

Structural analysis of leader peptide binding enables leader-free cyanobactin processing

Jesko Koehnke^{#1}, Greg Mann^{#1}, Andrew F Bent^{#1}, Hannes Ludewig¹, Sally Shirran¹, Catherine Botting¹, Tomas Lebl¹, Wael Houssen^{3,4,5}, Marcel Jaspars³, and James H Naismith^{1,6,*}

¹BSRC, University of St Andrews, St Andrews, KY16 9RH

³Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Meston Walk, Aberdeen, AB24 3UE

⁴Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, Scotland, UK

⁵Pharmacognosy Department, Faculty of Pharmacy, Mansoura University, Mansoura 35116, Egypt

⁶State Key Laboratory of Biotherapy and Collaborative Innovation Center for Biotherapy, West China Hospital, Sichuan University, China

These authors contributed equally to this work.

Abstract

Regioselective modification of amino acids within the context of a peptide is common to a number of biosynthetic pathways and many such products have potential as therapeutics. The ATP dependent enzyme LynD heterocyclizes multiple cysteine residues to thiazolines within a peptide substrate. The enzyme requires the substrate to have conserved N-terminal leader for full activity. Catalysis is almost insensitive to immediately flanking residues in the substrate suggesting recognition occurs distant from the active site. Nucleotide and peptide substrate co-complex structures of LynD reveal the substrate leader peptide binds to and extends the β -sheet of a conserved domain of LynD, whilst catalysis is accomplished in another conserved domain. The spatial segregation of catalysis from recognition combines seemingly contradictory properties of regioselectivity and promiscuity; it appears to be a conserved strategy in other peptide modifying

*To whom correspondence should be addressed: James Naismith, Naismith@st-andrews.ac.uk.

Author contributions

J.K. carried out crystallization and x-ray crystallography experiments, interpreted data and wrote the paper. G.M. carried out protein expression and purification, heterocyclization reactions, ITC measurements and selective ¹⁸O labeling experiments, interpreted data and wrote the paper. A.F.B. carried out site directed mutagenesis, protein expression and purification, heterocyclization reactions, MALDI-TOF-MS data acquisition experiments, interpreted data and wrote the paper. H.L. carried out protein expression and purification, crystallization and x-ray crystallography experiments and interpreted data. S.S. carried out MALDI-TOF-MS, LCMS and MS-MS data acquisition experiments and interpreted data. C.B. carried out MALDI-TOF-MS, LCMS and MS-MS data acquisition experiments and interpreted data. T.L. carried out ³¹P NMR experiments and interpreted data. W.H. interpreted data and wrote the paper. M.J. interpreted data and wrote the paper. J.H.N. carried out x-ray crystallography experiments, interpreted data and wrote the paper.

Competing financial interests

The Universities of St Andrews and Aberdeen have filed a patent application with the UK Patent Office on the use of modified LynD enzyme. J.K., A.F.B., G.M., W.E.H., M.J. and J.H.N. are among those listed as inventors.

enzymes. A variant of LynD that efficiently processes substrates without a leader peptide has been engineered.

Introduction

There has been a surge in interest in the ocean as a source of new therapeutics¹. This has in part been stimulated by high profile successes and by belief that the less well-explored marine environment contains many more unexploited resources^{2,3}. Ribosomally synthesized and post-translationally modified peptides (RiPPs) produced by marine organisms have been shown to possess anti-tumour, anti fungal, antibacterial and antiviral properties⁴. Cyanobactins, cyclic peptide natural products from cyanobacteria, are RiPPs in which one or more core peptides (it is the core peptide which becomes a natural product) are embedded into a larger precursor peptide. The most well-known example of this class are the patellamides, whose biosynthetic pathway was one of the first cyanobactin pathways to be described and cloned⁵⁻⁷. The precursor peptide has a conserved N-terminal leader, typically around 40 residues, which is disposed of during maturation⁸. Characterized modifications of the core peptide include heterocyclization (also known as cyclodehydration) of Cys (and in some cases Ser/Thr) residues to oxazolines (and thiazolines) (Figure 1A), epimerization of amino acids to give *D*-stereocenters, oxidation of these heterocycles to oxazoles and thiazoles, Ser/Thr/Tyr prenylation and macrocycle formation^{5,9,10}. The enzymes for each natural product are named XxxY; where Xxx is three letter abbreviation of the natural product (Pat, Lyn, Tru) and Y is a single letter that denotes the enzyme, A being the N-terminal protease (removes leader peptide), G the macrocyclase / oxidase, F the prenylase and D the cyclodehydratase⁸. The permissiveness of the modifying enzymes to sequence changes in the core peptide has been elegantly demonstrated by the creation of large libraries of novel macrocycles made *in vivo* by genetic engineering^{6,11}. An alternative combined *in vivo* and *in vitro* system for production of patellamides in larger quantities has also been reported¹². In the patellamide pathway the protease PatA, the cyclodehydratase PatD and macrocyclase PatG all recognize residues outside the core peptide to accomplish their transformations¹³⁻¹⁷. The recognition beyond the functional group being modified that governs the prenylase^{18,19}, oxidase²⁰ and hypothetical epimerase remain unknown.

Heterocyclization of core peptide Cys (and sometimes Ser/Thr) residues to thiazolines (and oxazolines) by a highly conserved class of ATP and Mg²⁺-dependent YcaO-domain containing cyclodehydratases (the D enzymes PatD, TruD, LynD etc) is presumed to be first chemical transformation in patellamide-like products^{21,22}. The site-selective introduction of heterocycles into peptide backbones alters both the conformation and reactivity of peptides; this tailoring of peptides is highly desirable in modifying their biological properties²³. It has long been known that the N-terminal leader of substrate peptides was key for processing by TruD / PatD^{21,22} but no molecular insight has been forthcoming. The apo structure of the cyanobactin heterocyclase TruD showed this enzyme to be a three-domain protein¹⁶. The first two domains share structural but limited sequence homology with MccB (an adenylating enzyme from the microcin pathway)²⁴ and the third domain (the 'YcaO' domain) had, at that time, no homology to known structures¹⁶. Interestingly the putative ATP binding site in TruD (based on homology with MccB and other adenylating enzymes)

has two loop deletions that meant if ATP bound this domain it would do so differently than in other adenylating enzymes¹⁶. The direct observation of pyrophosphate and AMP production during enzyme turnover, coupled with the homology between TruD and MccB, led us to propose an adenylating mechanism for the enzyme¹⁶, in contrast to the kinase mechanism proposed for BalhD^{25,26}(Figure 1B), a cyclodehydratase involved in thiazole/oxazole-modified microcins (TOMM) biosynthesis - another product of the growing family of RiPPs. BalhD is a YcaO domain and is homologous in sequence to the third domain of the D enzymes. Analysis of both the BalhD and TruD has shown they operate with a preferred order (start at C-terminus)^{16,27}. A substrate recognition motif within the leader peptide of PatE, the precursor peptide, has been identified (denoted 'minimal leader')¹⁶ and this motif is conserved in all related precursor peptides (TruE, LynE etc)²⁸. TruD was shown to process the C-terminal cysteine of test peptides which lacked the leader, albeit more slowly, but TruD was not, within the timescale of the experiment, able to process a second 'internal' cysteine¹⁶. A recent study reported that *trans* activation of the PatD enzyme by exogenous leader peptide restored processing activity for internal cysteines²⁹.

The structure of an *E. coli* YcaO domain bound to nucleotide has been reported (25 % sequence identity to TruD)³⁰. The residues that coordinate the AMP-PNP analog in this *E. coli* YcaO domain are largely conserved in cyanobactin cyclodehydratases pointing to a unified mechanism for all such YcaO domain proteins. In the absence of substrate *E. coli* YcaO hydrolyzed ATP to AMP and pyrophosphate, echoing the finding from TruD¹⁶.

Here, we report three crystal structures of the cyanobactin heterocyclase LynD (highly homologous to PatD and TruD) from the aestuaramide pathway (*Lynngbya sp.*). We used the previously described modified PatE¹⁶ (denoted PatE') for co-complex studies as it is a substrate of LynD. LynE (the authentic LynD substrate) has an identical minimal leader sequence²⁸ to that identified in PatE¹⁶. Structures are reported for LynD/AMP/PatE', LynD/ADP PO₄³⁻/PatE'C51A and LynD/ β,γ -imido-ATP/PatE'C51A co-complexes. The structures confirm that LynD has a nucleotide-binding site in its third (YcaO) domain of LynD, similar to that described for the *E. coli* YcaO structure³⁰. The molecular basis of leader recognition by LynD has been elucidated, rationalizing both *cis* / *trans* activation of PatD and the molecular basis of the 'minimal leader'. A novel LynD (FUSED LynD) has been engineered, which no longer requires leader on the substrate peptide and this has potential in biotechnology and chemical synthesis.

