Oxidative Peptide Backbone Cleavage by a HEXXH Enzyme During RiPP Biosynthesis

Yao Ouyang,[1] Yue Yu,[1] Lingyang Zhu,[2] and Wilfred A. van der Donk[1,3*]

[1] Department of Chemistry and Howard Hughes Medical Institute, 600 South Mathews Avenue, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA.

[2] School of Chemical Sciences NMR Laboratory, University of Illinois at Urbana-Champaign, Urbana, 61801, IL, USA.

[3] Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 West Gregory Drive, Urbana, Illinois, 61801, USA.

* E-mail: vddonk@illinois.edu

Abstract

Ribosomally synthesized and post-translationally modified peptides (RiPPs) rely on a diverse array of enzymes to tailor peptide backbones and side chains. In this study, we characterized enzymes from two different biosynthetic gene clusters (BGCs) from Pseudomonas strains (*pfl* and *pos*) that catalyze new transformations in RiPP biosynthesis. Two α-ketoglutarate-dependent HEXXH enzymes, PflC and PosC, perform hydroxylation of multiple consecutive glutamine residues and selectively recognize a C-terminal ARMD tetrapeptide to trigger oxidative backbone cleavage that generates an amide terminus. Mutational analysis pinpoints the first position of this motif as a critical determinant. Notably, PflC displays proteolytic activity in the absence of the leader peptide, indicating that leader peptide–enzyme interactions modulate the observed reaction selectivity. The biosynthetic gene clusters also encode a unique MNIO-nitroreductase fusion enzyme that installs a rare *Z*-dehydrophenylalanine and hydroxylates an Asp residue. Collectively, this work expands both the catalytic repertoire and structural diversity accessible through bacterial RiPP biosynthesis.

Introduction

Ribosomally synthesized and post-translationally modified peptides (RiPPs) represent one of the most structurally and functionally diverse families of natural products.[1] Their biosynthesis relies on genetically encoded precursor peptides that undergo extensive enzymatic tailoring, which introduces a broad spectrum of chemical modifications to generate highly diversified scaffolds that display a range of bioactivities.[2-4] Recent years have witnessed the discovery of numerous emerging RiPP-associated enzymes that mediate unprecedented

chemical transformations.[5,6] For example, multinuclear non-heme iron oxidative enzymes (MNIOs) catalyze a wide variety of different transformations.[7-9] These findings not only highlight the remarkable diversity of chemical logic underlying RiPP maturation, but also expand the known enzymatic repertoire by revealing new strategies for scaffold modification. Therefore, continuous mining and characterization of novel RiPP enzymes is valuable, both for understanding the biosynthetic pathways and for uncovering new paradigms in enzymatic chemistry with potential applications in biocatalysis and synthetic biology.[6]

Proteins harboring the conserved HEXXH motif constitute a large class of metalloproteases traditionally associated with peptide bond hydrolysis.[10,11] Recent studies have revealed that not all members conform to this canonical role. Morinaka and co-workers characterized a HEXXH-motif protein fused to a radical SAM domain, which contrary to its prior annotation as a metallopeptidase, functions as an α-ketoglutarate (αKG)-dependent mononuclear non-heme iron enzyme that installs a β-hydroxyl group onto aspartate and histidine residues of a RiPP precursor peptide.[12,13] This activity depends on a preinstalled cyclophane scaffold (Figure 1A). Structural analysis revealed a fold reminiscent of the zincin protease family yet distinct from any previously characterized αKG-dependent enzymes, underscoring an unusual evolutionary trajectory. While αKG-dependent enzymes are widespread in biology,[14,15] within RiPP biosynthesis their characterized activities remain rare and are largely confined to hydroxylation.[16-21] Recently, the first αKG-dependent-enzyme-catalyzed crosslink formation between a Cys and Tyr was reported, expanding the repertoire of these enzymes.[22] These findings highlight the untapped chemistries embedded in RiPP biosynthetic gene clusters.

Our laboratory recently reported that the αKG-HEXXH protein BpmH is involved in the biosynthesis of biphenomycin-like macrocyclic peptides.[8] In this study, we characterize two other novel examples of HEXXH proteins as well as an MNIO that is fused to a nitroreductase. The HEXXH enzymes unexpectedly catalyze two different transformations. They hydroxylate a string of consecutive glutamine residues in their peptide substrate and effect oxidative cleavage of the backbone after the last glutamine at a conserved ARMD motif. The MNIO-nitroreductase fusion enzyme hydroxylates an Asp residue and dehydrogenates a Phe residue to form a *Z*-dehydrophenylalanine, a transformation not seen previously in RiPPs. These new activities may provide starting points for development of useful biocatalysts.

Results

Biosynthetic Gene Cluster

We first generated a sequence similarity network (SSN) using the Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST)[23,24] using the αKG-dependent non-heme iron HEXXH-type β-hydroxylase domain family (IPR026337) as input with an alignment score of 30. Network analysis revealed that the largest cluster in the network belongs to rSAM-fused HEXXH enzymes, which contains the previously characterized ChlBH (Figure 1B, highlighted in a blue triangle).[12] The hydroxylase BpmH from the biphenomycin pathway is located in a different cluster (Figure 1B, highlighted in a green triangle). A smaller cluster features enzymes encoded in BGCs that contain genes encoding a MNIO-fused nitroreductase in addition to the predicted αKG-HEXXH enzyme (Figure 1B, highlighted in a purple circle). Flavin-dependent nitroreductases have been widely found in RiPP BGCs and usually co-localize with YcaO proteins that together typically install azoles.[25] Nitroreductases that are not associated with YcaO proteins or that are fused to MNIOs are rare and functionally uncharacterized suggesting the possibility of new chemistry. We chose a BGC from the genome of *Pseudomonas fluorescens* CH267 (*pfl,* Figure 2A). It encodes a RiPP precursor peptide (PflA), a member of the GCN-5 related acetyltransferase (GNAT) family (PflB), the HEXXH-motif protein (PflC), the MNIO-fused nitroreductase (PflD), and two transporters (PflE and PflF). No genes encoding proteases or protease domains were found nearby. Assuming that this system, like all previously characterized RiPP biosynthetic pathways,[1] contains a sequence in the precursor peptide that is used for recognition by the biosynthetic enzymes (typically called a leader or a follower peptide), the proteolytic machinery required to remove this sequence is likely encoded elsewhere in the genome, as observed for a subset of RiPPs.[26] ClusterBlast analysis[27] showed that this type of BGCs is prevalent among Pseudomonas species (Figure S1). Multisequence alignment of the precursor peptides shows a highly conserved C-terminal consecutive Gln motif followed by an ARMD sequence (Figures 2B and S2).

Co-expression of PflA with PflC and PflD

We heterologously expressed the N-terminally His_6-tagged precursor peptide (PflA) with individual enzymes of the *pfl* BGC in *Escherichia coli* using codon optimized genes (Table S1). After Ni^{2+} affinity purification, analysis by high-resolution mass spectrometry (HRMS) demonstrated that the peptide co-expressed with PflC (denoted PflAC) underwent a mass loss of −394 Da (Figure 2C). When PflA was co-expressed with PflD, two products were observed that were changed by +14 Da and −2 Da (Figure S3). Considering that PflD is a fusion of two enzymes,

the observed mass changes are likely a combination of −2 Da and +16 Da modifications. When PflA was co-expressed with both PflC and PflD, mass changes of −380 Da and −396 Da were observed, consistent with the individual mass shift assignments of PflC and PflD (Figure 2C and S4).

