Supplementary Information for:

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

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

\textsuperscript{1}Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. \textsuperscript{2}Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. \textsuperscript{3}Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. \textsuperscript{4}Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. \textsuperscript{5}Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA.

\textsuperscript{†}These authors contributed equally to this work

\textsuperscript{*}e-mail: douglasm@illinois.edu, phone: 1-217-333-1345, fax: 1-217-333-0508
\textsuperscript{*}e-mail: snair@uiuc.edu, phone: 1-217-333-0641, fax: 1-217-244-5858

**Supplementary Results**

**Table of Contents:**

Supplementary Table 1: Oligonucleotide primers used in this study

Supplementary Table 2: Crystallographic statistics for Ec-YcaO structures

Supplementary Table 3: List of peptides used in this study

Supplementary Figure 1: The biosynthetic gene clusters for trifolitoxin and bottromycin do not contain a recognizable TOMM C protein

Supplementary Figure 2: Coomassie-stained SDS-PAGE gel of proteins described in this study

Supplementary Figure 3: Ec-YcaO hydrolyzes ATP to AMP and pyrophosphate

Supplementary Figure 4: Structural homology between Ec-YcaO and TruD

Supplementary Figure 5: Cytoscape sequence similarity network for the YcaO superfamily

Supplementary Figure 6: Diversity oriented Maximum likelihood tree for the YcaO family

Supplementary Figure 7: The canonical E1 domain ATP-binding site is not conserved in TOMM C proteins

Supplementary Figure 8: The ATP-binding pocket is conserved in characterized TOMM YcaOs
Supplementary Figure 9: Multiple sequence alignment of Ec-YcaO with diverse TOMM YcaOs

Supplementary Figure 10: Mutations to the BalhD ATP-binding site affect heterocycle formation

Supplementary Figure 11: Mutations to the BalhD ATP-binding site affect ATP hydrolysis

Supplementary Figure 12: Mutations to the ATP-binding site that decrease cyclodehydratase activity also affect BalhD-only activity

Supplementary Figure 13: The $K_M$ for ATP does not depend on the concentration of BalhA1

Supplementary Figure 14: Mutations of the Mg$^{2+}$-binding residues prevent ATPase potentiation by BalhC

Supplementary Figure 15: Effect of the ATP-binding site mutations on BalhC binding

Supplementary Figure 16: The C-terminal PxPxP motif conserved in TOMM YcaOs is near the ATP-binding site in TruD

Supplementary Figure 17: The PxPxP motif of BalhD is critical for cyclodehydratase activity

Supplementary Figure 18: Mutations that relocate the C-terminus of BalhD prevent ATPase potentiation by BalhC

Supplementary Figure 19: C-terminal truncations of BalhD affect BalhC binding

Supplementary Figure 20: Removal of the C-terminal residue of BalhD dysregulates the order of azole formation

Supplementary Figure 21: Mutagenesis of the C-terminus of BalhD affects cyclodehydratase activity

Supplementary Figure 22: The sequence of the C-terminus of BalhD is important for binding BalhC

Supplementary Figure 23: BalhD P429G displays an altered order of heterocycle formation

Supplementary Figure 24: Mutations to the C-terminus of BalhD also affect BalhC-independent catalysis

Supplementary Figure 25: The conserved TOMM YcaO C-terminus is critical for activity in the microcin B17 cyclodehydratase

Supplementary Figure 26: A C-terminal His$_6$ tag on BalhD abolishes cyclodehydratase activity

Supplementary Figure 27: An updated model for azoline formation by TOMM cyclodehydratases
Supplementary Table 1. Oligonucleotide primers used in this study. Mutations are listed in parentheses, asterisks denote stop codons, and endonuclease cut sites are listed when appropriate. The *430A mutation indicates that the BalhD stop codon was mutated to alanine. This resulted in an additional 12 residues to the C-terminus of BalhD (-A418/419). F, forward primer; R, reverse primer.

