JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

# Ribosomal Route to Small-Molecule Diversity

Ma. Diarey B. Tianero,[†] Mohamed S. Donia,[†,§] Travis S. Young,[‡,⊥] Peter G. Schultz,[‡] and Eric W. Schmidt*[,†]

[†]Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112, United States

[‡]The Scripps Research Institute, La Jolla, California 92037, United States

Supporting Information

**ABSTRACT:** The cyanobactin ribosomal peptide (RP) natural product pathway was manipulated to incorporate multiple tandem mutations and non-proteinogenic amino acids, using eight heterologous components simultaneously expressed in *Escherichia coli*. These studies reveal the potential of RPs for the rational synthesis of complex, new small molecules over multiple-step biosynthetic pathways using simple genetic engineering.


## INTRODUCTION

A major goal of synthetic biology is to engineer the synthesis of organic compounds in vivo.[1] Ribosomal peptide (RP) natural products provide a relatively simple starting point for such studies. In RP pathways, short precursor peptides are ribosomally translated and subsequently modified by enzymes into complex natural products.[2] The products of RP pathways are often elaborately tailored, so that in extreme cases (such as pyrroloquinoline quinone),[3] the compounds are no longer easily recognizable as arising from amino acids. A further advantage is that RP biosynthetic enzymes generally exhibit relaxed substrate specificity.[2,4] Because of this, often sequence-identical enzymes are capable of accepting a vast array of peptide substrates.[5,6] Here, we seek to develop methods for making unnatural RPs and to better understand substrate selectivity of RP pathways. In so doing, we hope to enable synthetic biological approaches to the directed creation of diverse structures using ribosomally translated starting materials. In addition to these design and synthesis goals, RPs are often potently bioactive against diverse targets, and thus understanding and manipulating individual RP pathways is a major goal in its own right.

In Nature, RPs occupy many different chemical classes, with many different posttranslational modifications.[2] These classes have been examined by observing natural variation, as well as by manipulating biosynthetic genes and by performing biochemical and substrate analyses. Observation of Nature indicates that at least several pathways handle extremely diverse substitutions. For example, >60 cyanobactin genes have been isolated from symbiotic bacteria living in marine animals, where the enzymes are essentially identical yet the products are extremely sequence-diverse.[5] RP biosynthetic genes have already been modified in several ways, most of which have focused on point mutations of pathways that require one or two enzymatic steps for maturation.[7−10] Single or multiple mutations have been introduced into lantibiotics, and more recently, multiple tandem substitutions have been introduced into lasso peptides.[12] These mutations are usually aimed either to improve the known biological activity, or occasionally even to design a new biological activity.[13] A few mutational studies have focused on complex pathways with multiple biosynthetic steps.[5,14,15] For example, the integrin-binding RGD motif has been introduced into cyclic peptides using cyanobactin pathway enzymes[5] and lasso peptide enzymes.[13] However, overall the field is still distant from achieving the goal of wholesale introduction of new motifs at will, especially in the more complex, multistep pathways.

A further goal in this field has been to incorporate motifs other than the canonical 20 amino acids into the backbone of RPs. Since the posttranslational enzymes exhibit extremely broad substrate tolerance, such changes would greatly extend the capabilities of RP pathways in synthetic biology. In many cases, synthetic substrates have been fed to RP enzymes in vitro. A potentially more powerful approach from the synthetic biology point of view is to simply incorporate non-proteinogenic amino acids in vivo. Indeed, non-proteinogenic amino acids have been added to RPs in vivo, although only in single-step processes involving lanthionine synthase or inteins.[16,17] Incorporation of such "unnatural" amino acids is considered desirable because it provides unprecedented control over chemical structure in vivo.

Here, we sought to explore the mutation and addition of non-proteinogenic amino acids with a much more complex

**Received:** September 1, 2011
**Published:** November 22, 2011

   dx.doi.org/10.1021/ja208278k | J. Am. Chem.Soc. 2012, 134, 418−425

pathway to cyanobactins, which are highly modified, macrocyclic RPs.[18,19] Cyanobactins usually have four types of posttranslational modifications, with up to eight individual modified amino acids, encompassing up to 18 separate biochemical transformations (Figure 1). Heterocylization and

