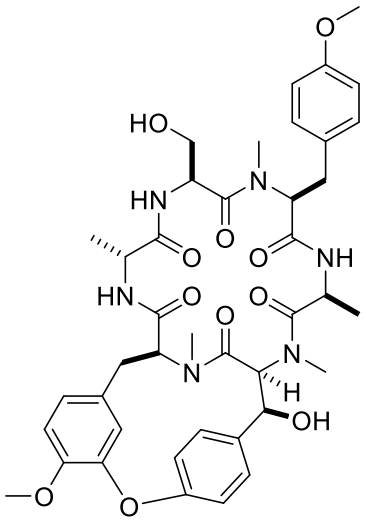
Problem Session

2020. 10.10.

Aoi Takeuchi

Cyclopeptide **1** was isolated from roots of *Galianthe thalictroides*. Major fragment ions of **1** observed in ESI-MS/MS analysis were listed in **Table 1**. Please provide structures of fragment ions listed in **Table 1** and describe possible mechanisms of fragmentation.



1

exact mass: 802.3538

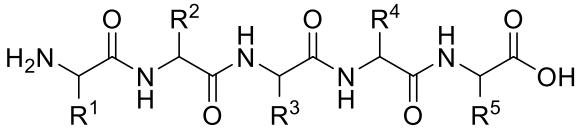
Table 1. major fragment ions in ESI-MS/MS analysis

m/z	fragment ions
825.3448	[**1**+Na]$^+$
803.3634	[**1**+H]$^+$
785.3556	[**1**+H-H_2O]$^+$
714.3138	
643.2819	
541.2301	
523.2172	
470.1916	
454.1978	
452.1810	
383.1591	
355.1640	

Note: The fragmentation was carried out in low-energy CID (collision-induced dissociation), so satellite ions like *d*-, *v*-, or *w*- ions were not generated.

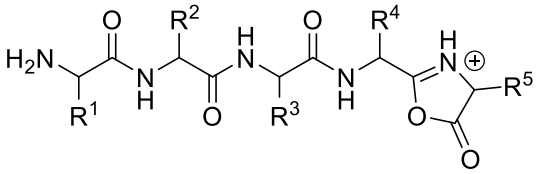
Supplementary Note:

It is often the case that CID of b_n ion (n≥5) leads to anomalous sequence fragment ions that cannot directly be derived from the original peptide structure (**S1**). The linear b_n ion with a C-terminal oxazolone ring (**S2**) is attacked by the N-terminal amino group to form a cyclic peptide b_n isomer (**S3**). The cyclic intermediate undergoes various proton transfer reactions, then get cleaved to form other fragment ions (**S2'**) leading to scrambling of sequence information.



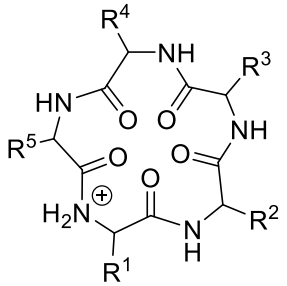
S1

fragmentation pathway
(*reaction mechanism was omitted*)



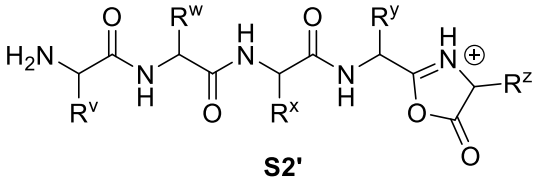
S2

cyclization



S3

ring opening



S2'

(v, w, x, y, z) = (1, 2, 3, 4, 5)
(5, 1, 2, 3, 4)
(4, 5, 1, 2, 3)
(3, 4, 5, 1, 2)
(2, 3, 4, 5, 1)
} **sequence scrambling**

For details, see also;

Harrison, A. G.; Young, A. B.; Bleiholder, C.; Suhai, S.; Paizs, B. *J. Am. Chem. Soc.* **2006**, *128*, 10364.

Problem Session Answer

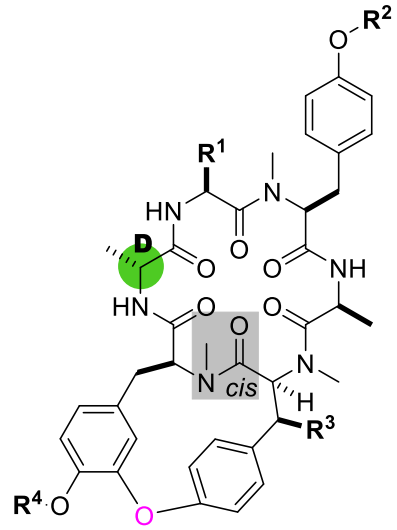
2020. 10.10.

Topic: *de novo* MS/MS sequencing of cyclic peptides

Aoi Takeuchi

Contents:

1. Rubiaceae-type cyclopeptide (RA)



isolation[1]
from *Rubia* species (**1**: from roots of *Galianthe thalictroides*)
structure
homodicyclohexapeptide with one D-Ala
14-membered ring fused to 18-membered cyclic hexapeptide, formed by a phenolic oxygen linkage with a cis-peptide bond
activity
potent antitumor activity (**1**: GI_{50} = 0.26 mg/mL against PC-3)

common structure in Rubiaceae-type cyclopeptides[2,3] (RAs)
5β-hydroxy-RA-III (**1**): **R**^{1} = CH_2OH, **R**^{2} = CH_3, **R**^{3} = OH, **R**^{4} = CH_3

The structure of **1** was determined by ^{1}H and ^{13}C NMR, and HRESI-MS/MS data. The absolute configurations of residues were characterized by applying Marfey's method[1,4].

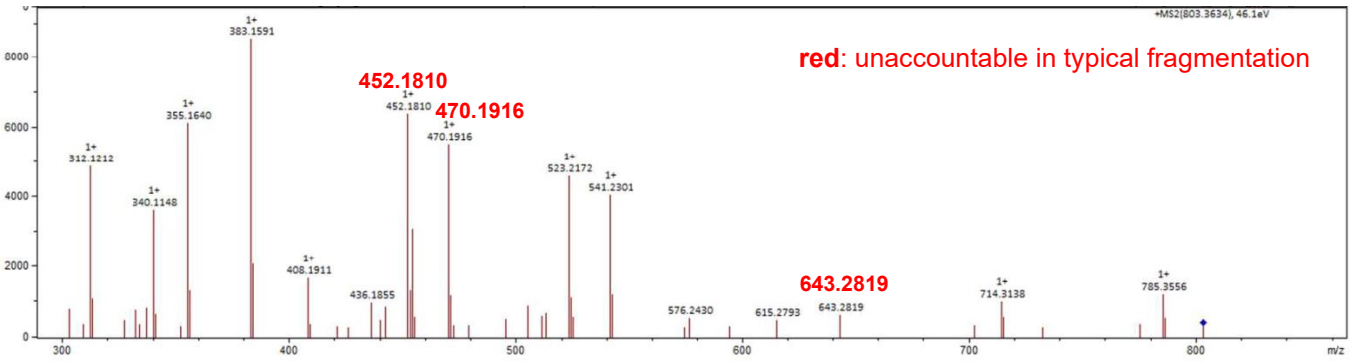


Figure 1. HRESI-MS/MS spectrum of **1**[1]

In general, tandem mass spectrometry (MS/MS) is an attractive method for structure elucidation of peptidic natural products as it can access to peptide sequence information from picograms of non-purified material[5,6].

