased and the isotope spacings are now clearly visible and can be confidently measured. So, only in the absorption mode spectra can the isotopic envelope of ruthenium be correctly identified for these peak clusters, and the charge state assigned.

Both peak clusters, at *m/z* 1440.854 and 1444.859, were calculated to have a charge state of +4, therefore allowing an accurate mass to be determined and the peaks to be assigned. The peak at *m/z* 1444.859 was identified (with a mass accuracy of −0.37 ppm) as insulin plus the ruthenium compound, but minus two amino acids: leucine and valine (indicated by asterisks in the figure). From the unphased data, it had been proposed that the ruthenium metal was bound to the histidine and glutamic acid residues of the insulin B-chain (His10 and Glu13), but there was no amino acid loss observed to support this conclusion. The assignment of the newly identified peak shows loss of both leucine and valine, which are the two connecting amino acids between His10 and Glu13 at positions 11 and 12 of the insulin B-chain respectively. Loss of these two amino acids suggests the ruthenium is in fact bound at histidine and glutamic acid, as this enables the chain to still be connected, whilst the leucine and valine are cleaved and lost in CAD. The peak at *m/z* 1440.854 corresponds to a loss of NH_2 from the previous peak, the loss of which could have occurred from a side chain or the N-terminus of either insulin chain. Therefore, the use of absorption mode spectra can be shown to allow new and valuable information to be discerned from mass spectra which had previously been analysed only in magnitude mode. The results of the complete study into the binding of the ruthenium complex and insulin will be published elsewhere.

Processing times for producing absorption mode spectra will depend on the computer used to undertake the calculations. On a Dell Optiplex 390 desktop, with a Core i5 processor, 8 GB Ram, 7200rpm hard drive, running Windows 7, it took 30(±~3) seconds to phase an 8M point spectrum using the Autophaser algorithm programmed in National Instruments LabVIEW 2012; this includes time to load the file and display the result. This time is based on 5 sequential runs each, phasing both the ruthenium bound insulin spectrum described above and a crude oil spectrum described previously.[18]

CONCLUSIONS

None of the results we have recorded can confirm that it is the image charge effect alone which is responsible for the apparent shift in ion frequency, hence the necessity of the modeling studies which were discussed earlier. However all spectra analysed to date have been seen to require a quadratic form of phase correction function and where the ion oscillation frequency is apparently reduced after excitation. Therefore, we have seen no evidence to date which contradicts the expected image charge hypothesis.

Regardless of the cause of the apparent shift in ion oscillation frequency, the method we have described in this manuscript provides a robust and rapid method of producing

absorption mode mass spectra from FT-ICR MS data with minimal user interaction. We have applied this technique to aid in the analysis of different sample types, where the improved mass resolving power, mass accuracy and improved signal-to-noise ratio has proven beneficial.

A version of this algorithm, coded in NI LabVIEW 2012 software, is available on request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge Matthias Witt (Bruker Daltonik GmbH, Bremen, Germany) and Pat Langridge-Smith (University of Edinburgh) for the use of the whisky data. The authors thank the EPSRC: Warwick Centre for Analytical Sciences (EPSRC: P/F034210/1), NIH (Defining the Isoaspartome), and the Department of Chemistry, University of Warwick for funding.

