Elucidating the Structure of Cyclotides by Partial Acid Hydrolysis and LC–MS/MS Analysis

ARTICLE *in* ANALYTICAL CHEMISTRY · MARCH 2009

Impact Factor: 5.64 · DOI: 10.1021/ac802175r · Source: PubMed

Article

Elucidating the Structure of Cyclotides by Partial Acid Hydrolysis and LC#MS/MS Analysis

Siu Kwan Sze, Wei Wang, Wei Meng, Randong Yuan, Tiannan Guo, Yi Zhu, and James P. Tam

Anal. Chem., **2009**, 81 (3), 1079-1088 • DOI: 10.1021/ac802175r • Publication Date (Web): 12 January 2009

Anal. Chem. **2009,** *81,* 1079–1088

Elucidating the Structure of Cyclotides by Partial Acid Hydrolysis and LC–MS/MS Analysis

Siu Kwan Sze,*,† Wei Wang,† Wei Meng,† Randong Yuan,† Tiannan Guo,† Yi Zhu,† and James P. Tam*,†,‡

School of Biological Sciences, Nanyang Technological University, Singapore, and Department of Chemistry, Scripps Research Institute, Jupiter, Florida

We describe here a rapid method to determine the cyclic structure and disulfide linkages of highly stable cyclotides via a combination of flash partial acid hydrolysis, LC–MS/MS, and computational tools. Briefly, a mixture of closely related cyclotides, kalata B1 and varv A purified from *Viola yedoensis* was partially hydrolyzed in 2 M HCl for 5 min by microwave-assisted hydrolysis or for 30 min in an autoclave oven (121 °C and 15 psi). The partially hydrolyzed peptide mixture was then subjected to LC–MS/MS analysis, with the disulfide linked-peptides fragmented by collision activated dissociation (CAD). A computer program written in-house (available for download at http://proteomics.sbs.ntu.edu.sg/cyclotide_SS) was used for interpreting LC–MS/MS spectra and assigning the disulfide bonds. Time-point analysis of single-disulfide fragments revealed that nonrandom acid catalyzed fragmentation mostly occurred at the turns which are solvent-exposed and often contain side chain functionalized amino acids such as Asx/Glx and Ser/Thr. In particular, the most susceptible bond for acid hydrolysis in kalata B1 and varv A was found to be the highly conserved N25–G26 which is also the head-to-tail ligation site of the linear precursor proteins, indicating that formation of the three disulfide bonds might precede cyclic structure closure by N25–G26 ligation. This observation is consistent with the recent report that the N25–G26 bond formation is the last step in the cyclotide biosynthetic pathway. The process demonstrated here can potentially be a high throughput method that is generally applicable to determine disulfide bonds of other relatively low-abundance cyclotides.

Cyclotides belong to a family of heat-stable and enzyme-resistant macrocyclic peptides with three intertwined disulfide bonds forming a cyclic cysteine knot (CCK).[1] They are widely distributed in plants and have recently attracted attention due to their high heat stability, resistance to enzymatic degradation, and possibility of being orally bioavailable.[2] These properties are highly desirable in engineering biologics for pharmaceutical applications. Cyclotides possess a diverse range of bioactivities, ranging from antipest, antiviral, antifungal, antibacterial to anticancer functions. These functions are associated with cytotoxity and suggest that cyclotides may play an important role in plant defense, although their physiological roles in plants remain to be determined. Nevertheless, understanding the structure–function relationship in cyclotides will facilitate the rational modification of their structures for therapeutic applications.

Since the discovery of kalata B1 and B2 in *Oldenlandia affinis* in 1973,[3] more than 95 cyclotides have been identified in the plant families of *Rubiaceae*, *Violaceae*, and *Cucurbitaceae*.[4] Recent studies showed that an individual plant can produce up to 100 different cyclotides, and there may be over 9000 different cyclotide sequences present in nature.[5] However, the abundance of cyclotides in a plant may vary by several orders of magnitude, and moreover, expression levels of these cyclotides are known to vary seasonally within a species.[6,7] As a result, only the relatively high-abundance cyclotides have been purified in sufficient quantities for detailed structural characterization. Of the 95 known cyclotides, many have their 3D structures determined by NMR. However, their disulfide bonds remain controversial because crowding in the cyclotide core area as a result of three disulfide (SS) bonds creates difficulties in characterizing the bonds.[8–10] So far, only three cyclotides have their disulfide bonds directly determined by biochemical methods.[11–14] Thus, there is a need to develop a rapid

* Corresponding authors. Siu Kwan Sze and James P. Tam, School of Biological Sciences, Nanyang Technological University, Singapore 637551. E-mail: SKSze@ntu.edu.sg (S.K.S.); JPTam@ntu.edu.sg (J.P.T.).

† Nanyang Technological University.

‡ Scripps Research Institute.

(1) Gruber, C. W.; Cemazar, M.; Anderson, M. A.; Craik, D. J. *Toxicon* **2007**, *49* (4), 561–575.

(2) Gillon, A. D.; Saska, I.; Jennings, C. V.; Guarino, R. F.; Craik, D. J.; Anderson, M. A. *Plant J.* **2008**, *53* (3), 505–515.

(3) Gran, L. *Lloydia* **1973**, *36* (2), 174–178.

(4) Wang, C. K.; Kaas, Q.; Chiche, L.; Craik, D. J. *Nucleic Acids Res.* **2008**, *36* (Database issue), D206-10.

(5) Simonsen, S. M.; Sando, L.; Ireland, D. C.; Colgrave, M. L.; Bharathi, R.; Goransson, U.; Craik, D. J. *Plant Cell* **2005**, *17* (11), 3176–3189.

(6) Trabi, M.; Svangard, E.; Herrmann, A.; Goransson, U.; Claeson, P.; Craik, D. J.; Bohlin, L. *J. Nat. Prod.* **2004**, *67* (5), 806–810.

(7) Ireland, D. C.; Colgrave, M. L.; Craik, D. J. *Biochem. J.* **2006**, *400* (1), 1–12.

(8) Craik, D. J.; Daly, N. L. *Mol. Biosyst.* **2007**, *3* (4), 257–265.

(9) Shenkarev, Z. O.; Nadezhdin, K. D.; Sobol, V. A.; Sobol, A. G.; Skjeldal, L.; Arseniev, A. S. *FEBS J.* **2006**, *273* (12), 2658–2672.

(10) Skjeldal, L.; Gran, L.; Sletten, K.; Volkman, B. F. *Arch. Biochem. Biophys.* **2002**, *399* (2), 142–148.

(11) Derua, R.; Gustafson, K. R.; Pannell, L. K. *Biochem. Biophys. Res. Commun.* **1996**, *228* (2), 632–638.

(12) Tam, J. P.; Lu, Y. A. *Protein Sci.* **1998**, *7* (7), 1583–1592.

10.1021/ac802175r CCC: $40.75
Published on Web 01/12/2009

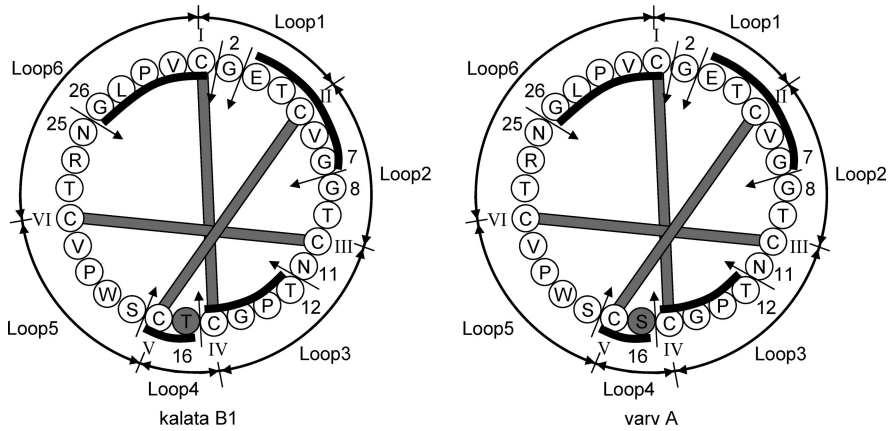

Figure 1. The amino acid sequences of kalata B1 (left) and varv A (right). The cyclotide varv A is identical to kalata B1 except the replacement of T16 with S16. The disulfide connectivity of kalata B1 is known to be Cys I−IV, II−V, and III−VI which is shared by varv A as determined by partial acid hydrolysis couple to LC−MS/MS methods (this work). The numbering of cyclotides sequences is according to the prevalent literature practice with Cys1. SS-connectivity is denoted by Roman numerals, and the loops 1−6 are the amino acid sequences flanked by two Cys in the circular permutated sequences.

alternative method for determining the disulfide bonds of low- and high-abundance cyclotides.

