Organic LETTERS

Letter

pubs.acs.org/OrgLett

Protoaculeine B, a Putative *N*-Terminal Residue for the Novel Peptide Toxin Aculeines

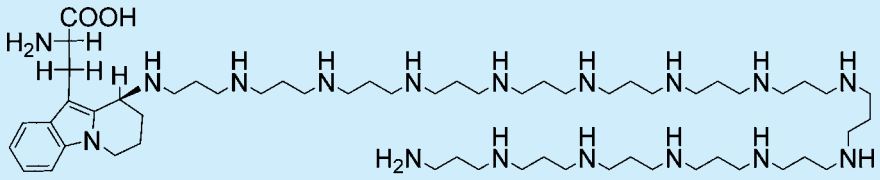
Satoko Matsunaga,[‡] Reimi Kishi,[†] Kazunori Otsuka,[∥] Masaki J Fujita,[§] Masato Oikawa,[∥] and Ryuichi Sakai*[,†]

[†]Faculty of Fisheries Sciences and [§]Creative Research Institution, Hokkaido University, Hakodate 041-8611, Japan

[‡]Hakodate National College of Technology, Hakodate 042-8501, Japan

[∥]Yokohama City University, Seto 22-2, Kanazawa-ku, Yokohama 236-0027, Japan

Supporting Information



protoaculeine B (1)

ABSTRACT: A new polyamine-modified indole derivative protoaculeine B (**1**) was isolated from Okinawan marine sponge *Axinyssa aculeata*. The structure of **1** was assigned on the basis of spectral data along with chemical transformations. Because the structure of **1** greatly inferred the *N*-terminal amino acid for highly modified peptide toxin aculeines, the probable structure for aculeine B was proposed on the basis of high-resolution mass spectral analysis.

Recently, we reported the isolation, biological activity, and amino acid sequence for novel peptide toxin aculeines (ACUs) from the Okinawan marine sponge *Axinyssa aculeata*.[1] ACUs are ribosomal peptides post-translationally modified by long-chain polyamines (LCPA). Three aculeines A (**2**), B (**3**), and C were obtained and their molecular sizes of approximately 6570, 5706, and 2800, respectively, were estimated on the basis of MALDI-TOF mass and gel electrophoresis data.[1] ACU-A and -B were purified to reasonable homogeneity containing a few polyamine homologues, while ACU-C was found to be a complicated mixture judging from the complex ion cluster observed in the mass spectral data.[1] We thus focused structure elucidation on **2** and **3** and disclosed the amino acid sequence for the peptide common in both molecules. The peptide (AcuPep) was composed of 44 amino acid residues and had a characteristic arrangement of cysteine residues known as a cysteine knot motif, which occurs often in small bioactive peptides such as conotoxins.[1] The *N*-terminal amino acid of ACUs, however, failed to be determined by Edman degradation, while tryptophan was suggested on the basis of analysis of the gene sequence that encodes ACUs.[1] These results suggested that the *N*-terminal amino acid residue for ACUs is the tryptophan derivative and also a site of modification by LCPA. This assumption was supported by isolating *N*-terminal tripeptide after protease treatment of ACU-A in that a disubstituted benzene ring and LCPA were recognized from the ^{1}H NMR data.[1] Further and more precise structure elucidation of the *N*-terminal amino acid of aculeines, however, was hampered because direct NMR analysis of the peptide was difficult due to minute quantity and structural complexity of the compound, given that it is so far not possible to prepare ^{13}C and ^{15}N labeled toxins.

In the present study, however, we found in the same sponge extract a molecule that exhibited the molecular weight corresponding to that of the *N*-terminal residue of ACU-B. The structure of this compound, denoted as protoaculeine B (pACU-B, **1**), provided us with further structural insight into the *N*-terminal residue of ACU-B. This information along with super-high-resolution mass spectral technique greatly encouraged us to overcome the structural problems that we have had so far, allowing further exploration toward the structure determination of aculeines. Here, we report isolation and structure determination of **1** and elucidate structures of ACUs further and then propose their probable structures.

