High Accuracy Molecular Weight Determination and Variation Characterization of Proteins Up To 80 ku by Ionspray Mass Spectrometry*

Rong Feng, Yasuo Konishi, and Alexander W. Bell
National Research Council of Canada, Biotechnology Research Institute, Montreal, Quebec, Canada

A quadrupole mass spectrometer with an ionspray interface was used to measure the molecular weight (MW) of proteins up to 80,000 u. With the improvements in instrument calibration by a statistical averaging method and in data analysis by a gaussian curve-fitting method, precision of MW determination as high as 12 ppm was achieved with equine myoglobin (MW 16,950.4 ± 0.2 u). Exact MW determination of three components in cerato-ulmin revealed that the two minor ones had lost amino acid residues Ser and Ser-Asp, respectively, from the major component (MW 7618.4 ± 0.2 u). MW classification of eight components in the Fab fragment of a monoclonal antibody revealed that one set of four had MW ~ 47,540 u and the other ~ 47,640 u. The MW difference of 100.2 ± 0.6 u between fragment 1 and 2, attributed to inhomogeneous cleavage at the Fab C-terminus, was probably due to one additional Thr in 1. The MW of bovine serum albumin (BSA) was found to be 66,431.5 ± 1.3 u, ~ 164 u higher than the calculated sequence MW, most probably because of the incorrectness in the previously reported BSA amino acid sequence. The MW of human serum transferrin (79,556.8 ± 1.7 u) was shown to be 4414 u higher than the sequence MW, pointing to a glycosylation of 22.7 sugar units in this protein. The greater complexity in bovine serum transferrin (MW 78,030.5 ± 1.8 and 78,326 ± 3.3 u for the two major components) was correlated with the heterogeneity in the glycosylation. *(J Am Soc Mass Spectrom 1991, 2, 387–401)*

Molecular weight (MW) information has been used in biotechnology for protein characterization. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) has been the traditional method for MW classification of these biopolymers. However, its low accuracy (~ 5%) has hindered its application to more elaborate studies, such as differentiation of protein variants having mutations or chemical modifications. Mass spectrometry, on the other hand, has been a well-established technique for accurate mass determination of gaseous ions. Recent development of soft ionization methods, such as matrix-assisted laser desorption [1–3], electrospray [4, 5], ionspray (pneumatically assisted electrospray) [6], and ^{252}Cf plasma desorption [7, 8], has made possible the mass spectrometric analysis of proteins up to 274,800 u [1, 9]. Impressively accurate MW determination for proteins (precision 0.005% ~ 0.2%) has been achieved with electrospray or ionspray ionization techniques combined with quadrupole [4–6, 10–12], Fourier transform (FT) [13], and ion trap (IT) [14] mass analyzers. Such accuracy has made it possible to detect and characterize posttranslational modifications, chemical modifications, mutations, and protease degradations in proteins. However, more accurate MW measurements are required in some cases. For example, mutations between Asp and Asn, Asn and Leu/Ile, Glu and Gln, Glu and Lys, produce mass differences by only 1 u (Table 1), which demand the precision of MW measurements be < 0.5 u for their differentiation.

High accuracy mass determinations of small- or medium-sized molecules have been done traditionally on high resolution instruments, such as double-focusing magnetic sector mass spectrometers or FTMS. However, the majority of the mass measurements of large biopolymers are currently done on quadrupole mass spectrometers with unit resolution. Therefore, it is necessary to explore means to improve the accuracy of high mass measurements on these instruments for solving research problems more precisely. In this article we will describe the approaches we have taken to

*Presented at the 38th ASMS Conference on Mass Spectrometry and Allied Topics, Tucson, Arizona, June 3–8, 1990. This is NRCC Publication No. 32408.

Address reprint requests to Rong Feng, National Research Council of Canada, Biotechnology Research Institute, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada.

1044-0305/91/$3.50

Received January 1, 1991
Accepted February 28, 1991

Table 1. The masses (in u) of 20 common amino acids in free state and in peptide chain

Amino acid	Letter codes	Molecular formula	Monoiso.MW[a] in free state	Aver.MW[b] in free state	Monoiso. mass in peptide chain[c]	Aver. mass in peptide chain[c]
Alanine	Ala (A)	$C_3H_7NO_2$	89.0477	89.0938	71.0371	71.0786
Arginine	Arg (R)	$C_6H_{14}N_4O_2$	174.1117	174.2022	156.1011	156.1870
Asparagine	Asn (N)	$C_4H_8N_2O_3$	132.0535	132.1188	114.0429	114.1036
Aspartic acid	Asp (D)	$C_4H_7NO_4$	133.0375	133.1036	115.0269	115.0884
Asn and/or Asp	Asx (B)	—	—	—	—	—
Cysteine	Cys (C)	$C_3H_7NO_2S$	121.0197	121.1538	103.0091	103.1386
Glutamine	Gln (Q)	$C_5H_{10}N_2O_3$	146.0691	146.1456	128.0585	128.1304
Glutamic acid	Glu (E)	$C_5H_9NO_4$	147.0532	147.1304	129.0426	129.1152
Gln and/or Glu	Glx (Z)	—	—	—	—	—
Glycine	Gly (G)	$C_2H_5NO_2$	75.0320	75.0670	57.0214	57.0518
Histidine	His (H)	$C_6H_9N_3O_2$	155.0695	155.1560	137.0589	137.1408
Isoleucine	Ile (I)	$C_6H_{13}NO_2$	131.0946	131.1742	113.0840	113.1590
Leucine	Leu (L)	$C_6H_{13}NO_2$	131.0946	131.1742	113.0840	113.1590
Lysine	Lys (K)	$C_6H_{14}N_2O_2$	146.1055	146.1888	128.0949	128.1736
Methionine	Met (M)	$C_5H_{11}NO_2S$	149.0510	149.2074	131.0404	131.1922
Phenylalanine	Phe (F)	$C_9H_{11}NO_2$	165.0790	165.1914	147.0684	147.1762
Proline	Pro (P)	$C_5H_9NO_2$	115.0633	115.1316	97.0527	97.1164
Serine	Ser (S)	$C_3H_7NO_3$	105.0426	105.0932	87.0320	87.0780
Threonine	Thr (T)	$C_4H_9NO_3$	119.0582	119.1200	101.0476	101.1048
Tryptophan	Trp (W)	$C_{11}H_{12}N_2O_2$	204.0899	204.2280	186.0793	186.2128
Tyrosine	Tyr (Y)	$C_9H_{11}NO_3$	181.0739	181.1908	163.0633	163.1756
Valine	Val (V)	$C_5H_{11}NO_2$	117.0790	117.1474	99.0684	99.1322

[a] The monoisotopic MW is calculated from the most abundant isotopes of the elements in the molecular formula.
[b] The average MW is the abundance-weighted average of all elemental masses in the molecular formula.
[c] To link N amino acids together to form a peptide chain, N-1H_2O molecules have to be removed. Therefore, the mass of each amino acid residue in a peptide chain is the MW of the free amino acid less the MW of H_2O (18.0106 for the monoisotopic and 18.0152 for the average). The monoisotopic and average MW of a protein can thus be simply calculated by summing up the corresponding masses of all residues in the protein sequence and then adding the corresponding MW of one H_2O; the mass of hydrogen has to be subtracted for each cysteine that forms disulfide bond.

increase protein MW accuracy on a low resolution (~ 2000) quadrupole instrument. The applications that utilize the information from such precise MW determinations for protein characterization and problem solving in biotechnology research will also be described.

Experimental

Instrumentation

A triple quadrupole mass spectrometer (the API III LC/MS/MS system, Sciex, Thornhill, Ontario, Canada) was used for all experiments. The instrument has a mass-to-charge ratio range of 0 ~ 2400 and allows up to 2^{16} (= 65,536) data points for each scan. The system is similar to the prototype instrument described previously [6, 15], except that the mass-to-charge ratio range has been extended from 1400 to 2400 and the instrument tuning, data acquisition, and analysis are now handled by a MacIntosh IIx computer. Multiply charged protein ions were generated by spraying the sample solution through a stainless steel capillary held at high potential. The voltage on the sprayer was usually set between +4 ~ 6 kV for positive ion production. A coaxial air flow along the sprayer was provided to assist the liquid nebulization [15]; the nebulizer pressure was usually adjusted in the range of 25 ~ 35 psi. The sample was delivered to the sprayer by a syringe infusion pump (Model 22, Harvard Apparatus, South Natick, MA) through a fused silica capillary of 100 μm ID. The liquid flow rate was usually set at 0.5 ~ 5 μL/min for sample introduction. The interface between the sprayer and the mass analyzer was made of a small conical orifice of 100 μm diameter. The potential on the orifice was set at +30 V during calibration and was raised to +80 ~ 150 V for proteins to enhance ion signals. A gas curtain formed by a continuous flow (0.6 ~ 1.8 L/min) of N_2 in the interface region served to evaporate the aerosol droplets and to break up the cluster formation from supersonic expansion. All of the above instrumental parameters were carefully adjusted prior to calibration, and also before each protein MW measurement, so as to achieve the highest ion intensity and the least fluctuations in ion flux. For MW determination, only the first rod set (Q1) of the triple quadrupole system was used, and the other two (Q2 and Q3) served only as ion guidance lenses. The instrument mass-to-charge ratio scale was calibrated

with the ammonium adduct ions of polypropyleneglycols (PPG). The unit resolution was maintained across the entire mass range for singly charged PPG calibrant ions, according to the 50% valley definition (i.e., in the group of two adjacent peaks the smaller peak was resolved to at least 50% of its full height). The same resolution setting was used for protein MW measurement to avoid possible interference on accuracy due to resolution adjustment. All protein mass spectra shown were from signal averaging of multiple scans.

