

Control of nucleophile chemoselectivity in cyanobactin YcaO heterocyclases PatD and TruD

Wenjia Gu,^a Yiwu Zheng,^b Taras Pogorelov,^{c, d, e} Satish K. Nair,^{b, d, f} and Eric W. Schmidt^{*a}

Members of the YcaO superfamily are among the most common post-translational modification enzymes in natural product biosynthesis, with wide usage in biotechnology and synthetic biology applications. Here, we define key amino acids that modulate the use of sulfhydryl and/or hydroxyl nucleophiles by the widely used cyanobactin YcaO enzymes. Using domain swapped chimeras we discovered unstructured regions in cyanobactin YcaOs that guide interactions with the substrates, governing access to interior amino acids in the substrates and explaining the chemoselectivity between PatD and TruD. These results demonstrate new strategies for precision engineering of YcaO catalysts and for the rapid discovery of catalysts with desired properties from nature.

Introduction

YcaO proteins are among the most common post-translational modification enzymes in bacteria/archaea, with about 30,000 entries in GenBank (2020). YcaO proteins play several roles in biochemistry, but they are perhaps best known for enabling modifications in the biosynthesis of bioactive natural products, the ribosomally synthesized and post-translationally modified peptides (RiPPs).^{1,2} Within the RiPP class, YcaOs catalyze diverse reactions to modify simple, inactive peptides into elaborate bioactive species. In an ATP-dependent manner, YcaOs use nucleophiles including amine,^{3–6} sulfide,⁷ hydroxyl,^{8–14} and sulfhydryl^{12,15–17} to displace the backbone carbonyl oxygen of peptides, releasing AMP and pyrophosphate or ADP and phosphate in the process. The resulting bioactive natural product peptides contain thioamide, amidine, heterocycles such as thiazoline and oxazoline, or other functional groups that are not found in most peptides.

Here, we sought to answer a seemingly simple question, but one that has long eluded the field: how do YcaO proteins control nucleophile selectivity? A key example of the importance of this question is found in cyanobactin heterocyclase/cyclodehydratase YcaO enzymes.¹² YcaO heterocyclases related to PatD use the β -nucleophilic side chains of Cys, Ser, or Thr residues to install both thiazoline and oxazoline moieties. By contrast, YcaO heterocyclases related to TruD are restricted to modifications on Cys side chains and synthesize only thiazoline (Fig. 1A). These enzymes are 88% sequence identical, potentially enabling the identification of individual residues responsible for nucleophile selectivity.

Peptide substrates of TruD and PatD have several key features that are essential to the enzyme-substrate interaction. The substrate precursor peptides contain recognition sequences (RSs) that guide the interactions with enzymes.^{11,16–21} Core peptides are short sequences, embedded within precursor peptides, which are the site of all enzymatic modifications. The core peptide encodes the bioactive natural product, while all other elements in the precursor peptide are

proteolytically cleaved during biosynthesis. Accordingly, core peptide sequences can be hypervariable, while RSs that are later cleaved and act to recruit the post-translational modification enzymes are highly conserved.^{11,16,22,23}

Crystal structures of representative YcaO heterocyclases have identified elements that guide the recognition of peptides and substrates.^{17,24} TruD (and the homologous PatD) contains an N-terminal RiPP precursor peptide recognition element (RRE; residues 2-85), which binds to a recognition sequence (RS) element in the substrate precursor peptide, PatE. In addition, TruD contains an E1-like domain containing a structural zinc ion (residues 86-316) and the C-terminal YcaO domain (residues 316-784) to which all catalytic functions have been attributed (Fig. 1B).

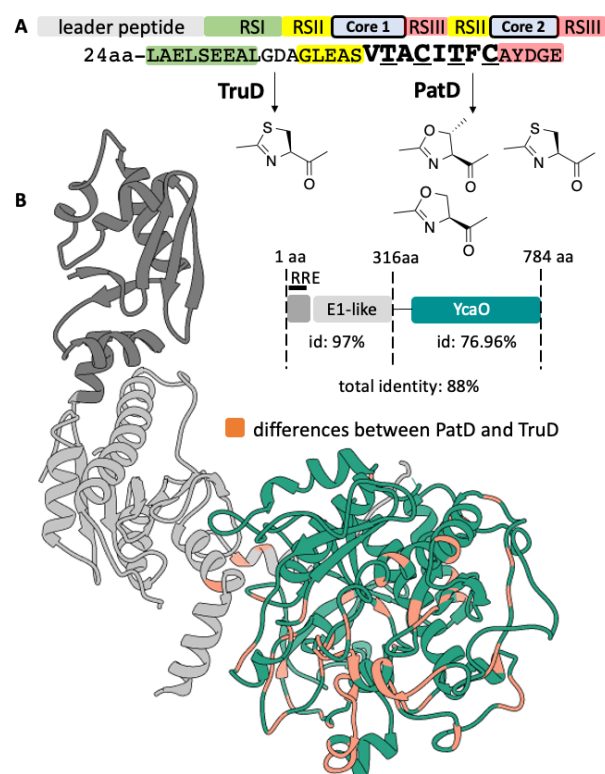


Fig 1 (A) Heterocyclases TruD and PatD convert the underlined Cys residues to thiazoline, but only PatD converts the underlined Thr/Ser residues to oxazoline. The PatE1-pC substrate sequence is shown. **(B)** Crystal structure of LynD, a heterocyclase that is similar to TruD. RRE domain is grey, E1-like domain is light grey, and YcaO domain is teal. Amino acids that differ between TruD and PatD are orange.

