J|A|C|S COMMUNICATIONS

Published on Web 01/23/2009

# Using Marine Natural Products to Discover a Protease that Catalyzes Peptide Macrocyclization of Diverse Substrates

Jaeheon Lee,[†] John McIntosh,[‡] Brian J. Hathaway,[‡] and Eric W. Schmidt*[,†,‡]

*Departments of Biology and Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112*

Received November 25, 2008; E-mail: ews1@utah.edu

Numerous N−C terminally cyclized ribosomal peptides have been isolated from natural sources (Figure 1), yet enzymatic routes to these peptides remain largely unknown.[1−3] Cyclic peptides have structural features that make them good drug candidates, and in fact synthetic cyclic peptides are used clinically.[3,4] Ribosomal peptides can be artificially cyclized using intein technology, which although groundbreaking suffers from certain sequence and mechanistic requirements.[5] In principle, proteases could also catalyze this process, as exemplified by the proposed pilin cyclization catalyst, which was examined *in vivo* by mutagenesis,[6] and an artificial cyclization of a trypsin inhibitor.[7] The potential synthesis of cyanobactins and cyclotides by proteases is also supported by *in vivo* analysis.[2,8−11] However, no proteins have yet been characterized *in vitro* whose native function is to catalyze N−C cyclization.

Cyanobactins provide excellent models for the study of ribosomal N−C terminal cyclization. These compounds are synthesized as ribosomal precursor peptides that are extensively modified by enzymes to yield cyclic natural products.[2] Cyanobactins are found in diverse cyanobacteria, including symbionts of marine animals. In the first report of this pathway type, a cyanobactin gene cluster, *pat*, was cloned from the metagenome of a coral reef animal and transferred to *E. coli*, leading to production in the laboratory of a cyclic peptide marine natural product previously attributed to an animal.[2] Subsequently, each gene in the *pat* cluster (*patA-G*) was independently expressed under control of a T7 promoter to determine that *patA,D,E,F,G* were necessary and sufficient for cyanobactin biosynthesis and *patB,C* were not required.[10] However, the enzymatic basis of each transformation could not be discerned because removing any essential enzyme led to a loss of detectable products.

Using this initial discovery, we proceeded to assign the function of genes in the pathway by comparing over 30 closely related pathways from the cyanobacterial symbionts that we showed earlier to be the true producers of the cyclic peptides—a novel approach to functional analysis, which we term "deep metagenome mining".[12] This method was applied to show that cyanobactins are made by enzymes with broad substrate selectivity (Figure 1 and Table S1).[10,12] We then showed that wholly artificial cyclic peptides could be generated by this system *in vivo*.[10] The putative cyanobactin peptide cyclase enzymes were demonstrated by these methods to be broadly substrate tolerant.

Most cyanobactin precursor peptides contain two product "cassettes" that are flanked by putative modifying enzyme recognition sequences (Figures 1 and 2).[2,10] Thus, the cassettes are cleaved out of the full-length precursor at both their C- and N-termini and then cyclized to form the cyanobactin products.

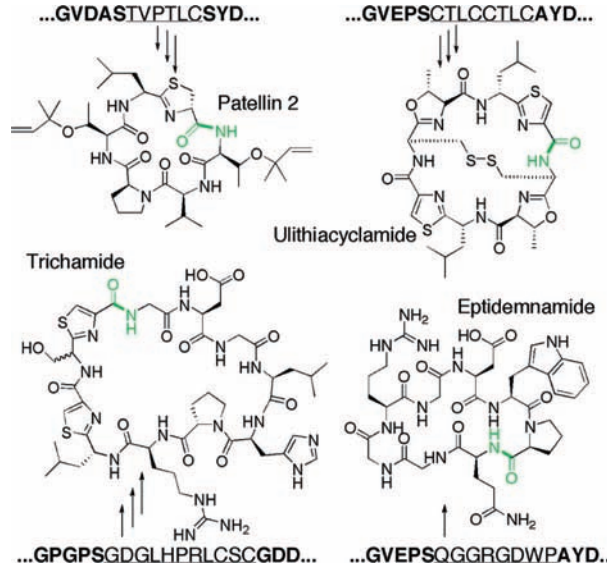


***Figure 1.*** N−C cyclic peptides of the cyanobactin group. The partial precursor peptide sequences are shown that give rise to the cyclic products. Amide bonds formed by PatG and its relatives are shown in green, while recognition sequences are shown in bold and cassette sequences leading to products are underlined. Patellin 2, ulithiacyclamide, and trichamide are natural products that require multiple enzymatic processing steps in their synthesis, while the eptidemnamide precursor was engineered and requires only cleavage and cyclization.

