Biochemistry

Insights into the Mechanism of the Cyanobactin Heterocyclase Enzyme

Ying Ge,[†] Clarissa Melo Czekster,[†] Ona K. Miller,[†,⊥] Catherine H. Botting,[†] Ulrich Schwarz-Linek,[†] and James H. Naismith[*,‡,§,∥]

[†]Biomedical Sciences Research Complex, University of St Andrews, St Andrews, Fife KY16 9ST, United Kingdom
[‡]Research Complex at Harwell, Didcot, Oxon OX11 0FA, United Kingdom
[§]Division of Structural Biology, University of Oxford, Oxford OX3 7BN, United Kingdom
[∥]Rosalind Franklin Institute, Harwell, Didcot, Oxon OX11 0FA, United Kingdom

S *Supporting Information*

ABSTRACT: Cyanobactin heterocyclases share the same catalytic domain (YcaO) as heterocyclases/cyclodehydratases from other ribosomal peptide (RiPPs) biosynthetic pathways. These enzymes process multiple residues (Cys/Thr/Ser) within the same substrate. The processing of cysteine residues proceeds with a known order. We show the order of reaction for threonines is different and depends in part on a leader peptide within the substrate. In contrast to other YcaO domains, which have been reported to exclusively break down ATP into ADP and inorganic phosphate, cyanobactin heterocyclases have been observed to produce AMP and inorganic pyrophosphate during catalysis. We dissect the nucleotide profiles associated with heterocyclization and propose a unifying mechanism, where the γ-phosphate of ATP is transferred in a kinase mechanism to the substrate to yield a phosphorylated intermediate common to all YcaO domains. In cyanobactin heterocyclases, this phosphorylated intermediate, in a proportion of turnovers, reacts with ADP to yield AMP and pyrophosphate.
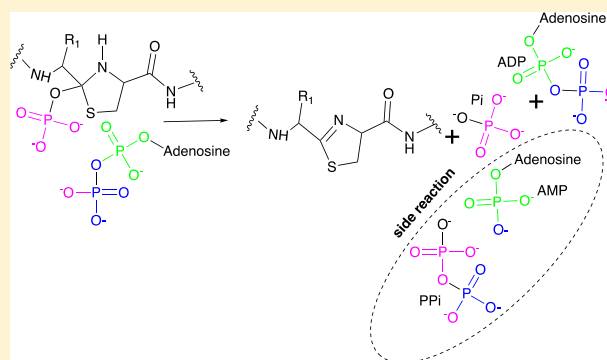
Ribosomal peptide natural products, known as RiPPs, are an intriguing class of genetically encoded post-translationally modified molecules produced by bacteria, plants, and fungi.[1] The substrate peptide has a core of amino acids that becomes the final product and is flanked by recognition sequences. In bacteria, the substrate peptide often contains multiple cores, each giving rise to different products. In contrast to the diversity of the amino acid sequence, the recognition sequences and the processing enzymes are often well conserved between bacteria. The collection of processing enzymes varies but almost always includes a protease and, for macrocyclic RiPPs, a macrocyclase. The exemplar system is that of patellamides, from the bacterium *Prochloron didemni*, a symbiont of *Lissoclinum patella*, where the entire biosynthetic gene cluster has been sequenced.[2−5] The patellamide cluster produces two eight-residue macrocycles, patellamide A and C, from a single substrate peptide, PatE. Each product contains two thiazoles and two oxazolines, as well as two D-amino acids. The structures of the enzymes that carry out the proteolysis (PatA),[6] formation of the azolines PatD homologues TruD and LynD,[7,8] oxidation of the thiazolines (PatG oxidase domain),[9] nonfunctional prenylating enzyme (PatF),[10] and macrocyclization (PatG macrocyclization domain)[11] have been

reported as well as the structure of a domain of unknown function found in both PatA and PatG.[12]

The five-membered heterocyclic rings (azol(in)es) are found in a wide range of RiPPs that include the linear azol(in)e-containing peptides (known as LAPS), cyanobactins, thiopeptides, and bottromycins.[1] The ring results from the formation of a bond between the amino acid side chain oxygen or sulfur atom (from cysteine, serine, or threonine) and the preceding amide bond.[13] ATP and Mg^{2+} are required for heterocyclization enzyme activity, and all heterocyclase enzymes share the same catalytic unit, the YcaO domain (named after the *E. coli* homologue).[14] The YcaO domain works in conjunction with a substrate recognition unit, known as the RiPP recognition element (RRE),[15] which can either be fused to the YcaO domain or occur as a separate protein. The combined system works on multiple residues within the same substrate in a distributive manner. Some enzymes catalyze the formation of both (methyl)oxazolines and thiazolines; others appear to yield only thiazolines. It has been shown for the thiazoline specific enzyme TruD[8] that installation of thiazolines follows a defined

order, with the cysteine closest to the C-terminus of the core being processed first in all substrate molecules, before the enzyme then moves to the next most C-terminal cysteine residue. The order is retained when multiple cassettes are present on the same substrate, with the most C-terminal cysteine on all cassettes being processed first.[16] BalhCD,[17] a heterocyclase involved in the biosynthesis of linear, azol(in)e-containing microcins, also showed in general a C-to-N reaction order. The study on the patellamide system did not address whether the heterocyclization of serine/threonine residues is also ordered, nor whether the leader peptide alone determines the order. This is significant because the leader peptide is known to play a crucial role in both recognition (binds to the RRE) and in regulating enzyme activity (binding orders the active site).[7] The fusion of the leader sequence to LynD created an enzyme that processed peptides without the need for the leader sequence,[7] allowing for the investigation of the role of the leader in the order of heterocyclization.

The ATP chemistry associated with the heterocyclase enzymes has been reported as yielding ADP + P_i or AMP + P_i + PP_i.[7,8,14,18] The most convincing mechanism has ATP phosphorylate, the hemiorthoamide that results from the attack of side chain of the Cys/Ser/Thr residue on the carbonyl of the preceding amide bond—a kinase mechanism (Scheme S1). However, a kinase mechanism is inconsistent with the production of AMP + PP_i (Scheme S2), rather these products would suggest adenylation and diphosphokinase type mechanisms. The nucleotide-bound structures of *Escherichia coli* YcaO protein and the cyanobactin heterocyclase LynD both show an arrangement of ATP that is consistent only with a kinase mechanism, all but eliminating the other mechanisms. Yet, the observation by different laboratories of the production of AMP and PP_i by different YcaO enzymes argues it is not a simple experimental error. Moreover, ^{18}O has been shown to transfer from carbonyl of the substrate peptide to PP_i during turnover.[7] An enzyme operating via two different mechanisms for the same substrate would seem highly unlikely; rather, the observations could point to something important that has been overlooked.

