MECHANISTIC INTERROGATIONS OF RIBOSOMAL AND NON-RIBOSOMAL NATURAL PRODUCT ENZYMES

BY

DILLON P COGAN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biochemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Doctoral Committee:

Professor Satish K. Nair, Chair and Director of Research
Professor Wilfred A. van der Donk
Professor John A. Gerlt
Professor Chad M. Rienstra

ABSTRACT

Described herein is my thesis work towards structural and biochemical characterization of bacterial proteins involved in metabolite regulation and natural product biosynthesis. In the first case, I will describe the structural basis for transcriptional regulation of a small molecule metabolite, *p*-coumarate, and the remaining cases involve enzymes responsible for biosyntheses of complex natural products. Among the natural product cases are those falling under the class of ribosomally synthesized and post-translationally modified peptides (RiPPs), with a particular emphasis on the thiopeptide subclass of RiPPs. Additionally, characterization of an off-loading enzyme involved in hybrid polyketide/non-ribosomal peptide (PK-NRP) natural product biosynthesis is described.

ACKNOWLEDGEMENTS

I dedicate this thesis to my parents John and Elizabeth Cogan for their sources of inspiration and support of my academic pursuits, both emotional and monetary. I additionally thank my academic mentors Satish Nair, Wilfred van der Donk, John Gerlt, Douglas Mitchell, William (Bill) Metcalf, and Chad Rienstra for their insights, written material, motivation, and resources, all of which were instrumental to conducting the research described here. Finally, I thank my collaborators Claudine Baraquet, Graham Hudson, Zhengan Zhang, Manuel Ortega, Linna An, and Ian Bothwell for providing written material, figures, and data which were crucial to developing the following thesis work. Some of these materials were used directly in the work presented herein.

TABLE OF CONTENTS

INTRODUCTION

Second to plants, bacteria represent a significant contribution to life on Earth, accounting for ~15% of the global biomass.[1] Estimates of the earliest existing bacteria date as far as 3.8 billion years ago suggesting ample time for emergence of advantageous fitness traits accompanied by divergence of these traits selected for in a variety of habitats.[2] With genome sequencing rates nearly doubling every year, the access to this biodiversity is exponentially growing and greatly outpacing the functional assignment of genes by experimentation.[3] Because many of the genes lacking correct annotation likely encode for proteins with novel function, this highlights the need to collaborate experimental and computational protein science to efficiently narrow the gap between sequence and function. Of particular interest to medicine are the bacterial enzymes involved in complex natural product biosynthesis, as many of the therapeutic small molecules (~50% in the case of cancer-specific drugs) are directly taken from nature or closely related analogs thereof.[4]

Bacteria are master organic chemists with elegant strategies to produce both the primary metabolites for energy production as well as the secondary metabolites involved in signaling, nutrient sequestration, and chemical weaponry, for example. Owing to their impressive ability to effectively synthesize bioactive compounds in resource limited environments, bacteria provide useful examples of synthetic routes towards potentially therapeutic compounds. As aforementioned, much of the sequence space remains understudied in terms of molecular function, suggesting enzymatic strategies to produce small molecules of interest remain undiscovered. In addition to the application of natural products in mitigating dangers posed by microbial and viral infections in humans, the bacterial enzymes involved could serve as environmentally useful, green biocatalysts. In contrast to synthetic routes which may have any one of the following drawbacks of hazardous waste disposal, lower atom efficiency, costly energetics, and lower safety; enzymes offer an environmentally less taxing, more renewable, and safer alternative to chemical syntheses.[5]

Enzymatic synthesis does, however, suffer from its own disadvantages of catalyst stability, solvent tolerance, and substrate permissiveness, calling for protein engineering approaches to improve catalyst performance as well as hybrid, chemoenzymatic strategies which utilize both chemical and enzymatic techniques. Developing analogs of natural bioactive compounds will be highly important in the face of rising antibiotic resistance, and protein engineering represents one potential strategy to modulate the substrate scope and activity to create natural product analogs.[6] Towards this aim, biophysical and biochemical characterization of enzymes will be critical to assess enzyme function and amenability to engineering for downstream drug production and development.

This thesis work presents several accounts of bacterial enzyme characterization whereby structural, binding, and reactivity experiments have been deployed in combination with mutagenic and bioinformatic analyses to enhance our understanding of natural small molecule biosynthesis.

CHAPTER 1: MOLECULAR BASIS OF P-COUMARATE REGULATION BY COUR, A MARR FAMILY TRANSCRIPTIONAL REPRESSOR[a]

1.1 ABSTRACT

The MarR-family transcriptional regulator CouR, from the soil bacterium *Rhodopseudomonas palustris* CGA009, has recently been shown to negatively regulate a *p*-coumarate catabolic operon. Unlike most characterized MarR repressors that respond to small metabolites at concentrations in the millimolar range, repression by CouR is alleviated by the 800 Da ligand *p*-coumaroyl-CoA with high affinity and specificity. Here we report the crystal structures of ligand free CouR as well as the complex with *p*-coumaroyl-CoA, each to 2.1 Å resolution, and the 2.85 Å resolution cocrystal structure of CouR bound to an oligonucleotide bearing the cognate DNA operator sequence. In combination with binding experiments which uncover specific residues important for ligand and DNA recognition, these structures provide glimpses of a MarR family repressor in all possible states, allowing for an understanding for the molecular basis of DNA binding and the conformation alterations that accompany ligand-induced dissociation for activation of the operon.[7]

1.2 INTRODUCTION

Proteins of the **M**ultiple **a**ntibiotic **r**esistance **R**egulator (MarR) family constitute a large group of transcription factors that are widespread in bacteria and archaea. These transcription

---

[a] This research was originally published in the *Journal of Biological Chemistry*. Part of this chapter was directly obtained, with permission, from:

Cogan DP, Baraquet C, Harwood CS, and Nair SK. Structural Basis of Transcriptional Regulation by CouR, a Repressor of Coumarate Catabolism in Rhodopseudomonas palustris. *J Biol Chem*. 2018 Jul 27;293(30):11727-11735. Copyright © 2018, the American Society for Biochemistry and Molecular Biology.

factors control gene expression to regulate such diverse processes as antibiotic resistance, stress-response, aromatic carbon source catabolism, and virulence. An analysis of 19 major families of bacterial transcription factors recently implicated the MarR family as a significant contributor, comprising roughly 8% of the total proteins analyzed. With ca. 130,000 deposited bacterial genomes to date nearly doubling annually, many MarR-family members have and will be annotated as such, however less than 1% of these sequences have been fully characterized with respect to physiological function. The founding member of this family was first identified in multidrug resistant *E. coli*, where the MarR protein regulates an operon encoding the AcrAB-TolC multidrug efflux system, in response to a wide range of antibiotics and phenolic.[8–14]

Most MarR homologs act as gene repressors, and a typical organization of locus orients the *marR* gene distally from the genes that are under transcriptional control. The intergenic region, typically at sites that overlap the -10 and -35 promoter elements, contains a 16-20 base-pair palindromic DNA sequence that harbors the binding site for the transcriptional regulator. Binding of a MarR homodimer to this palindromic region represses expression of both the gene that is regulated as well as the repressor itself. Dissociation of the repressor from DNA and activation of gene expression occurs when binding of a small molecule ligand alters the conformation of the MarR homodimer. As a result of this autoregulated expression, the physiological concentrations of the repressor itself change very slightly, allowing for an exceptionally sensitive response to ligand concentrations.

Structural studies of MarR homologs reveals two sets of N- and C-terminal α-helices that facilitate dimerization, and two winged helix-turn-helix motifs (wHTH) that bind to the palindromic DNA duplex via spacing that is established by dimer formation. Each monomer binds to one half-site of the palindromic DNA inverted repeat sequence, and residues along the dimerization interface help to establish the spacing between the two half-sites. Disruption of dimer formation results in a loss in DNA binding affinity and a loss of the corresponding drug resistance phenotype. A clade of MarR members responds to oxidative stress through the oxidation of Cys residues, which results in a change to a conformation that is incompatible with DNA binding. Biochemical and biophysical studies on MarR family members are limited by low binding affinity for cognate ligands, which often results in confounding data. For example, several small molecule ligands can bind to MarR factors at multiple sites, and the physiological relevance of these interactions is unclear.[15,16]

A recent analysis of the pathway for the catabolism of the aromatic compound *p*-coumarate by the alpha-proteobacterium *Rhodopseudomonas palustris* CGA009 revealed that genes encoding the enzymes for this pathway are regulated by a MarR-like transcriptional repressor. This soil

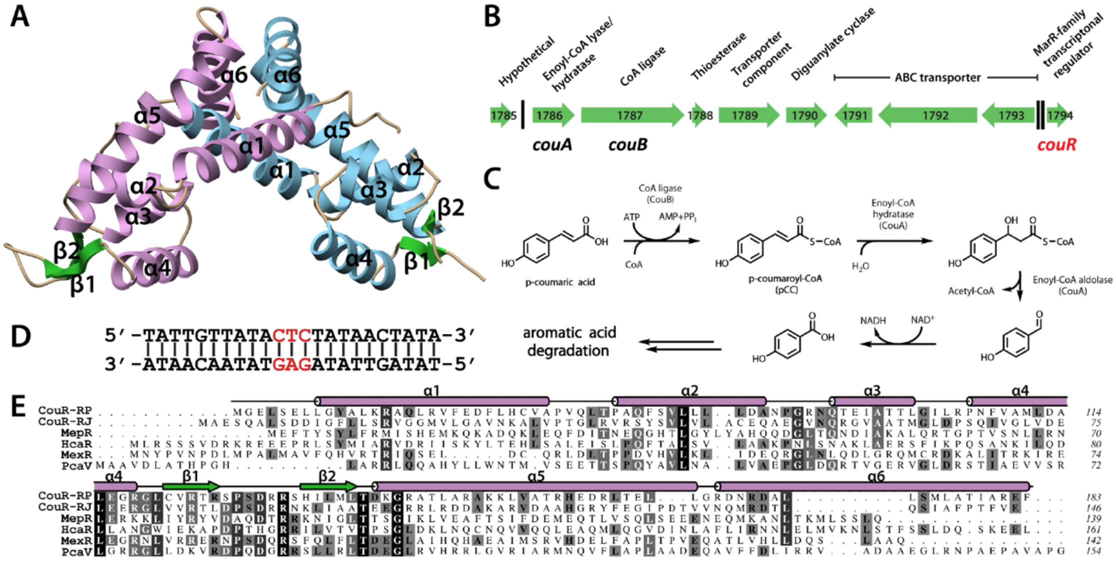


Fig. 1.1: Overall structure of CouR, the *p*-coumarate catabolic operon, and multisequence alignment of related MarR family proteins. *A*, overall structure of the apo CouR homodimer labeled with secondary structural elements (one monomer is colored in *plum* and the other in *sky blue*, wHTH = α3-α4-β1-β2). *B*, the gene organization of the *p*-coumarate catabolic operon regulated by CouR in *R. palustris* CGA009. *Vertical black bars* represent CouR binding motifs (GTTATAnnnTATAAC). *C*, coumarate catabolism pathway in *R. palustris* CGA009. The plant-derived starting material *p*-coumaric acid is converted to its aryl-CoA thioester (pCC) by an ATP-dependent CoA ligase (CouB), followed by hydration of the alkene and retro-aldol cleavage catalyzed by CouA to afford *p*-hydroxybenzaldehdye, which is readily oxidized to *p*-hydroxybenzoic acid and shunted into aromatic acid degradation or fatty acid metabolism. *D*, sequence of oligonucleotide duplex used in crystallization. *E*, a multisequence alignment of MarR family proteins featuring CouR from *R. palustris* CGA009 (CouR-RP), CouR from *R. jostii* RHA1 (CouR-RJ), MepR, HcaR, MexR, and PcaV.[10,11,20–22]

bacterium utilizes plant-derived phenylpropanoids, including *p*-coumarate, as carbon sources by first converting them to acyl/aryl-CoA and subsequently to *p*-hydroxybenzoates, which are degraded aerobically by an oxidative meta-ring cleavage pathway or anaerobically by a reductive aromatic ring degradation pathway.[17,18] In *R. palustris*, the transcription of an operon that encodes for enzymes involved in the catabolism of *p*-coumarate to *p*-hydroxybenzoate is under the control of a MarR family regulator named CouR. Recombinant CouR was shown to bind to an inverted repeated sequence in the -10 region of the promoter, and DNA binding was disrupted by the addition of low micromolar concentrations of *p*-coumaroyl-CoA (pCC) (Fig. 1.1 B,C).[8,17,19]

An orthologous pathway for aromatic acid catabolism has been identified in *Rhodococcus jostii* RHA1, which also utilizes a similar strategy of CoA-thioesterification, followed by β-oxidative deacetylation to generate hydroxybenzoates. However, the constituent enzymes encoded by this operon are different from those of *R. palustris*. For example, the *R. jostii* RHA1 cluster encodes an aryl-CoA dehydrogenase, and an aryl-CoA ligase (CouL) that can utilize dihydroferulate as a substrate, suggesting that the pathway functions on dihydro-*p*-hydroxycinnamic acids rather than their unsaturated *p*-hydroxycinnamate counterparts. Lastly, although the MarR family repressor that regulates the *R. jostii* RHA1 operon has been named CouR, the sequence identity of this polypeptide to *R. palustris* CouR is only 36%, and the DNA-binding site is divergent in sequence and size, as it contains a 5 nucleotide spacer between the two half-sites, rather than the 3 nucleotide spacer in the *R. palustris* promoter.[23]

The *R. palustris* CouR represents a rare example of a MarR family regulator for which a physiologically relevant ligand has been verified, and the cognate ligand is shown to bind to the receptor with low-micromolar affinity. In order to elucidate the mechanism for recognition of the target promoter, and how binding of a large ligand compromises a protein-DNA interaction to relieve transcriptional repression, we determined the cocrystal structures of CouR in complex with a 23-nucleotide duplex inverted repeat corresponding to the physiological DNA binding site (to 2.85 Å resolution; Fig. 1.2A) and the corresponding structure of the repressor in the inactive state, bound to pCC (to 2.1 Å resolution; Fig. 1.3 A,B). The function of specific amino acids in DNA or ligand recognition were probed by analysis of structure-based site-specific variants using electrophoretic mobility shift assays (EMSAs; for DNA binding) or differential scanning fluorimetry (DSF) assays (for ligand binding). The combined biochemical and structural biological data provide a framework for understanding transcriptional regulation by CouR, which further extends the existing knowledge of MarR family repressors.

1.3 RESULTS

1.3.1 Structure determination and overall structure

Cocrystallization efforts with CouR and various synthetic oligonucleotides bearing the inverted repeat DNA recognition sequence yielded several candidates, but most of these crystals did not diffract beyond 3.5 Å resolution. Altering the length and identity of the nucleotides flanking the recognition sequence finally produced crystals that diffracted to 2.85 Å resolution at an insertion device synchrotron beamline. Structures of ligand-free CouR and a complex with synthetic pCC each produced crystals that diffracted to 2.1 Å resolution. Due to ease of reproducibility, structure determination focused on using crystals of the CouR-DNA complex, and crystallographic phases were determined by the single wavelength anomalous diffraction method from data collected on crystals of selenomethionine labeled CouR. The apo-CouR and CouR-pCC structures were determined by molecular replacement using the coordinates of the polypeptide from the DNA-bound structure, followed by manual rebuilding of the polypeptide.

The overall structure of the CouR homodimer is triangular in shape with a pseudo two-fold axis of symmetry. Each monomer consists of six α-helices and two β-strands containing a wHTH motif, where helices α3-α4 and strands β1-β2 define the DNA binding elements (Fig. 1.1 A,D). Dimerization between the two monomers is mediated by interlocking interactions between helices α1, α5, and α6 from each monomer, resulting in a burial of 4,487-4,497 $Å^2$ of solvent-accessible surface area (depending on the conformational state). Dimerization between the two monomers is mediated by extensive burial of numerous hydrophobic residues to form the structural core. Residues at this interface include Leu51, Leu55, Val62, Phe66 from helix α1; Val152, Leu160, Leu164 from helix α5; and Leu172, Leu176, Ile179 from helix α6. Additional interactions include intramolecular (Asp65-Arg159), as well as intermolecular (Glu46-Lys149, Glu49-Arg169) salt bridges. Most notably, electron density is sparse for the β-strands that contain the wHTH motif, and the wings show different conformation for each of the two monomers. There is also weak density in the presumptive binding-pocket for one of the monomers, which may reflect a weakly bound molecule of HEPES from the buffer component, but this was too poorly defined and not modeled.

A DALI search of the CouR monomer against the Protein Data Bank reveals a strong conservation of secondary structural elements with other transcriptional regulators. The closest homologs include a putative regulator of unknown function from *Pseudomonas aeruginosa* (PDB code 2NNN; Z-score of 17.5; RMSD of 1.4 Å over 133 aligned Cα atoms); the *Streptomyces coelicolor* β-ketoadipate regulator PcaV (PDB code 4FHT; Z-score of 17.1; RMSD of 2.7 Å over 141 aligned Cα atoms); *Escherichia coli* MarR (PDB code 3VOE; Z-score of 17.1; RMSD of 2.7 Å over 136 aligned Cα atoms); and the multidrug efflux regulator MexR from *Pseudomonas aeruginosa* (PDB code 1LNW; Z-score of 17.1; RMSD of 1.7 Å over 140 aligned Cα atoms).

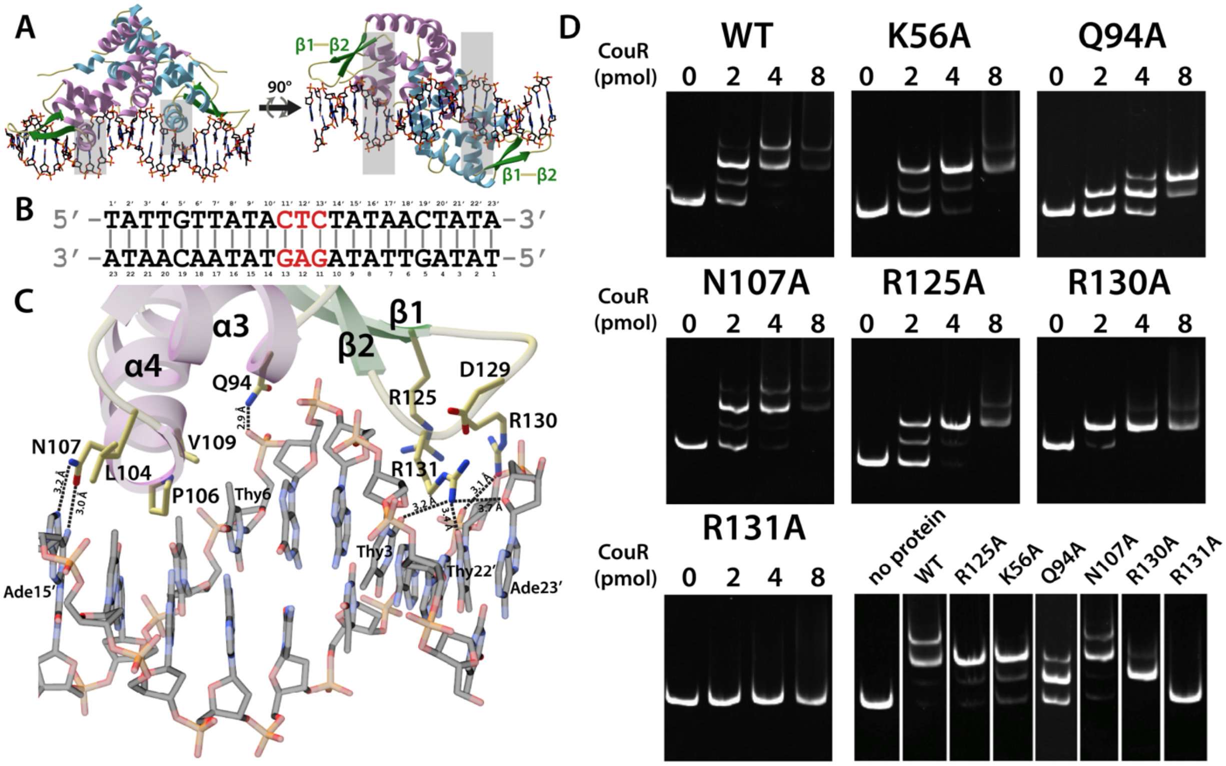


Fig. 1.2. CouR–DNA cocrystal structure and EMSA. *A*, overall structure of CouR bound to its *couR* operator DNA sequence represented in two 90° related orientations. *Gray rectangles* highlight the location of α4 in the major groove. Minor groove protein–DNA interactions are mediated by a looped region between strands β1-β2 (*khaki* and *green*). *B*, the 23-mer dsDNA duplex used in cocrystallization experiments with CouR. *C*, close-up view of the wHTH motif with specific contacts labeled. Asn-107 forms bidentate hydrogen bonds with $A^{15\prime}$, Gln-94 hydrogen-bonds with the phosphate of T^6, Arg-130 interacts electrostatically with the $T^{22\prime}$ phosphate, and Arg-131 forms multiple nucleobase interactions in the minor groove (*e.g.* T^3, $T^{22\prime}$, and $A^{23\prime}$; see Ref. 7 Fig. S4 for a more comprehensive 2D diagram of CouR–DNA contacts). *D*, EMSAs showing the binding of CouR variants to the *couA* promoter (see Ref. 7 Fig. S1 for SDS-PAGE analysis of CouR variants). Purified CouR proteins were incubated at different quantities (2, 4, or 8 pmol) with 0.3 pmol of a 317-bp DNA fragment containing the *couA* promoter (see "Experimental procedures"). The reaction mixtures were separated on polyacrylamide gels. The last panel summarizes the results obtained with 4 pmol of protein.

1.3.2 Cocrystal structure of CouR with operator DNA

The cocrystal structure of CouR bound to a 23-basepair duplex bearing the two half-sites was determined to 2.85 Å resolution from crystals containing 2 copies of the CouR dimer-DNA duplex complex in the asymmetric unit (ASU). The CouR homodimer is situated on the pseudo-palindromic duplex such that helix α4 from the wHTH motif is positioned into each of the two half-sites along consecutive major grooves that are roughly 34 Å apart. The DNA duplex is roughly B-form but is under-twisted by 1.3°, resulting in a shortening of the end-to-end distance by roughly

3.8 Å, relative to canonical B-DNA. Notably, these deviations arise from a widening of the major grooves, in order to accommodate binding of the CouR α4 recognition helix, and a corresponding slight narrowing of the minor groove (Fig. 1.2).

Each CouR monomer contacts each half of the inverted repeat sequence through interactions with the wHTH motif, with the α4 helix positioned in the major groove and strands β1-β2 of the wHTH making contacts with nucleotides on the outer periphery of the recognition sites. As in other structures of MarR family regulators in complex with their cognate operator sequences, the number of contacts between CouR and DNA bases of the inverted repeat is limited. The only base-specific interaction occurs at helix α4 between the carboxamide of Asn107 and N6 and N7 of Ade15. Van der Waals contacts with DNA bases of the recognition sequence are mediated via the wedging of Pro106 into the segment encompassing Thy6-Thy7-Ade8 of one strand, and the corresponding bases of the complementary strand. The Pro106-Asn107 sequence comprises the first two residues of the α4 helix, and the combination of extensive van der Waals contacts and sequence specific interactions mediated through these two residues establish the orientation of the recognition helix within the major groove of each half site.

The wing portion of the DNA-binding region, spanning residues Val122 through Leu135 (from strands β1-β2), mediates additional contacts with nucleotides at the periphery of the recognition sequence. Electron density for both strands, as well as for the intervening loop, is clear and continuous in the DNA cocrystal structure. Residues within the loop of this region are situated in the minor groove and are involved in extensive non-specific contacts with the DNA backbone. The phosphate backbone of the DNA duplex is within hydrogen bonding distance of several residues including the side chains of Arg123, Arg125, Arg130, Arg131, Ser132, and His133, as well as the main chain carbonyl of Ser128. The side chain of Arg131 is poised within hydrogen-bonding distance of Thy3, but this nucleotide is outside the recognition half-site, and is not part of the inverted repeat sequence of the naturally occurring *couA* operator. Arg131 interacts with the side chain carboxylate of Asp129 via a salt-bridge, and this Asp-X-Arg pairing, along with a similar interaction with thymine, is observed across various MarR family members.[20,23–26]

Outside of interactions with the wHTH motif, there are a few additional residues that make contact with the DNA backbone. Specifically, the side chain of Lys56 (equivalent to Arg10 in MepR and Tyr19 in OhrR) from helix α1 is situated directly above the phosphate of Thy14, which is located at the inner periphery of the inverted repeat. As a result, interactions between equivalent Lys56 residues from each monomer occur at either side of the minor groove. In the CouR cocrystal structure, Asn107 makes direct interactions with the Ade15 nucleobase, and equivalent residues in MepR (Thr63) and OhrR (Thr70) are also involved in base pair recognition. Lastly, Gln94 (equivalent to Gln50 in MepR) is located in helix α3 and is part of a framework of interactions with the phosphate of Thy6, which also engages Arg123 and His133 from each of the strands of the wing (for a detailed CouR-DNA interaction map, see Ref. 7 Fig. S4).

1.3.3 Electrophoretic mobility shift assay (EMSA)

In order to probe the CouR-DNA contacts in detail, we generated site-specific variants corresponding to a number of residues within the α4 helix, the β1-β2 wing, and two residues located outside of these elements, and probed the effects of each mutation on DNA binding using EMSAs (Fig. 1.2D). Four distinct bands are observed when the *couA* promoter fragment is presented with CouR, corresponding to four distinct oligomeric states, with the highest mobility band corresponding to free DNA. The presence of two bands observed at high protein concentrations (i.e. 8 pmol lanes for CouR wild-type and Asn107→Ala in Fig. 1.2D) is likely a result of quaternary interactions between CouR-*couA* promoter complexes in non-denaturing conditions. Mutation of the highly conserved Arg131 to Ala completely abolished DNA-binding activity, as previously observed in studies of the *S. aureus* multidrug resistance regulator MepR.[20] This result is reconciled by the crystal structure as numerous contacts are made between Arg131 and the minor groove of the DNA binding partner. The Arg130→Ala variant also shows a significant reduction in DNA binding, reflecting the multiple roles that this residue plays as both part of the Asp-X-Arg pairing common across MarR members, as well as facilitating interactions with the phosphate backbone. Notably, Arg130 is not conserved across other orthologs, and its provision in DNA interactions may be unique to CouR (Fig. 1.1D).

Other notable residues identified in the EMSA as important for binding include Lys56 and Arg125 which form electrostatic interactions with the DNA phosphate backbone, and Gln94 that interacts with the phosphate group of Thy6 (Fig. 1.2D and Ref. 7 S4). Alanine mutations at these residues resulted in observable decreases in the formation of higher molecular weight complexes. The modest effect of the Gln94→Ala variant suggests that while the DNA interaction mediated by this residue is important, it is not absolutely essential. Notably, the Asn107→Ala variant results in a mobility shift commensurate with wild-type CouR suggesting the base-pair interaction is not important for binding.

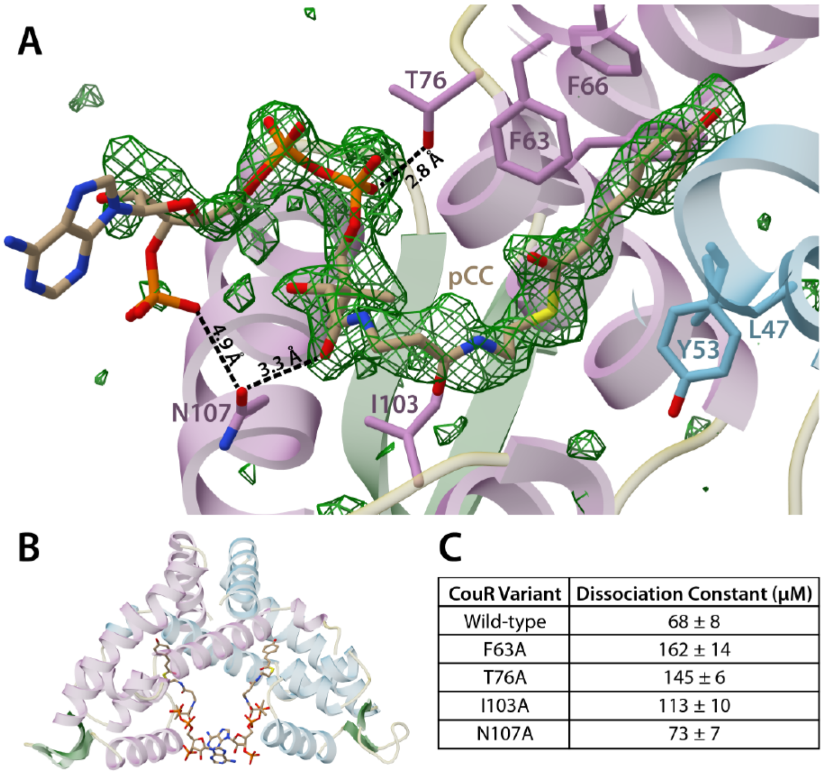
1.3.4 Cocrystal structure of CouR with ligand p-coumaroyl-CoA (pCC)

CouR was observed to bind pCC in a 1:1 stoichiometric ratio without disruption of the homodimeric interface also observed in the ligand-free CouR and CouR-DNA cocrystal structures (Fig. 1.3A,B). The ligand-free and pCC-bound structures are very similar demonstrating that binding of the ligand does not cause any appreciable changes in global or local structure of CouR (Fig. 1.4A). A major difference between the two structures is reflected in the orientation of the β-strands of the wHTH motif, which are partially disordered in the ligand-free structure but are well resolved in the pCC-bound structure. Moreover, the strands of each monomer in the pCC-bound structure are nearly superimposable, suggesting that ligand binding may restrict the flexibility of these strands.

Recognition of pCC appears to be driven primarily by hydrophobic interactions between the coumaroyl moiety of pCC and one of two equivalent clefts (one per monomer) in addition to some hydrogen bonding and hydrophobic interactions along the pantetheine group of pCC. The cleft exists adjacent to the dimeric interface surrounded by residues from helices α1, α2, and α5 (e.g. Phe63, Phe66, Pro77, and Phe80) of one monomer in addition to residues from helix α1 (e.g. Leu47 and Tyr53) of its neighboring monomer comprising the homodimer. CouR residue Asn107, which is involved in base pair recognition of Ade15 of the operator, also interacts with a backbone amide of pCC.

The location and secondary structural composition of this hydrophobic cleft is analogous to those observed in structures of MarR-family members bound to various phenolic acids and aldehyde ligands lacking coenzyme A.[10,15,21,27] Sufficient electron density was observed to model four pCCs bound to four CouR monomers in the crystallographic ASU. Additional interactions near the opening of the cleft include hydrophobic contacts between the side-chain of Ile103 and the geminal dimethyl group of pCC as well as hydrogen bonding between the side-chain of Thr76 and the β-phosphate of pCC (Fig. 1.3A).

The electron density corresponding to the adenosine nucleotide of pCC is ambiguous, indicative of disorder due to the minimal interactions of this moiety with the protein and its residence within a solvent pocket in the crystal lattice. Similar disorder has been observed for the adenosine group in other CoA bound structures.[28,29] These cross-subunit interactions with the pCC ligand result in an inward constriction of the second subunit upon binding of the ligand to the first subunit (Figs. 1.3A,B and 1.4A).



CouR Variant	Dissociation Constant (µM)
Wild-type	68 ± 8
F63A	162 ± 14
T76A	145 ± 6
I103A	113 ± 10
N107A	73 ± 7

Fig. 1.3: CouR–pCC cocrystal structure and summary of DSF binding constants. *A*, simulated annealing difference Fourier map (F_o-F_c) contoured at 3σ reveals an electron density corresponding to pCC (*dark khaki*) bound to CouR. Each ligand is embedded in a hydrophobic cleft comprised of residues from both monomers of the homodimer (colored either *plum* or *sky blue*). Additional residues from the monomer in *plum* engage the phosphopantetheine group of pCC (*i.e.* Thr-76, Ile-103, and Asn-107). The electron density corresponding to the adenine ring is spurious. *B*, overall structure of CouR bound to pCC retains the homodimeric oligomerization state and binds pCC symmetrically in a 1:1 fashion. *C*, summarized dissociation constants of site-directed CouR variants analyzed by DSF (see Ref. 7 Fig. S1 for SDS-PAGE analysis of CouR variants). Mutation of Phe-63 to Ala results in the greatest reduction of binding affinity (see "Experimental procedures" and Ref. 7 Fig. S2 for individual binding curves).

The overall structures of the CouR monomer in either DNA or pCC bound forms do not differ appreciably, with an RMSD of 0.85 Å over 891 atoms. However, superimposition of CouR dimers (RMSD of 2.35 Å over 2061 atoms) reveals notable changes that occur in the quaternary structure especially with regards to the wHTH motif that mediates DNA binding. Specifically, a superposition of dimers of the DNA bound CouR with that of the pCC bound structure reveals that the wHTH motif is shifted outwards by 5 Å in the former structure, which facilitates a disposition compatible with binding of the homodimer across consecutive major grooves of the operator (Fig. 1.4). Binding of DNA resulted in an outward shift of helix α4 that forms a portion of the pCC binding cavity, such that the wings of the wHTH motif are also suitably positioned for interactions with the duplex. The collapse of helix α4 necessary to form a viable pCC binding pocket would result in a CouR dimer that is not optimally aligned for binding across consecutive major grooves, suggesting a mechanism for ligand-mediated attenuation of operator binding. There are no relative rotational movements of the two CouR monomers between the DNA and pCC bound structures, such as that observed in comparisons of ligand and DNA bound structures of the MarR related Rv2887 from *M. tuberculosis*.[30]

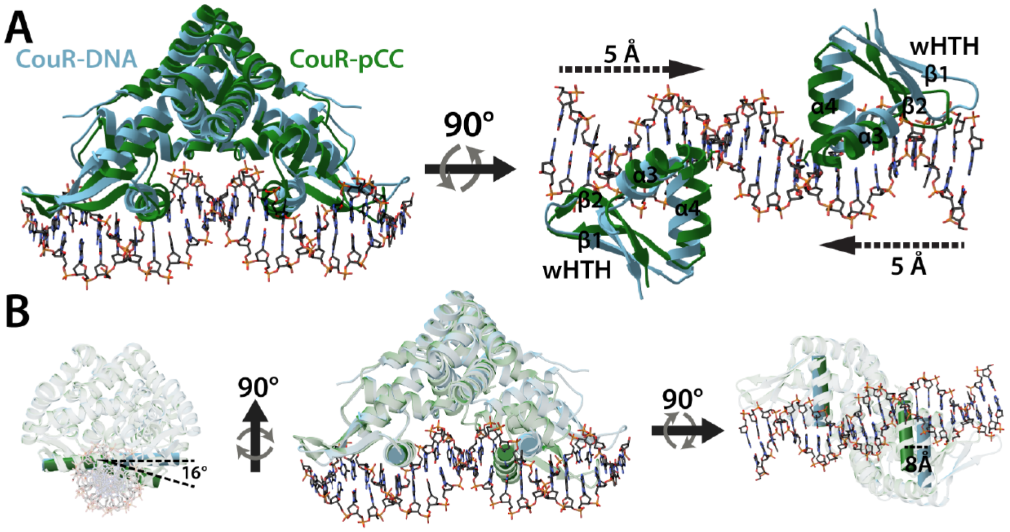


Fig. 1.4: Comparative analysis of CouR structures reveals the structural basis for ligand-induced disruption of DNA binding. *A*, superposition of DNA-bound (*sky blue*) and pCC-bound (*forest green*) CouR homodimers reveals differences in global structure. Two representations are related by a 90° rotation about the axis, depicted by a *black arrow*. The *top right* shows only the wHTH motifs for clarity and the symmetrical, ligand-induced 5-Å inward shift that compromises protein–DNA contacts in the major and minor grooves. *B*, superimposition of pCC-bound CouR onto one of the monomers in the CouR–DNA structure accentuates the motion of helix α4 at one-half of the symmetrical binding site. Three 90° related orientations of the superposition are displayed with measurements of relative angles (*bottom left*) and separation between helices α4 (*bottom right*).

1.3.5 Differential scanning fluorimetry (DSF)

On the basis of the observed crystallographic contacts between CouR and pCC, residues involved in hydrophobic contacts (Phe63 and Ile103) and hydrogen-bonding interactions (Thr76 and Asn107) were mutated to Ala. To probe the contributions of each residue to ligand binding, we carried out differential scanning fluorimetry (DSF) analyses for each variant. As the protein-ligand complex demonstrated an increased melting temperature (T_m), melting curves were monitored as a function of pCC concentration, enabling the measurement of dissociation constants (K_D) between pCC and each CouR variant.[31]

Analysis of wild-type CouR with pCC yielded a K_D value of 68 ± 8 μM. The Phe63→Ala and the Thr76→Ala variants showed the greatest effect on the K_D values for pCC binding. This result is consistent with crystallographic observations, which shows that Phe63 serves as a major contributor to the hydrophobic pocket that houses the ligand, and an Ala substitution at this site would likely compromise the integrity of the pCC binding pocket. Likewise, Thr76 is within hydrogen bonding distance of pCC, with a distance of 2.7 Å between the Oγ of this residue and the β-phosphate of pCC (Fig. 1.3).

Notably, the thermal denaturation midpoint (T_m) of the Thr76→Ala variant had decreased by 10 °C, relative to the wild-type. Consequently, this variant was tested to ensure it was properly folded and formed the requisite homodimeric assembly. Analytical size-exclusion chromatographic (SEC) analysis performed on CouR wild-type and the Thr76→Ala variant resulted in retention times expected for that of a 40 kDa homodimer (see Experimental procedures and Ref. 7 Fig. S3). The Ile103→Ala variant displays a slight binding impairment upon replacement of the sec-butyl group with a methyl group in the position found to associate with the geminal dimethyl group of pCC. No appreciable binding impairment was observed for the Asn107→Ala mutant suggesting that proximity of Asn107 to the amide carbonyl of pCC may not be critical for tight binding of the pCC ligand (Fig. 1.3).

1.4 DISCUSSION

Detailed biochemical studies on MarR family proteins are often hampered due to a lack of details regarding the physiological effector. Even though surrogate small molecules may function at the protein level, the relevance of binding by such effectors is sometimes not clear. Here, we present a detailed characterization of CouR, a repressor of a *p*-coumarate catabolic pathway for which the genetic, and microbiological analyses have been carried out in detail.[8,19] We complement the earlier studies using biochemical and structural biological approaches to elucidate details of a MarR member visualized in three different states: without any bound ligand, bound to the cognate operator element, and bound to the physiological effector molecule. Unlike other studies of MarR-family members, pCC the effector for CouR is not metabolically ubiquitous, and the extensive contacts made throughout the entirety of this structurally unique ligand affords an unbiased delineation of the structural details with a bound ligand. To further dissect the mechanism of ligand-induced dissociation, additional studies will be necessary to confirm whether pCC binds directly to the CouR-DNA complex or free CouR.

A prior effort also characterized a CouR-like pathway encoded in the soil bacterium *Rhodococcus josti* RHA1 (CouR-RJ).[23] That system is distinct in that not all of the genes proposed in the catabolic pathway are within the same operon. Based on the sequences of the different promoters, it was proposed that a single repressor may regulate transcription, even though the promoters are divergently transcribed. Notably, the promoter sequence identified in *R. josti* RHA1 (cATTGAnnnnnTCAATg) is entirely distinct from the *R. palustris* CGA009 CouR (CouR-RP) responsive promoter (GTTATAnnnTATAAC), and the two MarR-type regulators share only 36% sequence identity. The authors provide a molecular basis for DNA binding attenuation which invokes the sequestration of two Arg residues (numbered 36 and 38 in CouR-RJ; Fig. 1.1D) by pCC, however these residues are not present in CouR-RP, and similar contacts were not observed in the CouR structures presented here. Moreover, superimposing the CouR-RP DNA-bound structure with the orthologous CouR-RJ pCC-bound structure reveals that the separation of helices α4 are nearly equidistant, suggesting compatibility with major groove binding and a unique mechanism of repression (see Ref. 7 Fig. S5).

The crystal structure of CouR bound to the operator supports the idea of an indirect sequence readout, first suggested by Dolan et al. in the context of the SlyA-DNA structure.[25] This is a result of primarily non-base specific contacts mediating DNA recognition and is corroborated by existing MarR-family/DNA cocrystal structures in the Protein Data Bank.[20,24–26] The unexpected finding that many of the protein-DNA contacts are actually non-base-specific lends credence to the notion of an indirect sequence readout and suggests that newer *in silico* techniques may be necessary to predict three-dimensional topologies for correct assignment of protein-DNA binding partners. Indeed, computational techniques involving comparative and machine learning strategies have been utilized to predict protein-DNA binding interactions.[32] Predictive methods will hopefully benefit from the experimental structures in this work to understand similar MarR family regulators for which experimental data is lacking.

1.5 METHODS

1.5.1 Cloning, expression, and purification of CouR

The CouR gene was amplified by the polymerase chain reaction from purified *Rhodopseudomonas palustris* genomic DNA using the following set of primers: 5' GCACAGGATCCGTGACCTCGTCGAACAGGATC 3' (forward) and 5' TAACACTC GAGTCAGAACTCGCGGGCGATGG 3' (reverse). The PCR product was digested with BamHI and XhoI and ligated into a similarly digested pET28-MBP expression vector using T4 DNA ligase (NEB). The presence of CouR was tested by restriction digest analysis and verified by di-deoxy sequencing (ACGT Inc.) The resulting construct encodes a fused maltose-binding protein (MBP) protein followed by a hexahistidine (His_6) affinity tag preceding the CouR reading frame.

The recombinant plasmid described above was used to transform *Escherichia coli* (Rosetta) for overexpression of MBP-CouR. *E. coli* cultures were grown with shaking at 250 rpm in LB medium supplemented with kanamycin (50 μg/mL) and chloramphenicol (25 μg/mL) until an OD_{600} of 0.5-0.6 was reached. The cultures were induced with 0.5 mM IPTG after cooling on ice for 20 min and grown for an additional 16 h at 18 °C. Cells were harvested by centrifugation at 3500 rpm and resuspended in buffer A (500 mM NaCl, 25 mM Tris-HCl pH 8.0, and 10% glycerol) before lysing with a French-press at 8,000-10,000 psi for 4 cycles. The lysate was cleared by centrifugation at 14,000 rpm for 1 h at 4 °C and the supernatant was loaded on a 5 mL HisTrap nickel-nitrilotriacetic acid (Ni-NTA) affinity column (GE Healthcare) equilibrated with buffer B (1 M NaCl, 25 mM Tris-HCl pH 8.0, and 30 mM imidazole). The column was washed with 40 mL of buffer B before eluting with a linear gradient from 0-100% buffer C (1 M NaCl, 25 mM Tris-HCl pH 8.0, and 250 mM imidazole) over 20 min at 2 mL/min. CouR was then subjected to TEV protease proteolysis at 4 °C to remove the MBP-tag while being dialyzed into 250 mM NaCl, 20 mM Tris pH 7.5, and 1 mM DTT. The reaction ran to near completion after 24 h (as monitored by SDS-PAGE) before dialyzing into 250 mM NaCl, and 20 mM Tris-HCl pH 7.5 to remove DTT before subtractive nickel purification. Reaction contents were loaded onto a Ni-NTA column equilibrated with buffer A followed by a 30 mL wash with buffer A. Proteins were eluted in 4 mL increments using the following stepwise gradient of increasing buffer B: 5%, 10%, 20%, 30%, and 100%. This was followed by additional stepwise elution steps of increasing buffer C: 50% and 100%. Fractions containing CouR were dialyzed into 100 mM NaCl, 20 mM Tris-HCl pH 7.5 at 4 °C before loading onto a 5 mL HiTrap SP HP column (GE Healthcare) equilibrated with buffer D (100 mM NaCl, 50 mM bicine pH 7.5). The column was washed with 20 mL of buffer D before eluting with a linear gradient of increasing buffer E (1 M NaCl, 50 mM bicine pH 7.5) for 35 min at a flow-rate of 1.5 mL/min. CouR-containing fractions indicated ≥95% purity by SDS-PAGE analysis (see Ref. 7 Fig. S1).

1.5.2 Differential scanning fluorimetry (DSF) binding assay

CouR wild-type and mutant proteins were purified as described above followed by gel-filtration on a 120 mL Superdex 75 column (GE Healthcare) in 50 mM KCl, 0.5 mM 2-mercaptoethanol, 20 mM HEPES pH 7.0. The DSF assay was performed on a StepOnePlus RT-PCR instrument (Applied Biosystems) using a 96-well plate format. Wild-type and mutant proteins were subjected to increasing concentrations of pCC from 0-10 mM while maintaining a constant protein concentration of 68 μM and 6X SYPRO orange dye (prepared from a 5,000X stock in DMSO). Samples were allowed to equilibrate for 30 minutes prior to initiating the melt curve. The melt curve program initiated with a 2 minute 25 °C hold step followed by a ramp up to 99 °C at 1.7 °C/min (step and hold) and a final hold for 2 minutes at 99 °C. First derivatives of the melting curves were used to determine the T_m for each sample well. Plots were fit to a dose response curve using OriginPro 2015 and K_D was taken to be equal to the substrate concentration at which melting temperature achieved 50% of the total change.

1.5.3 Analytical size-exclusion chromatography (SEC)

Analytical SEC was performed using a 40 mL Superdex 200 10/300 GL column (GE Healthcare) equilibrated with 0.1 M NaCl, 20 mM Tris pH 7.0. A similar running buffer was used to elute protein standards prepared from commercially available dry lyophilized powders (Sigma-Aldrich Co.) at 1 mL/min. CouR variants were eluted similarly (see Ref. 7 Fig. S3).

1.5.4 Crystallization and structure solution of CouR-couR operator complex

Single-stranded DNA oligomers were ordered from Integrated DNA Technologies (IDT) that were made up to 10 mM by solvating in 20 mM Tris pH 7.5, 50 mM NaCl. The complementary ssDNA solutions were mixed 1:1 before using a thermocycler to heat at 95 °C for 5 minutes followed by a decrease to 25 °C at a rate of 1 °C per minute. A series of dsDNA sequences ranging from 16-29 bp and containing either blunt-end or single-nucleotide overhangs were screened for cocrystallization with CouR before arriving at the optimally diffracting 23 nucleotide sequence (23-mer) used in the final model. Crystals of CouR bound to the 23-mer were obtained by mixing 15 mg/mL CouR with 1.1 stoichiometric amounts of 23-mer and incubating on ice for 20 minutes. This solution was mixed with an equal volume of reservoir solution containing 0.1 M Tris pH 8.5, 0.1 M ammonium phosphate, and polyethylene glycol 6,000 to generate 2 µL hanging drops. Thin plate-shaped crystals grew at 16 °C to maximum size (about 100 x 100 µm) in less than a week.

Crystals were cryoprotected in crystallization buffer supplemented with 25% glycerol and vitrified by direct immersion into liquid nitrogen prior to data collection. Structure determination was carried out using single-wavelength anomalous dispersion data collected from crystals of selenomethionine derivatized CouR bound to DNA, using Phenix AutoSol with a Bayesian estimate of the map quality.[33] Data were collected at the Advanced Photon Source, Argonne National Lab using the Life-Science Collaborative Access Team (LS-CAT) 21-ID-F, 21-ID-G, and 21-ID-D beamlines. After determining the phases and initial model from Phenix AutoSol, each structure was rebuilt using Phenix AutoBuild with multiple rounds of manual intervention.[34] This new model was further refined using CCP4 refmac5 in combination with additional manual rebuilding in *COOT*.[35,36]

1.5.5 Crystallization and structure solution of CouR-pCC complex

CouR was concentrated to 24 mg/mL, combined with 3 mM pCC, and crystallized at 9 °C in similar 2 µL hanging drops containing a reservoir solution of 0.17 M Li_2SO_4, 23.5% polyethylene glycol 4,000, 1 mM DTT, 0.085 M Tris-HCl pH 8.5, and 15% glycerol. Rod-shaped crystals grew to maximum size in less than a week to about 200 × 50 µm at the largest face. Prior to data collection, crystals were directly flash-frozen in liquid nitrogen. Data were collected at the

Advanced Photon Source, Argonne National Lab using the Life-Science Collaborative Access Team (LS-CAT) 21-ID-G beamline. The structure was solved by molecular replacement using the protein coordinates from the CouR-*couR* operator structure. Molecular replacement derived phases were used to build an initial model in Phenix AutoBuild, followed by similar refinement procedures as described for the CouR-*couR* operator complex. Prior to fitting pCC into electron density difference maps, restraint parameters and geometry optimizations were produced by Phenix eLBOW.[37] Water molecules were incorporated in the CouR-pCC structure using Phenix Refine.[38]

1.5.6 Crystallization and structure solution of apo CouR

CouR was concentrated to 15 mg/mL and crystallized at 9 °C in 1 μL sitting drops by mixing 1:1 (*v/v*) protein solution and reservoir solution containing 60% (*v/*v) Tacsimate pH 7.0. Diamond shaped crystals grew to maximum size in less than a week. The structure was solved by molecular replacement in a similar manner to that described for the CouR-pCC cocrystal structure. Additional model building and refinement was carried out as described above.

CHAPTER 2: STRUCTURAL STUDIES OF A TRYPTOPHAN 5-HALOGENASE IN NAI-107 BIOSYNTHESIS[b]

2.1 ABSTRACT

Lantibiotics are ribosomally synthesized and post-translationally modified antimicrobial peptides containing thioether rings. In addition to these cross-links, the clinical candidate lantibiotic NAI-107 also possesses a C-terminal *S*-[(*Z*)-2-aminovinyl]-d-cysteine (AviCys) and a unique 5-chloro-l-tryptophan (ClTrp) moiety linked to its potent bioactivity. Bioinformatic and genetic analyses on the NAI-107 biosynthetic gene cluster identified *mibH* and *mibD* as genes encoding flavoenzymes responsible for the formation of ClTrp and AviCys, respectively. The biochemical basis for the installation of these modifications on NAI-107 and the substrate specificity of either enzyme is currently unknown. Using a combination of mass spectrometry, liquid chromatography, and bioinformatic analyses, we demonstrate that MibD is an FAD-dependent Cys decarboxylase and that MibH is an $FADH_2$-dependent Trp halogenase. Most $FADH_2$-dependent Trp halogenases halogenate free Trp, but MibH was only active when Trp was embedded within its cognate peptide substrate deschloro NAI-107. Structural comparison of the 1.88 Å resolution crystal structure of MibH with other flavin-dependent Trp halogenases revealed that subtle amino acid differences within the MibH substrate binding site generates a solvent exposed crevice presumably involved in determining the substrate specificity of this unusual peptide halogenase.

---

[b] This research was originally published in *ACS Chemical Biology*. Part of this chapter was directly obtained, with permission, from:

Ortega MA, Cogan DP, Mukherjee S, Garg N, Li B, Thibodeaux GN, Maffioli SI, Donadio S, Sosio M, Escano J, Smith L, Nair SK, van der Donk WA. Two Flavoenzymes Catalyze the Post-Translational Generation of 5-Chlorotryptophan and 2-Aminovinyl-Cysteine during NAI-107 Biosynthesis. *ACS Chem Biol.* 2017 Feb 17;12(2):548-557. Copyright © 2017, American Chemical Society.

2.2 INTRODUCTION

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a growing class of natural products containing a wide array of unique pharmacophores.[39] These moieties are introduced on a ribosomally synthesized precursor peptide by dedicated biosynthetic enzymes and bestow upon these molecules their biological activities. As such, characterization of RiPP biosynthetic enzymes provides an avenue to harness the potential of these catalysts as tools for the development of chemoenzymatic methodologies to generate novel therapeutic peptides.

Within RiPPs, lanthipeptides (lanthionine-containing peptides) encompass one of the most abundant and well-studied subfamilies to date.[39] The key structural features defining lanthipeptides are thioether rings formed by the bis-amino acids lanthionine (Lan) and/or methyllanthionine (MeLan; Fig. 2.1a). These cross-links are introduced enzymatically *via*dehydration of select Ser/Thr residues to 2,3-dehydroalanine (Dha) and (*Z*)-2,3-dehydrobutyrine (Dhb), respectively, within the C-terminal core region of a precursor peptide (LanA), followed by subsequent Michael-type additions of Cys thiols onto the newly generated dehydroamino acids (Fig. 2.1a).[40,41] The enzymes responsible for catalyzing both modifications require an N-terminal leader peptide sequence within the precursor peptide for efficient processing.[42,43] After thioether ring formation, one or more proteases remove the N-terminal leader sequence from the precursor peptide releasing the mature natural product. Lanthipeptides exhibiting antimicrobial activity are known as lantibiotics (lanthionine-containing antibiotics).

In addition to these common post-translational modifications, some lanthipeptides contain additional modified residues installed by tailoring enzymes. These modifications in many cases increase the stability and biological activity of the peptides. For instance, epilancin 15X contains an N-terminal lactate that serves to protect it against the action of aminopeptidases.[44] Similarly, the lantibiotic NAI-107 (also known as microbisporicin) contains a C-terminal AviCys predicted

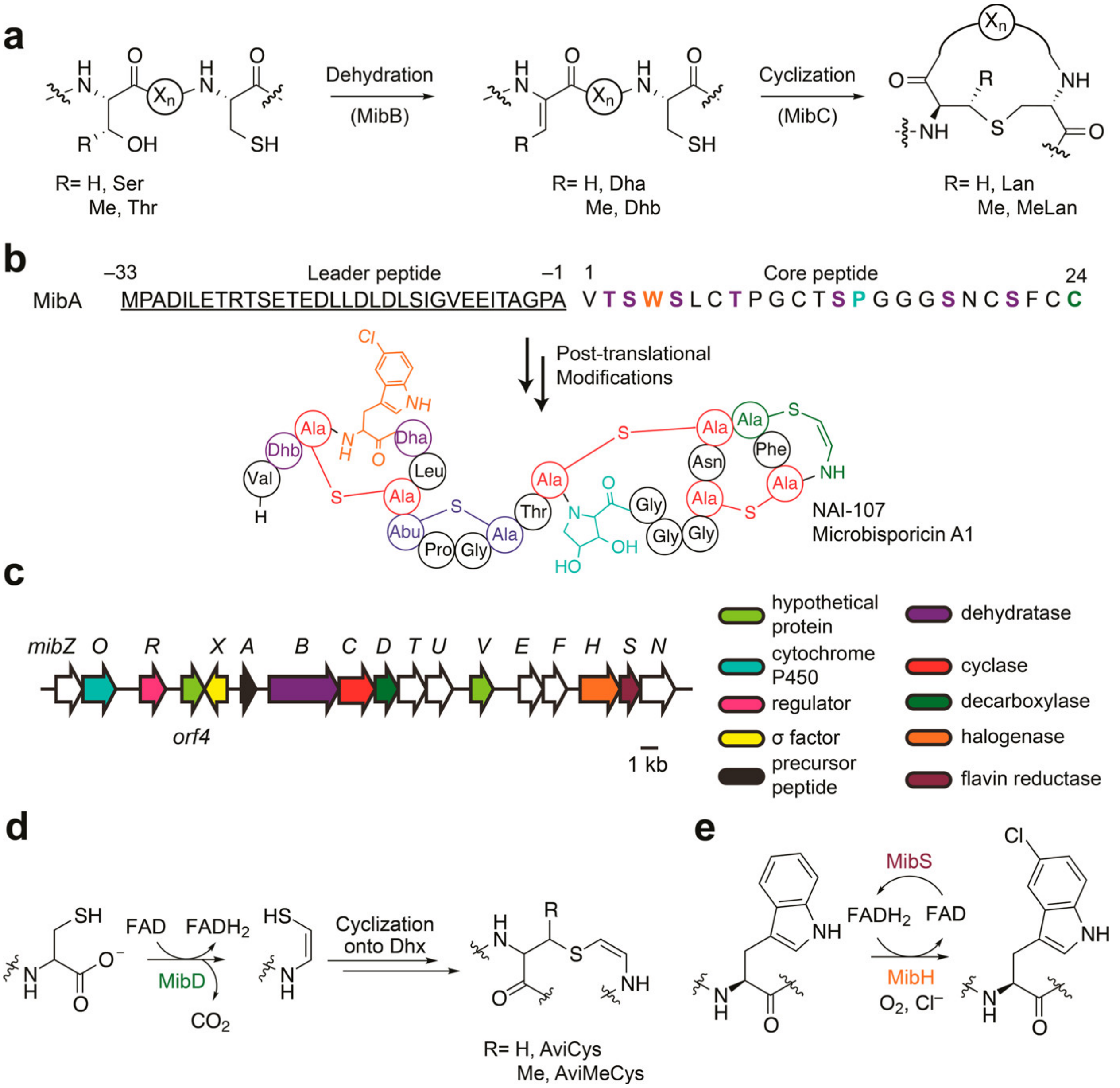


Fig. 2.1: Overview of NAI-107 biosynthesis. (a) Scheme for thioether ring formation during lanthipeptide biosynthesis. The enzymes that catalyze each step during NAI-107 biosynthesis are shown in parentheses. (Xn) connecting peptide. (b) Structure of NAI-107. The amino acid sequence of the precursor peptide MibA is shown. Negative numbers represent the position of amino acids within the leader peptide (underlined) with respect to the first amino acid in the core region. Dehydroamino acids, lanthionine, and methyllanthionine are shown in purple, red, and blue, respectively. The aminovinyl cysteine and 5-chlorotryptophan are colored green and orange. NAI-107 also contains a 3,4-dihydroxyproline moiety shown in cyan. (c) NAI-107 biosynthetic gene cluster. (d,e) Proposed reactions catalyzed by the enzymes (d) MibD and (e) MibH/MibS. Dhx, dehydroalanine or dehydrobutyrine; AviMeCys, S-[(Z)-2-aminovinyl]-(3S)-3-methyl-d-cysteine (Figure credit: Manuel Ortega and Wilfred van der Donk)

to arise from the oxidative decarboxylation of a Cys residue within the NAI-107 precursor peptide (Fig. 2.1b).[45] Presumably, this modification protects the peptide from carboxypeptidases and removes a negative charge, which has proven beneficial for bioactivity in semisynthetic lanthipeptide analogs.[46] In addition, NAI-107 contains a 3,4-dihydroxyproline (diHPro) and a unique ClTrp responsible for increasing its bioactivity 2-fold compared to its deschloro counterpart (Fig. 2.1b).[45,47] The benefits these additional modifications confer to the stability and bioactivity of lantibiotics prompted biochemical characterization of the tailoring enzymes responsible for installing AviCys and ClTrp during NAI-107 biosynthesis.

NAI-107 is a produced by the actinomycete *Microbispora* sp. 107891.[45,48] NAI-107 acts as a peptidoglycan biosynthesis inhibitor by sequestering the cell wall precursor lipid II and is currently in preclinical trials for the treatment of multidrug resistant Gram-positive bacterial infections.[49–51] NAI-107 biosynthesis commences with the translation of the precursor peptide MibA (Fig. 2.1b).[48,52] Then, MibB dehydrates seven Ser/Thr residues within the core region of MibA in a glutamyl-tRNA^{Glu}-dependent manner, and the lanthipeptide cyclase MibC catalyzes the nucleophilic addition of Cys residues onto five dehydroamino acids.[53] The biosynthetic details regarding AviCys, ClTrp, and diHPro installation during NAI-107 biosynthesis are not known.

Previous bioinformatic and genetic analyses of the NAI-107 biosynthetic gene cluster identified two genes coding for enzymes predicted to be responsible for the decarboxylation and halogenation (Fig. 2.1c).[48] One of the open reading frames (ORFs), *mibD*, encodes an enzyme that has sequence similarity to lanthipeptide decarboxylases (LanD).[54,55] These enzymes together with (*R*)-4′-phospho-*N*-pantothenoylcysteine (PPC) decarboxylases belong to the homo-oligomeric flavin-containing Cys decarboxylase (HFCD) superfamily that employs a flavin cofactor to catalyze the oxidative decarboxylation of C-terminal Cys-containing peptides (Fig. 2.1d).[54–56] Decarboxylation proceeds *via* oxidation of a Cys thiol to a thioaldehyde that subsequently serves to delocalize the negative charge developed at the former Cα upon decarboxylation.[57–61] Nucleophilic addition of the C-terminal aminoenethiolate to a nearby dehydroamino acid within the core peptide yields the AviCys moiety (Fig. 2.1d).[61] At present, it is unclear whether this latter reaction is catalyzed by lanthipeptide decarboxylases or lanthipeptide cyclases or if it proceeds nonenzymatically.

A second ORF, *mibH*, shows homology to flavin-dependent Trp halogenases.[48] Members of this enzyme class (Fig. 2.1e) use reduced flavin to activate molecular oxygen.[62] A halide ion is believed to perform a nucleophilic attack onto the distal oxygen atom of a 4a-flavin hydroperoxide forming a hypohalite that subsequently transfers the halogen to a nearby conserved Lys residue (Lys102 in MibH). This chloramine intermediate is then believed to halogenate Trp.[63–65] Within the NAI-107 biosynthetic gene cluster, *mibS* encodes a protein that shows sequence similarity to flavin reductases and is presumably responsible for providing reduced flavin to MibH (Fig. 2.1e).[48]

In this study, we reconstituted the activity of MibD, MibS, and MibH *in vitro*. Decarboxylation assays revealed MibD to be specific for the C-terminal sequence in the precursor peptide MibA and to not require the N-terminal leader peptide for activity. MibH and MibS form an FAD-dependent halogenase system. Interestingly, halogenation assays with a panel of different substrate analogs revealed MibH to be highly specific for the modified MibA core peptide carrying thioether rings suggesting halogenation to occur in one of the last steps of NAI-107 biosynthesis. Last, the 1.88-Å-resolution crystal structure of holo MibH provides a structural basis for its unusual substrate specificity.

2.3 RESULTS

2.3.1 Structural Characterization of MibH

To determine the structural basis of such unusual peptide specificity, the crystal structure of MibH bound to FAD was determined to 1.85-Å resolution (Fig. 2.2a and Ref. 66 Supporting Information Table S2). MibH crystallized with four molecules in the asymmetric unit, and clear density corresponding to FAD and chloride ions was observed (Fig. 2.2b) in two out of four of

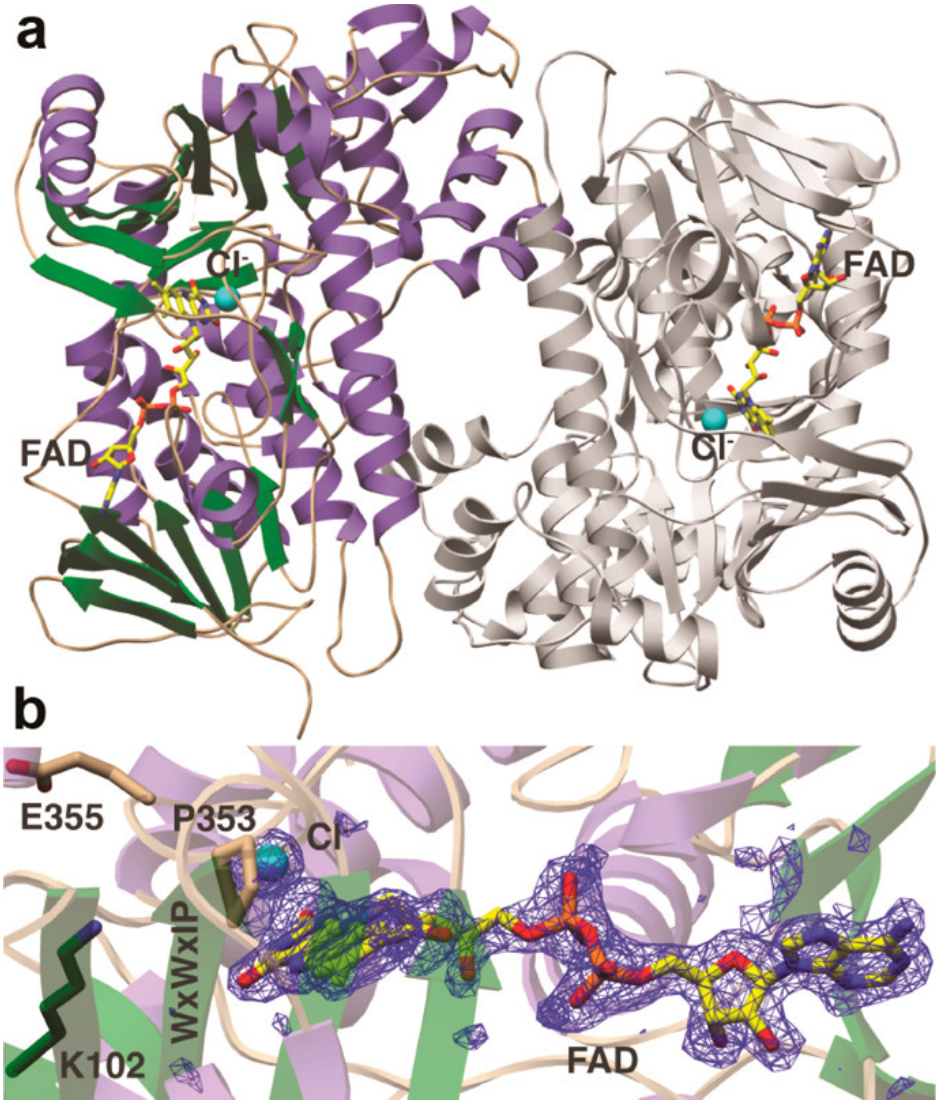


Fig. 2.2: Crystal structure of the lantibiotic halogenase MibH. (a) Overall cartoon representation of MibH. α helices are shown in purple. β sheets are shown in green. Flexible loops are shown in wheat, and the second protomer is shown in gray. FAD and the Cl– ion are shown as sticks and as a cyan sphere, respectively. (b) Simulated annealing omit difference Fourier map (Fo – Fc) contoured to 2.5σ of FAD and Cl–. The conserved Lys/Glu pair involved in chlorination is shown as sticks. The conserved WXWXIP motif is labeled in one amino acid code. (Figure credit: Satish Nair)

these molecules, whereas spurious electron density was observed for the other two molecules. The overall structure of the MibH monomer consists of a large single domain consisting of a rectangular arrangement of several α-helices surrounding two β-sheets, with a triangular helical protuberance appended to the base of this rectangle. The FAD cofactor is positioned in a solvent exposed groove, roughly parallel to the main body of each monomer, with the adenosine moiety positioned near one set of β-strands and the isoalloxazine ring situated at the other set of β-strands.

