ORGANIC LETTERS

2011
Vol. 13, No. 19
5144–5147

NRPS Substrate Promiscuity Diversifies the Xenematides

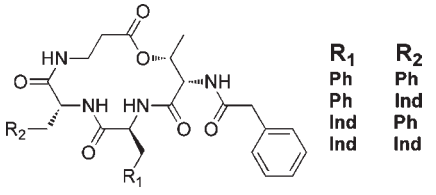
Jason M. Crawford, Cyril Portmann, Renee Kontnik, Christopher T. Walsh, and Jon Clardy*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, United States

jon_clardy@hms.harvard.edu

Received July 26, 2011

ABSTRACT



Xenematide, a cyclic depsipeptide antibiotic produced by *Xenorhabdus nematophila*, had a candidate nonribosomal peptide synthetase (NRPS) with atypical features. Differential metabolite analysis between a mutant and wildtype validated that this stand-alone NRPS was required for xenematide production, and further analysis led to a series of new xenematide derivatives encoded by the same NRPS. Our results indicate that adenylation domain promiscuity and relaxed downstream processing in the *X. nematophila* NRPS provide a conduit for xenematide diversification.

Bacteria belonging to the *Xenorhabdus* genus engage in a mutualistic symbiosis with insect-pathogenic *Steinernema* nematodes.[1] After invading an insect host, the nematode regurgitates the bacteria into the host's amino acid rich circulatory system (hemolymph). The released bacteria produce a mélange of small molecules that regulate the insect–bacteria–nematode symbiosis in which the bacteria and nematode prosper by cooperatively killing and consuming the dead insect. The bacteria also synthesize small molecule antimicrobials to reduce competition. We recently discovered that L-proline, an abundant amino acid in insect hemolymph, contributes to the bacterial proton motive force and upregulates the production of secondary metabolites in *Photorhabdus luminescens* TT01 and *Xenorhabdus nematophila* 19061.[2] Xenematide A (**9**), a cyclic depsipeptide with antibacterial activity against both Gram-negative and Gram-positive bacteria,[3] is one of the dramatically upregulated metabolites in *X. nematophila*.[2] Here we identify the nonribosomal peptide synthetase (NRPS) responsible for xenematide biosynthesis, discover a small family of new xenematide derivatives originating from substrate promiscuity in the NRPS pathway, and propose a unified biosynthesis for xenematide assembly.

Xenematide A (**9**), which incorporates the nonproteinogenic amino acid β-Ala and a phenylacetate (PAA) starter unit, would most plausibly arise from an NRPS pathway rather than a ribosomal pathway. NRPS pathways are modular enzymatic assembly lines whose products can largely be predicted from the number and order of their modules and the specific domains within each module according to the colinearity rule.[4] In an ATP-consuming reaction, adenylation (A) domains link free amino acid building blocks onto their cognate peptidyl carrier protein (PCP; or T, thiolation) domains through the formation of a thioester bond to a phosphopantetheinyl prosthetic group. The amino acid specificity encoded by an A domain can frequently be predicted from gene sequence alone.[5] Condensation (C) domains catalyze amide bond formation, and the nonproteinogenic peptide is elongated

(1) Richards, G. R.; Goodrich-Blair, H. *Cell Microbiol.* **2009**, *11*, 1025–1033.

(2) Crawford, J. M.; Kontnik, R.; Clardy, J. *Curr. Biol.* **2010**, *20*, 69–74.

(3) Lang, G.; Kalvelage, T.; Peters, A.; Wiese, J.; Imhoff, J. F. *J. Nat. Prod.* **2008**, *71*, 1074–1077.

(4) Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006**, *106*, 3468–3496.

(5) (a) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999**, *6*, 493–505. (b) Challis, G. L.; Ravel, J.; Townsend, C. A. *Chem. Biol.* **2000**, *7*, 211–224. (c) Rausch, C.; Weber, T.; Kohlbacher, O.; Wohlleben, W.; Huson, D. H. *Nucleic Acids Res.* **2005**, *33*, 5799–5808. (d) Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. *Nucleic Acids Res.* **2011**, *39*, W362–367.