| Primer | Sequence | Cut Site |
|--------|----------|----------|
| YcaO F | AAGGATCCATGACGCAAACATTTATCCCGCGGCAAG | BamHI |
| YcaO R | TTGGCGGCGGCTTATTTTGCGCGCGCGCGGCACGAGGCG | NotI |
| McbD (M392*) F | GAGAATCAAAGTAGGTACCATTCCCATAAAAGCTGCGCCCGGCCGAGGCG | |
| McbD (M392*) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD F | CGCGATCCATGGAAGGCGGCAAGTGGCAAAATTCTTTGAGGCGTGAATTC | |
| BalhD (S72A) F | CAGCACTAATAGCGCGCTTACATTCCCATAAAAGCTGCGCCCGGCCGAGGCG | |
| BalhD (S72A) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (E76A) F | CGCGATCCATGGAAGGCGGCAAGTGGCAAAATTCTTTGAGGCGTGAATTC | |
| BalhD (E76A) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (E79A) F | CGCGATCCATGGAAGGCGGCAAGTGGCAAAATTCTTTGAGGCGTGAATTC | |
| BalhD (E79A) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (R80A) F | CGCGATCCATGGAAGGCGGCAAGTGGCAAAATTCTTTGAGGCGTGAATTC | |
| BalhD (R80A) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (Q186A) F | CGCGATCCATGGAAGGCGGCAAGTGGCAAAATTCTTTGAGGCGTGAATTC | |
| BalhD (Q186A) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (R198A) F | CGCGATCCATGGAAGGCGGCAAGTGGCAAAATTCTTTGAGGCGTGAATTC | |
| BalhD (R198A) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (K281A) F | CGCGATCCATGGAAGGCGGCAAGTGGCAAAATTCTTTGAGGCGTGAATTC | |
| BalhD (K281A) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (P427G) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (F428A) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (H426A) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (PxPxPG) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (PxGxG) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (GxGxG) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (PxPxPG) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (PxGxG) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (GxGxG) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (PxPxPG) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (PxGxG) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (GxGxG) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (PxPxPG) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (PxGxG) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
| BalhD (GxGxG) R | GGGAATGGTACCTACTTTGATTCGTGACTTAAACGCTGCCCCCGCCG | |
Supplementary Table 2. Crystallographic statistics for Ec-YcaO structures.

|                           | Ec-YcaO APO   | Ec-YcaO AMP | Ec-YcaO AMPCPP |
|---------------------------|---------------|-------------|-----------------|
| **Data collection**       |               |             |                 |
| Space group               | P2 2 1        | P1          | P1              |
| Cell dimensions           | 68.96, 140.33, 163.49 | 111.25, 112.95, 132.89 | 110.28, 112.40, 130.68 |
| α, β, γ (°)               | 90.0, 90.0, 90.0 | 89.94, 73.51, 77.23 | 89.40, 73.62, 77.62 |
| Resolution (Å)            | 50.0-2.65 (2.70-2.65) | 127.14-2.25 (2.26-2.25) | 125.20-3.29 (3.30-3.29) |
| Rsym                      | 9.4 (68.2)    | 10.6 (71.0) | 17.6 (59.3)     |
| l / σl                    | 16.5 (2.3)    | 10.6 (2.0)  | 7.0 (2.1)       |
| Completeness (%)          | 99.5 (99.6)   | 93.1 (66.1) | 99.0 (98.2)     |
| Redundancy                | 6.0 (6.1)     | 3.8 (3.9)   | 2.5 (2.5)       |
| **Refinement**            |               |             |                 |
| Resolution (Å)            | 50.0-2.64     | 127.14-2.25 | 125.20-3.29     |
| No. reflections           | 43,214        | 251,531     | 84,143          |
| Rwork / Rfree             | 0.1628/0.2185 | 0.2178/0.2500 | 0.1896/0.2377 |
| No. atoms                 |               |             |                 |
| Protein                   | 9,027         | 36,574      | 35,815          |
| Nucleotide+Metal          | 4             | 182         | 174             |
| Water                     | 512           | 1,987       | 432             |
| B-factors                 |               |             |                 |
| Protein                   | 43.95         | 40.03       | 64.89           |
| Nucleotide                | 88.88         | 51.83       | 23.18           |
| Metal                     | 88.88         | 25.93       | 33.68           |
| Water                     | 39.21         | 33.71       |                 |
| R.m.s. deviations         |               |             |                 |
| Bond lengths (Å)          | 0.0141        | 0.0119      | 0.0113          |
| Bond angles (°)           | 1.5489        | 1.4810      | 1.5215          |

Each dataset was derived from one crystal.
Supplementary Table 3. List of peptides used in this study. Residues known to be cyclized in vitro or in vivo are colored orange. In McbA, the orange underlined serine (CSN) is the site of the ninth heterocycle that is installed in vitro and found as a minor microcin B17 species. A caret and an asterisk denote putative and known leader peptide cleavage sites, respectively. FP-BalhA1-LP is a synthetically prepared reagent designed for fluorescent polarization (FP) studies. The reagent contains a fluorescein label (installed via the amine reactive isothiocyanate, FITC) linked to an aminohexanoic-glycine spacer, which in turn is linked to the predicted leader peptide (LP) of BalhA1.

