UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Structural analysis and engineering of protein-protein interfaces in natural product biosynthetic pathways

Permalink

https://escholarship.org/uc/item/6sg2b3ks

Author

Corpuz, Joshua Catungal

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Structural analysis and engineering of protein-protein interfaces in natural product
biosynthetic pathways

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Chemistry

by

Joshua Catungal Corpuz

Committee in charge:

      Professor Michael D. Burkart, Chair
      Professor Eric E. Allen
      Professor Neal K. Devaraj
      Professor Partho Ghosh
      Professor Navtej Toor

2022

The Dissertation of Joshua Catungal Corpuz is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To my family, friends, loved ones, mentors, and anyone who have supported me

during this journey.

EPIGRAPH

I always think that what we need

is sunshine and fine weather and blue air

as the most dependable remedy

Vincent van Gogh

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

NRPS .... Non-ribosomal peptide synthetase

NRP .... Non-ribosomal peptide

FAS .... Fatty acid synthase

PKS .... Polyketide synthase

PCP .... Peptidyl carrier protein

ACP .... Acyl carrier protein

PPant .... Phosphopantetheine

PPTase .... Phosphopantetheinyl transferase

CoA .... Coenzyme A

A....................................... Adenylation

A_{sub} .... Adenylation subdomain

A_{core} .... Adenylation core

CTD .... C-terminal domain

NTD .... N-terminal domain

C .... Condensation

T....................................... Thiolation

DH .... Dehydrogenation

E....................................... Epimerization

F....................................... Formylation

R .... Reductase

TE .... Thioesterase

HTH .... Helix-turn-helix

PPI .... Protein-protein interface

HPLC .... High pressure liquid chromatography

FPLC................................. Fast performance liquid chromatography

A_{210}.................................... Absorbance at 210 nm

NMR.................................. Nuclear magnetic resonance

CSP................................... Chemical shift perturbation

EM..................................... Electron microscopy

MD .... Molecular dynamics

RMSD .... Root mean squared deviation

RMSF................................ Root mean squared fluctions

AMP .... Adenosine monophosphate

ATP .... Adenosine triphosphate

PP_i.................................... Pyrophosphate

AVS................................... Adenosine vinylsulfonamide

AVSN .... Adenosine vinylsulfonamide

NADPH .... Reduced nicotinamide adenine dinucleotide phosphate

TFA .... Trifluoroacetic acid

TLC .... Thin layer chromatography

ESI .... Electrospray ionization

PCR .... Polymerase chain reaction

RCF................................... Relative centrifugal force

LB...................................... Luria-Bertani

ACKNOWLEDGEMENTS

First, I would like to acknowledge my PhD advisor Professor Michael Burkart as well as Dr. Jim La Clair for their support and mentorship over the last 4.5 years. Their guidance in and out of the lab has been and will continue to be invaluable in my professional development.

I would also like to acknowledge members of the Burkart lab for all the support over the years. The collaborators I've had in this lab really helped achieve success in our projects, and have all been fantastic to work with. My mentors, peers, and juniors all have made life in the 6th floor of Pac Hall exciting and I am glad I was able to spend my PhD with great lab mates.

I would also like to acknowledge the professors that took the time to help me apply for multiple fellowship applications, especially those that helped me achieve the NIH NRSA Fellowship in addition to getting a spot in the Molecular Biophysics Training Grant. Because of your support, I was able to be fully funded during my entire PhD.

Special shout out to the friends from my cohort that I bonded with and became close to along our journey; because of your friendship, life inside and outside of lab has been enjoyable, worthwhile, and memorable.

Finally, thank you to my family and loved ones for your never-ending support during my PhD. With your support, I was able to have unwavering focus during my PhD.

I appreciate my collaborators allowing me to use the following publications in my dissertation:

Chapter 1 is a reprint of the material as it appears in: Corpuz, J. C.; Sanlley, J.; Burkart, M. D. Protein-protein interface analysis of the non-ribosomal peptide synthetase

peptidyl carrier protein and enzymatic domains. *Systems and Synthetic Biology,* **2022**, 7 (2), 677-688. The dissertation author was the primary author of this paper.

Chapter 2 is a reprint of the material as it appears in: Corpuz, J. C.; Podust L. M.; Davis, T. D.; Jaremko M. J.; Burkart, M. D. Dynamic visualization of type II peptidyl carrier protein recognition in pyoluteorin biosynthesis. *RSC Chemical Biology,* **2020**, 1, 8-12. The dissertation author was the primary author of this paper.

Chapter 3 is currently in preparation for submission: Corpuz, J. C.*; Patel, A.*; Davis, T. D.; Podust L. M.; McCammon, J. A.; Burkart M. D. Theory and experiment demonstrate the essential role of loop dynamics in type II NRPS biomolecular recognition. *In Preparation*. **2022**. The dissertation author was the primary co-author of this paper along with Dr. Ashay Patel.

Chapter 4 is currently in preparation for submission: Sztain, T.*; Corpuz, J. C.*; Bartholow, T.; LaClair, J. J.; Jiang Z.; Sanlley, J.; Heberlig, G. W.; McCammon, J. A.; Burkart, M. D. Computer-aided engineering of a protein-protein interface in carrier protein-dependent metabolic pathways. *In Preparation.* **2022**. The dissertation author was the primary co-author of this paper along with Dr. Terra Sztain.

VITA

2017   Bachelor of Science in Biochemistry, University of California San Diego

2022   Doctor of Philosophy in Chemistry, University of California San Diego

PUBLICATIONS

Sztain, T.*; Corpuz, J. C.*; Bartholow, T.; LaClair, J. J.; Jiang Z.; Sanlley, J.; Heberlig, G. W.; McCammon, J. A.; Burkart, M. D. Computer-aided engineering of a protein-protein interface in carrier protein-dependent metabolic pathways. *In Preparation.* **2022**.

Corpuz, J. C.*; Patel, A.*; Davis, T. D.; Podust L. M.; McCammon, J. A.; Burkart M. D. Theory and experiment demonstrate the essential role of loop dynamics in type II NRPS biomolecular recognition. *In Preparation*. **2022.**

Corpuz, J. C.; Sanlley, J.; Burkart, M. D. Protein-protein interface analysis of the non-ribosomal peptide synthetase peptidyl carrier protein and enzymatic domains. *Systems and Synthetic Biology*, **2022**, 7 (2), 677-688.

Corpuz, J. C.; Podust L. M.; Davis, T. D.; Jaremko M. J.; Burkart, M. D. Dynamic visualization of type II peptidyl carrier protein recognition in pyoluteorin biosynthesis. *RSC Chemical Biology,* **2020**, 1, 8-12.

Jaremko, M. J.*; Davis, T. D.*; Corpuz, J. C.; Burkart, M. D. Type II non-ribosomal peptide synthetase proteins: structure, mechanism, and protein-protein interactions. *Natural Product Reports,* **2020**, 37 (3), 355-379.

FIELD OF STUDY

Major Field: Chemistry

Studies in Biochemistry

Professor Michael D. Burkart

ABSTRACT OF THE DISSERTATION

Structural analysis and engineering of protein-protein

interfaces in natural product biosynthetic pathways

by

Joshua Catungal Corpuz

Doctor of Philosophy in Chemistry

University of California San Diego, 2022

Professor Michael D. Burkart, Chair

Carrier protein (CP) dependent biosynthetic pathways are attractive targets for biosynthetic pathway engineering due to their modular architecture and the therapeutic relevance of their natural products. These pathways, which include the fatty acid synthase (FAS), polyketide synthase (PKS), and non-ribosomal peptide synthetase (NRPS), have been targeted for engineering through substitution of modules, domains and subdomains. This method, termed combinatorial biosynthesis, has been met with limited success due to the lack of proper protein-protein interactions between noncognate proteins. With

catalysis mediated by specific protein-protein interactions between the carrier protein and its partner enzymes, enzymology and control remain fertile ground for discovery.

Here, I investigate the biomolecular recognition between the peptidyl carrier protein (PCP) and adenylation (A) domains of type II NRPS systems as my first step in engineering these synthases. The first chapter provides a recent review of the structural biology of transient NRPS PCP and partner protein complexes to identify the specific modes of PCP recognition in the type I and type II NRPS. The second and third chapter presents a thorough structural analysis of the PCP-A domain protein-protein interface from prodigiosin and pyoluteorin biosynthesis. The PCP-A domain complexes were stabilized using a mechanism-based inhibitor, which afforded crystallization and successful structure determination of two cognate and one noncognate PCP-A domain complexes. This high-resolution information was integrated with previous NMR titration data, MD simulations, and mutagenesis studies to reveal PCP dynamics and specific protein-protein interactions that govern PCP-A domain complex formation.

The fourth chapter involves application of the previously solved PCP-A domain structures towards a development of a computational protein-protein interface design protocol to create a hybrid natural product pathway. The first PCP-A domain structure solved in this work was used towards computational design of a new interface between the acyl-carrier protein (ACP) from *Escherichia coli* fatty acid biosynthesis. The optimized computational design protocol was able to improve noncognate A domain activity by ~1600 fold through the manipulation of electrostatic interactions to create a new protein-protein interface. Through these chapters, I have proven that the coupling of biophysical

data to computational methodologies can be the next platform towards re-engineering of carrier-protein dependent pathways to create novel natural products.

CHAPTER 1. PROTEIN-PROTEIN INTERFACE ANALYSIS OF THE NON-RIBOSOMAL PEPTIDE SYNTHETASE PEPTIDYL CARRIER PROTEIN AND ENZYMATIC DOMAINS

1.1. Introduction

Non-ribosomal peptides (NRPs) are secondary metabolites biosynthesized by microbes that are small peptides that are assembled outside of ribosomal translation. NRPs can act as metal chelators, pigments, and toxins given their wide scope of structural diversity.[1] NRPs also exhibit a variety of relevant therapeutic properties, such as antibiotic, antitumor, and immunosuppressant bioactivities.[1] Commonly used NRP therapeutics include vancomycin, cyclosporin A, bleomycin A_2, and polymyxin B.[2–5] The bioactivities of these compounds can be attributed to their complex molecular scaffolds installed by the non-ribosomal peptide synthetase (NRPS).

The NRPS is a modular collection of enzymes that catalyzes the biosynthesis and modification of short peptide products. Central to NRP biosynthesis is the peptidyl carrier protein (PCP) (commonly referred to as thiolation domain), which is a small ~80 residue protein that forms a conserved 4-helix bundle (Figure 1.2).[6] Separating helices 1 and 2 is the loop 1 region, an ordered 17-22 residue loop that immediately precedes a conserved serine at the beginning of helix 2. The proper assembly of NRPs requires a series of reactions catalyzed by different NRPS domains. First, the inactive apo-PCP requires the post-translational attachment of a 4'phosphopantetheine (PPant) arm to the conserved serine residue of the PCP via a phosphopantetheinyl transferase (PPTase) to form holo-PCP (Figure 1.1A).[7] Next, the adenylation (A) domain is responsible for the activation and covalent attachment of a specific amino acid onto the holo-PCP through the adenylation and thiolation reactions (Figure 1.1B), which encompass the activation of
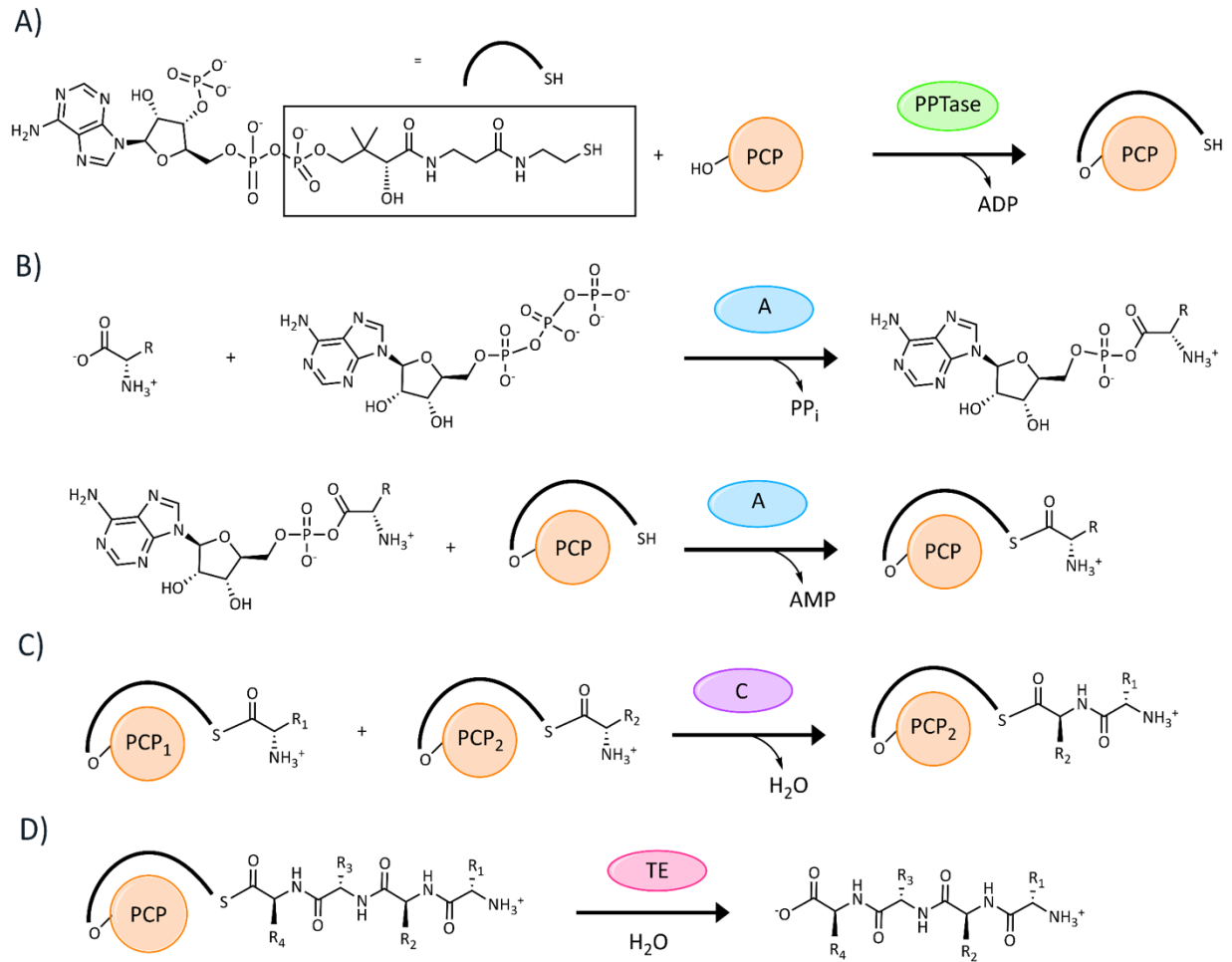
Figure 1.1: General reactions catalyzed by canonical NRPS enzymes. Reactions shown include the A) phosphopantetheinylation, B) aminoacylation, C) condensation, and D) thioesterification reactions.

the amino acid substrate with adenosine triphosphate (ATP),[6] followed by thioester linkage formation with to form the peptidyl-PCP. Once loaded, the peptidyl-PCPs from upstream and downstream NRPS modules bind to their respective donor or acceptor sites of the condensation (C) domain, which catalyzes peptide bond formation between the PCP-bound substrates (Figure 1.1C).[6] Upon reaching the termination module, the peptidyl-PCP transfers the elongated peptide chain to the thioesterase (TE), which catalyzes the hydrolysis or cyclization of the peptide product for product release (Figure 1.1D). In addition to the canonical A, C, and TE domains, tailoring domains that are fused to the enzymatic assembly line (*in cis*) or act as standalone domains (*in trans*) may also be included to install unique structural modifications to the peptide product.[8] These chemical modifications include halogenation, dehydrogenation (DH), hydroxylation, formylation (F), methylation, epimerization (E), or acylation. These functionalizations demonstrate the wide variety of partner proteins available for the PCP to generate productive protein-protein interactions to enable transformation of the nascent natural product.

The NRPS can be divided into Type I and Type II proteins. Type I NRPS proteins consist of the canonical enzymatic domains linked together in a single polypeptide chain, analogous to the Type I fatty acid synthase (FAS) and the Type I polyketide synthase (PKS).[9–11] Type II NRPS pathways exist as stand-alone proteins or di-domains that are expressed independently from multi-modular NRP biosynthetic enzymes. Type II NRPS proteins commonly exist as part of linear pathways, unlike the iterative nature of the Type II FAS and the Type II PKS. Although both Type I and II NRPS proteins contain tailoring

domains, Type II NRPS proteins commonly install diverse chemical groups, including dehydrogenated prolines, substituted aromatics, cyclopropanes, and halogenated aliphatics.[10] The Type II NRPS proteins are commonly found at the first initiation step of an NRPS pathway, which includes the A domain and PCP, and also may include subsequent tailoring domains, such as halogenation, dehydrogenation, hydroxylation, or cyclopropanation domains that may be associated with downstream Type I NRPS systems.

Due to the pharmaceutical relevance of NRPs and the modular architecture of the NRPS, NRPS biosynthesis has been a target for engineering in order to create new natural products with enhanced bioactivities. Initially, early attempts at engineering NRPS systems were met with limited success, where published combinatorial biosynthetic attempts reported low yields or no product formation.[12] Early efforts included swapping a cognate A domain with a non-cognate A domain to change the identity of the incorporated amino acid, a process coined combinatorial biosynthesis.[13] The lack of identified product formation has been attributed to many challenges, among them was the hypothesis that engineered systems may lack proper protein-protein interactions found in wild-type pathways.[14] Domain substitution with non-cognate partner proteins runs the risk of losing the specific protein-protein interactions made at the PCP-partner protein interface, potentially abrogating enzyme turnover. Recent efforts towards the re-engineering of NRPSs have included obtaining high-resolution structural data on the NRPS domains,[15] specifically within a PCP-partner protein complex to identify the exact interactions that govern protein recognition (Figure 1.2).[16] This review focuses on recent advancements in such efforts to discover the modes of PCP-partner protein recognition. Understanding the

various modes of interaction will prove critical to achieve high turnover, rationally designed NRPS biosynthesis.

1.2. Phosphopantetheinyl transferase and peptidyl carrier protein interface analysis

PPTases are essential enzymes due to their critical roles in both primary and secondary metabolism from all domains of life.[17] PPTases are responsible for a post-translational modification of FAS and PKS acyl carrier proteins (ACPs) in addition to NRPS PCPs. Due to their essential role in fatty acid synthesis, PPTases have served as a promising target for antibiotic drug development.[7] PPTases convert the inactive *apo*-carrier protein to the active *holo*-carrier protein through the covalent attachment of a 4'-phosphopantetheine moiety from coenzyme A (CoA) onto a conserved serine on all carrier proteins (Figure 1.1A), which is found at the beginning of helix 2 (Figure 1.2). In each carrier protein-dependent system, the PPant arm on the *holo*-carrier protein allows for the covalent tethering of the carboxylic acid substrate in the form of a thioester linkage, and the tethered substrate may then be shuttled by the carrier protein to various enzymatic domains for subsequent modifications and incorporation into the final natural product.

Due to their ability to attach CoA substrates directly onto the carrier protein, PPTases have been subjected to distinct applications in the field of biotechnology, which include attachment of fluorophores, chemical crosslinkers, and solid supports onto carrier proteins.[7] The PPTase from *Bacillus subtilis*, Sfp, demonstrates wide substrate scope with respect to both protein and CoA substrates.[7] Sfp has therefore been a heavily utilized
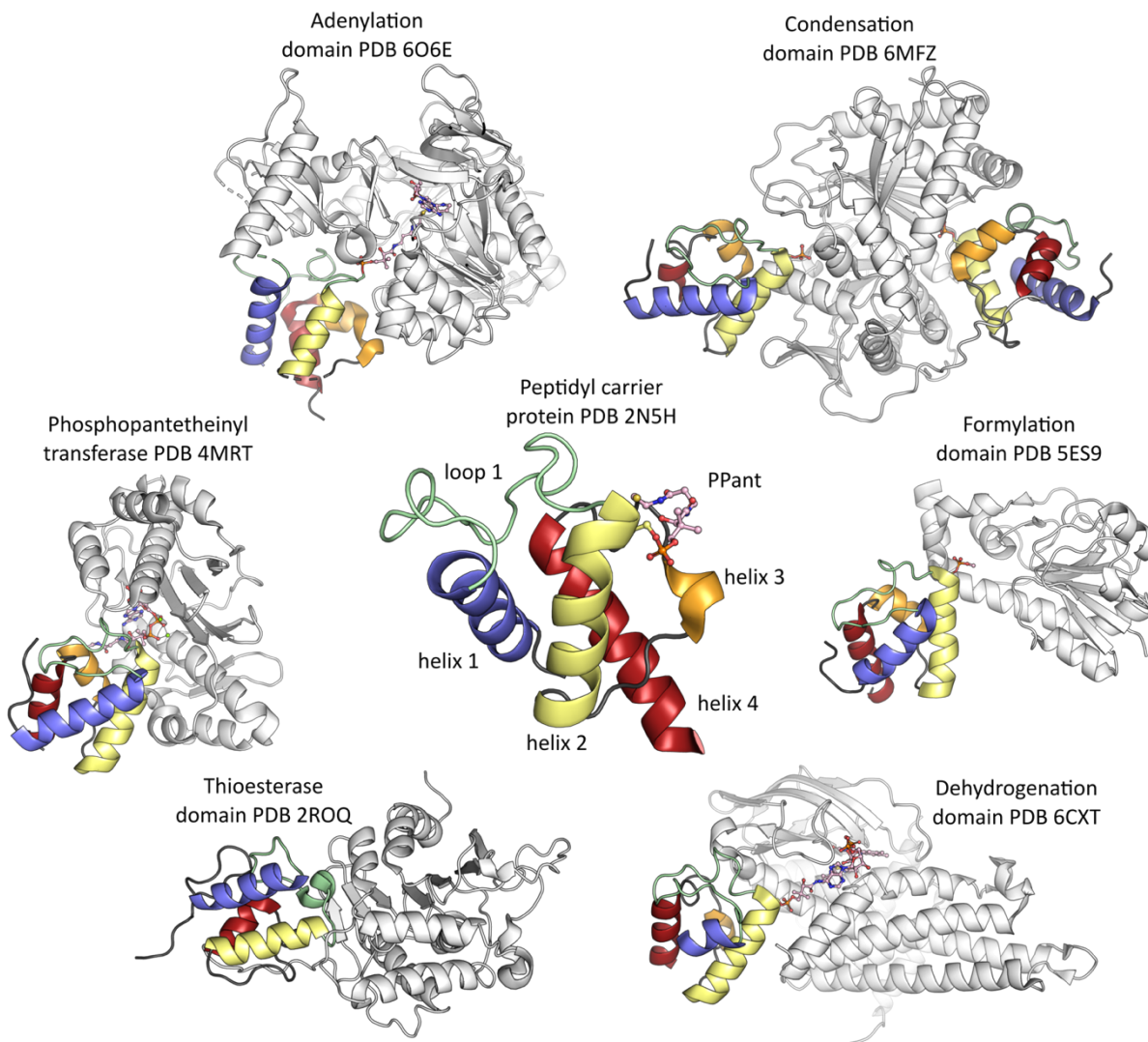
Figure 1.2: High-resolution structures of NRPS PCP-partner protein complexes. In the center is the PCP with its conserved secondary structure, where helix 1 (blue) is connected to helix 2 (yellow) by the 17-22 residue loop 1 (green), which is followed by a short helix 3 (orange) and helix 4 (red). At the beginning of helix 2 is the conserved serine, which is modified with a PPant arm (pink). Around the PCP are examples of the different PCP-partner protein complex structures that reveal the protein-protein interactions at the interface. These include the PCP-PPTase (PDB 4MRT), PCP-A domain (PDB 6O6E), PCP-C domain (PDB 6MFZ), PCP-F domain (PDB 5ES9), PCP-DH domain (PDB 6CXT), and PCP-TE domain (PDB 2ROQ).

tool to append carrier proteins with unnatural cargo and has been crucial for loading unnatural chemical probes onto the carrier protein to aid in stabilizing the carrier protein-partner protein complex for structural analysis. To understand the unique promiscuity observed in Sfp, the X-ray crystal structure of Sfp in complex with CoA and the PCP from tyrocidine NRPS, TycC3, was solved to a resolution of 2.0 Å.[18] As a means to promote complex formation, the conserved serine on the PCP was mutated to an alanine to prevent transfer of the PPant arm. Analysis of the PCP-Sfp protein-protein interface revealed a dependence on hydrophobic interactions and the presence of an intramolecular hydrogen bond similar to the recently solved structure of the ACP-PPTase complex from the *Mycobacterium abscessus* PKS PpsC.[19] Helix II in the PCP is mainly responsible for the hydrophobic interactions that occur at the interface where Leu46 and Met49 occupied a hydrophobic patch located in the C-terminal portion of Sfp. The single hydrogen bonding reaction is formed by the Gln40 located in the loop 1 region of the PCP. Mutagenic studies further support the importance of residues that comprise this hydrophobic patch and are necessary for sustained catalytic activity, where mutations that disrupt the hydrophobic interface residues result in abolished activity. Mutation of residues responsible for hydrogen bonding interactions retained enzymatic activity, which suggested the importance of the hydrophobic patch towards PCP recognition and the hydrogen bonding interaction responsible for Sfp promiscuity for non-cognate carrier proteins.[18]

Studies involving Sfp as a tool to tag and modify short peptide sequences further support the role of the hydrophobic patch for carrier protein substrate recognition in Sfp.[20] Experiments in which short segments from the PCP flanking the conserved serine were

incubated with Sfp led to identification of minimal peptide sequences that can be recognized and modified by Sfp. As supported by the Sfp-TycC3 PCP complex X-ray crystal structure, peptides that formed an α-helix analogous to the carrier protein helix 2 and its hydrophobic interaction with Sfp were able to be loaded with a fluorescently modified PPant.[20,21] These results have been further combined with computational approaches such as machine learning to develop the utility of Sfp and short peptide sequences as a tag to modify and functionalize proteins.[22]

1.3. Adenylation domain and peptidyl carrier protein interface analysis

The A domain is a critical player in NRP biosynthesis due to its role in the activation and attachment of a specific substrate onto the PCP prior to substrate incorporation into the natural product. The N-terminal core of the A domain, A_{core}, (residues ~1-400) houses the substrate binding pockets for ATP, a magnesium ion, and an amino acid.[23] While the ATP and magnesium ion binding are conserved across A domains, the substrate binding pocket for the amino acid varies, and has been demonstrated to distinguish the binding of the various acid substrates across A domain homologs. Upon binding of these three substrates, the A domain exists in the adenylation state and will catalyze the adenylation reaction through a conserved catalytic lysine located in the A10 motif of the A_{sub} domain (residues ~400-500). This forms an amino acid-adenylate intermediate in the active site upon loss of a pyrophosphate (Figure 1.1B). The A_{sub} domain undergoes a domain alternation to form the thiolation state, which is the rotation of the A_{sub} by ~140 degrees along a hinge region in the A8 loop to form a new catalytic active site for thiolation, as well as a protein-protein interface that can bind its cognate PCP. This A_{sub} domain rotation

has been uncovered in multiple A domain crystal structures bound to different substrates.[23–30] Upon holo-PCP binding to the A domain in the thiolation state, the PPant arm extends into the active site of the A domain, where the thiolation reaction is catalyzed to form a new thioester bond with the amino acid substrate with adenosine monophosphate (AMP) as a leaving group (Figure 1.3).[23] The AMP and substrate-loaded PCP dissociate and the A domain is ready to catalyze the next set of adenylation and thiolation reactions.

Because of the A domain's role as a gatekeeper in controlling substrate incorporation into the natural product, it has been the main target in NRPS engineering through A domain substitution or active site engineering.[12,31–35] Many of the early efforts in A domain substitution, however, were met with limited success, which was suggested to be due to the lack of proper protein-protein interactions.[36,37] Thus, large efforts have been made in determining the molecular basis of the PCP-A domain interaction; many of these efforts were spearheaded through the use of chemical biology tools in combination with structural biology to unveil the specific protein-protein interactions responsible for binding and therefore substrate loading. One of the main chemical probes utilized was the adenosine vinylsulfonamide (AVS) inhibitor (Figure 1.3C), which was initially designed as a substrate mimic to the aminoacyl-adenylate intermediate.[26,27,38–43] This inhibitor incorporates an electrophilic trap in the form of a Michael acceptor that would be attacked by the PPant thiol, thus covalently linking the PCP to the probe while it is non-covalently bound in the A domain in the thiolation state.

1.3.A. Initial structural analysis of type II PCP-A domain complexes

The first two X-ray crystal structures of the PCP-A domain complex were solved from the type II PCP and A domain from enterobactin biosynthesis, EntB-EntE, and the type II PCP-A di-domain from an unknown biosynthesis, PA1221.[26,27] In enterobactin biosynthesis, the initiation step involves the A domain, EntE, which activates a 2,3-dihydroxybenzoic acid and transfers it to the PCP, EntB.[27] In an unknown biosynthetic pathway, the A domain in PA1221 activates valine and loads it onto the PCP.[26] Both of these studies utilized the AVS inhibitor modified with the appropriate acyl substrate in AVS probes **1** and **2** (Figure 1.3C), respectively, which allowed the trapping, crystallization, and structure determination of the otherwise transient PCP-A domain complex in the thiolation state.[26,27]

Generally, both X-ray structures revealed that the PCP helix 2 and loop 1 regions formed specific protein-protein interactions with a composite interface formed by the A_{core} and A_{sub} domains, respectively. The EntB-EntE X-ray crystal structure uncovered a large dependence on hydrophobic interactions located on loop 1 and helix 2 in addition to three salt bridge interactions that exist on loop 1 and helix 2.[27] The PA1221 PCP-A domain X-ray crystal structure identified three hydrophobic interactions at the PCP helix 2 and multiple hydrogen bonding interactions found at helix 1, loop 1, and helix 2.[26] While the interface locations were consistent, the types of interactions at the interfaces varied across both structures. The high-resolution information of the EntB-EntE interface allowed the rational design of a homologous A domain from acinetobactin, BasE, to improve activity with the non-cognate PCP, EntB.[27] The successful BasE mutations involved swapping of potential BasE interface residues with that of EntE based on sequence

alignments, which replaced BasE hydrophobic interactions with electrostatic interactions observed in the EntB-EntE structure. These mutations improved BasE initial velocity rates with EntB by 15-fold.

1.3.B. Application of chemical probes to type I PCP-A domain systems

The next set of significant PCP-A domain studies involved applying chemical probes to study the larger type I NRPS systems from the first NRPS module of enterobactin biosynthesis, EntF, and the first NRPS module from linear gramicidin biosynthesis, LgrA. The EntF crystal structure was solved in the thiolation state using the AVS probe **3**, which revealed a PCP-A domain interface dependent on mainly hydrophobic interactions located at the EntF PCP loop 1, helix 2, and helix 3 regions.[28] Additionally, hydrogen bonds were found at EntF PCP helix 1, loop 1, and helix 2. In the case of the LgrA crystal structure in the thiolation state, which was solved using probe **5** (Figure 1.3C), a valyl-pantetheinamide probe, the PCP-A domain interface was identified to be similar to the EntF PCP-A domain interface in that the interface utilized the PCP loop 1, helix 2, and helix 3 regions.[30] While the LgrA PCP-A domain interface had less hydrophobic interactions and more hydrogen bonding interactions than EntF, LgrA contains a single salt bridge interaction at the start of helix 2 to aid in PCP binding. The trapped thiolation states of EntF and LgrA were compared to other module states to further dive into the modular architecture and movements during the NRPS biosynthetic cycles.

1.3.C. Recent interface analysis PCP-A domain complexes

Aside from conventional crystallographic studies on crosslinked carrier protein complexes, nuclear magnetic resonance (NMR) titration studies were performed to probe the residues involved in protein-protein recognition of the type II PCP-A domain interaction from pyoluteorin biosynthesis.[44] The A domain, PltF, activates and loads L-proline onto the PCP, PltL. An NMR titration was carried out utilizing N^{15}-labeled PltL loaded with an S-methyl PPant probe that allowed the PPant probe to access the A domain active site, but inhibited formation of the thioester and thus transfer of the aminoacyl moiety.[44] While this method identified PCP residues involved at the PCP-A domain interface, there was still uncertainty in how the A domain specifically interacts with the PCP. A follow up study of the PltL-PltF interaction utilized the AVS probe **4** (Figure 1.3C) to trap, crystallize, and solve the PltL-PltF complex structure through X-ray crystallography (Figure 1.3A).[45]

The crystal structure of the PltL-PltF complex revealed a similar mode of binding when compared to the previously solved PCP-A domain structures, however the main difference was the minimal role of PltL helix 2 in creating specific protein-protein interactions at the interface. Between PltL and PltF, the structure shows a single hydrophobic interaction between PltL Met43 and PltF Met257 at the beginning of helix 2 immediately following the conserved serine residue (Figure 1.3D). The remainder of the interface was located along the loop 1 region of PltL, which employed precise hydrophobic and hydrogen bonding interactions. PltF Ile454 is seen sitting inside a hydrophobic pocket formed by PltL loop 1 residues Ile19, and Trp37. Adjacent to this hydrophobic interaction is a hydrogen bond between PltF Lys457 with the backbone

carbonyl of PltL Gly38. PltL helix 3 was also observed forming hydrophobic interactions and a single hydrogen bonding interaction between PltL Ser62 and PltF Ser232. Alanine scanning of the A domain interface residues and comparison to the previous NMR titration experiments confirmed the importance of each specific interface interaction.

The covalent crosslinking of a PCP-A domain complex has been recently explored with success towards solving a PCP-A domain X-ray crystal structure. The ACP and partner protein interactions with the ketosynthase, acyltransferase, and TE from the FAS and PKS systems have been crosslinked and structurally analyzed using a variety of PPant probes such as the chloroacrylamide and bromoacetamide probes.[9,46] These probes take advantage of the nucleophilic active site cysteine or serine that attacks the PPant probe with a halide as a leaving group, which covalently crosslinks the PCP and partner protein. Since A domains do not have a nucleophilic residue as part of its catalytic mechanism, a cysteine mutation must be introduced in the A domain active site to enable crosslinking with these probes.[47]

The probes and respective mutations have been applied to the type II A domain and PCP of hitachimycin biosynthesis.[47] The A domain, HitB, activates and loads a (*S*)-β-phenylalanine onto the PCP, HitD. A conserved aspartate in the active site of HitB, which is involved in substrate binding of the amino group of the amino acid, was mutated to a cysteine. This mutation enabled crosslinking of the HitD-HitB complex with probe **6** (Figure 1.3C), which afforded crystallization and determination of the HitD-HitB X-ray crystal structure in the thiolation state (Figure 1.3B).[47] The complex crystal structure revealed an interface formed mainly by the HitD loop 1 and helix 2 regions, which is consistent with the previously solved PCP-A domain structures discussed above (Figure
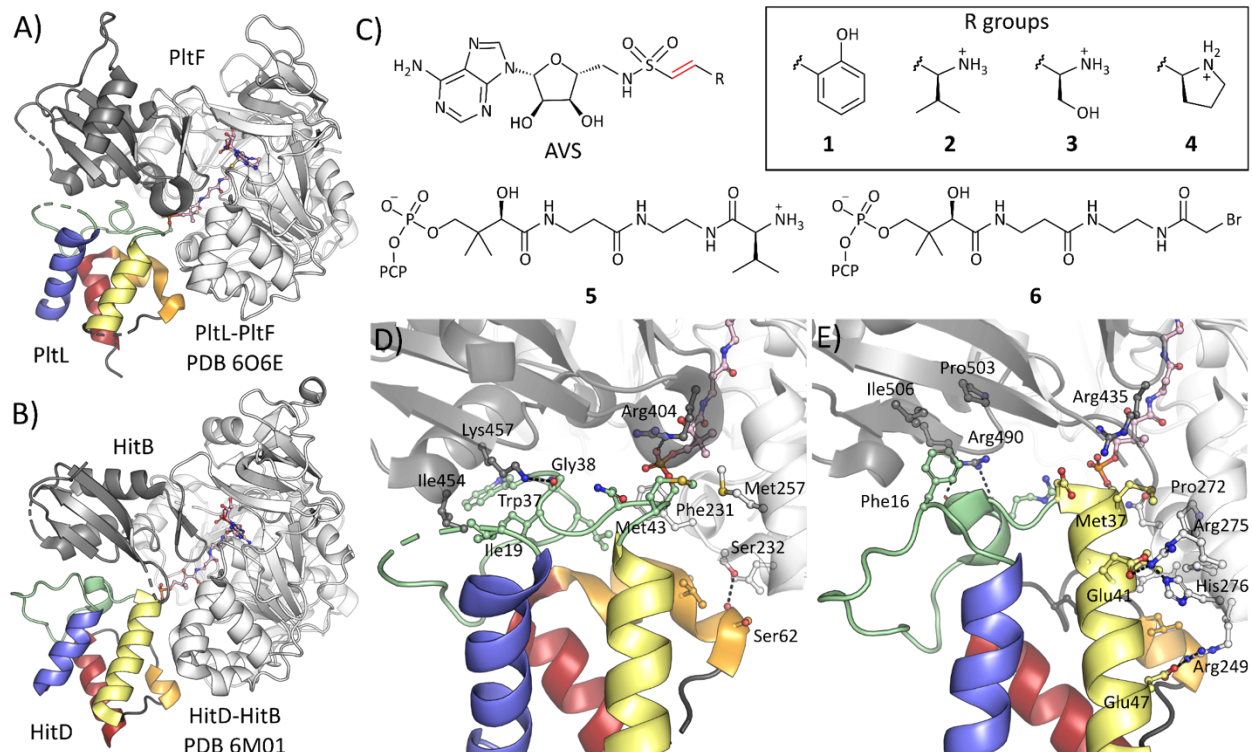
Figure 1.3: X-ray crystal structures of PCP-A domain complexes. The overall crystal structures of A) PltL-PltF complex (PDB 6O6E) and B) HitD-HitB (PDB 6M01), where the A domain is colored according to the A_{core} (white) and the A_{sub} (gray) at the C-terminal end. The structures of chemical probes C) utilized in PCP-A domain structural analysis: 1 is a salicylate-AVS inhibitor; 2 is a valyl-AVS inhibitor; 3 is a seryl-AVS inhibitor; 4 is a prolyl-AVS inhibitor (electrophilic trap in red); 5 is a valine pantetheineamide substrate mimic; and 6 is a bromoacetamide pantetheine crosslinker. The interface residues from the D) PltL-PltF complex and E) HitD-HitB complex are shown in ball and stick. Hydrogen bonding and electrostatic interactions are shown with black dashed lines. The PCPs from both structures are colored as previously described, where helix 1 is blue, loop 1 is green, helix 2 is yellow, helix 3 is orange, and helix 4 is red, and the PPant/ligands are are pink.

1.3E). Although the regions are consistent, the specific interactions differ. The HitD loop 1 uses a Phe16 to fit into a HitB hydrophobic pocket formed by A_{sub} residues Arg590, Pro503, and Ile506. Adjacent to the hydrophobic interaction are two hydrogen bond interactions formed between HitB Arg590 and the main chain carbonyls of HitD Arg30 and Asp31. On HitD helix 2, there are multiple salt bridge interactions with the HitD A_{core}. These salt bridge formations occur between HitD Glu41 with HitB Arg275/His276 and HitD Glu47 with HitB Arg249. Additionally, HitB Trp247 sits inside a HitD hydrophobic pocket formed between helix 2 and helix 3 consisting of Thr39, Leu43, Leu59, and Phe64.

While the general PCP and A domain regions in the HitD-HitB crystal structure are consistent with the previously solved PCP-A domain complex structures, the HitD-HitB interface utilizes a combination of hydrogen bonding, hydrophobic, and electrostatic interactions.[47] EntF, PA1221, and PltF-PltL interfaces depend on hydrogen bonding and hydrophobic interactions,[26,28,45] whereas EntE-EntB is dependent on electrostatic interactions and hydrophobic interactions.[27] HitD-HitB is most similar to LgrA in that it involves 1 electrostatic interaction, 2 hydrophobic interactions, and multiple hydrogen bonding interactions.[30]

1.3.D. Outlook on PCP-A domain interfaces

Although this extensive collection of PCP-A domain structures has surfaced recently, there remains questions regarding the mechanism of PCP-A domain binding. The multitude of PCP-A structures reveal the conformation formed during thiolation; however, it is important to note that the initial recognition and binding events are just as important. NMR structures of PCPs have shown the positioning of the substrate-loaded

PPant in a retracted state,[48] which may be the conformation that the A domain must recognize. Conversely, during thiolation, the PCP-A domain structures have the PPant in the extended state. A full understanding will require that a dynamic picture of the PCP-A domain binding mechanism be teased out, which is an ongoing investigation.[49,50] Furthermore, analysis of the PCP-A domain linker and its role in catalytic activity, in addition to its contributions to forming the PCP-A domain interface, will also aid efforts in A domain substitutions.[31,51]

1.4. Condensation domain and evolutionarily related enzyme interface analysis with the peptidyl carrier protein

The C domain catalyzes the peptide bond formation between adjacent PCP-linked substrates, and is responsible for the downstream transfer of the elongating peptide intermediate throughout the synthetase. Multiple C domain crystal structures revealed that the C domain is split into two halves, which are referred to as the N-terminal lobe and C-terminal lobe.[28–30,52–61] The two lobes are held together as a pseudo-dimer through conserved latch and floor loop motifs. To access the active site, the previously solved crystal structures have revealed two 15 Å tunnels from the donor and acceptor PCP binding sites by which the PPant can enter and present its substrate to the catalytic residues.[15] The conserved catalytic residues postulated to be responsible for the catalysis of condensation, HHxxxDG, are found on the N-terminal lobe at the interface of both lobes. While the exact mechanism of catalysis is still being discerned, it is postulated that the second conserved histidine deprotonates the amine of the acceptor substrate, which allows amine nucleophilic attack at the thioester carbon of the donor substrate.[62]

C domains have been debated as a secondary checkpoint, where the C domain must bind the correct substrates in the active site before catalyzing peptide bond formation in addition to creating specific protein-protein interactions with the appropriate donor and acceptor PCPs. While assessment of the C domain active site substrate selectivity is still underway,[63–67] information involving the protein-protein recognition of the C-domain with both acceptor and donor PCPs is critical when engineering C domains with non-cognate NRPS systems.

1.4.A. The first crystal structure of the PCP-C domain complex

The earliest studies that structurally analyzed the protein-protein interactions of C domains involved the type I NRPS SrfA-C and the type I NRPS AB3403.[28,54] Both SrfA-C and AB3403 are termination modules consisting of the domains C-A-PCP-TE from surfactin biosynthesis and an unknown biosynthetic pathway, respectively. The X-ray crystal structures of both NRPS modules have been solved with the PCP bound at the acceptor site of the C domain.[28,54] The SrfA-C acceptor PCP-C domain structure initially revealed a protein-protein interface formed by the C domain N-terminal lobe and C-terminal lobe with the PCP helix 2 and helix 3, respectively.[54] The specific PCP-C domain interactions consisted of nearly all hydrophobic interactions, with only one potential hydrogen bonding interaction. In this study, the conserved serine on the PCP was mutated to an alanine to prevent addition of a PPant arm, so specific interactions with the cofactor remain unresolved.

The AB3403 acceptor PCP-C domain structure revealed a similar protein-protein interface to SrfA-C, where the C domain N-terminal lobe and C-terminal lobe are

observed interacting with the PCP helix 2 and 3, respectively.[28] Similarly, the AB3403 PCP-C domain interface is very dependent on hydrophobic residues. Unlike the SrfA-C structure, the AB3403 PPant arm was observed extended into the active site with the C domain Arg344 forming an electrostatic interaction with the PPant phosphate. Despite the similar acceptor PCP-C domain interfaces, superposition of the C domains revealed that the bound PCPs differ by a 30° rotation. While this may be due to the lack of a PPant arm in the SrfA-C structure, the protein-protein interface created is feasible because the location of the conserved serine is still at the entrance to the C domain tunnel.

1.4.B. Recent success in the interface analysis of type I PCP-C domain complexes

Recent work in obtaining the structural snapshots of a di-modular NRPS has revealed multiple PCP-C domain bound structures, including the donor PCP-C domain complex as well as the first structure of a C domain with both the acceptor and donor PCPs bound simultaneously (Figure 1.4A).[29] This was performed on the type I NRPS from linear gramicidin synthesis, LgrA, which consists of F_1-A_1-PCP_1-C_2-A_2-PCP_2-E_2, where the subscript represents the module.[29] Of the multiple structures and conformations solved on this system, two structures utilized probe **7** (Figure 1.4D) to help crystallize and gain high-resolution insights into the protein-protein interactions of the donor PCP-C domains (PCP_1-C_2) complex. The crystal structures of the donor PCP-C domain complex revealed a protein-protein interface mainly dependent on hydrophobic interactions (Figure 1.4B). These interactions are located at the PCP loop 1, helix 2, and helix 3 regions and the C domain C-terminal lobe. PCP loop 1 contributes an electrostatic interaction with His721 and C domain Asp1011. Adjacent is a network of hydrophobic

interactions involving PCP loop 1 Leu723 and helix 3 Phe752 and Tyr748 with C domain Thr1013, Met1016, Leu1085, and Leu1088.

