

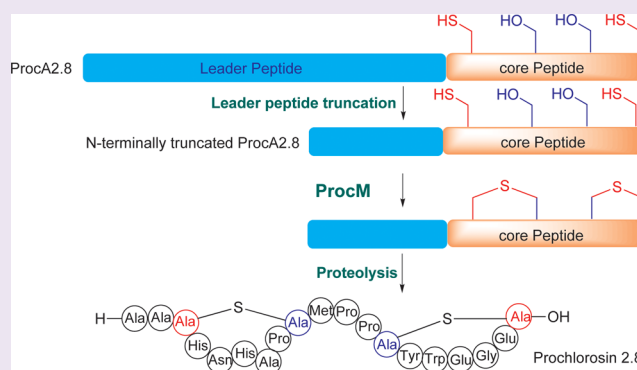
# High Divergence of the Precursor Peptides in Combinatorial Lanthipeptide Biosynthesis

Qi Zhang, Xiao Yang, Huan Wang, and Wilfred A. van der Donk\*

Department of Chemistry, Howard Hughes Medical Institute, University of Illinois at Urbana–Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, United States

## S Supporting Information

**ABSTRACT:** Lanthionine-containing peptides (lanthipeptides) are a rapidly growing family of polycyclic peptide natural products belonging to the large class of ribosomally synthesized and post-translationally modified peptides (RiPPs). These compounds are widely distributed in taxonomically distant species, and their biosynthetic systems and biological activities are diverse. A unique example of lanthipeptide biosynthesis is the prochlorosin synthetase ProcM from the marine cyanobacterium *Prochlorococcus* MIT9313, which transforms up to 29 different precursor peptides (ProcAs) into a library of lanthipeptides called prochlorosins (Pcns) with highly diverse sequences and ring topologies. Here, we show that many ProcM-like enzymes from a variety of bacteria have the capacity to carry out post-translational modifications on highly diverse precursor peptides, providing new examples of natural combinatorial biosynthesis. We also demonstrate that the leader peptides come from different evolutionary origins, suggesting that the combinatorial biosynthesis is tied to the enzyme and not a specific type of leader peptide. For some precursor peptides encoded in the genomes, the leader peptides apparently have been truncated at the N-termini, and we show that these N-terminally truncated peptides are still substrates of the enzymes. Consistent with this hypothesis, we demonstrate that about two-thirds of the ProcA N-terminal sequence is not essential for ProcM activity. Our results also highlight the potential of exploring this class of natural products by genome mining and bioengineering.



Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a major class of natural products, as revealed by the genome sequencing efforts of the past decade.<sup>1</sup> These compounds are produced in all three domains of life and possess vast structural diversity. Among the best-studied RiPPs are lanthipeptides, a class of compounds that are distinguished by the presence of thioether cross-linked amino acids named lanthionines and methylanthionines.<sup>2–7</sup> Many lanthipeptides, such as the commercially used food preservative nisin, have potent antimicrobial activity and are termed lantibiotics.<sup>8,9</sup> Lanthipeptides are widely distributed among taxonomically distant species<sup>10</sup> and are currently grouped into four distinct classes according to their biosynthetic machineries.<sup>3,10</sup> Like all RiPPs, lanthipeptides are generated from a linear precursor peptide, which is generically termed LanA. This precursor peptide consists of a C-terminal core peptide, where all post-translational modifications take place, and an N-terminal leader peptide that is important for post-translational modification and that is subsequently removed by proteolysis (Figure 1).<sup>1,11</sup> The installation of the (methyl)lanthionine thioether bridges is achieved by the initial dehydration of Ser and Thr residues in the precursor peptides, followed by stereoselective intramolecular Michael-type addition of Cys thiols to the newly formed dehydroamino acids (Figure 1).

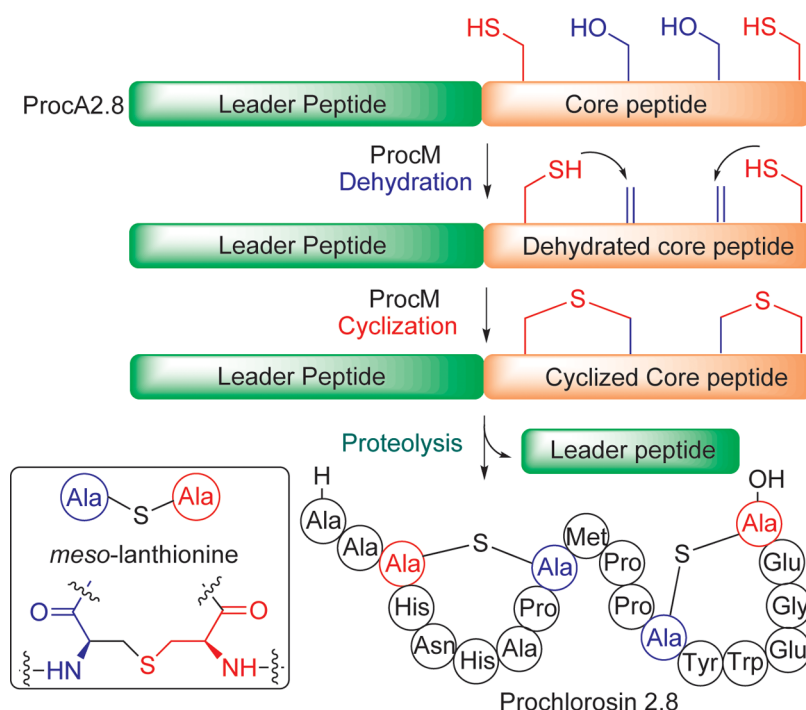
An intriguing example of a lanthipeptide synthetase is ProcM, a class II enzyme (generically termed LanM) from the planktonic marine cyanobacterium *Prochlorococcus* MIT9313.<sup>12</sup> ProcM acts on up to 29 different precursor peptide substrates (ProcAs) and produces a library of lanthipeptides termed prochlorosins (Pcns) that possess highly diverse sequences and ring topologies,<sup>12,13</sup> representing a remarkable example of natural combinatorial biosynthesis. The biological role of Pcns is currently elusive, but they are believed to be functional, as they were found to be produced in the host strain, and their biosynthetic genes were transcribed in response to changes in environmental conditions.<sup>12</sup>

The intriguing combinatorial biosynthesis of Pcns provides an interesting model to investigate the evolution of natural product diversity and the molecular origins for the remarkable substrate tolerance displayed by the enzyme. The ProcA substrates have an unusually long leader peptide compared to that of other lanthipeptide substrates, raising the question of whether this longer leader peptide might be correlated with the

Received: August 5, 2014

Accepted: September 22, 2014

Published: September 22, 2014



**Figure 1.** Schematic representation of the biosynthetic pathway of lanthipeptides exemplified by prochlorosin 2.8. A shorthand notation for lanthionine structures is shown in the box. Leader and core peptides are not shown in proportion to their actual lengths.

