HHS Public Access

Author manuscript

Nat Chem Biol. Author manuscript; available in PMC 2012 December 01.

Published in final edited form as:

Nat Chem Biol. ; 8(6): 569–575. doi:10.1038/nchembio.944.

YcaO domains utilize ATP to activate amide backbones during peptide cyclodehydrations

Kyle L Dunbar[1,2], **Joel O Melby**[1,2], and **Douglas A Mitchell**[1,2,3,*]

[1]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801, USA

[2]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801, USA

[3]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801, USA

Abstract

Thiazole/oxazole-modified microcins (TOMMs) encompass a recently defined class of ribosomally synthesized natural products with a diverse set of biological activities. Although TOMM biosynthesis has been investigated for over a decade, the mechanism of heterocycle formation by the synthetase enzymes remains poorly understood. Using substrate analogs and isotopic labeling, we demonstrate that adenosine 5′-triphosphate (ATP) is utilized to directly phosphorylate the peptide amide backbone during TOMM heterocycle formation. Moreover, we present the first experimental evidence that the D-protein component of the heterocycle-forming synthetase (YcaO/DUF181 family member), formerly annotated as a docking/scaffolding protein involved in complex formation and regulation, is able to perform the ATP-dependent cyclodehydration reaction in the absence of the other TOMM biosynthetic proteins. Together, these data provide a greater level of detail into the biosynthesis of azol(in)e heterocycles in ribosomal natural products and prompt a reclassification of the enzymes involved in their installation.

TOMMs constitute a recently grouped class of peptidic natural products of ribosomal origin containing Cys-, Ser-, and Thr-derived azol(in)e heterocycles. The installation of these heterocycles onto a precursor peptide endows the molecule with conformational rigidity and, in all cases examined to date, is absolutely required for biological activity[1]. Since the

*Correspondence and requests for materials should be addressed to D.A.M. phone: 1-217-333-1345, fax: 1-217-333-0508, douglasm@illinois.edu.

Author contributions

Experiments were designed by D.A.M., K.L.D. and J.O.M and performed by K.L.D and J.O.M. The manuscript was written by D.A.M. and K.L.D. with critical editorial input from J.O.M. The study was conceived and overseen by D.A.M.

Competing financial interests

The authors declare no competing financial interests.

enzymatic machinery responsible for TOMM biosynthesis was first characterized[2], it has become evident that this strategy for natural product biosynthesis has been extensively propagated in both bacteria and archaea[1,3]. Furthermore, the diverse array of biological activities and pharmacological potential displayed by the characterized members of the TOMM family has led to extensive investigations focused on elucidating the molecular underpinnings of their biosynthesis[2,4–8]. These studies have provided the foundation to understanding the complex nature of substrate processing, but many of the finer details regarding heterocycle formation have remained elusive.

For example, studies performed on the TOMM enzymatic machinery involved in microcin B17 (Mcb, Fig. 1a) and azol(in)e-containing cyanobactin biosynthesis have demonstrated that azole heterocycles are installed through two steps: a cyclodehydration to generate an azoline heterocycle and a subsequent flavin mononucleotide (FMN)-dependent dehydrogenation to afford the aromatic azole (Fig. 1b)[2–4]. While the enzymatic complex responsible for these transformations has been partially characterized, the dehydrogenase (B-protein) is the only enzyme with a definitive function[6]. Dissection of the biochemical functions of the C- and D-proteins has been stymied by the inability to obtain individual activities for either protein; however, cyclodehydratase activity has been demonstrated for a complex of the two proteins[4,9]. Clearly, the C- and D-proteins act cooperatively and thus, it is perhaps not surprising that both proteins are found fused as a single polypeptide in roughly half of all known TOMM clusters[1,3].

Despite the challenges in separating the enzymatic activities of the C- and D-proteins, functional assignments have been postulated. The discovery that the C-protein of the microcin B17 synthetase (McbB) purified with a stoichiometric amount of Zn^{2+} led to the hypothesis that the metal was serving a Lewis acid to facilitate cyclodehydration[6]. While the Zn^{2+} was later established to serve only a structural role[10], additional support for continuing the assignment of the cyclodehydratase to the C-protein comes from sequence similarity to adenylating enzymes involved in microcin C7 (MccB), molybdopterin (MoeB), and thiamin (ThiF) biosynthesis[9]. As a result of this similarity, TOMM C-proteins are routinely misannotated in public databases as being involved in molybdopterin or thiamin cofactor biosynthesis. In contrast, the D-protein is homologous only to members of the uncharacterized YcaO/DUF181 (DUF, domain of unknown function) protein family. Lacking both a biochemically and bioinformatically identifiable function, the D-protein has been proposed to play a regulatory role in TOMM biosynthesis by either facilitating synthetase complex assembly and iterative substrate processing, or providing substrate recognition[3,6,8]. On this basis, the D-protein has previously been referred to as a docking/scaffolding protein[3]. Confusing matters further, the cyclodehydration reaction has been shown to require the hydrolysis of ATP to ADP and phosphate (P_i)[5,8], yet no characterized C-protein contains a bioinformatically recognizable nucleotide-binding site. Rather the D-protein from the microcin B17 synthetase has been implicated in ATP utilization because it contains sequences of weak resemblance to nucleotide binding motifs[8]. These residues are not conserved in other D-proteins, which casts doubt on their importance in TOMM biosynthesis.