Interestingly, unlike the previously solved acceptor PCP-C domain structures that show the composite C domain interface formed by both N-terminal and C-terminal lobes, the donor PCP in LgrA is observed only interacting with the C-terminal lobe, where the donor PCP helix 2 contacts the floor loop region of the C-terminal lobe.[29] The crystal structure shows a lack of specific interactions commonly encountered between the PCP helix 2 and C domain floor loop. Instead, the PCP-C domain interface reveals a dependence on shape complementarity between the two helices. Using these high-resolution structures, the 6 Å resolution crystal structure of the LgrA C domain bound to both acceptor and donor PCPs (PCP_1-C_2-PCP_2) was resolved (Figure 1.4A).[29] The crystal structure revealed the first instance of both acceptor and donor PCPs occupying their respective sites on the C domain. The donor PCP maintains an identical binding interface to the other donor PCP-C domain complexes, while the acceptor PCP is supported through comparison to the AB3403 PCP-C domain interaction. Although a low-resolution structure, the acceptor PCP-C domain interface seems to be formed by the PCP helix 2 and loop 1 regions and the C domain N-terminal lobe and C-terminal lobe, respectively. Direct coupling analysis and mutagenesis of the protein-protein interactions between the LgrA acceptor PCP-C domain revealed significant decreases in C domain activity, thus supporting the interface interactions inferred from the model.

Recently, the crystal structure of the PCP_2-C_3 didomain from the fuscachelin type I NRPS, FscG, was solved utilizing probe **8** (Figure 1.4D) to aid in crystallization and visualization of active site residues.[66] The probe mimicked a glycyl-PPant, where the

thioester linkage was replaced with a more stable thioether. The glycyl-PPant moiety bound in the active site provided insight into the lack of a substrate binding pocket to control C domain specificity, which is consistent with recent C domain substrate analyses.[63–65,67] Surprisingly, the crystal structure revealed that the PCP_2 was bound at the acceptor site of the C domain instead of the expected donor site.[66] Comparison of the PCP_2 and PCP_3 revealed a sequence identity of 65% and a structural alignment with a root mean squared deviation of 2 Å, which supports the continued analysis of the PCP_2 bound at the opposite side of the C domain. This donor PCP-C domain binding interaction revealed an interface that was mainly hydrophobic located at the PCP helix 2 and helix 3 regions (Figure 1.4C). Specific hydrophobic interactions included the PCP helix 2 Leu2518 and Leu2515 with C domain N-terminal lobe Leu2580 and Trp2579, and also with the PCP helix 3 Phe2538 with Val2908. The C domain also utilizes Arg2906 to create a salt bridge interaction with the PPant phosphate moiety. Interestingly, the buried surface area at the interface is ~550 $Å^2$, which is small compared to the previous PCP-A domain and PCP-C domain interface areas. The donor PCP_2-C_3 domain complex structure was also solved with only a PPant arm attached to the PCP.[66] While the protein-protein interface remained generally the same, the C domain Arg2577 was observed blocking access to the C domain tunnel at the interface, and the PPant was unable to extend into the tunnel. On the other hand, the glycyl-PPant loaded PCP was observed inside the tunnel and active site of the C domain. The PCP_2-C_3 crystal structure with the Arg2577Gly mutation revealed an unloaded PPant extended into the C domain tunnel. This structure along with enzyme assays of the mutant supported the hypothesis that Arg2577 acts as a gating residue that may only be moved if the appropriate substrate is loaded.
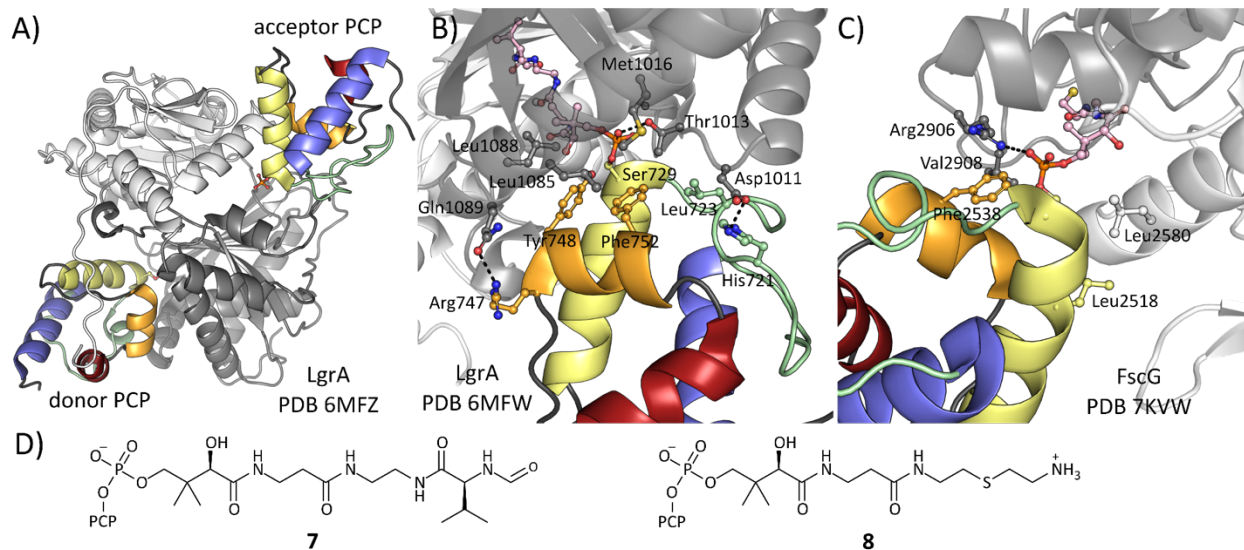
Figure 1.4: X-ray crystal structures of PCP-C domain complexes. A) Crystal structure of the LgRA C domain bound with PCPs at the donor and acceptor sites (PDB 6MFZ). B) Interface view of the donor PCP-C domain X-ray crystal structure from LgrA (PDB 6MFW). C) Interface view of the X-ray crystal structure of the donor PCP bound at the C domain acceptor position of FscG (PDB 7KVW). D) Chemical probes utilized to obtain PCP-C domain crystal structures: 7 is a formyl-valine pantetheineamide substrate mimic and 8 is a glycyl-ether panthetheine substrate mimic. The PCPs from both structures are colored as previously described, where helix 1 is blue, loop 1 is green, helix 2 is yellow, helix 3 is orange, and helix 4 is red, and the PPant is pink. The C domain is colored according to the N-terminal lobe (white) and C-terminal lobe (gray).

1.4.C. Evolutionarily divergent epimerization and termination domains

Tailoring and termination domains, such as the E domain and the C_T domains, have evolved from C domains.[68] The E domain is chiefly responsible for conversion of thiotemplated L-amino acids to D-amino acids, which contributes to the structural diversity of NRPs.[69] The C_T domain instead terminates NRP production through cyclization and release of a cyclic peptide product. Structurally, both the E and C_T domains conserve the canonical V-shaped fold seen in C domain structures.[56,70] The differences arise in subtle active site changes that confer different activities.

Although these enzymes catalyze different reactions, the mode of binding the donor PCP remains similar to previously solved donor PCP-C domain interfaces. In the gramicidin type I NRPS, module 1 consists of A-PCP-E, where the E domain epimerizes the L-phenyl-PCP to D-phenyl-PCP.[70] The crystal structure of the PCP-E di-domain was solved and revealed the PCP bound to the donor binding site of the E domain using hydrophobic, hydrogen bonding, and electrostatic interactions.[70] The location of the interactions mainly involved the E domain N-terminal lobe with the PCP loop 1, helix 2, and helix 3. The E domain C-terminal lobe also contacted the PCP helix 2 as part of the protein-protein interface. Additionally, a crystallographically ordered 20-residue linker region between the E domain and PCP was also identified as crucial for the formation of a protein-protein interface, as mutation of specific electrostatic residues decreased product formation.[70]

In fumiquinazoline F biosynthesis, the C_T domain, TqaA, is responsible for the cyclization of a ten-membered ring from a tripeptide.[56] The X-ray crystal structure of TqaA as a holo-PCP-C_T domain complex was solved and revealed an overall structure and

complex with the PCP bound at the donor site of the C_T domain similar to the PCP-E domain structure from GrsA.[56] This interface was formed by almost exclusively the C_T domain C-terminal lobe with the PCP helix 2 and helix 3 through mainly hydrophobic interactions in addition to a PPant phosphate hydrogen bonding interaction.

1.4.D. Outlook on the PCP-C domain interface

Obtaining structural information on the C domain interactions with its donor and acceptor PCPs to inform NRPS engineering has remained challenging due to multiple factors. PCP interactions with partner proteins are transient in nature and thus are difficult to crystallize in order to study the specific interfaces that enable peptide bond formation. Furthermore, two substrate-loaded PCPs are required to be bound at the donor and acceptor sites in order to evaluate the active site interactions that affect substrate selectivity. Additionally, C domain dynamics at the interface may play a role in PCP binding and substrate access.[61] Promising chemical biology tools are currently being developed to help stabilize the transient PCP-C domain complexes for structural analysis of the protein-protein interface and active site substrate binding.[46,57,71,72] Despite these challenges, X-ray crystal structures of C domains have guided successful re-engineering of type I NRPSs through identification of new areas susceptible to combinatorial biosynthesis.[73,74] Emerging techniques in structural biology such as cryo-electron microscopy (EM) could potentially aid in capturing the C domain with a combination of donor and acceptor PCPs, which can further shed light on the effect of PCP-bound substrates in forming a protein-protein interface. Starter condensation domains, which

condense a donor acyl chain with an acceptor amino acid in lipopeptide NRPSs, have also seen recent success in active site analysis and engineering.[75,76]

1.5. Peptidyl carrier protein and tailoring domain interface analysis

The tailoring domains encompass the groups of proteins that are not considered to be core NRPS domains (PCP, A domain, and C domain), yet have the capacity to chemically modify the growing peptide bound to the carrier protein.[6] The chemical modifications catalyzed by tailoring domains add diversity and functionality to the structure of NRPs, which may add protection against degradation by proteases, enhance binding affinity to specific targets, and increase NRP half-life upon release from the PCP.[77] Given the vast number of tailoring domains characterized to date, we limit our discussion to those that have been shown to form interfaces with carrier proteins in order to further illustrate the importance of protein-protein interactions to gain access to substrate functionalization. Some of the chemical modifications on the growing peptide include but are not limited to N-formylation, β-hydroxylation, dehydrogenation, and the aforementioned epimerization.[10] Being able to understand the interactions that are involved in coordinating these reactions, combined with an understanding of the governing protein-protein interactions provide a higher degree of spatiotemporal control in engineered NRPSs. In this section, we explore the protein-protein interfaces observed in the structures of different tailoring domains bound to the PCP, with an emphasis on the types of interactions that promote transient complex formation and guide reactivity in the biosynthesis of NRPs.

1.5.A. Formylation domain

After biosynthesis, the dimerization of gramicidin is stabilized by the N-formyl valine moiety, which enables the formation of pores that disrupt ion gradients in the membranes of gram-positive bacteria, thus highlighting the importance of formyl modifications and F domains.[89] Recently published crystal structures of LgrA from the linear gramicidin NRPS show different conformational states that illustrate the different stages of adenylation, thiolation and formylation involved in a type I NRPS module.[30] In the formylation state, the A_{sub} domain in LgrA positions the valyl-PCP at the active site of the F domain, where a single salt bridge between Arg758 in helix II of PCP and a Asp652 in a nearby loop in A_{sub} was observed. The F domain Met178 and Leu127 and the PCP Tyr748 form a hydrophobic patch that provides further stability to the complex, although interestingly the interface area is approximately 500 $Å^2$, which is small relative to other PCP-partner protein structures.[30]

1.5.B. Oxidation domain

Another important group of tailoring domains are the P450 oxygenases, or "Nature's blowtorch" as they are sometimes referred to, are oxygen dependent metalloproteins widely known for their capacity to install hydroxyl groups to certain substrates.[16] One example of a heavily hydroxylated NRP is skyllamycin, which is a cyclic depsipeptide with multiple β-hydroxylated amino acids, as well as hydroxylated aromatic rings.[78] In the skyllamycin biosynthetic pathway, P450s are selective towards the cognate PCPs from different modules within the synthetase.[78] P450s must recognize the competent binding interfaces that emerge from the PCPs loaded with different peptides.

The structure of the cytochrome P450 tailoring domain, P450$_{sky}$, bound to the PCP, PCP$_7$, reveals the protein-protein interface of a monooxygenase domain that binds to PCP;[78] electrostatic and hydrogen bonding interactions at the interface were observed in the PCP helices 2 and 3, where residues Arg63, Thr46, and Lys47 form interactions with the P450 Asp191, Asn197 and E235, respectively. Trp193 and Leu194 of the P450 form hydrophobic patches with residues of helix 2 and 3 in the PCP that also assist in accommodating the geminal dimethyl group of the PPant attached to the conserved Ser42. It is important to note that although the P450 is a standalone domain that binds to the PCP, it does so selectively and does not necessarily interact with all amino acids in the module, given that not all residues in the final product show hydroxylation at the β-carbon. Comparing the crystal structure to a computational model of other PCPs reveals slight conformational differences between the relative orientations of the helices in PCP, which could potentially account for the P450 selectivity for certain PCPs.[78]

1.5.C. Dehydrogenation domain

Another important tailoring domain is the DH domain involved in the biosynthesis of pyrrole containing NRPs. DH domains, which may also be re-classified as an oxidase,[79] are flavin-dependent proteins that use oxygen as the final electron acceptor in the process of dehydrogenating proline for the production of thiotemplated pyrroles.[10,16] In pentabromopseudilin biosynthesis, a type II NRPS DH domain oxidizes a PCP-bound proline to a pyrrole group.[79] To understand the mechanism of dehydrogenation and how the DH domain binds the PCP, the tetrameric X-ray crystal structure of a flavin dependent
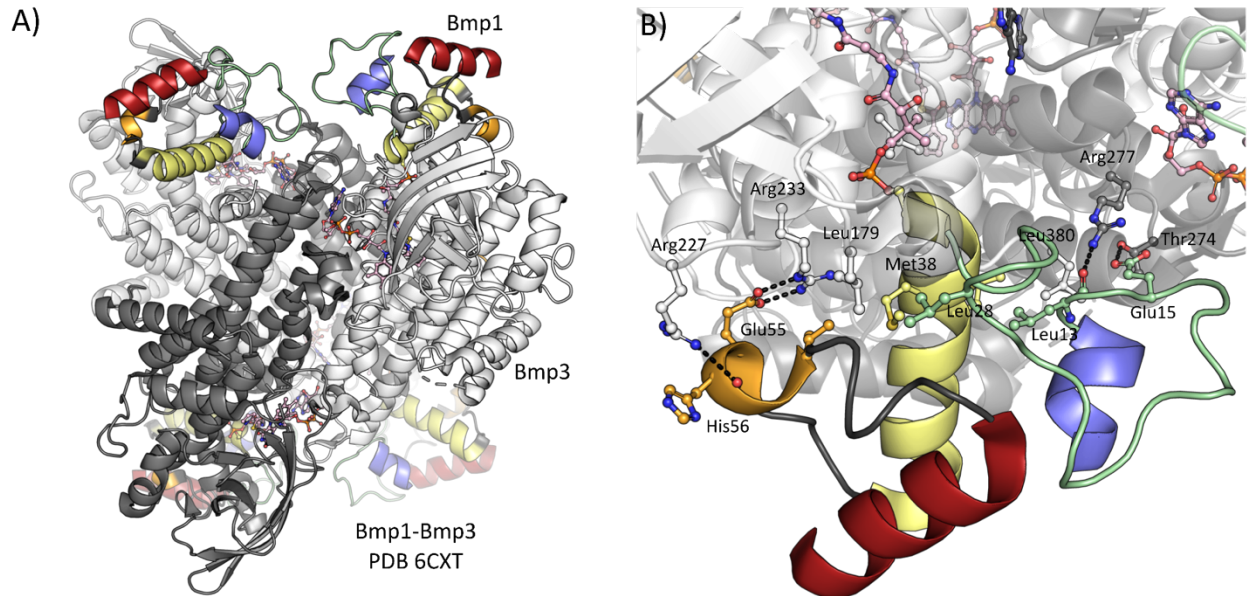
Figure 1.5: X-ray crystal structure of the PCP-DH domain complex. A) Overview of the tetrameric Bmp1-Bmp3 crystal structure (PDB 6CXT). The monomers of the DH domain, Bmp3, is alternating in white or gray. B) Close up of the Bmp1-Bmp3 interface. The PCP, Bmp1 is colored as previously described, where helix 1 is blue, loop 1 is green, helix 2 is yellow, helix 3 is orange, and helix 4 is red, and the PPant/ligands are pink.

DH domain, Bmp3, bound to FAD in complex with the PCP, Bmp1, was solved with either *holo*-Bmp1 or pyrrolyl-Bmp1 (Figure 1.5A).[79]

When comparing the DH domain active site between both structures, there were no differences in terms of the active site and cofactor spatial organization.[79] Hydrophobic residues in Bmp3 aid in aligning the PPant moiety with its active site, placing the proline in close proximity to FAD. The Bmp1 helix I interactions have proven to be important for DH domain activity, as demonstrated by mutations to Arg277 in Bmp3 that disrupt hydrogen bonds with the Leu13 main chain carbonyl and significantly decrease production of the pyrrole (Figure 1.5B).[79] Glu15 in the Bmp1 helix I also forms hydrogen bonds with the Thr274 main chain nitrogen, demonstrating how disruptions to one of the recognition helices in PCP can impact activity. The most disruptive mutations are those shown to interfere with helix II and III of PCP, where a hydrophobic patch consisting of Leu28, Met38, Ile58, Pro60, and Phe63 binds the side chains Tyr178 and Leu179 in Bmp3. Double mutations of these residues in Bmp3 result in complete elimination of product formation. The overall contribution of these interactions led to the conclusion that hydrophobic interactions govern the formation of interfaces with electrostatic and salt bridge interactions playing a minor role.

1.6. Peptidyl carrier protein and termination domain interface analysis

Termination domains are commonly found at the end of linear NRPS modules and are responsible for the release of the mature peptide from the PCP.[16] The mechanism of action varies between termination domains from different NRPSs by taking advantage of a diverse array of nucleophiles that may catalyze the intra- or intermolecular release from

the PCP, yielding linear or cyclized products with different functional groups. For instance, reductase (R) domains catalyze a reduction at the thioester linkage that can lead to the release of alcohols or aldehyde groups at the C-terminus of the peptide product. TEs cleave the thioester bond in PPant and can use different substrates as nucleophiles.[80] Additionally, the C_T domain can also catalyze the release and macrocyclization of NRPs as discussed previously.[56]

1.6.A. Thioesterase domains

The structures of PCPs in complex with TE domains give insight into the protein-protein interactions that lead to the timed release of substrate from PCP. In general, NRPS TEs belong to the α/β hydrolase family of enzymes, with an average size of 240-290 residues. Apart from possessing a Ser-His-Asp catalytic triad, they also have a 40-residue lid region that lines the substrate and alternates between open and closed states.[80] TEs are be further classified into type I, which hydrolyze a mature peptide from the PCP using diverse catalytic strategies, and type II TEs that recognize and hydrolyze PCPs with incorrectly loaded cargo that can stall the biosynthetic machinery.[88] While not directly involved with the core NRPS machinery, Type II (repair) TEs can act *in trans* on PCPs with similar structure and catalytic domains as their type I counterparts. [81]

The TE domain in the termination module EntF is an example of a type I TE responsible for the peptide releasing step from the PCP to produce the enterobactin in *Escherichia coli.* The cyclization reaction that ultimately forms a tri-lactone is a product of coordinated reactions between the PCP and TE that are situated as the final two domains of the synthetase. To identify the protein-protein interactions and TE conformational

29

changes upon binding, the EntF-TE complex was structurally analyzed in complex with the PCP.[82] The X-ray crystal structure of the EntF PCP-TE complex shows extensive interactions between the TE lid region and active site residues with the PCP helices II and III, encompassing over 1000 Å^2 of total buried surface area excluding the PPant arm. In terms of the PPant arm, the majority of contacts involve a loop region in the TE, while mutagenesis studies revealed the importance of specific residues at the interface essential for enterobactin production and release. For instance, mutation of the TE Trp1079 resulted in a disruption of the hydrophobic interactions with the PCP at the PPant cavity and inhibition of product formation.

Compared to type I TEs, the type II TE, SrfTEII from surfactin biosynthesis, also shares an α/β hydrolase fold.[81] However, the promiscuity observed in SrfTEII is due to the partial covering of its catalytic triad as well as other structural modifications that allow increased accessibility. Further comparisons with another type II TE structure from the colibactin synthase, ClbQ, further highlights the preference for smaller substrates, as evidenced by a smaller active site cavity when compared to type I.[82,83] Substrate specificity studies demonstrate that type II TEs favor hydrolysis of acetate, indicative of a proofreading role of PCPs that have been post-translationally modified with acetyl-CoA or malonyl-CoA by PPTases.[17]

1.6.B. Reductase domain

The biosynthesis of aureusimine from *Methanobrevibacter ruminatium* involves the NRPS Mru_0351, which utilizes an archaeal R domain to release the peptide product.[84] The recent structure of the Mru_0351 PCP-R domain shows the first archaeal R domain
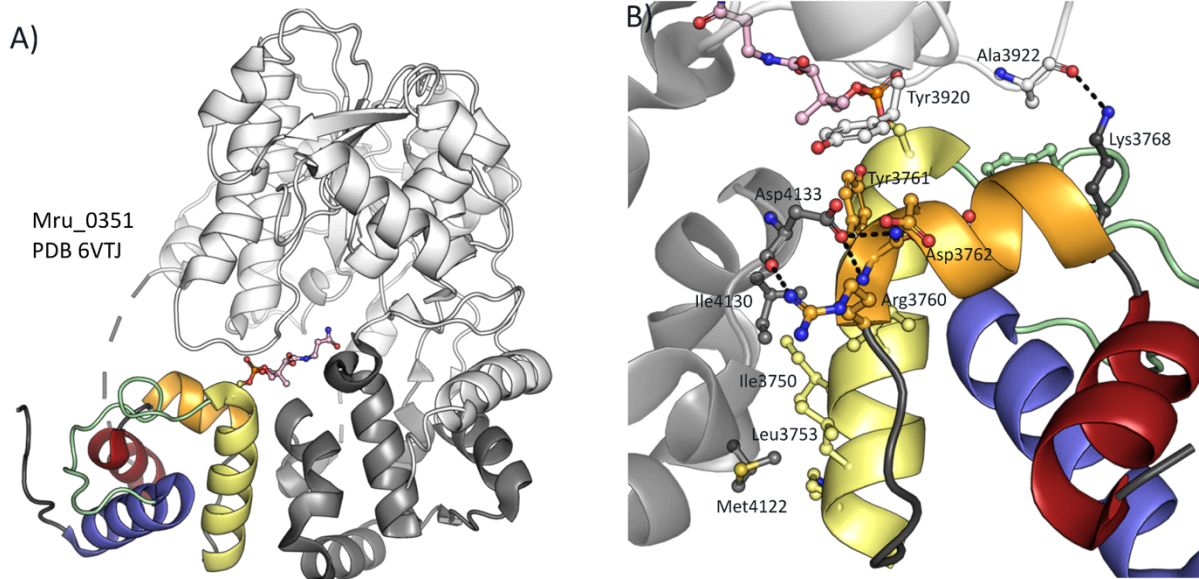
Figure 1.6: X-ray crystal structure of the PCP-R domain complex. A) Overview of the Mru_0351 PCP-R domain complex (PDB 6VTJ). B) Close up of the PCP-R domain interface. The R domain is colored according to the N-terminal (white) and the C-terminal (gray) regions. The PCP is colored as previously described, where helix 1 is blue, loop 1 is green, helix 2 is yellow, helix 3 is orange, and helix 4 is red, and the PPant is pink.

bound to the PCP (Figure 1.6A).[84] The R domain was compared to a carboxylic acid reductase module, CAR-PCP-R, which is the only other R domain structure complex reported to date.[85] The principal interactions between the R domain and the PCP include a novel helix-turn-helix (HTH) motif and a gating loop in the R domain that interact with the PCP helix residues.[84] A series of hydrogen bonds and hydrophobic interactions decorate the PCP and R domain interface. The hydrophobic interactions mainly consist of the PCP Phe3736, Leu3743, Ile3749, Ile3750, Leu3753, Tyr3761 and Phe3765 and the R domain Tyr4118, Met4122, Ile4126, Ile4130 and Tyr3920 from the HTH-motif and gating loop (Figure 1.6B). Hydrogen bonding networks occur between the novel HTH-motif and PCP helices II and III that extend to water molecules found at the interface.

Although it is not fully modeled in the crystal structure, the PPant participates in various interactions with the R domain as it extends into its active site. The geminal dimethyl group of PPant is stabilized by a hydrophobic pocket in the R domain composed of Tyr3920, Leu3743 and Tyr3761. Hydrogen bonding interactions also exist between PPant and main chain atoms in His3919, Thr4032, and Ala4304. The structure of the PCP-R domain also evaluates the role of the gating loop residues in stabilizing the PCP and positioning the PPant group close to the NAD(P)H binding site.

1.7. Outlook

The increase in high-resolution information of the transient PCP-partner protein complexes over the past decade is extremely insightful towards establishing guidelines to inform future efforts in engineering NRPS pathways. Compared to the ACP-partner protein interactions in the type II FAS, the interface regions on the carrier protein are

similar, however, studies of ACP-partner protein interactions are revealing protein-protein interfaces that are much more dependent on small, electrostatic interfaces.[9,86] The complex structures reviewed here provide static details on the mode of binding; however, to further understand the PCP-partner protein binding event, more dynamic information will be required through techniques such as NMR titrations, solution NMR structures, molecular dynamic simulations, and cryo-EM structures.

Nevertheless, the protein-protein interactions found at the PCP-partner protein interface can already be leveraged to understand and design new interactions with non-cognate partner proteins. During interface design, the wild-type PCP-partner protein structures can also be integrated with computational techniques, such as protein-protein docking and MD simulations, to create a model of a new non-cognate PCP-partner protein complex.[87] This model can then be used in rational design, semi-rational design, or in directed evolution to improve the binding interactions between the non-cognate proteins. It may be worthwhile to focus design on the interface of the partner protein, as mutations to the PCP will likely affect interactions with other partner proteins necessary in a pathway. Interface design can be performed in conjunction with current NRPS design methodologies, such as A domain substitution, A-PCP-C domain substitutions, or insertion of tailoring domains in type I and type II NRPS systems (Figure 1.7). While this review only covers PCP-partner protein interfaces, interface design can also be applied to the variety of interdomain interactions created throughout the NRPS biosynthetic cycle, such as between the A-C domain interfaces. Overall, designing protein-protein interfaces as part of combinatorial biosynthesis strategies is a promising way to enhance the
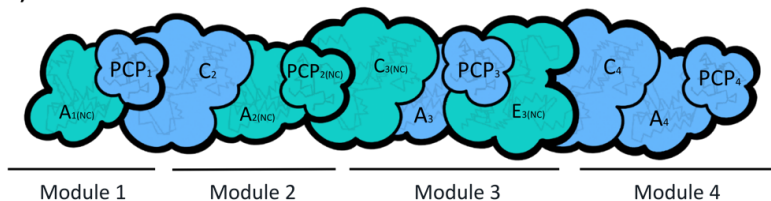
Figure 1.7: Design methodologies of the NRPS integrated with protein-protein interface design between non-cognate PCP and partner proteins. A) A general type I NRPS (blue) is shown with non-cognate (green) substitutions or insertions. The following panels highlights the new non-cognate protein-protein interfaces that have been introduced in B) A domain substitution, C) A-PCP-C domain substitution, and D) tailoring domain insertion. Movement of the PCP in C) is shown with arrows, where the initial PCP position is shown as more transparent. The new non-cognate interface can be optimized via mutations, which are depicted as complementary shapes similar to a puzzle piece.

success in NRPS engineering. A productive interface, and thus product formation, or improved pathway productivity, may only lie a few mutations away!

1.8. Acknowledgements

Chapter 1, entitled Protein-protein interface analysis of the non-ribosomal peptide synthetase peptidyl carrier protein and enzymatic domains, in full, is a reprint of the material as it appear in: Corpuz, J. C.; Sanlley, J.; Burkart, M. D. Protein-protein interface analysis of the non- ribosomal peptide synthetase peptidyl carrier protein and enzymatic domains. *Systems and Synthetic Biology,* 2022, 7 (2), 677-688. The dissertation author is the primary co-author of this manuscript.

1.9. References

(1) Schwarzer, D.; Finking, R.; Marahiel, M. A. Nonribosomal Peptides: From Genes to Products. *Nat. Prod. Rep.* **2003**, *20* (3), 275.

(2) Shevach, E. M. The Effects of Cyclosporin A on the Immune System. *Annual Review of Immunology*. **1985**, *3* (1), 397-423.

(3) Hecht, S. M. Bleomycin: New Perspectives on the Mechanism of Action. *J. Nat. Prod.* **2000**, *63* (1), 158–168.

(4) Zavascki, A. P.; Goldani, L. Z.; Li, J.; Nation, R. L. Polymyxin B for the Treatment of Multidrug-Resistant Pathogens: A Critical Review. *Journal of Antimicrobial Chemotherapy* **2007**, *60* (6), 1206–1215.

(5) Moellering, R. C. Vancomycin: A 50-Year Reassessment. *Clinical Infectious Diseases* **2006**, *42* (Supplement_1), S3–S4.

(6) Hur, G. H.; Vickery, C. R.; Burkart, M. D. Explorations of Catalytic Domains in Non-Ribosomal Peptide Synthetase Enzymology. *Nat. Prod. Rep.* **2012**, *29* (10), 1074.

(7) Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. The Phosphopantetheinyl Transferases: Catalysis of a Post-Translational Modification Crucial for Life. *Nat. Prod. Rep.* **2014**, *31* (1), 61–108.

(8) Süssmuth, R. D.; Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Ed.* **2017**, *56* (14), 3770–3821.

(9) Chen, A.; Re, R. N.; Burkart, M. D. Type II Fatty Acid and Polyketide Synthases: Deciphering Protein–Protein and Protein–Substrate Interactions. *Nat. Prod. Rep.* **2018**, *35* (10), 1029–1045.

(10) Jaremko, M. J.; Davis, T. D.; Corpuz, J. C.; Burkart, M. D. Type II Non-Ribosomal Peptide Synthetase Proteins: Structure, Mechanism, and Protein–Protein Interactions. *Nat. Prod. Rep.* **2020**, *37* (3), 355–379.

(11) Miyanaga, A.; Kudo, F.; Eguchi, T. Protein–Protein Interactions in Polyketide Synthase–Nonribosomal Peptide Synthetase Hybrid Assembly Lines. *Nat. Prod. Rep.* **2018**, *35* (11), 1185–1209.

(12) Brown, A. S.; Calcott, M. J.; Owen, J. G.; Ackerley, D. F. Structural, Functional and Evolutionary Perspectives on Effective Re-Engineering of Non-Ribosomal Peptide Synthetase Assembly Lines. *Nat. Prod. Rep.* **2018**, *35* (11), 1210–1228.

(13) Cane, D. E.; Walsh, C. T.; Khosla, C. Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations. *Science* **1998**, *282* (5386), 63–68.

(14) Staunton, J.; Wilkinson, B. Combinatorial Biosynthesis of Polyketides and Nonribosomal Peptides. *Current Opinion in Chemical Biology* **2001**, 5 (2), 159-164.

(15) Miller, B. R.; Gulick, A. M. Structural Biology of Nonribosomal Peptide Synthetases. In *Nonribosomal Peptide and Polyketide Biosynthesis*; Evans, B. S., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2016; Vol. 1401, pp 3–29.

(16) Izoré, T.; Cryle, M. J. The Many Faces and Important Roles of Protein–Protein Interactions during Non-Ribosomal Peptide Synthesis. *Nat. Prod. Rep.* **2018**, *35* (11), 1120–1139.

(17) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. A New Enzyme Superfamily — the Phosphopantetheinyl Transferases. *Chemistry & Biology* **1996**, *3* (11), 923–936.

(18) Tufar, P.; Rahighi, S.; Kraas, F. I.; Kirchner, D. K.; Löhr, F.; Henrich, E.; Köpke, J.; Dikic, I.; Güntert, P.; Marahiel, M. A.; Dötsch, V. Crystal Structure of a PCP/Sfp Complex Reveals the Structural Basis for Carrier Protein Posttranslational Modification. *Chemistry & Biology* **2014**, *21* (4), 552–562.

(19) Nguyen, M. C.; Saurel, O.; Carivenc, C.; Gavalda, S.; Saitta, S.; Tran, M. P.; Milon, A.; Chalut, C.; Guilhot, C.; Mourey, L.; Pedelacq, J. Conformational Flexibility of

Coenzyme A and Its Impact on the Post-translational Modification of Acyl Carrier Proteins by 4′-phosphopantetheinyl Transferases. *FEBS J.* **2020**, *287* (21), 4729–4746.

(20) Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. Genetically Encoded Short Peptide Tag for Versatile Protein Labeling by Sfp Phosphopantetheinyl Transferase. *Proceedings of the National Academy of Sciences* **2005**, *102* (44), 15815–15820.

(21) Zhou, Z.; Cironi, P.; Lin, A. J.; Xu, Y.; Hrvatin, S.; Golan, D. E.; Silver, P. A.; Walsh, C. T.; Yin, J. Genetically Encoded Short Peptide Tags for Orthogonal Protein Labeling by Sfp and AcpS Phosphopantetheinyl Transferases. *ACS Chem. Biol.* **2007**, *2* (5), 337–346.

(22) Tallorin, L.; Wang, J.; Kim, W. E.; Sahu, S.; Kosa, N. M.; Yang, P.; Thompson, M.; Gilson, M. K.; Frazier, P. I.; Burkart, M. D.; Gianneschi, N. C. Discovering de Novo Peptide Substrates for Enzymes Using Machine Learning. *Nat Commun* **2018**, *9* (1), 5253.

(23) Gulick, A. M. Conformational Dynamics in the Acyl-CoA Synthetases, Adenylation Domains of Non-Ribosomal Peptide Synthetases, and Firefly Luciferase. *ACS Chem. Biol.* **2009**, *4* (10), 811–827.

(24) Yonus, H.; Neumann, P.; Zimmermann, S.; May, J. J.; Marahiel, M. A.; Stubbs, M. T. Crystal Structure of DltA. *Journal of Biological Chemistry* **2008**, *283* (47), 32484–32491.

(25) Wu, R.; Reger, A. S.; Lu, X.; Gulick, A. M.; Dunaway-Mariano, D. The Mechanism of Domain Alternation in the Acyl-Adenylate Forming Ligase Superfamily Member 4-Chlorobenzoate: Coenzyme A Ligase ̕. *Biochemistry* **2009**, *48* (19), 4115–4125.

(26) Mitchell, C. A.; Shi, C.; Aldrich, C. C.; Gulick, A. M. Structure of PA1221, a Nonribosomal Peptide Synthetase Containing Adenylation and Peptidyl Carrier Protein Domains. *Biochemistry* **2012**, *51* (15), 3252–3263.

(27) Sundlov, J. A.; Shi, C.; Wilson, D. J.; Aldrich, C. C.; Gulick, A. M. Structural and Functional Investigation of the Intermolecular Interaction between NRPS Adenylation and Carrier Protein Domains. *Chemistry & Biology* **2012**, *19* (2), 188–198.

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

(29) Reimer, J. M.; Eivaskhani, M.; Harb, I.; Guarné, A.; Weigt, M.; Schmeing, T. M. Structures of a Dimodular Nonribosomal Peptide Synthetase Reveal Conformational Flexibility. *Science* **2019**, *366* (6466), eaaw4388.

(30) Reimer, J. M.; Aloise, M. N.; Harrison, P. M.; Martin Schmeing, T. Synthetic Cycle of the Initiation Module of a Formylating Nonribosomal Peptide Synthetase. *Nature* **2016**, *529* (7585), 239–242.

(31) Winn, M.; Fyans, J. K.; Zhuo, Y.; Micklefield, J. Recent Advances in Engineering Nonribosomal Peptide Assembly Lines. *Nat. Prod. Rep.* **2016**, *33* (2), 317–347.

(32) Alanjary, M.; Cano-Prieto, C.; Gross, H.; Medema, M. H. Computer-Aided Re-Engineering of Nonribosomal Peptide and Polyketide Biosynthetic Assembly Lines. *Nat. Prod. Rep.* **2019**, *36* (9), 1249–1261.

(33) Niquille, D. L.; Hansen, D. A.; Hilvert, D. Reprogramming Nonribosomal Peptide Synthesis by Surgical Mutation. *Synlett* **2019**, *30* (19), 2123–2130.

(34) Kaniusaite, M.; Kittilä, T.; Goode, R. J. A.; Schittenhelm, R. B.; Cryle, M. J. Redesign of Substrate Selection in Glycopeptide Antibiotic Biosynthesis Enables Effective Formation of Alternate Peptide Backbones. *ACS Chem. Biol.* **2020**, *15* (9), 2444–2455.

(35) Ishikawa, F.; Nohara, M.; Nakamura, S.; Nakanishi, I.; Tanabe, G. Precise Probing of Residue Roles by NRPS Code Swapping: Mutation, Enzymatic Characterization, Modeling, and Substrate Promiscuity of Aryl Acid Adenylation Domains. *Biochemistry* **2020**, *59* (4), 351–363.

(36) Mootz, H. D.; Schwarzer, D.; Marahiel, M. A. Construction of Hybrid Peptide Synthetases by Module and Domain Fusions. *Proceedings of the National Academy of Sciences* **2000**, *97* (11), 5848–5853.

(37) Nguyen, K. T.; Ritz, D.; Gu, J.-Q.; Alexander, D.; Chu, M.; Miao, V.; Brian, P.; Baltz, R. H. Combinatorial Biosynthesis of Novel Antibiotics Related to Daptomycin. *Proceedings of the National Academy of Sciences* **2006**, *103* (46), 17462–17467.

(38) Qiao, C.; Wilson, D. J.; Bennett, E. M.; Aldrich, C. C. A Mechanism-Based Aryl Carrier Protein/Thiolation Domain Affinity Probe. *J. Am. Chem. Soc.* **2007**, *129* (20), 6350–6351.

(39) Lu, X.; Zhang, H.; Tonge, P. J.; Tan, D. S. Mechanism-Based Inhibitors of MenE, an Acyl-CoA Synthetase Involved in Bacterial Menaquinone Biosynthesis. *Bioorganic & Medicinal Chemistry Letters* **2008**, *18* (22), 5963–5966.

(40) Olsen, S. K.; Capili, A. D.; Lu, X.; Tan, D. S.; Lima, C. D. Active Site Remodelling Accompanies Thioester Bond Formation in the SUMO E1. *Nature* **2010**, *463* (7283), 906–912.

(41) Lu, X.; Olsen, S. K.; Capili, A. D.; Cisar, J. S.; Lima, C. D.; Tan, D. S. Designed Semisynthetic Protein Inhibitors of Ub/Ubl E1 Activating Enzymes. *J. Am. Chem. Soc.* **2010**, *132* (6), 1748–1749.

(42) Ji, C.; Sharma, I.; Pratihar, D.; Hudson, L. L.; Maura, D.; Guney, T.; Rahme, L. G.; Pesci, E. C.; Coleman, J. P.; Tan, D. S. Designed Small-Molecule Inhibitors of the Anthranilyl-CoA Synthetase PqsA Block Quinolone Biosynthesis in *Pseudomonas Aeruginosa*. *ACS Chem. Biol.* **2016**, *11* (11), 3061–3067.

(43) Lux, M. C.; Standke, L. C.; Tan, D. S. Targeting Adenylate-Forming Enzymes with Designed Sulfonyladenosine Inhibitors. *J Antibiot* **2019**, *72* (6), 325–349.

(44) Jaremko, M. J.; Lee, D. J.; Patel, A.; Winslow, V.; Opella, S. J.; McCammon, J. A.; Burkart, M. D. Manipulating Protein–Protein Interactions in Nonribosomal Peptide Synthetase Type II Peptidyl Carrier Proteins. *Biochemistry* **2017**, *56* (40), 5269–5273.

(45) Corpuz, J. C.; Podust, L. M.; Davis, T. D.; Jaremko, M. J.; Burkart, M. D. Dynamic Visualization of Type II Peptidyl Carrier Protein Recognition in Pyoluteorin Biosynthesis. *RSC Chem. Biol.* **2020**, *1* (1), 8–12.

(46) Gulick, A. M.; Aldrich, C. C. Trapping Interactions between Catalytic Domains and Carrier Proteins of Modular Biosynthetic Enzymes with Chemical Probes. *Nat. Prod. Rep.* **2018**, *35* (11), 1156–1184.

(47) Miyanaga, A.; Kurihara, S.; Chisuga, T.; Kudo, F.; Eguchi, T. Structural Characterization of Complex of Adenylation Domain and Carrier Protein by Using Pantetheine Cross-Linking Probe. *ACS Chem. Biol.* **2020**, *15* (7), 1808–1812.

(48) Jaremko, M. J.; Lee, D. J.; Opella, S. J.; Burkart, M. D. Structure and Substrate Sequestration in the Pyoluteorin Type II Peptidyl Carrier Protein PltL. *J. Am. Chem. Soc.* **2015**, *137* (36), 11546–11549.

(49) Alfermann, J.; Sun, X.; Mayerthaler, F.; Morrell, T. E.; Dehling, E.; Volkmann, G.; Komatsuzaki, T.; Yang, H.; Mootz, H. D. FRET Monitoring of a Nonribosomal Peptide Synthetase. *Nat Chem Biol* **2017**, *13* (9), 1009–1015.

(50) Mayerthaler, F.; Feldberg, A.-L.; Alfermann, J.; Sun, X.; Steinchen, W.; Yang, H.; Mootz, H. D. Intermediary Conformations Linked to the Directionality of the Aminoacylation Pathway of Nonribosomal Peptide Synthetases. *RSC Chem. Biol.* **2021**, *2* (3), 843–854.

(51) Miller, B. R.; Sundlov, J. A.; Drake, E. J.; Makin, T. A.; Gulick, A. M. Analysis of the Linker Region Joining the Adenylation and Carrier Protein Domains of the Modular Nonribosomal Peptide Synthetases: Functional Analysis of NRPS Linkers. *Proteins* **2014**, *82* (10), 2691–2702.

(52) Keating, T. A.; Marshall, C. G.; Walsh, C. T.; Keating, A. E. The Structure of VibH Represents Nonribosomal Peptide Synthetase Condensation, Cyclization and Epimerization Domains. *Nat. Struct Biol.* **2002**.

(53) Samel, S. A.; Schoenafinger, G.; Knappe, T. A.; Marahiel, M. A.; Essen, L.-O. Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. *Structure* **2007**, *15* (7), 781–792.

(54) Tanovic, A.; Samel, S. A.; Essen, L.-O.; Marahiel, M. A. Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase. *Science* **2008**, *321* (5889), 659–663.

(55) Bloudoff, K.; Rodionov, D.; Schmeing, T. M. Crystal Structures of the First Condensation Domain of CDA Synthetase Suggest Conformational Changes during the Synthetic Cycle of Nonribosomal Peptide Synthetases. *Journal of Molecular Biology* **2013**, *425* (17), 3137–3150.

(56) Zhang, J.; Liu, N.; Cacho, R. A.; Gong, Z.; Liu, Z.; Qin, W.; Tang, C.; Tang, Y.; Zhou, J. Structural Basis of Nonribosomal Peptide Macrocyclization in Fungi. *Nat Chem Biol* **2016**, *12* (12), 1001–1003.

(57) Bloudoff, K.; Alonzo, D. A.; Schmeing, T. M. Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases. *Cell Chemical Biology* **2016**, *23* (3), 331–339.

(58) Wang, L.; Yuan, M.; Zheng, J. Crystal Structure of the Condensation Domain from Lovastatin Polyketide Synthase. *Synthetic and Systems Biotechnology* **2019**, *4* (1), 10–15.
(59) Kreitler, D. F.; Gemmell, E. M.; Schaffer, J. E.; Wencewicz, T. A.; Gulick, A. M. The Structural Basis of N-Acyl-α-Amino-β-Lactone Formation Catalyzed by a Nonribosomal Peptide Synthetase. *Nat Commun* **2019**, *10* (1), 3432.

(60) Kosol, S.; Gallo, A.; Griffiths, D.; Valentic, T. R.; Masschelein, J.; Jenner, M.; de los Santos, E. L. C.; Manzi, L.; Sydor, P. K.; Rea, D.; Zhou, S.; Fülöp, V.; Oldham, N. J.; Tsai, S.-C.; Challis, G. L.; Lewandowski, J. R. Structural Basis for Chain Release from the Enacyloxin Polyketide Synthase. *Nat. Chem.* **2019**, *11* (10), 913–923.

(61) Tan, K.; Zhou, M.; Jedrzejczak, R. P.; Wu, R.; Higuera, R. A.; Borek, D.; Babnigg, G.; Joachimiak, A. Structures of Teixobactin-Producing Nonribosomal Peptide Synthetase Condensation and Adenylation Domains. *Current Research in Structural Biology* **2020**, *2*, 14–24.

(62) Bloudoff, K.; Schmeing, T. M. Structural and Functional Aspects of the Nonribosomal Peptide Synthetase Condensation Domain Superfamily: Discovery, Dissection and Diversity. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2017**, *1865* (11), 1587–1604.