As noted above, the BGC does not contain a gene encoding a protease. Therefore, to help localize the modifications, we treated the PflCD-modified PflA (denoted as PflACD) with the commercial protease LysC. Analysis of the proteolysis products by MS showed that the modifications were all located in the C-terminal fragment that starts at a Leu that we designate residue 1 for this study (Figure S5). Co-expressions of PflA with PflBCD did not lead to an additional change in mass. In the Discussion section we provide possible explanations for the lack of PflB activity.

Isolation of purified peptides

To assign molecular structures to the observed mass changes introduced by PflC and PflD, we first sought to purify the proteolytic fragments. The LysC-digested PflACD fragment was poorly soluble in most common buffers and organic solvents with or without acidification. We then turned to analyze the PTMs in a homologous BGC from *Pseudomonas sp*. Os17 (*pos* BGC, Figure S6, Table S2). A −410 Da mass change was observed in PosA after co-expression with PosC in *E. coli* (Figure S7). LysC-digestion of the PosAC product resulted in precipitation of the C-terminal fragment with all other fragments remaining in solution. After three washes of the solid with ammonium acetate buffer (100 mM, pH 8), we isolated the pure modified fragment of the PosA peptide (PosAC), and after drying using a lyophilizer a white powder was obtained. Through this simple workup, we isolated 2-3 mg modified peptide from 1 L cell culture.

The same operation did not allow analogous isolation of LysC-digested PflACD. We did succeed in isolation of the PflD-modified PflA (denoted as PflAD) by co-expression of PflA and PflD in *E. coli*. After immobilized metal affinity chromatography (IMAC) purification, LysC digestion, and RP-HPLC purification, 2 mg of a product fragment was isolated from 2 L culture as a mixture of peptide that had undergone a −2 Da change (denoted as PflAD1) and unmodified PflA. We failed to isolate the minor product in PflAD with a mass gain of +14 Da due to its low concentration. We then characterized PosAC and PflAD1 by a series of one-dimensional (1D) and two-dimensional (2D) NMR experiments as described below.

NMR characterization of the modifications

The LysC-digested PosAC peptide obtained as described in the previous section was dissolved in d_6-DMSO. Twelve backbone amide protons were observed in the ^{1}H and 2D ^{1}H-^{1}H TOCSY spectra (Figures S8, S9). Together with 2D ^{1}H-^{1}H NOESY, ^{1}H-^{13}C HSQC, and ^{1}H-^{15}N-HSQC spectra (Figures S8, S9), a complete assignment of the peptide was achieved (Table S3), indicating a 13 amino acids sequence of IGVAAIFDTQQQQ. These NMR data indicated that the last four residues ARMD at the C-terminus of PosA were removed, explaining why the C-terminal LysC fragment that was expected to consist of 17 residues, only showed twelve backbone amide signals. In the ^{1}H-^{15}N HSQC spectrum, in addition to the four pairs of NH_2 side chain signals from the four Gln residues, an extra pair of amido NH_2 signals were detected (Figure S9A). These latter signals showed NOESY cross peaks with the side chain amide and α/β protons of the last Gln residue (Figure 3A), suggesting C-terminal amidation. The ^{1}H-^{13}C HSQC spectrum displayed multiple downfield CH (methine) cross peaks around 69 ppm that were not anticipated based on the sequence of the peptide, suggesting a nearby electronegative oxygen atom (Figure S8C). The TOCSY and ^{1}H-^{13}C HMBC data showed that each of the four methine signals were part of Gln spin systems in which one of the expected CH_2 groups was absent, as confirmed by the ^{1}H-^{13}C HSQC spectrum, consistent with hydroxylations of four Gln residues. Detailed analysis of the TOCSY and NOESY data (Figure S9B and S10) confirmed that the OH groups were installed at the β-position of the four Gln side chains. Collectively, these data demonstrate that PosC catalyzed β-hydroxylation of four Gln residues and removal of the ARMD-motif to generate a terminal amide. The net result of these modifications explain the observed mass change of −410 Da (4 x 16 = +64 Da; loss of ARMD = −490 Da; C-terminal amidation +16). The conclusions from the NMR data were corroborated by tandem mass spectrometry. HR-MS/MS analysis (Figure 3B) clearly demonstrated a 15.99 Da mass gain of the four Gln residues and confirmed an amidated C-terminus. These data suggest that the observed −394 Da change observed in PflAC is the result of oxidative removal of the ARMD motif in this peptide resulting in an amide, and hydroxylation of the five Gln residues.

For PflAD1, LysC digestion generated an 18-mer peptide with the sequence of LGTAAIFDQ-QQQIQARMD. The Phe-Asp segment within the sequence had undergone modifications that were unambiguously assigned based on TOCSY, NOESY and ^{1}H-^{13}C HSQC spectra (Table S4). The modified part of the peptide was identified via sequential NOESY cross peaks of amide protons at 8.06 ppm (Ala5), 7.91 ppm (Ile6), 9.79 ppm (Phe7), 8.28 ppm (Asp8), and 7.80 ppm (Gln9). The considerable downfield shift of the amide proton of Phe7 is consistent with a loss of 2 Da from this residue to form a dehydro-Phe. The absence of an α-proton at residue

7 and the observation of a singlet at 6.96 ppm consistent with a vinyl proton confirms generation of dehydroPhe. The configuration of the dehydro-Phe residue was established as *Z* by a strong NOE observed between the singlet vinyl proton of former Phe7 at 6.96 ppm and the amide proton at 8.28 ppm of Asp8 (Figure 3C and S11). NOE correlations were also detected between the aromatic side chain (7.55 ppm) of former Phe7 and the side chain protons of Ile6, further supporting a *Z* configuration.

All NMR assignments in PosAC and PflAD1 were corroborated in the fully modified PflACD product by high-resolution tandem electron ionization mass spectrometry (HR-ESI-MS/MS) analysis of the LysC-digested PflACD fragment. The fragmentation patterns are consistent with hydroxylations of all five Gln residues, a C-terminal amide and removal of the ARMD sequence, and loss of 2 Da at Phe7 (Figure 3D). These MS/MS data further showed a 15.99 Da gain at Asp8 (Figure 3D). DehydroPhe formation (−2 Da) is likely catalyzed by the nitroreductase domain of PflD whereas the hydroxylation of Asp8 (+16 Da) is likely catalyzed by the MNIO domain, explaining the net +14 Da change in mass in the minor product of PflAD.

Stereochemistry of hydroxylation by PflC and PflD

Enantiopure β-hydroxylated glutamine (or glutamic acid) is not commercially available. Therefore, we constructed a variant carrying a Q90N mutation in the PosA peptide (numbering here based on full length precursor peptide). When co-expressed with PosC, this mutant underwent the same modifications as the wild type (WT) PosA peptide, resulting in a −410 Da mass change (Figure S12). Marfey's analysis of this mutant product alongside authentic standards assigned the hydroxylated Asn residue as having the (2*S*,3*R*) configuration (Figure S13A). While we cannot completely exclude the possibility that the WT substrate may result in a different stereochemical outcome for Gln hydroxylation, the close structural and biosynthetic parallels between the WT and Q90N variant suggest that they likely share the same stereoselectivity. Marfey's analysis of the PflD-modified PflA alongside authentic standards assigned the hydroxylated Asp8 residue as having the (2*S*,3*S*) configuration (Figure S13B), which is different from the HEXXH PosC-hydroxylated Asn.