Results

Structure of LynD

Full-length LynD heterocyclase was expressed, purified and its ability to perform heterocyclization reactions with a precursor peptide PatE' in the presence of ATP/Mg²⁺ (Supplementary Results, Supplementary Table 1) confirmed as described previously¹⁶. PatE' is 64 amino acids long, has a full leader but possesses a single core peptide (ITACITFC) and C-terminal His₆-tag. The retention time of LynD in gel filtration suggested LynD was a dimer and diffraction quality crystals were only obtained when incubated with a nucleotide and PatE'. The same orthorhombic form (but different unit cells) with two monomers in the asymmetric unit was obtained for all three complexes (LynD/AMP/PatE' 2.86 Å,

LynD/ATP/PatE/C51A 2.14 Å, LynD/ β,γ -imido-ATP/PatE/C51A 3.01 Å). The two protomers are composed of the same three domains and form the same antiparallel dimer by head-to-tail association of domains 1 and 2 with a large interface (buried surface area of ~ 3280 Å²) first seen for TruD¹⁶ (Figure 2A). The structure of LynD/AMP/PatE' was determined by molecular replacement using TruD (PDB 4BS9) as the search model and partly refined. This structure was used as a search model for the 2.14 Å data, a complete model was built and refined with these data and then molecular replaced into the lower resolution data. Full structural statistics are given in Supplementary Table 2. PatE' binds to LynD in an identical fashion in all three structures and as the overall C $_{\alpha}$ rmsd between the structures was low (C $_{\alpha}$ rmsd of ≤ 1.06) our discussion refers to the high resolution LynD/ATP/PatE/C51A structure except where detailed. In LynD residues 6-143, 151-229, 240-336, and 342-775 in chain A and 4-226, 240-336, and 343-775 in chain B are ordered; the missing residues are presumed to be disordered. As seen in TruD, a Zn²⁺ ion in domain 2 is coordinated in the same manner¹⁶. The structures locate both the catalytic site and substrate recognition: The nucleotide at the active site is located in domain 3, while the leader peptide is bound at the interface between domain 1 from one monomer and domain 3 from the other. Simple distance constraints suggest that processing of the core peptide must occur in the same domain 3 that contacts the leader, and thus the dimer is a functional requirement (Figure 2B). Key molecular interactions governing nucleotide and substrate recognition have been identified.

Nucleotide binding site

Domain 3 (residues 316-775) is structurally homologous to the *E. coli* YcaO domain³⁰ (C $_{\alpha}$ rmsd of 2.34 over 282 residues) and binds nucleotide in an identical location. In LynD AMP complex (Figure 2C), the adenosine ring N7 atom makes a hydrogen bond with the side chain of Q415 and the N1 atom with N536. The ring sits between a cation- π stacking interaction with R344 on one face and van der Waals interactions to C419 and T351 on the other (Figure 2D). The lack of extensive hydrogen bond recognition of the ring is consistent with ability of these enzymes to use other nucleotide triphosphates¹⁶. The O2' and O3' atoms of the ribose molecule form hydrogen bonds with the main chain of A534 and the side chain of E426 whilst the α -phosphate makes a salt bridge with R636 and hydrogen bonds with Q544 and S419 (Figure 2D), very similar to the *E. coli* YcaO domain³⁰. The recognition of the adenosine is subtly different in *E. coli* YcaO³⁰ as the adenosine ring has been built in a different conformer in this structure.

The LynD ADP PO₄³⁻ (Figures 2D, 2E and 2F) and AMP-PNP (Figure 2G) complexes inform about the binding of the β and γ phosphates; the phosphate in the ADP PO₄³⁻ structure, we assume mimics the γ phosphate of ATP. There are differences in the precise positions of the β and γ phosphates but the similarities are informative; in both structures a Mg²⁺ ion, coordinated by E640, binds the β and γ phosphates and a second Mg²⁺ ion, bound by E548, ligates the β phosphate (Figure 2D). In the ADP PO₄³⁻ structure the second Mg²⁺ ion bridges the β and γ phosphate (Figure 2D). These two ions are also found in the *E. coli* YcaO complex³⁰, and are coordinated by equivalent glutamate residues. In the ADP PO₄³⁻ complex a third Mg²⁺ is present, which is coordinated by E423 and bridges the β and γ phosphates (Figure 2D). In addition to the metal ions, the β phosphate has salt links with

R636 and K409, whilst the γ phosphate salt bridges with R552 and R427 (Figure 2D). Very similar interactions are seen in the AMPCPP complex of *E. coli* YcaO³⁰. The ADP PO₄³⁻ has an unusual arrangement of Mg²⁺ ions, which may be a reflection of the 150 mM Mg formate present in the crystallization buffer.

Key residues for catalysis and nucleotide binding

Cyclodehydration of cysteine/serine/threonine proceeds via attack on the nucleotide triphosphate by a hemioorthoamide intermediate (Figure 1B). Our crystal structures reveal the orientation of the nucleotide within the active site, relative to the incoming substrate, is such that the α -phosphate is occluded from attack by the hemioorthoamide, thus precluding an adenylation mechanism that we had previously proposed¹⁶. To ensure the observed orientation of the nucleotide is mechanistically relevant and to eliminate any possibility it is an artifact of crystallization, residues identified as binding to nucleotide binding were mutated, and the subsequent mutants were tested for their ability to bind ATP/AMP and to process substrate PatE'. Nucleotide binding data for LynD and mutants at the active/nucleotide binding site and their ability to process substrates are shown in Table 1 and Supplementary Figure 2 (MS/MS in Supplementary Figure 3 and Supplementary Tables 1 and 3). K409A processes PatE' overnight to give two heterocycles, whilst K409E slows down the reaction such that only one heterocycle is observed overnight (Supplementary Table 1). Both mutants reduce binding to ATP and AMP below our isothermal calorimetry (ITC) detection limits (usually around 2 mM). Two mutants R636A and R636E slow down the reaction rate and impact observed processivity giving a mixture of 0, 1 and 2 heterocycles. Both mutants show a reduction in AMP binding to below ITC detection limits. ATP binding to R636E was not detected by ITC, whilst R636A gave a complex ITC trace with ATP that we have been unable to interpret. R427E slows down the reaction giving a mixture of 1 and 2 heterocycles and retains AMP binding but eliminates detectable ATP binding. Our assays typically use 5mM ATP; when this was reduced to 500 μ M ATP a reduction in rate of heterocyclization was observed for the LynD K409A mutant (Supplementary Figure 4). E423R was completely inactive in our assay. All binding and activity data support the orientation of the nucleotide within the crystal structure and rule out an adenylation mechanism¹⁶. As with TruD¹⁶, LynD was observed to produce AMP (not ADP), pyrophosphate (PPi) and phosphate when incubated with PatE' and like YcaO³⁰ the same products were produced in the absence of substrate (Supplementary Figure 5A). Following the procedure described in the BalhD study²⁶ PatE' we observed ¹⁸O incorporation not only into Pi as described before²⁶ but also into PPi (Supplementary Figures 5B and 5C).

Leader peptide recognition by LynD

Only residues 21-35 of PatE' are ordered in the structure: Q21 to S23 adopt a coil, S24 to E28 adopt a helical turn, S30 to A33 adopt a β -strand conformation and L34 and G35 adopt a coil (Figure 2H). The β -strand adds to the three-stranded anti-parallel β -sheet in domain 1 of LynD, converting it to a four-stranded anti-parallel β -sheet. In addition to β -sheet hydrogen bonds the side chains of PatE' E28 and E31 make both salt bridges and hydrogen bonds with LynD Q47 (main chain), side chains of Y39 and R74 (Figure 2H). Both L26 and L29 of PatE' insert into hydrophobic pockets in Domain 1 of LynD (Figure 2H). The

interactions provide a structural rationale for the minimal leader (residues 26 to 34 of PatE) and site directed mutants previously reported¹⁶. LynD/PatE' interaction affinities were measured for a number of mutants of LynD (Table 2, Supplementary Table 1 and Supplementary Figure 6). In the apo structure of TruD¹⁶ residues 371–415 (LynD numbering) in Domain 3 were disordered but in LynD these are now ordered forming a strand turn helix strand motif that makes the entrance to and part of the nucleotide binding site. We suggest it is the leader peptide that leads to this ordering of LynD by reducing the volume available for this region and secondly by making specific contacts; a salt bridge between PatE' E32 and LynD R399 and a hydrophobic cluster of PatE' L34 with V217 and L398 of LynD. These interactions depend upon the dimer, as it is the other subunit of LynD which supplies these residues and which is ordered by the leader peptide (Figure 2H). The extended β -sheet interaction between the leader and domain 1 we observe is similar to that observed in the lantibiotic dehydratase substrate complex³².