10.1021/ol2020237
Published on Web 09/02/2011

sequentially in the order dictated by the domain sequence in the NRPS protein(s). Some NRPS systems contain epimerization (E) or specialized condensation/epimerization (C/E) domains to introduce D-amino acids into the final product. Based on xenematide's 2-phenylacetamide-substituted cyclic tetrapeptide sequence, we initially expected a clustered five module NRPS system for the loading of a PAA starter unit and for the stepwise condensation of Thr, Trp, Trp, and β-Ala. Bioinformatic analysis of all of the NRPS genes in the *X. nematophila* 19061 genome led to a stand-alone *four* module NRPS (XNC1_2713, Accession YP_003712913) as the most likely candidate. Comparisons to the conserved domain database (CDD) indicated that the 4949 amino acid protein contained a C-A-T initiation module, two canonical C-A-T extender modules, and an atypical C-C/E-A-T-TE termination module. The termination module harbored tandem C-C/E domains with a small internal spacer sequence. Using the online tool NRPSpredictor,[5c] the adenylation domain sequences were closest to domains that accept Thr-Trp-Trp-β-Ala, as expected. An initiation module for the PAA starter unit was lacking.

To determine if NRPS XNC1_2713 was indeed responsible for xenematide biosynthesis, we knocked out the gene in *X. nematophila* by insertional inactivation. Differential metabolite analysis of organic extracts between the XNC1_2713 mutant and the wildtype cultures indicated that xenematide A production (**9**, 17.3 min) was abolished in the mutant (Figure 1A). Production of antibiotics not associated with the NRPS was largely unaffected (Figure 1AB), and the identified molecules included xenortide C (**1**, a new xenortide identified here, $C_{24}H_{34}N_3O_2$, $[M+H]^+$ at m/z 396.2642, calcd 396.2651, Table S1 and Figures S1−S5), rhabduscin (**2**),[2] **3**,[2] **5**,[6] **10**,[2] nematophin (**11**),[7] and xenocoumacin (**12**).[8] Unexpectedly, two peaks (**4**, 15.1 min and **6**−**8**, 16.8 min) more polar than xenematide A under reversed phase separation conditions were also lacking in the mutant. Careful separation of the peak at 16.8 min revealed three related products (**6**−**8**) produced only in wildtype cultures. High-resolution mass spectrometry (HR-MS) indicated that **6**−**8** had molecular formulas of $C_{33}H_{37}N_4O_6$ ($[M+H]^+$ at m/z 585.2697, calcd 585.2713), $C_{35}H_{38}N_5O_6$ ($[M+H]^+$ at m/z 624.2827, calcd 624.2822), and $C_{35}H_{38}N_5O_6$ ($[M+H]^+$ at m/z 624.2811, calcd 624.2822), respectively. One-(^{1}H) and two-(gCOSY, gHSQC, and gHMBC)-dimensional NMR of the two compounds with identical masses (**7**−**8**) indicated that both are composed of PAA, Thr, Trp, Phe, and β-Ala (Figures 2 and S6−S13, Tables S2−S3). The mass data and long-range HMBC correlations confirmed that the residues were cyclized, as in xenematide A, and MS/MS fragmentation analysis confirmed the positioning of the residues (Figures 2 and S14−S17). These results show that the two A domains that select Trp in xenematide A biosynthesis can accept either Trp or Phe to give a diversified product scaffold. Notably, this Trp/Phe promiscuity is accommodated in the downstream domains and results in a small family of cyclic depsipeptides. Low production of **6** precluded detailed NMR-based structural characterization, but HR-MS and MS/MS fragmentation analysis combined with the genetic correlation indicates that the compound contains Phe in both positions 3 and 4. HR-MS, MS/MS, and NMR analysis of the second peak at 15.1 min indicated that the major xenematide A underwent methanolysis during sample processing to the open-chain form **4** (Table S4 and Figures S18−S23). Under the conditions of the experiment, minor xenematides **6**−**8** represent 9% of the total xenematide peak area (**4**, **6**−**9**; 210 nm), indicating substantially relaxed substrate selectivity.