MibH contains two signatures motifs that are conserved among other $FADH_2$-dependent halogenases. The WxWxIP (Trp283-Pro288) motif, whose role in halogenation remains elusive, [67] and the GxGxxG (Gly36-Gly41) flavin consensus sequence that is located in the vicinity of the FAD. A Lys/Glu pair that is catalytically relevant in other $FADH_2$-dependent halogenases is also present in MibH (Lys102 and Glu355). In orthologous enzymes, the substrate Trp is located more than 10 Å away from the flavin and chlorination of the equivalent Lys facilitates halogenation *via* a covalent Lys chloramine intermediate.[65] The conserved Glu may either facilitate proton abstraction from the Trp substrate following halogenation[62] or enhance the electrophilicity of the chloramine. Electron density corresponding to a chloride ion is located at a pocket at the *Re* face of the isoalloxazine, where it interacts with the backbone amides of Ser357 and Gly358. Pro353 forms the opposing edge of the binding pocket (Fig. 2.2b). As in other $FADH_2$-dependent halogenases, the size of the binding pocket likely precludes the binding of halides such as iodide, but not bromide, consistent with the ability to purify brominated NAI-107 variants when the natural producer is supplemented with bromide sources.[68] Comparison of MibH with other enzymes in the protein data bank (PDB) *via* the Dali server reveals a conserved global architecture shared among other Trp halogenases. The closest structural relatives to MibH include the 5-Trp halogenase PyrH involved in pyrroindomycin biosynthesis,[69] and the 7-Trp halogenases PrnA and RebH from pyrrolnitrin and rebeccamycin biosynthesis, respectively,[62,63] all of which utilize free Trp as a substrate but exhibit different regioselectivity. Visual inspection of the proposed MibH active site based on conservation of important catalytic residues and the spatial arrangement of FAD and chloride reveals that MibH has a noticeable larger accessible surface area in comparison to free-tryptophan utilizing halogenases, which likely serves to accommodate its larger peptidic substrate (Fig. 2.3a). Interestingly, the similar structural topology of MibH to other $FADH_2$-dependent Trp halogenases suggests this larger crevice did not arise due to structural deletions. Instead, subtle amino acid variations within MibH are responsible for forming the cavity, as revealed by pairwise structural alignments of MibH with different members of this enzyme family.

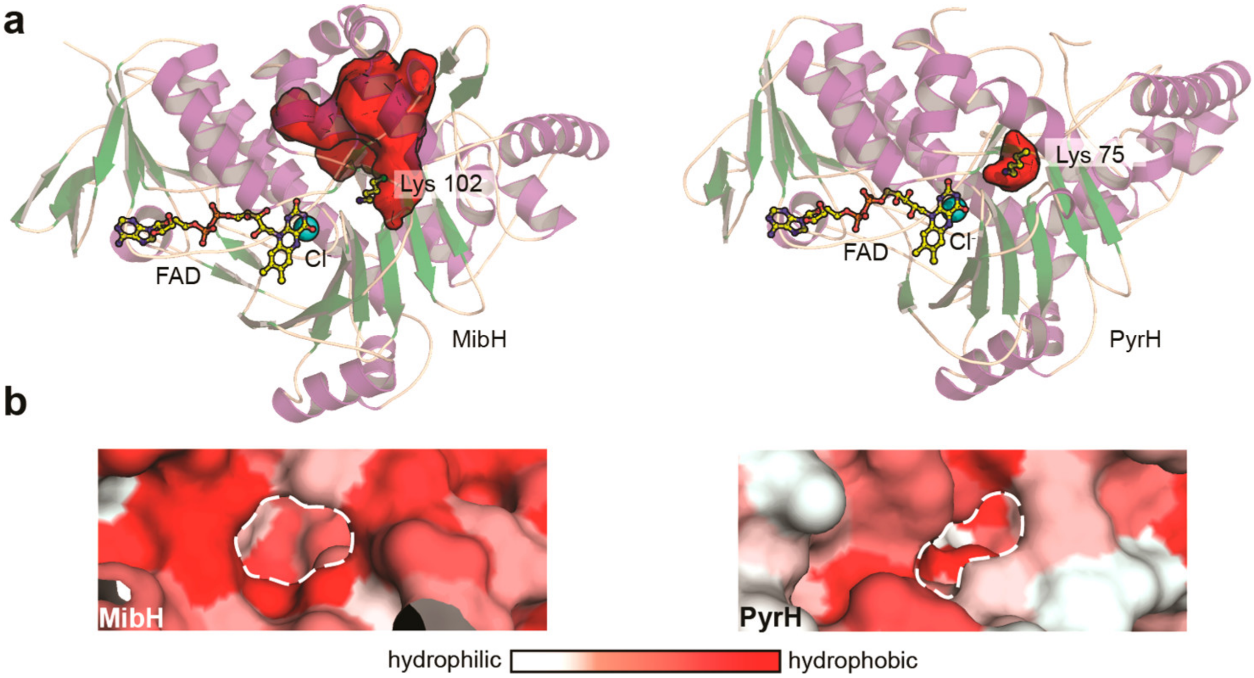


Fig. 2.3: Insights into the substrate specificity of the 5-tryptophan halogenase MibH. (a) Cartoon representation of the overall structure of (left) MibH and (right) PyrH(49) (PDB 2WET) with their respective substrate binding clefts shown as red surfaces. Color scheme is the same as in Figure 2.2. (b) Surface representation of the proposed Trp binding site (dashed white line) in (left) MibH and (right) PyrH colored by hydrophobicity (Figure credit: Manuel Ortega and Wilfred van der Donk).

In the absence of a substrate-bound structure, we propose this region to serve as the substrate binding cleft in MibH. The CASTp server was used to identify residues delineating the proposed peptide binding site,[70] allowing the calculation of its solvent accessible surface area in PyMOL, which reveals this region to be 718.5 $Å^2$ larger than that of other members of this enzyme family. In addition, the greater relative hydrophobicity of the MibH binding cleft also supports its peptide substrate preference over free tryptophan (Fig. 2.3b). We next performed a docking analysis of MibH and its substrate deschloro NAI-107. Docking poses were obtained where the C-

5 indole carbon is in closest proximity to the catalytic Lys102 with an 8.5-Å distance to its ε-nitrogen (Fig. 2.4). For catalysis to occur, we envision this distance to be reduced upon substrate binding.

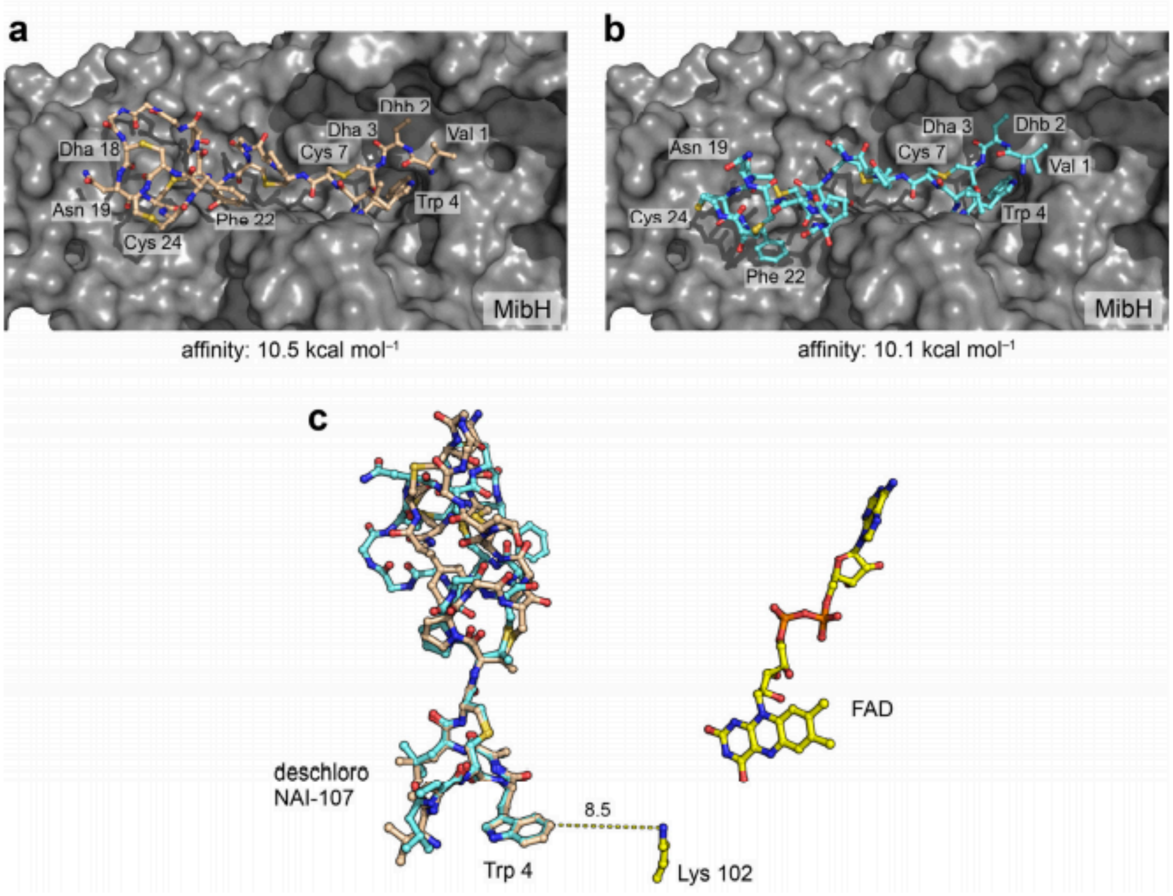


Fig. 2.4: Autodock/Vina models of deschloro NAI-107 bound to MibH. (a, b) Two different conformations for deschloro NAI-107 were obtained differing mainly at the spatial arrangement of its C-terminus. (c) Distance measurement of the resulting deschloro NAI-107 bound models to the catalytic conserved Lys102 in MibH involved in chlorination (Figure credit: Satish Nair).

2.4 METHODS

2.4.1 MibH Structure Determination

Following purification of MibH by nickel affinity chromatography as described in the SI, the protein was desalted by size exclusion chromatography using a HiLoad 16/600 Superdex 200 column (flow rate of 1 mL min^{-1}) connected to an ÄKTA purifier at 4 °C (buffer of 20 mM HEPES at pH 8.0 and 150 mM KCl). Eluted protein was then concentrated to 22 mg mL^{-1} in the presence of 2 mM FAD prior to crystallization. Crystals were obtained within 24 h in hanging drops containing 12.5% MPD, 12.5% PEG 3.35K, 0.1 M MES/imidazole at pH 6.5, and 0.09 M NPS (1:1:1 $NaNO_3$, Na_2HPO4, $(NH_4)_2SO_4$). Prior to data collection, crystals were transferred into the crystallization solution supplemented with 20% MPD and vitrified in liquid nitrogen. Data were collected at the Advanced Photon Source, Argonne National Lab using the Life-Science Collaborative Access Team (LS-CAT) 21-ID-G beamline. Raw data were integrated and scaled using the autoPROC software.[71] The structure was determined by molecular replacement using the PHASER software[72,73] and a modified model of PyrH (PDB: 2WET) with all nonidentical residues truncated to Ala. The initial model was further improved using Buccaneer[74] and interspersed with rounds of manual building using COOT.[36,75] Interim models were routinely adjusted through rounds of crystallographic refinement using Refmac5.[35,76] Refinement converged to the final model for which statistics are presented in Ref. 66 Supporting Information Table S2.

2.4.2 MibH-deschloro NAI-107 docking models

Automated docking of deschloro NAI-107 onto the MibH coordinates were carried out using the Autodock/Vina PyMOL plugin.[77] Ligand coordinates were obtained from the solution NMR structure of NAI-1079 and manually modified to generate deschloro NAI-107 with appropriate bond lengths. Details regarding two relevant poses are graphically represented in Figure 2.4.

CHAPTER 3: STRUCTURAL INSIGHT INTO PYRIDINE FORMATION IN THIOPEPTIDE BIOSYNTHESIS[c]

3.1 ABSTRACT

The [4+2] cycloaddition reaction is an enabling transformation in modern synthetic organic chemistry but there are only limited examples of dedicated natural enzymes that can catalyze this transformation. Thiopeptides (or more formally thiazolyl peptides) are a class of thiazole-containing, highly modified, macrocyclic secondary metabolites made from ribosomally synthesized precursor peptides. The characteristic feature of these natural products is a six-membered nitrogenous heterocycle that is assembled via a formal [4+2] cycloaddition between two dehydroalanine (Dha) residues. This heteroannulation is entirely contingent on enzyme activity, though the mechanism of the requisite pyridine/dehydropiperidine synthase remains to be elucidated. The unusual *aza*-cylic product is distinct from the more common carbocyclic products of synthetic and biosynthetic [4+2] cycloaddition reactions. To elucidate the mechanism of cycloaddition, we have determined atomic resolution structures of the pyridine synthases involved in the biosynthesis of the thiopeptides thiomuracin (TbtD) and GE2270A (PbtD), in complex with substrates and product analogs. Structure-guided biochemical, mutational, computational, and binding studies elucidate active site features that explain how orthologs can generate rigid macrocyclic scaffolds of different sizes. Notably, the pyridine synthases show structural similarity to the elimination domain of lanthipeptide dehydratases, wherein insertions of secondary structural

---

[c] This research was originally published in the *Proceedings of the National Academy of Sciences of the United States of America*. Part of this chapter was directly obtained, with permission, from:

Cogan DP, Hudson GA, Zhang Z, Pogorelov TV, van der Donk WA, Mitchell DA, Nair SK. Structural Insights into Enzymatic [4+2] Aza-Cycloaddition in Thiopeptide Antibiotic Biosynthesis. *Proc Natl Acad Sci U S A.* 2017 Dec 5;114(49):12928-12933. Copyright © 2017, National Academy of Sciences.

elements result in the formation of a distinct active site that catalyzes new chemistry. Comparative analysis identifies other catalysts that contain a shared core protein fold but whose active sites are located in entirely different regions, illustrating a principle predicted from efforts in *de novo* protein design.

3.2 INTRODUCTION

Thiopeptides are a class of ribosomally synthesized and post-translationally modified peptides (RiPPs)[39] that display a range of biological activities.[78–80] The identification of thiopeptides with impressive antibiotic activity towards drug-resistant Gram-positive bacteria, such as methicillin-resistant *Staphylococcus aureus* (MRSA), as well as others with antiproliferative and immunosuppressive activities, has spurred a resurgence in their research.[78] The antibacterial activity of thiopeptides is exerted through the inhibition of ribosomal protein synthesis. Curiously, the biological target of thiopeptides correlates with the size of the macrocyclic ring that embeds the heterocyclic core; thiostrepton and thiocillins (26-membered rings) inhibit the 50S ribosome, whereas GE2270 and thiomuracin (29-membered rings) (Fig. 3.1A) bind to elongation factor thermal unstable (EF-Tu).[81,82] To date, dozens of different types of thiopeptides have been identified, all of which share the class-defining 6-membered nitrogenous heterocycle that is at the center of a framework (pink circles) decorated with multiple thiazol(in)es (purple circles), and dehydro amino acids (green circles). The precursor peptide is of ribosomal origin and consists of a leader sequence that directs enzyme recruitment and a core sequence where the modifications occur.[42,43,83] Access to many thiopeptide variants was facilitated owing to the precursor peptide being a direct gene product.[80,84,85]

Reconstitution studies using purified enzymes have established both the nature and the timing of the modifications installed during the biosynthesis of thiomuracin from *Thermobispora*

bispora.[86] Specifically, TbtG, a YcaO-type cyclodehydratase[87] converts Cys residues in the core sequence of the TbtA precursor peptide to thiazolines, which are then dehydrogenated to the corresponding thiazoles. A split class I (LanB) lanthipeptide dehydratase[88] modifies the resultant

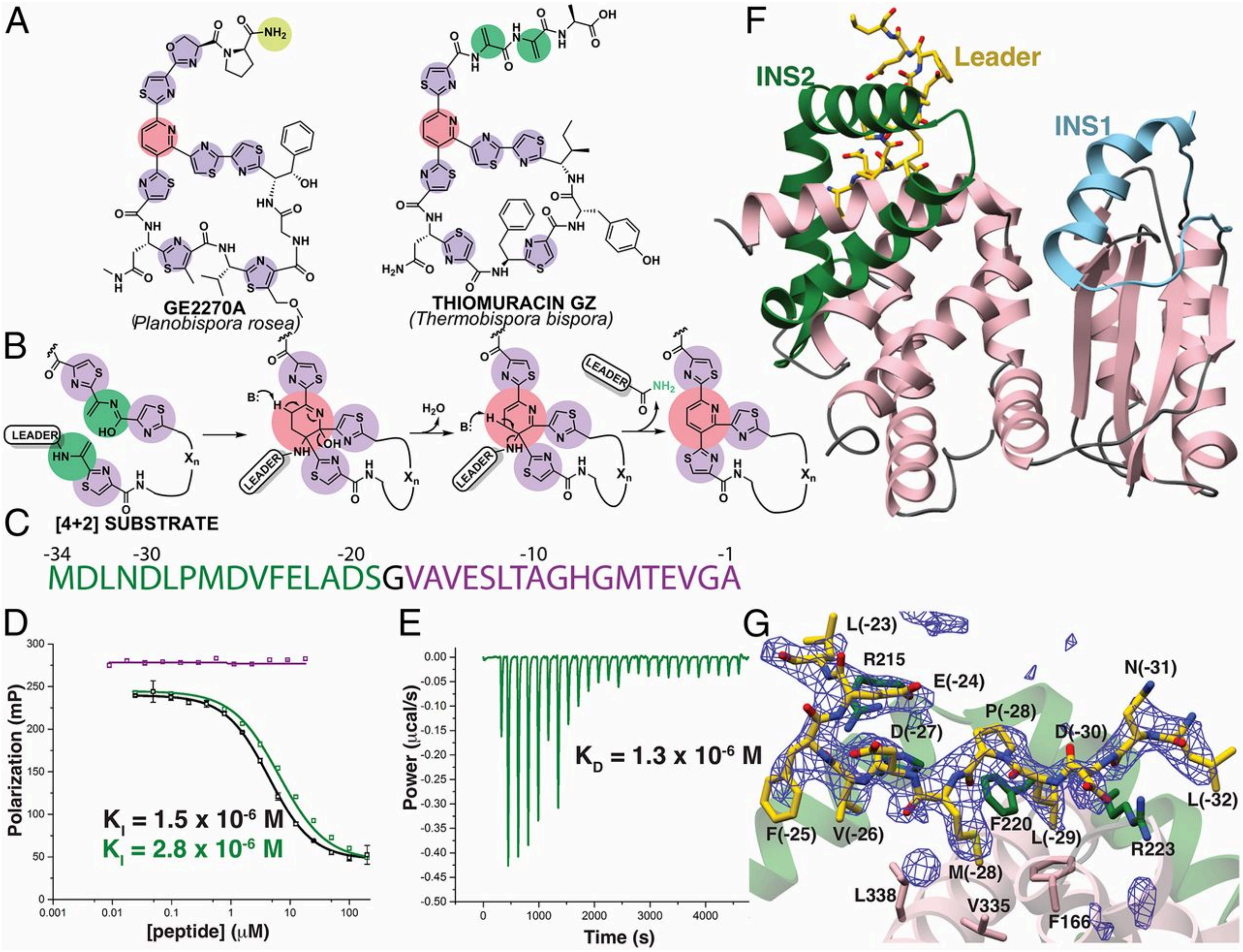


Fig. 3.1: Scheme of enzyme-catalyzed pyridine formation and leader peptide binding of the pyridine synthase, TbtD. (A) Chemical structures of representative thiopeptides with residues color-coded by posttranslational modification. (B) Proposed enzymatic route for the installation of the central nitrogenous six-membered ring from two dehydroalanine residues, followed by the elimination of water and the leader peptide. (C) Primary sequence and numbering scheme of the TbtA leader peptide with the fragment that is engaged by TbtD colored in green. (D) Competition fluorescence polarization binding experiments where TbtD saturated with fluorescently labeled full-length leader peptide is challenged with either unlabeled full-length leader peptide (black curves), or fragments corresponding to the N-terminal (green curve and sequence) or C-terminal (purple curve and sequence) regions of the leader peptide. The N-terminal fragment is sufficient to compete with avidity similar to that of the full-length leader peptide. (E) ITC shows that the N-terminal leader peptide fragment binds to TbtD. (F) Overall structure of TbtD in complex with the N-terminal leader fragment, showing insertions 1 (cyan) and 2 (green) that distinguish the [4+2] enzymes from the LanB elimination domains. (G) Simulated annealing difference Fourier map ($F_{obs} - F_{calc}$) contoured at 2.5σ showing residues from the leader peptide that bind to a region near insertion 2 (green) in TbtD (Figure credit: Satish Nair).

hexathiazole-containing peptide by first *O*-glutamylating Ser residues in the core sequence, followed by glutamate β-elimination, to yield the corresponding α,β-unsaturated amino acids termed dehydroalanine (Dha). Finally, the central six-membered nitrogenous ring is formed through the heteroannulation of two Dha residues in the modified core, possibly after tautomerization of one Dha to a conjugated iminol (Fig. 3.1B). Subsequent aromatization of the central pyridine is established with a dehydration and β-elimination of the leader peptide as a carboxamide (Fig. 3.1B). Recent *in vitro* studies using recombinant pyridine synthases demonstrated that a single enzyme plays a direct role in annulation via a formal [4+2] *aza*-cycloaddition.[86,89,90] Ala-scanning mutational analysis of TbtA suggested the recognition by TbtD to be most influenced by N-terminal residues of the leader peptide.[86]

While other enzymes have been identified as [4+2] cycloaddition catalysts,[91,92] they frequently serve only to enhance a reaction that can occur spontaneously and/or favor product stereochemistry via conformational restriction of the reactant(s).[91] In contrast, the reaction catalyzed by the thiopeptide pyridine synthase is not spontaneous and synthetically equivalent reactions require either prolonged incubation (~1 week) at high temperatures, or microwave irradiation for non-enzymatic catalysis, often with competing side reactions.[93] The thiopeptide-forming pyridine synthases are dedicated [4+2] cycloaddition catalysts whose only function is catalyzing the cyclization of two Dha residues within the precursor peptide.[94] Notably, pyridine synthases catalyze the unique formation of an *aza*-cyclic product, in contrast to the carbocyclic products formed in other biosynthetic pathways, such as those for spinosin A,[95,96] pyrroindomycins,[97] *trans*-decalin containing polyketides,[98] and leporin B.[99]

3.3 RESULTS AND DISCUSSION

3.3.1 TbtD binds the N-terminus of its leader peptide

We initially focused our studies on TbtD, a pyridine synthase from the thiomuracin biosynthetic cluster from *T. bispora*. We previously established TbtD activity *in vitro* utilizing the TbtA precursor substrate with six thiazoles and four Dha residues enzymatically pre-installed.[86,89] Proteolytic removal of the N-terminal 31 residues (out of 34) of the leader peptide from this substrate abolished pyridine formation by TbtD. A similar leader peptide dependency was observed for TclM, the orthologous enzyme from the thiocillin biosynthetic cluster.[100] Maltose-binding protein (MBP)-tagged TbtD bound to a fluorescently labeled TbtA leader peptide with a dissociation constant (K_D) of 1.4×10^{-7} M.[86] However, sequence analysis of the pyridine synthases failed to identify a precursor peptide recognition element (RRE) that mediates leader peptide-binding in many other RiPP enzymes (see Fig. 3.1C for leader peptide numbering scheme).[83] We carried out fluorescence polarization competition assays to identify the minimal leader peptide-binding motif and provide a quantitative affinity assessment for TbtD independent of any catalytic events (Fig. 3.1D). Competition studies using tag-free TbtD and unlabeled full-length TbtA leader peptide yielded an inhibition constant (K_I) value of 1.5×10^{-6} M. Prior studies demonstrated that engagement by TbtD is compromised by mutations at residues at the N terminus of the leader peptide.[86] Accordingly, a 17-residue peptide corresponding to the C-terminal portion of the leader peptide [from Val(-17) to Ala(-1)] failed to compete against the full-length leader peptide, while a 16-residue N-terminal portion [from Met(-34) to Ser(-19)] yielded a K_I of 2.8×10^{-6} M. The binding of this peptide by TbtD was corroborated using isothermal titration calorimetry (ITC), which yielded a K_d of 1.3×10^{-6} M (Fig. 3.1E), while the C-terminal region of the leader peptide did not produce an observable signal.

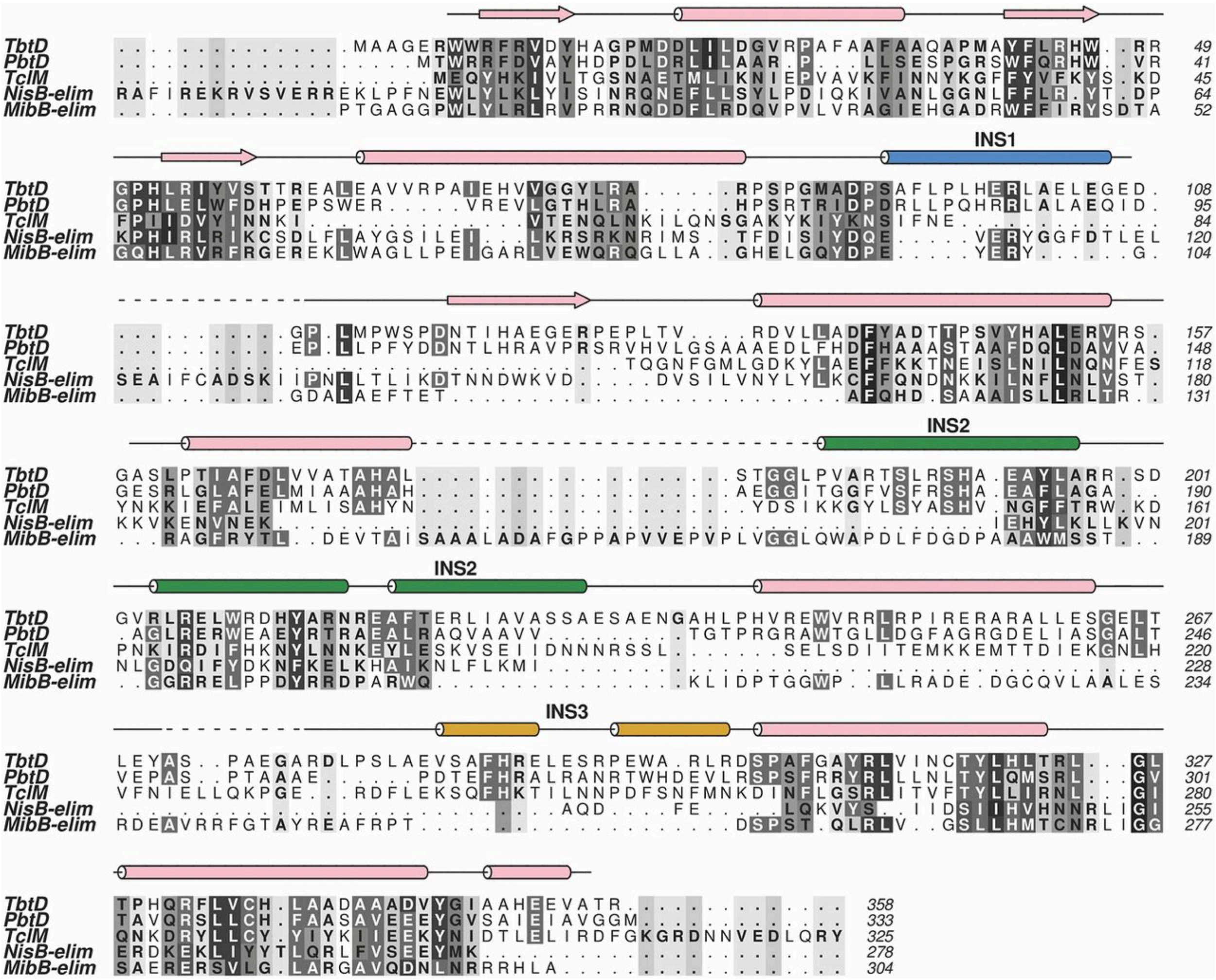


Fig. 3.2: Secondary-structure–based multiple sequence alignment (MSA). Included in the MSA are the pyridine synthases TbtD, PbtD, and TclM in addition to the class I LanB dehydratase elimination domains (-elim) NisB-elim, and MibB-elim generated using ALINE software (45). A majority of secondary structural elements are conserved in these functionally divergent enzymes and notable differences are observed in the insertion regions (INS 1, INS 2, and INS 3) found mainly in the pyridine synthases (Figure credit: Satish Nair).

3.3.2 Overall fold and new leader peptide binding mode of the pyridine synthase

Given the lack of an RRE, we determined crystal structures of TbtD both alone and in complex with the N-terminal region of the leader peptide to define its binding site. TbtD is structurally homologous to the core of the lanthipeptide dehydratase elimination domain (Pfam identifier PF14028; abbreviated as Lant_Dehydr_C) elucidated in NisB[101] and MibB,[53] but contains numerous secondary structural insertions (Figs. 3.1F and 3.2). Notable examples include a helical insertion in the region spanning Pro86 through Asp116 (insertion region 1), a reorganization of the region encompassing Pro183 through Pro241 (insertion region 2), and a large insertion spanning Gly264 through Arg301 (insertion region 3), which is a short connection loop in the Lant_Dehydr_C domains. Importantly, each of the insertions help to configure the binding sites for the substrates specific to the pyridine synthases, and insertion 3 is only visible in structures with bound pyridine core (see ***PbtD Binds a Trisubstituted Pyridine Product Analog of GE2270A***; Fig. 3.3C). The cocrystal structure with the N-terminal 16 residues of the TbtA leader peptide shows electron density corresponding to residues Leu(-32) through Leu(-22) in the vicinity of insertion region 2 (Fig. 3.1G). Residues Ser200 through Ala234 of TbtD form a large, bent helical insert that abuts the enzyme core and form the binding pocket for the leader peptide. The equivalent region in both MibB (Gly990 through Glu1035) and NisB (Asp882 through Gly919) is largely devoid of secondary structure (Fig. 3.2).

3.3.3 Mutagenesis and leader peptide binding studies

Previous binding experiments between TbtD and site-directed alanine variants of TbtA identified several residues within the leader peptide that are important for binding,[86] which can now be rationalized in light of the structural data. For example, an Ala substitution at Leu(-29) severely impairs TbtA binding, and in the cocrystal structure, this residue is docked into a hydrophobic pocket that is ensconced by residues from insertion 3. Similar hydrophobic docking interactions are important for leader peptide engagement by enzymes involved in the biosynthesis of lanthipeptides[101] and cyanobactins.[102] Additionally, binding by TbtD is compromised by substitutions at one of several acidic residues in TbtA, including Asp(-30), Asp(-26), and Glu(-23)[86] and the structural data reveal that TbtD engages some of these residues through electrostatic complementarity. We carried out mutational analysis of the binding site in TbtD (see Ref. 103 *SI Appendix*, Fig. S2 and Table S3) to establish the importance of residues in mediating interactions with TbtA, using a shorter peptide corresponding to the N-terminal 12 residues visualized in the structure. Binding of TbtA by the pyridine synthase is compromised upon substitutions at enzyme residues that define the hydrophobic pocket for Leu(-29), including Phe166, Phe220, and Val335. An Ala mutation at Arg223, which is involved in additional hydrophobic contacts with Leu(-29) and hydrogen-bonding interactions with main-chain amides of the leader peptide, also diminishes TbtA binding (see Ref. 103 *SI Appendix*, Fig. S2 and Table S3). These structural and mutational data establish the validity of a leader peptide-binding site in TbtD and demonstrate that the C-terminal region of the leader peptide is dispensable for precursor binding to the enzyme.

3.3.4 PbtD binds a tri-substituted pyridine product analog of GE2270A

Repeated attempts to crystallize a complex of hexazole-containing tetradehydrated TbtA (designated as the [4+2] substrate in Fig. 3.1B) bound to catalytically deficient variants of TbtD proved recalcitrant. Binding measurements using fluorescently tagged, modified TbtA hexazole demonstrated that TbtD binds with nanomolar affinity (see Ref. 103 *SI Appendix*, Fig. S3), ruling out weak binding as the cause of crystallization issues. TbtD also binds with micromolar affinity a fluorescently tagged and modified TbtA hexazole treated with endoproteinase GluC, containing only three leader peptide residues, Val(-3) through Ala(-1) (see Ref. 103 *SI Appendix*, Fig. S5), but crystallization attempts using this substrate were not fruitful. Likewise, attempts at cocrystallization of PbtD with intact GE2270A failed. Rationalizing that flexible regions of the modified peptide substrate may interfere with lattice packing, we focused efforts on studies using analogs that retained the central pyridine moiety (Fig. 3.3A). Prior studies have shown that acid hydrolysis of GE2270A produced a trithiazole-substituted pyridine (TSP);[81] therefore, we acid-treated GE2270A and purified two resultant TSP variants to near homogeneity (Fig. 3.3A and Ref. 103 *SI Appendix*, Fig. S8). We exploited the overlap in the absorption spectrum of TSP-F (λ_{ex} ~

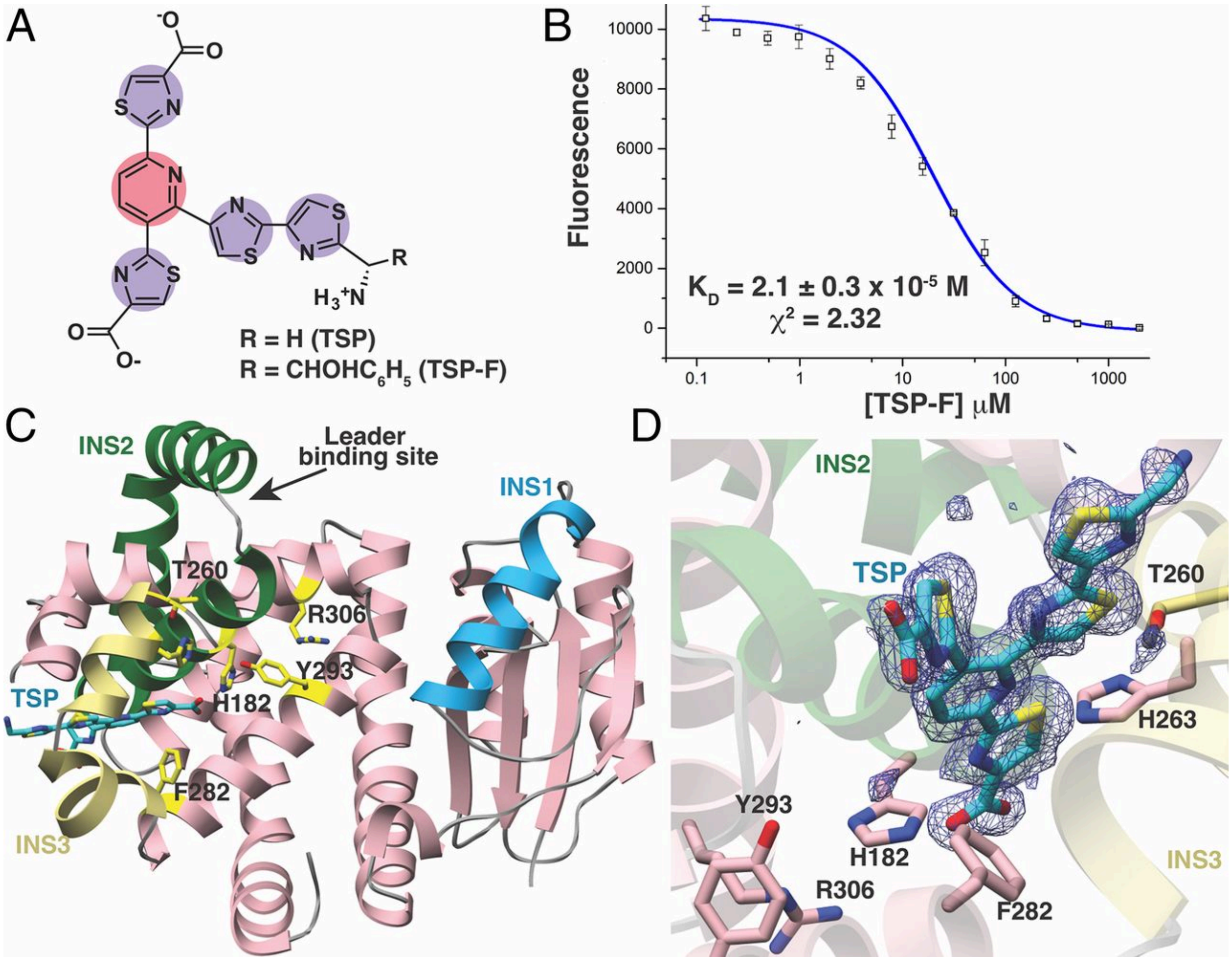


Fig. 3.3: Structural and binding studies of the PbtD–TSP complex. (A) Chemical structure of the trisubstituted pyridine (TSP) core produced by acid hydrolysis of GE2270A. (B) Binding of TSP to PbtD as measured by the changes in FRET upon complex formation. (C) Overall structure of PbtD in complex with TSP shows that residues in insertion 3 (yellow) become ordered upon ligand binding. (D) Simulated annealing difference Fourier map (Fobs – Fcalc) contoured at 2.5σ showing the TSP core binds to a region near insertion 3 (yellow) in PbtD. Active-site residues implicated in catalysis, without influencing leader peptide binding, are shown in stick figures (Figure credit: Satish Nair).

333 nm) and the emission spectrum for the corresponding PbtD pyridine synthase from its biosynthetic cluster (λ_{ex} = 275 nm; λ_{em} = 333 nm) to directly measure binding using Förster resonance energy transfer (FRET), yielding a K_d of 2.1×10^{-5} M (Figs. 3.3B and 3.7).

3.3.5 Structure-guided active site mutagenesis

To discern active-site details of the pyridine synthases, we determined the structure of PbtD both alone and in complex with the GE2270A TSP. The overall structure of PbtD recapitulates the architecture observed in the homologous TbtD, including the lack of interpretable electron density for residues Glu248 through His273 within insertion 3, despite the much higher resolution of this structure. In contrast, the structure with bound TSP reveals that these residues organize and form two short α-helices that define a periphery of the binding pocket for the ligand (Fig. 3.3C). The helix from insertion 2 forms a base at the bottom of the binding groove. Previous Ala-scanning analysis of TbtD implicated several residues in catalysis. Indeed, substitutions at His191 (equivalent PbtD numbering is given in parentheses; His182), Ser287 (Thr260), His290 (His263), Tyr319 (Tyr293), and Arg332 (Arg306) diminish activity.[89] Following a structure-guided mutagenesis approach, this study has revealed an additional residue, Phe308 (Phe282), to be important for product formation after a panel of conserved residues proximal to TSP were tested using similar Ala-scanning mutational approaches (see Ref. 103 *SI Appendix*, Fig. S11). Each of these residues is near the bound TSP but are located ∼5–7 Å from each other and are not all spatially clustered (Fig. 3.3D). Rather, these residues line alternate sides of a shallow cleft that is located at the juncture of the leader peptide-binding site (on TbtD) and bound TSP. It should be noted that some of these variants may carry out the [4+2] cyclization and only be deficient in dehydration activity but are refractory to identification as there would be no net change in mass.

3.3.6 Model of pyridine formation

A feasible model of pyridine formation involves a looped substrate embedded within the shallow cleft in PbtD, where the aforementioned residues may play roles in catalyzing iminol formation, water elimination, and/or leader peptide ejection as the peptide transverses along the binding cleft. In this model, the anchored N terminus of the leader peptide would allow a flexibly tethered core peptide to form this looped structure within the cleft, orienting the C terminus of the substrate away from its N terminus and at the interface of insertion 1 and the periphery of the cleft

(Fig. 3.4). To test this hypothesis, we carried out docking of the [4+2] substrate, as well as various intermediates, and evaluated the stability of each state over a 1-ns time course molecular-dynamics simulation (Figs. 3.8-3.11 and see Ref. 103 *SI Appendix*, Fig. S16). Among the lowest energy poses obtained, we observed the maturing substrate and product docking results to be consistent with the predicted substrate-binding site based on the TSP cocrystal structure (i.e., C terminus and macrocycle orientations). Residues within insertion 3 that are specific to the [4+2] enzymes are largely disordered in the

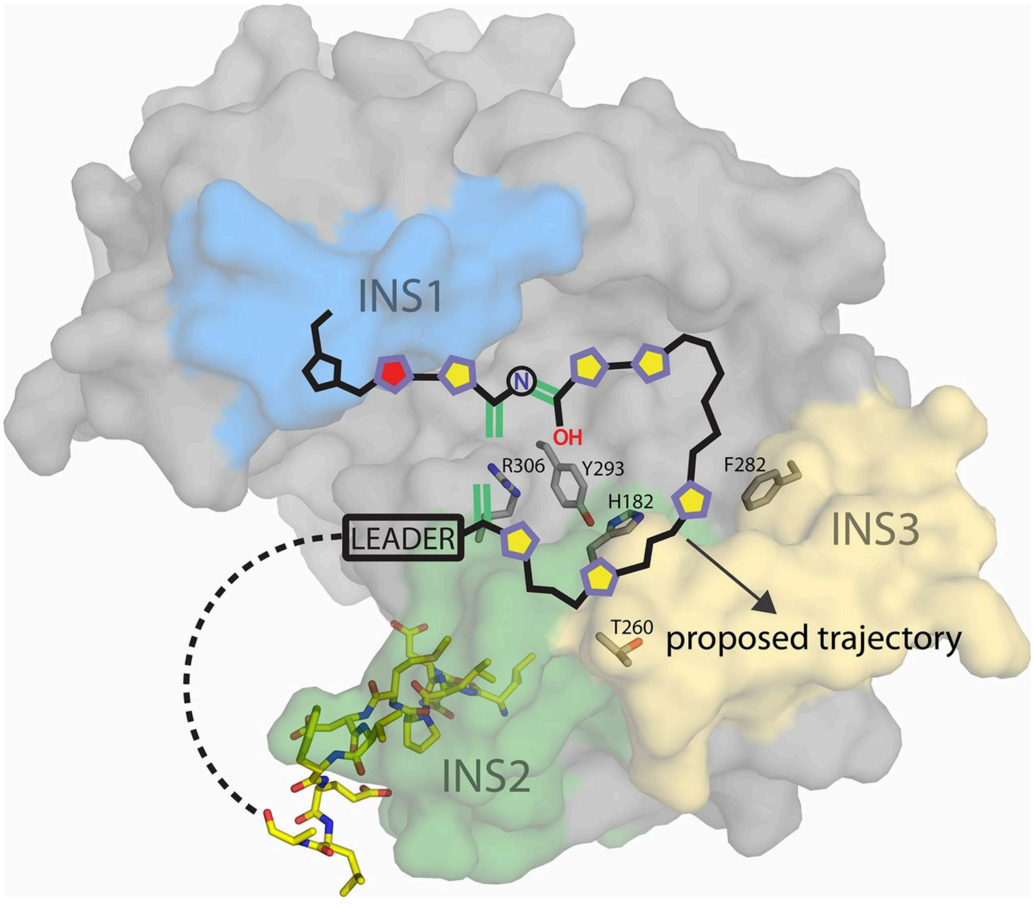


Fig. 3.4: Model of pyridine formation catalyzed by PbtD. The PbtA [4+2] substrate is superimposed onto the structure of PbtD, with active-site residues drawn as gray sticks (Figs. 3.8-3.10; state 1). Five-membered heterocycles along the substrate backbone are represented as pentagons (purple outline/yellow center, thiazole; purple outline/red center, oxazoline; black outline/blue center, proline). Dehydroalanine and iminol double bonds involved in a formal [4+2] cycloaddition are drawn in green. The N-terminal leader peptide fragment (N12-mer: LNDLPMDVFELA) observed in the TbtD cocrystal structure has also been superimposed and is represented as yellow sticks.

absence of substrate, and the proximity of this region to the active-site cleft suggests that these residues may serve a role in active-site capping and/or modified core binding. Indeed, ordering of insertion 3 upon binding of TSP is consistent with this proposal. Recruitment of the true substrate may lead to additional ordering of insertion 3 to facilitate substrate threading, whereupon the resultant macrocycle becomes solvent exposed and is directed away from the binding cleft. The central nitrogenous ring in thiopeptides varies both in the pattern of substitution, as well as in the oxidation state.[87] We generated a sequence similarity network[3] using the available sequences for thiopeptide [4+2] cycloaddition enzymes. This network revealed clear clustering into clades that correlate directly with the substitution pattern and oxidation state of the six-membered heterocycle, even at a fairly low primary sequence similarity cutoff of ~30% (Fig. 3.5). Sequence analysis reveals large differences in the composition and length of residues in insertion 3, consistent with

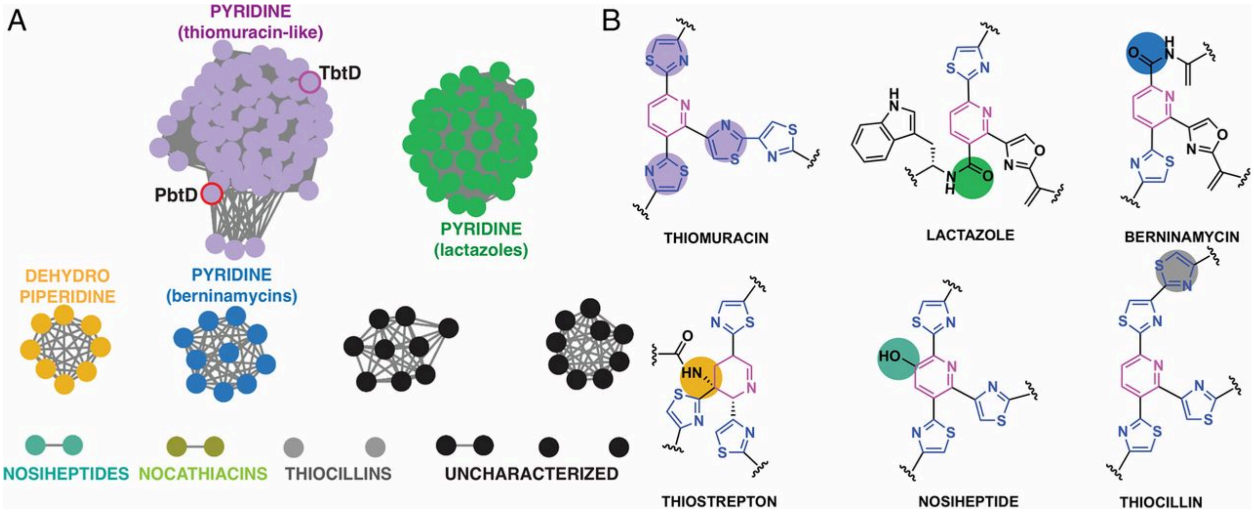


Fig. 3.5: (A) Sequence similarity network of [4+2] cycloaddition proteins found across various thiopeptide biosynthetic clusters showing that the sequences cluster based on the nature of the central nitrogenous ring. Nodes represent protein sequences while edges denote sequences with at least ~30% amino acid identity. PbtD and TbtD are highlighted. The macrocycle compositions are as follows: thiostrepton-like, 26-membered dehydropiperidines; thiocillin/nosiheptide-like, 26-membered pyridines; thiomuracin/GE2270-like, 29-membered pyridine; lactazole-like, 32-membered pyridine; berninamycin-like, 35-membered pyridine. (B) Structures of the central ring found in different thiopeptide natural products with colored circles to indicate changes in the substitution pattern. The colors correspond with the clusters shown in A. It is noted that the nocathiacin and nosiheptide [4+2] enzymes form two unique clusters, despite these thiopeptides bearing high structural homology and an identical pattern of pyridine substitution. This is likely due to limited sample size and some variation at the N terminus between sequences in these clusters (Figure credit: Graham Hudson).

our proposed role of this region in directing the path of the core region of the modified peptide substrate, and thus the size of the macrocycle.

3.3.7 A conserved scaffold catalyzes new chemistry by active site shifting

Mapping the primary sequences of nine pyridine synthases onto the structure of the PbtD–TSP cocrystal structure reveals a pattern of conservation at the active site (Fig. 3.6A). As noted, this active site of the [4+2] enzyme is distinct from that of other Lant_Dehydr_C domains, which

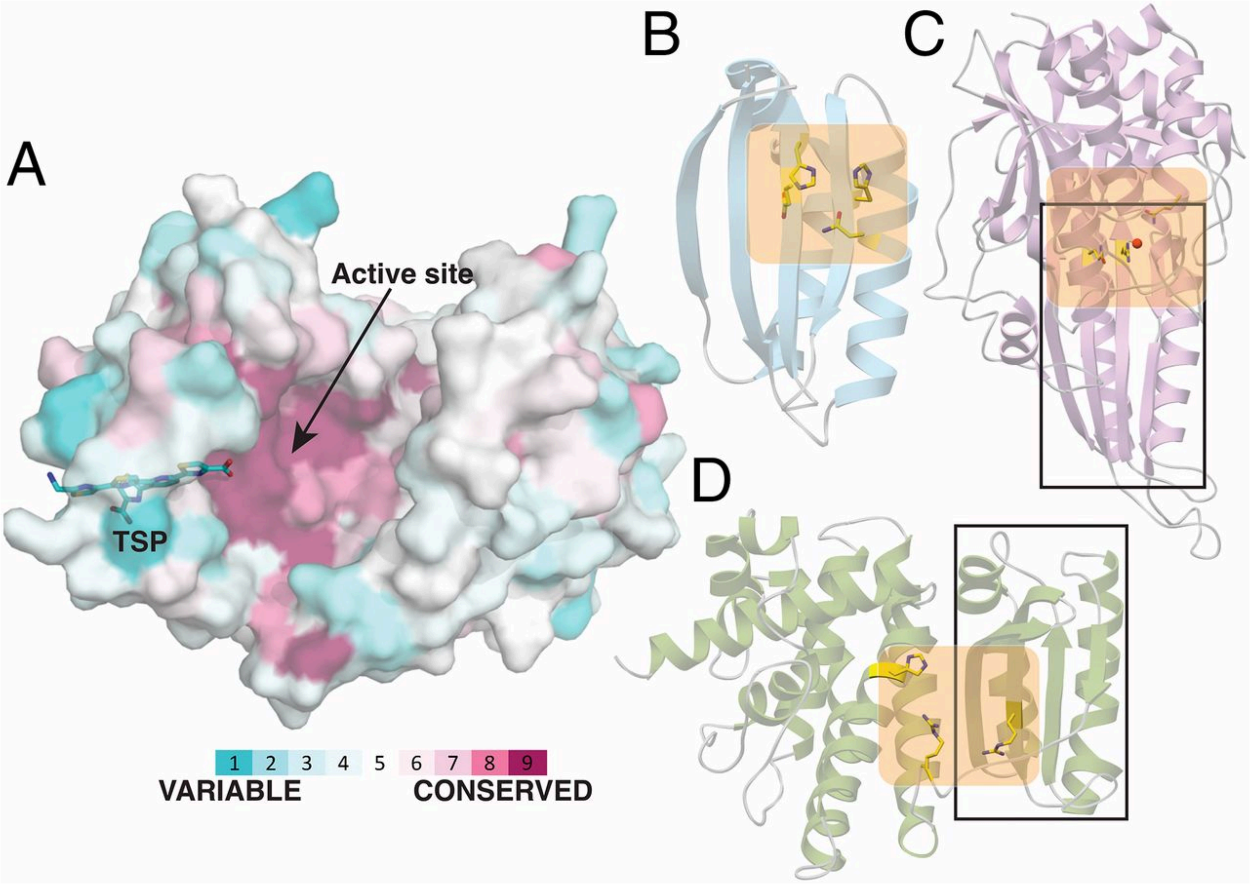


Fig. 3.6: Pyridine synthase scaffold is conserved in enzymes with divergent functions. (A) PbtD surface rendering with the bound TSP shown as a stick figure in cyan. Using the ConSurf server, the surface has been color-coded according to conservation in primary sequence based upon the alignment of nine sequences of pyridine synthases and colored according to regions of most variability (cyan), modest (white) or highest conservation (magenta) (see SI Appendix, SI Materials and Methods for a list of proteins and accession numbers used in alignment). The strongest sequence conservation occurs near the active site where the [4+2] cycloaddition occurs. (B) Ribbon diagram of the LsrG epimerase (PDB entry 3QMQ) with active-site residues shown as yellow sticks. (C) Ribbon diagram of the allantoate amidohydrolase AAH (PDB entry 4PXD) with active-site residues shown as yellow sticks and a manganese ion shown as a red sphere. (D) Ribbon diagram of the elimination domain of the lanthipeptide dehydratase MibB (PDB entry 5EHK) with active-site residues shown as yellow sticks. The ferredoxin-like α+β folds of AAH and MibB are demarcated within the rectangle, and the active sites are colored in orange (Figure credit: Satish Nair).

are built upon a common scaffold observed in the LsrG epimerase,[104] involved in quorum sensing. LsrG contains a ferredoxin-like domain, a prevalent α+β fold observed in numerous metabolic enzymes (Fig. 3.6B).[105] For example, the architecture of allantoate amidohydrolase, an enzyme involved in purine degradation in both plants and bacteria,[106] is formed through the fusion of the LsrG scaffold with a second domain containing a metal-binding active site (Fig. 3.6C).[107] Likewise, the elimination active site of NisB is formed by the insertion of an additional subdomain onto the LsrG ferredoxin-like fold (Fig. 3.6D). Here, we demonstrate that the thiopeptide [4+2] cycloaddition enzymes amend the Lant_Dehydr_C fold with functionally relevant insertion regions to achieve gain of function. A comparison of the TbtD/PbtD structures with that of MibB, and LsrG also provides an unexpected occurrence of proteins that contain active sites located at different regions of the polypeptide and catalyze different reactions, illustrating how a common scaffold can accommodate new chemistry at a new active-site region (Fig. 3.6). Recent de novo protein design efforts by Baker and coworkers[108] describe methodologies for engineering proteins with entirely new primary sequences through considerations of long-range interactions observed in prevalent folds such as the ferredoxin-like domain. These studies predict the possibility of engineering custom-designed large assemblies from such robust building blocks. The structure of the pyridine synthase provides a conspicuous example of how nature has already employed this strategy to yield new catalysts, based on an ancestral ferredoxin-like fold.

3.4 METHODS

3.4.1 Peptide purification

Peptides used for co-crystallization and fluorescence polarization (FP) experiments were synthesized by GenScript and purified by high-performance liquid chromatography (HPLC) using a Grace VisionHT C18 HL column with 5 µm particle size and 250 mm length (Part No. 5151988).

A gradient of 5-55% aqueous acetonitrile (MeCN) containing 0.1% formic acid (*v/v*) was used to separate peptides. Peptide identity was verified using a Bruker UltrafleXtreme matrix-assisted laser desporption/ionization time-of-flight/time-of-flight (MALDI TOF-TOF) mass spectrometer (Bruker Daltonics) in reflector positive mode at the University of Illinois School of Chemical Sciences Mass Spectrometry Laboratory.

3.4.2 Purification of wild-type and site-directed variants of TbtD and PbtD

All protein expression constructs were generated using the following procedure with the exception of N-terminal hexahistidine maltose-binding protein (His_6-MBP) tagged TbtD, TbtD-R332A, and PbtD which were cloned according to methods previously described.[89] Details of protein production and purification are provided in the Supporting Materials and Methods.

3.4.3 Fluorescence polarization (FP) binding assay

To determine the dissociation constants (KD) of the TbtD variants towards N-terminally fluorescein isothiocyanate-labeled leader peptide (FITC-TbtA leader), an FP assay was carried out in a 384-well format using flat-bottom black polystyrene non-binding plates (Corning) and a Synergy H1 plate reader (BioTek Instruments) equipped with a green FP filter cube (λex = 485 nm and λem = 528 nm). Samples were prepared using the following binding buffer: 0.2 M NaCl, 50 mM HEPES pH 7.5, 2.5% glycerol, and 2 mM β-mercaptoethanol. All other experimental procedures and data analyses were performed as previously described (see Ref. 103 *SI Appendix*, Figs. S2-S5 and Table S3).[86]

3.4.4 Competition fluorescence polarization (FP) binding assay

For each TbtD variant, protein concentrations were selected to achieve initial polarization values of 250-300 mP in the presence of 25 nM FITC-labeled peptide. Protein and labeled peptide concentrations were held constant while TbtA N12-mer peptide (LNDLPMDVFELA) was serially diluted from 1 mM to 5.6 nM in binding buffer. Data were collected in a similar manner to that described above. Inhibition constants (K_I) were determined by taking the IC_{50} values (concentration of peptide required to diminish maximum polarization by 50%) derived from a dose-response fit in OriginPro 2016 (OriginLab) and applying the following competitive inhibition equation:

$$K_I = \frac{IC_{50}}{\frac{L_{50}}{K_D} + \frac{P_0}{K_D} + 1} \tag{1}$$

Where L_{50} is the concentration of free labeled ligand at 50% inhibition, K_D is the dissociation constant between protein and labeled ligand, and P_o is the initial protein concentration (Fig. 3.1D, and see Ref. 103 *SI Appendix* Fig. S3 and Table S3).[109]

3.4.5 Isothermal titration calorimetry (ITC)

ITC data were collected on a Nano ITC calorimeter (TA Instruments) using the low volume sample cell (190 µL) maintained at 25 °C with stirring at 220 rpm. All protein and peptide dilutions were prepared using the same binding buffer as used in FP experiments. The sample cell contained TbtD at 40 µM and peptides at 0.5 mM were injected into the sample cell in 0.2 µL increments for 26 total injections (Fig. 3.1E). All downstream data processing was carried out using the AFFINImeter online data analysis tool (https://www.affinimeter.com/; Software for Science

Developments). The first and sixth injections from the left in Figure 3.1E were omitted in the calculations.

3.4.6 Production and purification of TSP, TSP-F, FITC TbtA-hexazole, FITC GluC TbtA-hexazole, and TbtA tetradehydrate

Tri-thiazole-substituted pyridines (TSP and TSP-F) were generated by acid hydrolysis of GE2270A (a gift from Novartis) (see Ref. 103 *SI Appendix* Fig. S9A). Hydrolysis was carried out at 110-120 °C for 12 h in the presence of 6 M HCl. The reaction mixture was subsequently lyophilized and extracted with 10% (*v/v*) MeCN. Insoluble material was removed by centrifugation at 10,000 × g for 15 min, and the supernatant was purified on a Shimadzu Prominence Preparative Liquid Chromatography system equipped with a Phenomenex Luna C18 column (250 × 10 mm, 10 μm particle size, 100 Å pore size). MeCN and 10 mM aq. NH_4HCO_3 were used as the mobile phases with a gradient of 2 - 80% aq. MeCN over 45 min at 7 mL/min for separation. TSPs eluted from the column with a retention time around 13 and 16 min (see Ref. 103 *SI Appendix* Fig. S9B). Fluorescein-labeled TbtA hexazole (see Ref. 103 *SI Appendix* Fig. S4; FITC TbtA-hexazole), and TbtA tetradehydrate (Fig. 3.8; compound **1**) were prepared and purified as previously described.[89,86] Fluorescein-labeled and GluC-treated TbtA-hexazole (FITC GluC TbtA-hexazole) was prepared using methods previously described[86] and HPLC purified using a Grace Vydac DENALI C18 column (250 × 4.6 mm, 5 μm particle size, 120 Å pore size) with a similar gradient of mobile phases described above for TSP and TSP-F and a 1 mL/min flow-rate. Desired product eluted with retention time around 27 min.

3.4.7 Förster resonance energy transfer (FRET) binding assay

Binding experiments were initially carried out using TSP; however, due to insolubility, we opted to use the more soluble TSP-F (see Ref. 103 *SI Appendix* Fig. S9). To ensure a spectral overlap between PbtD native protein fluorescence and TSP-F absorbance, emission and absorbance scans were performed on a Synergy H4 Hybrid plate reader (BioTek) and an Epoch microplate spectrophotometer (BioTek), respectively. PbtD (5 μM) was excited at 275 nm before collecting an emission scan from 300 to 600 nm. An absorbance spectrum of TSP-F (0.5 mM) was collected over a similar range (see Ref. 103 *SI Appendix* Fig. S7). A λ_{ex} of 275 nm and λ_{em} of 333 nm were selected to satisfy spectral overlap. Emission spectra were also taken of TSP-F in water and methanol by exciting at 333 nm and monitoring emission from 333 to 750 nm. Fluorescence of TSP-F was only observed in methanol, preventing detection of fluorescence in aqueous binding buffer. Instead, quenching of PbtD fluorescence (observed in aqueous binding buffer) at 333 nm was monitored as a function of TSP-F concentration. To ensure adequate fluorescence signal above the instrument background emission, PbtD was maintained at 5 μM, while TSP-F was serially diluted from 2 mM to 122 nM. To correct for background emission quenched by TSP-F, the following formula was applied:

$$Fl(x) + [Fl_b - Fl_b(x)] - Fl_b \tag{2}$$

where $Fl(x)$ is the total fluorescence in the presence of PbtD and some ligand concentration (x), Fl_b is the total background fluorescence in the absence of protein or ligand, and $Fl_b(x)$ is the total background fluorescence in absence of protein and presence of some concentration of ligand (x). In order to obtain values for Fl_b and $Fl_b(x)$, a blank solution containing only binding buffer was measured in addition to serially diluted TSP-F in the absence of PbtD. Measurements of K_D were approximated by fitting the curves to a dose-response curve using OriginPro 2016 (OriginLab) and taking K_D to be equal to the ligand concentration at 50% Fl_{max} (Fig. 3.2B and see Ref. 103 *SI Appendix* Fig. S8).[109,110]

3.4.8 Protein Crystallization

TbtD (apo): Following SEC, TbtD was concentrated to 5 mg/mL before crystallization in hanging drops at 16 °C whereby protein solution and reservoir solution (0.4 M NaH_2PO_4, 1 M KH_2PO_4 pH 6.5) were incubated 1:1 (2 µL drops). Cube-shaped crystals grew in 2-3 d.

TbtD-16mer: Following SEC, TbtD was concentrated to 6 mg/mL before equilibrating with 1 mM 16mer peptide (MDLNDLPMDVFELADS) for 30 min on ice. Thick plate-shaped crystals grew within 48 h in hanging drops at 16 °C by equilibrating 1:1 protein solution and reservoir solution (2 µL drops) containing 0.2 M magnesium formate and 17.5% polyethylene glycol (PEG) 3350.

PbtD (apo): Following SEC, PbtD was concentrated to 3 mg/mL before crystallization in hanging drops at 9 °C by equilibrating 1:1 protein solution and reservoir solution (2 µL drops) containing 15% 2-methyl-2,4-pentanediol (MPD), 8% PEG 6K, 0.1 M 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 6.5. Large, cube-shaped crystals grew within 18 h. Selenomethionine-containing PbtD (SeMet-PbtD) was crystallized in the same manner with the following reservoir solution: 17% MPD, 10% PEG 6K, and 0.1 M MES pH 6.5.

PbtD-TSP: Following SEC, PbtD was concentrated to 4.5 mg/mL before equilibrating with a 1.5 mM TSP suspension for 1 h at room temperature. Clustered rod-shaped crystals grew within 48 h in hanging drops at 9 °C by equilibrating 1:1 protein solution and reservoir solution (2 µL drops) containing 1.9 M Li_2SO_4, 2% PEG 400, and 0.1 M tris pH 8.5.

3.4.9 Crystallography data collection, structure determination, and refinement

Prior to data collection, crystals were cryoprotected in similar crystallization conditions supplemented with 20% ethylene glycol and flash-frozen in liquid nitrogen. Data were collected at the Advanced Photon Source (APS) at Argonne National Lab using the Life-Science Collaborative Access Team (LS-CAT) 21-ID-D, 21-ID-F, and 21-ID-G beamlines. Initial diffraction data were processed using autoPROC.[71] Experimental phases for PbtD were determined by using diffraction data from SeMet-PbtD and running Phenix AutoSol.[33] The initial structure was built using Phenix AutoBuild and further refined using REFMAC5[35,76] in combination with manual rebuilding using COOT[36,75]. Solvent molecules were incorporated using Phenix Refine.[38] Prior to fitting ligands into electron density difference maps, geometry optimization and restraint parameters for tri-thiazole-substituted pyridines (TSPs) were produced using Phenix eLBOW.[37] Crystallographic phases for TbtD were determined by molecular replacement using PbtD coordinates as a search model for Phaser.[111] Calculated phases were used to build an initial model, followed by manual and automated refinements as described for PbtD.

