The Cyanobactin Heterocyclase Enzyme: A Processive Adenylylase That Operates with a Defined Order of Reaction**

Jesko Koehnke, Andrew F. Bent, David Zollman, Kieran Smith, Wael E. Houssen, Xiaofeng Zhu, Greg Mann, Tomas Lebl, Richard Scharff, Sally Shirran, Catherine H. Botting, Marcel Jaspars, Ulrich Schwarz-Linek, and James H. Naismith*

Heterocyclic rings are a recurring and iconic motif in organic chemistry. The introduction of five-membered rings into protein backbones is synthetically challenging, but many peptide-based biologically active natural products contain such Ser-, Thr-, and Cys-derived heterocycles[1] and therefore peptide-based biologically active natural products contain protein backbones is synthetically challenging, but many chemistry. The introduction of five-membered rings into...
a kinase type mechanism in which a hemithioamide intermediate is phosphorylated\cite{6-7} (Figure S3).

The cyanobactin class of heterocyclases, exemplified by TruD, possess an almost unique combination of processivity, specificity, chemical versatility, and promiscuity, which thus makes them a fascinating subject for study. We show by biochemical assay that TruD is an adenylase, not a kinase, and that TruD processes cysteines in a defined order, which we describe. We demonstrate that entire substrate leader can be removed and TruD will process a single specific cysteine residue. However, we establish the role of leader is to permit processivity through a balance of recognition.

TruD overexpressed in E. coli BL21 (DE3) cells and purified to homogeneity (as judged by SDS-PAGE) using established procedures\cite{8} was used for all analyses. TruD (5 μm), when incubated with Mg\(^{2+}\), ATP (5 mM) and an engineered PatE2 (single core peptide sequence ITACITFC, Figure S1; 100 μm) results in a loss of mass in PatE2 of two water molecules (Figure S4a). Mass spectrometry (MS) fragmentation data confirms the formation of heterocycles\cite{5,8} validating TruD activity and substrate. 2.2 molar excess of ATP (1.1 for cysteines) was added to PatE2 at 20°C and incubated with TruD for 8 h and immediately the product of the reaction was analyzed by 1D \(^{31}\)P NMR spectroscopy. Authentic AMP, ADP, ATP, pyrophosphate, and phosphate in the same buffer at the same concentration and pH were used as standards (Figure 1a,b). Two new peaks matching AMP and pyrophosphate were identified in the product mixture. Unreacted ATP and a smaller amount of phosphate were observed, but no ADP. We repeated the experiment at 10°C and recorded spectra every 30 min for the first 12 h and every 2 h thereafter (Figure S5a). This time-course confirms AMP and pyrophosphate production (and small amount of phosphate). A control incubation of the mixture with ADP reveals that it only slowly degrades to AMP. These data can only be explained by an adenylation mechanism and rules out a kinase mechanism. Repeating the NMR after overnight incubation or at a temperature > 25°C shows only phosphate (Figure S5b), a fact we attribute to pyrophosphate degradation rather than a time- or temperature-dependent change in mechanism.

ATP, ADP, and AMP bind to TruD with \(K_D\) values of 19, 20, and 12 μm, respectively (Figure S5c), but as ADP decomposes slowly in a dissipated exothermic manner, the corresponding \(K_D\) value is an apparent value obtained by analyzing the rapidly emerging endothermic parts of the ITC signals. Although dTTP does not bind in ITC experiments, we find it and GTP can catalyze heterocyclization of cysteine (Figure S5d–f); BalhD has also been shown to use GTP\cite{9}.

The coupled NADH assay reported in the study of the adenylase enzyme AcsD\cite{10} uses lactate dehydrogenase to oxidize NADH during the reduction of pyruvate. Pyruvate is generated by pyruvate kinase, which requires ADP. The assay has a background rate in the absence of any substrate\cite{10} which for a slow enzyme such as TruD is significant. The addition of TruD and/or PatE2 to the mixture did not increase the rate of NADH consumption above background, indicating no enzymatic production of ADP. Addition of myokinase, which converts AMP and ATP into ADP, accelerates the rate of NADH consumption to 5.04 μm\(^{-1}\)min\(^{-1}\) (Figure 1c), confirming enzymatic production of AMP consistent with an adenylase mechanism.