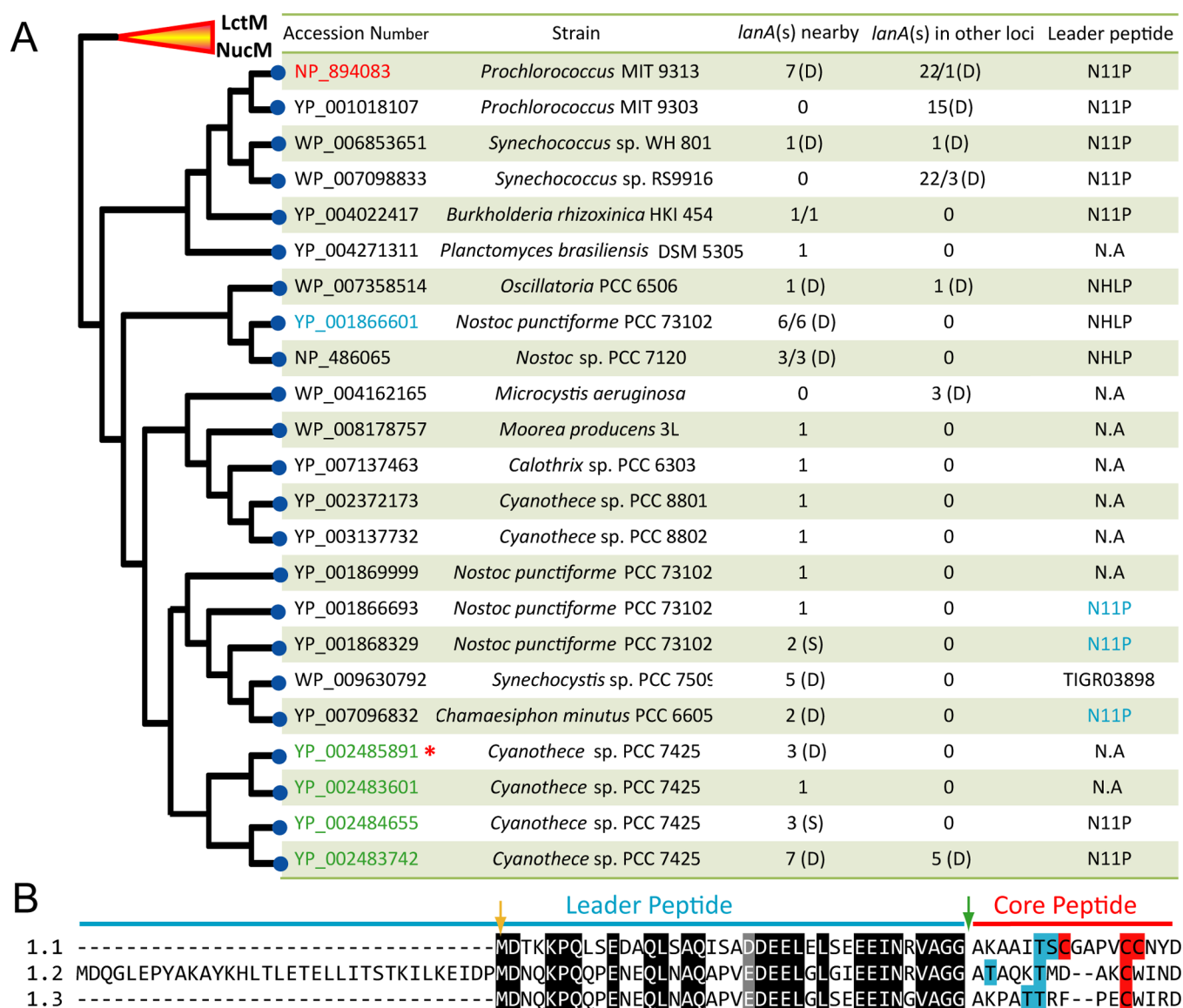
large diversity of substrates that ProcM processes. The ProcA leader peptides are also unique in that they have sequence homology with the Nif11 proteins.<sup>14</sup> The exact function of the Nif11 proteins is not known, but they are thought to play a role in nitrogen fixation, as their genes cluster with other nitrogen fixation genes.<sup>15</sup> An alternative model that has been proposed is that it is the cyclization active site of ProcM that is unique and that confers the ability to cyclize a wide variety of substrates. Here, we present bioinformatic and biochemical investigations on lanthipeptide biosynthetic systems employing ProcM-like enzymes. We show that the precursor peptides for lanthipeptide biosynthesis are highly divergent among different biosynthetic systems and that many ProcM-like lanthipeptide synthetases can engage in combinatorial biosynthesis by tolerating precursor peptides with highly diverse core sequences.

## RESULTS AND DISCUSSION

**Genome Mining of LanAs Associated with ProcM Analogues.** ProcM catalyzes both dehydration and cyclization reactions to transform the linear ProcA peptides into a panel of Pcns. The enzyme contains a conserved CCG motif, suggesting that it likely uses three Cys residues for binding of an active site zinc ion,<sup>10</sup> unlike other lanthipeptide cyclases that utilize two Cys and one His as zinc ligands. This zinc site has been shown to be important for the cyclization reaction.<sup>16,17</sup> The presence of three Cys ligands may account in part for the high substrate tolerance of ProcM, as model studies have demonstrated that the reactivity of thiolate nucleophiles ligated to  $Zn^{2+}$  is enhanced with an increased number of thiolate ligands.<sup>18</sup> We previously have shown that LanMs containing the CCG motif cluster together to form a distinct subclade in the LanM phylogenetic tree, suggesting that these ProcM-like enzymes evolved independently.<sup>10</sup> To investigate whether other ProcM-like enzymes also have multiple structurally diverse LanA

substrates, we performed a genome-wide examination of the LanAs associated with ProcM analogues. This investigation showed that similar lanthipeptide synthetases can have very diverse precursor peptides, which vary significantly in both the number of putative substrates and their amino acid sequence (Figure 2A and Supporting Information Table 1 for a detailed list). Unlike ProcM, many ProcM analogues have only a single LanA substrate, as found for most lanthipeptides. On the other hand, many other ProcM-like enzymes have several substrates, harboring either very similar or highly diverse core peptide sequences (Figure 2A and Supporting Information Table 1). A notable example is a ProcM analogue from *Prochlorococcus* MIT9303, which has almost the same sequence as ProcM (95% identity, 97% similarity). This organism encodes 15 putative substrate peptides (Figure 2A), and none of these substrates have a core peptide similar in sequence to the core peptides of the 29 ProcAs. This observation raises the possibility that many, or all, Pcns may lack a biological function and that they may represent intermediaries during evolution. Conversely, these two lanthipeptide biosynthetic systems may represent a remarkable example of convergent evolution of functional lanthipeptides, if the structurally diverse peptides fulfill similar roles in the closely related species.