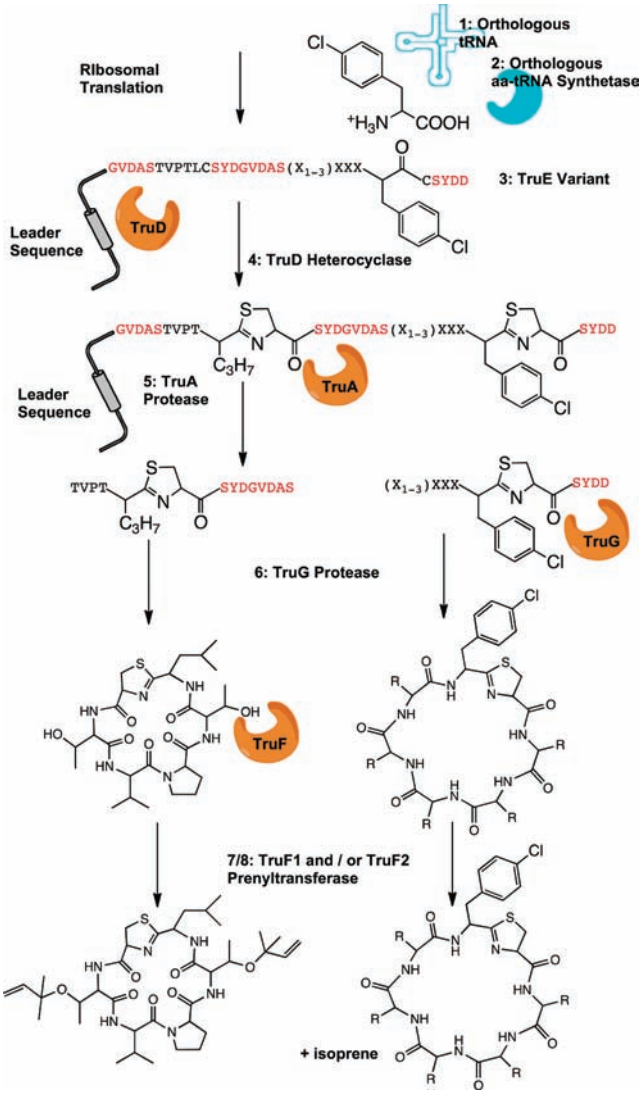


**Figure 1.** Biosynthesis of *tru* pathway derivatives in *Escherichia coli*. Numbered components indicate eight individual elements including enzymes, tRNA, or precursor peptide (TruE variants) that are required for synthesis of mature natural products. The orthologous tRNA and amino acyl tRNA synthetase components required to introduce non-proteinogenic amino acids are shown in blue. Enzymes from the *tru* pathway are indicated in orange. The precursor peptide TruE is shown with a helical, 35-amino acid leader sequence indicated schematically.[11] X indicates any amino acid.

oxidation, N- and C-terminal proteolysis, macrocyclization, and prenylation lead to final natural products.[5,6,20−24] In the *pat* and *tru* cyanobactin pathways, core sequences encoding six- to eight-amino acid cyclic peptides are contained in defined cassettes in a precursor peptide (Figures 2 and 3).[5,6] The 18−24 nucleotide hypervariable regions are flanked by invariant sequences that direct enzymatic modification.[5] Thus, the synthesis of new compounds requires only simple genetic engineering.[6]

In the *tru* pathway, originally cloned from uncultivated symbiotic bacteria in a marine animal, a ∼70 amino acid precursor peptide (TruE) is synthesized that encodes two cyanobactins (Figure 1).[6] Heterocyclase TruD acts on Cys residues in specific positions, synthesizing thiazoline.[23] Subsequently, TruA protease removes the leader sequence and affords two free N-termini for macrocyclization.[22] TruG protease cleaves C-terminal recognition sequences in tandem with macrocyclization to provide small, cyclic peptides.[20,22,23] Finally, TruF1/TruF2 decorate the cyclic peptides with isoprene on specific Ser and Thr residues.[24] In this manner, we have previously successfully produced coral reef-derived marine natural products patellin 2 (**1**), trunkamide (**2**) and patellin 3 (**3**) by heterologous expression in *E. coli*. Optimal production of wild type compounds **1**−**3** requires five days, emphasizing the complexity of the process.[25]

Some ascidian cyanobactins, and especially **2**, are considered as potential anticancer leads.[26,27] Cyanobactins in the patellamide class are potent inhibitors of drug efflux.[28,29] Therefore, synthesis of derivatives in the manner described herein is expected to lead to bioactive molecules that are amenable to drug development. More importantly, as shown by the integration of RGD motifs into cyanobactins, we are interested in exploring the mutability of these pathways because they can be used to install desired or random features into natural product-like structures, using simple genetic engineering.

## RESULTS

**Cloning and Expression Strategy.** The pUC-based vector ptru-SD1 (Symbion Discovery, Inc.) encodes *tru* biosynthetic enzymes and a copy of the precursor peptide TruE1, encoding patellins 2 (**1**) and 3 (**3**), under control of the *lac* promoter (Figure 2). To encode new cyanobactin derivatives, we constructed a second vector (ptruE), which is compatible with ptru-SD1 and can be used for coexpression experiments. ptruE contains only the *truE* gene, under control of the *lac* promoter. The first cassette of ptruE encodes **3**, while the second cassette can be varied to synthesize novel compounds (Figures 1 and 2, Tables 1 and 2). In this system, compounds **1** and **3** are always synthesized in *E. coli* and serve as internal controls to show that the *tru* pathway is functional. In addition, because both plasmids are under control of the constitutive *lac* promoter, no induction is necessary, and optimum production requires five days of fermentation at 30 °C. The general system was previously optimized, and we showed that addition of inducers or repressors serve to decrease yields.[25] Thus, after five days without induction, internal controls **1** and **3** are produced, and possibly new derivatives encoded on ptruE.

In this study, the sequence in the second cassette of ptruE was manipulated using cloning, while the first cassette was kept constant to produce positive control **1**. The fidelity of this process was determined by gene sequencing, where the ptruE derivatives were Sanger sequenced. Thus, each *E. coli* expression clone encoded all necessary *tru* enzymes, internal standards leading to production of **1** and **3**, and a gene for one of the new compounds **4**−**22**. In expression experiments, production of **1** and **3** would indicate that enzymes were functional and active and would also provide internal calibration for yield determination, while new compounds would be expressed only if the sequences were substrates for all *tru* enzymes.