In the case of cyclopeptides, it is sometimes complicated to assign of MS/MS spectrum as their propensity to break at all pairs of points in their cyclic backbone gives a far more complex series of ions than in linear peptides[7,8]. Also, it is possible that yielding fragment ions undergo sequence scrambling[9] (**Supplementary Notes**, colored in red in **Figure 1**), sometimes making the resulting MS/MS spectrum more complex. (**Figure 2**)

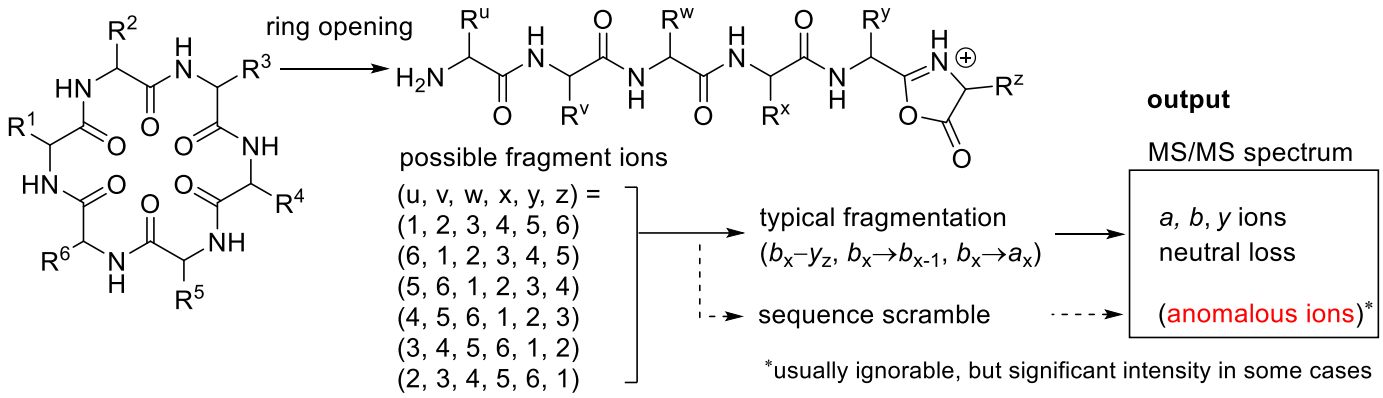


Figure 2. general complexity in fragmentation caused by multiple linearization patterns and sequence scramble

2. Nomenclature for peptidic fragment ions

- linear peptide

The universally accepted nomenclature for linear peptide fragment ions was first proposed by Roepstorff and Fohlman[10], and subsequently modified by Johnson and Biemann[11].

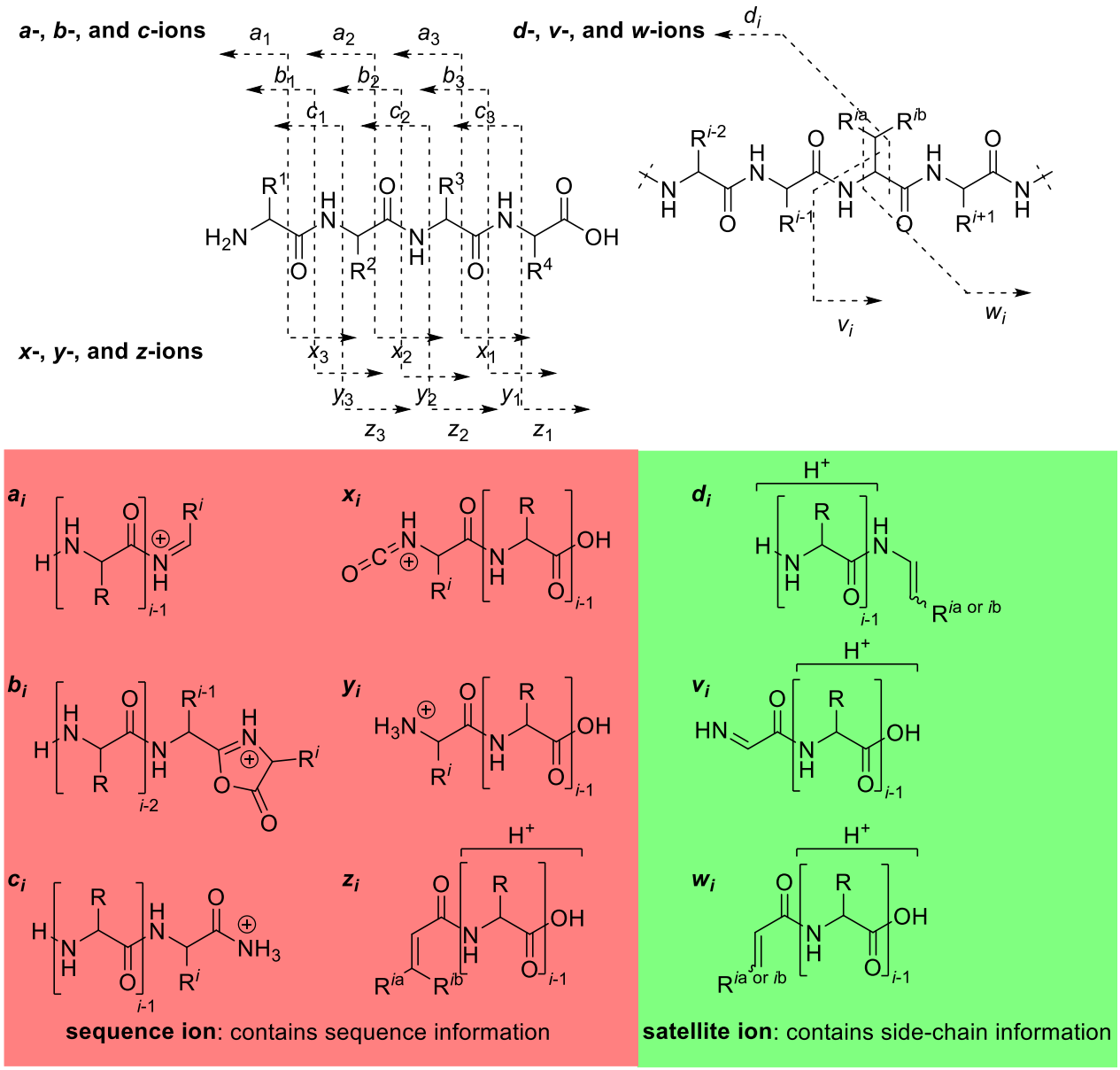


Figure 3. nomenclature for fragments of linear peptide ions

In low energy CID, cleavage of carbonyl C-N dominantly occurs to give *b*- and *y*- ions while charge-remote fragmentation at C_α-carbonyl C gives *a*- and *x*-ions in high energy CID[12]. *a*- Ions can also be generated from the degradation of *b*- ions. *c*- And *z*- ions dominantly appears at electron transfer dissociation (ETD)[13] and electron capture dissociation (ECD)[14] spectrum. Formation of satellite ions require high collision energy and/or radical processes. Thus, they can be observed in high energy CID and in some ECD/ETD experiments[12].

To sum up, *a*-, *b*-, and *y*- ions are dominantly observed in low energy CID.

- cyclic peptide

Nomenclature for cyclic peptides was proposed by Ngoka ang Gross[15]. A fragment ion is labeled with the four-part descriptor $X_{nJZ.}$ *X* stands for the ion and n is the number of amino acid residues in accordance with nomenclature of linear peptides. The symbol J and Z describe N- and C- terminal amino acid residues in one letter each, respectively, uniquely defining the cleaved bond when ring opening.