The vast majority of ProcM-like enzymes with the CCG motif were found in cyanobacteria, a rich source of RiPP natural products,<sup>19</sup> but they are also present in other phyla (Figure 2A). The precursor peptide genes associated with these ProcM-like synthetases appear to have different evolutionary origins. Many of the LanA leader peptides are members of the N11P family, but several other leader peptides belong to the nitrile hydratase leader peptide (NHLP) family (Figure 2A).<sup>14</sup> NHLP is highly similar to the  $\alpha$  subunit of nitrile hydratase (NHase) but lacks about 30 amino acids in the middle of the NHase sequence that are important for binding of a catalytic metal ion.<sup>14,20</sup> We note that many cyanobacteria (e.g., *Prochlorococcus*



**Figure 2.** Genome mining of precursor peptide genes associated with ProcM-like enzymes. (A) Bayesian MCMC phylogram of ProcM-like enzymes (protein sequence) and a summary of the number of their putative LanA substrates. The lacticin 481 synthetase LctM and nukacin synthetase NucM were used as an outgroup for Bayesian MCMC analysis, which is shown as an orange triangle. The detailed Bayesian MCMC tree is shown in Supporting Information Figure 29. The putative *lanA* genes were categorized into two groups based on whether they are spatially close to their associated *lanM* genes. For LanAs that lack Cys residues, the substrates are shown as total number of *lanAs*/the number of *lanAs* that do not code for Cys. If an enzyme had multiple LanA substrates, then the core peptide sequences were aligned to examine whether these precursor peptides are similar (S, for which the Ser/Thr and Cys residues are aligned well) or diverse (D, for which Ser/Thr and Cys residues did not align well). ProcM from *Prochlorococcus* MIT9313, NpnM from *Nostoc punctiforme* PCC 73102, and four LanMs (CyanM1–4) from *Cyanothece* sp. PCC 7425 are highlighted in red, blue, and green, respectively. CyanM1 is highlighted by an asterisk. Three groups of substrates shown in blue contain leader peptides that share very weak similarities with the N11P family ( $1 \times 10^{-4} < e\text{-value} < 0.1$ ). NA indicates that the precursors do not belong to any known protein family. (B) Sequence alignment of CyanA1.1–1.3, showing that CyanA1.1 and CyanA1.3 may have been truncated at their N-termini. Alternatively, the open reading frame (ORF) annotation of CyanA1.2 could be incorrect, and its translation start codon may instead be at the light brown arrow, like CyanA1.1 and 1.3. Completely conserved and highly conserved residues in the leader peptides are shown in black and gray boxes, respectively. Ser/Thr and Cys residues in the core peptides are shown in blue and red boxes, respectively. The proteolytic cleavage site is indicated by a green arrow. For detailed information on precursor peptide sequence and the procedures for bioinformatics analysis, see Supporting Information Table 1 and Supporting Information Methods.

strains) do not have Nif11 and/or NHase genes and therefore the evolutionary relationship between these enzymes and the leader peptides is not clear. In addition to the N11P and NHLP leader peptides, the LanAs from the planktonic cyanobacterium *Synechocystis* sp. PCC 7509 have leader peptides that belong to the TIGR03898 family (Figure 2A).<sup>21</sup> This family mostly consists of the leader peptides from LanAs in firmicutes (e.g.,

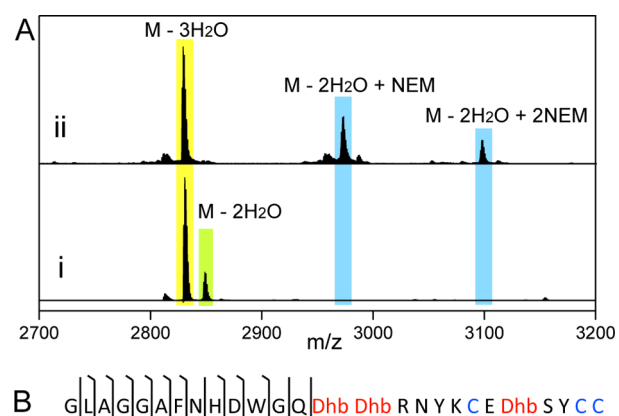
MrsA in mersacidin biosynthesis). The five LanAs in *Synechocystis* sp. PCC 7509 may thus be a result of horizontal gene transfer, possibly from a firmicute. Furthermore, many of the leader peptides of LanAs associated with ProcM-like enzymes do not belong to any known protein family (Figure 2A), illustrating the diverse origins of lanthipeptide precursor genes that cluster with the *lanM* genes in the ProcM clade.



Another interesting observation is that some LanAs may have been truncated at their N-termini. For example, three putative precursor genes (*cyanA1.1–1.3*) were found adjacent to a ProcM-like LanM gene (*cyanM1*) in the genome of the cyanobacterium *Cyanothece* sp. PCC 7425. Compared with CyanA1.2, CyanA1.1 and CyanA1.3 appear to be much shorter in length (Figure 2B). Analysis of the 5'-untranslated regions (UTRs) of CyanA1.1 and CyanA1.3 showed that their sequences are very similar to the coding region of the N-terminus of CyanA1.2 and that the putative truncation in CyanA1.1 might be caused by a point mutation resulting in a stop codon in the 5'-UTR (Supporting Information Figure 1). An alternative interpretation is that the longer CyanA1.2 is based on an incorrectly assigned start codon; translation initiation at a later start codon would result in a peptide that has the same length as that of CyanA1.1 and CyanA1.3 (Figure 2B). We did observe what appears to be a true N-terminally truncated *procA* gene (here named *procAt.1*) in the genome of *Prochlorococcus* MIT9313, which escaped our previous annotation for *procA* genes.<sup>12</sup> The truncation of ProcAt.1 seems to be caused by the transformation of the original start codon to an ochre stop codon, resulting in translation initiation at a different downstream start codon (Supporting Information Figure 2).

**Much of the N-Termini of ProcAs Are Dispensable for ProcM Activity.** Unlike the newly identified ProcAt.1, whose leader peptide has 45 amino acids, all of the other ProcA leader peptides have more than 60 amino acids (Supporting Information Figure 3). As mentioned above, ProcA leader peptides belong to the N11P family<sup>14</sup> and are much longer than those of other class II lantibiotics (e.g., as a typical example, the leader peptide of LctA for lactacin 481 biosynthesis has only 23 residues<sup>22,23</sup>). It has been shown that the lactacin 481 synthetase LctM recognizes both the LctA leader and core peptides<sup>24</sup> and thus one possible role of the much longer leader peptides of ProcAs is that they provide additional recognition elements for ProcM, allowing for the remarkably high substrate tolerance. In this scenario, N-terminally truncated LanAs are likely the vestiges of precursor peptide divergence during evolution and might not necessarily be real substrates. To test whether ProcAt.1 is a ProcM substrate, we coexpressed ProcAt.1 with ProcM in *Escherichia coli*, and the resulting peptide was digested by endoprotease Glu-C. High-resolution matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) mass spectrometry (MS) analysis showed that ProcAt.1 had been dehydrated up to three times (Figure 3 and Supporting Information Figure 4). N-Ethylmaleimide (NEM) derivatization, which was employed to derivatize any free cysteine residues, showed that although lanthionine formation for the 2-fold dehydration product is incomplete, the 3-fold dehydration product is fully cyclized (Figure 3), strongly suggesting that ProcAt.1 is a true substrate of ProcM.

