

Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*

Eric W. Schmidt^{*†}, James T. Nelson^{*}, David A. Rasko[‡], Sebastian Sudek[§], Jonathan A. Eisen[‡], Margo G. Haygood[§], and Jacques Ravel[‡]

^{*}Department of Medicinal Chemistry, University of Utah, Salt Lake City, UT 84112; [‡]The Institute for Genomic Research, Rockville, MD 20850; and [§]Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093

Edited by Robert Haselkorn, University of Chicago, Chicago, IL, and approved April 7, 2005 (received for review February 18, 2005)

Prochloron spp. are obligate cyanobacterial symbionts of many didemnid family ascidians. It has been proposed that the cyclic peptides of the patellamide class found in didemnid extracts are synthesized by *Prochloron* spp., but studies in which host and symbiont cells are separated and chemically analyzed to identify the biosynthetic source have yielded inconclusive results. As part of the *Prochloron didemni* sequencing project, we identified patellamide biosynthetic genes and confirmed their function by heterologous expression of the whole pathway in *Escherichia coli*. The primary sequence of patellamides A and C is encoded on a single ORF that resembles a precursor peptide. We propose that this prepatellamide is heterocyclized to form thiazole and oxazoline rings, and the peptide is cleaved to yield the two cyclic patellamides, A and C. This work represents the full sequencing and functional expression of a marine natural-product pathway from an obligate symbiont. In addition, a related cluster was identified in *Trichodesmium erythraeum* IMS101, an important bloom-forming cyanobacterium.

heterocycle | tunicate | ascidian | genome sequencing | lantibiotic

Marine invertebrates, particularly sponges and ascidians, are well known for their production of bioactive natural products, several of which are currently undergoing clinical trials (1, 2). A major hurdle in the development of many of these agents into drugs has been their supply, because collection or aquaculture of marine invertebrates pose many difficulties and may not be environmentally acceptable. Because marine invertebrate compounds often resemble molecules isolated from bacteria, many compounds could be synthesized by symbiotic bacteria and not by the animals themselves (3–6). These early speculations were borne out recently in the cloning and sequencing of genes from two symbiotic natural-product pathways (7, 8), opening a new era in marine natural-products discovery and development. However, before this report, rationally discovered biosynthetic pathways from obligate symbionts had not been heterologously expressed, nor have complete pathways been reported.

Ascidians in the family Didemnidae contain numerous structural classes of cyclic peptides (5) and harbor symbiotic cyanobacteria, *Prochloron* spp. (Fig. 1) (9, 10). Despite nearly 30 years of attempts, *Prochloron* spp. have eluded cultivation and thus are considered to be obligate symbionts. *Prochloron* spp., unlike the vast majority of cyanobacteria but like plants, use both chlorophylls *a* and *b* for photosynthesis, lack phycobilins, and have plant-like thylakoids (11). The cells are relatively large for bacteria (10–20 μm in diameter). *Prochloron* has also been implicated in the biosynthesis of cyclic peptides isolated from whole didemnid ascidians. In early cell-separation studies, it was reported that the peptides were localized in *Prochloron* cells (12, 13), but a later investigation found the molecules distributed throughout the ascidian tunic as well as in the cyanobacteria (14). Because of the unique biological and chemical features of the *Prochloron*–ascidian symbiosis, we initiated

a project to sequence the genome of *Prochloron didemni*, isolated from the ascidian *Lissoclinum patella*.

The patellamides and trunkamide (another didemnid product) are peptides that exemplify both the unique structural features and potent bioactivities of didemnid ascidian natural products (Fig. 1). Both groups have clinical potential, because patellamides are typically moderately cytotoxic, and patellamides B, C, and D reportedly reverse multidrug resistance (15, 16), whereas trunkamide was initially isolated because of specific and unusual activity against the multidrug-resistant UO-31 renal cell line (17). Patellamides are characteristically composed of pseudosymmetrical, cyclic dimers, with each substructure having the sequence thiazole–nonpolar amino acid–oxazoline–nonpolar amino acid. Trunkamide and related molecules often contain proline, thiazolines, and prenylated serine and threonine derivatives. These features could easily result from either a ribosomal or nonribosomal peptide biosynthetic pathway, because precedents exist for heterocyclization and cyclization in both cases (18–21). We initially chose to investigate the nonribosomal hypothesis of patellamide biosynthesis using a homology-based approach (22). Only a single nonribosomal peptide synthetase (NRPS) gene was identified in fosmid clones, but the gene was found in only a few strains, and its presence did not correlate with patellamide production. Analysis of a low-coverage draft sequence of the *P. didemni* genome confirmed the dearth of NRPS genes and led us to search the genome for alternative synthetic routes. Here we report the identification of ribosomally encoded patellamide biosynthesis genes and the confirmation of their function by heterologous expression in *Escherichia coli*.

Materials and Methods

Purification of *Prochloron* DNA. *L. patella* was collected in the Republic of Palau in 2002. One ascidian colony (25 cm^2) was washed with sterile seawater and gently pressed to release *Prochloron*, which were purified by centrifugation. Within 15 min of harvesting, *Prochloron* DNA was obtained from freshly released cells by using the Genomic-Tip kit (Qiagen, Valencia, CA), following the bacterial DNA-purification protocol. *P. didemni* was enriched to >95% homogeneity, as evidenced by light microscopy and denaturing gradient-gel electrophoresis of the 16S rRNA gene. The DNA obtained was ≈ 40 kbp, as indicated by gel electrophoresis. Two strains were obtained in this collection, both from single

This paper was submitted directly (Track II) to the PNAS office.

