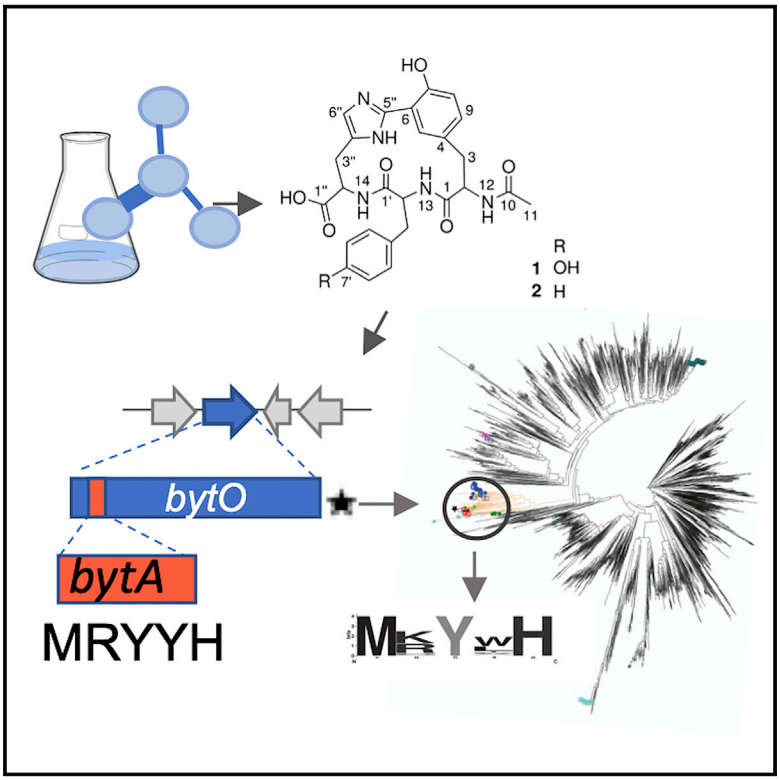
Brief Communication

Cell Chemical Biology

A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters

Graphical Abstract



Highlights

- N-Acetylated tripeptides with an unusual Tyr-His biaryl bridging
- Ribosomally synthetized from a pentapeptide precursor and modified by a cytochrome P450
- Encoded by the 18-bp *bycA* gene, the smallest coding gene ever reported
- Widespread distribution of related two-gene clusters in actinobacterial genomes

Authors

Mitja M. Zdouc,
Mohammad M. Alanjary,
Guadalupe S. Zarazúa, ...,
Marnix H. Medema, Stefano Donadio,
Margherita Sosio

Correspondence

mzdouc@naicons.com (M.M.Z.),
msosio@naicons.com (M.S.)

In Brief

Zdouc et al. describe the discovery of two crosslinked, cyclic tripeptides with unusual Tyr-His biaryl bridging. Their formation requires a minimal biosynthetic gene cluster consisting of a gene encoding a pentapeptide precursor and a gene encoding cytochrome P450 monooxygenase. Phylogenomic analysis revealed widespread distribution of related two-gene clusters in actinobacterial genomes.

Zdouc et al., 2021, Cell Chemical Biology *28*, 1–7
March 18, 2021
https://doi.org/10.1016/j.chembiol.2020.11.009

CellPress

Please cite this article in press as: Zdouc et al., A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters, Cell Chemical Biology (2020), https://doi.org/10.1016/j.chembiol.2020.11.009

Cell Chemical Biology

CellPress

Brief Communication

A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters

Mitja M. Zdouc,[1,2,*] Mohammad M. Alanjary,[3] Guadalupe S. Zarazúa,[4] Sonia I. Maffioli,[1] Max Crüsemann,[4] Marnix H. Medema,[3] Stefano Donadio,[1] and Margherita Sosio[1,5,*]

[1]Naicons Srl., Viale Ortles 22/4, 20139 Milano, Italy

[2]Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, Amsterdam, XH 1098, the Netherlands

[3]Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, Wageningen, PB 6708, the Netherlands

[4]Institut für Pharmazeutische Biologie, Rheinische Friedrich-Wilhelms-Universität, Nußallee 6, Bonn 53115, Germany

[5]Lead Contact

*Correspondence: mzdouc@naicons.com (M.M.Z.), msosio@naicons.com (M.S.)

https://doi.org/10.1016/j.chembiol.2020.11.009

SUMMARY

Microbial natural products impress by their bioactivity, structural diversity, and ingenious biosynthesis. While screening the less exploited actinobacterial genus *Planomonospora*, two cyclopeptides were discovered, featuring an unusual Tyr-His biaryl bridging across a tripeptide scaffold, with the sequences N-acetyl-Tyr-Tyr-His and N-acetyl-Tyr-Phe-His. *Planomonospora* genomes pointed toward a ribosomal synthesis of the cyclopeptide from a pentapeptide precursor encoded by 18-bp *bytA*, to our knowledge the smallest coding gene ever reported. Closely linked to *bytA* is *bytO*, encoding a cytochrome P450 monooxygenase likely responsible for biaryl installment. In *Streptomyces*, the *bytAO* segment was sufficient to direct production of the crosslinked N-acetylated Tyr-Tyr-His tripeptide. Bioinformatic analysis of related cytochrome P450 monooxygenases indicated that they constitute a widespread family of enzymes, and the corresponding genes are closely linked to 5-amino acid coding sequences in approximately 200 (actino)bacterial genomes, all with potential for biaryl linkage between amino acids 1 and 3. We propose the named biarylitides this family of RiPPs.

INTRODUCTION

Microbial natural products impress not only for their potent bioactivity, but also for their chemical diversity and originality. Structurally novel compounds have a high chance of interacting with novel targets or with mechanisms of action different from known compounds (Yera et al., 2011; Gfeller et al., 2013; Chen et al., 2020). In recent years, ribosomally synthesized and post-translationally modified peptides (RiPPs) have drawn attention due to their intricate, three-dimensional (3D) structures and wide distribution in bacteria. RiPP biosynthetic gene clusters (BGCs) are compact, encompassing only a precursor peptide, separated in an N-terminal leader and a C-terminal core peptide. After ribosomal translation, the precursor peptide is modified by dedicated enzymes, with consecutive cleavage of the leader peptide and release of the mature core peptide (Arnison, et al., 2013; Funk and van der Donk, 2017). RiPPs are grouped into families on the basis of their structural features and post-translational modifications (Funk and van der Donk, 2017). Several bioinformatic tools have been developed for the automatic recognition of RiPP BGCs in microbial genomes, which are usually based on recognizing elements, such as leader peptides or diagnostic RiPP-processing enzymes (Russell and Truman, 2020).

In this study, we identified two crosslinked, N-acetylated tripeptides featuring an unusual Tyr1-His3 biaryl bridging. Genome analysis of the producing strains led to the identification of the shortest RiPP BGC so far reported, encoding just a pentapeptide precursor and a closely linked cytochrome P450 monooxygenase. Phylogenomic analysis of the latter sequence revealed that cytochrome P450 monooxygenase genes closely linked to a 5-amino acid coding sequence represent a widespread family of minimal RiPP BGCs present especially in actinobacterial genomes.