Cyclotide disulfide bonds are often determined via partial reduction and stepwise alkylation followed by mass spectrometry analysis.[13] Generally, this method involves multiple manipulation steps, requires a sufficient amount of starting material, and is consequently limited to highly abundant cyclotides. Partial acid hydrolysis coupled with molecular weight determination of the hydrolyzed fragments by mass spectrometry is an alternative method. However, the ultrastable cyclotides necessitate prolonged acid hydrolysis, and furthermore, many hydrolyzed fragments are isoforms of a single cyclotide and molecular weights by themselves cannot be used to unambiguously assign the disulfide bonds.[11] Recently, microwave-assisted acid hydrolysis coupled with partial reduction, alkylation, and HPLC fractionation have been used to fragment the kalata B2 cyclotide structure, followed by FTMS analysis.[14] The method used multiple sample preparation steps, FTMS instrumentation, and NMR data for cyclotide structure determination.

We report here a sensitive and potentially high-throughput method for cyclotide disulfide bonds characterization based on flash partial hydrolysis in 2 M HCl or 5% TFA aqueous solutions accelerated by microwave irradiation or autoclaving. In this study, we used a cyclotide mixture consisting of kalata B1 and varv A purified from *Viola yedoensis* as a model. The amino acid sequences of varv A are similar to kalata B1 except that S16 in varv A is replaced by T16 in kalata B1. This high sequence identity makes it difficult for the mixture to be separated by C18 HPLC chromatography or other analytical methods. The disulfide connectivity of varv A is not known but is likely similar to kalata B1, which has a CCK-type motif of Cys I−IV, II−V, and III−VI (Figure 1).[13] Here we first determined if our method is able to correctly identify the known disulfide bonds of kalata B1, and subsequently, we tried to identify the previously uncharacterized disulfide bonds of varv A to test and validate the robustness of our method in determining the disulfide connectivity of highly similar cyclotides that cannot be easily separated.

After partial acid hydrolysis of a mixture of varv A and kalata B1, the hydrolysis products were separated and analyzed by LC−MS/MS in a 60 min LC gradient. Peptide fragments were sequenced online by collision activated dissociation in the MS/MS mode. We then developed a computer program to interpret the MS/MS spectra of single disulfide-linked fragments and to correctly assign the peptides from possible isoforms, leading to unambiguous characterization of disulfide bonds. As the area of a chromatography peak is generally proportional to the abundance of hydrolysis products, which, in turn reflects the acid susceptibility of peptide bonds, the relative susceptibility of peptide bonds within a given cyclotide can be used to infer its structure, as well as provide predictive rules for cyclotide fragmentation under acidic conditions to determine SS connectivity.

EXPERIMENTAL METHODS

All chemicals were of the highest grade and were purchased from Sigma unless stated otherwise. Formic acid was from Fluka (Buchs, Switzerland). HPLC-grade water and acetonitrile were obtained from JT Baker (Philipsburg, NJ).

Cyclotides were isolated from the aerial parts of *Viola yedoensis* collected in Beijing, China. The plant material was homogenized and extracted three times in dichloromethane/methanol (1:1, v/v) at room temperature, and the combined extracts were partitioned by addition of water. Next, the aqueous (water/methanol) layer was collected and concentrated by a rotary evaporator prior to lyophilization, while the organic extracts was subjected to a solvent−solvent partitioning between water and hexane, ethyl acetate, and 1-butanol, respectively. The 1-butanol fraction containing the active components was purified by reversed-phase HPLC on a preparative GRACE Prosphere (Grace Davison Discovery Sciences, Deerfield) C18 column (250 mm × 22 mm, 5 μm, 300 Å) using an 80 min 5 mL/min gradient from 0 to 60% buffer B. Buffer A was 0.05% (v/v) trifluoroacetic acid (TFA) in MilliQ H_2O, and buffer B was 0.05% (v/v) TFA and 90% (v/v) acetonitrile (ACN) in MilliQ H_2O. Further purification was performed with a semipreparative GRACE VYDAC (Grace Davison Discovery

(13) Goransson, U.; Craik, D. J. *J. Biol. Chem.* **2003**, *278* (48), 48188–48196.
(14) Nair, S. S.; Romanuka, J.; Billeter, M.; Skjeldal, L.; Emmett, M. R.; Nilsson, C. L.; Marshall, A. G. *Biochim. Biophys. Acta* **2006**, *1764* (10), 1568–1576.

Sciences, Deerfield) Protein & Peptide C18 column (250 mm × 10 mm, 5 μm, 300 Å) and an analytical GRACE VYDAC Protein & Peptide C18 column (250 mm × 4.6 mm, 5 μm, 300 Å).

A fraction of the LC eluants was collected and analyzed with ABI 4800 MALDI-TOF/TOF-MS. Two cyclotides of molecular weights 2877.13 (~60%) and 2891.15 Da (~40%) were found to coelute from the C18 analytical column (Figure S-1 in the Supporting Information). These cyclotides were later confirmed to be varv A and kalata B1, respectively. Amino acid sequences of kalata B1 and varv A were shown in Figure 1.

About 800 μg of the cyclotide mixture was dissolved in 1 mL of HPLC water. A total of 1 μL of this solution was then mixed with 100 μL of 2 M HCl in a glass HPLC vial, which was then placed into a 50 mL Falcon tube with a 0.5 mm hole cut into the cap. Degassing was carried out via a 100 μm i.d. fused silica tubing inserted into the solution, through which a slow flow of helium gas at 100 mL/min was bubbled for 15 min and consequently filled the Falcon tube with residual helium gas. The 0.5 mm hole on the cap of the Falcon tube was sealed with parafilm, and the sample was subjected to partial acid hydrolysis immediately after degassing.

For microwave-assisted partial acid hydrolysis, a 900 W domestic microwave oven (Sanyo, model EM-G4753) was used. The cap of the Falcon tube was opened just before closing the oven door, and microwave energy was applied at full power for 1−5 min. After 5 min, the solution had completely been dried, and the vial was allowed to cool down to room temperature. The hydrolyzed product was stored at 4 °C until LC−MS/MS analysis.

For partial acid hydrolysis in an autoclave oven, the sealed Falcon tube was placed in a holder inside an autoclave oven that was used for heat sterilization. The oven was set to operate under normal sterilization conditions for 15−30 min. Upon removal from the oven, the sample was immediately chilled to 4 °C and dried in a vacuum concentrator. The dried sample was kept at 4 °C until LC−MS/MS analysis.

LC−MS/MS Analysis. The partially acid-hydrolyzed sample was reconstituted in 20 μL of 0.1% formic acid, injected by an autosampler, concentrated, and desalted utilizing an online peptide trap (Michrom Bioresources, Auburn, CA). Chromatographic separation was carried out with a HPLC system (Shimadzu, Kyoto, Japan) using a 75 μm i.d. C18 column packed directly into a pico-frit nanospray tip (New Objective, MA). The gradient was set at 200 nL/min and ramped from 5% ACN to 35% in 40 min, then to 80% ACN in 45 min, kept at 80% ACN for 3 min, and ramped back to 5% ACN. The eluate was ionized by online nanospray at 1800 V into the LTQ-Orbitrap or LTQ-FT-Ultra (ThermoFinnigan, Bremen, Germany). The MS scan was performed in the Orbitrap at 60 000 resolution or in the ICR cell at 100 000 resolution. The LTQ was operated in a data-dependent mode by performing MS/MS scans for seven of the most intense peaks from each MS scan. Dynamic exclusion was activated for 20 s.

