ChemComm

Oxidation-guided and collision-induced linearization assists *de novo* sequencing of thioether macrocyclic peptides

Journal:	*ChemComm*
Manuscript ID	CC-COM-06-2024-003179.R1
Article Type:	Communication

Oxidation-guided and collision-induced linearization assists *de novo* sequencing of thioether macrocyclic peptides

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Ayaka Hayashi,[a] Yuki Goto,*[b,c,d] Yutaro Saito,[e] Hiroaki Suga,[b] Jumpei Morimoto,*[e] and Shinsuke Sando*[a,e]

Oxidation of a thioether linkage in thioether-closed macrocyclic peptides led to collision-induced site-selective linearization of the peptides. This method has allowed for *de novo* sequencing of thioether macrocyclic peptides. The utility of the sequencing method was demonstrated by identifying the correct peptide sequences from a virtually randomized thioether macrocyclic peptide library.

Macrocyclic peptides (MPs) are attracting a significant attention as a therapeutic modality[1]. Combinatorial MP libraries synthesized via the split and pool method[2] are a powerful means for structure optimization of lead MPs. In this approach, the libraries are constructed through chemical synthesis, allowing the incorporation of nonstandard amino acid residues possessing favorable pharmacokinetic properties. The constructed library can be used for various assays, including affinity selection on bead[3–6] and in solution[7–9], or membrane permeability assays[10–13]. Through these assays, the combinatorial libraries have enhanced the discovery of MPs with improved binding affinity, proteolytic resistance[14], and membrane permeability. To conduct these assays, post-assay sequence deconvolution of the library is required. Tandem mass spectrometry (MS/MS) is the most widely used method for peptide sequence deconvolution because of its high sensitivity and methodological simplicity. While high-fidelity sequencing is possible for linear peptides, accurately deconvoluting cyclic peptide sequences poses significant challenges. On MS/MS of cyclic peptides, concomitant fragmentations at two positions are required to observe the fragment ions available for sequencing. Consequently, MS/MS spectra of cyclic peptides often consist of a complex mixture of different series of fragment ions, and accordingly, the intensity of each fragment ion tends to become low. These issues make it difficult to reliably characterize the sequences of cyclic peptides by MS/MS (Fig. 1a). To overcome this problem, several methods for linearizing peptides at specific positions have been introduced, facilitating MS/MS-mediated sequencing of various types of cyclic peptides[7,15–18].

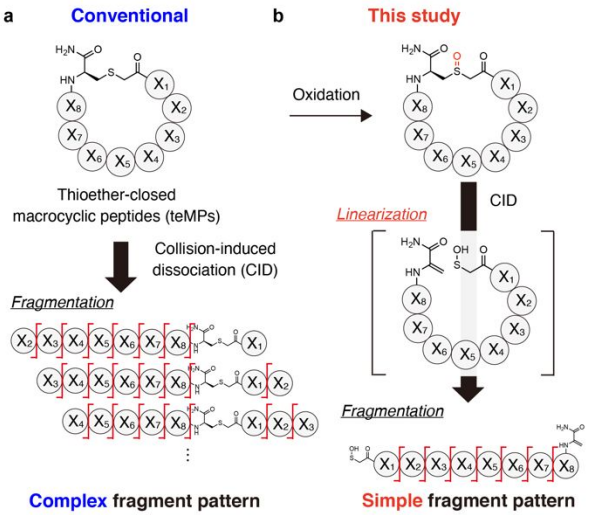

Fig. 1 (a) The structure of thioether-closed macrocyclic peptide and conventional LC-MS/MS analysis of macrocyclic peptides. (b) A schematic illustration of this study. The oxidized peptide is linearized by CID and, subsequently, fragmented.

The thioether-closed macrocyclic peptides (teMPs) (Fig. 1a, top) have been utilized as a reliable scaffold to devise *de novo* peptide ligands against target proteins of interest[19–22]. The teMPs can be expressed by an *in vitro* engineered translation system, in which an N-terminal chloroacetyl group incorporated via genetic code reprogramming posttranslationally undergoes spontaneous cyclization with a downstream Cys residue to give the ring-closing thioether linkage. The mRNA display-mediated *in vitro* selection campaigns have enabled the rapid discovery of teMP hits binding to given target proteins. Construction and screening of chemically synthesized focused libraries of teMP hits should be useful for their structural optimization to develop practical peptide agents. However, the lack of a reliable site-selective linearizing method for teMPs has hampered such an approach.