Materials and Sample Handling

Equine myoglobin was from Sigma Chemical Co. (St. Louis, MO). The albumins and transferrins were from both Calbiochem Corp. (San Diego, CA) and Sigma. Plant toxin cerato-ulmin and monoclonal antibody Fab fragments were our internal research samples. The PPG calibration standard was supplied by Sciex. No special care was given to optimize the sample concentrations for these measurements, and the infusion flow rates used were chosen to achieve the maximum ion flux stability, not the optimum sensitivity (lower flow rates usually gave greater ion signals). Unless otherwise stated, the following sample concentrations, solvent compositions (v/v), and flow rates were used in the experiments: PPG standard, an H_2O solution of 1:2 mixture of PPG 1000 (1×10^{-4} M) and PPG 2000 (2×10^{-4} M) with 2 mM ammonium acetate, 0.5 ~ 5 μL/min; myoglobin, 2×10^{-5} M, 1:1 H_2O/MeOH with 0.1% formic acid, 3 μL/min; cerato-ulmin, 5×10^{-5} M, 1:1, H_2O/MeOH with 0.1% trifloroacetic acid (TFA), 3 μL/min; Fab fragments, 3×10^{-5} M, 7:1 H_2O/MeOH with 0.1% TFA, 3μL/min; albumins, 3×10^{-5} M, 9:2 H_2O/MeOH with 0.2% TFA, 5 μL/min; transferrins, 5×10^{-5} M, 9:1 H_2O/acetic acid, 1μL/min.

Method of Curve-Fitting Data Analysis

The Cricket Graph (Cricket Software, Inc., Malvern, PA) was used for all gaussian curve-fitting analyses. The normal distribution function, $F(x; \mu, \sigma, H) = H/(\sqrt{2\pi}\sigma)\exp(-(x-\mu)^2/2\sigma^2)$, was used to approximate the top portion of an isolated peak, where x was the mass-to-charge ratio value of each data point, μ the position of the peak apex, σ the standard deviation defining the peak width, and $H/(\sqrt{2\pi}\sigma)$ the preexponential factor determining the peak height. The logarithm of the above function was taken for the least squares polynomial curve-fitting, i.e., $\ln(F) = Ax^2 + Bx + C$, where A, B, and C were constants defining an individual peak profile. The square of curve correlation coefficient (the R^2 value) [16] was used to evaluate the confidence level of a fitting result; the acceptance level was set at $R^2 \geq 0.90$.

The apex of the acquired gaussian profile was used to calculate the average mass of the ion in that particular charge state. No distinction was made between the average mass and the most abundant mass, although strictly speaking the apex of an isotope profile corresponded to the latter. The standard deviation of the calculated ion masses from a series of peaks in the same protein charge distribution was reported as the measurement precision for the protein MW.

Results and Discussion

Instrument Calibration: Single Scan Versus Multiple Scan Averaging

Precise mass calibration is the first step toward achieving high accuracy for MW determinations. On the Sciex instrument, the routine calibration is typically done with a mixture of PPG 1000 and 2000 in the single scan mode. Single scan usually gives adequate calibration precision for singly charged molecules of MW < 2000 u. As shown in Table 2, the single scan calibration of five PPG peaks at m/z 384.30, 520.40, 906.67, 1254.92, and 1545.13 has the average mass-to-charge ratio deviation of ± 0.1. For a singly charged ion, the error introduced to the MW measurement by this imprecision of ± 0.1 in mass-to-charge ratio calibration is negligible. However, for multiply charged ions from large biomolecules, the dependence of MW accuracy on instrument calibration is much greater. First, because of the multicharging phenomenon, the calibration deviation in the instrument mass-to-charge ratio scale is amplified by a factor equal to the number of charges on the ion and is then propagated to the MW error. For example, for a protein ion of mass 20,000 u carrying 20 charges (i.e., at m/z 1001, assuming the charges are due to protonation), a deviation as small as ± 0.1 in mass-to-charge ratio calibration would end up as a ± 2 u ($20 \times \pm 0.1$) error to the measured MW of the protein. It is obvious that the higher the charge state, the more susceptible it is to small experimental errors. Second, because every charge state is independent of each other for the purpose of MW determination (thus offering the convenience of multichannel detection), multiple charge states in a protein charge distribution are usually

Table 2. The effect of multiple scans on mass-to-charge ratio calibration accuracy[a]

		Dwell time on each data point		
Calibration mode	m/z Deviation	10 ms	25 ms	50 ms
Normal (single scan)	$\Delta(m/z)_1$	± 0.09	± 0.1	± 0.1
MCA (10 scans)	$\Delta(m/z)_2$	± 0.01	± 0.01	± 0.01

[a] Five polypropyleneglycol ions at m/z 384.30, 520.40, 906.67, 1254.92, and 1545.13 were selected. All isotope peaks of the polypropyleneglycol ions, except for those of m/z 520.40, which were doubly charged, were resolved to > 50% of their peak heights. It is important to maintain a minimum of 50% resolution during calibration to minimize peak shift due to the convolution effect of adjacent isotope peaks. The mass-to-charge ratio deviations shown, $\Delta(m/z)$, were the average values from all five ions.

taken as a series of independent measurements, with the MW calculated by solving simultaneous equations [6, 10, 17] and the measurement precision reported as the standard deviation of the MW values obtained from various charge states in the same charge distribution [4-6, 10-14, 17]. Therefore, different PPG calibration deviations at various mass-to-charge ratio values will bring MW errors of varying amplitudes to these protein charge states (with the higher charge ones subject to bigger errors), resulting in a poor precision for the final MW value. In principle, the mass-to-charge ratio calibration imprecision should be reduced to ± 0.01 in order to increase the MW precision of a 20 ku protein to ± 0.2 u (relative precision 10 ppm).

To minimize the errors introduced in calibration stage, an alternative calibration protocol was adopted by using the multichannel averaging (MCA) function provided in the instrument control software. The MCA function is essentially a signal averaging routine that accumulates multiple scans to increase the signal-to-noise (S/N) ratio of a spectrum. As shown in Table 2, when the MCA method was applied for instrument calibration, it resulted in much smaller random peak position drifting between calibrations; the average mass-to-charge ratio deviation in the MCA mode with 10 scans was ± 0.01, one order of magnitude smaller than those using the single scan mode.

Three sources of errors may have prevented single scan mass-to-charge ratio calibration from reaching the ± 0.01 precision. First, the statistical fluctuations in the quadrupole power supply may cause random mass drifting on the instrument mass-to-charge ratio scale. Second, the ion flux instability, which is probably due to inhomogeneity in the liquid nebulization process, may produce a virtual apex shift in a peak profile. Third, the small-amplitude high-frequency (SAHF) spikes on a peak profile, which are due to various electrical noises, may cause the apex misassignment by computer peak finding software.

The random fluctuations in the quadrupole power supply follow normal distribution, therefore the probability of a peak apex falling on a given mass-to-charge ratio value can be described by a gaussian function. By taking the average of multiple scans the probability of getting the correct mass-to-charge ratio assignment for each apex is thus increased, as described by the *t*-distribution [16]. Consequently, the mass-to-charge ratio deviations between the calibrations are reduced. In other words, averaging of multiple scans has reduced the random mass drifting on the instrument mass-to-charge ratio scale to a range much narrower, as compared to that of a single scan.

Large amplitude instability in ion flux can cause virtual apex shift in both low and high mass-to-charge ratio directions. As illustrated by the computer simulation shown in Figure 1, if the ion intensity drops by 25% before the mass-to-charge ratio scanning reaches the position of the true peak apex, the resulting new peak apex will shift toward lower mass-to-charge ratio value by $\sim 0.37\sigma$, where σ is the standard deviation in gaussian distribution [in a gaussian-type profile, 2σ corresponds to the peak width at 60.7% of the full peak height (see Figure 1b)]. Similarly, if the ion intensity rises by 25% after the mass-to-charge ratio scanning has passed the position of the true peak apex, the observed peak apex will have a mass-to-charge ratio value $\sim 0.27\sigma$ greater than the real one. If σ takes on a value of 0.5 mass-to-charge ratio unit (the corresponding peak widths at the 0.607 and half height are then $2\sigma = 1.0$ and $2.354\sigma = 1.18$ mass-to-charge ratio units, respectively), these $\pm 25\%$ intensity fluctuations would cause apex shifts equivalent to -0.18 and $+0.13$ mass-to-charge ratio unit, respectively, on the mass-to-charge ratio scale. It is apparent that the larger the intensity fluctuations, the greater the induced apex shifts. It is also clear from Figure 1

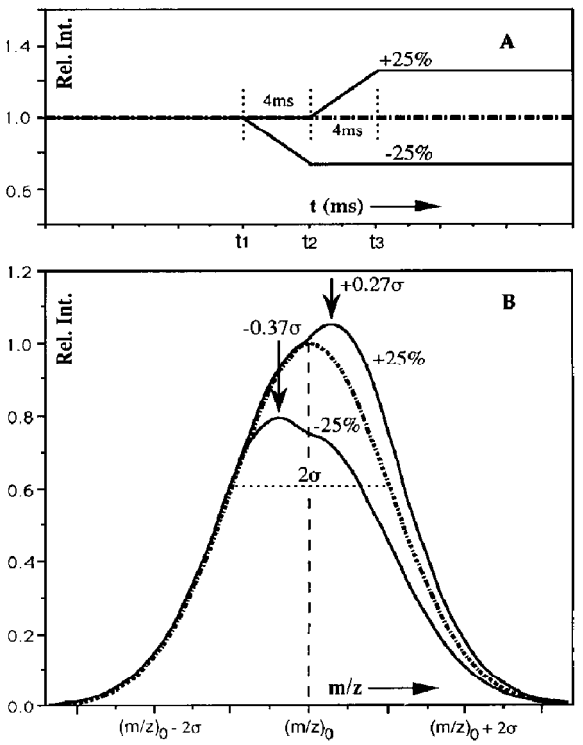

Figure 1. Computer simulation of peak apex shifts due to ion intensity changes during instrument scanning. (a) The ion flux is monitored as a function of time during the scanning of the single gaussian-type peak in (b); the dashed line represents an ideal case in which the ion flux is constant during the entire scanning; the two solid lines indicate the ion flux drops between t_1 and t_2 by 25%, or rises between t_2 and t_3 by 25%. (b) The dashed line represents the true peak profile with its apex positioned at $(m/z)_0$ (arbitrary value); the two solid profiles are the results of the drop and rise of ion flux by 25%, respectively. The vertical arrows indicate the positions of the new peak apexes.

that intensity fluctuations in both low and high directions affect peak shapes, and the distorted peaks appear to have shoulders. Multiscan averaging dilutes these adverse effects (apex shifting and shape distortion) simply by reducing the percentage of contributions from the bad scans.