RESULTS AND DISCUSSION

Identification of crosslinked N-Acetylated Tyr-Tyr-His

In the search for microbial secondary metabolites, we investigated 72 strains belonging to the poorly characterized actinomycete genus *Planomonospora* (Monciardini et al., 2014) by extensive liquid chromatography coupled to high-resolution ion electrospray ionization-tandem mass spectrometry-based metabolome mining of extracts (Zdouc et al., 2020). Molecular networking analysis revealed a group of related features (see Figure S1) observed in samples from four phylogenetically related strains, with 16S rRNA gene sequences having 99.5%–99.6% identity with *Planomonospora algeriensis* PM3

Please cite this article in press as: Zdouc et al., A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters, Cell Chemical Biology (2020), https://doi.org/10.1016/j.chembiol.2020.11.009

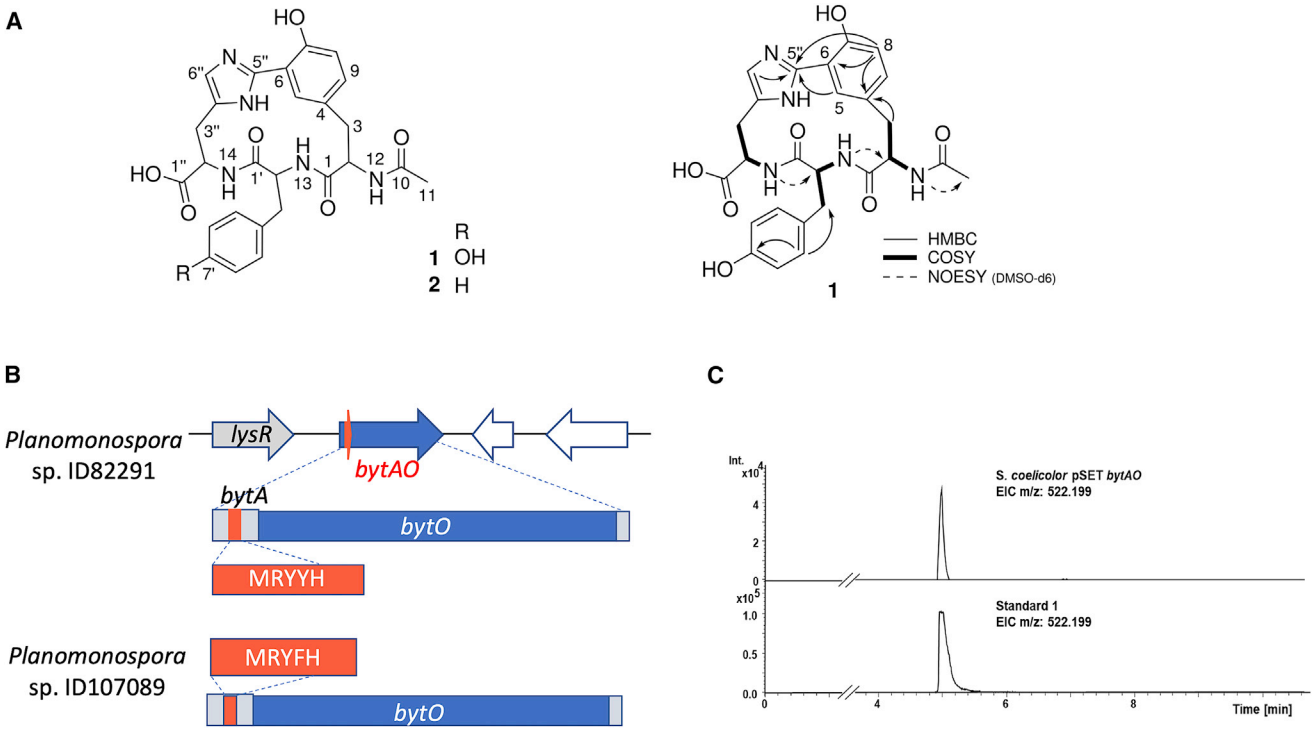


Figure 1. Chemical structures of 1 and 2, the *bytAO* region for 1 and 2 biosynthesis, and its heterologous expression

(A) Structures of 1 and 2 with key 2D NMR correlations for 1 (right side); (B) *bytA* encodes the 5-amino acid precursor peptide, while *bytO* encodes the cytochrome P450 monooxygenase PLM4_2056. A *lysR* family transcriptional regulator is in the vicinity of *bytAO* and the two empty arrows represent genes outside the 1 biosynthetic cluster. The corresponding segment from the genome of *Planomonospora* sp. ID107089 is shown at the bottom. Note amino acid difference in BytA; (C) extracted ion chromatograms (EIC) from heterologous expression of *bytAO* in *Streptomyces coelicolor* M1152 (top) and an authentic standard of 1 (bottom). See also Figures S1 and S4 and Table S1.

(Zdouc et al., 2020). One strain, *Planomonospora* sp. ID82291, produced a compound (**1**) with an *m/z* 522.198 $[M + H]^+$ (calculated molecular formula $C_{26}H_{28}N_5O_7$), while the other three strains produced a compound (**2**) with an *m/z* 506.203 $[M + H]^+$ (calculated molecular formula $C_{26}H_{28}N_5O_6$). Additional variants were found only in extracts from strain ID46116 (Figure S1) and were not further investigated. **1** and **2** showed UV absorption maxima at 270 and 315 nm, the latter indicating an unusual chromophore.

Planomonospora sp. ID82291 was selected for further investigation. From a 15-L culture, 5 mg pure **1** were isolated. Extensive 1D and 2D NMR analyses indicated an acetylated short peptide. A first set of experiments, performed in DMSO-*d6*, allowed the observation of three aromatic spin systems containing $CH\alpha$ and $CH_2\beta$ consistent with two tyrosine and one histidine residues. The correlations observed in a NOESY experiment suggested the sequence N-acetyl-Tyr-Tyr-His, but with an additional unsaturation, which could not be assigned due to overlapping aromatic signals. A second set of analyses in CD_3OD allowed the complete assignment of the non-exchangeable protons of **1**. TOCSY and HSQC experiments showed that the histidine residue lacked its characteristic aromatic proton at position 5″ but had a quaternary carbon at 142 ppm. In an HMBC experiment, a diagnostic cross-correlation was observed between both CH-5 and CH-8 and C5″, consistent with a carbon-carbon bond between C6 and C5″ (Figure 1A). The proton and carbon assignments and the 1D and 2D spectra are reported in Figures S2 and S3 and Table S1.

The existence of the crosslink was confirmed after hydrolysis of **1**, which led to the release of a compound with *m/z* 335 $[M + H]^+$, corresponding to the expected carbon-carbon-linked tyrosine and histidine, and a compound with *m/z* 480 $[M + H]^+$, consistent with the loss of the acetyl group from the intact tripeptide. The *m/z* 335 $[M + H]^+$ compound retained the UV maximum at 315 nm, as expected for the biaryl chromophore.

Compound **2**, produced by the other three *Planomonospora* isolates, was deduced to be N-acetyl-Tyr-Phe-His, based on the similarity of the tandem mass fragmentation of **1** and **2**, with a neutral loss corresponding to phenylalanine (−147.07 Da) being the only difference between the fragmentation spectra (see Table S2).

We could only identify one literature precedent for natural products with a tyrosine-histidine biaryl bridge: aciculitins, cyclic nonapeptides decorated with glycolipids from the marine sponge *Aciculites orientalis* (Bewley et al., 1996). They contain a carbon-carbon bond between tyrosine and histidine but with a connectivity different than in **1**. Interestingly, aciculitins, which lack additional aromatic residues, show UV absorption maxima at 270 and 310 nm, matching those of **1** and **2** (Bewley et al., 1996).