Data Analysis. The extracted ion chromatogram (XIC) was integrated over the MS scan with Qual Browser version 2.0 (ThermoFinnigan, Bremen, Germany) with a bandwidth of 0.05 Da at the specific precursor ion m/z. Alternatively, the raw data file was converted to the ASCII text file with XRawFileTestVB (Xcalibur 2.0 SR2, ThermoFinnigan, Bremen, Germany), and the XIC was integrated at the specific precursor ion m/z. The chromatographic peak areas of the precursor ions were used to measure the relative abundance of the partial acid hydrolysis products. This abundance was related to the susceptibility of acid hydrolysis and solvent accessibility of the peptide bond. The NMR (1NB1) structure of kalata B1 was downloaded from a protein database (http://www.rcsb.org), and the solvent-accessible surface area of each residue was calculated with GETAREA 1.1[15] based on the NMR structure.

The MS/MS spectra (dta files) were extracted from the raw file by the extract_msn program packaged with Bioworks Browser 3.3 (ThermoFinnigan, Bremen, Germany). A home-written program (available for downloading at http://proteomics.sbs.ntu.edu.sg/cyclotide_SS) was used for subsequent data analysis and disulfide bond determination. First, the program read the cyclotide amino acid sequence and calculated all possible combinations of disulfide-linked peptides with singly and doubly charged ions (Tables S-2a and S-2b in the Supporting Information). Next, the $M + H^+$ ion mass and charge state of each MS/MS spectrum were read from the dta files. The experimental ion mass was then matched to the tabulated ion mass. If the mass difference was within 5 ppm, the theoretical CAD fragment of the disulfide-linked peptides was calculated and compared to the peak list of the MS/MS spectrum. If the MS/MS spectra peak list matched more than 20% of the theoretical CAD b, y ions within a 0.8 Da mass error, the spectra were manually inspected to confirm the assignment. Figure 2 shows the schematic diagram of the cyclotide disulfide bond characterization.

RESULTS AND DISCUSSION

Partial Acid Hydrolysis of Cyclotides. Determining the SS connectivity of a cyclic structure with three SS bonds such as cyclotides requires the generation of suitable single-SS-linked fragments. In turn, this requires six specific backbone fragmentations, each in one of the loops (loops 1−6, Figure 1) framed by two Cys in the circular permutation of cyclotides. Specific fragmentation by enzymatic digestion for determining disulfide connectivity is prone to rearrangement under alkaline conditions and is thus performed under mildly acidic conditions with an appropriate enzyme such as pepsin, followed by molecular weight measurement of pepsin-digested and SS-linked fragments.[16] However, this strategy does not work well for cyclotides because of the lack of multiple substrate sites and their tightly compacted structure that often resists enzymatic digestion.

Enzymatic fragmentation of cyclotides after kinetic-controlled partial disulfide reduction, stepwise alkylation, followed by mass spectrometry analysis had been reported.[13] This method usually requires a large amount of starting material because the partially reduced intermediates (1 SS- or 2 SS-reduced forms) are generated in limiting amounts relative to the starting material. This is due in part to the ultrastability of

(15) Fraczkiewicz, R.; Braun, W. J. *Comput. Chem.* **1998**, *19*, 319–333.
(16) Gorman, J. J.; Wallis, T. P.; Pitt, J. J. *Mass Spectrom. Rev.* **2002**, *21* (3), 183–216.

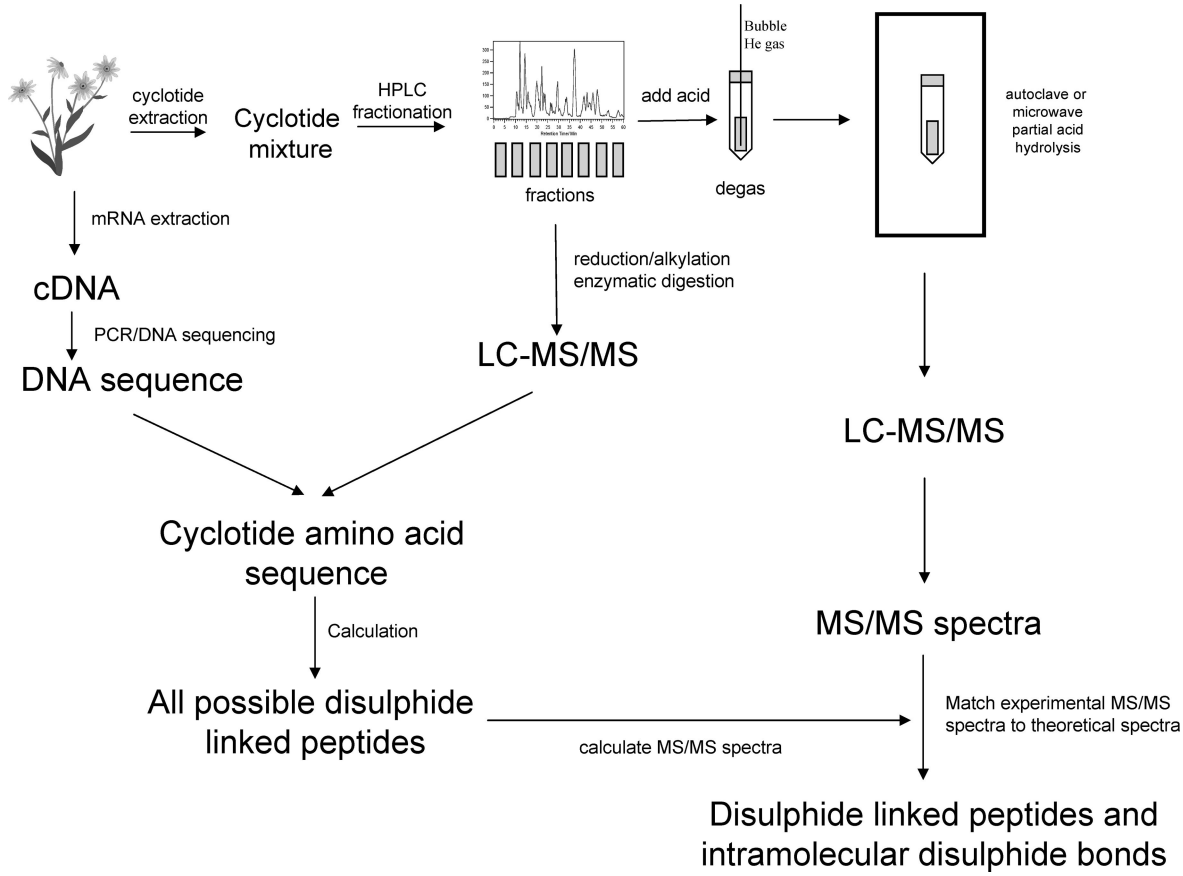

Figure 2. Schematic diagram of the characterization of disulfide bonds of cyclotides by partial acid hydrolysis.

CCK and the rapid complete reduction of partially reduced CCK intermediates (once the ultrastable CCK structure is destroyed by the first SS reduction). The multiple downstream manipulations by alkylation and enzymatic digestion further limit its sensitivity.

An alternative to enzymatic fragmentation is by chemical means using partial acid hydrolysis. This method traditionally requires prolonged hydrolysis under high temperature in a vacuum tube.[12] Recently, microwave-assisted digestion of proteins was used to accelerate or increase the selectivity of partial acid hydrolysis.[17−19] In contrast to the traditional enzymatic digestion, microwave-assisted partial acid hydrolysis can be performed in a detergent and chaotrope-free environment with significantly reduced hydrolysis time. Consequently, the hydrolysis products can be directly analyzed with mass spectrometry without extra peptide purification steps.

