J|A|C|S COMMUNICATIONS

Published on Web 10/20/2010

Circular Logic: Nonribosomal Peptide-like Macrocyclization with a Ribosomal Peptide Catalyst

John A. McIntosh,[†] Charles R. Robertson,[†] Vinayak Agarwal,[‡] Satish K. Nair,[‡,§] Grzegorz W. Bulaj,[†] and Eric W. Schmidt*[,†]

Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112, and Center for Biophysics and Computational Biology, Department of Biochemistry, University of Illinois, Urbana−Champaign, Illinois 61801

Received July 29, 2010; E-mail: ews1@utah.edu

Abstract: A protease from ribosomal peptide biosynthesis macrocyclizes diverse substrates, including those resembling nonribosomal peptide and hybrid polyketide−peptide products. The proposed mechanism is analogous to thioesterase-catalyzed chemistry, but the substrates are amide bonds rather than thioesters.

Macrocyclization is a common strategy to improve the rigidity and stability of bioactive metabolites.[1,2] In polyketide and nonribosomal peptide biosynthesis, macrocyclization via lactones or lactams is typically catalyzed by thioesterase (TE) domains, which contain a serine protease-like Asp-His-Ser catalytic triad (Figure 1).[1] The TE domain transfers the peptide/polyketide chain from a carrier protein to the active site serine, which is then displaced by a nucleophile, generating either a linear product or more commonly a macrocycle, as in tyrocidine A (**1**).

Quite interestingly, one of the major groups of macrocyclic ribosomal peptides, the cyanobactins,[3] is cyclized in a similar way: a subtilisin-like serine protease catalyzes cleavage of a C-terminal peptide sequence in tandem with N−C macrolactamization, leading to compounds such as patellamide C (**2**).[4−6] Among ribosomal peptides, both the cyanobactins and cyclotides are N−C cyclic, while other ribosomal peptides, such as capistruin,[7] microcin J25,[8] and the microviridins,[9] are cyclized via side-chain residues using ATP via wholly different biochemical mechanisms. There is indirect evidence that cyclotides are circularized in a similar fashion to cyanobactins,[10] and circular ribosomal peptides are common in diverse organisms. However, no definitive enzymatic or genetic studies of N−C macrocyclization have been performed on any ribosomal peptide system other than the cyanobactins.

Previously, we have shown that the subtilisin-like protease, PatG, is solely responsible for catalyzing macrocyclization in the patellamide pathway.[6] Metagenomic and biochemical analyses of the patellamide pathway showed that PatG is a broad-substrate enzyme, processing 29 known precursor peptide sequences encoding macrocycles of 7−8 amino acids.[4,5] Every natural product residue is mutated at least once in this series, and PatG could also produce the unnatural compound eptidemnamide (**3**) both *in vivo* and *in vitro* (Figure 2). Consequently, the cyanobactin macrocyclases exhibit exceptionally relaxed substrate specificity. Furthermore, unlike other ribosomal peptide natural product pathways,[11,12] the cyanobactin macrocyclases require only a C-terminal 4−5 amino acid recognition sequence, AYDG(E), which allowed us to employ short synthetic peptides as substrates.

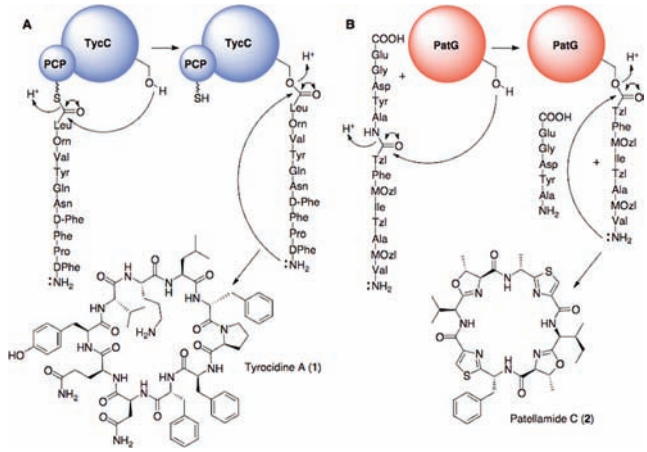

Figure 1. Macrocyclization in ribosomal and nonribosomal synthesis. (A) The nonribosomal TycC TE domain circularizes tyrocidines. PCP = peptidyl carrier protein. (B) The ribosomally acting PatG protease circularizes patellamides and many other compounds. The proposed catalytic mechanism is indicated here. Tzl = thiazol(ine); MOzl = methyloxazoline.