Here we report the order of oxazoline formation with an enzyme that processes both cysteine and serine/threonine residue. We have also investigated the role of the leader peptide in controlling the order of reaction. We have confirmed the production of AMP, ADP, PP_i, and P_i using different enzyme variants during catalysis. ATP analogues were used to eliminate mechanistic hypotheses. We propose that an enzyme-catalyzed event occurs during the breakdown of phosphorylated hemiorthoamide that gives rise to AMP and PP_i production. This event is not stoichiometric, and the extent of its occurrence appears to vary depending on the enzyme employed. We believe this proposal resolves the apparent contradictory results extant in the literature.

■ MATERIALS AND METHODS

General Methods. Materials were purchased from Sigma unless specified. Peptides were purchased from Biosynthesis or expressed and purified with a his-tag.

Protein Production. The construction of the expression plasmid for MicD Q219GA (pJExpress411-MicD Q219GA) was carried out in the following manner (Scheme S3). (Step 1) The plasmid pJExpress411-LynD fusion[7] was subjected to site-directed mutagenesis,[3] using the primer pair 5′-GGC-GCCAAGCTTATGCAATCTACCCCGTTGCTGCAAATT-

3′ and 5′-TGCATAAGCTTGGCGCCGGCGCCTGCA-CCCGCACC-3′ and the KOD hot start DNA polymerase kit (Merck). The mixture was subjected to 12 cycles of denaturation (94 °C for 1 min), annealing (55 °C for 1 min), and extension (68 °C for 1 min and 55 s), followed by 3 cycles of 1 min at 95 °C, 1 min at 47 °C, and 1 min and 55 s at 68 °C. The gene encoding LynD was removed with endonucleases *Hind*III and *Xho*I. (Step 2) The gene encoding MicD (heterocyclase from *Microcystis aeruginosa*) was amplified from pJExpress-411-MicD (DNA 2.0) using the KOD hot start DNA polymerase kit (Merck) following the manufacturer's instruction, with the primer pair 5′-CTTCTTAAGC-TTATGCAGTCGACCCCGCTGCTG-3′ and 5′-C-TTCTTCTCGAGTTAGAACGGGATGTTGGTCTG-3′. The amplified DNA fragment was subjected to endonucleases *Hind*III and *Xho*I and ligated with the vector prepared in step 1 using the DNA ligation kit 2.1 (TAKARA). Plasmids were recovered from *Escherichia coli* DH5α (DE3) cells using a Qiagen miniprep kit and subjected to digestion by *Nco*I and *Hind*III. (Step 3) The sequence QLSSQLAELSEEALGDA-(GA)$_9$KL was produced with pBMS-PatE2K as a template and the primer pair 5′-CTTCATATGAGCCATCATCAC-3′ and 5′-CTTCTTAAGCTTGGCGCCTGCGCCAGCACC-GGCACCGGCGCCGGCGCCTGCACCCGC-3′ using the KOD hot start DNA polymerase kit (Merck) following the manufacturer's instructions. The PCR product was subjected to the endonucleases *Nco*I and *Hind*III and ligated with the vector prepared in step 2 using the DNA ligation kit 2.1 (TAKARA).

Heterocyclases and their fusion variants were expressed and purified following an established protocol.[7] The "full length" substrate peptides PatE2K and PatE3KK (Scheme S4) were expressed with a C-terminal, noncleavable his-tag and purified as previously described.[19]

Alkylation of Peptides. PatE2K/PatE3KK (100 μM) and ITACITFCAYD/ITACITFCAYDG (100 μM) were reacted with MicD (5 μM) and MicD Q219GA (5 μM), respectively, for 3 h at 25 °C in 20 mM Tris pH 8.0 supplemented with 150 mM NaCl, 10 $MgCl_2$, 10 mM ATP, and 1 mM DTT. The reactions were terminated by the addition of EDTA to the final concentration of 50 mM. Five mM alkylating agent iodoacetamide (IAA) was added to an aliquot of each, which was kept at room temperature away from light for 30 min. Excess IAA was then quenched by the addition of 10 mM DTT. Samples were subjected to MALDI-MS for analysis of heterocyclization and alkylation. Doubly heterocylized Pa-tE3KK and ITACITFCAYDG were further analyzed by ESI-MS/MS; PatE3KK was treated with trypsin before analysis.

Radioactive Assay. ATP [α-^{32}P] was purchased from PerkinElmer. PEI cellulose plates (Macherey-Nagel) were prerinsed with water and used after drying. All experiments were performed in duplicate. Assays were performed in 100 mM Tris pH 8.0, 50 mM NaCl, 5 mM DTT (MicD) or 1 mM TCEP (LynD and LynD fusion), and 10 mM $MgCl_2$ at 25 °C. For MicD, assays were performed with or without 150 μM E2K and 20 μM enzyme; for LynD and LynD fusion, assays were performed with 100 μM E2K and 5 μM enzyme. All reactions and controls contained 250 nM cold ATP and 50 μCi ATP [α-^{32}P]. At desired time points, 6 μL aliquots from the reaction were removed and mixed with 1 μL of 500 mM EDTA (71 mM after quenching), and 1 μL of the quenched time point mixtures was directly spotted on a TLC plate. For MicD, the reaction was quenched at 15 s, 30 s, 1, 2, 5, 10, and

30 min. For LynD and LynD fusion, the reaction was quenched after 1 h. TLC plates ran in 0.9 M guanidinium HCl and were dried prior to exposure using a BA phosphor screen (GE Healthcare). The phosphor screen was imaged using a Typhoon FLA7000 (GE Healthcare), spots quantified using ImageJ, followed by data analysis using Prism. Reactions to be used as positive controls were performed by adding 5 μM hexokinase + 0.5 mM glucose (for ADP production) or 5 μM Aapyrase (for AMP production) to the same reaction mixture used for the reactions with LynD or MicD. Negative controls containing just ATP and each enzyme + ATP in the absence of E2K were performed. Reactions were carried out simultaneously with positive and negative controls and ran on the same day and same TLC plates using identical conditions. Spots on the TLC plate were used to quantify nucleotides using the relationship 250 nM = [ATP] + [ADP] + [AMP].

