STRUCTURAL STUDIES ON
THE BIOSYNTHESIS AND BIOLOGICAL FUNCTIONS OF NATURAL PRODUCTS

BY

ZHI LI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biochemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Satish K. Nair, Chair, Director of Research
Professor Susan A. Martinis
Professor Emad Tajkhorshid
Professor Wilfred A. van der Donk

Abstract

Natural products are chemical compounds that are synthesized by living organisms and have biological and pharmacological functions. Many natural products or their derivatives have been successfully developed as pharmaceuticals and have thus made great contributions to human health. However, there is still an urgent need for novel and more potent pharmaceuticals, largely due to the rapidly increasing incidence of drug resistance. Natural products and their derivatives continue to be an important source for the development of new drugs. Critical steps in the investigation of natural products include the dissecting of the biosynthesis pathways of these compounds and elucidating the biological processes affected by these compounds. These studies would facilitate the production of potential drugs through biosynthesis or semi-biosynthesis, and provide valuable guidance in engineering these compounds towards improved efficacy. In this dissertation I present such studies on several natural products produced by bacteria and plants.

Teicoplanin is a glycopeptide antibiotic produced by *Actinoplanes teichomyceticus* and has been shown to be potent against many vancomycin-resistant strains. Such improved efficacy is mainly attributed to two special tailoring modifications that are unique for teicoplanin-type glycopeptide antibiotics and are catalyzed by a P450 monooxygenase, Orf6*, and an acyltransferase, Orf11*, respectively. The structural characterization of these two enzymes presented in Chapter 2 provides insight into the substrate specificity of both enzymes.

4-coumarate-CoA and its derivatives represent the key branching point in the phenylpropanoid pathway in plants. They are the precursors to many molecules that are important either pharmaceutically or for plant development. Studies on 4-coumarate:CoA ligase (4CL), the enzyme that synthesizes 4-coumarate-CoA, are thus crucial for plant engineering and drug discovery. In Chapter 3, I present thorough biochemical and structural studies on a tobacco 4CL enzyme, and a successful example of engineering this enzyme.

Bis-(3’-5’) cyclic dimeric guanosine monophosphate (c-di-GMP) is a second messenger molecule that plays a vital role in the global regulation in bacteria. By binding to different effectors, it triggers diverse responses that confer adaptability to various environmental conditions. In the last chapter of this dissertation I present the structural characterization of a novel c-di-GMP effector PelD from *Pseudomonas aeruginosa* that is critical for the formation of pellicles. The PelD structures shown here reveal an unprecedented binding mode for c-di-GMP.

Acknowledgements

I would like to thank all those who have helped me throughout my PhD study. This dissertation would not have been possible without their support and encouragement.

My deep gratitude goes first to my PhD advisor, Professor Satish Nair, who expertly guided me through my graduate research. I joined Professor Nair's laboratory at a difficult time during my graduate school journey and I can never thank him enough for accepting me as his student. During the past four and half years in Professor Nair's laboratory, I had the opportunity to work on many exciting projects under his guidance. His enthusiasm and dedication in research have demonstrated to me what a great scientist should be. His encouragement, inspiration, and patience have been invaluable to me.

I would like to thank the former members in Professor Nair's group, including Drs. Houjin Zhang, Yaozhong Zou, Brian Bae, Jui-Hui Chen, and Vinayak Agarwal for teaching me the basics in X-ray crystallography. I would especially like to thank Dr. Brian Bae, for his kindness and encouragement in guiding me through the difficult projects.

My sincere appreciation also extends to my current laboratory colleagues. For his helpful suggestions on structure determination and refinement practices, I would like to thank Dr. Tiit Lukk. For his assistance in ordering reagents and proofreading this dissertation, I would like to thank Jonathan Chekan. My special thanks go to Yue Hao, a very helpful labmate and also a close friend, who has unselfishly provided her assistance in many ways. Also I would like to thank the undergraduate students, including John Ha, Trusha Patel, Samuel Lee, and Caitlin Ondera, who have assisted me in many experiments.

I would like to thank my Committee members, Professors Susan Martinis, Emad Tajkhorshid, and Wilfred van der Donk for their intellectual input in my graduate research. Also thanks to my collaborators Professors Mary Schuler, Huimin Zhao, and

William Metcalf, and Drs. Annette Erb and Despina Bougioukou, for the opportunity to work with them and for their helpful discussions.

I would also like to thank Professor Brian Freeman in the Department of Cellular and Development Biology. I spent my first two years of graduate study in his laboratory and received good training on molecular biology, which has proved very helpful.

I am truly grateful to Jeffery Goldberg and Cara Day for their tireless help in the logistics throughout the process of my graduate study. Also thanks to Alejandra Stenger and Nicholas Kirchner for their guidance in my Teaching Assistantship, which partly provided me the fiscal support during graduate school. I would also like to thank Mrs. Marjorie Kade for the prestigious Charles F. Kade Fellowship, which provided me one year of generous fiscal support.

I am indebted to my family, who has been a constant source of support from overseas. I thank my parents, my elder sister, and my elder brother, for their continuous encouragement through our weekly phone calls. Their emotional and moral support has been and will always be invaluable to my career and personality development.

Lastly, but most importantly, I would like to thank my wife, Fang Wan, who has always been there for me with love and endless support. She has blessed me with a life of joy and has always been my guiding light. I owe the success of my graduate study to her.

Table of Contents

Chapter 1 Introduction 1

Chapter 2 Structural studies of a P450 monooxygenase and two acyltransferases involved in the biosynthesis of teicoplanin-type glycopeptide antibiotics 13

Chapter 3 Biochemical and structural studies on the catalysis mechanism and substrate specificity of a tobacco 4-coumarate:CoA ligase 47

Chapter 4 Structures of a novel cyclic diguanylate effector PelD involved in pellicle formation in *Pseudomonas aeruginosa* PAO1 77

References 105

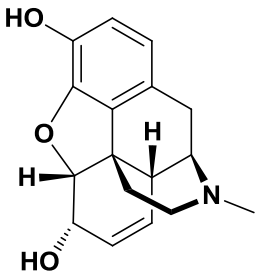
Chapter 1 Introduction[1]

Human society has a long history of using natural substances derived from animals, plants, microbes, and minerals to treat and prevent diseases. For example, the plant *Artemisia annua* (sweet wormwood) has been used in traditional Chinese medicine for more than two thousand years as the treatment for many diseases, most notably malaria (Miller and Su, 2011). The exploration of such natural medications has been greatly advanced since the purification of morphine from *Papaver somniferum* (opium poppy) at a German pharmacy in 1804, which demonstrated that the pharmacologically active compounds can be purified from their natural sources (Hamilton and Baskett, 2000) (Figure 1-1). The active anti-malaria ingredient in *Artemisia annua*, namely artemisinin, was purified from the leaves of this plant in 1972 and has since been widely used for the treatment of malaria (Miller and Su, 2011) (Figure 1-1). Such chemical compounds that are produced by living organisms and found to have pharmacological or biological activities have been collectively named *natural products*. By the end of twentieth century, the major source of new drugs has been natural products and other compounds derived from natural products, including semi-synthetic natural product derivatives and synthetic compounds based on natural products (Newman et al., 2003; Koehn and Carter, 2005; Li and Vederas, 2009; Newman and Cragg, 2012). Apart from the two drugs mentioned above, more examples include antibiotics (penicillin, tetracycline), cholesterol-lowering agent (lovastatin), and anti-cancer drugs (descodermolide) (Figure 1-1). Although research and development of natural product-based drugs are facing more economic and technical challenges in recent years, there is still tremendous ongoing interest in this field and natural products continue to be the source and inspiration for drug discovery (Koehn and Carter, 2005; Li and Vederas, 2009; Newman and Cragg, 2012).

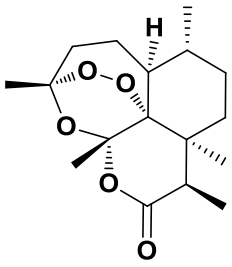
[1] **Parts of this chapter are adapted from the following published articles with the permission from the respective publisher:**

Li, Z., Rupasinghe, S.G., Schuler, M.A., and Nair, S.K. (2011). Crystal structure of a phenol-coupling P450 monooxygenase involved in teicoplanin biosynthesis. Proteins 79: 1728-1738.

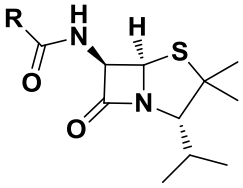
Li, Z., Chen, J.H., Hao, Y., and Nair, S.K. (2012). Structures of the PelD cyclic diguanylate effector involved in pellicle formation in *Pseudomonas aeruginosa* PAO1. J Biol Chem 287: 30191-30204.



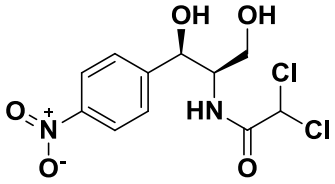
morphine
analgesic
Papaver somniferum
(opium poppy)



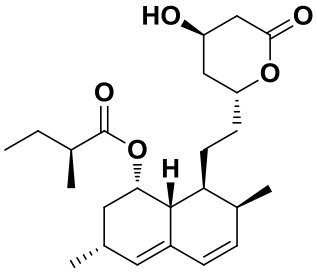
artemisinin
anti-malaria
Artemisia annua
(sweet wormwood)



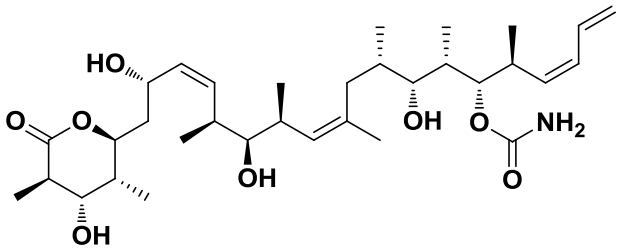
penicillin
anti-bacteria
Penicillium fungi



chloramphenicol
anti-bacteria
Streptomyces venezuelae



lovastatin
lipid-control
Pleurotus ostreatus
(oyster mushroom)



discodermolide
anti-tumor
Discodermia dissoluta
(Caribbean marine sponge)

Figure 1-1. Chemical structures of some examples of natural-product drugs. The name of the compound is shown below the structure. The pharmacological function is shown in blue text, below which is shown the producer organism.

Research in natural products has emphasized dissecting the biosynthesis pathways of these compounds. Advances in genome sequencing and manipulation in the last 15 years have greatly facilitated the identification and isolation of the biosynthesis gene cluster of interesting natural products from various organisms (Challis, 2008; Li et al., 2009; Doroghazi et al., 2011). Understanding of the biosynthesis mechanism of natural products, however, is still challenging and requires a cooperative effort in many disciplines including microbiology, chemistry, and biochemistry, as demonstrated by recent work on phosphonate antibiotics (Cicchillo et al., 2009; Lee et al., 2010).

Understanding the biological processes affected by natural products is also critical. Many natural products are secondary metabolites in the producer organisms. Investigating the interactions between these compounds and their producers, as well as their target organisms, provides important guidelines for a better understanding of their functions. For example, bis-(3'-5') cyclic dimeric guanosine monophosphate (c-di-GMP) (Figure 1-2) is a second messenger molecule produced by many bacteria. Recent studies revealed that c-di-GMP not only plays a vital role in the global regulation in the producer organisms (Ryder et al., 2007; Gjermansen et al., 2010; Li et al., 2012), but also is an important factor in triggering a series of responses in the target organisms (Burdette et al., 2011; Huang et al., 2012; Ouyang et al., 2012; Shang et al., 2012; Shu et al., 2012; Yin et al., 2012).

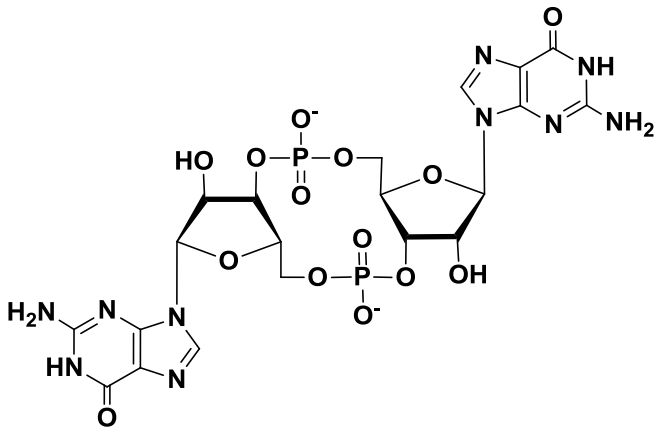


Figure 1-2. Chemical structure of c-di-GMP.

Notably, structural biology has played a crucial role in the investigation on natural products (Nair and van der Donk, 2011). Crystal structures of enzymes involved in the biosynthesis of natural products have provided a great deal of insight into the mechanisms of biosynthesis and mode of action of these compounds. Information on substrate specificity, catalysis mechanism, and compound-host interactions in the target organisms can all be investigated with the aid of crystal structures (Cicchillo et al., 2009; Lee et al., 2010; Burdette et al., 2011; Li et al., 2011; Huang et al., 2012; Li et al., 2012; Ouyang et al., 2012; Shang et al., 2012; Shu et al., 2012; Yin et al., 2012). Moreover, structural biology has also provided the framework for molecular docking and computer modeling in natural product based-drug development (Shoichet et al., 2002; Balamurugan et al., 2005; Dekker et al., 2005; Geromichalos, 2007; van Montfort and Workman, 2009).

In this dissertation, I present my research on the biosynthesis and function of three natural products, including teicoplanin (Li et al., 2011), 4-coumarate-CoA, and c-di-GMP (Li et al., 2012). For all these studies, I employed a combination of biochemical, biophysical, and structural biology approaches, with an emphasis on structural biology, in particular X-ray crystallography.

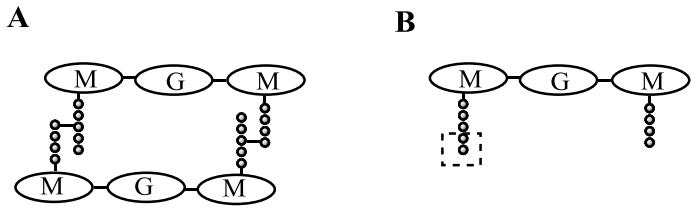


Figure 1-3. Mechanism of action of glycopeptide antibiotics. M: N-acetylmuramyl-pentapeptide. The pentapeptide is shown as spheres. G: N-acetylglucosamine. (A) Crosslinks between the peptidoglycan layers in normal bacterial cell wall. (B) Single layer of peptidoglycan, with the binding target of glycopeptide antibiotics highlighted in dashed frame.

Teicoplanin

Many glycopeptides have been shown to have antibiotic activities. Glycopeptide antibiotics exert their antimicrobial activity against gram-positive bacteria by targeting the N-acyl-D-Ala-D-Ala peptide termini of peptidoglycan precursors and inhibiting cell

wall synthesis (Kahne et al., 2005) (Figure 1-3). One representative of this class of antibacterials, vancomycin (compound **1** in Figure 1-4), has been approved for clinical use and is used as the "last-resort" treatment against multiple-drug-resistant bacteria (Moellering, 2006), including methicillin-resistant *Staphylococcus aureus* (MRSA) (Anstead and Owens, 2004; Anstead et al., 2007) and drug-resistant pneumococci (Cunha, 2006). However, a considerable number of vancomycin-resistant bacteria have emerged, including vancomycin-resistant enterococci (VRE) (Murray, 2000) and vancomycin-intermediate and -resistant *S. aureus* (VISA and VRSA, respectively) (Hiramatsu et al., 1997; Sieradzki et al., 1999). There are various mechanisms for vancomycin resistance, one of which is outlined in Figure 1-5.

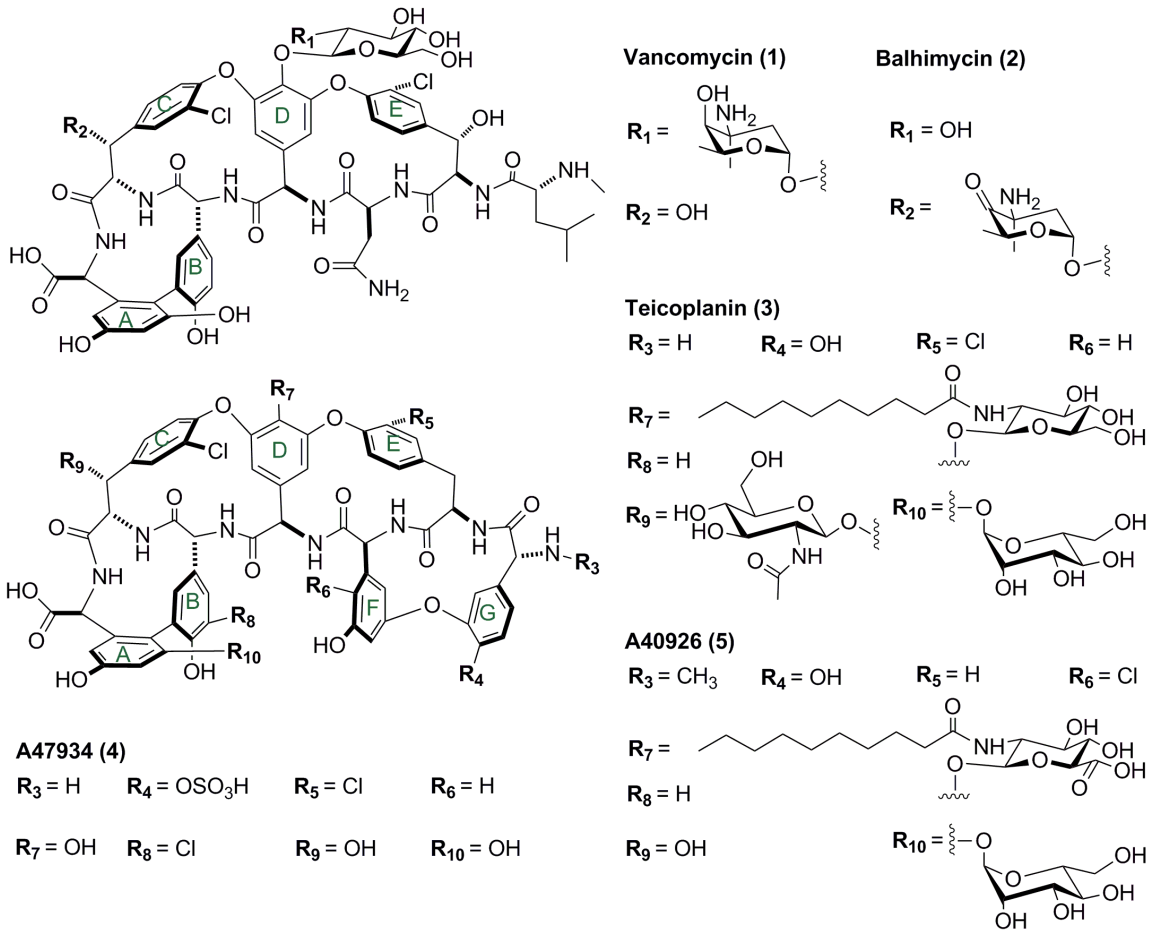


Figure 1-4. Chemical structures of type I (**1** vancomycin, **2** balhimycin), and type IV (**3** teicoplanin, **4** A47934, and **5** A40926) glycopeptide antibiotics.

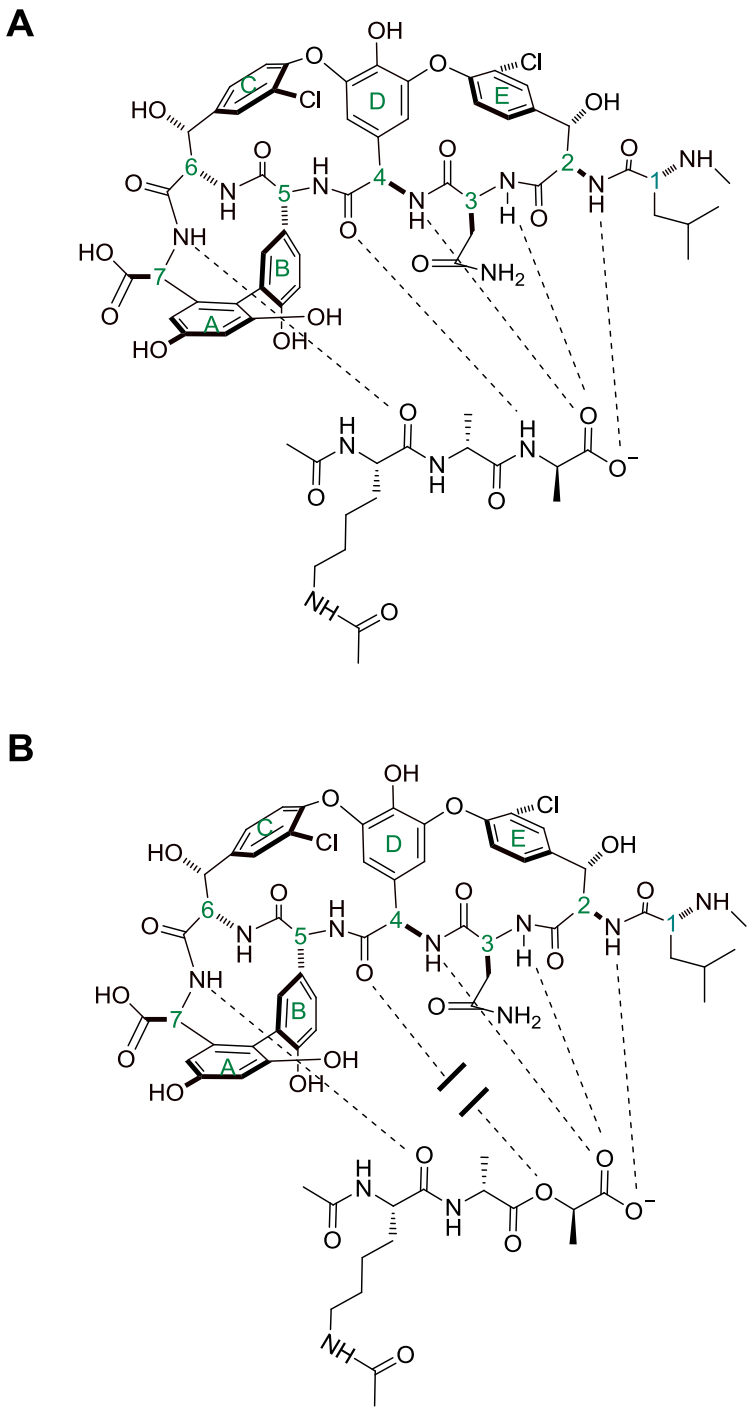


Figure 1-5. Illustration of one example of vancomycin resistance mechanism. (A) The hydrogen bonds between vancomycin and an analog to its normal binding target, di-acetyl-Lys-D-Ala-D-Ala. (B) The hydrogen bonds between vancomycin and an analog to its binding target in the resistant bacteria, di-acetyl-Lys-D-Ala-D-Lac. The loss of one hydrogen bond compared to the wild type strains results in over 1000-fold decrease in binding affinity (Kahne et al., 2005).

The *lipo*glycopeptide antibiotic teicoplanin (compound **3** in Figure 1-4) has proven efficacy against MRSA and VRSA (Van Bambeke, 2006), and has shown superior pharmacokinetic profile in comparison to vacomycin (Malabarba and Goldstein, 2005). The ability of teicoplanin to counter vancomycin resistant pathogens has prompted research efforts towards the design of lipoglycopeptide derivatives with increased antimicrobial properties relative to the parent compounds.

The general structure of glycopeptide and lipoglycopeptide antibiotics consists of a heptapeptide backbone containing aromatic amino acids whose side chains are oxidatively coupled via biaryl or biarylether bridges to generate an aglycone core (Donadio et al., 2005) (Figure 1-4). These cross-links force the peptide backbone into a rigid conformation that is crucial to the primary antibiotic activity of glycopeptide antibiotics (Hadatsch et al., 2007). Previous studies have shown that linear or various mono-cyclic or bi-cyclic intermediates of the corresponding antibiotic have minimal antimicrobial activity (Pelzer et al., 1997; Süssmuth et al., 1999; Bischoff et al., 2001b; Bischoff et al., 2001a; Stegmann et al., 2006). The cross-linked aglycone undergoes further tailoring modifications including N-methylation, halogenation, glycosylation, and/or acylation (Hubbard and Walsh, 2003; Sussmuth and Wohlleben, 2004).

The most distinguishing features of teicoplanin-type antibiotics compared to vancomycin-type antibiotics include an additional cross-link between residues 1 and 3 and an N-acyl chain on the glucosamine group linked to residue 4 (Figure 1-4). These two modifications have been found to be greatly responsible for the superior pharmacokinetic profile of teicoplanin (Beauregard et al., 1995; Malabarba et al., 1997; Cooper and Williams, 1999; Dong et al., 2002). In Chapter 2 of this dissertation, I present structural studies on the enzymes that catalyze these two modifications. These structures provide insight into the substrate specificity determinants of these enzymes and suggest possible catalysis mechanisms.

4-coumarate-CoA

Many important natural products are made by plants as primary or secondary metabolites. The phenylpropanoid pathway represents the central biosynthetic nexus for the production of an array of plant metabolites including monolignols, phytohormones, flavonoids, and phenylpropenes (Hahlbrock and Scheel, 1989). The pathway directs the flow of carbon from primary metabolism to diverse secondary metabolism branch pathways. These phenylpropanoids serve a range of functions *in planta*, including providing mechanical support, UV protection, defense against pathogens, and mediating interactions with pollinators (Hahlbrock and Scheel, 1989; Lee and Douglas, 1996).

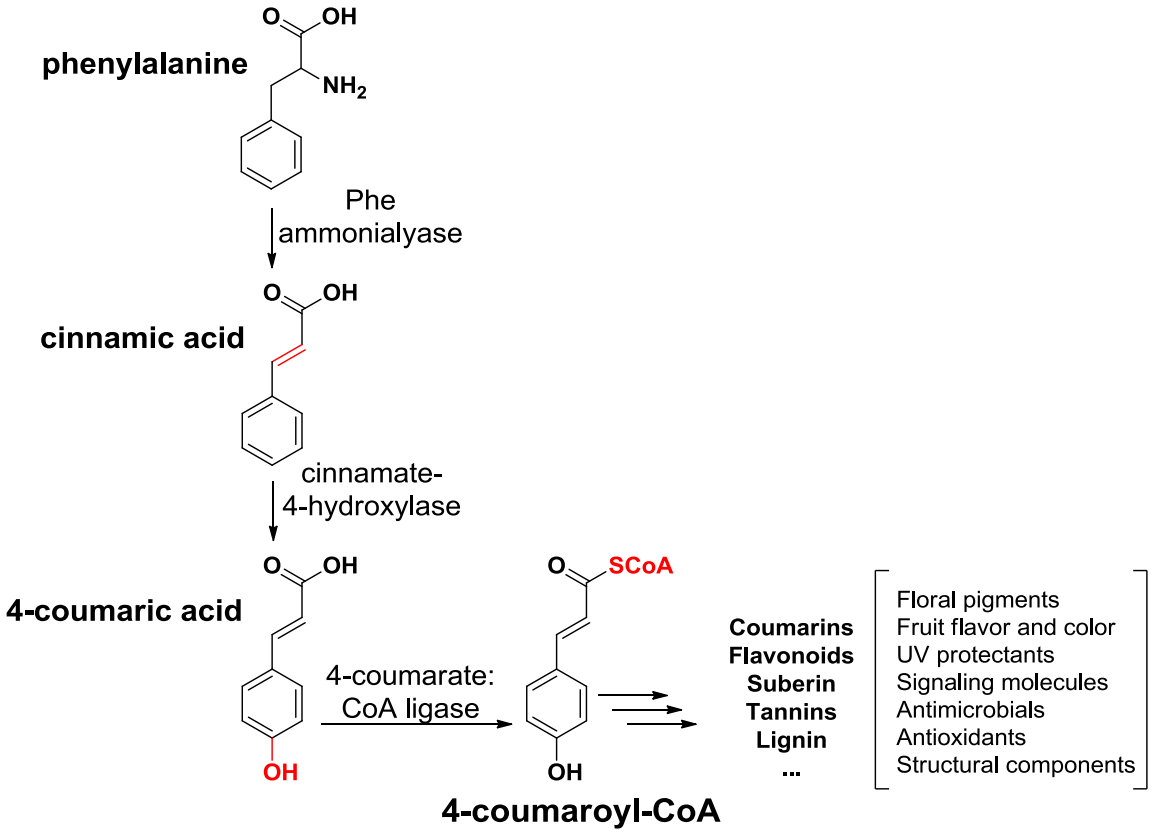


Figure 1-6. Early steps in the phenylpropanoid pathway in plants.

Several of these metabolites are also of pharmaceutical interest, including the stilbenoid resveratrol (lifespan extension) (Richard et al., 2011), the phenylpropene shikimol (a precursor in the synthesis of the entheogenic drug MDMA) (Gimeno et al., 2005), and the coumarin umbelliferone (antioxidant) (Kanimozhi et al., 2011). Lignin, the polymeric product of monolignols, has been the focus of research that targets agro-industrial uses of plant biomass (Boudet et al., 2003; Ragauskas et al., 2006).

The core reactions of the general phenylpropanoid pathway start with the deamination of phenylalanine (by phenylalanine ammonialyase) to yield *trans*-cinnamic acid (Figure 1-6), which is then *para*-hydroxylated (by cinnamate-4-hydroxylase) to form 4-coumaric acid. A second hydroxylation at the C3 position of the phenolic ring (by coumarate 3-hydrolase) affords caffeic acid. Caffeic acid can undergo *O*-methylation at C3 to yield ferulic acid or additional hydroxylation and *O*-methylation at C5 to yield sinapinic acid (Figure 1-7). A key branching point in the phenylpropanoid pathway is the formation of coenzyme A (CoA) thioesters of these hydroxycinnamic acid derivatives catalyzed by the enzyme 4-coumarate:CoA ligase (4CL, EC 6.2.1.12), which leads to the productions of a variety of important molecules (Figure 1-6).

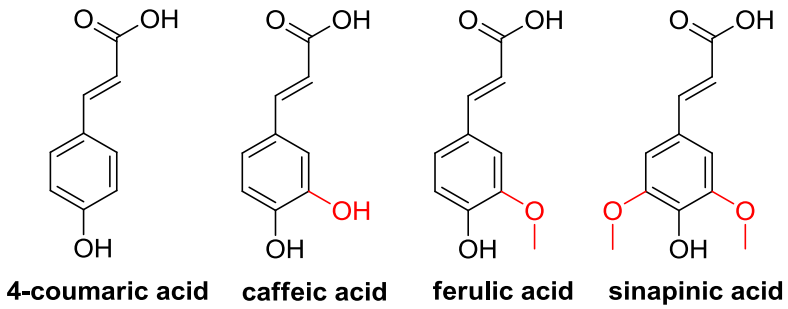


Figure 1-7. Examples of hydroxycinnamic acids.

4CL proteins play vital roles in regulating carbon flow in plant biosynthetic pathways, as their product hydroxycinnamate derived-CoA thioesters serve as precursors for various branching pathways of phenylpropanoid synthesis. For example, 4-coumaroyl-CoA is a substrate for the first committed step in flavonoid biosynthesis (Cukovic et al., 2001), and lignin monomers, including 4-coumaryl alcohol, corniferyl alcohol, and sinapyl alcohol (Figure 1-8), are derived from the hydroxycinnamoyl-CoA thioesters (Boerjan et al., 2003). As the level and composition of lignin in plants largely determine the efficiency of biomass utilization (Boudet et al., 2003), 4-coumarate:CoA ligases have been studied extensively in efforts to produce engineered plants with improved biomass utility. Although successful studies have been reported recently (Kajita et al., 1997; Lee et al., 1997; Hu et al., 1999), 4CL engineering remains challenging due to limited knowledge of its catalytic mechanism. In Chapter 3, I present a thorough biochemical and structural

studies on an 4CL isoform from tobacco. The results provide insights into the substrate specificity of 4CL enzymes and provide guidelines in 4CL engineering.

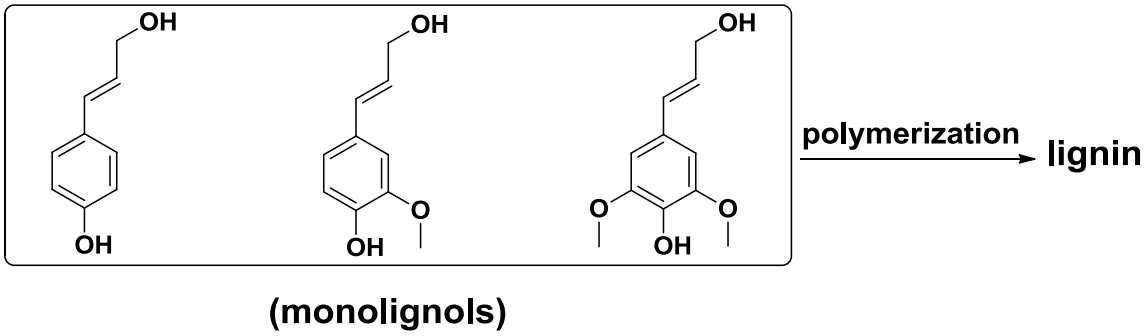


Figure 1-8. Chemical structures of lignin monomers (monolignols). From left to right are 4-coumaryl alcohol, corniferyl alcohol, and sinapyl alcohol, respectively. Lignin is the polymerization product of these three monomers.

c-di-GMP

Bis-(3'-5') cyclic dimeric guanosine monophosphate (c-di-GMP) (Figure 1-2) is a central regulator, which functions as an intracellular second messenger. In bacteria, this molecule confers adaptability to various environmental conditions, by coordinating the transition between the motile planktonic state to a sessile state associated with biofilm production (Jenal and Malone, 2006; Hengge, 2009; Schirmer and Jenal, 2009). Specifically, c-di-GMP stimulates the production of adhesins and exopolysaccharide matrix components and leads to biofilm formation to protect bacteria from host-defense, starvation conditions, and antibiotics (Ryder et al., 2007; Gjermansen et al., 2010). Additional roles for c-di-GMP include control of cell cycle progression (Duerig et al., 2009), antibiotic biosynthesis (Fineran et al., 2007), and expression of virulence genes (Parsek and Singh, 2003; Dow et al., 2006; Kulesekara et al., 2006; Tamayo et al., 2007). The bacterial signaling nodes that respond to the c-di-GMP message present targets for therapeutic intervention against pathogens.

Similar to other second messenger pathways, the c-di-GMP control module can be generally divided into four components that govern signal generation, degradation, recognition, and targeting, respectively (Hengge, 2009). The level of the signal molecule

is dynamically regulated by the opposing activities of diguanylate cyclases (DGCs) and phosphodiesterases (PDEs), which synthesize and degrade c-di-GMP, respectively. DGC activity is attributed to proteins that contain a characteristic GGDEF domain (named for the single-letter amino acid nomenclature of essential active site residues), while PDE activity is associated with either enzymes that contain either an EAL or a HD-GYP domain (Ausmees et al., 2001; Paul et al., 2004; Schmidt et al., 2005; Ryan et al., 2006). Interestingly, many DGCs have been shown to be subject to allosteric product inhibition, which is often caused by c-di-GMP binding to the so-called I-site in the GGDEF domain (Chan et al., 2004; Christen et al., 2006; De et al., 2009). The I-site is readily identified by the RxxD (Arg-X-X-Glu; where X is any amino acid) motif, which is connected N-terminally to GGDEF motif through a five-residue linker. These two motifs are antipodal to each other in the three-dimensional structure as shown by the structures of PleD from *Caulobacter vibriodes* (Chan et al., 2004; Wassmann et al., 2007) and WspR from *Pseudomonas aeruginosa* (De et al., 2008; De et al., 2009).

Direct recognition of the c-di-GMP signal occurs through an effector component that is often linked to a signal input component that regulates cellular functions at transcriptional, translational, or post-translational levels (Weber et al., 2006; Merighi et al., 2007; Monds et al., 2007; Duerig et al., 2009). Strikingly, c-di-GMP effectors are highly diverse and are responsible in the diversity of the cellular functions and processes controlled by c-di-GMP in bacteria (Hengge, 2009; Schirmer and Jenal, 2009). Several classes of c-di-GMP effectors have been identified through *in vivo* and *in vitro* studies, and some have been structurally characterized. A novel class of c-di-GMP effector consists of molecules that can bind c-di-GMP through an RxxD motif that resembles the I-site of GGDEF domain of DGCs, but do not show catalytic activity as their active sites lack the requisite GGDEF motif. Examples include PelD from *P. aeruginosa* (Lee et al., 2007), CdgG from *V. cholera* (Beyhan et al., 2008), and PopA from *C. vibriodes* (Duerig et al., 2009). In contrast to the effectors described in the previous paragraph, biochemical data for I-site-containing c-di-GMP effectors is sparse, and there are no crystal structures available for any members of this receptor class. In the last chapter of this dissertation, I present biochemical and structural studies on such a c-di-GMP effector. The results reveal a

novel binding mode of c-di-GMP by its effector and implied possible mechanisms for signal transduction in bacteria.

Chapter 2 Structural studies of a P450 monooxygenase and two acyltransferases involved in the biosynthesis of teicoplanin-type glycopeptide antibiotics[1]

Abstract

The lipoglycopeptide antibiotic teicoplanin has proven efficacy against gram-positive pathogens. Teicoplanin is distinguished from the vancomycin-type glycopeptide antibiotics, by the presence of an additional cross-link between the aromatic amino acids 1 and 3 that is catalyzed by the cytochrome P450 monooxygenase Orf6* (CYP165D3), and an acyl chain on a sugar moiety added by an acyltransferase Orf11* (tAtf). As a goal towards understanding teicoplanin biosynthesis, I present the biochemical and structural characterization of recombinant Orf6* from the teicoplanin producer *Actinoplanes teichomyceticus*. I also present the structures of tAtf from the same strain and its homolog, aAtf from *nonomuraea*, the producer of a teicoplanin-type glycopeptide A40926, in complex with octanoyl-CoA. The structure of Orf6* reveals the core fold common to other P450 monooxygenases but also shows novel features in the disposition of secondary structure elements near the active site cavity necessary to accommodate its complex heptapeptide substrate. The complex structures tAtf and aAtf display a circular binding tunnel for the acyl chain and explain the acyl chain specificity. These structures provide further insights into the mechanism of the biosynthesis of teicoplanin-type glycopeptide antibiotics.

Introduction

Glycopeptide antibiotics exert their antimicrobial activity against gram-positive bacteria by targeting the N-acyl-D-Ala-D-Ala peptide termini of peptidoglycan precursors and inhibiting cell wall synthesis (Kahne et al., 2005). One representative of this class of antibacterials, vancomycin (compound **1** in Figure 1-4), has been approved for clinical use and is used as the "last-resort" treatment against multiple-drug-resistant bacteria

[1] **Parts of this chapter are adapted from the following published article with the permission from the publisher:**

Li, Z., Rupasinghe, S.G., Schuler, M.A., and Nair, S.K. (2011). Crystal structure of a phenol-coupling P450 monooxygenase involved in teicoplanin biosynthesis. Proteins 79: 1728-1738.

(Moellering, 2006), including methicillin-resistant *Staphylococcus aureus* (MRSA) (Anstead and Owens, 2004; Anstead et al., 2007) and drug-resistant pneumococci (Cunha, 2006). However, a considerable number of vancomycin-resistant bacteria have emerged, including vancomycin-resistant enterococci (VRE) (Murray, 2000) and vancomycin-intermediate and -resistant *S. aureus* (VISA and VRSA, respectively) (Hiramatsu et al., 1997; Sieradzki et al., 1999). The *lipo*glycopeptide antibiotic teicoplanin (compound **3** in Figure 1-4) has proven efficacy against MRSA and VRSA (Van Bambeke, 2006), and has shown superior pharmacokinetic profile in comparison to vancomycin (Malabarba and Goldstein, 2005). The ability of teicoplanin to counter vancomycin resistant pathogens has prompted research efforts towards the design of lipoglycopeptide derivatives with increased antimicrobial properties relative to the parent compounds.

The general structure of glycopeptide and lipoglycopeptide antibiotics consists of a heptapeptide backbone containing aromatic amino acids whose side chains are oxidatively coupled via biaryl or biarylether bridges to generate an aglycone core (Donadio et al., 2005) (Figure 1-4). These cross-links force the peptide backbone into a rigid conformation that is crucial to the primary antibiotic activity of glycopeptide antibiotics (Hadatsch et al., 2007). Previous studies have shown that linear or various mono-cyclic or bi-cyclic intermediates of the corresponding antibiotic have minimal antimicrobial activity (Pelzer et al., 1997; Süssmuth et al., 1999; Bischoff et al., 2001b; Bischoff et al., 2001a; Stegmann et al., 2006). The cross-linked aglycone undergoes further tailoring modifications including N-methylation, halogenation, glycosylation, and/or acylation (Hubbard and Walsh, 2003; Sussmuth and Wohlleben, 2004).

Oxidative cross-links

The elucidation of biosynthesis gene clusters of several glycopeptides has established the general route of the complex structure assembly (Pelzer et al., 1999; Pootoolal et al., 2002; Sosio et al., 2003; Li et al., 2004). The building blocks, which are mostly non-proteinogenic amino acids, are assembled by nonribosomal peptide synthetase (NRPS) (Kahne et al., 2005). The electron-rich aromatic side chains facilitate oxidative cross-

linking reactions, which are carried out by a series of cytochrome P450 monooxygenases (P450s). For example, in the vancomycin-type (type I) glycopeptide antibiotic balhimycin (compound **2** in Figure 1-4) system, three heme-dependent P450s (OxyA, OxyB, OxyC) are responsible for the three cross-links in the heptapeptide substrate (Bischoff et al., 2001b; Bischoff et al., 2001a). Genetic and chemical analyses have defined the regiospecificity and timing of the oxidative phenol-coupling reactions in the balhimycin cluster (Süssmuth et al., 1999; Bischoff et al., 2001b; Bischoff et al., 2001a). The first cross-linking occurs between residues 4 and 6 (C-O-D ring) and is catalyzed by OxyB (Süssmuth et al., 1999; Bischoff et al., 2001b; Bischoff et al., 2001a), the second is between residues 2 and 4 (D-O-E ring) and is catalyzed by OxyA, and the third is between residues 5 and 7 (AB ring) and is catalyzed by OxyC (Süssmuth et al., 1999; Bischoff et al., 2001b; Bischoff et al., 2001a). It has been suggested that the cross-linking reactions occur when the heptapeptide is still attached to the peptidyl carrier domain (PCD) as a thioester through its carboxy-terminus (Puk et al., 2002). This is supported by the *in vitro* activity studies of OxyB (Woithe et al., 2007), using peptide substrates that are conjugated to a PCD domain.

