Supporting Information for:

Mechanistic Basis for Ribosomal Peptide Backbone Modifications

Shi-Hui Dong[1,#], Andi Liu[2,3,#], Nilkamal Mahanta[2,4], Douglas A. Mitchell[2,3,4*], and Satish K. Nair[1,2,5*]

[1]Department of Biochemistry, [2]Carl R. Woese Institute for Genomic Biology, [3]Department of Microbiology, [4]Department of Chemistry, and [5]Center for Biophysics and Quantitative Biology; University of Illinois, 600 S Mathews Ave, Urbana, IL, 61801, USA

#Denotes equal authorship contribution.
* Authors to whom correspondence should be addressed: douglasm@illinois.edu, snair@illinois.edu

Table of Contents:

Materials and Methods

Safety statement. No unexpected or unusually high safety hazards were encountered.

General materials and methods. Oligonucleotides were purchased from Integrated DNA Technologies. Restriction endonucleases, DNA polymerases, T4 DNA polymerase, and HiFi DNA Assembly Mater Mix were obtained from New England Biolabs. Amicon Ultra-4 centrifugal filters were purchased from EMD Millipore. Chemical reagents were purchased from Sigma Aldrich. All synthetic peptides were obtained from GenScript. *Escherichia coli* DH5α and *E. coli* BL21(DE3) cells were used for plasmid maintenance and protein overproduction, respectively. DNA sequencing was performed by ACGT, Inc and the DNA sequencing facility at the University of Illinois, Urbana-Champaign.

Molecular biology techniques. The sequence encoding residues that encompass five amino acids N- and C-terminal to the site of thioamidation ($_{443}$RLGFY**G**YDLQD$_{453}$) of *Mj*McrA (WP_010870360) was cloned from *Methanocaldococcus jannaschii* using the HindIII and NotI restriction enzymes into a modified pET28 vector that fuses the peptide to the C-terminus of maltose-binding protein (MBP). *Mj*YcaO was cloned from *Mj*gDNA obtained from DSMZ, Germany (DSM 2661) as reported before [1] into the pET MBP His_6 TEV LIC vector (gift from Scott Gradia, Addgene plasmid #29656). These two constructs were used for protein expression and as the templates for McrA peptide and *Mj*YcaO site-directed mutagenesis. Mutagenesis was performed using the QuikChange method (Agilent) or the NEBuilder HiFi DNA Assembly (NEB #E2621) method. The primers used in this study were listed in Table S4. All mutagenesis was verified by DNA sequencing.

Protein expression and purification. The plasmids with constructs of interest were transformed into *E. coli* BL21(DE3) Rosetta 2 cells for protein expression. Cells were grown for 24 h on Luria-Bertani (LB) agar plates containing 50 μg/mL kanamycin and 34 μg/mL chloramphenicol at 37 °C. Single colonies were picked to inoculate 10 mL of LB containing 50 μg/mL kanamycin and 34 μg/mL chloramphenicol and grown at 37 °C for 16-18 h. This culture was used to inoculate 1 L of LB containing 50 μg/mL kanamycin and 34 μg/mL chloramphenicol and grown to an optical density at 600 nm (OD_{600}) of 0.6-0.8 and then placed on ice for 15 min. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.5 mM, followed by an induction period at 18 °C for 18 h. Cells were harvested via centrifugation at 8,000 × g for 10 min and resuspended in 30 mL resuspension buffer [500 mM NaCl, 25 mM Tris pH 8.0, 10% glycerol (*v/v*)]. Cells were lysed by sonication for five cycles (40 s each) at 20% amplitude, with 1 min intervals of resting at 4 °C. Insoluble cell material was removed by centrifugation at 38,000 × g for 45 min at 4 °C. The resultant supernatant was loaded onto a pre-equilibrated HisTrap HP column (GE Healthcare Life Sciences; 5 mL of resin per L of cell culture). The column was washed with

40 mL buffer containing 1 M NaCl, 25 mM Tris pH 8.0, and 25 mM imidazole. Protein elution was performed on an ÄKTAprime chromatography system (GE Healthcare Life Sciences) with 40 mL buffer containing 1 M NaCl, 25 mM Tris pH 8.0, and a linear gradient of imidazole from 25 mM to 250 mM at a flow rate of 2 mL/min. The resultant fractions were examined visually by Coomassie-stained sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel and those with higher than 95% purity were combined and concentrated using a 30 kDa molecular weight cut-off (MWCO) Amicon Ultra centrifugal filter. A buffer exchange with 10 × volume of protein storage buffer [50 mM HEPES pH 7.5, 300 mM NaCl, 2.5% glycerol (*v/v*), 0.5 mM tris(2-carboxyethyl)phosphine] was performed prior to storage. Protein concentrations were determined using both 280 nm absorbance (theoretical extinction coefficients were calculated using the ExPASy ProtParam tool; http://web.expasy.org/protparam/) and a Bradford colorimetric assay using bovine serum albumin as a standard (Thermo Fisher Scientific). Purity was visually inspected by Coomassie-stained SDS-PAGE gel (SI Figure S2).

To obtain *Mj*YcaO protein without the MBP-His tag for crystallography study, the aforementioned combined fractions were digested by addition of 1:100 (*w/w*) tobacco etch virus (TEV) protease and dialyzed concomitantly against buffer containing 300 mM NaCl, 25 mM Tris pH 7.5, and 10% glycerol (*v/v*). The dialyzed proteins were loaded onto a pre-equilibrated Ni-NTA column to remove the MBP-His tag. The flow-through was collected and concentrated using a 30 kDa MWCO Amicon Ultra centrifugal filter. The resultant proteins were subjected to a Superdex 200 size-exclusion column (GE Healthcare Lifesciences) equilibrated with 100 mM KCl and 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5 for buffer exchange and purity examination: a single peak was observed by monitoring absorbance at 280 nm.

MBP-tagged McrA peptide expression and purification. MBP-tagged McrA peptide and its variants were expressed and purified similarly as stated above with the following alterations. The expression cultures were grown to an OD_{600} of 1.0 and cooled on ice for 15 min. IPTG was then added to a final concentration of 0.3 mM, followed by an induction period at 18 °C for 1.5 h. Cells were harvested via centrifugation at 4,000 × g for 10 min, washed in buffer containing 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4, and resuspended in 30 mL lysis buffer [50 mM Tris-HCl pH 7.5, 500 mM NaCl, 2.5% glycerol (*v/v*), 0.1% Triton X-100 (*v/v*)] containing 4 mg/mL lysozyme, 2 µM leupeptin, 2 µM benzamidine HCl, 2 µM E64, and 30 mM phenylmethylsulfonyl fluoride. Cells were lysed by sonication for three cycles (40 s each) at 20% amplitude, with 10 min intervals of gentle rocking at 4 °C. Insoluble cell material was removed by centrifugation at 17,000 × g for 1 h at 4 °C. The resultant supernatant was loaded onto a pre-equilibrated amylose resin column (NEB; 10 mL of resin per L of cell culture). The column was washed with 30 mL of lysis buffer and then eluted using 30 mL buffer

containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2.5% glycerol (*v/v*), and 10 mM maltose. The eluent was concentrated, buffer exchanged, and its concentration was determined as stated above. Purity was visually inspected by Coomassie-stained SDS-PAGE gel (SI Figure S2).

Crystallization, Data Collection, Phasing and Refinement. To produce crystals of high quality, 8-10 mg/mL tag-free *Mj*YcaO protein sample was subjected to the hanging drop vapor diffusion method [2]. The *Mj*YcaO-nucleotide crystals were obtained by equilibrating with 4 mM ATP and 2 mM $MgCl_2$ prior to the addition of the precipitate solution [0.1 M sodium acetate pH 4.6, 2.0 M ammonium sulfate, 0.02 M taurine] in 1:1 (*v/v*) ratio. To obtain the co-crystals, 4 mM ATP, 2 mM $MgCl_2$, and 1.5 mM of *Mj*McrA-11-mer were equilibrated with the *Mj*YcaO sample for over 30 min, which was then used for crystallization by adding the precipitate solution [0.02 M 1,6-hexanediol, 0.02 M 1-butanol, 0.02 M 1,2-propanediol, 0.02 M 2-propanol, 0.02 M 1,4-butanediol, 0.02 M 1,3-propanediol, 0.05 M imidazole and 2-(N-morpholino)ethanesulfonic acid (MES) pH 6.5, 20% (*v/v*) ethylene glycol, 10% (*w/v*) polyethylene glycol 8000, 0.02 M ethylenediaminetetraacetic acid disodium salt] in 1:1(*v/v*) ratio. Crystallization was observed in ~2 d for all samples, and the crystals reached their maximal size within ~7 d. The *Mj*YcaO-nucleotide crystals were coated with cryo protection solution [precipitate solution with 35% (*w/v*) threitol] prior to flash frozen in liquid nitrogen. The *Mj*YcaO-nucleotide-peptide crystals were flash frozen without cryo protection.

All crystallographic measurements were collected on a CCD device at an insertion device synchrotron source at Sector 21 ID (LS-CAT, Advanced Photon Source, Argonne National Labs, IL). Data were indexed, integrated, and scaled using XDS [3], as implemented in the autoPROC suite [4]. Crystallographic phases were determined by the molecular replacement method as implemented in Phaser [5] using the coordinates of previously published *Mk*YcaO structure (PDB entry 6CIB).[1] For each structure, iterative model building was carried out using phenix-refine [6] and further improved by manual fitting and adjustment using COOT [7]. Cross-validation with 5% of the data for the calculation of the free R factor [8] was utilized throughout model building process in order to monitor building bias. The stereochemistry of all models was routinely monitored throughout the course of refinement using PROCHECK [9]. Refinement of the structures made use of TLS parameters. Data collection and refinement parameters are provided in Table S2. Coordinates for the *Mj*YcaO structures can be accessed under PDB codes XXX.

Fluorescence polarization (FP) binding assay. Peptide binding to *Mj*YcaO and its variants was assessed by equilibrium FP at 25 °C in non-binding-surface, 384-black-well polystyrene microplates (Corning) using a Synergy H4 Hybrid plate reader (BioTek) with λ_{ex} = 485 nm and λ_{em} = 538 nm. The data were recorded using Gen5 software. For each titration, protein was serially diluted into the binding buffer [50

mM HEPES, pH 7.5, 300 mM NaCl, 2 mM ATP, 2.5% (*v/v*) glycerol], mixed with 25 nM fluorescein isothiocyanate (FITC)-labeled McrA 11-mer peptide as reported before,[1] and equilibrated for 1 h with shaking before data acquisition. Data from three independent titrations were background subtracted and fitted using a standard dose-response curve in OriginPro9.1 (OriginLab) to calculate the dissociation constants (K_D).

For competition FP binding experiments, serial dilutions of MBP-tagged McrA peptide and its variants were mixed with 25 nM FITC-labeled McrA 11-mer peptide and 3 μM MBP-tagged *Mj*YcaO. The mixture was equilibrated for 1 h with shaking before data acquisition as described above. IC_{50} values were determined from the average of triplicate data by fitting with the standard, dose-response curve in OriginPro9.1. Competitive binding constants K_i were calculated from experimentally-determined IC_{50} values and the K_D value as reported [10].

HPLC-based kinetics studies and conversion analysis. Reactions were initiated by adding 1 μM *Mj*YcaO into a mixture of 2 mM Na_2S, 5 mM ATP, and *Mj*McrA 11-mer peptide of concentration varying from 0-600 μM (final concentrations listed for all components) in the buffer containing 100 mM NaCl, 20 mM $MgCl_2$, 50 mM HEPES pH 7.5 and 5 mM dithiothreitol. The reactions were allowed to proceed at 37 °C for 1 min and then quenched by addition of one volume of acetonitrile and 2 volumes of dimethyl sulfoxide prior to the removal of precipitate by centrifugation. The supernatants were subjected (80 μL) to High Performance Liquid Chromatography (HPLC) analysis using a C18 column (Grace/Alltech, 4.6 × 250 mm, 5 μm particle size). Acetonitrile and 0.1% (*v/v*) formic acid were used as the mobile phase. A linear gradient of 5 to 100% acetonitrile over 25 minutes at 1 ml/min was used to separate the thioamide-containing peptide from the starting material. Their identities were confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) with 2,5-dihydroxybenzoic acid as matrix using a Bruker UltrafleXtreme instrument (Bruker Daltonics) in reflector positive mode at the University of Illinois School of Chemical Sciences Mass Spectrometry Laboratory. Data analysis was carried out using the Bruker FlexAnalysis software. Reactions were carried out in triplicate. The concentration of thioamidated product was determined based on the area under the corresponding peak (absorbance at 220 nm) and a standard curve obtained from substrate peptide of known concentrations. The initial rates were then determined, plotted against the substrate concentrations and fitted into the Michaelis-Menten model using GraphPad Prism7 to obtain the steady state kinetic parameters.