Time-course MS experiments showed that although both cysteines were heterocyclized after about 60 min at 37°C, there appeared to be an accumulation of a mono-heterocyclized intermediate during the reaction (Figure S4b). Using standard triple-resonance experiments, we assigned the spectra (except four of the six histidine residues of the C-terminal tag) of uniformly 13C, 15N-labeled PatE2 (Figure S6). The sharp, poorly dispersed NMR signals and 13C chemical shifts indicate PatE adopts a natively unfolded state in aqueous buffer (compare with the helical arrangement of N-terminal leader peptide in organic solvent\cite{11}). Real-time \(^{1}H-\(^{15}\)N HSQC-NMR reaction monitoring using uniformly \(^{15}\)N-labeled PatE2 showed a specific order with the first heterocyclization event characterized by the disappearance of eight cross-peaks (assigned to residues 48–55) and the emergence of seven new cross-peaks (Figure 2). In the second heterocyclization event, eight cross-peaks (residues 43–50) disappeared (three of which appeared with first heterocyclization) and seven new peaks appeared (Figure 2).

Heterocyclization of cysteine induces large changes in \(^{1}H\), \(^{15}\)N, C\(^{13}\), and C\(^{15}\) chemical shifts (Figure S6). This allowed us to establish that the terminal cysteine, C51, heterocyclizes
first, followed by the slower heterocyclization of the internal cysteine C47 (Figure 2; Supporting Information, Figure S6).

The reaction was highly temperature-dependent with completion in under 5 h at 20°C, but still incomplete after 44 h at 10°C (Figure 2). The internal cysteine of the PatE2 mutant C51P (chosen to mimic the mono-heterocyclized intermediate) was processed (within error) at the same rate as the internal cysteine in PatE2 (Figure S7). However, the internal cysteine of the C51A PatE2 mutant is processed very slowly (Figure S7). The conformational change at the C-terminus of core peptide introduced upon formation of the five-membered ring is thus critical for TruD to be able to process the internal cysteine of the core peptide.

We incubated a uniformly 15N-labeled PatE with a core peptide with three cysteines (sequence ICACITFC (PatE3C)) and monitored the reaction by HSQC NMR at 10°C and observed that C51 reacts first, followed by C47 and finally C45 (Figure S8), TruD thus works from the C-terminus within the core, but the accumulation of the intermediates shows that they are released and rebound, rather than held on the enzyme. The first two heterocyclizations of PatE3 occur on an identical timescale (at both 10°C and 37°C) to the two-cysteine PatE2 substrate, but the third heterocyclization is slower than the second heterocyclization (which is slower than the first).

PatE2 binds very tightly to TruD (Kd 80 nM) (Figure 3), but titrating TruD into 15N-PatE2 shows that 1H, 15N-HSQC cross-peaks of residues 1–15 are unaffected and thus do not bind to TruD (Figure S6c); they might therefore have no role in substrate recognition.

Using MS fragmentation analysis and ITC (where the substrate was sufficiently soluble), we sought to define the role of precursor peptide residues outside the core peptide in controlling substrate recognition. A synthetic peptide with the first 25 residues of PatE2 deleted (Δ25PatE2) retains the most conserved leader region, residues 26-35 (denoted “minimal leader”), and was processed by TruD only slightly more slowly than PatE2 (judged by MS; Figure S9). Insertion of up to six amino acids between the N-terminal protease signature, and the core peptide permitted normal heterocyclization (Figure S7, Table S1). Four further synthetic peptides were tested: Δ37PatE (retains five-residue protease signature) Δ42PatE (no leader), the eight-residue core peptide itself and the eight-residue core peptide with an additional C-terminal Gly (Figure S9). No reaction is seen with core peptide alone or core peptide with C-terminal Gly. The terminal cysteine only is however processed in both Δ37PatE and Δ42PatE peptides (more slowly than Δ25PatE but within an order of magnitude). From these data we conclude that recognition of the C-terminal cysteine of the cassette does not require the leader.

Radical mutants of individual residues within the center of the minimal leader region revealed S30F had no effect but mutants L29R and E31R did not allow processing of the PatE2 to completion. In both cases, we observed the mono-heterocyclized intermediate along with the product but no unreacted substrate, indicating that the rate of the second heterocyclization is significantly slower in these mutants but the rate of the first heterocyclization is largely unaffected. Both mutations significantly decreased binding affinity (Figure S7). Mutations G38I, L39N, A41I, and S42C within the GLEAS protease signature had little or no effect on processing (first and second heterocyclization; E40R was insoluble in our hands); although the binding was decreased (Figure S7). The “new” cysteine of S42C was not processed by TruD. Y33A and D54R (in the C-terminal macrocyclization signature) were processed normally. Processivity (heterocyclization of the second or internal cysteine of the core peptide), although very sensitive to changes in the minimal leader, is insensitive to changes in the core peptide flanking regions. We explored more radical mutants S42Q and A52D, which immediately flank the core peptide; both mutants significantly disrupted binding and processing. In contrast to changes in the leader peptide however, these mutations primarily produced either fully heterocyclized product or unreacted substrate. This suggests that these mutations affect the rate of the first heterocyclization (C-terminal cysteine), which has now become rate-limiting, but the mutations do not significantly perturb heterocyclization of the internal cysteine
The dimer partner is represented in corresponding faded colors.