The requirement for ATP hydrolysis with respect to azoline formation remains undescribed[5,8]. Reports of the super-stoichiometric hydrolysis of ATP relative to azole formation, and continued ATP hydrolysis after complete substrate modification, have been used as evidence for the role of ATP as a dynamic regulator of the cyclodehydratase complex, yet direct evidence for such a mechanism has not been reported[5,8]. Recently, this role for ATP utilization has been referred to as the "molecular machine" hypothesis (Fig. 1c)[5]. An alternative mechanism implicates ATP hydrolysis in the activation of the peptide backbone during cyclization[2,5]. Such an activation event would result in the elimination of P_i instead of water during the cyclization, and could potentially help drive the cyclodehydration reaction towards completion. Although isotopic labeling studies could readily differentiate between these mechanisms, the relative experimental intractability of previously characterized TOMM synthetases has precluded such a study[2,3,6]. The discovery of a TOMM biosynthetic cluster from *Bacillus* sp. Al Hakam (Balh; Fig. 1a) yielded robust heterocycle-forming enzymes for *in vitro* characterization[11] and allowed us to address two of the foremost questions regarding TOMM biosynthesis: (1) the role of ATP in the mechanism of heterocycle formation and (2) the individual roles of the C- and D-proteins in the synthetase complex. Our findings demonstrate that the YcaO/DUF181 family member, present in all TOMM clusters utilizes ATP to phosphorylate the amide backbone of peptide substrates during the cyclodehydration reaction. As such, the data reported herein provide an improved framework to understand the factors that govern biosynthesis in this class of natural products that has garnered significant attention[12–19].

RESULTS

Minimal requirements for cyclodehydratase activity

In order to determine the minimal set of proteins required for azoline formation, reactions were initiated with one or more reaction components omitted with progress monitored by matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS, Fig. 2a). Samples lacking ATP, BalhC or BalhD did not display a detectable level of BalhA1 substrate (sequence given in Fig. 1a) modification under the conditions employed. When the BalhA1 substrate was treated with both BalhC and BalhD, the majority of the peptide was converted to a penta-azoline species (−90 Da). Substrate treated with the full BCD synthetase was converted to a penta-azole species (−100 Da) as previously reported[11]. In congruence with the microcin B17 and cyanobactin synthetases, the Balh synthetase also hydrolyzed ATP to ADP and P_i during substrate processing (Supplementary Results, Supplementary Fig. 1)[2,5]. Subsequent studies demonstrated that the Balh enzymatic machinery was able to utilize both ATP and GTP (also analogous to the microcin B17 synthetase)[8], although ATP was preferred over GTP by approximately a factor of 10 due to differences mainly in K_m (Supplementary Fig. 2). Substitution of ATP with a non-hydrolyzable analog resulted in no ring formation, providing further evidence of the requirement of nucleotide triphosphate hydrolysis during heterocycle formation (Supplementary Fig. 3). The rate at which P_i was generated under varying reaction conditions was monitored by a purine nucleoside phosphorylase (PNP)-coupled assay[20]. Robust P_i generation was observed only when the C- and D-proteins were added to samples containing BalhA1 and ATP (Supplementary Fig. 4). Addition of the B-protein did not

significantly alter the rate of ATP hydrolysis, which is consistent with earlier reports of C–D fusion enzymes forming azoline heterocycles in the absence of the dehydrogenase[4]. In contrast to the Balh and the C–D fusion enzymes, the microcin B17 synthetase was not functional in the absence of the dehydrogenase[6]. The reason for this discrepancy is currently unknown but the available data suggest that the Mcb dehydrogenase may play a pivotal role in the assembly of an active synthetase complex.

A cyclizable residue stimulates rapid ATP hydrolysis

The "molecular machine" hypothesis asserts that ATP hydrolysis regulates the conformational dynamics/enzymatic activity of the synthetase complex[5]. While the mechanism for how this may work has not yet been explicitly stated, one possible route could be that upon peptide substrate binding, the synthetase may undergo a conformational change that leads to an active complex (Fig. 1c). Previous studies on microcin B17 biosynthesis have demonstrated that the leader sequence of the precursor peptide, in addition to a severely truncated (non-processed) substrate analog, were incapable of stimulating ATP hydrolysis[8]. These results suggested that the core sequence of TOMM precursor peptides must be important for the ATPase activity of the synthetase. Studies conducted in the absence of BalhA1 demonstrated that a complex of BalhC and BalhD had a measureable level of ATP-hydrolysis; roughly 100-fold lower than at saturating BalhA1 concentrations (Supplementary Fig. 5a). To test if a non-cyclizable (NC), full-length substrate analog could potentiate this basal ATPase activity of the TOMM enzymatic complex, we prepared BalhA1-NC, in which every heterocyclizable residue in the core region was mutated to a hydrophobic amino acid (Fig. 1a). Interestingly, the presence of BalhA1-NC decreased the rate of ATP hydrolysis four-fold compared to samples containing only BalhC and BalhD (Supplementary Fig. 5a). Addition of the predicted BalhA1 leader sequence (BalhA1-LS, Fig. 1a) to similar reactions showed a slight, but statistically significant, reduction in the basal ATPase activity. The marked difference between the levels of inhibition provided evidence that the core sequence was primarily responsible for suppressing the ATPase activity and indicated that there may be a direct interaction between residues of the core sequence and the ATP binding site. In support of this, our previous study demonstrated that the reinstatement of a naturally occurring cysteine (BalhNC-C40, Fig. 1a) yielded a substrate that was not only processed by the synthetase, but increased the rate of ATP hydrolysis close to wild-type levels [11], as shown again here (Supplementary Fig. 5). These data provided substantial evidence that the ATP binding site and the residues of the core sequence directly communicate. Moreover, the ability of BalhA1-NC to inhibit the basal ATPase activity suggested that the ATP site had greater accessibility in the absence of the precursor peptide. While not definitive, these results are difficult to reconcile with respect to an ATP-utilizing mechanism consistent with the molecular machine hypothesis, which prompted us to explore alternative mechanistic possibilities.

ATP hydrolysis is tightly coupled to azoline formation

The microcin B17 synthetase has been shown to consume approximately 5 moles of ATP per mole of heterocycle formed[8]. Although the super-stoichiometric consumption of ATP has not been observed for all TOMM biosynthetic enzymes, this observation casted doubt on the use of ATP to directly activate the substrate during processing and has been used as

supportive evidence of the molecular machine[5,8]. To evaluate the stoichiometry of products formed by the Balh synthetase, the rate of P_i and azole production were monitored in parallel time course assays. An overlay of P_i and azole synthesis illustrated that the formation of these products was tightly coupled during the reaction, and that the ATP to azole stoichiometry was essentially one–to–one over the entire six-hour time course (Fig. 2b). Furthermore, when BalhA1 was fully heterocyclized, the rate of ATP hydrolysis returned to approximately the basal rate (Fig. 2b). This result was consistent with an earlier report on trunkamide biosynthesis, which employs a C–D fusion protein[5].

