

# Bifurcating Fragmentation Behavior of Gas-Phase Tryptic Peptide Dications in Collisional Activation

Mikhail M. Savitski, Maria Fälth, Y. M. Eva Fung,  
Christopher M. Adams, and Roman A. Zubarev

Institute for Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

---

Collision-activated dissociation (CAD) of tryptic peptides is a cornerstone of mass spectrometry-based proteomics research. Principal component analysis of a database containing 15,000 high-resolution CAD mass spectra of gas-phase tryptic peptide dications revealed that they fall into two classes with a good separation between the classes. The main factor determining the class identity is the relative abundance of the peptide bond cleavage after the first two N-terminal residues. A possible scenario explaining this bifurcation involves *trans*- to *cis*-isomerization of the N-terminal peptide bond, which facilitates solvation of the N-terminal charge on the second backbone amide and formation of stable  $b_2$  ions in the form of protonated diketopiperazines. Evidence supporting this scenario is derived from statistical analysis of the high-resolution CAD MS/MS database. It includes the observation of the strong deficit of  $a_3$  ions and anomalous amino acid preferences for  $b_2$  ion formation. (J Am Soc Mass Spectrom 2008, 19, 1755–1763) © 2008 American Society for Mass Spectrometry

---

Collision-activated dissociation (CAD) of tryptic peptides is a cornerstone of mass spectrometry-based proteomics, in which proteins are identified by matching experimental datasets obtained with tandem mass spectrometry (MS/MS) of their peptides with predicted MS/MS patterns of theoretical peptides derived from a sequence database. Matching algorithms used in such popular search engines as Mascot [1] and Sequest [2] do not predict relative abundances of fragment ions and are mostly concerned with matching ionic masses of certain fragment types. At the same time, fragment ion abundances in CAD MS/MS spectra can be relatively reproducible and are far from homogeneous—and to use the patterns of these abundances for sequence verification has for a long time been a tempting idea. Two alternative approaches were pursued in implementation of this idea. One approach was to collect a massive database of tandem mass spectra of all possible peptides and to make direct comparison between the standard MS/MS datasets and the experimental ones [3]. Although this approach has many merits, it faces a huge challenge in the form of an astronomical number of peptide sequences existing in nature. The alternative approach is to predict fragment abundances for a given sequence. The abundance-predicting algorithm by Zhang has more than 200 adjustable parameters but still exhibits only 70% accuracy [4, 5]. The global model of CAD fragmentation

developed by us for tryptic peptide dications uses only 20 parameters, one for each amino acid [6]. In preliminary assessment of the prediction accuracy of that model we have found a bifurcating behavior. It appeared that, although a large group of mass spectra behaved in good agreement with the model's predictions, another large group showed significantly deviating features. Thus the question arose as to whether all tryptic peptide dications constitute one group in the statistical sense. This question is critical for every abundance-predicting algorithm because different models of fragmentation should be applied to describe the behavior of each group. The obvious hypothesis that one group represented peptides ending with Lys and another one with Arg was quickly tested and rejected. Another suggestion was that the division line was the presence or absence of a particular amino acid, e.g., proline, histidine, or aspartic acid, which are known for their effect on bond fragmentation [7, 8]. This hypothesis was also tested and rejected. Another anomaly in fragmentation could arise from the presence of basic residues in internal parts of the sequence for peptides with cleavage sites missed by trypsin. Thus the sequences with missed cleavages were also removed from consideration.

Herein we present evidence for the presence of two statistical ensembles in the CAD MS/MS database even after removal of all known sources of artifacts. We also suggest and discuss a hypothesis that is capable of explaining the bifurcating behavior of CAD mass spectra.

---

Address reprint requests to Prof. Roman A. Zubarev, Uppsala University, Division of Molecular Biometry, Institute for Cell and Molecular Biology, Box 596, SE-75 124 Uppsala, Sweden. E-mail: [roman.zubarev@icm.uu.se](mailto:roman.zubarev@icm.uu.se)

## Experimental

MS/MS data were collected from the proteomics analysis of lysates of human cell lines obtained from Santa Cruz Biotechnology (Santa Cruz, CA, USA) and *Escherichia coli* was obtained from Sigma performed on a linear ion trap/Fourier transform (LTQ-FT) mass spectrometer (ThermoFisher Scientific, Braunschweig, Germany) using consecutive electron capture dissociation (ECD) and CAD fragmentation of peptides eluting from the analytical column of a nano-LC (liquid chromatography) system (Agilent 1100, Agilent Technologies, Santa Clara, CA, USA). The peptides were identified by the Mascot search engine (Matrix Sciences, London, UK). The details of the experimental procedure and data processing are described in Nielsen and colleagues [9] and Savitski and colleagues [10].