***Trans* and *cis* activation of LynD by the leader**

Overnight incubation of LynD with a short test peptide ITACITFCAYDG (mimicking the core peptide and C-terminal macrocyclization signature of PatE) in the absence of leader predominantly yields species containing only one heterocycle (Figures 3A and B) as previously reported for TruD¹⁶. However co-incubation with exogenous full-length PatE leader (up to the core peptide) or minimal leader accelerates the turnover of the test peptide (6.7- and 5.0-fold, respectively), such that both heterocycles are observed overnight (Figure 3A and Supplementary Figure 7). The acceleration was more pronounced with the full-length leader peptide than with the minimal leader peptide and reproduces a previous report of *trans* activation of PatD by full-length leader²⁹. In our hands this *trans* activation remained slower than seen for full length PatE'. Guided by structural analysis, the leader peptide (residues 21 to 38) was fused to the N-terminus of LynD to construct a self (*cis*) activating enzyme (LynDfusion). This engineered enzyme expressed to a higher soluble yield than native protein and ran normally as a dimer on gel filtration (Supplementary Figure 1). It processed both cysteines in the test peptide at a rate comparable to wild-type LynD with the full length PatE' substrate (Figure 3, Supplementary Figures 7 and 8) and we have also demonstrated activity on a selection of short peptides (Supplementary Figures 9A-D). In addition we show LynDfusion can also process full-length PatE' substrate at a similar rate to wild-type LynD (Supplementary Figure 9E).

Discussion

The ability of heterocyclases to separate recognition from the catalytic site is highly desirable in biotechnology, conferring specificity by recognizing an invariant leader but tolerating a wide range of residues adjacent to the target cysteine (or serine/threonine). Such spatial separation and use of a leader is a feature of other post translational modifying enzymes such as those found in the lantibiotic pathway^{35,36}. Our structural data of the LynD complexes has enabled us to formally identify both the active site and the region responsible for substrate recognition in the PatD class of cyclodehydratases.

These data revealed that ATP binds in domain 3, (the conserved YcaO domain³⁰) (Figure 2D). Residues involved in nucleotide binding were identified, and mutational analysis of

these residues confirmed the nucleotide binding observed in the crystal structure to be correct. In these models structural shielding of the α -phosphate from the substrate peptide makes our previous suggestion of an adenylating mechanism untenable¹⁶, whereas the exposed γ -phosphate does support a kinase mechanism²⁶. However a simple kinase mechanism does not explain the observations that TruD¹⁶ produces AMP and PPi during turnover, as does LynD (both in the presence and absence of substrate) and the related YcaO enzyme produces AMP and PPi in the absence of substrate³⁰.

The leader peptide of the substrate is not required for processing of the terminal cysteine of the core peptide but it is essential for meaningful processing of internal cysteine residues (detectable within hours)^{16,29,37} (Figure 3 and Supplementary Figure 7). The LynD substrate complexes provide insight into the molecular details of substrate recognition, and the role of the leader peptide in efficient processing of heterocyclizable residues. LynD uses the β -sheet of domain 1 to bind a conserved region within the leader of PatE' (Figure 2H), similarly to the lantibiotic dehydratase enzyme³². By anchoring the substrate peptide it converts an intermolecular reaction into an intramolecular one. With the concomitant reduction in the entropy penalty and increase in local concentration, these two factors accelerate catalysis of both first and second heterocyclization (Figure 3). The substrate leader peptide also makes contacts with domain 3, the catalytic domain, where it stabilizes a loop involved in the active site. Importantly the leader peptide bridges domain 1 of LynD from one monomer to domain 3 of LynD from the other monomer (Figure 2H). This is reminiscent of but different from the MccB system, where domain 1 also acts a peptide clamp within a dimeric arrangement. However in MccB the region of domain 1 and the interactions between its 'leader' peptide are different from that observed here²⁴. The BalhC protein, which is required for efficient heterocyclization of cysteine residues in leader containing linear peptides by BalhD, clearly contains the peptide clamp domain and we propose it will operate in similar manner. The presence of a structurally and functionally conserved peptide clamp domain in three different enzymes that operate on different substrates suggests that the peptide clamp may be general motif for peptide processing enzymes that utilize leader peptides for recognition. This idea is supported by the observation of the same peptide clamp in the nisin dehydratase substrate complex³². The low level of sequence homology between MccB and TruD prevented confident identification of the conserved nature of the peptide clamp but with multiple structures of this domain, identification of homologs is now more reliable.

The stabilization of the active site loops by the leader peptide offers a molecular rationale for the *trans* activation of PatD by the leader peptide²⁹ and a molecular route by which catalysis and substrate binding/product release may be correlated. We show that the *trans* effect can be observed even with a very small minimal leader, but stronger *trans* activation occurs with a full-length leader (Figure 3 and Supplementary Figure 7). Following the example from lantibiotic synthetase where the need for a peptide substrate leader was removed by fusing the leader to the lantibiotic synthetase enzyme³⁸, an engineered LynD was designed based on our structural and biochemical insights into the role of the leader in LynD (Figure 4). The fused enzyme (LynDfusion) was indeed as active at processing two cysteines within a short test peptide ITACITFCAYDG as native LynD with PatE' substrate (Figure 3 and Supplementary Figure 7). Additionally, LynDfusion was able to process PatE' as efficiently

as wild-type LynD during an overnight incubation (Supplementary Figure 9E), which is encouraging given the much higher expression of LynDfusion compared with LynD in our hands. We confirmed the utility of the enzyme by evaluating it with other short peptide substrates with seven, eight or nine core peptide residues including one with three cysteines (Supplementary Figure 9A-D). The introduction of multiple heterocycles into a simple, short, linear peptide is a desirable chemical modification and requires enzymes. To this point, the substrate peptide would need at least 15 N-terminal residues added to ensure complete processing. Since these residues are discarded in making the final product, such 'waste' makes the chemical synthesis of the peptide substrates unattractive and favors bacterial production. However, the newly engineered LynD can act on peptides with no leader and for those products where macrocyclization is required only three C-terminal residues (AYD), which are disposed of, are required. This enables the chemical synthesis of substrates, rather than biological production, which is typically higher yielding and permits greater diversity through the introduction of multiple non-natural amino acids, providing the potential for the generation of a large variety of bio-active peptides.

Materials and methods

Protein cloning, expression and purification

Codon-optimized full-length LynD (*Lyngbya sp.* PCC 8106) with an N-terminal TEV protease-cleavable His₆-tag was purchased from DNA2.0 in the pJexpress411 plasmid. The protein was expressed in *Escherichia coli* BL21 (DE3) grown in auto-induction medium³⁹ for 48 h at 20 °C. Cells were harvested by centrifugation at 4,000 × g, 4 °C for 15 min and re-suspended in lysis buffer (500 mM NaCl, 20 mM Tris pH 8.0, 20 mM imidazole and 3 mM β-mercaptoethanol (BME)) with the addition of complete EDTA-free protease inhibitor tablets (Roche) and 0.4 mg DNase g⁻¹ wet cells (Sigma). Cells were lysed by passage through a cell disruptor at 30 kPSI (Constant Systems Ltd) and the lysate was cleared by centrifugation at 40,000 × g, 4 °C for 20 min. The cleared lysate was applied to a Nickel Sepharose 6 Fast Flow (GE Healthcare) column pre-washed with lysis buffer and protein eluted with 250 mM imidazole. The protein was then passed over a desalting column (Desalt 16/10, GE Healthcare) in 100 mM NaCl, 20 mM Tris pH 8.0, 20 mM imidazole, 3 mM BME. Tobacco etch virus (TEV) protease was added to the protein at a mass-to-mass ratio of 1:10 and the protein digested for 2 h at 20 °C to remove the His₆-tag. Digested protein was passed over a second Ni-column and the flow-through loaded onto a monoQ column (GE Healthcare) equilibrated in 100 mM NaCl, 20 mM Tris pH 8.0, 3 mM BME. Protein was eluted from the monoQ column through a linear NaCl gradient, eluting at 250 mM NaCl. Finally, the protein was subjected to size-exclusion chromatography (Superdex™ 200, GE Healthcare) in 150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP, and concentrated to 8 mg mL⁻¹. Integrity and identity were confirmed by SDS_PAGE and mass spectrometry.

