Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Mass Spectrometry 246 (2005) 1–9
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

Charge ratio analysis method to interpret high resolution electrospray Fourier transform—ion cyclotron resonance mass spectra

Simin D. Maleknia [a,b,*], Kevin M. Downard [c]

[a] *School of Science, Griffith University, Nathan Campus, Brisbane, Qld 4111, Australia*
[b] *Sciformatics, Sydney, Australia*
[c] *School of Molecular & Microbial Biosciences, The University of Sydney, Australia*

Received 11 July 2005; received in revised form 1 August 2005; accepted 1 August 2005
Available online 16 September 2005

Abstract

The charge ratio analysis method (CRAM) is a new approach for the interpretation of high resolution Fourier transform ion cyclotron resonance (FT-ICR) electrospray mass spectral data. The high resolution capability of FT-MS provides resolved isotopic peaks of multiply charged ions of biopolymers enabling their accurate and monoisotopic molecular mass determination. It does, however, require that the correct charge and isotope composition of these ions be assigned in order for this accuracy to be realized. The unique feature of the CRAM in processing the FT-ICR data is that the charge states of ions are identified from analysis of the ratios of m/z values of isotopic peaks of different multiply charged ions. In addition, the CRAM process correlates the isotopic peaks of different multiply charged ions that share the same isotopic composition. As the size of biopolymers increases, their isotope patterns become more uniform and more difficult to discern from one another. This impacts on the correct matching of a theoretical isotope distribution to experimental data particularly in the case of biopolymers of unknown elemental compositions. The significance of the CRAM is demonstrated in terms of correlating theoretical isotopic distributions to experimental data where this correlation could not always be achieved based on the relative intensities of isotopic peaks alone. While for high resolution FT-ICR mass spectral data, the ion charge can be otherwise determined from the reciprocal of the m/z difference between adjacent isotopic peaks, the CRAM approach is superior and determines ion charge with several orders of magnitude higher accuracy. The CRAM has been applied to high resolution FT-ICR mass spectral for several proteins (ubiquitin, cytochrome *c*, transthyretin, lysozyme and calmodulin) to demonstrate the general utility of this approach and its application to proteomics. The results have been discussed in terms of internally calibrated ions versus external calibration where the CRAM approach was superior in both cases.

Keywords: Charge ratio; Charge state; ESI; Isotope distribution; FT-ICR

1. Introduction

Electrospray ionization (ESI) [1] in combination with Fourier transform ion cyclotron resonance (FT-ICR) [2] mass spectrometry is now routinely utilized for accurate mass determination of biopolymers [3–7] and is incorporated in proteomics applications. The electrospray ionization process generates highly charged ions that facilitate the detection of macromolecules across a mass-to-charge (m/z) range of 500 to ~2000. FT-ICR mass analyzers afford mass resolving powers of up to 10^6 in direct proportion to the magnetic field strength [2]. This high mass resolution capability resolves the multiply charged ions according to their isotopic composition. Both the accurate determination of the ion charge (z), and the correct identification of an ion's isotopic composition are required to obtain a mass accuracy in the low part-per-million (ppm) range for biopolymers [3,4]. The ion charge is conventionally determined as the reciprocal of the separation between adjacent ion peaks of an isotope cluster [8]. This method of charge assignment is reliable for ions with up to 10 charges. Beyond this charge level, however, differentiating between consecutive charges requires a high

* Corresponding author. Tel.: +61 7 3735 7894; fax: +61 7 3735 6572.
E-mail addresses: s.maleknia@griffith.edu.au, maleknia@sciformatics.com (S.D. Maleknia).

1387-3806/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2005.08.002

mass accuracy. In the case of ions of charge $z = 19$ and 20, a mass accuracy of less than 3 ppm is needed which requires the application of careful calibration procedures even on high resolution mass spectrometers [8,3].

Here, we report on the application of the charge ratio analysis method (CRAM) [9] to determine the ion charge at several-orders of magnitude higher accuracy in comparison to the isotope spacing method. The CRAM is based on the simple ratio of m/z values of multiply charged ions and requires no prior knowledge or assumption of the nature of the charge carrying species. In addition, the CRAM can correlate isotope peaks for multiply charged ions that share a common isotopic composition, another vital factor in order to achieve a mass determination for biopolymers with low ppm mass accuracy. Biopolymers are primarily composed of the common elements carbon, hydrogen, nitrogen, oxygen and sulfur. Of these only carbon has a heavy isotope form with a natural abundance that exceeds 1%. Nonetheless, as biopolymers contain large numbers of these elements, they exhibit complex isotopic patterns that reflect the combined distributions of each naturally abundant isotope [10–12]. Within the isotopic distribution, only one resolved isotopic peak associated with molecules that contain only the lightest isotopes (i.e., $^{12}C_a{}^{1}H_b{}^{14}N_c{}^{16}O_d{}^{32}S_e$), from which the monoisotopic mass of a molecule is derived, has a unique isotopic composition [6,10]. The remaining isotopic peaks reflect molecules with some level of heavy isotope enrichment where each peak results from the superposition of several isotope variants. For example, the second peak of an isotope distribution ($M + 1$) results from a superimposition of ion peaks for molecules that contain one heavy isotope for each element, such as $^{12}C_{a-1}{}^{13}C_1{}^{1}H_b{}^{14}N_c{}^{16}O_d{}^{32}S_e$ (for carbon), $^{12}C_a{}^{1}H_{b-1}{}^{2}H_1{}^{14}N_c{}^{16}O_d{}^{32}S_e$ (for hydrogen), $^{12}C_a{}^{1}H_b{}^{14}N_{c-1}{}^{15}N_1{}^{16}O_d{}^{32}S_e$ (for nitrogen), $^{12}C_a{}^{1}H_b{}^{14}N_c{}^{16}O_{d-1}{}^{17}O_1{}^{32}S_e$ (for oxygen) and $^{12}C_a{}^{1}H_b{}^{14}N_c{}^{16}O_d{}^{32}S_{e-1}{}^{33}S_1$ (for sulfur). For biopolymers of <10 kD, the monoisotopic peak can easily be detected in mass spectral data with good signal-to-noise. However, as the molecular mass increases (>10 kD), the monoisotopic ion peak abundance decreases relative to the other heavy isotope enriched peaks of the isotope distribution and becomes undetectable [3,5]. The molecular mass must then be determined based on the m/z values of ions containing heavy isotopes. Therefore, it is absolutely necessary to correctly assign the isotopic composition to these multiply charged ions otherwise the molecular mass will be in error by several Daltons [4,13]. To achieve this, algorithms have been developed that align experimental isotopic distributions with those theoretically calculated [8,12,14]. These algorithms require a prior knowledge of the elemental composition of a molecule, and thus are only suited to known compounds. To overcome some of the limitations of this isotope fitting technique [14] and allow accurate molecular masses to be obtained in the case of unknown protein samples, the amino acid *averagine* [10] has been introduced for use in modeling isotopic distributions. The molecular formula of *averagine* $C_{4.9384}$ $H_{7.7583}$ $N_{1.3577}$ $O_{1.3577}$ $S_{0.0417}$ is based on the statistical occurrence of the 20 common amino acids within protein databases. Unknown proteins are assigned a particular number of averagine molecules required in order to generate a theoretical isotope distribution that "fit" the experimental distribution where the number of *averagine* units is corrected to round the number of each element to integer value [10]. Mass errors will occur when unknown protein samples do not follow this statistical occurrence, particularly in the case of proteins isolated from heavy isotope enriched environments. Furthermore, protein databases are biased toward known, characterized proteins that may differ markedly in sequence and composition from the many proteins that have yet to be characterized.