| Peptide   | Sequence                                                                          |
|-----------|------------------------------------------------------------------------------------|
| BalhA1    | MEQKKILDIKLTETGKINYAHKPDD^SGCAGCMGCAGGTGCAGTGCGIGQGVWKKCSGK                       |
| FP-BalhA1-LP | FITC-Ahx-GMEQKKILDIKLTETGKINYAHKPDD                           |
| McbA      | MELKASEFGVVLSDLKLSRQSPLG^VGIGGGGGGGGGGGGGCGGGQGGGCGGCGSNGCSGGNGGSGGSGSHI            |
Supplementary Figure 1. The biosynthetic gene clusters for trifolitoxin and bottromycin do not contain a recognizable TOMM C protein. (a) The gene clusters for bottromycin A2 and trifolitoxin biosynthesis are displayed along with the amino acid sequence of the precursor peptide. Although each cluster contains at least one YcaO homolog, neither cluster contains a recognizable TOMM C protein (homolog of the E1 ubiquitin-activating enzyme superfamily)\textsuperscript{2-6}. Asterisks indicate the leader and follower peptide cleavage sites. (b) The structure of bottromycin A2 and the partial structure of trifolitoxin are displayed. Post-translation modifications presumably installed by the YcaO proteins encoded in each cluster are colored blue. The red X’s in in the trifolitoxin structure denote an uncharacterized post-translational modification involving Arg and Gln\textsuperscript{7}.
Supplementary Figure 2. Coomassie-stained SDS-PAGE gel of proteins described in this study. MBP-tagged proteins were purified by amylose affinity chromatography. Coomassie-stained 12% SDS-PAGE gels for MBP-Ec-YcaO (a) and the YcaO mutants described in this study (b) are shown. Masses for pertinent bands of the molecular marker are shown. The faint bands appearing just below the 50 kDa marker band likely result from proteolysis of MBP fusion partner. Asterisks denote stop codons. The dashes (- - - - -) indicate an empty lane.
Supplementary Figure 3. Ec-YcaO hydrolyzes ATP to AMP and pyrophosphate. (a) The ATPase activity of Ec-YcaO was screened using the purine nucleoside phosphorylase (PNP)-coupled assay with and without 1 unit of pyrophosphatase (PPase). The addition of PPase increased the rate of chromophore production in the assay by 5-fold after the signal amplification achieved during PPi cleavage is taken into account. These data indicate that ATP is preferentially hydrolyzed to AMP and pyrophosphate by Ec-YcaO in vitro. This result is corroborated by the observation that co-crystallization of Ec-YcaO with ATP yielded an AMP-bound structure (Fig. 2A). Error bars represent the standard deviation from the mean (n=3). (b) An ATP kinetic curve was obtained for Ec-YcaO using the PPase-supplemented PNP assay. Error bars represent the standard deviation from the mean (n=3). Regression analyses to obtain Michaelis-Menten kinetic parameters were carried out in IGOR Pro version 6.12 (Wavemetrics). The error on the kinetic parameters represents the standard deviation from the curve fitting.

![Graph a](image1.png)

![Graph b](image2.png)
Supplementary Figure 4. Structural homology between Ec-YcaO and TruD. (a) Structural alignment of TruD (green) and Ec-YcaO (purple). The alignment demonstrates that the ATP-binding region of Ec-YcaO is also conserved in TruD. (b) TruD structure with the conserved domain shown in green. (c) Ec-YcaO structure with the conserved region shown in purple. (d) Structure-based sequence alignment of YcaO domain containing proteins with the secondary structure of Ec-YcaO superimposed. Residues mutated in BalhD for this study are denoted by black triangles. A summary of the BalhD mutants made, along with their predicted roles in ATP recognition is provided.
| Mutation | Rationale                              | Mutation | Rationale                              |
|----------|----------------------------------------|----------|----------------------------------------|
| S72A     | α-PO<sub>4</sub> binding               | E194A    | Mg<sup>1</sup> binding (AMPCPP)        |
| E76A     | α-PO<sub>4</sub> binding, Mg<sup>2</sup> binding (AMP), catalytic | E197A    | Mg<sup>2</sup> binding                |
| E79A     | Mg<sup>1</sup> binding                 | R198A    | y-PO<sub>4</sub> binding               |
| R80A     | Active site organization, catalytic    | K281A    | α-PO<sub>4</sub> binding (AMP), Mg<sup>2</sup> binding (AMPCPP) |
| Q186A    | Adenine binding                        | E286A    | Mg<sup>2</sup> Binding                |
| N190A    | 3'-OH binding, active site organization |          |                                        |