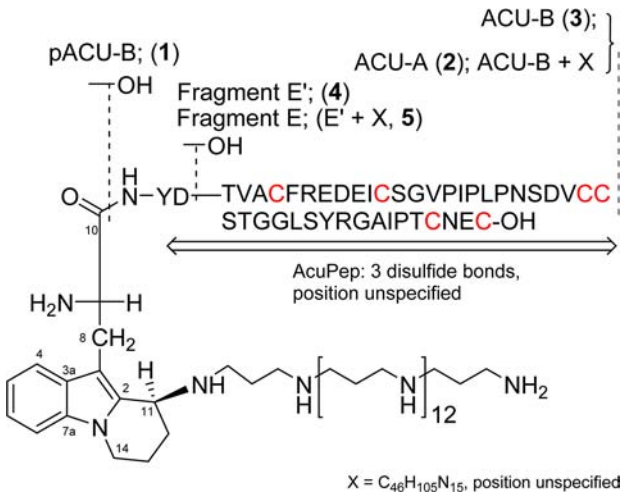
The *N*-terminal tripeptides of ACU-B (fragment E′, **4**) and that of ACU-A (fragment E, **5**), were produced by enzymatic hydrolysis of the toxins (Scheme 1).[1] Accurate molecular masses for these molecules were determined by Orbitrap-ESI mass analyses (Table S1, Supporting Information).[2,3] Because both fragment E and E′ contained two common amino acid residues, tyrosine-aspartic acid, the molecular weight of the putative *N*-terminal residue for ACU-A and -B (as the −COOH form) was calculated to be 1940 and 1072, respectively.

During the separation of ACUs, we noted the presence of a conspicuous molecular ion at *m*/*z* 1073 (MH^+, MALDI-TOFMS) that corresponded to the molecular mass for the theoretical *N*-terminal residue for ACU-B (Scheme 1). We thus

Received: April 26, 2014
Published: May 20, 2014

dx.doi.org/10.1021/ol5011888 | *Org. Lett.* 2014, 16, 3090−3093

Scheme 1. Structures of 1, 4, 5 and Working Structures of 2 and 3



elaborated on the compound in the hope that it would facilitate further structure elucidation of ACUs.

The aqueous extract of the sponge was chromatographed on a Sephadex LH-20 column using water as an eluent. The fractions that contain ACUs and free long-chain polyamines (LCPA-*Aa*) were combined and separated further by C18 column monitoring of the molecular ion at *m/z* 1073 in MALDI-TOFMS. Final purification was carried out using HPLC to give compound **1** as a TFA salt.

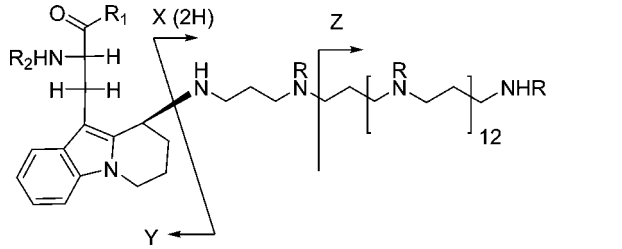
An Orbitrap ESIMS for **1** gave an exact molecular ion which corresponded well with the estimated *N*-terminal residue of **3** (Table S1, Supporting Information). MS^2 analysis of the molecular ion afforded a large product ion at *m/z* 816. Furthermore, MS^3 data on the ion at *m/z* 816 showed a series of product ions with 57 mu separation indicating that this product ion corresponded to the LCPA of the 14-meric propylamine (Figure S4, Supporting Information). The exact mass for the product ions obtained in the MS^2 data agreed well with the theoretical values for the LCPA, and the fragmentation pattern corresponded well with that of free LCPA reported previously.[4]

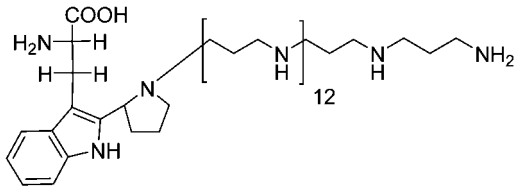
The presence of LCPA as a partial structure was also evident from ^{1}H NMR spectrum of **1**, since two characteristic signals at δ 3.02 and 1.95 corresponding to the amino methylene and center methylene of the propylamine unit for LCPA were observed. These data clearly indicated that a LCPA was incorporated in **1**. These data allow us to propose the molecular formula of **1** as $C_{57}H_{117}N_{17}O_2$.

^{1}H NMR spectrum of **1** had a typical disubstituted benzene system, composed of four ^{1}H signals at δ 7.11 (H5), 7.20 (H6), 7.40 (H7), and 7.55 (H4) and the corresponding protonated carbons resonated at δ 120.0 (C5), 122.9 (C6), 111.7 (C7), and 118.5 (C4), and two additional aromatic carbons at δ 125.6 (3a) and 136.8 (7a) were readily established by COSY, HSQC, and HMBC data (Table S2, Supporting Information).