A distinguishing feature of teicoplanin-type (type IV) glycopeptide antibiotics is an additional cross-link between residues 1 and 3 (F-O-G ring), which, in contrast to vancomycin-like glycopeptides, consists of non-proteinogenic aromatic amino acids (Malabarba et al., 1986) (Figure 1-4). Consequently, there are four P450s in biosynthetic systems of teicoplanin and related lipoglycopeptides (Pootoolal et al., 2002; Sosio et al., 2003; Li et al., 2004; Hadatsch et al., 2007), namely *orf5**, *6**, *7**, and *9** in the teicoplanin cluster (Li et al., 2004). While the different oxidative enzymes from the same biosynthetic cluster show only modest sequence identity (i.e. less than 40%), orthologous enzymes from different species are highly similar. For example, Orf5*, 7*, and 9* from the teicoplanin gene cluster show strong sequence correspondence to OxyA, OxyB, and OxyC from the vancomycin cluster (similarity: 90%, 80%, 84%; identity: 78%, 72%, and 74%). This conservation has been used to infer function for the teicoplanin oxidative enzymes in catalyzing the cross-link between residues 2-4, 4-6, and 5-7, respectively (Li et al., 2004). The remaining P450, Orf6* (CYP165D3), is likely to be responsible for the

additional cross-link between residue 1 and 3 that is unique to the teicoplanin aglycone (Li et al., 2004). Gene-knockout studies on the teicoplanin-type glycopeptide antibiotic A47934 (compound **4** in Figure 1-4) confirmed four P450 proteins that are responsible for the four cross-links in this type of compound and suggested a likely order of the four phenol-coupling reactions. Intriguingly, the additional 1-3 cross-link in A47934, catalyzed by the Orf6* orthologue StaG, is thought to be the second phenol-coupling reaction (Hadatsch et al., 2007).

N-Acylation

Another distinguishing feature of teicoplanin is the presence of an N-acyl chain on the glucosamine group linked to residue 4 (Figure 1-4), hence the name *lipo*glycopeptide. The superior pharmacokinetic profile of teicoplanin has been partly attributed to this N-acyl group (Beauregard et al., 1995; Malabarba et al., 1997; Cooper and Williams, 1999; Dong et al., 2002). For example, it was suggested to be responsible for the antibacterial activity of teicoplanin against the VRE VanB strains (Dong et al., 2002). It was proposed that this hydrophobic acyl chain anchors teicoplanin onto the bacterial cell membrane (Beauregard et al., 1995; Cooper and Williams, 1999) and thus increases the local concentration of teicoplanin near the peptidoglycan layer of bacterial cell wall, leading to a higher efficiency in inhibiting cell wall synthesis. In contrast, the vancomycin-type antibiotics, which lack such acyl chains, are more freely distributed (Beauregard et al., 1995; Cooper and Williams, 1999).

Studies of teicoplanin biosynthesis gene cluster (Li et al., 2004; Sosio et al., 2004) in *A. teichomyceticus* lead to the identification of an acyltransferase Orf11* that is responsible for N-acylation in teicoplanin tailoring. The protein shared no significant sequence homology to proteins of known function and its role as an acyltransferase was proposed based on the fact that *orf11** gene was adjacent to a glycosyltransferase gene *orf10** in the teicoplanin biosynthesis gene cluster (Li et al., 2004). In vitro analysis confirmed its activity, with medium-to-long chain acyl-CoA molecules as the acyl donor substrate, and 2-aminoglucosamine teicoplanin pseudoaglycone as the acyl acceptor substrate (Li et al., 2004; Kruger et al., 2005). An Orf11* homolog involved the biosynthesis of another type

IV lipoglycopeptide antibiotic A40926 by *nonomuraea* species was also identified (Sosio et al., 2003) and was shown to have similar activity (Kruger et al., 2005). For simplicity, these two acyltransferase (Atf) homologs will be referred to as *t*Atf and *a*Atf, respectively.

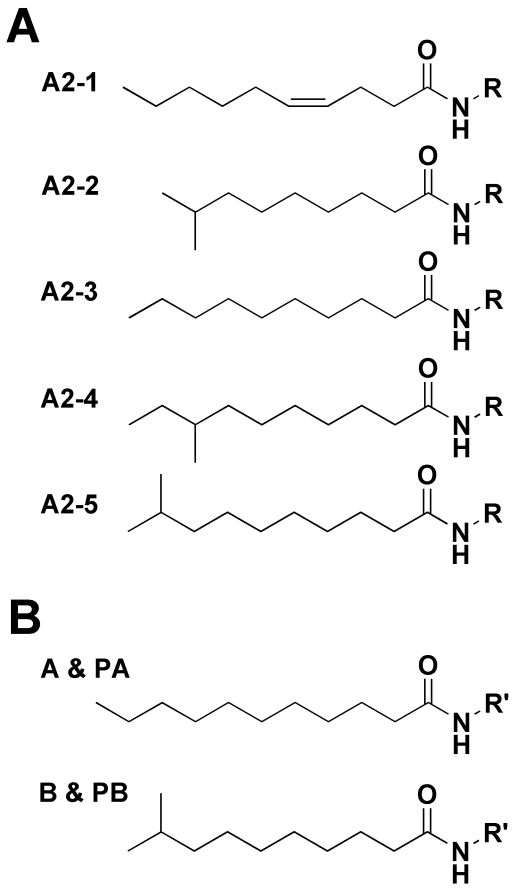


Figure 2-1. Acyl chains of the major species of teicoplanin (A) and A40926 (B). For the latter, each acyl chain has been identified in two different species.

Notably, both teicoplanin and A40926 are naturally produced as a mixture of compounds that differ almost only in the identities of the acyl chains. Teicoplanin has mainly five species (Borghi et al., 1984) and four species have been isolated for A40926 (Goldstein et al., 1987). These species have acyl chains that range from 10 carbons to 12 carbons and have different saturation or branching mode (Figure 2-1). In agreement with such acyl chain diversity, both tAtf and aAtf have been shown to possess significant activity against a series of acyl-CoA substrates of the chain lengths ranging from 6 to 14 carbons, with the best substrate of tAtf being decanoyl (C10)-CoA and the best for aAtf being lauroyl (C12)-CoA (Kruger et al., 2005). Understanding the molecular rationale for such wide

substrate tolerance would be crucial for developing novel lipoglycopeptide antibiotics based on the framework of these Atfs and their substrates.

Here we report the recombinant expression and characterization of Orf6* from the teicoplanin producer *Actinoplanes teichomyceticus*, and also tAtf and aAtf from the producers of teicoplanin and A40926, respectively. We present high resolution crystal structures of these proteins. The Orf6* crystal structure is the first structure for a P450 responsible for the additional 1-3 cross-link in the teicoplanin-type glycopeptide antibiotics. A comparison with the available structures of the phenol-coupling monooxygenases (OxyB and OxyC) in vancomycin system reveals a possible mechanism for Orf6* regiospecificity. The structures of the Atfs shed light on the acyl-chain specificity of these enzymes.

Experimental procedures

Cloning, Protein Expression, and Purification of Orf6*, aAtf, and tAtf

Actinoplanes teichomyceticus and *nonomuraea* were purchased from the American Tissue Culture Collection and used directly as a template for polymerase chain reaction amplifications without purification of the genomic DNA. The *orf6*, aAtf, tAtf* genes were cloned into pET28b (Novagen) for overexpression of each protein with an amino-terminal hexahistidine affinity tag.

For Orf6* protein production, the expression construct was transformed into the *Escherichia coli* Rosetta (DE3) strain. A pre-culture was grown overnight at 37 ℃ in Terrific Broth (TB) with kanamycin (50 μg/ml) and chloramphenicol (34 μg/ml) and used to inoculate (4%, v/v) 400 ml of TB medium with the antibiotics and the resultant culture was grown at 37 °C in shaker flasks. When the optical density at 600 nm (OD_{600}) reached 0.5, δ-aminolevulinic acid was added to 0.15 mM. Protein production was induced at OD_{600} of 1.0 with the addition of 0.5 mM isopropyl-1-thio-β-D-galactopyranoside and growth was continued at 18 ° C. A second aliquot of δ-aminolevulinic acid was added (0.15 mM) 20 hours after induction, followed by another 24 hours of growth before cell harvest.

For protein production of aAtf and tAtf, the expression constructs were transformed into the *Escherichia coli* Rosetta (DE3) strain. A pre-culture was grown overnight at 37 ℃ in Luria Broth (LB) with kanamycin (50 μg/ml) and chloramphenicol (34 μg/ml) and used to inoculate (0.3%, v/v) 1 L of TB medium with the antibiotics and the resultant culture was grown at 37 °C in shaker flasks. When the optical density at 600 nm (OD_{600}) reached 0.5, protein production was induced with the addition of isopropyl-1-thio-β-D-galactopyranoside to 0.5 mM and growth was continued at 18 °C for 20 hours before cell harvest.

The culture was harvested by centrifugation at 3,500 rpm for 25 minutes and the resultant pellet was resuspended in 20 mM Tris, pH 8.0, 1 M NaCl, 30 mM imidazole, and 10% (v/v) glycerol. Harvested cells were disrupted by four passes through an Avestin C5 Emulsiflex French Press and insoluble aggregates and cellular debris were removed by centrifugation at 15, 000 rpm for 1 hour.

Recombinant protein from the above-clarified supernatant was applied to a 10 ml Talon resin (Clontech) column that was charged with nickel sulfate and pre-equilibrated with 20 mM Tris, pH 8.0, 1 M NaCl, 30 mM imidazole. After elution from the nickel affinity resin with 200 mM imidazole, the polyhistidine tag was removed with thrombin (1 unit/mg, MP Biomedicals). The protein was dialyzed into 20 mM Tris, pH 8.0, 50 mM NaCl and applied to a 5 ml HiTrap Q column (GE Healthcare), and then eluted with a linear gradient of NaCl from 50 mM to 1 M over 60 ml. Orf6* was further purified by size-exclusion chromatography (Superdex 75 16/60, GE Healthcare) prior to crystallization.

Selenomethionine-incorporated Orf6* was produced by the method of van Duyne et al. (Van Duyne et al., 1993) and purified in the same manner as described above. Protein concentration was estimated by Bradford assay.

UV-visible spectroscopy

The UV-visible absorption spectra for Orf6* were measured with a Varian Cary 3 double-beam spectrophotometer. After thermal equilibration at 25 °C the baseline was zeroed between 200 and 900 nm and the UV-visible spectra of the substrate-free Orf6* (0.1 mg/ml in 20 mM HEPES, pH 7.5, 100 mM KCl) were recorded. The protein was reduced by addition of 2 mg of solid sodium dithionite and spectra were measured under the same condition. The reduced protein was divided into two tandem cuvettes (sample and reference) and CO was bubbled through the sample cuvette for 30 seconds, then the difference spectrum was measured. The protein-imidazole complex was generated by adding 100 mM imidazole to Orf6* in the same buffer and spectra were measured under the same condition.

Orf6* Crystallization

Orf6* crystals were grown by the hanging drop vapor diffusion method. 1 μl (8 mg/ml in a solution containing 20 mM HEPES, pH 7.5, 100 mM KCl) protein was mixed with 1 μl precipitant solution containing 0.2 M sodium acetate, 0.1 M Tris, pH 8.5, and 20% (w/v) polyethylene glycol 4, 000. The mixture drop was equilibrated over a well containing the same precipitant solution at 8 ℃, and crystals reached their maximum size after 3 days. Selenomethionine-incorporated Orf6* was crystallized under the same condition. The crystals were soaked in cryoprotectant solution containing the precipitant solution supplemented with 30% (v/v) glycerol anhydrous prior to vitrification in liquid nitrogen.

Orf6* Phasing and Structure Determination

Crystals of selenomethionine-incorporated Orf6* (SeMet Orf6*) consistently diffracted to higher resolutions, and was subsequently used for data collection and structure refinement. Diffraction data were collected to a limiting resolution of 2.2 Å at an insertion device line (LS-CAT-Sector 21 ID-D, Advanced Photon Source, Argonne, IL), and integrated and scaled using the HKL2000 (Otwinowski et al., 2003). SeMet Orf6* crystals occupy space group $P3_1$ with unit cell parameters a = 74.3 Å, b = 74.3 Å, c = 75.5 Å and contain one molecule per asymmetric unit. A four-fold redundant data set

was collected to a limiting resolution of 2.2 Å (overall R_{merge} 0.066, I/ (I) = 4.8 in the highest resolution shell).

Table 2-1. Data collection, phasing and refinement statistics of Orf6*.

	SeMet Orf6*
Data collection	
Space group	$P3_1$
Cell dimensions	
a, *b*, *c* (Å)	74.3, 74.3, 75.5
Resolution (Å)	50-2.2 (2.28-2.2)[a]
R_{sym} (%)	6.6 (29.7)
I / (I)	28.8 (4.8)
Completeness (%)	96.2 (74.0)
Redundancy	3.5 (3.1)
Refinement	
Resolution (Å)	25.0-2.2
Number of reflections	21,363
R_{work} / R_{free}[b]	20.1%/25.7%
Number of atoms	
Protein	2898
Solvent	184
Heme	43
Average B value	
Protein	30.8
Solvent	37.2
NADP	25.2
R.m.s deviations	
Bond angles (Å)	1.29
Bond lengths (°)	0.011

[a]Highest resolution shell is shown in parenthesis.
[b]R-factor = $(|F_{obs}|-k|F_{calc}|)/ |F_{obs}|$and R-free is the R value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement.

Selenium sites were identified using HySS (Adams et al., 2002; Grosse-Kunstleve and Adams, 2003) and heavy atom parameters were further refined using PHENIX (Adams et al., 2002; Grosse-Kunstleve and Adams, 2003) to yield an initial figure of merit of 0.390 to 2.2 Å resolution. Solvent flattening further improved the quality of the initial map permitting 80% of the main chain and 35% of the side chain residues to be automatically built by PHENIX. This initial model was further improved by automated building using ARP/wARP (Perrakis et al., 1999) to yield a model with nearly all of the main chain and 70% of the side chains. The remainder of the model was fitted using XtalView (McRee,

1999) and further improved by rounds of refinement with REFMAC5 (Murshudov et al., 1997) and manual building. Multiple rounds of manual model building were interspersed with refinement using REFMAC5 to complete structure refinement. Cross-validation used 5% of the data in the calculation of the free R factor (Kleywegt and Brunger, 1996). The stereochemistry of the model was routinely monitored throughout the course of refinement using PROCHECK (Laskowski et al., 1996). Relevant data collection and refinement statistics are summarized on Table 2-1.

Results

UV-visible Spectroscopy analysis of Orf6*

Amino-terminally hexahistidine-tagged Orf6* was heterologously expressed in recombinant *E. coli* Rosetta (DE3) and purified from the soluble fraction by Ni-NTA affinity chromatography, and was further purified by anion-exchange and size-exclusion chromatography after thrombolytic removal of the affinity tag. As shown in Figure 2-2A, Orf6* gave characteristic UV-visible absorption spectra of P450 hemeprotein (Hill et al., 1970). Specifically, the UV-visible absorption spectrum of substrate-free Orf6* has a Soret band at 419 nm, and β and α peaks at 536 and 570 nm, respectively, as is typical of oxidized P450s in the low-spin ferric state. Reduction of substrate-free Orf6* with sodium dithionite shifts the Soret band to 424 nm, and β and α peaks at 530 and 565 nm, respectively. The CO difference spectrum of reduced protein shows a prominent peak at 450 nm (Figure 2-2B). These spectra show that the purified Orf6* is in the oxidized form, and can be reduced with sodium dithionite. Addition of imidazole to substrate-free Orf6* leads to a typical red shift in the Soret band to 429 nm, and β and α peaks to 543 and 565 nm indicating that imidazole binds to the heme iron (III) atom as an axial ligand (Zerbe et al., 2002; Pylypenko et al., 2003), a typical feature of P450 hemeproteins (Jefcoate, 1978).

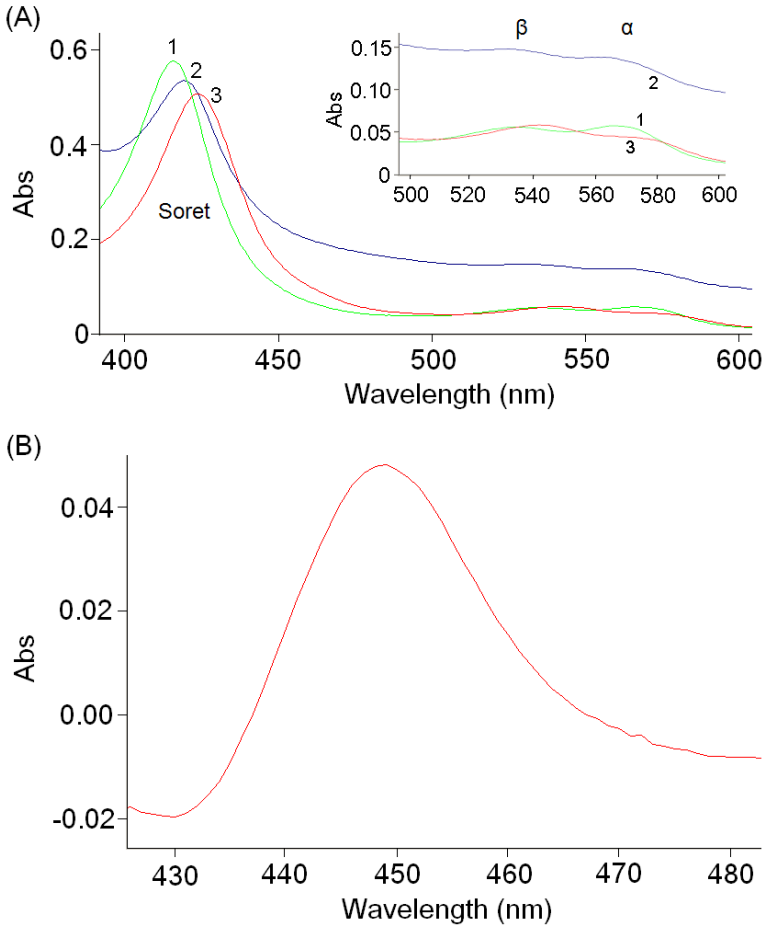


Figure 2-2. UV-visible absorption spectra of Orf6*. (A) Curve 1 (green) shows the spectrum of substrate-free Orf6*; Curve 2 (blue), after reduction with sodium dithionite; and Curve 3 (red), after addition of imidazole. (B) The CO-difference spectrum of reduced Orf6* showing the characteristic absorbance at 450 nm.

Overall Structure of Orf6*

The structure of Orf6* was determined to 2.2 Å using single-wavelength anomalous scattering from crystals grown with selenomethionine-incorporated protein (relevant data collection and refinement statistics are given in Table 2-1). As shown in Figure 2-3A and B, Orf6* exhibits the typical overall structure and protein topography of most P450 heme-proteins (Graham and Peterson, 1999). The protein has a triangular prism shape with predominantly α-helical secondary structure (see Figure 2-3B for a structure-based sequence alignment and secondary structure assignment) and the heme prosthetic group is embedded between the proximal L and distal I helices. The structure comprises 15 α-helices and 6 β-stands, and it can be roughly divided into α-rich and β-rich regions.

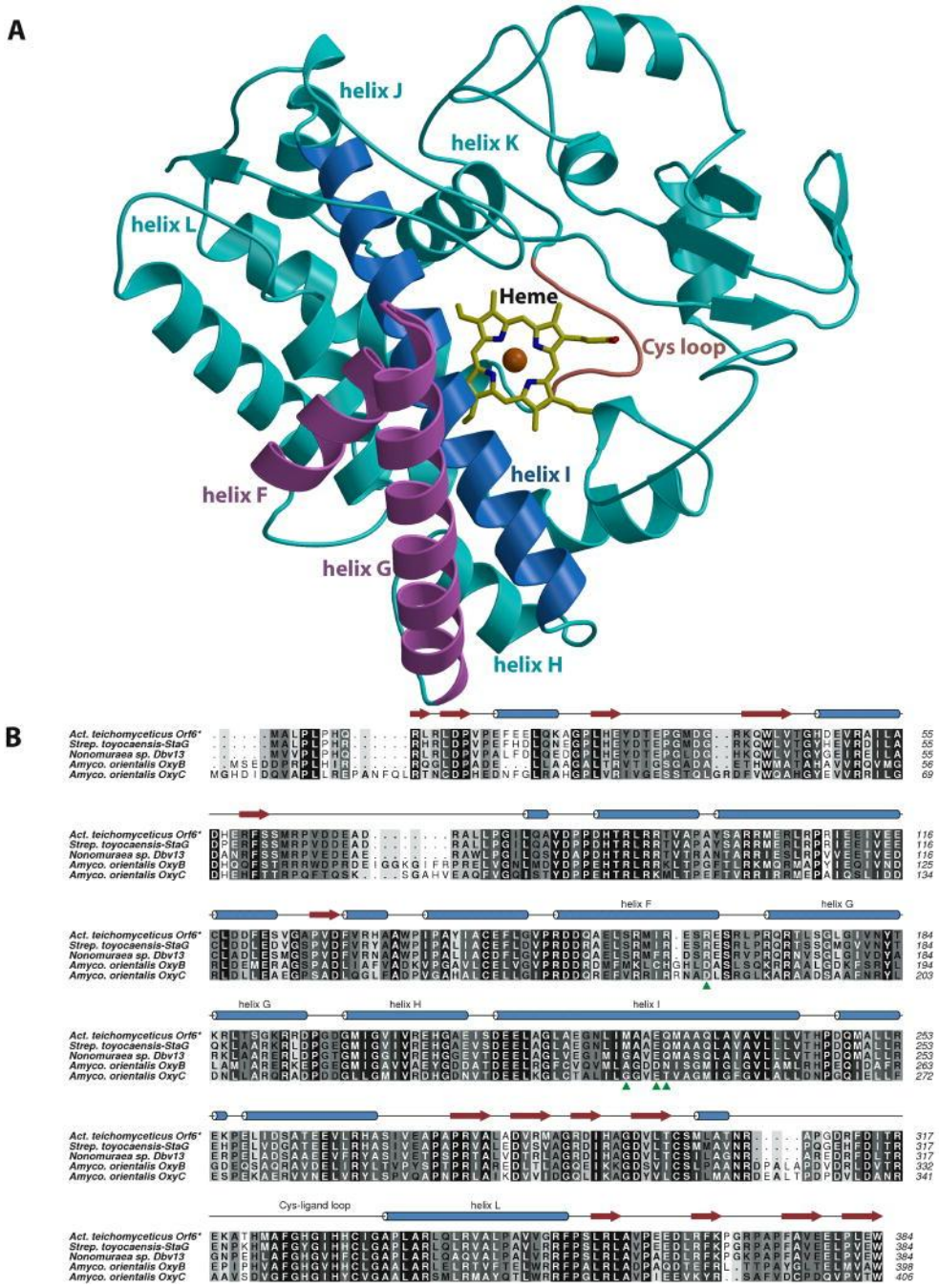


Figure 2-3. (A) Ribbon diagram showing the overall structure of Orf6*. Important secondary structural elements discussed in the text include F and G helices (colored in pink), I helix (colored in blue), and Cys loop (colored in red). The heme is shown in yellow sticks and the iron atom as a red sphere. Other secondary structural elements are colored in cyan. (B) Structure-based multiple sequence alignment of the type IV glycopeptide cross-linking cytochrome P450 monooxygenases Orf6* (teicoplanin), StaG (A47934) and Dbv13 (A40926) with those of the type I glycopeptide monooxygenases OxyB and OxyC (both from the vancomycin cluster). Residues that are unique to Orf6* and may be catalytically relevant are shown with a green arrowhead.

When compared with the crystal structures of other P450s using DALI (Holm and Rosenstrom, 2010), Orf6* shows the highest structural similarity to OxyB (Zerbe et al., 2002), OxyC (Pylypenko et al., 2003), P450nor (Park et al., 1997), CYP105P1 (Xu et al., 2009), and CYP105D6 (Xu et al., 2010), among the substrate-free wild-type P450 structures (Table 2-2).

Table 2-2. Similarity of structure and sequence between Orf6* and other P450s.

P450 name[a]	PDB[b] code	Z-score	RMSD[c] Å	Sequence identity %	FG[d] conformation relative to Orf6*
OxyB	1lg9	44.9	2.2	38	Open
CYP105 (MES)	2z36	44.8	1.9	32	Same
OxyC	1ued	44.7	2.1	37	Open
CYP105P1 (FLI)	3aba	43.8	2.5	36	Close
CYP105P1 (PIM)	3e5k	43.6	2.0	35	Same
P450nor	1rom	43.4	2.0	26	Same
CYP105P1	3e5j	43	2.1	35	Same
CYP105D6	3abb	42.8	1.9	33	Same

[a]P450s that show the highest structural similarity to Orf6*. Shown in parenthesis are bound substrates/ligands. MES: 2-(N-morpholino)-ethanesulfonic acid; FLI: Filipin I; PIM: 4-Phenyl-1H-imidazole.

[b]Protein Data Bank.

[c]Root mean square deviation.

[d]F and G helices of the corresponding P450s.

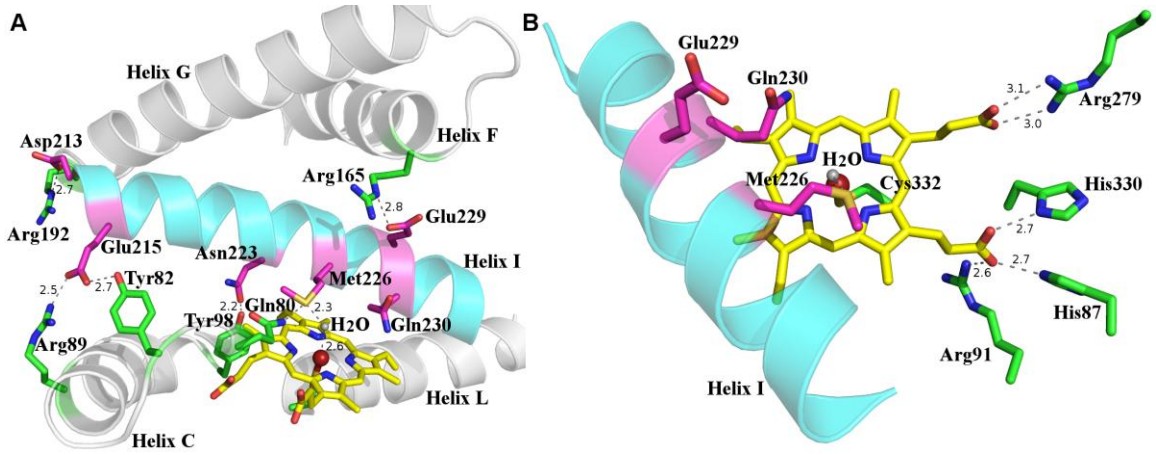


Figure 2-4. (A) Interactions between residues from I helix and those from other secondary structural elements. The I helix is shown in cyan and residues are colored in magenta, other helices are shown in gray and residues in these helices are colored in green. The heme is shown as a yellow stick, the iron shown as a maroon sphere, and the iron bound water as a gray sphere. Hydrogen bond interactions are shown as dash lines, next to which the distances are shown. (B) Active site of Orf6* showing critical catalytic residues and those interacting with the heme propionate groups. The I helix is shown in cyan, catalytically relevant residues conserved among type IV glycopeptide P450 monooxygenases are shown in magenta, Cys332 and other heme interacting residues shown in green. Hydrogen bond interactions are shown as dash lines, next to which the distances are shown.

Active Site Architecture of Orf6*

Orf6* shows the conserved four-helix bundle core, formed by D, E, I, and L helices, with the heme group confined between the I and L helices (Figure 2-4A), and above the Cys-ligand loop adjacent to the amino-terminus of the L-helix. This Cys-ligand loop contains the P450 signature amino acid sequence FxxGxHxCxG, with the absolutely conserved Cys332 being the proximal ligand to the heme iron. The iron atom is in the heme plane and its distance to the thiolate sulfur of Cys332 is 2.3 Å. A solvent molecule, located 2.6 Å from the heme iron, completes the coordination shell (Figure 2-4A), consistent with the low spin heme state of substrate-free Orf6* observed in the UV-visible spectra (Figure 2-2).

The long I-helix spans the entire catalytic site and forms a wall above the heme pocket. In most P450s, two highly conserved and catalytically important residues, an acidic residue immediately followed by a threonine, are usually found in the I-helix over the pyrrole ring B of the heme group (Graham and Peterson, 1999; Denisov et al., 2005). In contrast, in Orf6*, the acidic residue is Glu229 but the residue immediately to its carboxy-terminus is Gln230 (Figure 2-4B). This variation is not unprecedented as the corresponding residues in OxyB are Asp239-Asn240, with the latter forming hydrogen bonds with two water molecules in the active site (Zerbe et al., 2002). Although the side chain of Gln230 in Orf6* points to the heme group as in OxyB, there are no water molecules within hydrogen bonding distance to its backbone or side chain. This sequence variation is conserved at the corresponding place in the sequences of most of the phenol-coupling P450s involved in glycopeptide antibiotics biosynthesis (Figure 2-3C). The Gln230 side chain might play a similar role to the hydroxyl group of the usual threonine in forming a proton delivery channel during catalysis (Denisov et al., 2005) or be involved in hydrogen bond interactions with the extended peptide substrate.

In striking contrast to the small neutral amino acid (glycine or alanine) typically found three residues before the conserved acidic residue in the I-helix of P450s (Park et al., 1997; Zerbe et al., 2002; Pylypenko et al., 2003; Oshima et al., 2004), Orf6* contains a methionine residue (Met226) (Figure 2-4B) at this position. The side chain of Met226 points into the active site and the sulfur atom forms a hydrogen bond with the heme iron-coordinating water molecule. Although the function of Met226 is not yet clear, corresponding methionine residues are found in *S. toyocaensis* StaG and *Nonomuraea sp.* Dbv13 (Figure 2-3C), both of which are P450s cross-linking the side chains of the aromatic residues 1 and 3 in the type IV glycopeptide antibiotics A47934 and A40926(Pootoolal et al., 2002; Hadatsch et al., 2007), analogous to the Orf6* cross-linking reaction in teicoplanin biosynthesis(Li et al., 2004). All other putative phenol-coupling P450s in glycopeptide antibiotics biosynthesis have the more typical glycine or alanine at this position. The conservation of this methionine residue suggests that it plays a role in regiospecificity of the 1-3 cross-linking reaction. Attempts to carry out mutational analysis at Met226 are limited by the fact that attempts to produce the

synthetic, singly cross-linked, heptapeptide substrate coupled to a PCD domain have thus far been unsuccessful.

The active site conformation of Orf6* is stabilized by a number of hydrogen bonds between residues from I-helix and those from other secondary structure elements (Figure 2-4A). Hydrogen bonds from the side chain carboxylate oxygens of the I-helix residues Glu229 and Asp213 to the backbone amides of the F-helix residue Arg165 and the G-helix residue Arg192, mediate the interaction of the I-helix with the F and G helices. The side chain carboxylate oxygens of Glu215 from the I-helix form hydrogen bonds with the guanidine nitrogen atom of Arg89 from the N-terminal portion of the C-helix, and the hydroxyl group of Tyr82 from the B'-C loop, respectively. The C-D loop forms interaction with the I-helix through the hydrogen bond between the hydroxyl group of Tyr98 and the side chain carbonyl oxygen of the I-helix residue Asn223.

Orf6* Substrate Binding Pocket Comparison with Known P450s

A significant difference between the structures of Orf6* and other known P450s is in the orientation of F and G helices, which are believed to be important for substrate binding (Gotoh, 1992; Graham and Peterson, 1999; Denisov et al., 2005). Although the F and G helices in Orf6* are similar to those in OxyB (Zerbe et al., 2002) and OxyC (Pylypenko et al., 2003) in terms of length and fold, their relative orientations are significantly different. In particular, in Orf6* the F and G helices are rotated toward the active site, resulting in a much more closed substrate binding pocket (Figure 2-5A and B). This is unexpected because the Orf6* substrate, teicoplanin heptapeptide (Li et al., 2004), is nearly the same size as the vancomycin heptapeptide substrate of OxyB and OxyC (Pelzer et al., 1999; Zerbe et al., 2002; Stegmann et al., 2006) (Figure 2-5C). While the relative orientation of the F and G helices are similar in OxyB and OxyC (as would be expected given that they both work on the same substrate), the orientation of the F and G helices in Orf6* is much more similar to that of P450nor (Park et al., 1997), a nitric acid reductase whose substrate is substantially smaller compared to the teicoplanin aglycone substrate of Orf6*.

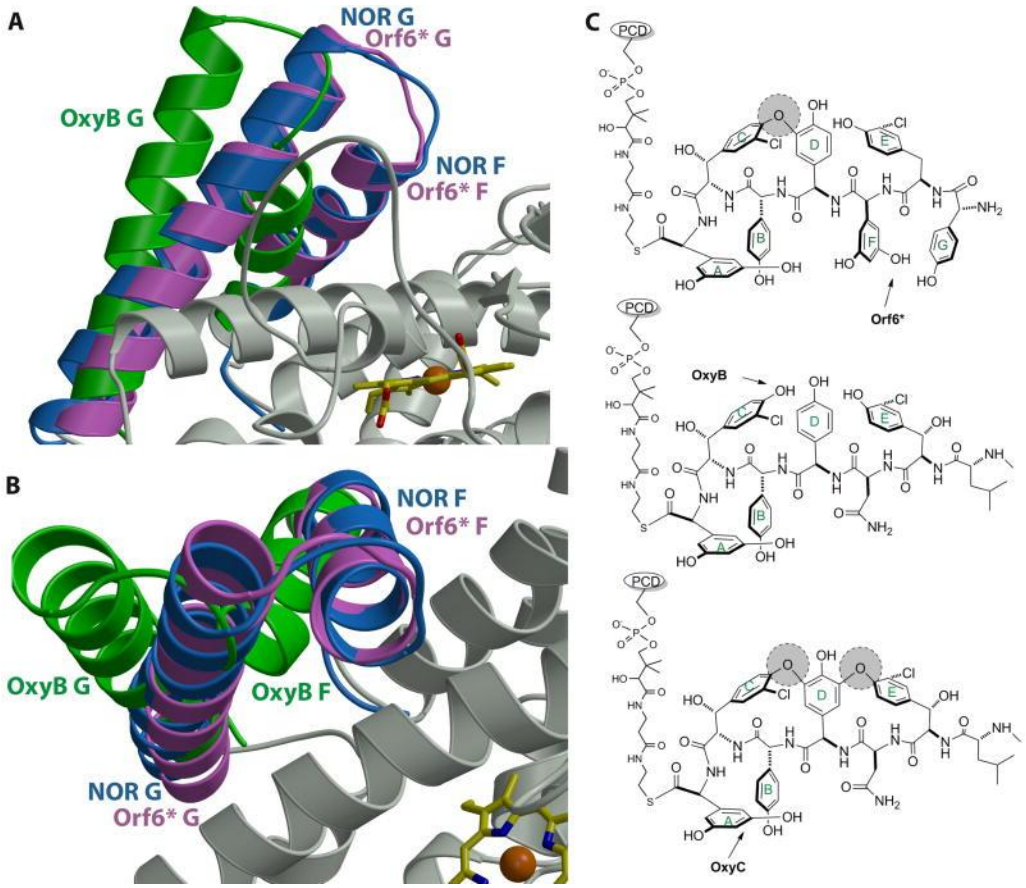


Figure 2-5. (A) and (B) Orthogonal views of a least squared superposition of the crystal structure of Orf6* (in pink) with those of P450nor (in blue) and OxyB (in green). Note that the F and G helices of Orf6* and OxyB deviate significantly, despite the fact that both enzymes work on large, peptidic substrates. The F-G helices of Orf6* are in a closed conformation, similar to those observed in structures of P450s that work on small molecule substrates (e.g., P450nor). (C) A comparison of the substrates of Orf6*, OxyB and OxyC, which may provide a rationale for the positioning of the F and G helices in these enzymes. Existing cross-links in the substrate are shown as shaded circles and an arrow indicates the target site for each enzyme.

Prior analysis of the substrate binding pockets of OxyB and P450nor lead to the speculation that the larger size of the OxyB substrate dictated the need for the more open binding pocket (Zerbe et al., 2002). However, our structure of Orf6* suggests that the binding pocket capped by the F and G helices may not only be determined by the substrate size, but also its *orientation*. Orf6* catalyzes the cross-linking at the amino-terminus of the teicoplanin heptapeptide (Li et al., 2004), while the target of OxyB is the bulky center of the vancomycin heptapeptide (Pelzer et al., 1999; Zerbe et al., 2002; Stegmann et al., 2006). Consequently, the heptapeptide substrate must be positioned in

the active site in substantially different orientations to afford the regiospecificities of each individual enzyme. While OxyC carries out the coupling reaction at the carboxy-terminus of the vancomycin heptapeptide, the open conformation of this enzyme can also be understood in the context of substrate orientation. Previous gene inactivation studies have suggested that the OxyC reaction is the last coupling reaction in vancomycin biosynthesis (Pylypenko et al., 2003; Stegmann et al., 2006), so its substrate, which is cross-linked between residues 2-4 and 4-6, has a rigid wide conformation that requires an open binding site (Figure 2-5C). In contrast, genetic studies of A47934 biosynthesis (Pootoolal et al., 2002; Hadatsch et al., 2007) imply that Orf6* catalyzes the second coupling reaction in teicoplanin biosynthesis. As a result, the Orf6* substrate is cross-linked only between residues 4-6 and the lack of a rigid structure in this substrate can be accommodated by a closed binding pocket in Orf6*.

In the Orf6* structure, the side chain of Glu229 in the I-helix forms a hydrogen bond with the backbone amide group of Arg165 from the F-helix (Figure 2-4A). This interaction likely stabilizes the inward conformation of F-helix, which in turn rotates the G-helix toward the active site. Such an interaction is not found in OxyB, OxyC, or P450nor but the Glu and Arg residues are conserved in both StaG and Dbv13 (Figure 2-3C), both of which catalyze similar 1-3 coupling during the biosynthesis of the teicoplanin-type glycopeptide antibiotics A47934 and A40926, respectively. The F and G helices of Orf6* might undergo a further closure of the active site upon substrate binding, as has been observed in some structurally similar P450s (Table 2-2), such as CYP105P1 that binds the antifungal macrolide antibiotic filipin (Xu et al., 2009; Xu et al., 2010).

Implications for Catalysis on PCD-coupled Substrate by Orf6*

Studies by Robinson and co-workers established that such oxidative coupling reaction only occurs on substrate peptides attached as thioesters to the NRPS module (Puk et al., 2002). In the presence of a suitable electron source, purified recombinant OxyB can catalyze the 4-6 crosslink of a heptapeptide substrate conjugated to a peptide carrier domain (PCD) from the cognate NRPS through a thioester linkage (Woithe et al., 2007).

As the Orf6* 1-3 cross-linking reaction occurs at the amino-terminus of the peptide substrate, and the PCD conjugation is through a carboxy-terminal linkage, the substrate can be correctly registered in the active site without requiring neither the PCD domain nor the phosphopantetheine arm to be substantially wedged into the binding pocket. In contrast, for the OxyB and OxyC reactions that occur at the center and carboxy-terminus of the peptide, their substrates are accommodated in an open binding pocket that is necessary to fit significant portions of both the heptapeptide substrate as well as the phosphopantetheine group (Figure 2-5C).

In order to gain further insights into the determinants of substrate recognition, we docked the co-crystal structure of the P450(Biol)-tetradecanoic acid-acyl carrier protein (ACP) complex (Cryle and Schlichting, 2008) onto that of Orf6* (Figure 2-6A and B). Despite the fact that P450(Biol) and Orf6* share less than 40% sequence identity, the ACP can be docked onto the structure of Orf6* without significant steric clashes. Although the PCD-peptide conjugate substrate of Orf6* is distinct from the tetradecanoic acid-ACP substrate of P450(Biol), this model yields a likely trajectory for the teicoplanin aglycone backbone. In this model, the unique Met226 present in Orf6* is located proximal to predicted locations of substrate residues 1 and 3, suggesting that this residue stabilizes the orientation of the substrate backbone to facilitate cross-linking (Figure 2-6B). The sulfur atom Met226 may form hydrogen bonds with the phenol hydroxyls on either residues 1 and/or residue 3 of the substrate aglycone. Similarly, residue Gln230, which is conserved only among P450s that catalyze 1-3 cross-links on glycopeptides, may form hydrogen bonds with the phenol hydroxyls on residues 2 or 4 to stabilize the extended substrate prior to the coupling reaction.