In the case amino acid residues are descripted with a suffix like A_1, or A_4, J and Z are lettered in a normal size with a space after X_n instead of subscripts[1,16]. Ex) $b_{5AW} \rightarrow b_5\ A_1W_5$

- fragment ions with scrambled sequences

Chawner et al. proposed an extension to the nomenclature explained above so that fragment ions deriving from macrocycle formation can be easily assigned[17]. To describe a scrambled fragment ion, a fragment ion to be scrambled is lettered inside a bracket, following the amide bond number at which ring opening occurs. Any ions resulting from further fragmentation are assigned outside the bracket. The amide bond number is given starting from the bond formed by N-terminal nucleophilic attack of the oxazolone ring as 0 and in ascending order toward C-terminal.

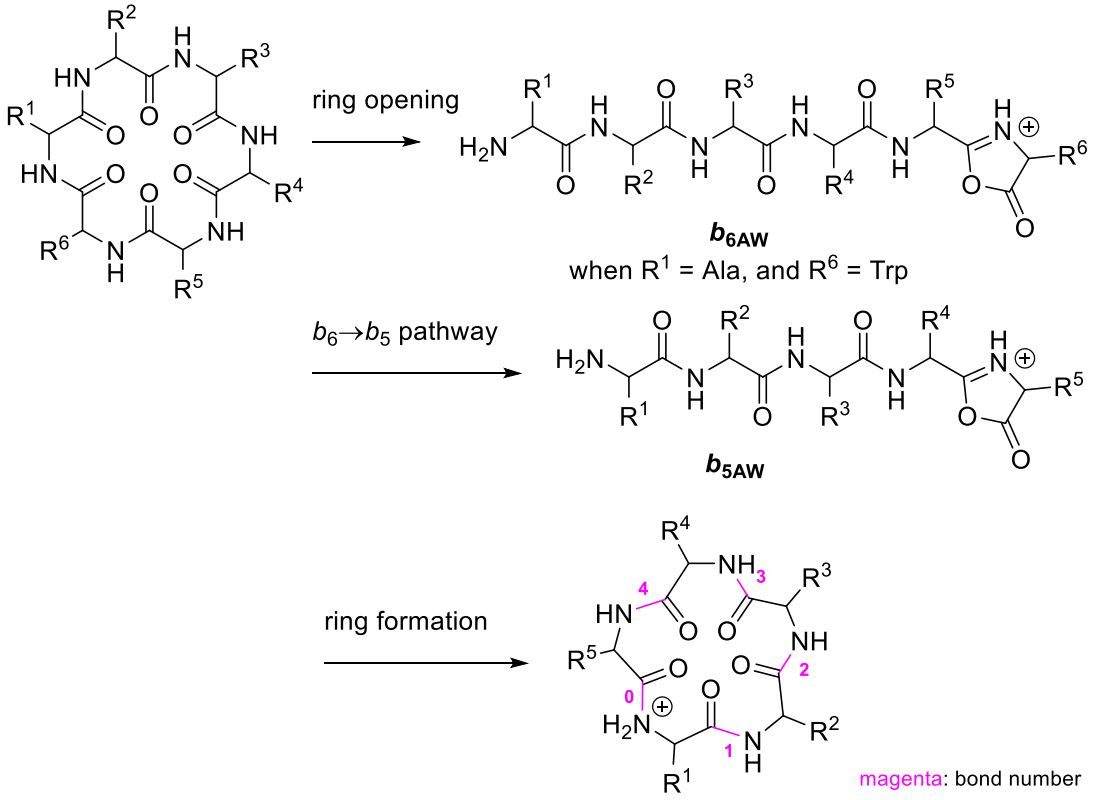


Figure 4. nomenclature for fragments of cyclic, and scrambled peptide ions

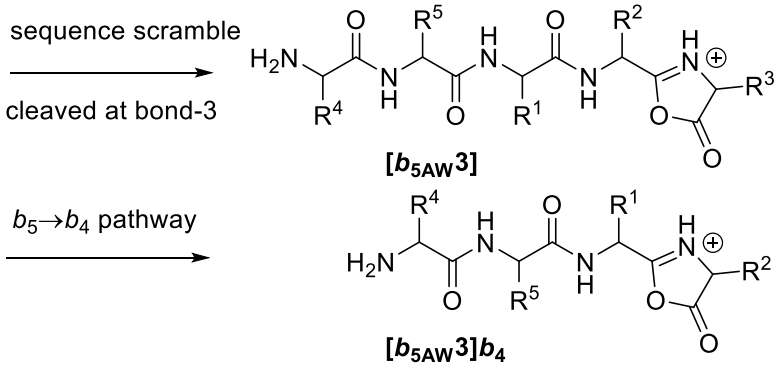


Figure 4 (continue). nomenclature for fragments of cyclic, and scrambled peptide ions

3. Reaction mechanisms of fragmentation

In low energy CID, peptides undergo three types of dominant fragmentation pathway[18]. First, ring opening takes place via the b_x-y_z pathway with oxazolone formation, resulting in a linear peptide ion having a free N-terminus and an oxazolone ring at the C-terminus[19]. The resulting linear ion can originate a smaller *b* fragment via $b_x \rightarrow b_{x-1}$ pathway[20,21], which in turn loses CO via the $b_x \rightarrow a_x$ pathway[19,22].

The b_x-y_z pathway is explained in "mobile proton model". In this model, fragmentation requires the transfer of a proton from a basic site to the amide nitrogen[23]. (for rationale of this model, see 190914_PS_Hiroaki_Itoh[24])