Here, we demonstrate that the oxidation of the thioether bond in teMPs to sulfoxide prior to the MS/MS analysis induces preferential fragmentation at the sulfoxide moiety through collision-induced dissociation (CID), facilitating sequence deconvolution of teMPs (Fig. 1b). In this method, CID-induced selective fragmentation of the sulfoxide results in producing a linearized product ion, which further undergoes fragmentations

[a] *Department of Bioengineering, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan.*
[b] *Department of Chemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033 Japan.*
[c] *Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa-Oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan.*
[d] *Toyota Riken Rising Fellow, Toyota Physical and Chemical Research Institute, Kitashirakawa-Oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan.*
[e] *Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan.*
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

to give a sole series of fragment ions. We have demonstrated the utility of the oxidation- and collision-induced linearization method for identifying the sequences of teMPs in sequence-randomized library[18,21]

We envisioned that teMPs could be cleaved by CID upon oxidation of the thioether linkage to sulfoxide. This hypothesis is based on the previous reports about sulfoxide-containing cross-linkers[23] and reporters[24]. The previous reports showed that CID induced the cleavage of the C–S bond of sulfoxide to generate an alkene and sulfenic acid, and the sulfoxide bond was more easily cleaved than amides in the peptide chain by CIDs.

To examine this idea, we first sought the optimal oxidation conditions using a model teMP **1**, which consists of a peptide sequence of WLTFPYNC closed by a thioether linkage (Fig. 2a). (The macrocyclic peptide will be described as te[WLTFPYNC] hereafter, and other thioether-closed macrocyclic peptides in the same format.) Initially, we tested two oxidizing agents: hydrogen peroxide and sodium periodate (Fig. S1). While hydrogen peroxide treatment produced only singly oxidized +16 product (Fig. S1b), sodium periodate treatment additionally produced multiply oxidized +48 product, presumably attributed to the oxidation of the Trp side chain (Fig. S1c). Therefore, we chose hydrogen peroxide as the oxidizing agent. Following the optimization of the conditions, we found that incubating the peptide with 1 M hydrogen peroxide in water for 7 h at 37 °C resulted in its quantitative conversion to the oxidized form (**1-oxidized**) (Fig. 2b and Fig. S2). Additionally, two more teMPs, te[AVKMDYQC] (**S1**) and te[HIERGYSC] (**S2**), were synthesized and oxidized under the same conditions to evaluate if the conditions consistently oxidize the ring-closing thioether linkage in various sequences without causing undesired modifications (Fig. S3 and S4). These three peptides (**1, S1**, and **S2**) combined encompass all proteinogenic amino acid residues.

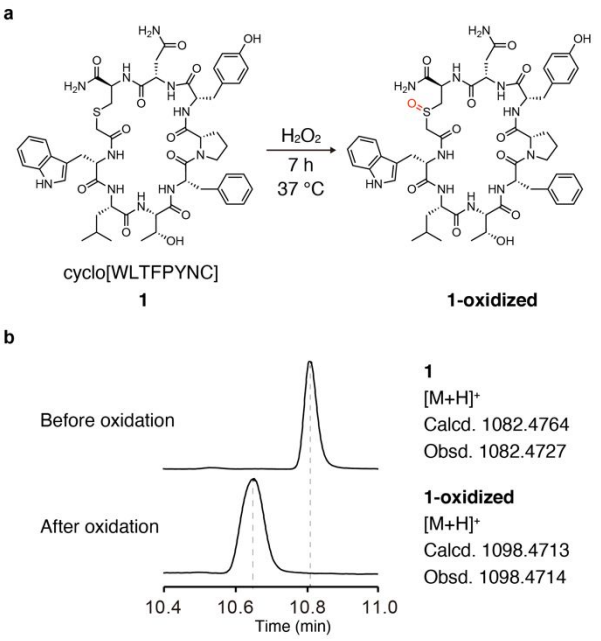

Fig. 2 (a) Oxidation of thioether linkage by hydrogen peroxide. (b) The chromatograms of the macrocyclic peptide before and after the oxidation. The elution was monitored by UV absorption at 220 nm.