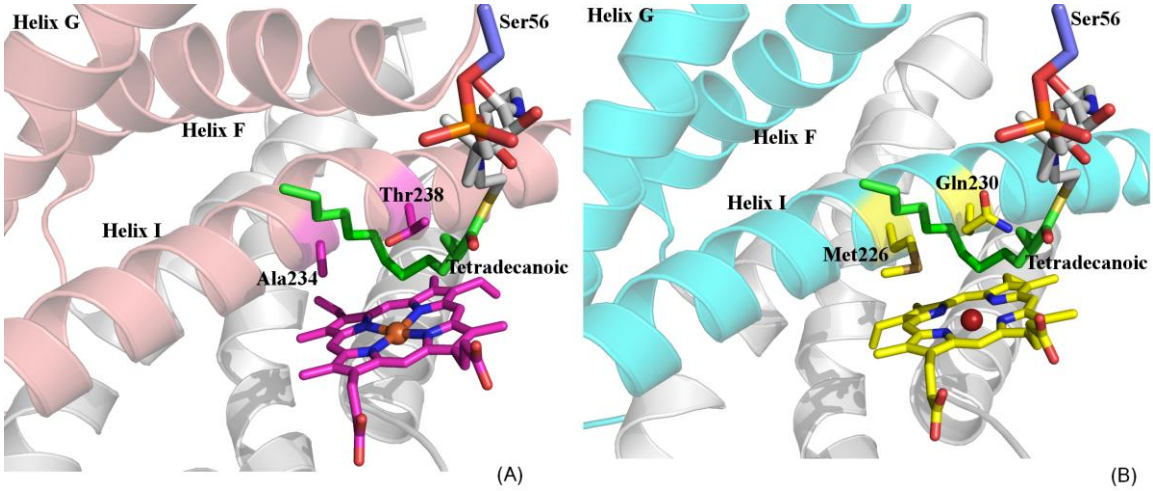


Figure 2-6. (A) Structure of the P450(Biol)-ACP-tetradecanoic acid complex near the vicinity of the active showing the F, G, and I helices (in pink), the ACP-bound substrate (in green) and catalytically important residues (in magenta). (B) A docking model of the putative Orf6* complex with the F, G, and I helices colored in cyan. This model suggests that Met226 and Gln230, which are conserved only among Orf6* orthologues, may play roles in steering and stabilizing the extended peptide substrate.

Overall structures of the acyltransferases tAtf and aAtf

The structures of tAtf and aAtf were determined in complex with the acyl donor substrate octanoyl-CoA to the resolution of 2.2 Å (tAtf-occ) and 1.7 Å (aAtf-occ), respectively. The two Atfs are highly homologous in both primary sequences (identity: 71%; similarity: 82%) (Altschul et al., 1997) and three-dimensional structures (RMSD: 1.1 Å over 318 aligned C_α atoms for the ligand-complexed structures) (Holm and Rosenstrom, 2010) (Figure 2-7). For simplicity, the description of the overall structure will be based on the high resolution aAtf-occ structure.

The overall structure is composed of two domains of similar sizes, with the amino-terminal domain (NTD) and the carboxy-terminal domain (CTD) consisting of approximately 170 residues and 150 residues, respectively (Figure 2-7). A DALI search (Holm and Rosenstrom, 2010) showed that the overall structure does not have significant similarity to any published protein structures in the Protein Data Bank.

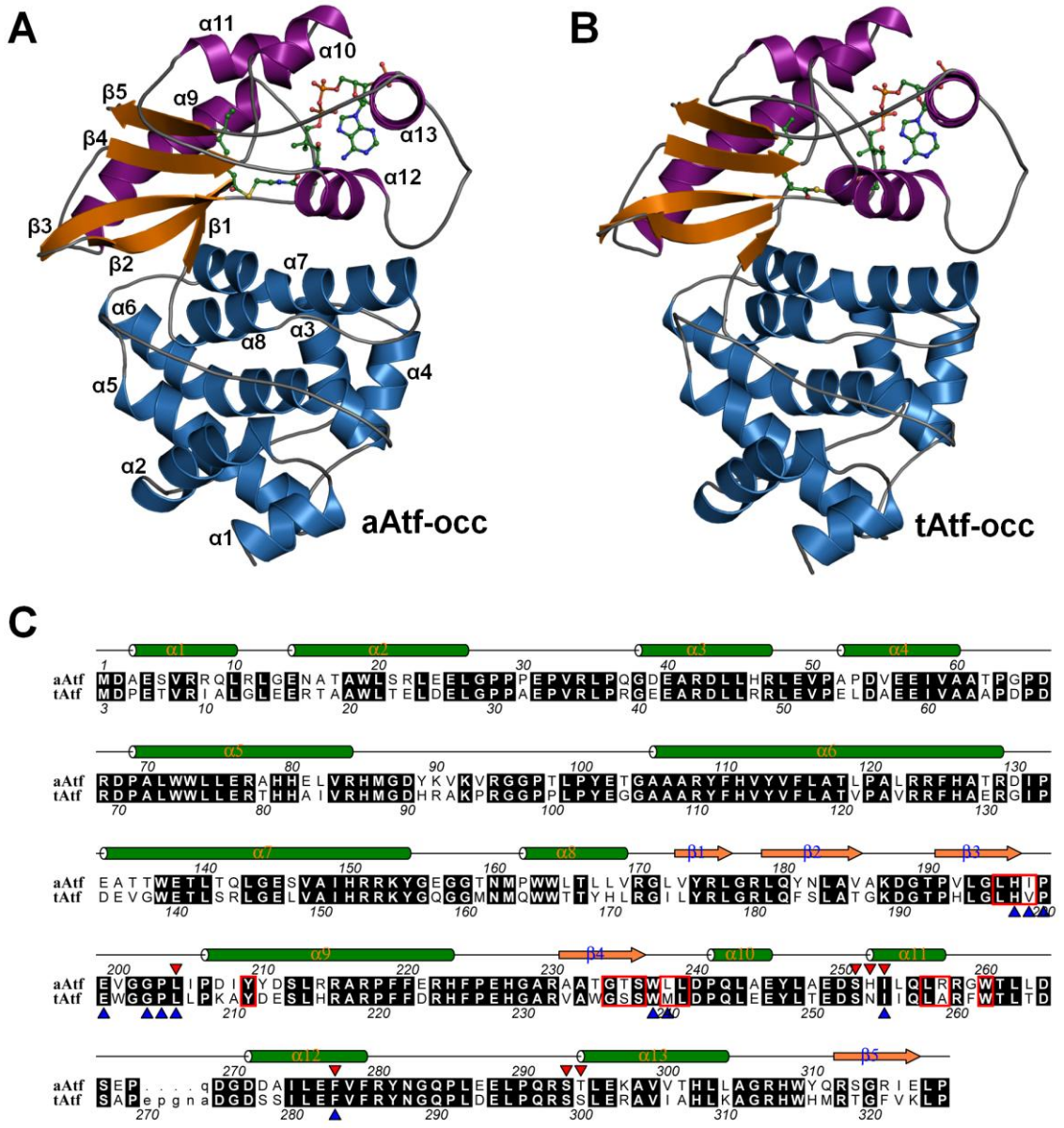


Figure 2-7. (A) (B) Overall structures of the acyltrasferases in complex with octanoyl-CoA (aAtf-occ and tAtf-occ). The N-terminal domain is shown in blue; the C-terminal domain is shown in purple for α-helices and orange for β-strands. The ligand is shown as green ball-and-stick representation. (C) Structure-based sequence alignment of the two Atfs. Secondary structures are based on aAtf-occ structure. Identical residues are highlighted by black background. Red downward triangles, blue upward triangles, and red frames indicate residues that are involved in the interactions with the 3'-phosphoadenosine-5'-diphosphate, the pantothenoylcysteamine, and the octanoate moieties, respectively.

The NTD is composed of eight α-helices, α1-α8 (Figure 2-7). It may represent a novel fold as no structures similar to these domains are found in the Protein Data Bank. The CTD is linked to the NTD through a loop (residues Arg170-Leu172) between helix α8 and strand β1. It is composed of five α-helices and a twisted five-stranded β-sheet (Figure 2-8A). The β-sheet is wrapped by the α-helices, with one side open to the NTD (Figure 2-7). The CTD structure shows only modest similarity to the structure of a glyphosate N-acetyltransferase (GAT; PDB code: 2JDD) (Siehl et al., 2007) (Figure 2-8B), with an RMSD of 2.9 Å over only 86 aligned C_α atoms between tAtf-apo and GAT. Most of the structurally similar residues are located in the β-sheet of the two proteins. GAT was identified by its overall fold as a member of the GCN5-related N-acetyltransferase (GNAT) superfamily (Neuwald and Landsman, 1997; Siehl et al., 2007).

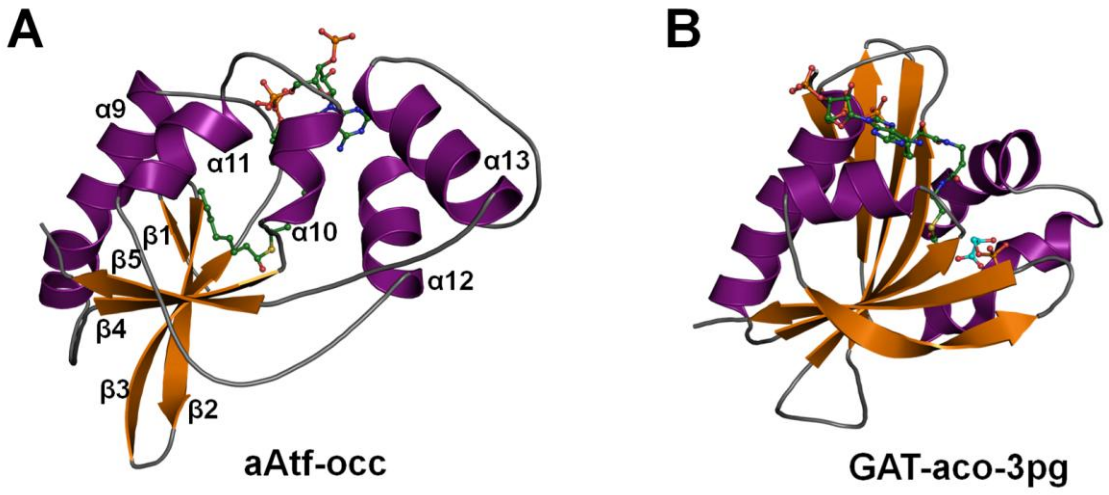


Figure 2-8. (A) CTD of aAtf-occ complex structure. α-helices are shown in purple and β-strands are shown in orange. Octanoyl-CoA is shown as green ball-and-stick representation. (B) Structure of glyphosate N-acetyltransferase (GAT; PDB code: 2JDD) in complex with acetyl-CoA (aco, colored in green) and 3-Phosphoglyceric acid (3pg, colored in cyan).

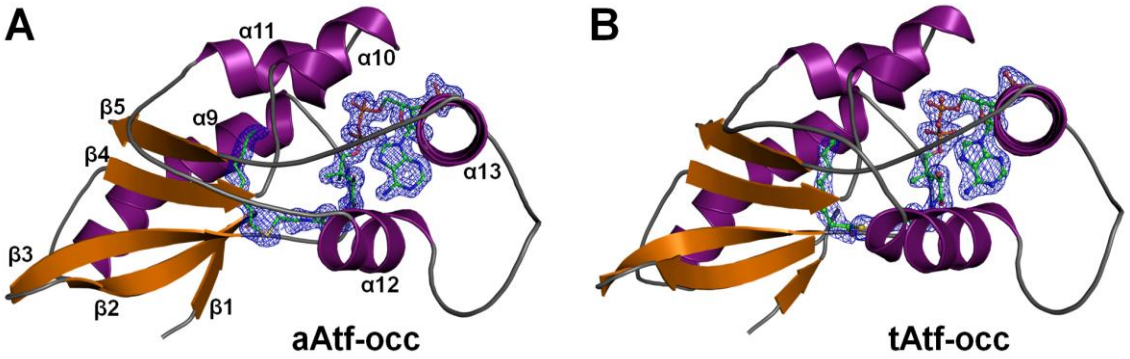


Figure 2-9. CTDs of the structures of (A) aAtf-occ and (B) tAtf-occ. α-helices are shown in purple and β-strands are shown in orange. Octanoyl-CoA is shown as green ball-and-stick representation. $2F_o$-F_c electron density map is shown as blue mesh for the ligands.

Octanoyl-CoA binding of the acyltransferases: CoA moiety

The structures of tAtf-occ and aAtf-occ show that an octanoyl-CoA molecule is bound to the CTDs of these Atfs (Figure 2-7). In both structures the acyl substrate is bound in almost identical modes (Figure 2-9). The ligand in the aAtf-occ structure shows unambiguous electron density for the entire molecule, while in the tAtf-occ structure the ligand shows relatively weaker density for the octanoyl moiety (Figure 2-9). Therefore, the dissection of the acyl substrate binding pocket will be based on the aAtf-occ structure.

The polar head group of the CoA molecule, 3'-phosphoadenosine-5'-diphosphate, is largely exposed to the solvent (Figure 2-10A). Each of the five α-helices in the CTD presents one end to contribute in stabilizing the CoA head group. The contacts between the head group and the protein are mostly polar interactions including hydrogen bonds and salt bridges, except the hydrophobic interactions between the adenine ring and the side chain of Leu295, and between the ribose ring and the side chain of Leu247 (Figure 2-10B). Both leucine residues are conserved in aAtf and tAtf (Figure 2-7C). The adenine ring forms two hydrogen bonds with the main chain oxygen atom of Phe277 (3.4 Å) and the side chain oxygen of Ser293 (2.9 Å). The 3'-phosphate group forms a salt bridge with the main chain nitrogen atom of Thr294 (2.9 Å). The 5'-diphosphate group forms a hydrogen bond with the side chain oxygen of Ser251 (2.6 Å) and several salt bridges with the main chain nitrogen atoms of Leu204 (2.8Å), His252 (2.9 Å), and Ile253 (2.9 Å), and

the side chain nitrogen atom of His252 (2.7 Å) (Figure 2-10B). All these residues are conserved or highly similar in aAtf and tAtf (Figure 2-7C).

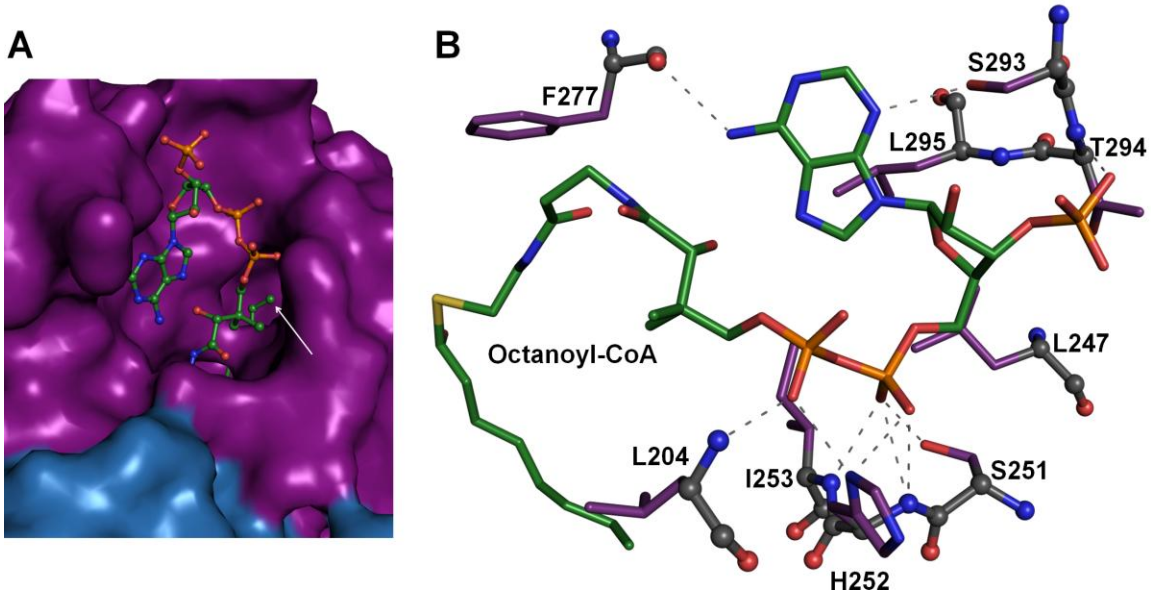


Figure 2-10. (A) Surface representation of aAtf-occ structure, showing the highly exposed 3'-phosphoadenosine-5'-diphosphate moiety of octanoyl-CoA. Regions from the CTD and NTD are colored in purple and blue, respectively. The ligand is shown as green ball-and-stick representation. Note that the tail of the octanoate group is visible, as indicated by the arrow. (B) Interactions between the 3'-phosphoadenosine-5'-diphosphate moiety and aAtf protein. The side chains of the interacting residues are shown as purple sticks; the main chain atoms are shown as ball-and-sticks, with the carbon atoms in black. Octanoyl-CoA is shown as green sticks. Polar interactions are indicated by dashed lines (see text for distances).

The pantothenoylcysteamine moiety of the CoA molecule spans the pocket formed by the five α-helices in the CTD (Figure 2-11). It bends at the center of the moiety by approximately 90 degrees and travels towards the β-sheets (Figure 2-11). It is accommodated mostly by hydrophobic interactions, except for the hydrogen bond between the terminal carbonyl oxygen of the pantothenate group and the indole nitrogen atom of Trp237 (3 Å), and the hydrogen bond between the cysteamine nitrogen atom and the main chain oxygen atom of Ile197 (3 Å) (Figure 2-11). The hydrophobic interactions on one side of the pantothenoylcysteamine moiety are contributed by the loop (His196-Leu204) connecting strand β3 and helix α9, including the contributions from the side chains of His196, Ile197, and Leu204, and the main chains of His196-Glu199 and Gly202-Leu204. On the other side, the side chains of Trp237, Leu238, Ile 253, and

Phe277 provided more hydrophobic interactions (Figure 2-11). All these residues are also conserved or highly similar in aAtf and tAtf (Figure 2-7).

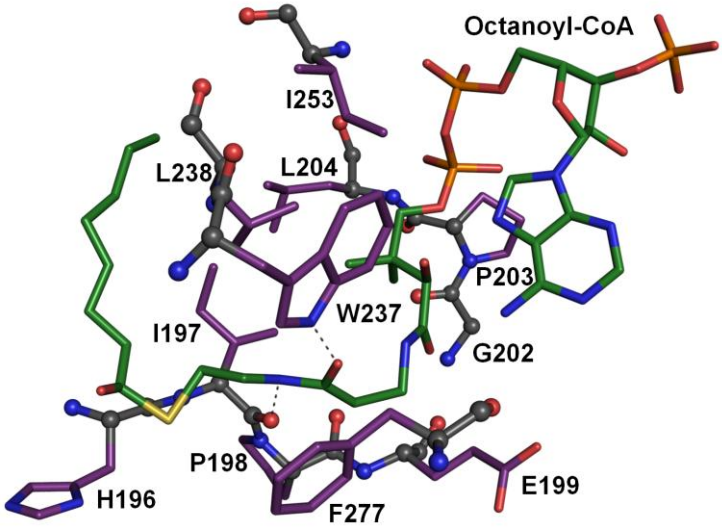


Figure 2-11. Interactions between the pantothenoylcysteamine moiety and aAtf protein. The side chains of the interacting residues are shown as purple sticks; the main chain atoms are shown as ball-and-sticks, with the carbon atoms in black. Octanoyl-CoA is shown as green sticks. Polar interactions are indicated by dashed lines (see text for distances).

Octanoyl-CoA binding of the acyltransferases: octanoate moiety

- Implications on the acyl chain specificity determinants

The thioester bond of the octanoyl-CoA molecule is positioned at the interface of CTD and NTD, with the octanoate group pointing into the CTD (Figure 2-7), forming an approximately 90-degree angle with the tail of the pantothenoylcysteamine moiety (Figure 2-11). Very interestingly, in aAtf the binding site for the pantothenoylcysteamine moiety and the octanoate group forms a continuous circle, displaying a donut-shape tunnel (Figure 2-12A). The circle starts and ends at C4 of the pantoate moiety. The pantothenoylcysteamine moiety and the octanoate group (20 atoms total in the main chain) travels along the circle and occupy about three fourth of the round tunnel. The remaining room in the tunnel between C8 of the octanoate group and C4 of the pantoate moiety appears to be suitable to accommodate around five more main chain atoms (Figure 2-12A), meaning that an acyl chain of around 13-carbon length could possibly be

accommodated. This is in good agreement to a recent report that aAtf cannot accept acyl chains longer than 14 carbons (Kruger et al., 2005).

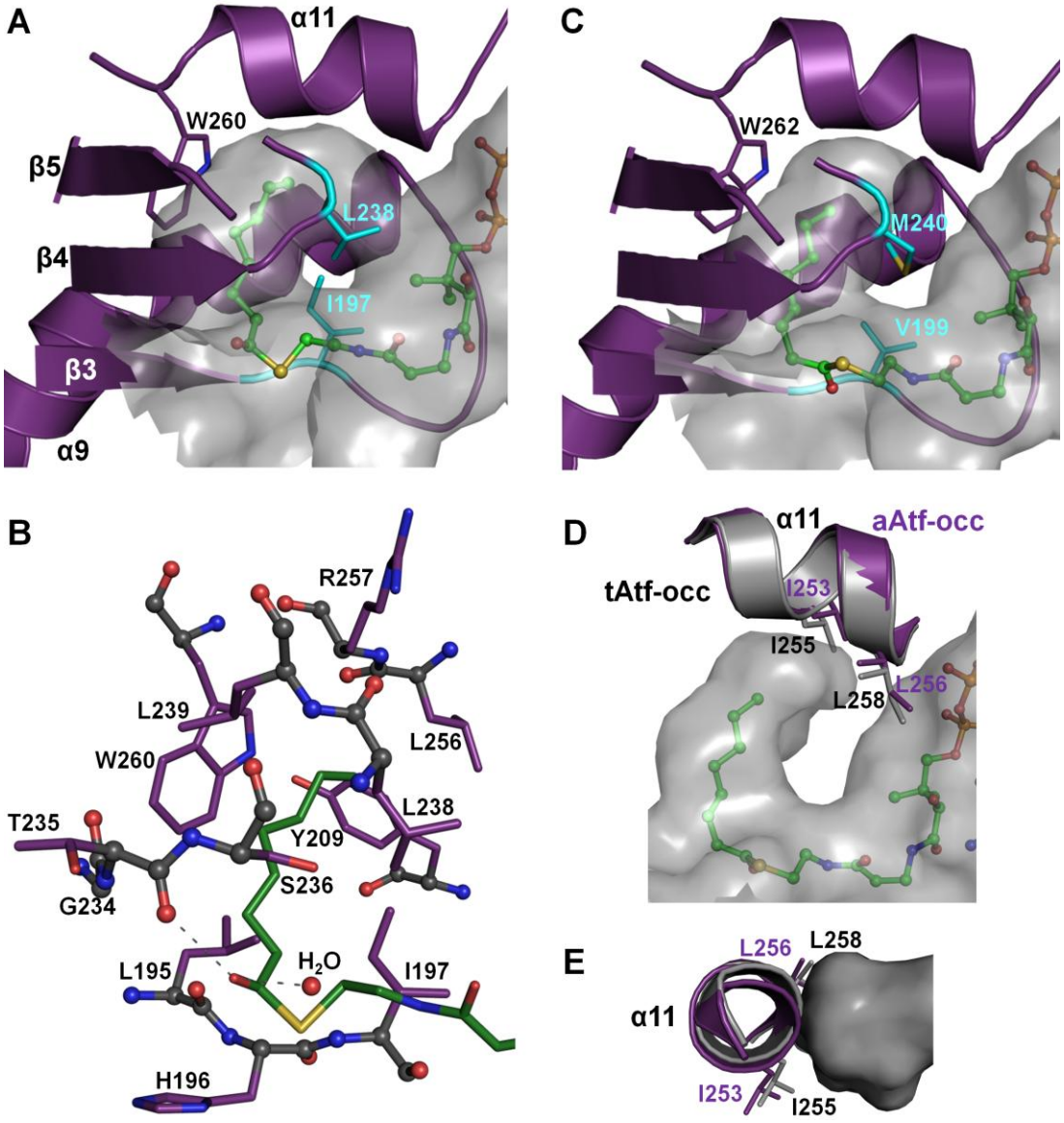


Figure 2-12. (A) The round binding tunnel for the pantothenoylcysteamine moiety and the octanoate group in aAtf-occ structure as shown by gray surface representation. Secondary structures involved in octanoate binding are shown in purple cartoon representation. The two "axel" residues are shown as cyan sticks. Octanoate-CoA is shown as green ball-and-stick representation. (B) Interactions between aAtf and the octanoate group of octanoyl-CoA (green sticks). The side chains of the residues are shown as purple sticks; the main chain atoms are shown as ball-and-sticks, with the carbon atoms in black. Polar interactions are indicated by dashed lines (see text for distances). (C) The round binding tunnel for the pantothenoylcysteamine moiety and the

octanoate group in tAtf-occ structure as shown by gray surface representation. The coloring is the same as in (A). (D) Superimposition of tAtf-occ (gray) and aAtf-occ (purple) structures, showing helix α11 of both structures and the round binding tunnel in tAtf-occ. (E) The same superimposed structures as in (D) viewed from the left side in (D).

The rotating axis, or "axel", of the round tunnel, is formed by the side chains of Ile197 and Leu238, which are located on the opposite faces of the "donut" (Figure 2-12A). Starting from the location where the thioester bond is positioned, the outer boundary of the acyl chain binding region of the tunnel is defined by helices α9 and α11 and strands β3-β5 (Figure 2-12A). The 5-stranded β-sheet in the CTD is twisted between strands β3 and β4, opening a cleft between the two strands (Figure 2-8A and Figure 2-12A). The octanoate group inserts into the pocket through this cleft and toward helix α11. This helix forms the boundary for the last portion of the round tunnel between C8 of the octanoate group and C4 of the pantoate moiety, presumably defining the path for the tail carbons of acyl chains longer than 8 carbon atoms (Figure 2-12A).

The octanoate group is accommodated almost exclusively by hydrophobic interactions, except that the carbonyl oxygen forms hydrogen bonds with the main chain oxygen of Thr235 and a water molecule (Figure 2-12B). The first four carbons are stabilized by the main chains of residues Leu195-Ile197 and Gly234-Ser236, and also the side chains of Leu195, His196, Ile197, and Leu238. On the other hand, carbons 5-8 are mainly stabilized by the side chains of Tyr209, Leu239, and Trp260, and the main chains of Leu256-Arg257. Notably, the major contribution comes from the indole ring of Trp260 (Figure 2-12B). The indole ring lies parallel and side by side to carbons 5-8, providing significant support for the acyl chain binding. Given that it is conserved in aAtf and tAtf (Figure 2-7C), the existence of this tryptophan might partly explain the preference of both Atfs on acyl chains longer than 5 carbons (Kruger et al., 2005).

All the octanoate-binding residues are conserved or highly similar in aAtf and tAtf (Figure 2-7C). tAtf displays a hydrophobic round tunnel similar to that in aAtf (Figure 2-12C). However, significant difference exists between the tunnels in the two Atfs

(Figure 2-12A, C, D, and E). The tunnel in tAtf is discontinuous between C8 of the octanoate group and C4 of the pantoate moiety, indicating relatively limited space in this region (Figure 2-12C). This is in good agreement with the recent findings that the aAtf has the highest activity on lauroyl-CoA, which has 12 carbons in the acyl chain, while the best substrate for tAtf is decanoyl-CoA, which has only 10 carbons in the acyl chain (Kruger et al., 2005).

Such difference in the acyl binding tunnel of the two Atfs could be attributed to mainly two aspects in their structures. First, in tAtf, the axel for the round tunnel is formed by the side chains of residues Met240 and Val199, while their equivalents in aAtf are Leu238 and Ile197, respectively (Figure 2-12A and C). Compared to its equivalent (Leu238), Met240 in tAtf inserts deeper along the axel, with the methylthio group pointing towards the end of the tunnel, which limits the space in this region (Figure 2-12C). Second, helix α11 of tAtf is slightly shifted towards the tunnel axel, inevitably bringing more constrains on the end of the tunnel (Figure 2-12D). For example, the side chains of Ile255 and Leu258 in tAtf are obviously closer to the tunnel compared to their equivalents (Ile253 and Leu256) in aAtf (Figure 2-12E).

Putative heptapeptide scaffold binding site in the acyltransferases

To investigate the activity of the Atfs, the acyl acceptor substrate was prepared by deacylating mature teicoplanin using the deacetylase/deacylase Orf2* from the teicoplanin biosynthesis pathway, a method reported in a recent study (Truman et al., 2006). The production of deacylated teicoplanin (DT) was confirmed by LC-MS (Figure 2-13A and Table 2-3). Octanoyl-CoA was used as the acyl donor and the activity of tAtf was tested. The formation of the product was indicated by a new peak in the LC trace and it was confirmed by mass spectrometry to be the re-acylated glycopeptide, octanoyl-teicoplanin (OT) (Figure 2-13A and Table 2-3). These results confirmed the acylation activity of tAtf on the teicoplanin scaffold.

Both aAtf and tAtf were reported to be promiscuous on the glycopeptide scaffold and they can accept both the teicoplanin and vancomycin scaffolds with significant activities

Table 2-3. **Mass spectrometry results for acylation reactions.**

	Calculated (g/mole)	Experimental (g/mole)	Difference (g/mole)
teicoplanin	1877.6	1916.6	39
deacylated teicoplanin	1723.4	1763.6	40.2
octanoyl-teicoplanin	1849.5	1888.6	39.1
vancomycin	1447.4	1448.8	1.4
octanoyl-vancomycin	1573.5	1575.7	2.2

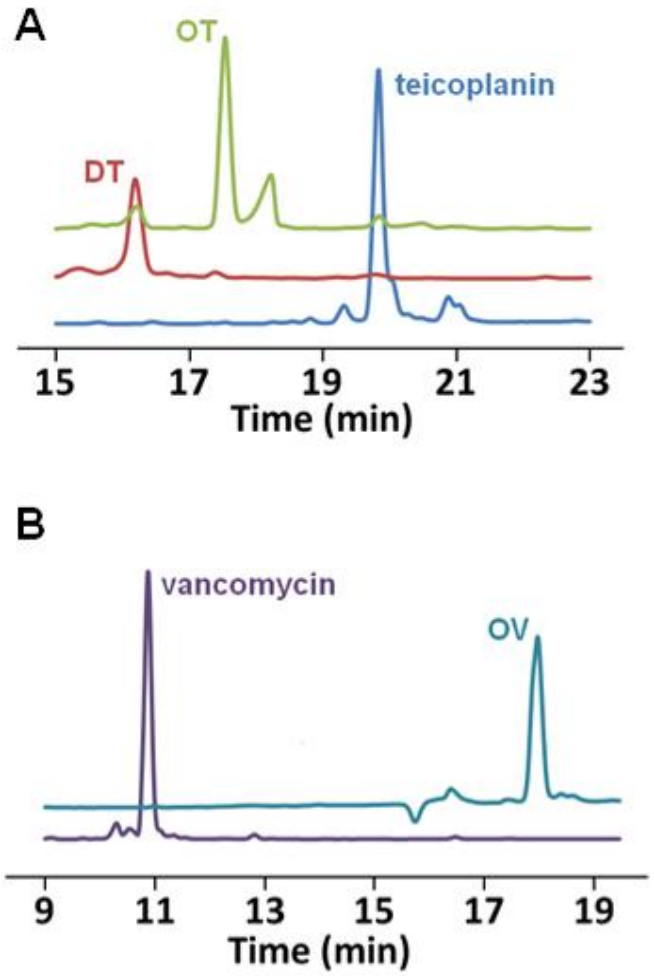


Figure 2-13. LC traces of heptapeptides. (A)DT: deacylated teicoplanin; OT: octanoyl-teicoplanin. (B) OV: octanoyl-vancomycin.

(Kruger et al., 2005). I tested the activity of tAtf using octanoyl-CoA and vancomycin, and indeed vancomycin was acylated to form octanoyl-vancomycin (OV) (Figure 2-13B and Table 2-3).

Attempts of co-crystallizing either of the Atfs with a heptapeptide scaffold have not been successful, largely due to solubility issues when mixing the protein and the heptapeptide. However, the structures of aAtf-occ and tAtf-occ provide valuable information on the putative binding site of the scaffold. They both display a clam-shape overall structure, with the NTD and CTD being the two shells and strand β1 being the "hinge" (Figure 2-14A). As a result, a large "mouth-like" open space exists between the two shells (Figure 2-14B). Moreover, the acyl substrate is bound in such a way that the thioester bond of the substrate is positioned at the interface of NTD and CTD, directly facing the large open space between the two domains (Figure 2-14). We speculate that the heptapeptide scaffold binds in this open space, with the sugar moiety of amino acid 4 (Figure 1-4) in close vicinity to the thioester bond for the acylation reaction to occur.

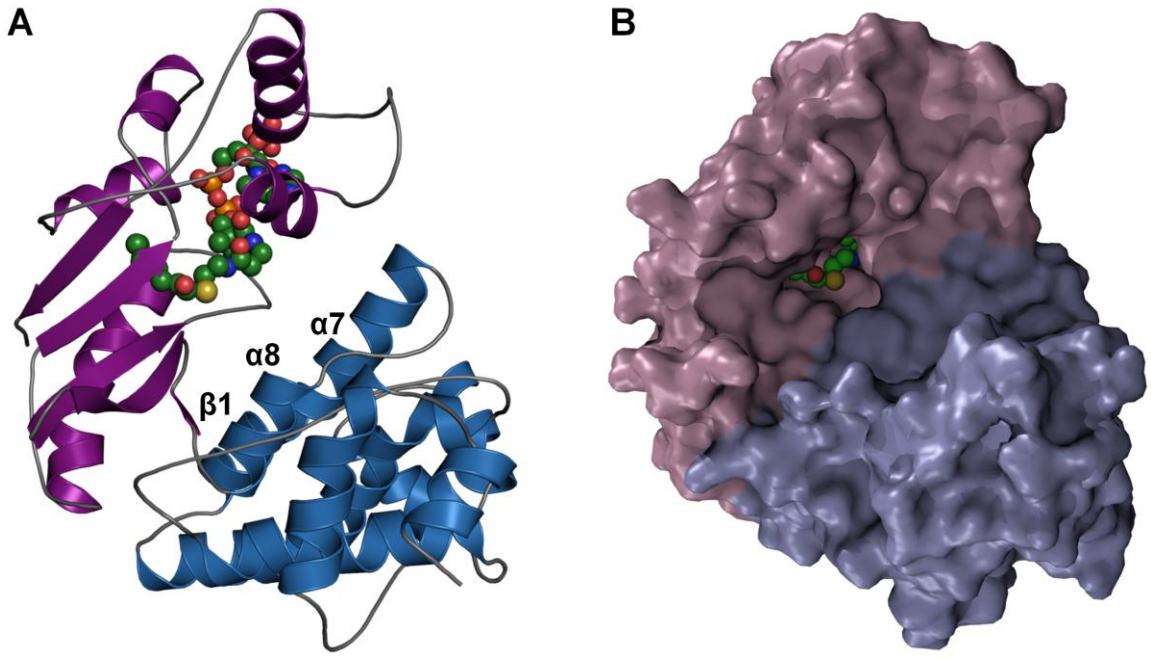


Figure 2-14. The putative heptapeptide scaffold binding site shown as (A) cartoon representation and (B) surface representation using the aAtf-occ structure. The NTD and

CTD are colored in blue and purple, respectively. Octanoyl-CoA is shown as green spheres.

Discussion

Monooxygenase Orf6*

In order to facilitate the development of novel glycopeptide antibiotics, we have investigated the cytochrome P450 monooxygenase Orf6* that is suggested to be the second coupling oxygenase in the biosynthesis of teicoplanin. Recombinant Orf6* yields a UV-visible absorption spectra characteristic of P450 heme-proteins, and indicates the existence of a low spin heme state for the substrate-free enzyme. The oxidized form of Orf6* can be reduced by sodium dithionite and bound via its heme iron (III) atom to imidazole. The 2.2 Å resolution crystal structure of Orf6* exhibits a typical P450-fold with a triangular prism shape. Among the substrate-free structures of known wild type P450s, OxyB, OxyC, P450nor, and CYP105D6 show the highest similarity to Orf6*.

Sequence comparisons of all the phenol-coupling P450s involved in glycopeptide antibiotics biosynthesis suggest primary sequence conservations that may reflect substrate preferences of this subgroup of P450s. Among these, residues Arg165, Met226, and Gln230 are conserved only in Orf6*, Dbv13, and StaG, which catalyze the extra 1-3 cross-link found in type IV glycopeptides. The conservations of these residues likely reflect their importance in stabilizing the extended peptide substrate of these particular P450 enzymes. The corresponding residues in OxyB, which catalyzes the 4-6 cross-link, are Leu174, Ala236, and Asn240, and these residues are absolutely conserved among all 4-6 cross-linking monooxygenases in the biosynthetic clusters of teicoplanin (Orf7*), A47934 (StaH), and A40926 (Dbv12). These patterns of conservation imply the importance of these residues in determining the regiospecificity and timing of the phenol-coupling P450s. Further analysis of structure-function relationships will require reconstitution of *in vitro* activities using various synthetic peptide substrates that have been singly cross-linked and then coupled to the appropriate peptide carrier domains.

Acyltransferases aAtf and tAtf

The acyltransferases involved in the biosynthesis of teicoplanin-type lipoglycopeptide antibiotics are valuable tools for semi-biosynthesis of novel lipoglycopeptide antibiotics. Initial attempts of using these acyltransferases combined with other tailoring enzymes to synthesize novel lipoglycopeptide derivatives proved to be successful (Kruger et al., 2005). To understand the mechanism of the acylation reaction and substrate recognition, we investigated two such acyltrasferases, aAtf and tAtf, which are involved in the biosynthesis of A40926 and teicoplanin, respectively. The structures of these two enzymes complexed with octanoyl-CoA provided detailed information on the determinants of the acyl chain specificity. The enzymes display a two-domain overall topology and the acyl-CoA substrate is bound to the CTD. While the polar head of the substrate is largely exposed to solvent, the hydrophobic portion, including the pantothenoylcysteamine moiety and the octanoate group, fits into a donut-shape tunnel deep in the center of CTD. These two groups occupy approximately three fourth of the tunnel. The length of the tunnel suggests that it can accommodate around 13 main chain carbon atoms in the acyl chain. On the other hand, a significant contribution for the acyl chain binding is provided by the interactions between the indole ring of Trp260 (aAtf) or Trp262 (tAtf) and carbons 5-8 in the acyl chain, suggesting that acyl chains shorter than 5 would have low binding affinity. These observations are in good agreement with recent report on the chain length preference of these acyltrasferases.

The octanoyl-CoA complexed structures revealed several residues important for the acyl chain binding and provided guidelines for engineering the acyl substrate binding pocket to accept different novel substrates. It is plausible to propose that mutating Trp260 to a smaller residue, such as alanine, would decrease the acyl chain binding affinity significantly. Mutating other hydrophobic residues in the pocket, including Leu195, Ile197, Tyr209, Leu238, and Leu239, to alanine, might also decrease the affinity to acyl substrates. However, the Ile197Ala and Leu238Ala mutations might have another effect on the pocket to allow larger substrates to be tolerated, for example, substrates with aromatic rings or large branches. This is because that these two residues insert into the center of the round tunnel to form the “axel”, and mutating them to smaller residues

might free up the center of the round tunnel and generate a “bun-shape” binding pocket, which might accommodate larger substrates. On the other hand, mutating Leu195, Ile197, and Leu238, which contribute to stabilizing the first four carbons on the acyl chain, to tryptophan or other aromatic residues, may increase the affinity to short chain acyl-CoA molecules. These mutations would provide useful information towards the development of novel lipoglycopeptides.

Both A40926 and teicoplanin are naturally produced as a mixture of compounds that differ almost only in the identities of the acyl chains. These acyl chains range from 10 to 12 carbons and several of them have a methyl branch after carbon 8 (Figure 2-1). This fact may be explained from the structures of aAtf and tAtf complexed with octanoyl-CoA. In both structures, the acyl binding tunnel show widened shape near the C8 position compared to the rest of the acyl-binding tunnel (Figure 2-15), which may account for the tolerance of a methyl branch in this region.

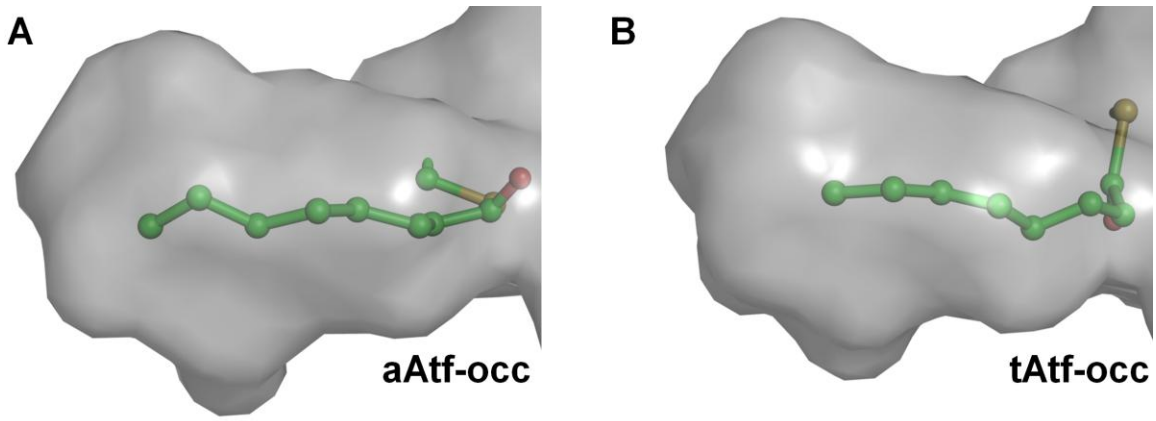


Figure 2-15. Side view of the octanoate-binding pocket shown as surface representation in the structures of (A) aAtf-occ and (B) tAtf-occ. The octanoate group is shown as green ball-and-sticks. For simplicity, most of the remaining part of the octanoyl-CoA molecule is not shown.

Forced by extensive hydrophobic interactions in the round tunnel, the octanoyl-CoA molecule forms two important turns, one at the center of the pantothenoylcysteamine moiety and the second one at the sulfur atom. These two turns restrict the long hydrophobic tail within the boundary of the CTD and positions the thioester bond at the

interface of CTD and NTD, presumably bringing it near to the target acylation site on the acyl chain acceptor. The interface between CTD and NTD resembles the open mouth of a clam, with the two domains being the two shells. This open mouth is very likely to be the binding site for the heptapeptide scaffold of the acyl chain acceptor. NTD might provide the major support for the scaffold binding, with helices α7-8 being the bed for the scaffold. There might be conformational change in the overall structure upon scaffold binding. For example, the two shells might close up to facilitate the scaffold binding. Complex structures of the Atfs with the scaffolds would be required to further understand the scaffold binding.