The same HPLC method was used to quantify the reactivity of different *Mj*YcaO variants. 10 μM MBP-tagged *Mj*YcaO protein, 100 μM McrA 11-mer peptide, 5 mM ATP, and 2 mM Na_2S were allowed to react for 5 h in the buffer containing 125 mM NaCl, 50 mM Tris pH 7.5, and 20 mM $MgCl_2$. Reactions

were quenched by heating at 95 °C for 5 min and the precipitate was removed by centrifugation. Acetonitrile was added into the supernatant to a final concentration of 5% (*v/v*). The mixture was filtered through a 0.22 μm filter and analyzed via the above HPLC method.

Cyclodehydration reaction and its characterization. Reaction were carried out with 100 μM synthetic peptides of sequences corresponding to the McrA 11-mer peptide with the (+1) residue replaced by Cys, Ser or Thr, 20 μM MBP-tagged *Mj*YcaO or *Mk*YcaO in a buffer containing 125 mM NaCl, 50 mM HEPES (pH 7.5), 20 mM $MgCl_2$, 5 mM ATP, 5 mM dithiothreitol (DTT), and 5 mM Na_2HPO_4 (pH 7.5) at 37 °C for 2 h. For the experiments under different pH conditions, Na_2HPO_4 and Tris were omitted and one of the following buffering reagents (50 mM) were instead used: MES for pH 6.5, HEPES for pH 7.0 and 7.5, Tris for pH 8.0 and 8.5, and Bicine for pH 9.0. Reactions were examined by MALDI-TOF-MS as described above.

ATP utilization analysis of the *Mj*YcaO reaction by HPLC. *Mj*YcaO thioamidation reactions were set up with 10 μM MBP-tagged *Mj*YcaO protein, 100 μM peptide, 5 mM ATP, and 2 mM Na_2S in a buffer containing 125 mM NaCl, 50 mM Tris pH 7.5, 20 mM $MgCl_2$, and 5 mM DTT. *Mj*YcaO cyclodehydration reactions were iniatied as above. The reactions were allowed to proceed at 37 °C for 2 h prior to termination by adding a volume of acetonitrile with 0.1% formic acid. The precipitate was removed by centrifugation. The supernatant was dried under vacuum using a Speedvac concentrator (ThermoFisher) and re-dissolved in 100 μL water for HPLC analysis. A C18 column (Grace/Alltech, 4.6 × 2500 mm, 5 μm particle size) was used with 5% acetonitrile in water with 20 mM ammonium acetate (pH 5.0) as the mobile phase isocratically at 1 mL min^{-1}. To confirm the identity of the resultant nucleotides, 120 μL of 20 μM ATP, ADP, and AMP standards were analyzed via the same method and compared to the reactions using their retention times.

Tandem mass spectrometric (MS/MS) characterization of the *Mj*McrA-Y(+1)C 11-mer. The reactions were desalted using a ZipTip and eluted into 80% aq. MeCN. To obtain the hydrolyzed peptide, the reaction mixture was lyophilized (Labconco) and reconstituted in 100 μL H_2O^{18} (Millipore-Sigma) with 1% formic acid. The subsequent sample was desalted using a ZipTip and eluted into 80% aq. MeCN. Desalted samples were directly infused into a ThermoFisher Scientific Orbitrap Fusion ESI-MS using an Advion TriVersa Nanomate 100. The MS was calibrated and tuned with Pierce LTQ Velos ESI Positive Ion Calibration Solution (ThermoFisher). The MS was operated using the following parameters: resolution, 100,000; isolation width (MS/MS), 1 *m/z*; normalized collision energy (MS/MS), 70; activation q value (MS/MS), 0.4; activation time (MS/MS), 30 ms. Fragmentation was performed using

higher-energy collision-induced dissociation (HCD) at 35%. The resulting data were averaged and analyzed using the Qualbrowser application of Xcalibur software (Thermo-Fisher Scientific).

Modeling studies of TruD and PatD. Each sequence of these YcaOs (Table S3) was modeled using the Phyre 2 web server (one-to-one threading mode) [11] using the LynD-leader peptide structure as the template (PDB entry 4V1T) [12]. The computational structure of each YcaO protein was superimposed with *Mj*YcaO-*Mj*McrA complex structure (PDB code: XXX using Chimera 1.10.2 [13]. The structure of *Mj*McrA 11-mer was retained in the model structures, and was modified manually for the purpose of demonstration.

Construction and annotation of sequence similarity networks and phylogenetic trees. All YcaO sequences as of mid-2018 (n = 17,077) were retrieved through the Conserved Domain Database from NCBI. The accession list was used as the "Option D" input for Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST, https://efi.igb.illinois.edu/efi-est) [14] resulting in 5,949 non-redundant sequences. The genes encoded within six open reading frames of the YcaO sequence were examined using the Rapid ORF Description & Evaluation Online (RODEO) bioinformatics tool [15]. TfuA proteins associated with thioamide-forming reactions were identified by profile Hidden-Markov Model (pHMM) PF07812 while the E1-like proteins were identified by TIGR03603 or a previously described [16], custom ocin-ThiF pHMM based on F proteins found in thiopeptide biosynthetic gene clusters [17]. Through this analysis, the YcaO sequences that have a local *tfuA* and methanogenic YcaOs are binned into the TfuA-associated group. Those with E1-like proteins in the local genomic region are known to catalyze cyclodehydration and are binned into the E1-associated group. The E1-associated group was manually curated owing to a high level of sequence divergence. YcaO proteins fused to a tetratricopeptide repeat (TPR, PF00515) do not have any reported function. YcaO sequences were considered to possess a Pro-rich C-terminus if there were at least two Pro residues present in its last ten residues.

Sequence Similarity Networks (SSN) were built using the EFI-EST tool with an alignment score of 40 serving as the edge cutoff (equivalent to a BLAST expectation value of 1×10^{-40}). Sequences with 90% or higher identity were conflated into a single node and the SSNs were visualized using the organic layout within Cytoscape [18]. Genomic neighborhood and phylogenetic information were mapped onto the SSN. A diversity-oriented maximum likelihood tree was generated with 1,912 representative YcaO sequences, none of which were more than 40% identical at the amino acid level. These sequences were aligned with MAFFT [19] and the tree was generated by FastTree2 [20] with default parameters and visualized using the online interactive tree of life tool (iTOL) [21].

Purine nucleoside phosphorylase (PNP)-based analysis. To investigate background phosphate production during the thioamidation reaction, reactions were set up as described previously.[1] Briefly, reactions were initiated via the addition of 1 μM MBP-tagged *Mj*YcaO and 0.2 U/mL PNP (Sigma Aldrich) to a mixture of 300 μM *Mj*McrA 11-mer peptide substrate, 400 μM of 2-amino-6-mercapto-7-methylpurineriboside (Berry and Associates), 5 mM ATP, and 2 mM Na_2S (final concentrations listed for all components) in a buffer containing 125 mM NaCl, 50 mM Tris pH 7.5, and 20 mM $MgCl_2$. Control reactions were initiated in the same way except for certain component(s) being omitted. Reaction temperature was 25 °C and reaction progress was monitored by absorbance changes at 360 nm on a Tecan Infinite M200 plate reader using the kinetic cycle module. Phosphate concentration was calculated based on the resulting purine analog (extinction coefficient at 360 nm = 11,000 M^{-1} cm^{-1}). Reactions were carried out in triplicate and the average curves were plotted in Microsoft Excel. The initial rates were calculated based on the slopes of the linear models constructed by the first four data points.

^{31}P NMR analysis. 5 μL MBP-tagged *Mj*YcaO (5 μM) or storage buffer was equilibrated with 5 mM ATP in a low-salt buffer [50 mM Tris pH 7.5, 80 mM NaCl, 20 mM $MgCl_2$] at 22 °C for 14 h for the experimental and control groups, respectively. The reaction was quenched by heating at 95 °C for 5 min and the precipitate was removed by centrifugation. D_2O (50 μL) was added to the resulting supernatant (450 μL) before transferring the sample into a standard 5-mm NMR tube. The ^{31}P NMR spectrum was obtained on a 600 MHz Agilent NMR at the Carl R. Woese Institute for Genomic Biology, University of Illinois with a 5-mm Agilent AutoTuneX probe, 512 transients, 32,768 points, and a spectral window of −50 to 50 ppm.

Additional software. The interaction map between *Mj*YcaO and *Mj*McrA was produced using LigPlot Plus (SI Figure S4) [22]. The structure-based sequence alignment (SI Figure S7) were generated using ALINE [23] and secondary structure deduced from the coordinates of the refined *Mj*YcaO-*Mj*McrA structure using DSSP [24].

Table S1: Substrate residues important for *Mj*YcaO recognition. Summary of the processing and binding of the *Mj*McrA 11-mer peptide variants by *Mj*YcaO. The McrA fragment [Arg(-5)–Asp(+5)] with the residue naturally thioamidated Gly(0) shown in red (bold text indicates peptide sequence deriving from McrA while italics indicates flanking sequences from heterologous expression). Relative thioamide installation was determined by an end-point MALDI-TOF-MS assay; ++, fully processed; +, partially processed. Binding constants were determined by competitive fluorescence polarization assay of MBP-tagged 11-mer peptide variants against FITC-labeled 11-mer peptide. Error represented as SEM (n = 3).

McrA 11-mer for reactivity assay: *SGS***RLGFYGYDLQD**

McrA 11-mer for binding assay: MBP-*SGSLESTSL***RLGFYGYDLQD***AAALEHHHHHH*

McrA 11-mer variant	K_i (μM)	Relative reactivity
wild-type	1.3 ± 0.1	++
R(-5)A	1.9 ± 0.2	++
L(-4)A	>15	+
G(-3)A	>6	+
F(-2)A	3.7 ± 0.9	+
Y(-1)A	1.7 ± 0.3	+
G(0)A	4.0 ± 0.2	+
Y(+1)A	1.6 ± 0.3	+
D(+2)A	4.2 ± 0.7	++
L(+3)A	6.7 ± 0.8	++
Q(+4)A	1.8 ± 0.3	++

Table S2. Data collection, phasing and refinement statistics

	*Mj*YcaO-ATP-Mg^{2+}	*Mj*YcaO-ATP-McrA (4)
Data collection		
Space group	$P6_322$	$P2_1$
Unit cell dimensions	106.8, 106.8, 182.6	59.3, 149.8, 105.7, 103.6°
Resolution	25.6-2.3 (2.34-2.3)	102.7-1.95 (1.98-1.95)
Total reflections	805,721	788,920
Unique reflections	26,776	125,182
R_{sym} (%)[1]	0.122 (1.57)	0.067 (0.845)
$I/\sigma(I)$ [1]	19.9 (2.2)	17.1 (2.1)
CC (1/2)	0.999 (0.780)	0.999 (0.833)
Completeness (%)[1]	95.8 (97.0)	95.9 (100)
Redundancy	30.1 (22.8)	6.3 (6.4)
Refinement		
Resolution (Å)	25.0-2.3	25.0-1.95
No. reflections	25,367	118,511
R_{work} / R_{free}[2]	18.9/23.8	17.7/22.7
R_{work} / R_{free} (outermost shell)	26.8/31.8	25.1/31.7
Number of atoms		
Protein	2,996	12,364
ATP	31	124
Peptide	-	396
Water	172	962
B-factors		
Protein	67.9	36.8
ATP	71.7	29.8
Peptide	-	35.3
Water	72.1	44.8
R.m.s deviations		
Bond lengths (Å)	0.009	0.007
Bond angles (°)	1.054	0.962
Molprobity clash score	6.71	7.25

1. Highest resolution shell is shown in parenthesis.

2. R-factor = $\Sigma(|F_{obs}|-k|F_{calc}|)/\Sigma\ |F_{obs}|$and R-free is the R value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement.

Table S3. Information of YcaO homologs used in the bioinformatics studies.

Name	Organism	NCBI Accession Identifier
*Mj*YcaO	*Methanocaldococcus jannaschii* DSM 2661	WP_010870606
*Mk*YcaO	*Methanopyrus kandleri* AV19	AAM01332
TvaH	*Streptomyces olivoviridis*	BAN83923
LynD	*Lyngbya* sp. PCC 8106	WP_009787121
TruD	*Prochloron* sp. 06037A	ACA04490
PatD	*Prochloron didemni*	AAY21153
BalhD	*Bacillus thuringiensis* str. Al Hakam	WP_000192967
BmbD	*Streptomyces bottropensis* ATCC 25435	WP_005486705
BmbE	*Streptomyces bottropensis* ATCC 25435	WP_020115555
KlpD	*Klebsiella pneumoniae* subsp. ozaenae ATCC 11296	WP_077257196
*Ec*YcaO	*Escherichia coli* strain K12	WP_001295344
TbtG	*Thermobispora bispora* DSM 43833	WP_013130815
TsrH	*Streptomyces laurentii* ATCC 31255	ACN52298

Table S4. Oligonucleotide primers used in this study. All sequences provided from 5' to 3', "F" indicates forward primers, "R" indicates reverse primers.