We therefore explored the use of microwave-assisted partial acid hydrolysis to fragment cyclotides, thus obtaining single-SS-linked peptides suitable for MS determination. Our initial attempts to perform cyclotide hydrolysis under microwave irradiation with weak acids such as formic or oxalic acid yielded insufficiently usable fragments for data interpretation (data not shown). Instead, 2 M HCl or 5% TFA were found to be efficient for hydrolysis of our cyclotide sample.

Figure 3 shows base peak chromatograms of SS-linked fragments generated in the microwave-assisted partial acid hydrolysis of the cyclotide mixture in 2 M HCl with two time points: 5 min (upper panel) and 2 min (lower panel), respectively. With comparison of these two chromatograms, peaks containing SS-linked fragments could be classified into three distinct groups. The first group (region 1) had a retention time range of 10−18 min, with peak intensities increasing with prolonged incubation from 2 to 5 min. The second group (region 2) eluted at a retention time of 18−45 min. These peaks were generally not well-defined and not well-resolved and appeared to be diminished when the cyclotide mixtures were subjected to 5 min of microwave-assisted partial acid hydrolysis. The third group (region 3) eluted at a retention time of 45−55 min. They were mostly impurities unrelated to cyclotide fragments and were not pursued further.

The early elution time of the first group of hydrolyzed peptide fragments in C18 column (10−18 min from 8−12% acetonitrile LC gradient) suggested that these fragments are relatively hydrophilic and carry multiple charges. These characteristics are consistent with fragments containing a single disulfide bond. A single-SS-bond peptide is essentially two linear peptides linked together by a disulfide bond, and this SS-linked peptide will have at least four charge-carrying groups from its two N- and two C-termini. Although these fragments were not tryptic peptides, i.e., their C-terminus residues are not necessarily lysine or arginine, LC−MS/MS analysis confirmed that most of these peptides carried double charges, indicating that the fragments have two amino termini. The second group of

(17) Hua, L.; Low, T. Y.; Sze, S. K. *Proteomics* **2006**, *6* (2), 586–591.
(18) Zhong, H.; Marcus, S. L.; Li, L. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (4), 471–481.
(19) Zhong, H.; Zhang, Y.; Wen, Z.; Li, L. *Nat. Biotechnol.* **2004**, *22* (10), 1291–1296.

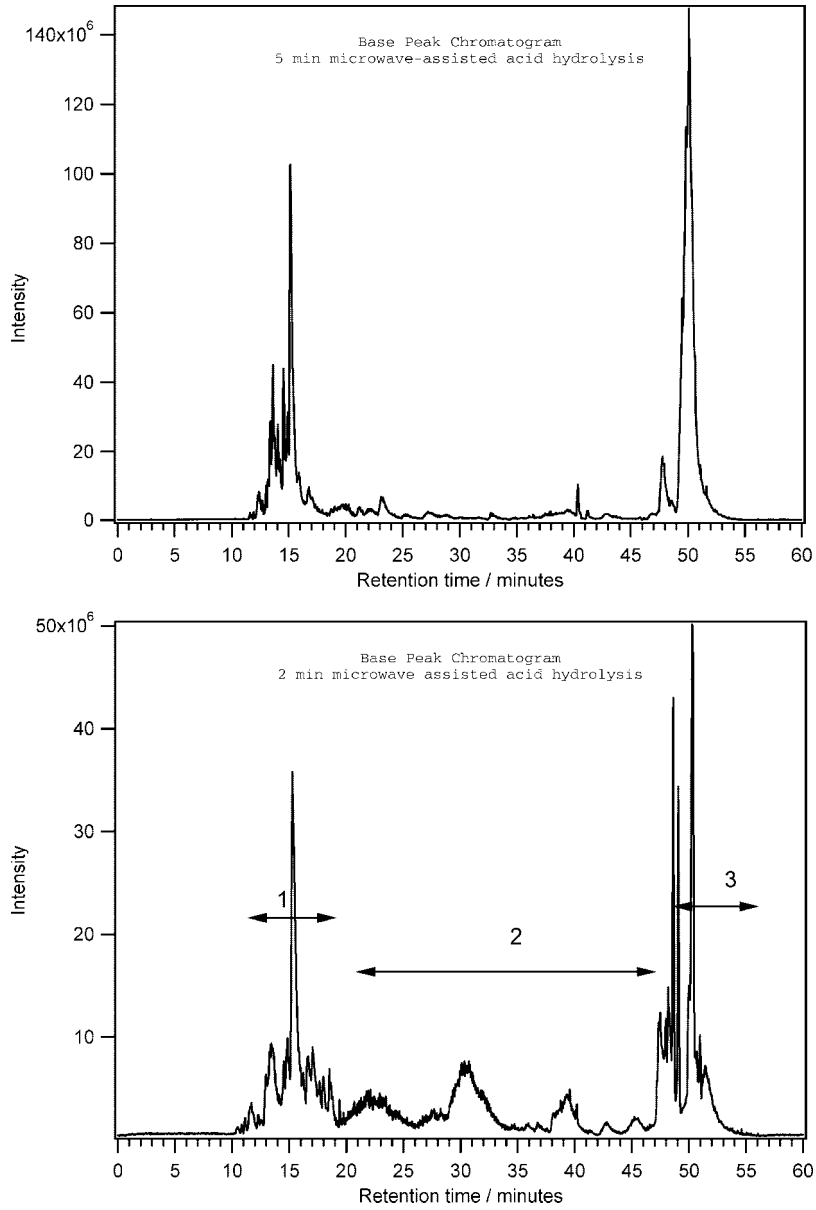

Figure 3. The base peak chromatograms of microwave-assisted partial acid hydrolysis of kalata B1 and varv A mixture in 2 M HCl: (upper) 5 min hydrolysis and (lower) 2 min hydrolysis. Three chromatographic features (1, 2, and 3) can be distinguished by comparing these two time points of acid hydrolysis.

peptide fragments with a retention time range of 18−45 min was comprised mainly of large partially hydrolyzed fragments with molecular weights greater than 1500 Da. This group of fragments could not be assigned to any single-SS-linked fragments. These are probably partially hydrolyzed cyclotide fragments with two or three SS bonds and are intermediates which undergo further hydrolyses to form the smaller fragments found in the first group (with a 10−18 min retention time range after 5 min microwave-assisted hydrolysis).

Both microwave radiation and autoclaving have been used for partial acid hydrolysis of cyclotides in this study. Both methods produced hydrolyzed products that lead to identification of the SS bonds by LC−MS/MS analysis. A time of 30 min of autoclaving at 121 °C and 15 psi or 5 min of microwave irradiation at full power hydrolyzed the kalata B1 and varv A mixture completely. It should be noted that the sample must be placed in the same position in the microwave oven for reproducible results because of the uneven distribution of microwave energy within the oven. Degassing of the sample prior to microwave heating or autoclaving is essential to prevent oxidization of peptides.

Tryptophan is an acid-labile amino acid that often decomposes in the absence of scavengers during conventional acid hydrolysis by 6 M HCl during amino acid analyses.[20] However, it is worth mentioning that several disulfide-linked fragments with intact Trp residue were found (Table S-1 in the Supporting Information), suggesting that this residue survived from the acid hydrolysis conditions employed in our experimental setup. Figure 4 shows an MS/MS spectrum of a triply charged precursor ion with m/z of 375.83 Da. The assignment of the weak fragments of the MS/MS spectrum is shown in Figure S-4 in the Supporting Information. The CAD ion peaks matched the calculated CAD fragments of the disulfide-linked hydrolysis product GTC(III)−SWPVC(VI)TR. This indicates that the

(20) Tsugita, A.; Scheffler, J. J. *Eur. J. Biochem.* **1982**, *124* (3), 585–588.