Central have an insertion between the first and second strand of the ence; all the adenylating enzymes that share a fold with TruD over 142 C

Thiamine pyrophosphate biosynthetic enzyme ThiF (2.1 /C138)

found in the microcin C7 antibiotic production pathway and structural fold with the enzyme MccB, an adenylase molecule that exists as a dimer (Figure 4a).

The structure of TruD was determined to 2.9 Å (PDB 4BS9) and can be decomposed into three domains (domain 1 residues 2–85, domain 2 residues 86–321, domain 3 residues 323–781), which combine to form an extended, curved molecule that exists as a dimer (Figure 4a).

Domains 1 and 2 of TruD share 21% sequence identity and structural fold with the enzyme MccB, an adenylase found in the microcin C7 antibiotic production pathway (MccB catalyzes two successive adenylation reactions; Figure S10).[12] Pairwise superposition of TruD with the MccB structure gives a Ca rmsd of 5.4 Å over 219 structurally equivalent residues (MccB lacks the third domain; Figure 4b; Supporting Information, Figure S11); individual domain superpositions are much lower due to a domain rotation (2.6 Å over 80 residues and 2.0 Å over 152 residues). TruD and MccB possess a similar dimeric arrangement (Figure 4a) and both share two CXXC motifs in domain 2 that bind a Zn2+ ligated in a tetrahedral arrangement of sulfur atoms. Both motifs are conserved in cyanobactin heterocyclases and these motifs (and associated zinc) are a conserved feature of one class of adenylating enzymes. The coordinatively saturated zinc plays a structural role only[12b,13] in this class of adenylating enzymes.

Domain 2 of TruD shares the fold of this class of adenylating enzymes; other members include human ubiquitin-activating enzyme 5 (1.9 Å rmsd over 134 Cαs) and thiamine pyrophosphate biosynthetic enzyme ThiF (2.1 Å over 142 Cαs). There is, however, a subtle topology difference; all the adenylating enzymes that share a fold with TruD have an insertion between the first and second strand of the central β-sheet relative to TruD domain 2. Significantly this means that the ATP binding site of MccB is not fully conserved in TruD (Figure S11). The ATP-bound structure of MccB (PDB 3H5N) would suggest R123, K190, R221, and R253 of TruD as possible ATP binding residues. R123A retained native activity and the three other TruD mutants were insoluble. D165 in TruD is conserved both in heterocyclases and in the adenylase superfamily but the TruD mutant D165A retained native activity. Structural data of course do not establish the mechanism but they are entirely consistent with TruD being an adenylase enzyme. The 20% sequence identity of BalhC to TruD (including the two CXXC motifs; Figure S2), establishes that BalhC will have the same fold as domains 1 and 2 of TruD and thus its fold will belong to the same class of adenylylating enzymes.

Domain 3 of TruD is a novel fold and has a large negatively charged central cleft (Figure S11). In our hands the third domain of TruD or PatD, expressed separately, is insoluble, but by proteolysis of PatD (highly homologous to TruD) we were able to purify part of the third domain (residues 410–784). In contrast to full-length PatD, which is active, the fragment possessed no catalytic or substrate binding activity (Figure S12). BalhD[5] with 23% identity will have the same fold as TruD domain 3 and thus possesses an entirely new fold for kinase chemistry.