PatE' was synthetically produced in the pBMS vector (a gift from H. Liu) with a C-terminal His₆-tag and expressed in *Escherichia coli* BL21 (DE3) cells grown in auto induction medium³⁹ for 24 h at 30 °C where the protein was driven to inclusion bodies. Amino acid sequence of PatE':

MDKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASKITACITFCAYDDEL
 EHHH HHH Cells were harvested by centrifugation at $4,000 \times g$ for 15 min at 4°C . Cells were re-suspended in urea lysis buffer (8 M urea, 500 mM NaCl, 20 mM Tris pH 8.0, 20 mM Imidazole and 3 mM β -mercaptoethanol (BME) and lysed by sonication at 15 microns (SoniPrep 150, MSE). The lysate was cleared by centrifugation at $40,000 \times g$, 20°C for 20 min followed by passage through a $0.45 \mu\text{m}$ filter. The cleared lysate was applied to a Ni-sepharose FF column (GE Healthcare) pre-washed with urea lysis buffer and protein eluted with 250 mM imidazole. The protein was then supplemented with 10 mM DTT and incubated at room temperature for 2 h before size-exclusion chromatography (Superdex™ 75, GE Healthcare) in 150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP. Peak fractions were pooled and concentrated to 1 mM.

To generate full length leader peptide for *trans* activation studies the PatE' sequence was mutated at 4 positions: D2Y, K3E, K4E and R16E. This quadruple mutant (PatE'4) was expressed and purified as described for PatE', and subjected to heterocyclization tests to ensure normal processing. The mutations K3E, K4E, and R16E enable full-length leader to be retained following digestion of PatE'4 with trypsin (D2Y permits quantification of the leader peptide). PatE'4 was digested with 1/100 trypsin at 37°C , 300 rpm for 3 h, and subsequently applied to a Ni-sepharose FF column (GE Healthcare) equilibrated in 150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP. The flow-through was collected, which was confirmed to contain full-length PatE'4 leader peptide by mass spectrometry, and concentrated to 1 mM.

LynD and PatE point mutants were produced using the Phusion® site-directed mutagenesis kit (Finnzymes) following the manufacturer's protocol. All mutant proteins were expressed and purified as above. LynD fusion enzyme was produced by adding residues 21-38 of the PatE leader and an eleven residue long linker to the N-terminus of LynD via PCR. The protein was expressed and purified as described for native LynD above.

Protein sequence of LynD fusion protein:

MSHHHHHDYDENLYFQGSQSLAELSEEALGDAGAGAGAGAGAMQSTPLLQIQP
 HFHVEVIEPKQVYLLGEQANHALTGQLYCQILPLLNGQYFLEQIVEKLDGVEPPEYIDYVLE
 RLAEKGYLTEAAPELSSEVAAFWSELGIAPPVAAEALRQPVTLPVGNISEVTVAAALTTALR
 DIGISVQTPTEAGSPTALNVVLTDDYLQPELAKINKQALESQQTWLLVKPVGSVLWLGPFV
 VPGKTGCWDCLAHLRLGNREVEASVLRQKQAQQQRNGQSGSVIGCLPTARATLPSTLQTG
 LQFAATEIAKWIVKYHVNATAPGTVFFPTLDGKIITLNHSILDLSKSHILIKRSQCPTCGDPKIL
 QHRGFEPKLKLESRPKQFTSDGGHRGTTPEQTVQKYQHLISPVTVGVVTELVRITDPANPLVHT
 YRAGHSFGSATSRLRLNRLKHKSSGKGKTDSSKASGLCEAVERYSGIFQGDEPRKRATL
 AELGDLAIHPEQCLCFSDGQYANRETLNEQATVAHDWIPQRFDAQIAEWTPVWSLTEQT
 HKYLPTALCYHYHLPPEHRFARGDSNGNAAGNTLEEAILQGMELVERDGVALLWVYNR
 LRRPAVDLGSFNPEYFVQLQQFYRENRDLWVLDLTADLGIPAFAGVSNRKTGSSERLILG
 FGAHLDPITAILRAVTEVNQIGLELDKVPDENLKS DATDWLITEKLADHPYLLPDTTQPLKT
 AQDYPKRWSSDDIYTDVMTCVNIAQQAGLETLVIDQTRPDIGLNVVKTVPGMRFHWSRFG
 EGRLYDVPVKGWLDEPLTEAQMNPMPF

Preparation of test substrates for LynDfusion enzyme

PatE' variants with seven (PatE'-IACITFC) and nine (PatE'-IITACIMAC) core peptide residues, and one with three cysteines (PatE'-ICACITFC) in the core peptide were cloned, expressed and purified as described for PatE'. Each peptide was digested with 1/100 trypsin at 37 °C, 300 rpm for 3 h and subsequently purified using size-exclusion chromatography (Superdex™ 30, GE Healthcare) in 150 mM NaCl, 10 mM HEPES pH 7.4, 2 mM TCEP.

Heterocyclization reactions

For all heterocyclization reactions, 100 µM PatE' (and variants,) was incubated with 5 µM enzyme in 150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP, 5 mM ATP and 5 mM MgCl₂ for 16 h at 37 °C. Samples were analyzed by ESI or MALDI MS (LCT, Micromass or 4800 MALDI TOF/TOF Analyzer, ABSciex). For *cis trans* activation studies, 100 µM ITACITFCAYDG synthetic test peptide (Peptide Protein Research Ltd) was reacted with 5 µM heterocyclase enzyme in the presence and absence of 5.5 µM LAELSEEAL synthetic peptide, or 5.5 µM full-length PatE'4 leader peptide. The reaction was carried out in 150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP, 5 mM ATP and 5 mM MgCl₂ for 16 h at 37 °C and analyzed as above.

To assess the relative rates of heterocyclization under various conditions, heterocyclization reactions were monitored at regular intervals using MALDI TOF MS. In each case, reactions of PatE' or ITACITFCAYDG synthetic peptide were prepared as described and incubated at 37 °C prior to addition of LynD (wild-type, mutants and fusion) allowing a 0 time-point to be recorded. Time points for each reaction were recorded as follows: For PatE' with LynD (standard conditions), and ITACITFCAYDG with LynD fusion, the reaction was monitored after 1, 5, 10, 15, 20, 30, 45, 60, 90, 120 and 180 minutes; For ITACITFCAYDG with LynD in the absence and presence of either minimal leader LAELSEEAL peptide, or full length PatE'4 leader peptide, the reaction was monitored after 15, 30, 60, 120, 240 and 1440 minutes; For PatE' with wild-type LynD and LynD K409A with 500 µM ATP, the reaction was monitored after 10, 30, 60, 120 and 360 minutes. Reactions were set up in triplicate, and each sample was analyzed by MALDI TOF MS in triplicate (9 spectra recorded in total for each reaction at each time point). For each time point, the total ion count for each species (0, 1 and 2 heterocycles) was recorded, averaged and the percentage of each species was calculated. The rate of each reaction relative to the wild-type reaction (PatE' + LynD) was determined.

LC-MS of reaction mixtures was performed on a Waters LC-MS system (LCT mass spectrometer and 2795 HPLC) using a Waters MassPrep column (2.1 × 10 mm). Solvent B was 0.1% formic acid and Solvent A was MeCN containing 0.1% formic acid. Gradient 0 - 0.5min 98 % B, 0.5 - 2.5 min linear to 2 % B, 2.5 - 4.5 min 2 % B, 4.5 - 4.6 min linear to 98 % B, 1.6 - 12 min 98 % B at 0.05 ml min⁻¹. The following conditions were used on the mass spectrometer: ESI +ve, capillary voltage 3.5 kV, cone voltage 40 V, mass range 500-2500 m/z, RF lens 500. The spectra were combined across the eluted protein peak and the charged ion series processed using Water's MaxEnt algorithm to give protein mass, using peak width at half height on the strongest peak in the ion envelope. The data was calibrated externally against horse heart myoglobin (16,951.5 Da).

MALDI MS was acquired using a 4800 MALDI TOF/TOF Analyzer (ABSciex, Foster City, CA) equipped with a Nd:YAG 355 nm laser in linear mode and calibrated using the $[M+H]^+$ and $[M+2H]^{2+}$ peaks of ubiquitin. The spot was analyzed in positive MS mode between 3000 and 10000 m/z , by averaging 1000 laser spots.