dx.doi.org/10.1021/ja208278k | J. Am. Chem.Soc. 2012, 134, 418−425

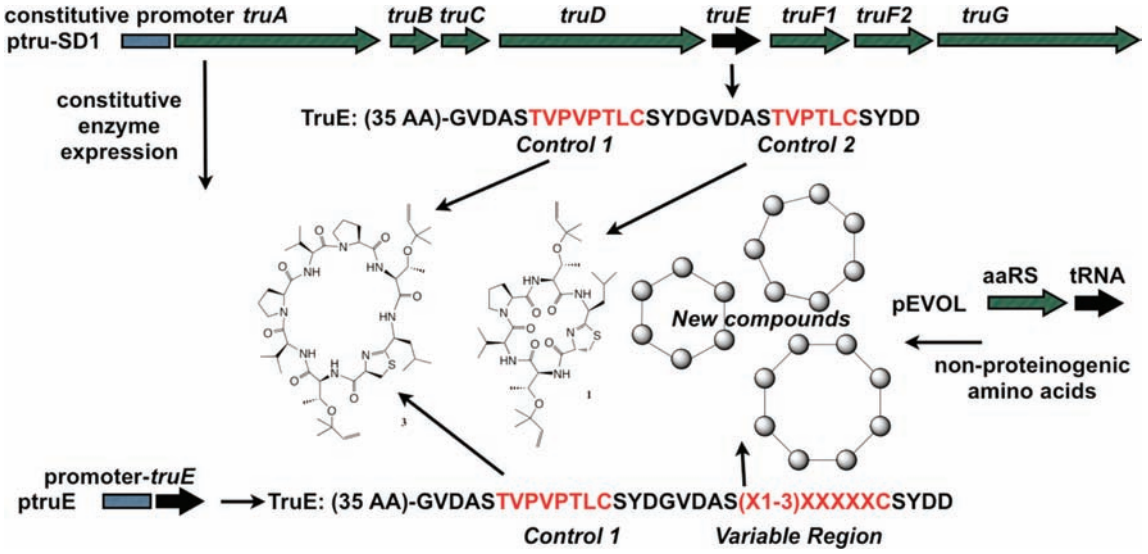


**Figure 2.** Expression strategy. Enzymes and two control molecules are synthesized constitutively from the vector ptru-SD1 (top). An internal control sequence and a variable region, that can lead to peptide libraries, is synthesized from ptruE (bottom). Optionally, non-proteinogenic amino acids can be included using a third vector, pEVOL (right).

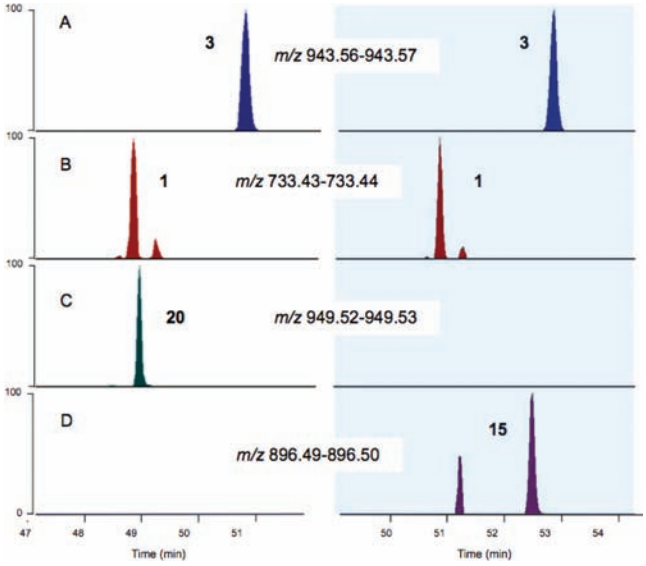


**Figure 3.** Expression of mutant cyanobactins. *E. coli* cultures expressing **20** (left) and **15** (right) were harvested and analyzed by FT-MS. Extracted ion chromatograms showed that each culture expressed control compounds **3** (A) and **1** (B). In cells containing plasmids encoding **20**, compound **20** could be detected (C), but not **15** (D), nor any other cyanobactin **4**–**22**. Similarly, in cells encoding **15**, only compound **15** (D) but not **20** (C), nor any other cyanobactin, could be detected. Thus, each experiment contained two internal positive controls and 19 external negative controls.

We constructed mutants based upon known natural products, compounds **1**-**3**, which we previously identified in uncultivated symbiotic bacteria living in marine animals and expressed successfully in *E. coli*.[6] In addition, **1**–**3** are hexa-, hepta-, and octameric, respectively, so that mutations explored a range of different substrate and product sizes. We sought primarily to make a series of mutants that would broadly explore sequence selectivity of the *tru* pathway. Out of the constellation of possible mutations that could answer this question, we picked representative derivatives that were also interesting to us because they helped to answer other pressing scientific questions about the cyanobactins pathways (see Discussion). Thus, the results of these experiments would have broad application in synthetic biology and specific application to understanding this interesting group of natural products, which are very broadly distributed.

**Table 1. Design Strategy for Mutants**

| # | sequence[a] | design principle[b] | size[c] |
|---|---|---|---|
| **1** | TVPTLC | Patellin 2 natural product[41] | 6 |
| 4 | TLATIC | Genome mining product[25] | 6 |
| 5 | IVPPFC | Genome mining product[30] | 6 |
| 6 | TTVTAC | Genome mining product[30] | 6 |
| 7 | VTPFVC | Mollamide B natural product[42] | 6 |
| 8 | IPGSLC | Keenamide natural product[37] | 6 |
| 9 | TVPT-pBrF-C | Patellin 2 + non-proteinogenic | 6 |
| 10 | TVPT-pClF-C | Patellin 2 + non-proteinogenic | 6 |
| 11 | TVPT-pOMeF-C | Patellin 2 + non-proteinogenic | 6 |
| 12 | TVPT-pN$_3$F-C | Patellin 2 + non-proteinogenic | 6 |
| **2** | TSIAPFC | Trunkamide natural product | 7 |
| 13 | TSIAPLC | Trunkamide point mutant | 7 |
| 14 | TSIASFC | Trunkamide point mutant | 7 |
| 15 | TVIAPFC | Trunkamide point mutant | 7 |
| 16 | ISFPCIP | Mollamide natural product[43] | 7 |
| 17 | IPISFPC | Mollamide natural product[43] | 7 |
| 18 | TSIAP-pBrF-C | Trunkamide + non-proteinogenic | 7 |
| **3** | TVPVPTLC | Patellin 3 natural product | 8 |
| 19 | TLPVPTVC | Patellin 4 natural product | 8 |
| 20 | TVPVPSFC | Patellin 5 natural product | 8 |
| 21 | CTLCCTLC | *pat-tru* hybrid (ulithiacyclamide)[44,5] | 8 |
| 22 | VTACITFC | *pat-tru* hybrid (patellamide C)[45,5] | 8 |

[a].Wild-type compounds **1**–**3** were mutated, with changes to wild type shown in bold. [b].Compounds were designed based on known natural products that had not been previously expressed, or based on sequences identified by genome mining (see Discussion for full description). [c].Macrocylic ring size, in number of amino acids.