^a Department of Medicinal Chemistry, University of Utah, Salt Lake City UT 84112 USA. E-mail: ews1@utah.edu

^b Department of Biochemistry

^c Department of Chemistry

^d Center for Biophysics and Quantitative Biology

^e Beckman Institute for Advanced Science and Technology and

^f Institute for Genomic Biology, University of Illinois, Urbana, IL 61801, USA

Although the RS recognition by the RRE interaction is important for substrate engagement, there are other subtle features that guide PatD and TruD activity. For example, PatD has been used as a catalyst to modify very long, non-native core peptide sequences appended to RSI, in which virtually all Thr/Ser/Cys residues (up to 22 amino acids) are modified to azolines.^{25,26} However, native substrates of PatD contain many Ser residues that in principle could be modified, but that are not. These residues are found flanking the core peptides in regions referred to as RSII and RSIII. While RSII and RSIII were initially identified as recognition sequences for other enzymes in the cyanobactin pathway, they exert a further constraint on heterocyclase catalysis. Unfortunately, there are no crystal structures of a heterocyclase with a substrate bound at the active site. The only YcaO-core peptide co-crystal structure is that for the thioamide-forming MjYcaO, but that substrate does not contain RSIII, and the MjYcaO enzyme is functionally distinct and sequence divergent from the heterocyclases.⁷

Here, we take advantage of the sequence similarity between PatD and TruD to carry out mutagenesis and domain swapping survey in order to define the residues important for chemoselectivity. Both PatD and TruD are capable of using Cys and Ser/Thr residues when present at the end of the core peptide, but discriminate in their ability to heterocyclize these amino acids at other positions in the core peptide. We show that two flexible arm regions adjacent to the PatD/TruD active site direct this discrimination. Molecular modeling indicates that these arms interact with RSII/RSIII, and modulate the ability of the enzymes to access internal residues in the substrate. Site-directed mutagenesis identified a hydrogen-bonding network responsible for chemoselectivity at Ser/Thr versus Cys residues. These features collectively define how YcaO heterocyclases govern reactivity.

PatD and TruD and close homologs are unusually promiscuous post-translational enzymes, and this promiscuity has been exploited in different biotechnological contexts aimed at synthesizing peptide libraries, discrete bioactive compounds, and for other bioengineering purposes. For example, PatD will accept nonproteinogenic analogs of Ser, Thr, and Cys residues.²⁷ Hence, the greater understanding of reactivity described here may provide an improved framework in directed biosynthetic applications.

Results and discussion

Roles of recognition sequences in substrate processing

Biochemical and structural data show that a nine residue sequence in the leader termed RSI (LAELSEEAL) engages with the RRE of both PatD and TruD to enable substrate processing.^{12,17} RSIII (S/AYD) was assumed not to influence heterocyclase activity, and it was instead defined as the recognition sequence for the PatG macrocyclase enzyme.^{28,29} However, increasing data indicated that RSIII may be involved in heterocyclase processing.³⁰ To further investigate substrate tolerance, we carried out studies using RSI-TruD, in which the RSI sequence was covalently fused to the TruD N-terminus. This

RSI-TruD fusion obviates the need for substrates with a leader sequence, allowing the heterocyclase activity to be investigated using simple, short synthetic peptides.

A series of short substrates was synthesized, in which the core peptide sequence from aeruginosamide (*age*) biosynthesis (FFPC) was fused to C-terminal RSIII sequences of varying lengths (**Fig. 2**). When a 7-residue C-terminal RSIII sequence (SYDGVDA) was used, or when the minimal RSIII sequence SYD or AYD was employed, full heterocyclization of the Cys residue was observed. However, further truncation of RSIII to SY or AY significantly hampered heterocyclization, and only the unmodified precursor peptide could be observed in reactions with this peptide (**Fig. S1**). When feeding RSI-TruD with core and RSIII in trans, no modification was observed (**Fig. S2**). These data demonstrate that TruD requires a C-terminal sequence fused to the core peptide in order to process substrate peptides and that the C-terminal sequence of S/AYD is both necessary and sufficient. Hence, TruD residues that are in contact with RSIII are critical for heterocyclase activity in a different mechanism comparing to RSI which accelerates the reaction in trans.³¹

Further experiments were performed using wild-type TruD and three full-length precursor peptides in which RSIII was progressively deleted. When RSIII was deleted, leaving a terminal Cys residue, or RSIII was replaced with a single Ala residue, the C-terminal Cys residue was no longer heterocyclized, and instead only an internal Cys residue was modified by TruD (**Fig. S3**). These results demonstrate that TruD requires interaction with at least two C-terminal residues, but that the sequence of these residues is not crucial.

Modeling the enzyme-substrate interaction

To gain insight into the possible interactions between PatD/TruD and its cognate core peptides, including the requirement for RSIII, we created a structural model using

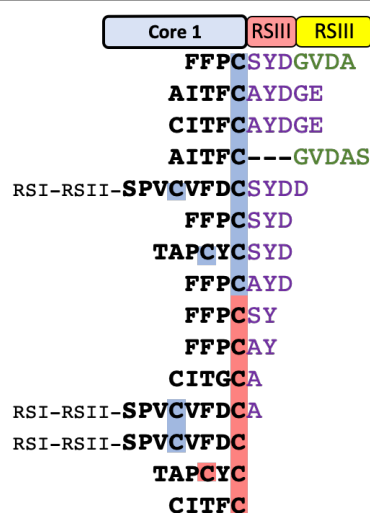


Fig. 2 Substrates that were modified by RSI-TruD or TruD (with RSI-RSII sequences on the N-terminus). RSIII (purple) on the C-terminus were gradually truncated. Some substrates contain the RSII sequence (green). Cys residues that were modified are highlighted in blue, and those not modified are in red.

MjYcaO bound to substrate (PDB 6PEU).⁷ The structure of the cyanobactin heterocyclase LynD (PDB 4V1T)¹⁷ was superimposed on the MjYcaO-McrA peptide cocrystal structure, and the MjYcaO substrate was replaced with a short sequence, FGFPCSYDG (Age_{Ecut}) related to cyanobactin aeruginosamide precursor sequence, as well as to the native sequence accepted by MjYcaO (**Fig. 3A**).³² The underlined Cys residue indicates the native site of heterocyclization. We chose these sequences because LynD is biochemically well defined and promiscuous, and Age_{Ecut} is a short and easily modeled substrate with the correct RSIII found in native LynD substrates. The enzyme-substrate conformation was optimized using molecular dynamics simulations. Amino acids that differed between PatD and TruD, and thus that are responsible for their different reactivity, are labeled on the LynD structure.