Fragmentation MS of the peptides were carried out on an ABSciex 5600 mass spectrometer with Eksigent nanoLC and ThermoScientific Aclaim Pepmap RSLC column 75 μ M \times 150mm. Trap and elute methodology was used with a 6 minute trap wash. Trapping solvent A was 98 % water, 2 % MeCN, 0.05 % Trifluoroacetic acid. Gradient solvent A was 98 % water, 2 % MeCN, and 0.1 % formic acid and gradient solvent B was 98 % MeCN, 2 % water, and 0.1 % formic acid. Gradient 0 -0.5 min 5 % B, 0.5 - 5 min linear to 40 % B, 5 - 6 min linear to 95 % B, 6 - 9 min 95 % B, 9 - 10 min linear to 2 % B, hold at 2 % B until 20 mins at 300 nl min^{-1} , and 45 °C. The following conditions were used on the mass spec: ESI +ve, 0.25 s MS accumulation and 0.15 s MSMS accumulation, collision energy of 45 V with rolling collision energy optimization applied. The data was calibrated externally prior to analysis with 8 peptides from a tryptic digest of 25 $\text{fmol } \mu\text{l}^{-1}$ β -galactosidase.

MALDI mass spectrometry for the time course was carried out on an ABSciex 4800 MALDI TOF/TOF mass spectrometer. 0.5 μl of sample was co-spotted with 0.5 μl of matrix (10 mg ml^{-1} alpha cyano- 4-hydroxycinnamic acid in 50% MeCN and 50% 0.1 % TFA) and the sample left to dry. Spectra were collected over the range 500 – 4000 m/z (reflectron for peptides < 4000 m/z) or 3000 – 10000 m/z (linear for peptides > 4000 m/z) with 20 subspectra of 50 shots accumulated randomly across the spot. The data was calibrated externally prior to analysis with 6 peptides from ABSciex standard 6 peptide mix (PN 4465940) in reflectron or the $[M+H]^+$ and $[M+2H]^{2+}$ peaks of ubiquitin in linear.

Crystallization, data collection, and crystallographic analysis

Co-crystallization trials of LynD with various peptide and nucleotide substrates were set up at 7.5 mg ml^{-1} LynD with a 1.1 molar excess of PatE' (or variants), 5 mM nucleotide and 3 mM MgCl_2 . LynD crystals in complex with PatE'-C51A and ATP were obtained in 16 % PEG 3350, 0.1 M Mg formate. The crystals were cryoprotected in 37 % PEG 3350, 0.1 M Mg formate, 1 mM ATP and flash-cooled in liquid nitrogen. These crystals belonged to space group $P2_12_12_1$ with cell dimensions $a = 81.8 \text{ \AA}$ $b = 116.1 \text{ \AA}$, $c = 183.5 \text{ \AA}$. Diffraction data was collected at Diamond beamline I24 at 100 K and processed with Xia2⁴⁰.

LynD crystals in complex with PatE' and AMP were obtained in 13 % (w/v) PEG 3350, 0.1 M Tris pH 8.5, 0.2 M CaCl_2 and 0.6 % (w/v) myo-inositol. The crystals were cryoprotected in 35 % PEG 4000, 0.2 M CaCl_2 , 0.1 M Tris pH 8.5, 1 mM AMP and flash-cooled in liquid nitrogen. These crystals belonged to space group $P2_12_12_1$ with cell dimensions $a = 65.8 \text{ \AA}$ $b = 152.8 \text{ \AA}$, $c = 182.8 \text{ \AA}$. Diffraction data was collected at Diamond beamline I02 at 100 K and processed with Xia2⁴⁰. LynD crystals in complex with PatE-C51A and β - γ -imido-ATP were obtained in 28 % PEG 4000, 0.1 M Tris pH 8.9, 0.2 M LiSO_4 . The crystals were cryoprotected in 40 % PEG 4000, 0.1 M Tris pH 8.9, 5 mM β - γ -imido-ATP and flash-cooled in liquid nitrogen. These crystals belonged to space group $P2_12_12_1$ with cell dimensions $a = 65.9 \text{ \AA}$ $b = 152.9 \text{ \AA}$, $c = 182.5 \text{ \AA}$. Diffraction data was collected at ESRF beamline ID29 at 100 K and processed with Xia2⁴⁰.

The structure of LynD in complex with PatE' and AMP was determined by molecular replacement using Phaser^{41,42} with TruD as a search model (PDB: 4BS9). The partially refined structure was then used as a search model for the remaining structures. In each case, complete manual rebuilding was performed with COOT⁴³ and refinement was performed using CCP4 REFMAC5⁴⁴ and Phenix Refine⁴⁵. The statistics of data collection and refinement are summarized in Supplementary Table 2. All molecular graphics figures were generated with the program Pymol⁴⁶.

NMR of ATP in the presence of LynD

³¹P observed NMR spectra were acquired using 1024 scans on a Bruker AVANCE III 500 MHz spectrometer equipped with a room temperature BBFO+ probe. Composite pulse sequence 'waltz-64' was used for ¹H power-gated broadband decoupling.

Selective ¹⁸O labeling of PatE'

PatE' was selectively ¹⁸O labeled at the carbonyl of the residue preceding cysteines based on a protocol described in a study of BalhD²⁶. PatE' was heterocyclized as described and the reaction mixture was purified using size-exclusion chromatography (Superdex™ 75, GE Healthcare). Pure heterocyclized-PatE' was concentrated to 0.5 mL and dried to a powder using a Thermo Savent Speed Vac. To accomplish selective ¹⁸O labeling the dried sample was re-suspended in H₂¹⁸O supplemented with 0.1 % formic acid to hydrolyze the heterocycles, and incubated at 20 °C. The ring opening was followed using MALDI TOF MS. Once complete, selectively labeled ¹⁸O-PatE' was dialyzed into 150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM DTT. To ensure all traces of acid and H₂¹⁸O were removed the sample was dialyzed 3 × 1:1000 (2 × 2h followed by o/n). The dialyzed sample was recovered, diluted to 7.5 mL in 8M urea containing 10 mM DTT and passed through a Superdex™ 75 size-exclusion column (GE Healthcare) pre-equilibrated in 150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM DTT to yield pure ¹⁸O₂-PatE' as confirmed by mass spectrometry.

Identification of ¹⁸O Pyrophosphate

Purification of pyrophosphate from heterocyclization reactions was achieved by passing the reaction mixture through a 5 kDa cut-off protein concentrator. The flow-through was applied to CaptoQ resin (GE Healthcare), washed with 15 column volumes dH₂O and pyrophosphate eluted with 1M Na acetate pH 7.0.

Accurate mass measurements were carried out on ABSciex 5600 mass spectrometer with nanospray source. Samples were diluted 50% with ACN and infused at 1ul/min using the inbuilt syringe pump. Negative ionization spectra were collected from 60- 400m/z in MS mode. Masses 176.9 and 178.9 m/z were isolated in Q1, fragmented with collision energy 30V in Q2 and the fragments measured from 50-200m/z on the TOF analyzer. The TOF was externally calibrated with a mixture of Periodate (190.8847 m/z) and sulphanilic acid (172.0074 m/z)

ITC analysis of LynD (and mutants) with substrate peptide, and nucleotide

ITC experiments were performed using a VP-ITC instrument (MicroCal) in heterocyclization reaction buffer (150 mM NaCl, 10 mM HEPES pH7.4, 1mM TCEP) supplemented with 5 mM MgCl₂. For substrate binding experiments a cell solution of 10 μM LynD (and mutants) and a syringe solution of 150 μM PatE' (and mutants) were prepared by diluting protein and peptide with the buffer used for dialysis. For nucleotide binding experiments, the cell concentration was increased to 20 μM, and the ATP/AMP concentration in the syringe solution was 300 μM. Substrate, and ATP experiments were performed at 20 °C, and AMP experiments at 25 °C. Cell and syringe solutions were degassed for 15 min at 18 and 23 °C. Titration method was as follows: one injection of 2 μl followed by injections of 5 μl at 0.5 μl/min, with a delay of 4 min between every injection. The stirring speed was 307 rpm throughout. Raw data was processed using MicroCal Origin software, and the baseline was adjusted and the integrations limits were selected manually. Data was non-linearly fitted to the one-site model (Origin) setting the stoichiometry to 1; allowing the stoichiometry to refine gave values above 0.9 and a two site model gave meaningless fits.

