Insights into Heterocyclization from Two Highly Similar Enzymes

John A. McIntosh, Mohamed S. Donia, and Eric W. Schmidt*

Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112

Received December 30, 2009; E-mail: ews1@utah.edu

A problem for chemists in the metagenomic era is how to approach the vast amounts of data coming from metagenome sequences. Metagenomes are mixed samples of multiple organisms from the environment, including symbiotic associations of bacteria with animals.[1−4] We have been developing the use of metagenomes from symbiotic bacteria for pathway engineering, allowing natural sequence variation to define parameters for gene-based organic synthesis.[5,6] Here, we provide an example of how metagenome sequence analysis contributes to understanding biosynthetic enzyme function using the thiazoline and oxazoline synthases, PatD and TruD. These proteins in turn are useful in the synthesis of diverse natural product libraries.

The first natural products pathways identified and characterized by sequencing of metagenomic DNA and comparison are the patellamide (*pat*) and trunkamide (*tru*) pathways to the cyanobactins, ribosomal peptides that are extensively posttranslationally modified (Figure 1A).[5,7] We previously used a metagenome sequence analysis method to determine that enzymes from *pat* and *tru* are extremely broad-substrate selective and can be used for *in vivo* synthesis of diverse natural and unnatural cyanobactin derivatives. Essentially identical (>99%) enzymes within the *tru* or *pat* pathway accept numerous substrates. Although enzymes are identical within pathways, there are differences between *pat* and *tru*. In particular, there are enzyme domains that are virtually identical between pathways, and there are more divergent domains. These changes have functional consequences that can be precisely tracked (Figure 1).

Both *pat* and *tru* were initially cloned from bacteria living symbiotically with marine animals, the ascidians. Over 60 cyanobactins, including the well-known metabolites trunkamide and patellamide C,[8,9] have been isolated from ascidians; sequencing of metagenomic DNA has identified more than 30 additional relatives.[6] Cyanobactins are ribosomal peptide natural products, and comparison of metagenomic sequencing data with known compounds provided a natural mutagenesis in which evolution has mutated every amino acid in the natural products while maintaining identical modifying enzymes.[5] The ribosomal peptides are encoded on precursor peptides with length of ∼70 amino acids. Precursors include leader sequences (∼35 amino acids) and two sets each of recognition and product-coding sequences.

Despite their demonstrated usefulness in synthesizing diverse compounds, the biosynthetic rules behind posttranslational modification patterns remained elusive. For example, products of the *pat* pathway contain Cys, Ser, and Thr residues that are mostly heterocyclic. By contrast, known representatives of the *tru* pathway are heterocyclic at Cys but reverse prenylated at Ser and Thr (Figure 1).

Previous work has shown that essentially identical enzymes within the *pat* or *tru* pathways lead to these diverse products.[5,6,10] The enzymes modify precursor peptides that have conserved leader sequences and enzyme recognition sequences. Between the conserved recognition sequences are cassettes that directly encode the
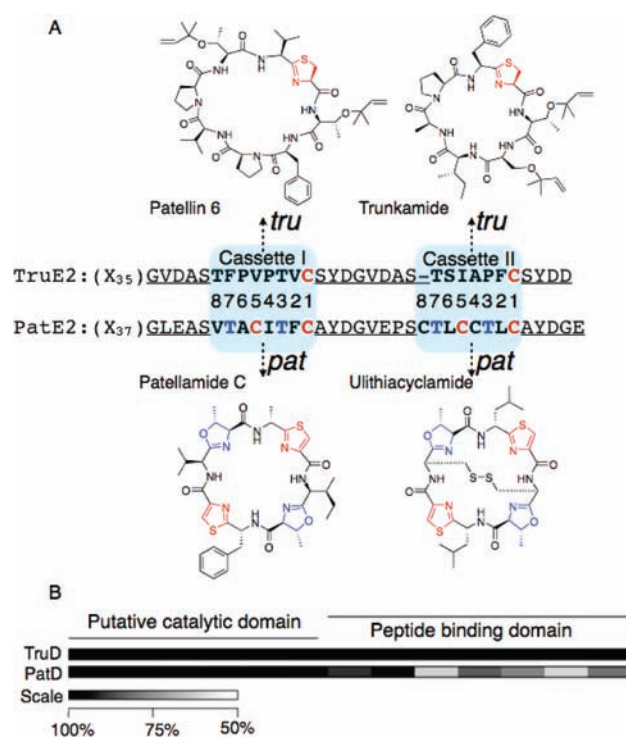