Primer name	Primer sequence
*Mj*YcaO-Y66F-F	GGCATCGCGATTCACTTTGGCAAGGGCGCGAAC
*Mj*YcaO-Y66F-R	GTTCGCGCCCTTGCCAAAGTGAATCGCGATGCC
*Mj*YcaO-Y66L-F	GGCATCGCGATTCACCTGGGCAAGGGCGCGAAC
*Mj*YcaO-Y66L-R	GTTCGCGCCCTTGCCCAGGTGAATCGCGATGCC
*Mj*YcaO-Y66G-F	GGCATCGCGATTCACGGTGGCAAGGGCGCGAAC
*Mj*YcaO-Y66G-R	GTTCGCGCCCTTGCCACCGTGAATCGCGATGCC
*Mj*YcaO-Y66A-F	TCGCGATTCACGCGGGCAAGGGCGCGAACGACATCCAAGC
*Mj*YcaO-Y66A-R	GCGCCCTTGCCCGCGTGAATCGCGATGCCCTCTTTACCGTC
*Mj*YcaO-R86K-F	CATGGAAGCGATTGAGAAGTTCAGCGCGAGCTACG
*Mj*YcaO-R86K-R	CGTAGCTCGCGCTGAACTTCTCAATCGCTTCCATG
*Mj*YcaO-R86M-F	CATGGAAGCGATTGAGATGTTCAGCGCGAGCTACG
*Mj*YcaO-R86M-R	CGTAGCTCGCGCTGAACATCTCAATCGCTTCCATG
*Mj*YcaO-R86A-F	AAGCGATTGAGGCGTTCAGCGCGAGCTACGATAAGAACAAAG
*Mj*YcaO-R86A-R	CTCGCGCTGAACGCCTCAATCGCTTCCATGCACGCGCTCAC
*Mj*YcaO-Y143F-F	CGGATGCGGTGTTCTTTCCGACCAGCGGCAAG
*Mj*YcaO-Y143F-R	CTTGCCGCTGGTCGGAAAGAACACCGCATCCG
*Mj*YcaO-Y143L-F	CGGATGCGGTGTTCCTGCCGACCAGCGGCAAG
*Mj*YcaO-Y143L-R	CTTGCCGCTGGTCGGCAGGAACACCGCATCCG
*Mj*YcaO-Y143A-F	CGGATGCGGTGTTCGCACCGACCAGCGGCAAG
*Mj*YcaO-Y143A-R	CTTGCCGCTGGTCGGTGCGAACACCGCATCCG
*Mj*YcaO-T154A-F	TTCGTGGTAACGCCAACGGCCTGGCGAGCGGTAACAACCTG
*Mj*YcaO-T154A-R	GCCAGGCCGTTGGCGTTACCACGAAACAGCTTGCCGCTGG
*Mj*YcaO-R177K-F	CCCTGGAAATCATTGAAAAGGATGCGTGGAGCCTGGC
*Mj*YcaO-R177K-R	GCCAGGCTCCACGCATCCTTTTCAATGATTTCCAGGG
*Mj*YcaO-R177M-F	CCCTGGAAATCATTGAAATGGATGCGTGGAGCCTGGC
*Mj*YcaO-R177M-R	GCCAGGCTCCACGCATCCATTTCAATGATTTCCAGGG
*Mj*YcaO-W180F-F	CATTGAACGTGATGCGTTTAGCCTGGCGGATCTGG
*Mj*YcaO-W180F-R	CCAGATCCGCCAGGCTAAACGCATCACGTTCAATG
*Mj*YcaO-W180V-F	CATTGAACGTGATGCGGTGAGCCTGGCGGATCTGG
*Mj*YcaO-W180V-R	CCAGATCCGCCAGGCTCACCGCATCACGTTCAATG
*Mj*YcaO-R278K-F	CAACTGCACGGCTTCAAGCGTGACGCGAAGCTG
*Mj*YcaO-R278K-R	CAGCTTCGCGTCACGCTTGAAGCCGTGCAGTTG
*Mj*YcaO-R278A-F	CAACTGCACGGCTTCGCACGTGACGCGAAGCTG
*Mj*YcaO-R278A-R	CAGCTTCGCGTCACGTGCGAAGCCGTGCAGTTG
*Mj*YcaO-R278D-F	TGCACGGCTTCGATCGTGACGCGAAGCTGCGTGAGGAATTTAC
*Mj*YcaO-R278D-R	TTCGCGTCACGATCGAAGCCGTGCAGTTGGCTCGCACGGC
*Mj*YcaO-W303A-F	TTCACCGTAAAGCGTTCGAGTTTGAAGGCGAGATCAACATTG
*Mj*YcaO-W303A-R	TCAAACTCGAACGCTTTACGGTGAATACGCTTCAGACGCTC

*Mj*YcaO-T366F-F	CATTCCGAAAATGGAAGTGTATTTCATTGACCGTGACCGTCTGAGCCG
*Mj*YcaO-T366F-R	CGGCTCAGACGGTCACGGTCAATGAAATACACTTCCATTTTCGGAATG
*Mj*YcaO-I367F-F	CCGAAAATGGAAGTGTATACCTTTGACCGTGACCGTCTGAGCCG
*Mj*YcaO-I367F-R	CGGCTCAGACGGTCACGGTCAAAGGTATACACTTCCATTTTCGG
*Mj*YcaO-I367W-F	CCGAAAATGGAAGTGTATACCTGGGACCGTGACCGTCTGAGCCG
*Mj*YcaO-I367W-R	CGGCTCAGACGGTCACGGTCCCAGGTATACACTTCCATTTTCGG
*Mj*YcaO-I367Y-F	CCGAAAATGGAAGTGTATACCTACGACCGTGACCGTCTGAGCCG
*Mj*YcaO-I367Y-R	CGGCTCAGACGGTCACGGTCGTAGGTATACACTTCCATTTTCGG
MJ-11mer-R(-5)A-F	CGACAAGCTTAGCCCTCGGCTTCTACGGTTACGACCTGCAG
MJ-11mer-R(-5)A-R	TAGAAGCCGAGGGCTAAGCTTGTCGACTCGAGGGATCCGGA
MJ-11mer-L(-4)A-F	CAAGCTTACGCGCCGGCTTCTACGGTTACGACCTGCAGGAC
MJ-11mer-L(-4)A-R	CCGTAGAAGCCGGCGCGTAAGCTTGTCGACTCGAGGGATCC
MJ-11mer-G(-3)A-F	GCTTACGCCTCGCCTTCTACGGTTACGACCTGCAGGACGCG
MJ-11mer-G(-3)A-R	TAACCGTAGAAGGCGAGGCGTAAGCTTGTCGACTCGAGGGA
MJ-11mer-F(-2)A-F	TACGCCTCGGCGCCTACGGTTACGACCTGCAGGACGCGGCCGCACTCGAG
MJ-11mer-F(-2)A-R	TCCTGCAGGTCGTAACCGTAGGCGCCGAGGCGTAAGCTTGTCGACTCGAG
MJ-11mer-Y(-1)A-F	GCCTCGGCTTCGCCGGTTACGACCTGCAGGACGCGGCCGCACTCGAG
MJ-11mer-Y(-1)A-R	TCCTGCAGGTCGTAACCGGCGAAGCCGAGGCGTAAGCTTGTCGACTC
MJ-11mer-G(0)A-F	GCCTCGGCTTCTACGCGTACGACCTGCAGGACGCGGCCGCACTCGAG
MJ-11mer-G(0)A-R	TCCTGCAGGTCGTACGCGTAGAAGCCGAGGCGTAAGCTTGTCGACTC
MJ-11mer-Y(+1)A-F	GCCTCGGCTTCTACGGTGCCGACCTGCAGGACGCGGCCGCACTCGAG
MJ-11mer-Y(+1)A-R	TCCTGCAGGTCGGCACCGTAGAAGCCGAGGCGTAAGCTTGTCGACTC
MJ-11mer-D(+2)A-F	GCCTCGGCTTCTACGGTTACGCCCTGCAGGACGCGGCCGCACTCGAGCAC
MJ-11mer-D(+2)A-R	GCGTCCTGCAGGGCGTAACCGTAGAAGCCGAGGCGTAAGCTTGTCGACTC
MJ-11mer-L(+3)A-F	ACGGTTACGACGCGCAGGACGCGGCCGCACTCGAGCACCAC
MJ-11mer-L(+3)A-R	GCCGCGTCCTGCGCGTCGTAACCGTAGAAGCCGAGGCGTAAG
MJ-11mer-Q(+4)A-F	GTTACGACCTGGCGGACGCGGCCGCACTCGAGCACCACCAC
MJ-11mer-Q(+4)A-R	GCGGCCGCGTCCGCCAGGTCGTAACCGTAGAAGCCGAGGCG
MJ-11mer-D(+5)A-F	ACGACCTGCAGGCCGCGGCCGCACTCGAGCACCACCACCAC
MJ-11mer-D(+5)A-R	AGTGCGGCCGCGGCCTGCAGGTCGTAACCGTAGAAGCCGAG
MJ-11mer-Y(+1)C-F	GCCTCGGCTTCTACGGTTGCGACCTGCAGGACGCGGCCGCACTCGAG
MJ-11mer-Y(+1)C-R	TCCTGCAGGTCGCAACCGTAGAAGCCGAGGCGTAAGCTTGTCGACTC

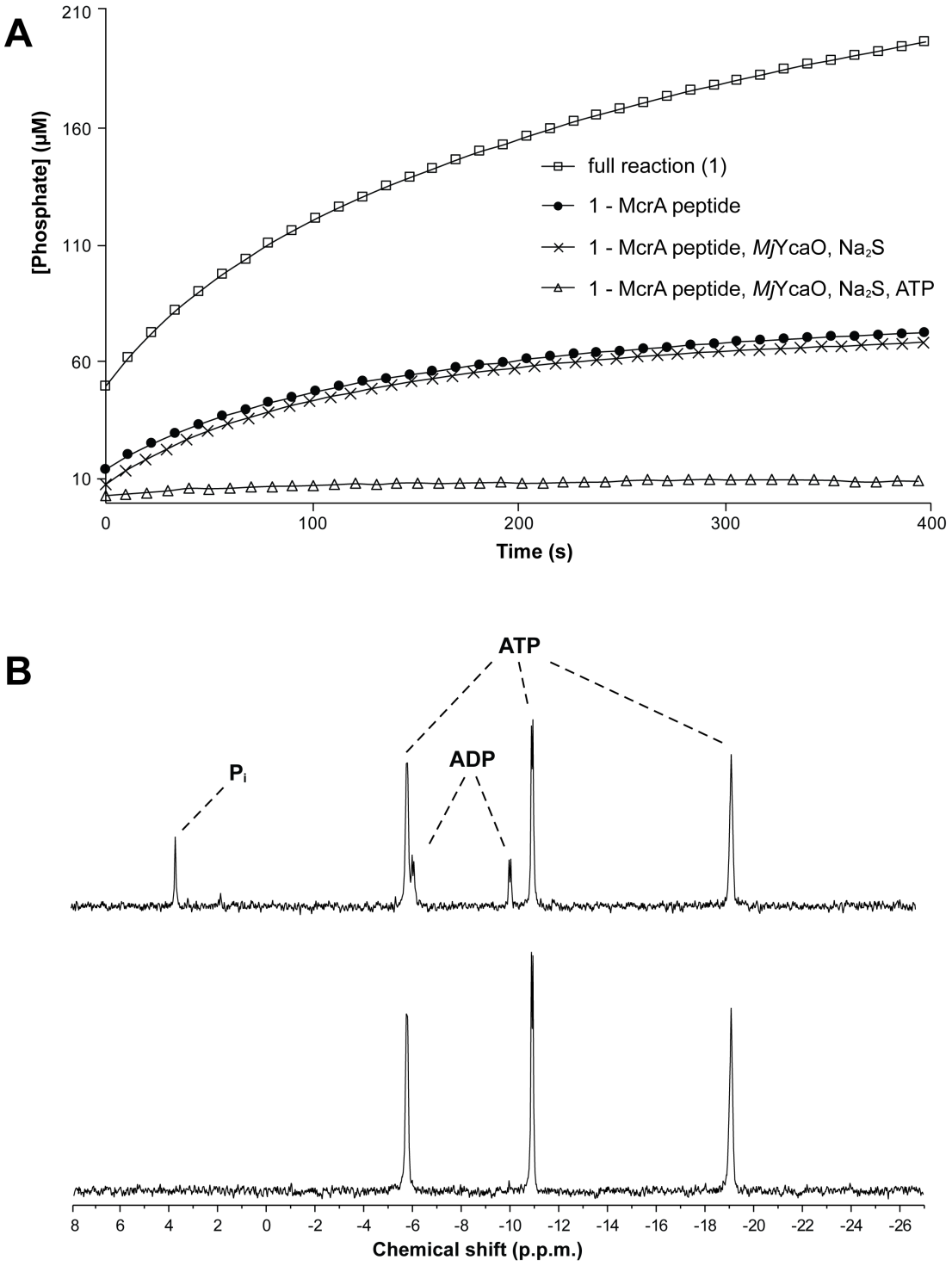


Figure S1. Non-enzymatic phosphate production and *Mj*YcaO background ATPase activity. (A) Phosphate generation was monitored by the PNP assay (n = 3). The initial rate for the full reaction was 9.5×10^{-1} s^{-1}. Upon omission of the peptide substrate the rate was 5.4×10^{-1} s^{-1} (approx. half). **(B)** ^{31}P NMR spectrum of the *Mj*YcaO incubation with ATP showing the formation of ADP and phosphate (P_i) from ATP (top) and the control with no enzyme added (bottom).