In vitro activity of PflC and PosC

To further investigate the unexpectedly complicated modification caused by the HEXXH enzymes, we reconstituted their activity in vitro. We first heterologously expressed His_6-tagged PflC and PosC, and the corresponding precursor peptides (PflA and PosA) in *E. coli*, followed by

IMAC purification. When incubated with their cognate modification enzymes (peptide to enzyme ratio 10:1) in the presence of Fe(II), ascorbic acid, and αKG, the fully modified peptides were observed (Figure 4A). Control experiments showed that αKG was essential for the activity, and ascorbic acid enhanced the catalytic efficiency (Figure 4A). Omission of Fe(II) resulted in minor activity, likely due to some trace metal in the protein during purification (Figure 4A). We also investigated cross-activity by incubating PflA with PosC or PosA and PflC under standard conditions, and in both cases, incomplete hydroxylation and deletion of the ARMD motif were observed (Figure S14). The reconstituted in vitro activity also allowed identification of the second product from peptide backbone cleavage. Based on other oxidative peptide backbone cleavage events, we anticipated production of a ketone generated from the ARMD sequence. Therefore, we treated the in vitro reaction products with *O*-benzylhydroxylamine and observed the anticipated mixture of *E* and *Z* oxime products for both PflC and PosC reactions (Figure S15).

Substrate selectivity of PosC

We next evaluated the sequence selectivity of PosC. To probe the role of the C-terminal ARMD sequence, we performed an alanine scan at each position (PosA-R92A, PosA-M93A, and PosA-D94A). Co-expression of these precursor variants with PosC demonstrated that the Arg, Met, and Asp residues are individually dispensable, as all three mutants were fully modified to a product bearing four hydroxylations of the consecutive Gln residues with concomitant oxidative removal of the last four residues (AXXX; Figure S16). We also generated the PosA-Q90A variant to form a C-terminal QQQ-AARMD sequence; co-expression with PosC yielded a major product corresponding to hydroxylations of the three Gln residues with deletion of ARMD and a minor product corresponding to hydroxylation of the three Gln residues with deletion of AARMD. These observations indicate that the enzyme preferentially acts on Ala in the ARMD motif rather than an Ala after a series of (hydroxylated) Gln residues (Figure 4B). Substituting Ala91 with Gln (PosA-A91Q) abolished peptide scission and resulted in a product with hydroxylation of all five Gln residues in the variant peptide (Figure 4C). Likewise, the double mutant A91Q/R92A afforded the same five-hydroxylated species without AMD cleavage (Figure S17). When the C-terminal ARMD was deleted (PosA_ΔARMD), hydroxylation of two Gln residues was observed as the major product, along with a minor triple hydroxylation product (Figure S18), suggesting the C-terminal ARMD sequence enhanced Gln hydroxylation efficiency.

To test whether ARMD must reside at the C-terminus for peptide backbone cleavage, we inserted an alanine after ARMD (PosA-ARMDA) and in another variant, swapped the positions of

ARMD and the four Gln residues (PosA-ARMDQQQQ). In both cases, PosC consistently cleaved immediately upstream of the ARMD sequence (Figure 4D and E), illustrating that the ARMD motif does not need to be at the C-terminus for oxidative cleavage. We also prepared a variant harboring two consecutive ARMD motifs by mutating the four consecutive Gln residues to ARMD (PosA-ARMDARMD). Coexpression of this variant with PosC produced exclusively the double-deletion product (Figure 4F). Conversely, two variants that contained an additional third ARMD motif further upstream did not show any cleavage at the first (most N-terminal) ARMD motif. These variants underwent cleavage at either the second or third ARMD sequence or both (Figure 4, bottom box and Figure S19). Collectively, these results suggest that productive cleavage requires an ARMD motif that does not need positioning at a specific location within the peptide as long as the motif is a certain distance of the N-terminal leader peptide. Cleavage also does not appear to be dependent on the residue before the cleavage site because the ARMD motif can be preceded by a hydroxylated Gln, a hydroxylated Asn, an Ala, or a Thr (Figures 4B,C,E,F and Figure S12).

Amidation by oxidative enzymes is a process that has been observed previously including for the peptidylglycine α-amidating enzyme[28,29] and the MNIO MovX.[30] These enzymes usually catalyze hydrogen atom transfer (HAT)[31] from the α-carbon of the amino acid that is removed. A β-HAT pathway is also possible for PosC/PflC, and would be consistent with the observed β-hydroxylation of the Gln residues. Therefore, we prepared PosA-A91S as a possible intermediate of Ala oxidation. Upon co-expression with PosC, the peptide displayed hydroxylation at the four Gln residues and an additional −2 Da mass shift, while the backbone remained intact. LysC digestion and HR-MS/MS localized the −2 Da shift to Ser91, consistent with oxidation of the Ser side-chain primary alcohol to an aldehyde (Figure S20 and S21). Although the outcome with the PosA-A91S variant suggests that PosC can abstract a hydrogen atom from the β-position at position 91, as with the Gln hydroxylations, the observation that this did not result in backbone cleavage for both A91S and A91Q variants suggests that amide bond cleavage is achieved by hydrogen atom abstraction from the α-carbon, as in other α-amidating enzymes, and that an Ala residue is required for this process. The identification of the ketone products originating from the ARMD motif (Figure S15) further supports this mechanistic proposal.

PflC can catalyze oxidative and hydrolytic peptide backbone cleavage

Considering the apparently high sequence specificity of the HEXXH enzymes for peptide backbone fragmentation, and to assess whether the N-terminal part of the substrate functions as

leader peptide, we turned to investigate the minimal sequence requirement of PflC. We used solid-phase-peptide synthesis (SPPS) to prepare several truncated N-acetylated PflA precursor peptides (Figure S22). PflC was incubated with the truncated substrates in the presence of Fe(II), ascorbic acid and αKG. Hydroxylation activity was not observed, but unexpectedly, instead C-terminal hydrolytic release of ARMD was detected (Figures S22, S23). This process left a C-terminal carboxylic acid instead of amide, as confirmed by HR-MS/MS (Figure 5 and Figure S24). When the reaction was performed in $H_2{}^{18}O$-containing buffer, a +2 Da mass increase in the product confirmed that the oxygen atom originated from water (Figure 5). This observation established peptidase activity of PflC in the absence of the N-terminal sequence of PflA, which in other RiPP pathways serves as a leader peptide. These observations suggest that in this pathway, the N-terminus of the precursor peptides also serves as a leader peptide that ensures proper biocatalytic transformations.

Discussion

The investigation of the *pfl* and *pos* BGCs in this study revealed unexpected catalytic activities that are unprecedented in RiPP biochemistry. Similar BGCs are found in other *Pseudomonas* strains (Figure S1). Given the lack of a protease gene in the BGCs, we were not able to investigate the function of the final natural product of the pathways. Although it is possible that the full length products of PflA and PosA are the mature products, in all previously characterized RiPPs removal of a leader/follower peptide is involved in formation of the final product.[1,26] The presence of a gene encoding an acetyl transferase PflB, which did not display any activity on full length peptides, also suggests that it will act after removal of an N-terminal leader peptide; N-terminal acylation by a GNAT-like enzyme has been reported for several RiPPs.[21,32-34] N-terminal acylation, as opposed to acylation of Lys side chains,[21,35] is supported by the observation that the sequence logo of the precursor peptides (Figure 2B) does not reveal a highly conserved Lys residues in the precursor peptides. Alternatively, we cannot rule out that a non-specific aminopeptidase encoded elsewhere in the genome may act on the modified peptides, as observed for several RiPPs.[25] In that scenario, the aminopeptidase would likely terminate at the *Z*-dehydroPhe, resulting in formation of an N-terminal ketone upon hydrolysis (Figure S25), similar to the known generation of N-terminal 2-oxobutyryl and 2-oxopropionyl groups from dehydrobutyrine and dehydroalanine, respectively.[36] If this were the case, the GNAT might acylate one of the introduced hydroxyl groups similar to the recently reported acylation of the hydroxyl group of a Tyr residue in a lasso peptide.[37]