In Figure 1a, the rise/fall times of the fluctuations that cause apex shifts in Figure 1b are estimated to be ~ 4 ms (assuming the peak width $2\sigma = 1.0$ mass-to-charge ratio unit, each mass-to-charge ratio unit is given 10 data points, and the instrument scans at the speed of 1 ms/point), which is within the fluctuation time scale monitored experimentally. It is also apparent that both the rise/fall time of fluctuation and the peak width have effects on the magnitude of the fluctuation-induced apex shift. Narrow peaks are less affected by the slow intensity fluctuations from the inhomogeneous nebulization of the sample fluid, because they take less time to scan and may experience only a fraction of the intensity variation. On the other hand, peaks from protein ions are usually much broader than those from PPG ions, e.g., the peak width at the half height for the 41^+ ion of bovine serum albumin (BSA) in Figure 9 is ~ 3 mass-to-charge ratio units, six times wider than the usual ~ 0.5 mass-to-charge ratio unit of PPG ions. Consequently, protein peak apexes are much more susceptible to intensity fluctuations. It is important to maintain a stable ion flux during both instrument calibration and protein MW measurement in order to achieve optimum mass accuracy. It also saves time and sample consumption by keeping ion flux stable, because less scans are required to obtain reproducible results. Fluctuations in ion flux can often be substantially reduced by a combination of instrumental adjustments involving the sprayer voltage and position, nebulization pressure, sample infusion rate and curtain gas flow rate.

As for the SAHF noise, the much smoother peak profiles due to the increased S/N ratio from signal averaging would reduce the probability of peak apex misassignment. Long dwell time on the data points takes advantage of the integration effect of the ion detection system and reduces the level of the SAHF noise, so one would expect that longer dwell time would give better calibration accuracy. However, the results in Table 2 indicate that increasing dwell time has little effect on mass accuracy in both single scan and MCA mode. The insensitivity of calibration accuracy to the SAHF noise in the single scan mode is probably because its amplitude is negligible in comparison to the strong ion flux from PPG, or the high frequency noise has been sufficiently averaged out by the relatively long dwell times used (10 ~ 50 ms). Because increasing dwell time has little effect in mass accuracy, it is advantageous to use short dwell times for rapid calibration. Using short dwell times is also advantageous for sample measurements, as multiple fast scans require much less time to accumulate and consume much less sample than slow scans do.

Data Analysis: Conventional Method Versus Curve Fitting Method

Equine myoglobin was chosen as the testing protein for MW determination (ionspray spectrum shown in Figure 2). The average MW of the protein was calculated to be 16,950.5 u, using its amino acid sequence and the average masses of amino acid residues in Table 1. (Unless otherwise noted, all protein MWs in this article are calculated by using the amino acid sequences from the Protein Sequence Database of the Protein Identification Resource, National Biomedical Research Foundation, Georgetown University Medical Center, Washington, DC). When the experimental data were analyzed by the conventional computer routine, i.e., selecting the highest point of each peak, the MW value of 16,949.8 ± 0.7 u was initially obtained. Although the relative error of the measurement was only 0.0041% (41 ppm), a close inspection of individual peak profiles revealed the inadequacy of such approach. Shown in Figure 3 are the top portions of three representative peak profiles taken from $(M + 18H)^{18+}$ (m/z 942.7), $(M + 13H)^{13+}$ (m/z 1304.9), and $(M + 12H)^{12+}$ (m/z 1413.6) ions, respectively. Computer peak-finding routine automatically picks up the highest points of these peaks, even though they may not be the real peak apexes. As in instrument calibration, any imprecision in mass-to-charge ratio assignment during data analysis of multi-

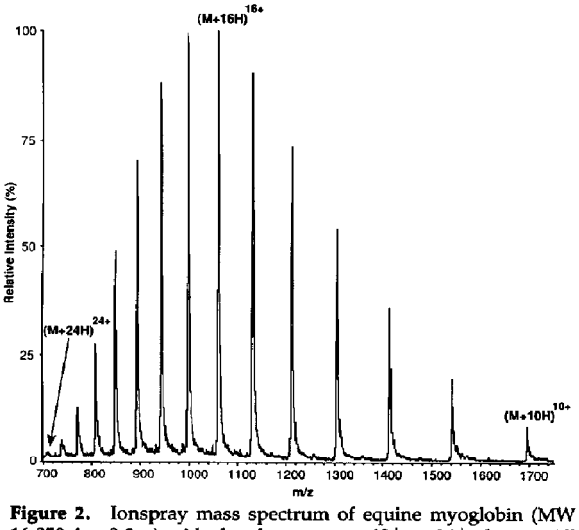

Figure 2. Ionspray mass spectrum of equine myoglobin (MW 16,950.4 ± 0.2 u) with the charge-states 10^+ to 24^+ shown. All positive ions in the charge distribution are due to protonation of the basic amino acid residues (Arg, Lys, and His) and the N-terminal amino group in the protein.

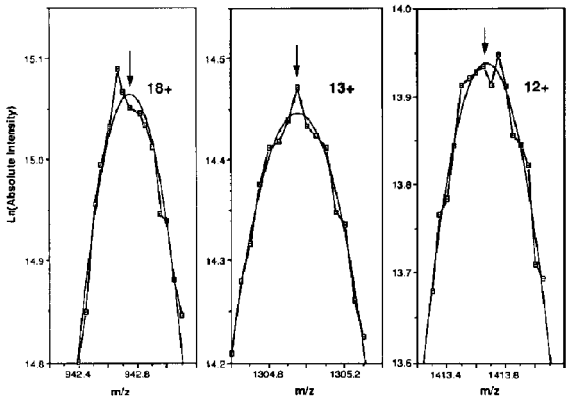

Figure 3. Detailed peak profiles of equine myoglobin ions carrying 18, 13, and 12 positive charges. The intensities are in logarithm scale. The smooth profiles are the fitted gaussian-type curves. The arrows indicate the positions of the expected peak apexes calculated from the amino acid sequence of equine myoglobin.

ply charged ions is multiplied by the number of charges and propagated to the final error of MW determination.

Due to low ion flux and intensity fluctuations, rarely do protein ions give smooth peak profiles, even after many scans of signal averaging. Therefore, apex assignment based purely on peak intensity can result in large errors for such rugged profiles. Obviously, better results would be obtained if these spikes can be "filtered out" by either data smoothing or curve-fitting. We have chosen the latter because of its simplicity and insensitivity to local intensity variations such as large spikes. In principle, the best curve-fitting would be obtained by matching the natural isotope profile against the experimental profile, if the protein sequence (elemental composition) is known and the mathematical expression which describes the profile has been derived. For general applications, however, approximations have to be made. A gaussian-type curve has been chosen because of its simplicity and the commercial availability of the fitting programs.

For multiply charged protein ions in the multi-ku mass range, all isotope peaks have degenerated into a single profile, with the peak apex corresponding to the mass of the most abundant ion in the isotope distribution. Therefore, the peak apex has to be located correctly during data analysis to achieve accurate MW determination. Although the monoisotopic mass of a protein is significantly smaller than its average mass (e.g., 10.5 u smaller in the case of equine myoglobin, as calculated from the monoisotopic masses given in Table 1), the most abundant mass is in fact very close to the average mass [18-20] (theoretical calculations have shown the difference between the two values for equine myoglobin is only 0.5 u [19]). Thus, from the practical point of view one can approximate the average mass of a protein with its most abundant mass.

Proteins in general have unsymmetrical isotope distributions due to the tailings on the high mass sides of their profiles. The degree of asymmetry, however, decreases rapidly with the increase of protein MW. Computer simulations [18-20] have indicated that protein isotope distributions at MW > 10,000 u are close to symmetrical. For experimental profiles obtained using the ionspray method, we have found that their top portions are less distorted from symmetry than their lower parts. The more obvious tailings on the lower parts, which can deviate far from the natural isotope profiles, are probably due to some unresolved small adduct peaks. As a first order approximation, top-fitting (neglecting the lower part of a profile) with gaussian curves appears to give sufficient accuracy. Because the fitted apex of the experimental peak profile, not the peak centroid, is used to derive the MW of a protein, strictly speaking the value obtained corresponds to the most abundant mass in the protein isotope distribution, which is slightly smaller than the protein average MW [18-20]. However, in reality the mass-to-charge ratio position of the fitted apex is slightly higher than the true apex of the experimental peak profile because the symmetrical gaussian curve has been used to fit the top portion of the experimental profile having a slight asymmetry on the high mass side. These two opposite effects cancel each other out to some degree, so the MW value derived from the fitted apex is essentially the average MW of the protein, rather than its most abundant mass. In other words, the net outcome of gaussian curve-fitting is close to averaging (centroiding) the slightly unsymmetrical protein isotope distribution. A small difference still exists between the fitted average MW and the true one; however, it appears rather small (< 0.2 u for equine myoglobin). It stays as a systematic error of the method and can thus be corrected for.

As shown in Figure 3, gaussian curve-fitting effectively eliminated the problems due to noises and spikes, revealing the true peak apexes. By using this method, the average MW of equine myoglobin was determined to be 16,950.4 ± 0.2 u (precision 12 ppm), which was in excellent agreement with the calculated value of 16,950.5 u, demonstrating the applicability of this method and its considerable improvement over the conventional data analysis method.