The structures of aAtf-occ and tAtf-occ showed that a conserved histidine residue (His196 in aAtf and His198 in tAtf) is in close vicinity of the carbonyl group of the octanoate moiety in the acyl-CoA substrate. Its side chain points towards the putative scaffold binding cavity in these two structures. Thus it is plausible that this imidazole ring would be forced to move away from the cavity upon scaffold binding and swing towards the carbonyl group of the octanoate moiety. This would stabilize the carbonyl oxygen atom and cause the carbonyl carbon to be vulnerable for a nucleophilic attack from the amino group on the acyl acceptor (Figure 2-16). Mutagenesis study and co-crystal structure with the acyl acceptor substrate would be required to examine the role of this histidine in the catalysis.

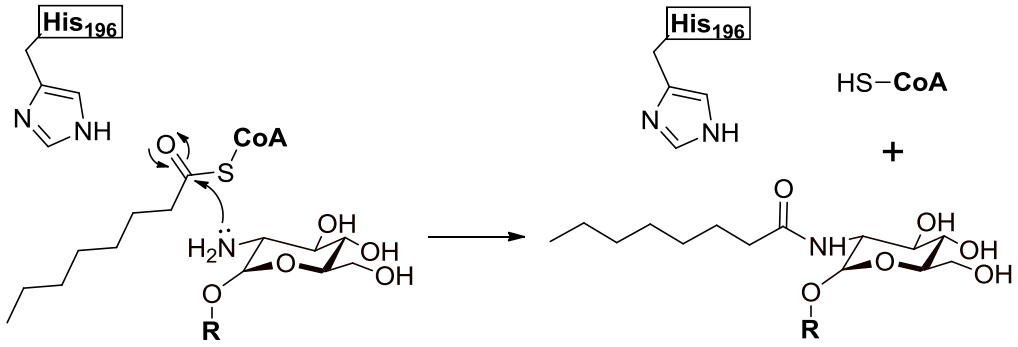


Figure 2-16. Putative mechanism of acylation by aAtf.

Chapter 3 Biochemical and structural studies on the catalysis mechanism and substrate specificity of a tobacco 4-coumarate:CoA ligase

Abstract

4-coumarate:CoA ligase (4CL; EC 6.2.1.12) belongs to the ANL superfamily of adenylating enzymes that contains three subfamilies: **a**cyl- and aryl-CoA synthetases, the adenylating domains of **n**on-ribosomal peptide synthetases, and **l**uciferases. As a central enzyme in the phenylpropanoid pathway in plants, 4CL provides precursors for numerous important metabolites and regulates the carbon flow in plants. The engineering of 4CL has attracted extensive interests, but great challenges exist due to lack of a thorough understanding of its catalytic mechanism and substrate-specificity determinants. Here we present high resolution crystal structures of *Nicotiana tabacum* 4CL isoform 2 (Nt4CL2) in complex with Mg^{2+} and ATP, with AMP and CoA, and with three different hydroxycinnamate adenylate intermediates, including 4-coumaroyl-AMP, caffeoyl-AMP, and feruloyl-AMP. Nt4CL2-Mg^{2+}-ATP structure displays the adenylate-forming conformation, whereas other structures are in the thioester-forming conformation. These structures not only represent a rare example of an ANL enzyme captured in both conformations, but also reveal several important structural features, including a highly ordered P-loop that interacts with the pyrophosphate group of ATP molecule, and the mainly hydrophobic CoA tunnel and hydroxycinnamate-binding pocket of a 4CL protein. Our structures combined with mutagenesis studies identified several crucial residues involved in catalysis, ATP binding, and hydroxycinnamate substrate specificity determination. Lastly, we generated a deletion mutant of Nt4CL2, ΔVal341, that possesses the unusual sinapate-utilizing activity, and explained the molecular rational for this activity through structural study of this mutant.

Introduction

The phenylpropanoid pathway represents the central biosynthetic nexus for the production of an array of plant metabolites including monolignols, phytohormones, flavonoids, and phenylpropenes (Hahlbrock and Scheel, 1989). The pathway directs the flow of carbon from primary metabolism to diverse secondary metabolism branch

pathways. These phenylpropanoids serve a range of functions *in planta*, including providing mechanical support, UV protection, defense against pathogens, and mediating interactions with pollinators (Hahlbrock and Scheel, 1989; Lee and Douglas, 1996). Several of these metabolites are also of pharmaceutical interest, including the stilbenoid resveratrol (lifespan extension) (Richard et al., 2011), the phenylpropene shikimol (a precursor in the synthesis of the entheogenic drug MDMA) (Gimeno et al., 2005), and the coumarin umbelliferone (antioxidant) (Kanimozhi et al., 2011). Lignin, the polymeric product of monolignols, has been the focus of research that targets agro-industrial uses of plant biomass (Boudet et al., 2003; Ragauskas et al., 2006).

The core reactions of the general phenylpropanoid pathway start with the deamination of phenylalanine (by phenylalanine ammonialyase) to yield *trans*-cinnamic acid (Figure 1-6), which is then *para*-hydroxylated (by cinnamate-4-hydroxylase) to form 4-coumaric acid. A second hydroxylation at the C3 position of the phenolic ring (by coumarate 3-hydrolase) affords caffeic acid. Caffeic acid can undergo *O*-methylation at C3 to yield ferulic acid or additional hydroxylation and *O*-methylation at C5 to yield sinapinic acid (Figure 1-7). A key branching point in the phenylpropanoid pathway is the formation of coenzyme A (CoA) thioesters of these hydroxycinnamic acid derivatives catalyzed by the enzyme 4-coumarate:CoA ligase (4CL, EC 6.2.1.12), which leads to the productions of a variety of important molecules (Figure 1-6).

4CL proteins play vital roles in regulating carbon flow in plant biosynthetic pathways, as their product hydroxycinnamate derived-CoA thioesters serve as precursors for various branching pathways of phenylpropanoid synthesis. For example, 4-coumaroyl-CoA is a substrate for the first committed step in flavonoid biosynthesis (Cukovic et al., 2001), and lignin monomers, including 4-coumaryl alcohol, corniferyl alcohol, and sinapyl alcohol, are derived from the hydroxycinnamoyl-CoA thioesters (Boerjan et al., 2003). As the level and composition of lignin in plants largely determine the efficiency of biomass utilization (Boudet et al., 2003), 4-coumarate:CoA ligases have been studied extensively in efforts to produce engineered plants with improved biomass utility. Although successful studies have been reported recently (Kajita et al., 1997; Lee et al., 1997; Hu et

al., 1999), 4CL engineering remains challenging due to limited knowledge of its catalytic mechanism.

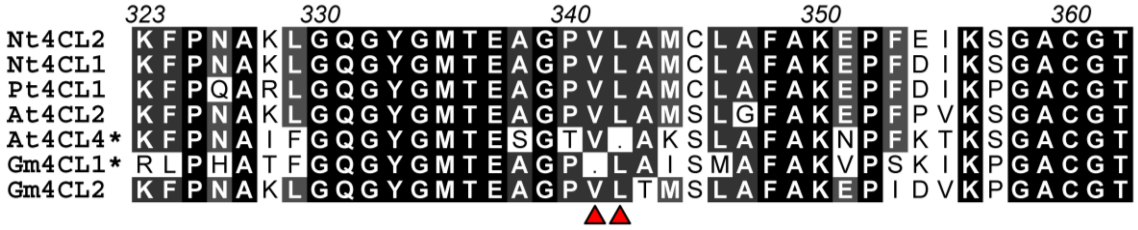


Figure 3-1. Sequence alignment of 4CL isoforms from different organisms. Nt: *Nicotiana tobacum*; Pt: *Populus tomentosa*; At: *Arabidopsis thaliana*; Gm: *Glycine max*.

Homologs of 4CL occur largely in higher plants, including *Glycine max* (soybean) (Knobloch and Hahlbrock, 1975; Lindermayr et al., 2002), *Petunia hgbrida* (petunia) (Ranjeva et al., 1976), *Pisum sativum* (pea) (Wallis and Rhodes, 1977), *Petroselinum crispum* (parsley) (Douglas et al., 1987), *Solanum tuberosum* (potato) (Becker-Andre et al., 1991), *Pinus taeda* (loblolly pine) (Zhang and Chiang, 1997), *Nicotiana tabacum* (tobacco) (Lee and Douglas, 1996), *Populus tremuloides* (aspen) (Hu et al., 1998), and *Arabidopsis thaliana (Ehlting et al., 1999).* In many species multiple isoforms are expressed in various levels in different tissues and at different development stages. These isoforms display divergent substrate specificity against various hydroxycinnamic acid derivatives, as exemplified by the differing substrate profiles for the four isoforms of 4CL found in soybean (Lindermayr et al., 2002). Intriguingly, one of these 4CL isozymes, *G. max* 4CL1 (Gm4CL1), is capable of utilizing sinapinic acid as its substrate (Lindermayr et al., 2002; Lindermayr et al., 2003), which is unusual because most 4CLs homologs lack activity against this metabolite (Becker-Andre et al., 1991; Lee and Douglas, 1996; Allina et al., 1998; Hu et al., 1998; Ehlting et al., 1999). The only other 4CL protein with activity against sinapinic acid is *A. thaliana* 4CL4 (At4CL4). Most remarkably, the sinapate-utilizing activity of Gm4CL1 and At4CL4 is attributed to the absence of a single amino acid within the active site (a valine between Pro343 and Leu344 of Gm4CL1 and a leucine between Val370 and Ala371 of At4CL4) (Figure 3-1) (Lindermayr et al., 2003; Hamberger and Hahlbrock, 2004). Deletion of the equivalent Val or Leu in other paralogs result in significant activity against sinapinic acid

(Lindermayr et al., 2003; Schneider et al., 2003), but the molecular rationale for this substrate specificity shift has yet to be determined.

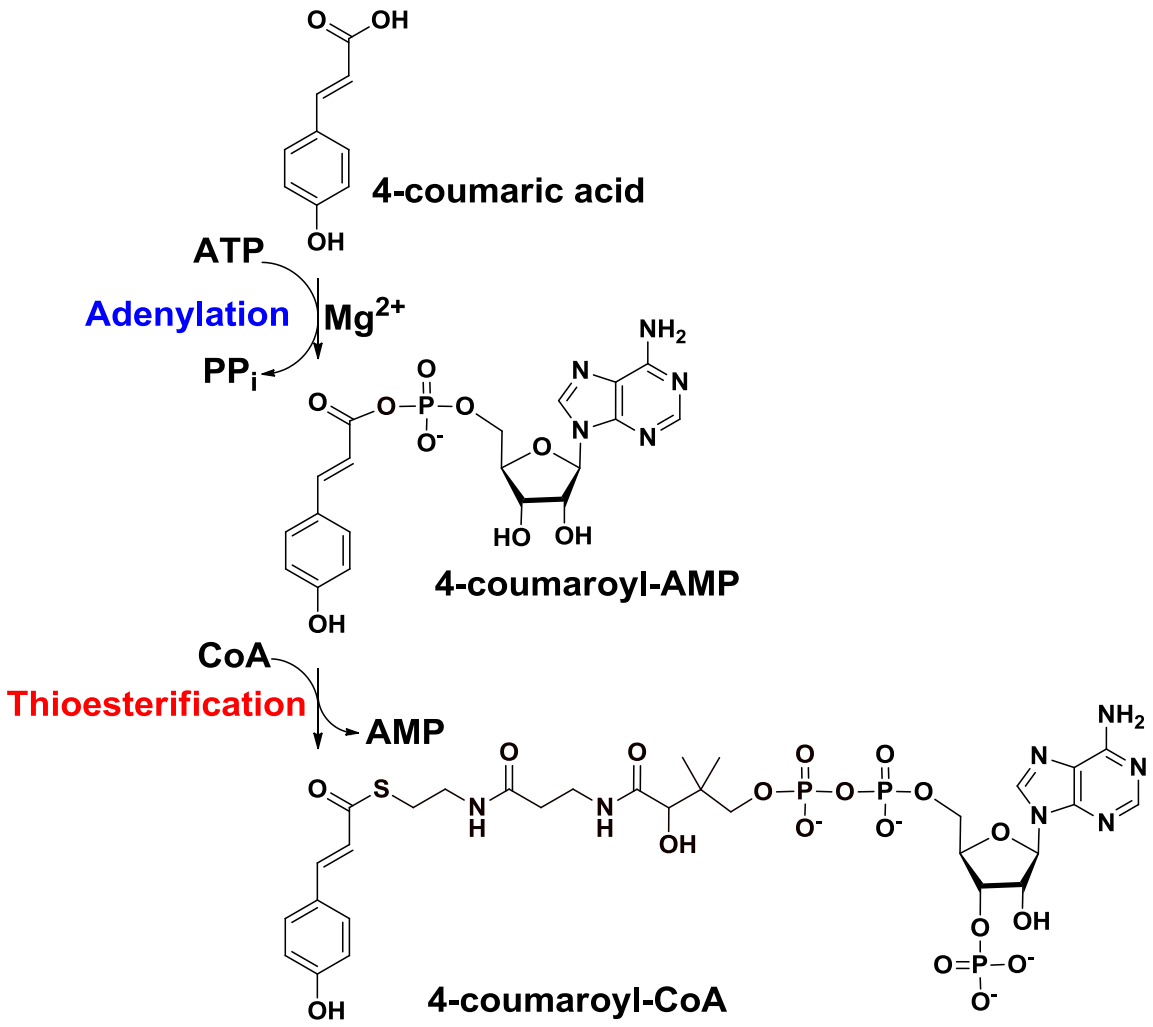


Figure 3-2. The two-step reaction catalyzed by 4CL enzyme.

Formation of CoA thioesters by 4CL occurs through a two-step reaction mechanism, involving the formation of a hydroxycinnamate-AMP anhydride in the presence of ATP and Mg^{2+} (adenylation step), followed by nucleophilic attack on the carbonyl carbon of the adenylate by the phosphopantetheine thiol of CoA (thioesterification step) to yield the product thioester (Cukovic et al., 2001) (Figure 3-2). 4CL belongs to the ANL superfamily of adenylating enzymes, which contains three subfamilies: **a**cyl- and aryl-CoA synthetases, the adenylating domains of **n**on-ribosomal peptide synthetases, and

firefly luciferases (Gulick, 2009). Although they catalyze diverse overall reactions, enzymes in the ANL superfamily all function through a two-step reaction scheme, in which the first adenylation step is conserved (Figure 3-3). While these enzymes are structurally similar, their amino acid sequence similarity is limited to typically less than 20% identity.

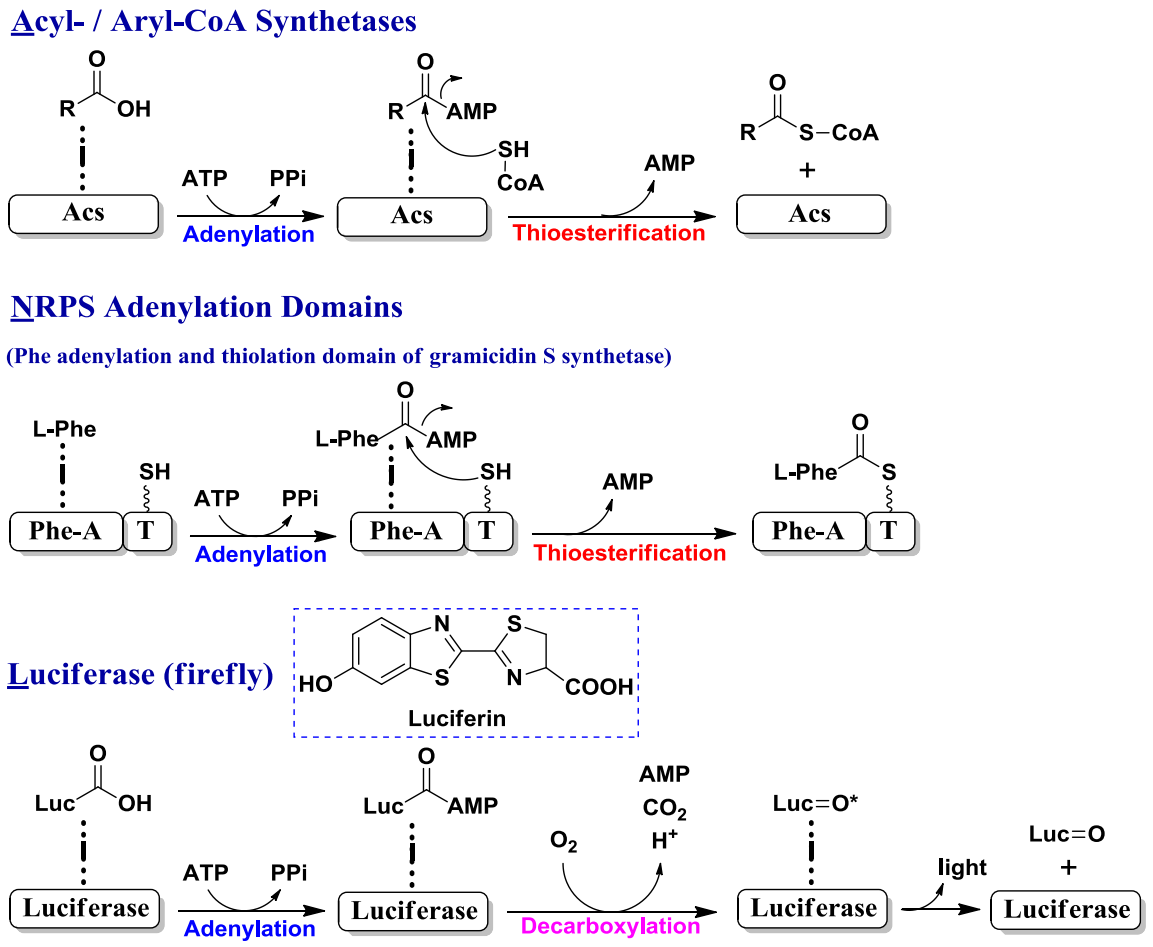


Figure 3-3. General reaction scheme for the ANL superfamily of adenylating enzymes.

Biochemical and structural biological studies demonstrate that members of the ANL superfamily undergo large-scale domain movement in order to facilitate the two catalytic partial reactions (McElroy et al., 1967; Bar-Tana and Rose, 1968; Bandarian et al., 2002; Gulick, 2009). These conformational changes are induced by the binding of the respective substrates for each of the reactions steps, and ANL superfamily members are proposed to adopt an adenylate-forming conformation during the first half-reaction, and a

thioester-forming conformation during the second step. Only two members in this superfamily have been structually characterized in both conformations, including 4-chlorobenzoyl-CoA ligase from *Alcaligenes sp. AL3007* (Gulick et al., 2004; Reger et al., 2008) and human medium-chain acyl-coenzyme A synthetase ACSM2A (Kochan et al., 2009), limiting thorough structure-function analysis. In addition, structural data of any ANL superfamily enzyme bound to unhydrolyzed ATP is restricted due to either partial occupancy of the bound ligand or weak electron density in highly conserved residues that are critical for ligand binding . Hence, the lack of structural data for an ANL superfamily enzyme in all states corresponding to its reaction coordinate restricts protein engineering experiments aimed at expanding the substrate scope of these ubiquitous catalysts.

Here we report the structural, and biochemical studies of isoform 2 of *N. tabacum* 4CL (Nt4CL2), along with further kinetic characterization of site-specific mutants identified by the structural data. We present crystal structures of Nt4CL2 in complex with Mg^{2+} and ATP (Nt4CL2-Mg^{2+}-ATP, 2.3 Å), with AMP and CoA (Nt4CL2-AMP-CoA, 1.6 Å), and with different adenylate intermediates, including the anhydrides of 4-coumaroyl-AMP (Nt4CL2-CMA, 1.6 Å), caffeoyl-AMP (Nt4CL2-CFA, 1.8 Å), and feruloyl-AMP (Nt4CL2-FRA, 1.7 Å). Lastly, we have generated a deletion variant at Val341 (ΔVal341) that expands the catalytic repertoire of 4CL to accommodate sinapinic acid as a substrate, and rationalize this gain-of function mutation through structural analysis of ΔVal341 in complex with AMP and CoA (ΔV341-AMP-CoA, 1.8 Å). These structures track each of the different conformations that are adopted by an enzyme from the ANL superfamily along the reaction coordinate.

Experimental procedures

Protein Purification

The expression clone of Nt4CL2 (pCRT7/CT-TOPO) (Beuerle and Pichersky, 2002) was a kind gift from Dr. Beuerle. It provides constitutive expression of Nt4CL2. A pre-culture of *Escherichia coli* Rosetta (DE3) transformed with the expression clone was grown overnight at 37 ^{o}C in Luria-Bertani (LB) medium with ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) and used to inoculate (0.5%, v/v) 1 liter of LB medium with

the antibiotics. The resultant culture was grown at 37 ^{o}C for 20-24 hours. Cells were harvested by centrifugation at 3,500g for 25 minutes at 4 ^{o}C. Cell pellet was resuspended in 20 mM Tris, pH 8.0, 500 mM NaCl, 30 mM imidazole, and 10% (v/v) glycerol, flash-frozen in liquid nitrogen, and stored at -80 ^{o}C until further use.

Cells were disrupted by four passes through an Avestin C5 Emulsiflex French Press and insoluble aggregates and cellular debris were removed by centrifugation at 15, 000 rpm for 1 hour. The clarified supernatant was applied to a 5ml HisTrap HP column (GE Healthcare) that was charged with nickel sulfate and pre-equilibrated with 20 mM Tris, pH 8.0, 1 M NaCl, 30 mM imidazole. The column was rinsed with 10 column volume of the same buffer and the protein was then eluted with a linear gradient of imidazole from 30 mM to 200 mM over 60 ml. Fractions with >90% pure Nt4CL2 protein were pooled and concentrated to 4 ml, which was further purified by size-exclusion chromatography (Superdex 75 16/60, GE Healthcare) in 20 mM HEPES pH7.5, 100 mM KCl. Protein was concentrated to 50-60 mg/ml and aliquoted. The aliquots were flash-frozen in liquid nitrogen, and stored at -80 ^{o}C. Protein concentration was measured by UV absorbance at 280 nm using the theoretical extinction coefficient of 31.5 mM^{-1} cm^{-1}. The final yield of Nt4CL2 was approximately 5 mg/L.

Mutagenesis

Nt4CL2 mutants were generated by Quickchange (Stratagene) site-directed mutagenesis using the aforementioned expression clone as template. The procedures were according to the manufacturer, except that mutagenesis PCR reactions were performed using Phusion Polymerase (Finnzymes). All the mutant proteins were purified following the same protocol described above for wild type Nt4CL2.

Enzyme Kinetics of Wild Type and Mutant Nt4CL2

The enzymatic kinetics of wild type and mutant Nt4CL2 were measured spectrophotometrically as described (Stuible et al., 2000) with a few modifications. Briefly, each 200 μl reaction contained 100 mM Tris pH 7.5, 2.5 mM $MgCl_2$, 2.5 mM ATP, 0.2 mM CoA, and various concentrations of corresponding hydroxycinnamate

substrates (4-coumaric acid, caffeic acid, ferulic acid, or sinapinic acid). Enzyme concentration was typically 15 nM unless a higher concentration was needed for some low activity mutants. The reaction was started by the addition of ATP and monitored using Varian Cary 4000 UV-Vis spectrophotometer at 25 ^{o}C. The formation of the thioester products were followed at the wavelengths of 333, 345, 346, and 352 nm for 4-coumaroyl-CoA, caffeoyl-CoA, feruloyl-CoA, and sinapoyl-CoA, respectively (Beuerle and Pichersky, 2002), and the concentration of these products were calculated using the corresponding extinction coefficients of 21 $mM^{-1}cm^{-1}$, 18 $mM^{-1}cm^{-1}$, 19 $mM^{-1}cm^{-1}$, and 20 $mM^{-1}cm^{-1}$, respectively (Obel and Scheller, 2000; Yamauchi et al., 2003). K_m and V_{max} were calculated by non-linear regression method using OriginPro 8.5 by OriginLab. The final values were averaged from three replicas.

LC-MS Analysis of the Thioester Products

Overnight reactions were set up at room temperature with wild type Nt4CL2 or the ΔV341 mutant and appropriate substrates under following concentrations: 100 mM Tris pH 7.5, 2.5 mM $MgCl_2$, 2.5 mM ATP, 0.2 mM CoA, 0.2 mM of corresponding hydroxycinnamic substrates (4-coumaric acid, caffeic acid, ferulic acid, or sinapinic acid), and 0.2 μM of enzyme. 10 μl of each reaction was subjected to positive/negative-ESI LC/MS using Agilent 1100 Series HPLC with a Phenomenex Luna C18 (2) (5μm, 100 Å, 4.6 mm × 150 mm) column coupled to Agilent LC/MSD Ion Trap XCT Plus mass spectrometer. Mobile phases consisted of water (solvent A) and acetonitrile (solvent B). A liner gradient of 5-95% B over 18 minutes at a flow rate of 0.4 ml/min was used. LC traces were recorded at 260 nm (adenine) and also the corresponding wavelengths for the thioester products.

To analyze the adenylate intermediate 4-coumaroyl-AMP, reactions were set up under the following concentrations: 100 mM Tris pH 7.5, 2.5 mM $MgCl_2$, 2.5 mM ATP, 0.2 mM of 4-coumaric acid, and 0.2 μM of wild type Nt4CL2. The reactions were incubated at room temperature for indicated time and were then subjected to LC-MS analysis following the same protocol as described above, except that LC traces were recorded at 340 nm, at which the intermediate absorbs the maximum.

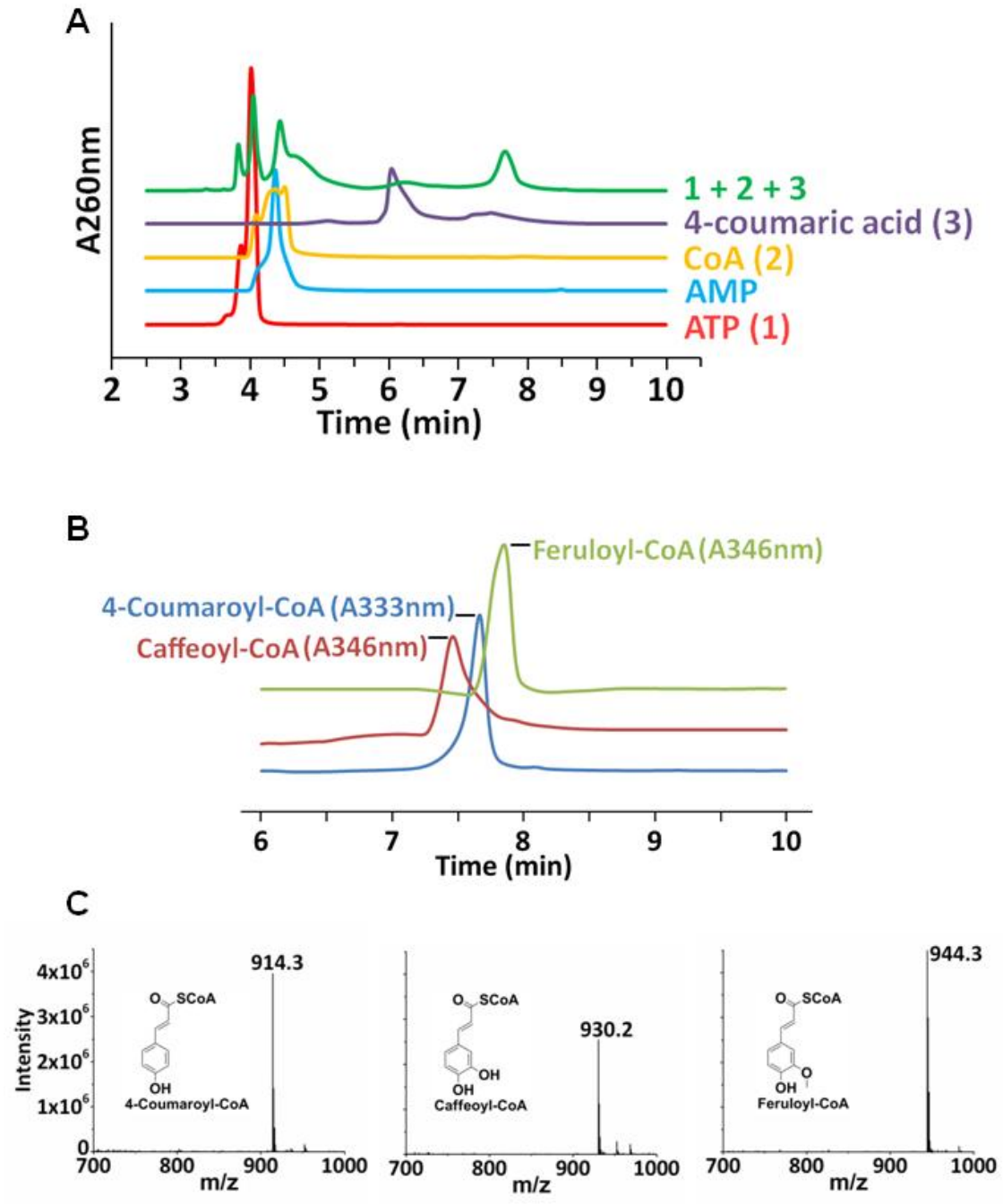


Figure 3-4. LC-MS analysis of the thioesters produced by wild type Nt4CL2. A) LC traces of reactants and 4-coumaroyl-CoA recorded at 260 nm; B) LC traces of 4-coumaroyl-CoA, caffeoyl-CoA, and feruloyl-CoA recorded at the indicated wavelengths; C) MS spectra of the thioester products.

Results

In vitro Kinetic Analysis of Nt4CL2 Activity

The enzymatic activity of wild type Nt4CL2 against various hydroxycinnamate substrates were assayed spectrophotometrically as described (Beuerle and Pichersky, 2002) with a few modifications. For each reaction, progress was monitored by measuring formation of the final product thioester, which is a two-step process. Consequently, the values reported here correspond to apparent V_{max}, rather than k_{cat}. Kinetic parameters were determined for Nt4CL2 using 4-coumaric acid, caffeic acid, and ferulic acid as substrates (Table 3-1). The formation of each thioester product was further confirmed by end-point analysis using reversed-phase liquid chromatography coupled with mass-spectrometry (LC-MS) (Figure 3-4). Nt4CL2 showed the highest activity against 4-coumaric acid (V_{max}/K_m = 44.2), and the lowest activity against ferulic acid (V_{max}/K_m = 20.7), which has a *meta*-methoxy substituent.

Table 3-1. Kinetic properties of Nt4CL2 against different hydroxycinnamic substrates.

	K_m μM	V_{max} nkat/mg	V_{max}/K_m *
4-Coumaric acid	1.5 ±0.4	66.3 ±4.2	44.2
Caffeic acid	1.0 ±0.3	44.2 ±2.7	44.2
Ferulic acid	2.9 ±0.4	59.9 ±2.2	20.7
Sinapinic acid	1321.9 ±103.9	0.3 ±0.0	0.0002

nkat/mg: nmole/mg Enzyme · sec
*Unit: L/g Enzyme · sec

Two Conformations Shown by Nt4CL2 Structures

The overall architecture of Nt4CL2 is similar to those of other ANL superfamily enzymes and consists of a large N-terminal domain (Val8 to Asp434) and a smaller C-terminal domain (Lys441 to Ala537) that are connected by a highly flexible linker consisting of residues Arg435 through Ile440 (the "hinge loop"). Two different conformations for Nt4CL2 are observed in the various co-crystal structures, consisting of an adenylate-forming and a thioester-forming conformation (Figure 3-5). The two conformations are related by a rigid body movement, consisting of a ~140-degree rotation of the C-terminal domain along the hinge loop. Despite the drastic rotational movement, there are minimal

changes in the isolated N- or C-terminal domains and the individual domains from each of the two conformations can be aligned with an RMSD of less than 0.9 Å for the N-terminal and 0.6 Å for the C-terminal domains.

The structure of Nt4CL2 in the adenylate-forming conformation (in complex with Mg^{2+} and ATP) is most similar to that of *Luciola cruciata* (Japanese firefly) luciferase in complex with AMP (PDB code: 2D1Q; sequence identity: 34%; RMSD = 2.0 over 514 aligned C_α atoms) (Nakatsu et al., 2006). The thioester-forming conformation of Nt4CL2 is most similar to the structure of *P. tomentosa* 4CL1 (Pt4CL1) complexed with an adenylate intermediate analog (PDB code: 3NI2; sequence identity: 77%; RMSD = 1.1 over 528 aligned C_α atoms) (Hu et al., 2010). Our structures of Nt4CL2 append to the recently reported structures of *Alcaligenes sp.* AL3007 4-chlorobenzoate-CoA ligase (CBL) (Gulick et al., 2004; Reger et al., 2008) and human medium-chain acyl-coenzyme A synthetase ACSM2A (Kochan et al., 2009) as one of three ANL superfamily enzymes with crystal structures that have been captured in both conformations.

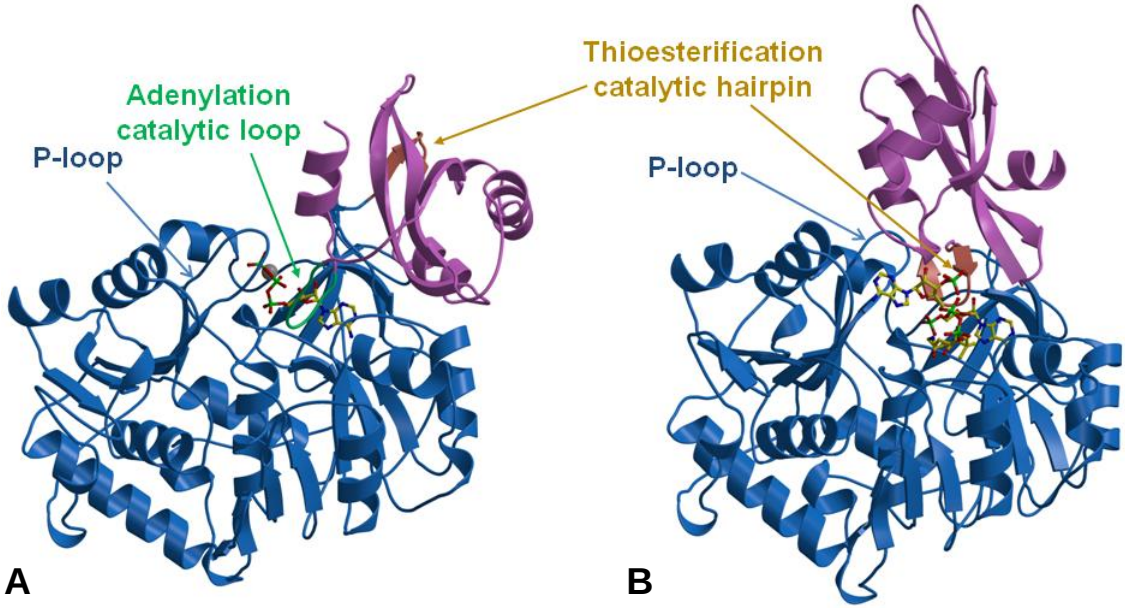


Figure 3-5. Two overall conformations of Nt4CL2. A) Adenylate-forming conformation assumed by Nt4CL2-Mg^{2+}-ATP complex structure. Mg^{2+} is shown as gray sphere and ATP is shown as yellow sticks. B) Thioester-forming conformation assumed by Nt4CL2-AMP-CoA complex structure. AMP and CoA are shown as yellow sticks. In both

conformations, N-terminal domain is shown in blue and C-terminal domain is shown in purple. Key secondary structures are indicated by arrows.

The Adenylation Conformation Reveals an Ordered P-loop

In contrast to previously reported ATP-bound structures of enzymes in the ANL superfamily (Kochan et al., 2009; Osman et al., 2009), our Nt4CL2-Mg^{2+}-ATP complex structure displays clear and unambiguous electron density for ATP, Mg^{2+}, as well as all the ATP-interacting residues, especially those within the pyrophosphate-binding loop (P-loop) (Figure 3-6). The ATP molecule is located at the interface of the N- and C- terminal domains, with the AMP portion lying on the top of the N-terminal domain and the pyrophosphate group pointing upwards. The adenosine is enclosed by a highly conserved region containing Gln331-Glu337, which is designated as the core motif A5 in the ANL superfamily (Marahiel et al., 1997) (Figure 3-7), and a divergent loop containing Ser307-Leu312. The α-phosphate group is coordinated by the side chains of His237 (core motif A4), Thr336 (core motif A5) , and Lys526 (core motif A10), and the main chain nitrogens of Ser190 (core motif A3) and Thr336. Notably, Lys526 is the only ATP-interacting residue that is contributed from the C-terminal domain (Figure 3-6). This Lys is a universally conserved residue in the core motif A10 (Figure 3-7), or the adenylation catalytic loop, and has been shown to be essential for the adenylation reaction (Gocht and Marahiel, 1994; Branchini et al., 2000; Horswill and Escalante-Semerena, 2002; Reger et al., 2007) but dispensible for the second half reaction (Branchini et al., 2000; Horswill and Escalante-Semerena, 2002). Its close vicinity with the α-phosphate and the ribose group in the Nt4CL2-Mg^{2+}-ATP structure is consistent with its important catalytic role and confirms that the structure represents the adenylate-forming conformation. As expected, the Lys526→Ala mutation abolished activity (3-8). In addition, mutations of other α-phosphate-binding residues, including His237→Ala and Thr336→Ala, all resulted in impaired catalysis (3-8).

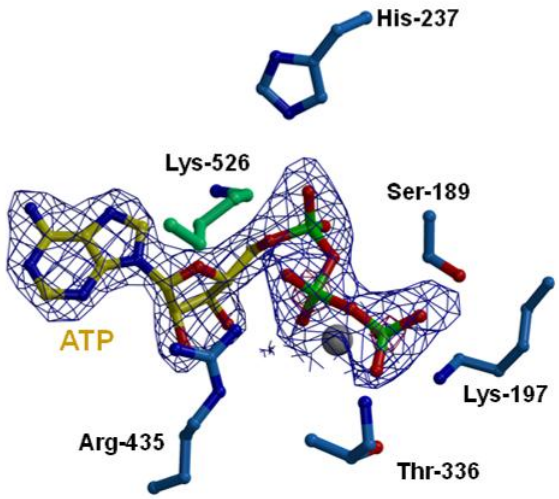


Figure 3-6. ATP binding in Nt4CL2-Mg^{2+}-ATP complex structure. Electron density map of ATP is shown as blue mesh. Mg^{2+} is shown as gray sphere.

The pyrophosphate moeity is mainly wrapped by the so-called P-loop containing Ser189-Lys197 (Figures 3-5A and 3-6). This glycine- and serine-/threonine-rich loop belongs to the conserved core motif A3, which is believed to play an important role in orientating the pyrophosphate during the adenylation reaction (Chang et al., 1997; Stuible et al., 2000; Horswill and Escalante-Semerena, 2002), in a manner reminiscent of the common Walker A motif in many ATP-/GTP-binding proteins (Saraste et al., 1990). In all prior structures of ANL enzymes in the adenylation conformation, the P-loop is partially disordered (Conti et al., 1996; Conti et al., 1997; May et al., 2002; Gulick et al., 2004; Du et al., 2008; Osman et al., 2009), implying that this loop is highly mobile. In contrast, in our Nt4CL2- Mg^{2+}-ATP structure, clear and continuous electron density can be observed for all residues within this loop (Figure 3-6). Residues within the P-loop form extensive hydrogen-bonds with the ATP pyrophosphate through the side chains of Ser189, Thr192, Thr193, and Lys197, and the main chain of Gly191. Mutations at residues within this region, such as Thr193→Ala and Lys197→Ala, reduced the catalytic efficiency (V_{max}/K_m) by five- to eight-fold (3-8), confirming the importance of this P-loop in the enzyme activity.

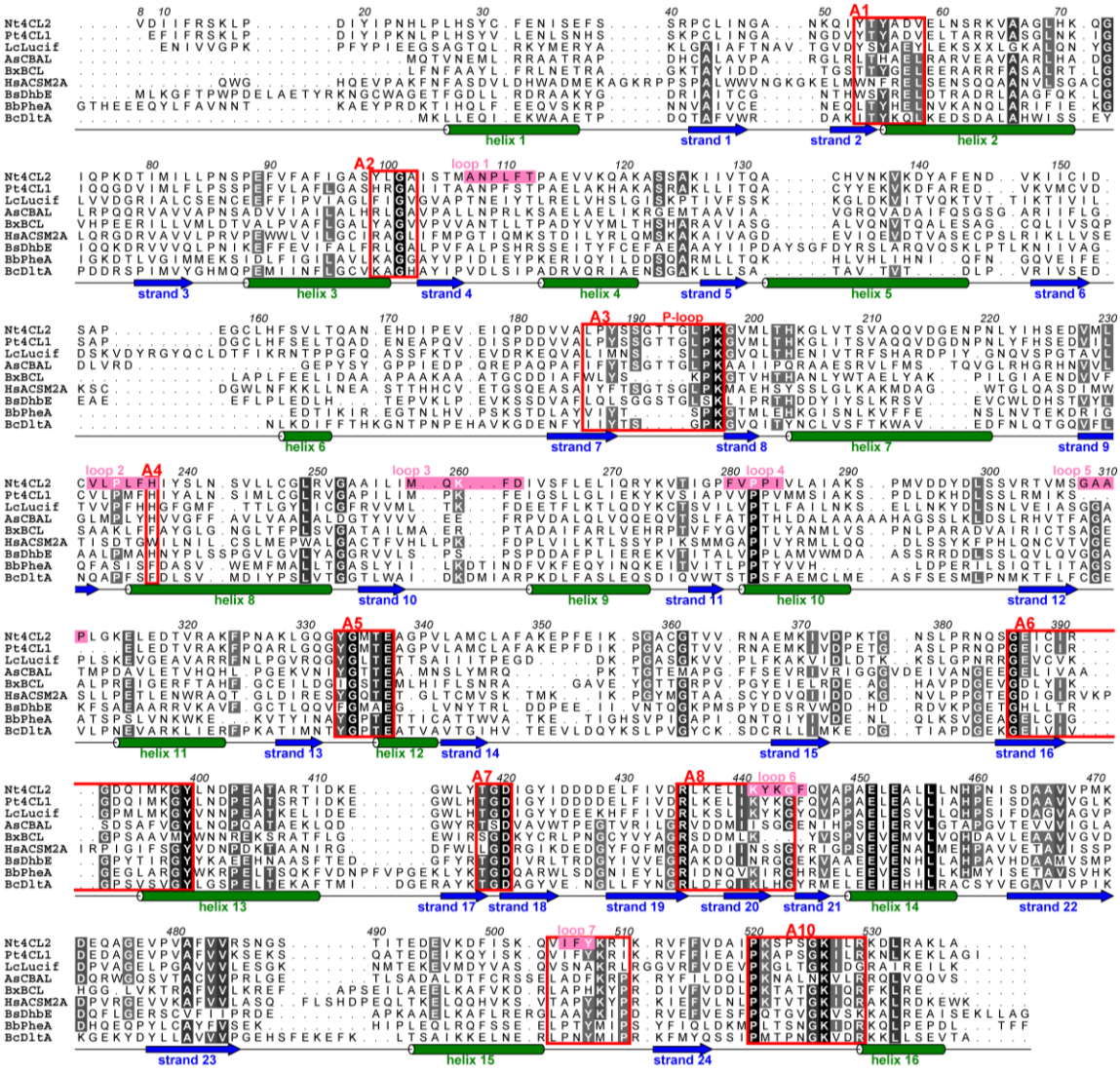


Figure 3-7. Structure-based sequence alignment of enzymes from the ANL superfamily. Secondary structures assignment is based on Nt4CL2 structure and shown below the sequences. Conserved core motifs (A1-A10) are shown in red frames. Residues that form the hydrophobic CoA-binding tunnel (loop1-loop7) are highlighted in pink. Nt: *Nicotiana tobacum*; Pt: *Populus tomentosa*; Lc: *Luciola cruciata*; As: *Alcaligenes sp.* AL3007; Bx: *Burkholderia xenovorans*; Hs: *Homo sapiens*; Bs: *Bacillus subtilis*; Bb: *Brevibacillus brevis*; Bc: *Bacillus cereus*.