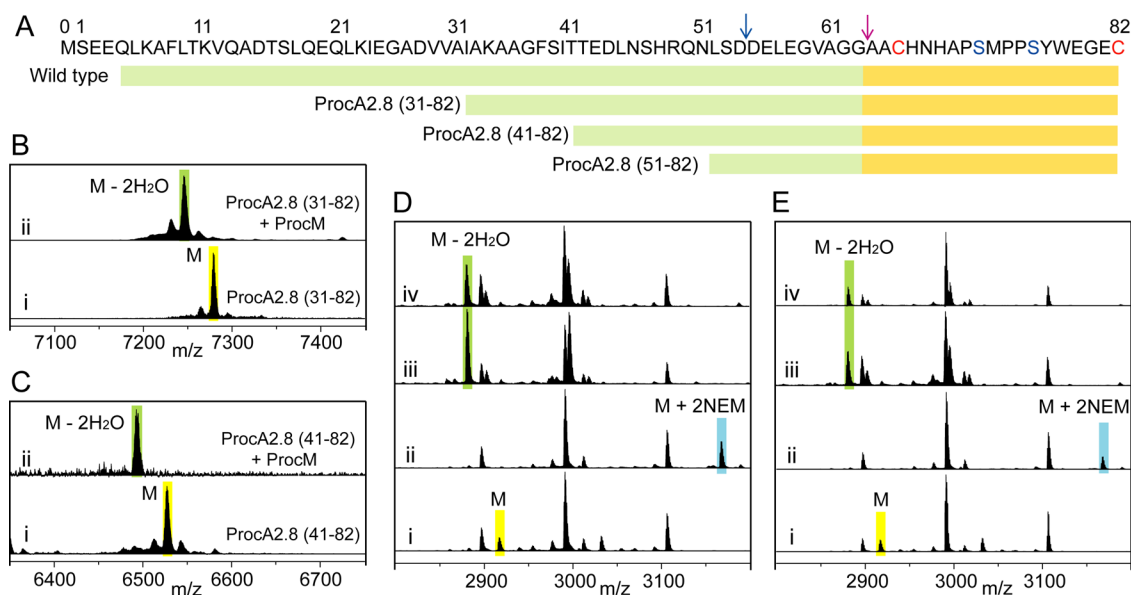
To further interrogate the importance of the full-length ProcA leader peptide, we made a series of ProcA2.8 mutants lacking 30, 40, and 50 N-terminal amino acids (Figure 4A) and coexpressed these mutants with ProcM in *E. coli*. MALDI-ToF-MS analysis clearly showed that mutants lacking the 30 and 40 N-terminal residues were fully dehydrated by ProcM (Figure 4B,C). The peptides were subsequently digested by endoprotease Asp-N and subjected to NEM derivatization. Compared with the control peptides that were expressed in the absence of ProcM and that were fully derivatized by NEM on their two free Cys residues, the ProcM-modified peptides did not react



**Figure 3.** Modification of ProcAt.1 by ProcM. (A) MALDI-ToF-MS analysis of ProcAt.1 that was obtained by coexpression with ProcM and treated with endoprotease Glu-C (trace i) and subsequently derivatized by NEM (trace ii). (B) Sequence of ProcAt.1 modified by ProcM and treated with Glu-C. The ESI-MS/MS fragmentation pattern for the 3-fold dehydrated species is shown (the MS/MS data is presented in Supporting Information Figure 4).

with NEM (Figure 4D,E), indicating that these peptides were fully cyclized. MS/MS analysis confirmed that the correct lanthionine rings of ProcA2.8 were produced in both truncated substrates (Supporting Information Figure 5). Hence, ProcM is able to perform catalysis not only on substrates with dramatically varied core sequences but also on mutants that are significantly truncated in the leader peptide. Although the dispensability of part of the N-terminal leader peptides has previously been shown in the biosynthesis of the lantibiotic lactacin 481,<sup>23</sup> the lasso peptide MccJ25,<sup>25</sup> and the biosynthetic enzymes of cyanobactin biosynthesis,<sup>26–28</sup> the fact that about two-thirds of the leader peptide is dispensable for ProcM activity is surprising considering the high substrate tolerance of the enzyme. The possibility that the long ProcA leader peptides are correlated with ProcM's tolerance of highly varied core sequences is thus not supported by our studies. Whether ProcM could modify the peptide lacking the 50 N-terminal residues could not be determined because this mutant peptide could not be obtained, regardless of whether the peptide was coexpressed with ProcM or not, suggesting that a certain minimum length of ProcA2.8 might be necessary for peptide stability, at least in *E. coli*.

**Lanthipeptides from *Cyanothece* sp. PCC 7425.** Since, at present, only the prototypical ProcM has been shown to actually carry out combinatorial biosynthesis, we decided to investigate for a subset of ProcM-like enzymes whether they are active with their associated LanA peptides with diverse leader peptides. We first investigated a particularly interesting lanthipeptide biosynthetic system in the marine cyanobacterium *Cyanothece* sp. PCC 7425, a strain that also produces several cyanobactins.<sup>29</sup> The genome of this organism encodes four LanMs (here termed CyanM1–4) that consist of a separate phylogenetic subclade (Figures 2A and 5A). Two sets of three *lanA* genes (*cyanA1.1–1.3* and *cyanA4.1–4.3*) were found adjacent to *cyanM1* and *cyanM4*, respectively, whereas *cyanM2* has only a single *lanA* gene (*cyanA2.0*) nearby (Figure 5A). Multiple *lanA* genes (*cyanA3.1–3.7*) were found adjacent to *cyanM3*. In addition, another locus approximately 2.8 Mb away from *cyanM3* encodes five more putative LanAs (*cyanA3.8–3.12*) with leader peptides that are very similar to