Given the markedly different ATP/azole stoichiometry in the Balh and Mcb systems, we revisited the ATP stoichiometry of microcin B17 processing. We hypothesized that the use of a truncated substrate and a synthetase complex with an abnormal ratio of the B-, C- and D-proteins (due to the purification strategy) could have led to an aberrant usage of ATP[8,12]. Therefore, the Mcb ATP stoichiometry was repeated using the full-length substrate (McbA; Fig. 1a) and a completely tag-free synthetase complex added in at a 1:1:1 molar ratio. As with the Balh synthetase, the concentration of P_i and azole heterocycles mapped almost perfectly over the time course assayed and an average stoichiometric ratio was calculated to be unity (Fig. 2c). Taken together, these data demonstrated that ATP hydrolysis is tightly linked to heterocycle formation, which appears to be a general feature of TOMM heterocycle formation.

ATP is hydrolyzed by a non-solvent oxygen nucleophile

Intrigued by the inability to artificially activate ATP hydrolysis and the 1:1 stoichiometry between ATP consumption and ring formation, isotopic labeling was used to determine the source of the oxygen in the P_i generated during the course of the reaction. If ATP were used to directly activate the substrate, the oxygen incorporated into the P_i byproduct would originate from the peptide backbone. Thus, running the reaction in [^{18}O]-H_2O would result in ^{16}O incorporation into phosphate, producing [$^{16}O_4$]-P_i (Fig. 1c). Alternatively, the molecular machine hypothesis demands that the oxygen incorporated into the P_i would originate from bulk solvent (water) in the same manner as with any ATPase (Fig. 1c). The molecular machine mechanism, if acting, would introduce a single ^{18}O label in the P_i produced when the reaction was run in [^{18}O]-H_2O, which is readily detectable by a roughly 0.02 ppm upfield shift in ^{31}P-NMR compared to [$^{16}O_4$]-P_i (Ref 22).

Reactions of the Balh TOMM biosynthetic enzymes with the BalhA1 substrate in [^{18}O]-H_2O resulted in the production of primarily [$^{16}O_4$]-P_i (Fig. 3a). As would be predicted from the earlier stoichiometry assay (Fig. 2b) and the basal rate of ATP hydrolysis (Supplementary Fig. 5a), these samples also contained a small amount (~15%) of [$^{18}O_1{}^{16}O_3$]-P_i. The majority production of [$^{16}O_4$]-P_i was a consequence of substrate processing and was neither a contaminant of the ATP buffer, nor the proteinaceous components of the synthetase reaction as assessed by the malachite green assay. While the direct activation of the amide carbonyl is the most logical explanation for this data, the possibility existed that [$^{16}O_4$]-P_i was generated as a consequence of nucleophilic attack at the β-phosphate of ATP or through isotope scrambling with the ADP byproduct. Thus, a liquid chromatography (LC)-MS analysis of the ADP formed during the reaction was carried out to determine if this

byproduct contained ^{18}O. The trace clearly demonstrated that ^{18}O was not present in the ADP and provided strong evidence that ATP consumption occurred due to the action of a non-solvent, oxygen nucleophile (Supplementary Fig. 6). This assay was repeated with the microcin B17 synthetase, which yielded the same result (Fig. 3b), though the requirement for a lower pH buffer gave a slight perturbation in the ^{31} P-phosphate chemical shift.

Our data have implicated ATP in the direct phosphorylation of the amide carbonyl oxygen preceding the residue undergoing cyclization and also have invalidated the molecular machine hypothesis as the major pathway for heterocycle formation. Furthermore, these results demonstrated that this mechanism of substrate modification is shared among distantly related TOMM synthetases (Fig. 1a).

The D-protein alone is sufficient for azoline formation

With the elucidation of the role of ATP in heterocycle formation, we sought to more precisely assign functions to the individual BalhC and BalhD proteins, which collectively comprise the cyclodehydratase (Fig. 2a, Supplementary Fig. 4). Given the homology between the C-protein, previously labeled as the cyclodehydratase5, and MccB, an adenylating protein involved in microcin C7 biosynthesis (Supplementary Fig. 7)[23], we hypothesized that BalhC catalyzed the direct activation of the peptidic substrate. BalhD was suspected to be involved with general substrate handling and enzymatic regulation, thus allowing BalhC to perform multiple turnovers. To test these putative assignments, we performed a pseudo-single turnover experiment with a stoichiometric amount of either BalhC or BalhD and the BalhA1 substrate, under saturating ATP. Unexpectedly, no modifications were detected when BalhC alone was added to BalhA1, but up to 4 dehydrations (−90 Da) were observed on BalhA1 treated with BalhD alone (Fig. 3c). Subsequent studies showed that these modifications were ATP-dependent and competent intermediates in the production of the penta-azole product (Supplementary Fig. 8). Similar reactions on substrates containing a reduced number of heterocyclizable residues (BalhNC-C40 and BalhNC-CCG*; Fig. 1a), subjected to either tandem MS spectroscopy or iodoacetamide labeling, proved that the BalhD-installed modifications were cyclodehydration products (to yield an azoline) and not dehydrations (to yield a dehydroalanine/dehydrobutyrine; Supplementary Fig. 9, 10). These data confirmed that BalhD, formerly assigned as a docking/scaffold protein, was necessary and sufficient for azoline formation. Importantly, the rate of BalhD-only catalyzed heterocyclization was nearly three orders of magnitude slower than when BalhC was present (Supplementary Fig. 11). Analysis of the BalhD-only catalyzed reaction products revealed that the ATP/azoline stoichiometric ratio significantly deviated from unity (Supplementary Fig. 12). Thus, we recommend that the term cyclodehydratase should be reserved for describing the C-D complex.