Figure 1. Biosynthesis of heterocyclic cyanobactins. (A) Enzymes modify ribosomal precursor peptides into mature products. Trunkamide group: top; patellamide group: bottom. (B) An alignment of PatD and TruD proteins, showing putative catalytic and peptide-binding domains. Catalytic domains are >99% identical; peptide-binding domains are >77% identical.

final natural products. These cassettes are modified by heterocyclization, prenylation, and N−C circularization to yield the mature natural products. Cassette sequences are hypervariable, with substitutions accepted at any amino acid position, leading to large natural product libraries.

Six, seven, or eight amino acids are found in ascidian cyanobactins. Numbering from C- to N-terminus, all positions except position 2 naturally contain the heterocyclizable residues, Cys, Ser, or Thr, in one or more of the 60 known cyanobactins. The *pat* group includes members in which heterocycles occur in positions 1, 3, 5, 6, and 7 (Figure 2). By contrast, cycles are only found in position 1 in the *tru* group. In *tru* products, only Cys is heterocyclized, while Ser and Thr are prenylated; Cys, Ser, or Thr are cyclic in the *pat* products, and prenylation is not found. In all of these peptides, there are numerous Ser residues outsitde of the cassettes that are not modified.

These natural substitution patterns are particularly striking given the overall enzyme conservation. Despite chemical differences between *pat* and *tru*, all enzymes are homologous between these pathways. In fact, the didomain heterocyclase enzymes, PatD and TruD, are >99% identical in their N-terminal domains and only
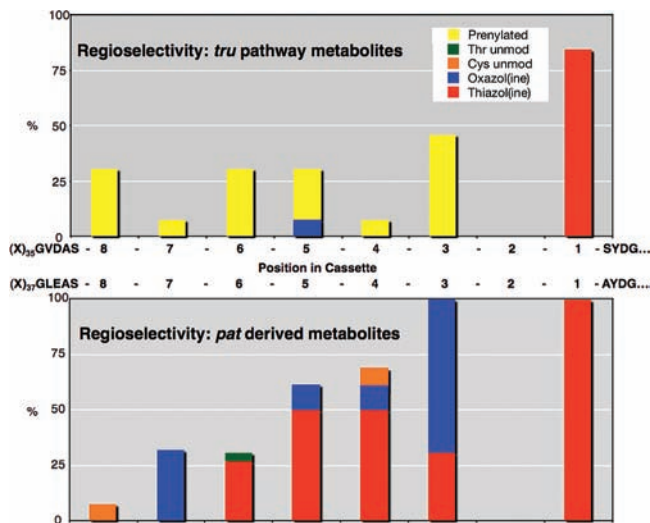
Figure 2. Enzyme selectivity defined by metagenome sequence and chemical analysis. Natural products from the *pat* (top) and *tru* (bottom) pathways contain modified Cys, Ser, and Thr residues in the defined positions. *y*-axis: % of natural products containing these modifications. *x*-axis: amino acid position in cassette. Empty space above the bars denotes hypervariability: diverse amino acids occupy these positions in isolated natural products.

Figure 3. Reaction characterization and substrate specificity of TruD and PatD. (A) MALDI MS of reaction products. (B) Position of modifications identified by ESI-MS/MS. (C) Representative SDS-PAGE gel showing band shifts of precursor peptides 1−4. The left of each triplet is a standard of the peptide, middle is the TruD-modified peptide, and right is the PatD-modified peptide.

~77% identical in their C-terminal domains. The questions addressed in this study are how such similar enzymes and pathways lead to different posttranslational products and how these enzymes can be used to synthesize diverse chemical derivatives.