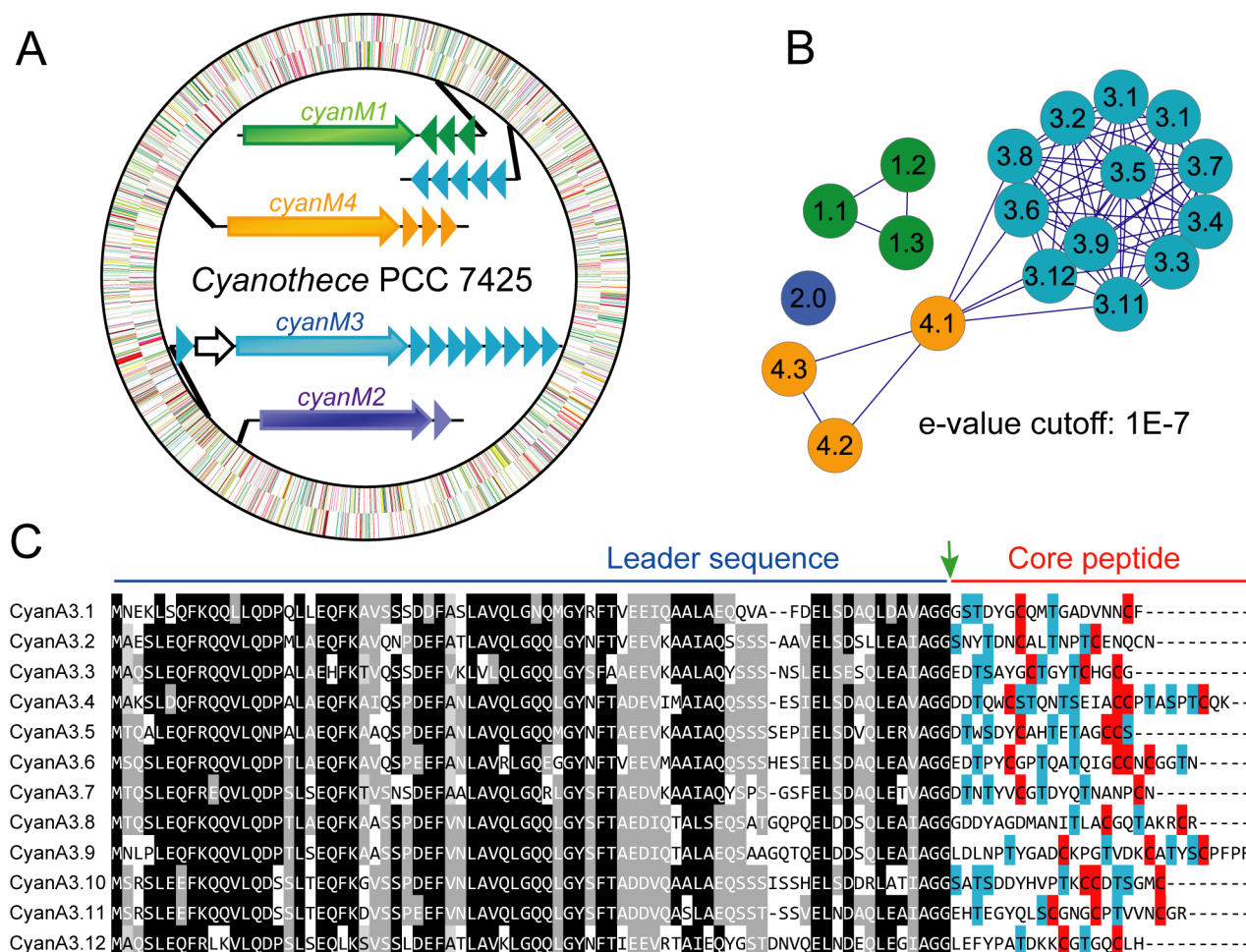
**Figure 4.** ProcM modification of truncated ProcA2.8 derivatives. (A) Sequence of ProcA2.8 and schematic representation of the truncation variants discussed in this study. The purple arrow shows the physiological proteolytic cleavage site for leader peptide removal. The blue arrow shows the endoprotease Asp-N site that was used in this study to shorten the peptide and allow better analysis of the post-translational modifications in the core peptide. (B) MALDI-ToF MS analysis of ProcA2.8-(31–82) that was obtained either by expressing the peptide alone (trace i) or by coexpression with ProcM (trace ii). (C) MALDI-ToF MS analysis of ProcA2.8-(41–82), presented in the same manner as for ProcA2.8-(31–82) in panel B. (D) ProcA2.8-(31–82) peptides (unmodified and modified) were digested by Asp-N and subsequently treated with NEM. Trace i shows the unmodified peptide before NEM derivatization, and trace ii demonstrates complete NEM derivatization of this peptide. Traces iii and iv show the ProcM-modified peptide before and after NEM treatment, respectively. No derivatization of the modified peptides is observed, strongly suggesting formation of lanthionine rings in the ProcM-modified peptide. (E) MALDI-ToF MS analysis of Asp-N-digested ProcA2.8-(41–82). The data are shown as in panel D. In all of the MS spectral data shown, the signals corresponding to the unmodified and the ProcM-modified peptides are highlighted in yellow and green, respectively, whereas the NEM-derivatized peptides are highlighted in light blue. Part of the nonhighlighted peaks are derived from proteolysis products of the leader peptide and Asp-N.

those of CyanA3.1–3.7 but with lower sequence homology to the leader peptides of the other CyanAs (Figure 5A and Supporting Information Table 1). To correlate these precursor peptides with their corresponding enzymes, a sequence similarity network was constructed based on their leader sequences. The results show that CyanAs can be divided into four groups (Figure 5B), suggesting that they are substrates of four different enzymes.

CyanA1.1, CyanA1.2, and CyanA1.3 were first tested as potential substrates for CyanM1 by coexpression of each *cyanA1.1–1.3* individually with *cyanM1* in *E. coli*. MALDI-ToF MS analysis showed that all three peptides were modified by the enzyme (Supporting Information Figures 6–8). The 12 peptides CyanA3.1–3.12 have highly conserved leader regions but very diverse core sequences (Figure 5C) and thus the CyanM3–CyanA3.x system could be similar to the combinatorial biosynthetic system of the prochlorocins. To test this hypothesis, the 12 genes, *cyanA3.1–3.12*, were each coexpressed individually with *cyanM3*, and MALDI-ToF MS analysis showed that all peptides were modified by the enzyme (Supporting Information Figures 9–20), demonstrating very high substrate tolerance for CyanM3, similar to the observations with ProcM.<sup>12</sup> Some precursor peptides have sequences that allowed cleavage by commercially available proteases to remove most of the leader peptides without proteolysis in the core peptide. MALDI-ToF MS analysis of the proteolytic products clearly demonstrated that the dehydrations take place in the core peptide (Supporting Information Figures 16–18).

Given that CyanM1–4 form a distinct phylogenetic subclade, we reasoned that the three other enzymes might also have high substrate tolerance. In line with this proposal, we showed that CyanM4 can modify not only its putative substrate CyanA4.1 (Supporting Information Figure 21) but also a chimeric peptide consisting of the CyanA4.1 leader and CyanA1.2 core peptide (Supporting Information Figure 22); the latter is very different from the CyanA4.1 core peptide. However, CyanM4 did not modify CyanA1.2 (Supporting Information Figure 23), supporting the model that it is the cognate leader peptide (i.e., the leader peptide of CyanA4.1) that plays a requisite role for enzyme catalysis,<sup>11</sup> although probably not for substrate promiscuity.

**LanAs That Do Not Have a Cysteine Residue.** Cysteine is a required residue for formation of the thioether bridges of lanthipeptides. However, we noted that several putative LanAs associated with ProcM-like enzymes only have Ser/Thr residues and lack any Cys residues (Figure 2A and Supporting Information Table 1). These peptides include ProcA4.1, which is the only Cys-free peptide among 30 ProcAs (including the newly characterized ProcAt.1). To investigate whether ProcA4.1 is a ProcM substrate, the peptide was coexpressed with ProcM, and MALDI-ToF-MS analysis showed that the resulting peptide was dehydrated despite the lack of Cys residues (Figure 6A and Supporting Information Figure 24). Since the 29 other ProcA peptides all contain at least one Cys (Supporting Information Figure 3), the absence of a Cys in ProcA4.1 likely resulted from mutations in the core sequence of an ancestor LanA.



**Figure 5.** Lanthipeptide biosynthesis in *Cyanotheca* sp. PCC 7425. (A) Four lanthipeptide biosynthetic systems in *Cyanotheca* sp. PCC 7425, showing the gene clusters of each system and their locations in the genome. (B) Sequence similarity network based on the leader peptide sequence of CyanAs. Each node represents a leader peptide, and each edge (line) indicates a pair of nodes (leader peptides) that have a BlastP e-value more stringent than the cutoff value used ( $1 \times 10^{-7}$ ). Different biosynthetic systems are depicted by different colors. (C) High level of conservation in the N-terminal leader sequence and hypervariability of the C-terminal core peptide of CyanA3.1–3.12. The GG/GA protease cleavage site for leader peptide removal is marked by a green arrow. For the sequences of CyanA1, CyanA2, and CyanA4, see Supporting Information Table 1.