Thioester-forming Conformation Results in Drastic Movements of the Catalytic Residues and the P-loop

Other than the Nt4CL2-Mg^{2+}-ATP complex structure described above, all other Nt4CL2 co-crystal structures reside in the thioester-forming conformation. As our Nt4CL2 structures represent a rare example of an ANL enzyme that is structurally characterized in both conformations, these structures provide an ideal model for tracking the differences between these two states. For the sake of simplicity, discussions of the overall architecture of the thioester-forming conformation will be based on the Nt4CL2-AMP-CoA structure (Figure 3-5).

As noted previously, the thioester-forming conformation is characterized by a 140-degree rotation of the C-terminal domain along the hinge region. This domain movement results in two major differences in the nucleotide-binding pocket. First, the core motif A10 (harboring the essential Lys526) moves 20Å away from the active site and becomes completely exposed to solvent (Figures 3-5 and 3-8). Core motif A8 (Arg435 - Gly444), or the thioesterification catalytic hairpin, replaces A10 to form the lid of the active site, resulting in an entirely new set of residues that stabilize the bound nucleotide. Specific interactions with the ribose are mediated through Arg435, Lys437, and Lys441, and Lys441 and Gln446 engage the α-phosphate. The C-terminal domain is stabilized in the thioester-forming conformation mainly through hydrogen bonding interactions between Arg435 and the backbone oxygen of Leu439 and carboxylate oxygens of Glu447. Consequently, the Arg435→Ala mutation reduced the catalysis efficiency by 500-fold (Table 3-2).

Table 3-2. Kinetic properties of Nt4CL2 variants against 4-coumaric acid.

	K_m μM	V_{max} nkat/mg	V_{max}/K_m *
wt	1.5 ±0.4	66.3 ±4.2	44.2
T193A	2.9 ±0.3	16.6 ±0.6	5.7
K197A	13.9 ±1.8	118.5 ±4.2	8.5
H237A	143.4 ±17.8	33.0 ±1.3	0.2
T336A	5.1 ±1.0	4.7 ±0.2	0.9
M344A	520.7 ±91.8	152.4 ±17.4	0.3
R435A	614.9 ±76.1	50.6 ±2.8	0.1
K441A	No conversion		
K443A	0.9 ±0.2	24.5 ±1.6	27.8
K526A	No conversion		

nkat/mg: nmole/mg Enzyme · sec
*Unit: L/g Enzyme · sec

As a consequence of the domain movement, a highly conserved Lys residue (Lys441) in the motif A8, or the thioesterification catalytic hairpin, is now within close distance (2.9 Å – 3.3 Å) to the oxygen atoms of the ribose and α phosphate (Figure 3-9), implying a crucial role for this residue in the second half reaction. Interestingly, both the position and interactions of Lys441 are reminiscent of those of the invariant catalytic residue Lys526 in the adenylate-forming conformation. A Lys441→Ala mutation completely abolished enzyme activity (Table 3-2).

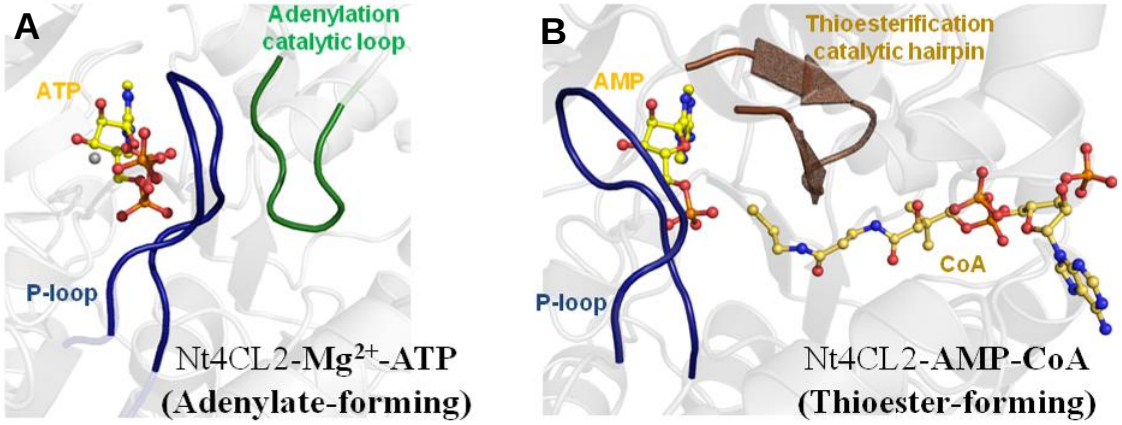


Figure 3-8. Domain alternation leads to the movement of the P-loop and the catalytic residues. A) Nt4CL2-Mg^{2+}-ATP complex structure (adenylate-forming conformation); B) Nt4CL2-AMP-CoA complex structure (thioester-forming conformation).

The second major difference between the two conformations in the nucleotide-binding pocket is the displacement of the P-loop by the thioesterification catalytic hairpin (Figures 3-5 and 3-8). Consequently, the P-loop bends upwards and adopts a more closed structure. This movement is the only significant difference between the two conformations at the N-terminal domain, and displacement of this P-loop by the thioesterification catalytic hairpin facilitates the shift to the thioester-forming conformation. The side chains of three residues within this hairpin, Leu439, Lys441, and Gln446, insert into the pyrophosphate-binding site. As this insertion can only occur following pyrophosphate removal, the dissociation of the pyrophosphate group is absolutely required for the transition from the adenylate-forming to the thioester-forming conformation.

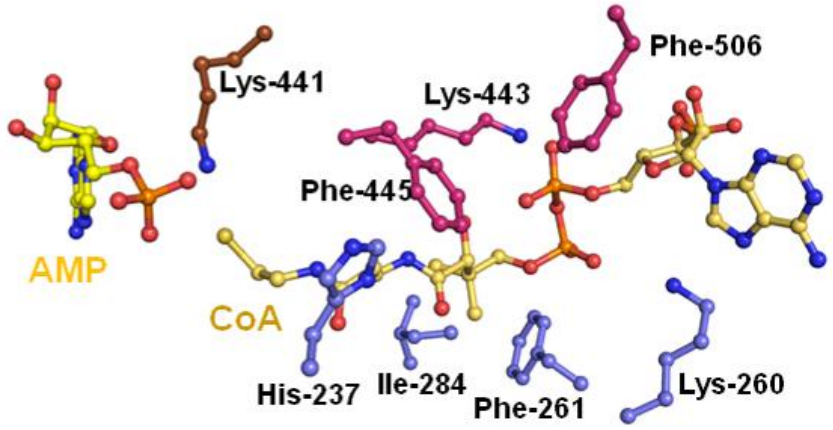


Figure 3-9. Residues interacting with CoA. N-terminal residues are shown in blue, C-terminal residues are shown in red, except that Lys-441 is shown in brown.

The CoA Binding Tunnel is Formed by Seven Loops

Structures of ANL enzymes with bound CoA are relatively rare and have not been reported for any 4CL proteins prior to this study. In the Nt4CL2-AMP-CoA structure, the CoA molecule spans a large portion of the protein between the N- and C-terminal domains, with the thiol group pointing into the active site. The sulfur atom is situated 3.6 Å away from an α-phosphate oxygen, where it would be poised for nucleophilic attack onto the anhydride intermediate (Figure 3-9). Clear and continuous electron density,

corresponding to the bound CoA ligand can be envisioned within a tunnel formed by seven loops (five from the N-terminal and two form the C-terminal domains) (Figure 3-7). The adenosine head group of CoA is mostly exposed to solvent and the pantetheine group is accommodated in the tunnel through van der Waals contacts (Figure 3-9). The only hydrogen bonds with CoA molecule are formed between the diphosphate group and residues Lys260 and Lys443 (Figure 3-9). These hydrogen bonds are likely not crucial for enzyme activity, as the Lys443→Ala mutation has a minimal effect on enzyme efficiency (Table 3-2).

A comparison between the adenylate-forming (Nt4CL2-Mg^{2+}-ATP) and thioester-forming (Nt4CL2-AMP-CoA) conformation reveals that the C-terminal domain rotation facilitates CoA binding in two ways. First, two loops that are distal to the CoA-binding site in the adenylation conformation are brought to the upper side of the tunnel upon domain rotation (Figure 3-9). Second, motif A10, which is inserted deeply into the CoA tunnel in the adenylation conformation (Figure 3-8), is displaced upon domain rotation. As a result of both of these changes, CoA binding can only occur in the thioester-forming conformation.

Nt4CL2 Forms a Stable High Energy Anhydride Complex

During attempts to characterize the two half-reactions using analytical chromatographic methods, recombinant wild-type Nt4CL42 was incubated with ATP and various hydroxycinnamates. To our surprise, the resulting (adenylated) acid anhydride products were stable for several days at room temperature in the absence of CoA (Figure 3-10). Upon addition of stoichiometric concentrations of CoA, the anhydride instantly decomposed into AMP and the corresponding hydroxycinnamoyl-CoA (Figure 3-10).

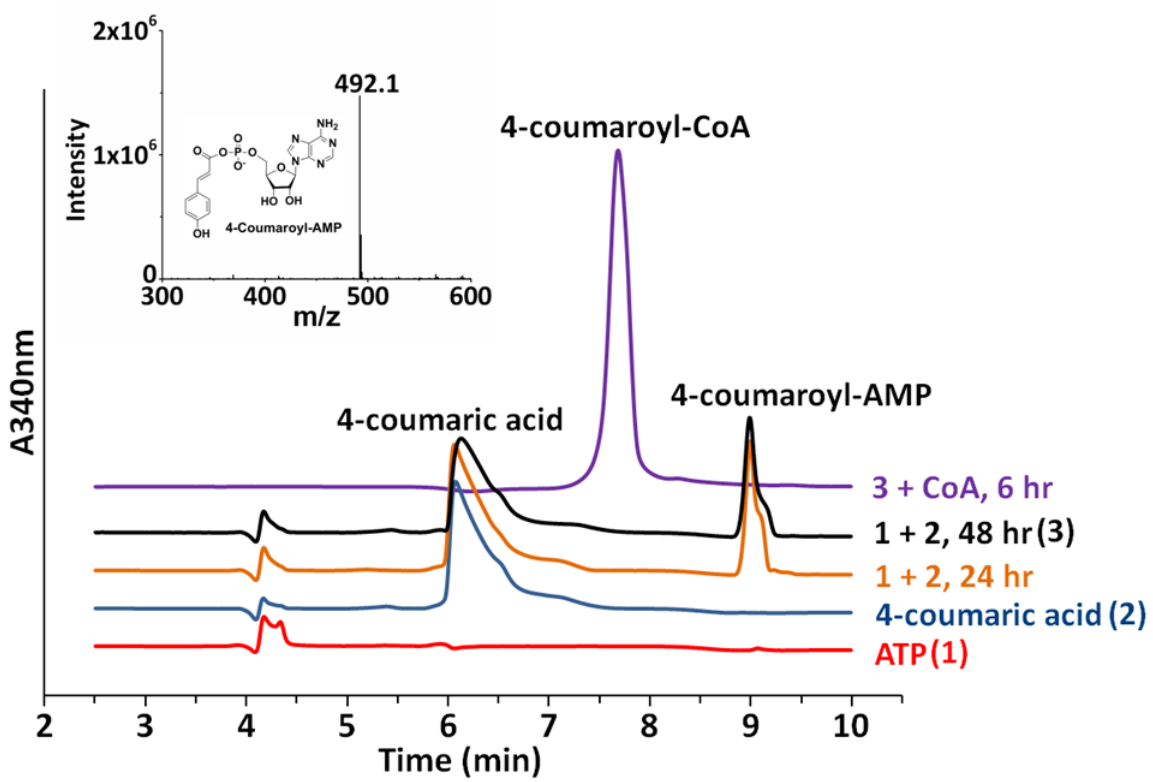


Figure 3-10. LC-MS analysis of the adenylate intermediate produced by Nt4CL2. The enzyme was incubated with the indicated reagents at room temperature. The MS result of 4-coumaroyl-CoA is shown in the insert.

Inspired by these observations, we carried out co-crystallization of Nt4CL2 in the presence of ATP and 4-coumaric, caffeic, and ferulic acids. The resultant structures all occupy the thioester-forming conformation and each reveals unambiguous electron density to the respective acid anhydride within the active site (Figure 3-11). One significant difference with the Nt4CL2-AMP-CoA structure is in the conformation of His237 (Figure 3-12), which composes the one-residue core motif A4 (Figure 3-7). In presence of CoA the side chain of this residue moves away from the active site to stabilize the phosphopantetheine of CoA (Figure 3-12B). Similar movement of the equivalent His has been observed in all CoA- or thioester product-bound structures of ANL enzymes (Gulick et al., 2003; Reger et al., 2007; Reger et al., 2008; Kochan et al., 2009; Hughes and Keatinge-Clay, 2011). The important role of His237 is supported by

our mutational studies that demonstrated that the His237→Ala mutant had nearly 800-fold decrease in catalytic efficiency relative to the wild-type enzyme (Table 3-2).

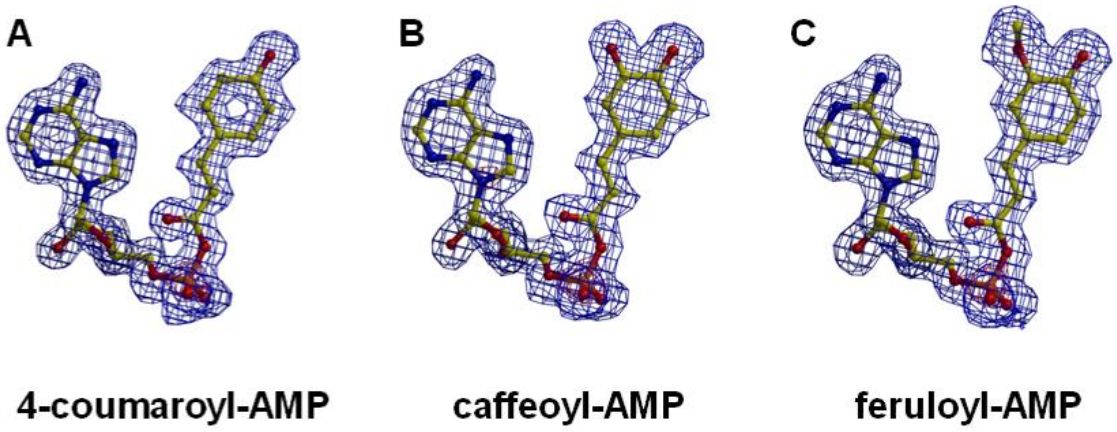


Figure 3-11. Electron density map of the bound adenylate-intermediates in Nt4CL2 structures.

In each acid anhydride complex structure, the hydroxycinnamate moiety is enclosed in a hydrophobic pocket composed of a five-stranded β-sheet and two α-helices (Figure 3-13A and B). The side chain of Ser243 provides the only hydrogen bond interactions with the hydroxycinnamate group through the 4-hydroxy group. Additionally, the aromatic side chain of Tyr239 engages in stacking interaction with the hydroxyphenyl ring (Figure 3-12A) and the main chains of Ile239 and Tyr239 are in van der Waals contact with the propanoate moiety. Mutating Tyr239 to Phe increases the K_m for the hydroxycinnamate substrates by less than ten-fold, whereas a Tyr239→Ala mutation results in a 100- to 250-fold increase of the K_m values (Table 3-2), indicating the importance of the stacking interaction in accommodating the substrate.

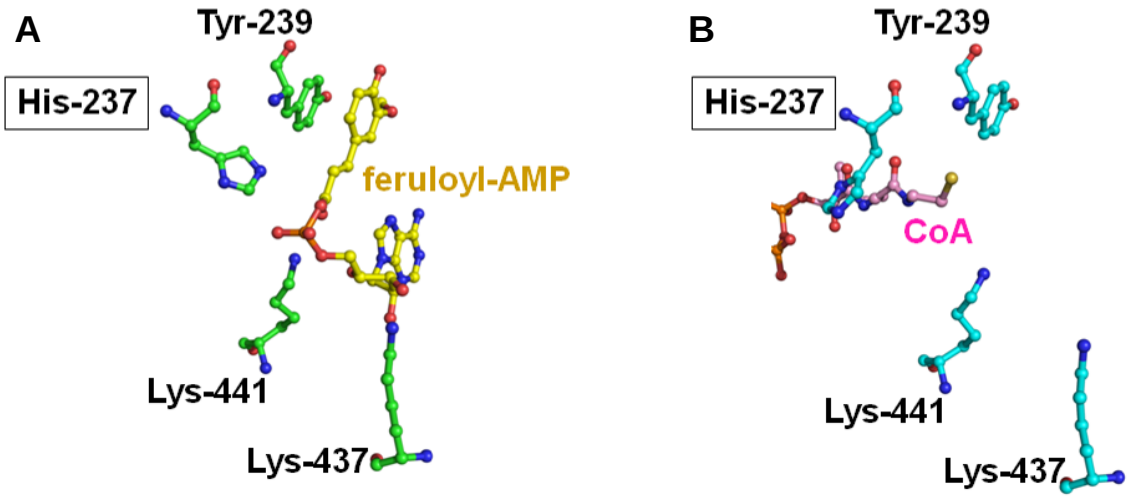


Figure 3-12. Comparison between Nt4CL2-feruloyl-AMP (A) and Nt4CL2-AMP-CoA (B) complex structures. For clarity, AMP is not shown in the latter structure.

A comparison of the Nt4CL2 acid anhydride complex structures reveals that the binding modes for the hydroxycinnamate moieties are similar. The main chain atoms of Met306 through Ala309 provide additional contacts with the ligand and define one side of the substrate ring-binding site. The substituents at C3 point to the side of Met306 and are not located in the vicinity of Val341 (Figure 3-13C), contrary to what was suggested by previous modeling studies (Hu et al., 2010). The α-carbon of Gly308 is only 3.9 Å to C2 of the hydroxyphenyl ring (Figure 3-13C), which indicates that the enzyme cannot tolerate substrates with substituents at C2. The bottom of the substrate-binding pocket is

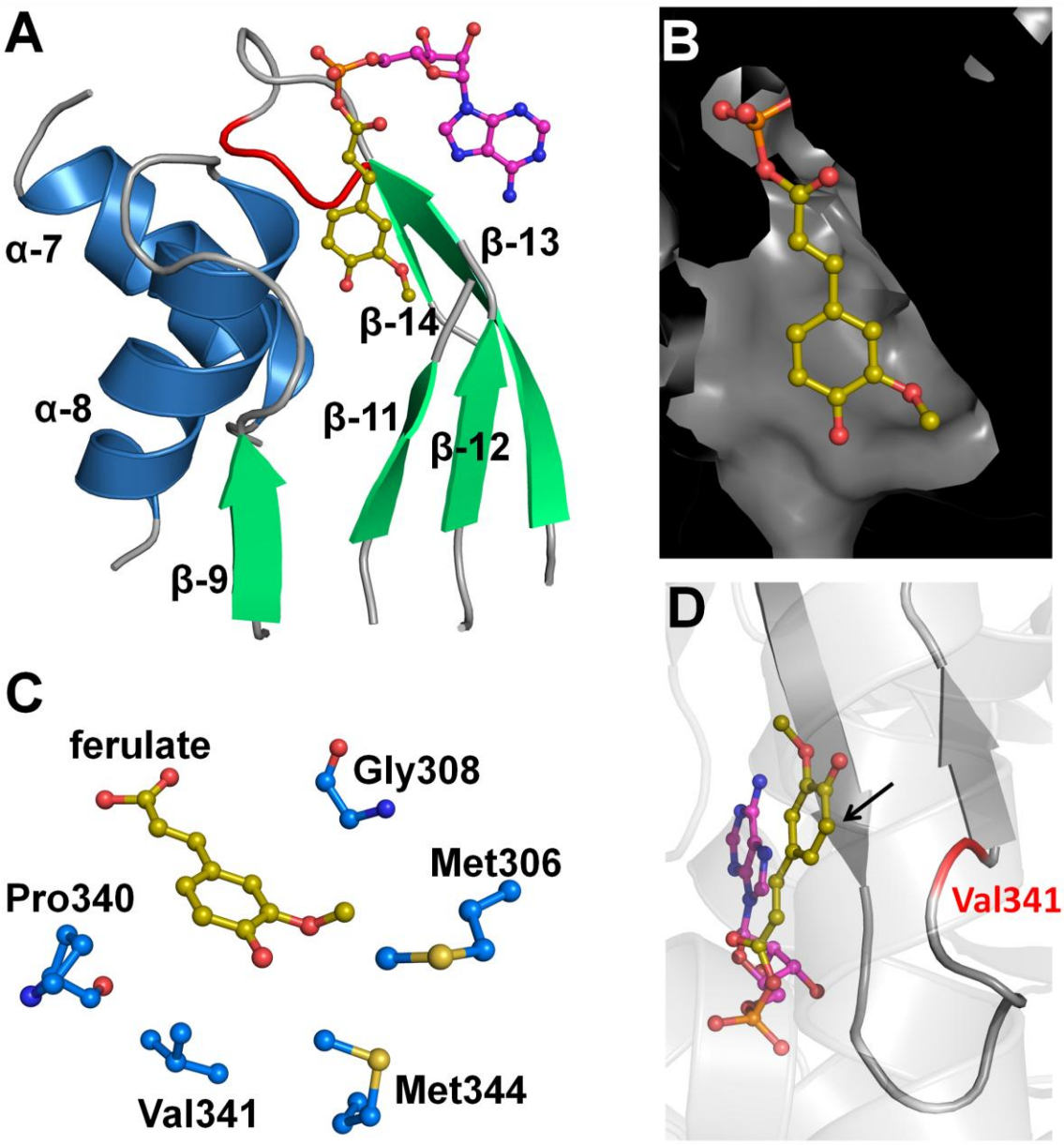


Figure 3-13. Hydroxycinnamate-binding in Nt4CL2-feruloyl-AMP complex structure. A) Hydroxycinnamate-binding pocket is composed of two α-helices (blue) and a five-strand β-sheet (green). The bulged loop formed by residues Gly399-Leu342 is shown in red. Numbering of the secondary structures is based on **Figure 3-7**. Feruloyl-AMP is shown as sphere-stick representation. The ferulate group is colored in yellow and the adenylate group is shown in magenta. B) Surface representation of the hydroxycinnamate-binding pocket. C) Residues in the hydroxycinnamate-binding pocket. D) The bulged loop formed by residues Gly399-Leu342 protrudes towards position 5 on the hydroxycinnamate ring, as indicated by the arrow. Val341 is shown in red.

formed by two anti-parallel β-strands, and the side chain of Met344 (located on the second strand) points towards the 4-hydroxyl group (3.5 Å distance) (Figure 3-13C). The Met344→Ala mutation dramatically increases the K_m for all three substrates, indicating the importance of Met344 in accommodating substrate (Table 3-2). Additionally, four residues in this region (Gly339-Leu342) protrude towards the C2' and C3' of the substrate (Figure 3-13A, C, and D) and presumably accounts for the inability to accommodate substrates with substitutions at these positions (such as sinapinic acid).

A comparison between all our Nt4CL2 structures showed that most residues of the hydroxycinnamate binding pocket are in the same conformation in our Nt4CL2 structures, indicating that the large domain rearrangement and the hydroxycinnamate substrate binding have little impact on each other.

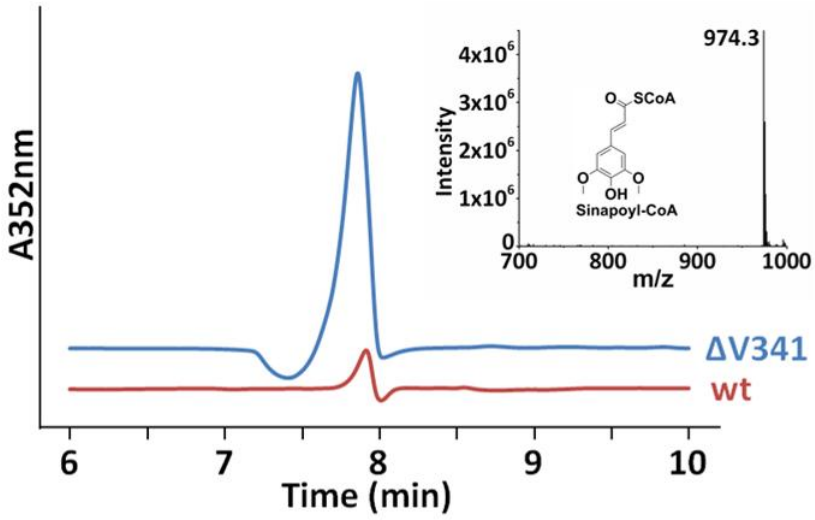


Figure 3-14. LC-MS analysis of sinapoyl-CoA formation catalyzed by Nt4CL2-ΔVal341. The MS result is shown in the insert.

Table 3-3. Kinetic properties of Nt4CL2 variants against different hydroxycinnamic substrates.

	4-Coumaric acid			Caffeic acid			Ferulic acid			Sinapinic acid		
	K_m μM	V_{max} nkat/mg	V_{max}/K_m *	K_m μM	V_{max} nkat/mg	V_{max}/K_m *	K_m μM	V_{max} nkat/mg	V_{max}/K_m *	K_m μM	V_{max} nkat/mg	V_{max}/K_m *
wt	1.5 ±0.4	66.3 ±4.2	44.2	1.0 ±0.3	44.2 ±2.7	44.2	2.9 ±0.4	59.9 ±2.2	20.7	1321.9 ±103.9	0.3 ±0.0	0.0002
Y239A	228.1 ±34.8	16.6 ±1.2	0.07	257.2 ±85.5	13.9 ±2.0	0.05	278.3 ±89.1	2.8 ±0.4	0.01	n.m.	n.m.	n.m.
Y239F	6.9 ±0.6	74.5 ±1.6	10.8	8.4 ±2.0	48.1 ±3.3	5.7	30.8 ±6.8	56.0 ±4.1	1.8	n.m.	n.m.	n.m.
V341G	4.3 ±2.2	42.9 ±4.3	10.0	4.3 ±0.9	33.9 ±1.5	7.9	4.7 ±0.8	41.3 ±1.5	8.8	1230.5 ±196.6	0.6 ±0.1	0.0005
dV341	13.7 ±2.3	40.1 ±1.6	2.9	15.4 ±4.9	21.2 ±1.6	1.4	31.1 ±2.1	38.1 ±0.7	1.2	312.8 ±24.9	12.9 ±0.5	0.04

nkat/mg: nmole/mg Enzyme · sec; **n.m.:** not measured
*Unit: L/g Enzyme · sec

70

An Engineered Nt4CL2 Variant That Can Utilize Sinapinic Acid as a Substrate

Most 4CLs are unable to utilize the *meta*-substituted sinapinic acid as a substrate (Becker-Andre et al., 1991; Lee and Douglas, 1996; Allina et al., 1998; Hu et al., 1998; Ehlting et al., 1999). The only known exceptions are the 4CL1 gene product from *G. max* and the 4CL4 protein from *A. thaliana*. Multiple sequence alignments of various 4CLs demonstrate that both of these paralogs have a single amino acid deletion that corresponds to either Val341 or Leu342 of Nt4CL2 (Figure 3-1). Moreover, deletion of the equivalent amino acid in canonical 4CLs expands the substrate scope of these enzymes to accommodate sinapinic acid as a substrate (Lindermayr et al., 2003; Schneider et al., 2003).

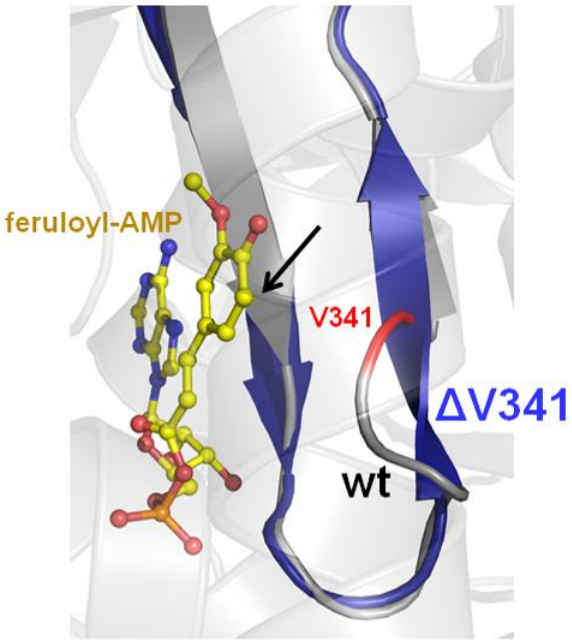


Figure 3-15. Structure comparison between wild type Nt4CL2 (gray) and ΔVal341 (blue). Val341 in the wild type structure is shown in red. Position 5 on the hydroxycinnamate ring of feruloyl-AMP (yellow) is indicated by the arrow.

In order to understand the rationale for the expanded substrate scope in the deletion variant, we generated Nt4CL2 Val341→Ala mutant and the single amino acid deletion variant Nt4CL2-ΔV341. Kinetic analysis of both variants demonstrates that only the Nt4CL2-ΔV341 deletion variant has appreciable activity with the sinapinic acid substrate (Table 3-3 and Figure 3-14). Hence, accommodation of *meta*-substituted substrates can

only be tolerated by alterations of the Nt4CL2 main chain (rather than the side chain) at residue 341. In addition, the crystal structure of this deletion variant in complex with AMP and CoA was determined to 1.8Å resolution (attempts at co-crystallization with sinapate substrate were unsuccessful, presumably due to relatively lower binding affinity) (Figure 3-15). Structural analysis of this variant reveals that deletion of Val341 significantly changes the topology of the hydroxycinnamate binding region. In particular, the region containing residues Gly339 to Leu341, which is one residue shorter in this variant, has flattened to form an extended β-strand (Figure 3-15). Consequently, the space near C3' of the substrate ring is expanded, and substrates bearing a methoxy substitution at this position can be accommodated at the active site. While the substrate scope of this deletion variant has been expanded to include sinapate, the mutation also eliminates some constrains on smaller substrates (coumarate, caffeate, and ferulate), resulting in an increase in the K_m values for these substrates by more than ten-fold (Table 3-3).

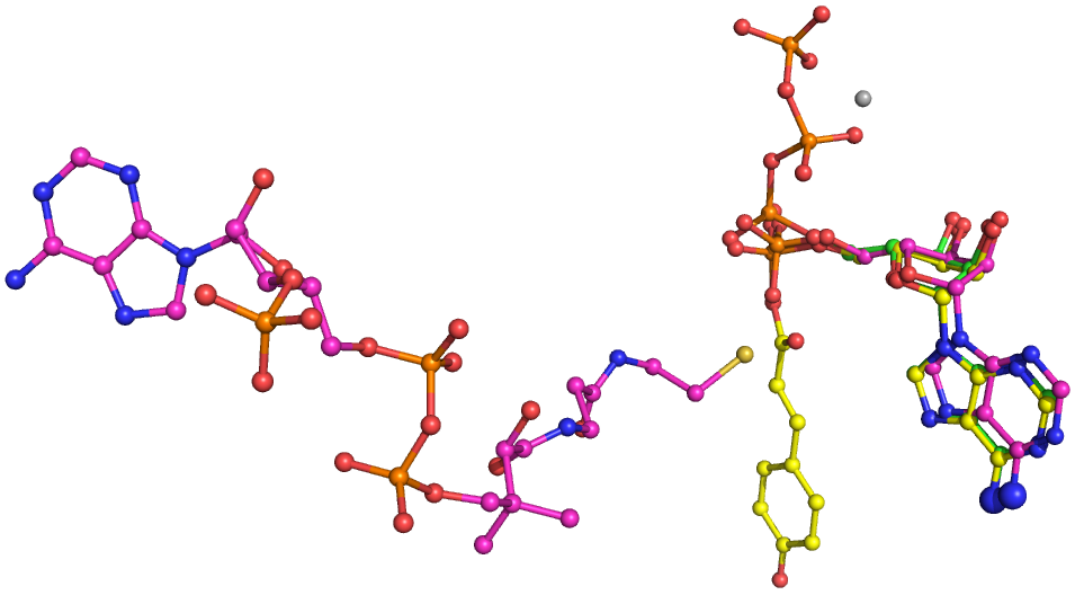


Figure 3-16. Orientation of the ligands bound with Nt4CL2. The structures of Nt4CL2-MgATP, -CMA, and -AMP-CoA are superimposed and only the ligands are shown. ATP is shown in green ball-stick model, 4-coumaroyl-AMP in yellow, and both AMP and CoA in magenta. Mg^{2+} is shown as a gray sphere.

Discussion

Domain Alternation

The several co-crystal structures of Nt4CL2 presented here establish four distinct substrate binding pockets at the active site that accommodate AMP, pyrophosphate, CoA, and hydroxycinnamate, respectively (Figure 3-16). The AMP- and hydroxycinnamate-binding pockets do not change appreciably between the adenylate- and thioester-forming conformations, suggesting that there is no obvious correlation between domain alternation and the binding of either AMP or the hydroxycinnamate. This notion is also borne out by the fact that previously reported structures of ANL enzymes with only AMP bound have been observed in both adenylate-forming conformation (Jogl and Tong, 2004; Nakatsu et al., 2006) and thioester-forming conformation (Yonus et al., 2008; Kochan et al., 2009; Hu et al., 2010).

At the active site, the pyrophosphate-binding pocket is delineated by a highly conserved P-loop that is dynamically disordered in most prior ANL enzyme structure. The only two reported ATP-bound structures of an ANL enzyme contains either a disordered P-loop (Osman et al., 2009) or partial occupancy of the pyrophosphate group (Kochan et al., 2009). In marked contrast, our Nt4CL2-Mg^{2+}-ATP co-crystal structure reveals clear and continuous electron density for both the P-loop and for all atoms of the bound nucleotide. In the adenylate-forming conformation, the P-loop displays an extended structure and wraps the pyrophosphate, whereas in the thioester-forming conformation, the P-loop bends upwards and occludes the pyrophosphate-binding pocket. This movement of the P-loop is a direct consequence of the C-terminal domain rotation in the thioester-forming conformation, suggesting that the transition from the adenylate-forming conformation must be preceded by pyrophosphate dissociation. Notably, all reported structures of ANL enzymes with bound ATP are in the adenylation conformation, reinforcing the notion that ATP (in particular the pyrophosphate moeity) can only be bound in this conformation (Kochan et al., 2009; Osman et al., 2009).

Hydrolysis of the pyrophosphate, however, is not sufficient to drive the transition to the thioester-forming confirmation, as complexes of ANL enzymes with their cognate adenylate intermediate (Conti et al., 1997; May et al., 2002; Du et al., 2008; Reger et al.,

2008) or intermediate analog (Nakatsu et al., 2006; Hu et al., 2010), which do not have pyrophosphate group bound, are in the adenylate-forming conformation. In contrast, the structure of *P. tomentosa* 4CL1 complexed with an intermediate analog (Hu et al., 2010) and the co-crystal structures of Nt4CL2 in complex with various adenylated acid anhydrides are observed in the thioester-forming conformation. Hence, neither the formation of the adenylated intermediate nor the dissociation of the pyrophosphate group is sufficient to prompt the transition of ANL enzymes into the thioester-forming conformation. Following pyrophosphate release, ANL enzymes can adopt either of the two conformations, and the population equilibria between the two may vary between members.

Subsequent binding of CoA induces the switch to the thioesterification coformation. As shown in the Nt4CL2-AMP-CoA structure, the shift to the thioesterification conformation results in the recruitment of two loops from the C-terminal domain to form the upper side of the CoA-binding tunnel and the displacement of the core motif A10 to unblock the tunnel. Also, several new catalytic residues (Arg435, Lys437, Lys441, and Gln446) are brought to the active site to carry out the thioesterification reaction. These features are consistent with prior observations that all CoA or thioester product bound structures of ANL enzymes are in the thioesterification conformation (Gulick et al., 2003; Reger et al., 2007; Reger et al., 2008; Kochan et al., 2009; Hughes and Keatinge-Clay, 2011).

In summary, our Nt4CL2 structures provide a clear illustration of the catalytic cycle of an acyl-/aryl-CoA ligase. Binding of ATP induces the enzyme to adopt the adenylate-forming conformation, and the A10 motif catalyzes the adenylation reaction in the presence of the acyl/aryl substrate. Upon the finish of the adenylation reaction and the release of pyrophosphate group, CoA dictates the enzyme to adopt the thioester-forming conformation, bringing new catalytic residues to the active site to carry out the thioesterification reaction. Lastly, the thioester product must dissociate before the protein can switch back to the adenylate-forming conformation and start a new cycle. This domain alternation strategy not only provides two sets of catalytic residues for the two

half reactions, but also coordinates the access of substrates and the release of products at different stages, which might be important to minimize the probability of reverse reaction and thus to increase the reaction efficiency.

Substrate Specificity

Previous attempts to investigate 4CL substrate specificity relied on sequence comparison and structural modelling using other ANL enzyme structures (Stuible and Kombrink, 2001; Schneider et al., 2003). A recent report of *P. tomentosa* 4CL1 structures did not provide thorough exploration of the substrate binding pocket due to the lack of structures with cognate substrates (Hu et al., 2010). Our three Nt4CL2 structures in complex with 4-coumaroyl-AMP, caffeoyl-AMP, and feruloyl-AMP allow us to study 4CL substrate specificity determinants in greater details.

Nt4CL2 accomadates the hydroxycinnamate substrate mainly through three segments: His237-Ser243, Met306-Ala309, and Gly332-Met344. The hydroxyphenol ring is sandwiched by the aromatic side chain of Tyr239 and the peptide plane of residues Gly332-Thr336. The side chain of Tyr239 plays a crucial role in stabilizing the substrate ring. Y239A mutation increases the K_m of hydroxycinnamate substrates by 100- to 250-fold and reduced the overall catalysis efficiency (V_{max}/K_m) by over 1000-fold. In contrast, Y239F mutation only results in less than 10-fold of either K_m increase or efficiency decrease. Surprisingly, these results are dramatically different than those reported for *P. tomentosa* 4CL1 (Hu et al., 2010). When the equivalent tyrosine in Pt4CL1 was mutated to alanine, the enzyme activity increased, whereas mutating it to phenylalanine caused the enzyme to be almost inactive against caffeic acid and ferulic acid. Although Pt4CL1 activity was measured using an end-point assay at only one substrate concentration and we measured Nt4CL2 kinetics using continuous assays at various substrate concentrations, such dramatic differences in the mutation effect is unlikely due to the different methods.

Our structures show that the 4-hydroxy group of the hydroxycinnamate substrate faces a large open space, with the side chain of Ser243 to be the only possible block. Mutating it

to alanine or glycine might allow a larger substituent at the *para*-position of the substrate. In contrast, the lateral sides of the pocket apply more constraints on the substrate. Residues that line the lateral sides of the hydroxycinnamte pocket determine the size of substituents allowed on *ortho*- and *meta*-positions of the substrate. On one side, residues Met306-Ala309 are lined closely (3.9 Å) against the *ortho*-position, disfavoring any substitution on this position. These residues belong to a loop (Met306-Gly313) between β-strand 12 and α-helix 11, which is part of both the adenosine binding pocket and CoA tunnel, as described earlier. Therefore, an attempt to changing or deleting any residues in this region to accommodate *ortho*-substituent on this side might risk lowering the affinity to ATP or CoA. In contrast, the *meta*-position of this side is fit into a larger space, thus allowing the 3-hydroxy group of caffeate or the 3-methoxy group of ferulate to be accommadated on this side. It is noteworthy that accomadation of the methoxy group of ferulate is accomplished by redirecting the side chain of Met306. Moreover, this methoxy group is very close to the main chains of Gly332 (2.6 Å) and Ser307 (3.8 Å), therefore a larger substitution might not be allowed at this position.

The orientation of the caffeate and ferulate groups in our structures confutes the prediction proposed in the *P. tomentosa* 4CL1 study (Hu et al., 2010), further underscoring the importance of these structures. Our structures show that the *meta*-substituent cannot be oriented to the other lateral side of the substrate pocket because the bulged loop (Gly339-Leu342) between α-helix12 and β-strand 14 protrudes towards the *ortho*- and *meta*- positions on this side and greatly limits the space. However, a *meta*-methoxy substituent can be tolerated on this side after shortening this loop by one residue, which flattens this region, as shown by our biochemical and structural analysis on a deletion variant Nt4CL2-ΔV341. It might also be possible for this mutant to accept an *ortho*-substitution on this side.