REFERENCES

1. Comisarow MB, Marshall AG. Chem. Phys. Lett. 1974; 25:282–283.
2. Amster IJ. J. Mass Spectrom. 1996; 31:1325–1337.
3. Marshall AG, Hendrickson CL, Jackson GS. Mass Spectrom. Rev. 1998; 17:1–35. [PubMed: 9768511]
4. Nikolaev E, Boldin I, Jertz R, Baykut G. J. Am. Soc. Mass. Spectrom. 2011; 22:1125–1133. [PubMed: 21953094]
5. Tolmachev AV, Robinson EW, Wu S, Kang H, Lourette NM, Paša-Tolić L, Smith RD. J. Am. Soc. Mass. Spectrom. 2008; 19:586–597. [PubMed: 18296061]
6. Tolmachev A, Robinson E, Wu S, Smith R, Pašs-Toli L. J. Am. Soc. Mass. Spectrom. 2011; 22:1334–1342. [PubMed: 21953187]
7. Tolmachev AV, Robinson EW, Smith RD, Leach Iii FE, Futrell JH, Pašsa-Tolić L. Int. J. Mass spectrom. 2012; 325–327:45–50.
8. Caravatti P, Allemann M. Org. Mass Spectrom. 1991; 26:514–518.
9. Brustkern AM, Rempel DL, Gross ML. J. Am. Soc. Mass. Spectrom. 2008; 19:1281–1285. [PubMed: 18599306]
10. Comisarow MB. J. Chem. Phys. 1971; 55:205.
11. Marshall AG. J. Chem. Phys. 1971; 55:1343.
12. Marshall, AG.; Verdun, FR. Fourier Transforms in Nmr, Optical, and Mass Spectrometry: A User's Handbook. Elsevier Science and Technology. Oxford, UK: Elsevier Science Ltd; 1990.
13. Xiang X, Grosshans PB, Marshall AG. Int. J. Mass Spectrom. Ion Processes. 1993; 125:33–43.
14. Comisarow MB, Marshall AG. Chem. Phys. Lett. 1974; 25:282–283.
15. Vining BA, Bossio RE, Marshall AG. Anal. Chem. 1999; 71:460–467. [PubMed: 9949733]
16. Craig EC, Marshall AG. J. Magn. Reson. 1988; 76:458–475.
17. Xian F, Hendrickson CL, Blakney GT, Beu SC, Marshall AG. Anal. Chem. 2010; 82:8807–8812.
18. Qi Y, Thompson C, Van Orden S, O'Connor P. J. Am. Soc. Mass. Spectrom. 2011; 22:138–147. [PubMed: 21472552]
19. Beu SC, Blakney GT, Quinn JP, Hendrickson CL, Marshall AG. Anal. Chem. 2004; 76:5756–5761. [PubMed: 15456295]
20. Leach F, Kharchenko A, Heeren R, Nikolaev E, Amster I. J. Am. Soc. Mass. Spectrom. 2010; 21:203–208. [PubMed: 19896390]
21. Qi Y, Barrow MP, Li H, Meier JE, Van Orden SL, Thompson CJ, O'Connor PB. Anal. Chem. 2012; 84:2923–2929. [PubMed: 22339804]

22. Qi Y, Barrow MP, Van Orden SL, Thompson CJ, Li H, Perez-Hurtado P, O'Connor PB. Anal. Chem. 2011; 83:8477–8483. [PubMed: 21975143]

23. Guan S, McIver RT. J. Chem. Phys. 1990; 92:5841–5846.

24. Xian F, Corilo YE, Hendrickson CL, Marshall AG. Int. J. Mass spectrom. 2012; 325–327:67–72.

25. Williams, B.; Cornett, S.; Dawant, B.; Crecelius, A.; Bodenheimer, B.; Caprioli, R. Proceedings of the 43rd annual Southeast regional conference - Volume 1. Kennesaw, Georgia: ACM; 2005. p. 137-142.

26. Friedman, J. A Variable Span Smoother. Stanford: Laboratory for Computational Statistics, Dept. of Statistics Stanford University; 1984. Technical Report 5

