STRUCTURAL STUDIES OF AMIDE-BOND FORMING DOMAINS IN NONRIBOSOMAL PEPTIDE SYNTHETASES

Kristjan Bloudoff
Department of Biochemistry
McGill University
Montreal, Quebec, Canada
February 2017

A thesis submitted to McGill university in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

“If I have seen further, it is by standing on the shoulders of giants”
— Sir Isaac Newton

Abstract

Nonribosomal peptide synthetases (NRPSs) are a family of multimodular enzymes that synthesize structurally and functionally diverse peptides, many of which are interesting secondary metabolites. NRPSs are organized into modules of ~1100 residues, with each module responsible for adding one specific amino acid to the peptide product. In general, a module consists of three domains that are essential to the elongation of a peptide: the adenylation (A) domain, the peptidyl carrier protein (PCP) domain and the condensation (C) domain. The A domain will activate a monomer, which is then carried by the PCP domain to the C domain, where amide bond formation occurs between the monomer and the growing peptide chain. In some NRPSs, a C domain can be replaced by a heterocyclization (Cy) domain, which performs amide bond formation, as well as an intramolecular cyclization of serine, cysteine, or threonine sidechains. In either case, specificity determinants and reaction mechanisms are not well understood.

In addition to crystallizing the first condensation domain of the viomycin synthetase, I solved the crystal structure of the first condensation domain of calcium-dependent antibiotic synthetase (CDA-C1) in two different space groups, to resolutions of 1.7 and 2.4 Å. The conformations adopted by CDA-C1 are quite similar in these two structures, yet distinct from those seen in other NRPS C domain structures. Activity assays verified that this protein is active, and SAXS analysis suggested that the conformation observed in these crystal structures could faithfully represent the conformation in solution. Using computational analyses, the implications of these conformational changes in the overall synthetic cycle are presented, as well as the observation that the “latch” that covers the active site is consistently formed in all studied C domains.

In order to study the condensation reaction specifically, chemical probes that covalently bind to an engineered cysteine residue located near the active site of CDA-C1 were designed, mimicking native acceptor substrate delivery to the site by carrier domains. Using mass spectrometry, I verified that the chemical probes were specific for the engineered cysteine and that they were competent for reaction only when the active site histidine is present. I determined the crystal structure of CDA-C1 in complex with these chemical probes, leading to the insight that the active site histidine acts to position the alpha-amino group of the acceptor substrate during catalysis. As well, I used the crystal structure to identify a mutation that is able to alter the domain’s specificity for the acceptor substrate.

Finally, I have solved the crystal structure of an NRPS Cy domain to resolution of 2.3 Å. Despite sharing the same fold, the active sites of C and Cy domains have important differences. To probe its reaction mechanism, I expressed and purified the entire NRPS synthetase containing this Cy domain (bacillamide synthetase from the thermophilic *Thermoactinomyces vulgaris*). Mutations were introduced to the Cy domain, and their effects on peptide biosynthesis was assayed. The drastically different results, interpreted using the structural and bioinformatics results, provided insight into the catalytic mechanisms of the Cy domain, and implicated a previously unexamined Asp-Thr dyad in catalysis of the cyclodehydration reaction.

Resumé

Les enzymes de synthèse de peptides non ribosomiques (aussi appelés NRPS, de l'anglais « nonribosomal peptide synthetase ») sont un groupe d'enzymes multi-modulaires qui synthétisent de nombreux peptides très diversifiés, tant structurellement que fonctionnellement, dont certains s'avèrent être d'intéressants métabolites secondaires. Les NRPS s'organisent comme une chaîne de modules d'environs 1100 résidus, où chaque module a pour rôle d'incorporer spécifiquement un acide aminé dans le peptide non-ribosomique ainsi produit. Dans la plupart des cas, un module de NRPS est composé de 3 domaines essentiels à l'élongation du peptide non-ribosomique : le domaine d'adénylation (ou domaine A), le domaine porteur de peptide (ou domaine PCP de l'anglais « peptidyl carrier protein ») et le domaine de condensation (ou domaine C). L'acide aminé est d'abord sélectionné par le domaine A qui l'active par adénylation, puis chargé sur le domaine PCP qui l'apporte au domaine C, où la formation du lien peptidique est catalysée, incorporant ainsi un nouvel acide aminé dans la chaîne naissante. Chez certains NRPS, le domaine C est remplacé par un domaine d'hétéro-cyclisation (ou domaine Cy), qui catalyse la formation d'un hétérocycle par la réaction intramoléculaire d'une sérine, d'une cystéine ou d'une thréonine, en plus de la formation du lien peptidique. Dans les deux cas décrits ci-dessus, le mécanisme d'action et de spécificité du substrat de ces modules ne sont toujours pas entièrement expliqués à ce jour.

Outre la cristallisation du premier domaine C de l'enzyme de synthèse de la viomicine, nous présentons ici la structure déterminée par cristallographie du premier domaine C de l'enzyme de synthèse d'antibiotiques dépendante au calcium (CDA-C1) dans deux groupes d'espaces différents, à une résolution de 1,7 Å et 2,4 Å. Les deux conformations adoptées par CDA-C1 dans ces deux structures sont similaires entre elles, mais différentes de celles observées chez d'autre domaines C. L'activité catalytique de l'enzyme est vérifiée par essai enzymatique, et les analyses par diffusion des rayons X aux petits angles (SAXS) suggèrent que la conformation adoptée par la protéine au sein de la structure cristalline est fidèle à celle observée en solution. L'implication de ces changements de conformation dans le cycle catalytique est présentée ici grâce à un model informatique, ainsi que l'observation d'une « boucle de sécurité » qui couvre le site actif de manière consistante chez tous les domaines C étudiés.

Dans le but de caractériser plus en profondeur la réaction de condensation, des sondes chimiques furent conçues pour se lier de façon covalente a une cystéine introduite par mutation contrôlée à proximité du site actif du domaine CDA-C1, imitant ainsi la livraison du substrat par le domaine PCP. Nous vérifions par spectrométrie de masse la spécificité des sondes chimiques pour cette cystéine, ainsi que la dépendance de leur activité à une histidine distincte au sein du site actif. Nous présentons ici la structure déterminée par cristallographie du domaine CDA-C1 en complexe avec ces sondes chimiques, menant à la conclusion que cette histidine spécifique a pour rôle de positionner le groupe fonctionnel amine en alpha du résidu au niveau du site accepteur de substrat du domaine CDA-C1 pour permettre la condensation. D'autre part, cette structure nous a permis d'identifier un site de mutation contrôlant la spécificité du domaine pour son substrat.

Enfin, nous présentons ici la structure déterminée par cristallographie d'un domaine Cy à une résolution de 2,3 Å. En dépit d'une grande similarité de structure tridimensionnelle, les domaines C et Cy présentent d'importantes variations au niveau de leur sites actifs. Afin d'examiner son mécanisme catalytique, nous avons exprimé et purifié l'ensemble du NRPS contenant ce domaine Cy, à savoir l'enzyme de synthèse de la bacillamide chez la bactérie thermophile *Thermoactinomyces vulgaris*. Les mutations apportées au domaines Cy ainsi que leur effet sur la biosynthèse du peptide non-ribosomique sont évalués par HPLC et par spectrométrie de masse. Ces résultats drastiquement différents, ainsi que les informations apportées par analyse structurelle et bio-informatique, permettent une meilleure appréciation globale du mécanisme catalytique de condensation effectuée par les domaines C et Cy, et décrivent aussi l'implication précédemment inconnue du dyade Threonine-Aspartate dans la réaction de d'hétéro-cyclisation.

Table of Contents

Table of Figures

Table of Tables

Preface

This is a manuscript-based thesis, which consists of four published research articles. In addition, the Introduction (**Chapter 1**) and General Conclusions (**Chapter 6**) will form the basis of a review submitted to *Biochimica et Biophysica Acta*: *Proteins and Proteomics*.

Chapter 2

Bloudoff K and Schmeing TM (2013). Crystallization and preliminary crystallographic analysis of the first condensation domain of viomycin synthetase. *Acta Crystallographica Section F* 69(4): 412–415.

Chapter 3

Bloudoff K, Rodionov D, and Schmeing TM (2013). Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. *Journal of Molecular Biology* 425(17): 3137–3150.

Chapter 4

Bloudoff K, Alonzo DA, and Schmeing TM (2016). Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases. *Cell Chemical Biology* 23(3):331-339.

Chapter 5

Bloudoff K, Fage CD, Marahiel MA, and Schmeing TM (2017). Structural and mutational analyses of the nonribosomal peptide synthetase heterocyclization domain provide insight into catalysis. *Proceedings of the National Academy of Sciences* 114(1): 95-100.

Contributions of Authors

Chapter 2:
I performed all experiments performed in the manuscript. T. Martin Schmeing and I designed the experiments and wrote the manuscript.

Chapter 3:
I performed all experiments performed in the manuscript with the exception of the Small Angle X-ray Scatter (SAXS), which was performed by Dmitry Rodionov and computation of conformational changes, which were performed by T. Martin Scheming. T. Martin Schmeing and I designed the experiments and wrote the manuscript.

Chapter 4:
I performed all experiments performed in the manuscript, with the exception of Diego A. Alonzo contributing significantly to the X-ray data processing (published and unpublished). T. Martin Schmeing and I designed the experiments and wrote the manuscript.

Chapter 5:
I performed all experiments performed in the manuscript, with the exception of T. Martin Schmeing performing the bioinformatics analysis. T. Martin Schmeing and I designed the experiments and wrote the manuscript. Christopher D. Fage and Mohamad A. Marahiel examined the data and provided input for the manuscript.

Original Contributions to Knowledge

Chapter 2: Crystallization and preliminary crystallographic analysis of the first condensation domain of viomycin synthetase.

- I have crystallized and collected a 2.9 Å of the nonribosomal peptide synthetase condensation domain from the viomycin synthetase

Chapter 3: Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases.

- I have solved the crystal structure of a nonribosomal peptide synthetase starter condensation domain from the calcium-dependent antibiotic synthetase, CDA-C1. The crystal structure was observed to have a significantly more “closed” conformation when compared to previously-solved condensation domains
- A HPLC/mass spectrometry-based assay confirmed that the protein was active, and SAXS analysis was consistent with the protein having a “closed” conformation in solution
- Using several computational analyses, the transition between “open” and “closed” conformations was found to be reasonable, putting forward the proposal that the conformation of condensation domains are dynamic during the synthetic cycle

Chapter 4: Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases.

- Using the solved crystal structure of CDA-C1, acceptor substrate analogues were designed to alkylate an engineered cysteine residue, mimicking delivery of the substrate to the active site.
- I definitively showed this analogue specifically alkylates the engineered cysteine, and that the analogue is able to participate in the condensation reaction. I was able to further show the length of the substrate analogue linker was optimally four carbons long.

- Using the C domain alkylated with the four carbon substrate analogue, I was able to solve the first condensation domain co-complex crystal structure. The structure and experimental data led to the novel proposal of the catalytic histidine acting to position the acceptor substrate for nucleophilic attack.
- In addition, using the co-complex crystal structure and HPLC/mass spectrometry-based assay, I was able to show the first specificity-altering mutation in condensation domains

Chapter 5: Structural and mutational analyses of the nonribosomal peptide synthetase heterocyclization domain provide insight into catalysis.

- I have solved the crystal structure of a nonribosomal peptide synthetase heterocyclization domain from the bacillamide synthetase, BmdB-Cy2
- I purified the intact three module nonribosomal peptide synthetase containing BmdB-Cy2 and developed an assay to measure bacillamide formation. Coupling the structure, assay, and bioinformatics analysis, I was able to show mutations that drastically affected condensation and cyclodehydration.
- This led to insights about the reaction mechanism of the Cy domain, and the proposal of an unexamined Asp-Thr dyad in catalysis of the cyclodehydration reaction

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Martin Schmeing, for giving me the opportunity to work on this project, and providing me with the environment and support to succeed in my studies. Being the first member of the Schmeing lab could have been viewed as a risk on both of our parts, but I believe it has resulted in a great success.

A large number of people have played important roles over the course of my PhD. First and foremost are members of the Schmeing Laboratory for their friendship and support over the years. In addition, I would like to highlight Janice Reimer for her substrate synthesis and mass spectrometry advice, Michael Tarry for valuable experimental advice and day-to-day administration of the laboratory, Diego Alonzo for his great support and effort in procession X-ray data (presented in Chapter 4) while I away, Asfarul Haque for performing ITC experiments, Camille Fortinez for adeptly continuing my project as I transition from the laboratory, and Clarisse Chiche-Lapierre for translating my thesis abstract to French. I would also like to thank Nancy Schmeing for proofreading my thesis introduction and translating a German thesis for me in order to gain a greater understanding about their experiments, in addition to the other behind-the-scenes work I know she has provided for the laboratory.

Several people outside of the laboratory have also been key to my success. I am grateful to Kurt Dejgaard, Alexander Wahba, and Mark Hancock for their mass spectrometry expertise and helping facilitate the experiments in Chapters 3, 4, and 5. Dmitry Rodionov, in addition to his SAXS contribution in Chapter 3, did an outstanding job in keeping the X-ray home source and various other instruments on the floor functional during his time here.

I would like to thank Drs. Bhushan Nagar and Anthony Mittermaier for serving on my Research Advisory Committee. They provided valuable insights and support throughout my PhD studies.

I received generous funding from McGill University, especially through the McGill University Scholarship in Chemical Biology, NSERC-CREATE Training Programme in Bionanomachines, and Groupe de Recherche Axé sur la Structure des Protéines. In

addition to supporting me, they facilitated travel to conferences in Hawaii (USA), Norwich (UK), and Oulu (Finland). Furthermore, NSERC-CREATE Training Programme in Bionanomachines funded an industrial internship in Cambridge (UK), a valuable learning experience.

I would finally like to thank my family for their continued love and support. Even though my mother, Heidi, has long had to deal with debilitating back pain, she has always remained a bedrock of support and ensured my wellbeing. My sister, Eva, has always been an empathetic presence in my life. Finally, my father, Michael, passed away before he could see me to the end of this process, but I hope that he would be proud.

CHAPTER 1: INTRODUCTION TO NONRIBOSOMAL PEPTIDE SYNTHETASES AND THE NECESSESSITY OF AMIDE BOND FORMATION

1.1 Overview of the nonribosomal peptide synthetase

Nonribosomal peptide synthetases (NRPSs) are a fascinating family of enzymes which assemble acyl substrates into bioactive secondary metabolites (Schwarzer et al., 2003; Walsh, 2004; Weissman, 2015). Their nonribosomal peptide products have important and diverse activities. They include antifungals (bacillomycin), antibacterials (daptomycin), antivirals (luzopeptin), antitumors (actinomycin D), siderophores (enterobactin), and immunosuppressants (cyclosporin) (Felnagle et al., 2008) (Figure 1.1).

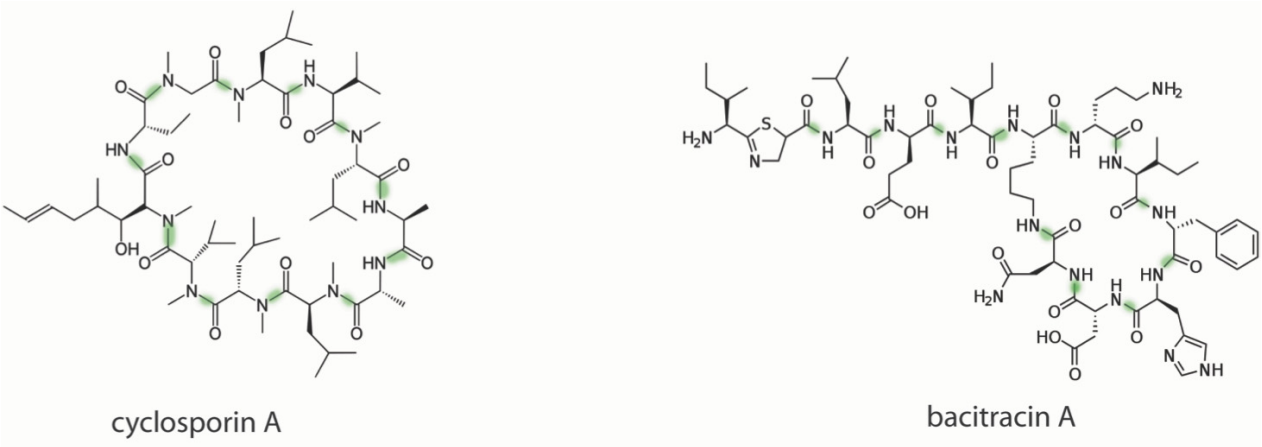


Figure 1.1: Examples of NRPS products. Peptide bonds are highlighted green.

These compounds occupy a huge area of chemical space because NRPSs can use over 500 different acyl monomer substrates (including proteogenic and nonproteogenic amino acids, fatty acids and hydroxy acids), can be co-synthetically or post-synthetically modified, and have linear, cyclic or branched topologies.

NRPSs themselves are large and elegant macromolecular machines which use a modular, assembly line synthetic logic. Evidence for the existence of a peptide biosynthesis pathway distinct from the ribosome was first presented by Mach *et al.* (Mach et al., 1963), who observed that the ribosome-targeting antibiotics

chloramphenicol and puromycin effectively inhibited protein synthesis but had no effect on the formation of tyrocidine. This study was quickly followed by several others (Berg et al., 1965; Eikhom et al., 1963; Fujikawa et al., 1966; Tomino and Kurahashi, 1964; Yukioka et al., 1965). In sum, these studies confirmed the results of Mach *et al*. (Mach et al., 1963) and extended similar observations to the biosynthesis of gramicidin S. Additionally, it was shown that this process was dependent on ATP and that there was a directionality to the order of amino acid incorporation (Berg et al., 1965).

The vast majority of NRPSs are found in bacteria and fungi. With respect to bacteria, NRPSs are most often found in the phyla Actinobacteria, Firmicutes, α-/β-/γ-/δ-Proteobacteria and Cyanobacteria, while fungal-based NRPSs are generally found in the phylum Ascomycota (Sussmuth and Mainz, 2017). However, a few NRPSs have been identified in higher-order organisms. The first example of such a protein was Ebony, which is a single module NRPS found in *Drosophilia melanogaster*. This NRPS synthesises β-alanyl-dopamine, a compound believed to be involved in histamine neurotransmitter metabolism at the photoreceptor synapse in the eye of *Drosophilia melanogaster* (Richardt et al., 2003). Another example is the nemamide synthetase, a hybrid polyketide synthase-NRPS found in *Caenorhabditis elegans*, whose products have been speculated to play a signalling role in development (Shou et al., 2016).

NRPSs are organized into modules of ~1100 residues with each module responsible for the addition of one specific amino acid to the peptide product (Walsh, 2008). The number and order of modules typically correspond directly to the number and order of amino acids in the peptide product, though there are multiple known cases of iteration, module skipping and intermediate oligomerization. NRPSs can be as small as one module or as large as eighteen modules, with molecular weights >2 MDa in a single protein chain. Typically, an elongation module consists of the three domains essential to the elongation of a peptide: the adenylation (A) domain, the peptidyl carrier protein (PCP) domain (alternatively referred to as the thiolation (T) domain) and the condensation (C) domain. Terminal modules also include domains dedicated to releasing the final peptide, such as a thioesterase (Te) or reductase domain. A typical organization of an NRPS is A-PCP-$(C\text{-}A\text{-}PCP)_x$-Te, with additional tailoring domains inserted into some of the modules.

In each round of the peptide synthetic cycle (Figure 1.2), the A domain binds specifically to its cognate monomer and first activates it by adenylation, then transfers it onto the 4'-phosphopantetheinyl (PPE) arm of the PCP domain. The PCP domains of the current and previous modules transport their covalently bound substrates into the active site of the C domain. The C domain catalyzes the formation of a peptide bond between the two substrates, which both elongates the growing peptide chain and passes it downstream. After this condensation reaction, the peptide (now attached to the PPE arm of the PCP domain of the current module) is carried to the C domain of the next module to be donated in that module's condensation reaction. This both frees the PCP domain to participate in the round of catalysis within the module and further elongates the peptide chain. The peptide is likewise elongated in each of the subsequent modules, which act like stations in an assembly line, until it reaches the termination module, where after a final elongation, the Te or R domain releases the peptide product.

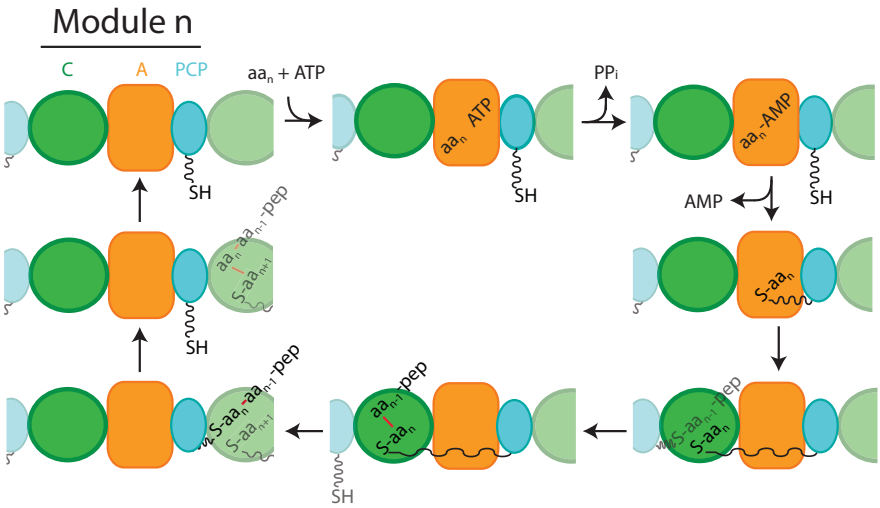


Figure 1.2: Schematic representation of the NRPS peptide elongation cycle. The elongation cycle begins with the A domain (orange) binding an amino acid and ATP, followed by amino acid adenylation/activation. The A domain then catalyzes the transfer of the activated amino acid to the PPE arm covalently linked to the PCP domain (blue). The PCP-linked amino acids of the current module and the previous module are delivered to the current module's C domain (green), where peptide bond formation occurs and the previous module's PCP-PPE is released. The elongated peptide is then delivered to the next module's C domain, where the cycle repeats.

Significant effort has been put into elucidating how these NRPS domains function individually, as well as how they function within the context of the larger NRPS. Several insightful reviews of the field have been published recently, including those by Süssmuth and Mainz (Sussmuth and Mainz, 2017), Miller *et al.* (Miller and Gulick, 2016), Payne *et al.* (Payne et al., 2016) and Weissman (Weissman, 2015). In addition to the motivation of fundamentally understanding the complicated and elegant way NRPSs function, much interest in NRPSs has been generated because, despite complex molecular mechanisms, their overall synthetic logic is quite simple, which raises the tantalizing possibility of bioengineering: Adding, removing and substituting substrate binding site residues, individual domains and entire modules should all have predictable effects on the peptide products, and allow production of myriads of novel bioactive compounds. This can work: there have been successes in bioengineering, including excellent work with the daptomycin pathway (Nguyen et al., 2006). However, bioengineering of NRPSs often fails, and when successful is typically accompanied by a marked decrease in peptide yield (Calcott and Ackerley, 2015; Calcott et al., 2014; Kries et al., 2015), underlining the fact that our understanding of NRPSs is not yet complete.

1.2 Other common NRPS domains

My thesis focuses on the condensation domain and related condensation domain superfamily members that have acquired specialized function. However, as C domains work in concert with the other NRPS domains, a brief discussion of the structure and function of those domains is presented here to provide context:

1.2.1 The adenylation domain

As described above, the A domain is responsible for selecting the cognate substrate and ligating it to the PCP domain to make it into an intermediate competent to participate in reactions catalyzed by other NRPS domains. Thus the A domain is the main selectivity determinant / gatekeeper in NRPSs (Mootz and Marahiel, 1997; Stachelhaus and Marahiel, 1995).

The A domain is also one of the best-characterized NRPS domains. The first crystal structure of any NRPS domain was of the phenylalanine-activating A domain from gramicidin S synthetase, determined by Elena Conti and colleagues in 1997 (Conti et al., 1997). Based on this structure and biochemical and phylogenic studies, Stachelhaus *et al.* (Stachelhaus et al., 1999) determined that a group of ~10 amino acids in the amino acid binding pocket dictates substrate specificity and deciphered the "specificity determining code", which allowed the identity of these 10 amino acids to be used to reliably predict which monomer substrate was cognate for any A domain. This code and its refinements (Challis et al., 2000; Stachelhaus et al., 1999) are incredibly useful, as they allow a reasonable guess as to what product a completely uncharacterized NRPS will synthesize, based solely on gene sequence information.

The A domain consists of two portions, the larger A_{core} (~450 amino acids) and smaller A_{sub} (~100 amino acids) subdomain. The A_{sub} subdomain is highly mobile between the various functional states in the two-reaction A domain cycle. The cycle was elucidated by excellent structural work from several groups (Conti et al., 1997; Du et al., 2008; Gulick et al., 2003; May et al., 2002) and first assembled by Yonus *et al.* (Yonus et al., 2008): an "open" conformation of the A_{sub} allows amino acid, ATP and Mg^{2+} binding, after which the A_{sub} closes with a ~30° rotation to donate residues for the adenylation reaction, producing the activated aminoacyl-adenylate. Next, the A_{sub} rotates ~140° and allows the PCP domain to bind. The A domain catalyzes transfer of an acyl group from the adenylate onto the terminal thiol of PPE arm of the PCP domain (Mitchell et al., 2012; Reger et al., 2008). Attached to the PCP, the acyl monomer is now in a form competent for incorporation into the nonribosomal peptide.

1.2.2 The peptidyl carrier protein domain

The PCP domain is responsible for the transportation and presentation of the aminoacyl and peptidyl substrates to different domains of the NRPS. PCP domains are small, four helix bundle domains of ~80 amino acids, homologous to acyl carrier protein domains of fatty acid synthases (FAS) and polyketide synthases (PKSs) (Weissman, 2015). For the NRPS to be functional, PCP domains must be post-translationally modified by a PPE-transferase, which attaches the PPE moiety from coenzyme A onto

a serine side chain of the conserved GGXS motif located at one end of the four helix bundle (Lambalot et al., 1996).

PCP domains are unique among common NRPS domains, as they do not possess their own catalytic activity, but rather fulfill a transportation role. (Their other name, “thiolation domains” is thus somewhat misleading, as PCP domains do not catalyze thiolation themselves, but rather provide the terminal PPE thiol as a substrate in the thiolation reaction, which is catalyzed by the A domain.) The PCP domain presents substrates to every other NRPS domain, and so during a catalytic cycle must engage in transient but productive protein-protein interactions with all domains of its module (including A, C, downstream C, tailoring, Te domains), despite limited surface area to do so (Tanovic et al., 2008).

1.2.3 The thioesterase domain

Te domains, found in the final module of many NRPSs, catalyze the release of the mature peptide from the NRPS (Schneider and Marahiel, 1998). NRPS Te domains are ~35 kDa domains, homologous to Te domains in FAS and PKSs, with an α/β hydrolase protein fold and an active site with a catalytic triad of serine, histidine, and aspartic acid, topped by a variable “lid” region (Bruner et al., 2002). Te domains catalyze a two half-step reaction: First, they catalyze transfer of the peptidyl group from the PCP domain onto their active site serine to form an O-acyl intermediate. Te domains then catalyze one of three types of second half-reactions that occur, which lead to 3 kinds of nonribosomal peptide products (Tseng et al., 2002): (1) attack of a water molecule, leading to hydrolysis and release of a linear peptide; (2) attack of a nucleophilic group from within the peptidyl moiety attached to the Te domain (such as the N-terminal amino group or a threonine side chain) leading to cyclization and release of a cyclic peptide; or (3) attack of a nucleophilic group from within a newly-synthesized peptidyl moiety attached to the PCP domain, which leads to oligomerization of the peptide, producing an oligomeric substrate for further processing and ultimate release by the Te domain (Hoyer et al., 2007). Individual Te domains are often very selective for the kind of reaction they perform (hydrolysis, cyclization, or oligomerization) and are also regio- and stereoselective (Trauger et al., 2000).

1.2.4 Tailoring domains

In addition to C, A, PCP, and Te domains, NRPSs can include several types of domain which co-synthetically modify the nonribosomal peptide. Those related to C domains are discussed later in this review. Other common NRPS tailoring domains include formylation (F) (Schoenafinger et al., 2006), methyltransferase (Lawen and Zocher, 1990), oxidase (Schneider et al., 2003), reductase (Reimmann et al., 2001), and halogenase domains (Dorrestein et al., 2005). Typically, these domains are inserted into the protein sequence adjacent to the canonical domains or within variable loops of a canonical domain, and act in cis during synthesis (Labby et al., 2015). There are also free-standing versions of these domains which act co-synthetically in trans in some NRPS systems (Miller et al., 2002). Tailoring domains generally look and function like their non-NRPS homologues, but they have evolved to work in the NRPS assembly line. For example, the fold of the first 171 amino acids of the F domain of LgrA is conserved with sugar formyltransferases, but the F domain also includes a new structural element, a link to the A domain, a hydrophobic patch for PCP domain binding and residues for interaction with the PPE phosphate (Reimer et al., 2016).

Tailoring domains likely arose from gene fusion events of single domain genes and existing NRPS genes, with advantageous fusions persisting and being fine-tuned for function within NRPSs. Analyses of genomic sequencing data show hundreds of other kinds of domain inserted into NRPS genes, but these can be assigned as true NRPS tailoring domains only if their tailoring activities have been established.

1.3 Introduction to the Condensation Domain

1.3.1 Discovery of the Condensation Domain

The condensation domain was first identified in 1995, when de Crécy-Lagard saw a conserved HHxxxDG motif in all of the 55 NRPS sequences examined (De Crecy-Lagard et al., 1995). She noticed that this motif is present in an NPRS the same number of times a condensation or epimerization reaction occurs during biosynthesis by that NRPS, and that this same motif is also found at the active sites of chloramphenicol acetyltransferase (CAT) and dihydrolipoamide acetyltransferase (E2p) (Guest, 1987).

Informed by the structural studies of CAT by Leslie (Leslie, 1990), and of E2p by Mattevi *et al.* (Mattevi et al., 1992), the second histidine of the motif was proposed to deprotonate the α-amino group of the acceptor amino acyl-PCP to catalyze nucleophilic attack on the carbonyl carbon of the donor peptidyl-PCP_{+1} (from the upstream module) in C domains.

The first direct biochemical evidence that the C domain catalyzes peptide bond formation (Figure 1.3) came in 1998 (Stachelhaus et al., 1998). Stachelhaus mixed a phenylalanine-activating module of A-PCP-E from GrsA, a proline-activating module of C-A-PCP from TycB, phenylalanine, proline and ATP, and detected a linear peptide of D-Phe-L-Pro (after hydrolysis from the PCP). The same experiment using protein harboring a mutation of the putative catalytic histidine to valine abolished peptide bond formation.

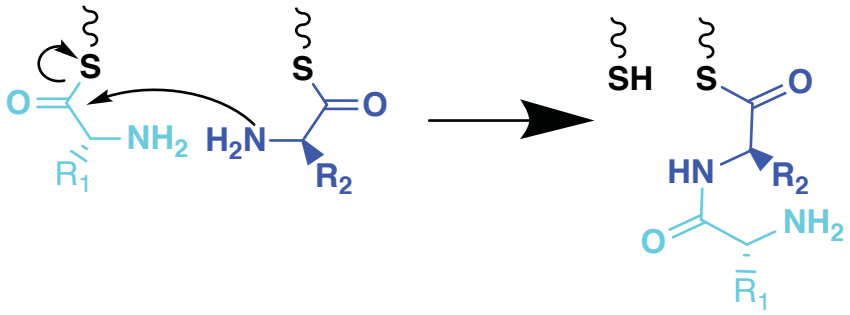


Figure 1.3 Schematic of the condensation reaction.

These experiments concretely established the C domain as the NRPS peptidyl transferase domain, highlighted the importance of the second histidine of HHxxxDG motif for reaction, and reinforced the idea that this histidine could be acting as a general base.

1.3.2 Mutational Insight into C Domains

Classical mutational analyses to investigate the importance of additional C domain residues soon followed. To facilitate the discussion of these mutational analyses, which were performed using many different C domains from many different NRPS proteins, we will provide both the residue number from the interrogated protein,

as it appears in the original publication, and the equivalent residue number for the first condensation domain of the calcium dependent antibiotic (CDA-C1) which we have used for our own structural and mutagenic work, in parentheses. The putative catalytic histidine from tyrocidine synthetase protein B (TycB) is thus denoted “H147 (*cda H157*)”.

First, Vater *et al.* interrogated the aspartate of the HHxxxDG motif in a C domain of surfactin synthetase (Vater et al., 1997). Mutation of D148 (*cda D161*) to alanine also completely abolished condensation activity, though no specific role for this aspartate was proposed. A more comprehensive study came just prior to the release of any C domain structure (Bergendahl et al., 2002): Bergendahl *et al*. aligned the sequence of 80 C domains and selected 10 residues that appeared very highly or completely conserved for mutational analyses. Three mutations, R67A (*cda C63*), H146A (*cda H156*), and W202L (*cda W216*), led to insoluble or misfolded proteins, and the residues were assigned to be structurally important. The Q19A (*cda Q15*) mutation had no effect on catalytic activity, while mutations C154A (*cda S164*), R284A (*cda R296*), and Y166W (*cda Y176*) decreased (but did not abolish) catalytic turnover of the enzyme. However, mutations H147A (*cda H157*), D151N (*cda D161*), and R62A (*cda G58*) completely abolished condensation activity. This reiterates the importance of the second histidine and the aspartate of the HHxxxDG motif and adds R62 as vital, at least for the C domain of TycB1. The refined catalytic mechanism proposal accompanying this study continued the role of H147 (*cda H157*) as a general base and suggested that no other residue would participate directly in the reaction, though Y166 (*cda Y176*) might play a favorable but non-essential role in transition state stabilization. Roles for D151 (*cda D161*) and R62 (*cda G58*) as forming a structure-stabilizing salt bridge were correctly assigned by analogy to corresponding residues in CAT (Leslie, 1990).

The determination of the first structure of a C domain (see below) naturally informed the mutational experiments that came thereafter and their interpretation (Di Lorenzo et al., 2004; Gatto et al., 2005; Samel et al., 2007), but it is nonetheless worth summarizing some of the key mutational data in the current section. The study of Roche & Walsh (Roche and Walsh, 2003) was particularly interesting because it combined *in vitro* and *in vivo* data. First, a multi-turnover assay with the excised C domain of EntF

and L-serinyl-SNAC as the acceptor substrate was used to interrogate the contributions of residues they suspected to be involved in catalysis, H138A/K/E (*cda H157*) and N350A/D (*cda N364*), or in maintaining the structural integrity of the active site, I14A (*cda V18*), D142A (*cda D161*), G143L (*cda G162*), and R278M/K (*cda R296*). Of all mutants assayed, only R278K (*cda R296*) and I14A (*cda V18*) retained observable (albeit diminished) condensation activity, despite that all but the D142A (*cda D161*) appeared to be well folded. Mutations were then introduced into full length EntF, and their capacity to participate in production of the iron siderophore enterobactin was measured by ability to grow in a low iron environment. The deleterious effects seen in the *in vitro* assay are tempered *in vivo*. For example, I14A (*cda V18*), which had greatly diminished *in vitro* activity, imparted only a minor growth defect, and N350D (*cda N364*), which had no observable *in vitro* activity, only slowed growth by ~10 fold. This mitigation can be explained by condensation activity not being the limiting factor in enterobactin synthesis *in vivo*. However, it is notable that all EntF C domain mutations support growth to some extent, including H138A/K (*cda H157*), and that this growth was more robust than when EntF is absent or with the most deleterious Te domain mutants.

Finally, in the study including structure determination of VibH, Keating *et al.,* (Keating et al., 2002) included accompanying mutational analysis of 7 residues. The D130A (*cda D161*) mutation rendered VibH inactive, as it had done in SrfB1 (Vater et al., 1997), EntF (Roche and Walsh, 2003) and TycB1 (Bergendahl et al., 2002), and W264K (*cda Y297*) was also deleterious. However, several mutations had little to moderate effects on catalysis, including Y132A (*cda T163*), N335L (*cda N364*), HHxxxDG residues H125L (*cda H156),* G131L (*cda G162),* and remarkably H126A/E (*cda H157*), the putative general base. Although striking, the result of a minor effect of mutation of the putative catalytic base is not an isolated case. In a study published the same year, the C domain from another protein in the same pathway, VibF, suffered only a 9-fold decrease in catalysis upon mutating the histidine to alanine (Marshall et al., 2002b). The demonstration that some C domains are able to retain catalytic activity despite a mutation of the second histidine of the HHxxxDG motif called into great doubt the previously proposed reactions, at least as a general catalytic mechanism for all C domains.

1.4 Structural Characterization of the C Domain

1.4.1 The structure of the C domain

A breakthrough in C domain knowledge came in 2002, when Keating *et al*. published the first study describing a crystal structure of an NRPS C domain (Keating et al., 2002) (Figure 1.4). This structure was of VibH, a stand-alone C domain involved in the synthesis of the siderophore vibriobactin. It provided a framework and structural context for interpretation of previous mutational results and aided in design of new experiments (Di Lorenzo et al., 2004; Gatto et al., 2005; Roche and Walsh, 2003; Samel et al., 2007). Furthermore, it identified putative locations of the binding sites for partner PCP domains and visualized and expanded the recognition of structural similarities to CAT.

The structures show that the C domain can be described as a pseudo-dimer of two lobes, the N-terminal lobe (N-lobe) and the C-terminal lobe (C-lobe) (Figure 1.4). Each lobe consists of one large central beta sheet, with most of the rest of the sequences made up of α-helices. The C-lobe is composed of a core sheet of 6 ß-strands, extended by two additional strands and covered on one side by 9 α-helices. The N-lobe is composed of a core sheet of 5 ß-strands, one of which is part of the "latch" feature (VibH 337-360; *cda 367–388*) donated from sequence from the C-lobe, surrounded by 5 α-helices and a small, peripheral, 2-strand, ß-sheet. The two lobes form head-to-tail pseudo-dimers, described as V-shaped (Samel et al., 2007). In addition to the latch, there is a second, smaller, cross-over region, the "floor loop" (VibH 262-276; *cda 295–309*) (Samel et al., 2007), which stretches from the C-lobe to nestle an α-helix against the N-lobe. The active site motif, HHxxxDG, is in a loop connecting the central strand of the N-lobe and one of its long α-helices. It sits at the center of a ~30 Å tunnel made by the interface of the two lobes, partially floored by the floor loop and roofed by the latch (Keating et al., 2002; Samel et al., 2007). The tunnel reaches the surface at two points along the N-lobe / C-lobe interface which are on completely opposite sides of the domain.

1.4.2 PCP binding sites on the C domain

The tunnel observed in the structure of VibH and subsequent excised C domains clearly indicated putative binding sites for the donor peptide-PPE-PCP and acceptor aminoacyl-PPE-PCP (Keating et al., 2002) on the front and back face of the domain (Figure 1.4). The distance from each of the surfaces to the second histidine of the active site motif (~15 Å) fits nicely with the length of the PPE arm. Although the path of the donor PPE could be inferred from superimposition of co-enzyme A bound to CAT (Keating et al., 2002; Leslie, 1990) and confirmed by mutagenesis (Keating et al., 2002; Lai et al., 2006a, b), direct visualization was long in coming. The structure of the TycC5-6 PCP-C didomain (Samel et al., 2007) included the right domains to visualize donor PCP binding. However, in this structure the domains were not captured in a productive conformation for donation; the PPE attachment point on the PCP domain was 47 Å away from the C domain active site and oriented in the opposite direction. Acceptor PCP domain binding was directly observed initially in the landmark structure of a NRPS termination module with domains C-A-PCP-Te (Tanovic et al., 2008). Here, the PCP domain is bound to the putative acceptor site of the C domain, despite the absence of PPE moiety. The interaction included PCP domain residues M1007 and F1027, known to be required for productive binding to the C domain (Lai et al., 2006b), interacting with F24 (*cda L20*), L28 (*cda L24*), and Y337 (*cda E347*).

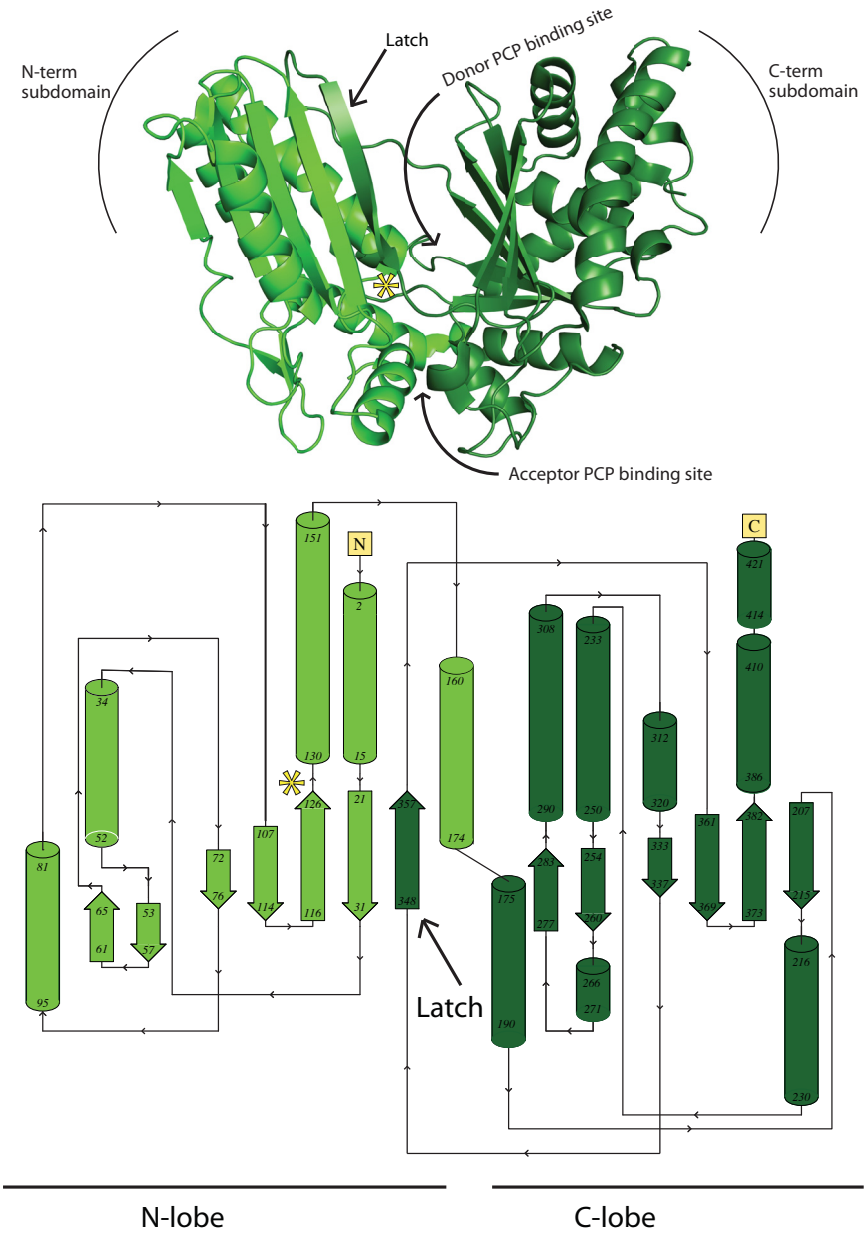


Figure 1.4 Crystal structure of VibH (PDB: 1L5A). N-lobe is light green and C-lobe is dark green. The C-lobe is composed of a core sheet of 4 ß-strands (residues 207-215, 333-337, 361-369, and 373-382) extended by two additional strands (residues 254-260 and 277-283) and covered on one side by 9 α-helices (residues 175-190, 202-206, 216-230, 233-250, 262-276, 290-308, 312-320, 386-410, and 414-421). The N-terminal lobe is composed of a core sheet of 5 ß-strands (residues 21-31, 72-76, 107-114, 116-126, 348-357), one of which (residues 348-357), called the “latch”, is donated from sequence in C-terminal lobe, surrounded by 5 α-helices (residues 2-15, 34-52, 81-95, 130-151, and 160-174) and a small, peripheral, 2-strand, ß-sheet (residues 53-57 and 61-65). The two lobes form head-to-tail pseudodimers, described at “V-shaped” (Samel et al., 2007).

1.4.3 C domain – A domain interactions

The impressive structure of the C-A-PCP-Te termination module from the Marahiel group (Tanovic et al., 2008) also showed the C domain in a more holistic context. In the SrfAC module, a large interface between the C and A_{core} domains buries ~800 $Å^2$ of each domain's surface area. This large interface, including a compact L-shaped linker, was proposed to hold the two domains in a "catalytic platform", rigid throughout the catalytic cycle and perhaps shared by all elongation modules (Figure 1.5). This is in contrast to the A_{sub} domain; while there were interactions found between the A_{sub} domain and a 6 amino acid stretch of the C domain, they must be transient because of the requirement of the A_{sub} domain to adopt different conformations during the adenylation cycle of the A domain.

Although the C-A interaction within this module appeared rigid, it remained to be seen whether (a) the same would be observed in other modules, and (b) if a static interaction between A and C domains would take place between modules. At this point in time, the answer to (a) would have to wait for additional NRPS module structures to be published. However, hints of an answer to (b) existed thanks to the electron microscopic image of native cyclosporin synthetase molecules (Hoppert et al., 2001). Here, negatively stained cyclosporin synthetase samples were observed to generally be grouped in larger, globular complexes and beads-on-a-chain-like complexes. In both groups, the individual modules of the NRPS were visible as condensed, structurally separate units. The conformational heterogeneity between the two groups, and within each group, was a result of the conformational flexibility between modules. Thus, it would not be likely for the A domain of one module and the C domain of the following module would form a rigid "catalytic platform".

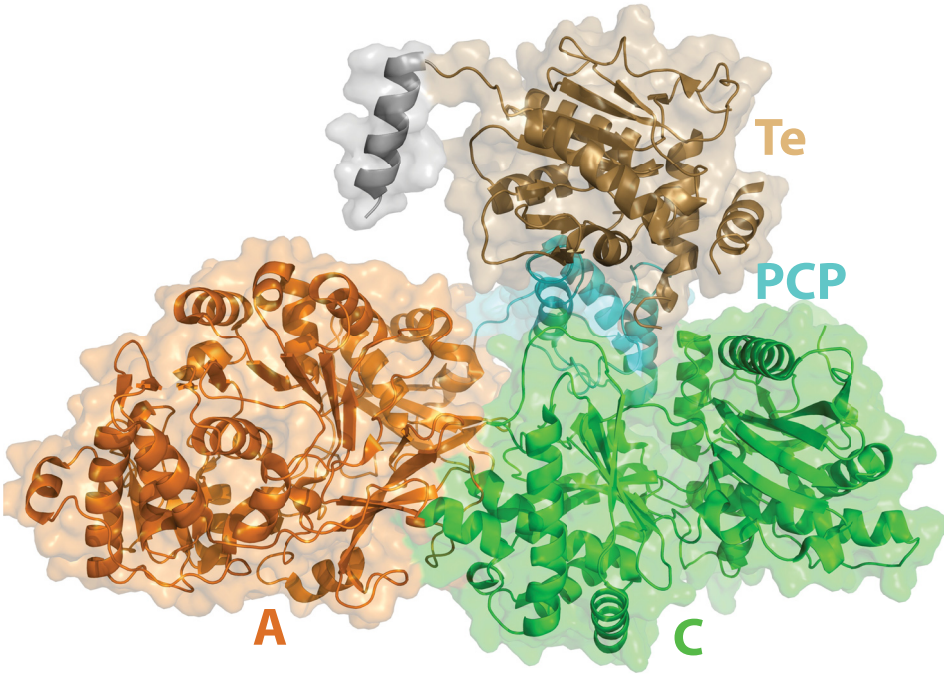


Figure 1.5. Crystal structure of the surfactin synthetase termination module (Tanovic et al., 2008). This module consists of a C domain (green), an A domain (orange), an acceptor PCP domain (blue) and a Te domain (brown), as well as a C-terminal tag helix (grey).

1.5 Reaction Mechanism of the C Domain

As discussed above, the original proposals for the catalytic mechanism of condensation in C domains centered on the proposed action of the second histidine in the HHxxxDG motif as a catalytic base to deprotonate the α-amino group of the acceptor amino acyl-PCP to allow nucleophilic attack on the donor peptidyl-PCP thioester carbonyl (Bergendahl et al., 2002; De Crecy-Lagard et al., 1995). However, mutation of this histidine is not greatly detrimental in several C domains (Keating et al., 2002; Marshall et al., 2002b; Roche and Walsh, 2003).

Samel *et al*. proposed an alternative hypothesis (Samel et al., 2007). The pKa of that histidine in TycC5-6 PCP-C (H224; *cda H157*) was predicted to be 11.8, using that structure and the H++ server (Gordon et al., 2005). This pKa suggested that both ring nitrogens of H224 *(cda H157*) are protonated and the residue is positively charged at physiological conditions, so unable to act as a catalytic base. Samel *et al.* proposed instead that the catalytic histidine, combined with dipole moment of the adjacent α-helix,

act as the oxyanion hole, stabilizing the reaction intermediate. A sulfate from the crystallization buffer was thought to mimic this tetrahedral center of the intermediate. As H224 *(cda H157*) was only partially responsible for this stabilization, its mutation may not always have a catastrophic effect. However, theoretical pKa calculations are not always robust, and the degree to which the sulfate mimics a reaction intermediate is not clear.

A major gap in the above discussion of catalytic mechanisms is the lack of co-complex structures of the C domain with any substrates or substrate analogs competent for reaction.

1.6 C Domain Specificity

Although A domains, which contain the binding sites for monomer substrates, act as the primary specificity determinants in NRPSs (Stachelhaus et al., 1999), C domains also can display specificity. That C domains do not catalyze peptide bond formation between any substrates presented to them was recognized early on: Belshaw *et al.* used chemically synthesized aminoacyl-CoA molecules and *in vitro* loading by the promiscuous PPE transferase Sfp and produced PCP domains bearing a variety of aminoacyl moieties on the PPE (Belshaw et al., 1999). This bypasses substrate selection by the A domain and allows interrogation of the selectivity of the C domain, both the donor and acceptor sites. In the TycB1 C-A-PCP module (Stachelhaus et al., 1998) interrogated, the acceptor site of the C domain is very selective for side chain and chirality, with cognate L-prolinyl-PCP and non-cognate L-alaninyl-PCP able to act as an acceptor, while D-alaninyl, L-leucyl, D/L-phenylalanyl-PCP cannot. The donor site is less selective, as cognate L-prolinyl-PCP and non-cognate L-phenylalanyl-, L-alaninyl-, D-alaninyl- or L-leucinyl-PCP act as donors (Belshaw et al., 1999). Conversely, the acceptor site is very selective for side chain and chirality, with only L-alaninyl-PCP other than cognate L-prolinyl-PCP competent to act as an acceptor in the condensation reaction.

Ehmann *et al.* then similarly produced and used small-molecule NRPS substrate analogues, aminoacyl-N-acetylcysteamine thioesters (aminoacyl-SNACs) (Ehmann et al., 2000). These are like the acyl-SNACs which had been used with polyketide

synthases (Gokhale et al., 1999; Holzbaur et al., 1999) and represent the terminal half of the aminoacyl-PPE. Aminoacyl-SNACs also bypass selection by A domains for direct analysis of C domain specificity. Ehmann also saw selectivity at the acceptor site, both for side chain and chirality, using the C domain of EntF: Aminoacyl-SNACs (alaninyl, theroninyl) similar to cognate serinyl-SNAC could react, though at slower rates, whereas others (leucinyl, prolinyl, and all D- amino acid assayed) were not competent substrates. SNAC molecules have drawbacks, however. They can hydrolyze rapidly (Luo et al., 2002) and they appear to require one PCP-bound substrate for reaction, as no published accounts of their use demonstrate condensation between two aminoacyl-/peptidyl-SNAC molecules (Ehmann et al., 2000; Luo et al., 2002; Roche and Walsh, 2003).

Peptidyl-SNACs in combination with PCP-bound substrates were useful in determining that the TycB3 C domain is selective for L-amino acids at the acceptor site and at the first position of the peptide of the donor substrate (Luo et al., 2002). The biochemical studies and bioinformatics (see below) led to the insight that in general, at the acceptor site C domains can have higher specificity for side chains and prefer L-amino acids (denoted ^{L}C), and at the donor site there can be lower specificity for side chain identity, and C domains can be sorted as to whether they prefer D or L chirality at the first position of the peptide of the donor substrate (denoted $^{L}C_{L}$ or $^{L}C_{D}$) (Clugston et al., 2003; Kallow et al., 1997; Luo et al., 2002; Stindl and Keller, 1994) (exceptions exist, such the requirement for the epoxide group in the donor acyl group of CDA-C1 (Kraas et al., 2012)).

Although these biochemical studies gave good insight into substrate specificity in C domains, understanding of the molecular mechanism underlying this specificity was lacking.

1.7 Bioinformatics Reveal Distinct C Superfamily Domain Subclasses

Following from the sequence gazing which led de Crécy-Lagard to discover the C domain (De Crecy-Lagard et al., 1995) and highlight conserved motifs (including HHxxDG), several other bioinformatics studies have led to insight into the domain. Initial bioinformatic efforts to classify C domains appeared first in 1999, by von Döhren *et al*.

(von Dohren et al., 1999), and were expanded first by Roongsawang *et al*. (Roongsawang et al., 2005) and then comprehensively by Rausch *et al.* (Rausch et al., 2007). The essential finding is that phylogenetic sorting of C domain superfamily sequences groups them according to biological function. This results in subclassification of C domains into $^{L}C_{L}$ (performs condensation between an L-aminoacyl/peptidyl-PCP donor and an L-aminoacyl-PCP acceptor), $^{D}C_{L}$ (performs condensation between D-aminoacyl/peptidyl-PCP donor and an L-aminoacyl-PCP acceptor (Luo et al., 2002)), starter C (performs condensation between a β-hydroxy-carboxylic acyl-ACP donor and an L-aminoacyl-PCP acceptor), E (performs epimerization of the peptidyl-PCP (De Crecy-Lagard et al., 1995)), and dual E/C (first performs epimerisation of the L-aminoacyl-PCP donor to D, and then performs condensation with L-aminoacyl-PCP acceptor (Balibar et al., 2005)), Cy (first performs condensation between aminoacyl/peptidyl-PCP and cysteinyl- threoninyl or serinyl-PCP, and then performs cyclodehydration with the thiol or hydroxyl side chain (Konz et al., 1997)), and X domains (which had no known function at that time).

Some very tangible trends in sequences appeared from the analyses. The conserved motifs were modified from previous versions (De Crecy-Lagard et al., 1995; Marahiel et al., 1997). $^{L}C_{L}$, $^{D}C_{L}$, E, dual E/C, and starter C all have the HHxxxD at motif C3, whereas Cy domains have DxxxxD and X domains, HRxxxDD. The C4 motif in $^{L}C_{L}$ domains differs within $^{D}C_{L}$ domains, and $^{L}C_{L}$ have an additional moderately-conserved LPxDxxRP motif downstream of C4. $^{D}C_{L}$ domains are also found immediately downstream of E domains (De Crecy-Lagard et al., 1995). E domains do not share any core motifs with the C domains besides HHxxxD. Dual E/C domains exhibit a second histidine motif of sequence HH[I/L]xxxxGD, located close to the N-terminus of the domain (Balibar et al., 2005).