The natural abundance of the ^{13}C and ^{12}C isotopes used for modeling isotopic distributions are based on a terrestrial average from total mass of mineral carbon and the mass of carbon in living organisms that bias the $^{13}C/^{12}C$ isotope ratio toward mineral carbon. Even among proteins from different organisms, $^{13}C/^{12}C$ ratio strictly varies [15]. An advantage of the CRAM is that no prior knowledge of the elemental composition of a compound is required [9]. The CRAM can correlate isotope peaks of different multiply charged ions that share a common isotopic composition simply based on the ratios of m/z values of their isotope peaks. This feature of the CRAM is particularly useful in the case of low signal-to-noise mass spectral data, where isotope distribution pattern recognition techniques may fail due to an inability to fit an ideal theoretical isotopic distribution pattern to a poorly defined experimental pattern. In this article, the CRAM has been applied to high resolution FT-ICR mass spectral data for several proteins to illustrate its advantages in terms of obtaining both the charge of ions, and correlating those ions that share common isotopic compositions.

2. Experimental

Protein and peptide samples were obtained from Sigma Chemicals (St. Louis, MO, USA), and were used without further purification. The sample of transthyretin was provided from the Scripps Research Institute and prepared accordingly from published methods [16]. The protein solutions were prepared at a concentration range of 1–5 μM in 50:50 water and methanol or acetonitrile containing 2–4% acetic acid or 0.1% TFA, accordingly. For desalting of the transthyretin sample from a 400 mM sodium phosphate buffer, C_{18} Sep-Pak (Waters Corporation, MA, USA) was used, and the protein was eluted with a 50% acetonitrile/water solution containing 0.1% TFA. Electrospray mass spectra were recorded on a 4.7 T magnet (APEX, Bruker Daltonics, Billerica, MA, USA) mass spectrometer in the positive ion mode and spectra were processed with 512 k or 1 M data points. The theoretical isotopic distributions were produced from the isotopic distribution utility of the Xmass software (Bruker Daltonics, Billerica, MA, USA). Protein solutions were infused at a rate of between 3 and 5 μL/min and a needle voltage in the

range of 4.2–4.5 kV was used. Spectra were mass calibrated with the most abundant isotopic peak of angiotensin-1 (2+ and 3+ ions) and ubiquitin (7+ to 13+) as external or internal calibration ions.

3. Results and discussion

The charge ratio analysis method is a unique approach for the interpretation of electrospray mass spectral that identifies charge states for multiply charged ions simply from the ratio of their m/z values [9]. The theoretical basis of the CRAM has been reported earlier, and is briefly described here. The m/z values for two multiply charged ions a and b originating from the same compound with a molecular mass (M) are represented as $(R_z)_a$ and $(R_z)_b$, respectively. These multiply charged ions of charge (z) originate from the addition or abstraction of charge carrying species (m_A), where $R_z = (M \pm zm_A)/z$. The ratio of the m/z values for two ions a and b is then represented by Eq. (1):

$$\frac{(R_z)_a}{(R_z)_b} = \frac{z_b(M \pm z_a m_A)}{z_a(M \pm z_b m_A)} \quad (1)$$

The CRAM approach makes an assumption that for large biopolymers $M > z_a m_A$ or $z_b m_A$, and Eq. (1) then simplifies to Eq. (2):

$$\frac{(R_z)_a}{(R_z)_b} = \frac{z_b}{z_a} \quad (2)$$

For relatively small molecules the assumption of $M > z_a m_A$ or $z_b m_A$ also works well since small molecules support fewer charge carrying species (i.e., $zm_A < M$) as compared to large biopolymers. Therefore, from the ratio of m/z values of any two multiply charged ions, the inverse ratio of their two ion charges is calculated. The unique property of the ratio of two integers is the basis of the CRAM where by correlating the ratio of the m/z values of two multiply charged ions to the unique ratio of two integers, the charge states of ions are identified without a prior knowledge or assumption of the nature of the charge carrying species. In contrast, most deconvolution [17] algorithms require information about the charge carrying species and consider a uniform charge carrying species, while CRAM does not assume that charge carrying species is uniform for all ions. The limitation of the CRAM approach is that at least two ions are required for the determination of the ion charge. Therefore, the CRAM is not be applicable to compounds where only one ion is detected in their electrospray mass spectra.