In the resulting model, the substrate's reactive Cys residue is bound adjacent to the ATP residue, and the entire substrate fits within the active site pocket (**Fig. 3B**). Two observations are immediately apparent: first, the series of residues in MjYcaO that were identified that were involved in a putative H-bonding network to deprotonate the substrate correspond to R427, S529, and R552 in the LynD model.⁷ Second, the amino acids that differ between PatD and TruD are concentrated along the probable substrate-binding pocket, with fewer mutations elsewhere.

A hydrogen-bonding network directs chemoselectivity

We hypothesized that the H-bonding/putative general base residues identified in MjYcaO are responsible for chemoselectivity of sulfur versus oxygen in TruD/PatD. Interestingly, these residues are identical in TruD/PatD, and indeed are highly conserved in YcaO enzymes.³³ To test this hypothesis, we synthesized three substrates, RS1-p6: *LAELSEEALGVDASTFPVPAVXAYD*. The substrate is so named because the italicized portion is RSI, responsible for binding to the heterocyclase RRE, while the underlined sequence is the core peptide encoding a mature compound similar to the natural product, patellin 6 (p6). The **X** residue is the site of heterocyclization, hence RSI-p6-Cys has a cysteine at the last position in the core. We also designed variants containing Ser (RSI-p6-Ser) and Thr (RS-P6-Thr) to investigate chemoselectivity.

Wild-type TruD enzyme modified both RSI-p6-Cys and RSI-p6-Thr, although it was much less efficient on RSI-p6-Ser (**Fig. 4**). This demonstrated that the observed natural chemoselectivity of TruD impacts only internal residues within the core peptide, while the C-terminal Cys may be substituted with Ser or Thr. Wild-type PatD functioned similarly on these substrates, except that RSI-p6-Ser was also an efficient substrate (**Fig. 4**).

We then generated mutants in the putative hydrogen-bonding network, including TruD S536A, R434K and R559K, and PatD R561A, R434M, R561M. We aimed to generate identical mutants in TruD and PatD but proceeded only with those that could be cloned and expressed in soluble fashion. The PatD mutants were nonfunctional, with only one mutant

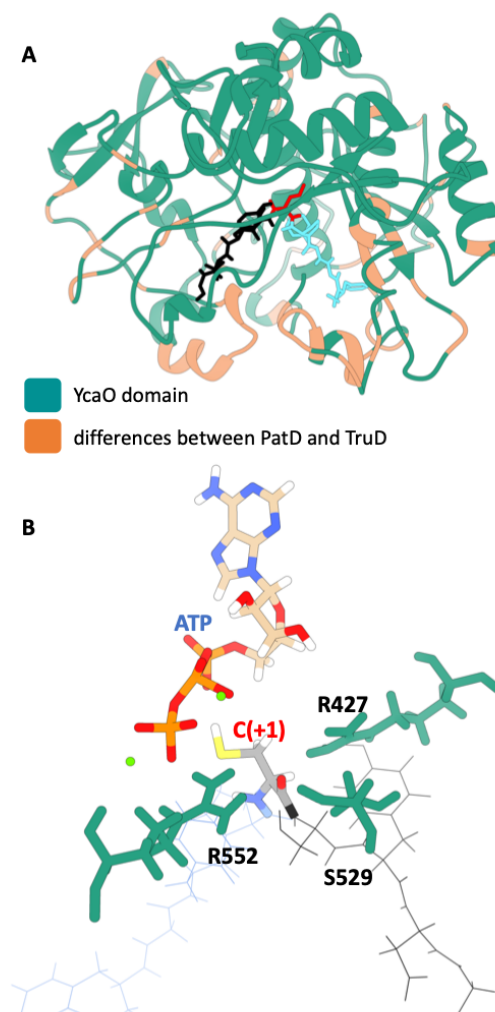


Fig. 3 YcaO-substrate model. **(A)** Homology model of YcaO domain (teal) from LynD (PDB Code 4V1T) and substrate Age_{Ecut} (FGFPCSYDG) using the MjYcaO-substrate co-structure (PDB Code 6PEU). Substrate core sequence (FGFP) is in cyan, reactive Cys is red, and RSIII (SYDG) is black. Amino acids that differ between TruD and PatD are orange. **(B)** Expansion of the YcaO active site, showing that the reactive Cys residue is in proximity to the ATP phosphate and important active site residues (LynD R427, S529 and R552).

(PatDR561A) showing slight modification of the Cys-containing substrate after 20 h (**Fig. S4B**). By contrast, all of the TruD variants efficiently heterocyclized RSI-p6-Cys, but variants S536A and R559K no longer processed RSI-p6-Thr or RSI-p6-Ser (**Fig. S4A**). Thus, mutation of basic residues into non-basic residues obviated enzyme function, whereas decreasing the basicity by mutating R to K led selectively to loss of β -hydroxyl heterocyclization but maintained β -thiolate heterocyclization. These results are consistent with a role for these residues in accelerating catalysis required for the more difficult β -hydroxyl heterocyclization, probably by aiding with the deprotonation of the less nucleophilic Ser/Thr side chains. Since TruD is incapable of modifying oxygen nucleophiles except at the C-terminal position, other sequence elements must be responsible for