Insight was also provided into C domain evolution. Starter C domains were found to share a common ancestor with and $^{L}C_{L}$ domains, as were dual E/C domains with $^{D}C_{L}$ domains. The latter pairing is not surprising, as they both catalyze condensation between a D-aminoacyl/peptidyl-PCP donor and an L-aminoacyl-PCP acceptor, the difference being that in dual E/C domains the epimerization activity is integrated within the domain.

We highlight some of these and other C domain family subtypes below.

1.8 Structures and Mechanisms of Non-Canonical C Superfamily Domain Members

1.8.1 The Cy Domain

An interesting set of modifications found in nonribosomal peptides are the five-membered thiazoline, oxazoline or methyloxazoline rings in the peptide backbone, derived from cyclodehydration of cysteine, serine or threonine residues. These heterocyclic rings are important for the bioactivity (Roy et al., 1999) of peptides such as bacitracin A (Konz et al., 1997), bleomycin (Shen et al., 2002), argyrin (Vollbrecht et al., 2002), yersiniabactin (Gehring et al., 1998) and colibactin (Vizcaino and Crawford, 2015).

Cy domains are responsible for the introduction of these heterocycles. Cy domains were first observed in bacitracin synthetase (Konz et al., 1997). These C superfamily domains take the place of C domains in certain NRPS elongation module (Rausch et al., 2007) and catalyze two separate reactions (Chen et al., 2001; Duerfahrt et al., 2004; Gehring et al., 1998; Keating et al., 2000a; Kelly et al., 2005; Konz et al., 1997; Marshall et al., 2001) (Figure 1.6). First, the Cy domain catalyzes condensation between the aminoacyl/peptidyl-PCP donor and the serinyl/threoninyl/cysteinyl-PCP acceptor. Then it catalyzes a two-step cyclodehydration between the thiol or hydroxyl of the side chain and the carbonyl of the newly-formed amide bond (Duerfahrt et al., 2004; Gehring et al., 1998). The peptidyl-PCP containing the thiazoline or (methyl)oxazoline ring then moves to the next step in assembly line synthesis.

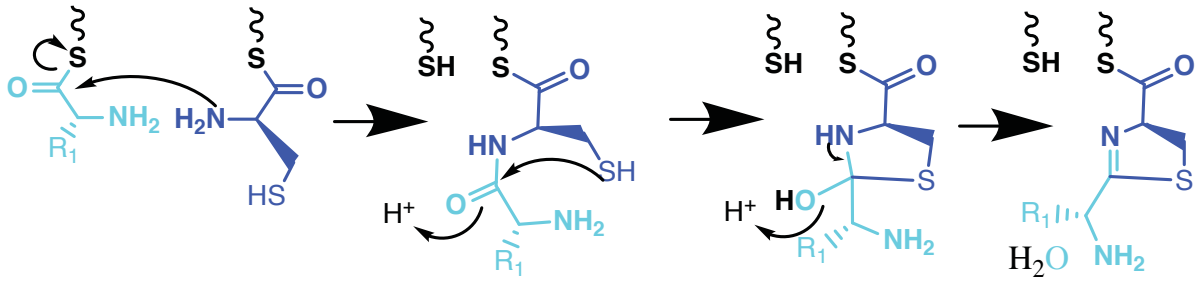


Figure 1.6 Schematic of the heterocyclization reaction.

Depending on the NRPS, this is oxidation or reduction by oxidase or reductase domains embedded into the module or working *in trans* (Patel and Walsh, 2001; Schneider et al., 2003), or peptide donation in the next module.

Strikingly, Cy domains have an extremely conserved DxxxxD motif in the place of the HHxxxDG (Konz et al., 1997). Mutating the aspartate residues diminishes or abolishes catalytic activity (Di Lorenzo et al., 2008; Keating et al., 2000b; Kelly et al., 2005; Marshall et al., 2001), suggesting a key, but undefined, role in catalysis. In addition, a triple mutation that replaces the DxxxxD motif with HHxxxDG also abolishes both condensation and heterocyclization (Keating et al., 2000b), showing Cy domains cannot be transformed into C domains simply by interchanging these active site motifs. Crucially, Duerfahrt *et al.* found that mutation of the N900 or of S984 in the Cy domain of BacA2 eliminated heterocyclization but maintained condensation (Duerfahrt et al., 2004). This clearly indicated that the two reactions are independent and possibly catalyzed by distinct sets of residues. Based on mapping of the Cy domain sequence to the C domain structure, these conserved residues were proposed to be involved in the formation of the solvent channel and not the heterocyclization reaction itself. The questions of how Cy domains perform condensation without the HHxxxDG motif and additionally perform cyclodehydration, remained.

1.8.2 The E Domain

Unlike ribosomal products, nonribosomal peptides often include D-amino acid residues. The presence of D-amino acids allows the peptides to adopt conformations

which are not permitted in all L- peptides, and are critical to their function (Kawai et al., 2004; Veatch, 1976). D-amino acid residues also confer resistance to degradation by cellular proteases (Bessalle et al., 1990). D-chirality can be introduced into a nonribosomal peptide by through A domains that are cognate for a D-amino acid as a monomer substrate (Dittmann et al., 1994), or by the action of a dual C/E domain which performs both condensation and epimerization (Balibar et al., 2005; Rausch et al., 2007)), but in the large majority of cases, the D-chirality stems from the action of a dedicated tailoring domain, the E domain (Figure 1.7).

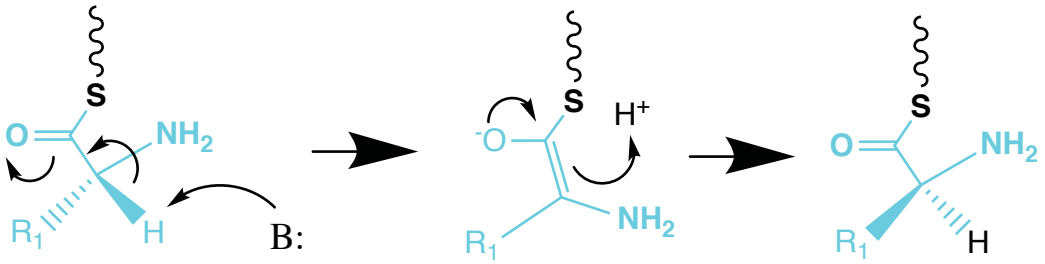


Figure 1.7 Schematic of the epimerization reaction.

Epimerization activity has long been known to associated with nonribosomal peptide synthesis, as it featured in one of the earliest known examples of a purified NRPS (albeit without the modern terminology). In 1968, Yamada & Kurahashi reported the purification of a 120 kDa "phenylalanine racemase" from *Bacillus brevis*, thought to be involved with gramicidin S synthesis (Yamada and Kurahashi, 1968). This protein is now known as gramicidin S synthetase 1 (GrsA). Nearly three decades later, the Vater and Maraheil labs found that epimerization was independent from the adenylation and thiolation activities of GrsA (Stachelhaus and Marahiel, 1995; Stein et al., 1995), and localized the activity to the C-terminal portion of GrsA (Stachelhaus and Marahiel, 1995). So, knowledge of the existence of the E domain actually predated that of the C domain. In de Crecy-Lagard's discovery of the C domain (De Crecy-Lagard et al., 1995), the E domain was also shown to have the same HHxxxDG motif as C domains, with lower conservation of the glycine. Also, despite the domains possessing detectable homology, this active site motif is the only core motif shared by the E and C domain (De Crecy-Lagard et al., 1995; Konz and Marahiel, 1999; Rausch et al., 2007;

Roongsawang et al., 2005). Instead of replacing C domains like dual C/E or Cy domains do, E domains are typically found as a fourth domain in a C-A-PCP-E module.

The E domain has been subject to good biochemical and mutagenic study. In a comprehensive interrogation of GrsA-E1, residues conserved across 45 E domains were targeted for mutagenesis (Stachelhaus and Walsh, 2000). The H753 (*cda H157*) and D757 (*cda D161*) of the active site motif, and Y976 (V365) were shown to be essential for deprotonation of the C_{α} of L-phenylalanine, which supported a common active site in C and E domains. Actual epimerization was decreased at least 200 fold in mutants H753A (*cda H157*), D757A (*cda D161*), E892A (*cda P282*), and R896A (*cda R296*). In addition, while wildtype GrsA-E1 promotes a ~2:1 ratio of D- to L-phenylalanine with rapid reaction kinetics in both directions (Stachelhaus and Walsh, 2000; Takahashi et al., 1971), a mutation of D767S (*cda R171*) or D673S (*cda D60*) selectively decreases the rate of D- to L- conversion by at least 200-fold relative through an unknown mechanism.

Specificity towards the potential substrates and products of the E domain play a particularly important role in NRPS synthesis. In addition to specific PCP – E domain protein-protein interactions (Linne et al., 2001), the E domain active site displays substrate specificity, and acts preferentially on peptidyl-PCP_n, rather than L-aminoacyl-PCP_n (which exists before condensation catalyzed by the C_n domain of the E domain's module) (Luo et al., 2001; Stein et al., 2005). Furthermore, the downstream C_{n+1} domain selects the peptidyl-PCP_n with the D- amino acid as the competent donor for its condensation reaction (Clugston et al., 2003; Stachelhaus and Walsh, 2000). This defines the timing of the epimerization reaction, as after condensation in the current (n) module and before condensation in the downstream (n+1) module, and reveals that these specifies could act as a mechanism to ensure aberrant initiation does not occur in elongation modules with E domains (Luo et al., 2002).

1.8.3 The Ester-bond Forming C Domains

While the monomers in most nonribosomal peptides are linked by amide bonds, some nonribosomal peptides feature ester links between monomers. Nonribosomal depsipeptides like kutzneride have a single ester link, (Fujimori et al., 2007), and others

like valinomycin (Cheng, 2006) and cereulide (Agata et al., 1995; Magarvey et al., 2006) have alternating ester and amide links. These ester linkages result from certain C domains in these systems catalyzing ester bond formation between an aminoacyl/peptidyl-PCP donor and a hydroxyl-containing accepter.

An example of ester-forming C domains has been studied *in vitro*. Zaleta-Rivera *et al.* (Zaleta-Rivera et al., 2006) deciphered that the Fum14p PCP-C didomain performs two condensation reactions using the hydroxyl groups of the free-standing hydroxyalkane fumonisin as the acceptor substrates and tricarballyl-PCP as the donor substrate. This C domain was also able use tricarballyl-SNACs as donors. The active site motif in Fum14p is DHTHCDA, which retains H245 (*cda H157*), but has an aspartate in place of a first histidine (*cda H156*), and the residues between H245 (*cda H157*) and D249 (*cda D161*) are unusually hydrophilic. However, it remained to be seen whether this is a typical feature of ester-bond forming C domains, or an outlier.

1.8.4 The Terminal C Domain of Fungal NRPSs

In bacteria, the final domain in an NRPS system which produce macrocyclic products is typically a Te domain, but in fungi, macrocyclic-producing NRPSs have been found to terminate with a specialized C domain (C_T) (Cramer et al., 2006; Du and Lou, 2010; Kopp and Marahiel, 2007; von Dohren, 2009; Weber et al., 1994). These C_T domains catalyze the attack of an amino group from the peptide its own thioester carbonyl carbon to release the peptide (Figure 1.8). For example, in cyclosporin synthetase, the nucleophile is the α-amino of the first amino acid in the peptide (D-alanine), resulting in the unbranched, macrocyclic form of cyclosporin.

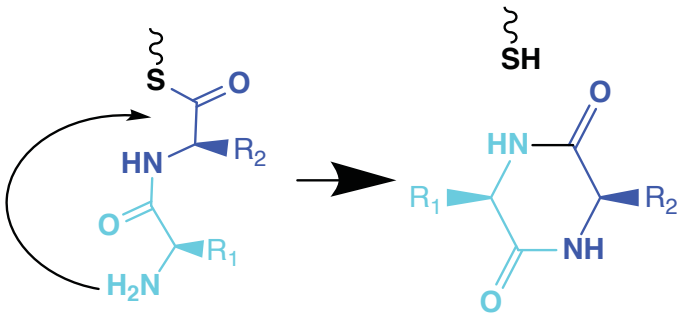


Figure 1.8 Schematic of the macrocyclization reaction.

1.8.5 The X Domain

Glycopeptide antibiotics (GPAs) like vancomycin and teicoplanin are nonribosomal peptides that not only are modified with sugar residues, but also feature several covalent crosslinks between amino acid side chains (Yim et al., 2014). These crosslinks result in multiple macrocycles and a characteristic 3D structure of the antibiotic that is critical for their function. A series of biochemical studies identified P450 monooxygenase enzymes that catalyze the cross-link formation while the peptide is attached to the NRPS (Bischoff et al., 2005; Zerbe et al., 2004). It was noted that all GPA-producing NRPSs contain terminal modules with a domain of unknown function(van Wageningen et al., 1998) homologous to a C domain, located between PCP and Te domains (Bischoff et al., 2005). This was called the X domain, and Rausch showed that it was evolutionarily related to $^{L}C_{L}$ domains, but because it did not have a typical HHxxxDG motif, concluded that it was an inactive domain. Stegmann *et al.* (Stegmann et al., 2010) speculated that it could recruit the P450 monooxygenase enzymes to the NRPS, but direct evidence had not been reported to show that.

1.9 Thesis Objectives and Overview

Amide bond formation is a critical part of nonribosomal peptide biosynthesis. Although known to be typically performed by the condensation domain, the reaction mechanism and specificity determinants of the protein are not known, potentially hindering NRPS bioengineering efforts. In addition, although the heterocyclication domain is believed to adopt a similar fold to the condensation domain, the former performs the same condensation reaction as the latter in addition to an intramolecular cyclodehydration reaction. Limited mutational studies have led to the proposal of hypotheses about how this can take place, but the actual reaction mechanism and residues specifically involved in the reaction are currently unknown as well.

In chapter 2, I briefly describe the crystallization of VioAC1, the first condensation domain of the viomycin synthetase. As I was unsuccessful in solving the structure, I changed my focus to CDA-C1, the first condensation domain in the calcium-dependent antibiotic NRPS in Chapter 3. Here, I describe the crystal structure of CDA-C1, which

was crystallized in two different space groups. The C domain maintains a similar conformation in both instances, but is different when compared to other C domains published at the time. Using an HPLC-mass spectrometry assay, I determined that CDA-C1 is active, and small angle X-ray scattering experiments suggest that the novel conformation observed in the crystal structures could faithfully represent the conformation in solution. Using targeted molecular dynamics simulations, normal mode analyses and energy minimized linear interpolation, I visualized possible conformational changes a C domain could go during the biosynthetic cycle, and discussed the possible implications.

With a crystallizable C domain and an activity assay, I set out to co-crystalize the protein with a substrate analogue in order to determine specificity determinants in Chapter 4. Unable to succeed with free substrates, I developed novel chemical probes that covalently bind to an engineered cysteine residue located near the active site of CDA-C1, mimicking native acceptor substrate delivery to the site by carrier domains. After verifying that the covently-linked probes were competent substrates for the condensation reaction, solved the crystal structure of the CDA-C1 in complex with the probes, which suggested that the principal role of the active site histidine is to position the alpha amino group of the acceptor substrate for nucleophilic attack and led to the identification of a mutation that altered the acceptor substrate specificity

Finally, in chapter 5, I solved the crystal structure of a heterocyclization domain from the bacillamide synthetase. Using the structure, I identified conserved residues which could potentially be involved in condensation and cyclodehydration. Using HPLC-mass spectrometry assay with intact bacillamide synthetase, I was able to probe the effect of these mutations on both reactions. With the assay results and in extensive bioinformatics analysis, I was able to implicate a previously unexamined Asp-Thr dyad in the cyclodehydration reaction.

CHAPTER 2: CRYSTALLIZATION AND PRELIMINARY CRYSTALLOGRAPHIC ANALYSIS OF THE FIRST CONDENSATION DOMAIN OF VIOMYCIN SYNTHETASE

Bloudoff K and Schmeing TM (2013). Crystallization and preliminary crystallographic analysis of the first condensation domain of viomycin synthetase. *Acta Crystallographica Section F* 69(4): 412–415.

2.1 Introduction

Nonribosomal peptide synthetases (NRPSs) are a family of large multimodular enzymes that synthesize peptide chemicals with interesting and useful biological properties which include anti-fungals, anti-bacterials, anti-virals, anti-tumours, siderophores, and immunosuppressants (Konz and Marahiel, 1999; Walsh, 2008). NRPSs are organized into modules of ~1100 residues, with each module responsible for the addition of one specific monomer to the peptide product (Strieker et al., 2010). In general, a module consists of three domains that are essential to the elongation of the peptide: the adenylation (A) domain, the peptidyl carrier protein (PCP) domain and the condensation (C) domain. In an elongation cycle, the A domain binds specifically to one monomer, activates it by adenylation, and then transfers it onto the 4'-phosphopantetheinyl arm of the PCP domain. The PCP domains of the current and previous modules transport their covalently-bound substrates into the active site of the C domain, which catalyzes formation of a peptide bond between the two substrates (Figure 2.1). The elongated molecule is then carried to the C domain in the next module by the current module's PCP domain, to continue this assembly line elongation.

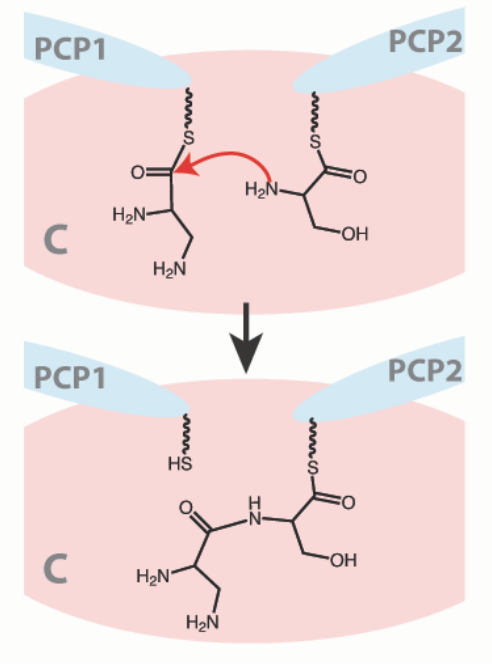


Figure 2.1. The peptidyl transferase reaction catalyzed by VioA-C1 (Helmetag et al., 2009)**.** The wavy lines represent the 4'-phosphopantetheinyl arms of the PCP domains.

Although the C domain plays the central chemical role in nonribosomal peptide synthesis and structures of C domains in their unliganded forms have been determined (Keating et al., 2002; Samel et al., 2007; Tanovic et al., 2008), key questions remain unanswered. The C domains possess a highly conserved histidine motif (HHxxxDG) (Marahiel et al., 1997) involved in catalysis, but there is disagreement about the precise enzymatic mechanism: It has been proposed that the second histidine of the

motif acts as a general base in the reaction (Bergendahl et al., 2002), but a computational analysis of the structure of a C domain from tyrocidine synthetase suggests that the pKa of this residue makes it incapable of acting as a general base, and an electrostatic role was put forward for it instead (Samel et al., 2007). In addition, although A domains are primarily responsible for substrate selection, C domains are also known to display some substrate specificity, the basis of which is unknown (Belshaw et al., 1999; Clugston et al., 2003; Ehmann et al., 2000; Kraas et al., 2012; Linne and Marahiel, 2000; Stachelhaus and Marahiel, 1995). Clarification of the catalytic mechanism and substrate selectivity of C domains would lead to better fundamental understanding of NRPSs and could potentiate future bioengineering efforts to produce designer peptides from modified NRPS systems.

Viomycin is a cyclic nonribosomal peptide antibiotic in the tuberactinomycin family. These drugs are an important class of antibiotics used specifically against multidrug-resistant tuberculosis (Caminero et al., 2010), some of which are included on the World Health Organization's Model List of Essential Medicines (World_Health_Organization, 2013). Viomycin acts by binding the small subunit of the bacterial ribosome and strongly stabilizes tRNA in the ribosomal A site, which blocks tRNA translocation to the P site, and thus inhibits protein synthesis (Modolell and Vazquez, 1977; Stanley et al., 2010). To study viomycin synthesis and NRPS condensation in general, we initiated structural studies of the first C domain of viomycin synthetase from *Streptomyces vinaceus*, VioA-C1 (Yin et al., 2003).

We cloned, expressed, purified and crystallized VioA-C1. We obtained a diffraction data set which extends to 2.9 Å resolution. Future structure determination of apo and co-complex forms of VioA-C1 should help clarify mechanisms of the key catalytic step for one of nature's most elegant macromolecular machines.

2.2 Materials and methods

2.2.1 Cloning and expression

VioA-C1 was cloned from gene *vioA* (European Nucleotide Archive accession number AAP92491.1) in *Streptomyces vinaceus* genomic DNA obtained from the

American Type Culture Collection (ATCC, Manassas, Virginia, USA; ATCC number 11861) using primers VioAC1-F (5'-AAAAAATCATGAGAACCGCGGAAGGCGAGCTGTC-3') and VioAC1-R (5'-AAAAAACTCGAGTTAGGACAGCAGAGGCAGCCGGG-3'). These primers contain BspH1 and XhoI restriction sites, respectively and allow placement into a pET28-derived vector featuring an N-terminal octa-histidine tag and tobacco etch virus (TEV) protease cleavage site of sequence MHHHHHHHHPDLGTGSENLYFQGAMR to give the expression vector pHTVioAC1. Protein production was performed in Rosetta2 (DE3) *E. coli* (EMD Millipore, Billerica, Massachusetts, USA) cells grown in 6L of LB medium supplemented by 300 μg ml^{-1} kanamycin and 340 μg ml^{-1} chloramphenicol. VioA-C1 expression was induced at an OD_{600} of ~0.4–0.6 with the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside, and expression continued for 5 hours at 295 K. Cells were then harvested and stored at 193 K.

2.2.2 Protein purification

VioA-C1 cell pellets were resuspended in a buffer containing 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM β-mercaptoethanol, 1 mM imidazole pH 8.0 and 1 mM phenylmethanesulfonyl fluoride. Cells were lysed using an EmulsiFlex-C3 homogenizer (Avestin Inc., Canada) and centrifuged for 30 minutes at 40,000 g and 277 K. The resulting supernatant was pooled and applied to a 5 mL HiTrap IMAC FF column (GE Healthcare) charged with Ni^{2+}. VioA-C1 was eluted using a buffer containing 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM β-mercaptoethanol, 250 mM imidazole pH 8.0 and 1 mM phenylmethanesulfonyl fluoride, and the purity assessed using SDS-polyacrylamide gel electrophoresis.

The fractions containing VioA-C1 were pooled and digested overnight with TEV protease, using a ratio of 1 mg TEV protease per 40 mg VioA-C1, while dialyzing against a buffer containing 50 mM Tris-HCl pH 7.5, 100 mM NaCl and 2 mM β-mercaptoethanol (dialysis buffer). Cleavage with TEV protease results in a remnant sequence of GAMR on the N-terminus of VioA-C1. The sample was reapplied to the 5 mL HiTrap IMAC FF column charged with Ni^{2+} and the flowthrough was collected and pooled. The sample was concentrated and applied to a HiPrep 26/60 Sephacryl S-300

HR column (GE Healthcare) equilibrated in dialysis buffer. VioA-C1–containing fractions were pooled and diluted 1:1 with 50 mM Tris-HCl pH 7.5 and 2 mM β-mercaptoethanol for anion exchange chromatography with a MonoQ 5/50 GL column (GE Healthcare). VioA-C1 was eluted from the MonoQ column by a gradient from 0–1 M NaCl. High and low salt MonoQ buffers also contained 50 mM Tris-HCl, pH 7.5 and 2 mM β-mercaptoethanol. Finally, pure VioA-C1 (Figure 2.2) was placed into a buffer of 20 mM Tris-HCl, pH 7.5, 100 mM NaCl and 2 mM β-mercaptoethanol, by serial dilution and concentration using an Amicon Ultra 10K centrifugal filter device (EMD Millipore, Billerica, Massachusetts, USA). At a concentration of 10 mg ml^{-1}, VioA-C1 was flash frozen in liquid nitrogen and stored at 193 K.

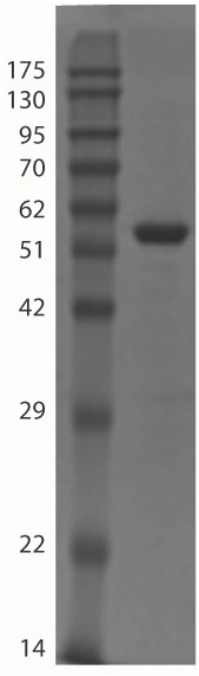


Figure 2.2. 12% SDS-polyacrylamide gel used to assess the purity of VioA-C1. Molecular weight markers present in the left lane and the sizes in units of kDa are indicated.

2.2.3 Crystallization

Sparse matrix crystallization trials of VioA-C1 were performed using the sitting-drop vapor-diffusion method with the aid of a Phoenix crystallization robot (Art Robbins, Sunnyvale, California, USA) in Intelli-Plates 96 (Art Robbins, Sunnyvale, California, USA). Crystallization drops with 0.2 µl of 5 or 10 mg ml^{-1} protein sample and 0.2µl crystallization solution were equilibrated against 75 µl reservoir volume. Crystallization solutions from the PACT, Classics I, Classics II, PEGs and JCSG+ crystallization suites (Qiagen, Germantown, Maryland, USA), were assayed for promotion of crystal growth at temperatures of 277K and 295 K.

Initial crystallization screens yielded small, fragile needles growing in heavy precipitate from an experiment using 0.2 M magnesium chloride, 0.1 M HEPES pH 7.5 and 25% (w/v PEG 3350) as the crystallization solution. Subsequent optimization was performed in larger format 24-well sitting-drop Cryschem Plates (Hampton Research, Aliso Viejo, California, USA) by manual two-dimensional grid screens around the initial conditions. In these screens, the drop contained 2 µl of 10 mg ml^{-1} protein sample and 2

µl of crystallization solution, and was equilibrated against a 400 µl reservoir of crystallization solution. Further optimization was done with multiple screenings of the Hampton Additive Screen (Hampton Research, Aliso Viejo, California). Crystals suitable for data collection were grown using a crystallization solution of 11-16% polyethylene glycol (PEG) 3350, 0.2 M ammonium acetate, 0.1 M HEPES, pH 7.5, and 0.01-0.025 mM $MnCl_2$ at 277 K. VioA-C1 crystals grew to a maximum size of approximately 400 x 30 x 10 µm, and grew only out of drops containing heavy precipitate (Figure 2.3). Crystals were cryoprotected by adding 25% PEG 400 to the equilibrated crystal drop, mounting into cryoloops and flash-cooled in liquid nitrogen.

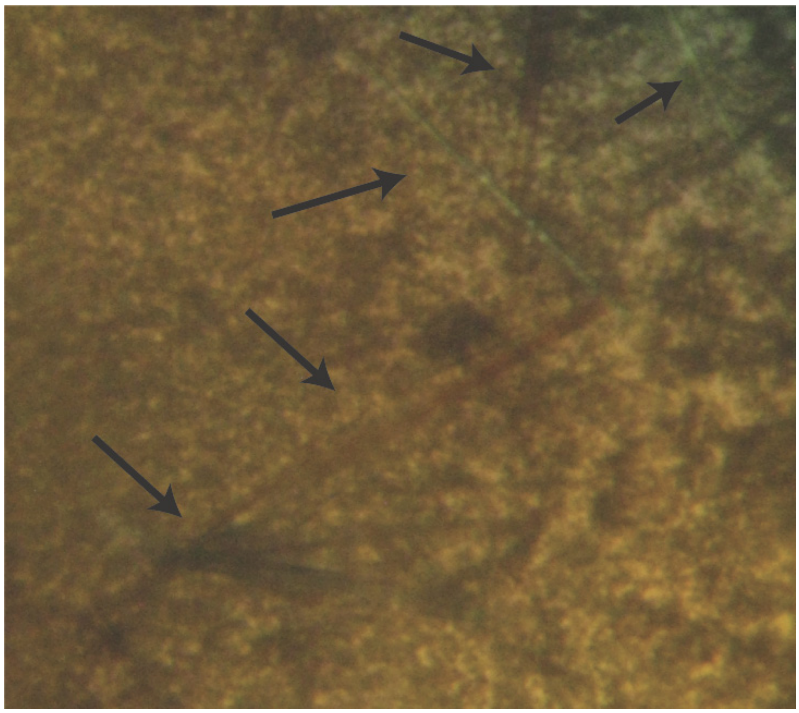


Figure 2.3. A drop containing heavy precipitate and crystals of VioA-C1. The largest crystals are approximately 400 x 30 x 10 µm. Arrows indicate the positions of crystals.

2.2.4 Data collection and processing

Diffraction data were collected using a Rigaku rotating copper-anode generator fitted with Osmic confocal optics and an R-AXIS IV^{++} image-plate detector (Rigaku, Tokyo, Japan) at the Biochemistry Department of McGill University, Montreal,

Canada. A 98° data set was collected with an image oscillation width of 1°, an exposure time of 10 minutes per image and a crystal-to-detector distance of 230 mm.

The raw data were processed using the program *HKL2000* (Otwinowski and Minor, 1997). The data were indexed in the orthorhombic space group $P2_12_12_1$, with unit cell parameters a = 46.165, b = 68.335, c = 146.423 Å. The statistics for data collection and processing are listed in Table 2.1. Calculations reveal that there is likely one monomer in the asymmetric unit, which would give a solvent content of 49.2% and a Matthews coefficient (V_M) of 2.42 $Å^3$ Da^{-1}.

Values in parentheses are for the highest resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 46.165, b = 68.335, c = 146.423
Wavelength (Å)	1.5418
Resolution (Å)	50.00-2.86 (2.91-2.86)
No. of unique reflections	11209 (527)
No. of observed reflections	39463
Completeness (%)	99.3 (98.7)
Multiplicity	3.5 (3.5)
R_{merge} † (%)	16.0 (67.1)
$\langle I/\sigma(I)\rangle$	7.0 (1.6)
Temperature (K)	100
Matthews coefficient ($Å^3$ Da^{-1})	2.42
Solvent content (%)	49.2
No. of molecules in the asymmetric unit	1

† $R_{merge} = \sum_{hkl}\sum_I | I_i(hkl) - \langle I(\mathrm{hkl})\rangle | / \sum_{hkl}\sum_I I_i(hkl)$, where $I_i(hkl)$ represents the intensity of the ith measurement and $\langle I(hkl)\rangle$ represents the average intensity of reflection hkl.

Table 2.1. Data-collection and processing statistics of VioAC1.

2.3 Results and discussion

VioA-C1 from *S. vinaceus* was heterologously expressed in Rosetta2 (DE3) *E. coli*, as an octa-histidine-tagged construct and purified to homogeneity by a four-step purification consisting of immobilized-metal affinity chromatography, reverse immobilized-metal affinity chromatography, size exclusion chromatography and anion exchange chromatography. Each litre of bacterial culture yielded approximately 2.1 mg of pure VioA-C1. Crystallization conditions obtained from initial screens were optimized

using sitting drop vapor-diffusion experiments; 2 µl of 10 mg ml^{-1} VioA-C1 in a buffer of 20 mM Tris-HCl, pH 7.5, 100 mM NaCl and 2 mM β-mercaptoethanol were mixed with 2 µl reservoir solution and equilibrated against 400 µl reservoir solution. The best diffracting crystals grew in 11-16% PEG 3350, 0.2 M ammonium acetate, 0.1 M HEPES, pH 7.5, and 0.01-0.025 mM $MnCl_2$, at a temperature of 277 K. This condition produced drops with both crystals and heavy precipitate, but attempts to grow crystals in conditions without precipitate using micro- or macroseeding did not succeed. Diffraction data were collected with the in-house X-ray source and the crystal diffracted to a resolution of 2.86 Å with an outer shell $\langle I/\sigma(I)\rangle$ of 1.6 (or 2.96 Å with an $\langle I/\sigma(I)\rangle$ of 2) (Figure 2.4).

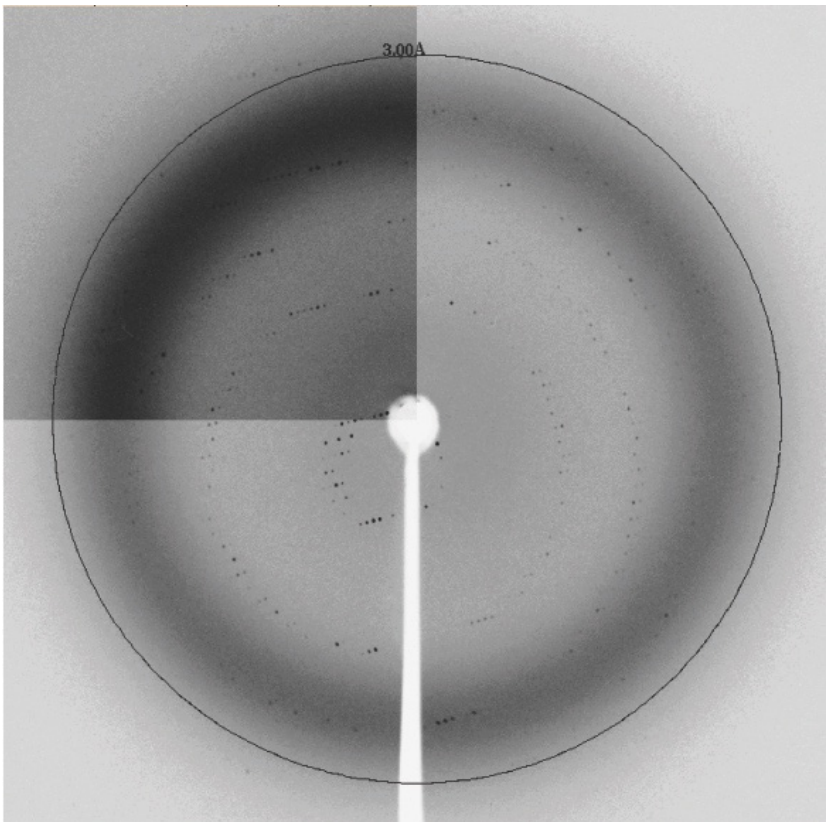


Figure 2.4. An X-ray diffraction image from a crystal of VioA-C1. The crystal-to-detector distance was 230 mm; exposure time was 10 minutes per image.

Indexing of the data indicated that the crystals belonged to the orthorhombic space group $P2_12_12_1$ with unit cell parameters *a* = 46.165, *b* = 68.335, *c* = 146.423 Å. The

Matthews coefficient calculation for one monomer in the asymmetric unit is 2.42 $Å^3$/Da, making it likely that there is one monomer in the asymmetric unit and a solvent content of 49.2%.

There are structures available for C domains from tyrocidine synthetase III (Protein Data Bank identification code (PDB ID) 2JGP, sequence identity with VioA-C1 of 25%) (Samel et al., 2007), vibriobactin synthase (PDB ID 1L5A, 17% sequence identity) (Keating et al., 2002) and surfactin A synthetase C (PDB ID 2VSQ, 20% sequence identity) (Tanovic et al., 2008). Thus far we have been unable to determine the structure of VioA-C1 using molecular replacement phasing, despite trials with all three full C domain structures and multiple subdomain models of them as search models. The lack of success is likely due to conformational changes between and within the N- and C-terminal subdomains of the C domains, which can be observed in the published C domain structures, as well as fairly low sequence conservation between C domains. We have grown crystals of selenomethionine-derivatized VioA-C1 for structure determination with multiple or single wavelength anomalous dispersion phasing techniques.

2.4 Acknowledgements

We thank our colleagues in the Schmeing lab for helpful advice and discussion. This research was supported by Canadian Institutes of Health Research (CIHR) Operating Grant MOP 106615 awarded to TMS, a Human Frontiers Science Program Organization Career Development Award to TMS and a CIHR Strategic Training Initiative in Chemical Biology studentship held by KB.

2.5 Bridge to Chapter 3

While the structure of VioAC1 remains unsolved at this time, novel structures of C domain superfamily proteins and new molecular replacement pipelines such as MoRDa (Vagin and Lebedev, 2015) may yet allow for its solution. However, for the purpose of my thesis, it was imperative to change my focus to a protein that produced crystals consistently and whose structure I was able to solve. Fortunately, I had been working with another C domain in parallel to VioAC1, which was the starter C domain from the calcium dependant antibiotic synthetase, CDA-C1. This C domain was a protein that had high expression, easily grew thick crystals, and was readily solvable using multiwavelength anomalous diffraction.

While I was waiting for small molecule substrates to be synthesized (for the purpose of soaking and co-crystallization experiments), I made the observation that this C domain was in a novel conformation, which formed the basis of the manuscript published in Chapter 3.

CHAPTER 3: CRYSTAL STRUCTURES OF THE FIRST CONDENSATION DOMAIN OF CDA SYNTHETASE SUGGEST CONFORMATIONAL CHANGES DURING THE SYNTHETIC CYCLE OF NONRIBOSOMAL PEPTIDE SYNTHETASES

Bloudoff K, Rodionov D, and Schmeing TM (2013). Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. *Journal of Molecular Biology* 425(17): 3137–3150.

3.1 Introduction

Nonribosomal peptide synthetases (NRPSs) are large macromolecular machines that catalyze the assembly of monomer substrates into biologically active secondary metabolites (Felnagle et al., 2008; Konz and Marahiel, 1999; Schwarzer et al., 2003). As the name implies, NRPS substrates are often amino acids, but over four hundred monomers are known to be used as substrates, including D-amino acids, aryl acids, keto acids, hydroxy acids, and fatty acids (Caboche et al., 2008). Nonribosomal peptides have important and diverse biological activity and include anti-fungals, anti-bacterials, anti-virals, anti-tumours, siderophores, and immunosuppressants (Konz and Marahiel, 1999), including well-known compounds such as penicillin (van Liempt et al., 1989), daptomycin (Wessels et al., 1996), and cyclosporin (Zocher et al., 1986).

NRPSs are organized into modules of >110 kDa, with each module responsible for the addition of one specific monomer. Modules contain multiple domains, each performing specific functions in product synthesis (Weber and Marahiel, 2001). A basic elongation module contains a condensation (C) domain, an adenylation (A) domain and a peptide carrier protein (PCP) domain. The A domain selects the cognate amino acid and adenylates it, then attaches it to a phosphopantetheinyl (PPE) group on the PCP domain. The PCP domain transports the amino acid to the C domain, which catalyzes peptide bond formation between this amino acid and the peptide attached to the PCP domain of the preceding module, thus elongating the peptide chain. Next, the PCP domain brings the elongated peptide chain to the downstream module, where it is passed off and further elongated in the next peptidyl transferase reaction.

In this reaction cycle, numerous conformational changes are known, or have been proposed, to occur. As described above, the PCP domain must completely translocate to interact with different partner domains (Frueh et al., 2008; Tanovic et al., 2008), and is also known to change conformation, depending on its functional state (Koglin et al., 2006). Large scale-movements are known to occur in the A domain between the conformation which catalyzes the adenylation reaction and that which catalyzes the thiolation reaction (attachment of substrate to PCP domain) (Gulick et al., 2003; Mitchell et al., 2012; Reger et al., 2008; Sundlov et al., 2012). Thus, conformational cycling is absolutely essential for NRPS function.

The C domain catalyzes the key catalytic event of NRPS function, peptide (amide) bond formation (De Crecy-Lagard et al., 1995; Stachelhaus et al., 1998). Three structures which include NRPS C domains have been determined by X-ray crystallography: a stand-alone C domain (Keating et al., 2002), a C–PCP didomain complex (Samel et al., 2007) and a C-A-PCP-Te termination module (Tanovic et al., 2008). The C-domain comprises ~450 amino acids and has a pseudo-dimer configuration, with both N and C-terminal subdomains displaying folds in the CoA dependent acyltransferase (CAT) superfamily. The active site sits at the bottom of a "canyon" (Samel et al., 2007) or "V" (Tanovic et al., 2008) formed by the two subdomains of the C domain, and is covered by a "latch" that crosses over from C to N subdomain. The catalytic center includes an HHxxDG sequence motif (Bergendahl et al., 2002; De Crecy-Lagard et al., 1995; Stachelhaus et al., 1998) and must have binding sites for its donor and acceptor substrates. The conformations of the C domain visualized in these three structures vary somewhat, though it is unclear whether these differences stem from the fact that different proteins were used in the three studies or from the possibility that they are in different functional states. Although there are hundreds of different C domains in NRPS biosynthetic clusters, they likely all share a common mode of action.

In this study we focus on the first C domain of the calcium dependent antibiotic (CDA) synthetase, CDA-C1 (Figure 3.1) (Chong et al., 1998; Hojati et al., 2002). The calcium dependent antibiotic is part of a class of acidic lipopeptides which includes the last resort antibiotic daptomycin (Strieker and Marahiel, 2009). These antibiotics work by binding to and disrupting cytosolic membranes in bacteria (Lakey et al., 1983). The CDA biosynthetic cluster in *Streptomyces coelicolor* includes an 11-module canonical NRPS spread over 3 proteins, which adds the 11 amino acids and cyclizes the product (Figure 3.1a) (Hojati et al., 2002).

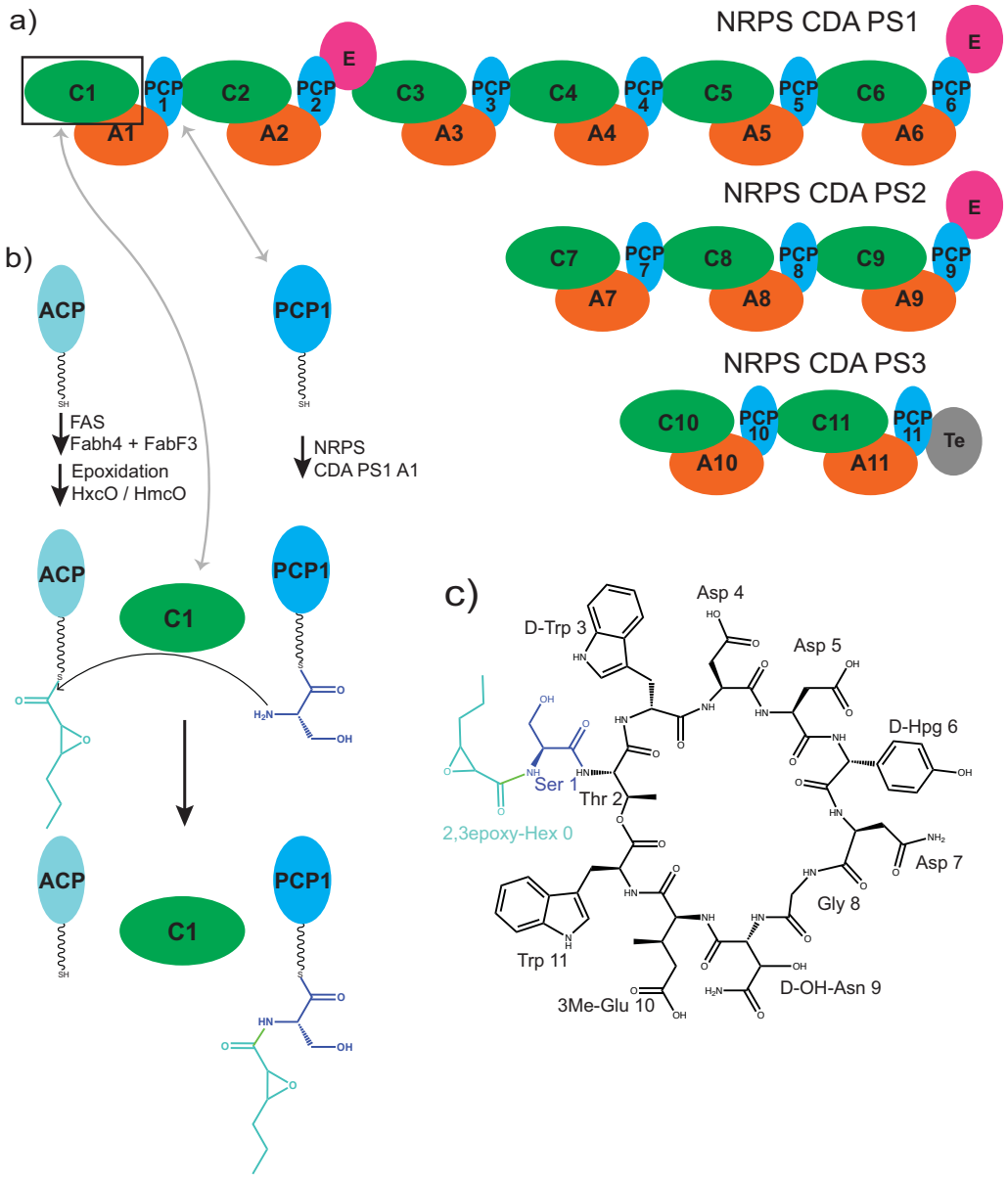


Figure 3.1. Schematic diagrams for CDA components. a) A schematic diagram of the proteins CDA peptide synthetase 1, 2 and 3 which make up the NRPS system for calcium dependent antibiotic synthesis in *Streptomyces coelicolor*. b) A schematic diagram illustrating the reaction catalyzed by CDA-C1. c) A schematic diagram of the chemical structure of CDA-4b, a representative CDA peptide synthesized in *Streptomyces coelicolor* (Caboche et al., 2008; Kraas et al., 2012).

The first monomer is not an amino acid, but a fatty acid, a 2,3-epoxyhexanoyl group which is synthesized as a hexanoyl fatty acid on an acyl carrier protein (ACP) by a fatty acid synthase, then modified to the epoxy form by epoxidation enzymes HxcO and/or

HcmO (Figure 3.1b) (Kopp et al., 2008; Kraas et al., 2012). The first C domain of CDA synthetase catalyses the transfer of the 2,3-epoxyhexanoyl group from 2,3-epoxyhexanoyl-ACP to the serinyl-PCP domain of the first module. Although normally present as part of the 799 kDa CDA PS1, excised CDA-C1 is active in catalysis (Kraas et al., 2012).

Here we present two structures of the first condensation domain of the calcium-dependent antibiotic (CDA) synthetase (CDA-C1), determined by X-ray crystallography at resolutions of 1.8 Å and 2.4 Å, and accompanying small angle X-ray scattering (SAXS), activity assays and computational analyses.

3.2 Results and discussion

3.2.1 Purification, crystallization and structure determination of CDA-C1

CDA-C1 is the N-terminal domain of CdaPS1, a 20 domain, 7463 residue NRPS protein (Kraas et al., 2012). To study CDA-C1, we designed a gene construct by aligning the sequences of CDA-C1 to C domains with known structure to enable C-terminal boundary definition. The resulting construct was heterologously expressed in *E. coli* with extremely high yields and purified to homogeneity using a four-column chromatographic protocol. CDA-C1 was subjected to high-throughput crystallization trials, which readily yielded crystals in several crystallization conditions. Two crystallization conditions were optimized and could be used reproducibly to give highly diffracting crystals in $P2_12_12_1$ and $P2_1$ space groups. Data collection at a rotating anode "home" source gave complete data sets with good statistics (Table S3.1). These were subjected to phasing trials using molecular replacement with many different search models derived from one or both subdomains of the known C domains, with no success. Selenomethionine-derivatized protein was then produced, which could be purified and crystallized by the same protocols. Multi-wavelength data sets were collected from $P2_12_12_1$ crystals at the National Synchrotron Light Source, and multiwavelength anomalous dispersion (MAD) phasing techniques were used to determine the structure to 1.8 Å resolution (Figure 3.2, Supplementary Figure S3.2, Table S3.1). Electron density maps showed that there are two molecules in asymmetric unit with very similar confirmation, and allowed building of residues 5 - 450 in molecule A and 4 - 449 in

molecule B. The use of this structure as a search model readily gave a molecular replacement solution for the $P2_1$ crystal form using home source data with resolution solved to 2.4 Å (Figure S3.2b, Table S3.1). This final model included residues 5 - 450 in molecule A and 7 - 449 in molecule B.

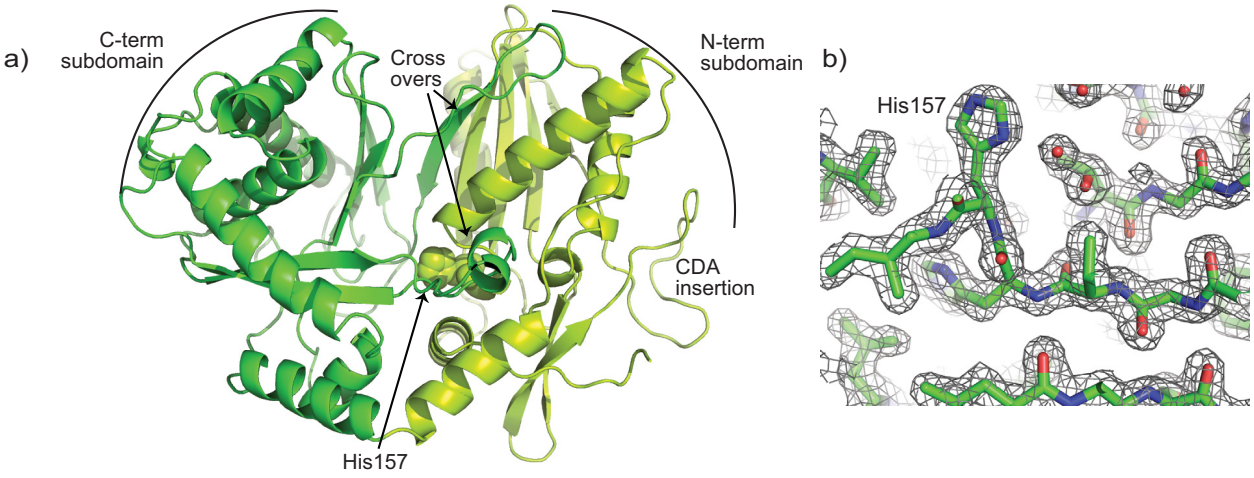


Figure 3.2. Structure of CDA-C1. a) Ribbons representation of CDA-C1 determined in the $P2_12_12_1$ space group. The N-terminal subdomain is chartreuse, the C-terminal subdomain is green, and the domain crossovers, CDA-specific insertion and active site H157 are all indicated. b) An F_O electron density map contoured at 1σ.

3.2.2 The structure of CDA-C1

CDA-C1 adopts the classical C domain conformation (Keating et al., 2002; Samel et al., 2007; Tanovic et al., 2008), with two subdomains, each having a core CAT fold, two points of crossover from the C-terminal subdomain to the N-terminal subdomain, and the active site in the center of the "V" formed by the subdomains (Figure 2.2). Superimposition of all four independent molecules determined here (two molecules from the $P2_12_12_1$ crystal form and two from the $P2_1$ crystal form) show that in all these molecules, the subdomains adopt approximately the same relative orientation to one another (Supplementary Figure S3.3). The only major difference between these structures is in the conformation of the loop 82-96. Loop 82-96 is in the linker between two beta sheets of the N-terminus. Multiple alignment of over 500 C domains (Rausch et al., 2007) shows that this loop is almost unique to CDA-C1, with only a protein 930752.1 from *Photorhabdus luminescens* also having an insertion in the same location

(albeit with no conservation between the two sequences) (Supplementary Figure S3.1). Since CDA-C1 and Plu 930752 are initiator C domains, this face may have less stringent conservation requirements: the C domain does not have to pack with a full upstream module, rather it needs to interact only with the small ACP proteins.

Comparison of the structures presented here with the structures of the three previously determined structures of C domains reveals significant conformational differences. When the C-terminal subdomains of the characterized C domains from tyrocidine synthetase III (TycC) (Samel et al., 2007), vibriobactin synthase (VibH) (Keating et al., 2002) and surfactin A synthetase C (SrfAC) (Tanovic et al., 2008) are superimposed upon CDA-C1, the N-terminal subdomains of these proteins are in strikingly different positions (Figure 3.3, Supplementary Figure S3.4).

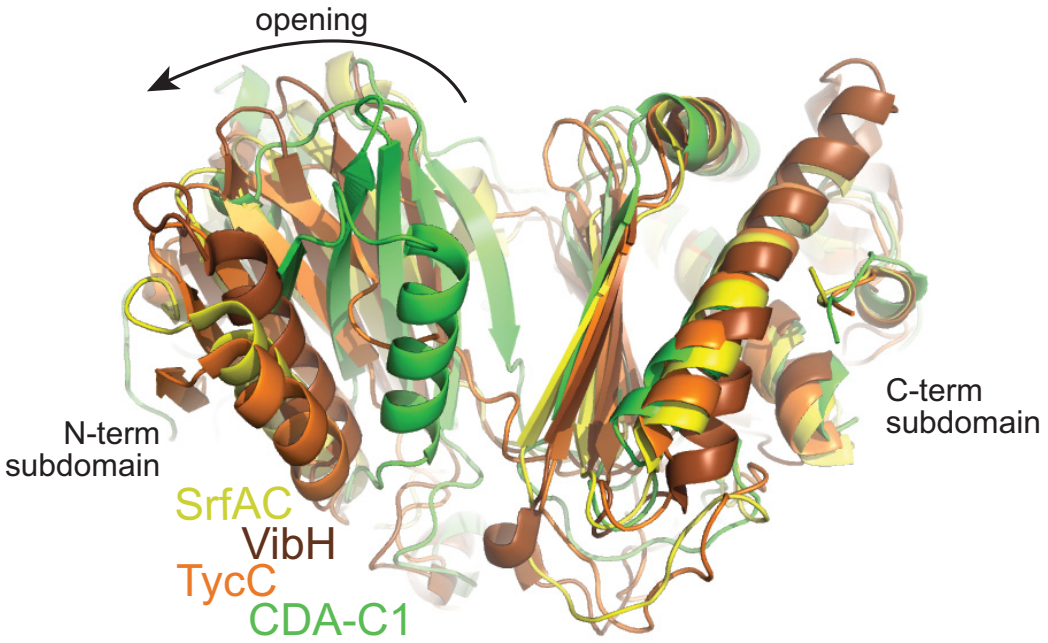


Figure 3.3. Comparison of CDA-C1 with other C domains. Alignment of the C-terminal subdomain shows different relative subdomain–subdomain orientation in the structures of C domains. CDA-C1 displays the most "closed" conformation. CDA-C1 is in green, the C domain from SrfAC is yellow, from VibH is brown and from TycC is orange. See supplementary Figure S3.4 for individual comparisons.