3.1. Determination of ion charge by the CRAM

High resolution FT-ICR mass spectral data for several proteins are selected to illustrate the utility of the CRAM. Table 1 represents mass spectral data (512 k data points) for five charge states of ubiquitin with an average resolution of 30 k that easily affords the isotopic separation of each charge state. The code listed in column 1 of Table 1 designates the isotope peaks for each charge state (i.e., A1 to A10 represent isotope peaks of the first charge state, and B1 to B13 represent isotope peaks of the next charge state in the series, etc.). For the application of the CRAM, the ratios of m/z values for isotope peaks of two different charge states are calculated as listed in columns 4 and 5 of Table 1. The m/z value of *one* of the isotope peaks of one charge state is divided consecutively by the m/z values of *every* isotope peak of the next charge state (column 4). This process is continued until the ratio of two m/z values is closest to a unique ratio of two integers (value shown in bold in column 4) [9]. This identifies the charge state of each set of ions. For example, from the ratios of A1/B1 to A1/B5, the ratio of A1/B3 shown in bold (0.923078) is in closest agreement with the ratio of integers 12/13 (0.923077). Therefore, the charge states for ions designated as A and B are 13 and 12, respectively. Once two isotope peaks have been correlated (i.e., A1 and B3), the ratios of other isotopes peaks (i.e., A2/B4 to A9/B11) are calculated as illustrated in column 5 of Table 1 in order to verify the accuracy of the charge state assignment. The CRAM further identifies charge states of +12 to 10 for the ion series designated as B, C and D.

3.2. Comparison of the CRAM and the isotope spacing method for determining the charge of ions in high resolution FT-ICR mass spectra

For high resolution electrospray mass spectra where multiply charged ions are resolved to their isotope peaks, the ion charge can be calculated from the reciprocal of the separation between adjacent isotopic peaks ($1/\Delta m/z$) [8]. For example, for a multiply charged ion with an isotope spacing of 0.1 u, the ion charge is 10. While this charge spacing method is mostly successful in assigning ion charge, it has limited success for ions with more than 20 charges. Here, a low ppm mass precision is required which can only be achieved on high resolution instruments with the use of internal standards that may suppress or mask the sample ion signal [3,13]. Moreover, the accuracy of the isotope spacing method is compromised when interpreting mass spectral data with low signal-to-noise and in the case of overlapping peaks from the analysis of mixtures. The CRAM approach is shown to improve the accuracy of charge state determination by several orders of magnitude [9].

For the high resolution FT-ICR mass spectral data of ubiquitin (Table 1), the m/z spacing between each isotope peak is calculated for successive isotope peaks as listed in column 7. The error based on the isotopic spacing method (column 8) is calculated from the $\Delta m/z$ of isotope peaks and the inverse ratio of their theoretical charge states ($1/z$). For the CRAM approach, errors (column 6) are calculated as the difference in the ratios of m/z values across two sets of isotope peaks (column 5) and the ratio of two integers corresponding to their charge. The average error calculated for the four multiply charged ions of ubiquitin based on

Table 1
Identification of the ion charge by the CRAM for high resolution electrospray FT-ICR mass spectra of protein ubiquitin (bovine)

	$(R_z)_n$	Intensity (%)[a]	$(R_z)_n/(R_z)_{n+1}$	$(R_z)_n/(R_z)_{n+1}$	Error[b] $\times 10^{-6}$	Isotopic spacing $\Delta(R_z)_n$	Error[c] $\times 10^{-3}$
			A1/B1 to A1/Bn	A1/B3 to A9/B11			
A1	659.594885	7.0	0.923291	0.923078	0.80	0.076649	0.27
A2	659.671534	16.2	0.923185	0.923078	0.94	0.076588	0.34
A3	659.748122	27.3	**0.923078**	0.923078	1.10	0.076972	0.05
A4	659.825094	34.5	0.922971	0.923078	1.37	0.076521	0.40
A5	659.901615	33.0	0.922864	0.923077	0.52	0.077185	0.26
A6	659.978800	27.2		0.923078	0.72	0.077134	0.21
A7	660.055934	19.0		0.923078	0.70	0.076760	0.16
A8	660.132694	12.3		0.923075	1.53	0.075688	1.24
A9	660.208382	7.7		0.923074	3.18	0.079033	2.11
A10	660.287415	4.5			Average 1.21		Average 0.56
			B1/C1 to B1/Cn	B1/C2 to B12/C13			
B1	714.395064	5.6	0.916772	0.916668	0.97	0.082550	0.78
B2	714.477614	21.7	**0.916668**	0.916667	0.14	0.082892	0.44
B3	714.560506	51.4	0.916561	0.916668	0.99	0.082930	0.40
B4	714.643436	81.7		0.916667	0.10	0.082846	0.49
B5	714.726282	98.3		0.916666	0.56	0.083176	0.16
B6	714.809458	93.3		0.916666	0.71	0.083553	0.22
B7	714.893011	76.0		0.916666	0.71	0.083461	0.13
B8	714.976472	52.8		0.916666	1.11	0.083579	0.25
B9	715.060051	35.9		0.916665	1.83	0.084888	1.55
B10	715.144939	20.7		0.916666	1.02	0.083274	0.06
B11	715.228213	10.4		0.916667	0.05	0.083087	0.25
B12	715.311300	6.5		0.916667	0.53	0.085516	2.18
B13	715.396816	3.9			Average 0.73		Average 0.57
			C1/D1 to C1/Dn	C1/D2 to C11/D12			
C1	779.250699	4.0	0.909197	0.909095	4.39	0.088550	2.36
C2	779.339249	19.7	**0.909095**	0.909092	1.38	0.090753	0.16
C3	779.430002	50.4	0.908989	0.909093	1.69	0.089711	1.20
C4	779.519713	84.0		0.909092	1.13	0.091221	0.31
C5	779.610934	100.0		0.909092	1.40	0.090942	0.03
C6	779.701876	96.6		0.909092	1.14	0.090863	0.05
C7	779.792739	85.4		0.909091	0.28	0.091149	0.24
C8	779.883888	61.4		0.909091	0.50	0.091389	0.48
C9	779.975277	39.5		0.909091	0.26	0.091791	0.88
C10	780.067068	24.9		0.909091	0.28	0.091919	1.01
C11	780.158987	11.8		0.909094	3.49	0.089930	0.98
C12	780.248917	6.8			Average 1.45		Average 0.69
C13	780.339150	3.1					
			D1/E1 to D1/En	D1/E1 to D10/E10			
D1	857.075855	2.7	**0.900005**	0.900005	4.92	0.095776	4.22
D2	857.171631	11.9	0.899902	0.900002	2.15	0.100238	0.24
D3	857.271869	27.7	0.899795	0.900001	1.20	0.099536	0.46
D4	857.371405	44.9		0.900002	2.06	0.099209	0.79
D5	857.470614	53.9		0.900001	0.53	0.100095	0.10
D6	857.570709	52.2		0.900002	1.51	0.100279	0.28
D7	857.670988	46.3		0.900000	0.26	0.100763	0.76
D8	857.771751	34.2		0.900001	1.00	0.100052	0.05
D9	857.871803	22.9		0.900001	0.96	0.101250	1.25
D10	857.973053	12.9		0.900005	5.04	0.100461	0.46
D11	858.073514	6.4			Average 1.96		Average 0.86
D12	858.171589	2.5					