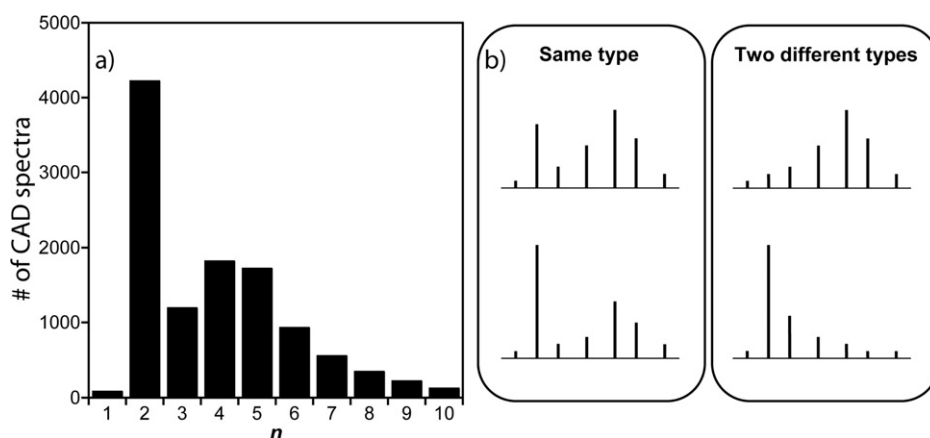
## Results and Discussion

Distribution of the positions of the maximum-intensity cleavage sites in tryptic peptides is presented in Figure 1a. Compared to Figure 1 in Savitski et al. [11], sequences were removed that contain internal basic residues Arg and Lys, appearing as a result of imperfect trypsin action (missed cleavage) and occurrence of Arg/Lys-Pro combinations. The bimodal character of the distribution manifests itself through the presence of two local maxima corresponding to the loss of two and four amino acids from the N-terminus, respectively. However, by itself such bimodality is not a definite proof of the presence of two distinct components. As illustrated in Figure 1b, such a distribution can be obtained both from the spectra of largely the same character (left panel) and from those of vastly different characters (right panel). In the first case, each spectrum exhibits intrinsic heterogeneity in the form of bimodality, with local maxima at  $y_{m-2}$  and  $y_{m-4}$  or  $y_{m-5}$ . In the alternative case there are two different classes of

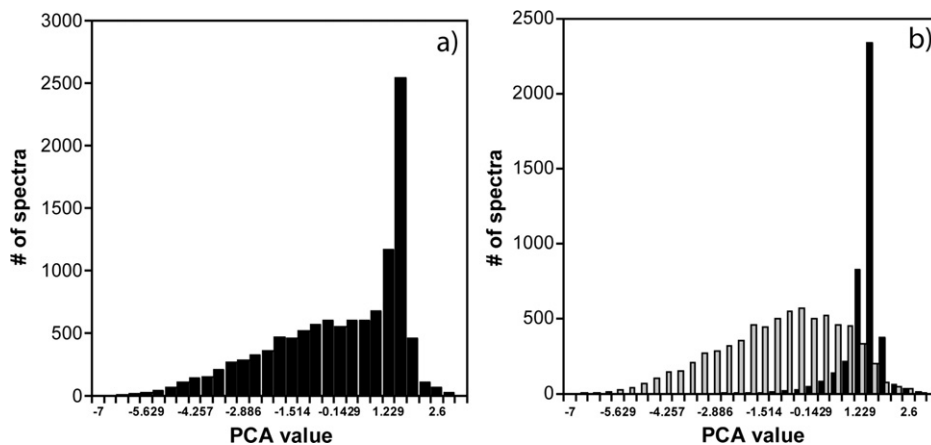
spectra, each of a homogeneous nature, one peaking at  $y_{m-2}$  and the other one at  $y_{m-4}$  or  $y_{m-5}$ . Both cases lead to the same distribution. To determine which of the two alternatives is present in reality, we applied principal component analysis (PCA) to the CAD mass spectra in the database.

### Principal Component Analysis

PCA (SIMCA-P software version 11, Umetrics AB, Umeå, Sweden) is used to reduce the number of dimensions of the analytical data in such a way as to maximize the variance of the reduced data. In our case, the whole space had 10 dimensions as we analyzed  $y_{m-n}$  ions, where  $m$  is the peptide length and  $1 \leq n \leq 10$ . For peptides with the length  $m \leq 10$ , the abundances of  $y_{m-n}$  ions with  $n > m$  were set to zero. Normalization was performed by dividing each  $y_{m-n}$  abundance by that of the most abundant  $y_{m-n}$  ion. PCA replaced the original 10 intensities by a single global variable, a linear combination of the 10 intensities. In a “blind” reduction, no assumption on the nature of data was made—i.e., it was implicitly assumed that all data points belong to the same statistical group. To test whether this assumption held, the histogram of the data points was investigated (Figure 2a). The clearly bimodal shape of the histogram revealed the presence of two major components: (1) a broad distribution that peaks at the value of approximately  $-0.6$  of the new global PCA variable and (2) a sharp distribution peaking at about  $1.5$ . Since the new global variable was a linear combination of  $y_{m-n}$  ion abundances, the coefficients of that combination reflected the relative contributions of each cleavage. The vector of the linear coefficients was  $[0.172, 0.410, 0.185, -0.044, -0.265, -0.367, -0.403, -0.395, -0.368, -0.327]$ . The largest absolute value,  $0.410$ , was thus acquired by the second coefficient in the vector, which corresponds to the  $y_{m-2}$  ion. Thus in the simplest



**Figure 1.** (a) Distribution of the positions of maximum intensity cleavage sites in the database of CAD mass spectra of tryptic peptides without missed cleavages. Each bin corresponds to a  $y_{m-n}$  ion where  $m$  is the peptide length, and  $n$  is the cleavage site counting from the N-terminus. (b) Examples of a homogeneous (left) and a heterogeneous with two classes (right) subsets of spectra that could give rise to the distribution in (a).



**Figure 2.** Results of the principal component analysis yielding one global parameter instead of the 10-dimensional input vector  $y_{m-1}, y_{m-2}, \dots, y_{m-10}$  (same data as in Figure 1a): (a) overall distribution; (b) black columns, CAD mass spectra with  $y_{m-2}$  as the most abundant peak; gray columns, all other spectra.