Microcin B17 and streptolysin S synthetases were the first and second heterocyclase enzymes to be characterized *in vitro*.[11−13] Although they are only distantly related to PatD and TruD, comparison of these enzymes showed that the PatD/TruD catalytic domain is at the N-terminus, while the C-terminus functions primarily to bind the substrate peptide. This led to the question, do these enzymes operate in a chemoselective fashion (*O* vs *S* in heterocycles), or is regioselectivity important, as might be implicated by peptide binding differences? Here, we use biochemical experiments to define PatD and TruD as regioselective heterocyclases, which catalyze thiazoline and oxazoline biosynthesis.

When the *pat* gene cluster encoding PatD was previously expressed in *Escherichia coli*, we observed products containing the natural heterocycle pattern, including thiazoline and oxazoline.[6,7] Similarly, heterologous expression of the *tru* pathway led to production of thiazoline-containing natural products with prenylated Ser/Thr residues.[5] Thus, PatD and TruD were implicated as probable heterocyclases, but they had not been purified or characterized. Both genes were cloned and expressed here as N-terminal His-tagged proteins (Figure S1, Supporting Information). The *patD* clone includes two point mutations in the N-terminal domain, which did not change enzyme function. The *truD* gene was cloned into the C-terminus of an existing *patD* construct because of toxicity problems. Practically, this cloning strategy ensured that PatD and TruD were in fact 100% identical (instead of merely >99%) in their presumed catalytic N-terminal domains. A series of substrates were constructed by cloning and expressed recombinantly, which was extremely efficient in comparison to peptide synthesis because of the large size of the substrates (~70 amino acids).

Purified PatD or TruD were used in experiments with the substrate analogues, PatEdm, PatEα, TruE2, TruE4, and TruE5 (Figure 3). All enzyme experiments in this study were done with varying substrate concentrations, with each condition performed independently at least in triplicate. We first noticed that PatD and
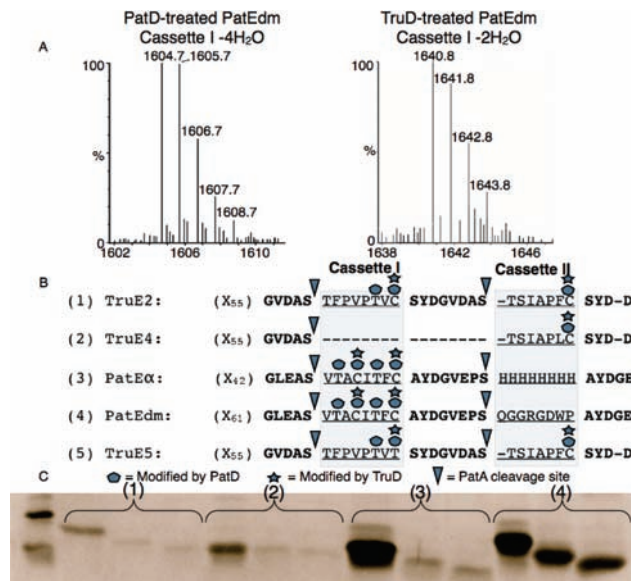
TruD products exhibited band shift differences by SDS-PAGE. In all cases, TruD products migrated more rapidly than unmodified peptides, while PatD products migrated more rapidly still (Figure 3). The mobility shift was consistent with formation of thiazoline and/or oxazoline rings given that these heterocycles introduce a significant conformational restraint on the peptide backbone.[14−17]