It seemed probable that proteases PatA and PatG could participate in these steps, but several other enzymes are also required to synthesize cyanobactin natural products.[13] Although much experimental evidence indicated that PatA and PatG were involved in this process, their precise roles were not defined. Here, we report on the *in vitro* activity of these enzymes, defining their roles in cyclization of natural and artificial peptides and providing the first *in vitro* study of natural ribosomal peptide N−C cyclization.

Previously, we genetically engineered an artificial precursor peptide, PatEdm, encoding the new compound eptidemnamide.[10] PatEdm encodes the sequence for the natural cyclic peptide patellamide C in cassette 1 and the sequence for the synthetic cyclic peptide eptidemnamide in cassette 2 (Figure 2). Eptidemnamide was produced by the full complement of cyanobactin biosynthetic proteins using heterologous *in vivo* experiments.[10] Because eptidemnamide requires only cyclization, and not the other enzymatic transformations required to produce the natural products, it was used for biochemical experiments described below.

Both PatA and PatG proteases are clearly paralogous and contain homologous domains, including a subtilisin-like serine

[†] Department of Biology.
[‡] Department of Medicinal Chemistry.



10.1021/ja8092168 CCC: $40.75 

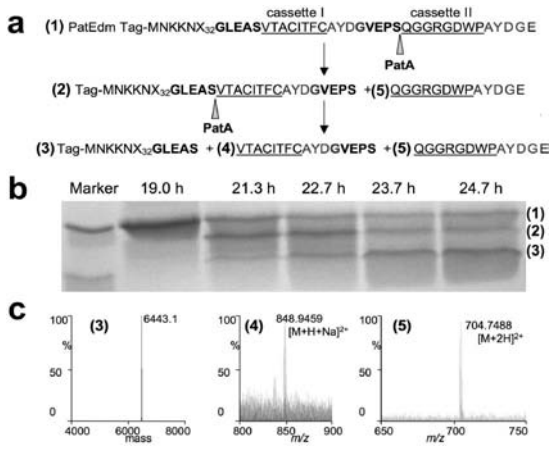


***Figure 2.*** PatA catalyzes two N-terminal cleavage reactions. (a) The precursor peptide PatEdm is cleaved by PatA at a recognition sequence to release first **5** and then **4**, with free N-termini for cyclization. (b) SDS-PAGE of a time-course experiment with PatA and PatEdm shows that the first site is slowly cleaved, followed by relatively rapid proteolysis of the second site. (c) All reaction products were observed by mass spectrometry.

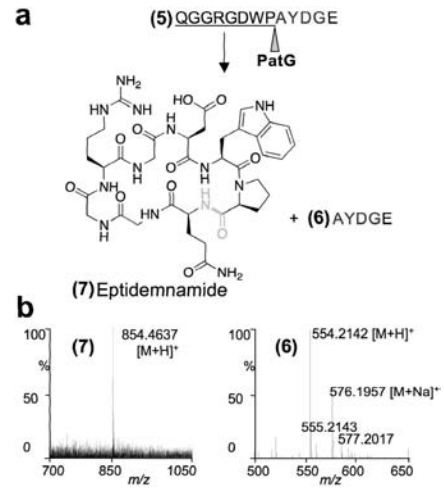


***Figure 3.*** PatG catalyzes cyclization by a transamidation mechanism. (a) Starting from product **5** of PatA, PatG cleaves this peptide into a cyclic compound, eptidemnamide, and a small linear fragment representing the recognition sequence. (b) High resolution mass spectrometry and comparison to authentic standards confirmed that this cyclic peptide was formed from compound **5**.

protease domain and a domain of unknown function. PatG also contains a N-terminal oxidase domain that is not involved in peptide cyclization as shown by metagenomic analysis.[13] The PatA and PatG proteases contain the Asp-Ser-His catalytic triad,[14] as shown by sequence alignment and protein threading (Figure S1).

To determine the potential roles of PatA and PatG in the cyclization of ribosomal peptide, we expressed and purified several constructs of PatA, PatG, and the artificial precursor peptide, PatEdm. Subsequent biochemical analysis defined the roles of each of these proteins. Expressed PatA was isolated as the full-length product with a mixture of fragments, some of which contained only the subtilisin-like protease domain (Figure S2). This domain was neatly cleaved at the linker region between the C-terminal protease and the domain of unknown function, as determined by tryptic-digest mass spectrometry. The protease domain could be purified to homogeneity. PatEdm was purified at high yield despite the inherent difficulties often encountered in expression of precursor peptides.