(63) Calcott, M. J.; Owen, J. G.; Ackerley, D. F. Efficient Rational Modification of Non-Ribosomal Peptides by Adenylation Domain Substitution. *Nat Commun* **2020**, *11* (1), 4554.

(64) Kranz, J.; Wenski, S. L.; Dichter, A. A.; Bode, H. B.; Bozhüyük, K. A. J. Influence of Condensation Domains on Activity and Specificity of Adenylation Domains. BioRxiv. Preprint, **2021**.

(65) Baunach, M.; Chowdhury, S.; Stallforth, P.; Dittmann, E. The Landscape of Recombination Events That Create Nonribosomal Peptide Diversity. *Molecular Biology and Evolution* **2021**, *38* (5), 2116–2130.

(66) Izoré, T.; Candace Ho, Y. T.; Kaczmarski, J. A.; Gavriilidou, A.; Chow, K. H.; Steer, D. L.; Goode, R. J. A.; Schittenhelm, R. B.; Tailhades, J.; Tosin, M.; Challis, G. L.; Krenske, E. H.; Ziemert, N.; Jackson, C. J.; Cryle, M. J. Structures of a Non-Ribosomal Peptide Synthetase Condensation Domain Suggest the Basis of Substrate Selectivity. *Nat Commun* **2021**, *12* (1), 2511.

(67) Stanisic, A.; Husken, A.; Stephan, P.; Niquille, D. L.; Reinstein, J.; Kries, H. An Engineered Nonribosomal Peptide Synthetase Shows Opposite Amino Acid Loading and Condensation Specificity. *ACS Catalysis* **2021**, 11 (14), 8692–8700.

(68) Wheadon, M. J.; Townsend, C. A. Evolutionary and Functional Analysis of an NRPS Condensation Domain Integrates β-Lactam, ᴅ-Amino Acid, and Dehydroamino Acid Synthesis. *Proc Natl Acad Sci USA* **2021**, *118* (17), e2026017118.

(69) Dekimpe, S.; Masschelein, J. Beyond Peptide Bond Formation: The Versatile Role of Condensation Domains in Natural Product Biosynthesis. *Nat. Prod. Rep.* **2021**, *38* (10), 1910–1937.

(70) Chen, W.-H.; Li, K.; Guntaka, N. S.; Bruner, S. D. Interdomain and Intermodule Organization in Epimerization Domain Containing Nonribosomal Peptide Synthetases. *ACS Chem. Biol.* **2016**, *11* (8), 2293–2303.

(71) Kim, W. E.; Patel, A.; Hur, G. H.; Tufar, P.; Wuo, M. G.; McCammon, J. A.; Burkart, M. D. Mechanistic Probes for the Epimerization Domain of Nonribosomal Peptide Synthetases. *ChemBioChem* **2019**, *20* (2), 147–152.

(72) Shi, C.; Miller, B. R.; Alexander, E. M.; Gulick, A. M.; Aldrich, C. C. Design, Synthesis, and Biophysical Evaluation of Mechanism-Based Probes for Condensation Domains of Nonribosomal Peptide Synthetases. *ACS Chem. Biol.* **2020**, *15* (7), 1813–1819.

(73) Bozhüyük, K. A. J.; Fleischhacker, F.; Linck, A.; Wesche, F.; Tietze, A.; Niesert, C.-P.; Bode, H. B. De Novo Design and Engineering of Non-Ribosomal Peptide Synthetases. *Nature Chem* **2018**, *10* (3), 275–281.

(74) Bozhüyük, K. A. J.; Linck, A.; Tietze, A.; Kranz, J.; Wesche, F.; Nowak, S.; Fleischhacker, F.; Shi, Y.-N.; Grün, P.; Bode, H. B. Modification and de Novo Design of Non-Ribosomal Peptide Synthetases Using Specific Assembly Points within Condensation Domains. *Nat. Chem.* **2019**, *11* (7), 653–661.

(75) Liu, Q.; Fan, W.; Zhao, Y.; Deng, Z.; Feng, Y. Probing and Engineering the Fatty Acyl Substrate Selectivity of Starter Condensation Domains of Nonribosomal Peptide Synthetases in Lipopeptide Biosynthesis. *Biotechnol. J.* **2020**, *15* (2), 1900175.

(76) Zhong, L.; Diao, X.; Zhang, N.; Li, F.; Zhou, H.; Chen, H.; Bai, X.; Ren, X.; Zhang, Y.; Wu, D.; Bian, X. Engineering and Elucidation of the Lipoinitiation Process in Nonribosomal Peptide Biosynthesis. *Nat Commun* **2021**, *12* (1), 296.

(77) Horsman, M. E.; Hari, T. P. A.; Boddy, C. N. Polyketide Synthase and Non-Ribosomal Peptide Synthetase Thioesterase Selectivity: Logic Gate or a Victim of Fate? *Nat. Prod. Rep.* **2016**, *33* (2), 183–202.

(78) Haslinger, K.; Brieke, C.; Uhlmann, S.; Sieverling, L.; Süssmuth, R. D.; Cryle, M. J. The Structure of a Transient Complex of a Nonribosomal Peptide Synthetase and a Cytochrome P450 Monooxygenase. *Angew. Chem. Int. Ed.* **2014**, *53* (32), 8518–8522.

(79) Thapa, H. R.; Robbins, J. M.; Moore, B. S.; Agarwal, V. Insights into Thiotemplated Pyrrole Biosynthesis Gained from the Crystal Structure of Flavin-Dependent Oxidase in Complex with Carrier Protein. *Biochemistry* **2019**, *58* (7), 918–929.

(80) Little, R. F.; Hertweck, C. Chain Release Mechanisms in Polyketide and Non-Ribosomal Peptide Biosynthesis. *Nat. Prod. Rep.* **2021**, 10.1039.D1NP00035G.

(81) Koglin, A.; Löhr, F.; Bernhard, F.; Rogov, V. V.; Frueh, D. P.; Strieter, E. R.; Mofid, M. R.; Güntert, P.; Wagner, G.; Walsh, C. T.; Marahiel, M. A.; Dötsch, V. Structural Basis for the Selectivity of the External Thioesterase of the Surfactin Synthetase. *Nature* **2008**, *454* (7206), 907–911.

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

(83) Guntaka, N. S.; Healy, A. R.; Crawford, J. M.; Herzon, S. B.; Bruner, S. D. Structure and Functional Analysis of ClbQ, an Unusual Intermediate-Releasing Thioesterase from the Colibactin Biosynthetic Pathway. *ACS Chem. Biol.* **2017**, *12* (10), 2598–2608.

(84) Deshpande, S.; Altermann, E.; Sarojini, V.; Lott, J. S.; Lee, T. V. Structural Characterization of a PCP–R Didomain from an Archaeal Nonribosomal Peptide Synthetase Reveals Novel Interdomain Interactions. *Journal of Biological Chemistry* **2021**, *296*, 100432.

(85) Gahloth, D.; Dunstan, M. S.; Quaglia, D.; Klumbys, E.; Lockhart-Cairns, M. P.; Hill, A. M.; Derrington, S. R.; Scrutton, N. S.; Turner, N. J.; Leys, D. Structures of Carboxylic Acid Reductase Reveal Domain Dynamics Underlying Catalysis. *Nat Chem Biol* **2017**, *13* (9), 975–981.

(86) Bartholow, T. G.; Sztain, T.; Patel, A.; Lee, D. J.; Young, M. A.; Abagyan, R.; Burkart, M. D. Elucidation of Transient Protein-Protein Interactions within Carrier Protein-Dependent Biosynthesis. *Commun Biol* **2021**, *4* (1), 340.

(87) Schaub, A. J.; Moreno, G. O.; Zhao, S.; Truong, H. V.; Luo, R.; Tsai, S.-C. Computational Structural Enzymology Methodologies for the Study and Engineering of Fatty Acid Synthases, Polyketide Synthases and Nonribosomal Peptide Synthetases. In *Methods in Enzymology*; Elsevier, **2019**; Vol. 622, pp 375–409.

(88) Kotowska, M.; Pawlik K. Roles of type II thioesterases and their application for secondary metabolite yield improvement, *Appl Microbiol Biotechnol* **2014,** *98*, 7735–7746.

(89) Sun, D., Peyear, T. A., Bennet, W. D., Andersen, O. S., Lightstone, F. C., Ingolfsson, H. I. Molecular mechanism for gramicidin dimerization and dissociation in bilayers of different thickness. *Biophysical Journal* **2019**, *117* (10), 1831-1844.

CHAPTER 2. DYNAMIC VISUALIZATION OF TYPE II PEPTIDYL CARRIER PROTEIN RECOGNITION IN PYOLUTEORIN BIOSYNTHESIS

2.1. Introduction

The type II non-ribosomal peptide synthase (NRPS) system consist of stand-alone enzymes that commonly participate in hybrid pathways along with fatty acid synthase (FAS) and polyketide synthase (PKS) enzymes.[1] The stand-alone architecture of type II NRPS proteins makes them promising candidates for metabolic engineering, as they commonly serve to install unique chemical functionality into growing metabolites. These hybrid pathways afford complex natural products that include antibiotic, antitumor, and antifungal agents.[2] Examples of functionalized natural products include those that utilize dehydrogenated prolines as a pharmacophore, including prodigiosin, pyoluteorin, and chlorizidine A, all of which require type II NRPS proteins to functionalize and incorporate a pyrrole into the natural product (Figure 2.1).[3-5]

All NRPS include an adenylation (A) and peptidyl carrier protein (PCP) domain to activate and load substrates. The A domain activates a specific amino acid substrate and subsequently installs the substrate onto the thiol of the 4'-phosphopantetheine (PPant) arm of the PCP (Figure 2.1A).[6] The PCP is a 10 kDa protein that consists of a 4-helix bundle.[7] The PPant modification is attached onto an invariant serine via a phosphopantetheinyl transferase to activate the PCP to the *holo-* form. The PCP can transport acyl or aminoacyl cargo to a variety of enzymatic domains for functionalization and incorporation of the substrate into the nascent natural product.

During pyoluteorin biosynthesis in *Pseudomonas fluorescens*, the type II NRPS PCP, *holo*-PltL, is loaded with L-Pro by the A domain, PltF (Figure 2.1). While protecting
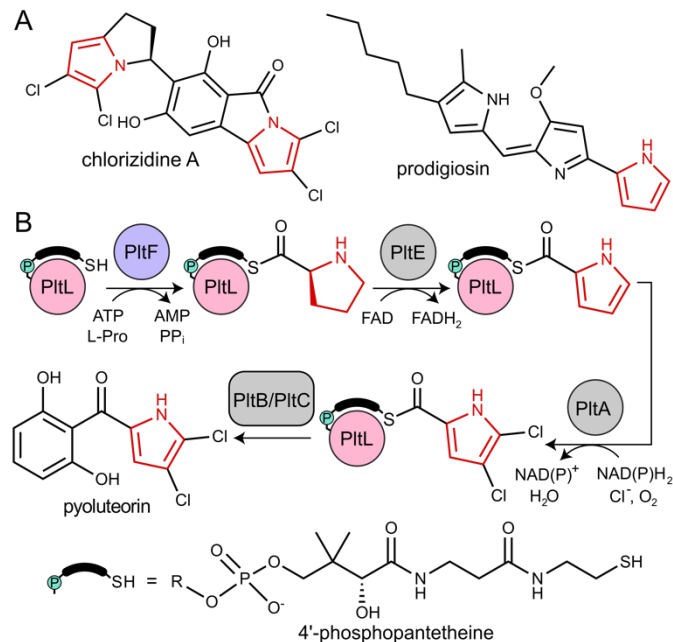
Figure 2.1: Pyrrolidine examples and functionalization. A) Natural product examples that incorporate pyrrole (red) via type II NRPS proteins. B) Pyrrole functionalization and installation in pyoluteorin via type II NRPS proteins. PltB and PltC is a type I PKS.

its substrate in a hydrophobic cleft,[8] prolyl-PltL transports the L-Pro for dehydrogenation and dichlorination before being off-loaded onto a type I PKS for the incorporation of the dichloropyrrolyl substrate into pyoluteorin.[9]

PltL has been shown to exhibit specificity towards PltF and no interactivity towards homologous A domains.[10,11] This suggested the requirement of a specific protein-protein interaction motif for A domain activity. Studies have attempted to control the partner protein specificity of PltL; solution-phase nuclear magnetic resonance (NMR) titration experiments revealed a region of PltL, loop 1 (residues 19-41), that was postulated to form the protein-protein interface with PltF.[10] Mutagenesis of residues in this region disrupted activity, however, these studies could not resolve the PCP-A domain interface clearly enough to accurately manipulate PltL specificity.

2.2. Determination and validation of the PltL-PltF X-ray crystal structure

In order to determine the basis of molecular recognition between PltL and PltF, we set out to structurally characterize the PltL-PltF complex with X-ray crystallography. Carrier proteins (CPs) and partner proteins often form weak transient interactions and high disassociation may impede co-crystallization.[12] To stabilize the interaction between PltL and PltF, a substrate mimic of the proline adenosine monophosphate (Pro-AMP) intermediate was deemed necessary. Based on a covalent inhibitor motif developed by the Aldrich and Tan groups,[13-18] the proline adenosine vinylsulfonamide (Pro-AVSN) was synthesized (see Supporting Information) and employed to trap PltL with PltF (Fig 2.2B). Incubation of *holo*-PltL, PltF and Pro-AVSN enabled crystallization and optimization led to a crystal that diffracted to 2.15 Å. Molecular replacement was carried out using the A
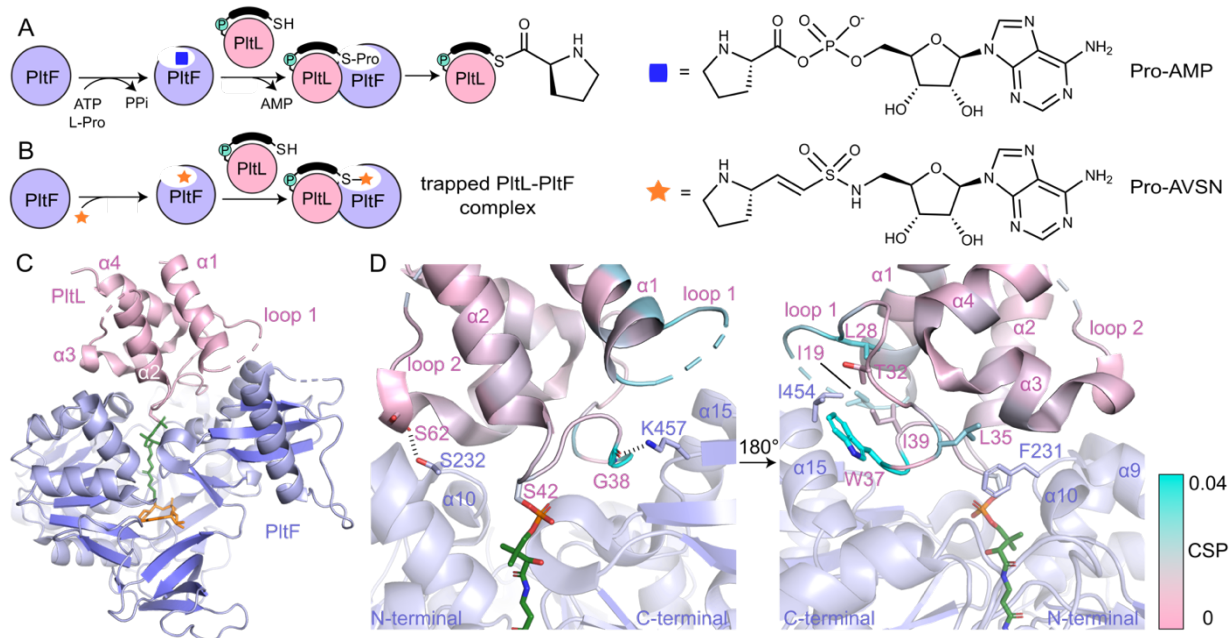
Figure 2.2: The PltL-PltF crystal structure. A) Adenylation and thiolation of L-Pro by PltF. B) Trapping of the PltL-PltF complex with Pro-AVSN (orange star). C) 2.15 Å resolution crystal structure of the PltL-PltF complex. PltL (pink) is trapped to PltF (purple) via Pro-AVSN (orange). PPant is shown in green. D) Close-up of the PltL-PltF interface. The left panel displays hydrogen bonding interactions. The right panel, which is rotated 180° from the left, displays hydrophobic interactions. The CSPs (cyan) from a recent NMR titration of PltL with PltF are mapped onto the bound PltL.

47

domain, DltA (PDB ID 3E7W), in the thiolation state[19] and with PltL (PDB ID 2N5H) (Figure 2.S2).[8]

The PltL-PltF crystal structure contains PltL trapped with PltF in the thiolation state via addition to Pro-AVSN (Figure 2.2C). Successful trapping of the complex is shown by electron density of the PPant extended into the active site of PltF, with its terminal thiol covalently linked to the sulfonamide β-carbon of Pro-AVSN (Figure 2.S2). The putative K486 responsible for adenylation is 25 Å away from the active site, consistent with the domain alternation hypothesis proposed by Gulick and coworkers (Figure 2.S3A).[20] The N-terminal domain (NTD) contains an AMP binding pocket and L-Pro binding pocket that is conserved amongst A domains that process L-Pro (Figure 2.S3A, Figure 2.S4).[21] The bound PltL maintains the conserved four-α helix bundle, with the PPant attached onto Ser42 at the top of helix 2. (Figure 2.S3B). Loop 1 residues 20-25 had poor electron density, so no atomic model was built for this region.

The PltL-PltF crystal structure reveals a protein-protein interface that is mediated by interactions between the loop 1 region of PltL with PltF helix 15 and the turn between helix 9 and 10 (Figure 2.2D, Figure 2.3B). A portion of PltL helix 2 and loop 2 also contribute to the interface. The backbone carbonyl of PltL Ser62 and Gly38 form hydrogen bonds (H-bonds) with PltF Ser232 and Lys457, respectively (Figure 2.2D). The phosphate of the PPant arm also forms an electrostatic interaction with PltF Arg404. PltF Ile454 is positioned inside a hydrophobic pocket created by PltL loop 1 residues Ile19, Leu28, Trp37, and Ile39 (Figure 2.2D, Figure 2.S5B). The indole ring of PltL Trp37 is buried along PltF helix 15 (Figure 2.S5A). PltF Phe231 is observed in a hydrophobic cleft between PltL helices 2, 3 and loop 1 residue Leu35 (Figure 2.2D, Figure 2.3C).

Previously reported NMR titration studies of *holo*-^{15}N-PltL with PltF revealed significant chemical shift perturbations (CSPs) in PltL loop 1 residues (Figure 2.S6), implicating their participation in, or response to, the binding event.[10] Of those residues, Ile19, Leu28, Leu35, Trp37, and Gly38 were seen at the protein-protein interface in the crystal structure (Figure 2.2D). The remaining residues with high CSPs are likely due to changes in the secondary structural interactions between helix 1 and loop 1 (Figure 2.S7).

2.3. Identification of PCP dynamics upon PltF binding

To gain insight into the highest relative CSPs found in PltL Trp37 and Gly38, the solution NMR structure of *holo*-PltL was aligned and superposed to PltL in the bound state with PltF (Figure 2.3B). Comparison of the different states reveal the lack of the loop 1 hydrophobic pocket in the unbound PltL that contacts PltF Ile454 (Figure 2.3A, Figure 2.S7). The superposition shows the pocket formation via a 5.5 Å displacement of the indole ring of PltL Trp37 by PltF Ile454. Adjacent to Trp37 is Gly38, which forms a H-bond via its backbone carbonyl with PltF Lys457 that may stabilize the displacement. The relevance of the hydrophobic pocket formation and H-bonding was demonstrated by mutagenesis of Ile454Ala and Lys457Ala, respectively, where individual mutations both stunted aminoacylation activity (Figure 2.3D). Furthermore, previous molecular dynamic simulations revealed the relative flexibility of PltL loop 1, which supports loop 1 reorganization upon binding.[10] These data provide evidence that specific H-bonding and hydrophobic interactions allowed by the conformational flexibility of PltL loop 1 is responsible for its recognition towards PltF.

49
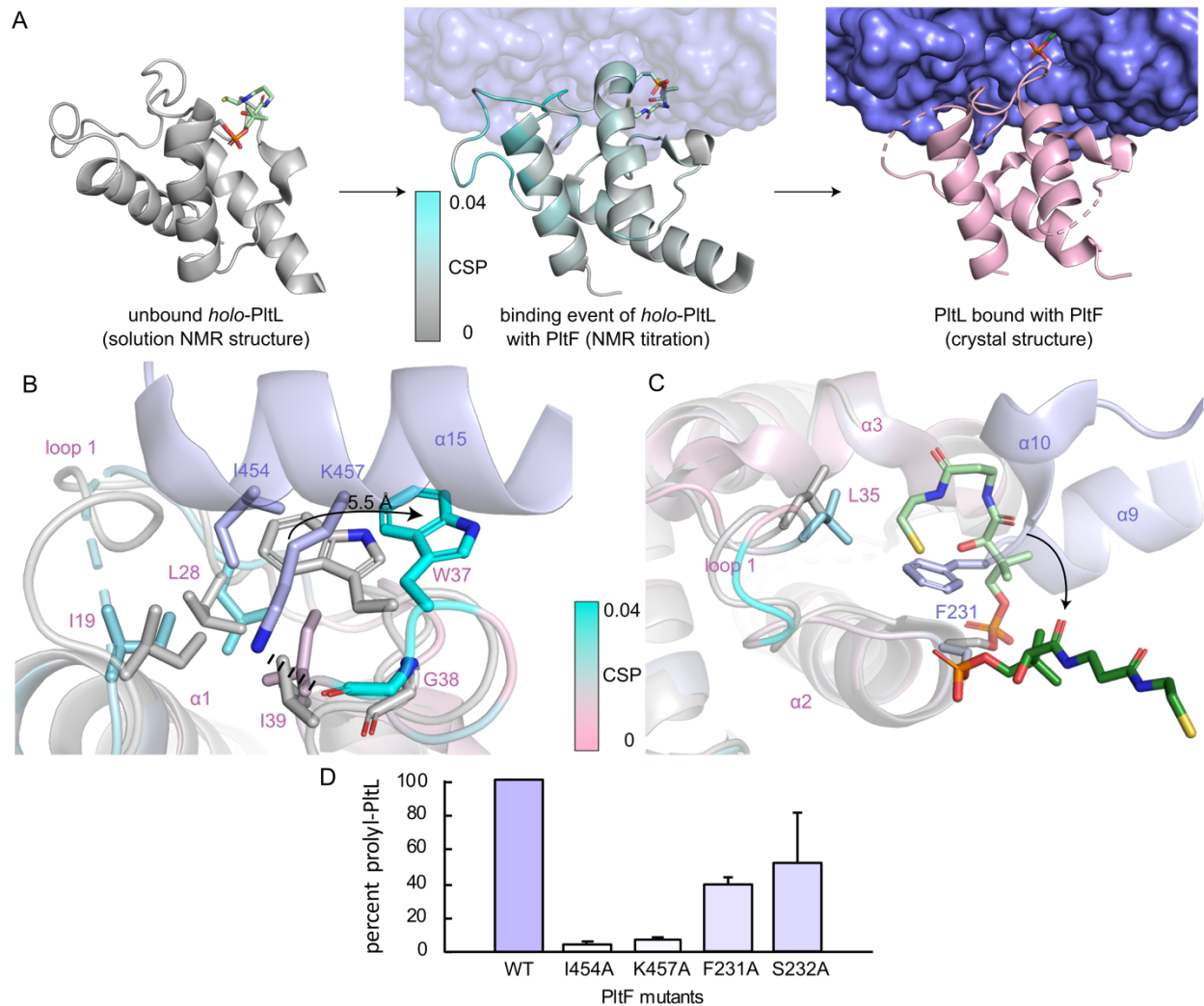
**Figure 2.3:** Visualization of the PltL-PltF binding event. A) Linking of the solution NMR structure of *holo*-PltL (gray, PDB ID 2N5H), the CSPs (cyan) from the NMR titration, and the PltL-PltF (pink, purple) complex. B) Superposition of the bound (pink, cyan) and unbound (gray) PltL with PltF Ile454 (purple) and C) Phe231 (purple). Pro-AVSN was omitted for clarity. D) PltF mutant aminoacylation assay with PltL.

Superposition of bound and unbound PltL also reveals PltF Phe231 residing between helices 2 and 3 of PltL, which occludes the hydrophobic cleft that protects the substrate (Figure 2.3C).[8] This suggests that the hydrophobic cleft is not only involved in substrate protection but also molecular recognition. Mutagenesis of PltF Phe231Ala and Ser232Ala results in decreased PltL aminoacylation, which supports its relevance in PltL-PltF recognition (Figure 2.3D). Recent structural analysis of the type II PCP of pentabromopseudilin biosynthesis, Bmp1, in complex with the oxidase, Bmp2, yields similar conclusions.[22]

2.4. Analysis of type I and type II PCP-A domain interfaces

Next, we compared the PltL-PltF interface against other PCP-A domain interfaces from the crystal structures of LgrA, EntF, EntE-EntB, and PA1221.[14,15,23,24] The most outstanding difference is the location of the interface interactions. The PCP interfaces of LgrA, EntF, PA1221, and EntE-EntB involve the loop 1 and, to a larger extent, helix 2 of the PCP (Figure 2.4). In contrast, the PltL-PltF structure reveals that only the first residue of PltL helix 2, Met43, is partially buried at the interface. The loop 1 region of PltL is instead the main contributor to the interface with PltF. This interface independent of helix 2 is surprising considering the established prevalence of helix 2 at the interface of CP-dependent pathways, such as the *E. coli* type II FAS, where the CP-enzyme interface primarily relies on helix 2 for binding (Figure 2.S9).[25-27]

While PCP-A domain interfaces consist of hydrophobic interactions, each pathway differs in number of H-bonds and salt bridges for partner protein recognition (Figure 2.4). The PltL-PltF structure reveals two H-bonds at the interface, whereas EntE-EntB employs
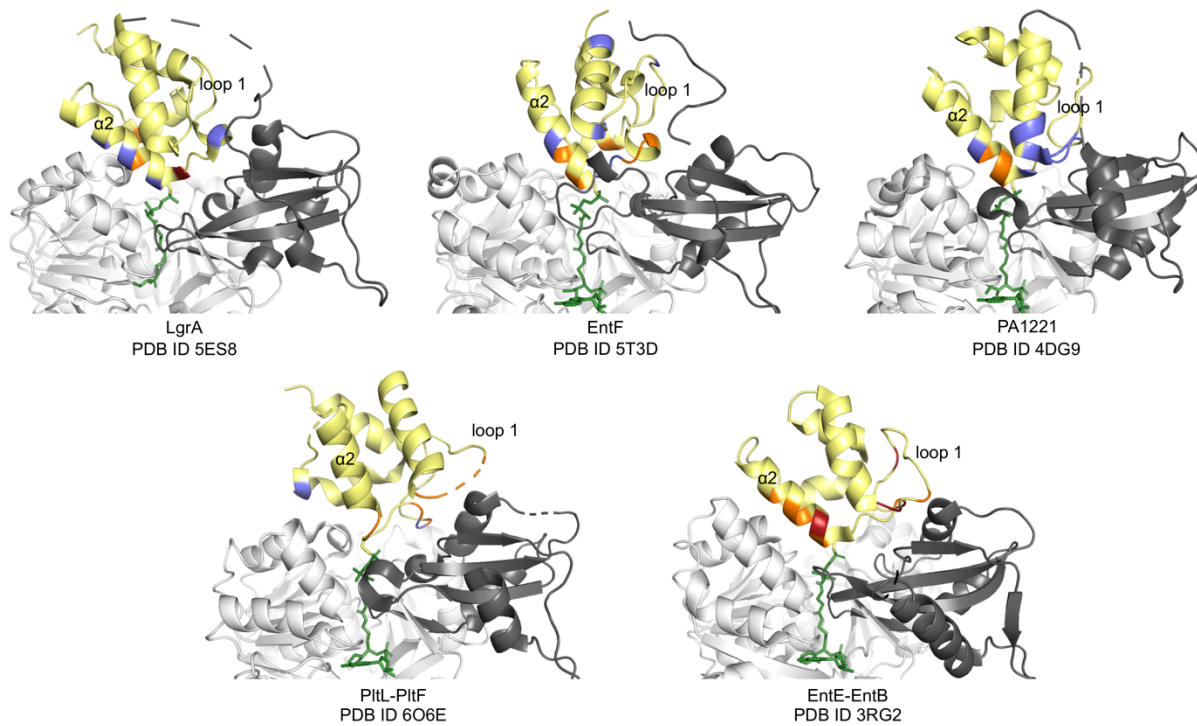
Figure 2.4: PCP-A domain interface comparison. Yellow, PCP; dark grey, A domain C-terminal domain; white, A domain N-terminal domain; blue, H-bonding; red, salt bridge; orange, hydrophobic.

multiple salt bridges. In contrast, the LgrA, EntF, and PA1221 structures contain four or more H-bonds. The type of interaction in addition to the location of each interaction presents a challenge towards the combinatorial biosynthesis of CP-dependent pathways.

2.5. Conclusions

The current structural investigation enhances our understanding of the molecular basis of PltL-PltF interactions. While our previous NMR titrations identified potential interface residues on PltL, mutagenesis studies did not afford noncognate activity. Here, combining NMR titrations, chemical trapping, X-ray crystallography, and mutagenesis assays allows us to more precisely identify the interface to understand specificity and predict functional mutations.

This structure of the trapped PltL-PltF complex, complemented by prior NMR studies, has revealed new insights into a distinct mechanism of recognition used by A domains and PCPs. Resolving structural features of the protein-protein interface revealed the differences in the type, location, and dynamics of interfacial interactions that govern CP and partner protein recognition. These differences may serve a fundamental reason behind the limited success of prior combinatorial biosynthetic efforts in NRPSs. With our discoveries on the type II PCP and A domain from pyoluteorin biosynthesis, a layer of complexity has been revealed that will inform the future combinatorial biosynthetic efforts and engineering of CP-dependent pathways.

2.6. Acknowledgements

Chapter 2, entitled Dynamic visualization of type II peptidyl carrier protein recognition in pyoluteorin biosynthesis, in full, is a reprint of the material as it appears in Corpuz, J. C.; Podust L. M.; Davis, T. D.; Jaremko M. J.; Burkart, M. D. Dynamic visualization of type II peptidyl carrier protein recognition in pyoluteorin biosynthesis. *RSC Chemical Biology,* **2020**, 1, 8-12. The dissertation author was the primary investigator and author of this paper.

2.7. Methods

2.7.A. Synthesis of prolyl-adenosine vinylsulfonamide probe

Chemical reagents were purchased from Acros, Fluka, Sigma-Aldrich, or TCI. Deuterated NMR solvents were purchased from Cambridge Isotope Laboratories. All reactions were conducted with vigorously dried anhydrous solvents that were obtained by passing through a solvent column exposed of activated A2 alumina. All reactions were performed under positive pressure of argon in flame-dried glassware sealed with septa and stirred with Teflon coated stir bars using an IKAMAG TCT-basic mechanical stirrer (IKA GmbH). Analytical Thin Layer Chromatograpy (TLC) was performed on Silica Gel 60 F254 precoated glass plates (EM Sciences). Visualization was achieved with UV light and/or appropriate stain (I_2 on SiO_2, $KMnO_4$, bromocresol green, dinitrophenylhydrazine, ninhydrin, or ceric ammonium molybdate). Flash column chromatography was carried out with Geduran Silica Gel 60 (40–63 mesh) from EM Biosciences. Yield and characterization data correspond to isolated, chromatographically, and spectroscopically homogeneous materials. ^{1}H NMR spectra were recorded on Varian Mercury 400, Varian

Mercury Plus 400, or JEOL ECA500 spectrometers. ^{13}C NMR spectra were recorded at 100 MHz on Varian Mercury 400 or Varian Mercury Plus 400 spectrometers. Chemical shifts for ^{1}H NMR and ^{13}C NMR analyses were referenced to the reported values of Gottlieb[28] using the signal from the residual solvent for ^{1}H spectra, or to the ^{13}C signal from the deuterated solvent. Chemical shift δ values for the ^{1}H and ^{13}C spectra are reported in parts per millions (ppm) relative to these referenced values, and multiplicities are abbreviated as s=singlet, d=doublet, t=triplet, q-quartet, m=multiplet, b=broad. All ^{13}C NMR spectra were recorded with complete proton decoupling. FID files were processed using MestreNova 10.0 (MestreLab Research). Electrospray ionization (ESI) mass spectrometric analyses were preformed using a ThermoFinnigan LCQ Deca spectrometer. Spectral data and procedures are provided for all new compounds and copies of spectra have been provided.
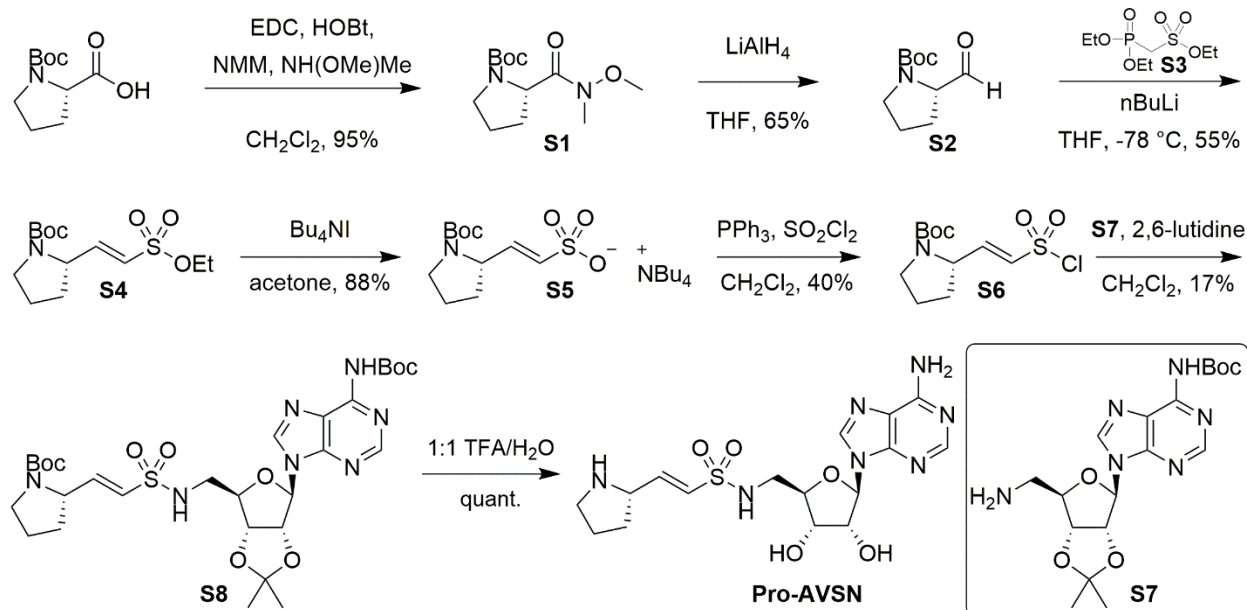
Figure 2.S1: Synthesis of prolyl-adenosine vinylsulfonamide (Pro-AVSN). EDC=1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide); HOBt=hydroxybenzotriazole; NMM=*N*-methylmorpholine; TFA=trifluoroacetic acid.

tert-butyl (S)-2-(methoxy(methyl)carbamoyl)pyrrolidine-1-carboxylate (S1).

In a 250 mL round-bottom flask, *N*-Boc-l-proline (5.11 g, 23.7 mmol, 1.0 equiv.) and 100 mL CH_2Cl_2 were added. The vessel was cooled to 0 °C before the addition of HOBt (3.54 g, 26.2 mmol, 1.1 equiv.) and EDC•HCl (5.50 g, 28.7 mmol, 1.2 equiv.). The reaction was stirred for 20 minutes at 0 °C before the addition of *N,O*-dimethylhydroxylamine-hydrochloride (2.69 g, 27.6 mmol, 1.2 equiv.) and 4-methylmorpholine (3.0 mL, 27.3 mmol, 1.2 equiv.). The reaction was warmed to room temperature and stirred overnight. After 14 h, the volatiles were removed by rotary evaporation and the resulting residue was dissolved in EtOAc (250 mL) and washed with 1 N HCl (2x50 mL), saturated $NaHCO_3$ (50 mL), and brine (50 mL). The organic phase was dried ($MgSO_4$), filtered, and concentrated by rotary evaporation. Purification by silica flash chromatography (1:1 hexanes/diethyl ether → diethyl ether) afforded Weinreb amide S1 (5.81 g, 95%) as a clear viscous oil.

TLC: R_f 0.57 (diethyl ether). ^{1}H-NMR (400 MHz, $CDCl_3$): δ4.58 (dd, *J* = 39.3, 8.7, 3.4 Hz, 1H), 3.71 (s, 2H), 3.65 (s, 1H), 3.55–3.44 (m, 1H), 3.44–3.29 (m, 1H), 3.12 (s, 3H), 2.22–2.03 (m, 1H), 1.98–1.86 (m, 1H), 1.85–1.71 (m, 2H), 1.38 (s, 5H), 1.34 (s, 4H). ^{13}C-NMR (100 MHz, $CDCl_3$): δ 154.50, 153.92, 79.62, 61.28, 56.83, 46.88, 32.42, 30.49, 29.61, 28.48, 24.05, 23.40. HR-ESI-MS *m/z* calcd. $[C_{12}H_{22}N_2O_4Na]^+$: 281.1472, found 281.1469.

tert-butyl (S)-2-formylpyrrolidine-1-carboxylate (S2).

In a 200 mL pear-shaped flask, *N*-Boc-l-proline Weinreb amide S1 (2.02 g, 7.81 mmol, 1.0 equiv.) and 100 mL THF were added. The vessel was cooled to 0 °C before the addition of $LiAlH_4$ (387.4 mg, 10.2 mmol, 1.3 equiv.). After stirring for 1 h at 0 °C, the

reaction was quenched by the slow, dropwise addition of an aqueous solution of 1.5 M NaHSO$_4$ (25 mL) and saturated Rochelle's salt (25 mL). The aqueous phase was extracted with EtOAc (3x50 mL) and the combined organic extracts were washed with saturated NaHCO$_3$ (50 mL), water (50 mL), and brine (50 mL). The organic phase was dried (MgSO$_4$), filtered, and concentrated by rotary evaporation to afford aldehyde S2 (1.56 g, quant.) as a clear liquid that was carried forward without additional purification. TLC: R$_f$ 0.55 (1:1 hexanes/EtOAc). ^{1}H-NMR (400 MHz, CDCl$_3$): δ 9.56 (s, 0.5 H), 9.45 (s, 0.5 H), 4.10–4.00 (m, 1H), 3.61–3.49 (m, 1H), 3.45–3.31 (m, 1H), 2.19–2.05 (m, 1H), 2.05–1.93 (m, 1H), 1.91–1.79 (m, 2H), 1.47 (s, 5H), 1.42 (s, 4H).ESI-MS m/z (rel int): (pos) 254.07 ([M+MeOH+Na]$^+$, 100); 222.04 ([M+Na]$^+$, 25). HR-ESI-MS *m/z* calcd. [C$_{10}$H$_{17}$NO$_3$Na]$^+$: 222.1101, found 222.1102.

tert-butyl (*S,E*)-2-(2-(ethoxysulfonyl)vinyl)pyrrolidine-1-carboxylate (S4).

In a 50 mL pear-shaped flask, ethyl(diethoxyphosphoryl)methane sulfonate S3[29] (2.12 g, 8.13 mmol, 1.2 equiv.) and 16 mL THF were added. The vessel was cooled to -78 °C before the dropwise addition of *n*BuLi (5.5 mL of 1.6 M solution in hexanes, 8.81 mmol. 1.3 equiv.). Stirring was continued for 1 h at -78 °C. In a separate 100 mL round-bottom flask, aldehyde S2 (1.36 g, 6.77 mmol, 1.0 equiv.) and 45 mL THF were added and the flask was cooled to -78 °C for 1 h. The phosphonate carbanion solution was cannulated into the solution containing S2. After stirring for 4 h at -78 °C, the reaction was quenched by the addition of H$_2$O (50 mL) and the volatiles were removed by rotary evarporation. The resulting aqueous phase was extracted with CH$_2$Cl$_2$ (3x50 mL) and the combined organic extracts were dried (MgSO$_4$), filtered, and concentrated by rotary

evaporation. Purification by silica flash chromatography (3:2 hexanes/diethyl ether) afforded ethyl vinyl sulfonate ester S4 (1.36 g, 55%) as a white solid.

TLC: R_f 0.54 (1:1 hexanes/EtOAc). ^{1}H-NMR (400 MHz, CDCl$_3$): δ 6.75 (dd, J = 15.0, 5.7 Hz, 1H), 6.18 (d, J = 15.0 Hz, 1H), 4.44 (d, J = 22.2 Hz, 1H), 4.21–4.09 (m, 2H), 3.48–3.33 (m, 2H), 1.94–1.74 (m, 4H), 1.41 (s, 5H), 1.36 (bs, 3H). ^{13}C-NMR (100 MHz, CDCl$_3$): δ 148.57, 124.30, 57.26, 46.77, 31.58, 30.65, 28.43, 23.79, 22.95, 14.86. ESI-MS m/z (rel int): (pos) 327.99([M+Na]$^+$, 100); 322.91 ([M+NH$_4$]$^+$, 30). HR-ESI-MS m/z calcd. [C$_{13}$H$_{23}$NO$_5$SNa]$^+$: 328.1189, found 328.1187; [C$_{13}$H$_{23}$NO$_5$SNH$_4$]$^+$: 323.1635, found 323.1634.

tert-butyl (*S,E*)-2-(2-(chlorosulfonyl)vinyl)pyrrolidine-1-carboxylate (S6).

In a 100 mL round-bottom flask, ethyl vinyl sulfonate ester S4 (255.6 mg, 0.8370 mmol, 1.0 equiv) and tetrabutylammonium iodide (410.1 mg, 1.1103 mmol, 1.3 equiv.) were dissolved in 17 mL acetone. The mixture was refluxed for 40 h. The volatiles were removed by rotary evaporation and the residue was dissolved in CH$_2$Cl$_2$ (50 mL) and washed with H$_2$O (2x10 mL). The organic extracts were dried (MgSO$_4$), filtered, and concentrated by rotary evaporation to afford crude vinyl sulfonate tetrabutylammonium salt S5, which was carried forward without additional purification.

In a 25 mL pear-shaped flask, triphenylphosphine (443.6 mg, 1.6913 mmol, 2.0 equiv.) was dissolved in 5 mL CH$_2$Cl$_2$ and cooled to 0 °C. To the flask was added sulfuryl chloride (0.15 mL, 1.8504 mmol, 2.2 equiv.), followed by a solution of crude vinyl sulfonate tetrabutylammonium salt S5 in 5 mL CH$_2$Cl$_2$. The mixture was stirred at 0 °C for 3 h, then

the volatiles were removed by rotary evaporation. Purification by silica flash chromatography (9:1 → 1:1 hexanes/EtOAc) afforded vinyl sulfonyl chloride S6 (98.8 mg, 40%, 2 steps) as a brown solid.

TLC: R_f 0.29 (3:1 hexanes/EtOAc). ^{1}H-NMR (400 MHz, CDCl$_3$): δ 6.99–6.89 (m, 1H), 6.71 (d, J = 13.9 Hz, 1H), 4.47 (bs, 1H), 3.47 (bs, 2H), 2.20 (bs, 1H), 1.99–1.77 (m, 4H), 1.44 (s, 9H). ^{13}C-NMR (100 MHz, CDCl$_3$): δ 149.45, 133.18, 57.00, 46.72, 31.57, 28.44, 23.15. HR-ESI-MS m/z calcd. [C$_{11}$H$_{18}$ ClNO$_4$SNa]$^+$: 318.0537, found 318.0538.

tert-butyl (S)-2-((E)-2-(N-(((3aR,4R,6R,6aR)-6-(6-((tert-butoxycarbonyl)amino)-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)methyl)sulfamoyl)vinyl)pyrrolidine-1-carboxylate (Boc-Pro-AVSN[N^6-Boc-2′,3′-isopropylidene adenosine]; S8).

In a 10 mL pear-shaped flask, N^6-Boc-2′,3′-isopropylidene 5′-aminodeoxyladenosine (tert-butyl (9-((3aR,4R,6R,6aR)-6-(aminomethyl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)-9H-purin-6-yl)carbamate) S7 (73.2 mg, 0.1801 mmol, 1.0 equiv.) and 2,6,-lutidine (60 µL, 0.5180 mmol, 2.9 equiv.) were dissolved in 2 mL CH$_2$Cl$_2$.[16] To the clear solution was added vinyl sulfonyl chloride S6 (53.7 mg, 0.1816 mmol, 1.0 equiv.) as a solution in 2 mL CH$_2$Cl$_2$. After stirring for 3 h at r.t. (21 °C), the volatiles were removed by rotary evaporation. Purification by silica flash chromatography (1:1 hexanes/EtOAc → EtOAc) afforded protected vinyl sulfonamide Boc-Pro-AVSN [N^6-Boc-2′,3′-isopropylidene] S8 (20.7 mg, 17%) as a white solid.