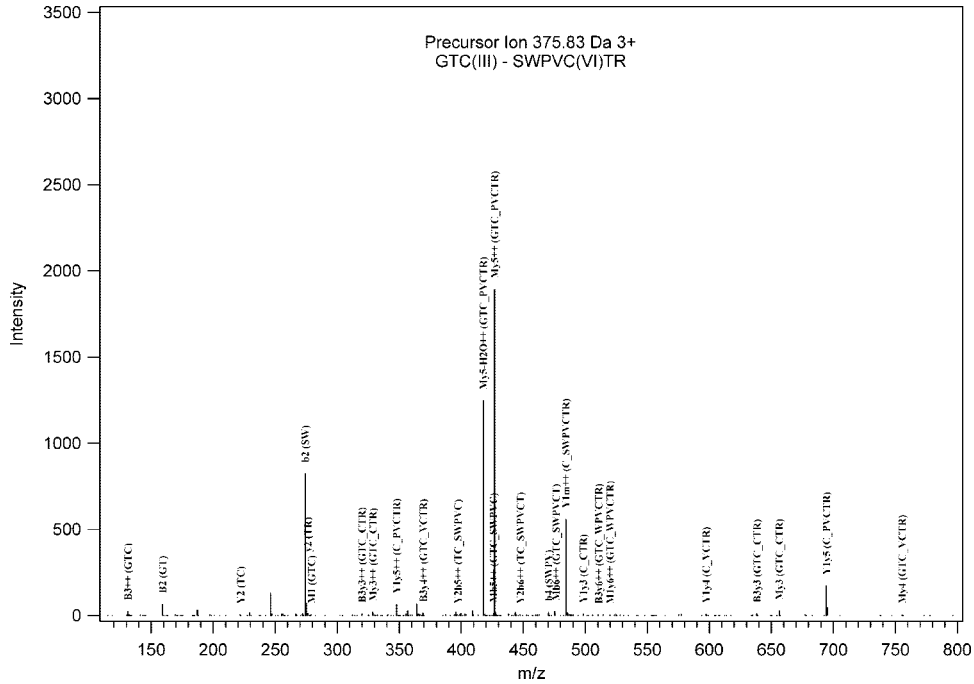

Figure 4. MS/MS spectrum of a hydrolysis product with precursor ion $m/z = 375.83$. The CAD ion peaks matched to the calculated fragments of GTC(III)−SWPVC(VI)TR.

condition employed in our study is strong enough for the partial acid hydrolysis of peptide bond of cyclotides but is mild enough to preserve amino acid integrity.

LC−MS/MS Analysis. Enzymatic digestion and partial acid hydrolysis coupled to molecular weight measurement of SS-linked peptides have been applied to characterize the disulfide bonds in proteins.[11,12,16] However, the fragment mass measurement alone may not definitely identify the disulfide bonds in cyclotides because many possible isoforms can be generated by the acid hydrolysis. The presence of isoforms is further complicated by the relatively small size of cyclotides with 28−37 amino acids and their composition which is predominated (>50%) by amino acids such as cysteine, glycine, threonine, and serine. Therefore, many possible isoforms can be produced by a disulfide-linked fragment in the six-loop permutation of a cyclotide (Figure 1). For example, kalata B1 is a 29 amino acid cyclotide with an amino acid composition of $C_6G_5T_5P_3V_3N_2E_1R_1S_1L_1W_1$. On the basis of its amino acid sequence, possible permutation of SS-linked fragments with a cysteine residue in each peptide was calculated to be 6790 (kalata B1, Supporting Information Table S-2a; varv A, Supporting Information Table S-2b). A total of 81.5% of all the possible fragments have at least two isoforms. As an example, there are 25 fragments with a molecular weight of 1265.57 Da. Therefore, molecular weight determination alone cannot be used to unambiguously characterize the disulfide bonds in the SS-rich and highly conserved cyclotides. To circumvent this problem, we used LC−MS/MS to determine the amino acid sequence of each potential SS-linked peptide.

Supporting Information Table S-1 shows the SS-linked fragments that were identified by LC−MS/MS. The LC−MS/MS was run in a typical data-dependent mode with dynamics exclusion of 20 s. This condition enables identification of both strong and weak ions. The ion intensity of these fragments varied in several orders of magnitude and represented the relative susceptibility of the peptide bonds by acid hydrolysis. The intensity of each peptide is a combination of two hydrolytic events; first being the primary fragmentation at the labile points of the cyclotide peptide backbone, and the second being secondary hydrolyses of the primary fragmentation products.

A typical LC−MS/MS experiment in the LTQ-Orbitrap with a 1 h LC gradient can generate up to 10 000 MS/MS spectra, rendering manual interpretation of the data set difficult. In a typical proteomics experiment, these MS/MS spectra are submitted to a protein database search algorithm for protein identification.[21,22] Although a database search engine for disulfide bond determination has been reported recently,[23] we wrote our own search algorithm to cater to disulfide-rich cyclic peptides. This algorithm extracts useful information from a large pool of LC−MS/MS data for cyclotide disulfide bond determination. First, all theoretical disulfide-linked fragments were calculated based on the sequence of the cyclotides kalata B1 (Supporting Information Table S-2a) and varv A (Supporting Information Table S-2b). There are 15 possible disulfide linkages in each cyclotide. All possible combinations of fragments with one pair of disulfide bonds and only one cysteine residue in each peptide were calculated and listed in Supporting Information Tables S-2a and S-2b. The theoretical molecular weight of each fragment was tabulated for comparison to the precursor mass of each MS/MS spectrum. The $M + H^+$ mass of each MS/MS spectrum was extracted from the dta file. If the mass error was less than 5 ppm, the fragmentation peaks of the dta file were compared to the theoretical CAD fragments of the peptides.

(21) Eng, J. K.; McCormack, A. L.; Yates, J. R., III *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.
(22) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III *Nat. Biotechnol.* **1999**, *17* (7), 676–682.
(23) Xu, H.; Zhang, L.; Freitas, M. A. *J. Proteome Res.* **2008**, *7* (1), 138–144.