CDA-C1 is in a much more "closed" conformation, with its N and C-terminal subdomains closer together. The N-terminal subdomain would require a shift and large rotation of

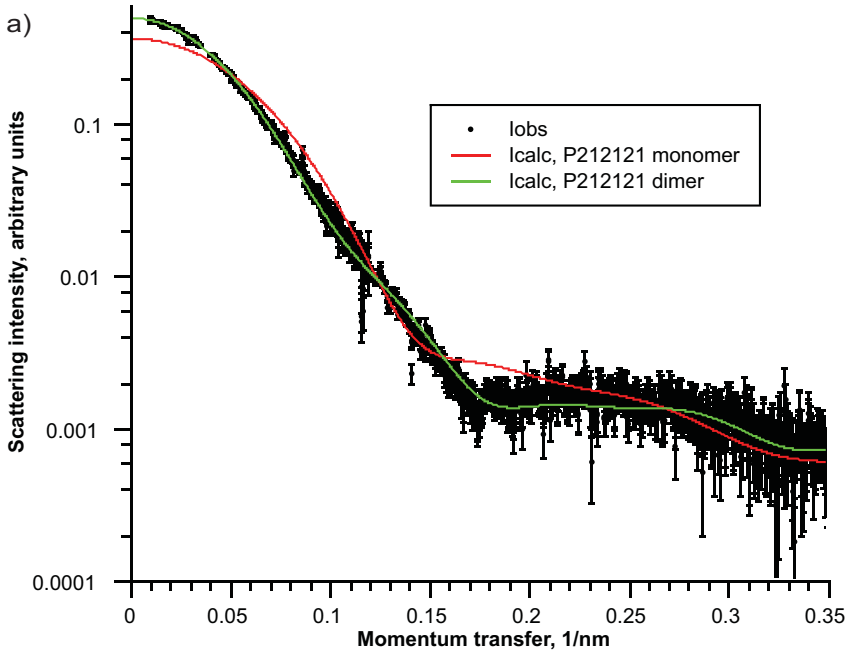
~15°, ~22°, and ~25° respectively to assume the conformations seen with VibH, TycC and SrfAC. This is a large movement – the alpha carbon positions in the N-terminal domain move by as much as 19, 20 and 23 Å and on average differ by 6, 8 and 11 Å.

3.2.3 SAXS analysis of CDA-C1

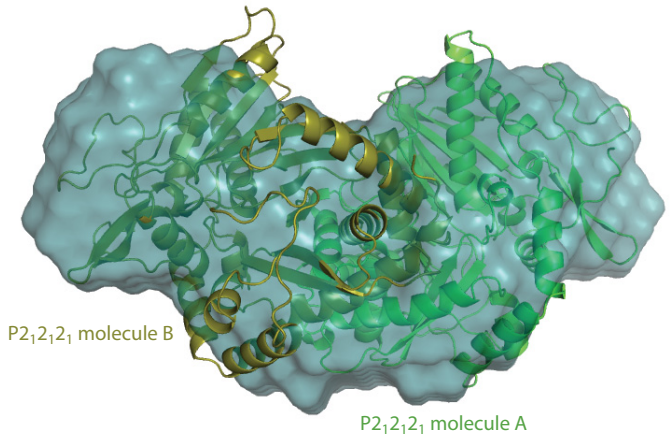
This "closed" conformation of CDA-C1 seemed unlikely to be artificially forced by crystal packing. All CDA-C1 monomers are observed in very similar conformations, even though the packing contacts between the two molecules in each asymmetric unit (crystallographic dimers) are solely through the C-terminal subdomain, and many of the other crystal contacts are different for the two space groups (Figure S3.3d). To address whether this "closed" conformation is indeed adopted in solution, we undertook a SAXS analysis of CDA-C1.

SAXS data were collected on CDA-C1 at multiple concentrations with each experiment giving consistent results (a typical scattering result is shown in Figure 3.4). Molecular weight of CDA-C1 in solution was found to be ~97 kDa using the Kratky plot method (comparing to BSA and lysozyme) which agrees well with calculated molecular weight of CDA-C1 dimer (96.8 kDa) (Figure S3.5). We calculated theoretical scattering curves from the structures of both a single CDA-C1 and the dimer of CDA-C1 observed in both $P2_12_12_1$ and $P2_1$ crystal forms. Comparison of the scattering data and the theoretical curves obviously indicates that CDA-C1 is a dimer in solution, and that the dimer is in a conformation very similar to that of the crystallographic dimer (Figure 3.4a). The envelope calculation also shows a good consistency between our crystallographic dimer and the solution conformations as assayed by SAXS (Figure 3.4b). It is not clear whether the dimerization of the excised C domain would have biological significance or occur in the intact NRPS. Some, but not all, intact NRPSs appear to dimerize (unpublished data). The interface involved in dimerization is not blocked by adjacent domains in the context of a full module (Tanovic et al., 2008), raising the possibility that this interface may be used to form (perhaps transient) intra- or inter- protein contacts in the intact NRPS.

a)



b)



c)

PDB model	χ^2
CDA-C1 conformation (dimer)	1.099
VibH conformation (dimer)	1.132
SrfAC conformation (dimer)	1.139
TycC conformation (dimer)	2.041
CDA-C1 conformation (monomer)	4.522

Figure 3.4. SAXS analyses of CDA-C1. a) Fitting of calculated scattering curves to merged experimental data indicating that CDA-C1 is a dimer in solution b) superposition of averaged filtered envelope on crystallographic CDA-C1 dimer c) quality of fit of various models to experimental scattering data suggesting that CDA-C1 in solution is best described by crystallographic dimer in its observed conformation.

To test whether the SAXS data could differentiate between the C domain conformation observed in the crystal structure reported here and those conformations seen in the previously reported structures (Figure 3.3, Figure S3.4), we created models of CDA-C1 in the conformations of VibH, TycC and SrfAC. These models were used to calculate theoretical scattering curves which were fitted to the experimental SAXS scattering data, with goodness of fit assessed by calculating χ^2 values. CDA-C1 had the lowest χ^2 to the SAXS data (Figure 3.4c), although the values for VibH and SrfAC were only slightly higher. TycC had a substantially higher χ^2. Thus, the conformation we observe in both $P2_12_12_1$ and $P2_1$ crystal forms is at least consistent with solution studies.

3.2.4 CDA-C1 is catalytically active

We next asked whether the construct of CDA-C1 we made, which adopts this "closed" conformation, is catalytically active. We first attempted to show activity using small molecule acyl-N-acetylcysteamine thioester analogues (acyl-SNACs) (Ehmann et al., 2000) of the two substrates (2,3-epoxy-hexanoyl-SNAC and serinyl-SNAC), but we were unable to detect product (2,3-epoxy-hexanoyl-serinyl-SNAC) formation. While we were performing our studies, Krass *et al*. published experiments with a similar construct of CDA-C1, which showed their construct to be active in an assay where the two substrates are delivered on carrier proteins and the product 2,3-epoxy-hexanoyl-serinyl-PCP is separated by HPLC and confirmed by mass spectrometry (Kraas et al., 2012). We could demonstrate catalytic activity of our CDA-C1 construct using a version of this assay (Figure 3.5). The assay is complicated by the fact that serinyl-PCP (and, in our experience, 2,3-epoxy-hexanoyl-ACP) is liable to hydrolysis, and that several PCP species migrate very similarly on reverse phase columns. Nonetheless, we were able to clearly identify product formation when CDA-C1 was in the reaction mix, but not when it was replaced by CDA-C1 harboring an H157A active site mutation (Figure 3.5) (Kraas et al., 2012).

Thus, the CDA-C1 construct we crystallized is catalytically active, and the conformation we observe in the crystal structures is consistent with SAXS results probing its solution conformation.

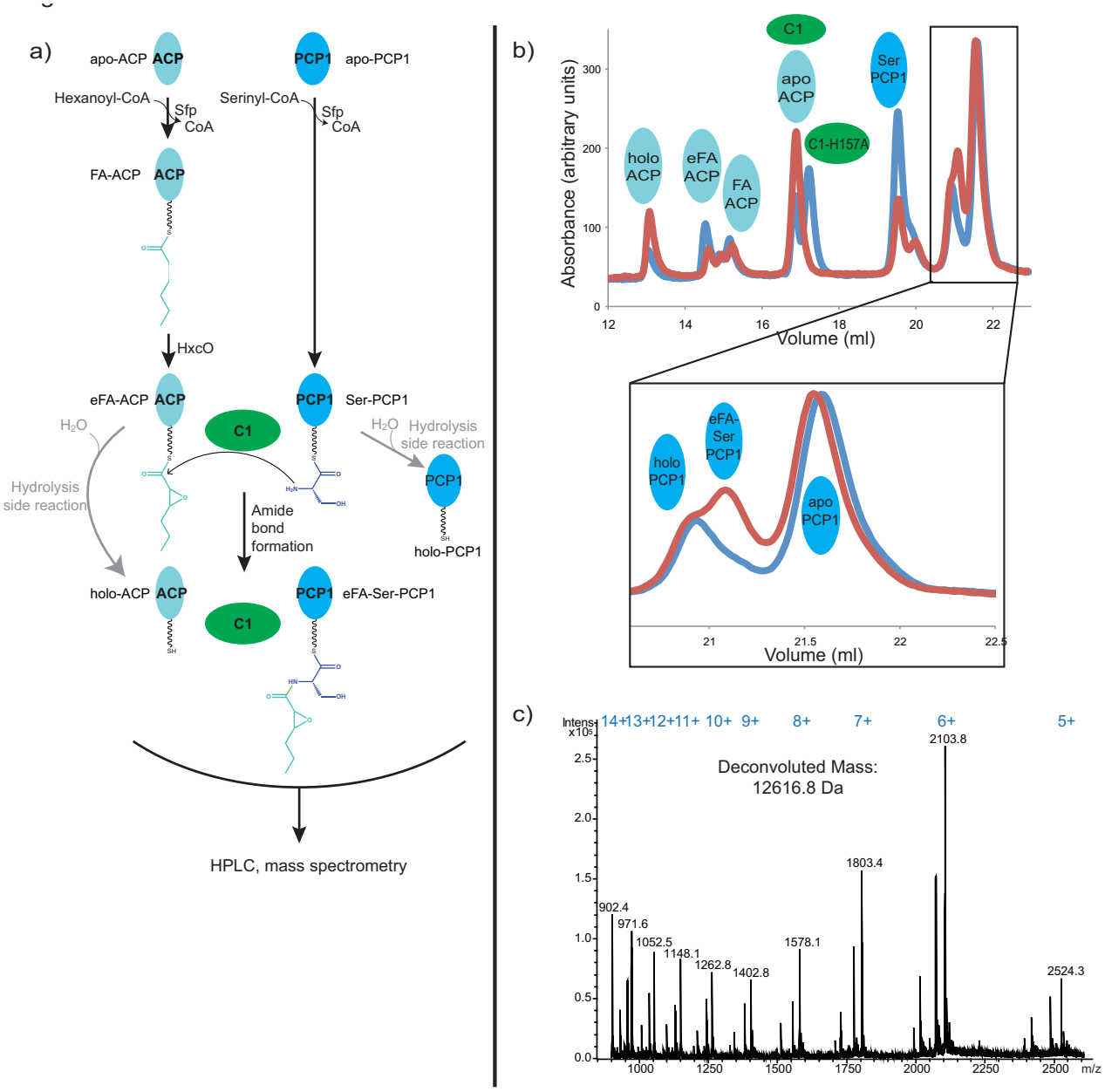


Figure 3.5. CDA-C1 is catalytically active. a) A schematic diagram explaining the C domain activity assay. b) HPLC traces and c) mass spectra of reaction assays show that product epoxyl-hexanoyl-serinyl-PCP1 is formed only when wild type CDA-C1 (red) is present in the reaction mixture. Reactions with CDA-C1 harboring the active-site mutation H157A (blue) do not lead to product formation. Expected mass of epoxyl-hexanoyl-serinyl-PCP1 is 12618.3 Da. Expected and observed masses for other carrier protein states are listed in Table S3.2.

3.2.5 Computational analyses of C domain movement

Although each structurally characterized C domain is seen in a different conformation, there appears to be sufficient space at the active site to bind substrate in each form. One could imagine that progressively more open C domain conformations would occur in C domains from later modules of NRPS to accommodate larger substrates, but that is unlikely to be the cause of the different conformations: VibH is an initiator C domain like CDA-C1, but found in an open conformation, and SrfAC is more open than TycC despite coming from an earlier module in their respective NRPS (Figure 3.3, Figure S3.4, and Movie S3.7). It is unclear if all these conformations are catalytically active, or if a C domain would sample each of the observed conformations during an NRPS catalytic cycle. There is extensive precedent for using crystal structures of homologous proteins to investigate conformational changes that a protein may undergo in solution (Marechal and Perahia, 2008; Van Wynsberghe et al., 2004; Vonrhein et al., 1995; Yonus et al., 2008). Therefore, we undertook analyses of the types of movement required to transition between observed conformations.

We subjected CDA-C1 to normal mode analysis to observe whether a transition from “closed” to “open” form would be replicated by normal modes, using the programs WEBnm@ (Hollup et al., 2005) and NOMAD-Ref (Lindahl et al., 2006). Both programs reveal that such a movement is reasonable. WEBnm@ replicates this movement very well (Movie S3.1), although in this simulation, the amplitude is not quite enough to bring it to a fully “open” state. Likewise, several NOMAD-Ref modes appear to replicate similar movements (Movie S3.2-S3.3).

To study and visualize the full transition between closed and open states as well as between the various open states, we carried out two further computational exercises. First, we performed linear interpolation with energy minimization refinement (“morphing”) to interconvert the conformations. The resulting animations show the transition from the conformation seen in this study to those of previously published structures (Movie S3.4-S3.6), as well as the movements required to transition between these states (Movie S3.7). Secondly, we performed targeted molecular dynamics simulations on the same transitions (Movie S3.8-S3.11). The targeted molecular dynamics simulations transitions appear smooth and do not seem to have to pass

through any obviously unfavorable conformations. The root mean square deviations between progressive models in the simulations and the target models in SrfAC, TycC and VibH conformation smoothly decrease for residues in the N terminal subdomain and crossover strands (Figure S3.5).

Together, the normal mode analyses, the morphing and targeted molecular dynamics simulations suggest that the conformation changes which would be required to interconvert the four conformations observed in NRPS C domains are reasonable. The putative conformational changes describe here involve relative movement between the two CAT fold containing subdomains, while the conformation of each CAT core fold remains unchanged. There are numerous proteins which contain CAT folds, three of which have been characterized structurally and shown to also contain two "pseudodimeric" CAT folds: murine carnitine acetyltransferase (CrAT) (Hsiao et al., 2006), human choline acetyltransferase (CHAT) (Kim et al., 2006) and polyketide-associated protein A5 (PapA5; which has a similar overall structure to C domains) (Buglino et al., 2004). We asked whether similar opening has been seen in these three proteins. CrAT does not undergo significant conformational changes upon binding substrate (Hsiao et al., 2006); CHAT undergoes small but significant conformational change (~1.5° opening) upon substrate binding (Kim et al., 2006) and PapA5 is hypothesized to undergo some kind of opening of the active site to allow substrate to bind, as an α helix blocks the active site in the apo conformation, but this opening has yet to be observed (Buglino et al., 2004). Therefore, the opening of two CAT fold subdomains is a feature in some, but not all proteins which contain them.

Though some C domains are expressed as stand-alone proteins, they are more usually part of a much larger protein, so the conformational changes must be considered in the context of a full NRPS. Within a module, an outside face of the C-terminal subdomain of the C domain forms a large binding interface with the A domain (Tanovic et al., 2008). The open or closing of the C domain would not abolish this C – A interface, though subtle changes within the contact area might occur. We speculate that conformational changes such as the opening of the C domain may be part of a communication network in NRPSs by which the functional state of one domain is conveyed to other domains. This may include sensing of the A_{sub} domain position,

which forms part of the CA interface in the adenylation state but rotates away in the thiolation state (Gulick et al., 2003; Mitchell et al., 2012; Reger et al., 2008; Sundlov et al., 2012; Tanovic et al., 2008), or PCP domain position, which has been observed bound at the acceptor site of an open C domain (Tanovic et al., 2008). Such a communication network could contribute to appropriate timing and coordination of the many reactions in the NRPS synthetic cycle, increasing the efficiency and rate of small molecule synthesis (Hahn and Stachelhaus, 2006; Hur et al., 2009). Confirmation of this theoretical network would require multiple structures of intact NRPSs and accompanying biochemical analyses.

3.2.6 The "latch" of the C domain

There are two points of crossover where a segment from the C-terminal subdomain crosses over to form part of the N-terminal subdomain. Residues ~295–309 cross over and form a small alpha helix, which backs against N-terminal subdomain helices h3 and h4. Residues ~367–388 cross and donate a beta strand to the major beta sheet in the N-terminal subdomain (formed of strands s1, s4, s5, s6) (Figure 3.2 and Figure S3.7a). This segment has been described as a "latch", forming a "roof" of the active site, and has been proposed to disengage from the N-terminal subdomain during the reaction cycle (Samel et al., 2007). However, in all structures of C-domains, the crossover "latch" interaction is intact (Figure S3.7). Furthermore, the interaction remains intact throughout every NMA, MD and morph simulation performed here (Movies S3.1-S3.11). Buried surface area calculations suggest that it is possible that the latch, at least in some C domains, could remain intact: In CDA-C1, SrfAC, TycC and VibH the latch buries 954 $Å^2$, 903 $Å^2$, 641 $Å^2$ and 927 $Å^2$ of surface area respectively, where interface on the order of 700 $Å^2$ is known to support heterodimer formation (Amit et al., 1986; Janin and Chothia, 1976). Finally, Samel *et al*. (Samel et al., 2007) argue that high B factors in the loop suggest that it could be a mobile element, but the latch residues in our structures display slightly lower than average B factors (average B factor for latch: 20.8 $Å^2$; average B factor for structure: 24.1 $Å^2$). Conceptually, it seems reasonable that the latch would open, because if it did not, the growing peptide chain would need to be

threaded through each C domain active site. However, we do not feel that there is at this time compelling evidence to support latch opening.

3.2.7 An active site tunnel and transition state model

As no structure of a C domain with substrates has been determined, the precise binding site and approach of the PPE-bound substrates are not definitively known, but can be reasonably guessed (Keating et al., 2002; Samel et al., 2007; Tanovic et al., 2008). The donor and acceptor substrates must approach and bind from opposite faces and meet at the active site H157. If the latch is not open, as is the case in all structures and models to date, the active site can be described as being at the center of a tunnel in the middle of the C domain (Figure 3.6). We identified this tunnel visually and also by using the program Caver (Chovancova et al., 2012). This tunnel, formed partially by the latch, stays intact through the morphing and MD simulations, and in all crystal structures. The tunnel is ~30 Å long and wide enough to accommodate substrates. Indeed, the tunnel entrances correspond to what is known about the binding sites for the upstream and downstream carrier protein domains. The SrfAC module has a substrate-less PCP domain positioned with its PPE attachment site at a reasonable distance from the C domain's catalytic histidine, and positioned at the tunnel entrance shown in Figure 3.6b (Stachelhaus et al., 1998). The structure of the TycC didomain has an upstream PCP domain in an unproductive conformation, but it is generally positioned near the tunnel entrance shown in Figure 3.6c, showing that the upstream carrier protein domain can reach this site. To give a holistic view of the C domain at the point of NRPS peptide bond formation, we constructed a model of the transition state of the reaction, including the C domain, 2,3-epoxy-hexenoyl-ACP and serinyl-PCP domain (Figure 3.6). The starting position of the PCP domain was taken from the SrfAC model, while the ACP was docked, using the program HADDOCK, taking into account the residues on the carrier proteins required for productive donor binding data (Lai et al., 2006a, b; Trivedi et al., 2005). The pantetheinyl substrates were roughly positioned along the tunnel, and the whole complex was subject to multiple rounds of Cartesian coordinate energy minimization in the program CNS.

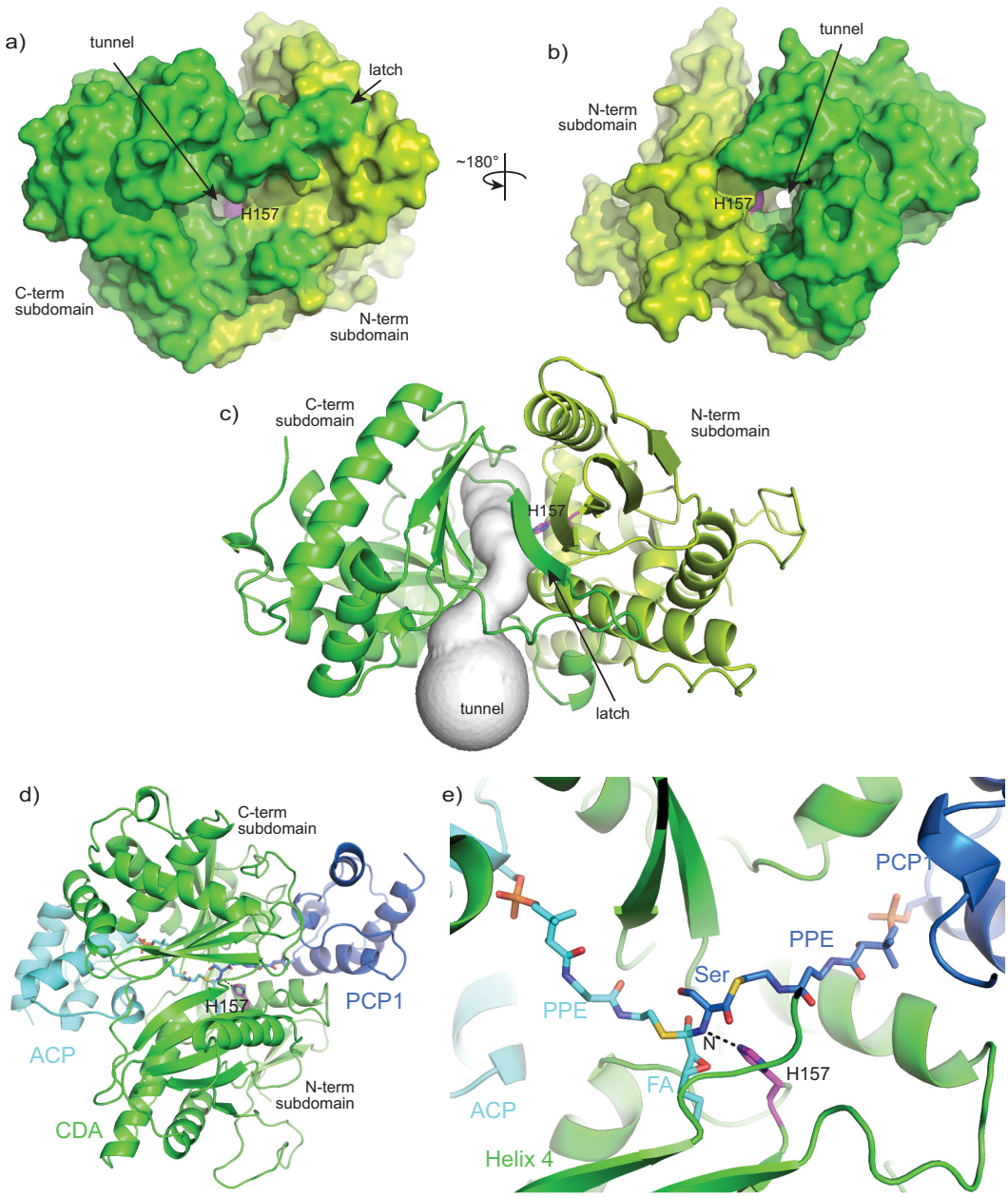


Figure 3.6. A tunnel to the active site of CDA-C1. a,b) Surface representations of CDA-C1 show a tunnel through CDA-C1, with the active site in its center. c) A representation of the tunnel from CAVER (Chovancova et al., 2012) output. d,e) A model of the transition state of the CDA-C1 catalyzed reaction including CDA-C1, ACP, PCP1 and covalently-attached transition state.

The model fits exactly into the described tunnel, with the nucleophilic alpha amino group of the serinyl-PCP domain modeled at 3 Å distance from the putative catalytic residue, as would be expected for the intermediate state of the reaction. The

fatty acid 2,3-epoxy hexanoyl side chain fits nicely into a pocket lined by α helix 4 and the major β sheet in the N-terminal subdomain, whereas the serine faces the C-terminal subdomain. Although precise analysis of substrate-C domain interactions awaits successful co-complex determination, this model will be useful in efforts to dissect the substrate specificity shown by C domains (Belshaw et al., 1999; Clugston et al., 2003; Ehmann et al., 2000; Kraas et al., 2012; Linne and Marahiel, 2000; Stachelhaus and Marahiel, 1995).

In summary, we have presented a structure of an active, previously undetermined NRPS condensation domain, which adopts a novel conformation. Conformational changes in C domains such as those modeled in our computational analyses are likely to occur in the catalytic cycle of NRPS C domains and may be important for peptide synthesis.

3.3 Methods

3.3.1 Cloning and expression of CDA-C1

The CDA-C1 construct was designed by aligning the sequence of first C domain of CDA peptide synthetase I of *Streptomyces coelicolor* A3(2) (NCBI NP_627443.1) to C domains of known structure (Figure S3.1) (Altschul et al., 1990). The CDA-C1 construct was synthesized by DNA 2.0, Inc. (Menlo Park, CA, USA), featuring an N-terminal octahistidine tag and tobacco etch virus (TEV) protease cleavage site of sequence MHHHHHHHHENLYFQG. Protein production was performed in BL21 (DE3) *E. coli* cells grown in LB medium supplemented by 300 μg ml^{-1} kanamycin (LB-kan) at 37°C. CDA-C1 expression was induced at an OD_{600} ~0.5-0.6 with the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and expression continued overnight at 16°C. Selenomethionine derivatized (SeMet) CDA-C1 was expressed in the same cell strain grown in kanamycin-containing M9 minimal medium supplemented with 0.2% glucose, 2 μM $MgSO_4$, and 0.1 μM $CaCl_2$. After reaching an OD_{600} of 0.5, the medium was further supplemented with the amino acids K, F, T (100 mg l^{-1} each), I, L, V (50 mg ml^{-1} each) and seleno-L-methionine (60 mg l^{-1}) (Van Duyne et al., 1993). Fifteen minutes after supplementation, cultures were induced with 1 mM IPTG and grown overnight at 25°C.

3.3.2 Protein purification of CDA-C1

CDA-C1 cell pellets were resuspended in a buffer of 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM β-mercaptoethanol (βME), 1 mM imidazole pH 8.0, and 1 mM phenylmethanesulfonyl fluoride (PMSF). Cells were lysed by sonication, and centrifuged for 30 min at 40 000 X g and 4°C. The supernatant was pooled and applied onto a 5 ml HiTrap IMAC FF column (GE Healthcare) charged with Ni^{2+}. CDA-C1 was eluted with 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM βME, 250 mM imidazole pH 8.0, and 1 mM PMSF; purity was assessed using SDS-PAGE. The fractions containing CDA-C1 were diluted 5-fold with a buffer of 50 mM Tris-HCl pH 7.5 and 2 mM β-mercaptoethanol, and applied to a MonoQ 5/50 GL column (GE Healthcare) and eluted with a gradient of 0–1 M NaCl. Relevant fractions were pooled and digested overnight at 4°C with N-His-TEV protease, using a ratio of 1 mg N-His-TEV protease per 40 mg CDA-C1. The cleaved sample was reapplied onto the 5 ml HiTrap IMAC FF column charged with Ni^{2+} and the flowthrough collected. The sample was brought to 1 M $(NH_4)_2SO_4$, applied to 2 x 1 ml HiTrap phenyl HP column (GE Healthcare) and eluted with a gradient of 1–0 M $(NH_4)_2SO_4$. After dialysis into 50 mM Tris-HCl pH 7.5, 100 mM NaCl and 2 mM dithiothreitol (DTT), CDA-C1 was concentrated to 10 mg ml^{-1}, flash-frozen and stored in liquid nitrogen.

3.3.3 Crystallization and data collection

Sparse-matrix crystallization trials of CDA-C1 and subsequent optimization identified two crystallization solutions that allowed growth of diffraction-quality crystals in sitting drop format using 2 µl of 10 mg ml^{-1} protein sample and 2 µl of crystallization solution in the drop and a 400 µl in the reservoir: (a) 25-27% PEG 3000, 0.2-0.25 M lithium sulfate, 0.1 M HEPES, pH 7.5, and (b) 9-13% PEG 10000 and 0.1 M HEPES pH 7.1-7.3. Crystals in condition (a) were directly mounted in cryoloops and flash-cooled in liquid nitrogen, while crystals in (b) were cryoprotected in a solution containing 20% PEG 10000 prior to mounting. SeMet-CDA-C1 crystals were grown in 26-31% PEG 3000, 0.2-0.26 M lithium sulfate and 0.1 M HEPES pH 7.5.

Data from SeMet-CDA-C1 crystals were collected at Se peak, inflection and remote wavelengths at X6A beamline at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory (BNL), Brookhaven, NY with a 0.5° frame width (Table S3.1). Native data sets were collected using a Rigaku RUH-3R rotating copper-anode source equipped with a R-AXIS IV++ image plate detector at McGill University, Montreal, Canada (GRASP) with a 1° frame width.

3.3.4 Structure determination

Data were indexed and scaled using the program *HKL-2000* (Otwinowski and Minor, 1997) The structure of SeMet-CDA-C1 was determined by a three-wavelength multiwavelength anomalous dispersion (MAD) phasing experiment using the program Phenix (Adams et al., 2010) (Table S3.1). The structure of CDA-C1 in the crystal form grown in the PEG 10000 – containing condition was determined by molecular replacement with SeMet-CDA-C1 as the reference model, using the program Phaser (McCoy et al., 2007). Structures were subjected to iterative rounds of modeling and refinement to give the final model (Table S3.1).

3.3.5 SAXS analyses

Purified CDA-C1 was subjected to gel filtration on Superdex200 10/300 GL column equilibrated with 25 mM Bis-TRIS pH 7.0 150 mM NaCl 10% glycerol. The most concentrated fraction (0.843 mg/ml) was used unaltered for SAXS data collection. The remaining fractions were pooled and concentrated on an Amicon(R) Ultra concentrator with 10kDa cut-off membrane to produce samples at 3.48, 9.82 and 23.04 mg/ml. Samples were filtered through a Millipore Ultrafree-MC VV 0.1 um filter prior to loading into SAXS cell. SAXS data was collected on an Anton Paar SAXSess MC2 CCD system on a PANalytical PW3830 generator with a Cu LFF tube (GRASP). Beam was collimated to 8mm length. Data points from 0.35 to 0.4 1/nm were covered by a CCD blemish and thus omitted. Primary data was processed and desmeared (Lake algorithm) using SAXSquant software. Scattering curves were merged using the program PRIMUS (Konarev et al., 2003). Theoretical scattering was calculated and fitted to experimental data using the program CRYSOL (Svergun et al., 1995). P(r) plots

were calculated with the program GNOM (Svergun et al., 1995). For *ab initio* shape reconstruction 50 models generated with the program DAMMIF (Franke and Svergun, 2009) were averaged with the program DAMAVER (Volkov and Svergun, 2003) assuming 2-fold symmetry. The resulting averaged and filtered envelope was superimposed onto crystal structure using the program SUPCOMB (Kozin and Svergun, 2001).

3.3.6 Expression and purification of enzymes required for reaction assay

ACP SCO3249 (ACP) (NCBI NP_627461.1, residues Met1-Ala89), DptA-PCP1 (PCP1) (NCBI WP_006122820.1, residues Asn929-Thr1030) and HxcO (NCBI NP_627459.1, residues Thr2-Pro600) constructs were designed based on the sequences reported by Marahiel and colleagues (Kopp et al., 2008; Kraas et al., 2012) and synthesized by DNA 2.0, Inc. (Menlo Park, CA, USA). ACP and HxcO both featured an N-terminal tag sequence MHHHHHHHHENLYFQG, and PCP1 featured a C-terminal tag sequence ENLYFQGHHHHHH. Protein production of HxcO, DptA and ACP was done as described for CDA-C1. Sfp from Bacillus subtilis was expressed and purified based on the protocol of Quadri *et al*. (Quadri et al., 1998).

The first step of purification of ACP, PCP1 and HxcO was nickel affinity chromatography as described for CDA-C1. HxcO was further purified using a Q Sepharose column (GE Healthcare) with a gradient of 0–1 M NaCl in a buffer of 50 mM Tris-HCl pH 7.5 and 1 mM tris(2-carboxyethyl)phosphine (TCEP). PCP1 and ACP were cleaved with TEV protease and reapplied to the 5 ml HiTrap IMAC FF column charged with Ni2+, then subjected to anion exchange chromatography using a 5 ml HiTrap Q Sepharose column (GE Healthcare) and a gradient of 0–1 M NaCl in a buffer of 50 mM Tris-HCl pH 7.5, 1 M NaCl and 2 mM βME. PCP1 was placed into a buffer of 25 mM HEPES pH 7.0, 50 mM NaCl, 1 mM TCEP, then concentrated and frozen. ACP was applied to a HiPrep 16/60 Sephacryl S-200 HR in buffer containing 25 mM HEPES pH 7.0 and 50 mM NaCl, then concentrated and frozen.

3.3.7 Assay for CDA-C1 activity

Assay for CDA-C1 activity was adapted from Kraas *et al.* (Kraas et al., 2012). ACP

(200 μM) was incubated for 60 minutes at room temperature with hexanoyl-coenzyme A (800 μM) (Sigma-Aldrich, Oakville, Canada) and Sfp (40.2 μM), in buffer containing 25 mM HEPES pH 7.0, 50 mM NaCl and 1 mM TCEP. HxcO (5 μM), FAD (25 μM) and HCl (3.75 mM) were added; the reaction was incubated for a further 60 minutes. In parallel, PCP1 was incubated at room temperature with serinyl-coenzyme A (800 μM) (Zamboni Chem Solutions, Montreal, Canada) and Sfp (40.2 μM) in buffer containing 25 mM HEPES pH 7.0, 50 mM NaCl and 1 mM TCEP for 20 minutes. At the completion of loading of the carrier domains, the reactions were mixed, CDA-C1 (6.2 μM) was added, and the solution was further incubated for one hour.

Separation of differentially loaded carrier proteins was performed using reverse phase chromatography (Varian, C4 4.6 x 250 mm Microsorb 300Å, 5 μM) at a flow rate of 0.5 ml/min using gradient elution from 40-50% B over 25 min, followed by a gradient up to 55% B over another 25 min (buffer A: 0.2% TFA with H_2O and buffer B: 0.2% TFA with acetonitrile). Protein masses were determined offline using ESI-MS (Esquire HCT Ultra; Bruker Daltonics) in positive-ion mode with an ESI nebulizer (Mass Spectrometry Core Facility, Bellini Life Science Complex, Montreal, Canada). The mass spectrometer was set to acquire spectra in the mass range 900-3000 m/z for an average of 3 minutes and the protein was infused through a syringe pump at 240 microliters per hour. Acquired spectra were averaged and the charge states of the protein was determined with the aid of a charge state ruler (Esquire Data Analysis; Bruker Daltonics).

3.3.8 Computational analyses

Models of CDA-C1 in the conformation observed in the structures SrfAC, TycC and VibH were produced to allow energy minimized refined linear interpolation, targeted molecular dynamics simulations and analysis of SAXS data using the SWISS-MODEL server (Arnold et al., 2006) in alignment mode. Energy minimized refined linear interpolations (“morphings”) were performed using the rigimol and refine functions in the program iPyMOL (Incentive PyMOL 2006 release, DeLano Scientific). Targeted molecular dynamics simulations were performed with the program NAMD (Phillips et al., 2005) using a hydrated sphere of 47.2 Å radius, a temperature of 310 K and an elastic constant TMDk of 200 kcal/mol/$Å^2$ for between 40 ps and 1 ns. Morphing and TMD was

performed between the CDA-C1 structure and models with sequence of CDA-C1 in the conformations of SrfAC, TycC and VibH, but note that the reciprocal transition using the structure of the more "open" C domains and a model with its sequence in the CDA-C1 conformation works equally well. Normal mode analyses were undertaken using the web-based programs WEBnm@ (Hollup et al., 2005) and NOMAD-Ref (Lindahl et al., 2006). Buried surface areas were calculated using the Lee & Richards buried surface accessibility calculation (Lee and Richards, 1971) implemented in the program CNS (Brunger, 2007). The active site tunnel was identified with the program CAVER (Chovancova et al., 2012), using the coordinates of H157 as the search seed.

For the model of the transition state of the C domain – catalyzed reaction, the server SWISS-MODEL was used to produce homology models of upstream ACP starting from the structure of ACP from *Staphylococcus aureus* (PDB ID 4DXE; Center for Structural Genomics of Infectious Diseases, unpublished), and the downstream PCP from the structure of the SrfAC module [10]. The initial position of ACP was derived from molecular docking using the program HADDOCK (de Vries et al., 2010), with outputs vetted by whether the resulting model agreed with experimental protein-protein contact data (Lai et al., 2006a, b). The initial position of PCP1 was achieved by superimposition of the SrfAC C domain on CDA-C1. Restraints parameters for the acyl-PPE transition state (TS) were generated using the program PRODRG (Schuttelkopf and van Aalten, 2004). Atoms of the TS were arbitrarily placed along the tunnel, and the entire system was subjected to conjugate gradient minimization with no experimental energy terms in the program CNS, to yield the final holo ACP–C–PCP1 model.

3.3.9 Accession numbers

The coordinates and structure factors for SeMet-CDA-C1 in $P2_12_12_1$ space group and CDA-C1 in $P2_1$ space group have been deposited in the PDB with accession codes 4JN3 and 4JN5.

3.3.10 Acknowledgements

We are indebted to Vivian Stojanoff and Edwinto Lazo for data collection at beamline X6A of NSLS, Brookhaven National labs, and to Kurt Dejgaard for advice and

training in mass spectrometry, as well as Janice Reimer for purification of Sfp protein, Michael Tarry, Diego Alonzo and Fabien Bergeret for advice and discussion, and Albert Berghuis for advice, critical reading of the manuscript and support of DR. The plasmid containing Sfp was a kind gift of Nathan Magarvey (McMaster University). We thank Alex Deiters, Qingyang Liu and Yan Zou (North Carolina State University) for preparation of SNAC analogues. We are indebted to John Colucci, Helmi Zaghdane and Robert Zamboni (Zamboni Chem Solutions) for an extremely generous gift of serinyl-CoA. This research was supported by the Canadian Institutes of Health Research (CIHR) Operating Grant No. MOP 106615 awarded to TMS, a Human Frontiers Science Program Organization Career Development Award to TMS, a Tier 2 Canada Research Chair in Macromolecular Machines held by TMS and a CIHR Strategic Training Initiative in Chemical Biology studentship held by KB.

3.4 Supporting information

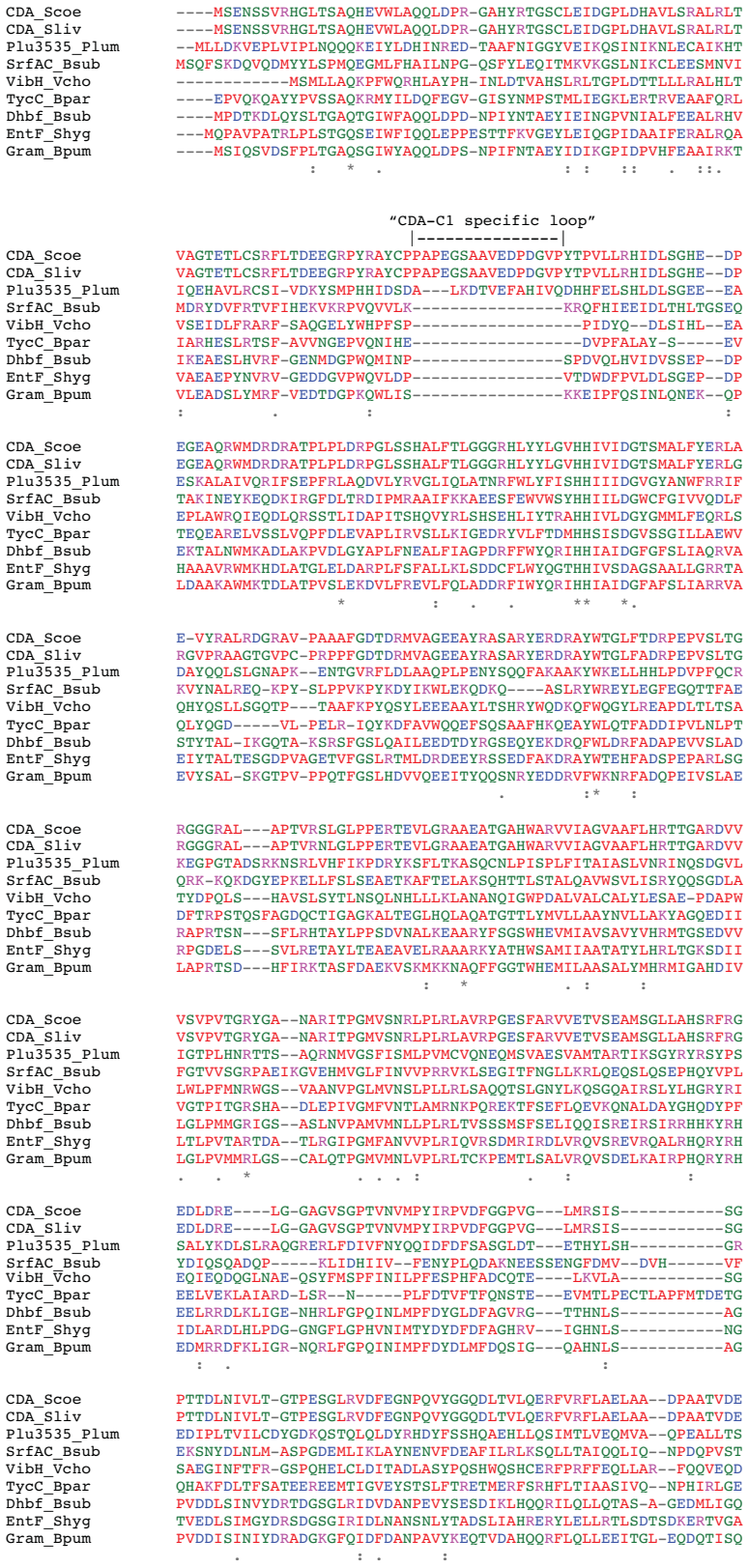


Figure S3.1. Sequence alignment of select C domains. An alignment of CDA-C1 with its orthologue from *Streptomyces lividans*, Plu3535, the three C domains with known structures (SrfAC, VibH, TycC), and three of the more homologous C domains as returned by BLAST [54] search (Dhbf, Entf, Gram).

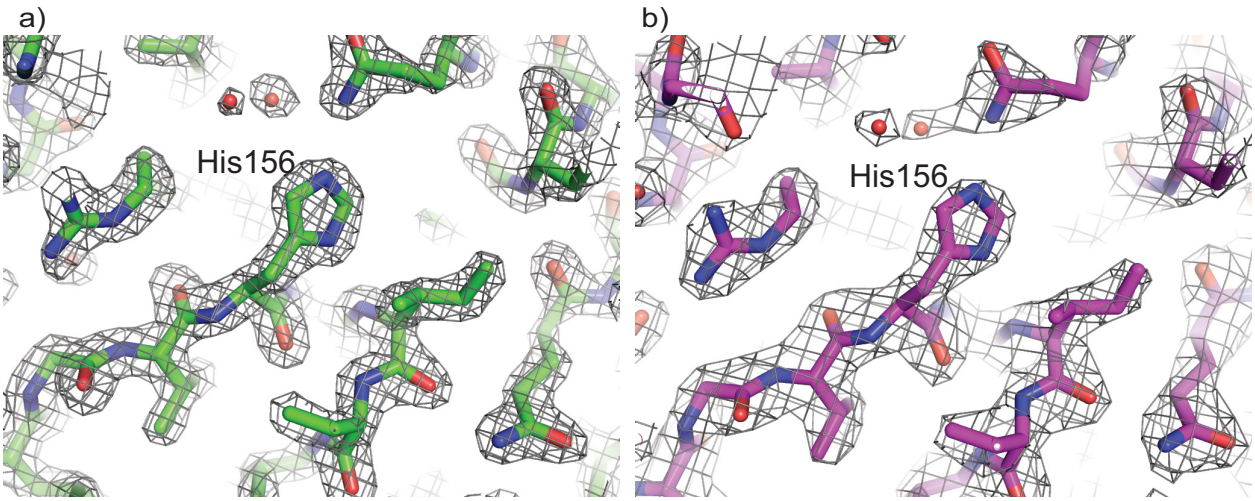


Figure S3.2. Electron density maps for the two crystal forms of CDA-C1. a) An F_O electron density map from the $P2_12_12_1$ crystal form contoured at 1σ. b) A $2F_O - F_C$ electron density map from the $P2_1$ crystal form, contoured at 1σ.

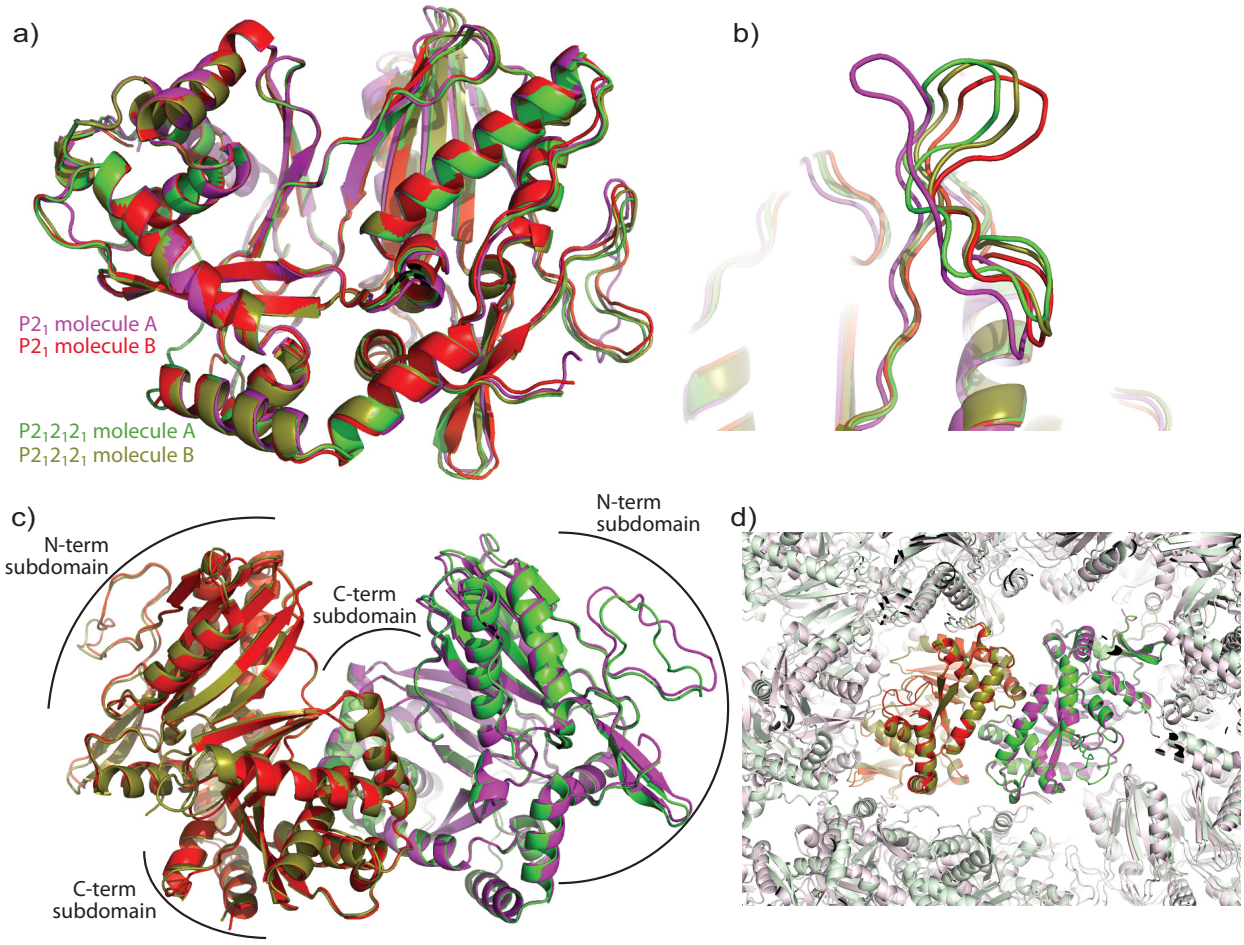


Figure S3.3. Comparison of the conformations of CDA-C1 in two crystal forms. a) Alignment of all four monomers in the asymmetric units of the $P2_12_12_1$ crystal form and the $P2_1$ crystal form. b) The conformations of a unique insert in the CDA-C1 sequence shows variable conformations in the four monomers. c) The dimers which are present in one asymmetric unit of the $P2_12_12_1$ crystal form and the $P2_1$ crystal form are in similar conformations. d) The packing of the dimers in the crystals are different for the $P2_12_12_1$ crystal form and the $P2_1$ crystal form.

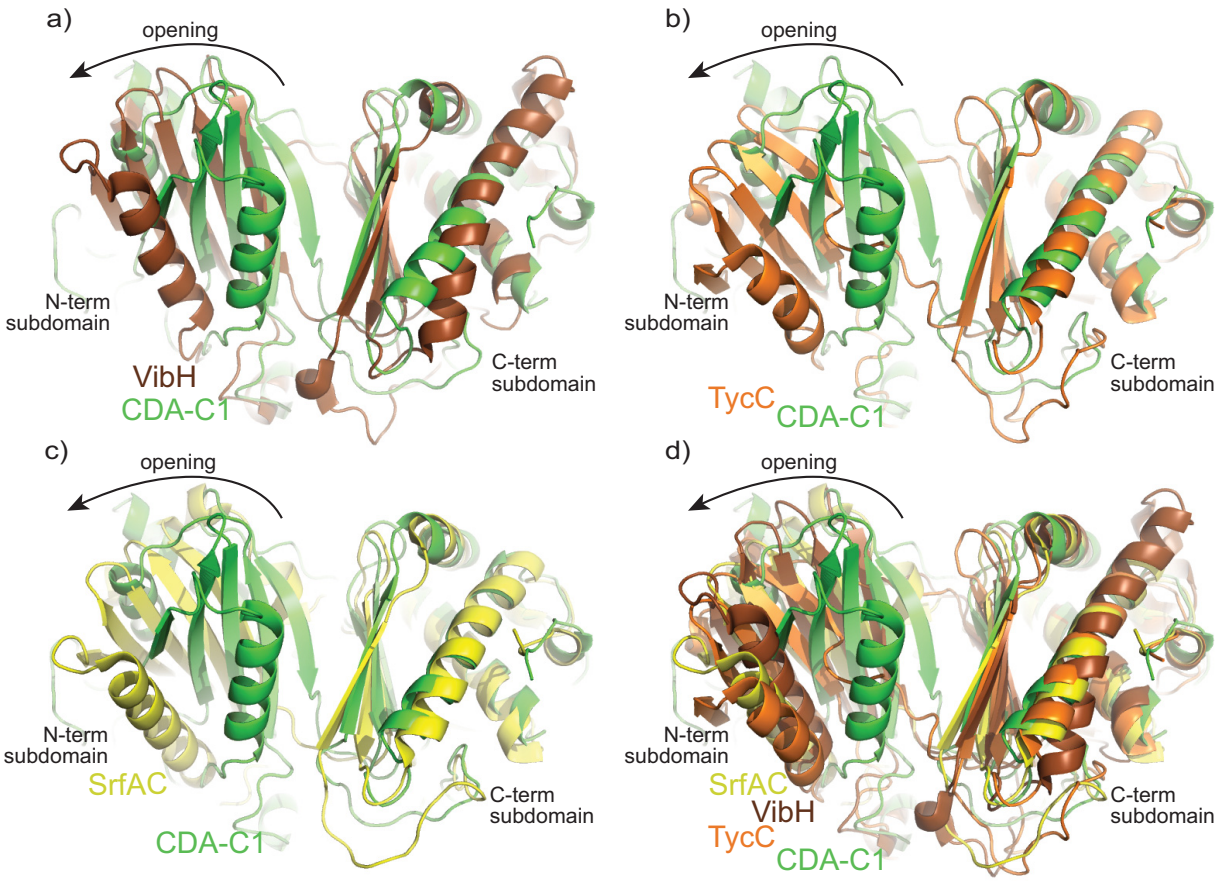


Figure S3.4. Individual comparisons of CDA-C1 with other C domains. Alignment of the C-terminal subdomain shows different relative subdomain–subdomain orientation in the structures of CDA-C1 (green) and those of a) VibH (brown), b) TycC (orange) and c) SrfAC (orange).

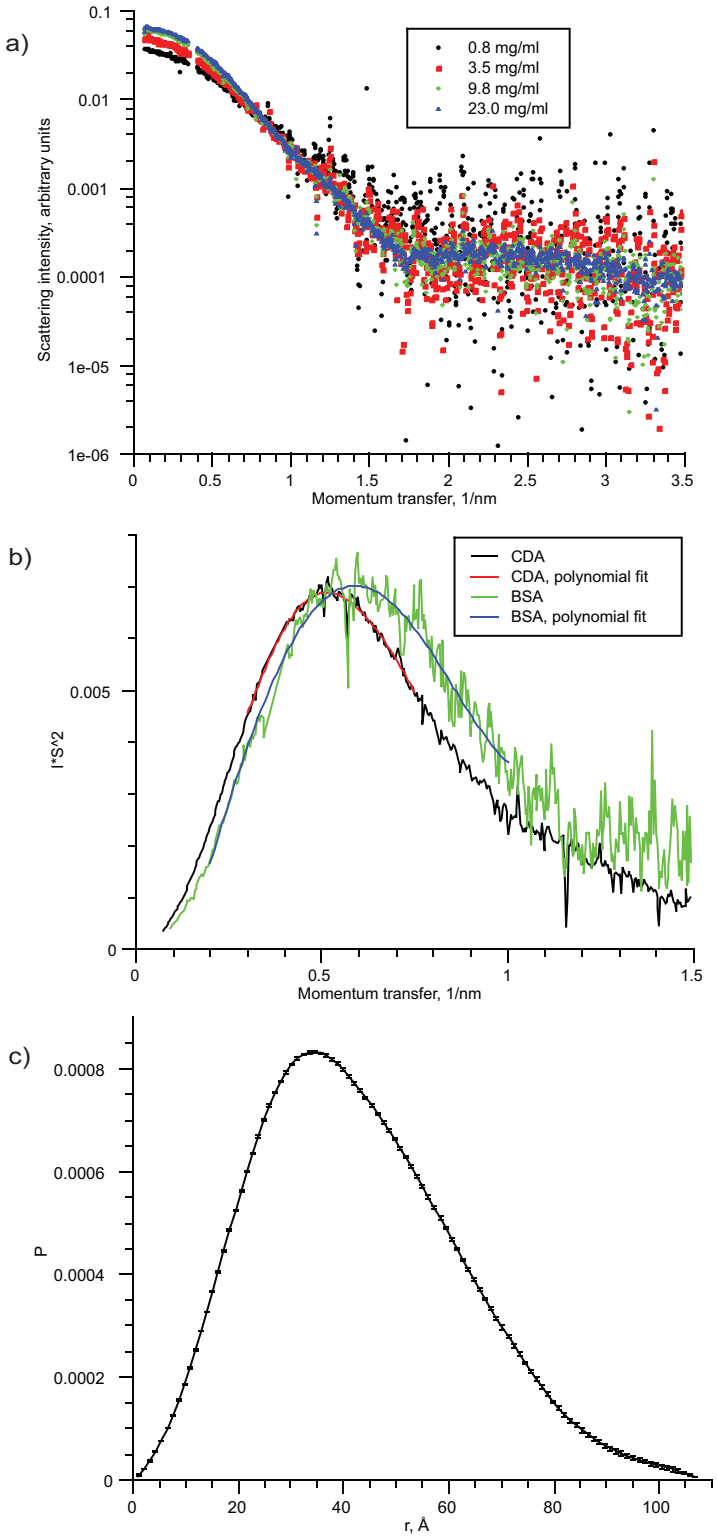


Figure S3.5. Additional SAXS data for CDA-C1. a) Normalized scattering curves of CDA-C1 at different concentrations showing little concentration dependence. b) Kratky plots of CDA-C1 and BSA indicating that CDA-C1 is a dimer. c) P(r) plot generated from merged data and used for envelope reconstruction.