Both **S1** and **S2** were converted to the oxidized forms with high conversion rates according to the ultraviolet (UV) chromatograms and mass spectra (96 ± 5% for **S1** and 84 ± 6% for **S2**). Peptide **S1** showed a +32 Da increase, attributed to the oxidation of the Met side chain and the ring-closing thioether linkage. We note that, for peptides **S1** and **S2**, multiple peaks with identical mass values appeared in the UV chromatograms. This is likely because the peptide diastereomers originated from the chiral nature of the sulfoxide structure were separated for these peptides during liquid chromatography (LC). To further confirm the selective oxidation of the ring-closing thioether bond under the condition, we also synthesized and oxidized a cyclic peptide **S3** (te[ALTFPYNC]) and the corresponding linear peptide lacking the ring-closing thioether linkage (**S3-linear**, Ac-ALTFPYN-NH_2). The peptide **S3** underwent oxidation to yield a product with a +16 mass shift (Fig. S5), while **S3-linear** remained unmodified (Fig. S6). Based on these results, we concluded that the oxidation condition leads to the selective oxidation of the ring-closing thioether bond without causing any modifications on proteinogenic amino acid residues except for a Met residue.

Next, we investigated if the oxidized thioether bond is more susceptible to cleavage by CID compared to amide bonds within the peptide. Since the cleavage of a thioether bond in a cyclic peptide generates the linearized product ion with the identical mass to the precursor ion, it cannot be monitored on the mass chromatogram. To circumvent this issue, we utilized a linear peptide containing a thioether linkage in the middle of the sequence (Ac-AC-[thioether bond]-AKAF-NH_2 (**2**)) (Fig. 3a) for

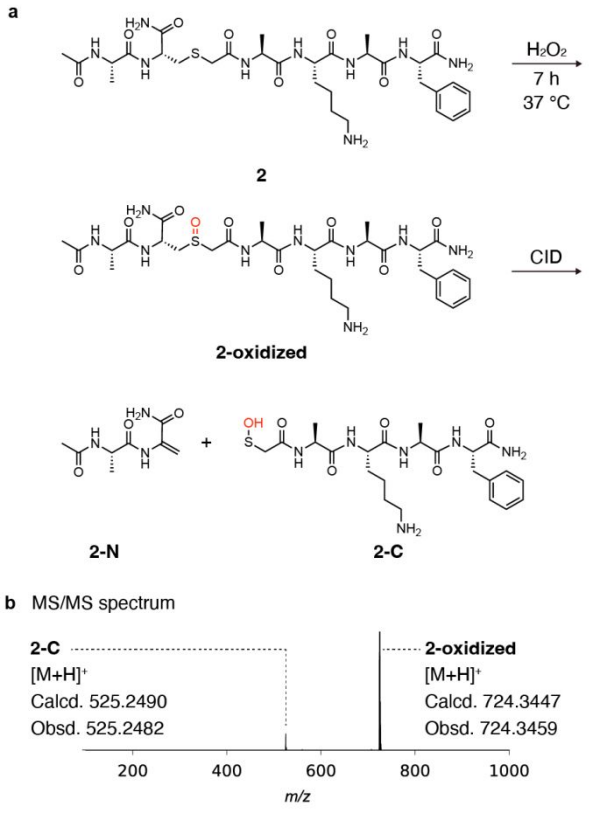

Fig. 3 (a) A proposed mechanism of collision-induced fragmentation of peptide **2** after oxidation. The **2-N** and **2-C** are expected to be generated. (b) MS/MS spectrum of **2-oxidized** with collision energy of 18 eV. MS/MS spectra with different collision energies are shown in Fig. S7.

this investigation. The peptide was oxidized and then subjected to LC-MS/MS analysis. The MS/MS spectra were recorded at various collision energies (Fig. S7). At lower collision energies, no significant fragment was observed on the spectra. As the collision energy was increased, the expected fragment peak generated by the cleavage at the sulfoxide group (**2-C**) emerged as the predominant product ion (Fig. 3b). On further increasing the collision energy, fragments resulting from cleavages at amide bonds started to appear (Fig. S7b). The preferential cleavage at the sulfoxide group supports the validity of our strategy that the CID of oxidized teMPs would facilitate sequence identification of given teMPs.