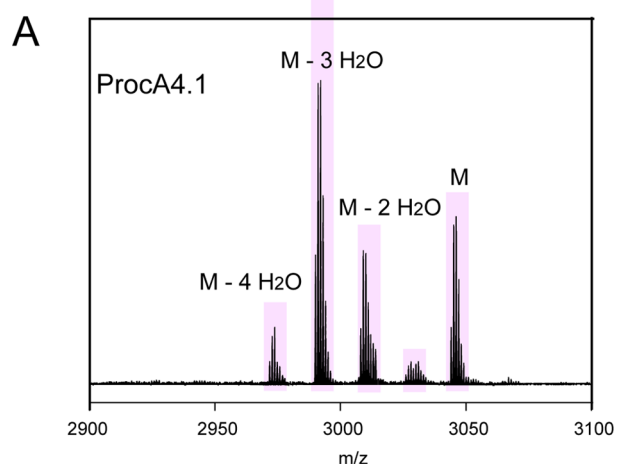
A more unusual example is found in *N. punctiforme* PCC 73102, which encodes putative RiPP precursor peptides that have very conserved putative leader peptides but highly diverse core peptides. Among the six peptides identified, none has a Cys residue (Supporting Information Table 1), but the genes encoding these peptides do cluster with a ProcM-like gene. Other genes encoding known RiPP biosynthetic enzymes (e.g., cyclodehydratases for thiazole/oxazole formation<sup>14,30</sup>) were not found in the genome of the organism, suggesting the peptides are putative LanM substrates. Notably, unlike ProcAs and CyanA3.x, whose leader peptides belong to the N11P protein family, the leader peptides of these putative RiPP precursors (here termed NpnAs) belong to the NHLP family (Figure 1A). To test whether these NHLP-containing and Cys-lacking peptides are LanM substrates, two NpnAs with very different core peptides (NpnA3 and NpnA6) were coexpressed with NpnM. MALDI-ToF MS analysis of the resulting peptides showed that both NpnAs were dehydrated (Figure 6B,C and Supporting Information Figures 25–26). Attempts to form a lanthionine by appending a five-amino acid sequence containing a Cys to NpnA3 were unsuccessful (Supporting Information Figure 27). Detailed sequence analysis showed that NpnM may have an impaired cyclase active site because the

enzyme has a QG motif instead of a HG motif found in almost all LanC and LanM proteins (Supporting Information Figure 28). The His in this motif is essential for cyclase activity of the class I lanthipeptide cyclase NisC,<sup>31</sup> and its mutation in NpnM may be a consequence of its substrates no longer requiring cyclization. These observations provide support for the model in which lanthipeptide synthetases coevolved with their substrates during the evolutionary process.<sup>10</sup>

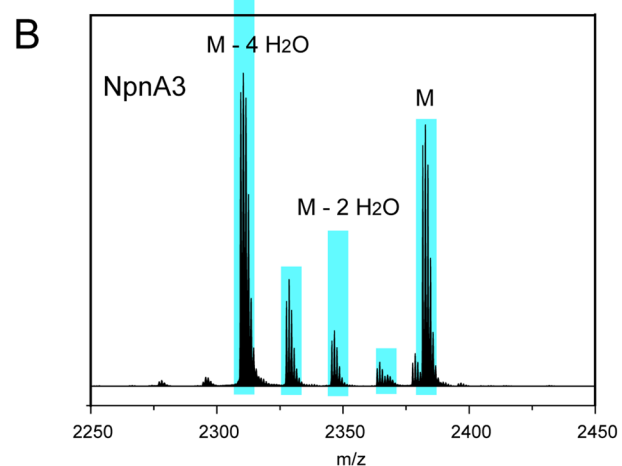
The function(s) of the Cys-lacking NpnAs is also an interesting question, since the mature products cannot be lanthipeptides. As previously noted,<sup>32,33</sup> a gene encoding a member of the zinc-dependent dehydrogenase enzymes (NpnJ) is encoded near the LanM gene. For the lantibiotics lactacin 3147 and carnolysin, the enzymes LtnJ and CrnJ convert Dha into D-Ala residues.<sup>33,34</sup> Hence, it is possible that the products of the gene cluster in *N. punctiforme* PCC 73102 are D-amino acid-containing peptides, if NpnJ hydrogenates the dehydroamino acids formed by NpnM. Unfortunately, we did not obtain soluble NpnJ by heterologous expression in *E. coli*. Similarly, previous attempts to complement a  $\Delta ltnJ$  mutant of *Lactococcus lactis* with *npnJ* were unsuccessful, and insoluble expression could not be ruled out.<sup>32</sup>



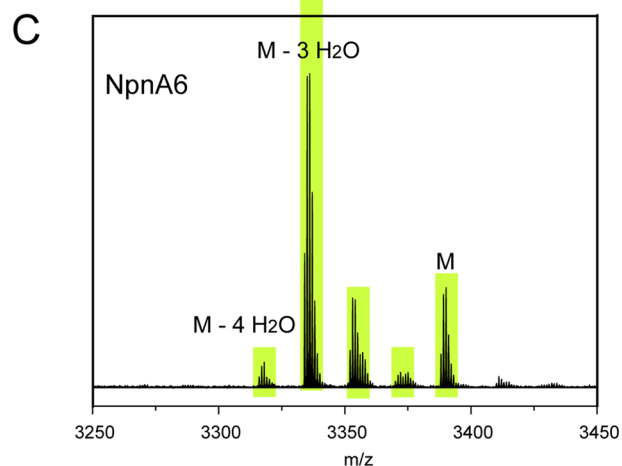
GGGA}R{Dhb}K{Dhb}A{N}V{PSD}L{PVR}A{PAM}S{Dhb}F{A}E{N}Q{T



LAK}G{A}L{Dha}V{G}L{A}I{Dhb}L{V}D{Dhb}L{K}R{L}DhbK



SVAG}G{F}I{GG}L{I}T{I}L{AV}GVTPF{Dhb}GDIVKA{Dhb}KKLDhbKK



**Figure 6.** Coexpression studies of Cys-lacking peptides with LanMs. (A) MALDI-ToF MS analysis of ProcA4.1 that was obtained by coexpression with ProcM. Also shown is the sequence of the ProcA4.1 core (obtained by TEV cleavage of a ProcA4.1 mutant containing an engineered TEV cleavage site just before the predicted core sequence) and the MS/MS fragmentation pattern for the 3-fold dehydrated species. (B) MALDI-ToF MS analysis of NpnA3 that was obtained by coexpression with NpnM in *E. coli*. Also shown is the sequence of

Figure 6. continued

endoproteinase Glu-C cleaved NpnA3 and the MS/MS fragmentation pattern for the 4-fold dehydrated species. (C) MALDI-ToF MS analysis of NpnA6 obtained similarly to that for NpnA3 in panel B. Also presented is the sequence and MS/MS fragmentation pattern for 3-fold dehydrated NpnA6. The MS/MS data for 3-fold dehydrated ProcA4.1, 4-fold dehydrated NpnA3, and 3-fold dehydrated NpnA6 are shown in Supporting Information Figures 24–26, respectively.