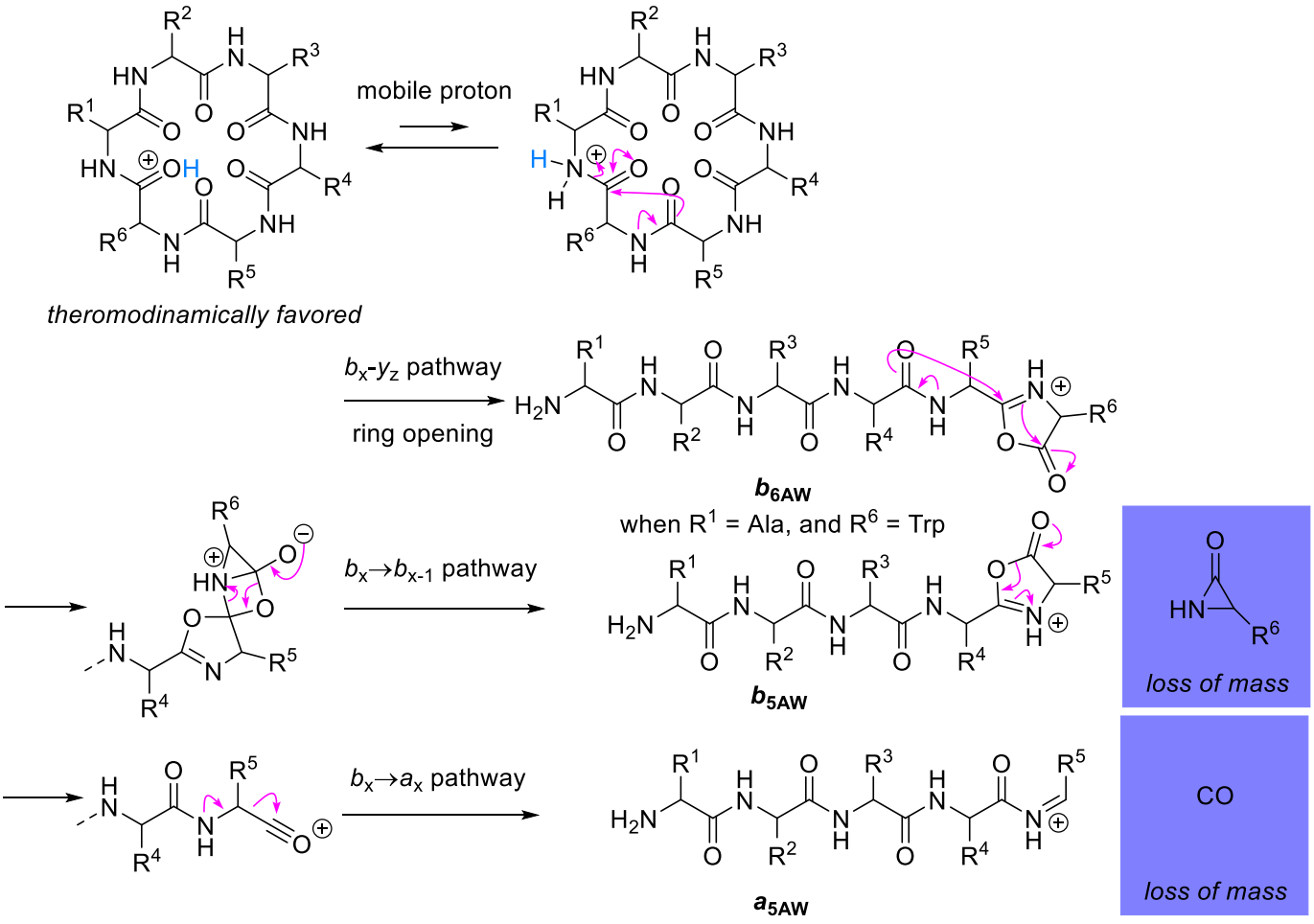


Figure 5. mechanisms of fragmentations of a cyclic peptide in low-energy CID.

A particularly common loss of H_2O occurs for protonated peptides containing a serine or threonine residue where there is a side-chain hydroxyl group[25]. Ser residue in N-terminal is likely to undergo dehydration in neighboring group participation, while dehydration of Ser residue in the peptide sequence is likely to proceed in *cis* 1,2 elimination[26,27]. Dehydroalanine (ΔAla, **3**), oxazoline (**5**), or aziridine (**6**) are possible products of dehydration (**Figure 6**). (Dehydroalanine products are suggested for structures of fragment ions in solutions of this session.)

Rationale of mechanisms of dehydration of Ser residue[26]

(i) MS^3 intensity patterns of dehydrated ions suggested generation of **6** from Ser, and **3** from Ac-Ser. (**Table 2**.)

(ii) H/D exchange experiment indicated dehydration of Ser proceeded in neighboring group participation, while that of Ac-Ser proceeded both in neighboring group participation, and *cis* 1,2 elimination. (**Table 3**.)

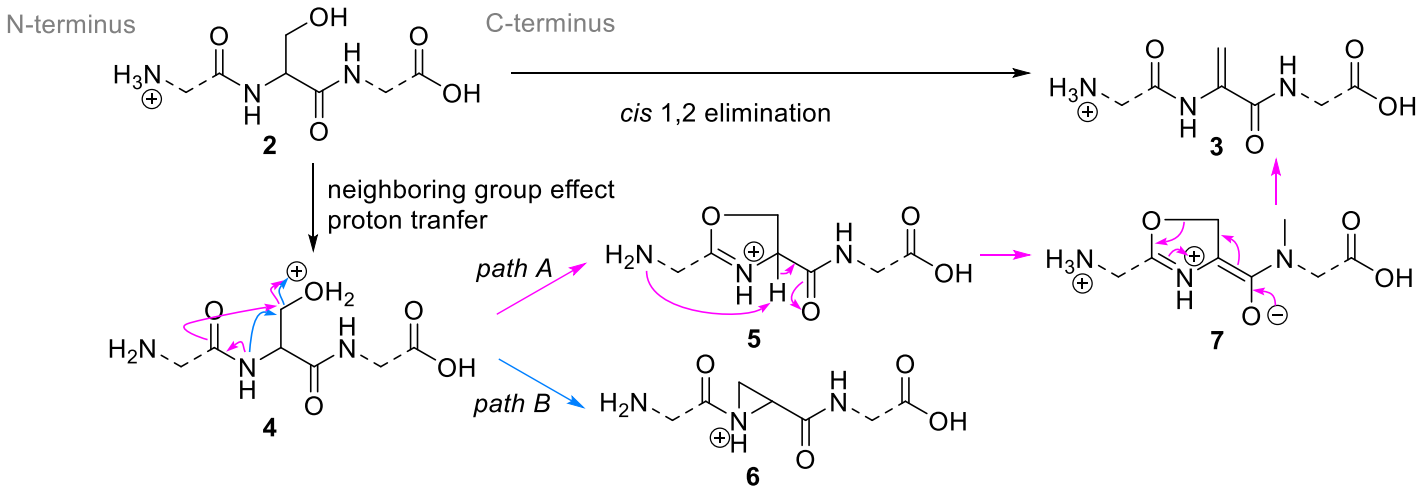


Figure 6. mechanisms of dehydration of Ser residue

Table 2. CID MS^3 spectra of side chain loss MS/MS product ions

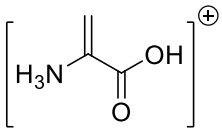
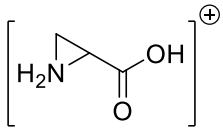
amino acid	MS^3 product ions (relative abundance)
$[Ser+H]^+$-H_2O	$[M-H_2O]$ (100), [M-CO] (4), $[M-H_2O-CO]$ (2)
$[Ac-Ser+H]^+$-H_2O	$[M-H_2O]$ (4), [M-CO] (32), $[M-H_2O-CO]$ (100)
	$[M-H_2O]$ (13), [M-CO] (9), $[M-H_2O-CO]$ (100)
	$[M-H_2O]$ (100), [M-CO] (8), $[M-H_2O-CO]$ (2)

Table 3. CID MS/MS spectra of fully deuterated serine and N-acetyl serine

amino acid	MS/MS product ions (relative abundance)
$[Ser+D]^+$	$[M-D_2O]$ (28), $[M-D_2O-CO]$ (100)
$[Ac-Ser+D]^+$	[M-HOD] (51), $[M-D_2O]$ (100), $[M-D_2O-CO]$ (1)

4. Solution for the problem

For sake of simplicity, the structure of [**1**+H]$^+$ is described as a 6-membered sequence information. (**Figure 7**)

Five possible reactions in low-energy CID are listed in **Table 4**. They provide two types of information, order of sequence of fragment ions, and mass numbers of leaving components from fragment ions. The latter can be accountable from mass numbers of fragment ions in **Table 1**, the only information we get from MS/MS analysis.

