Automated Assignment of Charge States from Resolved Isotopic Peaks for Multiply Charged Ions

Michael W. Senko, Steven C. Beu, and Fred W. McLafferty

Department of Chemistry, Cornell University, Ithaca, New York, USA

The recent proliferation of electrospray as an ionization method has greatly increased the ability to perform analyses of large biomolecules by using mass spectrometry. The major advantage of electrospray is the ability to produce multiply charged ions, which brings large molecules down to a mass-to-charge ratio range amenable to most instruments. Multiple charging is also a disadvantage because mass (m) becomes ambiguous unless charge (z) can be assigned. This is typically performed with simple algorithms that use multiple peaks of the same m and different z , but these methods are difficult to apply to complex mixtures and not applicable when only one z appears for each m . The use of mass analyzers with higher resolving powers, like the Fourier transform mass spectrometer, allows resolution of isotopic peaks, providing an internal 1-Da mass scale that can be used for unambiguous charge assignment. Manual assignment of charge state from the isotopic peaks is time consuming and becomes inaccurate when either the signal level or resolving power are low. For these cases, computer algorithms based on pattern recognition techniques have been developed to assist in assignment of charge states to isotopic clusters. These routines provide for more rapid analysis with higher accuracy than available manually. (*J Am Soc Mass Spectrom* 1995, 6, 52–56)

New techniques such as electrospray ionization (ESI) [1, 2] and matrix-assisted laser desorption-ionization (MALDI) [3] have made possible accurate molecular weight determinations for large biological molecules. ESI has the particular benefit that multiple charging shifts the signal of molecules as large as 200 Da down to the mass-to-charge ratio range accessible with most mass analyzers [4]. The tandem mass spectrometry (MS/MS) limit of about 3 Da for singly charged ions also appears to be removed; collisionally-activated dissociation (CAD) spectra are reported for 150-Da ions [5].

The benefit of working with ions amenable to dissociation and analysis in the low m/z range is somewhat offset by the ambiguity of the charge state (z) and therefore mass (m). Typically two peaks of different z , but the same m can be used to solve this ambiguity [6], but this method is difficult to apply to spectra that contain multiple mass values. Deconvolution algorithms have been developed to assist in these cases [7, 8], but are not applicable when only one z appears for each m . For true MS/MS, where one charge state is selected and dissociated, typically fewer z s appear for each m as compared to nozzle-skimmer dissociation,

which uses all molecular ions and provides a range of z s for the products. As an example, the MS/MS spectrum of the $[M + 19H]^{19+}$ of β^A chain of hemoglobin produced 13 peaks whose m/z values correspond to fragments expected from the known sequence, but for only one of these could the m value be assigned from multiple z values [9].

An alternate method for assigning m when only one z appears is the use of an adduct peak [10–12] of known mass that is large enough to be resolved from the primary peak. These methods typically require an additional experiment with the added disadvantage of reduction in signal-to-noise ratio (S/N) levels due to signal consumption in the adduct peaks. When ESI is combined with higher resolving power analyzers like the Fourier transform [13–16] or double-focusing sector mass spectrometer [17, 18], resolution of the isotopic peaks provides an internal 1-Da mass scale, and unambiguous charge assignment [19]. The Fourier transform mass spectrometer (FTMS) [20, 21] is perhaps the ideal mass analyzer for use with ESI because of its unmatched ability to obtain high mass accuracy, high sensitivity, and ultrahigh resolving powers simultaneously over a broad m/z range. Direct identification of z for fragment ions, from precursors as large as 29 kDa, has been extensively demonstrated with FTMS, such as for nozzle-skimmer dissociation [22, 23] and MS/MS by CAD [22–24], infrared multiphoton disso-

Address reprint requests to Dr Fred McLafferty, Cornell University, Department of Chemistry, Baker Laboratory, Ithaca, NY 14853-1301

ciation (IRMPD) [25], and surface induced dissociation [26].