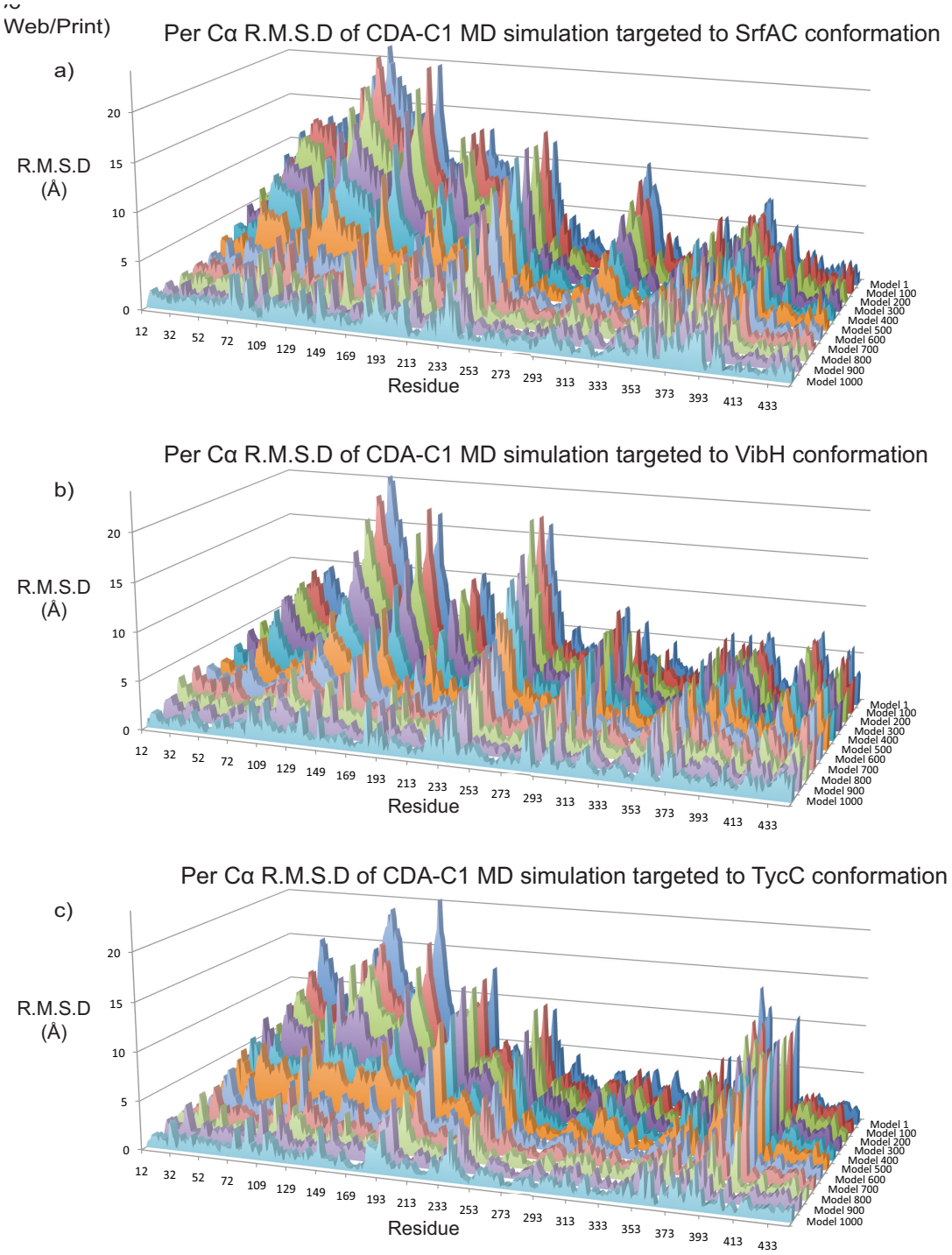


Figure S3.6. Per residue root mean square deviations during targeted molecular dynamics simulations. The root mean square deviation (rmsd) between Cα atoms of the model and target during targeted molecular dynamics simulations using a) SrfAC, b) TycC and c) VibH C domains as models. Simulations were 1 ns in length and rmsd values for every 100th model are shown. Residues ~295-309 and ~367-388 cross over to the N-terminal subdomain.

	Native	MAD (SeMet)		
		Peak	Inflection	Remote
Diffraction Data				
Wavelength (Å)	1.5418	0.97860	0.9792	0.9070
Space Group	$P2_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a=62.13, b=82.438, c=87.732; β=106.3	a=62.788, b=83.477, c=177.588	a=62.786, b=83.492, c=177.594	a=62.786, b=83.508, c=177.615
Resolution (Å)	50.00-2.44 (2.47-2.44)	50.00-1.80 (1.83-1.80)	50.00-1.80 (1.83-1.80)	50.00-1.80 (1.83-1.80)
$\langle I/\sigma(I)\rangle$	13.9 (2.6)	14.8 (2.5)	17.8 (2.4)	16.7 (2.0)
Temperature (K)	100	100	100	100
Measured Reflections (*n*)	119278	528361	550502	517345
Unique Reflections (*n*)	31620	87173	87248	87717
Completeness (%)	98.5 (82.1)	99.9 (99.8)	100.0 (100.0)	100.0 (99.9)
Multiplicity	3.8 (3.5)	6.1 (5.5)	6.3 (5.9)	5.9 (5.5)
R_{merge}[a] (%)	10.1 (49.1)	10.7 (70.2)	10.3 (75.7)	10.7 (81.2)
Sites (*n*)	-	8	8	8
Refinement Statistics				
Resolution range (Å)	33.91-2.44		34.11-1.69	
R-factor/R_{free} (%)[b,c]	19.8/25.0		19.5/21.7	
Waters	292		1230	
Average B-Factor (Å^2)	34.80		25.50	
R.M.S.D in bond lengths (Å)	0.0024		0.0054	
R.M.S.D in bond lengths (°)	0.654		0.945	
Ramachandran Plot (%)				
Favoured	97.6		98.4	
Allowed	2.14		1.24	
Outliers	0.24		0.34	

[a] $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th measurement and $\langle I(hkl)\rangle$ is the weighted average of all measured reflections.

[b] $R - factor = \sum \| F_{obs}(hkl)| - |F_{calc}(hkl) \| / \sum |F_{obs}(hkl)|$

[c] R_{free} is the R-factor computed for the test set of reflections omitted from the refinement process.

Table S3.1. Data collection and refinement statistics for the $P2_12_12_1$ and $P2_1$ crystal forms of CDA-C1.

Protein	Theoretical Mass (Da)	Observed Mass (Da)
Apo-ACP	8471.4	8471.0
Hexanoyl-ACP	8909.9	8909.1
2,3-expoxyhexanoyl-ACP	8923.9	8924.1
Holo-ACP	8830.8	8809.1
Apo-DptA-PCP1	12077.7	12077.3
Holo-DptA-PCP1	12419.0	12417.4
Seryl-DptA-PCP1	12506.1	12505.3
2,3-epoxyhexanoyl-Seryl-DptA-PCP1	12618.3	12615.4

Table S3.2. Identification of HPLC peaks by mass spectrometry. Deconvoluted masses of differentially-loaded carrier proteins identified from the HPLC assay.

Movie S3.1-S3.3. Normal mode analyses. Normal modes calculated using WEBNM@ (Movie S3.1) and NOMAD-Ref (Movie S3.2-S3.3) show conformational changes similar to what would be required to interconvert “closed” and “open” forms of the C domain.

Movie S3.4-S3.7. Morphing between observed C domain conformations. Energy minimized linear interpolation (“morphing”) of CDA-C1 between conformations observed in the crystal structures of CDA-C1 and SrfAC (Movie S3.4), CDA-C1 and TycC (Movie S3.5), CDA-C1 and TycC (Movie S3.6), and between all four conformations (Movie S3.7). The CDA-C1 insert is removed for these computations. The active site is colored red.

Movie S3.8-S3.11. Targeted molecular dynamics simulations between observed C domain conformations. Targeted molecular dynamics simulations of CDA-C1 between conformations observed in the crystal structures of CDA-C1 and SrfAC (Movie S3.8), CDA-C1 and TycC (Movie S3.9), CDA-C1 and TycC (Movie S3.10), and between all four conformations (Movie S3.11). The CDA-C1 insert is removed for these computations. The active site is colored red.

3.5 Bridge to Chapter 4

Although the structure of CDA-C1 produced a fascinating insight into the possible conformational dynamics during the synthetic cycle, I was still seeking to gain insights into the reaction mechanism and specificity determinants of the C domain. In order to do this, I would need to gain a structure of a C domain in complex with substrate analogues. Our collaborators at North Carolina State University (Yan Zou and Alex Dieters) produced several variations of aminoacyl-PPE and aminoacyl-SNAC analogues. However, all soaking and co-crystallization experiments failed, presumably due to the low binding affinity between the C domain and the analogues.

In order to solve this problem, I took a chemical biology approach to reach my goal in Chapter 4. This involved taking the transition state model I produced in Chapter 3, identifying a reasonable residue near the active site to mutate to cysteine, and design an acceptor substrate analogue that can alkylate that cysteine.

CHAPTER 4: CHEMICAL PROBES ALLOW STRUCTURAL INSIGHT INTO THE CONDENSATION REACTION OF NONRIBOSOMAL PEPTIDE SYNTHETASES

Bloudoff K, Alonzo DA, and Schmeing TM (2016). Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases. *Cell Chemical Biology* 23(3):331-339.

4.1 Introduction

Nonribosomal peptide synthetases (NRPSs) are a family of elegant macromolecular machines that assemble acyl substrates into interesting bioactive secondary metabolites (Schwarzer et al., 2003; Walsh, 2004; Weissman, 2015). Nonribosomal peptides have important and diverse activities; they include antifungals (bacillomycin), antibacterials (daptomycin), antivirals (luzopeptin), antitumors (actinomycin D), siderophores (enterobactin), and immunosuppressants (cyclosporin) (Felnagle et al., 2008). These compounds occupy a huge area of chemical space, because NRPSs can use over 500 different acyl monomer substrates, including L-amino acids, D-amino acids, aryl acids, fatty acids, hydroxy acids, and keto acids, and can further modify these moieties during peptide synthesis (Caboche et al., 2008; Hur et al., 2012).

The condensation (C) domain is responsible for linking these diverse substrates into nonribosomal peptide products during thio-templated synthesis (De Crecy-Lagard et al., 1995; Stachelhaus et al., 1998). In each NRPS elongation module, the substrates are delivered to the C domain as thioesters linked to the prosthetic phosphopantetheinyl (PPE) arm of carrier protein domains (peptide carrier protein (PCP) domains, or sometimes acyl carrier protein (ACP) domains). The C domain catalyzes amide bond formation between the donor aminoacyl, peptidyl or acyl group and the acceptor amino acyl-PCP, which transfers the donor group to the acceptor substrate, elongating the peptide chain (Figure 4.1a).

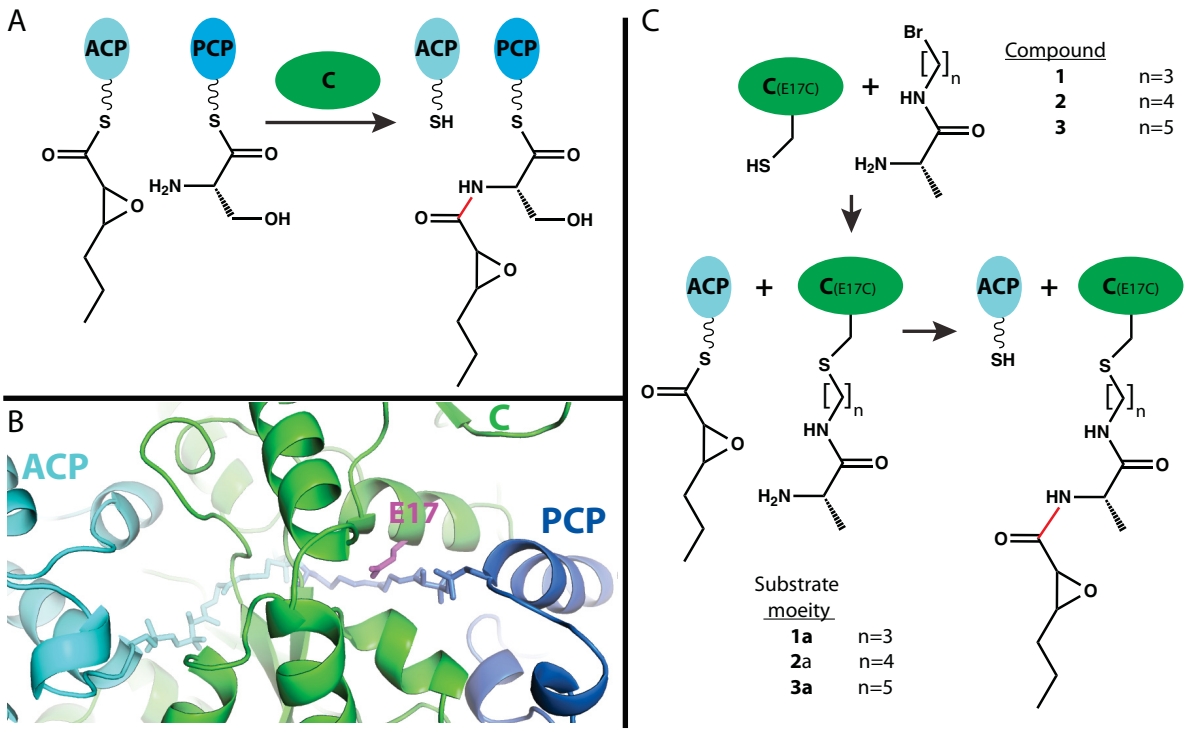


Figure 4.1. Schematic diagrams of (a) condensation by CDA-C1, (b) the chemical biology strategy presented here and (c) the transition state model, which suggests residue 17 as tethering point.

The newly elongated peptidyl-PCP domain then moves to the C domain of the next module, where it will donate the peptide chain in the next condensation reaction.

The C domain thus performs the central synthetic reaction in NRPSs. Although substrate selection occurs in the adenylation (A) domain, C domains can show specificity at both donor and acceptor substrates which could prevent peptide synthesis in bioengineered systems (Belshaw et al., 1999; Clugston et al., 2003; Ehmann et al., 2000; Kraas et al., 2012; Linne and Marahiel, 2000; Rausch et al., 2007; Stachelhaus and Marahiel, 1995). Bioengineering of NRPSs is attractive because their conceptually straightforward synthetic strategy should mean that domain- or module-swapped NRPS generate predictable new products. However, successful reengineering of NRPS is typically accompanied by a marked decrease in peptide yield (Calcott and Ackerley, 2015; Calcott et al., 2014; Kries et al., 2015). Despite the importance of this reaction for bioengineering efforts to produce novel compounds and for advancing fundamental understanding of NRPSs, the specificity determinants are not clear, and the chemical

mechanism is under debate. C domains possess a well-conserved HHxxxDG motif at their active site (De Crecy-Lagard et al., 1995). Mutational analysis showed that substitution of the second histidine or the aspartic acid in some C domains eliminated enzyme activity, leading to the proposal that this histidine acts as the general base, perhaps in a catalytic dyad (Bergendahl et al., 2002; Gaudelli et al., 2015; Roche and Walsh, 2003; Samel et al., 2007; Stachelhaus et al., 1998). However, in other C domains substantial activity is retained in mutants of the second histidine (Keating et al., 2002; Marshall et al., 2002b; Roche and Walsh, 2003; Samel et al., 2007). Along with theoretical pKa calculations, this data was used to suggest that it does not act as a general base, and an alternative, transition-state stabilization role was proposed (Samel et al., 2007).

The lack of structural information of C domains with their substrates hinders our understanding of the peptidyl transferase reaction. The published crystal structures show that the C domain is a V-shaped pseudo dimer that has its histidine motif in the middle of a central tunnel (Bloudoff et al., 2013; Drake et al., 2016; Haslinger et al., 2015; Keating et al., 2002; Samel et al., 2007; Tanovic et al., 2008). The structure of the C-A-PCP-Te termination module (Tanovic et al., 2008) identifies the binding site of the acceptor PCP, and the approximate position of the donor PCP can be modeled (Bloudoff et al., 2013; Samel et al., 2007). However, none of these structures include amino acyl- or peptidyl-PPE arm bound to the C domain, and so provide limited insight into catalytic mechanism or substrate specificity. Soaking or co-crystallization experiments with small molecule analogues of large substrates have provided excellent catalytic information in many other systems, but experiments with the small molecule analogues for the C domain, amino acyl SNACS (Belshaw et al., 1999; Ehmann et al., 2000), have failed to result in co-complex structures (Bloudoff et al., 2013; Samel et al., 2014), likely because for many C domains, the affinity for the substrates is too low when the PCP domain - C domain interaction is not present.

Here we present a chemical biology approach to tether reaction-competent acceptor substrates near the active site of the first C domain of calcium dependent antibiotic synthetase (CDA-C1). CDA-C1 catalyzes amide bond formation between the donor 2,3-epoxyhexanoyl ACP and the accepter serine PCP during synthesis of this

acidic lipopeptide in *Streptomyces coelicolor* (Hojati et al., 2002; Hopwood and Wright, 1983; Kopp et al., 2008). CDA-C1 is an ideal protein to use for this study, as its structure is known, it readily formed well-diffracting crystals in several space groups (Bloudoff et al., 2013) and it retains activity with its stand-alone donor ACP when excised from the full synthetase (Bloudoff et al., 2013; Kraas et al., 2012). Indeed, in this study it allowed structural determination of the C domain in complex with an acceptor substrate analogue and enabled insight into the catalytic mechanism and substrate specificity of the condensation reaction.

4.2 Results

4.2.1 A tethering approach to deliver C domain substrates

To structurally study the condensation reaction without full aminoacyl-PCPs, we tethered small molecule aminoacyl substrates near the active site of the first C domain of the calcium dependent antibiotic synthetase (CDA-C1). Tethering presents the substrates to the C domain active site at high local concentration and mimics covalent delivery by the acceptor PCP domain (Figure 4.1b). The tethering mechanism is based on alkylation of a thiol by alkyl halides (Armstrong and Lewis, 1951; Chowdhury et al., 2013) and requires a cysteine for ligation. We identified potential sites of tethering using a model we constructed of the CDA-C1 condensation transition state which includes C domain, acceptor amino acyl-PCP and donor fatty acyl-ACP (Figure 4.1c) (Bloudoff et al., 2013). Glu17 is situated in the channel between the acceptor PCP site and the histidine motif and is directly adjacent to the modeled PPE arm. Site directed mutagenesis was used to introduce a cysteine in this position (CDA-C1(E17C)). After our study was performed, the structure of the AB3403 C-A-PCP-TE termination module was published, which contained the PPE arm (without amino acid) of its PCP domain directed into the C domain active site (Drake et al., 2016). That PPE arm follows the same trajectory as our modeled PPE arm. Acceptor substrate analogs were commissioned for a custom synthesis, where alanine was coupled with bromoalkylamines to produce **1**, **2** and **3**, with varying linker length between the reactive bromine and the amino acid (Figure 4.1c).

4.2.2 Mass spectrometry shows tethering and condensation

CDA-C1(E17C) was incubated with either **1, 2, or 3,** and electrospray mass spectrometry (ESI) was performed on the intact protein (Figure 4.2a-c, respectively).

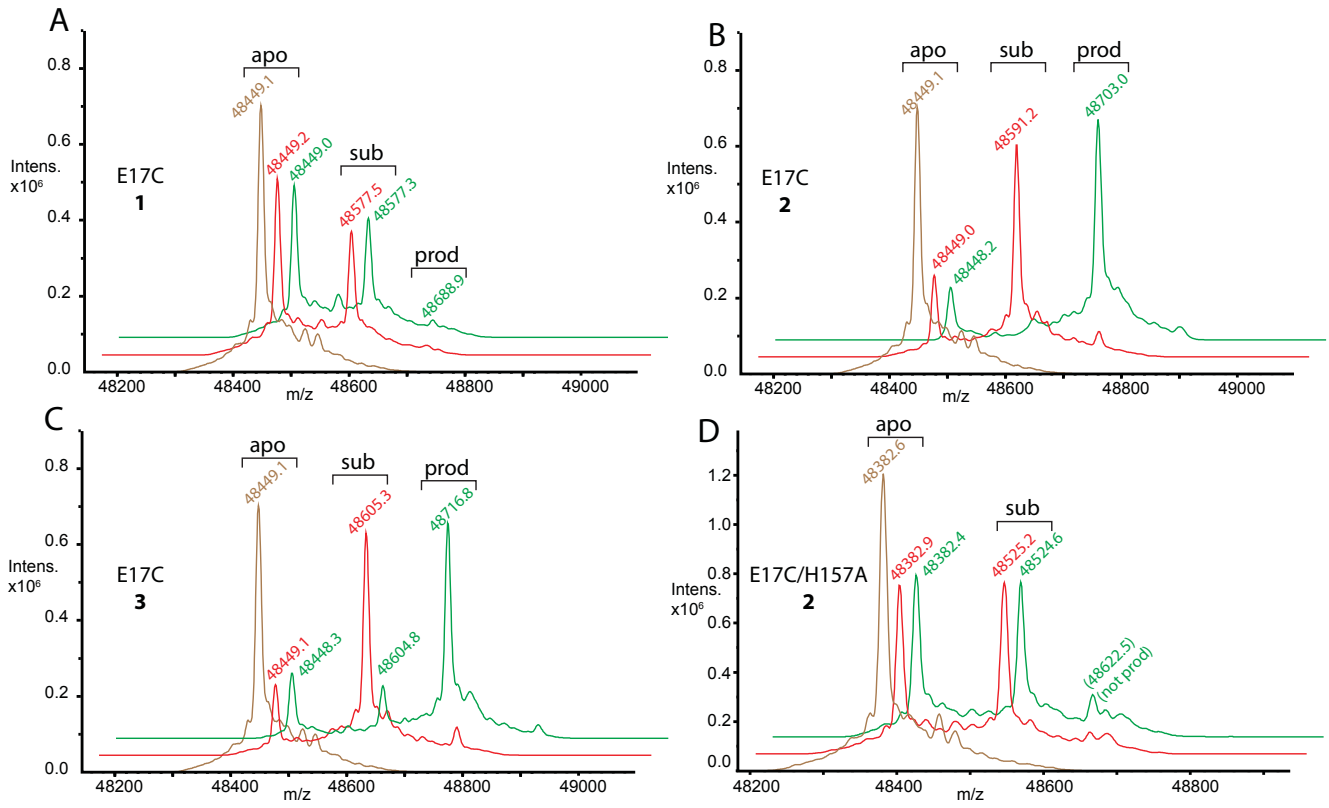


Figure 4.2. ESI-MS of intact CDA-C1 protein to assess the alkylation and condensation reactions. (brown: before reactions, red: after alkylation; green: after condensation). See Table S4.1 for expected masses.

When incubated with **2**, there was an increase in mass equal to that expected for a single alkylation with **2a** (Figure 4.2b). To verify the specific location of the modification, we performed in-source decay matrix-assisted laser desorption/ionization mass spectrometry (ISD-MALDI). This approach directly indicates the sequence of the protein and reveals the presence and exact position of modifications. ISD-MALDI proved that **2a** was specifically attached to residue C17 (Figure S4.1). As a control, we incubated WT CDA-C1 with **2**, and found very minimal modification by the probe (Figure S4.2a

and Table S4.1), emphasizing that **2** was highly specific for E17. Next, we incubated this CDA-C1(E17C)-**2a** complex with the full donor substrate, 2,3-epoxyhexanoyl-ACP (eFA-ACP) and assayed amide bond formation. Both ESI and ISD-MALDI showed that amide bond formation does indeed occur between 2a linked to C17 and the donor, 2,3-epoxyhexanoyl group (Figure 4.2c, Figure S4.1).

The alkylation and condensation with **2** (butyl linker) was compared to **1** (propyl) and **3** (pentyl). All three were able to alkylate CDA-C1(E17C), with **1** giving substantially lower alkylation (Figure 4.2a). Amide bond formation was also observed in each case, although **3a** was somewhat less efficient and **1a** gave only a very small product peak compared to **2a**, indicating a linker length of 4 carbons is optimal when the substrate analogue is linked at position 17.

Next, we produced a double mutation of E17C and H157A to assay our system without the second histidine in the active site motif, which is required for CDA-C1 condensation of full acceptor and donor substrates (Bloudoff et al., 2013; Kraas et al., 2012). Alkylation of CDA-C1(E17C;H157A) by **2** was observed, albeit with a reduced efficiency, but no amide bond formation occurred (Figure 4.2d), indicating that this minimal system behaves similarly to the full CDA-C1 condensation reaction in its requirement for H157.

4.2.3 Crystal structures of a C domain with tethered substrate

A large alkylation reaction of CDA-C1(E17C) was performed with **2,** the resulting complex was crystallized and its structure was determined at 1.6 Å resolution (Table S4.2). Clear density was observed continuous with C17 and could be easily fitted with **2a** (Figure 4.3a). The alkyl linker extends down the tunnel and allows the substrate amino acid to bind at the active site. The α-amino group of the substrate amino acid, which is the nucleophile in the condensation reaction, makes hydrogen bonds with the ε nitrogen of H157 in the catalytic motif and with the backbone carboxyl of S386.

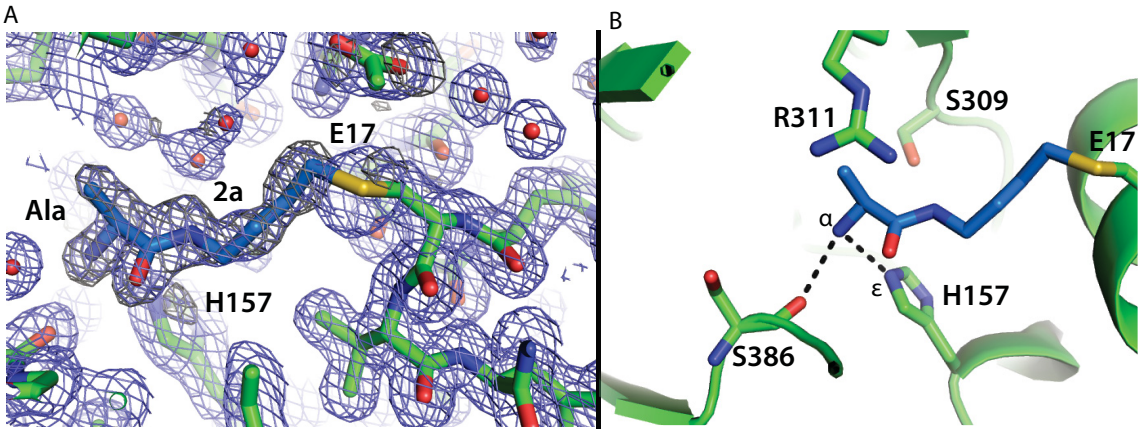


Figure 4.3. The crystal structure of CDA-C1(E17C)-2a. (a) Unbiased $2F_O$-F_C (slate; 1σ) and F_O-F_C (grey, 3σ) electron density maps. (b) The α-amino group of the substrate amino acid makes hydrogen bonds with the ε nitrogen of H157 of the catalytic motif and with the backbone carboxyl of S386. See Table S4.2 for crystallographic statistics.

We also performed the alkylation and crystallography with **3**. The substrate amino acid in **3a** binds at the same position, with the α-amino hydrogen bonding with the ε nitrogen of H157 and the backbone carboxyl of S386 (Figure S4.3). **3a** is marginally less efficient as an acceptor substrate, and the density for the alkyl linker portion is less defined than in **2a**. That **2a** and **3a** make the same interactions with H157 and S386, despite different linker lengths, strongly suggests that this conformation is representative of the native scenario (Figure S4.4).

4.2.4 Mutagenesis alters the specificity of CDA-C1

The determination of the co-complex of CDA-C1 with an acceptor substrate presented an opportunity to interrogate substrate specificity. CDA-C1(E17C) is bound with an alanine substrate, but the native serine side chain can easily be modeled, and could interact with S309 and/or R311 (Figure 4.3b). Because of this modeled interaction, we hypothesized that one of these residues can help impart specificity of the C domain for acceptor substrates. We analyzed the activity of wild type CDA-C1 and CDA-C1 with mutations of S309 and R311, using acceptor substrates with a variety of side chain sizes (serine-, alanine-, leucine and methionine-PCP), in an LC and MS-

based *in vitro* condensation assay (Figure 4.4, Supplemental Data 1) (Bloudoff et al., 2013; Kraas et al., 2012).

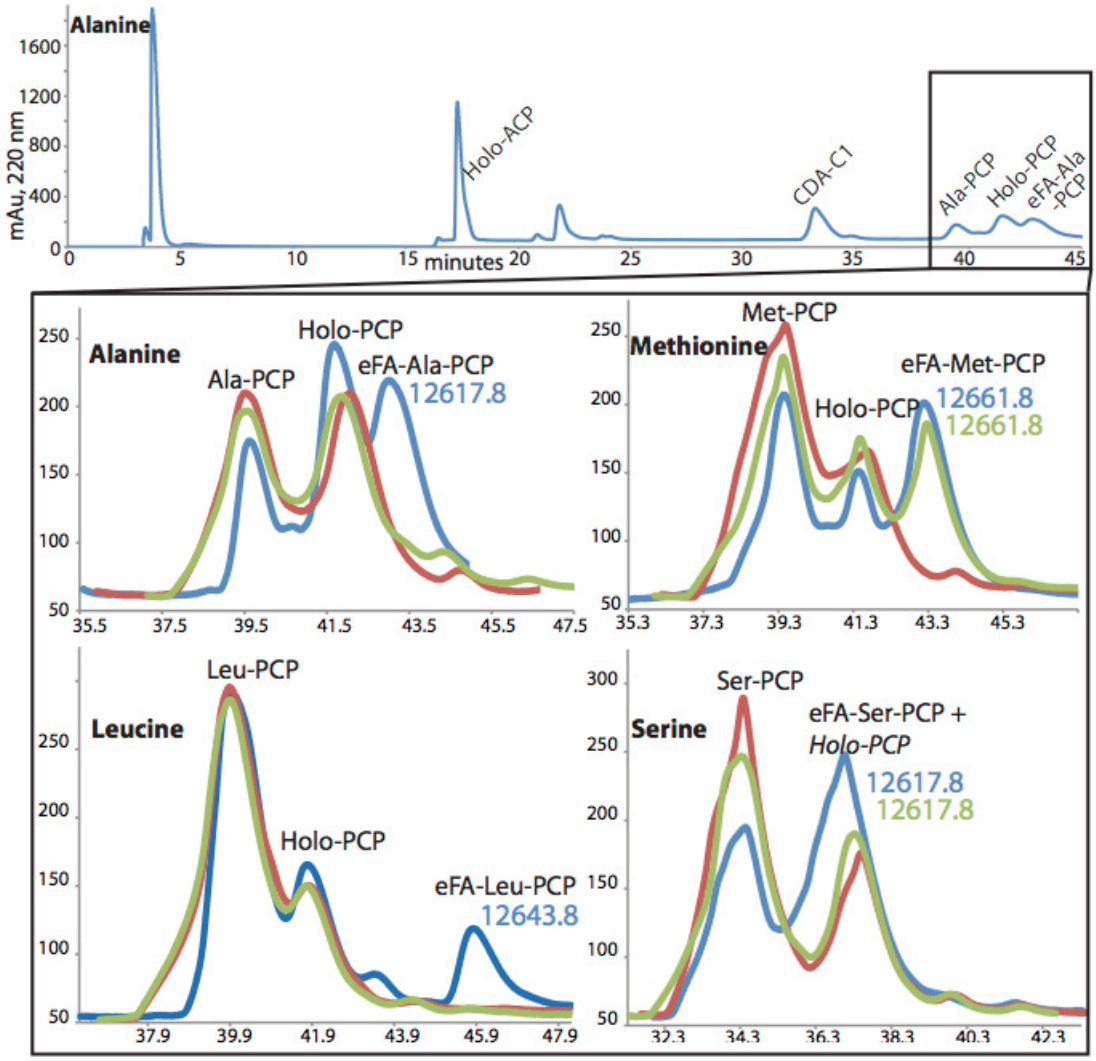


Figure 4.4. HPLC – MS assay of CDA-C1 and mutants with various acceptor aminoacyl substrates. (blue: wild type, red: S309V, green: S309G). Traces are overlaid by aminoacyl-PCP peak. Note that holo-PCP and eFA-Ser-PCP co-elute and MS is required to detect the presence of the eFA-Ser-PCP in this peak. See Table S4.3 for observed and expected masses and Supplemental Data 1 for all full HPLC traces and MS identification.

CDA-C1 was able to catalyze amide bond formation between 2,3-epoxyhexanoyl-ACP and each of the acceptor substrates assayed. The R311A mutant was also active with all substrates (unpublished observations). In contrast, the S309V mutation abolished

activity with all substrates assayed. Interestingly, another mutation at the same residue, S309G, allowed condensation reactions with serine- or methionine-PCP domain as the acceptor substrate, but not with alanine- or leucine-PCP. The S309G mutation thus imparts greater specificity to CDA-C1 for the acceptor substrate.

To investigate whether the high local concentration of the chemical probe could overcome some of the native specificity in CDA, we produced CDA-C1(E17C/S309V) and CDA-C1(E17C/S309G). We alkylated these double mutants with **2** and measured their ability to participate in the condensation reaction. CDA-C1(E17C/S309V) was not able to catalyze amide bond formation (Figure S4.2b). However, CDA-C1(E17C/S309G) was able to produce some condensation product (Figure S4.2c), although not nearly as much as CDA-C1(E17C), indicating the tethering can partially counteract inherent C domain specificity.

4.3 Discussion

Tethering of substrate analogs near active sites has enabled insight into the mechanisms or specificities of HIV RT (Huang et al., 1998), kinases (Maly et al., 2004), ribozymes (Xiao et al., 2008) and the ribosome (Green et al., 1998). Here, the covalent tethering of an active substrate analogue to its C domain allowed us to visualize the co-complex at high resolution and guided our experiments interrogating C domain acceptor substrate specificity. We used alkyl halides, as they are not as reactive as acyl halides, and we reasoned that the substrate's amino acyl moiety could help target it to the active site. Indeed, unless much higher compound concentration or much longer incubation time was used, substantial alkylation was observed only at C17, and not at other cysteines. Also, lower alkylation was observed when the linker length was shortened and with the H157A mutation (Figure 4.2), both of which could hinder targeting to C17.

A key reason for this study was to develop a new approach to provide insight into the condensation reaction in the C domain. Previous work has demonstrated that (residue numbering as in CDA-C1) His156, Asp161 and Gly162 are all known to be conserved for structural reasons, with His156 buried and Asp161 making a conserved salt bridge with Arg296 (Bloudoff et al., 2013; Keating et al., 2002; Samel et al., 2007; Tanovic et al., 2008), leaving only the His157 with a possible direct role in catalysis. In

some C domains, experimental mutation of the His157 abolishes catalysis (Bergendahl et al., 2002; Bloudoff et al., 2013; Roche and Walsh, 2003; Samel et al., 2007; Stachelhaus et al., 1998). Also, in the X domain (a catalytically inactive C domain used for protein-protein interaction), an Arg replaces the catalytic His, which could contribute to its lack of catalytic activity (Haslinger et al., 2015). However, in other C domains, the effect of mutating H157 is more modest, with as little as 2-fold difference (Keating et al., 2002; Marshall et al., 2002b; Roche and Walsh, 2003). These differential effects call into doubt the proposed role of H157 as a strong general base. An alternate theory was presented with the publication of a PCP-C didomain structure in the absence of substrates (Samel et al., 2007). This structure was used to calculate the theoretical pKa of the catalytic His at 11.8, which suggests that it is protonated at neutral pH and unable to extract a proton. The catalytic His in that structure is interacting with a buffer-derived sulfate ion, assumed to mimic the tetrahedral transition state, and it was proposed that the positively-charged His aids in transition state stabilization. However, these theoretical pKa calculations are not rigorous. Indeed, CDA-C1 will have the same catalytic mechanism, but the same calculation in the presence and absence of ligand found the theoretical pKa for H157 to be ~6.1-6.5 (Bashford and Karplus, 1990; Gordon et al., 2005). Furthermore, in our structure, H157 accepts a hydrogen bond from the substrate's α-amino group, which precludes H157 from participating in transition state stabilization.

The crystal structures presented here have allowed us to view the active site of CDA-C1 in presence of a reaction-competent acceptor substrate. These structures show that the α-amino group of the acceptor substrate hydrogen bonds with the ε nitrogen of H157 and the backbone carbonyl of S386 (Figures 4.3 and S4.4). These interactions suggest that the role of H157 in catalysis could principally be one of substrate positioning, a fundamental mechanism used in all enzymes (Jencks, 1969; Keating et al., 2002). The δ nitrogen of H157 is protonated and interacting with nearby backbone carbonyls, allowing the non-protonated ε nitrogen (along with the carbonyl of S386) to accept hydrogen bonds from the α-amino nucleophile, precisely orienting it for attack (Figure 4.3b, S4.3, S4.4). The neutral, reaction-competent $–NH_2$ form of the α-amino group may be present spontaneously at neutral pH, since the pKa of a model

thioester has been measured at ~7 (Anderson and Packer, 1974; Samel et al., 2007). Upon orientation of the $–NH_2$ nucleophile, reaction likely proceeds through a zwitterionic transition state to the amide product (Yang and Drueckhammer, 2000). A proton is lost from the amino group at some point during this pathway, but the mutational analyses performed by Keating et al., Roche et al. and Marshall et al. (Keating et al., 2002; Marshall et al., 2002b; Roche and Walsh, 2003) and the fact that the pKa of the alpha-amino grou[before reaction is extremely high (Green and Lorsch, 2002) dictate that the proton is not extracted by H157 before reaction. The proton may be facilely lost to solvent when the pKa of the alpha-amino group drops markedly as the reaction proceeds. It is possible that the His could accept this proton, but the minor effects of the His mutations (Keating et al., 2002; Marshall et al., 2002b; Roche and Walsh, 2003) indicate that would make only a small contribution to catalysis and that the proton can be transferred to solvent in absence of the His.

The proposed role of H157 in substrate positioning is attractive on several accounts. It is easy to rationalize that the interaction between H157 and the α-amino group of the acceptor substrate is absolutely critical in certain C domains, whereas other C domains may possess additional interactions that can partially compensate for the loss of this His, including with PCP domains, with backbone carbonyls like S386, or with the side chains of acceptor amino acids. Indeed, C domains in general have significant side chain selectivity and show a strong stereoselectivity for L-amino acids at the acceptor site (Clugston et al., 2003; Lautru and Challis, 2004; Tang et al., 2007; Yonus et al., 2008). Furthermore, it is worth noting that the other peptide bond-making macromolecular machine, the ribosome, likely uses substrate positioning and solvent-mediated proton extraction as its source of catalytic power (Kuhlenkoetter et al., 2011; Wallin and Aqvist, 2010)

The structure of CDA-C1 with tethered acceptor substrate also allowed us to identify, and confirm via mutagenesis, that S309 is involved with acceptor substrate specificity in CDA-C1. It was somewhat unexpected that while S309V completely abolished catalytic activity, the S309G mutation increased the selectivity of CDA-C1 towards acceptor substrates. The result can be explained by steric arguments, namely that the S309G mutation creates room which the methionine side chain (or serine and

water) fills, but desolvation of this area is unfavorable if a hydrophobic side chain is unable to make van der Waals contacts with the protein. However, this result, along with the fact that computational analyses of C domains have not been able to predict native substrates of a C domain from the active site sequence (M. Skinnider, N. Magarvey, and (Rausch et al., 2007)), might indicate that a loose specificity code similar in concept to the well-defined code in A domains (Rausch et al., 2005) is not achievable in C domains. Thus, the most promising strategy for module and domain-swapping initiatives (Baltz, 2014; Calcott et al., 2014; Duerfahrt et al., 2003; Kries et al., 2015; Mootz et al., 2002; Mootz et al., 2000; Schauwecker et al., 2000; Schneider et al., 1998; Stachelhaus et al., 1995) may be to find or experimentally develop a C domain with broad promiscuity, rather than to predict which substrates a C domain will use, based on the sequence of its active site.

Soluble small molecule substrate analogs such as aminoacyl-coenzyme A and aminoacyl-N-acetylcysteamine thioesters have been used to study the specificity of C domains (Belshaw et al., 1999; Ehmann et al., 2000; Zane et al., 2014). These molecules, like our chemical probes, do not have the full carrier protein-domain – C domain interactions that are present in the native case, but still are able to reveal important mechanistic features. However, these molecules are not competent for catalysis by excised CDA-C1 and did not bind to the crystallized protein (Bloudoff et al., 2013). Elegant covalent approaches were used to determine insightful A-PCP (Sundlov et al., 2012) and PCP-Te (Liu et al., 2011) didomain crystal structures. As well, an innovative click chemistry approach has been used to examine intermodular NRPS interactions via a covalent crosslink (Hur et al., 2009). In this approach, it was hypothesized that the specific crosslink could be forming within the TycB1 condensation domain substrate channel, thus mimicking the C domain during the transition state. However, it remains uncertain how faithful the mimicking is (Hur et al., 2009), and the co-complex is yet to be visualized.

The strategy presented here uses easily synthesized substrates in conjunction with a single site-directed mutation, preserves the native reactive atoms of the condensation reaction, and is simple to perform. We used the approach to visualize CDA-C1 in complex with a catalytically active acceptor substrate, which enabled a

targeted analysis of a residue whose mutation altered acceptor substrate specificity. In future studies, the use of similar chemical probes (for example, bromo-linked transition state analogues) will enable visualization of additional stages of the condensation reaction. The general approach could also be useful in other systems where substrate analogs have low affinity or are problematic, such as polyketide synthases, fatty acid synthases, chaperones and peptidoglycan-synthesizing enzymes.

4.4 Significance

Nonribosomal peptide synthetases are elegant macromolecular machines, which produce diverse peptides of biological and commercial significance. One of the crucial steps of nonribosomal peptide synthesis is the formation of a peptide bond between an incoming monomer and the growing peptide chain a step performed by the condensation domain. Although it plays a central role in peptide synthesis, the reaction mechanism and specificity determinants of the condensation domain remain unclear. Here, we have developed chemical probes that covalently bind to an engineered cysteine residue located near the active site of the first condensation domain of the calcium-dependent antibiotic synthetase (CDA-C1), thus mimicking native acceptor substrate delivery to the site by carrier domains. Using mass spectrometry, we verified that the covalently-bound acceptor substrate analogues were competent for condensation only when the active site histidine is present, making the system faithful to assays previously published. We determined the crystal structure of the condensation domain in complex with two similar chemical probes which suggest that the principal role of the active site histidine is to position the alpha amino group of the acceptor substrate for nucleophilic attack. In addition, the crystal structures lead to the identification of a mutation that altered the acceptor substrate specificity of CDA-C1. Further development of these chemical probes will lead to future studies interrogating other stages of the condensation domain reaction cycle, and can be developed for use in other systems with low binding affinity proteins.

4.5 Experimental Procedures

4.5.1 Cloning and expression of wild-type CDA-C1, E17C, E17C/H157A, S309V, S309G, E17C/S309V and E17C/S309G.

Site directed mutagenesis was performed on a plasmid containing the gene for CDA1-C1 using the QuikChange kit (Agilent) and the primers listed in Table S4.4. CDA-C1 protein (Bloudoff et al., 2013) was produced in BL21 (DE3) *E. coli* cells grown in LB medium supplemented with 300 μg ml^{-1} kanamycin at 37 °C. CDA-C1 expression was induced at an OD_{600} of ~0.5–0.6 with the addition of 1 mM IPTG, and expression continued overnight at 16 °C. CDA-C1 and mutational variants were lysed by sonication and centrifuged for 30 min at 40,000 g before being applied onto 2x5-ml HiTrap IMAC FF columns (GE Healthcare) charged with Ni^{2+}. After elution, CDA-C1 was applied and eluted from 2x5-ml HiTrap Q HP columns (GE Healthcare). CDA-C1 was cleaved overnight at 4 °C with N-His-TEV protease and reapplied to 2x5-ml HiTrap IMAC FF columns charged with Ni^{2+}. The flowthrough was pooled and applied to a 5 ml HiTrap phenyl HP column (GE Healthcare) and dialyzed into 50 mM Tris–HCl (pH 7.5), 100 mM NaCl and 1 mM Tris(2-carboxyethyl)phosphine (TCEP). Protein purity was assessed using SDS-PAGE.

4.5.2 Substrate syntheses

For Ala-, Met-, and Leu-CoA substrates, 16 mM coenzyme A sodium salt hydrate (Sigma-Aldrich) was incubated overnight with 128 mM Boc-protected L-amino acid–*N*-hydroxysuccinimide ester (Chem-Impex International Inc.) and 64 mM N,N-diisopropylethylamine in 1 mL, to form Boc-L-amino acid–CoA. Reactions were lyophilized and resuspended in 1.5 mL of a solution of 97% trifluoroacetic acid, 1.5% water and 1.5% triisopropylsilyl ether, and incubated at 25 °C for 2 hours while shaking at 750 rpm. Ice-cold diethyl ether (30 mL) was added and the solution incubated at -20 °C for another 2 hours. The solution was centrifuged at 4000 rpm for 20 minutes, and the pellet was resuspended in 5 mL of water and 0.1% TFA and immediately purified.

Purification took place by reverse phase chromatography, using a YMC-Pack ODS-A column (YMC America Inc., 250 mm x 20 mmI.D., 5 μm particle size, 30 nm pore size). At a flow rate of 40 mL min^{-1}, separation was achieved with a gradient of 0-

98 % B over 5 minutes (buffer A: 0.2% formic acid with H2O and buffer B: 0.2% formic acid with acetonitrile). Masses of -1 ions were determined for Ala-CoA as 837.134 (exact mass 837.15), for Met-CoA as 897.130 (exact mass 897.15) and for Leu-CoA as 879.184 (exact mass 879.19) using an Agilent QTOF 650 in ESI negative mode, a buffer of 50% acetonitrile and calibrant Agilent ESI-L G1969-8500 (exact mass 1033.988).

Serinyl-CoA and bromo-alkyl-alanine compounds **1, 2** and **3** were purchased from Zamboni Chemical Solutions (Montreal, QC). Alaninyl-PCP is as active in the condensation reaction as the native serinyl-PCP, and chemical synthesis of serinyl-amino alkyl-bromide failed, likely because of side chain hydroxyl reactivity.

4.5.3 Alkylation and Condensation Assay

In a reaction volume of 1000 µL, 52 µM CDA-C1(WT), CDA-C1(E17C), CDA-C1(E17C/S309V) or CDA-C1(E17C/S309G) was alkylated with 15 mM compound **1**, **2** or **3** in a buffer containing 100 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM TCEP, and 20% glycerol. The reactions were incubated at room temperature for 18 hours, and purified over a Superdex 200 10/300 GL size exclusion column (GE Healthcare) equilibrated in a buffer of 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, and 1 mM TCEP. Aliquots were taken for mass spectrometry to measure the extent of alkylation, then glycerol was added to 20% final before concentration for condensation reactions.

Condensation reactions, in a volume of 500 µL, were composed of 40 µM alkylated CDA-C1(E17C), CDA-C1(E17C/S309V) or CDA-C1(E17C/S309G), and 160 µM ACP loaded with 2,3-epoxyhexanoyl-fatty acid in a buffer containing 50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM TCEP, and 20% glycerol. Condensation reactions were incubated at room temperature for 5 hours before being separated on a 1 mL HiTrap Q HP column (GE Healthcare).

4.5.4 ESI-MS with intact CDA-C1

The intact CDA-C1 proteins and complexes were analyzed by LC-MS using a Dionex Ultimate 3000 UHPLC coupled to a Bruker Maxis Impact QTOF in positive ESI mode. Samples were separated on a Waters BEH300 C4 column (1.7 µM 300A 2.1 x

100 mm) using a gradient of 80% mobile phase A (0.1% formic acid in water) and 20% mobile phase B (0.1% formic acid in acetonitrile) to 0% mobile phase A and 100% mobile phase B in 15 minutes. The data was processed and deconvoluted using the Bruker Data Analysis software version 4.1.

4.5.5 MALDI-Top Down Sequencing

Alkylated protein samples (~3 mg/mL each) were desalted using 0.6 uL reverse-phased C4 ZipTips (EMD Millipore) according to the manufacturer's recommendations. Each desalted/concentrated sample was mixed 1:1 with super DHB matrix (SDHB, Bruker Daltonics; 50 mg/mL in 50% ACN/0.1% TFA) and applied to a ground or polished steel MALDI target using the dried droplet method.

Mass spectra were recorded on an ultrafleXtreme MALDI-TOF/TOF system (Bruker Daltonics, Bremen, Germany) using the reflector positive "ISD_1-8_kDa" method (ion source 1 = 25 kV; ion source 2 = 22.35 kV; lens = 8.85 kV; reflector = 26.45 kV; reflector 2 = 13.40 kV; pulsed ion extraction = 170 ns) and flexControl v1.4 and flexAnalysis v1.4 software. The intensity of ISD ions was evaluated by averaging three measurements of 4,000 shots each (12,000 shots total per sample). Sequences (with or without the Cys17 modifications; generated in Sequence Editor v3.2 (Klynt) software) were overlaid with ISD spectra in the BioTools v3.2 software to generate the depicted amino acid assignments.

4.5.6 Loading of ACP with 2,3-epoxyhexanoyl-fatty acid

ACP (120 μM) was incubated overnight at 30 °C with $MgCl_2$ (10 mM), hexanoyl-coenzyme A (800 μM) and His-Sfp (10 μM), in buffer containing 50 mM Bis-Tris (pH 6.5), 50 mM NaCl and 1 mM TCEP. His-HxcO (5 μM) and FAD (25 μM) were added, and the reaction was continued at room temperature for 7 hours. ACP-eFA was purified by running the reaction over a desalting column and nickel IMAC columns in tandem (GE Healthcare), into a final buffer of 25 mM HEPES pH 7.0, 50 mM NaCl.

4.5.7 Crystallography

CDA-C1(E17C)–**2a** and CDA-C1(E17C)–**3a** were crystallized by sitting drop vapour diffusion using an initial drop of 2 µl of 10 mg ml^{-1} protein sample and 2 µl of crystallization solution (consisting of 25–27% PEG 3000, 0.2–0.25 M lithium sulfate, and 0.1 M HEPES, pH 7.5) and a 400 µl reservoir. This crystallization condition produces crystals in $P2_12_12_1$ and R3:H space groups. Crystals of each complex in each space group were flash frozen and data were collected at Canadian Light Source beam line 08ID-1.

Data were indexed and scaled using the program mosflm (Leslie and Powell, 2007).

The structure of CDA-C1(E17C)–**2a** in the R3:H space group was determined by molecular replacement using the program Phaser (Adams et al., 2010) with selenomethionine-labeled CDA-C1 (PDB code: 4JN3(Bloudoff et al., 2013)) as the search model, followed by iterative modeling in the program Coot (Emsley et al., 2010) and refinement in the program Phenix (Adams et al., 2010) to 1.6 Å resolution. The structure of CDA-C1(E17C)–**3a** in the $P2_12_12_1$ space group was determined by rigid body refinement of the structure of the selenomethionine-labeled CDA-C1 (PDB code: 4JN3(Bloudoff et al., 2013)) followed by iterative modeling in the program Coot (Emsley et al., 2010) and refinement in the program Phenix (Adams et al., 2010) to 2 Å resolution. There are two copies of CDA-C1(E17C) in the asymmetric units of the crystals in both space groups. Unbiased electron density maps of CDA-C1(E17C)–**2a** show clear density for the entire substrate in both molecules. In the maps for CDA-C1(E17C)–**3a**, the entire substrate is visible in molecule A. In molecule B, the amino acid moiety is clearly visible at the same position interacting with H157 and the backbone carbonyl of S386, but a portion of the alkane linker is ill defined, presumably because it is flexible.

The same procedure was performed with the product complexes of CDA-C1(E17C) – 2,3-epoxyhexanoyl-**2a** and CDA-C1(E17C) – 2,3-epoxyhexanoyl-**3a,** but the density for the epoxyhexanoyl-**2a** and epoxyhexanoyl-**3a** moieties was too weak to be fit reliably, perhaps because of a reduced affinity of this product complex for the

active site. We performed mass spectrometry with washed and melted crystals to confirm that the crystallized protein did indeed contain the product complex.
The pKa of H157 was estimated using the H++ server (Anandakrishnan et al., 2012). Chain A of the refined CDA-C1(E17C) model, including waters and ligands, was used as input, with parameters set at 0.15 mM NaCl, pH 7.5, an external dielectric constant of 80, and an internal dielectric constant of 15.

4.5.8 In vitro condensation domain activity assay

DptA-PCP1 (100 μM) was incubated with 10 mM MgCl2, 40 μM Sfp, and 800 μM amino acid-CoA in a buffer containing 50 mM Bis-Tris, pH 6.5, 100 mM NaCl, and 1 mM TCEP at room temperature for 30 minutes. 25 μM each of loaded DptA-PCP1 and ACP were combined with 10 μM CDA-C1 (wild type, S309G, S309V) in a buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM NaCl and 1 mM TCEP and incubated at room temperature for 1 hour.

Separation of the condensation reaction was performed using reverse phase chromatography (C4, 4.6 mm × 250 mm Microsorb 300 Å, 5 μM; Varian). At a flow rate of 0.75 ml min^{-1}, protein was eluted using a gradient of 40% to 50% B over 25 min (buffer A: 0.2% trifluoroacetic acid with H2O and buffer B: 0.2% trifluoroacetic acid with acetonitrile). Protein masses were determined offline using ESI-MS (Agilent Technologies 6510 Q-TOF) in positive-ion mode with an ESI nebulizer (Mass Spectrometry Core Facility, Bellini Life Science Complex, Montreal, Canada). The mass spectrometer was set to acquire spectra in the mass range 600–1700 m/z for an average of 2 min. Acquired spectra were averaged, and the deconvoluted masses of the proteins were determined using ESIprot Online (Winkler, 2010).

4.5.9 Data deposition

The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank, www.pdb.org (accession code 5DU9 and 5DUA).