**Chemical Analysis and Isolation.** After five days of fermentation, the pelleted *E. coli* cells were extracted with methanol. The organic extracts were partially purified and then analyzed using HPLC-ESI-MS. In all cases, heterologous expression of control compounds **1** and **3** was confirmed, with the compounds eluting with the same profile as standards of authentic **1** and **3**, which we obtained from a marine animal


dx.doi.org/10.1021/ja208278k | *J. Am. Chem.Soc.* 2012, 134, 418–425

**Table 2. Expression Yields in *E. coli***

| # | sequence | yield new (μg/L)[a] | yield (1) (μg/L) | yield (3) (μg/L) | total (μg/L) | rel yield (%) |
|---|---|---|---|---|---|---|
| 1 | TVPTLC | Std | | | | |
| 4 | TLATIC | 8.8/16[a] | 26/13 | 15/5.9 | 85 | 64[b] |
| 5 | IVPPFC | 90/NA | 47/10 | 74/22 | 240 | 160 |
| 6 | TTVTAC | 12/19 | 87/24 | 140/50 | 330 | 28 |
| 7 | VTPFVC | ND | 190/74 | 320/110 | 700 | 0 |
| 8 | IPGSLC | ND | 150/37 | 170/41 | 400 | 0 |
| 9 | TVPT-BrF-C | ND/6.2 | 120/64 | 52/33 | 280 | 3.3 |
| 10 | TVPT-ClF-C | 3.2/28 | 170/130 | 100/70 | 500 | 10 |
| 11 | TVPT-OMeF-C | ND/5.8 | 120/32 | 120/70 | 350 | 4.0 |
| 12 | TVPT-N,F-C | ND/0.8 | 99/64 | 74/55 | 290 | 5.2 |
| 2 | TSIAPFC | Std | | | | |
| 13 | TSIAPLC | 44/130 | 86/37 | 44/6.6 | 350 | 140 |
| 14 | TSIASFC | 54/13 | 74/24 | 48/29 | 240 | 68 |
| 15 | TVIAPFC | ND | 51/11 | 56/16 | 130 | 0 |
| 16 | ISFPCIP | ND | 3.9/ND | 1.2/1.9 | 7.0 | 0 |
| 17 | IPISFPC | ND | 96/84 | 62/17 | 260 | 0 |
| 18 | TSIAP-BrF-C | 5.0[c] | 97/79 | 66/33 | 280 | 2.8 |
| 3 | TVPVPTLC | Std | | | | |
| 19 | TLPVPTVC | 7.7/10 | 40/68 | 38/10 | 170 | 17 |
| 20 | TVPVPSFC | 41/33 | 170/55 | 120/85 | 500 | 33 |
| 21 | CTLCCTLC | ND | 7.7/0.6 | 13/3.0 | 24 | 0 |
| 22 | VTACITFC | 4.8/2.7 | 30/29 | 10/1.9 | 78 | 13 |

[a]Yields shown are for new compounds, compound **1**, compound **3**, and the total of all compounds found in a single expression experiment. Yields are reported as fully prenylated/−1 prenyl; i.e., for compound **4**, 8.8 $\mu$g $L^{-1}$ of diprenylated and 16 $\mu$g $L^{-1}$ of monoprenylated compounds were produced. [b]Rel yield (%) indicates the total amount of new compounds, divided by the total amount of compound **1**, found in an individual sample (times 100 to give percentage). This gives the most accurate assessment of relative yield of new compounds between experiments. [c]Only the nonprenylated **18** was identified, indicating that **18** is not a substrate for TruF1/F2.

as previously described (Figure 3).[6] We previously showed that isoprene is readily lost from Ser and Thr cyanobactin derivatives under standard MS conditions.[25] Thus, loss of isoprene reliably indicates the formation of mature cyanobactins, and this loss is not observed in any natural *E. coli* compound. In addition, we observed ions representing a total of 16 out of 22 recombinant cyanobactins. A table was constructed in which the recombinant sequence determined by DNA sequencing was used to predict a unique mass for the new cyanobactin and for the loss of one or more isoprene groups from each predicted new compound. The prediction ions were only observed in expression experiments involving the sequence in question, and not in other experiments, so that we essentially had 19 negative control experiments for each compound produced (Figure 3). In addition, if prenylated, the ions readily fragmented to lose the predicted numbers of isoprene groups, in contrast to all other *E. coli* metabolites. Finally, in most cases we observed incomplete prenylation, so that mono-, di- and sometimes triprenylated derivatives were formed in *E. coli*. In total, we identified 21 new compounds, produced in *E. coli*, that matched the masses predicted from DNA sequencing.

To further confirm the expression of the predicted compounds, they were subjected to analysis by high-resolution LC-FT-ICR-MS/MS, using previously established methods.[24,25,30] The ions did indeed reflect the predicted compounds, to <2 ppm value, with loss of isoprene observed in MS/MS. When no isoprene was present on a compound, the fragmentation pattern reflected the sequence of the peptide. Thus, because of the numerous internal and external controls, the sequencing data, and the well-validated MS data, we had high confidence about the identity of expressed products.

Finally, to further demonstrate that the recombinant compounds were successfully produced, we selected a set of representative compounds, **1**, **10**, and **14**, for NMR analysis. These compounds were purified to homogeneity from the *E. coli* cell pellets, and their $^1$H NMR spectra were obtained (Figures S4−S7 and Table S1 in Supporting Information [SI]). In all cases, the NMR spectra matched those predicted for the new compounds, showing that they were the predicted cyanobactins.

**Expression of Mutant Cyanobactins.** Hexameric patellin 2 (**1**) was manipulated to create compounds **4**−**8**, which were triple or quadruple mutants of the wild type sequence (Tables 1 and 2). The triple mutants (**4**−**6**) were successfully synthesized in *E. coli*, while the quadruple mutant **7** was not detected. We previously reported successful expression of **4**,[31] while all other compounds **5**−**22** are reported here for the first time.