Further analysis of two-dimensional NMR data indicated that additional nonprotonated sp^2 carbons at δ 107.1 (C3) and 129.5 (C2) both showed correlation to one of the methylene protons at δ 3.37 (H8a), which was further connected to another methylene proton at δ 2.95 (H8b), and to a proton at δ 4.01 (H9) attached to a heteroatom bearing carbon at δ 58.2 (C9). Furthermore, H9 showed a cross peak to a carbonyl carbon resonating at δ 174 (C10) in the HMBC spectrum. A nitrogen atom was incorporated in the aromatic system between the C7a and C2 on the basis of chemical shift values to form an indole ring. A primary amino group was assigned for the α position to the carbonyl carbon, and finally a carboxylic acid was assigned for C10 to build a tryptophan substructure.[5] Because the molecular formula of **1** corresponded to eight degrees of unsaturation, the remaining atoms, C_4H_7, should construct a ring or a double bond. 1-D TOCSY NMR data along with other 2-D NMR spectral data revealed a consecutive spin system [−$CHCH_2CH_2CH_2$−] for respective C11−C14. ^{13}C chemical shift values δ 47.1 (C14) and 53.2 (C11) indicated those carbons are attached to a nitrogen atom or two nitrogen atoms. Because a proton attached to the amino-methyne C11 was rather lower field shifted, resonating at δ 4.77, C11 was connected to C2 of the indole ring. One of the nitrogen atoms of LCPA is best fit to substitute on C11. Because the amino methylene C14 can be connected either to a nitrogen atom of the indole or one of the nitrogen atoms of the LCPA, two possible candidate structures A and B were proposed for the structure of **1** (Scheme 2); however,

Scheme 2. Fragment Ions for 1, 4, and 9





B

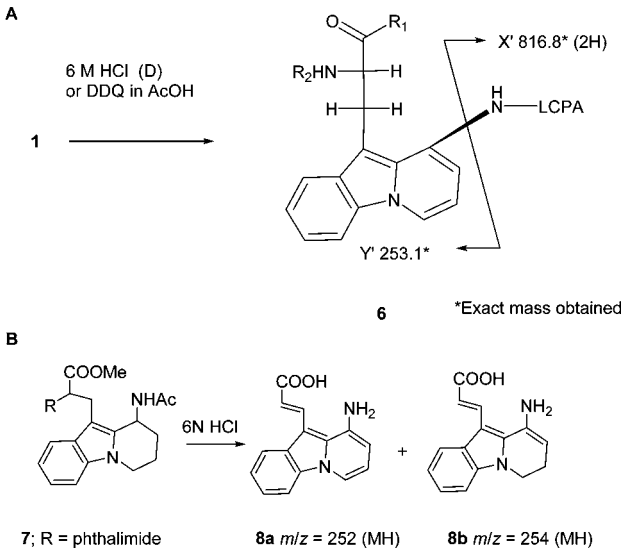
HMBC correlations which would support connectivity between C2−C11 or N1−C14 were missing. Therefore, mass spectral information was used to subsidize the structural assignment. Orbitrap MS^2 data from the molecular ion of **1** gave large product ions at *m/z* 816.8 (100%) and 257.1 (45%). These ions can be well accommodated to the candidate structure A. The candidate structure B, however, does not accommodate such fragment ions. Thus, the candidate A was assigned as the planar structure of **1** (Scheme 2).

A treatment of **1** with 6 M HCl (110 °C, 20 h) resulted in complete loss of ion at *m/z* 1073 (MH^+), but ion at *m/z* 1069 (MH^+) emerged (Figure S12, Supporting Information). An exact mass for the latter ion suggested a formula $C_{57}H_{113}N_{17}O_2$ (Δ5.4 ppm, Table S1, Supporting Information) for the product which corresponds to didehydro derivative **6**. Furthermore, the

dx.doi.org/10.1021/ol5011888 | *Org. Lett.* 2014, 16, 3090−3093

presence of characteristic fragment ions X′ and Y′ supported the assignment (Scheme 3A, Figure S12, Supporting