4.6 Acknowledgements

The authors thank Mark A. Hancock and Gerhard Multhaup (McGill SPR-MS Facility) for assistance with the in-source decay experiments performed on Canada Foundation for Innovation-funded MS equipment, Alex Wahba (MS Facility, Department of Chemistry, McGill) for intact protein ESI experiments, John Colucci and Robert Zamboni (Zamboni Chemical Solutions) for substrate syntheses and MS, Shaun Labuik and Pawel Grochulski (Canadian Light Source) for diffraction data collection, Janice Reimer for optimizing the synthesis conditions of amino acyl-CoA substrates, Michael Skinnider and Nathan Magarvey for large-scale computational analysis of C domains sequences, and Nathan Magarvey, Kalle Gehring and Jon Lorsch for critical reading of the manuscript. Funding support is from Canadian Institutes of Health Research (106615) and Canada Research Chairs to T.M.S. K.B. was supported by the NSERC CREATE Training Program in Bionanomachines

4.7 Supplemental Figures and Tables

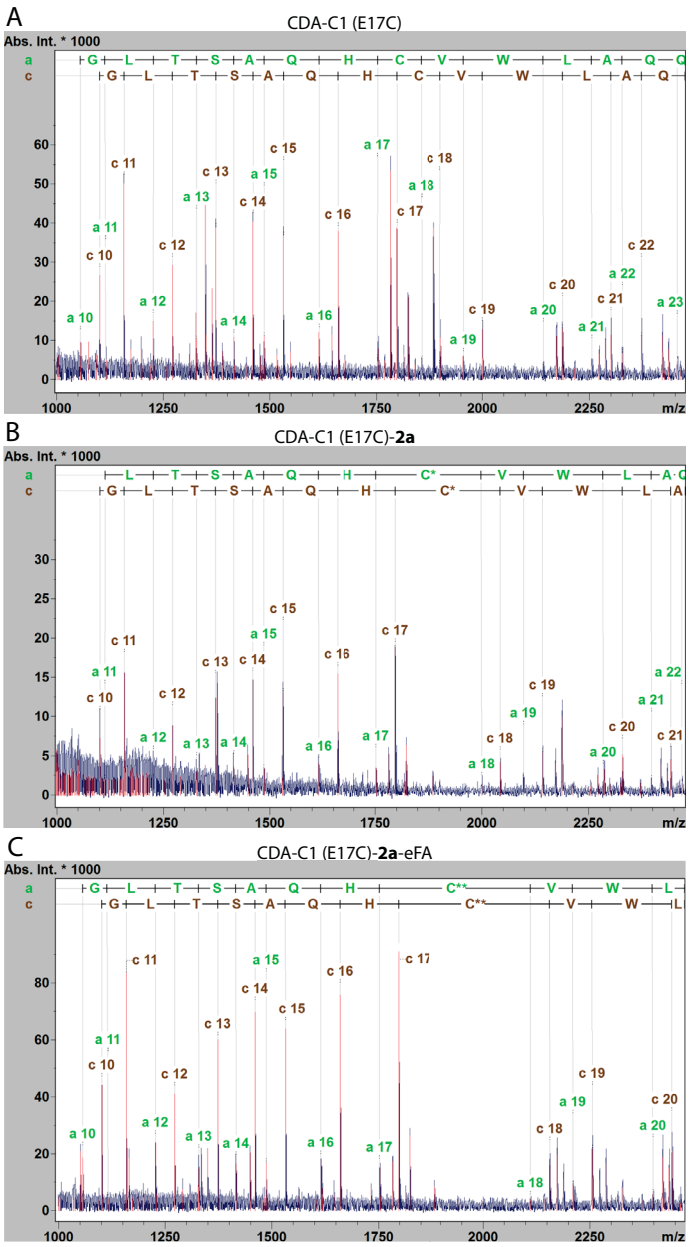


Figure S4.1. MALDI-ISD spectrum of CDA-C1(E17C) and complexes. (a) MALDI-ISD spectrum of apo CDA-C1(E17C). N-terminal fragmentation produced c- and a-ion series indicates C17 is unmodified. (b) MALDI-ISD spectrum of CDA-C1(E17C)-2a. C* and following residues in c- and a-ion series indicate C17 has increased by a mass equal to 2a. (b) MALDI-ISD spectrum of CDA-C1(E17C)-2a-eFA. C** and following residues in c- and a-ion series indicate C17 has increased by a mass equal to 2a-eFA. Note that the cloning introduced an extra N-terminal residue, so the residue numbering is shifted by +1.

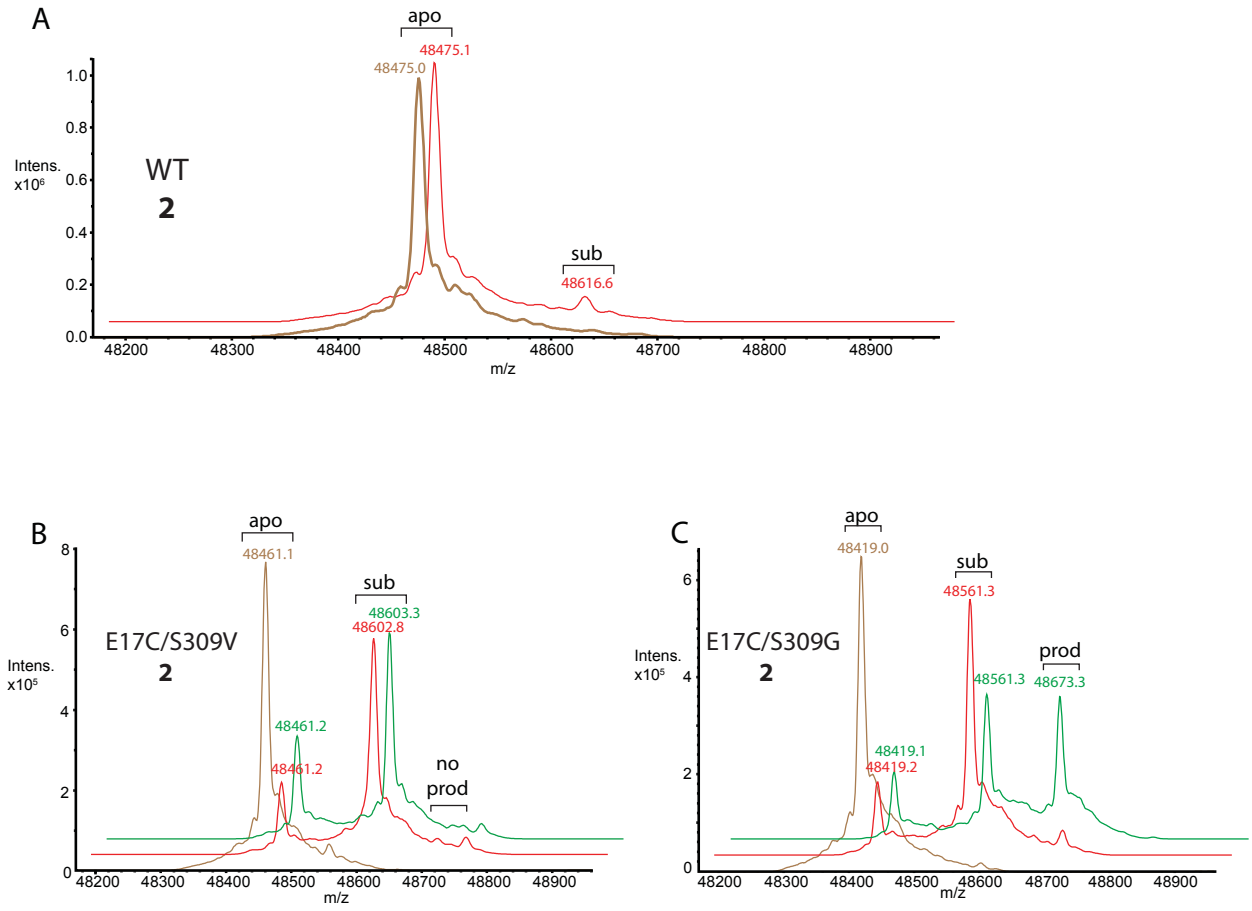


Figure S4.2. ESI-MS of intact CDA-C1 to assess alkylation and condensation reactions (brown: before reaction, red: after alkylation, green: after condensation). (a) ESI-MS of intact wild-type CDA-C1 protein. (b) ESI-MS of intact CDA-C1 (E17C/S309V). (c) ESI-MS of intact CDA-C1 (E17C/S309G). See Table S4.1 for expected masses.

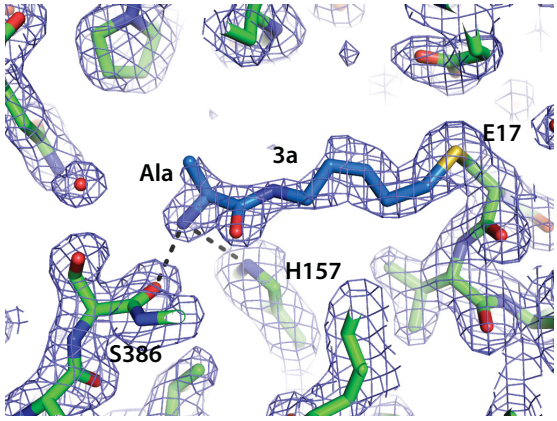


Figure S4.3. Structure and $2F_O$-F_C electron density maps for CDA-C1(E17C)-3a contoured at 1σ.

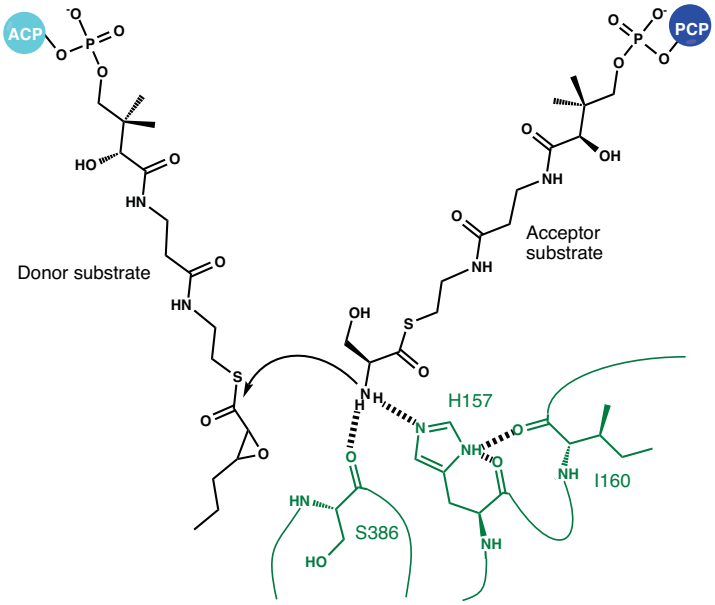


Figure S4.4. Schematic of the CDA-C1 active site.

CDA-C1 Complex	Expected Mass (Da)	Measured pre-reaction (Da)	Measured post-alkylation (Da)	Measured post-reaction (Da)
E17C + **1**				
E17C-Apo	48449	48441.1	48449.2	48449.0
E17C-**1a**	48577	-	48577.5	48577.3
E17C-**1a**-eFA	48689	-	-	48688.9
E17C + **2**				
E17C-Apo	48449	48449.1	48449.0	48449.2
E17C-**2a**	48591	-	48591.2	-
E17C-**2a**-eFA	48703	-	-	48703.0
E17C + **3**				
E17C-Apo	48449	48449.1	48449.1	48448.3
E17C-**3a**	48605	-	48605.3	48604.8
E17C-**3a**-eFA	48717	-	-	48716.8
E17C;H157A + **2**				
E17C;H157A-Apo	48383	48382.6	48382.9	48382.4
E17C;H157A-**2a**	48525	-	48525.2	48524.6
E17C;H157A-2a-eFA	48337	-	-	-
E17C/S309V + **2**				
E17C/S309V-Apo	48461	48461.1	48461.2	48461.2
E17C/S309V-**2a**	48603	-	48602.8	48603.3
E17C/S309V -**2a**-eFA	48715	-	-	-
E17C/S309G + **2**				
E17C/S309G -Apo	48419	48419.0	48419.2	48419.1
E17C/S309G -**2a**	48561	-	48561.3	48561.3
E17C/S309G -**2a**-eFA	48673	-	-	48673.3
WT + **2**				
WT-Apo	48476	48475.0	48475.1	-
WT-**2a**	48618	-	48616.6	-

Table S4.1. Expected and measured masses modified CDA-C1.

	CDA-C1(E17C)-**2a**	CDA-C1(E17C)-**3a**
Diffraction Data		
Space Group	R 3 H (No. 146)	P 2_1 2_1 2_1 (No. 19)
Unit-cell parameter (Å,°)	213.43, 213.43, 52.90, 90, 90, 120	62.66, 83.82, 178.07, 90, 90, 90
Resolution (Å)	42.17-1.53 (1.56-1.53)	75.84-1.88 (1.92-1.88)
I/σ(I)	6.6 (1.0)	5.4 (1.3)
Temperature (K)	100	100
Measured Reflections	514599 (21723)	395060 (19636)
Unique Reflections	135546 (6752)	75270 (4023)
Completeness (%)	100.0 (99.9)	97.6 (89.6)
Multiplicity	3.8 (3.2)	5.2 (4.9)
Rmerge (%)	8.1 (82.1)	15.1 (68.7)
Estimates of resolution limits (Å):		
from half-dataset correlation CC(1/2) > 0.50	1.59	1.88
from Mn(I/sd) > 2.00:	1.68	2.04
Refinement statistics		
Resolution range (Å)	34.8-1.6	50-1.9
R factor/R_{free} (%)	17.1 / 19.3	17.7 / 22.1
Average B Factor (Å^2)	31.8	31.8
R.M.S.D in bond lengths (Å)	0.003	0.008
R.M.S.D in bond lengths (°)	0.839	1.068
Ramachandran Plot (%)		
Favoured	98.9	98.8
Allowed	1.1	1.2
Outliers	0	0

Table S4.2. Crystallographic statistics for data collection and processing for modified CDA-C1. Values in parentheses are for the highest-resolution shell.

Donor Substrate	$M_{exp.}$ of eFA-aminoacyl-PCP (Da)	$M_{obs.}$ in CDA-C1 (wild type) assay	$M_{obs.}$ in CDA-C1 (S309V) assay	$M_{obs.}$ in CDA-C1 (S309G) assay
Serine	12617.3	12617.8	n.d.	12617.8
Alanine	12601.3	12601.8	n.d.	n.d.
Leucine	12643.3	12643.8	n.d.	n.d.
Methionine	12661.4	12661.8	n.d.	12661.8

Table S4.3. Expected and measured masses of products of the full condensation assay. Deconvoluted masses were calculated by ESIprot Online using peaks corresponding to protein charge states 11+, 12+, and 13+.

Primer Name	Sequence (5′ - 3′)
E17C-F	CTGACTAGCGCACAACATTGCGTTTGGCTGGCGCAGCAA
E17C-R	TTGCTGCGCCAGCCAAACGCAATGTTGTGCGCTAGTCAG
H157A-F	CTACTATCTGGGCGTCCATGCTATTGTTATCGATGGCACC
H157A-R	GGTGCCATCGATAACAATAGCATGGACGCCCAGATAGTAG
S309G-F	GCAGGCGGTTGCCCACCATGCCTGG
S309G-R	CCAGGCATGGTGGGCAACCGCCTGC
S309V-F	CGGCAGGCGGTTGACCACCATGCCTGGG
S309V-R	CCCAGGCATGGTGGTCAACCGCCTGCCG

Table S4.4. Primers used for site-directed mutagenesis.

4.8 Bridge to Chapter 5

As work on Chapter 4 was coming to a close, I wished to take my research in a slightly different direction. I believed studying the structure of the Cy domain could provide fascinating insights into how a protein with a highly (hypothetical) similar architecture to C domains could perform a cyclodehydration reaction in addition to a condensation reaction. Using the structure of a Cy domain, I would be able to model in its substrates, identify residues that could potentially be involved in either of its catalytic reaction, and measure their contributions to catalysis using a robust biochemical assay. I believed this would provide a well-needed update into the function of this intriguing domain.

In additon, a structural study of a Cy domain could complement the results presented in Chapter 4. Presumably, both C and Cy domains perform the condensation reaction in a similar manner. Therefore, if I could identify a residue (or residues) potentially involved in acceptor substrate positioning in the Cy domain condensation reaction, this would be a strong argument for condensation domain reaction mechanism that I proposed in the previous chapter.

CHAPTER 5: STRUCTURAL AND MUTATIONAL ANALYSIS OF THE NONRIBOSOMAL PEPTIDE SYNTHETASE HETEROCYCLIZATION DOMAIN PROVIDES INSIGHT INTO CATALYSIS

Bloudoff K, Fage CD, Marahiel MA, and Schmeing TM (2017). Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. *Proceedings of the National Academy of Sciences* 114(1): 95-100.

5.1 Introduction

Nonribosomal peptide synthetases (NRPSs) are a family of large, multimodular enzymes that produce a wide range of important bioactive secondary metabolites (Weissman, 2015). NRPS products have great diversity because they can use over 500 different acyl monomer substrates, including L-amino acids, D-amino acids, aryl acids, fatty acids, hydroxy acids, and keto acids, and they can subsequently modify these moieties during peptide synthesis.

NRPSs function in a modular, assembly line fashion. A typical elongation module consists of a condensation (C), an adenylation (A), and a peptidyl-carrier protein (PCP) domain. The A domain specifically recognizes and activates a monomer acyl substrate though adenylation, then transfers it onto the prosthetic phosphopantetheine arm of the PCP domain. This acyl-PCP then travels to the C domain acceptor site for condensation with the upstream module's acyl-PCP at the C domain donor site (Bergendahl et al., 2002; Bloudoff et al., 2013; Keating et al., 2002; Samel et al., 2007). The PCP domain brings the elongated peptide chain to the downstream module, where it is passed off and further elongated in the next condensation reaction. This process is repeated in each module until the termination module, where the final free product is released from the PCP domain, often by a thioesterase domain. However, most NRPSs, along with their C, A and PCP domains, also include tailoring and/or alternative domains, which co-synthetically modify the nonribosomal peptide.

One important modification that can occur during peptide synthesis is cyclodehydration of Cys, Ser or Thr residues into thiazoline, oxazoline or methyloxazoline rings, respectively, by the NRPS heterocyclization (Cy) domain (Chen et al., 2001; Duerfahrt et al., 2004; Gehring et al., 1998; Keating et al., 2000a; Kelly et al., 2005; Konz et al.; Marshall et al., 2001). These heterocyclic rings are found in many peptides with important clinical and research utility, such as the antibiotics bacitracin A (Konz et al., 1997) and zelkovamycin (Tabata et al., 1999), the antitumor agents bleomycin (Shen et al., 2002) and epothilone (Chen et al., 2001), the immunosuppressant argyrin (Vollbrecht et al., 2002), the siderophores mycobactin (Snow, 1970) and yersiniabactin (Drechsel et al., 1995), and the microbiome genotoxin colibactin (Vizcaino and Crawford, 2015). Introduction of the 5-membered heterocyclic

ring makes the peptide resistant to proteolytic cleavage, and can induce conformations in the peptide that favor interaction with biological targets (Roy et al., 1999).
In NRPSs that synthesize these heterocycle-containing products, the module specific for the Cys, Ser or Thr monomer substrate contains a Cy domain in place of the C domain. Cy domains are evolutionarily and structurally related to C domains (Rausch et al., 2007). The Cy domain first catalyzes nucleophilic attack on the thioester of a PCP-linked donor substrate by the α-amino group of a Cys-, Ser- or Thr-PCP substrate, presumably in a manner similar to C domains (Bloudoff et al., 2016; Duerfahrt et al., 2004; Gehring et al., 1998; Keating et al., 2000b; Kelly et al., 2005; Konz et al.; Marshall et al., 2001; Marshall et al., 2002b) (Figure 5.1). In the two-step cyclodehydration reaction that follows, the thiol of the Cys side chain or hydroxyl of the Ser or Thr side chain first attacks the carbonyl carbon of the newly-formed amide bond to form the heterocycle (Duerfahrt et al., 2004; Gehring et al., 1998) and then the former carbonyl oxygen is removed in a dehydration reaction, which introduces the carbon-nitrogen double bond of the thiazoline or (methyl)oxazoline ring. The nascent heterocyclic peptidyl-PCP can be used as the donor substrate by the next module's C domain, or is first oxidized or reduced by discrete oxidase or reductase domains (Duerfahrt et al., 2004; Patel and Walsh, 2001).

All C domain superfamily domains share the same protein fold, so the overall form of the Cy domain is not in doubt (Chen et al., 2016; Samel et al., 2014). However, the features that allow the Cy domain to catalyze two separate and very different reactions are not known. Cy domains contain a conserved Asp-motif, DxxxxD, which directly replaces the catalytic His-motif, HHxxxD, of C domains (Konz et al., 1997). Mutation of the aspartate residues of the Asp-motif diminishes or abolishes catalytic activity of the protein(Di Lorenzo et al., 2008; Keating et al., 2000b; Kelly et al., 2005; Marshall et al., 2002a). Furthermore, other mutations have differential effects on the condensation reaction and the cyclodehydration reaction, suggesting that the reactions are not catalyzed by a completely overlapping set of residues(Duerfahrt et al., 2004; Marshall et al., 2002b).

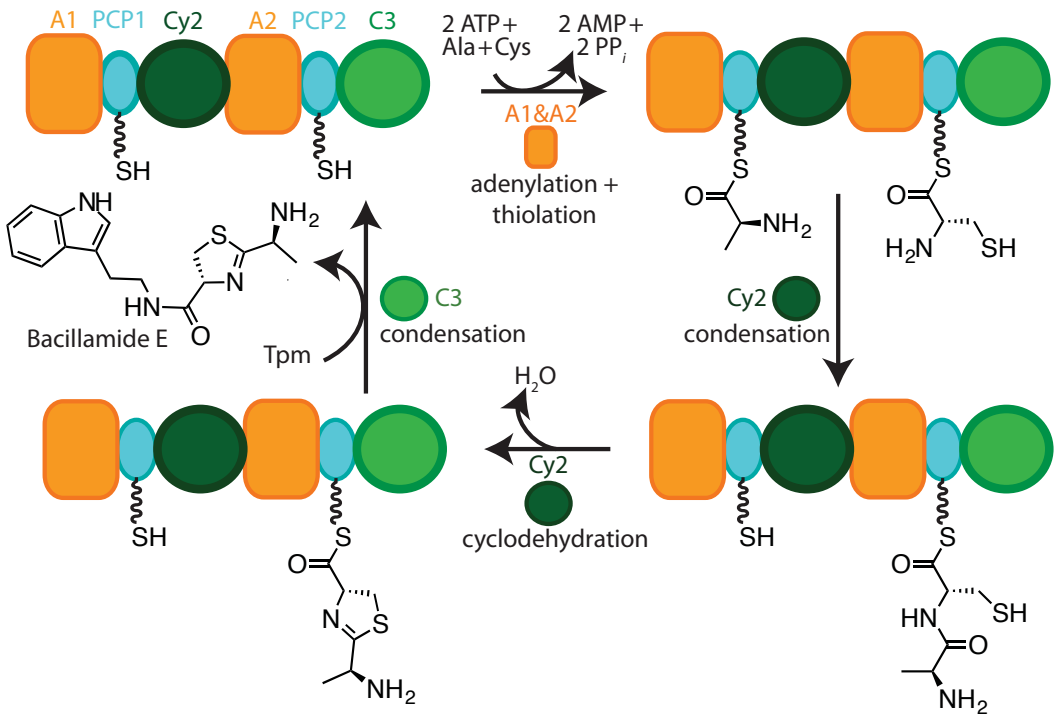


Figure 5.1. Schematic representation of BmdB, and the bacillamide E biosynthesis cycle. The A domains (orange) adenylate alanine and cysteine and transfer them onto the phosphopantetheine arm of the PCP domains (blue). The Cy2 domain (dark green) first catalyzes amide bond formation between the acyl-linked Ala and Cys residues, then catalyzes the intramolecular cyclodehydration reaction. The C3 domain (light green) catalyzes amide bond formation between the PCP-linked Ala-thiazoline moiety and free tryptamine, which releases the bacillamide E from the NRPS.

Since Cy domains are not larger than C domains, the added function in Cy domains must occur in a confined sequence space.

Here, we present the crystal structure of Cy2 of bacillamide synthetase at 2.3 Å resolution. Bacillamide synthetase (Figure 5.1) is a trimodular NRPS that produces bacillamide E, whose derivatives bacillamide A-D exhibit algicidal activity against dinoflagellates, raphidophytes, and cyanobacteria (Figure S5.1) (Churro et al., 2009; Jeong et al., 2003). Structural determination, along with mutagenic analysis of this Cy domain active site in the context of the full, intact and active bacillamide synthetase, and bioinformatic investigation allowed us to newly identify D1226 and T1196 as important residues for cyclodehydration, providing a better understanding of the structure and mechanism of the Cy domain.

5.2 Results and Discussion

5.2.1 The crystal structure of an NRPS heterocyclization domain

We have solved the crystal structure of an NRPS heterocyclization domain by X-ray crystallography to a resolution of 2.3 Å. Like all known C domain superfamily proteins, BmdB-Cy2 adopts two chloramphenicol acetyltransferase folds (Keating et al., 2002), with the N- and C-terminal lobes forming a pseudo-dimer (Figure 5.2).

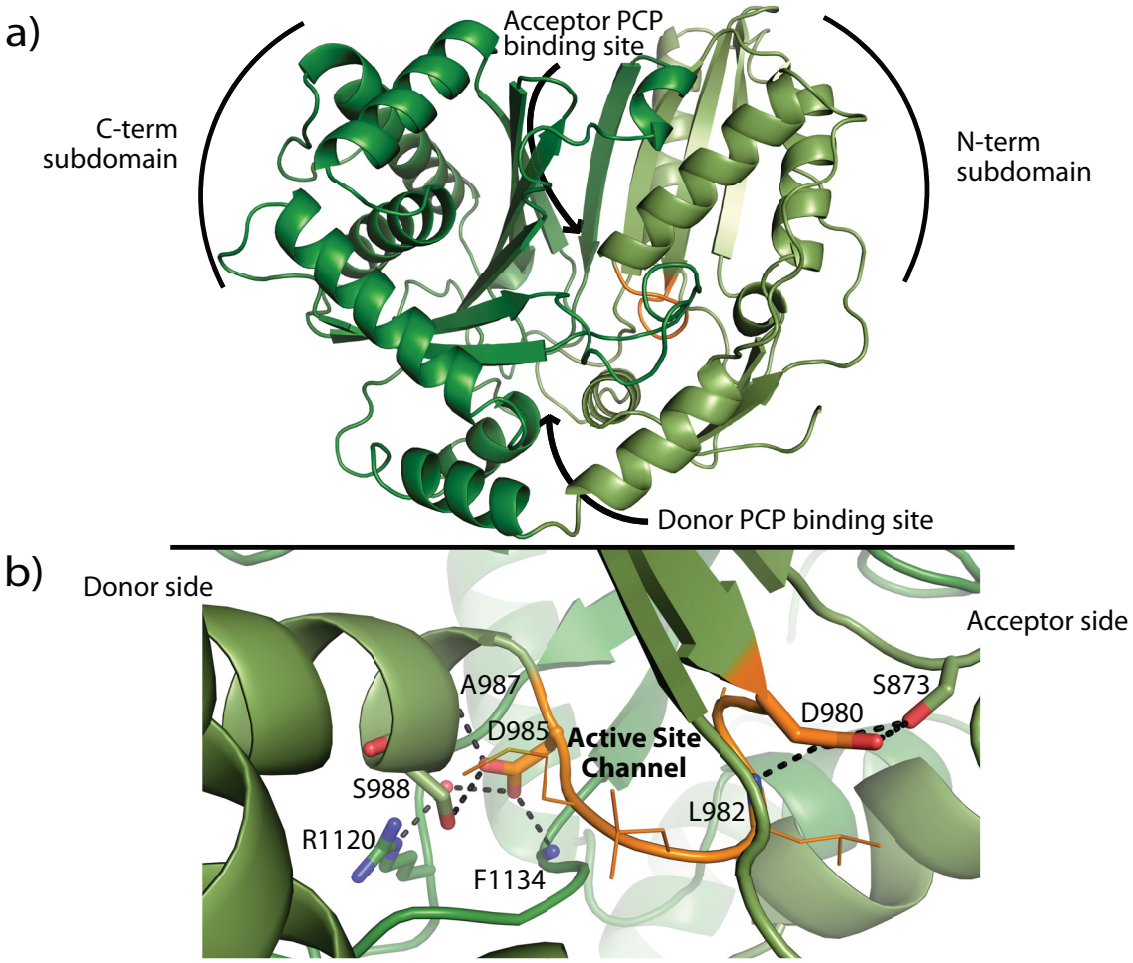


Figure 5.2. The crystal structure of BmdB-Cy2. (a) Ribbon representation of BmdB-Cy2. The cyclization domain adopts two chloramphenicol acetyltransferase folds, similar to C domains. (b) Close-up view of the BmdB-Cy2 DxxxxD motif (orange). D980 makes a bifurcated hydrogen bond with the hydroxyl of S873 and a hydrogen bond with the backbone nitrogen of L982. D985 makes hydrogen bonds with S988 and the amides of A987 and F1134, as well as water-mediated interaction with R1120.

In C domains, these two lobes assume a range of relative conformations (more “open” or “closed” (Bloudoff et al., 2013)). Superimposition of BmdB-Cy2 with each C domain shows that it fits within the observed range, and is quite similar to AB3403 (Drake et al., 2016) (Figure S5.2). The latch formed by a crossover between N- and C-terminal subdomains in C domains is present, meaning that the Cy domain active site is also situated in the center of a tunnel, between donor and acceptor PCP binding sites. (See Supplemental Results and Discussion for more detailed description of the overall Cy domain structure.)

5.2.2 BmdB-Cy2 Active Site

The main established signature that differentiates Cy domains from C domains is an active site motif: C domains have a catalytic His-motif of HHxxxD, and Cy domains have an Asp-motif of DxxxxD (Konz et al., 1997). The first aspartate of the Asp-motif is essential for catalytic activity in only some Cy domains. When the first aspartate is mutated to alanine in HMWP2-Cy1, AngN-Cy1, and AngN-Cy2, condensation and cyclodehydration are completely eliminated (Di Lorenzo et al., 2008; Keating et al., 2000b), but in VibF-Cy1, VibF-Cy2, and EpoB-Cy1, both condensation and cyclodehydration occur, although they are significantly diminished(Keating et al., 2000b; Kelly et al., 2005; Marshall et al., 2002a) (Table S5.1 lists current and previous Cy domain mutations). The second aspartate is critical for activity(Di Lorenzo et al., 2008; Keating et al., 2000b; Kelly et al., 2005; Marshall et al., 2002a), but whether its role is catalytic or structural (like the aspartate at that position in in C domains) (Roche and Walsh, 2003) was not definitively determined. Furthermore, a triple mutant of the Asp-motif of HMWP2-Cy1 to introduce a C domain His-motif results in a catalytically inactive protein (Keating et al., 2000b), demonstrating that a straight swap of the motifs is not sufficient to interconvert catalytic activities.

In the structure of BmdB-Cy2, both aspartate residues in DxxxxD (D980 and D985) are oriented away from the active site (Figure 5.2b). The side chain of D980 makes a bifurcated hydrogen bond with the hydroxyl of S873, and a hydrogen bond with the backbone amide of L982. D985 hydrogen bonds with the sidechain of S988, and the amides of A987 and F1134. The interaction this aspartate makes with a nearby

arginine in C domains is present, but is water-mediated for D985-R1120 of BmdB-Cy2. Thus, these aspartate residues directly replace the first histidine and the aspartate in the C domain HHxxxD motif, occupying the same position in their respective domains and making the same or similar interactions with the rest of the protein (Figure S5.2b). Indeed, the whole DxxxxD motif is essentially in the same conformation as the HHxxxD motif. The position of the second histidine of the HHxxxD motif, which is the most important residue for condensation in C domains (Bergendahl et al., 2002; Bloudoff et al., 2013; Keating et al., 2002; Samel et al., 2007), is occupied by A981 in BmdB-Cy2, and thus unable to contribute to catalysis. The HHxxxD motif in C domains does not reorient upon substrate binding (Bloudoff et al., 2013), and there is no indication that the Cy domain DxxxxD motif would do so. Overall, the structure strongly suggests that both of these aspartate residues play structural rather than catalytic roles, as has been established for the corresponding residues in C domains (Bergendahl et al., 2002; Bloudoff et al., 2013; Keating et al., 2002; Samel et al., 2007).

5.2.3 Bacillamide synthesis assay and mutational analysis

In order to determine the importance of other active site residues in condensation and cyclization, we established a bacillamide synthesis assay. Bacillamide synthetase is a 1-protein, 6-domain, 265 kDa NRPS (Figure 5.1) (Yuwen et al., 2013). It adenylates and thiolates its Ala and Cys substrates, after which the Cy2 domain performs condensation and cyclodehydration. The final domain in this NRPS is not a thioesterase domain, but a specialized C domain that condenses the thiazoline-containing intermediate with free tryptamine (Tpm) to release bacillamide E. Therefore, this terminal C domain plays all the roles typically associated with a 4-domain termination module – substrate selection, peptide bond formation and release of the peptide product – which is rare, but not unprecedented in NRPSs (Di Lorenzo et al., 2008; Kelly et al., 2005).

The Tpm in the product makes the reaction convenient to follow by HPLC at 280 nm (Figure 5.3, Figure S5.3).

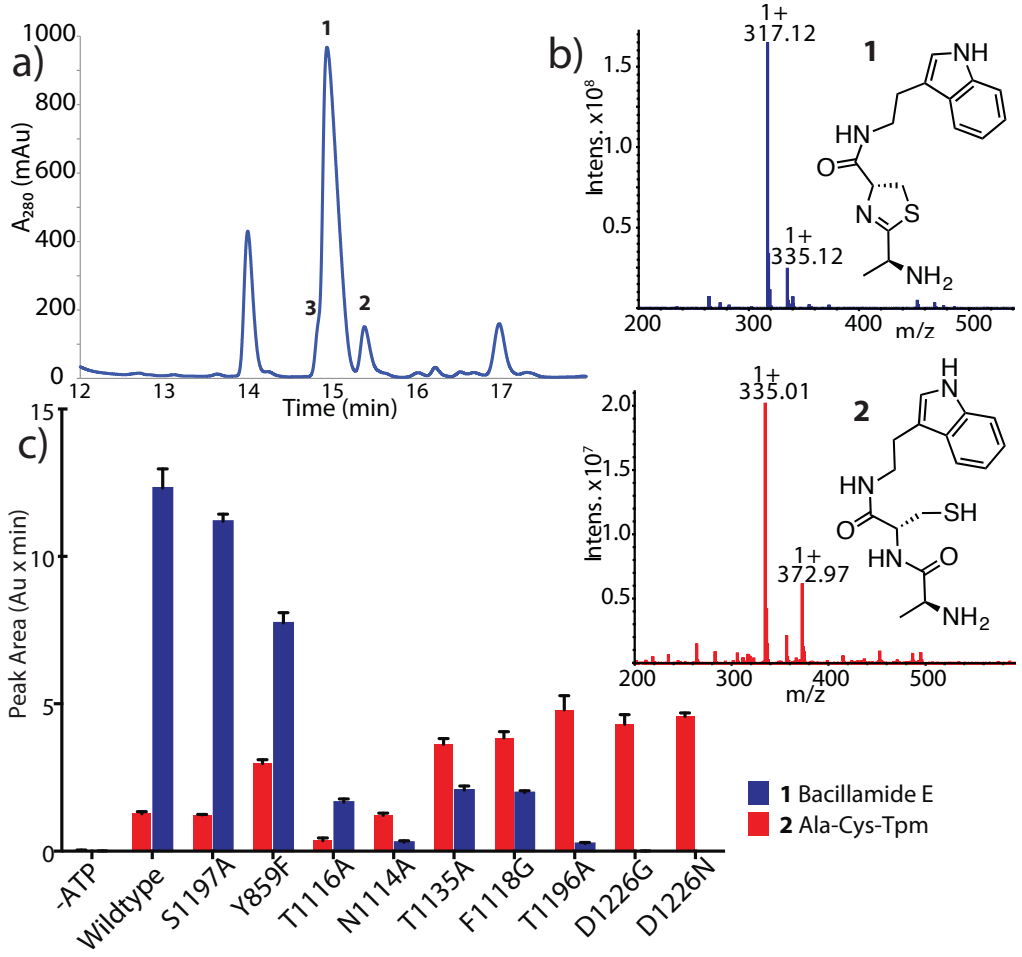


Figure 5.3. Effects of structure-guided mutations in BmdB-Cy2 on bacillamide E synthesis. (a) Representative HPLC trace of a BmdB activity assay. Compound 1 is bacillamide E, compound 2 is linear Ala-Cys-Tpm, and compound 3 is Cys-Tpm. (b) Representative mass spectra of 1 (blue) and 2 (red). A representative mass spectrum of 3 is shown in Figure S5.3a. (c) Quantification of the relative production of 1 (blue) and 2 (red) in reactions with mutant BmdB. All reactions were measured in triplicate, except F1118G, which was measured in duplicate. Error bars represent standard deviation.

Wild-type bacillamide synthetase showed a large peak at 15 min that corresponded by high resolution mass spectrometry, fragmentation mass spectrometry and NMR to bacillamide E (**1**) (Figure S5.3c). A minor peak at ~15.5 min, with 10% of the intensity of the main peak, corresponded to the uncyclized tripeptide Ala-Cys-Tpm (**2**). This shows that, at least *in vitro*, BmdB-Cy2 is not completely efficient at cyclizing all the intermediates it condenses, and that the terminal C3 domain is not completely selective for the cyclized substrate. Furthermore, a small peak for the dipeptide Cys-Tpm is also

evident (**3**, Figure S5.3), indicating that the terminal C3 domain also uses Cys-PCP2 as a donor substrate to some extent. Note that bacillamide E contains a thiazoline ring, and not a thiazole ring as in bacillamide A-D extracted from natural sources (Ivanova et al., 2007; Jeong et al., 2003; Socha et al., 2007). The adjacent oxidase BmdC likely performs the oxidation post-synthetically.

We next produced a series of bacillamide synthetase constructs with mutations in the Cy domain. We targeted residues in spatial proximity to the DxxxxD motif in the active site tunnel of BmdB-Cy2, with side chains that could play an important role in condensation or cyclization, and which are largely conserved in alignments of characterized Cy domains (Figure S5.4). Therefore, bacillamide synthetase with mutations Y859F, T1116A, F1118G, T1135A, T1196A, D1226G and D1226N, as well as N1114A and S1197A (two residues shown previously to selectively affect cyclodehydration (Duerfahrt et al., 2004)) were purified and assayed.

The bacillamide synthetase mutants had radically different effects on condensation and cyclodehydration (Figure 5.3, Figure S5.3, and Supplemental Results and Discussion). Mutant S1197A had almost no effect on production of linear or heterocyclized tripeptide. With Y859F, bacillamide E production decreased to 79% of wild-type, but linear Ala-Cys-Tpm production doubled. Although the hydroxyl of Y859 points directly into the heart of the active site, the relatively minor effect of its removal suggests it is unlikely to act chemically. Mutants N1114A and T1116A each drastically reduced synthesis of both the linear and heterocyclic product. T1116A maintained approximately the same ratio of heterocyclic to linear product as wild-type, whereas the little product made by N1114A was successfully heterocyclized, consistent with previously reported decoupling of condensation and heterocyclization by this residue (Duerfahrt et al., 2004). Both N1114 and T1116 thus appear important for condensation, and N1114 is also important for cyclodehydration.

For insight into the cyclization reaction, the most interesting mutations are those that selectively affect cyclodehydration. Mutation of residues F1118, T1135, T1196 and D1226 resulted in substantially more linear Ala-Cys-Tpm and less bacillamide E than the wild-type enzyme. F1118 is positioned along the tunnel that the phosphopantetheine arm of donor PCP1 occupies when presenting the Ala substrate to the active site

(Figure S5.5). In the structure of BmdB-Cy2, the F1118 side chain completely blocks that tunnel. This residue must move to allow alanyl-PCP1 to bind and participate in the condensation reaction. However, for cyclization, PCP1 likely departs, leaving the cyclization substrate Ala-Cys-PCP2 bound at the acceptor site. It is possible that F1118 aids cyclization by blocking the donor side of the tunnel to create a single-entry active site more optimal for cyclodehydration. This is reminiscent of the (permanent) blocking of the acceptor site by a tryptophan side chain to form a single-entry active site in related epimerization domains, observed in a recently-published crystal structure (Chen et al., 2016). Even more striking are the effects of the T1196A mutation, which allowed only minimal cyclization, and mutations of D1226 to glycine or asparagine, which obliterated cyclization while retaining robust condensation. These two residues are adjacent to one another, oriented towards the active site cavity, and hydrogen bond to one another. Their position and importance for cyclization activity make them most likely to play a direct role in catalysis (see below).

5.2.4 Bioinformatic analysis of Cy domains

We undertook a bioinformatic analysis of Cy domains. We retrieved 36,853 C domain superfamily sequences that had a maximum pairwise sequence identity of 90%, and sorted them using the Natural Products Domain Seeker web server(Ziemert et al., 2012). Multiple sequence alignment of 1790 Cy domains showed two motifs in the C-terminal region that stand out as highly conserved in Cy domains but absent from the other C domain superfamily proteins (Figure S5.6). These motifs bear the consensus sequences PVVFTS and SQTPQVxLD (Figure 4.4a), and are part of what had been recognized by Konz *et al.*(Konz et al., 1997) as conserved signature sequences 6 and 7. They exist in BmdB-Cy2 as PIVFTS (1192-1197) and ARTPQVYLD (1218-1226). These motifs are as close as 9 Å in space to the DxxxxD, but they are separated from it in sequence by over 200 amino acids. The motifs form a surface on the acceptor substrate side of the active site (Figure 5.4b, Figure S5.6a). Interestingly, this is the putative location of the Ala-Cys-phosphopantetheinyl substrate in the cyclization reaction, and includes the residues that displayed the most drastic effect on cyclization when mutated, namely T1196 of the PVVFTS motif and D1226 of the SQTPQVxLD

motif. D1226 is one of the most conserved residues in Cy domains, present in 96% of the 1790 sequences.

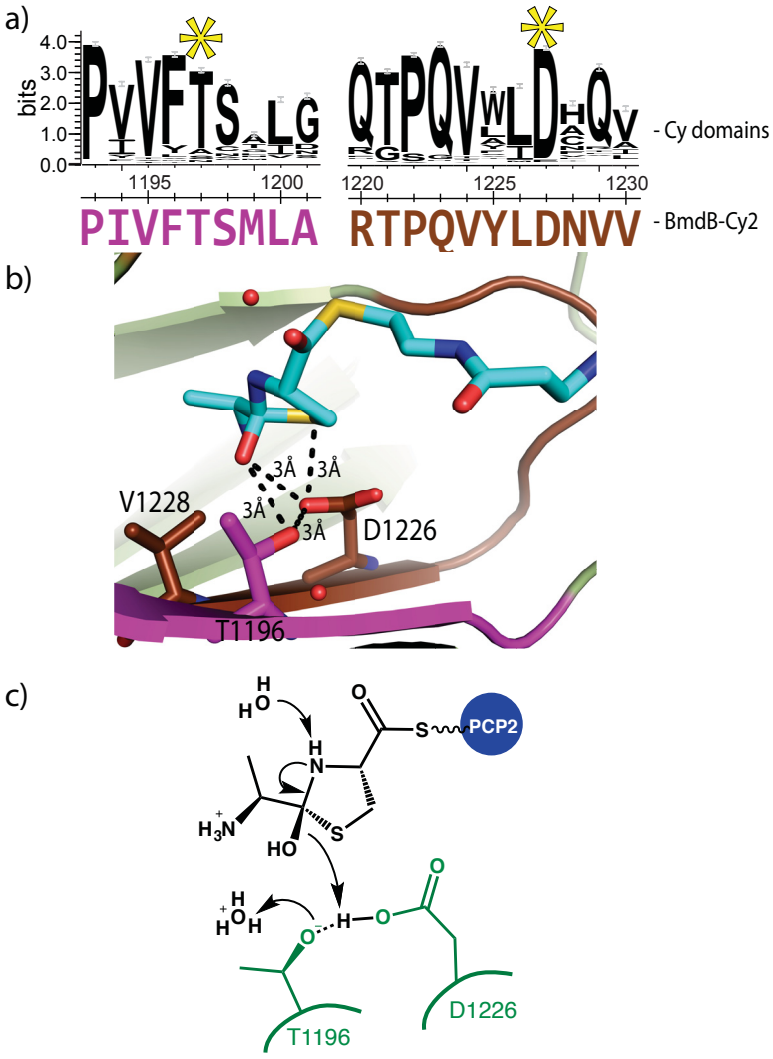


Figure 5.4. Model of cyclodehydration intermediate model and critical residues for cyclodehydration reaction. (a) WebLogo3 (Crooks et al., 2004) of Cy domain motifs PVVFTS (core Cy6) and SQTPQVxLD (part of core Cy7) compared to the sequence of BmdB-Cy2. Putative catalytic residues T1196 and D1226 are labeled with yellow asterisks. (b) BmdB-Cy2 (mostly green, with DxxxxD in orange, PVVFTS in purple, and SQTPQVxLD in brown) with a model of the cyclodehydration intermediate (blue). Putative catalytic residues T1196 and D1226 are shown in sticks, as is V1228, a position occupied by glutamine in most Cy domains. (c) Possible mechanism of the dehydration step. We suggest that deprotonation of the amino hydrogen occurs after dehydration and double bond formation. The full putative reaction pathway is diagramed in Figure S5.7.

T1196 is somewhat more variable: It is a threonine or serine in 88% of the sequences.

Putative functions had not previously been assigned to these motifs. S1197 was the only residue in either motif previously targeted by mutation, and it had differential effects on Cy domain reactivity in EpoB, BacA2, and BmdB, consistent with its moderate conservation (Figure 5.3, 5.4a, Table S5.1) (Duerfahrt et al., 2004; Kelly et al., 2005). Notably, a portion of Cy domain sequence that largely overlaps with the two new motifs is assigned PFam08415. It is annotated only as being found in NRPSs with C and PCP domains, and ends partway through the SQTPQVxLD motif, before the critical D1226. We propose that this be extended to residue 1241 (to incorporate the whole SQTPQVxLD and a conserved WD (Figure S5.5-5.6a)) and annotated as a signature for Cy domains.

5.2.5 A trend for tandem Cy domains in (methyl)oxazoline-forming modules

Cy domains can be subdivided by their use of thiol (Cys) or hydroxyl (Ser or Thr) groups as the nucleophile in the cyclization reaction. Each subset contains the highly conserved PVVFTS and SQTPQVxLD motifs, and there are no discernable differences between Cy domain sequences in each subset. However, one clear trend did emerge: In general, modules that incorporate Cys substrates contain only one Cy domain per module, whereas those using Thr or Ser have tandem Cy-Cy (or Cy-C) domains. Out of a set of 505 full-length Cy domain-containing proteins, 454 are predicted to use Cys (440 with single Cy domains, 14 with Cy-Cy domains), while 51 are predicted to use Ser or Thr (2 with a single Cy domain,49 with Cy-Cy domains). This trend is maintained after the proteins have been filtered for unique architecture and/or unique predicted product: Of 200 remaining proteins, 187 use Cys (177 with a single Cy domain, 10 with Cy-Cy domains) and 13 use Ser or Thr (1 with single a Cy domain, 12 with Cy-Cy domains).

The trend of tandem Cy domains for hydroxyl-containing substrates and single Cy domains for thiol-containing substrates holds for most characterized NRPS systems that feature Cy domains (Chen et al., 2001; Du et al., 2003; Duerfahrt et al., 2004; Keating et al., 2000b; Kelly et al., 2005; Konz et al., 1997; Marshall et al., 2001; Marshall et al., 2002b; Quadri et al., 1999; Seyedsayamdost et al., 2012; Silakowski et

al., 1999; Wang et al., 2014). The tandem Cy domain arrangement is exemplified by the well-characterized VibF protein. In VibF, Cy2 is primarily responsible for condensation and Cy1 for cyclodehydration of the Thr substrate (Marshall et al., 2002b). Furthermore, the tandem Cy arrangement can be maintained even when the module is split between two proteins, as in serratiochelin synthetase SchF1 and SchF2, which also cyclodehydrates Thr (Seyedsayamdost et al., 2012; Wang et al., 2014). Exceptions to the trend are mycobactin synthetase (McMahon et al., 2012) (Thr-specific module with a single Cy domain in MbtB) and anguibactin synthetase (Cys-specific module with Cy-Cy domains in AngN) (Di Lorenzo et al., 2008). The tandem Cy domains in anguibactin synthetase could be a remnant of evolution from a Thr-using ancestor, since, other than containing a Cys in place of a Thr, actinomycin is identical to acinetobactin, and the two synthetases share identical domain configurations (Dorsey et al., 2004). Furthermore, the Cy domains in AngN can each perform both condensation and cyclodehydration reactions and are largely redundant with one another (Di Lorenzo et al., 2008).

The Cys thiol is a better nucleophile than the Ser or Thr hydroxyl (Belshaw et al., 1998). This may be why thiazoline-forming NRPS modules are ~10x more prevalent than (methyl)oxazoline-forming modules, and may explain the trend to dedicate two domains in tandem in (methyl)oxazoline-forming modules. Tandem domains may increase the probability of cyclodehydration before the peptide is donated in the downstream module's condensation reaction. The downstream C domains are likely not completely selective for the heterocycle-containing peptide, as shown with BmdB (Figure 5.3) and VibF (Marshall et al., 2002a), so efficiency of cyclodehydration would be important for cognate heterocycle production.

4.2.6 Insight into catalysis and model of the cyclodehydration intermediate

The first reaction performed by Cy domains is condensation. The N1114A and T1116A mutants of BmdB-Cy2 exhibited the greatest effects on condensation without fully abolishing it (Figure 5.3), and neither residue is highly conserved (Figure S5.6a). Likewise, all mutations that were previously reported to decrease condensation in Cy domains are shown by the BmdB-Cy2 structure to be too distal to (residues 900, 988, 1089, 1114, 1120) or turned away from (residues 980, 985 of the DxxxxD motif) the

atoms participating in condensation (Table S1) (Di Lorenzo et al., 2008; Duerfahrt et al., 2004; Keating et al., 2000b; Kelly et al., 2005; Marshall et al., 2002a). Unless the DxxxxD motif radically reorients (which is possible, although there is currently no supporting data), there do not appear to be any critical, conserved residues that could act to abstract or donate a proton in the condensation reaction in Cy domains. We have argued that substrate orientation is the principal source of catalytic power for condensation in C domains (Bloudoff et al., 2016), and this appears to be even more likely in Cy domains. Indeed, comparison of the kinetic parameters of uncatalyzed peptide bond formation and ribosome-catalyzed peptide bond formation shows that the rate enhancement provided by the ribosome comes predominantly from the lowering of activation entropy via positioning of substrates (and perhaps water molecules) for reaction (Beringer et al., 2005). This may likewise be true for the more reactive thioester substrates of Cy domains.

The second reaction performed by Cy domains, cyclodehydration, is a non-simplistic, two-step reaction. Cy domains and at least two other types of enzymes perform this reaction to form 5-membered rings in independent natural product biosynthesis systems. YcaO- and TruD-type enzymes processively modify ribosomally synthesized and post-translationally modified peptides using cyclodehydration reactions to introduce thiazoline and (methyl)oxazoline rings in the synthesis of peptides such as microcin and the trunkamides (Dunbar et al., 2012). YcaO and TruD, which are structurally unrelated to Cy domains, covalently attach a phosphate or adenylate from ATP to the carbonyl oxygen to promote both cyclization and dehydration (Dunbar et al., 2012). In Cy domains, which do not use ATP, the energetics of cyclodehydration is presumably linked to the high-energy thioester bond that is broken in the subsequent step of bacillamide synthesis. Cy domains share this characteristic with the newly described specialty C domain, NocB-C5, that catalyzes β-lactam ring formation (Gaudelli et al., 2015). NocB-C5 is proposed to first dehydrate a serine to dehydroalanine and, before cyclization, use a histidine directly upstream of its HHxxxD motif. Dehydration before cyclization is the reaction order shared with lantibiotic cyclodehydratases, but not with Cy domains (Zhang et al., 2012). The position of the upstream histidine is occupied by a hydrophobic residue in Cy domains, and this V979

in BmdB-Cy2 is somewhat recessed from the active site. Thus, Cy domains rely on a completely different mechanism for catalysis than YcaO, TruD, NocB-C5 or the lantibiotic cyclodehydratases.

To help integrate our structural, mutagenic and conservation data, we created a model of the intermediate of the BmdB-catalyzed cyclodehydration reaction (Figure 5.4b). The configuration of the model and the present data are consistent with the absolutely critical catalytic action of D1226 in the first step of the cyclodehydration reaction, orienting and abstracting a proton from the thiol (or hydroxyl in Ser- or Thr-specific Cy domains) (Figure S5.7). T1196, 3.0 Å from D1226, appears well positioned to aid in catalysis by interacting with D1226, and may donate a proton to the former carbonyl oxygen to form the hydroxyl-thiazolidine. D1226 may protonate the same oxygen again to allow it to leave as water in the dehydration reaction (Figure 5.4b,c, Figure S5.7). Upon dehydration, the pKa of the amino proton is lowered such that it can be facilely lost to solvent to produce the final thiazoline moiety.

Our mutational results indicate that D1226 is critical for cyclodehydration, and T1196 is nearly as important (Figure 5.3a). However, T1196 is only a threonine or serine in 88% of Cy sequences, and appears as an alanine in 7% of cases, including in the functionally-characterized PchE-Cy1(Quadri et al., 1999). How can its apparent importance in the reaction and its relative lack of conservation be reconciled? Homology modelling of PchE-Cy1 onto BmdB-Cy2 provides a potential answer: The residue at the nearby position equivalent to 1228 of PchE-Cy1 is a glutamine, which could compensate for the missing T1196. A double mutant of T1196A/V1228Q did not restore cyclization activity to BmdB (Figure S5.8), but is not surprising given that BmdB-Cy has evolved to perform cyclodehydration with T1196 and not Q1228. In the analyzed Cy domain sequences which have an alanine at position 1196, all but two have a glutamine at 1228 (with the remaining two having glutamate or asparagine). In the majority of Cy domains, all three residues are present to form a T-D-Q triad. A similar S/T-D-Q catalytic triad in *E. coli* thioesterase II activates water for nucleophilic attack(Li et al., 2000). We suggest that mutation of the threonine or glutamine in the Cy domain triad may be tolerated in some instances, as chemistry is not rate-limiting in the overall slow synthetic cycle of NRPSs, and at least in thiazoline-forming Cy domains it would not be

difficult to deprotonate the Cys side chain (pKa of ~8.2 in aqueous solution). Deprotonation of Ser and Thr side chains is more difficult and could require the full catalytic triad. Variation in catalytic residues is not abnormal in nature, as exemplified by serine protease active sites (Ekici et al., 2008). Analogous to the observed variation of T1196, thioesterase II enzymes feature hydrophobic residues in place of the catalytic serine/threonine in over 5% of proteins.

Finally, parallels can be drawn between the role of D1226 in deprotonating the substrate side chain for nucleophilic attack and the mechanisms of microbial transglutaminase, cytosolic phospholipase A2, patatin, and TEM-1 beta-lactamase, in which the aspartate of a C-D or S-D dyad deprotonates the cysteine or serine for nucleophilic attack on a carbonyl carbon (Dessen et al., 1999; Minasov et al., 2002; Rydel et al., 2003).

During the review process for this study, a structure and mutational analysis of the heterocyclization domain of epothilone synthase (EpoB-Cy) was reported by Dowling *et al*. (Dowling et al., 2016). The structures of EpoB-Cy and BmdB-Cy2 are very similar. Their mutational data are consistent with those presented here, in particular the identification of D1226 (D449 in EpoB-Cy) as important for catalysis, though without differentiation between defects in condensation and heterocyclization (Supplemental Result and Discussion, Table S5.1). Their study and ours complement one another and provide a greater understanding of NRPS heterocyclization domains.

5.2.7 Conclusion

In summary, we have presented the crystal structure of an NRPS heterocyclization domain, solved to a resolution of 2.3 Å. We cloned, expressed, and purified the entire bacillamide synthetase containing this Cy domain, and used it to assay effects of Cy domain mutations on peptide production. We were able to identify two residues, T1196 and D1226, which are very important for the cyclodehydration reaction. Finally, we presented a putative mechanism for cyclodehydration in which D1226 acts as general acid/base catalyst.