A few metabolites have been described that consist of a cross-linked (N-acetylated) tripeptide: K-13 (**3**) from *Micromonospora halophytica* ssp. *exilisia* with a Tyr-Tyr-Tyr sequence and an ether

Please cite this article in press as: Zdouc et al., A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters, Cell Chemical Biology (2020), https://doi.org/10.1016/j.chembiol.2020.11.009

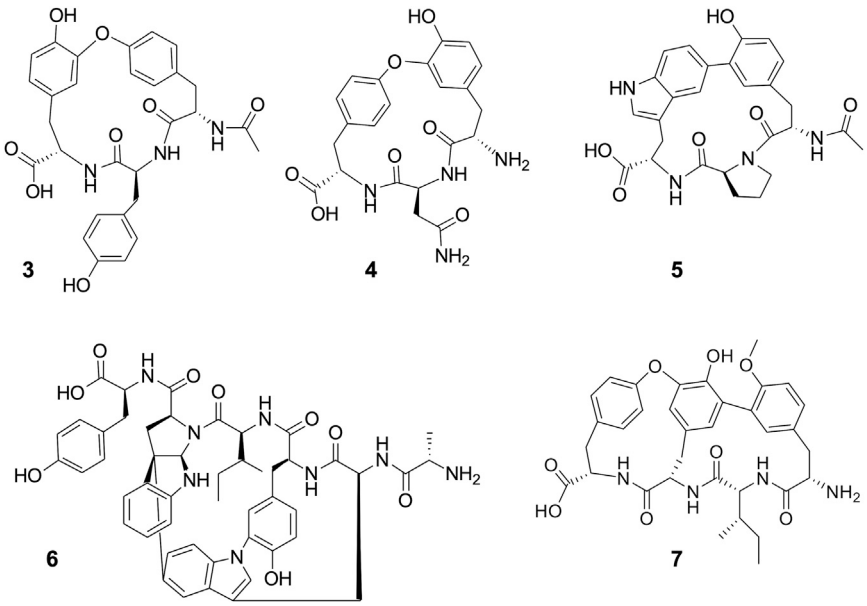


Figure 2. Structures of relevant metabolites mentioned in the text

K-13 (3), OF-4949 (4), pseudosporamide (5), tryptorubin (6), and cittilin A (7). Note that four variants of OF-4949 have been reported, with either H or Me on Tyr1-OH and H or OH on the Asn β carbon, but only one is shown for simplicity.

bond between the oxygen of Tyr-3 and C6 of Tyr-1 (Kase et al., 1987; Yasuzawa et al., 1987); OF-4949 (**4**) from *Penicillinum rugulosum* OF4949, which is identical to **3** except for asparagine in place of the central tyrosine and lack of the N-terminal acetyl group (Sano et al., 1986); and pseudosporamide (**5**) from *Pseudosporangium* sp. RD062863, reported while writing this article, with a Tyr-Pro-Trp sequence and a carbon-carbon bond between C6 of tyrosine and C-9 of tryptophan (Saito et al., 2020). Of note, compounds **3**, **4**, and **5** lack an absorption maximum at 315 nm.

It should be noted that compounds **3** to **5** have been shown to consist of L-amino acids only, consistent with a ribosomal synthesis. Among the **1**-related metabolites, **3** and **4** have been reported to inhibit the proteases angiotensin-converting enzyme (ACE) and aminopeptidase B, respectively, with apparent strict specificity, since **4** is not an ACE inhibitor. We were unable to observe inhibition of different proteases by **1** (data not shown), suggesting that the amino acid sequence and/or ring size can strongly influence activity. **5** was reported to have weak cytotoxic activity (Saito et al., 2020).

When the strain was grown in D_2O-supplemented medium, **1** was found to be extensively labeled with deuterium (Figure S4), indicating that its formation requires *de novo* amino acid synthesis.

We are unaware of any biosynthetic studies on **3**, **4**, or **5**. To get insights into **1** biosynthesis, we analyzed the 7.58-Mbp genome of *Planomonospora* sp. ID82291 (Zdouc et al., 2020) for the presence of antiSMASH-predicted NRPS or RiPP BGCs (Blin, et al., 2019), likely to specify a Tyr-Tyr-His peptide. However, this search proved unsuccessful, suggesting that **1** was formed by a BGC not recognized by the antiSMASH search tools. Searching the six-frame translation of the ID82291 genome led to a single plausible candidate: a short open reading frame encoding the pentapeptide MRYYH, preceded by a ribosomal binding site and followed by PLM4_2056, encoding a cytochrome P450 monooxygenase (see Figure 1B). The proposed BGC for **1**, encoding the two genes designated *bytA* and *bytO* (see below), is similar to the recently reported BGC for tryptorubin (**6**), a hexapeptide containing carbon-carbon and carbon-nitrogen bonds between aromatic amino acid residues (Wyche et al., 2017; Reisberg et al., 2020). Also for **6**, the BGC encodes just one cytochrome P450 enzyme and a precursor peptide, the latter, however, containing a 22-amino acid leader peptide. Similarly, the recently reported BGC for cittilin A (**7**), a crosslinked YIYY tetrapeptide from *Myxococcus xanthus*, encodes a 27-amino acid precursor peptide, an adjacent cytochrome P450 enzyme, a methyltransferase and a distant endopeptidase for cleavage (Hug et al., 2020). Cytochrome P450 monooxygenases are known to install crosslinks between aromatic residues of peptides, such as, e.g., in the NRPS-generated glycopeptides (Greule et al., 2018). These precedents made the BGC depicted in Figure 1B as the likely candidate for **1** biosynthesis. Apart from a possible regulator, the five genes upstream and downstream *bytAO* do not partake in **1** biosynthesis.

A similar two-gene organization was found in the genome of *Planomonospora* sp. ID107089, one of the producers of **2** (Figures 1B and S1). Of note, the two *bytAO* regions are almost identical: there is only one nucleotide (A-T) difference between the *bytA* genes, responsible for the Y-F difference in the peptides; both genes are preceded by equally spaced, identical ribosomal binding sites; the *bytO* sequences share 97% identity, while the *bytAO* intergenic regions (120 bp) are 95% identical.

Heterologous expression of *bytAO*

To prove that production of **1** is indeed governed by *bytAO*, we cloned a 1,303-bp sequence containing *bytA* and *bytO* into pSET152 and conjugated the plasmid into the expression host *Streptomyces coelicolor* M1152. Cultivation of *S. coelicolor* pSET *bytAO*, followed by HP20 extraction and high-performance liquid chromatography-tandem mass spectrometry analysis, resulted in the detection of a peak with an *m/z* 522.199, having the same UV spectrum, retention time, and tandem mass spectrometry spectrum as **1** (Figures 1C and S4). This unambiguously confirms the heterologous production of mature **1** by the minimal gene cluster *bytAO*, making it the shortest RiPP BGC so far reported. To our knowledge, the shortest reported peptide-encoding gene is the 24-bp *mccA*, which encodes the precursor peptide of microcin C7 in *Escherichia coli*, a RiPP also devoid of leader peptide (González-Pastor et al., 1994; Guijarro et al., 1995). With just 18 bp (including the stop codon), *bytA* is two codons smaller than *mccA* and thus the smallest peptide-coding gene ever reported across the Tree of Life. Since heterologous expression of *bytAO*

Please cite this article in press as: Zdouc et al., A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters, Cell Chemical Biology (2020), https://doi.org/10.1016/j.chembiol.2020.11.009