**Conclusions.** The tremendous structural and functional diversity of lanthipeptides raises many questions regarding their biological roles, origins, and evolutionary mechanisms.<sup>35–38</sup> The Pcn biosynthetic system is arguably one of the most remarkable examples of natural product combinatorial biosynthesis found in nature. By genome mining for putative lanthipeptides whose biosyntheses involve ProcM-like enzymes, we show that the LanA precursor peptides are highly diverse among different systems and that phylogenetically closely related lanthipeptide synthetases can be associated with very different sets of substrates. In addition, we demonstrated that much of the N-terminal ProcA leader peptide is not required for enzyme activity.

This work also extends the combinatorial biosynthesis paradigm to many other organisms by experimental demonstration that ProcM-like enzymes that have multiple nearby precursor peptides have the capacity to process a widely diverse set of core peptides. Given the sequence diversity of their leader peptides and the demonstration that much of the leader peptide is dispensable, this substrate tolerance is not imparted by the leader peptide and hence it is likely that it is a property of the ProcM enzyme clade. The findings herein thus highlight the potential of exploring lanthipeptides by genome mining and for biosynthetic engineering efforts to produce novel lanthipeptides with desired biological activities by mimicking the natural evolutionary process.

## METHODS

**Materials.** All oligonucleotides were purchased from Integrated DNA Technologies. Restriction endonucleases, DNA polymerases, T4 DNA ligase, and endoproteinase Asp-N and Glu-C were purchased from New England Biolabs. Media components for bacterial cultures were purchased from Difco laboratories. Chemicals were purchased from Fisher Scientific or from Aldrich unless noted otherwise. *E. coli* DH5 $\alpha$  was used as host for cloning and plasmid propagation, and *E. coli* BL21 (DE3) was used as a host for coexpression. *Cyanotheca* sp. PCC7425 was purchased from ATCC. Synthetic *npnM* gene was ordered from GeneArt, and synthetic *npnA* genes were ordered from Integrated DNA Technologies (IDT).

**General Methods.** All polymerase chain reactions (PCR) were carried out on a C1000 thermal cycler (Bio-Rad). DNA sequencing was performed by the Biotechnology Center at the University of Illinois at Urbana–Champaign, using appropriate primers. MALDI-ToF MS was carried out on Bruker Daltonics Ultraflex MALDI ToF/ToF mass spectrometer. ESI MS/MS analyses were performed on a Synapt ESI quadrupole ToF mass spectrometry system (Waters). Deisotoping and deconvolution of ESI MS/MS spectra were performed using the MaxEnt3 program (Waters). Detailed procedures for cloning are described in the Supporting Information. Primer sequences are included in Supporting Information Table 2.

**Genome Mining of ProcM-like Enzymes and Their Associated LanA Substrates.** To identify ProcM-like enzymes, BlastP searches were performed using the ProcM protein sequence as the query. Hits were selected with identity >30% and gap <8%. To identify putative *lanAs*, the annotated open reading frames (ORFs) around *lanMs* were inspected manually, and BlastP searches were performed

within this genome using either a ProcA leader peptide sequence, the leader sequence from other LanAs identified in the same genome, or the leader sequences from known LanAs as the queries. Detailed genome mining procedures are described in the Supporting Information. LanA sequences are shown with their corresponding LanM enzymes in Supporting Information Table 1.

**Phylogenetic and Network analysis.** Bayesian MCMC inference analyses were performed using the program MrBayes (version 3.2).<sup>39</sup> Final analyses consisted of two sets of eight chains each (one cold and seven heated), run for about 2 million generations with trees saved and parameters sampled every 100 generations. A mixed amino acid model was utilized, and analyses were run to reach a convergence with standard deviation of split frequencies <0.01. Posterior probabilities were averaged over the final 75% of trees (25% burn in). Network analysis was performed by BLAST searches comparing each CyanA leader peptide against another. A Matlab script was written to remove all duplicate comparisons, and the result was imported into the Cytoscape software package.<sup>40</sup> The nodes were arranged using the yFiles organic layout provided with Cytoscape version 2.8.3.

**General Procedure for LanA *in Vivo* Modification.** Electrocompetent *E. coli* BL21 (DE3) cells were transformed with the coexpression constructs. Cultures were inoculated from single colony transformants and grown overnight at 37 °C in 20 mL of LB broth supplemented with 50 mg L<sup>-1</sup> kanamycin. The overnight culture was used to inoculate 1 L of LB broth, and cells were grown at 37 °C to OD<sub>600</sub> ≈ 0.6–0.8. Expression was induced by the addition of 0.2 mM IPTG, and the culture was incubated at 18 °C for 18 h. After harvesting, the pellet was resuspended in 35 mL of denaturing LanA buffer 1 (6 M guanidine hydrochloride, 20 mM NaH<sub>2</sub>PO<sub>4</sub>, 500 mM NaCl, 0.5 mM imidazole, pH 7.5). The cell paste was subjected to sonication, and the cell debris was removed by centrifugation. The supernatant was purified by immobilized metal affinity chromatography (IMAC) using His60 Ni Superflow Resin (Clontech), and the elution fractions were desalted and purified by reversed-phase HPLC using a Waters Delta-pak C4 column. The fractions were either analyzed directly by MALDI-ToF-MS, or treated with appropriate endoprotease (Glu-C, Asp-N, or TEV protease) to remove the leader peptide before MALDI-ToF-MS and/or ESI-MS/MS analysis.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

Descriptions of all molecular biology procedures, peptide purifications, analytical procedures, proteolysis assays, and supporting figures and tables. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [vddonk@illinois.edu](mailto:vddonk@illinois.edu).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health (R01 GM 58822) to W.A.v.d.D. The Bruker UltrafleXtreme MALDI TOF/TOF mass spectrometer was purchased in part with a grant from the National Institutes of Health (S10 RR027109). We thank J. Zhu (UIUC) for assistance with the proteolytic degradation of some of the CyanA peptides and A. Cubillos-Ruiz (Chisholm laboratory, MIT) for helpful discussions about the CyanA ORFs.