5.3 Methods

5.3.1 Bacillamide synthetase Cy2 crystallography

The *Thermoactinomyces vulgaris* BmdB-Cy2 construct, with an N-terminal octahistidine tag, tobacco etch virus (TEV) protease cleavage site and BmdB residues 844-1287, was synthesized by DNA 2.0. BmdB-Cy2 was heterologously expressed in *E. coli*, and purified to homogeneity (SI Appendix, Supplementary Methods). BmdB-Cy2 was crystallized using 0.88% Tween 20, 1.62 M ammonium sulfate, 0.1 M HEPES pH 7.5, 2.67% PEG 400, and 3% 6-aminohexanoic acid, with 20% (v/v) ethylene glycol, added before flash-cooling. Diffraction data were collected at CLS beamline 08ID-1 (Table S5.2) and the structure of BmdB-Cy2 was determined by molecular replacement.

5.3.2 Full-length BmdB

bmdB was cloned from *Thermoactinomyces vulgaris* F-5595 genomic DNA into a pET21-derived vector containing an N-terminal TEV-cleavable calmodulin binding peptide tag and a C-terminal TEV-cleavable octa-histidine tag. Mutations were introduced via site-directed mutagenesis (Table S5.3). Wild-type and mutant protein was expressed in *E. coli* cells and purified to homogeneity.

5.3.4 Peptide synthesis assay for BmdB

In a 1 mL reaction, 50 mM Tris-HCl pH 7.5, 100 mM NaCl, and 10 mM $MgCl_2$, 1 mM Ala, 1 mM Cys, 10 mM tryptamine, 2 mM ATP, and 100 nM BmdB were incubated for 2 hours at 37 °C. Reactions were analyzed by HPLC using C18 media and a gradient of water/0.1% TFA to acetonitrile/0.1% TFA. High resolution MS and fragmentation MS analysis was performed at the Proteomics Platform at the RI-MUHC, and NMR analysis at QANUC.

5.3.5 Bioinformatic Analysis

Sequences of known C domains were queried against the nr90 database(Biegert and Soding, 2009), using low threshold search. Matched sequences were themselves filtered for maximum pairwise sequence identity of 90%, which gave 36,853 C domain

superfamily sequences. These were classified into subtypes using NaPDoS (Ziemert et al., 2012), and included 1790 Cy domains. To look for trends in Cy domains by substrates, all full-length protein sequences from which the 1790 Cy domains originate were retrieved. Sequences were discarded if there was not an A domain adjacent to the Cy domain, or if they started as a single Cy domain leaving a set of 505 proteins. The Cy domain acceptor / cyclodehydration substrates were inferred from substrate of the A domains in the same module, as predicted by the program ANTISMASH (Weber et al., 2015).

5.3.7 Data deposition

The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank, www.pdb.org (accession code 5T3E).

5.3.8 Author Contribution

K.B. performed all molecular biology, biochemistry and crystallography experiments. T.M.S planned the project and performed the bioinformatics. C.D.F and M.A.M. provided intellectual contributions, experimental advice and manuscript editing. K.B. and T.M.S. wrote the manuscript.

5.3.9 Acknowledgements

We thank Chaitan Khosla and James Kuo for the kind gift of BAP1 cells; Vikram Alva and Johannes Soeding for providing the current nr90 database, Christopher Boddy, Adrian Keatinge-Clay for conversations; Asfarul Haque for performing the ITC experiments; Janice Reimer for schematic figure design; the Proteomics Platform at the RI-MUHC (Montreal, Canada) for LC-MS analysis; members of the Schmeing lab for helpful discussions; Alexander Wahba for MS analysis, Varoujan Yaylayan for fragmentation MS assignment and Robin Stein for NMR analysis.

5.4 Supplemental Results and Discussion

5.4.1 Expanded structural description of Cy domains

Like known condensation domains, BmdB-Cy2 consists of two structurally similar subdomains, the N-terminal subdomain, consisting of residues Leu846-Gly1027 and the C-terminal subdomain, consisting of residues Thr1028-Gln1279 (Figure 5.2a). The two subdomains arrange themselves in a V-shaped fashion (Keating et al., 2002) (Figure 5.2a). In C domains, the angle of the V can vary as these two lobes assume a range of relative conformations termed more “open” or “closed” (Bloudoff et al., 2013). The conformations are achieved essentially by a pivot of one subdomain relative to the other centered on the residue corresponding to BmdB-Cy2 Gly1027. Superimposition of BmdB-Cy2 with each C domain (Bloudoff et al., 2013; Drake et al., 2016; Keating et al., 2002; Samel et al., 2007; Tanovic et al., 2008) and epimerization domain(Chen et al., 2016; Samel et al., 2014) structure shows that it fits within the observed range of conformations, and is quite similar to AB3403(Drake et al., 2016) (Figure S5.2). The active site is found at the interface between the two subdomains, in the center of a channel linking the donor and acceptor PCP binding sites, which are located at the “front” and “back” face of the V.

The core of both of these subdomains adopt the classic chloramphenicol acetyltransferase fold, which has a large central ß sheet with adjacent α-helices (Leslie, 1990). The C-terminal subdomain has a ß sheet with 6 strands (residues 1061-1072; 1111-1117; 1136-1142; 1194-1199; 1225-1232; 1235-1242) and 6 adjacent α-helices. The strands at residues 1194-1199 and 1225-1231 form one surface of the active site. The N-terminal subdomain has a core ß sheet with 5 strands (residues 875-882; 922-928; 961-969; 972-980; 1214-1219), 5 adjacent α-helices, and a second, smaller ß sheet. The DxxxxD active site motif (residues 980-985) is a loop directly adjacent to the strand at 972-980. Notably, the last strand in the core sheet, residues 1214-1219, is donated from the C-terminal subdomain. In all, residues 1204-1222 cross-over and form part of the N-terminal subdomain, despite being in the C-terminal half of the Cy domain sequence. This crossing-over forms a roof on the active site, which is known as the latch. It has been postulated that the latch could disengage from the N-terminal

subdomain during condensation (Samel et al., 2007), but to date that has not been observed (Bloudoff et al., 2013).

5.4.2 Comments on potential mode of action of Cy domain mutants

As discussed in the main text, the bacillamide synthetase mutants had radically different effects on condensation and cyclodehydration (Table S5.1, Figure S5.5). Without crystal structures of each mutant bound to a suite of reaction analogues, one cannot conclude why a particular mutation has moderate deleterious effects. Logical rationalizations of these effects are possible, and though speculative, a full discussion with rationalizations may be useful and appropriate to include here in the Supplemental Results and Discussion section: Mutant S1197A had almost no effect on production of linear or heterocyclized tripeptide, despite the corresponding mutation in BacA2 decoupling condensation and cyclization(Duerfahrt et al., 2004) (Table S5.1). We believe the discrepancy is from differential secondary effects on the adjacent T1196, mutation of which did decouple the reactions in BmdB (Duerfahrt et al., 2004; Kelly et al., 2005). With Y859F, bacillamide E production decreased to 79% of wild-type, but linear Ala-Cys-Tpm production doubled. Although the hydroxyl of Y859 points directly into the heart of the active site, the relatively minor effect of its removal suggests that it could aid in configuring the active site or substrates, but it is unlikely to act chemically. Conversely, mutants N1114A and T1116A each drastically reduced synthesis of both the linear and heterocyclic product. T1116A maintained approximately the same ratio of heterocyclic to linear product as wild-type, whereas the little product made by N1114A was successfully heterocyclized, consistent with previously reported decoupling of condensation and heterocyclization by this residue (Duerfahrt et al., 2004). Both N1114 and T1116 thus appear important for condensation, and N1114 is also important for cyclodehydration. They may help orient the substrates for reaction, a role similar to that proposed for the second histidine of HHxxxD in C domains (Bloudoff et al., 2016), and/or to maintain active site solvation (Duerfahrt et al., 2004).

For insight into the cyclization reaction, the most interesting mutations are those that selectively affect cyclodehydration. Mutation of residues F1118, T1135, T1196 and D1226 resulted in substantially more linear Ala-Cys-Tpm and less bacillamide E than

the wild-type enzyme. T1135 is adjacent to the binding site of the phosphopantetheine arm and its mutation may disfavour the proper conformation of the active site and substrate. F1118 is positioned along the channel that the phosphopantetheine arm of donor PCP1 occupies when presenting the Ala substrate to the active site (Figure S5.5). In the structure of BmdB-Cy2, the F1118 side chain completely blocks that channel. This residue must move to allow alanyl-PCP1 to bind and participate in the condensation reaction. However, for cyclization, PCP1 likely departs, leaving the cyclization substrate Ala-Cys-PCP2 bound at the acceptor site. It is possible that F1118 aids cyclization by blocking the donor channel to create a single-entry active site more optimal for cyclodehydration. This is reminiscent of the (permanent) blocking of the acceptor site by a tryptophan side chain to form a single-entry active site in related epimerization domains, observed in the recently-published crystal structure (Chen et al., 2016) (GrsA W632; BmdB A858).

5.4.3 Comparison to a recently-published Cy domain.

During the review process for this study, a structure of the heterocyclization domain of epothilone synthase (EpoB-Cy), along with its N-terminal PKS-NRPS docking sequence (EpoB-dd) was reported by Dowling *et al*. (Dowling et al., 2016). The structure and accompanying molecular dynamics simulations show that EpoBdd only acts as a covalent tether to the Cy domain and does not otherwise make important contacts with it. EpoB-Cy and BmdB-Cy2 share 30% sequence identity. The structures of the domains are very similar, with EpoB-Cy showing the typical C domain superfamily topology and putative PCP binding sites at either end of the active site tunnel. Dali pairwise comparison of these two Cy domains gives a root mean squared deviation (RMSD) of 2.0 Å over 423 C-alpha atoms (Holm and Rosenstrom, 2010). The difference is largely due to EpoB-Cy assuming a slightly more “open” configuration of its two subdomains (Bloudoff et al., 2013). The only other difference of note appears to be the conformation of single loop (BmdB-Cy2 1199-1209; EpoB-Cy 420-432). In EpoB-Cy, this loop largely blocks the previously-observed donor PCP binding site(Bloudoff et al., 2013; Chen et al., 2016) while in BmdB-Cy2 there is a smaller overlap. It is likely that this loop is mobile and may move to allow canonical donor PCP binding. The Asp-motifs

of both structures are in similar conformations, with the aspartates pointing away from the actives site, also leading Dowling *et al*. to conclude this motif is structural and does not participate directly in catalysis.

Dowling *et al*. present mutational data are consistent with those presented here, in particular identifying N1114 (N335 in EpoB-Cy) as important for substrate positioning and D1226 (D449 in EpoB-Cy) as important for catalysis (Table S5.1). They suggest that D1226 could act as a catalytic base during both condensation and cyclodehydration. However, our results clearly indicate that the D1226G and D1226N mutants have robust condensation activity, and are only unable to catalyze cyclodehydration (Figure 5.3). D1226 is thus critical for cyclodehydration and not condensation. Dowling *et al.* did not differentiate between defects in condensation and cyclodehydration, which explains the difference in proposals. Together, the Dowling *et al.* study and our present study complement one another and provide greater understanding of NRPS heterocyclization domains.

5.5 Supplemental Methods

5.5.1 Bacillamide synthetase Cy2 crystallography

We propose to call bacillamide synthetase BmdB (WP_054096387.1); the adjacent tryptophan decarboxylase that makes the tryptamine substrate, BmdA (WP_022737640.1); and the adjacent oxidation enzyme, BmdC (WP_022737638.1). The *Thermoactinomyces vulgaris* BmdB-Cy2 construct, with an N-terminal octahistidine tag, tobacco etch virus (TEV) protease cleavage site and BmdB residues 844-1287, was synthesized by DNA 2.0. BmdB-Cy2 was heterologously expressed in BL21 (DE3) *E. coli* cells grown in LB medium and 300 μg ml^{-1} kanamycin at 37 °C. When the culture reached an OD_{600} of ~0.5–0.6, the temperature was reduced to 16 °C, and expression was induced with the addition of 1 mM IPTG, overnight. Cell pellets were resuspended in IMAC buffer B (50 mM Tris–HCl pH 7.5, 100 mM NaCl, 2 mM β-mercaptoethanol (βME), 1 mM imidazole pH 8.0 and 1 mM phenylmethylsulfonyl fluoride (PMSF)), lysed by sonication and centrifuged for 30 min at 40,000g. The supernatant was pooled and applied onto a two tandem 5-ml HiTrap IMAC FF columns charged with Ni^{2+}. BmdB-Cy2 was eluted with IMAC buffer B (50 mM Tris–HCl pH 7.5, 100 mM NaCl, 2 mM βME, 250

mM imidazole pH 8.0 and 1 mM PMSF). BmdB-Cy2 was pooled and digested overnight at 4 °C in dialysis buffer (50 mM Tris–HCl pH 7.5, 100 mM NaCl, 2 mM βME) with N-His-TEV protease (1 mg per 40 mg BmdB-Cy2). Cleaved sample was reapplied onto the HiTrap IMAC FF columns, and the flow-through was collected. Sample was diluted 2-fold and applied to two tandem 5-ml HiTrap Q HP columns equilibrated with 50 mM Tris–HCl pH 7.5, 2 mM βME and eluted with 50 mM Tris–HCl pH 7.5, 1M NaCl, 2 mM βME. BmdB-Cy2 was then applied to a 16/60 Superdex 200 column equilibrated with 50 mM Tris-HCl pH 7.5, 100 mM NaCl, and 1 mM DTT. BmdB-Cy2 was pooled, concentrated, and stored at -80 °C.

BmdB-Cy2 (2 µl of 10 mg ml^{-1}) was crystallized by mixing with 2 µl of 0.88% Tween 20, 1.62 M ammonium sulfate, 0.1 M HEPES pH 7.5, 2.67% PEG 400, and 3% 6-aminohexanoic acid, and equilibrated against 500 µl of 0.88% Tween 20, 1.62 M ammonium sulfate, 0.1 M HEPES pH 7.5, 2.67% PEG 400, and 3% 6-aminohexanoic acid. Crystals were transferred into a solution of mother liquor plus 20% (v/v) ethylene glycol, and flash-cooled in liquid nitrogen.

Diffraction data were collected at Canadian Light Source beamline 08ID-1 (Table S5.2). The structure of BmdB-Cy2 was determined by molecular replacement with the program Phaser (Adams et al., 2010) using CDA-C1 (PDB code: 4JN3) (Bloudoff et al., 2013) as a search model, followed by iterative modelling using the program AutoBuild (Terwilliger et al., 2008) and the program Coot (Emsley et al., 2010) and refinement using CNS 1.3(Brunger et al., 1998) and Phenix (Terwilliger et al., 2008).

A homology model of BmdB-PCP2 was created using SWISS-MODEL (Arnold et al., 2006) and a homologous PCP domain (Drake et al., 2016). BmdB-PCP2 was placed by superimposition of the Cy / C domains (Drake et al., 2016), and subjected to gradient energy minimization using the program CNS 1.3 (Brunger et al., 1998). A model of the cyclodehydration intermediate of Ala-Cys- phosphopantetheinyl-PCP was constructed and restraints were generated using PRODRG (Schuttelkopf and van Aalten, 2004). This entire complex was subject to structure idealization refinement using the program Refmac (Murshudov et al., 1997).

5.5.2 Full-length BmdB

Thermoactinomyces vulgaris F-5595 (ARS Culture Collection) was grown on ISP2 media overnight at 50 °C and its genomic DNA was using the GenElute Bacterial Genomic DNA Kit. *bmdB* was amplified with primers VulgarNRPS_F and VulgarNRPS_R (Table S5.3) and cloned into a pET21-derived vector containing an N-terminal TEV-cleavable calmodulin binding peptide tag and a C-terminal TEV-cleavable octa-histidine tag. Mutations were introduced via site-directed mutagenesis (Table S5.3). Wild-type and mutant protein was expressed in BAP1 *E. coli* cells (Pfeifer et al., 2001) grown in LB medium with 300 μg ml^{-1} kanamycin at 37 °C. Protein expression was induced at an OD_{600} of ~0.5–0.6 with the addition of 1 mM IPTG, and continued overnight at 16 °C.

Cell pellets were resuspended in a IMAC buffer A plus 2 mM $CaCl_2$. Cells were lysed by sonication and centrifuged for 30 min at 40,000g and 4 °C. The supernatant was pooled and applied onto two tandem 5-ml HiTrap IMAC FF columns charged with Ni^{2+}. BmdB was eluted with IMAC buffer B plus 2 mM $CaCl_2$, applied to a 30 ml calmodulin sepharose 4B column and eluted with 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM EGTA, and 2 mM βME. BmdB was digested overnight at 4 °C in dialysis buffer with N-His-TEV protease (1 mg per 40 mg BmdB). Cleaved BmdB was passed back over the affinity columns, with the flow-through collected, concentrated and applied to a 16/60 Superdex 200 column equilibrated with 50 mM Tris-HCl pH 7.5, 100 mM NaCl, and 1 mM TCEP. BmdB was concentrated to 10 mg ml^{-1}, flash-frozen in liquid nitrogen and stored at −80 °C.

5.5.3 Peptide synthesis assay for BmdB

The peptide synthesis assay was adapted from that of Duerfahrt *et al*.(Duerfahrt et al., 2004) In a 1 mL reaction, 100 nM BmdB was incubated for 2 hours at 37 °C in 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM ATP, 1 mM Ala, 1 mM Cys, 10 mM tryptamine, and 1 mM (Figure S5.8) or 10 mM (all other reactions) $MgCl_2$. After 2 hours, 10 μL of 1 M DTT was added, followed by incubation for 15 minutes at 65°C, before 500 μL of 4:1 (v/v) mixture of n-butanol and chloroform was added. Samples were

evaporated to dryness and then resuspended in 10% methanol. Insoluble material was removed by centrifugation (13000 rpm, 15 min).

HPLC analysis (Figure 5.3, Figure S5.3a) of the reaction was performed with the Agilent 1260 infinity LC using reverse-phase chromatography (ZORBAX 80Å Extend-C18, 5µm, 4.6 x 250 mm; Agilent) at a flow rate of 0.5 ml min^{-1}, a sample injection volume of 45 µL and buffers A (water/0.1% TFA) and B (acetonitrile/0.1% TFA). Peptides were separated by a gradient of 10-60% buffer B over 30 minutes, followed by a gradient of 60-95% buffer B over 5 minutes and isocratic flow at 95% B for 5 minutes. Calculation of the area under the curves was performed by Agilent OpenLAB CDS ChemStation Edition software. Peptide masses were determined offline using ESI-MS (amaZon speed ETD; Bruker Daltonics) in positive-ion mode with an ESI nebulizer. The mass spectrometer was set to acquire spectra in the mass range 200–1000 m/z for an average of 1 min, and the sample was infused through a syringe pump at 240 µl h^{-1}. High resolution mass spectrometry of **1** (Figure S5.3c) was performed by the McGill University Department of Chemistry Mass Spectroscopy Facility.

In-line HPLC/MS (Figure S5.3b) of aliquots of the same samples of the peptide synthesis assay of BmdB was commissioned and performed by the Proteomics Platform at the RI-MUHC (Montreal, QC, Canada), where a different column, buffer system, gradient and protocol was used: Equal amounts of samples dissolved in 0.1% aqueous formic acid were analyzed using a Shimadzu 2080 LC system and a 2.1 m X 50 mm HALO2 PFP column operated at 200 uL/min using a gradient of 0%-30% acetonitrile/0.1% formic acid over 30 min into a 5600 TripleTOF (SCIEX) running in high resolution mode (25,000 FWHM). Both peptide species were targeted using an MRMHR workflow. [M+H] parent masses for both the linear (335.1) and cyclic (317.1) peptides were placed into an include list and product ions specific for each peptide were used to confirm identity for each peak. Extracted ion currents for each parent ion were extracted at 0.1amu to provide relative concentrations for each species.

A large bacillamide synthesis reaction was performed to obtain sample for NMR analysis. Bacillamide E and Ala-Cys-Tpm were not fully resolved by HPLC, so the ^{1}H NMR showed peaks for Bacillamide E as the major species (M) and Ala-Cys-Tpm as the minor (m) species (Figure S5.9): ^{1}H NMR (500 MHz, Acetonitrile-d3+D2O) δ 7.581

(m, 1.69, 1M+1m), 7.39 (d, J = 8.2 Hz, 1.66, 1M+1m), 7.13 (m, 3.12, 2M+2m), 7.049 (m, 1.65, 1M+1m), 5.025 (m, 0.54, 1m), 4.363 (m, 1.10, 1M), 4.245 (m, 0.63, 1m), 4.07 (s, D2O), 3.993 (q, J=6.6 Hz, 1.18, 1M), 3.686 (dd, J=11.27, J=9.91, 0.57, 1m), 3.58 – 3.36 (m, multiple peaks, 4.22, 2M+3m), 2.93 (m, multiple peaks, 3.22, 2M+2m) 2.72 (m, 2.15, 2M), 1.97 (q, CD3CN), 1.44 (d, J = 7.2 Hz, 1.41, 3m), 1.40 (d, J = 7.1 Hz, 3.0, 3M).

5.5.4 Bioinformatic Analysis

Sequences of characterized Cy domains were aligned with the program Muscle (Edgar, 2004). We then undertook a bioinformatic analysis of C superdomain family proteins, with special focus on Cy domains. Earlier analyses have been very useful for identifying and specifying the various clades of C superdomain families, which include Cy, starter C, $^{L}C^{L}$, $^{L}C^{D}$, E, dual C-E, and hybrid C domains (Rausch et al., 2005; Wang et al., 2014; Ziemert et al., 2012). The PF00668 domain (C domain superfamily) sequences and alignments did not cover all the relevant sequence of C domains. To produce a large set of full-length, non-redundant C domain superfamily sequences, sequences of known C domains of each major type were used as inputs to search using BLAST (Altschul et al., 1990) against a database of known proteins, filtered for a maximum pairwise sequence identity of 90% (nr90) (Biegert and Soding, 2009), using low threshold search parameters (E value of 1; word size of 2, gap open of 9 and gap extend of 1). Matched sequences of >400 amino acids were themselves filtered for maximum pairwise sequence identity of 90% using the EFI web server (Gerlt et al., 2015), which gave 36,853 C domain superfamily sequences. These were classified into subtypes using the Natural Product Domain Seeker server (Ziemert et al., 2012), and included 1790 Cy domains. The program WebLogo3 (Crooks et al., 2004) was used to draw sequence conservation schematics for all C domain types.

To look for trends in Cy domains by substrates, all full-length protein sequences from which the 1790 Cy domains originate were retrieved. Sequences were discarded if there was not an A domain adjacent to the Cy domain, or if they started as a single Cy domain (to rule out the uncertainty of a tandem Cy domain provided in trans), leaving a set of 505 proteins. The Cy domain acceptor / cyclodehydration substrates were

inferred from substrate of the A domains in the same module, as predicted by the program ANTISMASH (Weber et al., 2015).

5.6 Supplemental Figure and Table Legends

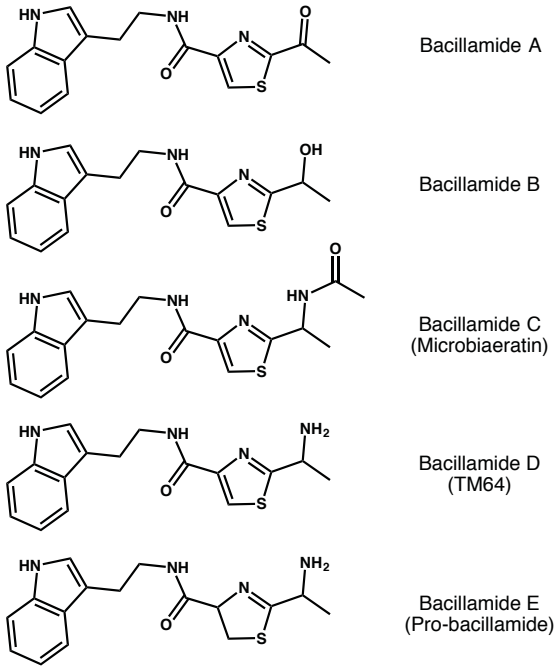


Figure S5.1. The bacillamide family of compounds. Bacillamide synthetase is shown in this study to make bacillamide E, which can be modified to produce bacillamides A-D(Ivanova et al., 2007; Jeong et al., 2003; Konda et al., 1976; Martinez and Davyt, 2013; Socha et al., 2007; Yuwen et al., 2013).

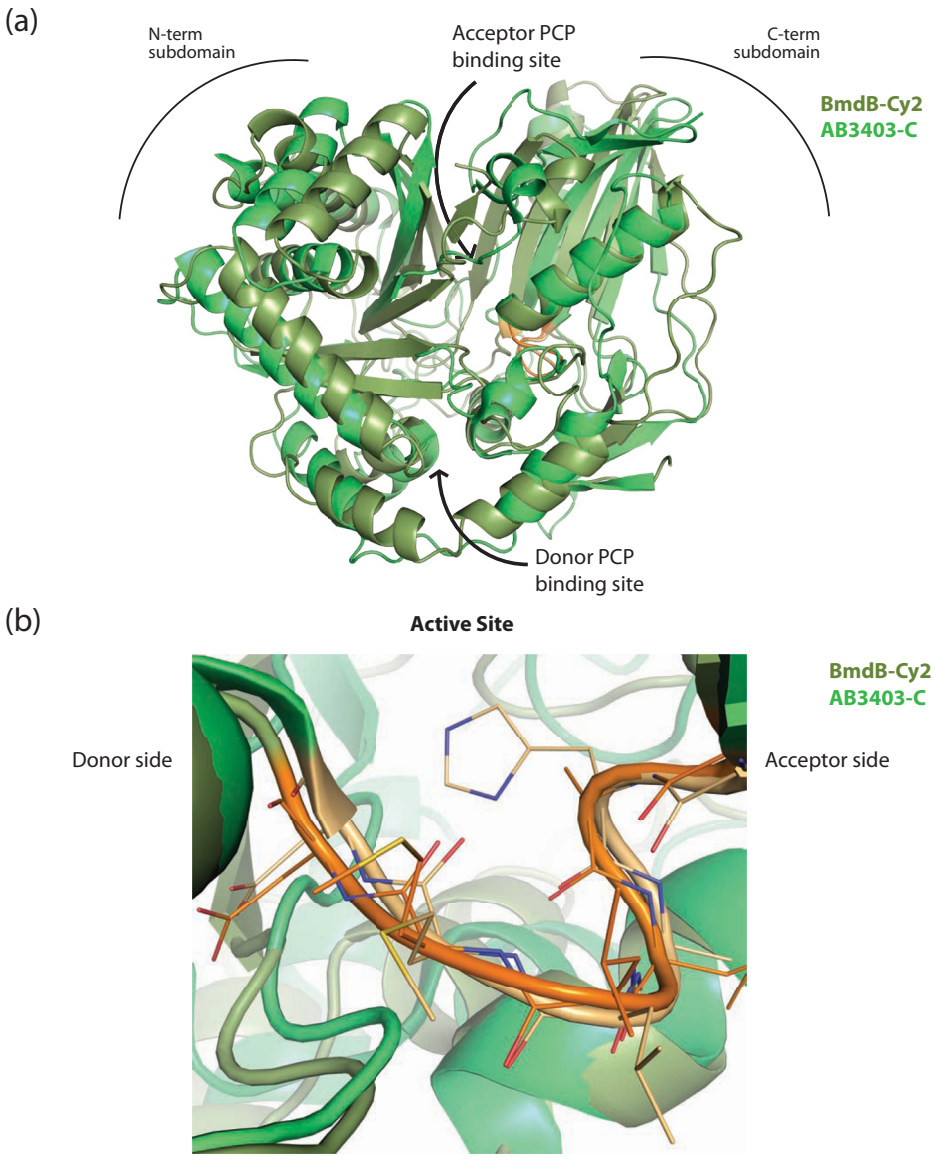
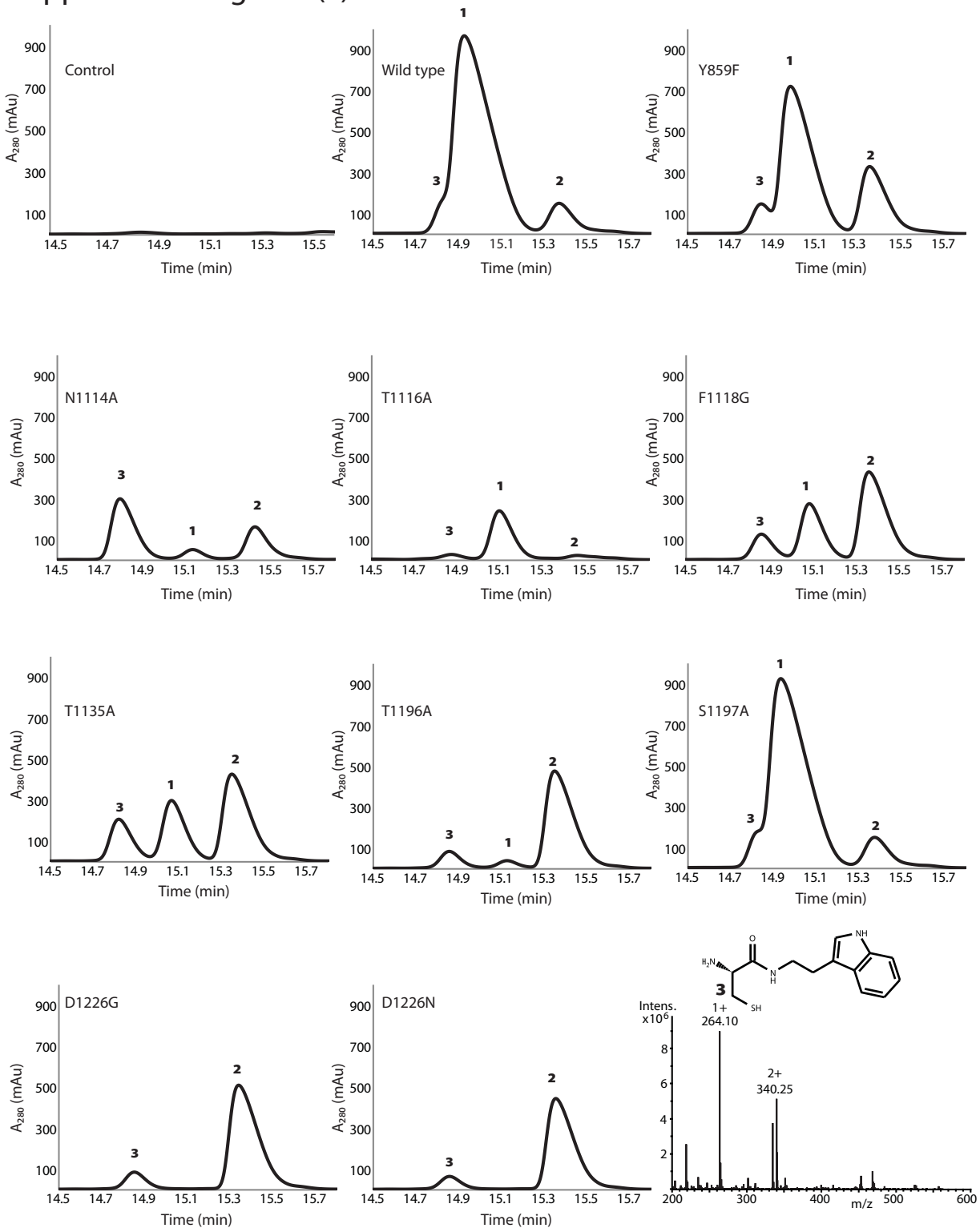
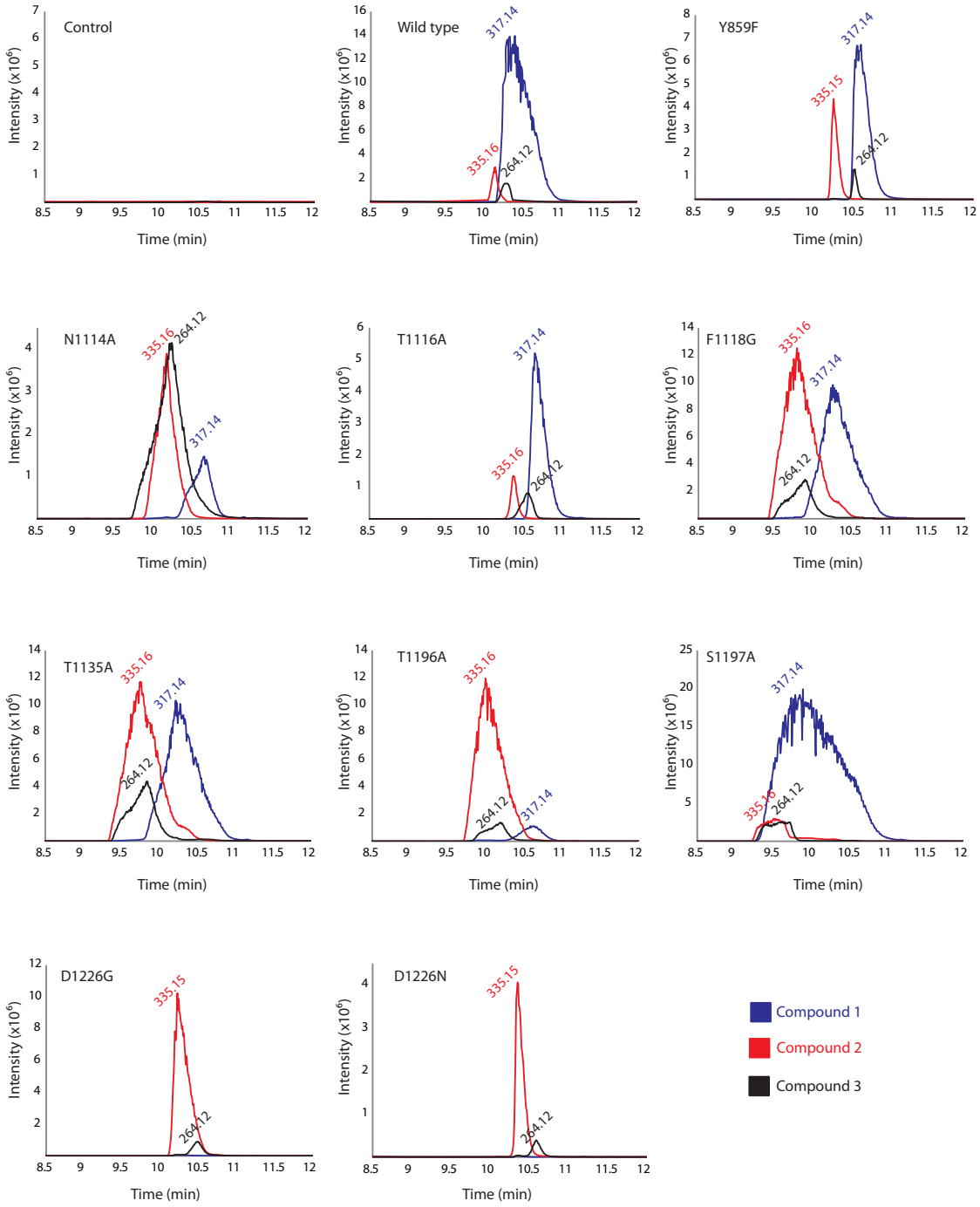


Figure S5.2: Structural comparison of BmdB-Cy2 and AB3403-C domain (PDB: 4ZXH). (a) Alignment of C-terminal lobes of BmdB-Cy2 (darker green) and AB3403-C(Drake et al., 2016) (lighter green). (b) Alignment of active sites of BmdB-Cy2 (darker green) and AB3403-C (lighter green). The Asp-motif of BmdB-Cy2 (DALLMD) is shown in orange cartoon and sticks and the His-motif of AB3403-C (HHALMD) is shown in beige cartoon and sticks. Dali pairwise comparison of the two structures gives an RMSD of 3.2 Å over 396 residues(Holm and Rosenstrom, 2010). Pairwise alignment of 107 residues in the N-terminal lobe results in an RMSD of 1.97 Å.

Supplemental Figure 3(a).



Supplemental Figure 3(b).



Supplemental Figure 3(c).

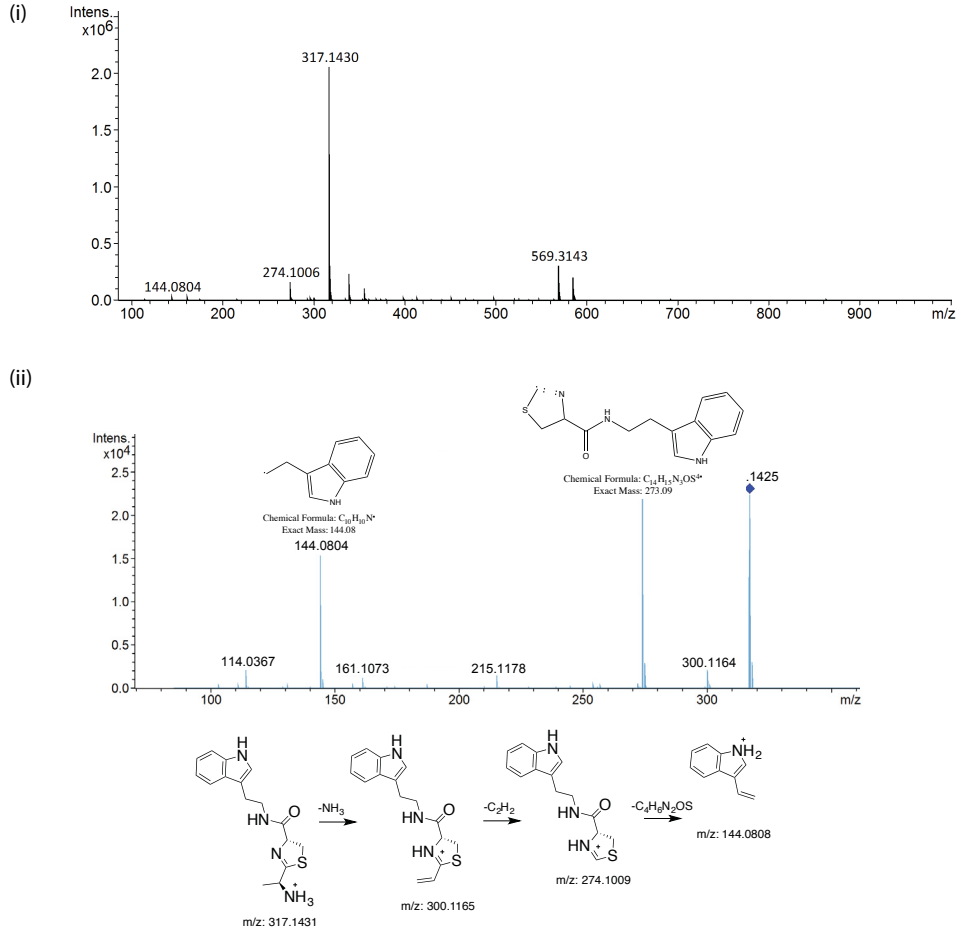


Figure S5.3. Representative traces for peptide synthesis assays for all BmdB mutants. (a) Representative HPLC traces for bacillamide synthetase assays. Peak 3 was determined to be Cys-Tpm, an assay byproduct likely produced by C3. MS for this compound is shown in the bottom right. We note that the amount of 3 produced varies in the different mutants. It is possible that higher production of Cys-Tpm may be caused by a decrease in cysteinyl-PCP2 binding to the mutated Cy2 active, which would the favour the cysteinyl-PCP2 binding to C3 that leads to production of Cys-Tpm. (b) Representative extracted ion currents for bacillamide synthetase assays. Mass value shown in traces has a +1 charge. Exact expected masses are 316.14 (for 1), 334.15 (for 2), and 264.12 (for 3). Note that due to the differences in column and buffer, the order of elution of peaks 1 and 2 is reverse between the analysis performed in the Schmeing laboratory (Figure 5.3b, Figure S5.3a) and that performed by the Proteomics Platform at the Research Institute of the McGill University Health Centre (this panel). (c) (i) High-resolution mass spectrometry of 1. Measured m/z of 317.1430 is predicted to have an ion formula of $C_{16}H_{21}N_4OS$ (m/z of 317.1431, with an error of 0.1 ppm, mSigma of 28.3, and score of 100.00). (ii) Fragmentation MS/MS of bacillamide E, with a putative fragmentation pathway.

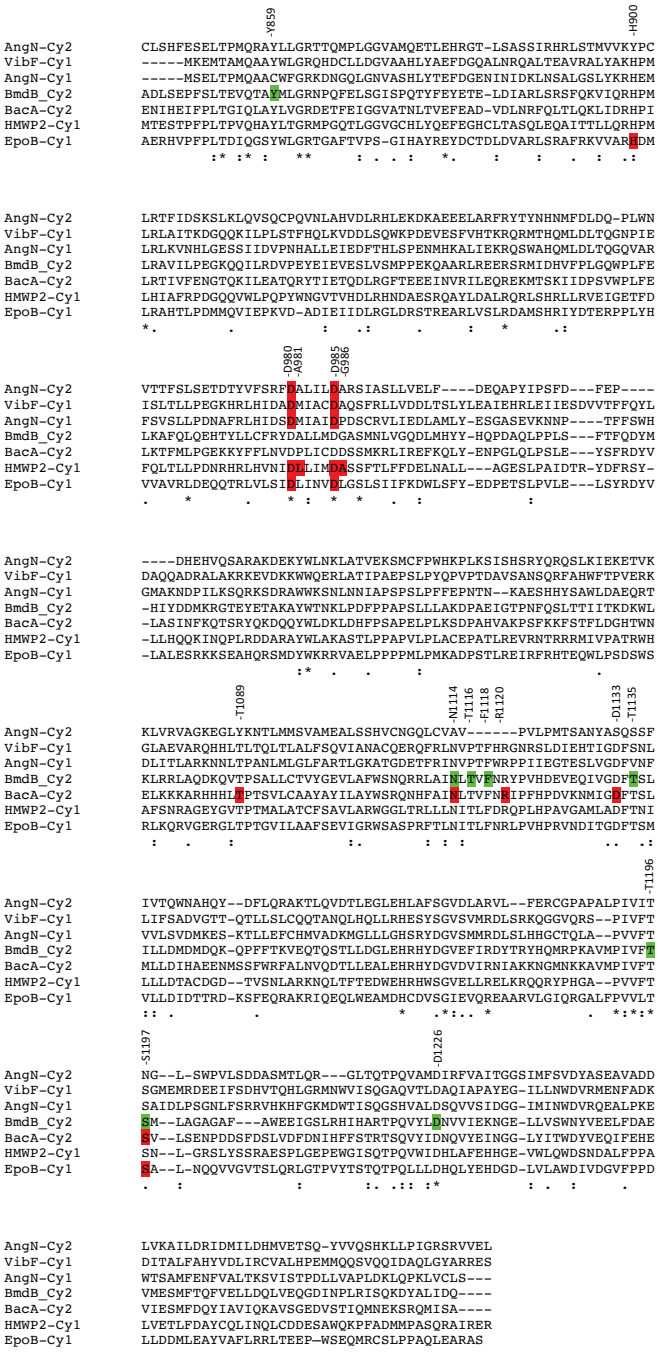


Figure S5.4. Sequence alignment of BmdB-Cy2 and other biochemically characterized Cy domains. Sequences of characterized Cy domains were aligned with the program Muscle (Edgar, 2004). Mutations performed in previous studies (red) (Bloudoff et al., 2016; Duerfahrt et al., 2004; Gehring et al., 1998; Keating et al., 2000b; Kelly et al., 2005; Konz et al.; Marshall et al., 2001; Marshall et al., 2002b) and mutations performed in this study (green) are highlighted, and numbered as in BmdB-Cy.

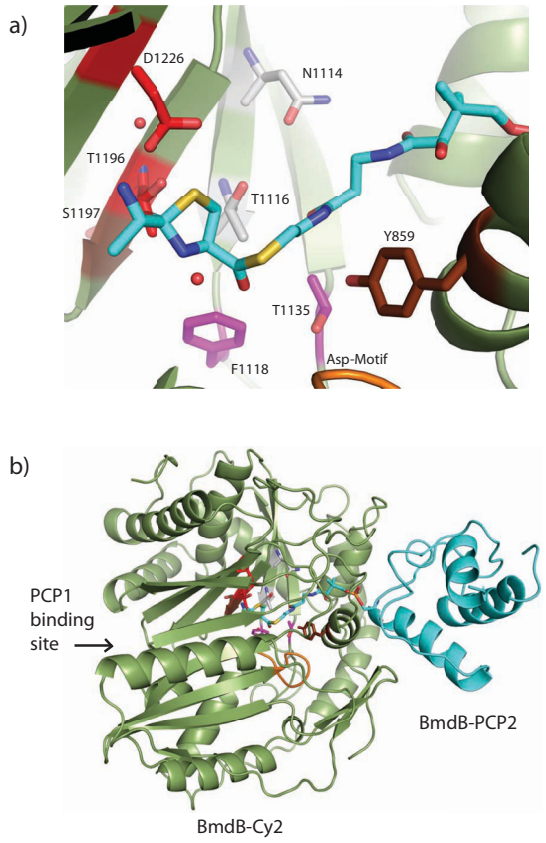
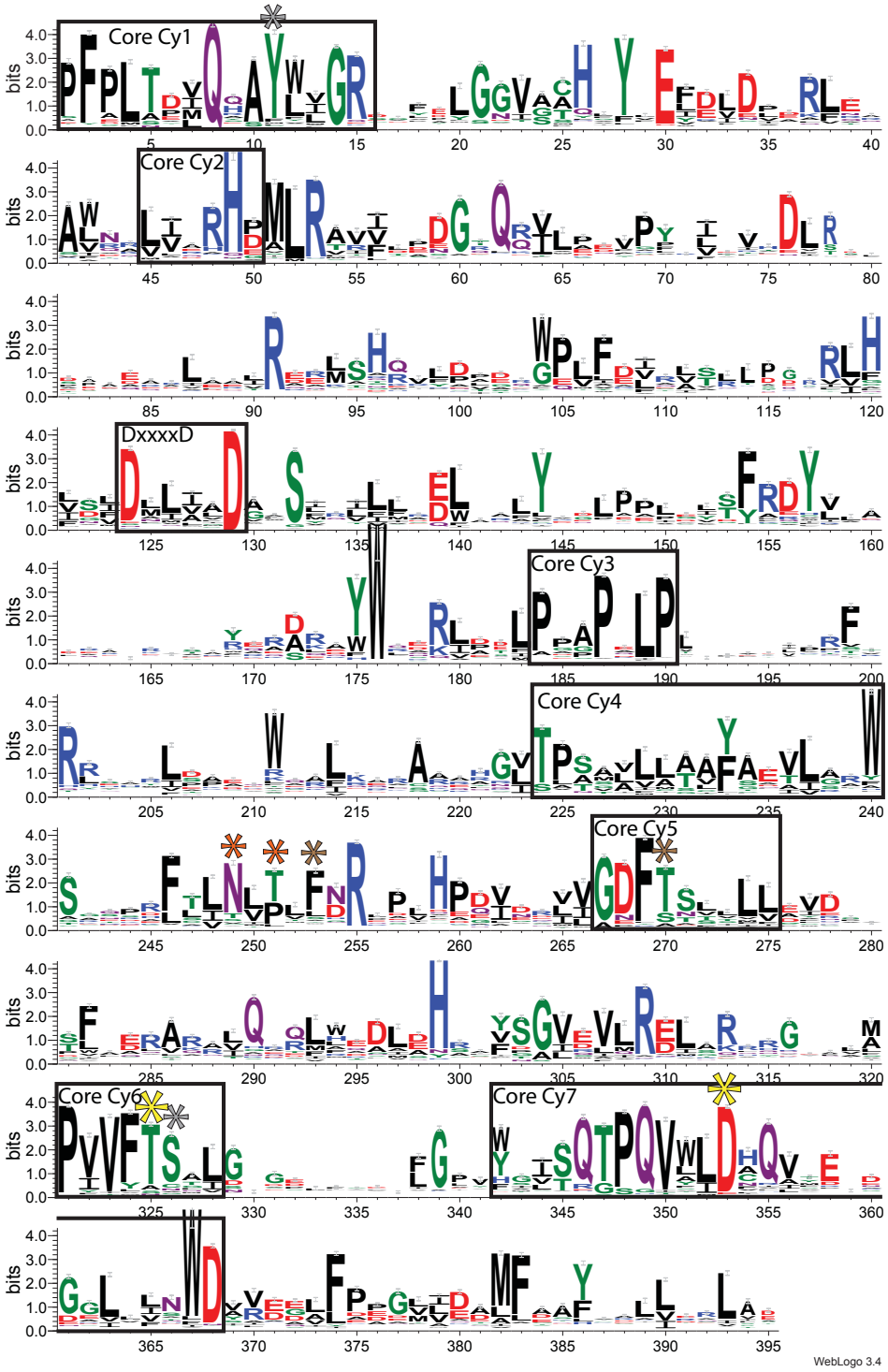


Figure S5.5. Model of BmdB-Cy2 in complex cyclodehydration intermediate. Superimposition of BmdB-Cy2 (mostly green) and a homology model of BmdB-PCP2 onto the C-(A)-PCP-(TE) structure (Drake et al., 2016) and refinement of the whole complex including the built cyclodehydration intermediate (cyan) produces a mid-reaction model. (a) Zoomed view of mutations performed in this study, colored according to their effect: minor effect on total product synthesis (brown; Y859F and S1197A), major effect on total product synthesis (grey; N1114A and T1116A), lowered cyclodehydration (magenta; T1135A and F1118G), and drastically lowered cyclodehydration (red; D1226G/N and T1196A). (b) Overview of modelled BmdB-Cy2 in complex with BmdB-PCP2 (cyan) shown in (a). Residues are colored as listed above.

Supplemental Figure 6a

Cy domain conservation



Supplemental Figure 6b

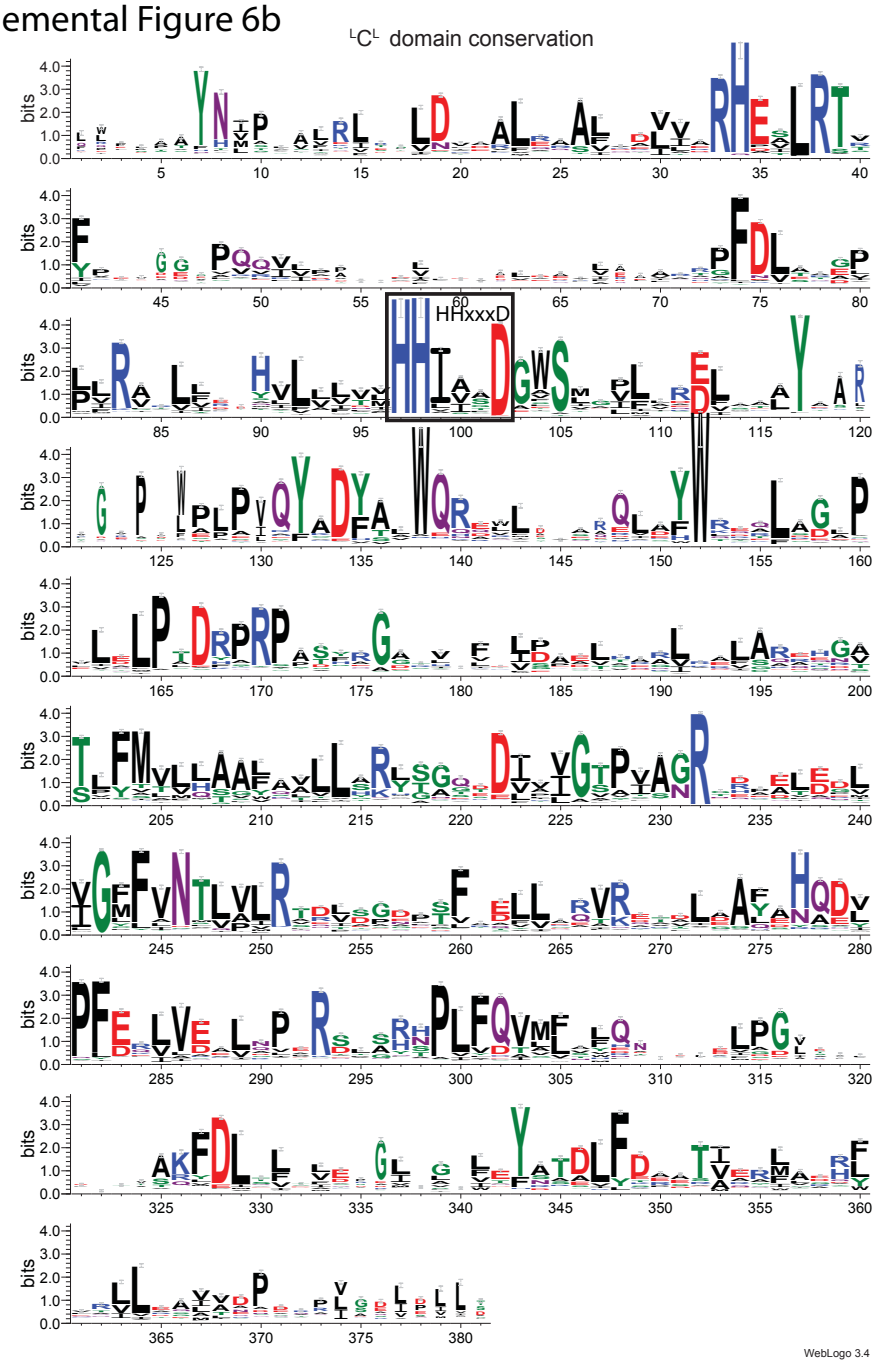


Figure S5.6. Sequence conservation in Cy and $^{L}C^{L}$ domains. (a) Conserved residues of Cy domains represented by WebLogo. Core motifs Cy1-Cy7 and DxxxxD motif are highlighted by black boxes. Mutations performed in this study are highlighted by asterisks. (b) Conserved residues of $^{L}C^{L}$ domains represented by WebLogo. The His-motif is highlighted by the black box.

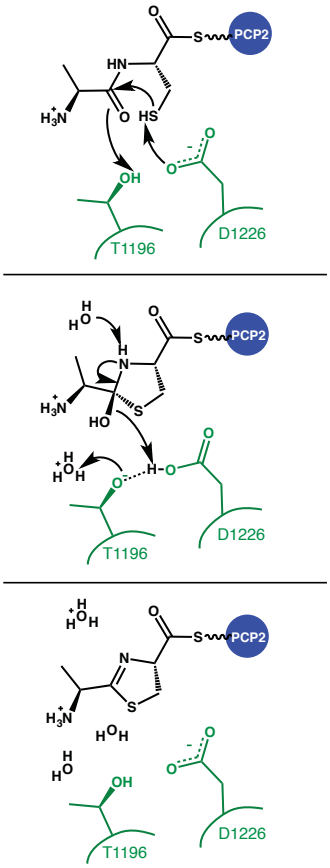


Figure S5.7. Putative reaction mechanism of heterocyclization in Cy domains. Nucleophilic attack of the substrate Cys side chain on the carbonyl carbon is promoted by orientation and deprotonation by D1226, and donation of a proton from T1196 to the carbonyl oxygen. D1226 can then donate the proton it has abstracted to that oxygen to promote dehydration. Upon dehydration, the pKa for the amino proton is lowered such that it can be facilely lost to solvent, to produce the final thiazoline moiety. When present, Q1228 (not displayed) could aid in one or more of these steps, or compensate for the lack of T1196 when absent, perhaps by ordering solvent molecules to take the place of its hydroxyl.

