Automated Charge State Determination of Complex Isotope-Resolved Mass Spectra by Peak-Target Fourier Transform

Li Chen and Yee Leng Yap

Bioinformatics Institute, 30 Biopolis Street, Singapore
Davos Life Science Pte Ltd., 11 Biopolis Way, Singapore

This study describes a new algorithm for charge state determination of complex isotope-resolved mass spectra. This algorithm is based on peak-target Fourier transform (PTFT) of isotope packets. It is modified from the widely used Fourier transform method because Fourier transform may give ambiguous charge state assignment for low signal-to-noise ratio (S/N) or overlapping isotopic clusters. The PTFT algorithm applies a novel “folding” strategy to enhance peaks that are symmetrically spaced about the targeted peak before applying the FT. The “folding” strategy multiplies each point to the high-
$m/z$
side of the targeted peak by its counterpart on the low-
$m/z$
side. A Fourier transform of this “folded” spectrum is thus simplified, emphasizing the charge state of the “chosen” ion, whereas ions of other charge states contribute less to the transformed data. An intensity-dependent technique is also proposed for charge state determination from frequency signals. The performance of PTFT is demonstrated using experimental electrospray ionization Fourier transform ion cyclotron resonance mass spectra. The results show that PTFT is robust for charge state determination of low S/N and overlapping isotopic clusters, and also useful for manual verification of potential hidden isotopic clusters that may be missed by the current analysis algorithms, i.e., AID-MS or THRASH. (J Am Soc Mass Spectrom 2008, 19, 46–54) © 2008 American Society for Mass Spectrometry

Electrospray ionization mass spectrometry (ESI-MS) [1–4] greatly extends the capability for measuring the masses of large biomolecules by generating multiply charged ions. Multiple charging shifts the signal of molecules as large as 200 kDa down to the mass-to-charge ratio (m/z) for facilitated detection by most mass analyzers [5]. However, it also complicates mass spectral interpretation because the charge state of each ion must be assigned to enable determination of ion mass, and ions of a given mass typically exhibit several charge states.

Charge state determination of a fully resolved isotopic cluster is relatively straightforward because the spacing between adjacent isotopic peaks is simply the reciprocal of the charge state [6]. However, mass spectra of large biomolecules are usually very complicated because thousands of isotopic peaks from a mixture of ions are packed in an m/z range from 500 to 18,000 Da [7]. Consequently, this produces a high probability that two or more isotopic clusters may overlap together. The determination of charge states of overlapping isotopic clusters is one of the greatest challenges for speeding up the interpretation of complex high-resolution mass spectra.

Some efforts have been made to infer charge state determination from high-resolution mass spectra. Senko and colleagues [5, 8] described a method that combines a Patterson pattern-recognition algorithm with a Fourier transform for charge determination of fragment ions from isotopically resolved collision-induced dissociation tandem mass spectra of intact proteins in a Fourier transform ion cyclotron resonance (FTICR) mass spectrometer [9, 10]. Their results suggest that the Patterson algorithm performs best on low charge (Z < 5) with low signal-to-noise ratio (S/N) and excessive resolving power. In contrast, the Fourier algorithm becomes superior on higher charge states (Z > 5) with low resolving power. They also demonstrated a successful example for charge determination of two overlapping isotopic clusters with significant abundance differences using Fourier transform, although the Patterson routine is not recommended because the broad peaks it produces will blend together and potentially produce inaccurate results. Zhang and Marshall [11] described a Z-score algorithm for routine charge state determination and spectral deconvolution, for either high- or low-resolution mass-to-charge ratio spectra, to yield unambiguous zero-charge mass spectra. Z-score assigns a score for each possible charge state according to an appropriate charge-scoring scheme, and then determines the charge state with the highest score as the correct one. Tabb and colleagues [6] investigated the Fourier transform approach for determining charge states from either mass spectra of peptides from a linear quadrupole ion trap mass spectrometer or mass spectra of intact proteins from an FTICR mass spectrometer. To
our dismay, study on the overlapping isotopic clusters was not performed. Horn and co-workers [12] developed a THRASH (thorough high-resolution analysis of spectra by Horn) algorithm for automated reduction and interpretation of high-resolution electrospray mass spectra of large molecules. This method uses the Fourier transform/Patterson method for primary charge determination in each moving window. However, if the ambiguous charge states are obtained by Fourier transform/Patterson, it is required to search each possible charge state in a user-defined charge range (i.e., 1–30). The charge state with the highest reliability value is assigned as the correct charge state. In our previous research, an AID-MS (automated intensity descent algorithm for analysis of mass spectrometry) algorithm [7, 14] for speeding up the interpretation of complex high-resolution mass spectra was developed. The results showed that AID-MS is fast in computational speed, robust in identification of overlapping isotopic clusters, and efficient in minimization of false positives. A peak selection method is proposed to identify isotopic peaks, and then the exact value or a searching range of charge state is determined by the reciprocal of the interpeak spacing of isotopic peaks. If a searching range is obtained, the charge state is determined by the optimal matching between average peaks with isotopic peaks.