27. Savitzky A, Golay MJE. Anal. Chem. 1964; 36:1627–1639.

28. Kilgour DPA, Mackay CL, Langridge-Smith PRR, O'Connor PB. Anal. Chem. 2012; 84:7431–7435. [PubMed: 22880549]

29. Kilgour DPA, Mackay CL, Langridge-Smith PRR, O'Connor PB. Anal. Chem. 2012; 84:7436–7439. [PubMed: 22881189]

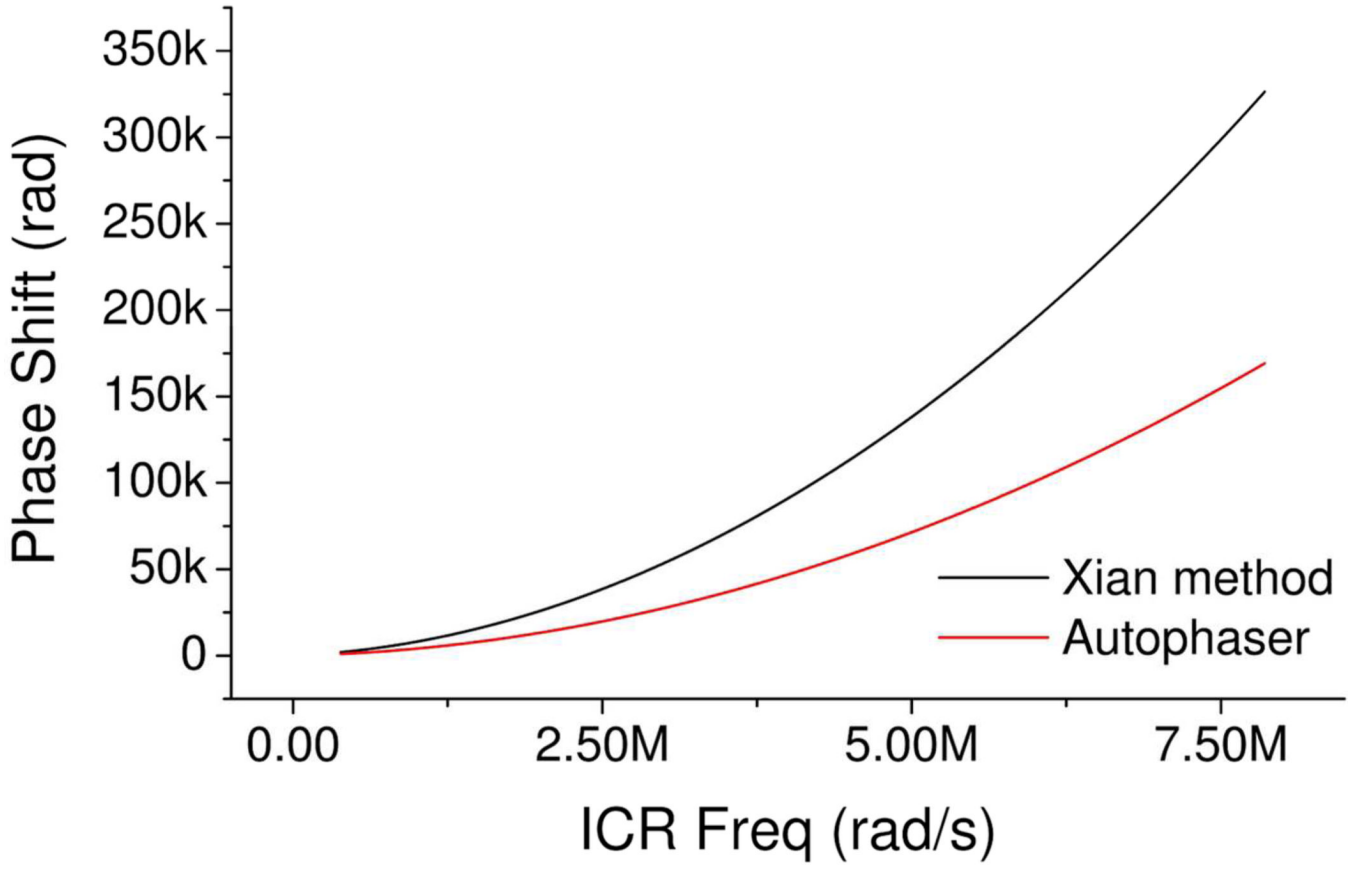

Figure 1.
Comparison of the phase correction functions calculated using the methods of Xian *et al.*[17] and the Autophaser algorithm. This data was generated from the spectrum of a Ruthenium bound insulin complex; experimental details are provided in the Supplementary Information.