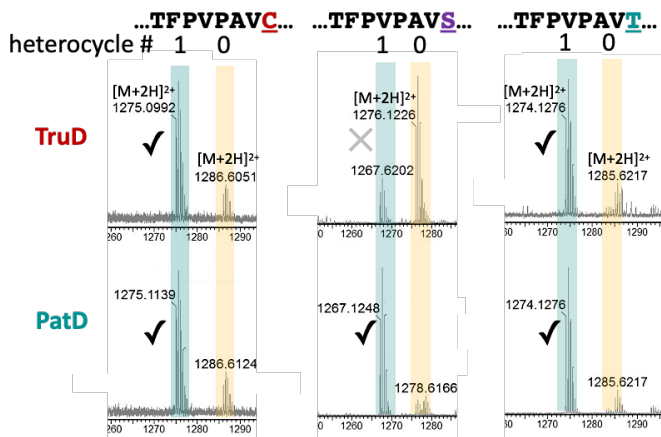


Fig. 4 MS spectra of TruD and PatD enzyme reaction mixtures at 24 h. Identical precursor peptides were used, but with either Ser, Thr, or Cys residues at the P1 heterocyclization site. Both PatD and TruD equally modify Cys and Thr residues at P1, whereas only PatD shows complete heterocyclization of serine. Chromatography conditions included ammonium formate (10 mM, pH 6.5) in H₂O and acetonitrile to prevent oxazoline ring opening.

explaining heterocycle selectivity in core residues that are internal.

We then took advantage of the close sequence similarity between TruD and PatD to create a series of chimeric proteins, using PatE1-pC as a substrate to determine which sequence elements enabled cyclization of Thr residues. Remarkably, out of 31 swap mutants created in this study, 24 were soluble and functional, defined as being able to modify at least one residue. Proteins could not be expressed in soluble form in *Escherichia coli* despite repeated attempts. In general, the percent identity to PatD was correlated with the ability to heterocyclize Ser/Thr residues, although the exceptions are highly revealing (see below).

Selectivity is localized to the YcaO domain

Fourteen of the chimeric proteins were swaps in which the N-terminus of PatD was progressively replaced with that of TruD. The first ten mutants replaced the RRE and E1 domains, as well as portions of the YcaO up to position PatD₅₁₂/TruD₅₁₀. All ten mutants were capable of processing both Cys and Thr residues (Fig. 5). Replacing the RRE/E1 domains had no effect on activity, nor did swap mutants in the YcaO until position TruD₄₇₈. Mutants including TruD₄₈₅₋₅₁₀ heterocyclized three residues, including one Thr residue (which was localized to substrate Thr7; Fig. S6), although processing of the Thr residue was less efficient and incomplete. The remaining mutants could not cyclize Thr residue. Thus, the ability to discriminate in the cyclization of internal β-hydroxyl nucleophiles was localized to the C-terminal 372 amino acids, comprising residues PatD₅₁₂₋₇₈₄.

Two short regions near the active site direct internal residue modification

Although several chimeras in which the C-terminus of PatD was progressively replaced with the sequence of TruD failed to

Table 1. Summary of chimeric enzyme activity

enzyme	%ID to TruD YcaO**	heterocyclized residues in PatE1-pC	RSI-pCys/pThr /pSer
PatD wild type	76.96	4	Y***/Y/Y
TruD ₁₋₄₁₄ PatD ₄₁₆₋₇₈₄ *	76.96	4	Y/Y/Y
PatD ₁₋₇₇₅ TruD ₇₇₃₋₇₈₁	77.78	1 (small amount of 2,3,4)	N***/N/Y
TruD ₁₋₄₇₈ PatD ₄₇₉₋₇₈₄	78.05	4	Y/Y/Y
TruD ₁₋₄₈₅ PatD ₄₈₈₋₇₈₄	79.89	2/3	N/Y/Y
TruD ₁₋₄₈₆ PatD ₄₈₉₋₇₈₄	80.16	2/3	N/N/Y
TruD ₁₋₄₈₉ PatD ₄₉₂₋₇₈₄	80.16	2/3	Y/Y/Y
TruD ₁₋₄₉₂ PatD ₄₉₄₋₇₈₄	80.43	2/3	N/Y/Y
TruD ₁₋₄₉₅ PatD ₄₉₈₋₇₈₄	80.71	2/3	N/Y/Y
PatD ₁₋₇₀₀ TruD ₆₉₉₋₇₈₁	80.98	2	N/N/Y
TruD ₁₋₅₀₂ PatD ₅₀₅₋₇₈₄	81.25	2/3	N/Y/Y
TruD ₁₋₅₀₅ PatD ₅₀₈₋₇₈₄	81.52	3	N/Y/Y
PatD ₁₋₅₆₃ TruD ₅₆₂₋₆₃₇ PatD ₆₄₀₋₇₈₄	81.57	4	N/Y/Y
TruD ₁₋₅₁₀ PatD ₅₁₂₋₇₈₄	81.79	2/3	N/Y/Y
PatD ₁₋₄₇₉ TruD ₄₇₉₋₅₃₀ PatD ₅₃₂₋₇₈₄	82.38	2	N/N/Y
TruD ₁₋₅₀₅ PatD ₅₀₈₋₇₆₀ TruD ₇₅₈₋₇₈₁	83.42	2	N/N/Y
TruD ₁₋₅₂₉ PatD ₅₃₂₋₇₈₄	83.7	2	N/Y/Y
TruD ₁₋₅₆₁ PatD ₅₆₄₋₇₈₄	84.51	2	Y/Y/Y
TruD ₁₋₆₃₇ PatD ₆₄₀₋₇₈₄	89.13	2	N/N/Y
TruD ₁₋₆₁₃ PatD ₆₁₆₋₇₀₀ TruD ₆₉₉₋₇₈₁	91.01	2	N/Y/Y
TruD ₁₋₄₇₈ PatD ₄₈₀₋₅₃₁ TruD ₅₃₀₋₇₇₂ PatD ₇₇₆₋₇₈₄	93.48	2	N/N/Y
TruD ₁₋₄₇₈ PatD ₄₈₀₋₅₃₁ TruD ₅₃₀₋₇₈₁	94.29	2	N/N/Y
TruD ₁₋₄₇₈ PatD ₄₈₀₋₄₈₈ TruD ₄₈₇₋₇₇₂ PatD ₇₇₆₋₇₈₄	97.01	2	n.d. [†] /Y/Y
TruD ₁₋₄₇₈ PatD ₄₈₀₋₄₈₈ TruD ₄₈₇₋₇₈₁	97.83	2	N/Y/Y
TruD ₁₋₇₅₇ PatD ₇₆₁₋₇₈₄	98.09	2	N/Y/Y
TruD wild type	100	2	N/Y/Y

* TruD₁₋₄₁₄PatD₄₁₆₋₇₈₄: a protein sequence with TruD amino acids 1-414 fused with PatD amino acids 416-784.

