JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Subscriber access provided by ISTANBUL TEKNIK UNIV

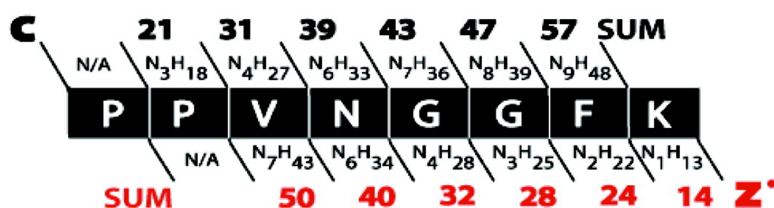
Article

Valence Parity Renders *z*-Type Ions Chemically Distinct

Shane L. Hubler, April Jue, Jason Keith, Graeme C. McAlister, Gheorghe Craciun, and Joshua J. Coon

J. Am. Chem. Soc., **2008**, 130 (20), 6388-6394 • DOI: 10.1021/ja7099985 • Publication Date (Web): 29 April 2008

Downloaded from http://pubs.acs.org on February 8, 2009



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 1 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

ACS Publications
High quality. High impact.

Journal of the American Chemical Society is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

J|A|C|S
ARTICLES

Published on Web 04/29/2008

Valence Parity Renders $z^{\bullet}$-Type Ions Chemically Distinct

Shane L. Hubler,[†] April Jue,[‡] Jason Keith,[‡] Graeme C. McAlister,[‡] Gheorghe Craciun,[*,†,§] and Joshua J. Coon[*,‡,§]

Departments of Mathematics, Chemistry, and Biomolecular Chemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706

Received November 5, 2007; E-mail: craciun@math.wisc.edu; jcoon@chem.wisc.edu

Abstract: Here we report that the odd electron $z^{\bullet}$-type ions formed by the electron-based peptide dissociation methods (electron capture or transfer, ECD or ETD) have distinctive chemical compositions from other common product ion types. Specifically, *b*-, *c*-, and *y*-type ions have an *odd* number of atoms with an odd valence (e.g., N and H), while $z^{\bullet}$-type ions contain an *even* number of atoms with an odd valence. This tenet, referred to as the valence parity rule, mandates that no *c*-type ion shall have the same chemical composition, and by extension mass, as a $z^{\bullet}$-type ion. By experiment we demonstrate that nearly half of all observed *c*- and $z^{\bullet}$-type product ions resulting from 226 ETD product ion spectra can be assigned to a single, correct, chemical composition and ion type by simple inspection of the *m*/*z* peaks. The assignments provide (1) a platform to directly determine amino acid composition, (2) an input for database search algorithms, or (3) a basis for *de novo* sequence analysis.

Introduction

For over 50 years accurate mass measurements have provided a means to assign chemical compositions to confirm structural assignments of synthetic or natural products.[1,2] During this time mass spectrometric (MS) instrumentation has advanced greatly so that today such measurements are among the expected rigors of small molecule validation. The field of protein sequence analysis—proteomics—has likewise been accelerated by MS-based technologies, though derivation of peptide chemical compositions from mass alone is, generally, not possible. Zubarev, Marshall, and others calculate that with 1 ppm mass accuracy a unique chemical composition can only be determined for peptides having masses less than ~750 Da; 550 Da is the approximate limit for determining unique amino acid compositions; beyond this mass the number of isomeric possibilities increases rapidly.[3,4]

Tandem mass spectrometry (MS/MS) obviates this problem by providing a subset of sequence-specific product ions following dissociation of a selected precursor.[5,6] Collision-activated dissociation (CAD) is the most commonly employed fragmentation method and generates *b*- and *y*-type product ions (fragments carrying either the peptide's N or C-terminus, respectively).[7] Direct sequence derivation (*de novo*) is accomplished by first deciphering which *m*/*z* peaks are *b*-type and which are *y*-type and then reading the sequence either forward (*b*-type series) or in reverse (*y*-type series).[8] In practice, however, such assignments are difficult to make, even for high mass accuracy data sets, primarily because *b*- and *y*-type ions often share the same chemical composition. For example, a *b*-type ion comprising the amino acids Pro, Pro, and Val has the same chemical composition as the *y*-type ion with Phe and Lys. This example indicates that every *y*-type ion that contains a Phe and a Lys could be confused with a *b*-type ion product (for more examples see Supporting Table 1); if only *b*- and *y*-type ions are measurable, it is possible for two different peptides to yield the same tandem mass spectrum (e.g., homeometric peptides).[9]

Electron-based methods are also used for fragmentation, i.e., electron capture or transfer dissociation (ECD or ETD).[10,11] Cleavage in these methods is directed to the N$-$C$_{\alpha}$ bond, is driven by free radical chemistry, and generates even-electron *c*-type fragments and odd-electron $z^{\bullet}$-type fragments.[12–16] Inspired by the radical nature of the $z^{\bullet}$-type ion and our recent

[†] Department of Mathematics.
[‡] Department of Chemistry.
[§] Department of Biomolecular Chemistry.