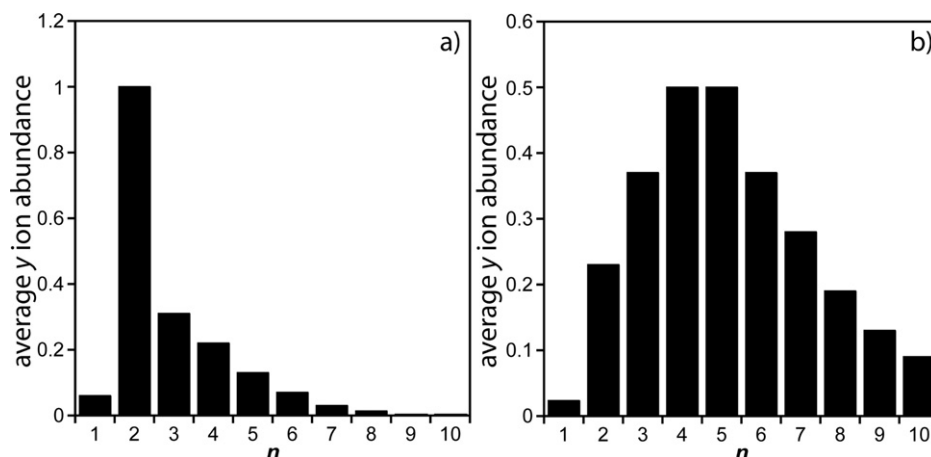
interpretation, the sharp distribution should correspond to the spectra with  $y_{m-2}$  ion being the most abundant of all  $y$  species, whereas the broad distribution should correspond to all other spectra. The validity of this interpretation was clearly confirmed when the dataset was split according to that principle (“the  $y_{m-2}$  rule”) and corresponding distributions were plotted separately (Figure 2b).

If the  $y_{m-2}$  ion abundance was the *only* feature separating the two classes of datasets, its removal from the input data for PCA would eliminate the bimodality of the output. However, upon removal of the  $y_{m-2}$  ion abundances and renormalizing the nine remaining abundances by their highest value, the bimodality was reduced but did not disappear: the peak of the sharp distribution was still higher than the peak of the broad component by a factor of 2.5. Thus not only the  $y_{m-2}$  ion abundance but the whole fragmentation pattern in any given MS/MS spectrum bears the signature of one of the two classes.

To reveal the patterns characteristic for both classes, we plotted the distributions of the average cleavage abundances in the two datasets separated by the  $y_{m-2}$  rule (Figure 3). In class 1 (~35% of the data), the  $n = 2$  ions in the  $y_{m-n}$  series are the most abundant (Figure 3a), with monotonously decreasing abundances for  $n > 2$ . In class 2 (65% of the data, Figure 3b), there is a broad distribution of ion abundances with maximum at  $n = 4-5$ .

#### Possible Origin of the Two Classes

The stark difference between the classes strongly suggests a single and persistent reason for their existence. Since this reason is not related to the presence or absence of a single amino acid (see preceding text), it is most likely structural. The most straightforward structural parameter of a linear peptide is its length. Indeed, in terms of the peptide lengths, class 1 peptides are shorter (average length 8.4 residues) than class 2 pep-



**Figure 3.** Average  $y_{m-n}$  ion abundances in CAD mass spectra where (a)  $y_{m-2}$  is the most abundant peak, and (b) all other spectra.

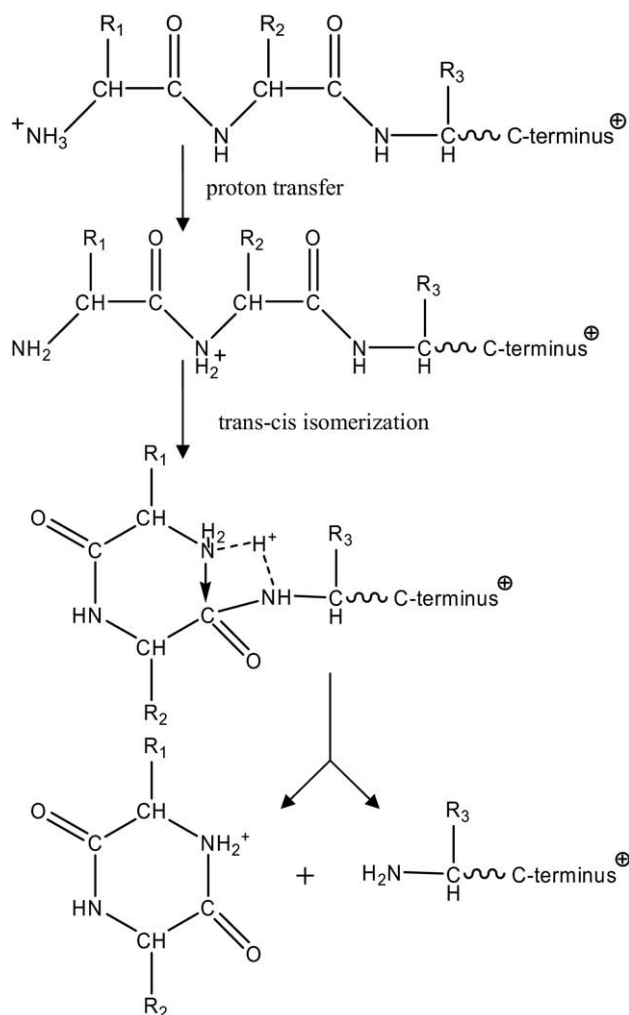
tides (average length 12.4 residues). Even for peptides of similar lengths (e.g., 8–10 residues), however, PCA showed strong division between the classes (data not shown).