TLC: R_f 0.39 (EtOAc). ^{1}H-NMR (400 MHz, CDCl$_3$): δ 8.82 (s, 1H), 7.99 (s, 1H), 6.59 (dt, J = 14.5, 8.3 Hz, 1H), 6.17 (t, J = 13.6 Hz, 1H), 5.84 (d, J = 4.8 Hz, 1H), 5.33–5.29 (m,

1H), 5.17–5.09 (m, 1H), 4.56 (d, J = 20.1 Hz, 1H), 3.50–3.28 (m, 4H), 2.17–2.05 (m, 1H), 1.91–1.82 (m, 4H), 1.80–1.74 (m, 2H), 1.62 (s, 3H), 1.57 (s, 7H), 1.39 (s, 4H), 1.36 (s, 3H). ^{13}C-NMR (100 MHz, CDCl$_3$): δ 153.12, 144.83, 142.00, 127.73, 123.06, 93.54, 83.14, 82.49, 81.69, 31.72, 31.01, 28.47, 28.23, 27.59, 25.33. HR-ESI-MS *m/z* calcd. [C$_{29}$H$_{44}$N$_7$O$_9$S]$^+$: 666.2916, found 666.2908.

(*E*)-*N*-(((2*R*,3*S*,4*R*,5*R*)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)-2-((*S*)-pyrrolidin-2-yl)ethene-1-sulfonamide (Pro-AVSN).

In a 10 mL conical flask, protected Pro-AVSN S8 (9.2 mg, 0.0138 mmol, 1.0 equiv.) was dissolved in 1 mL H$_2$O and cooled to 0 °C. TFA (1 mL) was added and stirring continued at 0 °C for 3 h. The vessel was warmed to 21 °C and stirring was continued for 5 h. TFA and water were removed by azeotroping the mixture in cyclohexane (3x4 mL), MeOH (3x4 mL), and benzene (3x4 mL). Purification by semi-preparative HPLC (C$_{18}$; 10 x 250 mm; gradient 5-95% CH$_3$CN in H$_2$O with 0.05% TFA over 30 min; 5 mL/min) afforded Pro-AVSN (5.9 mg, quant.) as a white solid.

^{1}H-NMR (400 MHz, CD$_3$OD): δ 8.39 (s, 1H), 8.33 (s, 1H), 6.86–6.70 (m, 2H), 5.99 (d, J = 5.8 Hz, 1H), 4.76 (t, 2H), 4.35 (t, 1H), 4.29 (q, J = 15.8, 7.2 Hz, 1H), 4.25–4.20 (m, 1H), 3.43–3.35 (m, 3H), 2.38–2.27 (m, 1H), 2.20–2.02 (m, 2H), 1.95–1.83 (m, 1H).

2.7.B. Protein expression and preparation of *holo*-PltL and PltF

Expression and purification of *holo*-PltL

The pET22b-PltL plasmid and the pREP4-sfp plasmid were obtained from the laboratory of Christopher Walsh at Harvard University. Both plasmids were co-

transformed into *Escherichia coli* BL21 cells. The BL21 cells were grown in Luria-Bertani (LB) media with 50 µg/mL kanamycin and 100 µg/mL ampicillin. Expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.6. The cells were grown for at 25 °C for 16 hours, then harvested at 2500 relative centrifugal force (RCF). The pelleted cells were resuspended in 30 mL of 150 mM NaCl, 50 mM Tris pH7.5, 10% glycerol, 0.25 mg/mL lysozyme. The cells were further lysed with the Ultrasonic Processor FS-600N. The lysate was centrifuged at 12000 RCF for 1 hour and the supernatant was passed through the Novagen Ni-NTA resin column. *holo*-PltL was eluted with 15 mL of 150 mM NaCl, 50 mM Tris pH 7.5, 10% glycerol, and 250 mM imidazole. The expressed sfp did not have a his-tag and was did not bind to the column well. The eluant was passed through a Superdex S75 16/600 pg size exclusion column on an AKTA Pure fast performance liquid chromatography (FPLC) with 20 mM Tris pH 7.5 and 20 mM NaCl. The fractions containing *holo*-PltL were collected and concentrated to a final concentration of 2.2 mg/mL.

Expression and purification of PltF

The pET37b-PltF plasmid was obtained from the laboratory of Christopher Walsh at Harvard University. The plasmid was transformed into *E. coli* BL21 cells. The cells were grown in LB media with 50 µg/mL kanamycin. Expression was induced with 0.5 mM IPTG at an OD_{600} of 0.6. The cells were grown at 16 °C for 16 hours then harvested at 2500 relative centrifugal force (RCF). The pelleted cells were resuspended in 30 mL of 150 mM NaCl, 50 mM Tris pH 7.5, 10% glycerol, 0.25 mg/mL lysozyme. The cells were further lysed with the Ultrasonic Processor FS-600N. The lysate was centrifuged at 12000 RCF

for 1 hour and the supernatant was passed through the Novagen Ni-NTA resin column. PltF was eluted with 15 mL of 150 mM NaCl, 50 mM Tris pH 7.5, 10% glycerol, and 250 mM imidazole. The eluant was passed through a Superdex S200 16/600 pg size exclusion column on an AKTA Pure FPLC with 20 mM Tris pH 7.5 and 20 mM NaCl. The fractions containing PltF were collected and concentrated to a final concentration of 30 mg/mL.

2.7.C. X-ray crystallography studies

Trapping of *holo*-PltL with PltF

In 30 uL, the trapping reaction between *holo*-PltL and PltF consisted of 30 mM Tris pH 7.5, 2.0 mM $MgCl_2$, 2.0 mM tris(2-carboxyethyl)phosphine (TCEP), 0.02% Triton X, 0.34 mM Pro-AVSN, 0.17 mM *holo*-PltL, and 0.17 mM PltF. The reaction was incubated for 24 hours at 4 °C prior to crystallization experiments.

Crystallization of the PltL-PltF complex

Initial crystallization conditions were identified by the Index high throughput sparse matrix screen from Hampton Research through hanging drop vapor diffusion. The initial condition included a 200 nL drop volume with a 1:1 ratio of 0.17 mM PltL-PltF (11 mg/mL) complex to 3.5 M sodium formate pH 7.0. Subsequent optimizations led to a crystal condition consisting of 400 nL drop volume with a 1:1 ratio of 0.17 mM PltL-PltF (11 mg/mL) complex to a solution of 3.18 M sodium formate and 0.1 M malonic acid pH 5.73. Single crystals were transferred into cryoprotectant consisting of 20% glycerol and 80% of the crystallization condition and immediately frozen in liquid nitrogen.

Data collection and structure determination

X-ray diffraction data were collected at the Advanced Light Source beam line 8.3.1 at the Lawrence Berkeley National Laboratory at a temperature of 110 K with a wavelength of 1.12 Å. The data was processed with XDS[30] and an initial structure was built using molecular replacement with MOLREP[31] using PDB ID 3E7W[19] as the search model for PltF. The structure was iteratively refined via automated refinement with REFMAC[32] and manual refinement with COOT.[33] Then, PDB ID 2N5H[8] was used as a search model for the second round of molecular replacement to identify PltL. Then the Ppant/Pro-AVSN ligand and waters were added. The coordinates and structure factor amplitudes were deposited in the Protein Data Bank with the accession code 6O6E.

2.7.D. Interface analysis

Protein Interfaces, Surfaces, and Assemblies (PISA)[34] from the European Bioinformatics Institute, in combination with manual identification, was used to identify the residues involved in the PCP-A domain interface. PISA was also used to calculate the protein-protein interface areas of each PCP-A domain interface and the *E. coli* AcpP-partner protein interface.

The NRPS PCP-A domain interfaces under comparison include the PltL-PltF complex in this study, the type II EntE-EntB complex (PDB ID 3RG2),[15] the type II didomain PA1221 (PDB ID 4DG9),[14] the type I LgrA (PDB ID 5ES8),[23] the type I EntF (PDB ID 5T3D).[24] The contribution of other proteins in the same complex for both type I systems were omitted in this analysis. The FAS AcpP-partner protein interfaces include

the AcpP-FabA (PDB ID 4KEH)[25] and AcpP-FabZ complexes (PDB ID 6N3P)[35] (Figure S11).

2.7.E. Alignment and superposition of PltL structures

The "align" command in PyMol was used for sequence alignment, superposition, and refinement for aligning the structures of PltL (Figure S8). All 20 lowest energy solution NMR structures of *holo*-PltL (PDB ID 2N5H) were aligned to the bound state of PltL from the PltL-PltF crystal structure.

2.7.F. Mutagenesis of PltF

Preparation of PltF mutants

The PltF point mutations were introduced via QuikChange PCR.[36] The primer sequences are listed in Figure S9. The PltF mutants were expressed and purified through the same methods as in B.2.

HPLC-based aminoacylation assay of PltF mutants

The aminoacylation reaction (30 uL) contained 50 mM Tris pH 7.5, 12.5 mM $MgCl_2$, 2.0 mM TCEP, 5 mM ATP, 5 mM L-Pro, 0.015 mM *holo*-PltL, and 0.27 uM PltF. Upon addition of ATP and L-Pro last, the reaction was mixed and incubated at 25 °C for 2 minutes. The reaction was quenched with 6 uL of formic acid and centrifuged for 10 minutes at 10,000 RCF. The supernatants were passed over the Ascentis Express Peptide ES-C18 column with solvent B for 2 minutes, then increasing 5-40% B over 8 minutes, and finally 40-44% over 10 minutes (Solvent A = H_2O, 0.05% trifluoroacetic acid

(TFA); solvent B = acetonitrile, 0.05% TFA). Analyses were conducted on an Agilent HP 1100 series HPLC. The protein elution was monitored by absorbance at 210 nm (A_{210}) (Figure S5). The percent of *prolyl-* and *holo*-PltL was calculated through the integration of the chromatographic peak areas and using the A_{210} of *prolyl*-PltL divided by the summation of A_{210} of *prolyl*-PltL with A_{210} of *holo*-PltL. Each assay was performed in triplicate (Figure S10).

2.8. Supplementary information

Table 2.S1: PltL-PltF Crystallography data. Numbers in parenthesis denotes highest resolution shell.

PDB Entry	6O6E
	Data Collection
Space group	P3$_2$21
a, b,c (Å)	170.6, 170.6, 64.9
α, β, γ (deg)	90.0, 90.0, 120.0
resolution (Å)	2.14 (2.14-2.19)
total no. of reflections	59600
CC 1/2	98.9 (10.3)
I/σ(I)	9.73 (0.31)
Completeness	99.2 (90.5)
Redundancy	10.03 (8.48)
	Refinement
Resolution (Å)	2.14 (2.14-2.19)
No. of reflections	55985
R$_{work}$/R$_{free}$	0.216/0.261
No. of atoms (non-hydrogen)	
Protein	4354
Solvent	104
Heteroatoms	111
Average B factor (Å^2)	64.0
Ramachandran analysis (%)	
Favored	94.2
Allowed	5.5
Outliers	0.3
Root-mean-square deviation	
Bond lengths (Å)	0.009
Bond angles (°)	1.688
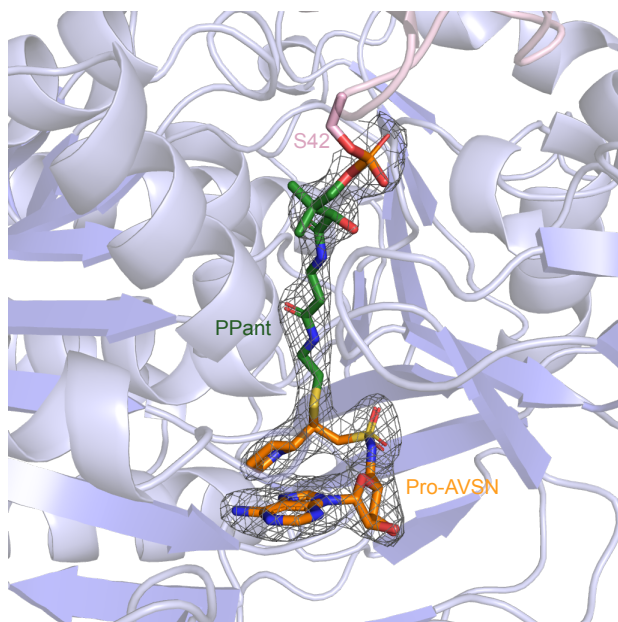
Figure 2.S2: 2fo-fc omit map of the phosphopantetheine (Ppant, green) and prolyl-adenosine vinylsulfonamide (Pro-AVSN, orange). Electron density is contoured to 2σ (gray mesh).
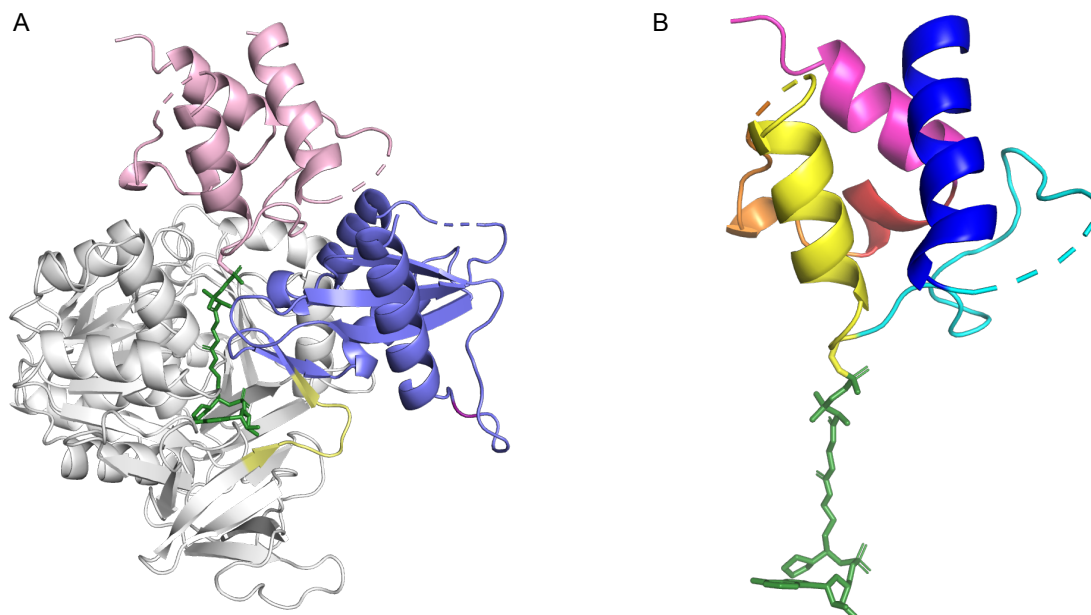
Figure 2.S3: Regions of PltL and PltF. A) PltF N-terminal domain (white) is connected to the C-terminal domain (blue) via the hinge region (yellow). The catalytic K486 responsible for the adenylation reaction is shown in purple. B) The secondary structure of PltL are as follows: α helix 1 (residues 1-19, blue), loop 1 (residues 20-41, cyan), α helix 2 (residues 42-55, yellow), loop 2 (residues 56-61, orange), α helix 3 (residues 62-65, red), α helix 4 (residues 66-90, pink), phosphopantetheine and proline adenosine vinylsulfonamide (green).
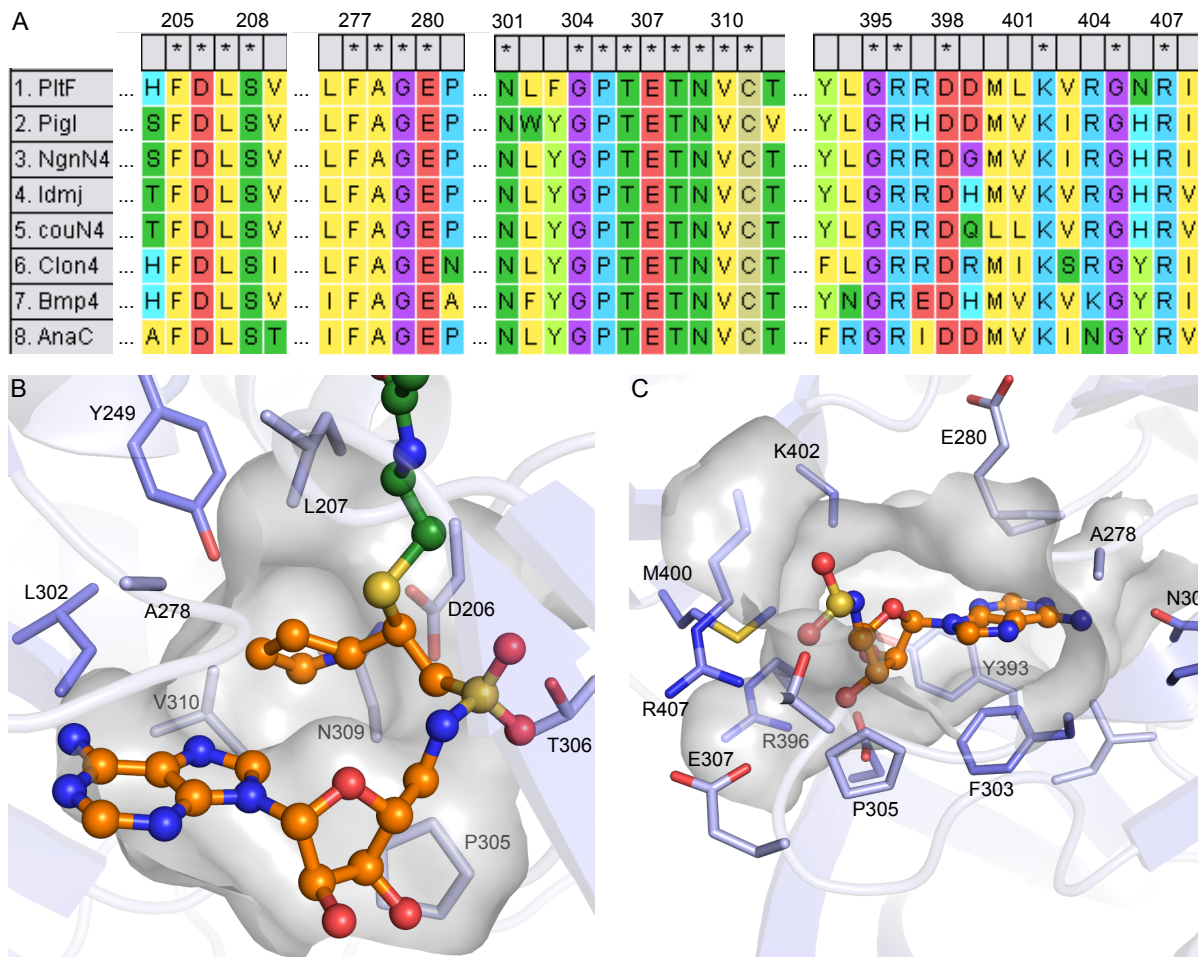
Figure 2.S4: Conservation of the PltF active site. A) Highly conserved regions in proline-loading A domains are shown with an asterisk (*). These regions correspond to the L-proline binding site and the AMP binding site. B) Binding pocket of L-proline in PltF. C) Binding pocket of AMP in PltF. The Pro and Ppant moieties were excluded for clarity. The residues directly involved in forming the binding pocket are also found in the conserved regions from panel A.
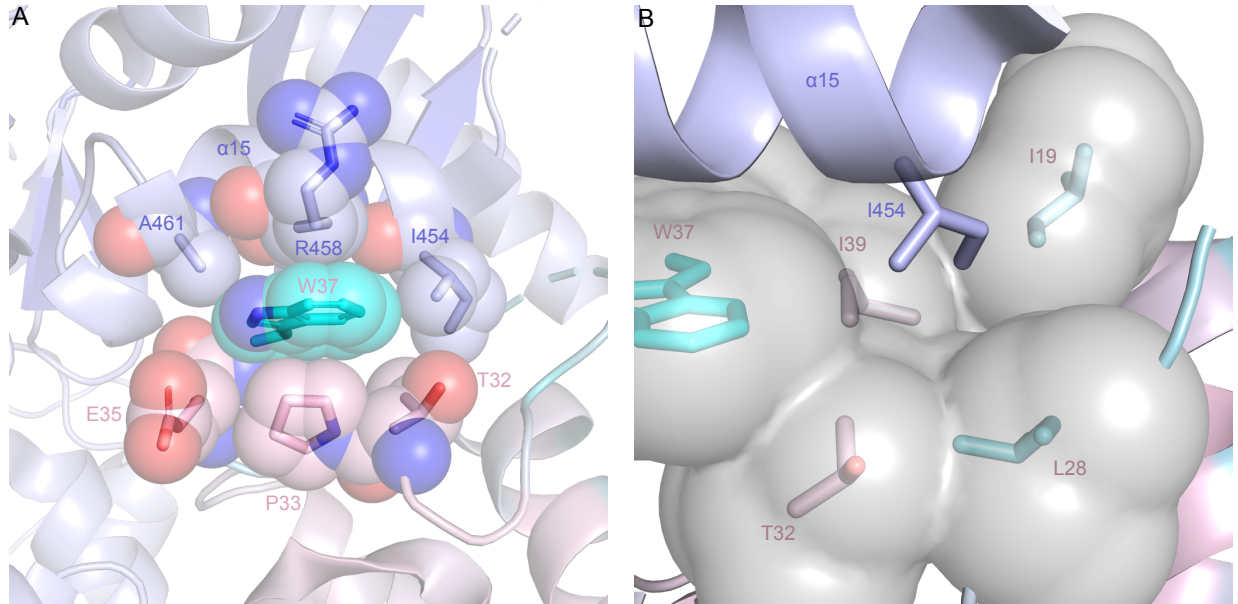
Figure 2.S5: Interface hydrophobic interactions involving PltL Trp37. A) Burial of the indole ring of PltL Trp37 between PltF helix 15 and PltL loop 1 residues. The spheres represent van der waals radius of each atom. B) Close up of the surface of the PltL hydrophobic pocket that is occupied by PltF Ile454.
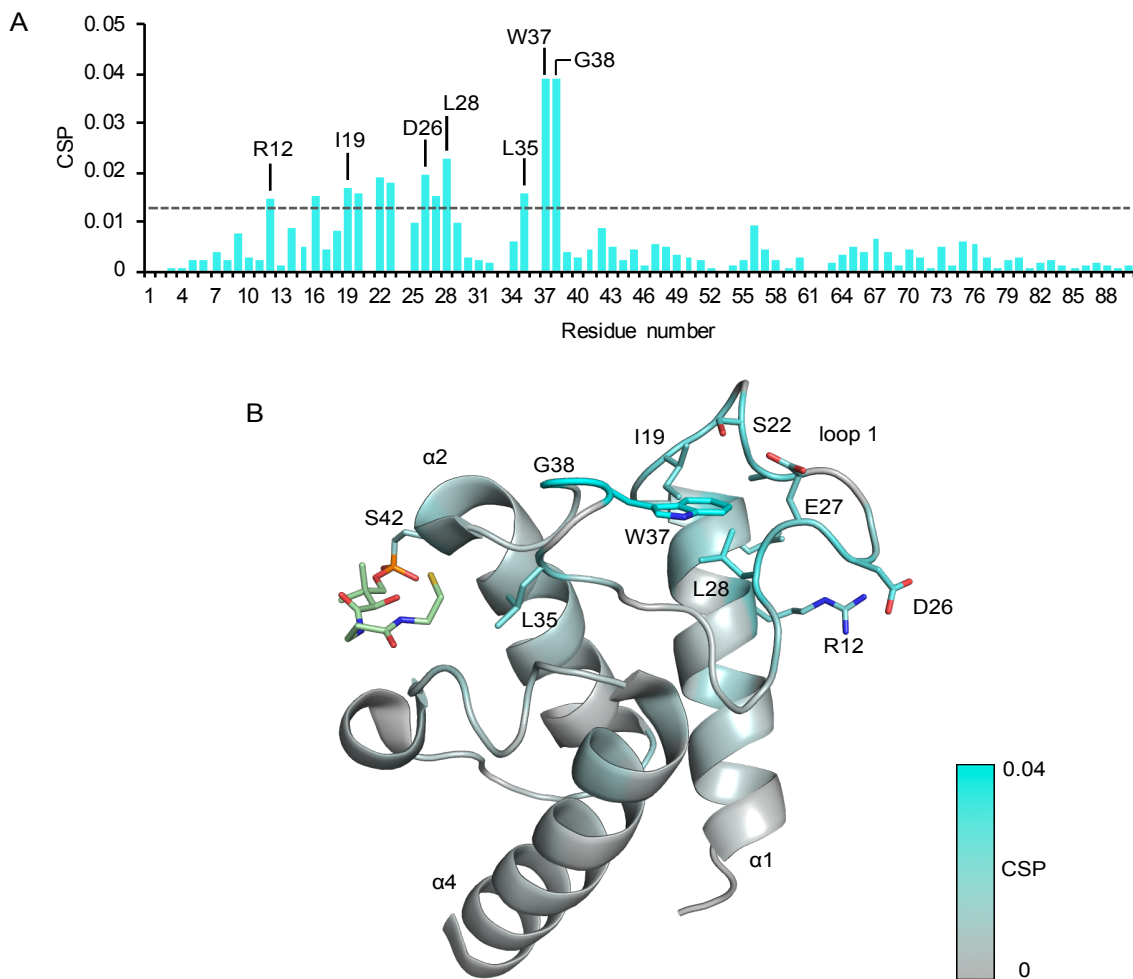
Figure 2.S6: NMR studies of PltL from Jaremko et. al. 2017. A) Plot of the chemical shift perturbations of an NMR titration with an ^{15}N-labeled PltL with increasing equivalents of PltF. The dotted line indicates one standard deviation above the mean. The corresponding residue to each perturbation is listed on the x-axis. B) Solution NMR structure of *holo*-PltL with the CSPs from A) mapped on the model. The Ppant arm was also ^{15}N labeled upon HSQC peak identification.
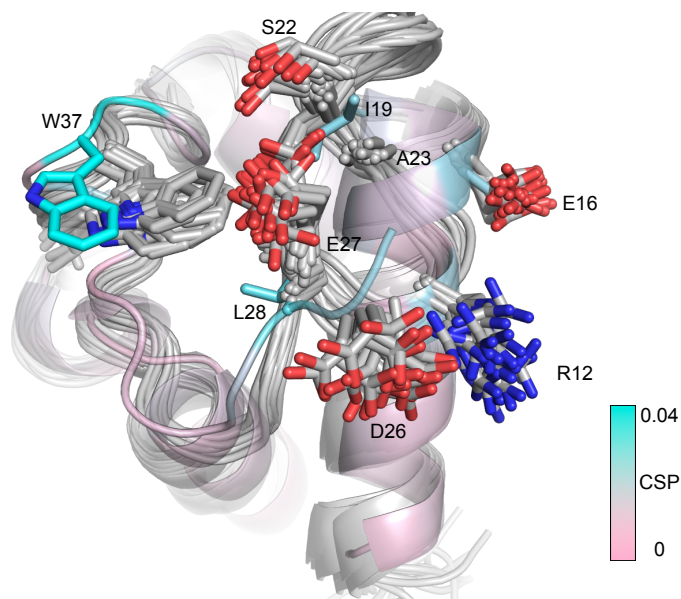
Figure 2.S7: Superposition of the bound state of PltL (this study) with the solution NMR structures of unbound PltL (PDB ID 2N5H). The 20 lowest energy models of the PltL solution NMR structure (gray) was aligned to PltL (pink) in the bound state. The residues with high CSPs are shown; a change in intraprotein loop and side chain interactions correlates with the high CSPs. This alignment also highlights the required dynamical movement of PltL W37 upon binding to PltF.

Table 2.S2: Primers for PltF alanine scanning.

PltF Mutants	Forward Primer	Reverse Primer
F231A	CGCTGCATCGCCGCGGCTGCTGAC	CGATGCAGCGATCGACTCGGGAACCAGG
S232A	GTTCGCACCGCGGCTGCTGACCGACTT	CGGTGCGAACGCGATCGACTCGGGAAC
I454A	GCTGGCAGACCTCAAGCGCCACTGCG	GTCTGCCAGCGAAAGAGCGTCCCGGGT
K457A	CTCGCACGCCACTGCGCCCAGCG	CTCGCACGCCACTGCGCCCAGCG
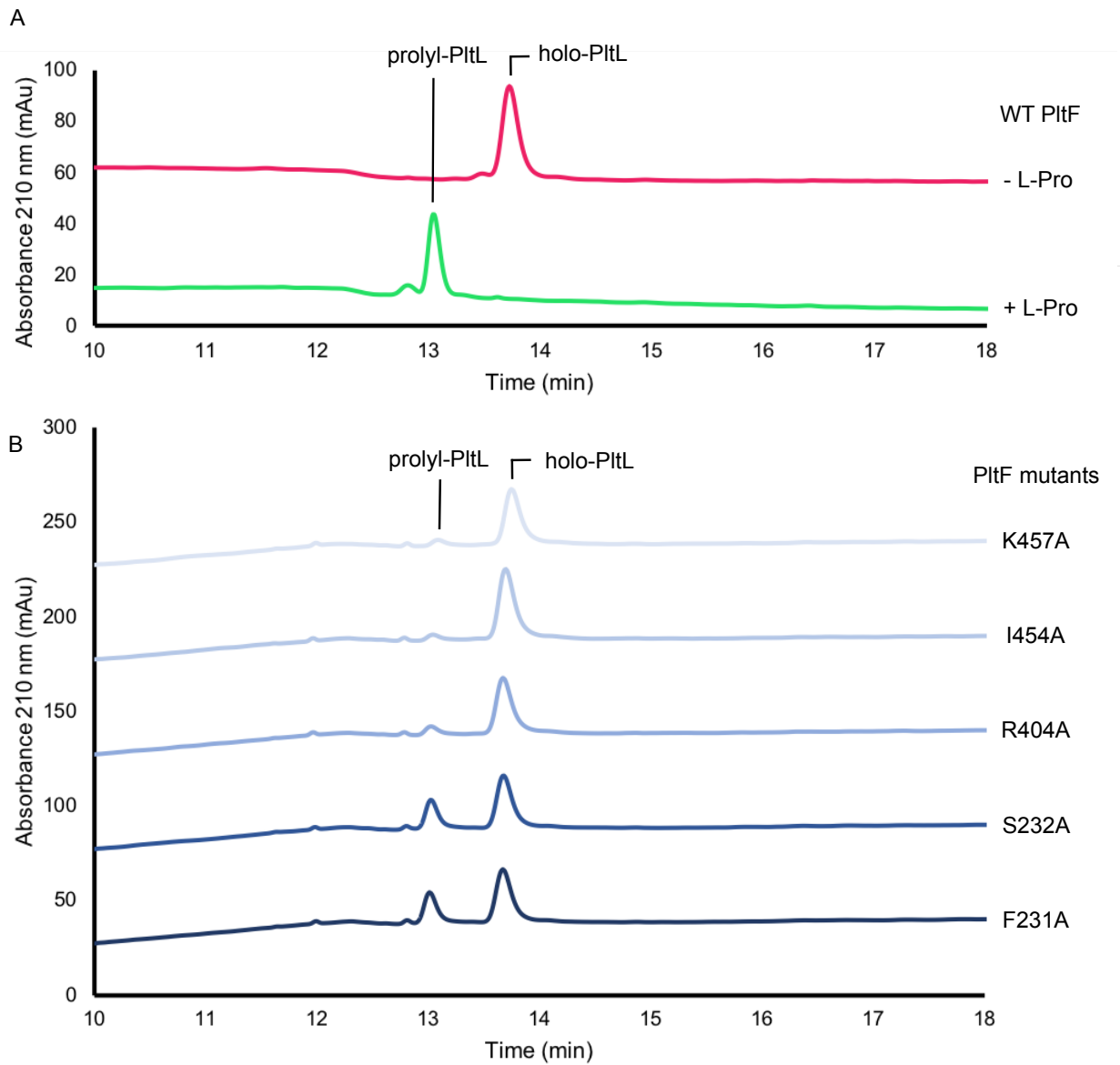
Figure 2.S 8: HPLC chromatograms of the aminoacylation of PltL by wild-type (WT) PltF and PltF mutants.
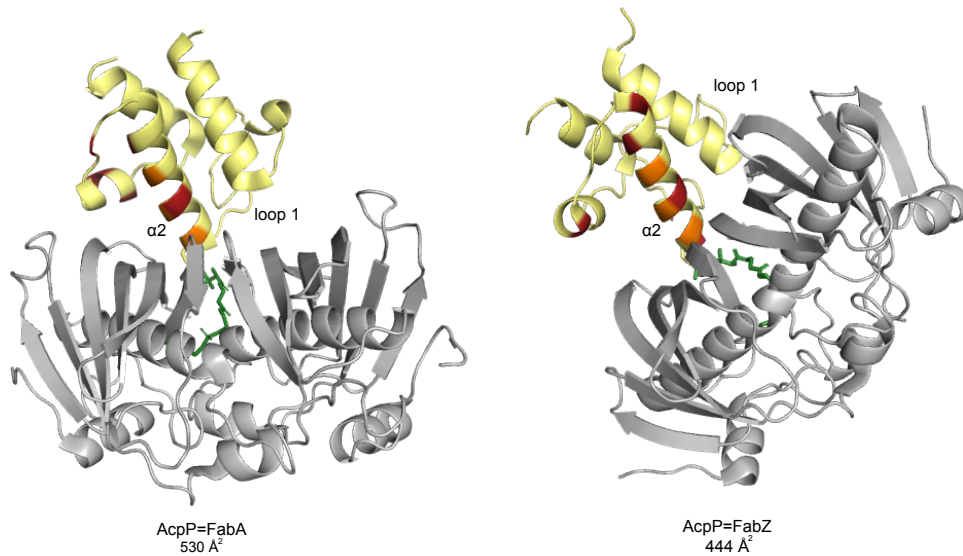
Figure 2.S 9: AcpP (yellow) interactions with its partner proteins (gray) in *E. coli* fatty acid biosynthesis. The protein-protein interface interactions of AcpP=FabA (left, PDB ID 4KEH) and of AcpP=FabZ (right, PDB ID 6N3P) consist of hydrophobic and salt bridge interactions and are almost exclusively located in alpha helix 2 of AcpP. The interface area was calculated, and protein-protein interactions identified via PISA. The interface areas of PCP-A domain interfaces are as follows: PltL-PltF 847 Å^2, EntE-EntB 928 Å^2, PA1221 675 Å^2, EntF 731 Å^2, LgrA 662 Å^2.

NMR characterization of S1

1H-NMR (400 MHz) of Weinreb amide S1 in CDCl3
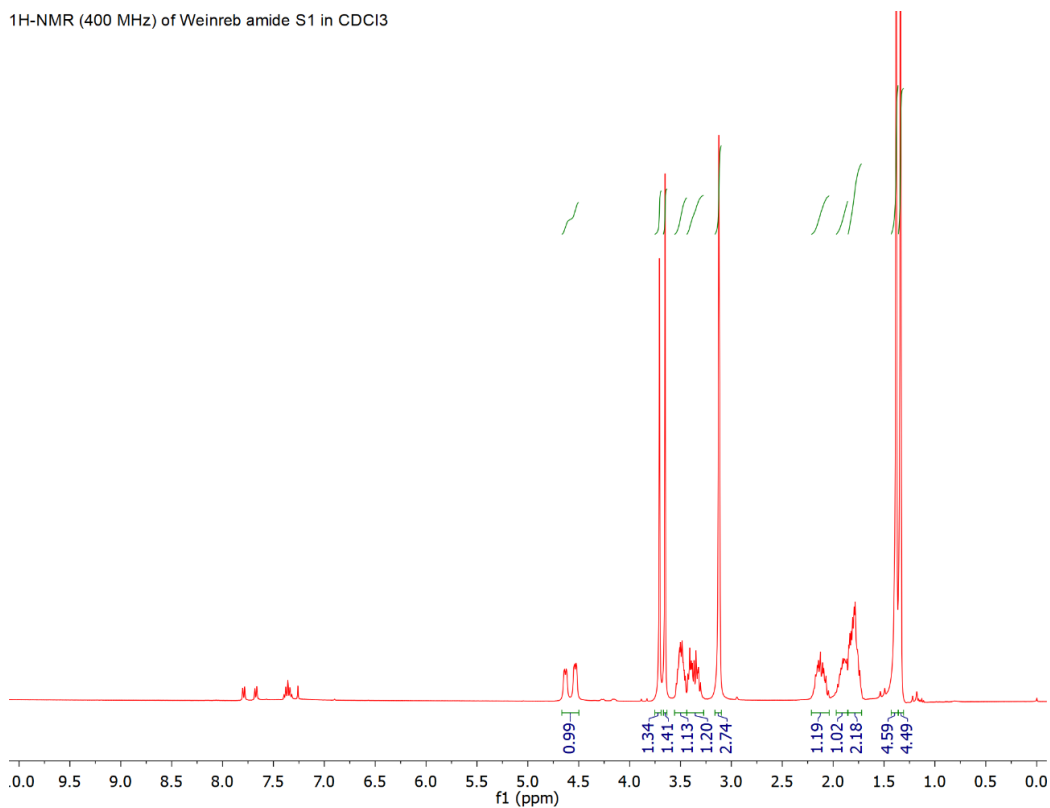
13C-NMR (100 MHz) spectra of Weinreb amide S1 in CDCl3
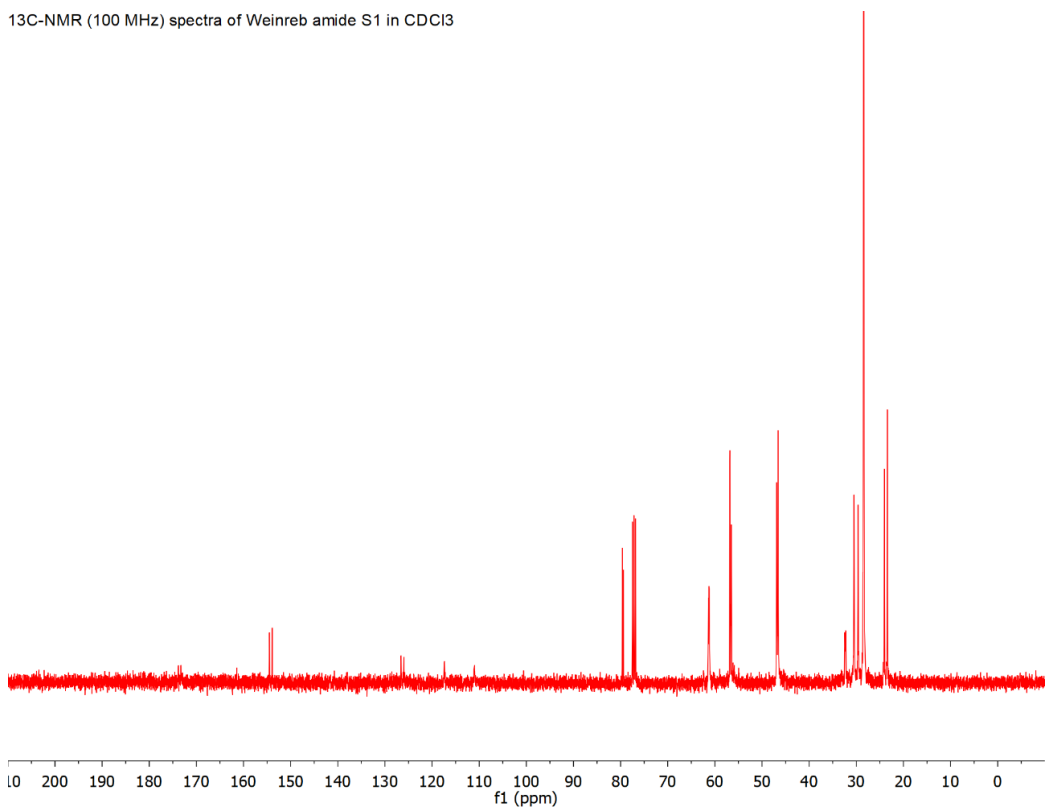
NMR characterization of S2

1H NMR (400 MHz) of aldehyde S2 in CDCl3
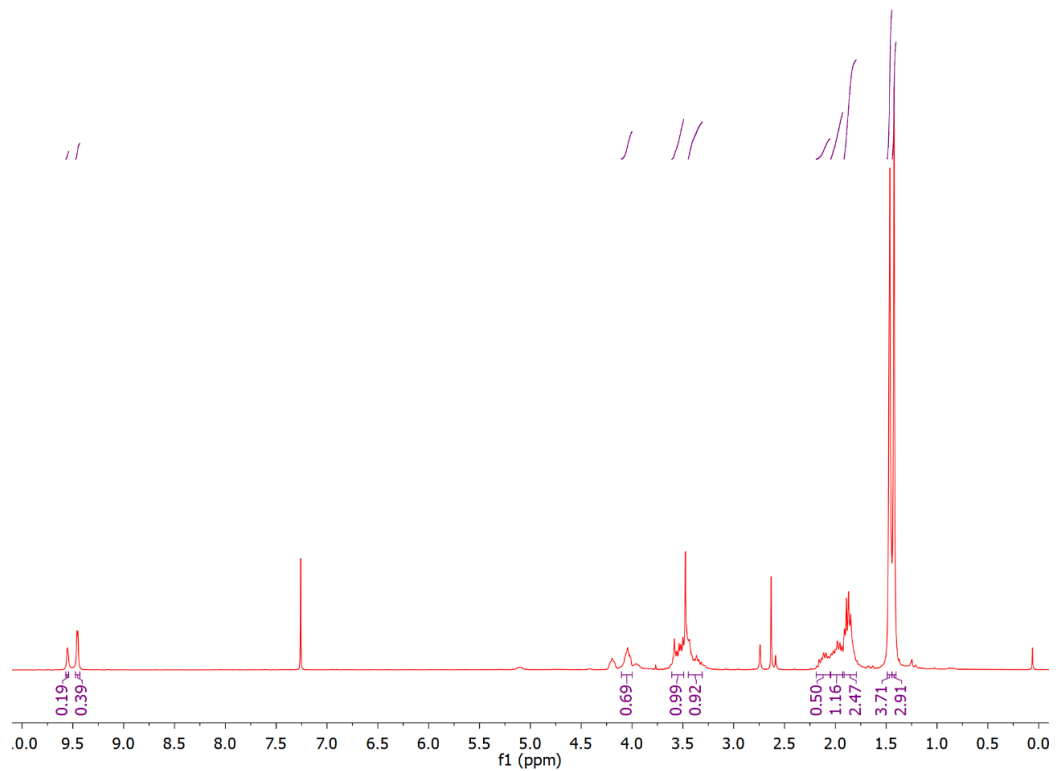
NMR characterization of S4

1H NMR (400 MHz) of ethyl vinyl sulfonate S4 in CDCl3
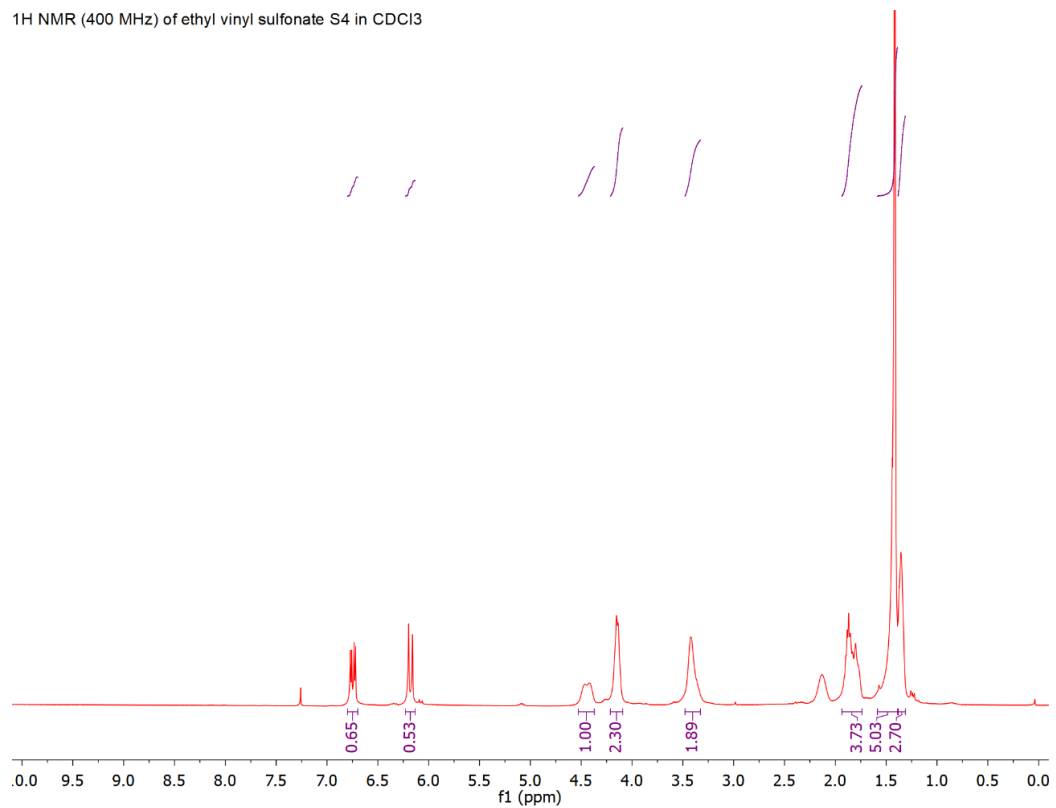
13C-NMR (100 MHz) spectra of ethyl vinyl sulfonate S4 in CDCl3
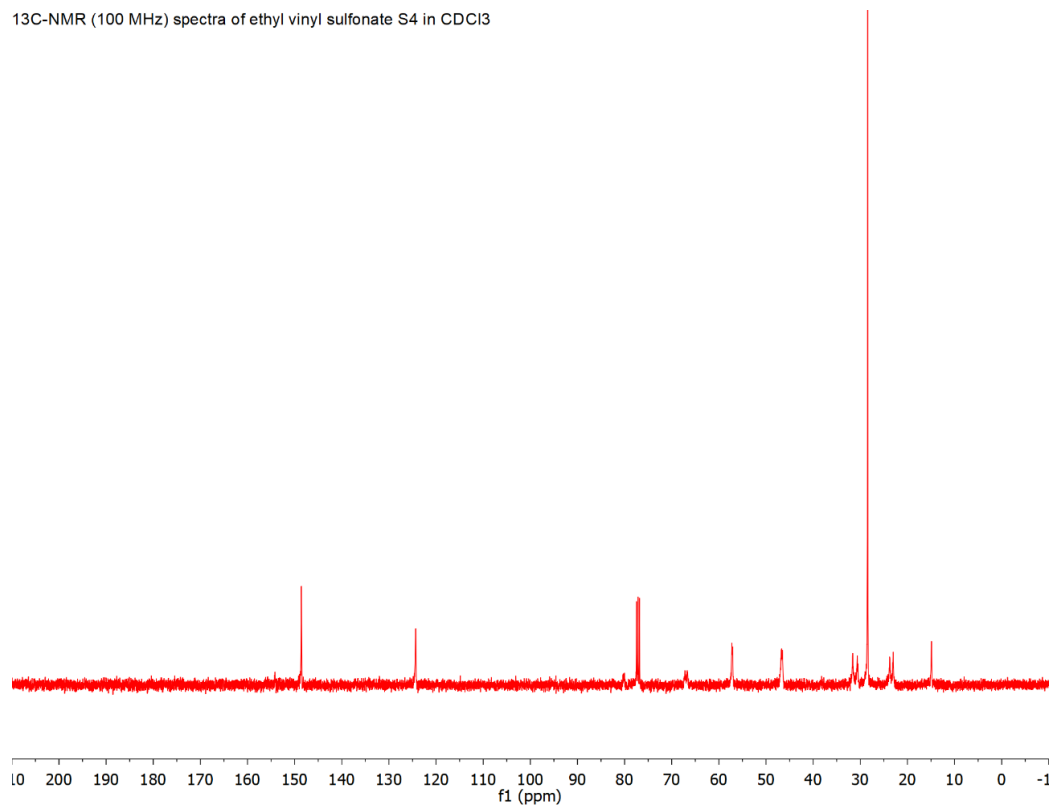
NMR characterization of S6