Scheme 3. Spontaneous Aromatization of 1 and 7



Information). These results indicate that a spontaneous dehydrogenation−aromatization took place when **1** was heated with HCl. Unfortunately, a diagnostic UV spectrum could not be obtained presumably due to relatively low amount of **6** presented in the reaction mixture.[6] The same mass spectral shift was observed when **1** was treated with DDQ (Figure S13, Supporting Information). Those data together supported the presence of 6,7,8,9-tetrahydropyrido[1,2-*a*]indole structure in **1**. Although some examples of spontaneous dehydrogenation in heterocycles under acidic conditions are described in the literature,[7] we reinforced our results by a reaction using synthetic model compound **7** (Scheme 3B).[8] Treatment of **7** with HCl gave products that show molecular ions corresponded with aromatized products **8a** and **8b** (see the Supporting Information for details).

Treatment of **1** by acetic anhydride and pyridine afforded pentadecaacetate **9** (Scheme 2). Incorporation of 15 acetyl groups was evident from the mass spectral data ($C_{87}H_{148}O_{17}N_{17}$, MH^+, Δ1.0 ppm, Table S1, Supporting Information). The linear nature of LCPA was suggested by consecutive loss of acetopropanamide groups in MALDI-TOF MS^2 data (Figure S14, Supporting Information). Along with the above data, product ions at *m*/*z* 1404.9 and 1388 (Scheme 2) indicated that the terminal nitrogen of LCPA attached to C11 left unreacted due presumably to steric hindrance.

Some stereochemical insights were revealed by further analysis of the NMR data for **1**. A pair of small and large coupling constants observed between H9 and H8a and H8b, respectively, and NOE observed between H9 and H8a, but not H8b in ROESY data indicated that C8−C9 bond is in a staggered conformation with H8b-H9 being *anti*-relationship (Figure 1). In a ROESY spectrum, a cross peak between H9 and H11 was observed. The spatial proximity of these protons suggested pseudoequatorial nature of H11. A small cross peak observed between H11 and protons resonated at δ 2.98 in COSY is assignable to a homoallylic ($^5J_{HH}$) coupling between H11 and H8b protons.[9] These data together led to assign the stereochemistry of **1** to be (9*R**,11*R**).

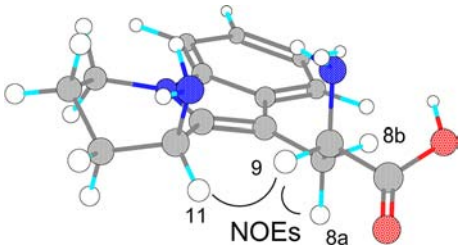


Figure 1. Stereochemistry of **1** estimated by NMR data.

With an implication of the *N*-terminal structure of **3** from the structure of **1** in hand, we further proceeded to elucidate the structure of ACUs employing the Orbitrap mass as a key tool. Although we have estimated the molecular weight of **2** and **3** on the basis of MALDI-TOF mass data to be 6569 and 5701, the resolution and accuracy were not high enough for the determination of molecular formulas.[1] Other mass spectral analyses including fast atom bombardment (FAB) or ESI-FT mass were not successful as no ions were detected in ether techniques: however, application of Orbitrap-ESIMS technique succeeded observation of multiply charged molecular ions at monoisotopic level with super high resolution and accuracy. Deconvoluted multiply charged molecular ions (M^+ equivalent) for each **2** and **3** were 6569.7960, and 5702.9343, respectively (Figure S15, Supporting Information).

Data from the present study suggested that **1** represents the structure of the *N*-terminal residue of aculeines, especially that of **3**. The peptide moiety of ACU contains six cysteine residues similar to the motif known as cysteine knots. Thus, it is reasonable to assume that the sulfur atoms of those Cys residues in ACUs are all involved in disulfide bonds. These information led us to propose a possible structure for **3** to be composed of the 44-amino acid peptide with three disulfide bonds, and **1** as a unique *N*-terminal residue being connected to the peptide by an amide bond (Scheme 1). The hypothetical molecular formula, $C_{252}H_{413}N_{69}O_{69}S_6$, for **3** thus can be calculated (*M* = 5701.9254), which agreed very well (Δ1.6 ppm) with the above mass result for **3** (Table S1, Supporting Information).