The D-protein utilizes ATP for amide backbone activation

The discovery that BalhD could independently install azolines on BalhA1 in an ATP-dependent manner, but that ATP consumption was dysregulated in the absence of BalhC, led us to investigate if BalhD could independently activate the amide backbone of BalhA1. As before, a reaction was carried out in [^{18}O]-H_2O and the $^{16}O/^{18}O$ isotopic ratio in the

resultant P_i was analyzed by ^{31}P-NMR spectroscopy. Samples containing BalhD, BalhA1 and ATP showed a significant [$^{16}O_4$]-P_i peak (Fig. 3d). However, if any of these essential components were omitted, [$^{16}O_4$]-P_i was not detectable (Supplementary Fig. 13). While these data clearly demonstrated that BalhD was responsible for the amide activation event, the nearly identical heights of the [$^{16}O_4$]-P_i and [$^{18}O_1{}^{16}O_3$]-P_i NMR signals provided further evidence that ATP consumption was dysregulated in the absence of BalhC (Fig. 3d and Supplementary Fig. 12). Unlike the BalhC-BalhD complex, which had a strong enhancement in the rate of ATP hydrolysis upon the addition of BalhA1, the ATPase activity of BalhD remained unchanged upon the addition of BalhA1 (Supplementary Fig. 11). These findings implicate the C-protein as a potentiator of the D-protein, which is the stark opposite of the original assignments.

DISCUSSION

Prior to this study, the utilization of ATP hydrolysis for cyclodehydration was poorly understood. Although it had been noted that the initial steps of this transformation could mimic the mechanism of protein splicing by intein domains (an ATP-independent process; Fig. 4), the absolute requirement for nucleotide triphosphate hydrolysis had never definitively been factored into the reaction[2,5]. Our findings make the TOMM D-proteins only the second enzyme class known that can utilize ATP to directly activate an amide carbonyl oxygen, the other being the PurM superfamily. PurM family members conduct similar chemistry on non-peptide substrates (Supplementary Fig. 14)[24,25] and in cases where crystal structures have been solved, all contain a non-canonical ATP-binding domain[26,28]. Despite the similarity in ATP usage between TOMM D-proteins and PurM family members, the two enzyme families share no sequence similarity and represent an example of convergent evolution towards a common mode of ATP utilization. Members of the PurM family are proposed to act via the formation of an iminophosphate intermediate[26], whereby phosphorylation occurs prior to nucleophilic attack (Supplementary Fig. 14). While a similar mechanism can be drawn for azoline formation by the TOMM synthetase (Supplementary Fig. 15), it would require a disfavored 5-endo-trig cyclization[29] and is inconsistent with the suppression of ATP hydrolysis in the presence of BalhA1-NC. Alternatively, we hypothesize that azoline formation by the TOMM synthetase occurs through a hemi-orthoamide intermediate analogous to that implicated in protein autoproteolytic pathways[30,31]. During intein splicing and other autoproteolytic events, the hemi-orthoamide is resolved by *N*-protonation and (thio)ester formation (Fig. 4)[32]. We assert that the phosphorylation of the amide oxygen (in lieu of *O*-protonation) would direct the hemi-orthoamide towards azoline formation and prevent the non-productive breakdown of the intermediate (Fig. 4). Phosphorylation would not only accelerate the elimination reaction (based on the lower pK_a of phosphate relative to water), but would directly couple ATP hydrolysis to cyclodehydration, providing a thermodynamic drive for the reaction. The resolution of the hemi-orthoamide via the inclusion of a thermodynamically favorable step is seen in all autoproteolytic pathways[32]. Further evidence supporting a mechanism where cyclization precedes phosphorylation (*i.e.* the intein-like mechanism) comes from a recent report of ester formation during microcin B17 biosynthesis[33] and the discovery that engineered intein domains can catalyze azoline formation[34]. In light of earlier reports, and

the data presented here, TOMM cyclodehydration and intein splicing proceed through a common intermediate, which we propose is the hemi-orthoamide depicted in Figure 4.

While previous studies on fused C–D proteins, implicated the YcaO/DUF181 domain in the cyclodehydration reaction, the exact role of the D-domain in this complex was never elucidated. Besides the TOMM D-protein, only one other YcaO/DUF181 family member has had a biological function reported. A recent study linked the function of *E. coli* YcaO to ribosomal thiomethylation, but did not reveal any mechanistic details or physiological ramifications of this process[35]. Thus, our data assign the first definitive function to a member of this uncharacterized protein family. As such, the YcaO/DUF181 protein that exists in all TOMM biosynthetic clusters has the potential to provide insight into the activity of an estimated 3000 conserved/hypothetical genes that remain largely unannotated in GenBank. At the time of writing, approximately 500 of these lie in bioinformatically recognizable TOMM clusters, all of which are likely to perform the reaction described in this report[1,3]. While we cannot confidently comment on the function of the remaining 2500 genes, it is possible that non-TOMM members of the YcaO/DUF181 families harbor ATP/GTP-binding sites that either facilitate or otherwise control enzymatic activity.

Our dissection of the cyclodehydratase complex and reclassification of the roles the C/D-proteins posed an interesting question as to the role of the C-protein in TOMM biosynthesis. While we have shown that BalhC acts cooperatively with BalhD to accelerate azoline formation and govern the proper utilization of ATP, this study does not allow us to conclusively deduce the role of BalhC. Nonetheless, it is conceivable that the C-protein could partake in one of three plausible functions: (1) BalhC activates BalhD through an allosteric mechanism, increasing the rate of heterocyclization, (2) BalhC catalyzes the nucleophilic attack of the proceeding side chain by providing the requisite general acid/base residues, or (3) BalhC forms key contacts with both the substrate and BalhD to facilitate the interaction of the core region of BalhA1 and the BalhD active site. Our data do not distinguish between these possibilities, which are not necessarily mutually exclusive.

In summary, the results reported herein have not only elucidated the role of ATP in the heterocyclization of TOMM substrates, but have also implicated the hydrolysis of this co-substrate in the activation of the peptide backbone during substrate processing. Furthermore, we have obtained evidence demonstrating that the D-protein is solely responsible for this ATP-dependent cyclodehydration reaction and, as such, have assigned an enzymatic function to the previously enigmatic YcaO/DUF181 protein family. These discoveries pave the way for future work focused on uncovering the complex quaternary interactions involved in regulation of cyclodehydratase activity and the characterization of the catalytic architecture for this poorly understood ATP-utilizing enzyme.