Chapter 4 Structures of a novel cyclic diguanylate effector PelD involved in pellicle formation in *Pseudomonas aeruginosa* PAO1[1]

Abstract

The second messenger bis-(3'-5') cyclic dimeric guanosine monophosphate (c-di-GMP) plays a vital role in the global regulation in bacteria. Here, we describe structural and biochemical characterization of a novel c-di-GMP effector PelD that is critical to the formation of pellicles by *Pseudomonas aeruginosa.* We present high-resolution structures of a cytosolic fragment of PelD in apo form and its complex with c-di-GMP. The structure contains a bi-domain architecture composed of a GAF domain (commonly found in cyclic nucleotide receptors) and a GGDEF domain (found in c-di-GMP synthesizing enzymes), with the latter binding to one molecule of c-di-GMP. The GGDEF domain has a degenerate active site but a conserved allosteric site (I-site), which we show binds c-di-GMP with a K_d of 0.5 μM. We identified a series of residues that are crucial for c-di-GMP binding, and confirmed the roles of these residues through biochemical characterization of site-specific variants. The structures of PelD represent a novel class of c-di-GMP effector and expand the knowledge of scaffolds that mediate c-di-GMP recognition.

Introduction

Bis-(3'-5') cyclic dimeric guanosine monophosphate (c-di-GMP) (Figure 1-2) is a central regulator, which functions as an intracellular second messenger. In bacteria, this molecule confers adaptability to various environmental conditions, by coordinating the transition between the motile planktonic state to a sessile state associated with biofilm production (Jenal and Malone, 2006; Hengge, 2009; Schirmer and Jenal, 2009). Specifically, c-di-GMP stimulates the production of adhesins and exopolysaccharide matrix components and leads to biofilm formation to protect bacteria from host-defense, starvation conditions, and antibiotics (Ryder et al., 2007; Gjermansen et al., 2010). Additional roles for c-di-

[1] **Parts of this chapter are adapted from the following published article with the permission from the publisher:**

Li, Z., Chen, J.H., Hao, Y., and Nair, S.K. (2012). Structures of the PelD cyclic diguanylate effector involved in pellicle formation in *Pseudomonas aeruginosa* PAO1. J Biol Chem 287: 30191-30204.

GMP include control of cell cycle progression (Duerig et al., 2009), antibiotic biosynthesis (Fineran et al., 2007), and expression of virulence genes (Parsek and Singh, 2003; Dow et al., 2006; Kulesekara et al., 2006; Tamayo et al., 2007). The bacterial signaling nodes that respond to the c-di-GMP message present targets for therapeutic intervention against pathogens.

Similar to other second messenger pathways, the c-di-GMP control module can be generally divided into four components that govern signal generation, degradation, recognition, and targeting, respectively (Hengge, 2009). The level of the signal molecule is dynamically regulated by the opposing activities of diguanylate cyclases (DGCs) and phosphodiesterases (PDEs), which synthesize and degrade c-di-GMP, respectively. DGC activity is attributed to proteins that contain a characteristic GGDEF domain (named for the single-letter amino acid nomenclature of essential active site residues), while PDE activity is associated with either enzymes that contain either an EAL or a HD-GYP domain (Ausmees et al., 2001; Paul et al., 2004; Schmidt et al., 2005; Ryan et al., 2006). Interestingly, many DGCs have been shown to be subject to allosteric product inhibition, which is often caused by c-di-GMP binding to the so-called I-site in the GGDEF domain (Chan et al., 2004; Christen et al., 2006; De et al., 2009). The I-site is readily identified by the RxxD (Arg-X-X-Glu; where X is any amino acid) motif, which is connected N-terminally to GGDEF motif through a five-residue linker. These two motifs are antipodal to each other in the three-dimensional structure as shown by the structures of PleD from *Caulobacter vibriodes* (Chan et al., 2004; Wassmann et al., 2007) and WspR from *Pseudomonas aeruginosa* (De et al., 2008; De et al., 2009).

Direct recognition of the c-di-GMP signal occurs through an effector component that is often linked to a signal input component that regulates cellular functions at transcriptional, translational, or post-translational levels (Weber et al., 2006; Merighi et al., 2007; Monds et al., 2007; Duerig et al., 2009). Strikingly, c-di-GMP effectors are highly diverse and are responsible in the diversity of the cellular functions and processes controlled by c-di-GMP in bacteria (Hengge, 2009; Schirmer and Jenal, 2009). The effectors identified so far encompass a variety of domains that are capable of recognizing c-di-GMP, including

the well-characterized PilZ domains (Amikam and Galperin, 2006; Ryjenkov et al., 2006; Benach et al., 2007; Merighi et al., 2007), an unusual receiver domain in VpsT from *Vibrio cholera* (Krasteva et al., 2010), the cyclic nucleotide monophosphate binding domain in Clp from *Xanthomonas campestris* (Leduc and Roberts, 2009; Chin et al., 2010; Tao et al., 2010), the AAA σ^{54} interaction domain in FleQ from *P. aeruginosa* (Hickman and Harwood, 2008), and a degenerate (non-canonical) EAL domain in FimX from *P. aeruginosa* (Navarro et al., 2009) and in LapD from *Pseudomononas fluorescens* (Newell et al., 2009; Navarro et al., 2011), as well as RNA riboswitches (Sudarsan et al., 2008). Most of the above-mentioned types of c-di-GMP effectors have been structurally characterized, including PilZ-domain-containing proteins (Benach et al., 2007; Gulick, 2009; Guzzo et al., 2009; Ko et al., 2010), VpsT (Krasteva et al., 2010), Clp (Chin et al., 2010), FimX (Navarro et al., 2009), LapD (Navarro et al., 2011), and class I and II c-di-GMP-binding riboswitches (Smith et al., 2009; Smith et al., 2010; Smith et al., 2011).

A distinct class of c-di-GMP effector consists of molecules that can bind c-di-GMP through an RxxD motif that resembles the I-site of GGDEF domain of DGCs, but do not show catalytic activity as their active sites lack the requisite GGDEF motif. Examples include PelD from *P. aeruginosa* (Lee et al., 2007), CdgG from *V. cholera* (Beyhan et al., 2008), and PopA from *C. vibriodes* (Duerig et al., 2009). In contrast to the effectors described in the previous paragraph, biochemical data for I-site-containing c-di-GMP effectors is sparse, and there are no crystal structures available for any members of this receptor class.

PelD is found in the Pel pathway involved in the formation of pellicles, one of the major biofilm types formed by *P. aeruginosa* (Friedman and Kolter, 2004). The Pel pathway synthesizes and exports PEL polysaccharides that play both structural and protective roles in *P. aeruginosa* biofilms (Colvin et al., 2011). This pathway contains seven proteins, namely PelABCDEFG, which are highly conserved in diverse microbes (Friedman and Kolter, 2004), and have been shown to be functionally conserved (Vasseur et al., 2005). Although a few attempts were made to assign functions to the seven proteins

(Friedman and Kolter, 2004; Vasseur et al., 2005; Vasseur et al., 2007), the mechanism of polysaccharides production by this pathway remains largely unknown.

Interestingly, the pellicle formation of *P. aeruginosa* was implied to be stimulated by c-di-GMP (Hickman et al., 2005; Kulesekara et al., 2006). Subsequently, PelD was shown to be the receptor of c-di-GMP in the Pel pathway and c-di-GMP binding to PelD is essential to pellicle production (Lee et al., 2007). *In vivo* studies showed that c-di-GMP enhanced transcription level of *pel* genes and deletion of *pelD* gene rendered *P. aeruginosa* PA14 incapable of producing pellicles (Lee et al., 2007). In vitro biochemical assays showed that PelD is the only protein in the Pel pathway that binds c-di-GMP (Lee et al., 2007). PelD is predicted to be an inner membrane protein with four trans-membrane helices and a large cytosolic region (Friedman and Kolter, 2004; Lee et al., 2007). Remarkably, it is proposed to bind c-di-GMP through an I-site-like motif and thus represent a novel family of c-di-GMP-binding proteins (Lee et al., 2007). Mutation of residues in the I-site-like motif abolished the binding of PelD to c-di-GMP and in vivo assays showed that these PelD mutants lack the ability to restore pellicle formation and Congo-red binding by *P. aeruginosa* PA14, indicating that binding of c-di-GMP to PelD is required for pellicle production.

Here, we present structural, biochemical, and mutational analyses of a soluble, cytosolic domain of PelD from *P. aeruginosa* PAO1 that harbors all of the necessary elements for c-di-GMP recognition. We have determined the crystal structures of this cytosolic domain both alone and in complex with c-di-GMP (each to 2.0 Å resolution), as well as that of a deletion variant in complex with c-di-GMP (to 1.7 Å resolution). Utilizing the structural data as a guide, we have carried out structure-function analysis of a number of residues at the c-di-GMP binding site and have quantitatively assessed their roles in ligand binding. The combined structural and biochemical data expand upon the current knowledge of c-di-GMP receptors and provide the first structural view of a c-di-GMP effector that recognizes cognate ligand only through the I-site.

Table 4-1. Data collection, phasing and refinement statistics.

	Apo PelD Native	Apo PelD Hg derivative	PelD-c-diGMP Form I	PelD-c-diGMP Form II	PelD-Δ̃loop c-di-GMP
PDB Codes	4ETX		4EUV	4ETZ	4EU0
Data collection					
Cell dimensions					
a, b, c (Å), β (°)	60.3, 42.4, 60.5 112.8	56.1, 102.3, 103.1 100.1	59.5, 41.5, 64.4 110.9	70.3, 41.3, 110.9 95.5	58.1, 41.4, 62.9 109.9
Resolution (Å)[1]	50-2.0 (2.1-2.0)	50-2.6 (2.7-2.6)	50-2.0 (2.07-2.0)	50-2.05 (2.15-2.05)	50-1.7 (1.76-1.7)
R_{sym} (%)[2]	4.9 (46.8)	9.2 (74.3)	7.5 (36.5)	6.4 (89.7)	6.0 (30.8)
I/σ(I)	24.7 (4.0)	12.0 (1.8)	23.0 (2.8)	15.6 (2.3)	19.8 (5.3)
Completeness (%)	99.8 (99.7)	99.2 (98.8)	99.1 (95.0)	99.7 (99.8)	97.6 (82.5)
Redundancy	7.4 (6.6)	4.3 (4.2)	6.2 (5.3)	5.8 (5.4)	37.7 (3.1)
Phasing					
FOM/DM FOM[3]		0.417/0.668			
Refinement					
Resolution (Å)	25.0-2.0		25.0-2.0	25.0-2.05	25.0-1.7
No. reflections	18,333		18,926	38,272	29,209
R_{work} / R_{free}[4]	23.2/28.5		22.0/28.1	23.0/27.7	22.9/26.6
Number of atoms					
Protein	2,370		2,256	4,536	2,247
c-di-GMP	-		46	92	46
Water	86		128	140	241
B-factors					
Protein	22.5		15.9	28.6	11.7
c-di-GMP	-		27.8	46.2	24.8
Water	28.0		19.3	34.1	25.4
R.m.s deviations					
Bond lengths (Å)	0.012		0.013	0.013	0.007
Bond angles (°)	1.35		1.52	1.51	1.21

1. Highest resolution shell is shown in parenthesis.
2. $R_{sym} = |(I_i - \langle I_i \rangle| \quad I_i$ where I_i = intensity of the *i*th reflection and $\langle I_i \rangle$ = mean intensity.
3. Mean figure of merit before and after density modification
4. R-factor = $(|F_{obs}|-k|F_{calc}|)/ \quad |F_{obs}|$ and R-free is the R value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement.

81

Experimental procedures

Protein purification and crystallization

$PelD_{158\text{-}CT}$ was cloned into pET28 vector and expressed in *E. coli* Rosetta2. Wild type and mutant proteins were purified using Ni-NTA affinity column followed by size-exclusion chromatography in 20 mM HEPES pH 7.5, 100 mM KCl. Protein was concentrated to 20-30 mg/ml and kept at 4 ^{o}C. Protein precipitated when stored in this condition but became soluble again at room temperature. Protein concentration was determined by UV absorbance at 280 nm using calculated extinction coefficient 11920 $M^{-1}cm^{-1}$.$PelD_{158\text{-}CT}$ crystals were grown by the hanging drop vapor diffusion method at room temperature. Each hanging drop contained 1 µl of protein solution and 1 µl of mother liquor. The mother liquor conditions are as follows. (i) $PelD_{158\text{-}CT}$ apo: 100 mM Tris, pH 8, 200 mM $MgCl_2$, and 10% (v/v) PEG 8000; (ii) $PelD_{158\text{-}CT}$ wild type in complex with c-di-GMP (two molecules in the asymmetric unit): 100 mM Tris, pH 8.5, 200 mM Li_2SO_4, and 1.26 M $(NH_4)_2SO_4$; (iii) $PelD_{158\text{-}CT}$ wild type or Δ-loop mutant in complex with c-di-GMP (1 molecule in the asymmetric unit): 50 mM sodium cacodylate, pH 6.5, 10 mM $MgSO_4$, 1.3 M Li_2SO_4. Protein concentration was 2-5 mg/ml in all cases. For co-crystallization, 2 mM c-di-GMP was incubated with the protein at room temperature for 30 minutes prior to crystallization. Crystals were cryo-protected in 15% ethylene glycol before flash-frozen in liquid nitrogen.

Data Collection and Structure Determination

Initial crystallographic studies were carried out using a construct that spanned residues Ile144 ($PelD_{144\text{-}CT}$) through the C-terminus. Flash cooled crystals of $PelD_{144\text{-}CT}$ diffract X-rays beyond a Bragg spacing of 2.5 Å, using an insertion device X-ray beam line (LS-CAT, Sector 21ID, Advanced Photon Source, Argonne, IL). A mercury derivative was prepared by treating crystals with 5 mM ethylmercury bromide for 24 hours. Crystallographic phases were determined by single wavelength anomalous diffraction from the mercurial derivative. A four-fold redundant data set was collected at 100K to a limiting resolution of 2.6 Å (overall R_{merge}= 9.2%, I/ (I)=1.8 in the highest resolution shell). All diffraction data were integrated and scaled using the HKL2000 package (Otwinowski et al., 2003). Heavy atom refinement and phase calculation were carried

out using PHASER (McCoy et al., 2007) as implemented in the PHENIX software suite (Adams et al., 2002; Grosse-Kunstleve and Adams, 2003), followed by density modification using DM (Cowtan and Main, 1993) and cycles of automated building using ARP/wARP (Perrakis et al., 1999) and manual rebuilding using XtalView (McRee, 1999). Continuous electron density could only be observed for two of the four molecules that were expected to be in the crystallographic asymmetric unit, and subsequent refinement of all models using REFMAC5 (Murshudov et al., 1997) stalled with a free R factor greater than 40%. As subsequent inspection of the model revealed that electron density for the amino terminus could only be observed starting from Asn158, all further crystallographic analysis utilized a construct spanning Asn158 through the C-terminus ($PelD_{158\text{-}CT}$). Structures of ligand complexes and deletion variants were determined using molecular replacement, as implemented in PHENIX. Ramachandran analysis shows that over 90% of the protein main chain dihedral angles are in the most favored regions, and the rest in generously allowed regions. Data collection, phasing, and refinement statistic are summarized in Table 4-1.

Analytical Size-Exclusion Chromatography

Oligomerization of $PelD_{158\text{-}CT}$ was examined using analytical size-exclusion chromatography (Superdex 200 HR 10/30, GE Healthcare) in 20 mM HEPES pH 7.5, 100 mM KCl. 600 µl of sample with a protein concentration of 0.5 mg/ml was applied to the column. For protein complexed with c-di-GMP or cGMP, 70 µM of c-di-GMP or 10 mM cGMP was included in the sample, and 10 µM of c-di-GMP or 1 mM cGMP was included in the mobile phase. The molecular weight standards were Blue Dextran (~2,000,000 Da), beta-amylase (~200,000 Da), Albumin (~66,200 Da), carbonic anhydrase (~29,000 Da), and cytochrome c (~12,400 Da), and were all purchased from Sigma.

[^{32}P]-c-di-GMP binding assay

[^{32}P]-c-di-GMP binding assay was adopted from the method by (Lee et al., 2007) with some major modifications. [^{32}P]-c-di-GMP was synthesized by YdeH (Zahringer et al., 2011) using [α-^{32}P]-GTP. It was then incubated with purified wild type PelD158 or each

of the mutants that have N-terminal His-tag under the following conditions: 7 nM of [^{32}P]-c-di-GMP, 30 µM of protein, 50% Ni-NTA agarose beads (GE Healthcare), 10 mM Tris pH 7.5, and 50 mM NaCl. The mixture was incubated at room temperature for 30 minutes and then transferred to a Spin-X 0.22 µm centrifuge tube filter (Costar). The remaining beads in the original tube were washed with 50 µl of wash buffer (10 mM Tris pH 7.5, and 50 mM NaCl) and were also transferred to the filter. Free [^{32}P]-c-di-GMP was removed from the mixture by centrifugation. The flow through was collected in the 2ml centrifuge tube that held the filter insert. The filter was washed twice with 300 µl wash buffer and the flow through was collected in the same centrifuge tube. The filter insert and the flow through were both counted in the scintillation counter and the fraction of bound [^{32}P]-c-di-GMP was calculated by dividing the filter counts with the sum of both counts.

Isothermal titration calorimetry

Measurements were carried out on a Nano ITC (TA Instruments - Waters LLC) with protein protomer typically at 30 µM in the cell and c-di-GMP at 0.5 mM in the syringe. For D370A mutant, protein concentration was 100 µM and c-di-GMP concentration was 3 mM. An initial 1 µl injection was followed by 24 injections of 2 µl each at 240-second intervals. c-di-GMP was synthesized by YdeH using GTP and purified according to the method by Zahringer et al, and quantitated based on UV absorbance using extinction coefficient of 26 $mM^{-1}cm^{-1}$ at 260 nm (Rao et al., 2008). Heat of dilution for c-di-GMP was estimated from the last 7-10 injections and subtracted from raw data before fitting the binding isotherm in NanoAnalyze (TA Instruments). Curve-fitting was conducted using single-independent-site binding model. When using the nominal ligand concentration we consistently obtain a stoichiometry of more than 2 c-di-GMP molecule per protomer of PelD. This physically unreasonable value of stoichiometry lead us suspect that the ligand concentration was overestimated due to the presence of UV-absorbing contaminants, a similar situation as reported previously (Benach et al., 2007). Therefore, the ligand concentration was empirically adjusted by a 2-fold reduction to yield a stoichiometry approximately at 1 when using the single-independent-site binding model. cGMP or cAMP binding experiments were carried out similarly except that

protein concentration was 100 μM and cNMP (Sigma-Aldrich) concentration was 10 mM. Curve fitting was conducted using single-independent-site binding model without any concentration adjustment.

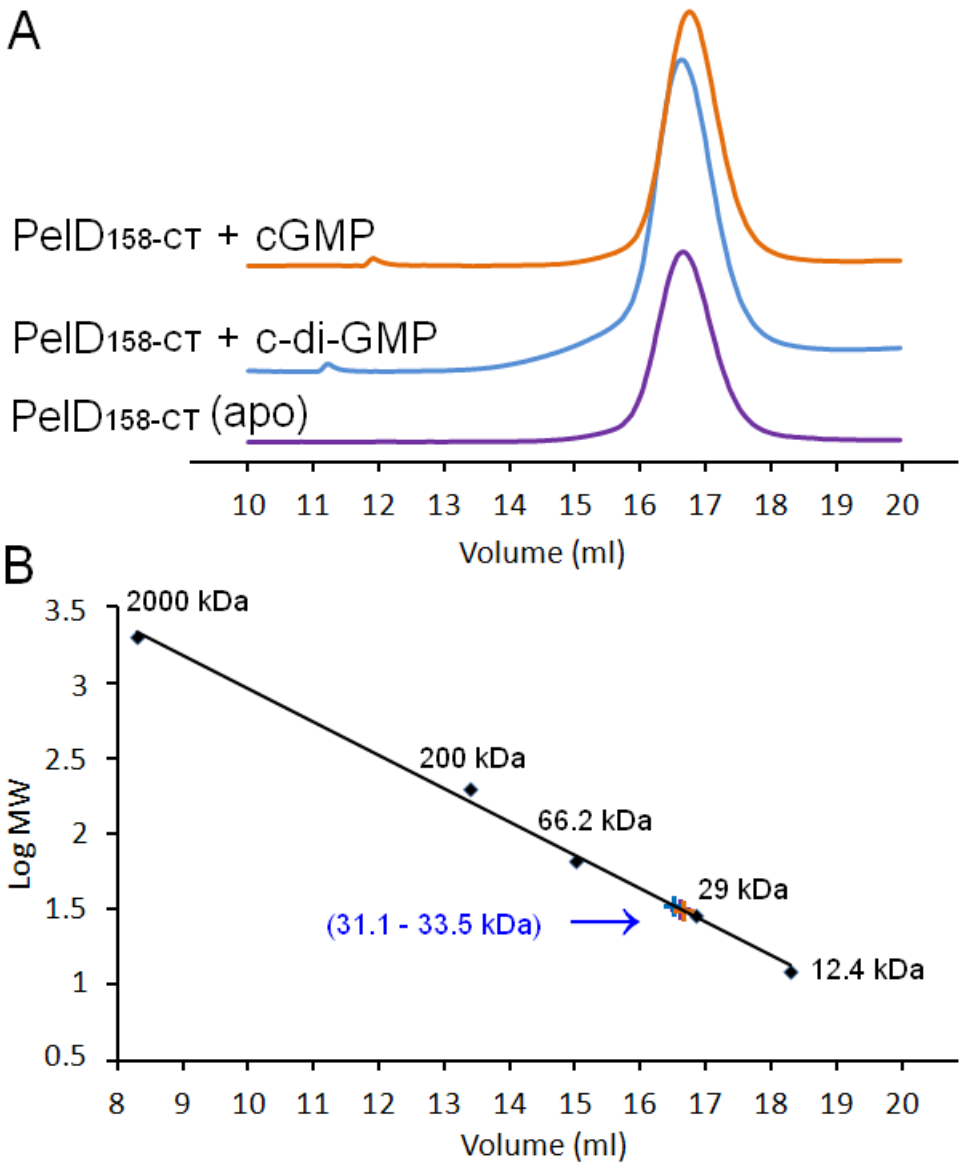


Figure 4-1. Analytical size-exclusion chromatography of $PelD_{158\text{-}CT}$. A) Chromatographic traces of $PelD_{158\text{-}CT}$ in the absence or presence of c-di-GMP or cGMP are shown as indicated. B) Molecular weight standard curve. The molecular weight of each standard is indicated. The positions and calculated molecular weight of protein samples in A) are indicated.

Results

Overall structures of $PelD_{158\text{-}CT}$

A series of soluble constructs encompassing the cytoplasmic region of PelD were purified and crystallized, and these constructs started from Met105, Leu123, Asp 133, or Ile144 ($PelD_{144\text{-}CT}$) through the C-terminus, respectively. Crystals of $PelD_{144\text{-}CT}$ diffracted beyond 2.5 Å resolution and crystallographic phases were determined to 2.6 Å using a mercurial derivative. Clear electron density could only be observed for two of the four molecules in the crystallographic asymmetric unit and while the quality of the experimental map was sufficient to allow building of an initial, near-complete model, the structure could not be satisfactorily refined. Subsequent inspection of the model revealed that electron density for the amino terminus could only be observed starting from Asn158. A new construct encompassing Asn158 through the C-terminus ($PelD_{158\text{-}CT}$) was generated and structural analysis was carried out on crystals of both $PelD_{158\text{-}CT}$ (2.0 Å resolution) and its complex with c-di-GMP in two different crystal forms (form I: 2.0 Å resolution, form II: 2.05 Å resolution) (see Table 4-1 for data collection and refinement statistics). The structures all occupy the same space group of $P2_1$, but with different unit cell dimensions and, consequently, different packing. There is one molecule in the asymmetric unit of the apo structure and in the form I co-crystal structure, while the form II co-crystal structure contacts two molecules in the asymmetric unit that appear to the result of crystal packing and are biologically irrelevant, as illustrated by both the small contact area between protomers (818 $Å^2$) (Krissinel and Henrick, 2007) and the solution behavior of $PelD_{158\text{-}CT}$ as a monomer, both in the presence or absence of c-di-GMP ligand (Figure 4-1) .

While the apo structure is well ordered throughout its entirety, the region between Glu251 and Val263 is poorly defined in both of the c-di-GMP complex structures. Hence, description of the overall fold will be based on the apo structure unless otherwise stated. The structure of $PelD_{158\text{-}CT}$ can be clearly divided into two domains of similar size, an N-terminal domain (composed of Gln158 through Ser309) and a C-terminal domain (encompassing Asp318 through Ala454), which are connected by a loop composed of

residues Asp310-Ala317 (Figure 4-2A). Binding of the ligand does not induce any local or global changes in the structure (Figure 4-2B) and the two structures can be aligned with an average RMSD of 1.0 Å over 285 C atoms.

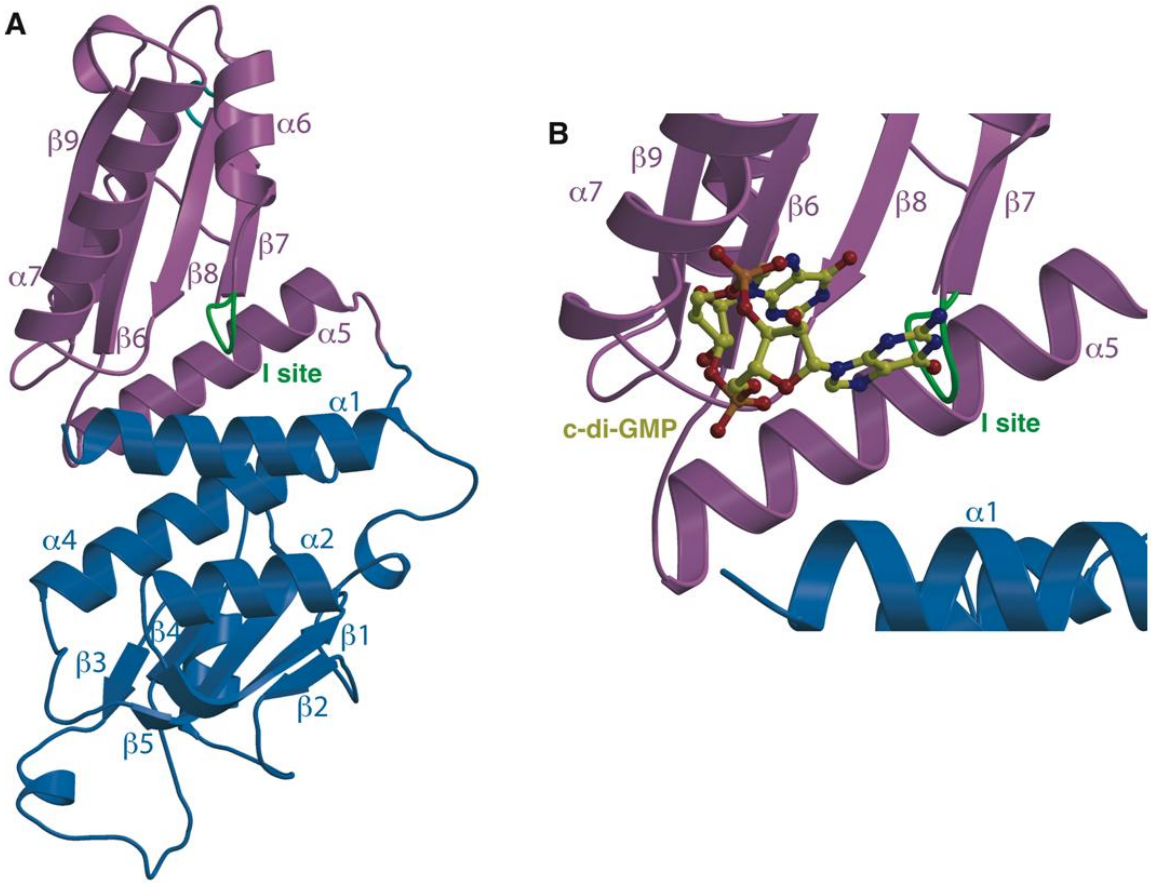


Figure 4-2. Overall structures of $PelD_{158\text{-}CT}$. A) Ribbon diagram of the overall structure of $PelD_{158\text{-}CT}$ illustrating the overall architecture and the secondary structural elements are numbered as illustrated. The disposition of the GGDEF domain (blue) and the GAF domain (pink) is shown with the I-site colored in green and the A-site GGDEF motif colored in cyan . B) Close-up view of the ligand-binding site in the $PelD_{158\text{-}CT}$-c-di-GMP co-crystal structure using the same color and numbering scheme as in panel A. The c-di-GMP ligand is shown as a ball-and-stick with yellow carbon atoms. Binding of the ligand does not result in any significant changes in the structure of the protein.

Table 4-2. DALI search (Holm and Rosenstrom, 2010) results using the C-terminal 142 residues (313-454) (GGDEF domain) of $PelD_{158\text{-}CT}$.

Protein	Organism	PDB code	Z-score	RMSD[a] (Å)	Sequence identity[b]	References
FimX	*Pseudomonas aeruginosa*	3HVA	10.6	3.9	15% (124)	(Navarro et al., 2009)
PleD	*Caulobacter vibrioides*	2V0N	9.6	4.1	17% (127)	(Wassmann et al., 2007)
PleD	*Caulobacter vibrioides*	1W25	9.6	3.9	17% (126)	(Chan et al., 2004)
WpsR	*Pseudomonas aeruginosa*	3I5B	9.4	3.7	14% (121)	(De et al., 2009)
XCC4471	*Xanthomonas campestris* pv. campestris	3QYY	9.2	4.0	15% (121)	(Yang et al., 2011)
WspR	*P.aeruginosa* PAO1	3BRE	9.0	4.0	15% (126)	(De et al., 2008)
LapD	*Pseudomonas fluorescence* Pf0-1	3PJX	8.9	4.3	15% (125)	(Navarro et al., 2011)

[a]RMSD: root-mean-square deviation.

[b]Percentage of identical amino acids over all structurally equivalent residues. The number of structurally equivalent residues is shown in parenthesis.

Table 4-3. DALI search results using the N-terminal 155 residues (158-312) (GAF domain) of $PelD_{158\text{-}CT}$.

Protein	Organism	PDB code	Z-score	RMSD[a] (Å)	Sequence identity[b]	References
PDE5A1	*Homo sapiens*	3MF0	12.7	2.7	13% (127)	(Wang et al., 2010)
Free methionine-R-sulfoxide reductase (fRMsR)	*Neisseria meningitidis*	3MMH	11.7	2.9	18% (131)	(Gruez et al., 2010)
YebR (fRMsR)	*Escherichia coli*	1VHM	11.3	2.9	17% (129)	(Badger et al., 2005; Lin et al., 2007)
Histidine kinase DevS	*Mycobacterium smegmatis*	2VKS	11.1	2.6	17% (109)	(Lee et al., 2008)
Ykg9 (fRMsR)	*Saccharomyces cerevisiae*	1F5M	11.0	2.8	18% (128)	(Ho et al., 2000)
Bacteriophytochrome	*Deinococcus radiodurans*	2O9C	10.9	2.9	20% (124)	(Wagner et al., 2007)
Histidine kinase DosS	*Mycobacterium tuberculosis*	2W3G	10.8	3.2	11% (118)	(Cho et al., 2009)
PDE10A	*Homo sapiens*	2ZMF	10.6	3.4	13% (137)	(Handa et al., 2008)
PDE2A	*Homo sapiens*	3IBJ	10.4	11.3	15% (129)	(Pandit et al., 2009)
PDE6C	*Gallus gallus*	3DBA	10.1	3.3	10% (129)	(Martinez et al., 2008)
Bacteriophytochrome	*Rhodopseudomonas palustris* CGA009	2OOL	10.1	3.2	9% (125)	(Yang et al., 2007)
Histidine kinase DosT	*Mycobacterium tuberculosis*	2VZW	9.9	2.8	16% (113)	(Podust et al., 2008)
Bacteriophytochrome	*Pseudomonas aeruginosa*	3G6O	9.6	12.1	12% (143)	(Yang et al., 2009)
cyaB2	*Anabaena* sp.	1YKD	9.5	3.9	11% (133)	(Martinez et al., 2005)

[a]RMSD: root-mean-square deviation.

[b]Percentage of identical amino acids over all structurally equivalent residues. The number of structurally equivalent residues is shown in parenthesis.

A DALI search (Holm and Rosenstrom, 2010) against the Protein Data Bank using the complete structure of $PelD_{158\text{-}CT}$ failed to identify any structure with significant similarities over the entire polypeptide. However, a search using either the N- or C-terminal domains identified several candidates that show significant structural homology to each of the individual domains (Tables Table 4-2 and Table 4-3). The PelD N-terminus consists of a GAF domain found in various cyclic nucleotide receptors including cyclic-GMP-regulated phosphodiesterases, adenylyl cyclases. The closest structural homolog is the GAF-A of human cyclic GMP (cGMP)-specific 3',5'-cyclic phosphodiesterase PDE5A1 (Wang et al., 2010) with an RMSD of 2.7 Å over 128 C atoms (sequence identity: 15%; PDB code: 3MF0). The C-terminus of PelD shows an architecture similar to GGDEF domain found in diguanylate cyclases and the closest homolog is the GGDEF domain of *P. aeruginosa* c-di-GMP receptor FimX (Navarro et al., 2009), with an RMSD of 3.9 Å over 124 aligned C atoms (sequence identity: 15%; PDB code: 3HVA). Consequently, we refer to the N- and C-terminal domains of $PelD_{158\text{-}CT}$ as the GAF and GGDEF domains, respectively.

The GGDEF domain of $PelD_{158\text{-}CT}$ shows significant differences to canonical GGDEF domains

The GGDEF domain of $PelD_{158\text{-}CT}$ is composed of a central four-stranded β-sheet, sandwiched between two pairs of α-helices (Figure 4-3A) and is topologically similar to other GGDEF domains such as those noted above (Figure 4-3). However, in PelD the GGDEF domain is approximately 20-25 residues shorter than the canonical equivalents found in these other polypeptides and lacks some of the highly conserved secondary structure features (Figure 4-3A and B). For example, a comparison of the GGDEF domain of $PelD_{158\text{-}CT}$ with that of PleD reveals that the canonical central β-sheet in PleD GGDEF domain possesses one extra strand (βi3). Additionally, canonical GGDEF domains contain an additional helix (αi1), as well as two additional anti-parallel β-strands (βi1 and βi2) that are peripheral to the core structure (Figure 4-3B and C). PelD also lacks the catalytically requisite GGDEF sequence characteristic of active DGCs, and instead contains RNDEG (Arg376 through Gly380) at the equivalent position (Figure 4-3C). In active DGCs such as PleD, this motif is located on a loop between two β-strands and is

involved in binding to GTP and metal ion. However, in PelD, both of the β-strands are extended and thus occlude the GTP-binding pocket. Even though the two catalytically requisite metal-ion-coordinating residues (Asp378, Glu379) are conserved in PelD, they point away from the position of the GTP-binding pocket and thus cannot contribute to either substrate binding or catalysis. The absence of both the requisite active site residues, as well as the additional secondary structural elements found in all active DGCs contribute to the lack of a competent active site in the GGDEF domain of PelD.

$PelD_{158\text{-}CT}$ binds to c-di-GMP molecule through the I-site

In the $PelD_{158\text{-}CT}$-c-di-GMP co-crystal structure, one molecule of c-di-GMP molecule is bound to the GGDEF domain through the I-site, in an open, shallow pocket, with only one guanine ring (Gua-1) in the pocket and the other guanine ring (Gua-2) completely exposed to the solvent (Figure 4-4A). The two adenine rings are parallel to each other and both vertical to the 12-membered macrocycle formed by two phosphodiester bonds between the two GMP molecules, engaging in a two-fold symmetrical clip-shaped manner. This binding configuration of c-di-GMP is similar to the one observed in the co-crystal structures PleD (Chan et al., 2004; Wassmann et al., 2007) and WspR (De et al., 2008; De et al., 2009), and the PilZ domain VCA0042 from *V. cholera* (Benach et al., 2007), but distinct from that in FimX (Navarro et al., 2009) and LapD (Navarro et al., 2011), where the c-di-GMP ligand adopts an extended conformation and is inserted in a deep binding pocket.

The interactions of $PelD_{158\text{-}CT}$ with c-di-GMP occur mainly through two segments: a loop composed of residues Arg367 through Asp-370, and a β-loop-α segment encompassing Leu387 to Arg402. Within these regions, Arg367, Asp370, and Arg402 are responsible for the majority of the interactions with c-di-GMP (Figure 4-4B). One of the carboxylate oxygen of Asp370 forms a hydrogen bond with N^2 of Gua-1, and the second forms a hydrogen bond with N-1 of Gua-1. The guanidinium nitrogen of Arg402 forms hydrogen bonds with O^6 and N-7 of Gua-1. The guanidinium group of Arg367 inserts under the Gua-2 ring and forms hydrogen bonds with the diphosphate backbone, and is in

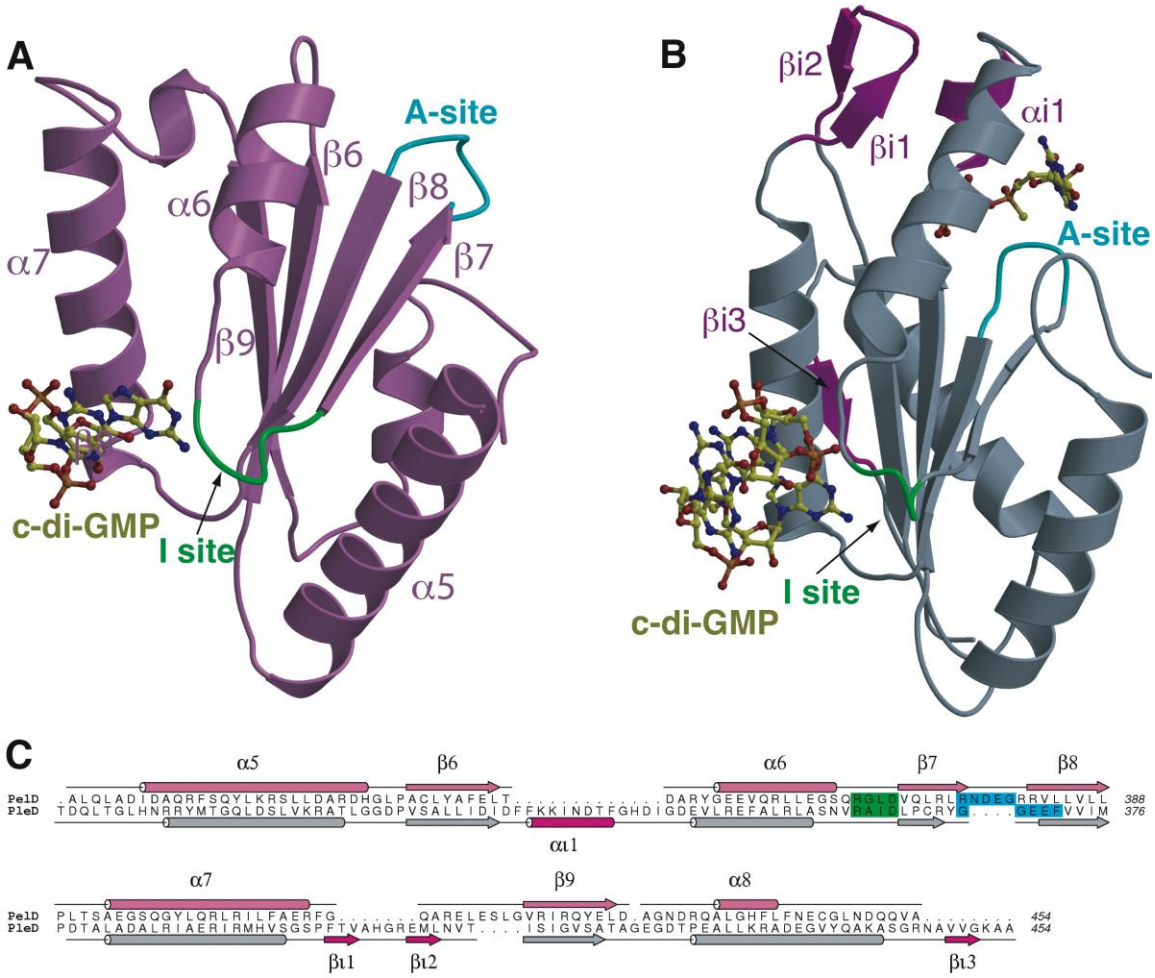


Figure 4-3. GGDEF domains of $PelD_{158\text{-}CT}$ and *Caulobacter vibriodes* PleD. A and B) A comparison of the GGDEF domain of (A) $PelD_{158\text{-}CT}$ (shown in pink with secondary structure numbered as in Figure 1) with that of (B) the active diguanylate cyclase *Caulobacter vibriodes* PleD (shown in gray), with the secondary structural elements that are found in most GGDEF domain but are lacking in $PelD_{158\text{-}CT}$ shown in magenta. The A-site is colored in cyan and the I-site is colored in green. In the PleD co-crystal structure, two molecules of the c-di-GMP product stack and occupy the autoinhibitory I-site. C) Comparison of secondary structural elements between PelD and PleD near the ligand binding sites. The I-site is highlighted in green, the A-site is colored in cyan, and the additional secondary structure elements in PleD (αi1, βi1 and βi2) are colored in magenta.