Freely available online through the PNAS open access option.

Abbreviations: NRPS, nonribosomal peptide synthetase; gDNA, genomic DNA; SRM, selective reaction monitoring; CDS, coding sequence.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AY986476).

[†]To whom correspondence should be addressed at: Department of Medicinal Chemistry, University of Utah, 30 South 2000 East Room 201, Salt Lake City, UT 84112. E-mail: ews1@utah.edu.

© 2005 by The National Academy of Sciences of the USA

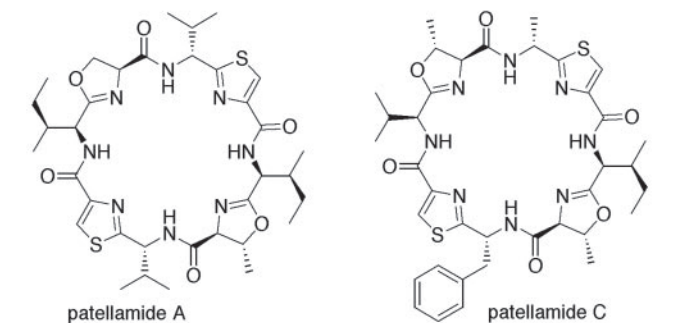
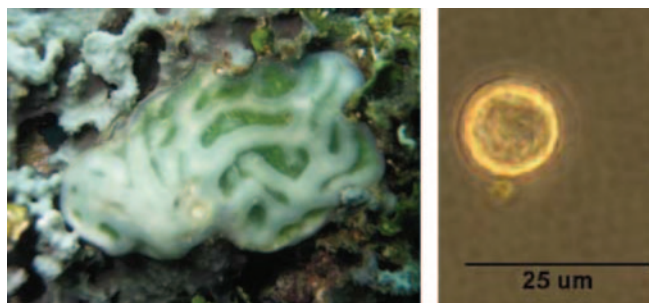


Fig. 1. Peptides and symbionts from *L. patella*. (Upper) Single cell of *P. didemni* (Right) isolated from the ascidian *L. patella* (Left) (photograph by Chris Ireland, University of Utah). The green pockets near the surface of *L. patella* are monocultures of *P. didemni*. (Lower) Patellamides A and C.

colonies of *L. patella*. One strain, designated “reef,” was collected from a fringe reef near Blue Corner, and the “Omodes” sample was collected in the Omodes sea-grass bed near Koror Island.

Genome Sequencing and Analysis. Two libraries were constructed from 12 μg of genomic DNA (gDNA) extracted from the reef sample with insert sizes of 2–5 and 6–8 kb. A total of 31,473 sequencing reads were obtained with an average length of 857 bp, corresponding to approximately three-times coverage of the predicted genome size (5 megabases). The genome was assembled by using the CELERA ASSEMBLER (23) into 734 scaffolds ranging in size from 1 to 77 kb. The scaffolds were concatenated randomly into a pseudomolecule, which was processed with the gene finder GLIMMER (24). Autoannotation was performed as described (25) by using a combination of hidden Markov models [TIGRFAM (26) and PFAM (27)] and BLAST analysis for predictive assignment of protein function.

The prepatellamide peptide was identified by using TBLASTN by querying each of the eight linear amino acid combinations of the cyclic peptide patellamide A in the genome scaffolds nucleotide sequence. Manual curation of the annotation of the cluster was performed by using the open-source MANATEE system (<http://manatee.sourceforge.net>).

Fosmid Library Construction. gDNA was prepared from the *Prochloron* gDNA reef sample by using the Genomic-Tip kit to yield DNA fragments of ≈ 40 kbp in size. This DNA was directly ligated to the pCC1FOS vector (Epicentre Technologies, Madison, WI) following manufacturer instructions. The titer from this method was very low, and the entire packaging extract was used to generate a 600-clone library. This library was representative of the sample and did not contain colonies with identical inserts, as judged by screening for a known *Prochloron* gene [*cao* (28)], and from fosmid end-sequencing.

Cloning of the *pat* Biosynthetic Cluster. The cluster was amplified from the *Prochloron* gDNA reef sample by PCR using primers

Lanti1f (5'-CGTGAAAATTGCTCTTTGAATAAAGG-3') and Lanti2r (5'-ACGGCAAAGGGAGTTTAAACGG-3') with PlatinumTaq HiFi (Invitrogen) and cloned into pCR2.1-TOPO (Invitrogen). Fosmids containing *pat* were identified from a pCC1FOS-arrayed library by using the previously reported methods with primers Lanti1f, Lanti2r, Lantimid1r (5'-CGCAGCTACGAGCAAAAACATTG-3'), and Lantimid1f (5'-CCACAGTTGAGGCCAGCAC-3'). The two sets of primers were also applied to whole gDNA samples extracted from the patellamide A/C-producing strain of *Prochloron* (reef) and from the nonproducing strain (Omodes).

Chemical Analysis. Patellamides A and C were isolated from whole *L. patella*, purified to homogeneity, and identified as reported (22). Briefly, the compounds were extracted with methanol and purified by partitioning followed by silica-gel chromatography. Fosmid clones containing *pat* were grown in 1 liter of LB-chloramphenicol (12.5 mg/liter) to an OD_{600} of ≈ 0.1 – 0.3 . Induction to a high copy number of the fosmid was performed according to manufacturer instructions and followed by an additional 24–48 h of incubation at 37°C. The pCR2.1-*pat* plasmid was transferred to *E. coli* BL21(DE3) pLys cells (Stratagene). Fresh colonies were grown in 1 liter of LB-ampicillin (50 mg/ml) for ≈ 24 h at 37°C with 1 mM isopropyl β -D-thiogalactoside.

