Enzymatic Synthesis Assisted Discovery of Proline-Rich Macrocyclic Peptides in Marine Sponges

Ipsita Mohanty[+],[a] Nguyet A. Nguyen[+],[a] Samuel G. Moore,[a] Jason S. Biggs,[b] David A. Gaul,[a] Neha Garg,[a] and Vinayak Agarwal*[a, c]

Proline-rich macrocyclic peptides (PRMPs) are natural products present in geographically and phylogenetically dispersed marine sponges. The large diversity and low abundance of PRMPs in sponge metabolomes precludes isolation and structure elucidation of each individual PRMP congener. Here, using standards developed via biomimetic enzymatic synthesis of PRMPs, a mass spectrometry-based workflow to sequence PRMPs was developed and validated to reveal that the diversity of PRMPs in marine sponges is much greater than that has been realized by natural product isolation-based strategies. Findings are placed in the context of diversity-oriented transamidative macrocyclization of peptide substrates in sponge holobionts.

Macrocyclization of ribosomal and non-ribosomal peptides is ubiquitous in natural product chemistry. Macrocyclization imparts pharmacologically desirable properties of stability and rigidity which reduces the entropic cost associated with target binding.[1] From neutrophil defensins and snake venom toxins to cyclic peptide natural products produced by bacteria, across all scales of life, macrocyclic peptides demonstrate potent bioactivities and serve various different biological roles.[2] Benthic marine invertebrates- sponges, are prolific sources of proline rich macrocyclic peptides (PRMPs).[3] PRMPs are detected in marine sponges without restriction of geographical location and are widespread among several sponge phyla. Cell growth inhibitory cytotoxic activities of sponge derived PRMPs has sustained interest in their development as drug candidates.[3] Unlike other natural product chemical classes, the abundance of which dominates the sponge metabolomes, PRMPs in sponges are present in trace amounts. For instance, in *Stylissa* spp. sponges the metabolomes of which are enriched in pyrrole imidazole alkaloids (PIAs),[4] PRMPs were isolated in 0.01 % yields (PRMP weight/sponge dry weight).[5] Low abundance of PRMPs interspersed with much higher abundance other natural products in sponge extracts encumbers their structure elucidation using natural product isolation-based strategies. Here, we devise a mass spectrometry-based workflow to sequence sponge-derived PRMPs and validate the spectrometric sequencing using standards furnished by biomimetic enzymatic peptide macrocyclization.

We previously described the detection of PIAs in *Stylissa* and *Axinella* spp. sponges using LC/MS-based untargeted metabolomics.[6] Screening these LC/MS data using the DEREPLICATOR tool[7] hinted at the presence of peptidic natural products in *Stylissa* and *Axinella* spp. extracts consistent with prior reports describing the isolation of PRMPs from these sponge genera.[3] The low abundance of PRMPs and sub-gram amounts of dry sponge biomass which were available to us precluded a natural product isolation-based effort. Thus, we decided to pursue PRMP structure elucidation from mass spectrometry fragmentation data.

In a proline containing peptide, fragmentation preferentially occurs at the N-terminus of the Pro residue.[8] Sequencing PRMPs detected in sponge extracts thus began with the annotation of the ^{N}Pro-Xaa^{C} dipeptide oxonium MS^2 fragment ions per the inventory shown in Table S1. Proceeding from the ^{N}Pro-Xaa^{C} dipeptide, to recover the PRMP sequence, we annotated mass shifts corresponding to proteinogenic amino acids till we reached the parent ion in the MS^2 spectra. As shown in Figure 1A for a PRMP detected in *Axinella* sp., PRMPs containing more than one proline residue yielded multiple ^{N}Pro-Xaa^{C} dipeptide MS^2 fragments. Progressing from each of these different dipeptide fragments, ^{N}Pro-Phe^{C} (MS^2 *m/z* 245) and ^{N}Pro-Glu^{C} (MS^2 *m/z* 227), the same cyclic sequence, cyclo (FFPE**X**WP), was recovered. Here, **X** denotes Leu/Ile residues that cannot be differentiated based on MS^2 fragmentation alone. Identities of the amino acid constituents in cyclo (FFPE**X**WP), Phe, Pro, Glu, Leu/Ile, and Trp, were supported by the detection of the individual amino acyl immonium ions in the MS^2 spectra (Figure S1, Table S2).

To resolve Leu and Ile residues, we developed standards for the two possible PRMPs- cyclo(FFPE**L**WP) and cyclo(FFPE**I**WP). For this, synthetic linear peptides FFPE**L**WPFQA and FFPE**I**WPFQA were obtained. The C-terminal FQA tripeptide was proteolyzed and transamidatively macrocyclized products were furnished by the plant peptidase PCY1.[9] Upon fragmentation of the cyclo(FFPE**L**WP) and cyclo(FFPE**I**WP) products, the MS^2 immonium ion for Leu/Ile, *m/z* 86, was isolated and fragmented.

[a] *I. Mohanty,[+] N. A. Nguyen,[+] S. G. Moore, Dr. D. A. Gaul, Dr. N. Garg, Dr. V. Agarwal*
School of Chemistry and Biochemistry, Georgia Institute of Technology
Atlanta, GA 30332 (USA)
E-mail: vagarwal@gatech.edu
[b] *Dr. J. S. Biggs*
University of Guam Marine Laboratory
UOG Station, Mangilao, Guam 96923 (USA)
[c] *Dr. V. Agarwal*
School of Biological Sciences, Georgia Institute of Technology
Atlanta, GA 30332 (USA)
[+] *These authors contributed equally to this work.*
Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202100275

© 2021 Wiley-VCH GmbH

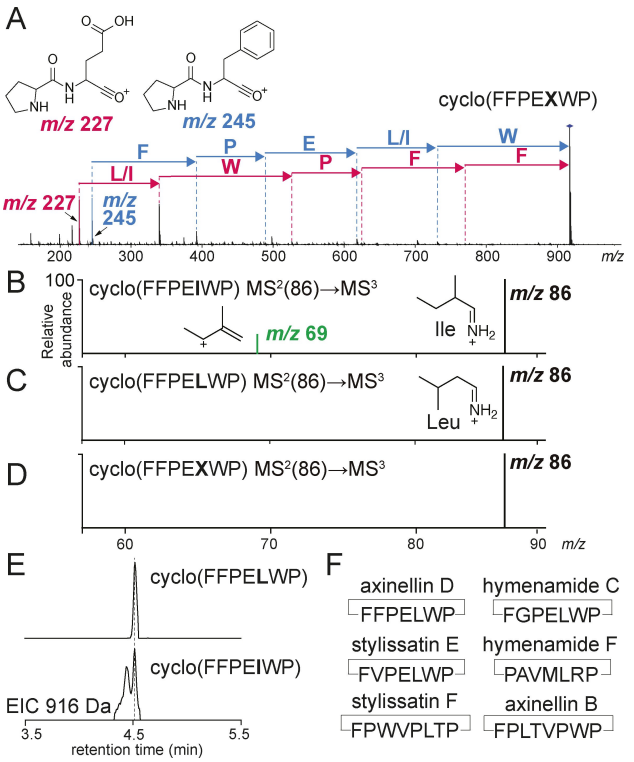


Figure 1. (**A**) MS^2 spectra for cyclo(FFPE**X**WP). $MS^2(86) \rightarrow MS^3$ spectra for the (**B**) Ile-derived MS^2 *m/z* 86 precursor ion (cyclo(FFPE**I**WP) standard), (**C**) Leu-derived MS^2 *m/z* 86 precursor ion (cyclo(FFPE**L**WP) standard), and the (**D**) Leu/Ile-derived MS^2 *m/z* 86 precursor ion (cyclo(FFPE**X**WP) PRMP). (**E**) Extracted ion chromatograms (EICs) showing co-injection of cyclo(FFPELWP) (top) and cyclo(FFPEIWP) (bottom) standards with the sponge extract. (**F**) Inventory of single Leu/Ile containing PRMPs sequenced in this study.

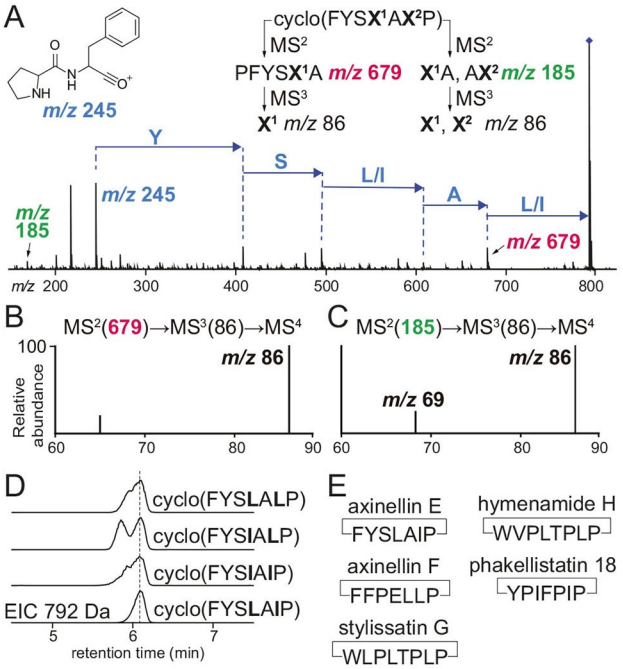


Figure 2. (**A**) MS^2 spectra for the PRMP cyclo(FYS**X**^{1}A**X**^{2}P) detected in *Axinella* sp. where **X** corresponds to Leu/Ile. MS^n fragmentation strategies to query **X**^{1} and **X**^{2} identities are illustrated. (**B**) Confirmation of **X**^{1} as Leu as MS^4 *m/z* 69 product ion is not detected. (**C**) Confirmation of **X**^{2} as Ile as MS^4 *m/z* 69 product ion is detected. (**D**) EICs showing co-injection of enzymatically synthesized PRMP standards with the sponge extract. (**E**) Inventory of multiple Leu/Ile containing PRMPs sequenced in this study.

Consistent with prior reports,[10] fragmentation of the Ile-derived MS^2 *m/z* 86 immonium ion generated a characteristic MS^3 *m/z* 69 product ion in high abundance, while the MS^3 *m/z* 69 product ion was detected in lesser abundance when the MS^2 *m/z* 86 precursor ion was derived from Leu. By modulating the MS^3 fragmentation energy (Figure S2), we generated an unambiguous binary result in which the MS^3 *m/z* 69 product ion was observed only for Ile (Figure 1B), but not observed for Leu (Figure 1C). Adopting the thus optimized methodology for the cyclo(FFPE**X**WP) PRMP detected in the sponge extract, we annotated **X** as Leu (Figure 1D). Validation of the mass spectrometry-based result was obtained by spiking the sponge extract with enzymatically synthesized cyclo(FFPE**L**WP) and cyclo(FFPE**I**WP) standards; the cyclo(FFPE**L**WP) standard coeluted with the sponge-derived PRMP (Figure 1E) which we term as axinellin D. PRMP congeners detected in this study are highlighted in Table S3.

Employing the now validated mass spectrometry workflow which could differentiate between Leu and Ile residues, we additionally sequenced the PRMPs stylissatin E cyclo(FVPELWP) and stylissatin F cyclo(FPWVPLTP) from *Stylissa* sp. specimens used in this study (Figures 1F, S3–6, Table S3). We also detected the presence of the PRMPs cyclo(FGPELWP), cyclo(PAVMLRP), and cyclo(FPLTVPWP) in *Axinella* sp. extracts (Figures S7–12). Primary sequences of these PRMPs correspond to the previously described hymenamide C, hymenamide F, and axinellin B. The presence of axinellin B serves as an internal control as it has been reported from an *Axinella* sponge previously.[11] However, hymenamides C and F were isolated from *Hymeniacidon* sp. sponges (Table S3).[12] *Hymeniacidon* sp. sponges belong to a different phylogenetic order (Suberitidia) as compared to *Axinella* sp. (order Axinellida) within the Demospongiae class of marine sponges.

Abovementioned workflow is applicable when a single Leu/Ile is present in the peptide sequence. When more than one Leu/Ile residues are present, multiple side chains will contribute to MS^2 *m/z* 86 ion which will complicate the $MS^2(86) \rightarrow MS^3$-based Leu/Ile distinction. This scenario presents itself for the cyclo(FYS**X**^{1}A**X**^{2}P) PRMP that we detected in *Axinella* sp.; **X**^{n} denote Leu/Ile residues. As before, the PRMP was sequenced starting from the ^{N}Pro-Phe^C MS^2 *m/z* 245 fragment ion (Figures 2A, S13). Next, the ion corresponding to the fragment PFYS**X**^{1}A, MS^2 *m/z* 679 (Figure 2A) was isolated and the MS^3 *m/z* 86 Leu/Ile immonium ion was generated. Here, only a single residue, **X**^{1}, contributes to the production of the MS^3 *m/z* 86 product ion. Further fragmentation of the MS^3 *m/z* 86 ion did not generate the characteristic MS^4 *m/z* 69 ion (Figure 2B), denoting that **X**^{1} was Leu.

Next, the cyclo(FYS**L**A**L**P), cyclo(FYS**I**A**L**P), cyclo(FYS**L**A**I**P), and cyclo(FYS**I**A**I**P) standards were generated by PCY1-mediated macrocyclization of the corresponding -FQA appended linear

© 2021 Wiley-VCH GmbH

peptides. Fragmenting cyclo(FYSIALP) and cyclo(FYSIAIP) using the $MS^2(679)\rightarrow MS^3(86)\rightarrow MS^4$ procedure yielded the characteristic MS^4 *m/z* 69 ion while fragmentation of cyclo(FYSLALP) and cyclo(FYSLAIP) did not (Figure S14). To query the identity of $\mathbf{X^2}$ in the PRMP cyclo(FYS$\mathbf{X^1}$A$\mathbf{X^2}$P), the MS^2 *m/z* 185 ion was isolated. This MS^2 ion can be generated by the $\mathbf{X^1}$A, or the A$\mathbf{X^2}$ dipeptides. The $MS^2(185)\rightarrow MS^3(86)\rightarrow MS^4$ fragmentation scheme demonstrated the production of the characteristic MS^4 *m/z* 69 ion (Figure 2C). As $\mathbf{X^1}$ was Leu (and hence the $\mathbf{X^1}$A fragment could not contribute to the production of the MS^4 *m/z* 69 product ion), these data denote that $\mathbf{X^2}$ was Ile. Analogous $MS^2(185)\rightarrow MS^3(86)\rightarrow MS^4$ fragmentation for the cyclo(FYSLAIP) standard generated the MS^4 *m/z* 69 ion while the cyclo (FYSLALP) standard did not (Figure S15). Validation of the mass spectrometry-based PRMP sequencing scheme was obtained by spiking the sponge extract with the PCY1 generated standards; the cyclo(FYSLAIP) standard coeluted with the sponge PRMP which is termed axinellin E (Figure 2D, Table S3).

Using the workflow developed above, we additionally sequenced the multiple Leu/Ile containing PRMPs axinellin F cyclo(FFPELLP) from *Axinella* sp., and stylissatin G cyclo (WLPLTPLP) from *Stylissa* sp. (Figures 2E, S16–19, Table S3). In addition to these PRMPs, we detected the presence of the previously described PRMPs hymenamide H[13]- cyclo (WVPLTPLP) and phakellistatin 18[14]- cyclo(YPIFPIP) from *Stylissa* sp. (Figure 2E, S20–23). Hymenamide H was described from *Hymeniacidon* sp. (order Suberitidia) while phakellistatin 18 was described from *Phakellia fusca* (order Axinellida), sponges belonging to orders different from *Stylissa* sp. (order Scopalinida). Together with the detection of hymenamide C in *Axinella* sp., it is evident that PRMP congeners are shared among phylogenetically different sponges.