As discussed previously, fluctuations in quadrupole power supply and ion intensity during the measurement, and misassignments of peak apexes in the data analysis can all bring deviations to the mass-to-charge ratio value of an ion. However, the MW values from ions of lower mass-to-charge ratio are relatively more affected by these deviations because of their higher charge states. Thus, it is often desirable to choose ions of lower charge states (higher mass-to-charge ratio)

for more accurate MW determination. This preference is also important for mixture analysis of closely related protein components, as the neighboring peaks of these components in the same charge state region are better separated at lower charge states, and their MW values are less affected by the convolution effect between the adjacent peaks.

Accumulation of multiple scans inevitably increases sample consumption (the spectra shown in this article usually consumed a few hundreds of picomole of sample; with the parameter optimization in flow rate, concentration and scanning speed, the sample consumption could be reduced to a few tens of picomole). However, the sensitivity of the MW measurements can be improved by using the following strategy. First, a fast scan (one data point for each mass unit and 1 ms dwell time for each data point) over the entire mass range is taken to determine approximately the MW of a protein and the mass-to-charge ratio values of all charge states in the charge distribution. Then, for more accurate MW determination the appropriate narrow MCA scan windows are chosen for the selected peaks (see Figure 6b for an example) and more data points are added (up to 20 data points for each mass unit). Because only limited areas were monitored, the time required to accumulate multiple scans was reduced tremendously, and consequently the sample consumption was kept low. For example, the high sensitivity MW determination was demonstrated with equine myoglobin (MW_{exp} 16,949 ± 5 u, precision 0.03%) by using only 3×10^{-15} mole of sample at the concentration of 2.4×10^{-8} M. Similarly, only 5×10^{-15} mole of sample at the concentration of 5×10^{-9} M was used to determine the MW (2981 ± 1 u, precision 0.03%) of a synthetic peptide. Because ion intensity is not sacrificed for achieving high resolution, using profile analysis on low resolution data also appears to have a sensitivity advantage.

N-Terminal Sequence Determination of Cerato-Ulmin

Cerato-ulmin is proposed to be the toxic agent involved in Dutch elm disease that kills elm trees; the toxin functions by stabilizing microscopic air bubbles that block nutrient passage. Ionspray mass spectrometry revealed that in addition to the major component at 7618.4 ± 0.2 u, there were two other species of lower MW in the sample (Figure 4). Curve-fitting analysis had further pinpointed the sequential MW differences of the three to be 87.1 ± 0.2 u and 115.1 ± 0.2 u, respectively. From the MW table of common amino acids (Table 1), it is apparent that the minor lower MW components have lost amino acid residues Ser (mass 87.1 u) and Asp (mass 115.1 u) sequentially from the native protein. An independent study [21] utilizing the Edman degradation method on a conventional gas-phase protein sequencer has determined that Ser-Asp is the N-terminal sequence of cerato-ulmin, consistent with the findings from the mass spectrometry measurement.

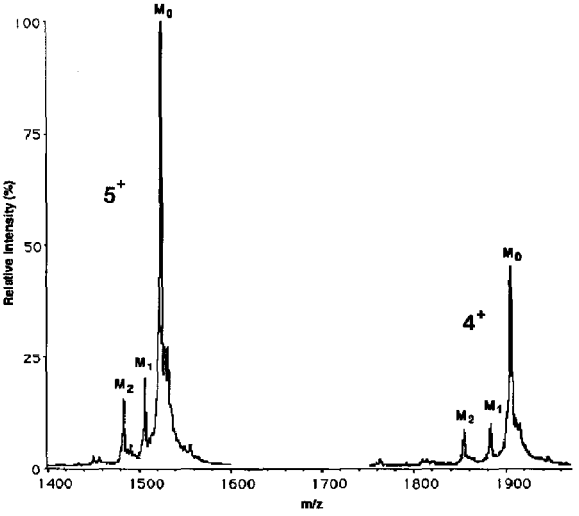

Figure 4. The partial ionspray mass spectrum from cerato-ulmin ions carrying four and five protons. M_0 are the signals from the major component, and M_1 and M_2 are due to two minor components. The peak separations in MW scale are M_0–M_1 = 87.1 ± 0.2 u and M_1–M_2 = 115.1 ± 0.2 u, respectively.

The above data demonstrate that the accurate determination of MW differences between protein components can be used to determine the protein terminal sequence. Although mass spectrometry alone cannot determine whether such partial sequence is from the N- or C-terminus, or from internal deletion, the combination of ionspray mass spectrometry and conventional degradation methods, such as N-terminal peptidase and carboxypeptidase, appears promising for partial sequence determination on either the N- or C-terminus of a protein. The same strategy has been used in the past with other mass spectrometry techniques such as fast atom bombardment (FAB) [22] and ^{252}Cf plasma desorption (PD) [23], and it is expected that ionspray mass spectrometry will be more versatile. Such a combination would offer not only high speed for sample analysis as compared to the conventional high performance liquid chromatography analysis of the liberated amino acid residues, but also diminish the dependence on the enzyme degradation kinetics because only the masses of the protein truncates are measured.

C-Terminal Variation Analysis of Antibody Fab Fragments

A monoclonal antibody (immunoglobulin G) was raised against a single strand RNA to study antibody-RNA interaction [24]. Anion-exchange chromatogra-

phy revealed eight components [24] in its Fab fragment [25, 26] prepared by papain digestion. Papain cleaves the antibody at the site above the disulfide hinge region where one or more disulfide bonds link the two heavy chains together, so the observed differences most probably reflected the variations at the C-terminus of the Fab fragment. Ionspray mass spectrometry analyses were done on all eight separated fractions and all gave similar mass spectra (Figure 5, the spectrum of fraction no. 1 is shown). One set of four components, fractions nos. 1, 3, 5, and 7, were found to have MW ~ 47,640 u and the other set, fractions nos. 2, 4, 6, and 8, ~ 47,540 u (more detailed experimental description and data analysis will be the subjects of a separate publication [24]). Fractions no. 1 and no. 2 were also mixed together (Figure 6) to measure their MW difference more accurately without the interference from occasional instrument drifting between the measurements. Curve-fitting analysis determined their MW to be 47,643.1 ± 0.6 u and 47,542.9 ± 0.5 u, respectively (measurement precision of 13 ppm and 11 ppm, respectively). The MW difference of 100.2 ± 0.6 u may be due to either Val or Thr (masses 99.1 and 101.1 u, respectively; see Table 1), which is probably caused by papain cleaving this immunoglobulin G at two adjacent sites.

As will be discussed in the next section, if two neighboring peaks are not well resolved, the convolution effect will make the observed mass difference between the two components smaller than the real. Taking into account the tailings from the peaks of Fab fragment no. 2 and the convolution effect, it is likely that the observed MW difference between fragment no. 1 and no. 2 is a bit smaller than the true difference, although in Figure 6a the smaller peak (no. 1) has been resolved to 53% of its full height. Thus, the lighter fragment no. 2 is more likely to be missing Thr (101.1 u), rather than Val (99.1 u), from the heavier fragment no. 1. It is obvious that the ambiguity encountered here can be better solved by increasing the instrument resolution to further resolve the neighboring peaks. However, as the limit of the instrument's resolving power has been reached in this case, it might be necessary to take an alternative approach to answer the question more definitively. For example, one might wish to reduce the disulfide bridge in the Fab fragment that links together the light chain (~ 24 ku) and the shortened heavy chain (~ 23 ku), and subject the dissociated smaller pieces to mass spectrometry analysis. This will not only ease the demand on the instrument resolution and give more accurate mass differences between the Fab fragments, but it will also localize the area having the variations. (If it is true, as speculated, that all the variations had occurred on the heavy chains, one would expect the light chains from all eight components to give the identical MW value.)

Mass spectrometry analysis also indicates that, within each set of Fab fragments, the four components differ in MW by only a few units. The fact that they can be separated by anion-exchange chromatography implies that there are small charge differences among them. It is likely that their differences are

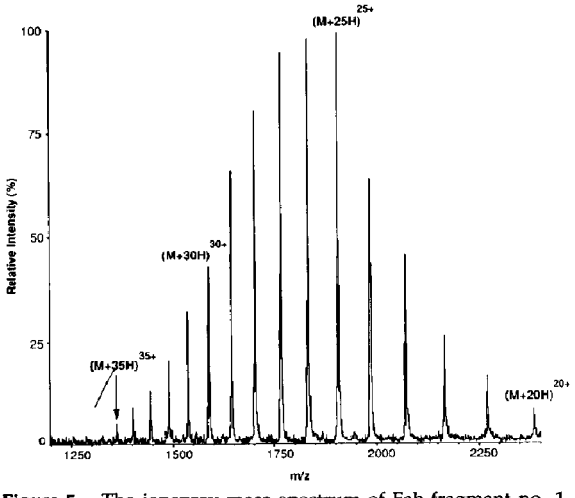

Figure 5. The ionspray mass spectrum of Fab fragment no. 1. Charge states due to protonation of its 20–35 basic amino acid residues are observed. The center of the charge distribution is significantly higher than those of other proteins, which may reflect its significant structural or amino acid composition difference from common proteins.