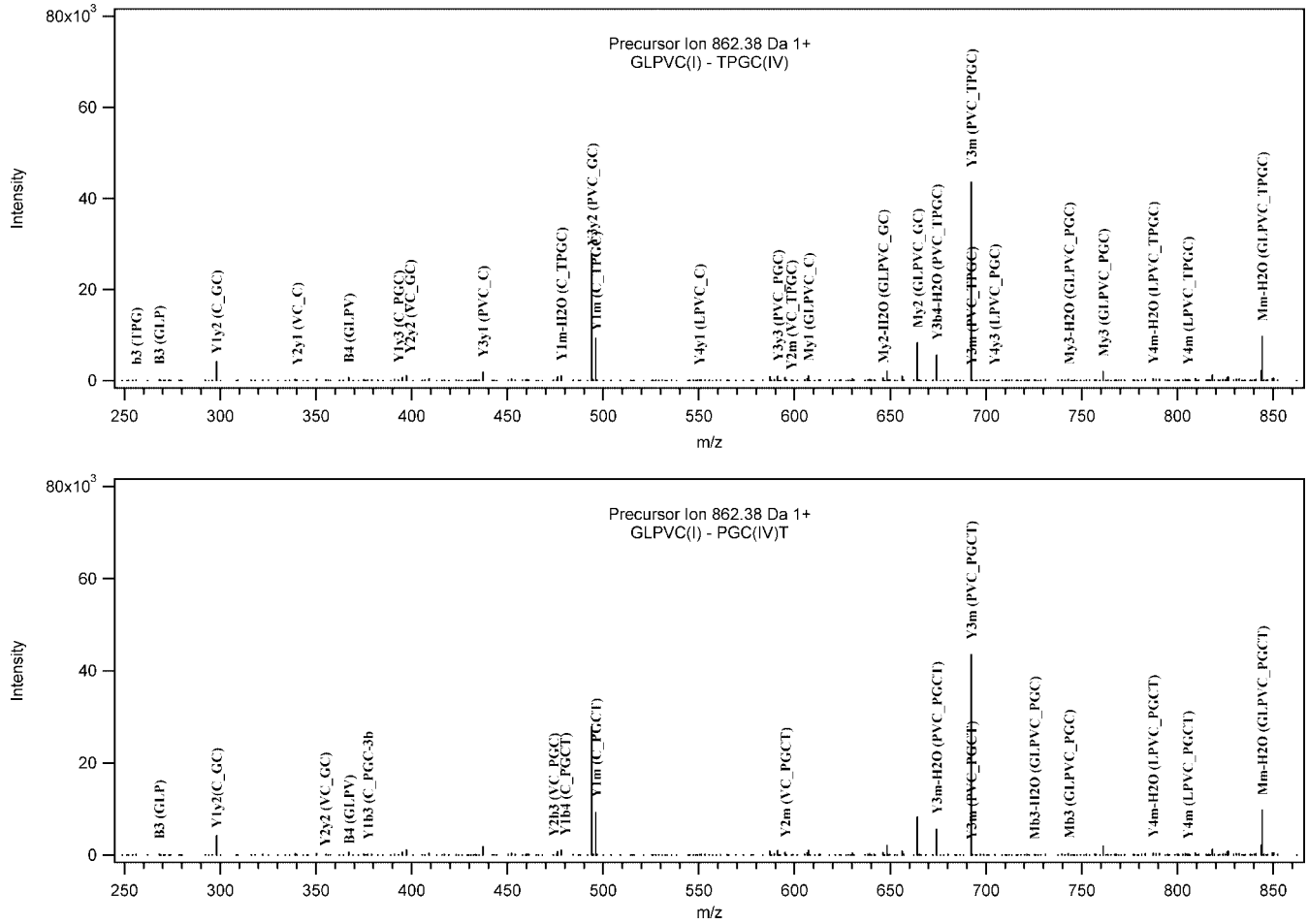

Figure 5. Assignment of CAD spectrum of a peptide with precursor ion mass = 862.38 Da. Two disulfide linked peptides, GLPVC(I)−TPGC(IV) and GLPVC(I)−PGC(IV)T, have their theoretical CAD fragments significantly similar to the experimental MS/MS spectrum. However, the first peptide is the correct hydrolyzed product.

Since cyclotides are largely comprised of only a few types of amino acids, many isoforms of hydrolyzed fragments with closely similar structures can match the experimentally determined molecular weight of such fragments. As shown in Figure 5, an MS/MS spectrum of a singly charged precursor ion of 862.38 Da could match up to 4 possible isoforms, GLPVC(I)−TPGC(IV), GLPVC(I)−PGC(IV)T, LPVC(I)G−TPGC(IV), and LPVC(I)G−PGC(IV)T. Interpretation of the MS/MS spectrum is important for the correct assignment of the isoforms. To interpret the MS/MS spectra of SS-linked peptides, we introduced a modified nomenclature to assign the CAD peaks because the conventional a,b,c and x,y,z nomenclature[24−26] for linear peptide fragments cannot be applied unambiguously to CAD fragments, as there are two peptides from the same protein joined by a SS bond between two cysteine residues. The peptide (and its CAD fragments) with a cysteine closest to the N-terminus was designated with the capital letter M (A,B,C and X,Y,Z for its fragments containing N- and C-terminals, respectively in the same format as conventional nomenclature of linear peptides); whereas the peptide closest to the C-terminal of the protein was designated as m (a,b,c and x,y,z for its fragments containing N- and C-terminals, respectively, in the same format as conventional nomenclature of linear peptides). Interpretation of CAD spectra of SS-linked peptides obtained from hydrolyzed cyclotides are more complicated than that of linear peptides because multiple isoforms can be assigned to a single fragment. For example, a fragment with 607.3 Da in Figure 5 can be assigned to either My1 (GLPVC(I)_C(IV)) or Y4y2 (LPVC(I)_GC(IV)) fragments of GLPVC(I)−TPGC(IV). Although disulfide bond determination in protein by LC−MS/MS has been described,[23,27−30] the assignment of isoform fragments to tandem MS spectra is not established. We have manually interpreted MS/MS spectra of fragments from the hydrolyzed cyclotide mixture in the LTQ-Orbitrap (MS/MS was done in the LTQ) and observed that the frequency of single bond cleavage was higher than cleavage of multiple bonds in SS-linked peptide fragments. Although the

(24) Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11* (11), 601.
(25) Hunt, D. F.; Yates, J. R., III; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83* (17), 6233–6237.
(26) Biemann, K. *Biomed. Environ. Mass Spectrom.* **1988**, *16* (1−12), 99–111.
(27) John, H.; Forssmann, W. G. *Rapid Commun. Mass Spectrom.* **2001**, *15* (14), 1222–1228.
(28) Kim, H. I.; Beauchamp, J. L. *J. Am. Chem. Soc.* **2008**, *130* (4), 1245–1257.
(29) Lee, T.; Singh, R.; Yen, T. Y.; Macher, B. *Comput. Syst. Bioinf. Conf.* **2007**, *6*, 41–51.
(30) Thakur, S. S.; Balaram, P. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (3), 358–366.
(31) Kruger, N. A.; Zubarev, R. A.; Carpenter, B. K.; Kelleher, N. L.; Horn, D. M.; McLafferty, F. W. *Int. J. Mass Spectrom.* **1999**, *182*, 1–5.
(32) Cemazar, M.; Craik, D. J. *Int. J. Peptide Res. Ther.* **2006**, *12* (3), 253–260.
(33) Tam, J. P.; Lu, Y. A.; Yang, J. L.; Chiu, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (16), 8913–8918.

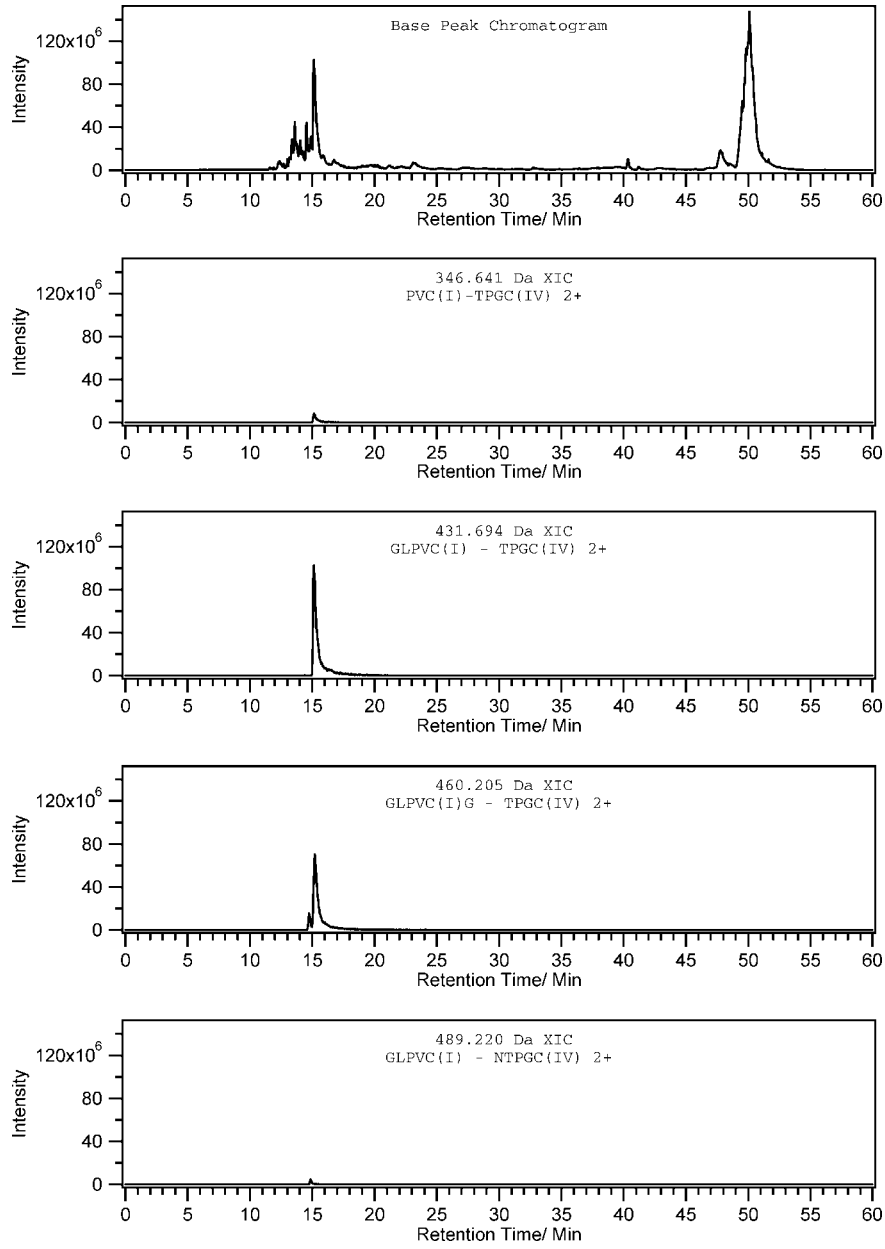

Figure 6. Extracted ion chromatograms of some hydrolyzed products. The fragments with *m*/*z* at 431.694 (GLPVC(I)—TPGC(IV)) and 460.205 (GLPVC(I)G—TPGC(IV)) were the most abundant fragments.