To each 1-liter culture broth, ≈ 50 g of HP20SS (Supelco) resin was added, and the mixture was incubated for 1–2 h. The resin was filtered and rinsed with deionized water (2×100 ml) and 25% aqueous methanol (2×100 ml). The resin then was rinsed twice with methanol (100 ml) and acetone (100 ml), and the organic fractions were combined and dried by rotary evaporation. The extract was partitioned between chloroform and 25% aqueous methanol. The chloroform layer was evaporated to dryness and extracted further with equal volumes of hexanes and methanol. The methanol fractions were dried, resuspended in methanol (250 μl), and used directly for HPLC/electrospray ionization-MS analysis. In addition, a standard was prepared in which 25 ml of culture broth containing HP20SS resin was set aside before filtration. To this broth was added a mixture of patellamides A and C (10 μg each), and the resulting mixture was treated by filtration and partition as described above.

HPLC/electrospray ionization-MS analysis was performed on a ThermoFinnigan LCQ Classic ion-trap mass spectrometer. For HPLC, an analytical C18 column (Microsorb-MV, Varian) was used with a methanol–water gradient. An initial 50:50 mixture of methanol and water (each containing 0.1% formic acid) was subjected to a gradient to 95% methanol over 15 min, followed by 10 min at 95% methanol. Electrospray ionization MS was performed in the positive mode, and selective reaction monitoring (SRM) was applied to patellamide peaks at $m/z = 743$ and 763.

Results and Discussion

***Prochloron* Preparation and Purity.** *Prochloron* cells were prepared from whole *L. patella* and determined to be $>95\%$ pure as described previously (22). This 95% purity represents a conservative estimate, because we have yet to identify DNA from any other source in the *Prochloron* samples.

Chemical Analysis of *L. patella* Reef and Omodes Samples. We previously reported that the reef sample contained patellamides A

$\text{>PatE } \textit{MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASVTACITFCAYD}$
GVEPSITVCISVCAYDGE

Fig. 2. PatE sequence. Italic type, the conserved leader sequence; bold type, the proposed start and stop cyclization sequences; underlined type, product CDs. Sequences corresponding to patellamide C (Upper) and A (Lower) are aligned for clarity.

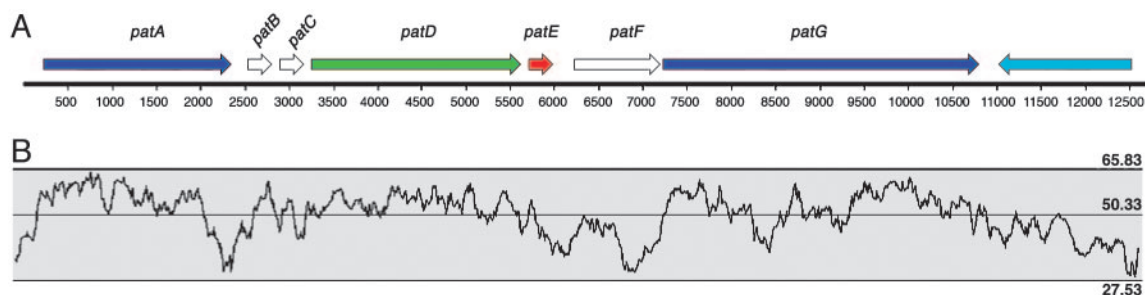


Fig. 3. The *pat* gene cluster (A) and GC skew (B). Colored genes represent those that can have a function assigned. White genes are those that have no significant homolog; blue genes contains protease activity. The G+C% skew below is altered where a coding region is present, as is common in many species and suggests that the gene predictions are correct (53). Additionally, the increase of the G+C% in this area suggests that this region may have been transferred into this species by means of horizontal gene transfer.

and C (29) in nearly equimolar amounts. Other patellamides were not detected as major products in the crude extract. The Omodes sample did not contain detectable patellamides, which was one of the criteria used to select reef for whole-genome sequencing.

Identification of *pat* Genes. Previously, an exhaustive PCR-based search for NRPS adenylation domains yielded only a single NRPS gene, *prnA* (GenBank accession no. AY590470). Detailed analysis of *prnA* revealed that it has the wrong domain architecture for patellamide biosynthesis. Furthermore, it has been found in some patellamide-producing strains but not in others (22). It was suggested that these results could indicate that *prnA* is not responsible for patellamide production; alternatively, *prnA*-like genes could be highly variable and thus were not detected in all peptide producers. The preliminary analysis of the draft genome sequence of *P. didemni* showed that *prnA* contained the only NRPS adenylation domain identified, bearing out the PCR data. Thus, a ribosomal synthesis of patellamides was a strong possibility. We performed a TBLASTN search of the draft genome sequence, querying for all eight possible peptides that could lead to the formation of the cyclic patellamide A. A single coding sequence (CDS) was identified, and strikingly this CDS also contained the required sequence for patellamide C. Because of the low probability that these sequences could co-occur by chance, this gene was identified as a candidate for the patellamide precursor peptide, *patE* (Fig. 2). No other oligopeptide 8-mers with an identical sequence to patellamide A or C could be identified in GenBank, and the entire *patE* precursor peptide was not closely related to any other known or predicted CDS. The presence of two peptide products on a single CDS suggests that synergy may be important to the patellamide mechanism of action (30).