EnzCheck (Pyro)phosphophosphate Assay. The EnzCheck Pyrophosphate Assay kit was purchased from Thermo Fisher. Each reaction contains the following components: 100 mM Tris pH 8.0, 50 mM NaCl, 10 mM MgCl$_2$, 5 mM DTT, 200 μM 2-amino-6-mercapto-7-methylpurine riboside (MESG), 1 U purine nucleoside phosphorylase (PNP), and when indicated 10 μM *Mycobacterium tuberculosis* pyrophosphatase (*Mt*PPase, a kind gift from Dr. Luiz Pedro de Carvalho of the Francis Crick Institute). Concentrations of enzymes, peptides, and nucleotides were varied in individual reactions. *Mt*PPase was purified following an established protocol.[20] Assays were carried out in 100 μL or 200 μL reaction volumes in 96-well plates, and the absorbance at 360 nm was monitored by a SpectraMax plate reader (Molecular Devices).

High-Performance Liquid Chromatography. Heterocyclization of ITACITFCAYD by MicD fusion in the presence of ATP, ADP, or AMP-CPP was carried out in HPLC assay buffer (100 mM Tris pH 8.0, 50 mM NaCl, 10 mM MgCl$_2$, 5 mM TCEP). The reactant concentrations, reaction time, and temperature varied and are individually specified. Reactions were terminated by the addition of two volumes of urea quench buffer (8 M urea, 20 mM Tris pH 8.0, 500 mM NaCl, 10 μM L-tryptophan (L-tryptophan added as an internal standard for MS analyses)) and then incubated with 120 μM Ni-NTA agarose resin (ABT) in a Corning centrifuge tube filter for 30 min at room temperature. After centrifugation at 4000g for 10 min, the flow through was collected and applied to an EC 250/4.6 NUCLEODUR 300-5 C18 column (Thames-Restek) connected to a G6130B Single Quad LC-MS instrument (Agilent Technologies). The column was pre-equilibrated in solvent A (5 mM Ammonium Bicarbonate pH 7.0), and after sample application, a stepwise isocratic program was run at a flow rate of 1 mL/min for a total of 11 min to separate the reactants and products: 80% solvent A, 20% solvent B (95% acetonitrile) for 3 min, followed by 50% A, 50% B for 3 min, and finally 100% B for 5 min. Separation of linear and heterocyclized peptides was not achieved by this method, and therefore, different selected ion monitoring (SIM) channels were set up to give separate ion chromatograms of these peptides. The following three SIM channels (negative mode) were set up: (1) AMP or AMP-CP and ITACITFCAYD, (2) ATP or AMP-CPP and ITACITFCAYD-1het, (3) ADP and L-tryptophan (internal standard). Ion intensity peaks were integrated using the Agilent ChemStation software and corrected against the area of the L-tryptophan mass peak. Standard curves for the quantitation of AMP, ADP, ATP, AMP-CP, AMP-CPP, and ITACITFCAYD were

obtained by diluting each compound to various concentrations in HPLC assay buffer and adding two volumes of urea quench buffer, before applying the samples to the HPLC-MS system.

For determining the degree of heterocyclization, reactions were set up between 25, 50, and 100 μM ITACITFCAYD and 1 mM AMP-CPP, catalyzed by MicD Q219GA (5 μM). The reactions were incubated at room temperature for 16 h and quenched with 2× volumes of urea quench buffer. The enzyme was removed using a filtration device, and the flow through was subjected to HPLC-MS as described above, except that a different solvent gradient was run: 20%−60% B from 0−10 min, 95% B from 10−15 min. Masses corresponding to the unmodified, singly and doubly heterocyclized peptides were entered into mass selective detectors (MSDs) 2, 3, and 4, respectively.

Duplicated time course experiments were carried out with MicD Q219GA (20 μM) with ITACITFCAYD (50 μM) and ATP or AMP-CPP (10 μM). The reaction containing ATP was performed at 25 °C, with samples withdrawn from the reaction mix at 15, 30, 60, 120, 300, and 600 s after the start of the reaction, and quenched by being mixed with two volumes of urea quench buffer. The AMP-CPP reaction was carried out in separate aliquots and incubated at 30 °C; reactions were quenched after 15, 30, 60, 120, and 240 min after initiation. Areas of ion intensity peaks (peptides, nucleotides, and L-tryptophan) were integrated and plotted, as a ratio of peptide or nucleotide over tryptophan, against time. Standard curves for ATP, ADP, AMP, and AMP-CPP were employed to calculate the concentrations of these nucleotides. Due to difficulties in obtaining a pure, singly heterocyclized peptide, the concentration of this compound was obtained from normalized data. The AMP-CPP reaction was assumed to have reached completion by the end of the incubation period, and the ion intensity readings of the 1het species at time 0 and 240 min (adjusted against tryptophan), respectively, were used as a minimum (0%) and maximum (100% or 10 μM), and a linear relationship between area ratio and peptide concentration existed within this range. The same slope (area ratio/concentration) was then used to normalize data from the ATP reaction. AMP-CP concentrations were estimated based on normalized data, with the peak ratio value at time 0 set as 0 and the peak ratio value at 240 min as 100% (10 μM), assuming linearity within this range.

Nuclear Magnetic Resonance Spectroscopy. Two sets of NMR experiments were performed: (set 1) the reaction between 2 mM uniformly labeled (u)-^{13}C,^{15}N-PatE2K and MicD, for sequential assignment as well as reaction monitoring; (set 2) the reaction between 100 μM u-^{15}N PatE2K-2het and MicD. Both sets of experiments were performed at 20 °C on a Bruker Ascend 700 MHz spectrometer equipped with a Prodigy TCI probe. The instrument was controlled by Topspin (Bruker).

For set 1, u-^{13}C,^{15}N-PatE2K was concentrated to 2 mM and exchanged to a buffer containing 50 mM HEPES pH 7.4 supplemented with 150 mM NaCl, 15 mM ATP, 10 mM MgCl$_2$, 5 mM DTT, 0.02% NaN$_3$, and 5% D$_2$O. ^{1}H,^{15}N-HSQC (heteronuclear single-quantum coherence spectroscopy) spectra using a standard Bruker pulse sequence incorporating water flip-back and PEP water suppression at 2028 × 128 points and a digital resolution of 12.3 and 31.0 Hz for the ^{1}H and ^{15}N dimensions, respectively, were recorded after 15 min, 30 min, 1 h, and each hour afterward until 66 h after the start of the reaction. HNCACB and CBCA(CO)NH spectra were

recorded using standard Bruker pulse sequences with $1536 \times 50 \times 104$ points and digital resolutions of 12.8, 59.6, and 257 Hz for the ^1H, ^{15}N, and ^{13}C dimensions, respectively, for the starting material and the product after 66 h. All spectra were processed with NMRPipe[21] and analyzed with CCPN Analysis 2.[22] All backbone amides, $C\alpha$ and $C\beta$ resonances of PatE2K, and product (after a 66 h incubation) were assigned, with the exception of the two N-terminal residues and the His$_6$-tag at the C-terminus.