E-series not shown.

[a] Abundance > 2%.

[b] Error = $|(R_z)_n/(R_z)_{n+1} - (z_n/z_{n+1})|$.

[c] Error = |isotopic spacing $(\Delta(R_z)_n - 1/z_n|$.

the isotopic spacing method is 0.7×10^{-3} while the average error based on the CRAM is 1.3×10^{-6}. In this example, the CRAM has identified the charge states with 500-fold greater accuracy.

In order to illustrate the utility of CRAM for interpreting high resolution mass spectra, data for several additional proteins are presented in Table 2. This set of proteins includes cytochrome *c*, transthyretin, lysozyme and calmodulin with molecular masses ranging from 12 to 16 kDa. The multiply charged ions for these proteins with different isotopic compositions can easily be resolved on a 4.7 T magnet FT-ICR instrument. Three most intense multiply charged ions for each protein are labeled in Table 2 as A, B and C. The A1 to A11 refer to *m*/*z* values for the first multiply charged ion of cytochrome *c* (column 2), and A1 to A13 refers to *m*/*z* values for the first multiply charged ion of lysozyme listed in column 8.

When applying the isotope spacing method, the difference in *m*/*z* values for isotope peaks of each multiply charged ion is calculated based on consecutive *m*/*z* values for each protein as listed in columns 3, 6, 9 and 12. For the CRAM, the charge states are correlated as described above for ubiquitin (Table 1) by calculating the ratios of *m*/*z* values for isotope peaks of two different charge states and finding the ratios of *m*/*z* values that are in closet agreement with the ratio of two integers [9]. The *m*/*z* ratios for the four proteins are listed in columns 4, 7, 10 and 13 of Table 2. The calculated ratios for cytochrome *c* (column 4) as A1/B3 to A10/B12, B1/C1 to B11/C10 and C1/D2 to C9/D10 identify charge states as 17, 16 and 15.

While the peaks in Table 2 are correlated in a consecutive manner, the CRAM is not limited to only consecutive ions [9]. Since the ratios of two non-consecutive integers are also unique, the CRAM approach can be used to analyze non-consecutive ions directly. For transthyretin (column 7 of Table 2), the ratios of A1 and B2, and B2 and C3 are in closest agreement, respectively, with the ratios of 17/18 and 16/17, which show that A1 and C3 correlate. The ratio of *m*/*z* values for the A1 isotope (18+) and the C3 isotope (16+) is 0.888882, which correlates these ions directly with an accuracy of 7×10^{-6} based on the ratio of integers 16 and 18.

Table 3 shows the charge states for multiply charged ions of all proteins in this study with a comparison of the error associated with the isotope spacing method and the CRAM. The results show that the ratios of *m*/*z* values calculated from isotope peaks of different multiply charged ions are identical to the ratios of integers to the fifth decimal for ubiquitin (Table 1, column 5) and to the fourth decimal (Table 2, columns 4, 7, 10 and 13) for the other four proteins. As a result, the CRAM identifies charge states with errors in the range of 10^{-6} (Table 3, column 6). By comparison, the values for $\Delta m/z$ calculated in the isotope spacing method diverge at the third decimal for ubiquitin (Table 1, column 7) and the second decimal (Table 2, columns 3, 6, 9 and 12) for the other four proteins. The errors in the isotope spacing method are determined to be in the range of 10^{-3} (Table 3, column 5). As summarized in Table 3, the charge states calculated by the CRAM approach are of the order of 100–1000-fold more accurate. The high accuracy of the CRAM approach in determining ion charge (Table 3) demonstrates that the CRAM approach would also be valuable for the analysis of protein mixtures.

The CRAM approach is superior to the isotope spacing method in assigning charge states to multiply charged ions regardless of whether the mass spectra are internally or externally mass calibrated. Mass spectral data recorded for protein ubiquitin were internally calibrated while the spectra for the other proteins were externally calibrated. Two sets of data for protein ubiquitin were acquired following the accumulation of 512 k and 1 M data points (Table 3). Spectra recorded with 1 M data points afford superior peak shapes and resolution and consequently improve the accuracy of *m*/*z* values by resolving protein signals from other chemical noise. This higher accuracy in *m*/*z* measurements would benefit both the isotope spacing method and the CRAM. As is evident in Table 3, the CRAM approach is superior at identifying the charge states of ions regardless of the data quality.