** Sequence identity compared to TruD YcaO domain.

*** "Y", the enzyme is able to modify the substrate; "N", the enzyme cannot modify the substrate.

[†] n.d.: not detected

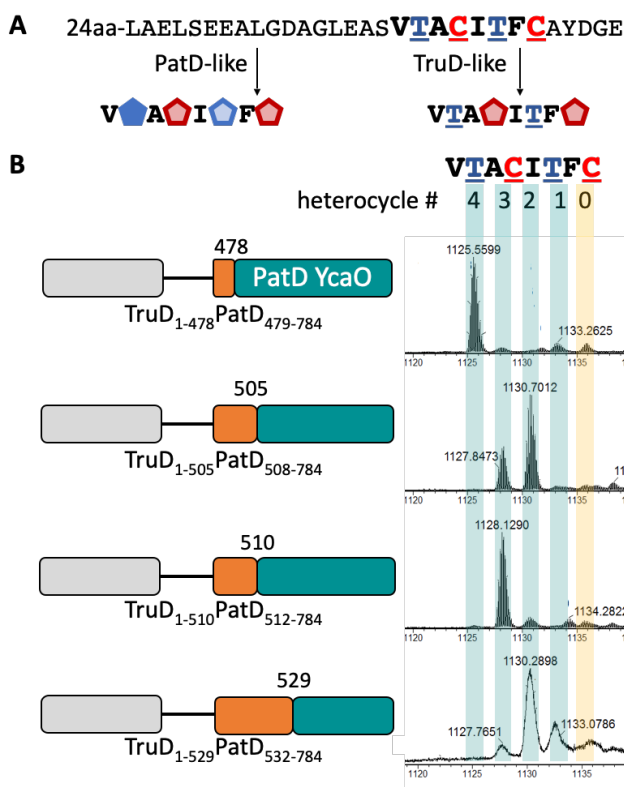


Fig. 5 PatD mutants past residue 512 lose the ability to cyclize internal Thr residues. **(A)** PatE1-pC is modified by TruD to produce thiazoline residues (red), while PatD also produces oxazoline residues (blue). In function-impaired PatD proteins, Thr7 (dark blue) is modified, and not Thr3 (light blue). **(B)** MS spectra of heterocyclization reactions with TruD-PatD chimeric mutants.

express, the two exceptions were extremely informative. In the first chimera, PatD₁₋₇₀₀TruD₆₉₉₋₇₈₁, only the last 82 residues were from TruD, and the mutant was 96% identical to PatD, and yet only two Cys residues of the PatE1-pC substrate were processed. By contrast, other mutants with identities to PatD as low as 95% could still cyclize Thr residues. In the second chimera PatD₁₋₇₇₅TruD₇₇₃₋₇₈₁, only three amino acids from PatD were replaced with those from TruD, and the protein was >99% identical to PatD. However, the predominant product of this variant consisted of cyclization at only the C-terminal Cys residue of PatE1-pC (**Fig. S7**). The variant was still capable of cyclizing the internal Cys and Thr residues, albeit with much lower yield. These data indicated that residues at the C-termini of PatD/TruD and other heterocyclases govern the modification of internal residues.

Based on the results of the domain swap experiments, it was clear that a region adjacent to TruD510/PatD512, and the C-terminal residues of TruD/PatD both played a role in the modification of internal Ser/Thr/Cys residues. A series of individual mutations aimed at teasing the responsible residues apart were unsuccessful. For example, point mutants of individual PatD or TruD residues in the region of TruD510-529 did not affect the selectivity of the enzymes (**Fig. S8A**). At the C-terminus, several point mutants of PatD failed to express, and

TruD5752P and TruD746 (in which an additional Thr residue present in PatD was inserted) were functional, but their selectivity for internal residues was not changed relative to wild-type TruD (**Fig. S8B**).

Therefore, to investigate these regions in greater detail, we performed a series of eight “sandwich” swaps, in which regions of TruD were flanked on both the C- and N-terminus by regions from PatD, or vice versa. In addition, two of these chimeras contained 4 protein pieces from TruD and PatD, spliced together. In general, these proteins behaved as expected, with the number of heterocycles produced corresponding to the percent identity of either TruD or PatD. There was one notable exception: swap mutant PatD₁₋₅₆₃TruD₅₆₂₋₆₃₇PatD₆₄₀₋₇₈₄ efficiently heterocyclized all four residues, much like wild-type PatD. The contrasting mutant, TruD₁₋₅₀₅PatD₅₀₈₋₇₆₀TruD₇₅₈₋₇₈₁ made only TruD-type modifications (**Fig. S9**). Therefore, the central region of the protein, between PatD564-639, does not seem to play a role in the heterocyclization of internal Thr residues. Based upon the mutants and results described above, amino acid residues in the region of PatD512-531 are responsible for this selectivity.

The regions corresponding to PatD512-531 and PatD777-784 were not localized adjacent to the active site in the LynD-AgeE_{cut} structural model. Instead, the region at PatD532 terminates in an unstructured region that interacts directly with RSIII. The PatD777-784 region points directly at the N-terminus of the substrate (**Fig. 6**). Most of these sequences are within relatively unstructured regions of the LynD crystal structure. These findings suggest that the ability to modify internal residues is governed at least in part by binding to the flanking RSII and RSIII and their presumed mobility (based on disorder observed in the crystal structures) would enable the PatD/TruD active site to modify multiple residues.