3.4.10 ConSurf analysis

Pyridine synthases selected for ConSurf[112] analysis are as follows with the protein name and corresponding thiopeptide listed (if applicable): AGY49590.1 (PbtD; GE2270 from Planobispora rosea), WP_000438834.1 (TclM; thiocillin), ACS83769.1 (TpdD; GE2270 from Nonomuraea sp. WU8817), AGN11669.1 (BerD; berninamycin), WP_025357078.1 (suspected thiomuracin-like), ACR48344.1 (NosO; nosiheptide), WP_013130812.1 (TbtD; thiomuracin), WP_052527475.1 (suspected thiomuracin-like), and ACS83784.1 (suspected thiomuracin-like).

3.4.11 Docking analysis

Preliminary docking studies of PbtD and modified PbtA substrates were carried out using the AutoDock Vina PyMOL plugin.[77] Prior to docking, ligands were generated and 3D optimized in MarvinSketch 16.4.11 (ChemAxon). The dehydration and leader peptide ejection requirements of the proposed reaction mechanism catalyzed by pyridine synthases were taken into account by docking different substrate intermediate states of modified PbtA (Fig. 3.8C). In addition to the full suite of azol(in)e modifications present in GE2270A, the remaining post-translational modifications (i.e. oxidations, methylations, and C-terminal amide) were also included in the PbtA-derived substrates (Fig. 3.8B). The leader peptide was truncated substantially to *N*-acetyl-L-alanine at the N-terminus to simplify the docking experiments and avoid complications of non-specific binding. The only chemical variation from states 1 through 4 involves the substitution and connectivity of atoms 1 through 6 involved in the formal [4+2] cycloaddition and constituting the final pyridine ring (Fig. 3.9A,B; labeled in red). Stereocenters at atom positions 2 and 3 in state 2 were selected as *R* and *S*, respectively, to reflect the conserved stereochemistry observed in all characterized (to the best of our knowledge) dehydropiperidine-containing thiopeptides (i.e. thiostrepton, siomycin, thiopeptin, 8 and Sch 18640).[113–117] Correspondingly, the *R* stereocenter at position 3 was chosen for state 3 (Fig. 3.9C,D). In state 2, we also considered the inverted configuration at the 2 position (Fig. 3.9C,D; corresponding to the (*R,R*)-stereoisomer) and comparative docking analyses reveals the (*R,S*) to be more favorable on the basis of binding energy and interatomic distance measurements between the oxygen of water to-be-ejected and terminal guanidinium nitrogen of Arg306 (Fig. 3.11). Receptor coordinates used in AutoDock Vina were generated from PbtD crystal structures (this study) to represent an “open” state derived from the apo PbtD structure lacking residues Val247 - Ser281. AutoDock Vina was performed with a large grid size (>27,000 Å^3) and a grid step of 0.375 Å centered near the active site residues identified in TbtD as important for catalysis (corresponding PbtD residues: His182, Thr260, Phe282, Tyr293, and Arg306).[89] Initial poses were selected on the basis of docking scores, proximity to active site residues, and expected substrate orientation with respect to the PbtD-TSP co-crystal structure. These poses were then subjected to further docking experiments using the Molecular Operating Environment 2016.08 (MOE) software (Chemical Computing Group Inc.). Prior to docking in MOE, ligands were solvated within a sphere of gas-phase water (10 Å radius) and energy minimized using the MMFF94x force field with a non-bonded interaction cutoff of 10-12 Å.[118,119] The interior region dielectric constant for electrostatic interactions was set to 1 and that for the exterior solvent region was set to 80. Due to the lack of structural resolution of 11 residues (Val247 – Glu257 in PbtD) in all PbtD and TbtD crystal structures, an unstructured loop was built prior to docking using the "Loop Modeler" function in MOE with the PDB loop search enabled. Initial coordinates containing the missing residues were generated using the SWISS-MODEL modeling tool with a user template file consisting of the PbtD-TSP cocrystal structure coordinates containing only protein atoms.[120–122] Using the "Dock" function in MOE, the energy-minimized ligands were docked onto both the loopmodified PbtD receptor and the PbtD "open" state receptor using the sites identified in AutoDock Vina. The triangle matcher placement method was used along with the London dG scoring methodology.[123,124] Low energy poses obtained in MOE satisfy the criteria mentioned above (i.e. docking scores, active site proximity, and orientation; Figs. 3.10-3.11).

3.4.12 Molecular dynamics (MD) simulations

All MD simulations were carried out using the "Dynamics" interface in MOE 2016.08. Poses obtained in MOE using both PbtD receptor variants ("open" state or containing MOE-derived loop region) and all four states containing the *R,S*-stereoisomer at state 2 and 9 *R*-stereoisomer at state 3 (Figs. 3.9C,D) were used as the protein-ligand complexes under study.

These complexes were solvated and energy minimized using the MMFF94x force field with the same parameters described above. The Nose-Poincare-Andersen (NPA) algorithm was selected to solve the equations of motion.[125] Waters within 10 Å from the solute atoms were made explicit throughout the simulation. An equilibration time of 50 ps followed by a production time of 1 ns was specified, and the system temperature was set to 300 K. A time step of 2 fs was used and data were written every 0.5 ps.

3.5 SUPPORTING FIGURES

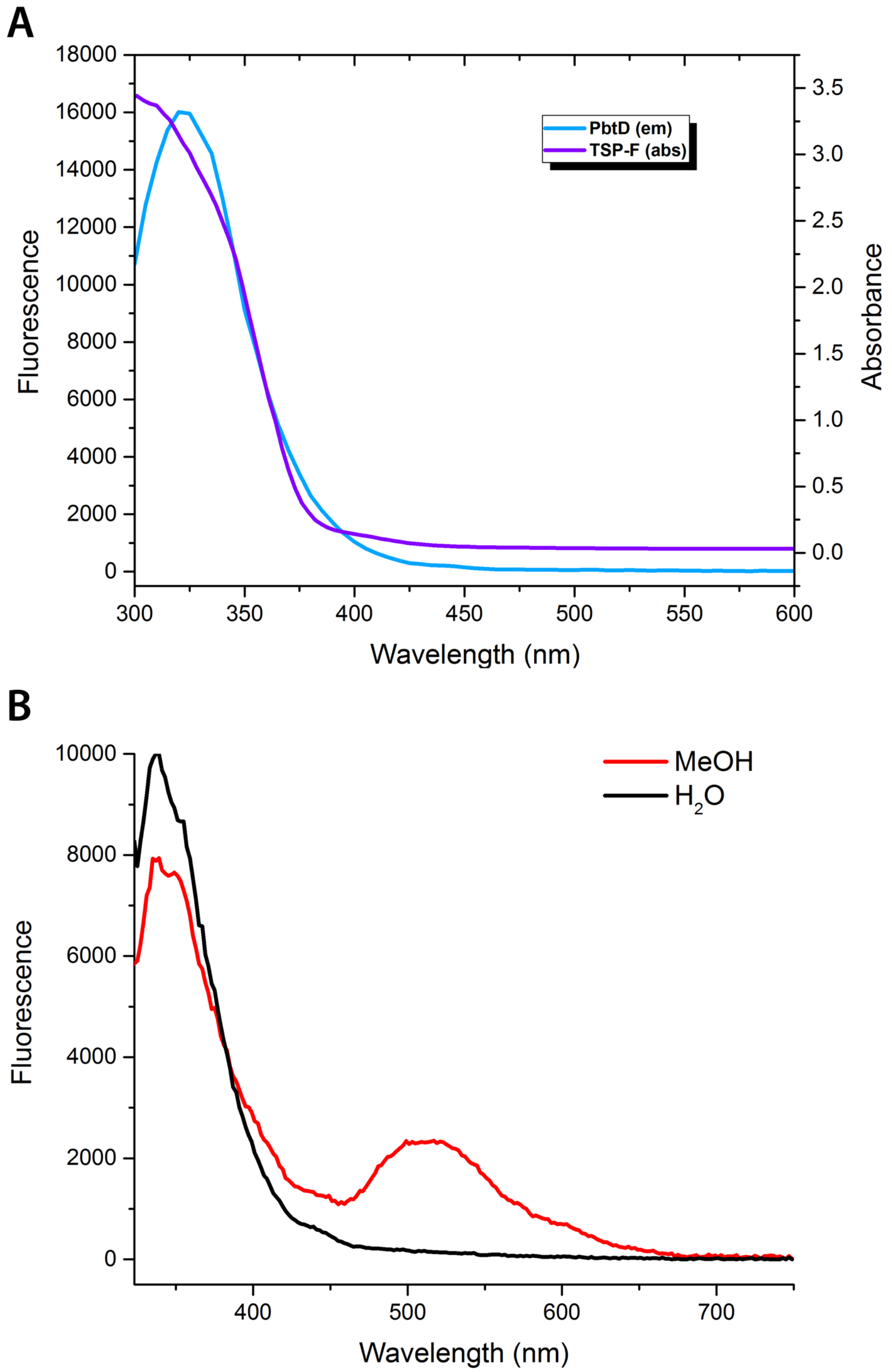


Fig. 3.7: Förster resonance energy transfer (FRET) analysis between PbtD and tri-thiazolesubstituted pyridine containing β-hydroxyphenylalanine (TSP-F; see Ref. 103 *SI Appendix* Fig. S8A). (A) Emission spectrum of native PbtD fluorescence (λex = 275 nm) and absorption spectrum of TSP-F indicates spectral overlap for FRET to take place. (B) Fluorescence spectrum of TSP-F when excited at 333 nm in water (black) and methanol (red) indicates a solvent dependent fluorescence around 510 nm.

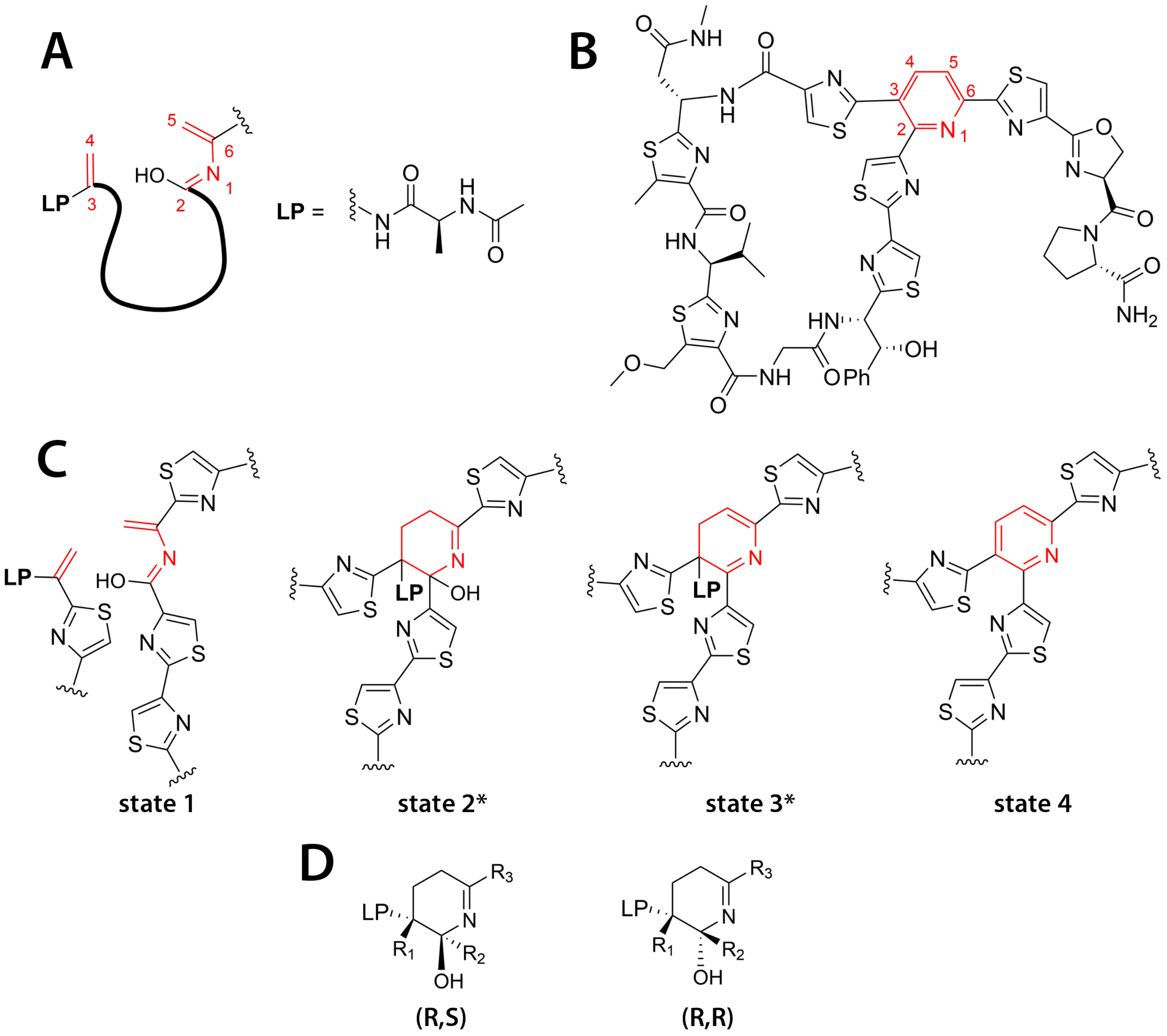
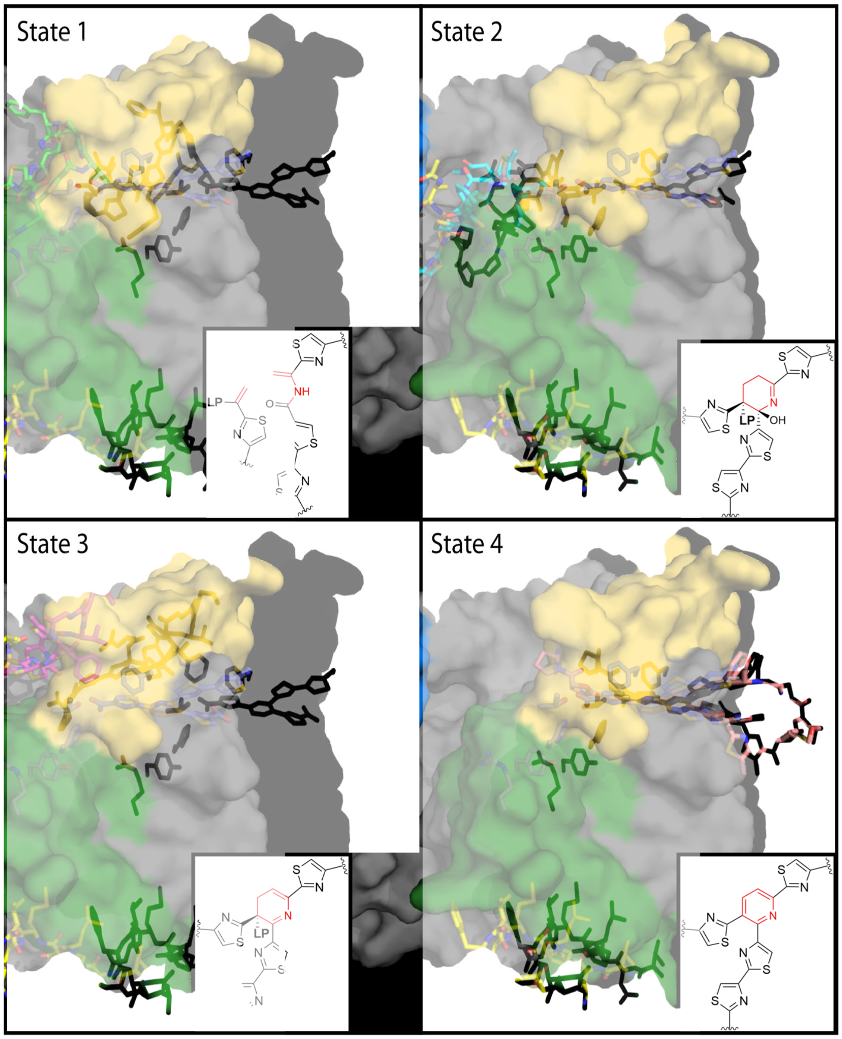
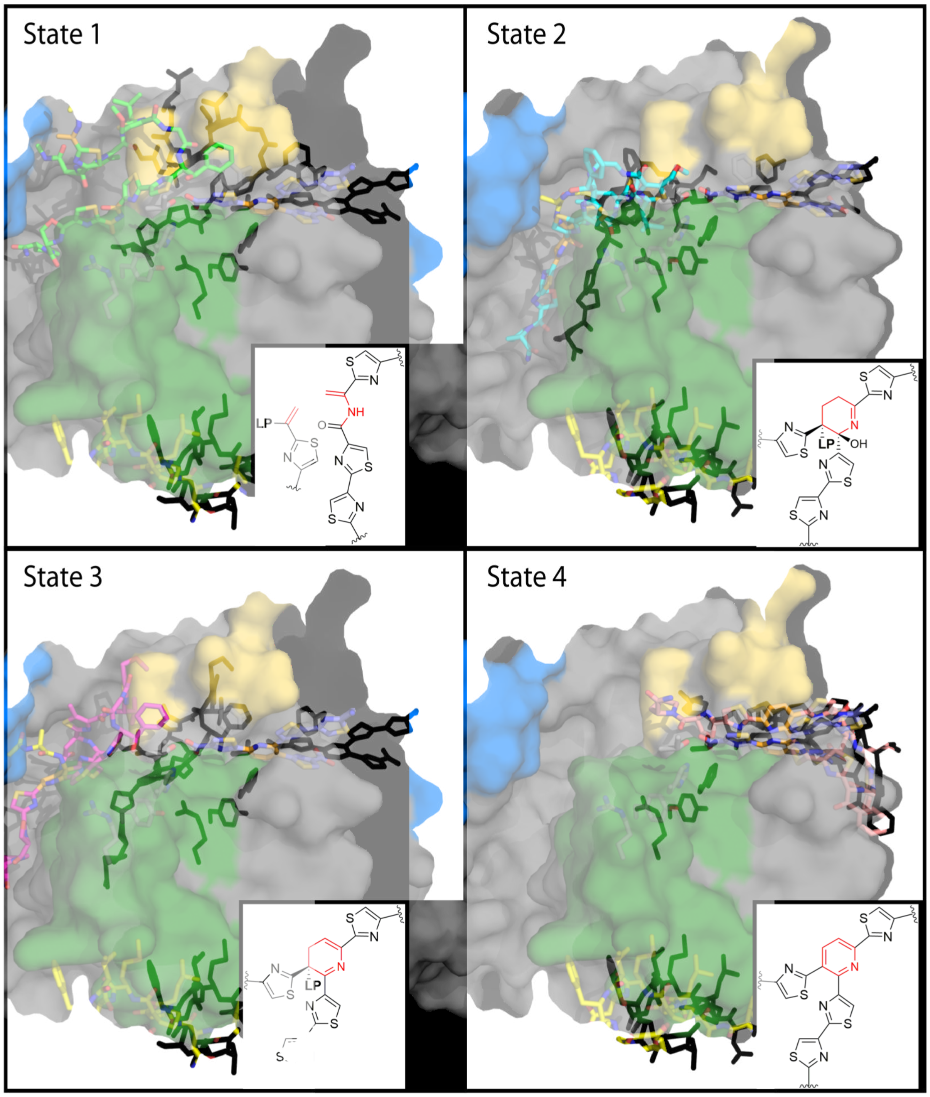


Fig. 3.8: Summary of PbtD ligand structures used in docking experiments. (A) Cartoon drawing of modified PbtA substrate where bold lines indicate portions of the macrocycle and LP represents the truncated leader peptide (*N*-acetyl-L-alanine). (B) Structure of GE2270A to highlight structural features of the macrocycle and C-terminus shared in all of the ligand states where the only differences arise from substitution pattern and connectivity of the atoms in red. (C) States of the modified PbtA substrates highlighting the structural differences at atom positions 1–6 (*MD data only shown for the (*R*,*S*)-stereoisomer at state 2 and R-stereoisomer at state 3). (D) Structures of two unique stereoisomers at state 2 analyzed in this study (R1 = R2 = R3 = thiazole).



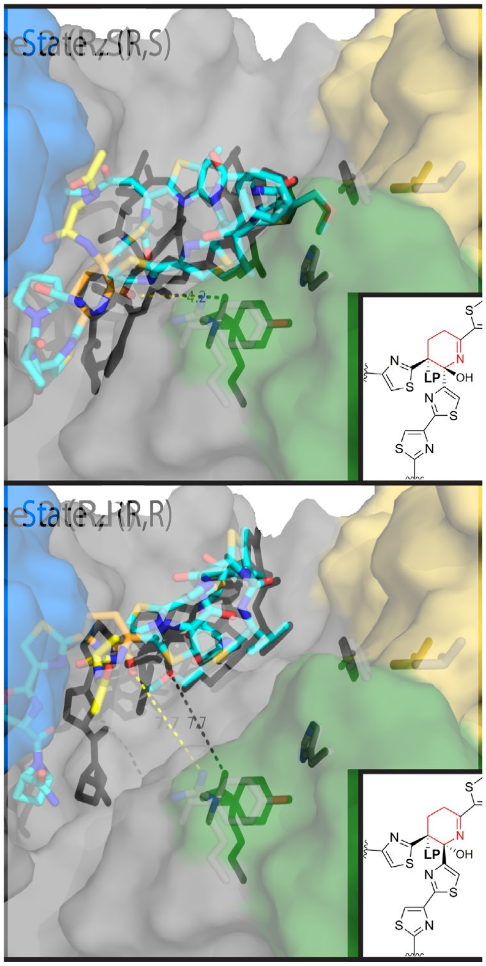
State 1	State 2	State 3	State 4
S (kcal/mol)	S (kcal/mol)	S (kcal/mol)	S (kcal/mol)
-9.6	-12.5	-10.3	-8.8
-7.7	-11.9	-10.2	-8.6
-6.8	-11.4	-9.9	-8.1
-6.5	-10.6	-9.6	-7.9
-6.4	-10.4	-9.6	-7.9

Fig. 3.9: Summary of docking energies (S) and poses obtained in MOE with built-in loop residues (Val247 – Glu257; omitted for clarity). The energies for poses used in the figure are colored red. State 2 corresponds to the (*R,S*) intermediate. In all relevant states, the leader peptide residues (including superimposed TbtA N12-mer from TbtD co-crystal structure) are displayed as yellow sticks and crystallographically observed TSP as blue slate. Atoms 1–6 (Fig. 3.8A) are colored orange, and active site residues H182, T260, F282, Y293, and R306 as grey sticks. Insertions 1-3 are color coded consistent with the main text and Figure 3.2 (i.e. INS 1 = blue; INS 2 = green; INS 3 = yellow; remaining PbtD surface = grey)



State 1	State 2	State 3	State 4
S (kcal/mol)	S (kcal/mol)	S (kcal/mol)	S (kcal/mol)
-13.3	-12.2	-12.3	-10.7
-12.3	-12.2	-10.3	-10.3
-10.6	-11.4	-9.9	-9.7
-9.0	-11.3	-9.6	-9.4
-8.8	-10.6	-9.3	-9.4

Fig. 3.10: Summary of docking energies (S) and poses obtained in MOE using the "open" receptor state as described in the experimental section. The energies for poses used in the figure are colored red. State 2 corresponds to the (*R,S*) intermediate. In all relevant states, the leader peptide residues (including superimposed TbtA N12-mer from TbtD co-crystal structure) are displayed as yellow sticks and crystallographically observed TSP as blue slate. Atoms 1–6 (Fig. 3.8A) are colored orange, and active site residues H182, T260, F282, Y293, and R306 as grey sticks. Insertions 1-3 are color coded consistent with the main text and Figure 3.2 (i.e. INS 1 = blue; INS 2 = green; INS 3 = yellow; remaining PbtD surface = grey).



State 2 (R,S)	State 2 (R,R)
S (kcal/mol)	S (kcal/mol)
-12.5	-11.5
-11.9	-10.6
-11.4	-10.2
-10.6	-10.2
-10.4	-9.4

Fig. 3.11: Comparative docking analysis performed in MOE of the state 2 (*R*,*S*) and (*R*,*R*) stereoisomers. Ligands were docked onto a loop-containing PbtD receptor (containing residues Val247 – Glu257; omitted for clarity) and docking pose energies (S) are tabulated where values in red correspond to graphically represented poses. Distance measurements between the oxygen of the water to-be-ejected and terminal guanidium nitrogen of Arg306 are displayed (in Å). The leader peptide residues (including superimposed TbtA N12-mer from TbtD co-crystal structure) are displayed as yellow sticks and crystallographically observed TSP as blue slate. Atoms 1–6 (Fig. 3.8A) are colored orange, and active site residues H182, T260, F282, Y293, and R306 as grey sticks. Insertions 1-3 are color coded consistent with the main text and Figure 3.2 (i.e. INS 1 = blue; INS 2 = green; INS 3 = yellow; remaining PbtD surface = grey).

CHAPTER 4: STRUCTURAL BASIS FOR SUBSTRATE-ASSISTED AND ENZYMATIC LYSINOALANINE FORMATION IN DURAMYCIN BIOSYNTHESIS[d]

4.1 ABSTRACT

Duramycin is a heavily post-translationally modified peptide that binds phosphatidylethanolamine. It has been investigated as an antibiotic, an inhibitor of viral entry, a therapeutic for cystic fibrosis, and a tumor and vasculature imaging agent. Duramycin contains a β-hydroxylated Asp (Hya) and four macrocycles, including an essential lysinoalanine (Lal) cross-link. The mechanism of Lal formation is not known. Here we show that Lal is installed stereospecifically by DurN via addition of Lys19 to a dehydroalanine. The structure of DurN reveals an unusual dimer with a new fold. Surprisingly, in the structure of duramycin bound to DurN, no residues of the enzyme are near the Lal cross-link. Instead, Hya15 of the substrate makes interactions with Lal, suggesting it acts as a base to deprotonate Lys19 during catalysis. Biochemical data suggest that DurN preorganizes the reactive conformation of the substrate, such that the Hya15 of the substrate can serve as the catalytic base for Lal formation.

4.2 INTRODUCTION

Duramycin and closely related compounds are ribosomally synthesized and post-translationally modified peptides (RiPPs) produced by various actinomycetes[126–130] and marine symbionts.[131] Duramycin binds phosphatidylethanolamine (PE), a major structural phospholipid in mammalian and microbial cell membranes.[132,133] Duramycin contains 19 amino acids, and five posttranslational modifications (PTMs), one lanthionine (Lan), two methyllanthionines (MeLan),

---

[d] This research was originally published in *Nature Chemical Biology*. Part of this chapter was directly obtained, with permission, from:

*An L, *Cogan DP, Navo CD, Jiménez-Osés G, Nair SK, van der Donk WA. Substrate-Assisted Enzymatic Formation of Lysinoalanine in Duramycin. *Nat Chem Biol.* 2018 Oct;14(10):928-933. Copyright © 2018, Springer Nature.

an *erythro*-3-hydroxy-L-aspartic acid (Hya) resulting from the hydroxylation of Asp15, and one (2*S*,9*S*)-Lal (Fig. 4.1). The carboxylate and hydroxyl groups of Hya15 make key interactions with the ethanolamine head group of PE, providing nanomolar affinity and high specificity.[133,134] As such, duramycin has been investigated as an antibiotic,[126–128] inhibitor of viral entry,[135] therapeutic for cystic fibrosis,[137] and tumor and vasculature imaging agent.[138,139] Recently, divamide A, a close homolog of duramycin discovered from the symbiotic microbiome of small marine animals, was reported to display potent anti-HIV activity.[131]

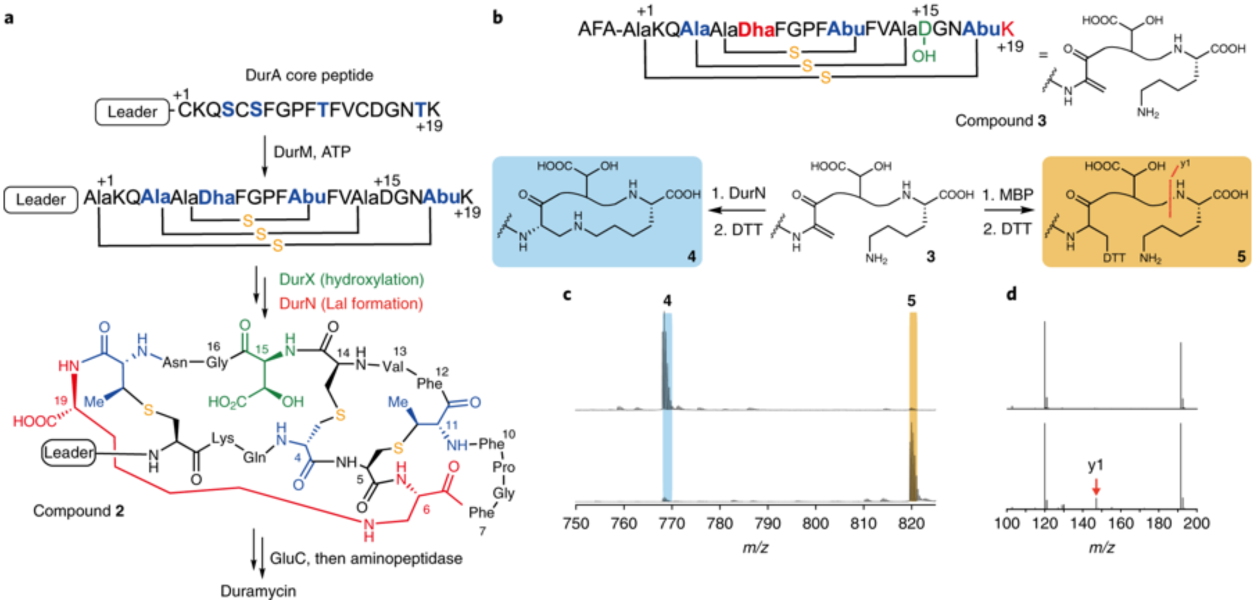


Fig. 4.1: Duramycin biosynthesis. **a**, Co-expression of a minimal biosynthetic gene cassette and in vitro proteolytic removal of the LP from peptide **2** produces duramycin. **b**, Peptide **3** is converted to **4** by DurN in vitro (blue). Formation of the Lal is monitored by reaction with DTT, which provides an adduct (**5**) if Lal is not generated (yellow). The characteristic y1 fragment ion is marked with an orange line. **c**, Compound **3** was reacted with MBP-DurN (top) or MBP (bottom), treated with DTT, and analyzed by LC–MS (see Ref. 136 Supplementary Fig. 2). Expected and observed masses are listed in Ref. 136 Supplementary Table 2. Experiment was performed independently three times with similar results. **d**, LC–MS/MS analysis of the products in **c**. Because of the extensive cross-linking, the only fragments observed in addition to the y1 ion are likely an a2 ion at *m/z* 191 and a phenyl fragment at *m/z*120 (Figure credit: Linna An and Wilfred van der Donk).

Previous studies have identified the biosynthetic genes (*dur*) that introduce these PTMs in *Streptomyces cinnamoneus* ATCC12686.[140] The precursor peptide DurA is synthesized with an N-terminal leader peptide (LP), which is removed after maturation of a C-terminal core peptide (CP) (Fig. 4.1a). Serine and threonine residues in the CP region of DurA are dehydrated by DurM to form dehydroalanine (Dha) and dehydrobutyrine (Dhb) residues. Subsequently, DurM catalyzes addition of cysteine thiols onto the β-carbon of Dha or Dhb to produce Lan or MeLan, respectively (Fig. 4.1a). The final two PTMs are hydroxylation of Asp15 by DurX to generate Hya15 and formation of (2*S*,9*S*)-Lal, putatively by DurN-catalyzed addition of Lys19 to Dha6 (Fig. 4.1a).[140] A previous study demonstrated that duramycin lacking Lal does not have antimicrobial activity,[140] indicating that the Lal linkage is critical to set up the PE binding pocket. DurN and its orthologs have no sequence homology with any characterized proteins, their mechanism of Lal formation is not known, and previous attempts to reconstitute activity of orthologs *in vitro* were unsuccessful.[141,142] Instead, prior studies have resorted to non-enzymatic base-catalyzed formation of Lal.[131,141]

In this study, we demonstrate DurN activity *in vitro* and present the co-crystal structures of the protein with a substrate analog and product. We compared Lal formation under enzymatic and non-enzymatic conditions, indicating that only DurN catalyzes stereospecific and diastereoselective Lal formation. Our data suggest that DurN promotes Lal formation via an unusual substrate-assisted mechanism.

4.3 RESULTS

4.3.1 The structure of DurN with a substrate analog

We removed the MBP tag from MBP-DurN (see Methods) and obtained crystal structures of DurN in complex with a substrate analog and in complex with duramycin. We generated an unreactive variant of **3** in which Dha6 (originating from Ser6 in DurA) was mutated to Ala (**3-**Dha6Ala, compound **6**, see Ref. 136 Supplementary Fig. 3a). The structure of DurN with **3-**Dha6Ala was solved to 1.90 Å resolution by single wavelength anomalous dispersion (SAD)

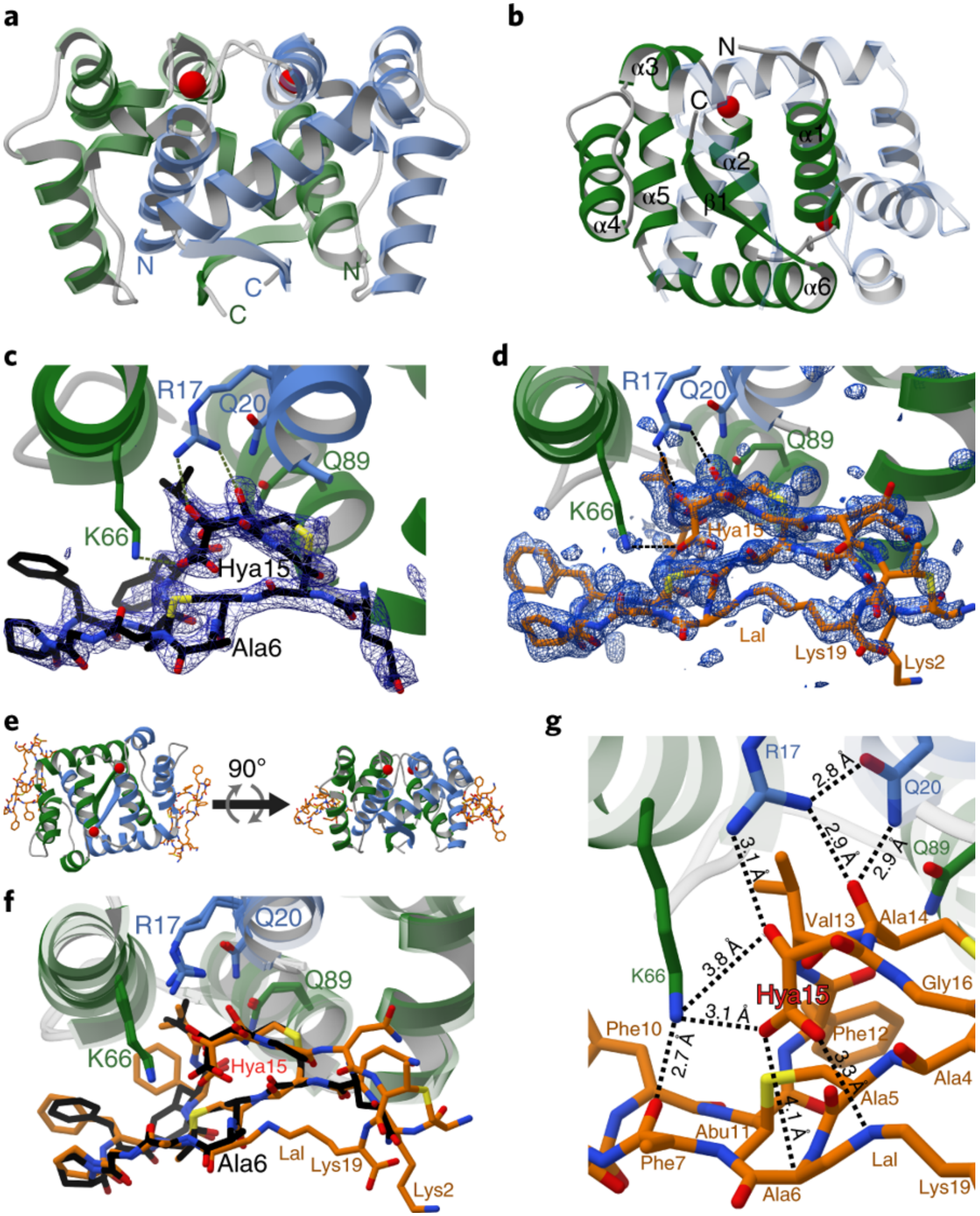


Fig. 4.2: Structural analysis of DurN. a, Apo-DurN with the two monomers represented as green and blue ribbons. Two potassium ions are depicted as red spheres. b, Secondary structure mapped onto the DurN homodimer. For clarity, only the green monomer is labeled. c, DurN bound to 6 (black sticks). A difference Fourier map (Fo—Fc) contoured at 2.5σ is superimposed and shows electron density for portions of the ligand, including one Lan, one MeLan, and Hya15. Critical catalytic DurN residues are represented as green and blue sticks. d, DurN bound to duramycin (orange

Fig 4.2 (cont.) sticks) with a difference Fourier map (Fo—Fc) contoured at 2.5σ revealing occupancy for nearly the entire ligand. e, Two representations of the DurN homodimer featuring two molecules of duramycin related by a 90° rotational translation about the axis, depicted by the black arrow. f, Superimposition of 6 and duramycin-bound structures of DurN. Phe7 in DurA is not modeled in the ligand structures (c, d, and f) because of insufficient electron density. g, Close-up view of the DurN–Hya15 interactions (interatomic distances are calculated averages between two ligand-binding sites comprising chains C and G in PDB 6C0Y). For residue numbering, see Fig. 4.1a.

methods using an iodide-soaked crystal. DurN also crystallized in the apo form and the structure was solved to 2.15 Å by molecular replacement using the SAD-derived coordinates (Fig. 4.2a). The overall structure of DurN consists of an interlaced homodimer where α-helix 1 of each monomer is engaged with residues from α-helices 2-6 in the adjacent monomer. The α-helices are followed by a single β-strand that forms an intermolecular, antiparallel interaction with its neighboring β-strand (Fig. 4.2a,b). A query against the PDB[142] using the DALI server[143] shows that DurN does not confidently resemble any known domain, indicating DurN contains a new protein fold and dimerization mode.

The homodimer contains two structural ions that are related by the pseudo two-fold symmetry axis of the DurN homodimer. Our data are consistent with potassium as the bound metal ions (see Ref. 136 Supplementary Figs. 4-5). DurN coordinates potassium at both sites via an octahedral array of oxygen ligands derived from amides in the backbone (see Ref. 136 Supplementary Fig. 4). In addition to potassium, we also considered the possibility of magnesium ions bound at the ion binding sites and carried out equivalent rounds of refinement in REFMAC5[35] after modeling either potassium or magnesium ions independently. Whereas no positive signal was observed in the difference Fourier map ($F_o - F_c$) at 3σ for the corresponding potassium bound model, a positive difference signal was observed for the magnesium bound model (see Ref. 136 Supplementary Fig. 4c). Additional support for structural potassium ions was gathered by performing differential scanning fluorimetry (DSF), whereby DurN melting curves were obtained in the presence of various potassium or magnesium ion concentrations. While potassium ions conferred an increase in thermal stability of DurN as a function of ion concentration, magnesium ions had no effect on DurN thermal stability (see Ref. 136 Supplementary Fig. 4a,b). Molecular dynamics (MD) simulations also favor a structural potassium ion over magnesium (see Ref. 136 Supplementary Fig. 5), and we needed potassium in our assays for enzymatic activity and to prevent precipitation.

Analysis of simulated annealing difference Fourier maps ($F_o - F_c$) of the apo-structure and the structure with the bound substrate analog revealed electron density for **3**-Dha6Ala from Gln3 to Gly16 (Fig. 4.2c), with the reactive Lys19 invisible due to disorder. Only one of the two symmetric binding sites of the homodimer is occupied by ligand because of occlusion of the other site by residues from the N-terminus of a different protomer. Recognition of the ligand is mediated by residues from α-helix 1 of one polypeptide and α-helices 4-6 of the second monomer. Surprisingly, the ligand bound to the periphery of DurN and Ala6 (corresponding to the Dha6 involved in Lal formation) is far removed from any residues of the enzyme (Fig. 4.2c).

4.3.2 The structure of DurN with duramycin

The observed binding pose of **3**-Dha6Ala does not provide an obvious mechanism of catalysis. We therefore determined the 1.66 Å resolution structure of DurN bound to duramycin. This structure revealed additional electron density in both binding sites of the DurN dimer, enabling modeling of the entire duramycin, including the Lal linkage (Fig. 4.2d,e). Globally, the two ligand-bound structures are highly similar with an RMSD of 0.45 Å over 1400 atoms (including ligands) (Fig. 4.2f). The duramycin-bound structure reaffirms that DurN does not provide any residues that could catalyze the Michael-type addition and suggests that DurN may mediate Lal formation by serving as a molecular scaffold to bring Dha6 and Lys19 in close proximity. Notably, a bidentate hydrogen-bond is observed between the guanidine of Arg17, the Hya15 hydroxyl group, and an amide backbone oxygen of Cys14 of duramycin (Fig. 4.2g). The

Gln20 and Gln89 side chains of DurN further extend the hydrogen-bond network, and an electrostatic interaction is also observed between the carboxylate of Hya15 and Lys66 (Fig. 4.2g). These interactions position one of the carboxylate oxygens of Hya15 at 3.3 Å from the Lal secondary amine nitrogen. Thus, it appears that DurN holds the Hya15 carboxylate of the substrate in a conformation that is poised to activate Lys19 (Fig. 4.3). The other Hya15 carboxylate oxygen is 4.1 Å from the α-carbon of Ala6. This suggests that it would shield the *Si* face of the reactive Dha6, thus promoting protonation by solvent of the enolate formed upon the addition of Lys19 from the more accessible *Re* face, resulting in the correct (2*S*,9*S*)-Lal diastereomer (Fig. 4.3). The structures do not provide any evidence for activation of the carbonyl oxygen of Dha6. In fact, the Ψ angle in Ala6 in duramycin is close to zero, very likely due to electrostatic repulsion between the carboxylate of Hya15 and the amide carbonyl of Ala6, which is directed towards the solvent. This strongly suggests that the reacting Dha6 is fixed in an *s-cis* conformation prior to cyclization, which has been found to be the more reactive form in other Michael-type additions occurring in lanthipeptide biosynthesis.[144]

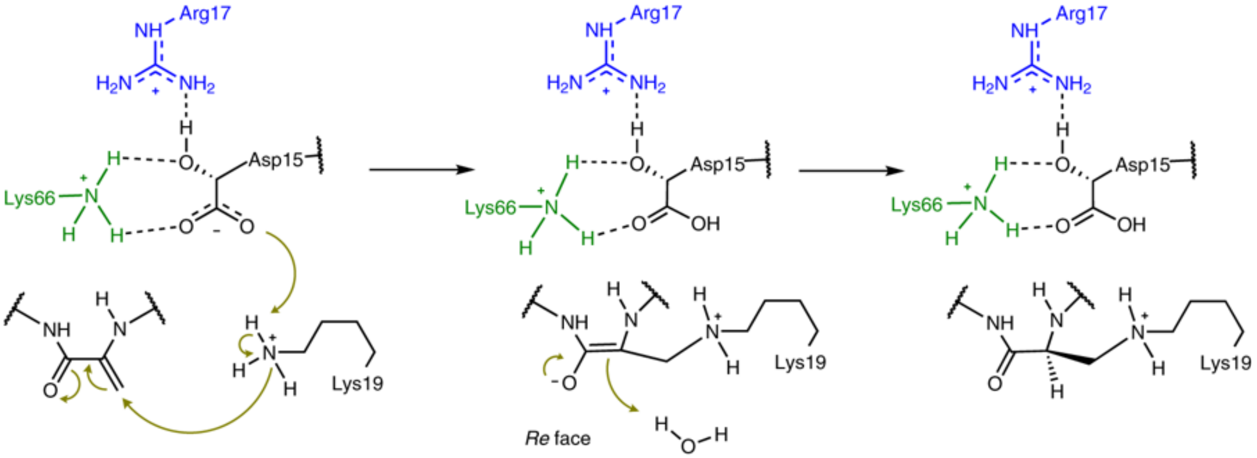


Fig. 4.3: Proposed substrate-assisted Lal formation catalyzed by DurN. Features in blue and green correspond to residues coming from the two monomers of DurN (see Fig. 4.2) and those in black correspond to its substrate/product (Figure credit: Linna An and Wilfred van der Donk).

4.4 DISCUSSION

In this study, we provide the first in vitro demonstration that DurN catalyzes Lal formation, a step that had been recalcitrant in previous studies.[131,141] By comparing enzymatic and nonenzymatic Lal generation, our data suggest that DurN greatly facilitates the reaction and controls its stereoselectivity. Furthermore, our data shows that the enzyme does not require the leader peptide for catalysis. DurN has a unique fold and dimeric architecture, and in co-crystal structures, the Lal is entirely solvent exposed and the Hya15 of the substrate is positioned to potentially act as a base during catalysis to deprotonate Lys19 to initiate the Michael-type addition. Additionally, Hya15 appears to shield the *Si* face of the enolate intermediate such that protonation will only occur from the *Re* face, resulting in the correct (2*S*,9*S*)-Lal product. Mutational studies show that substrates that lack either the hydroxyl or carboxylate groups of Hya15 are not processed by DurN despite maintaining binding affinity, suggesting that both of these groups on the substrate are indispensable for catalysis. Merging the biochemical findings with structural and computational insights points to a mechanism whereby the substrate employs its own Hya15 as a catalytic base to facilitate a Michael addition and stereo selective enolate protonation, resulting in Lal formation between Lys19 and Dha6.

The biosynthetic gene clusters of lysinoalanine-containing natural products such as duramycin and cinnamycin are very similar (see Ref. 136 Supplementary Fig. 12a) and all encode a DurN-like protein. This relatively small protein family is comprised of similar proteins, as shown by a tight sequence similarity network (SSN) of members identified by a BLAST search with DurN as query (see Ref. 136 Supplementary Fig. 12b).[3] All identified homologs belong to duramycin-like lanthipeptide biosynthesis clusters that contain a type II lanthionine synthetase-like protein, a duramycin-like LanA precursor peptide, and a DurX-like protein. The natural products for some of these clusters have been characterized, and some of the LanN proteins (CinN, MarN) have been demonstrated to be important for Lal installation in vivo.[129,131,140,141,145] Because all of the LanA precursor peptides have a conserved Asp and a Lys close to the C terminus, it is likely that the substrate-assisted mechanism reported here is general for Lal formation in the biosynthesis of duramycin-like lanthipeptides.

4.5 METHODS

4.5.1 Strains and plasmids

E. coli DH5α was used as host for cloning and plasmid propagation and *E. coli* BL21 (DE3) as host for expression of proteins/peptides. pRSFDuet-1, pACYCDuet-1, and pETDuet-1 were obtained from Novagen. The pET His_6 small ubiquitin-like modifier (Sumo) tobacco etch virus (TEV) protease LIC cloning vector (2S-T) was a gift from S. Gradia (Addgene plasmid #29711).

4.5.2 Production and purification of MBP-DurN and its mutants

E. coli BL21 (DE3) cells were transformed with pET28a/MBP-DurN (or mutants). Single colonies of chemically competent *E. coli* BL21(DE3) were grown in Luria-Broth (LB) media supplemented with kanamycin (50 µg/mL). A 5-mL starter culture was grown overnight and used to inoculate 1 L of LB media supplemented with kanamycin. Liquid cultures were grown at 37 °C with shaking (200 r.p.m.) and induced with 0.15 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) when the OD_{600} reached 0.75, which was followed by a 20 min ice-bath incubation and further shaking for 20 h at 18 °C and 200 r.p.m. Cells were harvested by centrifugation at 4 °C, followed by suspension of the pellet in ~30 mL of binding buffer (50 mM Tris pH 7.5, 0.5 M NaCl, 2.5% (v/v) glycerol, 0.1% (v/v) Triton X-100) and lysis by sonication. The lysates were clarified by centrifugation at 4 °C. The supernatant was loaded onto a 5 mL amylose resin pre-equilibrated with binding buffer. The column was washed with 50 mL of wash buffer (50 mM Tris pH 7.5, 0.5 M NaCl, 2.5% (v/v) glycerol) and eluted with 20 mL of elution buffer (50 mM Tris pH 7.5, 0.15 M NaCl, 10 mM maltose, 2.5% (v/v) glycerol). Fractions containing the highest purity protein as judged by SDS–PAGE were pooled and further purified by size-exclusion chromatography (SEC) using storage buffer (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 50 mM NaCl, 2.5% (v/v) glycerol). The purified protein was concentrated using Amicon Ultra-4 centrifugal filters (30 kDa molecular weight cut-off (MWCO); Millipore Sigma) and frozen in liquid nitrogen for storage at −80 °C. Protein concentration was quantified by NanoDrop (Thermo Fisher Scientific) using an extinction coefficient of 83,070 M^{-1} cm^{-1} at 280 nm.

4.5.3 Tag removal of MBP-DurN and its mutants

Tobacco etch virus (TEV) protease was added to MBP-tagged protein at a 1:10 (w/w) ratio in digestion buffer (50 mM HEPES pH 7.5, 300 mM KCl, 50 mM NaCl, 2.5% (v/v) glycerol), and the reaction was left at 4 °C overnight. The sample was loaded onto 5 mL of HisTrap resin pre-equilibrated with digestion buffer and washed with 10 mL of the same buffer. Tag-free proteins were eluted with elution buffer (10 mM imidazole, 50 mM HEPES pH 7.5, 300 mM KCl, 50 mM NaCl, 2.5% (v/v) glycerol). Fractions containing the highest purity protein as judged by SDS–PAGE were pooled and concentrated using Amicon Ultra-4 centrifugal filters (10 kDa MWCO, Millipore Sigma) to ≤5 mL and injected onto a 120 mL Superdex 200 10/300 GL column (GE Healthcare) to purify by SEC using 300 mM KCl, 20 mM HEPES pH 7.5 as running buffer.

4.5.4 Analytical size-exclusion chromatography

Following amylose-resin and size-exclusion chromatography, N-terminal MBP-tagged DurN variants were concentrated to ≥5 mg/mL before injecting 0.5 mL onto a 30 mL Superdex 200 10/300 GL column (GE Healthcare) equilibrated with 0.1 M KCl, 20 mM HEPES, pH 7.5. A similar running buffer was used to elute proteins at a flow rate of 1 mL/min. N-terminal Sumo-tagged and tag-free DurN variants were analyzed similarly following their purification by Ni-affinity and size-exclusion chromatography. The retention times of DurN variants were compared to those of commercially available protein standards: carbonic anhydrase (bovine erythrocytes; 29 kDa), albumin (bovine serum; 66 kDa), and alcohol dehydrogenase (yeast; 150 kDa) purchased from Sigma-Aldrich and analyzed using a similar procedure. Overlaid chromatograms for each DurN variant suggest similar homodimeric states (see Ref. 136 Supplementary Fig. 11).

4.5.5 Protein crystallization

Pooled fractions from MBP tag removal and purification of MBP-DurN were concentrated to ≤5 mL and injected onto a 120 mL Superdex 200 10/300 GL column (GE Healthcare) to purify by SEC using 300 mM KCl, 20 mM HEPES pH 7.5 as running buffer.

4.5.5.1 DurN (apo)

Following SEC, DurN was concentrated to 7 mg/mL and crystallized in 2 μL hanging drops at 16 °C, whereby DurN solution was mixed 1:1 (v:v) with a reservoir solution containing 25% polyethylene glycol (PEG) 3350, 0.2 M magnesium chloride, and 0.1 M Bis–Tris pH 6.5. Apo DurN also produced similar diffraction quality crystals using a different reservoir solution containing 20% PEG 8000, 0.2 M magnesium acetate, and 0.1 M sodium cacodylate pH 6.0. In both cases, rod-shaped crystals grew to a maximum size within 48 h.

4.5.5.2 DurN + compound ***6***

Following SEC, DurN was concentrated to 9 mg/mL and incubated with 1.8 mM compound **6** for 30 min on ice. This solution was mixed 1:1 (v:v) with a reservoir solution containing 25% PEG 3350, 0.2 M $MgCl_2$, and 0.1 M Bis–Tris pH 6.0 in 2 µL hanging drops incubated at 16 °C. Hexagonal shaped crystals grew to maximum size within 48 h.

4.5.5.3 DurN + duramycin

Following SEC, DurN was concentrated to 6 mg/mL and incubated with 1.5 mM duramycin for 30 min on ice. Insoluble particulates were pelleted by centrifugation and the supernatant was mixed 1:1 (v:v) with a reservoir solution containing 22.5% PEG 3350, and 0.1 M 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 5.6 in 2 µL hanging drops at 16 °C. Thick rod-shaped crystals grew to a maximum size within 72 h.

4.5.6 Crystallographic data collection, structure solutions, and refinement

Data were collected at the Advanced Photon Source at Argonne National Lab using the Life-Science Collaborative Access Team 21-ID-D, 21-ID-F, and 21-ID-G beamlines. Before collecting data, crystals were cryoprotected by immersing the crystals in similar solutions supplemented with 20% ethylene glycol and then flash-freezing in liquid nitrogen. To obtain phases experimentally, co-crystals of DurN and compound **6** were soaked with 0.5 M potassium iodide for 24 h at 16 °C and used to collect diffraction data at 8 keV. Data were initially processed using autoPROC and further processed in autoSHARP to obtain phase information.[71,146] The initial model was rebuilt using Buccaneer and refined using REFMAC5 in the CCP4 Software Suite.[35,74,76] Diffraction data sets corresponding to apo-DurN crystals and DurN-duramycin co-crystals were phased after similar processing in autoPROC followed by molecular replacement using Phaser MR.[73,111] Ligand parameters for compound **6** and duramycin were determined using Phenix eLBOW and modeled into their corresponding electron densities, followed by iterative rounds of automated REFMAC5 refinement and manual rebuilding in Coot.[36,37] Solvent molecules and potassium ions were first incorporated using Phenix Refine or REFMAC5 then manually curated in Coot.[38]

4.5.7 Differential scanning fluorimetry (DSF)

Tag-free DurN purified by SEC was solvent exchanged into a buffer lacking potassium or magnesium ions by subjecting to additional SEC on a 30 mL Superdex 200 10/300 GL column (GE Healthcare) equilibrated with 0.25 M NaCl and 20 mM HEPES pH 7.5. Protein was eluted using a similar running buffer, and pooled fractions were concentrated to 1 mg/mL. SYPRO orange dye was added to the protein at 8x final concentration and incubated at r.t. for 20 min. High-Profile 96-Well Semi-Skirted PCR plates (Bio-Rad Laboratories, Inc.) were used to serially titrate either KCl or $MgCl_2$ into the protein-dye mixture from 300 mM to 150 μM or 200 mM to 100 μM, respectively, and incubated for 25 min at 22–25 °C. Melting curves were obtained using a CFX Connect Real-Time PCR Detection System (Bio-Rad Laboratories, Inc.) for which the sample was initially held at 25 °C for 3 min and then incrementally ramped to 98 °C at 0.5 °C every 10 s with plate-reading steps at every increment. The channel was set to "FRET" to satisfy the excitation and emission wavelengths of the SYPRO orange dye (λ_{ex} 470 nm /λ_{em} 570 nm). Data corresponding to the melt curves and first derivatives of the melt curves were obtained in the Bio-Rad CFX Manager 3.1 Data Analysis tool and subsequently plotted in OriginPro 2016 (OriginLab).

4.5.8 Mutational studies of DurN

A number of residues make hydrophobic contacts at the dimeric interface (i.e., Leu22, Val24, Leu25, Leu28, Leu39, Leu43, Leu51, Ile54, Trp68, Leu72, Trp73, Leu85, Ile111, and Leu113). Additional electrostatic interactions are observed intermolecularly between Asp12–Lys103 and Arg18–Glu106, as well as intramolecularly between Lys34–Glu37 and Arg53–Glu82. Several MBP-tagged and site-directed DurN variants were constructed based on the co-crystal structures and sequence conservation between DurN homologs (see Ref. 136

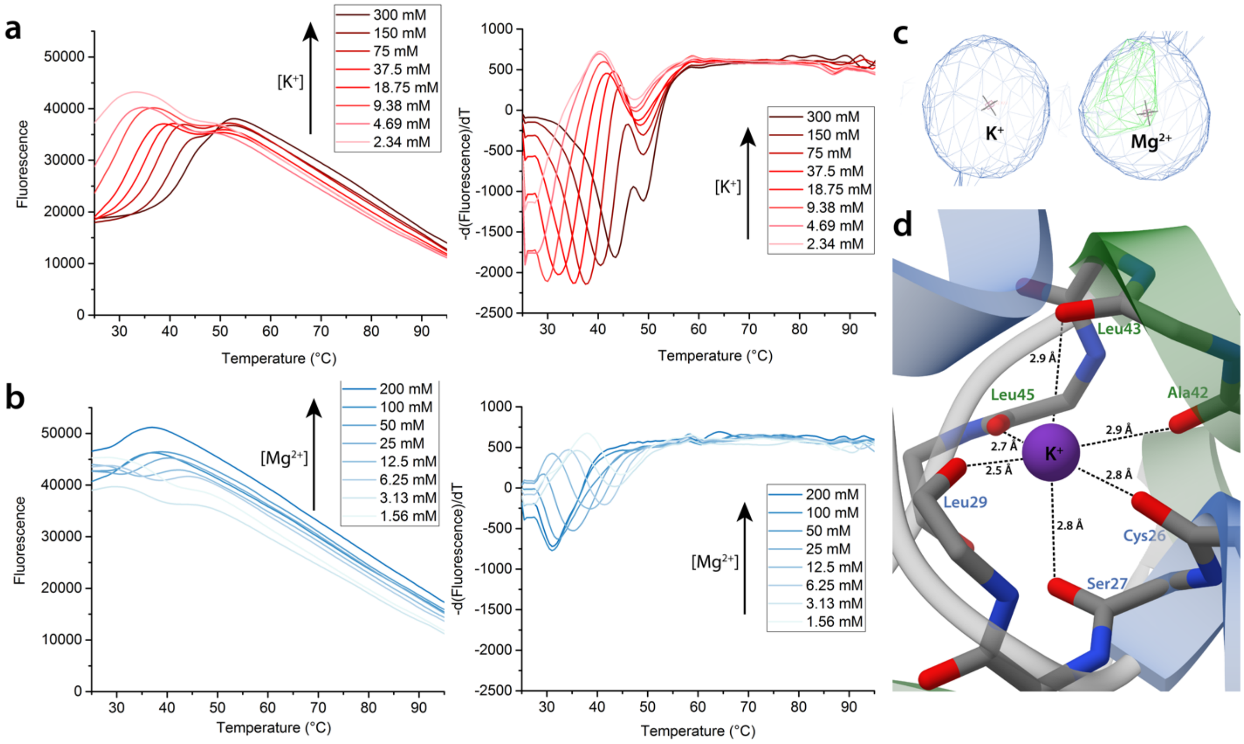


Fig. 4.4: Melt curves, negative first derivatives of the melt curves obtained by differential scanning fluorimetry (DSF), and potassium coordination by DurN. a, DSF data corresponding to DurN in the presence of increasing potassium ion concentrations (2.34 mM – 300 mM). The dose-dependent stabilization of DurN by potassium ions is made evident by the increasing melting temperature as a function of potassium ion concentration (i.e. right-shifting inflection point in the left panel and minima in the right panel). b, Similar DSF data corresponding to DurN in the presence of increasing magnesium ion concentration (i.e. 1.56 mM – 200 mM). No appreciable difference in the melt curves and persistence of high initial fluorescence across all magnesium ion concentrations suggests that DurN thermal stability is not influenced by magnesium ion concentration. c, 2Fo – Fc maps (blue) at 1.5σ and positive Fo – Fc maps (green) at 3σ of one structural ion binding site when potassium or magnesium ions are modeled and refined in REFMAC5.[35,76] d, Octahedral potassium ion coordination by DurN involving amide carbonyl oxygen ligands from the DurN backbone of both monomers color-coded green or light blue. Experiments were performed once but are supported by potassium requirement for activity and stability and MD simulations.

Supplementary Fig. 12). Specifically, Ala substitutions were generated at Arg17, Arg18, Gln20, Cys26, Lys66, Trp68, Ser79, Gln89, and Glu101 in addition to a Lys substitution at Arg17, and an end-point substrate conversion assay using compound **3** was performed (see Methods). Aside from DurN variants C26A and S79A, a diminution in activity relative to the wild-type DurN was observed for all DurN variants. These activity data are rationalized by the structural data, as residues Arg17, Gln20, and Lys66 are implicated in substrate binding, and residues Arg18, Trp68, and Glu101 are important for dimerization. Unsurprisingly, mutations at Cys26 and Ser79 exhibit no activity loss, as these residues are solvent exposed and distal to the ligand-binding site in the crystal structures.

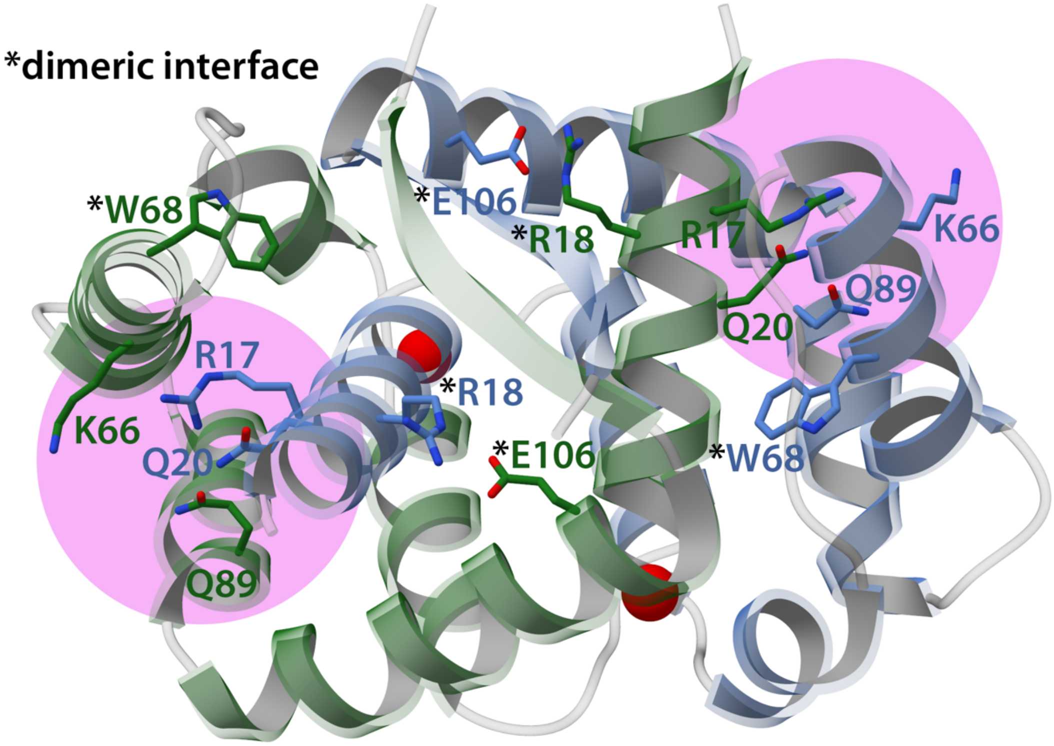


Fig. 4.5: DurN residues selected for site-directed mutagenesis. DurN activity is abolished upon mutating the selected residues (drawn as sticks) to Ala (i.e. R17A, R18A, Q20A, K66A, W68A, Q89A, and E106A). Loss in activity is attributed either to compromising substrate recognition or disruption of the homodimer interface. Residues involved in intermonomer contacts at the dimer interface are highlighted with an asterisk. A magenta circle is drawn to encapsulate residues important for substrate recognition. Potassium ions are represented as red spheres.

CHAPTER 5: STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF AN OFF-LOADING DIECKMANN CYCLASE IN NOCAMYCIN BIOSYNTHESIS

5.1 ABSTRACT

A recently identified family of stand-alone off-loading enzymes found in hybrid polyketide synthase/non-ribosomal peptide synthetase (PKS/NRPS) biosynthetic gene clusters catalyzes a Dieckmann condensation to afford five and six membered tetramate and pyridone heterocycles, respectively. While several tetramate containing natural products have been described, structural and mechanistic information on the enzymatic transformations that impart this functionality has largely been limited. Here we report the first structure of one such Dieckmann cyclase, NcmC, from nocamycin biosynthesis in its apo form (1.6 Å) and covalently bound to the epoxide containing inhibitor and substrate-mimic cerulenin (1.6 Å) guiding additional structure-based mutagenesis and activity analyses. The overall fold of NcmC closely resembles off-loading thioesterase and macrolactone synthase domains. Notably, only subtle active site changes are necessary to convert the Dieckmann cyclization activity to thioester hydrolysis. Bioinformatic analysis of NcmC and its homologs, combined with docking and modeling, identifies likely candidates for heterocycle formation in underexplored gene clusters and suggests a structural basis for substrate selectivity and additional engineering experiments.