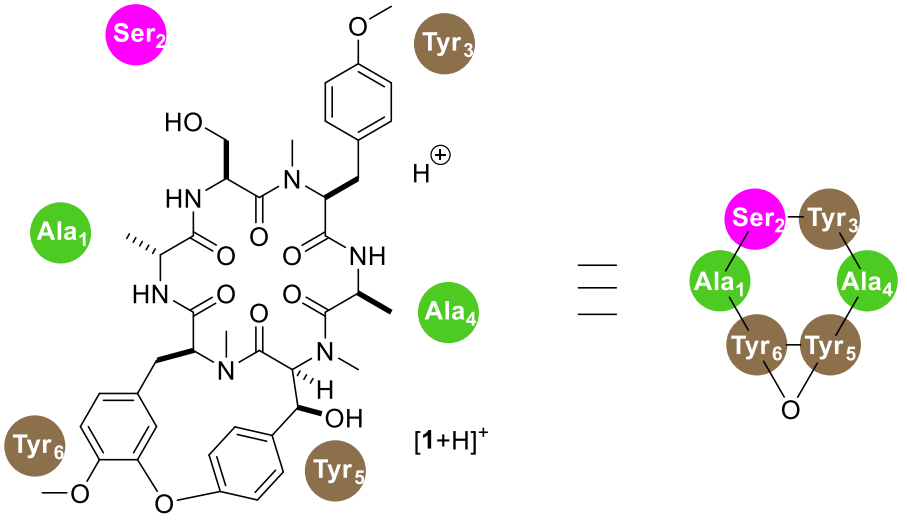


Figure 7. simplified sequence information of [**1**+H]$^+$

Table 4. possible reactions in low-energy CID

CID reaction	*providing information*	*accountability*
b_x-y_z	order of sequence	not directly accountable from **Table 1**
sequence scramble		
$b_x \rightarrow b_{x-1}$	leaving component	accountable from **Table 1**
$b_x \rightarrow a_x$		
dehydration		

Losses of masses of all fragment ions listed in **Table 1** are derived from either the $b_x \rightarrow b_{x-1}$ pathway, the $b_x \rightarrow a_x$ pathway, or dehydration. Only 5 mass number patterns are possible to be lost in the processes (**Figure 8**). This means all fragment ions in **Table 1** can be bridged with the 5 patterns. If a difference between mass numbers of two fragment ions is 71, it is probable that these two ions are b_x and b_{x-1}, and the leaving residue is Ala. All mass numbers in **Table 1** are thus arranged into two fragmentation pathways shown in **Figure 9**, starting from [**1**+H]$^+$, or [**1**+H-H_2O]$^+$.

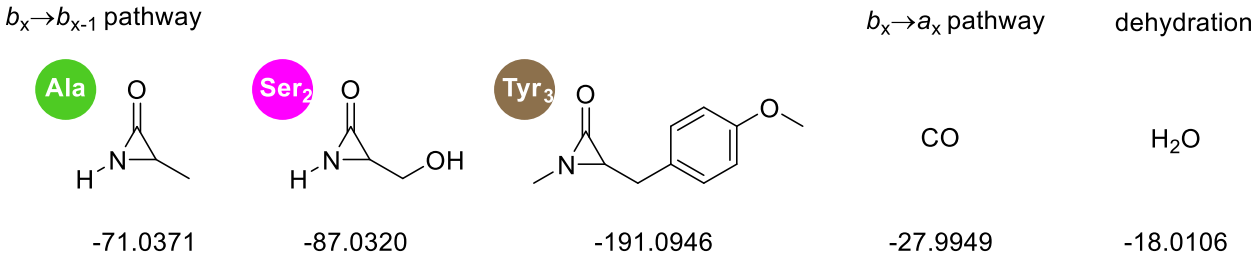


Figure 8. possible leaving fragments with exact masses during low energy CID fragmentation processes of [**1**+H]$^+$

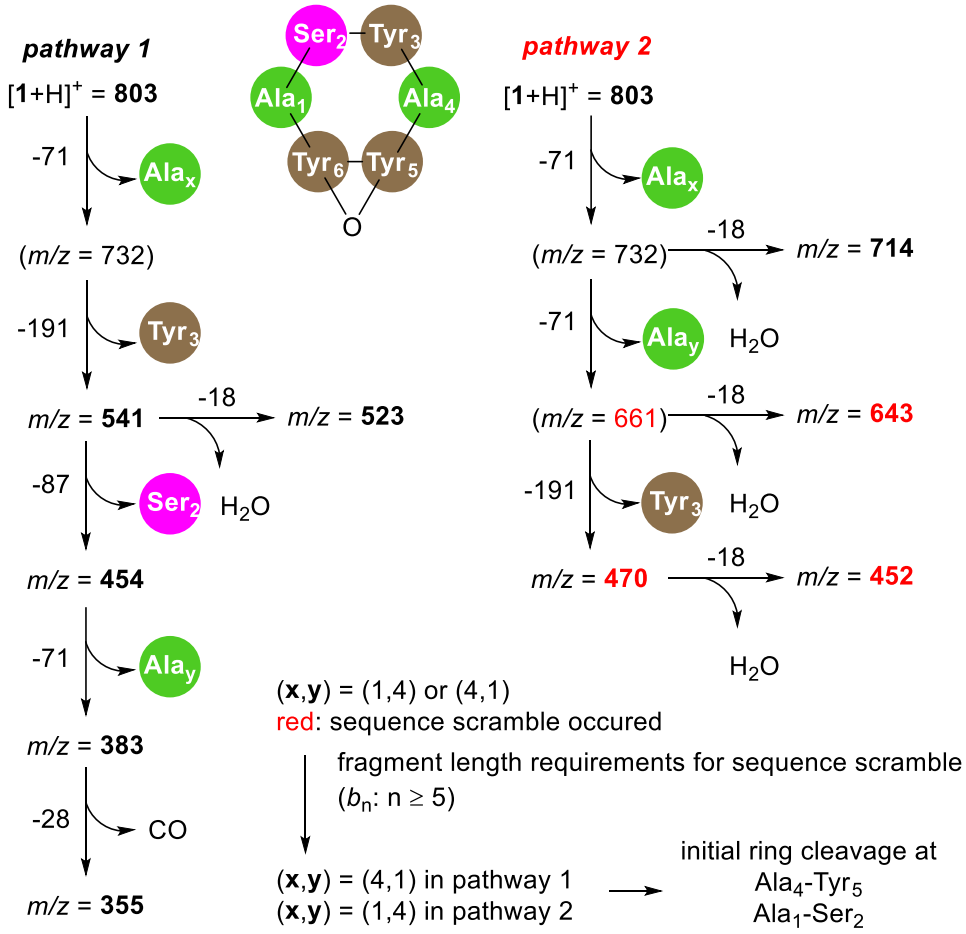


Figure 9. possible fragmentation pathways of [**1**+H]$^+$ (mass numbers are displayed in an integer)

It is necessary to assume sequence scramble to explain successive leaving of two Ala in pathway 2. As sequence scramble is reported to potentially occur in b_n ion (n≥5) so far[28], only possible combination of (**x**,**y**) is (1,4) in the case of pathway 2. In the same way, (**x**,**y**) is uniquely determined as (4,1) in the case of pathway 1 to avoid sequence scrambling of too short fragment ions. Thus, initial *b*-*y* pathway to lead ring opening occurs at Ala_4-Tyr_5 in the case of pathway 1, and Ala_1-Ser_2 in the case of pathway 2.

Fragmentation pathway and structures of fragment ions of pathway 1, and 2 are displayed in **Figure 10**, **11**, respectively.