(1) Gross, M. L. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 57–57.
(2) Bristow, A. W. T.; Webb, K. S. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 1086–1098.
(3) Zubarev, R. A.; Hakansson, P.; Sundqvist, B. *Anal. Chem.* **1996**, *68*, 4060–4063.
(4) He, F.; Emmett, M. R.; Hakansson, K.; Hendrickson, C. L.; Marshall, A. G. *J. Proteome Res.* **2004**, *3*, 61–67.
(5) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198–207.
(6) Coon, J. J.; Syka, J. E.; Shabanowitz, J.; Hunt, D. F. *Biotechniques* **2005**, *38*, 519, 521, 523.
(7) Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*, 601–601.
(8) Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233–6237.
(9) Frank, A. M.; Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A.; Pevzner, P. A. *J. Proteome Res.* **2007**, *6*, 114–123.
(10) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266.
(11) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528–9533.
(12) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10313–10317.
(13) McLafferty, F. W. *Int. J. Mass Spectrom.* **2001**, *212*, 81–87.
(14) McLafferty, F. W.; Horn, D. M.; Breuker, K.; Ge, Y.; Lewis, M. A.; Cerda, B.; Zubarev, R. A.; Carpenter, B. K. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 245–249.
(15) Coon, J. J.; Syka, J. E. P.; Schwartz, J. C.; Shabanowitz, J.; Hunt, D. F. *Int. J. Mass Spectrom.* **2004**, *236*, 33–42.
(16) Pitteri, S. J.; Chrisman, P. A.; Hogan, J. M.; McLuckey, S. A. *Anal. Chem.* **2005**, *77*, 1831–1839.

10.1021/ja7099985 CCC: $40.75 © 2008 American Chemical Society

coupling of ETD with the high mass accuracy Orbitrap mass analyzer, we pondered whether *c*- and *z*•-type product ions share chemical compositions, just as their *b*- and *y*-type relatives do. Using computation, theory, and experiment we reveal here that *z*•-type product ions are chemically distinct and are readily distinguished from all other common fragment ion types by mass alone. Below we outline the source of this chemical oddity, report on the required mass accuracy to utilize it, and discuss new bioinformatic opportunities to encode and exploit such information for automated protein sequence identification.

Experimental Details

Computation. Using Microsoft Access VBA and queries, we created tables containing all elemental compositions of all possible amino acid sequences up to 2000 Da. From the elemental composition table (hereafter referred to as a mass table) we derived four subtables, each containing all fragment ion masses for a particular product ion type, i.e., *b*-, *y*-, *c*-, or *z*•-type. Each of the tables had approximately four million records and included a mass index, mass number, and a computed mass for each entry. The mass index is an encoding of the chemical composition, guaranteeing correct comparisons between any two amino acid compositions. Note that chemical composition, not computed mass, is used to index these tables as round-off error makes comparing objects of similar masses problematic. Rows in these tables represent a chemical composition for a collection of amino acid residues of a given ion type.

Instrumentation. As described in detail elsewhere, we adapted a hybrid linear ion trap-Orbitrap MS (Thermo Scientific, Bremen, Gmbh) to accommodate a chemical ionization (CI) source and modified the QLT end lenses to allow superposition of a supplementary RF voltage (to generate reagent anions for ETD and to enable charge-sign independent storage of cations and anions, respectively).[17–19] Fluoranthene reagent anions commute to the QLT through an added octopole, the C-trap, and then a second octopole. With this instrument ETD product ions can be *m/z* analyzed in either the QLT or Orbitrap. The eluent of the nHPLC experiments described above were sampled via an integrated electrospray emitter, operated at 1.7 kV, for peptide–ionization. Online MS experiments were performed by full MS analysis in the Orbitrap followed by six data-dependent MS/MS analyses with product ion analysis also performed in the Orbitrap.

Materials. Yeast cells ("wild type" S288C strain, alpha mating type, diploid, *SUC2 mal gal2 CUP1*) were cultured on yeast nitrogen base (YNB; DIFCO, Detroit, MI) that lacked amino acids and ammonium sulfate but had additional glucose (2% v/v); a loop was dipped into a 100 mL solution of 1.7 g/L YNB that lacked amino acids and ammonium sulfate but had 4% glucose, 5 g ammonium sulfate and was adjusted to a pH of 5.6 using NaOH. The solution was shaken overnight at 30 °C. After ~18 h this solution was transferred to a Fernboch and shaken for ~18 h at 30 °C. The resulting solution was spun down, washed with deionized water (DI H_2O), and then resuspended in a lysis buffer (50 mM Tris base, 0.3 M sucrose, 5 mM Na_2EDTA, 1 mM EDTA-free acid, 1 mM PMSF from 100 mM IPA stock, pH to 7.5 with HCl). For lysis, the yeast cells were sonicated in 1 min intervals for a total of ~3 min. Organelles and membranes were pelleted using centrifugation (~2000 g and ~100 000 g, respectively). After an acetone precipitation, the soluble proteins were resuspended in 6 M

(17) McAlister, G. C.; Makarov, A.; Horning, S.; Phanstiel, D.; Swaney, D. L.; Syka, J. E.; Stafford, G. C.; Lange, O.; Coon, J. J. *55th American Society for Mass Spectrometry Conference*; Indianapolis, IN, 2007.
(18) McAlister, G. C.; Lange, O.; Griep-Raming, J.; Schwartz, J. C.; Syka, J. E. P.; Stafford, G. C.; Horning, S.; Phanstiel, D.; Swaney, D. L.; Coon, J. J. Submitted.
(19) McAlister, G. C.; Swaney, D. L.; Coon, J. J. Submitted.