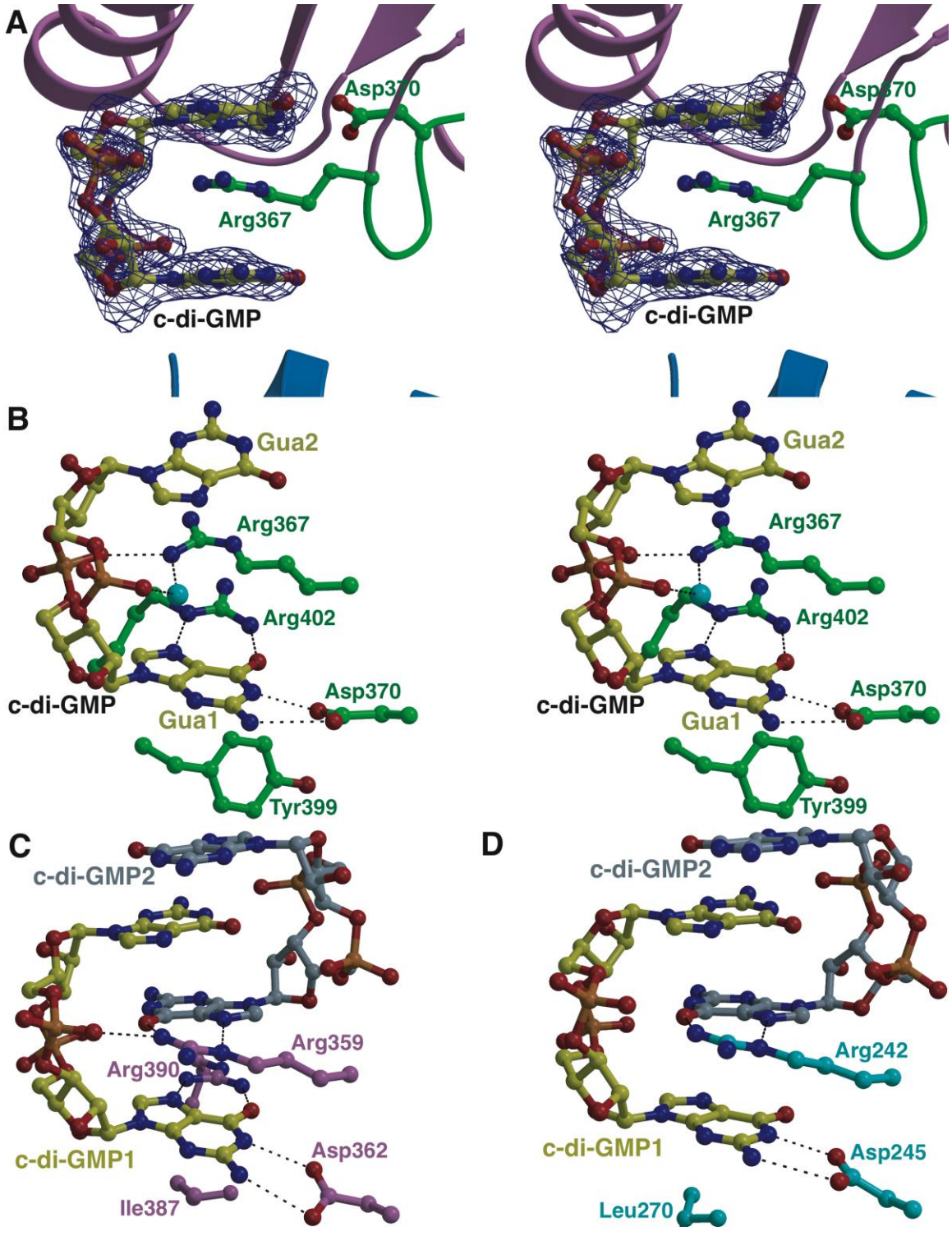


Figure 4-4. Cyclic di-GMP binding site in $PelD_{158\text{-}CT}$. A) Stereo-view of electron density maps calculated using Fourier coefficients $F_{obs} - F_{calc}$ with phases derived from the final refined model of the 1.7 Å resolution co-crystal structure of Δ-loop $PelD_{158\text{-}CT}$ with c-di-GMP. The map was calculated by omitting the coordinates of the cyclic di-GMP prior to

one round of crystallographic refinement and is contoured at 2.3σ (blue mesh) and 7σ(red mesh). A ribbon diagram of the co-crystal structure is superimposed, and the ligand is shown as yellow ball-and-stick and I-site residues Arg-367 and Asp-370 (of the RXXD motif) are shown in green. B) Stereo diagram showing the interactions between the cyclic di-GMP (in yellow) and PelD residues (in green) that are critical for ligand binding (as confirmed by biochemical analysis of site-specific variants; see text for further details). Note that Arg-367 wedges between the two guanines and helps to engage the ligand in a clip-like fashion. C and D) Comparison of the interactions between cyclic di-GMP and the I-site in the co-crystal structures of C) *Caulobacter vibriodes* PleD and D) *Pseudomonas aeruginosa* WspR. Note that in both of these co-crystal structures, two molecules of c-di-GMP stack in order and engage with their respective protein effectors in a manner analogous to the interactions provided by Arg-367 in PelD.

electrostatic proximity with both purine rings (Figure 4-4B). Importantly, Arg367 and Asp370 belong to the conserved RxxD motif and is equivalent to the I-sites in the GGDEF domain of many DGCs like PleD (Wassmann et al., 2007) and WspR (De et al., 2009), where the product c-di-GMP binds and allosterically inhibits DGC activity. In both of these DGCs, two c-di-GMP molecules are bound as intercalated dimers (Figure 4-4C and D). The interactions between the RxxD motif and one c-di-GMP are similar to those observed in PelD, but intercalated Gua rings stabilize each other in a manner similar to the interactions observed between Arg367 and c-di-GMP in the co-crystal structure of PelD (Figure 4-4C and D).

Binding affinity of PelD variants to c-di-GMP

In order to characterize the importance of the residues implicated in c-di-GMP binding in our co-crystal structure, we carried out mutagenesis studies and tested the binding affinity of wild type and variant PelD proteins to c-di-GMP. Residues Arg367, Asp370, and Arg402, which form extensive hydrogen bonds with the ligand, and Tyr399, which provides a hydrophobic floor of the binding pocket, were individually mutated to Ala. Gly395, which also lines the floor of the binding pocket, was mutated to Pro to introduce steric hinderance within the pocket. All these mutants, along with the wild type protein, were purified with polyHis tag. The protein variants were immobilized on Ni-NTA agarose beads, and their ability to retain [^{32}P]-c-di-GMP was tested (Figure 4-5A).

Arg367Ala, Tyr399Ala, and Arg402Ala failed to retain c-di-GMP, while the binding ability of Asp370Ala was reduced by more than 60%. Gly395Pro bound c-di-GMP at a similar level to the wild type.

In order to more accurately quantitatively assess the c-di-GMP binding affinities of Asp370Ala and Gly395Pro variants, we conducted isothermal calorimetry titration (ITC) analysis (Figure 4-5B) of these two mutants as well as on the wild-type. Wild type $PelD_{158\text{-}CT}$ binds c-di-GMP at a K_d of approximately 0.5 μM (See Experimental Procedures). These results are comparable with previously reported values (Lee et al., 2007). The K_d for Asp370Ala and Gly395Pro $PelD_{158\text{-}CT}$ are 28.6 μM and 3.4 μM, a 60-fold and 7-fold decrease in the affinity, respectively (Figure 4-5C and D). These results confirmed the importance of both Asp370 and Gly395 in c-di-GMP binding, which could not be determined directly from the [^{32}P]-c-di-GMP binding assay.

As noted, the only significant difference between the structure of unliganded $PelD_{158\text{-}CT}$ and the two c-di-GMP co-crystal structures is the lack of ordered electron density in the region spanning Glu251 and Val263 in the latter. In both crystal forms of the ligand bound structure, residues in this loop from the unliganded would clash with a bound c-di-GMP from a symmetry related molecule. In order to demonstrate that the crystal contacts are not physiologically relevant, we generated a deletion variant in which ten residues from this loop (Leu249 through His258) were replaced with a Gly-Gly linker (Δ-loop). The binding affinity of this mutant was tested using both [^{32}P]-c-di-GMP binding assay and ITC (Figure 4-5A and E). The Δ-loop variant showed a binding affinity similar to the wild type in the [^{32}P]-c-di-GMP binding assay, and ITC yielded a K_d of 0.4 μM, which is similar to that of the wild type. These results demonstrate that this loop does not interfere with c-di-GMP binding and the symmetry-related interactions are a consequence of crystal packing. To further corroborate these results, we solved the co-crystal structure of the Δ-loop-di-c-GMP complex to 1.7 Å resolution and showed that the binding mode of the ligand is identical to that observed in the wild-type structure.

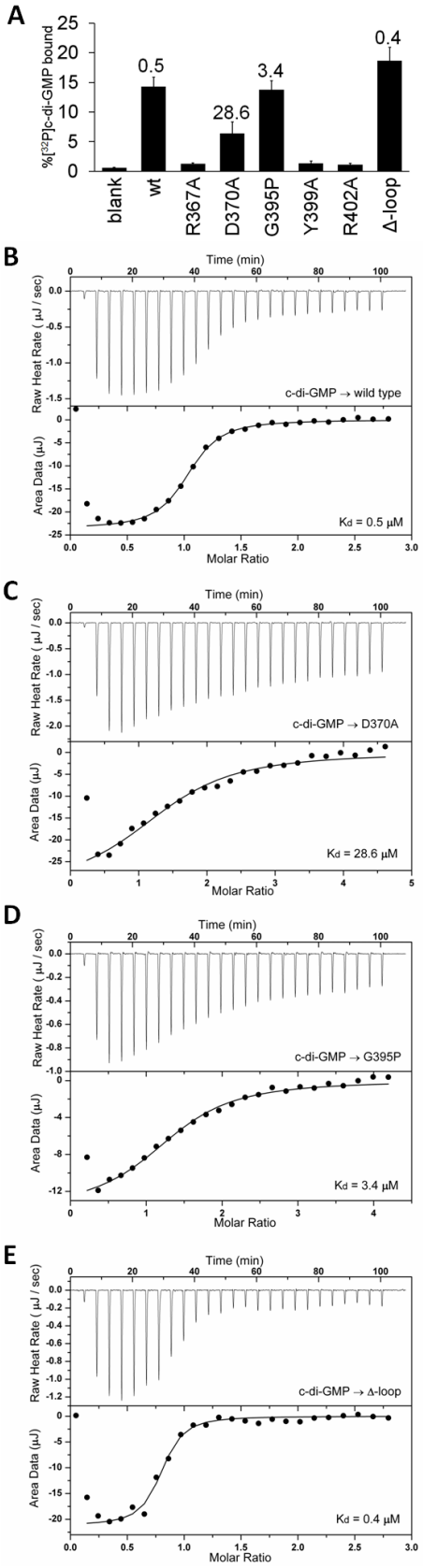


Figure 4-5. Binding affinity of $PelD_{158\text{-}CT}$ and variants to c-di-GMP. A) Filter binding analysis measuring the relative affinity of wild-type and variant $PelD_{158\text{-}CT}$ for [^{32}P]-c-di-GMP. Experiments were conducted in triplicate and the black bars represent average values with associated error bars. K_d values (µM) obtained from isothermal titration calorimetric experiments are shown on top of the corresponding columns. B) C) D) E) Isothermal titration calorimetric analysis of the binding of c-di-GMP to wild-type $PelD_{158\text{-}CT}$, D370A, G395P, and Δ-loop mutants, respectively. The binding isotherms for the titration experiment (top) and fitted curves based on single-independent-site binding model (bottom) are shown.

GAF domain of PelD$_{158\text{-}CT}$

The GAF domain is part of various multi-domain proteins that participate in numerous signal transduction processes (Aravind and Ponting, 1997; Zoraghi et al., 2004). The PelD$_{158\text{-}CT}$ GAF domain displays structural similarity to those of cAMP- or cGMP-specific phosphodiesterases (PDE), including PDE2A (Martinez et al., 2002; Pandit et al., 2009), 5A(Wang et al., 2010), 6C (Martinez et al., 2008), and 10A (Handa et al., 2008), as well as the adenylyl cyclase (AC) CyaB2 (Martinez et al., 2005) (Table 4-3). The GAF domains in PDEs and ACs have the capacity to bind cyclic nucleotide (cNMP), including cAMP and cGMP, which allosterically regulate the catalytic activity of these enzymes. Several PDEs are shown to bind cGMP or cAMP with nanomolar affinities (10-200 nM) (Hebert et al., 1998; Huang et al., 2004; Wu et al., 2004; Zoraghi et al., 2005; Heikaus et al., 2008; Heikaus et al., 2009) and the cyanobacterial adenylyl cyclases are shown to be activated exclusively by cAMP at sub-micromolar concentrations (Martinez et al., 2005).

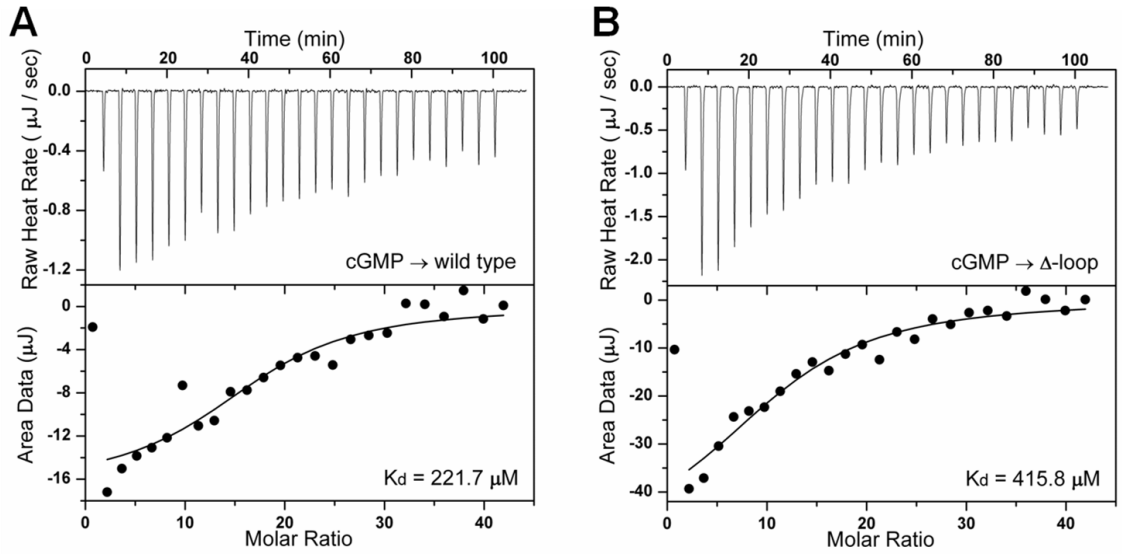


Figure 4-6. Calorimetric analysis of binding affinities of PelD$_{158\text{-}CT}$ and the Δ-loop variant for cyclic GMP. A and B) Binding isotherms and fitted curves from isothermal titration calorimetric analysis of the binding affinity of A) wild-type PelD$_{158\text{-}CT}$ and B) Δ-loop for cyclic GMP.

In order to determine if the GAF domain of PelD$_{158\text{-}CT}$ is functional in nucleotide binding, we carried out ITC analysis using either cAMP or cGMP. Calorimetric analysis demonstrated that PelD$_{158\text{-}CT}$ does not bind cAMP with any appreciable affinity (data not

shown), and only binds cGMP weakly, with a K_d of 221.7 μM that is several orders of magnitude higher than those reported previously for other GAF domains (Figure 4-6A). Given the low intracellular concentration of cGMP in bacteria (below 100 nM) (Black et al., 1980) and the experimentally determined high K_d value of PelD, the GAF-like domain of $PelD_{158\text{-}CT}$ likely does not bind cyclic nucleotides.

A structure-based comparison of the $PelD_{158\text{-}CT}$ GAF domain with those of nucleotide-activated GAF domains reveal several important features that may result in the low affinity of $PelD_{158\text{-}CT}$-GAF for cGMP (Figure 4-7). First, in $PelD_{158\text{-}CT}$-GAF a loop that connects strand β2 and helix α3 travels through the potential cyclic nucleotide-binding pocket and leaves very little room to accommodate any ligands (Figure 4-7A). Second, strand βi1 in the GAF domains of PDEs undergoes a significant movement towards the ligand pocket upon cGMP binding (Heikaus et al., 2008; Wang et al., 2010). However, this strand is absent in $PelD_{158\text{-}CT}$-GAF domain (Figure 4-7A and B). Lastly, several residues that are shown to be required for cyclic nucleotide binding are not conserved in $PelD_{158\text{-}CT}$-GAF domain (Figure 4-7C). For example, in PDE10A, residues Cys287, Phe304, Asp305, Phe352, Thr364, and Gln383 are involved in cAMP binding (Handa et al., 2008), but none of these residues are conserved in $PelD_{158\text{-}CT}$-GAF domain. In addition, a conserved NKFDE motif (Ho et al., 2000; Martinez et al., 2002; Pandit et al., 2009) is largely degenerate in $PelD_{158\text{-}CT}$-GAF domain.

Of particular note, the region in $PelD_{158\text{-}CT}$-GAF domain that is disordered in the c-di-GMP co-crystal structure (Glu251-Val263) corresponds to a portion of the cyclic nucleotide binding pocket in the PDEs. The structure of cAMP-bound PDE10A GAF-B showed that cAMP molecule is deeply buried in a pocket that uses the antiparallel β-sheet as the floor and a short helix (αi1, Asn353-Gly361 in PDE10A) as the roof (Handa et al., 2008) (Figure 4-7B). To test whether the region between Glu251 through Val263 plays a role in to the inability of $PelD_{158\text{-}CT}$ to bind cyclic nucleotides, we carried out ITC analysis on the Δ-loop mutant with cGMP (Figure 4-6B). The result showed that deleting this region led to only a modest increase of the K_d (415.8 μM). Hence, this loop region does not play any role in the inability of $PelD_{158\text{-}CT}$ GAF to bind ligands.

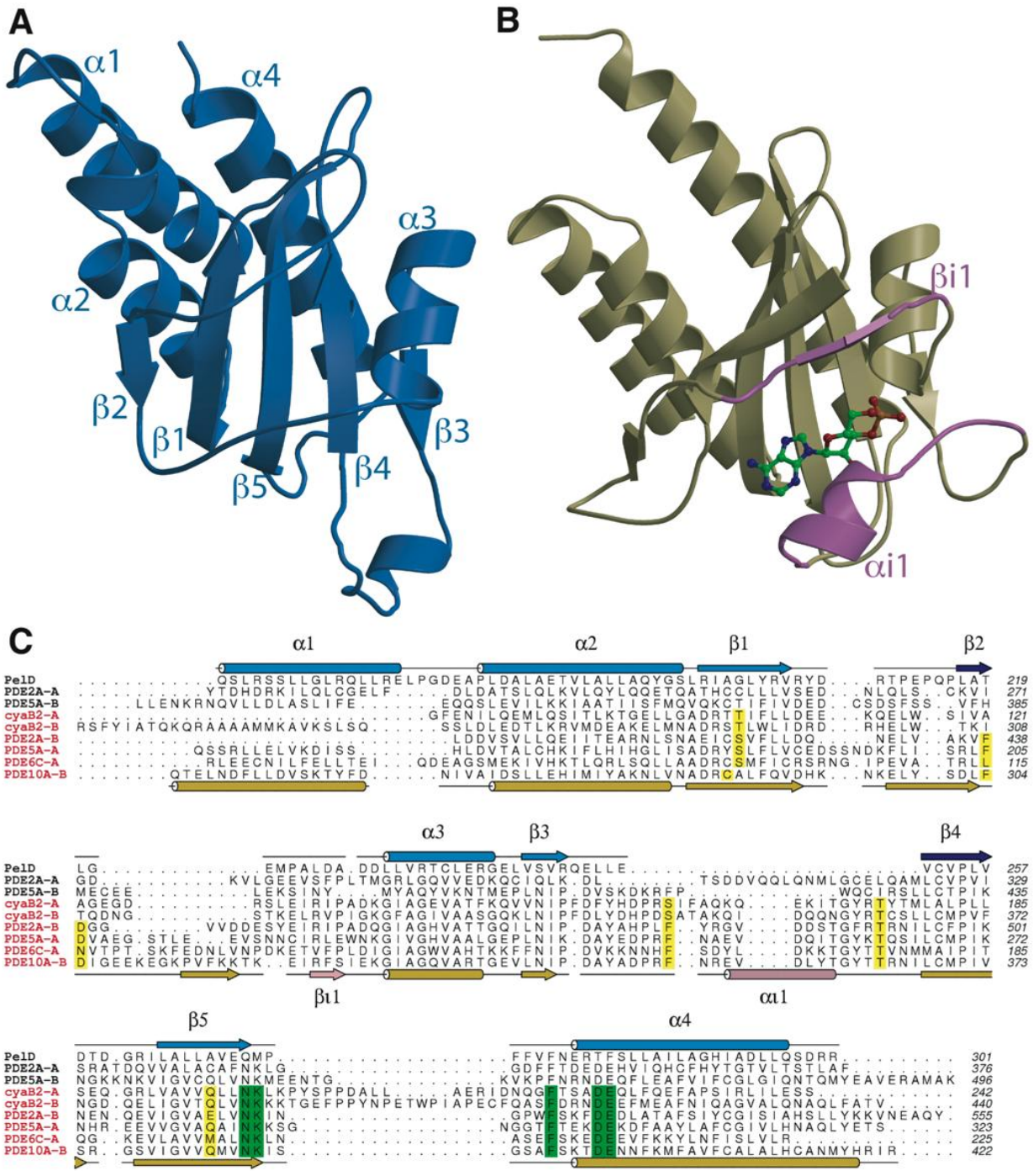


Figure 4-7. GAF domains of $PelD_{158\text{-}CT}$ and Phosphodiesterase 10A. A and B) Comparison of the GAF domains of A) $PelD_{158\text{-}CT}$ (shown in cyan with secondary structure numbered as in Figure 1) with the B) human PDE10A-cyclic GMP co-crystal structure (shown in brown with ligand colored in green ball-and-stick), with the secondary structural elements that are found in most GAF domain but are lacking in $PelD_{158\text{-}CT}$ shown in magenta. The binding pocket for cyclic GMP is partially occluded in

$PelD_{158\text{-}CT}$ by the loop that joins β2 and α3.C) Comparison of secondary structural elements between PelD and PDE10A near the cyclic GMP binding sites. The I-site is highlighted in green, the A-site is colored in cyan, and the additional secondary structure elements in PleD (αi1 and βi1) are colored in magenta. Residues in PDE10A that involved in direct contact with the cyclic nucleotide are highlighted in yellow and the NKFDE motif is highlighted in green.

We also examined the possibility that $PelD_{158\text{-}CT}$-GAF may mediate homodimerization, a typical feature of GAF domains (Heikaus et al., 2009). Analytical size-exclusion chromatographic analysis failed to identify any changes in the elution profile of $PelD_{158\text{-}CT}$ in the presence of high concentration of cGMP (Figure 4-1). Although the dimerization interface and the domain orientation differ for many GAF domains, they all involve the two or three helices (α1, 2, and 4) located on the opposite side of the ligand binding pocket (Martinez et al., 2002; Handa et al., 2008; Pandit et al., 2009; Gruez et al., 2010). However, the corresponding helices in $PelD_{158\text{-}CT}$-GAF domain, consisting of residues Gln158-Glu174 (α1), Leu231-Gly240 (α2), and Glu290-Leu307 (α4), are oriented towards the GGDEF domain and partly buried in the domain interface (Figure 4-2). Thus, the conformation of PelD observed in our structures is not competent to mediate dimerization, consistent with the results from our analytical size-exclusion data (Figure 4-1).

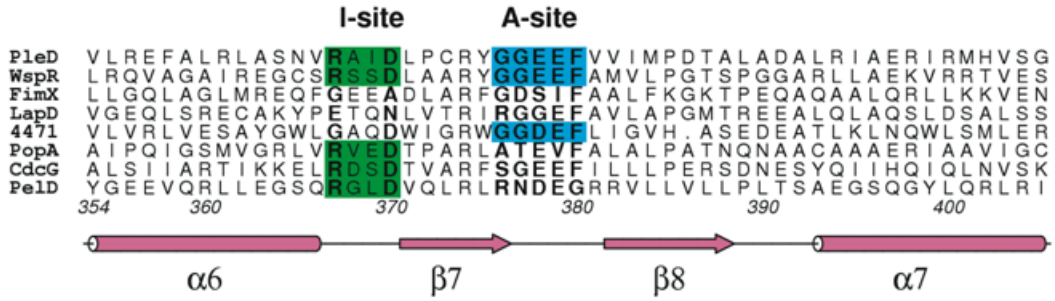


Figure 4-8. Multiple sequence alignment of the GGDEF domains discussed in this study (see text for further details). The I-site RXXD motif (highlighted in green) and the A-site GGDEF motif (in cyan) are shown in bold. The nomenclature of the secondary structural elements is based on the PelD structure.

Discussion

The GGDEF domain of $PelD_{158\text{-}CT}$ represents a novel class of c-di-GMP receptor that binds c-di-GMP through a conserved I-site. GGDEF domains constitute the active sites of diguanylate cyclases (DGCs), such as PleD from *C. vibriodes* (Wassmann et al., 2007), WspR from *P. aeruginosa* (De et al., 2009), and XCC4471 from *X. campestris* (Yang et al., 2011), but are also found in enzymatically inactive c-di-GMP receptors, such as FimX (Navarro et al., 2009). The GGDEF domains of PleD and WspR both have an active GGEEF motif (A-site) that binds to substrate and catalyzes the cyclization of two molecules of GTP into one molecule of cyclic di-GMP. Both PleD and WspR contain a conserved RxxD motif (I-site) located amino-terminal to the A-site, and this I-site binds to c-di-GMP and accounts for allosteric product inhibition (Wassmann et al., 2007; De et al., 2009). In contrast, XCC4471 has a conserved A-site but a degenerate I-site, and (competitive) product inhibition is achieved by c-di-GMP binding directly to the A-site (Yang et al., 2011). Lastly, the GGDEF domain of FimX is degenerate at both the A-site and the I-site, and lacks both DGC activity and c-di-GMP-binding capability (Navarro et al., 2009). A similar feature is also observed in another c-di-GMP receptor LapD from *P. fluorescens* (Navarro et al., 2011). Remarkably, distinct from all these domains, the GGDEF domain of $PelD_{158\text{-}CT}$ has a degenerate A-site (R_{375}NDEG) but a conserved I-site (R_{367}GLD) (Figure 4-8). This combination is also conserved in some other potential c-di-GMP receptors, for example, CdgG from *V. cholera* (Beyhan et al., 2008), and PopA from *C. vibriodes* (Duerig et al., 2009) (Figure 4-8). Thus the $PelD_{158\text{-}CT}$-GGDEF domain is representative of the class of c-di-GMP receptors that contain a degenerate GGDEF active site but a conserved I-site that can engage c-di-GMP.

Our structures showed that PelD binds to one molecule of c-di-GMP through the conserved I-site in the GGDEF domain. Mutations of the conserved residues in this site, Arg367 and Asp370, abolished the binding of PelD to c-di-GMP. Importantly, it was shown in previous *in vivo* studies that such mutants failed to restore the ability of *P. aeruginosa* PA14 to form pellicles and bind Congo red. Combining these functional studies and our structural data, a strong correlation may be concluded between the binding of PelD to c-di-GMP and the formation of pellicles *in vivo*.

Although the GAF domain of $PelD_{158\text{-}CT}$ is topologically similar to canonical GAF domains that can bind to cNMP molecules, our calorimetric studies show that $PelD_{158\text{-}CT}$ does not bind either cAMP or cGMP with affinities that are physiologically meaningful. A number of GAF-containing proteins have tandem GAF domains and the two GAF domains are proposed to have distinct functions (Heikaus et al., 2009). For example, only one of the two GAF domains in PDE2A (GAF-B) binds a cyclic nucleotide, while the second domain (GAF-A) lacks ligand-binding ability but is proposed to function as a dimerization locus (Martinez et al., 2002; Pandit et al., 2009). cGMP binding to GAF-B domain has been shown to allosterically increase the PDE activity at the catalytic domain (Martins et al., 1982). A second example is that cGMP binding to GAF-A domain of PDE5A stimulates phosphorylation through a cGMP-dependent protein kinase, which in turn increases the catalytic activity of PDE5A and cGMP binding affinity of its GAF-A domain (Francis et al., 2002; Rybalkin et al., 2003b; Rybalkin et al., 2003a).

Our results showed that the GAF domain of PelD has only very weak affinity to cyclic nucleotides, thus it is unlikely regulated by these molecules. On the other hand, the location of the $PelD_{158\text{-}CT}$ GAF domain between the C-terminal c-di-GMP-binding GGDEF domain and an N-terminal transmembrane region implies that it might serve as a c-di-GMP signal relay between these two domains. Sequence-based genome analysis demonstrates that the association of a GAF domain with a GGDEF domain occurs in many diguanylate cyclases (Schirmer and Jenal, 2009). However, only a few of these proteins have been biochemically characterized. The diguanylate cyclases DgcA from *Rhodobacter sphaeroides* (Ryjenkov et al., 2005), and MSDGC-1 from *Mycobacterium smegmatis* (Bharati et al., 2012) have been shown to largely lose their DGC activity after partial or complete removal of the GAF domain. However, the rationale behind this remains elusive, as a ligand-dependent function of the GAF domains in these proteins has not yet been established. In the case of DgcA, neither cAMP nor cGMP stimulated DGC activity (Ryjenkov et al., 2005).

To date, all PDEs that have been structurally characterized form dimers, although the functional significance of PDE dimerization remains unclear. The regulatory N-terminal region of these proteins, including the tandem GAF domains, are suggested to provide dimerization contacts as the isolated catalytic domains from PDE2A (Iffland et al., 2005), PDE5A (Sung et al., 2003; Wang et al., 2006), and PDE10A are monomeric (Chappie et al., 2007). Unlike the GAF domains found in PDEs, $PelD_{158\text{-}CT}$-GAF domain does not form a dimer, neither in the crystal nor in solution. The helices that provide the dimerization interface in other GAF domains are partially buried between the GAF and GGDEF domains of $PelD_{158\text{-}CT}$, and dimerization through the GAF domain would require significant conformational movements.

In most of the characterized receptors, c-di-GMP binding usually induces large conformational changes, and these changes have been proposed to propagate signal transduction. For example, binding of c-di_GMP to the PilZ domain of VCA0042 from *V. cholerae* induces a 123° rotation, resulting a more compact overall structure and drastically different accessible protein surface, which is proposed to interact directly with downstream effectors (Benach et al., 2007). Another example is a membrane bound effector LapD from P. *fluorescens* in which autoinhibitory interactions between the degenerate EAL and HAMP domains is relieved upon c-di-GMP binding to the EAL domain, promoting the interaction of the HAMP domain with other effectors (Newell et al., 2009; Navarro et al., 2011). Lastly, c-di-GMP binding to the I-site of two DGCs, PleD (Chan et al., 2004; Wassmann et al., 2007) and WspR (De et al., 2008; De et al., 2009), allosterically inhibits the enzyme activity by forcing the protein dimer into a non-productive conformation.

Unlike these other c-di-GMP binding targets, c-di-GMP binding to $PelD_{158\text{-}CT}$ does not result in any structural changes, either globally or local to the ligand-binding site. The relative orientation of the GAF and GGDEF domains are retained upon c-di-GMP binding, and $PelD_{158\text{-}CT}$ remains as a monomer regardless of the absence and presence of the ligand. Similarly, the c-di-GMP binding EAL domain from FimX from *P. aeruginosa* is monomeric, and does not undergo any structural changes upon ligand binding (Navarro

et al., 2009). It was proposed that c-di-GMP binding might facilitate complex formation between the EAL domain and an unidentified binding partner (Navarro et al., 2009). A similarly plausible model can also be considered for PelD, and this is further strengthened by the fact that c-di-GMP in the complex structure is significantly surface exposed. Binding of the ligand might either interrupt the interaction between the PelD GGDEF domain and an unidentified binding partner, or bridge the GGDEF domain to its binding partner. An example of protein-protein interactions triggered by c-di-GMP was reported recently on a PilZ domain-containing protein (Guzzo et al., 2009).

There are a large number of GGDEF domain-containing proteins in bacteria, which have diverse functions and thus account in part for the complexity of c-di-GMP signaling. While studies have begun to investigate the hierarchy and specificity of GGDEF domain-mediated signaling pathways (Kader et al., 2006; Sommerfeldt et al., 2009), the knowledge of protein binding partners of GGDEF domains is still limited. The binding or dissociation of a protein partner to the GGDEF domain might cause a conformational change in the N-terminal transmembrane domain of PelD, using the GAF domain as a signal relay. A BLAST search (Marchler-Bauer et al., 2009) reveals that the N-terminal transmembrane domain of PelD (N-terminus - Ala104) is categorized as a domain of unknown function under DUF4118 superfamily (Pfam 13493). Domains in this superfamily exist in a wide variety of bacterial signaling proteins, and these may play a role in signal transduction. Along with two other proteins in the Pel pathway, PelE and PelG (Friedman and Kolter, 2004), which are predicted to contain transmembrane helices, PelD could facilitate the export of carbohydrate-containing substance. Such export processes might require interactions between these transmembrane proteins and PelC, an outer-membrane lipoprotein that has been shown to facilitate exopolysaccharide transport (Vasseur et al., 2007). We are continuing additional biochemical, microbiological and structural studies aimed at identifying the effectors downstream of PelD that mediate pellicle formation.

References

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr **58:** 1948-1954.

Allina, S.M., Pri-Hadash, A., Theilmann, D.A., Ellis, B.E., and Douglas, C.J. (1998). 4-Coumarate:coenzyme A ligase in hybrid poplar. Properties of native enzymes, cDNA cloning, and analysis of recombinant enzymes. Plant Physiol **116:** 743-754.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389-3402.

Amikam, D., and Galperin, M.Y. (2006). PilZ domain is part of the bacterial c-di-GMP binding protein. Bioinformatics **22:** 3-6.

Anstead, G.M., and Owens, A.D. (2004). Recent advances in the treatment of infections due to resistant Staphylococcus aureus. Curr Opin Infect Dis **17:** 549-555.

Anstead, G.M., Quinones-Nazario, G., and Lewis, J.S., 2nd. (2007). Treatment of infections caused by resistant Staphylococcus aureus. Methods in Molecular Biology **391:** 227-258.

Aravind, L., and Ponting, C.P. (1997). The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem. Sci. **22:** 458-459.

Ausmees, N., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M., and Lindberg, M. (2001). Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. FEMS Microbiol. Lett. **204:** 163-167.

Badger, J., Sauder, J.M., Adams, J.M., Antonysamy, S., Bain, K., Bergseid, M.G., Buchanan, S.G., Buchanan, M.D., Batiyenko, Y., Christopher, J.A., Emtage, S., Eroshkina, A., Feil, I., Furlong, E.B., Gajiwala, K.S., Gao, X., He, D., Hendle, J., Huber, A., Hoda, K., Kearins, P., Kissinger, C., Laubert, B., Lewis, H.A., Lin, J., Loomis, K., Lorimer, D., Louie, G., Maletic, M., Marsh, C.D., Miller, I., Molinari, J., Muller-Dieckmann, H.J., Newman, J.M., Noland, B.W., Pagarigan, B., Park, F., Peat, T.S., Post, K.W., Radojicic, S., Ramos, A., Romero, R., Rutter, M.E., Sanderson, W.E., Schwinn, K.D., Tresser, J., Winhoven, J., Wright, T.A., Wu, L., Xu, J., and Harris, T.J. (2005). Structural analysis of a set of proteins resulting from a bacterial genomics project. Proteins **60:** 787-796.

Balamurugan, R., Dekker, F.J., and Waldmann, H. (2005). Design of compound libraries based on natural product scaffolds and protein structure similarity clustering (PSSC). Mol Biosyst **1:** 36-45.

Bandarian, V., Pattridge, K.A., Lennon, B.W., Huddler, D.P., Matthews, R.G., and Ludwig, M.L. (2002). Domain alternation switches B(12)-dependent methionine synthase to the activation conformation. Nat Struct Biol **9:** 53-56.

Bar-Tana, J., and Rose, G. (1968). Studies on medium-chain fatty acyl-coenzyme a synthetase. Enzyme fraction I: mechanism of reaction and allosteric properties. Biochem J **109:** 275-282.

Beauregard, D.A., Williams, D.H., Gwynn, M.N., and Knowles, D.J. (1995). Dimerization and membrane anchors in extracellular targeting of vancomycin group antibiotics. Antimicrob Agents Chemother **39:** 781-785.

Becker-Andre, M., Schulze-Lefert, P., and Hahlbrock, K. (1991). Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate: CoA ligase genes in potato. J Biol Chem **266:** 8551-8559.

Benach, J., Swaminathan, S.S., Tamayo, R., Handelman, S.K., Folta-Stogniew, E., Ramos, J.E., Forouhar, F., Neely, H., Seetharaman, J., Camilli, A., and Hunt, J.F. (2007). The structural basis of cyclic diguanylate signal transduction by PilZ domains. EMBO J. **26:** 5153-5166.

Beuerle, T., and Pichersky, E. (2002). Enzymatic synthesis and purification of aromatic coenzyme a esters. Anal Biochem **302:** 305-312.

Beyhan, S., Odell, L.S., and Yildiz, F.H. (2008). Identification and characterization of cyclic diguanylate signaling systems controlling rugosity in Vibrio cholerae. J. Bacteriol. **190:** 7392-7405.

Bharati, B.K., Sharma, I.M., Kasetty, S., Kumar, M., Mukherjee, R., and Chatterji, D. (2012). A full length bifunctional protein involved in c-di-GMP turnover is required for long term survival under nutrient starvation in Mycobacterium smegmatis. Microbiology**:** [Epub ahead of print].

Bischoff, D., Pelzer, S., Holtzel, A., Nicholson, G.J., Stockert, S., Wohlleben, W., Jung, G., and Sussmuth, R.D. (2001a). The Biosynthesis of Vancomycin-Type Glycopeptide Antibiotics-New Insights into the Cyclization Steps Angewandte Chemie. International Ed. In English **40:** 1693-1696.

Bischoff, D., Pelzer, S., Bister, B., Nicholson, G.J., Stockert, S., Schirle, M., Wohlleben, W., Jung, G., and Sussmuth, R.D. (2001b). The Biosynthesis of Vancomycin-Type Glycopeptide Antibiotics-The Order of the Cyclization Steps. Angewandte Chemie. International Ed. In English **40:** 4688-4691.

Black, R.A., Hobson, A.C., and Adler, J. (1980). Involvement of cyclic GMP in intracellular signaling in the chemotactic response of Escherichia coli. Proc. Natl. Acad. Sci. USA **77:** 3879-3883.

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. Annu Rev Plant Biol **54:** 519-546.

Borghi, A., Coronelli, C., Faniuolo, L., Allievi, G., Pallanza, R., and Gallo, G.G. (1984). Teichomycins, new antibiotics from Actinoplanes teichomyceticus nov. sp. IV. Separation and characterization of the components of teichomycin (teicoplanin). J Antibiot (Tokyo) **37:** 615-620.

Boudet, A.M., Kajita, S., Grima-Pettenati, J., and Goffner, D. (2003). Lignins and lignocellulosics: a better control of synthesis for new and improved uses. Trends Plant Sci **8:** 576-581.

Branchini, B.R., Murtiashaw, M.H., Magyar, R.A., and Anderson, S.M. (2000). The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase. Biochemistry **39:** 5433-5440.

Burdette, D.L., Monroe, K.M., Sotelo-Troha, K., Iwig, J.S., Eckert, B., Hyodo, M., Hayakawa, Y., and Vance, R.E. (2011). STING is a direct innate immune sensor of cyclic di-GMP. Nature **478:** 515-518.

Challis, G.L. (2008). Mining microbial genomes for new natural products and biosynthetic pathways. Microbiology **154:** 1555-1569.

Chan, C., Paul, R., Samoray, D., Amiot, N.C., Giese, B., Jenal, U., and Schirmer, T. (2004). Structural basis of activity and allosteric control of diguanylate cyclase. Proc. Natl. Acad. Sci. USA **101:** 17084-17089.

Chang, K.H., Xiang, H., and Dunaway-Mariano, D. (1997). Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: a site-directed mutagenesis study with the Pseudomonas sp. strain CBS3 4-chlorobenzoate:coenzyme A ligase. Biochemistry **36:** 15650-15659.

Chappie, T.A., Humphrey, J.M., Allen, M.P., Estep, K.G., Fox, C.B., Lebel, L.A., Liras, S., Marr, E.S., Menniti, F.S., Pandit, J., Schmidt, C.J., Tu, M., Williams, R.D., and Yang, F.V. (2007). Discovery of a series of 6,7-dimethoxy-4-pyrrolidylquinazoline PDE10A inhibitors. J. Med. Chem. **50:** 182-185.

Chin, K.H., Lee, Y.C., Tu, Z.L., Chen, C.H., Tseng, Y.H., Yang, J.M., Ryan, R.P., McCarthy, Y., Dow, J.M., Wang, A.H., and Chou, S.H. (2010). The cAMP receptor-like protein CLP is a novel c-di-GMP receptor linking cell-cell signaling to virulence gene expression in Xanthomonas campestris. J. Mol. Biol. **396:** 646-662.

Cho, H.Y., Cho, H.J., Kim, Y.M., Oh, J.I., and Kang, B.S. (2009). Structural insight into the heme-based redox sensing by DosS from Mycobacterium tuberculosis. J. Biol. Chem. **284:** 13057-13067.

Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., Meuwly, M., and Jenal, U. (2006). Allosteric control of cyclic di-GMP signaling. J. Biol. Chem. **281:** 32015-32024.

Cicchillo, R.M., Zhang, H., Blodgett, J.A., Whitteck, J.T., Li, G., Nair, S.K., van der Donk, W.A., and Metcalf, W.W. (2009). An unusual carbon-carbon bond cleavage reaction during phosphinothricin biosynthesis. Nature **459:** 871-874.

Colvin, K.M., Gordon, V.D., Murakami, K., Borlee, B.R., Wozniak, D.J., Wong, G.C., and Parsek, M.R. (2011). The pel polysaccharide can serve a structural and protective role in the biofilm matrix of Pseudomonas aeruginosa. PLoS Pathog. **7:** e1001264.

Conti, E., Franks, N.P., and Brick, P. (1996). Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. Structure **4:** 287-298.

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.

Cooper, M.A., and Williams, D.H. (1999). Binding of glycopeptide antibiotics to a model of a vancomycin-resistant bacterium. Chem Biol **6:** 891-899.

Cowtan, K.D., and Main, P. (1993). Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. Acta Crystallogr D Biol Crystallogr **49:** 148-157.

Cryle, M.J., and Schlichting, I. (2008). Structural insights from a P450 Carrier Protein complex reveal how specificity is achieved in the P450(BioI) ACP complex. Proceedings of the National Academy of Sciences of the United States of America **105:** 15696-15701.