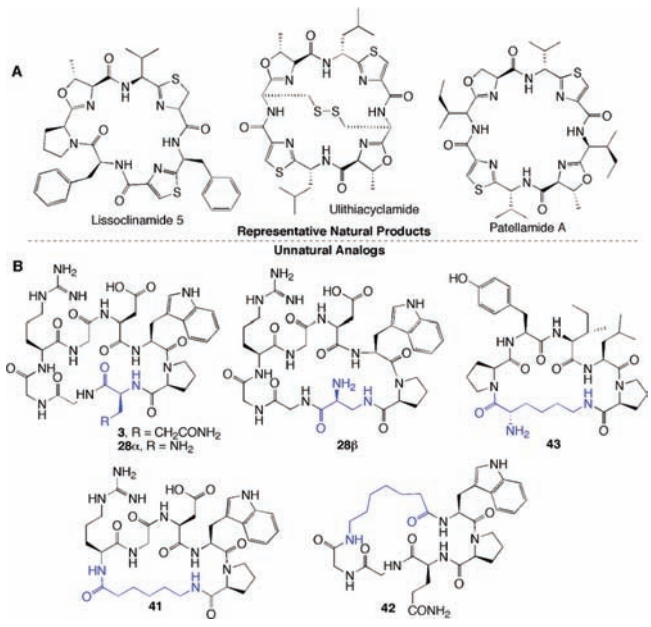

Figure 2. Products of PatG macrocyclization. (A) Representative known natural products cyclized by PatG, out of a total of 39 known natural products in this series. (B) Macrocyclization products from this study, showing side-chain circularization (**28**, **43**) and polyketide-like cyclization (**41**, **42**).

We proposed that PatG performs macrocyclization in a manner that is mechanistically analogous to TE domains, although the proteins and the substrates are quite different. In particular, while

[†] University of Utah.
[‡] Center for Biophysics and Computational Biology, University of Illinois.
[§] Department of Biochemistry, University of Illinois.

Table 1. Synthetic Peptide Substrates and Macrocyclic PatG Products[a]

Compd	Sequence	Compd	Products
4	QGGRGDWPAYDGE	**3**	C
5	QGGRGDW**A**AYDGE	–	NR
6	(**Dap**)GGRGDWPAYDGE	**28**	C $\alpha > \beta$
7	(**Orn**)GGRGDWPAYDGE	**29**	C α:δ ~1:1
8	**K**GGRGDWPAYDGE	**30**	C $\alpha \gg \epsilon$
9	(**Glyc**)GGRGDWPAYDGE	**31**	L
10	GGRGDWPAYDGE	**32**	C
11	**q**GGRGDWPAYDGE	**33**	L
12	QGG**r**GDWPAYDGE	**34**	C
13	QGGRG**d**WPAYDGE	**35**	C
14	QGGRGD**w**PAYDGE	–	NR
15	QGGRGDW**p**AYDGE	–	NR
16	QGGWPAYDGE	–	NR
17	Q**G**GGRGDWPAYDGE	**36**	C
18	Q**GG**GGRGDWPAYDGE	**37**	C
19	Q**GGG**GGRGDWPAYDGE	**38**	C
20	**G**GGRGDWPAYDGE	**39**	C
21	**GG**GGRGDWPAYDGE	**40**	C
22	(**Ahx**)RGDWPAYDGE	**41**	C:L 1:2
23	QGG(**Ahp**)WPAYDGE	**42**	C:L 1:2
24	KPYILPAYDGE	**43**	C:L 1:2, $\alpha \ll \epsilon$
25	**K**KPYILPAYDGE	**44**	C:L 1:2
26	**K**KPY**ll**PAYDGE	–	NR
27	GWTLNSAGYLLGPAYDGE	–	NR