For set 2, 100 μM uniformly labeled ^{15}N-PatE2K-2het (prepared as previously described)[19] was reacted with 5 μM MicD at 20 °C and the reaction was monitored by recording $^1H,^{15}N$-HSQC spectra every 30 min for a total of 16 h. Four transients were recorded at 2048×110 points with a spectral resolution of 12.3 and 36.1 Hz for the ^1H and ^{15}N dimensions, respectively. Spectral data were processed with Bruker Topspin and analyzed with CCPN Analysis 2.[22]

Mass Spectrometry. Molecular masses were determined using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) on a 4800 MALDI TOF/TOF analyzer (ABSciex).

Tandem mass spectrometry (MSMS) data of peptides were collected on an AB SCIEX Triple TOF 5600+ system equipped with an Eksigent nanoLC. Samples were first desalted by entering into a 5 μL/min flow of 98% H$_2$O, 2% acetonitrile (ACN), and 0.05% trifluoroacetic acid (TFA) and washed through a Thermo Pepmap 20 mm × 0.075 mm column (trap column) for 5 min. The trap column was then connected to an analytical column (150 mm × 0.075 mm), and both were equilibrated in 98% H$_2$O, 2% ACN, and 0.1% formic acid (FA). Peptides were eluted and separated by switching the solvent toward 98% ACN, 2% H$_2$O, and 0.1% FA in a linear gradient over the course of 6 min, with a flow rate of 300 nL/min, and the eluate was sprayed into the mass spectrometer. The most intense peaks within the 400−1250 m/z region were automatically selected in the information dependent acquisition (IDA) mode and directed to collision induced fragmentation (CID) followed by and ions collected in the 95−1800 m/z range. Alternatively, product ion scan (PIS) mode was employed where specified masses were entered into the program prior to the experiment so that they are selected for fragmentation. Sample application and data collection were performed by the University of St Andrews mass spectrometry facility.

■ RESULTS

Order of Heterocyclization. Previous work using TruD (heterocyclase from the trunkamides biosynthetic pathway) and the PatE2K substrate showed that heterocyclization of cysteine 51 precedes that of cysteine 47 (C8 and C4 of the core peptide, respectively).[8] PatD catalyzes both oxazoline and thiazoline formation and with PatE2K yields four heterocycles formed by two cysteine and two threonine residues. However, PatD has proven to be very difficult to keep stable in solution in our hands. PatE2K was therefore reacted with MicD (heterocyclase from the microcyclamides biosynthetic pathway), which can also modify both cysteines and serines/threonines. The reaction was analyzed by MALDI-MS without (Figure 1A) and with (Figure 1A,B) the sample being treated with iodoacetamide (IAA), an alkylating agent that covalently modifies cysteine residues resulting in the addition of 57 Da mass per free cysteine (Figure 1A,B). The experiment shows no evidence of the IAA modified peptide with two or more
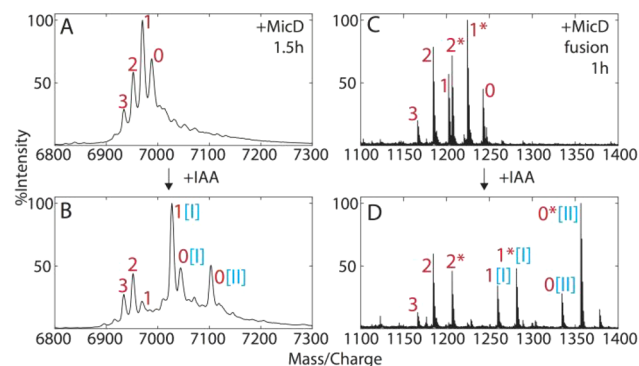

Figure 1. Alkylation of free cysteines on precursor peptides. PatE2K (A, B) and ITACITFCAYD (C, D) were reacted with MicD and MicD fusion, respectively, to a state with a mixed number of heterocycles. The samples were then reacted with the alkylating agent iodoacetamide to detect the presence of free cysteines. The numbers of dehydrations in the product in each m/z species are annotated in red. Asterisks signify $[M + Na]^+$ species, and the lack of an asterisk indicates $[M + H]^+$ species; [I] denotes an adduct of +57 Da (mono alkylation), while [II] denotes two alkylations (+114 Da). The same analysis, with essentially identical results, was carried out with PatE3KK (Figure S1), which has the same ITACITFC core sequence (Scheme S4).

heterocycles. In contrast, there is clear evidence for a double adduct of the peptide with no heterocycles and a single adduct of a peptide with a single heterocycle. Further, MSMS analysis of the doubly dehydrated species located the heterocycles at (Figures S1 and S2) positions C4 and C8 of the core peptide. Thus, we conclude that both cysteines are heterocyclized before either threonine.

To establish the order of heterocyclization, uniformly ^{15}N-labeled PatE2K was reacted with MicD, and the $^1H,^{15}N$-HSQC spectra were recorded. Interpretation of the HSQCs was based on a backbone assignment of uniformly $^{13}C,^{15}N$-labeled PatE2K before incubation with MicD (Figure S3, Table S1), and after 66 h of incubation, which under the conditions chosen yielded a mixture of peptides containing 1−4 heterocycles (Figure S4B, Figure S5). In the triple-resonance spectra, up to three different states were observed for residues affected by heterocyclization (Figure S5). Distinct changes in $C\alpha$ and $C\beta$ resonances reflect heterocyclization of C51, C47, T45, and T49. As observed previously,[8] cysteine $C\alpha$ and $C\beta$ resonances undergo dramatic downfield shift changes when the heterocycles form (strips for I48 and A52 in Figure S5). A similar effect has now been observed for threonine heterocyclization (strips for A46 and F50 in Figure S5). In addition, $C\alpha$ resonances of I44 and I48 have undergone upfield chemical shifts in one of their observed states, reflecting heterocyclization of T45 and T49 (Figure S5). These characteristic changes greatly facilitated interpretation of the time-dependent changes observed in the HSQCs. C51, judged by shifts in cross-peaks of I48, T49, F50, A52, Y53, D54, and G55, reacts first, consistent with studies of TruD (Figure S4A).[8] After the first modification, the situation was more complex and harder to interpret, with multiple shifts in cross-peaks (Figure S4B) being observed at similar rates. To simplify the interpretation of HSQC spectra, uniformly ^{15}N-labeled PatE2K was reacted with LynD to heterocyclize only the cysteines. This new peptide was used as a substrate for the MicD reaction. MicD modified both T45 and T49 at the same time (Figure S4C,D), based on shifts in cross-peaks S42, K43,

