Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria

Bo Li1,2, Daniel Sheh1, Libusha Kellyb, Yanxiang Shi1, Katherine Huangb, Patrick J. Knerrc, Ike Joevono1, Doug Rusch1, Sallie W. Chisholmb,2, and Wilfred A. van der Donk1,4,a

1Department of Biochemistry, 2Department of Chemistry, and 4the Howard Hughes Medical Institute, University of Illinois at Urbana-Champaign, 600 S. Mathews Avenue, Urbana, IL 61801; 1Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; and 5J. Craig Venter Institute, Rockville, MD 20850

Edited by Craig A. Townsend, Johns Hopkins University, Baltimore, MD, and accepted by the Editorial Board April 17, 2010 (received for review December 7, 2009)

Our understanding of secondary metabolite production in bacteria has been shaped primarily by studies of attached varieties such as symbionts, pathogens, and soil bacteria. Here we show that a strain of the single-celled, planktonic marine cyanobacterium Prochlorococcus—which conducts a sizable fraction of photosynthesis in the oceans—produces many cyclic, lanthionine-containing peptides (lantipeptides). Remarkably, in Prochlorococcus MIT9313 a single promiscuous enzyme transforms up to 29 different linear ribosomally synthesized peptides into a library of polycyclic, conformationally constrained products with highly diverse ring topologies. Genes encoding this system are found in variable abundances across the oceans—with a hot spot in a Galapagos hypersaline lagoon—suggesting they play a habitat- and/or community-specific role. The extraordinarily efficient pathway for generating structural diversity enables these cyanobacteria to produce as many secondary metabolites as model antibiotic-producing bacteria, but with much smaller genomes.

S
secondary metabolites are among the most functionally and structurally diverse molecules in nature and play critical roles in a wide variety of processes such as metal transport, cell-cell communication, and chemical defense (1). Two major pathways exist for the production of peptide secondary metabolites—ribosomal and nonribosomal. In the former case, the resulting ribosomally produced peptide is often modified by cognate enzymes to produce the mature metabolite. Recent genome-enabled studies have shown that many classes of peptide natural products initially suspected to be of nonribosomal origin are in fact gene-encoded (e.g., refs. 2–6) and that structural motifs previously thought to be found only in nonribosomal secondary metabolites are also found in ribosomally synthesized compounds (7). Clearly, synthesis of secondary metabolites via the ribosome turns out to be much more widespread than originally anticipated.

Lanthionine-containing peptides are a family of ribosomally synthesized microbial secondary metabolites characterized by intramolecular thioether cross-links formed between a dehydrated Ser or Thr residue (dehydroalanine or dehydrobutyryl, respectively) and a Cys residue (Fig. L4). Most of the known lanthionine-containing peptides have antimicrobial activity and are referred to as lantibiotics (8). Family members with other, often unknown, functions are termed lantipeptides (9). Their precursors are encoded by short genes, generically termed lanAs. The precursor LanA peptides translated from these genes contain an N-terminal leader peptide that is removed in the final step of biosynthesis (10); a C-terminal core peptide is transformed into the mature natural product through a dehydration and cyclization process (Fig. L4), which is catalyzed for class II lantibiotics by a bifunctional synthetase generically termed LanM (11).

In most lantibiotic-producing bacteria, the LanM lanthionine synthetase modifies only a single LanA precursor peptide. However, LanM enzymes reveal low substrate specificity under laboratory conditions, a property that has been exploited for the generation of variants of the natural products (12–14). Wondering whether organisms in nature have taken advantage of this substrate promiscuity to generate natural product libraries with a single streamlined biosynthetic toolset, we conducted a search of the sequenced bacterial genomes for cases where a single bacterium can produce multiple lanthionine-containing peptides using the same LanM enzyme. We discovered that the genomes of several strains of marine Prochlorococcus and Synechococcus contain multiple lanA-like genes but only a single lanM-like gene. This finding is of interest not only because these two cyanobacteria account for as much as half of the chlorophyll—i.e., photosynthetic capacity—in the tropical and subtropical oligotrophic oceans (15), but also because their “lifestyle” is very different from that of microbes known to produce these types of compounds; they are single-celled and free-floating and live in a very dilute habitat where the function of secondary metabolites is not readily apparent. Using a cultured strain of Prochlorococcus and available metagenomic databases from the oceans, we asked whether Prochlorococcus can indeed utilize this system to produce a diverse array of lantipeptides, whether there is evidence for this capability in ocean metagenomic databases, and whether it is found more commonly in specific oceanic regions or habitats than in others.