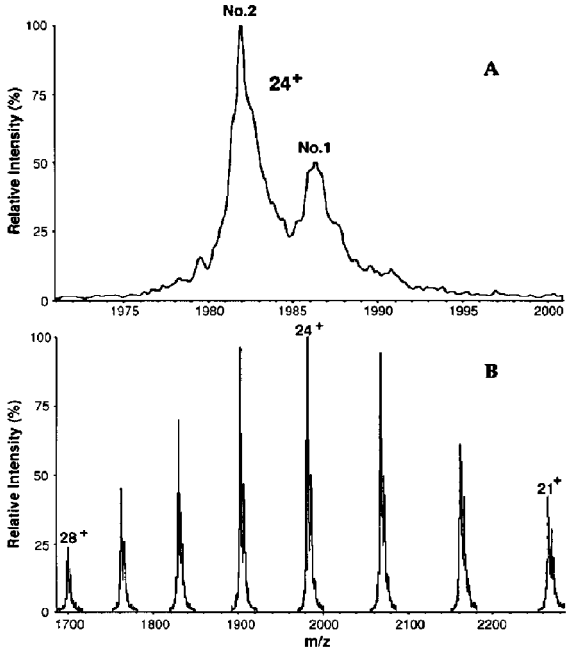

Figure 6. The partial ionspray mass spectrum from the mixture of Fab fragment no. 1 and no. 2. The average MW of the two fragments are 47,643.1 ± 0.6 u and 47,542.9 ± 0.5 u, respectively. The mass difference of 100.2 ± 0.6 u can be due either to Val or to Thr. (a) is the blowup of the 24^+ charge state region.

caused by deamidation of some Gln and/or Asn in the Fab fragment, as each mutation introduces a negatively charged carboxylic group, but brings only 1 u increase to the mass. Investigations are under way to find out exact MW differences of the components between and within the two sets and to probe the structural causes of the subtle MW differences between the four components in each set.

Inward Apex Shift of Partially Resolved Protein Peaks

As seen in Figure 6b, the peak separation between the two components varies according to charge state; the higher the charge state, the smaller the spacing between the two adjacent peaks. For example, the two adjacent peaks at the high mass-to-charge ratio end (21 charges) are separated by ~4.8 mass-to-charge ratio units, but the ones at the low mass-to-charge ratio end (28 charges) by only ~3.5 mass-to-charge ratio units. Although the peak width decreases at the same rate as the peak spacing with the increase of number of charges, the unit mass-to-charge ratio resolution of the instrument has virtually made the low mass-to-charge ratio peaks less resolved than the high mass-to-charge ratio ones.

When two closely positioned peaks are not well resolved, the mutual contribution from each other, i.e., the convolution of the two peaks, will shift their apexes closer toward each other. A simple case where the two adjacent peaks have equal intensity and width is illustrated by the gaussian-type peaks in Figure 7. In Figure 7b, where the two peaks are resolved down to the base line (100% resolution), the observed mass-to-charge ratio separation ($\Delta(m/z)_{obs}$) between the peak apexes reflects the true mass-to-charge ratio difference ($\Delta(m/z)_{true}$). However, when the peak width (2σ) is increased by 2.5 times, as shown in Figure 7a, the profiles of the two original components start to cross each other at 45% of their full height, and the intensities of the two resulting composite peaks increase by 5% due to the intensity contribution from each other, but the resolution between the two peaks decreases to only 13% of the full peak height. At the same time, the peak of lower mass has its apex shifted to a higher mass-to-charge ratio value and the peak of higher mass to a lower mass-to-charge ratio value. This inward apex shift has reduced the observed mass-to-charge ratio difference between the two by 12%. In other words, if the true mass difference in MW scale is 100 u, the observed one would be only 88 u! Plotted in Figure 8a is the computer-simulated relationship between $\Delta(m/z)_{obs}/\Delta(m/z)_{true}$, the ratio between the observed peak separation and the true peak separation, and $\sigma/\Delta(m/z)_{true}$, the peak width to true peak separation ratio, with $\Delta(m/z)_{true}$ kept constant. Similarly plotted in Figure 8b is the relationship between $\Delta(m/z)_{obs}/\sigma$, the observed peak separation to peak width ratio, and $\Delta(m/z)_{true}/\sigma$, the true peak

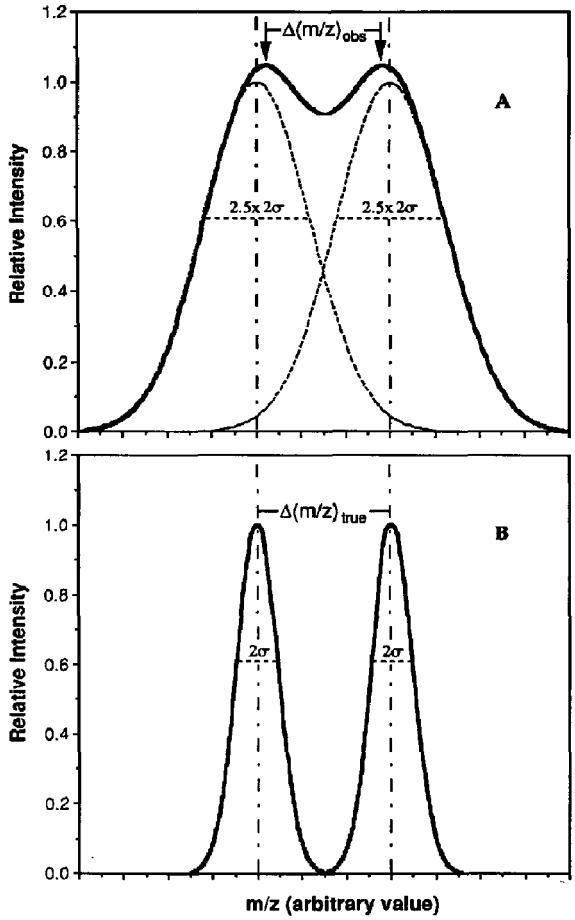

Figure 7. Schematic illustration of (a) how the convolution effect of two not well-resolved gaussian-type peaks reduce the observed peak separation between them, as in comparison to (b) which is the well-resolved peaks showing the true separation. The effect in (a) is generated by increasing the peak width in (b) by 2.5 times.

separation to peak width ratio, with the σ kept constant. As seen, upon reaching either a critical peak width [$\sigma \geq \sim 0.3\Delta(m/z)_{true}$ (Figure 8a)] or a critical peak separation [$\Delta(m/z)_{true} \leq \sim 3.3\sigma$ (Figure 8b)], $\Delta(m/z)_{obs}$ becomes less than $\Delta(m/z)_{true}$, although in the ideal case $\Delta(m/z)_{obs}/\Delta(m/z)_{true} = 1.0$ and it should remain constant in the entire range. With further increase of σ or decrease of $\Delta(m/z)_{true}$, $\Delta(m/z)_{obs}$ drops quickly to zero. It is interesting to note that in both cases if $\Delta(m/z)_{obs}/\Delta(m/z)_{true}$ is plotted against the observed peak resolution (Figure 8c), the critical resolution of ~50% is clearly revealed. If the peaks are not resolved to at least 50% of their full heights, $\Delta(m/z)_{obs}$ is smaller than $\Delta(m/z)_{true}$, and the lower the resolution, the smaller $\Delta(m/z)_{obs}/\Delta(m/z)_{true}$. Fifty percent resolution appears to be critical in general, even when the peaks have different intensities (in such case, 50% resolution

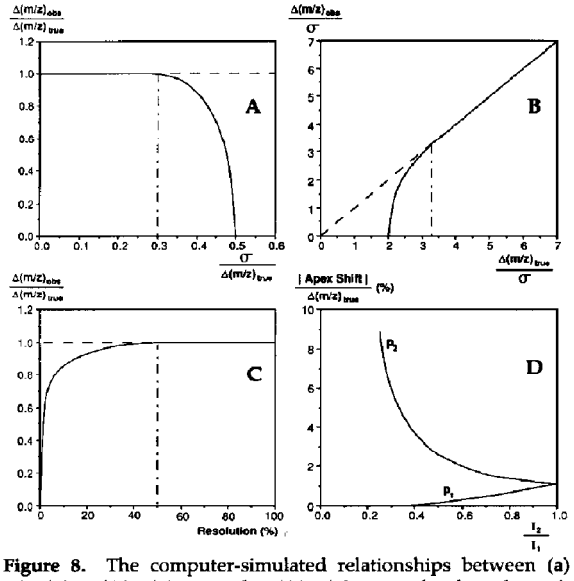

Figure 8. The computer-simulated relationships between (a) $\Delta(m/z)_{obs}/\Delta(m/z)_{true}$ and $\sigma/\Delta(m/z)_{true}$ at fixed peak positions; (b) $\Delta(m/z)_{obs}/\sigma$ and $\Delta(m/z)_{true}/\sigma$ at fixed peak width; (c) $\Delta(m/z)_{obs}/\Delta(m/z)_{true}$ and the observed peak resolution; and (d) the percent shift of each peak relative to $\Delta(m/z)_{true}$ and their relative intensity I_2/I_1, with $\Delta(m/z)_{true}$ kept constant. Gaussian peak profiles are assumed in all calculations. In (d), p_1 denotes the peak whose relative intensity is increasing, and p_2 the peak whose relative intensity is decreasing.

implies that the smaller peak should be resolved to 50% of its full height). However, if the two adjacent peaks are not of equal intensity, after the resolution drops below the 50% point, $\Delta(m/z)_{obs}$ decreases faster than in the case of equal intensity due to the much bigger shift of the smaller peak toward the larger one. Similar to the gravitational attraction, it appears that the smaller peak is "pulled" toward the larger one, while the larger one is relatively much less moved; the larger the intensity difference, the bigger "pulling" the smaller one gets. The effect is illustrated in Figure 8d, where the percent shifts of both peaks relative to $\Delta(m/z)_{true}$ are plotted against the peak intensity ratio I_2/I_1, with $\Delta(m/z)_{true}$ kept constant. At the starting point, the two peaks have equal intensity ($I_2/I_1 = 1.0$) and are resolved to ~36%, and both have ~1% inward apex shift relative to the true mass-to-charge ratio difference between the two. However, when the intensity (I_2) of peak 2 drops, its relative shift increases quickly and reaches to ~9% when the intensity decreases to one fourth of the original value. In contrary, the relative shift of peak 1 is actually diminishing because its relative intensity is increasing (therefore, the MW information from the larger peak is more reliable than from the smaller one). The combined inward apex shift at $I_2/I_1 = 0.25$, however, is now ~9% of the true mass difference, a 4.5 fold increase from the original 2% (= 1% + 1%) at $I_2/I_1 = 1.0$.