We then examined the efficiency of the fragment identification on an MS/MS spectrum after thioether oxidation, to evaluate the utility of the oxidation method for teMP sequence deconvolution. Peptide **1** was subjected to tandem mass spectrometry after oxidation (Fig. 4a). For peptide sequence deconvolution based on CID, identification of the fragments generated by amide bond cleavage is important. Without oxidation, fragments are generated via cleavages at arbitrary two amide bonds in the cyclic peptide. Therefore, many possible fragments (7 ring opening positions × 6 residual amide bonds × N/C-terminal fragments = 84 fragments) need to be considered (Fig. 4b, left). This increases the spectral complexity and decreases the intensity of each fragment ion as shown in Fig. 4c, which makes it challenging to identify the correct sequence with high confidence. On the other hand, oxidized teMPs can be first primarily cleaved at the sulfoxide moiety to generate the linearized ion with sulfenic acid and dehydroalanine residue, which further undergoes amide fragmentations to give controlled and limited fragmentation patterns (1 ring opening position × 7 amide bonds × N/C-terminal fragments = 14) (Fig. 4b, right). Upon the CID of **1-oxidized**, all possible N/C-terminal fragments after the linearization by the cleavage at sulfoxide were identified (Fig. 4d). The simpler fragmentation pattern facilitates identifying the peptide sequence. To examine the scope of the linearization-assisted fragmentation method, we conducted oxidation and fragmentation of two more teMPs (**S1** and **S2**). Both peptides gave high coverage of the fragments (Fig. S8). (93% for both **S1** and **S2**.) These results demonstrated the utility of oxidization for enhancing sequence deconvolution of teMPs.

Finally, we evaluated the utility of our linearization-assisted fragmentation method in identifying the sequence of teMPs from a sequence-randomized library. The second, fourth, and sixth residues of peptide **1** were virtually randomized using all proteinogenic amino acid residues except for Cys residue and Ile residue, which is indistinguishable from Leu due to the same mass value (Table 1). Similarly, the first, third, and fifth residues of peptides **S1** and **S2** were also virtually randomized (Table S1 and S2). Each virtual sequence library consists of 18^3 = 5,832 members. The MS/MS spectra of oxidized peptide **1**, **S1**, and **S2** obtained in the previous section were used as inputs, and the peptide sequences were deduced from the sequences in the virtual library using a *de novo* peptide sequencing program, UNIFI. As a result, the correct sequences were successfully selected as the top candidates among the sequences in the library. As described in the previous section, over 90% of the possible fragments generated by amide bond cleavage of the linearized product ion were observed for all three peptides. Besides, the sum of the intensities of the assigned fragment ions (assigned intensity) for the correct sequence was higher than other alternative assignments (67%, 62%, and 36% of the total observed ions for peptide **1**, **S1**, and **S2**, respectively). For example, as shown in Table 1, the correct assignment for

