



Published in final edited form as:

Nat Chem Biol. 2014 October ; 10(10): 823–829. doi:10.1038/nchembio.1608.

Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis

Kyle L. Dunbar^{1,2,†}, Jonathan R. Chekan^{3,†}, Courtney L. Cox^{2,4}, Brandon J. Burkhart^{1,2}, Satish K. Nair^{2,3,5,*}, and Douglas A. Mitchell^{1,2,4,*}

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

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

³Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

⁴Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

⁵Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Abstract

Despite intensive research, the cyclodehydratase responsible for azoline biogenesis in thiazole/oxazole-modified microcin (TOMM) natural products remains enigmatic. The collaboration of two proteins, C and D, is required for cyclodehydration. The C protein is homologous to E1 ubiquitin-activating enzymes, while the D protein is within the YcaO superfamily. Recent studies have demonstrated that TOMM YcaOs phosphorylate amide carbonyl oxygens to facilitate azoline formation. Here we report the X-ray crystal structure of an uncharacterized YcaO from *Escherichia coli* (Ec-YcaO). Ec-YcaO harbors an unprecedented fold and ATP-binding motif. This motif is conserved among TOMM YcaOs and is required for cyclodehydration. Furthermore, we demonstrate that the C protein regulates substrate binding and catalysis and that the proline-rich C-terminus of the D protein is involved in C protein recognition and catalysis. This study identifies the YcaO active site and paves the way for the characterization of the numerous YcaO domains not associated with TOMM biosynthesis.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence and requests for materials should be addressed to D.A.M. or S.K.N., douglasm@illinois.edu, phone: 1-217-333-1345, fax: 1-217-333-0508 snair@uiuc.edu, phone: 1-217-333-0641, fax: 1-217-244-5858.

†These authors contributed equally to this work

Author contributions

Experiments were designed by D.A.M., S.K.N., K.L.D., J.R.C., C.L.C. and B.J.B. and performed by K.L.D., J.R.C., C.L.C., and B.J.B. The manuscript was written by D.A.M., K.L.D. and J.R.C. with critical editorial input from S.K.N., C.L.C. and B.J.B. The overall study was conceived and managed by D.A.M. with S.K.N. overseeing all aspects of protein structure determination.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available online at <http://www.nature.com/naturechemicalbiology/>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

Accession codes. Coordinates for apo YcaO, AMP-bound YcaO, and AMPCPP-bound YcaO structures were deposited to the Protein Data Bank as codes 4Q84, 4Q86, and 4Q85 respectively.

The YcaO/DUF181 (domain of unknown function) family of proteins is currently comprised of nearly 5000 members distributed across the bacterial and archaeal domains. Disparate functions have been ascribed to members of the YcaO family. In *Escherichia coli*, the deletion or overexpression of the eponymous YcaO protein (Ec-YcaO; Fig. 1A) suggested that it potentiated the thiomethylation of ribosomal protein S12 and influenced biofilm formation, respectively^{1,2}. However, a molecular explanation for these observations is currently unavailable. Another YcaO-associated activity is the ATP-dependent cyclodehydration of Ser, Thr and Cys residues to azoline heterocycles, which is the defining modification of thiazole/oxazole-modified microcin (TOMM) natural products (Fig. 1A and B)³. TOMMs display diverse structures and activities^{3,4}, with some implicated in bacterial pathogenesis⁵, making the ~1000 bioinformatically identifiable TOMM YcaO proteins noteworthy members of the larger superfamily.

Although the TOMM YcaO domain was first implicated in cyclodehydration reactions in the mid-1990s⁶, its exact role remains unclear^{7,8}. The function of the TOMM YcaO (D protein) is intimately linked to members of the E1 ubiquitin-activating enzyme family (C protein) found in canonical TOMM biosynthetic clusters^{6,9,10}. Underscoring this linked function, roughly half of known TOMM clusters express C and D as a single polypeptide^{3,10}. Studies on both fused and unfused cyclodehydratases have demonstrated that these domains are necessary and sufficient for TOMM azoline formation^{7,11}. Consequently, the C–D complex is referred to as the TOMM cyclodehydratase (or alternatively, heterocyclase^{8,10}). As early studies on the cyclodehydratase were unable to observe activity from either protein in isolation^{6,9,12}, the respective contributions of C and D were inferred by bioinformatics. Based on the ATP dependence of the reaction¹³, and the homology of C to members of the E1 ubiquitin-activating superfamily¹⁰, which includes other ATP-utilizing enzymes (*e.g.* MccB, ThiF and MoeB)¹⁴, it was assumed that C was responsible for cyclodehydration, while the uncharacterized YcaO (D protein) played a regulatory or scaffolding role^{9,10}.

In 2012, research from our group challenged these assignments with the characterization of the TOMM cyclodehydratase from *Bacillus* sp. Al Hakam (Balh; Fig. 1A)⁷. This YcaO protein (BalhD) displayed ATP-dependent cyclodehydratase activity in the absence of the cognate C protein (BalhC); however, BalhC potentiated cyclodehydration by nearly 1000-fold. Considering that the C protein had been implicated in precursor peptide recognition in streptolysin S biosynthesis¹⁵, and that BalhC dictated the regio- and chemoselectivity of the Balh cyclodehydratase¹⁶, we hypothesized that the YcaO contained the active site residues while the C protein was responsible for binding the peptide substrate. Although YcaO proteins lack recognizable ATP-binding motifs, the presence of one or more YcaOs in the bottromycin and trifolitoxin biosynthetic clusters, which lack recognizable C proteins, supports these functional assignments (Supplementary Results, Supplementary Fig. 1)^{17–21}.

While the above studies assigned a putative activity to TOMM YcaOs, a molecular understanding of cyclodehydratase catalysis remained elusive. Recently, the X-ray crystal structure of a fused cyclodehydratase was reported (TruD; Protein Data Bank, PDB code 4BS9), providing the first structural glimpse of a TOMM cyclodehydratase⁸. The C domain adopted the expected E1 fold while the YcaO fold was unique. As the structure lacked both the ATP and peptide substrates, no information regarding substrate engagement and

catalysis could be gleaned. However, the lack of structural homology of YcaO to known ATP-binding proteins led to the reassertion that the C domain was responsible for ATP binding and carbonyl activation, while the YcaO domain catalyzed the requisite nucleophilic attack⁸.

Here, we report the structure of a non-TOMM YcaO from *E. coli* in various nucleotide-bound and -free forms and demonstrate that the most conserved residues in YcaOs comprise a previously uncharacterized ATP-binding motif. We show that these ATP-binding residues are critical for catalysis in TOMM YcaOs using BalhD as a model cyclodehydratase. We further identified the active site of TOMM cyclodehydratases and demonstrated that the conserved, proline-rich C-termini are involved in active site organization and C protein binding. Our results strongly support a model where ATP utilization is a universal feature of YcaOs (TOMM and non-TOMM) and that TOMM C proteins recognize the peptide substrate and potentiate the activity of the cognate YcaO.

RESULTS

E. coli YcaO hydrolyzes ATP to AMP and PP_i

With the ATP-dependent cyclodehydratase activity of BalhD previously established⁷, we attempted to locate the ATP binding site in BalhD by 8-azido-ATP crosslinking; however, these experiments were unsuccessful. Furthermore, the TruD crystal structure did not reveal an obvious ATP-binding site⁸ and BalhD was refractory to numerous crystallization attempts. We reasoned that since TOMM YcaOs evolved to interact with their cognate C proteins, working with a non-TOMM (with no C protein partner) might alleviate the previously encountered challenges. The local genomic environment of Ec-YcaO does not contain an E1 homolog (*i.e.* TOMM C protein, Fig. 1A), nor is Ec-YcaO known to interact with an E1 homolog, making it an attractive candidate for structural and biochemical characterization. Ec-YcaO was cloned into a tobacco etch virus (TEV) protease-cleavable maltose-binding protein (MBP)-fusion vector and expressed in *E. coli* (Supplementary Fig. 2). While the function of Ec-YcaO was unknown, characterized cyclodehydratases hydrolyze ATP in the absence of peptide substrates⁷. Consequently, we measured the ATPase activity of Ec-YcaO using an established purine nucleoside phosphorylase assay²². This assay revealed that Ec-YcaO indeed hydrolyzed ATP, preferentially generating AMP and PP_i (Supplementary Fig. 3). Although ATP hydrolysis was slow, perhaps because the native substrate was not present, Ec-YcaO displayed a K_M for ATP of ~80 μ M (Supplementary Fig. 3), comparable to several characterized cyclodehydratases^{7,11,13}.