Unlike axinellin F, cyclo(FFPELLP), where Ile residues occur next to Leu/Ile residues in a PRMP sequence, mass spectrometry alone may be insufficient to provide unambiguous results. This scenario presented itself during the sequencing of the PRMP cyclo($\mathbf{X^1}$PHPY$\mathbf{X^2X^3}$GP) that we detected in *Axinella* sp. As before, the PRMP was sequenced starting from the ^{N}Pro-His^C (*m/z* 235), ^{N}Pro-Tyr^C (*m/z* 261), and ^{N}Pro-$\mathbf{X^1}^C$ (*m/z* 211) dipeptide MS^2 ions (Figure 3A). Using PCY1 and appropriately designed linear peptide substrates, we generated standards for all eight possible Leu/Ile combinations for cyclo($\mathbf{X^1}$PHPY$\mathbf{X^2X^3}$GP). Residue $\mathbf{X^1}$ was annotated as Ile by isolating the MS^2 *m/z* 211 ion corresponding to the P$\mathbf{X^1}$ dipeptide and following the $MS^2(211)\rightarrow MS^3(86)\rightarrow MS^4$ fragmentation scheme (Figure 3B). Annotation of $\mathbf{X^1}$ as Ile was supported by analogous fragmentation of the cyclo(LPHPY$\mathbf{X^2X^3}$GP) and cyclo(IPHPY$\mathbf{X^2X^3}$GP) standards (Figure S24–25). Similarly, identity of $\mathbf{X^2}$ as Ile was decided by isolating the MS^2 *m/z* 608 ion corresponding to the PHPY$\mathbf{X^2}$ pentapeptide and following the $MS^2(608)\rightarrow MS^3(86)\rightarrow MS^4$ fragmentation scheme (Figure 3C) and comparing the MS^4 spectra to cyclo(IPHPYL$\mathbf{X^3}$GP) and cyclo(IPHPYI$\mathbf{X^3}$GP) standards (Figure S26). Because $\mathbf{X^2}$ was Ile (which generates the MS^4 *m/z* 69 diagnostic ion) and no MS^2 fragment could be detected which contained only $\mathbf{X^3}$ without $\mathbf{X^1}$ or $\mathbf{X^2}$, identity of $\mathbf{X^3}$ could not be determined by mass spectral fragmentation. To annotate $\mathbf{X^3}$, retention times for the cyclo(IPHPYILGP) (X^1,X^2 = Ile; X^3 = Leu)

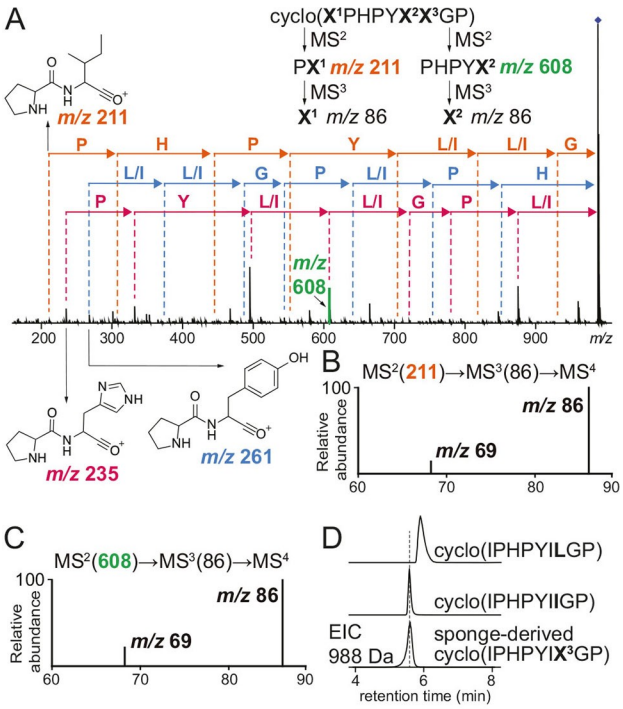


Figure 3. (**A**) Sequencing of the cyclo($\mathbf{X^1}$PHPY$\mathbf{X^2X^3}$GP) based on MS^2 spectra. Note that three Pro residues in this octapeptide generate three different ^{N}Pro-Xaa^C MS^2 dipeptide product ions. Fragmentation strategies to query identities of $\mathbf{X^1}$ and $\mathbf{X^2}$ residues are illustrated. Annotation of (**B**) $\mathbf{X^1}$ and (**C**) $\mathbf{X^2}$ residues as Ile by the detection of the MS^4 *m/z* 69 product ion. (**D**) Annotation of $\mathbf{X^3}$ as Ile by retention time comparison of the sponge detected PRMP against enzymatically synthesized cyclo(IPHPYILGP) and cyclo (IPHPYIIGP) standards.

and cyclo(IPHPYIIGP) (X^1,X^2 = Ile; X^3 = Ile) standards were compared to reveal the sponge PRMP as cyclo(IPHPYIIGP) which we term axinellin G (Figure 3D, Table S3). Among the inventory of PRMPs, cyclo-nonapeptides such as axinellin G are relatively rare (Table S3). Overall, we demonstrate the utility of enzymatic synthesis of cyclic peptide standards for discovery and sequencing of sponge PRMPs. At present, we cannot determine the *cis/trans* conformation of the Pro residues.

Amino acid residues reported to date for sponge PRMPs are proteinogenic amino acids in the L- configuration (to the best of our knowledge the only exception is Trp derived L-kynurenine in phakefustatins A–C[15]). It is thus conceivable that sponge-derived PRMPs are ribosomally synthesized and post-translationally modified peptides[16] and that the PCY1 catalyzed proteolysis of a C-terminal recognition sequence (RS in Figure 4A; PCY1 RS is FQA[9]) and intramolecular transamidative macrocyclization mimics their physiological biosynthetic route. Peptide macrocyclization by prolyl oligopeptidases and cyanobactin macrocyclases is well established.[16] A recurring feature in biosynthesis of macrocyclic peptides is the requirement of a proline residue or a Ser/Thr/Cys-derived azol(in)e heterocycle immediately upstream of the RS (shaded green in Figure 4A).[9,17] Organizing the sponge PRMP octa- and heptapeptide sequences containing a single Pro residue such that the Pro is at the C-terminus leads to the observation that the transamidating N-

© 2021 Wiley-VCH GmbH

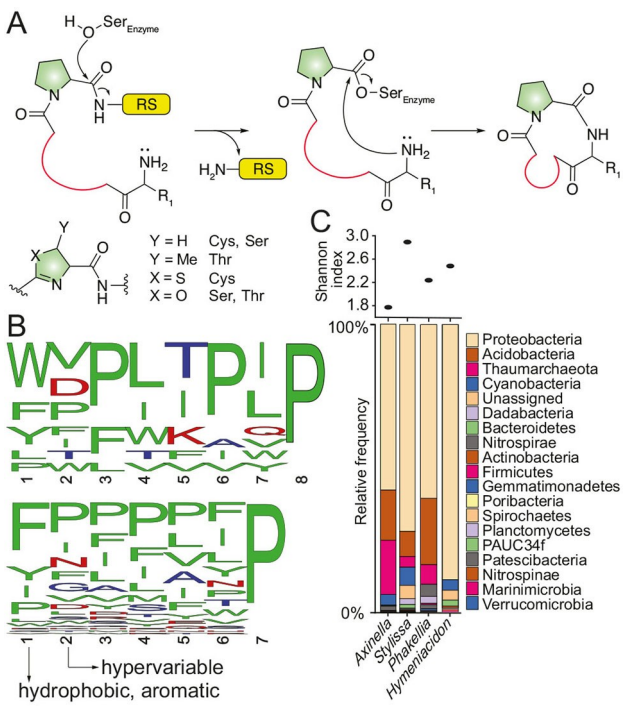


Figure 4. (**A**) Intramolecular transamidative peptide cyclization catalyzed by serine proteases such as PCY1. (**B**) WebLogo representation of PRMP octa- (top) and heptapeptides (bottom) with a C-terminal Pro. (**C**) Representative phylum-level microbiome architectures for *Axinella*,[6] *Stylissa*,[6] *Phakellia*,[20] and *Hymeniacidon*[20] spp. sponges. Microbiome diversity is denoted by Shannon indices. Shannon indices for HMA sponges are typically greater than 4.

terminal residues in PRMP linear peptides are hydrophobic amino acids with Trp, Phe, and Tyr predominating (Figure 4B). Further organizing PRMP sequences containing multiple Pro residues in this sequence alignment then identifies position 2 to be a site for hypervariability.

It is as yet unknown whether the sponge host or a bacterial symbiont produces PRMPs in sponge holobionts. To date, natural products that have been shown to be produced by bacterial symbionts within marine sponges are all derived from high microbial abundance (HMA) sponges.[18] As microbial diversity correlates with microbial abundance for sponge microbiomes that are not dominated by cyanobacteria,[19] PRMP harboring sponges of the genera- *Axinella*, *Stylissa*, *Phakellia*, and *Hymeniacidon* spp. are all LMA sponges (Figure 4C).

The diversity of natural products in marine sponges is much greater than that has been realized.[21] Here, with the sequencing of 12 macrolactam cyclopeptide sequences of which seven were novel PRMP sequences including a rare cyclononapeptide (Table S3), we demonstrate that inventorying natural product diversity using mass spectrometry and supplementing structural information using enzymatic synthesis can rapidly expand the natural product chemical space without recourse to sacrificing large amounts of sponge biomass which is otherwise necessary for the isolation of low abundance molecules.

Acknowledgements

The authors acknowledge support from the National Science Foundation (NSF, CHE-2004030) and National Institutes of Health (NIH, GM142882) to V.A., and support from the Georgia Institute of Technology's Systems Mass Spectrometry Core Facility. The authors also thank Dr. S. K. Nair for the gift of PCY1 expression plasmid.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: chemoenzymatic synthesis · cyclopeptides · fragmentation · marine sponges

[1] a) A. Zorzi, K. Deyle, C. Heinis, *Curr. Opin. Chem. Biol.* **2017**, *38*, 24–29; b) E. M. Driggers, S. P. Hale, J. Lee, N. K. Terrett, *Nat. Rev. Drug Discovery* **2008**, *7*, 608–624.
[2] T. B. Andrew, M. M. Cayla, R. S. Lokey, *Curr. Top. Med. Chem.* **2013**, *13*, 821–836.
[3] W.-Y. Fang, R. Dahiya, H.-L. Qin, R. Mourya, S. Maharaj, *Mar. Drugs* **2016**, *14*.
[4] A. Al-Mourabit, M. A. Zancanella, S. Tilvi, D. Romo, *Nat. Prod. Rep.* **2011**, *28*, 1229–1260.
[5] a) G. Schmidt, A. Grube, M. Köck, *Eur. J. Org. Chem.* **2007**, 4103–4110; b) C. Cychon, M. Köck, *J. Nat. Prod.* **2010**, *73*, 738–742; c) X. Wang, B. I. Morinaka, T. F. Molinski, *J. Nat. Prod.* **2014**, *77*, 625–630; d) M. Arai, Y. Yamano, M. Fujita, A. Setiawan, M. Kobayashi, *Bioorg. Med. Chem. Lett.* **2012**, *22*, 1818–1821.
[6] I. Mohanty, S. G. Moore, D. Yi, J. S. Biggs, D. A. Gaul, N. Garg, V. Agarwal, *ACS Chem. Biol.* **2020**, *15*, 2185–2194.
[7] H. Mohimani, A. Gurevich, A. Shlemov, A. Mikheenko, A. Korobeynikov, L. Cao, E. Shcherbin, L.-F. Nothias, P. C. Dorrestein, P. A. Pevzner, *Nat. Commun.* **2018**, *9*, 4035.
[8] a) L. A. Breci, D. L. Tabb, J. R. Yates, V. H. Wysocki, *Anal. Chem.* **2003**, *75*, 1963–1971; b) E. A. Kapp, F. Schütz, G. E. Reid, J. S. Eddes, R. L. Moritz, R. A. J. O'Hair, T. P. Speed, R. J. Simpson, *Anal. Chem.* **2003**, *75*, 6251–6264.
[9] H. Ludewig, C. M. Czekster, E. Oueis, E. S. Munday, M. Arshad, S. A. Synowsky, A. F. Bent, J. H. Naismith, *ACS Chem. Biol.* **2018**, *13*, 801–811.
[10] a) Y. Xiao, M. M. Vecchi, D. Wen, *Anal. Chem.* **2016**, *88*, 10757–10766; b) T. B. Moyer, N. C. Parsley, P. W. Sadecki, W. J. Schug, L. M. Hicks, *Nat. Prod. Rep.* **2021**, *38*, 489–509.
[11] A. Randazzo, F. Dal Piaz, S. Orrù, C. Debitus, C. Roussakis, P. Pucci, L. Gomez-Paloma, *Eur. J. Org. Chem.* **1998**, 2659–2665.
[12] a) M. Tsuda, H. Shigemori, Y. Mikami, J. i. Kobayashi, *Tetrahedron* **1993**, *49*, 6785–6796; b) J. i. Kobayashi, T. Nakamura, M. Tsuda, *Tetrahedron* **1996**, *52*, 6355–6360.
[13] M. Tsuda, T. Sasaki, J. i. Kobayashi, *Tetrahedron* **1994**, *50*, 4667–4680.
[14] H.-J. Zhang, Y.-H. Yi, G.-J. Yang, M.-Y. Hu, G.-D. Cao, F. Yang, H.-W. Lin, *J. Nat. Prod.* **2010**, *73*, 650–655.
[15] Y. Wu, H. Liao, L.-Y. Liu, F. Sun, H.-F. Chen, W.-H. Jiao, H.-R. Zhu, F. Yang, G. Huang, D.-Q. Zeng, M. Zhou, S.-P. Wang, H.-W. Lin, *Org. Lett.* **2020**, *22*, 6703–6708.
[16] M. Montalbán-López, T. A. Scott, S. Ramesh, I. R. Rahman, A. J. van Heel, J. H. Viel, V. Bandarian, E. Dittmann, O. Genilloud, Y. Goto, M. J. Grande Burgos, C. Hill, S. Kim, J. Koehnke, J. A. Latham, A. J. Link, B. Martínez, S. K. Nair, Y. Nicolet, S. Rebuffat, H.-G. Sahl, D. Sareen, E. W. Schmidt, L. Schmitt, K. Severinov, R. D. Süssmuth, A. W. Truman, H. Wang, J.-K. Weng, G. P. van Wezel, Q. Zhang, J. Zhong, J. Piel, D. A. Mitchell, O. P. Kuipers, W. A. van der Donk, *Nat. Prod. Rep.* **2021**, *38*, 130–239.
[17] a) S. Sarkar, W. Gu, E. W. Schmidt, *ACS Catal.* **2020**, *10*, 7146–7153; b) H. Luo, S.-Y. Hong, R. M. Sgambelluri, E. Angelos, X. Li, J. D. Walton, *Chem. Biol.* **2014**, *21*, 1610–1617; c) D. Sardar, Z. Lin, E. W. Schmidt, *Chem. Biol.* **2015**, *22*, 907–916.

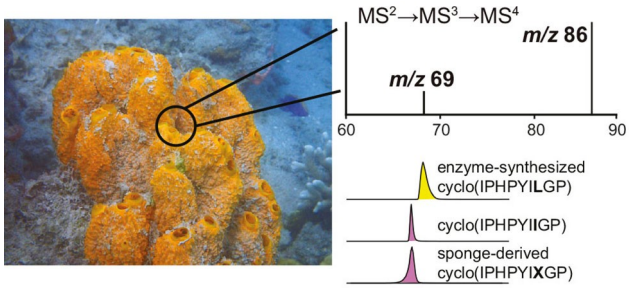
© 2021 Wiley-VCH GmbH

[18] a) M. Morita, E. W. Schmidt, *Nat. Prod. Rep.* **2018**, *35*, 357–378; b) E. P. McCauley, I. C. Piña, A. D. Thompson, K. Bashir, M. Weinberg, S. L. Kurz, P. Crews, *J. Antibiot.* **2020**, *73*, 504–525.
[19] L. Moitinho-Silva, G. Steinert, S. Nielsen, C. C. P. Hardoim, Y.-C. Wu, G. P. McCormack, S. López-Legentil, R. Marchant, N. Webster, T. Thomas, U. Hentschel, *Front. Microbiol.* **2017**, *8*, 752.
[20] L. Moitinho-Silva, S. Nielsen, T. Thomas, J. J. Bell, J. Vicente, J. R. Björk, J. M. Montoya, J. B. Olson, J. Reveillaud, L. Steindler, M.-C. Pineda, N. S. Webster, M. Ilan, P. M. Erwin, S. Lopez-Legentil, A. Amir, A. Gonzalez, G. L. Ackermann, R. Knight, P. J. Schupp, R. L. Simister, R. W. Thacker, R. Costa, R. T. Hill, T. Ravasi, U. Hentschel, C. Cerrano, C. Astudillo-Garcia, M. W. Taylor, C. Easson, D. Sipkema, G. Steinert, F. Liu, G. Feng, Z. Li, G. Kotoulas, P. Polymenakou, T. Dailianis, G. P. McCormack, M. V. Marra, *Gigascience* **2017**, *6*, 1–7.
[21] a) M. Reverter, S. Rohde, C. Parchemin, N. Tapissier-Bontemps, P. J. Schupp, *Front. Mar. Sci.* **2020**, *7*, 1062; b) V. J. Paul, C. J. Freeman, V. Agarwal, *Integr. Comp. Biol.* **2019**, *59*, 765–776.

Manuscript received: June 9, 2021
Revised manuscript received: June 25, 2021
Accepted manuscript online: June 29, 2021
Version of record online: ■■, ■■■■

© 2021 Wiley-VCH GmbH

COMMUNICATIONS



*I. Mohanty, N. A. Nguyen, S. G. Moore, Dr. J. S. Biggs, Dr. D. A. Gaul, Dr. N. Garg, Dr. V. Agarwal**

1 – 6

Enzymatic Synthesis Assisted Discovery of Proline-Rich Macrocyclic Peptides in Marine Sponges

By adopting an integrated approach of chemoenzymatic synthesis of standards and HCD-MS^n fragmentation, we have sequenced proline-rich cyclopeptide natural products from marine sponges. Our MS^n fragmentation workflow successfully differentiated between leucine and isoleucine residues in the cyclopeptides and can be used to sequence low abundance cyclic peptides from sub-gram amounts of sponge biomass.