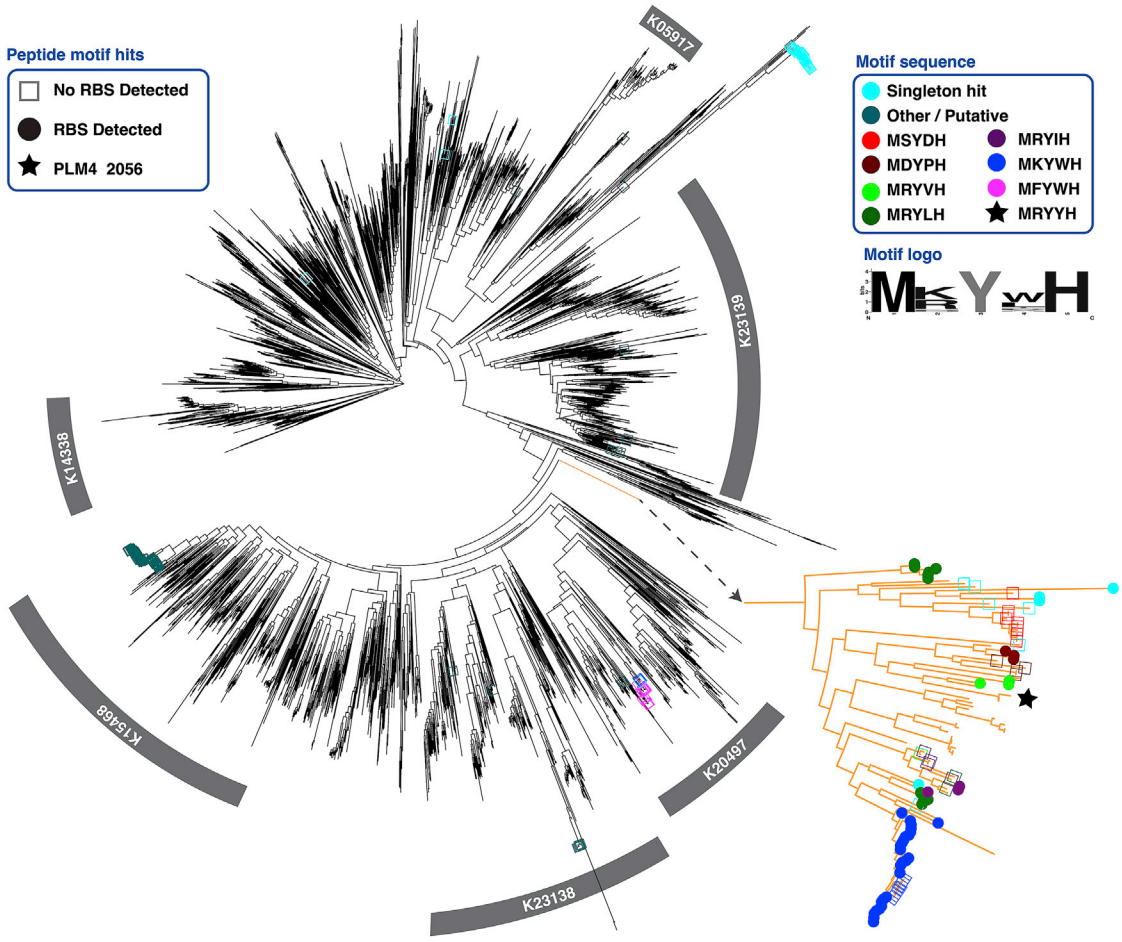


Figure 3. Overview phylogeny of cytochrome P450s from a variety of bacterial genera used in this study

KEGG orthology IDs in gray boxes highlight selected functional groupings (K20497: methyl-branched lipid omega-hydroxylase; K23138: family CYP109; K23139: family CYP110; K05917: steroid biosynthesis; K15468: polyketide biosynthesis; K14338: fatty acid degradation). Positive matches for pentapeptide motif hits are indicated as circles, where filled ones visualize a fully validated Shine-Dalgarno (SD) ribosomal binding site (RBS), while for empty ones an SD sequence was not detected. Putative singleton hits are marked in light or dark cyan if the same motif is not seen in other P450s, with <95% amino acid identity, or if the hit overlaps with the P450 coding region, respectively. Identical motifs correspond to the same color shown in the legend. P450s likely associated with biarylitides are highlighted in orange. This clade also contains the *bytAO* sequence marked with a filled star. The related *bytAO* sequence from strain ID107089 has not been included for simplicity. Consensus sequences of all motifs with validated RBS sites are illustrated with the sequence logo at the bottom right. More annotations and interactive tree can be accessed at: https://itol.embl.de/tree/21312710611256691590408100.
See also Figure S4.

in *S. coelicolor* resulted in the mature product, we infer that **1** biosynthesis recruits one or more generic enzymes, involved in trimming the additional amino acid(s) at the N terminus and in N-acetylation. Of note, the existence of cluster-encoded proteases in RiPP BGCs from *Actinobacteria* seems to be the exception (Chen et al., 2019). As an additional observation, the N-acetylated compounds depicted in Figure 2 are all produced by *Actinobacteria*, while those with a free N terminus were observed in fungi or myxobacteria.

Biarylitides: a widespread family of RiPPs

We reasoned that the cytochrome P450 monooxygenase PLM4_2056 would be selective in installing a biaryl crosslink on short peptides, and that it could belong to a distinct phylogenetic branch of this enzyme family. If the cognate pentapeptide-encoding gene was closely linked to the cytochrome monooxygenase gene, then one could easily identify additional BGCs related to those depicted in Figure 1B. We thus conducted a search for analogous small CDSs (coding DNA sequences) in the flanking regions of genes related to PLM4_2056. Assuming structural conservation of the carbon-carbon bond, we restricted our search to pentapeptides with tyrosine at position 3 and tyrosine/histidine at position 5. From approximately 3,300 PLM_2056-related sequences, CDSs encoding the specified peptides were identified in approximately 200 genomes, the majority from *Actinobacteria*. To increase our confidence in the hits, we filtered the detected CDS for a Shine-Dalgarno motif 6–8 bp upstream of the peptide sequence, as indicated in the phylogenetic trees shown in Figures 3 and 4. Interestingly, such a ribosomal binding site was only detected in sequences from *Actinobacteria*, where it occurred in the vicinity of approximately half of the detected peptides (indicated by filled circles

Please cite this article in press as: Zdouc et al., A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters, Cell Chemical Biology (2020), https://doi.org/10.1016/j.chembiol.2020.11.009