Crystallization of Ec-YcaO

The structure of nucleotide-free Ec-YcaO (contains a mercurial salt for phasing) was determined to a Bragg limit of 2.63 Å and revealed a circularly symmetric homodimer in the asymmetric unit (Supplementary Table 2). The overall structure consists of an N-terminal YcaO domain of ~400 residues and a 150 residue C-terminal domain resembling a tetratricopeptide repeat (TPR) that mediates dimerization. A structure-based comparison against the PDB revealed similarity solely with TruD, the only other solved YcaO structure

(RMSD of 3.1 Å over 279 aligned Ca atoms)²³, confirming that YcaOs constitute a new structural fold (Supplementary Fig. 4).

To identify the ATP-binding site, we determined the structure of Ec-YcaO in complex with multiple nucleotides. Co-crystallization of Ec-YcaO with ATP produced an AMP-bound structure (2.25 Å), suggesting that *in situ* hydrolysis had occurred (Fig. 2A). To clarify the residues involved in ATP binding, we also determined the co-structure of Ec-YcaO with α , β -methyleneadenosine 5'-triphosphate (AMPCPP, a non-hydrolyzable ATP analog). These three structures facilitated the characterization of the ATP-binding site in the YcaO superfamily.

Structural characterization of ATP-binding in Ec-YcaO

Analysis of the 2.25 Å resolution AMP-bound and 3.29 Å resolution AMPCPP-bound co-crystal structures revealed that the adenine ring is recognized by Glu191 and Asn187 via interactions through the N7 nitrogen and between Ser16 and the exocyclic-N6 (Fig. 2B). Additionally, Lys9 resides above one face of the adenine ring, while Ala70, Ser71, and Gly74 are found within an α -helix that extends below the adenine and ribose rings, forming a hydrophobic surface. The ribose within the AMP- and AMPCPP-bound structures is oriented perpendicular to the adenine ring, with Ser184 and Glu78 coordinating the 2'- and 3'-hydroxyls, respectively (Fig. 2C and D). While Ser71 coordinates the α -phosphate in both structures, Arg286 coordinates to the α -phosphate in only the AMP-bound form (Fig. 2C and 2D). Surprisingly, two Mg²⁺ ions are found in the nucleotide-binding pocket in both structures. In the AMP structure, Glu199 and Glu78 ligate one Mg²⁺ ion while Glu290 and Glu75 bind the second (Fig. 2C). The Mg²⁺ ions are coordinated in a similar fashion in the AMPCPP structure with the subtle difference that Glu202, rather than Glu75, coordinated the second Mg²⁺ (Fig. 2D). This slight change in the coordination of the second Mg²⁺ ion positions the metal ions on opposite sides of the β - and γ -phosphates. Furthermore, Arg203 coordinates the γ -phosphate of AMPCPP (Fig. 2D). The interactions between Ec-YcaO and AMPCPP are summarized in Figure 2E.

The ATP-binding site is conserved in TOMM YcaOs

Utilizing the nucleotide-bound structures of Ec-YcaO, we established the conservation of the ATP-binding residues across the superfamily. First, we generated a Cytoscape sequence similarity network²⁴ of all YcaO members in InterPro (IPR003776)²⁵. During assembly, redundant sequences were removed, leaving ~2000 sequences in the network. While the sequence of TOMM precursor peptide dictates the structure of the natural product, there is also a strong correlation between TOMM structure and the sequence similarity of the cognate YcaO³. For the network, all YcaO sequences were manually annotated based on neighboring genes. YcaOs were categorized as being involved in TOMM biosynthesis if there was a gene encoding for a recognizable C protein in the local region (~10 kb on either side of the *ycaO* gene) or if the protein had an experimentally verified link to a known TOMM (*e.g.* bottromycin¹⁷⁻²⁰ and trifolitoxin²¹). The remaining YcaOs were separated into two other categories, non-TOMM YcaOs (*e.g.* Ec-YcaO) and TfuA-associated non-TOMM YcaOs. The latter were found within 10 kb of a gene encoding for the protein TfuA, which is implicated in trifolitoxin biosynthesis²⁶. Whenever possible, TOMM YcaOs were further

subdivided by expected structural class. Based on these classifications, we determined that an expectation value of 10^{-80} gave an optimal separation of YcaO sequences into isofunctional clusters (Supplementary Fig. 5).

Using the sequence similarity network as a guide, 349 of the ~2000 members in the non-redundant network were selected from across all clusters and a maximum likelihood tree was generated (Supplementary Fig. 6). Among the 349 were all singletons, defined as divergent family members not grouping with any other YcaO at an e-value of 10^{-80} . Using this diversity-maximized tree, a sequence logo for each of the regions involved in ATP binding was generated using WebLogo (Fig. 3A)²⁷. The logos clearly demonstrate that the ATP-binding pocket is highly conserved in the YcaO family across all three groups (*i.e.* TOMM, non-TOMM, and TfuA-associated non-TOMM). The ATP-binding residues were found to be the most conserved feature in the YcaO superfamily (Fig. 3B). This is in stark contrast to TOMM C proteins, which have lost the ATP-binding residues conserved in all characterized non-TOMM E1 ubiquitin-activating superfamily members¹⁴ (Supplementary Fig. 7). Furthermore, the conservation in the YcaO ATP-binding residues is maintained in all characterized TOMMs (Supplementary Fig. 8 and 9), suggesting that the previously reported carbonyl activation mechanism is likely a universal biosynthetic feature^{7,16}.

The Ec-YcaO ATP-binding site is vital for BalhD activity

Because the native substrate of Ec-YcaO is unknown, we validated the ATP-binding residues by conducting structure-function studies on BalhD. An alignment of BalhD and Ec-YcaO permitted the mapping of the nucleotide- and Mg^{2+} -binding residues onto BalhD (Supplementary Fig. 4 and 9). Subsequently, an alanine mutagenesis scan was performed on the polar residues of BalhD predicted to bind ATP. Every mutation was well tolerated in terms of protein yield/stability (Supplementary Fig. 2). The effect on heterocycle formation on BalhA1 (the peptide substrate) by the mutant BalhD proteins, in the presence of BalhC, was monitored in a 16 h endpoint assay (Supplementary Fig. 10). Of the 11 mutated residues in the ATP-binding pocket, four were able to convert BalhA1 to the previously reported penta-azoline species⁷, three showed intermediate levels of processing (2–4 heterocycles), while the remaining four generated no heterocyclic products within the limit of detection (Table 1). To quantify the effect of each mutation to BalhC/D activity, the rate of ATP hydrolysis was monitored using the K_M concentration of BalhA1 (15 μ M) and a concentration of ATP that would be saturating for wild-type BalhD (3 mM). Mutants unable to cyclize BalhA1, even after extended reaction times, displayed no detectable ATP hydrolysis over the assay background (Supplementary Fig. 11). Likewise, mutants that installed five azolines on BalhA1 in the endpoint assay had the highest ATP hydrolysis rates. These data are congruent with our earlier work showing that ATP hydrolysis is tightly coupled to heterocycle formation⁷. The YcaO mutations examined here did not appear to disrupt this feature of TOMM cyclodehydration.

While mutation of the BalhD ATP-binding pocket reduced cyclodehydratase activity, an alternative interpretation of the above data could be that these mutations interfered with the association of BalhC and BalhD. A unique feature of the Balh cyclodehydratase is that BalhD is catalytically active in the absence of BalhC⁷. This permitted the use of BalhD-only

activity measurements to determine if the alanine mutations affected the intrinsic cyclodehydratase activity. Heterocycle formation endpoint assays (16 h) were again conducted, but this time with 50 μM BalhA1 and 25 μM BalhD mutant to account for the expected $\sim 10^3$ drop in catalytic activity in the absence of BalhC⁷. The resultant mass spectra confirmed that the decrease in cyclodehydration arose from a perturbation in BalhD activity (Table 1 and Supplementary Fig. 12).