5.2 INTRODUCTION

Natural products containing a core tetramic acid (pyrrolidine-2,4,-dione) or pyridone heterocycle are found throughout all ecology, and have been purified from marine and terrestrial species including cyanobacteria, actinobacteria, fungi, and sponges (Fig. 5.1A). While pharmacological profiles of tetramate-based natural products reveal a wide breadth of bioactivities including antiviral, antimicrobial, and antineoplastic properties, knowledge of the detailed

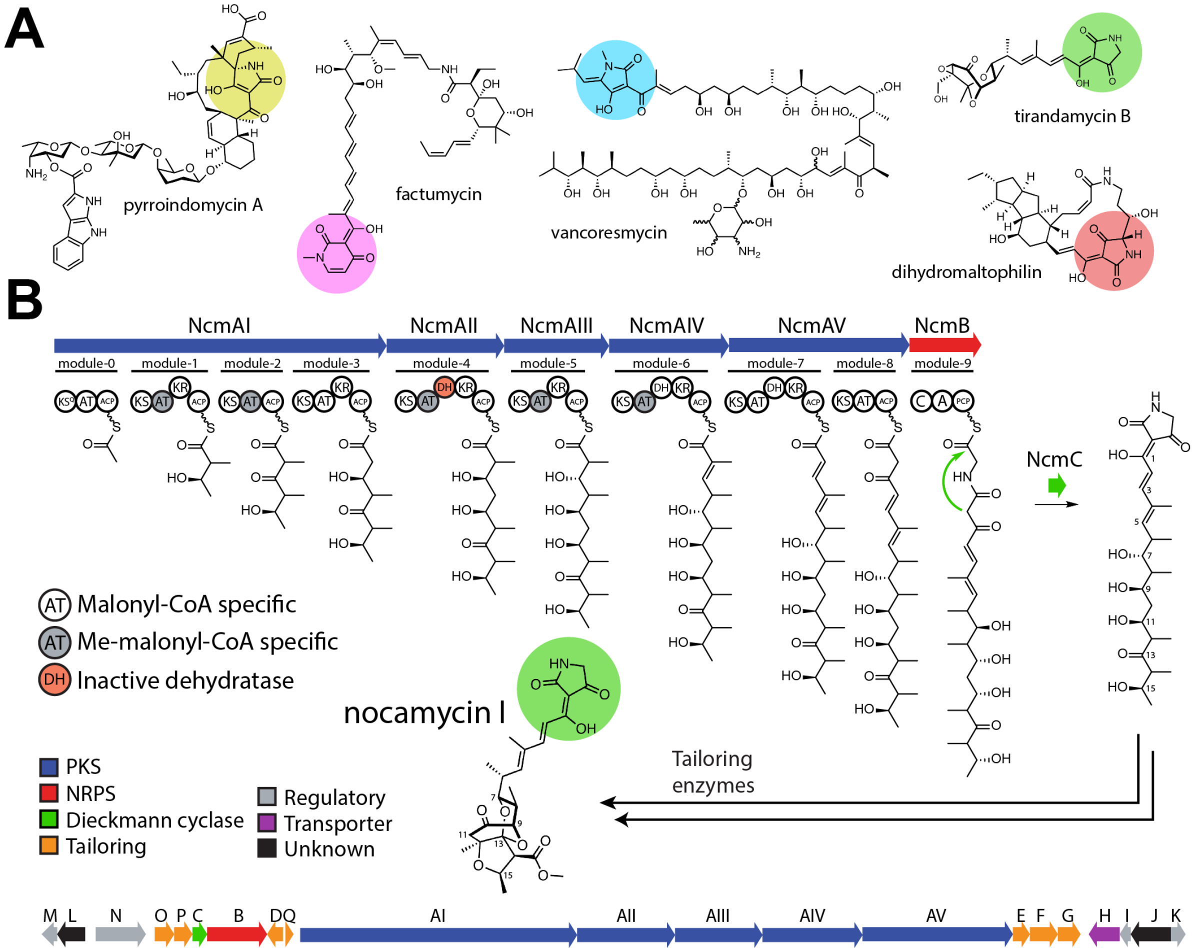


Fig. 5.1: Tetramate and pyridone containing natural products and nocamycin biosynthesis. (A) Representative set of bacterial natural products containing tetramate and pyridone heterocycles. (B) PKS/NRPS domain assembly in nocamycin biosynthesis. Stereochemical assignments for intermediate structures are based on bioinformatic predictions using the specificity conferring motifs of ketoreductase (KR) domains (57). The dehydratase (DH) domain from module 4 (DH4) is considered to be catalytically inactive in accordance with retention of the oxygen from C9 in nocamycin I and prior studies on the related natural product tirandamycin (58). Following NcmC catalyzed off-loading by Dieckmann condensation, a variety of tailoring enzymes (i.e. NcmDEFGOPQ) are putatively involved in formation of nocamycin I. The overall gene organization of the nocamycin BGC is drawn below (~61 kb).

biosynthetic steps towards such molecules remains limited.[147–157] In general, these scaffolds are secondary metabolites assembled from polyketide and α/β-amino acid precursors by the concerted actions of type I polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) machines. In most cases, the carbon skeleton is assembled from acetyl- and malonyl-CoA precursors by the actions of the PKS, which in fungi are often iterative and highly reducing. In tetramate and pyridone containing natural products, the NRPS module then appends the α- or β-amino acid to generate an *N*-β-ketoamide aminoacyl thioester. The mechanism for the subsequent carbon-carbon bond forming reaction necessary to produce the lactam ring has only been elucidated for a small number of the ~300 substituted tetramate and ring-expanded pyridone secondary metabolites isolated to date.[155,158–160]

There are currently four biosynthetic routes that have been characterized for the installation of the carbon-carbon bond that elaborates the tetramate core. Biochemical reconstitution studies of equisetin biosynthesis illustrates that off-loading of the enzyme-bound thioester intermediate from the fungal PKS-NRPS hybrid is catalyzed by an R domain (R*) that resembles redox-active NAD(P)H-dependent modules but carries out a cofactor-independent Dieckmann condensation to produce the pyrrolidine-2-one.[161] Analogous non-reductive R* domains install the tetramic acid in other fungal metabolites fusaridione A and cyclopiazonic acid.[149,162] A second route towards tetramate formation is orchestrated by the terminal thioesterase (TE) in bacterial PKS-NRPS hybrids. These modules catalyze carbon-carbon bond formation reactions via Dieckmann condensation in contrast to hydrolytic release, as demonstrated for the polycyclic tetramate macrolactams frontalamides A and B, dihydromaltophilin, and ikarugamycin.[155,163–166] The TEs are conceptually analogous to the fungal Claisen cyclase modules from non-reducing PKSs that catalyze the formation of aromatic polyketides by sequestering the substrate in a hydrophobic pocket that favors cyclization over hydrolysis.[167–169] A third biosynthetic route consists of two proteins that catalyze tetramate formation in the biosynthesis of spirotetramate-containing pyrroindomycins through a yet to be established mechanism.[170]

Examination of the putative pathways for actinomycete-derived metabolites bearing tetramic acid or pyridone moieties failed to identify either an off-loading TE or homologs of the R* domains. Bioinformatics analysis identifies an open reading frame adjacent to the PKS-NRPS in the tirandamycin pathway that encodes for a ~280 residue polypeptide, and homologous genes could be located directly upstream of NRPS genes in the assembly lines of several 3-acyltetramate/pyridone biosynthetic gene clusters (BGCs) including those encoding for the biosynthesis of streptolydigin, α-lipomycin, factumycin, and kirromycin (2, 25-30). Enzymatic reactions using recombinant proteins and *N*-acetylcysteamine thioesters of *N*-acetoacetyl-Gly, -Ala, or -β-Ala as substrate surrogates resulted in conversion into the tetramic acid or pyridone scaffolds establishing these polypeptides as bona fide Dieckmann cyclases (DCs).[171] While sequence based bioinformatic searches failed to reveal sequence or structural similarities to characterized genes in the NCBI or UniProt databases, hidden Markov Model profiling detected similarities to TE domains. Notably, the active site catalytic triad Ser of canonical TEs is replaced with a Cys in the DCs and the Cys→Ser replacement did not demonstrate the anticipated hydrolytic activity.

Utilizing the primary sequences of these DCs as a query seed, we identified a putative biosynthetic pathway for a tetramate or pyridone natural product in *Saccharothrix syringae* NRRL B-16468. Fermentation and tandem mass spectrometry (MS/MS) analysis of culture supernatants and methanolic extracts identified nocamycin I as the product, consistent with the organization of the BGC (Figs. 5.1B and 5.6-5.7). A recent study also identified the nocamycin BGC using a cosmid library approach and corroborates our fermentation analysis.[172] In order to elucidate the mechanistic basis for Dieckmann cyclization, we characterized the stand-alone catalyst for tetramate formation from this BGC using chemical and structural biological approaches. We further demonstrate that the mycotoxin cerulenin modifies NcmC through formation of a covalent adduct with the active site Cys89, consistent with the involvement of this residue in covalent catalysis. The cocrystal structure reveals that cerulenin binds to NcmC as a cyclic hemiaminal, providing a plausible mimic of a tetramate-like reaction intermediate. Combining the structural data with product profile analysis and docking studies points to residues which may be involved in acid-base catalysis, as well as solvent shielding to facilitate carbon-carbon bond formation. Notably, a single amino acid change (Q128A) is sufficient to partially convert this Dieckmann cyclase enzyme into a TE. Structure-based modeling of NcmC homologs offers a potential explanation for substrate selection, which may guide future engineering efforts.

5.3 RESULTS

5.3.1 In vitro characterization of the Dieckmann cyclase NcmC

The location of the *ncmC* gene immediately adjacent to the NRPS module in the nocamycin BGC, and the high sequence identity (45%) shared with the biochemically established DC (TrdC) from the tirandamycin biosynthetic cluster suggests that NcmC likely catalyzes Dieckmann condensation to facilitate off-loading of the hybrid PK-NRP metabolite.[171] Given the biosynthetic logic and structure of nocamcyin, we reasoned that the NRPS adenylation (A) domain should select for a Gly substrate for adenylation and thiolation in vivo, and the natural substrate of NcmC should be a PCP-linked PK-NRP thioester (Fig. 5.1B). We used a minimal substrate mimic, *N*-acetoacetyl-glycl-SNAC (**1**), to measure NcmC activity in vitro, as performed previously (Figs. 5.2A and 5.8-5.16).[161,171] Following incubation of **1** in the presence or absence of wild-type NcmC, an enzyme-dependent rate enhancement was observed by liquid chromatography-mass spectrometry (LC-MS) analysis, demonstrating that NcmC catalyzes Dieckmann condensation in vitro (Fig. 5.17).

5.3.2 Overall structure of NcmC

To gain further insights into the mechanism of the enzymatic Dieckmann condensation, we determined the crystal structure of NcmC to 1.6 Å resolution, using phases calculated from data

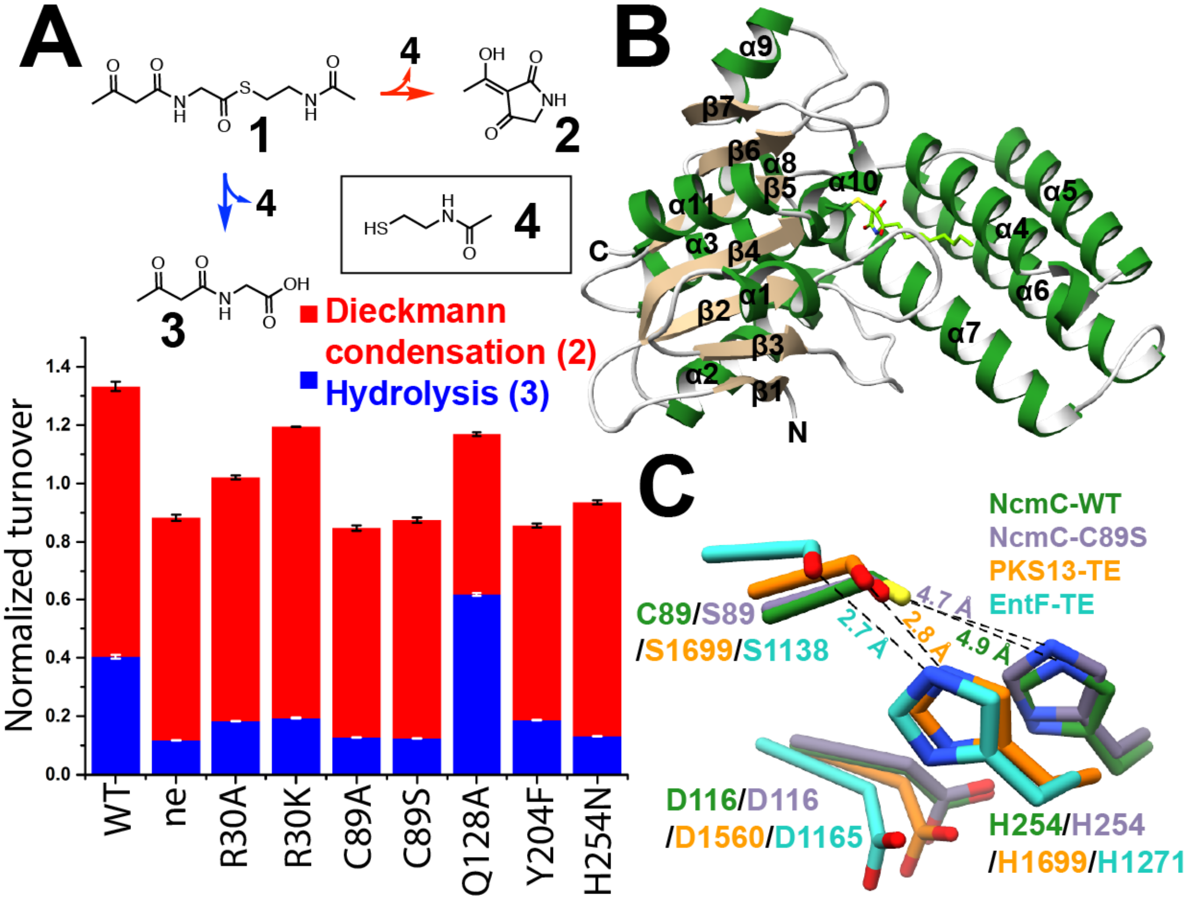


Fig. 5.2: Structure and activity of NcmC. (A) Reaction scheme for enzymatic or non-enzymatic Dieckmann condensation (red) or hydrolysis (blue) of **1**. Normalized **1** turnover determined by quantitative LC-MS is plotted for eight NcmC variants and the non-enzymatic reaction (ne). Error bars represent standard deviation (n=3). (B) NcmC-cerulenin cocrystal structure with secondary structural elements mapped onto the ribbon diagram (PDB 6E6T). (C) Comparison of the DC and TE catalytic triads. NcmC is superimposed with TEs from EntF (PDB 3TEJ) and PKS13 (PDB 5V3W).[173,174]

collected on a mercurial derivative. The overall structure contains two subdomains consisting of an α/β hydrolase fold (composed of residues Arg5-Ser119, and Val214-Ala272) with a four-helical bundle (composed of residues Thr121-Ala210) inserted between strands β5 and β6 (Figs. 5.2B and 5.18). Surprisingly, the α/β hydrolase fold subdomain resembles that found in chain termination TEs from bacterial NRPSs, PKSs, and the multi-subunit mammalian fatty acid synthase complex, despite the lack of any appreciable sequence similarity with these other proteins. The four-helical bundle subdomain is unique to the DCs and is not a feature common to any of the TEs. A DALI search against the Protein Data Bank identifies various off-loading domains as the closest structural homologs, due to the similarities in the α/β hydrolase fold subdomain, and examples include the termination module from the NRPS AB3403 (PDB 4ZXH; Z-score of 17.0; RMSD of 3.4 Å over 215 aligned Cα atoms),[175] the TE domain from NRPS EntF, which catalyzes macrolactonization (PDB 3TEJ; Z-score of 16.7, RMSD of 3.7 Å over 207 aligned Cα atoms),[173] the termination TE domain from PKS13 (PDB 5V42; Z-score of 15.9 RMSD of 4.0 Å over 219 aligned Cα atoms),[174] and the TE domain from the human fatty acid synthase (hFAS) (PDB 1XKT; Z-score of 15.7 RMSD of 3.7 Å over 217 aligned Cα atoms).[143,176] While each of these subdomains contain an insertion between the equivalent strands β5 and β6, the helical subdomains themselves are structurally divergent and cannot be accurately aligned with each other (Fig. 5.19). Prior structural studies of the hFAS-TE domain in complex with the inhibitor orlistat led to the proposal that this helical subdomain is responsible for recognition of the aliphatic tail region of its substrate palmitoyl-ACP.[177]

The α/β hydrolase fold domain houses residues that are shown to function as the catalytic triad for the cyclization reaction, and consists of Cys89, Asp116, and His254. In canonical TE domains, the Asp/His pair acts to lower the pKa of the nucleophilic Ser in order to promote reactivity of the alkoxide towards its thioester substrate, and the proximity of the catalytic His serves this function. For example, the distance between the catalytic His2481 and Ser2308 in the TE domain of hFAS is 2.7 Å, and His1699 and Ser1533 are 2.8 Å apart in the PKS13 TE.[174] In contrast, the distance between His254 and Cys89 is 4.9 Å in NcmC, and the orientation of the His254 imidazole side chain is sub-optimal, relative to the thiol of Cys89. Crystallographic analysis of the C89S variant reveals a comparable separation of 4.7 Å between Ser89 and the His254 imidazole side chain, suggesting a sulfur to oxygen swap is not sufficient to recapitulate the triad geometry observed in catalytic Ser-containing TEs (Fig. 5.2C). Notably, Cys89 is located at the N-terminus of helix α3 and the α-helical dipole may be sufficient to lower the pKa of the thiol to promote reactivity (Figs. 5.2B and 5.18). Prior studies on elongation β-ketoacyl-ACP synthase III (FabH) support a role for the helix dipole in tuning the pKa of the nucleophilic Cys, and experiments with model peptides show a pKa decrease of 1.6 for Cys residues positioned at the N-terminus of an α-helix.[28]

5.3.3 Structure of a covalently bound mimic of the cyclization reaction

With the goal of visualizing NcmC with a bound substrate, we generated the Cys89→Ala and Ser variants and incubated crystals of each with **1** but were unable to observe convincing electron density corresponding to features of **1**. As the catalytic triad and active site elements were reminiscent of features observed in structures of fatty acid biosynthetic enzymes, we tested a panel of small molecule fatty-acid biosynthesis inhibitors against NcmC. Surprisingly, cerulenin, a mycotoxic, covalent inhibitor of β-ketoacyl-ACP synthase (KAS)[178] formed a complex with NcmC as detected by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS)

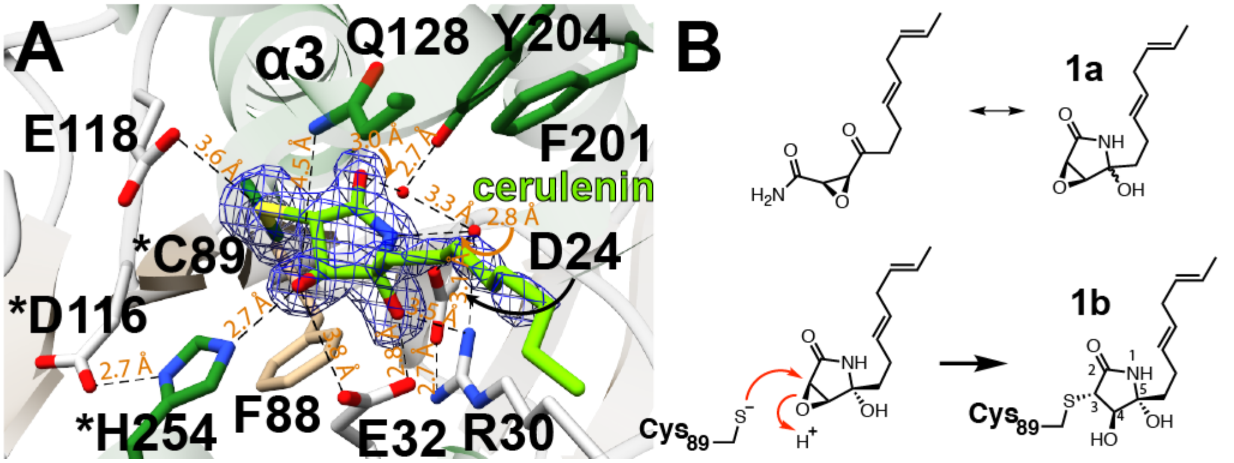


Fig. 5.3: Covalent modification of NcmC by cerulenin. (A) Close-up view of the cerulenin-modified active site of NcmC with residues targeted for mutagenesis represented as sticks (residues derived from looped regions in white and helical regions in green). A difference Fourier map (F_{obs} - F_{calc}) contoured at 2σ is shown for Cys-cerulenin in blue mesh. Catalytic triad residues are denoted with an asterisk. (B) Solution $^{1}H/^{13}C$ NMR previously demonstrated cerulenin prefers its cyclic hemiaminal form (**1a**) in polar protic solvents.[185] NcmC is proposed to bind the cyclic form (**1a**) in aqueous buffer with stereospecific ring opening to afford the product (**1b**) supported by the crystal structure.

(Fig. 5.20), and crystals of the covalent complex diffracted to 1.6 Å resolution (Fig. 5.3A). The cerulenin mycotoxin contains a 12-carbon acyl chain that mimics a fatty acyl chain substrate and extends into the hydrophobic pocket of various KASs.[178–184] It also contains a reactive epoxide that undergoes ring opening by a nucleophilic Cys to form a covalent complex.

By crystallographic observations, NcmC reacts with cerulenin stereospecifically to give the (3*S*,4*R*,5*R*) cyclic hemiaminal product (Fig. 5.3B). The stereospecificity is achieved as a result of favorable hydrophilic interactions occurring between the 4*R*-hydroxyl and Glu32/His254 where Phe88 is situated at the opposite face engaged in hydrophobic interactions with the C4-hydrogen. Similarly, specificity for the 5*R* position is established by H-bonding between Arg30 and the 5*R*-hydroxyl, and favorable hydrophobic interactions take place along the aliphatic tail between side chains of Phe201 and the methylene groups of Arg30 (Fig. 5.3A).

5.3.4 Structure-guided mutagenesis and end-point analysis of NcmC

To understand the contribution of individual residues to catalysis, we generated site-directed variants of NcmC residues based on sequence conservation and their proximity to the active site. NcmC variants were incubated with **1** for 1.5 h at 28 °C before quenching the reactions and quantifying amounts of **3** and **4** by LC-MS using authentic standards. Assuming that production of **4** results only from one of two mechanisms (i.e. hydrolysis or Dieckmann condensation) the amount of tetramate product (**2**) was taken to be equal to **4** – **3** (Fig. 5.2A). Production of **4** could also result from the formation of a covalent enzyme-bound intermediate, limiting accurate quantitation of **2** by this assumption. However, long incubation times ensured that near steady-state conditions were achieved as monitored by a spectrophotometric 2,6-dichlorophenolindophenol (DCPIP) coupled assay (Fig. 5.21), thus the quantity of any covalent intermediate is likely to be very low.[186,187] Consistent with the proposed role in hydrogen-bonding interactions, Arg30 shows a high degree of conservation across all 98 NcmC homologs identified by NCBI BLAST.[188] Consequently, the R30A variant resulted in a 56% decrease in enzymatic **2** formation. In the crystal structures, Arg30 forms an electrostatic interaction with Asp24, and an Nη of the guanidium is involved in H-bonding with water and cerulenin (Fig. 5.3A). Surprisingly, the R30K variant, which would be expected to maintain the electrostatic environment but with impaired H-bonding interactions, resulted in a 42% increase in enzymatic **2** production with a 10% decrease in total turnover (including hydrolysis). This suggests that Arg30 may play more of a structural role and maintenance of the electrostatic interaction may help forge the integrity of the substrate binding site. Both mutations at Cys89 (i.e. C89A and C89S) resulted in **2** production commensurate with the non-enzymatic reactions, supporting that Cys89 is critical for the enzyme catalyzed reaction.

Notably, the Q128A mutation results in a 74% increase in the hydrolytic product (**3**), relative to wild-type NcmC (5.2-fold increase in **3** relative to non-enzymatic hydrolysis). In the cocrystal structure, Gln128 is situated adjacent to the C-S bond formed in the epoxide ring opening reaction with NcmC, and there are no bridging water molecules in between, suggesting that Gln128 serves to shield the enzyme-bound thioester intermediate against solvent hydrolysis. Similar to the C89A/C89S variants, the Y204F variant resulted in **2** formation comparable with the reaction lacking enzyme, indicating its loss in enzymatic DC activity. Considering the structurally observed H-bond network between the side chain of Tyr204, water, and cerulenin, the high degree of sequence conservation (conserved across 67 sequences identified by BLAST), and the apparent importance of Tyr204 for tetramate formation determined by quantitative LC-MS, we propose that Tyr204 may participate in a proton-shuttle mechanism to activate the substrate enol for nucleophilic attack and cyclization (Fig. 5.4). The H254N mutation nearly diminishes all enzymatic **2** formation with a 77% decrease relative to the wild-type reaction. This observation suggests that His254 may lower the pKa of Cys89, in spite of the relatively long distance between these residues, or it may stabilize an intermediate by interacting with the thioester, amide, or β-keto/enol moieties during the transformation (Fig. 5.4).

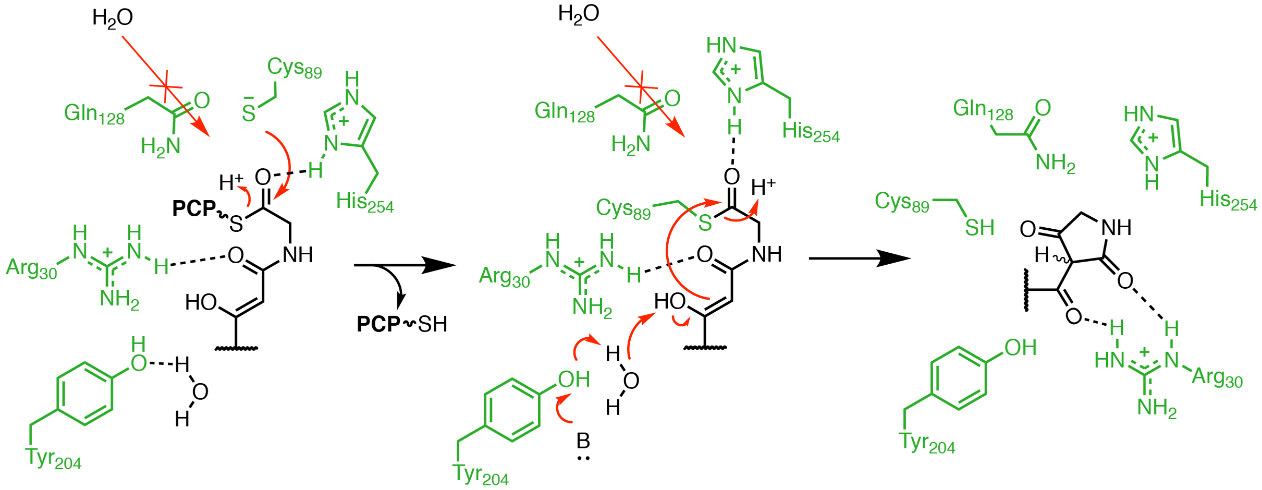


Fig. 5.4: Proposed mechanism of NcmC-catalyzed tetramate formation.

While the wild-type enzyme is the most proficient at total turnover, it is not the most proficient tetramate producer (i.e. second to the R30K variant), and a considerable amount of hydrolysis occurs (i.e. 3.4-fold increase relative to the enzyme-free reaction). We hypothesized this may arise due to the differences in minimal substrate (**1**) and the natural PCP-linked substrate, causing alterations in substrate recognition that may compromise protection from solvent hydrolysis. To investigate this further, we synthesized the *N*-acetoacetyl-glycyl-*S*-PCP (**1.8**) substrate chemoenzymatically using a synthetic intermediate (**1.4**) from the synthesis of **1** and employing the promiscuous phosphopantetheinyl transferase, Sfp, from surfactin biosynthesis (Fig. 5.22).[189]

Previous studies of the structural homolog EntF demonstrated hydrophobic interdomain contacts between the PCP and helical bundle region of TE (*in cis*) which appropriately positions the phosphopantetheine (PPant) arm into the TE active site.[173,190] Reasoning that NcmC catalysis may be driven by similar interdomain interactions, especially as it acts *in trans*, we probed the interaction between NcmC and its *holo*-PCP using size-exclusion chromatography (SEC). Co-incubating the two proteins and subjecting them to SEC gave rise to a peak with shorter elution time and increased absorbance relative to that of NcmC alone, indicative of an intermolecular interaction (Fig. 5.23). We modeled this interaction by superimposing NcmC and a homology model of its cognate *apo*-PCP ($\mathrm{NcmB}_{1020\text{-}1094}$) onto the structures of NRPSs EntF and AB3403. Additional inspection of the electrostatic potential maps reveals complementarity between the helical bundle of NcmC and *apo*-PCP, providing a feasible picture of the interaction (Fig. 5.24-5.25).

Despite the wild-type enzyme's proclivity to catalyze the TE reaction on **1** in vitro, MS analysis of the *S. syringae* NRRL B-16468 extracts failed to detect any of the corresponding linear nocamycin that would result from hydrolytic off-loading, suggesting the PCP-linked substrate is predominantly off-loaded via Dieckmann condensation in vivo. Attempts to produce additional NcmC variants heterologously in *E. coli* resulted in insoluble protein (i.e. D116A, D116N, H254A, and H254L) or the protein degraded during purification (i.e. E32A, E118A, R30A/E32A, and R30A/H254N) preventing further analysis.

5.3.5 Bioinformatics and modeling to understand evolution of function

In order to develop an understanding of the functional evolution of this conserved scaffold, we employed a combination of bioinformatics, docking, and modeling approaches to structurally rationalize differences in reactivity and substrate selection. Using the Enzyme Function Initiative's Enzyme Similarity Tool (EFI-EST), we created a sequence similarity network (SSN) using the BLAST results from NcmC as a query sequence (Fig. 5.5).[3] Owing to the relatively small population (78 sequences) and high sequence conservation in the SSN, we used a stringent alignment score (E-value = 10^{-56}) to separate clusters and begin annotation (*SI Appendix* Materials and Methods). The tetramate-containing natural product vancoresmycin is represented in the SSN with a corresponding node (Var15) that is likely responsible for heterocycle formation.[191] Curiously, the BGC for the tetramate-containing natural product pyrroindomycin contains an NcmC homolog (PyrG2), in contradiction to prior assignment of the enzymes PyrD3/PyrD4 as responsible for tetramate formation.[170] The natural products WS9326 and bleomycin have no

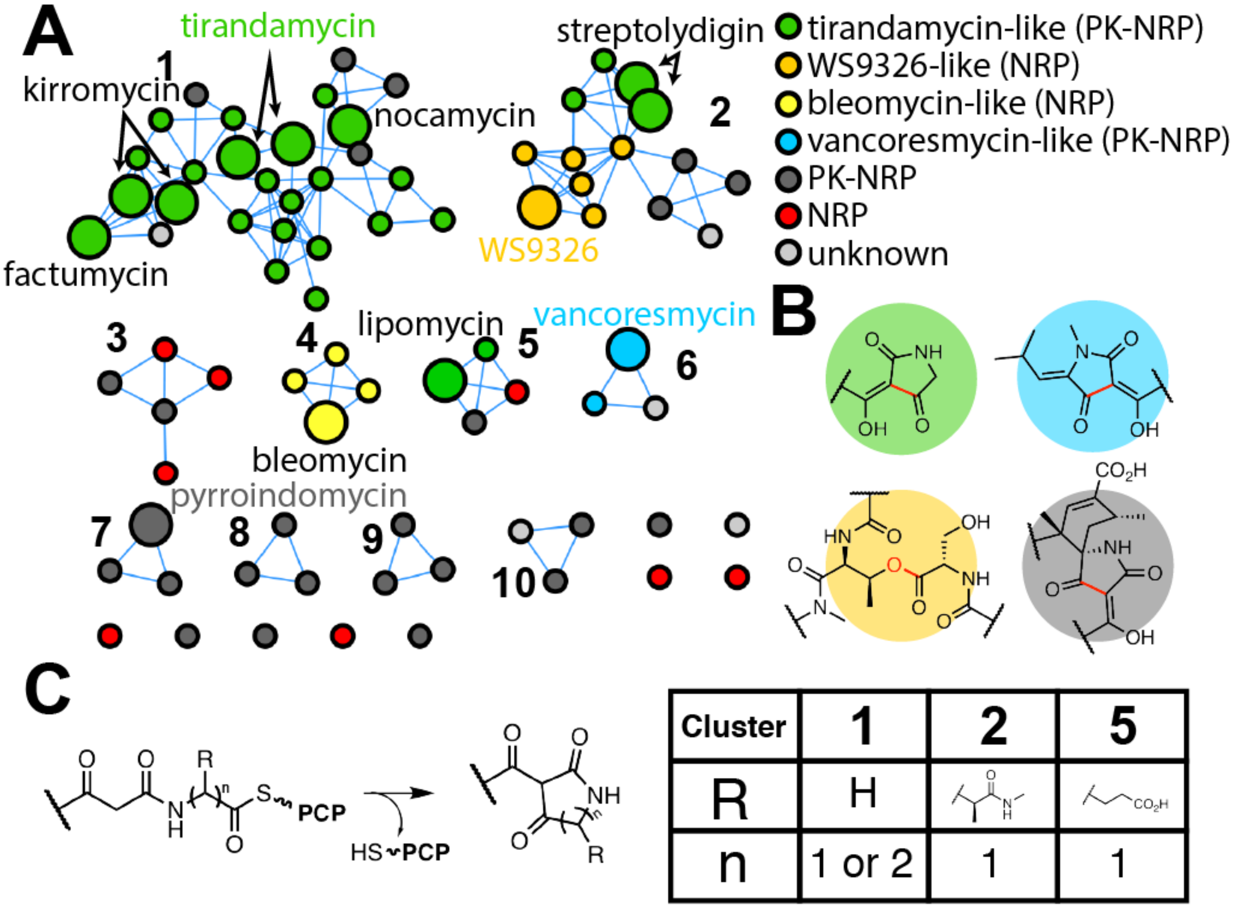


Fig. 5.5: Sequence similarity network (SSN) of NcmC homologs reveals clustering by natural product type and substrate specificity. (A) SSN using the BLAST results from query sequence NcmC (alignment score = 56). Larger circles denote proteins found in BGCs for which the natural product has been characterized. The genome context of each protein was analyzed to assign the natural product type color-coded according to homology with characterized natural products (Table 5.1). (B) Core structures of the different natural product types. Bonds in red are proposed to be formed by the NcmC homologs in the SSN. (C) General reaction scheme for tirandamycin-like DCs. The aminoacyl side chain specificity of the A-domains and DCs are tabulated for the tirandamycin-like DCs as revealed by the SSN. The specificity conferring "Stachelhaus" codes for each adenylation domain in cluster 1 were analyzed by the PKS/NRPS Analysis webtool (http://nrps.igs.umaryland.edu/)[193] and the resulting codes were input into the MEME Suite 5.0.1 Motif Discovery tool "MEME" to generate a consensus motif (DILQLGVI) predicting for Gly incorporation (Table 5.2).[194]

tetramate or pyridone rings yet each of these BGCs contain sequences represented in the SSN suggesting a similar scaffold may be used for alternative chemistry (e.g. macrolactonization). This is consistent with our observation that NcmC exhibits structural homology to enzymes with alternative function (e.g. PKS13-TE and EntF-TE).[173,174]

In cases where sequences have no obvious associated natural product structure or homology to characterized BGCs, we analyzed the local genomic regions using the antiSMASH web tool.[192] Many of the closest related natural products identified were macrolide and macrolactam polyketides, and some include tetronate oxo-heterocycles in their final structures (i.e. maklamicin, tetronasin, and chlorothricin), perhaps indicating that these DC homologs can utilize an ester substrate. Interestingly, with respect to the tirandamycin-like nodes, the SSN also reveals a clustering that is dependent on the NRPS-derived aminoacyl side chain of the substrate (Fig. 5.5C and Table 5.2).

To better visualize how NcmC might engage its substrate during catalysis, we performed docking with truncated analogs of the tetramate product in AutoDock.[77] Inspection of the lowest energy poses for both the keto and enol forms of the product shows the tetramate head group situated in the active site nearby the Cys89 of the catalytic triad and the aliphatic tail embedded along the outer face of helices α4 and α7 in the helical bundle (Fig. 5.26). Relative to the orientation of cerulenin in the NcmC cocrystal structure, consistencies were observed for the orientation of the head and tail groups as well as the proximity of Arg30 and Glu32 for H-bonding. Given the docking poses, in addition to the bioinformatic and structural data, we sought to understand the basis for substrate selection by the tetramate-forming DCs.

We generated NcmC template-guided homology models (HMs) using the SWISS-MODEL server.[195] Considering the lack of structural information with respect to the polyketide tail of the NcmC substrate, we focused our attention on recognition of the aminoacyl group of the substrate.

Following inspection of the docking and structural data, we identified a region of the active site (occupied by Glu118 in NcmC) that is most likely to accommodate the aminoacyl side chain of the substrate (Figs. 5.3 and 5.26).

Comparison of the TamC, FacHI, and KirHI HMs with NcmC revealed no considerable difference in this region, consistent with their function in selecting similar non-bulky Gly or β-Ala derived substrates. However, HMs for SlgL and Var15 (putative DC in the vancoresmycin BGC) whose substrates contain bulkier *N*,β-dimethylasparagine and (dehydro)leucine side chains, respectively, contain some notable differences in the proposed aminoacyl binding site, offering a potential explanation for substrate selection. Specifically, when overlaid with NcmC, the SlgL HM contains Ser132 in place of Glu118 (NcmC) potentially providing a more spacious active site, and this Ser is conserved in SlgL putative orthologs identified by BLAST and antiSMASH (Table 5.1 and Fig. 5.27). Similarly, the Var15 HM reveals a Thr117 in this position that is also conserved in its putative orthologs identified by BLAST and antiSMASH (i.e. AUO14856.1, WP_103348671.1, and WP_086863360.1; Table 5.1 and Fig. 5.27). Curiously, the LipX2 HM provided no obvious difference in the putative aminoacyl binding site, despite its preference for an L-Glu side chain. However, scanning of additional active site residues around the catalytic Cys revealed a His129 in place of Thr122 (NcmC) and an Arg241 in place of Ala242 (NcmC) which could feasibly interact with the carboxylate side chain of the substrate (Fig. 5.27). These His/Arg residues are also conserved in the putative LipX2 ortholog WP_061441796 (Table 5.1). Notably, the multisequence alignment reveals an 11-residue deletion in SlgL, LipX2, Var15, and WP_061441796 spanning helix α9 in NcmC and resulting in an apparent opening of the active site cleft near the putative aminoacyl binding site (~800 $Å^2$ change in total area; Figs. 5.18 and 5.26).

5.4 DISCUSSION

Metabolite off-loading from PKS/NRPS assembly lines represents an important step in the production and manipulation of bioactive compounds. Here, we examine a reoccurring protein scaffold that has been selected for a variety of chemical transformations, imparting chemical diversity in a number of off-loaded products found in nature. Specifically, we use structural and mechanistic methods to address how one type of stand-alone, off-loading enzyme catalyzes a Dieckmann condensation to afford a five-membered tetramate heterocycle. The noted structural homology to enzymes that catalyze alternative reactions, namely thioester hydrolysis and macrolactonization, suggests that these protein scaffolds can be engineered to alter their reactivities but maintain substrate specificity. Indeed, a single point mutation (Q128A) of NcmC was sufficient to increase hydrolysis to 5.2-fold greater than the non-enzymatic reaction with a corresponding 41% decrease in its natural DC activity, demonstrating partial conversion of this enzyme to a TE.

Structural and biochemical considerations point to why the DCs employ a catalytic Cys rather than a Ser found in the TEs. Additional effects on enzyme activity using structure-based NcmC variants enable a mechanistic proposal for activation of a nucleophilic enol involved in cyclization. Finally, NcmC homologs were assembled into a SSN that reveals a clustering of DCs from similar BGCs that is dependent on the identity of the aminoacyl side chains of the substrate. A deeper look into the structural features underpinning the substrate selectivity was facilitated by template-guided homology modeling and active site comparisons. Our data suggest that polyketide tail recognition is governed by the four-helical bundle, consistent with studies of the hFAS-TE structural homolog, and recognition of aminoacyl functionality is segmented into the α/β-hydrolase domain.[176] The apparent modularity of the polyketide and aminoacyl recognition sites may lend itself to engineering of hybrid DCs where recognition features are swapped in related biosynthetic pathways to assist in off-loading artificial, hybrid metabolites.

5.5 METHODS

5.5.1 Materials

Oligonucleotides were purchased from Integrated DNA Technologies. Enzymes used for PCR, plasmid digestion, and Gibson Assembly were purchased from New England Biolabs. Buffer and media components were purchased from Thermo Fisher Scientific and other reagents were purchased from Sigma-Aldrich (MilliporeSigma).

5.5.2 General Methods

Purification of protein was performed using an ÄKTAprime plus purifier (GE Healthcare). High-performance liquid chromatography (HPLC) was carried out using a Shimadzu Prominence Preparative Liquid Chromatography system. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) experiments were conducted using an UltrafleXtreme mass spectrometer (Bruker Daltonics).

5.5.3 Homologous production of nocamycin I from Saccharothrix syringae NRRL B-16468

A lyophilized mycelial stock of *Saccharothrix syringae* NRRL B-16468 was used to inoculate 6 x 5 mL solid R5A medium agar plates incubated at 28 °C for one week.[196] The media became dark red in color and agar slices were excised to inoculate 2 x 500 mL liquid R5A medium, and the liquid cultures were grown at 28 °C with shaking at 220 rpm. The cells were harvested by centrifugation at 3000 × g and cells were soaked in MeOH for 3 h followed by filtration on Whatman filter paper. The supernatant from centrifugation was extracted twice with 2 x 500 mL of ethyl acetate. The MeOH filtrate and ethyl acetate fractions were pooled and concentrated *in vacuo*. The crude, concentrated extract was used for HR-MS and MS/MS analysis.

5.5.4 Tandem mass-spectrometry (MS/MS) analysis of nocamycin I

Organic extracts derived from culture supernatant and methanolic cell soaks of *S. syringae* NRRL B-16468 were directly infused into a ThermoFisher Scientific Orbitrap Fusion electrospray-ionization mass-spectrometer (ESI-MS) using an Advion TriVersa Nanomate 100. The following parameters were used: resolution = 120,000; isolation width (MS/MS) = 2 *m/z*; normalized collision energy (MS/MS) = 35; activation q value (MS/MS) = 0.4; and activation time (MS/MS) = 30 ms. Collision-induced dissociation (CID) and high-energy collisional dissociation (HCD) were used to fragment the parent ion at 35%. Data analysis was performed in the Qualbrowser application of Xcalibur software (Thermo Fisher Scientific).

5.5.5 Cloning and heterologous expression of NcmC, NcmC site-directed variants, and NcmB (di)domains.

All heterologous expression plasmids were generated via the pET hexahis (His_6) small ubiquitin-like modifier (SUMO) tobacco etch virus (TEV) protease ligation-independent cloning (LIC) vector (2S-T). This vector was a gift from Scott Gradia (Addgene plasmid # 29711). Primers were designed to include the LICv1 tags specific to this vector. Genomic DNA template for PCR was obtained via cultivation of *Sacchrothrix syringae* NRRL B-16468 in ISP2 medium while shaking at 30 °C (15 mL) followed by isolation using the DNeasy UltraClean Microbial Kit (QIAGEN, Inc). PCR products were ligated into SspI-linearized 2S-T via the Gibson Assembly method to construct an N-terminally labeled His_6-SUMO-NcmC, His_6-SUMO-NcmB-PCP, or His_6-SUMO-NcmB-A-PCP expression plasmid.[197] Site-directed mutants of *ncmC* were generated via three-component assemblies of the SspI-linearized 2S-T and two *ncmC* gene fragments containing the PCR-derived mutations. Gibson assembled plasmids were maintained and propagated in *E. coli* DH5α. Sequencing services were provided by ACGT Inc. (Wheeling, IL). *E. coli* Rosetta (DE3) competent cells were subsequently transformed with the sequence-verified plasmids and used to overexpress all proteins by growing at 37 °C in Luria-Broth (LB) to an optical density at 600 nm (OD_{600}) of 0.5-0.6. Protein expression was then induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and allowed to grow for an additional 18-20 h at 18 °C.

5.5.6 NcmC protein purification

2-4 L of LB media was used for overexpression in *E. coli* Rosetta (DE3) for each NcmC variant. After induction, cells were centrifuged at 3,000 × g and resuspended in buffer A (500 mM NaCl, 25 mM tris-HCl pH 8.0, and 10% glycerol (*v/v*)) before subjecting to French press lysis at 8,000 – 10,000 psi for five total cycles. Lysate was cleared by centrifugation at 15,400 × g for 1 h at 4 °C and the supernatant was loaded onto a 5 mL HisTrap nickel-nitrilotriacetic acid (Ni-NTA) affinity column (GE Healthcare) equilibrated with buffer B (1 M NaCl, 25 mM tris-HCl pH 8.0, and 30 mM imidazole). This was followed by a 50 mL wash with buffer B before eluting the His_6-tagged protein by employing a linear gradient of 0-100% buffer C (1 M NaCl, 25 mM tris-HCl pH 8.0, and 500 mM imidazole) over 35 min at 2 mL/min. Pooled fractions containing His_6-SUMO-NcmC were proteolyzed using 30 µg/mL of tobacco etch virus (TEV) protease to remove the His_6-SUMO tags. Proteolysis was carried out while simultaneously dialyzing into 200 mM NaCl, 20 mM tris pH 8.2, and 2.5 mM β-mercaptoethanol (BME) for 12 h at 4 °C. Protein solution was dialyzed again into a similar buffer lacking BME for 4 h at 4 °C before loading onto a Ni-NTA column equilibrated with buffer A for subtractive Ni-NTA affinity purification. The column was washed with 35 mL of buffer A before using a stepwise gradient of increasing buffer B to elute tag-free NcmC variants. Steps of buffer B increased as follows: 10%, 20%, 30%, 40%, 50%, and 100% in 5-10 mL increments depending on the absorption at 280 nm (peaks were allowed to fully elute before moving to the next increment). NcmC eluted primarily in the flow-through and wash fractions. Prior to crystallography experiments, NcmC variants were purified by size-exclusion chromatography (SEC) using a 120 mL Superdex 200 10/300 GL column (GE Healthcare) and 100 mM KCl, 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5 as running buffer. NcmC eluted with ~90% purity at this stage by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 5.28).

5.5.7 NcmB PCP-, and A-PCP-(di)domain purification

Cells were harvested, lysed, and supernatant-cleared similar to NcmC samples. Supernatant was applied to 2 mL of HisPur nickel-nitrilotriacetic (Ni-NTA) resin (Thermo Scientific) and injected by gravity elution. The resin was washed with 40 mL of buffer B prior to eluting the protein with buffer C in 2 mL increments for 10 total fractions. His_6-SUMO-NcmB-PCP was treated with TEV protease and subjected to subtractive nickel affinity purification as described for NcmC. Tag-free NcmB-PCP (*apo*-PCP) eluted primarily in the flow-through, wash, and 10-30% buffer B fractions. Concentrated fractions were used directly in enzymatic reactions to synthesize *holo*-PCP or purified by acetone precipitation before synthesizing *N*-acetoacetyl-glycyl-*S*-PCP (see *5.5.9* In vitro enzymatic synthesis of *holo*-PCP and *N*-acetoacetyl-glycyl-*S*-PCP). A majority of His_6-SUMO-NcmB-A-PCP was observed in the cell pellet following lysis, and the small amount of soluble protein purified by Ni-NTA was observed in an aggregate state by SEC, preventing its further analysis.

5.5.8 Synthesis of N-acetoacetyl-glycyl-SNAC (1) and N-acetoacetyl-glycyl-S-CoA (1.7)

5.5.8.1 3-oxobutanoic acid; acetoacetic acid **(1.1)**

To a stirring solution of 200 mL 1 N NaOH in a 500 mL round-bottom flask was added 25 mL (196 mmol; 1 eq) of ethyl acetoacetate. The reaction was stirred at room temperature for 12 h before adding saturating amounts of $(NH_4)_2SO_4$ and then acidifying to pH 2 with concentrated HCl. The solution was then extracted twice with 150 mL of diethyl ether, dried with Na_2SO_4, and concentrated *in vacuo* to afford 15.6 g (78%) of a pale-yellow liquid. ^{1}H NMR (400 MHz, $CDCl_3$): δ 11.55 (1H, s), 3.47 (2H, s), 2.22 (3H, s) ppm (Fig. 5.10).

5.5.8.2 methyl (3-oxobutanoyl)glycinate **(1.3)**

In a flame-dried 100 mL round-bottom flask, 1.3 g (12.7 mmol; 1 eq) of **1.1** and 3.2 g (12.7 mmol; 1 eq) of disuccinimidyl carbonate (DSC) were added to 100 mL of stirring acetonitrile (ACN) cooled at 0 °C. Following a 20 min incubation, 1.5 mL (19.1 mmol; 1.5 eq) of pyridine was added dropwise and a cloudy white suspension was allowed to stir for another 20 min at 0 °C. The reaction was heated to room temperature and stirred for 1 h while the solution became clear. The solution was reduced to approximately half of its initial volume *in vacuo* and cooled to 0 °C after adding an equal volume of dry CH_2Cl_2. *In situ* generation of **1.2** was followed by addition of 1.6 g (12.7 mmol; 1 eq) glycine methyl ester hydrochloride. While maintaining at 0 °C, 2.7 mL (19.1 mmol; 1.5 eq) of Et_3N was added dropwise over 3 min. After stirring on ice for 1 h, the reaction was heated to room temperature and stirred for an additional 12 h. The reaction was concentrated *in vacuo* followed by resuspension in CH_2Cl_2. The solution was washed once with an equal volume of 1 N HCl and once with an equal volume of H_2O. After drying the organic phase with Na_2SO_4, the contents were concentrated *in vacuo* and subjected to silica gel flash-chromatography. A mobile phase of 80:15:5 EtOAc:CH_2Cl_2:MeOH was used and **1.3** eluted with an $R_f = 0.4$. 650 mg

(30%) of **1.3** was obtained as a light-yellow solid. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.53 (1H, s), 4.01 (2H, d, *J* = 5.2 Hz), 3.70 (3H, s), 3.45 (2H, s), 2.23 (3H, s) ppm (Fig. 5.11).

5.5.8.3 (3-oxobutanoyl)glycine **(1.4)**

In a 10 mL round-bottom flask 650 mg (3.8 mmol; 1 eq) of **1.3** was stirred in 2 mL of EtOH at room temperature followed by the addition of 4 mL of 1 N NaOH (4 mmol; 1.1 eq). The reaction was stirred for 1 h at room temperature before removing the EtOH *in vacuo* and acidifying the solution to pH 3. The aqueous contents were then extracted using ethyl acetate (5 x 20 mL). Organic phases were pooled, dried using Na_2SO_4, and concentrated *in vacuo* to yield 350 mg (59%) of **1.4** as a yellowish white solid. ^{1}H NMR (500 MHz, DMSO-d_6): δ 3.77 (2H, d, *J* = 5.5 Hz), 3.35 (2H, s), 2.15 (3H, s) ppm (Fig. 5.12).

5.5.8.4 S-phenyl 2-(3-oxobutanamido)ethanethioate **(1.5)**

Following slight modification to the procedure of Sims et al., 50 mg of **1.3** (0.31 mmol; 1 eq) was added to 5 mL of ACN in a 15 mL flame-dried round-bottom flask along with 96 μL of thiophenol (0.94 mmol; 3 eq) and 196 mg of (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP) (0.38 mmol; 1.2 eq).[161] After stirring for 1 h at room temperature, 25 mg of K_2CO_3 (0.19 mmol; 0.6 eq) was added and stirred for an additional 2 h. After concentrating *in vacuo,* the contents were dissolved in minimal EtOAc for purification by silica gel flash-chromatography with a gradient mobile phase of 0-20% hexanes in EtOAc over 100 mL. The product eluted with an R_f = 0.5 in 1:4 hexanes:EtOAc. Fractions containing crudely purified material were concentrated *in vacuo* and resuspended in DCM. Filtration of the suspension afforded pure product in the filtrate (50 mg; 63%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.70 (1H, m), 7.62 (1H, m), 7.41 (1H, m), 7.26 (2H, m), 4.18 (2H, d, *J* = 8 Hz), 3.39 (2H, s), 2.14 (3H, s) ppm (Fig. 5.13).

5.5.8.5 3-acetyl-4-hydroxy-1,5-dihydro-2H-pyrrol-2-one **(1.6)**

Following slight modification to the procedure of Sims et al., 4.6 mg of sodium methoxide (86 μmol; 1.2 eq) was added to a 10 mL flame-dried round-bottom flask and dissolved in 2 mL of dry benzene, followed by the addition of 18 mg of **1.5** (72 μmol; 1 eq).[161] The reaction was refluxed for 5 h then extracted with water. The aqueous layer was acidified to pH 2 using concentrated H_2SO_4 and extracted into diethyl ether. After concentration in vacuo, the product was purified by HPLC using a Grace Vydac DENALI C18 column (250 × 4.6 mm, 5 μm particle size, 120 Å pore size) and monitoring by absorbance at 230 nm. A two-component mobile phase consisted of H_2O + 0.1% formic acid (A) and ACN + 0.1% formic acid (B). Initially 5% B (*v/v*) was held for 10 min followed by a linear gradient of 5 - 60% B (*v/v*) over 21 min. **1.6** eluted around 15 min and was concentrated *in vacuo* to afford a white solid (2.0 mg, 20%) ^{1}H NMR (500 MHz, $CDCl_3$): δ 5.71 (1H, s), 3.82 (2H, s), 2.48 (3H, s) ppm (Fig. 5.14).

5.5.8.6 *S*-(2-acetamidoethyl) 2-(3-oxobutanamido)ethanethioate **(1)**

100 mg (0.63 mmol; 1 eq) of **1.4** was added to a flame-dried 25 mL round-bottom flask containing 5 mL of DMF and stirred briefly at 0 °C followed by addition of 133 mg (0.69 mmol; 1.1 eq) *N*-(3-dimethylaminopropyl)-*N′*-ethylcarbodiimide hydrochloride (EDC) and 106 mg (0.69 mmol; 1.1 eq) 1-hydroxybenzotriazole (HOBt). After stirring at 0 °C for 10 min 164 μL (0.94 mmol; 1.5 eq) of *N,N*-diisopropylethylamine (DIPEA) was added dropwise over 2 min followed by dropwise addition of 67 μL (0.63 mmol; 1 eq) *N*-acetylcysteamine over 5 min. The reaction was stirred at 0 °C for 30 min then heated to room temperature and stirred for another 6 h. DMF was removed from the reaction by running silica gel flash-chromatography with a 60:35:5 EtOAc:CH_2Cl_2:MeOH mobile phase. Crudely purified **1** eluted with an R_f = 0.2-0.3 as verified by ^{1}H NMR. **1** was further purified using HPLC with monitoring by absorbance at 230 nm and a Grace VisionHT C18 HL column (5 μm particle size; 250 mm length; part No. 5151988). The same elution program for **1.6** was used with mobile phases lacking formic acid. **1** eluted around 3

min and was concentrated *in vacuo* to afford a white solid (4 mg; 2%). ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.72 (1H, s), 5.85 (1H, s), 4.22 (2H, d, *J* = 6.0 Hz), 3.52 (2H, s), 3.43 (2H, q, *J* = 6.0 Hz), 3.07 (2H, t, *J* = 6.3 Hz), 2.30 (3H, s), 1.97 (3H, s) ppm (Fig. 5.15). HRMS: calculated $[M+H]^+$ = 261.0909 Da; observed $[M+H]^+$ = 261.0909 Da (Fig. 5.16).

5.5.8.7 *N*-acetoacetyl-glycyl-*S*-CoA **(1.7)**

Following slight modification to the procedure of Sims et al., 12 mg of **1.5** (48 µmol; 1.5 eq) was added to 25 mg of coenzyme A (CoA) trilithium salt dissolved in 20 mM sodium phosphate buffer (200 µL; pH 8).[161] After 30 min of stirring at room temperature, 10% THF was added to increase solubility of **1.5** followed by additional stirring for 2.5 h. The reaction was neutralized by addition of 10 µL of 1 N HCl and products were purified by HPLC intially using a Grace VisionHT C18 HL column (5 µm particle size; 250 mm length; part No. 5151988) and an isocratic elution of 2% (*v/v*) aqueous ACN over 35 min. Crudely purified product eluted from 3.5 – 8 min was subject to additional HPLC purification using a YMC-Triart Diol-HILIC column (5 µm particle size; 250 mm length; part No. 0425087425) with a two-component mobile phase (i.e. A = 10 mM ammonium bicarbonate pH 7.0; B = ACN) and an intial linear descent from 95 – 50% B over 10 min followed by a 15 min hold at 50% B. The product **1.7** co-eluted with free CoA from 16-20 min and was not further purified for downstream phosphopantetheinylation reactions (see **1.8** synthesis).

5.5.9 In vitro enzymatic synthesis of holo-PCP and N-acetoacetyl-glycyl-S-PCP (1.8)

The surfactin phosphopantetheinyl transferase (Sfp) from *Bacillus subtilis* (gifted from Dr. John Cronan) was employed in the modification of the *apo* peptidyl carrier protein (PCP) domain from the NRPS NcmB to generate *N*-acetoacetyl-glycyl-*S*-PCP and *holo*-PCP. *Holo*-PCP was synthesized by adding 8 µM Sfp to a mixture of 500 µM *apo*-PCP (from subtractive Nipurification), 2 mM CoA, 1 mM TCEP, 2.5 mM $MgCl_2$, and 50 mM tris pH 7.0 (500 µL total), and the reaction contents were incubated at 30 °C for 12 h. *Holo*-PCP was purified by adding 500 µL ice-cold acetone to precipitate protein impurities. The precipitate was removed by centrifugation, and the acetone layer was concentrated by applying a stream of N_2. The concentrate was resuspended in 500 µL H_2O and buffer exchanged into 50 mM tris pH 7.0 using a Sephadex G-25 in PD-10 Desalting Column (GE Healthcare). The eluate was concetrated by lyophilization to yield 1.5 mg (71%) *holo*-PCP (≥95% pure by SDS-PAGE; Fig. 5.29) quantified by the Bradford assay.[198] *N*-acetoacetyl-glycyl-*S*-PCP (**1.8**) was synthesized in a nearly identical manner to *holo*-PCP, except 125 µM *apo*-PCP substrate was used (purified by acetone precipitation as described for *holo*-PCP), 625 µM **1.7** was used rather than 2 mM CoA, and the reaction was terminated by flash-freezing in liquid N_2 after 3 h. The reaction mixture was used directly for downstream NcmC reactions.

5.5.10 MALDI-TOF MS analysis of Sfp reactions (holo-PCP and 1.8)

A small aliquot of 5 µL from Sfp enzymatic reactions was desalted following the manufacturer protocol using a C4 ZipTip (Millipore Sigma). Desalted samples contained 50% aq. ACN + 0.1% trifluoroacetic acid (*v/v*) and were spotted onto a Bruker 384-well MALDI plate using a 2,5-dihydroxybenzoic acid maxtrix for analysis by MALDI-TOF MS in linear positive mode (Fig. 5.22).

5.5.11 In vitro enzymatic reactions with NcmC and N-acetoacetyl-glycyl-SNAC (1)

Following subtractive Ni-NTA affinity purification, NcmC variants were concentrated to ≤ 5 mL and incubated with 1 mM dithiothreitol (DTT) for 30 minutes before injecting onto a 120 mL Superdex 75 10/300 GL column (GE Healthcare) and eluting with 0.1 M NaCl, 20 mM HEPES pH 7.0, 30 μM DTT running buffer. To minimize the amount of non-enzymatic consumption and maximize the enzymatic consumption of **1**, a panel of buffers was tested for optimal product production determined by relative HPLC peak areas. Both tris and HEPES buffers ranging from pH 7 - 7.5 were tested before arriving at an optimal tris pH 7.2 buffer. Additionally, it was found that NcmC was most active in the presence of DTT and was included at a low concentration (i.e. 30 μM) to minimize potential transthioesterification of **1**, thus an optimal reaction buffer of 100 mM NaCl, 20 mM tris pH 7.2, and 30 μM DTT was used for all in vitro enzyme reactions. 50 μL reactions were run at 28 °C for 1.5 h with 23 μM enzyme and 100 μM substrate (**1**) in reaction buffer. To quench the reactions, an equal volume of ACN was added and incubated for 15 min at room temperature to precipitate the proteins. Precipitates were pelleted by centrifugation, followed by addition of 0.1% formic acid to the supernatants for downstream HPLC and LC-MS analysis.

5.5.12 Analytical size-exclusion chromatography

5.5.12.1 NcmC variants: The SEC-purified NcmC variants used in in vitro assays were further purified by SEC on a 30 mL Superdex 200 10/300 GL column (GE Healthcare) equilibrated with 0.3 M KCl, 20 mM HEPES pH 7.5 and using a similar running buffer at a flow rate of 0.6 mL/min. Commercially available protein standards (Millipore Sigma) were run under similar conditions, enabling standard curve generation (Fig. 5.28). NcmC variants eluted as a monomer in solution (in agreement with the crystallographic data).

5.5.12.2 NcmC + holo-PCP: Wild-type NcmC (20 nmol), *holo*-PCP (20 nmol), or NcmC + *holo*-PCP (20 + 22 nmol) were separately injected in 200 μL volumes onto a 30 mL Superdex 200 10/300 GL column (GE Healthcare) equilibrated with 0.3 M KCl, 2 mM BME, 20 mM HEPES pH 7.5 and eluted with a similar running buffer while monitoring absorbance at 254 nm (Fig. 5.23).

Before co-injecting NcmC + *holo*-PCP, the sample was allowed to incubate at room-temperature for 20 min.

5.5.13 HPLC analysis of the in vitro NcmC reactions

Quenched reactions were analyzed by HPLC using a Grace Vydac DENALI C18 column (250 × 4.6 mm, 5 µm particle size, 120 Å pore size). Elution consisted of a two-component mobile phase (i.e. A = H_2O + 0.1% formic acid; B = ACN + 0.1% formic acid) at 1 mL/min with the following linear gradient: 0-5 min, 2% B; 5-15 min, 2-35% B; 15-27 min, 90% B. LabSolutions 5.81 software was used for data analysis.

5.5.14 LC-MS analysis of the in vitro NcmC reactions

Samples were analyzed with the 5500 QTRAP LC/MS/MS system (Sciex, Framingham, MA) in the Metabolomics Lab of Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign. Software Analyst 1.6.2 was used for data acquisition and analysis. The 1200 series HPLC system (Agilent Technologies, Santa Clara, CA) includes a degasser, an autosampler, and a binary pump. The LC separation was performed on an Agilent SB-Aq (4.6 x 50mm, 5µm) with a two-component mobile phase (A = H_2O + 0.1% formic acid; B = ACN + 0.1% formic acid). The flow rate was 0.3 mL/min. The linear gradient was as follows: 0-2 min, 0% B; 2-6 min, 0-98% B; 6-9 min, 98%B; 9-10 min 98-0% B; 10-14min, 0%B. The autosampler was set at 10 °C. The injection volume was 5 µL. Mass spectra were acquired under both positive (ion spray voltage was +5500 V) and negative (ion spray voltage was -4500 V) ESI modes. The source temperature was 450 °C. The curtain gas, ion source gas 1, and ion source gas 2 were 33, 65, and 60, respectively. Multiple reaction monitoring (MRM) was used for quantitation: in positive mode, compound **4** was monitored at *m/z* 120.0 --> *m/z* 86.0; in negative mode, compound **3** was monitored at *m/z* 158.1 --> *m/z* 100.0.

5.5.15 Monitoring reaction progress: 2,6-dichlorophenolindophenol (DCPIP) spectrophotometric coupled assay

In vitro NcmC reactions with **1** were initiated as described above with 15 μL reaction volumes and containing 150 μM DCPIP. Reactions were contained in a tape-sealed Corning 384-well black with clear bottom, low flange polystyrene plates (MilliPore Sigma; CLS3764). Absorbance at 600 nm was measured using an Epoch Microplate Spectrophotometer equipped with Gen5 2.09 software (BioTek Instruments, Inc.).

5.5.16 Covalent modification of NcmC by cerulenin

Wild-type NcmC (40 μM) purified by SEC as described for the in vitro reactions was incubated with 1 mM cerulenin at 28 °C for 4 h in reaction buffer (50 μL total volume). Half of the sample was treated with 7.5 μg trypsin and incubated at 24 °C for 30 min prior to desalting both the trypsinized and non-trypsinized reactions using a C4 resin ZipTip (MilliPore Sigma). Samples were spotted on a Bruker matrix-assisted laser desorption/ionization (MALDI) plate with 2,5-dihydroxybenzoic acid matrix and analyzed using a Bruker Daltonics UltrafleXtreme MALDI time-of-flight mass spectrometer (MALDI-TOF MS) in reflector positive mode at the University of Illinois School of Chemical Sciences Mass Spectrometry Laboratory.