Assignment of z from resolved isotopic peaks for lower charged ions ($z < 10$) is simple when the resolving power and signal-to-noise ratio are high. The reciprocal of the spacing between adjacent isotopic peaks unambiguously provides z . For higher charge states, however, the requirement of mass-measuring accuracy increases with the square of z . For example, the difference in isotopic spacing between components of four and five charges is m/z 0.05, which requires a mass-measuring error of only < 50 ppm at m/z 1000. However, differentiating between z 19 and 20 requires a mass-measuring error of < 3 ppm, which can become challenging even for high performance mass analyzers. In these cases, higher accuracy assignments can be obtained by counting the number of isotopic peaks over a specific m/z range (typically m/z 0.5). The major impediment to obtaining accurate results by using this technique is the necessity to determine what is and what is not an isotopic peak. When either signal or resolving power is low, production of an accurate peak table becomes the leading cause of error. This peak counting method also will fail if two isotopic clusters of different z happen to overlap.

In these cases, pattern recognition techniques, which do not require the production of a peak table, can be exploited for determination of z . Three algorithms that are more applicable to automated analysis have been created to help assign z under these more difficult circumstances. These algorithms all employ information from the entire isotopic envelope and also eliminate the need for a peak table, which thus eliminates a major source of error. Each of these functions is expected to produce maxima that correspond to the repetitive spacing between isotopic peaks, and thus can provide charge state information. The algorithms are even applicable when two isotopic clusters of different z overlap, because each cluster produces an independent maximum. The first method involves the Patterson function [27], which typically is used to construct electron density maps in x-ray crystallography. The second method uses a Fourier transform, which considers isotopic peaks in terms of their frequency of occurrence, and not their spacing. The third method is a combination of the above methods, which makes up for weaknesses when either is used alone.

Experimental

Equine myoglobin and bovine ubiquitin were obtained from Sigma Chemical Co. (St. Louis, MO) and used without further purification. Sustained off-resonance irradiation (SORI) CAD [24, 28] was performed on the molecular ions after selection of individual charge states by stored waveform inverse Fourier transform isolation [29] by using a 6-T ESI-FTMS system previously described [30, 31]. Charge state determina-

tion routines were implemented by using PV-WAVE (Visual Numerics, Houston, TX), a visual data analysis (VDA) system, and the Odyssey VDA interface for PV-WAVE (Waters Extrel, Madison, WI).

The Patterson routine is based on the function

$$P(\Delta M) = \sum_{i=1}^k f(M_i - \Delta M/2) * f(M_i + \Delta M/2),$$

where ΔM is the inverse of the charge being evaluated, M_i is the m/z value of interest, and $f(M_i)$ is the intensity at the m/z value. This function is expected to have a maximum when ΔM is equal to the spacing of the isotopic peaks for the charge state of the cluster. Summation is performed over a user-selected range (M_1 to M_k), and distinct intensity values are determined by linear interpolation between the discrete data points. To provide a smooth charge map, three values are evaluated at each charge ($z - 1/3$, z , $z + 1/3$). Evaluation at $z \pm 1/2$ would speed interpretation, but could provide ambiguous results.

The Fourier routine takes data points from the user-selected area, baseline corrects, pads the data set with zeros to the next power of 2, and produces the magnitude mode spectrum by using a fast Fourier transform (FFT) provided with the VDA. FFTs require that data be spaced linearly on the abscissa, so for FTMS the spectral point spacing is linearized first by using a cubic spline. The Fourier routine has been employed without initially linearizing the data points with the only observed effect being a slight increase in peak widths (but not peak centers) in the charge map. All analysis routines take less than 1 s when performed on a Sun Sparc II.