To assess the identity between **1** and the *N*-terminal residue of **3** further, the MS^2 data for the enzyme digest fragment E′ (**4**) obtained by V8 peptidase digestion of **3**[1] was examined (Figure S16, Supporting Information). The MS^2 data from the molecular ion at *m*/*z* 1351 $(MH)^+$ of **4** gave large product ions at *m*/*z* 816 and 535 corresponding to the fragments X and Y, respectively (Scheme 2D), suggesting the presence of the 6,7,8,9-tetrahydropyrido[1,2-*a*]indole structure and LCPA in **3**. Furthermore, an acid hydrolysate of **3** showed hallmark molecular ions for **6** at *m*/*z* 1069 and 816 in MALDI-TOFMS (Figure S17, Supporting Information). This result suggested that the product contained the didehydro derivative **6**, supporting an identity of **1** as a substructure of **3**.

The possible molecular formula for **2** and the enzyme digest, fragment E (**5**), was likewise deduced to be $C_{298}H_{518}N_{84}O_{69}S_6$ and $C_{116}H_{236}N_{34}O_7$, respectively (Table S1, Supporting Information).

The difference of the molecular weight between **2** and **3** (or **4** and **5**) corresponds to one LCPA (14-mer) plus C_4H_5 (X in Scheme 1). Although the ^1H NMR for the fragment E (**5**) indicated the spin systems corresponding to those in **1**, detailed NMR peak assignments were not possible due to peak broadening and a minute amount of material available.[1] The

dx.doi.org/10.1021/ol5011888 | *Org. Lett.* 2014, 16, 3090−3093

MS^2 spectrum of **5** showed major product ion at m/z 1351 (MH^+, Figure S19, Supporting Information). An exact mass of the peak agreed with the formula for **4**. Thus, the structure of ACU-A can be constructed by attaching another LCPA unit X (Scheme 1) to the terminal pACU residue, though the structure and site of the extra unit are not clear.

The present result suggested post-translational modification (PTM) of tryptophan by polyamine since the terminal Trp residue was encoded in the cDNA of ACU.[1] Although some PTMs of amino acid by polyamine[10,11] and PTMs of tryptophan residue by sugar[12] are known, the present example differs from any of the known PTMs as it incorporates unusual long-chain polyamine into the Trp via both C−N and C−C bonds. It is thus of particular interest to further investigate biosynthesis of ACUs in the sponge.

■ ASSOCIATED CONTENT

Supporting Information

Spectral data and experimental details. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail ryu.sakai@fish.hokudai.ac.jp.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

NMR was measured by Mr. Katsuhiko Kushida, Agilent Technologies Japan. This research was supported by grant from JSPS to R.S. (22380114) and a grant for 2010-2012 Strategic Research Promotion (nos. T2202, T2309, T2401) of Yokohama City University (MO).

■ REFERENCES

(1) Matsunaga, S.; Jimbo, M.; Gill, M. B.; Wyhe, L. L.; Murata, M.; Nonomura, K.; Swanson, G. T.; Sakai, R. *ChemBioChem* **2011**, *12*, 2191−200.
(2) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. *J. Mass Spectrom.* **2005**, *40*, 430−43.
(3) Zubarev, R. A.; Makarov, A. *Anal. Chem.* **2013**, *85*, 5288−96.
(4) Matsunaga, S.; Sakai, R.; Jimbo, M.; Kamiya, H. *ChemBioChem* **2007**, *8*, 1729−1735.
(5) Pretsch, E.; Bühlmann, P.; Affolter, C. *Tables of Spectral Data for Structure Determination of Organic Compounds*, 3rd ed.; Springer: New York, 1989; p 151.
(6) The reaction mixture that afforded molecular ion for **6** did not show any UV absorption for defined chromophores, indicating that the sample was mostly decomposed by the acid treatment (see Figure S12, Supporting Information).
(7) Attanasi, O. A.; Favi, G.; Filippone, P.; Lillini, S.; Mantellini, F.; Spinelli, D.; Stenta, M. *Adv. Synth. Catal.* **2007**, *349*, 907−915.
(8) Synthetic work for the model compounds will be published elsewhere.
(9) Barfield, M.; Sternhell, S. *J. Am. Chem. Soc.* **1972**, *94*, 1905−1913.
(10) Kröger, N.; Deutzmann, R.; Sumper, M. *Science* **1999**, *286*, 1129−32.
(11) Park, M. H. *J. Biochem.* **2006**, *139*, 161−9.
(12) Hofsteenge, J.; Muller, D. R.; de Beer, T.; Loffler, A.; Richter, W. J.; Vliegenthart, J. F. *Biochemistry* **1994**, *33*, 13524−30.