Cukovic, D., Ehlting, J., VanZiffle, J.A., and Douglas, C.J. (2001). Structure and evolution of 4-coumarate:coenzyme A ligase (4CL) gene families. Biol Chem **382:** 645-654.

Cunha, B.A. (2006). Antimicrobial therapy of multidrug-resistant Streptococcus pneumoniae, vancomycin-resistant enterococci, and methicillin-resistant Staphylococcus aureus. Medical Clinics of North America **90:** 1165-1182.

De, N., Navarro, M.V., Raghavan, R.V., and Sondermann, H. (2009). Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. J. Mol. Biol. **393:** 619-633.

De, N., Pirruccello, M., Krasteva, P.V., Bae, N., Raghavan, R.V., and Sondermann, H. (2008). Phosphorylation-independent regulation of the diguanylate cyclase WspR. PLoS Biol. **6:** e67.

Dekker, F.J., Koch, M.A., and Waldmann, H. (2005). Protein structure similarity clustering (PSSC) and natural product structure as inspiration sources for drug development and chemical genomics. Curr Opin Chem Biol **9:** 232-239.

Denisov, I.G., Makris, T.M., Sligar, S.G., and Schlichting, I. (2005). Structure and chemistry of cytochrome P450. Chemical Reviews **105:** 2253-2277.

Donadio, S., Sosio, M., Stegmann, E., Weber, T., and Wohlleben, W. (2005). Comparative analysis and insights into the evolution of gene clusters for glycopeptide antibiotic biosynthesis. Mol Genet Genomics **274:** 40-50.

Dong, S.D., Oberthur, M., Losey, H.C., Anderson, J.W., Eggert, U.S., Peczuh, M.W., Walsh, C.T., and Kahne, D. (2002). The structural basis for induction of VanB resistance. J Am Chem Soc **124:** 9064-9065.

Doroghazi, J.R., Ju, K.S., Brown, D.W., Labeda, D.P., Deng, Z., Metcalf, W.W., Chen, W., and Price, N.P. (2011). Genome sequences of three tunicamycin-producing Streptomyces Strains, S. chartreusis NRRL 12338, S. chartreusis NRRL 3882, and S. lysosuperificus ATCC 31396. J Bacteriol **193:** 7021-7022.

Douglas, C., Hoffmann, H., Schulz, W., and Hahlbrock, K. (1987). Structure and elicitor or u.v.-light-stimulated expression of two 4-coumarate:CoA ligase genes in parsley. EMBO J **6:** 1189-1195.

Dow, J.M., Fouhy, Y., Lucey, J.F., and Ryan, R.P. (2006). The HD-GYP domain, cyclic di-GMP signaling, and bacterial virulence to plants. Mol. Plant. Microbe. Interact. **19:** 1378-1384.

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.

Duerig, A., Abel, S., Folcher, M., Nicollier, M., Schwede, T., Amiot, N., Giese, B., and Jenal, U. (2009). Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression. Genes Dev. **23:** 93-104.

Ehlting, J., Buttner, D., Wang, Q., Douglas, C.J., Somssich, I.E., and Kombrink, E. (1999). Three 4-coumarate:coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms. Plant J **19:** 9-20.

Fineran, P.C., Williamson, N.R., Lilley, K.S., and Salmond, G.P. (2007). Virulence and prodigiosin antibiotic biosynthesis in Serratia are regulated pleiotropically by the GGDEF/EAL domain protein, PigX. J. Bacteriol. **189:** 7653-7662.

Francis, S.H., Bessay, E.P., Kotera, J., Grimes, K.A., Liu, L., Thompson, W.J., and Corbin, J.D. (2002). Phosphorylation of isolated human phosphodiesterase-5 regulatory domain induces an apparent conformational change and increases cGMP binding affinity. J. Biol. Chem. **277:** 47581-47587.

Friedman, L., and Kolter, R. (2004). Genes involved in matrix formation in Pseudomonas aeruginosa PA14 biofilms. Mol. Microbiol. **51:** 675-690.

Geromichalos, G.D. (2007). Importance of molecular computer modeling in anticancer drug development. J BUON **12 Suppl 1:** S101-118.

Gimeno, P., Besacier, F., Bottex, M., Dujourdy, L., and Chaudron-Thozet, H. (2005). A study of impurities in intermediates and 3,4-methylenedioxymethamphetamine (MDMA) samples produced via reductive amination routes. Forensic Sci Int **155:** 141-157.

Gjermansen, M., Nilsson, M., Yang, L., and Tolker-Nielsen, T. (2010). Characterization of starvation-induced dispersion in Pseudomonas putida biofilms: genetic elements and molecular mechanisms. Mol. Microbiol. **75:** 815-826.

Gocht, M., and Marahiel, M.A. (1994). Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthetase TycA by site-directed mutagenesis. J Bacteriol **176:** 2654-2662.

Goldstein, B.P., Selva, E., Gastaldo, L., Berti, M., Pallanza, R., Ripamonti, F., Ferrari, P., Denaro, M., Arioli, V., and Cassani, G. (1987). A40926, a new glycopeptide antibiotic with anti-Neisseria activity. Antimicrob Agents Chemother **31:** 1961-1966.

Gotoh, O. (1992). Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. Journal of Biological Chemistry **267:** 83-90.

Graham, S.E., and Peterson, J.A. (1999). How Similar Are P450s and What Can Their Differences Teach Us? Archives of Biochemistry and Biophysics **369:** 24-29.

Grosse-Kunstleve, R.W., and Adams, P.D. (2003). Substructure search procedures for macromolecular structures. Acta Crystallogr D Biol Crystallogr **59:** 1966-1973.

Gruez, A., Libiad, M., Boschi-Muller, S., and Branlant, G. (2010). Structural and biochemical characterization of free methionine-R-sulfoxide reductase from Neisseria meningitidis. J. Biol. Chem. **285:** 25033-25043.

Gulick, A.M. (2009). Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. ACS Chem Biol **4:** 811-827.

Gulick, A.M., Lu, X., and Dunaway-Mariano, D. (2004). Crystal structure of 4-chlorobenzoate:CoA ligase/synthetase in the unliganded and aryl substrate-bound states. Biochemistry **43:** 8670-8679.

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.

Guzzo, C.R., Salinas, R.K., Andrade, M.O., and Farah, C.S. (2009). PILZ protein structure and interactions with PILB and the FIMX EAL domain: implications for control of type IV pilus biogenesis. J. Mol. Biol. **393:** 848-866.

Hadatsch, B., Butz, D., Schmiederer, T., Steudle, J., Wohlleben, W., Sussmuth, R., and Stegmann, E. (2007). The biosynthesis of teicoplanin-type glycopeptide antibiotics: assignment of p450 mono-oxygenases to side chain cyclizations of glycopeptide a47934. Chemistry and Biology **14:** 1078-1089.

Hahlbrock, K., and Scheel, D. (1989). Physiology and molecular biology of phenylpropanoid metabolism. Annu. Rev. Plant Physiol. Plant Mol. Biol. **40:** 347-369.

Hamberger, B., and Hahlbrock, K. (2004). The 4-coumarate:CoA ligase gene family in Arabidopsis thaliana comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proc Natl Acad Sci U S A **101:** 2209-2214.

Hamilton, G.R., and Baskett, T.F. (2000). In the arms of Morpheus the development of morphine for postoperative pain relief. Can J Anaesth **47:** 367-374.

Handa, N., Mizohata, E., Kishishita, S., Toyama, M., Morita, S., Uchikubo-Kamo, T., Akasaka, R., Omori, K., Kotera, J., Terada, T., Shirouzu, M., and Yokoyama, S. (2008). Crystal structure of the GAF-B domain from human phosphodiesterase 10A complexed with its ligand, cAMP. J. Biol. Chem. **283:** 19657-19664.

Hebert, M.C., Schwede, F., Jastorff, B., and Cote, R.H. (1998). Structural features of the noncatalytic cGMP binding sites of frog photoreceptor phosphodiesterase using cGMP analogs. J. Biol. Chem. **273:** 5557-5565.

Heikaus, C.C., Pandit, J., and Klevit, R.E. (2009). Cyclic nucleotide binding GAF domains from phosphodiesterases: structural and mechanistic insights. Structure **17:** 1551-1557.

Heikaus, C.C., Stout, J.R., Sekharan, M.R., Eakin, C.M., Rajagopal, P., Brzovic, P.S., Beavo, J.A., and Klevit, R.E. (2008). Solution structure of the cGMP binding GAF domain from phosphodiesterase 5: insights into nucleotide specificity, dimerization, and cGMP-dependent conformational change. J. Biol. Chem. **283:** 22749-22759.

Hengge, R. (2009). Principles of c-di-GMP signalling in bacteria. Nat. Rev. Microbiol. **7:** 263-273.

Hickman, J.W., and Harwood, C.S. (2008). Identification of FleQ from Pseudomonas aeruginosa as a c-di-GMP-responsive transcription factor. Mol. Microbiol. **69:** 376-389.

Hickman, J.W., Tifrea, D.F., and Harwood, C.S. (2005). A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. Proc. Natl. Acad. Sci. USA **102:** 14422-14427.

Hill, H.A.O., Röder, A., and Williams, R.J.P. (1970). Cytochrome P-450 suggestions as to the structure and mechanism of action. Naturwissenschaften **57:** 69-72.

Hiramatsu, K., Hanaki, H., Ino, T., Yabuta, K., Oguri, T., and Tenover, F.C. (1997). Methicillin-resistant Staphylococcus aureus clinical strain with reduced vancomycin susceptibility. Journal of Antimicrobial Chemotherapy **40:** 135-136.

Ho, Y.S., Burden, L.M., and Hurley, J.H. (2000). Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. EMBO J. **19:** 5288-5299.

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

Horswill, A.R., and Escalante-Semerena, J.C. (2002). Characterization of the propionyl-CoA synthetase (PrpE) enzyme of Salmonella enterica: residue Lys592 is required for propionyl-AMP synthesis. Biochemistry **41:** 2379-2387.

Hu, W.J., Kawaoka, A., Tsai, C.J., Lung, J., Osakabe, K., Ebinuma, H., and Chiang, V.L. (1998). Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (Populus tremuloides). Proc Natl Acad Sci U S A **95:** 5407-5412.

Hu, W.J., Harding, S.A., Lung, J., Popko, J.L., Ralph, J., Stokke, D.D., Tsai, C.J., and Chiang, V.L. (1999). Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nat Biotechnol **17:** 808-812.

Hu, Y., Gai, Y., Yin, L., Wang, X., Feng, C., Feng, L., Li, D., Jiang, X.N., and Wang, D.C. (2010). Crystal structures of a Populus tomentosa 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. Plant Cell **22:** 3093-3104.

Huang, D., Hinds, T.R., Martinez, S.E., Doneanu, C., and Beavo, J.A. (2004). Molecular determinants of cGMP binding to chicken cone photoreceptor phosphodiesterase. J. Biol. Chem. **279:** 48143-48151.

Huang, Y.H., Liu, X.Y., Du, X.X., Jiang, Z.F., and Su, X.D. (2012). The structural basis for the sensing and binding of cyclic di-GMP by STING. Nat Struct Mol Biol **19:** 728-730.

Hubbard, B.K., and Walsh, C.T. (2003). Vancomycin assembly: nature's way. Angewandte Chemie. International Ed. In English **42:** 730-765.

Hughes, A.J., and Keatinge-Clay, A. (2011). Enzymatic extender unit generation for in vitro polyketide synthase reactions: structural and functional showcasing of Streptomyces coelicolor MatB. Chem Biol **18:** 165-176.

Iffland, A., Kohls, D., Low, S., Luan, J., Zhang, Y., Kothe, M., Cao, Q., Kamath, A.V., Ding, Y.H., and Ellenberger, T. (2005). Structural determinants for inhibitor specificity and selectivity in PDE2A using the wheat germ in vitro translation system. Biochemistry **44:** 8312-8325.

Jefcoate, C.R. (1978). Measurement of substrate and inhibitor binding to microsomal cytochrome P-450 by optical-difference spectroscopy. Methods in Enzymology **52:** 258-279.

Jenal, U., and Malone, J. (2006). Mechanisms of cyclic-di-GMP signaling in bacteria. Annu. Rev. Genet. **40:** 385-407.

Jogl, G., and Tong, L. (2004). Crystal structure of yeast acetyl-coenzyme A synthetase in complex with AMP. Biochemistry **43:** 1425-1431.

Kader, A., Simm, R., Gerstel, U., Morr, M., and Romling, U. (2006). Hierarchical involvement of various GGDEF domain proteins in rdar morphotype development of Salmonella enterica serovar Typhimurium. Mol. Microbiol. **60:** 602-616.

Kahne, D., Leimkuhler, C., Lu, W., and Walsh, C. (2005). Glycopeptide and lipoglycopeptide antibiotics. Chemical Reviews **105:** 425-448.

Kajita, S., Hishiyama, S., Tomimura, Y., Katayama, Y., and Omori, S. (1997). Structural Characterization of Modified Lignin in Transgenic Tobacco Plants in Which the Activity of 4-Coumarate:Coenzyme A Ligase Is Depressed. Plant Physiol **114:** 871-879.

Kanimozhi, G., Prasad, N.R., Ramachandran, S., and Pugalendi, K.V. (2011). Umbelliferone modulates gamma-radiation induced reactive oxygen species generation and subsequent oxidative damage in human blood lymphocytes. Eur J Pharmacol **672:** 20-29.

Kleywegt, G.J., and Brunger, A.T. (1996). Checking your imagination: applications of the free R value. Structure **4:** 897-904.

Knobloch, K.H., and Hahlbrock, K. (1975). Isoenzymes of p-coumarate: CoA ligase from cell suspension cultures of Glycine max. Eur J Biochem **52:** 311-320.

Ko, J., Ryu, K.S., Kim, H., Shin, J.S., Lee, J.O., Cheong, C., and Choi, B.S. (2010). Structure of PP4397 reveals the molecular basis for different c-di-GMP binding modes by Pilz domain proteins. J. Mol. Biol. **398:** 97-110.

Kochan, G., Pilka, E.S., von Delft, F., Oppermann, U., and Yue, W.W. (2009). Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A. J Mol Biol **388:** 997-1008.

Koehn, F.E., and Carter, G.T. (2005). The evolving role of natural products in drug discovery. Nat Rev Drug Discov **4:** 206-220.

Krasteva, P.V., Fong, J.C., Shikuma, N.J., Beyhan, S., Navarro, M.V., Yildiz, F.H., and Sondermann, H. (2010). Vibrio cholerae VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. Science **327:** 866-868.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. **372:** 774-797.

Kruger, R.G., Lu, W., Oberthur, M., Tao, J., Kahne, D., and Walsh, C.T. (2005). Tailoring of glycopeptide scaffolds by the acyltransferases from the teicoplanin and A-40,926 biosynthetic operons. Chem Biol **12:** 131-140.

Kulesekara, H., Lee, V., Brencic, A., Liberati, N., Urbach, J., Miyata, S., Lee, D.G., Neely, A.N., Hyodo, M., Hayakawa, Y., Ausubel, F.M., and Lory, S. (2006). Analysis of Pseudomonas aeruginosa diguanylate cyclases and phosphodiesterases reveals a role for bis-(3'-5')-cyclic-GMP in virulence. Proc. Natl. Acad. Sci. USA **103:** 2839-2844.

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

Leduc, J.L., and Roberts, G.P. (2009). Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of Xanthomonas axonopodis pv. citri. J. Bacteriol. **191:** 7121-7122.

Lee, D., and Douglas, C.J. (1996). Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family. cDNA structure, gene inheritance and expression, and properties of recombinant proteins. Plant Physiol. **112:** 193-205.

Lee, D., Meyer, K., Chapple, C., and Douglas, C.J. (1997). Antisense suppression of 4-coumarate:coenzyme A ligase activity in Arabidopsis leads to altered lignin subunit composition. Plant Cell **9:** 1985-1998.

Lee, J.H., Bae, B., Kuemin, M., Circello, B.T., Metcalf, W.W., Nair, S.K., and van der Donk, W.A. (2010). Characterization and structure of DhpI, a phosphonate O-methyltransferase involved in dehydrophos biosynthesis. Proc Natl Acad Sci U S A **107:** 17557-17562.

Lee, J.M., Cho, H.Y., Cho, H.J., Ko, I.J., Park, S.W., Baik, H.S., Oh, J.H., Eom, C.Y., Kim, Y.M., Kang, B.S., and Oh, J.I. (2008). O2- and NO-sensing mechanism through the DevSR two-component system in Mycobacterium smegmatis. J. Bacteriol. **190:** 6795-6804.

Lee, V.T., Matewish, J.M., Kessler, J.L., Hyodo, M., Hayakawa, Y., and Lory, S. (2007). A cyclic-di-GMP receptor required for bacterial exopolysaccharide production. Mol. Microbiol. **65:** 1474-1484.

Li, J.W., and Vederas, J.C. (2009). Drug discovery and natural products: end of an era or an endless frontier? Science **325:** 161-165.

Li, M.H., Ung, P.M., Zajkowski, J., Garneau-Tsodikova, S., and Sherman, D.H. (2009). Automated genome mining for natural products. BMC Bioinformatics **10:** 185.

Li, T.L., Huang, F., Haydock, S.F., Mironenko, T., Leadlay, P.F., and Spencer, J.B. (2004). Biosynthetic gene cluster of the glycopeptide antibiotic teicoplanin: characterization of two glycosyltransferases and the key acyltransferase. Chemistry and Biology **11:** 107-119.

Li, Z., Rupasinghe, S.G., Schuler, M.A., and Nair, S.K. (2011). Crystal structure of a phenol-coupling P450 monooxygenase involved in teicoplanin biosynthesis. Proteins **79:** 1728-1738.

Li, Z., Chen, J.H., Hao, Y., and Nair, S.K. (2012). Structures of the PelD cyclic diguanylate effector involved in pellicle formation in Pseudomonas aeruginosa PAO1. J Biol Chem **287:** 30191-30204.

Lin, Z., Johnson, L.C., Weissbach, H., Brot, N., Lively, M.O., and Lowther, W.T. (2007). Free methionine-(R)-sulfoxide reductase from Escherichia coli reveals a new GAF domain function. Proc. Natl. Acad. Sci. USA **104:** 9597-9602.

Lindermayr, C., Fliegmann, J., and Ebel, J. (2003). Deletion of a single amino acid residue from different 4-coumarate:CoA ligases from soybean results in the generation of new substrate specificities. J Biol Chem **278:** 2781-2786.

Lindermayr, C., Mollers, B., Fliegmann, J., Uhlmann, A., Lottspeich, F., Meimberg, H., and Ebel, J. (2002). Divergent members of a soybean (Glycine max L.) 4-coumarate:coenzyme A ligase gene family. Eur J Biochem **269:** 1304-1315.

Malabarba, A., and Goldstein, B.P. (2005). Origin, structure, and activity in vitro and in vivo of dalbavancin. Journal of Antimicrobial Chemotherapy **55 Suppl 2:** ii15-20.

Malabarba, A., Nicas, T.I., and Thompson, R.C. (1997). Structural modifications of glycopeptide antibiotics. Med Res Rev **17:** 69-137.

Malabarba, A., Ferrari, P., Gallo, G.G., Kettenring, J., and Cavalleri, B. (1986). Teicoplanin, antibiotics from Actinoplanes teichomyceticus nov. sp. VII.

Preparation and NMR characteristics of the aglycone of teicoplanin. Journal of Antibiotics **39:** 1430-1442.

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

Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Tasneem, A., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., and Bryant, S.H. (2009). CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res. **37:** D205-210.

Martinez, S.E., Heikaus, C.C., Klevit, R.E., and Beavo, J.A. (2008). The structure of the GAF A domain from phosphodiesterase 6C reveals determinants of cGMP binding, a conserved binding surface, and a large cGMP-dependent conformational change. J. Biol. Chem. **283:** 25913-25919.

Martinez, S.E., Wu, A.Y., Glavas, N.A., Tang, X.B., Turley, S., Hol, W.G., and Beavo, J.A. (2002). The two GAF domains in phosphodiesterase 2A have distinct roles in dimerization and in cGMP binding. Proc. Natl. Acad. Sci. USA **99:** 13260-13265.

Martinez, S.E., Bruder, S., Schultz, A., Zheng, N., Schultz, J.E., Beavo, J.A., and Linder, J.U. (2005). Crystal structure of the tandem GAF domains from a cyanobacterial adenylyl cyclase: modes of ligand binding and dimerization. Proc. Natl. Acad. Sci. USA **102:** 3082-3087.

Martins, T.J., Mumby, M.C., and Beavo, J.A. (1982). Purification and characterization of a cyclic GMP-stimulated cyclic nucleotide phosphodiesterase from bovine tissues. J. Biol. Chem. **257:** 1973-1979.

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.

McElroy, W.D., DeLuca, M., and Travis, J. (1967). Molecular uniformity in biological catalyses. The enzymes concerned with firefly luciferin, amino acid, and fatty acid utilization are compared. Science **157:** 150-160.

McRee, D.E. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. J Struct Biol **125:** 156-165.

Merighi, M., Lee, V.T., Hyodo, M., Hayakawa, Y., and Lory, S. (2007). The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in Pseudomonas aeruginosa. Mol. Microbiol. **65:** 876-895.

Miller, L.H., and Su, X. (2011). Artemisinin: discovery from the Chinese herbal garden. Cell **146:** 855-858.

Moellering, R.C., Jr. (2006). Vancomycin: a 50-year reassessment. Clinical Infectious Diseases **42 Suppl 1:** S3-4.

Monds, R.D., Newell, P.D., Gross, R.H., and O'Toole, G.A. (2007). Phosphate-dependent modulation of c-di-GMP levels regulates Pseudomonas fluorescens Pf0-1 biofilm formation by controlling secretion of the adhesin LapA. Mol. Microbiol. **63:** 656-679.

Murray, B.E. (2000). Vancomycin-resistant enterococcal infections. New England Journal of Medicine **342:** 710-721.

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.

Nair, S.K., and van der Donk, W.A. (2011). Structure and mechanism of enzymes involved in biosynthesis and breakdown of the phosphonates fosfomycin, dehydrophos, and phosphinothricin. Arch Biochem Biophys **505:** 13-21.

Nakatsu, T., Ichiyama, S., Hiratake, J., Saldanha, A., Kobashi, N., Sakata, K., and Kato, H. (2006). Structural basis for the spectral difference in luciferase bioluminescence. Nature **440:** 372-376.

Navarro, M.V., De, N., Bae, N., Wang, Q., and Sondermann, H. (2009). Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX. Structure **17:** 1104-1116.

Navarro, M.V., Newell, P.D., Krasteva, P.V., Chatterjee, D., Madden, D.R., O'Toole, G.A., and Sondermann, H. (2011). Structural basis for c-di-GMP-mediated inside-out signaling controlling periplasmic proteolysis. PLoS Biol. **9:** e1000588.

Neuwald, A.F., and Landsman, D. (1997). GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. Trends Biochem Sci **22:** 154-155.

Newell, P.D., Monds, R.D., and O'Toole, G.A. (2009). LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by Pseudomonas fluorescens Pf0-1. Proc. Natl. Acad. Sci. USA **106:** 3461-3466.

Newman, D.J., and Cragg, G.M. (2012). Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod **75:** 311-335.

Newman, D.J., Cragg, G.M., and Snader, K.M. (2003). Natural products as sources of new drugs over the period 1981-2002. J Nat Prod **66:** 1022-1037.

Obel, N., and Scheller, H.V. (2000). Enzymatic synthesis and purification of caffeoyl-CoA, p-coumaroyl-CoA, and feruloyl-CoA. Anal Biochem **286:** 38-44.

Oshima, R., Fushinobu, S., Su, F., Zhang, L., Takaya, N., and Shoun, H. (2004). Structural evidence for direct hydride transfer from NADH to cytochrome P450nor. Journal of Molecular Biology **342:** 207-217.

Osman, K.T., Du, L., He, Y., and Luo, Y. (2009). Crystal structure of Bacillus cereus D-alanyl carrier protein ligase (DltA) in complex with ATP. J Mol Biol **388:** 345-355.

Otwinowski, Z., Borek, D., Majewski, W., and Minor, W. (2003). Multiparametric scaling of diffraction intensities. Acta Crystallogr A **59:** 228-234.

Ouyang, S., Song, X., Wang, Y., Ru, H., Shaw, N., Jiang, Y., Niu, F., Zhu, Y., Qiu, W., Parvatiyar, K., Li, Y., Zhang, R., Cheng, G., and Liu, Z.J. (2012). Structural analysis of the STING adaptor protein reveals a hydrophobic dimer interface and mode of cyclic di-GMP binding. Immunity **36:** 1073-1086.

Pandit, J., Forman, M.D., Fennell, K.F., Dillman, K.S., and Menniti, F.S. (2009). Mechanism for the allosteric regulation of phosphodiesterase 2A deduced from the X-ray structure of a near full-length construct. Proc. Natl. Acad. Sci. USA **106:** 18225-18230.

Park, S.Y., Shimizu, H., Adachi, S., Nakagawa, A., Tanaka, I., Nakahara, K., Shoun, H., Obayashi, E., Nakamura, H., Iizuka, T., and Shiro, Y. (1997). Crystal structure of nitric oxide reductase from denitrifying fungus Fusarium oxysporum. Nature Structural Biology **4:** 827-832.

Parsek, M.R., and Singh, P.K. (2003). Bacterial biofilms: an emerging link to disease pathogenesis. Annu. Rev. Microbiol. **57:** 677-701.

Paul, R., Weiser, S., Amiot, N.C., Chan, C., Schirmer, T., Giese, B., and Jenal, U. (2004). Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. Genes Dev. **18:** 715-727.

Pelzer, S., Reichert, W., Huppert, M., Heckmann, D., and Wohlleben, W. (1997). Cloning and analysis of a peptide synthetase gene of the balhimycin producer Amycolatopsis mediterranei DSM5908 and development of a gene disruption/replacement system. Journal of Biotechnology **56:** 115-128.

Pelzer, S., Sussmuth, R., Heckmann, D., Recktenwald, J., Huber, P., Jung, G., and Wohlleben, W. (1999). Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in Amycolatopsis mediterranei DSM5908. Antimicrobial Agents and Chemotherapy **43:** 1565-1573.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat Struct Biol **6:** 458-463.

Podust, L.M., Ioanoviciu, A., and Ortiz de Montellano, P.R. (2008). 2.3 A X-ray structure of the heme-bound GAF domain of sensory histidine kinase DosT of Mycobacterium tuberculosis. Biochemistry **47:** 12523-12531.

Pootoolal, J., Thomas, M.G., Marshall, C.G., Neu, J.M., Hubbard, B.K., Walsh, C.T., and Wright, G.D. (2002). Assembling the glycopeptide antibiotic scaffold: The biosynthesis of A47934 from Streptomyces toyocaensis NRRL15009. Proceedings of the National Academy of Sciences of the United States of America **99:** 8962-8967.

Puk, O., Huber, P., Bischoff, D., Recktenwald, J., Jung, G., Sussmuth, R.D., van Pee, K.H., Wohlleben, W., and Pelzer, S. (2002). Glycopeptide biosynthesis in Amycolatopsis mediterranei DSM5908: function of a halogenase and a haloperoxidase/perhydrolase. Chemistry and Biology **9:** 225-235.

Pylypenko, O., Vitali, F., Zerbe, K., Robinson, J.A., and Schlichting, I. (2003). Crystal structure of OxyC, a cytochrome P450 implicated in an oxidative C-C coupling reaction during vancomycin biosynthesis. Journal of Biological Chemistry **278:** 46727-46733.

Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick, W.J., Jr., Hallett, J.P., Leak, D.J., Liotta, C.L., Mielenz, J.R., Murphy, R., Templer, R., and Tschaplinski, T. (2006). The path forward for biofuels and biomaterials. Science **311:** 484-489.

Ranjeva, R., Boudet, A.M., and Faggion, R. (1976). Phenolic metabolism in petunia tissues. IV. - Properties of p-coumarate : coenzyme A ligase isoenzymes. Biochimie **58:** 1255-1262.

Rao, F., Yang, Y., Qi, Y., and Liang, Z.X. (2008). Catalytic mechanism of cyclic di-GMP-specific phosphodiesterase: a study of the EAL domain-containing RocR from Pseudomonas aeruginosa. J. Bacteriol. **190:** 3622-3631.

Reger, A.S., Carney, J.M., and Gulick, A.M. (2007). Biochemical and crystallographic analysis of substrate binding and conformational changes in acetyl-CoA synthetase. Biochemistry **46:** 6536-6546.

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.

Richard, T., Pawlus, A.D., Iglesias, M.L., Pedrot, E., Waffo-Teguo, P., Merillon, J.M., and Monti, J.P. (2011). Neuroprotective properties of resveratrol and derivatives. Ann N Y Acad Sci **1215:** 103-108.

Ryan, R.P., Fouhy, Y., Lucey, J.F., Crossman, L.C., Spiro, S., He, Y.W., Zhang, L.H., Heeb, S., Camara, M., Williams, P., and Dow, J.M. (2006). Cell-cell signaling in Xanthomonas campestris involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. Proc. Natl. Acad. Sci. USA **103:** 6712-6717.

Rybalkin, S.D., Yan, C., Bornfeldt, K.E., and Beavo, J.A. (2003a). Cyclic GMP phosphodiesterases and regulation of smooth muscle function. Circ. Res. **93:** 280-291.

Rybalkin, S.D., Rybalkina, I.G., Shimizu-Albergine, M., Tang, X.B., and Beavo, J.A. (2003b). PDE5 is converted to an activated state upon cGMP binding to the GAF A domain. EMBO J. **22:** 469-478.

Ryder, C., Byrd, M., and Wozniak, D.J. (2007). Role of polysaccharides in Pseudomonas aeruginosa biofilm development. Curr. Opin. Microbiol. **10:** 644-648.

Ryjenkov, D.A., Tarutina, M., Moskvin, O.V., and Gomelsky, M. (2005). Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. J. Bacteriol. **187:** 1792-1798.

Ryjenkov, D.A., Simm, R., Romling, U., and Gomelsky, M. (2006). The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria. J. Biol. Chem. **281:** 30310-30314.

Saraste, M., Sibbald, P.R., and Wittinghofer, A. (1990). The P-loop--a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci **15:** 430-434.

Schirmer, T., and Jenal, U. (2009). Structural and mechanistic determinants of c-di-GMP signalling. Nat. Rev. Microbiol. **7:** 724-735.

Schmidt, A.J., Ryjenkov, D.A., and Gomelsky, M. (2005). The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. J. Bacteriol. **187:** 4774-4781.

Schneider, K., Hovel, K., Witzel, K., Hamberger, B., Schomburg, D., Kombrink, E., and Stuible, H.P. (2003). The substrate specificity-determining amino acid code of 4-coumarate:CoA ligase. Proc Natl Acad Sci U S A **100:** 8601-8606.

Shang, G., Zhu, D., Li, N., Zhang, J., Zhu, C., Lu, D., Liu, C., Yu, Q., Zhao, Y., Xu, S., and Gu, L. (2012). Crystal structures of STING protein reveal basis for recognition of cyclic di-GMP. Nat Struct Mol Biol **19:** 725-727.

Shoichet, B.K., McGovern, S.L., Wei, B., and Irwin, J.J. (2002). Lead discovery using molecular docking. Curr Opin Chem Biol **6:** 439-446.

Shu, C., Yi, G., Watts, T., Kao, C.C., and Li, P. (2012). Structure of STING bound to cyclic di-GMP reveals the mechanism of cyclic dinucleotide recognition by the immune system. Nat Struct Mol Biol **19:** 722-724.

Siehl, D.L., Castle, L.A., Gorton, R., and Keenan, R.J. (2007). The molecular basis of glyphosate resistance by an optimized microbial acetyltransferase. J Biol Chem **282:** 11446-11455.

Sieradzki, K., Roberts, R.B., Haber, S.W., and Tomasz, A. (1999). The development of vancomycin resistance in a patient with methicillin-resistant Staphylococcus aureus infection. New England Journal of Medicine **340:** 517-523.

Smith, K.D., Lipchock, S.V., Livingston, A.L., Shanahan, C.A., and Strobel, S.A. (2010). Structural and biochemical determinants of ligand binding by the c-di-GMP riboswitch. Biochemistry **49:** 7351-7359.

Smith, K.D., Shanahan, C.A., Moore, E.L., Simon, A.C., and Strobel, S.A. (2011). Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches. Proc. Natl. Acad. Sci. USA **108:** 7757-7762.

Smith, K.D., Lipchock, S.V., Ames, T.D., Wang, J., Breaker, R.R., and Strobel, S.A. (2009). Structural basis of ligand binding by a c-di-GMP riboswitch. Nat. Struct. Mol. Biol. **16:** 1218-1223.

Sommerfeldt, N., Possling, A., Becker, G., Pesavento, C., Tschowri, N., and Hengge, R. (2009). Gene expression patterns and differential input into curli fimbriae regulation of all GGDEF/EAL domain proteins in Escherichia coli. Microbiology **155:** 1318-1331.

Sosio, M., Stinchi, S., Beltrametti, F., Lazzarini, A., and Donadio, S. (2003). The gene cluster for the biosynthesis of the glycopeptide antibiotic A40926 by nonomuraea species. Chemistry and Biology **10:** 541-549.

Sosio, M., Kloosterman, H., Bianchi, A., de Vreugd, P., Dijkhuizen, L., and Donadio, S. (2004). Organization of the teicoplanin gene cluster in Actinoplanes teichomyceticus. Microbiology **150:** 95-102.

Stegmann, E., Pelzer, S., Bischoff, D., Puk, O., Stockert, S., Butz, D., Zerbe, K., Robinson, J., Sussmuth, R.D., and Wohlleben, W. (2006). Genetic analysis of the balhimycin (vancomycin-type) oxygenase genes. Journal of Biotechnology **124:** 640-653.

Stuible, H., Buttner, D., Ehlting, J., Hahlbrock, K., and Kombrink, E. (2000). Mutational analysis of 4-coumarate:CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. FEBS Lett **467:** 117-122.

Stuible, H.P., and Kombrink, E. (2001). Identification of the substrate specificity-conferring amino acid residues of 4-coumarate:coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. J Biol Chem **276:** 26893-26897.

Sudarsan, N., Lee, E.R., Weinberg, Z., Moy, R.H., Kim, J.N., Link, K.H., and Breaker, R.R. (2008). Riboswitches in eubacteria sense the second messenger cyclic di-GMP. Science **321:** 411-413.

Sung, B.J., Hwang, K.Y., Jeon, Y.H., Lee, J.I., Heo, Y.S., Kim, J.H., Moon, J., Yoon, J.M., Hyun, Y.L., Kim, E., Eum, S.J., Park, S.Y., Lee, J.O., Lee, T.G., Ro, S., and Cho, J.M. (2003). Structure of the catalytic domain of human phosphodiesterase 5 with bound drug molecules. Nature **425:** 98-102.

Sussmuth, R.D., and Wohlleben, W. (2004). The biosynthesis of glycopeptide antibiotics--a model for complex, non-ribosomally synthesized, peptidic secondary metabolites. Applied Microbiology and Biotechnology **63:** 344-350.

Süssmuth, R.D., Pelzer, S., Nicholson, G., Walk, T., Wohlleben, W., and Jung, G. (1999). New Advances in the Biosynthesis of Glycopeptide Antibiotics of the Vancomycin Type from Amycolatopsis mediterranei. Angewandte Chemie. International Ed. In English **38:** 1976-1979.

Tamayo, R., Pratt, J.T., and Camilli, A. (2007). Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. Annu. Rev. Microbiol. **61:** 131-148.

Tao, F., He, Y.W., Wu, D.H., Swarup, S., and Zhang, L.H. (2010). The cyclic nucleotide monophosphate domain of Xanthomonas campestris global regulator Clp defines a new class of cyclic di-GMP effectors. J. Bacteriol. **192:** 1020-1029.

Truman, A.W., Robinson, L., and Spencer, J.B. (2006). Identification of a deacetylase involved in the maturation of teicoplanin. Chembiochem **7:** 1670-1675.

Van Bambeke, F. (2006). Glycopeptides and glycodepsipeptides in clinical development: a comparative review of their antibacterial spectrum, pharmacokinetics and clinical efficacy. Curr Opin Investig Drugs **7:** 740-749.

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. Journal of Molecular Biology **229:** 105-124.

van Montfort, R.L., and Workman, P. (2009). Structure-based design of molecular cancer therapeutics. Trends Biotechnol **27:** 315-328.

Vasseur, P., Soscia, C., Voulhoux, R., and Filloux, A. (2007). PelC is a Pseudomonas aeruginosa outer membrane lipoprotein of the OMA family of proteins involved in exopolysaccharide transport. Biochimie **89:** 903-915.

Vasseur, P., Vallet-Gely, I., Soscia, C., Genin, S., and Filloux, A. (2005). The pel genes of the Pseudomonas aeruginosa PAK strain are involved at early and late stages of biofilm formation. Microbiology **151:** 985-997.

Wagner, J.R., Zhang, J., Brunzelle, J.S., Vierstra, R.D., and Forest, K.T. (2007). High resolution structure of Deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution. J. Biol. Chem. **282:** 12298-12309.

Wallis, P.J., and Rhodes, M.J.C. (1977). Multiple forms of hydroxycinnamate : CoA ligase in etiolated pea seedlings. Phytochemistry **16:** 1891-1894.

Wang, H., Robinson, H., and Ke, H. (2010). Conformation changes, N-terminal involvement, and cGMP signal relay in the phosphodiesterase-5 GAF domain. J. Biol. Chem. **285:** 38149-38156.

Wang, H., Liu, Y., Huai, Q., Cai, J., Zoraghi, R., Francis, S.H., Corbin, J.D., Robinson, H., Xin, Z., Lin, G., and Ke, H. (2006). Multiple conformations of

phosphodiesterase-5: implications for enzyme function and drug development. J. Biol. Chem. **281:** 21469-21479.

Wassmann, P., Chan, C., Paul, R., Beck, A., Heerklotz, H., Jenal, U., and Schirmer, T. (2007). Structure of BeF3- -modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. Structure **15:** 915-927.

Weber, H., Pesavento, C., Possling, A., Tischendorf, G., and Hengge, R. (2006). Cyclic-di-GMP-mediated signalling within the sigma network of Escherichia coli. Mol. Microbiol. **62:** 1014-1034.

Woithe, K., Geib, N., Zerbe, K., Li, D.B., Heck, M., Fournier-Rousset, S., Meyer, O., Vitali, F., Matoba, N., Abou-Hadeed, K., and Robinson, J.A. (2007). Oxidative phenol coupling reactions catalyzed by OxyB: a cytochrome P450 from the vancomycin producing organism. implications for vancomycin biosynthesis. Journal of the American Chemical Society **129:** 6887-6895.

Wu, A.Y., Tang, X.B., Martinez, S.E., Ikeda, K., and Beavo, J.A. (2004). Molecular determinants for cyclic nucleotide binding to the regulatory domains of phosphodiesterase 2A. J. Biol. Chem. **279:** 37928-37938.

Xu, L.H., Fushinobu, S., Ikeda, H., Wakagi, T., and Shoun, H. (2009). Crystal structures of cytochrome P450 105P1 from Streptomyces avermitilis: conformational flexibility and histidine ligation state. Journal of Bacteriology **191:** 1211-1219.

Xu, L.H., Fushinobu, S., Takamatsu, S., Wakagi, T., Ikeda, H., and Shoun, H. (2010). Regio- and stereospecificity of filipin hydroxylation sites revealed by crystal structures of cytochrome P450 105P1 and 105D6 from Streptomyces avermitilis. Journal of Biological Chemistry **285:** 16844-16853.

Yamauchi, K., Yasuda, S., Hamada, K., Tsutsumi, Y., and Fukushima, K. (2003). Multiform biosynthetic pathway of syringyl lignin in angiosperms. Planta **216:** 496-501.

Yang, C.Y., Chin, K.H., Chuah, M.L., Liang, Z.X., Wang, A.H., and Chou, S.H. (2011). The structure and inhibition of a GGDEF diguanylate cyclase complexed with (c-di-GMP)(2) at the active site. Acta Crystallogr. D Biol. Crystallogr. **67:** 997-1008.

Yang, X., Kuk, J., and Moffat, K. (2009). Conformational differences between the Pfr and Pr states in Pseudomonas aeruginosa bacteriophytochrome. Proc. Natl. Acad. Sci. USA **106:** 15639-15644.

Yang, X., Stojkovic, E.A., Kuk, J., and Moffat, K. (2007). Crystal structure of the chromophore binding domain of an unusual bacteriophytochrome, RpBphP3, reveals residues that modulate photoconversion. Proc. Natl. Acad. Sci. USA **104:** 12571-12576.

Yin, Q., Tian, Y., Kabaleeswaran, V., Jiang, X., Tu, D., Eck, M.J., Chen, Z.J., and Wu, H. (2012). Cyclic di-GMP sensing via the innate immune signaling protein STING. Mol Cell **46:** 735-745.

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.

Zahringer, F., Massa, C., and Schirmer, T. (2011). Efficient enzymatic production of the bacterial second messenger c-di-GMP by the diguanylate cyclase YdeH from E. coli. Appl. Biochem. Biotechnol. **163:** 71-79.

Zerbe, K., Pylypenko, O., Vitali, F., Zhang, W., Rouset, S., Heck, M., Vrijbloed, J.W., Bischoff, D., Bister, B., Sussmuth, R.D., Pelzer, S., Wohlleben, W., Robinson, J.A., and Schlichting, I. (2002). Crystal structure of OxyB, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis. Journal of Biological Chemistry **277:** 47476-47485.

Zhang, X.H., and Chiang, V.L. (1997). Molecular cloning of 4-coumarate:coenzyme A ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. Plant Physiol **113:** 65-74.

Zoraghi, R., Corbin, J.D., and Francis, S.H. (2004). Properties and functions of GAF domains in cyclic nucleotide phosphodiesterases and other proteins. Mol. Pharmacol. **65:** 267-278.

Zoraghi, R., Bessay, E.P., Corbin, J.D., and Francis, S.H. (2005). Structural and functional features in human PDE5A1 regulatory domain that provide for allosteric cGMP binding, dimerization, and regulation. J. Biol. Chem. **280:** 12051-12063.