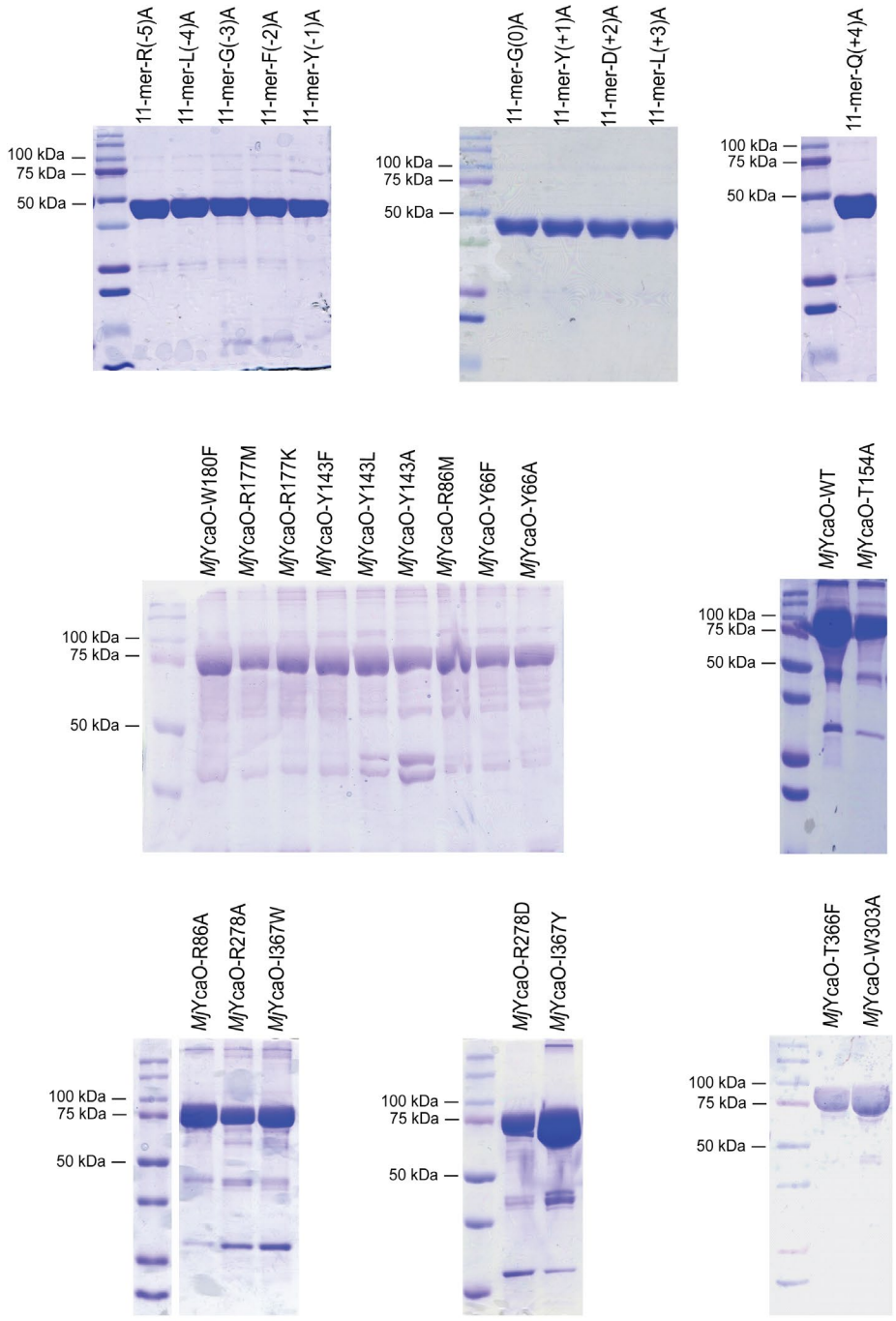


Figure S2. SDS-PAGE analysis of the proteins and peptides used in the study. The purity of MBP-tagged 11-mer peptide variants and MBP-tagged *Mj*YcaO variants are analyzed by 12% SDS-PAGE gel.

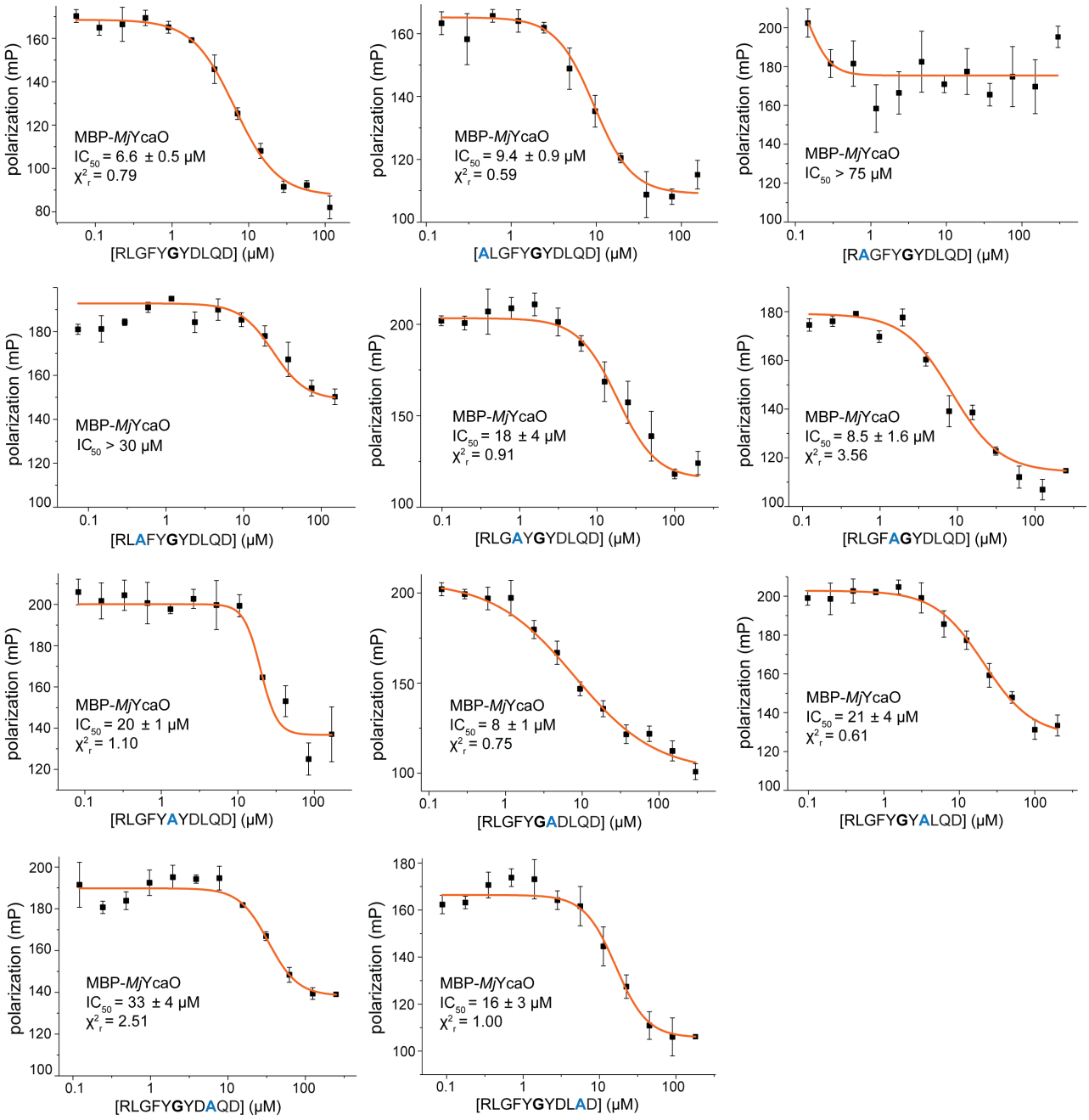


Figure S3. Competitive binding of *Mj*McrA peptide and its variants to *Mj*YcaO. Competitive fluorescence polarization traces for wild-type (WT) MBP-tagged *Mj*McrA 11-mer and its variants are shown. These peptides were titrated against *Mj*YcaO pre-complexed with FITC-labeled *Mj*McrA 11-mer. Error bars represent standard deviation (s.d.) with n = 3. Errors on IC_{50} are the standard error of the mean given by regression analysis. Reduced Chi-square (χ^2_r) values are reported for all regression analyses with eight degrees of freedom.

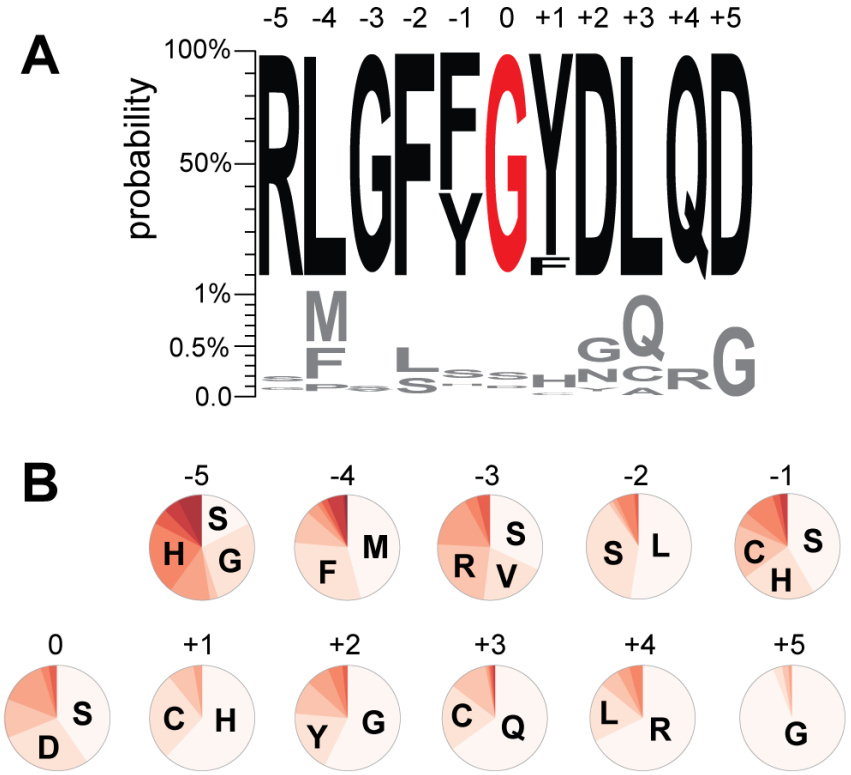
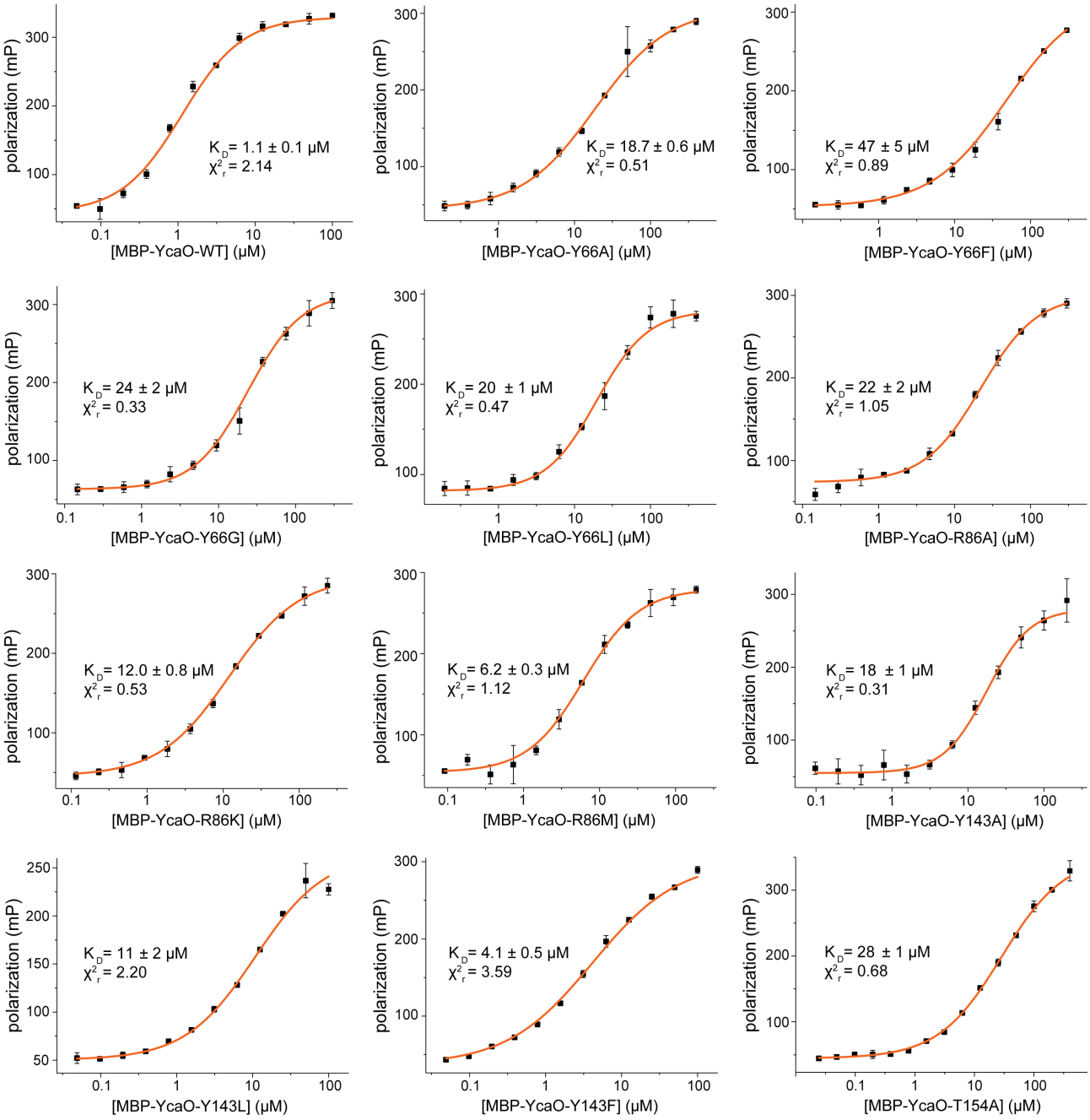