SUPPLEMENTARY INFORMATION

Supplementary Materials and Methods

Supplementary Tables 1–43

Supplementary Figures 1–26

Supplementary References

SUPPLEMENTARY MATERIALS AND METHODS

Expression and purification of PCY1 enzyme

Purification of PCY1 was adopted from the literature.[1] Overexpression of PCY1 was performed in *E. coli* BL21(DE3) grown in terrific broth media. Cultures were inoculated and grown at 30 °C until the OD_{600} reached 0.6. Cultures were cooled at 18 °C for 1 h prior to induction of protein expression by the addition of 0.3 mM isopropyl-β-d-thiogalactopyranoside (IPTG). Bacteria were grown for 20 h at 18 °C, and cells were harvested by centrifugation at 7,000×g for 25 min at 4 °C. Cell pellets were resuspended in lysis buffer (20 mM Tris-Cl (pH 8.0), 500 mM NaCl). Cells were lysed by sonication and the lysate was clarified by centrifugation at 30,000×g for 45 min at 4 °C. The supernatant was loaded on to a 5 mL His-Trap Ni-NTA column. Column was washed extensively with wash buffer (20 mM Tris-Cl (pH 8.0), 500 mM NaCl, 30 mM imidazole) and protein was eluted using a linear gradient to 100% elution buffer (20 mM Tris-Cl (pH 8.0), 500 mM NaCl, 250 mM imidazole). Purity of eluent fractions were checked by SDS-PAGE, and fractions containing protein of interest were pooled and dialyzed into 1 L of dialysis buffer (20 mM Tris-Cl (pH 8.9), 100 mM KCl, 10% glycerol) at 4 °C overnight. Protein concentrations were measured by Bradford assay and aliquots were frozen in liquid nitrogen and stored at -80 °C.

Metabolite extraction and reverse-phase LC/MS data acquisition

Stylissa sp. sponge specimens were collected from Anae Island in Guam in July 2015 at a depth of 10 feet. *Axinella* sp. sponge specimens were collected from the Solomon Islands in April 2018 at a depth of 20 feet. Frozen sponge tissues were lyophilized to dryness and metabolites extracted in 1:1 v/v DCM/MeOH (1 mL solvent/100 mg dry sponge tissue) for 48 h at room temperature. The organic extract was clarified by centrifugation, dried, resuspended in MeOH, and analyzed using an Agilent 1290 Infinity II ultra-performance liquid chromatography (UPLC) coupled to a Bruker ImpactII ultra-high-resolution

Qq-ToF mass spectrometer equipped with an electron spray ionization (ESI) source. A Kinetex 1.7 µm C_{18} reversed phase UPLC column (50×2.1 mm) was used for chromatographic separation. Data were acquired in positive ionization mode, *m/z* 50–2000 Da. An active exclusion of four spectra was employed. Chromatography solvent A: water + 0.1% v/v formic acid, solvent B: MeCN + 0.1% v/v formic acid. Flow rate was held constant at 0.5 mL/min. The elution profile employed was as follows: 2% solvent B for 3 min, a linear gradient from 2% to 50% B in 5 min, 50% B for 2 min, from 50% to 100% B in 5 min, 100% B for 3 min, from 100% to 5% B in 1 min, 5% B for 1 min, from 5% to 100% B in 1 min, 100% B for 3 min, from 100% to 5% B in 1 min, 5% B for 5 min. LC/MS data were converted to lock mass corrected mzXML format using Bruker DataAnalysis software.

Peptide macrocyclization assays with PCY1

Peptide macrocyclization assays were set up using previously described conditions[1]: 300 µM substrate peptide, 2 µM PCY1, 20 mM CAPSO (pH 9.5), 100 mM NaCl, 2 mM DMSO and 5 mM DTT at 30 °C. After 2 h, assays were quenched by addition of equal volume of 5% v/v trichloroacetic acid (TCA) in MeCN and centrifugated to remove precipitates. Samples were analyzed using 1290 Infinity II UPLC system (Agilent Technologies) coupled to a high-resolution Impact II Qq-TOF mass spectrometer (Bruker Daltonics). Mass spectrometry data were collected in the positive ionization mode in the mass range *m/z* 100–2000 Da. A Kinetex™ 1.7 µm C18 reversed phase UHPLC column (50 × 2.1 mm) at a flow rate of 0.5 mL/min was employed and the chromatographic separation was achieved using solvent A: H_2O + 0.1% (v/v) formic acid, and solvent B: MeCN + 0.1% (v/v) formic acid. The chromatography profile was as follows: 5% solvent B from 0–5 min, linear gradient to 100% solvent B from 5–27 min, 100% solvent B from 27–29 min, linear gradient to 5% solvent B from 29–30 min, 5% solvent B from 30–31 min, linear gradient to 100% solvent B from 31–32 min, 100% solvent B from 32–33 min, linear gradient to 5% solvent B from 33–34 min, and 5% solvent B from 34–35 min.

LC-MS method development for one Leu/Ile containing PRMPs

LC/MS data were acquired using a Waters Corporation CORTECS UPLC T3 column (2.1×150 mm, 1.6 µm particle size) coupled to a high-resolution accurate mass Orbitrap ID-X tribrid mass spectrometer. The chromatographic method for sample analysis involved elution with water with 10 mM ammonium formate and 0.1% formic acid (mobile phase A) and MeCN and 0.1% formic acid (mobile phase B) using the following gradient program: 0 min 10% B; 0.5 min 10% B; 8 min 97% B; 10.9 min

97% B; 11 min 10% B; 12 min 10% B. The flow rate was set at 0.4 mL/min. The column temperature was set to 40 °C, and the injection volume was 2.0 µL.

The Orbitrap ID-X is a tribrid spectrometer that utilizes quadrupole isolation with dual detectors, an orbitrap and an ion trap, with a maximum resolving power of 500,000 full width at half maximum (FWHM) at *m/z* 200 and mass accuracy of <1 ppm. The heated electrospray ionization (HESI) source was operated at a vaporizer temperature of 275 ºC, a spray voltage of 3.5 kV, and sheath, auxiliary, and sweep gas flows of 40, 8, and 1 in arbitrary units, respectively. The instrument acquired full MS data in the 100–1000 *m/z* range in positive ionization mode with an orbitrap resolution of 120,000. Targeted MS^3 data was acquired in the ion trap, with the MS^1 precursor ions isolated in the quadrupole with an isolation window of 1.6 *m/z* and activated at Higher-energy collisional dissociation (HCD) energy of 55%; and the MS^2 precursor ions isolated in the ion trap with an isolation window of 2 *m/z* and activated at HCD collision energy of 8%. This method was conserved across all single Leu/Ile containing cyclopeptides reported in this study.

For retention time matching and spiking experiments, we used the Thermo Scientific Accucore Phenyl-X column (2.1 x 50 mm, 2.6 µM particle size) to achieve better chromatographic separation. The chromatographic method involved elution with water with 10 mM ammonium formate and 0.1% formic acid (mobile phase A) and MeCN and 0.1% formic acid (mobile phase B) using the following gradient program: 0 min 10% B; 0.5 min 25% B; 4.7 min 60% B; 5 min 97% B; 7 min 97% B; 7.1 min 10% B; 8 min 10% B. The flow rate was set at 0.5 mL/min. The column temperature was set to 40 °C, and the injection volume was 5.0 µL.

LC-MS method development for two Leu/Ile containing PRMPs

LC/MS data were acquired using a Thermo Scientific Accucore C8 column (2.1×150 mm, 2.6 µm particle size) coupled to a high-resolution accurate mass Orbitrap ID-X tribrid mass spectrometer. The chromatographic method for sample analysis involved elution with water with 10 mM ammonium acetate and 0.1% acetic acid (mobile phase A) and 90:10 isopropanol:water with 10 mm ammonium acetate (mobile phase B) using the following gradient program: 0 min 5% B; 0.2 min 10% B; 1 min 20% B; 7.2 min 40% B; 9 min 97% B; 10.9 min 97% B; 11 min 5% B; 12 min 5% B. The flow rate was set at 0.5 mL/min. The column temperature was set to 40 °C, and the injection volume was 5.0 µL.

The orbitrap was operated with the same ESI source parameters as described in previous section. The instrument acquired full MS data in the 100–1000 *m/z* range in positive ionization mode with an orbitrap resolution of 120,000. Targeted MS^2 data were collected for the MS^1 precursor ions

corresponding to the cyclopeptide with both activation by HCD collision energy at 30% and CID collision energy at 35%. This was done to ensure detection of most of the MS^2 fragments required for downstream fragmentation analysis. Based on the selection of fragment ions detected in the MS^2 spectra, the hybrid tMS^3 and dd-MS^4 method was developed for each of the selected MS^2 precursor ions. For the *m/z* 679 Da MS^2 precursor ion for PRMP cyclo(FYS$\mathbf{X^1}$A$\mathbf{X^2}$P), targeted MS^3 data were acquired in the ion trap, with the MS^1 precursor ions isolated in the quadrupole with an isolation window of 1.6 *m/z* and activated at CID collision energy of 35% (CID activation time of 10 ms); and the MS^2 precursor ions were isolated in the ion trap with an isolation window of 2 *m/z* and activated at HCD collision energy of 55%. MS^4 data were acquired in a data-dependent acquisition method, with the MS^3 precursor ions activated with HCD collision energy of 15% and isolated with an isolation window of 2 *m/z*. For the *m/z* 185 Da MS^2 precursor ion for cyclo(FYS$\mathbf{X^1}$A$\mathbf{X^2}$P), targeted MS^3 data were acquired in the ion trap, with the MS^1 precursor ions isolated in the quadrupole with an isolation window of 1.6 *m/z* and activated at HCD collision energy of 35%; and the MS^2 precursor ions isolated in the ion trap with an isolation window of 2 *m/z* and activated at HCD collision energy of 55%. dd-MS^4 data were acquired with the MS^3 precursor ions activated with HCD collision energy of 15 % and isolated with an isolation window of 3 *m/z*.

For PRMP cyclo(FFPELLP), two fragments *m/z* 340 Da and *m/z* 211 Da were selected for tMS^3 and dd-MS^4 analysis. For both the ions, tMS^3 data were acquired in the ion trap, with the MS^1 precursor ions isolated in the quadrupole with an isolation window of 5 *m/z* and activated at HCD collision energy of 40%; and the MS^2 precursor ions isolated in the ion trap with an isolation window of 5 *m/z* and activated at HCD collision energy of 55%. dd-MS^4 data were acquired with the MS^3 precursor ions activated with HCD collision energy of 15% and isolated with an isolation window of 5 *m/z*.

For PRMP cyclo(WVPLTPLP), two fragments *m/z* 694 Da and *m/z* 308 Da were selected for tMS^3 and dd-MS^4 analysis. For the *m/z* 694 Da fragment, tMS^3 data were acquired in the ion trap, with the MS^1 precursor ions isolated in the quadrupole with an isolation window of 5 *m/z* and activated at CID collision energy of 35%; and the MS^2 precursor ions isolated in the ion trap with an isolation window of 5 *m/z* and activated at HCD collision energy of 55%. dd-MS^4 data were acquired with the MS^3 precursor ions activated with HCD collision energy of 15% and isolated with an isolation window of 5 *m/z*. For the *m/z* 308 Da fragment, tMS^3 data were acquired in the ion trap, with the MS^1 precursor ions isolated in the quadrupole with an isolation window of 5 *m/z* and activated at HCD collision energy of 30%; and the MS^2 precursor ions isolated in the ion trap with an isolation window of 5 *m/z* and activated at HCD collision energy of 55%. dd-MS^4 data were acquired with the MS^3 precursor ions activated with HCD collision energy of 15% and isolated with an isolation window of 5 *m/z*.

For PRMP cyclo(YPIFPIP), two fragments *m/z* 455 Da and *m/z* 308 Da were selected for tMS^3 and dd-MS^4 analysis. For both *m/z* 455 Da and *m/z* 308 Da, tMS^3 data were acquired in the ion trap, with the MS^1 precursor ions isolated in the quadrupole with an isolation window of 5 *m/z* and activated at HCD collision energy of 30%; and the MS^2 precursor ions isolated in the ion trap with an isolation window of 5 *m/z* and activated at HCD collision energy of 55%. dd-MS^4 data were acquired with the MS^3 precursor ions activated with HCD collision energy of 15% and isolated with an isolation window of 5 *m/z*.

Samples for the spiking experiment were prepared by adding crude sponge extract to the synthetic standards in 1:10 (v/v) ratio and the data were acquired using a Thermo Scientific Accucore C8 column (2.1×150 mm, 2.6 µm particle size) with the same LC gradient as described previously. **LC-MS method development for three Leu/Ile containing PRMPs**

LC/MS data were acquired using a Thermo Scientific Accucore Phenyl-X column (2.1 x 50 mm, 2.6 µM particle size) coupled to a high-resolution accurate mass Orbitrap ID-X tribrid mass spectrometer. The chromatographic method for sample analysis involved elution with water with 10 mM ammonium formate and 0.1% formic acid (mobile phase A) and MeCN and 0.1% formic acid (mobile phase B) using the following gradient program: 0 min 20% B; 7 min 100% B; 10.5 min 100% B; 10.7 min 20% B; 12 min 20% B. The flow rate was set at 0.4 mL/min. The column temperature was set to 40 °C, and the injection volume was 2.0 µL.

The orbitrap was operated with the same ESI source parameters as described in previous sections. The instrument acquired full MS data in the 100–1000 *m/z* range in positive ionization mode with an orbitrap resolution of 120,000. tMS^2, tMS^3 and dd-MS^4 data were acquired for the MS^1 precursor ion corresponding to the cyclopeptide similar to the method as described above for two Leu/Ile containing cyclopeptides. For the *m/z* 875 Da and *m/z* 608 Da MS^2 precursor ions for cyclo(**X^1**PHPY**X^2X^3**GP), targeted MS^3 data were acquired in the ion trap, with the MS^1 precursor ions isolated in the quadrupole with an isolation window of 1.6 *m/z* and activated at CID collision energy of 35% (CID activation time of 10 ms); and the MS^2 precursor ions isolated in the ion trap with an isolation window of 2 *m/z* and activated at HCD collision energy of 55%. MS^4 data were acquired in a data-dependent acquisition mode, with the MS^3 precursor ions activated with HCD collision energy of 15% and isolated with an isolation window of 2 *m/z*. For the *m/z* 211 Da MS^2 precursor ion, tMS^3 data were acquired in the ion trap, with the MS^1 precursor ions isolated in the quadrupole with an isolation window of 1.6 *m/z* and activated at HCD collision energy of 40%; and the MS^2 precursor ions isolated in the ion trap with an isolation window of 2 *m/z* and activated at HCD collision energy of 55%. dd-MS^4 data were acquired with the MS^3 precursor ions activated with HCD collision energy of 15% and isolated with an isolation window of 5 *m/z*.

For the cyclo(WLPLTPLP) PRMP, three MS^2 fragment ions, *m/z* 308, *m/z* 312, and *m/z* 397 Da were selected for tMS^3 and dd-MS^4 analysis. The same method was used for all the precursor ions with tMS^3 data acquired in the ion trap, and the MS^1 precursor ions isolated in the quadrupole with an isolation window of 1.6 *m/z* and activated at HCD collision energy of 40%; and the MS^2 precursor ions isolated in the ion trap with an isolation window of 2 *m/z* and activated at HCD collision energy of 55%. dd-MS^4 data were acquired with the MS^3 precursor ions activated with HCD collision energy of 15% and isolated with an isolation window of 5 *m/z*.

For the retention time matching experiment, data was acquired on a Thermo Scientific Accucore C30 column (2.1×150 mm, 2.6 µm particle size) for better resolving power between the standards. The chromatographic method for sample analysis involved elution with water with 10 mM ammonium acetate and 0.1% acetic acid (mobile phase A) and 90:10 isopropanol:water with 10 mm ammonium acetate (mobile phase B) using the following gradient program: 0 min 15% B; 1 min 15% B; 7 min 32% B; 7.8 min 100% B; 10.5 min 100% B; 10.7 min 15% B; 12 min 15% B. The flow rate was set at 0.4 mL/min. The column temperature was set to 40 °C, and the injection volume was 2.0 µL.

Microbiome sequencing and analyses

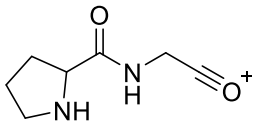
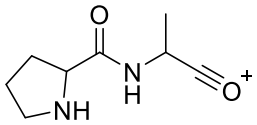
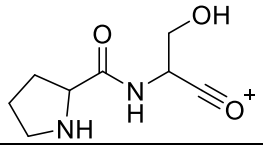
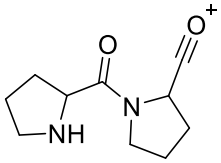
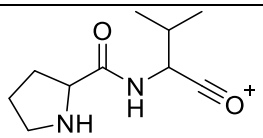
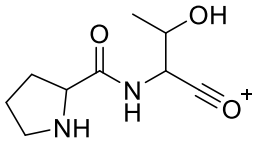
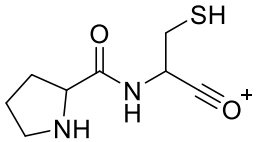
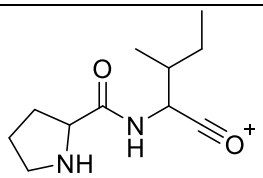
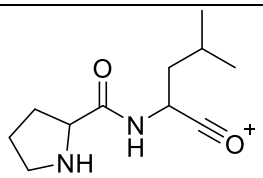
The procedure for querying the sponge-associated microbiome was based on our previous reports.[2] The raw sequence reads were demultiplexed and sequence variants (SVs) generated by QIIME2 using the qiime tools import script, qiime demux script and the DADA2 plugin,[3] respectively. Based on quality scores, the forward and reverse reads were truncated at 180 bp using the qiime dada2 denoise script. Taxonomy was assigned using the SILVA pre-trained classifier using the qiime feature classifier plug-in.[4] The qiime taxa barplot script was used to generate the bar plots representing the taxonomic distribution. The α-diversity of the microbiome was quantified by computing Shannon index using the q2_diversity plugin in QIIME2.