Figure 1. The valence parity rule dictates that the sum of N and H atoms is always odd for *c*-type, *b*-type, and *y*-type ions and even for *z*•-type ions.

guanidinium hydrochloride (GHCl). The protein extracts were equally aliquoted and digested with Lys-C (0.5 M GHCl for ~18 h at pH ~8).

For the liquid chromatography tandem MS analysis approximately 8 pmol of yeast digest was bomb loaded onto a 360 μm × 75 μm self-prepared microcapillary precolumn, which was fritted (Lichrosorb Si60, EM Separations Technology, Gibbstown NJ) and packed in-house with reversed-phase C_{18}material to 5 cm (Alltima 5 μm beads from Alltech Associates, Inc., Deerfield IL). A separate 360 μm × 50 μm microcapillary column was attached to the precolumn, also packed with 7 cm of reversed-phase C_{18} material, with a Teflon tubing butt-joint. The second column had an integrated ESI tip, made using a laser puller (model P-2000, Sutter Instrument Company, Novato CA).[20] A 65 min gradient was run over these attached columns using an Agilent 1100 series HPLC and a micro-T for flow splitting (Agilent Technologies, Palo Alta CA). Solvent A consisted of 100 mM of aqueous acetic acid; solvent B consisted of 70:30 acetonitrile to water with 100 mM of acetic acid. The gradient began with 100% A but then went to 70:30 A/B in 1 min. Next, the gradient was increased linearly over the next 44 min to 60% B. The system held at 40:60 for 15 min, and then it made a final push to 100% B over the last 5 min.

Results and Discussion

Valence Parity Rule. The free radical-driven cleavage observed from ECD/ETD prompted us to closely examine *c*- and *z*•-type fragment ion chemical compositions and to compare them to their *b*- and *y*-type kin. As such, we constructed a database consisting of all nonredundant peptide chemical compositions up to 2000 Da (considering all 20 amino acids, 3 962 682 entries, *vide supra*). From these entries we tabulated the corresponding chemical compositions of the four fragment ion types for each (8 447 133 nonredundant entries). Through careful inspection of these formulas we noticed that *z*•-type fragment ions do indeed have distinctive chemical compositions. Specifically, *b*-, *c*-, and *y*-type ions have an *odd* number of atoms with an odd valence (e.g., N and H), while *z*•-type ions contain an *even* number of atoms with an odd valence (see Figure 1). More generally, the sum of the valences for all atoms comprising *b*-, *c*-, and *y*-type ions is *odd*, and the sum of the valences for all atoms comprising *z*•-type ions is *even*.

Valence Parity Proof. Let us notice that, according to the standard way of representing any amino acid graphically, where the nodes are atoms of H, O, S, N, C and the edges are the chemical bonds, the following equation holds:

$h + 2o + 2s + 3n + 4c = 2$(number of chemical bonds), where h = the number of H atoms, o = the number of O atoms, etc. The equation above follows from the observation that

Table 1. Chemical Composition of Individual Amino Acid Residues and the Atomic Contribution of the Four Ion Fragment Types Discussed in the Text[a]

Abbr	Amino Acid	Integer Residue Mass (mono-isotopic)	Accurate Residue Mass (mono-isotopic)	S	O	N	C	H	N+H
G	Glycine	57	57.0214637	0	1	1	2	3	4
A	Alanine	71	71.0371138	0	1	1	3	5	6
S	Serine	87	87.0320284	0	2	1	3	5	6
P	Proline	97	97.0527638	0	1	1	5	7	8
V	Valine	99	99.0684139	0	1	1	5	9	10
T	Threonine	101	101.047678	0	2	1	4	7	8
C	Cysteine	103	103.009185	1	1	1	3	5	6
L	Leucine	113	113.084064	0	1	1	6	11	12
I	Iso-leucine	113	113.084064	0	1	1	6	11	12
N	Asparagine	114	114.042927	0	2	2	4	6	8
D	Aspartic Acid	115	115.026943	0	3	1	4	5	6
Q	Glutamine	128	128.058578	0	2	2	5	8	10
K	Lysine	128	128.094963	0	1	2	6	12	14
E	Glutamic Acid	129	129.042593	0	3	1	5	7	8
M	Methionine	131	131.040485	1	1	1	5	9	10
H	Histidine	137	137.058912	0	1	3	6	7	10
F	Phenylalanine	147	147.068414	0	1	1	9	9	10
R	Arginine	156	156.101111	0	1	4	6	12	16
Y	Tyrosine	163	163.063329	0	2	1	9	9	10
W	Tryptophan	186	186.079313	0	1	2	11	10	12