Figure S4. Natural variance of McrA near the thioamidation site. (A) Sequence log (WebLogo 3) of the 11 residues of McrA centered on the thioglycine (R443-D453, *Mj*McrA numbering), with the consensus sequence shown on top with the logo of rare variances appended below. McrA sequences (n = 15,301) were retrieved from GenBank via BLASTp against the *Mj*McrA 11-mer region. There were 14,503 sequences (95%) that match the consensus sequence, which were used to generate the top logo. **(B)** Rarer variations (798 sequences, 5%) were identified and aligned to generate the lower logo and scaled according to their frequencies. Variants with Asp and Ser at position (0) suggest that thioaspartic acid and thioserine will naturally occur in MCR proteins.


polarization (mP)
[MBP-YcaO-WT] (μM)
K_D= 1.1 ± 0.1 μM
χ^2_r = 2.14
[MBP-YcaO-Y66A] (μM)
K_D= 18.7 ± 0.6 μM
χ^2_r = 0.51
[MBP-YcaO-Y66F] (μM)
K_D= 47 ± 5 μM
χ^2_r = 0.89
[MBP-YcaO-Y66G] (μM)
K_D= 24 ± 2 μM
χ^2_r = 0.33
[MBP-YcaO-Y66L] (μM)
K_D= 20 ± 1 μM
χ^2_r = 0.47
[MBP-YcaO-R86A] (μM)
K_D= 22 ± 2 μM
χ^2_r = 1.05
[MBP-YcaO-R86K] (μM)
K_D= 12.0 ± 0.8 μM
χ^2_r = 0.53
[MBP-YcaO-R86M] (μM)
K_D= 6.2 ± 0.3 μM
χ^2_r = 1.12
[MBP-YcaO-Y143A] (μM)
K_D= 18 ± 1 μM
χ^2_r = 0.31
[MBP-YcaO-Y143L] (μM)
K_D= 11 ± 2 μM
χ^2_r = 2.20
[MBP-YcaO-Y143F] (μM)
K_D= 4.1 ± 0.5 μM
χ^2_r = 3.59
[MBP-YcaO-T154A] (μM)
K_D= 28 ± 1 μM
χ^2_r = 0.68

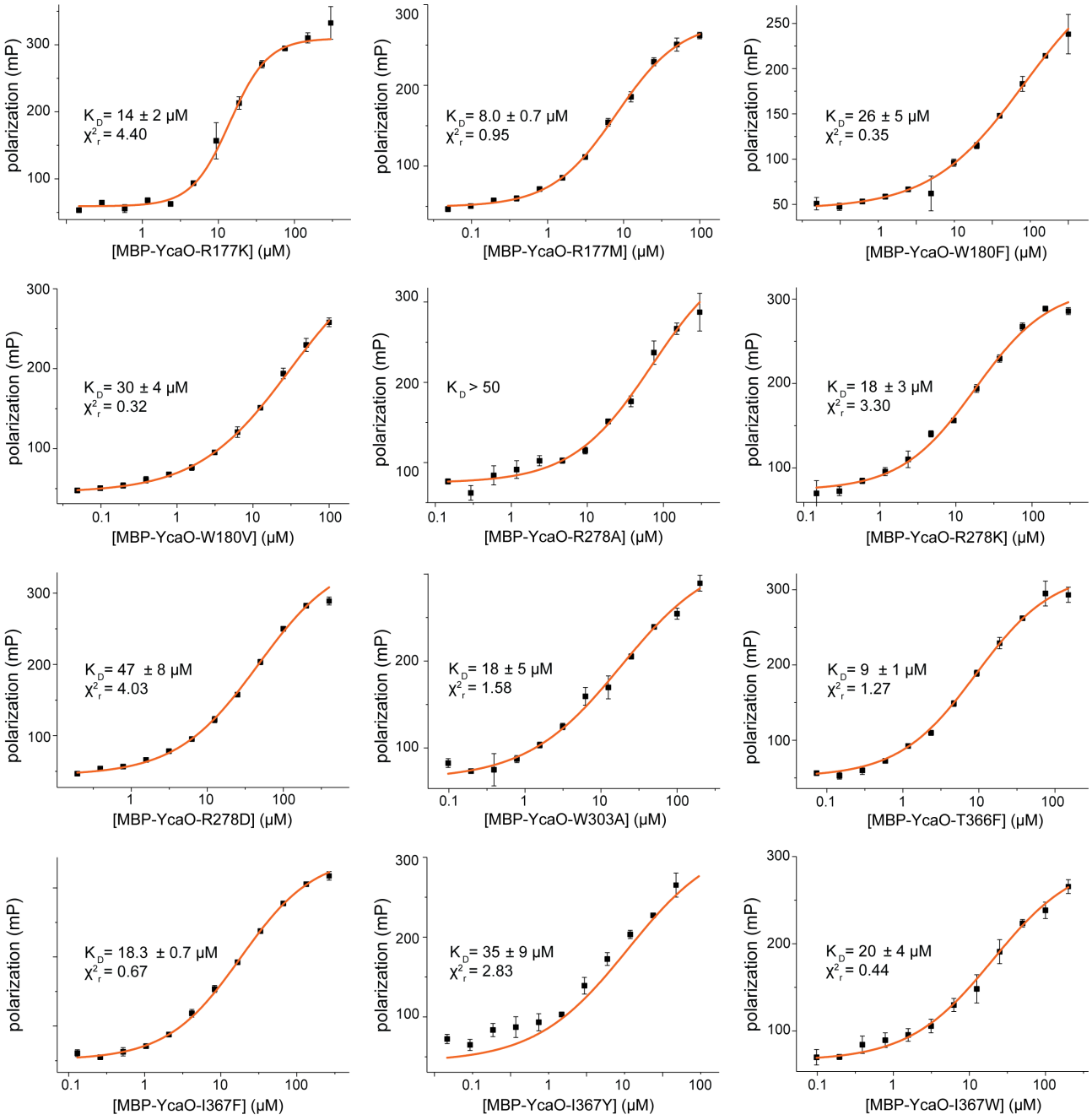


Figure S5. Fluorescence polarization analysis of *Mj*YcaO and its variants. Fluorescence polarization curves for *Mj*YcaO and its variants binding the FITC-labeled *Mj*McrA 11-mer are shown. The MBP-tagged *Mj*YcaO and variants were titrated against the FITC-labeled peptide. Error bars represent standard deviation (s.d.) of three independent replicates. Errors on K_D is the standard error of the mean given by regression analysis. Reduced Chi-square (χ^2_r) values are reported for all regression analyses with eight degrees of freedom.

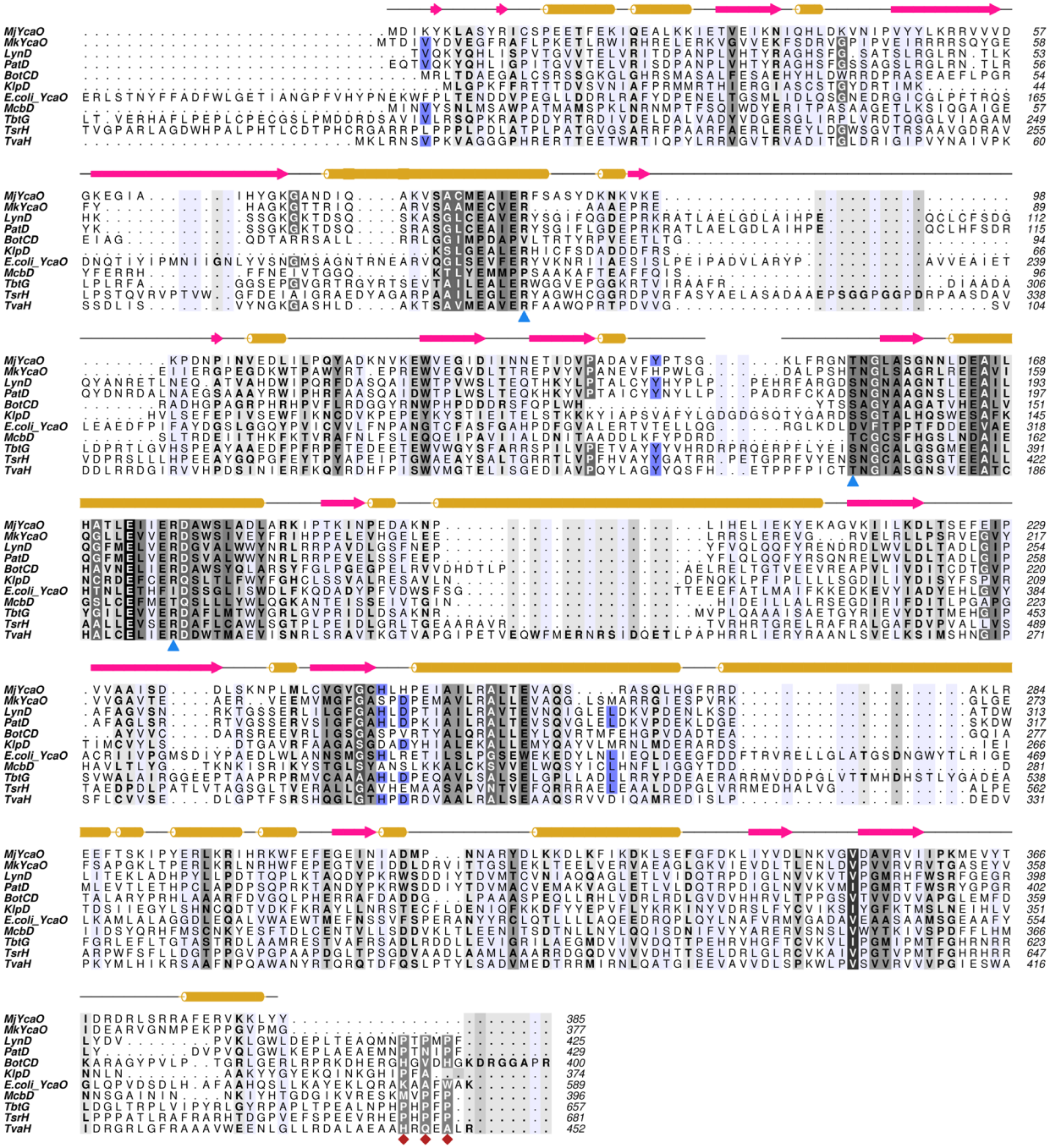


Figure S6. Structure-based multiple sequence alignments of YcaOs. Listed YcaO enzymes (**Table S3**) include members that catalyze the formation of thioamides (*Mj*YcaO, *Mk*YcaO, and TvaH), azoline heterocycles (LynD, PatD, McbD, KlpD, TbtG, and TsrH), and macrolactamidines (BmbD). The founding member *Ec*YcaO is also shown. Residues implicated in catalysis are shown with a blue triangle, and the PXPXP motif is indicated with red diamonds. Figure generated using ALINE[23] and secondary structure elements derived using DSSP.