As demonstrated above, the amplitude of the inward apex shift of each peak depends on the resolution between the two peaks and is inversely weighted by their relative intensities, with the smaller peak shifting more than the larger one. Therefore, special care must be given when calculating the true MW from the peaks in a poorly resolved peak group, or measuring their mass difference. The peaks must be resolved to at least 50% of their full heights to make the information reliable. The resolution requirement is even higher for unsymmetrical peaks with tailings, as the contribution from the leading peak to the following one is much greater than when they are symmetrical.

Inward apex shifting is a common problem associated with the mixture analysis of closely related protein components. Higher resolution mass analyzers, such as FTMS and double-focusing magnetic instruments, may provide help to solve such problems. On the other hand, mathematical deconvolution may be performed on low resolution data [27] to reveal the true mass difference, as illustrated by the broken profiles in Figure 7a. Work is in progress to find practical ways to circumvent the difficulties involved with deconvoluting the composite peaks made of unsymmetrical profiles whose mathematical expressions are not readily available. (Unlike in curve-fitting where only the top-portion of a profile is used, in deconvolution the entire composite profile has to be decomposed into the original peaks to determine quantitatively their intensity contribution to each other.)

Discrepancy Between Experimental MW and the Literature Sequence MW of Bovine Serum Albumin

Albumin is an abundant blood plasma protein serving as an agent for osmotic regulation and fatty acid transportation. Bovine serum albumin has been subject to electrospray analyses previously [10, 12, 14, 28-32], and the MW values ranging from ~65,280 to ~66,600 u have been reported. In this study the BSA (from Sigma Chemical Co.; the spectrum shown in Figure 9) MW was determined to be 66,431.5 ± 1.3 u (precision 20 ppm), whereas the calculated BSA MW is 66,263.7 ~ 66,267.7 u; the 4 u uncertainty is due to the unknown amidation state of three Asx (Asp or Asn) and one Glx (Glu or Gln) residues. The difference of 164 ~ 168 u between the experimental value and the calculated one could not be due to the measurement error because independent measurements on another instrument half a year later gave the same result. At first, a sample substitution due to a mistake by the supplier was suspected, as the measured MW was very close to that of human serum albumin (HSA) (MW_{calc} 66,436.1 u, MW_{obs} 66,443.8 ± 2.0 u). The supplier performed the cross antibody-affinity test between this BSA sample and HSA, and we did the partial N-terminal sequence determination

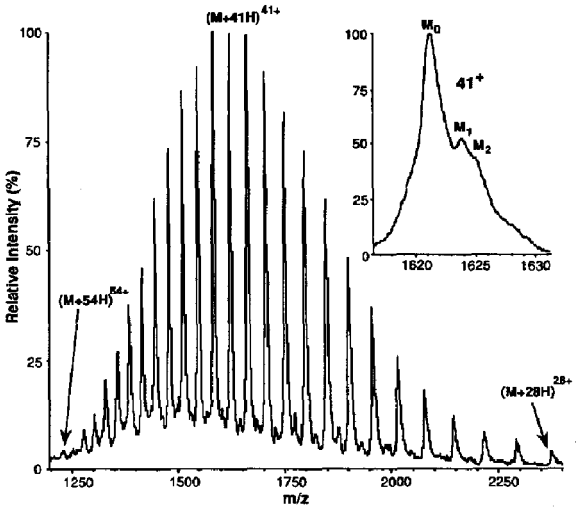

Figure 9. The ionspray mass spectrum of bovine serum albumin (BSA) (MW 66,431.5 ± 1.3 u for M_0). Charge states ranging from 28^+ to 54^+ are observed. The weak peaks between the major peaks are the signals from the odd-number charge states, $(2n + 1)^+$, of the BSA dimer (MW ~ 132,900 u). The inset shows the detailed peak profile of 41^+ charge state region of the monomer. Two minor components (M_1 and its shoulder peak M_2) of unknown origins, at MW ~ 108 and ~ 149 u higher than M_0, are also visible.

by using the gas-phase sequencer in our laboratory. Both experiments, however, indicated the sample to be BSA, not HSA. Further, the BSA sample from another supplier (Calbiochem Corp.) also gave the same MW. Thus, the sample substitution possibility was ruled out. This led to the speculation that BSA might have gone through posttranslational modification, which might have attributed to the observed MW discrepancy. One possibility is the attachment of a hexose unit (Δmass = 180 − 18 = 162 u) to the protein, which would increase its MW to 66,425.7 ~ 66,429.7 u, a value very close to the observed one. Therefore, the sample was further tested utilizing a saccharide specific antibody (Glycan Detection Kit, Boehringer Mannheim Canada Ltd., Laval, Quebec, Canada) along with known glycoproteins as positive controls and nonglycoproteins as negative controls. The positive antibody response obtained for the BSA sample suggested the presence of sugar in this protein, whereas an independent study on BSA [32] has recently shown that a Tyr residue (mass = 163.2 u) is missing at the 156th position from the previously reported BSA sequence, and the corrected sequence MW should be 66,430.3 u. This revised BSA MW is extremely close to our observed value of 66,431.5 ± 1.3 u, demonstrating that the MW value from the ionspray measurement is very dependable and such an MW accuracy can also lead to the finding of the incorrectness in the reported sequence.

However, there seems still to be a conflict between the two studies; the revised BSA MW allows no room for the attachment of one sugar unit to the protein, as has been suggested by the glycan test result. On the other hand, it is possible that the positive antibody response may have been due to the composition inhomogeneity in the BSA sample. As shown in the Figure 9 inset, there are clearly present two other minor components in the sample (M_1 and M_2), with their observed MW being ~ 108 and ~ 149 u, respectively, higher than that of the major component (M_0). Because the intensities of these minor components are only about one half of that of the major component, and M_1 is resolved by only 9% from M_0, and M_2 is nearly unresolved from M_1, there is no doubt that they have been affected by the inward apex shift. The observed MW differences between M_1 and M_0, and M_2 and M_0, are thus significantly smaller than the true ones, mainly due to the downward shifts of the minor components. It is likely that the true MW difference between M_2 and M_0 is actually around 162 u, possibly due to the attachment of a hexose or a hexosamine unit to M_0. Thus, this possible glycosylation on M_2 might have triggered the positive antibody response in the glycan test. However, it is still not clear whether M_2 or M_1 are indeed due to posttranslational modifications on M_0, or if they are just two genetic variants of M_0. Separating all three components and then subjecting them individually to the mass spectrometry and glycan analysis may give more definitive answers on the nature of their differences.

Also visible in Figure 9 are another series of weak signals lying between the peaks of the major component. Mass and charge analyses of these small peaks have revealed that they are actually only one half of the charge states, i.e., the odd-number ones $(2n + 1)^+$, in the full charge distribution of the BSA dimer that has a MW ~ 132,900 u [10]. The other half of the charge states, i.e., the even-number ones $2n^+$, are buried under the strong monomer signals, as the $2n^+$ charge states of the dimer overlap with the n^+ charge states of the monomer.

Characterization of Posttranslational Modifications in Transferrins

Transferrin is also a plasma protein which transports iron to cells and bone marrow. It is known that all transferrins from various species are glycosylated via posttranslational modification [33, 34]. The positive responses to the glycan detection analyses indicated the presence of sugar in these proteins, but gave no information about the degree of glycosylation in them. Mass spectrometry analyses (Figure 10a and b) were performed on the apo-type samples (from Calbiochem Corp.), in which the noncovalently bound ferric ions had been removed, of both human and bovine serum transferrin (apo-HST and apo-BST, respectively). Apo-HST gave an MW value of 79,556.8 ± 1.7 u (relative precision 29 ppm), which was 4413.9 u higher than the calculated sequence MW (75,142.9 u, based

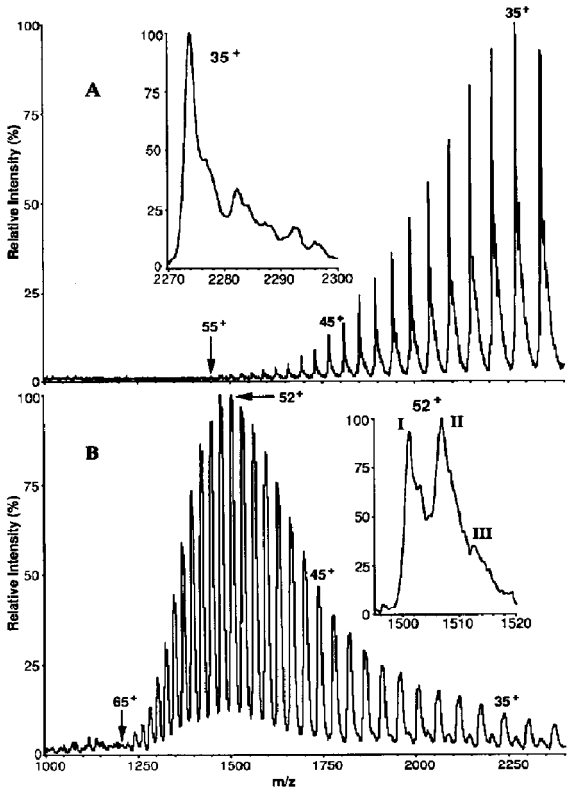

Figure 10. The ionspray spectra of (a) human serum apotransferrin (MW 79,556.8 ± 1.7 u) and (b) bovine serum apotransferrin (MW 78,030.5 ± 1.8 and 78,326.8 ± 3.3 u for the two major components). The center of the charge profile in (a) is much higher than that in (b), possibly due to the differences in the glycosylations or conformations between the two protein homologues. The insets are the blowups of the regions around (a) charge state 35^+ and (b) 52^+, respectively. A second series of small peaks (MW ~ 77,170 u) in (b), which degenerate into the major series at $m/z < 1800$, are also visible in the m/z range of 1800 ~ 2400.

on the C variant amino acid sequence of HST [35] and the mass values in Table 1), indicating a large degree of glycosylation in this protein (5.6% by weight). (Our experimental HST MW value was lower than the one reported in ref 35 by ~13 u. We recalculated the whole MW (sequence MW + masses of modifications) of HST by using the amino acid sequence in ref 35 and the glycosequence of the sugar attachments in ref 36 and obtained an MW value 13.1 u smaller than the one reported in ref 35. It appears that the MW value in ref 35 was based on the less precise mass values of amino acids, which might have introduced a calculation error of ~13 u.) Based on the average mass of 194.7 u for the common sugars found in glycoproteins [37] (fucose, 146.1; galactose, 162.1; mannose, 162.1; N-acetylgalactosamine, 203.2; N-acetylglucosamine, 203.2; N-acetylneuraminic acid, 291.3), it can be estimated that the number of sugar units per protein molecule in HST is approximately 22.7 (= 4413.9/194.7). The mass spectrometry result is in very good agreement with a previous study [36] that shows the modifications to HST are due to two identical oligosaccharide chains (2207.0 u each, total 4414.0 u) attached to two Asn residues in the protein, and each glycochain is composed of 11 sugar units (total 22 units). A few minor peaks have also been observed in the sample of apo-HST (see Figure 10a, inset), which however appear to be due to some noncovalent adducts of unknown identities as their intensities vary with sample preparation.