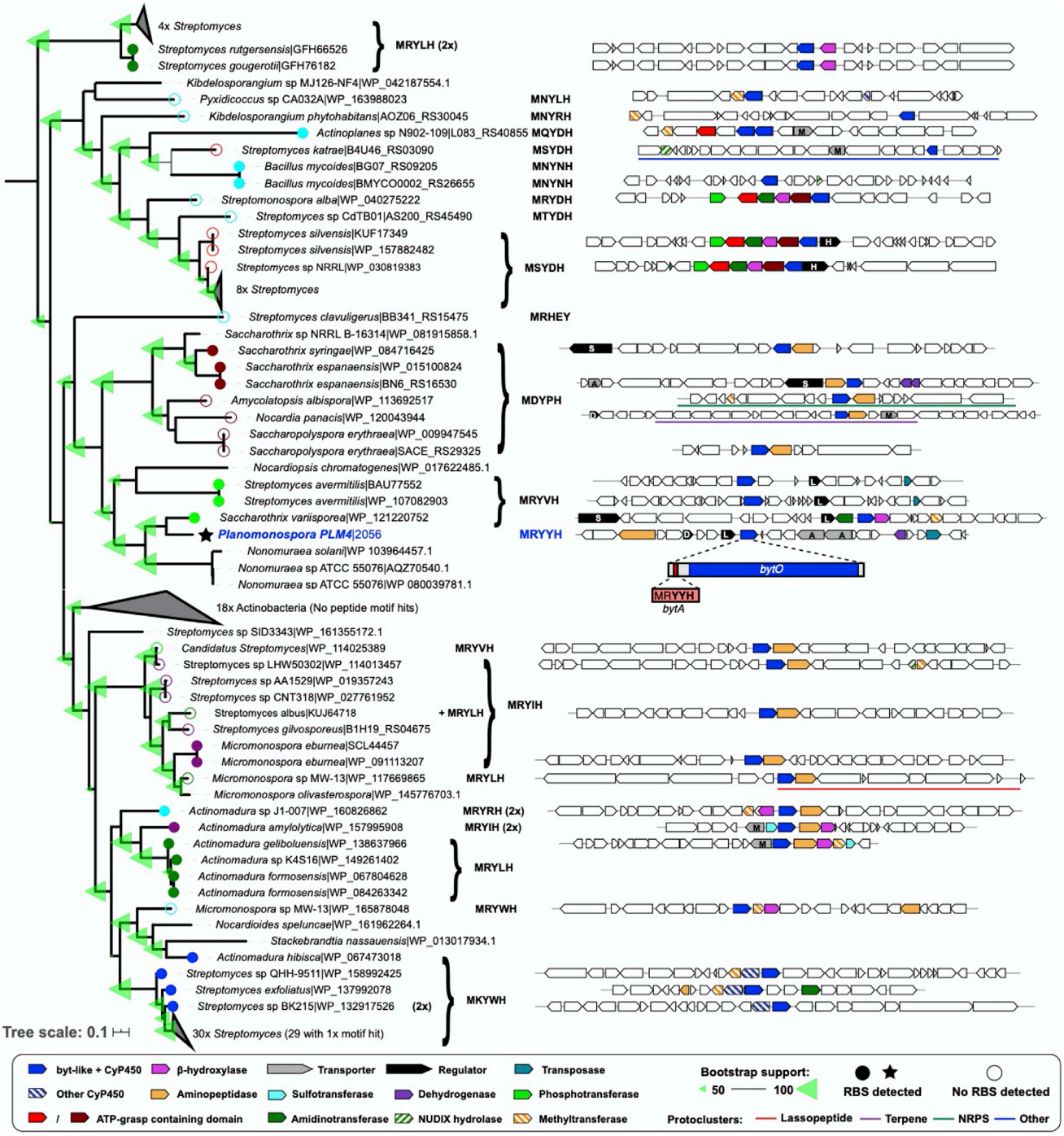


Figure 4. Summary phylogeny of cytochrome P450s, likely associated with biarylitide biosynthesis

Identical pentapeptide motif hits are shown as colored circles corresponding to the motif printed to the right of the tree. Filled and unfilled circles indicate those with and without a detected Shine-Dalgarno RBS, respectively. Selected regions of *byt*-like clusters are shown to the right of motifs, illustrating a variety of contexts and shared coding sequences potentially linked to the minimal P450 cluster (highlighted in blue). Expanded clades and detailed annotations are accessible at: https://itol.embl.de/tree/2131271061176991591098016 (Letunic and Bork, 2019).

in Figures 3 and 4). Furthermore, this clade was seen to have the only example of replicate motifs (denoted "2x" in Figure 4). To get insight into the amino acid distribution, detected peptides were aligned using the program MUSCLE. Peptides with a preceding ribosomal binding site showed less variation, with a strong preference for a basic amino acid at position 2, in comparison with the global alignment (see Figures 3 and S4). Intriguingly, the diversity of genomic loci in which PLM_2056 homologs

were found together with *bytA*-like motifs, suggests that these BGCs encode the production of a much wider diversity of molecules with distinct chemical modifications, as suggested by the co-localization of methyltransferases, sulfotransferases, ATP-grasp enzymes, and other tailoring enzymes in these loci (Figure 4). Together, this represents a treasure trove for further discovery efforts in this unexplored family.

As a side observation, sequence identity between BytO and actinobacterial P450 monooxygenases from the biarylitides clade ranges between 40% and 50%. In contrast, BytO shows 20%–25% identity with P450 monooxygenases involved in glycopeptide crosslinks.

We propose the name biarylitides for this family of N-acetylated tripeptides, carrying a crosslink between the aromatic side chains of amino acids 1 and 3. Accordingly, compounds **1**, **2**, K-13, and pseudosporamide would become biarylitide YYH, YFH, YYY, and YPW, respectively. Since the first comprehensive literature overview of just a few years ago (Arnison et al., 2013), several new RiPP families have been uncovered, mostly through searching precursor peptide-related features in genome sequences (Kloosterman et al., 2020; Montalbán-López et al., 2020). RiPPs such as the biarylitides can, however, not be detected by current bioinformatic search strategies because of the extremely small size of the *bytA* gene. This highlights the importance of metabolomics approaches to unveil novel chemistry and unexpected BGCs. As shown here and recently described in Russell and Truman (2020), with hindsight from genomic data, these minimal RiPPs can be identified via cytochrome P450 homology and screening for the short peptide sequences in many bacterial genomes. While occurring mostly in *Actinobacteria*, one such BGC was also detected in a *Pyxidicoccus* (myxobacteria) strain. While we restricted our searches to a pentapeptide MxYxY/H-encoding gene within ±500 bp from a cytochrome P450 monooxygenase gene, the similar architecture of BGCs for tryptorubin and cittilin A suggests that many further variations in short, crosslinked peptides exists in microbial genomes. Additional search strategies are thus likely to further expand these RiPP families.

While the *byt* BGC is extremely small, it poses nonetheless some biochemical challenges for the specificity of the biaryl crosslink. The basic amino acid, often found at position −1, might act as a signal for processing. Further studies will be needed to understand the timing and sequence requirements for installing a crosslink and trimming the additional amino acid(s) at the N terminus. Since the biaryl crosslink imparts proteolytic stability, understanding its specificity might have important applications in stabilizing peptides for different applications. Our understanding of the biological activities of biarylitides is extremely limited. The different protease inhibitory activities seen with **1**, **3**, and **4** suggest that the amino acid sequence and/or ring size can strongly influence bioactivity. Thus, their biological properties and their role in the producing organisms await discovery.

SIGNIFICANCE

We have identified crosslinked, cyclic tripeptides produced by different strains of *Planomonospora* and established that its formation in a heterologous host requires a minimal biosynthetic gene cluster consisting of just two genes: one encoding a pentapeptide, representing the smallest peptide-coding gene ever reported, and the other encoding a cytochrome P450 monooxygenase. Phylogenomic analysis of related cytochrome P450 sequences revealed that this class of minimal RiPP gene clusters is widespread in actinobacterial genomes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bacterial strains and growth conditions
  - Cultivation in D_2O
- METHOD DETAILS
  - Isolation and purification of **1**
  - Analytical procedures
  - Hydrolysis of **1**
  - Biarylitide YYH (**1**)
  - Cloning and heterologous Expression of *bytAO*
  - Protease activity assay
  - Phylogenetic analyses
  - Peptide motif search
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.chembiol.2020.11.009.

ACKNOWLEDGMENT

The authors thank Stefan Kehraus for measuring optical rotation of **1**. This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 721484 (Train2Target). G.S.Z. acknowledges the DAAD for a PhD scholarship. M.C. received funding from the German Research Foundation (DFG) grant no. FOR2372. M.M.A. and M.H.M. were supported by an ERA NET CoBiotech (Bestbiosurf) grant through the Netherlands Organization for Scientific Research (NWO) (053.80.739).

AUTHOR CONTRIBUTIONS

M.M.Z., S.D., and M.S. designed the work, identified metabolites, analyzed the data, and wrote the paper. M.M.A. and M.H.M. performed bioinformatics analyses. G.S.Z. and M.C. performed heterologous expression and analyzed the data. S.I.M. identified metabolites and analyzed the data. All authors contributed to the critical discussion of the manuscript, and read and approved the final manuscript.

DECLARATION OF INTERESTS

M.M.Z., M.S., S.I.M., and S.D. have been and/or are employees and/or shareholders of Naicons Srl. M.H.M. is a co-founder of Design Pharmaceuticals and a member of the scientific advisory board of Hexagon Bio. The other authors declare that no competing interests exist.

Received: September 11, 2020
Revised: October 23, 2020
Accepted: November 20, 2020
Published: December 14, 2020

REFERENCES

Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G., Camarero, J.A., Campopiano, D.J., Challis, G.L., Clardy, J., et al. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat. Prod. Rep. *30*, 108–160.

Bewley, C.A., He, H., Williams, D.H., and Faulkner, D.J. (1996). Aciculitins A–C: cytotoxic and antifungal cyclic peptides from the lithistid sponge *Aciculites orientalis*. J. Am. Chem. Soc. *118*, 4314–4321.