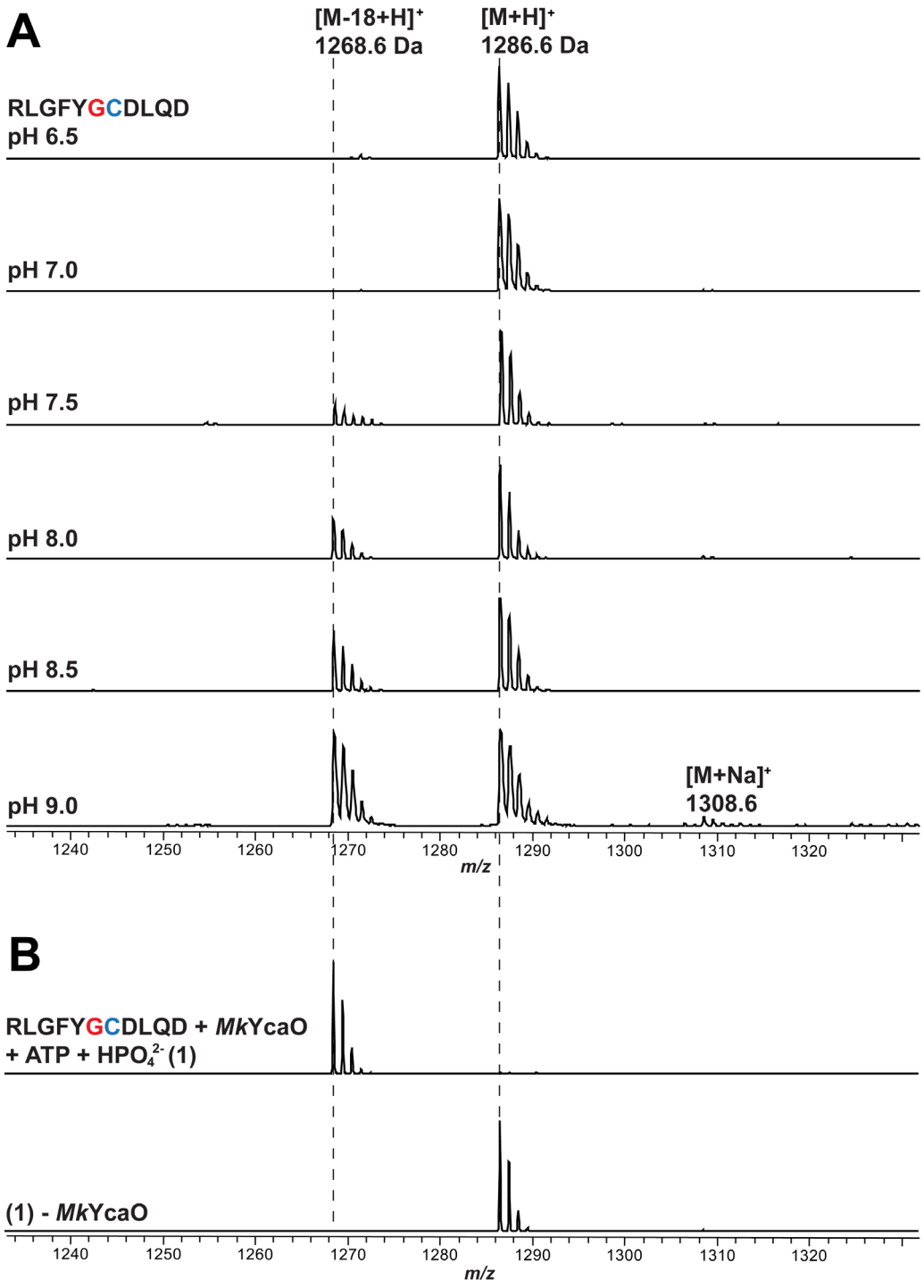


Figure S7. *Mj*McrA-Y(+1)C 11-mer peptide is cyclodehydrated by *Mj*YcaO or *Mk*YcaO. The sequence of the peptide substrate is shown with the thioamidation site in red and the Cys(+1) colored blue. The *m/z* 1286.6 Da peak corresponds to the unmodified peptide and the *m/z* 1268.6 Da peak indicates cyclodehydration (loss of water). **(A)** The peptide substrate is reacted with *Mj*YcaO with increasing pH. **(B)** The peptide substrate is reacted with *Mk*YcaO at pH 7.0 with additional phosphate present (lower spectrum is a control that omitted the YcaO enzyme).

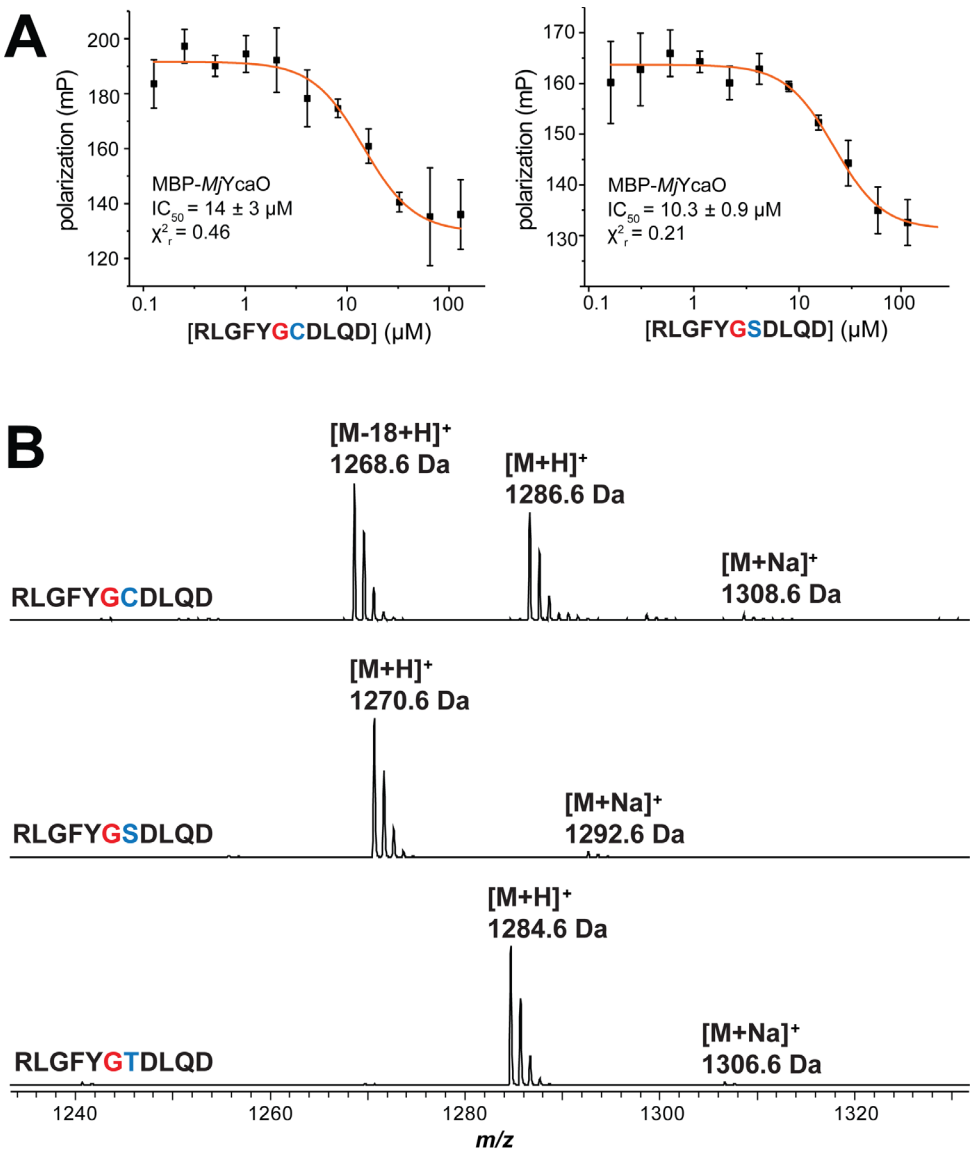
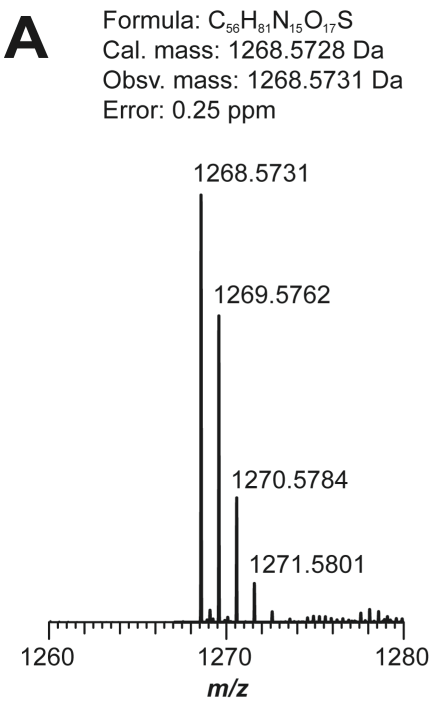


Figure S8. Binding and reaction of *Mj*McrA 11-mer (+1) variants with *Mj*YcaO. The sequence of the peptide substrate is shown with the thioamidation site in red and the (+1) residue colored blue. **(A)** Competitive FP traces for MBP-tagged 11-mer with sequence shown. The peptides were titrated against *Mj*YcaO pre-complexed with FITC-labeled *Mj*McrA 11-mer. Error bars represent standard deviation (s.d.) with n = 3. Errors on IC_{50} are the standard error of the mean given by regression analysis. Reduced Chi-square (χ^2_r) values are reported for all regression analyses with eight degrees of freedom. Y(+1)C and Y(+1)S peptides display similar IC_{50}. **(B)** For the Y(+1)C 11-mer, the *m/z* 1286.6 Da peak corresponds to the unmodified peptide and the *m/z* 1268.6 Da peak indicates cyclodehydration. For the Y(+1)S 11-mer, the *m/z* 1270.6 Da peak corresponds to the unmodified peptide. For the Y(+1)S 11-mer, the *m/z* 1284.6 Da peak corresponds to the unmodified peptide.



B

Species	Calculated mass (Da)	Observed mass (Da)	Error (ppm)
b1	157.10843	157.1088	2.36
b2	270.19250	270.1929	1.48
b3	327.21396	327.2144	1.34
b4	474.28237	474.2829	1.12
b5	637.34570	637.3461	0.63
b7	779.36579	779.3663	0.65
deaminated b8	877.36618	877.3665	0.36
b8	894.39273	894.3931	0.41
b9	1007.47680	1007.4772	0.40
deaminated b10	1118.50882	1118.5095	0.61
b10	1135.53537	1135.5360	0.55
y2	262.10341	262.1039	1.87
y3	375.18747	375.1879	1.15
y7	795.29783	795.2980	0.21
y9	999.38771	999.3887	0.99
y10	1112.47177	1112.4729	1.02
deaminated a2	225.1710	225.1714	1.42
deaminated a4	429.2609	429.2614	1.12
a4	446.2875	446.2880	1.25
deaminated a5	592.3242	592.3248	1.01
a5	609.3508	609.3514	1.02