Here, we present a simple and fast algorithm based on peak-target Fourier transform (PTFT) for automated charge state determination of isotope-resolved mass spectra. The PTFT algorithm selects a targeted peak in a studied m/z window, and then applies a “folding” strategy on the low-m/z side and high-m/z side of the targeted peak to generate a “folded” spectrum. The charge state is determined from the Fourier transform of this “folded” spectrum. The performance of the PTFT algorithm is demonstrated using complex isotopically resolved mass spectra. The results show that PTFT can identify the charge states with high accuracy, especially for low S/N and overlapping isotopic clusters.

**Algorithm**

For a theoretical isotopic cluster, the spacing between adjacent isotopic peaks is equidistant, and the charge state can be simply determined by the reciprocal of the spacing. However, due to the interference of noise signals and overlapping complexities, experimental isotopic peaks may not exhibit exactly equal spacing. Normally, a tolerance of peak shifting is required when selecting isotopic peaks from spectra; thus, the interpeak spacing between different adjacent isotopic peaks may not be same. When using the reciprocal of the interpeak spacing to estimate the charge state, the errors may be produced. When charge state (Z) is high, the trivial errors of spacing may cause a huge difference in charge states because 1/Z and 1/(Z + 1) differ only slightly. Therefore, higher charge states may give rise to higher error rates. However, it is worth pointing out that the reciprocal of the interpeak spacing works very well for the ions with low charge states.

The Fourier transform (FT) provides a means for determining the frequency of the periodic features. As described earlier, isotopic peaks of an ion are spaced by 1/Z in the isotope-resolved mass spectrum, so Fourier transform of isotope packets can generate useful frequency signals related to the charge state. As shown in Figure 1 a and b, FT is applied to an isolated isotopic cluster with the charge state 5+ . It is observed that its frequency signal exhibits a series of peaks spaced by the charge state. The “DC” signal at zero frequency is ignored because Z < 1 is not allowable. Generally, the first peak in the frequency signal corresponds to the charge state, and the second peak gives the double of the charge state, and so on. However, with the addition of noise signals in-between the isotopic peaks, the frequency peaks may be distorted and not exhibit a clear pattern as shown in Figure 1 c and d. Inaccurate assignment of charge states may be caused. When two or more isotopic clusters are overlapped, FT may fail to determine the charge state using the frequency signal. An example of three overlapping isotopic clusters is shown in Figure 1 e and f. The FT of overlapping clusters illustrates a messy frequency signal without an obvious pattern, from which it is very difficult to determine the charge state.

We develop a novel algorithm based on peak-target Fourier transform for fast charge state determination of complex overlapping isotopic clusters. To illustrate the use of PTFT for charge state determination of ions, two overlapping isotopic clusters with known charge states (5+ and 8+) are simulated and the frequency signal is achieved by FT as shown in Figure 2 a and b. It is known that the intensities of frequency signals are proportional to the abundances of isotopic clusters. When the abundances of two overlapping isotopic clusters differ substantially, the FT still can indicate an accurate charge state for the most abundant cluster as investigated by Senko and colleagues [5]. However, if the abundances of overlapping isotopic clusters are comparable or more than

**Experimental**

All examples of isotopic clusters are selected from three high-resolution FTMS spectra of different complexities, which had been analyzed by AID-MS in our previous study [7, 14]. Spectrum A is an FTICR mass spectrum of a liquid chromatography–separated fraction of commercial acid-extracted histone mixture (Sigma Chemicals) recorded by Bruker Daltonics 9.4T Q-FTMS. The other two spectra were selected from previously published plasma electron capture dissociation (ECD) of large proteins recorded by Cornell 6T FTMS [13]. Spectrum B is a plasma ECD spectrum of ubiquitin and Spectrum C is a plasma ECD spectrum of carbonic anhydrase. All of the time-domain spectra were processed by two zero-fill and no apodization. More details have been described in our previous study. The MATLAB source codes of the PTFT algorithm can be requested from the corresponding author.