3.3. Identification of the charge carrying species

The mass (m_A) and subsequent identity of the charge carrying species is determined by the CRAM as reported earlier according to Eq. (3):

$$((R_z)_a z_a - (R_z)_b z_b) = \pm m_A(z_a - z_b) \quad (3)$$

In the case of resolved isotope peaks, the $(R_z)_a$ and $(R_z)_b$ values for two different multiply charged ions in Eq. (3) must have the same elemental composition, otherwise the mass of charge carrying species (m_A) will differ from its true value. If $(R_z)_a$ and $(R_z)_b$ represent *m*/*z* values with a different number of the heavy isotopes, the value for m_A will also reflect this difference.

Where mass spectra of good signal-to-noise are recorded for relatively low molecular weight proteins or other biopolymers (up to 10 kD), monoisotopic ions containing only the lightest isotopes for each element (e.g., ^{12}C) are detected. Thus, isotope peaks of different multiply charged ions of the same elemental composition can easily be correlated based on these monoisotopic peaks with values for $(R_z)_a$ and $(R_z)_b$ assigned. As either the biopolymer becomes larger (>10 kD) and/or the signal-to-noise mass spectra deteriorates, the monoisotopic ions [3,5] are not detected. Values for $(R_z)_a$ and $(R_z)_b$ can then be assigned based on the intensity weighted average *m*/*z* value of all isotopic peaks corresponding to the two multiply charged ions. In this process, the high resolution mass spectral data are converted to low resolution data solely for the purpose of calculating the mass of the charge carrying species.

In the case of the protein ubiquitin (data shown in Table 1), the intensity weighted average *m*/*z* values of the

Table 2
Application of the CRAM to high resolution FT-ICR mass spectra for a set of diverse proteins

	Cytochrome *c*			Transthyretin			Lysozyme			Calmodulin		
	$(R_z)_n$	$\Delta(R_z)_n$	A1/B3 to A10/B12	$(R_z)_n$	$\Delta(R_z)_n$	A1/B2 to A9/B10	$(R_z)_n$	$\Delta(R_z)_n$	A1/B1 to A12/B12	$(R_z)_n$	$\Delta(R_z)_n$	A1/B3 to A9/B11
A1	727.754594	0.054531	0.941178	772.599554	0.055569	0.944442	1192.664104	0.084899	0.916659	1292.338148	0.075857	0.923073
A2	727.809125	0.059133	0.941175	772.655123	0.054954	0.944442	1192.749003	0.085631	0.916660	1292.414005	0.077268	0.923074
A3	727.868258	0.058265	0.941176	772.710077	0.056138	0.944441	1192.834634	0.080977	0.916662	1292.491273	0.073710	0.923075
A4	727.926523	0.056744	0.941177	772.766215	0.055920	0.944441	1192.915611	0.084256	0.916661	1292.564983	0.079755	0.923072
A5	727.983267	0.060184	0.941175	772.822135	0.055537	0.944441	1192.999867	0.084203	0.916661	1292.644738	0.078210	0.923074
A6	728.043451	0.059552	0.941178	772.877672	0.055450	0.944442	1193.084070	0.083234	0.916661	1292.722948	0.072045	0.923075
A7	728.103003	0.058255	0.941181	772.933122	0.055375	0.944441	1193.167304	0.083661	0.916661	1292.794993	0.075367	0.923069
A8	728.161258	0.058481	0.941178	772.988497	0.057330	0.944440	1193.250965	0.083121	0.916661	1292.870360	0.072540	0.923065
A9	728.219739	0.060308	0.941177	773.045827	0.053806	0.944445	1193.334086	0.082486	0.916660	1292.942900	0.093038	0.923062
A10	728.280047	0.058057	0.941180	773.099633			1193.416572	0.084261	0.916659	1293.035938		
A11	728.338104						1193.500833	0.084424	0.916659			
A12							1193.585257	0.082215	0.916661			
A13							1193.667472					
			B1/C1 to B11/C10			B1/C2 to B9/C10			B1/C2 to B9/C10			B1/C2 to B11/C12
B1	773.114083	0.060962	0.937503	817.991353	0.057530	0.941175	1301.099045	0.090498	0.909083	1399.871469	0.081037	0.916656
B2	773.175045	0.062805	0.937501	818.048883	0.058585	0.941171	1301.189543	0.091507	0.909084	1399.952506	0.086747	0.916658
B3	773.237850	0.060200	0.937503	818.107468	0.058847	0.941173	1301.281050	0.089361	0.909083	1400.039253	0.079723	0.916664
B4	773.298050	0.062237	0.937502	818.166315	0.059355	0.941172	1301.370411	0.092543	0.909083	1400.118976	0.083008	0.916657
B5	773.360287	0.060997	0.937501	818.225670	0.059296	0.941172	1301.462954	0.091174	0.909084	1400.201984	0.084442	0.916657
B6	773.421284	0.061922	0.937500	818.284966	0.058234	0.941173	1301.554128	0.091326	0.909082	1400.286426	0.082796	0.916664
B7	773.483206	0.062050	0.937500	818.343200	0.059146	0.941172	1301.645454	0.090570	0.909083	1400.369222	0.083965	0.916658
B8	773.545256	0.060623	0.937500	818.402346	0.059649	0.941167	1301.736024	0.092109	0.909080	1400.453187	0.086182	0.916656
B9	773.605879	0.064495	0.937495	818.461995	0.056983	0.941174	1301.828133	0.092338	0.909081	1400.539369	0.088178	0.916663
B10	773.670374	0.062338	0.937496	818.518978			1301.920471			1400.627547	0.082742	0.916666
B11	773.732712	0.062344	0.937503				1302.012342			1400.710289	0.076736	0.916660
B12	773.795056						1302.101617			1400.787025		
			C1/D2 to C9/D10[a]			C1/D1 to C8/D8[a]			C1/D2 to C10/D11[a]			C1/D2 to C11/D12[a]
C1	824.652201	0.066974	0.933330	869.058481	0.058442	0.937496	1431.117270	0.103689	0.899989	1527.065026	0.085291	0.909092
C2	824.719175	0.065100	0.933332	869.116923	0.064669	0.937491	1431.220959	0.098825	0.899989	1527.150317	0.084471	0.909090
C3	824.784275	0.065479	0.933331	869.181592	0.060868	0.937498	1431.319784	0.101664	0.899990	1527.234788	0.085394	0.909085
C4	824.849754	0.066829	0.933329	869.242460	0.063436	0.937495	1431.421448	0.098657	0.899990	1527.320182	0.098368	0.909081
C5	824.916583	0.065888	0.933331	869.305896	0.062908	0.937497	1431.520105	0.100316	0.899990	1527.418550	0.090847	0.090847
C6	824.982471	0.065912	0.933329	869.368804	0.062004	0.937498	1431.620421	0.102430	0.899990	1527.509397	0.080347	0.909082
C7	825.048383	0.066679	0.933328	869.430808	0.063128	0.937496	1431.722851	0.099817	0.899990	1527.589744	0.100554	0.909074
C8	825.115062	0.068684	0.933327	869.493936	0.067020	0.937492	1431.822668	0.103419	0.899990	1527.690298	0.094563	0.909080
C9	825.183746	0.068489	0.933332	869.560956			1431.926087	0.100555	0.899992	1527.784861	0.082065	0.909082
C10	825.252235			869.617781			1432.026642	0.097298	0.899992	1527.866926	0.091526	0.909074
C11							1432.123940			1527.958452	0.100311	0.909080
C12										1528.058763		
C13										1528.140164		