Ion Type	Mass	S	O	N	C	H	N+H
c	18.0343741	0	0	1	0	4	5
z•	2.99966565	0	1	-1	0	1	0
b	1.00782503	0	0	0	0	1	1
y	19.0183897	0	1	0	0	3	3

[a] Colored text illuminates the cases where the number of N + H is odd. An ion fragment consists of a collection of items from the top table and one item from the bottom table.

the sum $h + 2o + 2s + 3n + 4c$ represents the total number of bonds adjacent to all the atoms, and the result is twice the total number of chemical bonds because each bond joins exactly two atoms. Therefore the number $h + 2o + 2s + 3n + 4c$ must be *even*, which implies that $h + 3n$ must be *even*, thus $h + n$ must be *even*. This property holds not only for amino acids but also for any neutral nonradical molecules. The general result is that the total number of odd valence atoms must be *even*. This fact and other similar rules have been discussed previously and are collectively known as Senior's rules.[21–23] The nitrogen rule and the rings plus double bonds (r + db) formula are examples of more familiar, but related, mass spectrometry deduction tools.[24]

Senior's rule that the total number of odd valence atoms must be *even* holds when applied to neutral masses. And, in the case of product ions derived from peptidic precursors, a similar tenet holds. The chemical compositions of such product ions comprise two components: the collection of amino acids they make up (residue masses) and (2) a common modification specific to each product ion type (we shall call this the ion cap).[8] For example, a *c*-type ion containing only glycine would have the chemical composition of the glycine residue ($O_1N_1C_2H_3$) plus the *c*-type ion cap: one nitrogen atom and four hydrogen atoms (see Table 1), summing to $O_1N_2C_2H_7$. For the $z^•$-type ion equivalent we must subtract a nitrogen atom and add one hydrogen atom and one oxygen atom yielding $O_2N_0C_2H_4$. More complex ion fragments would include more amino acids but only one ion cap. The last column of Table 1 illustrates that $h + n$ is *even* for each of the 20 amino acid residues (following Senior's rules). The ion caps for *b*-, *c*-, or *y*-type ions, however, have an $h + n$ value that is *odd*, while the $z^•$-type ion cap exhibits an *even* $h + n$ sum. Since a sum of even numbers is *even*, the $h + n$ count is *always odd* for *b*-, *c*-, and *y*-type ions and is *always even* for $z^•$-type ions.

Valence Parity Rule − Theory. This tenet, hereafter referred to as the valence parity rule, mandates that no *c*-type ion shall have the same chemical composition, and by extension mass, as a $z^•$-type ion. In contrast, many *b*-type ions share their chemical compositions with *y*-type ions. Figure 1 illustrates the confusion that can arise during peak assignment for CAD fragments as two of the *b*-type ions have the same chemical composition as *y*-type ions. Note that sequence PPV**NGG**FK is also homeometric with the sequence PPV**GGN**FK. The valence parity rule eliminates the possibility of homeometric peptides when considering *c*-type and $z^•$-type ions with perfect accuracy.

To determine the probability of unambiguous product ion-type assignment for either CAD or ECD/ETD spectra we plotted the frequency at which ion pairs (*b*/*y*- or *c*/$z^•$-type, all possible unordered amino acid compositions, *vide supra*) could be separated as a function of fragment size at various mass accuracies (Figure 2). To do this we created four tables, one for each fragment ion type (*b*-, *c*-, *y*-, and $z^•$-type), containing every possible chemical composition for that ion type up to 2000

(20) Martin, S. E.; Shabanowitz, J.; Hunt, D. F.; Marto, J. A. *Anal. Chem.* **2000**, *72*, 4266–4274.
(21) Senior, J. K. *Am. J. Mathematics* **1951**, *73*, 663–689.
(22) Meija, J. *Anal. Bioanal. Chem.* **2006**, *385*, 486–499.
(23) Kind, T.; Fiehn, O. *BMC Bioinf.* **2007**, 8.
(24) McLafferty, F. W.; Turecek, F. *Interpretation of Mass Spectra*; University Science Books: 1993.