I44, A46, I48, F50, and A52, thus lacking a definite order for the third heterocyclization; however, T45 was depleted at a higher rate, suggesting a partial order (Figure S4C,D). A preference for C51 over C47, and for T45 over T49 is also seen by tandem mass spectrometry (MSMS) analyses (Figure 2A,B; Figures S6 and S7). The triply heterocyclized peptide
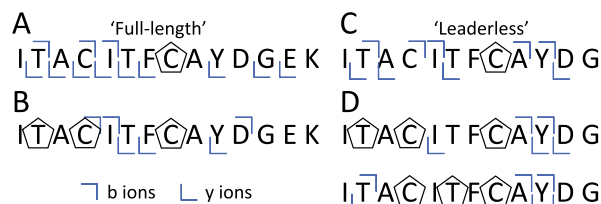

Figure 2. Tandem MS analyses of peptides modified by MicD or MicD fusion. Peptides with a variable number of heterocycles were subjected to MSMS to reveal the location of heterocyclization. (A) singly and (B) triply dehydrated ITACITFCAYDGEK (produced from PatE3KK) and (C) singly and (D) triply dehydrated ITACITFCAYDG. Pentagons signify the modification of residues into heterocycles; left- and right-facing right angles represent b and y ions, respectively. Full spectra and calculated masses are shown in Figures S6, S7, S9, and S10. "Full-length" and "leaderless" refer to the peptide substrate undergoing heterocyclization.

showed strong evidence for azolines at T45, C47, and C51 but not for T49, C47, and C51 (Figure S7). However, ionization propensity may account for the lack of the T49-containing species.

We explored the role of the leader peptide in determining the order of heterocyclization by using the fused enzyme MicD Q219GA (analogous to the previously reported LynD fusion[7]) and ITACITFCAYDG (a leaderless substrate). IAA labeling shows that cysteines are once again heterocyclized before threonines (Figure 1C,D, Figures S1 and S8). MS/MS analysis (Figure 2C, Figure S9) indicates that the reaction order of the cysteines is retained, as the fragmentation pattern of the singly dehydrated species corresponds to the presence of a heterocycle at C8 (equivalent to C51 of PatE2K) rather than C4 (equivalent to C47 of PatE2K). In contrast, any order of threonine heterocyclization appears lost, as T2 (T45 of PatE2K) and T6 (T49 of PatE2K) heterocycles were both found in the triply dehydrated species (Figure 2D, Figure S10).

Assay under Single Turnover Conditions. To identify if ATP consumption was leading to AMP and/or ADP formation in the first heterocyclization event, we performed experiments in the presence and absence of PatE2K using ATP [α-^{32}P]. Our experiments (Figure 3) were conducted under single turnover conditions and show that, both in the presence and absence of PatE2K, ADP is formed as a reaction intermediate, but with distinct rates of formation and decay. In the absence of PatE2K, ADP forms quickly (rate of formation (0.044 s^{-1})), being completely converted to AMP in 500 s (rate of ADP decay is 0.005 s^{-1}). In contrast, in the presence of PatE2K, ADP forms at a slightly slower rate (0.033 s^{-1}) but takes much longer to be converted into AMP (rate of ADP decay is 0.00013 s^{-1}). AMP is formed quickly in the absence of PatE2K (rate of formation 0.006 s^{-1}) and slower in the presence of PatE2K (0.0014 s^{-1}). In the absence of PatE2K, MicD is likely catalyzing solely ATP hydrolysis to ADP and then to AMP, while in the presence of PatE2K, ATP consumption is coupled to heterocyclization, and hence the slower rates observed for ADP decrease. The reaction with LynD and LynD fusion also
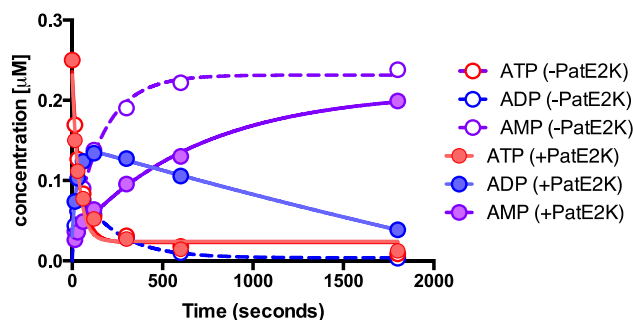

Figure 3. Rate of nucleotide usage under single turnover conditions. Concentrations of ATP, ADP, and AMP following incubation of ATP [α-^{32}P] with MicD, in the presence or absence of 100 μM PatE2K. The concentrations of ATP (red), ADP (blue), and AMP (purple) at each time point are shown as circles that are empty or filled to indicate whether PatE2K (substrate) is absent or present in the reaction, respectively. Lines are fits to a single (ATP and AMP) or double (ADP) exponential equation with the general format $y(t) = \Sigma_i A_i e^{-k_i t} + y_0$ where $y(t)$ is the observed signal at time t, i is the number of transients, A_i is the amplitude of the ith transient, k_i is the observed rate constant for the ith transient, and y_0 is the y-axis offset.

showed ADP formation at similar levels for both enzymes in agreement with the results obtained for MicD (Figure S11).

Chemical Analogues of ATP. As a positive control, PatE2K was reacted with ATP for 16 h and up to four dehydrations were observed by MS (Figure 4A). AMP-CPP
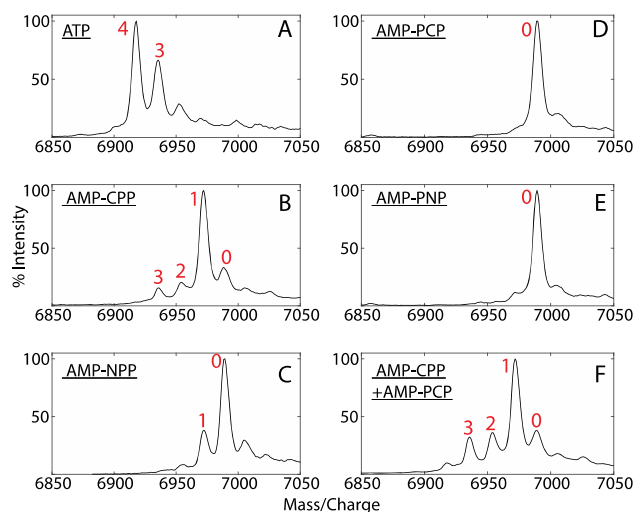