The HEXXH enzymes PflC and PosC both catalyze hydroxylation of the β-carbon of a string of Gln residues, as well as oxidative removal of a C-terminal ARMD tetrapeptide. These two transformations appear mostly independent because hydroxylation of Gln was still observed when the ARMD motif was deleted, and backbone cleavage was still observed when the Gln residues were absent. Because we only see a product with four Gln hydroxylations and backbone cleavage with the wild type PosA substrate, whereas a peptide lacking the ARMD motif produced mostly a product with two hydroxylations (Figure S15), we infer that Gln hydroxylations precede cleavage of the backbone. However, we cannot rule out that the amide-terminated backbone-cleaved peptide is a more efficient substrate for Gln hydroxylation than the carboxylate-terminated peptide. The length of the tetrapeptide appears not critical as extension at the C-terminus still resulted in backbone cleavage. Other than the Ala residue, the individual residues of the ARMD motif also did not appear critical, despite their complete conservation (Figure 2B). A possible mechanism for backbone scission in which an initial radical generated on the β-carbon of the last Gln would abstract a hydrogen atom from the Ala residue of the ARMD motif to initiate cleavage also seems unlikely because for such a mechanism, insertion of an additional Ala (AARMD) would have been expected to result in cleavage of AARMD, but removal of ARMD was still observed. These observations further corroborate that modification of these Gln residues is not required for oxidative backbone cleavage. Furthermore, the variant in which all Gln residues were replaced by another ARMD sequence was still cleaved.

To rationalize the site-specific backbone cleavage and amide formation, we propose that Ala91 undergoes α-C–H abstraction by a high-valent iron–oxo species to generate an α-amino carbon radical stabilized by the neighboring amide nitrogen. This intermediate could either (i) undergo oxygen rebound to form an α-hydroxylated species that eliminates water to an iminium, followed by hydrolytic cleavage of the peptide backbone, or (ii) be oxidized by one electron directly to the same iminium intermediate (Figure S26). The observed hydrolysis at the exact same site as the oxidative backbone cleavage when a truncated substrate was used suggests that the truncated peptide binds in the exact same position near the metal ion, but that without the N-terminal substrate sequence, oxygen activation and/or αKG binding does not occur. Substrate gating has been shown to control canonical αKG-dependent enzymes,[38] and the observations with PflC/PosC suggest that the αKG-dependent HEXXH enzymes studied here are similarly substrate-gated for oxidative chemistry. Furthermore, the data suggest that binding of the N-terminal sequence of the substrate appears critical for oxygen activation, consistent with this segment of the substrate acting as a leader peptide, as in other RiPP biosynthetic pathways.[39]

Other observations are also suggestive of an important role of the N-terminal sequence for the regioselectivity of backbone cleavage. When three ARMD motifs were introduced into the precursor peptides, backbone cleavage was only observed at positions that are located towards the C-terminus of the substrates. Installation of the ARMD motif at positions that were more towards the center of the peptide did not result in cleavage (Figure S19). One plausible explanation is that an N-terminal leader peptide positions the C-terminal part of the substrate near the Fe(II) site and that ARMD motifs that are closer to the leader peptide cannot reach the active site. The observation that PosA was not a good substrate for PflC and PflA was not a good substrate for PosC could also be explained by differences in their N-terminal sequences. An AlphaFold3 model of PosC in complex with PosA places the iron center right next to the Ala residue in the C-terminal ARMD motif (Figure S27), consistent with the observed site-selective cleavage.

The nitroreductase domain of PflD catalyzes an atypical dehydrogenation of phenylalanine. Dehydroamino acids (dhAAs) play central roles in the biological activity of both RiPPs and non-ribosomal peptide natural products.[40,41] In bacterial RiPPs, dhAAs are overwhelmingly represented by dehydroalanine (Dha) and dehydrobutyrine (Dhb), typically generated by dehydration/dehydrothiolation of serine or cysteine residues (for Dha), and threonine residues (for Dhb).[1] To the best of our knowledge, this study reports the first dehydrophenylalanine formation in RiPPs. Other nitroreductases in RiPP BGCs that are not associated with YcaO enzymes could be involved in the formation of other dehydroamino acids.

The MNIO domain of PflD catalyzes a reaction that has been observed in several other RiPPs, the hydroxylation of the β-carbon of an Asp residue. In other RiPPs, this reaction is usually catalyzed by either canonical αKG dependent enzymes,[16,18] or by the recently discovered αKG-dependent HEXXH enzymes.[12] The observation that an MNIO catalyzes this reaction was unexpected because most enzymes in this family catalyze four-electron oxidations, which is thought to involve first hydrogen abstraction from a weak C-H bond adjacent to a heteroatom by a ferric-superoxo species, and then a second oxidation involving an iron(IV)-oxo intermediate.[7] The only examples of (potential) two-electron oxidations that are catalyzed by MNIO enzymes, like the hydroxylation of Asp observed here, involve substrates that either can eliminate hydrogen peroxide,[30] or that have electron rich π-type electrons.[8] However, for hydroxylation of Asp, hydrogen atom abstraction is required from a non-activated C-H bond with a relatively high bond strength, a step that appears unlikely for a ferric-superoxo intermediate.[46] Hence, we propose that

the reaction catalyzed by the MNIO domain of PflD likely involves a cellular reductant that allows formation of a Fe(IV)-oxo intermediate.

Conclusions

Two αKG-dependent HEXXH enzymes (PflC and PosC) were characterized in the biosynthesis of a RiPP family from Pseudomonas. Both enzymes catalyze hydroxylation of consecutive Gln residues and specifically recognize a C-terminal tetrapeptide ARMD to promote oxidative peptide backbone cleavage and generation of a C-terminal amide and a ketone containing by-product originating from the tetrapeptide. The enzyme also exhibits proteolytic activity in the absence of the leader peptide, indicating that leader peptide–enzyme interactions can regulate catalytic outcomes. Further investigation is required to elucidate the mechanisms underlying these distinct catalytic selectivities. In addition, we identified a unique MNIO–nitroreductase fusion enzyme that produces a rare dehydrophenylalanine and hydroxylates aspartate. Together, these findings describe new catalytic activities that expand the catalog of post-translational modifications and set up studies to characterize the function of the product of the BGCs.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/

Experimental procedures, Figures S3-S27 showing spectroscopic data, NMR annotations, and AlphaFold3 modeling, and Tables S1-S4 with gene sequences and primers.

AUTHOR INFORMATION

Corresponding Author

Wilfred A. van der Donk −Department of Chemistry and Howard Hughes Medical Institute, University of Illinois at Urbana-Champaign, Urbana, IL, United States; orcid.org/0000-0002-5467-7071;

Email: vddonk@illinois.edu.

Authors

Yao Ouyang – Department of Chemistry and Howard Hughes Medical Institute, University of Illinois at Urbana-Champaign, Urbana, IL, United States; orcid: 0000-0002-8379-5843.

Yue Yu – Department of Chemistry and the Howard Hughes Medical Institute at the University of Illinois at Urbana-Champaign, 1206 W Gregory Drive Urbana, IL 61801, United States; orcid: 0000-0001-8306-4648.

Lingyang Zhu – School of Chemical Sciences NMR Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, United States; orcid: 0000-0002-6657-271X.

Funding

This work was supported in part by a grant from the National Institutes of Health (GM058822 to W.A.vdD.). W.A.vdD is an Investigator of the Howard Hughes Medical Institute. A Bruker UltrafleXtreme mass spectrometer used was purchased with support from the National Institutes of Health (S10 RR027109).