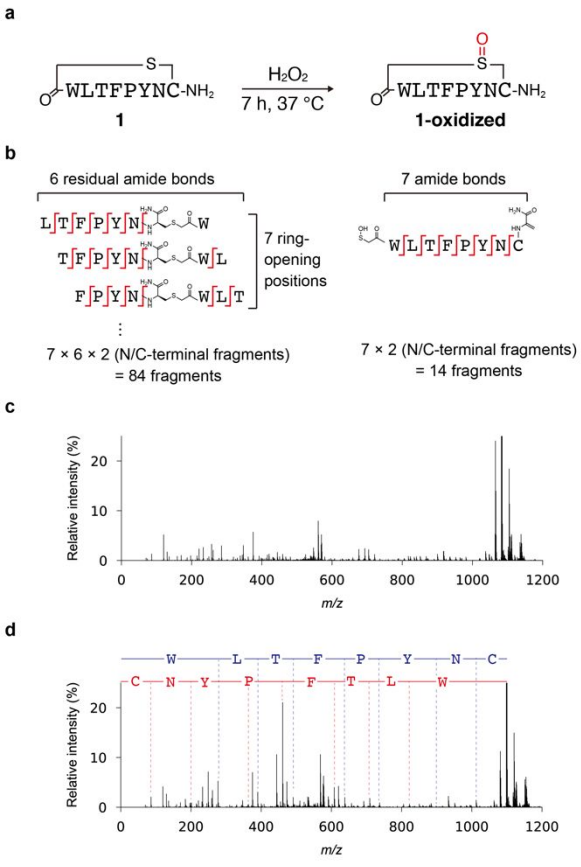

Fig. 4 The MS/MS analysis of **1** after oxidation. (a) Scheme of the oxidation of **1** to generate **1-oxidized**. (b) Comparison of the number of possible b and y fragments generated from **1** (left) and **1-oxidized** (right). (c, d) MS/MS spectrum of (c) **1** and (d) **1-oxidized**. N-terminal and C-terminal fragments are indicated by dashed lines in blue and red, respectively. For better visualization, the spectra have been vertically expanded, and the highest parent ion peaks are not shown up to their full height."

Table 1. Sequence identification of peptide **1** from a virtual library.

A virtual library of peptide **1**

	Sequence	$[M+H]^+$ (Calculated)		
Correct sequence	WLTFPYNC	1098.4713		
Candidate[a]	Sequence	Mass error (Δppm)[b]	Matched fragments	Assigned intensity (%)
1	WLTFPYNC	–0.15	14 (100%)	67
2	WFTLPYNC	–0.15	10 (71%)	62
3	WLTYPFNC	–0.15	10 (71%)	59

[a]Only sequences corresponding to the strongest parent ion peak with more than 50% matched fragments are shown.

[b]Mass errors from the mass value of the correct sequence are shown.

peptide **1** (candidate 1) has higher matched fragments (100%) and assigned intensity (67%) than alternative assignments (candidates 2 and 3). The similar lists of assignments for peptides **S1** and **S2** are shown in Tables S1 and S2, respectively. The result that the correct sequence was selected as the top candidate for all three peptides demonstrates the reliability of the linearization-assisted sequencing method. The results have demonstrated the utility of the oxidation-assisted linearization method for *de novo* sequencing of teMPs.

In summary, we established a method for the *de novo* sequencing thioether macrocyclic peptides using LC-MS/MS. Application of the sequencing method for affinity selections, membrane permeability measurements, and other assays of a large teMP library would facilitate peptide drug discovery in the future. This sequencing method requires a thioether linkage in MPs, and thus, it is not a general method applicable to any MP. However, the method is not only applicable to teMPs bearing a thioether linkage between N-terminus and a Cys side chain demonstrated here, but also potentially applicable to other thioether-closed macrocyclic peptides, e.g., teMPs with a side chain–side chain thioether linkage[25], monocyclic or bicyclic peptides closed via benzyl sulfide linkages[26–29], and naturally occurring lanthipeptides closed by lanthionine residues[30,31].

S.S. acknowledges financial support from CREST (JPMJCR21N5), Japan Science and Technology Agency. J.M. acknowledges financial support from PRESTO (JPMJPR21AF), Japan Science and Technology Agency. Y.G. acknowledges financial support from JSPS KAKENHI (JP20H02866, JP23H04546, 24H01754, and 24K01634). H.S. acknowledges financial support from JSPS KAKENHI (JP20H05618).

Conceptualization: Y.G., S.S. Investigation: A.H., Y.G., J.M. Project Administration: J.M., S.S. Resources: Y.G., H.S., J.M., S.S. Supervision: Y.G., J.M., Y.S., S.S. Writing – original draft: A.H., Y.G., J.M., S.S. Writing – review & editing: Y.S., H.S.