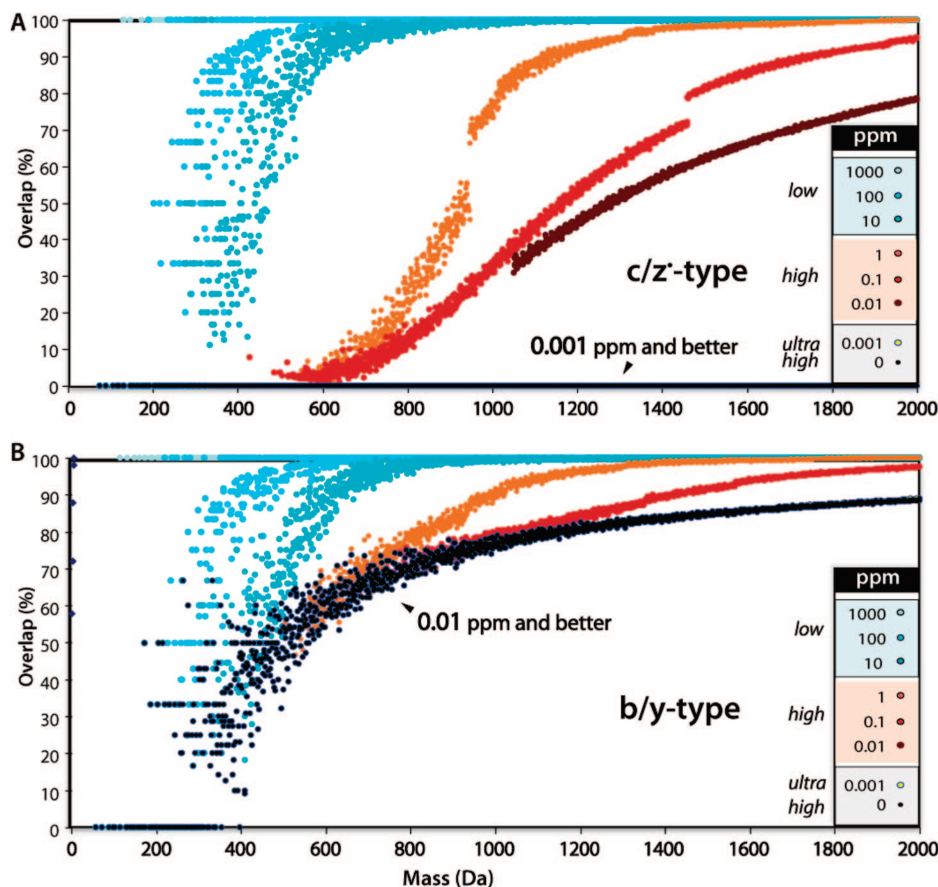


Figure 2. Comparative analysis of ambiguity in complementary ion pairs at various mass accuracies.

Da. Next we computed the difference, in ppm, between the nearest theoretical $z^{\bullet}$-type product ion mass to each *c*-type product ion mass (we also made a similar computation for the *b*- and *y*-ion pair combinations, panels A and B, respectively). We then aggregated the product ions by mass number and plotted the frequency of their separation as a function of fragment ion mass at various mass accuracies. For example, in a tandem mass spectrum containing *c*- and $z^{\bullet}$-type product ions collected at 0.1 ppm mass accuracy, 50% of these ions can be assigned an unambiguous type at mass 1200 Da (Figure 2, panel A, red curve). More specifically, at this mass accuracy (0.1 ppm) half of the *c*-type product ions at mass number 1200 are within ±0.05 ppm of a theoretical $z^{\bullet}$-type ion.

The black curves shown in Figure 2 demonstrate the valence parity rule; with perfect mass accuracy *c*- and $z^{\bullet}$-type ions are always distinguished (panel A), while *b*- and *y*-type ions having masses greater than 600 Da share identical chemical compositions more than 50% of the time (panel B). The chemical distinction of $z^{\bullet}$-type ions only becomes relevant, however, if product ion masses are known to within ~5 ppm. In fact, the eight mass accuracies, spanning seven decades, shown in Figure 2 indicate natural breakpoints for ion type resolving power. The group shown in cool colors (low; 1000–10 ppm, blue) is not adequate to exploit the valence parity rule. The high mass accuracy group (1–0.01 ppm, warm colors) performs well in this regard with the ion type separation capacity slowly fading with increasing mass. Complete separation of *c*- and $z^{\bullet}$-type ions (up to 2000 Da) is readily achieved by the ultrahigh group, which begins at 0.001 ppm (grayscale).

Next we asked, for a given mass accuracy, what is the first mass at which ion type assignment is ambiguous (i.e., overlap percentage)? The answer for *b*- and *y*-type ions is 173.0926 Da, provided that the accuracy is 100 ppm or better (Figure 3, panel A). Heightened mass accuracy will not improve the situation since the *b*-type ion (TA) has exactly the same chemical composition as the confounding *y*-type ion (PG). The *c*- and $z^{\bullet}$-type ion case is much more interesting. Mass accuracies between 7.2 and 160 ppm will have difficulties separating a particular ion pair at 188 Da, but increasing mass accuracy (better than 7.2 ppm) yields an improved range of unambiguous identifications, hitting another plateau at 428 Da for accuracies from 1.3 to 0.013 ppm (Figure 3, panel A). Panels B and C of Figure 3 display at which masses are either 50% or 90% of the ion type assignments ambiguous. These data show that with 1 ppm mass accuracy 50% of all *c*- and $z^{\bullet}$-type ions at mass 1050 are unambiguously assigned; with the same accuracy this limit is reached at 602 Da using *b*- and *y*-type ions.

Post-translational modifications (PTMs) of amino acid side chains occur and, in principle, could disrupt the valence parity rule. We examined four common PTMs—acetylation, methylation, oxidation, and phosphorylation—and verified that the valence parity rule holds for all of them. And, because it is defined by valence parity, we reason the rule will hold for all modifications.