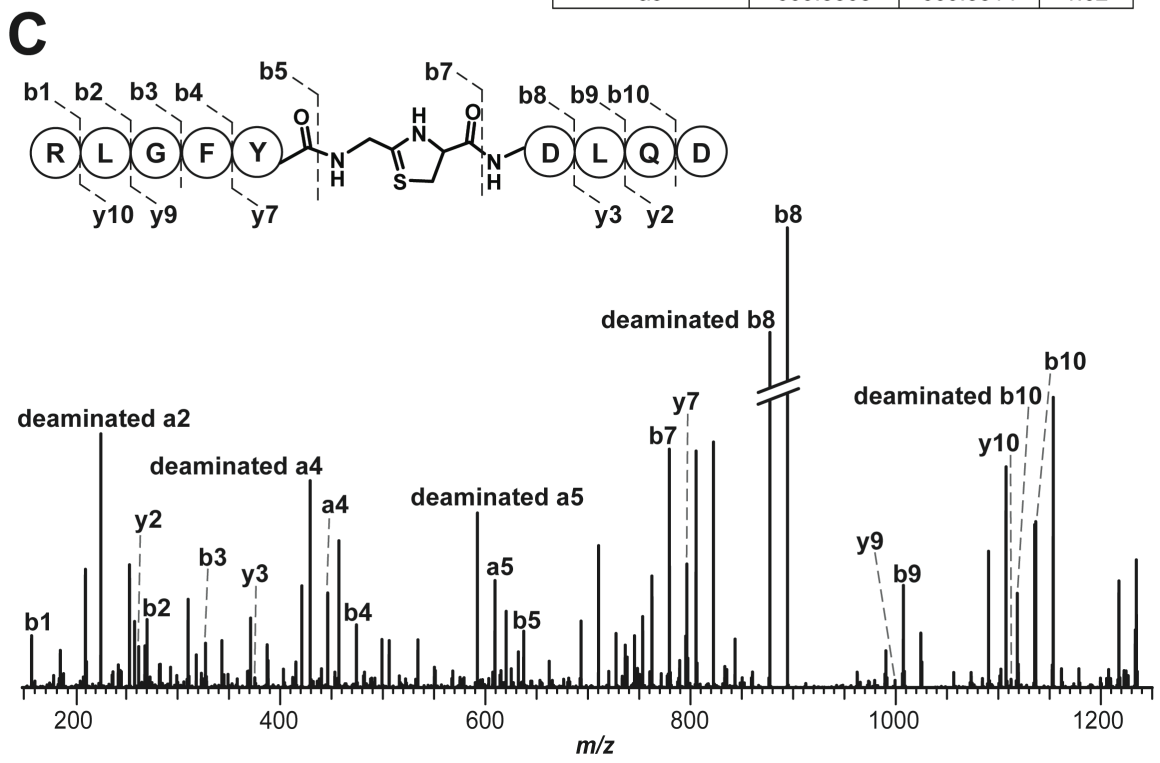
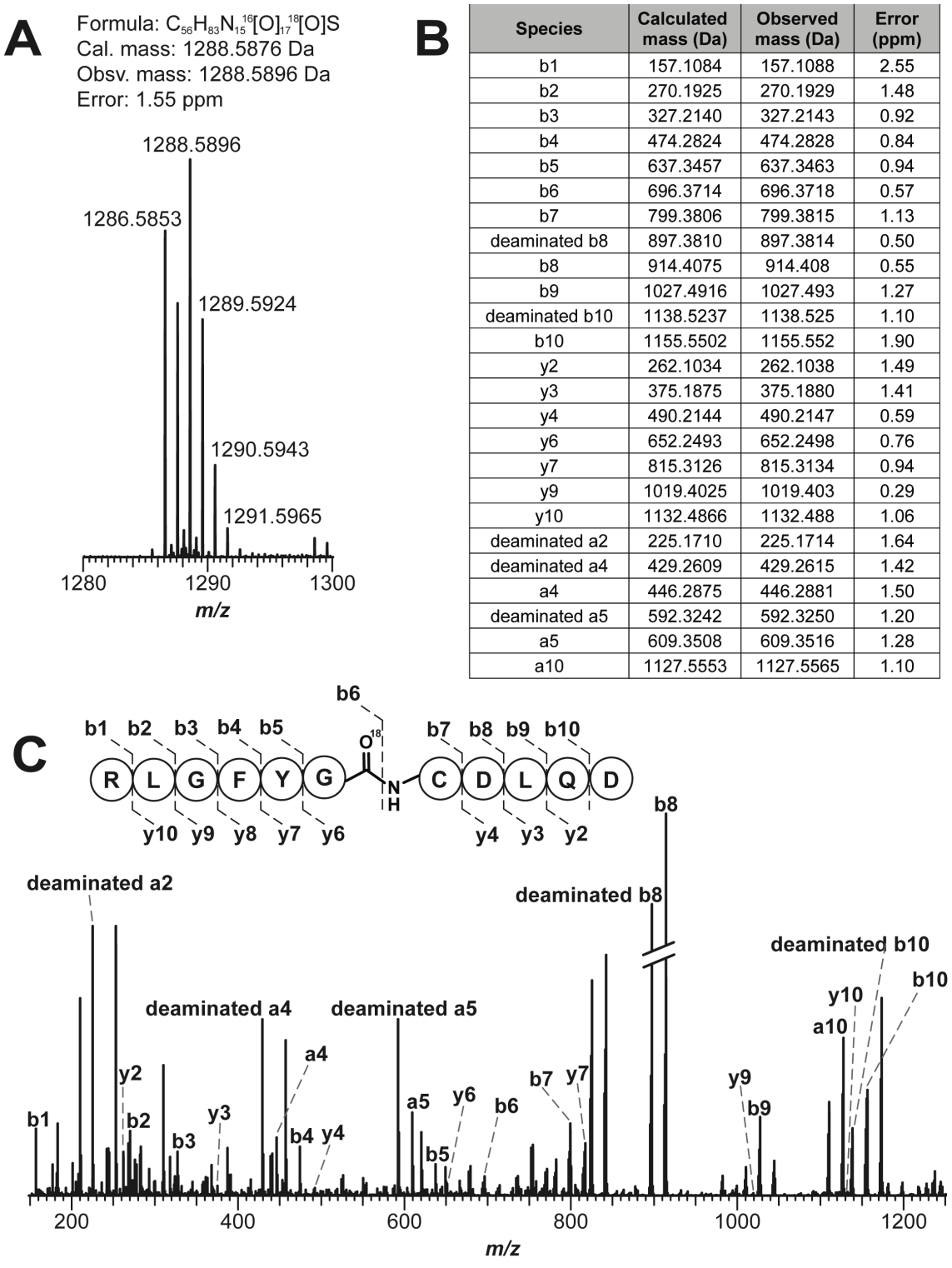


Figure S9. *Mj*YcaO reaction with *Mj*McrA-Y(+1)C 11-mer peptide. *Mj*YcaO catalyzes thiazoline formation on the *Mj*McrA-Y(+1)C 11-mer peptide. **(A)** The product was observed by high-resolution mass spectrometry. **(B)** The 1268.57 Da species was subjected to CID (collision-induced dissociation) conditions with assigned ions indicated in tabular form. **(C)** MS/MS spectral data localize the thiazoline to the C(+1) position.



Species	Calculated mass (Da)	Observed mass (Da)	Error (ppm)
b1	157.1084	157.1088	2.55
b2	270.1925	270.1929	1.48
b3	327.2140	327.2143	0.92
b4	474.2824	474.2828	0.84
b5	637.3457	637.3463	0.94
b6	696.3714	696.3718	0.57
b7	799.3806	799.3815	1.13
deaminated b8	897.3810	897.3814	0.50
b8	914.4075	914.408	0.55
b9	1027.4916	1027.493	1.27
deaminated b10	1138.5237	1138.525	1.10
b10	1155.5502	1155.552	1.90
y2	262.1034	262.1038	1.49
y3	375.1875	375.1880	1.41
y4	490.2144	490.2147	0.59
y6	652.2493	652.2498	0.76
y7	815.3126	815.3134	0.94
y9	1019.4025	1019.403	0.29
y10	1132.4866	1132.488	1.06
deaminated a2	225.1710	225.1714	1.64
deaminated a4	429.2609	429.2615	1.42
a4	446.2875	446.2881	1.50
deaminated a5	592.3242	592.3250	1.20
a5	609.3508	609.3516	1.28
a10	1127.5553	1127.5565	1.10

Figure S10. *Mj*YcaO-modified *Mj*McrA-Y(+1)C 11-mer peptide hydrolyzed in H_2O^{18}. Acid-mediated hydrolysis in H_2O^{18} was performed on the thiazoline-containing *Mj*McrA-Y(+1)C 11-mer peptide (Fig S8) to confirm its identity. **(A)** The product shows incorporation of one O^{18} (1288.59 Da). **(B)** The 1288.59 Da species was subjected to CID with assigned ions indicated in tabular form. **(C)** MS/MS spectral data localize the O^{18} to the amide bond between G(0) and C(+1).

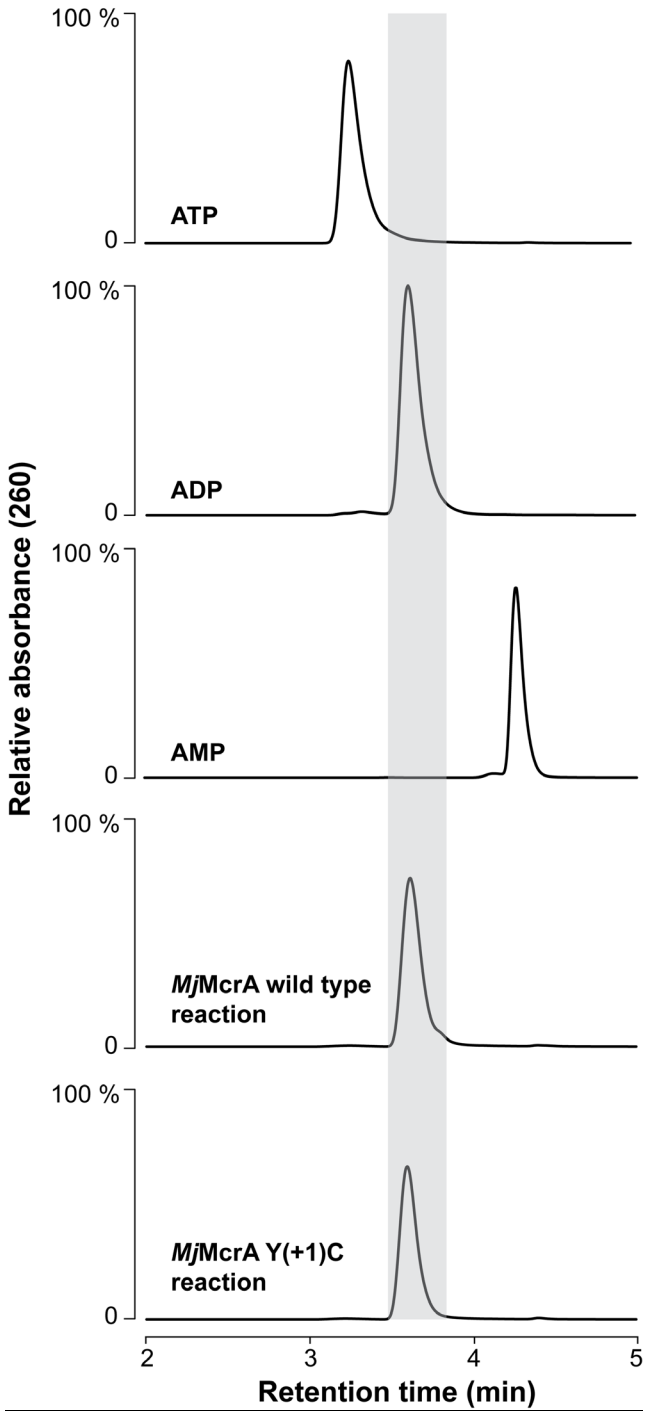


Figure S11. ATP utilization analysis of the *Mj*YcaO reactions. The HPLC traces (monitoring 260 nm absorbance) of ATP, ADP, and AMP standards, as well as the *Mj*YcaO reactions with *Mj*McrA 11-mer and *Mj*McrA-Y(+1)C 11-mer, are stacked. The *Mj*YcaO reaction with *Mj*McrA 11-mer was previously reported to generate ADP by ^{31}P NMR.[1]