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENTS

We thank Dr. Chandrashekhar Padhi for assistance with bioinformatic analysis, Dr. Lide Cha for providing standards for Marfey's analysis, and Youran Luo for assistance with Marfey's analysis. We also thank Dr. Chandrashekhar Padhi, Ceilia Leso, and Pelham Kogut for assistance with mass spectrometry. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

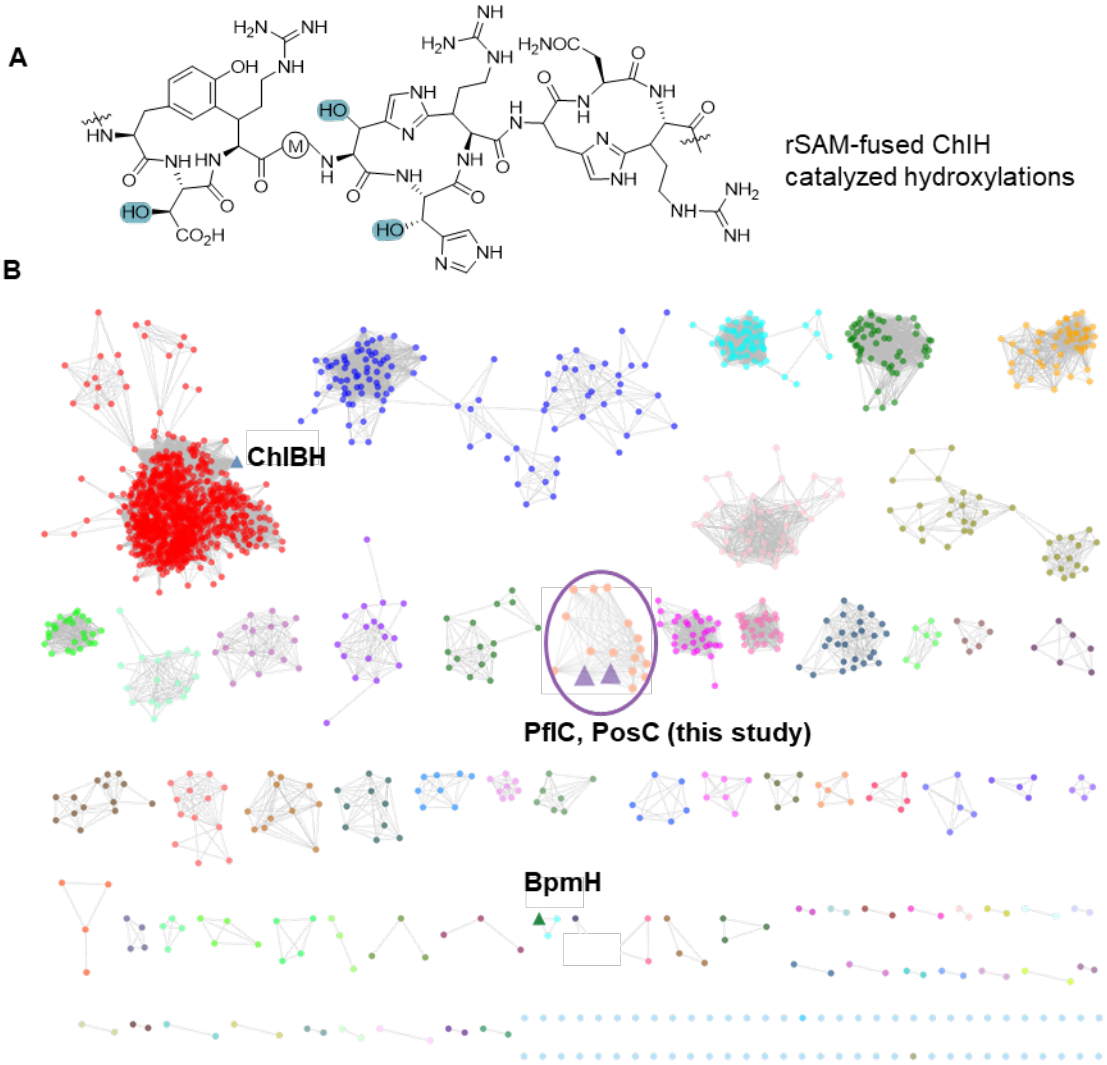


Figure 1. Genome mining of emerging αKG-dependent HEXXH motif enzymes. (A) ChlH catalyzes hydroxylation of His and Asp residues (highlighted in blue) in a peptide with a preinstalled cyclophane scaffold. (B) An SSN of the αKG HEXXH motif enzyme family was generated with the EFI tools using the UniProt database. RepNode 40 is shown as visualized in Cytoscape v3.10 (see Supporting Information). Reported αKG-dependent HEXXH enzymes are highlighted in blue and green triangles. PflC and PosC sequences (Table S2) were added manually because the originally reported Uniprot IDs were withdrawn during recent UniProt consolidation.

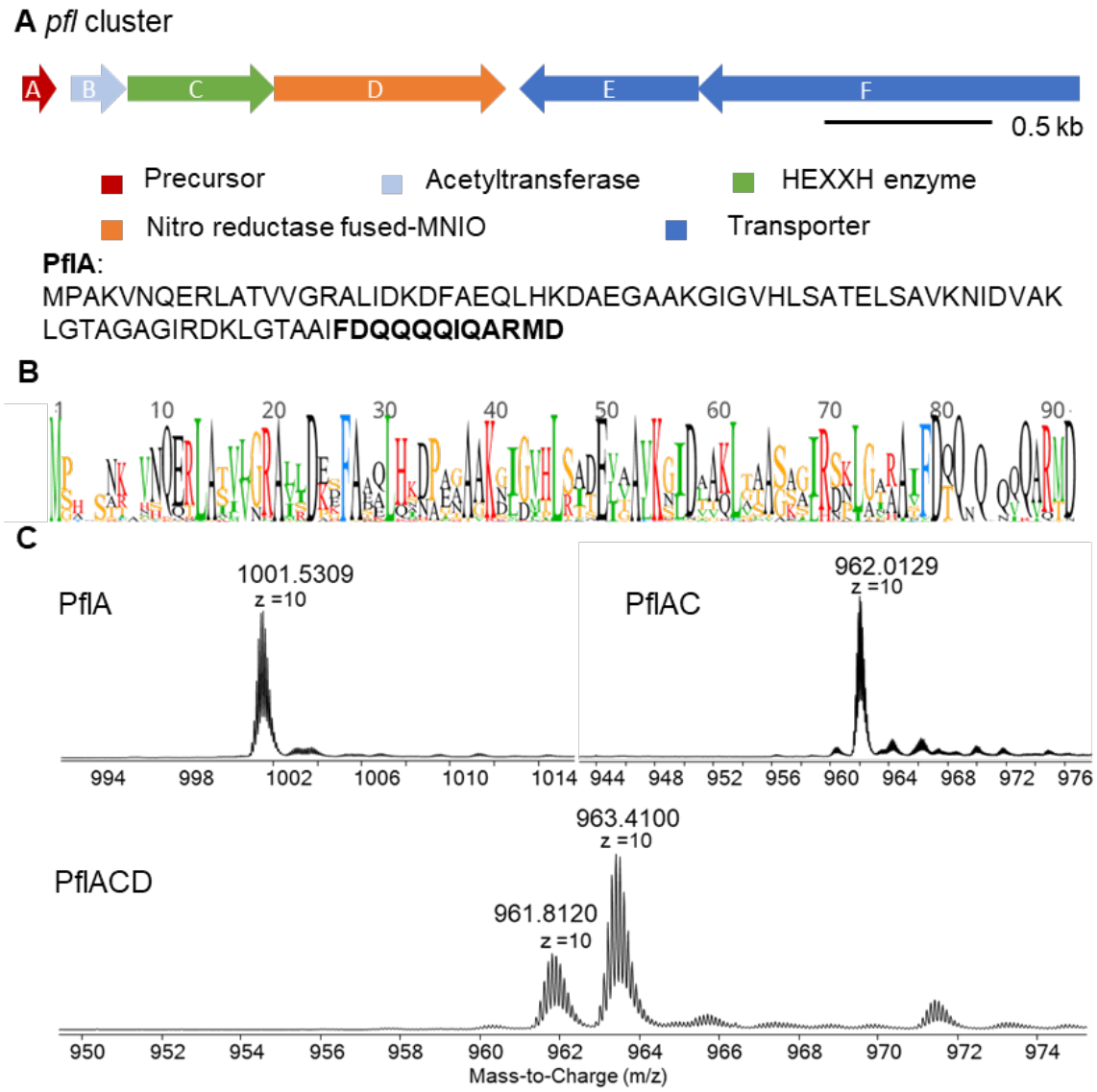


Figure 2. (A) Pfl BGC with the precursor peptide sequence. (B) Sequence logo of precursor peptides. (C) HR-LC-MS spectra of unmodified PflA (calculated m/z 1001.5178, z = 10), co-expressed PflAC (calculated average m/z 962.0599, z = 10), and co-expressed PflACD (calculated average m/z 963.4579, 961.8584, z = 10). Observed average masses are shown.