Both whole PatA, containing a mixture of fragments, and the purified protease domain of PatA were used for biochemical experiments. When incubated overnight with purified PatA or PatA protease domain, PatEdm was slowly cleaved to release first the eptidemnamide-encoding sequence and then the internally located patellamide-C sequence (Figure 2 and Table S3). To further confirm that this protease activity originated in PatA, a Ser-Ala active-site mutant was purified and shown to be inactive in this assay. In time-course experiments, the larger fragment was clearly produced first, followed by accumulation of the shorter fragment (Figure 2). By high-resolution mass spectrometry, we could observe the N-terminal leader sequence, the patellamide-C coding sequence, and the eptidemnamide-coding sequence (Figure 2 and S4). A synthetic standard of the eptidemnamide-coding sequence was used to further confirm this analysis (Figure S3). The enzyme reaction was very slow, but it led to eventual quantitative cleavage of the precursor peptide no matter what enzyme/substrate ratio was employed. This enzyme reaction is not optimal as is discussed further in the paragraphs concerning the PatG protein.

The precursor peptide, PatEdm, was cleaved neatly at the predicted N-terminal recognition sequence, G(L/V)E(A/P)S in both instances (Figure 2 and Figure S4). Thus, the cyanobactin cassettes fused to their conserved C-terminal recognition sequences, AYDG(E), were released upon incubation with PatA. By contrast, when incubated with PatEdm, PatG did not catalyze any detectable proteolysis in numerous different experimental conditions (Table S3). We therefore predicted that PatG would catalyze cleavage and cyclization of the C-terminal recognition sequence and that the products of PatA would be the substrates of PatG.

To assess the PatG reaction, we used synthetic peptide substrate QGGRGDWP-AYDGE. The substrate was identical to the product released by PatA from cassette II, as shown by high resolution LC−MS (Figures 2, 3, S3, and S4).

Several PatG variants were used with the synthetic substrate. Wild-type PatG was expressed, and in addition we constructed a variant in which the N-terminal oxidase domain was absent. Finally, TruG is a related protein that naturally lacks the oxidase domain (Figure S2). These three proteins were separately incubated with the synthetic substrate, either with or without the addition of PatA protease. All of these combinations led to production of cyclic eptidemnamide (Figure 3 and S5, Table S3). The protease recognition sequence, AYDGE, was also clearly discernible in these reaction mixtures as a discrete cleavage product. Cofactors and added metals were not required to catalyze the reaction. In the absence of PatG, with PatA only or without enzyme for example, only starting synthetic peptide was observed.

The presence of eptidemnamide was confirmed by comparing diode array-HPLC and high resolution MS/MS data with an authentic standard (Figures 3, S3, and S5). After 41 h, approximately 1.5 $\mu$g of eptidemnamide were synthesized from 50 $\mu$g of precursor peptide, representing a ~5% yield. No linear eptidemnamide analogue was detected in these experiments, although unreacted synthetic precursor was still clearly observed. This yield could be optimized by addition of more enzyme. Doubling the enzyme led to double the amount of product (~10% yield), while an increase to a 3:1 substrate/enzyme ratio led to a 75% ± 10% yield.

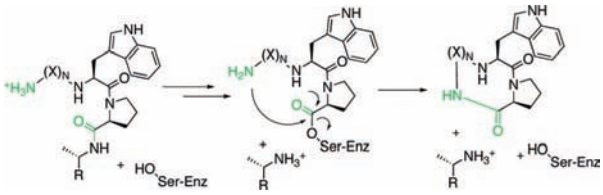

***Figure 4.*** Proposed transamidation mechanism of peptide cyclization. In this simplified scheme, the active-site Ser of PatG is shown forming an enzyme−substrate tetrahedral intermediate with loss of the AYDGE recognition sequence. Subsequently, the N-terminus of the substrate attacks the activated ester bond to release a cyclic peptide and the free enzyme.

This process is clearly far from optimal, since only ~1 turnover is observed per enzyme molecule per day. Despite this slow rate, each monomeric enzyme unit turns over substrate between >2 to 3 times during the course of the experiment under current conditions. Therefore, PatG is shown to be the cyclization catalyst, as this turnover exceeds a stoichiometric reaction. There are several possibilities that could explain low turnover. The substrate is artificial, and the reaction proceeds *in vitro*. Many *in vitro* enzyme reactions require other enzyme partners or buffer optimization to achieve a maximum rate, since they are highly artificial in nature. As one of many possibilities, perhaps the proteases require partnering with an export protein, as is common in lantibiotics. Finally, it should be noted that a single 1 L expression of PatG provides enough protease to synthesize several milligrams of peptide even without optimization.