In contrast to hexapeptide derivatives, for which diverse sequences are known, there are few known heptameric *tru* derivatives. Therefore, we created single-point mutants **13**−**15** of heptameric trunkamide (**2**). Only **13** and **14** were successfully processed in vivo. Pentuple mutants **16** and **17** were also not formed. Octapeptide selectivity was examined by synthesizing double mutants **19** and **20**, pentuple mutant **21**, and hextuple mutant **22**. Surprisingly in comparison with our experience with heptapeptides, **19**, **20**, and **22** were synthesized.

If more than one prenylation event was possible, we detected both singly and multiply prenylated derivatives. (For derivatives **9**, **11**, and **12** we detected only monoprenylated compounds, while for **18**, no prenylation was detected, indicating that the molecule was not a substrate for TruF1/F2.) This was true even for the wild type compounds in *E. coli*, showing that complete prenylation is a limiting step in this host. We also observed this pattern in systems that produce a lower cyanobactin yield, indicating that it may be an intrinsic pathway property and not linked to the amount of dimethylallylpyr-

ophosphate available in *E. coli*. In known natural products, TruD heterocyclase only modifies the C-terminal Cys residue. However, we previously showed that TruD modifies internal Cys residues in unnatural substrates in vitro.[16] The primary sequence of compound **22** was derived from the *pat* biosynthetic pathway,[23] in which all Cys and Thr/Ser residues are heterocyclic. However, in **22**, two Cys residues were heterocyclic, but the two Thr residues were prenylated. This reveals that the *pat* and *tru* pathways may be hybridized to create diverse new derivatives.

**Incorporation of Non-Proteinogenic Amino Acids.** We sought to incorporate non-proteinogenic amino acids using an orthogonal tRNA/aminoacyl-tRNA synthetase (tRNA/aaRS) pair that incorporates a specific unnatural amino acid in response to nonsense or frameshift mutation.[24] First, we showed that non-proteinogenic amino acids could be incorporated into the TruE precursor peptide. The plasmid pEVOL-pAcF which can specifically incorporate The non-proteinogenic amino acid *p*-acetyl-phenylalanine (*p*-AcF) was coexpressed with plasmids encoding *truE* derivatives, which were C-terminally His-tagged and in which each position in patellin 2 was individually mutated to the amber codon TAG (Figure S3, SI). *p*-AcF was added to the fermentations. SDS-PAGE of Ni purified lysates revealed that *p*-AcF could be incorporated in all six positions in the context of a precursor peptide (Figure S3, SI). These experiments showed that there was no problem in placing non-proteinogenic amino acids into the TruE precursor peptide backbone, in the absence of modifying enzymes. However, we still needed to determine whether the modifying enzymes could accept the highly unnatural amino acids and carry them through 4−5 biosynthetic steps to the mature unnatural natural products.

We therefore attempted to coexpress this same system with ptru-SD1 and other ptru vectors. However, in extensive initial experiments with non-proteinogenic amino acids, including *p*-AcF and others, could not be incorporated into cyanobactins using standard conditions. Several enzymes must be temporally coordinated in order to produce the mature cyanobactins over the course of our five-day fermentation. In extensive previous work with this pathway, changing the coordinated regulation of this pathway in any way severely impacted product yield.[25] For example, even providing each gene in the eight-gene *tru* pathway with its own heterologous promoter was unsuccessful. It is clear that the complex regulation of this multistep pathway remains poorly understood.

Previously, it was noted that incorporation of non-proteinogenic amino acids into single proteins over relatively long (14 h) induction times could be optimized with two copies of aaRS, one of which was constitutive and one of which was inducible with arabinose.[32] In fact, ∼1 day is the longest expression period to which this methodology has been previously applied, to the best of our knowledge. The five-day fermentation time for heterologous expression of *tru* is much longer than this, and in addition, the complex pathway regulation of *tru* is poorly understood. Therefore, we reasoned that adjustment of aaRS induction might improve cyanobactin production. Specifically, we thought that constitutive expression of aaRS might be better matched to the constitutive expression of *tru*. The plasmid pEVOL-pCNF (which encodes a polyspecific aaRS that incorporates a variety of Phe derivatives but not Phe or Tyr to any significant extent)[25] was used. We varied expression of the arabinose-inducible copy of aaRS in pEVOL-pCNF using different concentrations of inducer (L-arabinose). Optimal incorporation of *p*-chloro-phenylalanine was achieved in the absence of arabinose, while increasing arabinose concentrations and thus the flux of aaRS incrementally decreased product yield (Figure 4). With optimized conditions in hand, pEVOL-pCNF was used in the preparative synthesis of **9**−**12**, in which Leu of patellin 2 (**1**) was replaced with substituted phenylalanine derivatives, and **18**, in which Phe of trunkamide (**2**) was replaced with *p*-bromophenylalanine (Figure 2 and Table 1). Out of the many possible amino acids that might be incorporated into engineered natural products using pEVOL-pCNF, we chose a limited subset for this work not including *p*-AcF.

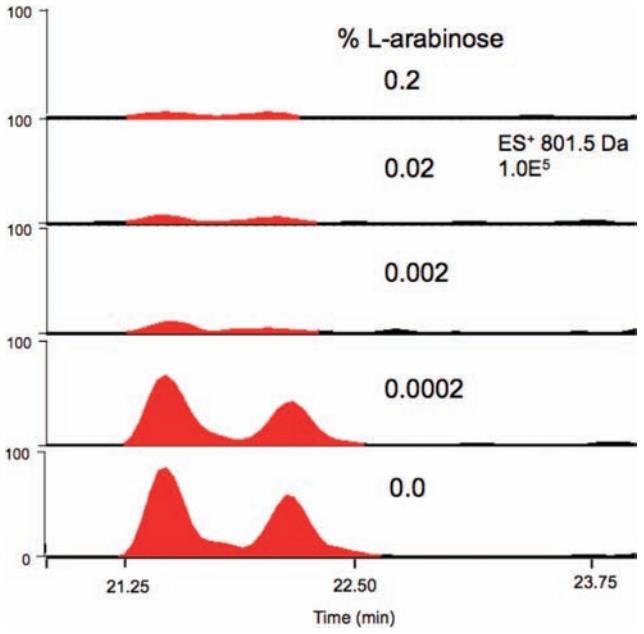


**Figure 4.** Optimization of non-proteinogenic amino acid incorporation. Compound **10** was synthesized in *E. coli* and analyzed by LC-ESI-MS. Optimum production was achieved in the absence of inducer (arabinose), whereas higher concentrations of inducer completely repressed synthesis. Two peaks are present because multiple stereoisomers are present (the $\alpha$-proton adjacent to thiazoline is labile, and Pro undergoes *cis*−*trans* isomerization in this family).[26].

