Published in final edited form as:

J Am Chem Soc. 2018 November 28; 140(47): 16213–16221. doi:10.1021/jacs.8b09328.

Roads to Rome: Role of Multiple Cassettes in Cyanobactin RiPP Biosynthesis

Wenjia Gu, **Debosmita Sardar**[†], **Elizabeth Pierce**[‡], and **Eric W. Schmidt**[*]

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

Abstract

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are ubiquitous natural products. Bioactive RiPPs are produced from a precursor peptide, which is modified by enzymes. Usually, a single product is encoded in a precursor peptide. However, in cyanobactins and several other RiPP pathways, a single precursor peptide encodes multiple bioactive products flanking with recognition sequences known as “cassettes”. The role of multiple cassettes in one peptide is mysterious, but in general their presence is a marker of biosynthetic plasticity. Here, we show that in cyanobactin biosynthesis the presence of multiple cassettes confers distributive enzyme processing to multiple steps of the pathway, a feature we propose to be a hallmark of multi-cassette RiPPs. TruD heterocyclase is stochastic and distributive. Although a canonical biosynthetic route is favored with certain substrates, every conceivable biosynthetic route is accepted. Together, these factors afford greater plasticity to the biosynthetic pathway by equalizing the processing of each cassette, enabling access to chemical diversity.

Graphical Abstract

[*]**Corresponding Author**: ews1@utah.edu.

Author Contributions

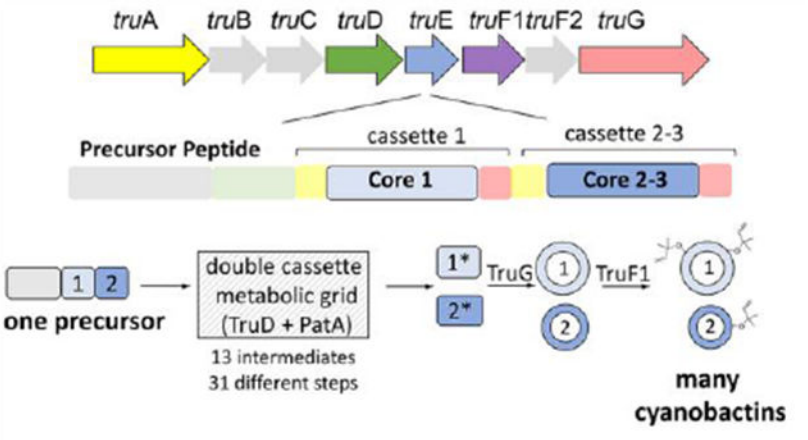
The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

[†]**Present Address**: Centre for Cell and Gene Therapy, Baylor College of Medicine, Houston, Texas 77030, USA

[‡]**Present Address**: Department of Physical Science, Southern Utah University, Cedar City, Utah, 84720

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website. Experimental method; additional figures, schemes, and tables (PDF).



INTRODUCTION

The paradigm of gene to peptide or protein is complicated by the large diversity of post-translational modifications (PTMs) that decorate polypeptides, imparting an additional level of chemical diversity.[1] PTMs are common on both large polypeptides, such as enzymes or structural proteins, and small polypeptides, such as the ribosomally synthesized and post-translationally modified peptide (RiPP) natural products. A common feature of PTM chemistry is that enzymes catalyze modification of multiple substrates, implying a degree of flexibility in substrate recognition. Understanding the factors guiding substrate choice is essential both to biology of PTMs and to applying the enzymes to directed bioengineering.

In RiPP biosynthesis, the natural product is encoded in the core peptide, a short sequence embedded within a longer precursor peptide.[2] Posttranslational enzymes modify the core peptide, which is later proteolytically cleaved from the precursor to afford the mature, bioactive natural product. Most RiPP pathways encode precursors with a single core, producing a single natural product. However, a subset of pathways encodes multiple cores, and thus makes multiple products, using a single precursor peptide.[3–9] These cores are often flanked by recognition sequences (RSs), which recruit enzymes. Together, core-RS combinations are referred to as “cassettes”. A hallmark of multiple cassette RiPPs is that they are highly substrate permissive, enabling the synthesis of many derivatives.[3,10–13]

A comparison between protein (collagen) and RiPP (cyanobactin) biosynthesis is illustrative (Figure S1).[10,14–16] In collagen and cyanobactins, PTM enzymes bind both to a substrate sequence that is modified, and to a distal recognition sequence (RS) on the polypeptide that is not modified. In prolyl hydroxylase modification of collagen, the primary site of binding is within the substrate, where X-Pro-Gly is hydroxylated in certain β-turn structures. By contrast, in cyanobactins the primary enzyme binding site is the RS, with relatively little contribution from the substrate. This frees the substrate in cyanobactins to be hypervariable, as long as the extrinsic RS is conserved. Among the many substrate proteins and RiPPs, there is a continuum of binding preferences, with varying degrees of contributions from the RS and the substrate.[17–19] Another question for RiPP substrates and collagen is the degree to which processing of multiple cassettes is processive (the enzyme stays on the substrate) or

distributive (substrate leaves the enzyme between steps). This is crucial because it gets to the heart of the substrate recognition question.

The canonical trunkamide (*tru*) cyanobactin pathway begins with the ribosomal production of the TruE precursor peptide (Figure 1).[10] TruE natively contains two to three cassettes consisting of different core peptide sequences, each flanked by RSs. Placement of multiple cassettes in the context of a precursor peptide such as TruE enables the synthesis of multiple natural and unnatural products. After ribosomal synthesis, a heterocyclase TruD binds to RSI, leading to the conversion of cysteine to thiazoline in each cassette (Figure 3A).[20–24] Protease TruA recognizes RSII and cleaves the N-terminus of each cassette (Figure 4A and 5A).[25,26] Protease TruG recognizes RSIII and circularizes the peptides to afford cyclic peptides (Figure 6A).[25–28] Prenyltransferase TruF1 prenylates Thr and Ser residues at last.

TruE variants containing one, two, or three cassettes are known to be substrates for all enzymatic steps.[10,21,22,24,29] Because single cassettes are accepted, a question that arises is why multiple cassettes might be advantageous or what role they might play in biology. Previously, in many heterologous expression experiments in *E. coli*, we observed that expression of unnatural sequences was context dependent (unpublished observations). That is, the yield differed depending upon how many cassettes were present, and which position the unnatural sequence cassette was placed in the precursor peptide. Because of this observation, we hypothesized that there may be a biochemical role for multiple-cassette precursor peptides, specifically that the appropriate order of discrete core peptides would improve their conversion to natural products.

Here, we sought to test this hypothesis and to understand the biochemical role of multiple cassettes in the *tru* pathway, with the idea that the results might be used to improve yields of recombinant products. This is the first work that characterizes the fate of multi-cassette cyanobactin precursors through the entire biosynthetic pathway. Much of the previous mechanistic work has been performed with artificial, single-cassette substrates that are technically less demanding to work with. Biochemistry of multi-cassette processing has rarely been examined, and only with single enzymes.[20,30] Through extensive characterization of the natural series of steps using multiple enzymes, here we demonstrate an unexpected biosynthetic plasticity.

RESULTS

Substrate design principles.

Previously described *tru* pathways contain TruE substrates encoding the natural products trunkamide (tk; **1**), patellin 2 (p2; **2**), patellin 3 (p3; **3**), and patellin 6 (p6) (Figure 1).[10] These compounds are cyclic hexa-, hepta-, and octapeptides. Natively, they are found in the cassette combinations p6-tk, tk-tk-tk, and p3-p2. In this nomenclature, “p6-tk” refers to compounds patellin 6 and **1** encoded on two different cassettes in a single precursor peptide. Since **3** and patellin 6 are both octapeptides, we selected **1**, **2** and **3** to cover most of the native product size ranges and sequence variations encoded in *tru* biosynthesis.