Bierman, M.I., Stein, K.L., and Snyder, J.V. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. Gene *116*, 43–49.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., and Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. *2*, W81–W87, https://doi.org/10.1093/nar/gkz310.

Chen, Y., Mathai, N., and Kirchmair, J. (2020). Scope of 3D shape-based approaches in predicting the macromolecular targets of structurally complex small molecules including natural products and macrocyclic ligands. J. Chem. Inf. Model. *60*, 2858–2875, https://doi.org/10.1021/acs.jcim.0c00161.

Chen, S., Xu, B., Chen, E., Wang, J., Lu, J., Donadio, S., Ge, H., and Wang, H. (2019). Zn-dependent bifunctional proteases are responsible for leader peptide processing of class III lanthipeptides. Proc. Natl. Acad. Sci. U S A *116*, 2533–2538, https://doi.org/10.1073/pnas.1815594116.

Donadio, S., Monciardini, P., and Sosio, M. (2009). Chapter 1. Approaches to discovering novel antibacterial and antifungal agents. Methods Enzymol. *458*, 3–28.

Du, L., Liu, R.H., Ying, L., and Zhao, G.R. (2012). An efficient intergeneric conjugation of DNA from *Escherichia coli* to mycelia of the lincomycin-producer *Streptomyces lincolnensis*. Int. J. Mol. Sci. *13*, 4797–4806, https://doi.org/10.3390/ijms13044797.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29–W37.

Funk, M.A., and van der Donk, W.A. (2017). Ribosomal natural products, tailored to fit. Acc. Chem. Res. *50*, 1577–1586, https://doi.org/10.1021/acs.accounts.7b00175.

Gfeller, D., Michielin, O., and Zoete, V. (2013). Shaping the interaction landscape of bioactive molecules. Bioinformatics *29*, 3073–3079, https://doi.org/10.1093/bioinformatics/btt540.

González-Pastor, J.E., San Millán, J.L., and Moreno, F. (1994). The smallest known gene. Nature *369*, 281, https://doi.org/10.1038/369281a0.

Greule, A., Stok, J.E., De Voss, J.J., and Cryle, M.J. (2018). Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism. Nat. Prod. Rep. *35*, 757–791.

Guijarro, J.I., González-Pastor, J.E., Baleux, F., San Millán, J.L., Castilla, M.A., Rico, M., Moreno, F., and Delepierre, M. (1995). Chemical structure and translation inhibition studies of the antibiotic microcin C7. J. Biol. Chem. *270*, 23520–23532.

Hug, J.J., Dastbaz, J., Adam, S., Revermann, O., Koehnke, J., Krug, D., and Mueller, R. (2020). Biosynthesis of cittilins, unusual ribosomally synthesized and post-translationally modified peptides from *Myxococcus xanthus*. ACS Chem. Biol. *15*, 2221–2231, https://doi.org/10.1021/acschembio.0c00430.

Iorio, M., Tocchetti, A., Santos Cruz, J.C., Del Gatto, G., Brunati, C., Maffioli, S.I., Sosio, M., and Donadio, S. (2018). Novel polyethers from screening *Actinoallomurus* spp. Antibiotics (Basel) *7*, 47, https://doi.org/10.3390/antibiotics7020047.

Kase, H., Kaneko, M., and Yamada, K. (1987). K-13, a novel inhibitor of angiotensin I converting enzyme produced by *Micromonospora halophytica* subsp. *exilisia*. I. J. Antibiot. *40*, 450–454.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kloosterman, A.M., Shelton, K.E., van Wezel, G.P., Medema, M.H., and Mitchell, D.A. (2020). RRE-finder: a genome-mining tool for class-independent RiPP discovery. bioRxiv *14*, 992123, https://doi.org/10.1101/2020.03.14.992123.

Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. *2*, W256–W259, https://doi.org/10.1093/nar/gkz239.

Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I., Chooi, Y.H., Claesen, J., Coates, R.C., et al. (2015). Minimum information about a biosynthetic gene cluster. Nat. Chem. Biol. *11*, 625–631.

Monciardini, P., Iorio, M., Maffioli, S.I., Sosio, M., and Donadio, S. (2014). Discovering new bioactive molecules from microbial sources. Microb. Biotechnol. *7*, 209–220.

Montalbán-López, M., Scott, T.A., Ramesh, S., Rahman, I.R., van Heel, A.J., Viel, J.H., Bandarian, V., Dittmann, E., Genilloud, O., Goto, Y., et al. (2020). New developments in RiPP discovery, enzymology and engineering. Nat. Prod. Rep.. https://doi.org/10.1039/d0np00027b.

Pastor, I., Vilaseca, E., Madurga, S., Gargès, J.L., Cascante, M., and Mas, F. (2011). Effect of crowding by dextrans on the hydrolysis of N-succinyl-L-phenyl-Ala-p-nitroanilide catalyzed byr-chymotrypsin. J. Phys. Chem. B *115* (5), 1115–1121.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490, https://doi.org/10.1371/journal.pone.0009490.

Reisberg, S.H., Gao, Y., Walker, A.S., Helfrich, E.J., Clardy, J., and Baran, P.S. (2020). Total synthesis reveals atypical atropisomerism in a small-molecule natural product, tryptorubin A. Science *367*, 458–463.

Russell, A.H., and Truman, A.W. (2020). Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. Computat. Struct. Biotechnol. J. *18*, 1838–1851.

Saito, S., Atsumi, K., Zhou, T., Fukaya, K., Urabe, D., Oku, N., Karim, M.R.U., Komaki, H., and Igarashi, Y. (2020). A cyclopeptide and three oligomycin-class polyketides produced by an underexplored actinomycete of the genus *Pseudosporangium*. Beilstein J. Org. Chem. *16*, 1100–1110.

Sano, S., Ikai, K., Katayama, K., Takesako, K., Nakamura, T., Obayashi, A., Ezure, Y., and Enomoto, H. (1986). OF4949, new inhibitors of aminopeptidase B II. Elucidation of structure. J. Antibiot. *39*, 1685–1696.

Wyche, T.P., Ruzzini, A.C., Schwab, L., Currie, C.R., and Clardy, J. (2017). Tryptorubin A: a polycyclic peptide from a fungus-derived *Streptomycete*. J. Am. Chem. Soc. *139*, 12899–12902.

Yasuzawa, T., Shirahata, K., and Sano, H. (1987). K-13, a novel inhibitor of angiotensin I converting enzyme produced by *Micromonospora halophytica* subsp. *exilisia*. II. J. Antibiot. *40*, 455–458.

Yera, E.R., Cleves, A.E., and Jain, A.N. (2011). Chemical structural novelty: on-targets and off-targets. J. Med. Chem. *54*, 6771–6785.

Zdouc, M.M., Iorio, M., Maffioli, S.I., Crüsemann, M., Donadio, S., and Sosio, M. (2020). Preprint *bioRxiv*, p. 210815, https://doi.org/10.1101/2020.07.19.210815.