Surrounding *patE*, there were several other CDSs with intriguing sequences, comprising the *patA*–*patG* genes in an \approx 11-kbp cluster (Fig. 3 and Table 1). In particular, a protease (*patA*), a possible adenylation enzyme–hydrolase hybrid (*patD*), and an oxidoreductase–protease hybrid (*patG*) immediately surround *patE*. Three other CDSs with very low or no similarity to other proteins of

known function (*patB*, *patC*, and *patF*) are also found in this cluster. On one side, this cluster ends with a gene that can be clearly assigned to primary metabolism (a DNA photolyase homolog), and on the other side a putative structural gene was identified extending \approx 1 kbp upstream of *patA*. These genes and the organization of the cluster are reminiscent of the lantibiotic and microcin biosynthetic machinery, which has been characterized in other bacteria (30–32). In particular, the microcin B17 peptide contains heterocycles (18, 33), whereas microcin J25 (21, 34) is cyclic. However, the combination of functionality (heterocyclization and cyclization) is thus far unique to the *pat* cluster.

Functional Expression of *pat* Genes. By using PCR, four fosmid clones containing *patE* were identified in a 576-clone arrayed library. From analysis of the fosmid end sequences, three of them (designated 21A, 28C, and 55F) were found to contain the complete pathway. Additionally, the region encompassing *patA*–*patG*, including putative regulatory regions, was amplified by PCR from whole reef gDNA and cloned into the pCR2.1-TOPO vector (Invitrogen). One-liter cultures from these fosmids and PCR clones were extracted and partially purified. Positive controls were established by adding patellamides A and C (0.4 mg/liter each) directly to *E. coli* culture broths containing vectors and then extracting these cultures in the same way that other samples were processed. We used an HPLC/electrospray ionization-MS approach to identify patellamides in our extracts.

Two standards were used to set up MS conditions. In the first standard, pure patellamides A and C, positively identified by NMR (^1H and ^{13}C) and mass spectrometry, were used for direct infusion and HPLC-MS experiments. In the second standard, a standard containing an initial 0.4 mg/liter of each patellamide was used for HPLC-MS. From both standards, molecular ions for patellamides A and C could readily be recognized in the mass spectrum (Fig. 4). Partially purified samples from fosmid and PCR clones were then injected. In all cases, blank or negative runs followed the injection of standards and did not contain the relevant ions. Ions of the appropriate mass could be identified at the correct elution time

Table 1. Predicted CDSs from the *pat* cluster

Protein	Amino acids	Proposed function	Sequence similarity	Identity/similarity	GenBank accession no.
PatA	702	Subtilisin-like protease	Hypothetical; <i>T. erythraeum</i>	57%/69%	ZP_00326030.1
PatB	78	Unknown	Hypothetical; <i>T. erythraeum</i>	52%/70%	ZP_00326032.1
PatC	64	Unknown	None	—	—
PatD	784	Adenylation/heterocyclization	Hypothetical; <i>T. erythraeum</i>	57%/70%	ZP_00326023.1
PatE	71	Patellamide precursor protein	None	—	—
PatF	320	Unknown	None	—	—
PatG	1187	Thiazoline oxidase/subtilisin-like protease	Two hypothetical proteins; <i>T. erythraeum</i>	43%/57% (N terminus) 48%/63% (C terminus)	ZP_00326026.1 (N terminus) ZP_00326033.1 (C terminus)

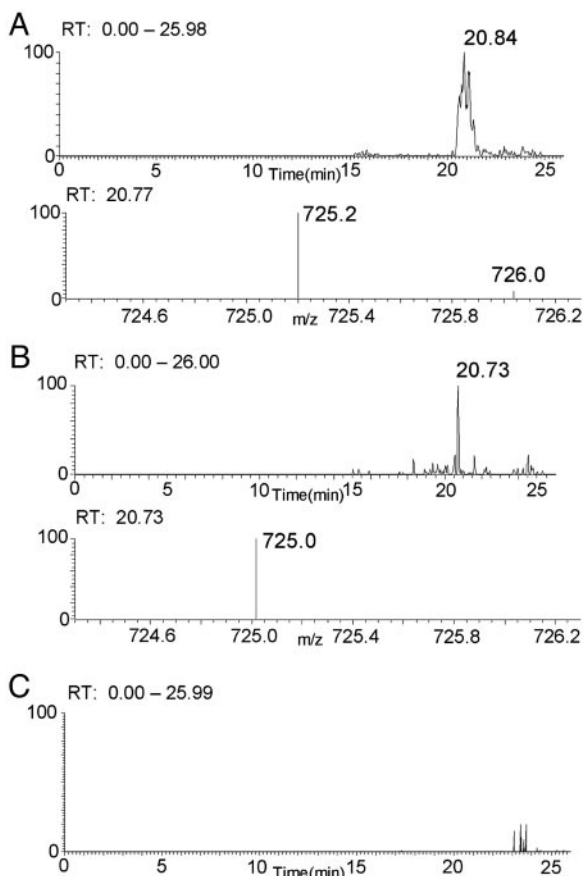


Fig. 4. Proof of function of the *pat* cluster. (A) Standard from 25 ml of culture broth containing 20 μ g of patellamides under SRM conditions observing $m/z = 725$ (patellamide A daughter ion). (B) Two-liter sample of pCR2.1-pat no. 9 under SRM conditions for $m/z = 725$. (C) Blind control: SRM using a sample identical to that shown in B except that empty pCR2.1 vector was used. y-axis scales are in units of relative abundance (0–100%).