Figure 4. Nucleotide usage by the heterocyclase. MALDI-MS spectra of PatE2K (100 μM) reacted with 2 mM each of (A) ATP, (B) AMP-CPP, (C) AMP-NPP, (D) AMP-PCP, (E) AMP-PNP, and (F) a mix of 2 mM AMP-CPP and 2 mM AMP-PCP. The reactions were catalyzed by MicD (5 μM). SM, starting material. The number of dehydrations that are present in each m/z species is annotated in red.

and AMP-NPP, both of which contain a hydrolyzable $\beta-\gamma$ and a nonhydrolyzable $\alpha-\beta$ phosphate bond, supported multiple heterocyclization of the test substrate peptide PatE2K by the heterocyclase MicD (Figure 4B,C). In contrast, AMP-PCP and AMP-PNP, which have a nonhydrolyzable $\beta-\gamma$ and hydrolyzable $\alpha-\beta$ phosphate bond (Figure 4D,E), did not support catalysis. The analysis of catalysis with AMP-CPP and AMP-PCP was repeated with heterocyclases LynD and OscD and gave the same results: AMP-CPP supported catalysis, while AMP-PCP did not (Figure S12). Mixing AMP-PCP and AMP-

CPP gave the same result as AMP-CPP (Figure 4F). HPLC analysis of the ATP and AMP-CPP reaction (Figure S15) was carried out with the fused enzyme MicD Q219GA as a catalyst with a leaderless peptide, ITACITFCAYD, at various time points. This reaction is slower than for the native enzyme PatE substrate combination, which allowed us to monitor ITACITFCAYD depletion and nucleotide decomposition more easily. A species with a mass corresponding to singly heterocyclized peptide appeared as a substrate was consumed. When ATP is included in the reaction, ADP was observed to initially accumulate before decomposing to AMP (Figure S4C), while AMP-CPP gave only AMP-CP (Figure S4D).

We thus concluded that catalysis (only) requires the breakage of the $\beta-\gamma$ phosphate bond, which is the kinase type mechanism. We do note that the presence of the nonhydrolyzable $\alpha-\beta$ phosphate bond does slow the enzyme very significantly when compared to ATP (Figure 4A–C).

Pi and PPi Production. If AMP was produced solely from ADP hydrolysis, then there should be no PP_i production, only P_i. The amount of P_i released from the heterocyclization reaction was measured using a coupled assay for both LynD and MicD enzymes (Figure 2). In both cases, a significant increase in Pi was observed when the enzyme pyrophosphatase was included, indicating the presence of PPi (Figure 5A,B).
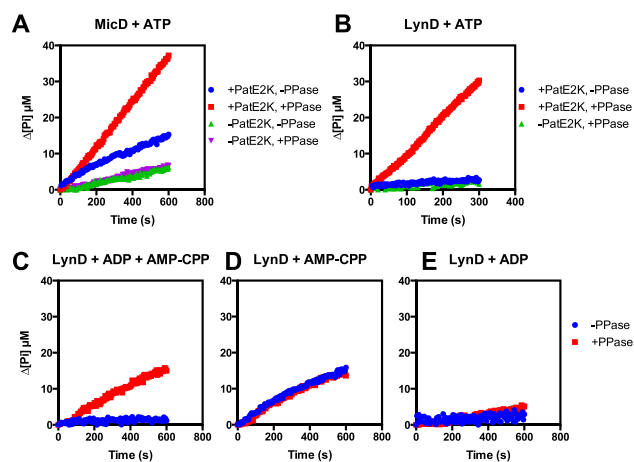

Figure 5. Production of PPi and Pi under different conditions. Level of phosphate produced following heterocyclase reactions using different nucleotides, in the presence and in the absence of pyrophosphatase enzyme (PPase). (A) 5 μM MicD reacted with 10 mM ATP and where indicated 20 μM PatE2K. (B) 5 μM LynD reacted with 1 mM ATP and where indicated, 100 μM PatE2K. (C–E) 5 μM LynD incubated with 100 μM PatE2K and a mix of 10 mM AMP-CPP and 10 mM ADP. The reaction was repeated with 10 mM AMP-CPP or ADP on their own. Reactions were carried out in duplicates.

The presence of PP_i in these reactions was dependent on the presence of peptide and so is not the result of some side reaction of ATP in solution (Figure 5A,B). The production of PP_i during catalysis is consistent with previous ^{31}P NMR experiments[8] and the previous report of [18]O incorporation into PP_i.[7] The *E. coli* YcaO protein was observed to produce AMP and PP_i in the absence of a substrate, but since the substrate is unknown, the relative rates cannot be estimated for YcaO.[18] Interestingly, the result for MicD is different to that for LynD. Ignoring the degradation of PP_i in the time of reaction and assuming all of the increase in P_i from the addition of PPase

comes from PP_i produced by the enzyme during turnover, it is calculated that MicD produces P_i/PP_i in the ratio 1:1 whereas LynD has a ratio of 1:5. Notably, PP_i was observed when ADP was added to a reaction mix containing LynD, AMP-CPP, and PatE2K (Figure 5C). PPi was not when either nucleotide was present on its own (Figure 5D,E). The rate of reaction with AMP-CPP (0.006 s^{-1}) was not, within the error of our measurements, affected by the presence of ADP (Figure 5C vs D).

■ DISCUSSION

We have shown that MicD, a heterocyclase capable of processing both cysteines and threonines, heterocyclizes two cysteines in the substrate peptide with the same order (C-to N-terminus) observed previously for TruD and BalhCD.[8,17] In a peptide with two cysteines and two threonines, the cysteine residues are hetereocyclized before either of the two threonine residues (Figures 1 and 2). The data suggest that for the threonines an N-to-C order in the heterocyclization is preferred but not obligatory, in contrast to the situation with cysteines where the order seems immutable (Scheme 1). We

Scheme 1. Proposed Order of Heterocyclization by MicD[a]
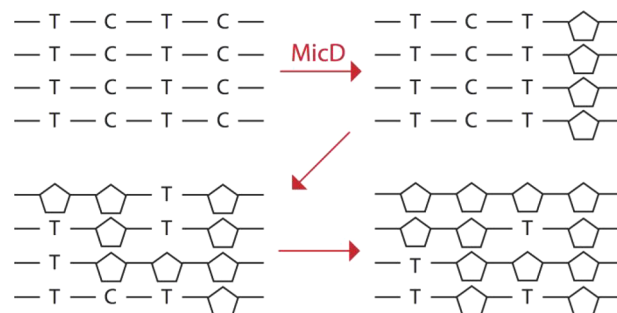
[a]T and C represent heterocyclizable residues, whereas five-membered rings represent azolines.