Please cite this article in press as: Zdouc et al., A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters, Cell Chemical Biology (2020), https://doi.org/10.1016/j.chembiol.2020.11.009

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Escherichia coli DH5α cells	Bioline	Bio-85026
Escherichia coli ET12567/pUZ8002	Macneil et al., 1992	NA
Streptomyces coelicolor strain M1152	Flinspach et al., 2010	NA
Planomonospora sp. ID82291	Naicons collection	NA
Planomonospora sp. ID107089	Naicons collection	NA
Planomonospora sp. ID114239	Naicons collection	NA
Planomonospora sp. ID46116	Naicons collection	NA
Chemicals, peptides, and recombinant proteins		
Apramycin sulfate	Sigma-Aldrich	Cat# A2024-5G
Nalidixic acid	Sigma-Aldrich	Cat# 1451000
Diaion HP-20	Sigma-Aldrich	Cat# 13,605
α-chymotrypsin from bovine pancreas	Sigma-Adrich	Cat# C4129
Chymostatin, microbial	Sigma-Aldrich	Cat# C7268
Deuterium oxide	Sigma-Aldrich	Cat# 151882
Critical commercial assays		
Phusion DNA polymerase	NEB	Cat# M0530S
Taq DNA ligase	NEB	Cat# M0208S
T5 exonuclease	NEB	Cat# M0663S
Angiotensin I Converting Enzyme (ACE) Activity Assay Kit	Sigma-Aldrich	Cat# CS0002
Deposited data		
Planomonospora sp. ID82291 genome	Zdouc et al. (2020)	GenBank: JABTEX000000000
biarylitide-YYH gene cluster	this paper	GenBank: MW201788
biarylitide-YFH gene cluster	this paper	GenBank: MW201789
iTOL-webserver	this paper	https://itol.embl.de/tree/213127106112566915904081 00
iTOL- Expanded clades	this paper	https://itol.embl.de/tree/213127106117699159109801 6
Oligonucleotides		
pSET_fwd: CCTCTCTAGAGTCGACCTGCAGC	this paper	NA
pSET_rev: CCTTCCGTACCTCCGTTGCT	this paper	NA
bytAO_fwd: AGCAACGGAGGTACGGAAGGAGGAGG TGTGCGATGCGC	this paper	NA
bytAO_rev: GCAGGTCGACTCTAGAGAGGCTAGCGG GGGAGAAGGAC	this paper	NA
pSET-bytAO_fwd: GCGTCGATTTTTGTGATGCTCG	this paper	NA
pSET-bytAO_rev: CAGCGAATTCGGAAAACGGC	this paper	NA
Recombinant DNA		
pSET152_*ermE**	Bierman et al. (1992)	NA
pSET-bytAO	this paper	NA

(*Continued on next page*)

Please cite this article in press as: Zdouc et al., A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters, Cell Chemical Biology (2020), https://doi.org/10.1016/j.chembiol.2020.11.009

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
antiSMASH 5.0	Blin et al. (2019)	https://antismash.secondarymetabolites.org/
MIBiG database	Medema et al. (2015)	https://mibig.secondarymetabolites.org/
MAFFT	Katoh, et al., 2013	https://mafft.cbrc.jp/alignment/software/
HMMER	Finn, et al., 2011	http://hmmer.org/
FastTree2	Price, et al., 2010	http://microbesonline.org/fasttree/
iTOL server	Letunic I., Bork P., 2019	https://itol.embl.de/
Other		
Mass Spectrometer	Thermo Scientific	LCQ Fleet
Dionex Ultimate 3000 with diode array detector	Thermo Scientific	NA
Column Atlantis T3 C18 5 μm × 4.6 mm x 50 mm	Waters	Cat# 186003726
HR-LCMS	Bruker	MicroTOF-QIII
Enspire Multimode Plate Reader	PerkinElmer	NA
CombiFlash Rf flash liquid chromatography system	Teledyne Isco	Cat# 68-5230-015
SNAP Ultra HP-Sphere C18 25 μm 12g	Biotage	NA
LC-2010 AHT	Shimadzu	NA
Column Nucleodur C18 Pyramid 5 μm × 250 mm x 10 mm	Macherey-Nagel	Cat# 762272.100
300 MHz Avance NMR Spectrometer	Bruker	NA

RESOURCE AVAILABILITY

Lead contact

Further information and requests for reagents should be directed to and will be fulfilled by the Lead Contact, Margherita Sosio (msosio@naicons.com)

Materials availability

Materials generated in this study can be obtained with a Materials Transfer Agreement.

Data and code availability

The software used in this study are listed in the Key Resourses Table. The *Planomonospora* sp. ID82291 genome sequence was deposited into the GenBank with the accession number JABTEX000000000. The sequences of the biarylitide-YYH and biarylitide-YFH gene clusters have been deposited in the NCBI under the accession number GenBank MW201788 and MW201789, respectively. Expanded clades and detailed annotations are accessible at: https://itol.embl.de/tree/2131271061176991591098016. Further data are available from the corresponding author upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial strains and growth conditions

The bacterial strains used in this study are listed in the Key Resources Table. Planomonospora strains from frozen stocks (−80°C) were propagated on S1 plates at 28°C for two to three weeks. The grown mycelium was then homogenized with a sterile pestle and used to inoculate 15 mL AF medium (20 g/L dextrose, 8 g/L soybean meal, 2 g/L yeast extract, 4 g/L $CaCO_3$, 1 g/L NaCl, pH 7.2) in a 50 mL baffled flask. After cultivation on a rotary shaker (200 rpm) at 30°C for 72 hr, 1.5 mL of the exponentially growing culture was used to inoculate each 15 mL of AF medium, as described (Donadio et al., 2009; Zdouc et al., 2020). After 7 days of cultivation as before, cultures were harvested and extracted by centrifugation and the supernatant, after adsorption on 200 μL of HP20 resin (Diaion) for 1 hr, was eluted with 200 μL of methanol (Sigma). Samples were analyzed by LC-MS as described below. For isolation and purification of **1**, 1.5 mL of a frozen stock of strain ID82291 was used to inoculate 15 mL AF medium in a 50 mL baffled flask (Donadio et al., 2009). After 72 hr, 10 mL were used to inoculate 100 mL fresh AF medium in a 500 mL baffled flask. After further 72 hr, 100 mL was used to inoculate 900 mL AF medium in a 3L flask, containing 50 g 5-mm glass beads.

Please cite this article in press as: Zdouc et al., A biaryl-linked tripeptide from *Planomonospora* reveals a widespread class of minimal RiPP gene clusters, Cell Chemical Biology (2020), https://doi.org/10.1016/j.chembiol.2020.11.009

Cultivation in D_2O

1.5 mL of a frozen stock culture of strain ID82291 was used to inoculate 15 mL AF medium in a 50 mL baffled flask. After 72 hr, 1.5 mL were used to inoculate each 15 mL AF medium prepared in 30% D_2O (v/v) in 50 mL baffled flasks and further cultivated under identical conditions. Production of **1** was monitored by sampling 1 mL of broth every 24 hr. Samples were centrifuged and the supernatant, after adsorption on 200 μL of HP20 resin (Diaion) for 1 hr, was eluted with 200 μL of methanol (Sigma). Samples were analyzed by LC-MS as described below.

METHOD DETAILS

Isolation and purification of 1

After 7 days, the culture (as described above) was first centrifuged and then filtered, and the resulting cleared broth was adsorbed on 200 mL HP20 resin (Diaion). The loaded resin was washed with demineralized water and eluted with aqueous 50% MeOH (V/V). The eluate was brought to dryness. The dried extract was taken up in 10% acetonitrile/90% water and pre-fractionated on a Biotage SNAP Ultra HP-Sphere C18 25 μm 12g with a CombiFlash system (Teledyne ISCO, Nebraska, USA). All chromatographic steps were performed using purified water (MilliQ, Merck). Phases A and B were water and acetonitrile (Sigma-Aldrich), respectively. A gradient with steps of 5, 5, 30, 100 and 100% phase B at 0, 5, 20, 25 and 30 min, respectively, was applied. After screening fractions by LC-MS, **1**-containing fractions were dried, taken up in 0.05% trifluoroacetic acid (TFA) in 10% MeCN/90% water and processed further by semi-preparative HPLC on a Shimadzu LC-2010 CHT instrument, equipped with a 250/10 Nucleodur C18 Pyramid 5M column (Macherey-Nagel) and a C18 pre-column. An isocratic method with 10% phase B (A: 0.05% TFA in water, pH 2.5; B: MeCN) at 3 mL/min, column temperature 40°C was used for purification. Due to co-eluting contaminants, the semi-purified substance was purified for a second time under the same conditions, except for the change of phase A to water. Drying of the eluent yielded 5 mg of a white solid.