Results

Putative Lantibiotic Gene Clusters in Cyanobacteria. By using the mersacidin synthetase gene mrsM from Bacillus sp. strain HIL Y-85 as a query, a single lanM gene was found in the genomes of two closely related strains of Prochlorococcus (MIT9313 and MIT9303) and in one distantly related (16) strain of marine Synechococcus (RSW916) (noted also by ref. 17). The genomes of these strains contained 29 (Fig. 1B), 15, and 10 (Fig. S1) putative lanA genes, respectively. We focused our work on Prochlorococcus MIT9313, because this strain contains the largest...
number of putative \textit{lanA} genes. Its genome contains one gene encoding a \textit{lanM} homolog, designated \textit{procM}, which is located in a cluster with seven genes that encode putative substrate peptides. Twenty-two additional \textit{procA} genes are found elsewhere in the genome, 20 of which are clustered together in three other regions on the genome (Fig. 1B). We named the \textit{lanA}-like genes \textit{procA}s, followed by the cluster number on the genome and the order of the specific \textit{procA} gene within this cluster (e.g., \textit{procA1.1-procA1.7}; see Fig. 1B). The products of the \textit{procA} genes display very high sequence identity in their putative leader peptides but strikingly high diversity in their putative core peptides, the part of the molecule that will form the mature product (Fig. 1C). The leader and core peptides are separated by a GG or GA cleavage motif, used by ABC transporters with an N-terminal cysteine protease domain to remove leader peptides during class II lantibiotic secretion (18). The genome of MIT9313 encodes one homolog of these transporters (Fig. S1). Intriguingly, not a single pair of core peptides shows detectable sequence identity, and their length varies greatly (from 13 to 32 amino acids). Furthermore, collectively the core peptides contain serines and threonines, the amino acids that are dehydrated in lantibiotic substrates, at every position from the 1st residue to the 24th residue following the leader peptide (Fig. 1C). Similarly, cysteine residues are found in every position from residue 1 of the core peptide to residue 23. For a single ProcM protein to process all 29 peptides, it would have to be extraordinarily promiscuous.
In Vitro Dehydration and Cyclization Activity. To test whether one ProcM enzyme could possibly process these 29 highly diverse core peptides, seven procA genes located in the vicinity of procM, ten procA genes located distally on the genome, and procM itself were cloned and heterologously expressed in Escherichia coli as N-terminally Hitagged fusion proteins. Remarkably, incubation of each of the 17 purified peptides with ProcM in the presence of ATP and Mg\(^{2+}\) resulted in efficient dehydration (Fig. 2 and Figs. S2 and S3) as determined by MALDI-TOF MS after proteolytic removal of most of the leader peptide (for detailed experimental procedures, see the SI Appendix). For many substrates, the number of dehydrations equaled the number of Ser/Thr present in the core peptide (Figs. 1C and 2). For some substrates such as ProcA1.2, one of the Ser/Thr in the core peptide escapes dehydration, which is not uncommon in lanthionine-containing peptides (18). Analysis by tandem electrospray ionization MS (ESI-MSMS) showed regions in the ProcA products that are resistant to fragmentation (Fig. 3A), suggestive of thioether cross-links (11). To provide further evidence that these rings consist of lanthionine linkages, larger scale generation of the product of ProcA2.8 was carried out, most of the leader peptide was removed by using protease GluC, and the C-terminal proteolytic fragment corresponding to the core peptide was purified by HPLC. Complete hydrolysis of the peptide and derivatization of the resulting amino acids as described previously (19), followed by GC-MS analysis and comparison with a meso-lanthionine standard confirmed the lanthionine linkages (Figs. S4–S6). It is likely, given the high sequence conservation in the leader peptides, that all 29 ProcA peptides are substrates for ProcM. We propose the topology of thioether cross-links. After removal of most of the leader peptide by using the commercial proteases LysC, trypsin, or GluC, the modified ProcAs were analyzed by ESI-MSMS. For the products of ProcA1.1, ProcA2.8, and ProcA4.3 modified by ProcM (Fig. 3A and Figs. S7 and S8), the observed fragmentation pattern demonstrates ring systems that do not overlap. In contrast, the constellation of Cys and Ser/Thr residues in ProcA1.7 requires overlapping rings, and indeed no fragmentation was observed between residues 8 and 22 (Fig. 3A). To determine the connectivity of the thioether rings, a series of mutant peptides of ProcA1.7 was generated to disrupt each individual ring by mutation of a Ser or Thr to Ala (Fig. S9). ESI-MSMS analysis of these ProcA1.7 mutants after modification by ProcM support the ring structure shown in Fig. 3B. Overlapping rings were also found in prochlorosin 1.5, 2.11 and 3.3 (Fig. 3B and Figs. S10–S12). We cannot rule out the presence of minor products with alternative ring topologies for which the intensities of the fragment ions would fall below our detection limit.