**Yield Measurement.** Yields were determined for all heterologous expression experiments. First, purified **1** was quantified by NMR integration using several different concentrations of the standard, 1,4-dinitrobenzene,[33] to give an isolated yield of 260 $\mu$g from a 10-L fermentation. This method is much more accurate than other methods for microgram quantities of material, since it measures the relative molar amounts of compound directly using a calibrant.

With a standard concentration of **1** available, we used known concentrations of **1** as internal and external standards for HPLC-ESI-MS experiments. These experiments used crude fractions containing recombinant peptides, so that they much more closely resembled the native production yield, and not merely the isolated yield. For example, the area under the curve was integrated to show that 1.7 mg of **1** was produced in the 10 L *E. coli* expression described above, but the isolated yield was only ∼10% after several purification steps. An additional 1.3 mg of the monoprenylated variant of **1** was also produced in this 10 L fermentation, leading to an overall yield of 3 mg of **1** variants.


dx.doi.org/10.1021/ja208278k | *J. Am. Chem.Soc.* 2012, 134, 418−425

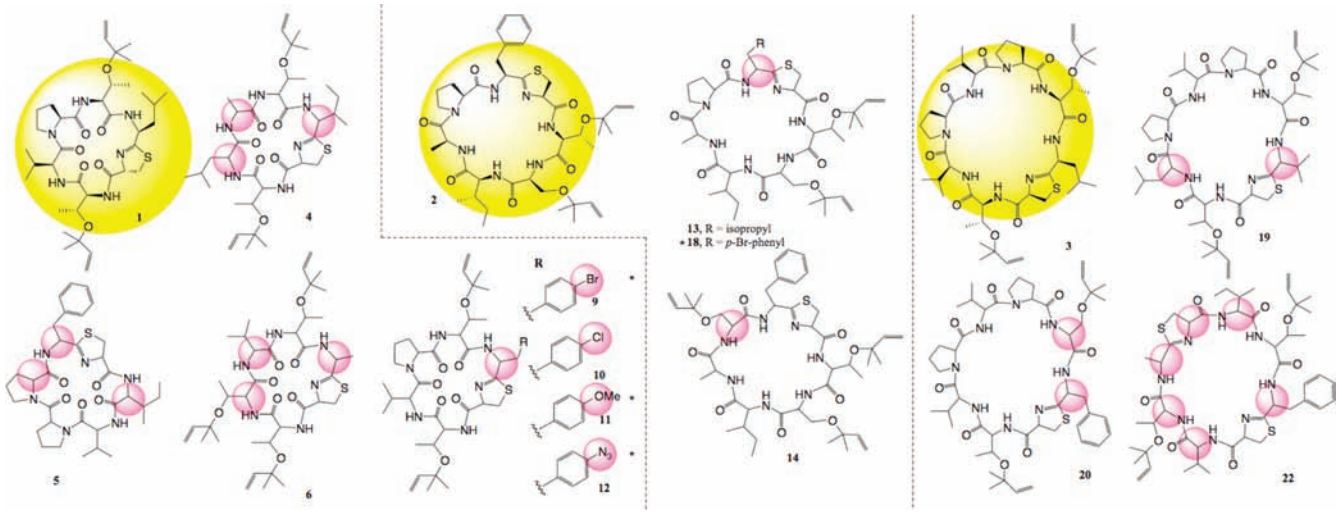


**Figure 5.** Compounds synthesized in this study. Yellow indicates wild-type compounds, while pink bubbles indicate mutations that deviate from wild type. Hexa- (left), hepta- (mid), and octa- (right) peptide derivatives were synthesized using these methods. (*) Indicates compounds for which only monoprenylated (**9**, **11**, and **12**) or nonprenylated (**18**) derivatives were identified. For all other compounds, both singly, doubly, and sometimes triply prenylated products were identified.

Because **1** was produced in all expression experiments reported herein, the compound could also be used to estimate the production level of other recombinant products. Yields of most compounds were in the range of 1−170 $\mu$g $L^{-1}$. (see Table 1). Compound **2** was also present as an internal control. A potential weakness of this method is that the relative ionization efficiencies of the compounds are likely to differ. However, the ratios were also compared to those achieved using FT-ICR-MS, showing a consistent relative ratio of ions in both methods. Moreover, in previous studies we have not seen a large range of ionization efficiencies for these similar compounds, indicating that the reported yields are good estimates. The relative yields determined by MS were also reflected in the isolated yields of **10** and **14**.

## DISCUSSION

In this study, we exploited the relaxed substrate selectivity of the *tru* pathway to generate diverse mutants, despite the need to go through a complex biosynthetic route. Although not all derivatives can be made, this study shows that the *tru* posttranslational machinery exhibits broad substrate tolerance. We also report the first in vivo incorporation of non-proteinogenic amino acids into multistep RP products. Notably, control of six different enzymes, two precursor peptides, and a tRNA molecule was required for synthesis of the reported derivatives. Incorporated, non-proteinogenic amino acids were successfully carried through a multistep pathway with 12 individual enzymatic transformations. This was accomplished using eight enzyme *tru* domains encoded in six *tru* proteins, whereas previously at most one enzyme with two domains has been used.[11] We expect that the methods here will be widely useful in the synthesis of diverse RP derivatives encoding non-proteinogenic amino acids.