Biochemical analysis of TruD conclusively rules out a kinase mechanism for this enzyme as there is no ADP production. As both AMP and pyrophosphate are directly detected during turnover, the data can only be interpreted as an adenylylation-type mechanism. We did observe phosphate (rather than pyrophosphate) at higher temperature or after prolonged incubation, but this is consistent with degradation of the reactive pyrophosphate. Chemical reasoning would suggest the carbonyl oxygen of amides is too unreactive a nucleophile to react directly with ATP. In any event, such a mechanism would seem to invite futile ATP consumption (adenylation of carbonyl followed by hydrolysis), which is not observed. The most plausible mechanism therefore requires the formation of a five-membered hemiorthoamide ring in a reversible step (Scheme 2), proposed by Dunbar et al.[6] in the BalhD mechanism. The general base required for the first

![Figure 4. Structure of TruD. a) Structure of the biological TruD dimer. Domain 1 (blue), domain 2 (yellow), domain 3 (red), and the covalently bound Zn (gray sphere in domain 2, near domain 3) are shown. The dimer partner is represented in corresponding faded colors. b) Structural alignment of TruD (domain 1 blue, domain 2 yellow) with MccB (magenta, PDB 3H5A). The MccB ATP binding site is boxed.](image)

![Scheme 2. Proposed mechanism for heterocyclization by the TruD class of enzyme. The sulfhydryl proton is removed by the formation of a hemiorthoamide, which is adenylated. Irreversible elimination of AMP leads to the formation of the thiazoline. PPI = pyrophosphate.](image)
step could come from PatE but the observations that the entire leader region can be removed, and the conserved flanking residues extensively mutated argues against this. We suggest the hemithioamide intermediate attacks the \( \alpha \)-phosphate of ATP displacing pyrophosphate giving an adenylated hemithioamide. This highly reactive intermediate decomposes to thiazoline and by rapid irreversible elimination of AMP (Scheme 2).

The difference in mechanism of phosphorylation versus adenylation between BalhD and TruD is unexpected; especially given their sequence and by extension structural homology and their shared chemical reaction. We speculate that BalhD and the third domain of TruD both catalyze the reversible formation of the hemithioamide ring. In TruD, domain 2 then adenylates the intermediate whilst in the TOMM pathway, BalhC docks to BalhD and somehow accelerates the phosphorylation. These data suggest that heterocyclases join a very small number of enzymes exemplified by the type I and type II aldolases, which catalyze the same reaction, share a fold yet operate by entirely different chemical mechanisms.

TruD and its homologues have to process multiple (> 8) cysteines scattered throughout a larger precursor peptide. NMR showed TruD operates in strict order with the terminal cysteine in the core peptide reacting first; the next most C-terminal cysteine then reacts, and so on (Scheme 3), thus avoiding a combinatorial array of intermediates. The starting material binds tightly (80 nM) whilst the mono-heterocyclized intermediate (mimicked by PatE2 C51P) binds more weakly (\( K_D \) 1500 nm). Thus the intermediate only binds once all the initial substrate is processed. The requirement for a five-membered ring at the end of the core peptide for processing mirrors the situation for the macrocyclase PatG[14] but for the opposite reason. In the macrocyclase the ring is essential for binding of the substrate, but for the heterocyclase it is essential to prevent unproductive substrate binding at an already processed residue: a form of chemical check point (Scheme 3).

The entire leader peptide could be dispensed with and the terminal cysteine still processed (something similar has been reported in the TOMM pathway[7]). This suggests somewhat special recognition for the terminal cysteine focused on its immediate environment, consistent with our mutational analysis (Scheme 3). What is the role of the leader sequence? We have shown that the leader sequence (crucially residues 26–37) is required for processivity (heterocyclization of subsequent cysteines). We advance a model in which substrate recognition involves both the target residue and the leader, the contribution of each of these to overall substrate binding shifts during processing and it is this differential binding that controls the order of heterocyclization.

As multiple core peptides share a single leader, the signatures flanking the core peptide can be mutated extensively and insertion of residues between the core and leader can be tolerated; we conclude that the spatial relationship between the minimal leader and target residue is highly flexible. This separation of recognition and catalysis is rare and should help drive the use of the enzyme in chemical synthesis.

\[ \text{Keywords: adenylation · cyanobactins · heterocyclase · ribosomal peptide pathways} \]

Received: July 19, 2013
Revised: September 3, 2013
Published online: [link]


The Cyanobactin Heterocyclase Enzyme: A Processive Adenylase That Operates with a Defined Order of Reaction

Counting backwards: The cyanobactin class of heterocyclases, exemplified by TruD, possess an almost unique combination of processivity, specificity, chemical versatility, and promiscuity. TruD is shown by biochemical assay to be an adenylase, and processes cysteines in a defined order. The entire substrate leader can be removed and TruD will process a single specific cysteine residue; however the role of leader is to permit processivity through a balance of recognition. ATP/AMP = adenosine tri/monophosphate; PPI = pyrophosphate.