Based on previous observations that yield in *E. coli* differed depending upon cassette context, we wondered whether **1)** the *tru* pathway displays a preference for one cassette position, **2)** the *tru* pathway displays a preference for particular cassette sequences independent of position, **3)** both position and sequence affect the yield of a particular compound from the pathway, or **4)** cassette processing order is random, with no preference for particular sequences or position. To test these possibilities, we designed a series of precursor peptides containing two core sequence positions. These were based on the native p3p2 arrangement of cores, and included all eight other possible combinations of **1**, **2** and **3** in the first and second cassettes (p3p3, p3tk, p2p2, p2p3, p2tk, tkp2, tkp3, tktk; Figure S2). By using only native core peptides, but in unnatural orders, we would be able to disentangle preferences for core peptide sequences from preferences for cassette order. If there is a cassette order dependence, we anticipated that we would see higher yield of each compound when placed in the preferred position; if there is no cassette order dependence, then no yield difference would be observed. Similarly, the cassette sequence dependence could be verified by comparing precursors where the cassettes are directly switched (e.g. p3p2 compared to p2p3). Finally, we synthesized single cassette *tru* pathway variants, encoding only tk, p2 or p3. While not directly relevant to core question of cassette preference, this is helpful in understanding and optimizing factors that impact product yield.

Production in *E. coli* does not depend on cassette order.

We expressed artificial cyanobactin pathway constructs in *E. coli*, extracted the resulting cyanobactins, and analyzed them under established conditions.[16] Relative yield was estimated by comparing relative areas under the curve using high-performance liquid chromatography mass spectrometry (HPLC-MS), normalized in comparison to internal standard. Normally, cyanobactin expression in *E. coli* leads to a series of products with zero, one, or two isoprenylations. Products were thus considered as sums of all prenylation variants (Figure 2 and S3). Here we show the highest yield obtained on day three (Figure 2). We found no correlation between the cassette position of a core peptide and its production level in *E. coli*, although production varied slightly with different constitutions of the precursor peptide. Placing identical substrates into each cassette, thereby doubling the substrate stoichiometry, generally gave higher yield but did not double it. This conclusion is also confirmed with the single cassette *tru* pathway variants (Figure S4). TruE variants containing only tk, p2, or p3 were expressed, and compared with double-cassette vectors. In comparing tk to tktk and p2 to p2p2, the yield was greatly increased in the double-cassette vectors. However, comparison of p3 to p3p3 revealed no increase in production on day 3, although the yield was nearly double on day 4. Thus, there is a trend that copy number leads to increased yield in *E. coli* in these experimental conditions. We further examined two alanine mutants of **2** (Figure S5). With these mutants, when two copies of the mutant sequence were used in a single precursor, the yield was doubled in comparison to experiments in which a single copy was used in combination with the p2 sequence.

These results show a general trend of cassette order independence, in terms of yield of substrate. However, these results were not conclusive despite substantial experimental effort, since biochemistry inside of living cells is complicated by many factors that cannot be controlled. For example, as can be seen in Figure S3, yield of **2** and its prenylation variants

was at maximum in the p2p2 construct on day four of fermentation, while the maximum of **2** and its prenylation variants in the tkp2 construct was reached on day three. Similar, but less obvious, trends can be observed in each of the constructs synthesized. These results can be interpreted as indicating that the precise timing of growth and production of compounds varies depending upon which exact construct is being synthesized in *E. coli* cells. We have previously reported in detail the growth changes that occur with different vectors.[31] Remarkably, these relatively large changes are accomplished by changing just a few nucleotides within a 14 kb plasmid inside of a cell. Moreover, intermediates in synthesis may be more or less stable within *E. coli* over the long time period of production, since previous research indicates that the precursor peptide is synthesized on day one, during log phase of growth.[31] Because of these and many other complications, despite some observed general trends, we felt uncomfortable overinterpreting the *in vivo* data. Therefore, we moved to biochemical experiments *in vitro* to explore the substrate preference of each enzyme step in the pathway.

TruD heterocyclase is distributive and stochastic.

We examined purified TruE derivative substrates in combination with purified TruD *in vitro* (Figure 3A). Previous characterization of PatD and TruD using native, double-cassette substrates indicated a distributive mechanism, since accumulation of intermediates was observed.[20,21] A later study also observed release and rebinding of artificial single-cassette substrates with TruD.[22] Thus, existing data suggest that TruD is distributive, although confusion abounds since one manuscript uses the term “processive” to describe directional processing of the substrate, rather than using the term in the correct sense of remaining on enzyme between biochemical steps. In addition to lingering confusion about TruD’s distributivity, the directionality of processing is not well addressed. In a previous study using a highly artificial substrate, directionality was observed, in which the C-terminal cysteine was processed first.[22] However, there are no native TruD substrates with this feature, calling into question the directionality of the enzyme in the natural context of multiple cassettes.

Here, we sought to examine cassette preference using native sequences in the natural double-cassette context. We expressed and purified precursor peptides with the p2 and p3 sequences in all possible combinations (p3p2, p2p3, p2p2 and p3p3). The resulting double-cassette substrates were used at saturating concentrations in reactions with TruD, and the products were measured by HPLC-MS. If TruD were processive, then one would expect the concentration of singly modified substrate not to greatly exceed the concentration of the enzyme. By contrast, we found an accumulation of singly modified substrates at early time points, despite a large excess of substrate (Figure 3B and S6) that are ultimately modified to completion (data not shown). This result of accumulating reaction intermediates reconfirms that TruD is distributive.

To further demonstrate the distributivity of the enzyme, as well as to determine the cassette order preference, we used enzymatic digestion with PatA protease at different time-points of the TruD reaction. PatA recognizes RSII, cleaving the TruD treated precursor peptides into four possible discrete cassette fragments with different molecular mass, which are detectable on MS: fragment-1, fragment-1*, fragment-2 and fragment-2* (Figure 3C; * indicates a

heterocycle). If one cassette was preferred by TruD, we would expect that heterocyclized cassette to be present in greater amounts, at least in early time points when singly-modified TruE is prevalent in the reaction mixture. However, we found that the heterocyclized cassettes were similarly modified at all time points (Figure 3C), demonstrating that either cassette-1 or cassette-2 can be heterocyclized first, with equal preference. This result held true no matter what substrate combination was attempted (p3p2, p2p3, p2p2 and p3p3). Therefore, TruD is distributive and stochastic, with no preferred modification order.

To further determine whether the enzyme is truly distributive in many different conditions, we designed and expressed a single cassette precursor peptide (TruE-p23) in which the p2 and p3 sequences were fused without an intervening RSIII-RSII sequence (Figure S7A). The TruE-p23 precursor peptide contains two cysteines for heterocyclization within a single core. The resulting heterocyclization pattern also showed accumulation of singly heterocyclized compounds (Figure S7B), which is consistent with a distributive mechanism.

Although these experiments were not designed to measure reaction rates, but instead to determine modification order, additional trends can be observed in the data shown in **Figures** 3C, S6, and S7B. It is clear that the efficiency of the enzyme modestly varies depending upon the precise details of the substrate, with the natural substrate TruE-p3p2 being slightly more rapidly modified than other substrates. Similar trends are seen in PatA reactions, in which the natural substrate TruE-p3p2 was preferred (Figure 4B). Further work is required using enzyme kinetics to determine how cassette order might lead to subtle rate differences.

PatA prefers cassette-1 under non-reductive conditions.

PatA and TruA are nearly identical and are used interchangeably in *pat/tru* pathways.[3,10] Previously, we found that PatA has a C-to-N directionality under reductive conditions with an unnatural substrate, although we noted that the enzymatic activity was not optimized.[26] PatA only operates efficiently under non-reduced conditions.[29] Here, we did not employ reducing agents, and we used only native sequences p2 and p3 (Figure 4A). The substrates were fully heterocyclized by TruD in advance of PatA treatment to generate native PatA substrates. Reducing agents were removed from the TruD reactions using desalting columns.