[a] The PatG recognition element, AYDGE, is underlined. Residues in bold indicate differences from canonical type sequences **4**, **24**, and **27**, while lower-case residues are in the D-configuration. C = circular; NR = no reaction; L = linear with predicted cleavage between P and A. If only C or L is indicated, >90% of products were in that form only.

as their name implies TEs require activated thioesters (or esters)[13] to catalyze circularization, the PatG substrates are simple amides readily accessible through standard solid phase peptide synthesis. Based on this mechanistic hypothesis, in this study we probed the capacity of PatG to circularize peptides containing nonproteinogenic amino acids and polyketide-like linkers. We show that PatG synthesizes macrocycles that are similar to those from the thiotemplate-based pathways,[13] lending weight to the mechanistic hypothesis and providing a significant step toward bridging ribosomal and nonribosomal worlds for the synthesis of complex peptide metabolites.

In this study we tested 23 analogues (**5**−**27**) to define the substrate tolerance of PatG. The analogues are of variable length and amino acid composition, but nearly all contain Pro followed by the macrocyclase recognition sequence AYDGE (Table 1). Pro was used because all of the natural compounds contain heterocycles immediately prior to AYDGE. These heterocycles are either Pro or thiazol(in)e/oxazol(in)e derived from Cys, Ser, or Thr. One exception (**5**) contained Ala in place of Pro to test this putative heterocycle requirement.

Analogues **5**−**23** are based on the previously reported PatG substrate **4**, which leads to eptidemnamide (**3**). **6**−**9** contain N-terminal residue substitutions that explore the potential for cyclization via side-chain nucleophiles, including OH in glycolate (Glyc) and NH_2 in diaminopropionate (Dap), Lys, and ornithine (Orn). **11**−**15** explore the tolerance for D-amino acids. In **10**, **16**−**21**, **24**, and **27** cyclizable sequence lengths from 5 to 11 amino acids are explored. **22** and **23** resemble polyketide−peptide hybrids, containing alkyl spacers aminohexanoic acid (Ahx) and aminoheptanoic acid (Ahp) that replace portions of the eptidemnamide sequence. Finally, to further explore substrate selectivity several substrates with wholly different sequences were attempted. **24**−**26** are based on the hormone neurotensin, while **27** mimics the neuropeptide galanin. The peptides were synthesized via standard solid phase synthesis (Supporting Information).

To assay macrocyclization, analogues were incubated with the PatG protease domain (Supporting Information). As controls, substrates were also incubated in an equivalent manner using a site-directed mutant in which the active-site Ser of PatG was replaced with Ala. Products were analyzed via matrix-assisted laser desorption ionization-mass spectrometry (MALDI-MS), which allowed us to readily observe the loss of starting material and the accumulation of products of the expected mass; for the most part these products are the only reasonable chemical entities that fit the mass. Representative products from each compound family were confirmed by high-resolution MS and MS-MS on a Fourier transform ion cyclotron resonance (FT-ICR) instrument. This method unambiguously establishes the position of circularization because numerous ions are present that only could arise if the amide bond was synthesized as shown in Figure 2. For full details, see Supporting Information. The identity of **3** was previously further confirmed by 2D NMR.[5]

For example, for compound **41** containing the polyketide-like Ahx, the predicted linear *m*/*z* would be 743 while the cyclic product would be *m*/*z* = 725, indicating a peptide with a mass decreased by 18 Da ($-H_2O$). Both ions were observed as the major nonmatrix peaks in the MALDI spectrum. This sample was then applied to LC-MS using an FT-ICR instrument, providing parent ions for both molecules that confirmed their molecular formulas, deviating by δ 0.5 ppm for the linear and δ 0.91 ppm for the cyclic variants. The MS/MS data for the cyclic product provided a very complex series of ions that overlapped the new cyclizing amide bond (Pro-Ahx) and that were inconsistent with possible alternative structures. For the linear product these ions arising from cyclization were absent, while the peaks for the linear portion were much more intense, and the resulting spectrum was much simpler.