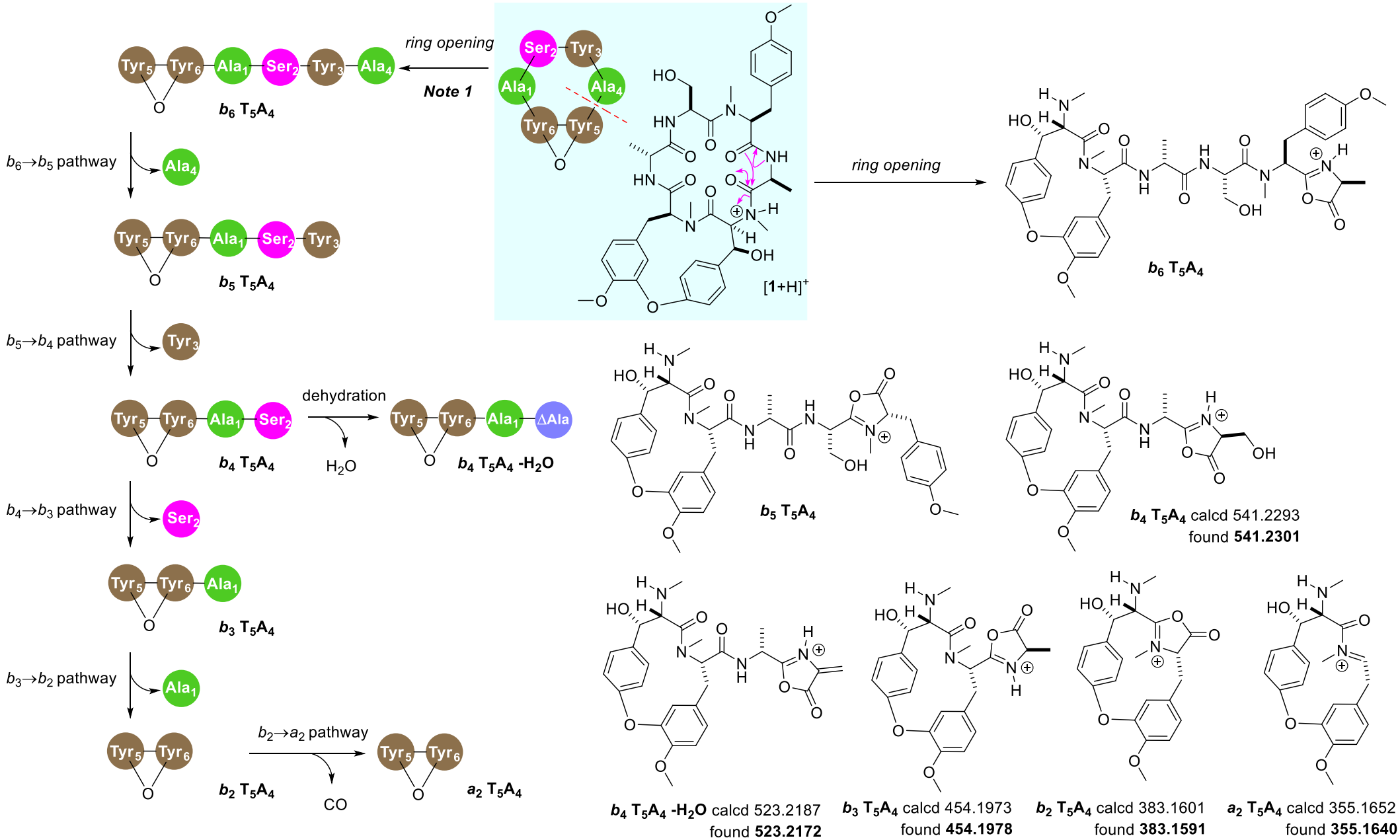


Figure 10. fragmentation pathway 1, and structures of fragment ions

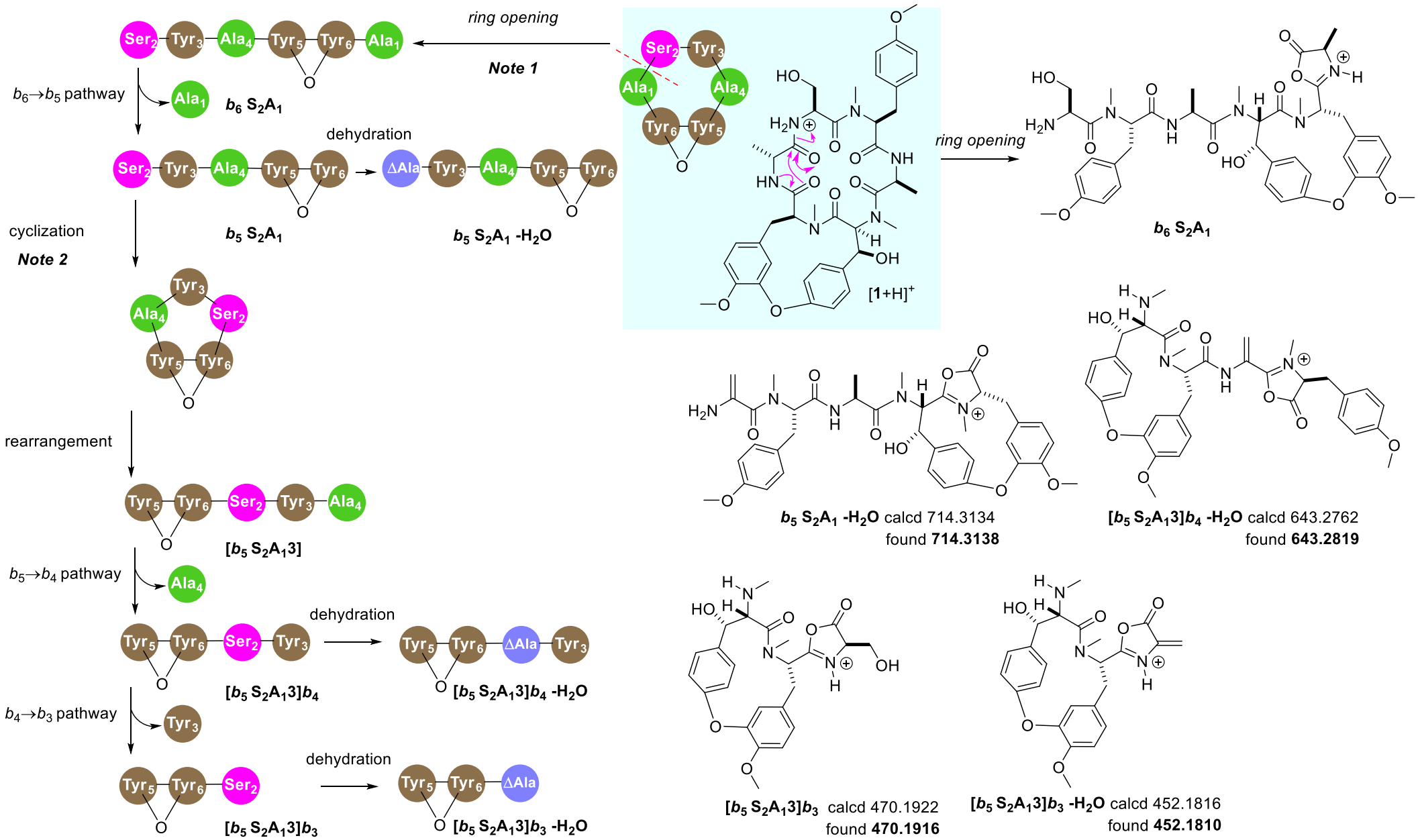


Figure 11. fragmentation pathway 2, and structures of fragment ions

Note 1. trends of site-selectivity of ring opening

Bond cleavage can potentially occur at any amide bonds and the fragility or stability of an individual peptide bond is dependent upon the amino acid residues flanking it. Statistical study of CID fragmentation of 1,465 tryptic peptides exhibited residual bond cleavage trends in b_x-y_z pathway[28], which is supporting that ring opening of **1** occurred at C-terminal amide bond of Ala.