It was reasonable to suggest that the fragmentation behavior of peptides is linked to their secondary structure, which for linear gas-phase peptides is held together by hydrogen bonds that can have both a neutral and a charged nature. Neutral H-bonds are much weaker than covalent bonds, but by acting in concert they can give rise to more or less defined gas-phase structures, such as  $\alpha$ -helices [12]. However, the presence of a mobile proton near the N-terminus plays a destabilizing role for helical structures, especially in shorter peptides [12]. At the same time, based on the widths of the distributions in Figures 2 and 3, the conclusion can be made that it is the generally shorter class 1 peptides (narrow distributions) that possess a common, well-defined structure, whereas the generally longer class 2 peptides (broad distributions) lack a defined common conformation. Furthermore, collisional excitation initiates unfolding of gas-phase ions and thus tends to erase or severely mask the secondary structure differences [13]. Thus the emergence of two classes has to be attributed to a more energetically stable structural feature than hydrogen bonding.

After reviewing other possible sources of structural differences between peptides, we arrived at the hypothesis that the stark effect on the fragmentation pattern may arise from the *cis/trans* isomerization of the first N-terminal peptide bond. The transition state for isomerization requires the partial double bond to be broken, with the activation energy in a free linear peptide chain of roughly 20 kcal/mol [14]. Such a large energy barrier may be the reason for the stark separation between the classes.

### Initiation of Peptide Bond Isomerization

The *trans* configuration initially dominates in all bonds of peptide dications. The N-terminal amine is usually protonated in dications of tryptic peptides because it is the most basic site after the basic C-terminal residue. The N-terminal proton tends to be solvated on one or several backbone carbonyl oxygens [15]. In one possible scenario, upon collisional dissociation this proton becomes mobile and induces fragmentation in the broad vicinity of the solvation site, which will give rise to a fragmentation pattern of class 2 (Figure 3b). Alternatively, the mobile proton may relocate from the N-terminus to the nitrogen atom of the first peptide bond. Such a process is shown in Scheme 1, which is a modified version of Scheme 4 from Paisz and Suhai [7], who previously presented computational evidence [16] that such relocation can induce *trans-cis* isomerization of the peptide bond followed by formation of a diketopiperazine intermediate and cleavage of the second peptide bond.



Scheme 1

There may be an alternative initiation mechanism for *trans-cis* isomerization. Solvation of the N-terminal proton onto backbone carbonyl oxygens involves bending of the polypeptide chain into a loop structure. As has been found by Hamelberg and McCammon [14], the stress induced on the peptide backbone by a hairpin loop reduces the isomerization barrier and enhances the rate of *trans* to *cis* conversion of the stressed peptide bond by several orders of magnitude. When the first peptide bond acquires *cis* configuration, the potential energy release reaches 17–18 kcal/mol [14]. The hairpin loop becomes more compact and the proton becomes preferentially solvated on the second carbonyl, giving rise to a cyclic diketopiperazine intermediate similar to that in Scheme 1. Such a cyclic structure is prone to facile C–N bond cleavage releasing protonated diketopiperazine [7, 16–18] from the N-terminus ( $b_2$  ion) and forming an abundant  $y_{n-2}$  ion, a class 1 behavior.

The above-mentioned hypothetical mechanism hinges on the initial formation of the N-terminal loop. The existence of the N-terminal charge-stabilized loop has been reported in the literature, e.g., by Counterman



and Clemmer [19], who simulated a singly protonated GPHH peptide in both *cis* and *trans* isoforms. In the lowest-energy *trans*-Pro isomer, the loop was large because the proton was shared between the N-terminus and the third and fourth carbonyls. The *cis*-conformer was lower in energy and the loop in it was more compact. The N-terminal proton in that structure was shared with the second carbonyl, whereas the third and fourth carbonyls were involved in charge solvation to a significantly lesser degree than in the *trans* isoform. With ion mobility spectroscopy, Counterman and Clemmens observed both *cis* and *trans* isomers for proline-containing peptides and only one isomer for peptides without a proline. Their measurements could not distinguish which (*cis* or *trans*) configuration the unique isomer has acquired. All 645 peptides they investigated were short ( $\leq 10$  residues long) and the majority were singly charged. Thus it is still an open question whether proline-free tryptic peptide dications can produce a mixture of gas-phase *cis*- and *trans*-isomers in comparable proportions.

#### Proline Effect in $b_2$ Ion Formation

Research linking  $b_2/y_{n-2}$  fragment formation with diketopiperazine formation in proline-containing peptides has been done by several authors, e.g., Smith et al. [17]. Proline is in general known to induce very strong cleavage at its N-terminal site [20], whereas the cleavage at the C-terminal site is suppressed [21]. Smith et al. [17] found that in the cationic peptide VPDPR cleavage C-terminal to the proline in the second position was unusually abundant and attributed this fact to cyclization of the first two residues into a diketopiperazine structure. It has been pointed out by Paizs and Suhai that the necessary condition for cyclization is the *cis* isomerization of the first peptide bond [16].

The enhanced cleavage after Pro-2 turned out to be a general phenomenon. We calculated the average

cleavage preferences C-terminal to proline in different positions  $n$  from our extensive dataset. As expected, the C-terminal cleavage to proline was very scarce for  $n > 2$ , whereas for  $n = 2$  the cleavage was very prominent—in fact, more prominent than the C-terminal cleavage to glycine and serine in the same position (data not shown). Thus for the Xaa-Pro N-terminal sequences, abundant  $b_2/y_{n-2}$  cleavage is a clear signature of *cis/trans* isomerization of the first peptide bond. Our data, however, support the hypothesis [18, 22] that such an isomerization frequently occurs for residues other than proline.

#### Proton Retention

Paizs and Suhai developed their model of *trans/cis* bond isomerization for neutral loss of the diketopiperazine fragment [7, 16], whereas our data point toward facile formation of ionic  $b_2$ . For the proton to be retained by  $b_2$  fragment, the diketopiperazine derivative should be more basic than the  $y_{m-2}^+$  ion. We found a strong correlation between the basicity of the two N-terminal amino acids and  $y_{m-2}^+$  formation frequencies (Figure 4, left). For reliable statistics, amino acids exhibiting strong ammonia and water losses (E, D, N, Q, S, and T) were omitted, as were amino acids rarely appearing in internal or N-terminal regions of tryptic peptides (P, C, W, M, R, and K). The dependence on the amino acid basicity is much stronger for  $n = 2$  than that for  $n > 2$ , for which case backbone carbonyl basicities are of prime importance [6].