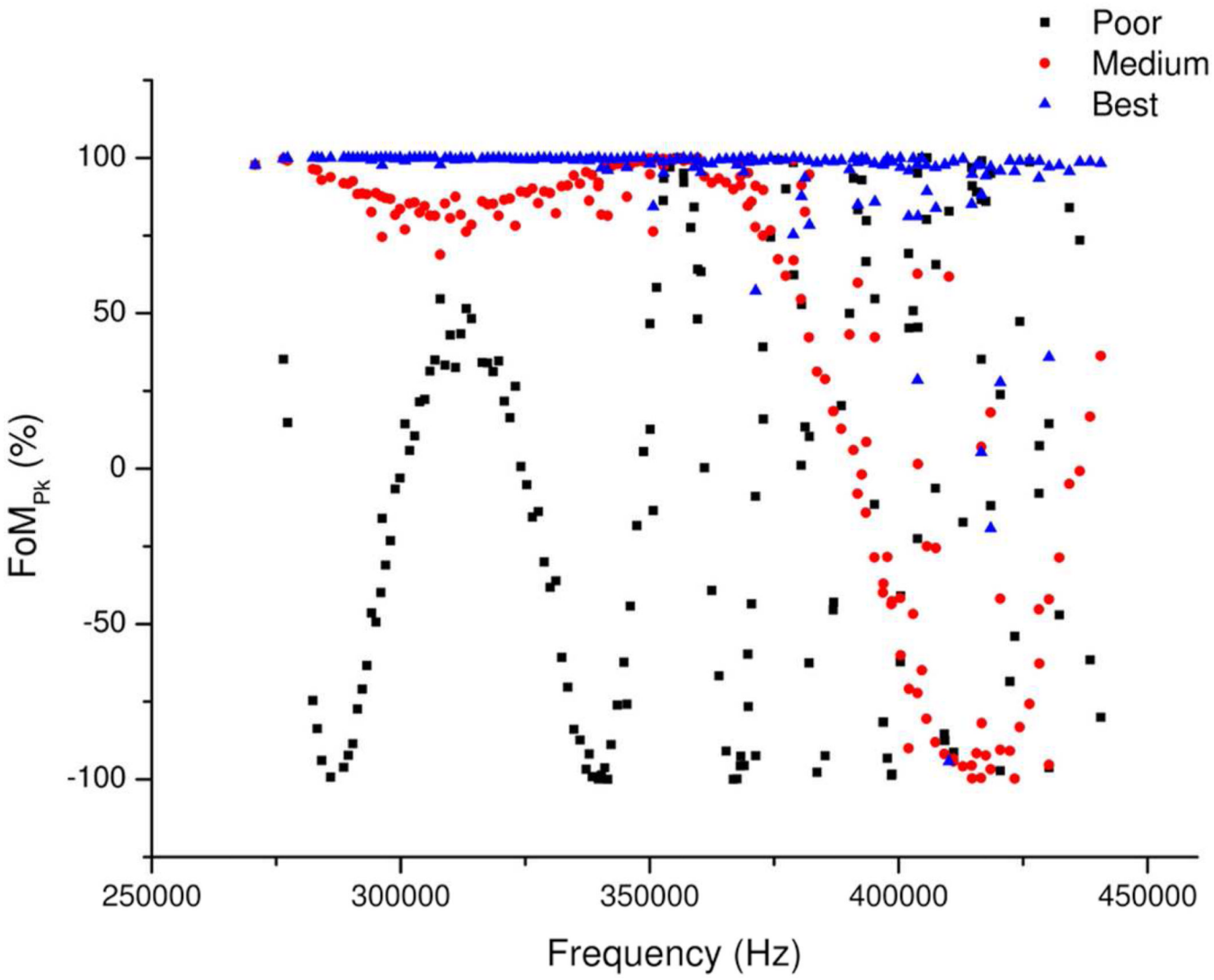

Figure 2.
Showing the improvement in the frequency versus FoM function as the quality of the phase calibration function is improved.