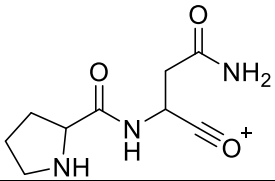
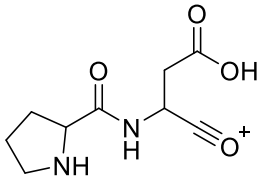
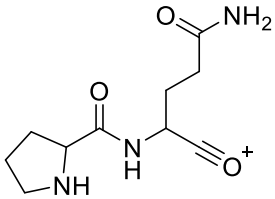
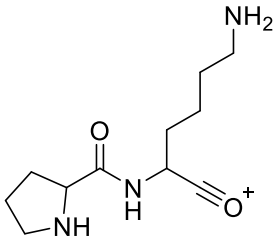
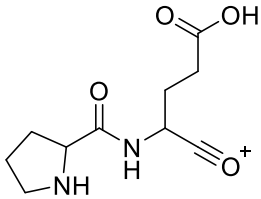
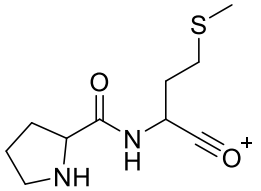
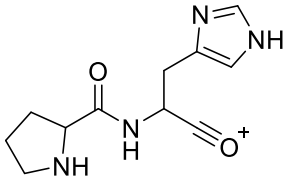
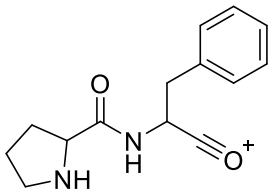
Data availability

Mass spectrometry data has been deposited to the MassIVE repository (**https://massive.ucsd.edu/**) under the accession code MSV000087277.

SUPPLEMENTARY TABLES

Table S1: $[M+H]^{1+}$ masses and structures of the ^{N}Pro-Xaa^{C} dipeptide MS^2 ions

^{N}Pro-Xaa^{C} $[M+H]^{1+}$	**Structure**
^{N}Pro-Gly^{C} 155.0815	
^{N}Pro-Ala^{C} 169.0972	
^{N}Pro-Ser^{C} 185.0921	
^{N}Pro-Pro^{C} 195.1128	
^{N}Pro-Val^{C} 197.1285	
^{N}Pro-Thr^{C} 199.1077	
^{N}Pro-Cys^{C} 201.0692	
^{N}Pro-Ile^{C} 211.1441	
^{N}Pro-Leu^{C} 211.1441	

^{N}Pro-Asn^{C} 212.1030	
^{N}Pro-Asp^{C} 213.0870	
^{N}Pro-Gln^{C} 226.1186	
^{N}Pro-Lys^{C} 226.1550	
^{N}Pro-Glu^{C} 227.1026	
^{N}Pro-Met^{C} 229.1005	
^{N}Pro-His^{C} 235.1190	
^{N}Pro-Phe^{C} 245.1285	

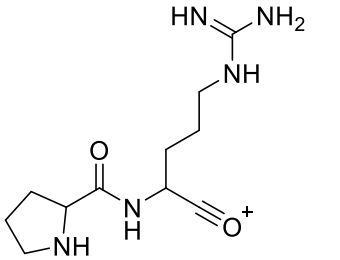
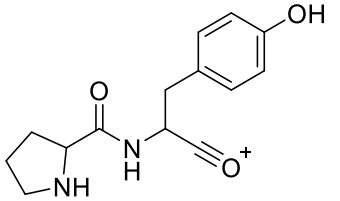
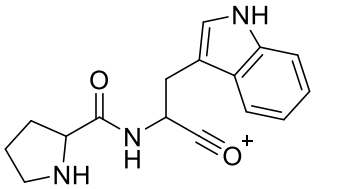
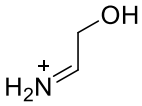
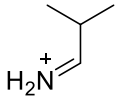
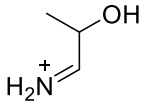
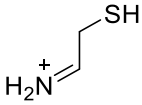
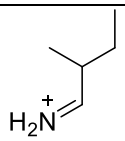
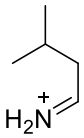
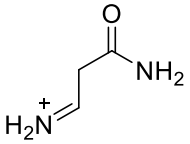
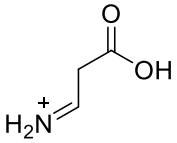
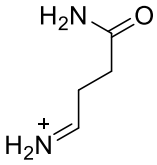
^{N}Pro-Arg^{C} 254.1612	
^{N}Pro-Tyr^{C} 261.1234	
^{N}Pro-Trp^{C} 284.1394	

Table S2: $[M+H]^{1+}$ masses and structures of the amino acyl immonium MS^2 ions

$[M+H]^{1+}$	**Structure**
30.0338	
44.0495	
60.0444	
70.0651	
72.0808	
74.0600	
76.0215	
86.0964	
86.0964	
87.0553	
88.0393	
101.0709	

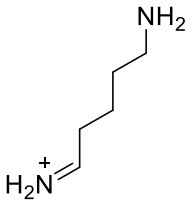
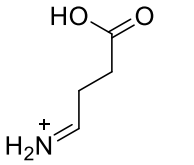
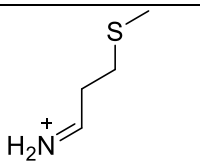
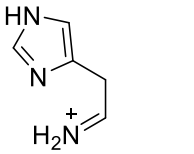
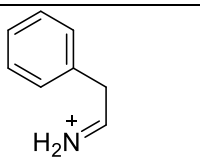
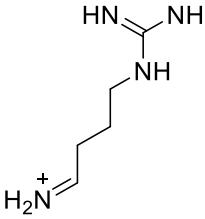
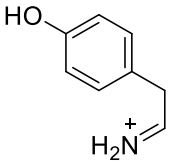
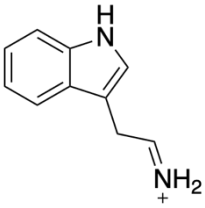
101.1073	
102.0550	
104.0528	
110.0713	
120.0808	
129.1135	
136.0757	
159.0917	

Table S3: Compound names, sequences, sources, and literature references for sponge derived PRMPs. PRMPs sequenced in this study are highlighted in yellow.

PRMP	Sequence	Sponge	Reference DOI
Phakellistatin 1	YPIPIFP	*Phakellia costata*	10.1021/np50092a011
Phakellistatin 2	FPIIPYP	*Phakellia carteri*	10.1016/S0960-894X(01)80781-2
Phakellistatin 3	FGPTLWP	*Phakellia carteri*	10.1021/jo00086a001
Phakellistatin 4	FIFSPTP	*Phakellia costata*	10.3987/COM-94-S45
Phakellistatin 5	FNAMAIP	*Phakellia costata*	10.1016/S0960-894X(01)80108-6
Phakellistatin 6	FPWLPIP	*Phakellia costata*	10.1016/S0960-894X(01)80695-8
Phakellistatin 7	FALPPYIPPI	*Phakellia costata*	10.1016/0960-894X(95)00219-J
Phakellistatin 8	FVLPPYIPPI	*Phakellia costata*	10.1016/0960-894X(95)00219-J
Phakellistatin 9	FVLPPYVPPI	*Phakellia costata*	10.1016/0960-894X(95)00219-J
Phakellistatin 10	WVPLTPIP	*Phakellia* sp.	10.1021/np50120a025
Phakellistatin 11	FPFIFPQP	*Phakellia* sp.	10.1021/np50120a025
Phakellistatin 12	FTLPPYIPPI	*Phakellia* sp.	10.1016/S0960-894X(02)01054-5
Phakellistatin 13	FGPTLWP	*Phakellia fusca*	10.1021/np020223y
Phakellistatin 14	FDAMAIP	*Phakellia* sp.	10.1021/np040092w
Phakellistatin 15	WIPLTPLP	*Phakellia fusca*	10.1021/np9008267
Phakellistatin 16	FDSRAVTYP	*Phakellia fusca*	10.1021/np9008267
Phakellistatin 17	WVPLIPIP	*Phakellia fusca*	10.1021/np9008267
Phakellistatin 18	YPIFPIP	*Phakellia fusca*	10.1021/np9008267
Phakellistatin 20[1]	FNAMAIP	*Stylissa flabelliformis*	10.1021/acs.jnatprod.8b00121
Phakellistatin 21[2]	FELPPYIPPM	*Stylissa flabelliformis*	10.1021/acs.jnatprod.8b00121
Phakellistatin 22	FELPPYIPPM	*Stylissa flabelliformis*	10.1021/acs.jnatprod.8b00121
Fuscasin A	FVLPPPP	*Phakellia fusca*	10.1021/acs.jnatprod.8b01033
Fuscasin B	YALPVNR	*Phakellia fusca*	10.1021/acs.jnatprod.8b01033
Fuscasin C	YPYAPAL	*Phakellia fusca*	10.1021/acs.jnatprod.8b01033
Fuscasin D	YLLPLPD	*Phakellia fusca*	10.1021/acs.jnatprod.8b01033
Hymenamide A	FWRPPVP	*Hymeniacidon* sp.	10.1016/S0040-4020(01)86318-3
Hymenamide B	PNFVEFP	*Hymeniacidon* sp.	10.1016/S0040-4020(01)86318-3
Hymenamide C	FGPELWP	*Hymeniacidon* sp.	10.1016/S0040-4020(01)80422-1
Hymenamide D	YDPLAIP	*Hymeniacidon* sp.	10.1016/S0040-4020(01)80422-1
Hymenamide E	YFFPTTP	*Hymeniacidon* sp.	10.1016/S0040-4020(01)80422-1
Hymenamide F	PAVMLRP	*Hymeniacidon* sp.	10.1016/0040-4020(96)00281-5
Hymenamide G	YVPLILPP	*Hymeniacidon* sp.	10.1016/S0040-4020(01)85006-7
Hymenamide H	WVPLTPLP	*Hymeniacidon* sp.	10.1016/S0040-4020(01)85006-7
Hymenamide J	YDFWKVYP	*Hymeniacidon* sp.	10.1016/S0040-4020(01)85006-7
Hymenamide K	YDFWKAVP	*Hymeniacidon* sp.	10.1016/S0040-4020(01)85006-7
Stylissamide A	YKPPVYP	*Stylissa caribica*	10.1002/ejoc.200700013
Stylissamide B	FPPPIYP	*Stylissa caribica*	10.1002/ejoc.200700013
Stylissamide C	FPFIPYP	*Stylissa caribica*	10.1002/ejoc.200700013
Stylissamide D	FIFYPLP	*Stylissa caribica*	10.1002/ejoc.200700013
Stylissamide E	YPAIQIP	*Stylissa caribica*	10.1021/np900664f
Stylissamide F	FDPRFPP	*Stylissa caribica*	10.1021/np900664f
Stylissamide G	FPLIPFP	*Stylissa caribica*	10.1021/np400891s

[1] The methionine residue in Phakellistatins 20, 21, 22 is oxidized.
[2] Phakellistatins 21 and 22 are epimers at the sulfur atom.

Stylissamide H	WPISFVP	*Stylissa caribica*	10.1021/np400891s
Stylissamide L	FPPSYPQ	*Stylissa caribica*	10.3390/md18090443
Stylissamide X	WFPLTPIP	*Stylissa* sp.	10.1016/j.bmcl.2011.10.023
Stylisin 1	YPIFPLP	*Stylissa caribica*	10.1021/np060006n
Stylisin 2	FPIPYPP	*Stylissa caribica*	10.1021/np060006n
Stylissatin A	FIYFPIP	*Stylissa massa*	10.1016/j.tetlet.2013.10.003
Stylissatin B	FPPGHPL	*Stylissa massa*	10.1016/j.tetlet.2016.08.024
Stylissatin C	FVLPPPD	*Stylissa massa*	10.1016/j.tetlet.2016.08.024
Stylissatin D	FVLPPPD	*Stylissa massa*	10.1016/j.tetlet.2016.08.024
Stylissatin E	FVPELWP	*Stylissa* sp.	This work
Stylissatin F	FPWVPLTP	*Stylissa* sp.	This work
Stylissatin G	WLPLTPLP	*Stylissa* sp.	This work
Stylostatin 1	FNSLAIP	*Stylissa massa*	10.1021/jo00052a041
Axinellin C	FPLTVPWP	*Stylissa massa*	10.1016/S0040-4020(02)00898-0
Wainunuamide	FPHPPGL	*Stylissa massa*	10.1016/S0040-4039(01)01993-1
Carteritin A	FIPPEYP	*Stylissa carteri*	10.1016/j.tetlet.2016.02.031
Carteritin B	YYPSLYP	*Stylissa carteri*	10.1016/j.tetlet.2016.02.031
Euryjanicin A	WPISFVP	*Prosuberites laughlini*	10.1016/j.tetlet.2009.05.067
Euryjanicin B	FVPPATP	*Prosuberites laughlini*	10.1021/np9004135
Euryjanicin C	FPISIPL	*Prosuberites laughlini*	10.1021/np9004135
Euryjanicin D	FSPIFPI	*Prosuberites laughlini*	10.1021/np9004135
Euryjanicin E	FPIFPIN	*Prosuberites laughlini*	10.1016/j.tet.2013.10.095
Euryjanicin F	FFPIFPI	*Prosuberites laughlini*	10.1016/j.tet.2013.10.095
Euryjanicin G	FPFPIFP	*Prosuberites laughlini*	10.1016/j.tet.2013.10.095
Axinastatin 1	FVVPVNP	*Axinella* sp.	10.1021/jm00115a027
Axinastatin 2	FVLPVNP	*Axinella* sp.	10.1021/jm00034a014
Axinastatin 3	FILPVNP	*Axinella* sp.	10.1021/jm00034a014
Axinastatin 4	LWVPLTP	*Axinella carteri*	10.1021/np970139w
Axinelllin A	FTLFPNP	*Axinella carteri*	10.1002/(SICI)1099-0690(199811)1998:11<2659::AID-EJOC2659>3.0.CO;2-H
Axinellin B	FPTLVPWP	*Axinella carteri*	10.1002/(SICI)1099-0690(199811)1998:11<2659::AID-EJOC2659>3.0.CO;2-H
Axinellin C	FPLTVPWP	*Stylotella aurantium*	10.1016/S0040-4020(02)00898-0
Axinellin D	FFPELWP	*Axinella* sp.	This work
Axinellin E	FYSLAIP	*Axinella* sp.	This work
Axinellin F	FFPELLP	*Axinella* sp.	This work
Axinellin G	IPHPYIIGP	*Axinella* sp.	This work
Reniochalistatin A	LLPNVIP	*Reniochalina stalagmitis*	10.1021/np5006778
Reniochalistatin B	FYLPLPI	*Reniochalina stalagmitis*	10.1021/np5006778
Reniochalistatin C	FPIYFPI	*Reniochalina stalagmitis*	10.1021/np5006778
Reniochalistatin D	FPFIFPP	*Reniochalina stalagmitis*	10.1021/np5006778
Reniochalistatin E	WPVIPIPL	*Reniochalina stalagmitis*	10.1021/np5006778
Stylopeptide 1	FSPIPLI	*Stylotella sp.*	10.1021/jo00130a027
Stylopeptide 2	FPYPTPFLLP	*Stylotella sp.*	10.1021/np0704856
Malaysiatin	FVVVNPP	*Pseudaxinyssa sp.*	10.1016/S0040-4039(00)61115-2
Dominicin	PTPIIILP	*Eurypon laughlini*	10.1021/np049711r

SUPPLEMENTARY FIGURES

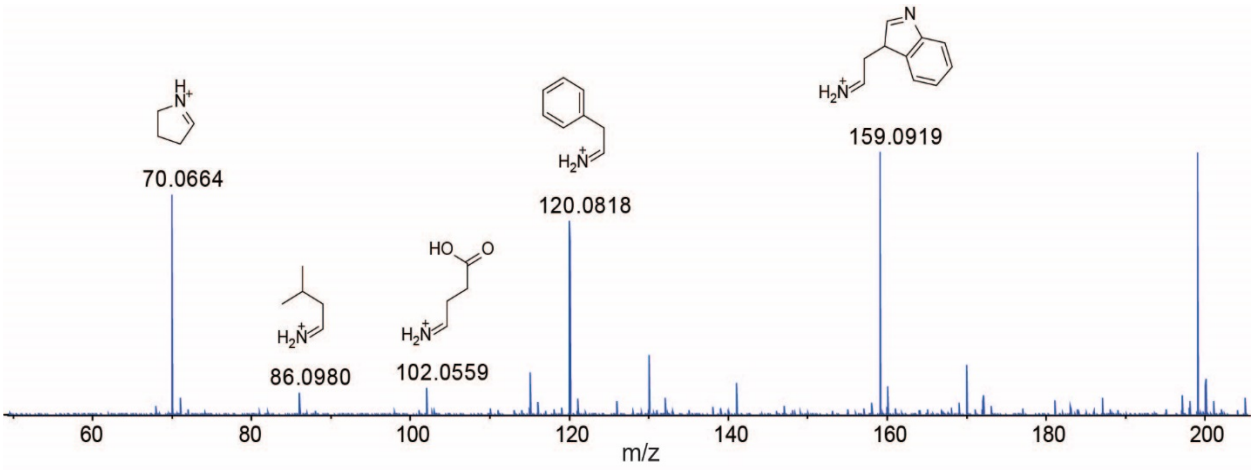


Fig. S1: Annotation of MS^2 amino acyl immonium ions detected for PRMP cyclo(FFPE**X**WP). Residue **X** was determined to be Leu.