from these samples, but the signal-to-noise ratio was not sufficient to conclusively prove the presence of patellamides. To confirm that these peaks resulted from patellamides, we used SRM, a commonly used technique in which sought ions are captured and fragmented by tandem MS. The mass spectrometer then scans only for a single daughter ion. This technique is extremely sensitive and less subject to error, because three pieces of data are obtained from a single experiment (elution time; presence of the parent ion; and fragmentation to a very specific daughter ion). By using this technique, patellamide A could be observed in the standard by monitoring for a major daughter ion at $m/z = 725$ (Fig. 4). In addition, patellamide C was seen in the standard by monitoring for the daughter ion at $m/z = 680$, although with much less sensitivity than for patellamide A. The patellamide A peak at $m/z = 725$ was observed in PCR clones and in fosmid extracts in a peak centered at 20.7 min (Fig. 4), indicating that patellamide A can be produced heterologously in *E. coli*. In particular, a 2-liter fermentation of a PCR clone led to a very clear identification of patellamide A, as shown in Fig. 4. It is difficult to determine exactly the amount of produced compound by this method, but our estimate is that at most 20 μ g/liter of patellamide A are produced under these conditions. Because of low sensitivity, we could not definitively confirm the presence of patellamide C in our extracts.

These data unambiguously confirm that the *patA–patG* gene cluster is responsible for patellamide biosynthesis in *P. didemni*. Because patellamide A is produced by clones containing the

\approx 11-kbp PCR product, we also have correctly identified the limits of the biosynthetic gene cluster.

Correlation of the Presence of the *pat* Gene Cluster with Patellamide Production. Although the *pat* pathway could be amplified by using DNA from the patellamide-producing reef strain, no products were amplified from the nonproducing *Omodes* strain. DNA quantity and quality from these two strains were identical, as judged by multiple PCR techniques, denaturing gradient-gel electrophoresis of 16S rDNA, UV spectroscopy, and quantitative gel electrophoresis (22). Thus, the patellamide cluster was found in a producing strain but was not present in a nonproducer. Because these two strains seem to be very similar by sequencing of several gene classes (chlorophyll *a* oxidase, 16S rDNA, and the *prn* NRPS operon), it is possible that *pat* and similar clusters in *Prochloron* originate through horizontal gene transfer, as has been proposed for other lantibiotic pathways (35). In fact, the *cao* and 16S rDNA genes are identical between *Prochloron* strains, whereas *prn* is >98% identical. Additional research is required to determine the origin and role of these pathways in *Prochloron*.

PatE: A Precursor Peptide Encoding Patellamides A and C. *patE* encodes a peptide of 71 aa, the first 37 of which are proposed to serve as a leader sequence for processing (Fig. 2). Of the remaining 34 aa, 16 directly encode the patellamide C and A sequences, whereas 18 make up motifs that we propose direct the cyclization of patellamides. The patellamide C peptide is located 8 aa upstream of the patellamide A sequence. Before both peptides, there is a 5-aa conserved region consisting of the consensus G(L/V)E(A/P)S. The sequence AYDGE terminates the patellamide A sequence and directly precedes the stop codon. Between the two patellamides, the 8-aa sequence AYDGVPEPS seems to encode for both a start and stop cyclization sequence, with the consensus stop sequence being AYDG(E/V). These sequences are of biotechnological interest, because they imply that diverse sequences could be synthesized to take advantage of these consensus regions, leading to the biosynthesis of a library of patellamides. It should be emphasized that the roles of these start/stop motifs are putative, and additional characterization is required. However, the microcin B17 prepeptide has been shown to be essential for proper posttranslational modification (36). Conserved residues in leader sequences are known to be important in the modification of some lantibiotics (30, 37, 38), and a consensus sequence (GAEPR) found in these prepeptides bears a striking resemblance to the PatE start consensus motif, G(L/V)E(A/P)S. Class I lantibiotics usually seem to possess a proline residue at the -2 position, although in the case of nisin this proline could be substituted with glycine and valine without impacting production (30, 37). Another general feature of class I lantibiotic leader peptides also found in PatE is a high proportion of charged residues.

Lantibiotics also contain C-terminal propeptide sequences that are cleaved by proteases, often in tandem with secretion from the cell. *pat* is, to the best of our knowledge, the first characterized lantibiotic- or microcin-like sequence from cyanobacteria.

The Patellamide Posttranslational Machinery. The *pat* cluster encompasses 7 genes, *patA–patG*, all of which are transcribed in the same direction and may comprise an operon. Sequence analysis of these genes allows the proposal of a biosynthetic pathway to patellamides (Fig. 5). PatA, PatD, and PatG (Table 1) are most similar to predicted proteins found clustered in the genome of *Trichodesmium erythraeum* IMS101 (GenBank accession no. AABK00000000). In addition, PatB also is most related to a *T. erythraeum* gene, although the *T. erythraeum* PatB homolog is not closely clustered with the PatA, PatD, and PatG homologs. The significance of this clustering in *T. erythraeum* will be discussed in the next section.