Previously, it was proposed that cyanobactin cyclization could proceed spontaneously.[15] In this event, PatG would act merely as a protease, and the small linear product would cyclize on its own or via PatG catalysis at a nonproteolytic site. To rule out these possibilities, we obtained the linear precursor, QG-GRGDWP, which lacked the AYDGE sequence. In enzyme reactions with PatG in various combinations and in comparison to a full set of control reactions, this peptide was neither modified nor cyclized (Table S3). Thus, cyclization is not spontaneous, and several lines of evidence favor a one-step transamidation mechanism to cyclization. The data currently support the mechanism shown in Figure 4, although further experiments are required for confirmation.

From these experiments, it is clear that PatA catalyzes the proteolytic cleavage of the N-terminal recognition sequence, while PatG catalyzes cleavage of the C-terminal sequence in tandem with cyclization. No addition of energy (in the form of ATP, for example) is required to drive this process, making PatG the first natural peptide-cyclizing enzyme to be characterized as a pure protein. Aside from the lack of an energy requirement, PatG has another potential technological advantage in that it is highly tolerant of diverse substrate sequences as long as the C-terminal recognition domain is present. So far, we have identified >30 natural peptide sequences that appear to be cleaved and cyclized by PatG (Table S1).[10,12] These data resemble results of alanine-scanning mutagenesis experiments, in that every amino acid position is substituted at least once by a nonconserved amino acid residue. Six, seven, and eight amino acid peptides are cyclized by PatG *in vivo* based upon this same analysis. In addition, close relatives of PatG catalyze the synthesis of >60 additional known natural products (Table S1). Finally, the wholly synthetic peptide, eptidemnamide, is cyclized by PatG *in vivo*[10] and *in vitro*. The exact limitation of product size and sequence content are not yet known. If this enzyme process can be improved, and if broad substrate activity can be demonstrated *in vitro*, the enzyme could serve as a useful general catalyst for the cyclization of peptides.

**Acknowledgment.** This work was funded by NIH GM071425. We thank T. Bugni and J. Reppart for help with mass spectrometry and G. Bulaj, C. Robertson, and B. Green for synthesis of the PatG substrate peptide.

**Supporting Information Available:** Experimental details, alignment data, tables, and MS data. The material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Trabi, M.; Craik, D. J. *Trends Biochem. Sci.* **2002**, *27*, 132–138.
(2) Schmidt, E. W.; Nelson, J. T.; Rasko, D. A.; Sudek, S.; Eisen, J. A.; Haygood, M. G.; Ravel, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7315–7320.
(3) Wang, C. K.; Kaas, Q.; Chiche, L.; Craik, D. J. *Nucleic Acids Res.* **2008**, *36*, 210.
(4) Scarborough, R. M. *Am. Heart J.* **1999**, *138*, 1093–1104.
(5) Scott, C. P.; Abel-Santos, E.; Wall, M.; Wahnon, D. C.; Benkovic, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13638–13643.
(6) Eisenbrandt, R.; Kalkum, M.; Lurz, R.; Lanka, E. *J. Bacteriol.* **2000**, *182*, 6751–6761.
(7) Marx, U. C.; Korsinczky, M. L.; Schirra, H. J.; Jones, A.; Condie, B.; Otvos, L., Jr.; Craik, D. J. *J. Biol. Chem.* **2003**, *278*, 21782–21789.
(8) Long, P. F.; Dunlap, W. C.; Battershill, C. N.; Jaspars, M. *ChemBioChem*, **2005**, *6*, 1760–1765.
(9) Saska, I.; Gillon, A. D.; Hatsugai, N.; Didetzgen, R. G.; Hara-Nishimura, I.; Anderson, M. A.; Craik, D. J. *J. Biol. Chem.* **2007**, *282*, 29721–29728.
(10) Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2006**, *2*, 729–735.
(11) Gillon, A. D.; Saska, I.; Jennings, C. V.; Guarino, R. F.; Craik, D. J.; Anderson, M. A. *Plant J.* **2008**, *53*, 505–515.
(12) Donia, M. S.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2008**, *4*, 341–343.
(13) Sudek, S.; Haygood, M. G.; Youssef, D. T.; Schmidt, E. W. *Appl. Environ. Microbiol.* **2006**, *72*, 4382–4387.
(14) Siezen, R. J.; Renckens, B.; Boekhorst, J. *Proteins* **2007**, *67*, 681–694.
(15) Milne, B. F.; Long, P. F.; Starcevic, A.; Hranueli, D.; Jaspars, M. *Org. Biomol. Chem.* **2006**, *4*, 631–638.

JA8092168