disulfide bond can be cleaved by CAD, it is not the preferred cleavage site for disulfide-linked peptides. Thus, the 607.3 Da fragment (shown in Figure 5) should be assigned to My1 which involved a single bond cleavage at the G−C bond of the TPGC(IV) peptide. Taken together, if a peak can be assigned to more than one possible CAD cleavage products, the product with minimal cleavage sites is more likely to occur than fragments produced by multiple backbone cleavages in low-energy CAD.

Through LC−MS/MS analysis, two out of four possible hydrolysis products (861.372 Da), GLPVC(I)−TPGC(IV) and GLPVC(I)−PGC(IV)T, have their theoretical CAD fragments agreeing significantly with the experimental MS/MS spectra. To assign the correct isoforms, the CAD dissociation probability of the amino acid residues was applied to guide our assignment.[31] There is a proline residue in each of the disulfide linked peptides. As the peptide bond at the N-terminus side of proline is usually the most labile bond for CAD fragmentation, the correct hydrolysis product is expected to produce strong CAD fragments with cleavage at the proline N-terminus. The two strongest CAD peaks in Figure 5 were 692.3 and 494.2 Da which matched to the Y3m (PVC(I)_TPGC(IV)) and Y3y2 (PVC(I)_GC(IV)) CAD fragments of GLPVC(I)−TPGC(IV), respectively. This agrees with the favored dissociation sites at the N-termini of proline and glycine. Therefore, the spectrum was assigned to GLPVC(I)−TPGC(IV). As three proline residues are distributed in different loops of the kalata B1 and the peptide backbone fragmentation at the N-termini of proline is highly favorable, as a result, most of the MS/MS spectra showed one or two very strong fragments. The detailed assignment of the relatively weak peaks in the MS/MS spectrum of singly charged GLPVC(I)−TPGC(IV) precursor ion is shown in Figure S-2 in the Supporting Information. The MS/MS spectrum of the corresponding doubly charged precursor ions of the same peptide is shown in Figure S-3 in the Supporting Information.

Table 1. Solvent Accessible Surface of Individual Amino Acid Residues of Kalata B1 Based on the NMR Structure (1NB1)[a]

residue	sequence number	total	apolar	backbone	side chain	ratio (%)	in (i)/ out (o)
Cys	1	20.31	0.83	20.31	0	0	i
Gly	2	90.81	56.47	90.81	0	100	o
Glu	3	34.18	30.96	8.99	25.19	17.8	i
Thr	4	76.83	39.31	12.49	64.34	60.6	o
Cys	5	1.12	1.12	1.12	0	0	i
Val	6	92.03	78.22	13.81	78.22	64	o
Gly	7	69.4	45.86	69.4	0	79.6	o
Gly	8	49.2	26.88	49.2	0	56.4	o
Thr	9	103.35	86.07	8.89	94.46	88.9	o
Cys	10	34.88	6.31	28.96	5.92	5.8	i
Asn	11	132.47	41.87	23.74	108.73	95.1	o
Thr	12	37.12	26.79	7.08	30.04	28.3	
Pro	13	134.72	116.9	31.03	103.68	98.6	o
Gly	14	62.9	47.27	62.9	0	72.1	o
Cys	15	12.5	0.07	12.5	0	0	i
Thr	16	94.29	86.08	5.22	89.07	83.9	o
Cys	17	59.09	24.29	25.27	33.82	33.1	
Ser	18	48.98	37.07	15.21	33.77	43.6	
Trp	19	170.94	145	22.71	148.23	66	o
Pro	20	54	54	0	54	51.3	o
Val	21	55.53	55.53	0	55.53	45.4	
Cys	22	0.75	0	0.75	0	0	i
Thr	23	18.11	14.99	0	18.11	17.1	i
Arg	24	79.41	37.07	0	79.41	40.6	
Asn	25	120.4	65.23	41.07	79.33	69.4	o
Gly	26	74.76	43.82	74.76	0	85.7	o
Leu	27	88.95	88.95	0.3	88.64	60.6	o
Pro	28	82.57	64.33	18.36	64.21	61	o
Val	29	81.36	66.04	25.63	55.73	45.6	

[a] The result was calculated by the GetArea program (http://pauli.utmb.edu/cgi-bin/get_a_form.tcl).

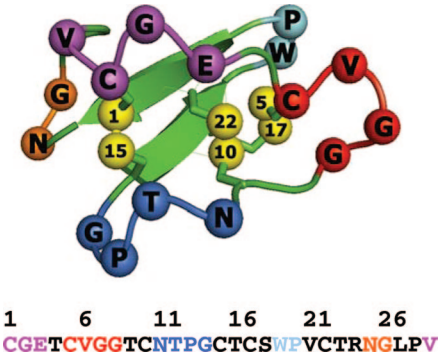

Figure 7. Illustration of the kalata B1 structure (PDB code 1NB1) Cα of the following residues are represented with spheres: NG, orange; VCGE, magenta; CVGG, red; NTPG, marine; WP, cyan. Sulfur atoms are represented with yellow spheres. The structure was processed by PyMol (http://pymol.sourceforge.net/).

With the above-described method, Supporting Information Table S-1 shows the identified hydrolysis fragments. There are 13 and 15 single-SS-linked fragments unique to kalata B1 and varv A, respectively, differing only at position 16 (S for varv A and T for kalata B1), and 60 fragments shared by both cyclotides. On the basis of these fragments, we can conclude that kalata B1 and varv A share similar SS connectivity of Cys I−IV, II−V, and III−IV as shown in Figure 1.

The fragmentation patterns of both cyclotides based on time-point analysis of 2 and 5 min under microwave-assisted conditions also yielded important clues to the susceptibility of peptide bonds in the tightly packed cyclotide structure. The relative abundance of fragmentation reflects on the liability of peptide bonds to acid hydrolysis, which in turn, should depend on the cyclotide structure and provide clues to the specificity of backbone fragmentation under acidic conditions.

Specific Backbone Fragmentations and Structure Information. *Asn25−Gly26.* Cyclotides are gene-coded post-translational products. Thus, their amino acid sequences can be determined by mass spectrometry peptide sequencing and validated by cDNA sequencing. The DNA sequence encodes the linear amino acid sequence which subsequently cyclizes head-to-tail at Gly26 and Asn25 with three disulfide bonds. G−N head-tail ligation is found in >90% of all cyclotide sequences (the remaining sequences contain the G−D bond). Since cyclization involves elimination of a water molecule and as the formation of a disulfide bond pair loses two hydrogen atoms from the two cysteine residues, the cyclic cyclotide containing three disulfide bonds can be deduced from the calculated molecular weight based on the amino acid residues and the measured molecular weight by MALDI-MS. However, the presence of many serine and theronine amino acids in cyclotides complicates structure determination because serine and threonine in peptides or proteins are prone to water loss during mass spectrometry analysis. An observed −18 Da decrease in molecular weight can be due to either the formation of a cyclic structure or as a result of water loss. The DNA sequence of the cyclotide precursor predicts that G26 is the N-terminus, while N25 is the C-terminus of the ligation site. Some SS-linked fragments determined under our conditions by LC−MS/MS (Supporting Information Table S-1) contain G26−N25 residues. These fragments, together with the presence of three SS linkages and the pattern of Cys spacing, provide confirmation for cyclotides.