A simple but important advance was the observation that low-level, constitutive expression provides excellent incorporation of non-proteinogenic amino acids in natural product derivatives. Prior to this report, the described technology had been used routinely for incorporation of amino acids into proteins, and a few reports exist for incorporation into short peptides which undergo one further posttranslational modification. In these cases, the relative simplicity of the heterologous expression systems made them very tolerant of overexpression. In more complex systems, our data suggest that producing too much of the aaRS (or any other component in a complex pathway) is detrimental to efficient product synthesis.

There are now many different methods for implanting non-proteinogenic amino acids into ribosomal peptides, some of which involve changing the ribosomal code and some of which involve synthetic modifications.[34] Notably, recently the concepts of multiple orthogonal substitution and residue-specific substitution have been elaborated.[35,36] The former enables different non-proteinogenic amino acids to be introduced into single peptides, while the latter enables all residues of a specific type to be replaced by non-proteinogenic amino acids. These methods are compatible with the ideas developed here, but they will undoubtedly require some further study to be applied to highly modified RPs, since so far they have not been applied to such compounds.

In addition to exploring incorporation of diverse substitutions, these experiments were designed to provide proof-of-concept for key technological advances in marine natural products, in five major areas. First, we have recently applied a genomic method, in which we sequence genes from marine animals and their symbiotic bacteria to discover new TruE-like peptides.[25,30] Subsequently, we find that the predicted peptides are the major natural products found in whole animals. However, the compounds are exceptionally limited in supply, and at most micrograms are available from the natural sources. As mentioned above, previously we expressed **4** to supply the compound. Here, we wished to test the broad applicability of the method by expressing compounds **4**-**6**, which were first identified by genomics and which were successfully made by *E. coli* here.

Second, there were a series of previously described, rare compounds from Nature, including mollamide (encoded by **16** and **17**) and keenamide (encoded by **8**).[19,37] These compounds were reported to exhibit interesting bioactivity but are in very short supply. However, these compounds were not produced by our *tru* pathway in *E. coli*, indicating that they are probably synthesized by a different variation of the *tru* pathway found in


dx.doi.org/10.1021/ja208278k | J. Am. Chem.Soc. 2012, 134, 418−425

Nature. Third, we wished to express known marine animal compounds, patellins 4 and 5 (**19** and **20**),[19,38] for which we had not previously found genes; these were successfully made.

Fourth, trunkamide is a potential anticancer agent of interest because of its unusual profile in the National Cancer Institute's 60-cell line panel. We tested its potential for derivatization by making a series of point mutants, **13**−**15**. Interestingly, addition of a Ser led to a third prenylation event in **14**, and **13** was successfully produced. However, a simple Ser-Val substitution in **15** led to loss of product.

Fifth, for compounds **21** and **22**, we wished to cross the *tru* pathway with the *pat* pathway that leads to patellamides. Patellamides are not prenylated, and instead Ser and Thr residues are cyclized to yield oxazoline residues.[5,39,40] Moreover, often 2 thiazole residues exist in the *pat* products, instead of 1 thiazoline as found in all known *tru* pathway relatives. The genetic and biochemical basis of these differences have been thoroughly established. We have been very interested in determining the "portability" of enzymes; that is, can individual RP enzymes be moved from one pathway to another? As a first step to establish this fact, we cloned patellamide precursor peptides that would normally yield the compounds ulithiacyclamide and patellamide C into the *tru* pathway background, to give derivatives **21** and **22**. **21** could not be produced, but **22** was successfully synthesized and contained 2 thiazoline residues, the first time this modification has been reported in *tru* derivatives. Both Thr residues were prenylated, rather than heterocyclic as they would be in the natural product.

This report represents a significant expansion of the toolkit available for directed, posttranslational modification of peptides in living cells. Additionally, it provides the first technical guidelines for performing complex manipulations with multistep RP pathways. Ultimately, we hope to improve control over RP processing for directed synthetic biology of fine chemicals and pharmaceuticals. This technology has numerous possible applications, including supply of marine natural products and improvement of anticancer properties of cyanobactins. More importantly, such technology allows simple genetic engineering tools to be applied to creation of wholly new drug motifs that combine the advantages of peptide technology and synthetic chemistry. For example, intein-circularized and phage display libraries encode an enormous sequence diversity which can be used to attack diverse biological targets,[40] but the resulting compounds are generally not drugs because of poor pharmacological properties. The addition of many posttranslational modifications promises to afford the same sequence diversity, but with products that are more "drug-like". Here, we use the multistep cyanobactin pathway to explore the chemistry that necessarily underlies these downstream applications.

## ■ ASSOCIATED CONTENT

### Supporting Information

Methods, supporting data, and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION


### Corresponding Author

ews1@utah.edu

### Present Addresses

[§]Department of Bioengineering and Therapeutic Sciences and California Institute for Quantitative Biosciences, University of California, San Francisco, San Francisco, California 94158, United States.

[⊥]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, United States.


### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

### Notes

E.W.S. is a co-owner of Symbion Discovery, Inc.

## ■ ACKNOWLEDGMENTS

This work was funded by NIH GM071425 and GM071425-S1 to E.W.S. and NIH GM097206 to P.G.S. We thank K. Parsawar and C. Nelson (University of Utah Mass Spectrometry Core Facility) for acquiring high-resolution mass spectra, J. Skalicky (University of Utah) for help with NMR, and D. E. Ruffner (Symbion Discovery, Inc.) for providing ptru-SD1.