The loss of a neutral diketopiperazine could be relatively more important for nonbasic N-termini. However, of the 78 spectra found for the extreme  $G_1G_2$  case (glycine is the least basic amino acid),  $y_{m-2}^+$  ions were present in 14 cases and  $y_{m-2}^{2+}$  ions were found in only 8 cases. Thus even for least basic N-terminal sequences, formation of protonated diketopiperazine is relatively more important than neutral diketopiperazine.

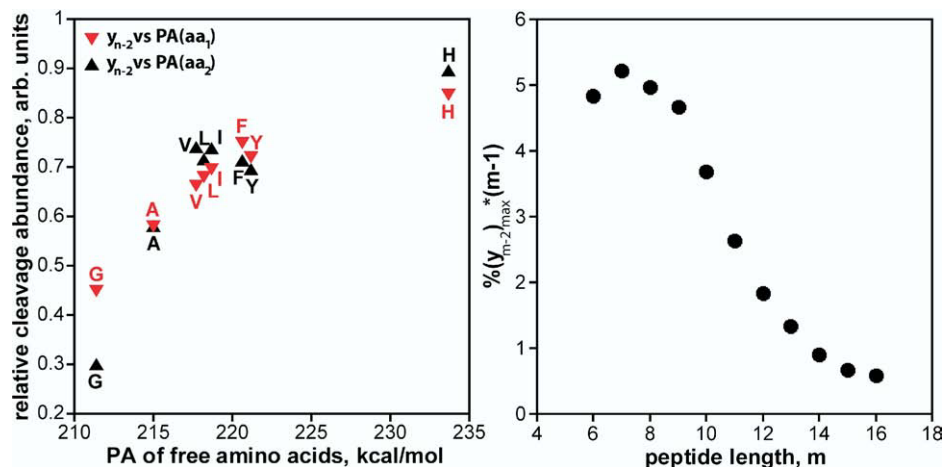


Figure 4. Left panel: dependence of the appearance frequency of dominant  $y_{m-2}^+$  ions as a function of the proton affinity of the first two N-terminal amino acids, aa<sub>1</sub> and aa<sub>2</sub>. Right panel: appearance frequency of dominant  $y_{m-2}^+$  ions multiplied by the number of peptide bonds ( $m - 1$ ).

zine loss. This result raises the issue of the reason for the surprisingly high basicity of the N-terminal leaving group. One reason for that effect could be the coulombic repulsion between the two ionizing protons. In agreement with the preference of shorter peptides for class 1 behavior, the dominance of  $y_{m-2}^+$  ions was found diminished for peptides with  $m > 10$  (right panel in Figure 4). The potential coulombic energy in a fully stretched 10-residue peptide (an average size for a tryptic peptide) protonated at both termini is approximately 9 kcal/mol, a significant value according to the left panel in Figure 4. The fall-off in the right panel of Figure 4 resembles a parabola, exactly as expected for Coulomb's law.

### Distinguishing Between the Cyclic and Oxazolone $b_2^+$ Structures

One way the cyclic and oxazolone  $b_2^+$  structures are expected to differ from each other is in the effect of the chemical nature of the first and second amino acids on the  $b_2^+$  formation probability. Oxazolone is formed via an asymmetric intermediate, which should reflect in vastly different, and indeed opposite, reactivities of these two residues during the bond cleavage, whereas in a cyclic diketopiperazine intermediate the difference between the first and second residues should be small. Influence of an amino acid in a specific position on bond cleavage can be evaluated through variance analysis [6]. The most important residue determining the frequency of a C–N bond cleavage is the one on its N-terminal side, with the same variance  $v = 0.030$  for both  $aa_2$  in  $y_{m-2}^+$  and  $aa_3$  in  $y_{m-3}^+$ . For  $n > 2$ , the effect of the next preceding amino acid is an order of magnitude smaller ( $v = 0.003$  for  $aa_2$  in  $y_{m-3}^+$ ), whereas for  $n = 2$  it is comparable ( $v = 0.014$  for  $aa_1$  in  $y_{m-2}^+$ ). The comparable variances mean that both N-terminal residues are involved in the ( $b_2^+$ ,  $y_{m-2}^+$ ) ion pair formation to a similar extent, as would be expected for diketopiperazine, but not for oxazolone (otherwise the variance for  $aa_2$  in  $y_{m-3}^+$  would also be large).

Even more telling than the variance analysis is the direct assessment of the effect of the amino acid nature

on the  $y_{m-2}^+$  formation probability. As would be expected for a symmetric  $b_2^+$ , the N-terminal sequences  $X_1Y_2$  and  $Y_1X_2$  tend to give similar frequencies of  $y_{m-2}^+$  formation (Figure 5, left). The overall cleavage propensities for  $aa_1$  (integrated rows) strongly correlate ( $r = +0.95$ ) with the overall propensities for  $aa_2$  (integrated columns). At the same time, sequences  $X_1Y_2Z_3$  and  $X_1Z_2Y_3$  gave strongly anti-correlating frequencies for  $aa_2$  and  $aa_3$  ( $r = -0.86$ ) in  $y_{m-3}^+$  formation (Figure 5, right), as expected for an asymmetric oxazolone intermediate. In  $y_{m-4}^+$  formation, the two N-terminal to the cleavage site amino acids  $aa_3$  and  $aa_4$  also behave in an anti-symmetric way.