METHODS

Protein overexpression and purification

All proteins were overexpressed as tobacco etch virus (TEV) protease-cleavable fusions to maltose binding protein (MBP) and purified by amylose affinity chromatography by methodology similar to that previously reported[3]. A detailed protocol can also be found in

the supplementary methods. Unless otherwise stated, all proteins and substrates were used as MBP-fusions. The Balh TOMM synthetase proteins are highly active while MBP-tagged and largely insoluble when untagged[11].

Purine nucleotide phosphorylase-based phosphate detection assay

In general, the indicated synthetase proteins were added to a cuvette for a final concentration of 1 μM each. Reactions were initiated via the addition of a room temperature mixture of peptide substrate, 200 μM 2-amino-6-mercapto-7-methylpurine riboside (Berry and Associates), and 0.2 units of PNP in synthetase buffer [50 mM Tris pH 7.5, 125 mM NaCl, 20 mM $MgCl_2$, 10 mM dithiothreitol (DTT) and 3 mM ATP]. Reaction progress was monitored by the change in absorbance at 360 nm on a Cary 4000 UV-Vis spectrophotometer (Agilent). Initial rates of phosphate production were calculated based on the linear absorbance change during the first 3 min of the reaction and the extinction coefficient of the resulting guanine analog (11,000 $M^{-1}cm^{-1}$). For determining the nucleotide triphosphate kinetic constants, a synthetase buffer was used that lacked ATP, and the reactions were initiated by the addition of the indicated nucleotide. Given the slow rate of ATP hydrolysis without the presence the precursor peptide, the background ATPase activity was measured using 15 μM MBP-tagged BalhC/D instead of 1 μM. Reactions were carried out as above ($n \geq 3$). Regression analyses to obtain the kinetic parameters were carried out using IGOR Pro version 6.12.

Subtractive activity studies

The effect of each protein on heterocyclization was monitored by MALDI-TOF-MS. Reactions were carried out for 16 h at 23 °C with 1 μM of the indicated components, 50 μM MBP-BalhA1 and the synthetase buffer listed above. MBP was proteolytically removed from the substrate by the addition of 2 μg/ml recombinant TEV protease and a 30 min incubation at 30 °C. Samples were desalted via C_{18} ZipTip (Millipore) according to the manufacturer's instructions and analyzed on a Bruker Daltonics UltrafleXtreme MALDI-TOF spectrometer. Spectra were obtained in positive reflector mode using α-cyano-hydroxycinnamic acid as the matrix.

Substrate analog assays

Inhibition studies were performed with 15 μM MBP-BalhCD and either 50 μM MBP-BalhA1-NC or MBP-BalhA1-LS. Rates were measured with the PNP assay as described above following a 15 min equilibration at 23 °C. MBP-BalhNC-C40 reactions were carried out with 1 μM MBP-tagged BcerB, BalhC, BalhD and 100 μM substrate for 16 h at 23 °C. Products were analyzed via MALDI-TOF-MS as described above.

Stoichiometry of ATP utilization

Samples were set up with 2 μM MBP-BalhC/D, 10 μM MBP-BcerB and 50 μM MBP-BalhA1 in synthetase buffer. Aliquots were removed at the indicated time points and frozen in liquid nitrogen. The samples were analyzed by LCMS to detect azole heterocycles and a malachite green assay to detect phosphate (Supplementary methods). The Mcb reactions were carried out in a similar fashion with the following exception: samples containing 1 μM

MBP-tagged McbBCD and 20 μM MBP-McbA were digested with thrombin for 4 h before the reaction was initiated with the addition of ATP. The background ATPase activities of both synthetase complexes were analyzed using both the PNP and malachite green phosphate assays.

^{31}P-NMR analysis of phosphate isotope incorporation

Two samples containing either the MBP-tagged BCD (4 μM) synthetase components or MBP-BalhA1 substrate (100 μM) with low-salt synthetase buffer (50 mM Tris pH 8.5, 25 mM NaCl, 5 mM $MgCl_2$, 10 mM DTT, 2 mM ATP) were lyophilized separately for 16 h. The resulting solid was reconstituted in 500 μL of 97 atom% [^{18}O]-H_2O (Cambridge Isotope Laboratories) and incubated at 23 °C with rocking for 8.5 h before quenching the reaction with 5.5 μL of 500 mM ethylenediaminetetraacetic acid (EDTA; pH 8.0). 50 μL of D_2O was added and the sample was transferred to a standard 5 mm NMR tube. The ^{31}P-NMR spectrum was obtained on a 600 MHz Varian Unity Inova NMR with a 5 mm Varian AutoTuneX probe, 512 transients, 32768 points, and a spectral window of −15 to 5 ppm.

The BalhD only ^{31}P-NMR experiment was set up in a similar manner except that the protein concentration was 20 μM and the substrate concentration was 60 μM (to partly circumvent the exceedingly slow reaction and dysregulated ATP usage in the absence of BalhC). After a 3 h incubation at 23 °C, the reaction was quenched by the addition of 25% acetonitrile to precipitate the protein components and dried via Savant SpeedVac (Thermo Fisher). The sample was reconstituted in 220 μL D_2O with 5 mM EDTA and placed in a D_2O-matched Shigemi NMR tube. ^{31}P-NMR analysis was carried out as before with 1000 transients.

The MBP-tagged McbBCD proteins were pooled at a 1:1:1 molar ratio and concentrated to 100 μM using a 3 kDa amicon centrifugal filter (Millipore). To this sample, 0.2 μg of thrombin (from bovine plasma) was added and the sample was proteolytically digested for 4 h at 23 °C to remove the MBP tags. A 600 μL sample of 20 μM MBP-McbA and low-salt synthetase buffer (pH 8.0) was lyophilized for 16 h. The lyophilized protein was reconstituted in 600 μL of [^{18}O]-H_2O and the above mixture of McbBCD was added to a final concentration of 5 μM to initiate the reaction. After a 2 h incubation at 23 °C, the sample was frozen in liquid nitrogen and lyophilized to dryness. The sample was reconstituted in 220 μL of D_2O with 15 mM EDTA and placed in a D_2O-matched Shigemi NMR tube. Analysis by ^{31}P-NMR was carried out as before with 2000 transients. Conditions describing the pertinent controls can be found in the supplementary methods.