1H-NMR (400 MHz) of vinyl sulfonyl chloride S6 in CDCl3
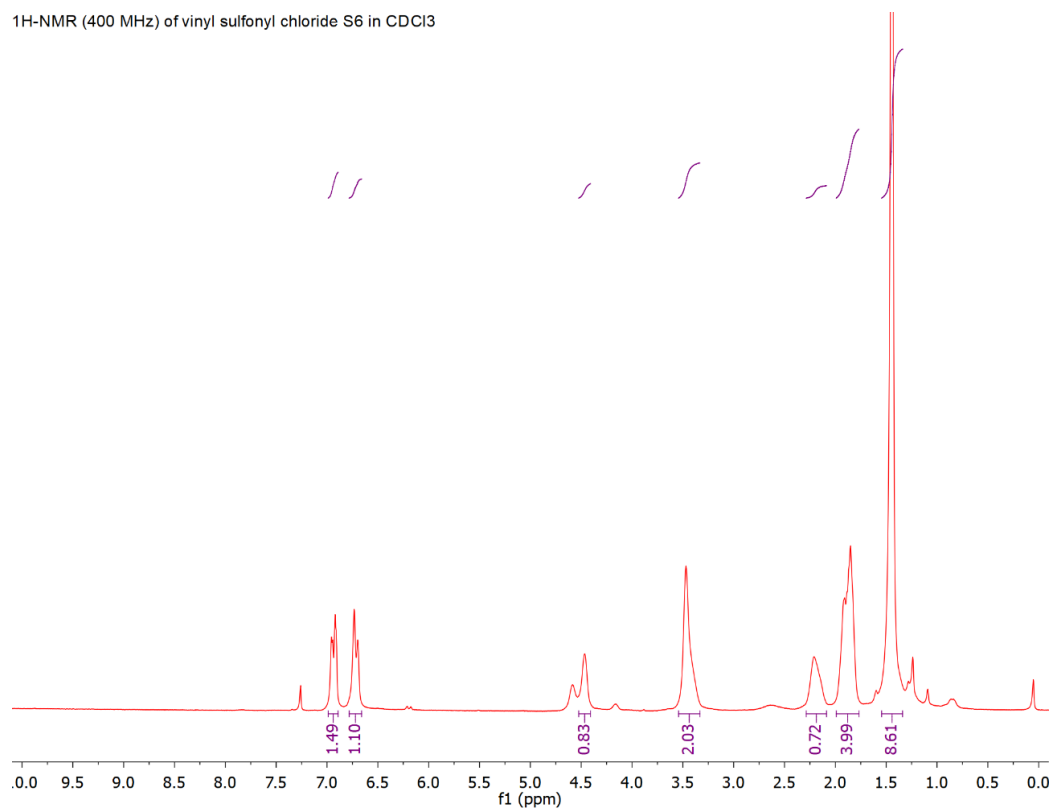
13C-NMR (100 MHz) spectra of vinyl sulfonyl chloride S6 in CDCl3
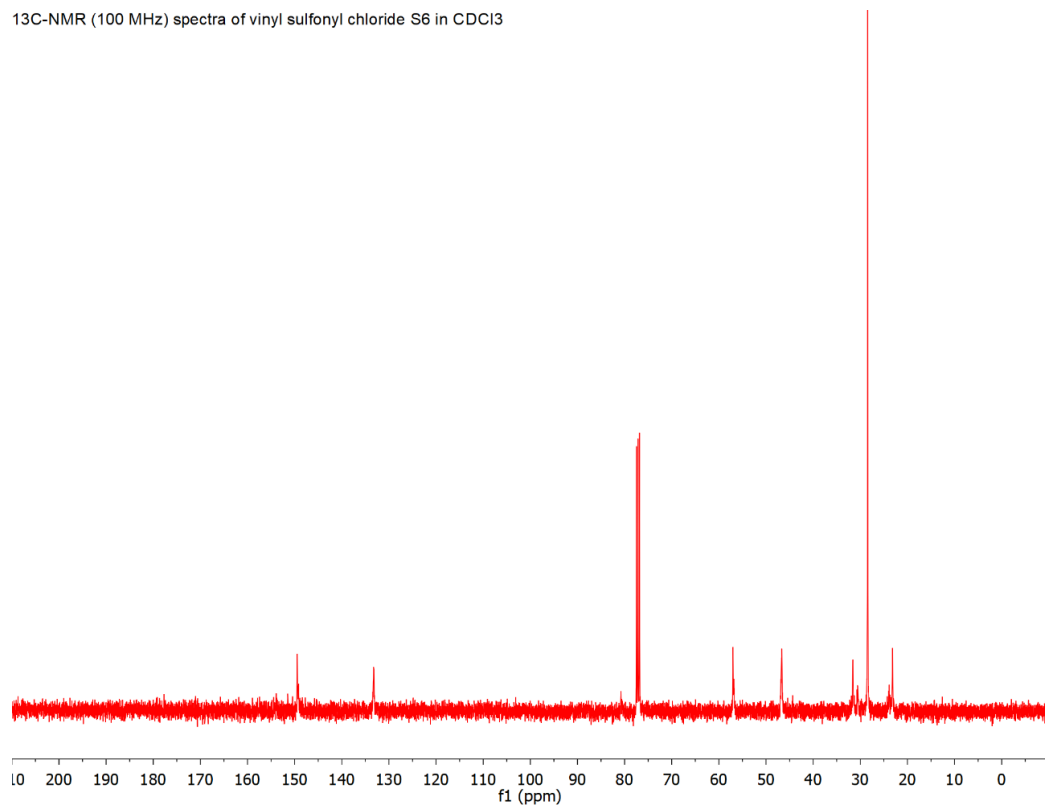
NMR characterization of S8

1H NMR (500 MHz) Boc-Pro-AVSN S8 in CDCl3
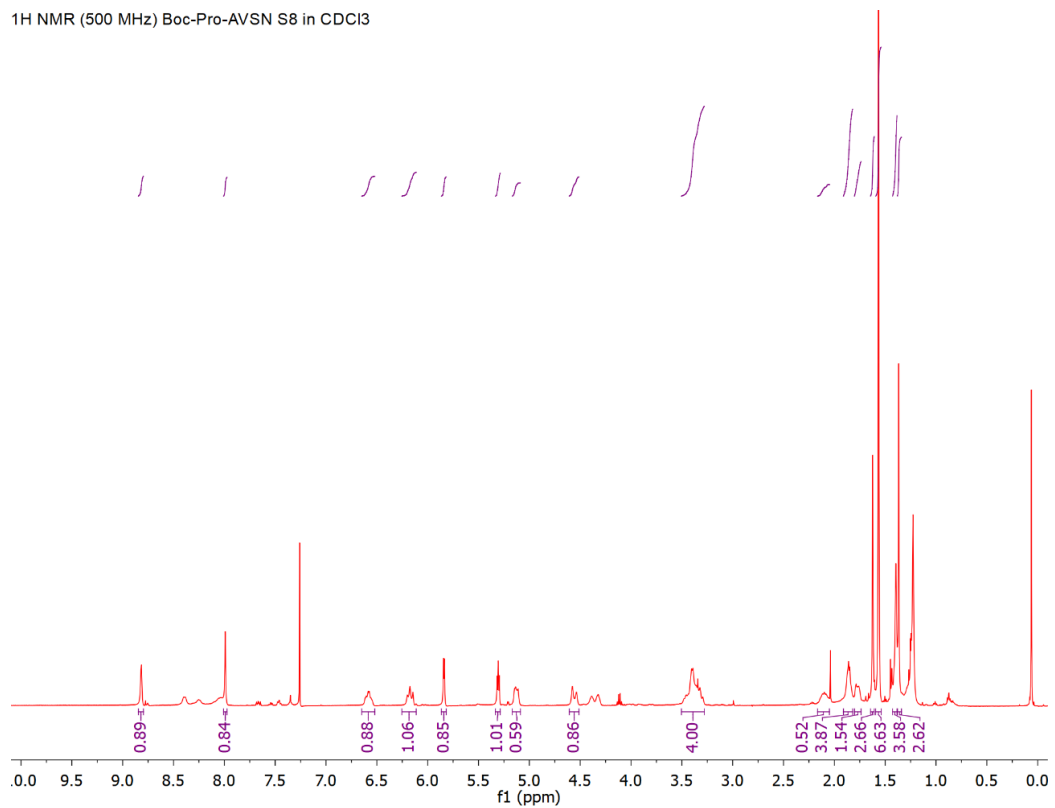
13C-NMR (100 MHz) spectra of Boc-L-Pro-AVSN S8 in CDCl3
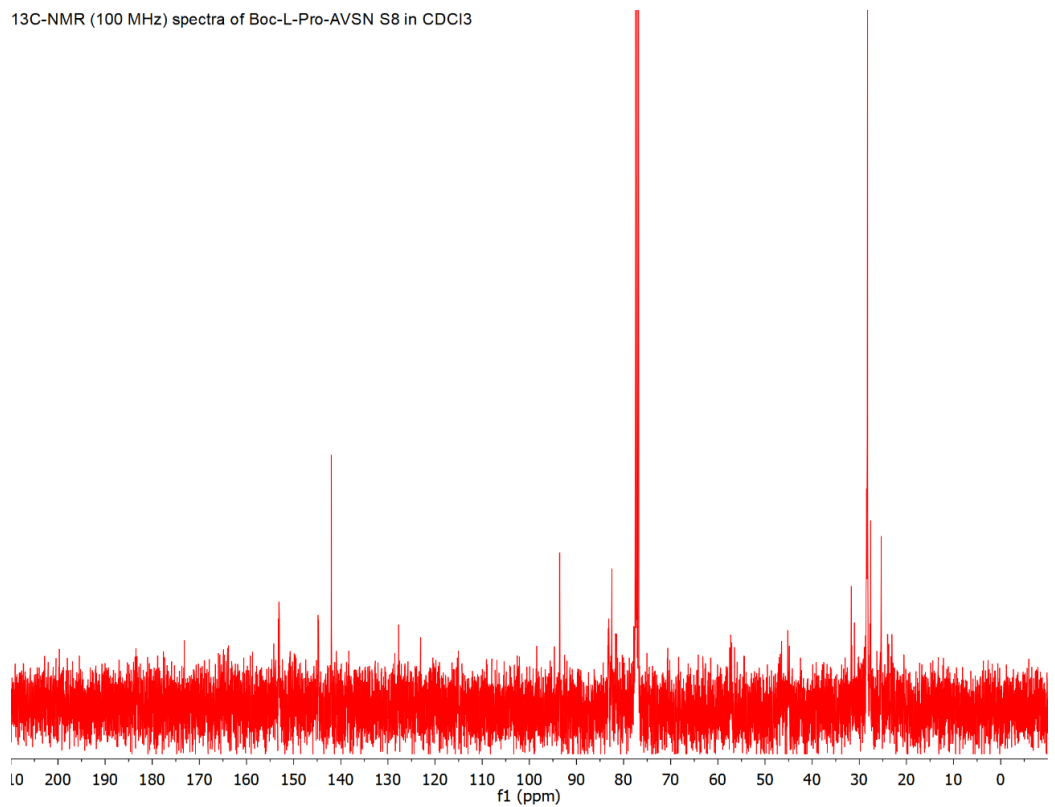
NMR characterization of Pro-AVSN
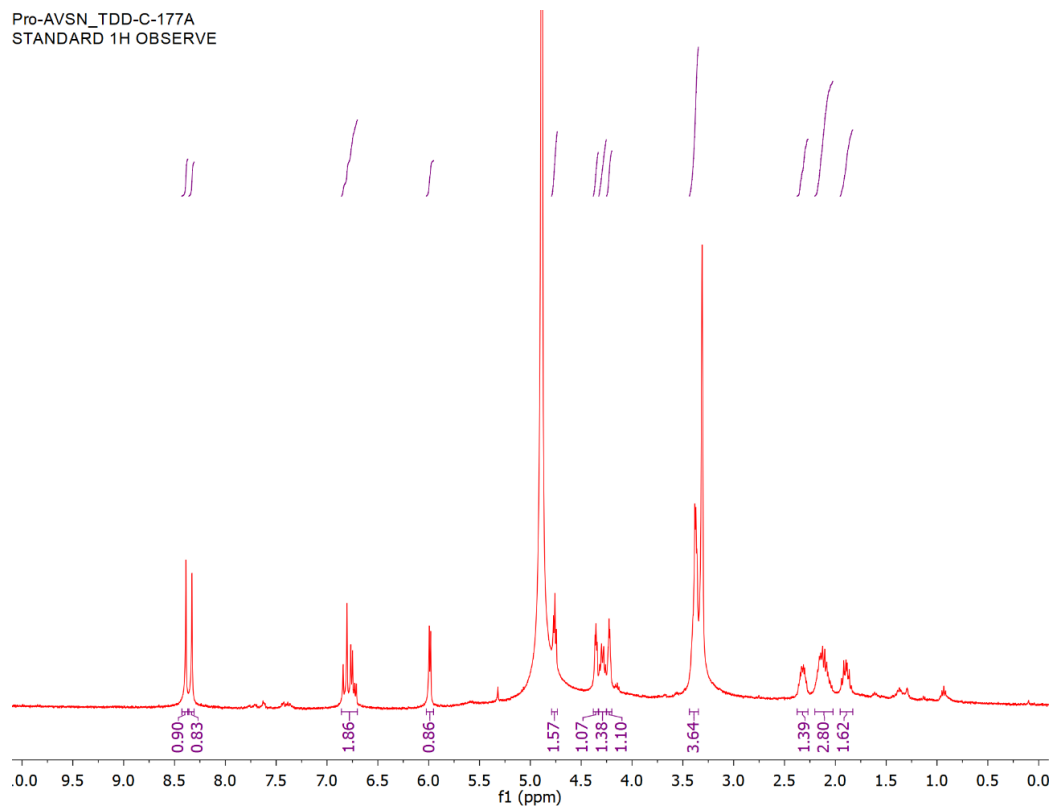
2.9. References

(1) Jaremko, M. J.; Davis, T. D.; Corpuz, J. C.; Burkart, M. D. Type II Non-Ribosomal Peptide Synthetase Proteins: Structure, Mechanism, and Protein–Protein Interactions. *Nat. Prod. Rep.* **2020**, *37* (3), 355–379.

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

(3) Williamson, N. R.; Simonsen, H. T.; Ahmed, R. A. A.; Goldet, G.; Slater, H.; Woodley, L.; Leeper, F. J.; Salmond, G. P. C. Biosynthesis of the red antibiotic, prodigiosin, in Serratia: identification of a novel 2-methyl-3-n-amyl-pyrrole (MAP) assembly pathway, definition of the terminal condensing enzyme, and implications for undecylprodigiosin biosynthesis in Streptomyces: The biosynthetic pathway of prodigiosin in Serratia. *Mol Microbiol*. **2005**, *56* (4), 971–989.

(4) Thomas, M. G.; Burkart, M. D.; Walsh, C. T. Conversion of L-Proline to Pyrrolyl-2-Carboxyl-S-PCP during Undecylprodigiosin and Pyoluteorin Biosynthesis. *Chem Biol*. **2002**, *9* (2), 171–184.

(5). Álvarez-Micó, X.; Rocha, D.D.; Guimarães, L. A.; Ambrose, A.; Chapman, E.; Costa-Lotufo, L. V.; LaClair J. J.; Fenical W. The Hybrid Pyrroloisoindolone-Dehydropyrrolizine Alkaloid (−)-Chlorizidine A Targets Proteins within the Glycolytic Pathway. *ChemBioChem*. **2015**, *16* (14), 2002–2006.

(6) Hur, G. H.; Vickery, C. R.; Burkart, M.D. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep*. **2012**, *29* (10),1074.

(7) Mercer, A. C.; Burkart, M. D. The ubiquitous carrier protein—a window to metabolite biosynthesis. *Nat Prod Rep*. **2007**, *24* (4), 750.

(8) Jaremko, M. J.; Lee, D. J.; Opella, S. J.; Burkart, M. D. Structure and Substrate Sequestration in the Pyoluteorin Type II Peptidyl Carrier Protein PltL. *J Am Chem Soc*. **2015**, *137*, (36), 11546–11549.

(9) Gross, H.; Loper, J. E. Genomics of secondary metabolite production by Pseudomonas spp. *Nat Prod Rep*. **2009**, *26* (11),1408.

(10) Jaremko, M. J.; Lee, D. J.; Patel, A.; Winslow, V.; Opella, S. J.; McCammon J. A.; Burkart, M. D. Manipulating Protein–Protein Interactions in Nonribosomal Peptide Synthetase Type II Peptidyl Carrier Proteins. *Biochemistry*. **2017**, *56*, (40), 5269–5273.

(11) Beld, J.; Finzel, K.; Burkart, M. D. Versatility of Acyl-Acyl Carrier Protein Synthetases. *Chem Biol*. **2014**, *21* (10), 1293–1299.

(12) Gulick, A. M.; Aldrich, C.C. Trapping interactions between catalytic domains and carrier proteins of modular biosynthetic enzymes with chemical probes. *Nat Prod Rep*. **2018**, *35* (11), 1156–1184.

(13) Qiao, C.; Wilson, D. J.; Bennett, E. M.; Aldrich, C. C. A Mechanism-Based Aryl Carrier Protein/Thiolation Domain Affinity Probe. *J Am Chem Soc*. **2007**, *129* (20),6350–6351.

(14) Mitchell, C. A.; Shi, C.; Aldrich, C. C.; Gulick, A. M. Structure of PA1221, a Nonribosomal Peptide Synthetase Containing Adenylation and Peptidyl Carrier Protein Domains. *Biochemistry*. **2012**, *51* (15), 3252–3263.

(15) Sundlov, J. A.; Shi, C.; Wilson, D.J.; Aldrich, C. C.; Gulick, A. M. Structural and Functional Investigation of the Intermolecular Interaction between NRPS Adenylation and Carrier Protein Domains. *Chem Biol*. **2012**, *19* (2),188–198.

(16) Lu, X.; Zhang, H.; Tonge, P. J.; Tan, D. S. Mechanism-based inhibitors of MenE, an acyl-CoA synthetase involved in bacterial menaquinone biosynthesis. *Bioorg Med Chem Lett*. **2008**, *18* (22), 5963–6.

(17) Lu, X.; Olsen, S. K.; Capili, A. D.; Cisar, J.S.; Lima, C. D.; Tan D. S. Designed Semisynthetic Protein Inhibitors of Ub/Ubl E1 Activating Enzymes. *J Am Chem Soc*. **2010**, *132* (6),1748–1749.

(18) Ji, C.; Sharma, I.; Pratihar, D.; Hudson, L.L.; Maura, D.; Guney, T.; Rahme, L. G.; Pesci, E. C.; Coleman, J. P.; Tan, D. S. Designed Small-Molecule Inhibitors of the Anthranilyl-CoA Synthetase PqsA Block Quinolone Biosynthesis in *Pseudomonas aeruginosa*. *ACS Chem Biol*. **2016**, *11* (11), 3061–3067.

(19) Yonus, H.; Neumann, P.; Zimmermann, S.; May, J.J.; Marahiel, M. A.; Stubbs, M. T. Crystal Structure of DltA: IMPLICATIONS FOR THE REACTION MECHANISM OF NON-RIBOSOMAL PEPTIDE SYNTHETASE ADENYLATION DOMAINS. *J Biol Chem*. **2008**, *283* (47), 32484–32491.

(20) Gulick, A. M. Conformational Dynamics in the Acyl-CoA Synthetases, Adenylation Domains of Non-ribosomal Peptide Synthetases, and Firefly Luciferase. *ACS Chem Biol*. **2009**, *4* (10),811–827.

(21) Schmelz, S.; Naismith, J.H. Adenylate-forming enzymes. *Curr Opin Struct Biol*. **2009**,*19* (6),666–671.

(22) Thapa, H. R.; Robbins, J. M.; Moore, B. S.; Agarwal, V. Insights into Thiotemplated Pyrrole Biosynthesis Gained from the Crystal Structure of Flavin-Dependent Oxidase in Complex with Carrier Protein. *Biochemistry*. **2019**,*58* (7),918–929.

(23) Reimer, J.M.; Aloise, M.N.; Harrison, P.M.; Schmeing, M. T. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature*. **2016**, *529* (7585),239–242.

(24) Drake, E.J.; Miller, B. R.; Shi, C.; Tarrasch, J. T.; Sundlov, J. A.; Allen C. L.; Skiniotis, G.; Aldrich, C. C.; Gulick, A. M. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature*. **2016**,. *529* (7585), 235–238.

(25) Nguyen, C.; Haushalter, R. W.; Lee, D. J.; Markwick, P. R. L.; Bruegger, J.; Caldara-Festin, G.; Finzel, K.; Jackson, D. R.; Ishikawa, F.; O'Dowd, Bing.; McCammon J. A.; Opella, S. J.; Tsai S.; Burkart, M. D. Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature*. **2014**, *505* (7483), 427–431.

(26) Izoré, T.; Cryle, M. J. The many faces and important roles of protein–protein interactions during non-ribosomal peptide synthesis. *Nat Prod Rep.* **2018**, *35* (11), 1120–1139.

(27) Crosby, J.; Crump, M. P. The structural role of the carrier protein – active controller or passive carrier. *Nat Prod Rep*. **2012**, *29* (10),1111.

(28) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *J Org Chem*. **1997**, *62* (21), 7512–7515.

(29) Carretero, J. C.; Oemillequand, M.; Ghosez, L. Synthesis of α, β-unsaturated sulphonates via the Wittig-Horner reaction. *Tetrahedron*. **1987**, *43* (21), 5125-5134.

(30) Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr*. **2010**, *66* (2),125–132.

(31) Vagin, A.; Teplyakov, A. MOLREP : an Automated Program for Molecular Replacement. *J Appl Crystallogr*. **1997**, *30* (6), 1022–1025.

(32) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr D Biol Crystallogr.* **1997**, *53* (3), 240–255.

(33) Emsley P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. **2004**, *60* (12), 2126–2132.

(34) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J Mol Biol*. **2007**, *372* (3),774–797.

(35) Dodge, G. J.; Patel, A.; Jaremko, K. L.; McCammon, J. A.; Smith, J. L.; Burkart, M. D. Structural and dynamical rationale for fatty acid unsaturation in *Escherichia coli*. *Proc Natl Acad Sci*. **2019**, *116* (14), 6775–6783.

(36) Liu H.; Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol*. **2008**, *8* (1),91.

CHAPTER 3. THEORY AND EXPERIMENT DEMONSTRATE THE ESSENTIAL ROLE OF LOOP DYNAMICS IN TYPE II NRPS BIOMOLECULAR RECOGNITION

3.1. Introduction

To date, nearly half of approved drugs are directly derived from or based upon natural products.[1] Natural product biosynthesis yields a vast library of structurally complex, highly functionalized, stereodense compounds. Among these compounds, non-ribosomal peptides (NRPs) have established a privileged status as drug-like molecules.[2,3] Consequently, considerable effort has been exhausted to engineer non-ribosomal peptide synthetases (NRPSs), with the expectation these efforts will not only yield novel therapeutics but also establish biosynthetic routes capable of producing novel metabolites efficiently enough to address human health concerns.[4] Though sound in concept, it has proven difficult in practice to engineer microbial metabolism to produce non-native products. These difficulties can largely be attributed to our limited understanding of the molecular basis of natural product biosynthesis; namely the sensitive and precise interplay of the key elements and proteins, that govern natural product biosynthesis.[5] The work reported herein represents the culmination of a longstanding effort to demonstrate that efforts to rationally engineer metabolic pathways to function in novel ways can prove fruitful.

Herein, we report work to clarify how type II NRPSs associate as required for the biosynthesis of proline-derived natural products. While naturally, these metabolites have been found to possess cellular roles of signaling molecules, pigments, and toxins, these NRPs also possess antitumor, antibacterial, and immunosuppressant therapeutic properties.[6] Clinically relevant examples of proline-derived natural products include

prodigiosin, pyoluteorin, and chlorozidine (Figure 3.1A).[6] The proline moieties present in these metabolites are incorporated into these natural products through the coordinated effort of multiple enzymes. Specifically, these pathways utilize type II NRPS enzymatic domains to prepare the pyrrolidine ring of these metabolites. Due to their ability to biosynthesize complex molecules with therapeutic bioactivities, these pathways have been targeted for engineering to create new natural products. Each of these proline-incorporating pathways employ similar type II NRPS proteins with the adenylation (A) domain acting as a gatekeeper that activates L-proline in an ATP dependent manner in order to transfer the activated proline to the phosphopantetheine (PPant) arm of the peptidyl carrier protein (PCP).[2,7] The PCP is a small ~10 kDa protein that is post-translationally phosphopantetheinylated, yielding a *holo*-PCP that, through the thiol terminus of its phosphopantetheine prosthetic group, can tether peptide substrates to the carrier protein via a thioester linkage.[7] Once loaded with proline, the PCP shuttles the prolyl scaffold to a dehydrogenation (DH) domain, where the DH domain catalyzes the FAD dependent reduction of the proline moiety in the biosyntheses of anatoxin, prodigiosin, pyoluteorin, and chlorizidine with the resulting species undergoing further chemical modification (Figure 3.1A).[6,8] In the type II NRPS system responsible for prodigiosin biosynthesis, the PCP, PigG, is loaded with a prolyl moiety by a cognate A domain, PigI. Similarly, the NRPS that produce pyoluteorin includes a PCP, PltL, that is converted to prolyl-PltL by the A domain, PltF. Prior studies of these systems have revealed the central role of proper protein-protein interactions in the substrate-charging of the PCP.[9] Further experimental work showed that both cognate (PigI) and noncognate (PltF) A domains can aminoacylate PigG, while PltL exclusively reacts with PltF despite
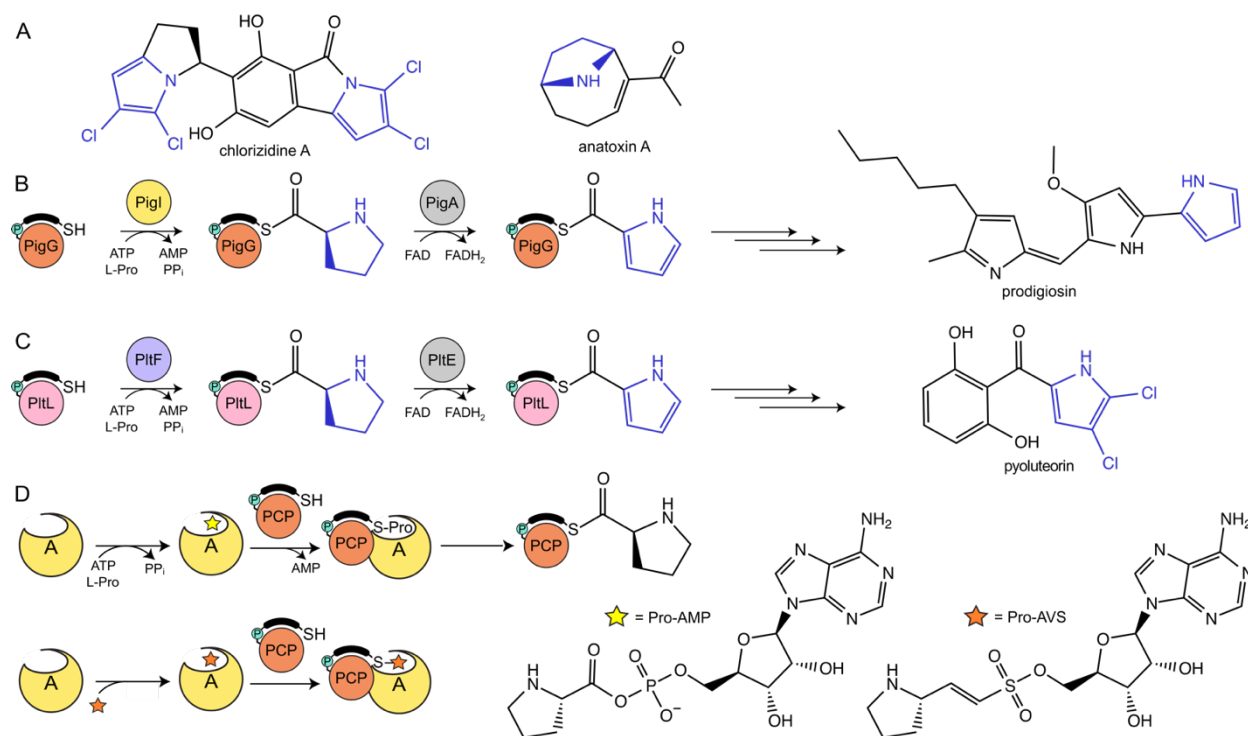
Figure 3.1: Prevalence and preparation of proline in NRP biosynthesis. A) Examples of NRPs that apply a functionalized proline ring. Proline is activated and dehydrogenated in B) prodigiosin and C) pyoluteorin biosynthesis. D) Adenylation of L-proline to form a proline-adenosine monophosphate intermediate (yellow star) and thiolation of the holo-PCP to form prolyl-PCP. Below, the PCP-A domain complex can also be trapped with a mechanism-based inhibitor, proline adenosine vinylsulfonamide (orange star).

homology and conserved PCP structures. Thus, understanding the precise levels of specificity between PCPs and A domains is crucial towards the design of new NRPS protein-protein interfaces for the development of new biosynthetic pathways.

This work builds upon previous PCP-A domain biophysical assays and complex crystal structures by the chemical trapping and X-ray crystal structure determination of the cognate PigG-PigI complex and the non-cognate PigG-PltF complex. The protein-protein interface in each PCP-A domain complex is supported through previous NMR titrations in addition to alanine scanning of the interface. The PltL-PltF, PigG-PigI, PigG-PltF, *holo*-PigG, and *holo*-PltL structures were subjected to molecular dynamics simulations, which revealed the difference in dynamism in the beginning of the PCP loop 1 region. In addition to molecular dynamics (MD) simulations, sequence alignments and structural superpositions guided mutagenesis of the PltL and PigI interface residues in the PCP loop 1 region, which enabled the enhancement of the non-cognate PltL-PigI activity by 20-fold.

3.2. Structural analysis of cognate and noncognate PCP-A domain complexes

In order to elucidate the molecular basis of PigG's promiscuity towards homologous A domains, we solved the X-ray crystal structures of the cognate PigG-PigI and the noncognate PigG-PltF complexes. A mechanism-based crosslinker[10] was used to trap *holo*-PigG in association with either PigI or PltF, which stabilized these complexes and facilitated their crystallization. The PigG-PigI crystal diffracted to 1.61 Å, while the PigG-PltF crystal diffracted to 2.46 Å. Initial phases for both data sets were solved using the adenylation domain, PltF, from a previously reported crosslinking PCP-A complex

(PDB ID: 6O6E)[10] as a model for molecular replacement. The NMR solution structure of PigG (PDB ID: 5JDX)[9] was then used to fit PigG into the remaining electron density. PigG and PigI crystallized as a single complex in the asymmetric unit, whereas two PigG-PltF complexes crystallized in a single asymmetric unit. The PigG-PigI and PigG-PltF complexes were refined to R_{free} of 0.193 and 0.266 respectively (Table 3.S1). The thiol of the PPant arm is observed covalently linked to the AVS crosslinker (Figure 3.S1).

The A domains, PigI and PltF, are comprised of a N-terminal subdomain (A_{core}) and the C-terminal subdomain (A_{sub}). In the crosslinked complexes, both A domains have been resolved in the thiolation conformation (Figure 3.2A), in which both subdomains are positioned to form a protein interface capable of binding a partner PCP.[5] The A_{core} of PigI and PltF are responsible for binding substrates ATP and L-proline, whereas the A_{sub} possess residues that are important for catalysis of both the adenylation and thiolation half-reactions. In both crosslinked structures, PigG maintains the conserved 4 α-helix bundle with the PPant attached to Ser36 of PigG and extended into the active site of the A domain, where it is covalently linked to the prolyl-adenosine vinylsulfonamide (Pro-AVSN, **1**).

The active site of the PigG-PigI and PigG-PltF structures resembles the active site of the previously reported crosslinked PltL-PltF complex (PDB ID: 6O6E).[10] Similar to what is observed in the PltL-PltF structure, the A_{sub} domain catalytic lysine responsible for adenylation in PltF and PigI, Lys486 and Lys477, respectively, are both ~25 Å away from the active site, in accordance with the proposed domain reorganization process that occurs between the catalysis of the adenylation and thiolation half-reactions.[11] The A_{core} domain of PigI and PltF provide a binding pocket for ATP and proline substrates that in
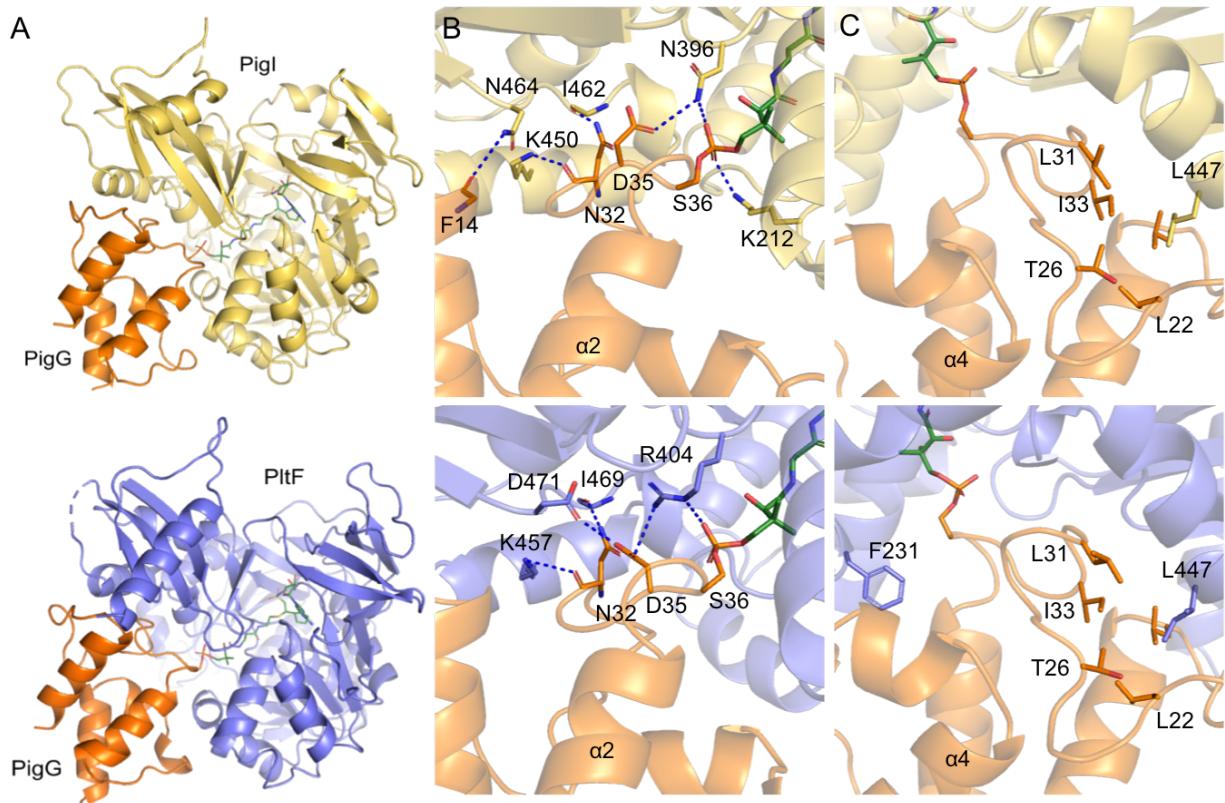
Figure 3.2: PCP-A domain X-ray crystal structure, interface analysis, and validation. A) Overall X-ray crystal structure of the PigG-PigI and PigG-PltF complex trapped with the Pro-AVS probe. Zoom in of the PCP-A domain protein-protein interface of B) hydrogen bonding interactions and C) hydrophobic interactions.

thiolation conformation stabilize a prolyl-adenylate or a proline AVSN adduct in the native and crosslinked structures, respectively. Analysis of the PigG-PigI, PigG-PltF, and the PltL-PltF structure indicates that Glu289 and Arg399 of PigI (Glu307 and Arg407 of PltF) form a salt bridge interaction in the proximity of the sulfonyl moiety of the crosslinking probe. In the native amino-adenylate bound structure, this salt bridge likely anchors the phosphodiester moiety of the intermediate and stabilizes the thiolation state.[12] Additional interactions that are observed between the prolyl moiety of the probe include a single interaction involving nitrogen of the pyrrolidine sidechain and Asp187 of PigI (Asp206 of PltF). Lastly, a Mg^{2+} ion required for catalysis is resolved in the 1.8 Å crosslinked PigG-PigI structure, in which the Mg^{2+} ion interacts with the sulfonyl group.

3.3. PCP-A domain protein-protein interface analysis

The protein-protein interface between PigG and both A domains is formed by PigG loop 1, which is a region of 20 residues connecting helices 1 and 2 (Figure 3.2), that mainly contacts the A_{sub} domain. The protein-protein interfaces of both PigG-PigI and PigG-PltF feature specific hydrogen-bonding and hydrophobic interactions. Loop 1 of PigG interacts with PigI utilizing a network of hydrogen bonds that include interactions between the sidechain carboxamide PigG's Asn32 and the backbone carbonyl of PigI's Ile462, the sidechain amine of PigI's Lys450 and the backbone carbonyls of PigG's Leu31 and Asn32, the amide of PigI's Asn464 and the backbone carbonyls of PigG's Phe14 and Ile33, and PigG's Asp35 sidechain carboxylate with the sidechain amide of PigI's Asn396 (Figure 3.2B). The protein-protein interface of PigG-PltF features a similar, but less extensive hydrogen-bonding network that includes hydrogen bonding of the sidechain

amide of PigG's Asn32 with the sidechain carboxyl of PltF's Asp471 and the backbone carbonyls of PltF's Ile469 and Ile470, the amine of PltF's Lys472 and the backbone carbonyl of PigG's Gln13 and Phe14. Arg404 of PltF forms a salt-bridge interaction with Asp35 of PigG (Figure 3.2B). Both the PigG-PigI and PigG-PltF interfaces share a common hydrophobic interaction, where an aliphatic residue, either Leu447 of PigI or Ile454 of PltF, sits inside a hydrophobic pocket formed by Leu15, Thr26, Ile31, Ile33, and Gly21 of PigG's loop 1 (Figure 3.2C). The PigG-PltF crystal structure shows an additional hydrophobic interaction involving the aromatic sidechain of Phe231 of PltF, which sits in between helix 1 and 3 of PigG (Figure 3.2C).

Alanine scanning mutagenesis of PigI and PltF residues that directly participate in molecular recognition demonstrates the essential role of specific interactions in mediating PCP-A domain complexation (Figure 3.S8). Disruption of specific hydrogen-bonding and hydrophobic interactions by mutagenesis results in stunted activity with PigG. Interestingly, the F231A mutant of PltF exhibited an increased rate of aminoacylation of PigG. Previous NMR titration experiments suggested that PigG binds more tightly to the noncognate PltF than its cognate partner PigI,[9] which may be attributed to hydrophobic interactions involving F231A. Furthermore, the same NMR titration studies reveal significant chemical shift perturbations (CSPs) in loop 1 residues of PigG upon binding to PigI and PltF, which include PigG Leu31, Asn32 and Asp35 upon PigI binding and PigG Leu31 upon PltF binding.[9] These perturbed residues are also observed at the interfaces of the crosslinked PigG-PigI and PigG-PltF structures, supporting that these residues are involved in forming a productive interaction (Figure 3.S7). Lastly, these findings are consistent with the previously reported PltL-PltF structure, which was crosslinked in a

similar manner using an identical proline AVSN probe.[10] The PCP, PltL interacts through its loop 1 to PltF's A_{sub} domain, forming similar hydrogen-bonding and hydrophobic interactions found in the PigG-PigI and PigG-PltF interfaces (Figure 3.S7). Interestingly, unlike PigG, PltL is specific towards its cognate A domain, PltF, and is not prolylated efficiently by PigI despite the formation of similar PCP-A domain interfaces.

In order to tease out the differences in the molecular basis of partner protein specificity of PigG and PltL, the solution NMR structures of standalone (unbound) *holo*-PigG (PDB 5JDX)[9] and *holo*-PltL (PDB 2N5H)[13] were compared to the structures of bound PCPs observed in the X-ray crystal structures of the crosslinked PCP-A domain complexes, which revealed dependence of the PCP-A domain complex formation on the PCP's loop 1 dynamics (Figure 3.3). Closer inspection of the PltL-PltF structure reveals a change in the orientation of the beginning of PltL's loop 1 region that distinguishes the conformations of loop 1 in the bound and unbound states. This conformational change creates a hydrophobic pocket that is filled with PltF's Ile454 (Figure 3.3I). In the bound structure, the beginning of loop 1 of PltL was disordered and unresolved in the crystal structure.[10] When compared to standalone PigG, PigG in the PigG-PigI and the PigG-PltF structures show a similar, but more modest change in the conformation of the N-terminal portion of PigG's loop 1 (residues 17-25), as this region shifts slightly to form a hydrophobic pocket that accommodates PigI's Leu447 or PltF's Ile454, in the PigG-PigI and PigG-PltF complexes (Figure 3.3b, Figure 3.3g). Differences in the loop dynamics and conformation of the beginning part of loop 1 likely contribute to the promiscuity of PigG and the specificity of PltL. PCP-A domain specificity may be installed by a difference in loop lengths and amino acid composition. PltL loop 1 is 22 residues long, two residues
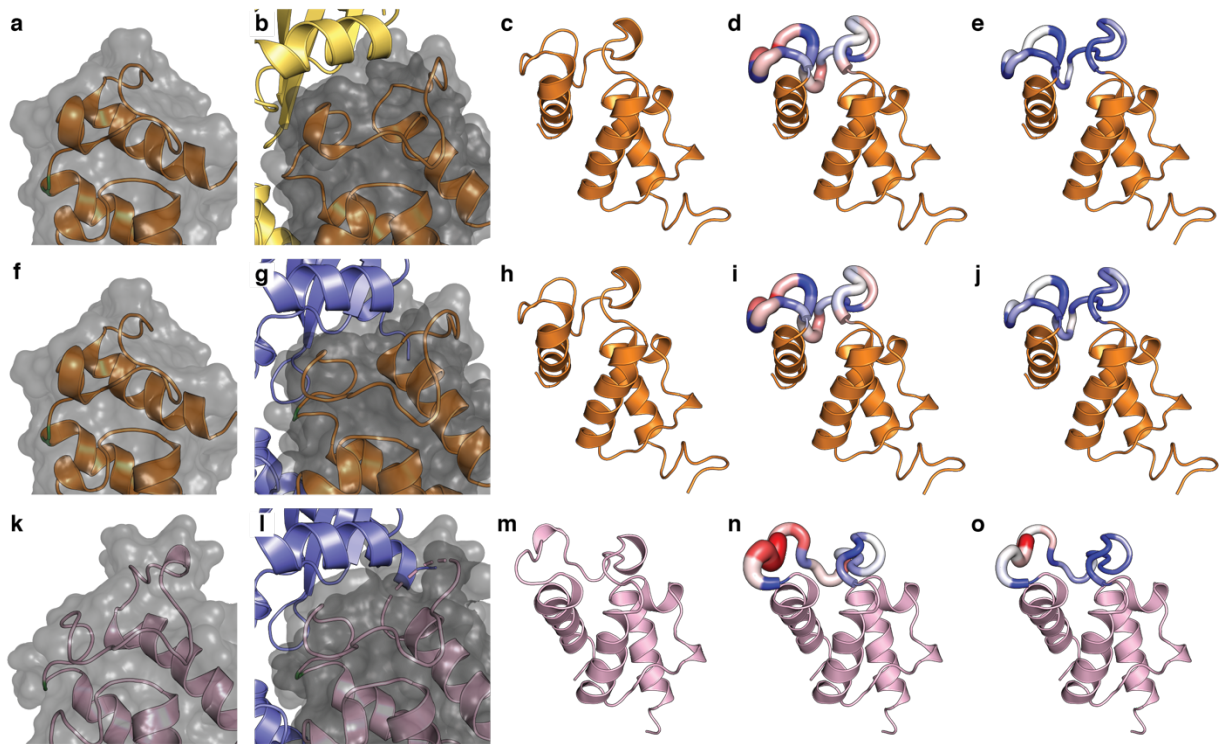
Figure 3.3: Analysis of loop dynamics. a,f,k) Illustrations of the loop orientation and "Connelly" surface of the PigG (a & f) and PltL (k) in their unbound states. Images generated using NMR structures of PigG (PDB 5JDX) and PltL (PDB 2N5H). b,g,l). Illustrations of the loop orientation and "Connelly" surface of the PigG (a & f) and PltL (k) in their PigI, PltF, and PltF-bound states. Loops have undergone structural reorganization upon PCP-A domain compex formation. Images generated using the X-ray structures of PigG-PigI (PDB 7THN) and PigG-PltF (PDB 7THQ) as well as PltL-PltF (PDB 6O6E). c,h,m. NMR solution structures of PigG (c & h) and PltL (m) oriented to clearly present the PCPs' key loop. d, i, n) Root mean square fluctuations (RMSFs) of the standalone PigG (d & h) and PltL (k) sampled over the course of MD simulations mapped onto the loops of the PCPs. e, j, k) RMSFs of the PigI and PltF-bound PigG (e & h, respectively) and PltF-bound PltL (k) sampled over the course of MD simulations mapped onto the loops of the PCPs. Color spectrum shows increasing backbone RMSFs calculated on a per-residue basis, whereas loop thickness from thin to thick indicates increasing sidechain RMSFs calculated on a per-residue basis. Note that PigG, PltL, PigI, and PltF are colored orange, pink, yellow, and purple, respectively.

longer than PigG's loop 1, where PltL contains an additional Pro21 and Ser22; a proline residue is not found at the corresponding position in loop 1 of PigG (Figure 3.S4). A comparison of the solution NMR structures of PigG and PltL demonstrate that this proline and a following glycine kinks PltL's loop 1 in a manner that distinguishes it from PigG's loop 1 and may play a role in PCP-A domain specificity.