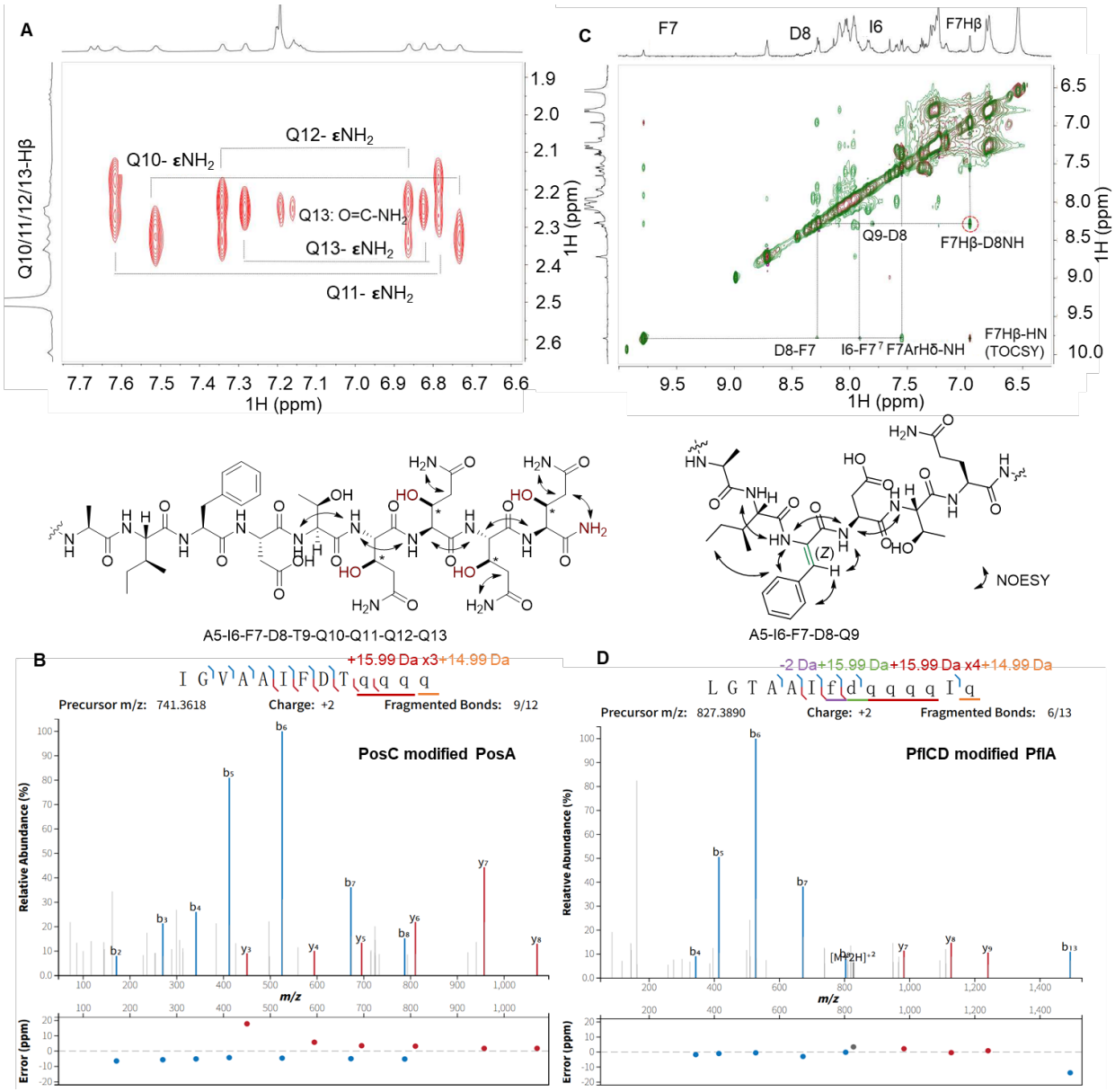


Figure 3. Structure determination of PosAC and PflAD after digestion with LysC. (A) NOESY correlations of the C-terminal PosAC LysC-fragment. For the TOCSY spectrum, see Figure S9B. (B) HR-MS/MS spectrum of the C-terminal PosAC LysC-fragment. (C) Overlay of the aromatic region of the TOCSY-NOESY spectra of the dehydro-Phe peptide (PflAD1). (D) HR-MS/MS spectrum of the C-terminal PflACD LysC-fragment. Fragment ion annotation and ppm errors were generated using the interactive peptide spectral annotator.[42]