Single-turnover activity studies

Individual MBP-tagged synthetase components (45 μM) were set up with MBP-BalhA1 substrate at a 1:1 ratio in synthetase buffer. After an 18 h reaction at 23 °C, samples were analyzed by MALDI-TOF-MS as described above.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Wilfred van der Donk for technical advice and for suggestions in the preparation of the manuscript. We thank members of Mitchell lab and Michael Marletta for critical review of this manuscript. This work was supported in part by the institutional funds provided by the University of Illinois and the National Institutes of Health (1R01 GM097142 to D.A.M). D.A.M. is also the recipient of the NIH Director's New Innovator Award (DP2 OD008463). Information on the New Innovator Award Program is at http://nihroadmap.nih.gov/newinnovator/. K.L.D. was supported by the NIH Training Program in the Chemistry –Biology Interface (2T32 GM070421). J.O.M. was supported by the University of Illinois Department of Chemistry Chinoree T. Kimiyo Enta Fellowship.

References

1. Melby JO, Nard NJ, Mitchell DA. Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. Curr Opin Chem Biol. 2011; 15:369–378. [PubMed: 21429787]
2. Li YM, Milne JC, Madison LL, Kolter R, Walsh CT. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. Science. 1996; 274:1188–1193. [PubMed: 8895467]
3. Lee SW, et al. Discovery of a widely distributed toxin biosynthetic gene cluster. Proc Natl Acad Sci USA. 2008; 105:5879–5884. [PubMed: 18375757]
4. McIntosh JA, Donia MS, Schmidt EW. Insights into heterocyclization from two highly similar enzymes. J Am Chem Soc. 2010; 132:4089–4091. [PubMed: 20210311]
5. McIntosh JA, Schmidt EW. Marine molecular machines: heterocyclization in cyanobactin biosynthesis. Chembiochem. 2010; 11:1413–1421. [PubMed: 20540059]
6. Milne JC, et al. Cofactor requirements and reconstitution of microcin B17 synthetase: a multienzyme complex that catalyzes the formation of oxazoles and thiazoles in the antibiotic microcin B17. Biochemistry. 1999; 38:4768–4781. [PubMed: 10200165]
7. Kelleher NL, Hendrickson CL, Walsh CT. Posttranslational heterocyclization of cysteine and serine residues in the antibiotic microcin B17: distributivity and directionality. Biochemistry. 1999; 38:15623–15630. [PubMed: 10569947]
8. Milne JC, Eliot AC, Kelleher NL, Walsh CT. ATP/GTP hydrolysis is required for oxazole and thiazole biosynthesis in the peptide antibiotic microcin B17. Biochemistry. 1998; 37:13250–13261. [PubMed: 9748332]
9. 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 USA. 2005; 102:7315–7320. [PubMed: 15883371]
10. Zamble DB, McClure CP, Penner-Hahn JE, Walsh CT. The McbB component of microcin B17 synthetase is a zinc metalloprotein. Biochemistry. 2000; 39:16190–16199. [PubMed: 11123948]
11. Melby JO, Dunbar KL, Mitchell DA. Selectivity, directionality, and promiscuity in peptide processing from a *Bacillus* sp. Al Hakam cyclodehydratase. J Am Chem Soc. 2012 In Press. 10.1021/ja211675n
12. Yu Y, et al. Nosiheptide biosynthesis featuring a unique indole side ring formation on the characteristic thiopeptide framework. ACS Chem Biol. 2009; 4:855–864. [PubMed: 19678698]
13. Wieland Brown LC, Acker MG, Clardy J, Walsh CT, Fischbach MA. Thirteen posttranslational modifications convert a 14-residue peptide into the antibiotic thiocillin. Proc Natl Acad Sci USA. 2009; 106:2549–2553. [PubMed: 19196969]
14. Morris RP, et al. Ribosomally synthesized thiopeptide antibiotics targeting elongation factor Tu. J Am Chem Soc. 2009; 131:5946–5955. [PubMed: 19338336]
15. Liao R, et al. Thiopeptide biosynthesis featuring ribosomally synthesized precursor peptides and conserved posttranslational modifications. Chem Biol. 2009; 16:141–147. [PubMed: 19246004]
16. Kelly WL, Pan L, Li C. Thiostrepton biosynthesis: prototype for a new family of bacteriocins. J Am Chem Soc. 2009; 131:4327–4334. [PubMed: 19265401]
17. Myers CL, Hang PC, Ng G, Yuen J, Honek JF. Semi-synthetic analogues of thiostrepton delimit the critical nature of tail region modifications in the control of protein biosynthesis and antibacterial activity. Bioorg Med Chem. 2010; 18:4231–4237. [PubMed: 20510619]