**Mutation Rationale:****

- **S72A**: α-PO<sub>4</sub> binding
- **E76A**: α-PO<sub>4</sub> binding, Mg<sup>2</sup> binding (AMP), catalytic
- **E79A**: Mg<sup>1</sup> binding
- **R80A**: Active site organization, catalytic
- **Q186A**: Adenine binding
- **N190A**: 3'-OH binding, active site organization
Supplementary Figure 5. Cytoscape sequence similarity network for the YcaO superfamily. A similarity network for non-redundant YcaO sequences is displayed. Each node represents a unique YcaO, while lines between nodes exist if the proteins bear significant similarity (in this case, a BLAST expectation value lower than $10^{-80}$). At this e-value, YcaO proteins from characterized TOMM biosynthetic clusters primarily form isofunctional clusters. TOMM YcaOs are defined as having a bioinformatically identifiable C protein within 10 kb of the YcaO protein. Stand-alone TOMM YcaOs are proteins found in a characterized TOMM biosynthetic clusters that lack a C protein (e.g. bottromycin and trifolitoxin). It is noteworthy that many non-TOMM YcaOs appear to exist in RiPP biosynthetic gene clusters, indicating that additional stand-alone YcaO cyclodehydratases may exist. Nodes are colored according to the legend and the clusters containing Ec-YcaO and BalhD are indicated.
Supplementary Figure 6. Diversity oriented Maximum likelihood tree for the YcaO family. A diversity-maximized tree for the YcaO superfamily is displayed. Red asterisks denote YcaOs associated clusters of functionally or structurally characterized natural products, while the proteins used in this study are circled and colored blue. The members of this tree were used to generate WebLogo\textsuperscript{9} frequency plots for the conservation of the ATP-binding residues in TOMM, non-TOMM, and TfuA-associated non-TOMM subclasses of the YcaO superfamily (Fig. 3A).
Supplementary Figure 7. The canonical E1 domain ATP-binding site is not conserved in TOMM C proteins. (a) Residues responsible for ATP- and Mg\(^{2+}\)-binding in E1 superfamily members were identified in the nucleotide-bound crystal structures for the following E1 superfamily members: ThiF (1ZFN), MoeB (1JWA), and MccB (3H5N). An alignment of diverse non-TOMM E1 proteins was generated and the ATP- and Mg\(^{2+}\)-binding residues identified in these crystal structures were highlighted in orange. The two CxxC motifs responsible for coordination of the structural Zn\(^{2+}\) found in all characterized E1 ubiquitin-activating family members are highlighted in blue.\(^{15,17}\) To improve the alignments, the highly variable N-terminal “peptide clamp” domain found in MccB homologs was manually removed. A sequence identity matrix is provided to demonstrate the level of divergence between members of the family. In spite of the divergence, the alignments demonstrate that the ATP-binding pocket is remarkably conserved among these non-TOMM E1 members. (b) Alignments of TOMM C proteins associated with cyanobactin biosynthesis to E. coli MoeB, ThiF and MccB are displayed. The highly variable N-terminal residues of each C protein were removed as above. Additionally, the YcaO domain of each CD fusion was removed. Unlike the non-TOMM E1 superfamily members, the TOMM C proteins show little conservation in the canonical E1-like ATP-binding site. This suggests that the ability to bind/utilize ATP has been lost in the TOMM C proteins. While non-TOMM E1 sequences are identified by GI number and the subclass of the E1 superfamily they belong to, TOMM E1 sequences are identified by their natural product that they produce.
Supplementary Figure 8. The ATP-binding pocket is conserved in characterized TOMM YcaOs. All of the YcaO sequences from the indicated TOMM subclasses were aligned with Clustal Omega and the multiple sequence alignments were used to generate sequence logos (WebLogos) of the regions involved in ATP binding. The size of the letter is proportional to its level of conservation. The ATP-binding motif identified in Ec-YcaO is displayed above each of the ATP-binding regions. Residues that are similar to the ATP-binding motif identified in Ec-YcaO are colored orange. Due to the high level of diversity in the sequences, WebLogos for the N-terminal ATP-binding residues could not be generated. The number of sequences represented in each WebLogo is displayed in parentheses. The ATP-binding motif of Ec-YcaO is conserved in diverse TOMM YcaOs.
Supplementary Figure 9. Multiple sequence alignment of Ec-YcaO with diverse TOMM YcaOs. Sequences of TOMM YcaOs from characterized natural product clusters were aligned with Ec-YcaO. The C domain of the naturally occurring CD fusion proteins (TruD, GodD, TsrH) was manually removed prior to alignment. Residues implicated in ATP- and Mg\(^{2+}\)-binding in Ec-YcaO are highlighted in orange. The conserved proline rich C-terminus of TOMM YcaOs is colored blue. A sequence identity matrix is displayed to show the divergence of the proteins in the alignment. The alignment indicates that the ATP-binding site is under intense selective pressure. McbD (microncin B17), YcaO (Ec-YcaO), TruD (trunkamide), GodD (goadsporin), TsrH (thiostrepton), SagD (streptolysin S), BalhD (unknown), PznD (plantazolicin).
Supplementary Figure 10. Mutations to the BalhD ATP-binding site affect heterocycle formation. A MALDI-TOF MS spectral overlay for BalhA1 treated with BalhC and either wild-type (WT) or mutant BalhD is displayed. Many of the mutations to the ATP-binding residues, which are listed on the right, decreased cyclodehydratase activity. The level of processing is summarized in Table 1. †, laser-induced deamination. A loss of 90 Da indicates the formation of 5 azolines, a full in vitro processed BalhA1 substrate. 