3.4. MD simulations to uncover PCP loop dynamics responsible for specificity

To elucidate the PCP loop 1 dynamics, MD simulations were performed on *holo*-PigG, *holo*-PltL, prolyl-PigG, and prolyl-PltL as well as the PigG-PigI, PigG-PltF, and PltL-PltF complexes (Figure 3.3, Figure 3.S9). Initial root mean squared fluctuation (RMSF) analysis revealed the beginning of the PCP loop 1 of both *holo*-PltL and *holo*-PigI experienced larger backbone and side chain fluctuations compared to other regions of loop 1. MD simulations of each PCP-A domain complex revealed lower backbone fluctuations in all loop 1 regions relative to the unbound PCP. Bound PltL in the PltL-PltF complex, however, has higher loop 1 fluctuations at the beginning of loop 1 compared to bound PigG in the PigG-PigI and PigG-PltF complex. These increased fluctuations suggest a lack of a single low energy conformation of the beginning region of PltL loop 1. This is also supported by the PltL-PltF crystal structure, where the beginning region PltL loop 1 did not have any electron density.[10]

3.5. Noncognate PCP-A domain interface manipulation for enhanced activity

The MD simulations, in addition to A domain and PCP sequence alignments and structural superpositions, identified residues that can be targeted to introduce and

enhance noncognate activity between PltL and PigI. Site directed mutagenesis was performed to swap residues between PigG and PltL to give mutant PltL (mPltL) W37L, G38N, and ΔS21 ΔP22 as well as between PltF and PigI to give mutant PigI (mPigI) K212F, N396R, L447I. The PltL ΔS21 ΔP22 and PltL W37L mutations did not significantly affect activity with PigI, whereas the G38N mutation increased prolyl-PltL formation to 32% from the initial 15% between wild-type PigI and PltL (Figure 3.4). Combination of single mutations to form the PltL double mutants generally did not affect activity, whereas surprisingly the PltL ΔS21 ΔP22 G38N mutation increased conversion to 59%. Remarkably, combination of all three PltL mutations increased conversion to 72% with a turnover rate of 4.54 mPltL/hr, where the turnover rate of PigI with PltL was calculated to be 0.92 PltL/hr (Table 3.S2). Of the individual PigI mutations, the single N396R mutation significantly improved activity with WT PltL to 74% conversion, with a turnover rate of 6.17 PltL/hr. The PigI double and triple mutations did not provide any more prolyl-PltL formation compared to the single N396R mutation. Incubation of PigI N396R with PltL ΔS21 ΔP22 W37L G38N provided an increased product formation at 93% completion with a calculated turnover rate of 18.87 mPltL/hr, which is a 20-fold increase compared to the wild-type proteins.

To better understand how loop dynamics of PltL are altered by mutagenesis, we performed simulations of the holo forms of PltL ΔS21 ΔP22 and PltL ΔS21 ΔP22, W37L, G38N mutants using the methodology briefly summarized above (and more exhaustively described in the Methods section of this article). To generate coordinates for the deletion mutants we used RoseTTAFold, an AI-driven structure prediction tool to predict the conformation of the altered loops, as we anticipated that residue deletions would

98
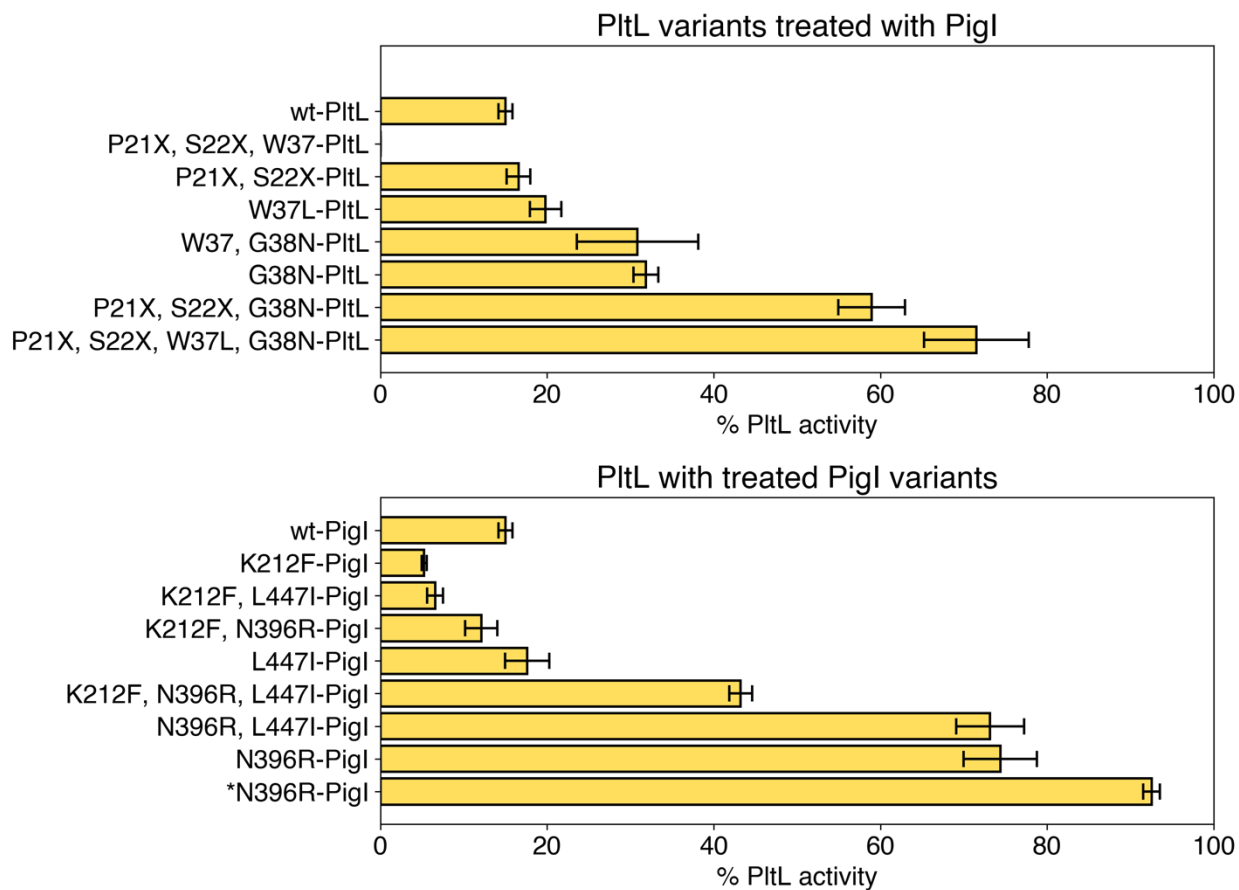
Figure 3.4: The initial activity of rationally designed PigI or PltL mutants. Mutant PltL was incubated with WT PigI (top) and WT PltL was also incubated with mutant PigI (bottom). Activities were monitored using HPLC, and % prolyl-PltL was calculated as described previously. The *N396R-PigI was incubated with the PltL ΔS21 ΔP22 W37L G38N mutant.

necessarily influence the backbone conformational preferences of the loop. Interestingly and perhaps unsurprisingly, a template-based comparative modeling approach using RosettaCM predicted approximately identical structures of the two PltL mutants. Analysis of the resulting RoseTTAFold-determined structures demonstrated that the deletion of PltL's Ser21 and Pro22 residues have a dramatic effect on the geometry of the loop 1 (Figure 3.S11). These modeled structures feature a loop orientation that more closely resembles that of PigG's loop 1 arrangement in both the unbound NMR structure as well as the crosslinked PigG-PigI and PigG-PltF structures (Figure 3.S10), thus supporting the hypothesis that the backbone conformation is influenced by PltL's Ser21 and Pro22 and that the loop 1 conformation can modulate PCP-A domain recognition.

3.6. Discussion

The experimental and computational data enhances our understanding of the protein-protein interactions that mediate catalysis and molecular recognition in type II NRPS biosynthesis, while clarifying the essential role of a PCP loop element in determining PCP-A domain binding specificities.

3.6.A. PCP-A domain interface in relation to other carrier protein dependent pathways

X-ray crystal structures, mutagenesis studies, and computer simulations demonstrate that the PCP-A domain interfaces are largely stabilized by hydrophobic interactions. The interfaces of PltL-PltF, PigG-PigI, and PigG-PltF measure in terms of buried surface area, 900 750, and 750 Å^2 , and are much larger than the interfaces

100

observed between type II acyl carrier proteins and type II fatty acid synthases, which are dominated by electrostatic and polar interactions.[14–17] Taken together these findings suggest that type II NRPS proteins may have evolved to form relatively hydrophobic interactions to maintain orthogonality between carrier protein-mediated secondary metabolic pathways and type II fatty acid biosynthesis, where type II fatty acid biosynthesis is dependent on electrostatic protein-protein interactions (Figure 3.S12).[18]

Interestingly, key polar residues on both the PCPs and A domains appear recognize the PPant cofactor to form productive complexes. This phenomenon is evidenced by the effect of the introduction of a positively charged residue, N396R, at the entrance of PigI's substrate binding pocket, which yielded a mutant PigI that had ~6-fold more when treated with wild type PltL; this mutation introduces a long-range electrostatic interaction involving the arginine's guanidinium group of N396R and the phosphate of the PPant cofactor.[19] Sequence alignments of 75 A domains reveal that positively charged residues, like Lys212 of PigI or Arg404 of PltF, at the entrance of the PPant binding tunnel are conserved amongst type I and type II NRPS A domains (Figure 3.S2, Figure 3.S5, Figure 3.S6). Structural analysis of previously solved PCP-A domain structures also reveal a positively charged amino acid interacting with the phosphate of the PPant, thus supporting the sequence alignments and importance of PPant recognition in PCP-A domain binding.

3.6.B. Importance of the PCP loop 1 region in A domain binding

Conversely, a single G38N mutation of PltL increases its activity toward PigI. Surprisingly, subsequent mutation additions of ΔS21 ΔP22 and W37L to the G38N

mutation revealed a synergistic 4-fold increase in activity with PigI. The G38N mutation seems to add additional hydrogen bonding interactions with the A domain, whereas the ΔS21 ΔP22 mutation changes the loop 1 structure to adopt a conformation more accessible to the A domain hydrophobic residue. The PCP loop 1 length is generally between 17-21 residues from 84 aligned PCP sequences, however, PltL was the only PCP with a loop 1 length of 22 residues (Figure 3.S3). Lastly, decreasing the size and thus steric clash of the hydrophobic residue (PltL W37L) involved in forming the PCP loop 1 hydrophobic pocket may also enhance PCP and A domain recognition.

3.6.C. Towards a more complete understanding of the adenylation mechanism

The step-wise increase in activity observed in additional PltL mutations may provide hints towards the mechanism of PCP-A domain binding. First, initial PCP attraction to the A domain can be influenced by the long-range electrostatic attraction between the PPant phosphate and positively charged residues at the entrance to the PPant tunnel (PltF Arg404, PigI Lys212, PigI N396R). Subsequently, adjacent hydrogen bonds may begin to form between the PCP loop 1 region (PigG Asn32, PltL Gly38, and PltL G38N) with the A_{sub} domain (PigI Ile462 Asn464, PltF Ile469 Ile470 Asp471). This intermediate binding conformation may prime the PCP loop 1 hydrophobic pocket (PltL ΔS21 ΔP22) to more readily accommodate the A_{sub} domain hydrophobic residue (PigI Leu447, PltF Ile454). Lastly, the PCP loop 1 hydrophobic pocket is then formed (PigG Leu31, PltL Trp37, PltL W37L) and accessed by the A domain hydrophobic residue. The order of PCP-A domain binding can be hypothesized as follows, where 1) initial electrostatics attract the two proteins, then 2) adjacent specific hydrogen bonding

interactions form, which allows for 3) neighboring hydrophobic interactions to develop, thus creating an optimized protein-protein interface that can enables PCP and A domain binding and for thiolation to occur.

The addition of positive charge at the PPant tunnel entrance has proven to be significant and powerful in enhancing a non-cognate PCP-A domain interaction, which may be utilized as an initial step in not only designing a new PCP-A domain interface, but also other carrier protein dependent pathways such as fatty acid synthases and polyketide synthases, which generally rely on electrostatic interactions to form a protein-protein interface.[18] Although the A domain prolylation activity of our designed PCP-A domain interface is still two orders of magnitude slower than the wild-type complex, further interface design can be performed to reach wild-type levels of prolylation. We envision that the protein-protein interface design of PCP-A domain interfaces and other PCP-partner protein interfaces will enhance future success in the combinatorial biosynthesis of carrier protein dependent pathways.

3.7. Conclusion

The X-ray crystal structures, mutagenesis studies, and MD simulations in conjunction with prior NMR experiments and X-ray crystallography demonstrates the integral role of carrier protein loop 1 dynamics in mediating molecular recognition and catalysis in type II NRP biosynthesis. This work also provides insight into the noncovalent interactions that promote PCP-A domain complexation. Perhaps most importantly, these studies provide evidence that rational interface design guided by structural, chemical, and

computational biology can be used to manipulate or alter the binding specificities of proteins responsible for the biosynthesis of complex natural products.

3.8. Acknowledgments

Chapter 3, entitled "Theory and experiment demonstrate the essential role of loop dynamics in type II NRPS biomolecular recognition", in full, is a reprint of the material as it appears in: Corpuz, J. C.*; Patel, A.*; Davis, T. D.; Podust L. M.; McCammon, J. A.; Burkart M. D. Theory and experiment demonstrate the essential role of loop dynamics in type II NRPS biomolecular recognition. *In Preparation*. **2022**. The dissertation author was the primary co-author of this paper along with Dr. Ashay Patel.

3.9. Methods

3.9.A. Protein expression and purification

Expression and purification of *holo*-PigG and *holo*-PltL. The pET22b-PltL plasmid and the pREP4-sfp plasmid were obtained from the laboratory of Christopher Walsh at Harvard University. Both plasmids were co-transformed into *Escherichia coli* BL21 cells. The BL21 cells were grown in Luria-Bertani (LB) media with 50 µg/mL kanamycin and 100 µg/mL ampicillin. Expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.6. The cells were grown for at 25 °C for 16 hours, then harvested at 2500 relative centrifugal force (RCF). The pelleted cells were resuspended in 30 mL of 50 mM NaCl, 50 mM Tris pH 7.50, 0.25 mg/mL lysozyme. The cells were further lysed with the Ultrasonic Processor FS-600N. The lysate was centrifuged at 12000 RCF for 1 hour and the supernatant was passed through the

Novagen Ni-NTA resin column. *holo*-PltL was eluted with 15 mL of 50 mM NaCl, 50 mM Tris pH 7.50, and 250 mM imidazole. The eluant was passed through a Superdex S75 16/600 pg size exclusion column on an AKTA Pure fast performance liquid chromatography (FPLC) with 20 mM Tris pH 7.5 and 20 mM NaCl. The concentration of *holo*-PltL was determined using the Bradford Assay.

The pET22b-PigG plasmid and the pET29b-sfp plasmid were co-transformed into *Escherichia coli* BL21 cells. The cells were grown identically as previously described for *holo*-PltL, except the induced cells were grown at 37 °C for 16 hours. *holo*-PigG was purified as previously described for *holo*-PltL.

Expression and purification of PigI and PltF. The pET37b-PltF plasmid and pET22-PigI was obtained from the laboratory of Christopher Walsh at Harvard University. The plasmids were transformed into *E. coli* BL21 cells. The cells were grown in LB media with 50 µg/mL kanamycin. Expression was induced with 0.5 mM IPTG at an OD_{600} of 0.6. The cells were grown at 16 °C for 16 hours then harvested at 2500 relative centrifugal force (RCF). The pelleted cells were resuspended in 30 mL of 50 mM NaCl, 50 mM Tris pH 7.50, 0.25 mg/mL lysozyme. The cells were further lysed with the Ultrasonic Processor FS-600N. The lysate was centrifuged at 12000 RCF for 1 hour and the supernatant was passed through the Novagen Ni-NTA resin column. PltF and PigI were eluted with 15 mL of 50 mM NaCl, 50 mM Tris pH 7.5, and 250 mM imidazole. The eluant was passed through a Superdex S200 16/600 pg size exclusion column on an AKTA Pure FPLC with 20 mM Tris pH 7.5 and 20 mM NaCl.

3.9.B. Synthesis of Pro-AVS probe.

The synthesis of prolyl-adenosine vinylsulfonamide probe was performed as described previously.[20]

3.9.C. X-ray crystallography

Crystallization of the PCP-A domain complexes. To trap the PigG-PigI and PigG-PltF complexes, a trapping reaction was performed, which consisted of 30 mM Tris pH 7.5, 2.0 mM $MgCl_2$, 2.0 mM tris(2-carboxyethyl)phosphine (TCEP), 0.02% Triton X, 0.34 mM Pro-AVSN, 0.17 mM *holo*-PigG, and 0.17 mM PigI or PltF. The reaction was incubated overnight at 4 °C. Initial crystallization conditions were identified using the Hampton Research Index and Crystal screens through hanging drop vapor diffusion. The PigG-PigI complex was crystallized in 0.3 M $MgCl_2$, 24.57% PEG 3350, and 0.1 M BIS-TRIS pH 6.23. The final PigG-PltF complex was crystallized in 35% PEG 3350, 0.50 M NaCl, 0.1 M BTP pH 7.50, and 0.01 M sodium citrate tribasic. Each crystal was cryoprotected in 20% glycerol and 80% of the crystallization condition and frozen in liquid nitrogen.

Data collection and structure Determination. X-ray diffraction data were collected at the Advanced Light Source beam line 8.3.1 at the Lawrence Berkeley National Laboratory at a temperature of 110 K with a wavelength of 1.12 Å. Each data set was processed with XDS[21]. An initial structure of the PigG-PigI complex was solved using molecular replacement with MOLREP[10] using PltF from PDB ID 6O6E[22] as the search model for PigI. The structure was iteratively refined via automated refinement with REFMAC[23] and manual refinement with COOT[13]. Then, the solution NMR structure of PigG PDB ID 5JDX

[10] was used as a search model for the second round of molecular replacement to model PigG. Lastly, the PPant/Pro-AVS ligand and waters were added. The coordinates and structure factor amplitudes were deposited in the Protein Data Bank with the accession code 7THN. The PigG-PltF complex was solved in the same manner as the PigG-PigI complex and deposited in the PDB as 7THQ.

3.9.D. Alanine scanning of the protein-protein interface

Alanine scanning mutagenesis of PigI and PltF. Alanine scanning of PigI and PltF was prepared by introducing single mutations to alanine using the QuickChange PCR method as described previously.[10] The mutants were expressed and purified as described for PigI and PltF, except scaled down to a 100 mL culture.

The mutant PigI assay with PigG contained 50 mM Tris pH 7.5, 12.5 mM $MgCl_2$, 2.0 mM TCEP, 5 mM ATP, 5 mM L-Pro, 0.15 mM *holo*-PigG, and 0.054 uM PigI. The mutant PltF assay with PigG contained the same reaction mixture, except with 1.19 uM PltF. The A domain was added last, then the reaction was mixed and incubated at 25 °C for 4 minutes. The reaction was quenched to 1.33 % formic acid and centrifuged for 10 minutes at 10,000 RCF. 8 uL of each supernatant was passed over the Ascentis Express Peptide ES-C18 column with solvent B for 2 minutes, then increasing 5-40% B over 8 minutes, and finally 40-44% over 10 minutes (Solvent A = H_2O, 0.05% trifluoroacetic acid (TFA); solvent B = acetonitrile, 0.05% TFA). Analyses were conducted on an Agilent HP 1100 series HPLC. The protein elution was monitored by absorbance at 210 nm (A_{210}) (Figure S5). The percent of *prolyl-* and *holo*-PltL was calculated through the baseline integration of the chromatographic peak areas and using the A_{210} of *prolyl*-PltL divided by

the summation of A_{210} of *prolyl*-PltL with A_{210} of *holo*-PltL. Each assay was performed in triplicate (Figure S10).

3.9.E. MD Simulations

Structure preparation. The following carrier proteins were modeled in their *holo* and pyrrolyl-forms of standalone carrier proteins, PigG, PltL and two mutant PltL variants, ΔP21, ΔS22-PltL and ΔP21, ΔS22, W37L, G38N-PltL. In addition, the following PCP•A complexes were simulated, PltL•PltF, PigG•PigI, and PigI•PltF. For simulation work, the coordinates for the wildtype PCPs, PigG and PltL were generated using previously reported solution-phase NMR structures of holo-PltL (PDB ID: holo-PigG), prolyl-PltL (PDB ID: 2N5h), and holo-PigG (PDB ID: 5JDX). The crosslinked PCP-A didomain complexes reported herein and previously, PltL-PltF (PDB ID: 6O6E), PigG-PigI, and PigG-PltF were used to generate initial coordinates for simulations of the three PCP•A complexes. The prolyl-PCPs and PCP•A complexes possess a substrate or intermediate species that could exist in either a neutral or protonated form; in all case both forms were considered. These structures are show in Scheme X. In total, 20 systems were subjected to MD simulation.

The variants of PltL, ΔP21, ΔS22-PltL and ΔP21, ΔS22, W37L, G38N-PltL, were prepared using RoseTTAFold using Robetta[24,25], a continually evaluated (through CAMEO) protein structure prediction service. Schrodinger's Protein Preparation Wizard (https://www.schrodinger.com/protein-preparation-wizard and https://www.schrodinger.com/prime) were used to add missing C-, N-terminal residues and missing side chains not resolvable from the experimental density. Hydrogen atoms

were added to all heavy atoms to cap all open valences and to predict the protonation states of the titratable residues in each structure assuming a pH of 7.4. To optimize their orientation, all waters resolved crystallographically were preserved for computer simulation. Histidine protonation states were inspected by hand.

In an analogous manner, TLEAP (https://ambermd.org/CiteAmber.php) were used to generate AMBER (ff14SB/GAFF2) topology and the parameters files for the simulation cell of all 18 systems of interest. ANTECHAMBER was used to parameterize the phosphopantetheine, aminoacyl substrates, **X** and **Y**, aminoadenylate intermediate, **A** and **B,** as nonstandard residues, using GAFF2 force field (https://ambermd.org/CiteAmber.php).[26] Topology and parameter files were prepared using TLEAP. Utilizing TLEAP, all simulation cells were prepared using TIP3P water molecules[27] to generate isometric cell to ensure that the cell walls are 10 Å away from the closest portion of the coordinates derived from experimental data in either data. The restrained electrostatic potential method[24] was used to determined partial atomic charges for the ANTECHAMBER parameterization of all nonstandard residues.[24] The potential was computed at the HF/6-31G(d) level of theory using Gaussian 09 ([28]).

Simulation methodology. All simulations were performed using GPU-accelerated Amber20. The Particle Mesh Ewald method was used to treat all non-bonded interactions with a 10 Å cutoff. The SHAKE algorithm was used to constrain all nonpolar bonds involving a hydrogen atom. Simulations were performed using a 2 fs time-step. All solvated PCP and PCP•A domain complexes were minimized using a two-step protocol, in which the solvated was allowed to relax while the protein was restrained using a

harmonic potential (k = 500 kcal mol^{-1} Å^{-2}). After this step, each system was minimized

in an unrestrained manner. Next, the system was heated to a physiological temperature

of 305 K over the course of 0.4 ns NVT ensemble simulation. All solvated acyl-AcpPs was

equilibrated at 305 K over the course of unbiased 2.0 ns NPT simulation. The equilibrated

acyl-AcpPs systems were subjected to conventional molecular dynamics simulations. For

each system, five independent production-grade 1.0 µs cMD simulations were performed.

A total of 100 µs of production-grade simulation data was collected. For both NVT and

NPT simulations, the Langevin thermostat (λ = 5.0 ps^{-1}) was used to maintain

temperature control. Pressure regulation in NPT simulations (target pressure of 1 atm)

was achieved by isotropic position scaling of the simulation cell volume using a

Berendsen barostat. Coordinate data was written to disk every 0.5 ps.

Analysis and visualization of simulation data. Analysis was performed using

CPPTRAJ,[29–31] PYTRAJ, a Python front-end for the CPPTRAJ analysis code

(https://amber-md.github.io/pytraj/latest/overview.html#citations). Trajectories were

visualized using NGLview[29–31] and Pymol v2.3 (https://gaussian.com/g09citation/). All

data was plotted using matplotlib library of Python.

3.9.F. Interface design of PltL and PigI

Interface design between PltL and PigI was prepared by introducing single

mutations as described previously in alanine scanning of PigI and PltF. Expression and

purification of each PigI mutant was performed similarly to as described previously for

PigI and PltF alanine scanning, whereas each PltL mutant was expressed and purified similarly as *holo*-PigG and *holo*-PltL.

The designed PigI with the designed PltL initial assays contained 50 mM Tris pH 7.5, 12.5 mM MgCl$_2$, 2.0 mM TCEP, 5 mM ATP, 5 mM L-Pro, 0.05 mM *holo*-PltL, and 0.0238 mM mPigI. The reaction was incubated at 25 °C, 300 rpm, for 1 hour and quenched with 1.3% formic acid. The reaction was prepared, passed, and analyzed through the same HPLC protocol as outlined in Alanine scanning of PigI and PltF.

The time course assays of the designed PigI with the designed PltL contained 50 mM Tris pH 7.5, 12.5 mM MgCl$_2$, 2.0 mM TCEP, 5 mM ATP, 5 mM L-Pro, 0.079 mM *holo*-PltL, and 0.0238 mM mPigI. Each time course utilized between 5-7 time points and was repeated in at least duplicate. Each time point was quenched with 1.67% formic acid and subjected to HPLC analysis identically to Alanine scanning and Interface design assays.

3.9.G. Sequence alignment of PCPs and A domains

The curated PCP or A domain protein sequences from the phylogenetic analysis was separated into categories: pyrrole-incorporating PCP or A domains, type II PCP or A domains (including pyrrole-incorporating PCP or A domains), and type I PCP or A domains. The protein sequence from each category was aligned using Clustal Omega Multiple Sequence Alignment (https://www.ebi.ac.uk/Tools/msa/clustalo/). Sequence alignments and conservation were visualized in JalView.

3.10. Supplementary Information

Table 3.S1: X-ray data collection and refinement statistics for PigG-PltF and PigG-PigI. Highest resolution shell is in parenthesis.

	PigG-PltF	PigG-PigI
Wavelength	1.1 Å	1.1 Å
Resolution range	96.61 - 2.456 (2.544 - 2.456) Å	62.55 - 1.605 (1.662 - 1.605) Å
Space group	P 21 21 21	P 1 21 1
Unit cell	65.73 132.68 140.94 90 90 90	50.67 92.66 66.77 90 110.469 90
Total reflections	580603 (40335)	510904 (46791)
Unique reflections	45208 (3958)	75390 (7504)
Multiplicity	12.8 (10.2)	6.8 (6.2)
Completeness (%)	98.16 (81.89)	99.93 (99.93)
Mean I/sigma(I)	13.99 (0.78)	12.66 (1.62)
Wilson B-factor	61.81	15.90
R-merge	0.1532 (2.684)	0.1141 (1.045)
R-meas	0.1595 (2.819)	0.1237 (1.142)
R-pim	0.04389 (0.8344)	0.04718 (0.4542)
CC1/2	0.999 (0.274)	0.998 (0.603)
CC*	1 (0.656)	1 (0.867)
Reflections used in refinement	44937 (3698)	75380 (7504)
Reflections used for R-free	2303 (183)	3615 (358)
R-work	0.2061 (0.3424)	0.1633 (0.2662)
R-free	0.2550 (0.3838)	0.1949 (0.2760)
CC(work)	0.961 (0.537)	0.967 (0.743)
CC(free)	0.915 (0.315)	0.952 (0.773)
Number of non-hydrogen atoms	8613	4947
macromolecules	8446	4222
ligands	118	105
solvent	49	620
Protein residues	1169	575
RMS(bonds)	0.005	0.015
RMS(angles)	1.23	1.83
Ramachandran favored (%)	97.48	99.08
Ramachandran allowed (%)	2.17	0.92
Ramachandran outliers (%)	0.35	0.00

Rotamer outliers (%)	0.25	0.68
Clashscore	7.03	2.93
Average B-factor	68.52	20.43
macromolecules	68.55	18.71
ligands	65.10	25.59
solvent	72.35	31.27
Number of TLS groups	5	

Table 3.S2: Turnover numbers of select PigI and PltL mutants.

PigI Mutant	PltL Mutant	turnover (PltL/hr)
WT	WT	0.92 ± 0.07
WT	ΔP21 ΔS22 W37L G38N	4.53 ± 0.81
N396R	WT	6.17 ± 3.20
N396R	ΔP21 ΔS22 W37L G38N	18.87 ± 0.44
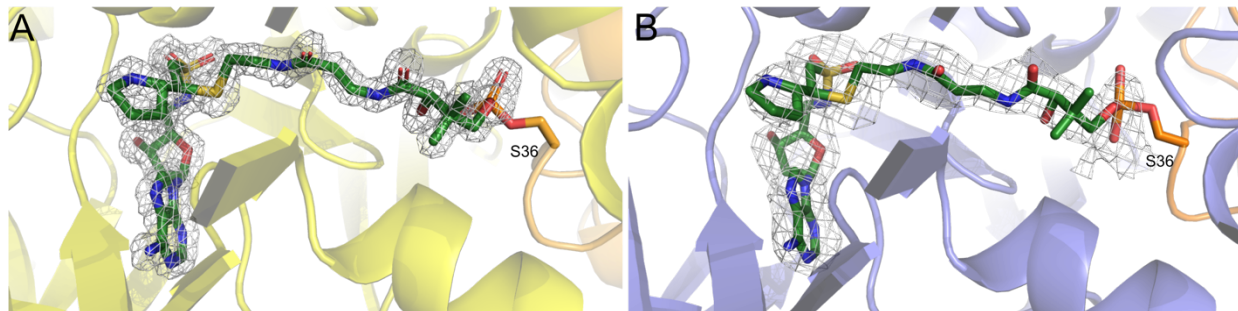
Figure 3.S1: fo-fc Omit map of PigG-PigI and PigG-PltF active site. The PPant-Pro-AVS ligand (I5M) was deleted, the structure was refined, and the (A) PigG-PigI fo-fc map (white) was contoured to 3.5 σ. The (B) PigG-PltF fo-fc was contoured to 3.0 σ.

A domains — Conservation analysis (Figure 3.S2)

prolyl A domains (12 proteins)

Site (PltF/LgrA)								
ppant binding F231/Y421	Ala 0%	Val 8%	Glu 0%	Thr 0%	0%			
	Gly 0%	Phe 8%	Arg 8%	Cys 0%	0%			
	Ile 0%	Trp 0%	His 0%	Met 0%	0%			
	Leu 0%	Tyr 58%	Lys 17%	Asn 0%	0%			
	Pro 0%	Asp 0%	Ser 0%	Gln 0%	0%			
ppant binding R404/R591	Ala 0%	Val 0%	Glu 0%	Thr 0%	0%			
	Gly 0%	Phe 0%	Arg 75%	Cys 0%	0%			
	Ile 0%	Trp 0%	His 0%	Met 0%	0%			
	Leu 0%	Tyr 0%	Lys 0%	Asn 8%				
	Pro 0%	Asp 0%	Ser 17%	Gln 0%	0%			
pcp loop 1 binding I454/G640	Ala 0%	Val 33%	Glu 0%	Thr 0%	0%			
	Gly 0%	Phe 0%	Arg 0%	Cys 0%	0%			
	Ile 8%	Trp 0%	His 0%	Met 0%	0%			
	Leu 58%	Tyr 0%	Lys 0%	Asn 0%	0%			
	Pro 0%	Asp 0%	Ser 0%	Gln 0%	0%			
pcp loop 1 binding K457/R643	Ala 0%	Val 8%	Glu 0%	Thr 0%	0%			
	Gly 0%	Phe 0%	Arg 0%	Cys 17%	0%			
	Ile 0%	Trp 0%	His 0%	Met 0%	0%			
	Leu 0%	Tyr 0%	Lys 0%	Asn 75%	0%			
	Pro 0%	Asp 0%	Ser 0%	Gln 0%	0%			

type II A domains (38 proteins)

Site				
ppant binding F231/Y421	Ala 3%	Val 13%	Glu 0%	Thr 0%
	Gly 0%	Phe 5%	Arg 16%	Cys 0%
	Ile 5%	Trp 0%	His 3%	Met 0%
	Leu 13%	Tyr 18%	Lys 5%	Asn 0%
	Pro 0%	Asp 0%	Ser 0%	Gln 3%
ppant binding R404/R591	Ala 0%	Val 0%	Glu 0%	Thr 0%
	Gly 18%	Phe 5%	Arg 45%	Cys 0%
	Ile 0%	Trp 0%	His 0%	Met 0%
	Leu 0%	Tyr 0%	Lys 3%	Asn 13%
	Pro 0%	Asp 0%	Ser 3%	Gln 11%
pcp loop 1 binding I454/G640	Ala 13%	Val 5%	Glu 18%	Thr 5%
	Gly 0%	Phe 0%	Arg 3%	Cys 0%
	Ile 0%	Trp 0%	His 0%	Met 0%
	Leu 8%	Tyr 0%	Lys 0%	Asn 8%
	Pro 0%	Asp 5%	Ser 5%	Gln 8%
pcp loop 1 binding K457/R643	Ala 5%	Val 0%	Glu 0%	Thr 3%
	Gly 0%	Phe 0%	Arg 39%	Cys 0%
	Ile 3%	Trp 0%	His 0%	Met 3%
	Leu 8%	Tyr 0%	Lys 37%	Asn 0%
	Pro 0%	Asp 0%	Ser 0%	Gln 0%

type I A domains (37 proteins)

Site				
ppant binding F231/Y421	Ala 0%	Val 3%	Glu 3%	Thr 3%
	Gly 0%	Phe 0%	Arg 41%	Cys 0%
	Ile 0%	Trp 0%	His 0%	Met 5%
	Leu 11%	Tyr 3%	Lys 16%	Asn 5%
	Pro 0%	Asp 0%	Ser 0%	Gln 8%
ppant binding R404/R591	Ala 0%	Val 0%	Glu 3%	Thr 0%
	Gly 3%	Phe 0%	Arg 92%	Cys 0%
	Ile 0%	Trp 0%	His 0%	Met 0%
	Leu 0%	Tyr 0%	Lys 0%	Asn 3%
	Pro 0%	Asp 0%	Ser 0%	Gln 0%
pcp loop 1 binding I454/G640	Ala 35%	Val 0%	Glu 11%	Thr 3%
	Gly 5%	Phe 0%	Arg 0%	Cys 0%
	Ile 3%	Trp 0%	His 0%	Met 0%
	Leu 3%	Tyr 0%	Lys 0%	Asn 0%
	Pro 3%	Asp 5%	Ser 27%	Gln 3%
pcp loop 1 binding K457/R643	Ala 0%	Val 0%	Glu 0%	Thr 0%
	Gly 0%	Phe 0%	Arg 70%	Cys 0%
	Ile 0%	Trp 0%	His 3%	Met 0%
	Leu 0%	Tyr 0%	Lys 11%	Asn 0%
	Pro 3%	Asp 0%	Ser 3%	Gln 3%

0 % — Consensus — 100 %

Figure 3.S2: Conservation analysis of aligned A domains. The amino acid sequences of proline-activating A domains, type II NRPS A domains, and type I NRPS A domains were gathered from previous phylogenetic analysis.[6] The sequences were aligned using Clustal Omega, and the consensus sequence of specific interface residues were analyzed, which include two PPant phosphate binding sites, A_{sub} domain hydrophobic residues that sit in the PCP loop 1, and A_{sub} domain hydrogen bond donors that form hydrogen bonds with the PCP loop 1. To ensure proper sequence alignment across classes, the X-ray crystal structures of PltF, PigI, and LgrA (A domain) were aligned and superposed in PyMol, which verified the 3D location of the residues. The type II A domains were include the proline-activating A domains.

| PCPs	prolyl PCPs (16 proteins)				type II PCPs (45 proteins)				type I PCPs (39 proteins)			
loop 1 **hydrophobic** **pocket** **W37/R725**	Ala 0%	Val 0%	Glu 6%	Thr 0%	Ala 2%	Val 2%	Glu 2%	Thr 2%	Ala 0%	Val 0%	Glu 0%	Thr 0%
	Gly 0%	Phe 0%	Arg 0%	Cys 0%	Gly 0%	Phe 2%	Arg 2%	Cys 2%	Gly 0%	Phe 0%	Arg 5%	Cys 0%
	Ile 0%	Trp 31%	His 0%	Met 0%	Ile 9%	Trp 15%	His 0%	Met 2%	Ile 8%	Trp 0%	His 0%	Met 3%
	Leu 56%	Tyr 0%	Lys 0%	Asn 0%	Leu 43%	Tyr 9%	Lys 0%	Asn 0%	Leu 79%	Tyr 0%	Lys 0%	Asn 3%
	Pro 0%	Asp 0%	Ser 0%	Gln 6%	Pro 0%	Asp 0%	Ser 0%	Gln 2%	Pro 0%	Asp 0%	Ser 0%	Gln 0%
loop 1 **hydrogen bond** **G38/G726**	Ala 0%	Val 0%	Glu 0%	Thr 0%	Ala 0%	Val 0%	Glu 0%	Thr 0%	Ala 0%	Val 0%	Glu 0%	Thr 0%
	Gly 63%	Phe 0%	Arg 0%	Cys 6%	Gly 80%	Phe 0%	Arg 4%	Cys 0%	Gly 100%	Phe 0%	Arg 0%	Cys 0%
	Ile 0%	Trp 0%	His 0%	Met 0%	Ile 0%	Trp 0%	His 0%	Met 0%	Ile 0%	Trp 0%	His 0%	Met 0%
	Leu 0%	Tyr 0%	Lys 0%	Asn 31%	Leu 0%	Tyr 0%	Lys 0%	Asn 13%	Leu 0%	Tyr 0%	Lys 0%	Asn 0%
	Pro 0%	Asp 0%	Ser 0%	Gln 0%	Pro 0%	Asp 0%	Ser 0%	Gln 0%	Pro 0%	Asp 0%	Ser 0%	Gln 0%
loop length **G20-S42/** **L711-S728**	22 6%	18 0%			22 2%	18 7%			22 0%	18 0%		
	21 38%	17 0%			21 13%	17 42%			21 0%	17 95%		
	20 56%	16 0%			20 22%	16 4%			20 0%	16 0%		
	19 0%				19 9%				19 5%			