5.5.17 Protein crystallization

5.5.17.1 NcmC: Following SEC, NcmC was concentrated to 20 mg/mL prior to crystallization in 2 μL hanging drops by incubation at 9 °C of 1:1 (protein:reservoir solution) above a reservoir solution containing 1.75 M $(NH_4)_2SO_4$, 1.9% (*w/v*) polyethylene glycol (PEG) 400, 0.1 M HEPES pH 7.0, and 3 mM DTT. Long rod-shaped crystals grew to maximum size within 36 h.

5.5.17.2 NcmC-C89S: The NcmC-C89S variant was crystallized similarly to wild-type NcmC with the exception of using a reservoir solution of 1.70 M $(NH_4)_2SO_4$, 2.5% (*w/v*) PEG 400, 0.1 M HEPES pH 7.0, and 5 mM tris(2-carboxyethyl)phosphine (TCEP). Long rod-shaped crystals grew to maximum size within 48 h.

5.5.17.3 NcmC-cerulenin: Following SEC and concentration of NcmC to 20 mg/mL, NcmC was incubated with 2 mM DTT for 30 min followed by incubation with 4 mM cerulenin for 1 h at room temperature. Covalently modified NcmC (Figs. 3 and S14) was then crystallized at 9 °C in 2 μL hanging drops by incubation of 1:1 (protein:reservoir solution) above a reservoir solution containing 1.8 M $(NH_4)_2SO_4$, 1.8% (*w/v*) PEG 400, 4% (*v/v*) formamide, 0.1 M HEPES pH 7.0, and 3 mM DTT. Long rod-shaped crystals grew to maximum size within 36 h.

5.5.18 X-ray data collection, and structure solution

Prior to flash-freezing in liquid nitrogen, crystals were immersed in their crystallization reservoir solution supplemented with 25% ethylene glycol. Data were collected using the 21-ID-D, 21-ID-F, and 21-ID-G beamlines at the Life Science Collaborative Access Team (LS-CAT) at the Advanced Photon Source (Argonne National Laboratory). Crystals mounted on an MD2 diffractometer (Arinax Scientific Instrumentation) were maintained at 100 K under an aerosolized nitrogen stream. NcmC crystals soaked in 5 mM mercury (II) acetate for 5 h at room temperature were used to collect diffraction data at 12.4 keV. Initial diffraction data were processed using XDS and autoPROC to obtain scaled and merged intensities.[71,199] Experimental phases were obtained by the single-wavelength anomalous diffraction method using Phenix AutoSol.[33] An initial model was built in Phenix AutoBuild and further refined using Phenix Refine and REFMAC5 along with manual fitting in Coot.[34–36,38] Diffraction data for the NcmC-C89S and NcmC-cerulenin variants were collected at wavelengths of 0.97856 Å and 0.97872 Å, respectively, and their structures were solved by similar XDS/autoPROC data processing followed by molecular replacement using the NcmC coordinates (PDB 6E6Y) in Phaser.[72] Initial models were similarly refined in Phenix Refine and REFMAC5 and manually refined in Coot. The crystallographic information file (cif) library and coordinate (pdb) files were created in PRODRG for the ring-opened and cyclic hemiaminal form of cerulenin (structure **1b** in Fig. 5.3B omitting the protein-derived atoms).[200] The resultant cif/pdb files were used to fit the ligand manually into the difference Fourier map ($F_{obs} – F_{calc}$) and further refined in REFMAC5. Overall Ramachandran statistics are as follows: NcmC (98% favored, 1.9% allowed, and 0.2% outliers); NcmC-C89S (98% favored, 1.5% allowed, and 0.0% outliers); NcmC-cerulenin (97% favored, 2.4% allowed, and 0.4% outliers).

5.5.19 Sequence Similarity Network (SSN) of NcmC homologs

The Enzyme Function Initiative's Enzyme Similarity Tool (EFI-EST; https://efi.igb.illinois.edu/efi-est/) was used to perform a BLAST search of the UniProt database with NcmC as a query amino acid sequence. An alignment score of 56 was used to generate the SSN containing 78 sequences restricted in length between 240-292 residues. Analysis of the network was carried out in Cytoscape 3.6.1.[201]

5.5.20 SSN node and associated NRPS adenylation (A) domain annotation

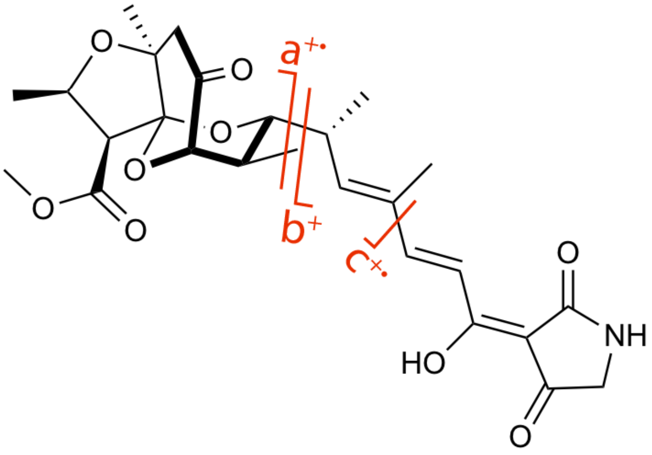
We initially identified the in vitro characterized DCs (i.e. NcmC, TrdC, SlgL, and KirHI), bioinformatically inferred DCs (i.e. FacHI and LipX2), and additional nodes in the SSN which are associated with characterized BGCs (larger circles in Fig. 5.5A). The RODEO web tool was then used to assign the genomic context of each node (for node details see Tables 5.1-5.2).[202] Based on natural product structure and co-occurrence of genome neighbors, we assigned each node to a particular cluster type. In silico prediction of NRPS A-domain specificity was carried out using the webtool developed by Bachmann et al. (http://nrps.igs.umaryland.edu/).[193] Analysis of cluster 1 predicts a consensus Gly motif (i.e. DILQLGVI; Table 5.2).[203,204] Similar A-domain prediction failed for cluster 2, but prior work on the methyltransferase SlgM (from streptolydigin biosynthesis) suggests an NRPS-bound *N*,β-dimethylasparagine to be a substrate prior to off-loading.[205] The lipomycin A-domain in cluster 5 was also resistant to specificity prediction but is presumed to select L-Glu on the basis of the final product structure.[206] Analysis of the genomic contexts of uncharacterized nodes in clusters 2 and 5 by antiSMASH suggests that these are additional streptolydigin and lipomycin producers, respectively (Table 5.1).

5.5.21 Docking analysis with NcmC and truncated tetramate product

The nocamycin I structure was truncated in ChemDraw at C7 to eliminate the fused tricyclic spiroketal moiety. The structure-data file (sdf) was opened in Molecular Operating Environment (MOE) 2016.08 and energy minimized in the gas phase using the MMFF94x forcefield (dielectric = 1; exterior = 80).[118] The (*R/S*) forms of the keto tautomer were manually built in MOE and similarly energy minimized. The enol and (*R/S*) keto structures were saved as sdf files and opened in the PyMOL Molecular Graphics System (1.7.0.0) containing the Autodock/Vina plugin (2.2.0). A ~27,000 Å^3 box was centered around the cerulenin binding site with a grid spacing of 0.375.[77] The NcmC receptor coordinates were prepared by removing cerulenin and water from the cocrystal structure. The sdf files were used to generate the enol and

(*R/S*) keto ligands prior to running Vina, and the resultant poses were analyzed in Chimera 1.13.0 (Fig. 5.26).

5.6 SUPPORTING FIGURES AND TABLES



Fragment Ion	Fragment Calculated Mass (Da)	Fragment(s) Observed Mass (Da)	Mass Error (ppm)	Fragmentation technique
$a^{+\bullet}$	284.1254	284.1278	8.4	HCD
b^{+}	220.0968	220.0966	0.9	CID
b^{+}	220.0968	220.0964	1.8	HCD
$c^{+\bullet}$	352.1880	352.1901	6.0	HCD

Fig. 5.6: Tandem-MS (MS/MS) analysis of nocamycin I from extracts of *Saccharothrix syringae* NRRL B-16468. Fragmentation data were collected using a ThermoFisher Scientific Orbitrap Fusion ESI-MS. Collision-induced dissociation (CID) and/or high-energy collisional dissociation (HCD) were used to fragment the parent ion at 35%. A ppm error of less than 10 was designated as the threshold for peak assignment.

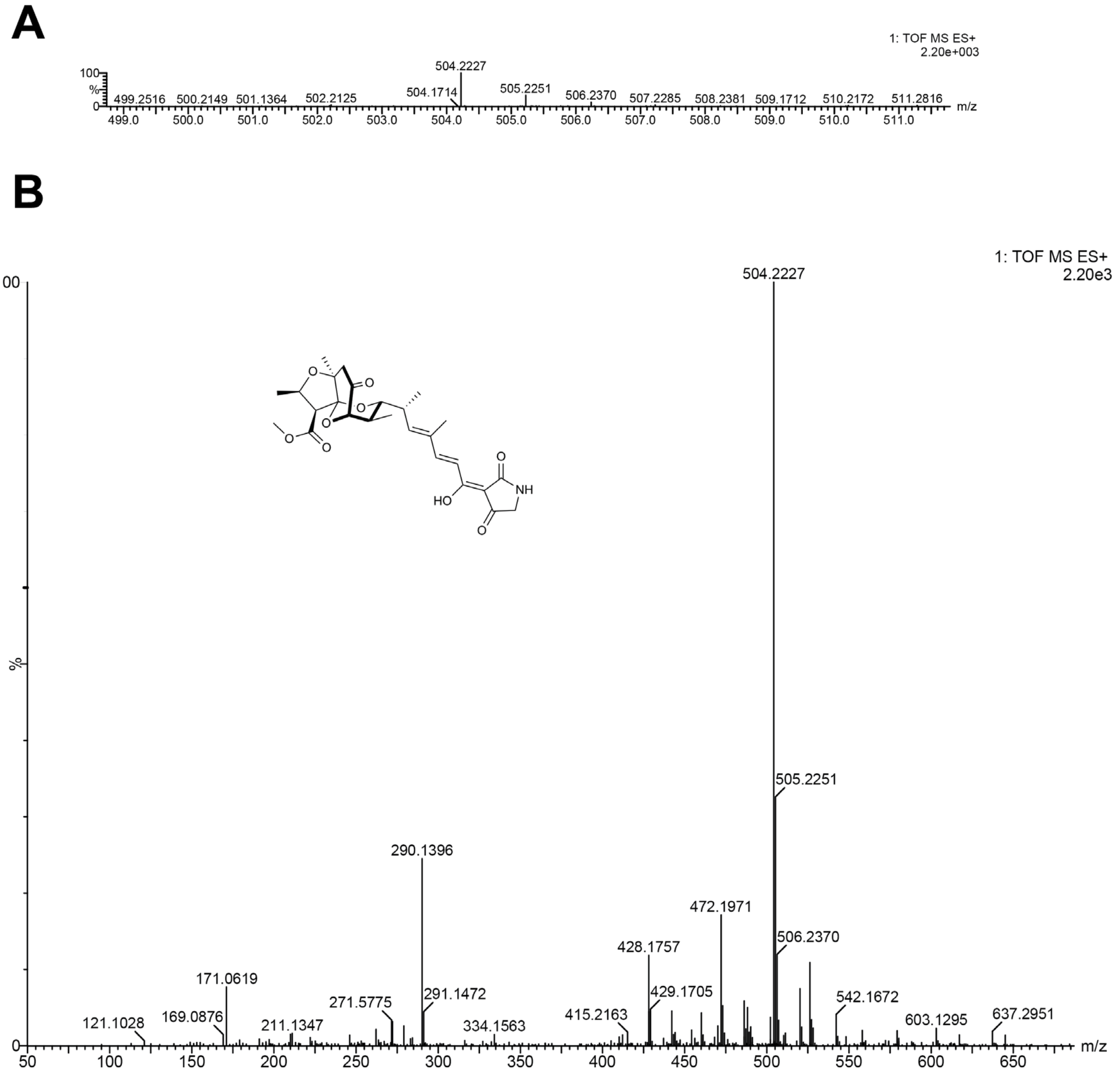


Fig. 5.7: High-resolution mass spectrum of nocamycin I. (A) Single mass analysis of nocamycin I (calculated mass = 504.2234 Da; observed mass = 504.2227; ppm error = 1.4). (B) Full spectrum mass analysis of nocamycin I. Both spectra were collected using a Waters Synapt G2-Si ESI-MS equipped with a quadrupole time-of-flight detector.

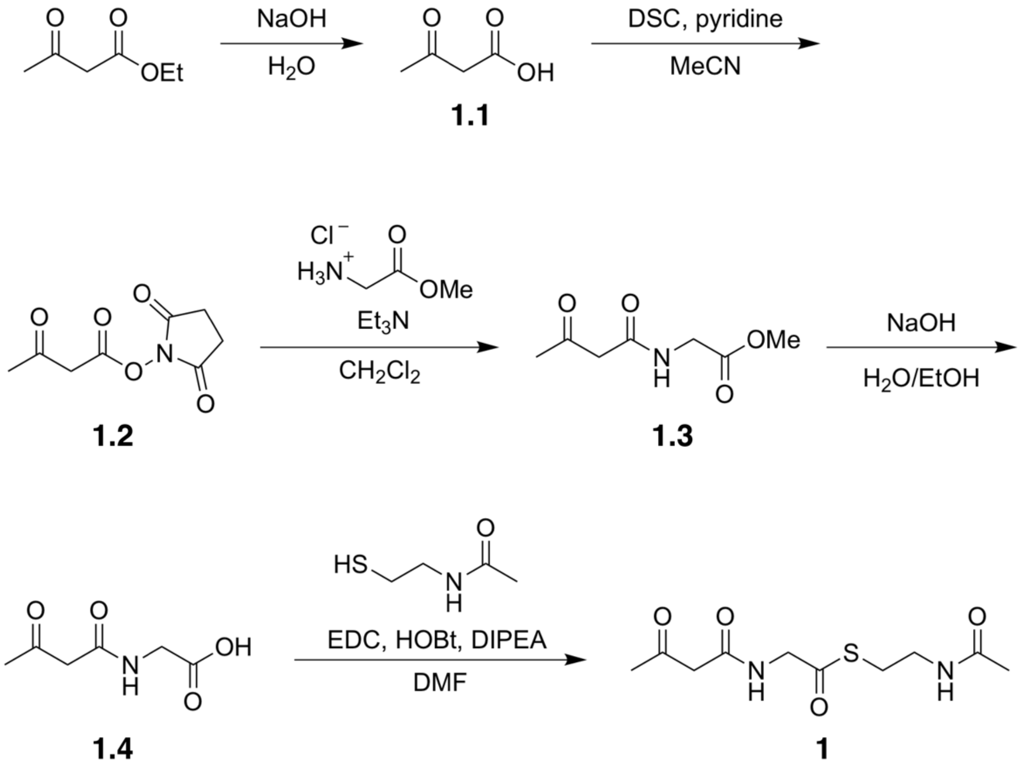


Fig. 5.8: Synthetic route towards *N*-acetoacetyl-glycyl-SNAC (1). See the Methods for detailed synthetic procedures.

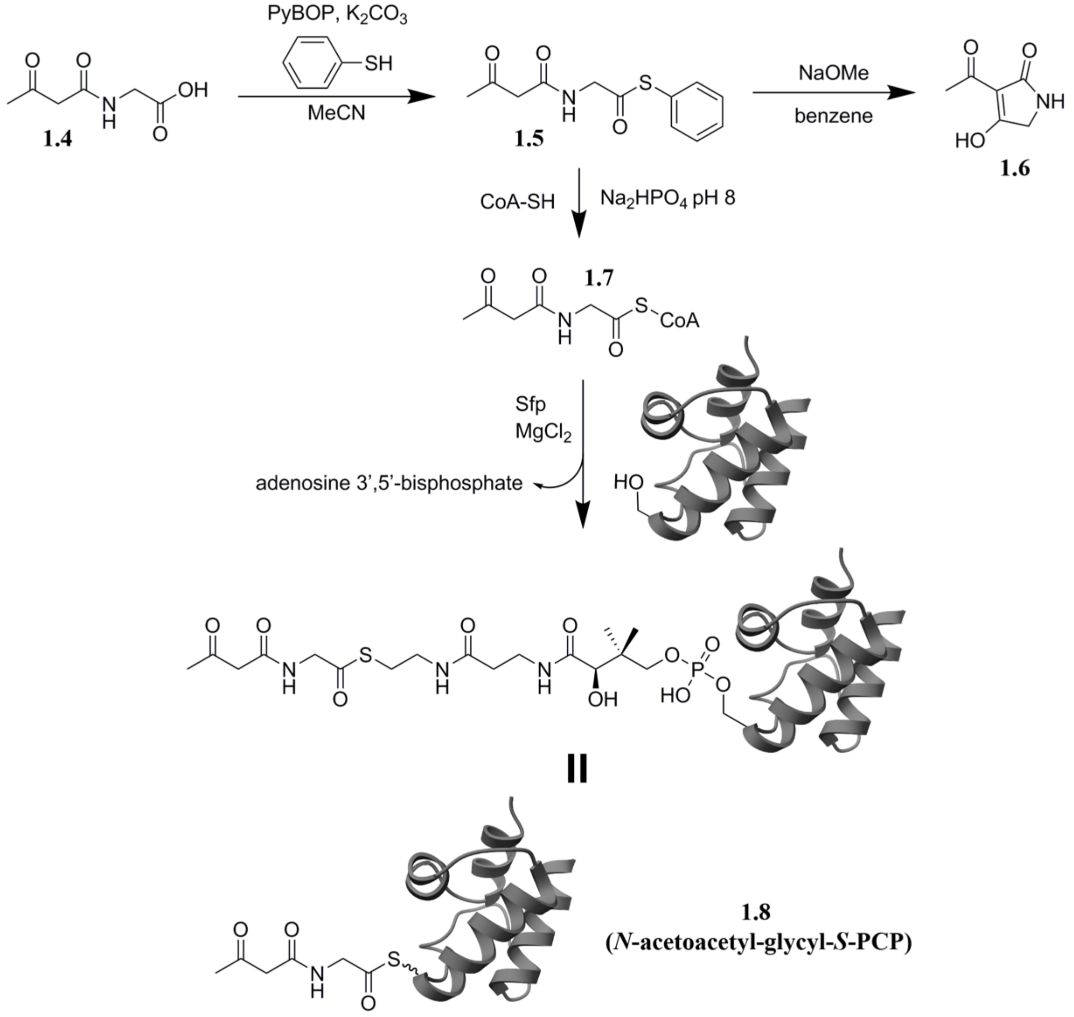


Fig. 5.9: Synthetic route towards 1.6 and *N*-acetoacetyl-glycyl-S-PCP (1.8). See the Methods for detailed synthetic procedures.

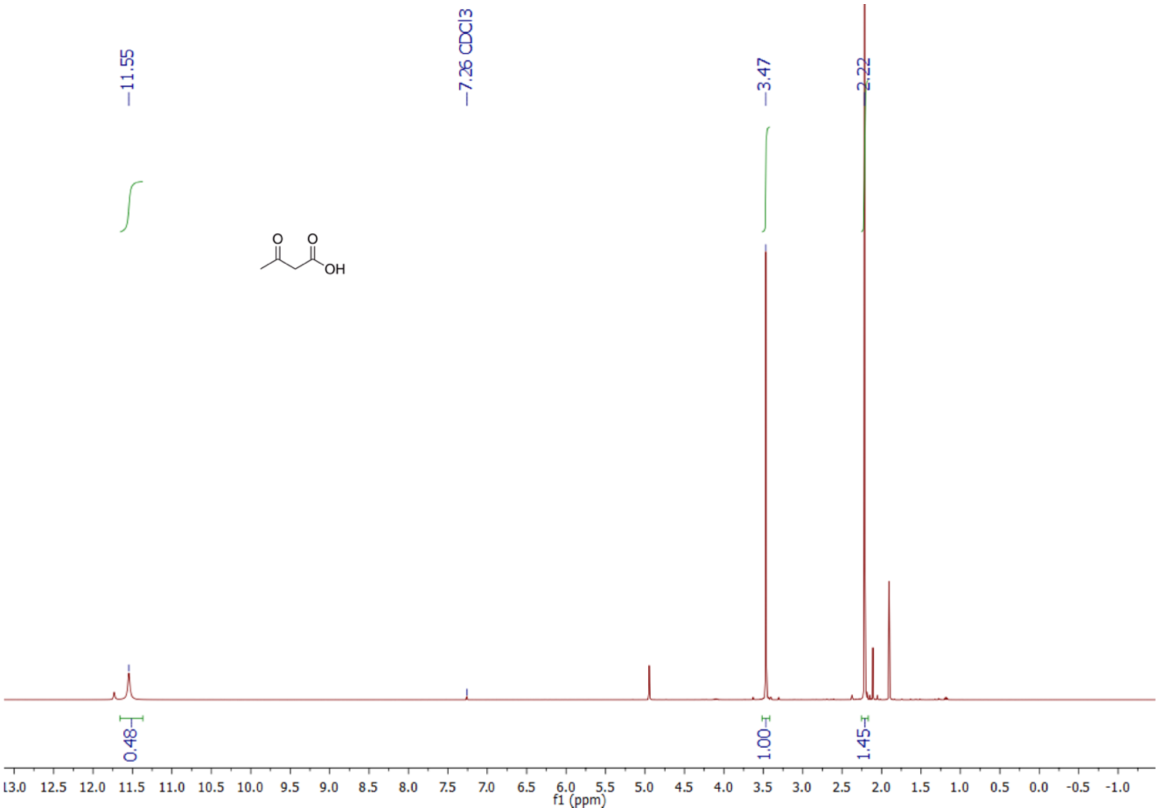


Fig. 5.10: ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of **1.1**.

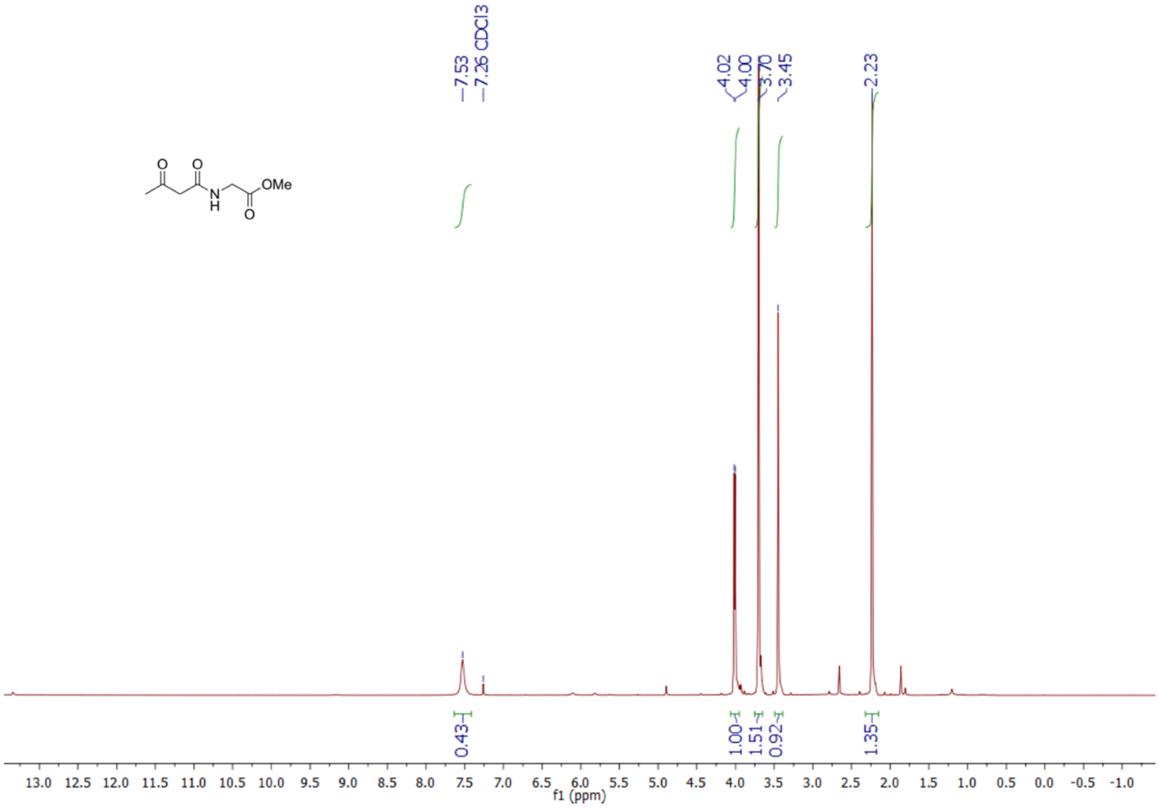


Fig. 5.11: ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of **1.3**.

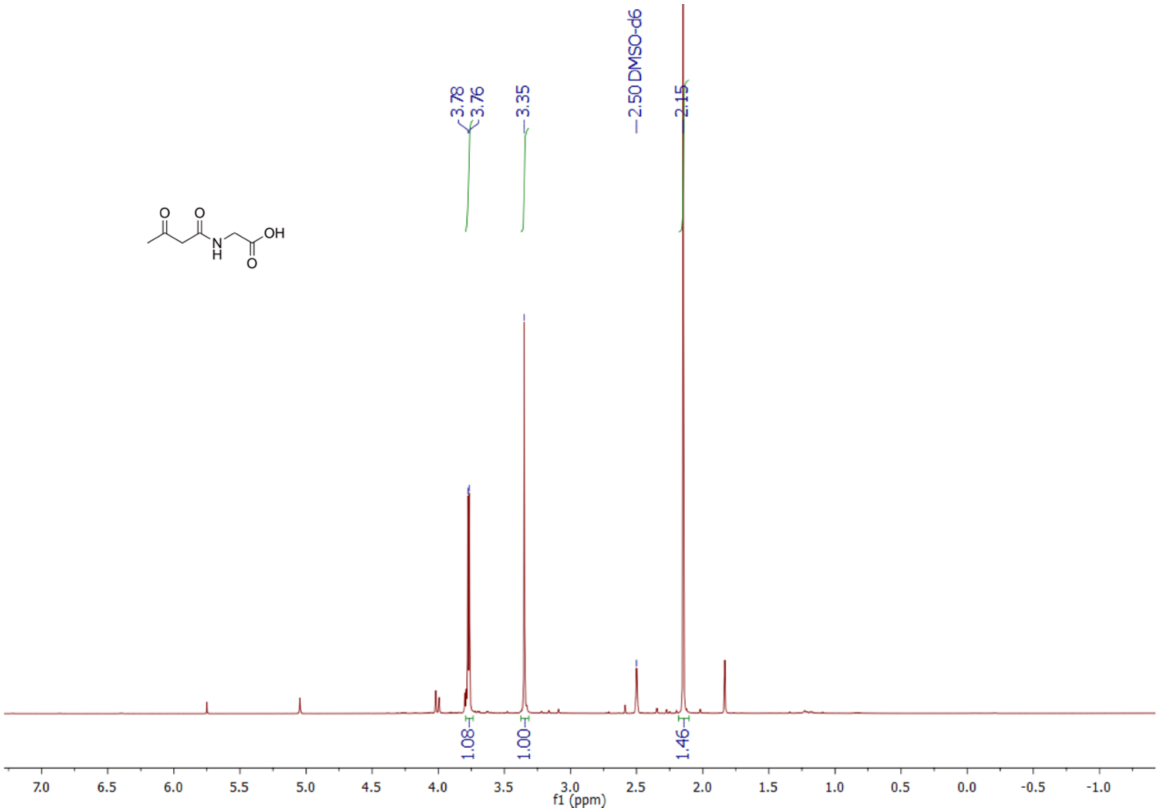


Fig. 5.12: ^{1}H NMR (500 MHz, DMSO-d_6) spectrum of **1.4**.

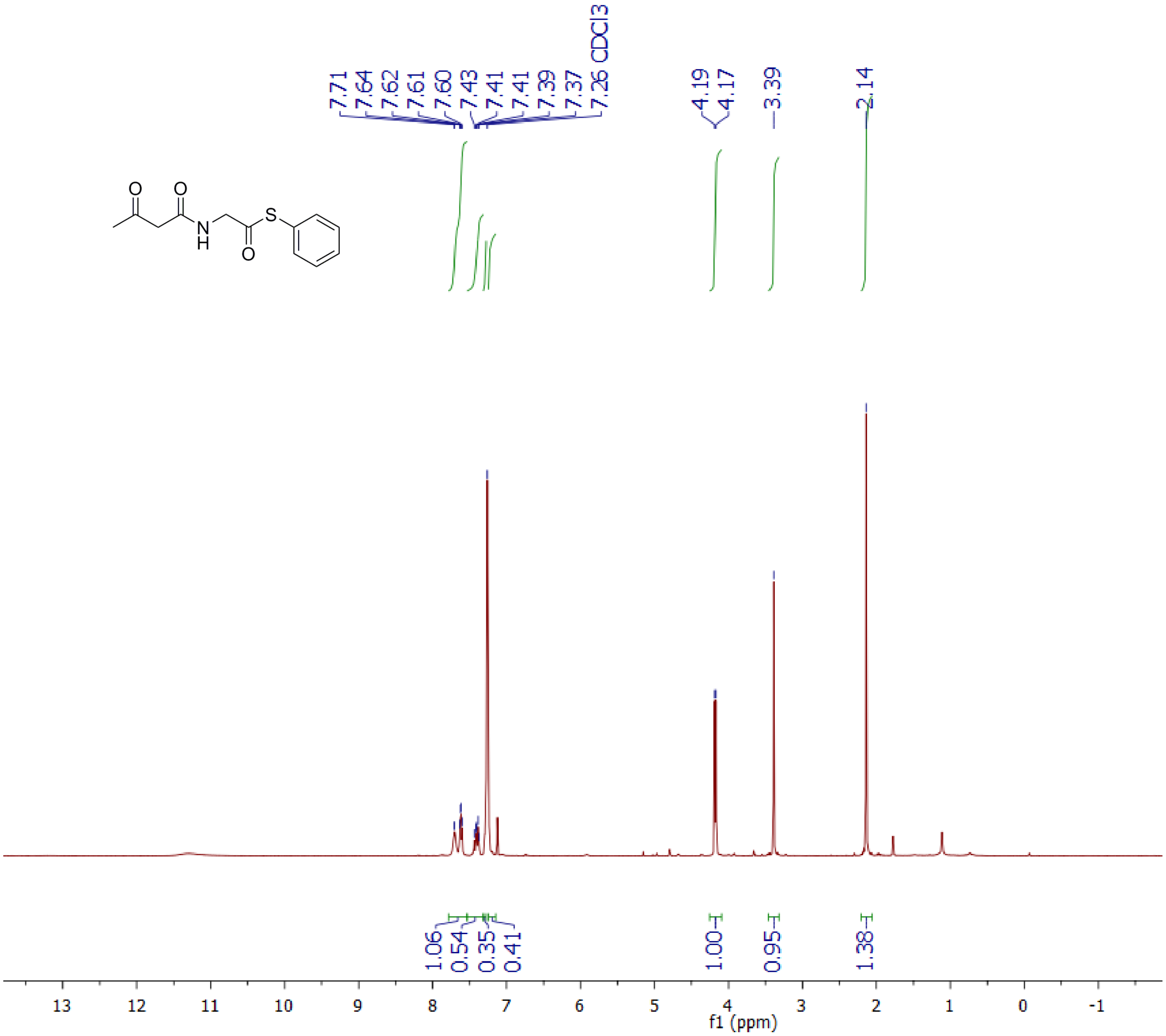


Fig. 5.13: ^{1}H NMR (400 MHz, CDCl$_3$) spectrum of **1.5**.

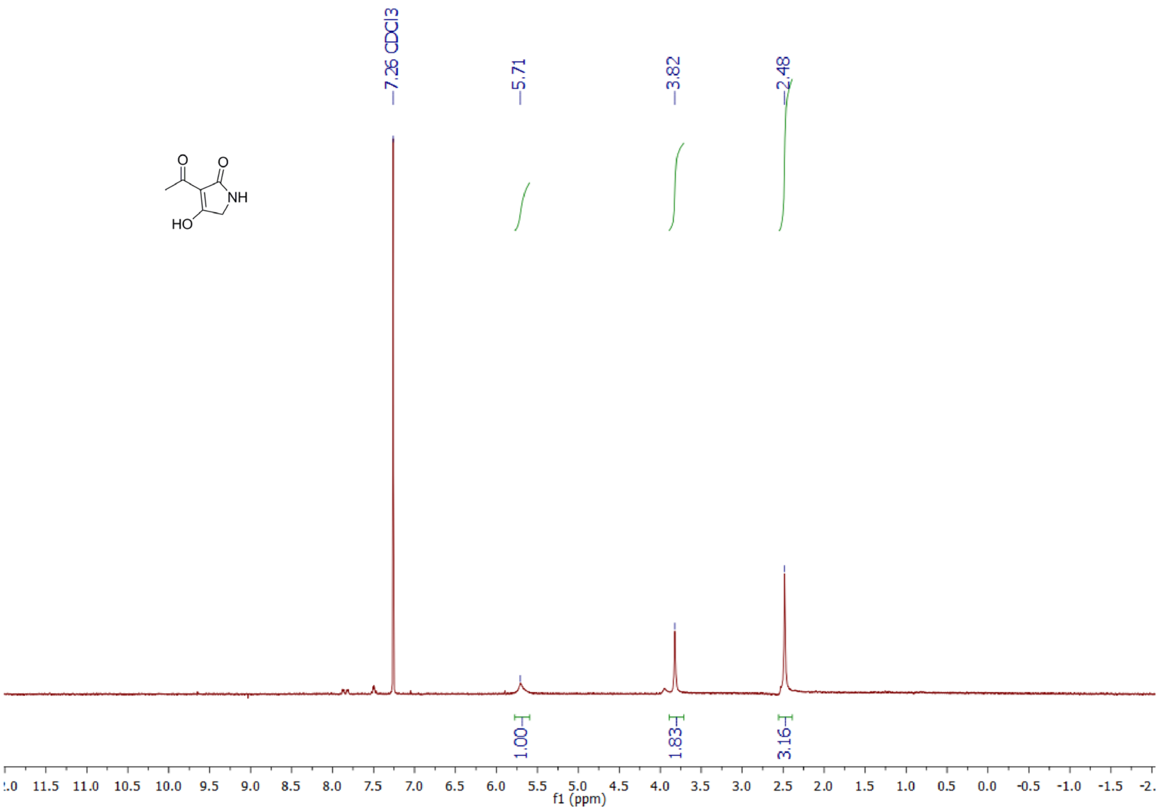


Fig. 5.14: ^{1}H NMR (500 MHz, $CDCl_3$) spectrum of **1.6**.

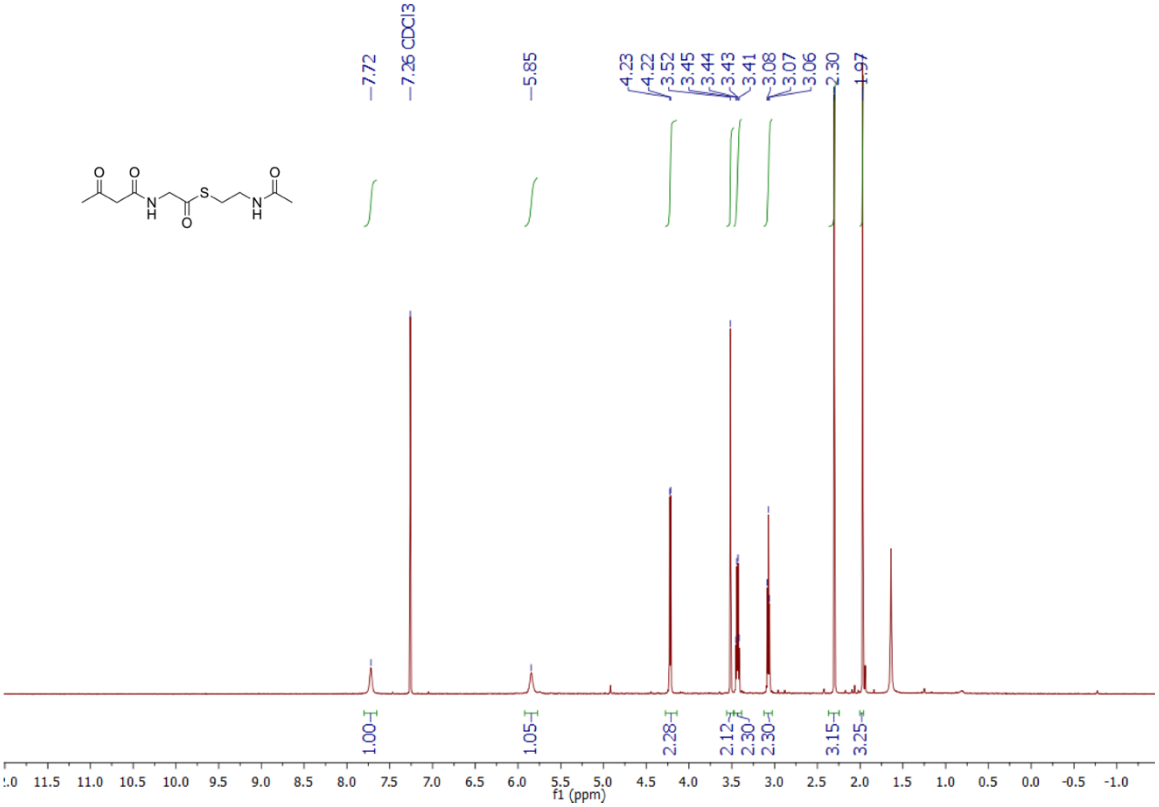


Fig. 5.15: ^{1}H NMR (500 MHz, $CDCl_3$) spectrum of **1**.

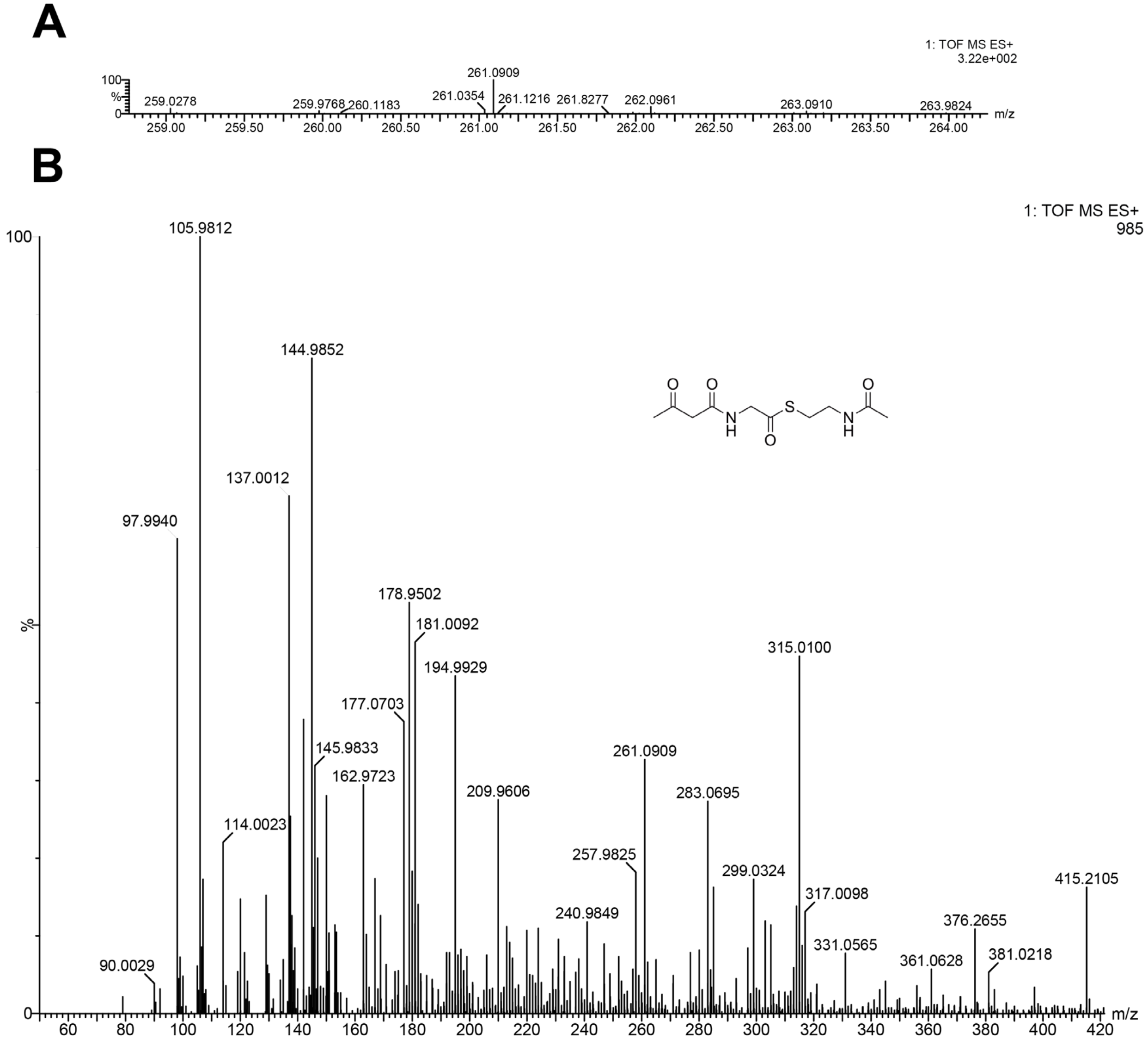


Fig. 5.16: High-resolution mass spectrum of 1. (A) Single mass analysis of **1** (calculated mass = 261.0909 Da; observed mass = 261.0909; ppm error = 0.0). (B) Full spectrum mass analysis of **1**. Both spectra were collected using a Waters Synapt G2-Si ESI-MS equipped with a quadrupole time-of-flight detector.

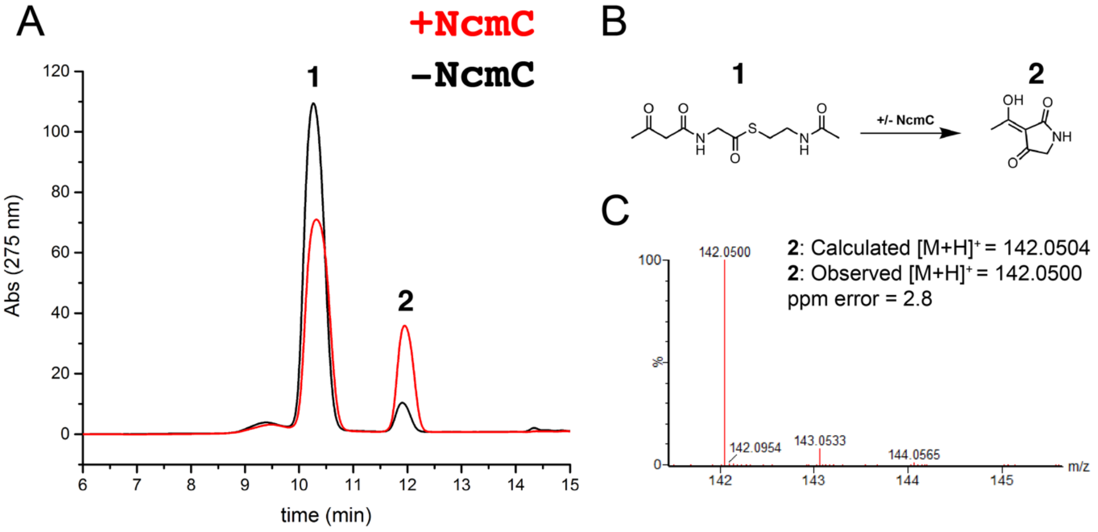


Fig. 5.17: LC-MS analysis of the NcmC catalyzed Dieckmann condensation. (A) HPLC trace of the reaction with or without enzyme demonstrates an NcmC dependent rate enhancement. (B) Reaction scheme. The hydrolytic product **3** was also observed but not shown here (Fig. 2A). (C) High resolution mass spectrum of **2** was collected on a Waters Synapt G2-Si ESI-MS equipped with a quadrupole time-of-flight detector.

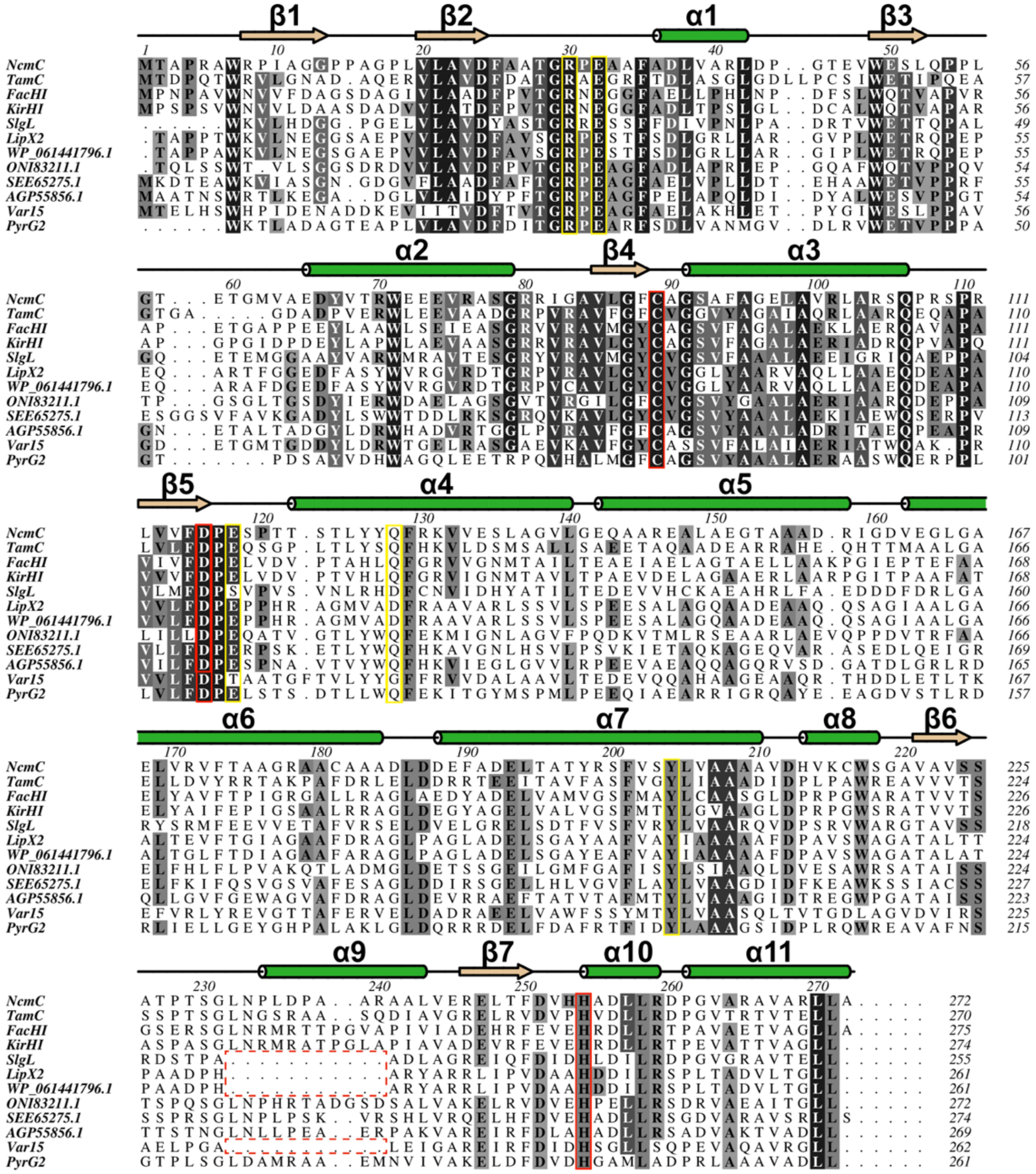


Fig. 5.18: Multisequence alignment of NcmC homologs with overlaid secondary structure using Aline.[207] Positions selected for mutation and constituting part of the Cys-Asp-His catalytic triad are boxed in red. Additional residues selected for mutation are boxed in yellow. The deletion of helix α9 in SlgL, LipX2, WP_061441796.1 (LipX2 ortholog), and Var15 is highlighted.

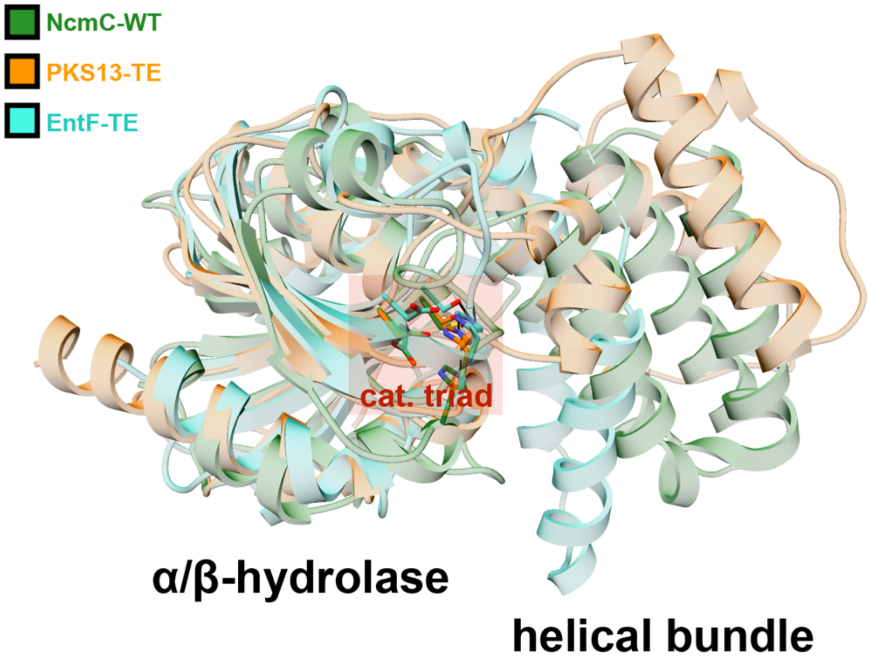


Fig. 5.19: Superimposition of NcmC and TE domains of EntF (PDB 3TEJ; Z-score of 16.7, RMSD of 3.7 Å over 207 aligned Cα atoms) and PKS13 (PDB 5V42; Z-score of 15.9 RMSD of 4.0 Å over 219 aligned Cα atoms).[173,174] Structural conservation is limited primarily to the catalytic α/β-hydrolase subdomain and not the helical-bundle subdomain.

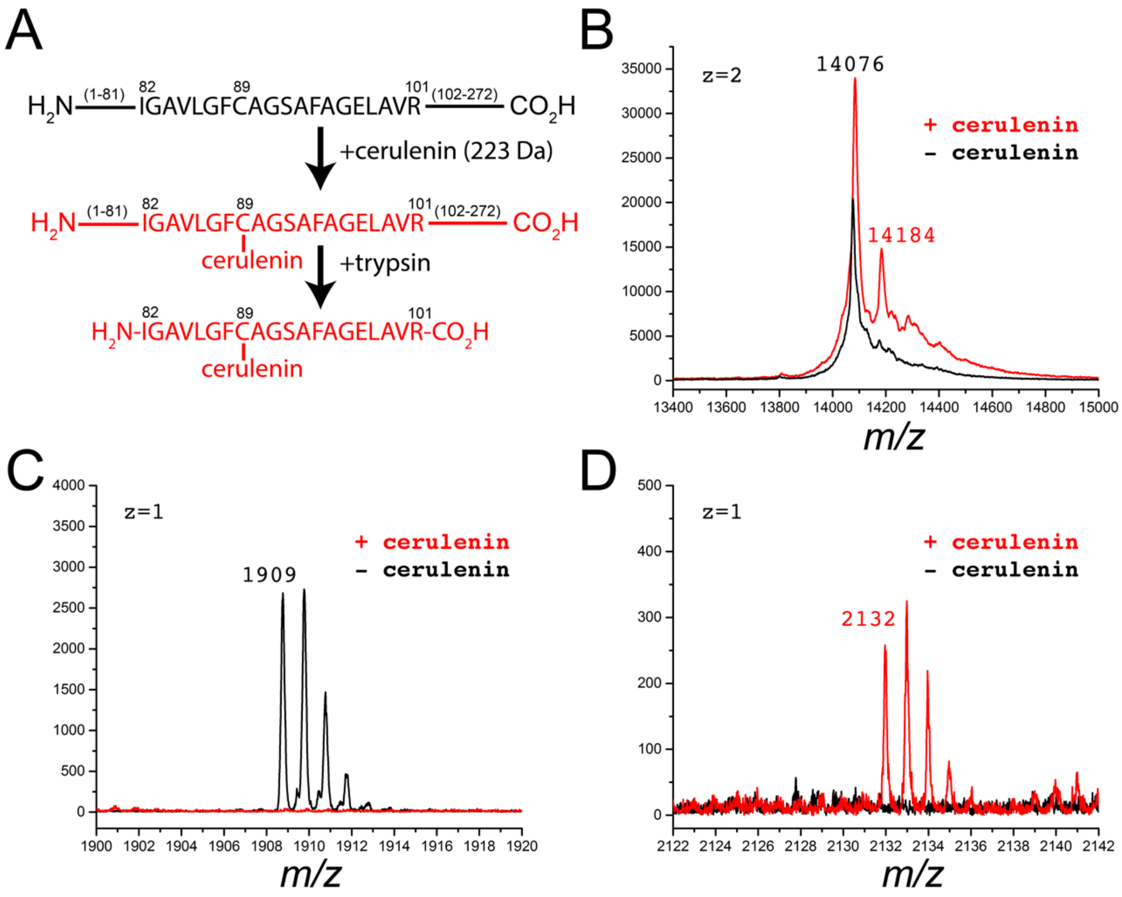


Fig. 5.20: Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) analysis of cerulenin-modified NcmC. (A) Schematic representation of the modification reaction. The region of interest (82-101) of NcmC, containing the catalytic Cys89, is represented in sequence form, and flanking N/C-terminal residues are represented as thick lines. (B) NcmC (28.5 kDa) was treated with cerulenin (223 Da) and subjected to MALDI-MS analysis, (C,D) or additionally treated with trypsin and analyzed by MALDI-MS. A mass shift of 108 Da was observed in the full-length NcmC+cerulenin reactions (z=2), and a mass shift of 223 Da was observed corresponding to the tryptic fragment (82-101) covalently modified with cerulenin (z=1).

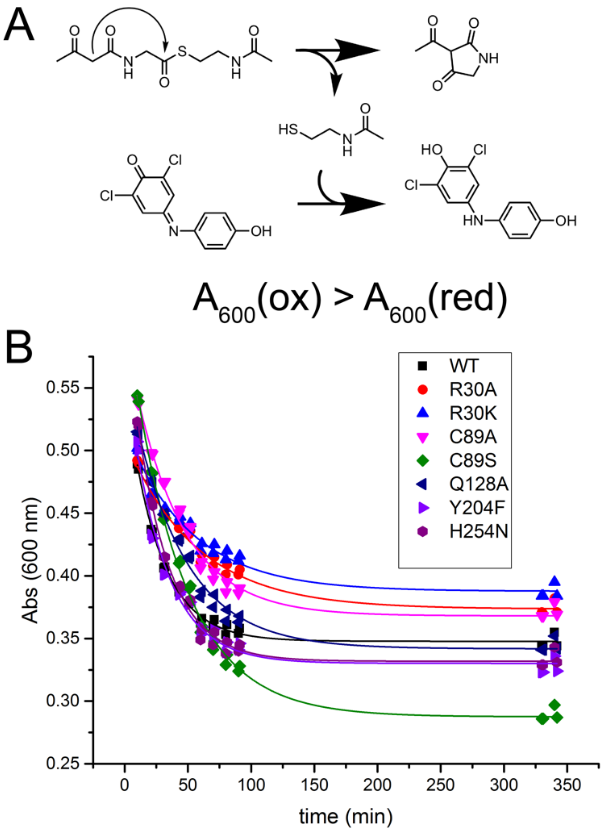


Fig. 5.21: Spectrophotometric monitoring of NcmC reactions with 1 by a 2,6-dichlorophenolindophenol (DCPIP) coupled assay.[186,187] Generation of a free thiol (i.e. *N*-acetylcysteamine) by tetramate formation (or hydrolysis; not shown) reduces DCPIP in the reaction mixture to afford the colorless, reduced form of DCPIP and attenuating absorbance at 600 nm (A_{600}). Reaction progress was monitored over time for each NcmC variant using an Epoch Microplate Spectrophotometer (BioTek Instruments, Inc.) in a 384-well plate format. Differences in absorbance at individual time points between NcmC variants is attributed to differences in protein-induced reduction of DCPIP by Cys89, Cys181, and/or Cys216 (all surface-exposed).

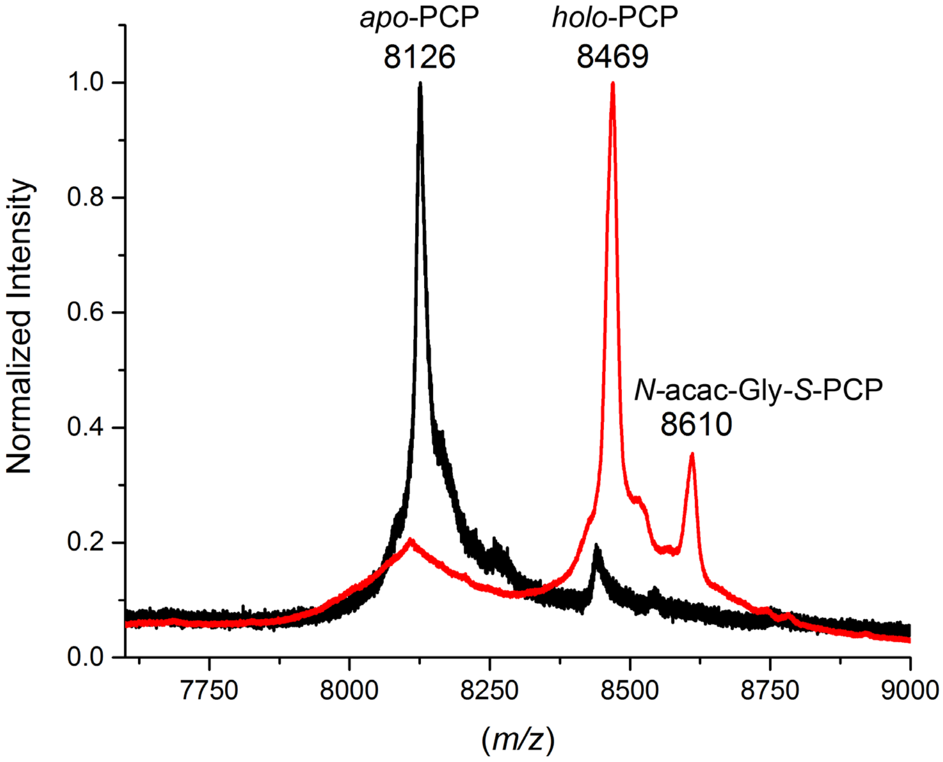


Fig. 5.22: MALDI-MS analysis of the Sfp-catalyzed phosphopantetheinylation of *apo*-PCP using **1.7** (mixture of CoA and *N*-acetoacetyl-glycyl-*S*-CoA) as acyl-donor substrate.

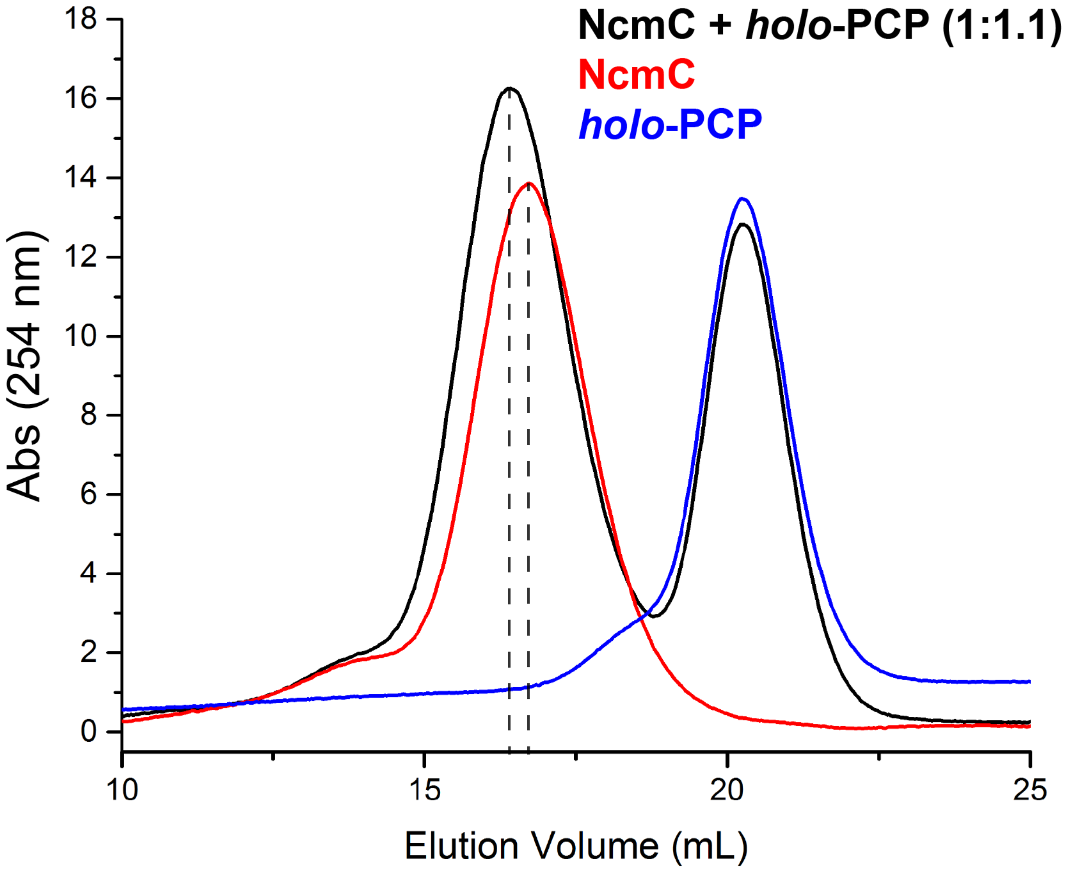


Fig. 5.23: Probing complex formation between NcmC and *holo*-PCP by analytical size-exclusion chromatography. Pre-incubating NcmC and *holo*-PCP before analyzing by gel-filtration results in their co-migration (first black peak) evidenced by relative absorbances and a shorter elution time relative to NcmC alone (red), signifying complex formation. See SI Material and Methods for detailed procedure.

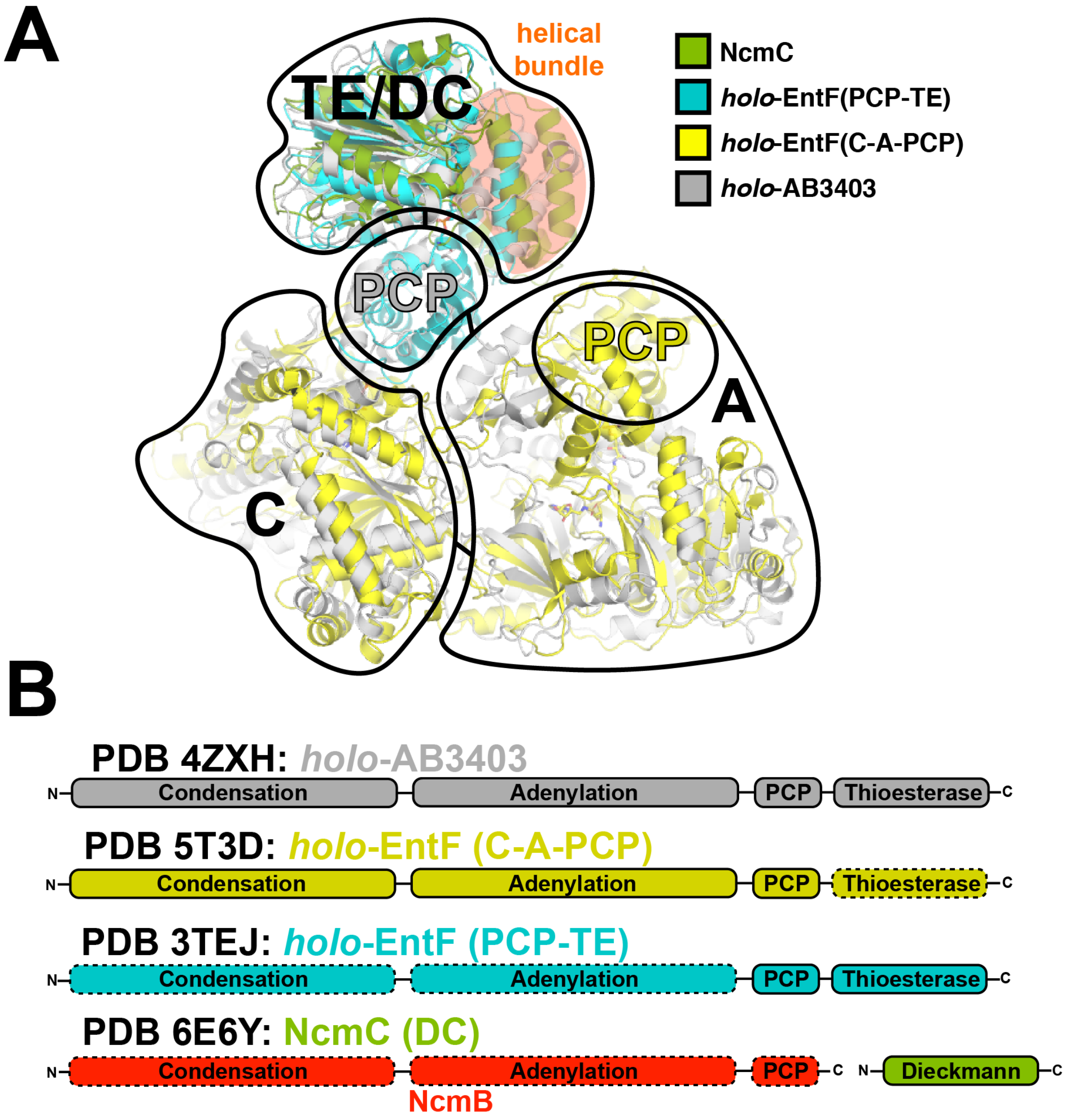


Fig. 5.24: Model of a trans-acting DC (NcmC) interacting with an NRPS. (A) NcmC (DC), *holo*-EntF (PCP-TE), and *holo*-EntF(C-A-PCP) were superimposed onto *holo*-AB3403 (see panel B for PDB IDs).[173,175] Note the highly separated PCP domains (~38 Å) reflecting different states of the NRPS catalytic cycle. The helical bundle regions of EntF and NcmC are highlighted in orange and are mostly proximal to the PCP. (B) Domain organization of the NRPSs used in structural alignment. Solid and dashed lines represent the domain regions for which there is structural resolution or not, respectively. NcmC is the only case where the off-loading domain is encoded as a separate polypeptide.