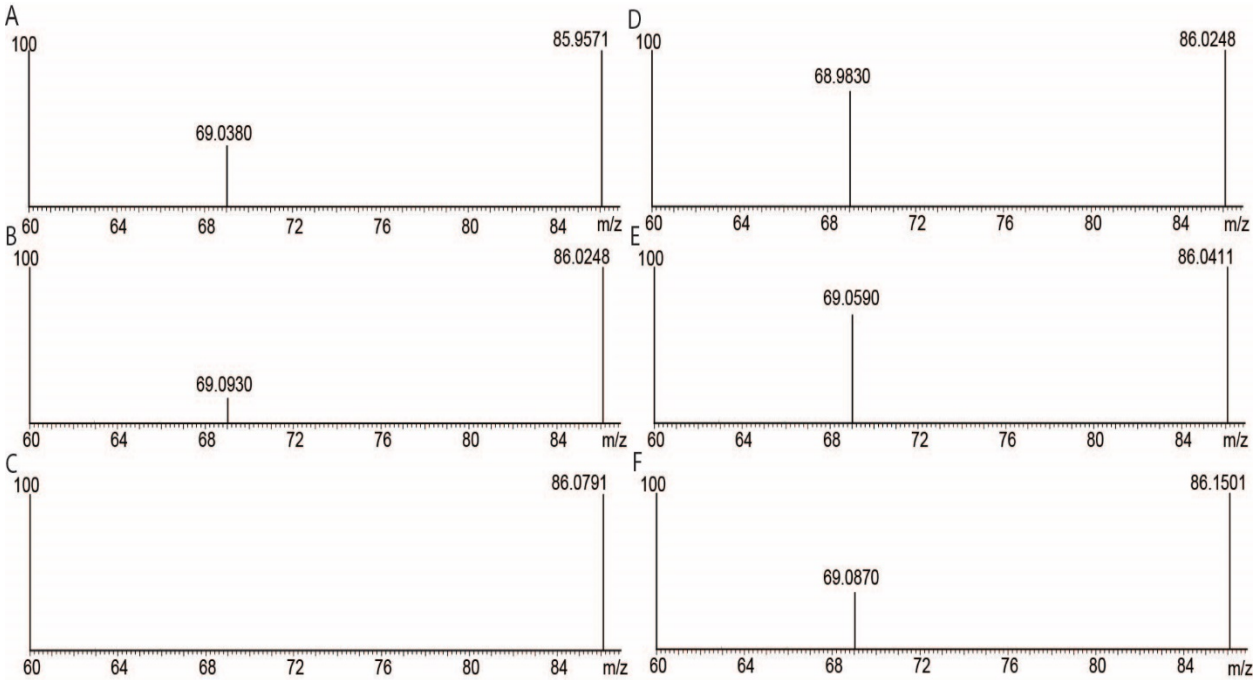


Fig. S2: Abundance of MS^3 *m/z* 69 Da product ion detected for synthetic standards with one Leu residue at HCD 15 (**A**), HCD 11 (**B**), HCD 8 (**C**) and one Ile residue at HCD 15 (**D**), HCD 11 (**E**), HCD 8 (**F**). Note that at HCD 8, panels **C** and **F**, the MS^3 *m/z* 69 Da product ion is not detected for Leu, but only detected for Ile. For all other HCD energies, some abundance of the MS^3 *m/z* 69 Da is detected for the Leu residue.

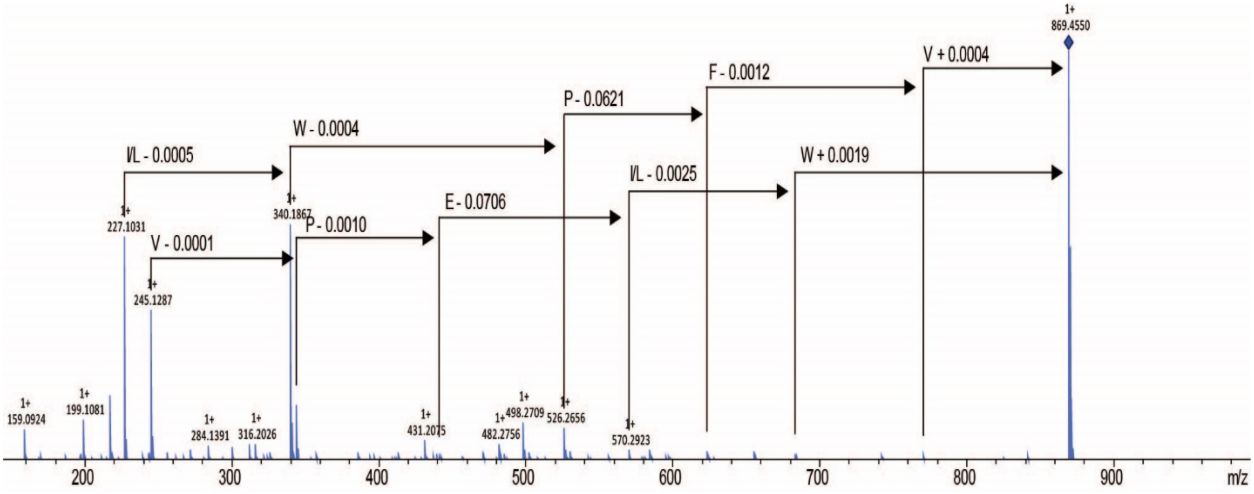


Fig. S3: MS^2 sequencing of PRMP stylissatin E cyclo(FVPE**L**WP).

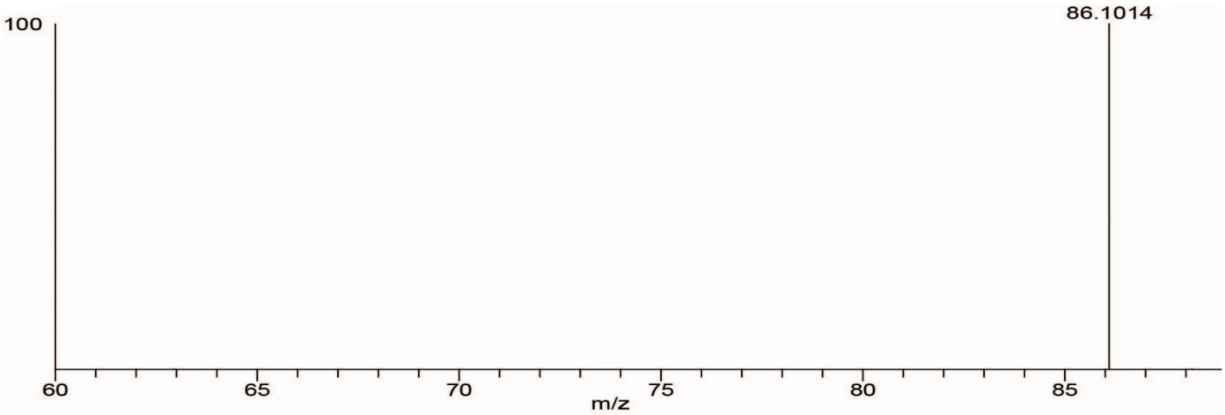


Fig. S4: $MS^2(86) \rightarrow MS^3$ spectra for PRMP stylissatin E cyclo(FVPE**L**WP).

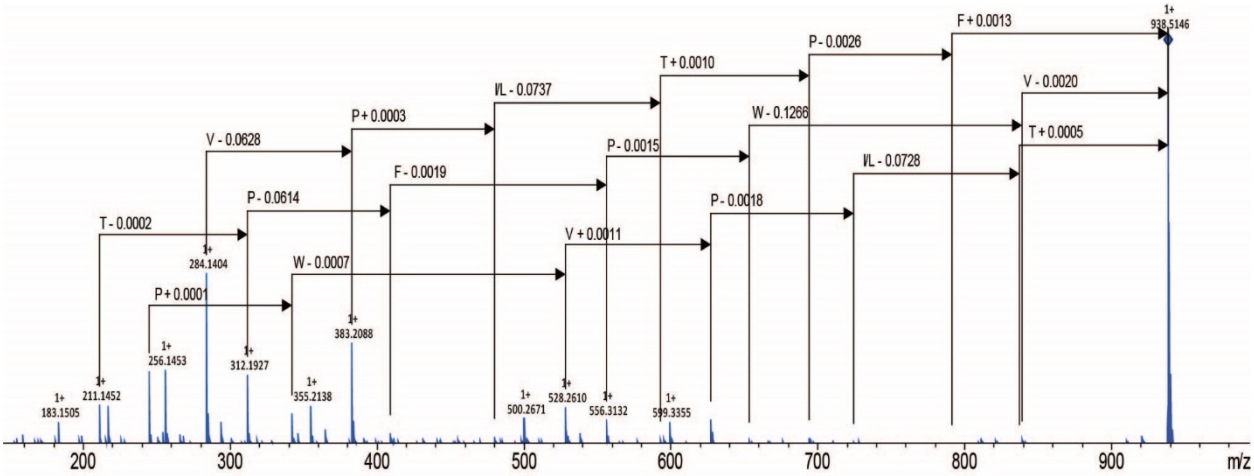


Fig. S5: MS^2 sequencing of PRMP stylissatin F cyclo(FPWVP**L**TP).

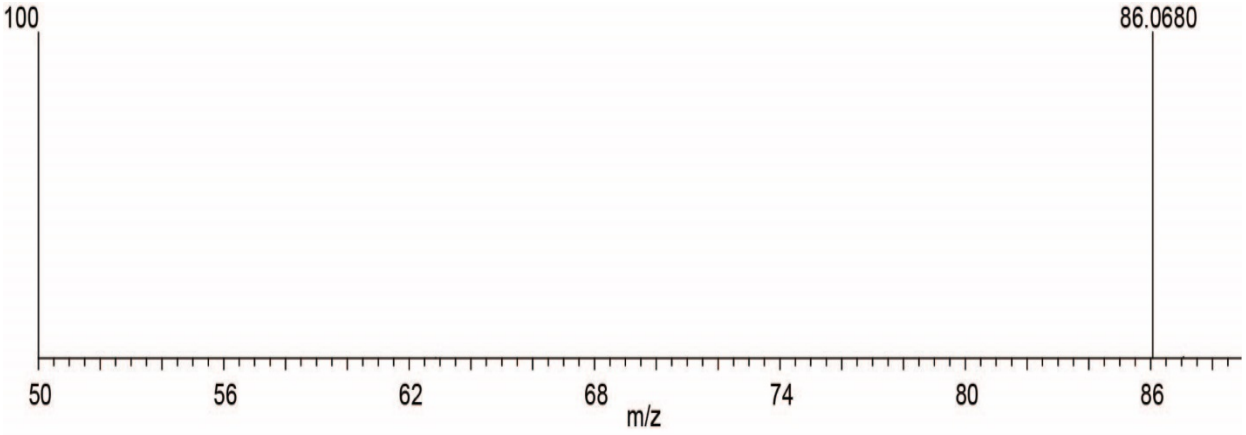


Fig. S6: $MS^2(86) \rightarrow MS^3$ spectra for PRMP stylissatin F cyclo(FPWVP**L**TP).

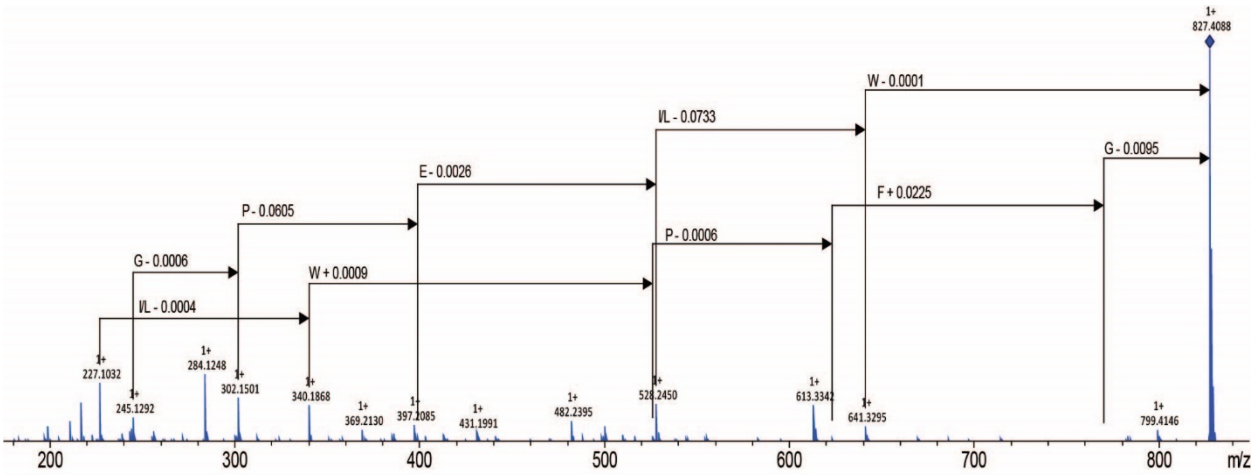


Fig. S7: MS^2 sequencing for PRMP hymenamide C cyclo(FGPE**L**WP).

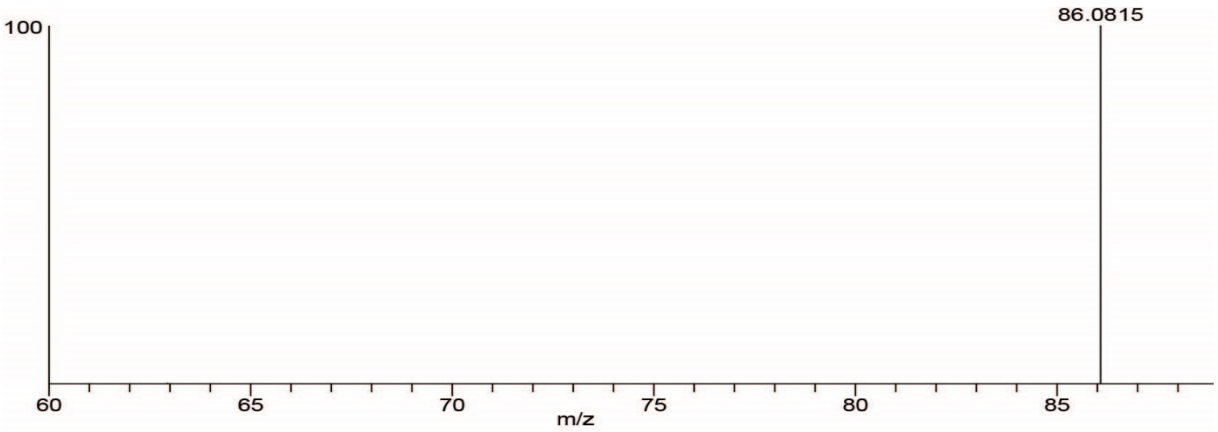


Fig. S8: $MS^2(86) \rightarrow MS^3$ spectra for hymenamide C cyclo(FGPE**L**WP).

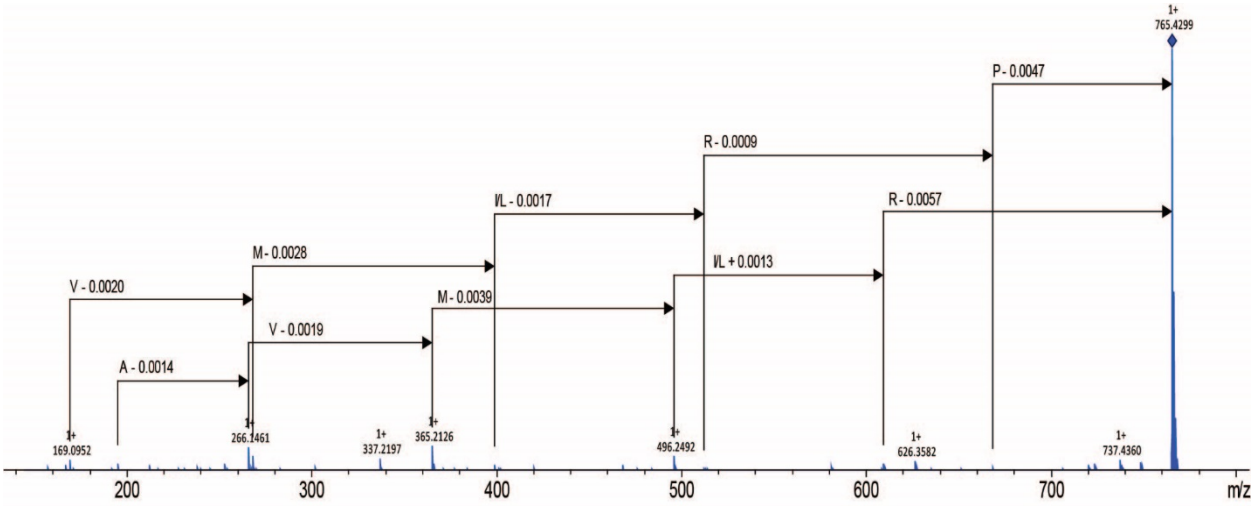


Fig. S9: MS^2 sequencing for PRMP hymenamide F cyclo(PAVM**L**RP).