0 % Consensus 100 %

Figure 3.S3: Conservation analysis of aligned PCPs. The amino acid sequences of proline-accepting PCPs, type II NRPS PCPs, and type I NRPS PCPs were gathered from previous phylogenetic analysis.[6] The sequences were aligned using Clustal Omega, and the consensus sequence of specific interface residues were analyzed, which includes a hydrophobic residue involved in creating the loop 1 hydrophobic pocket that accepts the A_{sub} hydrophobic residue and the loop 1 hydrogen bonding acceptor/donor that forms hydrogen bonds with the A_{sub} domain. Additionally, the length of the PCP loop 1 was analyzed. To ensure proper sequence alignment across classes, the X-ray crystal structures of PltL, PigG, and LgrA (PCP$_1$) were aligned and superposed in PyMol, which verified the 3D location of the residues. The type II PCPs include the proline-accepting PCPs.
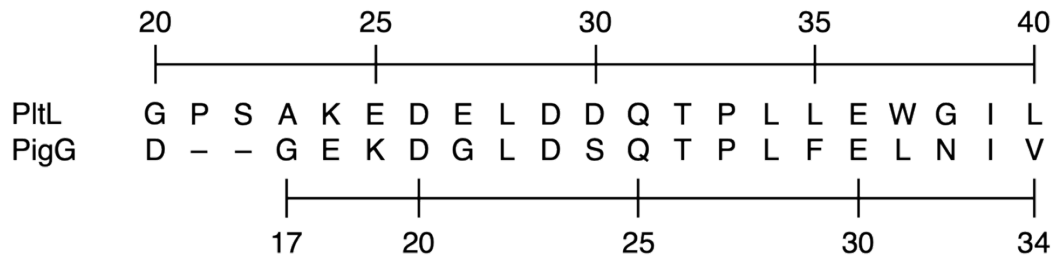
Figure 3.S4: Sequence alignment of the loop 1 regions of PltL and PigG.
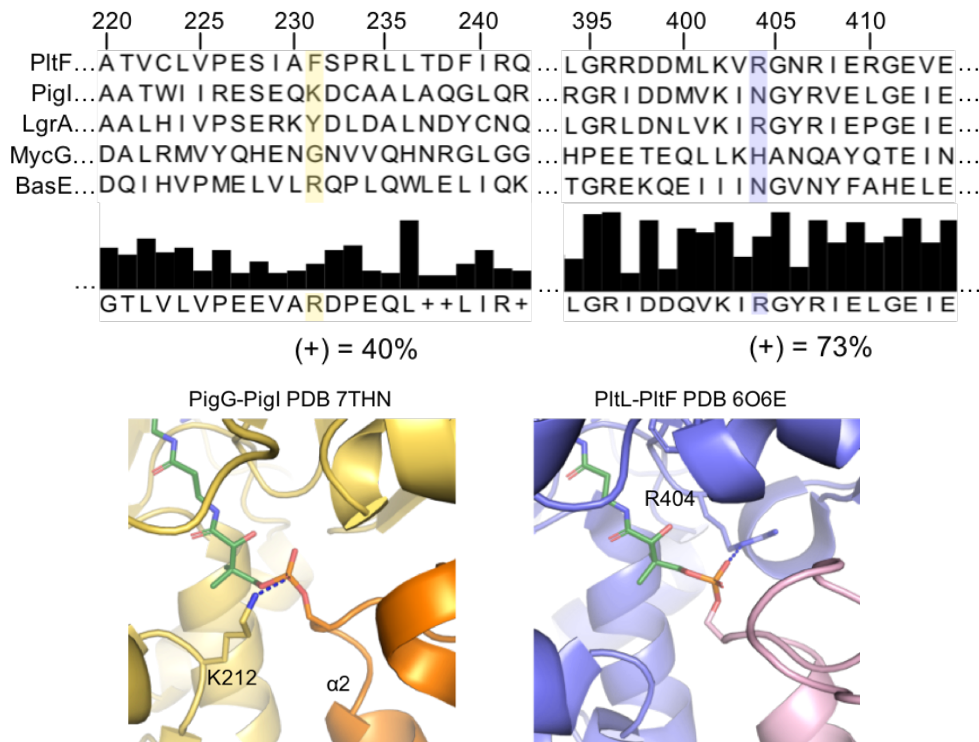
Figure 3.S5: Sequence alignment of 75 A domains. Highlighted in yellow at PltF F231 is the position of PigI K212 in secondary sequence space. Highlighted in purple at PltF R404 is the position of PltF R404 in secondary sequence space. Below is the consensus sequence and the percentage refer to the sum of Lys, Arg, and His residues in the alignment at the highlighted position. Below are the structures of PigG-PigI and PltL-PltF that reveals the electrostatic interaction between PigI K212 with the phosphate of the PPant and PltF R404 with the phosphate of the PPant.
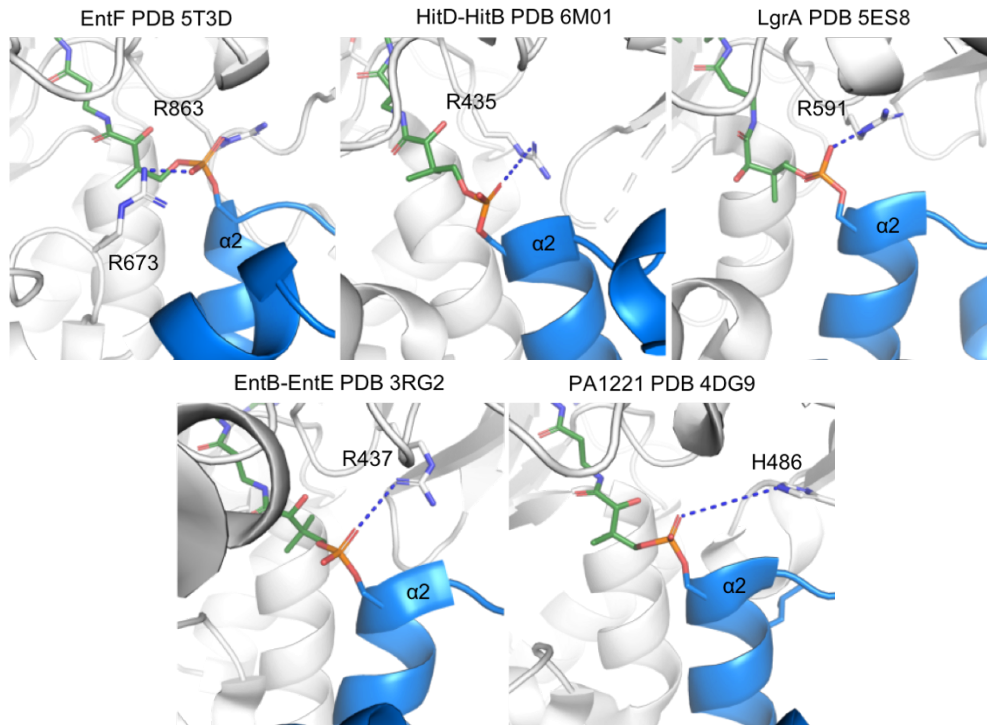
Figure 3.S6: Analysis of co-factor interactions in the PCP-A domain crystal structures. The A domains (white) are shown bound to the PCP (blue) with electrostatic interactions highlighted.
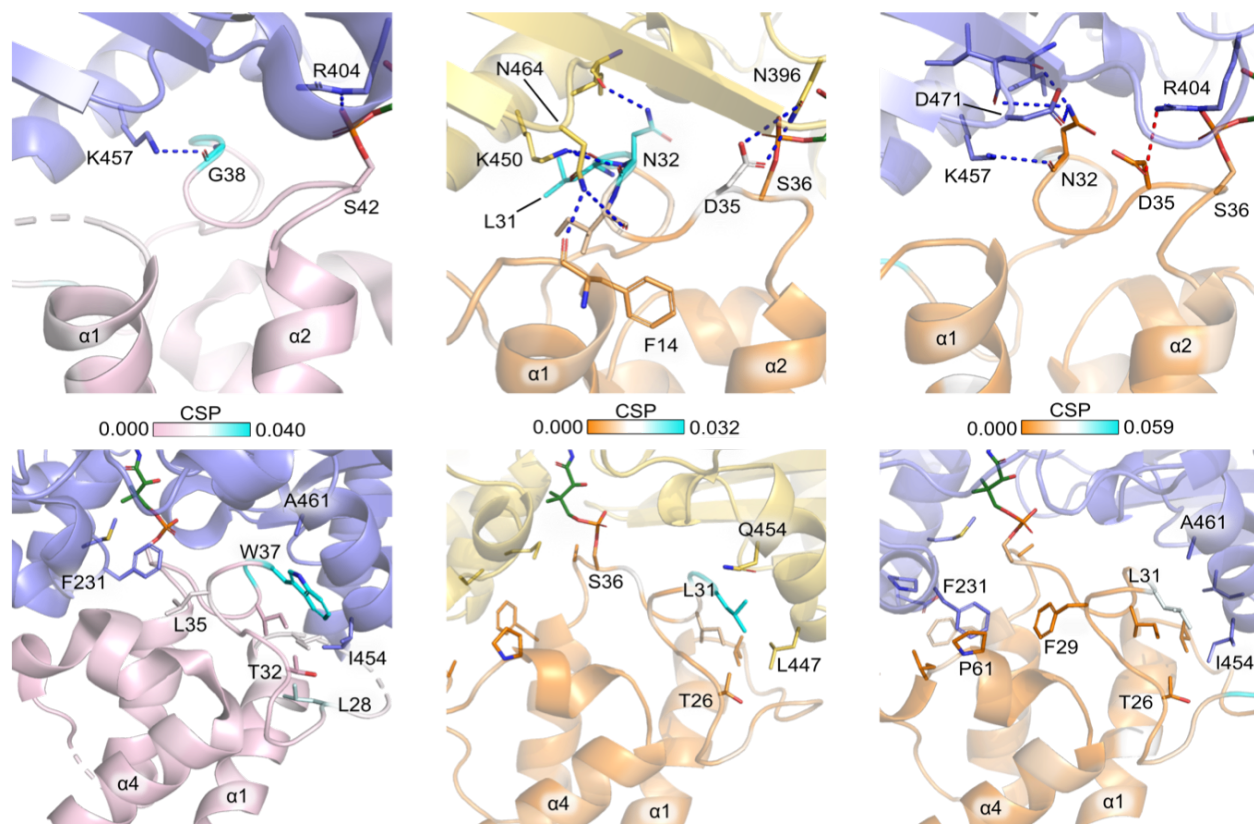
Figure 3.S7: Comparison of the PltL-PltF structure to the PigG-PigI and PigG-PltF structures. Top panels are the hydrogen bonding and electrostatic interactions, and bottom panels are the hydrophobic intereactions. Mapped onto the carrier proteins are their respective NMR titrations published previously.[9]
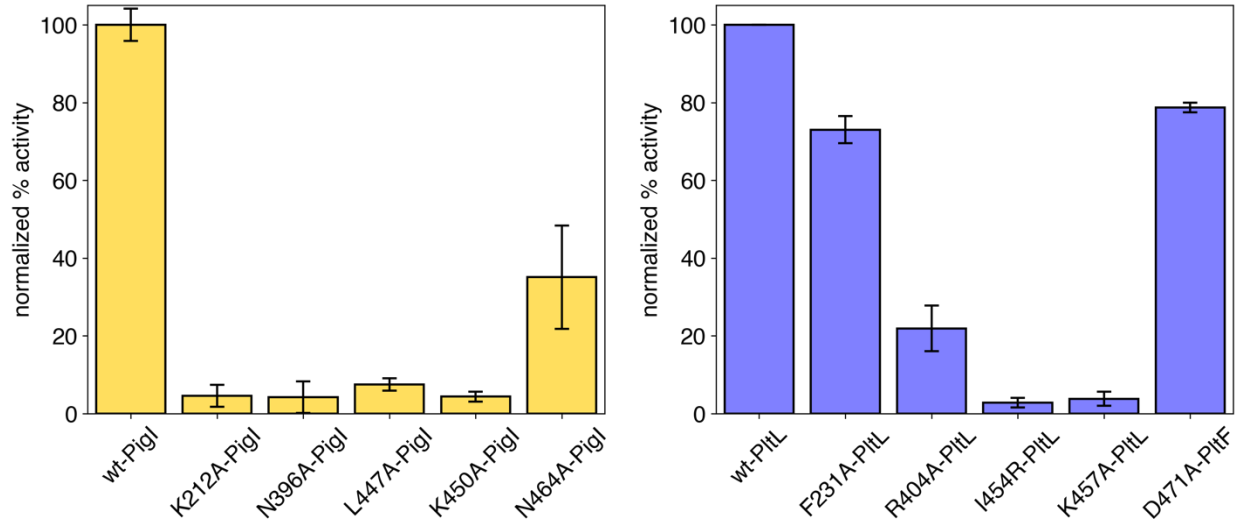
Figure 3.S8: Alanine scanning assays of A domain residues found at the interface with PigG. The %prolyl-PigG was normalized to the WT %prolyl-PigG for each enzyme.
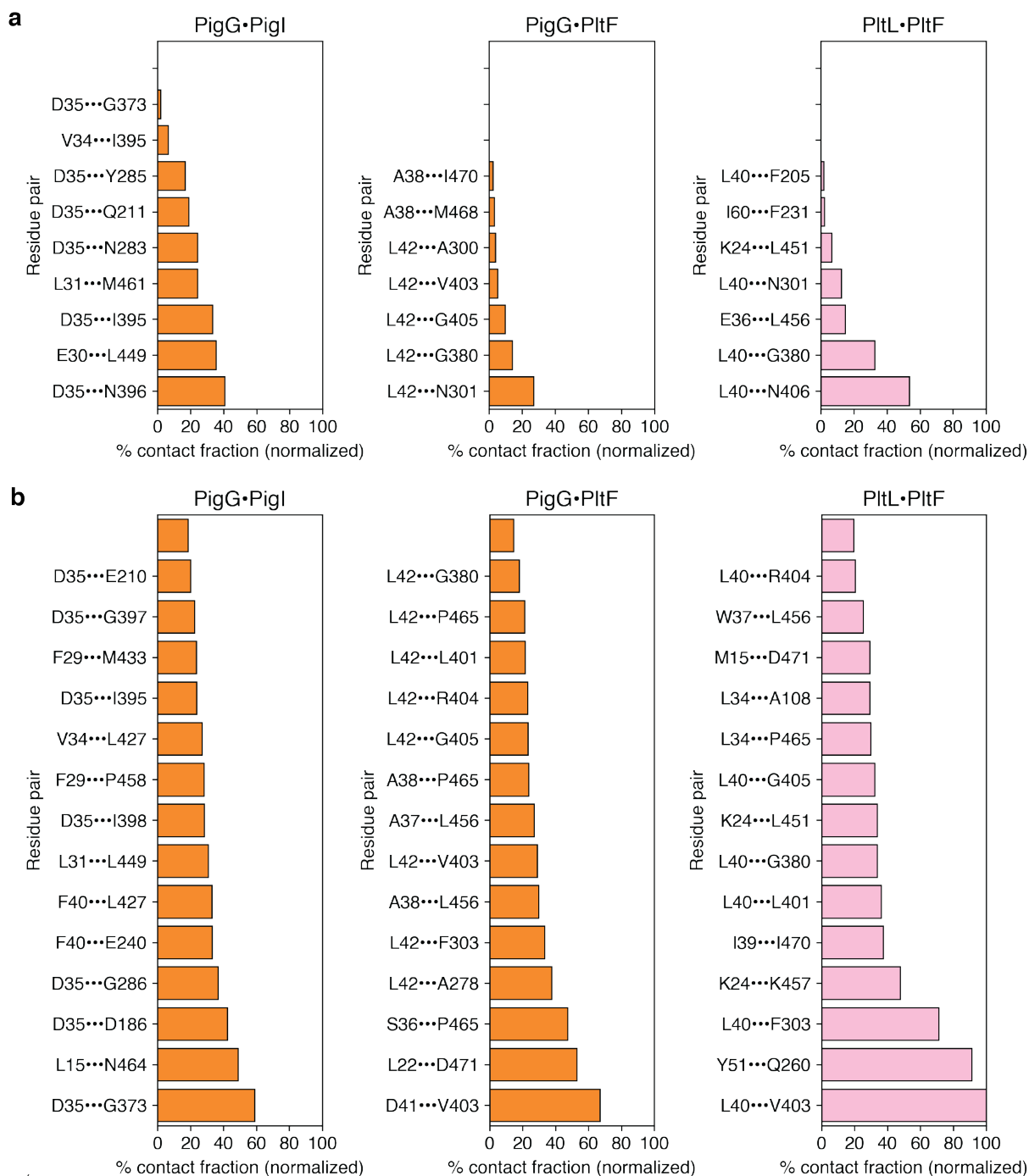
Figure 3.S9: Time-resolved analysis of the contacts formed between PCP and A domains. Contacts from the initial conditions and contacts formed during the course of the simulations are shown in panels a and b, respectively. The PCP•A domain complex is labeled above each bar graph. Residues listed on the left side of the interaction pair correspond to PCP and those on the right side of the pair correspond to the A domain. The bar graphs indicate the contact fraction normalized relative to the most frequently sampled contact. The contact fraction is defined as the total fraction of simulation data in which a residue pair is engaged in an intermolecular contact. A distance criterion of 3.0

Å or less between a pair of heavy atoms defines such a contact. Only pairwise contacts with contact fractions (prior to normalization) greater than or equal to 0.10 are included in the plots above. For brevity, no more than 15 contacts are shown in each plot.
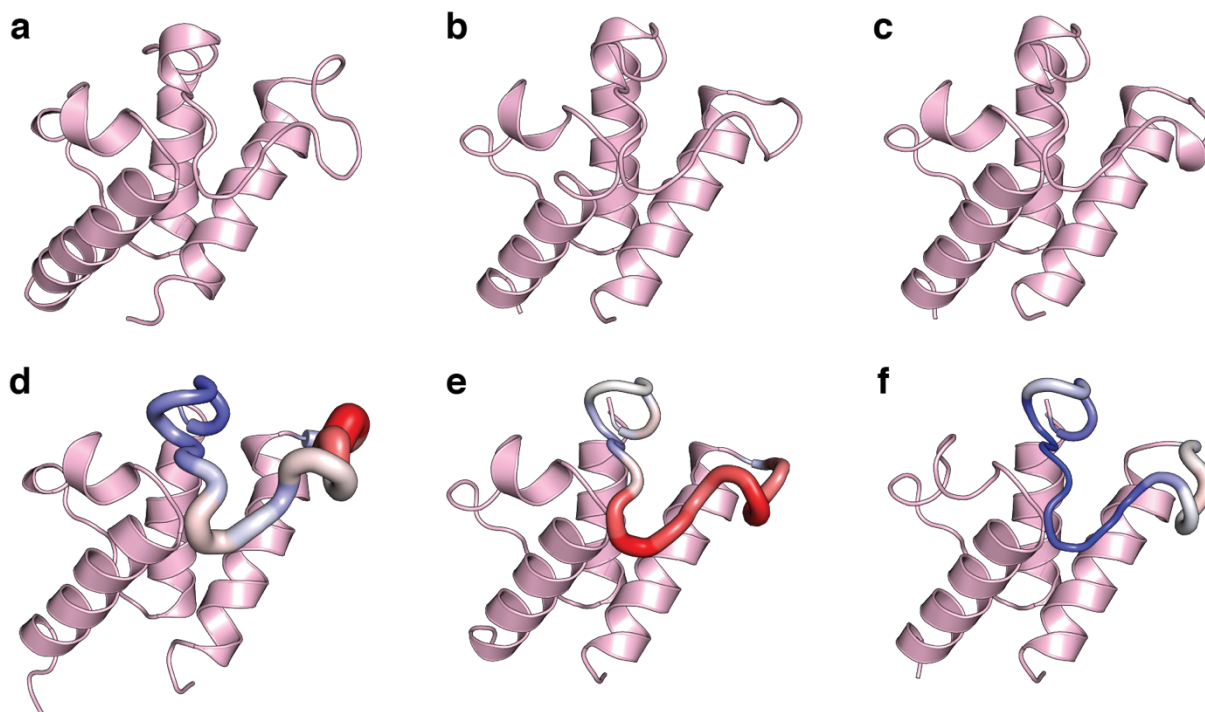
Figure 3.S10: Analysis of the loop conformation and dynamics of mutant PltL. NMR structure of wt prolyl-PltL (a), and computational models of ΔP21,ΔS22-PltL (b), and ΔP21,ΔS22,W37L,G38N-PltL Root mean square fluctuations (RMSFs) of wt prolyl-PltL (d), and computational models of ΔP21,ΔS22-PltL (e), and ΔP21,ΔS22,W37L,G38N-PltL (f) sampled over the course of MD simulations mapped onto the loops of the PCPs. Color spectrum shows increasing backbone RMSFs calculated on a per-residue basis, whereas loop thickness from thin to thick indicates increasing sidechain RMSFs calculated on a per-residue basis
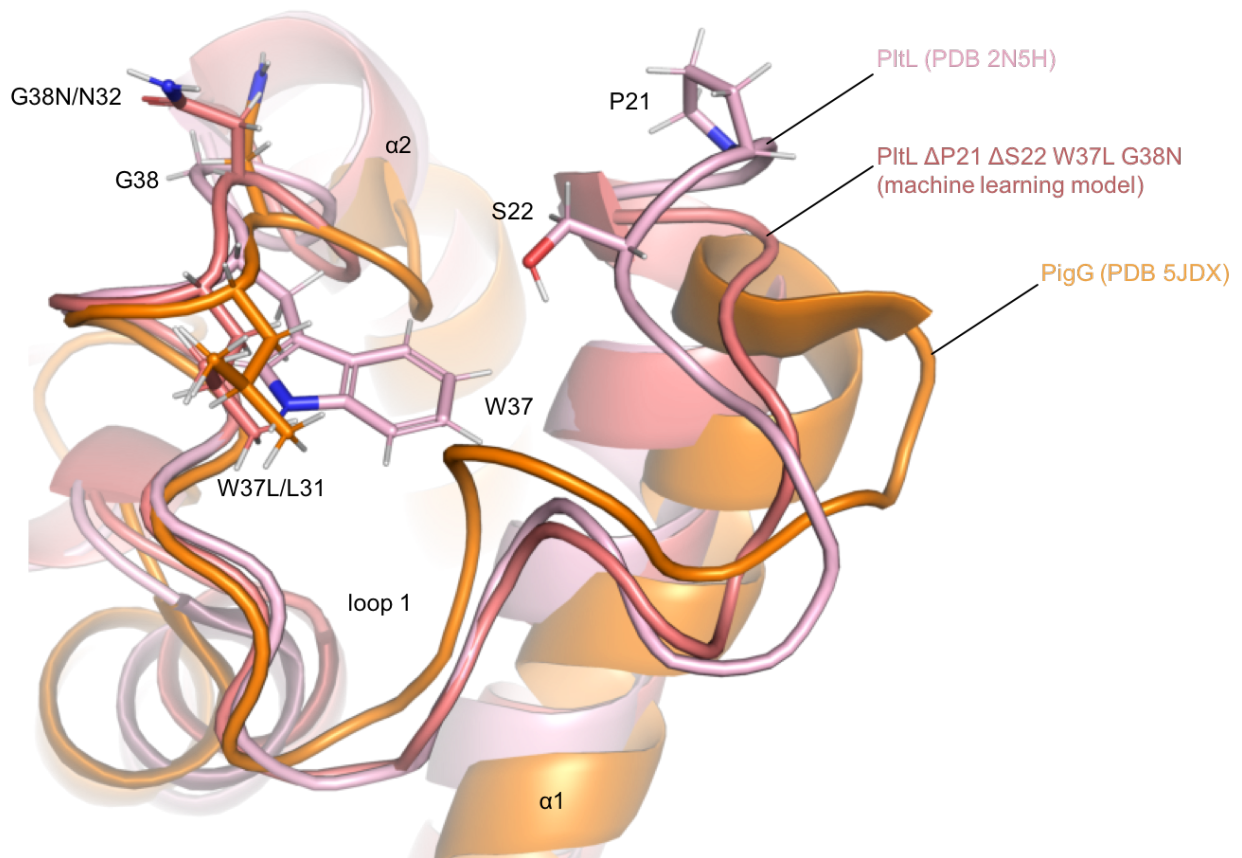
Figure 3.S11: Superposition of the loop 1 regions of PigG, PltL, and mPltL. The mPltL was prepared using the Robetta software.
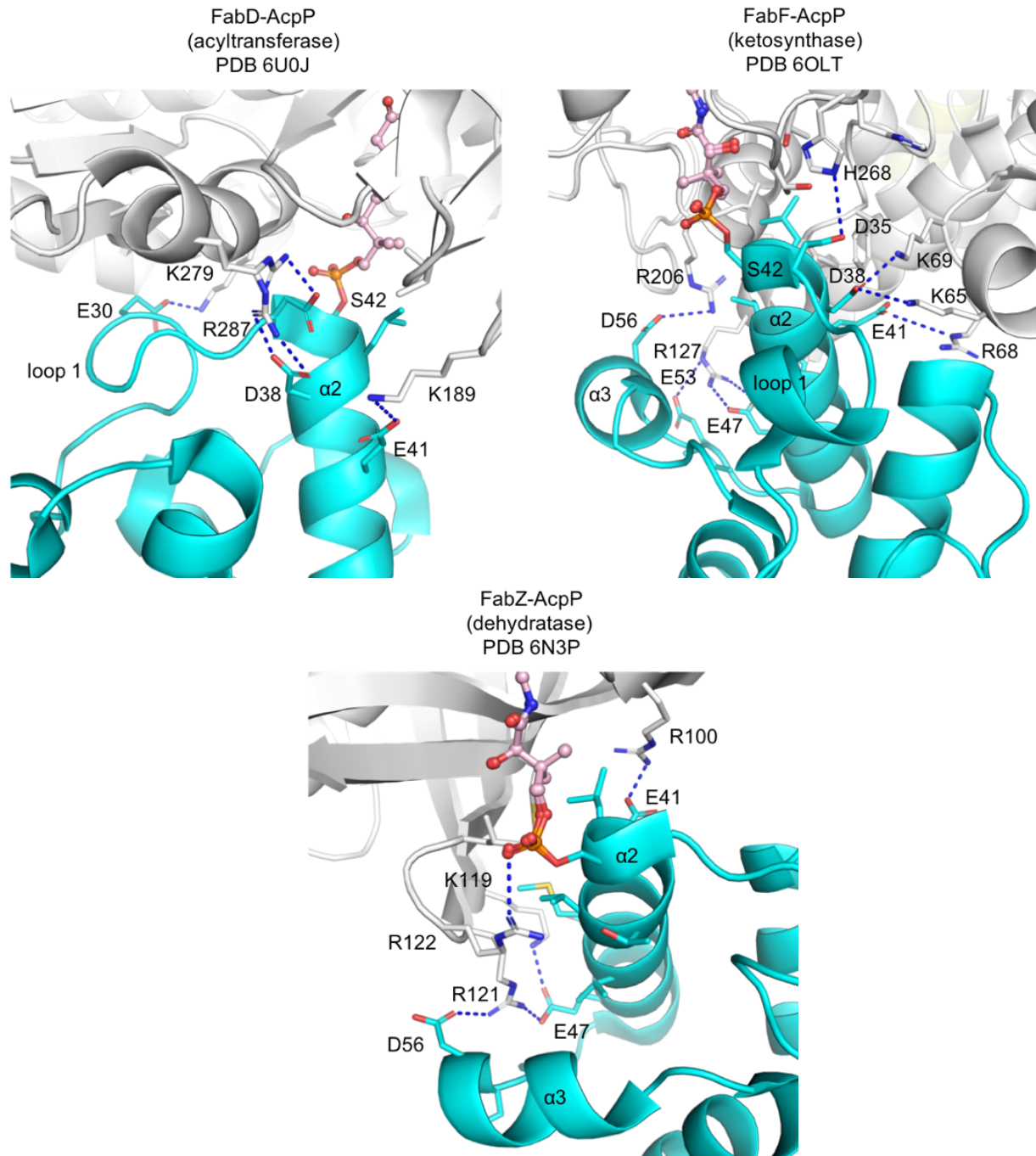
Figure 3.S12: Electrostatic interfacial interactions of E. coli AcpP with its partner proteins from the FAS.

3.11. References

(1) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83* (3), 770–803.

(2) Hur, G. H.; Vickery, C. R.; Burkart, M. D. Explorations of Catalytic Domains in Non-Ribosomal Peptide Synthetase Enzymology. *Nat. Prod. Rep.* **2012**, *29* (10), 1074.

(3) Finking, R.; Marahiel, M. A. Biosynthesis of Nonribosomal Peptides. *Annual Review of Microbiology* **2004**, *58* (1), 453–488.

(4) Brown, A. S.; Calcott, M. J.; Owen, J. G.; Ackerley, D. F. Structural, Functional and Evolutionary Perspectives on Effective Re-Engineering of Non-Ribosomal Peptide Synthetase Assembly Lines. *Nat. Prod. Rep.* **2018**, *35* (11), 1210–1228.

(5) Corpuz, J. C.; Sanlley, J. O.; Burkart, M. D. Protein-Protein Interface Analysis of the Non-Ribosomal Peptide Synthetase Peptidyl Carrier Protein and Enzymatic Domains. *Synthetic and Systems Biotechnology* **2022**, *7* (2), 677–688.

(6) Jaremko, M. J.; Davis, T. D.; Corpuz, J. C.; Burkart, M. D. Type II Non-Ribosomal Peptide Synthetase Proteins: Structure, Mechanism, and Protein–Protein Interactions. *Nat. Prod. Rep.* **2020**, *37* (3), 355–379.

(7) Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. The Phosphopantetheinyl Transferases: Catalysis of a Post-Translational Modification Crucial for Life. *Nat. Prod. Rep.* **2014**, *31* (1), 61–108.

(8) Acharya, A.; Yi, D.; Pavlova, A.; Agarwal, V.; Gumbart, J. C. Resolving the Hydride Transfer Pathway in Oxidative Conversion of Proline to Pyrrole. *Biochemistry* **2022**, *61* (3), 206–215.

(9) Jaremko, M. J.; Lee, D. J.; Patel, A.; Winslow, V.; Opella, S. J.; McCammon, J. A.; Burkart, M. D. Manipulating Protein–Protein Interactions in Nonribosomal Peptide Synthetase Type II Peptidyl Carrier Proteins. *Biochemistry* **2017**, *56* (40), 5269–5273.

(10) Corpuz, J. C.; Podust, L. M.; Davis, T. D.; Jaremko, M. J.; Burkart, M. D. Dynamic Visualization of Type II Peptidyl Carrier Protein Recognition in Pyoluteorin Biosynthesis. *RSC Chem. Biol.* **2020**, *1* (1), 8–12.

(11) Gulick, A. M. Conformational Dynamics in the Acyl-CoA Synthetases, Adenylation Domains of Non-Ribosomal Peptide Synthetases, and Firefly Luciferase. *ACS Chem. Biol.* **2009**, *4* (10), 811–827.

(12) Mayerthaler, F.; Feldberg, A.-L.; Alfermann, J.; Sun, X.; Steinchen, W.; Yang, H.; Mootz, H. D. Intermediary Conformations Linked to the Directionality of the

Aminoacylation Pathway of Nonribosomal Peptide Synthetases. *RSC Chem. Biol.* **2021**, *2* (3), 843–854.

(13) Jaremko, M. J.; Lee, D. J.; Opella, S. J.; Burkart, M. D. Structure and Substrate Sequestration in the Pyoluteorin Type II Peptidyl Carrier Protein PltL. *Journal of the American Chemical Society* **2015**, *137* (36), 11546–11549.

(14) Chen, A.; Re, R. N.; Burkart, M. D. Type II Fatty Acid and Polyketide Synthases: Deciphering Protein–Protein and Protein–Substrate Interactions. *Nat. Prod. Rep.* **2018**, *35* (10), 1029–1045.

(15) Misson, L. E.; Mindrebo, J. T.; Davis, T. D.; Patel, A.; McCammon, J. A.; Noel, J. P.; Burkart, M. D. Interfacial Plasticity Facilitates High Reaction Rate of *E. Coli* FAS Malonyl-CoA:ACP Transacylase, FabD. *Proc Natl Acad Sci USA* **2020**, *117* (39), 24224–24233.

(16) Mindrebo, J. T.; Misson, L. E.; Johnson, C.; Noel, J. P.; Burkart, M. D. Activity Mapping the Acyl Carrier Protein: Elongating Ketosynthase Interaction in Fatty Acid Biosynthesis. *Biochemistry* **2020**, *59* (38), 3626–3638.

(17) Mindrebo, J. T.; Patel, A.; Kim, W. E.; Davis, T. D.; Chen, A.; Bartholow, T. G.; La Clair, J. J.; McCammon, J. A.; Noel, J. P.; Burkart, M. D. Gating Mechanism of Elongating β-Ketoacyl-ACP Synthases. *Nat Commun* **2020**, *11* (1), 1727.

(18) Chen, A.; Re, R. N.; Burkart, M. D. Type II Fatty Acid and Polyketide Synthases: Deciphering Protein–Protein and Protein–Substrate Interactions. *Nat. Prod. Rep.* **2018**, *35* (10), 1029–1045.

(19) Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* **2018**, *118* (4), 1691–1741.

(20) Kabsch, W. *XDS. Acta Crystallogr D Biol Crystallogr* **2010**, *66* (2), 125–132.

(21) Vagin, A.; Teplyakov, A. *MOLREP* : An Automated Program for Molecular Replacement. *J Appl Crystallogr* **1997**, *30* (6), 1022–1025.

(22) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr D Biol Crystallogr* **1997**, *53* (3), 240–255.

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

(24) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *Journal of molecular graphics & modelling* **2006**, *25* (2), 247–260.

(25) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *Journal of chemical theory and computation* **2015**, *11* (8), 3696–3713.

(26) Price, D. J.; Brooks, C. L. A Modified TIP3P Water Potential for Simulation with Ewald Summation. *The Journal of chemical physics* **2004**, *121* (20), 10096–10103.

(27) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *The Journal of Physical Chemistry* **1993**, *97* (40), 10269–10280.

(28) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of chemical theory and computation* **2013**, *9* (7), 3084–3095.

(29) Rose, A. S.; Hildebrand, P. W. NGL Viewer: A Web Application for Molecular Visualization. *Nucleic acids research* **2015**, *43* (W1), W576-9.

(30) Nguyen, H.; Case, D. A.; Rose, A. S. NGLview-Interactive Molecular Graphics for Jupyter Notebooks. *Bioinformatics (Oxford, England)* **2018**, *34* (7), 1241–1242.

(31) Tiemann, J. K. S.; Guixà-González, R.; Hildebrand, P. W.; Rose, A. S. MDsrv: Viewing and Sharing Molecular Dynamics Simulations on the Web. *Nature methods* **2017**, *14* (12), 1123–1124.

CHAPTER 4. COMPUTER-AIDED ENGINEERING OF A PROTEIN-PROTEIN INTERFACE IN CARRIER PROTEIN DEPENDENT METABOLIC PATHWAYS

4.1. Introduction

The ability to design and evolve catalysis between unrelated primary and secondary metabolic players remains a long-elusive goal for the carrier protein dependent pathways that include fatty acid synthase (FAS), polyketide synthase (PKS), and non-ribosomal peptide synthetase (NRPS). Although not broadly appreciated until recently, protein-protein interactions (PPIs) have been demonstrated to be responsible for enzyme selectivity and pathway organization in these carrier protein dependent pathways.[1] Within recent years, these PPIs have been structurally characterized in solution,[1,2] crystal structures,[3–6] and through microscopy;[7] predicted in silico;[2] abrogated by mutagenesis;[1] and improved by rational design.[8] What has remained uncharted, however, is the design of PPIs between non-reactive carrier protein–enzyme pairs for gain-of-function activity. The ability to rapidly engineer enzyme selectivity has the potential to revolutionize our ability to design and control these metabolic pathways. Here we demonstrate development of a computational platform for evolving PPIs between carrier proteins and enzymes from previously incompatible pathways. We show how iterating computational docking and mutagenesis with experimental feedback can be leveraged to accelerate PPI engineering and demonstrate production of a functional unnatural product through a chimeric FAS-NRPS pathway construction.

In *Pseudomonas fluorescens,* the biosynthesis of the antibiotic, pyoluteorin, is initiated by the type II NRPS adenylation (A) domain, PltF, which activates and attaches proline to the thiol of the phosphopantetheine (PPant) of the peptidyl carrier protein

(PCP), *holo*-PltL, via an ATP dependent mechanism. (Figure 4.1A).[9] PltL transfers the prolyl group between tailoring domains for functionalization prior to transfer to a type I PKS. Recent experiments were able to deduce the PPIs between PltF and PltL through nuclear magnetic resonance (NMR) experiments and X-ray crystallography,[10,11] thus providing dynamic and high-resolution information of the PltL-PltF protein-protein interface, which can be used as a model to guide the design of new PPIs between A domains and non-cognate carrier proteins.

The acyl carrier protein (ACP), AcpP, from *Escherichia coli* fatty acid biosynthesis was identified as a promising non-cognate carrier protein in the design of a new PPI with PltF. The *E. coli* fatty acid biosynthesis is a well-studied primary metabolic pathway due to its relevance as a drug target and fuel precursor production.[12,13] AcpP is responsible for the shuttling of the thio-templated fatty acid chain to a variety of partner enzymes, which includes but is not limited to the ketosynthase for fatty acid chain extension and the ketoreductase, dehydratase, and enoyl reductase for reduction of the beta-ketone fatty acid to a fully saturated fatty acid.[14] Upon reaching a certain fatty acid chain length, the acyl chain is transferred into the production of the lipid bilayer or into the biosynthesis of secondary metabolites.[15,16] AcpP's role in shuttling substrates through the FAS cycle is critical towards production of the fatty acid product, thus making it a target towards loading unnatural substrates for extension and creation of unnatural fatty acid products.[17]

The success in chain extension of a variety of substrates prompted the idea to modify AcpP with amino acids using an NRPS A domain, specifically with PltF. To do so, a productive protein-protein interface between PltF and AcpP must be designed to allow prolylation of AcpP, which would enable biosynthesis of an unnatural ω-pyrrolidine fatty
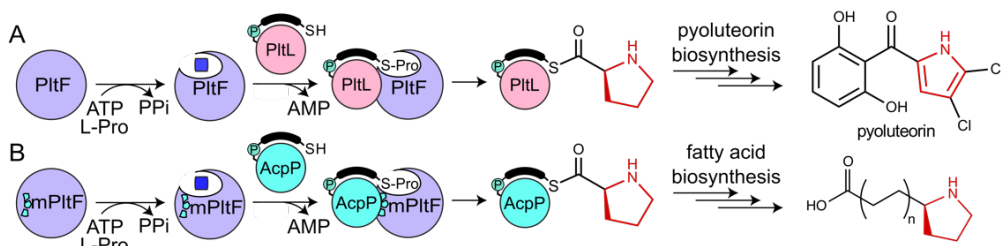
Figure 4.1: Adenylation and thiolation reactions by PltF. (A) Wild type reaction of PltF. PltF adenylates proline to create a prolyl-AMP intermediate (blue square), then transfers the prolyl moiety to *holo*-PltL. The pyrrolidine group is passed off and incorporated into pyoluteorin. (B) Designed PltF reaction. The interface mutant PltF (mPltF) instead transfers ATP-activated proline to *holo*-AcpP. Prolyl-AcpP then shuttles the pyrrolidine group through the *E. coli* FAS for carbon chain elongation of the product. Depicted is a pyrrolidine fatty acid, a potential product of this hybrid metabolic pathway.

acid upon chain elongation of the prolyl-AcpP by the *E. coli* FAS (Figure 4.1B). These functionalized fatty acids may be utilized as polymer precursors or further functionalized as potential therapeutic candidates with scaffolds similar to lanneaquinol and anaephene B.[18–21]

In this work, a computational workflow was developed using Rosetta to design new protein-protein interactions using the PltL-PltF crystal structure as a starting structure. The AcpP structure was aligned to PltL to mimic a binding interaction for AcpP and PltF, then subsequent *in silico* mutagenesis provided a library of PltF mutants that were predicted to form a protein-protein interface with AcpP. The library of PltF mutants was prepared using site directed mutagenesis and each mutant was assayed for prolylation activity with AcpP using an HPLC based assay. After three rounds of iterative mutagenesis with Rosetta, the mutant with the highest activity *in vitro* achieved a 182-fold increase in turnover rate for AcpP prolylation compared to wildtype PltF. Refining the default Rosetta score function using data from our experimental assay improved design success rate, including a design with 1690-fold increased turnover rate. Analysis of a model of this most active PltF mutant revealed the utilization of 3 positively charged residues that can electrostatically complement the negatively charged AcpP. Overall, our computational interface design methodology successfully identified evolved PltF mutants for activity with AcpP within a few rounds of iterative mutagenesis.

4.2. Iterative computational and experimental design of PltF-AcpP interface

Increasing evidence has demonstrated the role of protein-protein interactions (PPIs) for controlling substrate specificity in carrier protein dependent pathways.[1,8] We
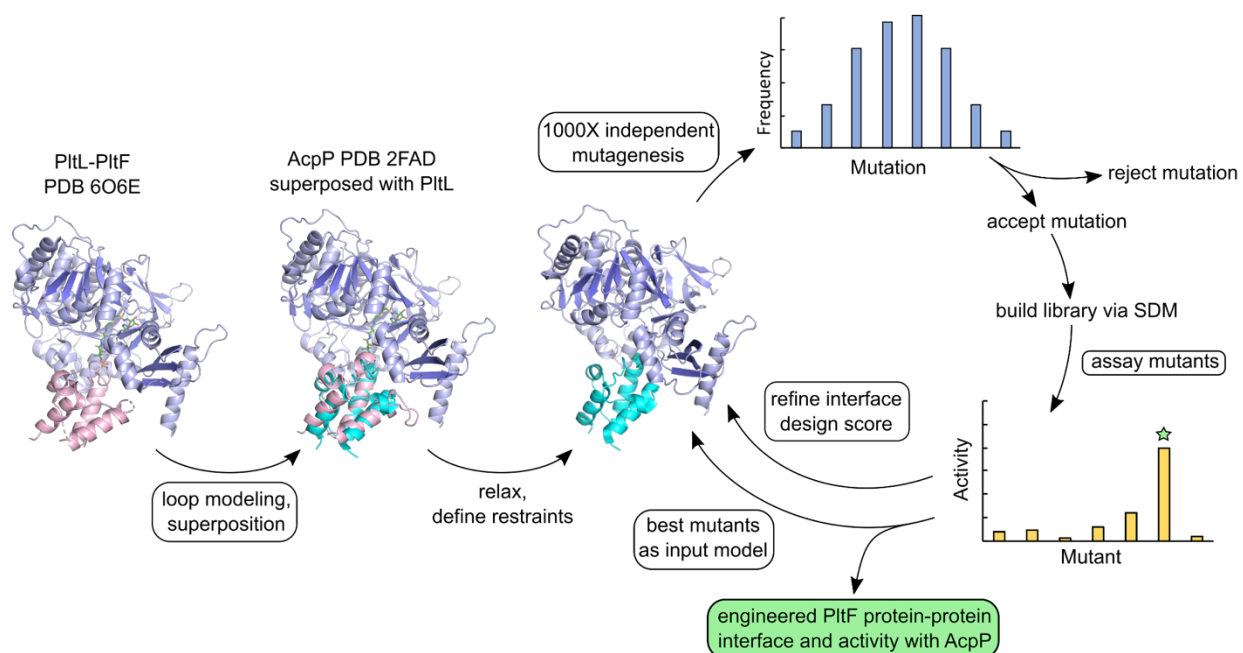
Figure 4.2: Schematic of computational and experimental interface design workflow. An initial interface with PltF (purple) is generated through carrier protein superposition of AcpP (cyan) onto PltL (pink). Then, the structures are prepared by modeling missing loops, and performing relaxation. Residue or design restraints are selected followed by random mutagenesis (red) of the interface to produce 1000 mutants which pass the score threshold. Mutants which significantly persist throughout the mutagenesis were selected for *in vitro* assay. Information from the experimental assay is then used to refine the Rosetta score function.

therefore sought to design a new PPI between the non-cognate pairs, PltF and AcpP, to facilitate a new activity for PltF by attaching proline to *holo*-AcpP. We leveraged information from the recently characterized, cognate PltF-PltL interface resolved by X-ray crystallography (PDB ID 6O6E).[22] AcpP shares 27% sequence identify with PltL, and aligns with 1.8 Å RMSD (Figure 4.S1). To avoid deleterious pathway effects, we chose to focus mutations to PltF rather than AcpP.

The RosettaScripts interface[23] of the Rosetta Software Suite was used to determine initial PltF mutations to form a more favorable PltF-AcpP interface, using a protocol which iterates through 1,000 cycles of docking, design, minimization, and scoring with the REF15 score function[24] (Figure 4.2). Only PltF residues within 8 Å of AcpP were re-designed, and PltF residues K402 and K486 were prevented from re-design due to importance for PltF catalysis.[22] An additional bonus score was applied to favor native residues to avoid excess mutations. Six mutations significantly persisted throughout all cycles and were assayed for activity as point mutations in PltF: L235A, S253A, M257Q, N436Y, Q438R, and K457I. Two mutants showed increased activity compared to the wild-type baseline of 5% conversion, with N436Y prolylating 8% of the AcpP, and Q438R prolylating 38% of the AcpP (Figure 4.3).

Encouraged by these initial results, we tested whether combining these initial mutants would produce an additive or synergistic effect. Double and triple mutants did not show increased prolylation activity (Figure 4.S2). Next, we tested whether subjecting single mutants through a subsequent round of *in silico* mutations would result in newer mutants with increased activity. Surprisingly, the majority of new mutants identified by Rosetta were unique compared to the first round of mutagenesis (Table 4.S1). All of the

second round mutants identified for PltF N436Y showed decreased activity compared to the 8% single mutant (Figure 4.3, Table 4.S1). Three of the second-round mutants for Q438R showed similar or increased activity compared to the 38% single mutant, with the Q438R K472R and Q438R D263H double mutants prolylating > 60% AcpP (Figure 4.3, Table 4.S1). A third round with each of these double mutants identified the triple mutant Q438R K472R D263N, which prolylated 83% AcpP.

Application of the *in silico* Rosetta PPI design protocol in an iterative fashion identified a triple mutant of PltF with significantly increased activity compared to the wild type. However, testing each of the designs *in vitro* is a laborious process, and many of the mutants did not show any activity. Furthermore, the sequence profile identified by Rosetta was not correlated with mutant activity. Improvement of computational score functions to predict PPIs *in vitro* remains an ongoing challenge.[24,25] Therefore, we sought to refine the Rosetta score function, such that the PPI design protocol would be more predictive of desired activity. Examination of the individual score terms (Data 4.S1) revealed that the electrostatic score was more predictive of activity than the interface score as suggested by the initial Q438R mutation. AcpP is a negatively charged protein, and many native partners have been found to contain a basic binding patch, so electrostatic interactions were anticipated to play an important role.[26]

4.3. Refinement of Rosetta score function for improved PPI prediction

Several electrostatic coefficient weights were tested with the design protocol for the wild type PltF-AcpP interface. The standard weight for the electrostatic term is 0.875, but we found that increasing the weight to 1.500 gave the best prediction of *in vitro* activity
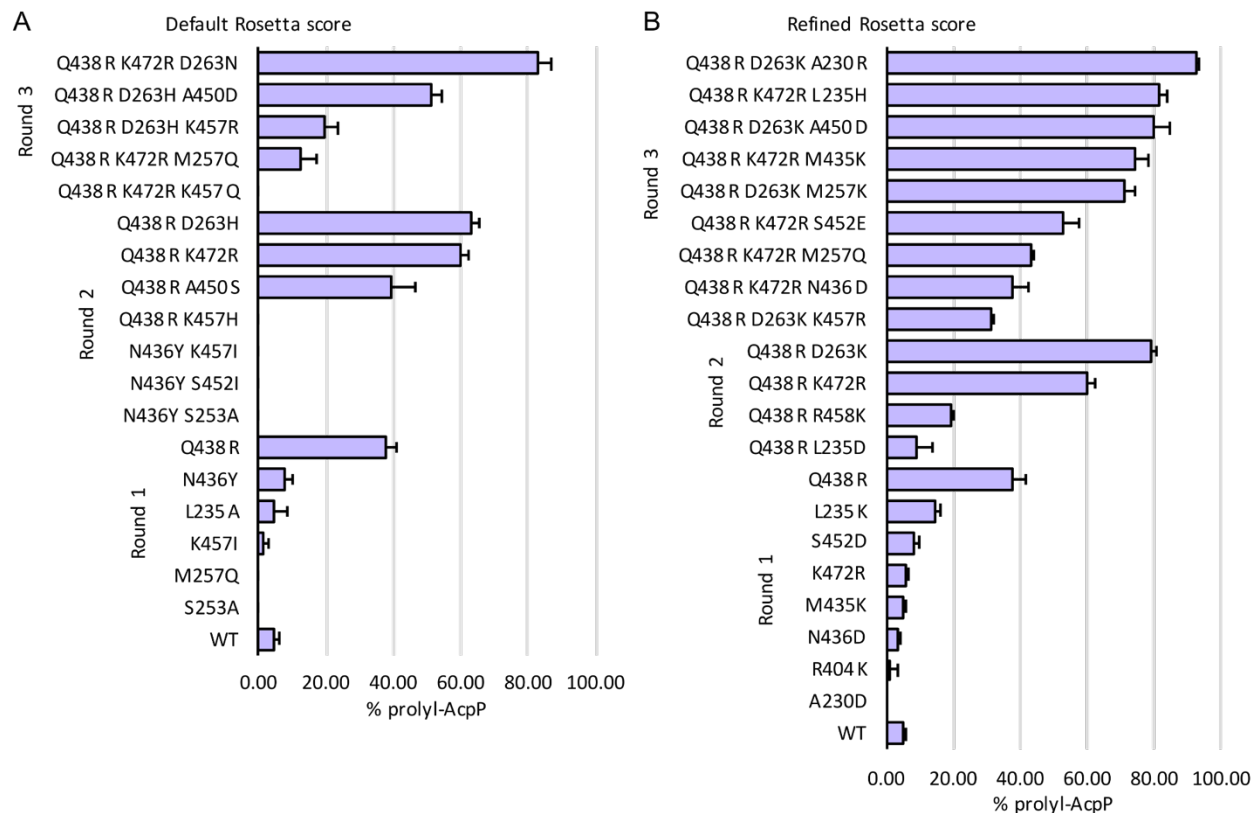
Figure 4.3: Prolylation activity of designed PltF mutants. PltF activity was monitored through HPLC chromatogram peak integration after incubation of 0.010 uM mPltF with 0.015 uM holo-AcpP, 5 mM ATP, 12.5 mM MgCl$_2$, and 5mM L-proline for 2 hours at 25° C. The (A) default or (B) refined score function was used to identify mutants through iterative rounds of Rosetta interface design.