explored whether the substrate leader influences the order of catalysis by using the fused variant of the MicD enzyme. The fused enzyme using a leaderless substrate showed cysteines were processed before threonines (consistent with the relative nucleophilicity). The data did indicate that the C-to-N order of the cysteine heterocyclization was conserved, but any preference for the order of threonines was lost (Figure 2). The fact that the leader was only one component contributing to the ordering of heterocyclization echoes the observations made of the lantibiotic synthetase LctM, where the leader plays a partial role in determining the reaction order.[23] A study of cyanobactin biosynthetic dehydrogenase enzyme ArtGox also showed that the order of the reaction (conversion of thiazolines to thiazoles) was unaffected by the presence or absence of the leader peptide.[19]

The seemingly contradictory observations around ATP usage by the heterocyclase class of enzymes has been discussed.[5,7,8,14,18,24] We confirm multiple previous studies that a cyanobactin heterocyclase produces AMP and PP_i during catalysis. At the same time, a detailed dissection of the nucleotide usage using nonhydrolyzable analogues establishes that catalysis requires only the cleavage of the $\beta-\gamma$ phosphate bond (Figure 4). The mechanism is a kinase type consistent with earlier work,[14] but the production of PP_i is unexplained by a simple kinase mechanism. Since the

formation of PP$_i$ is not essential for catalysis, we considered it might arise from a second, off pathway reaction. An estimate of the ratio of PP$_i$ to P$_i$ production during catalysis showed different values for different homologues (MicD, LynD) and different conditions, consistent with the concept of an off path reaction. We suspect that, depending on the specific YcaO type enzyme and conditions employed, PPi production may vary from negligible to dominant.

In rationalizing the production of PPi, we note especially the production of PP$_i$ when ADP was added to an enzyme AMP-CPP reaction (Figure 5C). AMP-CPP breaks down to AMP-CP and P$_i$ (Figure 5D); AMP-CP cannot undergo further reaction, and ADP does not support catalysis or produce PPi (Figure 5E). We propose that the most plausible source of PP$_i$ is from an enzyme-catalyzed reaction between P$_i$ (originating from AMP-CPP) and ADP to generate AMP and PP$_i$, essentially a disproportionation or transphosphorylation reaction (Scheme 2). Single turnover experiments also point

Scheme 2. Proposed Mechanism[a]
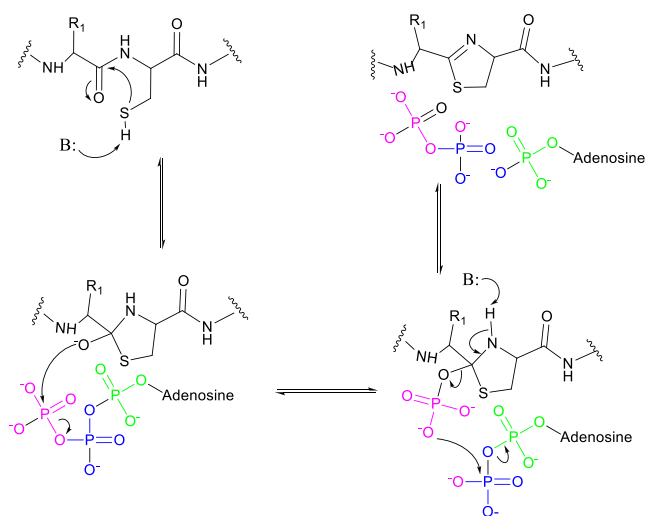
[a]A transphosphorylation reaction that occurs during thiazoline formation gives rise to PPi. Our data do not determine whether the phosphate bond to hemiorthoamide is broken before or during the reaction with ADP.

to ADP being an intermediate during catalysis (Figure 3). Our data do not determine whether the P$_i$ is chemically bound to the hemiorthoamide or not during the disproportionation reaction; although given the coupling to enzyme turnover, we favor the former. We speculate the extensive and unusual metal coordination of the nucleotide phosphate groups observed in YcaO domains[7,18] is responsible for this unusual chemistry. This mechanistic proposal reconciles the contradictions in the literature and divulges a new enzyme-catalyzed transphosphorylation.

ASSOCIATED CONTENT

ⓈSupporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.9b00084.

Kinase mechanism, bond breaking reactions during ATP hydrolysis, (MS)MS analysis, constructs, amino acid sequences, alkylation of free cysteines on precursor

peptides, sequential assignment of unreacted PatE2K, HSQC spectra, triple-resonance strips for PatE2K residues, radioactivity, heterocyclization of PatE2K, ion chromatograms, MALDI-MS spectra, HPLC, and chemical shifts (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: naismith@strubi.ox.ac.uk. Tel: +44-1334-463792. Fax: +44-1334-467229.

ORCID

Clarissa Melo Czekster: 0000-0002-7163-4057
James H. Naismith: 0000-0001-6744-5061

Present Address

⊥Departments of Ophthalmology, Microbiology and Immunology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114, USA.

Author Contributions

Y.G., C.C., U.S., and O.M. performed experiments. All authors analyzed data and wrote the paper. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

ADP, adenosine diphosphate; AMP, adenosine monophosphate; AMP-CP, α,β-methyleneadenosine diphosphate; AMP-CPP, α,β-methyleneadenosine triphosphate; AMP-NPP, adenosine 5′-(α,β-imido)triphosphate; AMP-PCP, β,γ-methyleneadenosine triphosphate; AMP-PNP, adenosine 5′-(β,γ-imido)triphosphate; ATP, adenosine triphosphate; HPLC, high-performance liquid chromatography; HSQC, heteronuclear single-quantum correlation spectroscopy; IAA, iodoacetamide; MALDI, matrix-assisted laser desorption/ionization; MESG, 2-amino-6-mercapto-7-methylpurine riboside; MS, mass spectrometry; NMR, nuclear magnetic resonance spectroscopy; P$_i$, inorganic phosphate; PNP, purine nucleoside phosphorylase; PPase, pyrophosphatase; PP$_i$, inorganic pyrophosphate; RiPPs, ribosomally synthesized and post-translationally modified peptides

REFERENCES

(1) 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., et al. (2013) Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature. *Nat. Prod. Rep. 30* (1), 108−160.