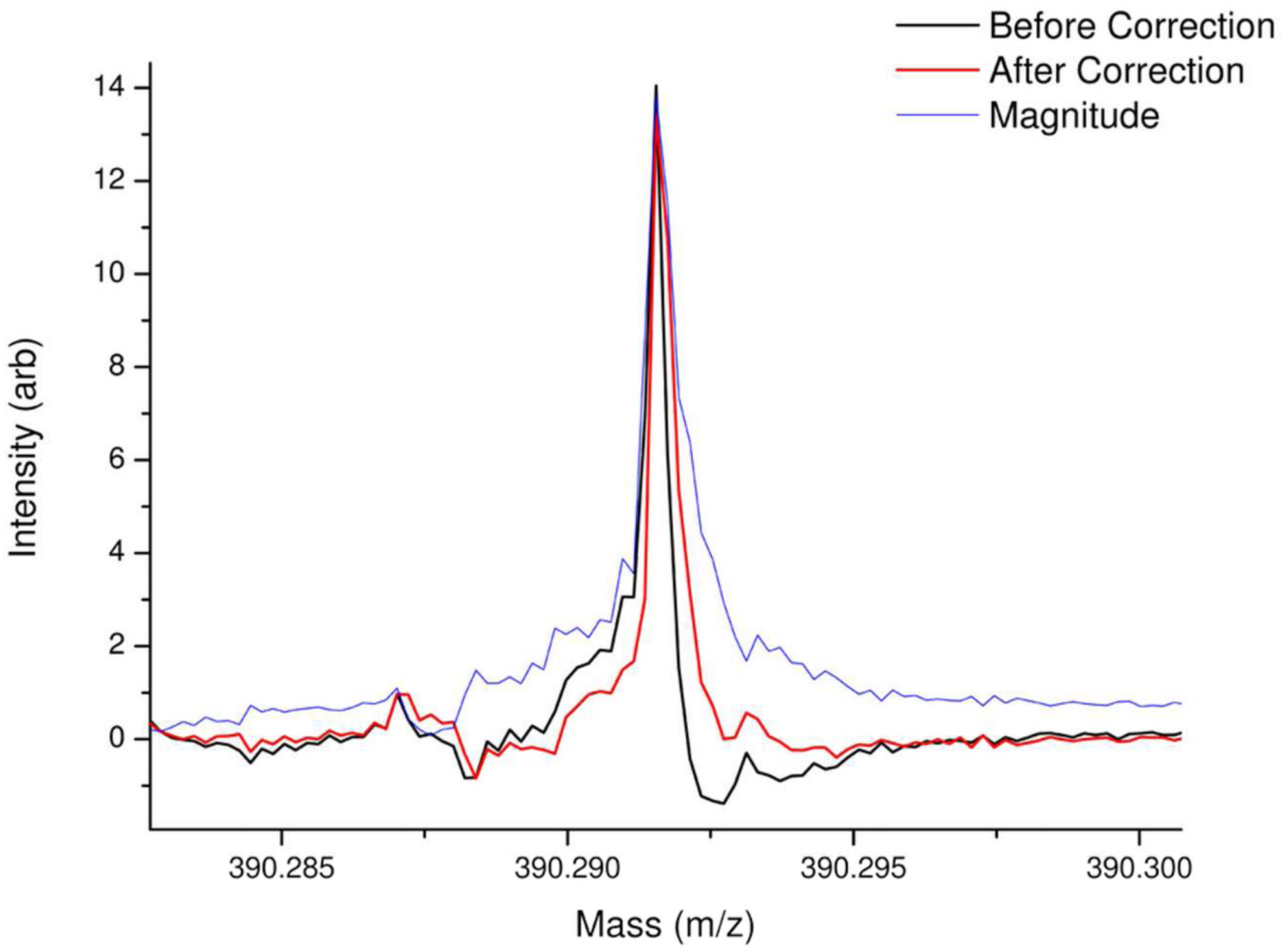

Figure 3.
Mass spectral portion showing the difference between the peak shape for a single example peak (extracted from a mass spectrum generated by atmospheric pressure photoionization of a crude oil) both before and after peak symmetry correction. The magnitude mode of the same peak is also shown for comparison.