Results and Discussion

Resolved isotopic envelopes are often of sufficient signal-to-noise ratio and resolving power (Figure 1a) so that charge can be determined directly from the spacing between two adjacent peaks. The two center isotopic peaks are located at m/z 1018.3756 and 1018.4755, $\Delta = 0.0999$, which corresponds to $z = 10$. The direct use of isotopic spacing requires the initial determination of what is and what is not an isotopic peak. With the large number of products possible from nozzle-skimmer [22, 23] and MS/MS spectra [22-26], some isotopic clusters will appear with fairly low S/N and/or resolving power (Figure 1b). Manual assignment in this case, by using the large peaks located at m/z 903.7565 and 903.8300, $\Delta = 0.0735$, would correspond to $z = 13.6$.

Figure 2a shows the results of applying the Patterson technique to the high quality cluster of Figure 1a. The most abundant peak in the charge map corresponds to the correct assignment of $z = 10$. Two distinct features of the charge map are the false maximum

Figure 1. High resolution spectra of the (a) y_{93} and (b) y_{99} fragments from the 17+ molecular ion of equine myoglobin produced by SORI-CAD. Arrows indicate positions of isotopic peaks.

at low z and the rise in the baseline at high z . False maxima appear in Patterson charge maps at even divisions of the true charge, which arise from nonadjacent isotopic peaks separated by distances that correspond to the false charge. There are fewer occurrences of these false separations than of the true separation, so the global maximum should indicate the true charge. In most cases these false maxima can be disregarded, because most appear at nonintegral charges, but it is interesting to note that these false maxima can provide confirmation of the correct charge. The rise in the

Figure 2. Patterson charge maps for (a) high quality and (b) low quality isotopic clusters.

baseline at high z occurs when ΔM becomes smaller than the width of any single isotopic peak and starts at lower z for isotopic clusters acquired with lower resolving powers. Figure 2b shows the Patterson charge map for lower quality isotopic cluster of Figure 1b. The charge map in this case resembles that of Figure 2a, but has less distinct peaks because of the poor S/N levels. Contrary to the manual assignment, automated analysis produces $z = 12$, which corresponds to a preferred fragmentation adjacent to a glutamic acid residue [24]. Successful analysis via the Patterson technique requires the use of sufficiently small step sizes during the summation. It has been determined empirically that 400 points per m/z unit provide satisfactory results. A higher point density does not provide more accurate results, but does slow analysis, whereas lower densities provide faster analysis, but potentially inaccurate results.

The Fourier charge maps for the isotopic clusters (Figure 3) show sharper, more distinct peaks than the Patterson charge map. The Fourier method unfortunately also produces false maxima. All Fourier charge maps have a rise in the baseline at $z < 1$. This is attributed to the Gaussian shape of the isotopic envelope, which shows up as a low frequency component in the Fourier transform. This rise can be disregarded because $z < 1$ is not allowable. For the low quality cluster, analysis was performed over a range $m/z < 1$, which allows the algorithm to ignore any $z < 2$ and provide the correct assignment, although the largest peak in the charge map at $z = 1$ does not correspond to the correct assignment. For spectra with high S/N and resolving powers, false maxima also are observed at multiples of the true charge. These are

Figure 3. Fourier charge maps for (a) high quality and (b) low quality isotopic clusters

caused by the FFT's representation of the isotopic peak shape, which produces harmonics of the fundamental frequency in the Fourier charge map. If the isotopic envelope being analyzed is produced by tandem mass spectrometry, then the charge of the precursor ion is probably known. The charge of the product ion must be less than or equal to the charge of the precursor if the product appears at a lower m/z value than the precursor, and less than the charge of the precursor if the product appears at a higher m/z value than the precursor. This gives the possibility of ignoring most of the anomalous maxima, although the additional peaks again can provide complementary information.

It has been determined that for low charge states ($z < 5$) with high resolving power and low S/N the Patterson routine works best. For high charge states ($z > 5$) with low resolving power as well as low S/N the Fourier routine is best. The Fourier routine does not benefit from excessive resolving power, as the spectral components that define peak shape do not add to the fundamental component, but it will benefit when more isotopic peaks are observed (i.e., higher m and therefore z).