The PatA N-terminal region is similar to subtilisin-like proteases, which usually are involved in the recognition of signature sequences

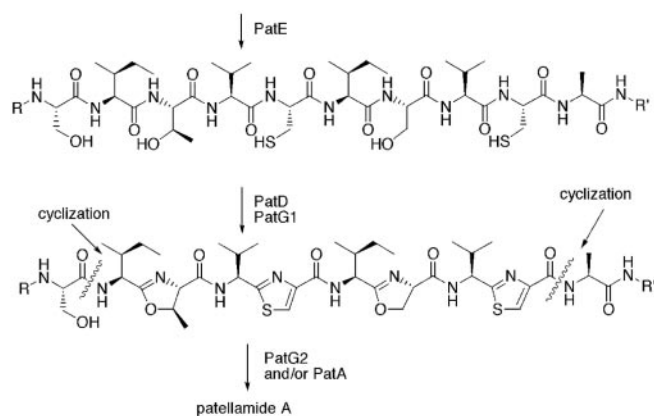


Fig. 5. Proposed pathway to patellamides showing the route to patellamide A. Timing of epimerization of alanine remains unclear; it may take place in tandem with thiazole formation (PatD2-catalyzed), at another biosynthetic step, or nonenzymatically (see *Results and Discussion*).

in hormone precursor peptides and the cleavage of these peptides near a signature motif (39, 40). The C-terminal region of the predicted protein shares no domain homology with proteins of known function, although it is related to hypothetical protein Orf4 from the cyanobacterium *Fremyella diplosiphon* (41). The predicted protein has a proline-rich region (amino acids 343–401), although the significance of this motif is unknown. Over its entire length, it is 59% identical to *T. erythraeum* subtilisin-like serine protease ZP_00326030.1. Because of the protease sequence homology, it is proposed that PatA is involved in cleavage of the PatE precursor peptide.

PatD, similar to PatA, seems to contain 2 domains. The N-terminal domain (PatD1) shares weak similarity with adenylating enzymes such as acyl-CoA ligases, and with MccB, the adenylating enzyme responsible for the biosynthesis of the microcins C51 and C7 (42). The PatD C terminus (PatD2) is similar to YcaO-like conserved domains of unknown function but also to SagD from *Streptococcus iniae*, which may serve as a hydrolase (43). PatD2 shows similarity to TtufA, a protein involved in the synthesis of the ribosomally derived trifolotoxin (44). The entire PatD peptide sequence is similar to only a handful of proteins, including the *T. erythraeum* homolog, a protein annotated as AknN (a hydrolase) from *Streptomyces galilaeus*, and Orf12, a predicted protein of unknown function from the granaticin biosynthetic pathway (45). Two possible roles are thus proposed for PatD. PatD2 may be involved in the cyclization of the cysteine and threonine residues of PatE, leading to thiazoline and oxazoline ring formation. PatD1 could activate cleaved patellamide precursors as adenylates, which then would cyclize to form the final patellamide structures. Alternatively, the ATP-binding region could have an as-yet-unknown function. For example, it is known that the microcin B17 heterocyclization complex includes an ATP-requiring enzyme, McbD, which is of unknown function (46). PatD1 does not show significant sequence homology to McbD, but it is often the case in microcin machinery that distantly related peptides serve similar functions.

PatG is a large, multidomain predicted protein. An N-terminal domain has homology to NAD(P)H oxidoreductases (PatG1). It is intriguing that the N-terminal region is distantly related to McbC from microcin B17 biosynthesis (18). McbC functions to oxidize thiazoline rings to the thiazole oxidation state, and it is likely that this is also the function of this region of PatG. This domain is also similar to an oxidase in the pathway to trifolotoxin, another thiazole-containing microcin. The C-terminal half of PatG (PatG2) is highly similar to PatA, containing subtilisin-like protease and *F. diplosiphon* Orf4-like regions. From this domain architecture, it seems that PatG is involved in the oxidation and maturation of PatE.

PatB, PatC, and PatF do not have obvious roles in patellamide biosynthesis. In addition, the protein responsible for epimerization is not evident from the sequence analysis, although it seems likely that epimerization could occur in tandem with heterocycle oxidation. The stereocenters adjacent to thiazole rings are highly labile and could also be subject to nonenzymatic epimerization. In some lantibiotic pathways, D-alanine residues are derived from serine (47), but in patellamides, they originate from L-alanine.

Related Pathways. The closest homologs of the pat cluster are CDSs of unknown function from the draft genome sequence of *T. erythraeum* IMS101. *patA*, *patD*, and *patG* are most similar to four clustered CDSs found in *T. erythraeum* (see Table 1). In fact, the *patG* homolog in *T. erythraeum* is split into two separate CDSs, comprising an oxidoreductase and a protease. In addition, a short peptide is present in this cluster that shares some structural features with *patE*. Furthermore, a transposase is found within the *T. erythraeum* gene cluster, possibly indicating that this cluster may move between strains by horizontal transfer. Several other CDSs of unknown function and not homologous to *pat* genes lie within the identified cluster. At the time of this writing, no other homologs to the *pat* pathway were found in GenBank.

In addition, streptolysin S has been known as an important mediator of pathogenesis in the “flesh-eating” *Streptococcus* spp. since its discovery 50 years ago (48), yet its structure has not been elucidated. The presence of a McbC-like oxidase and a PatD2-like hydrolase (SagD) in the streptolysin S biosynthetic gene cluster (43) indicate that streptolysin S likely contains thiazole rings. Indeed, the predicted streptolysin S prepropeptide contains numerous cysteine residues that could be cyclized.

Symbiosis and Secondary Metabolism. In the course of our studies, we found that some didemnid ascidians (including *L. patella*) contained bioactive secondary metabolites, whereas others did not contain these compounds. All the ascidians contained *Prochloron*, but not all of the *Prochloron* contain *pat*-like pathways. Patellamides are often produced in large amounts (up to several percent of animal dry weight), and presumably some selection pressure must be necessary to maintain such a large-scale synthesis. It is interesting that, because many *Prochloron* strains lack these pathways, other unknown selection mechanisms must be important to maintain symbiosis, and there are no obvious visible morphological differences between peptide-producer and nonproducer ascidians.