18. Bowers AA, Acker MG, Koglin A, Walsh CT. Manipulation of thiocillin variants by prepeptide gene replacement: structure, conformation, and activity of heterocycle substitution mutants. J Am Chem Soc. 2010; 132:7519–7527. [PubMed: 20455532]
19. Molohon KJ, et al. Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics. ACS Chem Biol. 2011; 6:1307–1313. [PubMed: 21950656]
20. Webb MR. A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. Proc Natl Acad Sci USA. 1992; 89:4884–4887. [PubMed: 1534409]
21. Roy RS, Belshaw PJ, Walsh CT. Mutational analysis of posttranslational heterocycle biosynthesis in the gyrase inhibitor microcin B17: distance dependence from propeptide and tolerance for substitution in a GSCG cyclizable sequence. Biochemistry. 1998; 37:4125–4136. [PubMed: 9521734]
22. Cohn M, Hu A. Isotopic (18O) shift in 31P nuclear magnetic resonance applied to a study of enzyme-catalyzed phosphate-phosphate exchange and phosphate (oxygen)-water exchange reactions. Proc Natl Acad Sci USA. 1978; 75:200–203. [PubMed: 203929]
23. Roush RF, Nolan EM, Lohr F, Walsh CT. Maturation of an *Escherichia coli* ribosomal peptide antibiotic by ATP-consuming N-P bond formation in microcin C7. J Am Chem Soc. 2008; 130:3603–3609. [PubMed: 18290647]
24. Schrimsher JL, Schendel FJ, Stubbe J, Smith JM. Purification and characterization of aminoimidazole ribonucleotide synthetase from *Escherichia coli*. Biochemistry. 1986; 25:4366–4371. [PubMed: 3530323]
25. Schendel FJ, Mueller E, Stubbe J, Shiau A, Smith JM. Formylglycinamide ribonucleotide synthetase from *Escherichia coli*: cloning, sequencing, overproduction, isolation, and characterization. Biochemistry. 1989; 28:2459–2471. [PubMed: 2659070]
26. Zhang Y, Morar M, Ealick SE. Structural biology of the purine biosynthetic pathway. Cell Mol Life Sci. 2008; 65:3699–3724. [PubMed: 18712276]
27. Li C, Kappock TJ, Stubbe J, Weaver TM, Ealick SE. X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the *Escherichia coli* purine biosynthetic pathway at 2.5 A resolution. Structure. 1999; 7:1155–1166. [PubMed: 10508786]
28. Morar M, Anand R, Hoskins AA, Stubbe J, Ealick SE. Complexed structures of formylglycinamide ribonucleotide amidotransferase from *Thermotoga maritima* describe a novel ATP binding protein superfamily. Biochemistry. 2006; 45:14880–14895. [PubMed: 17154526]
29. Baldwin JE. Rules for ring-closure. J Chem Soc Chem Comm. 1976:734–736.
30. Chong S, et al. Protein splicing involving the *Saccharomyces cerevisiae* VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. J Biol Chem. 1996; 271:22159–22168. [PubMed: 8703028]
31. Lee JJ, et al. Autoproteolysis in hedgehog protein biogenesis. Science. 1994; 266:1528–1537. [PubMed: 7985023]
32. Perler FB, Xu MQ, Paulus H. Protein splicing and autoproteolysis mechanisms. Curr Opin Chem Biol. 1997; 1:292–299. [PubMed: 9667864]
33. Ghilarov D, Serebryakova M, Shkundina I, Severinov K. A Major portion of DNA gyrase inhibitor microcin B17 undergoes an N,O-peptidyl shift during synthesis. J Biol Chem. 2011; 286:26308–26318. [PubMed: 21628468]
34. Ludwig C, Schwarzer D, Mootz HD. Interaction studies and alanine scanning analysis of a semi-synthetic split intein reveal thiazoline ring formation from an intermediate of the protein splicing reaction. J Biol Chem. 2008; 283:25264–25272. [PubMed: 18625708]
35. Strader MB, et al. A proteomic and transcriptomic approach reveals new insight into beta-methylthiolation of *Escherichia coli* ribosomal protein S12. Mol Cell Proteomics. 2011; 10:M110–005199. [PubMed: 21169565]

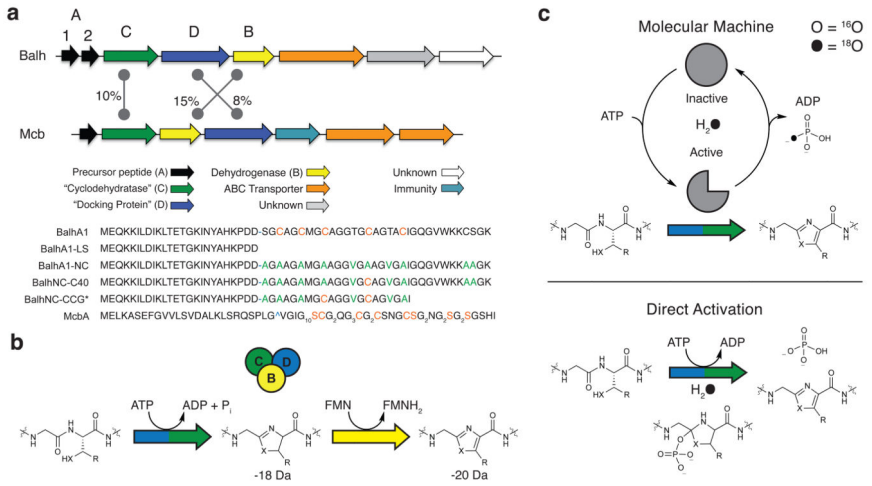

Figure 1. TOMM biosynthetic gene clusters and possible mechanisms of ATP utilization during azole formation

(a) The TOMM clusters from *Bacillus* **sp.** Al Hakam (Balh) and *Escherichia coli* (Mcb) are depicted along with the percent amino acid identity for each of the three key proteins. Gene assignments are given below. Note that the dehydrogenase and the "cyclodehydratase" in the microcin B17 cluster are assigned as McbC and McbB, respectively. The sequences of the peptide substrates used in this study are shown. Color-coding: green, point mutations; orange, residues known to be cyclized *in vitro*[11] blue hyphen, putative leader sequence cleavage site; blue caret, known leader sequence cleavage site[2]. **(b)** Heterocycles are installed on a precursor peptide by a heterotrimeric complex composed of a "cyclodehydratase" (C), a "docking/scaffolding" protein (D) and a dehydrogenase (B). The cumulative mass change for each step is shown below the modification. **(c)** The two leading hypotheses for ATP utilization. ATP hydrolysis could be used to control conformational dynamics (molecular machine) or to directly activate the peptidic substrate as shown. Reactions carried out in [^{18}O]-H_2O will give different products, allowing for distinguishing between these mechanisms. X = S, O; R = H, CH_3.