![MALDI-TOF MS overlay with Spectral Peaks](image-url)
Supplementary Figure 11. Mutations to the BalhD ATP-binding site affect ATP hydrolysis. The rate of ATP hydrolysis was measured with BalhC and either wild-type (WT) or mutant BalhD using the PNP assay. Error represents the standard deviation from the mean (n ≥ 3). Mutations showing the slowest rate of ATP hydrolysis also installed fewer heterocycles on BalhA1 in an endpoint assay (Supplementary Fig. 10).
Supplementary Figure 12. Mutations to the ATP-binding site that decrease cyclodehydratase activity also affect BalhD-only activity. MALDI-TOF MS spectra for BalhA1 treated with wild-type (WT) or mutant BalhD is displayed. Apart from BalhD S72A, all of the mutants displayed a decreased level of D-only processing consistent with the rate of ATP hydrolysis measured in the presence of the C protein (Supplementary Fig. 11). The level of processing is summarized in Table 1. †, laser-induced deamination.
Supplementary Figure 13. The $K_M$ for ATP does not depend on the concentration of BalhA1. Due to solubility limitations, saturating concentrations of BalhA1 were not obtainable for all of the BalhD mutants. To determine if the $K_M$ for ATP changed at non-saturating concentrations of BalhA1, an ATP kinetic curve was carried out at the $K_M$ for BalhA1 (15 µM). The resultant $K_M$ is within error of the previously reported $K_M$ for ATP, 240 ± 20 µM, obtained with a saturating concentration of BalhA1\textsuperscript{18}. Error on the Michaelis-Menten parameters represents the standard deviation from the regression analysis.

\begin{align*}
V_0, \text{µM P min}^{-1} & = 6.75 \pm 0.01 \text{ min}^{-1} \\
K_M & = 210 \pm 10 \text{ µM}
\end{align*}
Supplementary Figure 14. Mutations of the Mg$^{2+}$-binding residues prevent ATPase potentiation by BalhC. The background ATP hydrolysis rate of mutant BalhD proteins lacking the ability to modify BalhA1 was measured with and without BalhC using the PNP assay. Although the ATPase activity of wild-type BalhD was increased 2.5-fold by the addition of BalhC, none of the mutants displayed similar increases in ATP hydrolysis. Error bars represent the standard deviation to the mean (n ≥ 3).
Supplementary Figure 15. Effect of the ATP-binding site mutations on BalhC binding. (a) Size-exclusion chromatography was used to monitor the effect the ATP-binding site mutations had on complex formation. Although BalhD retained as a monomer (3), BalhC had a retention volume consistent with a dimer (2). When the proteins were mixed, the intensity of the BalhC dimer peak decreased and a new peak corresponding to the BalhC/D complex appeared (1). Absorbance values were normalized to the peak at ~3.5 mL. (b) To better quantify the BalhC binding affinity of the mutant BalhD proteins, ATP hydrolysis rates were monitored for wild-type cyclodehydratase reactions with the addition of variable concentrations of inactive BalhD mutants. As the reactions were carried out with 1 µM wild-type BalhD, mutants with an unperturbed BalhC interaction would display IC$_{50}$ values of approximately 1 µM. Error bars represent the standard deviation from the mean (n ≥ 3). Regression analysis to determine the IC$_{50}$ values for each inhibition curve was determined in IGOR Pro version 6.12 (Wavemetrics). The error on the IC$_{50}$ values represents the standard deviation from the curve fitting.