For all BalhD mutants with measureable cyclodehydratase activity, we obtained the Michaelis-Menten kinetic parameters for BalhA1 and ATP (Table 1 and Supplementary Fig. 13). Every mutation negatively affected the observed k_{cat} (k_{obs}), indicating that the selected residues were of catalytic importance. Apart from K281A, and to a lesser extent S72A, all ATP-binding site mutations of BalhD substantially increased the K_{M} for ATP. In contrast, the only mutant in this series to substantially raise the K_{M} for BalhA1 was R198A (Table 1).

Four BalhD mutants (*i.e.* E76A, E79A, E194A, E197A) did not exhibit detectable cyclodehydratase activity. Potential explanations include an inability to bind the substrates (BalhA1 and/or ATP), hydrolyze ATP, or a structural perturbation with these mutants. Previous work demonstrated that BalhC potentiates the BalhA1-independent ATPase activity of BalhD by 2.5-fold⁷; however, when the four inactive mutants were assayed, no potentiation was observed (Supplementary Fig. 14). This suggested that the lack of BalhD activity was due to a structural perturbation or the inability to bind/hydrolyze ATP. Unfortunately, attempts to directly measure ATP binding or a secondary structure perturbation of BalhD with isothermal titration calorimetry or circular dichroism spectroscopy, respectively, were problematic owing to the solubility characteristics of BalhD. However, we reasoned that the latter could be assayed indirectly by monitoring the interaction between BalhC and a mutant BalhD through a competition assay and size-exclusion chromatography. While all of the BalhD mutants were able to associate with the BalhC (Supplementary Fig. 15), E76A and E194A did so with reduced affinity, suggesting that these mutations affected the BalhC-BalhD interaction surface. Conversely, the wild-type-like affinity that BalhD E79A and E197A displayed for BalhC suggested that these mutants were inactive due to inability to bind/hydrolyze ATP.

The BalhD C-terminus affects BalhC binding and catalysis

In addition to the conserved ATP-binding site, TOMM YcaO proteins have a highly conserved, proline-rich C-terminus. In the most pronounced cases, the final five residues of the YcaO are PxPxP (Supplementary Fig. 16)⁹. The proline-rich C-terminus is not conserved in non-TOMM or TfuA-associated YcaO domains (Supplementary Fig. 16), implicating the motif in either C protein recognition or cyclodehydratase activity. This hypothesis is supported by the observation that the C-terminus of TruD is in close proximity to the YcaO ATP-binding site and is surface-accessible (Supplementary Fig. 16). We first interrogated the importance of this motif by truncating 5 residues from the BalhD C-terminus. This minor perturbation abolished the catalytic activity of BalhC/D (Table 2 and Supplementary Fig. 17). Removing the C-terminal three residues of BalhD produced an identical result, while removal of the C-terminal residue of BalhD (BalhD P429*; * stop codon) decreased activity by > 100 -fold (Table 2 and Supplementary Fig. 17). Similarly, extending the C-terminus by

a single amino acid (BalhD P_xP_xPG), or deleting two amino acids upstream of the P_xP_xP motif (BalhD 418–419), also resulted in inactive cyclodehydratases (Table 2 and Supplementary Fig. 17).

To establish if altering the BalhD C-terminus affected the interaction with BalhC, we assessed the ability of BalhC to potentiate the background ATPase activity of BalhD mutants lacking detectable activity. Potentiation was not observed in any case (Supplementary Fig. 18), indicating that mutants of the BalhD P_xP_xP motif had lost the ability to bind/hydrolyze ATP or to bind BalhC. We next assessed the ability of each P_xP_xP mutant to bind BalhC. Utilizing a combination of size-exclusion chromatography and a competition assay, all truncations to the BalhD P_xP_xP motif were shown to have decreased affinity to BalhC (Supplementary Fig. 19). Moreover, the order of heterocycle formation was dysregulated in BalhD P429*, reminiscent of wild-type BalhD reactions lacking BalhC (Supplementary Fig. 20)¹⁶.

Intrigued by the loss of activity observed upon extending or truncating the C-terminus, we next investigated the importance of the amino acid composition of the BalhD P_xP_xP motif (PHFPF₄₂₉). As with the truncations, any mutation to the C-terminal 5 residues of BalhD decreased cyclodehydratase activity (Table 2 and Supplementary Fig. 21). The decrease in activity ranged from ~2.5-fold (P429G) to 100-fold (H426A), with the severity diminishing the closer the mutation was to the C-terminus. This result was consistent with the observation that the C-terminal residues of TruD are located in a channel leading to the active site. As with the P_xP_xP truncations, every mutant tested, apart from F428A, displayed a decreased affinity for BalhC (Supplementary Fig. 22). Consistent with this observation, mutation of the terminal amino acid (P429G) resulted in an aberrant order of heterocycle formation (Supplementary Fig. 23). Increasing the flexibility of the PHFPF motif, by substituting it with GHGFG, yielded an inactive cyclodehydratase (Table 2 and Supplementary Fig. 21).

While these results implicated the C-terminus of BalhD in BalhC recognition, a decrease in BalhC affinity could not explain the data in its entirety. For example, both BalhD P425* and P429* displayed reduced interactions with BalhC, but only P425* was catalytically inactive (Table 2 and Supplementary Fig. 19). Furthermore, BalhD P_xP_xPG displayed a wild-type level of interaction with BalhC, despite being catalytically inactive. Based on these results, the activity of each mutant was tested in the absence of BalhC. Analogous to the mutations to the ATP-binding pocket, mutations to the P_xP_xP motif affected the intrinsic activity of BalhD (Supplementary Fig. 24). For all tractable BalhD mutants, BalhA1 and ATP Michaelis-Menten kinetic curves were obtained for the mutant BalhC-D complexes. While the largest effects were on k_{obs} , the mutations also affected the K_{M} for ATP and BalhA1 (Table 2). Unlike the ATP-binding mutants, the changes to K_{M} for the two substrates were similar in the P_xP_xP mutants, suggesting that the C-terminus is involved in active site organization/catalysis, not substrate binding. Furthermore, the importance of the YcaO C-terminus appears to be general for TOMM biosynthesis, given that the cyclodehydratase activity of McbD (microcin B17 YcaO protein) was also abolished when the C-terminus was truncated (Supplementary Fig. 25).

BalhC regulates BalhD ATP-binding and catalysis

To further characterize the C-terminus of BalhD, a derivative containing a C-terminal His₆ tag was generated (PA₄LEH₆; where P is the last residue of wild-type BalhD). As expected from earlier experiments with BalhD PxPxPG, addition of the longer tag abolished heterocycle formation (Supplementary Fig. 26). Although heterocycle formation is stoichiometric with ATP hydrolysis for wild-type cyclodehydratase⁷, the C-terminal His₆ tagged BalhD displayed robust ATPase activity, irrespective of the presence of BalhA1 (Fig. 4A). Moreover, this high level of unproductive ATP hydrolysis was potentiated by the addition of BalhC to the same extent observed with wild-type BalhD (2.5-fold increase), indicating that the His₆ tag did not interfere with BalhC recognition (Fig. 4A). With a BalhD derivative displaying robust BalhC-independent ATPase activity in-hand, we evaluated the role of BalhC on ATP utilization by BalhD by obtaining ATP Michaelis-Menten kinetics parameters for BalhD-A₄LEH₆ alone and in complex with BalhC (Fig. 4B). These data indicated that the addition of BalhC modulated BalhD activity by increasing the k_{obs} and decreasing the K_M for ATP.