Using precursor peptide combinations p3p2, p2p2, and p3p3, the only products and intermediates observed were fragment-1*, fragment-2*, and fragment-1*+2* (Figure 4B). In precursor TruE-p2p3, we observed those peptides as well as a small amount of intermediate fragment-leader-1* (Figure S8). Moreover, fragment-1* and fragment-2* were observed in nearly equal ratios at all time points measured. Thus, PatA prefers to cleave at cassette-1 first, although this represents preference and not an absolute rule, since all possible intermediates were observed at least once. The resulting intermediate fragment-1*+2* ensures equal amount of production of both cassettes for the next enzymatic step.

The order of the PatA and TruD reactions is redox dependent.

In the canonical *tru* pathway, TruD acts on TruE prior to the action of TruA.[16,20] Previous studies suggested that the order in which the protease and heterocyclase act might depend

upon the oxidation state of the pathway. Briefly, PatA is inhibited under reducing conditions, which are required for efficient modification by TruD.[29] Here, we investigated this effect by varying oxidation state and substrate order. We first examined the effects of PatA on unmodified TruE derivatives (not treated with TruD) (Figure 5A). Without reducing agents, PatA cleaved at cassette-1.

However, this does not indicate preference because digestion of cassette-2 was inhibited, and the product consisting of fragment-1+2 accumulated over a time course (Figure 5B and S9). Since there are two cysteines in a double-cassette precursor, it is possible that the cysteines form disulfide bridges, blocking the cleavage site on cassette-2 under non-reductive conditions. Indeed, we observed a *m/z* consistent with disulfide bond formation. To further verify disulfide formation, we added small amount of reducing agent tris(2-carboxyethyl)phosphine (TCEP) to the reaction (1:1 TCEP:substrate molar ratio), and the digestion product of cassette-2 was observed. PatA was increasingly inhibited above a 1:1 ratio (Figure 5C). We then digested TruE derivatives with PatA prior to TruD reactions. Under reductive conditions, TruD reacted with the cleaved peptide fragments and heterocyclized them to completion (Figure S10). Without the addition of reducing agents, the reaction was inhibited. Further, using synthetic substrates fragment-1 and fragment-2, we replicated previous studies showing that RSI is not absolutely required for TruD activity, but greatly accelerates the reaction (Figure S11).[22–24,32] In sum, the results showed that the order of early enzymes TruD and PatA is controlled by redox state, so that potentially in cyanobacteria the order of enzymatic reaction may differ under different cellular redox conditions.

TruG and PatG macrocyclases have no cassette order preference.

Although macrocyclases TruG and PatG have been extensively investigated with the broad array of fragment-2* sequences, and analogs thereof,[26,28,33–35] the macrocyclization of fragment-1* has never been investigated. In addition, the relative enzymatic preferences for these two natural sequences have not been determined. We synthesized four different, heterocyclic TruG substrates: p3*-RSIII-RSII, p3*-RSIII, p2*-RSIII-RSII, p2*-RSII. The substrates were added in combinations that mimicked what one would find in the native pathways, wherein fragment-1* and fragment-2* sequences should be found in equal concentrations (Figure 6A). These enzymatic reactions showed small differences in cassette preference over a 37-hour time course (Figure 6B and S12). In cases where p3 was in the first cassette, those differences were not statistically significant, but the second cassette was modestly preferred. When p2 was in the first cassette, cassette-1 was slightly preferred with statistical significance. PatG was also tested and showed similar results as found with TruG (Figure S13), although it should be pointed out that p2 and p3 are not native PatG substrates. Although modest differences in substrate preference are evident for both PatG and TruG, overall a stochastic reaction preference is prevalent for this biosynthetic step, leading to a very similar production from both cassettes.

Finally, previous studies demonstrate that TruF1 does not appear to have a preferred order of prenylation on the natural substrates.[12,16] Like other biosynthetic steps in the pathway, the

final two prenylation events are also catalyzed by an enzyme that is distributive and stochastic.

DISCUSSION

We sought to determine why there are multiple cassettes in cyanobactin *tru* biosynthesis. Here, we show that the presence of multiple cassettes does not significantly alter the biochemistry of the pathway. TruD acts stochastically and distributively, such that cassette order does not matter. PatA (TruA) prefers cassette-1 thus providing equal amount of two cassette substrates for TruG, and TruG is also highly flexible and broadly accepting of different substrates resulting from differential processing at cassette-1 or cassette-2. Since previous work exhaustively evaluates artificial single-cassette substrates, demonstrating that two cassettes are not required for efficient *tru* biosynthesis, this work rules out a biochemical advantage for the presence of multiple cassettes. We expected that such an advantage would help to improve the yield of products depending upon the placement of their core sequence in the precursor, but such an advantage was not observed. Although multiple lines of evidence support this conclusion, it is possible that there are some differences in the natural setting that would provide a biochemical rationale.

Previous work with TruD and related heterocyclases revealed a potentially distributive mechanism of action,[20,21] despite some controversy based upon a potential misuse of the term “processive”.[22] Here, we add further evidence using native substrates and close analogs, confirming that TruD is distributive. Previously, TruD was described as directional, from C- to N-terminus in an artificial substrate. We show that there is no preferred order for Cys heterocyclization between multiple cassettes (the directionality is stochastic). It remains possible that TruD follows a defined order *within* individual cassettes for some substrates, similar to many other YcaO enzymes.[36] If so, a similar phenomenon was recently described in another multi-cassette biosynthetic pathway to microviridins.[30] There, the enzyme AMdnC exhibited ordered processing *within* cores and stochastic processing *between* cores. Stochastic processing may thus represent a common mechanism in multi-cassette RiPP pathways.

The next step in the canonical pathway is recognition of RSII and proteolysis by TruA/PatA to remove RSI and free the N-terminus for circularization. Initially, it was observed that TruD strictly requires RSI, and if RSI is removed by PatA prior to TruD modification, no heterocycles are observed.[20] Later studies showed that in certain cases the leader sequence is not strictly required, but that the enzymatic reaction is faster if RSI is provided *in trans*, and faster still if provided *in cis*.[24,32] It might be speculated that this has to do with different enzyme activation barriers under different experimental conditions within these studies. Moreover, in previous studies a strict redox dependence of the PatA and TruD reactions were observed, in which PatA requires a more oxidized environment and TruD requires a more reduced environment.[21,29] PatA is a slow enzyme under reducing conditions and exhibits a C-to-N directionality.[26] In general, our observations here are consistent with these earlier studies. What is new, and what we have learned here concerning the role of multiple cassettes, is that for both heterocyclized and unheterocyclized precursors, the first cassette is favored so cassette-1 and −2 fragments can be generated at the same time. Nonetheless,

other intermediates are detected in some conditions, so that directionality is not strict. Digestion on the TruD time course peptides showed PatA can also fully proteolyze the partially heterocyclized double cassette substrates. The combination of PatA and TruD resemble a metabolic grid, which is known to increase the generation of chemical diversity in other systems.[37] Interestingly, here the metabolic grid is controlled by redox, which might be relevant to the natural biosynthesis in symbiotic cyanobacteria on the coral reef.[38]

The final enzyme that we examined is TruG, which is known to be a highly promiscuous enzyme, capable of circularizing many different substrates.[11,12,29] Intriguingly, it may be the only subtilisin-like protease that recognizes a C-terminal region that is cleaved off during processing (RSIII), rather than the usual S' region. Several studies have examined selectivity of RSIII using the AYD/SYD motif and synthetic variants thereof.[26,28,33–35] Here, for the first time we examine both the RSIII motif that is appended to the second cassette and the RSIII-RSII hybrid motif that is appended to the first cassette. We show that both TruG and PatG essentially equally accept these two natural recognition sequences, and that there is no preferred order of circularization. This result, along with previous studies of isoprenylation, demonstrate that every step in the *tru* pathway, with the exception of PatA/TruA, is stochastic and distributive. It should be noted that the stochastic/distributive model disproves our initial hypothesis that the order of cores would affect natural product yield, and therefore makes a biochemical role for multiple cassettes unlikely.