Xenematide A (**9**) was previously shown to contain both an L- and a D-Trp, but it was unclear which Trp in the sequence was L or D.[3] The enzyme identified here lacks an

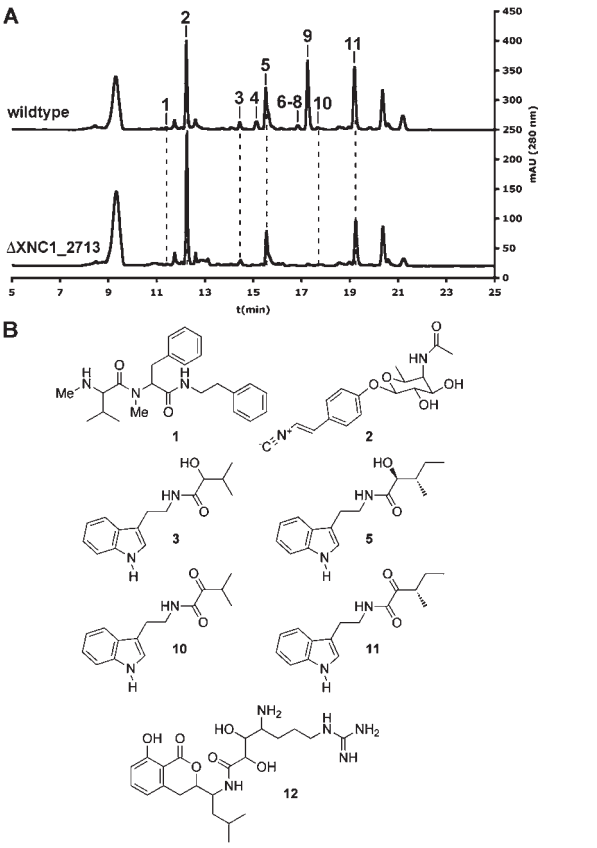


Figure 1. (A) The cyclic peptides (**4**, **6**−**9**, Figure 2) are present in wildtype but absent in the NRPS XNC1_2713 inactivant. Levels of other organic extractable metabolites were largely unaffected. (B) L-Proline upregulates production of various metabolites in the organic extract. Xenocoumacin (**12**) remains in the aqueous fraction.

(6) (a) Li, J.; Chen, G.; Webster, J. M. *J. Nat. Prod.* **1996**, *59*, 1157–1158. (b) Paik, S.; Park, M. K.; Jhun, S. H.; Park, H. K.; Lee, C. S.; Cho, B. R.; Byun, H. S.; Choe, S. B.; Suh, S. I. *Bull. Korean Chem. Soc.* **2003**, *24*, 623–626.

(7) (a) Li, J.; Chen, G.; Webster, J. M. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 1349–1352. (b) Li, J.; Chen, G.; Webster, J. M. *Can. J. Microbiol.* **1997**, *43*, 770–773.

(8) (a) McInerney, B. V.; Taylor, W. C.; Lacey, M. J.; Akhurst, R. J.; Gregson, R. P. *J. Nat. Prod.* **1991**, *54*, 785–795. (b) Reimer, D.; Luxenburger, E.; Brachmann, A. O.; Bode, H. B. *ChemBioChem* **2009**, *10*, 1997–2001.