Production of Prochlorosins in Vivo. We next investigated whether Prochlorococcus MIT9313 makes prochlorosins under laboratory conditions, by using several approaches. procM and several procA genes are transcribed by exponentially growing MIT9313 cultures (Fig. 4A). Analysis of previously published whole genome expression studies shows that their transcription is down-regulated under nitrogen starvation (Fig. S13) but is not responsive to phosphate (20) or iron (21) starvation. These findings show that these genes are integrated into the MIT9313 transcriptional circuitry, can respond to changes in environmental conditions, and hence are likely functional. Furthermore, several prochlorosins were detected in spent media of late-exponential stage MIT9313 cells, showing that they are produced. ESI-MSMS
analysis of in vivo produced prochlorosin 2.1 (MIT9313) showed two key fragment ions (y’17 and y’18; for nomenclature, see Scheme S1) that agree with the fragmentation observed for the compound produced in vitro (Fig. 4B). Similar fragmentation patterns for the in vivo and in vitro produced compounds were also observed for prochlorosin 2.11 and 3.2 (MIT9313) (Fig. S14). Thus, these three prochlorosins are produced in vivo, the leader cleavage site is located at the anticipated motif, and the in vitro prepared compounds have the same ring pattern as the natural products.

**Distribution of Prochlorosin Gene Clusters in the Global Ocean Metagenomic Survey.** The recent global ocean metagenomic survey (GOS), in which Prochlorococcus and Synechococcus genes are relatively abundant (22), offers an opportunity to assess the prevalence of lanthionine biosynthetic genes in these and other microbes and examine their geographic distributions. To this end, the GOS database was searched by using tBLASTn, with the MIT9313 procM sequence as a query. We also analyzed the occurrence of seven single-copy, reference genes from Prochlorococcus and Synechococcus at each station (Table S1), as a metric for the number of genome equivalents from these groups captured in the sample (23). We found 21 distinct procM-like sequences predicted to originate from either Prochlorococcus or Synechococcus (the resolution of the phylogenetic tree does not enable us to determine from which of these two genera the sequences originate; Fig. S15). We also found eight additional procM-like sequences, some of which are similar to those found in the genomes of proteobacteria, firmicutes, and actinobacteria (17). By using the conserved leader peptides of the procA genes from MIT9313 as a query, we found 152 distinct procA-like sequences in GOS (Table S2). As many as three different procA genes were found in tandem on the same GOS read. Together with the high ratio of procA to procM genes detected, this suggests that, among the Prochlorococcus and Synechococcus cells sampled by GOS, the genetic capability for promiscuous biosynthesis of lanthipeptides is the norm rather than an exception.

Fig. 3. In vitro cyclization of ProcAs by ProcM. (A) ESI-MS/MS analysis of ProcA1.1 G-1E (Upper) and ProcA1.7 (Lower) treated successively with ProcM and GluC. The labeled fragment ions support two nonoverlapping thioether rings for prochlorosin 1.1 (MIT9313) and show that, in prochlorosin 1.7 (MIT9313), Dhb1 and Dhb5 are not involved in rings. (B) Proposed structures of Pcn1.1, 1.7, 2.8, 3.3, 4.3, and 2.11 (MIT9313); see SI Appendix. Arrows illustrate the start of the putative core peptide. Asterisks indicate prochlorosins containing residues from the leader peptide (underlined).

Fig. 4. In vivo production of prochlorosins. (A) Transcription of procM and procA genes by a log-phase culture of MIT9313. RT-reverse-transcriptase added (+) or omitted (−). rnpB is a control housekeeping gene. (B) Production of prochlorosin 2.1 (MIT9313) in vivo. The upper panel depicts ESI-MSMS analysis of ProcA2.1 modified in vitro by ProcM and digested with GluC; the lower panel shows ESI-MSMS analysis for prochlorosin 2.1 (MIT9313) from the spent media of a late-exponential stage culture of MIT9313. Fragment ions are indicated. The b2 and y’28-31 ions in the upper panel are derived from fragmentation in a stretch of residues (underlined) remaining from the leader peptide after GluC cleavage. Therefore, they are not observed in the product isolated from spent medium. Site directed mutagenesis studies suggest that Ser7 of Pcn2.1 (MIT9313) escapes dehydration (Fig. S18).
The isoprenoid cyclases can produce multiple products from a single substrate to generate a group of diverse compounds (25, 26). Alternatively, a spectrum of structurally related secondary products produced by different members of a microbial consortium can contribute to the defensive capability of the consortium as a whole [e.g., cyanobactins (27)]. The lantipeptide producing machinery described here for prochlorosin biosynthesis by Prochlorococcus MIT9313 is unique in that it shows that a single organism can produce as many as 29 different secondary metabolites from distinct gene-derived precursors by using one promiscuous biosynthetic enzyme.