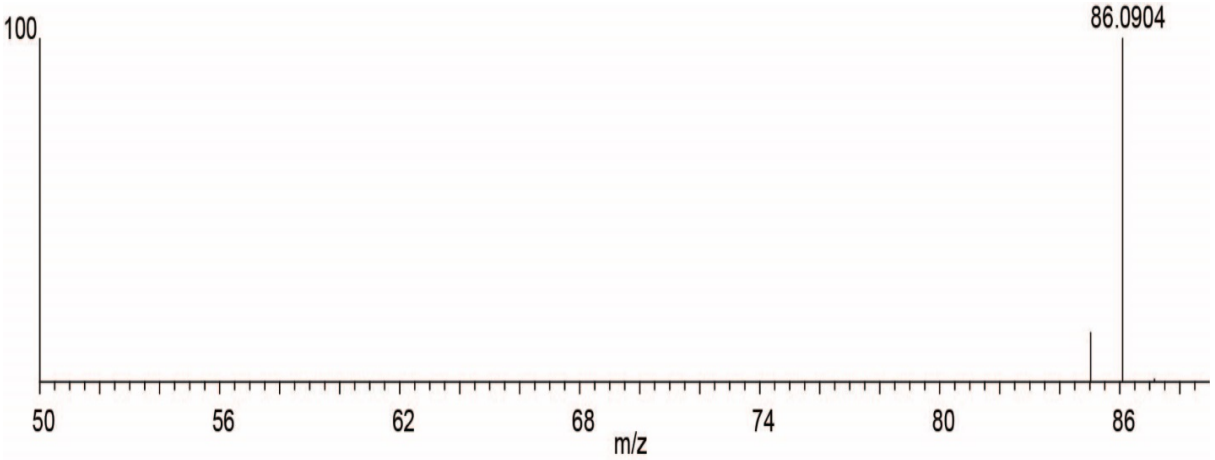


Fig. S10: $MS^2(86) \rightarrow MS^3$ spectra for PRMP hymenamide F cyclo(PAVM**L**RP).

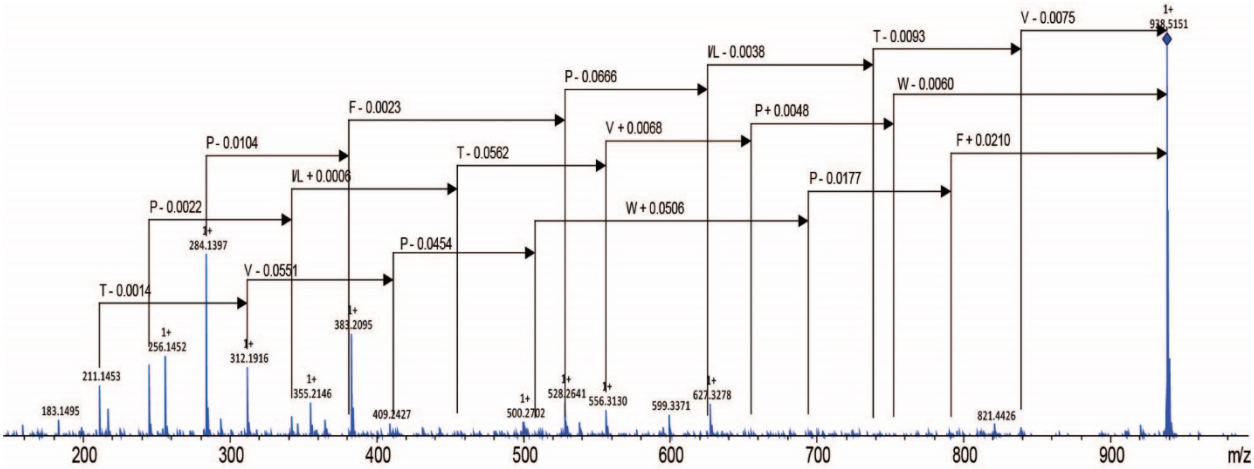


Fig. S11: MS^2 sequencing for PRMP axinellin B cyclo(FP**L**TVPWP).

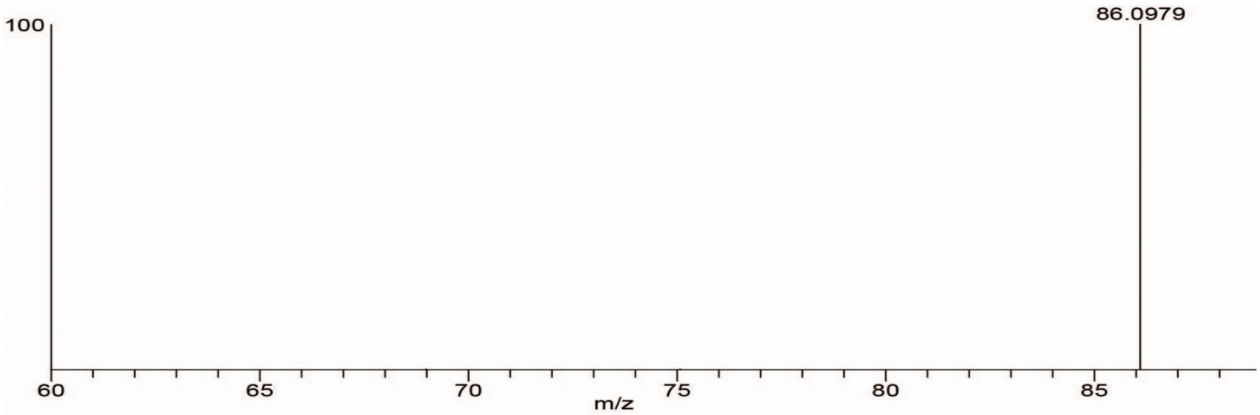


Fig. S12: $MS^2(86) \rightarrow MS^3$ spectra for PRMP axinellin B cyclo(FP**L**TVPWP).

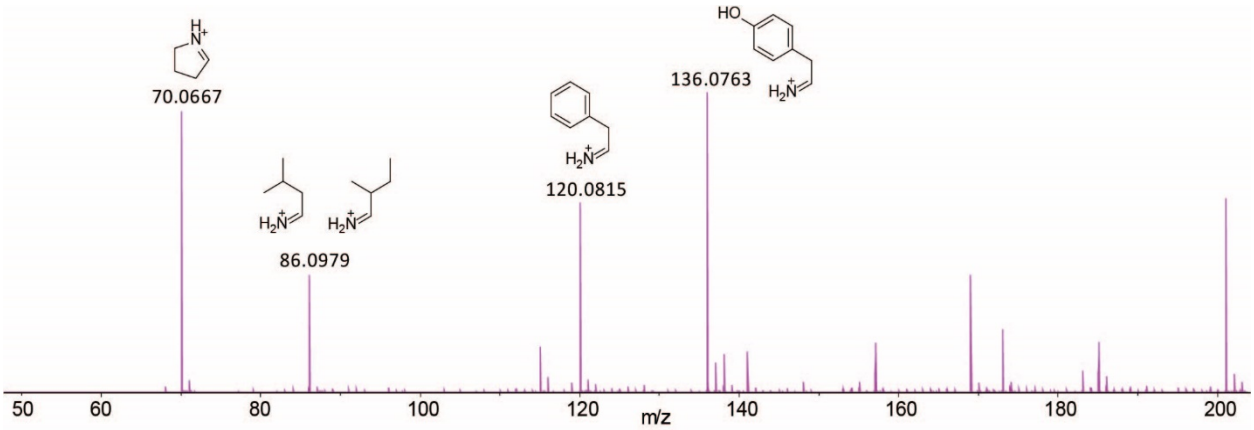


Fig. S13: Annotation of amino acyl immonium ions detected in the MS^2 spectra for PRMP axinellin E cyclo(FYS**L**A**I**P). Immonium product ions for Ala and Ser residues were not detected.

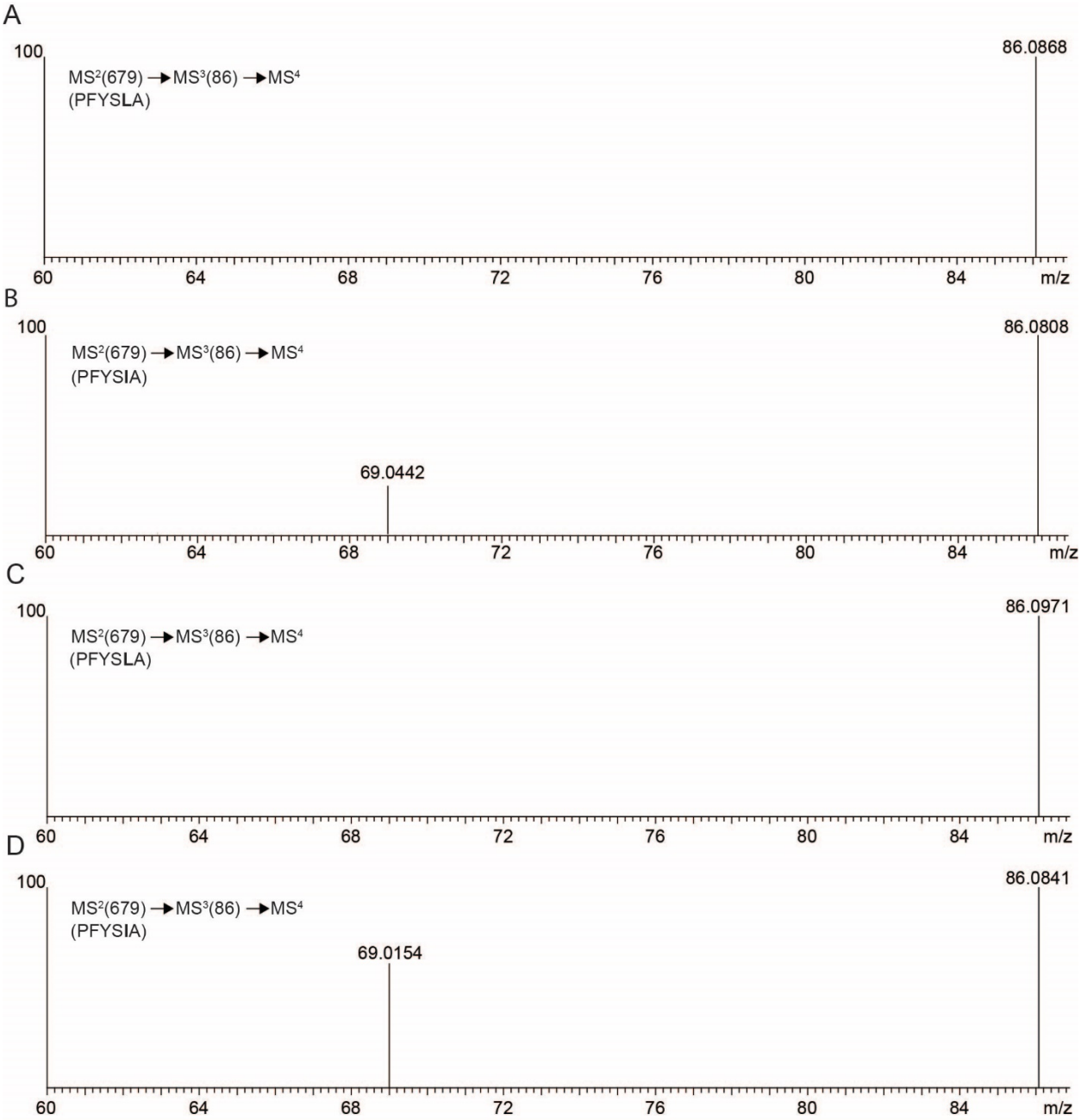


Fig. S14: $MS^2(679) \rightarrow MS^3(86) \rightarrow MS^4$ spectra for standards, (**A**) cyclo(FYS**L**A**L**P), (**B**) cyclo(FYS**I**A**L**P), (**C**) cyclo(FYS**L**A**I**P), and (**D**) cyclo(FYS**I**A**I**P).

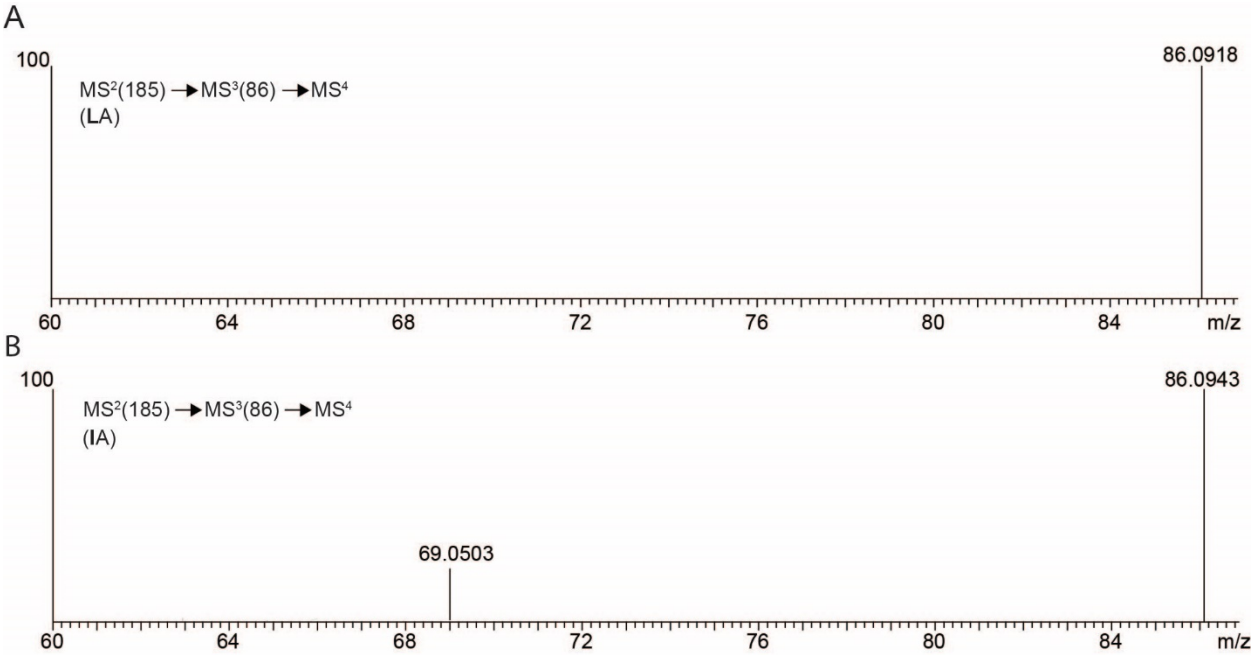


Fig. S15: $MS^2(185) \rightarrow MS^3(86) \rightarrow MS^4$ spectra for standards (**A**) cyclo(FYSLA**L**P), and (**B**) cyclo(FYSLA**I**P). Note that for both these standards, X^1=Leu.

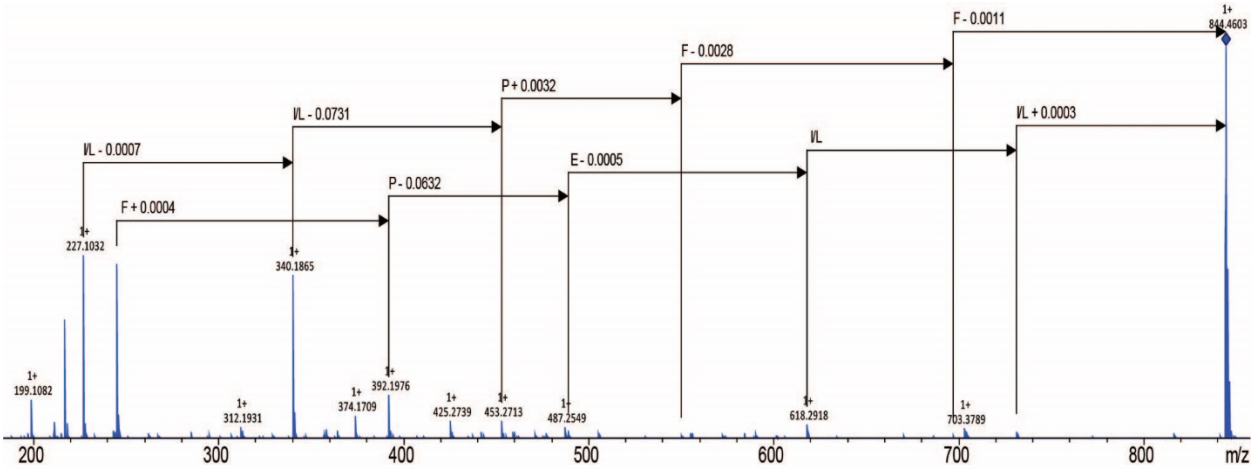


Fig. S16: MS^2 sequencing of PRMP axinellin F cyclo(FFPE**LL**P).