Valence Parity Rule − Application. ECD tandem mass spectra are nearly always acquired by Fourier transform ion cyclotron resonance mass spectrometers (FT-ICR-MS) where achieving

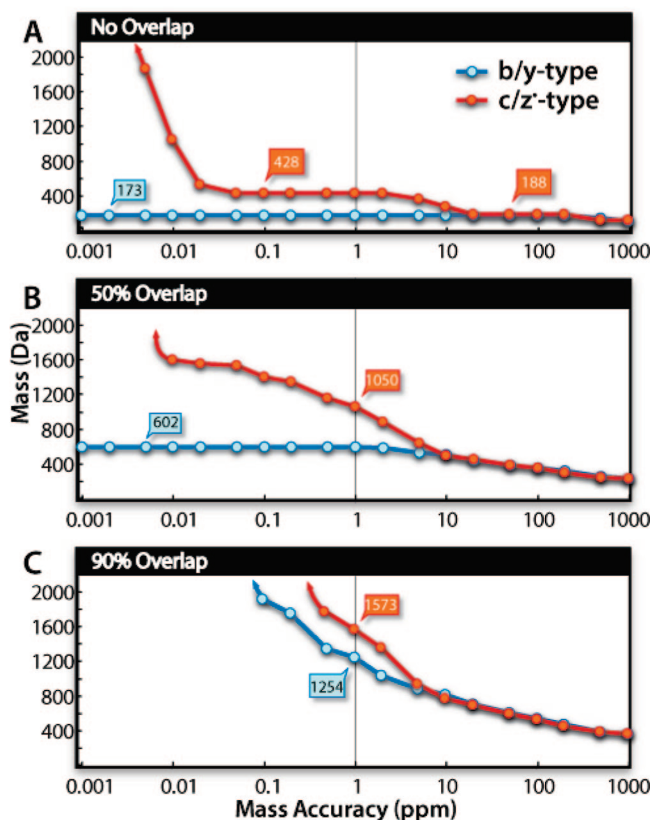


Figure 3. Relationship between mass accuracy and certainty of peak assignment to ion type where the cumulative ambiguity of peak assignment is 0% (A), 50% (B), and 90% (C). Arrows indicate the point at which the curve jumps beyond our analysis, at 2000 Da.

1 ppm (or better) mass accuracy is routine.[25] Recently, we have adapted a hybrid linear ion trap-Orbitrap mass spectrometer to perform ETD reactions and have demonstrated its ability to measure *c*- and $z^{\bullet}$-type product ion masses to within 2 ppm.[17–19,26] Using our ETD-enabled Orbitrap we tested our theory by examining 226 single-scan ETD-MS/MS spectra for ion type and chemical composition assignment. The spectra resulted from a single nanoflow-LC-MS/MS analysis of a complex mixture of yeast peptides (endo-LysC digest). Prior to examination, sequences had been assigned to each of the 226 spectra with high confidence using a database correlation algorithm. Manual inspection of these spectra confirmed the presence of 1962 distinct *c*- and $z^{\bullet}$-type ions with a signal-to-noise (*S*/*N*) ratio greater than 2:1. Using a 4 ppm window, we could identify one or more candidate chemical compositions for 92% of these *m*/*z* peaks (1,812), with 48% given an unambiguous chemical composition and ion-type assignment (937 peaks).

Panel A of Figure 4 displays that the ambiguity rates for these experimental data closely correlate with theory. An automated peak annotation algorithm produced an average of 8.0 chemically identifiable peaks per spectrum, for which 4.1 could be assigned an unambiguous ion type and chemical composition. Note that a minority of these are false positives. The mass