## ■ REFERENCES


(1) Mitchell, W. *Curr. Opin. Chem. Biol.* **2011**, *15*, 505.
(2) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *Nat. Prod. Rep.* **2009**, *26*, 537.
(3) Velterop, J. S.; Sellink, E.; Meulenberg, J. J.; David, S.; Bulder, I.; Postma, P. W. *J. Bacteriol.* **1995**, *177*, 5088.
(4) Oman, T. J.; van der Donk, W. A. *Nat. Chem. Biol.* **2010**, *6*, 9.
(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.
(6) Donia, M. S.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2008**, *4*, 341.
(7) Widdick, D. A.; Dodd, H. M.; Barraille, P.; White, J.; Stein, T. H.; Chater, K. F.; Gasson, M. J.; Bibb, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4316.
(8) Kelleher, N. L.; Hendrickson, C. L.; Walsh, C. T. *Biochemistry* **1999**, *38*, 15623.
(9) Willey, J. M.; van der Donk, W. A. *Annu. Rev. Microbiol.* **2007**, *61*, 477.
(10) Melby, J. O.; Nard, N. J.; Mitchell, D. A. *Curr. Opin. Chem. Biol.* **2011**, *15*, 369.
(11) Houssen, W. E.; Wright, S. H.; Kalverda, A. P.; Thompson, G. S.; Kelly, S. M.; Jaspars, M. *ChemBioChem* **2010**, *11*, 1867.
(12) Pan, S. J.; Link, A. J. *J. Am. Chem. Soc.* **2011**, *133*, 5016.
(13) Knappe, T. A.; Manzenrieder, F.; Mas-Moruno, C.; Linne, U.; Sasse, F.; Kessler, H.; Xie, X.; Marahiel, M. A. *Angew. Chem., Int. Ed.* **2011**, *50*, 8714.
(14) Bowers, A. A.; Acker, M. G.; Koglin, A.; Walsh, C. T. *J. Am. Chem. Soc.* **2010**, *132*, 7519.
(15) Li, C.; Zhang, F.; Kelly, W. L. *Mol. Biosyst.* **2011**, *7*, 82.
(16) Shi, Y.; Yang, X.; Garg, N.; van der Donk, W. A. *J. Am. Chem. Soc.* **2011**, *133*, 2338.
(17) Young, T. S.; Young, D. D.; Ahmad, I.; Louis, J. M.; Benkovic, S. J.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 11052.
(18) Sivonen, K.; Leikoski, N.; Fewer, D. P.; Jokela, J. *Appl. Microbiol. Biotechnol.* **2010**, *86*, 1213.
(19) Donia, M. S.; Schmidt, E. W. In *Comprehensive Natural Products II*; Mander, L, Liu, H.-W., Eds.; Elsevier: Oxford, 2010; Vol 2, p 539.
(20) Lee, J.; McIntosh, J.; Hathaway, B. J.; Schmidt, E. W. *J. Am. Chem. Soc.* **2009**, *131*, 2122.
(21) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *J. Am. Chem. Soc.* **2010**, *132*, 4089.
(22) McIntosh, J. A.; Robertson, C. R.; Agarwal, V.; Nair, S. K.; Bulaj, G. W.; Schmidt, E. W. *J. Am. Chem. Soc.* **2010**, *132*, 15499.
(23) McIntosh, J. A.; Schmidt, E. W. *ChemBioChem* **2010**, *11*, 1413.
(24) McIntosh, J. A.; Donia, M. S.; Nair, S. K.; Schmidt, E. W. *J. Am. Chem. Soc.* **2011**, *133*, 13698.

(25) Donia, M. S.; Ruffner, D. E.; Cao, S.; Schmidt, E. W. *ChemBioChem* **2011**, *12*, 1230.
(26) Carroll, A. R.; Coll, J. C.; Bourne, D. J.; MacLeod, J. K.; Zabriskie, T. M. *Aust. J. Chem.* **1996**, *49*, 659.
(27) Wipf, P.; Uto, Y. *J. Org. Chem.* **2000**, *65*, 1037.
(28) Aller, S. G.; Yu, J.; Ward, A.; Weng, Y.; Chittaboina, S.; Zhuo, R.; Harrell, P. M.; Trinh, Y. T.; Zhang, Q.; Urbatsch, I. L.; Chang, G. *Science* **2009**, *323*, 1718.
(29) Williams, A. B.; Jacobs, R. S. *Cancer Lett.* **1993**, *71*, 97.
(30) Donia, M. S.; Fricke, W. F.; Ravel, J.; Schmidt, E. W. *PLoS One* **2011**, *6*, e17897.
(31) Donia, M. S.; Schmidt, E. W. *Chem. Biol.* **2011**, *18*, 508.
(32) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. *J. Mol. Biol.* **2010**, *395*, 361.
(33) Rundlof, T.; Mathiasson, M.; Bekiroglu, S.; Hakkarainen, B.; Bowden, T.; Arvidsson, T. *J. Pharm. Biomed. Anal.* **2010**, *52*, 645.
(34) Budisa, N. *Angew. Chem., Int. Ed.* **2004**, *43*, 6426.
(35) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. *Nature* **2010**, *464*, 441.
(36) Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. *Curr. Opin. Chem. Biol.* **2010**, *14*, 774.
(37) Wesson, K. J.; Hamann, M. T. *J. Nat. Prod.* **1996**, *59*, 629.
(38) Fu, X.; Do, T.; Schmitz, F. J.; Andrusevich, V.; Engel, M. H. *J. Nat. Prod.* **1998**, *61*, 1547.
(39) 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.
(40) Tavassoli, A.; Benkovic, S. J. *Nat. Protoc.* **2007**, *2*, 1126.
(41) Zabriskie, T. M.; Foster, M. P.; Stout, T. J.; Clardy, J.; Ireland, C. M. *J. Am. Chem. Soc.* **1990**, *112*, 8080.
(42) Donia, M. S.; Wang, B.; Dunbar, D. C.; Desai, P. V.; Patny, A.; Avery, M.; Hamann, M. T. *J. Nat. Prod.* **2008**, *71*, 941.
(43) Carroll, A. R.; Bowden, B. F.; Coll, J. C.; Hockless, D. C. R.; Skelton, B. W.; White, A. H. *Aust. J. Chem.* **1994**, *47*, 61−69.
(44) Ireland, C.; Scheuer, P. J. *J. Am. Chem. Soc.* **1980**, *102*, 5688.
(45) Ireland, C. M.; Durso, A. R.; Newman, R. A.; Hacker, M. P. *J. Org. Chem.* **1982**, *47*, 1807.