Because one method gives false maxima at high z and another gives false maxima at low z , the obvious progression is to combine the two techniques. This is done by multiplying the two functions [$C(z) = F(z) * P(z)$], because only the true maximum should be present in both maps, and thus should be most abundant in the combination map. This improves the appearance of the charge maps (Figure 4), and thus the confidence in the correct answer. For the high quality cluster, the false maxima of both the Patterson and

Fourier maps have been reduced substantially. For the low quality isotopic cluster, confidence in the assignment also is improved.

For MS/MS spectra produced by either CAD or IRMPD, up to 100 products can be observed between m/z 700 and 1200 [23-25]. This produces a high probability that two isotopic clusters will overlap partially. In these cases, analysis can be confined to the nonoverlapping sections of each cluster. In the unusual case where two clusters completely overlap (Figure 5a), the algorithms are still applicable. Each isotopic cluster produces an independent maximum in the charge state maps, and therefore the overlap of the clusters does not interfere with proper charge state determination (Figure 5b). The abundance of the different peaks in the charge state map will be proportional to the abundances of the corresponding isotopic clusters in the mass spectrum. In a completely automated mode, the algorithms report only the most abundant value, so only the charge state of the most abundant cluster would be reported, with no indication of a second component. Manual investigation of the charge map, however, clearly indicates evidence of two separate components. For analysis of overlapped clusters, the Patterson routine is not recommended because the broad peaks it produces will blend together and potentially produce inaccurate results.

With the large number of peaks produced by MS/MS and nozzle-skimmer dissociation, manual spectral analysis would require several hours and a large number of isotopic clusters would be uninterpretable. By using automated charge state determination, the complete analysis takes less than 2 min, with

Figure 4. Combination charge maps for (a) high quality and (b) low quality isotopic clusters.

Figure 5. (a) Overlapping isotopic clusters of the incompletely ejected 9+ molecular ion of bovine ubiquitin (■) and the y_{59}^{7+} (●) produced by SORI-CAD of the 10+ molecular ion. (b) Combination charge map of the overlapping clusters.

more than 95% of the charge assignments being accurate. Incorrect assignments are usually easy to identify and can be corrected manually with limited effort.

Conclusion

Automated charge state determination algorithms provide a great benefit in speed and accuracy for the analysis of MS/MS spectra of multiply charged ions. The algorithms provide fast, accurate assignments that in many cases would not be possible manually. For isotopic clusters acquired at poor S/N levels, the construction of a peak table has been shown to be the limiting factor in accurate charge assignment, so in these cases it is impractical to base charge state assignment on the spacing between two adjacent isotopic peaks. The use of pattern recognition techniques eliminates the need for a peak table, and thus the primary source of error in charge assignment. The algorithms presented are not redundant, but complementary. In many cases, one method may provide a very ambiguous answer, whereas another provides a more definite assignment of charge. For low charge ($z < 5$) with low S/N and excessive resolving power the Patterson algorithm is best. For higher charge states ($z > 5$) with low resolving power the Fourier algorithm is the best choice. The combination algorithm typically will provide the correct assignment in all circumstances and is recommended for most situations.

Acknowledgments

We gratefully acknowledge the valuable advice and assistance of D. E. Drinkwater, K. D. Henry, P. B. O'Connor, J. P. Quinn, and D. B. Zax, and the financial support of the National Institutes of Health (Grant GM16609) and the Society of Analytical Chemists of Pittsburgh (summer fellowship for MWS).