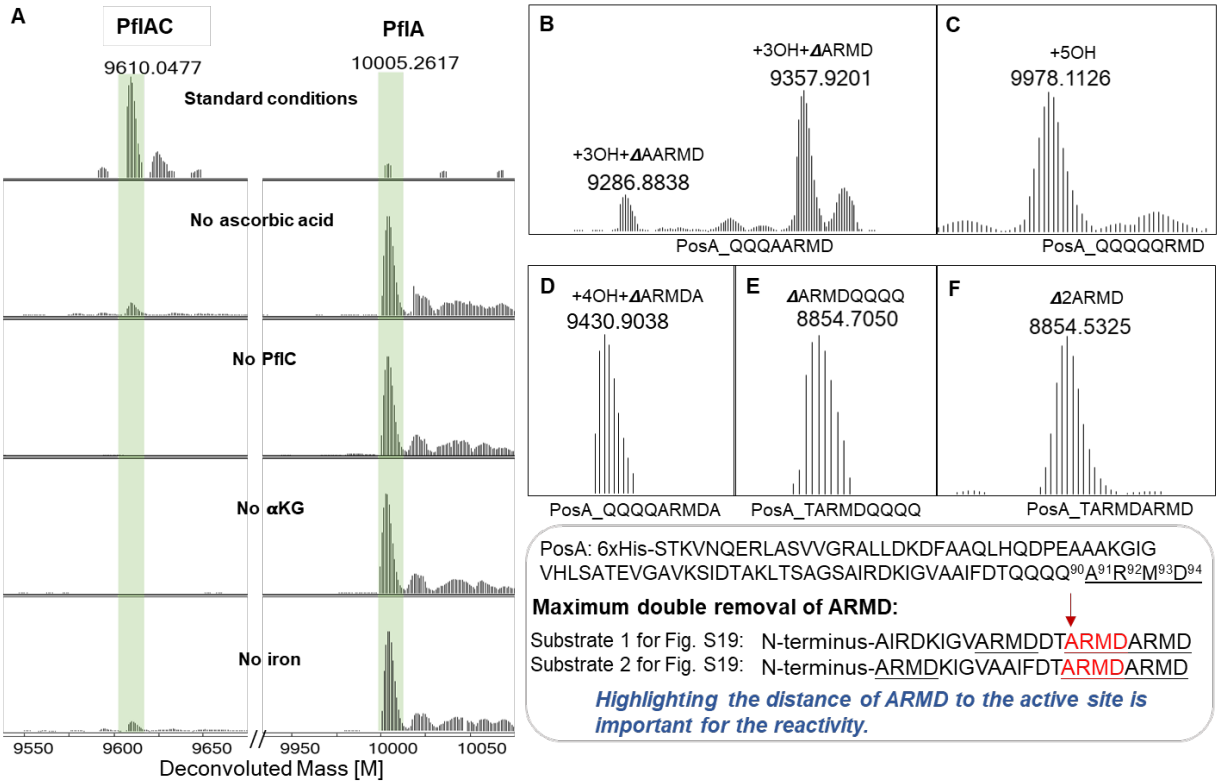


Figure 4. Peptide sequence tolerance test of PflC and PosC. (A) Experiments with full length PflA to determine the requirements for in vitro activity. (B) ESI-MS of PosA-Q90A modified in *E. coli* by PosC. (C) ESI-MS of PosA-A91Q modified in *E. coli* by PosC. (D) ESI-MS of PosA-QARMDA modified in *E. coli* by PosC. (E) ESI-MS of PosA-ARMDQQQQ modified in *E. coli* by PosC. (F) ESI-MS of PosA-2ARMD modified in *E. coli* by PosC. The sequences of the substrate are shown below the spectra, which display deconvoluted masses.

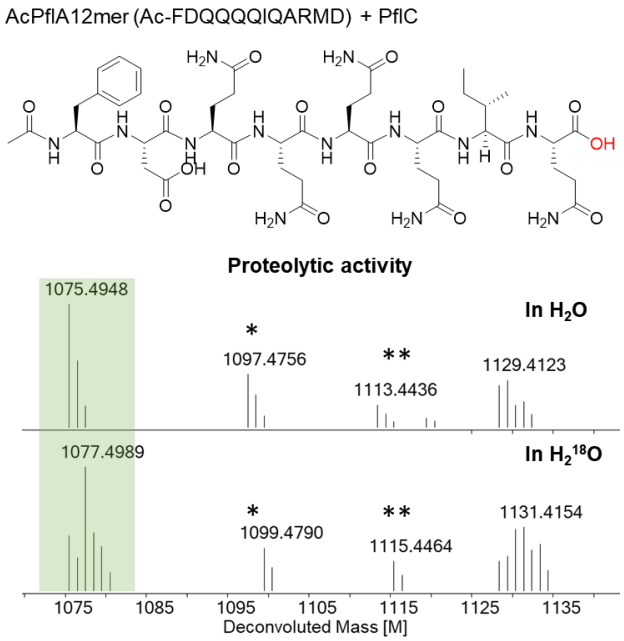


Figure 5. Proteolytic product obtained from a synthetic truncated PflA peptide shown at the top (also see Figures S23 and S24). An isotopic labeling experiment in $H_2{}^{18}O$ labeled buffer indicated the source of the oxygen in the product is water. Calculated [M] 1075.4934 in H_2O, [M] 1077.4977 in $H_2{}^{18}O$. * Na^+ form of the ion; ** K^+ form of the ion.

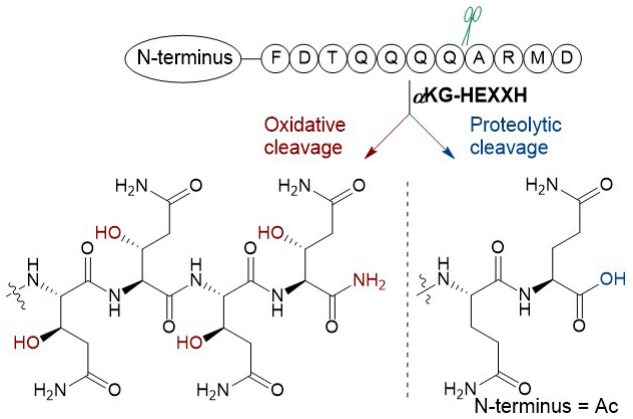
References

(1) 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.; Grande Burgos, M. J.; Hill, C.; Kim, S.; Koehnke, J.; Latham, J. A.; Link, A. J.; Martínez, B.; Nair, S. K.; Nicolet, Y.; Rebuffat, S.; Sahl, H.-G.; Sareen, D.; Schmidt, E. W.; Schmitt, L.; Severinov, K.; Süssmuth, R. D.; Truman, A. W.; Wang, H.; Weng, J.-K.; van Wezel, G. P.; Zhang, Q.; Zhong, J.; Piel, J.; Mitchell, D. A.; Kuipers, O. P.; van der Donk, W. A. New developments in RiPP discovery, enzymology and engineering. *Nat. Prod. Rep.* **2021**, *38*, 130-239.
(2) Cao, L.; Do, T.; Link, A. J. Mechanisms of action of ribosomally synthesized and posttranslationally modified peptides (RiPPs). *J. Ind. Microbiol. Biotechnol.* **2021**, *48*, kuab005.
(3) Li, Y.; Rebuffat, S. The manifold roles of microbial ribosomal peptide-based natural products in physiology and ecology. *J. Biol. Chem.* **2020**, *295*, 34-54.
(4) Ongpipattanakul, C.; Desormeaux, E. K.; DiCaprio, A.; van der Donk, W. A.; Mitchell, D. A.; Nair, S. K. Mechanism of action of ribosomally synthesized and post-translationally modified peptides. *Chem. Rev.* **2022**, *122*, 14722-814.
(5) Richter, D.; Piel, J. Novel types of RiPP-modifying enzymes. *Curr. Opin. Chem. Biol.* **2024**, *80*, 102463.
(6) Nguyen, D. T.; Mitchell, D. A.; van der Donk, W. A. Genome mining for new enzyme chemistry. *ACS Catal.* **2024**, *14*, 4536-53.
(7) Chen, J. Y.; van der Donk, W. A. Multinuclear non-heme iron dependent oxidative enzymes (MNIOs) involved in unusual peptide modifications. *Curr. Opin. Chem. Biol.* **2024**, *80*, 102467.
(8) Padhi, C.; Zhu, L.; Chen, J. Y.; Huang, C.; Moreira, R.; Challis, G. L.; Cryle, M. J.; van der Donk, W. A. Biosynthesis of biphenomycin-like macrocyclic peptides by formation and cross-linking of ortho-tyrosines. *J. Am. Chem. Soc.* **2025**, *147*, 23781-96.
(9) Strunk, E.; Lobert, A.; Khorovich, T.; Guzman Lucio, K. M.; Richarz, R.; Hohmann, M.; D'Agostino, P. M.; Gulder, T. A. M. Biosynthesis of the biphenomycin family of potent antibiotics *ChemRxiv* **2025**, doi:10.26434/chemrxiv-2025-vkzfb.
(10) Hooper, N. M. Families of zinc metalloproteases. *FEBS Lett.* **1994**, *354*, 1-6.
(11) Lipscomb, W. N.; Sträter, N. Recent advances in zinc enzymology. *Chem. Rev.* **1996**, *96*, 2375-434.
(12) Morishita, Y.; Ma, S.; De La Mora, E.; Li, H.; Chen, H.; Ji, X.; Usclat, A.; Amara, P.; Sugiyama, R.; Tooh, Y. W.; Gunawan, G.; Pérard, J.; Nicolet, Y.; Zhang, Q.; Morinaka, B. I. Fused radical SAM and αKG-HExxH domain proteins contain a distinct structural fold and catalyse cyclophane formation and β-hydroxylation. *Nat. Chem.* **2024**, *16*, 1882-93.
(13) Ruel, J.; Nguyen, T. Q. N.; Morishita, Y.; Usclat, A.; Martin, L.; Amara, P.; Kieffer-Jaquinod, S.; Stefanoiu, M. C.; de la Mora, E.; Morinaka, B. I.; Nicolet, Y. Peptide recognition and mechanism of the radical S-adenosyl-L-methionine multiple cyclophane synthase ChlB. *J. Am. Chem. Soc.* **2025**, *147*, 16850-63.
(14) Loenarz, C.; Schofield, C. J. Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* **2008**, *4*, 152-6.