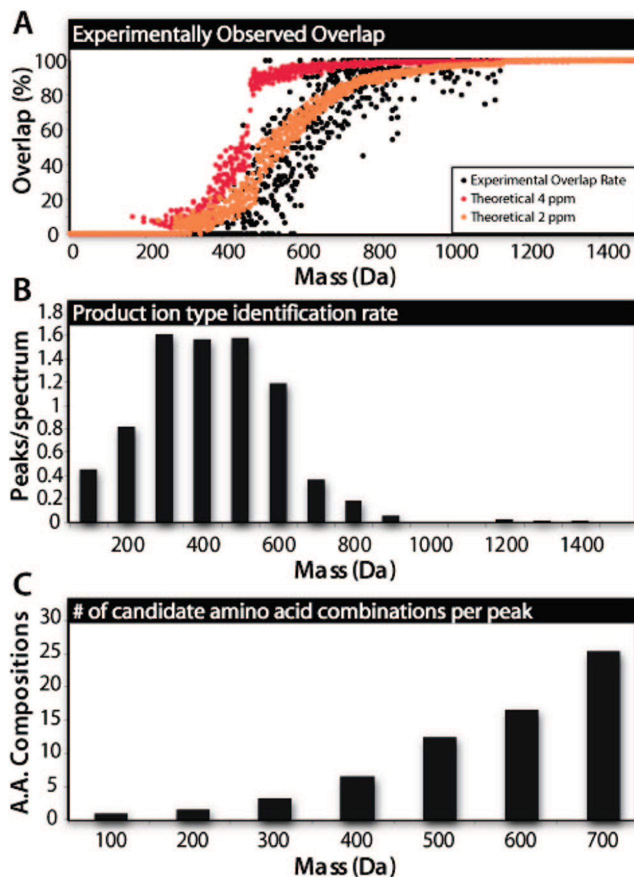


Figure 4. Ambiguity of ion type assignments for experimental data. Panel A displays the percentage of peaks for which there was both a *c*- and a $z^{\bullet}$-ion fragment assignment for each mass (Da) using a window of 4 ppm. Theoretical accuracy curves for 4 ppm and 2 ppm are shown for reference. Note the data match the 2 ppm theoretical results instead of the 4 ppm. This is because half of the measured mass values will fall on the far-side of the true mass from the peak that may confound the detection. Panel B displays the average number of peaks per spectrum identified in the indicated mass range. The average number of candidate amino acid combinations per identified peak is shown in panel C.

distribution of these identified peaks is presented in panel B of that figure. We then computed the average number of unordered amino acid compositions that could result from the identified *c*- or $z^{\bullet}$-type ions chemical composition (Figure 4, panel C). From these data we conclude that, for even relatively high mass identified peaks (e.g., 700 Da), there exist no more than 30 possible amino acid compositions on average. Of course, the presence of four identified peaks per spectrum allows one to rapidly scan these short lists for common amino acid combinations or to assign the previously ambiguous peaks.

Discussion

The above application of the valence parity rule required the following implicit assumptions: the *m*/*z* candidate peaks are singly charged, monoisotopic, and either *c*- or $z^{\bullet}$-type. The presence of *m*/*z* peaks not derived from any of these categories can, of course, confound analysis. Most proteomic applications rely on the use of proteases to generate short, lowly charged peptide precursors—where fragment ions are most frequently singly charged—our ETD-MS/MS data set being no exception. Nonetheless, the presence of multiply charged product ions can complicate ion type identification in two ways. First, it is possible for a multiply charged fragment to have a total mass

(25) Marshall, A. G.; Hendrickson, C. L.; Ernmetta, M. R.; Rodgers, R. P.; Blakney, G. T.; Nilsson, C. L. *Eur. J. Mass Spectrom.* **2007**, *13*, 57–59.
(26) McAlister, G. C.; Phanstiel, D.; Good, D. M.; Berggren, W. T.; Coon, J. J. *Anal. Chem.* **2007**, 3525–3534.

that is exactly an integer multiple of a theoretical, singly charged fragment. In practice this is not common since it requires that all of the elements of the multiply charged fragment have a common factor. The second obstacle, which is more difficult to review, is the possibility of a multiply charged fragment being close, but not identical, to the same mass as a singly charged product ion. Because of the tremendous number of possible chemical compositions at higher masses, the theoretical coverage by multiply charged fragments is much greater than that of singly charged ions. Thus, the likelihood that a multiply charged ion *m/z* is close to a theoretical singly charged product ion mass is quite low; however, the reverse situation (a singly charged product ion mass being close to a theoretical multiply charged product ion *m/z*) is quite high.

The relatively low abundance of multiply charged product ion peaks in our data suggests that the multiple isotopic peaks displayed by product ions is likely the most important issue to consider when separating *c*- and *z*•-type ions based on mass. There are two reasons for this: (1) nearly every product ion *m/z* displays at least one ^{13}C isotope peak, and (2) the addition of a single neutron is very close in mass to that from addition of a single hydrogen atom. Thus, an isotopic form of a *c*-type fragment can potentially be confounded with a *z*•-type fragment and vice versa. Fortunately, mass analyzers that achieve the mass accuracies required to distinguish *c*- and *z*•-type ions also afford high mass resolving power. Such resolution makes it possible to decode the isotopic cluster *m/z* peaks and to determine the charge state of product ions *en masse* in an automated fashion.[27–31] These deconvolution programs continue to evolve, and their use has become relatively routine so that multiple charging and the presence of isotopic cluster *m/z* peaks should not present an obstacle for valence-parity-based ion type assignment.

While multiple charges and isotopes can be easily dealt with, both ECD and ETD spectra can be complicated by the presence of *a*•- and *y*-type ions, a minor fragmentation pathway.[32] The existence of either can complicate matters. First, the radical *a*•-type ion has a valence parity that matches that of the *z*•-type fragment; i.e., the sum of the odd valence atoms is always *even*. This means that occasionally *a*•- and *z*•-type fragments share identical chemical compositions. Likewise *c*- and *y*-type fragments have *odd* valence parities, which again permits chemical composition overlap. Fortunately, from our experience with ETD we note that *a*•- and *y*-type product ions exist at very low abundance or, more often, are not present at all. To quantify this we closely examined (manually) the 226 tandem mass spectra used above for the presence of *a*•- and *y*-type product ions. Recall these spectra contained 1962 distinct *c*- and *z*•-type ions (S/*N* > 2:1). Using the same criterion we identified 35 *a*•-type and 46 *y*-type products. From these spectra *a*•- and *y*-type ions represent only 3.9% of the total product ion population. Thus, for ETD, our assumption that all product ions are either *c*- or *z*•-type is reasonable and will rarely (<∼ 4%) lead to false positive peak identification. Finally, elevation of precursor ion internal energy in ECD or supplemental activation with ETD can lead to the production of odd electron *c*-type ions and even electron *z*-type ions (*c*•- or *z*-type)-an obvious complication for application of valence parity that could elevate false identification rates.[33–35]