Database entries

PDB codes 4v1v 4v1u and 4v1t.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants from the ERC 339367 (J.H.N. and M.J.) and BBSRC BB/K015508/1 (J.H.N and M.J.). We acknowledge use of the Diamond (beamlines I02 and I24) and ESRF (beamline ID29) synchrotrons. JHN is Royal Society Wolfson Merit Award Holder and a 1000 talent scholar of the Chinese Government at Sichuan University.

References

1. Blunt JW, Copp BR, Keyzers RA, Munro MH, Prinsep MR. Marine natural products. *Nat Prod Rep.* 2012; 31:160–258.
2. Driggers EM, Hale SP, Lee J, Terrett NK. The exploration of macrocycles for drug discovery--an underexploited structural class. *Nat Rev Drug Discov.* 2008; 7:608–624. [PubMed: 18591981]
3. Mayer AM, Rodriguez AD, Tagliatalata-Scafati O, Fusetani N. Marine pharmacology in 2009–2011: marine compounds with antibacterial, antidiabetic, antifungal, anti-inflammatory, antiprotozoal, antituberculosis, and antiviral activities; affecting the immune and nervous systems, and other miscellaneous mechanisms of action. *Mar Drugs.* 2013; 11:2510–2573. [PubMed: 23880931]
4. Sivonen K, Leikoski N, Fewer DP, Jokela J. Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. *Appl Microbiol Biotechnol.* 2010; 86:1213–1225. [PubMed: 20195859]
5. Schmidt EW, et al. 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.* 2005; 102:7315–7320. [PubMed: 15883371]
6. Donia MS, et al. Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat Chem Biol.* 2006; 2:729–735. [PubMed: 17086177]

7. Long PF, Dunlap WC, Battershill CN, Jaspars M. Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production. *Chembiochem*. 2005; 6:1760–1765. [PubMed: 15988766]
8. Arnison PG, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep*. 2013; 30:108–160. [PubMed: 23165928]
9. Milne BF, Long PF, Starcevic A, Hranueli D, Jaspars M. Spontaneity in the patellamide biosynthetic pathway. *Org Biomol Chem*. 2006; 4:631–638. [PubMed: 16467937]
10. Schmidt EW, Donia MS. Chapter 23. Cyanobactin ribosomally synthesized peptides--a case of deep metagenome mining. *Methods Enzymol*. 2009; 458:575–596. [PubMed: 19374999]
11. Donia MS, Schmidt EW. Linking chemistry and genetics in the growing cyanobactin natural products family. *Chem Biol*. 2011; 18:508–519. [PubMed: 21513887]
12. Housen WE, et al. An Efficient Method for the In Vitro Production of Azol (in) e - Based Cyclic Peptides. *Angew Chem Int Ed Engl*. 2014; 126:14395–14398.
13. Housen WE, et al. The discovery of new cyanobactins from *Cyanothece* PCC 7425 defines a new signature for processing of patellamides. *Chembiochem*. 2012; 13:2683–2689. [PubMed: 23169461]
14. Koehnke J, et al. The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nat. Struct. Mol. Biol*. 2012; 19:767–772. [PubMed: 22796963]
15. Koehnke J, et al. An enzymatic route to selenazolines. *Chembiochem*. 2013; 14:564–567. [PubMed: 23483642]
16. Koehnke J, et al. The cyanobactin heterocyclase enzyme: a processive adenylyase that operates with a defined order of reaction. *Angew Chem Int Ed Engl*. 2013; 52:13991–13996. [PubMed: 24214017]
17. Agarwal V, Pierce E, McIntosh J, Schmidt EW, Nair SK. Structures of cyanobactin maturation enzymes define a family of transamidating proteases. *Chem Biol*. 2012; 19:1411–1422. [PubMed: 23177196]
18. Bent AF, et al. Structure of PatF from *Prochloron didemni*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2013; 69:618–623. [PubMed: 23722837]
19. Majmudar JD, Gibbs RA. Pericyclic prenylation: peptide modification through a Claisen rearrangement. *Chembiochem*. 2011; 12:2723–2726. [PubMed: 22114066]
20. Melby JO, Li X, Mitchell DA. Orchestration of enzymatic processing by thiazole/oxazole-modified microcin dehydrogenases. *Biochemistry*. 2014; 53:413–422. [PubMed: 24364559]
21. McIntosh JA, Schmidt EW. Marine molecular machines: heterocyclization in cyanobactin biosynthesis. *Chembiochem*. 2010; 11:1413–1421. [PubMed: 20540059]
22. McIntosh JA, Donia MS, Schmidt EW. Insights into heterocyclization from two highly similar enzymes. *J. Am. Chem. Soc*. 2010; 132:4089–4091. [PubMed: 20210311]
23. Nielsen DS, et al. Improving on Nature: Making a Cyclic Heptapeptide Orally Bioavailable. *Angew Chem Int Ed Engl*. 2014; 53:12059–12063. [PubMed: 25219505]
24. Regni CA, et al. How the MccB bacterial ancestor of ubiquitin E1 initiates biosynthesis of the microcin C7 antibiotic. *EMBO J*. 2009; 28:1953–1964. [PubMed: 19494832]
25. Dunbar KL, Mitchell DA. Insights into the mechanism of peptide cyclodehydrations achieved through the chemoenzymatic generation of amide derivatives. *J. Am. Chem. Soc*. 2013; 135:8692–8701. [PubMed: 23721104]
26. Dunbar KL, Melby JO, Mitchell DA. YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nat Chem Biol*. 2012; 8:569–575. [PubMed: 22522320]
27. Melby JO, Dunbar KL, Trinh NQ, Mitchell DA. Selectivity, directionality, and promiscuity in peptide processing from a *Bacillus* sp. Al Hakam cyclodehydratase. *J. Am. Chem. Soc*. 2012; 134:5309–5316. [PubMed: 22401305]
28. Leikoski N, et al. Highly diverse cyanobactins in strains of the genus *Anabaena*. *Appl. Environ. Microbiol*. 2010; 76:701–709. [PubMed: 20008171]

29. Goto Y, Ito Y, Kato Y, Tsunoda S, Suga H. One-pot synthesis of azoline-containing peptides in a cell-free translation system integrated with a posttranslational cyclodehydratase. *Chem Biol.* 2014; 21:766–774. [PubMed: 24856821]
30. Dunbar KL, et al. Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. *Nat Chem Biol.* 2014; 10:823–829. [PubMed: 25129028]
31. Blumenthal E, Herbert JBM. Interchange reactions of oxygen. 1 Interchange of oxygen between water and potassium phosphate in solution. *Trans. Faraday Soc.* 1937; 33:849–852.
32. Ortega MA, et al. Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature.* 2015; 517:509–512. [PubMed: 25363770]
33. Miller GA Jr, Rosenzweig S, Switzer RL. Oxygen-18 studies of the mechanism of pyrophosphoryl group transfer catalyzed by phosphoribosylpyrophosphate synthetase. *Arch. Biochem. Biophys.* 1975; 171:732–736. [PubMed: 173242]
34. McLennan AG, Taylor GE, Prescott M, Blackburn GM. Recognition of beta beta'-substituted and alpha beta, alpha'beta'-disubstituted phosphonate analogues of bis(5'-adenosyl) tetraphosphate by the bis(5'-nucleosidyl)-tetraphosphate pyrophosphohydrolases from *Artemia* embryos and *Escherichia coli*. *Biochemistry.* 1989; 28:3868–3875. [PubMed: 2546583]
35. Willey JM, van der Donk WA. Lantibiotics: peptides of diverse structure and function. *Annu. Rev. Microbiol.* 2007; 61:477–501. [PubMed: 17506681]
36. Oman TJ, van der Donk WA. Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat Chem Biol.* 2010; 6:9–18. [PubMed: 20016494]
37. Ruffner DE, Schmidt EW, Heemstra JR. Assessing the combinatorial potential of the RiPP cyanobactin tru pathway. *ACS Synth Biol.* Aug 20.2014 Epub. DOI: 10.1021/sb500267d. [PubMed: 25140729]
38. Oman TJ, Knerr PJ, Bindman NA, Velasquez JE, van der Donk WA. An engineered lantibiotic synthetase that does not require a leader peptide on its substrate. *J. Am. Chem. Soc.* 134:6952–6955. [PubMed: 22480178]
39. Studier FW. Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification.* 2005; 41:207–234. [PubMed: 15915565]
40. Winter G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography.* 2009; 43:186–190.
41. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Likelihood-enhanced fast translation functions. *Acta Crystallographica Section D-Biological Crystallography.* 2005; 61:458–464.
42. Storoni LC, McCoy AJ, Read RJ. Likelihood-enhanced fast rotation functions. *Acta Crystallographica Section D-Biological Crystallography.* 2004; 60:432–438.
43. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography.* 2004; 60:2126–2132.
44. Murshudov GN, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr.* 67:355–367. [PubMed: 21460454]
45. Adams PD, et al. Recent developments in the PHENIX software for automated crystallographic structure determination. *Journal of Synchrotron Radiation.* 2004; 11:53–55. [PubMed: 14646133]
46. DeLano WL. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA. 2002