Given that heterocycle formation causes a loss of 18 Da, we employed an MS approach to confirm that each of the five precursor peptides was modified by heterocyclization. First, intact mass electrospray ionization (ESI) was used to determine the total number of dehydrations catalyzed by PatD or TruD on these ~9 kDa substrates, in comparison to unmodified control (Figure S2, Supporting Information). Subsequently, the enzyme products were treated by a specific protease, PatA, that cleaves upstream from each cassette's start site.[10] These smaller fragments were analyzed by MALDI and ESI to localize dehydrations to single cassettes (Figures S3 and S8, Supporting Information). Finally, these PatA-digested cassettes were also subjected to LC-Fourier transform ion cyclotron resonance (FT-ICR) and MS/MS to determine which amino acids within cassettes were heterocyclized (Figures S4 and S5, Supporting Information). For Cys, MS data confirmed that thiazoline was formed and ruled out other possible dehydration routes. In addition, for Thr, all available evidence supported oxazoline formation. This evidence included observation of the same type of fragmentation suppression seen in other heterocycle-containing peptides,[18] the absence of observed MS/MS fragments consistent with other modifications, and the SDS-PAGE mobility shift, which as noted above, is consistent with heterocycle formation. Additionally, the facts that these genes lead to oxazoline formation *in vivo* and that no other type of Thr dehydration has been observed in this compound family also support this interpretation. Despite this evidence, we nevertheless sought to completely eliminate the possibility of a reverse-Michael reaction (Figure S6, Supporting Information). Because activated double bonds are not very reactive with acids, while oxazolines are labile in acidic conditions, substrates were subjected to very mild acidic conditions. The resulting rehydrated products confirmed that all Thr modifications

were indeed due to oxazolines (Figure S6, Supporting Information). Therefore, PatD and TruD catalyzed the synthesis of thiazoline and oxazoline; other products were not observed in extensive experimental analysis.

TruE2 is a natural substrate that contains two Cys residues that are found as thiazoline in the final natural products, as well as a number of Ser and Thr residues that are prenylated naturally (Figure 1A). TruE2 contains two cassettes, with Cys in position 1 of each cassette. When the *tru* cluster was expressed in *E. coli*, these natural products were synthesized,[5] indicating that *tru* gene products modify TruE2 to produce thiazoline. When treated with TruD, both Cys residues were heterocyclized as expected, while none of the Ser/Thr residues in the molecule were modified. By contrast, when treated with PatD, which is not normally associated with TruE2, both Cys residues and an additional Thr residue in position 3 of cassette I were cyclized (Figure 3). An unnatural analog of TruE2, TruE4, contained only cassette II and only a single Cys residue; this residue was also cyclized by both PatD and TruD.

We next analyzed reactivity using unnatural substrate analogs from the *pat* pathway. In cassette I, both PatEdm and PatEα encode the natural patellamide C sequence which would normally be modified to contain two thiazole and two oxazoline residues (Figure 1). Indeed, patellamide C is synthesized in *E. coli* when PatEdm is coexpressed with *pat* enzymes.[6,7] When these peptide substrates were treated with PatD, two Cys and two Thr residues were cyclized in both PatEdm and PatEα, as found *in vivo* in *E. coli* expression. TruD, which is not normally associated with the *pat* pathway, behaves differently. None of the *tru* products we have so far examined contain Cys in position 5, as found in PatEdm. However, TruD readily cyclized both Cys residues in position 1 and position 5 (Figure 3).

From these experiments, we could not determine whether TruD was truly chemoselective for Cys or whether regioselectivity played a role. We therefore synthesized an unnatural variant, TruE5, that contained Thr in place of Cys at position 1 of cassette I. Both PatD and TruD were able to cyclize this new Thr residue, indicating that the reaction specificity of these enzymes is mainly due to regioselectivity (Figure S7, Supporting Information), though this reaction was much slower than Cys heterocyclization. Thus, although the enzymes select residues for modification based primarily on their position within cassettes, the chemical features of the modified residue may influence selectivity as well.

The regioselectivity of these enzymes clearly explains the observed product patterns in the ∼60 ascidian-derived cyanobactins. In fact, previously the cyanobactin comoramide A was isolated with Thr heterocyclized in position 5 and prenylated in position 3.[19] Although genes for comoramide synthesis have not been cloned, the data described here allow this to be defined as a *tru*-like pathway. Another pathway type contains a prenylated Ser in position 5. In TruE2, PatD did not modify a Ser in position 6.

These experiments also indicate that TruD and PatD do not determine the regioselectivity of prenylation, since TruD leaves unmodified the residues that are presumably later prenylated. Although there are no prenyltransferase homologues in the pathways, our current hypothesis is that the TruF1 and TruF2 proteins are involved in this step.