From these data we conclude that the valence parity rule will allow for the rapid and accurate annotation of ECD/ETD tandem mass spectra. Specifically, at 4 ppm mass accuracy nearly half of all observed *c*- and *z*•-type products were assigned an ion type and unambiguous chemical composition. Such capabilities will afford new bioinformatic approaches for peptide sequence derivation. First, one can envision submission of tandem mass spectra to a database correlation algorithm with embedded ion type annotation. As an example, McLuckey et al. have recently examined the benefits of knowing *z*•-type product ions *a priori* and using them to exclude putative identifications based on solely peak matching.[36] *De novo* sequence analysis is a second field that will likely be propelled by application of the valence parity rule. Spengler, for example, has reported on the use of high mass accuracy spectra (CAD-type) for *de novo* sequencing through a process referred to as peptide composition analysis.[37,38] Here the peptide's amino acid composition is determined by using the precursor ion mass and a few selected fragment ion masses. A difficulty the strategy faces is that, for large masses, *b*- and *y*-type fragments are often confounded (Figure 2). Doubtless this approach will be advanced by *a priori* knowledge of fragment ion type and chemical composition. We also envision this information will allow for the development of newer, chemical composition-based *de novo* sequencing strategies.

Conclusions

Chemical derivation of the N- or C-terminus, gas-phase chemical reactions and, more recently, the collection of sequential tandem mass spectra have been proposed as methods for ion type assignment.[39–41] Here we present a more expedient and direct route—namely, that the major product ions formed from ETD and ECD are chemically distinct, regardless of amino acid composition or the presence of common PTMs. This difference becomes relevant, and exploitable, at mass accuracies better than 5 ppm. With experimental data we show that nearly half of all peaks attributable to *c*- and *z*•-type ions can be assigned an ion type and unambiguous chemical composition. Such assignments provide (1) a platform to directly determine amino acid composition, (2) an input for database search algorithms, or (3) a basis for *de novo* sequencing. Finally, by restricting candidate *m/z* peaks in spectral processing to only

(27) Mann, M.; Meng, C. K.; Fenn, J. B. *Anal. Chem.* **1989**, *61*, 1702–1708.
(28) Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 52–56.
(29) Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229–233.
(30) Zhang, Z. Q.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 225–233.
(31) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320–332.
(32) Kruger, N. A.; Zubarev, R. A.; Horn, D. M.; McLafferty, F. W. *Int. J. Mass Spectrom.* **1999**, *187*, 787–793.
(33) Horn, D. M.; Ge, Y.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 4778–4784.
(34) Oh, H.; Breuker, K.; Sze, S. K.; Ge, Y.; Carpenter, B. K.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15863–15868.
(35) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. *Anal. Chem.* **2007**, *79*, 477–485.
(36) Liu, J.; Liang, X. R.; McLuckey, S. A. *J. Proteome Res.* **2008**, *7*, 130–137.
(37) Spengler, B. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 703–714.
(38) Spengler, B. *Eur. J. Mass Spectrom.* **2007**, *13*, 83–87.
(39) Xia, Y.; Chrisman, P. A.; Pitteri, S. J.; Erickson, D. E.; McLuckey, S. A. *J. Am. Chem. Soc.* **2006**, *128*, 11792–11798.
(40) Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. *Mol. Cell. Proteomics* **2005**, *4*, 835–845.
(41) Tsybin, Y. O.; He, H.; Emmett, M. R.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2007**, *79*, 7596–7602.

those with ion type and chemical composition assignment, one should expect increases in the speed and accuracy of sequence assignment.

Acknowledgment. We thank the University of Wisconsin-Madison, Thermo Fisher, the Beckman Foundation, the American Society of Mass Spectrometry, Eli Lilly, the National Science Foundation (0701846 to G.C. and J.J.C.; 0747990 to J.J.C.), and the NIH (1R01GM080148 to J.J.C.) for financial support of this work. S.L.H. and G.C.M. gratefully acknowledge support from NIH predoctoral fellowships: Computational Bioinformatics (NLM Training Grant 5T15LM007359 to S.L.H.) and the Biotechnology Training Program (NIH 5T32GM08349 to G.C.M.). Finally, we thank John Syka and Mike Senko for helpful discussions.

Supporting Information Available: Table of amino acid combinations where the *b*-type ion fragment has the same m/z value (where $z = 1$) as a *y*-type ion fragment. This material is available free of charge via the Internet at http://pubs.acs.org.

JA7099985