TOMM C proteins provide leader peptide binding

Precursor peptide recognition in the majority of ribosomally synthesized post-translationally modified peptides (RiPPs) occurs in a bipartite fashion. An N-terminal sequence (leader peptide), serves as the recognition sequence by the modification enzymes, while the C-terminal sequence (core peptide) contains the sites of post-translational modification^{4,28}. We have previously demonstrated that the regioselectivity and directionality of BalhD azoline formation is dependent on BalhC¹⁶. Thus, we hypothesized that BalhC was responsible for presenting the core peptide to BalhD, most likely by engaging the leader peptide. However, the identification of BalhD mutants with a perturbed K_M for BalhA1 suggested that the YcaO domain might play a role in substrate recognition. To assess the role of the C and D proteins in peptide substrate recognition, a fluorescein-labeled BalhA1 leader peptide was employed to monitor binding to BalhC and BalhD by fluorescence polarization. While BalhD was found to not bind the BalhA1 leader peptide to any significant extent, BalhC displayed a K_d of 11 μ M (Fig. 4C), near the previously measured K_M for BalhA1 of 16 μ M²⁹. Moreover, the addition of BalhD did not significantly alter BalhC's affinity for the BalhA1 leader peptide ($p > 0.05$). Consequently, these data support a model where *i.* BalhD does not engage the leader region of BalhA1 and *ii.* the elevated K_M 's of select BalhD mutants for BalhA1 was due to a decreased affinity for the core region of BalhA1.

Discussion

We have discovered that Ec-YcaO contains a novel ATP-binding fold. Based on steric and electrostatic complimentary requirements, the YcaO strategy for binding ATP is reminiscent of other structurally characterized ATP-binding proteins. For example, the Lys/ α -helix "sandwich" involved in adenine recognition is similar to the conserved Arg and Gly motif found in class I amino acyl tRNA synthetases³⁰. Furthermore, select members of the ATP-grasp and PurM families have been shown to bind ATP through the use of multiple divalent cations^{31,32}; however, in these proteins, the Mg²⁺ ions are coordinated to all three phosphates, not just the β - and γ -phosphates. Since these similarities in ATP-binding occur

despite a lack of structural and primary homology between YcaO and all other known ATP-binding proteins, this represents an example of convergent evolution in ATP-binding domains.

The ATP-binding residues are the most highly conserved motifs in the YcaO superfamily and, appropriately, represent a prominent signature for the hidden Markov model that bioinformatically defines the YcaO/DUF181 family (IPR003776). Our extensive bioinformatics analysis, X-ray crystallographic data on Ec-YcaO, and biochemical characterization of BalhD confirm that ATP-utilization is a conserved feature in the superfamily. In spite of the low level of overall similarity between Ec-YcaO and BalhD, we were able to demonstrate that the YcaO ATP-binding motif was critical for cyclodehydratase activity. While the mutations affected BalhD activity to differing extents, the impact of mutating a particular residue on Balh cyclodehydratase activity was proportional to the level of conservation within the YcaO family (Fig. 3A).

During our analysis of the YcaO ATP-binding motif, we observed a striking difference between the TOMM and non-TOMM YcaO domains. TOMM YcaOs (D proteins) almost invariably harbor proline-rich C-termini, with PxPxP most often serving as the terminal five residues of the protein. While the widespread nature of the PxPxP motif had been previously recognized⁹, prior to this work, it was unclear if this motif played any role in TOMM biogenesis. Our data indicate that the C-terminus of TOMM YcaOs assists in both C protein recognition and cyclodehydration. It is rare for the C-terminus of an enzyme to be important for catalysis^{33–37}. In fact, the terminal regions are often highly sequence variable within a protein family. Intriguingly, the C-terminal proline content of a YcaO has powerful predictive value. If present, the YcaO is quite probably involved in TOMM biosynthesis. This tentative assignment can be confidently made even without knowledge of the flanking genes. As such, we hypothesize that a subset of the 249 (~8%) non-TOMM YcaOs that contain a proline-rich C-terminus may actually be stand-alone TOMM YcaOs (akin to the bottomycin YcaOs).

Previously, BalhC was shown to potentiate BalhD via an unknown mechanism⁷. The current study indicates that this potentiation occurs via two distinct mechanisms. First, the serendipitous identification of a C-terminal His-tagged construct of BalhD with robust ATP hydrolysis (BalhD-A₄LEH₆) allowed us to show that the presence of BalhC increases k_{obs} and lowers the K_M for ATP. Although our data suggest that C protein potentiation occurs via allosteric activation, follow-up studies will be required to validate this hypothesis. Secondly, our data demonstrate that BalhC is responsible for binding the leader peptide of BalhA1, thus efficiently bringing the substrate in close proximity to the BalhD active site. This result is in accord with a previous study implicating SagC in leader peptide recognition during streptolysin S biosynthesis¹⁵. In further support of a general role for TOMM C proteins in peptide substrate binding, the “C portion” (homologous to E1 ubiquitin-activating enzymes) of TruD has a MccB-like N-terminal “peptide clamp”⁸, which is responsible for leader peptide binding in microcin C7 biosynthesis³⁸. Combined with the fact that TOMM C proteins lack the ATP-binding site that is conserved in all characterized non-TOMM E1 ubiquitin-activating family members, all lines of evidence suggest that TOMM C proteins

engage the leader peptide while simultaneously potentiating the carbonyl activation chemistry of their cognate YcaO domain (D protein; Supplementary Fig. 27).

In light of this, it is not yet clear how stand-alone TOMM YcaO proteins (*i.e.* for bottromycin and trifolitoxin production) perform cyclodehydrations in the absence of a C protein. Given the diversity between these stand-alone and canonical (C protein-containing) TOMM YcaOs, we envision that multiple solutions to the substrate recognition problem could exist. For example, it is possible that these biosynthetic pathways utilize an unidentified companion protein to bind the precursor peptide. Alternatively, these YcaO proteins may have evolved to bind a specific motif within the core peptide and modify the substrate a single time. Of note is the fact that the bottromycin and trifolitoxin stand-alone YcaO domains are each predicted to install a single heterocycle^{17–21}. This is in stark contrast to canonical TOMMs that process a wide array of core peptides, often at numerous locations^{3,29,39}. Such promiscuity is common in other RiPPs and likely accounts for the existence of leader peptides (*i.e.* the modification enzymes can be specific for motifs within the leader peptide, but promiscuous on the core once the enzyme-substrate complex is formed)^{4,28}. Further work, including the reconstitution of a stand-alone YcaO, will be required to test these claims.

The capacity to bind ATP (or possibly other nucleotide triphosphates) appears to be ubiquitous in the YcaO superfamily, but it remains unclear whether the TOMM cyclodehydratase-like direct activation of carbonyls is a universal feature. It is intriguing that YcaOs have recently been implicated in the formation of thioamides⁴⁰ and macroamidine rings^{17–20}, as both of these modifications could conceivably occur through carbonyl activation. In addition to providing significant insight into the mechanics of TOMM cyclodehydration, the results presented here provide an initial framework to explore the elusive functions of the 4,000 uncharacterized non-TOMM YcaOs.

ONLINE METHODS

General methods

Unless otherwise specified, all chemicals were purchased from Sigma or Fisher Scientific. DNA sequencing was performed by either the Roy J. Carver Biotechnology Center (UIUC) or ACGT Inc. Restriction enzymes were purchased from New England Biolabs (NEB), Pfu Turbo was purchased from Agilent, and dNTPs were purchased from either NEB or GenScript. Oligonucleotide primers were synthesized by either Integrated DNA Technologies (IDT) or Eurofins MWG Operon. Fluorescein labeled BalhA1 leader peptide was purchased from GenScript as an *N*-terminal FITC-Ahx (fluorescein isothiocyanate, aminoethyl linker) conjugate with a single glycine spacer. Unless otherwise stated, all proteins and substrates were used as MBP fusions to circumvent solubility issues. A table of the peptide substrates used in this study can be found in Supplementary Table 3.

Cloning of MBP-YcaO

Ec-YcaO was amplified by PCR from *Escherichia coli* BL21 cells using the forward and reverse primers listed in Supplementary Table 1. Polymerase reactions were carried out with Pfu Turbo and the amplified product was digested with BamHI and NotI following a gel

extraction. The digested gene was PCR-purified and ligated into an appropriately digested, modified pET28 vector containing a tobacco etch virus (TEV) protease cleavable, *N*-terminal MBP-tag.

Preparation of BalhD PxPxP mutants

BalhD was amplified by PCR from the previously described pET28-BalhD plasmid²⁹ using the primers listed in Supplementary Table 1. Polymerase reactions were performed with PFU Turbo and the amplified product was digested with BamHI and NotI following gel extraction. The digested gene was PCR-purified and ligated into an appropriately digested, modified pET28 vector containing a tobacco etch virus (TEV) protease cleavable, *N*-terminal MBP-tag.