Given that there is no readily defined biochemical necessity for multiple cassettes, it could be asked why pathways have multiple cassettes. We envision three potential roles: substrate efficiency; substrate evolution; and bioactivity synergy. In terms of substrate efficiency, it is clear that it is biochemically cheaper to make a single peptide encoding multiple products than to have to produce a full-length substrate for each desired product. Since in some of our previous *E. coli* expression experiments, an estimated 25% of the cell dry weight would have been required for precursor peptide synthesis,[16] this could be metabolically costly. The metabolic grid by PatA/TruA and TruD helps to obtain maximum substrate efficiency.[37] Substrate evolution may also drive this phenomenon.[13,24] It is remarkable that within some families of cyanobactins the biosynthetic genes, precursor peptides, and intergenic sequences are essentially identical, while core peptides are hypervariable.[24] The presence of multiple cores could thus be part of a recombinational event underlying this feature. The conserved enzyme recognition sequences flanking the cores might be a selection feature to obtain core modification while maintaining substrate efficiency since the RSs are always present in each cassette. Finally, the biological roles of cyanobactins are poorly known, but it could be speculated that compounds work together to exert a phenotype. In this regard, conopeptides/conotoxins provide a good example,[39] where the biological roles are better defined, wherein multiple toxin peptide act together to generate a defense mechanism.

Given the observed properties of cyanobactin pathways as diversity-generating, we believe that all three factors may be important. Specifically, we have proposed that *tru* and related pathways have evolved to synthesize many derivatives in nature. Indeed, *tru* and related proteins have seen numerous applications to synthetic biology, where they are predicted to make millions of compounds.[12,13] Here, we show that the flexibility of the pathway is reinforced by a mechanism in which each enzyme can act stochastically, forming a

metabolic grid in which all possible intermediates are accepted. This provides a fascinating and unexpected mechanism enabling RiPP pathways to generate chemical diversity (Figure 7 and S14).

Finally, these studies are also useful in defining how to better engineer the increased yield of compounds *in vivo*. During expression of *tru* pathway variants in *E. coli*, we show that production varies in a counterintuitive manner. While simply doubling gene dosage usually doubles the amount of product, (Figure S4) changes made to multiple cassette substrates are not straightforward and must be measured empirically for each substrate. For example, in comparing substrates p2tk versus tkp2, the yield of **2** varies significantly, while the yield of **3** does not. Thus, there is not a predictable trend of yield increase. The duration of culture before harvesting also leads to unpredictable effects (Figure S3), with the optimum harvest time occurring on different days depending upon which substrate is used. Overall, the data indicate that manipulating the cassette dosage, position, or the *E. coli* culture timeline for an individual cyanobactin product can change the yield. This result has implications for synthetic biology, showing that it is worth varying position and dosage to improve the yield of desired products, with the expectation that doubling the cassette dosage will probably double the amount of a compound produced in culture. It also shows that optimizing heterologous expression duration is crucial for maximum production.

CONCLUSION

All roads lead to Rome in *tru* cyanobactin biosynthesis – every biochemically possible series of steps occurs *in vitro*. This property may be intrinsic to biochemical plasticity and diversity-generating biosynthesis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

We thank Alan Maschek and Thomas Smith for help with mass spectra and Maho Morita for helpful discussions.

Funding Sources

This work was funded by NIH GM122521 and GM102602; and a Kuramoto Graduate Research Fellowship to W.G.

REFERENCES

(1). Walsh C (2006) Posttranslational Modification of Proteins: Expanding Nature’s Inventory. W. H. Freeman.

(2). Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian K-D, Fischbach MA, Garavelli JS, Göransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Müller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJT, Rebuffat S, Ross RP, Sahl H-G, Schmidt EW, Selsted ME, Severinov K, Shen B, Sivonen K, Smith L, Stein T, Süssmuth RD, Tagg JR, Tang G-L, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, and van der Donk WA (2013) Ribosomally synthesized and post-translationally modified peptide

natural products: overview and recommendations for a universal nomenclature. Nat. Prod. Rep 30, 108–160. [PubMed: 23165928]

(3). Donia M, Hathaway B, Sudek S, Haygood M, Rosovitz M, Ravel J, and Schmidt E (2006) Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. Nat. Chem. Biol 2, 729–735. [PubMed: 17086177]

(4). Craik DJ, and Malik U (2013) Cyclotide biosynthesis. Curr. Opin. Chem. Biol 17, 546–554. [PubMed: 23809361]

(5). Ziemert N, Ishida K, Liaimer A, Hertweck C, and Dittmann E (2008) Ribosomal synthesis of tricyclic depsipeptides in bloom-forming cyanobacteria. Angew. Chem., Int. Ed 47, 7756–7759.

(6). Ye Y, Minami A, Igarashi Y, Izumikawa M, Umemura M, Nagano N, Machida M, Kawahara T, Shin-ya K, Gomi K, and Oikawa H (2016) Unveiling the Biosynthetic Pathway of the Ribosomally Synthesized and Post-translationally Modified Peptide Ustiloxin B in Filamentous Fungi. Angew. Chem., Int. Ed 55, 8072–8075.

(7). Noike M, Matsui T, Ooya K, Sasaki I, Ohtaki S, Hamano Y, Maruyama C, Ishikawa J, Satoh Y, Ito H, Morita H, and Dairi T (2014) A peptide ligase and the ribosome cooperate to synthesize the peptide pheganomycin. Nat. Chem. Biol 11, 71–76. [PubMed: 25402768]

(8). Shim YY, Young LW, Arnison PG, Gilding E, and Reaney MJT (2015) Proposed systematic nomenclature for orbitides. J. Nat. Prod 78, 645–652. [PubMed: 25785712]

(9). Ding W, Liu W, Jia Y, Li Y, van der Donk WA, and Zhang Q (2016) Biosynthetic investigation of phomopsins reveals a widespread pathway for ribosomal natural products in Ascomycetes. Proc. Natl. Acad. Sci. U. S. A 113, 3521–3526. [PubMed: 26979951]

(10). Donia MS, Ravel J, and Schmidt EW (2008) A global assembly line for cyanobactins. Nat. Chem. Biol 4, 341–343. [PubMed: 18425112]

(11). Tianero MDB, Donia MS, Young TS, Schultz PG, and Schmidt EW (2012) Ribosomal route to small-molecule diversity. J. Am. Chem. Soc 134, 418–25. [PubMed: 22107593]

(12). Ruffner DE, Schmidt EW, and Heemstra JR (2015) Assessing the Combinatorial Potential of the RiPP Cyanobactin *tru* Pathway. ACS Synth Biol 4, 482–492. [PubMed: 25140729]

(13). Gu W, and Schmidt EW (2017) Three Principles of Diversity-Generating Biosynthesis. Acc. Chem. Res 50, 2569–2576. [PubMed: 28891639]

(14). Yamauchi M, and Sricholpech M (2012) Lysine post-translational modifications of collagen. Essays Biochem. 52, 113–133. [PubMed: 22708567]

(15). Gorres KL, and Raines RT (2010) Prolyl 4-hydroxylase. Crit. Rev. Biochem. Mol. Biol 45, 106–124. [PubMed: 20199358]

(16). Tianero MD, Pierce E, Raghuraman S, Sardar D, McIntosh JA, Heemstra JR, Schonrock Z, Covington BC, Maschek JA, Cox JE, Bachmann BO, Olivera BM, Ruffner DE, and Schmidt EW (2016) Metabolic model for diversity-generating biosynthesis. Proc. Natl. Acad. Sci. U. S. A 113, 1772–1777. [PubMed: 26831074]

(17). Levengood MR, Patton GC, and Van Der Donk WA (2007) The leader peptide is not required for post-translational modification by lacticin 481 synthetase. J. Am. Chem. Soc 129, 10314–10315. [PubMed: 17676843]

(18). Thibodeaux GN, McClerren AL, Ma Y, Gancayco MR, and Van Der Donk WA (2015) Synergistic binding of the leader and core peptides by the lantibiotic synthetase HalM2. ACS Chem. Biol 10, 970–977. [PubMed: 25619528]

(19). Zhang Q, Yang X, Wang H, and Van Der Donk WA (2014) High divergence of the precursor peptides in combinatorial lanthipeptide biosynthesis. ACS Chem. Biol 9, 2686–2694. [PubMed: 25244001]

(20). McIntosh JA, and Schmidt EW (2010) Marine molecular machines: Heterocyclization in cyanobactin biosynthesis. ChemBioChem 11, 1413–1421. [PubMed: 20540059]

(21). Mcintosh JA, Donia MS, and Schmidt EW (2010) Insights into heterocyclization from two highly similar enzymes. J. Am. Chem. Soc 132, 4089–4091. [PubMed: 20210311]

(22). Koehnke J, Bent AF, Zollman D, Smith K, Houssen WE, Zhu X, Mann G, Lebl T, Scharff R, Shirran S, Botting CH, Jaspars M, Schwarz-Linek U, and Naismith JH (2013) The cyanobactin heterocyclase enzyme: A processive adenylase that operates with a defined order of reaction. Angew. Chem., Int. Ed 52, 13991–13996.