### Special Cases

A clearly special case is when the first N-terminal amino acid is lysine or arginine. Sequestering the N-terminal proton by the side chain of either of these two very basic residues derails the mechanism in Scheme 1. Among the 1197 entries in the full CAD database with  $Arg_1$  and  $Lys_1$ , the most abundant species are  $y_{m-1}^+$  ions, although for the less basic lysine the dominance of  $y_{m-1}^+$  ions is not overwhelming (322 cases versus 215 cases of dominant  $y_{m-2}^+$  ions).

Gly<sub>1</sub>Gly<sub>2</sub> may be another special case [23, 24]. Based on Figure 5, such a combination is the least favorable for the formation of  $b_2^+/y_{m-2}^+$  ions. Yet even this combination frequently gives rise to class 1 behavior for short, doubly charged peptides. It is worth mentioning that the mechanism of fast *trans* to *cis* isomerization of non-prolyl peptide bonds has been suggested by Hamelberg and McCammon for Gly–Gly bonds [14]. In extensive molecular dynamics simulations, they have found that the stress induced inside a  $\beta$ -hairpin loop forces Gly–Gly bond isomerization to *cis* configuration on a microsecond to millisecond timescale. To study the details of *cis/trans* isomerization mechanism, they carried out simulations on linear and cyclic peptides with the same sequence and calculated the free-energy profile along the Gly–Gly bond. The activation energy for isomerization in linear peptides, 18 kcal/mol, has been found greatly reduced by mechanical stress in cyclic

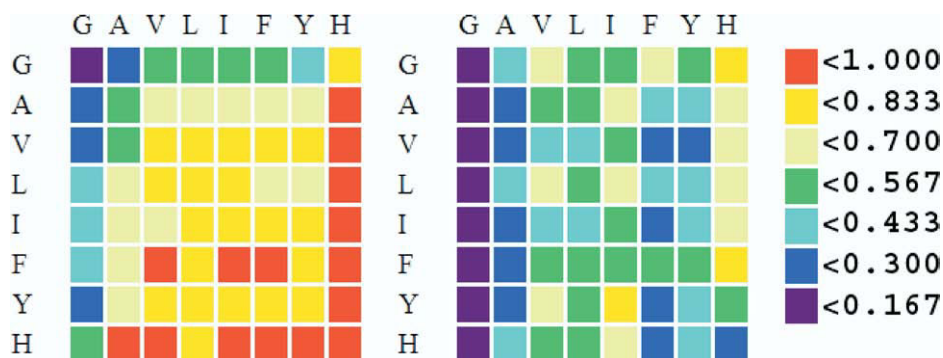
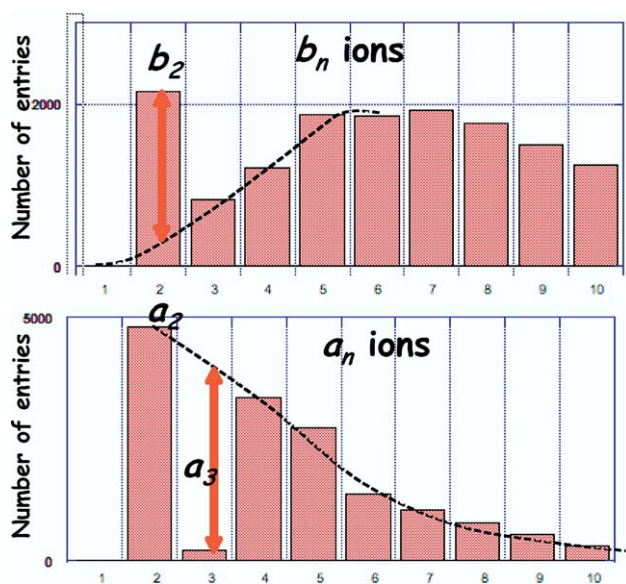


Figure 5. Frequencies of cleavage dominance (left,  $y_{m-2}^+$ ; right,  $y_{m-3}^+$ ) as a function of amino acid sequence N-terminal to the cleavage site. Columns represent the adjacent amino acid; rows, the preceding amino acid.

peptides. As a result, the calculated rate of isomerization in cyclic peptides has been found to increase by five orders of magnitude compared to linear peptides. Thus we cannot exclude that even Gly<sub>1</sub>Gly<sub>2</sub> combination leads in some cases to protonated diketopiperazine formation.

### *a*<sub>3</sub> Anomaly

In the instrument used, low-mass discrimination during CAD fragmentation as well as mass discrimination during the ion transfer to the FT analyzer decimates the detected signal from *b*<sub>2</sub> ions. Yet analysis of the occurrence frequency of these species (Figure 6, top panel) clearly shows an enhanced abundance compared to the expected value. Although this result closely matches the behavior of complementary *y* ions and thus was expected, the behavior of *a*<sub>3</sub> ions (Figure 6, bottom panel) came as a complete surprise. The general trend of abundance increase as the size decreases is understood in terms of lesser relative stability of smaller oxazolone *b*<sub>*n*</sub> ions, *n* > 2, from which *a*<sub>*n*</sub> ions arise through the CO loss. In view of that trend, nearly complete absence of *a*<sub>3</sub> species was staggering. There is no known reason that *a* ions would not form if their size was exactly three residues. The most logical explanation of the observed enormous deficit is thus not the low rate of formation but the fast decay rate of these species to *b*<sub>2</sub> ions. The reason for such a facile decay can be a great stability of *b*<sub>2</sub> species, and the reason for that can be their cyclic, symmetric nature as opposed to the cyclic asymmetric oxazolone structure of larger *b* ions. Thus we consider the deficit of *a*<sub>3</sub> ions as an additional argument for the cyclic diketopiperazine nature of *b*<sub>2</sub> ions.