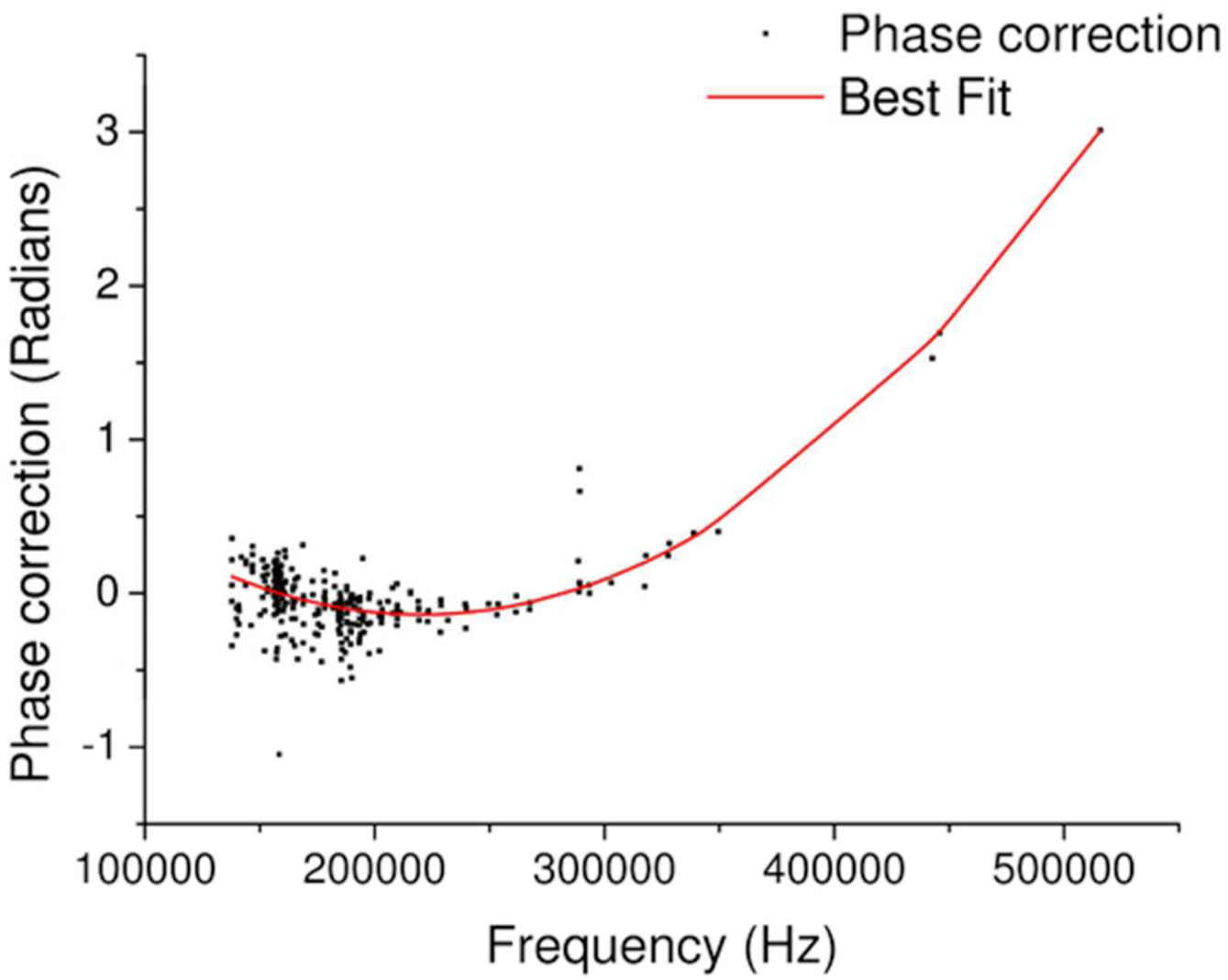

Figure 4.
Showing the quadratic least squares fit to the peak symmetry adjustment values. Points which lie off the line are due to error in correctly registering the phase and or frequency of the peak apices; this is often found for peaks with low signal-to-noise ratios.

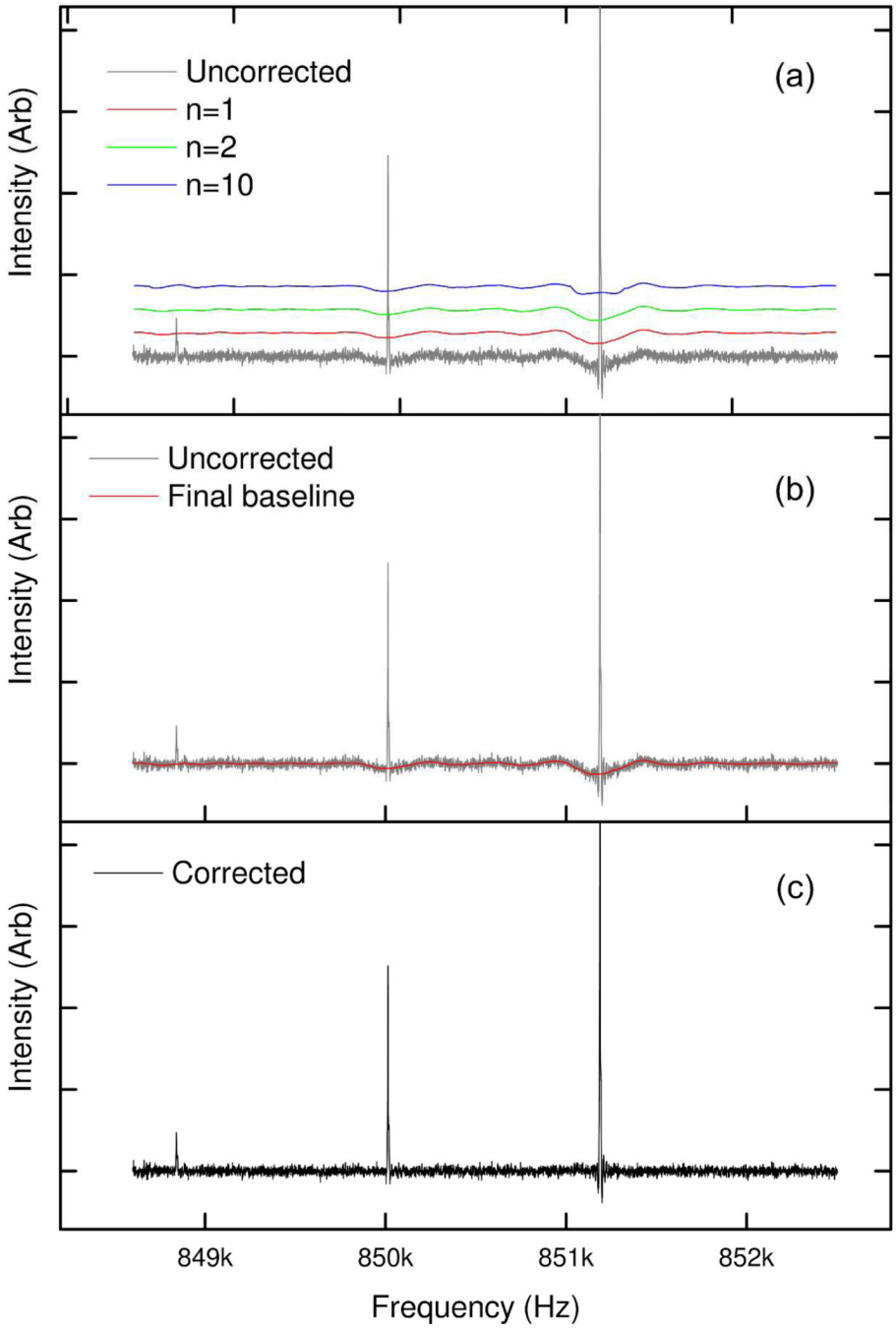

Figure 5.