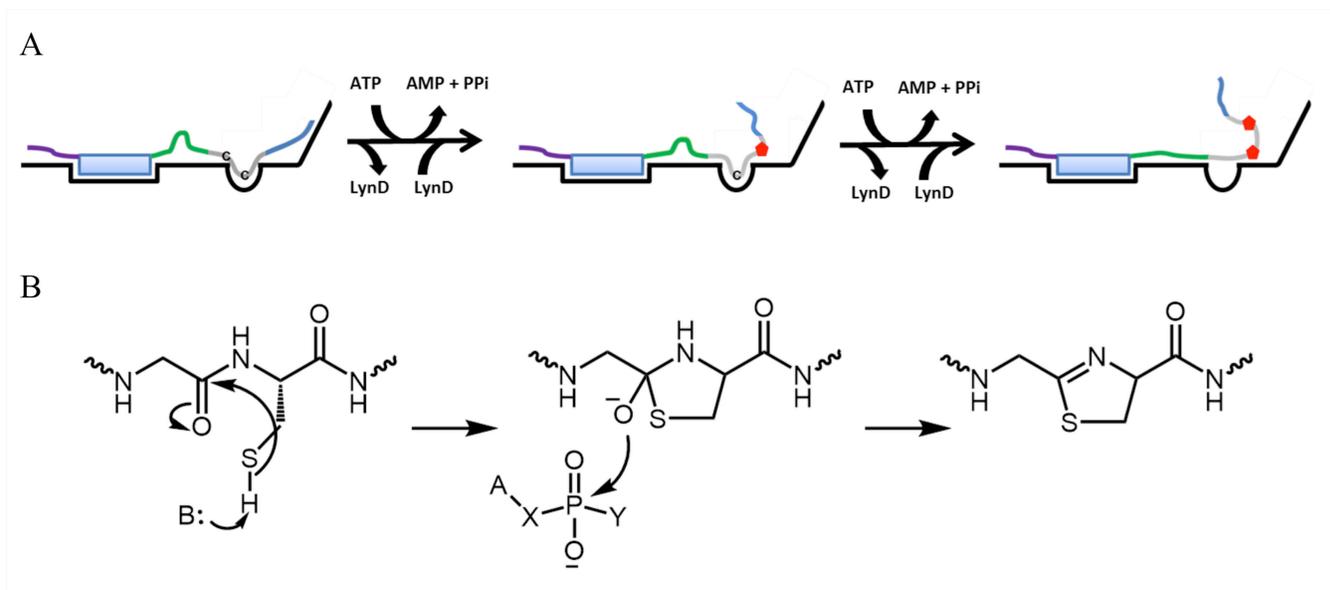


Figure 1. The heterocyclase reaction

A Cartoon schematic summarizing the processing cysteine residues within the core peptide of PatE precursor peptide by the ATP-dependent cyclodehydratase (heterocyclase) enzyme LynD, highlighting the defined order of heterocyclization. The enzyme works by binding substrate, processing a single heterocycle then dissociating, this is denoted by the arrows in the diagram.

B The chemical steps in cyclodehydration begin with abstraction of the sulfhydryl proton followed by attack at the adjacent (N-terminal) carbonyl carbon generating a hemiothoamide. In an irreversible step this intermediate attacks ATP, and the resulting phosphate species is eliminated, forming a thiazoline. Broadly there are three possible routes by which ATP reacts: an adenylation mechanism where $X = O$ and $Y = \text{PPi}$; a kinase mechanism where $X = \text{PPi}$ and $Y = O$; and a pyrophosphorylation mechanism where $X = Y = \text{PO}_3$.

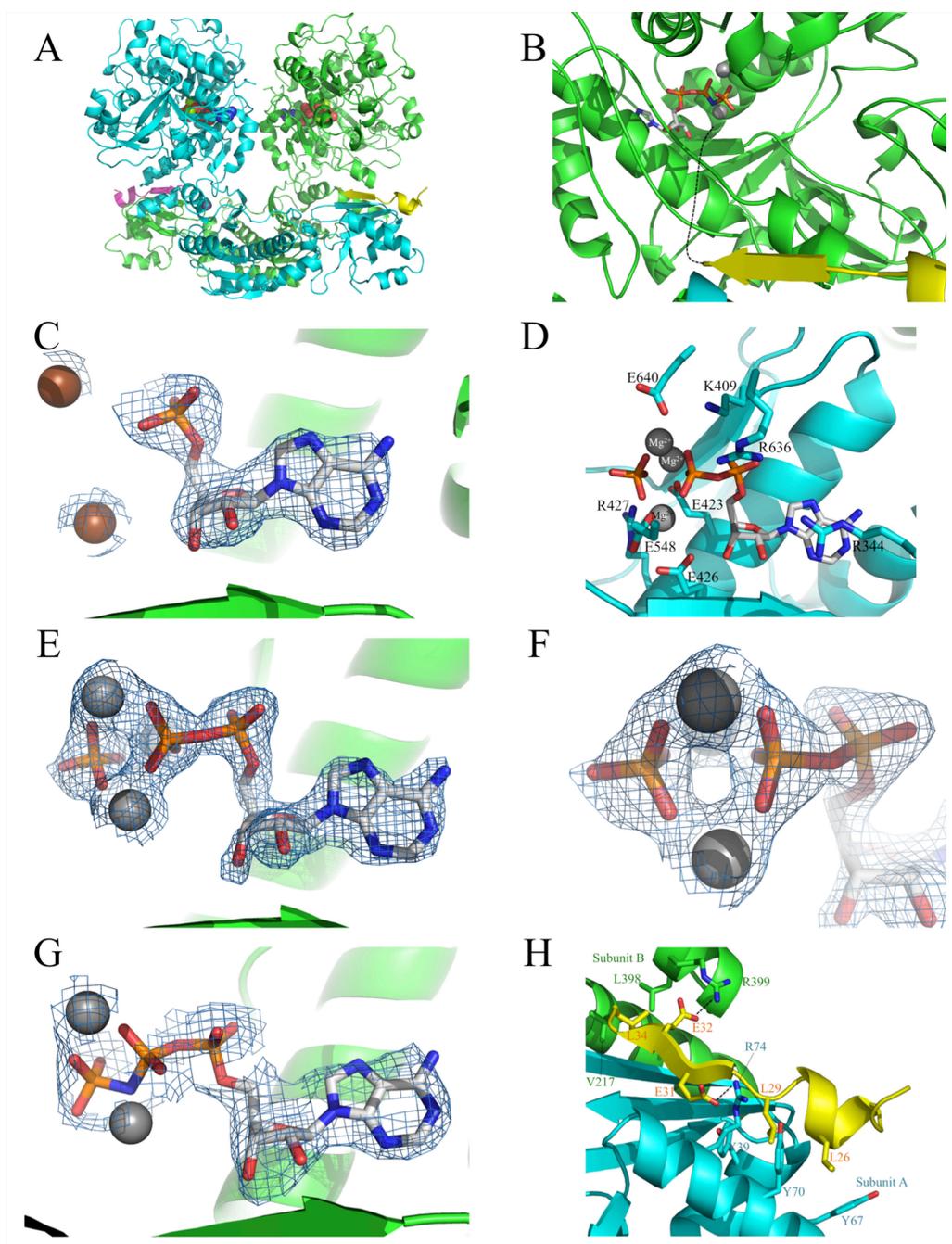


Figure 2. Crystal structures of LynD complexes

A Cartoon representation of the LynD/ATP/PatE/C51A complex structure. LynD forms a functional dimer, with monomers colored green and cyan. The ordered residues of PatE/C51A are shown in yellow and magenta. ATP bound in domain 3 is shown as colored spheres and structural zinc ions are shown as grey spheres.