(23). Goto Y, Ito Y, Kato Y, Tsunoda S, and Suga H (2014) One-pot synthesis of azoline-containing peptides in a cell-free translation system integrated with a posttranslational cyclodehydratase. Chem. Biol 21, 766–774. [PubMed: 24856821]

(24). Sardar D, Pierce E, McIntosh JA, and Schmidt EW (2015) Recognition sequences and substrate evolution in cyanobactin biosynthesis. ACS Synth. Biol 4, 167–176. [PubMed: 24625112]

(25). Agarwal V, Pierce E, McIntosh J, Schmidt EW, and Nair SK (2012) Structures of cyanobactin maturation enzymes define a family of transamidating proteases. Chem. Biol 19, 1411–1422. [PubMed: 23177196]

(26). Lee J, Mcintosh J, Hathaway BJ, and Schmidt EW (2009) Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. J. Am. Chem. Soc 131, 2122–2124. [PubMed: 19166292]

(27). Koehnke J, Bent A, Houssen WE, Zollman D, Morawitz F, Shirran S, Vendome J, Nneoyiegbe AF, Trembleau L, Botting CH, Smith MCM, Jaspars M, and Naismith JH (2012) The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. Nat. Struct. Mol. Biol 19, 767–772. [PubMed: 22796963]

(28). McIntosh JA, Robertson CR, Agarwal V, Nair SK, Bulaj GW, and Schmidt EW (2010) Circular logic: Nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst. J. Am. Chem. Soc 132, 15499–15501. [PubMed: 20961047]

(29). Sardar D, Lin Z, and Schmidt EW (2015) Modularity of RiPP Enzymes Enables Designed Synthesis of Decorated Peptides. Chem. Biol 22, 907–916. [PubMed: 26165156]

(30). Zhang Y, Li K, Yang G, McBride JL, Bruner SD, and Ding Y (2018) A distributive peptide cyclase processes multiple microviridin core peptides within a single polypeptide substrate. Nat. Commun 9, 1780. [PubMed: 29725007]

(31). Tianero MD, Pierce E, Raghuraman S, Sardar D, McIntosh JA, Heemstra JR, Schonrock Z, Covington BC, Maschek JA, Cox JE, Bachmann BO, Olivera BM, Ruffner DE, and Schmidt EW (2016) Metabolic model for diversity-generating biosynthesis. Proc. Natl. Acad. Sci. U. S. A 113, 1772–1777. [PubMed: 26831074]

(32). Koehnke J, Mann G, Bent AF, Ludewig H, Shirran S, Botting C, Lebl T, Houssen WE, Jaspars M, and Naismith JH (2015) Structural analysis of leader peptide binding enables leader-free cyanobactin processing. Nat. Chem. Biol 11, 558–563. [PubMed: 26098679]

(33). Oueis E, Jaspars M, Westwood NJ, and Naismith JH (2016) Enzymatic Macrocyclization of 1,2,3-Triazole Peptide Mimetics. Angew. Chem., Int. Ed 55, 5842–5845.

(34). Oueis E, Stevenson H, Jaspars M, Westwood NJ, and Naismith JH (2017) Bypassing the proline/thiazoline requirement of the macrocyclase PatG. Chem. Commun 53, 12274–12277.

(35). Oueis E, Nardone B, Jaspars M, Westwood NJ, and Naismith JH (2017) Synthesis of Hybrid Cyclopeptides through Enzymatic Macrocyclization. ChemistryOpen 6, 11–14. [PubMed: 28168143]

(36). Burkhart BJ, Schwalen CJ, Mann G, Naismith JH, and Mitchell DA (2017) YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function. Chem. Rev 117, 5389–5456. [PubMed: 28256131]

(37). Firn RD, and Jones CG (2003) Natural products - a simple model to explain chemical diversity. Nat. Prod. Rep 20, 382. [PubMed: 12964834]

(38). Behrendt L, Raina J-B, Lutz A, Kot W, Albertsen M, Halkjxr-Nielsen P, Sørensen SJ, Larkum AW, and Kühl M (2018) In situ metabolomic- and transcriptomic-profiling of the host-associated cyanobacteria Prochloron and Acaryochloris marina. ISME J. 12, 556–567.

(39). Teichert RW, Schmidt EW, and Olivera BM (2015) Constellation pharmacology: a new paradigm for drug discovery. Annu. Rev. Pharmacol. Toxicol 55, 573–89. [PubMed: 25562646]

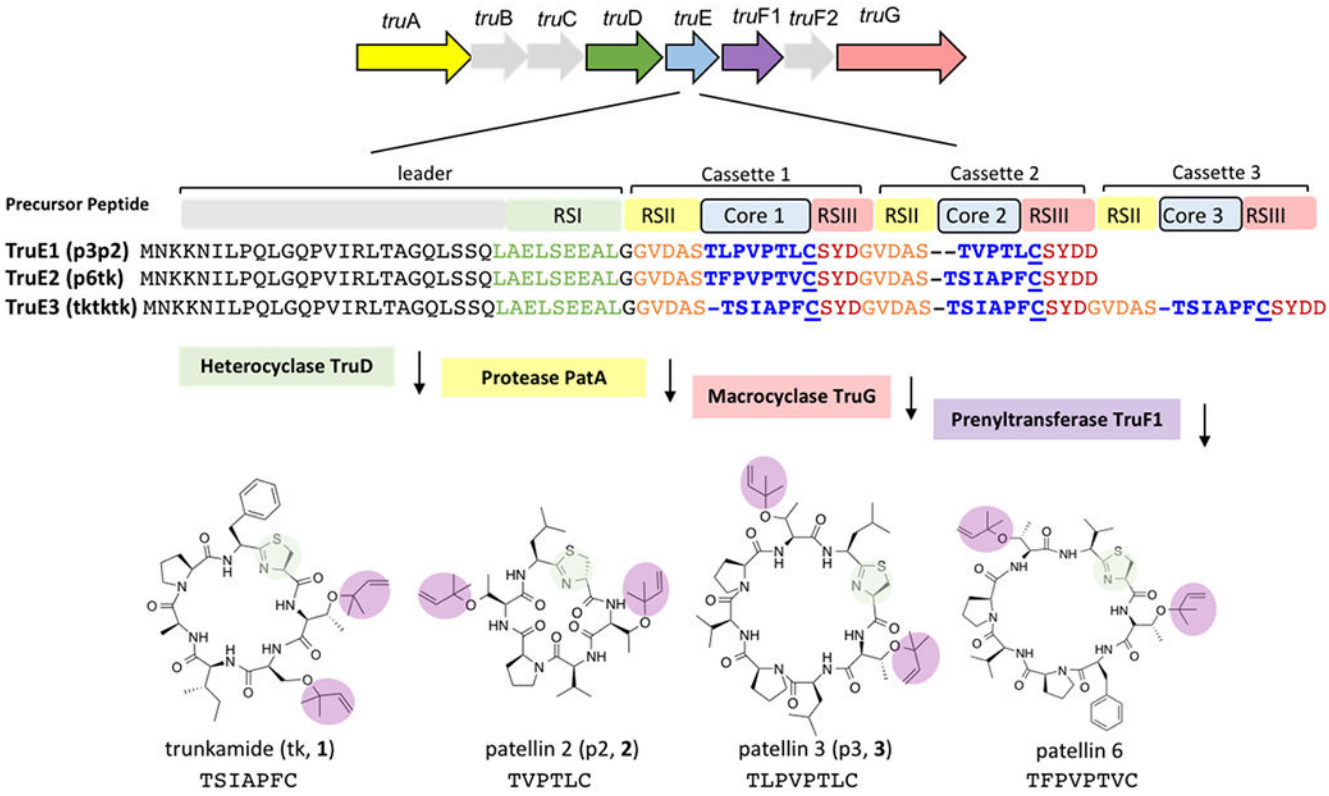


Figure 1.

tru pathway. The pathway genes encode modifying enzymes and precursor. Three different TruE precursors were discovered in nature. After PTMs, cyanobactins trunkamide (tk, **1**), patellins 2 (p2, **2**), 3 (p3, **3**) and 6 are produced.