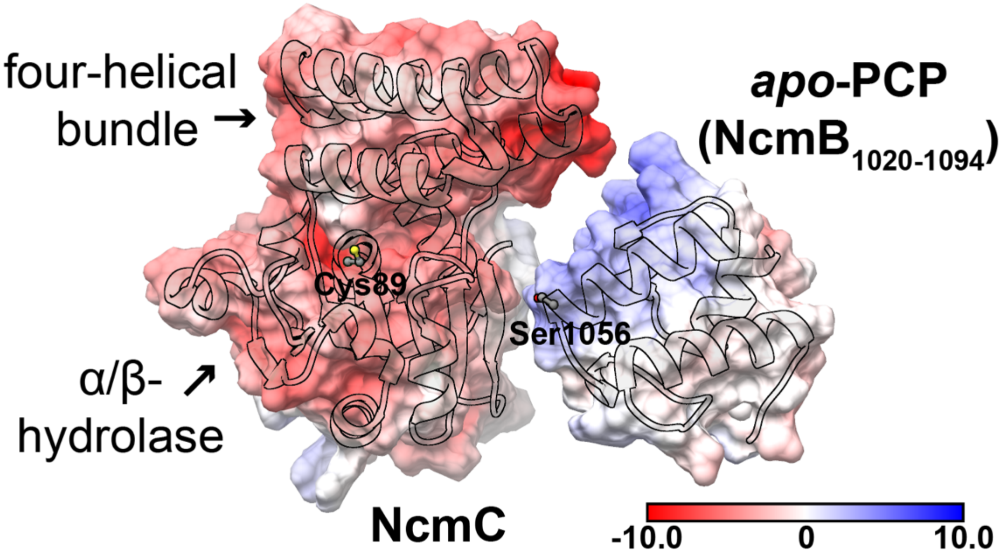
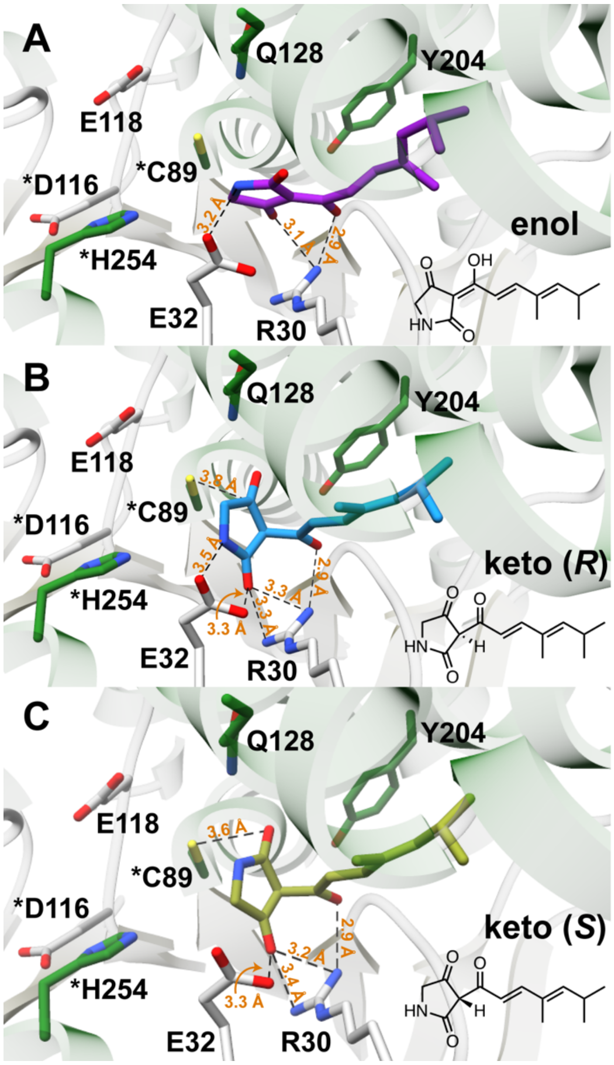


Fig. 5.25: Electrostatic complementarity between NcmC and *apo*-PCP (NcmB$_{1020\text{-}1094}$). Electrostatic potential maps were generated using the Adaptive Poisson-Boltzmann Solver (APBS) functionality in UCSF Chimera 1.13.[208,209] Active site residues Cys89 and Ser1056 in NcmC and NcmB, respectively, are represented as ball and sticks. The coordinates were generated by superposing a SWISS-MODEL-generated homology model of *apo*-PCP (NcmB$_{1020\text{-}1094}$) onto the PCP-domain of EntF in the previous structural alignment (Fig. S17) and translating NcmC and *apo*-PCP (NcmB$_{1020\text{-}1094}$) apart (~10 Å, without rotation) to reveal the complementary charged surfaces at the putative interface.



Enol	
Pose	Affinity (kcal/mol)
1	-7.7
2	-7.5
3	-7.3
4	-7.3
5	-7.1
6	-7.0
7	-7.0
8	-7.0
9	-6.6
10	-6.4

Keto (*R*)	
Pose	Affinity (kcal/mol)
1	-7.4
2	-7.1
3	-7.0
4	-7.0
5	-6.9
6	-6.6
7	-6.6
8	-6.5
9	-6.4
10	-6.4

Keto (*S*)	
Pose	Affinity (kcal/mol)
1	-7.4
2	-7.3
3	-7.0
4	-7.0
5	-6.9
6	-6.9
7	-6.9
8	-6.6
9	-6.5
10	-6.4

Fig. 5.26: Docking analysis of NcmC keto/enol truncated tetramate products. Docking was performed using the AutoDock/Vina plugin for PyMOL.[77] Lowest energy poses are represented as sticks in purple (enol), light blue (*R*-keto), or chartreuse (*S*-keto) and top 10 poses are tabulated. NcmC active site residues are colored in green or light grey and catalytic triad residues are marked with an asterisk.

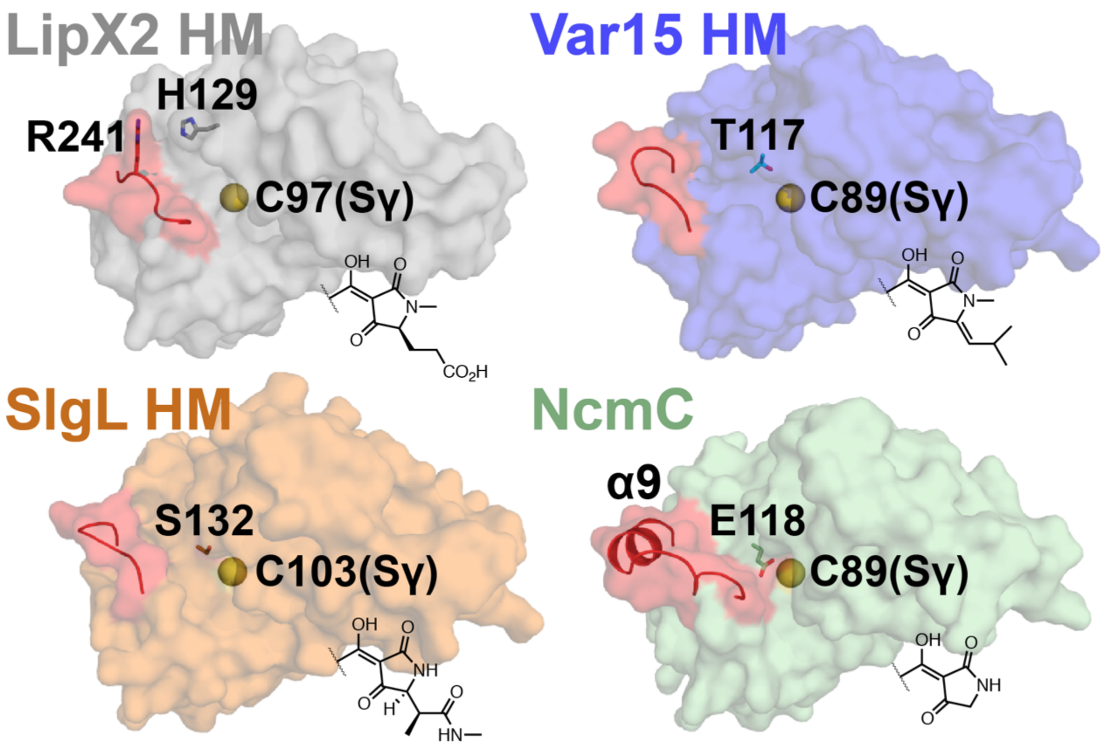


Fig. 5.27: Template-guided homology models (HMs) of NcmC homologs. HMs for LipX2 (grey), Var15 (blue), SlgL (orange), and the crystal structure of NcmC (green) with key residues in the putative aminoacyl binding site represented as sticks. The region spanning helix α9 in NcmC and equivalent regions in the HMs are colored red to highlight the difference in surface area (~800 Å^2 less area on average, compared to NcmC). The Sγ atoms from catalytic Cys residues are represented as yellow spheres. 2D structures of corresponding tetramate final products are drawn to the bottom-right of each protein.

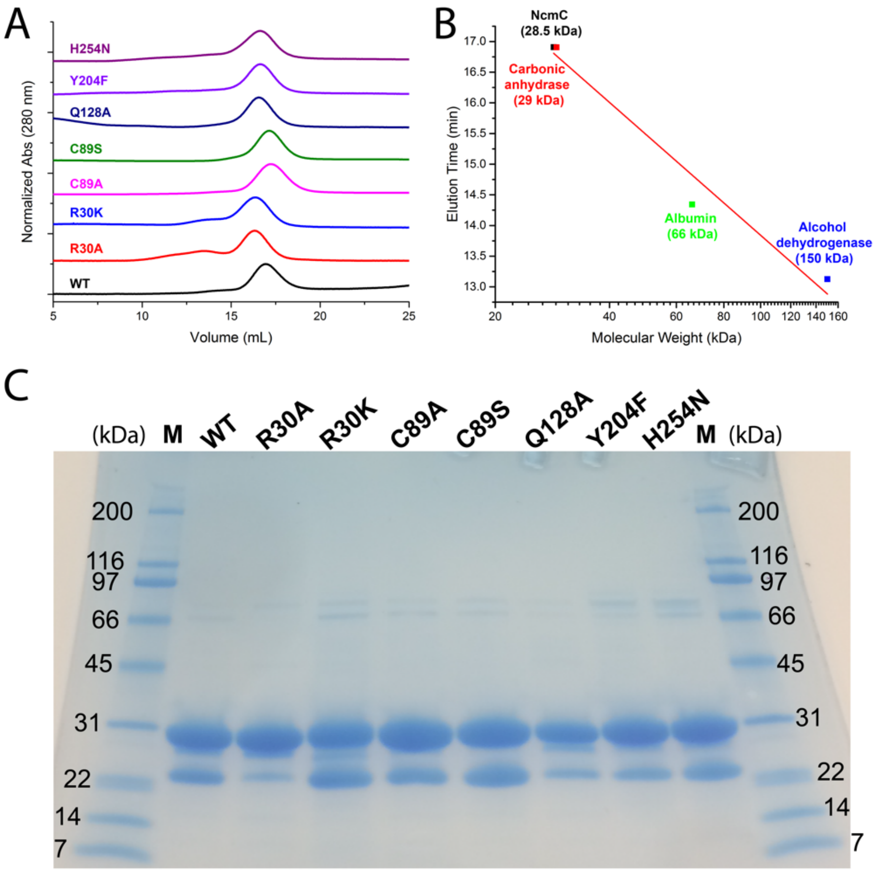


Fig. 5.28: (A) Individual SEC chromatograms for NcmC variants. (B) Standard curve (adjusted R^2=0.96) generated by running similar SEC with protein standards (Millipore Sigma), indicating NcmC elutes as a monomer. (C) SDS-PAGE analysis of NcmC variants. "M" denotes the SDS Broad Range Ladder (Bio-Rad Laboratories, Inc.) protein standard. Predominant bands correspond to NcmC variants (28.5 kDa). All NcmC variants contain His$_6$-SUMO impurity (between 22 and 14 kDa bands) which co-eluted in all stages of purification following its removal by TEV protease addition.

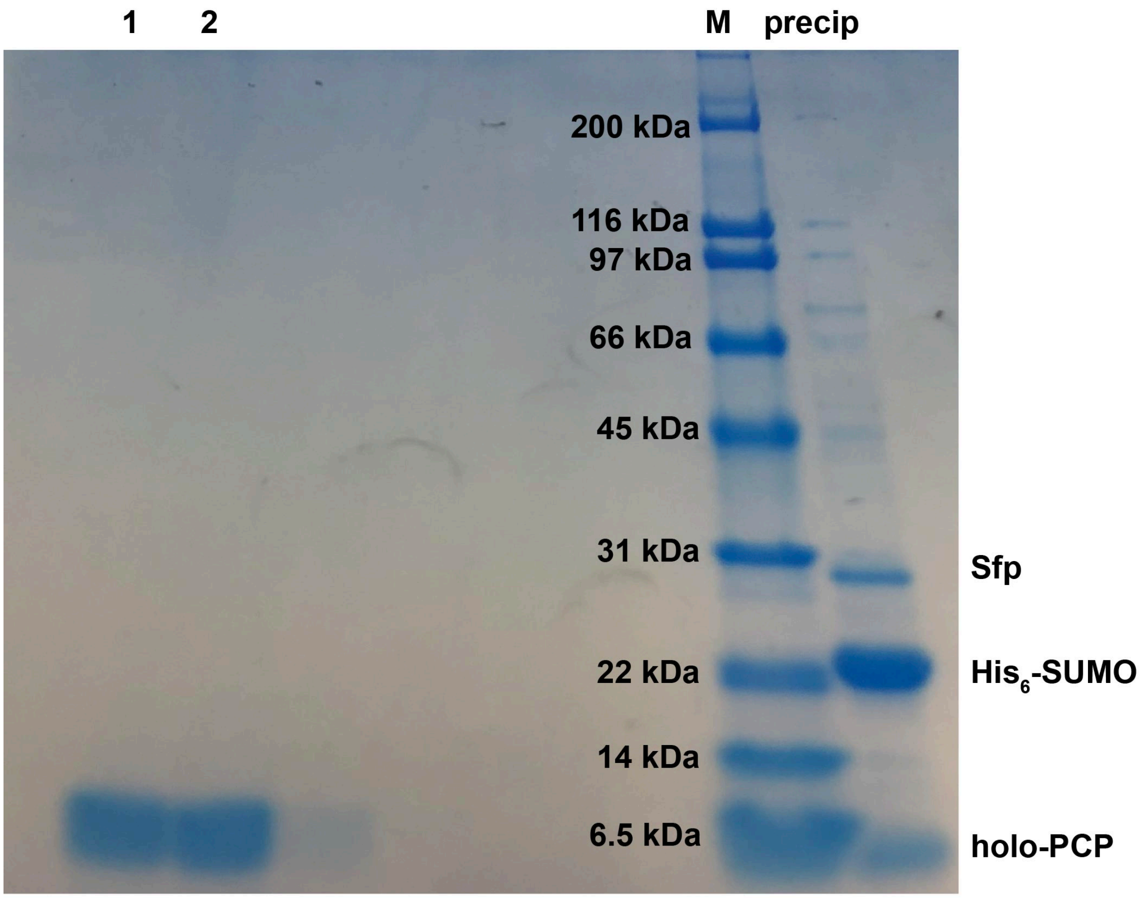


Fig. 5.29: SDS-PAGE analysis of *holo*-PCP standard. Lanes “1” and “2” correspond to the first two desalting fractions containing *holo*-PCP. “M” denotes the SDS Broad Range Ladder (Bio-Rad Laboratories, Inc.) protein standard, and “precip” denotes the precipitate fraction following acetone addition (see SI Materials and Methods for detailed purification procedure).

Table 5.1: Annotation of nodes in the sequence similarity network (SSN). Adjacent PKS/NRPS genes were determined by RODEO analysis [210]. Color coding is consistent with the nodes in Figure 4A. Adjacent PKS/NRPS genes are listed as upstream (up) or downstream (down) to the gene of interest. Reversal of directionality in the adjacent PKS/NRPS open reading frames is denoted as "-rev".

Cluster	UniProt accession	GenBank accession	Gene name	Characterized or putative (put.) natural product	Adjacent PKS/NRPS	PKS/NRPS genes in cluster
⬛ 1	A0A0F4K5Z7	WP_045933948.1			PKS (up)	PKS, NRPS
🟩 1	A0A0M8V9Z7	WP_053663421.1			NRPS (down)	PKS
🟩 1	A0A0Q8ZXB5	WP_057581822.1			NRPS (down)	PKS
🟩 1	A0A0K9XBY7	WP_049717535.1			NRPS (down)	PKS
🟩 1	A0A1Q5D7W2	WP_073865700.1			NRPS (down)	PKS
🟩 1	A0A1Q5JDB0	WP_073865700.1			NRPS (down)	PKS
⬛ 1	A0A0C1X845	WP_041991951.1				PKS, NRPS
🟩 1	A0A0N6ZGX9	WP_061441807.1			NRPS (down)	PKS
🟩 1	A0A1Q5MN15	WP_073920103.1			NRPS (down)	PKS
🟩 1	A0A1X9WEN9	WP_033435885.1	*ncmC*	nocamycin	NRPS (down)	PKS
⬛ 1	A0A1H2H606	WP_063274578.1				PKS, NRPS
⬛ 1	A0A1W2LL55	WP_063274578.1				PKS, NRPS
🟩 1	A0A290Z691	WP_096493684.1			NRPS (down)	PKS
🟩 1	A0A0A0NF48	AGP55856.1			NRPS (down)	PKS
🟩 1	F1DI35	WP_108908231.1	*trdC*	tirandamycin	NRPS (down)	PKS
🟩 1	A0A1H5KKH3	SEE65275.1			NRPS (down)	PKS
🟩 1	A0A2M8XPG9	SEE65275.1			NRPS (down)	PKS
🟩 1	D3Y1I7	ADC79641.1	*tamC*	tirandamycin	NRPS (down)	PKS
🟩 1	A0A1Y2NJQ2	OSY47118.1			PKS (up), NRPS (down)	PKS
🟩 1	A0A1V2Q4P3	ONI83211.1			NRPS (down)	PKS
🟩 1	A0A0P4R466	GAO07620.1			PKS (up), NRPS (down)	PKS
🟩 1	A0A1C6N4C4	SCK25965.1			PKS (up), NRPS (down)	PKS
🟩 1	A0A1V3ZZ37	WP_077974123.1			NRPS (down)	PKS
🟩 1	E4NIK2	BAJ32800.1	*facHI*	factumycin	NRPS (down)	PKS
🟩 1	B0B508	CAN89637.1	*kirHI*	kirromycin	PKS (up), NRPS (down)	PKS
🟩 1	S5VFF6	CAN89637.1	*kirHI*	kirromycin	PKS (up), NRPS (down)	PKS
⬜ 1	A0A1D2I704	ODA70255.1			PKS (up)	PKS
🟩 2	A0A1D7VJN8	WP_069568946.1		streptolydigin (put.)	NRPS (down)	PKS
🟩 2	A0A2G6YFS2	WP_033271195.1	*slgL*	streptolydigin	NRPS (down)	PKS
🟩 2	D1GLT3	WP_033271195.1	*slgL*	streptolydigin	NRPS (down)	PKS
🟨 2	A0A081XIT3	WP_037940301.1				NRPS
🟨 2	A0A1Q4ZI93	WP_073719472.1				NRPS
🟨 2	A0A1Q5FGX4	WP_073878011.1				NRPS
🟨 2	A0A1H9JGW2	SEQ85765.1				NRPS
🟨 2	A0A0N9LX19	ALG65322.1	*cal14*	WS9326		NRPS
🟩 2	A0A1W5XJW7	WP_084770204.1		streptolydigin (put.)	NRPS (down)	PKS
🟨 2	D9XLE7	EFL41629.1				NRPS

2	A0A1C5GDX1	WP_089001688.1			NRPS (down)	PKS, NRPS
2	A0A1C6U8U6	WP_091621404.1			NRPS (down)	PKS, NRPS
2	A0A0F4K9S3	KJY42834.1				
2	A0A2N3ZBK3	WP_101414461.1			NRPS (down)	PKS, NRPS
3	S3BZ19	EPD93919.1			NRPS (down)	
3	A0A0N0HWT3	WP_031032748.1			NRPS (down)	
3	A0A1D8BYT1	AOS63965.1			NRPS (down)	
3	A0A2M9IFQ0	WP_100570700.1			NRPS (down)	PKS, NRPS
3	A0A0P4RCI1	WP_042159217.1			NRPS (down)	PKS, NRPS
4	A0A243S3G0	WP_086602300.1			NRPS (down)	
4	Q9FB38	AAG02344.1	none	bleomycin	NRPS (down)	NRPS
4	A0A0T6LQ39	WP_018384057.1			NRPS (down)	NRPS
4	M3B3U0	WP_004942709.1			NRPS (down)	NRPS
5	A0A2G7F626	WP_099931426.1				PKS, NRPS
5	Q30CS4	ABB05100.1	*lipX2*	lipomycin	NRPS (down)	PKS
5	A0A0N1NPE0	WP_054213280.1				NRPS
5	A0A0N6W4Z6	WP_061441796.1			NRPS (down)	PKS
6	A0A1H9XTL3	WP_093781955.1				PKS
6	A0A2I6SBD0	AUO14856.1		vancoresmycin (put.)		PKS, NRPS
6	R1HZ43	WP_003069915.1	*var15*	vancoresmycin		PKS, NRPS
7	A0A248YPN6	WP_095565839.1			NRPS (down)	PKS
7	A0A2I0SD64	WP_103554186.1			NRPS (up-rev)	PKS
7	K7QRJ3	AFV71307.1	*pyrG2*	pyrroindomycin	NRPS (up-rev)	PKS
8	R4LG32	WP_015620564.1			PKS (up), NRPS (down)	PKS
8	A0A101N5D4	WP_067010779.1			NRPS (down)	PKS
8	A0A2N8MH05	WP_101607287.1			NRPS (down)	PKS
9	A0A2A2D3K0	WP_095583395.1			PKS (up-rev), NRPS (down)	PKS, NRPS
9	A0A1Z2KZ73	WP_087925822.1			PKS (up-rev), NRPS (down)	PKS, NRPS
9	A0A2N8NW41	WP_102919109.1			PKS (up-rev), NRPS (down)	PKS, NRPS
10	A0A1B1ZJB9	WP_068796317.1				
10	A0A1C4RPF7	WP_100564308.1				PKS, NRPS
10	A0A2M9L770	WP_100564308.1				PKS, NRPS
singleton	A0A1Q4XTQ2	WP_073893670.1			NRPS (up)	NRPS
singleton	A0A1C6UKU8	WP_091623539.1			NRPS (down-rev)	PKS, NRPS
singleton	A0A1H9WUG6	WP_090069547.1				PKS, NRPS
singleton	A0A1C5GC05	SCG17082.1				NRPS
singleton	A0A1C6V190	SCL60065.1			NRPS (up)	PKS, NRPS
singleton	A0A285QT70	WP_097237205.1				PKS, NRPS
singleton	A0A1H9XTC3	WP_093781959.1				PKS
singleton	A0A209CPB4	WP_087771923.1				NRPS
singleton	A0A2N3UYX7	WP_101384298.1			NRPS (down)	NRPS

Table 5.1 cont. (part 2 of 2)

Table 5.2: Adenylation (A) domain specificity prediction for A-domains in tirandamycin-like biosynthetic gene clusters. Color coding is consistent with the nodes in Figure 4A. Inferences in the A-domain specificity based on the natural product structures are noted in parentheses (experimental).

Cluster	UniProt accession	GenBank accession	Gene name	Characterized natural product	A-domain specificity code	A-domain specificity prediction (experimental)
1	A0A0M8V9Z7	WP_053663421.1			D I L Q L G V I	Gly
1	A0A0Q8ZXB5	WP_057581822.1			D I L Q I G M I	Gly
1	A0A0K9XBY7	WP_049717535.1			D I L Q - G V V	N/A
1	A0A1Q5D7W2	WP_073865700.1			D I L Q V G A I	Gly
1	A0A1Q5JDB0	WP_073865700.1			D I L Q V G A I	Gly
1	A0A0N6ZGX9	WP_061441807.1			D A W E A G L V	Gln
1	A0A1Q5MN15	WP_073920103.1			D T L Q L G V I	Gly
1	A0A1X9WEN9	WP_033435885.1	*ncmC*	nocamycin	D I L Q L G V I	Gly (Gly)
1	A0A290Z691	WP_096493684.1			D I L Q L G P I	Gly
1	A0A0A0NF48	AGP55856.1			D I L Q L G V I	Gly
1	F1DI35	WP_108908231.1	*trdC*	tirandamycin	D I L Q V G V I	Gly (Gly)
1	A0A1H5KKH3	SEE65275.1			D I L Q L G V I	Gly
1	A0A2M8XPG9	SEE65275.1			D I L Q L G V I	Gly
1	D3Y1I7	ADC79641.1	*tamC*	tirandamycin	D I L Q V G V I	Gly (Gly)
1	A0A1Y2NJQ2	OSY47118.1			D I L A L G V I	Gly
1	A0A1V2Q4P3	ONI83211.1			D I L Q L G V I	Gly
1	A0A0P4R466	GAO07620.1			D I L A L G V I	Gly
1	A0A1C6N4C4	SCK25965.1			D I L Q L G V I	Gly
1	A0A1V3ZZ37	WP_077974123.1			D I L X L G V I	Gly
1	E4NIK2	BAJ32800.1	*facHI*	factumycin	D T L Q L G V I	Gly (β-Ala)
1	B0B508	CAN89637.1	*kirHI*	kirromycin	D T L Q L G V I	Gly (β-Ala)
1	S5VFF6	CAN89637.1	*kirHI*	kirromycin	D T L Q L G V I	Gly (β-Ala)
2	A0A1D7VJN8	WP_069568946.1			unknown	N/A
2	A0A2G6YFS2	WP_033271195.1	*slgL*	streptolydigin	unknown	N/A (*N*,β-dimethyl-L-Asn)
2	D1GLT3	WP_033271195.1	*slgL*	streptolydigin	unknown	N/A (*N*,β-dimethyl-L-Asn)
2	A0A1W5XJW7	WP_084770204.1			unknown	N/A
5	Q30CS4	ABB05100.1	*lipX2*	lipomycin	unknown	N/A (L-Glu)
5	A0A0N6W4Z6	WP_061441796.1			unknown	N/A

CHAPTER 6: STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF AN ITERATIVE ASPARAGINE γ-N-METHYLTRANSFERASE IN AERONAMIDE BIOSYNTHESIS

6.1 ABSTRACT

Stemming from its impressively high degree of post-translational modification (PTM), the proteusin natural product, polytheonamide B (pTB), presents fertile ground in areas of synthetic biology and biosynthetic engineering of peptidic drugs. In addition to its extreme cytotoxic potency, the 49-residue modified-peptide assumes a unique β-helical structure giving rise to its unimolecular ion-channel function. With *N*- and *C*-methylations at the exterior providing a hydrophobic coating, the 0.4 x 4.5 nm β-helical barrel spans cellular membranes allowing passive diffusion of monovalent cations, thereby disrupting electrostatic potentials; the primary driver of its toxicity. Structural features of the β-helical toxin include alternating α-epimerizations, D-Asn N_γ-methylations, β-*C*-methylations, as well as β-*C*-hydroxylations. The former two have been highlighted as early-stage modifications and most critical for ion-channel activity, however, the minimal set of PTMs necessary for ion-channel function remains to be definitively understood. Here we take a detailed look into the D-Asn N_γ-methylation stage of aeronamide biosynthesis (a terrestrial ortholog of pTB) providing insight both into the mechanism of iterative D-Asn N_γ-methylation as well as ion-pore maturation. We show that the *N*-methyltransferase, AerE, penta-*N*-methylates its peptidic substrate with pseudo-(N-to-C)-directionality. Finally, our structural and biochemical data support a recent *in silico* analysis which points to the importance of *N*-methylation in maintaining β-helicity.[211]

6.2 INTRODUCTION

The prototypical proteusin, polytheonamide B (pTB), has been isolated from a bacterial symbiont of the marine sponge *Theonella swinhoei* and is among the most highly post-translationally modified peptides known.[212–216] Just seven enzymes have been implicated in carrying out 42 post-translational modifications (PTMs) within a 49-residue core peptide (CP) resulting in 18 α-epimerizations, 4 β-*C*-hydroxylations, 12 β-*C*-methylations, and 8 *N*-methylations collectively transforming the peptide into a bioactive 6.3-residue-turn β-helical structure.[217] The resultant wide-bore (0.4 nm) helix has been shown to facilitate ion passage of monovalent cations (H^+, Na^+, K^+, Rb^+, and Cs^+) by insertion of pTB into planar bilayers, suggesting a toxification strategy by disruption of membrane potentials.[214] This hypothesis was recently tested by administration of fluorescently labeled pTB to mammalian cells, revealing a two-pronged mechanism whereby membrane depolarization is accompanied by endocytic internalization of pTB.[218] Specifically, pTB localizes to lysosomes inside MCF-7 human breast cancer cells, triggering the transportation of H^+ across the organellar membrane and neutralization of the pH gradient necessary for cellular homeostasis.

A new proteusin, aeronamide, from the free-living, terrestrial bacterium *Microvirgula aerodenitrificans* DSM 15089 is currently under investigation and putatively assigned to be an ortholog of pTB in terms of its structure and function. Similar to the pTB biosynthetic machinery, *M. aerodenitrificans* DSM 15089 harbors corresponding genes encoding for the radical *S*-adenosyl methionine (SAM) epimerase (*aerD*), SAM-dependent *N*-methyltransferase (NMT) (*aerE*), LanM-like dehydratase (*aerF*), B12-dependent *C*-methyltransferase (*aerB*), and C1A cysteine protease (*aerH*) in the aeronamide biosynthetic gene cluster (BGC) (Fig. 6.1a-c).

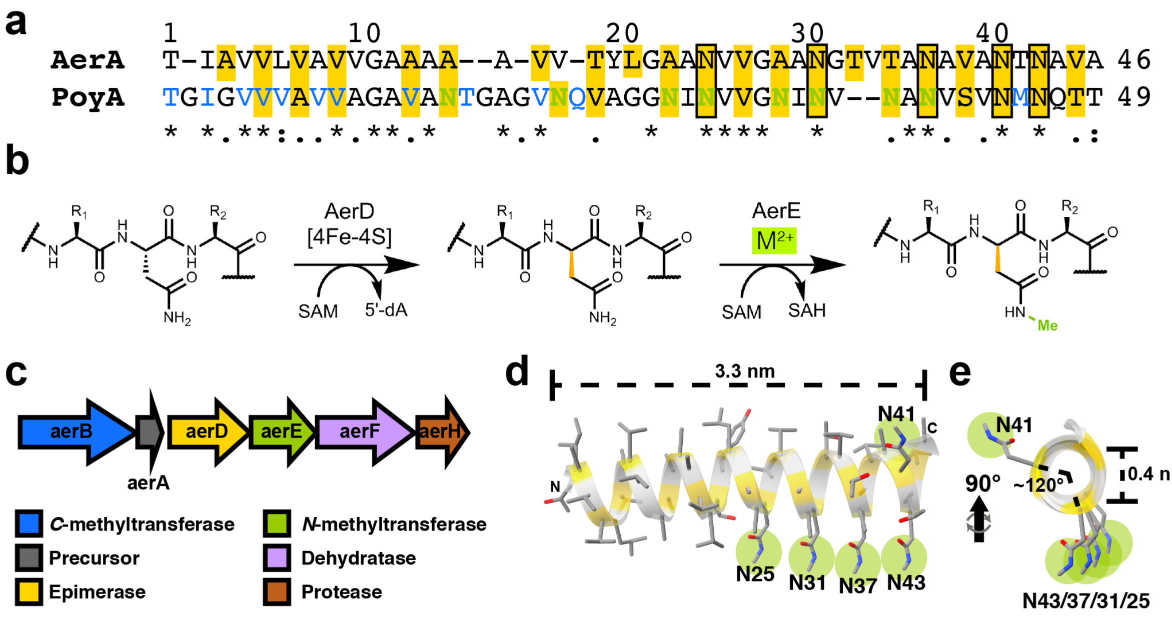


Fig. 6.1: Core peptide alignment, AerD/AerE-catalyzed modifications, aeronamide biosynthetic gene cluster, and β-helical model of aeronamide. a, AerA/PoyA core peptide alignments with epimerized residues boxed in yellow. Sites of *C*-methylation and *N*-methylation are colored blue and green, respectively. The five D-Asn residues methylated in AerA are fully conserved and boxed in black. **b,** Reaction scheme depicting the unidirectional α-epimerizations installed by the radical SAM-dependent, [4Fe-4S]-containing epimerase, AerD, and D-Asn N_γ-methylations catalyzed by SAM *N*-methyltransferase, AerE (R_1=Ala or Thr side chains; R_2=Val, Ala, Thr side chains, or H [Gly]). **c,** Aeronamide biosynthetic gene cluster. **d-e,** Model of β-helical aeronamide in two, 90°-related, orientations.

pTB biosynthesis, the activity of NMT PoyE has been suggested to parallel the timing of α-epimerizations catalyzed by the radical SAM-dependent epimerase, PoyD. *In vitro* experiments demonstrate that an epimerized PoyA substrate is sufficient for full octa-*N*-methylation catalyzed by PoyE and *N*-methylation efficiency corelates with extent of epimerization.[217]

What has remained unclear is how and when during proteusin maturation the β-helicity is established, and what are the minimal set of PTMs sufficient for ion-channel activity. Molecular dynamics simulations point to the importance of *N*-methylations in maintaining the co-axial H-bond network after systematically removing PTMs in pTB and evaluating hypothetical precursors for their stability *in silico*.[211] Future design and enzymatic syntheses of proteusin analogs with tailored membrane selectivities will rely on a thorough understanding of the substrate tolerances and requisite chemical features for bioactivity. Thus, detailed biochemical and biophysical characterization of the critical biosynthetic enzymes will be highly important in meeting these aims.

In this study, we look into the mechanism of AerE-catalyzed *N*-methylation of five D-Asn residues in the aeronamide CP. We provide evidence that AerE utilizes SAM to iteratively methylate the lowly-nucleophilic γ-nitrogen of D-Asn side-chains by a putative structural-ion mediated Lewis acid type activation. Guided by cocrystal structures of AerE in complex with its epimerized AerA substrate, our data also support notions from Renevey et al. that β-helicity is driven by *N*-methylation.[211]

6.3 RESULTS

6.3.1 In vitro penta-N-methylation

To confirm the expected Asn-*N*-methyltransferase activity of AerE *vitro*, we heterologously expressed an epimerized His$_6$-tagged AerA substrate (AerA$_{DL}$) in *E. coli* by co-expression with the radical SAM epimerase AerD, as previously described.[217] Purification of AerA$_{DL}$ under denaturing conditions afforded a putative substrate of AerE with alternating epimerizations along the CP (by homology to pTB). AerE was similarly expressed in *E. coli* as an N-terminal SUMO fusion-protein and purified to homogeneity as a tag-free monomer by size-exclusion chromatography (SEC). Following incubation of AerE with AerA$_{DL}$ in the presence of SAM and trypsinization of the peptide products, we observed a +70 Da mass shift in the CP-containing fragment by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), suggestive of full penta-*N*-methylation catalyzed by AerE (Fig. 6.2a,b). Compared to *in vitro* methylation assays with PoyA and the PoyE NMT lasting upwards of 10 days to observe full octa- *N*-methylation, it is noted that penta-*N*-methylation of AerA_{DL} required much shorter (1.5 h) incubation times.

6.3.2 Leader peptide and partial core peptide independence

To delineate a minimal substrate of AerE, we took advantage of previous observations that core-epimerizations protect PoyA against proteinase K digestion to yield LP-free PoyA CP-fragments.[217] Correspondingly, incubation of AerA_{DL} with proteinase K and subsequent HPLC purification provided a panel of AerA_{DL} CP-fragments ranging from 31-35 out of 46 CP-residues and containing all five D-Asn methyl-acceptors (Fig. 6.2a). Owing to the difficulty in separating these fragments by reversed-phase HPLC with C4/C18 resins, we incubated them as an HPLC-purified mixture with AerE and SAM. Following overnight incubation gave +70.2 Da mass shifts for all five species, thereby revealing full leader-peptide and partial core-peptide independence of AerE (Figs. 6.2c and 6.6). To confirm this observation with a homogenous and fully epimerized substrate, we synthesized the AerA_{DL}(34mer) peptide (Fig. 6.2a) containing all appropriate D-residues by solid-phase peptide synthesis (SPPS) and again observed full penta-*N*-methylation, supporting these minimal substrate requirements (Fig. 6.5a,b).

6.3.3 Progress and directionality of AerE D-Asn-N-methylation

The progress of D-Asn-*N*-methylation was probed by heat-precipitating the enzyme at various time points between 0-16.5 hours and quantifying relative peak intensities of methylated AerA_{DL} intermediates by MALDI-TOF MS. Comparing the normalized peak intensities of mono-, di-, tri-, tetra-, and penta-*N*-methylated AerA_{DL} over time is indicative of a loss in processivity of AerE as more extensively methylated AerA_{DL} intermediates appear to have longer lifetimes (Fig. 6.2d,e). Given the time-dependent relative populations of mono- through penta-*N*-methylated

AerA$_{DL}$, after reaction heat-quenching and trypsinization, we HPLC-isolated mono-, di-, and tri-*N*-methylated AerA$_{DL21}$ intermediates to study the directionality of AerE by electrospray ionization (ESI) tandem MS (MS/MS) (Fig. 6.3). In the case of mono- *N*-methylated AerA$_{DL21}$, we predominantly observed y-ions with no mono-*N*-methylation beyond D-Asn25 and only two y-ions (y19+Me/y19-NH$_3$+Me) with methylation of a fragment born of cleavage between D-Asn25 and D-Asn31, suggesting D-Asn25 is an early methyl acceptor. Fragmentation of di-, and tri-*N*-methylated AerA$_{DL21}$ reveals a lack of regioselectivity with respect to D-Asn31, D-Asn37, and D-Asn41 for the second and third methylation events. Importantly, in all cases, only one y-ion (y5+Me) was observed in the tri-*N*-methylated AerA$_{DL21}$ species corresponding to methylation at D-Asn43, suggesting it is the slowest to be methylated (Fig. 6.3b). To reconcile the observed pseudo-(N-to-C)-directionality of AerE, we manually built a β-helical model of the AerA CP based on the pTB solution NMR structure (Fig. 6.1d,e).[214] Inspection of the model reveals D-Asn25, D-Asn31, D-Asn37, and D-Asn43 to exist on the same face of the β-helix with D-Asn41 separated by ~120° about the helical axis (Fig. 1d,e), indicating that D-Asn recognition by AerE may be more a function of its local residue environment than position within a β-helical structure. If β-helical structure influenced accessibility of D-Asn to AerE, then we might expect all D-Asn residues on the same face (i.e., D-Asn25, D-Asn31, D-Asn37, and D-Asn43) to be methylated before D-Asn41, but this was not the case. A potential explanation for the sluggish methylation of D-Asn43 relative to D-Asn41 is explained in light of the structural data below.

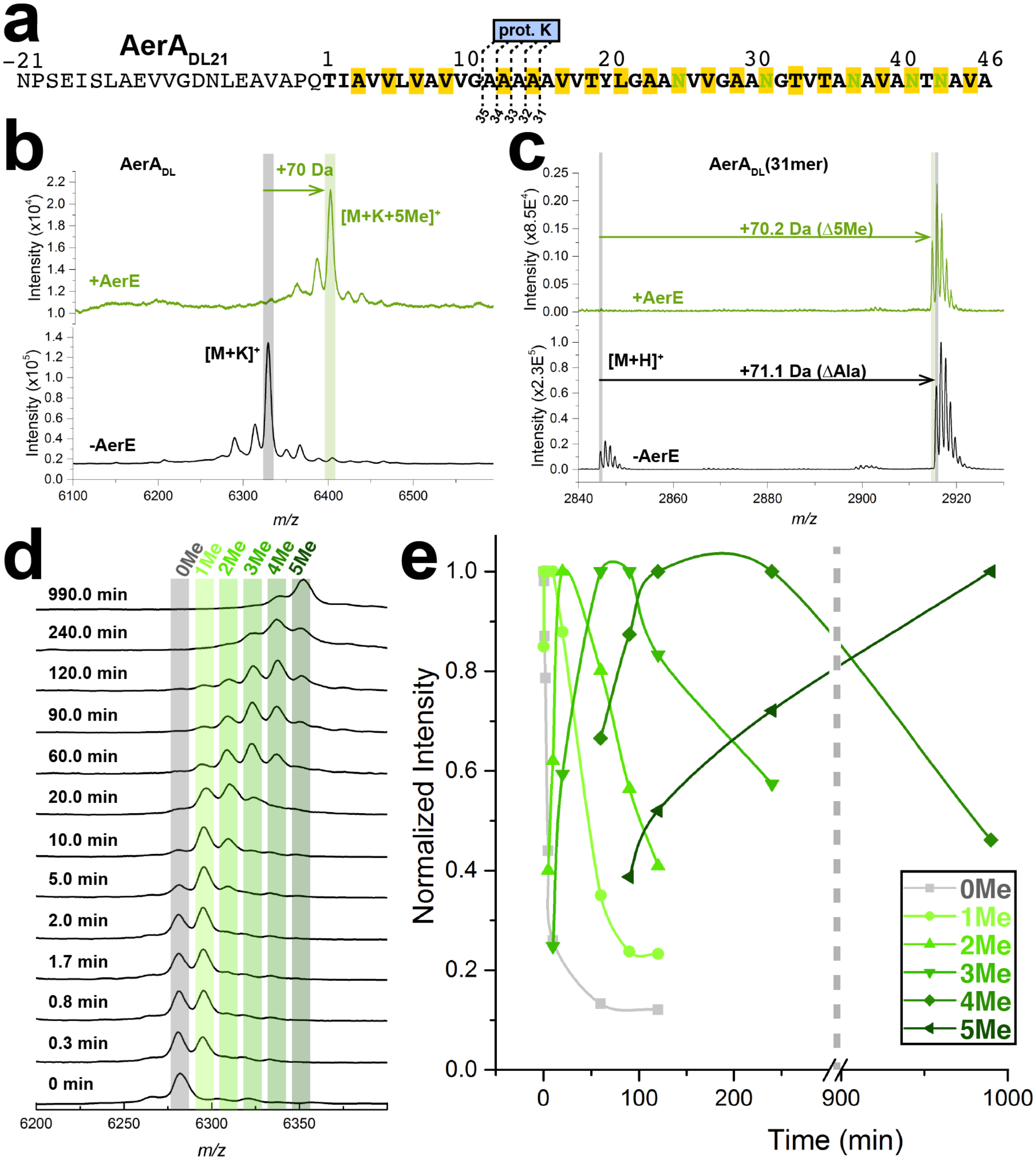


Fig. 6.2: Probing the minimal substrate and processivity of AerE-catalyzed *N*-methylation. a, Sequence representation of AerA$_{DL21}$ (tryptic fragment) with the aeronamide core peptide residues in bold and sites of proteinase K cleavage yielding 31-35mer epimerized AerA$_{DL}$ peptides. **b,** AerA$_{DL}$ treated with (green) or without (black) AerE confirms the AerE-dependent +70 Da mass shift suggesting penta-*N*-methylation observed by MALDI-TOF MS. **c,** The shortest proteinase K fragment, AerA$_{DL}$(31mer), isolated by HPLC is similarly penta-*N*-methylated in the presence of AerE (green) as observed by MALDI-TOF MS (Note: each proteinase K fragment differs by one Ala at the N-terminus corresponding to a +71.1 Da mass shift which is close in mass to that calculated for five added methyl groups [+70.2 Da], however, the mass accuracy was sufficient to resolve the monoisotopic differences). **d,** Reactions containing AerE (tag-free), AerA$_{DL}$, and SAM were heat-quenched at various time points between 0-16.5 h and treated with trypsin to analyze AerA$_{DL21}$ fragments containing 0 through 5 methylations (0Me – 5Me) by MALDI-TOF MS. **e,** AerE processivity towards AerA$_{DL}$ appears to diminish as the substrate becomes more methylated. For each time point, normalized peak intensities are plotted against time. From 300-900 min there is a break in the X-axis demarcated with a grey dotted line.

6.3.4 Overall structure of AerE and in complex with its substrates

To gain a framework for further mechanistic studies, we solved the crystal structure of AerE in complex with *S*-adenosyl homocysteine (SAH) to 1.51 Å resolution in the trigonal space group P 3_2 2 1 (Fig. 6.4a). Overall, AerE resembles two other characterized class I SAM-dependent NMTs: GenN (PDB 5U18; Dali Z-score of 24.9, root mean square deviation [RMSD] of 2.8 Å over 274 aligned Cα atoms) and PrmC (formerly HemK; PDB 2B3T; Dali Z-score of 17.9, RMSD of 3.9 Å over 216 aligned Cα atoms).[143] The former catalyzes *N*-methylation of a secondary amine in biosynthesis of the aminoglycoside gentamycin,[219] while the latter *N5*-methylates a Gln within a tripeptide motif (GGQ) of prokaryotic translation termination release factors (RF1 and RF2).[220,221] Superimposing the structures reveals two insertions in AerE (Gln23-Arg60 and Ala315-Leu350) which are lacking in GenN and PrmC; accounted for by 86 additional residues in AerE, on average (Fig. 6.7). Features of the C-terminal Rossmann-like subdomain bearing the SAM binding-site (Ile138-Ser316) are well aligned, whereas the primarily α-helical N-terminal subdomain (Met1-Glu131) only partially aligns with GenN and is highly divergent with respect to PrmC. In PrmC, this α-helical N-terminal region employs 13 residues (out of 29) to make protein-protein contact with its methyl-acceptor substrate RF1.[221] In the case of GenN, it was proposed that the N-terminal domain interfaces with other enzymes in the gentamycin BGC.[219]

To better understand the AerA CP recognition of AerE, we co-crystallized a ternary complex between AerE, AerA_{DL}(34mer), and SAH to 1.71 Å resolution in the same space group (P 3_2 2 1) (Fig. 6.4b-e). Global comparison of the two structures suggests AerE binds to its peptide substrate in a highly rigid manner with an RMSD of 0.15 Å over 2561 atoms (87% of the non-hydrogen

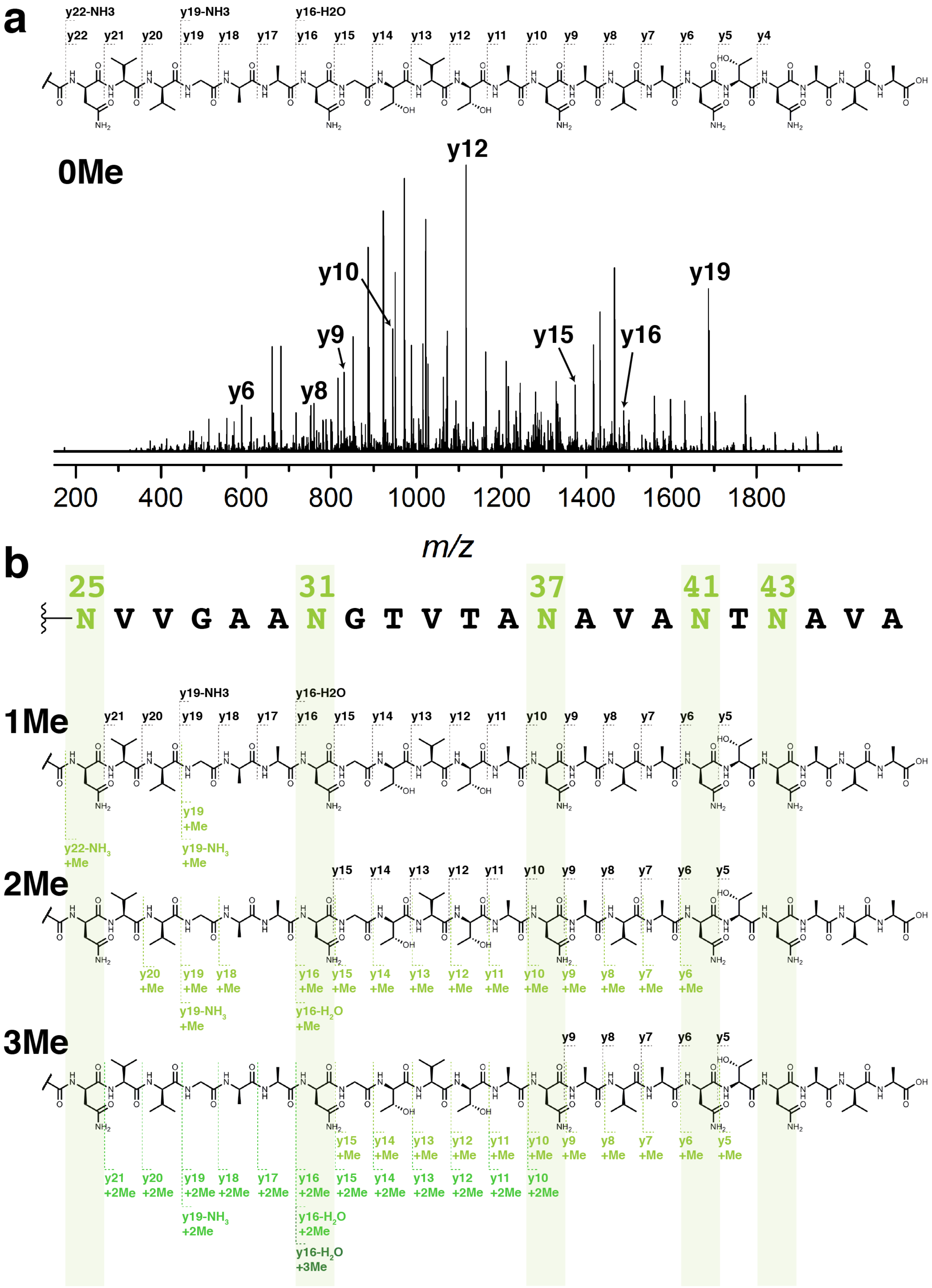

a
y22-NH3
y19-NH3
y16-H2O
y22 y21 y20 y19 y18 y17 y16 y15 y14 y13 y12 y11 y10 y9 y8 y7 y6 y5 y4
0Me
y12
y10
y9
y6
y8
y15
y16
y19
200
400
600
800
1000
1200
1400
1600
1800
m/z
b
25 31 37 41 43
N V V G A A N G T V T A N A V A N T N A V A
1Me
y19-NH3
y16-H2O
y21 y20 y19 y18 y17 y16 y15 y14 y13 y12 y11 y10 y9 y8 y7 y6 y5
y19 +Me
y22-NH$_3$ +Me
y19-NH$_3$ +Me
2Me
y15 y14 y13 y12 y11 y10 y9 y8 y7 y6 y5
y20 +Me
y19 +Me
y18 +Me
y16 +Me
y15 +Me
y14 +Me
y13 +Me
y12 +Me
y11 +Me
y10 +Me
y9 +Me
y8 +Me
y7 +Me
y6 +Me
y19-NH$_3$ +Me
y16-H$_2$O +Me
3Me
y9 y8 y7 y6 y5
y15 +Me
y14 +Me
y13 +Me
y12 +Me
y11 +Me
y10 +Me
y9 +Me
y8 +Me
y7 +Me
y6 +Me
y5 +Me
y21 +2Me
y20 +2Me
y19 +2Me
y18 +2Me
y17 +2Me
y16 +2Me
y15 +2Me
y14 +2Me
y13 +2Me
y12 +2Me
y11 +2Me
y10 +2Me
y19-NH$_3$ +2Me
y16-H$_2$O +2Me
y16-H$_2$O +3Me

Fig. 6.3

Fig. 6.3 (cont.): Pseudo-(N-to-C)-Directionality of AerE-catalyzed D-Asn *N_{γ}*-methylation analyzed by ESI-MS/MS. a, Fragmentation analysis of non-methylated AerA$_{DL}$ with observed y-ions mapped onto the structure and corresponding collision-induced dissociation (CID) mass spectrum below (representative set of high-abundance y-ions are labeled). **b,** Similar fragmentation analysis for the HPLC-purified mono-, di-, and tri-*N*-methylated AerA$_{DL}$ intermediates (1Me, 2Me, and 3Me). Observed y-ions with no methylation are demarcated above the peptide in black dotted lines, whereas methylated intermediates are below the peptide in green. A threshold of 5 ppm mass error was designated for peak assignment.

atoms including SAH). The peptide binds as an unstructured loop embedded into the methyl-acceptor binding site shared with other NMTs and akin to the interaction between the looped GGQ-motif of RF1 bound to PrmC.[219–221] Ten of the 34 residues in AerA$_{DL}$(34mer) are visible by electron density corresponding to Thr19-Gly28 of the AerA CP (Fig. 6.4c). D-Asn25 is unambiguously located closest to the site of SAM-mediated methylation, in agreement with the tandem MS data which suggest its role as an early methyl acceptor. Moreover, the ternary complex structure offers a potential explanation for the slow methylation of D-Asn43 by inspection of the neighboring residue environments. Specifically, survey of the -2 through +2 residues with respect to each D-Asn in the core peptide reveals a preference for Gly, D/L-Ala, D/L-Val, and D/L-Thr excepting D-Asn41 and D-Asn43 which have D-Asn in their +2 and -2 positions, respectively. In the crystal structure, the -2 residue (D-Ala23) of D-Asn25 is oriented with its side-chain pointing into a hydrophobic surface of AerE (towards Pro321 and Ile322), whereas the +2 residue (D-Val27) is more solvent exposed and directed away from AerE, perhaps explaining why a D-Asn is more tolerated at the +2 position.

In the active site, the γ-nitrogen of D-Asn25 is 3.7 Å away from the sulfur of SAH where it would be directly poised for nucleophilic attack of the SAM *S*-methyl group (Fig. 6.4e). However, given the low nucleophilicity of an sp^2-hybridized amide nitrogen owing to its participation in

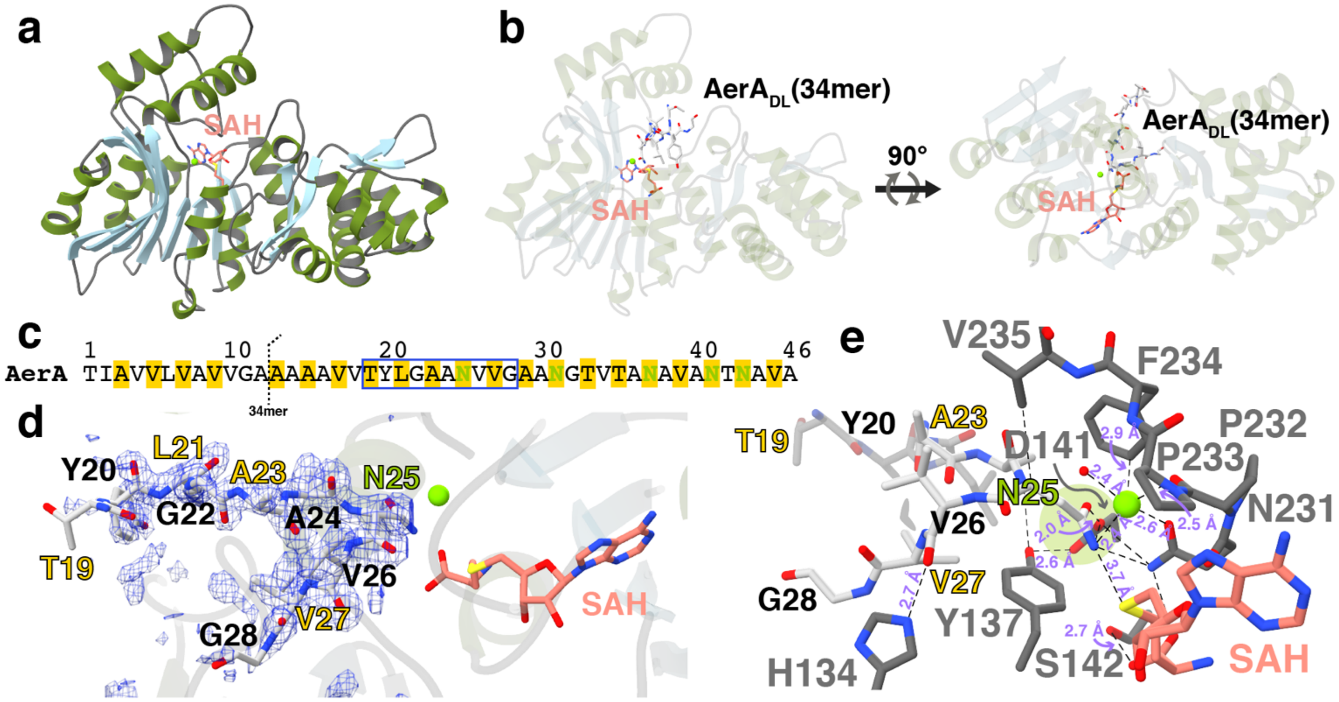


Fig. 6.4: Structural analysis of AerE bound to SAH and its methyl-acceptor substrate AerA$_{DL}$(34mer). a, Overall ribbon diagram of AerE bound to SAH (salmon). **b,** Two (90°-related) orientations of the AerE-SAH-AerA$_{DL}$(34mer) ternary complex. **c,** Sequence representation of AerA CP where amide bond cleavage between Ala12-Ala13 results in the C-terminal fragment AerA$_{DL}$(34mer) synthesized by SPPS with the epimerized residues boxed in yellow, *N*-methyl acceptor D-Asn residues in green, and the 10 residue region observed by electron density boxed in blue. **d,** Simulated annealing difference Fourier map (F_{obs}-F_{calc}) contoured at 2σ centered around the 10 blue-boxed residues in AerA$_{DL}$(34mer). D-configured residues are yellow (or green for D-Asn25) with a black outline. **e,** Close-up of the AerE active site where AerE residues are in grey, SAH in salmon, and AerA$_{DL}$(34mer) coloring-scheme consistent with panel d. An unidentified structural-ion is depicted as a green sphere. With the exception of D-Asn25 γ-nitrogen and SAH sulfur bond distance of 3.7 Å, only ≤ 3 Å bond distances are labeled for clarity.

resonance, we considered the possibility of a Lewis acid type activation where coordination of the amide oxygen and a heavier atom (e.g. Na^+, K^+, Mg^{2+}, or Ca^{2+}) might serve to lower the rotational barrier of the amide C-N bond; thus freeing up the lone-pair electrons for increased nucleophilicity. Indeed, the difference Fourier map (F_{obs} - F_{calc}) presents sufficient electron density adjacent to the D-Asn25 amide to model the above heavy atoms, however, the true atom identity remained ambiguous. We attempted to inhibit AerE by incubation with an excess of EDTA accompanied by overnight dialysis. While methylation activity was attenuated by comparison to the untreated enzyme at various time points, AerE remained fully capable of penta-*N*-methylating AerA$_{DL}$(34mer) following overnight incubation.

Interestingly, AerE is near fully inhibited by low stoichiometric amounts (4 eq.) of Zn^{2+} in reactions with SAM/AerA$_{DL}$(34mer), and the inhibition is not rescued by addition of excess Na^{+}, K^{+}, Mg^{2+}, Ca^{2+}, or Mn^{2+} (Figs. 6.5a and 6.8).

6.3.5 Site-directed mutagenesis and N-methylation activity of AerE

Considering sequence conservation and active-site proximity, we generated point mutations of AerE corresponding to H134A, Y137A, Y137F, D141A, S142A, and N231A AerE variants, and tested their *N*-methylation activities with the simple substrate AerA$_{DL}$(34mer) (Fig. 6.5b). After prolonged incubation of the wild-type and above AerE variants with SAM and AerA$_{DL}$(34mer), analysis of the MALDI-TOF MS data reveal Y137A, Y137F, D141A, S142A, and N231A variants are nearly devoid of all activity with trace amounts of mono- and di-methylation. The H134A variant is attenuated considerably, but still capable of achieving tri-methylation under these reaction conditions. Structurally, H134 appears to play a substrate binding role, as it makes contact (2.7 Å) with the methyl-acceptor +1 position by H-bonding with the backbone amide oxygen of Val26 in AerA$_{DL}$(34mer) (Fig. 6.4e). Given that the Y137A and Y137F variants gave comparable losses in methylation activity, we looked at the Y137 side chain hydroxyl interactions and observed H-bonding (2.6 Å) with the side chain carboxylate of D141, as well as H-bonding (3.2 Å) with the backbone amide nitrogen of D-Asn25. D141 further adheres the substrate-enzyme H-bond network by directly connecting to the substrate D-Asn25 side chain amide oxygen (2.0 Å) and the N231 side chain amide (3.1 Å); interactions which are collectively essential for activity. Moreover, the AerE-D141A/SAH co-crystal structure reveals a 1.3 Å motion of the Y137 side chain towards SAH that would be deleterious for the Y137/substrate-backbone interactions; a potential contributor to activity loss. The dysfunctional S142A variant may be a result of impaired SAM binding, as it forms a 2.7 Å H-bond with the carboxylate of SAH. The N231 side chain may also serve this role of stabilizing the bound SAM by making similar H-bond contacts (3.1 Å), however it likely exerts additional influence by structural-ion coordination (2.6 Å), aforementioned D141 interaction (3.1 Å), as well as side chain γ-nitrogen interaction with D-Asn25 of the substrate (3.3 Å). AerE-N231A was crystallized under similar conditions to the wild-type enzyme and appears to have lost structural-ion binding ability by comparison of the difference Fourier ($F_{obs} - F_{calc}$) maps; perhaps the most critical feature contributing its activity loss.

6.4 METHODS

6.4.1 Materials

Oligonucleotides were purchased from Integrated DNA Technologies, Inc. (IDT) and enzymes for PCR, plasmid digestion, and Gibson Assemblies were purchased from New England Biolabs (NEB). Buffers, media, and other reagents were purchased from Thermo Fisher Scientific, Sigma-Aldrich (MilliporeSigma), and Cayman Chemical.

6.4.2 General Methods

Protein affinity and size-exclusion chromatography were carried out using an ÄKTAprime plus purifier (GE Healthcare). High-performance liquid chromatography (HPLC) was performed with a Shimadzu Prominence Liquid Chromatography system. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) data were collected on UltrafleXtreme and Autoflex Speed MS instruments (Bruker Daltonics).

6.4.3 Cloning and heterologous expression of AerA, AerD, and AerE (wild-type and site-directed variants)

AerA and AerD: The *aerA* and *aerD* genes from *M. aerodenitrificans* DSM 15089 were cloned as previously described into expression vectors encoding for His_6-AerA in pACYCDuet-1 and tag-free AerD in a pCDFDuet-1 derivative plasmid (pBADcdf) containing the arabinose-inducible promoter (pBAD). Following co-transformation of *E. coli* BL21 (DE3) using the above plasmids, we grew transformants in 4 L of terrific broth (TB) medium until an optical density at 600 nm (OD_{600}) of 1.6-2.0 was achieved. Cells were cooled on ice for 30 min and His_6-AerA expression was induced by addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) with growth at 18 °C for 1 h followed by addition of 0.2% L-arabinose for induction of AerD and continued growth for 20 h at 18 °C.