Site-Directed mutagenesis

Site-directed mutagenesis of BalhD and McbD was carried out using the QuikChange method according to the manufacturer's instructions. All mutagenesis primers are listed in Supplementary Table 1.

Overexpression and purification of MBP-tagged proteins

All proteins were purified with amylose resin (NEB) according to previously described procedures⁷.

Multiple sequence alignments

Alignments were made with Clustal Omega using the standard parameters⁴¹.

Cytoscape sequence similarity network

A sequence similarity network was created using the Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST, enzymefunction.org)⁴². Sequences from the YcaO superfamily (InterPro number IPR003776) were used for the analysis. The network was constructed at an expectation-value (e-value) of 10^{-80} . Networks were visualized by Cytoscape using the organic layout²⁴. Sequences with 100% identity were visualized as a single node in the network.

Maximum likelihood phylogenetic analysis

A set of YcaO sequences representing the full diversity of the YcaO family was selected based on the Cytoscape sequence similarity network (Supplementary Fig. 5). This included at least one protein from each cluster and all singletons (proteins that fail to cluster with any other YcaO) for a total of 349 proteins (~17% of the non-redundant Cytoscape sequence similarity network). The phylogenetic analysis was performed with Molecular Evolutionary Genetics Analysis (MEGA5)⁴³. An amino acid sequence alignment was created using the standard parameters of ClustalW⁴⁴ and a maximum likelihood phylogenetic tree was created using standard parameters in MEGA5.

ATP-binding site conservation

The ligand interaction network for AMPCPP was generated using LigPlot plus⁴⁵ using the standard parameters. A WebLogo²⁷ frequency plot for the ATP-binding motif was generated from a Clustal Omega alignment of all of the sequences from the specified family using the standard parameters. The conservation map for the YcaO family was generated by aligning 150 unique YcaO sequences with at least 35% sequence similarity to Ec-YcaO (e-value < 10^{-5}) and mapping the resulting conservation data onto the Ec-YcaO structure using ConSurf⁴⁶. The structure based YcaO alignment was generated using Clustal Omega⁴¹ and ALINE⁴⁷, while the structural overlay of YcaO and TruD was generated using PyMOL version 1.5 (Schrödinger).

Proline-rich C-terminus analysis

The proline content of the C-termini of all YcaOs in the Cytoscape sequence similarity network was determined. Proteins were deemed to have a proline-rich (P-rich) C-terminus if at least 4 of the final 30 residues were Pro. In the most pronounced cases, the terminal 6 residues of the YcaO contain a PxPxP motif. Proteins were identified as containing a PxPxP C-terminus if they contained a PxP motif in the final 6 residues and at least 3 Pro in the final 30 residues.

Ec-YcaO crystallization

Purified MBP-tagged Ec-YcaO was treated with TEV protease at for 18 h at 4 °C. Successful cleavage was confirmed by SDS-PAGE and Ec-YcaO was subsequently separated from the His₆-tagged MBP by subtraction Ni²⁺ affinity chromatography. Fractions with a purity of at least 90% were combined and purified by size-exclusion chromatography using a GE Superdex 200 column equilibrated in a buffer containing 300 mM KCl and 20 mM HEPES pH 7.5. Fractions were collected and concentrated for crystallization screening experiments. The APO Ec-YcaO crystals were obtained by sitting drop crystallization using a mother liquor containing 1.8 M ammonium citrate pH 7.0 and 8 mg/mL Ec-YcaO in a 1:1 ratio. Incubation at 4°C yielded crystals after several days. Soaking with 1 mM PCMBBA for 4 hours prior to vitrification was used to generate phases. Immediately before vitrification, crystals were soaked in a cryoprotectant containing 1.8 M ammonium citrate pH 7.0 and 30% trehalose. AMP Ec-YcaO crystals were grown using hanging drop vapor diffusion with a mother liquor of 18% PEG 8,000, 0.1 M magnesium acetate, and 0.1 M sodium cacodylate pH 6.5. An Ec-YcaO concentration of 8 mg/mL and substrate concentration of 1 mM ATP and 1 mM MgCl₂ was used for the formation of crystals at 4°C. Directly before vitrification, crystals were immersed in a cryoprotectant containing the mother liquor supplemented with 30% MPD. AMPCPP Ec-YcaO crystals were grown and frozen in an identical manner with the exception of using 6 mg/mL Ec-YcaO, 1 mM AMPCPP instead of 1 mM ATP, and 20% PEG 8,000 instead of 18% PEG 8,000. All data was collected at LS-CAT sector 21 of the Argonne Nation Labs Advanced Photon Source at 100 K using wavelengths of 0.97872 Å for the APO and AMP bound structures and 0.97857 Å for the AMPCPP structure.

Ec-YcaO structure solution and refinement

Collected data was integrated and scaled using HKL2000 or autoProc⁴⁸. The PCBMA soaked APO Ec-YcaO crystals were used as a source of anomalous signal in SAD phasing using the PHENIX software suite⁴⁹. Automated building of the structure was accomplished by the arp/wARP server⁵⁰. Manual refinement was performed using COOT⁵¹ and REFMAC5⁵². For the AMP and AMPCPP containing structures, the APO Ec-YcaO was used as a search model to obtain phases using the PHENIX software suite. PHENIX was also used for automated building of the structures. Manual refinement was again performed using COOT and REFMAC5. Final Ramachandran statistics as determined by PROCHECK⁵³ are as follows: 97.2% favored, 2.8% allowed, and 0.0% outliers for the APO structure; 97.8% favored, 2.2% allowed, and 0.0% outliers for the AMP bound structure; and 96.3% favored, 3.7% allowed, and 0.0% outliers for the AMPCPP bound structure.

Endpoint heterocycle formation assays

For the reactions with the BalhD mutants, 50 μM MBP-BalhA1 was mixed with either 2 μM MBP-BalhC/D (CD activity) or 25 μM MBP-BalhD (D-only activity) in synthetase buffer [50 mM Tris (pH 7.5), 125 mM NaCl, 10 mM DTT, 20 mM MgCl_2 and 3 mM ATP]. The MBP tags were removed using 0.05 mg/mL of TEV protease and reactions were carried out for 18 h at 25 °C. Reactions with the Mcb enzymes were carried out identically except that the dehydrogenase (McbC) was also added to the reaction and proteins were cleaved with 0.1 $\mu\text{g/mL}$ thrombin protease (from bovine plasma). Processing was monitored by MALDI-TOF MS.

PNP-Based kinetic studies

Substrate processing kinetics for the BalhD mutants were determined using a previously described purine nucleoside phosphorylase (PNP)-coupled assay^{7,22}. For BalhA1 kinetic experiments, variable concentrations of MBP-BalhA1 (1–120 μM) were reacted with 1–10 μM MBP-tagged BalhC/D while ATP was held constant at 3 mM. ATP kinetic experiments were carried out in an identical fashion except that MBP-BalhA1 was fixed at 80 μM and variable concentrations of ATP (0.1–5 mM) were used. Although this does not provide a saturating level of BalhA1 for all mutants, the K_M for ATP does not change with varied BalhA1 concentration (Supplementary Fig. 13). Reactions were carried out in triplicate. Regression analyses to obtain the kinetic parameters for both substrates were carried out with IGOR Pro version 6.12 (WaveMetrics).

Fluorescence polarization

Equilibrium BalhC/BalhD fluorescence polarization binding assays were performed at 25 °C in non-binding surface, 384 black well, polystyrene microplates (Corning) and measured using a FilterMax F5 multi-mode microplate reader (Molecular Devices) with default settings. For each titration, protein was serially diluted into binding buffer [50 mM HEPES pH 7.5, 300 mM NaCl, 2.5% (v/v) glycerol, 0.5 mM TCEP], mixed with 10 nM fluorescein-labeled BalhA1 leader peptide (FP-BalhA1-LP), and equilibrated for 15 minutes with shaking prior to measurement. Data from three independent titrations were background subtracted and fitted using a non-linear dose response curve in OriginPro9 (OriginLab).