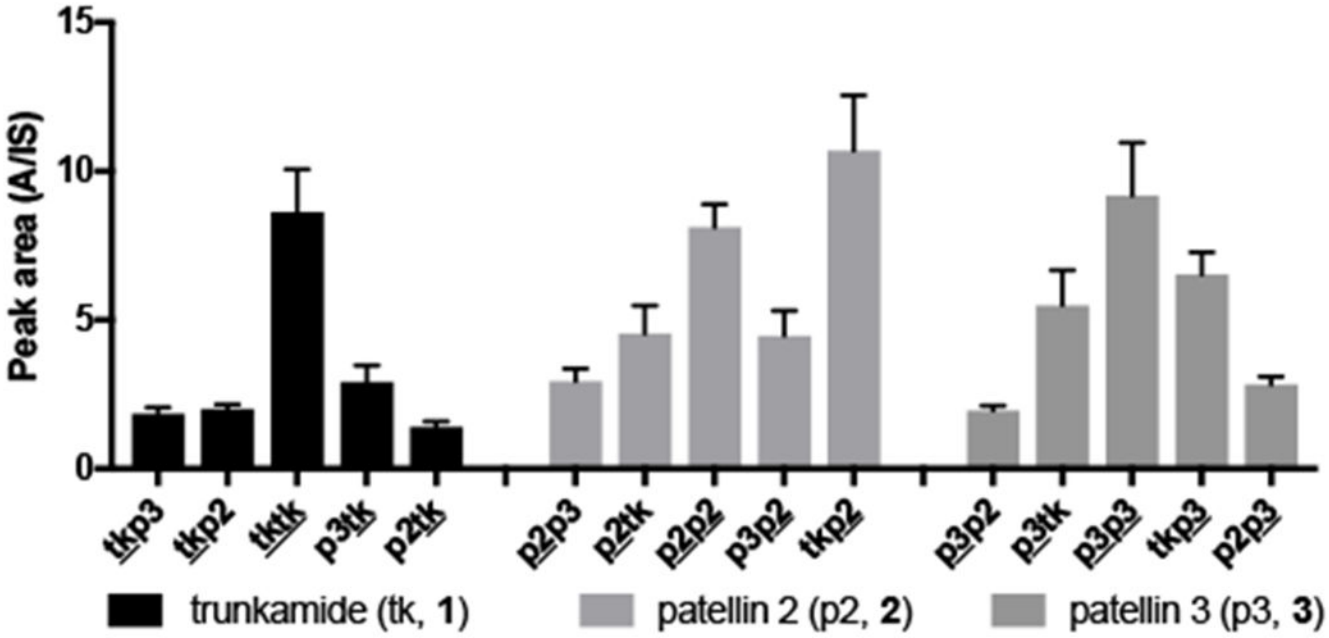


Figure 2.

Sum of compound production of prenylated, single prenylated and non-prenylated products on day three. Compound yields were determined by averaging the extracted areas under the curve, normalized to an internal standard. Although there are ionization differences between products, by using the same products in each condition, their relative abundance is normalized.

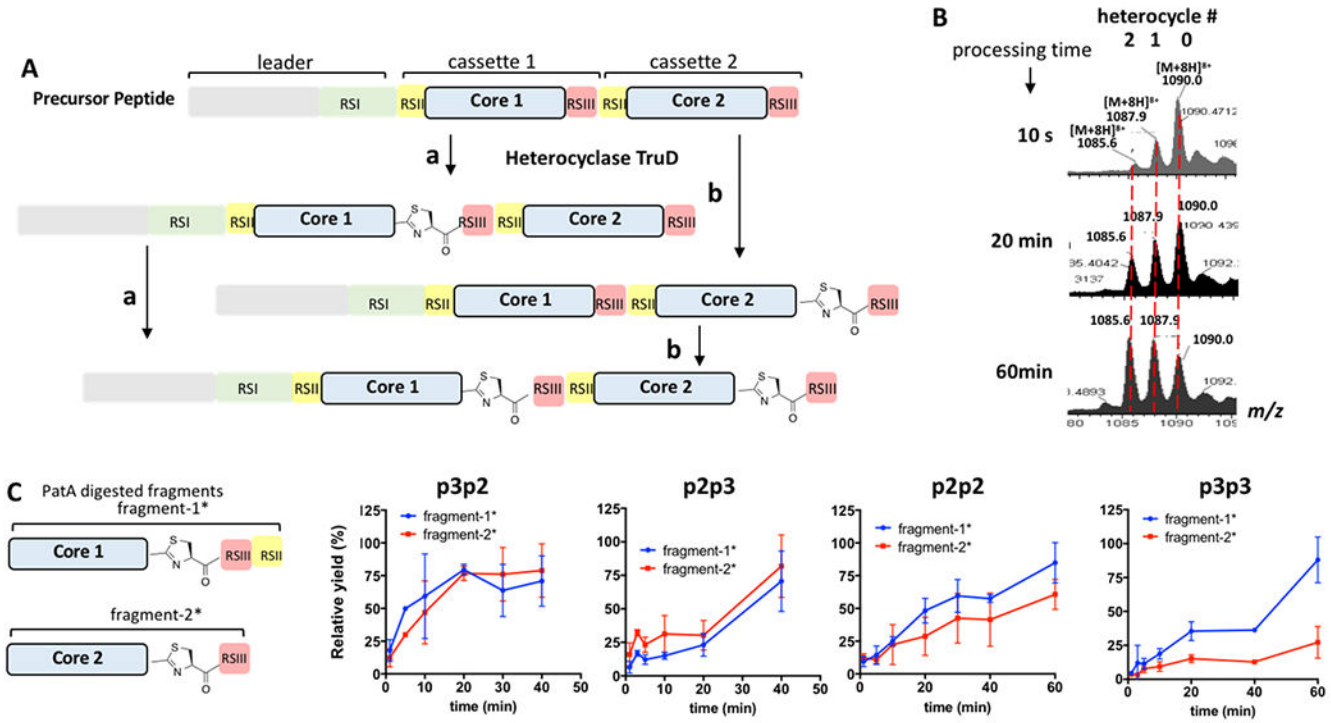


Figure 3.

TruD is distributive and stochastic. **(A)** Two possible routes **a** and **b** for TruD heterocyclization, which were both observed from the experiments. **(B)** MS spectra of precursor peptide TruE-p2p2 in reaction with TruD. Time points were taken at 10 s, 1, 5, 10, 20, 30, 40 and 60 min, of which a subset is shown here. The number of heterocycles is indicated above the red dashed lines. **(C)** Extent of modification in cassette-1 or −2. Precursor peptides were treated with TruD for the indicated times on x-axis, then digested to completion with PatA, allowing identification of fragment-1* and fragment-2* (*indicates a heterocycle) by mass. The amount of each fragment that was heterocyclized is shown as area under the curve, which was normalized to internal standards and corrected for ionization differences using standards of known quantity. The peak area was then normalized to the last time point taken.