AerE: All *aerE* expression plasmids were constructed using the pET hexahis (His_6) small ubiquitin-like modifier (SUMO) tobacco etch virus (TEV) protease ligation-independent cloning (LIC) vector (2S-T). This vector was a gift from Scott Gradia (Addgene plasmid # 29711). Primers designed to include the LICv1 tags were used for PCR amplification of *aerE* from *Microvirgula aerodenitrificans* DSM 15089 gDNA. Amplicons were used in a Gibson Assembly reaction with SspI-linearized 2S-T following the manufacturer (NEB) protocol and products were used to transform *E. coli* DH5a. Site-directed mutations were incorporated by three-component Gibson Assemblies containing two *aerE* gene-fragments containing the PCR-derived mutations. Transformant plasmids were isolated using a QIAprep Spin Miniprep Kit (QIAGEN) and sequenced by services from ACGT, Inc. (Wheeling, IL). Sequence-verified His_6-SUMO-AerE plasmids were used to transform *E. coli* Rosetta (DE3) competent cells for protein overexpression. Protein was overexpressed by growth at 37 °C in Luria-Broth (LB) to an OD_{600} of 0.4-0.6. Protein expression was then induced by addition of 0.5 mM IPTG and growth for 18-20 h at 18 °C. 2-4 L of LB media was used for the overexpression of wild-type and site-directed variants of His_6-SUMO-AerE.

6.4.4 Protein purification

6.4.4.1 AerE (tag-free)

Cells from a 4 L culture were harvested by centrifugation at 3,000 × g for 30 min at 4 °C and resuspended in buffer A (0.5 M NaCl, 20 mM tris-HCl pH 8.0, and 10% glycerol (*v/v*)) before lysing via sonication. The lysate was cleared by centrifugation at 15,400 × g for 1 h at 4 °C and the supernatant was injected onto a 5 mL HisTrap nickel-nitrilotriacetic acid (Ni-NTA) affinity column (GE Healthcare) equilibrated with buffer B (1 M NaCl, 30 mM imidazole, 20 mM tris-HCl pH 8.0). The column was washed with 70 mL of buffer B prior to eluting the bound protein with a linear gradient of increasing buffer C (1 M NaCl, 500 mM imidazole, 20 mM tris-HCl pH 8.0). Fractions containing His_6-SUMO-AerE were pooled and treated with 30 µg/mL of tobacco etch virus (TEV) protease while dialyzing against 0.2 M NaCl, 2.5 mM β-mercaptoethanol (BME), and 20 mM tris-HCl pH 8.1 for 12 h at 4 °C. Dialysis was continued for another 4 h against a similar buffer lacking BME before injecting the sample onto a 5 mL Ni-NTA affinity column (GE Healthcare) equilibrated with buffer A for subtractive Ni-affinity purification. The column was subsequently washed with buffer A (50 mL) and eluted using the following step-wise gradient of increasing buffer B: 10%, 20%, 30%, 40%, 50%, and 100% in 5-10 mL increments. Tag-free AerE primarily eluted at 30% B. Pooled AerE fractions were concentrated to ≤5 mL and purified by size-exclusion chromatography (SEC) by injecting the sample onto a 120 mL Superdex 200 10/300 GL column (GE Healthcare) with 300 mM KCl, 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid HEPES pH 7.5 used for equilibration and running buffer.

6.4.4.2 SUMO-AerE (wild-type and SDMs)

Cells from a 2 L culture were harvested, lysed, and centrifuged as described for tag-free AerE. The clarified lysate was added to 3 mL of Ni-NTA resin (Thermo Fisher Scientific) and allowed to bind by gravity elution. The loaded resin was washed with 70 mL of buffer B and eluted with 2 mL fractions of 10% C (2x), 50% C (2x), and 100% C (5x). Pooled fractions were concentrated and used directly in *N*-methylation assays (Fig. 6.5b).

6.4.5 In vitro N-methylation assays

Initial *N*-methylation assays with substrate $AerA_{DL}$ were carried out at 22 °C with 20 μL reaction volumes containing 25 μM AerE (tag-free; following SEC), 100 μM $AerA_{DL}$, and 1 mM SAM in reaction buffer 1 (50 mM KCl, 50 mM K_2HPO4 pH 8, 200 μM $MgCl_2$, 10% glycerol). Reactions were terminated by heat-precipitating the enzyme at 95 °C for 5 min. The precipitate was removed by centrifugation, and the peptide products in the supernatant were trypsinized (0.4 μg/μL) for 30 min at 22 °C before desalting with a C4 ZipTip (Millipore Sigma) following the manufacturer instructions. Full extent of methylation (+70 Da) was observed by MALDI-TOF MS after 16 h incubation (Fig. 6.2b).

6.4.6 Processivity analysis of AerE

Reactions were carried out and terminated as described (see *6.4.5 In vitro N-methylation assays*), except 20 μM enzyme was used and a larger volume of 300 μL. At each time point, 20 μL were extracted for heat-quenching, trypsinization, desalting, and MALDI-TOF MS analysis. The peak intensities were calculated using Bruker Analysis software and normalized before plotting in OriginPro 2016 (OriginLab) software.

6.4.7 Tandem MS analysis (ESI MS/MS) of methylated AerA$_{DL}$ intermediates

Four 100 µL reactions were setup as described (see *6.4.5 In vitro N-methylation assays*) and quenched at four different time points (i.e. 5, 30, 75, and 200 min) to isolate mono-, di-, tri-, and tetra-*N*-methylated AerA$_{DL}$ intermediates. After heat-denaturing and trypsinization, the reaction products were isolated by HPLC using a GRACE Denali C18 (120 A; 5 um; Part No. 238DE54; length 250 mm; ID 4.6 mm), flow-rate of 1 mL/min, and a detection wavelength of 280 nm. Elution consisted of a mobile phase of H_2O/ACN (A/B) + 0.1% formic acid and a linear gradient of 2-70% B over 35 min followed by a linear step to 90% B held for 10 min. Mono-, di-, and tri-methylated AerA$_{DL21}$ peptides eluted between 35-42 min (determined by MALDI-TOF MS). HPLC-isolated peptides were directly infused into a ThermoFisher Scientific Orbitrap Fusion electrospray-ionization mass-spectrometer (ESI-MS) using an Advion TriVersa Nanomate 100. The following parameters were used: resolution = 120,000; isolation width (MS/MS) = 2 *m/z*; normalized collision energy (MS/MS) = 35; activation q value (MS/MS) = 0.4; and activation time (MS/MS) = 30 ms. Collision-induced dissociation (CID) and high-energy collisional dissociation (HCD) were used to fragment the parent ion at 35%. Data analysis was performed in the Qualbrowser application of Xcalibur software (Thermo Fisher Scientific). Only peaks with less than 5 ppm error were tabulated and included in the analysis.

6.4.8 Solid-phase peptide synthesis (SPPS) of AerA$_{DL}$(34mer)

Automated SPPS was carried out using a Rainin PS3 synthesizer on a 0.1 mM scale with Fmoc-Ala Wang resin. Coupling reactions contained 4 eq of the appropriate Fmoc-amino acid and 4 eq of HCTU dissolved in DMF + 0.4 M *N*-methylmorpholine. Each coupling step was performed twice prior to deprotection with 20% (*v/v*) piperidine in DMF. Following all rounds of double-coupling/deprotection, the protected resin-bound product corresponded to: Fmoc-a-A-a-A-v-V-

t^{tBu}-Y^{tBu}-l-G-a-A-n^{Trt}-V-v-G-a-A-n^{Trt}-G-t^{tBu}-V-t^{tBu}-A-n^{Trt}-A-v-A-n-T^{tBu}-n^{Trt}-A-v-A -- Wang resin (D-configured amino acids in lowercase and side-chain protecting groups in superscript). The product was washed with DCM and dried with a N_2 stream. Resin-bound, crude peptide was globally deprotected by incubation with 4 mL of 95% trifluoroacetic acid (TFA), 2.5% triisopropylsilane (TIPS), and 2.5% H_2O for 1 h at 20 °C. The resin was removed by filtration through glass wool and the peptide was precipitated by collecting the filtrate in cold ether (-20 °C). The precipitate was spun down (3,000 × g for 10 min at 4 °C), supernatant discarded, and the pellet resuspended in ice-cold ether. This process was iterated two more times before the crude-peptide pellet was dried under a stream of N_2.

6.4.9 HPLC purification of AerA$_{DL}$(34mer) from SPPS, AerA$_{DL21}$ (trypsin fragment), and AerA$_{DL}$ (proteinase K fragments; 31-35mer)

6.4.9.1 AerA$_{DL}$(34mer) from SPPS

The deprotected, crude peptide was purified by reversed phase HPLC using a Macherey Nagel VP 250/10 NUCLEODUR C8 Gravity (5 um) column with a mobile phase consisting of H2O/ACN (A/B) + 0.1% FA, flow-rate of 3 mL/min, and a detection wavelength of 280 nm. A linear gradient of 2-50% B over 35 min eluted the peptide at 26 min as detected by MALDI-TOF MS.

6.4.9.2 AerA$_{DL21}$ (trypsin fragment)

Following trypsinization of AerA$_{DL}$ to afford AerA$_{DL21}$, (reaction quenching protocol) the peptide products were purified by HPLC using a GRACE Denali C18 (120 A; 5 um; Part No. 238DE54; length 250 mm; ID 4.6 mm), flow-rate of 1 mL/min, and a detection wavelength of 220 nm. Elution consisted of a mobile phase of H2O/ACN (A/B) + 0.1% FA with the following linear gradient: 2-70% B over 40 min. AerA$_{DL21}$ eluted at 22.5 min as detected by MALDI-MS TOF.

6.4.9.3 AerA$_{DL}$ (proteinase K fragments; 31-35mer)

Following overnight proteinase K treatment, (reaction quenching protocol) a mixture of proteinase K fragments (Fig. 6.2) were purified by HPLC as described for AerA$_{DL21}$ and co-eluted at 30 min as detected by MALDI-TOF MS. The mixture was used directly in in vitro *N*-methylation assays.

6.5 SUPPORTING FIGURES

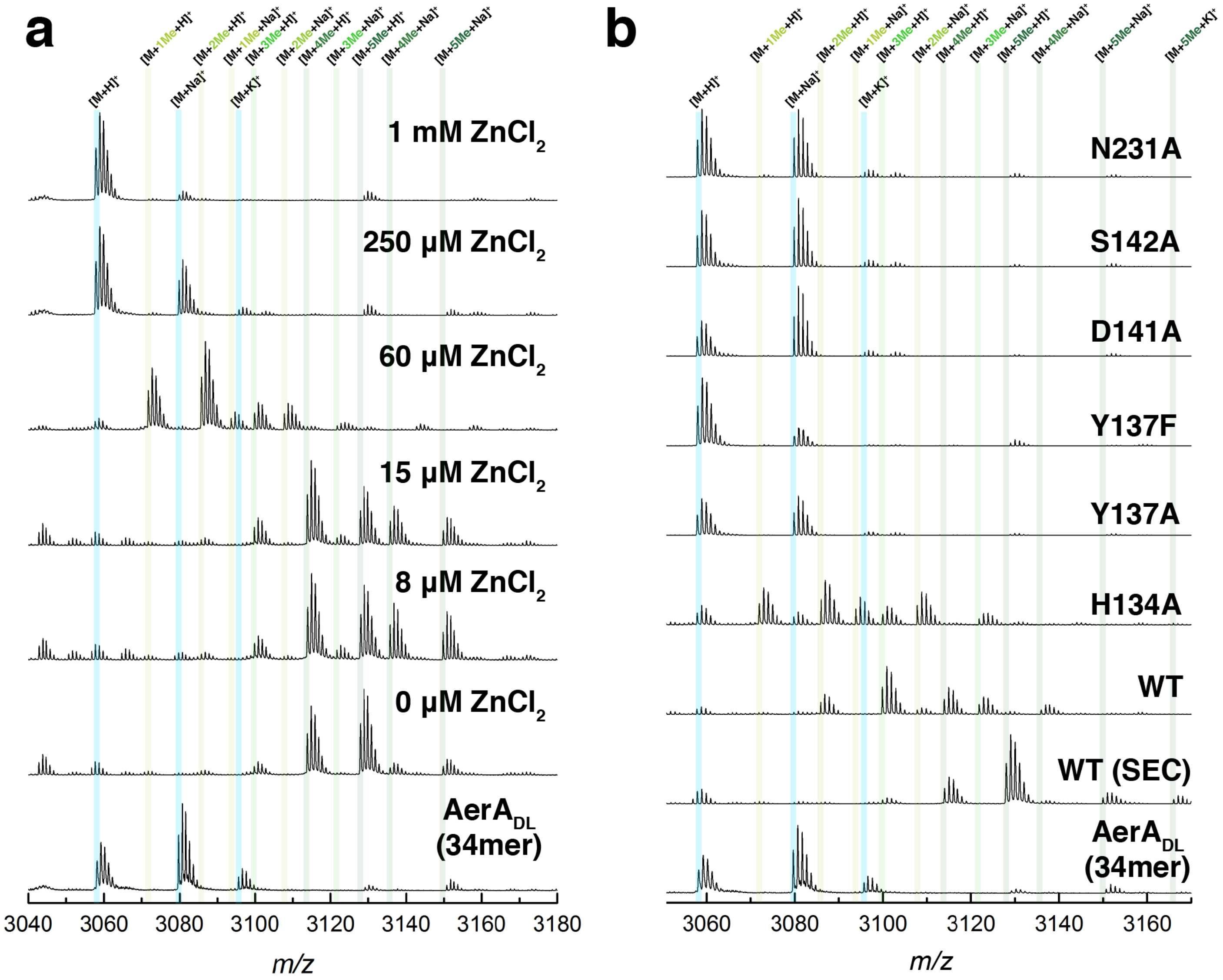


Fig. 6.5: Zinc inhibition and site-directed mutagenesis of AerE. a, Dose-dependent inhibition of AerE-catalyzed AerA$_{DL}$(34mer) *N*-methylation by $ZnCl_2$. Full inhibition is observed between 60-250 µM, as observed by MALDI-TOF MS. **b,** Methylation activity of site-directed AerE variants towards substrate AerA$_{DL}$(34mer). Enzymes were assayed directly following Ni-NTA purification, except for WT (SEC) which denotes the wild-type AerE enzyme

Fig 6.5 (cont.) which underwent multiple purification steps terminating in SEC (see Methods). The substrate, AerADL(34mer), was analyzed alone for comparison in the bottom spectrum of panels a-b.

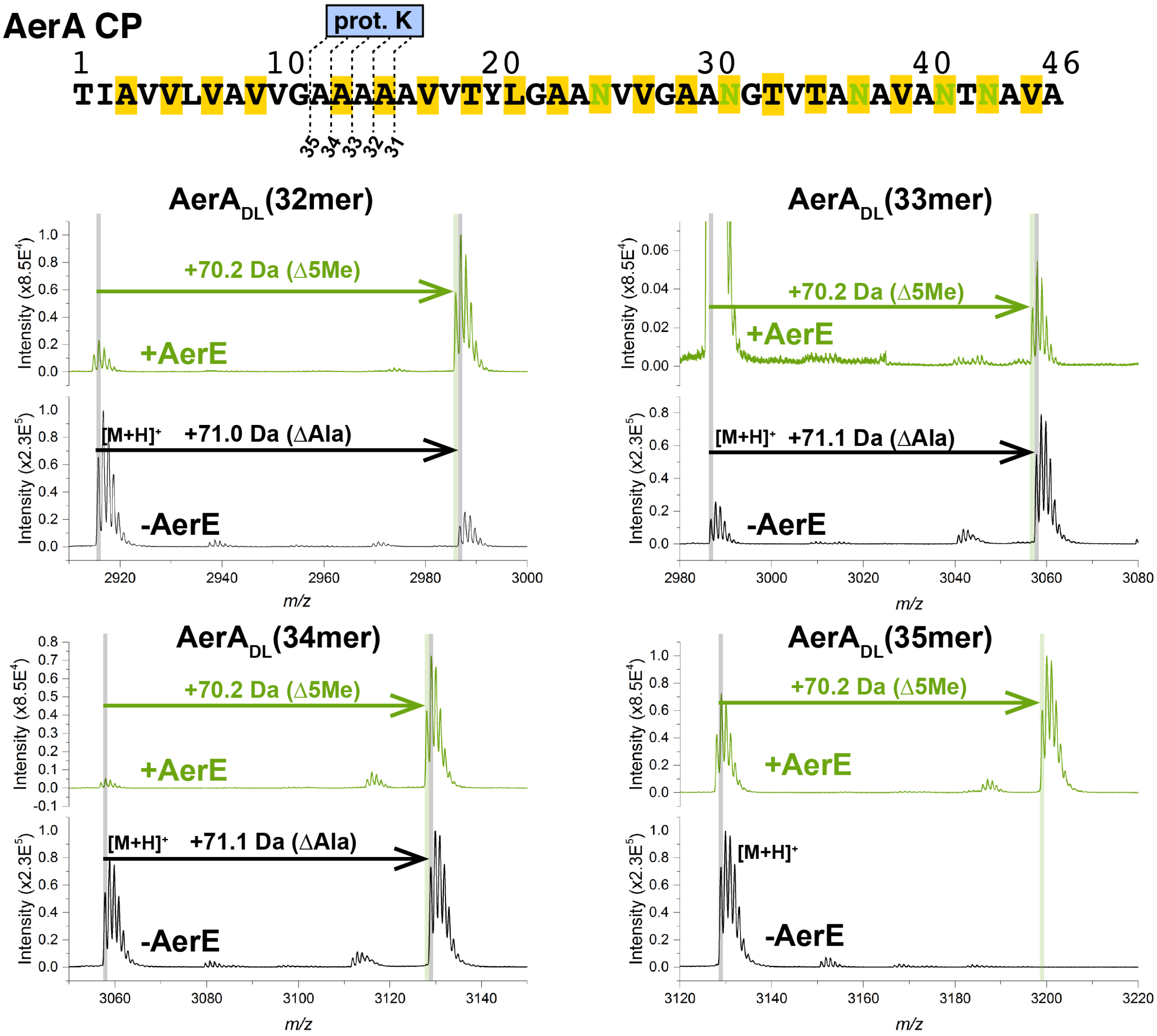


Fig. 6.6: MALDI-MS analysis of the AerE-catalyzed penta-*N*-methylation of proteinase K-derived fragments (32-35mer) of AerADL. The 31-35mer fragments co-eluted by HPLC and were incubated as a mixture with AerE/SAM. All species gave rise to a 70.2 Da mass shift, indicative of penta-*N*-methylation. Note: each species varies by an Ala, which is very close in mass to (71.1 Da) the calculated mass of 5 methylations (70.2 Da). Despite this, the monoisotopic resolution enabled their distinguishing.

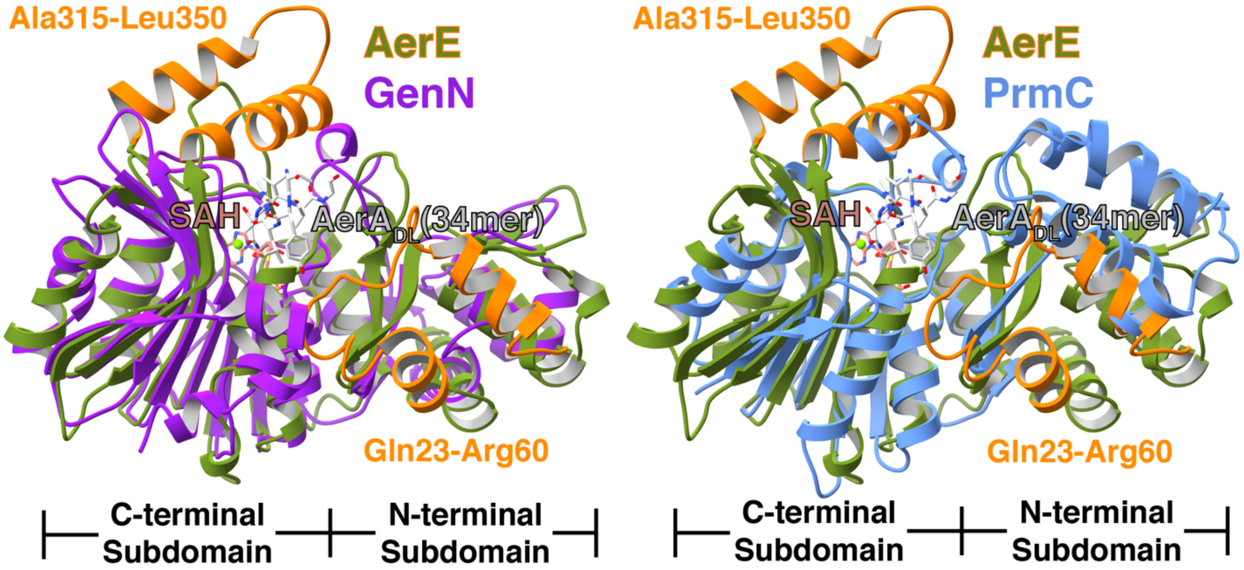


Fig. 6.7: Structural alignment of AerE (green and orange) with *N*-methyltransferases GenN (purple; left) and PrmC (blue; right). The orange regions (i.e. Gln23-Arg60; Ala315-Leu350) represent the structural insertions relative to GenN/PrmC.

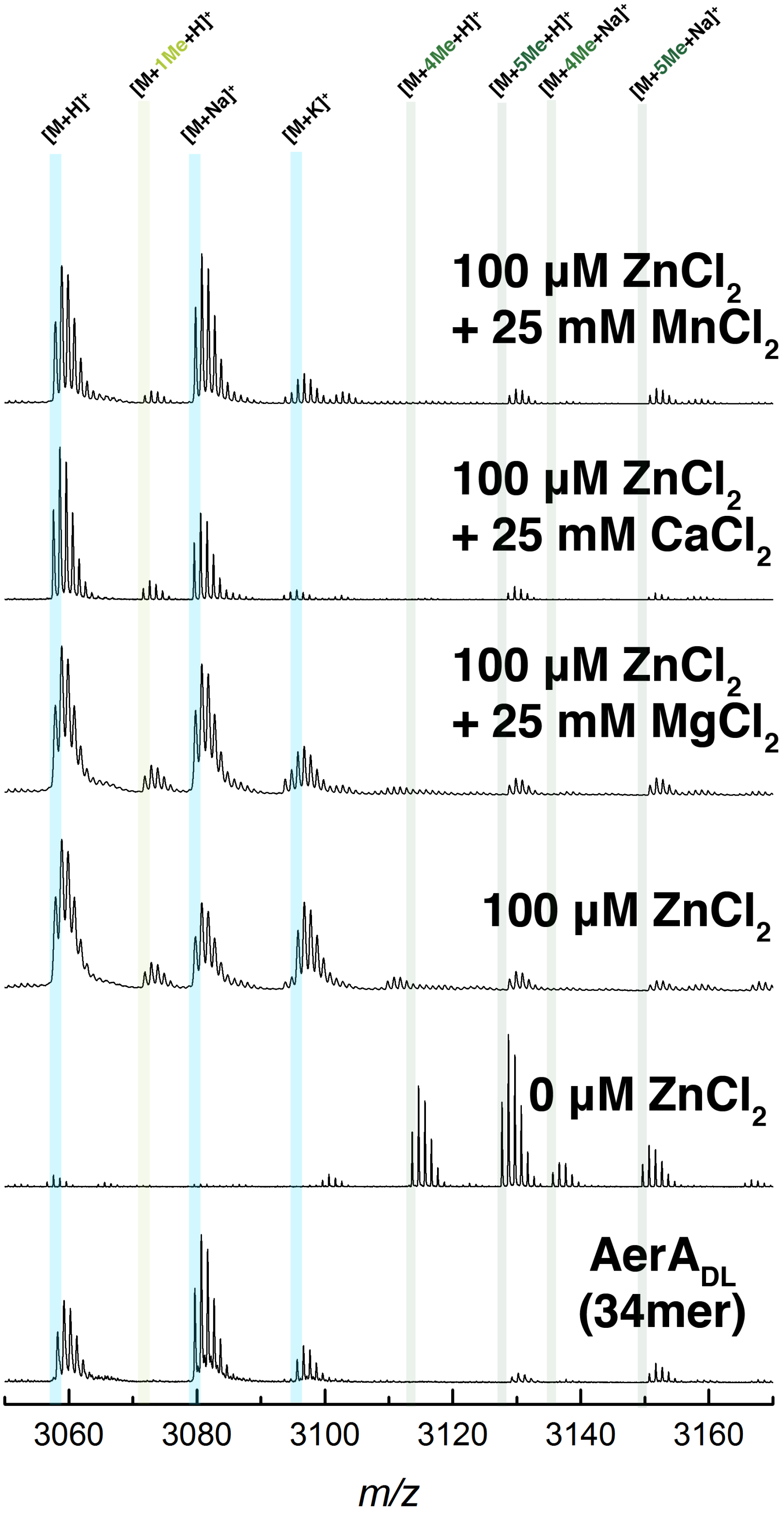


Fig. 6.8: $MgCl_2$, $CaCl_2$, and $MnCl_2$ do not rescue Zn-inhibited AerE *N*-methylation. MALDI-TOF MS analysis of reactions containing AerE, $AerA_{DL}$(34mer), $ZnCl_2$, and SAM supplemented with either 25 mM $MgCl_2$, $CaCl_2$, or $MnCl_2$. The substrate, $AerA_{DL}$(34mer), was analyzed alone for comparison in the bottom spectrum.

CHAPTER 7: MECHANSTIC INSIGHT INTO TRNA-DEPENDENT LANTHIPEPTIDE AND THIOPEPTIDE SIDE-CHAIN DEHYDRATION

7.1 ABSTRACT

The class I lantibiotic nisin, from *Lactococcus lactis*, has been employed extensively for agricultural and food preservative applications since the early/mid 20th century with minimal development of resistance. Despite decades of study, however, the first insights into the biosynthetic mechanisms of class I lantibiotics had remained elusive until recently. Herein, we describe our efforts to further examine the structure and mechanism of class I lantibiotic and thiopeptide dehydratases through the use of non-reactive substrate mimics. In particular, we report the development of a glutamylated-NisA analogue as a structural probe for the elimination domain of NisB, as well as a puromycin-like analogue, 5′-phosphoryl-*N6*-desmethylglutamycin, to mimic glutamyl-tRNA^{Glu} and probe the structure of the glutamylation domain, TbtB, involved in thiomuracin biosynthesis (from *Thermobispora bispora*). Our results support current hypotheses regarding the mode of substrate recognition and provide insight into the mechanism of this enzymatic transformation.

7.2 INTRODUCTION

Ribosomally synthesized and posttranslationally modified peptides (RiPPs) are an important class of genetically encoded peptide natural products with numerous biological functions, most notable are their antimicrobial properties. Broadly speaking, RiPP biosynthesis involves the enzymatic tailoring of a genetically encoded precursor peptide, or LanA. This peptide is composed of a C-terminal core peptide (CP) region, where posttranslational modifications (PTMs) occur, and an N-terminal leader peptide (LP) region, which is often essential for recognition by the modification machinery and ultimately removed in the mature RiPP. Chemical modification of the core region usually involves PTM by one or more tailoring enzyme included in the gene cluster in an iterative or sequential fashion. Of the many tailoring enzymes known, major interest lies with the class I lantibiotic dehydratases, which are involved in the biosynthesis of several classes of RiPPs, including lanthipeptides and thiopeptides.

Class I lantibiotic dehydratases (generally referred to as LanBs) are responsible for converting target serine and threonine residues in the core region of the precursor peptide to dehydroalanine (Dha) and dehydrobutyrine (Dhb), respectively. This class of enzyme can be further categorized based on homology and domain architecture into subclasses referred to as: full-length LanBs, split-LanBs, and small-LanBs (Figure 1a). Although the general function of small-LanBs is still under investigation, it has been well established that full-length and split-LanBs utilize aminoacyl-tRNA to install Dha and Dhb into the maturing LanA. Specifically, these enzymes transfer glutamate from glutamyl-tRNA^{Glu} to the side-chain hydroxyl-groups of target serines and threonines in the CP. These glutamylated intermediates are subsequently eliminated by specialized domains or stand-alone enzymes, yielding Dha and Dhb products. This reliance on aminoacyl-tRNA stands in stark contrast to class II, III, and IV lantibiotic synthetases, which utilize NTPs to phosphorylate Ser/Thr side-chains for activation and subsequent elimination to afford the dehydrated residues.[39]

Structural insight into how these transformations are achieved was originally derived from crystallographic studies of NisB and MibB, which are full-length LanBs involved in the biosynthesis of the nisin and microbisporicin (NAI-107) lanthipeptides, respectively.[53,101] Despite relatively low amino acid sequence identity (~20%), these proteins are quite similar in terms of their overall domain structure, with a smaller (~30 kDa) C-terminal domain responsible for elimination of the glutamyl-LanA intermediates and a larger (~90 kDa) N-terminal glutamylation domain that installs them (Fig. 7.1). Furthermore, the N-terminal region contains an embedded subdomain, the RiPP recognition element (RRE), which is often essential for binding to the leader region of the LanA and promoting modification of the CP.[83] While the structure for NisB was originally obtained with sufficient electron density to model a portion of the NisA LP bound, the MibB structure was determined in its apo-form. In both cases, neither tRNA^{Glu} nor the CP regions of their respective LanAs were bound to the enzyme at either the glutamylation or elimination active sites. Thus, our current understanding of active site substrate recognition and catalysis is restricted to inference based on mutational analysis of conserved amino acids of the enzyme and nucleotides of the tRNA^{Glu} cosubstrate [101,222]

In addition to addressing the gaps in knowledge outlined above, we also set out to expand our structural knowledge through the examination of the related split-LanB system from *Thermobispora bispora* involved in thiomuracin biosynthesis (TbtBC). The dehydratases from the split-LanB subclass are often found in thiopeptide biosynthetic gene clusters and their activity is a prerequisite for the ultimate formation of the six-membered nitrogenous heterocycles that are a structural hallmark of this class of RiPPs.[39] As the name suggests, split-LanBs are composed of two separate proteins: the C-terminal elimination domain (TbtC) and N-terminal RRE/glutamylation domain (TbtB) that are thought to act in complex along with additional biosynthetic machinery. In the case of thiomuracin biosynthesis, it was previously demonstrated that the installation of thiazole/oxazole modifications in the maturing CP precedes dehydration by TbtB and TbtC.[89] This implies that TbtB is capable of recognizing structural features of the partially modified CP that are not found in lanthipeptides modified by full-length LanBs. Furthermore, given that TbtB exhibits glutamyl-transferase activity *in vitro* independent of both the LP and TbtC, this enzyme provides a valuable and simplified platform for examining how the glutamylation domain engages its tRNA and peptidic substrates.[86]

The interactions between tRNA, CP, and LanBs are thought to be very transient and facilitated in large part through binding of the LP with the RRE.[83,223] This is likely a mechanistic imperative for LanBs given both their iterative nature and diverse product structures. However, this also presents a challenge for the development of structural probes with sufficient affinity to target the active site of interest. Central to our goal of better defining the active sites of both the elimination and glutamylation domains was the synthesis of non-reactive substrate mimics for these enzymes. These molecules were designed to be chemically inert but near-isosteric with either the glutamyl-tRNA^{Glu} cosubstrate that binds to the glutamylation domain, or the glutamylated-LanA intermediate that subsequently interacts with the elimination domain. The general structural concept for the design of these mimics is shown in Figure 7.1. Substitution of an amide linkage in place of an ester for both molecules was expected to prevent transesterification by the glutamylation domain and β-elimination by the elimination domain while keeping the steric features nearly identical to the natural substrates.

Herein, we report the structure of the glutamylation domain of TbtB bound to a small-molecule aminoacyl-tRNA mimic inspired by puromycin, which we refer to as PDG (5′-phosphoryl-*N*6-desmethylglutamycin; Fig. 7.1B). Moreover, we also report the structure of NisB bound to a non-reactive glutamyl-NisA mimic (NisA-Dap$_3^{Glu}$; Fig. 7.1C) in the elimination domain. Together, these data reveal key molecular features involved in LanB substrate recognition and active site structure. Accordingly, we carried out structure-guided mutagenesis and activity experiments to understand the relevance of individual residues to binding and/or catalysis. These findings expand upon our prior understanding of enzyme-catalyzed Dha/Dhb formation by enabling a more detailed mechanistic understanding and may serve to inform future works aimed at heterologous expression or bioengineering of RiPP systems.

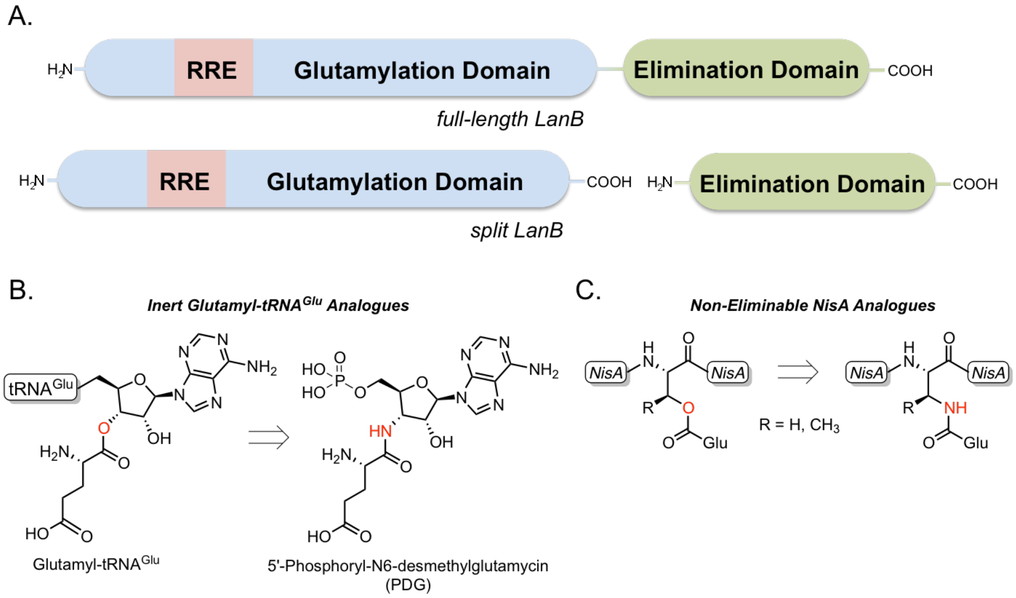


Fig. 7.1: LanB domain architecture and substrate mimic design philosophy. (A) General domain layout for full-length and split-LanB enzymes. (B) glutamyl-tRNA^{Glu} served as inspiration for the design of the glutamylation domain substrate mimic, PDG. (C) glutamylated-NisA provided inspiration for the design of a non-eliminable substrate mimic to probe the LanB elimination domain (Figure credit: Ian Bothwell).

7.3 RESULTS

7.3.1 Overall structure of TbtB

Structural investigations were initiated on the glutamylation domain, TbtB, from the split dehydratase involved in thiomuracin biosynthesis. Apo TbtB was crystallized and its structure solved to 2.1 Å resolution using diffraction data from crystals soaked in thiomersal. The overall structure of TbtB resembles the NisB and MibB glutamylation domains ($NisB_{5-729}$, RMSD = 5.0 Å, over 2889 atoms; $MibB_{4-809}$, RMSD = 3.5 Å, over 2586 atoms) with a distinguishing coiled-coil subdomain at the N-terminus (residues 36-153; helices $\alpha 2-\alpha 5$) which is absent in the type I lanthipeptide dehydratases NisB and MibB (Fig. 7.2; for structure-based multisequence alignment see Fig. 7.3).[53,101] Genome neighborhood analysis using RODEO suggests this extended N-terminal subdomain is a feature common to thiopeptide and other split LanB containing biosynthetic clusters (Fig. 7.4).[202] The PqqD-like RRE domain in TbtB, implicated in LP recognition,[83] recapitulates the structural organization observed in NisB and MibB (TbtB strands β6-β7-β8, helices α10-α11, and strand β9; Fig. 7.3), however, previous bioinformatic and binding experiments suggest a LP independent substrate recognition instead driven by features of the azole-containing core region of TbtA.[86]

7.3.2 Co-crystal structure of TbtB and PDG

Upon soaking apo TbtB crystals with PDG and obtaining diffraction data, the corresponding electron density maps revealed occupancy of PDG in the same AMP binding site, providing the first glimpse of a RiPP dehydratase bound to a glutamyl donor substrate-mimic. In the cocrystal structure, the adenine ring of PDG wedges between the phenyl side-chain of Phe783 and the guanidinium of Arg743 forming a π-π-cation interaction (Fig. 7.5). The exocyclic amine of the adenine ring also H-bonds with the Glu829 carboxylate and backbone amide oxygen of Glu830. The 2′-hydroxyl and Thr202 side-chain engage by H-bonding, and an electrostatic interaction is observed between the α-ammonium of PDG and the carboxylate of Glu851 in TbtB. Additional electrostatic interactions are present between the 5′-phosphate of ligand and side-chains of Arg197 and Lys201. Somewhat unexpectedly, the γ-carboxylate of the 3′-Glu of PDG is disordered in the crystal structure, despite being situated near to Arg22.

7.3.3 Structural analysis of NisB and non-eliminable glutamyl-NisA mimic

Previous studies identified several residues important for glutamate elimination in NisB.[101,222] To better understand their context in the mechanism of elimination, we determined the structure of the covalent complex between NisB variant V169C bound to NisA-$\mathrm{Dap}_3^{\mathrm{Glu}}$. The overall structure of NisB is highly similar to that of the previously elucidated structure with an

RMSD of 0.4 Å over 9196 atoms. The same region of the LP (i.e. residues -21 to -10) is visible and interfaces as an antiparallel β-sheet with β7 of the NisB RRE (Fig. 7.2). Electron density supports the existence of the engineered disulfide linkage at residues Cys169 and Cys(-12) of the protein and substrate mimic, respectively. In addition, residues 2-4 of the CP are visible in the elimination domain active site with the side-chain of the glutamate adduct inserted into a cavity containing residues previously shown to be important for catalysis (Fig. 7.6). Notably, the imidazole τ-nitrogen of His961 is situated 3.0 Å away from the α-carbon of the $\mathrm{Dap}_3^{\mathrm{Glu}}$ (D3G) residue where it is directly poised for proton abstraction, suggesting it plays the role of catalytic base in the glutamate elimination reaction. Moreover, Arg786 forms an electrostatic interaction with the carboxylate of the D3G side-chain, reaffirming its importance for binding and providing a structural basis for glutamate selectivity.

7.4 METHODS

7.4.1 Cloning and heterologous TbtB protein expression

Plasmids containing *tbtB* WT or site-directed mutations for heterologous protein expression in *E. coli* were constructed using the pET hexahis (His_6) small ubiquitin-like modifier (SUMO) tobacco etch virus (TEV) protease ligation-independent cloning (LIC) vector (2S-T). This vector was a gift from Scott Gradia (Addgene plasmid #29711). Mutations were generated via PCR using primers containing the LICv1 cloning tags and ligated into SspI-treated 2S-T via the Gibson Assembly.[197] The resulting His_6-SUMO-TbtB encoding plasmids were propagated in *E. coli* DH5α for plasmid isolation, and sequence-verified plasmids were used to transform *E. coli* Rosetta (DE3) competent cells for protein expression. Transformants were used to inoculate 50 mL or 4 L (for downstream activity or structural applications, respectively) of Luria-Broth (LB) medium and cells were grown at 37 °C with shaking (220 rpm) until an optical density at 600

(OD_{600}) of 0.4-0.6 was achieved. The cells were then cooled in an ice bath for 15 min and protein expression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and further incubated at 18 °C with shaking (220 rpm) for 20 h.

7.4.2 Preparation of SeMet-NisB-V169C

SeMet-derivatized NisB-V169C was generated using similar growth conditions to TbtB variants with substitution of LB medium for M9 minimal medium containing (0.4% (*w/v*) glucose, 11.2 g/L $Na_2HPO_4 \cdot 7H_2O$, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.12 g/L $MgSO_4$, and 0.5 g/L NaCl) supplemented with 4.2 mg/L $FeSO_4$, 0.5 mg/L thiamine hydrochloride, 0.1 g/L L-Lys, 0.1 g/L L-Phe, 0.1 g/L-Thr, 50 mg/L L-Leu, 50 mg/L L-Ile, 50 mg/L L-Val, and 50 mg/L L-SeMet.

7.4.3 Protein purification

7.4.3.1 SUMO-TbtB wild type and site-directed variants: Cells were harvested by centrifugation at 3,000 × g and cell pellets were suspended in 25 mL of buffer A (0.5 M NaCl, 10% glycerol (*v/v*), and 20 mM Tris-HCl pH 8.0). Suspended cells were lysed via sonication and the lysate was cleared by centrifugation at 15,400 × g for 1 h at 4 °C. Supernatant was applied to 2 mL of HisPur nickel-nitrilotriacetic (Ni-NTA) resin (Thermo Scientific) by gravity elution. The resin was washed with 40 mL of buffer B (1 M NaCl, 30 mM imidazole, and 20 mM Tris-HCl pH 8.0) prior to eluting the protein with buffer C (1 M NaCl, 0.1 M imidazole, and 20 mM Tris-HCl pH 8.0) in 1 mL increments for 2 total fractions followed by buffer D (1 M NaCl, 0.5 M imidazole, and 20 mM Tris-HCl pH 8.0) in 1 mL increments for 8 total fractions. Proteins at this point appeared ≥95% pure by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 7.7).

7.4.3.2 His$_6$-TbtB: Cells were harvested and lysed as described for the His$_6$-SUMO-TbtB variants. Supernatant was injected onto a 5 mL Ni-NTA HisTrap HP column (GE Healthcare) equilibrated with buffer A using an ÄKTApurifier (GE Healthcare). The column was washed with 50 mL of buffer B followed by a linear gradient elution of 0-100% buffer D over 40 mL (2 mL/min). Thrombin protease was added (1 U/mg substrate; Millipore Sigma) to elution fractions to remove the His$_6$-tag while dialyzing into 0.3 M NaCl, 3 mM β-mercaptoethanol (BME), 20 mM Tris-HCl pH 8.0 for 12 h. Dialysis was repeated in a similar buffer lacking BME for another 4 h before injecting protein onto a Ni-NTA HisTrap HP column (GE Healthcare) equilibrated with buffer A for subtractive Ni-affinity purification. The column was washed with 40 mL of buffer A before eluting with a stepwise gradient of increasing buffer B (i.e. 5, 10, 15, 20, 25, 30, 50, and 100 %B). Increments of 5 mL were used between each step unless a peak was observed by absorbance at 280 nm, in which case, the %B composition was maintained for the entire peak elution. The majority of tag-free TbtB protein eluted between 15-25 %B as monitored by SDS-PAGE. Pooled fractions of tag-free TbtB were concentrated to ≤5 mL and injected onto a 120 mL Superdex 200 10/300 GL column (GE Healthcare) to purify by size-exclusion chromatography (SEC) using 300 mM KCl, 20 mM 4-(2-hydroxylethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5 as running buffer. TbtB eluted as a monomer and appeared ≥90% pure by SDS-PAGE (Fig. 7.7).

7.4.3.3 His$_6$-SeMet-NisB-V169C (SM-NisB-V169C): Cells were harvested, lysed, and protein from the supernatant was purified by Ni-affinity chromatography as described for His$_6$-TbtB. Pooled fractions of SM-NisB-V169C were used to methylate surface exposed Lys residues following reported procedures.[224] The reductive amination reaction was quenched by addition of 0.1 M tris pH 7.5, and the protein was further purified by SEC as described for TbtB. Methylated SM-NisB-V169C eluted as a dimer and appeared ≥90% pure by SDS-PAGE (Fig. 7.7).

7.4.4 Protein crystallization and substrate analogue soaking

7.4.4.1 TbtB: TbtB purified by SEC was concentrated to 2 mg/mL prior to crystallizing in 2 µL hanging drops by mixing (1:1; *v/v*) protein and reservoir solution (0.2 M ammonium citrate pH 7.0 and 15.0% polyethylene glycol (PEG) 3350) and incubating at 20 °C. Ellipsoid-shaped crystals grew to maximum size within a week.

7.4.4.2 TbtB+PDG (soak): Drops containing apo TbtB crystals were supplemented with 10 mM PDG (dissolved in H_2O) and incubated for 8 h at 20 °C.

7.4.4.3 SM-NisB-V169C+NisA-Dap_3^{Glu}*:* Methylated SM-NisB-V169C purified by SEC was concentrated to 9 mg/mL and incubated with 2x NisA-Dap_3^{Glu} (150 µM) for 30 min on ice. The covalent complex was crystallized in 2 µL hanging drops by mixing (1:1; *v/v*) protein and reservoir solution (18.3% PEG 6000, 0.1 M bicine pH 8.0, and 1 mM dithiothreitol (DTT)) and incubating at 9 °C. Hexagonal-shaped crystals grew to maximum size with 72 h. Under these conditions the RRE-peptide disulfide linkage does not appear to be reduced in the crystals, on the basis of electron density.

7.4.5 X-ray diffraction data collection and structure solution

Crystals were cryoprotected prior to data collection by immersing them briefly in similar reservoir solutions supplemented with 20% ethylene glycol and plunging into liquid nitrogen. Data were collected at sector 21 of the Advanced Photon Source (Argonne National Laboratory) using the Life Science Collaborative Access Team (LS-CAT) 21-ID-D, 21-ID-F, and 21-ID-G beamlines. An MD2 diffractometer (Arinax Scientific Instrumentation) was used for crystal mounting and 100 K temperature maintenance under an aerosolized nitrogen stream. TbtB crystals were soaked with 10 mM thiomersal for 8 h at room temperature prior to cryoprotection and data collection at 12.35 keV to obtain anomalous signal for phasing. Data were processed using XDS and autoPROC to obtain structure factors, and phases were determined by the single-wavelength anomalous diffraction technique implemented in Phenix AutoSol.[33,71,199] The initial TbtB model was rebuilt in Phenix AutoBuild and refined using Phenix Refine and REFMAC5 with additional manual refining in Coot.[34–36,38] Diffraction data for SM-NisB-V169C crystals were collected at 12.67 keV and structure factors were obtained similarly to TbtB. Phases were determined by molecular replacement using the NisB coordinates previously reported (PDB 4WD9) in Phaser MR.[72] The protein coordinates were refined in a similar fashion to TbtB. Ligand parameters for PDG and residue NisA-$\text{Dap}_3^{\text{Glu}}$ were generated by creating a structure-data file (sdf) in ChemDraw Professional 17.1. The sdf file was input into Phenix eLBOW, and the output cif and pdb files were used in refinement.[37]

7.5 SUPPORTING FIGURES

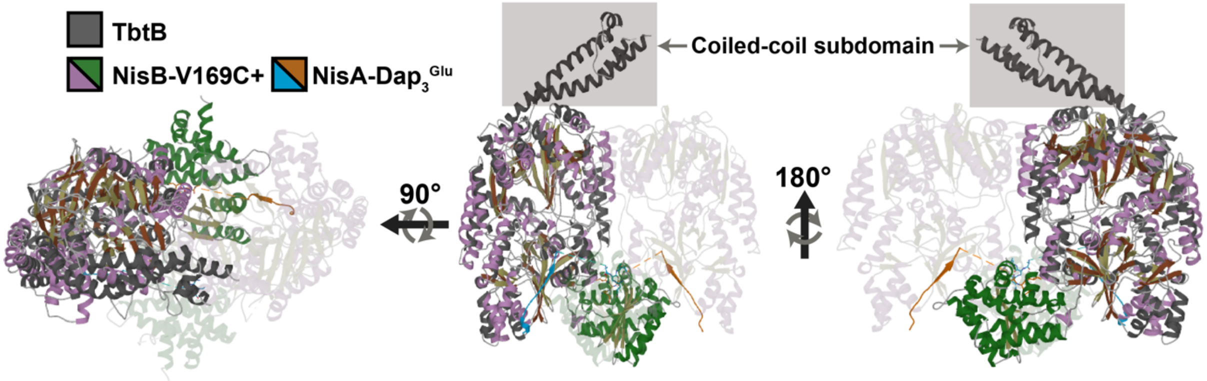


Fig. 7.2: Apo TbtB superimposed onto NisB-V169C bound to NisA-Dap$_3^{Glu}$ (β–sheets colored brown for TbtB and khaki for NisB-V169C). One NisB-V169C monomer is made 80% transparent for clarity. Helices of the glutamylation and elimination domains of NisB-V169C are colored purple and green, respectively. Two molecules of NisA-Dap$_3^{Glu}$ are shown in blue and orange where electron density is supported (i.e. -21 through -10 of the LP and 2-4 of the CP). Note that each molecule is modeled as a bidentate interaction across NisB monomers (i.e. LP bound at the NisB-RRE of one monomer and the core region bound at the elimination domain of the other monomer). Distance and entropic considerations favor this intermonomer interaction, where the corresponding intramonomer interaction would separate the Lys(-10) C-terminus an additional 18 Å from the N-terminus of Thr2 in the core. Moreover, the intramonomer interaction would require a back-and-forth threading of the maturing NisA peptide through a narrow passage (~20 Å) at the homodimeric interface.

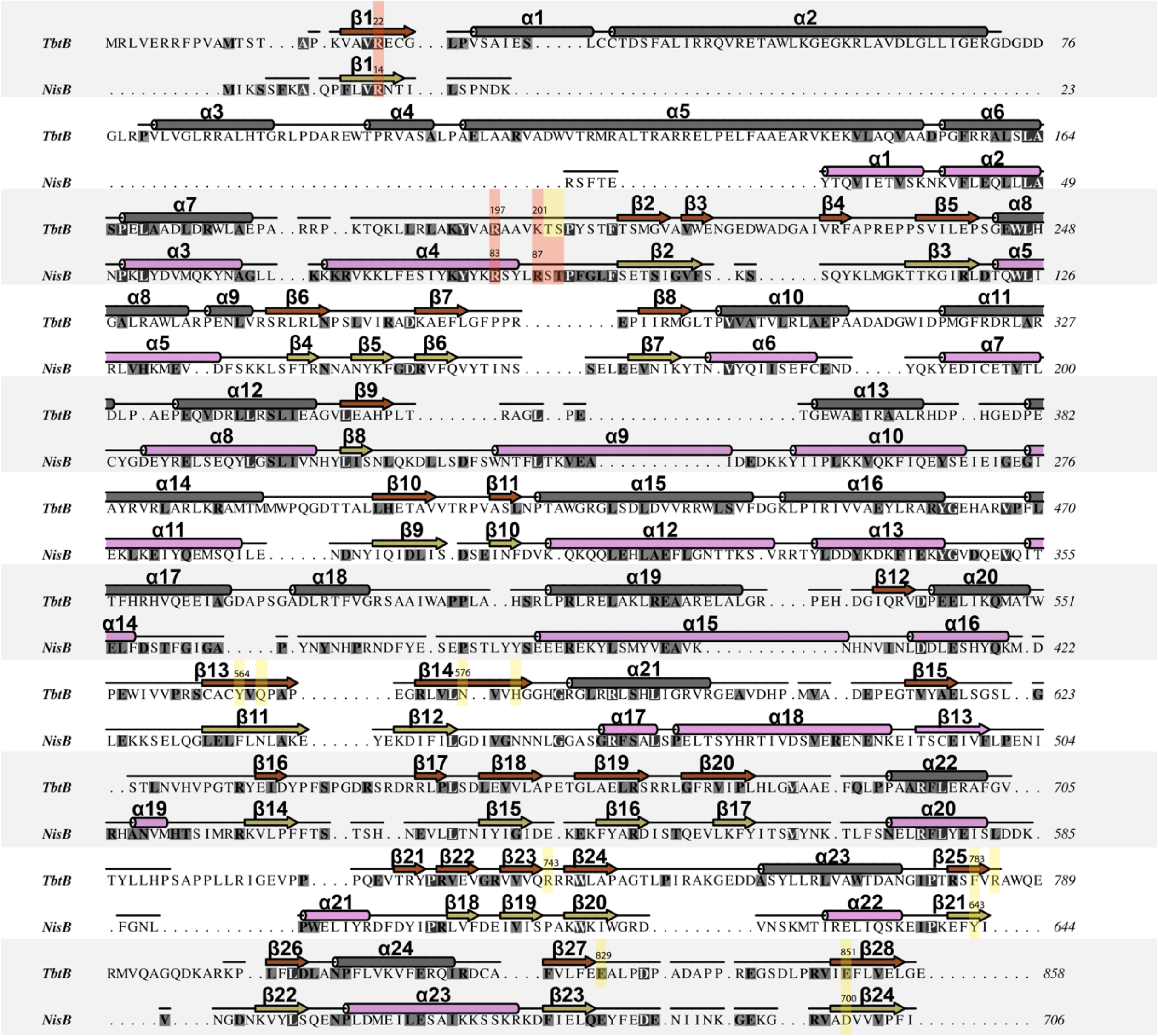


Fig. 7.3: A structural-based multiple sequence alignment between NisB$_{1\text{-}706}$ and TbtB generated by PROMALS3D.[225] Residues targeted for mutagenesis which gave partial or no impairment to activity are highlighted in yellow, and residues critical for activity are highlighted in red.[101,222]

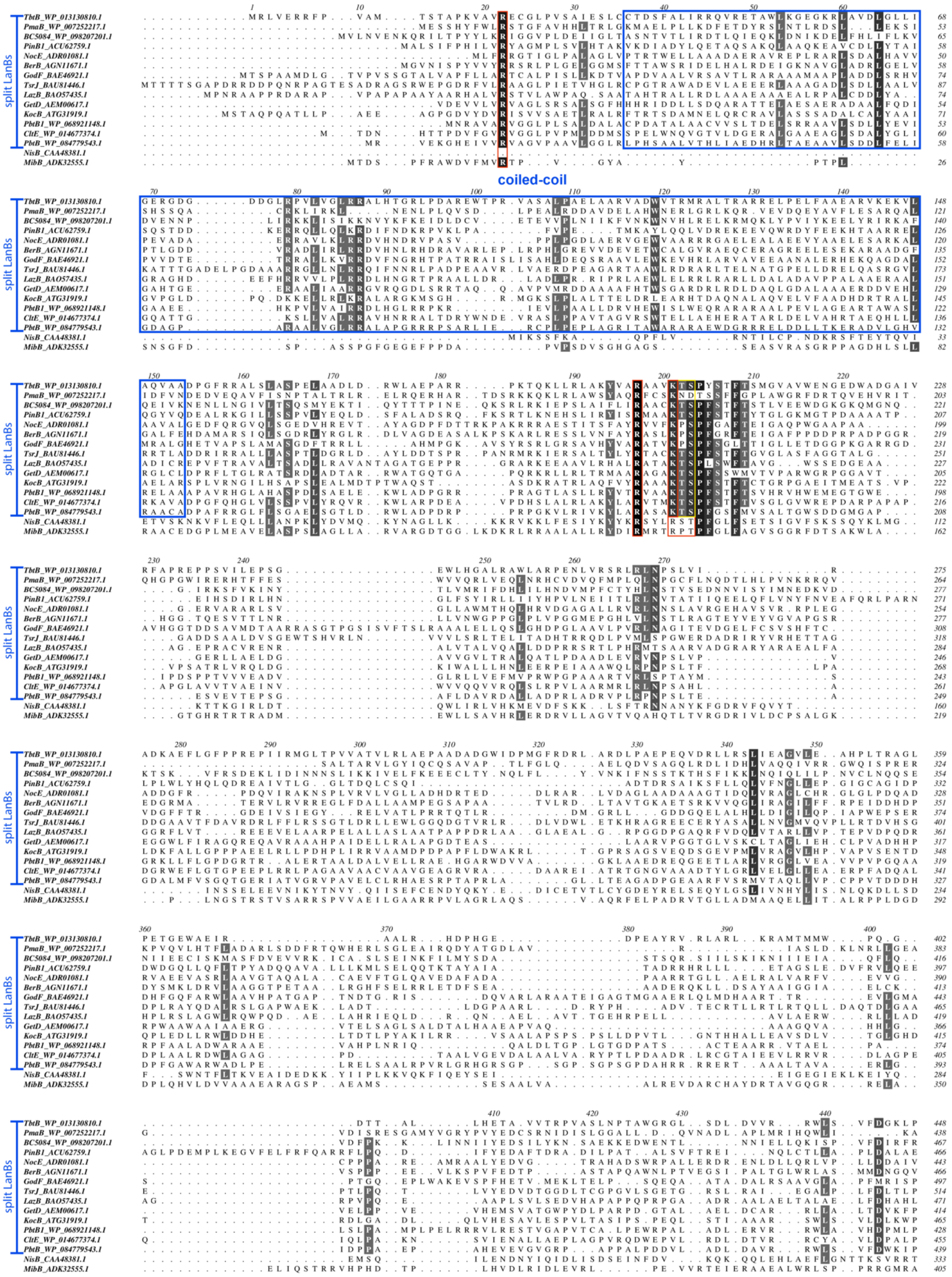


Fig. 7.4 (part 1 of 2)

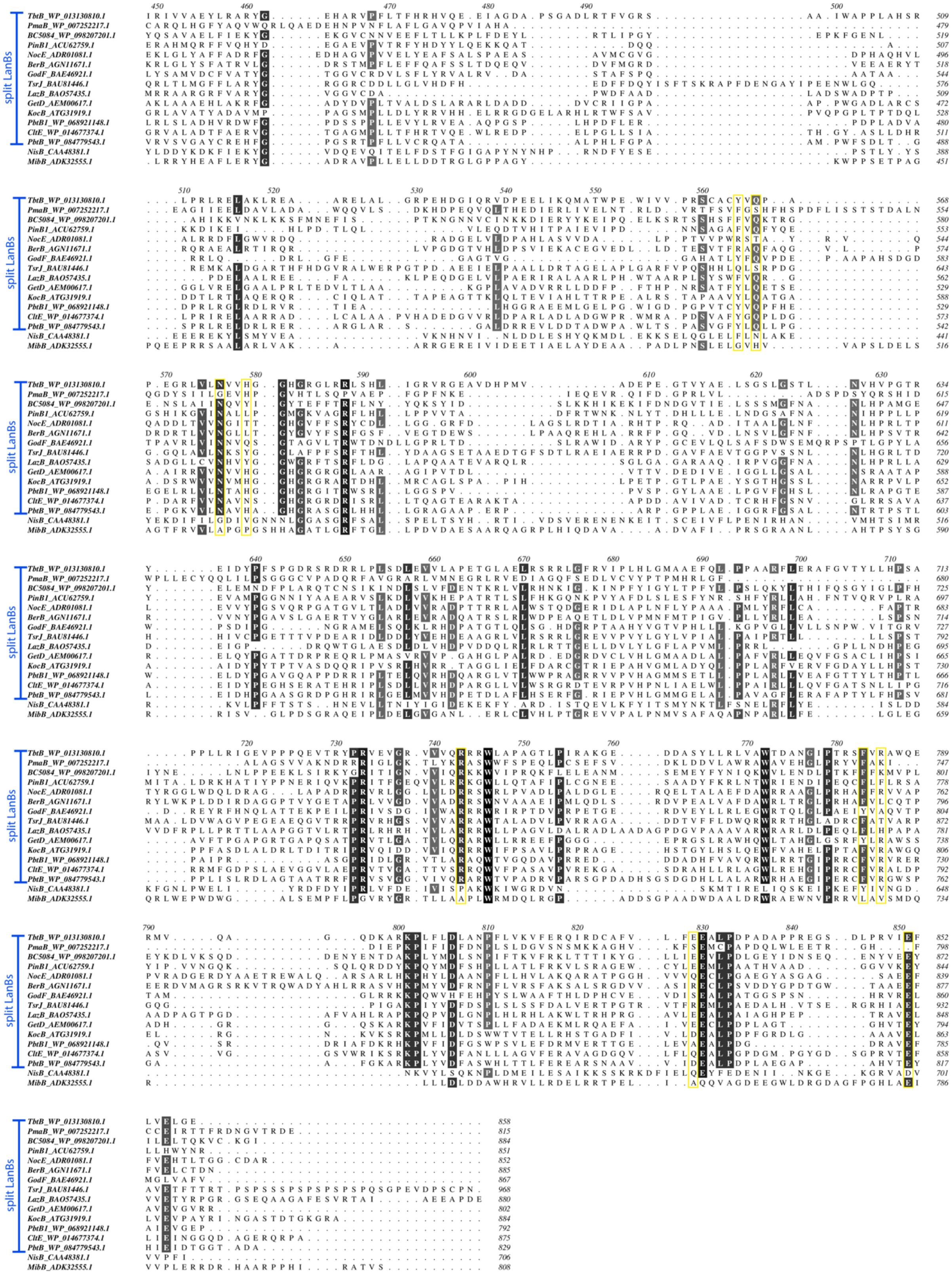


Fig 7.4 cont. (part 2 of 2)

Fig. 7.4 (cont.) Multisequence alignment of split LanBs and NisB/MibB glutamylation domains generated in Aline.[207] TbtB and NisB residues (numbered according to TbtB) targeted for mutagenesis which gave partial or no impairment to activity are highlighted in yellow, and residues critical for activity are highlighted in red (associated homologous sequences are also boxed). The coiled-coil subdomain boundaries are boxed in blue for the split LanB sequences. A majority (11/14) of the split LanB sequences are derived from thiopeptide biosynthetic gene clusters (BGCs) with the remaining three sequences from the pinensin BGC (PinB1), goadsporin BGC (GodF), and the as-yet uncharacterized PmaB from *Pseudomonas syringae.*[226–228]

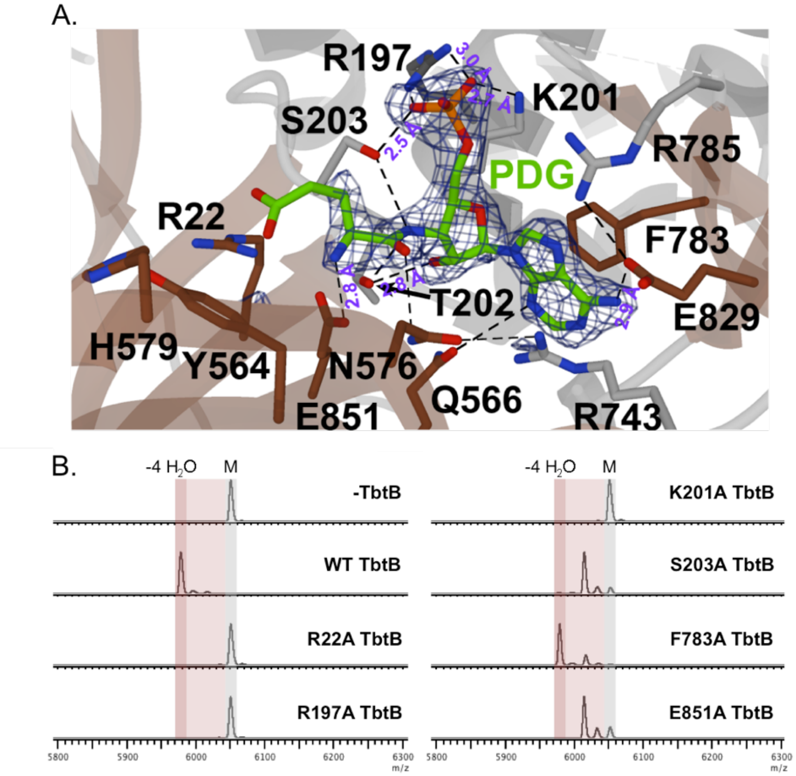


Fig. 7.5: Structural analysis of the glutamylation active site of TbtB. (A) TbtB bound to PDG (2.15 Å; PDB 6EC8) with active-site residues targeted for mutagenesis represented as sticks. A simulated annealing difference Fourier map (F_{obs} - F_{calc}) contoured at 2 σ is shown for the PDG ligand (green sticks). Interatomic distances ≤ 3 Å are labeled in purple. (B) Analysis of TbtB mutant activity *in vitro*. His_6 SUMO-TbtB mutants were incubated with TbtC, TbtA-hexazole, *T. thermobispora* tRNAGlu and GluRS before analysis by MALDI-MS (Figure credit [panel B]: Ian Bothwell).

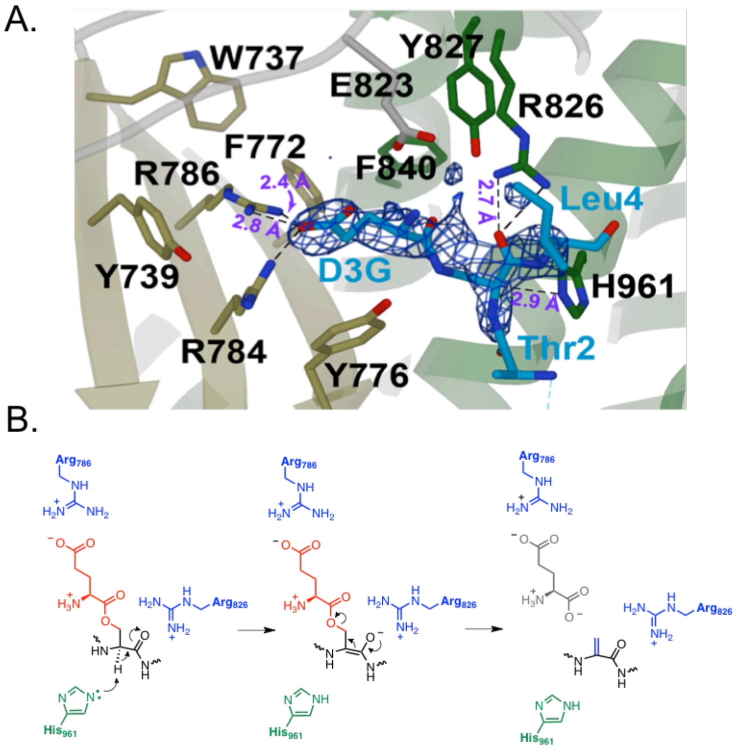


Fig. 7.6: (A) NisB-V169C bound to NisA-Dap$_3^{Glu}$ (2.79 Å; PDB 6M7Y) showing elimination domain active-site residues tested previously (Y776, R784, R786, E823, R826, Y827, and H961;[101,222] and in this study (W737, Y739, F772, and F840) as sticks. A simulated annealing difference Fourier map (F_{obs} - F_{calc}) contoured at 2 σ is shown for three residues of the CP (2-4 in blue sticks) containing the non-eliminable Dap$_3^{Glu}$ adduct. Interatomic distances < 3 Å are labeled in purple. (B) Proposed mechanism of elimination catalyzed by NisB (Figure credit [panel B]: Ian Bothwell).