Results are summarized in Table 1 and Figure 2b. As seen from precursor **5**, Ala is not accepted in place of Pro, supporting a requirement for a heterocyclic motif at the last position. All of the side-chain nucleophile peptides (**6**−**9**, **24**, and **25**) are substrates for the enzyme, although the OH-containing glycolate peptide **9** is linearized. Surprisingly, D-amino acids can be tolerated in some central positions (**12** and **13**), but apparently not too close to the C- or N-termini (**11**, **14**, **15**, and **26**). However, even when epimerization could be tolerated, MALDI-MS indicated that the reactions did not approach completion. PatG synthesized products in lengths from 6 to 11 amino acids (**10**, **17**−**21**, **24**), although there was some substrate-dependence as **27** was not a substrate. Peptide **16**, which could only form a five amino acid cycle, was not circularized.

Most significantly, peptides **22** and **23** containing polyketide-like linker regions were also substrates for circularization by PatG. Previously, it was shown that the tyrocidine TycC TE domain could cyclize hybrid polyketide−peptide esters, indicating a fundamental similarity in the biochemistry of these two enzyme classes despite wholly different sequences and substrates.[13]

TE domains often catalyze hydrolysis rather than macrocyclization − unnatural substrate analogues often trigger the hydrolysis reaction in place of a natural cyclization − and PatG was similar in this regard. Depending upon the substrate, PatG synthesized cyclic/linear products in ratios between 1:2 (for **22**) and >10:1 (for **8**) as determined by fluorescence HPLC and mass spectrometry. These cyclic or linear (-AYDGE) products were not formed in the active site Ser-Ala mutant of PatG. Products closely related to eptidemnamide (**3**) were always either completely (>90%) circularized or linearized, or else they were nonreactive. In short, they did not exhibit combinations of linear and cyclic products. The only

exceptions in this group were **22** and **23**, with highly flexible linkers, which were hydrolyzed in ratios reminiscent of nonribosomal TE domains.

The regioselectivity of macrocyclization for **6**−**8** and **24** was determined by derivatizing the HPLC purified cyclic peptide with 1-fluoro-2,4-dinitrobenzene, followed by acid hydrolysis. The hydrolysates were compared using HPLC and/or LC-MS with authentic standards of α- and *β*-amino dinitrophenyl (DNP) Dap, α- and *δ*-amino DNP-Orn, or α- and *ε*-amino DNP-Lys. The results show that side-chain cyclization via Dap, Orn, or Lys did occur in some substrates (**6**−**7**, **24**). Substrates with N-terminal Dap or Orn residues yielded a mixture of regioisomers, while substrates with N-terminal Lys showed nearly total selectivity for either backbone- (as in **8**) or side-chain-cyclized (as in **24**) peptides.

Taken together, these results show that PatG circularizes a broad array of substrates, including those with nonproteinogenic and D-amino acids and those containing polyketide-like linkers. It should be remarked that, based upon known natural products, PatG is also known to circularize an additional 29 natural substrates that encapsulate many extremely different amino acid sequences than those examined in this study. There are some limitations to substrate selectivity that are apparent in unreacted substrates, but the substrate specificity is remarkably broad and comparable to TE domains, especially given that lengths of 6−11 amino acids are effectively circularized through the terminal amino acid. It is also a remarkably broad-substrate enzyme, with a relatively short (4−5 amino acid) recognition sequence directing reactions with many different substrates and precluding water from the active site despite extreme differences in substrate length and constitution.

In this and previous *in vitro* studies using PatG, a limitation is its exceptionally slow rate, with some reactions complete in ∼24 h using a 50% catalyst load. In previous work with PatG, lower catalyst loads led to complete conversion in longer reaction periods, demonstrating that PatG acts catalytically, albeit slowly *in vitro*.[6] Moreover, *in vivo* other cyanobactin macrocyclases seem relatively efficient, where yields of up to 2.5 mg of compound per L of culture have been observed,[14] indicating that *in vitro* reactions could likely be improved. To the best of our knowledge, this report represents the first time that a ribosomal natural product catalyst has been shown to accept such extremely diverse, unnatural substrates *in vitro*.

