Structure prediction and synthesis of a new class of macrocyclic peptide natural products

Graham A. Hudson†, Annie R. Hooper†, Adam J. DiCaprio‡, David Sarlah†, Douglas A. Mitchell†‡*  

†Department of Chemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, USA.  
‡Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 West Gregory Drive, Urbana, Illinois 61801, USA.

ABSTRACT

Owing to advances in genomic sequencing and bioinformatics, the breadth of natural product biosynthetic gene clusters (BGCs) has meteorically risen. This remains true for ribosomally synthesized and post-translationally modified peptides (RiPPs), where the rate of bioinformatically identifying clusters vastly outpaces characterization efforts. Uniting bioinformatics and enzymological knowledge to predict the chemical product(s) of a RiPP BGC with total chemical synthesis to obtain the natural compound is an effective platform for investigating cryptic gene clusters. Herein, we report the bioinformatic identification of a biosynthetically divergent class of RiPP bearing a subset of enzymes involved in thiopeptide biosynthesis. These natural products were predicted based on BGC architecture to undergo a formal, enzymatic [4+2]-cycloaddition with subsequent elimination of the leader peptide and water to produce a tri-substituted pyridine-based macrocycle. Bearing a pyridine similar to thiopeptides but lacking the ubiquitous thiazole heterocycles, these new RiPPs were termed pyritides. One of the predicted natural products was chemically synthesized using an 11-step synthesis. This structure was verified to be chemically identical by an orthogonal chemoenzymatic synthesis utilizing the precursor peptide and the cognate [4+2]-cycloaddition enzyme. The chemoenzymatic platform was used to synthesize a second in-cluster pyritide product as well as analogs from other bioinformatically identified pyritide BGCs. This work exemplifies complementary bioinformatic, enzymological, and synthetic techniques to characterize a structurally distinct class of RiPP natural product.

INTRODUCTION

Advances in sequencing technology and analysis have rapidly expanded our knowledge of natural product biosynthetic space.1,2 This is notably true for the ribosomally synthesized and post-translationally modified peptide (RiPP) natural products, where characterization of new natural products is extensively outpaced by an ever-increasing amount of sequenced biosynthetic gene clusters (BGCs).3–6 Characterizing novel compounds often involves laborious and time-consuming techniques endemic to natural product discovery, such as the traditional “grind-and-find” approach or targeted heterologous expression of BGCs in tractable hosts.7 Given sufficient knowledge of the enzymes involved in biosynthetic maturation, the structure of a mature natural product can often be predicted, permitting access to the natural compound via
chemical synthesis. This approach provides a new entry point in exploring natural product biosynthetic space. RiPPs are attractive targets for genome-guided total synthesis, as their structures are more readily predicted than other natural product classes given their direct ribosomal origin. Synthetic approaches also boast unparalleled capacity for generating analogs, often with significant scalability, obviating potential issues with isolation from producing organisms/hosts.

Recently, we reported a survey of the biosynthetic space for the thiopeptides, an extensively modified RiPP class. All thiopeptides feature azol(in)e heterocycles, dehydroamino acids, and a (dehydro)piperidine or pyridine heterocycle, but many contain a significant number of ancillary modifications beyond these three ubiquitous post-translational modifications (PTMs). The class-defining PTM for thiopeptides is the 6-membered nitrogenous heterocycle, which arises from a formal [4+2]-cycloaddition. We sought to determine if homologs of the thiopeptide thiomuracin [4+2]-cycloaddition enzyme, TbtD, occurred in any context beyond thiopeptide BGCs and performed a bioinformatic search for homologs that did not cooccur with the normal repertoire of thiopeptide enzymes. This search yielded several BGCs lacking a local azoline-forming cyclodehydratase while still possessing a split LanB, putatively responsible for Ser/Thr glutamylation and elimination to yield dehydroalanine (Dha) or dehydrobutyrine (Dhb), respectively, as well as a [4+2]-cycloaddition enzyme (Figures 1A and S2). The precursor peptides for these BGCs generally lacked Cys. Therefore, we reasoned that these BGCs would produce a structurally unique class of RiPP featuring a tri-substituted pyridine-based macrocycle and termed this still theoretical class as "pyritides."

Figure 1. Biosynthesis of pyritides. (A) BGC from *Micromonospora rosaria*, which encodes a split LanB dehydratase and [4+2]-cycloaddition enzyme and notably lacks an azoline-forming cyclodehydratase. (B) Hypothetical biosynthetic route to produce pyritide A1 (1). (C) Predicted structure of pyritide A2 (2). LP, leader peptide.