(15) Gao, S. S.; Naowarojna, N.; Cheng, R.; Liu, X.; Liu, P. Recent examples of alpha-ketoglutarate-dependent mononuclear non-haem iron enzymes in natural product biosyntheses. *Nat. Prod. Rep.* **2018**, *35*, 792-837.
(16) Ökesli, A.; Cooper, L. E.; Fogle, E. J.; van der Donk, W. A. Nine post-translational modifications during the biosynthesis of cinnamycin. *J. Am. Chem. Soc.* **2011**, *133*, 13753-60.
(17) Huo, L.; Ökesli, A.; Zhao, M.; van der Donk, W. A. Insights into the biosynthesis of duramycin. *Appl. Environ. Microbiol.* **2017**, *83*, e02698-16.
(18) Zhang, C.; Seyedsayamdost, M. R. CanE, an iron/2-oxoglutarate-dependent lasso peptide hydroxylase from *Streptomyces canus*. *ACS Chem. Biol.* **2020**, *15*, 890-4.
(19) Kim, H. W.; Kang, S.; Kim, S.; Lee, H.; Hur, Y.; Song, W. J.; Oh, D. C.; Kim, S. Discovery of a two-step enzyme cascade converting aspartate to aminomalonate in peptide natural product biosynthesis. *J. Am. Chem. Soc.* **2025**, *147*, 20909-18.
(20) Hollands, S.; Tasch, J.; Simon, D. J.; Wassouf, D.; Barber, I.; Gessner, A.; Bechthold, A.; Zechel, D. L. Analysis of the cryptic biosynthetic gene cluster encoding the RiPP curacozole reveals a phenylalanine-specific peptide hydroxylase. *Chem. Sci.* **2024**, *15*, 19858-69.
(21) Hubrich, F.; Bösch, N. M.; Chepkirui, C.; Morinaka, B. I.; Rust, M.; Gugger, M.; Robinson, S. L.; Vagstad, A. L.; Piel, J. Ribosomally derived lipopeptides containing distinct fatty acyl moieties. *Proc. Natl. Acad. Sci. U S A* **2022**, *119*.
(22) Xia, Y.; Shi, Y.; Xiao, Z.; Xie, X.; Wang, B.; Yao, H.; Wang, H. α-Ketoglutarate-dependent dioxygenase-catalyzed macrocyclization in RiPPs. *J. Am. Chem. Soc.* **2025**, *147*, 26693-703.
(23) Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI web resource for genomic enzymology tools: Leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **2019**, *58*, 4169-82.
(24) Oberg, N.; Zallot, R.; Gerlt, J. A. EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) web resource for genomic enzymology tools. *J. Mol. Biol.* **2023**, *435*, 168018.
(25) Burkhart, B. J.; Schwalen, C.; Mann, G.; Naismith, J. H.; Mitchell, D. A. YcaO-dependent posttranslational amide activation: biosynthesis, structure, and function. *Chem. Rev.* **2017**, *117*, 5389-456.
(26) Eslami, S. M.; van der Donk, W. A. Proteases involved in leader peptide removal during RiPP biosynthesis. *ACS Bio. Med. Chem. Au* **2023**, *4*, 20-36.
(27) Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, M. H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29-w35.
(28) Francisco, W. A.; Merkler, D. J.; Blackburn, N. J.; Klinman, J. P. Kinetic mechanism and intrinsic isotope effects for the peptidylglycine alpha-amidating enzyme reaction. *Biochemistry* **1998**, *37*, 8244-52.
(29) Yu, Y.; Guo, H.; Zhang, Q.; Duan, L.; Ding, Y.; Liao, R.; Lei, C.; Shen, B.; Liu, W. NosA catalyzing carboxyl-terminal amide formation in nosiheptide maturation via an enamine dealkylation on the serine-extended precursor peptide. *J. Am. Chem. Soc.* **2010**, *132*, 16324-6.
(30) Chioti, V. T.; Clark, K. A.; Ganley, J. G.; Han, E. J.; Seyedsayamdost, M. R. N-Cα bond cleavage catalyzed by a multinuclear iron oxygenase from a divergent methanobactin-like RiPP gene cluster. *J. Am. Chem. Soc.* **2024**, *146*, 7313-23.

(31) Martinez, S.; Hausinger, R. P. Catalytic mechanisms of Fe(II)- and 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* **2015**, *290*, 20702-11.
(32) He, Z.; Yuan, C.; Zhang, L.; Yousef, A. E. N-Terminal acetylation in paenibacillin, a novel lantibiotic. *FEBS Lett.* **2008**, *582*, 2787-92.
(33) Ziemert, N.; Ishida, K.; Liaimer, A.; Hertweck, C.; Dittmann, E. Ribosomal synthesis of tricyclic depsipeptides in bloom-forming cyanobacteria. *Angew. Chem., Int. Ed.* **2008**, *47*, 7756-9.
(34) Zhu, S.; Hegemann, J. D.; Fage, C. D.; Zimmermann, M.; Xie, X.; Linne, U.; Marahiel, M. A. Insights into the unique phosphorylation of the lasso peptide paeninodin. *J. Biol. Chem.* **2016**, *291*, 13662-78.
(35) Asamizu, S. Recent advances in discovery and biosynthesis of ribosomally synthesized and post-translationally modified peptides (RiPP)-derived lipopeptides. *Nat. Prod. Rep.* **2025**, *42*, 1622-38.
(36) Repka, L. M.; Chekan, J. R.; Nair, S. K.; van der Donk, W. A. Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chem. Rev.* **2017**, *117*, 5457-520.
(37) Li, Y. Y.; Li, X. R.; Lu, J. L.; Cui, J. J.; Zhai, W.; Xiong, J.; Han, H.; Gao, K.; Lei, X.; Zhang, C. R.; Luo, S.; Dong, S. H. Discovery and biosynthesis of lipolasso peptides reveal CoA-transferase for tyrosine lipidation of ribosomal peptides. *Org. Lett.* **2025**, doi:10.1021/acs.orglett.5c03981.
(38) Goudarzi, S.; Iyer, S. R.; Babicz, J. T., Jr.; Yan, J. J.; Peters, G. H. J.; Christensen, H. E. M.; Hedman, B.; Hodgson, K. O.; Solomon, E. I. Evaluation of a concerted vs. sequential oxygen activation mechanism in α-ketoglutarate-dependent nonheme ferrous enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 5152-9.
(39) Oman, T. J.; van der Donk, W. A. Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat. Chem. Biol.* **2010**, *6*, 9-18.
(40) Siodłak, D. α,β-Dehydroamino acids in naturally occurring peptides. *Amino Acids* **2015**, *47*, 1-17.
(41) Wang, S.; Wu, K.; Tang, Y. J.; Deng, H. Dehydroamino acid residues in bioactive natural products. *Nat. Prod. Rep.* **2024**, *41*, 273-97.
(42) Brademan, D. R.; Riley, N. M.; Kwiecien, N. W.; Coon, J. J. Interactive peptide spectral annotator: A versatile web-based tool for proteomic applications. *Mol. Cell. Proteom.* **2019**, *18*, S193-S201.

TOC graphic


N-terminus
F D T Q Q Q Q A R M D
αKG-HEXXH
Oxidative cleavage
Proteolytic cleavage
N-terminus = Ac