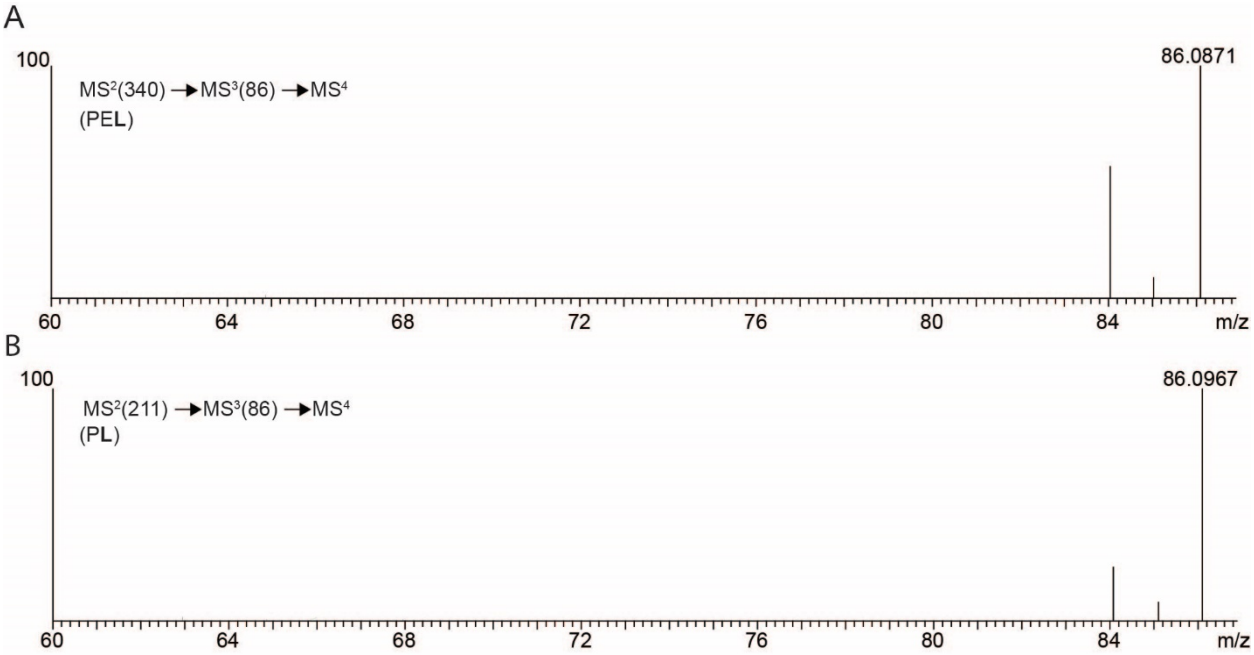


Fig. S17: (**A**) $MS^2(340) \rightarrow MS^3(86) \rightarrow MS^4$ and (**B**) $MS^2(211) \rightarrow MS^3(86) \rightarrow MS^4$ spectra for PRMP cyclo(FFPE**LL**P).

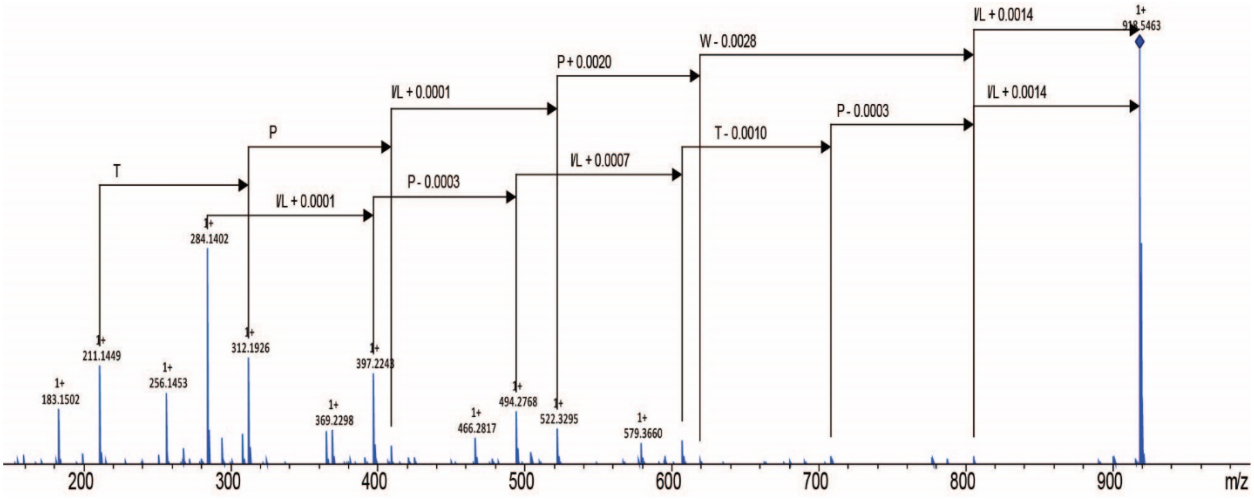


Fig. S18: MS^2 sequencing of PRMP stylissatin G cyclo(W**L**P**L**TP**L**P).

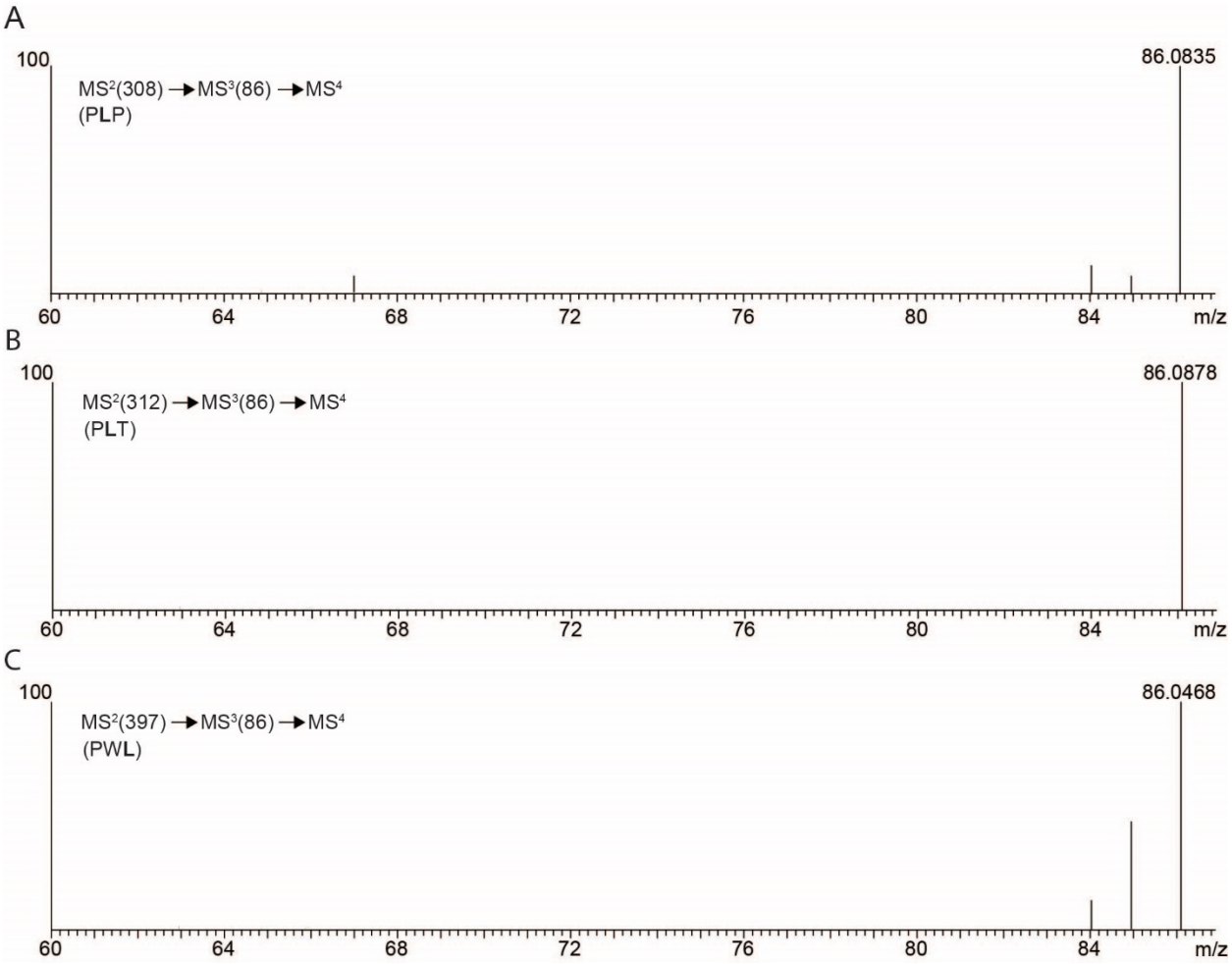


Fig. S19: (**A**) $MS^2(308) \rightarrow MS^3(86) \rightarrow MS^4$ (**B**) $MS^2(312) \rightarrow MS^3(86) \rightarrow MS^4$ (**C**) $MS^2(397) \rightarrow MS^3(86) \rightarrow MS^4$ spectra for PRMP cyclo(W**L**P**L**TP**L**P).

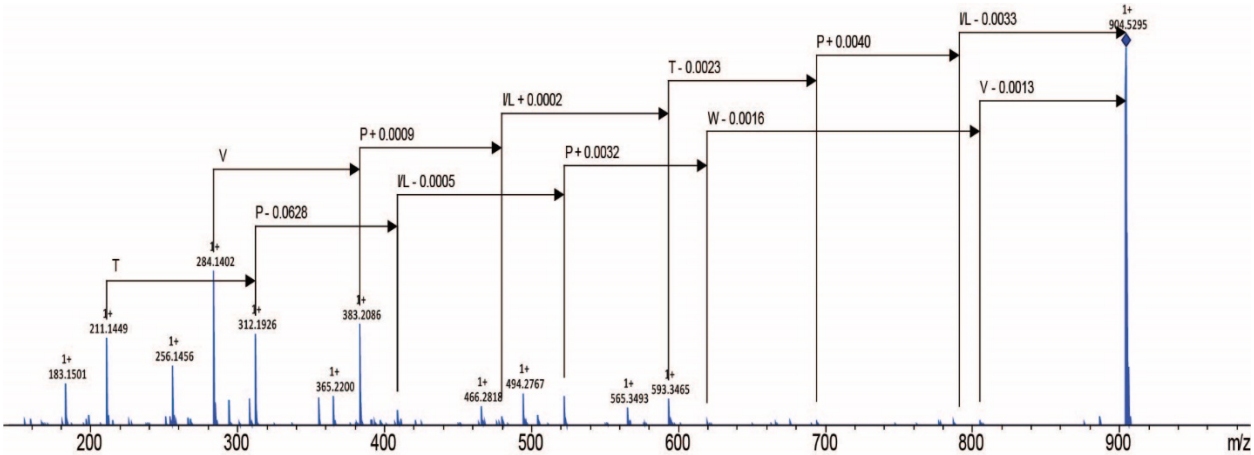


Fig. S20: MS^2 sequencing of PRMP hymenamide H cyclo(WVP**L**TP**L**P).

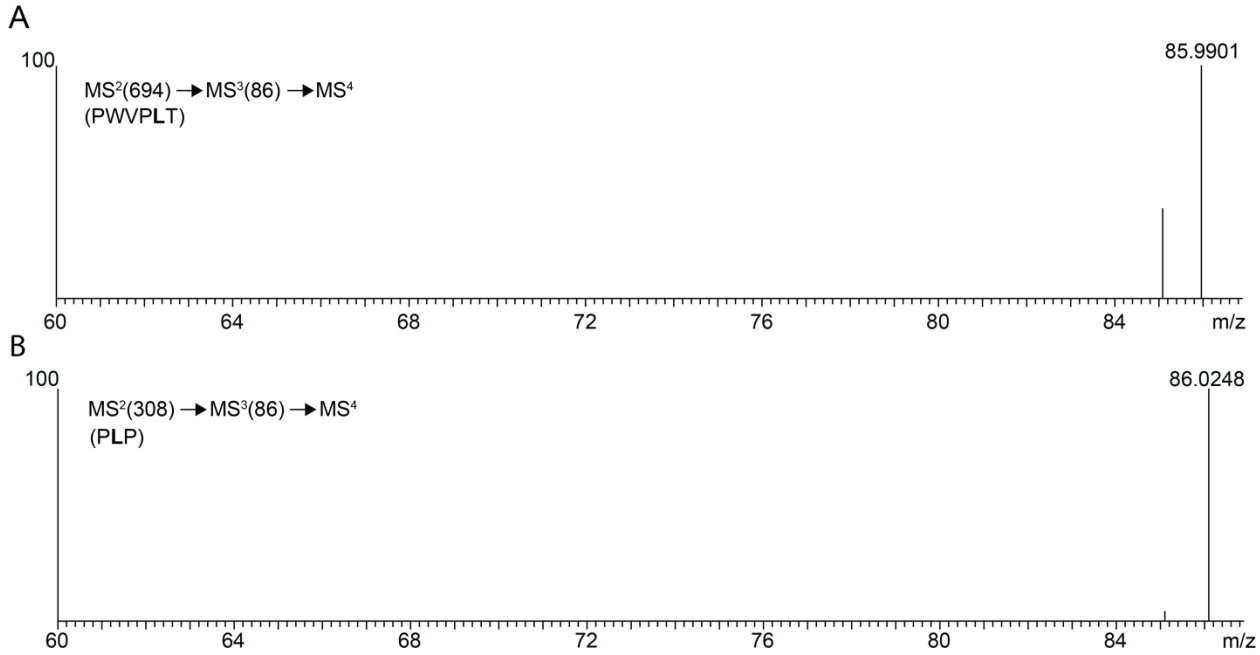


Fig. S21: (**A**) $MS^2(694) \rightarrow MS^3(86) \rightarrow MS^4$ (**B**) $MS^2(308) \rightarrow MS^3(86) \rightarrow MS^4$ analysis of PRMP cyclo(WVP**L**TP**L**P).

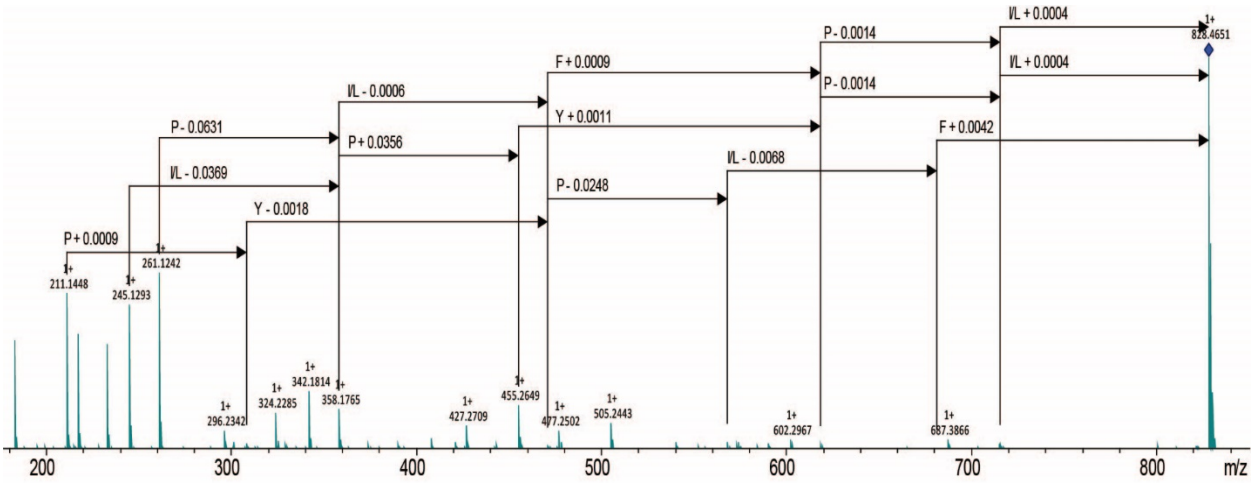


Fig. S22: MS^2 sequencing of PRMP phakellistatin 18 cyclo(YP**I**FP**I**P).