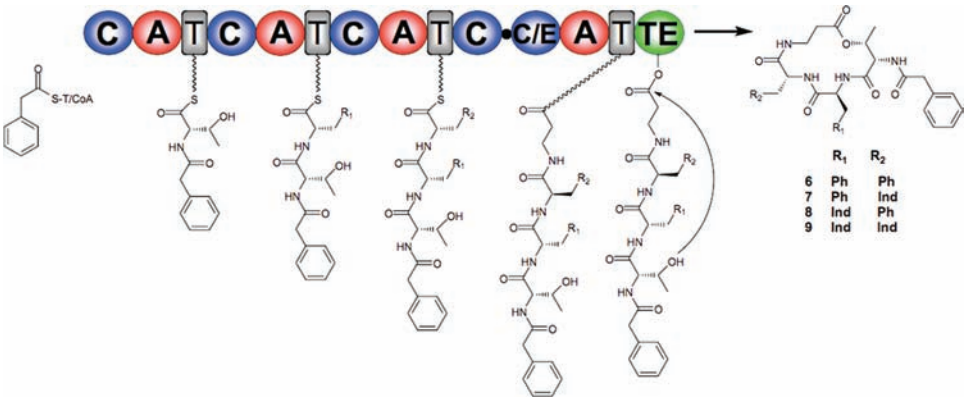


Figure 2. Proposed xenematide biosynthesis. Xenematide A (**9**) and several new xenematide derivatives B (**6**), C (**7**), and D (**8**) are shown. Compound **4** is the open chain methanolysis product of xenematide A, which is not shown. Ph, phenyl; Ind, indole.

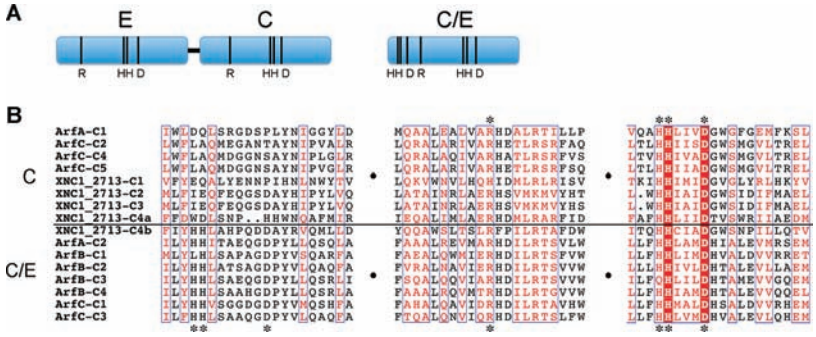


Figure 3. (A) Key active site residue organization for E, C, and C/E domains. (B) Alignment of active site regions from C and C/E domains in xenematide and arthrofactin biosynthesis. Active site residues are marked with asterisks.

obvious epimerization (E) domain to flip the stereochemistry from L to D in either module. Downstream C domains have sequence-based stereospecificity signatures that facilitate the prediction of product stereochemistry.[9] Additionally, specialized C/E domains, particularly in Gram-negative bacteria, have been shown to both epimerize the enzyme-bound substrate and subsequently condense the D-product during chain assembly.[10] Conserved domain database analysis of the termination module 4 indicated tandem C domains ($C4_a$ and $C4_b$) with $C4_b$ containing the sequence signature for dual C/E activities (Figures 3 and S24). Recently both possible xenematide diastereomers were synthesized, and spectroscopic comparison confirmed that D-Trp occupies position 4.[11]

While the stereochemistry of xenematide A was resolved by synthesis, the stereochemistries of the new xenematides were unknown. Peptides (**4**, **6**–**9**) were subjected to acid hydrolysis and advanced Marfey's analysis.[12] All peptides were found to incorporate L-Thr. Xenematide A (**9**) and its methanolysis product **4** contained L- and D-Trp as expected. Xenematide B (**6**) contained L- and D-Phe, and based on the positions of D/L-residues in xenematide A (**9**) and the bioinformatic analysis, it is likely that D-Phe occupies position 4. Since xenematides C and D (**7** and **8**) consist of a unique selection of amino acids, Marfey's analysis alone was sufficient to determine stereochemical assignments. Irrespective of whether a Trp or Phe occupied position 4, the amino acid in this position has the D-configuration.