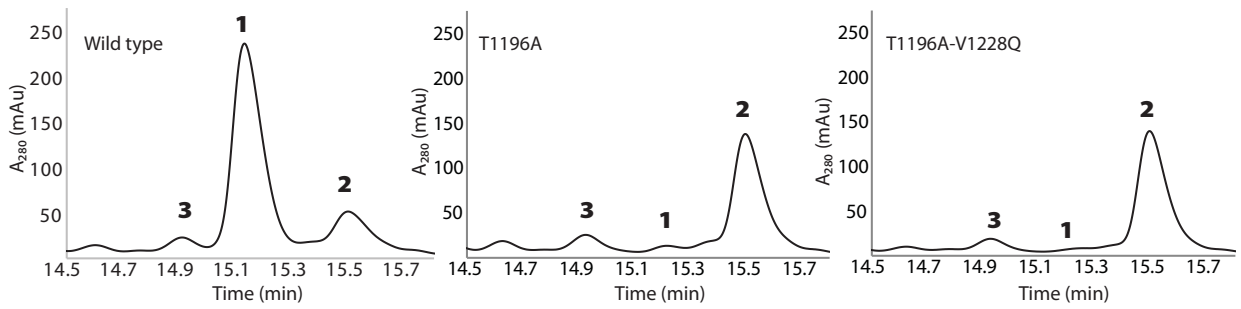


Figure S5.8. HPLC traces for peptide synthesis assays for all BmdB wild-type, T1196A, and T1196A/V1228Q. Single peptide assay reactions done in parallel for wild-type, T1196A and T1196A/V1228Q show the single mutant and double mutant possess the same catalytic ability.

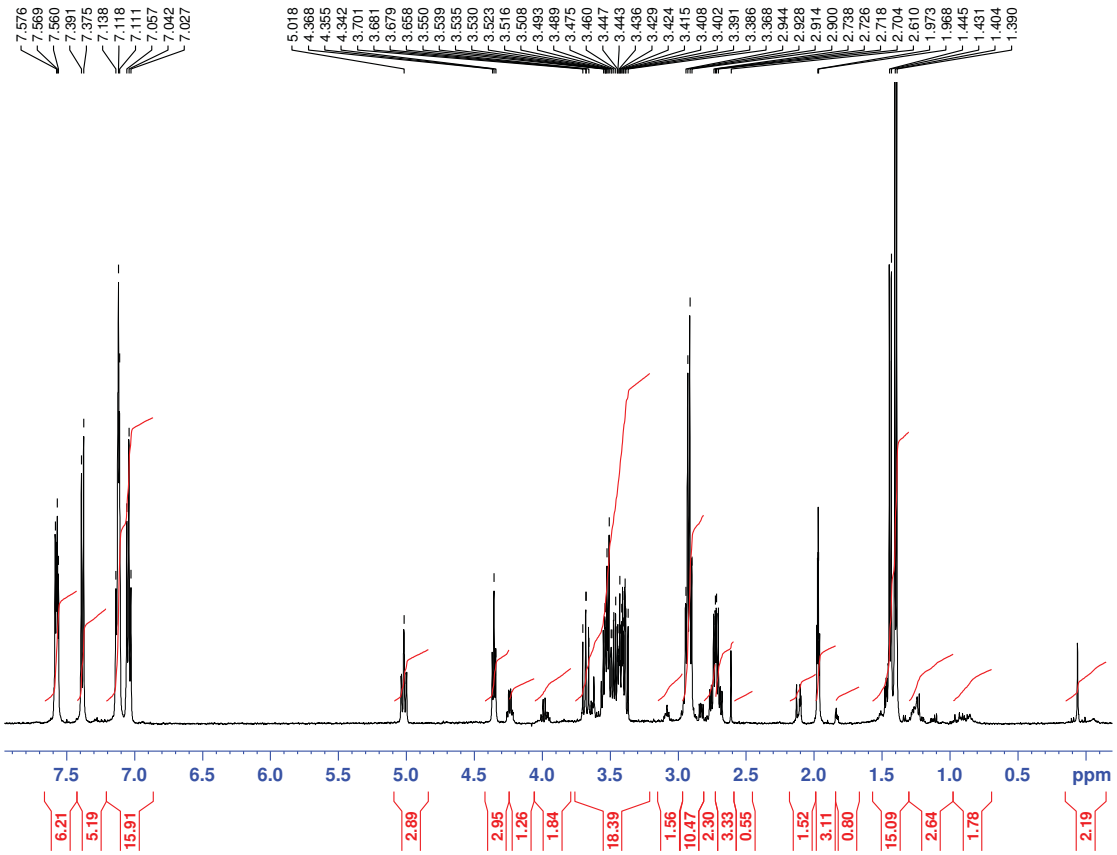


Figure S5.9. ^{1}H NMR spectrum of Bacillamide E. See Supplemental Methods for experimental description and peak assignment.

BmdB residue	Protein	Mutation	Effect	Study	Our structure-aided interpretation	Protein-specific considerations
A858	EpoBcy	S80A	moderate decrease in product formation*	Dowling, 2016	Mutation affects positioning of acceptor PPE minorly	
Y859	BmdB-Cy2	Y859F	moderately reduced cyclodehydration	This study	Mutation affects positioning of acceptor PPE minorly	
Y859	EpoBcy	Y81F	moderate decrease in product formation*	Dowling, 2016		
H900	EpoB-Cy1	H122A	greatly reduced total apparent rate*	Kelly, 2005	Mutation of buried residue disrupts folding	
D980	AngN-Cy1	D133A	product in media reduced moderately*	Di Lorenzo, 2008	Mutation disrupts folding of active site	Cy1, Cy2 largely redundant
D980	AngN-Cy2	D575A	product in media reduced moderately*	Di Lorenzo, 2008		Cy1, Cy2 largely redundant
D980;D980	AngN-Cy1-Cy2	D133A/D575A	product in media eliminated*	Di Lorenzo, 2008		Cy1, Cy2 largely redundant
D980	EpoB-Cy1	D201A	greatly reduction total apparent rate*	Kelly, 2005		
D980	HMWP2-Cy1	D246A	no condensation	Keating, 2000		
D980	VibF-Cy1	D133A	reduced cyclization	Marshall, 2002		Cy2 performs condensation; Cy1 cyclodehydration
D980	VibF-Cy2	D590A	reduced condensation	Marshall, 2002		Cy2 performs condensation; Cy1 cyclodehydration
D980;D980	VibF-Cy1-Cy2	D133A/D590A	reduced condensation and cyclodehydration	Marshall, 2002		Cy2 performs condensation; Cy1 cyclodehydration
D980;A981;G986	HMWP2-Cy1	D246H/L247H/A252G	no condensation	Keating, 2000		Mutations designed to introduce HHxxxD(G) motif
D985	AngN-Cy1	D138A	product in media reduced moderately*	Di Lorenzo, 2008	Mutation disrupts folding of active site	Cy1, Cy2 largely redundant
D985;D985	AngN-Cy1-Cy2	D138A/D580A	product in media eliminated*	Di Lorenzo, 2008		Cy1, Cy2 largely redundant
D985	AngN-Cy2	D580A	product in media reduced moderately*	Di Lorenzo, 2008		Cy1, Cy2 largely redundant
D985	EpoB-Cy1	D206N	less protein produced*	Kelly, 2005		
D985	HMWP2-Cy1	D251A	no condensation	Keating, 2000		
D985	VibF-Cy2	D595A	greatly reduced condensation	Marshall, 2002		Cy2 performs condensation; Cy1 cyclodehydration
S988	EpoB-Cy1	S209A	reduced total apparent rate*	Kelly, 2005	Mutation disrupts folding near donor PCP binding site	
S988	HMWP2-Cy1	S254A	wildtype activity	Keating, 2000		
T1089	BacA-Cy2	T875A	no condensation	Duerfahrt, 2004	Mutation disrupts donor PCP binding	
N1114	BacA-Cy2	N900A	condensation; no cyclodehydration	Duerfahrt, 2004	Mutation affects positioning of acceptor PPE	
N1114	BmdB-Cy2	N1114A	greatly reduced condensation; reduced cyclodehydration	This study		
N1114	EpoBcy	N335A	great decrease in product formation*	Dowling, 2016		
T1116	BmdB-Cy2	T1116A	reduced condensation	This study	Mutation affects positioning of acceptor PPE	
F1118	BmdB-Cy2	F1118G	reduced condensation; reduced cyclodehydration	This study	Mutation affects positioning of donor PPE; prevents isolation of active site for cyclodehydration	
R1120	BacA-Cy2	R906A	no condensation	Duerfahrt, 2004	Mutation disrupts donor PCP binding site	
D1133	BacA-Cy2	D919A	wildtype activity	Duerfahrt, 2004	Residue is second-layer from active site	
D1133	EpoBcy	D354A	moderate decrease in product formation*	Dowling, 2016		
T1135	BmdB-Cy2	T1135A	reduced condensation; reduced cyclodehydration	This study	Mutation affects conformation of adjacent active site residues	
T1196	BmdB-Cy2	T1196A	condensation, greatly reduced cyclization	This study	Mutation affects nucleophilic attack of thiol during cyclodehydration	
S1197	BacA-Cy2	S984A	condensation; no cyclodehydration	Duerfahrt, 2004	Residue is facing away from the active site and towards donor PPE binding site; mutation disrupts cyclodehydration through adjacent T1196	
S1197	BmdB-Cy2	S1197A	wildtype activity	This study		BmdB-Cy2 not as sensitive to this mutation
S1197	EpoB-Cy1	S418A	reduced total apparent rate*	Kelly, 2005		
S1197	EpoB-Cy1	S418N	greatly reduced total apparent rate*	Kelly, 2005		
Q1222	EpoBcy	Q445A	great decrease in product formation*	Dowling, 2016	Mutation affects positioning of the acceptor PPE	
D1226	BmdB-Cy2	D1226G	condensation; no cyclization	This study	Mutation affects nucleophilic attack of thiol during cyclodehydration	
D1226	BmdB-Cy2	D1226N	condensation; no cyclization	This study		
D1226	EpoB-Cy1	D449A	great decrease in product formation*	Dowling, 2016		
	VibF-Cy1	DELTA-CY1	condensation; no cyclodehydration	Marshall, 2002	Deletion eliminates domain performing cyclodehydration	Cy2 performs condensation; Cy1 cyclodehydration

Table S5.1. List of Cy domain mutations. List of Cy domain mutations and their effects as determined from the current study and previous studies(Bloudoff et al., 2016; Dowling et al., 2016; Duerfahrt et al., 2004; Gehring et al., 1998; Keating et al., 2000b; Kelly et al., 2005; Konz et al.; Marshall et al., 2001; Marshall et al., 2002b). The studies marked by an asterisk did not differentiate between defects in condensation and cyclodehydration.

	Native
Diffraction Data	
Wavelength (Å)	0.97949
Space Group	C 1 2 1
Unit-cell parameters (Å)	a = 139.7, b = 124.9, c = 68.9; β = 96
Resolution (Å)	39.9 – 2.3
<I/σ(I)>	15.1 (1.6)
Temperature (K)	100
Measured Reflections (*n*)	161560
Unique Reflections (*n*)	52064 (2557)
Completeness (%)	99.8 (99.5)
Multiplicity	3.1 (2.9)
R_{merge} [a] (%)	6.2 (70.2)
$CC_{1/2}$ (%)	92.3(66.6)
Refinement Statistics	
Resolution range (Å)	39.9-2.3
R-factor/R_{free} (%)[b,c]	20.1/23.7
Waters	217
Wilson B-Factor (Å^2)	49.2
R.M.S.D in bond lengths (Å)	0.004
R.M.S.D in bond angles (°)	0.773
Ramachandran Plot (%)	
Favoured	97.6
Allowed	2.2
Outliers	0.2

[a] $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - < I(hkl) > | \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th measurement and <*I*(*hkl*)> is the weighted average of all measured reflections.

[b] $R - factor = \sum \| F_{obs}(hkl)| - |F_{calc}(hkl) \| / \sum | F_{obs}(hkl)|$

[c] R_{free} is the R-factor computed for the test set of reflections omitted from the refinement process.

Table S5.2. Crystallization and refinement statistics for BmdB-Cy2.

Primer Name	Sequence (5'-3')
VulgarNRPS_F	AAAAAACCATGGAAATGCGAATAAGAGATTTGGATTTG
VulgarNRPS_R	TTTTTTGCGGCCGCTATCTGAAAATCACCGATCGTTT
Vul_Y859F_F	GTGCAGACCGCTTTCATGTTGGGGCGC
Vul_Y859F_R	GCGCCCCAACATGAAAGCGGTCTGCAC
Vul_F1118G_F	GTTGGCGATCAATTTGACTGTTGGTAACCGTTATCCGGTACATGAT
Vul_F1118G_R	ATCATGTACCGGATAACGGTTACCAACAGTCAAATTGATCGCCAAC
Vul_N1114A_F	CGTCGGTTGGCGATCGCTTTGACTGTTTTTAACCGTTATCC
Vul_N1114A_R	GGATAACGGTTAAAAACAGTCAAAGCGATCGCCAACCGACG
Vul_S1197A_F	CATGCCGATTGTGTTCACTGCCATGCTGGCTGGTG
Vul_S1197A_R	CACCAGCCAGCATGGCAGTGAACACAATCGGCATG
Vul_D1226N_F	GACGCCGCAAGTGTATTTGAATAATGTAGTGATTGAAAAAAACGG
Vul_D1226N_R	CCGTTTTTTTCAATCACTACATTATTCAAATACACTTGCGGCGTC
Vul_D1226G_F	GACGCCGCAAGTGTATTTGGGTAATGTAGTGATTGAAAAAAACGG
Vul_D1226G_R	CCGTTTTTTTCAATCACTACATTACCCAAATACACTTGCGGCGTC
Vul_T1116A_F	CGGTTGGCGATCAATTTGGCTGTTTTTAACCGTTATCCGG
Vul_T1116A_R	CCGGATAACGGTTAAAAACAGCCAAATTGATCGCCAACCG
Vul_T1196A_F	GGTCATGCCGATTGTGTTCGCTTCCATGCTGGCTGG
Vul_T1196A_R	CCAGCCAGCATGGAAGCGAACACAATCGGCATGACC
Vul_T1135A_F	CAGATCGTGGGGGATTTCGCTTCGCTCATCCTCTTGG
Vul_T1135A_R	CCAAGAGGATGAGCGAAGCGAAATCCCCCACGATCTG

Table S5.3. List of primers used in this study.

CHAPTER 6: GENERAL CONCLUSIONS

6.1 Structure of the C domain.

6.1.1 Pseudo-evolutionary comparison of the C domain with CAT

Since the beginning of my thesis research, several NRPS constructs that include the C domain have been crystallised and published (Bloudoff et al., 2016; Bloudoff et al., 2013; Bloudoff and Schmeing, 2013; Drake et al., 2016; Miller et al., 2016). In order to gain a higher level of appreciation for the structure of the C domain, it would be of interest to the reader to imagine conceptually how these domains could have arisen.

De Crecy-Lagard had noticed the similarity of the N-lobe sequence to CAT and E2p (De Crecy-Lagard et al., 1995), but only with the VibH structure was it evident that both the N- and C-lobes showed this structural similarity. Keating *et al.* (Keating et al., 2002) pointed out that type III CAT (Leslie, 1990) and E2p (Guest, 1987) both exist as head-to-tail trimers, and the two lobes of the C domains take approximately the positions of two of the three protomers in a CAT trimer. A “pseudo-evolutionary pathway” of the secondary structure elements of CAT to those of the C domain using topology diagrams (Figure 6.1) begins as follows: A single CAT protomer has a central, 6-stranded ß-sheet, one strand of which is donated from the adjacent protomer in the trimer. This sheet is surrounded by 5 α-helices and a smaller, 3-stranded ß-sheet. The HHxxxDG active site motif is in a loop adjacent to the central sheet (Lewendon et al., 1994). A gene duplication event is required to provide two sequences that could later become the N- and C-lobes. The duplication could have led directly to a fused pseudo-dimer, or fusion could have occurred later. The duplicated monomers diverged, each losing different sequences. The “proto-N-lobe” lost 4 ß-strands and 2 α-helices, while the proto-C-lobe lost 3 ß-strands and 1 α-helix, but only the loss of the first, very short ß-strand is common to both monomers. The HHxxxDG motif remains in the proto-N-lobe, while this completely degrades in the proto-C-lobe (this sequence is LASYPQS in VibH).

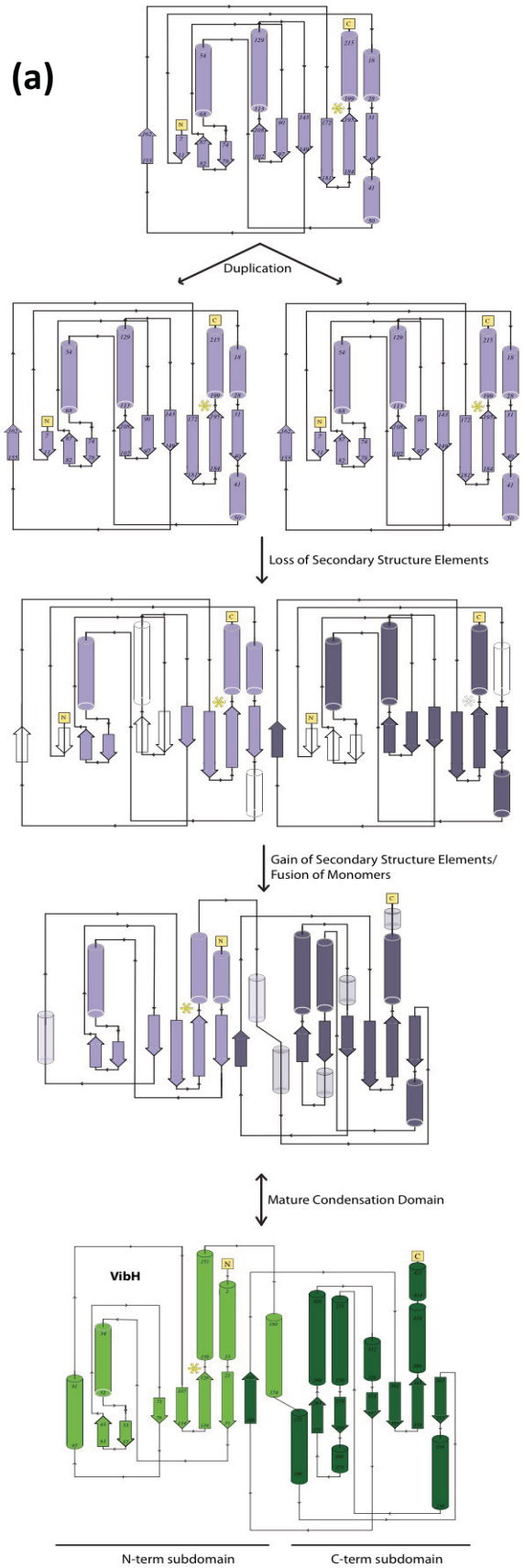
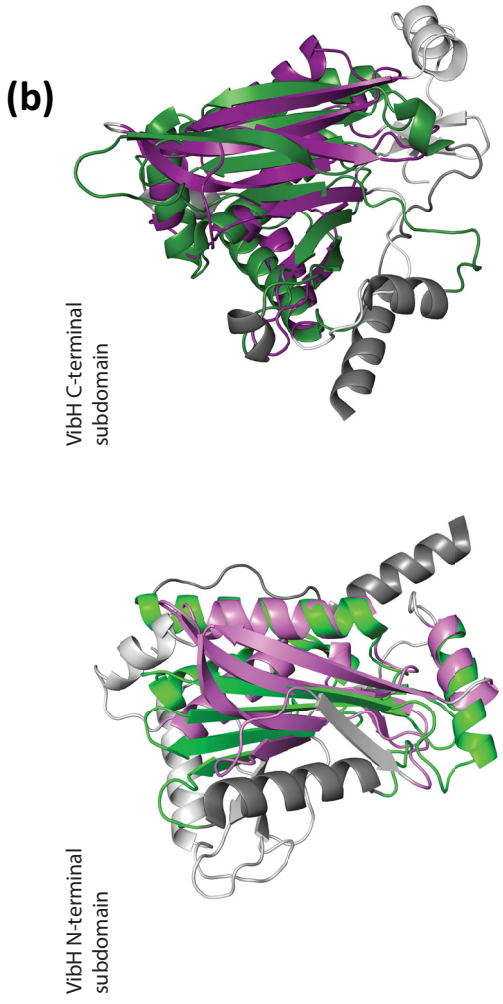


Figure 6.1. "Pseudo-evolution" of the C domain from CAT. (a) CAT topology is shown in purple and C topology diagram is shown in green, with the lighter shades representing the N-lobe and the darker shades representing the C-lobe. The active site His-motif is identified with a yellow star and the degraded His-motif is identified with a white star. White secondary structure elements are those that have been lost over time, while grey/translucent secondary structure elements are those that have been gained over time. (b) Comparison of N- and C-lobes of VibH (PDB: 1L5A) and CAT monomers (PDB: 1CIA). Colouring is same as (a)



The proto-N- and proto-C-lobes are joined by two α-helices (VibH 160-174 and 175-190, *cda 192-206 and 207-220*); alternatively described as one long, kinked α-helix) which shows their own, independent pseudo-symmetry. Three α-helices are also gained in various positions in the C-lobe and one in the N-lobe yield the mature C domain.

This exercise highlights the conceptual history of the latch. In the CAT trimer, the ß-strand (CAT 155-162) is donated to the adjacent monomer of the CAT and is a major contributor to trimerization. In the CAT monomer that becomes the C-lobe, this ß-strand maintains its interaction with the adjacent monomer's central sheet as it evolves into the N-lobe and becomes the latch. Conversely, as there is no third lobe in C domains for it to interact with, this element is lost in the N-lobe sequence and replaced by α-helix which packs against the N-lobe's own central sheet (VibH 81-95, *cda 113-127*). In C domains, the latch (including the ß-strand and residues) forms the roof over the active site. Samel *et al*. (Samel et al., 2007) suggested that the latch may disengage, opening the active site for product to leave. However, that this has not been observed in any C domain or related domain, and that this interaction is conserved through evolution is strong evidence that the latch is always engaged.

The comparison with CAT also informs analysis of the relative positions of the N- and C-lobes. We have noted that relative orientations of the two lobes vary substantially, from relatively more "open" or "closed" conformations of the V shape (Bloudoff et al., 2013). This openness did not correlate with whether the C domain originates from an early or late module along the NRPS assembly line. It does not seem to be a crystallization artifact, and normal mode analyses recapitulated the open to closed transition (Bloudoff et al., 2013). We proposed that this observed conformational variation may play an as of yet unidentified role, perhaps in domain-domain communication. Notably, in the recent termination module structures from AB3403 and EntF, the C domains are in similar, closed conformations (Drake et al., 2016). It is interesting to see, however, that the "closed" conformation of CDA-C1 overlays extremely well two CAT monomers in its native trimer (Figure 6.2), whereas in more "open" C domain structures like VibH, only one lobe at a time can be aligned. It is also possible that the degree of openness is a feature of the degree to which a particular C domain has maintained this conformation through evolution.

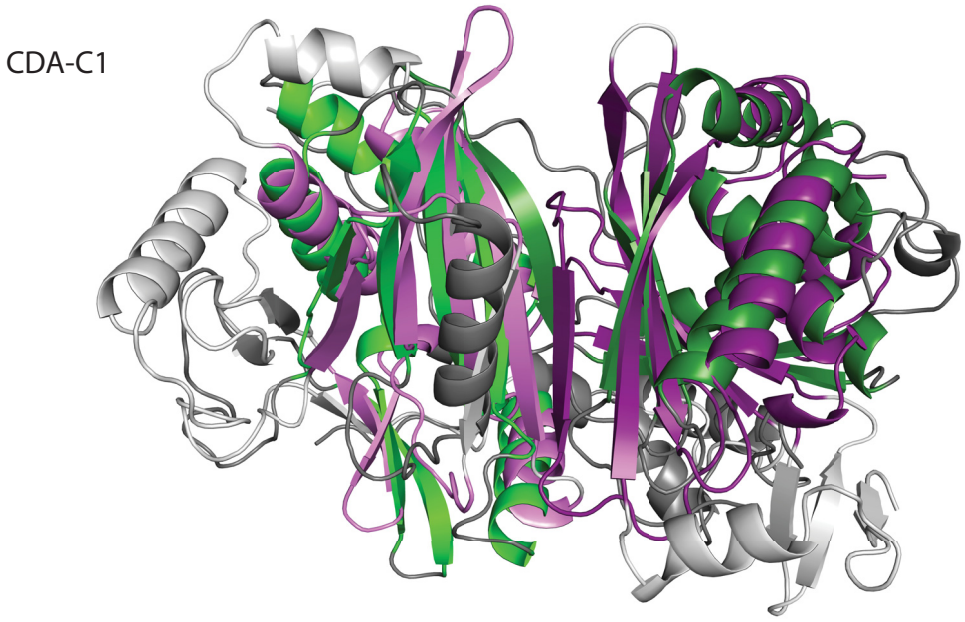


Figure 6.2. CDA-C1 vs. CAT. CDA-C1 (PDB: 4JN3, coloured green) is overlaid by a two CAT monomers in its native trimer (PDB: 1CIA, coloured purple). The lighter shades represent the N-lobe and the darker shades represent the C-lobe. White secondary structure elements are those that have been lost over time, while grey/translucent secondary structure elements are those that have been gained over time

One excellent way to address this question would be with NMR studies into dynamics of the C domain. Fortunately, it seems such studies are may be coming soon (Harden et al., 2015; Mishra and Frueh, 2015).

6.1.2 An improved understanding of PCP binding sites on the C domain and the NRPS

At the time of writing this thesis, the crystal structure of a donor PCP domain and C domain in a productive interaction still does not exist. However, proxies of such an interaction have now been described. Structures of a PCP-E didomain (Chen et al., 2016) and a PCP-C_T didomain (with and without the PPE present) (Zhang et al., 2016), show donor PCP binding to the C superfamily domain at the proposed site (described in Section 6.3). The acceptor PCP binding to the C domain is much more well established. In addition to this productive interaction which had previously been published (Tanovic et al., 2008), a more recent C-A-PCP-Te termination module structure of AB3403

included a covalently-linked PPE arm at the acceptor site (Drake et al., 2016). The PCP domain of AB3403 is rotated ~30° relative to the orientation of the PCP domain of SrfAC (Tanovic et al., 2008), but the same general interactions are observed. The acceptor PPE in AB3403 threads into the tunnel and makes contacts with a C domain α-helix (E20-L30, *cda A14-L24*), and Y37 (*cda Y31*).

In addition to interacting with the C domain, it has been interesting to learn about and visualize the vast distances the PCP domain must travel in order to fulfil its functions. The different active sites the growing peptide is required to interact with are tens of angstroms apart (e.g., ~ 50 Å in the initiation module of linear gramicidin synthetase) (Drake et al., 2016; Lohman et al., 2014; Reimer et al., 2016; Tanovic et al., 2008). The flexible linker sequences adjacent to the PCP domains are typically too short to enable these conformational changes alone, so repositioning of adjacent domains is also required for PCP to deliver their covalently bound substrates.

6.1.3 Hints of a more dynamic C domain – A domain interaction

New crystal structures of NRPS modules from the Gulick group (Drake et al., 2016; Miller et al., 2016) have provided new points of discussion regarding the "catalytic platform" provided by the C – A_{core} interface (Tanovic et al., 2008). The C domain – A domain interface in AB3404 is approximately the same size, but the EntF C domain – A domain interface is, somewhat surprisingly, substantially smaller. Additionally, the three structures displayed somewhat different relative C domain – A domain orientations (Drake et al., 2016; Tanovic et al., 2008), and a subsequent structure of EntF (Miller et al., 2016) showed another similar, yet distinct orientation. This suggests that that the C domain – A domain platform may be more dynamic than first thought.

A very recent study from our group includes the structure of a cross-module construct of DhbF, with A and PCP domains from module 1 and the C domain from module 2 (Tarry et al., In revision). Unlike C and A domains from the same module, these A1 and C2 domains from adjacent modules make absolutely no contact with each other. Furthermore, accompanying low resolution EM of the dimodular DhbF shows that the two modules can assume many different relative orientations, and that there is no

consistent A1 – C2 interaction. These data hint at a lack of rigid supermodular architecture in NRPSs mediated by C domains.

6.2 A Novel Reaction Mechanism of the C Domain

An important step towards determining an enzyme's reaction mechanism is visualizing it in complex with substrates or substrate analogues. However, for C domains, this did not appear to be a trivial problem. Even at extremely high substrate concentrations (>50 mM), no significant density was observed at the active site of a C domain (Bloudoff et al., 2013) for small molecule SNACs or aminoacyl-PPEs despite their utility in biochemical experiments with a handful of C domains (Ehmann et al., 2000; Haynes et al., 2011; Luo et al., 2002; Roche and Walsh, 2003). This is likely because in the crystallography experiments, the contribution imparted by covalent substrate delivery by the PCP domain could not be overcome by extremely high concentration. Furthermore, it is notable that though several didomain and module structures contain the PPE arm (Chen et al., 2016; Drake et al., 2016), none also include a substrate moiety attached to this PPE arm.

To attempt to resolve the lack of C domain - substrate complexes and investigate the catalytic mechanism, we recently developed a chemical biology approach using small molecule analogues which maintain the feature of the covalent delivery to the active site. Instead of being delivered by a PCP domain, the small molecule analogs are covalently tethered to a residue along the PPE tunnel for presentation at high local concentration to the C domain active site. We introduced a cysteine at position 17 of CDA-C1, which is part of the tunnel between the acceptor PCP domain binding site and H157. We then designed acceptor substrate analogs consisting of alanine as the acceptor amino acid, linked by alkyl chains of varying lengths to a bromine. Reaction of the introduced cysteine with the analogue displaces the bromine, producing a stable thioester-linked protein-substrate complex, able to place the alanine moiety in the acceptor position in the active site(Bloudoff et al., 2016). Critically, if these complexes are faithful mimics of the aminoacyl-PPE-PCP acceptor, they would be competent substrates in a reaction with the native donor, acyl-PPE-PCP. Electrospray and in-source decay matrix-assisted laser desorption/ionization mass spectrometry showed

that the substrate analogue could indeed be covalently tethered to C17, and then successfully act as a substrate in condensation. Notably, the length of the alkyl chain between the alanine and bromine moieties had substantial effects on the efficiency of C17 alkylation, hinting that interactions with the active site helped target the substrate. The length of the alkyl chain also affected the efficiency of condensation, as would be expected. Mutation of H157 abolished condensation of covalently tethered substrate, which is the same result this mutation has in the native synthetase system (Bloudoff et al., 2013), indicating again that though minimalistic, this approach is a faithful mimic from which insight into condensation can be drawn.

The major advantage of this chemical biology approach is that it produces reaction-competent substrate – C-domain complexes suitable for crystallization and visualization of the substrate. Several such complexes were crystallized and the structures determined. The complex with the most active substrate was resolved to 1.6 Å, which allows unambiguous positioning of all non-hydrogen atoms. In all complexes, the α-amino group of the acceptor amino acid (the nucleophile in the condensation reaction) makes hydrogen bonds with the ε nitrogen of H157 and with the backbone carboxyl of S386. No additional residue side chains are in position to act directly in catalysis, which is consistent with conservation and mutagenesis data. Furthermore, the ε nitrogen of H157 appears to accept a hydrogen bond from the substrate's α-amino group, as the δ nitrogen of H157 is protonated and interacting with nearby backbone carbonyls, suggesting H157 is not positively charged and thus not able to participate in transition state stabilization.

In all, the mutagenic and structural data draw together towards the conclusion that H157 cannot act as an obligate strong general base in the condensation reaction throughout NRPSs. This is further supported by the action of the Cy domain (see below), which catalyzes condensation effectively, but contains a DxxxxD motif in place of the HHxxxDG motif, where the position of H157 is usually taken by a hydrophobic residue.

The main catalytic power for C domains is thus likely substrate positioning, a fundamental mechanism used in all enzymes (Jencks, 1969; Keating et al., 2002). H157 plays an important role in this positioning in CDA-C1 and other C domains, whereas

some C domains (and Cy domains) do not absolutely require this interaction to achieve productive substrate positioning, perhaps due to additional contacts with active site residues. Deprotonation of the α-amino group from $–NH_3$ to $–NH_2$ may not be required to initiate nucleophilic attack, as its pKa has been estimated at ~7 (Anderson and Packer, 1974; Samel et al., 2007), so the neutral, reaction-competent $–NH_2$ form is likely to be present at neutral pH. Upon attack, the reaction proceeds through a zwitterionic transition state to the amide product (Yang and Drueckhammer, 2000). The proton from $–NH_2$ is likely lost facilely to solvent when its pKa drops markedly as the reaction proceeds (Green and Lorsch, 2002). When present, the H157 could accept this proton and thus formally act as a weak general base, and make a minor (Keating et al., 2002; Marshall et al., 2002b; Roche and Walsh, 2003) contribution to catalysis in this way, in addition to its positioning role. This mechanism of catalysis, principally through orientation, echoes that of peptide bond formation by the ribosome (Beringer et al., 2005; Kuhlenkoetter et al., 2011; Wallin and Aqvist, 2010).

6.3 New Insights into the Specificity of C Domains

Our structure of CDA-C1 with a novel tethered acceptor substrate also allowed us to identify, and confirm via mutagenesis, that a serine in the active site, S309, contributes to acceptor specificity in CDA-C1 (Bloudoff et al., 2016). A S309G mutation increased the selectivity of CDA-C1 towards acceptor substrates: Wild type CDA-C1 used serinyl-, methioninyl-, alaninyl- or leucinyl-PCP, whereas CDA-C1 S309G only used serinyl and methioninyl-PCP. The result can be explained by steric arguments, but was impossible to predict without the structure of this particular C domain, and S309 is not conserved throughout C domains which use serinyl-PCP as the acceptor substrate. Moreover, it exemplifies why computational analyses of C domains have not been able to predict native substrates or specificity restrictions of a C domain from the active site sequence (Rausch et al., 2007). Development of a rough specificity code along the lines of the well-defined code in A domains (Rausch et al., 2005) may not be achievable in C domains.

In addition to its own specificity, the C domain can also influence the specificity of the adjacent A domain. Meyer *et al.* (Meyer et al., 2016) assayed the adenylation

specificity of the A domain in constructs that included or lacked the C domains: in the three C-A-PCP modules tested, MycB adenylates leucine preferentially, MycC adenylates only arginine and $MycB_C$ adenylates arginine, leucine, tryptophan and tyrosine. Strikingly, when the C domains are not included in the constructs, all three A-PCP didomains can adenylate five or more amino acids, but adenylation of cognate arginine by MycC is decreased. Including just the C-lobe of the C domain in the constructs ($C_{C\text{-}lobe}$-A-PCP) largely returns the specificity of MycB, and enables MycC to adenylate arginine efficiently, but does not return native specificity to MycC or $MycB_C$. In a similar result, Li *et al.* (Li et al., 2016) demonstrated that in SulM module 2, the excised A-PCP construct adenylates only non-cognate L-2,3-diaminopropionate, whereas C-A-PCP adenylates the cognate L-alanine. These results serve as a warning that experiments from excised domains may carry caveats, but the *in vivo* relevance is unclear. In all these NRPSs, the C domain is an obligate, covalent partner of the A domain, and thus always bound to it. Theoretically, an NRPS system in which a module is split between subsequent proteins (NrpsA = ……-C; NrpsB = A-PCP-……) could use this feature to produce two distinct products, depending on whether the two proteins are bound together or not, but we are not aware of any such system.

The influence of the C domain on neighboring domain activities is long known. Linne and Marahiel demonstrated that the addition of a C domain in TycB2 module constructs decreases (but did not eliminate) the capacity of that module to donate phenylalanine to the downstream module (Linne and Marahiel, 2000). It was suggested that interactions of the aminoacyl-PCP domain with that C domain cause the reduction, and that this effect of the C domain is the reason aberrant initiation does not start in each elongation module(Ackerley and Lamont, 2004; Belshaw et al., 1999; Linne and Marahiel, 2000; Stachelhaus et al., 1998).

This question of why an NRPS of *x* modules initiates nearly exclusively at the first module, producing one length of peptide, and not at every module, producing peptides of *x* lengths, is non-trivial (Ackerley and Lamont, 2004; Belshaw et al., 1999; Linne and Marahiel, 2000; Stachelhaus et al., 1998). For an NRPS to synthesize a single product, the aminoacyl-PCP_n in each elongation module *n* must first act as the acceptor substrate in the C_n domain (in that module) before being allowed act as a donor

substrate in the C_{n+1} domain (in the next module). As PCP domain - C domain binding interactions are necessarily transient, it is unlikely that the aminoacyl-PCP_n is fully engaged in the C_n domain donor site while waiting for the upstream peptidyl-PCP_{-1} (Linne and Marahiel, 2000). As discussed above, the C_{n+1} donor site is typically not exquisitely selective. It is possible that the identity of the moiety attached to the PCP domain may influence the conformation of that domain and its affinity to its partner C domains (Goodrich et al., 2015; Jaremko et al., 2015), biasing the PCP domain towards the C_n or C_{n+1} to some extent, but there is no NTP-driven power stroke pushing the PCP domain downstream after condensation in C_n (Schmeing and Ramakrishnan, 2009). It is thus probable that the PCP domain will visit all its partner sites, independent of what it bears on the PPE arm. It seems likely that a combination of relative affinities for the aminoacyl-/peptidyl-PCPs at the C_n acceptor and C_{n+1} donor site is sufficient to prevent substantial mis-initiation.

6.4 An Updated View of Non-Canonical C Superfamily Domain Members

Understanding the subtle differences in how structurally-similar C superfamily domains are able to facilitate vastly different chemical reactions will ultimately provide a greater understanding in how these enzymes specifically work. In addition, the discovery of new C superfamily functionalities will help aid in future bioengineering efforts by providing simple tools to increase the chemical diversity of peptide products. While crystal structures are not known for all C superfamily domains described below, representative and recently published structures can be seen in Figure 6.3.

6.4.1 Cy Domain

Although the NRPS field had begun to elucidate hints of how the Cy domain was able to catalyze two independent reactions (Duerfahrt et al., 2004), definitive understand into how this was able to occur could not be achieved until structures (and accompanying mutagenesis data) of the Cy domain were obtained. This goal was recently reached, when we (Bloudoff et al., 2017) and Dowling *et al.* (Dowling et al.,

2016) determined the first structures of Cy domains, from bacillamide and epothilone synthetases. As expected, the overall structure of the Cy domain is quite similar to C domains. The donor and acceptor binding sites appear competent for PCP binding.

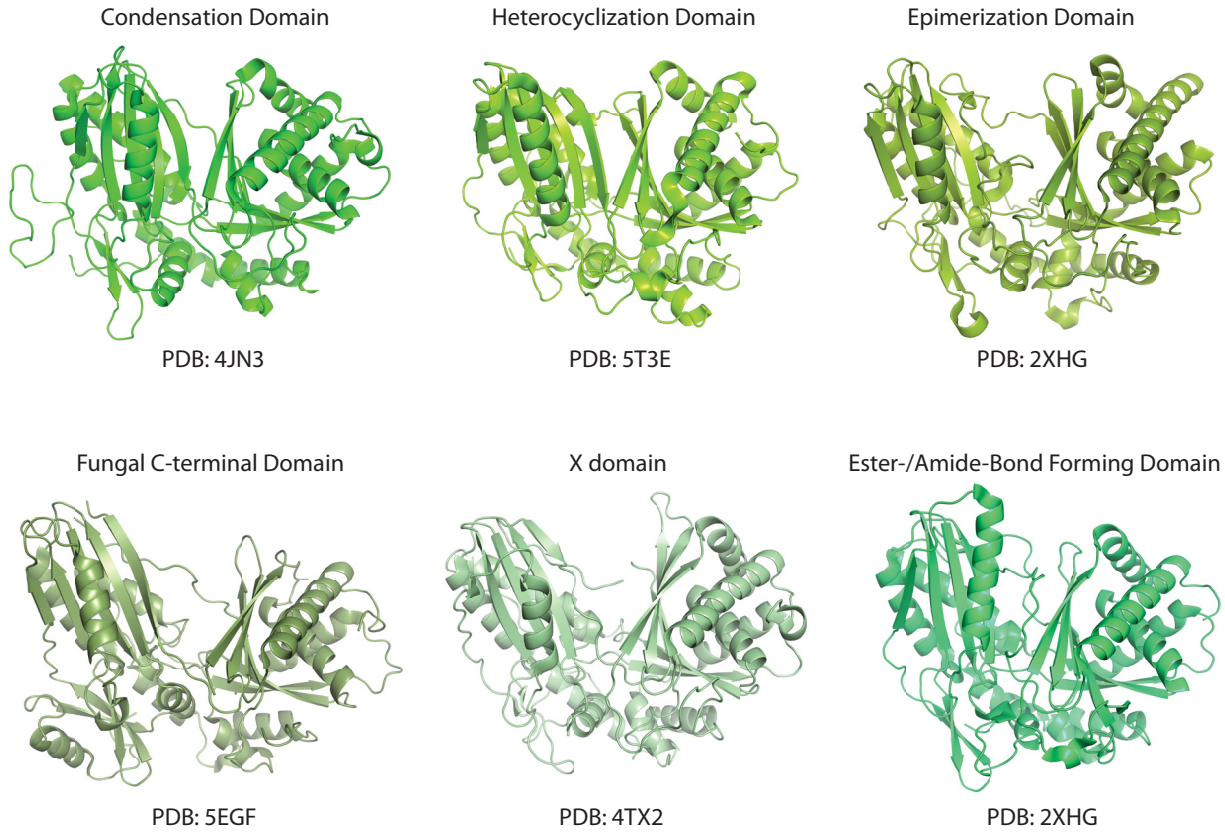


Figure 6.3. Recently solved crystal structures of C domain superfamily members.

Furthermore, the loop containing the DxxxxD motif is in near identical conformation to the C domain HHxxxDG, including D980 (*cda H156*) and D985 (*cda D161*) oriented with their side chains tucked into the protein and away from the active site cavity. As there is no indication this loop will change orientation, these residues are likely important for structural purposes rather than catalysis. The position of the second histidine (*cda H157*) in C domains is occupied by A981 in BmdB Cy2 and L202 in EpoB Cy, consistent with the overwhelming Cy domain preference for hydrophobic residues there. For insight into how the Cy domain can perform condensation and cyclodehydration if not

through the DxxxxD motif, we undertook a new bioinformatic analysis of 1790 Cy domains. This highlighted two conserved motifs with consensus sequences PVVFTS (BmdB 1192-1197) and SQTPQVxLD (BmdB 1218-1226) (Konz et al., 1997). The PVVFTS and SQTPQVxLD motifs are adjacent to one another and form a surface on the acceptor substrate side of the active site.

We and Dowling *et al.* each performed structure-guided mutational analyses, which gave consistent results. Both studies saw a large decrease in catalytic activity upon mutation of N1114 (Epo N335) to alanine. As our assay monitored formation of both linear and heterocycle-containing peptide, it was discernable that this mutation diminished both condensation and cyclodehydration. Both studies also identified the nearly completely conserved D1226 (EpoB D449) of the SQTPQVxLD motif as critical to Cy domain function, and we could pinpoint that it is specifically required for heterocyclization and not condensation. We also mutated T1196 (EpoB T417) of the PVVFTS motif to and saw robust condensation but near complete reduction of heterocyclization. D1226 and T1196 are 2.9Å apart (2.7 Å in EpoB Cy) and directly interact with one another.

The structure of BmdB-Cy2, the bioinformatics and the mutational data led us to propose putative reaction mechanisms: The condensation reaction in Cy domains is, like in C domains, largely promoted by substrate positioning, and N1114 may play a facilitating, though not essential role in this positioning. D1226 very likely acts as a catalytic base/acid in the heterocyclization reaction. It would first abstract a proton from the thiol / hydroxyl side chain. T1196 (or in the minority of cases where T1196 is absent, Q1228) could then donate a proton to the former carbonyl oxygen to form the hydroxyl-thiazolidine / oxazolidine intermediate. D1226 could protonate the same oxygen again to allow it to leave as water in the dehydration step, giving the thiazoline / oxazoline product. The Cy domain studies highlight the impressive versatility of C superfamily domains.

6.4.2 E domain

As a feature of one of the earliest known examples of a purified NRPS in Yamada & Kurahashi's 1968 publication (Yamada and Kurahashi, 1968), the crystal structure of

the E domain went unknown for 46 years, at which point the E domain from TycA module 1 was described (Samel et al., 2014). The E domain was found to have the expected overall structural similarity to the C domain (Keating et al., 2002). However, there are two substantial structural differences, located in the two crossover regions where sequence from the C-lobe crosses over to the N-lobe. In the E domain, the small "floor loop" crossover (TycA-E1 885-903; *cda 295–309*), is enlarged by ~5 residues relative to C domains. Because it is positioned near the PCP binding site, the floor loop was proposed to influence PCP binding. Indeed, the subsequent structure of the GrsA-PCP1-E1 didomain showed the PCP domain and interdomain linker interacting with the floor loop during productive binding (Chen et al., 2016). This didomain is the first C superfamily domain to be visualized with a PCP domain productively interacting at the donor site. The structure includes the PPE, which makes several hydrogen bonds with the backbones of residues lining the tunnel, as well as with the side chain of N975 (*cda N364*).

The latch is enlarged and modified in the structures of the E domains to an even greater extent than the floor loop (Chen et al., 2016; Samel et al., 2014). GrsA residues 980-989 (*~cda 368-380*) cross from the C-lobe to the central ß-sheet of the N-lobe in a manner not unusual for C domains. However, from there, the E domain latch does not donate a ß-strand to the N-lobe ß-sheet but instead adopts a long, irregular conformation (GrsA 990-1016; *~cda 381-391*). This pushes the N-terminal helix of the N-lobe (GrsA 627-636; *~cda 13-24*) into what would be the acceptor PCP binding site, blocking it. Chen *et al*. observed that the C-terminus of the latch is 9 residues shorter in some E domains found in elongation modules, perhaps to allow longer peptidyl-moieties to bind to the active site (Chen et al., 2016).

The E domain active site motif is in a near identical conformation to that of C domains. However, a residue conserved in E domains, GrsA E892 (*cda P282*) is positioned across the active site from GrsA H753 (*cda H157*) of the active site motif in both structures, making it a candidate for direct involvement in catalysis of epimerization. Samel *et al.* and Chen *et al.* both propose reaction mechanisms in which the C_{α} proton is abstracted by the residue acting as the general base, producing an enolate intermediate that is resolved by donation of a proton by the residue acting as

general acid. However, the proposals disagree about which, the histidine or the glutamate, is the general acid and which is the general base. Considering that, while mutation of either residue is extremely detrimental for epimerization, mutation of the histidine eliminates E domain-mediated C_α proton extraction and mutation of the glutamate only reduces it ~6 fold (Stachelhaus and Walsh, 2000), it is more likely that the histidine is acting as the general base (Chen et al., 2016).

6.4.3 X Domain

The speculation from Stegmann *et al.* (Stegmann et al., 2010), which hypothesized that the X domain could recruit the P450 monooxygenase enzymes to the NRPS, was recently proven in a comprehensive and definitive study by Haslinger et al. (Haslinger et al., 2015). The crystal structure of the X domain showed it to adopt the same overall fold as the other C superfamily domain proteins, but strikingly lack features the other proteins require for function. The position of the second histidine in the active site motif was occupied by an arginine (R141, *cda H157*) which protrudes into the active site and blocks the usual binding site of the donor PPE arm. Several other long, hydrophilic residues also fill space near the donor site, and, despite similar positions of secondary structure elements in C and X domains, loops in the X domain block the acceptor PCP binding site. Together this means the X domain does not have a substrate binding tunnel. Haslinger *et al.* then used analytical size-exclusion chromatography and native PAGE to show binding between the X domain and the P450 oxygenase which introduces the first GPA crosslink (OxyB) and determined the co-complex structure. OxyB binds the N-lobe of the X domain mostly through hydrogen bonds and salt bridges. A key binding interaction was observed between the important PRDD motif in OxyB and the X domain loop region containing R167 and R171 (*cda R183*, *A187*) which has been shown to be essential for crosslinking activity (Haslinger et al., 2015; Peschke et al., 2016b)). The PRDD motif thus acts as a recognition determinant for binding to the X-domains, as this motif is found only in P450 monooxygenase associated with GPA synthesis. Interestingly, the strength of interaction between the P450 monooxygenase and the X-domain can be reciprocally related to the efficiency of GPA biosynthesis (Ulrich et al., 2016). Ulrich *et al.* showed

that higher affinity interactions between P450 monooxygenase StaH and X-domain led to lower GPA biosynthesis, likely due to decreased substrate turnover.

Together, these X domain studies revealed and characterized a novel function for C superfamily domains, as a dynamic recruitment platform for a trans-acting modifier of peptidyl-PCP (reviewed in (Peschke et al., 2016a)).

6.4.4 Terminal C Domain of Fungal NRPSs

Although fungal C_T domains had been known to facilitate relatively simple peptide release via macrocyclization, as in the cyclosporin synthetase, it was more recently found that products with more complicated ring systems, such as the tryptoquialanine and fumiquinazoline peptidyl alkaloids, are also produced by NRPSs with C_T domains. In these systems, C_T domains are highly stereospecific (Haynes et al., 2011), and catalyze an attack with internal amino groups that lead to larger, 7-10 membered rings, which then rearrange to multiple 5-6 membered rings, likely spontaneously (Gao et al., 2012; Zhang et al., 2016). In the catalytic motif of five C_T domains, the first histidine (*cda H156*) is replaced by a serine (3 times) or asparagine (once), whereas the first (TqaA H3766; *cda 157*) is present in each of the five and required for condensation. (Gao et al., 2012). Only peptide substrates tethered to the natural PCP domain partner of the C_T domain are efficient substrates, and not SNAC- or CoA-substrate analogues, which underlines the importance of PCP domain delivery.

Recently, the first crystal structures containing a C_T, domain were determined. Zang *et al*. visualized the C_T domain of TqaA on its own and as part of a didomain PCP-C_T construct (Zhang et al., 2016). Overall, the fold of the C_T domain was like the folds of previously determined C domains. The most significant difference is that the N-terminal loop in C domains is replaced by an α-helix in this C_T domain, which pushes the truncated adjacent α-helix in, blocking what would have been the acceptor site of the C_T domain. This lack of acceptor site is expected, as the C_T domain has only one PCP-linked substrate. The didomain PCP-C_T structure shows the PCP at the donor site, its PPE arm in the tunnel towards H3766 (*cda H157*). The residues positioned around this donor side of the active site are mostly aromatic, making it a suitable environment to accommodate the aromatic substrate for macrocyclization. Zang *et al*. suggest, based

on the positon of the PPE thiol, that H3766 (*cda H157*) and backbone amide of G3771 (cda of the HHxxxDG motif) may aid in catalysis by polarizing the thioester oxygen of the reaction intermediate.

6.4.5 Ester-bond Forming C Domains

Contrary to the previously described Fum14p PCP-C didomain (Zaleta-Rivera et al., 2006), additional ester-forming C domains have been indentified with more traditional His-motifs (Marxen et al., 2015), suggesting that perhaps the irregular His-motif of Fum14p is not the source of its unique catalytic capability. One such ester-bond forming C domain is SgcC5, the free-standing C domain involved in C-1027 biosynthesis (Lin et al., 2009) and catalyzes condensation between a donor 3-chloro-5-hydroxy-ß-tyrosinyl-PCP and a hydroxyl from the enediyne core of C-1027. Interestingly, this C domain was also able to catalyze amide formation, albeit with significantly reduced efficiency. We await description of the structure of SgcC5 in complex with small molecule analogues. In examples more analogous to canonical NRPS systems, ester bond formation between peptidyl-PCP donor and a hydroxacyl-PCP acceptor substrates has been demonstrated *in vitro* for the cryptophycin (Ding et al., 2011), valinomycin (Jaitzig et al., 2014) and cereulide (Alonzo et al., 2015) synthetases.

6.4.6 β-lactam Ring Forming C Domains

A C domain with novel function has recently been described in nocardicin synthetase (Gaudelli et al., 2015; Gunsior et al., 2004). The NocB-C5 domain has donor substrate L- hydroxyphenylglycine–L-arginine–L-hydroxyphenylglycine–L-serinyl–PCP and acceptor substrate L-hydroxyphenylglycinyl–PCP. NocB-C5 not only condenses the substrates, but also introduces a β-lactam ring formed between the β-carbon of the serine side chain and the amino group of the L-hydroxyphenylglycinyl–PCP. NocB-C5 contains a typical active site motif, and its histidine, H792 (*cda H157*) was shown to be required for catalysis. This C domain displays features of a $^{D}C_{L}$, with no prominent insertions or deletions obviously accounting for this distinctive catalytic activity. However, the authors observed that there is a third histidine residue, H790 (*cda V155*)

directly upstream of the active site motif. When this residue was mutated to alanine, all C domain catalytic activity ceased. Remarkably, β–lactam formation is returned when NocB-C5 with H790A (*cda 155*) was presented with a tetrapeptidyl-PCP with dehydroalanine in place of serine as the donor substrate and native acceptor substrate. This led to the proposal that H790 (*cda 155*) acts as a general base to promote removal of the hydroxyl of the serine, forming dehydroalanine. H790 (*cda 155*) would then act as a general acid to enable the α-amino of the acceptor L-hydroxyphenylglycinyl–PCP to attack not the thioester, but this dehydroalanine side chain. Finally, the former α-amino group, now a secondary amine, would act as the nucleophile in intramolecular condensation, attacking the thioester carbonyl carbon, to form the β–lactam ring and release the donor PCP. This is an exciting and unconventional mechanism and role for the C domain, as other β–lactam antibiotic pathways, including those which feature NRPS components, generally use non-NRPS enzymes to introduce the β–lactam rings (Townsend, 2016). Furthermore, a structure of NocB-C5 would be especially exciting because in other C domains, the position of H790 (*cda V155*) is somewhat recessed from the active site, so it is not immediately evident how it would be positioned to act as a general base/acid. Even a structure without substrates bound might give insight into the basis of the unique specialization of NocB-C5.

6.5 Outlook

Since their discovery by de Crécy-Lagard (De Crecy-Lagard et al., 1995) in 1995, tremendous progress has been in understanding C domains. Excellent biochemical studies have been performed and a good number of crystal structures of C domains have been solved, as isolated domains, didomains, or in the context of a module. Together this has provided insight into the structure and function of the domain, including questions of catalytic mechanism, specificity, potential conformational flexibility and interaction with partner domains and proteins. The discovery of specialized C domains and other C superfamily domain families has also been exciting. Although they share the same overall fold, these alternative domains can perform significantly different reactions or functions. The above discussion is not exhaustive for

all family members and there is no doubt there are new members with new roles and new chemistry, waiting to be discovered.