In summary, although PatG is a catalyst from ribosomal peptide natural product synthesis that operates on amide bonds, its behavior is reminiscent of TE domains from nonribosomal and polyketide synthesis. Although this enzyme is not optimal for *in vitro* use, ultimately, it is hoped that this and other studies will provide a toolkit for genetic engineering of diverse small molecules *in vivo* and will help to bridge nonribosomal and ribosomal biosynthesis.[15] It is relatively straightforward to engineer peptide production *in vivo* using ribosomal synthesis, but nonribosomal machinery leads to much more chemically diverse products. Nonproteinogenic amino acids, such as those found in nonribosomal peptides, can already be ribosomally encoded using existing technology.[16] Ultimately, a combination of tools such as the enzymatic methods described here in concert with the ability to add unusual functions using the ribosome itself will enable the production of chemically diverse products *in vivo*. By bridging the biochemistry of these two worlds, the goal is to take advantage of the engineering simplicity of the ribosome while synthesizing the elaborate products more typical of complex nonribosomal peptides.

Acknowledgment. At U. Utah, we thank S. Endicott and R. Schackmann for synthesizing many of the analogues used in this study; C. Hubbard and A. Barrios for peptide purification assistance; D. Davis for helpful discussions; J. Muller, K. Parsawar, and C. Nelson for mass spectrometry assistance; and J. Olsen and D. Tianero for help with NMR. This work was partly funded by an ACS Medicinal Chemistry predoctoral fellowship funded by Sanofi Aventis to J.A.M. as well as NIH Grant GM071425 to E.W.S.

Supporting Information Available: Full methods and data are available. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kopp, F.; Marahiel, M. A. *Nat. Prod. Rep.* **2007**, *24*, 735–49.
(2) Burton, P. S.; Conradi, R. A.; Ho, N. F.; Hilgers, A. R.; Borchardt, R. T. *J. Pharm. Sci.* **1996**, *85*, 1336–40.
(3) Donia, M. S.; Schmidt, E. W.; Lew, M.; Hung-Wen, L. In *Comprehensive Natural Products II*; Elsevier: Oxford, 2010; pp 539−558.
(4) Donia, M. S.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2008**, *4*, 341–3.
(5) Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2006**, *2*, 729–35.
(6) Lee, J.; McIntosh, J. A.; Hathaway, B. J.; Schmidt, E. W. *J. Am. Chem. Soc.* **2009**, *131*, 2122–4.
(7) Knappe, T. A.; Linne, U.; Robbel, L.; Marahiel, M. A. *Chem. Biol.* **2009**, *16*, 1290–8.
(8) Duquesne, S.; Destoumieux-Garzon, D.; Zirah, S.; Goulard, C.; Peduzzi, J.; Rebuffat, S. *Chem. Biol.* **2007**, *14*, 793–803.
(9) Ziemert, N.; Ishida, K.; Liaimer, A.; Hertweck, C.; Dittmann, E. *Angew. Chem., Int. Ed.* **2008**, *47*, 7756–9.
(10) Henriques, S. T.; Craik, D. J. *Drug Discov. Today* **2010**, *15*, 57–64.
(11) Cheung, W. L.; Pan, S. J.; Link, A. J. *J. Am. Chem. Soc.* **2010**, *132*, 2514–5.
(12) Philmus, B.; Guerrette, J. P.; Hemscheidt, T. K. *ACS Chem. Biol.* **2009**, *4*, 429–34.
(13) Kohli, R. M.; Burke, M. D.; Tao, J.; Walsh, C. T. *J. Am. Chem. Soc.* **2003**, *125*, 7160–1.
(14) Donia, M. S.; Ruffner, D. E.; Cao, S.; Schmidt, E. W. **2010**, *Appl. Environ. Microbiol.* Submitted.
(15) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *Nat. Prod. Rep.* **2009**, *26*, 537–559.
(16) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem.* **2010**, *79*, 413–44.

JA1067806