We endeavored to test our structural hypothesis using a pyritide BGC identified in *Micromonospora rosaria* owing to strain availability and the fact that the two precursor peptides each contained only two Ser residues
(Figure 1A). Thus, we felt confident in the structural prediction that both Ser would be converted to Dha prior to the formation of 14- and 17-atom macrocyclic pyritides by the [4+2]-cycloaddition enzyme, yielding pyritide A1 (1) and pyritide A2 (2), respectively (Figure 1B-C). Notably, 1 would contain a macrocycle 12-atoms smaller than any characterized thiopeptide. As the predicted structure of the final natural products would be unambiguous, these structures would also be amenable to total chemical synthesis using the genome-guided structural prediction.8–10

RESULTS

Using the predicted structure of 1, a synthetic scheme consisting of 11 steps was devised (Figure 2A). To assess the validity of the structural prediction, we attempted to isolate 1 from the native producer; however, upon testing many different cultivation conditions, we were unsuccessful in eliciting production of the compound. Instead, building on our earlier work on thiomuracin biosynthesis,15,16 we opted to carry out a similar in vitro reconstitution of 1 using purified recombinant enzymes. In analogy to thiomuracin biosynthesis, we surmised Dha formation would require a tRNA-dependent glutamylation of Ser with subsequent elimination. In our hands, the enzymes responsible for this transformation have presented significant challenges for in vitro reconstitution.15,16 To obviate this issue and to leverage the fact that the precursor peptide of 1 (MroA1) lacked Cys, we hypothesized that replacement of Ser1 and Ser6 residues with Cys would enable a convenient desulfurization using dibromohexanediamide (3, DBHDA).20–23 The product of this reaction would be the putative substrate for an MroD-catalyzed [4+2]-cycloaddition (Figure 2B).

Figure 2. Orthogonal routes to obtain pyritide A1 (1). (A) Chemical synthesis beginning from pyridine-2,5-dicarboxylic acid and protected tripeptides. See Figure S3 for the enumerated route. (B) Chemoenzymatic synthesis using MroA1 precursor isolated from E. coli heterologous expression. The two Ser appearing in the wild-type sequence have been substituted with Cys to facilitate chemical desulfurization with DBHDA (3) and subsequent enzymatic [4+2]-cycloaddition at the resultant Dha residues.
Pursuant to this, the maltose-binding protein (MBP)-tagged MroA1 precursor peptide featuring double substitution of Ser to Cys (4) as well as MroD were separately expressed in Escherichia coli and purified by affinity chromatography (Figure S1). After treatment with 3, we noted the formation of the doubly desulfurized species (5) by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS, Figure 3). 5 was then presented to MroD, which resulted in the appearance of two new masses congruent with the predicted structure of 1 and the leader peptide with a C-terminal carboxamide, 6 (Figure 3).

![Figure 3. MALDI-TOF mass spectra detailing the chemoenzymatic synthesis of 1. The doubly Cys-substituted precursor peptide (4, m/z 5978 Da) was chemically desulfurized (5, m/z 5910 Da) and subsequently reacted with the [4+2]-cycloaddition enzyme MroD. The resulting masses correspond to the [M+H]^+ ions predicted for 1 (m/z 933 Da) and the eliminated carboxamide leader peptide (6, m/z 4959 Da).](image)

The above chemoenzymatic route was then carried out on a larger scale to produce sufficient quantities of 1 to allow a detailed spectroscopic comparison to the synthetic standard. Approximately 20 L of MBP-MroA1 and 8 L of MBP-MroD were used to produce ~2.5 mg of 1. Several analytical techniques were used to characterize the synthetic and chemoenzymatic samples of 1, including high-resolution and tandem mass spectrometry (HR-MS/MS), C\textsubscript{18} analytical HPLC retention times, and ^1H NMR (Figures S4-S8, S16). Data from all three methods confirmed that the structure of 1 was correctly predicted and the synthetic and chemoenzymatically prepared samples were chemically indistinguishable.

Using the chemoenzymatic route, we next sought to obtain a serviceable quantity of pyritide A2 (2, Figure 1C). The principle difference between 1 and 2 is the presence of an Arg in 2 adjacent to the pyridine. This position is conserved in several other bioinformatically predicted pyritides (Figure S3). Akin to 1, the MroA2 precursor was doubly substituted with Cys at both Ser positions, heterologously expressed, purified,
and chemically desulfurized by 3. This resulted in the production of a mass consistent with the predicted structure of 2, which was verified by \(^1\)H NMR and HR-MS/MS (Figures S9-S12), and the ejected carboxamide leader peptide 6. Analysis by HR-MS/MS confirmed that the leader peptide of MroA2 was eliminated as a C-terminal carboxamide (Figure S13). Purified 1 and 2 were then subjected to a brief panel of growth-suppression assays; however, we did not detect such activity for either pyritide tested alone or in combination (Table S2).