Several outstanding questions about C domains and C superfamily domain proteins do remain. Despite substantial effort, full visualization of the reaction cycle has not yet been achieved. A series of co-complexes of the C domain with donor and acceptor analogs together, with transition state mimics and with products, would undoubtedly reveal new information (Schmeing et al., 2005a; Schmeing et al., 2005b). Analogous complexes with other C superfamily domains with relevant substrate analogues (for example, the Cy domain with a mimic of the hydroxyl-thiazolidine cyclodehydration intermediate) would also certainly be informative and reveal residues whose roles should be further interrogated.

In the context of excised C domains, approaches like our tethering of substrates near the active site (Bloudoff et al., 2016) seem most promising. Even such approaches are not guarantees, as despite producing a homogenous sample of CDA-C1 with tethered with *in situ*-made product, the low affinity of the product for the active site meant no corresponding electron density was visible. This problem is unlikely to occur with a faithful analog of the tetrahedral transition state, as transition states have higher affinity to their enzymes than substrates or products (Fersht, 1974). In addition, co-complex structures of C domain that are cognate for small molecules rather than aminoacyl-/peptidyl tRNAs (Lin et al., 2012) will be interesting, if not quite as generalizable to the canonical case.

In the context of multidomain and multimodular NRPS constructs, visualization of C domains with full aminoacyl- and peptidyl-PCPs at acceptor and donor sites is desired. Again, work from our lab and others are actively targeting these complexes. As mentioned, it is notable that, though by now there are a handful of structures of C superfamily domains with PPE-PCP engaged at acceptor and donor sites, none include an aminoacyl or peptidyl group (Chen et al., 2016). This could indicate that the C domain does not possesses the ability to protect thioester from hydrolysis (as, for example, seen in the ribosome with peptidyl tRNAs (Schmeing et al., 2005b)), at least during the time scale required for crystallization. Because the sizes of the NRPS constructs that could include both acceptor and donor PCPs interacting with the C

domain are necessarily fairly large, it will be interesting to see whether these complexes will be visualized by X-ray crystallography, new cryo-electron microscopy approaches (Bai et al., 2015), or both.

The specificity determinants in C domains are still unresolved. The residue in CDA-C1 we saw to influence specificity is not conserved and thus likely not a general control mechanism throughout C domains. Many additional structures of many different C domains with substrates bound could start to provide broader insight into specificity-determining mechanisms. Alternatively, or complementarily, saturation mutagenesis approaches (Evans et al., 2011) targeting the active site of the C domain, though labour intensive and requiring a high-throughput assay, could provide invaluable insight. Increased understanding of specify determination should facilitate bioengineering efforts to produce novel nonribosomal peptides. Bioengineering of NRPSs is conceptually straightforward as domain- or module-swapped NRPS should generate predictable new products. However, reengineering of NRPS usually results decrease in peptide yield (Calcott and Ackerley, 2015; Calcott et al., 2014; Kries et al., 2015), and C domain specificity is one likely cause. One could envision using specificity information gained in the above-proposed experiments to construct a C domain with a fully promiscuous active site for use in bioengineering efforts. However, as C domain specificity is implicated in gatekeeping and ensuring initiation of peptide synthesis only at the first module and not each elongation module, these are yet more processes that must be better understood in C superfamily domain proteins.

REFERENCES

Ackerley, D.F., and Lamont, I.L. (2004). Characterization and genetic manipulation of peptide synthetases in Pseudomonas aeruginosa PAO1 in order to generate novel pyoverdines. Chem Biol *11*, 971-980.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Agata, N., Ohta, M., Mori, M., and Isobe, M. (1995). A Novel Dodecadepsipeptide, Cereulide, Is an Emetic Toxin of Bacillus-Cereus. Fems Microbiology Letters *129*, 17-19.

Alonzo, D.A., Magarvey, N.A., and Schmeing, T.M. (2015). Characterization of cereulide synthetase, a toxin-producing macromolecular machine. PLoS One *10*, e0128569.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Amit, A.G., Mariuzza, R.A., Phillips, S.E., and Poljak, R.J. (1986). Three-dimensional structure of an antigen-antibody complex at 2.8 A resolution. Science *233*, 747-753.

Anandakrishnan, R., Aguilar, B., and Onufriev, A.V. (2012). H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res *40*, W537-541.

Anderson, R.F., and Packer, J.E. (1974). Radiolysis of Aqueous-Solutions of Homocysteinethiolactone Hydrochloride. Int J Radiat Phys Ch *6*, 33-46.

Armstrong, M.D., and Lewis, J.D. (1951). Thioether Derivatives of Cysteine and Homocysteine. J Org Chem *16*, 749-753.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics *22*, 195-201.

Bai, X.-c., McMullan, G., and Scheres, S.H.W. (2015). How cryo-EM is revolutionizing structural biology. Trends Biochem Sci *40*, 49-57.

Balibar, C.J., Vaillancourt, F.H., and Walsh, C.T. (2005). Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. Chem Biol *12*, 1189-1200.

Baltz, R.H. (2014). Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. ACS synthetic biology *3*, 748-758.

Bashford, D., and Karplus, M. (1990). pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. Biochemistry *29*, 10219-10225.

Belshaw, P.J., Roy, R.S., Kelleher, N.L., and Walsh, C.T. (1998). Kinetics and regioselectivity of peptide-to-heterocycle conversions by microcin B17 synthetase. Chem Biol *5*, 373-384.

Belshaw, P.J., Walsh, C.T., and Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. Science *284*, 486-489.

Berg, T.L., Froholm, L.O., and Laland, S.G. (1965). The Biosynthesis of Gramicidin S in a Cell-Free System. Biochem J *96*, 43-52.

Bergendahl, V., Linne, U., and Marahiel, M.A. (2002). Mutational analysis of the C-domain in nonribosomal peptide synthesis. Eur J Biochem *269*, 620-629.

Beringer, M., Bruell, C., Xiong, L., Pfister, P., Bieling, P., Katunin, V.I., Mankin, A.S., Bottger, E.C., and Rodnina, M.V. (2005). Essential mechanisms in the catalysis of peptide bond formation on the ribosome. J Biol Chem *280*, 36065-36072.

Bessalle, R., Kapitkovsky, A., Gorea, A., Shalit, I., and Fridkin, M. (1990). All-D-magainin: chirality, antimicrobial activity and proteolytic resistance. FEBS letters *274*, 151-155.

Biegert, A., and Soding, J. (2009). Sequence context-specific profiles for homology searching. Proc Natl Acad Sci U S A *106*, 3770-3775.

Bischoff, D., Bister, B., Bertazzo, M., Pfeifer, V., Stegmann, E., Nicholson, G.J., Keller, S., Pelzer, S., Wohlleben, W., and Sussmuth, R.D. (2005). The biosynthesis of vancomycin-type glycopeptide antibiotics--a model for oxidative side-chain cross-linking by oxygenases coupled to the action of peptide synthetases. Chembiochem *6*, 267-272.

Bloudoff, K., Alonzo, D.A., and Schmeing, T.M. (2016). Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases. Cell Chem Biol *23*, 331-339.

Bloudoff, K., Fage, C.D., Marahiel, M.A., and Schmeing, T.M. (2017). Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. Proc Natl Acad Sci U S A *114*, 95-100.

Bloudoff, K., Rodionov, D., and Schmeing, T.M. (2013). Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. J Mol Biol *425*, 3137-3150.

Bloudoff, K., and Schmeing, T.M. (2013). Crystallization and preliminary crystallographic analysis of the first condensation domain of viomycin synthetase. Acta Crystallogr Sect F Struct Biol Cryst Commun *69*, 412-415.

Bruner, S.D., Weber, T., Kohli, R.M., Schwarzer, D., Marahiel, M.A., Walsh, C.T., and Stubbs, M.T. (2002). Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. Structure *10*, 301-310.

Brunger, A.T. (2007). Version 1.2 of the crystallography and NMR system. Nat Protoc *2*, 2728-2733.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S.*, et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54*, 905-921.

Buglino, J., Onwueme, K.C., Ferreras, J.A., Quadri, L.E., and Lima, C.D. (2004). Crystal structure of PapA5, a phthiocerol dimycocerosyl transferase from Mycobacterium tuberculosis. J Biol Chem *279*, 30634-30642.

Caboche, S., Pupin, M., Leclere, V., Fontaine, A., Jacques, P., and Kucherov, G. (2008). NORINE: a database of nonribosomal peptides. Nucleic Acids Res *36*, D326-331.

Calcott, M.J., and Ackerley, D.F. (2015). Portability of the thiolation domain in recombinant pyoverdine non-ribosomal peptide synthetases. BMC Microbiol *15*, 162.

Calcott, M.J., Owen, J.G., Lamont, I.L., and Ackerley, D.F. (2014). Biosynthesis of novel Pyoverdines by domain substitution in a nonribosomal peptide synthetase of Pseudomonas aeruginosa. Appl Environ Microbiol *80*, 5723-5731.

Caminero, J.A., Sotgiu, G., Zumla, A., and Migliori, G.B. (2010). Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis. Lancet Infect Dis *10*, 621-629.

Challis, G.L., Ravel, J., and Townsend, C.A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem Biol *7*, 211-224.

Chen, H., O'Connor, S., Cane, D.E., and Walsh, C.T. (2001). Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. Chem Biol *8*, 899-912.

Chen, W.-H., Li, K., Guntaka, N.S., and Bruner, S.D. (2016). Interdomain and Intermodule Organization in Epimerization Domain Containing Nonribosomal Peptide Synthetases. ACS chemical biology *11*, 2293-2303.

Cheng, Y.Q. (2006). Deciphering the biosynthetic codes for the potent anti-SARS-CoV cyclodepsipeptide valinomycin in Streptomyces tsusimaensis ATCC 15141. Chembiochem *7*, 471-477.

Chong, P.P., Podmore, S.M., Kieser, H.M., Redenbach, M., Turgay, K., Marahiel, M., Hopwood, D.A., and Smith, C.P. (1998). Physical identification of a chromosomal locus encoding biosynthetic genes for the lipopeptide calcium-dependent antibiotic (CDA) of Streptomyces coelicolor A3(2). Microbiology *144 (Pt 1)*, 193-199.

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P.*, et al.* (2012). CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. PLoS Comput Biol *8*, e1002708.

Chowdhury, G., Cho, S.H., Pegg, A.E., and Guengerich, F.P. (2013). Detection and characterization of 1,2-dibromoethane-derived DNA crosslinks formed with O(6) -alkylguanine-DNA alkyltransferase. Angew Chem Int Ed Engl *52*, 12879-12882.

Churro, C., Alverca, E., Sam-Bento, F., Paulino, S., Figueira, V.C., Bento, A.J., Prabhakar, S., Lobo, A.M., Calado, A.J., and Pereira, P. (2009). Effects of bacillamide and newly synthesized derivatives on the growth of cyanobacteria and microalgae cultures. J Appl Phycol *21*, 429-442.

Clugston, S.L., Sieber, S.A., Marahiel, M.A., and Walsh, C.T. (2003). Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a (D)C(L) catalyst. Biochemistry *42*, 12095-12104.

Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. EMBO J *16*, 4174-4183.

Cramer, R.A., Jr., Stajich, J.E., Yamanaka, Y., Dietrich, F.S., Steinbach, W.J., and Perfect, J.R. (2006). Phylogenomic analysis of non-ribosomal peptide synthetases in the genus Aspergillus. Gene *383*, 24-32.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

De Crecy-Lagard, V., Marliere, P., and Saurin, W. (1995). Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation. C R Acad Sci III *318*, 927-936.

de Vries, S.J., van Dijk, M., and Bonvin, A.M. (2010). The HADDOCK web server for data-driven biomolecular docking. Nat Protoc *5*, 883-897.

Dessen, A., Tang, J., Schmidt, H., Stahl, M., Clark, J.D., Seehra, J., and Somers, W.S. (1999). Crystal structure of human cytosolic phospholipase A2 reveals a novel topology and catalytic mechanism. Cell *97*, 349-360.

Di Lorenzo, M., Poppelaars, S., Stork, M., Nagasawa, M., Tolmasky, M.E., and Crosa, J.H. (2004). A nonribosomal peptide synthetase with a novel domain organization is essential for siderophore biosynthesis in Vibrio anguillarum. J Bacteriol *186*, 7327-7336.

Di Lorenzo, M., Stork, M., Naka, H., Tolmasky, M.E., and Crosa, J.H. (2008). Tandem heterocyclization domains in a nonribosomal peptide synthetase essential for siderophore biosynthesis in Vibrio anguillarum. Biometals *21*, 635-648.

Ding, Y., Rath, C.M., Bolduc, K.L., Hakansson, K., and Sherman, D.H. (2011). Chemoenzymatic synthesis of cryptophycin anticancer agents by an ester bond-forming non-ribosomal peptide synthetase module. J Am Chem Soc *133*, 14492-14495.

Dittmann, J., Wenger, R.M., Kleinkauf, H., and Lawen, A. (1994). Mechanism of cyclosporin A biosynthesis. Evidence for synthesis via a single linear undecapeptide precursor. J Biol Chem *269*, 2841-2846.

Dorrestein, P.C., Yeh, E., Garneau-Tsodikova, S., Kelleher, N.L., and Walsh, C.T. (2005). Dichlorination of a pyrrolyl-S-carrier protein by FADH2-dependent halogenase PltA during pyoluteorin biosynthesis. Proc Natl Acad Sci U S A *102*, 13843-13848.

Dorsey, C.W., Tomaras, A.P., Connerly, P.L., Tolmasky, M.E., Crosa, J.H., and Actis, L.A. (2004). The siderophore-mediated iron acquisition systems of Acinetobacter baumannii ATCC 19606 and Vibrio anguillarum 775 are structurally and functionally related. Microbiology *150*, 3657-3667.

Dowling, D.P., Kung, Y., Croft, A.K., Taghizadeh, K., Kelly, W.L., Walsh, C.T., and Drennan, C.L. (2016). Structural elements of an NRPS cyclization domain and its intermodule docking domain. Proc Natl Acad Sci U S A *113*, 12432-12437.

Drake, E.J., Miller, B.R., Shi, C., Tarrasch, J.T., Sundlov, J.A., Allen, C.L., Skiniotis, G., Aldrich, C.C., and Gulick, A.M. (2016). Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. Nature *529*, 235-238.

Drechsel, H., Stephan, H., Lotz, R., Haag, H., Zahner, H., Hantke, K., and Jung, G. (1995). Structure Elucidation of Yersiniabactin, a Siderophore from Highly Virulent Yersinia Strains. Liebigs Ann, 1727-1733.

Du, L., Chen, M., Zhang, Y., and Shen, B. (2003). BlmIII and BlmIV nonribosomal peptide synthetase-catalyzed biosynthesis of the bleomycin bithiazole moiety involving both in cis and in trans aminoacylation. Biochemistry *42*, 9731-9740.

Du, L., He, Y., and Luo, Y. (2008). Crystal structure and enantiomer selection by D-alanyl carrier protein ligase DltA from Bacillus cereus. Biochemistry *47*, 11473-11480.

Du, L., and Lou, L. (2010). PKS and NRPS release mechanisms. Nat Prod Rep *27*, 255-278.

Duerfahrt, T., Doekel, S., Sonke, T., Quaedflieg, P.J., and Marahiel, M.A. (2003). Construction of hybrid peptide synthetases for the production of alpha-l-aspartyl-l-phenylalanine, a precursor for the high-intensity sweetener aspartame. Eur J Biochem *270*, 4555-4563.

Duerfahrt, T., Eppelmann, K., Muller, R., and Marahiel, M.A. (2004). Rational design of a bimodular model system for the investigation of heterocyclization in nonribosomal peptide biosynthesis. Chem Biol *11*, 261-271.

Dunbar, K.L., Melby, J.O., and Mitchell, D.A. (2012). YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. Nat Chem Biol *8*, 569-575.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Ehmann, D.E., Trauger, J.W., Stachelhaus, T., and Walsh, C.T. (2000). Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. Chem Biol *7*, 765-772.

Eikhom, T.S., Jonsen, J., Laland, S., and Refsvik, T. (1963). On the Biosynthesis of Gramicidin S. Biochimica et biophysica acta *76*, 465-468.

Ekici, O.D., Paetzel, M., and Dalbey, R.E. (2008). Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. Protein Sci *17*, 2023-2037.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr Sect D-Biol Crystallogr *66*, 486-501.

Evans, B.S., Chen, Y., Metcalf, W.W., Zhao, H., and Kelleher, N.L. (2011). Directed evolution of the nonribosomal peptide synthetase AdmK generates new andrimid derivatives in vivo. Chem Biol *18*, 601-607.

Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, A.D., McMahon, M.D., and Thomas, M.G. (2008). Nonribosomal peptide synthetases involved in the production of medically relevant natural products. Mol Pharm *5*, 191-211.

Fersht, A.R. (1974). Catalysis, binding and enzyme-substrate complementarity. Proc R Soc Lond B Biol Sci *187*, 397-407.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapidab-initioshape determination in small-angle scattering. Journal of Applied Crystallography *42*, 342-346.

Frueh, D.P., Arthanari, H., Koglin, A., Vosburg, D.A., Bennett, A.E., Walsh, C.T., and Wagner, G. (2008). Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. Nature *454*, 903-906.

Fujikawa, K., Suzuki, T., and Kurahashi, K. (1966). Incorporation of L-leucine-C14 into tyrocidine by a cell-free preparation of Bacillus brevis Dubos strain. J Biochem *60*, 216-218.

Fujimori, D.G., Hrvatin, S., Neumann, C.S., Strieker, M., Marahiel, M.A., and Walsh, C.T. (2007). Cloning and characterization of the biosynthetic gene cluster for kutznerides. Proc Natl Acad Sci U S A *104*, 16498-16503.

Gao, X., Haynes, S.W., Ames, B.D., Wang, P., Vien, L.P., Walsh, C.T., and Tang, Y. (2012). Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. Nat Chem Biol *8*, 823-830.

Gatto, G.J., Jr., McLoughlin, S.M., Kelleher, N.L., and Walsh, C.T. (2005). Elucidating the substrate specificity and condensation domain activity of FkbP, the FK520 pipecolate-incorporating enzyme. Biochemistry *44*, 5993-6002.

Gaudelli, N.M., Long, D.H., and Townsend, C.A. (2015). beta-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. Nature *520*, 383-387.

Gehring, A.M., Mori, I.I., Perry, R.D., and Walsh, C.T. (1998). The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of yersinia pestis. Biochemistry *37*, 17104.

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

Gokhale, R.S., Hunziker, D., Cane, D.E., and Khosla, C. (1999). Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. Chem Biol *6*, 117-125.

Goodrich, A.C., Harden, B.J., and Frueh, D.P. (2015). Solution structure of a nonribosomal peptide synthetase carrier protein loaded with Its substrate reveals transient, well-defined contacts. J Am Chem Soc *137*, 12100-12109.

Gordon, J.C., Myers, J.B., Folta, T., Shoja, V., Heath, L.S., and Onufriev, A. (2005). H++: a server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Res *33*, W368-371.

Green, R., and Lorsch, J.R. (2002). The path to perdition is paved with protons. Cell *110*, 665-668.

Green, R., Switzer, C., and Noller, H.F. (1998). Ribosome-catalyzed peptide-bond formation with an A-site substrate covalently linked to 23S ribosomal RNA. Science *280*, 286-289.

Guest, J.R. (1987). Functional Implications of Structural Homologies between Chloramphenicol Acetyltransferase and Dihydrolipoamide Acetyltransferase. Fems Microbiology Letters *44*, 417-422.

Gulick, A.M., Starai, V.J., Horswill, A.R., Homick, K.M., and Escalante-Semerena, J.C. (2003). The 1.75 A crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. Biochemistry *42*, 2866-2873.

Gunsior, M., Breazeale, S.D., Lind, A.J., Ravel, J., Janc, J.W., and Townsend, C.A. (2004). The biosynthetic gene cluster for a monocyclic beta-lactam antibiotic, nocardicin A. Chem Biol *11*, 927-938.

Hahn, M., and Stachelhaus, T. (2006). Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. Proc Natl Acad Sci U S A *103*, 275-280.

Harden, B.J., Mishra, S.H., and Frueh, D.P. (2015). Effortless assignment with 4D covariance sequential correlation maps. J Magn Reson *260*, 83-88.

Haslinger, K., Peschke, M., Brieke, C., Maximowitsch, E., and Cryle, M.J. (2015). X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. Nature *521*, 105-109.

Haynes, S.W., Ames, B.D., Gao, X., Tang, Y., and Walsh, C.T. (2011). Unraveling terminal C-domain-mediated condensation in fungal biosynthesis of imidazoindolone metabolites. Biochemistry *50*, 5668-5679.

Helmetag, V., Samel, S.A., Thomas, M.G., Marahiel, M.A., and Essen, L.O. (2009). Structural basis for the erythro-stereospecificity of the L-arginine oxygenase VioC in viomycin biosynthesis. FEBS J *276*, 3669-3682.

Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F., Wilkinson, B., Sidebottom, P.J., Rudd, B.A., Hayes, M.A*., et al.* (2002). Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from Streptomyces coelicolor. Chem Biol *9*, 1175-1187.

Hollup, S.M., Salensminde, G., and Reuter, N. (2005). WEBnm@: a web application for normal mode analyses of proteins. BMC Bioinformatics *6*, 52.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res *38*, W545-549.

Holzbaur, I.E., Harris, R.C., Bycroft, M., Cortes, J., Bisang, C., Staunton, J., Rudd, B.A., and Leadlay, P.F. (1999). Molecular basis of Celmer's rules: the role of two ketoreductase domains in the control of chirality by the erythromycin modular polyketide synthase. Chem Biol *6*, 189-195.

Hoppert, M., Gentzsch, C., and Schorgendorfer, K. (2001). Structure and localization of cyclosporin synthetase, the key enzyme of cyclosporin biosynthesis in Tolypocladium inflatum. Arch Microbiol *176*, 285-293.

Hopwood, D.A., and Wright, H.M. (1983). CDA is a new chromosomally-determined antibiotic from Streptomyces coelicolor A3(2). J Gen Microbiol *129*, 3575-3579.

Hoyer, K.M., Mahlert, C., and Marahiel, M.A. (2007). The iterative gramicidin s thioesterase catalyzes peptide ligation and cyclization. Chem Biol *14*, 13-22.

Hsiao, Y.S., Jogl, G., and Tong, L. (2006). Crystal structures of murine carnitine acetyltransferase in ternary complexes with its substrates. J Biol Chem *281*, 28480-28487.

Huang, H., Chopra, R., Verdine, G.L., and Harrison, S.C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science *282*, 1669-1675.

Hur, G.H., Meier, J.L., Baskin, J., Codelli, J.A., Bertozzi, C.R., Marahiel, M.A., and Burkart, M.D. (2009). Crosslinking studies of protein-protein interactions in nonribosomal peptide biosynthesis. Chem Biol *16*, 372-381.

Hur, G.H., Vickery, C.R., and Burkart, M.D. (2012). Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. Nat Prod Rep *29*, 1074-1098.

Ivanova, V., Kolarova, M., Aleksieva, K., Graffe, U., Dahse, H.M., and Laatsch, H. (2007). Microbiaeratin, a new natural indole alkaloid from a Microbispora aerata strain, isolated from Livingston Island, Antarctica. Prep Biochem Biotech *37*, 161-168.

Jaitzig, J., Li, J., Sussmuth, R.D., and Neubauer, P. (2014). Reconstituted biosynthesis of the nonribosomal macrolactone antibiotic valinomycin in Escherichia coli. ACS synthetic biology *3*, 432-438.

Janin, J., and Chothia, C. (1976). Stability and specificity of protein-protein interactions: the case of the trypsin-trypsin inhibitor complexes. J Mol Biol *100*, 197-211.

Jaremko, M.J., Lee, D.J., Opella, S.J., and Burkart, M.D. (2015). Structure and substrate sequestration in the pyoluteorin type II peptidyl carrier protein PltL. J Am Chem Soc *137*, 11546-11549.

Jencks, W.P. (1969). Catalysis in chemistry and enzymology (New York,: McGraw-Hill).

Jeong, S.Y., Ishida, K., Ito, Y., Okada, S., and Murakami, M. (2003). Bacillamide, a novel algicide from the marine bacterium, Bacillus sp SY-1, against the harmful dinoflagellate, Cochlodinium polykrikoides. Tetrahedron Letters *44*, 8005-8007.

Kallow, W., Neuhof, T., Arezi, B., Jungblut, P., and von Dohren, H. (1997). Penicillin biosynthesis: intermediates of biosynthesis of delta-L-alpha-aminoadipyl-L-cysteinyl-D-valine formed by ACV synthetase from Acremonium chrysogenum. FEBS Lett *414*, 74-78.

Kawai, Y., Ishii, Y., Arakawa, K., Uemura, K., Saitoh, B., Nishimura, J., Kitazawa, H., Yamazaki, Y., Tateno, Y., Itoh, T.*, et al.* (2004). Structural and functional differences in two cyclic bacteriocins with the same sequences produced by lactobacilli. Appl Environ Microbiol *70*, 2906-2911.

Keating, T.A., Marshall, C.G., and Walsh, C.T. (2000a). Reconstitution and characterization of the Vibrio cholerae vibriobactin synthetase from VibB, VibE, VibF, and VibH. Biochemistry *39*, 15522-15530.

Keating, T.A., Marshall, C.G., Walsh, C.T., and Keating, A.E. (2002). The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. Nat Struct Biol *9*, 522-526.

Keating, T.A., Miller, D.A., and Walsh, C.T. (2000b). Expression, purification, and characterization of HMWP2, a 229 kDa, six domain protein subunit of Yersiniabactin synthetase. Biochemistry *39*, 4729-4739.

Kelly, W.L., Hillson, N.J., and Walsh, C.T. (2005). Excision of the epothilone synthetase B cyclization domain and demonstration of in trans condensation/cyclodehydration activity. Biochemistry *44*, 13385-13393.

Kim, A.R., Rylett, R.J., and Shilton, B.H. (2006). Substrate binding and catalytic mechanism of human choline acetyltransferase. Biochemistry *45*, 14621-14631.

Koglin, A., Mofid, M.R., Lohr, F., Schafer, B., Rogov, V.V., Blum, M.M., Mittag, T., Marahiel, M.A., Bernhard, F., and Dotsch, V. (2006). Conformational switches modulate protein interactions in peptide antibiotic synthetases. Science *312*, 273-276.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography *36*, 1277-1282.

Konda, Y., Suzuki, Y., Omura, S., and Onda, M. (1976). Alkaloid from Thermoactinomyces species. Chem Pharm Bull (Tokyo) *24*, 92-96.

Konz, D., Klens, A., Schorgendorfer, K., and Marahiel, M.A. (1997). The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases. Chem Biol *4*, 927-937.

Konz, D., and Marahiel, M.A. (1999). How do peptide synthetases generate structural diversity? Chem Biol *6*, R39-48.

Kopp, F., Linne, U., Oberthur, M., and Marahiel, M.A. (2008). Harnessing the chemical activation inherent to carrier protein-bound thioesters for the characterization of lipopeptide fatty acid tailoring enzymes. J Am Chem Soc *130*, 2656-2666.

Kopp, F., and Marahiel, M.A. (2007). Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. Nat Prod Rep *24*, 735-749.

Kozin, M.B., and Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models. Journal of Applied Crystallography *34*, 33-41.

Kraas, F.I., Giessen, T.W., and Marahiel, M.A. (2012). Exploring the mechanism of lipid transfer during biosynthesis of the acidic lipopeptide antibiotic CDA. FEBS Lett *586*, 283-288.

Kries, H., Niquille, D.L., and Hilvert, D. (2015). A subdomain swap strategy for reengineering nonribosomal peptides. Chem Biol *22*, 640-648.

Kuhlenkoetter, S., Wintermeyer, W., and Rodnina, M.V. (2011). Different substrate-dependent transition states in the active site of the ribosome. Nature *476*, 351-354.

Labby, K.J., Watsula, S.G., and Garneau-Tsodikova, S. (2015). Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. Nat Prod Rep *32*, 641-653.

Lai, J.R., Fischbach, M.A., Liu, D.R., and Walsh, C.T. (2006a). Localized protein interaction surfaces on the EntB carrier protein revealed by combinatorial mutagenesis and selection. J Am Chem Soc *128*, 11002-11003.

Lai, J.R., Fischbach, M.A., Liu, D.R., and Walsh, C.T. (2006b). A protein interaction surface in nonribosomal peptide synthesis mapped by combinatorial mutagenesis and selection. Proc Natl Acad Sci U S A *103*, 5314-5319.

Lakey, J.H., Lea, E.J., Rudd, B.A., Wright, H.M., and Hopwood, D.A. (1983). A new channel-forming antibiotic from Streptomyces coelicolor A3(2) which requires calcium for its activity. J Gen Microbiol *129*, 3565-3573.

Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., and Walsh, C.T. (1996). A new enzyme superfamily - the phosphopantetheinyl transferases. Chem Biol *3*, 923-936.

Lautru, S., and Challis, G.L. (2004). Substrate recognition by nonribosomal peptide synthetase multi-enzymes. Microbiology *150*, 1629-1636.

Lawen, A., and Zocher, R. (1990). Cyclosporin synthetase. The most complex peptide synthesizing multienzyme polypeptide so far described. J Biol Chem *265*, 11355-11360.

Lee, B., and Richards, F.M. (1971). The interpretation of protein structures: estimation of static accessibility. J Mol Biol *55*, 379-400.

Leslie, A.G. (1990). Refined crystal structure of type III chloramphenicol acetyltransferase at 1.75 A resolution. J Mol Biol *213*, 167-186.

Leslie, A.G.W., and Powell, H.R. (2007). Processing diffraction data with MOSFLM. Nato Sci Ser Ii Math *245*, 41-51.

Lewendon, A., Murray, I.A., Shaw, W.V., Gibbs, M.R., and Leslie, A.G. (1994). Replacement of catalytic histidine-195 of chloramphenicol acetyltransferase: evidence for a general base role for glutamate. Biochemistry *33*, 1944-1950.

Li, J., Derewenda, U., Dauter, Z., Smith, S., and Derewenda, Z.S. (2000). Crystal structure of the Escherichia coli thioesterase II, a homolog of the human Nef binding enzyme. Nat Struct Biol *7*, 555-559.

Li, R., Oliver, R.A., and Townsend, C.A. (2016). Identification and Characterization of the Sulfazecin Monobactam Biosynthetic Gene Cluster. Cell Chem Biol.

Lin, S., Huang, T., Horsman, G.P., Huang, S.X., Guo, X., and Shen, B. (2012). Specificity of the ester bond forming condensation enzyme SgcC5 in C-1027 biosynthesis. Org Lett *14*, 2300-2303.

Lin, S., Van Lanen, S.G., and Shen, B. (2009). A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis. Proc Natl Acad Sci U S A *106*, 4183-4188.

Lindahl, E., Azuara, C., Koehl, P., and Delarue, M. (2006). NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. Nucleic Acids Res *34*, W52-56.

Linne, U., Doekel, S., and Marahiel, M.A. (2001). Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases. Biochemistry *40*, 15824-15834.

Linne, U., and Marahiel, M.A. (2000). Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. Biochemistry *39*, 10439-10447.

Liu, Y., Zheng, T., and Bruner, S.D. (2011). Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. Chem Biol *18*, 1482-1488.

Lohman, J.R., Ma, M., Cuff, M.E., Bigelow, L., Bearden, J., Babnigg, G., Joachimiak, A., Phillips, G.N., and Shen, B. (2014). The crystal structure of BlmI as a model for nonribosomal peptide synthetase peptidyl carrier proteins. Proteins *82*, 1210-1218.

Luo, L., Burkart, M.D., Stachelhaus, T., and Walsh, C.T. (2001). Substrate recognition and selection by the initiation module PheATE of gramicidin S synthetase. J Am Chem Soc *123*, 11208-11218.

Luo, L., Kohli, R.M., Onishi, M., Linne, U., Marahiel, M.A., and Walsh, C.T. (2002). Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. Biochemistry *41*, 9184-9196.

Mach, B., Reich, E., and Tatum, E.L. (1963). Separation of the Biosynthesis of the Antibiotic Polypeptide Tyrocidine from Protein Biosynthesis. Proc Natl Acad Sci U S A *50*, 175-181.

Magarvey, N.A., Ehling-Schulz, M., and Walsh, C.T. (2006). Characterization of the cereulide NRPS alpha-hydroxy acid specifying modules: activation of alpha-keto

acids and chiral reduction on the assembly line. J Am Chem Soc *128*, 10698-10699.

Maly, D.J., Allen, J.A., and Shokat, K.M. (2004). A mechanism-based cross-linker for the identification of kinase-substrate pairs. J Am Chem Soc *126*, 9160-9161.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. Chem Rev *97*, 2651-2674.

Marechal, J.D., and Perahia, D. (2008). Use of normal modes for structural modeling of proteins: the case study of rat heme oxygenase 1. Eur Biophys J *37*, 1157-1165.

Marshall, C.G., Burkart, M.D., Keating, T.A., and Walsh, C.T. (2001). Heterocycle formation in vibriobactin biosynthesis: alternative substrate utilization and identification of a condensed intermediate. Biochemistry *40*, 10655-10663.

Marshall, C.G., Burkart, M.D., Meray, R.K., and Walsh, C.T. (2002a). Carrier protein recognition in siderophore-producing nonribosomal peptide synthetases. Biochemistry *41*, 8429-8437.

Marshall, C.G., Hillson, N.J., and Walsh, C.T. (2002b). Catalytic mapping of the vibriobactin biosynthetic enzyme VibF. Biochemistry *41*, 244-250.

Martinez, V., and Davyt, D. (2013). Total syntheses of bacillamide C and neobacillamide A; revision of their absolute configurations. Tetrahedron-Asymmetr *24*, 1572-1575.

Marxen, S., Stark, T.D., Rutschle, A., Lucking, G., Frenzel, E., Scherer, S., Ehling-Schulz, M., and Hofmann, T. (2015). Depsipeptide Intermediates Interrogate Proposed Biosynthesis of Cereulide, the Emetic Toxin of Bacillus cereus. Sci Rep *5*, 10637.

Mattevi, A., Obmolova, G., Schulze, E., Kalk, K.H., Westphal, A.H., de Kok, A., and Hol, W.G. (1992). Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. Science *255*, 1544-1550.

May, J.J., Kessler, N., Marahiel, M.A., and Stubbs, M.T. (2002). Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. Proc Natl Acad Sci U S A *99*, 12120-12125.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

McMahon, M.D., Rush, J.S., and Thomas, M.G. (2012). Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in Mycobacterium tuberculosis. J Bacteriol *194*, 2809-2818.

Meyer, S., Kehr, J.C., Mainz, A., Dehm, D., Petras, D., Sussmuth, R.D., and Dittmann, E. (2016). Biochemical Dissection of the Natural Diversification of Microcystin Provides Lessons for Synthetic Biology of NRPS. Cell Chem Biol *23*, 462-471.

Miller, B.R., Drake, E.J., Shi, C., Aldrich, C.C., and Gulick, A.M. (2016). Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture. J Biol Chem *291*, 22559-22571.

Miller, B.R., and Gulick, A.M. (2016). Structural Biology of Nonribosomal Peptide Synthetases. Methods Mol Biol *1401*, 3-29.

Miller, D.A., Luo, L., Hillson, N., Keating, T.A., and Walsh, C.T. (2002). Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. Chem Biol *9*, 333-344.

Minasov, G., Wang, X., and Shoichet, B.K. (2002). An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation. J Am Chem Soc *124*, 5333-5340.

Mishra, S.H., and Frueh, D.P. (2015). Assignment of methyl NMR resonances of a 52 kDa protein with residue-specific 4D correlation maps. J Biomol NMR *62*, 281-290.

Mitchell, C.A., Shi, C., Aldrich, C.C., and Gulick, A.M. (2012). Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. Biochemistry *51*, 3252-3263.

Modolell, J., and Vazquez (1977). The inhibition of ribosomal translocation by viomycin. Eur J Biochem *81*, 491-497.

Mootz, H.D., Kessler, N., Linne, U., Eppelmann, K., Schwarzer, D., and Marahiel, M.A. (2002). Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. J Am Chem Soc *124*, 10980-10981.

Mootz, H.D., and Marahiel, M.A. (1997). The tyrocidine biosynthesis operon of Bacillus brevis: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. J Bacteriol *179*, 6843-6850.

Mootz, H.D., Schwarzer, D., and Marahiel, M.A. (2000). Construction of hybrid peptide synthetases by module and domain fusions. Proc Natl Acad Sci U S A *97*, 5848-5853.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr *53*, 240-255.

Nguyen, K.T., Ritz, D., Gu, J.Q., Alexander, D., Chu, M., Miao, V., Brian, P., and Baltz, R.H. (2006). Combinatorial biosynthesis of novel antibiotics related to daptomycin. Proc Natl Acad Sci U S A *103*, 17462-17467.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol *276*, 307-326.

Patel, H.M., and Walsh, C.T. (2001). In vitro reconstitution of the Pseudomonas aeruginosa nonribosomal peptide synthesis of pyochelin: characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities. Biochemistry *40*, 9023-9031.

Payne, J.A., Schoppet, M., Hansen, M.H., and Cryle, M.J. (2016). Diversity of nature's assembly lines - recent discoveries in non-ribosomal peptide synthesis. Mol Biosyst *13*, 9-22.

Peschke, M., Gonsior, M., Sussmuth, R.D., and Cryle, M.J. (2016a). Understanding the crucial interactions between Cytochrome P450s and non-ribosomal peptide synthetases during glycopeptide antibiotic biosynthesis. Curr Opin Struct Biol *41*, 46-53.

Peschke, M., Haslinger, K., Brieke, C., Reinstein, J., and Cryle, M.J. (2016b). Regulation of the P450 Oxygenation Cascade Involved in Glycopeptide Antibiotic Biosynthesis. Journal of the American Chemical Society *138*, 6746-6753.

Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., and Khosla, C. (2001). Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. Science *291*, 1790-1792.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J Comput Chem *26*, 1781-1802.

Quadri, L.E., Keating, T.A., Patel, H.M., and Walsh, C.T. (1999). Assembly of the Pseudomonas aeruginosa nonribosomal peptide siderophore pyochelin: In vitro reconstitution of aryl-4, 2-bisthiazoline synthetase activity from PchD, PchE, and PchF. Biochemistry *38*, 14941-14954.

Quadri, L.E., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P., and Walsh, C.T. (1998). Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. Biochemistry *37*, 1585-1595.

Rausch, C., Hoof, I., Weber, T., Wohlleben, W., and Huson, D.H. (2007). Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC Evol Biol *7*, 78.

Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D.H. (2005). Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucleic Acids Res *33*, 5799-5808.

Reger, A.S., Wu, R., Dunaway-Mariano, D., and Gulick, A.M. (2008). Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. Biochemistry *47*, 8016-8025.

Reimer, J.M., Aloise, M.N., Harrison, P.M., and Schmeing, T.M. (2016). Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. Nature *529*, 239-242.

Reimmann, C., Patel, H.M., Serino, L., Barone, M., Walsh, C.T., and Haas, D. (2001). Essential PchG-dependent reduction in pyochelin biosynthesis of Pseudomonas aeruginosa. J Bacteriol *183*, 813-820.

Richardt, A., Kemme, T., Wagner, S., Schwarzer, D., Marahiel, M.A., and Hovemann, B.T. (2003). Ebony, a novel nonribosomal peptide synthetase for beta-alanine conjugation with biogenic amines in Drosophila. J Biol Chem *278*, 41160-41166.

Roche, E.D., and Walsh, C.T. (2003). Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. Biochemistry *42*, 1334-1344.

Roongsawang, N., Lim, S.P., Washio, K., Takano, K., Kanaya, S., and Morikawa, M. (2005). Phylogenetic analysis of condensation domains in the nonribosomal peptide synthetases. FEMS Microbiol Lett *252*, 143-151.

Roy, R.S., Gehring, A.M., Milne, J.C., Belshaw, P.J., and Walsh, C.T. (1999). Thiazole and oxazole peptides: biosynthesis and molecular machinery. Nat Prod Rep *16*, 249-263.

Rydel, T.J., Williams, J.M., Krieger, E., Moshiri, F., Stallings, W.C., Brown, S.M., Pershing, J.C., Purcell, J.P., and Alibhai, M.F. (2003). The crystal structure, mutagenesis, and activity studies reveal that patatin is a lipid acyl hydrolase with a Ser-Asp catalytic dyad. Biochemistry *42*, 6696-6708.

Samel, S.A., Czodrowski, P., and Essen, L.O. (2014). Structure of the epimerization domain of tyrocidine synthetase A. Acta Crystallogr D Biol Crystallogr *70*, 1442-1452.

Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A., and Essen, L.O. (2007). Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. Structure *15*, 781-792.

Schauwecker, F., Pfennig, F., Grammel, N., and Keller, U. (2000). Construction and in vitro analysis of a new bi-modular polypeptide synthetase for synthesis of N-methylated acyl peptides. Chem Biol *7*, 287-297.

Schmeing, T.M., Huang, K.S., Kitchen, D.E., Strobel, S.A., and Steitz, T.A. (2005a). Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. Mol Cell *20*, 437-448.

Schmeing, T.M., Huang, K.S., Strobel, S.A., and Steitz, T.A. (2005b). An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. Nature *438*, 520-524.

Schmeing, T.M., and Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. Nature *461*, 1234-1242.

Schneider, A., and Marahiel, M.A. (1998). Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in Bacillus subtilis. Arch Microbiol *169*, 404-410.

Schneider, A., Stachelhaus, T., and Marahiel, M.A. (1998). Targeted alteration of the substrate specificity of peptide synthetases by rational module swapping. Mol Gen Genet *257*, 308-318.

Schneider, T.L., Shen, B., and Walsh, C.T. (2003). Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. Biochemistry *42*, 9722-9730.

Schoenafinger, G., Schracke, N., Linne, U., and Marahiel, M.A. (2006). Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. J Am Chem Soc *128*, 7406-7407.

Schuttelkopf, A.W., and van Aalten, D.M. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr *60*, 1355-1363.

Schwarzer, D., Finking, R., and Marahiel, M.A. (2003). Nonribosomal peptides: from genes to products. Nat Prod Rep *20*, 275-287.

Seyedsayamdost, M.R., Cleto, S., Carr, G., Vlamakis, H., Joao Vieira, M., Kolter, R., and Clardy, J. (2012). Mixing and matching siderophore clusters: structure and biosynthesis of serratiochelins from Serratia sp. V4. J Am Chem Soc *134*, 13550-13553.

Shen, B., Du, L., Sanchez, C., Edwards, D.J., Chen, M., and Murrell, J.M. (2002). Cloning and characterization of the bleomycin biosynthetic gene cluster from Streptomyces verticillus ATCC15003. J Nat Prod *65*, 422-431.

Shou, Q.Y., Feng, L.K., Long, Y.L., Han, J., Nunnery, J.K., Powell, D.H., and Butcher, R.A. (2016). A hybrid polyketide-nonribosomal peptide in nematodes that promotes larval survival. Nature Chemical Biology *12*, 770-+.

Silakowski, B., Schairer, H.U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S*., et al.* (1999). New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca DW4/3-1. J Biol Chem *274*, 37391-37399.

Snow, G.A. (1970). Mycobactins: iron-chelating growth factors from mycobacteria. Bacteriol Rev *34*, 99-125.

Socha, A.M., Long, R.A., and Rowley, D.C. (2007). Bacillamides from a hypersaline microbial mat bacterium. Journal of Natural Products *70*, 1793-1795.

Stachelhaus, T., and Marahiel, M.A. (1995). Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. J Biol Chem *270*, 6163-6169.

Stachelhaus, T., Mootz, H.D., Bergendahl, V., and Marahiel, M.A. (1998). Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. J Biol Chem *273*, 22773-22781.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol *6*, 493-505.

Stachelhaus, T., Schneider, A., and Marahiel, M.A. (1995). Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. Science *269*, 69-72.

Stachelhaus, T., and Walsh, C.T. (2000). Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. Biochemistry *39*, 5775-5787.

Stanley, R.E., Blaha, G., Grodzicki, R.L., Strickler, M.D., and Steitz, T.A. (2010). The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. Nat Struct Mol Biol *17*, 289-293.

Stegmann, E., Frasch, H.J., and Wohlleben, W. (2010). Glycopeptide biosynthesis in the context of basic cellular functions. Curr Opin Microbiol *13*, 595-602.

Stein, D.B., Linne, U., and Marahiel, M.A. (2005). Utility of epimerization domains for the redesign of nonribosomal peptide synthetases. FEBS J *272*, 4506-4520.

Stein, T., Kluge, B., Vater, J., Franke, P., Otto, A., and Wittmann-Liebold, B. (1995). Gramicidin S synthetase 1 (phenylalanine racemase), a prototype of amino acid racemases containing the cofactor 4'-phosphopantetheine. Biochemistry *34*, 4633-4642.

Stindl, A., and Keller, U. (1994). Epimerization of the D-valine portion in the biosynthesis of actinomycin D. Biochemistry *33*, 9358-9364.

Strieker, M., and Marahiel, M.A. (2009). The structural diversity of acidic lipopeptide antibiotics. Chembiochem *10*, 607-616.

Strieker, M., Tanovic, A., and Marahiel, M.A. (2010). Nonribosomal peptide synthetases: structures and dynamics. Curr Opin Struct Biol *20*, 234-240.

Sundlov, J.A., Shi, C., Wilson, D.J., Aldrich, C.C., and Gulick, A.M. (2012). Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. Chem Biol *19*, 188-198.

Sussmuth, R.D., and Mainz, A. (2017). Nonribosomal Peptide Synthesis-Principles and Prospects. Angew Chem Int Ed Engl *56*, 3770-3821.

Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. Journal of Applied Crystallography *28*, 768-773.

Tabata, N., Tomoda, H., Zhang, H., Uchida, R., and Omura, S. (1999). Zelkovamycin, a new cyclic peptide antibiotic from Streptomyces sp. K96-0670. II. Structure elucidation. J Antibiot (Tokyo) *52*, 34-39.

Takahashi, H., Sato, E., and Kurahashi, K. (1971). Racemization of phenylalanine by adenosine triphosphate-dependent phenylalanine racemase of Bacillus brevis Nagano. J Biochem *69*, 973-976.

Tang, G.L., Cheng, Y.Q., and Shen, B. (2007). Chain initiation in the leinamycin-producing hybrid nonribosomal peptide/polyketide synthetase from Streptomyces atroolivaceus S-140. Discrete, monofunctional adenylation enzyme and peptidyl carrier protein that directly load D-alanine. J Biol Chem *282*, 20273-20282.

Tanovic, A., Samel, S.A., Essen, L.O., and Marahiel, M.A. (2008). Crystal structure of the termination module of a nonribosomal peptide synthetase. Science *321*, 659-663.

Tarry, M.J., Haque, A.S., Bui, K.H., and Schmeing, T.M. (In revision). Structures of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture. Structure.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.W., Read, R.J., and Adams, P.D. (2008). Iterative model building,

structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallogr D Biol Crystallogr *64*, 61-69.

Tomino, S., and Kurahashi, K. (1964). Enzymic synthesis of D-phenylalanyl-L-prolyl-L-valine, a peptide sequence present in gramicidin S. Biochem Biophys Res Commun *17*, 288-293.

Townsend, C.A. (2016). Convergent biosynthetic pathways to beta-lactam antibiotics. Curr Opin Chem Biol *35*, 97-108.

Trauger, J.W., Kohli, R.M., Mootz, H.D., Marahiel, M.A., and Walsh, C.T. (2000). Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. Nature *407*, 215-218.

Trivedi, O.A., Arora, P., Vats, A., Ansari, M.Z., Tickoo, R., Sridharan, V., Mohanty, D., and Gokhale, R.S. (2005). Dissecting the mechanism and assembly of a complex virulence mycobacterial lipid. Mol Cell *17*, 631-643.

Tseng, C.C., Bruner, S.D., Kohli, R.M., Marahiel, M.A., Walsh, C.T., and Sieber, S.A. (2002). Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase. Biochemistry *41*, 13350-13359.

Ulrich, V., Peschke, M., Brieke, C., and Cryle, M.J. (2016). More than just recruitment: the X-domain influences catalysis of the first phenolic coupling reaction in A47934 biosynthesis by Cytochrome P450 StaH. Mol Biosyst.

Vagin, A., and Lebedev, A. (2015). MoRDa, an automatic molecular replacement pipeline. Acta Crystallographica Section A *A71*.

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J Mol Biol *229*, 105-124.

van Liempt, H., von Döhren, H., and Kleinkauf, H. (1989). delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from Aspergillus nidulans. The first enzyme in penicillin biosynthesis is a multifunctional peptide synthetase. Journal of Biological Chemistry *264*, 3680-3684.

van Wageningen, A.M., Kirkpatrick, P.N., Williams, D.H., Harris, B.R., Kershaw, J.K., Lennard, N.J., Jones, M., Jones, S.J., and Solenberg, P.J. (1998). Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. Chem Biol *5*, 155-162.

Van Wynsberghe, A., Li, G., and Cui, Q. (2004). Normal-mode analysis suggests protein flexibility modulation throughout RNA polymerase's functional cycle. Biochemistry *43*, 13083-13096.

Vater, J., Stein, T., Vollenbroich, D., Kruft, V., Wittmann-Liebold, B., Franke, P., Liu, L., and Zuber, P. (1997). The modular organization of multifunctional peptide synthetases. J Protein Chem *16*, 557-564.

Veatch, W. (1976). The structure of the gramicidin A transmembrane channel. J Supramol Struct *5*, 431(435)-451(455).

Vizcaino, M.I., and Crawford, J.M. (2015). The colibactin warhead crosslinks DNA. Nat Chem *7*, 411-417.

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness ofab initioshape determination in small-angle scattering. Journal of Applied Crystallography *36*, 860-864.

Vollbrecht, L., Steinmetz, H., Hofle, G., Oberer, L., Rihs, G., Bovermann, G., and von Matt, P. (2002). Argyrins, immunosuppressive cyclic peptides from myxobacteria. II. Structure elucidation and stereochemistry. J Antibiot (Tokyo) *55*, 715-721.

von Dohren, H. (2009). A survey of nonribosomal peptide synthetase (NRPS) genes in Aspergillus nidulans. Fungal Genet Biol *46 Suppl 1*, S45-52.

von Dohren, H., Dieckmann, R., and Pavela-Vrancic, M. (1999). The nonribosomal code. Chem Biol *6*, R273-279.

Vonrhein, C., Schlauderer, G.J., and Schulz, G.E. (1995). Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. Structure *3*, 483-490.

Wallin, G., and Aqvist, J. (2010). The transition state for peptide bond formation reveals the ribosome as a water trap. Proc Natl Acad Sci U S A *107*, 1888-1893.

Walsh, C.T. (2004). Polyketide and nonribosomal peptide antibiotics: modularity and versatility. Science *303*, 1805-1810.

Walsh, C.T. (2008). The chemical versatility of natural-product assembly lines. Acc Chem Res *41*, 4-10.

Wang, H., Fewer, D.P., Holm, L., Rouhiainen, L., and Sivonen, K. (2014). Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. Proc Natl Acad Sci U S A *111*, 9259-9264.

Weber, G., Schörgendorfer, K., Schneider-Scherzer, E., and Leitner, E. (1994). The peptide synthetase catalyzing cyclosporine production in Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading frame. Current Genetics *26*, 120-125.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Muller, R., Wohlleben, W.*, et al.* (2015). antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res *43*, W237-243.

Weber, T., and Marahiel, M.A. (2001). Exploring the domain structure of modular nonribosomal peptide synthetases. Structure *9*, R3-9.

Weissman, K.J. (2015). The structural biology of biosynthetic megaenzymes. Nat Chem Biol *11*, 660-670.

Wessels, P., von Dohren, H., and Kleinkauf, H. (1996). Biosynthesis of acylpeptidolactones of the daptomycin type. A comparative analysis of peptide synthetases forming A21978C and A54145. Eur J Biochem *242*, 665-673.

Winkler, R. (2010). ESIprot: a universal tool for charge state determination and molecular weight calculation of proteins from electrospray ionization mass spectrometry data. Rapid Commun Mass Spectrom *24*, 285-294.

World_Health_Organization (2013). WHO Model List of Essential Medicines, http://apps.who.int/iris/bitstream/10665/93142/10661/EML_10618_eng.pdf.

Xiao, H., Murakami, H., Suga, H., and Ferre-D'Amare, A.R. (2008). Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. Nature *454*, 358-361.

Yamada, M., and Kurahashi, K. (1968). Adenosine triphosphate and pyrophosphate dependent phenylalanine racemase of Bacillus brevis Nagano. J Biochem *63*, 59-69.

Yang, W., and Drueckhammer, D.G. (2000). Computational studies of the aminolysis of oxoesters and thioesters in aqueous solution. Org Lett *2*, 4133-4136.

Yim, G., Thaker, M.N., Koteva, K., and Wright, G. (2014). Glycopeptide antibiotic biosynthesis. J Antibiot (Tokyo) *67*, 31-41.

Yin, X., O'Hare, T., Gould, S.J., and Zabriskie, T.M. (2003). Identification and cloning of genes encoding viomycin biosynthesis from Streptomyces vinaceus and evidence for involvement of a rare oxygenase. Gene *312*, 215-224.

Yonus, H., Neumann, P., Zimmermann, S., May, J.J., Marahiel, M.A., and Stubbs, M.T. (2008). Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. J Biol Chem *283*, 32484-32491.

Yukioka, M., Tsukamoto, Y., Saito, Y., Tsuji, T., Otani, S., and Otani, S. (1965). Biosynthesis of Gramicidin S by a Cell-Free System of Bacillus Brevis. Biochem Biophys Res Commun *19*, 204-208.

Yuwen, L., Zhang, F.L., Chen, Q.H., Lin, S.J., Zhao, Y.L., and Li, Z.Y. (2013). The role of aromatic L-amino acid decarboxylase in bacillamide C biosynthesis by Bacillus atrophaeus C89. Sci Rep *3*, 1753.

Zaleta-Rivera, K., Xu, C., Yu, F., Butchko, R.A., Proctor, R.H., Hidalgo-Lara, M.E., Raza, A., Dussault, P.H., and Du, L. (2006). A bidomain nonribosomal peptide synthetase encoded by FUM14 catalyzes the formation of tricarballylic esters in the biosynthesis of fumonisins. Biochemistry *45*, 2561-2569.

Zane, H.K., Naka, H., Rosconi, F., Sandy, M., Haygood, M.G., and Butler, A. (2014). Biosynthesis of amphi-enterobactin siderophores by Vibrio harveyi BAA-1116: identification of a bifunctional nonribosomal peptide synthetase condensation domain. J Am Chem Soc *136*, 5615-5618.

Zerbe, K., Woithe, K., Li, D.B., Vitali, F., Bigler, L., and Robinson, J.A. (2004). An oxidative phenol coupling reaction catalyzed by oxyB, a cytochrome P450 from the vancomycin-producing microorganism. Angew Chem Int Ed Engl *43*, 6709-6713.

Zhang, J., Liu, N., Cacho, R.A., Gong, Z., Liu, Z., Qin, W., Tang, C., Tang, Y., and Zhou, J. (2016). Structural basis of nonribosomal peptide macrocyclization in fungi. Nat Chem Biol.

Zhang, Q., Yu, Y., Velasquez, J.E., and van der Donk, W.A. (2012). Evolution of lanthipeptide synthetases. Proc Natl Acad Sci U S A *109*, 18361-18366.

Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E., and Jensen, P.R. (2012). The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLoS One *7*, e34064.

Zocher, R., Nihira, T., Paul, E., Madry, N., Peeters, H., Kleinkauf, H., and Keller, U. (1986). Biosynthesis of cyclosporin A: partial purification and properties of a multifunctional enzyme from Tolypocladium inflatum. Biochemistry *25*, 550-553.