B The last ordered C-terminal residue of PatE' is 26 Å distant from the nearest ATP (in the other subunit to which the leader is bound) with a clear route through protein. Shielding of

the α -phosphate precludes an adenylation mechanism. Magnesium ions are shown as grey spheres.

C Difference electron density for LynD with AMP, Ca^{2+} ions are shown as brown spheres, atoms are colored as Fig 1C. $F_o - F_c$ maps were contoured at 3σ with phases calculated from a model, which was refined with no nucleotide or metal ions present after simulated annealing (5000 K to 300 K in 100 K steps, Cartesian and torsion).

D Close up of nucleotide site, atoms colored as Fig 2C.

E LynD crystallized with ATP shows ADP, Pi and Mg^{2+} with maps calculated and atoms colored as Fig 2C.

F Zoom view of Mg^{2+} coordination in Figure 2E.

G LynD with β,γ -imido-ATP and Mg^{2+} with maps calculated and atoms colored as Figure 2C.

H The peptide clamp between domains 1 (cyan) and 3 (green) that binds the PatE' leader peptide (yellow). Amino acids directly involved in substrate recognition are shown as sticks.

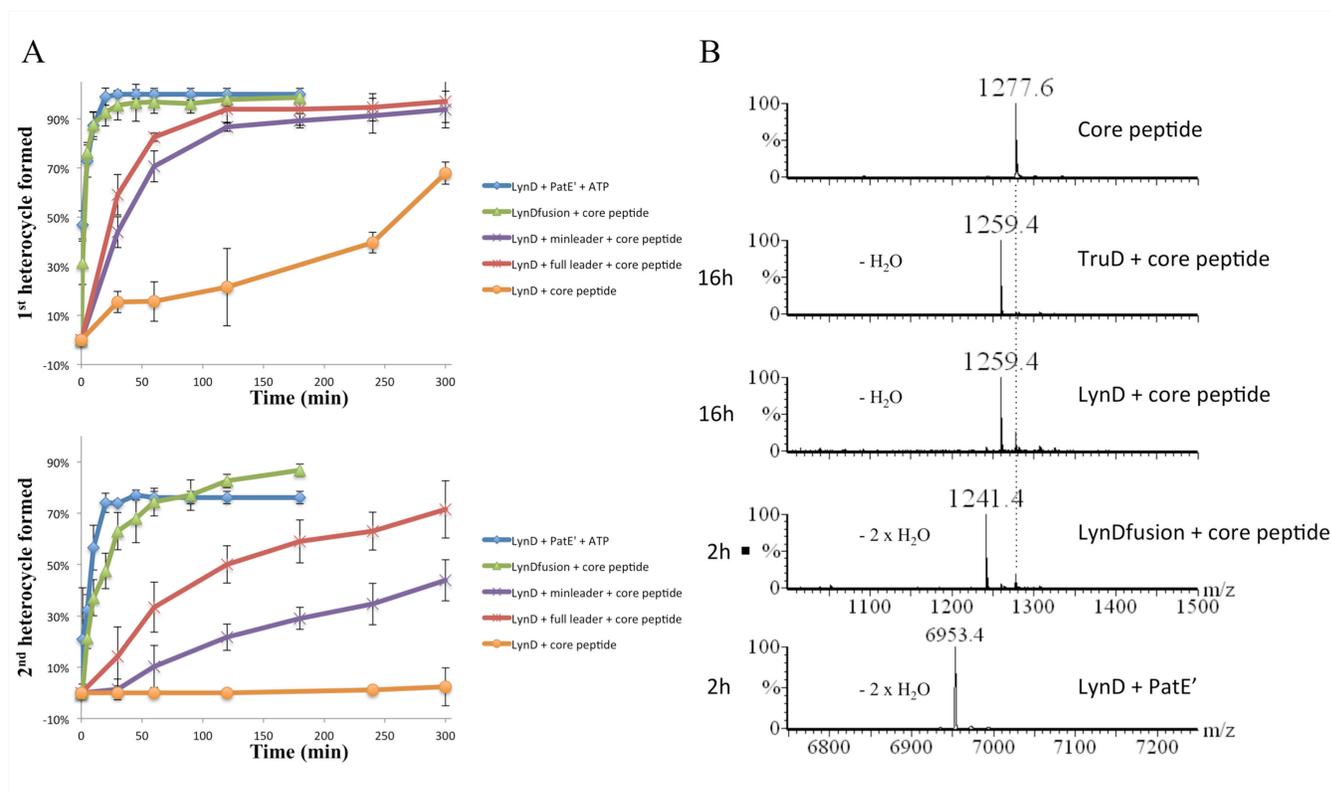


Figure 3. The role of the substrate leader in promoting catalysis

A Relative rates of various *cis* and *trans* activated heterocyclization reactions analyzed by MALDI TOF MS. The top graph shows the time taken to complete 1st heterocycle and the bottom graph shows the time taken to form the 2nd heterocycle. Each time point was analysed by three measurements, the experiment was repeated three times; thus each point represents 9 measurements. Errors are plotted as ± 1 s.d..

B LCMS analysis of heterocyclization reactions of core peptide (ITACITFCAYDG) incubated with TruD, LynD and LynDfusion. After 16 h core peptide incubated with either TruD or LynD only 1 heterocycle is formed. In contrast when incubated with LynDfusion, the heterocyclization reaction is nearly complete – the sample containing predominantly 2 heterocycles after just 2 h. PatE' and LynD reaction after 2 h is shown as a reference.

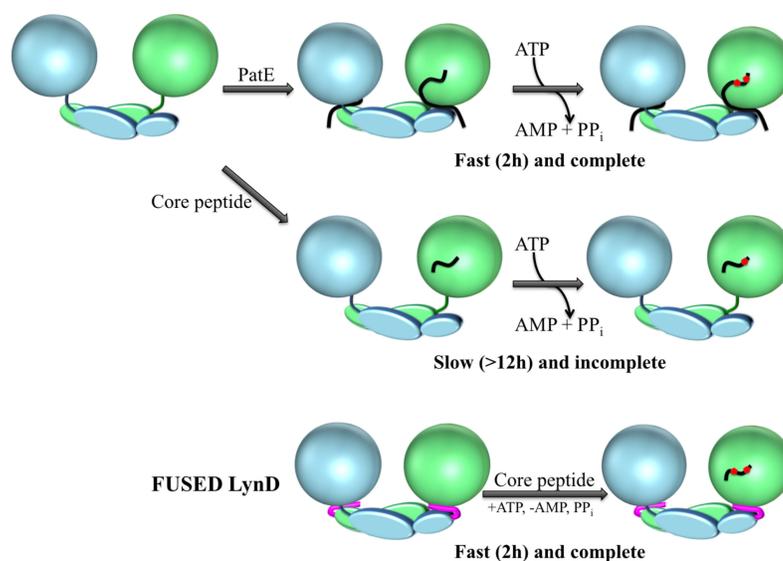


Figure 4. An engineered heterocyclase enzyme

Cartoon schematic highlighting the significant conformational rearrangement of the LynD dimer (blue and green) following PatE' binding (black), generating an 'active' enzyme. In the absence of a leader peptide the conformational rearrangement is not possible and heterocyclization of the core peptide is inefficient. In this way covalently bound leader peptide (pink) in the LynDfusion mimics the 'active' enzyme, restoring efficient 'wild type-like' processing of the core peptide.

Table 1
Nucleotide binding and processing of LynD mutants

Enzyme	Nucleotide K_D (μ M)		No. of heterocycles (PatE')
	ATP	AMP	
LynD	50.25	14.41	2
LynD K409E	no binding	no binding	2
LynD K409A	no binding	no binding	1
LynD E423R	no binding	5.65	0
LynD R427E	no binding	23.04	1 and 2
LynD R636E	no binding	no binding	0, 1 and 2
LynD R636A	N/A	no binding	0, 1 and 2

Table 2
LynD and PatE' mutations and their effects on binding and processing

Enzyme	Peptide	K _D (μM)	No. of heterocycles
LynD	PatE'	1.49	2
LynD	PatE' L26R	no binding	2
LynD	PatE' L29R	no binding	1 and 2
LynD	PatE' E31R	no binding	1 and 2
LynD	PatE' E32R	18.35	2
LynD Y67D	PatE'	12.92	2
LynD Y67D	PatE' L26R	15.22	2
LynD R74E	PatE'	no binding	2
LynD R74E	PatE' E31R	no binding	1 and 2
LynD R399E	PatE'	10.74	2
LynD R399E	PatE' E32R	8.55	2
LynD Fusion	PatE'	5.5	2