Specific Backbone Fragmentation. Figure 3 shows that cyclotide fragmentation under partial acid hydrolysis was not random. Certain peptide bonds were hydrolyzed first to generate large fragments (region 2 of Figure 3). With longer acid hydrolyses, the larger fragments were converted to the smaller single-SS-linked fragments in region 1 (Figure 3). The strongest chromatographic peak eluted at around 15 min in correlation with the most abundant hydrolysis product. Its intensity increased quickly when hydrolysis proceeded from 0 to 5 min of microwave irradiation. Two coeluted fragments with molecular weights of 861.38 (GLPVC(I)−TPGC(IV)) and 919.40 Da (GLPVC(I)G−TPGC(IV)) contributed to this peak.

Figure 6 shows the base peaks and four extracted ion chromatograms of kalata B1 and varv A under a 5 min microwave-assisted acid hydrolysis in 2 M HCl. Although the amount of hydrolyzed products varied widely in the mixture, the partial acid hydrolysis mainly generated several highly abundant primary fragments, as well as some secondary fragments due to further hydrolysis of the primary fragments by acid. Interestingly, two dominant hydrolyzed products have been detected (431.69 Da, doubly charged GLPVC(I)—TPGC(IV) ion; 460.204 Da, doubly charged GLPVC(I)G−TPGC(IV) ion); many low-abundance SS-linked (I−IV, II−V, and II−VI linked) peptides have been identified in the hydrolyzed mixture as shown in Figure 6.

Two factors may contribute to the acid-susceptible bond fragmentation:[12] (1) steric factors or the solvent accessible surface of the peptide bond and (2) anchimeric assistance of the amino acid side chain functional group.

Steric and Solvent Accessible Sites. Gly and Pro are two amino acids that are susceptible to acid hydrolysis. Gly is favored by

being the least sterically hindered amino acid. Pro, the only amino acid giving a secondary amide, is favored by hydrolysis to release its bond strain. Interestingly, both are often solvent exposed and occur frequently in turns. In kalata B1 and varv A, there are five Gly and three Pro. Combinations such as G7−G8 and P13−G14 make these bonds highly susceptible to acid hydrolysis. Indeed, many fragments listed in Supporting Information Table S-1 contain N-terminus Gly or Pro.

Table 1 shows the solvent accessible surface of individual amino acid residues in kalata B1 based on the NMR structure (1NB1) calculated with the GetArea program (http://pauli.utmb.edu/cgi-bin/get_a_form.tcl). The result showed that G2 and G26 were the most accessible of the backbone amino acids to solvent. On the basis of the steric factor and the solvent accessibility of the backbone, these two peptide bonds are expected (and indeed found) to be highly susceptible sites for acid hydrolysis.

Side Chain Assistance. Previous studies have shown that peptide bond acid fragmentation of a protein is not a random process.[12] Several side chain functional groups of amino acids such as carboxylic and hydroxyl groups provide anchimeric assistance to accelerate regiospecific acid hydrolysis of peptide bonds. Thus, the acid-susceptible peptide bonds adjacent to these amino acids comprising of Asp, Glu, Ser, Thr, Asn, and Gln provide specific fragmentation in cyclotides. With inspection of the 3D NMR structure of kalata B1, the hydrophobic and polar amino acids are clustered into different regions on the surface. Residues such as Ser, Thr, Arg, and Asn are clustered in the hydrophilic regions. As the Asn amide side chain is pointed out of the surface as revealed by the NMR structure, it is susceptible to deamidation to Asp. Consistent with the known structure of kalata B1, all of the identified hydrolysis products containing Asn were deamidated to aspartic acid, with no observable intact Asn remaining, suggesting that Asn deamidation may precede backbone fragmentation. The carboxylic group of aspartic acid is a functional group known to strongly catalyze acid hydrolysis of peptide bonds. Therefore, aspartic acid may catalyze the hydrolysis of the D25−G26 peptide bond after N25 to D25 deamidation. The liability of the N25−G26 bond was further enhanced by the high solvent accessibilities of the N25 side chain and the G26 backbone.

NMR structures show that the carboxylic functional group of E3 (Glu), an invariant amino acid residue found in all cyclotides, forms intramolecular hydrogen bonds with backbone amides of N11 (Asn) and T12 (Thr) to stabilize the 3D structure of kalata B1.[32] These hydrogen bonds can catalyze the hydrolysis of the N11−T12 peptide bond. The partial acid hydrolysis results reflected the susceptibility of individual peptide bonds to acid hydrolysis, which in turn, may be attributed to the compactness of the CCK-structure of the cyclotide. Figures 3 and 6 show that GLPVC(I)G−TPGC(IV), 460.204 Da, and GLPVC(I)−TPGC(IV), 431.69 Da, are the two most abundant acid hydrolysis products cleaved at N11−T12, with both containing the CysI−IV linkage. Time-point hydrolysis experiments also showed that these two single SS-linked fragments were among the first fragments produced from the acid hydrolysis. Taken together, our results show that the peptide bonds, N25−G26, N11−T12, C15−T16, C1−G2, and G2−E3, are the most susceptible sites for acid hydrolysis under our experimental conditions.

Turns. Interestingly, the combinations of steric, solvent accessible, and side chain catalysis factors place most of the labile bonds at turns (Figure 7). Thus, our results suggest that acid-catalyzed fragmentations of closely SS-packed, CCK-structured cyclotides occur frequently at solvent-exposed turns. Since all cyclotides contain the near invariant N25−G26 and G2−E3 bonds, single-SS-linked fragments containing CysI−IV are likely dominant fragments produced by partial acid hydrolysis under our conditions and could serve as a biomarker for cyclotide screening.

N25 and G26 are where the putative cyclotide ligase forms the cyclic structure of cyclotides. Interestingly, it is also one of the most acid susceptible sites. This may imply that this site contains the weakest bond and the most solvent-accessible region in the cyclotide structure. If the cyclotide cyclized before disulfide bond formation, the cyclization bond may be buried or shielded from the solvent and may not be the most water-accessible bond for initial acid hydrolysis-mediated bond cleavage. Therefore, the N25−G26 peptide bond formation is likely the last step in the cyclotide biosynthetic pathway, and formation of the three disulfide bonds might precede the cyclic structure closure by N25−G26 ligation. This observation is consistent with the recent report that the N25−G26 bond formation is the last step in the cyclotide biosynthetic pathway.[2]

In conclusion, we have developed a potentially high-throughput method for cyclotide disulfide bond characterization based on flash partial acid hydrolysis, LC−MS/MS and a computer program for disulfide bond assignment. This method should be generally applicable to disulfide bond determination of other natural-occurring or synthetic SS-rich cyclic peptides and cyclotides.[33] Besides the unambiguous assignment of the SS bonds, the abundance of hydrolytic fragments can be measured by integration of the corresponding extracted ion chromatogram area. The acid-catalyzed backbone fragmentation is predictable and frequently occurs at the solvent-exposed turns. Our results could provide the order of bond breakage so as to anticipate the identities of SS-linked fragments for finger printing, and, to a certain extent, to infer the biosynthetic cyclic formation of the ultrastable cyclotide structure.

ACKNOWLEDGMENT

We thank Sai Kiang Lim and Paul Tan for editorial help. This work was in part supported by grants from Nanyang Technological University (URC, Grant RG61/06, S.S.K.), Ministry of Education of Singapore (ARC, Grant T206B3211, S.S.K.), and the U.S. Public Health Service (Grant EB001986, J.P.T.).

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review October 14, 2008. Accepted December 11, 2008.

AC802175R