![Chromatography Peaks](image)

![Activity Graphs](image)
Supplementary Figure 16. The C-terminal PxPxP motif conserved in TOMM YcaOs is near the ATP-binding site in TruD. (a) Alignment of C-termini of select TOMM YcaO proteins. Prolines are colored blue and CD fusion proteins are underlined. McbD: microcin B17, SagD: streptolysin S, TruD: trunkamide, BalhD: Bacillus sp Al Hakam, Pzn: plantazolicin, TsrH: thiostrepton. (b) Graphs displaying the proline content of the C-terminus of all members of the YcaO superfamily (TOMM and non-TOMM) are shown. (c) The structure TruD<sup>12</sup> (PDB code 4BS9) is displayed. Residues implicated in ATP-binding are colored cyan while the PxPxP motif is colored orange. A zoom in of the YcaO active site (red circle) is displayed in panel d. The C-terminus of TruD is proximal (<8 Å) to the conserved ExxERD and RxxE motifs.

![Figure 16a](image1.png)

![Figure 16b](image2.png)

![Figure 16c](image3.png)

![Figure 16d](image4.png)
Supplementary Figure 17. The PxPxP motif of BalhD is critical for cyclodehydratase activity. (a) A MALDI-TOF MS spectral overlay for BalhA1 treated with BalhC and either wild-type (WT) or mutant BalhD is displayed. The level of processing is summarized in Table 2. (b) The rate of ATP hydrolysis was measured with BalhC and either wild-type (WT) or mutant BalhD using the PNP assay. Error represents the standard deviation from the mean (n ≥ 3). *, stop codons; †, laser-induced deamination; Δ2 AA, BalhD (ΔA418, K419).
Supplementary Figure 18. Mutations that relocate the C-terminus of BalhD prevent ATPase potentiation by BalhC. The background ATP hydrolysis rate of BalhD C-terminus mutants lacking the ability to modify BalhA1 was measured with and without BalhC. Although the mutants displayed a higher rate of ATP hydrolysis than wild-type BalhD, none of the mutants could be potentiated by the addition of BalhC (P > 0.1). Error bars represent the standard deviation to the mean (n ≥ 3). For reactions with BalhC, the horizontal lines indicate the expected rate of ATP hydrolysis with the background ATPase rate of BalhC. *, stop codons; Δ2 AA, BalhD (ΔA418, K419).
Supplementary Figure 19. C-terminal truncations of BalhD affect BalhC binding. (a) Size-exclusion chromatography was used to monitor the effect the C-terminal truncations had on complex formation. Although BalhD retained as a monomer (3), BalhC had a retention volume consistent with a dimer (2). When the proteins were mixed, the intensity of the BalhC dimer peak decreased and a new peak corresponding to the BalhC/D complex appeared (1). Absorbance values were normalized to the peak at ~3.5 mL. (b) ATP hydrolysis rates were monitored for wild-type cyclodehydratase reactions with the addition of variable concentrations of a C-terminal BalhD truncated protein. As the reactions were carried out with 1 µM wild-type BalhD, mutants with an unperturbed BalhC interaction would display IC$_{50}$ values of approximately 1 µM. *, stop codon. Error bars represent the standard deviation from the mean (n ≥ 3). Regression analysis to determine the IC$_{50}$ values for each inhibition curve was determined in IGOR Pro version 6.12 (Wavemetrics). The error on the IC$_{50}$ values represents the standard deviation from the curve fitting.
Supplementary Figure 20. Removal of the C-terminal residue of BalhD dysregulates the order of azole formation. BalhA1 was treated with BcerB, BalhC, and BalhD P429* for 16 h at 25 °C. (a) A high-resolution, intact mass spectrum of the processed substrate is displayed. The error (ppm) associated with each measurement from the calculated m/z is shown. (b) MS/MS fragmentation of the 0, 1 and 2 ring species was performed and the resultant fragmentation maps are displayed. The b- and y-ions are colored based on the number of azoles in the fragment (black = 0, orange = 1, purple = 2). Stars indicate the location of the azole heterocycles. In the di-azole species, the second heterocycle is found at both Cys45 and Cys31. The MS/MS spectrum for the di-azole species is shown in (c). Cys45 is the penultimate residue cyclized by a wild-type Balh cyclodehydratase. As such, these data indicate that the ring order of the complex has been perturbed in the P429* mutant.