## ■ REFERENCES

- (1) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D.; Fischbach, M. A.; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süßmuth, R. E.; Tagg, J. R.; Tang, G.-L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; and van der Donk, W. A. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* 30, 108–160.
- (2) Willey, J. M., and van der Donk, W. A. (2007) Lantibiotics: peptides of diverse structure and function. *Annu. Rev. Microbiol.* 61, 477–501.
- (3) Knerr, P. J., and van der Donk, W. A. (2012) Discovery, biosynthesis, and engineering of lantipeptides. *Annu. Rev. Biochem.* 81, 479–505.
- (4) Bierbaum, G., and Sahl, H. G. (2009) Lantibiotics: mode of action, biosynthesis and bioengineering. *Curr. Pharm. Biotechnol.* 10, 2–18.
- (5) Piper, C., Cotter, P. D., Ross, R. P., and Hill, C. (2009) Discovery of medically significant lantibiotics. *Curr. Drug Discovery Technol.* 6, 1–18.
- (6) Chatterjee, C., Paul, M., Xie, L., and van der Donk, W. A. (2005) Biosynthesis and mode of action of lantibiotics. *Chem. Rev.* 105, 633–684.
- (7) Ross, A. C., and Vederas, J. C. (2011) Fundamental functionality: recent developments in understanding the structure–activity relationships of lantibiotic peptides. *J. Antibiot.* 64, 27–34.
- (8) Schnell, N., Entian, K.-D., Schneider, U., Götz, F., Zahner, H., Kellner, R., and Jung, G. (1988) Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. *Nature* 333, 276–278.
- (9) Montalban-Lopez, M., Zhou, L., Buivydas, A., van Heel, A. J., and Kuipers, O. P. (2012) Increasing the success rate of lantibiotic drug discovery by synthetic biology. *Expert Opin. Drug Discovery* 7, 695–709.
- (10) Zhang, Q., Yu, Y., Velásquez, J. E., and van der Donk, W. A. (2012) Evolution of lantipeptide synthetases. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18361–18366.
- (11) Oman, T. J., and van der Donk, W. A. (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat. Chem. Biol.* 6, 9–18.
- (12) Li, B., Sher, D., Kelly, L., Shi, Y., Huang, K., Knerr, P. J., Joewono, I., Rusch, D., Chisholm, S. W., and van der Donk, W. A. (2010) Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10430–10435.
- (13) Tang, W., and van der Donk, W. A. (2012) Structural characterization of four prochlorosins: a novel class of lantipeptides produced by planktonic marine cyanobacteria. *Biochemistry* 51, 4271–4279.
- (14) Haft, D. H., Basu, M. K., and Mitchell, D. A. (2010) Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. *BMC Biol.* 8, 70.
- (15) Jacobson, M. R., Brigle, K. E., Bennett, L. T., Setterquist, R. A., Wilson, M. S., Cash, V. L., Beynon, J., Newton, W. E., and Dean, D. R. (1989) Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. *J. Bacteriol.* 171, 1017–1027.
- (16) Paul, M., Patton, G. C., and van der Donk, W. A. (2007) Mutants of the zinc ligands of lacticin 481 synthetase retain dehydration activity but have impaired cyclization activity. *Biochemistry* 46, 6268–6276.
- (17) Li, B., Yu, J. P., Brunzelle, J. S., Moll, G. N., van der Donk, W. A., and Nair, S. K. (2006) Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. *Science* 311, 1464–1467.



- (18) Penner-Hahn, J. (2007) Zinc-promoted alkyl transfer: a new role for zinc. *Curr. Opin. Chem. Biol.* 11, 166–171.
- (19) Wang, H., Fewer, D. P., and Sivonen, K. (2011) Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. *PLoS One* 6, e22384.
- (20) Hourai, S., Miki, M., Takashima, Y., Mitsuda, S., and Yanagi, K. (2003) Crystal structure of nitrile hydratase from a thermophilic *Bacillus smithii*. *Biochem. Biophys. Res. Commun.* 312, 340–345.
- (21) Haft, D. H., Loftus, B. J., Richardson, D. L., Yang, F., Eisen, J. A., Paulsen, I. T., and White, O. (2001) TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res.* 29, 41–43.
- (22) Piard, J. C., Kuipers, O. P., Rollema, H. S., Desmazeaud, M. J., and de Vos, W. M. (1993) Structure, organization, and expression of the *lct* gene for lactacin 481, a novel lantibiotic produced by *Lactococcus lactis*. *J. Biol. Chem.* 268, 16361–16368.
- (23) Xie, L., Miller, L. M., Chatterjee, C., Averin, O., Kelleher, N. L., and van der Donk, W. A. (2004) Lactacin 481: *in vitro* reconstitution of lantibiotic synthetase activity. *Science* 303, 679–681.
- (24) Levengood, M. R., Patton, G. C., and van der Donk, W. A. (2007) The leader peptide is not required for post-translational modification by lactacin 481 synthetase. *J. Am. Chem. Soc.* 129, 10314–10315.
- (25) Cheung, W. L., Pan, S. J., and Link, A. J. (2010) Much of the microcin J25 leader peptide is dispensable. *J. Am. Chem. Soc.* 132, 2514–2515.
- (26) Sardar, D., Pierce, E., McIntosh, J. A., and Schmidt, E. W. (2014) Recognition sequences and substrate evolution in cyanobactin biosynthesis. *ACS Synth. Biol.*, DOI: 10.1021/sb500019b.
- (27) Goto, Y., Ito, Y., Kato, Y., Tsunoda, S., and Suga, H. (2014) One-pot synthesis of azoline-containing peptides in a cell-free translation system integrated with a posttranslational cyclodehydratase. *Chem. Biol.* 21, 766–774.
- (28) Koehnke, J., Bent, A. F., Zollman, D., Smith, K., Houssen, W. E., Zhu, X., Mann, G., Lebl, T., Scharff, R., Shirran, S., Botting, C. H., Jaspars, M., Schwarz-Linek, U., and Naismith, J. H. (2013) The cyanobactin heterocyclase enzyme: a processive adenylase that operates with a defined order of reaction. *Angew. Chem., Int. Ed.* 52, 13991–13996.
- (29) Houssen, W. E., Koehnke, J., Zollman, D., Vendome, J., Raab, A., Smith, M. C., Naismith, J. H., and Jaspars, M. (2012) The discovery of new cyanobactins from *Cyanothece* PCC 7425 defines a new signature for processing of patellamides. *ChemBioChem* 13, 2683–2689.
- (30) Melby, J. O., Nard, N. J., and Mitchell, D. A. (2011) Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. *Curr. Opin. Chem. Biol.* 15, 369–378.
- (31) Li, B., and van der Donk, W. A. (2007) Identification of essential catalytic residues of the cyclase NisC involved in the biosynthesis of nisin. *J. Biol. Chem.* 282, 21169–21175.
- (32) Suda, S., Lawton, E. M., Wistuba, D., Cotter, P. D., Hill, C., and Ross, R. P. (2012) Homologues and bioengineered derivatives of LtnJ vary in ability to form D-alanine in the lantibiotic lactacin 3147. *J. Bacteriol.* 194, 708–714.
- (33) Cotter, P. D., O'Connor, P. M., Draper, L. A., Lawton, E. M., Deegan, L. H., Hill, C., and Ross, R. P. (2005) Posttranslational conversion of L-serines to D-alanines is vital for optimal production and activity of the lantibiotic lactacin 3147. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18584–18589.
- (34) Lohans, C. T., Li, J. L., and Vederas, J. C. (2014) Structure and biosynthesis of carnolysin, a homologue of enterococcal cytolysin with D-amino acids. *J. Am. Chem. Soc.* 136, 13150–13153.
- (35) Firn, R. D., and Jones, C. G. (2003) Natural products—a simple model to explain chemical diversity. *Nat. Prod. Rep.* 20, 382–391.
- (36) Jenke-Kodama, H., Müller, R., and Dittmann, E. (2008) Evolutionary mechanisms underlying secondary metabolite diversity. *Prog. Drug Res.* 65, 121–140.
- (37) Yu, Y., Zhang, Q., and van der Donk, W. A. (2013) Insights into the evolution of lanthipeptide biosynthesis. *Protein Sci.* 22, 1478–1489.
- (38) Yang, X., and van der Donk, W. A. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: new insights into the role of leader and core peptides during biosynthesis. *Chem.—Eur. J.* 19, 7662–7677.
- (39) Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542.
- (40) Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A. R., Vailaya, A., Wang, P. L., Adler, A., Conklin, B. R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G. J., Ideker, T., and Bader, G. D. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* 2, 2366–2382.