Nutrient exchange has been demonstrated to be important for some didemnid–*Prochloron* associations. Photosynthesis by *Prochloron* has been shown to provide 60–100% of the organic carbon theoretically needed by the host (49), and there is evidence for nitrogen cycling between host and symbiont in addition to nitrogen fixation (50). *L. patella* actively optimizes growth conditions for its symbiont by moving to regions with proper illumination and by modifying the structure of the tunic covering the upper surface of the colony (51).

Although these adaptations clearly show the importance of *Prochloron* for its didemnid host, the ecological function of the patellamides remains unclear. Because the patellamides are cytotoxic, they may serve as feeding deterrents. Patellamides are also known to chelate metals such as copper and zinc (52), and thus they may have a role in detoxification. These and other possibilities can be tested rigorously by using genetic, chemical, and ecological techniques.

The complete genome sequence of *P. didemni* will offer clues to the mechanisms responsible for maintaining symbiosis and will direct additional ecological studies. Comparative genomics, in tandem with field research, will help to determine the ecological importance of the secondary metabolites. Because of the visibility, ease of collection, and ubiquity of didemnid ascidians harboring *Prochloron*, these organisms may serve as ideal models for studying the ecological function (selection

pressure) and evolution of natural products in marine symbioses. These studies will be important, because they will improve our understanding of reef ecology and help direct the search for new drugs derived from marine invertebrates.

Conclusions. This work is the first example of genetics-based identification, transfer, and expression of a whole biosynthetic pathway from marine microbial symbionts. Because many, if not most, of the highly potent and bioactive marine natural products isolated from invertebrates have a symbiotic origin, this study shows that marine natural products may be accessed and supplied through genetic

manipulation. Although they are potent and have great pharmaceutical potential, marine natural products generally suffer from a problem of supply, and this study indicates the promise of genetic techniques in the alleviation of this problem.

We thank the Republic of Palau and the Coral Reef Research Foundation for providing the opportunity to collect samples and perform field research, B. O. Bachmann for helpful discussions, and E. M. Rachlin for mass-spectrometry training and help with selective reaction monitoring. This work was supported by the National Science Foundation's Emerging Frontier Program (EF-0412226) and Seed Incentive Fund and startup grants from the University of Utah.