References

1. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.
2. Loo, J. A.; Udseth, H. R.; Smith, R. D. *Anal Biochem* **1990**, *59*, 158.
3. Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal Chem* **1991**, *63*, 1193A-1202A.
4. Feng, R.; Konishi, Y. *Anal Chem* **1992**, *64*, 2090-2095.
5. Feng, R.; Konishi, Y. *Anal Chem* **1993**, *65*, 645-649.
6. Mann, M.; Meng, C. K.; Fenn, J. B. *Anal Chem* **1989**, *61*, 1702-1708.
7. Reinhold, B. B.; Reinhold, V. N. *J Am Soc Mass Spectrom* **1992**, *3*, 207-215.
8. Labowsky, M.; Whitehouse, C.; Fenn, J. B. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 71-84.
9. Light-Wahl, K. J.; Loo, J. A.; Edmonds, C. G.; Smith, R. D.; Witkowska, H. E.; Shackleton, C. H. L.; Wu, C. C. *Biol Mass Spectrom* **1993**, *22*, 112-120.
10. Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J Am Soc Mass Spectrom* **1993**, *4*, 828-830.
11. Neubauer, G.; Anderegg, R. J. *Anal Chem* **1994**, *66*, 1056-1061.
12. McLuckey, S. A.; Glush, G. L.; Van Berkel, G. J. *Anal Chem* **1991**, *63*, 1971-1978.
13. Henry, K. D.; Williams, E. R.; Wang, B. H.; McLafferty, F. W.; Shabanowitz, J.; Hunt, D. F. *Proc Natl Acad Sci USA* **1989**, *86*, 9075-9078.
14. Henry, K. D.; Quinn, J. P.; McLafferty, F. W. *J Am Chem Soc.* **1991**, *113*, 5447-5449.
15. Hofstadler, S. A.; Laude, D. A. *Anal Chem* **1992**, *64*, 572-575.
16. Winger, B. E.; Hofstadler, S. A.; Bruce, J. E.; Udseth, H. R.; Smith, R. D. *J Am Soc Mass Spectrom* **1993**, *4*, 566-577.
17. Cody, R. B.; Tamura, J.; Musselman, B. D. *Anal Chem* **1992**, *64*, 1561-1570.
18. Dobberstein, P.; Schroeder, E. *Rapid Commun. Mass Spectrom* **1993**, *7*, 861-864.
19. Henry, K. D.; McLafferty, F. W. *Org Mass Spectrom* **1990**, *25*, 490-492.
20. Marshall, A. G.; Grosshans, P. B. *Anal Chem* **1991**, *63*, 215A-229A.
21. Köster, C.; Kahr, M. S.; Castoro, J. A.; Wilkins, C. L. *Mass Spectrom. Rev* **1992**, *11*, 495-512.
22. Loo, J. A.; Quinn, J. P.; Ryu, S. I.; Henry, K. D.; Senko, M. W.; McLafferty, F. W. *Proc Natl Acad Sci. U.S.A* **1992**, *89*, 286-289.
23. Senko, M. W.; Beu, S. C.; McLafferty, F. W. *Anal Chem* **1994**, *66*, 415-417.
24. Senko, M. W.; Speir, J. P.; McLafferty, F. W. *Anal Chem* **1994**, *66*, 2801-2808.
25. Little, D. P.; Speir, J. P.; Senko, M. W.; McLafferty, F. W. *Anal Chem* **1994**, *66*, 2809-2815.
26. Chorus, R. A.; Little, D. P.; Beu, S. C.; Wood, T. D.; McLafferty, F. W. Unpublished.
27. Glusker, J. P.; Patterson, B. K.; Rossi, M. *Patterson and Pattersons*, Oxford Science Publications, 1987, pp 3-19.
28. Gauthier, J. W.; Trautman, T. R.; Jacobson, D. B. *Anal Chim Acta* **1991**, *246*, 211-225.
29. Marshall, A. G.; Wang, T.-C. L.; Ricca, T. L. *J Am Chem Soc* **1985**, *107*, 7893-7897.
30. Beu, S. C.; Senko, M. W.; Quinn, J. P.; McLafferty, F. W. *J Am Soc Mass Spectrom* **1993**, *4*, 190-192.
31. Beu, S. C.; Senko, M. W.; Quinn, J. P.; Wampler, F. M.; McLafferty, F. W. *J Am Soc Mass Spectrom* **1993**, *4*, 557-565.