Analytical procedures

High-resolution MS spectra acquisition was performed on a micrOTOF-Q III (Bruker) instrument equipped with an electrospray interface (ESI) coupled with a Dionex UltiMate 3000 (Thermo Scientific), using an Atlantis T3 C18 5 μm × 4.6 mm x 50 mm column (Zdouc et al., 2020). LC-MS measurements for monitoring metabolite production were performed on a Dionex UltiMate 3000 (Thermo Scientific) coupled with an LCQ Fleet (Thermo Scientific) mass spectrometer equipped with an electrospray interface (ESI) on an Atlantis T3 C18 5 μm × 4.6 mm x 50 mm column, as reported elsewhere (Iorio et al., 2018). LC-MS measurements for analysis of **1** after acid hydrolysis were performed on the same instrument, applying a hydrophilic gradient with 0, 0, 25, 95, 95, 0 and 0% phase B (MeCN) at 0, 1, 7, 8, 12, 12.5 and 14 min, respectively, with the *m/z* range set to 110–1000 Da. Phase A was 0.05% TFA in water. UV-VIS signals (190–600 nm) were acquired using the built-in diode array detector of the Dionex UltiMate 3000 (Thermo Scientific). Mono- and bi-dimensional NMR spectra were measured in DMSO-*d6* or CD_3OD at 298K using a Bruker Avance 300 MHz spectrometer.

Hydrolysis of 1

A small amount of purified compound **1** (<1 mg) was dissolved in 250 μL 0.1 M sodium borate buffer (pH 9), followed by addition of 250 μL fresh 6N HCl (Sigma-Aldrich). The solution was heated to 120°C and sampled at 24, 48 and 96 hr. The samples were analyzed by LC-MS analysis, using the hydrophilic gradient as described above.

Biarylitide YYH (1)

white crystalline powder; $[\alpha]_D = -165$ (c 0.11, H_2O pH10, 25°C); ^1H and ^{13}C NMR data in Supporting Information (Figures S2 and S3); HR-ESI-MS *m/z* 522.198 $[M + H]^+$ (calcd for $C_{26}H_{27}N_5O_7$, 522.1988); ESI-MS/MS ($[M + H]^+$) *m/z* (%) 269.139 (100), 359.134 (84), 315.143 (69), 494.201 (66), 289.128 (61), 285.133 (56), 331.140 (54), 243.122 (51), 313.129 (37), 448.198 (27), 476.191 (17), 228.111 (10), 406.181 (9), 214.094 (9), 199.081 (7).

Cloning and heterologous Expression of *bytAO*

A DNA stretch containing *bytAO* (1303 bp) was cloned into pSET152_*ermE** via Gibson assembly. Linear DNA fragments were generated by PCR using primers designed to include 20 bp overlap sequences (pSET_fwd: CCTCTCTAGAGTCGACCTGCAGC, pSET_rev: CCTTCCGTACCTCCGTTGCT, *bytAO*_fwd: AGCAACGGAGGTACGGAAGGAGGAGGTGTGCGATGCGC and *bytAO*_rev: GCAGGTCGACTCTAGAGAGGCTAGCGGGGGAGAAGGAC). Purified linear vector and insert were mixed (ratio 1:3) with 15 μL of Gibson assembly master mixture (1X ISO buffer, 10 U/L T5 exonuclease, 2 U/L Phusion polymerase, and 40 U/μL Taq ligase) and incubated at 50°C for 1 hr. Then, 5 μL of reaction mix were used to transform chemically competent *E. coli* DH5α cells, which were plated onto LB agar supplemented with 50 μg/mL apramycin. Positive clones containing the plasmid pSET-bytAO were selected by colony PCR (primers: pSET-bytAO_fwd: GCGTCGATTTTTGTGATGCTCG and pSET-bytAO_rev: CAGCGAATTCGGAAAACGGC) and confirmed using restriction digest and sequencing. pSET-bytAO was conjugated into *S. coelicolor* M1152 using a standard intergeneric conjugation protocol (Du et al., 2012) with the methylation-deficient *E. coli* ET12567/pUZ8002/pSET-bytAO as the donor. Cells were spread onto MS agar (20 g/L soya flour, 20 g/L mannitol, 20 g/L agar, 50 mM $MgCl_2$) and incubated at 30°C. After 20 hr, 0.5 mg nalidixic acid and 1 mg apramycin were overlaid on the plates and incubation continued until ex-conjugants showed up. Single colonies were spread on fresh plates containing 25 mg/L nalidixic acid and 50 mg/L apramycin to confirm the ex-conjugants. Colonies containing *bytAO*

were validated by colony PCR and cultivated in TSB broth for 12 days at 30°C. Cultures were extracted as described above and analyzed by HR-LCMS, as described above.

Protease activity assay

Angiotensin I Converting Enzyme (ACE) inhibition was measured using the Sigma-Aldrich Angiotensin I Converting Enzyme (ACE) Activity Fluorescence Assay Kit (CS0002), following instruction. Briefly, a serial dilution of compound 1, starting at 100 μM, was tested for inhibitory activity against ACE. As a positive control, teptrotide (Sigma-Aldrich A0773) at 100 μM was used. Reactions were incubated at 37°C and measured every 5 min, using a PerkinElmer Enspire Multimode Plate Reader. All reactions were performed in duplicates and the mean calculated. Chymotrypsin inhibition was measured by a fluorescent-based assay (Pastor et al., 2011). Briefly, a serial dilution of compound 1, starting at 100 μM, was tested for inhibitory activity against α-chymotrypsin from bovine pancreas (Sigma-Adrich C4129). As a positive control, chymostatin from microbial origin (Sigma-Aldrich C7268) at 40 μg/mL was used. Reactions were incubated at 25°C and measured every 5 min, using a PerkinElmer Enspire Multimode Plate Reader. All reactions were performed in duplicates and the mean calculated.

Phylogenetic analyses

A set of complete genomes from the antiSMASH database were used along with top blast results to PLM4_2056 from *Planomonospora sp*. ID82291 (NCBI reference JABTEX000000000). Cytochrome P450 enzymes were identified via a hidden Markov model search (Finn et al., 2011) using the Pfam (PF00067) P450 model and trusted bit-score cutoffs. After ensuring model and query alignment coverage over 70%, similar sequences (>95% amino acid ID) were clustered via an all vs. all blast and one representative was used for the final dataset, which yielded over 3,300 sequences. Any hit to the motif search analysis (see below) were also included in this dataset. All sequences were aligned via MAFFT (Katoh et al., 2013) using the local iterative (-linsi) option followed by alignment trimming using the trimal application with automated option to improve maximum-likelihood tree inference. Due to the large number of sequences FastTree (Price et al., 2010) was used to infer the final phylogenetic gene tree.

Peptide motif search

A custom python script (https://git.wur.nl/snippets/65) was used to examine 6-frame translations of corresponding genomic regions +/−500bp of identified cytochrome P450 enzymes. This reported any motif matching MxYx[Y/H]*, where "x" is any amino acid and * is a stop codon. Reported positions were then cross-referenced to highlight motifs occurring up/downstream of the P450 CDS. An additional screen of these results for the Shine-Dalgarno sequence "AGGAGG" was preformed upstream of all motif positions to annotate RBS sites. Identified peptides were mapped onto the generated, P450-based phylogenetic tree and visualized using the iTOL-webserver (Letunic and Bork, 2019; iTOL Webserver. https://itol.embl.de/, Accessed: 06/11/2019).

QUANTIFICATION AND STATISTICAL ANALYSIS

PerkinElmer Enspire Multimode Plate Reader was used for the Angiotensin I Converting Enzyme (ACE) Activity Fluorescence Assay and for the Chymotrypsin inhibition Assay, in METHOD DETAILS (Protease Activity Assay): experiments were conducted in duplicates and the mean calculated using Microsoft Excel.