Source of proteins (cytochrome *c* (horse heart), transthyretin (human wild-type), lysozyme (chicken egg white), calmodulin (bovine brain)).

[a] $(R_z)n$ values for the D-series not shown.

Table 3
Improvements in the determination of the ion charge by the CRAM vs. the isotope spacing method for a set of proteins from mass spectra of different data quality

Protein	Ion charge, z	Data points	Resolution × 10^3	$\lvert(\Delta R_z)_n - 1/z_n\rvert$[a] × 10^{-3}	$\lvert(R_z)_n/(R_z)_{n+1} - (z_n/z_{n+1})\rvert$[b] × 10^{-6}	Improvements in calculating ion charge
Ubiquitin[c]	13	512 k	36	0.56	1.21	460
	12		32	0.57	0.73	780
	11		29	0.70	1.45	480
	10		27	0.86	1.96	430
Ubiquitin[c]	13	1 M	69	0.43	1.99	215
	12		62	0.65	0.82	810
	11		58	0.47	0.40	1170
	10		54	0.53	0.48	1100
Cytochrome *c*	17	512 k	38	1.25	1.56	800
	16		47	1.01	2.01	500
	15		45	0.97	3.49	270
Transthyretin	18	1 M	113	0.60	2.67	220
	17		108	0.68	4.20	160
	16		91	2.28	4.14	550
Lysozyme	12	512 k	35	1.05	6.32	160
	11		32	0.87	8.26	105
	10		27	1.75	9.66	180
Calmodulin	13	1 M	46	3.96	5.85	670
	12		39	2.44	6.76	360
	11		36	6.16	8.80	700

Error (average for all isotope peaks of one ion charge from Table 1 and 2).
[a] The charge spacing method.
[b] The CRAM.
[c] Data for ubiquitin are based on internal calibration and external calibration has been applied to data for the other proteins.

five multiply charged ions (13+ to 9+) are calculated to be 659.897670, 714.806993, 779.706679, 857.569482 and 952.742971, respectively. The mass of the charge carrying species (m_A) using two of these consecutive m/z values is calculated from Eq. (3) to be 0.985792, 0.910450, 1.07864 and 1.008074, respectively. The average value for m_A is thus calculated to be 0.995741 identifying the charge carrying species to be a proton.

The elemental compositions of different multiply charged ions vary according to the addition of an extra charge carrying species. For example, the 13+ and 12+ protonated ions of a protein have elemental compositions that vary by one proton while the 13+ and 11+ ions vary by two protons. Since the contribution in mass of the added charge carrying species is not significant relative to the total mass of a protein, the CRAM can still identify the charge of its ions using Eq. (2) with a high degree of accuracy. In order to correlate two isotopic peaks across two different multiply charged ion distributions of charge z_a and z_b where $z_a > z_b$ and $a = b + n$, the m/z value for $(R_z)_a$ in Eqs. (1) to (3) are corrected for the additional mass of the charge carrying species by substituting it with $(R_z)_a - n(m_A/z_a)$. Under these circumstances, the correct representation of Eq. (2) is shown in Eq. (4).

$$\frac{((R_z)_a - n(m_A/z_a))}{(R_z)_b} = \frac{z_b}{z_a} \quad (4)$$

The value for n represents the difference in charge of the multiply charged ions that are being correlated ($n = a - b$). For ions of charge 13+ and 11+, the value for n is 2.

The process by which the CRAM procedure correlates ions of different isotopic distributions that share a common elemental composition is defined as follows. The charges for any pair of ions (a and b) are first identified as described above for the ubiquitin data of Table 1 using Eq. (2). The mass of the charge carrying species m_A is then calculated using Eq. (3). Ions sharing a common elemental composition are then correlated by applying Eq. (4) in which m_A is given its theoretical value (i.e., for a proton $m_A = 1.007825$). This process is achieved by subtracting the value for $n(m_A/z_n)$ from all m/z values for $(R_z)_a$. The ratios for $((R_z)_a - n(m_A/z_a))/(R_z)_b$ are then calculated and compared with the ratio of two integers by applying the CRAM [9]. Those closest in value to a ratio of two integers represent m/z values for ions of protein molecules that share a common isotope composition.