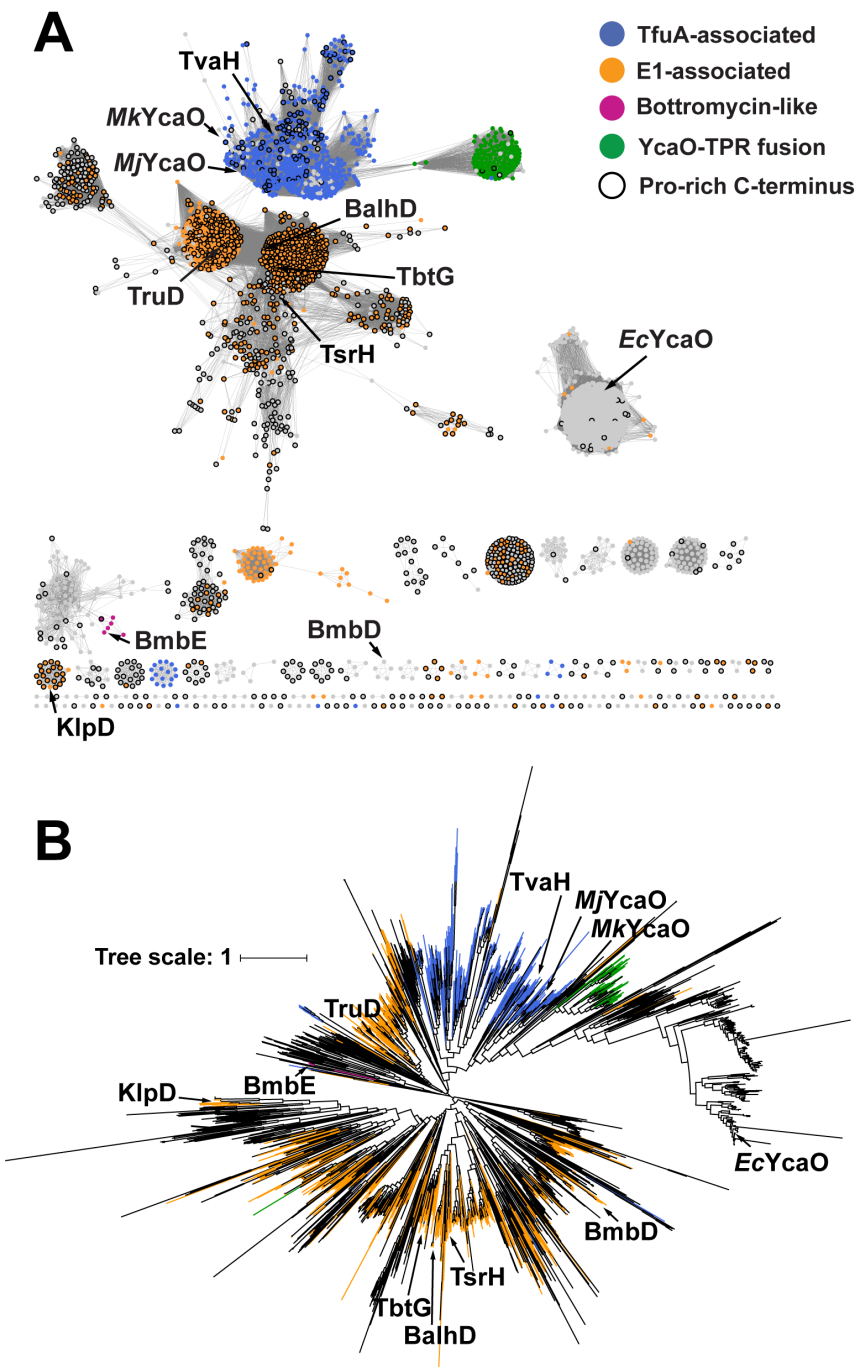


Figure S12. Function analysis of the YcaO superfamily. (A) The SSN was generated for 5,949 YcaO sequences retrieved from UniProt database. Protein sequences of 90% or higher identity are conflated to a single node and an alignment score of 40 (BLAST-P expectation value of 10^{-40}) was used as the edge cutoff. The TfuA- and E1-associated YcaO nodes are colored based on the local co-occurrence of the partner proteins (within 6 open-reading frames). YcaO sequences with Pro-rich C-termini are highlighted with thicker outlines. Representative YcaOs from Table S1 are indicated. **(B)** A diversity-oriented maximum likelihood tree was generated with 1,912 representative YcaO sequences, none of which are more than 40% identical. These sequences were aligned with MAFFT[19] and the tree was generated using FastTree2[20] with default parameters. Annotations are the same as panel A.

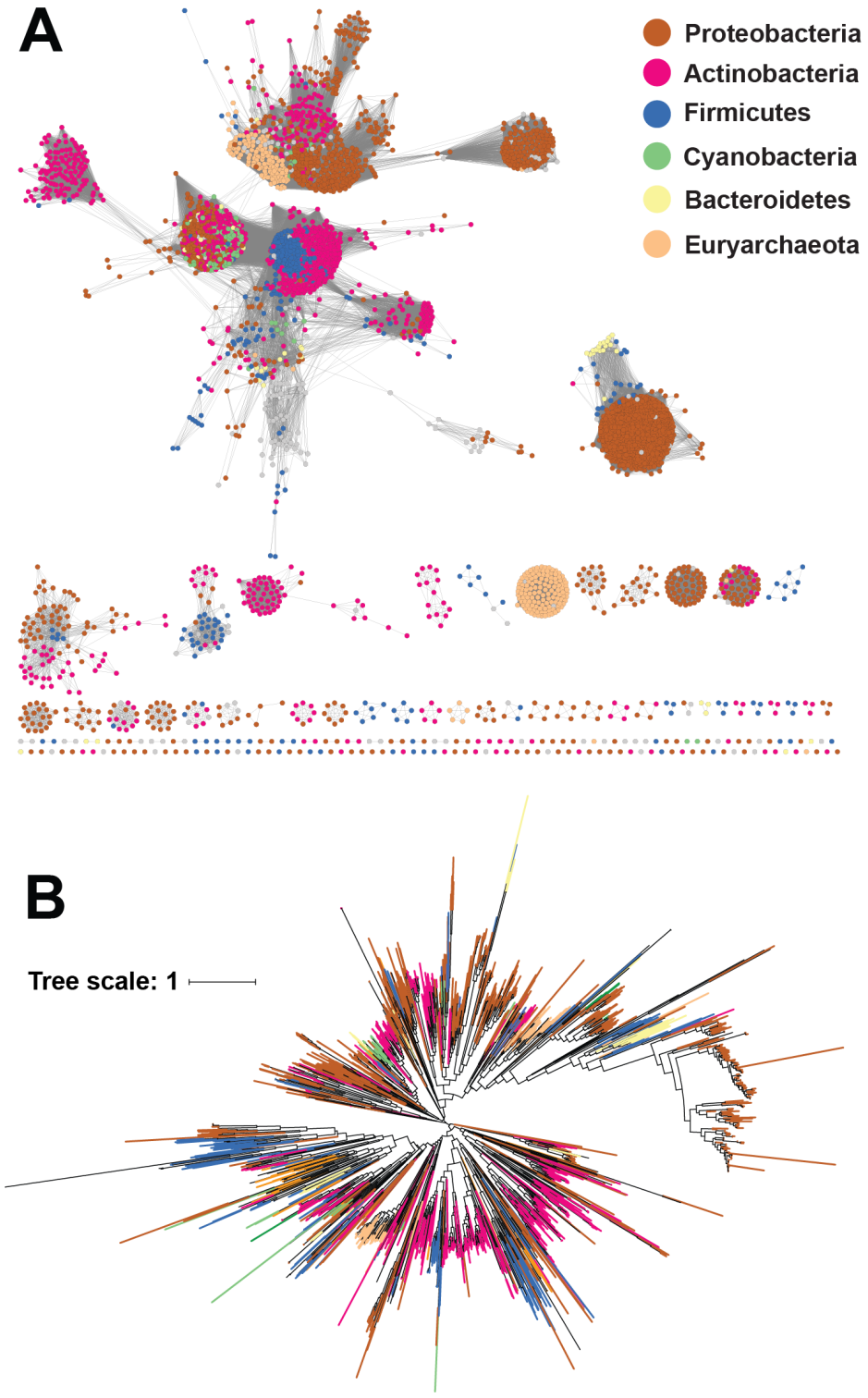


Figure S13. Phylogenetic analysis of the YcaO superfamily. The same SSN and phylogenetic tree as in Figure S13 are recolored here based on phylum. **(A)** SSN with the most prevalent six phyla colored as indicated. **(B)** Phylogenetic tree, similarly colored.

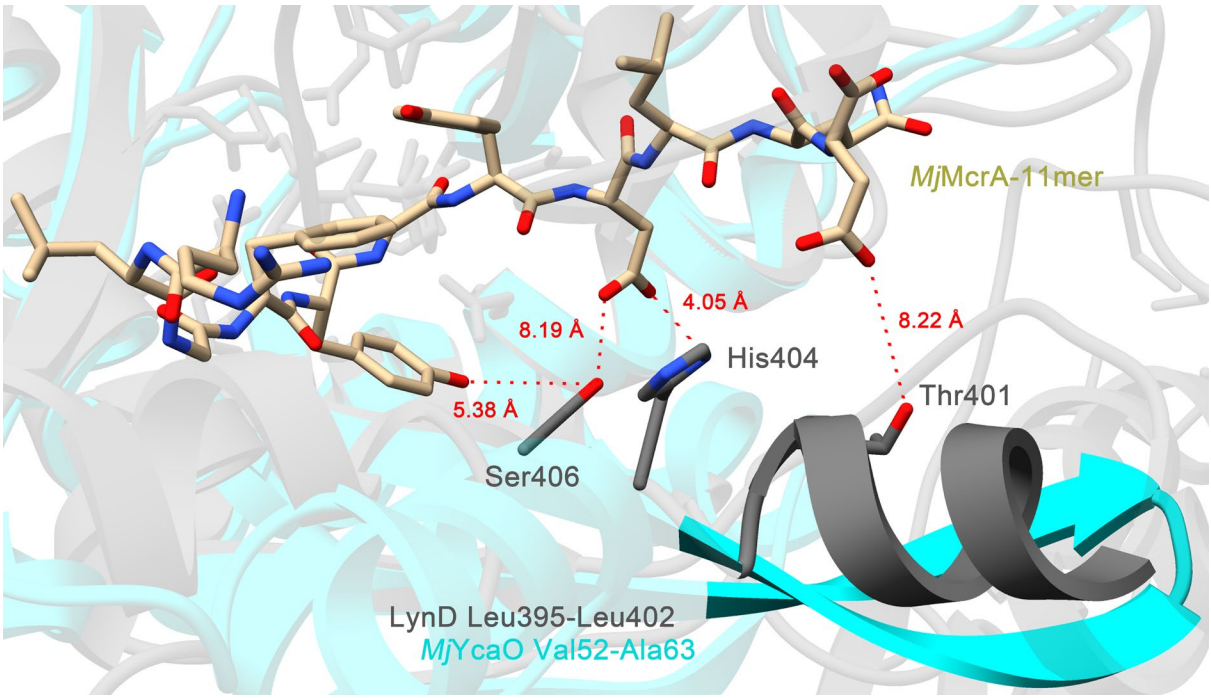


Figure S14. Structure-based comparison of *Mj*YcaO and LynD. The structure of the *Mj*YcaO (cyan) and *Mj*McrA (tan sticks) complex is superimposed with LynD-leader peptide structure (gray, PDB entry 4V1T).[12]

REFERENCES

1. Mahanta, N.; Liu, A.; Dong, S.; Nair, S. K.; Mitchell, D. A. (2018) Enzymatic reconstitution of ribosomal peptide backbone thioamidation, *Proc. Natl. Acad. Sci. USA 115*, 3030-3035.

2. McPherson, A.; Gavira, J. A. (2014) Introduction to protein crystallization, *Acta Crystallogr F Struct Biol Commun 70*, 2-20.

3. Kabsch, W. (2014) Processing of X-ray snapshots from crystals in random orientations, *Acta Crystallogr D Biol Crystallogr 70*, 2204-2216.

4. Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.; Bricogne, G. (2011) Data processing and analysis with the autoPROC toolbox, *Acta Crystallogr D Biol Crystallogr 67*, 293-302.

5. McCoy, A. J. (2007) Solving structures of protein complexes by molecular replacement with Phaser, *Acta Crystallogr D Biol Crystallogr 63*, 32-41.

6. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine, *Acta Crystallogr D Biol Crystallogr 68*, 352-367.

7. Emsley, P.; Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.

8. Brunger, A. T. (1997) Free R value: cross-validation in crystallography, *Methods Enzymol 277*, 366-396.

9. Laskowski, R. A.; Rullmannn, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J Biomol NMR 8*, 477-486.

10. Dunbar, K. L.; Tietz, J. I.; Cox, C. L.; Burkhart, B. J.; Mitchell, D. A. (2015) Identification of an auxiliary leader peptide-binding protein required for azoline formation in ribosomal natural products, *J. Am. Chem. Soc. 137*, 7672-7677.

11. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis, *Nat Protoc 10*, 845-858.

12. Koehnke, J.; Mann, G.; Bent, A. F.; Ludewig, H.; Shirran, S.; Botting, C.; Lebl, T.; Houssen, W.; Jaspars, M.; Naismith, J. H. (2015) Structural analysis of leader peptide binding enables leader-free cyanobactin processing, *Nat. Chem. Biol. 11*, 558-563.

13. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem 25*, 1605-1612.

14. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. (2015) Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks, *Biochim Biophys Acta 1854*, 1019-1037.

15. Tietz, J. I.; Schwalen, C. J.; Patel, P. S.; Maxson, T.; Blair, P. M.; Tai, H. C.; Zakai, U. I.; Mitchell, D. A. (2017) A new genome-mining tool redefines the lasso peptide biosynthetic landscape, *Nat Chem Biol 13*, 470-478.

16. Burroughs, A. M.; Iyer, L. M.; Aravind, L. (2009) Natural history of the E1-like superfamily: implication for adenylation, sulfur transfer, and ubiquitin conjugation, *Proteins 75*, 895-910.

17. Schwalen, C. J.; Hudson, G. A.; Kille, B.; Mitchell, D. A. (2018) Bioinformatic expansion and discovery of thiopeptide antibiotics, *J Am Chem Soc 140*, 9494-9501.

18. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res 13*, 2498-2504.

19. Katoh, K.; Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Mol Biol Evol 30*, 772-780.

20. Price, M. N.; Dehal, P. S.; Arkin, A. P. (2010) FastTree 2--approximately maximum-likelihood trees for large alignments, *PLoS One 5*, e9490.

21. Letunic, I.; Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees, *Nucleic Acids Res 44*, W242-245.

22. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, *Protein Eng 8*, 127-134.

23. Bond, C. S.; Schuttelkopf, A. W. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments, *Acta Crystallogr D Biol Crystallogr 65*, 510-512.

24. Kabsch, W.; Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577-2637.