The usefulness of these enzymes is primarily for *in vivo* synthesis of new compounds, not necessarily for the *in vitro* modification of discrete substrates. Previously, we demonstrated that, in principle, large libraries of natural and unnatural cyanobactin derivatives could be synthesized and screened in *E. coli*.[6] The enzymatic specificity results reported here will be greatly helpful in determining which sequences are appropriate for the development of a chemically diverse cyanobactin library. Metagenome sequence analysis enabled the discovery of methods to synthesize these libraries, and in this study the sequencing methods allowed us to obtain new insights into enzyme function. These methods should be applicable to other enzyme systems. Because the underlying genetic methods are now extremely fast and inexpensive, they are of practical utility in the enzymatic synthesis of new molecules.

Supporting Information Available: Additional mass spectrometry data including FT-ICR, MS-MS, ESI-MS, and MALDI data, and SDS-PAGE analyses. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Stein, J. L.; Marshall, T. L.; Wu, K. Y.; Shizuya, H.; DeLong, E. F. *J. Bacteriol.* **1996**, *178*, 591–599.
(2) Haygood, M. G.; Davidson, S. K. *Appl. Environ. Microbiol.* **1997**, *63*, 4612–4616.
(3) Schloss, P. D.; Handelsman, J. *Curr. Opin. Biotechnol.* **2003**, *14*, 303–310.
(4) Handelsman, J.; Rondon, M. R.; Brady, S. F.; Clardy, J.; Goodman, R. M. *Chem. Biol.* **1998**, *5*, R245–R249.
(5) Donia, M. S.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2008**, *4*, 341–343.
(6) 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.
(7) 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.
(8) Ireland, C. M.; Durso, A. R.; Newman, R. A.; Hacker, M. P. *J. Org. Chem.* **1982**, *47*, 1807–1811.
(9) Carroll, A. R.; Coll, J. C.; Bourne, D. J.; MacLeod, J. K.; Zabriskie, T.; Ireland, C. M.; Bowden, B. F. *Aust. J. Chem.* **1996**, *49*, 659–667.
(10) Lee, J.; McIntosh, J. A.; Hathaway, B. J.; Schmidt, E. W. *J. Am. Chem. Soc.* **2009**, *131*, 2122–2124.
(11) Lee, S. W.; Mitchell, D. A.; Markley, A. L.; Hensler, M. E.; Gonzalez, D.; Wohlrab, A.; Dorrestein, P. C.; Nizet, V.; Dixon, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5879–5884.
(12) Milne, J. C.; Roy, R. S.; Eliot, A. C.; Kelleher, N. L.; Wokhlu, A.; Nickels, B.; Walsh, C. T. *Biochemistry* **1999**, *38*, 4768–4781.
(13) Li, Y. M.; Milne, J. C.; Madison, L. L.; Kolter, R.; Walsh, C. T. *Science* **1996**, *274*, 1188–1193.
(14) Milne, B. F.; Long, P. F.; Starcevic, A.; Hranueli, D.; Jaspars, M. *Org. Biomol. Chem.* **2006**, *4*, 631–638.
(15) Abbenante, G.; Fairlie, D. P.; Gahan, L. R.; Hanson, G. R.; Piersens, G. K.; van den Brenk, A. L. *J. Am. Chem. Soc.* **1996**, *118*, 10382–10388.
(16) Bernhardt, P. V.; Comba, P.; Fairlie, D. P.; Gahan, L. R.; Hanson, G. R.; Lotzbeyer, L. *Chemistry* **2002**, *8*, 1527–1536.
(17) Walsh, C. T.; Nolan, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5655–5656.
(18) Belshaw, P. J.; Roy, R. S.; Kelleher, N. L.; Walsh, C. T. *Chem. Biol.* **1998**, *5*, 373–384.
(19) Rudi, A.; Aknin, M.; Gaydou, E. M.; Kashman, Y. *Tetrahedron* **1998**, *54*, 13203–13210.

JA9107116