An example portion of an uncorrected absorption mode spectrum of insulin, showing the effect of baseline correction. (a) Showing the original uncorrected spectrum with three different putative baselines, for three different values of n (at n = 1, 2 and 10 times the standard deviation respectively). The putative baselines are shown with an offset to improve readability. (b) The uncorrected spectrum with the final baseline. (c) The baseline corrected spectrum.

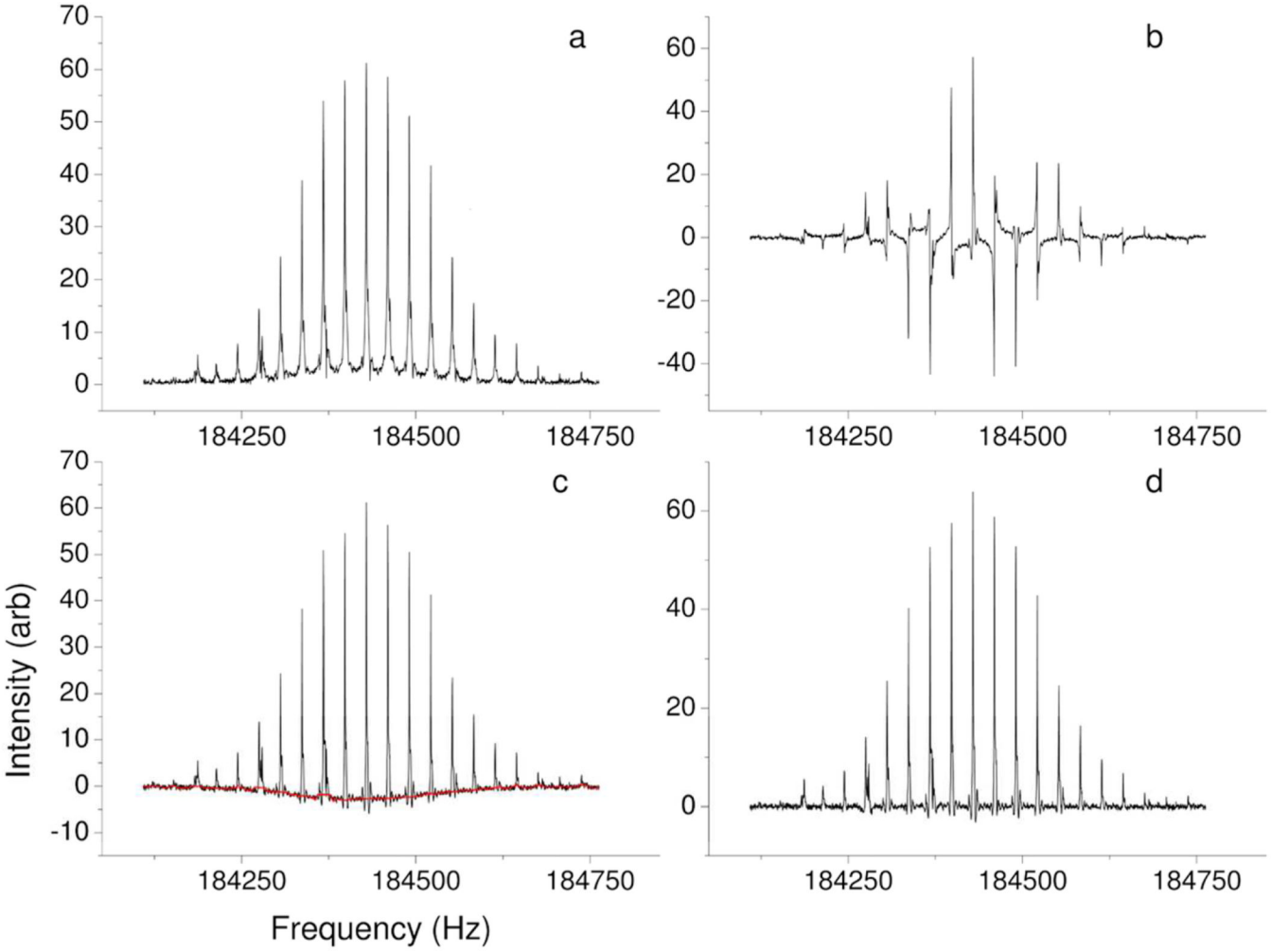

Figure 6.