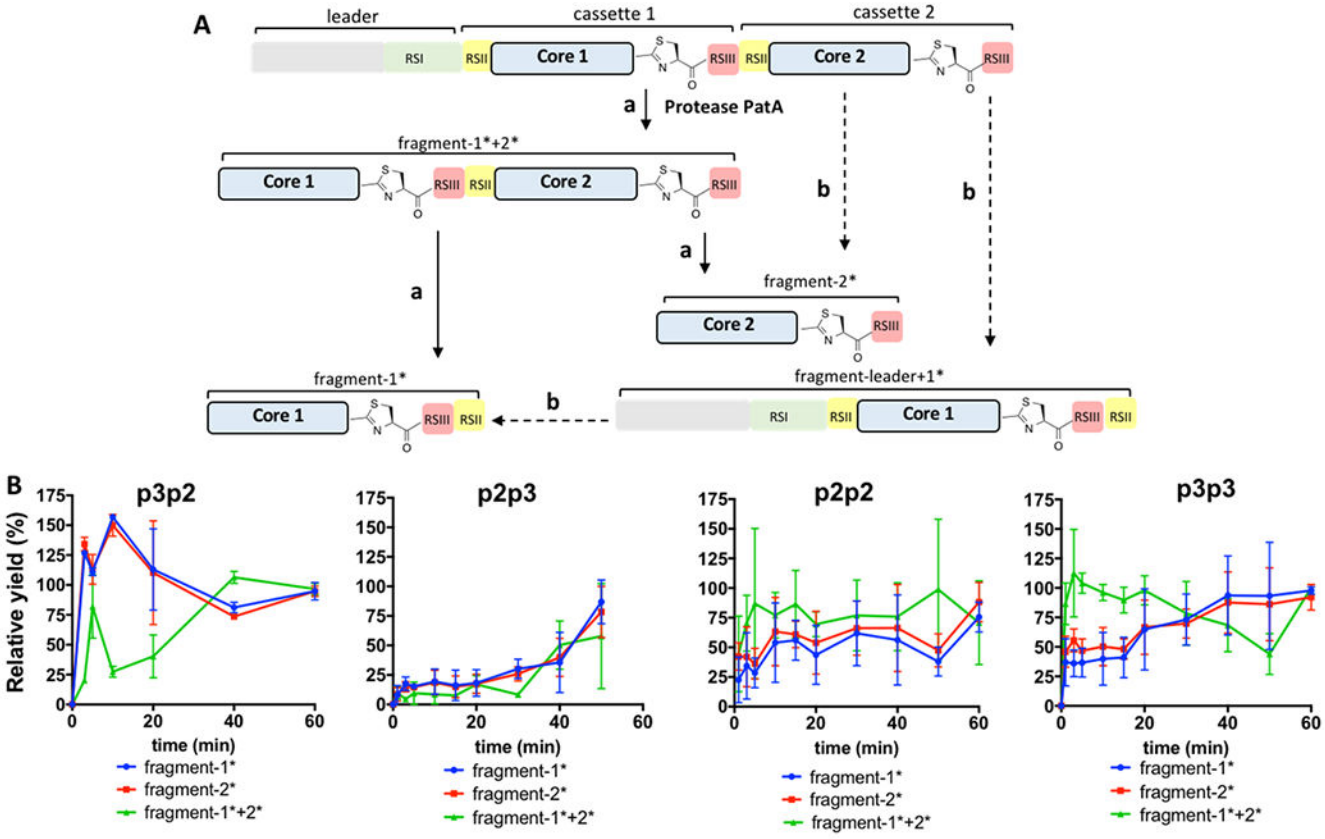


Figure 4.

PatA prefers cassette-1 when proteolyzing heterocyclized precursor peptides. **(A)** Proteolysis steps. **(B)** Extracted peak area is normalized to internal standards. Fragment-1* and -2* were also normalized to two different synthetic peptides to correct for ionization differences. The peak area was then normalized to the last time point taken as the relative yield percentage.

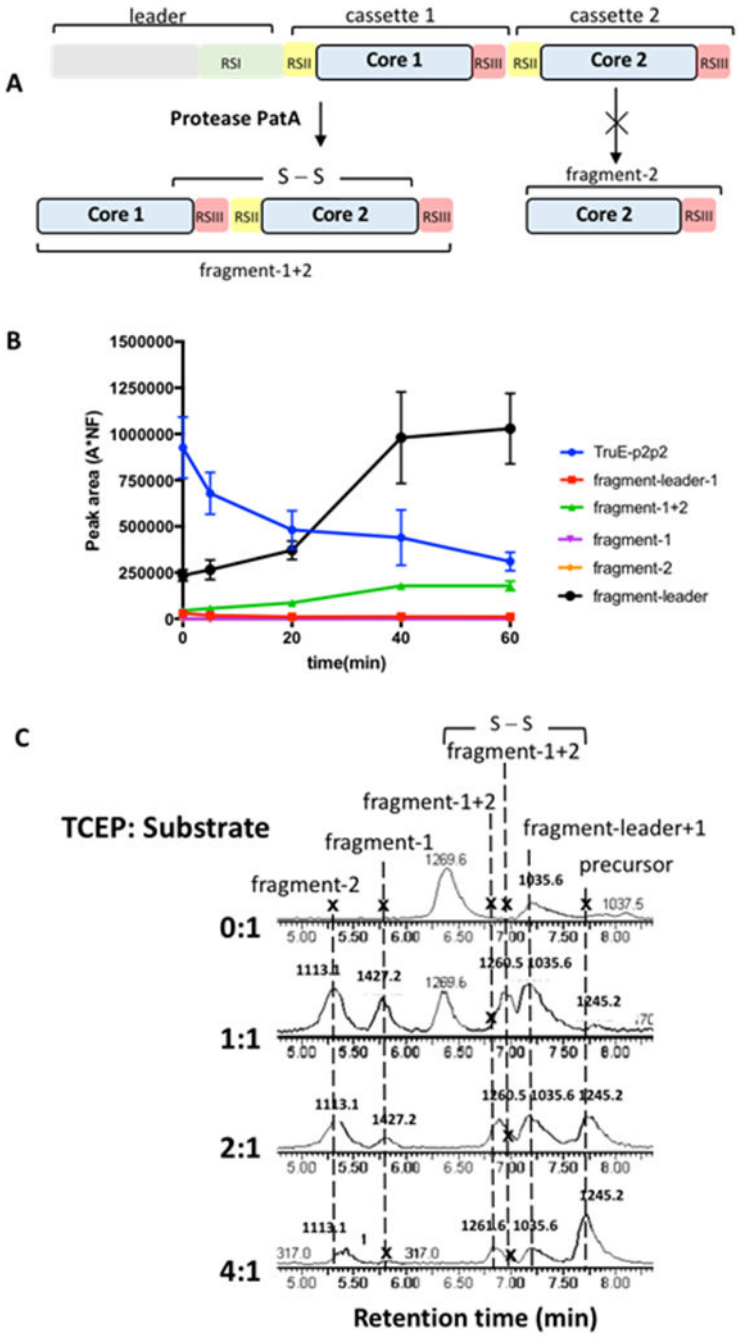


Figure 5.

PatA proteolysis over intact precursor peptides was inhibited by disulfide bonds between cassettes. **(A)** Proteolysis steps. **(B)** In the absence of added reductant, proteolysis on cassette-2 was inhibited by an intramolecular disulfide bridge. **(C)** 3 h incubation of the substrate and PatA with and without TCEP. **x** indicates that the corresponding species is not observed in the MS trace. Without TCEP, no fragment-1 or -2 was observed. With 1:1 TCEP:substrate molar ratio, the disulfide bridge is broken, and proteolysis of cassette-1 and -2 is restored. Above 1:1 ratio, the enzyme is inhibited by the reducing agent. The extra peak

(*m/z* 1269.6) in 0:1 and 1:1 ratio mass trace matches the sulfenic acid form of the thiol group from cysteine.