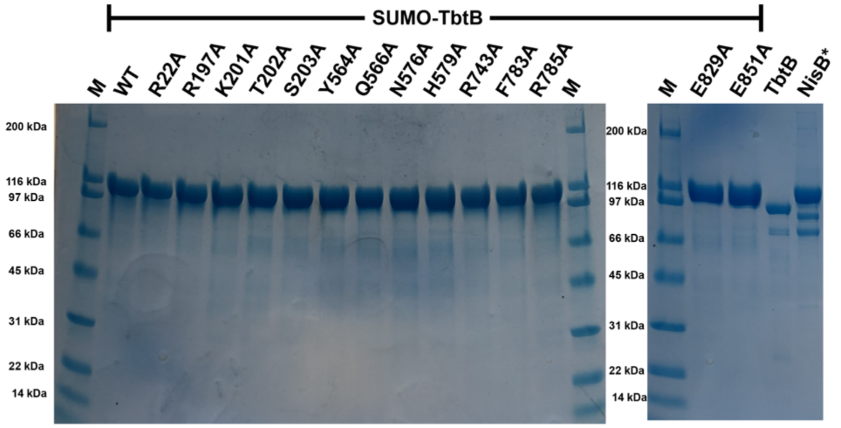


Fig. 7.7: SDS-PAGE analysis of N-terminal His$_6$-SUMO tagged TbtB (SUMO-TbtB) variants used in activity assays (Fig. 7.5B) in addition to tag-free TbtB and methylated His$_6$-SeMet-NisB-V169C (NisB*) used for crystallization. “M” denotes protein standard.

REFERENCES

(1) Bar-On, Y. M.; Phillips, R.; Milo, R. The Biomass Distribution on Earth. *Proc. Natl. Acad. Sci.* **2018**, *115* (25), 6506–6511.

(2) DeLong, E. F.; Pace, N. R.; Kane, M. Environmental Diversity of Bacteria and Archaea. *Syst. Biol.* **2001**, *50* (4), 470–478.

(3) Gerlt, J. a.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks. *Biochim. Biophys. Acta - Proteins Proteomics* **2015**, *1854* (8), 1019–1037.

(4) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79* (3), 629–661.

(5) Sheldon, R. A.; Woodley, J. M. Role of Biocatalysis in Sustainable Chemistry. *Chem. Rev.* **2018**, *118* (2), 801–838.

(6) Zaman, S. Bin; Hussain, M. A.; Nye, R.; Mehta, V.; Mamun, K. T.; Hossain, N. A Review on Antibiotic Resistance: Alarm Bells Are Ringing. *Cureus* **2017**, *9* (6).

(7) Cogan, D. P.; Baraquet, C.; Harwood, C. S.; Nair, S. K. Structural Basis of Transcriptional Regulation by Cour, a Repressor of Coumarate Catabolism, in Rhodopseudomonas Palustris. *J. Biol. Chem.* **2018**, *293* (30), 11727–11735.

(8) Hirakawa, H.; Schaefer, A. L.; Greenberg, E. P.; Harwood, C. S. Anaerobic P-Coumarate Degradation by Rhodopseudomonas Palustris and Identification of CouR, a MarR Repressor Protein That Binds p-Coumaroyl Coenzyme A. *J. Bacteriol.* **2012**, *194* (8), 1960–1967.

(9) Haque, M. M.; Kabir, M. S.; Aini, L. Q.; Hirata, H.; Tsuyumu, S. SlyA, a MarR Family Transcriptional Regulator, Is Essential for Virulence in Dickeya Dadantii 3937. *J. Bacteriol.* **2009**, *191* (17), 5409–5418.

(10) Davis, J. R.; Brown, B. L.; Page, R.; Sello, J. K. Study of PcaV from Streptomyces Coelicolor Yields New Insights into Ligand-Responsive MarR Family Transcription Factors. *Nucleic Acids Res.* **2013**, *41* (6), 3888–3900.

(11) Kaatz, G. W.; Demarco, C. E.; Seo, S. M. MepR , a Repressor of the Staphylococcus Aureus MATE Family Multidrug Efflux Pump MepA , Is a Substrate-Responsive Regulatory Protein MepR , a Repressor of the Staphylococcus Aureus MATE Family Multidrug Efflux Pump MepA , Is a Substrate-Responsive Regula. *Society* **2006**, *50* (4), 1276–1281.

(12) Hao, Z.; Lou, H.; Zhu, R.; Zhu, J.; Zhang, D.; Zhao, B. S.; Zeng, S.; Chen, X.; Chan, J.; He, C.; et al. The Multiple Antibiotic Resistance Regulator MarR Is a Copper Sensor in Escherichia Coli. *Nat. Chem. Biol.* **2014**, *10* (1), 21–28.

(13) Perez-Rueda, E.; Hernandez-Guerrero, R.; Martinez-Nuñez, M. A.; Armenta-Medina, D.; Sanchez, I.; Ibarra, J. A. Abundance, Diversity and Domain Architecture Variability in Prokaryotic DNA-Binding Transcription Factors. *PLoS One* **2018**, *13* (4), 1–16.

(14) Land, M.; Hauser, L.; Jun, S.-R.; Nookaew, I.; Leuze, M. R.; Ahn, T.-H.; Karpinets, T.;

Lund, O.; Kora, G.; Wassenaar, T.; et al. Insights from 20 Years of Bacterial Genome Sequencing. *Funct. Integr. Genomics* **2015**, *15* (2), 141–161.

(15) Chang, Y.-M.; Jeng, W.-Y.; Ko, T.-P.; Yeh, Y.-J.; Chen, C. K.-M.; Wang, A. H.-J. Structural Study of TcaR and Its Complexes with Multiple Antibiotics from Staphylococcus Epidermidis. *Proc. Natl. Acad. Sci.* **2010**, *107* (19), 8617–8622.

(16) Alekshun, M. N.; Levy, S. B.; Mealy, T. R.; Seaton, B. A.; Head, J. F. The Crystal Structure of MarR, a Regulator of Multiple Antibiotic Resistance, at 2.3 A Resolution. *Nat. Struct. Biol.* **2001**, *8* (8), 710–714.

(17) Larimer, F. W.; Chain, P.; Hauser, L.; Lamerdin, J.; Malfatti, S.; Do, L.; Land, M. L.; Pelletier, D. A.; Beatty, J. T.; Lang, A. S.; et al. Complete Genome Sequence of the Metabolically Versatile Photosynthetic Bacterium Rhodopseudomonas Palustris. *Nat. Biotechnol.* **2004**, *22* (1), 55–61.

(18) Egland, P. G.; Pelletier, D. A.; Dispensa, M.; Gibson, J.; Harwood, C. S. A Cluster of Bacterial Genes for Anaerobic Benzene Ring Biodegradation. *Microbiology* **1997**, *94* (June), 6484–6489.

(19) Phattarasukol, S.; Radey, M. C.; Lappala, C. R.; Oda, Y.; Hirakawa, H.; Brittnacher, M. J.; Harwood, C. S. Identification of a P-Coumarate Degradation Regulon in Rhodopseudomonas Palustris by Xpression, an Integrated Tool for Prokaryotic RNA-Seq Data Processing. *Appl. Environ. Microbiol.* **2012**, *78* (19), 6812–6818.

(20) Birukou, I.; Seo, S. M.; Schindler, B. D.; Kaatz, G. W.; Brennan, R. G. Structural Mechanism of Transcription Regulation of the Staphylococcus Aureus Multidrug Efflux Operon MepRA by the MarR Family Repressor MepR. *Nucleic Acids Res.* **2014**, *42* (4), 2774–2788.

(21) Kim, Y.; Joachimiak, G.; Bigelow, L.; Babnigg, G.; Joachimiak, A. How Aromatic Compounds Block Dna Binding of Hcar Catabolite Regulator. *J. Biol. Chem.* **2016**, *291* (25), 13243–13256.

(22) Lim, D.; Poole, K.; Strynadka, N. C. J. Crystal Structure of the MexR Repressor of the MexRAB-OprM Multidrug Efflux Operon of Pseudomonas Aeruginosa. *J. Biol. Chem.* **2002**, *277* (32), 29253–29259.

(23) Otani, H.; Stogios, P. J.; Xu, X.; Nocek, B.; Li, S. N.; Savchenko, A.; Eltis, L. D. The Activity of CouR, a MarR Family Transcriptional Regulator, Is Modulated through a Novel Molecular Mechanism. *Nucleic Acids Res.* **2016**, *44* (2), 595–607.

(24) Hong, M.; Fuangthong, M.; Helmann, J. D.; Brennan, R. G. Structure of an OhrR-OhrA Operator Complex Reveals the DNA Binding Mechanism of the MarR Family. *Mol. Cell* **2005**, *20* (1), 131–141.

(25) Dolan, K. T.; Duguid, E. M.; He, C. Crystal Structures of SlyA Protein, a Master Virulence Regulator of Salmonella, in Free and DNA-Bound States. *J. Biol. Chem.* **2011**, *286* (25), 22178–22185.

(26) Liu, G.; Liu, X.; Xu, H.; Liu, X.; Zhou, H.; Huang, Z.; Gan, J.; Chen, H.; Lan, L.; Yang, C. G. Structural Insights into the Redox-Sensing Mechanism of MarR-Type Regulator AbfR. *J. Am. Chem. Soc.* **2017**, *139* (4), 1598–1608.

(27) Duval, V.; McMurry, L. M.; Foster, K.; Head, J. F.; Levy, S. B. Mutational Analysis of the Multiple-Antibiotic Resistance Regulator MarR Reveals a Ligand Binding Pocket at the Interface between the Dimerization and DNA Binding Domains. *J. Bacteriol.* **2013**, *195* (15), 3341–3351.

(28) Davies, C.; Heath, R. J.; White, S. W.; Rock, C. O. The 1.8 Å Crystal Structure and Active-Site Architecture of β- Ketoacyl-Acyl Carrier Protein Synthase III (FabH) from Escherichia Coli. *Structure* **2000**, *8* (2), 185–195.

(29) Edwards, T. E.; Leibly, D. J.; Bhandari, J.; Statnekov, J. B.; Phan, I.; Dieterich, S. H.; Abendroth, J.; Staker, B. L.; Van Voorhis, W. C.; Myler, P. J.; et al. Structures of Phosphopantetheine Adenylyltransferase from Burkholderia Pseudomallei. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2011**, *67* (9), 1032–1037.

(30) Gao, Y. R.; Li, D. F.; Fleming, J.; Zhou, Y. F.; Liu, Y.; Deng, J. Y.; Zhou, L.; Zhou, J.; Zhu, G. F.; Zhang, X. E.; et al. Structural Analysis of the Regulatory Mechanism of MarR Protein Rv2887 in M. Tuberculosis. *Sci. Rep.* **2017**, *7* (1), 1–13.

(31) Vivoli, M.; Novak, H. R.; Littlechild, J. a; Harmer, N. J. Determination of Protein-Ligand Interactions Using Differential Scanning Fluorimetry. *J. Vis. Exp.* **2014**, No. 91, 51809.

(32) Xiong, Y.; Zhu, X.; Dai, H.; Wei, D.-Q. Survey of Computational Approaches for Prediction of DNA-Binding Residues on Protein Surfaces. In *Computational Systems Biology: Methods and Protocols*; Huang, T., Ed.; Springer New York: New York, NY, **2018**; pp 223–234.

(33) Terwilliger, T. C.; Adams, P. D.; Read, R. J.; McCoy, A. J.; Moriarty, N. W.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Zwart, P. H.; Hung, L. W. Decision-Making in Structure Solution Using Bayesian Estimates of Map Quality: The PHENIX AutoSol Wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2009**, *65* (6), 582–601.

(34) Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Zwart, P. H.; Hung, L. W.; Read, R. J.; Adams, P. D. Iterative Model Building, Structure Refinement and Density Modification with the PHENIX AutoBuild Wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2007**, *64* (1), 61–69.

(35) Vagin, A. a.; Steiner, R. a.; Lebedev, A. a.; Potterton, L.; McNicholas, S.; Long, F.; Murshudov, G. N. REFMAC5 Dictionary: Organization of Prior Chemical Knowledge and Guidelines for Its Use. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (12 I), 2184–2195.

(36) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (4), 486–501.

(37) Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D. Electronic Ligand Builder and Optimization Workbench (ELBOW): A Tool for Ligand Coordinate and Restraint Generation. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2009**, *65* (10), 1074–1080.

(38) 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. Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68* (4), 352–367.

(39) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; et al. Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature. *Nat. Prod. Rep.* **2013**, *30* (1), 108–160.

(40) Knerr, P. J.; van der Donk, W. A. Discovery, Biosynthesis, and Engineering of Lantipeptides. *Annu. Rev. Biochem.* **2012**, *81* (1), 479–505.

(41) Bierbaum, G.; Sahl, H.-G. Lantibiotics: Mode of Action, Biosynthesis and Bioengineering. *Curr. Pharm. Biotechnol.* **2009**, *10* (1), 2–18.

(42) Oman, T. J.; Van Der Donk, W. A. Follow the Leader: The Use of Leader Peptides to Guide Natural Product Biosynthesis. *Nat. Chem. Biol.* **2010**, *6* (1), 9–18.

(43) Yang, X.; Van Der Donk, W. A. Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: New Insights into the Role of Leader and Core Peptides during Biosynthesis. *Chem. - A Eur. J.* **2013**, *19* (24), 7662–7677.

(44) Velásquez, J. E.; Zhang, X.; Van Der Donk, W. A. Biosynthesis of the Antimicrobial Peptide Epilancin 15X and Its N-Terminal Lactate. *Chem. Biol.* **2011**, *18* (7), 857–867.

(45) Castiglione, F.; Lazzarini, A.; Carrano, L.; Corti, E.; Ciciliato, I.; Gastaldo, L.; Candiani, P.; Losi, D.; Marinelli, F.; Selva, E.; et al. Determining the Structure and Mode of Action of Microbisporicin, a Potent Lantibiotic Active Against Multiresistant Pathogens. *Chem. Biol.* **2008**, *15* (1), 22–31.

(46) Maffioli, S. I.; Monciardini, P.; Catacchio, B.; Mazzetti, C.; Münch, D.; Brunati, C.; Sahl, H. G.; Donadio, S. Family of Class i Lantibiotics from Actinomycetes and Improvement of Their Antibacterial Activities. *ACS Chem. Biol.* **2015**, *10* (4), 1034–1042.

(47) Maffioli, S. I.; Iorio, M.; Sosio, M.; Monciardini, P.; Gaspari, E.; Donadio, S. Characterization of the Congeners in the Lantibiotic NAI-107 Complex. *J. Nat. Prod.* **2014**, *77* (1), 79–84.

(48) Foulston, L. C.; Bibb, M. J. Microbisporicin Gene Cluster Reveals Unusual Features of Lantibiotic Biosynthesis in Actinomycetes. *Proc. Natl. Acad. Sci.* **2010**, *107* (30), 13461–13466.

(49) M̈nch, D.; Müller, A.; Schneider, T.; Kohl, B.; Wenzel, M.; Bandow, J. E.; Maffioli, S.; Sosio, M.; Donadio, S.; Wimmer, R.; et al. The Lantibiotic NAI-107 Binds to Bactoprenol-Bound Cell Wall Precursors and Impairs Membrane Functions. *J. Biol. Chem.* **2014**, *289* (17), 12063–12076.

(50) Jabés, D.; Brunati, C.; Candiani, G.; Riva, S.; Romanó, G.; Donadio, S. Efficacy of the New Lantibiotic NAI-107 in Experimental Infections Induced by Multidrug-Resistant Gram-Positive Pathogens. *Antimicrob. Agents Chemother.* **2011**, *55* (4), 1671–1676.

(51) Maffioli, S. I.; Cruz, J. C. S.; Monciardini, P.; Sosio, M.; Donadio, S. Advancing Cell Wall Inhibitors towards Clinical Applications. *J. Ind. Microbiol. Biotechnol.* **2016**, *43* (2–3), 177–184.

(52) Foulston, L.; Bibb, M. Feed-Forward Regulation of Microbisporicin Biosynthesis in Microbispora Corallina. *J. Bacteriol.* **2011**, *193* (12), 3064–3071.

(53) Ortega, M. A.; Hao, Y.; Walker, M. C.; Donadio, S.; Sosio, M.; Nair, S. K.; Van Der Donk, W. A. Structure and TRNA Specificity of MibB, a Lantibiotic Dehydratase from Actinobacteria Involved in NAI-107 Biosynthesis. *Cell Chem. Biol.* **2016**, *23* (3), 370–380.

(54) Kupke, T.; Stevanovic, S.; Sahl, H. G.; Gotz, F. Purification and Characterization of EpiD, a Flavoprotein Involved in the Biosynthesis of the Lantibiotic Epidermin. *J. Bacteriol.* **1992**, *174* (16), 5354–5361.

(55) Majer, F.; Schmid, D. G.; Altena, K.; Bierbaum, G.; Kupke, T. The Flavoprotein MrsD Catalyzes the Oxidative Decarboxylation Reaction Involved in Formation of the Peptidoglycan Biosynthesis Inhibitor Mersacidin. *J. Bacteriol.* **2002**, *184* (5), 1234–1243.

(56) Kupke, T.; Uebele, M.; Schmid, D.; Jung, G.; Blaessei, M.; Steinbacher, S. Molecular Characterization of Lantibiotic-Synthesizing Enzyme EpiD Reveals a Function for Bacterial Dfp Proteins in Coenzyme A Biosynthesis. *J. Biol. Chem.* **2000**, *275* (41), 31838–31846.

(57) Strauss, E.; Zhai, H.; Brand, L. A.; McLafferty, F. W.; Begley, T. P. Mechanistic Studies on Phosphopantothenoylcysteine Decarboxylase: Trapping of an Enethiolate Intermediate with a Mechanism-Based Inactivating Agent. *Biochemistry* **2004**, *43* (49), 15520–15533.

(58) Strauss, E.; Begley, T. P. Mechanistic Studies on Phosphopantothenoylcysteine Decarboxylase. *J. Am. Chem. Soc.* **2001**, *123* (26), 6449–6450.

(59) Blaesse, M.; Kupke, T.; Huber, R.; Steinbacher, S. Crystal Structure of the Peptidyl-Cysteine Decarboxylase EpiD Complexed with a Pentapeptide Substrate. *EMBO J.* **2000**, *19* (23), 6299–6310.

(60) Blaesse, M.; Kupke, T.; Huber, R.; Steinbacher, S. Structure of MrsD, an FAD-Binding Protein of the HFCD Family. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **2003**, *59* (8), 1414–1421.

(61) Kempter, C.; Kupke, T.; Kaiser, D.; Metzger, J. W.; Jung, G. Thioenols from Peptidyl Cysteines: Oxidative Decarboxylation of a LaC-Labeled Substrate. *Angew. Chemie (International Ed. English)* **1996**, *35* (18), 2104–2107.

(62) Dong, C.; Flecks, S.; Unversucht, S.; Haupt, C.; Naismith, J. H. The Structure of Tryptophan 7-Halogenase (PrnA) Suggests a Mechanism for Regioselective Chlorination. **2012**, *309* (5744), 2216–2219.

(63) Yeh, E.; Garneau, S.; Walsh, C. T. Robust in Vitro Activity of RebF and RebH, a Two-Component Reductase/Halogenase, Generating 7-Chlorotryptophan during Rebeccamycin Biosynthesis. *Proc. Natl. Acad. Sci.* **2005**, *102* (11), 3960–3965.

(64) Yeh, E.; Cole, L. J.; Barr, E. W.; Bollinger, J. M.; Ballou, D. P.; Walsh, C. T. Flavin Redox Chemistry Precedes Substrate Chlorination during the Reaction of the Flavin-Dependent Halogenase RebH. *Biochemistry* **2006**, *45* (25), 7904–7912.

(65) Yeh, E.; Blasiak, L. C.; Koglin, A.; Drennan, C. L.; Walsh, C. T. Chlorination by a Long-Lived Intermediate in the Mechanism of Flavin-Dependent Halogenases. *Biochemistry* **2007**, *46* (5), 1284–1292.

(66) Ortega, M. A.; Cogan, D. P.; Mukherjee, S.; Garg, N.; Li, B.; Thibodeaux, G. N.; Maffioli, S. I.; Donadio, S.; Sosio, M.; Escano, J.; et al. Two Flavoenzymes Catalyze the Post-Translational Generation of 5-Chlorotryptophan and 2-Aminovinyl-Cysteine during NAI-107 Biosynthesis. *ACS Chem. Biol.* **2017**, *12* (2), 548–557.

(67) Flecks, S.; Patallo, E. P.; Zhu, X.; Ernyei, A. J.; Seifert, G.; Schneider, A.; Dong, C.; Naismith, J. H.; Van Pée, K. H. New Insights into the Mechanism of Enzymatic Chlorination of Tryptophan. *Angew. Chemie - Int. Ed.* **2008**, *47* (49), 9533–9536.

(68) Cruz, J. C. S.; Iorio, M.; Monciardini, P.; Simone, M.; Brunati, C.; Gaspari, E.; Maffioli, S. I.; Wellington, E.; Sosio, M.; Donadio, S. Brominated Variant of the Lantibiotic NAI-107 with Enhanced Antibacterial Potency. *J. Nat. Prod.* **2015**, *78* (11), 2642–2647.

(69) Zhu, X.; De Laurentis, W.; Leang, K.; Herrmann, J.; Ihlefeld, K.; van Pée, K. H.; Naismith, J. H. Structural Insights into Regioselectivity in the Enzymatic Chlorination of Tryptophan. *J. Mol. Biol.* **2009**, *391* (1), 74–85.

(70) Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. CASTp: Computed Atlas of Surface Topography of Proteins with Structural and Topographical Mapping of Functionally Annotated Residues. *Nucleic Acids Res.* **2006**, *34* (WEB. SERV. ISS.), 116–118.

(71) Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.; Bricogne, G. Data Processing and Analysis with the AutoPROC Toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 293–302.

(72) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658–674.

(73) Bunkóczi, G.; Echols, N.; McCoy, A. J.; Oeffner, R. D.; Adams, P. D.; Read, R. J. Phaser.MRage: Automated Molecular Replacement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69* (11), 2276–2286.

(74) Cowtan, K. The Buccaneer Software for Automated Model Building. 1. Tracing Protein Chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, *62* (9), 1002–1011.

(75) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (12 I), 2126–2132.

(76) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 355–367.

(77) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461.

(78) Just-Baringo, X.; Albericio, F.; Álvarez, M. Thiopeptide Engineering: A Multidisciplinary Effort towards Future Drugs. *Angew. Chemie - Int. Ed.* **2014**, *53* (26), 6602–6616.

(79) Li, C.; Kelly, W. L. Recent Advances in Thiopeptide Antibiotic Biosynthesis. *Nat. Prod. Rep.* **2010**, *27* (2), 153–164.

(80) Zhang, Q.; Liu, W. Biosynthesis of Thiopeptide Antibiotics and Their Pathway Engineering. *Nat. Prod. Rep.* **2013**, *30* (2), 218–226.

(81) KETTENRING, J.; COLOMBO, L.; FERRARI, P.; TAVECCHIA, P.; NEBULONI, M.; VÉREY, K.; GALLO, G. G.; Selva, E. ANTIBIOTIC GE2270 A: A NOVEL INHIBITOR OF BACTERIAL PROTEIN SYNTHESIS. II. STRUCTURE ELUCIDATION. *J. Antibiot. (Tokyo).* **1991**, *44* (7).

(82) Morris, R. P.; Leeds, J. A.; Naegeli, H. U.; Oberer, L.; Memmert, K.; Weber, E.; Lamarche, M. J.; Parker, C. N.; Burrer, N.; Esterow, S.; et al. Elongation Factor Tu. **2009**, 5946–5955.

(83) Burkhart, B. J.; Hudson, G. A.; Dunbar, K. L.; Mitchell, D. A. A Prevalent Peptide-Binding Domain Guides Ribosomal Natural Product Biosynthesis. *Nat. Chem. Biol.* **2015**, *11* (8), 564–570.

(84) Zhang, F.; Kelly, W. L. *In Vivo Production of Thiopeptide Variants*, 1st ed.; Elsevier Inc., **2012**; Vol. 516.

(85) Tran, H. L.; Lexa, K. W.; Julien, O.; Young, T. S.; Walsh, C. T.; Jacobson, M. P.; Wells, J. A. Structure-Activity Relationship and Molecular Mechanics Reveal the Importance of Ring Entropy in the Biosynthesis and Activity of a Natural Product. *J. Am. Chem. Soc.* **2017**, *139* (7), 2541–2544.

(86) Zhang, Z.; Hudson, G. a; Mahanta, N.; Tietz, J. I.; van der Donk, W. a; Mitchell, D. a. Biosynthetic Timing and Substrate Specificity for the Thiopeptide Thiomuracin. *J. Am. Chem. Soc.* **2016**, 4–7.

(87) Burkhart, B. J.; Schwalen, C. J.; Mann, G.; Naismith, J. H.; Mitchell, D. A. YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function. *Chem. Rev.* **2017**, *117* (8), 5389–5456.

(88) Repka, L. M.; Chekan, J. R.; Nair, S. K.; Van Der Donk, W. A. Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. *Chem. Rev.* **2017**, *117* (8), 5457–5520.

(89) Hudson, G. a.; Zhang, Z.; Tietz, J. I.; Mitchell, D. a.; Van Der Donk, W. a. In Vitro Biosynthesis of the Core Scaffold of the Thiopeptide Thiomuracin. *J. Am. Chem. Soc.* **2015**, *137* (51), 16012–16015.

(90) Wever, W. J.; Bogart, J. W.; Baccile, J. A.; Chan, A. N.; Schroeder, F. C.; Bowers, A. A. Chemoenzymatic Synthesis of Thiazolyl Peptide Natural Products Featuring an Enzyme-Catalyzed Formal [4 + 2] Cycloaddition. *J. Am. Chem. Soc.* **2015**, *137* (10), 3494–3497.

(91) Jeon, B. S.; Wang, S. A.; Ruszczycky, M. W.; Liu, H. W. Natural [4 + 2]-Cyclases. *Chem. Rev.* **2017**, *117* (8), 5367–5388.

(92) Kim, H. J.; Ruszczycky, M. W.; Liu, H. wen. Current Developments and Challenges in the Search for a Naturally Selected Diels-Alderase. *Curr. Opin. Chem. Biol.* **2012**, *16* (1–2), 124–131.

(93) Hughes, R. A.; Thompson, S. P.; Alcaraz, L.; Moody, C. J. Total Synthesis of the Thiopeptide Antibiotic Amythiamicin D. *J. Am. Chem. Soc.* **2005**, *127* (44), 15644–

15651.

(94) Bowers, A. A.; Walsh, C. T.; Acker, M. G. Genetic Interception and Structural Characterization of Thiopeptide Cyclization Precursors from Bacillus Cereus. *J. Am. Chem. Soc.* **2010**, *132* (35), 12182–12184.

(95) Fage, C. D.; Isiorho, E. A.; Liu, Y.; Wagner, D. T.; Liu, H. W.; Keatinge-Clay, A. T. The Structure of SpnF, a Standalone Enzyme That Catalyzes [4 + 2] Cycloaddition. *Nat. Chem. Biol.* **2015**, *11* (4), 256–258.

(96) Kim, H. J.; Ruszczycky, M. W.; Choi, S. H.; Liu, Y. N.; Liu, H. W. Enzyme-Catalysed [4+2] Cycloaddition Is a Key Step in the Biosynthesis of Spinosyn A. *Nature* **2011**, *473* (7345), 109–112.

(97) Tian, Z.; Sun, P.; Yan, Y.; Wu, Z.; Zheng, Q.; Zhou, S.; Zhang, H.; Yu, F.; Jia, X.; Chen, D.; et al. An Enzymatic [4+2] Cyclization Cascade Creates the Pentacyclic Core of Pyrroindomycins. *Nat. Chem. Biol.* **2015**, *11* (4), 259–265.

(98) Li, L.; Yu, P.; Tang, M. C.; Zou, Y.; Gao, S. S.; Hung, Y. S.; Zhao, M.; Watanabe, K.; Houk, K. N.; Tang, Y. Biochemical Characterization of a Eukaryotic Decalin-Forming Diels-Alderase. *J. Am. Chem. Soc.* **2016**, *138* (49), 15837–15840.

(99) Ohashi, M.; Liu, F.; Hai, Y.; Chen, M.; Tang, M. C.; Yang, Z.; Sato, M.; Watanabe, K.; Houk, K. N.; Tang, Y. SAM-Dependent Enzyme-Catalysed Pericyclic Reactions in Natural Product Biosynthesis. *Nature* **2017**, *549* (7673), 502–506.

(100) Wever, W. J.; Bogart, J. W.; Bowers, A. A. Identification of Pyridine Synthase Recognition Sequences Allows a Modular Solid-Phase Route to Thiopeptide Variants. *J. Am. Chem. Soc.* **2016**, *138* (41), 13461–13464.

(101) Ortega, M. a; Hao, Y.; Zhang, Q.; Walker, M. C.; van der Donk, W. a; Nair, S. K. Structure and Mechanism of the TRNA-Dependent Lantibiotic Dehydratase NisB. *Nature* **2014**, *517* (7535), 509–512.

(102) Koehnke, J.; Mann, G.; Bent, A. F.; Ludewig, H.; Shirran, S.; Botting, C.; Lebl, T.; Houssen, W. E.; Jaspars, M.; Naismith, J. H. Structural Analysis of Leader Peptide Binding Enables Leader-Free Cyanobactin Processing. *Nat. Chem. Biol.* **2015**, *11* (8), 558–563.

(103) Cogan, D. P.; Hudson, G. A.; Zhang, Z.; Pogorelov, T. V.; van der Donk, W. A.; Mitchell, D. A.; Nair, S. K. Structural Insights into Enzymatic [4+2] *Aza* -Cycloaddition in Thiopeptide Antibiotic Biosynthesis. *Proc. Natl. Acad. Sci.* **2017**, 201716035.

(104) Marques, J. C.; Lamosa, P.; Russell, C.; Ventura, R.; Maycock, C.; Semmelhack, M. F.; Miller, S. T.; Xavier, K. B. Processing the Interspecies Quorum-Sensing Signal Autoinducer-2 (AI-2): Characterization of Phospho-(S)-4,5-Dihydroxy-2,3-Pentanedione Isomerization by LsrG Protein. *J. Biol. Chem.* **2011**, *286* (20), 18331–18343.

(105) Ma, B. G.; Chen, L.; Ji, H. F.; Chen, Z. H.; Yang, F. R.; Wang, L.; Qu, G.; Jiang, Y. Y.; Ji, C.; Zhang, H. Y. Characters of Very Ancient Proteins. *Biochem. Biophys. Res. Commun.* **2008**, *366* (3), 607–611.

(106) Werner, A. K.; Romeis, T.; Witte, C. P. Ureide Catabolism in Arabidopsis Thaliana and

Escherichia Coli. *Nat. Chem. Biol.* **2010**, *6* (1), 19–21.

(107) Shin, I.; Percudani, R.; Rhee, S. Structural and Functional Insights into (S)-Ureidoglycine Aminohydrolase, Keyenzyme of Purine Catabolism in Arabidopsis Thaliana. *J. Biol. Chem.* **2012**, *287* (22), 18796–18805.

(108) Koga, N.; Tatsumi-Koga, R.; Liu, G.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D. Principles for Designing Ideal Protein Structures. *Nature* **2012**, *491* (7423), 222–227.

(109) Nikolovska-Coleska, Z.; Wang, R.; Fang, X.; Pan, H.; Tomita, Y.; Li, P.; Roller, P. P.; Krajewski, K.; Saito, N. G.; Stuckey, J. a.; et al. Development and Optimization of a Binding Assay for the XIAP BIR3 Domain Using Fluorescence Polarization. *Anal. Biochem.* **2004**, *332* (2), 261–273.

(110) Martin, S. F.; Tatham, M. H.; Hay, R. T.; Samuel, I. D. W. Quantitative Analysis of Multi-Protein Interactions Using FRET: Application to the SUMO Pathway. *Protein Sci.* **2008**, *17* (4), 777–784.

(111) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658–674.

(112) Ashkenazy, H.; Erez, E.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2010: Calculating Evolutionary Conservation in Sequence and Structure of Proteins and Nucleic Acids. *Nucleic Acids Res.* **2010**, *38* (SUPPL. 2), 529–533.

(113) Donovick, R.; Pagano, J.; Stout, H.; Weinstein, M. Thiostrepton, a New Antibiotic. I. In Vitro Studies. *Antibiot. Annu.* **1955,** *3*, 554–559.

(114) Vandeputte, J.; Dutcher, J. Thiostrepton, a New Antibiotic. II. Isolation and Chemical Characterization. *Antibiot. Annu.* **1955,** 560–561.

(115) Tokura, K.; Tori, K.; Yoshimura, Y.; Okabe, K.; Otsuka, H.; Matsushita, K.; Inagaki, F.; Miyazawa, T. The Structure of Siomycin-D1, Peptede Antibiotic Isolated from Streptomyces Sioyaensis. *J. Antibiot. (Tokyo).* **1980**, *33* (12), 1563–1567.

(116) Miyairi, N.; Miyoshi, T.; Aoki, H.; Kohsaka, M.; Ikushima, H.; Kunugita, K.; Sakai, H.; Imanaka, H. Studies on Thiopeptin Antibiotics. I. Characteristics of Thiopeptin B. *J. Antibiot.* **1970**, *23*, 113–119.

(117) Puar, M.; Ganguly, A.; Afonso, A.; Brambilla, R.; Mangiaracina, P.; Sarre, O.; MacFarlane, R. Sch 18640. A New Thiostrepton-Type Antibiotic. *J. Am. Chem. Soc.* **1981**, *103* (17), 5231–5233.

(118) Halgren, T. a. Merck Molecular Force Field. *J. Comput. Chem.* **1996**, *17* (5–6), 490–519.

(119) Halgren, T. A. MMFF VI. MMFF94s Option for Energy Minimization Studies. *J Comput Chem* **1999**, *20* (7), 720–729.

(120) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling. *Bioinformatics* **2006**, *22* (2), 195–201.

(121) Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T. Protein Structure Homology Modeling Using SWISS-MODEL Workspace. *Nat. Protoc.* **2009**, *4* (1), 1–13.

(122) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; et al. SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information. *Nucleic Acids Res.* **2014**, *42* (W1), 252–258.

(123) Leach, A. R.; Kuntz, I. D. Conformational Analysis of Flexible Ligands in Macromolecular Receptor Sites. *J. Comput. Chem.* **1992**, *13* (6), 730–748.

(124) Corbeil, C. R.; Williams, C. I.; Labute, P. Variability in Docking Success Rates Due to Dataset Preparation. *J. Comput. Aided. Mol. Des.* **2012**, *26* (6), 775–786.

(125) Bond, S. D.; Leimkuhler, B. J.; Laird, B. B. The Nosé–Poincaré Method for Constant Temperature Molecular Dynamics. *J. Comput. Phys.* **1999**, *151* (1), 114–134.

(126) Kessler, H.; Steuernagel, S.; Will, M.; Jung, G.; Kellner, R.; Gillessen, D.; Kamiyama, T. The Structure of the Polycyclic Nonadecapeptide Ro 09-0198. *Helv. Chim. Acta* **1988**, *71* (8), 1924–1929.

(127) HAYASHI, F.; NAGASHIMA, K.; TERUI, Y.; KAWAMURA, Y.; MATSUMOTO, K.; ITAZAKI, H. The Structure of PA48009: The Revised Structure of Duramycin. *J. Antibiot. (Tokyo).* **1990**, *43* (11), 1421–1430.

(128) FREDENHAGEN, A.; FENDRICH, G.; MÄRKI, F.; MÄRKI, W.; GRUNER, J.; RASCHDORF, F.; PETER, H. H. Duramycins B and C, Two New Lanthionine Containing Antibiotics as Inhibitors of Phospholipase A2. Structural Revision of Duramycin and Cinnamycin. *J. Antibiot. (Tokyo).* **1990**, *43* (11), 1403–1412.

(129) Actinomycete, M. Crossm Mathermycin , a Lantibiotic from the Marine Actinomycete Marinactinospora. **2017**, *83* (15), 1–10.

(130) Kodani, S.; Komaki, H.; Ishimura, S.; Hemmi, H.; Ohnishi-Kameyama, M. Isolation and Structure Determination of a New Lantibiotic Cinnamycin B from Actinomadura Atramentaria Based on Genome Mining. *J. Ind. Microbiol. Biotechnol.* **2016**, *43* (8), 1159–1165.

(131) Smith, T. E.; Pond, C. D.; Pierce, E.; Harmer, Z. P.; Kwan, J.; Zachariah, M. M.; Harper, M. K.; Wyche, T. P.; Matainaho, T. K.; Bugni, T. S.; et al. Accessing Chemical Diversity from the Uncultivated Symbionts of Small Marine Animals. *Nat. Chem. Biol.* **2018**, *14* (2), 179–185.

(132) Pomorski, T.; Hrafnsdóttir, S.; Devaux, P. F.; Van Meer, G. Lipid Distribution and Transport across Cellular Membranes. *Semin. Cell Dev. Biol.* **2001**, *12* (2), 139–148.

(133) Iwamoto, K.; Hayakawa, T.; Murate, M.; Makino, A.; Ito, K.; Fujisawa, T.; Kobayashi, T. Curvature-Dependent Recognition of Ethanolamine Phospholipids by Duramycin and Cinnamycin. *Biophys. J.* **2007**, *93* (5), 1608–1619.

(134) Hosoda, K.; Ohya, M.; Kohno, T.; Maeda, T.; Endo, S.; Wakamatsu, K. Structure Determination of an Immunopotentiator Peptide, Cinnamycin, Complexed with Lysophosphatidylethanolamine By1H-NMR. *J. Biochem.* **1996**, *119* (2), 226–230.

(135) Richard, A. S.; Zhang, A.; Park, S.-J.; Farzan, M.; Zong, M.; Choe, H. Virion-Associated Phosphatidylethanolamine Promotes TIM1-Mediated Infection by Ebola, Dengue, and

West Nile Viruses. *Proc. Natl. Acad. Sci.* **2015**, *112* (47), 14682–14687.

(136) An, L.; Cogan, D. P.; Navo, C. D.; Jiménez-Osés, G.; Nair, S. K.; van der Donk, W. A. Substrate-Assisted Enzymatic Formation of Lysinoalanine in Duramycin. *Nat. Chem. Biol.* **2018,** 14 (10), 928-933.

(137) Oliynyk, I.; Varelogianni, G.; Roomans, G. M.; Johannesson, M. Effect of Duramycin on Chloride Transport and Intracellular Calcium Concentration in Cystic Fibrosis and Non-Cystic Fibrosis Epithelia. *Apmis* **2010**, *118* (12), 982–990.

(138) Zhao, M.; Li, Z.; Bugenhagen, S. 99mTc-Labeled Duramycin as a Novel Phosphatidylethanolamine-Binding Molecular Probe. *J. Nucl. Med.* **2008**, *49* (8), 1345–1352.

(139) Delvaeye, T.; wyffels, L.; Deleye, S.; Lemeire, K.; Gonçalves, A.; Decrock, E.; Staelens, S.; Leybaert, L.; Vandenabeele, P.; Krysko, D. Noninvasive Whole-Body Imaging of Phosphatidylethanolamine as a Cell Death Marker Using ^{99m} Tc-Duramycin during TNF-Induced SIRS. *J. Nucl. Med.* **2018**, *59* (7), jnumed.117.205815.

(140) Huo, L.; Ökesli, A.; Zhao, M.; van der Donk, W. A. Insights into the Biosynthesis of Duramycin. *Appl. Environ. Microbiol.* **2017**, *83* (3), 1–12.

(141) Ökesli, A.; Cooper, L. E.; Fogle, E. J.; van der Donk, W. A. Nine Post-Translational Modifications during the Biosynthesis of Cinnamycin. *J. Am. Chem. Soc.* **2011**, *133* (34), 13753–13760.

(142) Burley, S. K.; Berman, H. M.; Christie, C.; Duarte, J. M.; Feng, Z.; Westbrook, J.; Young, J.; Zardecki, C. RCSB Protein Data Bank: Sustaining a Living Digital Data Resource That Enables Breakthroughs in Scientific Research and Biomedical Education. *Protein Sci.* **2018**, *27* (1), 316–330.

(143) Holm, L.; Rosenström, P. Dali Server: Conservation Mapping in 3D. *Nucleic Acids Res.* **2010**, *38* (Web Server issue), W545-9.

(144) Tang, W.; Jiménez-Osés, G.; Houk, K. N.; Van Der Donk, W. A. Substrate Control in Stereoselective Lanthionine Biosynthesis. *Nat. Chem.* **2015**, *7* (1), 57–64.

(145) Widdick, D. A.; Dodd, H. M.; Barraille, P.; White, J.; Stein, T. H.; Chater, K. F.; Gasson, M. J.; Bibb, M. J. Cloning and Engineering of the Cinnamycin Biosynthetic Gene Cluster from Streptomyces Cinnamoneus Cinnamoneus DSM 40005. *Proc. Natl. Acad. Sci.* **2003**, *100* (7), 4316–4321.

(146) Vonrhein, C.; Blanc, E.; Roversi, P.; Bricogne, G. Automated Structure Solution With AutoSHARP BT - Macromolecular Crystallography Protocols: Volume 2: Structure Determination; Doublié, S., Ed.; Humana Press: Totowa, NJ, **2007;** pp 215–230.

(147) Ding, W.; Williams, D. R.; Northcote, P.; Siegel, M. M.; Tsao, R.; Ashcroft, J.; Morton, G. O.; Alluri, M.; Abbanat, D.; Maiese, W. M. Pyrroindomycins, Novel Antibiotics Produced by Streptomyces Rugosporus Sp. LL-42D005. I. Isolation and Structure Determination. *J. Antibiot. (Tokyo).* **1994**, *47* (11), 1250–1257.

(148) Gullo, V. P.; Zimmerman, S. B.; Dewey, R. S.; Hensens, O.; Cassidy, P. J.; Oiwa, R.; Omura, S. Factumycin, a New Antibiotic (A40A): Fermentation, Isolation and

Antibacterial Spectrum. *J Antibiot (Tokyo).* **1982**, *35* (12), 1705–7.

(149) Kanazawa, S.; Fusetani, N.; Matsunaga, S. Cylindramide: Cytotoxic Tetramic Acid Lactam from the Marine Sponge Halichondria Cylindrata Tanita & Hoshino. *Tetrahedron Lett.* **1993**, *34* (6), 1065–1068.

(150) Gunasekera, S. P.; Gunasekera, M.; McCarthy, P. Discodermide: A New Bioactive Macrocyclic Lactam from the Marine Sponge Discodermia Dissoluta. *J. Org. Chem.* **1991**, *56* (16), 4830–4833.

(151) Kakule, T. B.; Sardar, D.; Lin, Z.; Schmidt, E. W. Two Related Pyrrolidinedione Synthetase Loci in Fusarium Heterosporum ATCC 74349 Produce Divergent Metabolites. *ACS Chem. Biol.* **2013**, *8* (7), 1549–1557.

(152) Hopmann, C.; Kurz, M.; Brönstrup, M.; Wink, J.; LeBeller, D. Isolation and Structure Elucidation of Vancoresmycin—a New Antibiotic from Amycolatopsis Sp. ST 101170. *Tetrahedron Lett.* **2002**, *43* (3), 435–438.

(153) Graupner, P. R.; Thornburgh, S.; Mathieson, J. T.; Chapin, E. L.; Kemmitt, G. M.; Brown, J. M.; Snipes, C. E. Dihydromaltophilin; a Novel Fungicidal Tetramic Acid Containing Metabolite from Streptomyces Sp. *J. Antibiot. (Tokyo).* **1997**, *50* (12), 1014–1019.

(154) MacKellar, F. A.; Grostic, M. F.; Olson, E. C.; Wnuk, R. J.; Branfman, A. R.; Rinehart, K. L. Tirandamycin. I. Structure Assignment. *J. Am. Chem. Soc.* **1971**, *93* (19), 4943–4945.

(155) Linington, R. G.; Clark, B. R.; Trimble, E. E.; Uren, L.; Kyle, D. E.; Gerwick, W. H.; Kyle, D. E.; Gerwick, W. H. Antimalarial Peptides from Marine Cyanobacteria : Isolation and Structural Elucidation of Gallinamide A Antimalarial Peptides from Marine Cyanobacteria : Isolation and Structural Elucidation of Gallinamide A. **2009**, 14–17.

(156) Kakule, T. B.; Zhang, S.; Zhan, J.; Schmidt, E. W. Biosynthesis of the Tetramic Acids Sch210971 and Sch210972. *Org. Lett.* **2015**, *17* (10), 2295–2297.

(157) Blodgett, J. A. V; Oh, D.; Cao, S.; Currie, C. R.; Kolter, R.; Clardy, J. Common Biosynthetic Origins for Polycyclic Tetramate Macrolactams from Phylogenetically Diverse Bacteria. *Proc. Natl. Acad. Sci.* **2010,** 107 (26), 11692-11697.

(158) Royles, B. J. L. Naturally Occurring Tetramic Acids: Structure, Isolation, and Synthesis. *Chem. Rev.* **1995**, *95* (6), 1981–2001.

(159) Schobert, R.; Schlenk, A. Tetramic and Tetronic Acids: An Update on New Derivatives and Biological Aspects. *Bioorganic Med. Chem.* **2008**, *16* (8), 4203–4221.

(160) De Silva, E. D.; Geiermann, A. S.; Mitova, M. I.; Kuegler, P.; Blunt, J. W.; Cole, A. L. J.; Mumo, M. H. G. Isolation of 2-Pyridone Alkaloids from a New Zealand Marine-Derived Penicillium Species. *J. Nat. Prod.* **2009**, *72* (3), 477–479.

(161) Sims, J. W.; Schmidt, E. W. Thioesterase-like Role for Fungal PKS-NRPS Hybrid Reductive Domains. *J. Am. Chem. Soc.* **2008**, *130* (33), 11149–11155.

(162) Liu, X.; Walsh, C. T. Cyclopiazonic Acid Biosynthesis in Aspergillus Sp.: Characterization of a Reductase-like R* Domain in Cyclopiazonate Synthetase That Forms and Releases Cyclo-Acetoacetyl-L-Tryptophan. *Biochemistry* **2009**, *48* (36), 8746–8757.

(163) Li, Y.; Chen, H.; Ding, Y.; Xie, Y.; Wang, H.; Cerny, R. L.; Shen, Y.; Du, L. Iterative Assembly of Two Separate Polyketide Chains by the Same Single-Module Bacterial Polyketide Synthase in the Biosynthesis of HSAF. *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (29), 7524–7530.

(164) Lou, L.; Qian, G.; Xie, Y.; Hang, J.; Chen, H.; Zaleta-Rivera, K.; Li, Y.; Shen, Y.; Dussault, P. H.; Liu, F.; et al. Biosynthesis of HSAF, a Tetramic Acid-Containing Macrolactam from Lysobacter Enzymogenes. *J. Am. Chem. Soc.* **2011**, *133* (4), 643–645.

(165) Antosch, J.; Schaefers, F.; Gulder, T. a M. Heterologous Reconstitution of Ikarugamycin Biosynthesis in E. Coli. *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (11), 3011–3014.

(166) Zhang, G.; Zhang, W.; Zhang, Q.; Shi, T.; Ma, L.; Zhu, Y.; Li, S.; Zhang, H.; Zhao, Y. L.; Shi, R.; et al. Mechanistic Insights into Polycycle Formation by Reductive Cyclization in Ikarugamycin Biosynthesis. *Angew. Chemie - Int. Ed.* **2014**, *53* (19), 4840–4844.

(167) Fujii, I.; Watanabe, A.; Sankawa, U.; Ebizuka, Y. Identification of Claisen Cyclase Domain in Fungal Polyketide Synthase WA, a Naphthopyrone Synthase of Aspergillus Nidulans. *Chem. Biol.* **2001**, *8* (2), 189–197.

(168) Vagstad, A. L.; Hill, E. a; Labonte, J. W.; Townsend, C. a. Characterization of a Fungal Thioesterase Having Claisen Cyclase and Deacetylase Activities in Melanin Biosynthesis. *Chem. Biol.* **2012**, *19* (12), 1525–1534.

(169) Korman, T. P.; Crawford, J. M.; Labonte, J. W.; Newman, A. G.; Wong, J.; Townsend, C. A.; Tsai, S.-C. Structure and Function of an Iterative Polyketide Synthase Thioesterase Domain Catalyzing Claisen Cyclization in Aflatoxin Biosynthesis. *Proc. Natl. Acad. Sci.* **2010**, *107* (14), 6246–6251.

(170) Wu, Q.; Wu, Z.; Qu, X.; Liu, W. Insights into Pyrroindomycin Biosynthesis Reveal a Uniform Paradigm for Tetramate/Tetronate Formation. *J. Am. Chem. Soc.* **2012**, *134* (42), 17342–17345.

(171) Gui, C.; Li, Q.; Mo, X.; Qin, X.; Ma, J.; Ju, J. Discovery of a New Family of Dieckmann Cyclases Essential to Tetramic Acid and Pyridone-Based Natural Products Biosynthesis. *Org. Lett.* **2015**, *17* (3), 628–631.

(172) Mo, X.; Shi, C.; Gui, C.; Zhang, Y.; Ju, J.; Wang, Q. Identification of Nocamycin Biosynthetic Gene Cluster from Saccharothrix Syringae NRRL B-16468 and Generation of New Nocamycin Derivatives by Manipulating Gene Cluster. *Microb. Cell Fact.* **2017**, *16* (1), 1–13.

(173) Liu, Y.; Zheng, T.; Bruner, S. D. Structural Basis for Phosphopantetheinyl Carrier Domain Interactions in the Terminal Module of Nonribosomal Peptide Synthetases. *Chem. Biol.* **2011**, *18* (11), 1482–1488.

(174) Aggarwal, A.; Parai, M. K.; Shetty, N.; Wallis, D.; Woolhiser, L.; Hastings, C.; Dutta, N. K.; Galaviz, S.; Dhakal, R. C.; Shrestha, R.; et al. Development of a Novel Lead That Targets M. Tuberculosis Polyketide Synthase 13. *Cell* **2017**, *170* (2), 249–259.e25.

(175) Drake, E. J.; Miller, B. R.; Shi, C.; Tarrasch, J. T.; Sundlov, J. A.; Leigh Allen, C.; Skiniotis, G.; Aldrich, C. C.; Gulick, A. M. Structures of Two Distinct Conformations of Holo-Non-Ribosomal Peptide Synthetases. *Nature* **2016**, *529* (7585), 235–238.

(176) Chakravarty, B.; Gu, Z.; Chirala, S. S.; Wakil, S. J.; Quiocho, F. A. Human Fatty Acid Synthase: Structure and Substrate Selectivity of the Thioesterase Domain. *Proc Natl Acad Sci U S A* **2004**, *101* (44), 15567–15572.

(177) Pemble, C. W.; Johnson, L. C.; Kridel, S. J.; Lowther, W. T. Crystal Structure of the Thioesterase Domain of Human Fatty Acid Synthase Inhibited by Orlistat. *Nat. Struct. Mol. Biol.* **2007**, *14* (8), 704–709.

(178) Moche, M.; Schneider, G.; Edwards, P.; Dehesh, K.; Lindqvist, Y. Structure of the Complex between the Antibiotic Cerulenin and Its Target, β-Ketoacyl-Acyl Carrier Protein Synthase. *J. Biol. Chem.* **1999**, *274* (10), 6031–6034.

(179) Johansson, P.; Mulinacci, B.; Koestler, C.; Vollrath, R.; Oesterhelt, D.; Grininger, M. Multimeric Options for the Auto-Activation of the Saccharomyces Cerevisiae FAS Type I Megasynthase. *Structure* **2009**, *17* (8), 1063–1074.

(180) Tang, Y.; Chen, A. Y.; Kim, C. Y.; Cane, D. E.; Khosla, C. Structural and Mechanistic Analysis of Protein Interactions in Module 3 of the 6-Deoxyerythronolide B Synthase. *Chem. Biol.* **2007**, *14* (8), 931–943.

(181) Trajtenberg, F.; Altabe, S.; Larrieux, N.; Ficarra, F.; De Mendoza, D.; Buschiazzo, A.; Schujman, G. E. Structural Insights into Bacterial Resistance to Cerulenin. *FEBS J.* **2014**, *281* (10), 2324–2338.

(182) Price, A. C.; Choi, K. H.; Heath, R. J.; Li, Z.; White, S. W.; Rock, C. O. Inhibition of β-Ketoacyl-Acyl Carrier Protein Synthases by Thiolactomycin and Cerulenin: Structure and Mechanism. *J. Biol. Chem.* **2001**, *276* (9), 6551–6559.

(183) Goblirsch, B. R.; Frias, J. A.; Wackett, L. P.; Wilmot, C. M. Crystal Structures of Xanthomonas Campestris OleA Reveal Features That Promote Head-to-Head Condensation of Two Long-Chain Fatty Acids. *Biochemistry* **2012**, *51* (20), 4138–4146.

(184) Jensen, M. R.; Goblirsch, B. R.; Christenson, J. K.; Esler, M. A.; Mohamed, F. A.; Wackett, L. P.; Wilmot, C. M. OleA Glu117 Is Key to Condensation of Two Fatty-Acyl Coenzyme A Substrates in Long-Chain Olefin Biosynthesis. *Biochem. J.* **2017**, *474* (23), 3871–3886.

(185) Funabashi, H.; Iwasaki, S.; Okuda, S. A Model Study on the Mechanism of Fatty Acid Synthetase Inhibition by Antibiotic Cerulenin. *Tetrahedron Lett.* **1983**, *24* (26), 2673–2676.

(186) Nemeria, N.; Yan, Y.; Zhang, Z.; Brown, A. M.; Arjunan, P.; Furey, W.; Guest, J. R.; Jordan, F. Inhibition of the Escherichia Coli Pyruvate Dehydrogenase Complex E1 Subunit and Its Tyrosine 177 Variants by Thiamin 2-Thiazolone and Thiamin 2-Thiothiazolone Diphosphates: Evidence for Reversible Tight-Binding Inhibition. *J. Biol. Chem.* **2001**, *276* (49), 45969–45978.

(187) Raychaudhuri, A.; Jerga, A.; Tipton, P. A. Chemical Mechanism and Substrate Specificity of RhlI, an Acylhomoserine Lactone Synthase from Pseudomonas Aeruginosa. *Biochemistry* **2005**, *44* (8), 2974–2981.

(188) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.

(189) Quadri, L. E. N.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. Characterization of Sfp, a Bacillus Subtilis Phosphopantetheinyl Transferase for Peptidyl Carder Protein Domains in Peptide Synthetases. *Biochemistry* **1998**, *37* (6), 1585–1595.

(190) Frueh, D. P.; Arthanari, H.; Koglin, A.; Vosburg, D. A.; Bennett, A. E.; Walsh, C. T.; Wagner, G. Dynamic Thiolation-Thioesterase Structure of a Non-Ribosomal Peptide Synthetase. *Nature* **2008**, *454* (7206), 903–906.

(191) Kepplinger, B.; Morton-Laing, S.; Seistrup, K. H.; Marrs, E. C. L.; Hopkins, A. P.; Perry, J. D.; Strahl, H.; Hall, M. J.; Errington, J.; Ellis Allenby, N. E. Mode of Action and Heterologous Expression of the Natural Product Antibiotic Vancoresmycin. *ACS Chem. Biol.* **2018**, *13* (1), 207–214.

(192) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; et al. AntiSMASH 3.0-A Comprehensive Resource for the Genome Mining of Biosynthetic Gene Clusters. *Nucleic Acids Res.* **2015**, *43* (W1), W237–W243.

(193) Bachmann, B. O.; Ravel, J. *Chapter 8 Methods for In Silico Prediction of Microbial Polyketide and Nonribosomal Peptide Biosynthetic Pathways from DNA Sequence Data*, 1st ed.; Elsevier Inc., **2009;** Vol. 458.

(194) Bailey, T. L.; Boden, M.; Buske, F. a.; Frith, M.; Grant, C. E.; Clementi, L.; Ren, J.; Li, W. W.; Noble, W. S. MEME Suite: Tools for Motif Discovery and Searching. *Nucleic Acids Res.* **2009**, *37* (SUPPL. 2), 202–208.

(195) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46* (May), gky427.

(196) Bosserman, M. A.; Shepherd, M. D.; Kharel, M. K.; Rohr, J. Laboratory Maintenance of Streptomyces Argillaceus and Streptomyces Griseus. *Curr. Protoc. Microbiol.* **2010**, No. SUPPL.19, 1–4.

(197) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. a; Smith, H. O.; Iii, C. a H.; America, N. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6* (5), 343–345.

(198) Bradford, M. M. A Rapid and Sensitive Method for the Quantitation Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, *72*, 248–254.

(199) Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (2), 125–132.

(200) Schüttelkopf, A. W.; Van Aalten, D. M. F. PRODRG: A Tool for High-Throughput Crystallography of Protein-Ligand Complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (8), 1355–1363.

(201) Christmas, Rowan; Avila-Campillo, Iliana; Bolouri, Hamid; Schwikowski, Benno; Anderson, Mark; Kelley, Ryan; Landys, Nerius; Workman, Chris; Ideker, Trey; Cerami, Ethan; Sheridan, Rob; Bader, Gary D.; Sander, C. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Am. Assoc. Cancer Res.*

Educ. B. **2005**, No. Karp 2001, 12–16.

(202) Tietz, J. I. T.; Schwalen, C. J.; Patel, P. S.; Maxon, T.; Blair, P. M.; Tai, H.-C.; Zakai, U. I. Z.; Mitchell, D. a. A New Genome Mining Tool Redefines the Lasso Peptide Biosynthetic Landscape. *Nat. Chem. Biol.* **2017**, 13 (5), 470-478.

(203) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. The Specificity-Conferring Code of Adenylation Domains in Nonribosomal Peptide Synthetases. *Chem. Biol.* **1999**, *6* (8), 493–505.

(204) Challis, G. L.; Ravel, J.; Townsend, C. A. Predictive, Structure-Based Model of Amino Acid Recognition by Nonribosomal Peptide Synthetase Adenylation Domains. *Chem. Biol.* **2000**, *7* (3), 211–224.

(205) Horna, D. H.; Gómez, C.; Olano, C.; Palomino-Schötzlein, M.; Pineda-Lucena, A.; Carbajo, R. J.; Braña, A. F.; Méndez, C.; Salas, J. A. Biosynthesis of the RNA Polymerase Inhibitor Streptolydigin in Streptomyces Lydicus: Tailoring Modification of 3-Methyl-Aspartate. *J. Bacteriol.* **2011**, *193* (10), 2647–2651.

(206) Brückner, R.; Hofferberth, M. L. *The Tetramic Acid Antibiotics α- and β-Lipomycin: Total Synthesis and Assignment of the Absolute Configuration of All Stereogenic Centers. Making Doubly Sure*; **2017;** Vol. 12.

(207) Bond, C. S.; Schüttelkopf, A. W. ALINE: A WYSIWYG Protein-Sequence Alignment Editor for Publication-Quality Alignments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2009**, *65* (5), 510–512.

(208) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci.* **2001**, *98* (18), 10037–10041.

(209) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.

(210) Tietz, J. I.; Schwalen, C. J.; Patel, P. S.; Maxson, T.; Blair, P. M.; Tai, H. C.; Zakai, U. I.; Mitchell, D. A. A New Genome-Mining Tool Redefines the Lasso Peptide Biosynthetic Landscape. *Nat. Chem. Biol.* **2017**, *13* (5), 470–478.

(211) Renevey, A.; Riniker, S. The Importance of N-Methylations for the Stability of the B6.3-Helical Conformation of Polytheonamide B. *Eur. Biophys. J.* **2017**, *46* (4), 363–374.

(212) Hamada, T.; Sugawara, T.; Matsunaga, S.; Fusetani, N. Polytheonamides, Unprecedented Highly Cytotoxic Polypeptides, from the Marine Sponge Theonella Swinhoei: 1. Isolation and Component Amino Acids. *Tetrahedron Lett.* **1994**, *35* (5), 719–720.

(213) Hamada, T.; Matsunaga, S.; Yano, G.; Fusetani, N. Polytheonamides A and B, Highly Cytotoxic, Linear Polypeptides with Unprecedented Structural Features, from the Marine Sponge, Theonella Swinhoei. *J. Am. Chem. Soc.* **2005**, *127* (1), 110–118.

(214) Hamada, T.; Matsunaga, S.; Fujiwara, M.; Fujita, K.; Hirota, H.; Schmucki, R.; Güntert, P.; Fusetani, N. Solution Structure of Polytheonamide B, a Highly Cytotoxic Nonribosomal Polypeptide from Marine Sponge. *J. Am. Chem. Soc.* **2010**, *132* (37),

12941–12945.

(215) Freeman, M. F.; Gurgui, C.; Helf, M. J.; Morinaka, B. I.; Uria, A. R.; Oldham, N. J.; Sahl, H.; Matsunaga, S.; Piel, J. Metagenome Mining Reveals Polytheonamides as Modified Ribosomal Peptides. **2012**, No. October, 387–391.

(216) Wilson, M. C.; Mori, T.; Rückert, C.; Uria, A. R.; Helf, M. J.; Takada, K.; Gernert, C.; Steffens, U. A. E.; Heycke, N.; Schmitt, S.; et al. An Environmental Bacterial Taxon with a Large and Distinct Metabolic Repertoire. *Nature* **2014**, *506* (7486), 58–62.

(217) Freeman, M. F.; Helf, M. J.; Bhushan, A.; Morinaka, B. I.; Piel, J. Seven Enzymes Create Extraordinary Molecular Complexity in an Uncultivated Bacterium. *Nat. Chem.* **2017**, *9* (4), 387–395.

(218) Hayata, A.; Itoh, H.; Inoue, M. Solid-Phase Total Synthesis and Dual Mechanism of Action of the Channel-Forming 48-Mer Peptide Polytheonamide B. *J. Am. Chem. Soc.* **2018**, *140* (33), 10602–10611.

(219) Bury, P. D. S.; Huang, F.; Li, S.; Sun, Y.; Leadlay, P. F.; Dias, M. V. B. Structural Basis of the Selectivity of GenN, an Aminoglycoside N-Methyltransferase Involved in Gentamicin Biosynthesis. *ACS Chem. Biol.* **2017**, *12* (11), 2779–2787.

(220) Schubert, H. L.; Phillips, J. D.; Hill, C. P. Structures along the Catalytic Pathway of PrmC/HemK, an N5-Glutamine AdoMet-Dependent Methyltransferase. *Biochemistry* **2003**, *42* (19), 5592–5599.

(221) Graille, M.; Heurgué-Hamard, V.; Champ, S.; Mora, L.; Scrima, N.; Ulryck, N.; Van Tilbeurgh, H.; Buckingham, R. H. Molecular Basis for Bacterial Class I Release Factor Methylation by PrmC. *Mol. Cell* **2005**, *20* (6), 917–927.

(222) Garg, N.; Salazar-Ocampo, L. M. A.; van der Donk, W. A. In Vitro Activity of the Nisin Dehydratase NisB. *Proc. Natl. Acad. Sci.* **2013**, *110* (18), 7258–7263.

(223) Oman, T. J.; Van Der Donk, W. A. Insights into the Mode of Action of the Two-Peptide Lantibiotic Haloduracin. *ACS Chem. Biol.* **2009**, *4* (10), 865–874.

(224) Walter, T. S.; Meier, C.; Assenberg, R.; Au, K. F.; Ren, J.; Verma, A.; Nettleship, J. E. E.; Owens, R. J.; Stuart, D. I. I.; Grimes, J. M. Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. *Structure* **2006**, *14* (11), 1617–1622.

(225) Pei, J.; Kim, B. H.; Grishin, N. V. PROMALS3D: A Tool for Multiple Protein Sequence and Structure Alignments. *Nucleic Acids Res.* **2008**, *36* (7), 2295–2300.

(226) Mohr, K. I.; Volz, C.; Jansen, R.; Wray, V.; Hoffmann, J.; Bernecker, S.; Wink, J.; Gerth, K.; Stadler, M.; Müller, R. Pinensins: The First Antifungal Lantibiotics. *Angew. Chemie - Int. Ed.* **2015**, *54* (38), 11254–11258.

(227) Onaka, H.; Nakaho, M.; Hayashi, K.; Igarashi, Y.; Furumai, T. Cloning and Characterization of the Goadsporin Biosynthetic Gene Cluster from Streptomyces Sp. TP-A0584. *Microbiology* **2005**, *151* (12), 3923–3933.

(228) Ozaki, T.; Yamashita, K.; Goto, Y.; Shimomura, M.; Hayashi, S.; Asamizu, S.; Sugai, Y.; Ikeda, H.; Suga, H.; Onaka, H. Dissection of Goadsporin Biosynthesis by in Vitro Reconstitution Leading to Designer Analogues Expressed in Vivo. *Nat. Commun.* **2017**, *8*, 1–13.