A model for chemoselectivity and application to new catalyst discovery

We show that manipulating the conserved residues involved in the putative H-bonding active-site network destroys the enzyme ability to cyclize Ser and Thr, presumably because Cys is more nucleophilic and reactive than Ser and Thr residues. These conserved residues are also present in other YcaO enzymes. The homologous Arg and Thr pairs are critical catalytic residues in the thioamide forming MjYcaO.⁷ In MjYcaO, these residues were observed flanking the reactive amide carbonyl, which might potentially act as general bases in the reaction pathway. In McbBCD structure, it was also hypothesized that Ser536 relates to substrate binding or deprotonation due to its proximity to the active site and its nucleophilicity.⁹ The similarity of the active site residues between different types of YcaO enzymes led to the homology modeling of the LynD-AgeE_{cut} based on the co-crystal of MjYcaO and its substrate.

Using structural modeling and mutant chimeras, we identified two regions that are relatively unstructured and sandwich the RSII/III elements at either end of the putative peptide binding site. The key positions are localized to 20 amino acid residues toward the N-terminus of YcaO, and nine residues

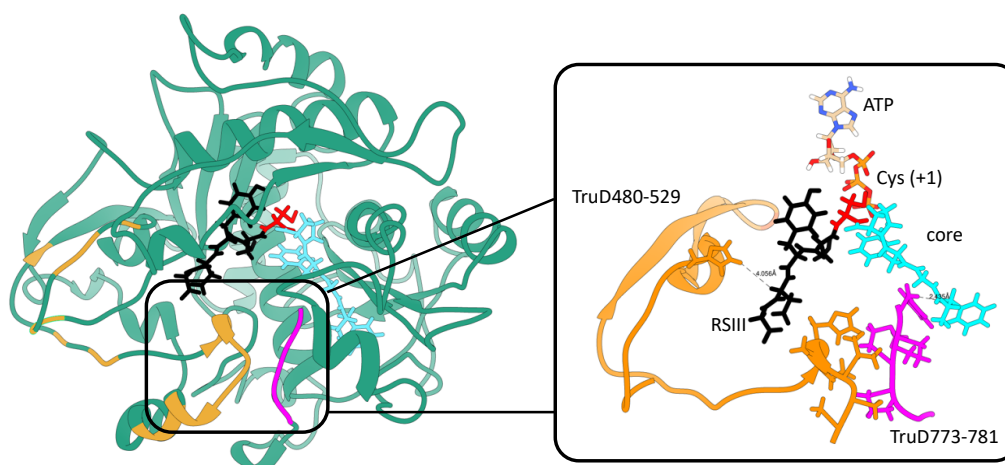


Fig. 6 Dynamic regions interacting with the substrate. The orange (residues TruD480-529) and magenta (C-terminal amino acids TruD773-781) regions interact with the substrate RSII and RSIII, leading to a model in which the binding of these regions to RSs guides access to internal nucleophiles in the substrate

at the C-terminus. The five residues on the C-terminus termed as PxPxP motif were also proposed to be involved active site organization and catalysis in thiazole/oxazole-modified microcin (TOMM) YcaO heterocyclases.³⁴ We propose that these regions first help orient the precursor peptide in the active site by forming a flexible hinge that binds to RSII/RSIII, placing the P1 Cys residue in position for heterocyclization. The first heterocyclization reaction changes the shape of the peptide, enabling access to internal residues in the core, facilitated by the flexible hinges. Because the β -thiol of Cys is much more reactive than the β -hydroxyl of Ser/Thr toward heterocyclization, minor differences in the kinetics of hinge movements likely dictate the ability to circularize Ser/Thr.

In support of this gated access model, we show that disruption of either RSIII on the substrate, or of the C-terminal amino acids on the enzyme, completely alters the selectivity, without blocking the catalysis of at least one residue. Modification of PatD residues distant from the active site, yet adjacent to flexible hinge regions lining RSII and RSIII, abolishes the enzyme's reaction on internal Ser/Thr residues. Dynamic regions are also observed in other RiPP enzymes such as the lanthionine synthetases ProcM and HalM, in the regions that control substrate access.^{35,36}

In recent years, promiscuous heterocyclases including PatD and LynD have been widely used in synthesis of diverse compounds. This work has been facilitated by fusion of RSI to activate the enzymes so that long precursor peptides are not needed.^{17,32,37-40} A limitation is that most current enzymes are "all-or-nothing", in that they modify at every Cys or Cys/Thr/Ser with one recent exception from thiopeptide biosynthesis.¹³ Moreover, despite their promiscuity there are still regions of sequence space in which they are not functional, or too slow to be useful. We suggest that the thousands of related YcaO-encoding genes that have been sequenced may be of service to solve this problem. Sequence alignments of all characterized cyanobactin YcaOs reveal that hinge region residues are predictive of PatD-like or TruD-like activity (**Fig. 7**). Thus,

TenD	RHKYLPTGLCYHYHYPMEAN	AEMNPMNIPF
SagD	KHKYVPTAFICYGYPLPEEQ	EELNQTNIIPF
McaD/MicD	KHKYVPTAFICYGYPLPEEQ	EELNQTNIIPF
PatD	KHKYVPTAICYNYLLPPAD	AEMNPTNIIPF
BisD	KHKYVPTAICYSYPLPEAD	AQMNPNTIIPF
ThcD	AHKYLPTAFICYNYKLPKQH	DQMNPMMPVF
MusD	THKYIPTAYCYGYPLPEDH	EQMNPISVPI
RivD	SHKYLPTAFICYEYPRPKDH	QQMNSTAMPF
ArtD	CHKYLPTAFICYNYMPKGG	TEMNPMAMPF
TriD	THKYLPTAWCYAYKLPKHH	EEMNSMPMPVF
OscD	THKYLPTAYCYDYPSPKDH	EEMNPMMPMF
VirD	THKYVPIAMCYHYHYPSPNH	AQMNPPTMPF
TruD	THKYLPTALCYRYPFPEEH	AQMNPPTMPF
TrfD	THKYLPTALCYRYPFPEEH	AQMNPPTMPF
LynD	THKYLPTALCYHYPLPEEH	AQMNPPTMPF
AgeD	THKYLPTALCYHYPLPEEH	AQMNPPTMPF