Size-Exclusion chromatography

A 200 μ L sample containing 25 μ M MBP-tagged BalhC/D was prepared in cleavage buffer [20 mM Tris (pH 7.5), 500 mM NaCl, 10 % glycerol (v/v), 0.5 mM tris(2-carboxyethyl)phosphine (Gold Biotechnology)] and treated with 0.05 mg/mL TEV protease for 12 h at 4 °C. The amount of protein in a BalhC/BalhD complex was assessed on a Flexar HPLC (Perkin Elmer) equipped with an analytical Yarra SEC-3000 (300 \times 4.6 mm, Phenomenex) equilibrated with cleavage buffer. Peaks of interest were collected and their composition was determined via a Coomassie stained 12% SDS-PAGE gel. The approximate molecular weights were determined by generating a standard curve with a 12–200 kDa molecular weight standard kit (Sigma). Control runs were also performed in which one of the two proteins was omitted. Chromatograms were analyzed using Flexar Manager (Perkin Elmer).

Mutant BalhD IC₅₀ determination

1 μ M MBP-tagged BalhC and BalhD were mixed in synthetase buffer [50 mM Tris (pH 7.5), 125 mM NaCl, 10 mM DTT, 20 mM MgCl₂ and 3 mM ATP] with 0.25–20 μ M mutant BalhD protein. Reactions were initiated by the addition of 15 μ M MBP-BalhA1 and progress was measured using the PNP phosphate detection assay. All reactions were performed in triplicate. IC₅₀ values were calculated with IGOR Pro version 6.12 (WaveMetrics).

BalhC “K_M” for active BalhD mutants

The affinity of catalytically active BalhD mutants for BalhC was determined using the PNP phosphate detection assay and a previously described procedure¹². 25 μ M MBP-BalhA1 was mixed with 1 μ M MBP-BalhD in synthetase buffer and reactions were initiated via the addition of 0.15–4 μ M BalhC. All reactions were performed in triplicate. Regression analyses to obtain kinetic parameters for BalhC were carried out with IGOR Pro version 6.12 (WaveMetrics).

Heterocycle localization via FT-MS/MS

50 μ M MBP-BalhA1 was modified by 2 μ M MBP-tagged BcerB, BalhC, and BalhD in synthetase buffer for 18 h at 25 °C. Proteins were digested with 0.02 mg/mL sequencing grade trypsin (Promega) in 50 mM NH₄CO₃ (pH 8.0) for 30 min at 37 °C before the sample was quenched via the addition of formic acid to a final concentration of 10% (v/v) and precipitate was removed via centrifugation at 17,000 \times g. FT-MS/MS analysis was carried out as previously described¹⁶.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Caitlin Deane and Kellie Taylor for the generation of select BalhD mutants and Joel Melby for assistance with collecting MS/MS data. This work was supported by the US National Institutes of Health (NIH) (1R01 GM097142 to D.A.M. and 2T32 GM070421 to K.L.D., B.J.B., and J.R.C.). Additional support was from the

Harold R. Snyder Fellowship (UIUC-Dept. of Chemistry to K.L.D.), the Robert C. and Carolyn J. Springborn Endowment (UIUC-Dept. of Chemistry to B.J.B.), the National Science Foundation Graduate Research Fellowship (DGE-1144245 to B.J.B.), and the University of Illinois Distinguished Fellowship (UIUC-Graduate College to J.R.C.) The Bruker UltrafleXtreme MALDI TOF/TOF mass spectrometer was purchased in part with a grant from the National Center for Research Resources, NIH (S10 RR027109 A).

References

1. 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:M110005199. [PubMed: 21169565]
2. Tenorio E, et al. Systematic characterization of *Escherichia coli* genes/ORFs affecting biofilm formation. *FEMS Microbiol Lett*. 2003; 225:107–114. [PubMed: 12900028]
3. 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]
4. 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]
5. Molloy EM, Cotter PD, Hill C, Mitchell DA, Ross RP. Streptolysin S-like virulence factors: the continuing saga. *Nat Rev Microbiol*. 2011; 9:670–681. [PubMed: 21822292]
6. 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]
7. 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]
8. Koehnke J, et al. The cyanobactin heterocyclase enzyme: a processive adenylase that operates with a defined order of reaction. *Angew Chem Int Ed*. 2013; 52:13991–13996.
9. Lee SW, et al. Discovery of a widely distributed toxin biosynthetic gene cluster. *Proc Natl Acad Sci USA*. 2008; 105:5879–5884. [PubMed: 18375757]
10. 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]
11. McIntosh JA, Schmidt EW. Marine molecular machines: heterocyclization in cyanobactin biosynthesis. *ChemBiochem*. 2010; 11:1413–1421. [PubMed: 20540059]
12. 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]
13. 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]
14. Schulman BA, Harper JW. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Biol*. 2009; 10:319–331. [PubMed: 19352404]
15. Mitchell DA, et al. Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S. *J Biol Chem*. 2009; 284:13004–13012. [PubMed: 19286651]
16. 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]
17. Huo L, Rachid S, Stadler M, Wenzel SC, Muller R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem Biol*. 2012; 19:1278–1287. [PubMed: 23021914]
18. Hou Y, et al. Structure and Biosynthesis of the Antibiotic Bottromycin D. *Org Lett*. 2012; 14:5050–5053. [PubMed: 22984777]
19. Gomez-Escribano JP, Song L, Bibb MJ, Challis GL. Posttranslational β -methylation and macrolactamidation in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. *Chem Sci*. 2012; 3:3522–3525.

20. Crone WJK, Leeper FJ, Truman AW. Identification and characterization of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides. *Chem Sci*. 2012; 3:3516–3521.
21. Breil BT, Ludden PW, Triplett EW. DNA sequence and mutational analysis of genes involved in the production and resistance of the antibiotic peptide trifolitin. *J Bacteriol*. 1993; 175:3693–3702. [PubMed: 8509324]
22. 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]
23. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. *Nucleic Acids Res*. 2010; 38:W545–549. [PubMed: 20457744]
24. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13:2498–2504. [PubMed: 14597658]
25. Hunter S, et al. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res*. 2012; 40:306–312.
26. Breil B, Borneman J, Triplett EW. A newly discovered gene, *tfuA*, involved in the production of the ribosomally synthesized peptide antibiotic trifolitin. *J Bacteriol*. 1996; 178:4150–4156. [PubMed: 8763943]
27. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004; 14:1188–1190. [PubMed: 15173120]
28. 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]
29. Melby JO, Dunbar KL, Trinh NQ, Mitchell DA. Selectivity, directionality, and promiscuity in peptide processing from a *Bacillus* sp AI Hakam cyclodehydratase. *J Am Chem Soc*. 2012; 134:5309–5316. [PubMed: 22401305]
30. Denessiouk KA, Johnson MS. When fold is not important: a common structural framework for adenine and AMP binding in 12 unrelated protein families. *Proteins*. 2000; 38:310–326. [PubMed: 10713991]
31. Zhang Y, Morar M, Ealick SE. Structural biology of the purine biosynthetic pathway. *Cell Mol Life Sci*. 2008; 65:3699–3724. [PubMed: 18712276]
32. Fawaz MV, Topper ME, Firestone SM. The ATP-grasp enzymes. *Bioorg Chem*. 2011; 39:185–191. [PubMed: 21920581]
33. Rakus JF, et al. Evolution of enzymatic activities in the enolase superfamily: D-Mannonate dehydratase from *Novosphingobium aromaticivorans*. *Biochemistry*. 2007; 46:12896–12908. [PubMed: 17944491]
34. Chen W, Biswas T, Porter VR, Tsodikov OV, Garneau-Tsodikova S. Unusual regioversatility of acetyltransferase Eis, a cause of drug resistance in XDR-TB. *Proc Natl Acad Sci USA*. 2011; 108:9804–9808. [PubMed: 21628583]
35. Selvy PE, Lavieri RR, Lindsley CW, Brown HA. Phospholipase D: enzymology, functionality, and chemical modulation. *Chem Rev*. 2011; 111:6064–6119. [PubMed: 21936578]
36. Bhatnagar RS, Futterer K, Waksman G, Gordon JI. The structure of myristoyl-CoA:protein N-myristoyltransferase. *Biochim Biophys Acta*. 1999; 1441:162–172. [PubMed: 10570244]
37. Climie SC, Carreras CW, Santi DV. Complete replacement set of amino acids at the C-terminus of thymidylate synthase: quantitative structure-activity relationship of mutants of an enzyme. *Biochemistry*. 1992; 31:6032–6038. [PubMed: 1627545]
38. 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]
39. Donia MS, et al. Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat Chem Biol*. 2006; 2:729–735. [PubMed: 17086177]
40. Izawa M, Kawasaki T, Hayakawa Y. Cloning and heterologous expression of the thioviridamide biosynthesis gene cluster from *Streptomyces olivoviridis*. *Appl Environ Microbiol*. 2013; 79:7110–7113. [PubMed: 23995943]
41. Sievers F, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011; 7:539. [PubMed: 21988835]

42. Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLOS One*. 2009; 4:e4345. [PubMed: 19190775]
43. Tamura K, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011; 28:2731–2739. [PubMed: 21546353]
44. Larkin MA, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 23:2947–2948. [PubMed: 17846036]
45. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model*. 2011; 51:2778–2786. [PubMed: 21919503]
46. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res*. 2010; 38:W529–533. [PubMed: 20478830]
47. Bond CS, Schuttelkopf AW. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr, Sect D: Biol Crystallogr*. 2009; 65:510–512. [PubMed: 19390156]
48. Vonrhein C, et al. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr, Sect D: Biol Crystallogr*. 2011; 67:293–302. [PubMed: 21460447]
49. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr, Sect D: Biol Crystallogr*. 2010; 66:213–221. [PubMed: 20124702]
50. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc*. 2008; 3:1171–1179. [PubMed: 18600222]
51. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr, Sect D: Biol Crystallogr*. 2010; 66:486–501. [PubMed: 20383002]
52. Vagin AA, et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr, Sect D: Biol Crystallogr*. 2004; 60:2184–2195. [PubMed: 15572771]
53. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst*. 1993; 26:283–291.

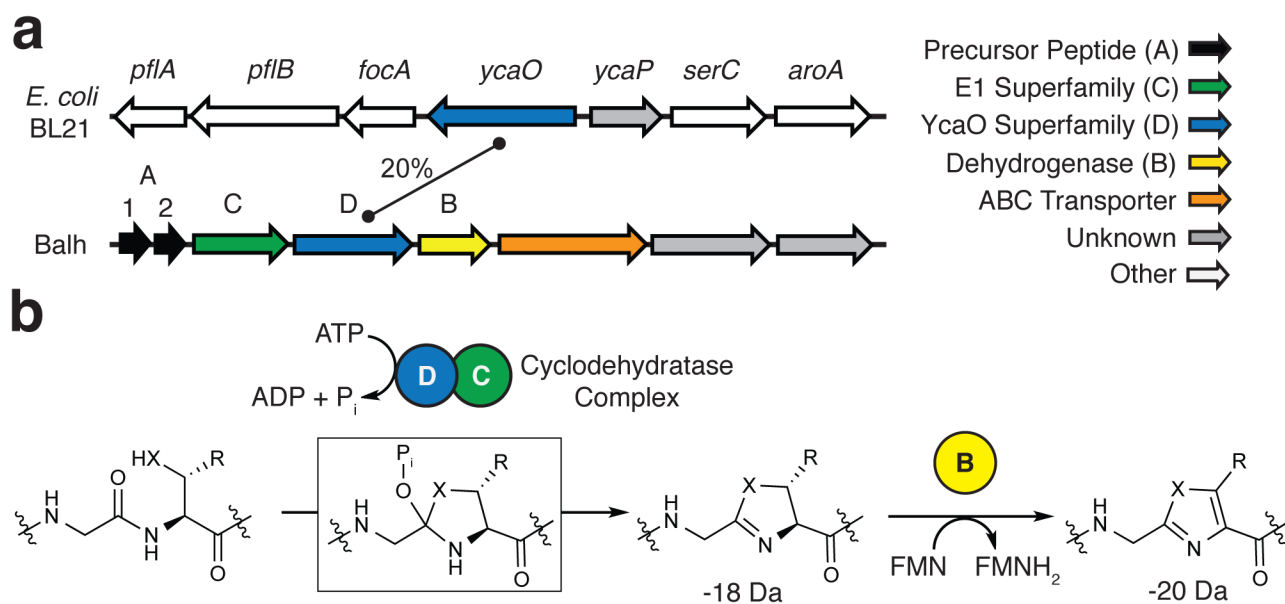


Figure 1. YcaO gene clusters and characterized roles of YcaO proteins

(a) The local genomic environment for the *E. coli* non-TOMM YcaO (Ec-YcaO) and the *Bacillus* sp. Al Hakam TOMM YcaO (BalhD) is depicted along with the % amino acid identity for the YcaOs. Gene assignments are shown. (b) Azole heterocycles in TOMM natural products are installed by the successive action of a cyclodehydratase (C and D proteins) and a flavin mononucleotide (FMN)-dependent dehydrogenase (B protein). The cumulative mass change for each step is shown below the modification. X = S, O; R = H, CH₃.

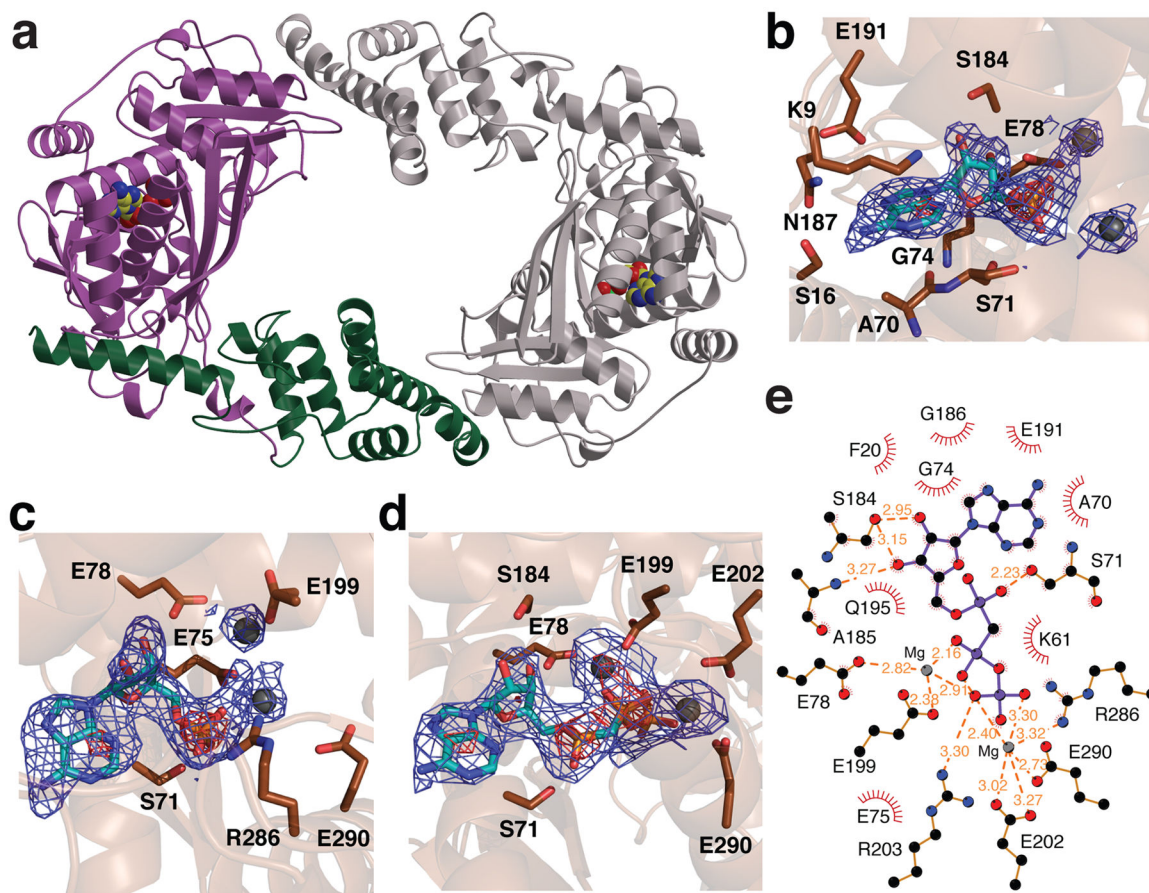


Figure 2. Structure of Ec-YcaO and ATP-binding residues of Ec-YcaO

(a) Structure of the Ec-YcaO homodimer with one monomer colored in purple (ATP-binding domain) and green (TPR domain) and the other monomer colored in gray. AMP is displayed as spheres in both monomers. Orthogonal views of the Ec-YcaO AMP-bound structure in the vicinity of the active site, showing residues responsible for adenine/ribose binding (b) and Mg^{2+}/P_i binding (c). AMP and Mg^{2+} ions are colored cyan and gray, respectively. The superimposed difference Fourier maps are contoured at levels of 2.0σ (blue) and 8.0σ (red). (d) AMPCPP (cyan) and Mg^{2+} (grey) bound in the Ec-YcaO active site coordinates with superimposed difference Fourier maps contoured at levels of 2.0σ (blue) and 6.0σ (red). Residues responsible for P_i , ribose and Mg^{2+} binding are indicated. (e) A ligand interaction diagram for the AMPCPP-bound structure is shown. Putative hydrogen bonds are displayed in orange with distances indicated, while red arcs denote hydrophobic interactions. Due to slight differences in residue orientation in the monomer subunits, only a subset of the interactions is displayed for clarity.