(2) Long, P. F., Dunlap, W. C., Battershill, C. N., and Jaspars, M. (2005) Shotgun Cloning and Heterologous Expression of the

Patellamide Gene Cluster as a Strategy to Achieving Sustained Metabolite Production. *ChemBioChem* 6 (10), 1760−1765.

(3) Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A., Haygood, M. G., and Ravel, J. (2005) Patellamide A and C Biosynthesis by a Microcin-like Pathway in Prochloron Didemni, the Cyanobacterial Symbiont of Lissoclinum Patella. *Proc. Natl. Acad. Sci. U. S. A.* 102 (20), 7315−7320.

(4) Koehnke, J., Bent, A. F., Houssen, W. E., Mann, G., Jaspars, M., and Naismith, J. H. (2014) The Structural Biology of Patellamide Biosynthesis. *Curr. Opin. Struct. Biol.* 29, 112−121.

(5) Czekster, C. M., Ge, Y., and Naismith, J. H. (2016) Mechanisms of Cyanobactin Biosynthesis. *Curr. Opin. Chem. Biol.* 35, 80−88.

(6) Agarwal, V., Pierce, E., McIntosh, J., Schmidt, E. W., and Nair, S. K. (2012) Structures of Cyanobactin Maturation Enzymes Define a Family of Transamidating Proteases. *Chem. Biol.* 19 (11), 1411−1422.

(7) Koehnke, J., Mann, G., Bent, A. F., Ludewig, H., Shirran, S., Botting, C., Lebl, T., Houssen, W. E., Jaspars, M., and Naismith, J. H. (2015) Structural Analysis of Leader Peptide Binding Enables Leader-Free Cyanobactin Processing. *Nat. Chem. Biol.* 11 (8), 558−563.

(8) Koehnke, J., Bent, A. F., Zollman, D., Smith, K., Houssen, W. E., Zhu, X., Mann, G., Lebl, T., Scharff, R., Shirran, S., et al. (2013) The Cyanobactin Heterocyclase Enzyme: A Processive Adenylase That Operates with a Defined Order of Reaction. *Angew. Chem., Int. Ed.* 52 (52), 13991−13996.

(9) Bent, A. F., Mann, G., Houssen, W. E., Mykhaylyk, V., Duman, R., Thomas, L., Jaspars, M., Wagner, A., and Naismith, J. H. (2016) Structure of the Cyanobactin Oxidase ThcOx from Cyanothece Sp. PCC 7425, the First Structure to Be Solved at Diamond Light Source Beamline I23 by Means of S-SAD. *Acta Crystallogr. D Struct Biol.* 72 (11), 1174−1180.

(10) Bent, A. F., Koehnke, J., Houssen, W. E., Smith, M. C. M., Jaspars, M., and Naismith, J. H. (2013) Structure of PatF from Prochloron Didemni. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* 69 (6), 618−623.

(11) Koehnke, J., Bent, A., Houssen, W. E., Zollman, D., Morawitz, F., Shirran, S., Vendome, J., Nneoyiegbe, A. F., Trembleau, L., Botting, C. H., et al. (2012) The Mechanism of Patellamide Macrocyclization Revealed by the Characterization of the PatG Macrocyclase Domain. *Nat. Struct. Mol. Biol.* 19 (8), 767−772.

(12) Mann, G., Koehnke, J., Bent, A. F., Graham, R., Houssen, W., Jaspars, M., Schwarz-Linek, U., and Naismith, J. H. (2014) The Structure of the Cyanobactin Domain of Unknown Function from PatG in the Patellamide Gene Cluster. *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* 70 (12), 1597−1603.

(13) Roy, R. S., Gehring, A. M., Milne, J. C., Belshaw, P. J., Walsh, C. T., et al. (1999) Thiazole and Oxazole Peptides: Biosynthesis and Molecular Machinery. *Nat. Prod. Rep.* 16 (2), 249−263.

(14) Dunbar, K. L., Melby, J. O., and Mitchell, D. A. (2012) YcaO Domains Use ATP to Activate Amide Backbones during Peptide Cyclodehydrations. *Nat. Chem. Biol.* 8 (6), 569−575.

(15) Burkhart, B. J., Hudson, G. A., Dunbar, K. L., and Mitchell, D. A. (2015) A Prevalent Peptide-Binding Domain Guides Ribosomal Natural Product Biosynthesis. *Nat. Chem. Biol.* 11, 564.

(16) Gu, W., Sardar, D., Pierce, E., and Schmidt, E. W. (2018) Roads to Rome: Role of Multiple Cassettes in Cyanobactin RiPP Biosynthesis. *J. Am. Chem. Soc.* 140 (47), 16213−16221.

(17) Melby, J. O., Dunbar, K. L., Trinh, N. Q., and Mitchell, D. A. (2012) Selectivity, Directionality, and Promiscuity in Peptide Processing from a Bacillus Sp. Al Hakam Cyclodehydratase. *J. Am. Chem. Soc.* 134 (11), 5309−5316.

(18) Dunbar, K. L., Chekan, J. R., Cox, C. L., Burkhart, B. J., Nair, S. K., and Mitchell, D. A. (2014) Discovery of a New ATP-Binding Motif Involved in Peptidic Azoline Biosynthesis. *Nat. Chem. Biol.* 10 (10), 823−829.

(19) Gao, S., Ge, Y., Bent, A. F., Schwarz-Linek, U., and Naismith, J. H. (2018) Oxidation of Cyanobactin Precursor Peptide Is Independent of Leader Peptide and Operates in a Defined Order. *Biochemistry* 57, 5996−6002.

(20) Pedreño, S., Pisco, J. P., Larrouy-Maumus, G., Kelly, G., and de Carvalho, L. P. S. (2012) Mechanism of Feedback Allosteric Inhibition of ATP Phosphoribosyltransferase. *Biochemistry* 51 (40), 8027−8038.

(21) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* 6 (3), 277−293.

(22) Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins: Struct., Funct., Genet.* 59 (4), 687−696.

(23) Thibodeaux, C. J., Wagoner, J., Yu, Y., and van der Donk, W. A. (2016) Leader Peptide Establishes Dehydration Order, Promotes Efficiency, and Ensures Fidelity During Lacticin 481 Biosynthesis. *J. Am. Chem. Soc.* 138 (20), 6436−6444.

(24) Burkhart, B. J., Schwalen, C. J., Mann, G., Naismith, J. H., and Mitchell, D. A. (2017) YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function. *Chem. Rev.* 117 (8), 5389−5456.