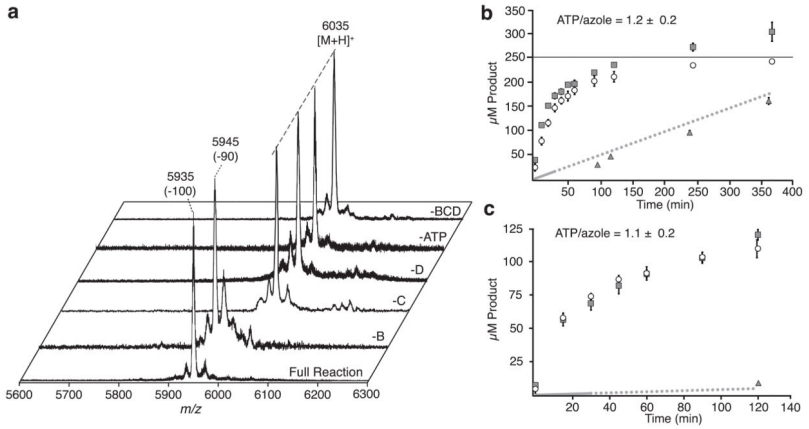

Figure 2. Minimal requirement for cyclodehydratase activity and ATP/azole stoichiometry for substrate processing

(a) MALDI-TOF spectra monitoring heterocycle formation. Labels represent the component omitted from the reaction (*e.g.* –C lacked BalhC). The mass shift for the major species relative to the unmodified substrate is displayed above the appropriate peak. The stoichiometry of ATP hydrolysis to azole heterocycle formation in **(b)** the Balh TOMM and **(c)** the Mcb TOMM is shown. Phosphate (squares), heterocycles (circles), BCD background ATPase activity (solid line, PNP assay; dashed line, extrapolated PNP assay; triangles, malachite green assay), maximum product produced based on the concentration of substrate (solid horizontal line). Error bars represent the standard deviation from the mean (n=3). The ATP/azole stoichiometric ratio is displayed and is the average of all time points.

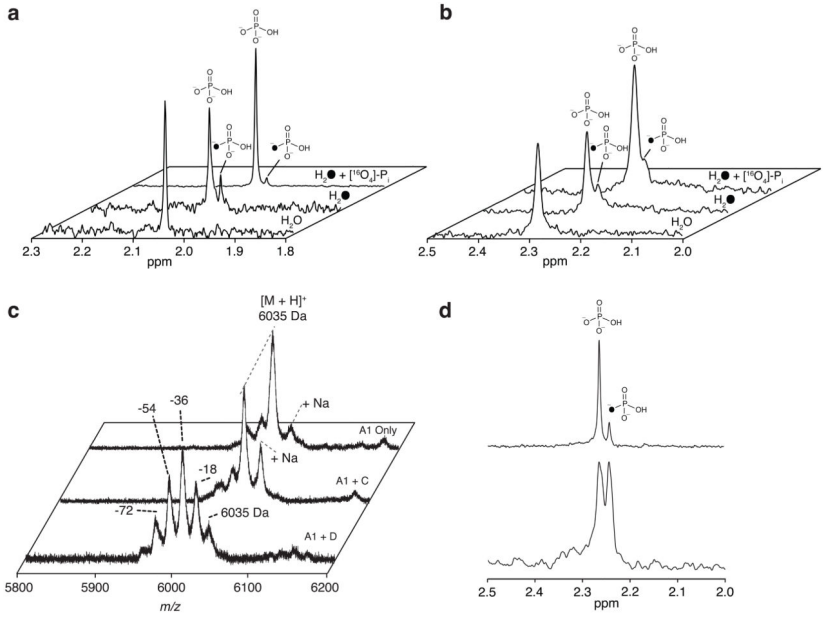

Figure 3. ATP is utilized by BalhD to directly activate the amide backbone of the substrate

^{31}P-NMR spectra of reactions with **(a)** the Balh or **(b)** the Mcb synthetase. The identity of the major product, [^{16}O]-P_i, was confirmed by spiking with authentic [^{16}O]-P_i. Slight pH differences between the Balh and Mcb samples (8.5 and 8.0, respectively) account for the altered chemical shifts. **(c)** MALDI-TO-MS spectra of BalhA1 treated with a stoichiometric amount of either BalhC or BalhD are displayed. The +1 charge state is shown. Mass shifts relative to the BalhA1 control are labeled. **(d)** A ^{31}P-NMR spectrum of a reaction conducted in [^{18}O]-water confirmed the presence of both [$^{16}O_4$]-P_i and [$^{18}O_1$,$^{16}O_3$]-P_i (bottom). The identity of the [$^{16}O_4$]-P_i peak was confirmed by spiking the sample with authentic [$^{16}O_4$]-P_i (top).

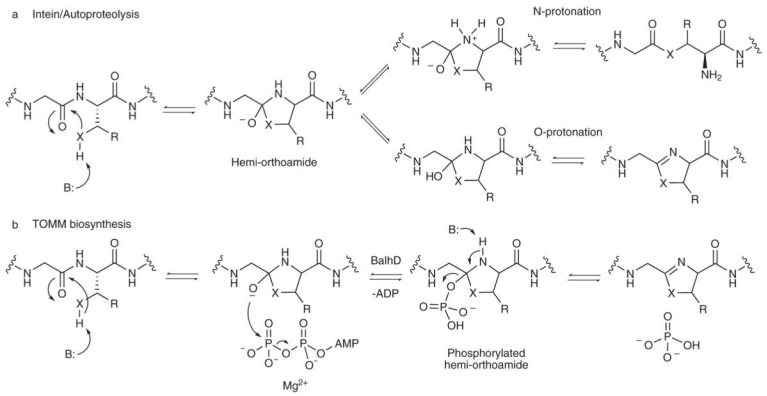

Figure 4. TOMM azoline installation is reminiscent of intein-mediated protein splicing

(a) The mechanism of protein splicing and autoproteolysis follows the *N*-protonation pathway to generate a (thio)ester. An azoline heterocycle could also be formed from the hemi-orthoamide intermediate if *O*-protonation occurred**. (b)** In TOMM biosynthesis, the use of ATP to phosphorylate the substrate could be used to drive the reaction down an *O*-elimination pathway to generate an azoline. X = S, O; R = H, CH_3.