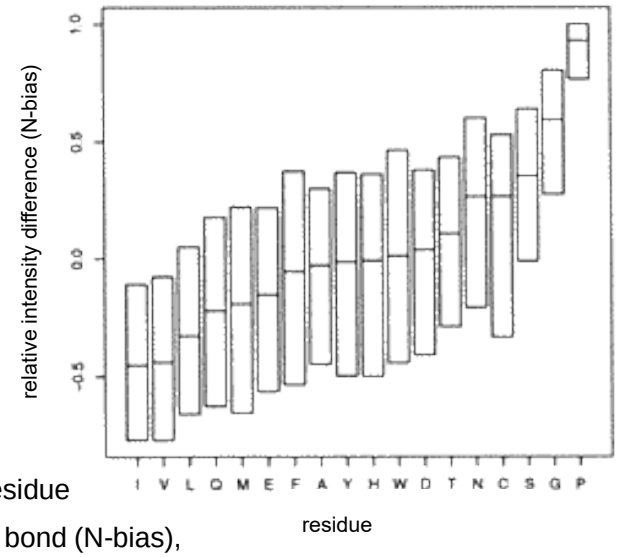


Figure 12 (right). the extent to which each residue directionally enhances cleavage at its N-terminal amide bond (N-bias), which is calculated by subtracting intensity of the C-terminal fragment peak from that of N terminal

Note 2. reaction mechanism and rationale of sequence scramble

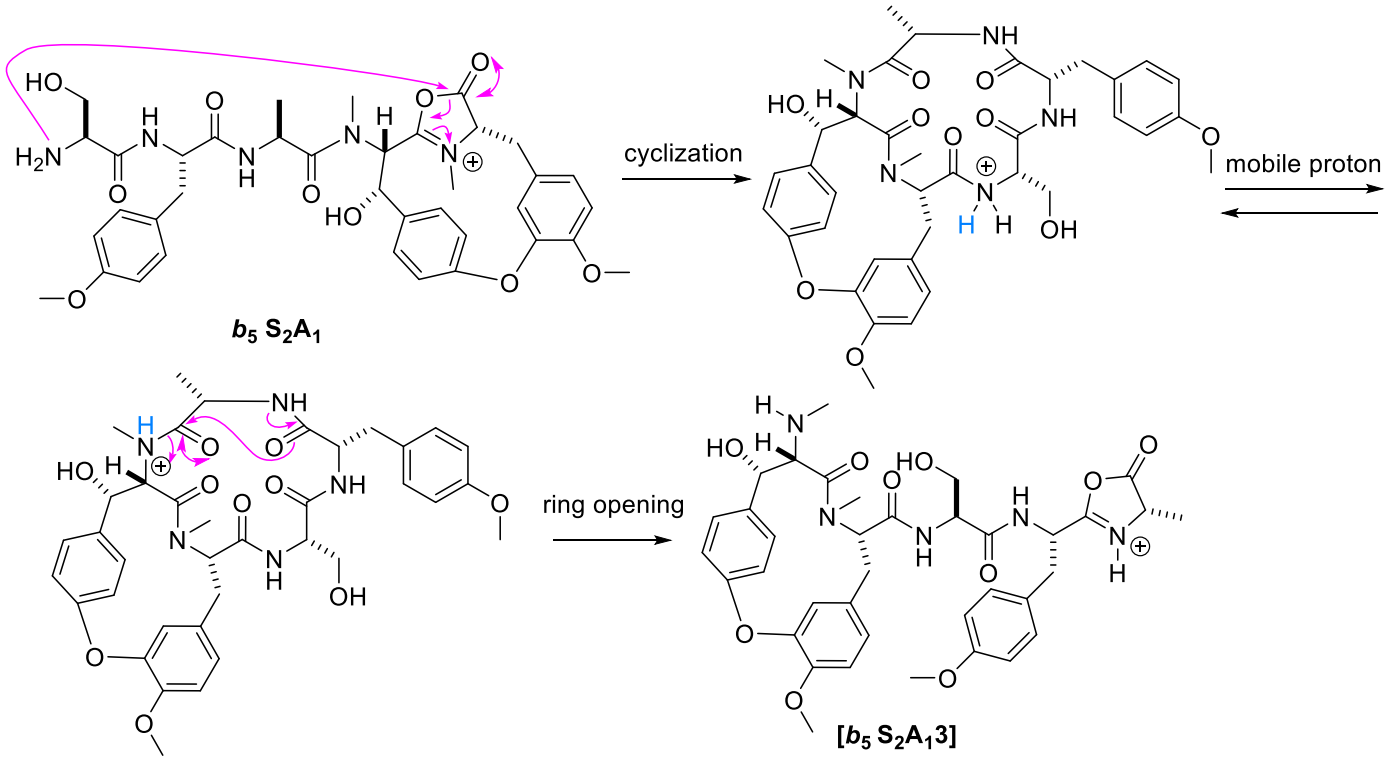


Figure 13. reaction mechanism of sequence scramble in pathway 2.

(i) MS^3 analysis: several anomalous di-, tri-, tetrapeptide were expulsed from b_5 ion of oligopeptide[29]

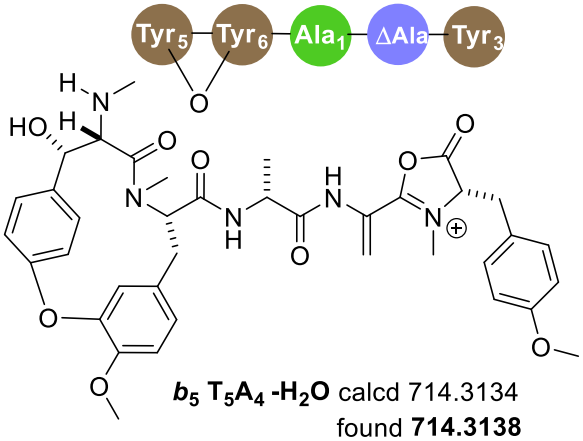
$$\textbf{AQVEL}PY \xrightarrow{\text{extract } b_5 \text{ for } MS^3} b_5\text{-}\textbf{V},\ b_5\text{-}\textbf{Q},\ b_5\text{-}(\textbf{VE}),\ b_5\text{-}(\textbf{QVE})$$

(ii) response to variable collisional energy: breakdown graph for the b_5 ion of linear hexapeptide YAGFL was similar to that of corresponding cyclopentapeptide[9].

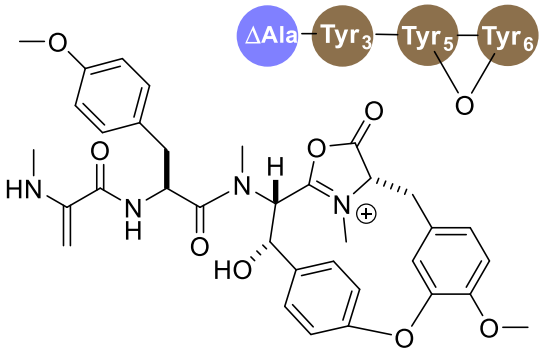
(iii) N-acetylation: acetyl-capping of N-terminal prevented sequence scrambling[30].