We next assessed the robustness of the chemoenzymatic route to determine if the MroD [4+2]-cycloaddition enzyme could provide access to additional predicted pyritides without requiring the native producer. Thus, we prepared chimeric precursor peptides consisting of the leader region of MroA2 fused to the core region of other predicted precursors with the Ser-to-Cys replacement for positions that form the pyridine (Table S1). In the case *Nonomuraea* sp. WAC 01424, the precursor contained Cys, thus this residue was replaced with Ser to prevent over-desulfurization. These chimeric substrate peptides were heterologously expressed, purified, chemically desulfurized, and reacted with MroD. Upon MS analysis, we observed a pyridine-containing product for only two of the four chimeric substrates tested. The examples from *Micromonospora yangpuensis* and *Nonomuraea* sp. WAC 01424 were successful and produced analogs 7 and 8 respectively, while those from *Herbidospora mongoliensis* and *Herbidospora yilanensis* failed to generate product. For the former examples, HR-MS/MS was used to verify the composition of the product and the sites of macrocyclization (Figures S14-S15).

![Figure 4](image-url)

**Figure 4.** Pyritide compounds produced through chemoenzymatic synthesis utilizing chimeric precursor peptides. (A) Structure of pyritide Y (7) from *Micromonospora yangpuensis*. Similar to 1 and 2, the structure of 7 can be unambiguously deduced as it only contains two pyridine-bound Ser (orange). (B) Structural analog of the 20-atom pyritide produced by *Nonomuraea* sp. WAC 01424 (8). This compound has numerous additional residues that may be modified (Ser, Thr, and Cys) and thus this structure will not be identical to the natural product. Cys3 was substituted with Ser (blue) to prevent over-desulfurization.

Our data suggest that MroD displays some tolerance towards non-cognate substrates and is capable of producing pyridine-based macrocycles at least in the range of 14 to 20 atoms. Further studies will be required to fully elucidate the substrate tolerance of MroD and other pyritide [4+2]-cycloaddition enzymes. It should be noted that while the chemoenzymatic route offers convenient access to pyritide
analogs, a drawback is the inability to install Dhb at Thr in ribosomally synthesized core peptides. The relative simplicity with which pyritide scaffolds can be accessed through total synthesis suggests an avenue to produce Dhb-containing scaffolds or analogs. The chemoenzymatic approach also relies on assuming the extent of dehydration to a precursor peptide. In the cases of 1, 2, and 7, the resultant structures can be unambiguously deduced as only two Ser are present and both incorporated into pyridine. For other structures containing additional Ser or Thr residues, the natural structures are less obvious and require further exploration. In the unique case of the Nonomuraea precursor peptide, Cys is present and thus caution is warranted, as desulfurization chemistry may lead to an unnatural Dha.

CONCLUSION

The genomics revolution has led to a renaissance in natural product discovery. Massive growth in genome databases combined with the development of utilities to analyze and interpret these data has bolstered our knowledge of natural product biosynthetic space, most of which remains unexplored. Together, these advances have led to the discovery of numerous new natural products and, in the case of RiPPs, new post-translational modifications. Bioinformatics recently has been utilized to discover new RiPP classes, such as the α-keto β-amino acid-containing peptides, RiPPs generating thiaglutamate by peptide-acyl tRNA ligases (PEARLs), aliphatic ether-containing rotapeptides and non-α-thioether-containing ranthipeptides, as well as the pyritides, described herein. Bioinformatics offers a means to prioritize the characterization of divergent BGCs that give rise to novel compounds while also mitigating the historical issue of rediscovery. This paradigm shift is further enhanced by the unification of enzymology and synthetic chemistry as it provides a fruitful platform to access new molecules. Here, we have applied this methodology by identifying unusual context for homologs of thiopeptide [4+2]-cycloaddition enzymes, leading to the structural prediction, chemical synthesis, and enzymatic verification of the founding members of a new RiPP class, the pyritides. We anticipate further investigations into pyritides will yield greater insights into the enzymology and biological function which will undoubtedly be augmented by complementary chemical synthesis. Moreover, we predict that the interplay between natural product enzymology and chemical synthesis will become an increasingly useful means producing new natural products or analogs thereof.

AUTHOR INFORMATION

Corresponding Author
* douglasm@illinois.edu

Author Contributions

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

Funding Sources

This work was supported in part by grants from the National Institutes of Health (GM097142 to D.A.M.).
Notes
The authors declare no competing financial interests.

ACKNOWLEDGMENT
We would like to thank Andraž Oštrek for assistance in synthetic scale up of 1.

REFERENCES