These bioinformatic, genetic, and structural results highlight the unusual aspects of the xenematide biosynthetic pathway: an absent initiation module, A domain Trp/Phe promiscuity and its toleration in subsequent steps, and the role of the tandem C/CE domains in module 4 (Figure 2). Unlike canonical initiation modules that begin with A-T,

(9) (a) von Dohren, H.; Dieckmann, R.; Pavela-Vrancic, M. *Chem. Biol.* **1999**, *6*, R273–279. (b) Clugston, S. L.; Sieber, S. A.; Marahiel, M. A.; Walsh, C. T. *Biochemistry* **2003**, *42*, 12095–12104.

(10) Balibar, C. J.; Vaillancourt, F. H.; Walsh, C. T. *Chem. Biol.* **2005**, *12*, 1189–1200.

(11) Hung, K. Y.; Harris, P. W.; Heapy, A. M.; Brimble, M. A. *Org. Biomol. Chem.* **2011**, *9*, 236–242.

(12) Fujii, K.; Ikai, Y.; Oka, H.; Suzuki, M.; Harada, K. *Anal. Chem.* **1997**, *69*, 5146–5151.

NRPS XNC1_2713 begins with a C-A-T module. This starter module arrangement can be found in NRPS products that are N-acylated, which is consistent with xenematide's 2-phenylacetamide substituent. Free standing A domains are known to activate aromatic carboxylic acids and transfer them to aryl-carrier proteins (ArCP), which are used as starter substrates in, for example, quinoxaline antibiotics.[13] Since no free-standing A domain and/or an ArCP are clustered with the XNC1_2713 gene, the potential starter proteins are encoded elsewhere in the genome or phenylacetyl-CoA is used as a starter unit directly. Module 1 condenses the PAA starter unit with an L-Thr. Modules 2 and 3 would variably condense L-Trp and L-Phe with a preference for L-Trp. Similarly, select modules in the tyrocidine biosynthetic pathway that activate Phe or Tyr in tyrocidine A can also activate Trp to produce tyrocidines B−D, and the relative amount of each tyrocidine is dependent upon amino acid concentration in the culture medium.[14] Based on literature precedent, it is not obvious how the domain arrangement in module 4 could catalyze condensation of a β-Ala unit (position 5) while epimerizing the upstream position 4 (L-Trp or L-Phe). Specialized C/E domains are capable of upstream epimerization as indicated by the arthrofactin, syringopeptin, and ramoplanin biosynthetic proteins.[10] The XNC1_2713 C domains were aligned with the C domains involved in the biosynthesis of the potent cyclic lipopeptide biosurfactant arthrofactin (Figure 3),[15] and $C4_a$ appears to be a typical C domain whereas $C4_b$ has the C/E sequence signature. In one recent case, multiple adjacent C domains were proposed to condense different lengths of a poly-Lys polymer (variable substrates) in poly-Lys biosynthesis.[16] However, with the D-configuration in xenematide position 4 and the C/E sequence activity only in $C4_b$, it is unlikely that both domains redundantly catalyze condensation. It is possible that $C4_a$ is nonfunctional, but sequence analysis suggests an intact domain with active site residues conserved. An intriguing possibility is that $C4_a$ catalyzes condensation while the apparent $C4_b$ domain is a dedicated epimerase, with both exhibiting relaxed substrate specificity to accommodate upstream adenylation domain promiscuity. The last noncanonical residue, β-Ala, serves as a spacer unit that results in a 14-member rather than a 13-member macrolactone that would arise from α-amino acids such as Gly or Ala. Finally, the terminal thioesterase (TE) domain would catalyze O−C macrocyclization and xenematide product release.