**Figure 6.** Frequencies of cleavage dominance among *b* ions (top panel) and *a* ions (bottom panel). The excess of *b*<sub>2</sub> ions and deficit of *a*<sub>3</sub> ions compared to the dominant trends are shown by an arrow.

### Intermediate-Order Structure?

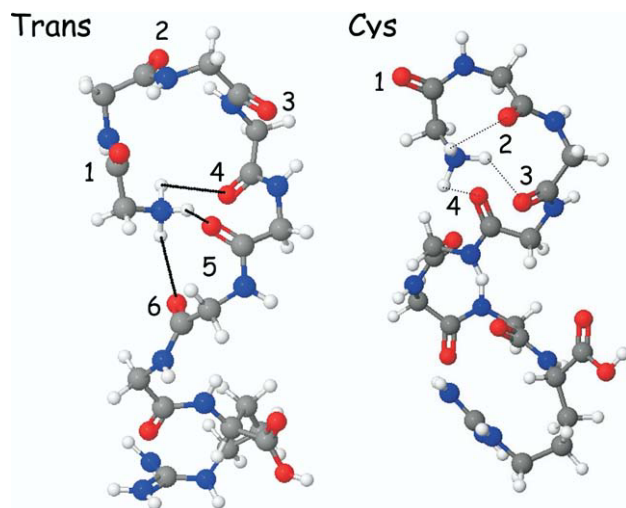
The primary structure of polypeptides, i.e., their sequence, is sensitive to isomerization. Indeed, amino acids Leu and Ile are constitutional isomers that can in principle interconvert, and Gly + Gly or Gly + Ala combinations can, at least in theory, isomerize to Asn or Gln, respectively. In general, however, single-bond isomerization (e.g., chiral isomerization of L- to D-amino acids, or  $\alpha$ -Asp conversion into  $\beta$ -Asp) is not considered a primary structure change. Under typical CAD conditions, such isomers behave very similarly. By slightly simplifying the situation and ignoring disulfide bonds, one can say that subcritical low-energy collisional excitation does not change the primary structure of a polypeptide as long as its mass does not change. However, the same regime of activation can completely destroy the secondary structure. This reactivity difference is a result of the strong nature of the covalent bonds with their high barriers for isomerization and dissociation, and the weak nature of the hydrogen bonding determining the secondary structure. Therefore, CAD can be viewed as a probe of the structural order, with first-order structures resistant to low-level, subcritical collisional activation, and second-order structures destroyed by it.

Intriguingly, the *cis-trans* isomerization of the first peptide bond does not fit into this Boolean picture. Indeed, according to Scheme 1 and the above-cited research such isomerization proceeds *before* bond fragmentation, i.e., under subcritical activation. Yet such isomerization is a persistent phenomenon (once isomerized, the bond can keep its conformation for a long time) and has a profound effect on peptide fragmentation pattern. Thus it is tempting to suggest that such isomerization determines a structure of an intermediate order.

As an example of the latter suggestion, consider local minimum-energy structures of the doubly protonated Gly<sub>7</sub>Arg peptide optimized by a force-field method (Scheme 2). When the first peptide bond is in the original *trans* configuration (left structure), the N-terminal proton is shared with carbonyls of residues 4 and 5. In such a state the peptide can be collisionally activated, which will result in acquiring a multitude of polypeptide chain conformations followed by proton mobilization and eventual peptide bond cleavage near the original solvation sites, i.e., class 2 behavior. Alternatively, if the peptide bond is isomerized to the *cis* state, the N-terminal proton-bound loop is much more compact and the proton is preferentially solvated on carbonyls 2, 3, and to some extent 4. Activation of such a structure will also cause it to acquire a multitude of conformations, but these will be different to those of the *trans*-isomer. Very soon, however, the N-terminal loop is bound to form an intermediate structure as in Scheme 1, with ejection of a protonated diketopiperazine.

There is no evidence suggesting that, once isomerized into *cis* configuration, the first (non-prolyl) peptide





Scheme 2

bond would easily return back to the *trans* form. If such reversed isomerization were a fast process, two ion populations would mix and no stark division between the two statistical groups would exist. Facile formation of  $b_2$  ions is the “escape route” for *cis* forms; reverse isomerization must be entropically disfavored. The presence of separate, largely isolated ion populations is the argument for considering *cis* and *trans* isomers as different structures of intermediate order.

## Conclusions

We have presented evidence that CAD mass spectra of tryptic peptide dications show bifurcating behavior. Class 1 spectra are dominated by  $b_2/y_{n-2}$  cleavage, whereas class 2 spectra show a broad distribution of cleavages preferentially peaking at residue 4 or 5 from the N-terminus. Such a bifurcating behavior is hypothesized to be linked to the *trans*-to-*cis* isomerization of the first peptide bond and subsequent formation of  $b_2$  ions in the form of cyclic, symmetrical diketopiperazines. Our data clearly show preference of shorter peptides for the class 1 behavior. Several effects contribute to this preference. In shorter dications, coulombic repulsion between the N- and C-terminal charges reduces the size of the loop formed by the protonated N-terminal amine solvated onto backbone carbonyls. In a smaller loop, an enhanced stress in the peptide bonds reduces the barrier for *cis*-*trans* isomerization that thus occurs faster. Additionally, the shorter distance between the charges increases the basicity of the N-terminus and the balance between the loss of a neutral and protonated diketopiperazine shifts toward charge separation.