Supplementary Figure 21. Mutagenesis of the C-terminus of BalhD affects cyclodehydratase activity. (a) MALDI-TOF MS spectra for BalhA1 treated with BalhC and either wild-type (WT) or mutant BalhD are displayed. These data indicate that the sequence of the C-terminus of BalhD is critical for cyclodehydratase activity. The level of processing is summarized in Table 2. (b) The rate of ATP hydrolysis was measured with BalhC and either wild-type (WT) or mutant BalhD using the PNP assay. Error represents the standard deviation from the mean (n ≥ 3). These data provided further evidence that the C-terminal residues of BalhD are critical for cyclodehydratase activity. †, laser-induced deamination. The wild-type sequence for the C-terminal PxPxP is 425PHPFP429.
Supplemental Figure 22. The sequence of the C-terminus of BalhD is important for binding BalhC. (a) Size-exclusion chromatography was used to monitor the effect the PxPxP-tail mutations had on complex formation. Although BalhD retained as a monomer (3), BalhC had a retention volume consistent with a dimer (2). When the proteins were mixed, the intensity of the BalhC dimer peak decreased and a new peak corresponding to the BalhC/D complex appeared (1). Absorbance values were normalized to the peak at ~3.5 mL. (b) ATP hydrolysis rates were monitored for wild-type cyclodehydratase reactions with the addition of variable concentrations of a catalytically deficient BalhD mutant. As the reactions were carried out with 1 µM wild-type BalhD, mutants with an unperturbed BalhC interaction would display IC\textsubscript{50} values of approximately 1 µM. All of the BalhD mutants had a disrupted BalhC affinity. (c) As the competition assay is only compatible with BalhD mutations displaying almost no activity, the interaction between F428A and P429G could not be determined with this assay. For these active mutants, the interaction between BalhC and mutant BalhD was interrogated by measuring an apparent “K\textsubscript{M}” for BalhC. A BalhC “K\textsubscript{M}” was obtained by holding the concentration of mutant BalhD and BalhA1 constant and adding variable concentrations of BalhC. The rate of processing was monitored using the PNP-phosphate detection assay. For clarity, the initial processing rates were normalized to the rate obtained with a 1:1 ratio of BalhC and mutant BalhD. Together, these data indicated that the prolines in the PxPxP motif are more important for BalhC recognition than the intervening “x” residues and that residues closer to the active site (more C-terminal) are less important for BalhC binding. Error bars represent the standard deviation from the mean (n ≥ 3). Regression analysis to determine the IC\textsubscript{50} and K\textsubscript{M} values for each inhibition curve was determined in IGOR Pro version 6.12 (Wavemetrics). The error on the IC\textsubscript{50} and K\textsubscript{M} values represents the standard deviation from the curve fitting.
Supplementary Figure 23. BalhD P429G displays an altered order of heterocycle formation. Although BalhD P429G had an initial rate of substrate processing that is 1/3 of wild-type BalhD, the mutant produced primarily a tetra-ring species following a 16 h reaction. Based on the aberrant ring order observed in BalhD P429* reactions, we hypothesized that the inability to generate the penta-azole species could be due to perturbation of ring order. To locate the sites of heterocyclization, BalhA1 was treated with BcerB, BalhC, and BalhD P429G for 16 h at 25 ºC. (a) A high-resolution, intact mass spectrum of the processed substrate is displayed. The error (ppm) associated with each measurement from the calculated m/z is shown. (b) MS/MS fragmentation of the 5, 4 and 3 ring species was performed and the resultant fragmentation maps are displayed. The b- and y-ions are colored based on the number of azoles heterocycles in the fragment (black = 0, orange = 1, purple = 2, blue = 3, pink = 4, green = 5). Stars indicate the location of the azoles. In the tri-azole species, the third heterocycle was found at both Cys45 and Cys34. The MS/MS spectrum for the tri-azole species is shown in (c). Based on the b- and y-ion intensities, both positions appeared to be cyclized to approximately the same extent. In wild-type reactions, the tri-azole species has rings at Cys31, Cys34, and Cys40. The deviation from the wild-type processing order was not observed in the tetra- or penta-azole species. Although less severe than the dysregulation in cyclization order found with BalhD P429* treated BalhA1 (Supplementary Fig. 20), the identification of partial processing at Cys45 in the tri-azole species suggested that the BalhC/D interaction was perturbed.

| Azole Type | calc. m/z | error |
|------------|-----------|-------|
| 5-Azole    | 1,115.750 | 8.9 ppm |
| 4-Azole    | 1,122.430 | 8.9 ppm |
| 3-Azole    | 1,129.110 | 6.2 ppm |