Zoomed portion of a frequency spectrum showing the different stages in the Autophaser process for a Ruthenium-bound insulin sample analysed by electrospray ionization FT-ICR MS. a) Magnitude mode. b) Absorption mode before phase correction. c) Absorption mode after phase correction and peak symmetry adjustment, showing final detected baseline. d) Absorption mode after baseline correction.

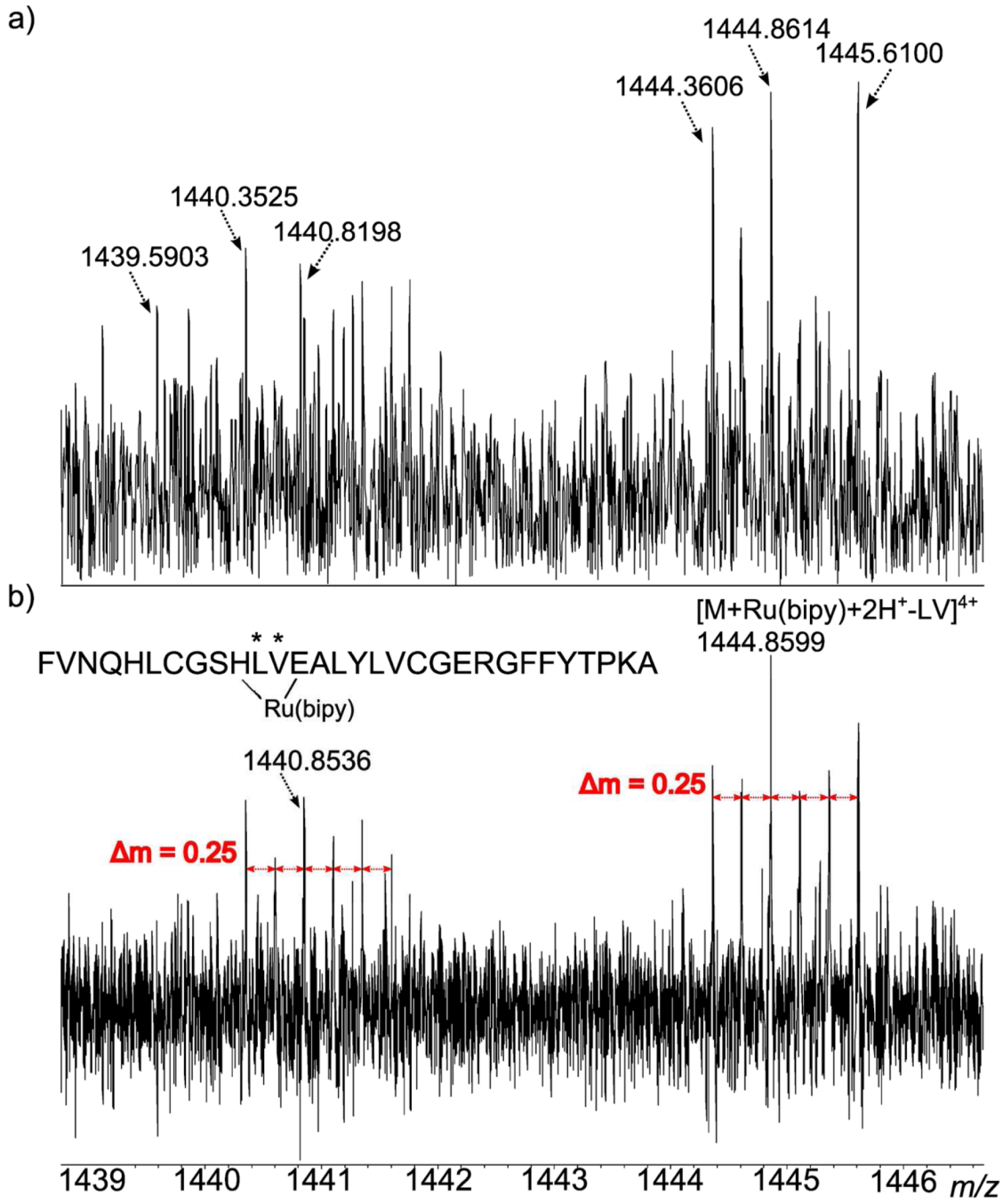

Figure 7.
Magnitude (a) and absorption (b) mode spectra of the same region in the mass spectrum of a ruthenium compound, bound to insulin (after collisionally induced dissociation).