Table 4 shows the 13+ and 11+ ions of two different sized proteins ubiquitin (8.5 kD) and calmodulin (16.8 kD). The isotope peaks of these multiply charged ions were correlated by applying Eq. (4) and utilizing the m/z values of Table 1 (column 2) for ubiquitin and of Table 2 (column 11) for calmodulin. Alignment of isotopes were accomplished by

Table 4
Correlation of isotope peaks of different multiply charged ions by the CRAM for proteins ubiquitin and calmodulin

	$(R_z)_a$	Intensity (%)	$(R_z)_a - n(m_A/z_a)$		$(R_z)_b$	Intensity (%)	$[(R_z)_a - n(m_A/z_a)]/(R_z)_b$	Error[a] $\times 10^{-6}$
Ubiquitin (13+)				Ubiquitin (11+)				
				C1	779.250699	4.0		
A1	659.594885	7.0	659.439835	C2	779.339249	19.7	0.846152	1.37
A2	659.671534	16.2	659.516484	C3	779.430002	50.4	0.846152	1.55
A3	659.748122	27.3	659.593072	C4	779.519713	84.0	0.846153	0.68
A4	659.825094	34.5	659.670044	C5	779.610934	100.0	0.846153	0.96
A5	659.901615	33.0	659.746565	C6	779.701876	96.6	0.846152	1.51
A6	659.978800	27.2	659.823750	C7	779.792739	85.4	0.846153	1.12
A7	660.055934	19.0	659.900884	C8	779.883888	61.4	0.846153	1.11
A8	660.132694	12.3	659.977644	C9	779.975277	39.5	0.846152	1.84
A9	660.208382	7.7	660.053332	C10	780.067068	24.9	0.846149	4.38
A10	660.287415	4.5	660.132365	C11	780.158987	11.8	0.846151	2.77
				C12	780.248917	6.8		
				C13	780.339150	3.1		
Calmodulin (13+)				Calmodulin (11+)				
				C1	1527.065026	25.2		
A1	1292.338148	24.8	1292.183098	C2	1527.150317	46.0	0.846140	13.76
A2	1292.414005	29.9	1292.258955	C3	1527.234788	70.2	0.846143	10.89
A3	1292.491273	40.1	1292.336223	C4	1527.320182	90.6	0.846146	7.61
A4	1292.564983	50.1	1292.409933	C5	1527.418550	100.0	0.846140	13.85
A5	1292.644738	48.2	1292.489688	C6	1527.509397	99.8	0.846142	11.96
A6	1292.722948	45.4	1292.567898	C7	1527.589744	90.8	0.846149	5.26
A7	1292.794993	34.2	1292.639943	C8	1527.690298	79.8	0.846140	13.80
A8	1292.870360	28.5	1292.715310	C9	1527.784861	59.8	0.846137	16.84
A9	1292.942900	22.7	1292.787850	C10	1527.866926	51.1	0.846139	14.81
A10	1293.035938	19.7	1292.880888	C11	1527.958452	42.2	0.846149	4.60
				C12	1528.058763	24.3		
				C13	1528.140164	20.2		

[a] Error = $|[(R_z)_a - n(m_A/z_a)]/(R_z)_b - (11/13)|$.

adjusting the 13+ and 11+ ions to a common elemental composition with a correction factor of $(2 \times 1.007825)/13$. The ratios shown in column 8 are calculated from Eq. (4) and are the closest values to the ratio of 11/13. For ubiquitin and calmodulin, the A1 to A10 isotopes of the 13+ correlate with the C2 to C11 isotopes of the 11+.

The isotope peaks of two different multiply charged ions could alternatively be correlated by predicting the identity of the charge carrying species, and utilizing the Eq. (3) while solving for a pre-assigned mass of the charge carrier (m_A). The m/z values for two isotope peaks (i.e., $(R_z)_a$ and $(R_z)_b$) corresponding to two different multiply charged ions that result in a pre-assigned mass of the charge carrier subsequently identify the isotope peaks that correlate. For example, from Table 1, the difference in the product of the m/z values for the first (A1) and the second (B2) isotope peaks and their charges is in closest match to the mass of a proton (i.e., $A1 \times 13 - B2 \times 12 = 1.002137$). This indicates that the A1 and B2 isotope peaks correlate and share a common isotopic composition. In contrast, the use of m/z values for A1 and B3 isotope peaks would result in a mass of 0.007433 for the charge carrying species and suggest that these isotope peaks do not correlate. In this case, contributions from extra heavy isotopes shift the m/z value for the B3 peak and cause the mass of the charge carrying species to deviate from its theoretical value.

3.4. Utility of the CRAM in addressing challenges with the matching of theoretical and experimental resolved isotopic profiles

Partial mass spectra for the two proteins ubiquitin (13+ and 11+ ions) and calmodulin (11+ and 13+ ions) and their theoretical isotope profiles are shown in Fig. 1. All multiply charged ions in Fig. 1 are displayed in a 1.6 u window. It is important to note that theoretical isotopic distributions of representative multiply charged ions for these two different proteins are very similar and show bell-shape (i.e., semi-Gaussian) [11,12,14] profiles. The theoretical profiles of 13+ and 11+ (Fig. 1) are symmetrical, while the experimental isotope peak intensities are unsymmetrical and vary greatly across the peak profiles. These features will affect the accurate matching of the theoretical and experimental distributions. For ubiquitin and calmodulin data of Fig. 1, the CRAM (Table 4) has established a correlation for these multiply charged ions (i.e., A1 isotope peak of the 13+ correlates with the C2 isotope of the 11+ ions), whereas correlation of the theoretical profiles to experimental data would not easily be achieved based on the relative intensity of isotope peaks.