***:* . *** * * . ::* : :

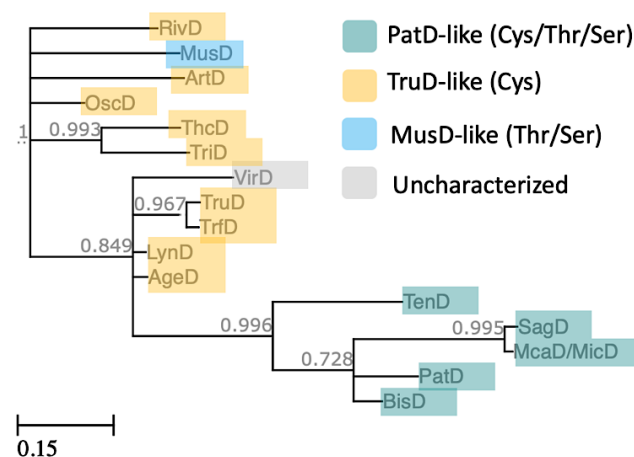


Fig. 7 Phylogenetic prediction of function using the RS-lining residues in cyanobactin YcaOs. Top: Only the flexible sequences predicted to interact with RSs (TruD 510-529 and 772-781) are aligned. Bottom: Bayesian analysis shows that these residues are predictive of function, with all enzymes that modify both Ser/Thr and Cys residues clustered (teal). When whole proteins are aligned, the resulting tree is not predictive of function. (Note: MusD is an unusual enzyme that is still not predictable from current knowledge.⁴¹)

selection of YcaO proteins that have the correct hinge-region sequence features will afford new catalysts capable of heterocyclizing Cys/Ser/Thr residues. The engineering principles that we derive here will thus be broadly useful in designing precision catalysts that are nucleophile-specific, and in overcoming the sequence limitations encountered with currently used heterocyclases.

Conclusions

Members of the YcaO superfamily catalyze a complex series of chemical reactions, involving the orchestration of exquisite chemo- and regioselectivity for multiple Cys, Ser, and Thr residues in precursor peptides. Here we use a structure-function approach dissect the chemoselectivity of PatD and TruD, two highly similar enzymes involved in cyanobactin biosynthesis. These methods are useful in engineered YcaO catalyst design and in discovering new catalysts with desired properties from nature, which may be widely useful in synthetic biology.

Author Contributions

W.G. designed and performed the experiments, analyzed the data and wrote the original draft. Y.Z. performed part of the RSIII related experiments. S.K.N. generated the protein-substrate model, and T.P. conducted the MD simulations. E.W.S. supervised the research and finalized the manuscript. All authors contributed to the manuscript editing and have given approval to the final version of the manuscript.

Conflicts of interest

EWS is a co-owner of Synthetic Biodesign, which uses RiPP engineering to produce drugs.

Acknowledgements

This work is funded by NIH R35GM122521 (EWS lab) and NIH R01GM131347 (SKN lab). We thank Paola Estrada for providing the C-terminal point mutant plasmids. We also thank Diao Chen and Danny Hung-Chieh Chou with the help of solid-phase peptide synthesis.