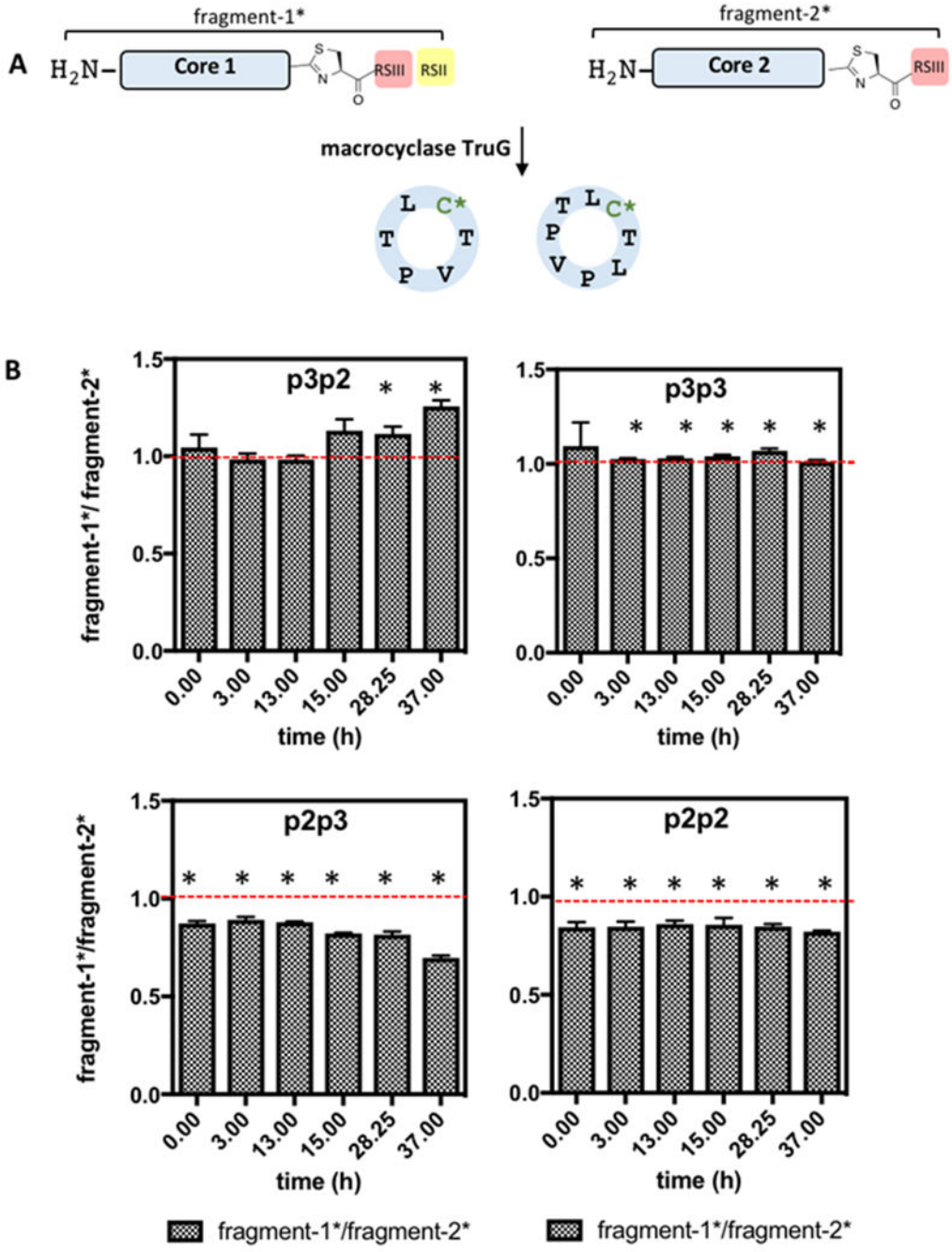


Figure 6.

TruG reaction with various substrates. **(A)** Macrocyclization steps. **(B)** Cassette-1 and cassette-2 substrate ratio with different combinations: p3p2, p3p3, p2p3 and p2p2. Each time point is hypothesis tested with one sample *t* test over 1 and the significant values are marked with *.

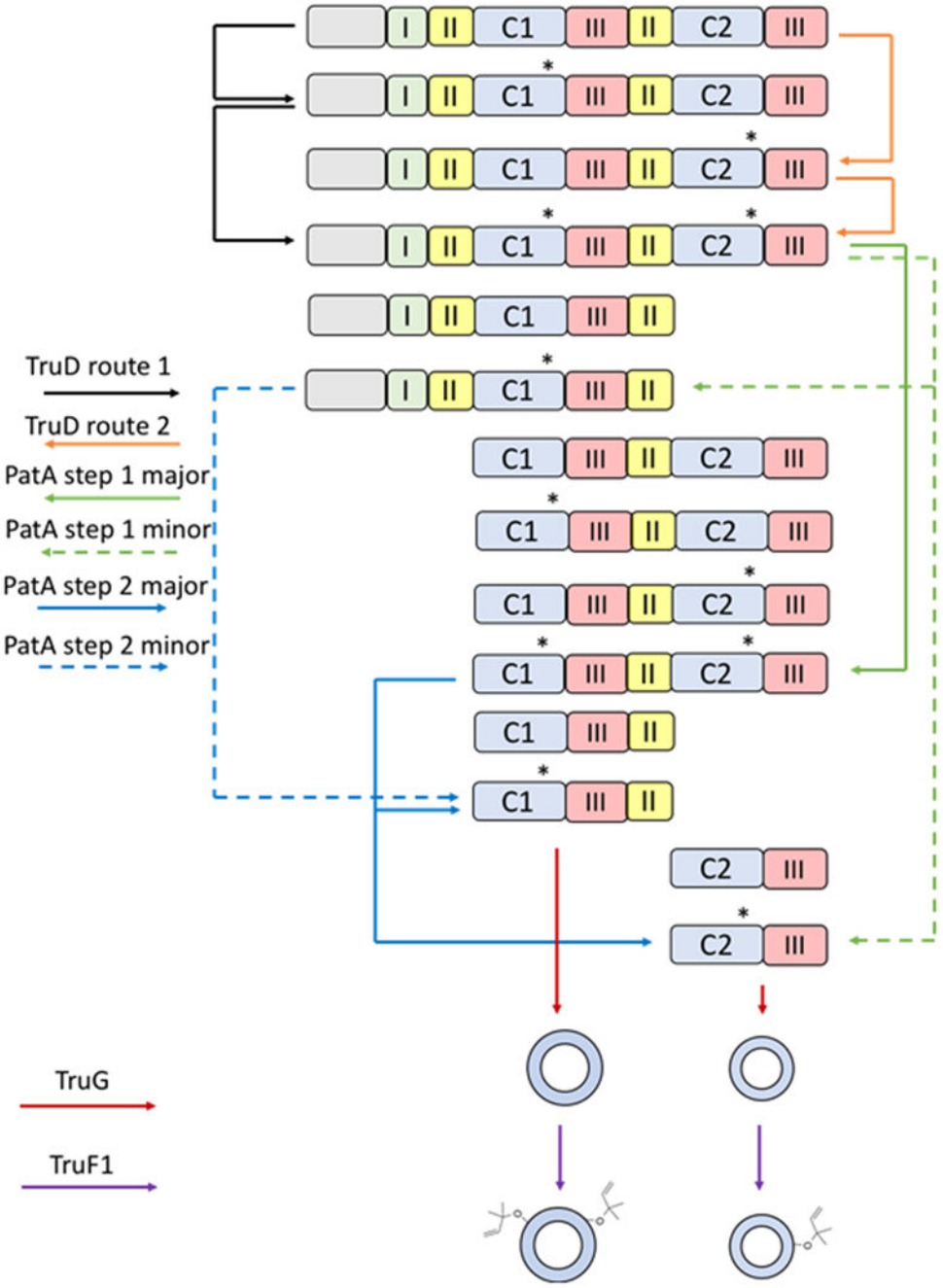


Figure 7.
Biosynthetic pathway model for a double cassette precursor peptide. * indicates a heterocycle.