Theoretical profiles are calculated by expansion of a polynomial expression. For a molecule containing carbon, hydrogen, nitrogen, oxygen and sulfur (i.e., with the elemental composition $C_aH_bN_cO_dS_e$), the distribution is

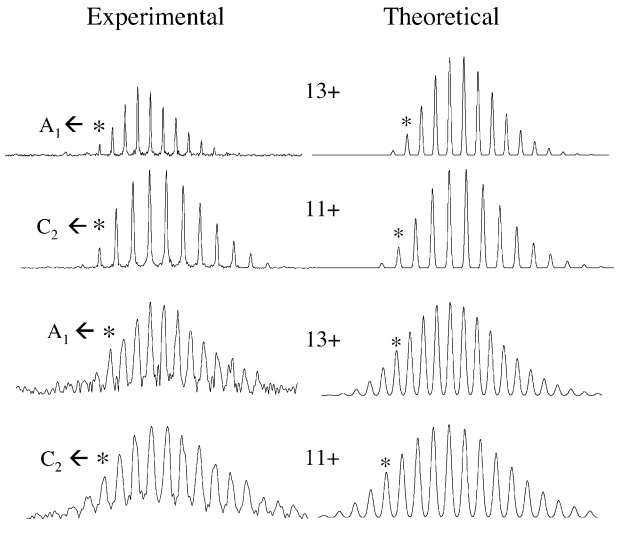

Fig. 1. Partial mass spectra for multiply charged ions (13+ and 11+) of ubiquitin and calmodulin with their corresponding theoretical isotopic distributions (asterisks correspond to isotopes peaks of Table 3).

calculated from the product of polynomials associated with the relative abundances of each isotope (i.e., $(^{12}C + ^{13}C)^a(^1H + ^2H)^b(^{14}N + ^{15}N)^c(^{16}O + ^{17}O + ^{18}O)^d(^{32}S + ^{33}S + ^{34}S + ^{36}S)^e)^{12}$. As the number of atoms increases with the size of proteins, the isotopic profiles expand on the m/z axis and the intensities of isotope peaks become more uniform toward the center of the distribution. The effect of isotopic peak spread can be seen by comparison of the isotopic profiles of ubiquitin and calmodulin (Fig. 1). For large molecules, correlation of theoretical isotopic distributions to experimental data based on relative intensities becomes more challenging particularly for spectra with low signal-to-noise. Thus, it is important to develop new approaches to interpret the isotopic distribution patterns in order to accurately determine molecular mass of compounds. The utility of the CRAM approach is that isotope peaks of different multiply charged ions can be correlated from the ratios of their m/z values and not from their relative ion intensities.

This feature of the CRAM in correlating isotope peaks of different multiply charged ions is also useful for the interpretation of mass spectral data of complex mixtures when ion signals overlap and for the analysis of samples whose elemental compositions are not known. When combined with other isotope pattern matching techniques [14], the CRAM approach would result in a more accurate molecular weight determination of biopolymers. For the analysis of samples with unknown elemental compositions, once a theoretical elemental composition has been estimated [10] that matches the m/z values and ion abundances of one of the charge states, the application of the CRAM would ensure that isotopic profiles of all charge states correlate for that particular elemental composition, and accordingly, adjustments in the estimated elemental composition [10] could be made so that isotope peaks of all charge states would correlate according to the CRAM procedure.

4. Conclusions

The charge ratio analysis method has been applied to high resolution FT-ICR mass spectral of several proteins with superior results for charge determination in comparison to the isotope spacing method. We have demonstrated that the CRAM provides more accurate charge state determination on the order of 100–1000-fold in comparison to the isotope spacing method. For accurate molecular weight determination of biopolymers by electrospray mass spectrometry, both the determination of the ion charge and the correct assignment of isotopic composition of peaks are required. The CRAM also correlates multiply charged ions according to their isotopic compositions. This feature of the CRAM is especially useful when combined with other isotope pattern matching protocols for the analysis of biopolymers with unknown elemental compositions.

Acknowledgements

The FT-ICR mass spectrometer has been purchased with funds provided by the Australian Research Council through the Linkage Infrastructure, Equipment and Facilities grant (ARC LE02/37908) and Griffith University. Author Maleknia thanks Joel Buxbaum and Natalia Reixach of the Scripps Research Institute for providing samples of protein transthyretin.

References

[1] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, G.M. Whitehouse, Science 246 (1989) 64.
[2] A.G. Marshall, Acc. Chem. Res. 18 (1985) 316.
[3] F.W. McLafferty, Acc. Chem. Res. 27 (1994) 379.
[4] M.K. Green, M.M. Vestling, M.V. Johnston, B.S. Larson, Anal. Biochem. 260 (1998) 204.
[5] N.L. Kelleher, M.W. Senko, M.M. Siegel, F.W. McLafferty, J. Am. Soc. Mass Spectrom. 8 (1997) 380.
[6] A.G. Marshall, M.W. Senko, W. Li, M. Li, S. Dillon, S. Guan, T.M. Logan, J. Am. Chem. Soc. 119 (1997) 433.
[7] B.E. Winger, S.A. Hofstadler, J.E. Bruce, H.R. Udesth, R.D. Smith, J. Am. Soc. Mass Spectrom. 4 (1993) 566.
[8] M.W. Senko, S.C. Beu, F.W. McLafferty, J. Am. Soc. Mass Spectrom. 6 (1995) 52.
[9] S.D. Maleknia, K.M. Downard, Anal. Chem. 77 (2005) 111.
[10] M.W. Senko, S.C. Beu, F.W. McLafferty, J. Am. Soc. Mass Spectrom. 6 (1995) 229.
[11] J. Yergey, D. Heller, G. Hansen, R.J. Cotter, C. Fenselau, Anal. Chem. 55 (1983) 353.
[12] J.A. Yergy, Int. J. Mass Spectrom. Ion Phys. 52 (1983) 337.
[13] S.C. Beu, M.W. Senko, J.P. Quinn, F.W. McLafferty, J. Am. Soc. Mass Spectrom. 4 (1993) 190.
[14] D.M. Horn, R.A. Zubarev, F.W. McLafferty, J. Am. Soc. Mass Spectrom. 11 (2000) 320.
[15] R.C. Beavis, Anal. Chem. 65 (1993) 496.
[16] H.A. Lashuel, C. Wurth, L. Woo, J.W. Kelly, Biochemistry 38 (1999) 13560.
[17] M. Mann, C.K. Meng, J.B. Fenn, Anal. Chem. 61 (1989) 1702.