Alternative answers at *m/z* = 714.3138, 643.2819, 541.2301, and 454.1978

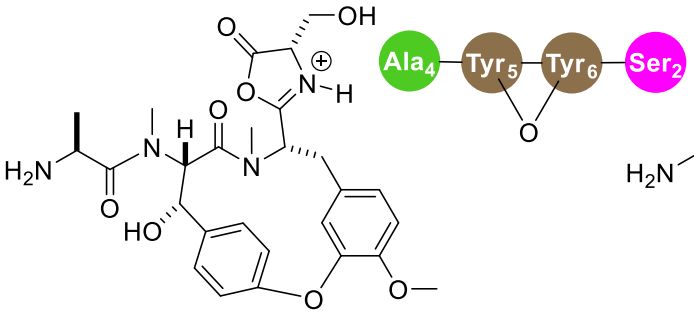
It also gives corresponding mass numbers to assume that the combination of (**x**,**y**) in **Figure 9** gets reversed. Note that fragmentation pathways of these fragments do not account for the entire ion series shown in **Table 1**.



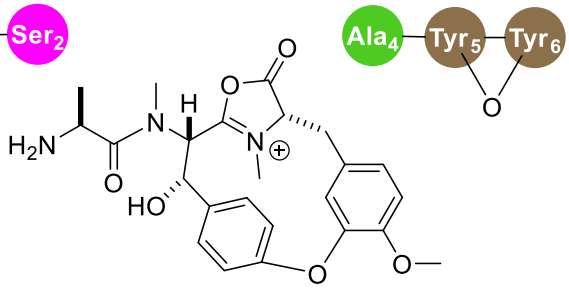
***b*$_5$ T$_5$A$_4$ -H$_2$O** calcd 714.3134
found **714.3138**



[*b*$_5$ T$_5$A$_4$3]*b*$_4$ -H$_2$O calcd 643.2762
found **643.2819**



[*b*$_5$ S$_2$A$_1$3]*b*$_4$ calcd 541.2293
found **541.2301**



***b*$_3$ A$_4$T$_3$** calcd 454.1973
found **454.1978**

Table 1 (completed). major fragment ions in ESI-MS/MS analysis

m/z	fragment ions
825.3448	[**1**+Na]$^+$
803.3634	[**1**+H]$^+$
785.3556	[**1**+H-H$_2$O]$^+$
714.3138	*b*$_5$ S$_2$A$_1$ -H$_2$O; *b*$_5$ T$_5$A$_4$ -H$_2$O
643.2819	[*b*$_5$ S$_2$A$_1$3]*b*$_4$ -H$_2$O; [*b*$_5$ T$_5$A$_4$3]*b*$_4$ -H$_2$O
541.2301	*b*$_4$ T$_5$A$_4$; [*b*$_5$ S$_2$A$_1$3]*b*$_4$
523.2172	*b*$_4$ T$_5$A$_4$ -H$_2$O
470.1916	[*b*$_5$ S$_2$A$_1$3]*b*$_3$
454.1978	*b*$_3$ T$_5$A$_4$; *b*$_3$ A$_4$T$_3$
452.1810	[*b*$_5$ S$_2$A$_1$3]*b*$_3$ -H$_2$O
383.1591	*b*$_2$ T$_5$A$_4$
355.1640	*a*$_2$ T$_5$A$_4$

References

1. Patrícia, O. F.; Maria, F. C. M.; Renata, T. P.; Wilson, H. K.; Marcos, V. G. O. B.; Fernanda, R. G.; Walmir, S. G. *J. Nat. Prod.* **2016,** *79*, 1165.
2. Tan, N.-H.; Zhou, J. *Chem. Rev.* **2006**, *106*, 840.
3. Xuejia, Z.; Qirui B.; Xingdong, W.; Zhe W.; Yuanyuan M.; Ninghua, T. *J. Chromatogr. A* **2018,** *1581-1582*, 43.
4. Bhushan, R.; Brückner, H. *Amino Acids* **2004**, *27*, 231.
5. Li, J.; Vederas, J. *Science* **2009**, *325*, 161.
6. Mohimani, H.; Liu, W.-T.; Mylne, J. S.; Poth, A. G.; Colgrave, M. L.; Tran, D.; Selsted, M. E.; Dorrestein, P. C.; Pevzner, P. A. *J. Proteome Res.* **2011**, *10*, 4505.
7. Tomer, K. B.; Crow, F. W.; Gross, M. L.; Kopple, K. D. *Anal. Chem.* **1984**, *56*, 880.
8. Eckart, K.; Schwarz, H.; Tomer, K. B.; Gross, M. L. *J. Am. Chem. Soc.* **1985**, *107*, 6765.
9. Harrison, A. G.; Young, A. B.; Bleiholder, C.; Suhai, S.; Paizs, B. *J. Am. Chem. Soc.* **2006**, *128*, 10364.
10. Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*, 601.
11. Johnson, R. S.; Martin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T. *Anal. Chem.* **1987**, *59*, 2621.
12. Medzihradszky, K. F.; Chalkley, R. J. *Mass Spectrom. Rev.* **2015**, *34*, 43.
13. Han, H.; Xia, Y.; McLuckey, S. A. *J. Proteome Res.* **2007**, *6*, 3062.
14. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265.
15. Ngoka, L. C. M.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 360.
16. Priscila, M. C. -B.; Patrícia, O. F.; Maria, F. C. M.; Fernanda, R. G.; Walmir, S. G. *J. Nat. Prod.* **2015,** *78*, 2754.
17. Chawner, R.; Gaskell, S. J.; Eyers, C. E. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 205.
18. Jegorov, A.; Paizs, B.; Žabka, M.; Kuzma, M.; Giannakopulos, A, E.; Derrick, P, J.; Havlíček, V. *Eur. J. Mass. Spectrom.* **2003**, *9*, 105.
19. Paizs, B.; Suhai, S. *Mass Spectrom. Rev.* **2005**, *24*, 508.
20. Harrison, A. G.; Young, A. B.; Bleiholder, C.; Suhai, S.; Paizs, B. *J. Am. Chem. Soc.* **2006**, *128*, 10364.
21. Yalcin, T.; Harrison, A. G. *J. Mass. Spectrom.* **1996**, *31*, 1237.
22. Yalcin, T.; Khouw, C.; Csizmadia, I. G.; Peterson, M. R.; Harrison, A. G. *J. Am. Soc. Mass. Spectrom.* **1995**, *6*, 1165.
23. Tsaprailis, G.; Nair, H.; Somogyi, A.; Wysocki, V. H.; Zhong, W. Q.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. *J. Am. Chem. Soc.* **1999**, *121*, 5142.
24. 190914_PS_Hiroaki_Itoh
25. Neta, P.; Pu, Q.-L.; Yang, X.; Stein, S. E. *Int. J. Mass Spectrom.* **2007**, *267*, 295.
26. Reid, G. E.; Simpson, R. J.; O'Hair, R. A. J. *J. Am. Soc. Spectrom.* **2000**, *11*, 1047.
27. DeGnore, J. P.; Qin J. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 1175.
28. Tabb, D. L.; Smith, L. L.; Breci, L. A.; Wysocki, V. H.; Lin, D.; Yates, J. R. *Anal. Chem.* **2003**, *75*, 1155.
29. L. Mouls.; Aubagnac, J. L.; Martinez. J.; Enjalbal, C. *J. Proteome Res.* **2007**, *6*, 1378.
30. Harrison, A. G. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1776.