Data availability

The data supporting this article have been included as part of the Supplementary Information.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 A. A. Vinogradov, Y. Yi and H. Suga, *J. Am. Chem. Soc.*, 2019, **141**, 4167–4181.
2 K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski and R. J. Knapp, *Nature*, 1991, **354**, 82–84.
3 T. Liu, Z. Qian, Q. Xiao and D. Pei, *ACS Comb. Sci.*, 2011, **13**, 537–546.
4 Y. Zhang, S. Zhou, A. S. Wavreille, J. DeWille and D. Pei, *J. Comb. Chem.*, 2008, **10**, 247–255.
5 D. H. Lau, L. Guo, R. Liu, A. Song, C. Shao and K. S. Lam, *Biotechnol. Lett.*, 2002, **24**, 497–500.
6 Y. Gao and T. Kodadek, *ACS Comb Sci*, 2015, **17**, 190–195.
7 M. A. Lee, J. S. Brown, C. E. Farquhar, A. Loas and B. L. Pentelute, *ChemRxiv*, DOI:10.26434/chemrxiv-2024-vpstm.
8 A. Bruce, V. Adebomi, P. Czabala, J. Palmer, W. M. McFadden, Z. C. Lorson, R. L. Slack, G. Bhardwaj, S. G. Sarafianos and M. Raj, *Angew. Chem., Int. Ed.*, 2024, **136**, e202320045.
9 M. Garrigou, B. Sauvagnat, R. Duggal, N. Boo, P. Gopal, J. M. Johnston, A. Partridge, T. Sawyer, K. Biswas and N. Boyer, *J. Med. Chem.*, 2022, **65**, 8961–8974.
10 W. M. Hewitt, S. S. F. Leung, C. R. Pye, A. R. Ponkey, M. Bednarek, M. P. Jacobson and R. S. Lokey, *J. Am. Chem. Soc.*, 2015, **137**, 715–721.
11 C. N. Kelly, C. E. Townsend, A. N. Jain, M. R. Naylor, C. R. Pye, J. Schwochert and R. S. Lokey, *J. Am. Chem. Soc.*, 2021, **143**, 705–714.
12 J. H. Faris, E. Adaligil, N. Popovych, S. Ono, M. Takahashi, H. Nguyen, E. Plise, J. Taechalertpaisarn, H.-W. Lee, M. F. T. Koehler, C. N. Cunningham and R. S. Lokey, *J. Am. Chem. Soc.*, 2024, **146**, 4582–4591.
13 Y. Hosono, J. Morimoto and S. Sando, *Org. Biomol. Chem.*, 2021, **19**, 10326–10331.
14 J. Morimoto, Y. Hosono and S. Sando, *Bioorg. Med. Chem. Lett.*, 2018, **28**, 231–234.
15 J. H. Lee, A. M. Meyer and H. S. Lim, *Chem. Commun.*, 2010, **46**, 8615–8617.
16 L. S. Simpson and T. Kodadek, *Tetrahedron Lett.*, 2012, **53**, 2341–2344.
17 K. J. Lee and H.-S. Lim, *Org Lett*, 2014, **16**, 5710–5713.
18 S. Das and A. Nag, in *Methods Enzymol*., 2024, **698**, 141–167.
19 T. Passioura and H. Suga, *Chem. Commun.*, 2017, **53**, 1931–1940.
20 Y. Huang, M. M. Wiedmann and H. Suga, *Chem. Rev.*, 2019, **119**, 10360–10391.
21 Y. Goto and H. Suga, *Acc. Chem. Res.*, 2021, **54**, 3604–3617.
22 Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami and H. Suga, *ACS Chem. Biol.*, 2008, **3**, 120–129.
23 A. Kao, C. Chiu, D. Vellucci, Y. Yang, V. R. Patel, S. Guan, A. Randall, P. Baldi, S. D. Rychnovsky and L. Huang, *Mol. Cell. Proteomics*, 2011, **10**, M110.002170.
24 S. V. Winter, F. Meier, C. Wichmann, J. Cox, M. Mann and F. Meissner, *Nat. Methods*, 2018, **15**, 527–530.
25 Y. Sako, Y. Goto, H. Murakami and H. Suga, *ACS Chem. Biol.*, 2008, **3**, 241–249.
26 P. Timmerman, J. Beld, W. C. Puijk and R. H. Meloen, *ChemBioChem*, 2005, **6**, 821–824.
27 F. P. Seebeck and J. W. Szostak, *J. Am. Chem. Soc.*, 2006, **128**, 7150–7151.
28 C. Heinis, T. Rutherford, S. Freund and G. Winter, *Nat. Chem. Biol.*, 2009, **5**, 502–507.
29 Y. V. Guillen Schlippe, M. C. T. Hartman, K. Josephson and J. W. Szostak, *J. Am. Chem. Soc.*, 2012, **134**, 10469–10477.
30 L. An and W. A. van der Donk, *Comprehensive Natural Products III: Chemistry and Biology*, Elsevier, 2020, 119–165.
31 K. J. Hetrick, M. C. Walker and W. A. Van Der Donk, *ACS Cent. Sci.*, 2018, **4**, 458–467.

Data Availability Statement

The data supporting this article have been included as part of the Supplementary Information.