![Supplementary Figure 23](image-url)
Supplementary Figure 24. Mutations to the C-terminus of BalhD also affect BalhC-independent catalysis. A MALDI-TOF MS spectral overlay for BalhA1 treated with either wild-type (WT) or mutant BalhD is displayed. Only BalhD P429G showed detectable levels of azoline formation. These data indicated that the C-terminus of BalhD is important for catalysis in addition to BalhC recognition. The level of processing is summarized in Table 2. *, stop codon; †, laser-induced deamination; Δ2 AA, BalhD ΔA418-K419.
Supplementary Figure 25. The conserved TOMM YcaO C-terminus is critical for activity in the microcin B17 cyclodehydratase. A MALDI-TOF MS spectral overlay for McbA treated with McbB/C and either wild-type (WT) or mutant McbD lacking the C-terminal 5 amino acids is displayed. *, stop codon; †, laser-induced deamination.
Supplementary Figure 26. A C-terminal His₆ tag on BalhD abolishes cyclodehydratase activity. A MALDI-TOF MS spectral overlay for BalhA1 treated with BalhD-A₄LEH₆ either with (a) or without (b) BalhC is displayed. A reaction carried out with wild-type (WT) BalhD is shown for a comparison. †, laser-induced deamination.
Supplementary Figure 27. An updated model for azoline formation by TOMM cyclodehydratases. Based on the data from this manuscript, a model for azoline formation is provided. The TOMM E1 homolog (C protein) allosterically activates the YcaO domain (D protein) and binds the leader peptide (LP). Leader peptide binding by the C protein places the core peptide in proximity of the YcaO active site to facilitate the cyclodehydration reaction.
References
1 Sinha Roy, R., Kelleher, N. L., Milne, J. C. & Walsh, C. T. In vivo processing and antibiotic activity of microcin B17 analogs with varying ring content and altered bisheterocyclic sites. Chem. Biol. 6, 305-318 (1999).
2 Huo, L., Rachid, S., Stadler, M., Wenzel, S. C. & Muller, R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. Chem. Biol. 19, 1278-1287 (2012).
3 Hou, Y. et al. Structure and Biosynthesis of the Antibiotic Bottromycin D. Org. Lett. 14, 5050-5053 (2012).
4 Gomez-Escribano, J. P., Song, L., Bibb, M. J. & Challis, G. L. Posttranslational β-methylation and macrolactamidination in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. Chem. Sci. 3, 3522-3525 (2012).
5 Crone, W. J. K., Leeper, F. J. & Truman, A. W. Identification and characterization of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversitiy of ribosomal peptides. Chem. Sci. 3, 3516-3521 (2012).
6 Breil, B. T., Ludden, P. W. & Triplett, E. W. DNA sequence and mutational analysis of genes involved in the production and resistance of the antibiotic peptide trifolitoxin. J. Bacteriol. 175, 3693-3702 (1993).
7 Herlache, T. C. & Triplett, E. W. Biological control of crown gall disease. US 7141395 B2 (2006).
8 Webb, M. R. A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. Proc. Natl. Acad. Sci. U.S.A. 89, 4884-4887 (1992).
9 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. Genome Res. 14, 1188-1190 (2004).
10 Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539 (2011).
11 Melby, J. O., Dunbar, K. L., Trinh, N. Q. & Mitchell, D. A. Selectivity, directionality, and promiscuity in peptide processing from a Bacillus sp. Al Hakam cyclodehydratase. J. Am. Chem. Soc. 134, 5309-5316 (2012).
12 Koehnke, J. et al. The cyanobactin heterocyclase enzyme: a processive adenylase that operates with a defined order of reaction. Angew. Chem. Int. Ed. 52, 13991-13996 (2013).
13 Melby, J. O., Li, X. & Mitchell, D. A. Orchestration of enzymatic processing by thiazole/oxazole-modified microcin dehydrogenases. Biochemistry 53, 413-422 (2014).
14 Dunbar, K. L. & Mitchell, D. A. Insights into the mechanism of peptide cyclodehydrations achieved through the chemoenzymatic generation of amide derivatives. J. Am. Chem. Soc. 135, 8692-8701 (2013).
15 Schulman, B. A. & Harper, J. W. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. Nat. Rev. Mol. Cell Biol. 10, 319-331 (2009).
16 Lee, S. W. et al. Discovery of a widely distributed toxin biosynthetic gene cluster. Proc. Natl. Acad. Sci. U.S.A. 105, 5879-5884 (2008).
17 Zamble, D. B., McClure, C. P., Penner-Hahn, J. E. & Walsh, C. T. The McbB component of microcin B17 synthetase is a zinc metalloprotein. Biochemistry 39, 16190-16199 (2000).
18 Dunbar, K. L., Melby, J. O. & Mitchell, D. A. YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. Nat. Chem. Biol. 8, 569-575 (2012).