Similar to the antibody Fab fragments (see Figures 5 and 6), the center of charge distribution in the ionspray spectrum of apo-HST, located around m/z 2200 ~ 2300, is much higher than those of other proteins. For apo-HST it is not due to the lack of basic amino acid residues in the protein, as it has 58 Lys, 26 Arg, and 19 His and their protonation is more than sufficient to bring the center of the charge distribution to around m/z 1500, where the apo-HST ion needs to have only 53 charges. Although protein backbone cross linkages due to multiple disulfide bridges (S—S bonds) sometimes cause protein charge profiles to shift to higher mass-to-charge ratio values [13, 38, 39], it is not clear whether they have a significant influence here. The number of S—S bonds and their structural arrangement are usually conserved within the same protein family, but the charge distribution of apo-HST's homologue apo-BST obtained under the same experimental conditions (i.e., the same solvent composition and orifice voltage, etc.) has the "normal" appearance (see Figure 10b). It is plausible that the profile difference between the two transferrins may have been caused by their difference in glycosylation, such as the content and the structural arrangement of the sugars. The sugar chains in HST may have prevented the protein from achieving full denaturation under the solvent conditions used (H_2O solution with 10% acidic acid), thus resulting in less basic amino acids exposed for protonation. If this is true, it may also account for the unusually high mass-to-charge ratio profiles of intact 150 ku antibodies observed on our instrument [40], as it is well known that antibodies are heavily glycosylated [25]. On the other hand, it is also possible that apo-HST may assume a conformation tighter than that of apo-BST under the same solvent condition, as it has been shown that protein conformational changes can also induce charge profile shift [41]. Mass spectrometry analysis of these proteins after stripping off their sugars and/or systematically varying solvent compositions may give more definitive answers.

Although the MW of apo-HST is > 10 times higher than that of certoa-ulmin (MW 7618.4 u), surprisingly its peak width (see Figure 10a, inset) at 60.7% of the full height (corresponding to 2σ) is only 2.0 mass-to-charge ratio units, which has changed little from the

1.8 mass-to-charge ratio unit peak width of ceratoulmin (Figure 4). On the contrary, the iron-free HST sample from another supplier (Sigma Chemical Co., spectrum not shown) exhibited very broad peaks (2σ = 36 mass-to-charge ratio units). This peak broadening was certainly not due to insufficient instrument resolution as the experiments were done under similar conditions. Run under the highest resolving power possible, multiple slightly resolved peaks with equal spacing between them became visible on top of those broad peaks, giving indication of multi-adducts formation. These observations suggest that under the ionspray conditions the increase of protein MW itself brings little change to the peak widths, due to the compensation effect of the growing number of charges, and the major causes of peak broadening come from adducts attachment and/or sample composition heterogeneity. Peak broadening of such a nature not only severely hinders the extraction of analytically useful information from the spectra, but also reduces the effective mass range of a mass spectrometer. For example, if a protein sample contains only a single component and all noncovalent adducts have been removed, it can be estimated that with peak widths keeping at 2σ = 2.0 mass-to-charge ratio units, the highest mass that can be measured with high accuracy in the instrument range of m/z 2000 ~ 2400 is around 1.2 ~ 1.8 million u (see Appendix). However, the values are reduced by 18 times to 67 ~ 97 ku, if adducts formation broadens the peak widths to 2σ = 36 mass-to-charge ratio units.

In contrast to apo-HST, apo-BST, which is known to possess the most complex heterogeneity among all transferrins [42], displays two major components of approximately equal intensity (peaks I and II in Figure 10b, inset), in line with the mass spectrometry observation reported earlier [10, 43]. In our study the MW values of the two major components were determined to be 78,030.5 ± 1.8 and 78,326.8 ± 3.3 u, respectively. Although the complete amino acid sequence and the identity of the sugar attachment of BST are still not available, if one assumes that BST also has ~5.6% sugar content, then the unmodified protein probably has an MW around 74,000 u, a value not far away from 75,142.9 u of its human homologue HST. The MW difference of 296 ± 3.3 u between the two major apo-BST components is probably due to the heterogeneity in the glycosylation, and the value seems to suggest one additional sialic acid (N-acetylneuraminic acid, mass = 291.3 u) in the second component. A third minor component (peak III), whose observed MW is ~282 u higher than the second, is also visible in the Figure 10b inset. Taking into account its low intensity and the consequent inward apex shift due to the convolution effect, the actual MW difference is surely bigger than what is observed; it is probably also around 291 u. Similarly, the apex of peak II may have also shifted slightly toward peak III, making the observed MW difference between peaks I and II appear a bit larger than the expected 291 u. The mass spectrometry results here agree fairly well with the earlier electrophoresis studies [42, 44, 45], which have shown two pairs of main bands and one pair of minor bands, with the electromobility differences between the three pairs being attributed to the presence, on top of other forms of sugars, of two and three sialic acids in the two pairs of major components, respectively, and four sialic acids in the third pair of minor components. These three pairs of bands probably correspond to peaks I, II, and III in the Figure 10b inset.

The electromobility differences within the three pairs have been demonstrated [42] to be all due to the same internal cleavage in the C-terminal domain of the BST polypeptide chain. However, the data give no clear indication whether any amino acids have been expelled at the site of scission. The posttranslational modification due to a simple protein backbone scission without loss of amino acids would cause an MW difference of 18 u between the two components within the pair. But a mixture of such cannot be resolved by mass spectrometry at the current instrumental resolution. However, a loss of one or more amino acid residues would be detected, provided that the loss causes a mass difference ~100 u between the two components. On the other hand, the following proposed reduction-alkylation experiments [39] may help to distinguish them mass spectrometrically and reveal exactly where the internal cleavage takes place and which amino acid residues have been expelled (if any). Upon reduction of the S—S bond that holds together the two polypeptide chains and end-capping of the resulting free thiols with an alkylating agent, the variant with the internal cleavage will separate into two pieces of smaller sizes. Exact MW determination of the two fragments would enable one to calculate the MW difference between the intact protein and the one with cleavage, which would in turn reveal the size of the excision (i.e., the number of amino acids expelled). By comparing the MW of the smaller C-domain peptide with the available C-terminal sequence, one may also locate the site of the scission.

The small shoulder peaks on the big ones in the Figure 10b inset are most likely due to the sulfate or phosphate adducts [46] as their intensities vary with sample preparation. Although the observed mass differences are smaller than the expected 98 u for the sulfate or phosphate binding, for times they go up (from 73 to 95 u) along with the intensity increase of these adducts (again demonstrating the effect of intensity on the amplitude of inward apex shifting). However, our attempts to use higher orifice voltage to remove collisionally [46] these noncovalent adducts have not resulted in significant reduction in their abundances, in contrast to our success with a sample of sulfate-infested myoglobin. It is possible that in

apo-BST, unlike in myoglobin which has an extended linear structure once fully denatured, some of the noncovalent adducts may be trapped inside the pockets formed by the tight backbone loops that may not be fully relaxed due to the multi-disulfide cross linkages and may therefore be less sensitive to the mild collisional shake-off. The same attempts with an adducts-contaminated sample of papain were similarly less successful [R. Feng et al., unpublished results], possibly due to the same kind of structural restriction.

It is worth noting that there is a second series of broad small peaks in the range of m/z 1800 ~ 2400 in Figure 10b (the peaks below m/z 1800 have all collapsed into the major series), which are probably all unresolved doublets (or triplets) as in the Figure 10b inset. These weak signals (MW 77,167 ± 13 u) may correspond to a minor component (or components) with less sugar content than the major ones. The MW difference of 863 u between the two series suggests that the minor component(s) have ~4 sugar units less than the component(s) under peak I.

Studies are under way to characterize further the posttranslational modification in transferrins.

Conclusions

It has been demonstrated that a low resolution quadrupole mass spectrometer is capable of high accuracy MW determination for proteins. By carefully calibrating the instrument using multiple scan averaging and by analyzing the data with a more precise method such as curve-fitting, measurement precision as high as 10 ppm can be achieved. Although isotope peaks are not resolved, profile analysis with simple gaussian curve-fitting is sufficient for determining the average MW of a protein. Because ion intensity is not sacrificed for achieving high resolution, using profile analysis on low resolution data may also have the sensitivity advantage.

The exact MW differences between protein components can be used, in conjunction with other techniques, for the determination of protein partial sequence and for the analysis of terminal inhomogeneity. High accuracy MW values can be used for the detection and characterization of posttranslational modifications in proteins and may also lead to the findings of protein sequence errors in the literature.