- Newman, D. J. & Cragg, G. M. (2004) *J. Nat. Prod.* **67**, 1216–1238.
- Simmons, L. T., Andrianasolo, E., McPhail, K., Flatt, P. & Gerwick, W. H. (2005) *Mol. Cancer Ther.* **4**, 333–342.
- Faulkner, D. J., He, H. Y., Unson, M. D. & Bewley, C. A. (1993) *Gazz. Chim. Ital.* **123**, 301–307.
- Kobayashi, J. & Ishibashi, M. (1993) *Chem. Rev. (Washington, D.C.)* **93**, 1753–1770.
- Sings, H. L. & Rinehart, K. L. (1996) *J. Ind. Microbiol.* **17**, 385–396.
- Haygood, M. G., Schmidt, E. W., Davidson, S. K. & Faulkner, D. J. (1999) *J. Mol. Microbiol. Biotechnol.* **1**, 33–34.
- Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N. & Matsunaga, S. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 16222–16227.
- Hildebrand, M., Waggoner, L. E., Liu, H., Sudek, S., Allen, S. W., Anderson, C. M., Sherman, D. H. & Haygood, M. G. (2004) *Chem. Biol.* **11**, 1543–1552.
- Withers, N., Vidaver, W. & Lewin, R. A. (1978) *Phycologia* **17**, 167–171.
- Lewin, R. A. & Cheng, L. (1989) *Prochloron: A Microbial Enigma* (Chapman & Hall, New York).
- Withers, N. W., Alberte, R. S., Lewin, R. A., Thornber, J. P., Britton, G. & Goodwin, T. W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2301–2305.
- Degnan, B. M., Hawkins, C. J., Lavin, M. F., McCaffrey, E. J., Parry, D. L., van den Brenk, A. L. & Watters, D. J. (1989) *J. Med. Chem.* **32**, 1349–1354.
- Biard, J. F., Grivois, C., Verbist, J. F., Debitus, C. & Carre, J. B. (1990) *J. Mar. Biol. Assoc. U.K.* **70**, 741–746.
- Salomon, C. E. & Faulkner, D. J. (2002) *J. Nat. Prod.* **65**, 689–692.
- Williams, A. B. & Jacobs, R. S. (1993) *Cancer Lett.* **71**, 97–102.
- Fu, X., Do, T., Schmitz, F. J., Andrushevich, V. & Engel, M. H. (1998) *J. Nat. Prod.* **61**, 1547–1551.
- Carroll, A. R., Coll, J. C., Bourne, D. J., MacLeod, J. K., Zabriskie, T. M., Ireland, C. M. & Bowden, B. F. (1996) *Aust. J. Chem.* **49**, 659–667.
- Gehring, A. M., Mori, I., Parry, R. D. & Walsh, C. T. (1998) *Biochemistry* **37**, 11637–11650.
- Trauger, J. W., Kohli, R. M., Mootz, H. D., Marahiel, M. A. & Walsh, C. T. (2000) *Nature* **407**, 215–218.
- Li, Y. M., Milne, J. C., Madison, L. L., Kolter, R. & Walsh, C. T. (1996) *Science* **274**, 1188–1193.
- Solbiati, J. O., Ciaccio, M., Fariás, R. N., González-Pastor, J. E., Moreno, F. & Salomón, R. A. (1999) *J. Bacteriol.* **181**, 2659–2662.
- Schmidt, E. W., Sudek, S. & Haygood, M. G. (2004) *J. Nat. Prod.* **67**, 1341–1345.
- Huson, H. D., Reinert, K., Kravitz, Z. A., Remington, K. A., Delcher, A. L., Dew, I. M., Flanagan, M., Halpern, A. L., Lai, Z., Mobarry, C. M., *et al.* (2001) *Bioinformatics* **17**, S132–139.
- Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. (1999) *Nucleic Acids Res.* **27**, 4636–4641.
- Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., DeBoy, R. T., Haft, D. H., Dodson, R. J., *et al.* (2001) *Science* **293**, 498–506.
- Haft, D. H., Loftus, B. J., Richardson, D. L., Yang, F., Eisen, J. A., Paulsen, I. T. & White, O. (2001) *Nucleic Acids Res.* **29**, 41–43.
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. (2002) *Nucleic Acids Res.* **30**, 276–280.
- Tomitani, A., Okada, K., Miyashita, H., Matthijs, H. C. P., Ohno, T. & Tanaka, A. (1999) *Nature* **400**, 159–162.
- Ireland, C. M., Durso, A. R., Newman, R. A. & Hacker, M. P. (1982) *J. Org. Chem.* **47**, 1807–1811.
- Chatterjee, C., Paul, M., Xie, L. & van der Donk, W. A. (2005) *Chem. Rev. (Washington, D.C.)* **105**, 633–684.
- Garneau, S., Martin, N. I. & Vederas, J. C. (2002) *Biochimie* **84**, 577–592.
- Jack, R. W. & Jung, G. (2000) *Curr. Opin. Chem. Biol.* **4**, 310–317.
- Yorgey, P., Lee, J., Kordel, J., Vivas, E., Warner, P., Jebaratnam, D. & Kolter, R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 4519–4523.
- Wilson, K. A., Kalkum, M., Ottesen, J., Yuzenkova, J., Chait, B. T., Landick, R., Muir, T., Severinov, K. & Darst, S. A. (2003) *J. Am. Chem. Soc.* **125**, 12475–12483.
- Fomenko, D. E., Metlitskaya, A. Z., Peduzzi, J., Goulard, C., Katrukha, G. S., Gening, L. V., Rebuffat, S. & Khmel, I. A. (2003) *Antimicrob. Agents Chemother.* **47**, 2868–2874.
- Madison, L. L., Vivas, E. I., Li, Y.-M., Walsh, C. T. & Kolter, R. (1997) *Mol. Microbiol.* **23**, 161–168.
- van der Meer, J. R., Rollema, H. S., Siezen, R. J., Beerthuyzen, M. M., Kuipers, O. P. & de Vos, W. M. (1994) *J. Biol. Chem.* **269**, 3555–3562.
- Xie, L., Miller, L. M., Chatterjee, C., Averin, O., Kelleher, N. L. & van der Donk, W. A. (2004) *Science* **303**, 679–682.
- Schnell, N., Entian, K.-D., Schneider, U., Götz, F., Zähler, H., Kellner, R. & Jung, G. (1988) *Nature* **333**, 276–278.
- van der Meer, J. R., Polman, J., Beerthuyzen, M. M., Siezen, R. J., Kuipers, O. P. & De Vos, W. M. (1993) *J. Bacteriol.* **175**, 2578–2588.
- Balabas, B. E., Montgomery, B. L., Ong, L. E. & Kehoe, D. M. (2003) *Mol. Microbiol.* **50**, 781–793.
- Gonzalez-Pastor, J. E., San Millan, J. L., Castilla, M. A. & Moreno, F. (1995) *J. Bacteriol.* **177**, 7131–7140.
- Fuller, J. D., Camus, A. C., Duncan, C. L., Nizet, V., Bast, D. J., Thune, R. L., Low, D. E. & De Azavedo, J. C. (2002) *Infect. Immun.* **70**, 5730–5739.
- Breil, B., Borneman, J. & Triplett, E. W. (1996) *J. Bacteriol.* **178**, 4150–4156.
- Ichinose, K., Bedford, D. J., Tornus, D., Bechthold, A., Bibb, M. J., Revill, W. P., Floss, H. G. & Hopwood, D. A. (1998) *Chem. Biol.* **5**, 647–659.
- Milne, J. C., Roy, R. S., Eliot, A. C., Kelleher, N. L., Wokhlu, A., Nickels, B. & Walsh, C. T. (1999) *Biochemistry* **38**, 4768–4781.
- Banerjee, S. & Hansen, J. N. (1988) *J. Biol. Chem.* **263**, 9508–9514.
- Cinader, B. & Pillemer, L. (1950) *J. Exp. Med.* **92**, 219–237.
- Koike, I., Yamamuro, M. & Pollard, P. C. (1993) *Aust. J. Mar. Freshw. Res.* **44**, 173–182.
- Kline, T. C. & Lewin, R. A. (1999) *Symbiosis* **26**, 193–198.
- Swift, H. & Robertson, D. L. (1991) *Symbiosis* **10**, 95–113.
- van den Brenk, A. L., Fairlie, D. P., Hanson, G. R., Gahan, L. R., Hawkins, C. J. & Jones, A. (1994) *Inorg. Chem.* **33**, 2280–2289.
- Besemer, J., Lomsadze, A. & Borodovsky, M. (2001) *Nucleic Acids Res.* **29**, 2607–2618.