C domains exhibit sequence and structural homology to E, C/E, and cyclization (Cy) domains.[17] Cy domains typically catalyze condensation, cyclization, and dehydration to introduce thiazoline or oxazoline rings into the peptide product. In vibriobactin biosynthesis, tandem Cy domains in VibF decouple the condensation and heterocyclization activities of standard Cy domains, and both domains are required for vibriobactin synthesis.[18] It has been proposed that E domains evolved from C domains by domain duplication and divergent evolution, and that C/E domains could have been an E domain precursor.[10] If, as suggested above, the C and E activities were decoupled in xenematide biosynthesis, the C−C/E pair of XNC1_2713 would provide a functional link supporting this evolutionary hypothesis.

The peculiar features of xenematide biosynthesis are not limited to *X. nematophila*, as close protein homologs can be identified in the related *Xenorhabdus bovienii* (*id*/*sim*/*gap*: 65/77/3%) and even in *Photorhabdus asymbiotica* (79/87/1%), which forms a mutualistic symbiosis with *Heterorhabditis* nematodes that can infect both humans and insects.[19] The conservation of this NRPS pathway across bacterial genera that infect both human and invertebrate hosts suggests a selective advantage for maintaining its ability to produce xenematides. Unlike primary metabolism that follows a "one enzyme, one reaction, one product" dogma due to large selective pressures for energy conservation and survival, secondary metabolism frequently features pathways that produce multiple related products,[20] and this biosynthetic flexibility must represent some sort of an evolutionary advantage. The variety of ways in which these naturally occurring pathways diversify their output also reflects the general importance of the strategy, and cases from promiscuous processing of the products of modular pathways are well-known. The xenematide diversification strategy—A domain promiscuity and downstream tolerance—represents another implementation of the same overall goal.

Acknowledgment. This work was supported by the U.S. National Institutes of Health (Grant R01 GM086258 to J.C.) and the New England Regional Center of Excellence for Biodefense and Emerging Infectious Diseases (Grant U54 AI057159 to J.C.). J.M.C. is supported by a U.S. National Institutes of Health Pathway to Independence award (Grant 1K99 GM097096-01). We thank John Dickson at Harvard Medical School for analyzing xenocoumacin production during his lab rotation.

Supporting Information Available. Experimental section and NMR and MS/MS spectra of new compounds **1**, **4**, **6**–**8**. This material is available free of charge via the Internet at http://pubs.acs.org.

(13) Schmoock, G.; Pfennig, F.; Jewiarz, J.; Schlumbohm, W.; Laubinger, W.; Schauwecker, F.; Keller, U. *J. Biol. Chem.* **2005**, *280*, 4339–4349.

(14) Mootz, H. D.; Marahiel, M. A. *J. Bacteriol.* **1997**, *179*, 6843–6850.

(15) Roongsawang, N.; Hase, K.; Haruki, M.; Imanaka, T.; Morikawa, M.; Kanaya, S. *Chem. Biol.* **2003**, *10*, 869–880.

(16) Yamanaka, K.; Maruyama, C.; Takagi, H.; Hamano, Y. *Nat. Chem. Biol.* **2008**, *4*, 766–772.

(17) Keating, T. A.; Marshall, C. G.; Walsh, C. T.; Keating, A. E. *Nat. Struct. Biol.* **2002**, *9*, 522–526.

(18) Marshall, C. G.; Hillson, N. J.; Walsh, C. T. *Biochemistry* **2002**, *41*, 244–250.

(19) Waterfield, N. R.; Ciche, T.; Clarke, D. *Annu. Rev. Microbiol.* **2009**, *63*, 557–574.

(20) (a) Clardy, J.; Walsh, C. *Nature* **2004**, *432*, 829–837. (b) Clardy, J.; Fischbach, M. A.; Walsh, C. T. *Nat. Biotechnol.* **2006**, *24*, 1541–1550. (c) Fischbach, M. A.; Clardy, J. *Nat. Chem. Biol.* **2007**, *3*, 353–355.