The balance between isomerization before fragmentation (class 1) and direct fragmentation (class 2) must also depend on the ionic temperature after excitation, i.e., the type of the tandem mass spectrometer used. In linear ion traps, one third of all MS/MS spectra of

tryptic peptide dications are found to belong to class 1; in other instruments this fraction may be smaller or larger. In this work we have argued for the plausibility of the isomerization hypothesis and provided supporting arguments. However, we recognize that these arguments may not be enough to resolve the issue once and for all. At any rate, realization of the bifurcating fragmentation behavior of peptides must have an impact on the efforts to create a quantitative model for CAD fragmentation.

## Acknowledgments

This work was supported by the Knut and Alice Wallenberg Foundation, European Union (grant to consortium EPITOPE) as well as the Swedish research council (Grants 621-2004-4897 and 621-2003-4877 to RZ).

## References

- Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20*, 3551–3567.
- Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
- Lam, H.; Deutsch, E. W.; Eddes, J. S.; Eng, J. K.; King, N.; Stein, S. E.; Aebersold, R. Development and Validation of a Spectral Library Searching Method for Peptide Identification from MS/MS. *Proteomics* **2007**, *7*, 655–667.
- Zhang, Z. Q. Prediction of Low-energy Collision-induced Dissociation Spectra of Peptides. *Anal. Chem.* **2004**, *76*, 3908–3922.
- Zhang, Z. Q. Prediction of Low-energy Collision-induced Dissociation Spectra of Peptides with Three or More Charges. *Anal. Chem.* **2005**, *77*, 6364–6373.
- Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Garbuzynskiy, S. O.; Galzitskaya, O. V.; Surin, A. K.; Zubarev, R. A. Backbone Carbonyl Group Basicities Are Related to Gas-phase Fragmentation of Peptides and Protein Folding. *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 1481–1484.
- Paizs, B.; Suhai, S. Fragmentation Pathways of Protonated Peptides. *Mass Spectrom. Rev.* **2005**, *24*, 508–548.
- Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Brechi, L. A. Special Feature: Commentary—Mobile and Localized Protons: A Framework for Understanding Peptide Dissociation. *J. Mass Spectrom.* **2000**, *35*, 1399–1406.
- Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. Improving Protein Identification Using Complementary Fragmentation Techniques in Fourier Transform Mass Spectrometry. *Mol. Cell. Proteomics* **2005**, *4*, 835–845.
- Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A. New Data Base-independent, Sequence Tag-based Scoring of Peptide MS/MS Data Validates Mowse Scores, Recovers below Threshold Data, Singles Out Modified Peptides, and Assesses the Quality of MS/MS Techniques. *Mol. Cell. Proteomics* **2005**, *4*, 1180–1188.
- Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A. Complementary Sequence Preferences of Electron-Capture Dissociation and Vibrational Excitation in Fragmentation of Polypeptide Polycations. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 5301–5303.
- Shelimov, K. B.; Jarrold, M. F. Conformations, Unfolding, and Refolding of Apomyoglobin in Vacuum: An Activation Barrier for Gas-Phase Protein Folding. *J. Am. Chem. Soc.* **1997**, *119*, 2987–2994.
- Burlet, O.; Yang, C. Y.; Gaskell, S. J. Influence of Cysteine to Cysteine Acid Oxidation on the Collision-Activated Decomposition of Protonated Peptides - Evidence for Intraionic Interactions. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 337–344.
- Hamelberg, D.; McCammon, J. A. Fast Peptidyl *cis*-*trans* Isomerization within the Flexible Gly-rich Flaps of HIV-1 Protease. *J. Am. Chem. Soc.* **2005**, *127*, 13778–13779.
- Samuelson, S.; Martyna, G. J. Computer Simulation Studies of Finite Temperature Conformational Equilibrium in Alanine-based Peptides. *J. Phys. Chem. B* **1999**, *103*, 1752–1766.
- Paizs, B.; Suhai, S. Combined Quantum Chemical and RRKM Modeling of the Main Fragmentation Pathways of Protonated GGG. I. *Cis*-*trans* Isomerization Around Protonated Amide Bonds. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2307–2323.
- Smith, L. L.; Herrmann, K. A.; Wysocki, V. H. Investigation of Gas Phase Ion Structure for Proline-Containing  $b(2)$  Ion. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 20–28.



18. Farrugia, J. M.; O'Hair, R. A. J.; Reid, G. E. Do All b(2) Ions Have Oxazolone Structures? Multistage Mass Spectrometry and Ab Initio Studies on Protonated N-Acyl Amino Acid Methyl Ester Model Systems. *Int. J. Mass Spectrom.* **2001**, *210*, 71–87.
19. Counterman, A. E.; Clemmer, D. E. Anhydrous Polyproline Helices and Globules. *J. Phys. Chem. B* **2004**, *108*, 4885–4898.
20. Hunt, D. F.; Yates, Y. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R. Protein Sequencing by Tandem Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 6233–6237.
21. Huang, Y.; Triscari, J. M.; Wysocki, V. H.; Pasa-Tolic, L.; Anderson, G. A.; Lipton, M. S.; Smith, R. D. Dissociation Behaviors of Doubly-Charged Tryptic Peptides: Correlation of Gas-Phase Cleavage Abundance with Ramachandran Plots. *J. Am. Chem. Soc.* **2004**, *126*, 3034–3035.
22. Cordero, M. M.; Houser, J. J.; Wesdemiotis, C. The Neutral Products Formed during Backbone Fragmentations of Protonated Peptides in Tandem Mass Spectrometry. *Anal. Chem.* **1993**, *65*, 1594–1601.
23. Paizs, B.; Suhai, S. Combined Quantum Chemical and RRKM Modeling of the Main Fragmentation Pathways of Protonated GGG. II. Formation of b(2), y(1), and y(2) Ions. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 375–389.
24. Paizs, B.; Suhai, S. Towards Understanding the Tandem Mass Spectra of Protonated Oligopeptides. 1: Mechanism of Amide Bond Cleavage. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 103–113.