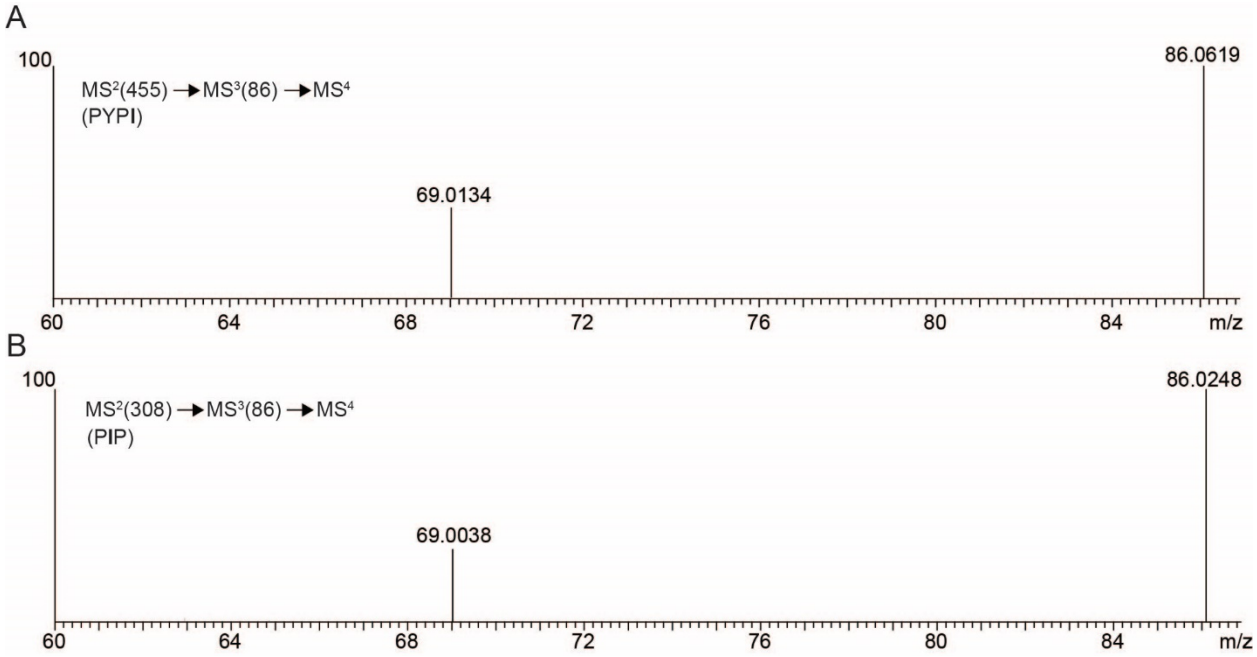


Fig. S23: (**A**) $MS^2(455) \rightarrow MS^3(86) \rightarrow MS^4$ (**B**) $MS^2(308) \rightarrow MS^3(86) \rightarrow MS^4$ spectra for cyclo(YP**I**FP**I**P).

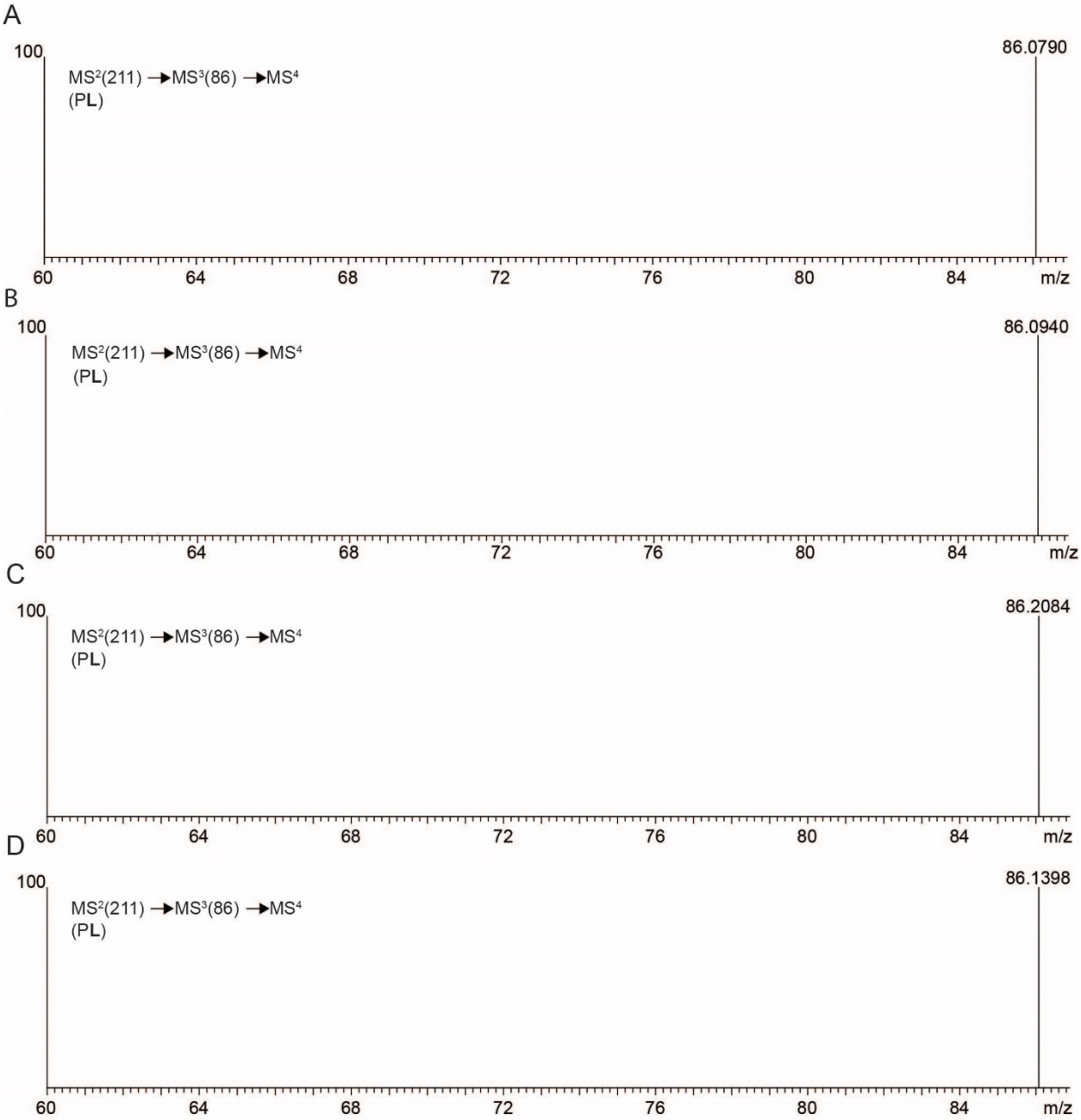


Fig. S24: $MS^2(211) \rightarrow MS^3(86) \rightarrow MS^4$ spectra for (**A**) cyclo(**L**PHPY**LL**GP), (**B**) cyclo(**L**PHPY**LI**GP), (**C**) cyclo(**L**PHPY**IL**GP), and (**D**) cyclo(**L**PHPY**II**GP).

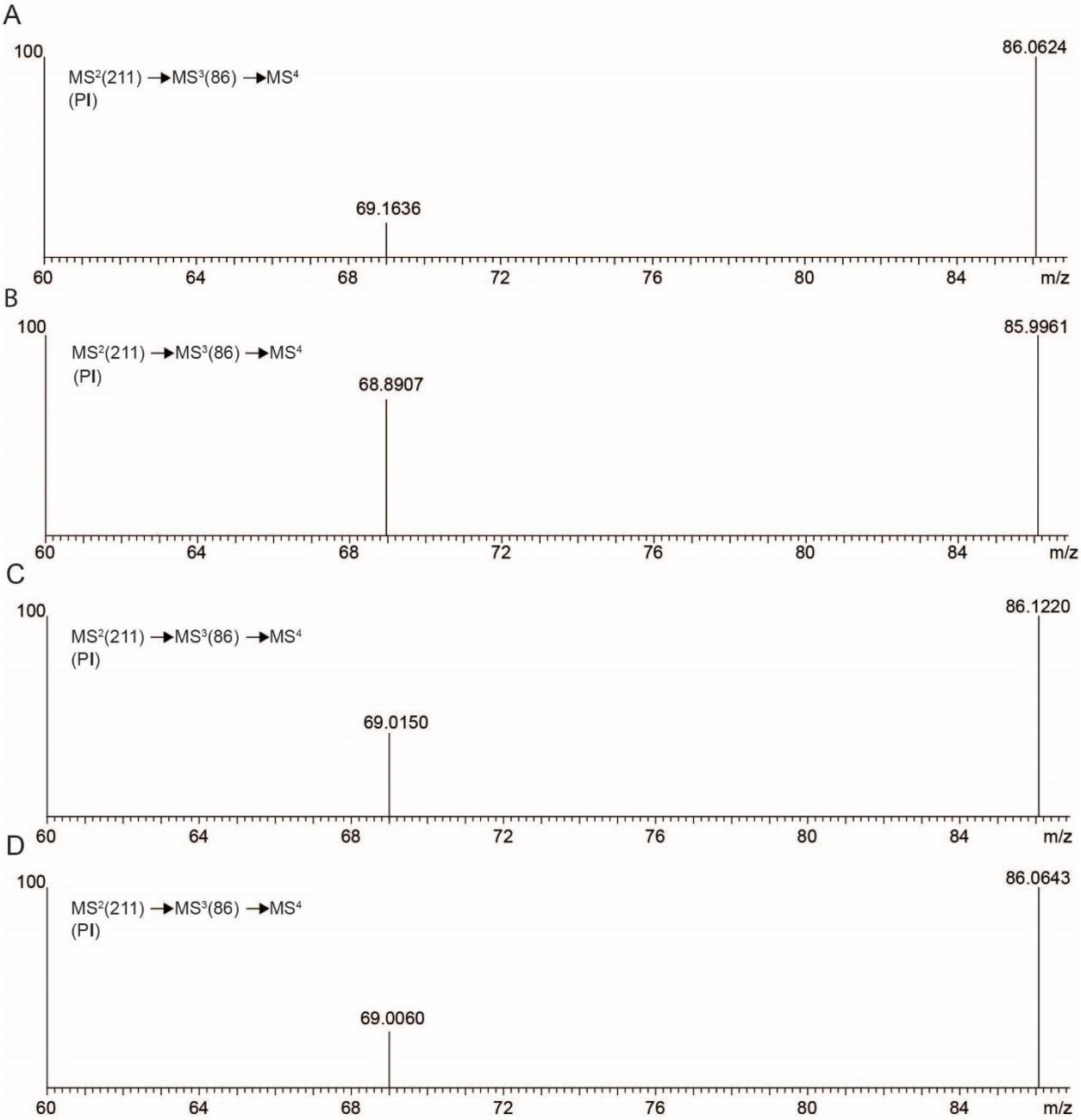


Fig. S25: $MS^2(211) \rightarrow MS^3(86) \rightarrow MS^4$ spectra for (**A**) cyclo(**I**PHPY**LL**GP), (**B**) cyclo(**I**PHPY**LI**GP), (**C**) cyclo(**I**PHPY**IL**GP), and (**D**) cyclo(**I**PHPY**II**GP).

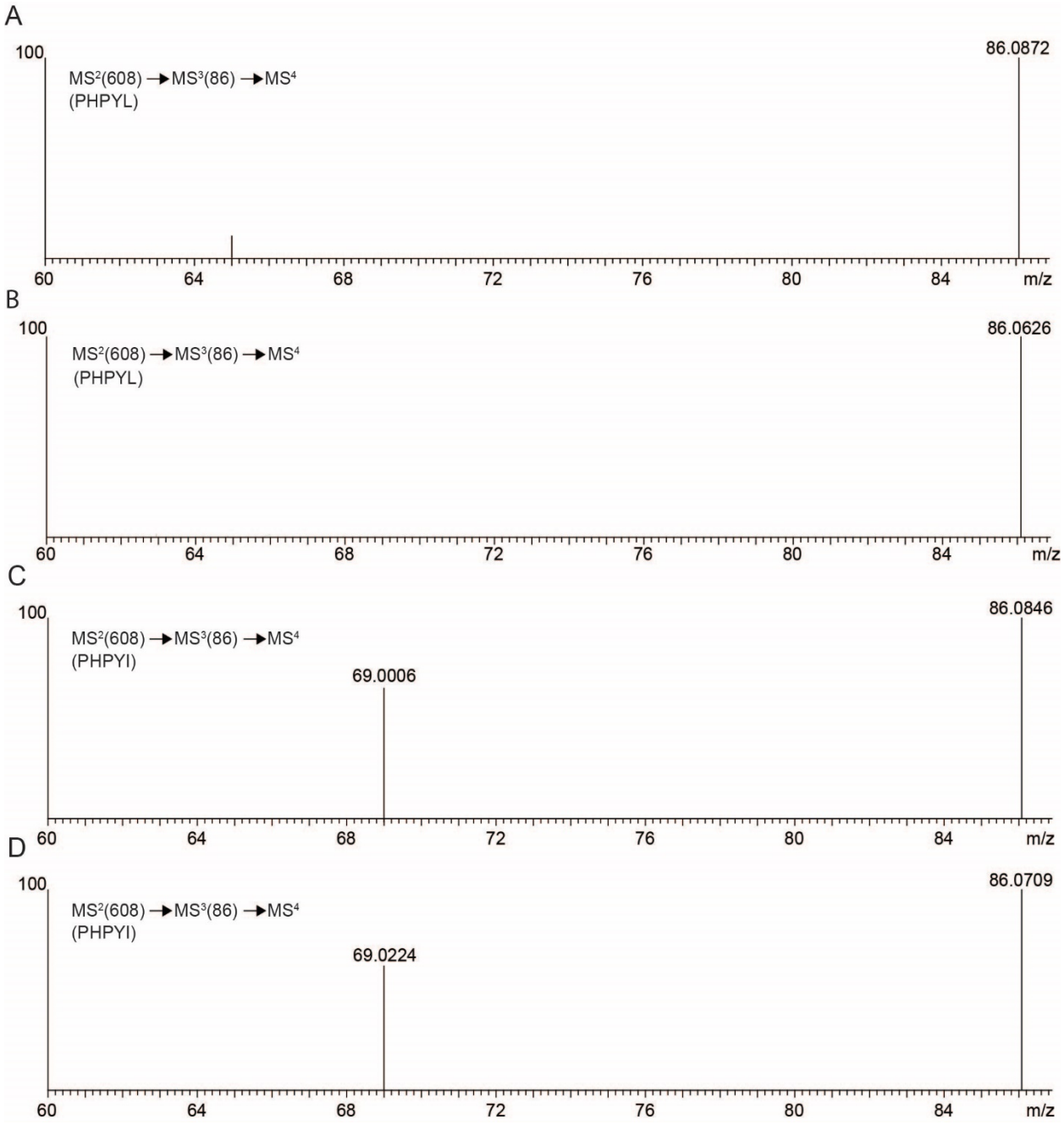


Fig. S26: (**A**) $MS^2(608) \rightarrow MS^3(86) \rightarrow MS^4$ spectra for (**A**) cyclo(**I**PHPY**LL**GP), (**B**) cyclo(**I**PHPY**LI**GP), (**C**) cyclo(**I**PHPY**IL**GP), and (**D**) cyclo(**I**PHPY**II**GP).

SUPPLEMENTARY REFERENCES

[1] H. Ludewig, C. M. Czekster, E. Oueis, E. S. Munday, M. Arshad, S. A. Synowsky, A. F. Bent, J. H. Naismith, *ACS Chemical Biology* **2018**, *13*, 801-811.

[2] aI. Mohanty, S. Podell, J. S. Biggs, N. Garg, E. E. Allen, V. Agarwal, *Mar. Drugs* **2020**, *18*, 124; bI. Mohanty, S. G. Moore, D. Yi, J. S. Biggs, D. A. Gaul, N. Garg, V. Agarwal, *ACS Chemical Biology* **2020**, *15*, 2185-2194.

[3] B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, S. P. Holmes, *Nature Methods* **2016**, *13*, 581-583.

[4] E. Bolyen, J. R. Rideout, M. R. Dillon, N. A. Bokulich, C. C. Abnet, G. A. Al-Ghalith, H. Alexander, E. J. Alm, M. Arumugam, F. Asnicar, Y. Bai, J. E. Bisanz, K. Bittinger, A. Brejnrod, C. J. Brislawn, C. T. Brown, B. J. Callahan, A. M. Caraballo-Rodríguez, J. Chase, E. K. Cope, R. Da Silva, C. Diener, P. C. Dorrestein, G. M. Douglas, D. M. Durall, C. Duvallet, C. F. Edwardson, M. Ernst, M. Estaki, J. Fouquier, J. M. Gauglitz, S. M. Gibbons, D. L. Gibson, A. Gonzalez, K. Gorlick, J. Guo, B. Hillmann, S. Holmes, H. Holste, C. Huttenhower, G. A. Huttley, S. Janssen, A. K. Jarmusch, L. Jiang, B. D. Kaehler, K. B. Kang, C. R. Keefe, P. Keim, S. T. Kelley, D. Knights, I. Koester, T. Kosciolek, J. Kreps, M. G. I. Langille, J. Lee, R. Ley, Y.-X. Liu, E. Loftfield, C. Lozupone, M. Maher, C. Marotz, B. D. Martin, D. McDonald, L. J. McIver, A. V. Melnik, J. L. Metcalf, S. C. Morgan, J. T. Morton, A. T. Naimey, J. A. Navas-Molina, L. F. Nothias, S. B. Orchanian, T. Pearson, S. L. Peoples, D. Petras, M. L. Preuss, E. Pruesse, L. B. Rasmussen, A. Rivers, M. S. Robeson, P. Rosenthal, N. Segata, M. Shaffer, A. Shiffer, R. Sinha, S. J. Song, J. R. Spear, A. D. Swafford, L. R. Thompson, P. J. Torres, P. Trinh, A. Tripathi, P. J. Turnbaugh, S. Ul-Hasan, J. J. J. van der Hooft, F. Vargas, Y. Vázquez-Baeza, E. Vogtmann, M. von Hippel, W. Walters, et al., *Nature Biotechnology* **2019**, *37*, 852-857.