(Table 4.S2). We next repeated the computational protocol for three rounds, and found a remarkable increase in successful mutant predictability (Figure 4.3). Many fewer false positives (predicted mutants with no *in vitro* activity) and greater true positive results were generated. By the third round, all designs prolylated > 30%, five prolylated > 70% and the best design, Q438R D263K A230R, prolylated 92% of the *holo*-AcpP (Figure 4.3).

To better quantify and compare the mutant activities, the turnover rates of select mutants were determined. The best triple mutant, Q438R D263K A230R, is able to convert 18.5 AcpP hr^{-1} (Table 4.S4), compared to 0.81 and 0.87 AcpP hr^{-1} for the best double mutants, Q438R D263K and Q438R K472R, and 0.26 AcpP hr^{-1} for the single Q438R mutant. Over three rounds of mutagenesis from the wild-type PltF to the Default Rosetta Score triple mutant, Q438R K472R D263N, the activity was improved 182-fold. Remarkably, the Refined Rosetta Score triple mutant, Q438R D263K A230R, exhibited an improved activity by 1690-fold relative to the wild-type enzyme. This improvement in activity, however, is still less than the cognate PltF-PltL reaction, where PltF can process about 4,000 PltL hr^{-1}.[27]

4.4. Model of the designed Pltf-AcpP interface

The cognate PltF-PltL interface was found to be dependent on hydrophobic interactions, where the X-ray crystal structure revealed minimal hydrogen bonding interactions at the interface, which includes the sidechain amine of PltF K457 with PltL G38 main chain carbonyl.[22] Interestingly, the initial superposition of AcpP onto PltL and relaxation with the Rosetta score function revealed salt bridges between PltF R234, R404, and R458 with AcpP E57, D35, and E21, respectively (Figure 4.4A). The AcpP-PltF
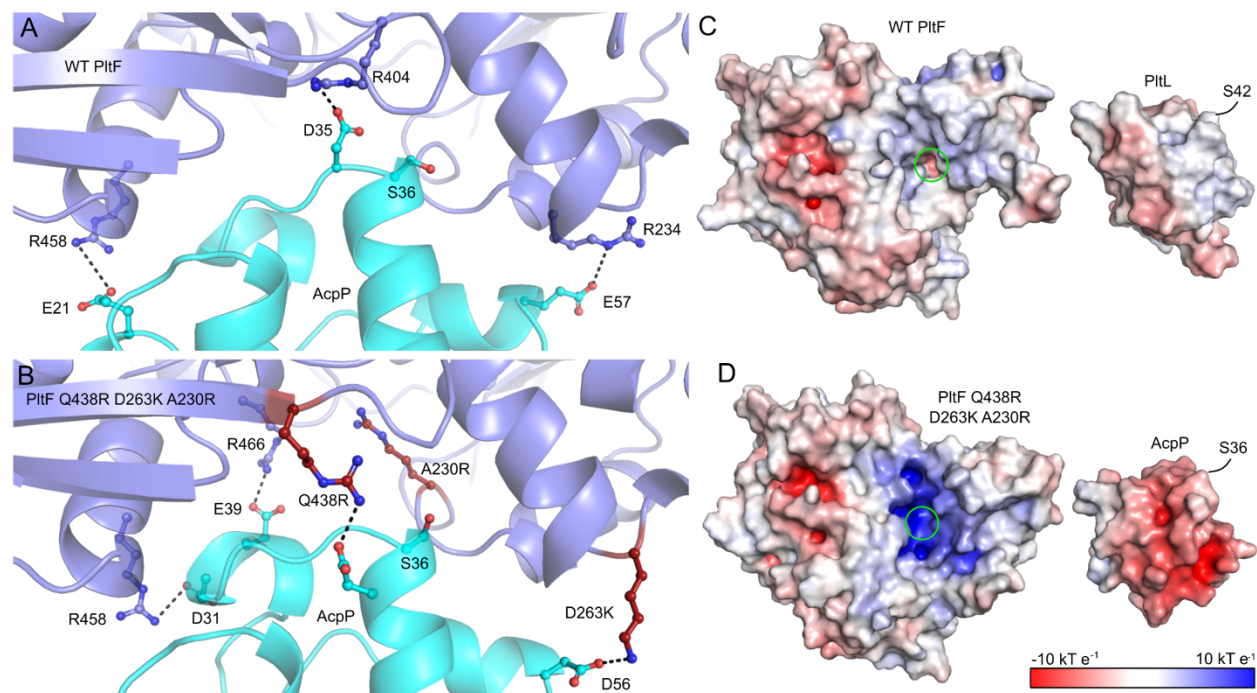
Figure 4.4: Comparison of wild-type and designed protein-protein interfaces. The predicted salt bridge interactions at the protein-protein interface of the (A) relaxed model of WT PltF (purple) docked to AcpP (cyan) and (B) the Rosetta model of PltF Q438R D263K A230R (purple, red mutations) bound to AcpP (cyan). Electrostatic potentials mapped onto the surface of (C) PltL (PDB ID 2N5H), PltF (PDB ID 6O6E), (D) AcpP (PDB ID 2FAD), and a Rosetta model of PltF Q438R D263K A230R. Circled in green is the entrance to the PPant tunnel.

Q438R D263K A230R model reveals additional putative salt bridges between PltF Q438R and D263K with AcpP D38 and D56, respectively (Figure 4.4B). With the refined Rosetta score, three design rounds introduced three positively charged residues while replacing a negatively charged residue, which increased the positive electrostatic potential at the entrance to the PPant tunnel compared to the slightly positive interface of the WT PltF (Figure 4.4C-D, Figure 4.S4). The PltF A230R mutation is in a position that can form a long-range electrostatic interaction with the PPant phosphate (Figure 4.S5).[28] These three synergistic mutations show that the designs generated with the refined Rosetta score successfully created an electrostatic-dependent interface from a hydrophobic interface to complement and attract the negatively charged surface of AcpP. The addition of electrostatic interface interactions resembles the PPIs of native AcpP FAS partner enzymes revealed from previously solved X-ray crystal structures, supporting the protocol's ability to design productive protein-protein interfaces.[14]

4.5. Discussion

The ability to re-engineer carrier protein-dependent pathways to create custom products has been a long sought-after goal. Here we develop an experimentally guided computational workflow, which incorporates both structure and function data to design mutations more efficiently than the current state of the art Rosetta REF2015 score function.[24] We engineered a hybrid FAS/NRPS pathway through interface mutations of the A domain from a NRPS to interact with the ACP from a FAS. While the default Rosetta protocol generated only 3 mutants with > 50% activity and 7 false positives, refining the score function based on the experimental assay resulted in 8 mutants with > 50% activity

and only 1 false positive. Additionally, the most active mutant identified with the refined score function was almost 10 times faster than the mutant identified with the default score function. This method advances the current state of interface engineering and provides a promising strategy for harnessing control over carrier protein-dependent pathways for designing new natural products.

4.6. Acknowledgements

Chapter 4, entitled "Computer-aided engineering of a protein-protein interface in carrier protein-dependent metabolic pathways", in full, is a reprint of the material as it appears in: Sztain, T.*; Corpuz, J. C.*; Bartholow, T.; LaClair, J. J.; Jiang Z.; Sanlley, J.; Heberlig, G. W.; McCammon, J. A.; Burkart, M. D. Computer-aided engineering of a protein-protein interface in carrier protein-dependent metabolic pathways. *In Preparation.* **2022**. The dissertation author was the primary co-author of this paper along with Dr. Terra Sztain.

4.7. Methods

4.7.A. Computational methods

Structure preparation. The crystal structure of the trapped PltF-PltL complex PDB: 6O6E[22] was used to obtain initial coordinates for PltF. AcpP from PDB: 2FAD[29] was aligned to the PltL in PyMOL to generate an initial PltF-AcpP complex (Figure 4.S1). PltL was then deleted and not used. Rosetta version 3.9 was used for all computational designs with the ref2015 score function. The rotamer packing flags -ex1 and -ex2 were used in all cases to include additional $\chi 1$ and $\chi 2$ rotamers. First, PltF residues 449-450

which are not resolved in the crystal structure were modeled using Rosetta loop modeling. Next the structure was relaxed using the FastRelax protocol.

Interface design. A RosettaScripts design protocol was used to generate 1,000 structures per design round. The PltF residues K402 and K486 were restricted to repacking and not designed due to their importance for catalytic activity.[22] Only residues of PltF within 8 Å of AcpP were allowed to be mutated. A favored native residue bonus of 1 was applied to avoid excessive mutations. For 1,000 iterations, the protocol first performs a local refinement stage of full atom docking, followed by interface design, then full atom backbone and side chain minimization. Acceptance criteria for each iteration required a negative binding score and an interface solvent accessible surface area greater than 800 Å^2.

Analysis and score refinement. Designs were analyzed based on sequence profile (Table 4.S1) Mutations persisting in over 40% of the structures were prepared for *in vitro* analysis.

4.7.B. Biochemical methods

PltF Mutagenesis. The PltF point mutations were introduced via QuikChange PCR.[30] The primer sequences are listed in Table 4.S3.

Protein Growth and Expression. Wild-type PltF or the mutant PltF plasmids were transformed into *E. coli* BL21 cells. The cells were grown in 50 mL LB media with 50

µg/mL kanamycin at 37 °C. Expression was induced with 0.5 mM IPTG at an OD600 of 0.6. The cells were grown at 16 °C for 16 hours then harvested at 4000 relative centrifugal force (RCF). The cells were resuspended in 1.4 mL of 50 mM NaCl, 50 mM Tris pH 7.5, and 1X BugBuster Protein Extraction Reagent (Millipore Sigma, 70921-3) and incubated at room temperature, 500 rpm. The lysate was centrifuged at 12000 RCF at 12 C for 45 minutes and the supernatant was passed through 0.15 mL bed volume of Novagen Ni-NTA resin. The column was then washed with 2 mL of 50 mM NaCl, 50 mM Tris pH 7.50 and then 2 mL of the same buffer with 20 mM imidazole. PltF was eluted with 0.25 mL of 50 mM NaCl, 50 mM Tris pH 7.5, and 250 mM imidazole, then dialyzed using Slide-A-Lyzer Dialysis Casettes (Pierce, 66383) overnight into 50 mM NaCl, 50 mM Tris pH 7.5, 10% glycerol at 4 C.

holo-AcpP was prepared by co-transforming C-terminal His-tagged AcpP from *E. coli* and sfp from *Bacillus subtilis* into *E. coli* BL21 cells. The cells were grown in LB broth with 50 µg/mL kanamycin and 100 µg/mL ampicillin at 37 °C and induced with 0.5 mM IPTG at an OD600 of 0.6. The cells were then grown for 4 hours at 37 °C then harvested at 2500 relative centrifugal force (RCF). The cells were resuspended in 50 mM NaCl, 50 mM Tris pH 7.5 and lysed using a Ultrasonic Processor FS-600N (4 seconds off, 1 second on for 10 minutes) on ice. The lysate was centrifuged at 12000 RCF at 4 °C for 45 minutes and the supernatant was passed through 2 mL bed volume of Novagen Ni-NTA resin. The column was washed with 50 mL of buffer and AcpP was eluted as described for the PltF purification. Passing of the nickel pure AcpP sample through an Ascentis C18 column on an Agilent 1100 HPLC using an gradient of increasing acetonitrile, starting 25% at acetonitrile and 75% water and increasing until 95% acetonitrile, revealed 90% *holo-*

AcpP, with the remaining 10% being apo-AcpP. All solvents for HPLC contained 0.05% trifluoroacetic acid. Full holofication was achieved by incubating the apo/holo mixture with 0.004 mM sfp, 50 mM Tris pH 7.5, 12.5 mM MgCl2, 1 mM TCEP, 0.5 mM CoA, and 150 mM NaCl overnight at 37 °C . The reaction was dialyzed into 50 mM Tris pH 7.50 and run through a HiTrapQ HP 5 mL (Cytiva, 17115401) column for pure *holo*-AcpP.

PltF activity assays. Initial mPltF activity assays were performed by incubating 0.010 mM mPltF, 0.015 mM *holo*-AcpP, 50 mM Tris pH 7.50, 12.5 mM MgCl2, 2 mM TCEP, 5 mM ATP, and 5 mM L-Proline for 2 hours at 25 °C. The reaction was quenched with 2% formic acid and centrifuged for 10 minutes, 10k rpm, at 4 °C. The supernatant was passed through an Ascentis C18 column on an Agilent 1100 HPLC, starting with 25% acetonitrile for 2 minutes, then increasing to 56% acetonitrile over 13 minutes, then increasing to 95% acetonitrile over 3 minutes, and finally decreased to 25% acetonitrile over 5 minutes. Absorbance at 210 nm was used for integration of the *holo*-AcpP and prolyl-AcpP peaks, which eluted at 8.40 and 8.10 minutes respectively. Percent prolyl-AcpP was calculated by dividing the integrated prolyl-AcpP peak by the summation of the integrated prolyl- and *holo*-AcpP peaks. Each assay was performed in triplicate.

Time course experiments of mPltF were performed by incubating 1.01 uM mPltF with 45 uM *holo*-AcpP, 50 mM Tris pH 7.50, 12.5 mM MgCl2, 2 mM TCEP, 5 mM ATP, and 5 mM L-Proline for 90 minutes, where an aliquot of each reaction was quenched with 2% formic acid at a 15 minute interval. The quenched reactions were analyzed through HPLC as described previously. The linear portion of the time course experiments were used to calculate the turnover rate. Each time-course experiment was performed in

duplicate. The exception is the WT PltF activity for AcpP, where the turnover rate was estimated from the initial activity assay described above due to very low prolylation activity.

4.7.C. Interface analysis

Electrostatic potential calculation. The ABPS program[31] in PyMOL was used to calculate the electrostatic surface potentials of the relaxed protein model outputs from the RosettaScripts design using the default parameters.

Protein-protein interface analysis. Protein Interfaces, Surfaces, and Assemblies (PISA)[32] from the European Bioinformatics Institute was used to identify the residues involved in the AcpP-PltF and AcpP-PltF Q438R D263K A230R interface models.

PyMOL mutagenesis. The Mutagenesis Wizard in PyMOL was used to mutate, fit, and visualize the A230R mutation. The rotamer was chosen based on a low clash score and proximity to the PPant arm.

4.8. Supplementary Figures

Table 4.S1: Summary of computationally predicted mutants

0.875		1.5	
Round 1		**Round 1**	
mutation	sequence profile	mutation	sequence profile
L235A	0.74	A230K	0.16
S253A	1.00	A230D	0.72
M257Q	0.80	L235K	1.00
N436Y	1.00	S253A	0.13
Q438R	0.42	V254K	0.73
K457I	1.00	V254E	0.03
Round 2 Q438R		V254D	0.13
mutation	sequence profile	M257Q	0.99
L235D	0.21	M257E	0.01
D263H	1.00	M435K	0.50
A450S	1.00	N436Y	0.50
K457H	0.83	N436D	0.50
K457W	0.17	Q438R	1.00
K472R	1.00	**Round 2 Q438R**	
Round 2 N436Y		mutation	sequence profile
mutation	sequence profile	L235D	1.00
F231Y	0.04	D263K	1.00
R234K	0.17	A540S	0.95
S253A	1.00	A450E	0.05
S452I	1.00	K457H	0.80
I454V	0.07	K457W	0.20
K457I	1.00	R438K	1.00
R466W	0.96	K472R	1.00
Round 3 Q438R+D263H		**Round 3 Q438R+D263K**	
mutation	sequence profile	mutation	sequence profile
A450D	1.00		
K456R	1.00		
Round 3 Q438R+K457H			
mutation	sequence profile		

Table 4.S2: Comparison of predicted mutants from various electrostatic weight coefficients

		electrostatic weight						
		0.875	0.9	0.95	1	1.5	2	2.5
mutation	% prolyl AcpP	sequence profile						
Q438R	39%	0.42	0.14	0.23	0.45	1.00	0.84	0.57
L235K	14%	0.00	0.00	0.00	0.00	1.00	1.00	1.00
N436Y	12%	1.00	1.00	1.00	1.00	0.50	0.04	0.01
S452D	8%	0.00	0.00	0.00	0.00	0.00	1.00	1.00
K472R	6%	0.00	0.00	0.00	0.00	0.00	0.80	0.50
M257Q	6%	0.80	0.60	0.71	0.80	0.99	0.46	0.27
M435K	5%	0.00	0.00	0.00	0.00	0.50	0.96	0.96
L235A	4%	0.74	0.71	0.73	0.73	0.00	0.00	0.00
N436D	3%	0.00	0.00	0.00	0.00	0.50	0.96	0.96
K457I	2%	1.00	1.00	1.00	1.00	0.00	0.00	0.00
S253A	1%	1.00	1.00	1.00	1.00	0.13	0.00	0.00
R404K	1%	0.00	0.00	0.00	0.00	0.00	0.70	0.57
A230D	0%	0.00	0.00	0.00	0.00	0.72	0.67	0.71

Table 4.S3: PCR Primers for site directed mutagenesis of PltF

| Mutant	Oligo Sequence	Mutant	Oligo Sequence
M257Q_F	TCATGCAGATGATGCAAGACGGCGACTTGC	Q438R_R458K_F	CTCAAGAAGCACTGCGCCCAGCG
M257Q_R	ATCATCTGCATGAGTACCGAGGGCACCGA	Q438R_R458K_R	AGTGCTTCTTGAGGTCGATCAGCGAAAGAGCGTC
K457I_F	CCTCATTCGCCACTGCGCCCAGCG	Q438R_D263H_A450D_F	GACGACCTTTCGCTGATCGACCTCAAGCGCC
K457I_R	GGCGAATGAGGTCGATCAGCGAAAGAGCGTC	Q438R_D263H_A450D_R	CGAAAGGTCGTCCCGGGTGTGAGCGAC
N436Y_F	GATGTATGCGCAGTTGTGGGGCGTGCTG	A230D_F	ATCGATTTCTCGCCGCGGCTG
N436Y_R	TGCGCATACATCCCCTCCCCGACCACC	A230D_R	AGAAATCGATCGACTCGGGAACCA
S253A_F	CCCGCAGTACTCATGATGATGATGCAAGACGGC	L235K_F	CCGCGGAAACTGACCGACTTCATCC
S253A_R	GTACTGCGGGCACCGAGTACCAGATGCTGATTTC	L235K_R	CAGTTTCCGCGGCGAGAACG
Q438R_F	GCGCGTTTGTGGGGCGTGCTGGTC	V254K_F	CCTCGAAACTCATGATGATGATGCAAGACGGCGA
Q438R_R	ACAAACGCGCGTTCATCCCCTCCCCGAC	V254K_R	TGAGTTTCGAGGGCACCGAGTACCAGATGCTGATTT
L235A_F	CGGGCACTGACCGACTTCATCCGCCAGC	R404K_F	GGTGAAAGGCAACCGGATCGAG
L235A_R	TCAGTGCCCGCGGCGAGAACGCGATC	R404K_R	GCCTTTCACCTTGAGCATGTCGTCAC
Q438R_D263H_F	ACGGCCACTTGCTCAGCGATA	M435K_F	GGGAAAAACGCGCAGTTGTGG
Q438R_D263H_R	AGTGGCCGTCTTGCATCATCATC	M435K_R	GCGTTTTTCCCCTCCCCGACC
Q438R_A450S_F	GGACAGTCTTTCGCTGATCGAC	N436D_F	GATGGATGCGCAGTTGTGGGGC
Q438R_A450S_R	AAGACTGTCCCGGGTGTGAGCGA	N436D_R	CGCATCCATCCCCTCCCCGACCA
Q438R_K457H_F	ACCTCCATCGCCACTGCGCCCAG	S452D_F	CAGATCAAGAGCGTCCCGGGTGTG
Q438R_K457H_R	GCGATGGAGGTCGATCAGCGAAAGA	S452D_R	CTCTTGATCTGATCGACCTCAAGC
Q438R_K472R_F	CGACCGTGTGCTGTGCCTGGAC	Q438R_D263K_A230R_F	CGATCCGTTTCTCGCCGCGGCTG
Q438R_K472R_R	GCACACGGTCGATGATCATGTAGCG	Q438R_D263K_A230R_R	AGAAACGGATCGACTCGGGAACCAGGCAC
N436Y_S452I_F	CTCTTATTCTGATCGACCTCAAGCG	Q438R_D263K_M257K_F	CTCATGAAAATGATGCAAGACGGCAAATTGCTCAGC
N436Y_S452I_R	CAGAATAAGAGCGTCCCGGGTGTGA	Q438R_D263K_M257K_R	CATCATTTTCATGAGTACCGAGGGCACCGAGTAC
Q438R_D263H_K457R_F	CTCCGTCGCCACTGCGCCCAGCG	Q438R_K472R_L235H_F	GCGGCATCTGACCGACTTCATCCGCCAG
Q438R_D263H_K457R_R	GCGACGGAGGTCGATCAGCGAAAGAGCGT	Q438R_K472R_L235H_R	GTCAGATGCCGCGGCGAGAACG
Q438R_K472R_D263N_F	CGGCAATTTGCTCAGCGATATCC	Q438R_K472R_M435K_F	GGGGAAAAACGCGCGTTTGTGGGGCG
Q438R_K472R_D263N_R	CAAATTGCCGTCTTGCATCATCATCATGA	Q438R_K472R_M435K_R	CGTTTTTCCCCTCCCCGACCACCAGC
Q438R_D263H_K457Q_F	CTCCAACGCCACTGCGCCCAGCG	Q438R_K472R_N436D_F	GATGGATGCGCGTTTGTGGGGCGTGCTG
Q438R_D263H_K457Q_R	GGCGTTGGAGGTCGATCAGCGAAAGAGCGTC	Q438R_K472R_N436D_R	GCGCATCCATCCCCTCCCCGACCACC
Q438R_D263K_F	CGGCAAATTGCTCAGCGATATCCAGGACACC	Q438R_K472R_S452E_F	CTCTTGAACTGATCGACCTCAAGCGCCACTG
Q438R_D263K_R	GCAATTTGCCGTCTTGCATCATCATCATGAGTACCGAG	Q438R_K472R_S452E_R	GATCAGTTCAAGAGCGTCCCGGGTGTGAG
Q438R_L235D_F	GCGGGATCTGACCGACTTCATCCGCCAG		
Q438R_L235D_R	CAGATCCCGCGGCGAGAACGCGATCGA		

149

**Table 4.S4:** Turnover rate of select PltF mutants with AcpP

| Mutagenesis Round	PltF Mutant	Turnover (prolyl-AcpP/hr)
-	WT PltF	0.0109 ± 0.0022
Initial/Refined Rosetta score round 1	Q438R	0.264 ± 0.001
Initial Rosetta score round 2	Q438R K472R	0.874 ± 0.026
Initial Rosetta score round 3	Q438R K472R D263N	1.98 ± 0.10
Refined Rosetta score round 2	Q438R D263K	0.806 ± 0.169
Refined Rosetta score round 3	Q438R K472R L235H	1.20 ± 0.26
Refined Rosetta score round 3	Q438R D263K A450D	1.22 ± 0.06
Refined Rosetta score round 3	Q438R D263K A230R	18.5 ± 0.4
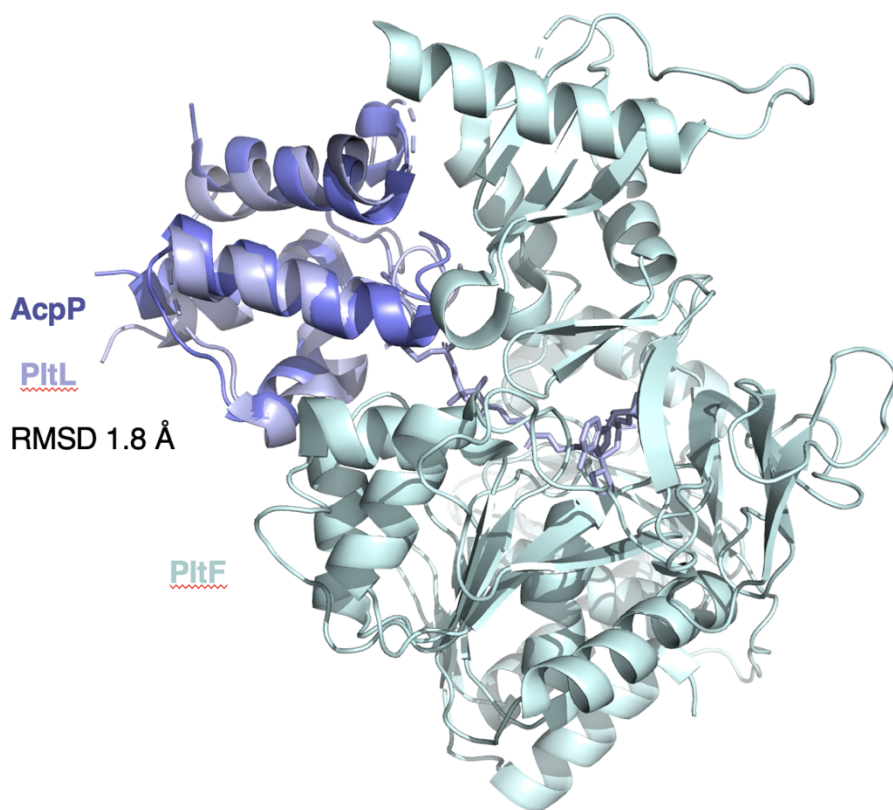
Figure 4.S1: Alignment of AcpP to PltL-PltF crystal structure. AcpP from PDB ID: 2FAD aligned in PyMOL with PltL from PltL-PltF structure from PDB ID: 6O6E.
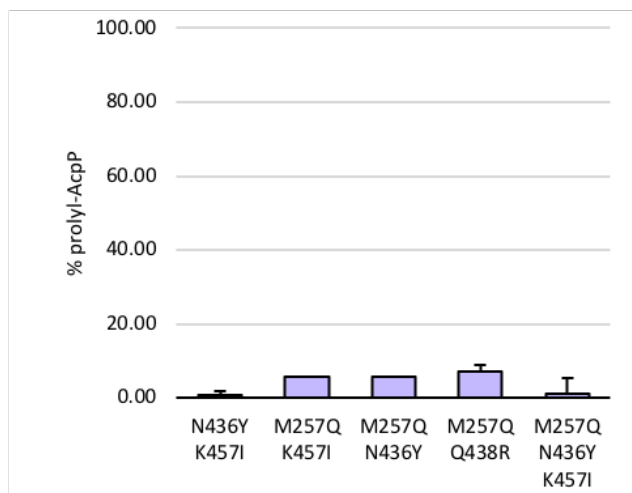
**Figure 4.S2:** Prolylation activity of combined round 1 PltF mutants.
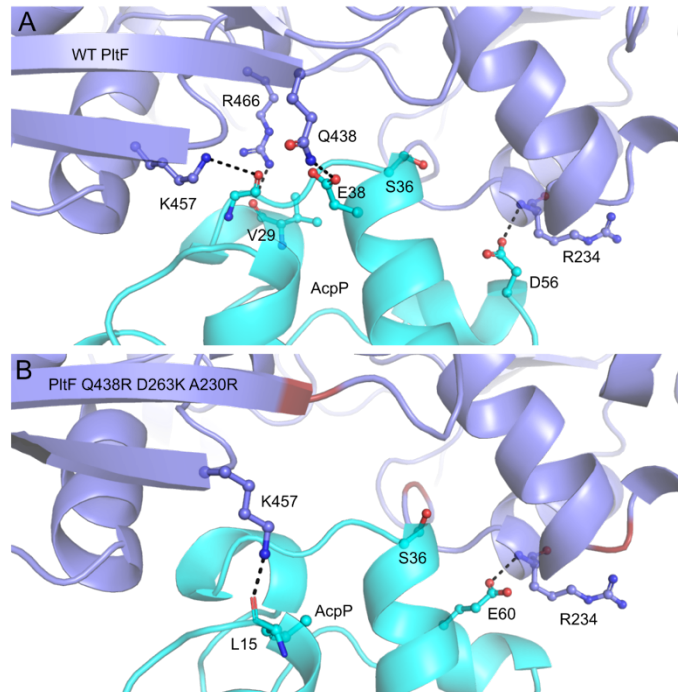
Figure 4.S3: Hydrogen bonding interactions at the protein-protein interface of (A) AcpP-PltF and (B) AcpP-PltF Q438R D263K A230R models.
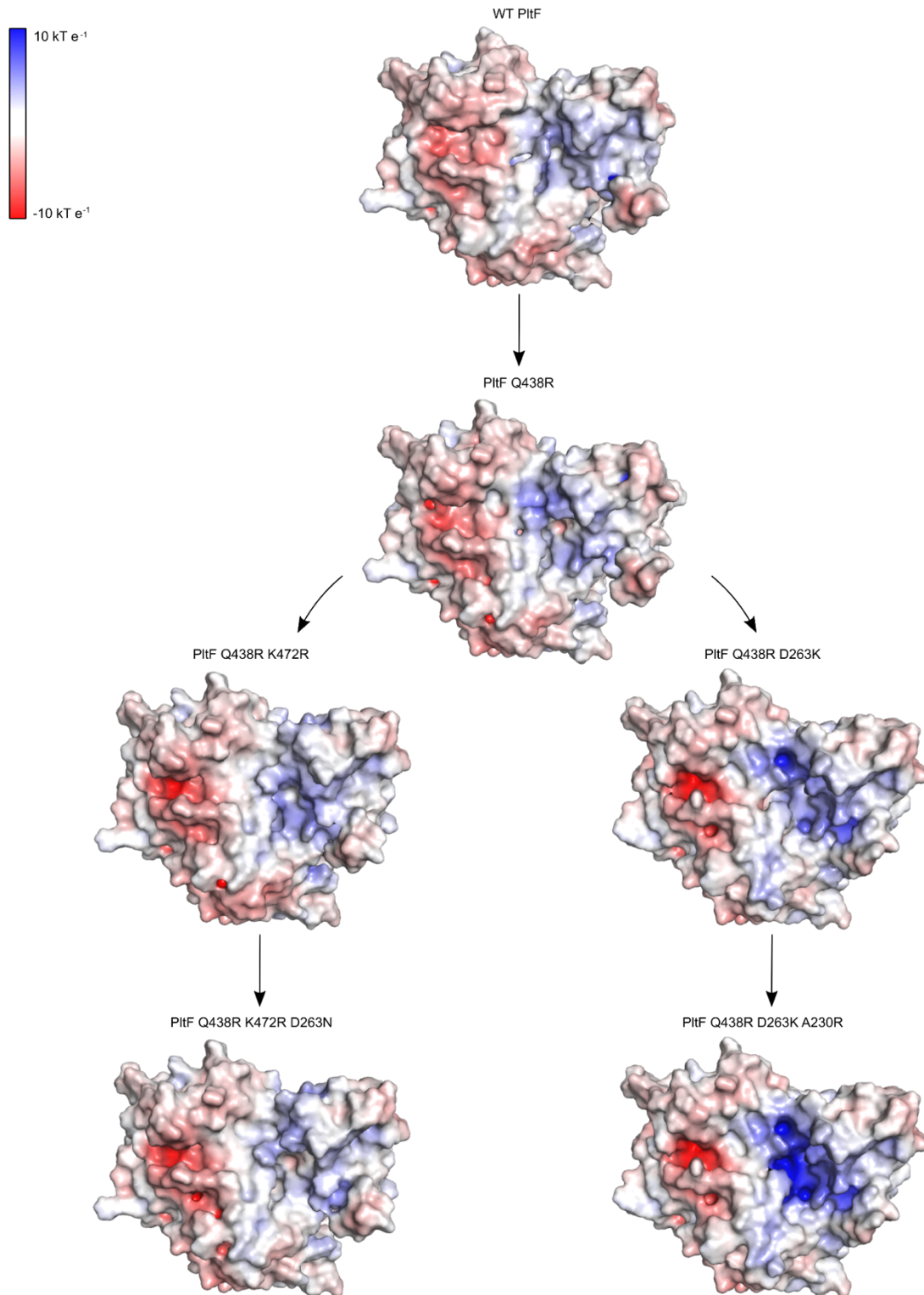
Figure 4.S4: ABPS electrostatic surface calculations using models of Rosetta mutants. The increase in size and positive potential of the interface on PltF can be seen through additional mutations from each round of design, which correlates with increased PltF activity towards AcpP.
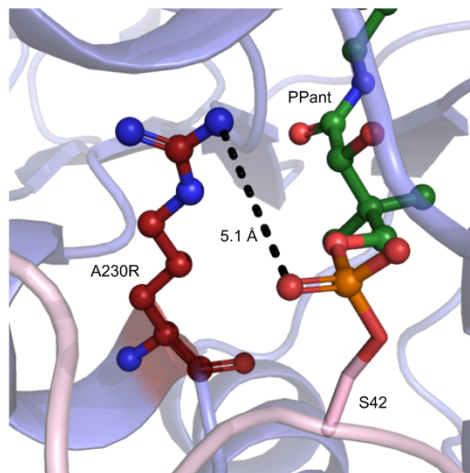
Figure 4.S5: PltF A230R interaction with the PPant phosphate. The A230R mutation (red) was visualized in PyMOL on crystal structure of PltF (purple) bound to PltL (pink) (PDB ID 6O6E) to show proximity of the guanidine group to the PPant phosphate moiety.
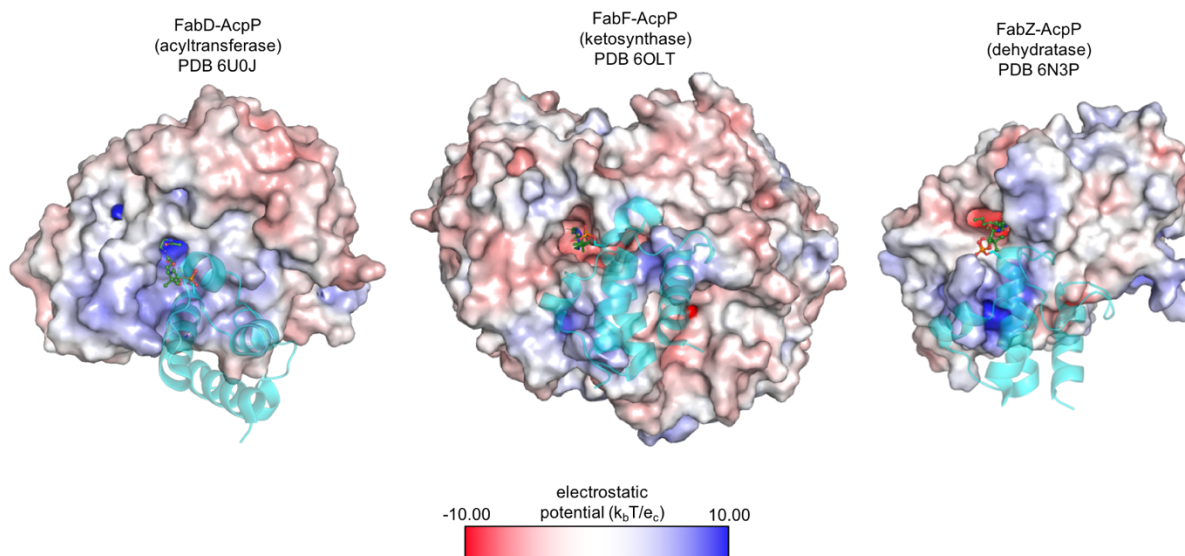
Figure 4.S6: Electrostatic surface visualization of the FAS partner proteins of AcpP.

4.9. References

(1) Sztain, T.; Bartholow, T. G.; Lee, D. J.; Casalino, L.; Mitchell, A.; Young, M. A.; Wang, J.; McCammon, J. A.; Burkart, M. D. Decoding Allosteric Regulation by the Acyl Carrier Protein. *Proc Natl Acad Sci USA* **2021**, *118* (16), e2025597118.

(2) Bartholow, T. G.; Sztain, T.; Patel, A.; Lee, D. J.; Young, M. A.; Abagyan, R.; Burkart, M. D. Elucidation of Transient Protein-Protein Interactions within Carrier Protein-Dependent Biosynthesis. *Commun Biol* **2021**, *4* (1), 340.

(3) Nguyen, C.; Haushalter, R. W.; Lee, D. J.; Markwick, P. R. L.; Bruegger, J.; Caldara-Festin, G.; Finzel, K.; Jackson, D. R.; Ishikawa, F.; O'Dowd, B.; McCammon, J. A.; Opella, S. J.; Tsai, S.-C.; Burkart, M. D. Trapping the Dynamic Acyl Carrier Protein in Fatty Acid Biosynthesis. *Nature* **2014**, *505* (7483), 427–431.

(4) Milligan, J. C.; Lee, D. J.; Jackson, D. R.; Schaub, A. J.; Beld, J.; Barajas, J. F.; Hale, J. J.; Luo, R.; Burkart, M. D.; Tsai, S.-C. Molecular Basis for Interactions between an Acyl Carrier Protein and a Ketosynthase. *Nat Chem Biol* **2019**, *15* (7), 669–671.

(5) Dodge, G. J.; Patel, A.; Jaremko, K. L.; McCammon, J. A.; Smith, J. L.; Burkart, M. D. Structural and Dynamical Rationale for Fatty Acid Unsaturation in *Escherichia Coli*. *Proc Natl Acad Sci USA* **2019**, *116* (14), 6775–6783.

(6) Mindrebo, J. T.; Patel, A.; Kim, W. E.; Davis, T. D.; Chen, A.; Bartholow, T. G.; La Clair, J. J.; McCammon, J. A.; Noel, J. P.; Burkart, M. D. Gating Mechanism of Elongating β-Ketoacyl-ACP Synthases. *Nat Commun* **2020**, *11* (1), 1727.

(7) Lou, J. W.; Iyer, K. R.; Hasan, S. M. N.; Cowen, L. E.; Mazhab-Jafari, M. T. Electron Cryomicroscopy Observation of Acyl Carrier Protein Translocation in Type I Fungal Fatty Acid Synthase. *Sci Rep* **2019**, *9* (1), 12987.

(8) Sarria, S.; Bartholow, T. G.; Verga, A.; Burkart, M. D.; Peralta-Yahya, P. Matching Protein Interfaces for Improved Medium-Chain Fatty Acid Production. *ACS Synth. Biol.* **2018**, *7* (5), 1179–1187.

(9) Jaremko, M. J.; Davis, T. D.; Corpuz, J. C.; Burkart, M. D. Type II Non-Ribosomal Peptide Synthetase Proteins: Structure, Mechanism, and Protein–Protein Interactions. *Nat. Prod. Rep.* **2020**, *37* (3), 355–379.

(10) Jaremko, M. J.; Lee, D. J.; Patel, A.; Winslow, V.; Opella, S. J.; McCammon, J. A.; Burkart, M. D. Manipulating Protein–Protein Interactions in Nonribosomal Peptide Synthetase Type II Peptidyl Carrier Proteins. *Biochemistry* **2017**, *56* (40), 5269–5273.

(11) Corpuz, J. C.; Podust, L. M.; Davis, T. D.; Jaremko, M. J.; Burkart, M. D. Dynamic Visualization of Type II Peptidyl Carrier Protein Recognition in Pyoluteorin Biosynthesis. *RSC Chem. Biol.* **2020**, *1* (1), 8–12.

(12) Campbell, J. W.; Cronan, J. E. Bacterial Fatty Acid Biosynthesis: Targets for Antibacterial Drug Discovery. *Annu. Rev. Microbiol.* **2001**, *55* (1), 305–332.

(13) Janßen, H.; Steinbüchel, A. Fatty Acid Synthesis in Escherichia Coli and Its Applications towards the Production of Fatty Acid Based Biofuels. *Biotechnol Biofuels* **2014**, *7* (1), 7.

(14) Chen, A.; Re, R. N.; Burkart, M. D. Type II Fatty Acid and Polyketide Synthases: Deciphering Protein–Protein and Protein–Substrate Interactions. *Nat. Prod. Rep.* **2018**, *35* (10), 1029–1045.

(15) Magnuson, K.; Jackowski, S.; Rock, C. O.; Cronan, J. E. Regulation of Fatty Acid Biosynthesis in Escherichia Coli. *MICROBIOL. REV.* **1993**, *57*, 21.
(16) Lin, S.; Cronan, J. E. Closing in on Complete Pathways of Biotin Biosynthesis. *Mol. BioSyst.* **2011**, *7* (6), 1811.

(17) Beld, J.; Finzel, K.; Burkart, M. D. Versatility of Acyl-Acyl Carrier Protein Synthetases. *Chemistry & Biology* **2014**, *21* (10), 1293–1299.

(18) Metzger, J. O. Fats and Oils as Renewable Feedstock for Chemistry. *Eur. J. Lipid Sci. Technol.* **2009**, *111* (9), 865–876.

(19) Biermann, U.; Bornscheuer, U.; Meier, M. A. R.; Metzger, J. O.; Schäfer, H. J. Oils and Fats as Renewable Raw Materials in Chemistry. *Angew. Chem. Int. Ed.* **2011**, *50* (17), 3854–3871.

(20) Groweiss, A.; Cardellina, J. H.; Pannell, L. K.; Uyakul, D.; Kashman, Y.; Boyd, M. R. Novel Cytotoxic, Alkylated Hydroquinones from *Lannea Welwitschii*. *J. Nat. Prod.* **1997**, *60* (2), 116–121.

(21) Brumley, D.; Spencer, K. A.; Gunasekera, S. P.; Sauvage, T.; Biggs, J.; Paul, V. J.; Luesch, H. Isolation and Characterization of Anaephenes A–C, Alkylphenols from a Filamentous Cyanobacterium (*Hormoscilla* Sp., Oscillatoriales). *J. Nat. Prod.* **2018**, *81* (12), 2716–2721.

(22) Corpuz, J. C.; Podust, L. M.; Davis, T. D.; Jaremko, M. J.; Burkart, M. D. Dynamic Visualization of Type II Peptidyl Carrier Protein Recognition in Pyoluteorin Biosynthesis. *RSC Chem. Biol.* **2020**, *1* (1), 8–12.

(23) Fleishman, S. J.; Leaver-Fay, A.; Corn, J. E.; Strauch, E.-M.; Khare, S. D.; Koga, N.; Ashworth, J.; Murphy, P.; Richter, F.; Lemmon, G.; Meiler, J.; Baker, D. RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLoS ONE* **2011**, *6* (6), e20161.

(24) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park,

H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. J. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13* (6), 3031–3048.

(25) Stranges, P. B.; Kuhlman, B. A Comparison of Successful and Failed Protein Interface Designs Highlights the Challenges of Designing Buried Hydrogen Bonds. *Protein Science* **2013**, *22* (1), 74–82.

(26) Byers, D. M.; Gong, H. Acyl Carrier Protein: Structure–Function Relationships in a Conserved Multifunctional Protein Family. *Biochem. Cell Biol.* **2007**, *85* (6), 649–662.

(27) Thomas, M. G.; Burkart, M. D.; Walsh, C. T. Conversion of L-Proline to Pyrrolyl-2-Carboxyl-S-PCP during Undecylprodigiosin and Pyoluteorin Biosynthesis. *Chemistry & Biology* **2002**, *9* (2), 171–184.

(28) Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* **2018**, *118* (4), 1691–1741.

(29) Roujeinikova, A.; Simon, W. J.; Gilroy, J.; Rice, D. W.; Rafferty, J. B.; Slabas, A. R. Structural Studies of Fatty Acyl-(Acyl Carrier Protein) Thioesters Reveal a Hydrophobic Binding Cavity That Can Expand to Fit Longer Substrates. *Journal of Molecular Biology* **2007**, *365* (1), 135–145.

(30) Liu, H.; Naismith, J. H. An Efficient One-Step Site-Directed Deletion, Insertion, Single and Multiple-Site Plasmid Mutagenesis Protocol. *BMC Biotechnol* **2008**, *8* (1), 91.

(31) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS Biomolecular Solvation Software Suite. *Protein Science* **2018**, *27* (1), 112–128.

(32) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* **2007**, *372* (3), 774–797.