Remarkably, Prochlorococcus MIT9313 has the genetic capacity to produce as many secondary metabolites as the model antibiotic-producing actinomycetes Streptomyces coelicolor and Streptomyces avermitilis (28, 29), but with a genome less than one-third in size (2.4 MB compared with 8.6–9.0 MB) (30). Rather than maintaining a large (~20–60 KB) gene cluster for the synthesis of each secondary metabolite, Prochlorococcus and Synechococcus maintain a single biosynthetic enzyme gene and a series of short, ~300 bp, precursor genes to generate a topologically diverse set of conformationally constrained peptides in a remarkably effective two-step reaction sequence. When taking into account all of the cyanobacterial genomes analyzed here and in the GOS dataset, we have detected 150 unique putative lantipeptide sequences illustrating the remarkable diversity of products that can be generated by this pathway (see the SI Appendix). For instance, the ring patterns determined for all six lantipeptides analyzed in this study (Fig. 3) are different from the ~20 previously documented lantibiotic/lantipeptide ring topologies, which originate from a multitude of different organisms (Fig. S16).

How can the structural diversity manifested in prochlorosins be introduced into a series of linear peptides by a single enzyme? The number and position of Ser/Thr and Cys residues varies greatly in the 17 investigated core peptides, and yet all of these peptides are dehydrated and cyclized. As depicted in Fig. 3B, the rings have very different sizes and are formed from Cys residues located both C-c and N-terminal to the dehydro amino acid partners with which they react. It is difficult to envision an active site geometry of ProCm that would actively catalyze each of these cyclizations, and we hypothesize that only a subset of rings are generated by the enzyme, reorganizing the resulting product for further regioselective, nonenzymatic cyclization; such nonenzymatic cyclization has been demonstrated with synthetic peptides (31). Notably, the modification process is likely guided by the conserved leader peptide, because ProCm did not modify truncated ProCA substrates lacking the leader peptides (Fig. S17). The promiscuity demonstrated here may be raveled by the relaxed substrate specificity of P450 enzymes that decorate terpenoid products in plants (32). However, for those systems, the molecular scaffolds have already been put in place by terpenoid cyclases, whereas ProCm generates the topologically diverse ring structures itself.

The promiscuous nature of the posttranslational modification machinery is poised to allow rapid evolution of diversity, because any nonsynonymous mutation in the gene encoding a peptide-derived product potentially results in a new cyclic product. Such evolvability stands in contrast to systems where the diversity is generated by an enzyme, in which only a limited subset of nonsynonymous mutations (those that affect the enzymatic activity without perturbing the enzyme structure) result in the production of a new compound. Combined with the ease at which short precursor peptides are thought to duplicate within a genome (3, 33), rapid diversification is facilitated as has been shown for the conopeptide family (33), and this may have been the mechanism which resulted in 29 different procA genes being present in MIT9313. It is noteworthy that the genomic regions in MIT9313 that contain the procA genes also contain phage integrases, transposases, and fragments of the latter (Fig. S1); genomic cluster 3

**Fig. 5.** Prochlorococcus and Synechococcus lantipeptide genes in the Global Ocean Survey database. (A) Lantipeptide biosynthetic clusters in assembled metagenomic reads from the hypersaline lagoon site, GOS33. Blue—procA, red—procM, green—ABC transporters (red **—peptidase domain), light green—toIC, yellow—response regulator. (B) Distribution of procM- and procA-like genes in the GOS sites. Upper: Average number of core gene copies (normalized for gene length to 1,000 aa and to the database size at each GOS site in Mb), representing the relative abundance of Prochlorococcus and Synechococcus cells at each site. Lower: Ratio of procM and procA gene copies to core genes at each site. Wedges denote the location of the sampling sites in the bar graph. The sites of isolation of the Prochlorococcus MIT9313 and MIT9303 and Synechococcus RS9916 strains that contain lantipeptide genes are shown as yellow triangles.