Notes and references

- 1 M. Montalbán-López, T. A. Scott, S. Ramesh, I. R. Rahman, A. J. van Heel, J. H. Viel, V. Bandarian, E. Dittmann, O. Genilloud, Y. Goto, M. J. Grande Burgos, C. Hill, S. Kim, J. Koehnke, J. A. Latham, A. J. Link, B. Martínez, S. K. Nair, Y. Nicolet, S. Rebuffat, H.-G. Sahl, D. Sareen, E. W. Schmidt, L. Schmitt, K. Severinov, R. D. Süßmuth, A. W. Truman, H. Wang, J.-K. Weng, G. P. van Wezel, Q. Zhang, J. Zhong, J. Piel, D. A. Mitchell, O. P. Kuipers and W. A. van der Donk, *Nat. Prod. Rep.*, 2021, **37**, 919–961.
- 2 B. J. Burkhart, C. J. Schwalen, G. Mann, J. H. Naismith and D. A. Mitchell, *Chem. Rev.*, 2017, **117**, 5389–5456.
- 3 L. Franz, S. Adam, J. Santos-Aberturas, A. W. Truman and J. Koehnke, *J. Am. Chem. Soc.*, 2017, **139**, 18158–18161.
- 4 C. J. Schwalen, G. A. Hudson, S. Kosol, N. Mahanta, G. L. Challis and D. A. Mitchell, *J. Am. Chem. Soc.*, 2017, **139**, 18154–18157.
- 5 L. Franz, U. Kazmaier, A. W. Truman and J. Koehnke, *Nat. Prod. Rep.*, 2021, **23**, 9–10.
- 6 D. Y. Travin, M. Metelev, M. Serebryakova, E. S. Komarova, I. A. Osterman, D. Ghilarov and K. Severinov, *J. Am. Chem. Soc.*, 2018, **140**, 5625–5633.
- 7 S. H. Dong, A. Liu, N. Mahanta, D. A. Mitchell and S. K. Nair, *ACS Cent. Sci.*, 2019, **5**, 842–851.
- 8 Y.-M. Li, J. C. Milne, L. L. Madison, R. Kolter and C. T. Walsh, *Science (80-.)*, 1996, **274**, 1188–1193.
- 9 D. Ghilarov, C. E. M. Stevenson, D. Y. Travin, J. Piskunova, M. Serebryakova, A. Maxwell, D. M. Lawson and K. Severinov, *Mol. Cell*, 2019, **0**, 1–14.
- 10 K. L. Dunbar, J. O. Melby and D. A. Mitchell, *Nat. Chem. Biol.*, 2012, **8**, 569–575.
- 11 M. S. Donia, B. J. Hathaway, S. Sudek, M. G. Haygood, M. J. Rosovitz, J. Ravel and E. W. Schmidt, *Nat. Chem. Biol.*, 2006, **2**, 729–735.
- 12 J. A. McIntosh, M. S. Donia and E. W. Schmidt, *J. Am. Chem. Soc.*, 2010, **132**, 4089–4091.
- 13 Y. Du, Y. Qiu, X. Meng, J. Feng, J. Tao and W. Liu, *J. Am. Chem. Soc.*, 2020, **142**, 8454–8463.
- 14 A. Vinogradov, M. Shimomura, Y. Goto, T. Ozaki, S. Asamizu, Y. Sugai, H. Suga and H. Onaka, *Nat. Commun.*, 2020, **11**, 1–13.
- 15 J. A. McIntosh and E. W. Schmidt, *ChemBioChem*, 2010, **11**, 1413–1421.
- 16 M. S. Donia, J. Ravel and E. W. Schmidt, *Nat. Chem. Biol.*, 2008, **4**, 341–343.
- 17 J. Koehnke, G. Mann, A. F. Bent, H. Ludewig, S. Shirran, C. Botting, T. Lebl, W. E. Housen, M. Jaspars and J. H. Naismith, *Nat. Chem. Biol.*, 2015, **11**, 558–563.
- 18 M. S. Donia and E. W. Schmidt, *Chem. Biol.*, 2011, **18**, 508–519.
- 19 D. Sardar, E. Pierce, J. A. McIntosh and E. W. Schmidt, *Am. Chemical Soc. Synth. Biol.*, 2015, **4**, 167–176.
- 20 B. J. Burkhart, G. A. Hudson, K. L. Dunbar and D. A. Mitchell, *Nat. Chem. Biol.*, 2015, **11**, 564–570.
- 21 W. Gu, S.-H. Dong, S. Sarkar, S. K. Nair and E. W. Schmidt, in *Methods in Enzymology*, Elsevier Inc., 1st edn., 2018, vol. 604, pp. 113–163.
- 22 D. Sardar and E. W. Schmidt, *Curr. Opin. Chem. Biol.*, 2016, **31**, 15–21.
- 23 W. Gu and E. W. Schmidt, *Acc. Chem. Res.*, 2017, **50**, 2569–2576.
- 24 J. Koehnke, A. F. Bent, D. Zollman, K. Smith, W. E. Housen, X. Zhu, G. Mann, T. Lebl, R. Scharff, S. Shirran, C. H. Botting, M. Jaspars, U. Schwarz-Linek and J. H. Naismith, *Angew. Chemie - Int. Ed.*, 2013, **52**, 13991–13996.
- 25 Y. Goto, Y. Ito, Y. Kato, S. Tsunoda and H. Suga, *Chem. Biol.*, 2014, **21**, 766–774.
- 26 D. Sardar, Z. Lin and E. W. Schmidt, *Chem. Biol.*, 2015, **22**, 907–916.
- 27 J. Koehnke, F. Morawitz, A. F. Bent, W. E. Housen, S. L. Shirran, M. A. Fuszard, I. A. Smellie, C. H. Botting, M. C. M. Smith, M. Jaspars and J. H. Naismith, *ChemBioChem*, 2013, **14**, 564–567.
- 28 J. A. McIntosh, C. R. Robertson, V. Agarwal, S. K. Nair, G. W. Bulaj and E. W. Schmidt, *J. Am. Chem. Soc.*, 2010, **132**, 15499–15501.
- 29 J. Lee, J. McIntosh, B. J. Hathaway and E. W. Schmidt, *J. Am. Chem. Soc.*, 2009, **131**, 2122–2124.
- 30 Y. Goto and H. Suga, *Chem. Lett.*, 2016, **45**, 1247–1249.
- 31 W. Gu, D. Sardar, E. Pierce and E. W. Schmidt, *J. Am. Chem. Soc.*,

- 2018, **140**, 16213–16221.
- 32 D. Sardar, Y. Hao, Z. Lin, M. Morita, S. K. Nair and E. W. Schmidt, *J. Am. Chem. Soc.*, 2017, **139**, 2884–2887.
- 33 S. H. Dong, A. Liu, N. Mahanta, D. A. Mitchell and S. K. Nair, *ACS Cent. Sci.*, 2019, **5**, 842–851.
- 34 K. L. Dunbar, J. R. Chekan, C. L. Cox, B. J. Burkhardt, S. K. Nair and D. a Mitchell, *Nat. Chem. Biol.*, 2014, **10**, 1–35.
- 35 Y. Habibi, K. A. Uggowitzer, H. Issak and C. J. Thibodeaux, *J. Am. Chem. Soc.*, 2019, **141**, 14661–14672.
- 36 K. A. Uggowitzer, Y. Habibi, W. Wei, N. Moitessier and C. J. Thibodeaux, *Biochemistry*, 2021, **60**, 412–430.
- 37 T. J. Oman, P. J. Knerr, N. A. Bindman, J. E. Velásquez and W. A. Van Der Donk, *J. Am. Chem. Soc.*, 2012, **134**, 6952–6955.
- 38 E. Reyna-Gonzalez, B. Schmid, D. Petras, R. D. Sussmuth and E. Dittmann, *Angew. Chemie - Int. Ed.*, 2016, **55**, 9398–9401.
- 39 M. N. Ahmed, E. Reyna-González, B. Schmid, V. Wiebach, R. D. Süssmuth, E. Dittmann and D. P. Fewer, *ACS Chem. Biol.*, 2017, **12**, 1538–1546.
- 40 C. Lee, H. Lee, J. U. Park and S. Kim, *Biochemistry*, 2020, **59**, 285–289.
- 41 A. Mattila, R.-M. Andsten, M. Jumppanen, M. Assante, J. Jokela, M. Wahlsten, K. M. Mikula, C. Sigindere, D. H. Kwak, M. Gugger, H. Koskela, K. Sivonen, X. Liu, J. Yli-Kauhaluoma, H. Iwai and D. P. Fewer, *ACS Chem. Biol.*, 2019, **14**, 2683–2690.