Figure 3. Conservation of the Ec-YcaO ATP-binding residues in the superfamily

(a) WebLogo frequency plots for the ATP- and Mg²⁺-binding residues of YcaO domains for each subclass (TOMM, non-TOMM, TfuA non-TOMM). Due to the high level of diversity in the sequences, WebLogos for the N-terminal ATP-binding residues were not generated. The ATP-binding motif identified in Ec-YcaO is displayed above each of the ATP-binding regions with the number representing the residue in Ec-YcaO, and conserved residues colored orange. (b) YcaO superfamily sequence conservation was mapped onto the structure of Ec-YcaO, which highlights strong conservation of the ATP-binding region.

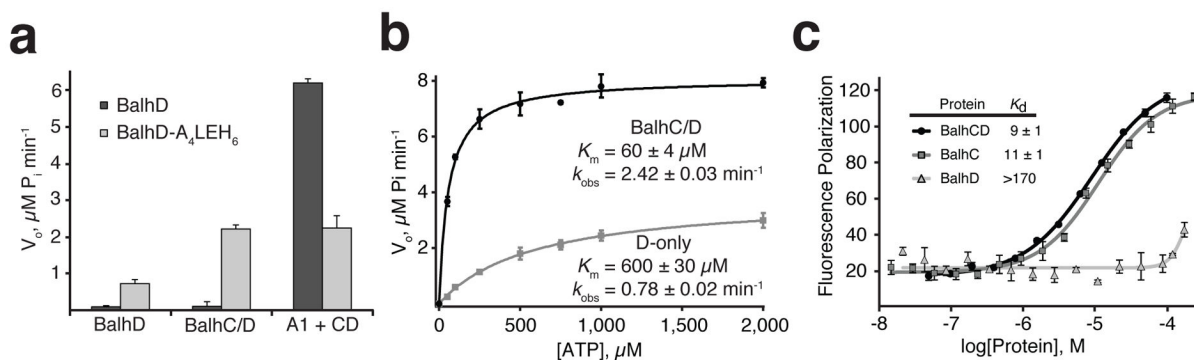


Figure 4. BalhC modulates ATP binding and hydrolysis by BalhD and is responsible for leader peptide binding

(a) ATP hydrolysis rates measured by the PNP assay. **(b)** ATP kinetic curves for BalhD- $A_4\text{LEH}_6$ with and without BalhC. Error bars represent the standard deviation from the mean ($n=3$), while error on the Michaelis-Menten parameters represents the standard deviation from the regression analysis. **(c)** A fluorescent polarization curve for fluorescein-labeled BalhA1 leader peptide recognition by BalhC and BalhD is displayed. Error bars represent the standard error of the mean of three independent titrations. Errors on the K_d values represent the error from curve fitting.

Table 1

Mutations to the ATP-binding pocket of BalhD decrease cyclodehydratase activity.

Mutation	Ring Formation ^a			BalhAI Kinetics			ATP Kinetics		
	CD	D-only	k_{obs} , min ⁻¹	K_{M} , μM^b	$k_{\text{obs}}/K_{\text{M}}$, M ⁻¹ s ⁻¹	k_{obs} , min ⁻¹	K_{M} , μM^b	$k_{\text{obs}}/K_{\text{M}}$, M ⁻¹ s ⁻¹	
WT	(5); 100%	(2-3); 45%	12.9 ± 0.4	16 ± 2	13,000	12.2 ± 0.3	240 ± 20	850	
S72A	(5); 100%	(2); 40%	8.3 ± 0.2	11 ± 1	12,500	8.0 ± 0.4	360 ± 60	370	
E76A	(0); 0%	(0); 0%	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	
E79A	(0); 0%	(0); 0%	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	
R80A	(2-4); 63%	(0-1); 10%	0.79 ± 0.05	16 ± 5	823	0.88 ± 0.02	620 ± 50	24	
Q186A	(5); 100%	(0-2); 10%	3.9 ± 0.1	12 ± 1	5,400	7.5 ± 0.3	2,500 ± 200	50	
N190A	(5); 100%	(0-2); 15%	5.1 ± 0.2	27 ± 3	3,150	4.7 ± 0.2	920 ± 110	85	
E194A	(0); 0%	(0); 0%	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	
E197A	(0); 0%	(0); 0%	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	
R198A	(2-4); 40%	(0); 0%	3.3 ± 0.2	50 ± 8	1,100	2.4 ± 0.2	760 ± 65	52	
K281A	(5); 100%	(2-3); 45%	7.9 ± 0.2	16 ± 2	8,200	6.1 ± 0.1	230 ± 20	440	
E286A	(2-4); 64%	(0); 0%	0.49 ± 0.03	27 ± 3	272	0.42 ± 0.01	1,200 ± 100	6	

^a % processing = (5·P₅ + 4·P₄ + 3·P₃ + 2·P₂ + 1·P₁)/(5); where P_X is the percentage of the substrate with X number of azolines. The number of heterocycles formed in the assay is listed in parentheses.

^b Apparent K_{M} .

^c Not determined.

Error on the Michaelis-Menten parameters represents the standard deviation from the regression analysis.

Table 2

Mutations to the C-terminus of BalhD disrupt catalysis.

Mutation	Ring Formation ^a			BalhA1 Kinetics			ATP Kinetics		
	CD	D-only	k_{obs} , min ⁻¹	K_{M1} , μ M ^b	k_{obs}/K_{M1} , M ⁻¹ s ⁻¹	k_{obs} , min ⁻¹	K_{M1} , μ M ^b	k_{obs}/K_{M1} , M ⁻¹ s ⁻¹	
WT	(5); 100%	(1-3); 45%	12.9 ± 0.4	16 ± 2	13,000	12.2 ± 0.3	240 ± 20	850	
P425*	(0); 0%	(0); 0%	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	
P427*	(0); 0%	(0); 0%	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	
P429*	(1-2); 36%	(0); 0%	0.19 ± 0.03	80 ± 20	31	0.21 ± 0.05	110 ± 10	32	
2 AA	(0); 0%	(0); 0%	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	
PxPxPG	(0); 0%	(0); 0%	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	
H426A	(0-2); 25%	(0); 0%	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	
P427G	(3-5); 80%	(0); 0%	0.43 ± 0.04	35 ± 8	204	0.41 ± 0.02	700 ± 100	10	
F428A	(3-4); 80%	(0); 0%	0.70 ± 0.02	22 ± 2	530	0.69 ± 0.03	480 ± 60	24	
P429G	(4); 80%	(0-2); 6%	4.7 ± 0.1	14 ± 1	5,600	5.0 ± 0.1	75 ± 4	1,100	
GxGxG	(0-3); 45%	(0); 0%	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	n.d. ^c	

^a % processing = (5·P5 + 4·P4 + 3·P3 + 2·P2 + 1·P1)/(5); where Px is the percentage of the substrate with × number of azolines. The number of heterocycles formed in the assay is listed in parentheses.^b Apparent K_M .^c Not determined.

* stop codon.

Error on the Michaelis-Menten parameters represents the standard deviation from the regression analysis.