**Discussion**

Several different forms of natural combinatorial biosynthesis have been recognized as contributing to the diversity in secondary metabolites. For instance, nonribosomal peptide synthetases and polyketide synthases create secondary metabolites from multienzyme “assembly lines” composed of a series of functional modules acting sequentially, with new domain combinations leading to the generation of unique products (24). In addition, enzymes such as the isopenoid cyclases can produce multiple products from a
also contains a putative siphovirus integration site ([34]—the peptides annotated as nif1-like are procA3.1-procA3.5). Together, these observations suggest a mechanism for transfer of these genes within and between genomes. At present, we can only speculate as to the biological function of prochlorosins. Whereas the vast majority of known lanthionine-containing peptides are bacteriocidal (12), lanthionine-containing peptides can also act as signaling molecules (35) or morphogenetic peptides (36). We have not observed any bacteriocidal activity in tests of four recombinant prochlorosins (Proc1.1, Proc1.2, Proc1.3, and Proc1.5) against Lactococcus lactis 117 and Bacillus subtilis 6633 (SI Appendix). However, these strains and assay conditions do not even remotely represent those encountered by Prochlorococcus cells in the wild. Further tests have been hampered by very low yields of prochlorosins from Prochlorococcus cultures (<10 μg from 20 L of cell culture) and by the need to remove the leader peptide from recombinantly expressed and modified peptides (SI Appendix)—obstacles that we will attempt to surmount in future studies.

Although procM/procA genes found in the GOS database were disproportionately abundant in a hypersaline lagoon near the Galapagos where microbial densities are relatively high and community diversity low (22), 42% of procM/procA genes detected were from nutrient-poor regions of the oceans, where total microbial cell densities are relatively low (∼10^9 bacteria mL^-1 seawater). In these habitats, the average distance between individual Prochlorococcus cells and their nearest microbial neighbor is approximately 200 Prochlorococcus cell lengths. It is difficult to imagine how prochlorosins could mediate intercellular functions at these distances. Perhaps they fulfill intracellular roles, and if so, one wonders what function would utilize such a diversity of secondary compounds.

**Materials and Methods**

General materials and protocols used for molecular biology, protein purification, MS, and bioinformatics are provided in the SI Appendix. procM was cloned into pET28b, and procA genes were cloned into pET-15b. Mutants of procA were generated by QuikChange (Strategene) or overlapping PCR. His6-ProCAs were overexpressed in insoluble form in E. coli and purified as previously described for other LanA peptides (37). Purification of His6-ProcM was achieved by cobalt-affinity chromatography resulting in 5–10 mg/L of cell culture. Activity assays were carried out in 50 mM Hepes (pH 7.5), 10 mM MgCl_2, 2.5 mM ATP, and 0.5 mM tris(2-carboxyethyl)phosphine with 25 μM ProcA and 0.5 μM ProcM. The assay mixture was incubated at 25 °C for 15–20 h and analyzed by MALDI-TOF MS (Voyager) or ESI-MS with a Synapt ESI quadrupole TOF System (Waters). For transcription assays of procM and procA, an axenic culture was grown to late-exponential stage, the cells harvested, and RNA isolated by using the Mirvana RNA isolation kit (Ambion). RNA was reverse-transcribed by using Superscript II (Invitrogen), DNA removed by using Ambion Turbo DNA-free, and PCR performed by using Platinum Taq DNA polymerase (Invitrogen). For detection of prochlorosins, a 20 L late-exponential axenic culture was harvested, the supernatant from the cell culture was filtered (Sopur 0.22 μm filter), and the cell-free spent media absorbed with 100 g of Amberlite XAD-16 resin (Sigma). The column was eluted with a stepwise gradient of aqueous isopropanol containing 0.1% trifluoroacetic acid. The elution fractions were concentrated and purified by using a C18 solid phase extraction column eluting with a stepwise gradient of aqueous acetonitrile containing 0.1% formic acid.

**ACKNOWLEDGMENTS.** This study was supported in part by grants from the Gordon and Betty Moore Foundation, the National Science Foundation, and the United States Department of Energy Genomics:GT program (S.W.C) and the National Institutes of Health (GM08222 to W.A.v.D.D.). D.S. is supported by postdoctoral fellowships from the Fulbright Foundation and the United States–Israel Binational Agricultural Research and Development Fund (Vaadia-BARD Postdoctoral Fellowship Award Fi-399-2007). I.J. was supported by the Howard Hughes Medical Institute.