Acknowledgments

We thank Sciex, Thornhill, Ontario, Canada, for inviting R. Feng as a visiting scientist and for making available the API III LC/MS/MS system during the early stage of this study; the valuable contributions from T. Covey, W. Davidson, G. King, G. Scott, B. Shushan, B. Thomson, and many others at Sciex are gratefully acknowledged. We are also grateful to our colleagues F. Bouthillier, M. Cygler, Y. Li, A. Storer, S. Takai, and M. Yaguchi for bringing to us their intriguing research problems and for insightful discussions, and to F. Dumas and and B. Gibbs for their valuable experimental assistance.

References

1. Karas, M.; Bahr, U.; Ingendoh, A.; Hillenkamp, F. *Agnew. Chem., Int. Ed. Engl.* **1989** *28*, 760-761.
2. Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1990**, *62*, 1836-1840.
3. Spengler, B.; Cotter, R. J. *Anal. Chem.* **1990**, *62*, 793-796.
4. Fenn, J.B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.
5. Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R. *Anal. Chem.* **1990**, *62*, 882-899.
6. Covey, T. R.; Bonner, R. F.; Shushan, B. I.; Henion, J. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 249-256.
7. Macfarlane, R. D.; Sundqvist, B. U. R. *Mass Spectrom. Rev.* **1985**, *4*, 421-460.
8. Jonsson, G.; Hedin, A.; Hakansson, P.; Sundqvist, B. U. R.; Bennich, H.; Roepstorff, P. *Rapid Commun. Mass Spectrom.* **1989**, *3*, 190-191.
9. Hillenkamp, F. *Proceedings of the 37th ASMS Conference on Mass Spectrometry and Allied Topics*, Miami Beach, FL, May 21-26, 1989, pp 1168-1169.
10. Loo, J. A.; Udseth, H. R.; Smith, R. D. *Anal. Biochem.* **1989**, *179*, 404-412.
11. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Mass Spectrom. Rev.* **1990**, *9*, 37-70.
12. Chowdhury, S. K.; Katta, V.; Chait, B. T. *Rapid Commun. Mass Spectrom.* **1990**, *4*, 81-87.
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. Berkel, G. J. V.; Glish, G. L.; McLuckey, S. A. *Anal. Chem.* **1990**, *62*, 1284-1295.
15. Bruins, A. P.; Covey, T. R.; Henion, J. D. *Anal. Chem.* **1987**, *59*, 2642-2646.
16. Miller, I.; Freund, J. E. *Probability and Statistics for Engineers*, 3rd ed.; Prentice-Hall: Englewood Cliffs, NJ, 1985.
17. Mann, M.; Meng, C. K.; Fenn, J. B. *Anal. Chem.* **1989**, *61*, 1702-1708.
18. Yergey, J.; Heller, D.; Hansen, G.; Cotter, R. J.; Fenselau, C. *Anal. Chem.* **1983**, *55*, 353-356.
19. Hsu, C. S. et al. Private communications.
20. Wang, B. H.; Williams, E. R.; Henry, K. D.; McLafferty, F. W. *Proceedings of the 37th ASMS Conference on Mass Spectrometry and Allied Topics*, Miami Beach, FL, May 21-26, 1989, pp 236-237.
21. Yaguchi, M. et al. Private communications.
22. Bradley, C. V.; Williams, D. H.; Hanley, M. R. *Biochem. Biophys. Res. Commun.* **1982**, *104*, 1223-1230.
23. Klarskov, K.; Breddam, K.; Roepstorff, P. *Anal. Biochem.* **1989**, *180*, 28-37.
24. Cygler, M. et al. Manuscript in preparation.
25. Lehninger, A. L. *Biochemistry*, 2nd ed.; Worth Publishers: New York, 1975; p 1002.
26. Walker, J. M.; Cox, M. *The Language of Biotechnology*; American Chemical Society: Washington, DC, 1988, p 127.
27. Siegel, M. M.; Tsao, R.; Doelling, V. W.; Hollander, I. J. *Anal. Chem.* **1990**, *62*, 1536-1542.
28. Berkel, G. J. V.; McLuckey, S. A.; Glish, G. L. *Proceedings of the 38th ASMS Conference on Mass Spectrometry and Allied Topics*; Tucson, AZ, June 3-8, 1990, pp 12-13.
29. Chowdhury, S. K.; Katta, V.; Chait, B. T. *Proceedings of the 38th ASMS Conference on Mass Spectrometry and Allied Topics*, Tucson, AZ, June 3-8, 1990, pp 138-139.
30. Feng, R.; Konishi, Y.; Bell, A. *Proceedings of the 38th ASMS*

Conference on Mass Spectrometry and Allied Topics, Tucson, AZ, June 3–8, 1990, pp 275–276.
31. Henry, K. D.; Wang, B. H.; Williams, E. R.; Alexander, A. J.; Quinn, J. P.; McLafferty, F. W. *Proceedings of the 38th ASMS Conference on Mass Spectrometry and Allied Topics*, Tucson, AZ, June 3–8, 1990, pp 433–434.
32. Hirayama, K.; Akashi, S.; Furuya, M.; Fukuhara, K. *Biochem. Biophy. Res. Commun.* **1990**, *173*, 639–646.
33. Aisen, P.; Listowsky, I. *Ann. Rev. Biochem.* **1980**, *49*, 357–393.
34. Brock, J. H. In *Metalloproteins, Part 2*; Harrison, P. M., Ed.; Verlag Chemie: Weinhein, 1985; p 183.
35. MacGillivray, R. T. A.; Mendez, E.; Shewale, J. G.; Sinha, S. K.; Lineback-Zins, J.; Brew, K. *J. Biol. Chem.* **1983**, *258*, 3543–3553.
36. Dorland, L.; Haverkamp, J.; Schut, B. L.; Vliegenthart, J. F. G.; Spik, G.; Strecker, G.; Fournet, B.; Montreuil, J. *FEBS Lett.* **1977**, *77*, 15–20.
37. Stryer, L. *Biochemistry*, 3rd ed.; W. H. Freeman: New York, 1988, p 344.
38. Loo, J. A.; Edmonds, C. G.; Udseth, H. R.; Smith, R. D. *Anal. Chem.* **1990**, *62*, 693–698.
39. Feng, R.; Bell, A.; Dumas, F.; Konishi, Y. *Proceedings of the 38th ASMS Conference on Mass Spectrometry and Allied Topics*, Tucson, AZ, June 3–8, 1990, pp 273–274.
40. Feng, R.; Konishi, Y. *Proceedings of the 39th ASMS Conference on Mass Spectrometry and Allied Topics*, Nashville, TN, May 19–24, 1991, in press.
41. Chowdhury, S. K.; Katta, V.; Chait, B. T. *J. Am. Chem. Soc.* **1990**, *112*, 9012–9013.
42. Maeda, K.; McKenzie, H. A.; Shaw, D. C. *Anim. Blood Grps. Biochem. Genet.* **1980**, *11*, 63–75.
43. Edmonds, C. G.; Loo, J. A.; Barinaga, C. J.; Udseth, H. R.; Smith, R. D. *J. Chromatog.* **1989**, *474*, 21–37.
44. Richardson, N. E.; Buttress, N.; Feinstein, A.; Stratil, A.; Spooner, R. L. *Biochem. J.* **1973**, *135*, 87–92.
45. Stratil, A.; Spooner, R. L. *Biochem. Genet.* **1971**, *5*, 347–365.
46. Chowdhury, S. K.; Katta, V.; Beavis, R. C.; Chait, B. T. *J. Am. Soc. Mass Spectrom.* **1990**, *1*, 382–388.

Appendix

Assuming two peaks of neighboring charge states are at

$$(m/z)_n = (M + nM_c)/ne \quad (1)$$

and

$$(m/z)_{n+1} = [M + (n + 1)M_c]/(n + 1)e \quad (2)$$

where M is the mass of the protein, n the number of charging species it carries, M_c the mass of the charging species, and e its charge unit, then the separation between the two peaks is simply

$$\Delta(m/z) = (m/z)_n - (m/z)_{n+1} = M/n(n + 1)e \quad (3)$$

With a simple mathematical manipulation, eq 3 can be transformed to

$$\Delta(m/z) = [(m/z)_{n+1} - M_c/e]/n \quad (4)$$

Rearranging eq 4, one obtains

$$n = [(m/z)_{n+1} - M_c/e]/\Delta(m/z) \quad (5)$$

Substituting eq 5 back into eq 2 and solving it for M, one obtains

$$M = e[(m/z)_{n+1} - M_c/e] \times [(m/z)_{n+1} - M_c/e + \Delta(m/z)]/\Delta(m/z) \quad (6)$$

For large protein ions $(m/z)_{n+1} \gg M_c/e$ and $\Delta(m/z)$, so in eq 6 these two minor terms can be neglected and the expression simplifies to

$$M = e(m/z)^2_{n+1}/\Delta(m/z) \quad (7)$$

For the charges due to protonation, then e = 1 and eq 7 changes to:

$$M = (m/z)^2_{n+1}/\Delta(m/z) \quad (8)$$

If the critical peak separation, at which no peak apex shifting occurs, is taken as $\Delta(m/z) = 3.3\sigma$ (i.e., the peaks are resolved to at least 50% of their full heights, see Figure 8), one would have:

$$M = (m/z)^2_{n+1}/3.3\sigma \quad (9)$$

It is obvious from eq 9 that at a given mass to charge ratio value, the narrower the peak width, the higher the mass that can be measured accurately (not just detected). For peak width $2\sigma = 2.0$ mass-to-charge ratio units and the selected instrument working range of m/z 2000 ~ 2400, the highest mass measurable would be 1,212,121 ~ 1,745,455 u. In contrast, if the peak width increases by 18 times to $2\sigma = 36$ mass-to-charge ratio units due to adducts formation or sample heterogeneity, then the highest mass measurable would be reduced to 67,340 ~ 96,970 u. It is also apparent in eq 9 that increasing the instrument mass-to-charge range will also greatly raise the limit on the highest measurable mass, as M is in a square relationship with mass-to-charge ratio.